Title
Designing custom organelles in bacteria

Author
Jakobson, Christopher Matthew

Publication Date
2016

Peer reviewed|Thesis/dissertation

Designing custom organelles in bacteria

By

Christopher Matthew Jakobson

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

In

Chemical Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:
Professor Danielle Tullman-Ercek, Chair
Professor Wenjun Zhang
Professor Michi Taga

Summer 2016

Abstract

Designing custom organelles in bacteria

by

Christopher Matthew Jakobson

Doctor of Philosophy in Chemical Engineering

University of California, Berkeley

Professor Danielle Tullman-Ercek, Chair

Despite the vast array of enzymatic activities available to metabolic engineers, many biosyntheses are impractical due to the production of toxic intermediates and off-target products, poor kinetic performance, and cofactor imbalances. We aim to address these roadblocks by organizing biosynthetic pathways inside custom organelles within cells. Bacterial microcompartments, porous proteinaceous organelles found in *Salmonella* and other bacteria, provide an ideal starting point for the engineering of subcellular bacterial organelles. To this end, we develop tools for the control of microcompartment formation, enzyme encapsulation, and small molecule transport across the compartment shell. Of particular interest is the encapsulation of enzymatic pathways that synthesize fuels, pharmaceuticals, and other products which are challenging to produce in the cytosol.

In order to tune the catalytic activity of the microcompartments, we orchestrate microcompartment formation at the transcriptional level to control when the organelles are formed and how much heterologous cargo is encapsulated. We also create new localization signals that allow fine control over the loading of multiple enzymes to the compartments simultaneously. We are investigating the role of pore residues in small molecule transport across the shells of virus-like particles and the microcompartments themselves. Lastly, we use computational tools to model microcompartment function in order to facilitate the selection of pathways for encapsulation. Together, these tools allow the creation of custom biosynthetic organelles with user-defined enzyme content and transport properties, enabling the use of biosynthetic pathways which otherwise fail to function.

For
Grandma & Grandpa
and
Nan & Grandad
Who laid the foundation for where I am today.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

MCP, microcompartment
Pdu, 1,2-propanediol utilization
Eut, ethanolamine utlilization
Grp, glycyl radical-generating protein
Etu, ethanol utilization
TEM, transmission electron microscopy
SEC, size exclusion chromatography
WT, wild type
VLP, virus-like particle
GFP, green fluorescent protein
1,2-PD, 1,2-propanediol
EA, ethanolamine
PhoA, alkaline phosphatase
LacZ, beta-galactosidase
Ald, aldehyde dehydrogenase
MBP, maltose binding protein

Acknowledgements

This work was generously supported by the National Science Foundation (MCB1150567), the Army Research Office (W911NF-15-1-0144), a gift from the ExxonMobil Corporation, and a U.C. Berkeley Berkeley Fellowship.

Many people have contributed to the work described herein: undergraduate students Jacob Cota, Vanessa Mackley, Ashley Kermah, Chuchu Zhang, and Yiqun Chen; graduate rotational students Alex Chien and Jorge Marchand; and graduate students Eddy Kim, Mary Slininger, Jeff Glasgow, Michael Asensio, Norma Morella, and Emily Hartman.

I owe a tremendous debt to everyone who mentored me in the Tullman-Ercek Lab, including Mike Fisher, Eddy Kim, Kevin Metcalf, Sergey Boyarskiy, Stephanie Davis-Lopez, and Anum Azam. I am also grateful to Niall Mangan for mentorship and guidance in the computational aspects of this work.

I would also like to thank Stanley Herrmann, Allie Landry, David Liu, Peter Lobaccaro, Michele Sarazen, and Kevin Wujcik for fruitful discussions, support throughout qualifying exam preparation, and critical reading of manuscripts.

My research advisor, Danielle Tullman-Ercek, gave me the scientific freedom to pursue my interests, the help and guidance to shape my ideas into scientific investigations, and a group of colleagues full of scientific curiosity. I will always be grateful for an amazing graduate school experience.

Above all, I would like to thank my family, particularly Kath, Paul, and Nicholas Jakobson and my partner Leah Sibener, for their unwavering support.

Chapter 1: Introduction

1.1 Discovery and characterization

Bacterial microcompartments (MCPs) are subcellular organelles bound by a proteinaceous membrane consisting of a monolayer of protein pentamers and hexamers. Bacterial MCPs are a promising target for engineering because they offer the potential to avoid several common roadblocks to enzymatic biosynthesis *in vivo*: toxicity of intermediate compounds; the generation of off-target products; imbalances in cofactor usage and abundance; and poor kinetic performance. Here, we summarize the current hypotheses on the native function of MCPs and their role in pathogenesis, discuss efforts on several fronts to engineer MCPs for biosynthesis, and present a roadmap of the remaining challenges in realizing custom-designed subcellular bacterial organelles.

The ability of *Salmonella enterica* to metabolize 1,2-propanediol was found in 1999 to be linked to the formation of polyhedral organelles [1], which were subsequently purified and found to contain 8 structural proteins and 13 proteins of enzymatic or unknown function [2]. Similar organelles were subsequently found which confer the ability to metabolize ethanolamine [3]. The 1,2-propanediol utilization (Pdu) and ethanolamine utilization (Eut) MCPs of *Salmonella* are roughly icosahedral structures of diameter 100-200 nm, though irregular in shape and size. MCPs homologues have since been found to be widespread throughout bacteria [4], and are likely to be diverse in their metabolic function. These organelles are an existing biological structure which can serve as the basis for engineering efforts directed towards creating subcellular compartments in bacteria, analogous to the membrane-bound organelles that are ubiquitous in eukarya.

1.2 Native function of MCPs

The native function of bacterial MCPs in bacteria is the subject of several hypotheses which are not mutually exclusive. These include the following: (1) the MCP sequesters toxic metabolic intermediates in the MCP lumen, protecting the cellular milieu; (2) the MCP contains a private pool of cofactors for the exclusive use of the encapsulated pathway; (3) the organization of enzymes in the MCP enhances flux through the encapsulated pathway through kinetic enhancement; (4) the MCP prevents the loss of carbon by sequestering a metabolic intermediate that would otherwise escape. There is experimental evidence for several of these hypotheses in the Pdu and Eut MCP systems, and it is possible that the function of the MCPs encompasses some or all of these mechanisms.

Experiments using a PduA shell protein mutant defective for MCP assembly provide evidence that sequestration of the toxic intermediate propionaldehyde plays a role in Pdu MCP function: an MCP-defective strain showed delayed growth on 1,2-PD and an increased mutation rate by the Ames test [5]. A genetic study of the Eut MCP system, on the other hand, suggested that a private pool of cofactors (NAD+/NADH and coenzyme A) is recycled internally to the MCP [6]. It has also been demonstrated in many scaffolded enzyme systems that arranging enzymes in proximity to one another can increase the flux through the scaffolded pathway [7–9]. Lastly, it has been proposed that a volatile intermediate of ethanolamine metabolism, acetaldehyde, may be lost from the cell if not contained within an MCP. It is unclear whether related MCP structures such as the Pdu and Eut MCPs share a mode of action, or have evolved to exploit different benefits of encapsulation. It does seem likely, however, that they have evolved distinct transport properties (Slininger, unpublished results).

Understanding the native function of the Pdu and Eut MCPs will inform the selection of heterologous pathway candidates for encapsulation. Ideally, those pathways will be chosen for encapsulation which are hypothesized to benefit from the natively observed effects of encapsulation.

1.3 Phylogeny and role in pathogenicity

Recent genomic studies have suggested that MCPs are widespread among bacteria, with more than 20% of available genome sequences containing ORFs encoding MCP-protein homologues [4]. At least 23 different types of MCP loci have been proposed, with putative MCPs occurring in the genomes of human and livestock pathogens such as *Salmonella enterica*, *Listeria monocytogenes*, *Yersinia enterocolitica*, and *Shigella flexneri* [10]. Only a subset of these MCPs has been structurally or metabolically characterized; the remainder are of unknown metabolic function. MCPs are known which metabolize 1,2-propanediol [1], ethanolamine [3], and L-fucose and L-rhamnose [11]; each of these metabolic pathways proceeds through an aldehyde intermediate. It is possible that the presence of a toxic intermediate such as an aldehyde is a conserved feature of bacterial MCPs.

The precise role of MCPs in bacterial pathogenesis in the host gut remains enigmatic, but there is evidence that Eut MCPs contribute to *S. enterica* proliferation in food products, in a *C. elegans* model of infection, and in the gut of mice [12,13]. It is proposed that gut inflammation induced by *S. enterica* causes the production of tetrathionate, which is in turn used as a terminal electron acceptor which allows anaerobic growth on ethanolamine and 1,2-PD [13]. This metabolic faculty thus allows the invading pathogen to avoid direct nutritional competition with the resident gut microbiota.

The following is adapted from Jakobson and Tullman-Ercek; PLoS Pathogens, 2016 with permission.

Upon arriving in the host gut, enteric pathogens face a daunting challenge: to proliferate in an environment already rich in commensal microbes and poor in available nutrients. Recent evidence suggests that *Salmonella* and other enteric bacteria conduct a coordinated assault employing two complimentary systems: bacterial microcompartments and the type III secretion system. While a portion of invading bacteria construct subcellular metabolic organelles designed to utilize unique nutrients, the remaining invading cells induce intestinal inflammation, remodeling the chemical environment of the gut to render it more favorable to *Salmonella* proliferation (Figure 1.1).

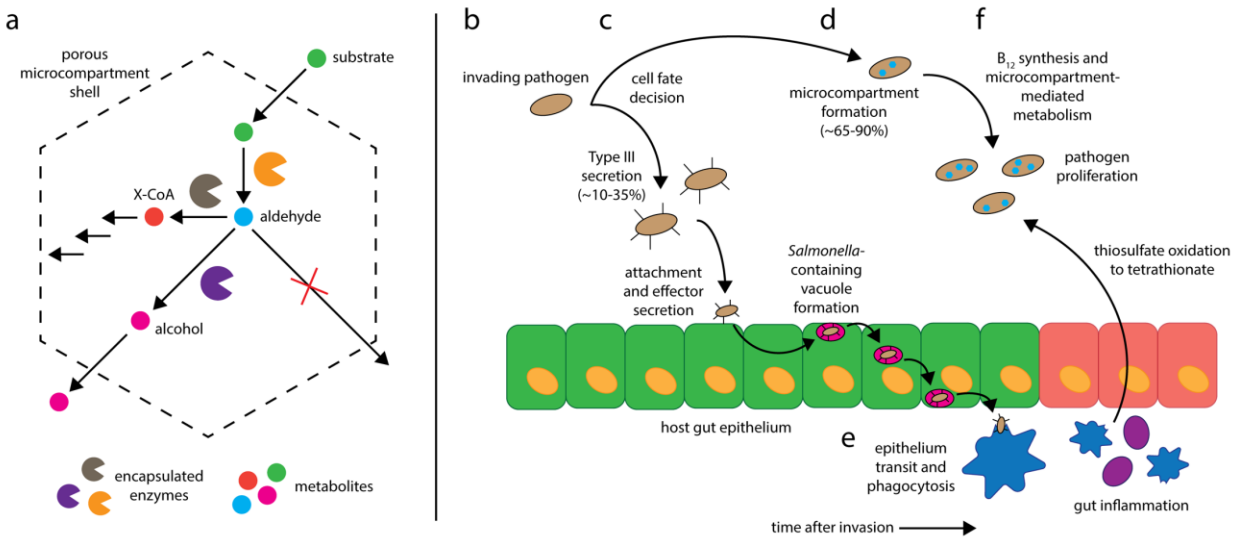


Figure 1.1: Bacterial microcompartment function and the coordinated invasion of the host gut by *Salmonella enterica*. (a): A substrate molecule enters the microcompartment and is converted to an aldehyde species, which is trapped in the microcompartment shell before being converted either to an alcohol or to a Coenzyme A-conjugated species [14]. (b) The invading pathogen population enters the gut; (c) each pathogen cell undergoes a fate decision between type III secretion (~10-35% of cells) and microcompartment formation (~65-90% of cells); (d) type III secretion-competent cells invade the host epithelium while microcompartment-competent cells form microcompartments and synthesize vitamin B_{12}; (e) type III-secretion competent cells traverse the epithelium and undergo phagocytosis in the lamina propria and (f) gut inflammation causes thiosulfate oxidation to tetrathionate, allowing microcompartment-mediated metabolism and pathogen proliferation.

What are bacterial microcompartments and what are they for?

Despite the received wisdom that eukarya possess intracellular organelles and bacteria do not, bacteria do use organelles called bacterial microcompartments to spatially segregate metabolism. Rather than a phospholipid membrane, however, these organelles are bound by a porous protein monolayer made up of trimeric, pentameric, and hexameric shell proteins. A suite of metabolic enzymes, including those required for cofactor regeneration, are encapsulated in the microcompartment lumen. The mechanism of microcompartment assembly remains elusive, but it is known that some enzymes are localized to the microcompartment through interactions with the inner face of the microcompartment shell [15,16]. Many pathogens possess microcompartments, including *Salmonella enterica*, *Escherichia coli*, *Listeria monocytogenes*, *Yersinia enterocolitica*, and *Shigella flexneri*, and microcompartment genes have been found in as many as 20% of sequenced bacterial genomes [4,10]. Bacteria are known to use a variety of microcompartment systems to metabolize compounds such as 1,2-propanediol [1], ethanolamine [3], and L-fucose and L-rhamnose [11]. All of these metabolic pathways proceed through toxic aldehyde intermediates, and it is proposed that the microcompartment shell functions to protect the rest of the bacterial cell contents from these toxic compounds, as well as to sequester a private pool of the requisite cofactor molecules [5]. Compartmentalized metabolic processes may impart a competitive advantage to invading pathogens over the existing gut microbiota, which

typically lack microcompartment operons and thus are unable to utilize the substrates metabolized in the microcompartments [17,18]. For example, *S. enterica* subsp. e*nterica* serovar Typhimurium dedicates approximately 2% of its genome to 1,2-propanediol and ethanolamine metabolism and the synthesis of associated cofactors, suggesting that these processes confer a significant competitive advantage at some point in the pathogen's life cycle [19,20].

Microcompartment systems, such as the ethanolamine utilization microcompartment, enhance *E. coli* and *S. enterica* proliferation in diverse settings, including: in food products; in a *C. elegans* model of infection; during growth on bovine intestinal content; and in the gut of a mouse model of *Salmonella* infection [12,13,17]. This indicates that there are many circumstances in which microcompartments may provide a competitive advantage to pathogens. Evidence suggests that the gut microbiota plays a central role in preventing host colonization by pathogens, possibly by sequestering critical nutrients [21]. In a model of *S.* Typhimurium infection of the mouse gut, for example, antibiotic treatment to reduce the abundance of native gut microbes renders the host more susceptible to *Salmonella* infection [22], and mice with a compromised microbiota clear *S.* Typhymurium from the gut much less effectively following nonfatal infection [23]. Gaining a unique metabolic capacity may help pathogens sidestep microbiotic defense mechanisms by creating a new nutritional niche in the host gut, but microcompartment-mediated metabolism also requires a unique micronutrient: the cofactor vitamin B_{12}.

The B_{12} synthesis paradox: is 1,2-propanediol utilization an aerobic or anaerobic process?

1,2-propanediol and ethanolamine utilization both require vitamin B_{12}, and the B_{12} biosynthetic genes in *S. enterica* were found to be transcriptionally coregulated with the 1,2-propanediol utilization operon [24]. This raised an apparent paradox: 1,2-propanediol and ethanolamine metabolism were once thought to occur only in aerobic conditions, whereas B_{12} synthesis is a strictly anaerobic process [19,25]. This puzzle was resolved by the discovery that 1,2-propanediol and ethanolamine metabolism can proceed using tetrathionate as an electron acceptor in place of molecular oxygen [20]. Tetrathionate, in turn, is a product of the oxidation of thiosulfate, an abundant molecule in the gut produced by the inactivation of H_2S. *Salmonella* or other pathogenic bacteria in the gut might thus synthesize vitamin B_{12} anaerobically while simultaneously respiring to tetrathionate instead of oxygen. Under normal gut conditions oxidation of thiosulfate to tetrathionate is minimal; however, this reaction is accelerated by inflammation, when the gut is rendered a more oxidizing environment [26]. The oxidizing environment of the inflamed gut may favor microcompartment-mediated metabolism by *Salmonella* and other pathogens, allowing their proliferation at the expense of the gut microbiota [13]. Indeed, many pathogenic bacteria possess a mechanism to induce just such an oxidative dysbiosis- the type III secretion system [27].

Type III secretion and microcompartment-mediated metabolism: does *Salmonella* conduct a coordinated assault on the gut microbiota?

Upon encountering environmental cues that are indicative of the host intestinal tract (*e.g.* high osmolarity, low pH, and low oxygen concentration), many enteric pathogens, including *Salmonella* and *Shigella* spp., express type III secretion systems which function to mechanically penetrate the intestinal epithelium and translocate various effector proteins into host cells [28]. These effectors not only mediate internalization of the *Salmonella* cell into a *Salmonella*-containing vacuole, but also induce an inflammatory response throughout the host gut. This

inflammatory response increases the rate of thiosulfate oxidation and hence the concentration of tetrathionate in the gut [29,26]. Not all the invading *Salmonella* cells, however, express the type III secretion system. Experiments examining the transcriptional regulation of type III secretion system promoters indicate that only a fraction of a given population expresses the type III secretion system even in appropriate inducing conditions [30]. What, then, is the role of the non-induced cells? This population is believed to remain in the gut in order to exploit the ensuing inflammation and gain a foothold in the metabolic competition between invaders and commensals [31,32].

Interestingly, 1,2-propanediol represses expression of the type III secretion system master regulator *hilA*, suggesting that cells may undergo a fate decision between type III secretion system-mediated epithelium invasion (leading to bacterial cell death) and 1,2-propanediol or ethanolamine metabolism (leading to proliferation) [29,33]. Furthermore, propionate, a downstream product of 1,2-propanediol metabolism, downregulates another type III secretion system master regulator, HilD, at the post-translational level; it is proposed that endogenous propionate in the gut is primarily responsible for this phenomenon [34]. We additionally propose that post-translational modification of HilD as a result of intracellular propionate production may be a means of downregulating type III secretion in response to 1,2-propanediol utilization microcompartment expression.

Does the paradigm of nutritional competition extend beyond micronutrients?

Nutritional immunity, the modulation of micronutrient concentrations by the host to prevent colonization by pathogens, is well characterized for transition metal micronutrients [35,36]. It seems likely that this paradigm, in which host and commensal processes are under selective pressure to sequester critical nutrients from pathogens, extends beyond micronutrients such as iron and copper ions to other small molecules, as well. Bacterial microcompartments may therefore represent another step in the nutritional arms race between pathogens and commensal species. The coordinated induction of the type III secretion system and bacterial microcompartments in separate bacterial populations allows the pathogen population to induce and exploit an inflamed state in the host gut, allowing colonization followed by diarrhea favorable for subsequent transmission to other hosts [37]. The proposed interplay between type III secretion and bacterial microcompartments suggests that pathogens “dumpster diving” in the gut can develop specialized metabolic mechanisms to utilize compounds otherwise considered to be waste by the gut microbiota. These strategies may involve the coordinated action of multiple cellular processes across the invading bacterial population.

1.4 Microcompartment engineering

In order to engineer a customizable subcellular organelle for catalysis, it is desirable to control both the structural characteristics of the MCP and the passage of small molecules across the proteinaceous MCP membrane. To this end, many researchers have investigated the structural and transport properties of naturally occurring MCP systems, and attempted to engineer new structural and transport characteristics. Crystal structures of several MCP shell proteins have been solved, including those of PduA and PduT [38], PduU [39], EutL [40,41], a shell protein of the putative Grp MCP, GrpN [42], and a shell protein of the putative Etu MCP, EtuB [43]. These structures have allowed the identification of residues predicted to be involved in protein-protein interactions and in small molecule transport through the pores present in shell protein multimers.

Due to the significant structural heterogeneity in MCPs, no crystal structure of the entire macromolecular assembly can be determined, as would be preferred for a viral capsid, for instance. Electron cryotomography of Pdu MCPs, however, has been conducted and confirmed the observation from transmission electron microscopy of thin cell sections that there is significant heterogeneity in MCP shape and size, even within a single cell [44].

Mutational strategies guided by structural data have elucidated residues in PduA important for Pdu MCP assembly [45] and implicated certain residues in the central pore of the PduA hexamer in small molecule transport in and out of the Pdu MCP [46]. It remains uncertain, however, whether the mutants found to influence small molecule transport do so in a general or selective fashion; that is, whether the mutated PduA hexamer pores favor the passage of one small molecule over another. *In vitro* experiments employing an enzyme encapsulated in a virus-like-particle indicate that mutations to the pore region of the VLP affect small molecule transport, suggesting that such a mechanism is feasible for MCPs [47]. In the case of EutL, two conformations, one apparently open and one apparently closed, have been observed crystallographically, raising the possibility of a gated mechanism by which transport can be allosterically regulated [41]. Further investigation will be necessary to establish whether both of these conformers are relevant *in vivo*. Of crucial importance to the field will be further characterization of the abundance of the various shell proteins in the complete MCP assembly; this has been characterized in a bulk sample of Pdu MCPs [2] but not for individual MCPs nor for MCPs of other types. Engineering of transport properties will likely hinge not only on the transport properties of the pore of each shell protein hexamer or pentamer, but also on the abundance of the hexamers and pentamers themselves.

More radical engineering strategies have also been demonstrated. The fusion of four gene products from the *Synechococcus elongatus* carboxysome into a single peptide resulted in the formation of functional carboxysomes [48], and the shell of a computationally identified MCP of unknown function from *Haliangium ochraceum* has been synthesized in *E. coli* [49]. These studies demonstrate that the increasing availability structural and genomic data will continue to increase the breadth and depth of structures available for engineering.

A subcellular biosynthetic organelle will necessarily contain heterologous enzymes, which must be localized in a controlled fashion to the interior of the organelle. At least two, and perhaps many more, MCP systems accomplish this localization through the interaction of N-terminal signal sequences of cargo proteins with C-terminal binding sequences of structural proteins [15,16,50,51]. The first N-terminal signal sequence discovered was that of PduP, the aldehyde dehydrogenase of the Pdu MCP. An 18-amino-acid sequence from the N-terminus of this protein is sufficient to direct heterologous proteins to the Pdu MCP [15]; subsequently, the 20 N-terminal amino acids of the diol dehydratase medium subunit PduD were found to function similarly, as well as mediating the encapsulation of the other 2 subunits of the holoenzyme [50]. Encapsulation was later shown to occur through an interaction of these N-terminal sequences with the C-terminus of the shell protein PduA [16]. The authors suggested based on computational studies that the interaction occurred along the hydrophobic faces of an alpha-helical signal sequence and an alpha-helical C-terminal shell protein domain. NMR studies have confirmed that the PduP signal sequence peptide adopts an alpha-helical confirmation in solution [52].

The N-terminus of EutC has similarly been shown to mediate the encapsulation of heterologous proteins in the Eut MCP [51]. Furthermore, signal sequences from the Eut MCP and the putative Grp MCP localize heterologous protein to the Pdu MCP, mediated by a common

hydrophobic residue motif suggestive of a common amphipathic alpha-helical character [53]. Genomic studies also suggest that this N-terminal amphipathic motif is conserved throughout Pdu- and Eut-associated proteins [54]. Together, these results suggest that a wide array of signal sequences can be generated which mediate the loading of heterologous proteins to MCPs at specified concentrations and stoichiometries. Also of interest is engineering of the MCP shell itself to increase the abundance of cargo binding partners; this may increase the enzyme capacity and hence the catalytic activity of the MCP.

There remain several Pdu enzymes which do not bear N-terminal signal sequences, but which are nonetheless apparently localized to the Pdu MCP lumen. The mechanism of this localization could involve signal sequences in regions away from the N-termini of the enzymes or other protein-protein interactions between the enzymes and the shell proteins of the MCP. Elucidation of these alternate modes of encapsulation may provide additional flexibility in the loading of heterologous proteins to MCPs.

In addition to engineering the means to direct heterologous proteins to the MCP, it is important to gain control over the formation and morphology of the MCPs themselves. The native MCP systems are induced by the presence of the metabolite they degrade; e.g. 1,2-PD in the case of the Pdu MCP [24] and ethanolamine in the case of the Eut MCP [55]. Secondly, both of these operons are subject to catabolite repression and are not induced in media containing glucose or other high-energy carbon sources [56]. Thirdly, the experimental host for both systems is typically a *S.* Typhimurium strain, which requires more stringent biosafety measures than domesticated laboratory *E. coli* strains. For these reasons, researchers have been interested in both controlling the formation of MCPs in their native host and developing plasmid-borne MCP expression systems that can be used in *E. coli* and other less-pathogenic host organisms which do not already bear MCP operons. Episomal expression of the genes for only six Pdu proteins (*pduABB'JKN*) was found to be sufficient for Pdu MCP formation in E. *coli* [57] and expression of the genes for all five Eut shell proteins (*eutSMNLK*) was shown to be sufficient for Eut MCP formation in *E. coli* [51]. Surprisingly, the latter study also found that expression of the *eutS* gene alone formed MCP-like structures, although these were morphologically distinct from the product of the expression of all five genes. Other studies have focused on controlling the formation of Pdu MCPs in the native *Salmonella* host. Episomal expression of the Pdu master regulator gene *pocR* was found to lead to Pdu MCP formation in the absence of 1,2-PD, and even in the presence of glucose [58]. MCPs formed in this manner were found to be functional for 1,2-PD metabolism.

It is as yet unknown whether these artificially controlled MCP systems have differences in shell protein content or transport properties as compared to the native system. Further investigation will be required to determine which aspects of MCP morphology and function can be tuned by using these inducible systems.

1.5 Applications outside biosynthesis

The knowledge of MCP engineering gained in efforts to create biosynthetic organelles can also be applied to a range of other applications. Increasing understanding of the structure of both the pentameric and hexameric MCP subunits and of the MCP structure as a whole will inform the design of self-assembling nanoscale protein structures; recent advances have begun to develop techniques for the design of protein quaternary structures *in silico* [59,60]. Bacterial MCPs also provide a fascinating potential alternative to viral capsids for drug delivery: since they are

typically composed of many unique peptides, each MCP bears multiple uniquely addressable sites for protein post-translational modification or functionalization. Furthermore, MCPs could form the structural basis of self-assembling scaffolds for *in vitro* catalysis in a manner similar to approaches using DNA, RNA, or proteins for scaffolding [7–9].

1.6 Outstanding challenges

There is as yet only one example of the encapsulation of a heterologous biosynthetic pathway in MCPs. The study demonstrated a modest increase in ethanol titer upon encapsulation of an ethanol biosynthesis pathway involving pyruvate decarboxylase and aldehyde dehydrogenase in Pdu MCPs in *E. coli* [52]. This result is promising for the application of MCPs to biosynthesis, but pathway choice remains a challenging problem. It is likely that successfully encapsulated pathways will exploit the native capabilities of the MCPs, but we do not yet thoroughly understand their native mechanism of function (see above for hypotheses as to their mode of action).

Many other engineering challenges remain, and the solution of engineering dilemmas in MCP systems will likely depend on a more thorough understanding of the microbiology and biochemistry of MCPs in their native hosts. We see the engineering objectives as four-fold: (1) control of MCP formation and shell protein content; (2) control of small molecule transport in and out of the MCP lumen; (3) control of enzyme content and stoichiometry; (4) selection of appropriate pathways for encapsulation. Of the four objectives, substantial progress has been made in controlling enzyme content, and there is a growing understanding of means to control MCP formation both episomally and genomically. There is tantalizing evidence for the control of small molecule transport in Pdu MCPs and in model systems, but extensive investigation will be required to further characterize this phenomenon and develop generalizable design rules for engineering. Lastly, pathway choice has been little-addressed, and depends heavily on the resolution of existing uncertainty regarding the native function of MCPs. The area of pathway choice, in particular, will benefit from increasing knowledge of MCP biology and its role in pathogenesis.

Chapter 2: Modulating transcriptional activation of the microcompartment operon to control microcompartment phenotype

2.1 Background

In the Tullman-Ercek lab, we have previously demonstrated that overexpression of the Pdu master regulator *pocR*, even in the absence of its native inducing molecule 1,2-PD, is sufficient to cause morphologically correct Pdu MCP formation [58]. We extended the idea of transcriptional control of Pdu MCP formation in the following experiments to control not only formation but also MCP cargo content.

The following is under consideration at the Journal of Molecular Biology (Jakobson, Chen, Slininger, Kim, and Tullman-Ercek).

2.2 Materials and methods

Plasmid and strain construction- All plasmids were constructed using the GoldenGate one-pot assembly method [61] using amplicons generated with primers as described in Table S3. Knockouts and genomic integrations were generated by the methods of Court [62] using amplicons generated with primers as described in Table 2.1. All plasmid assembly was conducted in *E. coli* DH10B. Molecular biology reagents were purchased from New England Biosciences.

Instruments used in this study- Phase contrast and epifluorescence microscopy were conducted on a Nikon Ni-U upright microscope with a 100x, 1.45 numerical aperture plan apochromat oil immersion objective. Images were captured using an Andor Clara-Lite digital camera. Fluorescence images were collected using a C-FL Endow GFP HYQ band pass filter. Images were captured using Nikon NIS Elements software. All images intended for direct comparison were captured using the same exposure and aperture at room temperature and were adjusted identically in the Adobe Photoshop software for contrast. Bulk fluorescence measurements were acquired using flat-bottomed UV transparent 96-well plates (Corning, Inc.) in a BioTek Synergy HTX multimode plate reader. Flow cytometry measurements were used to confirm these bulk measurements; these were conducted on a Millipore Guava easyCyte 5HT instrument.

Transmission electron microscopy of purified Pdu MCPs- Samples were placed on 400 mesh formvar coated copper grids with a carbon film after glow discharge treatment. 10 μL of purified MCPs at a concentration of approximately 0.1 mg/mL was placed on grids for 2 min. The grids were washed three times with deionized water before fixation. 10 μL of 2% glutaraldehyde in water was placed on grids for 1 min. Grids were then washed an additional three times with deionized water. Finally, samples were stained in 1.6% aqueous uranyl acetate for 1 min. Samples were imaged with an REI Tecnai T12 transmission electron microscope and a Gatan Ultrascan 1000 camera (Gatan, Pleasanton, CA).

Primer	Sequence	Amplicon
EYKP616	cagggccaggataatggtaataaaaagtgaatgtaaataatgtgtttgtttataacaataaattaacgtgtaggctggagctgcttc	Δ*pocR::KanR* cassette
EYKP617	attttctttctgaaatgattaagatacaaaagactatcaaaaatcggcaatagcaaaatattgctatattccggggatccgtcgac	Δ*pocR::KanR* cassette
CMJP155	tagtcccaactatcggaacactccatgcgaggtctttATCAAAGGGAAAACTGTCCATAT	Δ*pduA::cat/sacB* cassette
CMJP156	atcacctgcgccatgatctgttccaccagctcattgctgcTGTGACGGAAGATCACTTCG	Δ*pduA::cat/sacB* cassette
CMJP182	ctttcgggatctccatgcttaatcacaggagaacggcagtTGTGACGGAAGATCACTTCG	Δ*pduJ::cat/sacB* cassette
CMJP183	tgctccttattcgccatcgattaggctgatttcggtaATCAAAGGGAAAACTGTCCATAT	Δ*pduJ::cat/sacB* cassette
CMJP266	tagtcccaactatcggaacactccatgcgaggtctttatgcaacaagaagcactaggaatg	*pduA::pduA-FLAG*
CMJP267	atcacctgcgccatgatctgttccaccagctcattgctgctcatttacttgtcatcgtcatctttataatcttg	*pduA::pduA-FLAG*
CMJP270	ctttcgggatctccatgcttaatcacaggagaacggcagtatgaataacgcactgggactggttg	*pduJ::pduJ-FLAG*
CMJP271	tgctccttattcgccatcgattaggctgatttcggtattacttgtcatcgtcatctttataatcgg	*pduJ::pduJ-FLAG*
CMJP036	ATTGGTCTCATAGAagtaaaggagaagaacttttcactggag	*gfpmut2-ssrA*
CMJP037	ATTGGTCTCATttaagctgctaaagcgtagttttcg	*gfpmut2-ssrA*
CMJP009	ATTGGTCTCACatggaaattaatgaaaaattgctgcgcc	$pduD^{1-20}$
CMJP010	ATTGGTCTCATCTAGActtcatatcgcggagtacgtcttc	$pduD^{1-20}$
EYKP403	ATT GGTCTCACatgaatacttctgaactcgaaaccc	$pduP^{1-18}$
EYKP404	ATTGGTCTCACTAGAtaattgctcgctaagaatggtg	$pduP^{1-18}$
CMJP233	ATTGGTCTCATAGAaccatgattacggattcactgg	*lacZ-1xFLAG-6xHis*
CMJP234	ATTGGTCTCATTTAGTGATGATGATGATGATGCTTGTCATCGTCATCTTTATAATCttttttgacaccagaccaactgg	*lacZ-1xFLAG-6xHis*
EYKP545	ATTGGTCTCACatgatttctgcgagcgctctg	*pocR*
EYKP546	ATTGGTCTCATttatcataacgatggaggatgagaattctc	*pocR*
EYKP534	ATTGGTCTCACatgaaaatcgaagaaggtaaactggtaatctg	*mbp*
EYKP523	ATTGGTCTCATttacttggtgatacgagtctgcg	*mbp*

Table 2.1: Primers used in this study.

2.3 Results

Bacterial microcompartments (MCPs) are protein-bound organelles used by enteric bacteria to organize their metabolism, and are of increasing interest for the organization of heterologous biosynthetic pathways [4,63–66]. These organelles natively function to protect the host from toxic metabolic intermediates and to sequester private pools of cofactors using a

protein membrane [6,67]. Synthetic biologists aim to exploit MCPs to gain these same benefits for heterologous pathways--the titer of an ethanol biosynthesis pathway was shown to increase moderately upon encapsulation--but much remains to be explored before bacterial MCPs can be used as a general, tunable catalytic system [52].

Several N-terminal signal sequences [15,50,16,53] are sufficient to localize heterologous proteins to the 1,2-propanediol utilization (Pdu) MCPs. There is also a growing structural and functional understanding of the protein membrane that encloses the MCP contents, including crystal structures of several shell proteins of the Pdu and other microcompartments [68,39,43,69,70,38]. There is, however, little quantitative understanding of the loading of heterologous proteins to the Pdu MCPs, or of the factors that determine the level of encapsulation. Research on other methods of subcellular enzyme organization, such as RNA scaffolding [7], DNA scaffolding [9], and protein scaffolding [8] has consistently demonstrated that the abundance and stoichiometry of the scaffolded enzymes plays a role in pathway performance, so we seek similar tools to control these variables in the context of bacterial microcompartments.

We have previously demonstrated that the use of alternate signal sequences allows the modulation of the stoichiometric ratio of two cargo proteins [53], but it is also desirable to increase or decrease the absolute amount of heterologous protein loaded to the Pdu MCP in a manner independent of signal sequence. Here we describe transcriptionally and temporally based methods to tune the loading of heterologous proteins, including enzymes, to the Pdu MCP in a controllable fashion without altering the signal sequence or expression level of the cargo protein of interest.

Controlling encapsulation of heterologous cargo by modulating container induction

We speculated that the overexpression of *pocR*, the allosterically regulated transcriptional activator of P_{Pdu} transcription, might affect the loading of heterologous proteins to the MCPs, since we previously found that expression of *pocR* is sufficient to induce Pdu MCP formation in the absence of its allosteric partner, 1,2-propanediol (1,2-PD) [58]. We therefore explored whether episomal expression of *pocR* (denoted P_{TET} *pocR*) in the presence of varying concentrations of 1,2-PD altered the levels of encapsulation of fluorescent reporters of heterologous protein encapsulation. We used two reporters, $PduD^{1-20}$-GFP-SsrA and $PduP^{1-18}$-GFP-SsrA, which bear an N-terminal signal sequence to direct encapsulation and a C-terminal SsrA tag which directs cytosolic (non-encapsulated) GFP for degradation by ClpXP, allowing levels of GFP encapsulation to be measured as reported previously [71]. In all cases, we expressed the cargo protein of interest from a P_{BAD} promoter by the addition of 0.02% w/v arabinose. The primers, plasmids, and strains used throughout this and the following experiments are listed in Tables 2.1-2.3. Unless otherwise indicated, all *Salmonella* strains bear a chromosomal *pocR* allele in addition to any episomal copies. Based on fluorescence measurements, we found that GFP reporter encapsulation increased with increasing 1,2-PD concentration when *pocR* was overexpressed (Figure 2.1A), and we confirmed these results by purification of MCPs according to published methods followed by electrophoresis and western blotting against GFP (Figure 2.1A) [2,53]. Purified MCPs containing GFP were examined by TEM to confirm normal morphology as compared to wild-type MCPs (Figure 2.2). Fluorescence measurements of a Δ*pocR::FRT* strain when *pocR* was overexpressed indicated that GFP encapsulation increased with increasing 1,2-PD concentration, independently of the chromosomal *pocR* allele (Figure 2.3).

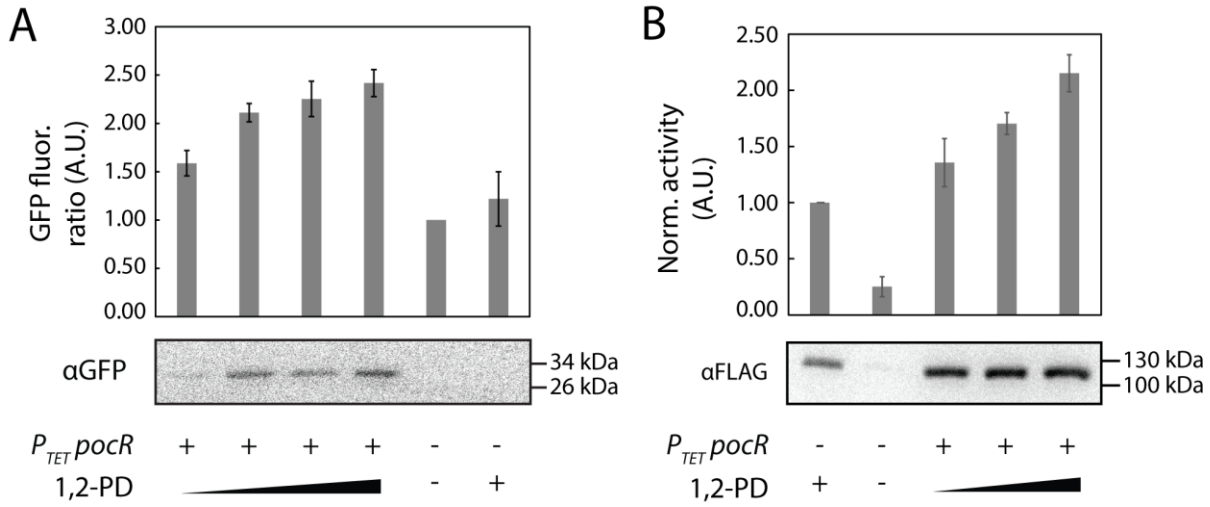


Figure 2.1: Overexpression of *pocR* in the presence of 1,2-PD increases cargo loading to the Pdu MCPs. (A) Top: GFP fluorescence of *S. enterica* cultures expressing the PduD$^{1-20}$-GFP-SsrA encapsulation reporter; values shown are normalized to a culture in which no MCPs are produced. 1,2-PD was added to *pocR*-overexpressing (P_{TET} *pocR*) cultures in increasing concentrations as follows: 0.0055 mM, 0.055 mM, 0.55 mM, 55 mM. aTc was added to all cultures to a final concentration of 1 ng/mL to induce overexpression of *pocR* or *mbp* (as a control). 1,2-PD was added to the culture without *pocR* to a final concentration of 55 mM as indicated. Values shown are the arithmetic mean of n=3 or more replicates; error bars represent the 95% confidence interval. Bottom: Western blotting against GFP of MCPs purified from cultures prepared as above. Loading was normalized to MCP sample total protein concentration. (B) Top: Enzymatic activity of MCPs purified from *S. enterica* cultures expressing the PduP$^{1-18}$-LacZ-1xFLAG-6xHis enzyme fusion; values shown are normalized first by the total protein content of the MCP sample as determined by BCA assay, and then to a culture in which *pocR* was not overexpressed. 1,2-PD was added to *pocR*-overexpressing cultures (P_{TET} *pocR*) in increasing concentrations as follows: 0.0055 mM, 0.055 mM, 55 mM. aTc was added to all cultures to a final concentration of 1 ng/mL to induce overexpression of *pocR* or *mbp* (as a control). 1,2-PD was added to the culture without *pocR* to a final concentration of 55 mM as indicated. Values shown are the arithmetic mean of n=6 or more replicates; error bars represent the 95% confidence interval. Bottom: Western blotting against the FLAG epitope of the purified MCPs examined above. Loading was normalized to MCP sample total protein concentration.

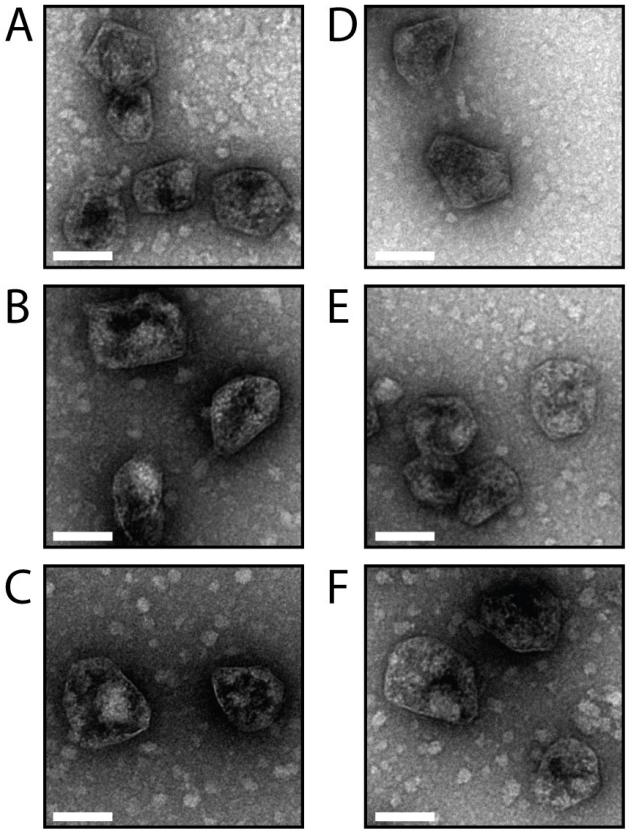


Figure 2.2: MCPs encapsulating heterologous cargo have normal morphology. Transmission electron micrographs of Pdu MCPs from strains encapsulating (A) no heterologous cargo in wild-type MCPs, (B) no heterologous cargo in wild-type MCPs when *pocR* is overexpressed, (C) PduD$^{1-20}$-GFP-SsrA in wild-type MCPs, (D) no heterologous cargo in *pduA::pduA-FLAG* MCPs, (E) no heterologous cargo in *pduJ::pduJ-FLAG* MCPs, and (F) PduP$^{1-18}$-LacZ-1xFLAG-6xHis in wild-type MCPs. Scale bars indicate 100 nm; samples stained with uranyl acetate.

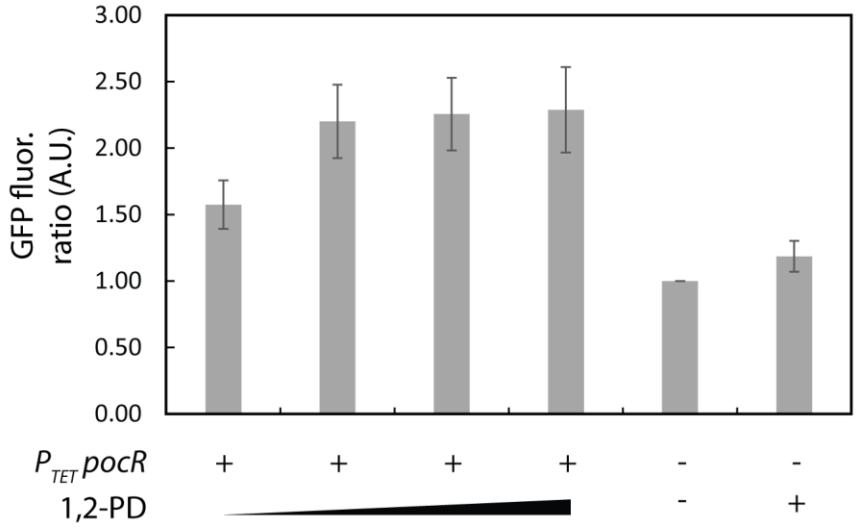


Figure 2.3: Increased loading upon *pocR* overexpression is independent of the chromosomal *pocR* allele. GFP fluorescence of *S. enterica* Δ*pocR::FRT* cultures expressing the PduD$^{1\text{-}20}$-GFP-SsrA encapsulation reporter; values shown are normalized to a culture in which no MCPs are produced. 1,2-PD was added to *pocR*-overexpressing (P_{TET} *pocR*) cultures in increasing concentrations as follows: 0.0055 mM, 0.055 mM, 0.55 mM, 55 mM. aTc was added to all cultures to a final concentration of 1 ng/mL to induce overexpression of *pocR* or *mbp* (as a control); arabinose was added to cultures to a final concentration of 0.02% w/v. 1,2-PD was added to the culture without *pocR* to a final concentration of 55 mM as indicated. Values shown are the arithmetic mean of n=3 or more replicates; error bars represent the 95% confidence interval.

Strain	Organism	Genotype
CMJS271	*Salmonella enterica* serovar Typhimurium LT2	Wild type
CMJS256	*Salmonella enterica* serovar Typhimurium LT2	Δ*pocR::FRT*
CMJS402	*Salmonella enterica* serovar Typhimurium LT2	Δ*pduA::pduA-FLAG*
CMJS404	*Salmonella enterica* serovar Typhimurium LT2	Δ*pduJ::pduJ-FLAG*
EYKS003	*Escherichia coli* DH10B	Wild type
TUC01 [62]	*Escherichia coli* DH10B	*cat/sacB*

Table 2.2: Bacterial strains used in this study.

Plasmid	Genotype	Antibiotic resistance
EYK054	P_{BAD} *pduP*$^{1-18}$*-gfp-ssrA*	Chloramphenicol
CMJ038	P_{BAD} *pduD*$^{1-20}$*-gfp-ssrA*	Chloramphenicol
CMJ172	P_{BAD} *pduP*$^{1-18}$*-lacZ-1xFLAG-6xHis*	Chloramphenicol
EYK352	P_{TET} *pocR*	Kanamycin
CMJ125	P_{TET} *mbp*	Kanamycin
pSIM6 [62]	λ-red, TS (30ºC)	Carbenicillin

Table 2.3: Plasmids used in this study.

Since increasing 1,2-PD concentration increased cargo loading when *pocR* was overexpressed, we also tested whether increased *pocR* expression in the presence of a constant concentration of 1,2-PD also increased encapsulation of a PduD$^{1-20}$-GFP-SsrA reporter. Encapsulation did not increase with increasing concentrations of the *pocR*-inducing molecule (anhydrous tetracycline; aTc), as determined by bulk fluorescence measurements (Figure 2.4). This suggests that even the lowest level of *pocR* expression tested (upon the addition of 1 ng/mL aTc) is sufficient to realize the increase in loading, with further overexpression having little or no effect. This is likely because even the lowest level of heterologous expression is much higher than the physiological expression level of *pocR*.

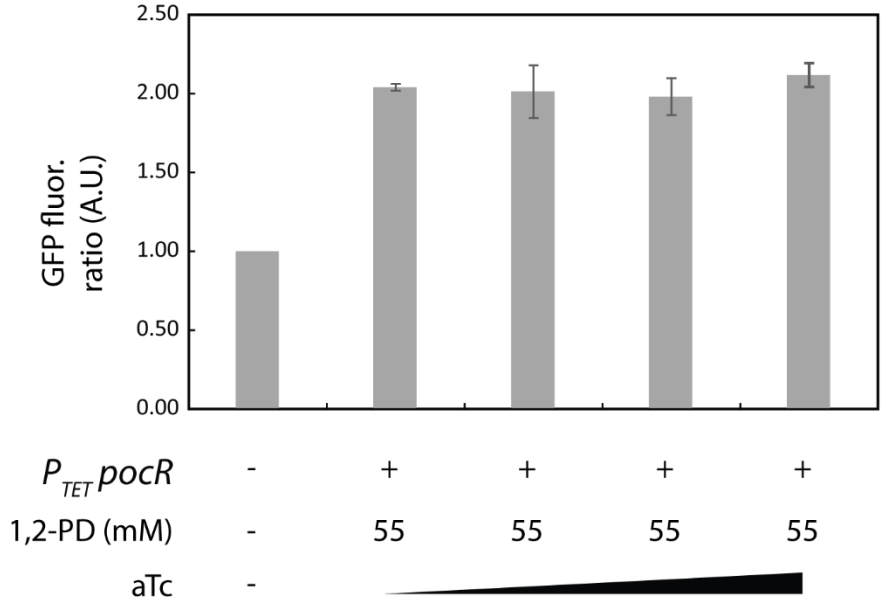


Figure 2.4: Cargo encapsulation does not depend on *pocR* inducer concentration. GFP fluorescence of *S. enterica* cultures expressing the PduD$^{1-20}$-GFP-SsrA encapsulation reporter and overexpressing *pocR* (P_{TET} *pocR*); values shown are normalized to a culture in which no MCPs are produced. 1,2-PD was added to cultures to a final concentration of 55 mM; aTc was added to cultures to final concentrations of 1 ng/mL, 10 ng/mL, 50 ng/mL, 100 ng/mL; arabinose was added to cultures to a final concentration of 0.02% w/v. Values shown are the arithmetic mean of n=3 or more replicates; error bars represent the 95% confidence interval.

In order to confirm that the loading phenotype observed for GFP was generalizable to enzymes, we extended our investigation to the tetrameric beta-galactosidase enzyme encoded by *lacZ*. A PduP$^{1-18}$-LacZ-1xFLAG-6xHis fusion protein was used in these experiments. We found that β-galactosidase encapsulation also increased with increasing 1,2-PD concentration when *pocR* is overexpressed (Figure 2.1B), as determined by anti-FLAG western blots of purified MCPs containing the enzyme. Furthermore, we conducted *in vitro* measurements of the enzymatic activity of purified MCPs containing the enzyme using the colorimetric substrate 2-nitrophenyl-β-galactoside (Sigma, performed according to the manufacturer's directions) and found that the normalized enzymatic activity of purified MCPs varied in a similar fashion as the 1,2-PD concentration increased (Figure 2.1B). Purified MCPs containing PduP$^{1-18}$-LacZ-1xFLAG-6xHis were found to have normal morphology by TEM (Figure 2.2). Thus, we can tune not only GFP loading, but also the heterologous catalytic activity of the MCPs in a general fashion using this transcriptional technique.

Increased encapsulation is due to higher occupancy of cargo binding sites

In order to explore the mechanism of the loading variation observed above, we constructed *Salmonella* strains with the FLAG epitope tag translationally fused to the 3' end of each of two Pdu ORFs, *pduA* and *pduJ*. These two shell proteins are the binding partners of PduD$^{1-20}$- and PduP$^{1-18}$-tagged heterologous cargo, and we speculated that the increased cargo loading observed in the presence of 1,2-PD and *pocR* overexpression could be due to an increased abundance of these proteins in MCPs [16]. We therefore used the Δ*pduA::pduA-FLAG*

and Δ*pduJ::pduJ-FLAG* strains to measure the abundance of cargo binding sites in MCPs by western blot. In order to confirm that these strains formed MCPs, we expressed the fluorescent $PduD^{1-20}$-GFP-SsrA reporter in the presence of 1,2-PD and observed punctate fluorescence indicative of MCP formation (Figure 2.5A). Proper MCP formation was further confirmed by TEM (Figure 2.2). We next observed the expression of each of the *pduA* and *pduJ* ORFs in the whole cell lysate in the presence of 1,2-PD with and without overexpression of *pocR* and found that, interestingly, the total expression levels of *pduA-FLAG* and *pduJ-FLAG* were comparable across both conditions (Figure 2.5B). Expression of the maltose binding protein (*mbp*) gene was used as a negative control in the case when *pocR* was not expressed. These experiments indicated that overall cellular expression levels of *pduA* and *pduJ* were not strongly affected by *pocR* overexpression; however, it remained a possibility that the fully assembled MCPs in each culture contained different levels of the PduA and PduJ cargo-binding partners.

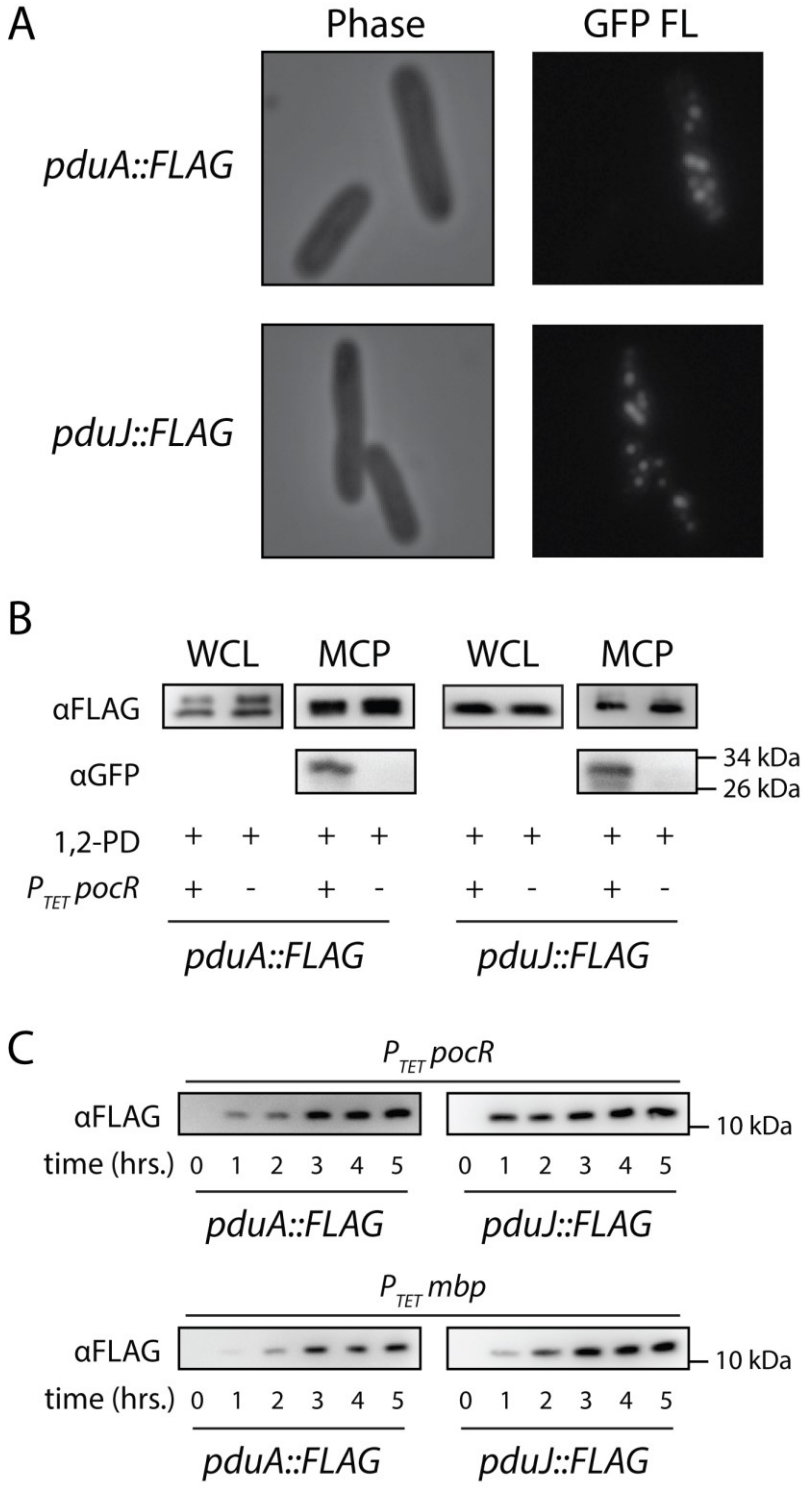


Figure 2.5: Increased cargo encapsulation is due to increased occupancy of cargo binding sites. (A) Phase contrast (Phase) and GFP epifluorescence (GFP FL) microscopy of *S. enterica* strains bearing either *pduA::pduA-FLAG* or *pduJ::pduJ-FLAG* alleles and expressing the PduD$^{1-20}$-GFP-SsrA encapsulation reporter. (B) Western blotting against GFP (αGFP) or against the FLAG epitope (αFLAG) of whole culture lysate (WCL) or purified MCPs (MCP) from strains

bearing either Δ*pduA::pduA-FLAG* or Δ*pduJ::pduJ-FLAG* and overexpressing either *pocR* (P_{TET} *pocR*) or *mbp* (as a control) under the control of a P_{TET} promoter. 1,2-PD was added to all cultures to a final concentration of 55 mM. Loading was normalized to culture OD_{600} at the time of collection or MCP sample total protein concentration, as appropriate. (C) Western blotting against the FLAG epitope (αFLAG) of whole culture lysate collected at one hour intervals as indicated from strains bearing Δ*pduA::pduA-FLAG* or Δ*pduJ::pduJ-FLAG* and overexpressing *pocR* (P_{TET} *pocR*) or *mbp* (P_{TET} *mbp*), as indicated. 1,2-PD was added to all cultures to a final concentration of 55 mM; aTc was added to all cultures to a final concentration of 1 ng/mL. Loading was normalized to culture OD_{600} at the time of collection.

To test this hypothesis, we purified MCPs formed in strains bearing either a Δ*pduA::pduA-FLAG* or Δ*pduJ::pduJ-FLAG* allele. This allowed us to observe PduA-FLAG or PduJ-FLAG abundance in assembled MCPs by western blot. We found that the levels of incorporation of each protein into MCPs were comparable in the presence of 1,2-PD whether or not *pocR* was overexpressed, commensurate with the comparable levels of PduA-FLAG and PduJ-FLAG observed in whole culture lysate (Figure 2.7B). Furthermore, western blotting against GFP in samples of MCPs purified from Δ*pduA::pduA-FLAG* or Δ*pduJ::pduJ-FLAG* strains expressing $PduD^{1-20}$-GFP-SsrA confirmed that cargo loading remained much higher when *pocR* was overexpressed in these strains (Figure 2.7B). This ruled out the possibility of an altered abundance of cargo binding partners (PduA or PduJ) altering the cargo loading capacity of the MCPs and resulting in the greater cargo encapsulation observed in the case of *pocR* overexpression; *pocR* overexpression instead appears to result in an increased occupancy of the available binding sites with cargo.

We next hypothesized that the difference in binding site occupancy observed upon expression of *pocR* in the presence of 1,2-PD might be due to altered transcriptional activation of the Pdu operon, leading to changes in the MCP assembly process that resulted in increased cargo loading. We previously described a fluorescent reporter of P_{Pdu} transcription and measured its response both to 1,2-PD alone and to the expression of *pocR*, so we employed it here to measure P_{Pdu} transcription with increasing 1,2-PD concentration when *pocR* is overexpressed [58]. Upon expression of *pocR* in the presence of 55 mM, 0.5 mM, 0.05 mM, 0.005 mM, and no 1,2-PD, we found that transcriptional activation of the Pdu promoter was similar across all 1,2-PD concentrations, as indicated by the P_{Pdu}-*gfp* transcriptional reporter, and slightly higher than in the absence of 1,2-PD [Fig. S4A]. The activity of the P_{Pdu}-*gfp* transcriptional reporter increased rapidly within 1 hour in all cases in which *pocR* was overexpressed, and the behavior was similar when *pocR* was overexpressed episomally in a Δ*pocR::FRT* strain, again suggesting that the chromosomal *pocR* allele is not implicated (Figure 2.6B). In both cases the normalized fluorescence was much higher throughout the course of the experiment than when 1,2-PD was added to a wild-type *Salmonella* culture overexpressing *mbp* (Figure 2.6C).

We confirmed that the trends observed using the P_{Pdu}-*gfp* episomal transcriptional reporter were consistent with the expression of the chromosomal *pduA* and *pduJ* ORFs by western blotting of whole culture lysate against the FLAG epitope. Samples of Δ*pduA::pduA-FLAG* and Δ*pduJ::pduJ-FLAG* strains expressing either P_{TET} *pocR* or P_{TET} *mbp* were collected at one hour intervals after the addition of 55 mM 1,2-PD and 1 ng/mL aTc (Figure 2.5C). The abundance of PduA::FLAG and PduJ::FLAG in each case followed the same trends observed for the P_{Pdu}-*gfp* transcriptional reporter: the abundance of each protein increased more rapidly in the *pocR*-overexpressing strain, and reached a similar level in each case after 5 hours.

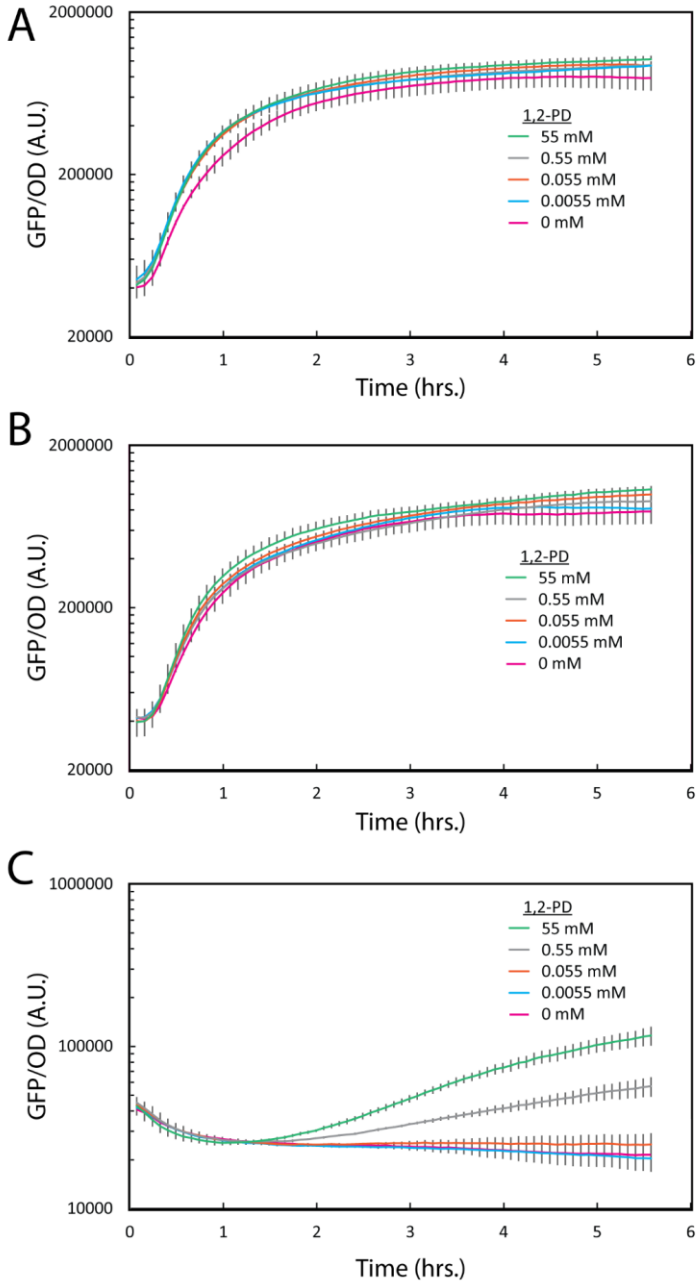


Figure 2.6: Transcriptional activation of P_{Pdu} is more rapid when *pocR* is overexpressed. GFP fluorescence normalized by OD_{600} of *S. enterica* cultures bearing the P_{Pdu}-*gfp* transcriptional reporter (A) expressing *pocR*, (B) expressing *pocR* in a Δ*pocR::FRT* strain, and (C) expressing *mbp*. 1,2-PD was added to cultures to the final concentrations indicated; aTc was added to cultures to a final concentration of 1 ng/mL. Values shown are the arithmetic mean of n=4 or more replicates; error bars represent the 95% confidence interval; the data have been joined with lines to guide the eye.

We therefore concluded that *pocR* overexpression alters the dynamics of P_{Pdu} transcriptional activation and Pdu ORF translation in a manner independent of the chromosomal *pocR* allele, in keeping with our above observation that even the lowest episomal expression level of *pocR* tested was likely much higher than the native level. This does not, however, account for the variation in loading when *pocR* is overexpressed in the presence of varying concentrations of 1,2-PD; this phenomenon is apparently due to an effect downstream of transcription initiation. Further investigation will be required to elucidate the mechanism of this effect.

Controlling encapsulation using cargo and container induction timing

We further hypothesized that changes in the timing of Pdu operon induction and MCP assembly relative to the timing of cargo induction and expression might cause the loading increase observed upon the overexpression of *pocR*. In order to test this hypothesis, we measured the encapsulation of the PduD$^{1-20}$-GFP-SsrA and PduP$^{1-18}$-GFP-SsrA reporters when compartment formation and cargo expression were induced separately at various times relative to the standard induction time (OD_{600}~0.4). In all cases, 1,2-PD was added to a final concentration of 55 mM and *pocR* was overexpressed. Cargo expression (controlled by arabinose at 0.02% w/v) and compartment formation (controlled by 1,2-PD at 55 mM and aTc at 1 ng/mL) were induced at the times indicated. The encapsulation of PduD$^{1-20}$-GFP-SsrA and PduP$^{1-18}$-GFP-SsrA was measured 5.5 hours after the standard induction time, and was found to increase with later cargo induction and earlier compartment induction in both cases, up to a limit that may reflect the maximum cargo capacity of the MCPs under this induction scheme (Figure 2.7).

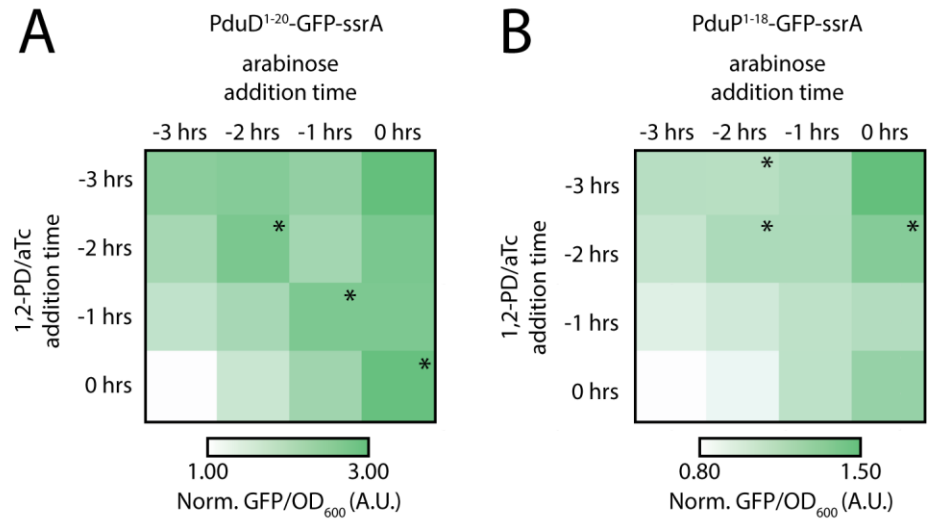


Figure 2.7: Cargo encapsulation depends on the relative timing of cargo and container expression. GFP fluorescence of *S. enterica* cultures expressing (A) the PduD$^{1-20}$-GFP-SsrA encapsulation reporter or (B) the PduP$^{1-18}$-GFP-SsrA encapsulation reporter; values shown are normalized to a culture in which no MCPs are produced. 1,2-PD was added to the cultures at the times indicated to a final concentration of 55 mM; aTc was added to the cultures to induce *pocR* overexpression at the times indicated to a final concentration of 1 ng/mL; arabinose was added to the cultures to induce cargo expression at the times indicated to a final concentration of 0.02% w/v. The times shown are relative to the standard induction time, at which time the OD_{600} of the cultures reaches ~0.4. Values shown are the arithmetic mean of n=3 or more replicates. Samples

in (A) have a 95% confidence interval of less than 0.50, apart from the samples marked with *, for which the 95% confidence interval is less than 0.80; samples in (B) have a 95% confidence interval of less than 0.20, apart from the samples marked with *, for which the 95% confidence interval is less than 0.30.

In order to further characterize this phenomenon, we measured the induction of the P_{Pdu}-*gfp* fusion and the expression of the PduD$^{1-20}$-GFP-SsrA reporter protein separately. The dynamics of induction were similar in each case regardless of the time of induction (Figure 2.8A, 2.8B), leading us to believe that the offset in induction timing explained the trends in loading we observed. This characteristic time difference was quantified as the difference between the time of half-maximal P_{Pdu} activation (as determined from the P_{Pdu}-*gfp* fusion) and the time of maximal PduD$^{1-20}$-GFP-SsrA fluorescence for each induction time. The trend in loading observed for each encapsulation reporter can be seen to collapse to a linear function of this characteristic time until the putative maximum loading limit is reached (Figure 2.8C, 2,8D), indicating that this characteristic time difference, not the absolute timing of induction, is most important in determining cargo loading. This in turn suggests that coordinating cargo and container expression, whether directly by altering the timing of inducer addition or indirectly by altering transcriptional activation, synchronizes key steps in cargo expression and encapsulation.

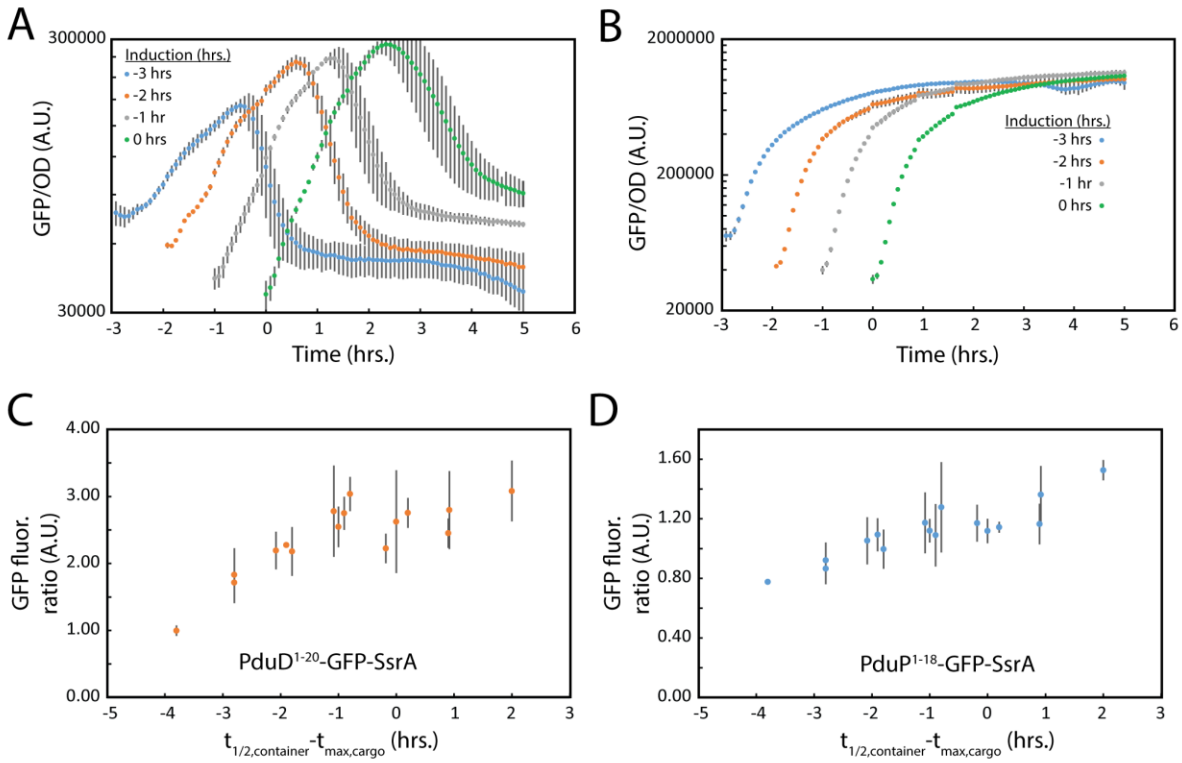


Figure 2.8: The timing dependence of cargo encapsulation is independent of absolute culture time. (A) GFP fluorescence normalized by OD_{600} of *S. enterica* cultures expressing the PduD$^{1-20}$-GFP-SsrA encapsulation reporter to which arabinose was added to a final concentration of 0.02% at the times indicated (relative to the standard induction time). Values shown are the arithmetic mean of n=4 or more replicates; error bars represent the 95% confidence interval. (B) GFP fluorescence normalized by OD_{600} of cultures bearing the P_{Pdu}-*gfp* transcriptional reporter to

which 1,2-PD was added to a final concentration of 55 mM and aTc added to a final concentration of 1 ng/mL at the times indicated (relative to the standard induction time). Values shown are the arithmetic mean of n=4 or more replicates; error bars represent the 95% confidence interval. GFP fluorescence of *S. enterica* cultures expressing (C) the PduD$^{1-20}$-GFP-SsrA encapsulation reporter or (D) the PduP$^{1-18}$-GFP-SsrA encapsulation reporter when cargo and container formation are induced at various times; values shown are normalized to a culture in which no MCPs are produced. Values shown are the arithmetic mean of n=3 or more replicates; error bars represent the 95% confidence interval. Normalized GFP fluorescence is plotted as a function of the difference in encapsulation reporter maximum fluorescence time and transcriptional reporter half-maximum fluorescence time. These times were calculated from the data shown in (A) and (B) above.

2.4 Discussion

Here we demonstrate the first quantitative measurement and control of the loading of heterologous enzymes to the Pdu MCPs. In conjunction with signal sequence tools to be discussed in the next chapter, it is now possible to independently control both the relative and absolute loading of multiple enzymes to subcellular nanoreactors in bacteria [53]. The molecular mechanisms of microcompartment assembly and cargo encapsulation remain obscure, but our results suggest that the relative timing of expression of the microcomparment shell proteins and the cognate enzymatic cargo is of critical importance in creating a custom nanoreactor, and may be important in the native system. Previous studies have indicated that the structure and function of the Pdu microcompartment are highly sensitive to changes in the expression levels of its constituents; we suspect that this is due to the disruption of a finely orchestrated assembly and loading process [72].

Chapter 3: Identifying the key N-terminal motif that directs proteins to the Pdu microcompartments and designing new signal sequences

The following is adapted from Jakobson, Kim, Slininger, Chien, and Tullman-Ercek; Journal of Biological Chemistry, 2015 with permission.

3.1 Background

Having established techniques to control Pdu MCP formation and cargo loading on-demand, we next investigated the sequence and structural characteristics of the N-terminal signal sequences that mediate protein localization to the MCP lumen. Interactions between the N-termini of encapsulated enzymes and the structural proteins of the associated MCP shell are a general mode of enzyme localization to diverse MCP systems, including the Pdu MCP [16,73], the ethanolamine utilization (Eut) MCP [51], and the smaller MCP of *Haliangium ochraceum* [49].

One of these interactions, between the N-terminal signal sequence of the 1,2-propanediol utilitzation (Pdu) enzyme PduP and the C-terminus of the Pdu shell protein PduA, is modeled to be mediated by interactions between residues E7, I10, and L14 presented on the alpha-helical N-terminus of PduP and H81, V84, and L88 on the C-terminus of PduA [16]. Further investigation by NMR confirmed that the PduP N-terminal signal sequence adopts an alpha-helical conformation, providing support for this model of signal sequence-shell protein interaction [52]. Only two signal sequences (both from natively encapsulated Pdu enzymes) have been identified and shown experimentally to localize proteins to the Pdu MCP, however, and a more diverse set of signal sequences with a variety of encapsulation levels is desirable for the encapsulation of heterologous pathways in order to allow tuning of the loading and stoichiometry of multiple heterologous enzymes in the Pdu MCP. Also desirable is a system of orthogonal compartments in which certain proteins are localized exclusively to one type of compartment, and other proteins exclusively to another compartment.

Here, we show that heterologous proteins are encapsulated within the Pdu MCP when fused to N-terminal targeting sequences from several other systems, including the Eut system from the same organism, and that at high expression levels these sequences compete with the native targeting sequences for encapsulation. Interestingly, the ratio of encapsulated proteins can be controlled by modulating the expression levels of these individual signal sequence:cargo protein fusions. These results are useful as a guide to engineering the Pdu MCP to house biosynthetic pathways, and also raise questions as to whether the native Pdu and Eut systems maintain unique cargo protein content when natively co-expressed.

3.2 Materials and methods

Bacterial strains, media, and growth conditions- Salmonella enterica serovar Typhimurium LT2 was used in this study, along with *Escherichia coli* DH10B. *S. enterica* LT2 Δ*pocR::FRT* and Δ*eutR::FRT* strains were constructed by the Lambda Red-based method described previously [74]. The kanamycin resistance cassette was amplified from pKD13 with primers EYKP616 and EYKP617 to yield the amplicon used to create the Δ*pocR::FRT* strain, and with primers CMJP132 and CMJP133 to yield the amplicon used to create the Δ*eutR::FRT* strain. The *S. enterica* LT2 Δ*pduP::cat/sacB* and *S. enterica* LT2 Δ*pduP::eutE* strains were constructed by the Lambda Red-based method of Court [62]. The *cat/sacB* cassette was amplified from the TUC01 genome with primers CMJP228 and CMJP229 to yield the amplicon used to create the

Δ*pduP::cat/sacB* strain, and the *eutE* allele was amplified from the *S. enterica* genome with primers CMJP257 and CMJP258 to yield the amplicon used to create the Δ*pduP::eutE* strain. The genotype of these strains was confirmed by Sanger sequencing of DNA amplified by PCR from the appropriate regions of the *S. enterica* genome (see Table 3.1).

Strain	**Organism**	**Genotype**
CMJS271	*Salmonella enterica* serovar Typhimurium LT2	Wild type
CMJS256	*Salmonella enterica* serovar Typhimurium LT2	Δ*pocR::FRT*
CMJS273	*Salmonella enterica* serovar Typhimurium LT2	Δ*eutR::FRT*
CMJS374	*Salmonella enterica* serovar Typhimurium LT2	Δ*pduP::cat/sacB*
CMJS377	*Salmonella enterica* serovar Typhimurium LT2	Δ*pduP::eutE*
EYKS003	*Escherichia coli* DH10B	Wild type
TUC01 [62]	*Escherichia coli* DH10B	*cat/sacB*

Table 3.1: Bacterial strains used in this study.

S. enterica strains were grown in No-Carbon E (NCE) medium (23 mM monobasic potassium phosphate (Fisher), 48 mM dibasic potassium phosphate (Fisher), 17 mM sodium magnesium sulfate (Fisher), and 50 μM ferric citrate (Sigma)) with 42 mM succinate (Sigma) and either 55 mM 1,2-propanediol (Spectrum) or 30 mM ethanolamine (Alfa) and 150 nM vitamin B12 (Sigma), as indicated [3,75]. Growth on 1,2-propanediol as the sole carbon source was assayed in the absence of succinate and with the addition of 150 mM vitamin B12. *E. coli* was grown in Lysogeny Broth (LB)-Miller medium (10 g/L NaCl, EMD Chemicals). For LB cultures, kanamycin (50 μg/ml, Fisher), chloramphenicol (34 μg/ml, Fisher), and carbenicillin (50 μg/ml, Fisher) were added when necessary for plasmid maintenance. For NCE cultures, kanamycin (25 μg/ml), chloramphenicol (17 μg/ml), and carbenicillin (25 μg/ml) were added when necessary. When indicated, anhydrous tetracycline (aTc, Fisher) and arabinose (ara, CalBioChem) were added at the concentrations indicated when cultures reached OD_{600}~0.4. Cultures were grown at 37°C and 225 rpm orbital shaking unless otherwise indicated.

MCP expression and purification- 5 mL cultures were grown in LB-Miller from a single colony for 24 hours at 30°C, then subcultured 1:1000 into 400 mL of NCE with 55 mM 1,2-PD and grown for 13-15 hours at 37°C. At OD_{600}~0.4, appropriate inducer was added at the concentrations indicated and the cultures were grown for a further 5.5 hours. MCPs were purified by sedimentation as previously described [76], with the following modification: in the place of BPER-II bacterial lysis solution (Thermo), a solution of 1% w/v octylthioglucoside (Santa Cruz Biotech) in 20 mM Tris (Fisher) pH 7.5 in water was used for cell lysis.

Transcriptional regulation analysis- 5 mL cultures were grown in LB-Miller from a single colony for 24 hours at 30°C, then subcultured 1:1000 into 5 mL of NCE and grown for 14-15 hours at 37°C. At OD_{600}~0.4, inducing molecules were added (50 mM 1,2-PD, 30 mM ethanolamine, and 150 nM vitamin B12, as appropriate) and the cultures were grown for the time indicated with time points being collected every hour for analysis by flow cytometry as described below.

Competition analysis- 5 mL cultures were grown in LB-Miller from a single colony for 24 hours at 30°C, then subcultured 1:1000 into 5 mL of NCE with 55 mM 1,2-PD and grown for 14-15 hours at 37°C. At OD_{600}~0.4, appropriate inducers were added (aTc and arabinose) at the concentrations indicated and the cultures were grown for a further 5.5 hours before samples were collected for analysis by flow cytometry as described below.

Reference genomes for Grp and Etu MCPs- Genes of the Grp MCP operon of *Clostridium beijerinckii* and the Etu MCP operon of *Clostridium kluyveri* were identified in a previous bioinformatics study [77]. Gene sequences from the putative Grp MCP operon were retrieved from the NCBI reference genome NC_009617 of *Clostridium beijerinckii*. The gene encoding the aldehyde dehydrogenase of the Grp operon, here denoted GrpAld, has accession number Cbei_4045 in the genome NC_009617. Gene sequences from the Etu operon were retrieved from the NCBI reference genome NC_011837 of *Clostridium kluyveri*. The first aldehyde dehydrogenase encoded in the Etu operon, denoted EtuAld in this manuscript, has the accession number CKR_0977 in this genome. Its first 60 nucleotide bases are identical to those of the second aldehyde dehydrogenase, CKR_0979, of the Etu operon in the genome NC_011837. Multiple sequences alignments were conducted using the Clustal Omega tool (http://www.ebi.ac.uk/Tools/msa/clustalo/) [78].

SDS-PAGE and western blotting- Polyacrylamide gel electrophoresis was carried out by standard procedures [79] with 12.5% or 15% acrylamide gels in a denaturing buffer system. Whole culture lysate sample loading was normalized by culture OD_{600} at time of sample collection. Purified MCP sample loading was normalized by total protein concentration as judged by bicinchoninic acid (BCA) assay performed according to the manufacturers instructions (Thermo). Proteins were transferred to a PVDF membrane for western blotting. Samples were probed with a ClonTech mouse anti-GFP primary antibody, a LifeTech rat anti-mCherry primary antibody, or a Sigma mouse anti-FLAG primary antibody diluted 1:2000 in 50 mM Tris 150 mM NaCl pH 7.6 with 0.05% Tween-20 (TBST) with 1% w/v dry milk, and then with a Thermo HRP-conjugated goat anti-mouse or a Santa Cruz Bitoech HRP-conjugated goat anti-rat secondary antibody diluted 1:1000 in TBST. Labeling was visualized with Thermo west-pico chemiluminescent substrate using a Bio-Rad ChemiDoc XRS+.

Fluorescence microscopy- Bacteria were viewed using a Nikon Ni-U upright microscope with a 100x, 1.45 n.a. plan apochromat oil immersion objective. Images were captured using an Andor Clara-Lite digital camera. Fluorescence images were collected using a C-FL Endow GFP HYQ band pass filter. Images were captured using the Nikon NIS Elements software. All images intended for direct comparison (*e.g.* images of the same strain in Pdu MCP inducing and non-inducing conditions) were captured using the same exposure and aperture at room temperature, and were adjusted identically in the Adobe Photoshop software for contrast.

Transmission electron microscopy of purified Pdu MCPs- MCPs were purified from *S. enterica* cultures as described above. 10 µl of purified MCP samples, at a concentration of 100 µg/ml, were placed on 400 mesh formvar coated copper grids with a carbon film (Electron Microscopy Sciences) for two minutes. The grids were washed three times with deionized water, then stained with 2% aqueous uranyl acetate for two minutes. Samples were observed and photographed with a Gatan Ultrascan 1000 camera (Gatan, Inc., Pleasanton, CA) on a FEI Tecnai T12 transmission electron microscope.

Plasmid construction- Plasmids bearing genes encoding fluorescent reporters and competitor fusion proteins were prepared by the Golden Gate assembly method [61]. Briefly, for the construction of fluorescent reporters, a *gfpmut2* gene was generated with a silent mutation GAC

to GAT at residue 237 to abrogate a BsaI restriction site and enhance assembly efficiency. For cases when a genomic template was available, signal sequence coding regions were amplified by PCR from the *S. enterica* LT2 genome. When such templates were not available (for the Etu, Grp, and *H. ochraceum* genes), signal sequence coding regions were synthesized by appending them directly to the 5' end of the *gfp* gene by PCR using 5' extensions to the primers. An analogous synthesis-by-extension approach was used to append signal sequences to the gene encoding $PduD^{21-224}$. Once appropriate PCR products were prepared, pTET or pBAD inducible plasmids were assembled by a standard Golden Gate temperature cycling protocol (25x[2 minutes at 37ºC, 5 minutes at 16ºC]; 5 minutes at 50ºC; 5 minutes at 80ºC) using T4 DNA ligase and BsaI restriction endonuclease and transformed into *E. coli* DH10B. The sequences of assembled constructs were confirmed by Sanger sequencing of the reporter- or competitor-encoding regions. Enzymes and other molecular biology reagents were obtained from New England Biolabs. Sequences of primers used can be found in Table 3.2.

Primer	**Sequence**	**Amplicon**
EYKP616	cagggccaggataatggtaataaaaagtgaatgtaaataatgtgttttgtttataacaataa attaac gtgtaggctggagctgcttc	Δ*pocR::Kan*^{R} cassette
EYKP617	attttctttctgaaatgattaagatacaaaagactatcaaaaatcggcaatagcaaaatattg ctat attccggggatccgtcgac	Δ*pocR::Kan*^{R} cassette
CMJP132	aaacgcagtagtcgctatcgcatcaaaaattaatctgtcggagggccgggcgtcctggc gggcgt gtgtaggctggagctgcttc	Δ*eutR::Kan*^{R} cassette
CMJP133	agctgccaccgggcaattcccgacacgcgcacgttatcagcaaccggagagcctccc catcaata attccggggatccgtcgac	Δ*eutR::Kan*^{R} cassette
CMJP228	gaccgccatcgcggctattaacgtgggaactcatcaatga TGTGACGGAAGATCACTTCG	Δ*pduP::cat/sacB* cassette
CMJP229	aaccgcgtttgtagtgagaaggtattcatcgcgacctcag ATCAAAGGGAAAACTGTCCATAT	Δ*pduP::cat/sacB* cassette
CMJP257	CACAGACCGCCATCGCGGCTATTAACGTGGGAACTC ATCA atgaatcaacaggatattgaacagg	Δ*pduP::eutE* cassette
CMJP258	AACCGCGTTTGTAGTGAGAAGGTATTCATCGCGACCT CAG ttatacaatgcgaaacgcatcc	Δ*pduP::eutE* cassette
CMJP036	ATT GGTCTC A TAGA agtaaaggagaagaacttttcactggag	*gfpmut2-ssrA*
CMJP037	ATT GGTCTC AT ttaagctgctaaagcgtagttttcg	*gfpmut2-ssrA*
CMJP009	ATT GGTCTC A C atg gaaattaatgaaaaattgctgcgcc	$pduD^{1-20}$
CMJP010	ATT GGTCTCATCTAGA cttcatatcgcggagtacgtcttc	$pduD^{1-20}$
EYKP403	ATT GGTCTC A Catgaatacttctgaactcgaaaccc	$pduP^{1-18}$
EYKP404	ATT GGTCTC A TCTAGAtaattgctcgctaagaatggtg	$pduP^{1-18}$

CMJP117	ATT GGTCTC A Catg gatcaaaaacagattgaagaaattgtacgtag	*eutC*$^{1-20}$
CMJP118	ATT GGTCTC A TCTAGA gtcctgtcccattgacgc	*eutC*$^{1-20}$
CMJP120	ATT GGTCTC A Catg aatcaacaggatattgaacaggtg	*eutE*$^{1-20}$
CMJP121	ATT GGTCTC A TCTAGA gctgtctttcattttcagcagtacc	*eutE*$^{1-20}$
CMJP149	ATT GGTCTC A C ATG GATAAGGACTTACAGTCAATACAGGAAGTAAGAACT CTTATAGCAAAAGCAAAGAAA tctagaagtaaaggagaagaacttttcac	*etuAld*$^{1-20}$*- gfpmut2-ssrA*
CMJP037	ATT GGTCTC AT ttaagctgctaaagcgtagttttcg	*etuAld*$^{1-20}$*- gfpmut2-ssrA*
CMJP151	CTTGTTAGGCAAGCAATAGAGGAAGTGAAA tctagaagtaaaggagaagaacttttcac	*grpAld*$^{1-20}$*-gfpmut2-ssrA*
CMJP152	ATT GGTCTC A C ATG GATGTAGATGTAGTTTTAGTTGAAAAA cttgttaggcaagcaatagaggaa	*grpAld*$^{1-20}$*-gfpmut2-ssrA*
CMJP037	ATT GGTCTC AT ttaagctgctaaagcgtagttttcg	*grpAld*$^{1-20}$*-gfpmut2-ssrA*
CMJP153	ATCGTCGAGCGCGTGCTCGCGCGCCTCGAC tctagaagtaaaggagaagaacttttcac	*HochAld*$^{1-20}$*-gfpmut2-ssrA*
CMJP154	ATT GGTCTC A C ATG GCTCTGCGCGAGGATCGCATCGCCGAG atcgtcgagcgcgtg	*HochAld*$^{1-20}$*-gfpmut2-ssrA*
CMJP037	ATT GGTCTC AT ttaagctgctaaagcgtagttttcg	*HochAld*$^{1-20}$*-gfpmut2-ssrA*
CMJP173	ATT GGTCTC A C ATG GATAAGGACTTACAGTCAATACAGGAAATT gtaagaactcttatagcaaaagcaaagaaa	*etuAld+Ile*$^{1-21}$*-GFPmut2-ssrA*
CMJP037	ATT GGTCTC AT ttaagctgctaaagcgtagttttcg	*etuAld+Ile*$^{1-21}$*-GFPmut2-ssrA*
CMJP217	ATT GGTCTC A C ATG GATCAAAAACAGATTGAAGAAATTGTACGTAGCGTG ATGGCGTCAATGGGACAGGAC ggcagcgataaacccgtc	*eutC*$^{1-20}$*-pduD*$^{21-224}$*-FLAG*
CMJP225	ATT GGTCTC A T TTA CTTGTCATCGTCATCTTTATAATC tcaaagcgccacgcg	*eutC*$^{1-20}$*-pduD*$^{21-224}$*-FLAG*
CMJP218	ATT GGTCTC A C ATG AATCAACAGGATATTGAACAGGTGGTGAAAGCGGTA CTGCTGAAAATGAAAGACAGC ggcagcgataaacccgtc	*eutE*$^{1-20}$*-pduD*$^{21-224}$*-FLAG*

CMJP225	ATT GGTCTC A T TTA CTTGTCATCGTCATCTTTATAATC tcaaagcgccacgcg	*eutE*$^{1-20}$-*pduD*$^{21-224}$-*FLAG*
CMJP219	ATT GGTCTC A C ATG GATGTAGATGTAGTTTTAGTTGAAAAACTTGTTAGGC AAGCAATAGAGGAAGTGAAA ggcagcgataaacccgtc	*grpAld*$^{1-20}$-*pduD*$^{21-224}$-*FLAG*
CMJP225	ATT GGTCTC A T TTA CTTGTCATCGTCATCTTTATAATC tcaaagcgccacgcg	*grpAld*$^{1-20}$-*pduD*$^{21-224}$-*FLAG*
CMJP220	ATT GGTCTC A C ATG GCTCTGCGCGAGGATCGCATCGCCGAGATCGTCGAG CGCGTGCTCGCGCGCCTCGACtct ggcagcgataaacccgtc	*HochAld*$^{1-20}$-*pduD*$^{21-224}$-*FLAG*
CMJP225	ATT GGTCTC A T TTA CTTGTCATCGTCATCTTTATAATC tcaaagcgccacgcg	*HochAld*$^{1-20}$-*pduD*$^{21-224}$-*FLAG*
CMJP222	ATT GGTCTC A C ATG GATAAGGACTTACAGTCAATACAGGAAGTAAGAACT CTTATAGCAAAAGCAAAGAAA ggcagcgataaacccgtc	*etuAld1-20-pduD21-224-FLAG*
CMJP225	ATT GGTCTC A T TTA CTTGTCATCGTCATCTTTATAATC tcaaagcgccacgcg	*etuAld1-20-pduD21-224-FLAG*
CMJP223	ATT GGTCTC A C ATG GATAAGGACTTACAGTCAATACAGGAAATTGTAAGA ACTCTTATAGCAAAAGCAAAGAAA ggcagcgataaacccgtc	*etuAld+Ile1-20-pduD21-224-FLAG*
CMJP225	ATT GGTCTC A T TTA CTTGTCATCGTCATCTTTATAATC tcaaagcgccacgcg	*etuAld+Ile1-20-pduD21-224-FLAG*
CMJP303	ATT GGTCTC A C ATG AATACTTCTGAACTCGAAACCCTGATTCGCACCATTC TTAGCGAGCAATTA gtgagcaagggcgaggagg	*pduP*$^{1-18}$-*mCherry-ssrA*
CMJP305	ATT GGTCTC A T TTA AGCTGCTAAAGCGTAGTTTTCGTCGTTTGCTGCGTCG AC cttgtacagctcgtccatgccg	*pduP*$^{1-18}$-*mCherry-ssrA*
CMJP304	ATT GGTCTC A C ATG GAAATTAATGAAAAATTGCTGCGCCAGATAATTGAA GACGTACTCCGCGATATGAAG gtgagcaagggcgaggagg	*pduD*$^{1-20}$-*mCherry-ssrA*
CMJP305	ATT GGTCTC A T TTA AGCTGCTAAAGCGTAGTTTTCGTCGTTTGCTGCGTCG AC cttgtacagctcgtccatgccg	*pduD*$^{1-20}$-*mCherry-ssrA*

Table 3.2: Primers used in this study.

The P_{eut} transcriptional reporter was synthesized by the same method described previously for the reporter of P_{pdu} transcription [58]. Briefly, Golden Gate assembly was used to assemble the promoter region of interest, a fluorescent reporter gene *gfpmut2*, and pPROTET vector backbone to form the reporter plasmid. Correct assembly was once again confirmed by Sanger sequencing. Sequences for each plasmid (Table 3.3) are available for download at the AddGene database.

Plasmid	**Genotype**	**Antibiotic resistance**
EYK054	pBAD *pduP*$^{1\text{-}18}$*-gfpmut2-ssrA*	Chloramphenicol
EYK345	pTET *pduD*$^{21\text{-}224}$	Kanamycin
EYK346	pTET *pduD*$^{1\text{-}224}$	Kanamycin
CMJ038	pBAD *pduD*$^{1\text{-}20}$*-gfpmut2-ssrA*	Chloramphenicol
CMJ116	pBAD *eutC*$^{1\text{-}20}$*- gfpmut2-ssrA*	Chloramphenicol
CMJ119	pBAD *eutE*$^{1\text{-}20}$*- gfpmut2-ssrA*	Chloramphenicol
CMJ129	pBAD *etuAld*$^{1\text{-}20}$*- gfpmut2-ssrA*	Chloramphenicol
CMJ130	pBAD *grpAld*$^{1\text{-}20}$*- gfpmut2-ssrA*	Chloramphenicol
CMJ131	pBAD *HochAld*$^{1\text{-}20}$*- gfpmut2-ssrA*	Chloramphenicol
CMJ137	pBAD *etuAld+Ile*$^{1\text{-}21}$*-gfpmut2-ssrA*	Chloramphenicol
CMJ156	pTET *eutC*$^{1\text{-}20}$*-pduD*$^{21\text{-}224}$*-FLAG*	Kanamycin
CMJ157	pTET *eutE*$^{1\text{-}20}$*-pduD*$^{21\text{-}224}$*-FLAG*	Kanamycin
CMJ158	pTET *grpAld*$^{1\text{-}20}$*-pduD*$^{21\text{-}224}$*-FLAG*	Kanamycin
CMJ159	pTET *HochAld*$^{1\text{-}20}$*-pduD*$^{21\text{-}224}$*-FLAG*	Kanamycin
CMJ161	pTET *etuAld*$^{1\text{-}20}$*-pduD*$^{21\text{-}224}$*-FLAG*	Kanamycin
CMJ162	pTET *etuAld+Ile*$^{1\text{-}20}$*-pduD*$^{21\text{-}224}$*-FLAG*	Kanamycin
CMJ205	pTET *pduP*$^{1\text{-}18}$*-mCherry-ssrA*	Kanamycin
CMJ206	pTET *pduD*$^{1\text{-}20}$*-mCherry-ssrA*	Kanamycin
pKD13 [74]	*FRT-Kanamycin*^{R}*-FRT*	Carbenicillin
pKD46 [74]	pBAD λ-red, TS (30ºC)	Carbenicillin
pSIM6 [62]	λ-red, TS (30ºC)	Carbenicillin

PCP20 [74]	*flp* recombinase	Carbenicillin

Table 3.3: Plasmids used in this study.

Flow cytometry- Cultures were grown as described above, and at the indicated time points, aliquots of the samples were diluted to OD_{600}~0.01 in phosphate-buffered saline (PBS) supplemented with 2 g/L kanamycin (to halt translation), and stored at 4°C [80]. Samples were then diluted 1:40 into PBS supplemented with 2 g/L kanamycin in 96-well plates for flow cytometry. The GFP fluorophore was allowed to mature for 30 minutes following collection of the last sample, and 10,000 events were collected for each sample on a Millipore Guava easyCyte 5HT instrument. Cells were distinguished from debris by gating on the forward and side scatter channels using the FlowJo software. Reported fluorescence values are the arithmetic mean of the geometric mean green fluorescence of three independent samples acquired on three different days. Unless otherwise noted, error bars represent one standard deviation.
Two-color competition assays- Cultures were grown as described, with appropriate inducing molecules added when the OD_{600} reached ~0.4. At 5.5 hours post-induction, cells were diluted 1:4 into PBS with 2 g/L kanamycin and we measured OD_{600} as well as bulk fluorescence in the GFP and mCherry channels for each sample. Measurements were collected using flat-bottomed UV-transparent 96-well plates (Corning, Inc.) in a BioTek Synergy HTX Multi-Mode plate reader. Reported fluorescence values are normalized to OD_{600} and subsequently normalized to the fluorescence of a control sample without 1,2-PD and with the same concentrations of inducing molecules. Values reported are the arithmetic mean of three biological replicates.

3.3 Results

S. enterica Δ*eutR::FRT does not form Eut MCPs, and can form Pdu MCPs.* We hypothesized that since the known or putative N-terminal signal sequences from other MCP systems bear similar hydrophobic residue motifs (Figure 3.1), these N-terminal signal sequences would localize heterologous cargo to the Pdu MCPs. *S. enterica* LT2, however, can express both Pdu and Eut MCPs, and we wished only to observe Pdu MCP formation and cargo encapsulation. The transcriptional regulator PocR is necessary for Pdu MCP formation [24,58,81,82]. We confirmed that the analogous transcriptional regulator EutR is necessary for Eut MCP formation by two methods: first using a fluorescent reporter of P_{eut} transcription [55,83], and then using a fluorescent reporter of encapsulation in the Eut or Pdu MCP [51]. We observed activation of the P_{eut} promoter by 30 mM ethanolamine and 150 nM vitamin B12, as indicated by increased GFP fluorescence in a *S. enterica* strain containing the P_{eut}-*gfpmut2* transcriptional reporter as compared to a control culture to which no ethanolamine and vitamin B12 was added. This activation was abrogated in a *S. enterica* Δ*eutR::FRT* strain (Figure 3.2). Activation could be complemented in the *S. enterica* Δ*eutR::FRT* strain by expression of EutR from a secondary plasmid (Figure 3.2). Episomal maltose binding protein (MBP) was expressed as a negative control for these experiments. To observe Eut MCP formation, we used a fluorescent reporter of encapsulation in which $EutC^{1-20}$ is fused to GFP followed by a C-terminal ssrA tag, which mediates degradation of the fluorophore by the ClpXP protease in the cytosol (encapsulation in an MCP thus rescues the reporter from proteolysis). When expression of the encapsulation reporter construct pBAD-$eutC^{1-20}$-*gfp-ssrA* was induced by the addition of 0.02% arabinose, fluorescent puncta were not observed in a Δ*eutR::FRT* strain upon the addition of 30 mM ethanolamine and 150 nM vitamin B12, but puncta were observed in the Δ*pocR::FRT* strain under the same conditions (Figure 3.3, 3.2). We therefore conducted our subsequent experiments

in the *S. enterica* Δ*eutR::FRT* strain in order to ensure that we observed only Pdu MCP formation and encapsulation.

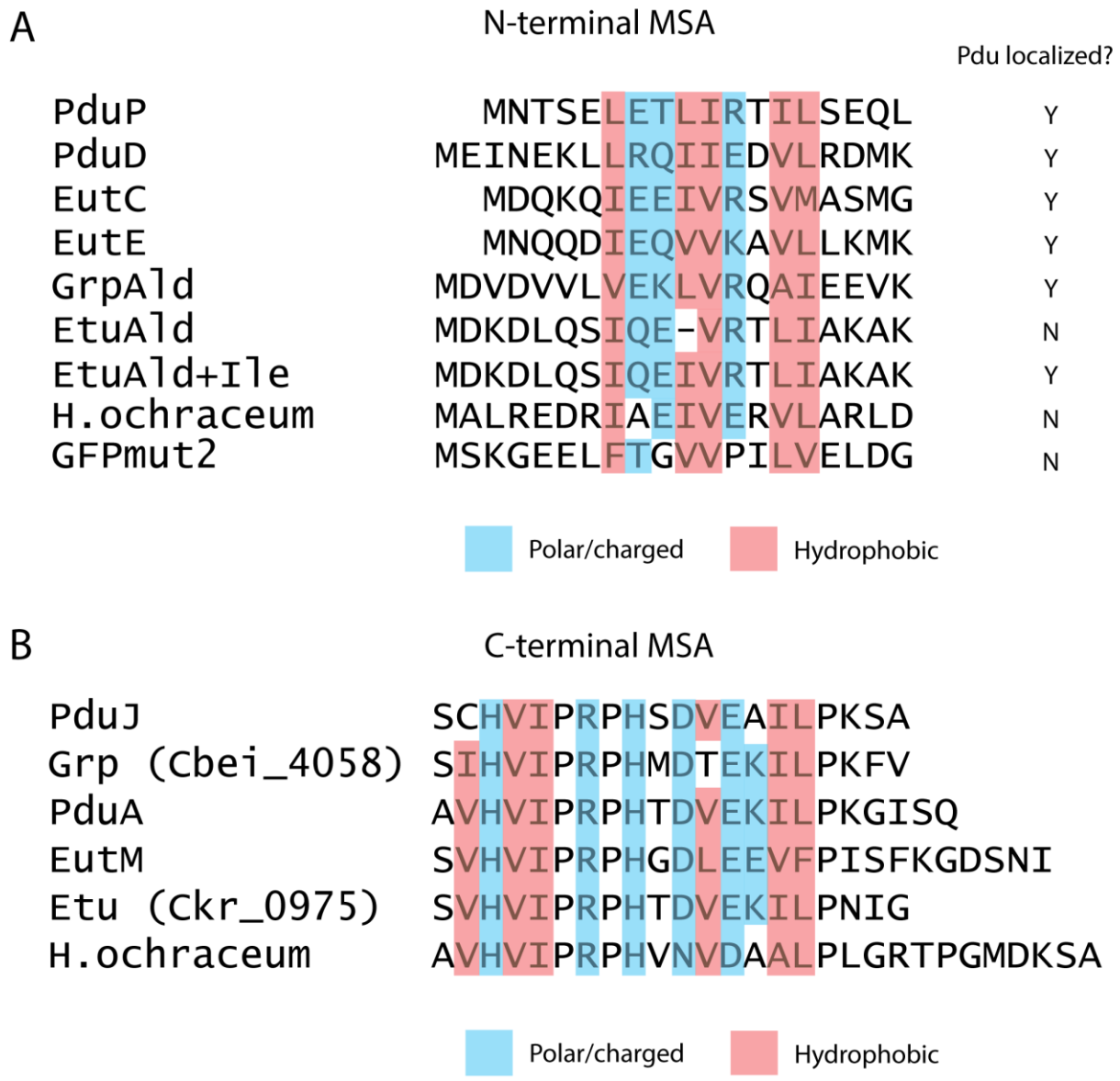
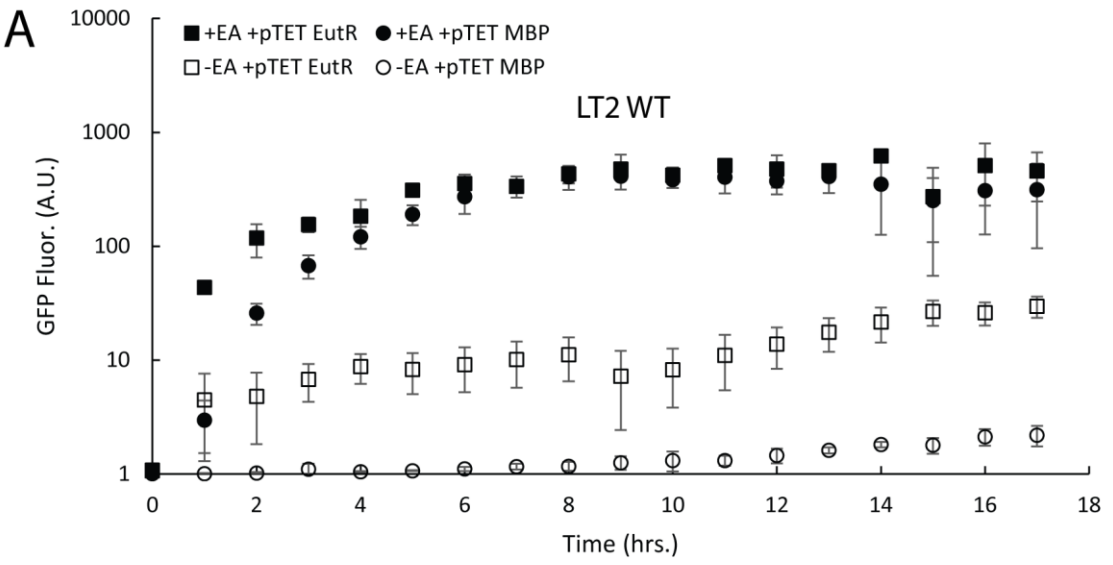
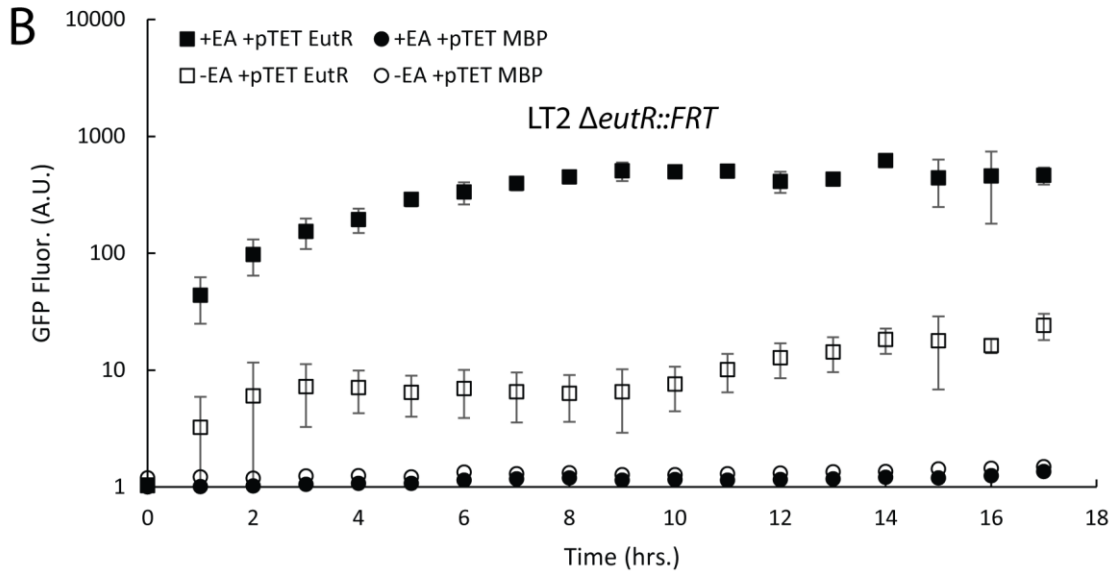


Figure 3.1: Sequence alignment of N-terminal sequences of encapsulated enzymes and C-terminal sequences of structural proteins. (A) A multiple-sequence alignment (MSA) of the N-terminal amino acid residues of the Pdu enzymes PduD and PduP, the Eut enzymes EutC and EutE, the Grp aldehyde dehydrogenase, the Etu aldehyde dehydrogenase, the *H. ochraceum* compartment signal sequence, the Etu aldehyde dehydrogenase with Ile added at residue 11, and the first 20 amino acids of the GFP protein, for comparison. (B) A multiple sequence alignment of the C-terminal amino acid residues of the Pdu structural proteins PduJ and PduA, the Eut structural protein EutM, the putative Grp structural protein from *C. beijerinckii* denoted Cbei_4058, the putative Etu structural protein from *C. kluyveri* denoted Ckr_0975, and the hexagonal *H. ochraceum* structural protein.


A
+EA +pTET EutR
+EA +pTET MBP
-EA +pTET EutR
-EA +pTET MBP
LT2 WT
GFP Fluor. (A.U.)
Time (hrs.)


B
+EA +pTET EutR
+EA +pTET MBP
-EA +pTET EutR
-EA +pTET MBP
LT2 ΔeutR::FRT
GFP Fluor. (A.U.)
Time (hrs.)

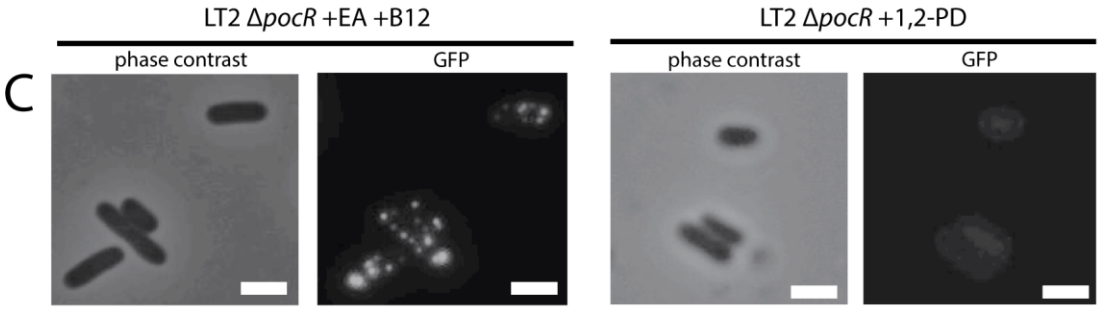

C
LT2 ΔpocR +EA +B12
LT2 ΔpocR +1,2-PD
phase contrast
GFP
phase contrast
GFP

Figure 3.2: EutR is necessary for Eut operon induction and Eut MCP formation. Flow cytometry of *S. enterica* cultures bearing a plasmid encoding the *Peut-gfpmut2* reporter construct and either a pTET *eutR* plasmid or a pTET *mbp* control vector in (A) wild type *S. enterica* LT2 and (B) *S. enterica* LT2 Δ*eutR*. Inducing molecules (30 mM ethanolamine, and 150 nM vitamin B12, as denoted by +EA) were added at OD600~0.4. Anhydrous tetracycline was added to all cultures at 1 ng/mL at this time. Values shown are the arithmetic mean of the geometric mean fluorescence of three independent replicates measured at one-hour intervals after induction, as indicated; error bars indicate the standard deviation. (C) Phase contrast and fluorescence microscopy of *S. enterica* LT2 Δ*pocR* expressing a $EutC^{1-20}$-GFPmut2-ssrA fluorescent reporter in the presence of 30 mM ethanolamine and 150 nM vitamin B12 or 55 mM 1,2-PD as indicated.

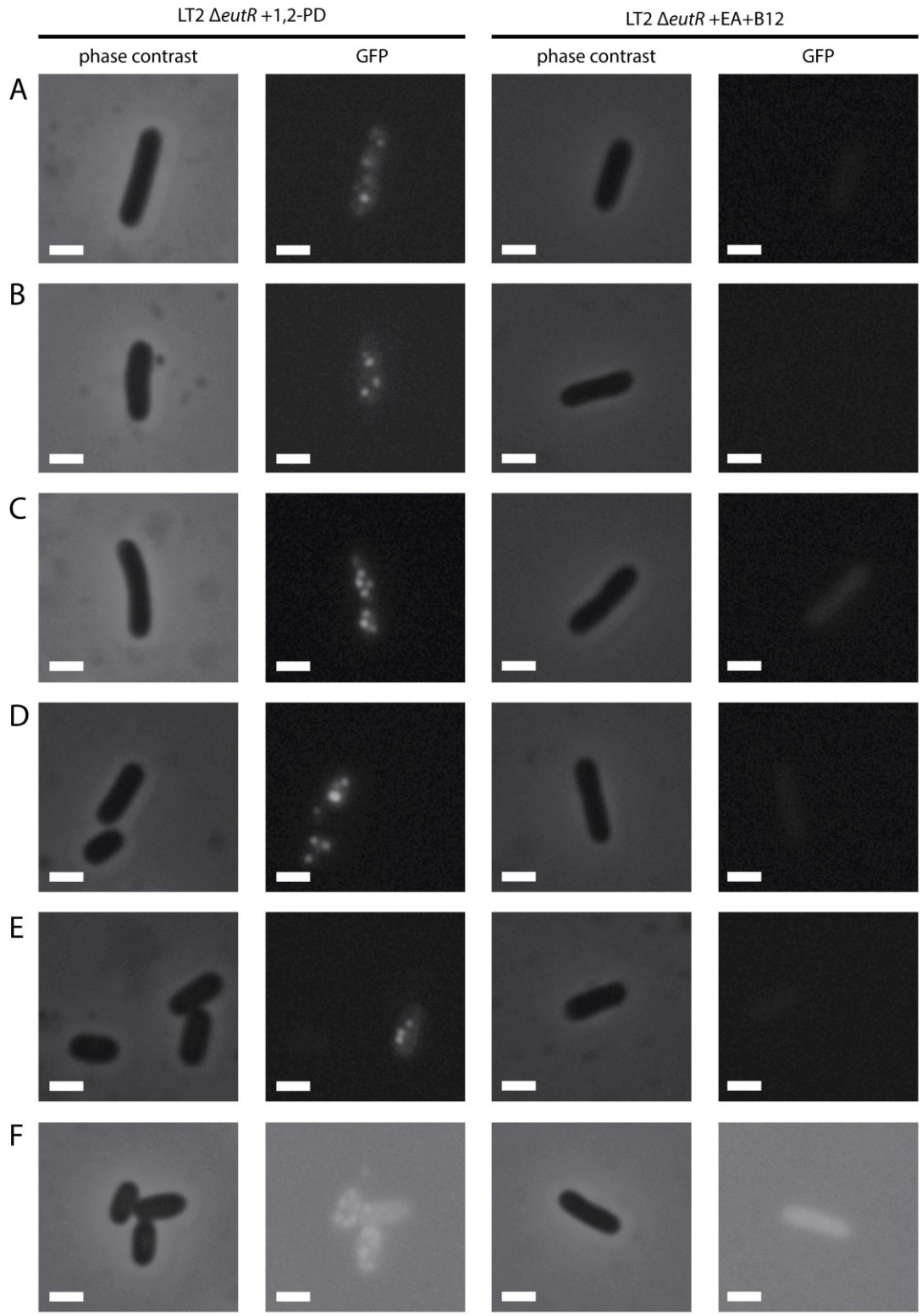


Figure 3.3: Several non-native targeting sequences appear to localize GFP to Pdu MCPs in *S. enterica*. Phase contrast and fluorescence microscopy of *S. enterica* LT2 Δ*eutR* expressing (A) PduD$^{1-20}$-GFP-ssrA, (B) PduP$^{1-18}$-GFP-ssrA, (C) EutC$^{1-20}$-GFP-ssrA, (D) EutE$^{1-20}$-GFP-ssrA,

(E) GrpAld$^{1-20}$-GFP-ssrA, (F) EtuAld+Ile$^{1-21}$-GFP-ssrA in the presence of 55 mM 1,2-PD (left columns) or 30 mM ethanolamine and 150 nM vitamin B12 (right columns). Scale bars represent 1 μm.

EutC$^{1-20}$ and EutE$^{1-20}$ direct heterologous proteins to the Pdu MCP – We next explored whether two signals sequences from the ethanolamine utilization (Eut) MCP are promiscuous in their localization of heterologous proteins; that is, whether the N-terminal signal sequences EutC$^{1-20}$ and EutE$^{1-20}$ localize proteins to the Pdu MCPs. When modeled as alpha helices, the Eut signal sequences present similar hydrophobic motifs to those of the Pdu signal sequences, leading us to suspect that they would interact with the Pdu MCP shell proteins in a similar manner (Figure 3.1). We therefore tested whether the EutC and EutE signal sequences function to mediate the encapsulation of the GFP-ssrA reporter in the Pdu MCP. Upon the induction of either pBAD-*eutC$^{1-20}$-gfp-ssrA* or pBAD-*eutE$^{1-20}$-gfp-ssrA* plasmids with 0.02% arabinose in *S. enterica* LT2 Δ*eutR::FRT*, punctate fluorescence was observed when 55 mM 1,2-propanediol was added to the culture media to induce the formation of Pdu MCPs. Cells grown in ethanolamine MCP-inducing conditions- in the presence of 0.02% arabinose, 30 mM ethanolamine, and 150 nM vitamin B12- showed no punctate fluorescence (Figure 3.3). A western blot against GFP of purified Pdu MCPs from *S. enterica* LT2 Δ*eutR::FRT* expressing EutC$^{1-20}$-GFP-ssrA and EutE$^{1-20}$-GFP-ssrA confirmed that these two Eut signal sequences target GFP to the Pdu MCP (Figure 3.4). Transmission electron micrographs of purified Pdu MCPs confirmed that MCPs are morphologically normal when EutC$^{1-20}$-GFP-ssrA and EutE$^{1-20}$-GFP-ssrA are encapsulated (Figure 3.4).

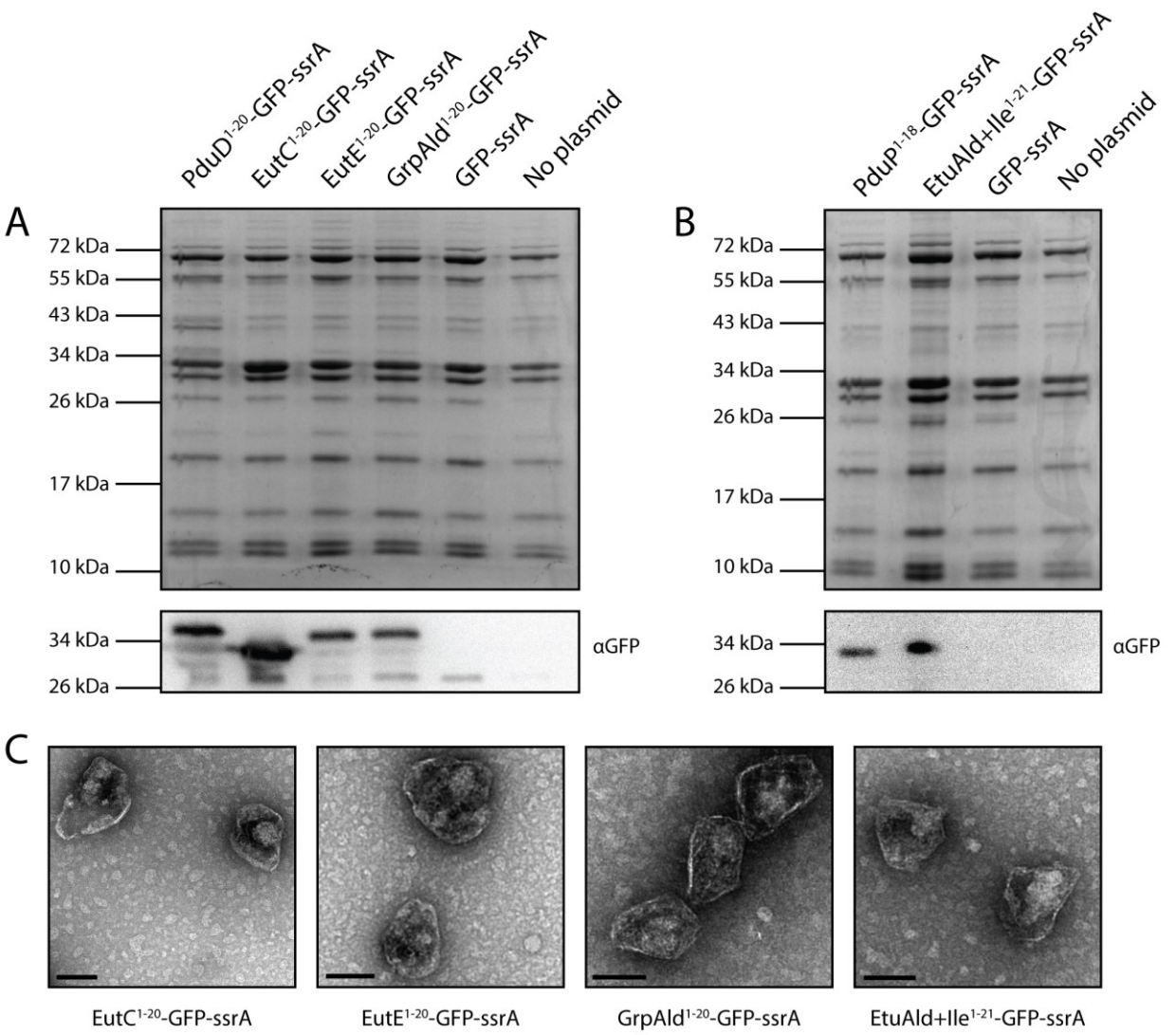


Figure 3.4: Characterization of purified Pdu MCPs containing encapsulation reporters. (A) Coomassie stain and western blot against GFP of SDS-PAGE gel of purified Pdu MCPs (with loading normalized by total protein concentration as judged by BCA assay) from *S. enterica* LT2 Δ*eutR* cultures expressing PduD$^{1-20}$-GFP-ssrA, EutC$^{1-20}$-GFP-ssrA, EutE$^{1-20}$-GFP-ssrA, GrpAld$^{1-20}$-GFP-ssrA, untagged GFP-ssrA, and from a strain bearing no plasmid. (B) Coomassie stain and western blot against GFP of SDS-PAGE gel of purified Pdu MCPs (with loading normalized by total protein concentration as judged by BCA assay) from *S. enterica* LT2 Δ*eutR* cultures expressing PduP$^{1-18}$-GFP-ssrA, EtuAld+Ile$^{1-21}$-GFP-ssrA, untagged GFP-ssrA, and from a strain bearing no plasmid. The western blots for the two sets of purified MCP samples were conducted separately due to large variations in expression and encapsulation between different GFP fusion constructs. (C) Transmission electron micrographs of Pdu MCPs purified from *S. enterica* LT2 Δ*eutR* cultures expressing EutC$^{1-20}$-GFP-ssrA, EutE$^{1-20}$-GFP-ssrA, GrpAld$^{1-20}$-GFP-ssrA, and EtuAld+Ile$^{1-21}$-GFP-ssrA, as indicated. Samples stained with uranyl acetate. Scale bars represent 100 nm.

GrpAld$^{1-20}$ from the Grp MCP directs heterologous proteins to the Pdu MCP- The aldehyde dehydrogenase enzyme from the computationally predicted glycyl radical-based propanediol utilization (Grp) MCP [77] contains a putative N-terminal signal sequence, as identified by comparison to the known Pdu and Eut signal sequences (Figure 3.1). This putative signal sequence is referred to here as GrpAld$^{1-20}$. We hypothesized that this signal sequence would also mediate encapsulation of GFP in the Pdu MCP. The induction of a pBAD-*grpAld$^{1-20}$-gfp-ssrA* plasmid with 0.02% arabinose in *S. enterica* LT2 Δ*eutR::FRT* yielded punctate fluorescence when 1,2-propanediol was added to the culture media (Figure 3.3). An anti-GFP western blot of purified Pdu MCPs from *S. enterica* LT2 Δ*eutR::FRT* expressing GrpAld$^{1-20}$-GFP-ssrA confirmed that the putative Grp signal sequence localizes protein to the Pdu MCP (Figure 3.4), as we observed for the EutC$^{1-20}$-GFP-ssrA and EutE$^{1-20}$-GFP-ssrA constructs. Transmission electron micrographs again confirmed that Pdu MCPs are morphologically normal when GrpAld$^{1-20}$-GFP-ssrA is encapsulated (Figure 3.4). *S. enterica* LT2 does not possess the Grp MCP genes, so the use of the *S. enterica* LT2 Δ*eutR::FRT* strain was sufficient to ensure that the puncta observed arose from Pdu MCPs alone.

Recapitulation of the hydrophobic motif confers signal sequence function on a peptide that did not previously mediate encapsulation- The aldehyde dehydrogenase enzyme Ald1 from the predicted ethanol utilization (Etu) MCP [43,77] also bears a putative N-terminal signal sequence, here referred to as EtuAld$^{1-20}$. We found that this peptide does not mediate detectable encapsulation of the heterologous fluorescent reporter protein EtuAld$^{1-20}$-GFP-ssrA in the Pdu MCP. We noted that the Etu signal sequence had an apparent gap that disrupted the hydrophobic motif, so we inserted an Ile residue at position 11 of the putative EtuAld sequence (denoted EtuAld+Ile$^{1-21}$) in order to recapitulate the putative signal sequence motif discussed above (Figure 3.1). Remarkably, this engineered peptide functions as a Pdu signal sequence, as indicated by the appearance of punctate fluorescence when we express the peptide as a genetic fusion to GFP under Pdu MCP-forming conditions (Fig. 3). Localization was confirmed by an anti-GFP western blot of purified Pdu MCPs from *S. enterica* LT2 Δ*eutR::FRT* expressing EtuAld+Ile$^{1-21}$-GFP-ssrA (Fig. 4). Transmission electron micrographs confirmed that Pdu MCPs are morphologically normal when EtuAld+Ile$^{1-21}$-GFP-ssrA is encapsulated (Figure 3.4). Again, since *S. enterica* LT2 does not possess the Etu MCP genes, the use of the *S. enterica* LT2 Δ*eutR::FRT* strain ensures that the puncta are attributable to Pdu MCPs.

A signal peptide from a smaller MCP does not direct heterologous proteins to the Pdu MCP- A smaller MCP (approximately 40 nm in diameter) of unknown function from *Haliangium ochraceum* has been heterologously expressed in *E. coli*, and its N-terminal signal peptide characterized [49]. The first 20 amino acids of the *H. ochraceum* N-terminal signal peptide (HochAld$^{1-20}$) also include the characteristic hydrophobic motif. We tested this signal peptide for the ability to target GFP to the Pdu MCP, but found no evidence of GFP encapsulation as assessed by fluorescence microscopy (data not shown).

Non-native Pdu-localized signal sequences compete with the native Pdu signal sequences for encapsulation- In order to investigate whether the localization of the various signal sequences tested above to the Pdu MCP is mediated by the same interaction with the Pdu shell proteins as mediates the encapsulation of PduD$^{1-20}$- and PduP$^{1-18}$-tagged proteins, we devised a competition assay for Pdu targeting. We constructed translational fusions of each signal peptide to PduD$^{21-224}$-FLAG on an aTc-inducible pTET plasmid, and then co-expressed these proteins at varying expression levels with a constant level of PduD$^{1-20}$- and PduP$^{1-18}$ -tagged GFP-ssrA expression from an arabinose-inducible pBAD plasmid. We measured the associated levels of PduD$^{1-20}$-

GFP-ssrA or PduP$^{1-18}$-GFP-ssrA encapsulation by flow cytometry [84]. If the GFP fluorescence decreases with increasing levels of competitor expression, this indicates competition between the two proteins for encapsulation, and suggests that the proteins may be encapsulated via the same mechanism or interaction.

We found that EutC$^{1-20}$ and EutE$^{1-20}$ compete significantly (p<0.05) with PduD$^{1-20}$-GFP-ssrA for encapsulation as compared to a negative control PduD$^{21-224}$-FLAG, as did GrpAld$^{1-20}$, EtuAld+Ile$^{1-21}$, and the positive control PduD$^{1-224}$-FLAG (Figure 3.5). As expected, EtuAld$^{1-20}$ and HochAld$^{1-20}$ exhibit no competition by this assay since they do not function as Pdu signal sequences (Figure 3.5). The assay using PduP$^{1-20}$-GFP-ssrA was somewhat less sensitive, likely due to lower expression of the PduP$^{1-20}$-tagged construct, but significant competition (p<0.05) was still observed for EutC$^{1-20}$, and for the positive control PduD$^{1-224}$-FLAG. Flow cytometry results were confirmed for the EutC$^{1-20}$-PduD$^{21-224}$-FLAG case by western blot of purified Pdu MCPs and cell lysates against GFP and against the FLAG epitope, which show the same trends as the flow cytometry data (Figure 3.6). We therefore conclude that all of the Pdu- localized sequences compete with the native Pdu signal sequences for encapsulation.

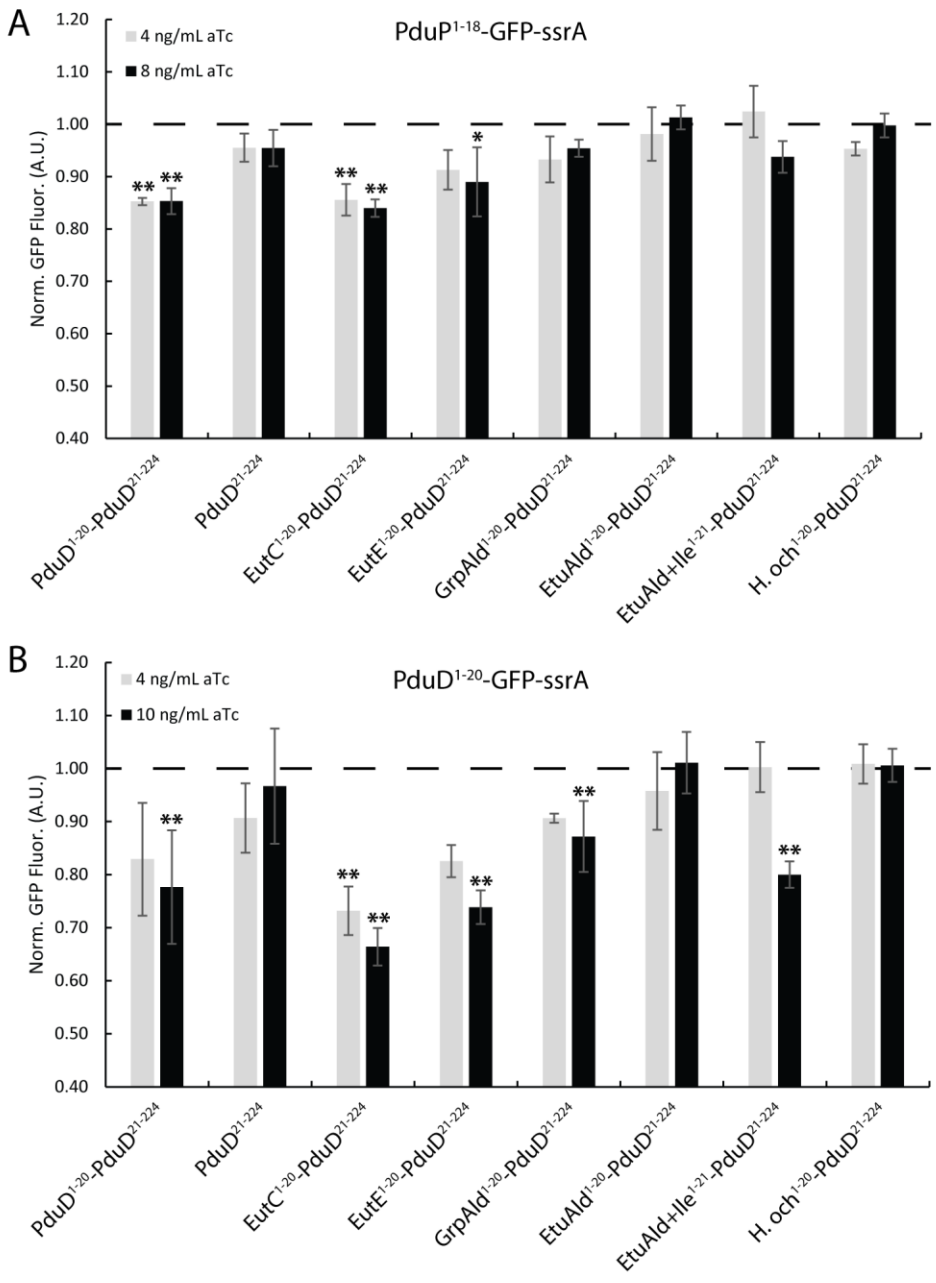


Figure 3.5: Several non-native targeting sequences appear to compete with PduP$^{1-18}$- and PduD$^{1-20}$–tagged fusions for localization to Pdu MCPs in *S. enterica*. Flow cytometry of *S. enterica* LT2 Δ*eutR* cultures grown in the presence of 55 mM 1,2-PD coexpressing (A) the PduP$^{1-18}$-GFP-ssrA encapsulation reporter or (B) the PduD$^{1-20}$-GFP-ssrA encapsulation reporter and N-terminal signal sequence fusions to PduD$^{21-224}$, as indicated. Expression of the fluorescent encapsulation reporter was induced with 0.02% arabinose (ara) in all cases, and expression of the competitor proteins was induced with 4, 8, or 10 ng/mL anhydrous tetracycline (aTc), as indicated, at OD_{600}~0.4. Values shown are the arithmetic mean of the geometric mean fluorescence of three independent replicates as measured 5.5 hours after induction, normalized to the fluorescence when no aTc was added; error bars indicate one standard deviation. * indicates

significance of $p < 0.1$ and ** indicates significance of $p < 0.05$ by Student's one-tailed T-test as compared to the PduD$^{21-224}$ control at the relevant induction level.

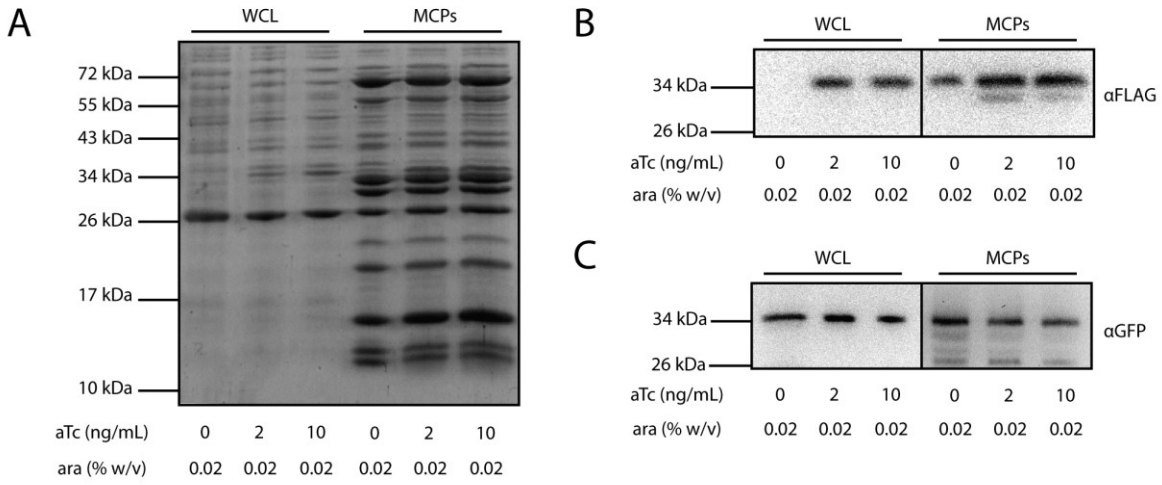


Figure 3.6: Characterization of whole-culture lysate and purified Pdu MCPs from cells expressing an encapsulation reporter and a competitor protein. (A) Coomassie stain of an SDS-PAGE separation of purified MCPs purified from *S. enterica* LT2 expressing PduD$^{1-20}$-GFP-ssrA and bearing a secondary aTc-inducible plasmid encoding EutC$^{1-20}$-PduD$^{21-224}$. Whole culture lysate (WCL) was reserved prior to purification and prepared by heating whole culture in Laemmli buffer at 95ºC for 5 minutes. Inducing molecules (55 mM 1,2-PD, 0.02% arabinose, and aTc as indicated) were added at OD600~0.4. Loading normalized by culture OD_{600} for WCL and by total protein concentration as judged by BCA assay for MCP samples. (B) Western blot of the same samples against the FLAG epitope. (C) Western blot of the same samples against GFP.

S. enterica *LT2 Δ*pduP::eutE *retains the ability to grow on 1,2-PD as the sole carbon source*- We speculated that, since EutE$^{1-20}$-GFP-ssrA is localized to the Pdu MCP, EutE might likewise be encapsulated in the Pdu MCP *in vivo* if the two MCP systems were expressed contemporaneously in a cell. Furthermore, a multiple sequence alignment reveals that the full-length *pduP* and *eutE* aldehyde dehydrogenase genes are homologs with 45% sequence identity. We therefore tested whether *S. enterica* LT2 Δ*pduP::eutE* retained the ability to grow on 1,2-PD as the sole carbon source, presumably with EutE encapsulated in the Pdu MCP and carrying out the metabolic function of PduP. We found that *S. enterica* LT2 Δ*pduP::eutE* retains the ability to grow on 1,2-PD as the sole carbon source in NCE media, exhibiting significantly greater growth than a control strain *S. enterica* LT2 Δ*pduP::cat/sacB*, although growth was somewhat slower than that of the wild type (Figure 3.7). A Δ*pduP::cat/sacB* strain retained some growth, as previously reported, but growth of the knockout was significantly slower than that of the wild type and the Δ*pduP::eutE* strain [85]. The observed residual growth in the Δ*pduP::cat/sacB* strain as compared to a Δ*pocR::FRT* strain (which forms no Pdu MCPs) is most likely due to cytosolic aldehyde dehydrogenases acting on propionaldehyde which escaped the Pdu MCP. We therefore conclude that EutE is likely encapsulated in the Pdu MCP when expressed from the

Pdu locus, and can complement the metabolic function of PduP, albeit with slightly decreased growth.

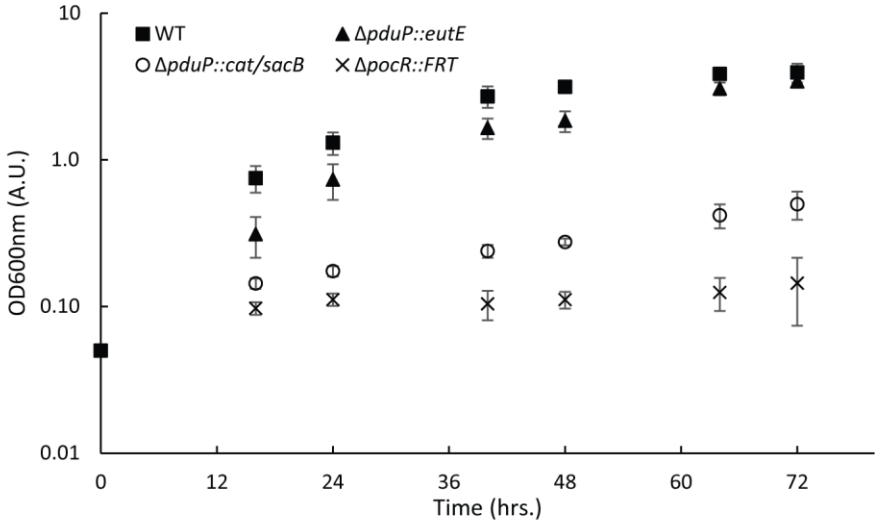


Figure 3.7: Replacing *pduP* with *eutE* in the Pdu locus can partially restore growth of *S. enterica* Δ*pduP* on 1,2-PD. Cultures of wild type *S. enterica* LT2, *S. enterica* LT2 Δ*pocR::FRT, S. enterica* LT2 Δ*pduP::cat/sacB,* and *S. enterica* LT2 Δ*pduP::eutE* were grown overnight in LB, then resuspended to an OD_{600}~0.05 in NCE media supplemented with 55 mM 1,2-PD and 150 nM vitamin B12. OD_{600} was subsequently measured at the time points indicated. Values shown are the arithmetic mean of three independent replicates; error bars indicate one standard deviation.

The P_{pdu} promoter is activated by 1,2-PD in the presence of ethanolsamine and vitamin B12, and the P_{eut} promoter is activated by ethanolamine and vitamin B12 in the presence of 1,2-propanediol- Given the apparent interaction of various signal sequences with the Pdu MCP shell proteins, we wondered what cellular mechanisms might be in place to avoid cross-localization of compartmentalized enzymes such as those observed in the Δ*pduP::eutE* strain. We hypothesized that transcriptional regulation may limit expression to only one compartment system in each cell, when conditions would otherwise induce the expression of multiple types of MCPs. To test this idea, we made use of a set of reporters for promoter activity. We previously constructed a fluorescent reporter of P_{pdu} promoter activation [58], and in this study we coupled that reporter with an analogous fluorescent reporter of P_{eut} transcriptional activation in order to compare the induction of P_{pdu} and P_{eut} in response to 1,2-propanediol and to ethanolamine and vitamin B12. The *P_{pdu}-gfpmut2* reporter is identical in sequence to that reported previously, consisting of the 373 bases immediately 5' of the *pduA* open reading frame transcriptionally fused to *gfpmut2* [58]. The *P_{eut}-gfpmut2* reporter contains a 300-base portion of the promoter of the *S. enterica* Eut operon immediately 5' of the *eutS* open reading frame transcriptionally fused to *gfpmut2*.

We measured the response of each reporter to the presence of 1,2-propanediol, to the presence of ethanolamine and vitamin B12, and to neither or both conditions by sampling cultures grown in appropriate media hourly and interrogating the samples by flow cytometry [80]. The P_{pdu} reporter showed activation in response to 55 mM 1,2-propanediol, but not in response to 30 mM ethanolamine and 150 nM vitamin B12, as expected (Figure 3.8;

representative histograms shown in Figure 3.9). Similarly, the P_{eut} reporter showed the expected activation in response to 30 mM ethanolamine and 150 nM vitamin B12, but not in response to 30 mM 1,2-propanediol (Figure 3.8).

Significantly, we observed that both promoters retain activation over the course of the experiment upon the addition of both sets of inducing molecules (1,2-PD, and ethanolamine and vitamin B12), and we conclude that there are no readily apparent mechanisms to abrogate the transcription of one MCP polycistron when the other is expressed (Figure 3.8).

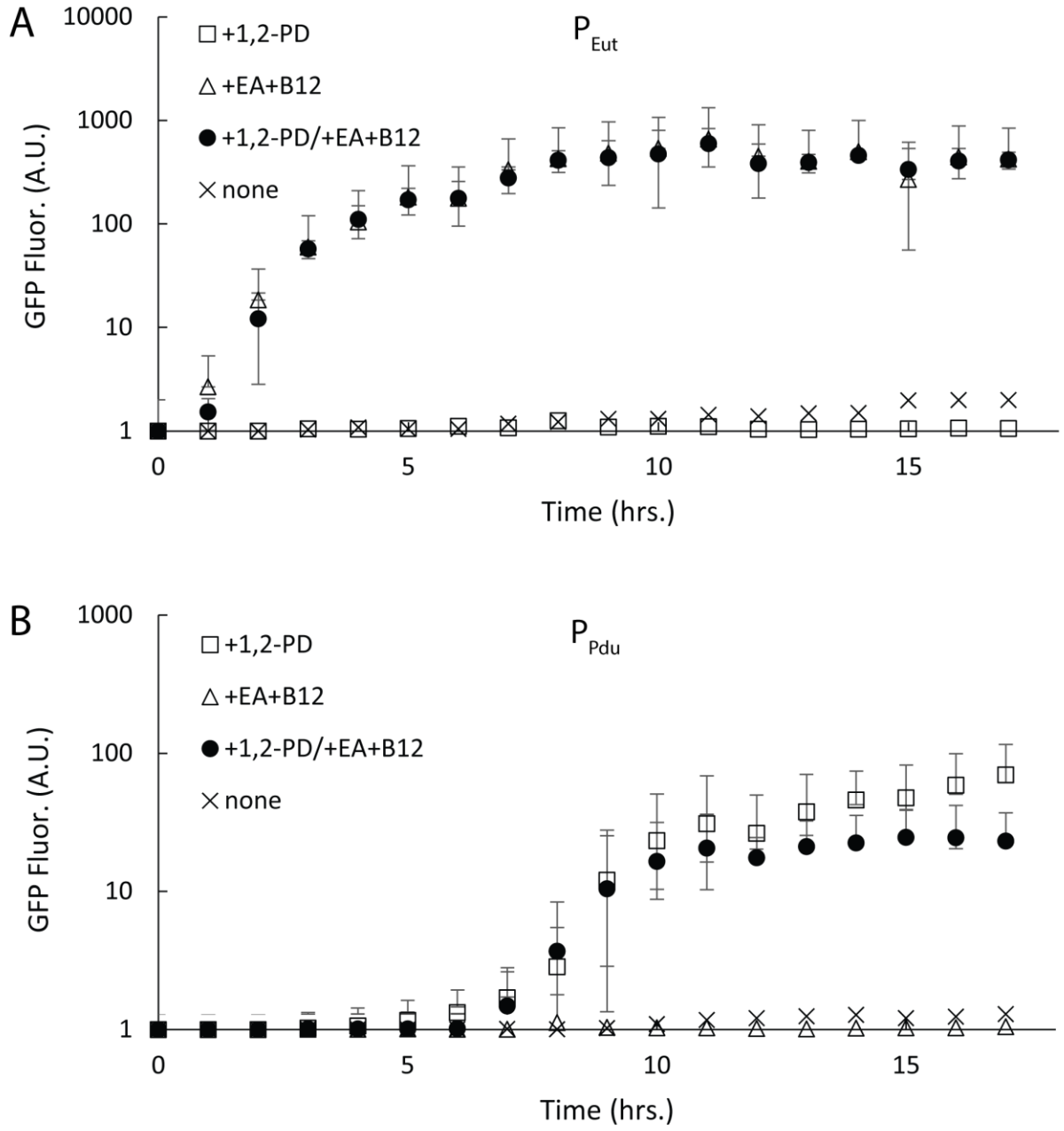


Figure 3.8: P_{pdu} and P_{eut} are transcriptionally activated when both 1,2-PD and ethanolamine are present. Flow cytometry of *S. enterica* LT2 cultures bearing plasmids

encoding (A) the P_{eut}-*gfpmut2* or (B) the P_{pdu}-*gfpmut2* reporter constructs. Inducing molecules (55 mM 1,2-PD, 30 mM ethanolamine, and 150 nm vitamin B12, as indicated) were added at OD_{600}~0.4. Values shown are the arithmetic mean of the geometric mean fluorescence of three independent replicates measured at one-hour intervals after induction, as indicated; error bars indicate one standard deviation. Representative histograms can be found in Figure 9.

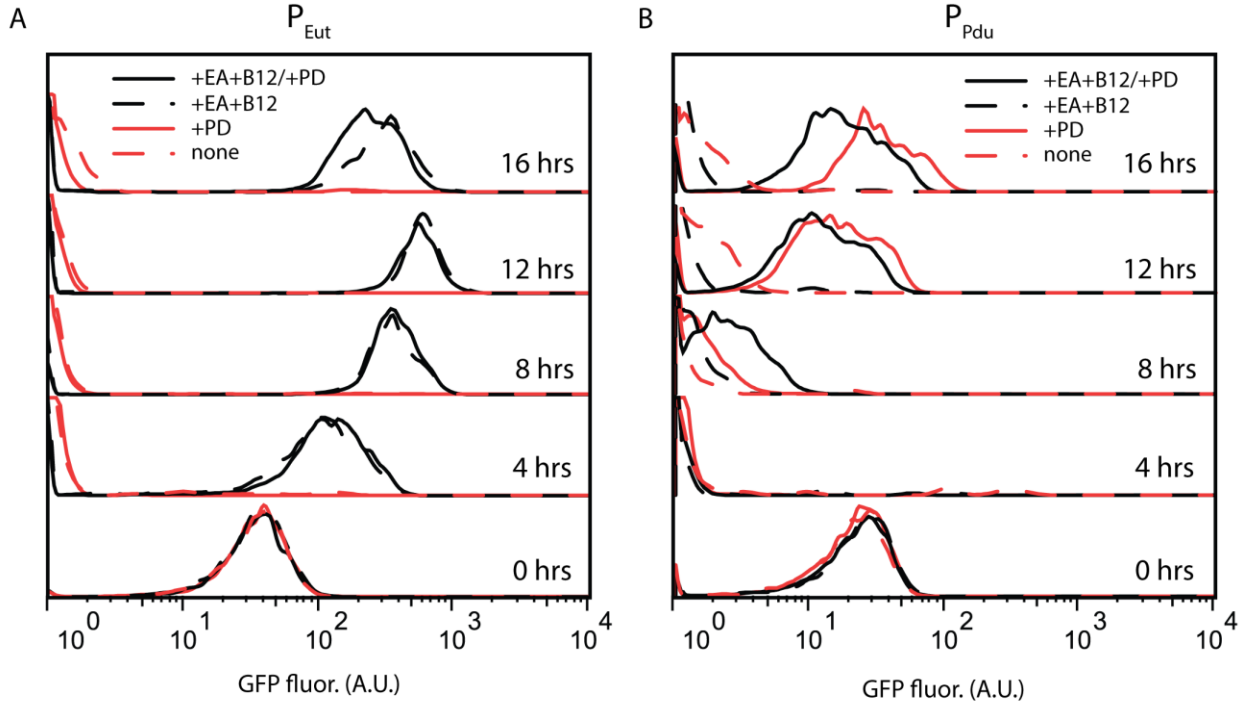


Figure 3.9: Flow cytometry histograms of *S. enterica* expressing fluorescent reporters of P_{pdu} and P_{eut} transcriptional activation. Flow cytometry of *S. enterica* LT2 cultures bearing plasmids encoding (A) the P_{eut}-*gfpmut2* or (B) the P_{pdu}-*gfpmut2* reporter constructs. Inducing molecules (55 mM 1,2-PD, 30 mM ethanolamine, and 150 nm vitamin B12, as indicated) were added at OD_{600}~0.4 (0 hours). Representative histograms are shown from the 0, 4, 8, 12, and 16 hour time points. The gain at the 0 hour time point is increased relative to later time points to ensure that samples are identical prior to induction.

Manipulating signal sequence and induction level allows the tuning of cargo protein ratios- We next explored whether combinations of cargo proteins bearing different N-terminal signal sequences cause different ratios of cargo proteins to be loaded to the Pdu MCP. In order to make these measurements, we employed the same fluorescent encapsulation reporter scheme described above, but in this case used two fluorophores (GFPmut2 and mCherry) simultaneously, each bearing an N-terminal signal sequence and a C-terminal ssrA degradation tag. This approach allowed us to quantify encapsulation of both proteins simultaneously using measurements of cellular fluorescence. We tested several combinations of cargo proteins (*pduP*$^{1\text{-}18}$*-mCherry-ssrA* and *pduD*$^{1\text{-}20}$*-mCherry-ssrA* each in combination with *pduP*$^{1\text{-}18}$*-gfp-ssrA*, *pduD*$^{1\text{-}20}$*-gfp-ssrA, eutC*$^{1\text{-}20}$*-gfp-ssrA,* or *eutE*$^{1\text{-}20}$*-gfp-ssrA*) with varying levels of transcriptional activation for each cargo protein, as modulated by varying concentrations of inducing molecules. As above, increasing ratios of fluorescence in the case with Pdu MCPs as compared to the case

without Pdu MCPs indicate increased cargo encapsulation. We further characterized the relative ratio of the two cargo fluorophores by taking the quotient of these two independent fluorescence ratios. This measurement is not an absolute measurement of the stoichiometric ratio of the cargo proteins, but is indicative of the stoichiometric ratio, and can be made in high throughput in order to characterize many signal-sequence and induction-level combinations.

The encapsulation of tagged GFP-ssrA protein under the control of the pBAD promoter, as indicated by the fluorescence ratio, increases with increasing concentrations of arabinose, as does the encapsulation of tagged mCherry-ssrA under the control of the pTET promoter with increasing concentrations of anhydrous tetracycline (Figure 3.10).Notably, the fluorescence ratios observed for each reporter do not vary significantly with respect to the induction of the other reporter protein (Figure 3.10). We calculated the relative ratio of the fluorescence ratios for these various combinations of GFP and mCherry reporters as a proxy for the stoichiometric ratio of the two cargo proteins, and found that a wide range of relative cargo fluorescence ratios can be achieved using the native Pdu signal sequences and the $EutC^{1-20}$ and $EutE^{1-20}$ sequences (Figure 3.11). The observations made by fluorescence for the $PduD^{1-20}$-GFP-ssrA/$EutE^{1-20}$-mCherry-ssrA reporter pair were confirmed by MCP purification and western blot against GFP and mCherry. As observed using fluorescence measurements, encapsulation of the $PduD^{1-20}$-GFP-ssrA or $EutE^{1-20}$-mCherry-ssrA protein was not significantly decreased by encapsulation of the other reporter protein (Figure 3.12).Colocalization of $PduD^{1-20}$-GFP-ssrA and $EutE^{1-20}$-mCherry-ssrA to Pdu MCPs in *S. enterica* cells expressing both reporters was confirmed by fluorescence microscopy (Figure 3.12).

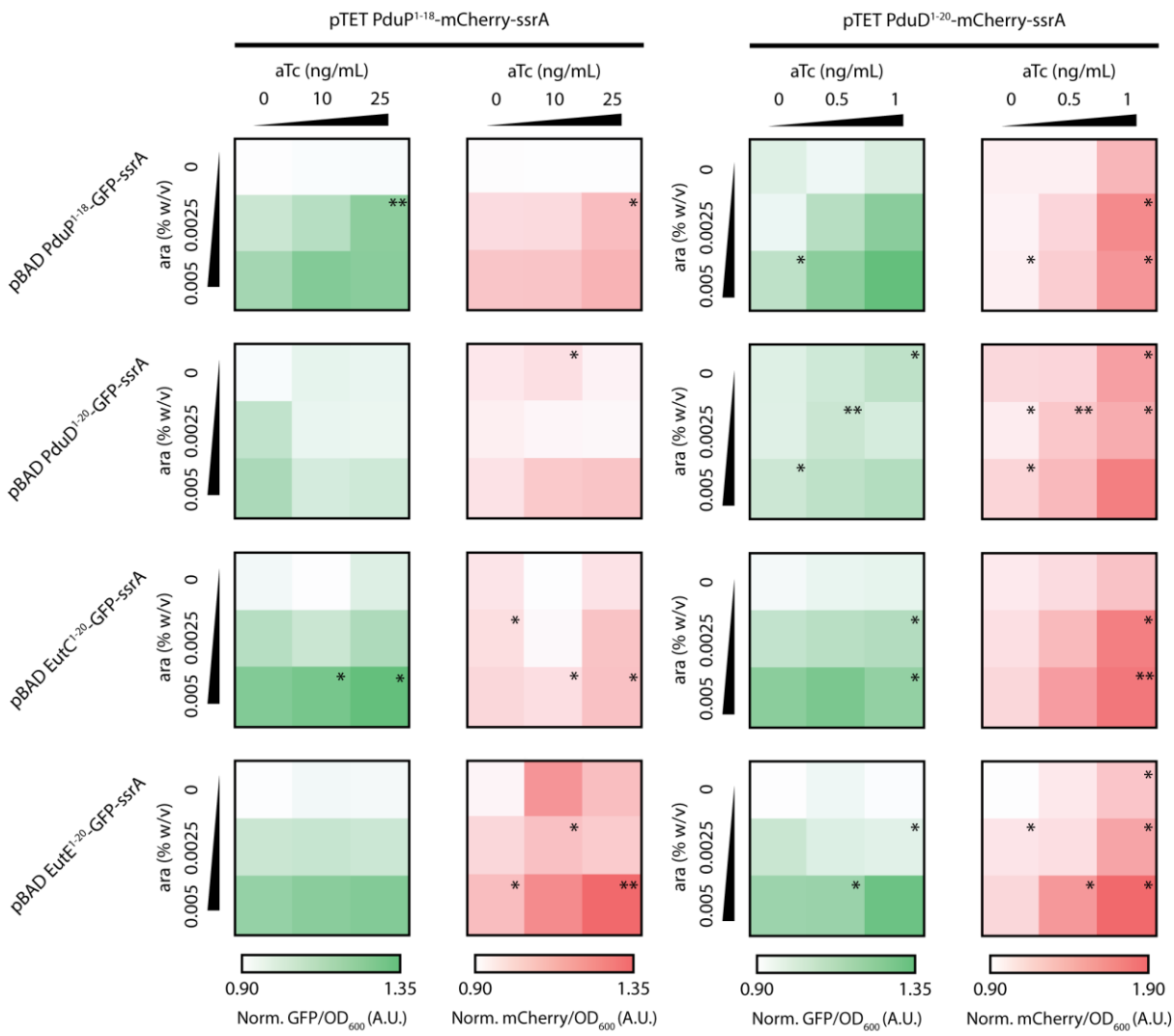


Figure 3.10: Fluorescence measurements of *S. enterica* expressing two fluorescent reporters of encapsulation. Plate reader fluorescence measurements of *S. enterica* LT2 cultures bearing plasmids encoding fluorescent reporters of encapsulation as indicated. Inducing molecules (55 mM 1,2-PD, arabinose, and aTc, as indicated) were added at OD_{600}~0.4. Values shown are the green (left column) or red (right column) fluorescence normalized to OD_{600} and subsequently to the fluorescence/OD_{600} of a culture to which no 1,2-PD was added. Values are the arithmetic mean of measurements from three independent replicates. The standard deviation of all samples was less than 0.12, apart from samples labeled with the * symbol, for which the standard deviation was less than 0.20, and samples labeled with the ** symbol, for which the standard deviation was less than 0.30.

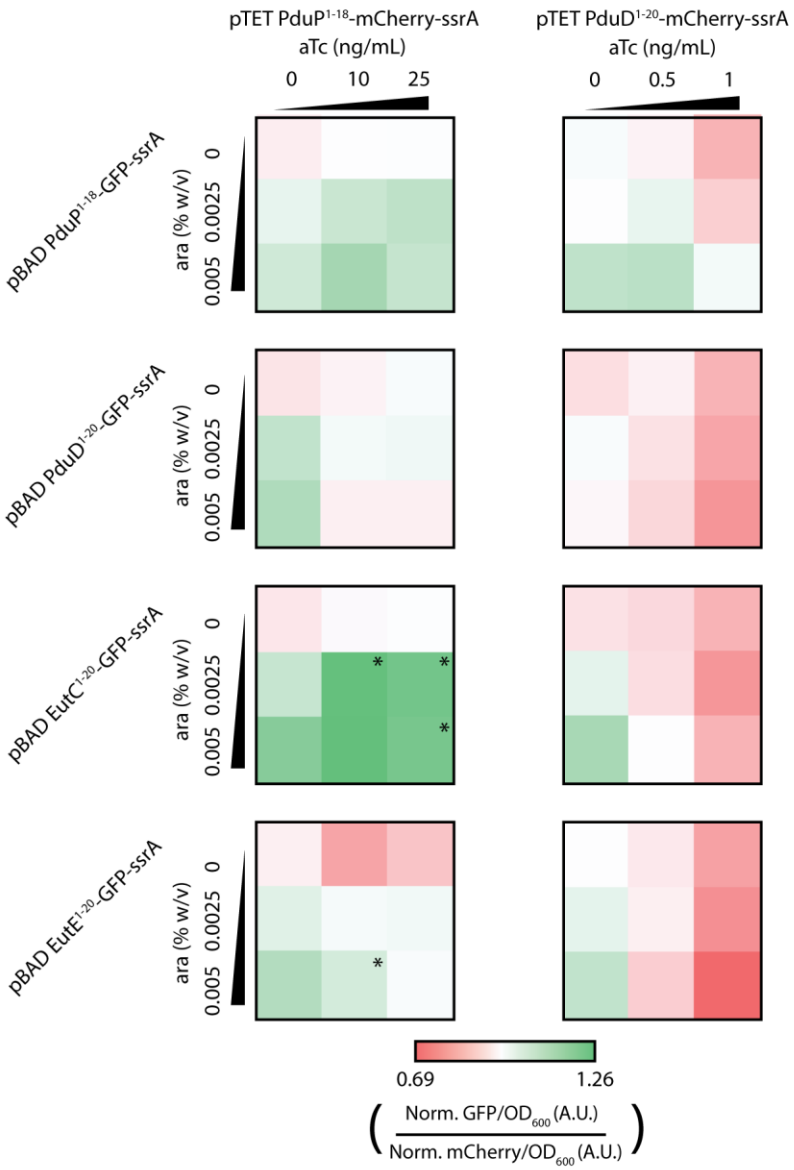


Figure 3.11: Relative fluorescence measurements of *S. enterica* expressing two fluorescent reporters of encapsulation reveal that a range of cargo ratios can be achieved. Plate reader fluorescence measurements of *S. enterica* LT2 cultures containing plasmids encoding fluorescent reporters of encapsulation as indicated. Inducing molecules (55mM 1,2-PD, arabinose, and aTc, as indicated) were added at OD_{600}~0.4. Values shown are the quotient of the normalized green and red fluorescence of each culture (see Figure 9). Higher values indicate a greater ratio of green to red normalized fluorescence; lower values indicate a greater ratio of red to green normalized fluorescence. Values are the arithmetic mean of measurements from three independent replicates. The standard deviation of all samples was less than 0.10, apart from samples labeled with the * symbol, for which the standard deviation was less than 0.20.

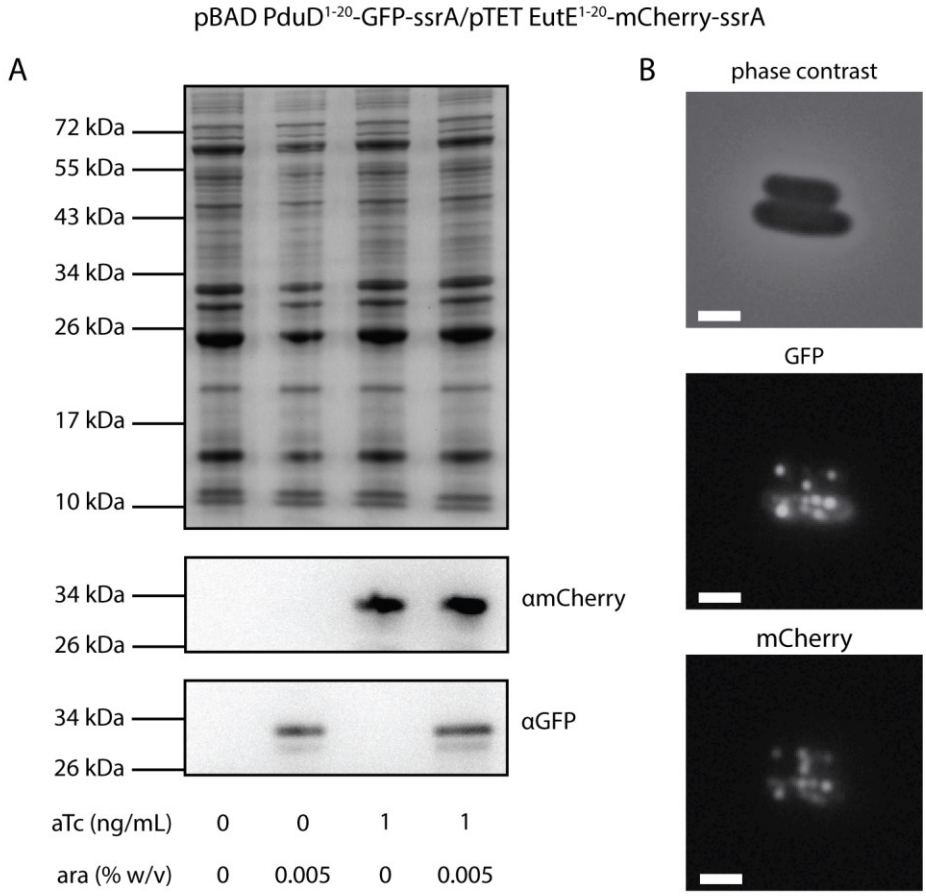


Figure 3.12: Characterization of purified MCPs from *S. enterica* expressing two fluorescent encapsulation reporters. (A) Top, Coomassie stain of an SDS-PAGE separation of purified MCPs purified from *S. enterica* LT2 bearing an arabinose-inducible plasmid encoding PduD$^{1-20}$-GFP-ssrA and bearing a secondary aTc-inducible plasmid encoding EutE$^{1-20}$-mCherry-ssrA. Inducing molecules (55 mM 1,2-PD, arabinose, and aTc, as indicated) were added at OD600~0.4. Loading normalized by total protein concentration as judged by BCA assay. Below, western blot of the same samples against mCherry. Bottom, western blot of the same samples against GFP. (B) Phase contrast and fluorescence microsopy of *S. enterica* LT2 expressing PduD$^{1-20}$-GFP-ssrA and EutE$^{1-20}$-mCherry-ssrA. Inducing molecules (55 mM 1,2-PD, 0.005% arabinose, and 1 ng/mL aTc) were added at OD600~0.4. Scale bars represent 1 μm.

3.4 Discussion

We demonstrated that heterologous proteins tagged with several known or putative N-terminal signal sequences from the Eut and Grp bacterial MCP systems can be localized to the Pdu MCP. These signal sequences include EutC$^{1-20}$, a known signal sequence for the Eut MCP, and EutE$^{1-20}$, a peptide that has not previously been shown to localize heterologous proteins to the Eut MCP. The fact that the aldehyde dehydrogenase EutE has an N-terminal signal sequence is not surprising considering its homology and functional similarity to the encapsulated Pdu MCP aldehyde dehydrogenase PduP. Rather than the I10/L14 motif of the PduP N-terminus, however, the Eut signal sequences bear the motifs V10/M14 and V10/L14 at the amino acid positions

previously shown to be most important for encapsulation (Figure 3.1) [16,84]. We further tested the putative N-terminal signal sequence from the aldehyde dehydrogenase of the Grp MCP, and found that it, too, mediates the encapsulation of proteins in the Pdu MCP. This finding aligns well with the hypothesis that MCP systems are evolutionarily centered around "signature" enzymes, such as aldehyde dehydrogenases, which must be encapsulated and form the basis of many evolutionarily related MCP operons [10,11,86]. The N-terminal signal sequence from the smaller *H. ochraceum* MCP did not mediate Pdu MCP encapsulation as judged by fluorescence microcscopy, perhaps due to structural differences between Pdu shell proteins and the shell proteins of its smaller compartment, particularly at the C-termini of the putative binding partners.

The EtuAld+Ile$^{1-21}$ peptide, which recapitulated the appropriate hydrophobic motif upon insertion of Ile at position 11, also localizes GFP to the Pdu MCP. This hydrophobic motif alone, however, is apparently not sufficient for a peptide to function as a Pdu MCP signal sequence. Alignment of the first 20 amino acids of GFP to the signal peptides (Figure 3.1) demonstrates that the N-terminus of a non-localizing protein can strongly resemble the motif in question, but not confer encapsulation. This suggests that alpha-helical structure, as has been shown by NMR for the PduP$^{1-18}$ signal sequence [52], in addition to primary amino acid sequence, is important to the signal sequence-shell protein interaction that mediates encapsulation. The signal sequences which we show here to be localized to the Pdu MCP are all predicted to form alpha helices by the Jpred 4 secondary structure prediction tool, whereas the first 20 amino acids of GFP are not [87]. Interestingly, the *H. ochraceum* N-terminal signal sequence is predicted to adopt an alpha-helical conformation and includes the appropriate hydrophobic motif, but does not mediate localization. Together, this evidence suggests that alpha-helical structure and the appropriate hydrophobic residue motif are necessary, but not sufficient, for Pdu MCP encapsulation by N-terminal signal sequences. Recent computational studies confirm that an N-terminal amphipathic, alpha-helical motif is widespread among encapsulated protein genes in Pdu and Eut loci in various organisms [54].

In order to investigate whether the various Pdu-localized signal sequences interact with the Pdu MCP by the same mechanism as the native Pdu N-terminal signal sequences, we tested whether these signal sequences competed with the Pdu signal sequences PduD$^{1-20}$ and PduP$^{1-18}$ for encapsulation. We found that they do compete for encapsulation, suggesting that the encapsulation of proteins bearing non-native signal sequences in Pdu MCPs is mediated by the same signal sequence-shell protein interaction as native Pdu enzyme encapsulation. It has previously been demonstrated computationally that multiple Pdu cargo enzymes may interact with PduA or PduJ [73]. The localization of Eut and Grp signal sequences to the Pdu MCP is perhaps not surprising considering the similarity in C-terminal amino acid sequence between the PduP binding partner PduA, the Pdu shell protein PduJ, the Eut shell protein EutM, and the Grp MCP shell protein homologue denoted Cbei_4058 in the *C. beijerinckii* genome (Figure 3.1). If the C-termini of the shell proteins of these various compartments all interact with the N-termini of their respective cargo in the same manner as the Pdu shell proteins interact with their cargo enzymes, then the similarities in C-terminal amino acid sequences of the shell proteins dictate the similar patterns of residues at the N-termini of cargo proteins. For this reason, we suspect that the Pdu-localized signal sequences in this study interact with PduA, PduJ, or both proteins. It remains possible that non-native signal sequence-tagged proteins are encapsulated via another interaction, but still compete with native Pdu signal sequences due to effects of steric exclusion during Pdu MCP loading; the assays described here cannot discriminate between these two possibilities.

The localization of proteins to the Pdu MCPs by the N-terminal signal sequences from other MCP systems raises important microbiological and engineering questions with respect to the function and use of bacterial MCPs. From a microbiological perspective, we speculated that a regulatory mechanism may dictate that each *S. enterica* cell expresses only one type of MCP (Pdu or Eut) at a time, avoiding the potential for mislocalization. Using fluorescence-based reporters of transcriptional activation, we found no evidence of such a mechanism at the transcriptional level, as both the P_{pdu} and P_{eut} promoters are activated similarly by their cognate inducing molecules in the presence and absence of the other operon's inducing molecule. We therefore conclude that it is possible for a *S. enterica* cell in the gut endothelium to encounter 1,2-propanediol and ethanolamine simultaneously and form both the Pdu and Eut MCPs. Both of these metabolites have been shown to contribute to bacterial proliferation in models of *Salmonella* and *E. coli* infection, so it may be advantageous for a cell to metabolize them both concurrently [17,88]. Furthermore, since the Pdu operon is coregulated with the Cob vitamin B12 synthesis operon by PocR, and the Eut operon is not, the coexpression of both MCP systems along with the Cob operon would allow the Eut MCP access to its requisite vitamin B12 cofactor.

In the absence of regulation to prevent MCP system co-expression in an individual cell, it is important to explore whether the catalytic enzymes associated with the Pdu and Eut MCPs are exclusively localized to their cognate MCP shells, as previously assumed, or whether, upon co-expression of both the Pdu and Eut MCPs, enzymes from each operon are localized to both types of MCPs. While we do not directly demonstrate here that the N-terminal signal sequences from Pdu enzymes localize heterologous proteins to the Eut MCPs, we believe that further investigation will confirm this phenomenon. Promiscuous localization of Eut proteins to the Pdu MCP may be avoided *in vivo*, however, because the heterologous fusions tested above are expressed at levels significantly higher than those of native cargo proteins. On the other hand, we demonstrated that *S. enterica* LT2 Δ*pduP::eutE* can grow on 1,2-PD as its sole carbon source, indicating that EutE can serve the same metabolic function as PduP in the Pdu MCP and suggesting that a natively encapsulated protein such as EutE can be encapsulated in a non-native compartment such as the Pdu MCP when expressed at near-native levels. If this behavior extends to the other cargo enzymes and also for Pdu enzymes targeted to the Eut MCP, this could indicate that selective transport of certain metabolites attributable to the characteristics of shell protein pores, *e.g.* of 1,2-propanediol over propionaldehyde [38,46,89], is not as important for MCP function as general retardation of diffusion [90], sequestration of cofactor pools [6,40], or simple colocalization of the relevant metabolic enzymes [8,91–93].

It is important to note that the replacement of *pduP* in favor of *eutE* relied on a non-native genomic context. It is possible that the genomic context of the cargo proteins is important to localization; that is, that enzymes are encapsulated within the MCPs forming from shell proteins expressed from the same genetic locus. If cargo proteins quickly and strongly associate with their cognate shell protein binding partners upon expression, Pdu enzymes may be confined to the Pdu MCP simply by virtue of being translated from the same mRNA. The crosstalk observed here for cargo encoded on plasmids would then be an informative and interesting engineering observation, but not a biologically relevant phenomenon.

We demonstrate an engineering application of these signal sequences by characterizing the ratio of two fluorescent cargo proteins encapsulated in the Pdu MCP when they are encapsulated using various signal sequences and expressed at various levels. Different combinations of signal sequences allow the loading of the Pdu MCPs with different absolute and

relative amounts of each cargo protein. We show that the stoichiometric ratio of two cargo proteins in the Pdu MCPs can be controlled by altering N-terminal signal sequences and expression levels. No previous study has demonstrated systematic measurement or control of these ratios, and we anticipate that future studies will reveal that the performance of heterologous enzymatic pathways is dependent on the tuning of this cargo-protein ratio using the approaches described here. Interestingly, we did not observe competition between the tagged fluorophores for localization; we therefore suspect that competition is not a significant factor in cargo localization at the moderate expression levels required for our dual fluorophore assay (which requires that all non-encapsulated reporter proteins be degraded by a limited number of ClpXP proteases). In contrast, expression levels of the signal sequence-tagged Pdu^{21-224} protein in our competition assay could be increased without regard for ClpXP degradation capacity until competition was observed.

From an engineering perspective, the ability of non-native signal sequences to target cargo to the Pdu MCP presents challenges to the development of a system with multiple, orthogonal MCPs in a single cell. The investigation presented here grew from an effort to localize proteins exclusively to either the Pdu or Eut MCPs within single cells; this proved impossible using currently available signal peptides due to the localization of Eut signal sequences to the Pdu MCP as detailed above. The strong sequence similarity between the C-termini of homologous shell proteins in the Pdu and Eut systems, for example (Figure 3.1), suggests that any efforts to generate orthogonal signal sequence-shell protein pairs will likely require engineering both the N-terminal signal sequences themselves and their cognate C-terminal binding partners in order to create orthogonal interactions, presenting a significantly greater challenge than the development of novel signal sequences alone. It is possible, however, that the emerging MCP system from *H. ochraceum* may offer such an orthogonal system, if the interaction of its N-terminal signal sequence is indeed mediated by a different interaction than that which mediates Pdu MCP encapsulation. We also anticipate that the common hydrophobic residue motif observed for the various Pdu-localized sequences will inform the discovery of a wide array of functional signal sequences by library-based protein engineering techniques for applications that do not require orthogonal MCPs.

Chapter 4: Controlling small molecule diffusion in a virus-like particle nanoreactor

The following is adapted from Glasgow, Asensio, Jakobson, Francis, and Tullman-Ercek; ACS Synthetic Biology, 2015 with permission. The experimental portions of this chapter were included in Jeff Glasgow's dissertation; they are included here in order to provide context for the modeling aspects, which were the focus of my contribution.

4.1 Background

Many hypotheses seeking to explain Pdu MCP function invoke the control of small molecule transport by the ~1 nm diameter pores in the MCP shell. We therefore sought to characterize the possibility of the control of small molecule transport through the similarly sized pores of the simpler MS2 virus-like particle (VLP) model system.

Natural metabolic systems provide numerous examples of enzyme co-localization occurring through genetic fusion, direct binding interactions, scaffolding, or compartmentalization.[94–97] This is thought to facilitate pathway flux by increasing the local concentrations of intermediate molecules around downstream enzymes while limiting off-target reactions that can lead to unproductive accumulation or even toxicity.[90,98–100] Many researchers have attempted to recapitulate and use this effect *in vitro* and *in vivo* by adding artificial scaffolding,[7,8,91,92,101] localizing enzymes to specific organelles,[102,103] engineering the enzymes,[104–106] or encapsulating them in protein compartments[107–111] and polymersomes.[112–116] Despite numerous technological advances, one aspect of compartmentalization that has yet to be fully explored is the ability to influence the passive diffusion of substrates, products, and intermediates. Decreasing the diffusion of intermediates compared to substrates and products across the shell of an enzyme-containing compartment may be a method to enhance pathway flux without altering individual proteins.

Permeability effects were examined in a variety of inorganic and polymeric nanoreactors,[117–121] but only a handful of protein complexes have been studied for such properties. Ferritin cages, for example, direct Fe(II) ions for oxidation using negatively charged pores.[122] Natural membrane protein ion channels have both ion-specific and nonspecific electrostatic interactions with their substrates, affecting their transfer kinetics and selectivity.[123] The carboxysome, a protein-based bacterial microcompartment, is also hypothesized to use positively charged pores to allow passive diffusion of bicarbonate ions while retaining hydrophobic CO_2 and excluding molecular oxygen, which can lead to undesired side reactions.[124–126] Inspired by these natural systems, we sought to develop a protein-based nanoreactor with mutable pores to examine the effects of electrostatics on substrate and product flux (Figure 4.1). For this proof-of-concept study, we chose to analyze the effect on a single enzyme reaction, with the expectation that the approach can be expanded to multi-enzyme pathways in the future.

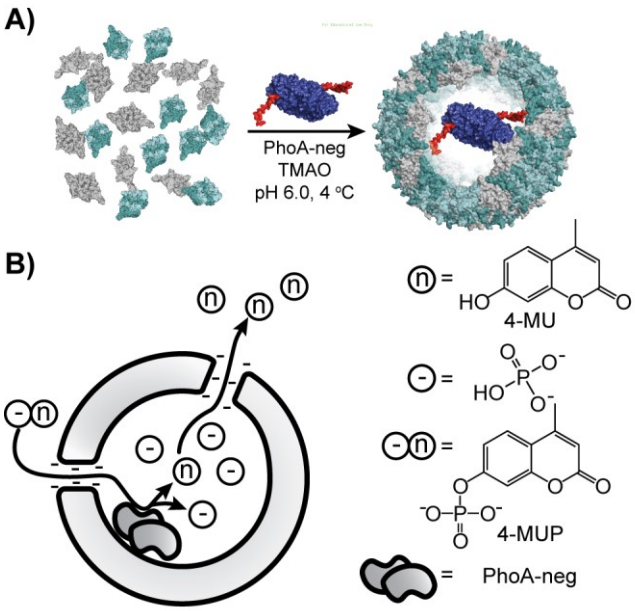


Figure 4.1: Testing the effects of pore charge on substrate and product flux. A) Bacteriophage MS2-based enzymatic nanoreactors are created through self-assembly of the capsid around negatively charged alkaline phosphatase (PhoA-neg) variants. B) Capsid pore charge is varied by mutagenesis, and reactions of negatively charged phosphatase substrates are monitored.

The bacteriophage MS2 forms an icosahedral capsid from 180 monomers that is 27 nm in diameter with triangulation number T = 3 and pores at its vertices. We previously developed the MS2 viral capsid as a stable, versatile compartment for encapsulation of negatively charged macromolecules enhanced by the presence of a protein-stabilizing osmolyte, trimethylamine-*N*-oxide (TMAO).[127,128] The capsid is also well-characterized for tolerance to mutations, particularly in the assembly-directing pore residues of the FG-loop.[129,130] This loop forms the inter-subunit contacts lining the 1.8 nm pores at the 5-fold and quasi-6-fold axes of the icosahedral capsid. Point mutations in this loop can thus add charge directly around the capsid pores, allowing controllable, genetically encoded modifications to the pore environment. Here, we show that capsids mutated in this way are able to encapsulate alkaline phosphatase enzymes bearing a localization tag. When these enzymes are inside capsids with pores having no charge or a charge opposite to that of the substrate, the apparent K_M ($K_{M,app}$) is similar to that of the free enzyme; however when pore and substrate charge are the same, the $K_{M,app}$ of the enzyme increases significantly. Kinetic modeling suggests this could be caused primarily by reduced product efflux, leading to inhibition of the enzyme. These experiments represent a first step in creating selective nanoreactors with the potential either to protect cargo by inhibiting entrance of interfering molecules or to enhance multi-step pathways by concentrating internal intermediates.

4.2 Materials and methods

Reagents. Unless otherwise noted, reagents were used as received from commercial sources. In all experiments, water was deionized using a MilliQ system (Millipore). Chemicals were obtained from Fisher Chemical except Tris base (VWR), TMAO

dihydrate (Alpha Aesar), triethanolamine (Sigma), 4-methylumbelliferyl phosphate (Invitrogen) and 4-methylumbelliferone (Sigma). Molecular biology reagents were obtained from New England Biolabs.

High Performance Liquid Chromatography. HPLC was performed using an Agilent 1100 series HPLC using ST buffer (50 mM Tris pH 7.2, 100 mM NaCl) or sodium phosphate buffer (10 mM, pH 7.2) as mobile phase. All size exclusion chromatography was performed using a Biosep SEC-4000 column (Phenomenex). Liquid Chromatography/Mass Spectrometry (LC/MS) was performed using an Agilent 1200 series HPLC connected to an Agilent 6224 Time-of-Flight (TOF) LC/MS system equipped with a Turbospray ion source. For purification, samples were filtered through 0.22 μm filters (Millipore). HPLC-based quantification was performed by comparing integrated fluorescence intensity of tryptophan (ex. 280/ em. 330) over time of untreated MS2 to reassembled samples.

Dynamic Light Scattering. DLS was performed using a Zetasizer Nano (Malvern Instruments). Before analysis samples were filtered through 0.22 μm filters (Millipore). Measurements were taken in 50 mM Tris pH 7.2, 100 mM NaCl.

Enzyme Kinetics. Enzyme assays were performed in 96-well plates 10 wells at a time. Fluorescence was read using a Spectramax M2 plate reader (Molecular Devices). The excitation wavelength was 360 nm, and the emission wavelength was 449 nm. Substrate stock solutions of 4-methylumbelliferyl phosphate and standards of 4-methylumbelliferone were prepared fresh daily to account for background substrate hydrolysis, bleaching, or variability in plate reader signal. Michaelis-Menten parameters were determined by fitting to the Michaelis-Menten equation in OriginPro 8 and R software.

Gel Analyses. Sodium dodecylsulfate-poly(acrylamide) gel electrophoresis (SDS-PAGE) analysis of all protein samples was carried out on a Mini Protean apparatus (Bio-Rad, Hercules, CA) using 12.5% poly(acrylamide) gels prepared according to the manufacturer's specifications. Samples were heated to 95 °C in the presence of Laemmli buffer containing β-mercaptoethanol for 3-5 minutes. Gels were run at 125 V for approximately 75 minutes, stained with Coomassie R-250, and imaged using a Chemidoc imager (Bio-Rad). Molecular masses were estimated by comparison to the EZ-run Prestained Rec Protein Ladder (Fisher). Quantification of PhoA-neg bands was accomplished by densitometry using the Chemidoc software using known quantities of free enzyme as standards.

UV-vis Spectroscopy. UV-vis spectroscopy was performed using a Nanodrop 2000C (Thermo Scientific). Free enzyme was quantified by absorbance at 280 using an extinction coefficient of $\varepsilon = 0.71$ cm·mL/mg. Nucleic acid-free MS2 was quantified using $\varepsilon = 1$ cm·mL/mg.

Transmission Electron Microscopy. TEM images were taken using an FEI Technai 12 transmission electron microscope with an accelerating voltage of 120 kV. Samples were

desalted using NAP-5 desalting columns (GE Healthcare), concentrated to approximately 50 μM using 100 kDa MWCO spin filters (Millipore), incubated on Formvar-coated copper mesh grids for 5 minutes and wicked off using filter paper and dried in air briefly. These grids were then quickly washed with water and immediately wicked again. The samples were stained with 1% $UO_2(OAc)_2$ for 2 minutes, and again wicked and dried.

Generation of MS2 mutants. The residues around the pore of the capsid, specifically T71 and V72 (numbering based on crystal structure 1ZDK) were targeted for mutation using the Quikchange method. The mutagenesis primers were found to interfere with ColE1 plasmid origins, so the gene for wild-type MS2 was cloned into pBAD33 with a p15A origin. The gene for wild-type MS2 coat protein was PCR-amplified from pBAD-*myc-his*-WTMS2, retaining that construct's 5' XbaI site and 3' HindIII site (forward: 5'-AGTCAGCCGTTTTCTAGACTAACAGG-3'; reverse: 5'-AGTCAGCCCAAGCTTAGTAG-ATGCCG-3'). The gene was digested, ligated into pBAD33, and transformed into DH10B cells.

Plasmid pBAD33-WTMS2 was used to generate mutants T71E, T71E/V72D, and T71K/V72R using the primers in Table 4.1. For optimal expression, the mutant genes were cloned into pTrc99a using the same method outlined above.

Mutant	Primer
T71E	5'-GGTGCCTAAAGTGGCAACCCAGGAGGTTGGTGGTGTAGAGC-3'
	5'-GCTCTACACCACCAACCTCCTGGGTTGCCACTTTAGGCACC-3'
T71E/V72D	5'-GGTGCCTAAAGTGGCAACCCAGGAGGACGGTGGTGTAGAGC-3'
	5'-GCTCTACACCACCGTCCTCCTGGGTTGCCACTTTAGGCACC-3'
T71K/V72R	5'-GGTGCCTAAAGTGGCAACCCAGAAACGCGGTGGTGTAGAGC-3'
	5'-GCTCTACACCACCGCGTTTCTGGGTTGCCACTTTAGGCACC-3'

Table 4.1: Mutagenesis primers for MS2 capsid pore mutants.

Expression and Purification of MS2. Wild-type MS2 capsid and pore variants were expressed and purified as previously described[39]. Plasmid pBAD-*myc-his* containing the gene for the wild-type capsid or pTrc99a containing the desired pore MS2 variant was transformed into DH10B cells and plated on LB + 50 mg/L carbenicillin (Cb). A single colony was grown overnight in 5 mL of LB + Cb and used to inoculate 500 mL 2xYT + Cb at a 1:500 dilution. This culture was shaken at 37 °C until it reached an OD_{600} of 0.5, at which time it was induced with 0.2 % arabinose (for pBAD) or 1 mM IPTG (for pTrc) and allowed to express overnight at the same temperature.

Cells were harvested, resuspended in 10 mM taurine pH 9.0 and lysed by sonication. The clarified lysate was applied to a DEAE column and the flowthrough containing the MS2 mutant was retained. To precipitate the capsid, solid poly(ethylene glycol) 8k and 5 M NaCl were added to final concentrations of 10% w/v and 0.5 M, respectively, and the mixture was incubated at 4 °C for 1 hour on an orbital shaker and centrifuged at 14,000 x *g* for 30 minutes. The pelleted protein was resuspended in a small amount of 10 mM phosphate buffer pH 7.2 and centrifuged again. The re-solubilized capsid was then applied to a S200 size exclusion chromatography column and eluted in 10 mM phosphate buffer. Fractions containing MS2 were re-precipitated and stored at 4 °C.

Generation of PhoA^{WT}-neg and PhoA$^{D153G/D330N}$-neg. The *phoA* gene was amplified from E. coli DH10B genomic DNA, inserted into a pTrc99a vector, re-amplified with PCR primers to add a C-terminal sequence of EEEEDDDEDDDDEEDD (forward primer 5'-ACACTGTCTAGAGTGAAACAAAGCACTATTGCAC-3', reverse primer 5'-ACACTGAAGCTTTTAATCGTCTTCCTCGTCATCGTCATCTTCGTCATCATCCTC TTCTTCCTCTTTCAGCCCCAGAGCG) and reinserted into pTrc99a. To generate PhoA$^{D153G/D330N}$-neg, plasmid pTrc99a harboring the gene for PhoA^{WT}-neg was subjected to two successive rounds of Quikchange mutagenesis. First, the D330N mutation was made using forward primer 5'-GTGCGTCAATCGATAAACAGAATCATGCTGCGAATCCTTG-3' and reverse primer 5'-CAAGGATTCGCAGCATGATTCTGTTTATCGATTGACGCAC-3' and the resulting plasmid was transformed into DH10B cells, isolated and sequenced. The D153G mutation was then made using forward primer 5'-CGTTTCTACCGCAGAGTTGCAGGGCGCCACGCCCGCTGCGCTGGTGGC-3' and reverse primer 5'-GCCACCAGCGCAGCGGGCGTGGCGCCCTGCAACTCTGCGGTAGAAACG-3' and resulting plasmids were again transformed into DH10B cells, isolated and sequenced.

*Expression and Purification of PhoA^{WT}-neg and PhoA$^{D153G/D330N}$-neg***.** PhoA^{WT}-neg and PhoA$^{D153G/D330N}$-neg were expressed and purified as previously described[39]. Plasmid pTrc99a harboring the gene for the desired enzyme was transformed into DH1 cells and plated on LB-Cb. A single colony was grown overnight in 5 mL of LB + Cb and used to inoculate 1 L 2xYT + Cb supplemented with 1 mM $MgSO_4$ and 0.1 mM $ZnSO_4$ at a 1:500 dilution. This culture was shaken at 37 °C until it reached an OD_{600} of 0.5, at which time it was induced with 1 mM IPTG and allowed to express overnight at the same temperature.

Cells were harvested and washed once with 100 mL of 50 mM Tris HCl pH 8.0, and resuspended in 50 mL of the Tris buffer. Periplasmic proteins were isolated by addition of solid sucrose to 500 mM, 500 mM EDTA to 2.5 mM, and lysozyme to 0.6 mg/mL. The resulting suspension was gently mixed and incubated at 37 °C for 30 minutes, followed by centrifugation at 5000 x *g* for 20 minutes. Solutions of $MgSO_4$ and $ZnSO_4$ were added to 10 mM and 1 mM final concentration, respectively, and the mixture was heated to 80 °C for 10 minutes, with intermittent mixing. Precipitated protein was pelleted by centrifugation at 10,000 x *g* for 20 minutes followed by filtration through a 0.22 μm filter. This was then applied to a 5 mL bed of DEAE-sepharose and washed with 20 mL aliquots of 20 mM Tris containing 0 mM, 50 mM, and 100 mM NaCl. The protein was eluted with 250 mM NaCl and stored at 4 °C in the same buffer with $MgSO_4$ and $ZnSO_4$ added to 1 mM and 0.1 mM final concentrations, respectively. This protocol yields approximately 40 mg of PhoA^{WT}-neg and 25 mg of PhoA$^{D153G/D330N}$-neg per liter of culture.

FITC Labeling of PhoA^{WT}-neg. PhoA^{WT}-neg was dialyzed into triethanolamine buffer pH 8.3 for labeling. Fluorescein 5(6)-isothiocyanate (50 mM freshly dissolved in DMF) was added to 1.3 mg/mL enzyme in 50-fold molar excess and incubated at 4 °C overnight. Excess FITC was removed by dialysis using a 10 kDa MWCO dialysis cassette (Pierce). Attachment of the fluorophore was confirmed by electrospray ionization-time of flight

mass spectrometry and size exclusion chromatography using a Biosep-SEC-4000 column at a flow rate of 1 mL/min.

Encapsulation of Enzymes in MS2 Mutants. A modified protocol was used to encapsulate enzymes inside MS2 pore mutants. MS2 was disassembled in 66% acetic acid and desalted into 1 mM acetic acid as described. The coat protein was added to a chilled solution of enzyme (0.1 mg/mL), Bis-Tris buffer (50 mM, pH 6.0), and TMAO (1.8 M). Ionic strength was altered with the addition of varying amounts of (5.0 M) NaCl to the reassembly reaction. The capsid was allowed to reassemble for 48 hours at 4 °C before purification and analysis.

Purification of Encapsulated Enzymes. Reassembled capsid containing the desired enzyme derivative was purified by precipitation with PEG 8k followed by SEC-HPLC. Solid PEG 8k and 5 M NaCl were added to the reassembly reaction to a final concentration of 10% and 0.5 M, respectively, and the mixture was incubated at 4 °C on an orbital shaker and centrifuged at 21,000 x *g* for 10 minutes. The precipitated capsids were resuspended in ST buffer (50 mM Tris pH 7.2, 100 mM NaCl) and centrifuged again to pellet any aggregated protein. The supernatant was filtered through a 0.22 μm filter and injected on a Biosep-SEC-4000 column. ST buffer was pumped over the column at 1 mL/min, and reassembled MS2, which elutes at ~6.3 minutes was collected and concentrated using 100 kDa MWCO centrifugal filters (Millipore).

Evaluation of Stability of MS2-Encapsulated PhoA^{WT}-neg. We examined thermal stability of our system using dynamic light scattering (DLS). Size trends upon change in temperature were recorded for each capsid, starting at 35 °C in 4 °C steps with a 5 minute incubation at each temperature.

Enzyme Kinetics. Kinetics of free and encapsulated enzymes were measured by monitoring the increase in fluorescence ($\lambda_{ex.}$ 360 nm/$\lambda_{em.}$ 449 nm) upon hydrolysis of the phosphoester bond in 4-methyl-umbelliferyl phosphate. The enzyme (100μl, 2 nM monomer) in 2x buffer was added to the 100 μl H_2O containing the substrate at increasing concentrations, and initial rates were recorded. Standards of 4-methylumbelliferone were used to generate a standard curve in the relevant signal range to convert fluorescent signal to concentration. Data were fit to the Michaelis-Menten equation using OriginPro 8 or R fitting software. Assays were done in the following buffers: 50 mM MOPS pH 8.0, 500 mM NaCl, 1 mM $MgSO_4$, 0.1 mM $ZnSO_4$ (high salt); 50 mM MOPS pH 8.0, 1 mM $MgSO_4$, 0.1 mM $ZnSO_4$ (low salt); 10 mM Tris pH 8.0, 4 mM NaCl, 1 mM $MgSO_4$, 0.1 mM $ZnSO_4$ (low salt); 10 mM Tris pH 8.0, 100 mM NaCl, 1 mM $MgSO_4$, 0.1 mM $ZnSO_4$ (medium salt); 10 mM Tris pH 8.0, 500 mM NaCl, 1 mM $MgSO_4$, 0.1 mM $ZnSO_4$ (high salt).

Experimental Determination of Inhibition Constants (K_I) for Free Enzymes. Free PhoA^{WT}-neg and PhoA$^{D153G/D330N}$-neg were incubated with increasing concentrations of phosphate and assayed in medium salt buffer for kinetic parameters as described above. The resulting $K_{M,app}$ values were plotted against phosphate concentrations and fit to the K_I equation.

Nanoreactor Kinetics Modeling. Modeling of kinetic parameters was performed in MATLAB using the reaction scheme shown in Figure 4.2A. The model incorporated substrate diffusion between the bulk and the capsid (k_1), Michaelis-Menten enzyme kinetics (k_{cat}) with competitive product inhibition by phosphate (K_I), and phosphate diffusion out of the capsid, as represented by the first-order rate constant k_2. Kinetic studies of the PhoA^{WT}-neg and PhoA$^{D153G/D330N}$-neg enzymes in the MS2^{T71E} capsid were simulated by numerical integration of the reaction scheme for each substrate concentration tested experimentally, producing a plot of product concentration against time for each substrate concentration (Figure 4.3). A linear regression was then used to fit the initial reaction rate for comparison to the experimental data (Figure 4.3). The model thus generates a set of rates of product formation as a function of substrate concentration. The four kinetic parameters (k_1, k_2, $K_{I,PhoA(WT)}$, and $K_{I,PhoA(D153G/D330N)}$) were then fitted to the experimental data for both the PhoA^{WT}-neg and PhoA$^{D153G/D330N}$-neg enzymes simultaneously by minimization of the sum of the square errors between the simulated data and the data from experiment, using the built-in lsqnonlin minimization function in MATLAB. For a detailed description of the model, see Appendix A.

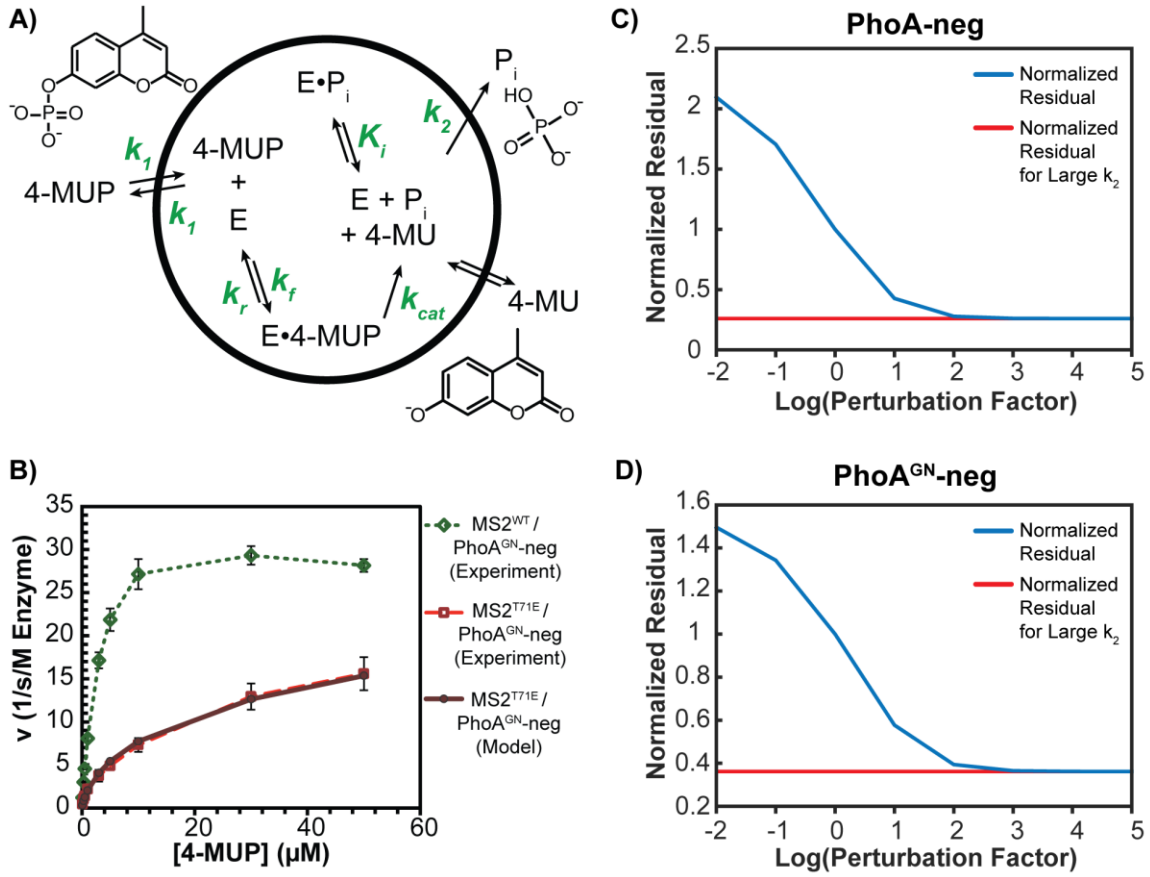


Figure 4.2: Kinetic model of encapsulated reactions. A) System of reactions used as parameters for the kinetic model. Variables highlighted in green define the kinetics of the system. B) Results of the model for encapsulated PhoA$^{D153G/D330N}$-neg. C) Relaxation of parameter k_2 to approximate MS2^{WT} behavior for encapsulated PhoA^{WT}-neg. The blue line represents the residual between the MS2^{WT} experimental data and the best fit to the

$MS2^{T71E}$ case, normalized to the residual at the fitted k_2 value, plotted as the fitted value of k_2 multiplied by a perturbation factor ranging from 0.001 to 1000. The red line represents the minimum residual value in the limit of large k_2. D) The same calculation as (C), performed for encapsulated $PhoA^{D153G/D330N}$-neg.

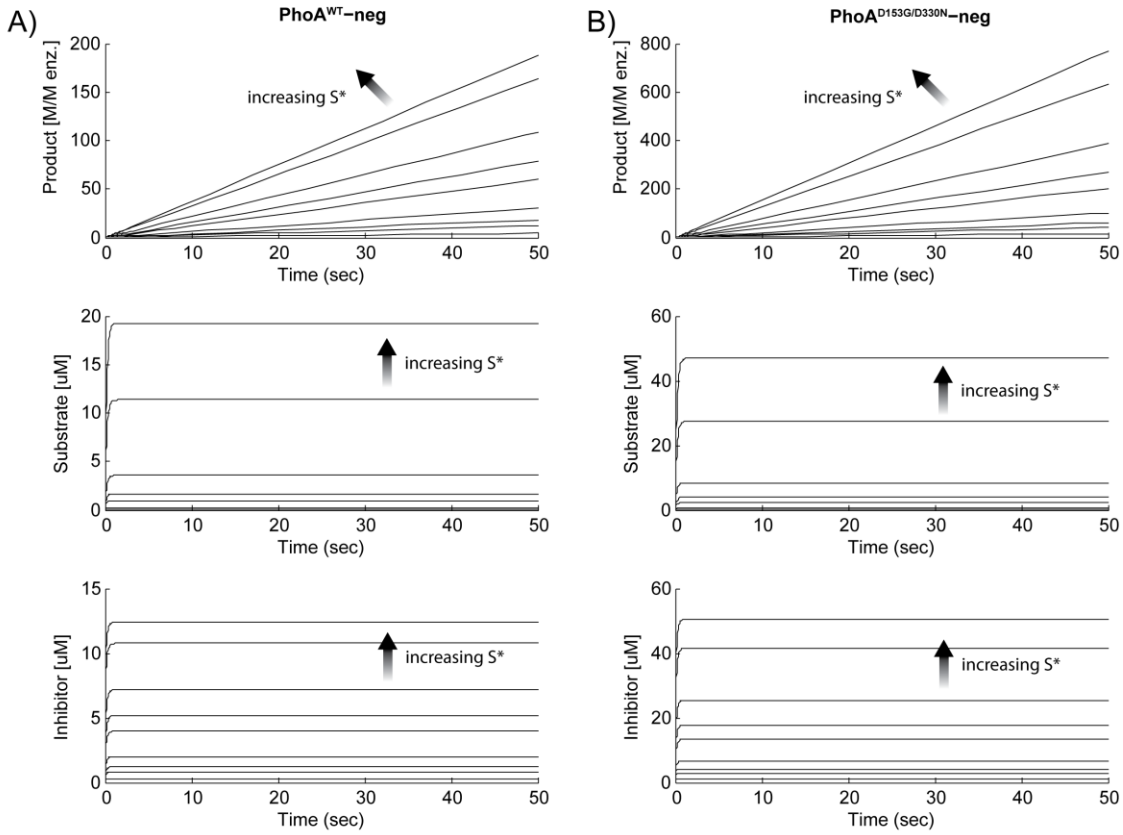


Figure 4.3: Simulated concentration profiles for the product, substrate, and competitive inhibitor are shown for each experimental substrate concentration for the A.) PhoA-neg and B.) PhoA-neg-GN cases. Product concentration is shown for the assay bulk and capsid phases combined, as observed by experiment; substrate and inhibitor concentrations are the intra-capsid concentrations. A linear regression was performed to determine the initial rate of product formation at each substrate concentration. Substrate and inhibitor concentrations were both observed to reach a pseudo-steady state over the timescale of the simulation.

4.3 Results

Encapsulation of Alkaline Phosphatase in Mutated MS2 capsids. We generated several mutants of the wild-type MS2 ($MS2^{WT}$) capsid by Quikchange mutagenesis and expressed them as described[127] (Table 4.1). The T71E ($MS2^{T71E}$) and T71E/V72D ($MS2^{T71E/V72D}$) mutations add one or two negative charges to each subunit, respectively, while the T71K/V72R ($MS2^{T71K/V72R}$) mutation adds two positive charges. This results in capsids with increased negative (in the case of $MS2^{T71E}$ and $MS2^{T71E/V72D}$) or positive (in

the case of MS2$^{T71K/V72R}$) charge around the 5- and quasi-6-fold axis pores on each capsid. Figure 4.4 depicts one possible conformation of these residues and their predicted electrostatics. The mutations have little or no effect on expression and assembly in *Escherichia coli*, and have only a small effect on the thermal stability of the capsids (Figures 4.5 and 4.6).

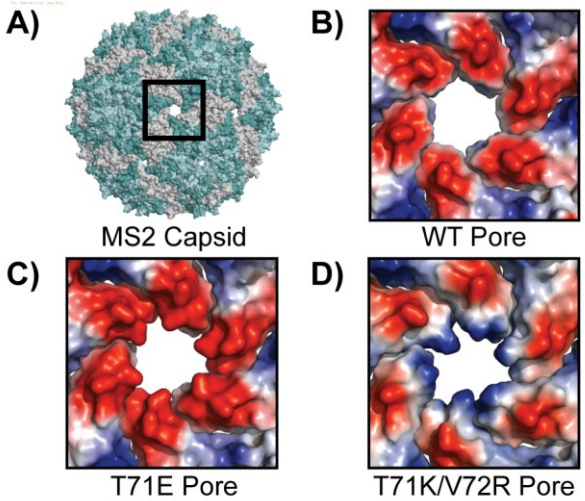


Figure 4.4: Vacuum electrostatic representations of pore mutants modeled in PyMol. Negative potential is represented by red shading, and positive by blue. A) The capsid features ~1.8 nm pores at each of its 5- or quasi-6-fold axes. B) Wild-type MS2 has negative charge around the pore periphery, but not inside. C) Mutant T71E has significant negative charge throughout the pore. D) Mutant T71K/V72R has positive charge in the pore.

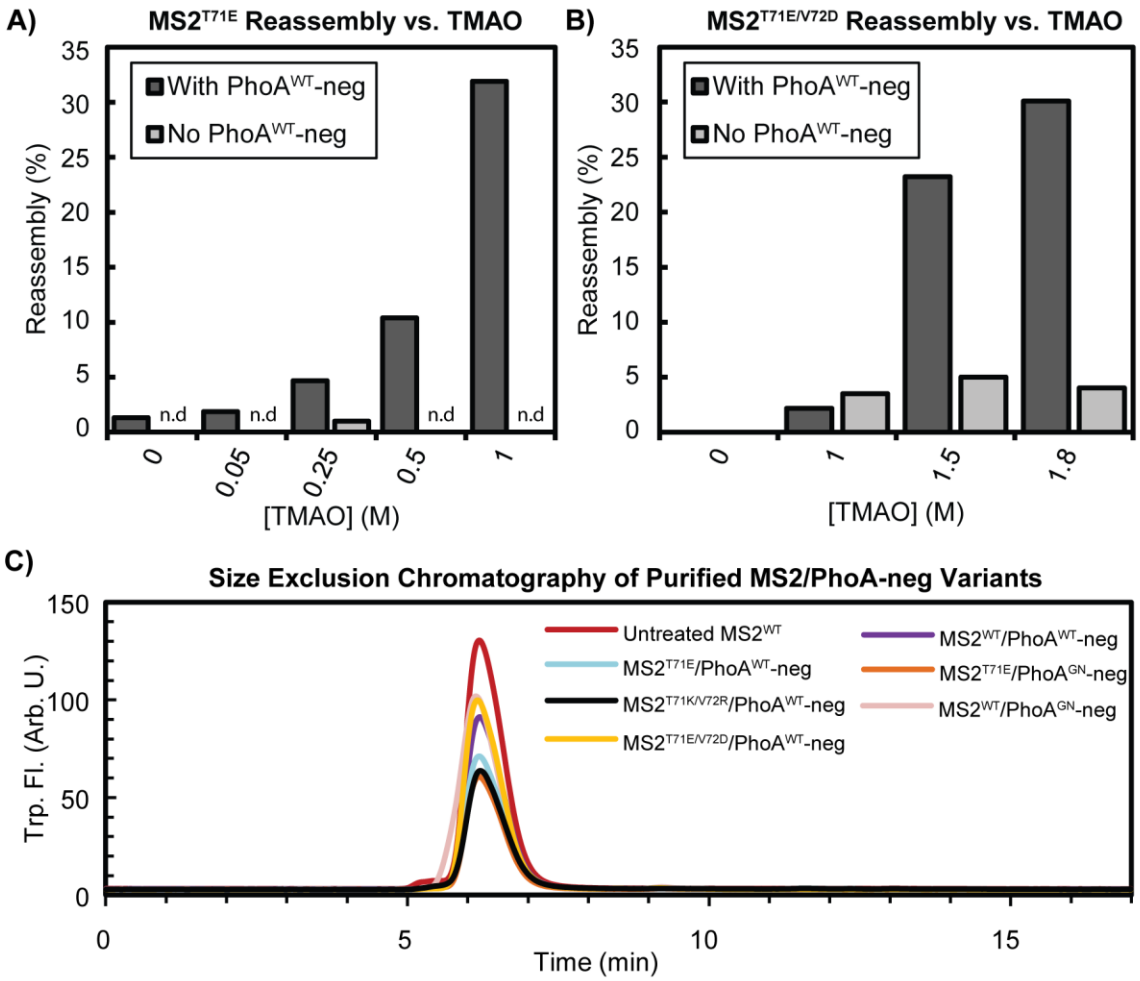
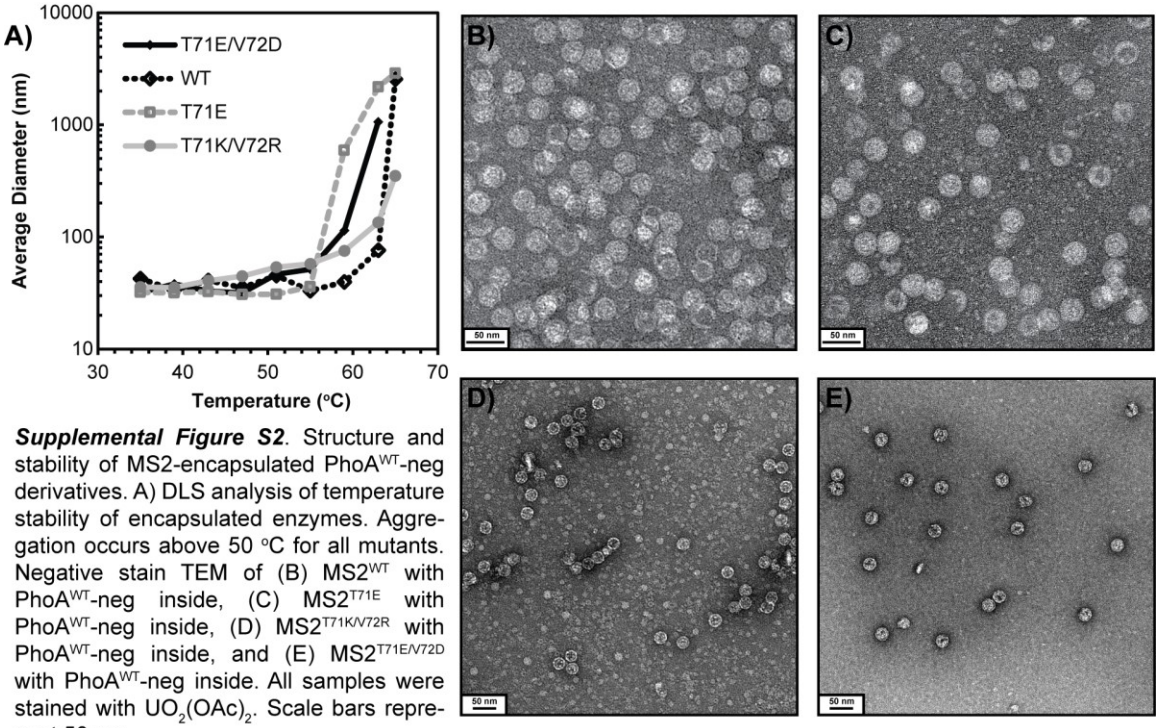


Figure 4.5: Reassembly of capsid mutants around PhoAWT-neg. A) MS2T71E reassembly around PhoAWT-neg in Bis Tris pH 6.0 is greatly enhanced by high concentrations of trimethylamine-N-oxide (TMAO). B) MS2T71E/V72D reassembly is also enhanced by TMAO. No significant reassembly was observed below 1 M. Background reassembly, though minimal, may result in some empty capsids. C) Size exclusion chromatography of all purified enzyme/capsid combinations. All have the same retention time and show no detectable free enzyme, even after 2 months of storage at 4 oC.



Supplemental Figure S2. Structure and stability of MS2-encapsulated PhoA^{WT}-neg derivatives. A) DLS analysis of temperature stability of encapsulated enzymes. Aggregation occurs above 50 °C for all mutants. Negative stain TEM of (B) MS2^{WT} with PhoA^{WT}-neg inside, (C) MS2^{T71E} with PhoA^{WT}-neg inside, (D) MS2$^{T71K/V72R}$ with PhoA^{WT}-neg inside, and (E) MS2$^{T71E/V72D}$ with PhoA^{WT}-neg inside. All samples were stained with $UO_2(OAc)_2$. Scale bars represent 50 nm.

Figure 4.6: Structure and stability of MS2-encapsulated PhoAWT-neg derivatives. A) DLS analysis of temperature stability of encapsulated enzymes. Aggregation occurs above 50 oC for all mutants. Negative stain TEM of (B) MS2WT with PhoAWT-neg inside, (C) MS2T71E with PhoAWT-neg inside, (D) MS2T71K/V72R with PhoAWT-neg inside, and (E) MS2T71E/V72D with PhoAWT-neg inside. All samples were stained with UO2(OAc)2. Scale bars represent 50 nm.

We subjected each of these mutants to our previously developed disassembly/reassembly protocol, which encapsulates negatively charged molecules, such as enzymes, based on electrostatic interactions with the positive residues on the capsid interior surface. To initiate reassembly, we used a model enzyme, *E. coli* alkaline phosphatase, with a C-terminal sequence of 16 negatively charged amino acids (PhoA^{WT}-neg).[127] The MS2$^{T71K/V72R}$ mutant reassembled similarly to the MS2^{WT} capsid in these conditions (50 mM Tris HCl pH 7.2, 100 mM NaCl, 36-48 hours, 4°C; Figure 4.5). The MS2^{T71E} and MS2$^{T71E/V72D}$ mutants, however, gave significantly lower yields of reassembled product. We varied the salt, pH, and TMAO concentrations of the reassembly reaction to determine what conditions favored reassembly of these mutants. While salt concentration did not significantly affect the yield of reassembled capsid, lowering the pH from 7.2 to 6.0 and increasing the TMAO concentration from 0.25 M to 1.8 M enhanced the yield of reassembled MS2^{T71E} and MS2$^{T71E/V72D}$ capsids by 8-fold. Using these improved reassembly conditions, we obtained between three and ten enzymes encapsulated per capsid in each of the MS2 mutants. The reassembly conditions did not affect the activity of the enzyme after purification (Figure 4.7). Subsequent enzyme assays were performed in identical buffers for each capsid mutant, as described below.

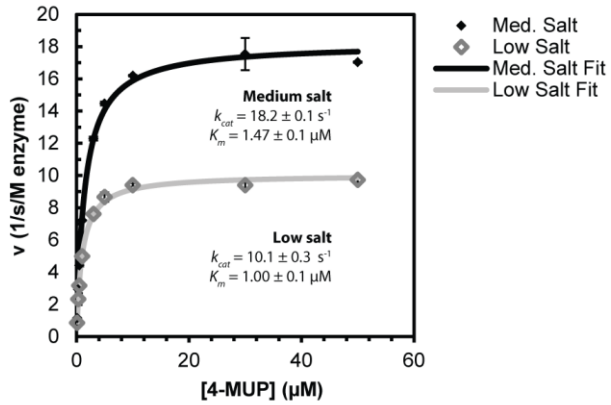


Figure 4.7: Kinetics of free PhoAWT-neg in medium and low salt Tris buffer after treatment with reassembly and purification conditions. Slight changes were observed in the kinetic parameters kcat and KM, but do not account for changes observed in encapsulated derivatives.

To ensure complete encapsulation of $PhoA^{WT}$-neg, the enzyme was labeled with fluorescein 5(6) isothiocyanate (FITC). The labeling had no effect on the enzymatic activity (Figure 4.8). The labeled enzyme was encapsulated in each of the capsid mutants. Upon encapsulation, a change in optical properties of the dye was observed; therefore the fluorescein was only used as a qualitative indicator of presence of the enzyme within the capsid. Enzyme quantification was performed by running the purified capsid constructs on SDS-PAGE gels followed by Coomassie staining and optical densitometry. The capsids each contained a similar amount of enzyme, with the $MS2^{WT}$ and $MS2^{T71K/V72R}$ capsids carrying slightly more, followed by the $MS2^{T71E}$ then $MS2^{T71E/V72D}$ capsids (Table 4.2). It is not surprising that added negative charge decreases the amount of encapsulated enzyme due to the electrostatic nature of the interaction between $PhoA^{WT}$-neg and the capsid. We also found that increasing the salt concentration lowered the number of enzymes encapsulated per $MS2^{WT}$ capsid (Figure 4.9), supporting the importance of electrostatics in the enzyme/capsid interaction during assembly. Moreover, $MS2^{WT}$ capsids with varying numbers of enzymes encapsulated showed indistinguishable $k_{cat,app}$ and $K_{M,app}$ kinetic parameters, as expected (Figure 4.10).

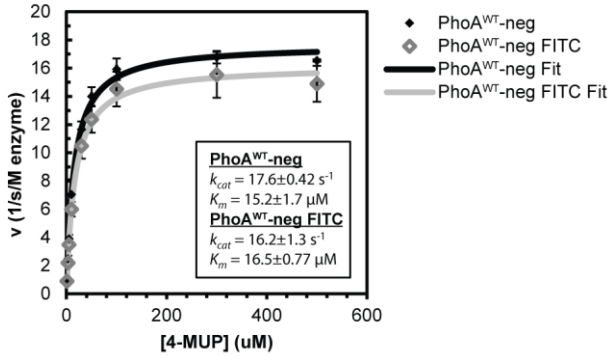


Figure 4.8: Analysis of kinetics of FITC-labeled and unlabeled PhoAWT-neg in high salt MOPS.

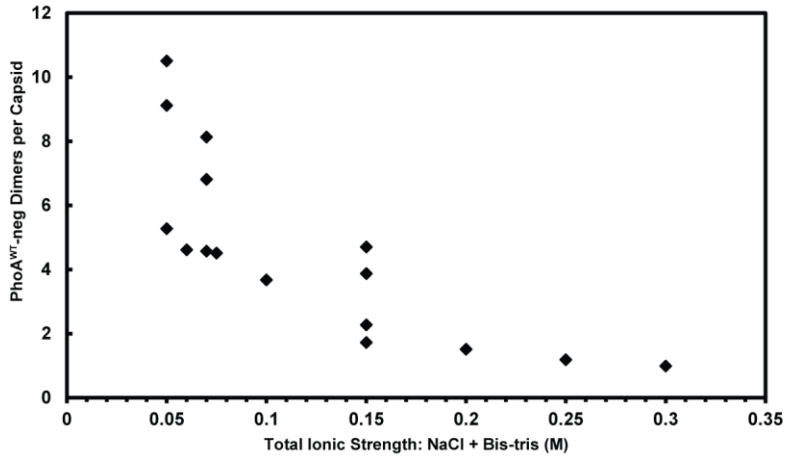


Figure 4.9: Loading of PhoAWT-neg in MS2WT varies inversely with ionic strength of the reassembly reaction. NaCl was added to increase the ionic strength; 0.05 M Bis-tris was present in all reassembly conditions. Each data point represents a separate reassembly reaction. Error of enzyme quantification is not shown.

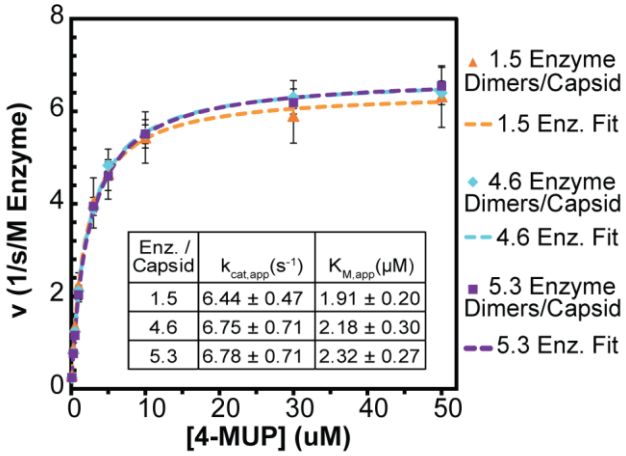


Figure 4.10: Enzyme assays with PhoA^{WT}-neg encapsulated into MS2^{WT} at different levels of enzyme loading. Error bars represent standard deviation of three assays for three separate encapsulations each. The inset displays calculated Michaelis-Menten parameters $K_{M,app}$ and $k_{cat,app}$, which were found to be not significantly different between the three loading levels.

Capsid and Enzyme Type	Average Enzyme Dimers/Capsid
WT/PhoA-neg	9.6
T71E/PhoA-neg	5.8
T71E/V72D/PhoA-neg	3
T71K/V72R/PhoA-neg	9.4
WT/PhoA$^{D153G/D330N}$-neg	9.5
T71E/ PhoA$^{D153G/D330N}$-neg	6

Table 4.2: Enzyme loading in MS2 derivatives. Enzyme concentration was measured by SDS-PAGE densitometry. Total protein was measured by A_{280} using $\varepsilon = 0.71$ cm·mL/mg for PhoA-neg and $\varepsilon = 1$ cm·mL/mg for MS2.

Capsid Mutations Impact Kinetics of Encapsulated PhoA. We evaluated the Michaelis-Menten kinetics of the hydrolysis of 4-methylumbelliferyl phosphate to 4-methylumbelliferone catalyzed by PhoA^{WT}-neg in each of the capsids, as shown in Figure 4.11A. We chose to do the experiments in low concentration Tris buffer, although similar trends were observed in MOPS (Table 4.3). Although the presence of Tris can lead to transphosphorylation onto the buffer, artificially increasing k_{cat}, this effect is negligible at the low concentrations used here (10 mM).[131] At the pH of the reaction, both the substrate and the product are negatively charged, which could lead to electrostatic effects on flux both into and out of the compartment.

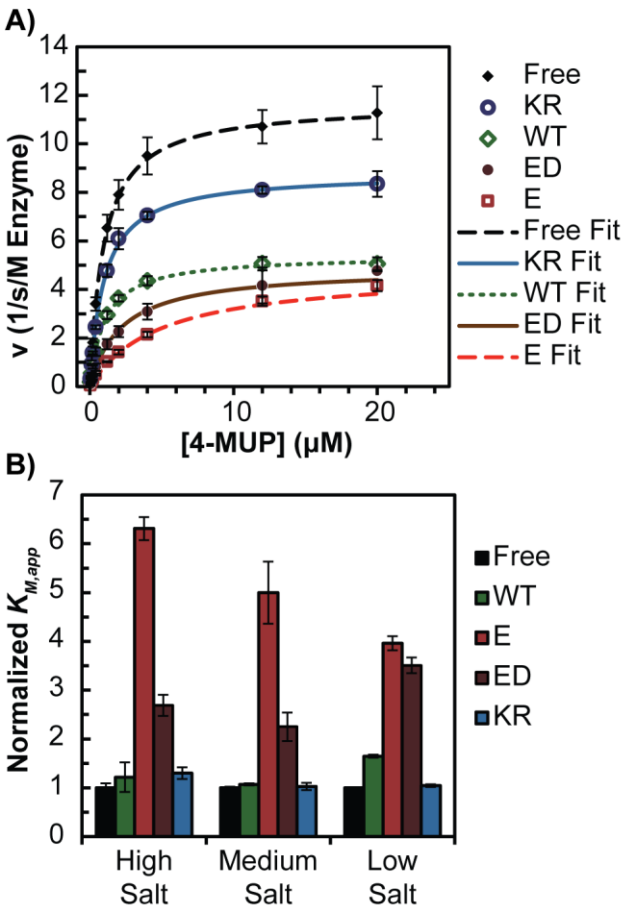


Figure 4.11: Enzyme assays with free and encapsulated PhoA^{WT}-neg. A) Michaelis-Menten curves of the enzyme encapsulated in each MS2 pore variant in 10 mM Tris pH 8.0, 100 mM NaCl (medium salt). B) Apparent K_M of the enzyme in each pore variant in 10 mM Tris pH 8.0 with 4 mM (low salt), 100 mM (medium salt), or 500 mM NaCl (high salt). The $K_{M,app}$ of both negative pore mutants, E and ED, increases significantly. Each $K_{M,app}$ was normalized to the free enzyme in the corresponding buffer conditions. Error bars represent standard deviation of 3 assays.

		Free	WT	E	ED	KR
Tris High Salt	k_{cat} (s^{-1})	23.8 ± 3.8	13.21 ± 1.0	19.7 ± 1.6	16.9 ± 2.0	18.3 ± 2.46
	$K_{M,app}$ (μM)	2.13 ± 0.09	2.59 ± 0.30	13.4 ± 0.24	5.72 ± 0.22	2.76 ± 0.12
Tris Medium Salt	k_{cat} (s^{-1})	11.6 ± 0.83	5.40 ± 0.34	4.75 ± 0.57	4.87 ± 1.0	8.79 ± 1.1
	$K_{M,app}$ (μM)	0.97 ± 0.02	1.0 ± 0.02	4.8 ± 0.63	2.18 ± 0.29	1.0 ± 0.07
Tris Low Salt	k_{cat} (s^{-1})	5.23 ± 0.25	2.51 ± 0.12	2.74 ± 0.30	2.59 ± 0.30	4.68 ± 0.57
	$K_{M,app}$ (μM)	0.54 ±0.003	0.89 ± 0.03	2.13 ± 0.15	1.89 ± 0.16	0.56 ± 0.03
MOPS Low Salt	k_{cat} (s^{-1})	7.58 ± 0.02	4.89 ± 0.33	4.40 ± 0.34	4.06 ± 0.57	n.d.
	$K_{M,app}$ (μM)	2.31 ± 0.07	3.79 ± 0.29	6.87 ± 0.71	6.42 ± 0.50	n.d.
Tris Medium Salt ($PhoA^{D153G/D330N}$ mutant)	k_{cat} (s^{-1})	43.3 ± 8.1	31.4 ± 5.3	19.8 ± 5.3	n.d.	n.d.
	$K_{M,app}$ (μM)	1.43 ± 0.16	2.45 ± 0.10	15.0 ± 1.8	n.d.	n.d.

Table 4.3: Kinetic constants for free and encapsulated PhoA-neg and $PhoA^{D153G/D330N}$-neg derivatives in several conditions all show substantial increases in $K_{M,app}$. Error represents aggregation of standard deviations from at least three kinetic assays using nine substrate concentrations and at least three SDS-PAGE densitometry measurements to determine enzyme concentration.

As we observed previously, the $MS2^{WT}$ capsid had no effect on the apparent K_M ($K_{M,app}$) but a modest inhibitory effect on the k_{cat}. Interestingly, encapsulation in the $MS2^{T71K/V72R}$ capsid had little effect on the $K_{M,app}$ and a less inhibitory effect on the k_{cat} than encapsulation within $MS2^{WT}$. On the contrary, both the negative pore mutants significantly increased the $K_{M,app}$ of the reaction while decreasing the overall k_{cat}. The increase in $K_{M,app}$ persists at a variety of salt concentrations (Figure 4.11B). Theoretical studies suggest electrostatic barriers in nanopores can be overcome by high ionic strength even when the substrate size approaches the Debye length of the pore.[132] Our pores are much smaller than previously studied channels and our data suggest that our system is not in this regime, even at 0.5 M NaCl.[133] The Debye length in this condition is 0.43 nm compared to a substrate radius of approximately 0.3 nm and a pore radius of 0.9 nm.

Curiously, encapsulation in the $MS2^{T71E}$ capsid increased the $K_{M,app}$ significantly more than encapsulation within the $MS2^{T71E/V72D}$ capsid. It is possible that the double mutation changed the pore diameter or electrostatic environment through electrostatic repulsion or other steric effects. We note that we did not directly measure the electrostatic landscape of the mutated pores, and quantification of the level of negative charge around the pore is difficult. Our results suggest that two mutations in the $MS2^{T71E/V72D}$ capsid may cause changes in the pore environment to form an electrostatic barrier similar to that of the $MS2^{T71E}$ capsid.

In order to extend our exploration of the effects of encapsulation, we next examined a mutant of alkaline phosphatase that has different kinetic parameters. This mutant, D153G/D330N ($PhoA^{D153G/D330N}$-neg), was developed to have significantly higher values of k_{cat} and K_M, while suffering less inhibition from free phosphate.[134] We speculated that product inhibition by free phosphate was affecting the kinetic

behavior observed in the various capsids, and this enzyme permitted further investigation of this effect. We expressed the $PhoA^{D153G/D330N}$-neg mutant and encapsulated it in $MS2^{WT}$ and $MS2^{T71E}$ capsids. The kinetics of the free $PhoA^{D153G/D330N}$-neg enzyme, as expected, showed an increase in both $k_{cat,app}$ and $K_{M,app}$ relative to the free wild-type enzyme (Table 4.3). When encapsulated in $MS2^{WT}$ capsid, this enzyme had a moderate increase in $K_{M,app}$; encapsulation in the $MS2^{T71E}$ capsid led to an even more pronounced increase in the $K_{M,app}$, suggesting that charge repulsion remains important with the mutant enzyme.

We hypothesized that the $MS2^{T71E}$ capsid pores could inhibit the transport of negatively charged species more than those of the $MS2^{WT}$ capsid, resulting in increased inhibitory effects on k_{cat} and $K_{M,app}$. Direct measurements of the phosphate ion concentration inside the capsid relative to the assay bulk were impractical, so we pursued a modeling strategy to demonstrate the feasibility of this ion-trapping mechanism. The kinetic behavior of the MS2/PhoA-neg nanoreactor system is different from the case of the free enzyme, due to the effects both of small molecule transport and of enzyme immobilization on kinetic behavior. Since we could not determine the quantitative effects of immobilization, we developed a model based on the enzyme kinetic parameters observed for alkaline phosphatase within the $MS2^{WT}$ capsid and attempted to describe the altered kinetics of the $MS2^{T71E}$ encapsidated enzyme solely by altering transport parameters.

Modeling Suggests a Role for Phosphate Transport in Kinetic Differences. The MS2 nanoreactor was modeled using the reaction scheme shown in Figure 4.2, treating the MS2 capsid interior and the assay bulk as two distinct, well-mixed phases. The model incorporates substrate diffusion between the bulk and the capsid, as represented by the permeability parameter k_1, classical Michaelis-Menten enzyme kinetics with competitive product inhibition by phosphate[135], and phosphate diffusion out of the capsid, as represented by the first-order rate constant k_2. We regarded the phosphate concentration in the bulk as essentially zero over the timescale modeled, due to the large assay volume relative to the volume of the capsid phase. The intrinsic Michaelis-Menten parameters $k_{cat,app}$ and $K_{M,app}$ for the $PhoA^{WT}$-neg and $PhoA^{D153G/D330N}$-neg enzymes, as fitted to the $MS2^{WT}$ capsid case, were held constant in the $MS2^{T71E}$ capsid case for each enzyme, as we assume that the intrinsic kinetic properties of the encapsulated enzymes differed from the free enzyme but are not altered by the capsid mutations.[136] Experimental results and fitted models are shown in Figure 4.2B.

The transport parameters k_1 and k_2 capture the effects of the T71E pore mutation on substrate and inhibitor transport, respectively, relative to that of $MS2^{WT}$ capsid (not relative to the free enzyme). The two competitive inhibition parameters, $K_{I,PhoA(WT)}$ and $K_{I,PhoA(D153G/D330N)}$, capture phosphate inhibition of the $PhoA^{WT}$-neg enzyme and the $PhoA^{D153G/D330N}$-neg mutant. The modeled enzyme kinetics for the $PhoA^{D153G/D330N}$-neg enzyme in the $MS2^{T71E}$ capsid are shown alongside the $MS2^{WT}$ and $MS2^{T71E}$ experimental data in Figure 4.2B. $K_{I,PhoA(WT)}$ and $K_{I,PhoA(D153G/D330N)}$ were fitted as 1.74 μM and 2.67 μM, respectively. These values are comparable to the experimentally determined values of 1.4 μM and 6.4 μM (Figure 4.12) found in the case of free enzyme inhibition by phosphate, validating the applicability of this modeling approach. Simulated concentration profiles for the product, substrate, and competitive inhibitor are shown for

each experimental substrate concentration in Figure 4.3. Product concentrations are shown normalized to the concentration of enzyme, for comparison with experiment; the substrate and inhibitor concentrations shown are those for the capsid interior phase. Both the substrate and inhibitor concentrations in the capsid are observed to reach a steady state within the first few seconds of the simulation.

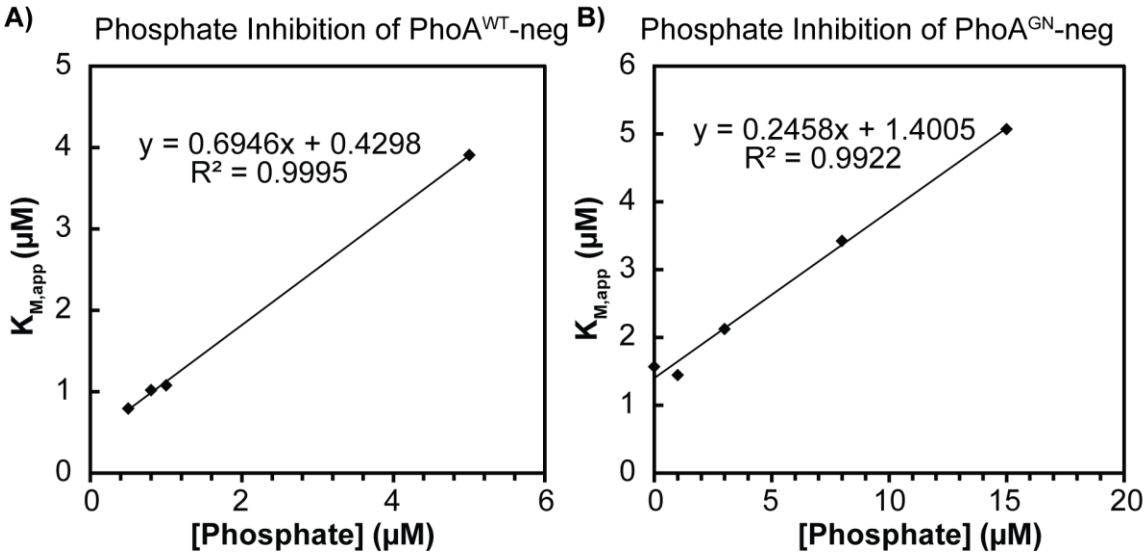


Figure 4.12: Experimental determination of K_i for A) Free PhoA-neg and B) PhoA-neg-GN in medium salt Tris buffer.

The transport parameters k_1 and k_2 and the competitive inhibition constants K_I fitted for the $PhoA^{WT}$-neg case and for the case using the $PhoA^{D153G/D330N}$-neg enzyme are shown in Table 4.1. By considering the geometry of the capsids, the phosphate transport parameter k_2 can be reinterpreted as an apparent diffusion coefficient for transport from the capsid interior across the pores of the capsid shell into the assay bulk (Table 4.4; see also Appendix A for calculations). We note that this conversion is dependent on many physical parameters that are not known with certainty; the apparent diffusivity should therefore be regarded as an estimate. The apparent diffusivity calculated in this way is much smaller than the diffusion coefficient of free phosphate in water (5×10^{-6} cm^2/s). It is possible that there are significant Coulombic effects on the diffusion of charged species through the pores, since the radius of the pores of the MS2 capsid is on the same order of magnitude as the Debye length of the system. Therefore, these lower apparent diffusivities are not unreasonable, and might be expected to be modulated by the amino acid character of the residues surrounding the pore, as we observe.

Parameter	Fitted value
k_1 (1/s)	$\geq 1.02\times10^{4}$
k_2 (1/s)	6.08×10^{2}
D_2 (cm^2/s)	1.64×10^{-9}
$K_{I(WT)}$ (μM)	1.74
$K_{I(GN)}$ (μM)	2.67

Table 4.4: Fitted parameters from kinetic model and calculated apparent diffusivity D_2.

We assessed the numerical uncertainty associated with each fitted parameter by calculating the residual error between the model and the experimental data for a range of values for each fitted parameter. A more rapid increase in the residual as the parameter value moves away from its fitted value indicates lower uncertainty in its estimated value from fitting. Plots of the residual error for perturbations in each parameter for the $PhoA^{WT}$-neg and the $PhoA^{D153G/D330N}$-neg cases can be found in Figure 4.13. In particular, we find that the fitted value of k_1 represents at best a lower bound on its value: the residual error is not observed to increase significantly even with large increases in the value of k_1. The intracapsid substrate concentration was observed to rapidly equilibrate with the substrate concentration in the bulk due to the small volume of the capsid phase relative to the assay bulk; increases in the substrate transport parameter k_1, therefore, have little effect on the steady state substrate concentration encountered by encapsulated enzymes. Due to this numerical uncertainty, we did not attempt to interpret this parameter as an apparent diffusivity. The phosphate escape parameter k_2, on the other hand, has a strong impact on the steady state inhibitor concentration in the capsid and hence the enzymatic rate.

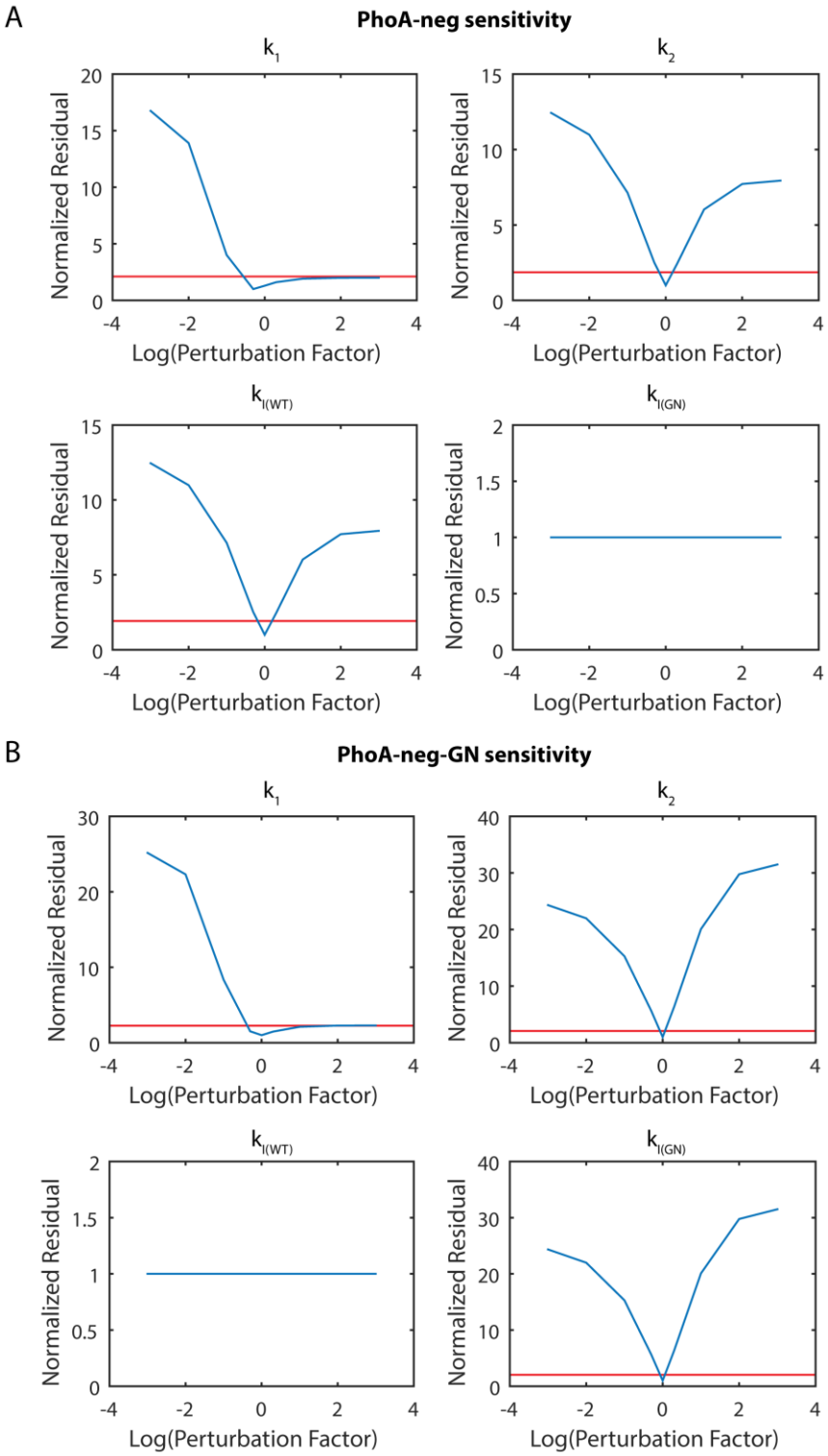


Figure 4.13: Sensitivity analysis for kinetics parameters. The residual (sum of the square errors) of T71E MS2 capsid kinetics model relative to the experimental data is shown for A.) PhoA-neg and B.) PhoA-neg-GN with each fitted kinetic parameter (k_1, k_2, KI_{wt}, and KI_{GN}) multiplied by the perturbation factor (ranging from 0.001 to 1000). The residuals are normalized to the best fit found by the nonlinear minimization routine. The

red line on each plot indicates a normalized residual of 2. The goodness of fit of the model is found not to be sensitive to increases in the substrate diffusion parameter k_1, but is found to be sensitive to decreases in the substrate diffusion parameter k_1, and to the phosphate diffusion parameter k_2 and the competitive inhibition parameters KI_{wt} and KI_{GN}.

In order to test whether our model could describe the behavior of both the $MS2^{WT}$ and $MS2^{T71E}$ systems, we estimated the change in k_2 required to recapitulate the behavior of the $MS2^{WT}$ system by computing the residual error between the model and the $MS2^{WT}$ experimental data as a function of the value of k_2, beginning with the value found by fitting to the $MS2^{T71E}$ experimental data (Figure 4.13). Since the intrinsic enzyme parameters were initially fitted to the $MS2^{WT}$ system, the model will approximate the $MS2^{WT}$ system in the limit of very high k_2 (no transport limitation). We find that, in the case of the $PhoA^{WT}$-neg enzyme, the residual approaches its limiting value upon a 100-fold increase in k_2, and upon a 100- to 1000-fold increase in k_2 for the $PhoA^{D153G/D330N}$-neg case. Thus, changes in k_2 alone, as might be caused by the amino acid mutation T71E in the MS2 capsid, are sufficient to describe the behavior of both systems in our model.

In addition to the assumption of two well-mixed phases and the assumption that the intrinsic kinetics of an enzyme in the $MS2^{WT}$ capsid are unaffected by the T71E pore mutation, the model does not account for modes of inhibition other than product inhibition by phosphate, nor for variations in the amount of enzyme loaded in each individual capsid in a sample. Despite these simplifications, we find that our model, in which substrate and inhibitor diffusion through the pore are inhibited in the negatively charged $MS2^{T71E}$ pore relative to $MS2^{WT}$, can account for the decrease in $k_{cat,app}$ observed in the $MS2^{T71E}$ capsid, as well as the increase in $K_{M,app}$, for both $PhoA^{WT}$-neg and $PhoA^{D153G/D330N}$-neg. Typically, competitive inhibition in a Michaelis-Menten reaction scheme increases the $K_{M,app}$ but does not affect $k_{cat,app}$; in this case however, trapping of the phosphate ion leads to increased inhibitor concentrations at higher rates of reaction (Figure 4.3), causing a decrease in the observed $k_{cat,app}$. The results of this highly simplified model are consistent with the hypothesis that changes in pore charge can affect transport into and out of the nanoreactor and thus affect observed enzyme kinetics.

4.4 Discussion

Here we show that a porous capsid-based nanocompartment can be modified to have specific effects on encapsulated enzymatic reactions by regulating substrate and product transport into and out of the system. We expect this type of transport control to be especially relevant in the development of multi-enzyme systems, where local substrate concentrations are important to enzymatic rates. In our model system, a large negatively charged substrate must overcome a Coulombic barrier to reach the enzyme. Similarly, the product must overcome a barrier to escape, potentially leading to product inhibition of the enzyme. It is important to note that this system cannot enhance the enzymatic rate; we instead demonstrate the potential of both controlled access to the interior of a protein compartment and product build-up. If the system is expanded to the study of compartmentalization of a multi-enzyme pathway, this effect could be useful to increasing local concentration of intermediates. Increased local intermediate concentration may alleviate bottlenecks in compartmentalized metabolons caused by rate-

limiting downstream enzymes. This is a key advantage in the enhancement of flux through synthetic enzymatic pathways. In the future, this study can be used to develop computational models for other designed compartments with selectivity based on size, charge, or hydrophobicity. Theories generated can then be applied to systems such as bacterial microcompartments for the production of fuels, chemical feedstocks, and pharmaceuticals in bacterial hosts.

Chapter 5: Studying microcompartment assembly using a virus-like particle model system

The following is under consideration at the Proceedings of the National Academy of Sciences (USA) (Asensio, Morella, Jakobson, Hartman, Glasgow, Sankaran, Zwart, and Tullman-Ercek).

5.1 Background

In addition to modeling transport through the pores of the Pdu MCP shell, we sought to use our MS2 VLP model system to explore the fundamental principles governing MCP and other protein structure assembly. To this end, we conducted a high-throughput selection for VLP assembly to identify assembly-competent MS2 *cp* alleles using deep sequencing. Of particular interest in this approach is the comparison of deep sequencing analysis of the naïve and selected pools, allowing rigorous determination of enrichment and selection. An analysis of Shannon information theoretic entropy at each residue of the ORF readily identified a key amino position at which selection resulted in diversification away from the wild-type allele, and the predominant mutation at this position proved to have fascinating structural properties.

5.2 Materials and methods

MS2 capsid library generation

Molecular biology reagents were purchased from NEB and chemicals from Fisher, unless otherwise indicated. The MS2 coat protein was amplified using primers (MA_014F and MA_015R) containing BamHI and NcoI restriction enzyme sites in preparation for ligation into an expression vector. Error-prone PCR was performed with these primers using a GoTaq® polymerase in 1x GoTaq® buffer with 5.5nmM $MgCl_2$, 0.4 mM dATP, 0.4 mM dGTP, 2 mM dCTP, 2 mM dTTP, 0.5 μM primers and 0.5 ng template DNA. One round of 30 PCR cycles generated an average of 1.6 mutations per gene (an error rate of 0.4%), as determined by Sanger sequencing of 10 library members. The library was mutagenized three times using this protocol, leading to an average of five mutations per gene. The PCR product of the capsid protein gene mutagenesis was then ligated into a pTRC99a (Ampicillin resistant, Lac promoter) expression plasmid as follows: the PCR product and plasmid vector were digested with the BamHI and NcoI restriction enzymes, purified by agarose gel electrophoresis and extraction, then ligated using T4 DNA Ligase. The plasmid library was transformed by electroporation into *E. coli* DH10B, and plated onto large Lysogeny Broth (LB) agar plates containing carbenicillin antibiotic at a concentration of 50 mg/L. Transformants were allowed to grow on plates overnight at 30°C (to prevent lawn formation), after which colonies were pooled and innoculated into carbenicillin-containing LB media and mixed. 125 μL aliquots of each pooled library of transformants were stored at -80°C in 15% glycerol. The final library size was estimated to be approximately 1.3×10^5 unique clones based on a viable cell count of transformants. We used the PEDEL-AA tool[137] to calculate an expected number of approximately 9×10^4 distinct full-length proteins encoded by the library of mutants.

VLP mutant expression and purification

Coat protein gene mutants were expressed from the pooled library members by thawing a frozen stock of library transformants containing approximately 1.3×10^5 unique *cp* alleles, inoculating the entire aliquot into carbencilllin-containing LB media, and incubating the culture overnight at 37°C with shaking. 400 μL of the overnight culture, containing approximately 10^9 bacterial cells, was added to 400 mL of antibiotic-containing LB media and incubated with shaking at 37°C until the culture reached an optical density at 600 nm wavelength (OD_{600}) of ~0.4. At this time, expression of mutant *cp* alleles in the culture was induced by the addition of IPTG to a final concentration of 1 mM, and the culture was allowed to grow to saturation overnight at 37°C. The culture was harvested by centrifugation at 6,000 × *g* for 10 min; the cell pellet was resuspended in 10 mM phosphate buffer pH 8.0, and cells were disrupted by sonication. Insoluble material released from cells was sedimented by centrifugation at 15,000 × *g* for 10 min, and the soluble fraction containing MS2 VLPs retained.

Intact capsids precipitated overnight at 4°C from the soluble fraction upon the addition of 40% PEG-8000 and 5 M NaCl to a final concentration of 10% PEG and 0.5 M NaCl. The capsids were further purified using column chromatography, first by anion exchange chomatography (DEAE-Sephadex® resin) and next by size-exclusion chromatography (Sephacryl S-1000 SF resin). Eluent fractions were collected at various time points and assayed for absorbance at 260 nm, indicating the presence of the nucleic acid cargo of intact capsids, and examined by polyacrylamide gel electrophoresis (12.5% acrylamide gel, 120 V applied potential for 1 hr) and Coomassie staining to detect the 13.7 kDa CP protein.

Nucleic acid sequencing

Libraries were prepared for next-generation sequencing by PCR with custom primers with NGS-specific adapters and barcodes (Table 5.1). An RNEasy kit (Qiagen) was used to isolate template RNA from libraries of intact capsids, and reverse transcription and PCR amplification (RT-PCR) was completed using a OneStep RT-PCR kit (Qiagen). Sample preparation and sequencing was completed on a MiSeq instrument (Illumina) by the QB3 Vincent J. Coates Genomics Sequencing Laboratory in Berkeley, CA.

QuickChange Mutation	Forward Primer	Reverse Primer
S37P	gaatggatcagctctaacCcgcgttcacaggcttac	gtaagcctgtgaacgcgGgttagagctgatccattc
N87C	catggcgttcgtacttatgcatggaactaaccattcc	ggaatggttagttccatgcataagtacgaacgccatg
N36A	ctgaatggatcagctctGCCtcgcgttcacaggcttacaaagtaac	gttactttgtaagcctgtgaacgcgaGGCagagctgatccattcag
N98A	ccattccaattttcgctacgGCTtccgactgcgagcttattg	caataagctcgcagtcggaAGCcgtagcgaaaattggaatgg
36A/37P	ggtcgctgaatggatcagctctGCCCCGcgttcacaggcttacaaagtaacc	gttactttgtaagcctgtgaacgCGGGGCagagcgatccattcagcgac
37P/98A	N98A primers were used with S37P template	
36A/37P/98A	N98A primers were used with N36A/S37P template	

MiSeq Primers	
Forward	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTgaccatggggagatatacatatg
Reverse 1	CAAGCAGAAGACGGCATACGAGATTAGTGCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGATCCcgcaagcttag
Reverse 2	CAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGATCCcgcaagcttag
Reverse 3	CAAGCAGAAGACGGCATACGAGATTGGTCAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGATCCcgcaagcttag
Reverse 4	CAAGCAGAAGACGGCATACGAGATATTGGCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGATCCcgcaagcttag
Reverse 5	CAAGCAGAAGACGGCATACGAGATTCCATGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGATCCcgcaagcttag

Table 5.1: QuikChange and MiSeq primers used in this study.

Analysis of nucleic acid sequencing data

Raw sequencing reads were trimmed using cutadapt [138] merged with FLASH [139] and translated with the Biostrings R package. Over 2 million reads were analyzed for both the naïve and selected VLP pools to calculate the amino acid content at each position in the *cp* gene. The Shannon entropy $S = \sum_{i=1}^{20} P_i \log(P_i), i \in A, \ldots, Y)$, where P_i is the probability of residue i occurring at the position in question, was calculated for each amino acid position in each pool. A Shannon entropy of 0 indicates no sequence diversity (all reads at the position code for the same amino acid) and a Shannon entropy of 1.3 indicates maximum diversity (a given read is equally likely to code for any amino acid at that position). The Shannon entropy of each position in the naïve pool was subtracted from the Shannon entropy of that position in the selected pool to give a measure of the sequence diversification away from the wild-type allele that occurred at each position as a result of the selection. Increased Shannon entropy indicates enrichment of a mutant allele in this case, since the wild type residue greatly predominates at each position in the naïve library.

Recapitulation of cp *mutants*

Mutants of the MS2 *cp* gene were recapitulated by Quickchange mutagenesis using the primers below. All VLPs were purified using the same lysis, precipitation and chromatography steps as for wild-type MS2 capsids.

Crystallization of Cp^{S37P} VLP

Protein used for crystallization experiments was expressed and purified as before, then subjected to an extra two rounds of size-exclusion chromatography and precipitation. The resulting protein was approximately 95% pure as determined by densitometry of protein separated by polyacrylamide gel electrophoresis and Coomassie staining. Capsid protein was resuspended from precipitation to a concentration of approximately 10 mg/ml and screened using the hanging drop method against a Molecular Dimensions HT- ProPlex™ crystallization screen. Larger crystals were grown from promising conditions (0.1 M Tris, 1.5 M ammonium sulfate pH 8.0 and 0.1 M Tris,

1.6 M lithium sulfate pH 8.0) in 4 µl hanging drops. Crystals grown in 0.1 M Tris, 1.5 M ammonium sulfate pH 8.0 yielded the best diffraction.

X-ray crystallography data acquisition, structure determination and refinement
X-ray diffraction data were obtained at the Advanced Light Source at beamline 501 at Lawrence Berkeley National Lab. Diffraction data was indexed and processed using the HKL2000 package (Otwinowski and Minor). Data quality and twin checks were performed with Phenix.xtriage (Zwart et al., 2005). The structure was solved fully automatically using MRage of the phenix package (Adams et al., 2010), utilizing phaser (McCoy et al., 2007) for molecular replacement. The structure of the wild type [PDBID 2MS2] was used as a search model and an unambiguous solution was readily obtained. Further structure refinement was carried out with phenix.refine (Afonine et al., 2012), which was iterated with rounds of manual model adjustment in Coot (Emsley & Cowtan, 2004). Root-mean-square deviation differences from ideal geometries for bond lengths, angles and dihedrals were calculated with Phenix (Adams et al., 2010). The overall stereochemical quality of all final models was assessed using the program MOLPROBITY (Davis et al., 2007).

Generation of VLPs incorporating the Mat protein

Capsid protein was co-expressed with maturation protein from pTrc IPTG inducible plasmid. VLPs were purified using techniques described above. Approximately 50 µg of VLPs were heated to 85°C for 15 min and analyzed via protein electrophoresis as described below. Bands with the molecular weight of the MS2 maturation protein were analyzed via tandem mass spec as described below.

Analysis of VLP attachment to the F-pilus of E. coli

F^+ XL1-Blue cells (Stratagene) were grown overnight with shaking in LB media at 37°C and after 16 h, subcultured with the addition of VLPs as follows. VLPs were added to culture at a MOI of 1000 VLPs to 1 *E coli* bacterium, assuming a final OD_{600} of ~0.4. Cultures were incubated, unshaken, at 37°C for 1 h. Samples were centrifuged at 4000 x g for 3 min, re-suspended in 100 µl of distilled water and prepared for TEM as described below.

Protein electrophoresis

Denaturing protein electrophoresis was performed in a 12.5-15% polyacrylamide gel with 120V applied potential for 1 h to detect the 13.7 kDa CP protein and Mat protein. Staining was performed with Coomassie dye by the method of Studier (36).

Dynamic light scattering

Purified capsids were analyzed by Dynamic Light Scattering (DLS) using a Zetasizer Nano ZS (Malvern) and SoftMax Pro software. A volume distribution was used to calculate particle diameters.

Transmission electron microscopy

Purified capsids were diluted to approximately 0.5 mg/ml concentration, incubated on Formvar-coated copper mesh grids for 5 min, wicked off using filter paper

and briefly dried. The grids were washed with water, which was immediately wicked off and dried. Next, a 1.6% $UO_2(OAc)_2$ stain was applied to the grids for 90 s, then wicked off and dried. TEM images were acquired using an FEI Technai 12 transmission electron microscope with an accelerating voltage of 120 kV. Measurement of particle diameter was performed using ImageJ software.

Chemical conjugation to VLPs

Cysteine residues of CP^{N87C} and $CP^{S37P/N87C}$ were alkylated with AlexaFluor 488 (Life Technologies) as previously described (37). 5 equivalents of AlexaFluor 488 maleimide in a 20 mM stock solution in DMSO were added to a solution of CP^{N87C} or $CP^{S37P/N87C}$ MS2 (final monomer concentration of 35 μM) in 10 mM pH 7.2 phosphate buffer. The reaction mixture was incubated at room temperature for 2 h. After 2 h, the excess dye was removed using 10 kDa molecular weight cutoff spin concentrators (EMD Millipore). Samples were analyzed via mass spec as described below (analysis for CP protein).

Mass spectrometry analysis for CP *Protein*

Liquid Chromatography Mass Spectrometry analysis was performed as previously described (36). Briefly, electrospray ionization mass spectrometry was performed with an Agilent 1260 series HPLC connected to an Agilent 6224 time-of-flight TOF mass spectrometer (Santa Clara, CA). Proteins were separated on a Proswift RP-4H C18 column (1.0 mm x 50 mm, Thermo). A gradient from 20% to 95% acetonitrile in .1% formic acid was run at .4 mL/min. Approximately 35 pm of protein was injected on column. Data was analyzed with Agilent MassHunter Qualitative Analysis B.05.00.

Mass spectrometry analysis for Maturation Protein

Performed at University of California Davis's campus mass spec facilities.
Protein bands were excised manually and transferred to Eppendorf tubes (0.2 mL). Protein-containing gel pieces were washed with a series of washes that included alternating additions of 100 μL 0.1 M ammonium bicarbonate followed by 100 μL of 100% ACN in order to remove traces of coomassie staining. Once gel slices were thoroughly destained, they were dried using a speed-vac. Gel pieces were then subjected to 10 mM DTT for 20 min at 56 °C and subsequently alkylated with 20 mM iodoacetamide for 1 h at room temperature in the dark. Acetonitrile was then added to shrink the gel pieces and the pieces were thoroughly dried in the speed-vac prior to addition of 0.1 ml of a solution containing 12.5 ng/μl trypsin in 50 mM ammonium bicarbonate. Gel pieces were allowed to swell at 4°C for 3 h prior to overnight digestion at 37°C. Liquid was collected following digestion and the gel pieces were bathed in 100% ACN to adequately shrink the gel pieces. All liquid was collected into one tube and evaporated to dryness using a speed-vac. The peptide solution was reconstituted into 10 μl of 2% ACN with 0.1% TFA and assessed for peptide concentration based on the absorbance at 280 nm using a Nanodrop spectrophotometer. One microgram total peptide was loaded onto the nano-ACQUITY UPLC™ chromatographic system. Peptides were loaded and separated on a C18 Trizaic Naotile using a 60 min RP gradient at 450 nL/min (3–40% ACN over 40 min). The column temperature was set at 45°C. Lock mass (Leucine enkephalin (556.2771 Da), 250 fmol/mL) was constantly

infused by the NanoAcquity auxiliary pump at a constant flow rate of 1 μL/min with lockspray scans set at intervals of 45 s. The Xevo QTof ™ mass spectrometer (Waters) was programmed to switch between low (6 eV) and high (18–42 eV) energies in the collision cell, with a scan time of 1 s per function over a mass range of 50–2000 Da. LC-MS^E data were processed with ProteinLynx GlobalServer v2.3 (Waters) and searched in the associated *E. coli* protein database (UniProtKB/SwissProt Protein Knowledge Base).

4.3 Results

Supramolecular structures are ubiquitous in nature; almost all proteins are thought to interact with other proteins or biomacromolecules in the course of their function. An understanding of the fundamental rules that govern the assembly of protein monomers into multi-protein structures, such as viral capsids, would be a powerful tool in understanding the structure, function, and evolution of proteins that form such supramolecular assemblies. In 1994, Creamer and Rose issued the Paracelsus challenge: to change the fold adopted by a monomeric protein while changing less than half of its amino acid sequence [140]. Over the intervening years, many minimal changes in protein sequence, some as small as a single amino acid, have been found to confer dramatic changes in the fold of monomeric proteins [141]. However, while the structures of viral capsids and related assemblies can be altered by substantial changes in the monomer sequence [142–145], no single amino acid mutations have yet been discovered to stably alter the structure of a multimeric assembly.

To explore this possibility, we employed a simple model of multiprotein assembly, the bacteriophage MS2. Each infectious MS2 particle consists of an icosahedral capsid of approximately 27 nm in diameter that encapsidates its positive-sense single-stranded RNA genome. The capsid comprises 178 copies of the coat protein (CP) and one copy of the maturation protein (Mat, or A-protein) [146,147]. The CP forms dimers that then arrange with the Mat protein such that there are 12 pentamers and 20 hexamers in the icosahedron, in a configuration described as a T=3 geometry. MS2 is extensively structurally characterized [148,149,129,150], and presents an attractive target for engineering because its dimeric subunits can reassemble *in vitro* to form an icosahedral virus-like-particle (VLP). The VLP is structurally similar to MS2 phage with an identical T=3 geometry, and the VLP can form even in the absence of the Mat protein [127]. This T=3 icosahedral structure is common to many virus families, including the *Nodaviridae*, *Caliciviridae*, and *Leviviridae* (of which bacteriophage MS2 is a member). It is proposed that these families arose from ancestors exhibiting a T=1 geometry composed entirely of pentamers, though this has not been proven [151]. The MS2 VLP has been broadly investigated for nanoscale patterning, drug delivery, and the study of enzyme encapsulation in part due to its amenability to chemical modifications and its ability to disassemble and reassemble *in vitro* around diverse forms of cargo [152–155,127,156]. The CP can tolerate several amino acid mutations that render it more amenable to targeted post-translational modifications for such applications [152–154,156,157]. As with other multimeric assemblies, however, no point mutations are known to confer uniform changes in the icosahedral supramolecular structure adopted by the CP. If these could be identified, the mutants would not only shed light on the structural evolutionary landscape of a common viral geometry, but also provide useful

tools for exploring the effects of VLP size on engineered MS2 VLP performance for drug delivery and imaging applications.

Here we report that a single amino acid change to the MS2 CP confers a radical change in the MS2 VLP supramolecular structure from a T=3 icosahedron of 27 nm diameter to a T=1 icosahedron of approximately 17 nm diameter consisting of 12 pentamers. Additionally, this MS2 CP variant retains a similar dimer conformation to the wild type, despite adopting a smaller capsid structure. Moreover, it remains amenable to *in vitro* reassembly, supports the incorporation of the Mat protein that is required for infection, and tolerates a mutation that has been previously used to perform *in vitro* coupling of small molecules to the interior of the MS2 VLP and thus highlights that many potential applications of MS2 VLPs will extend to this new structure. These results provide new insights into long-standing debates on the evolution of viruses and gene-transfer agents, as well as the possibility for multiple stable forms of multi-protein assemblies. Finally, we expect this work to inspire new possibilities for the computational design of made-to-order protein-based nanomaterials.

In order to identify mutants of the MS2 *cp* gene that encode CP variants retaining the ability to assemble into VLPs, error-prone PCR was used to generate a pool of mutated MS2 *cp* alleles. Approximately five nucleotide mutations were introduced in each allele relative to the wild-type, and the final library size was approximately 10^5 unique *cp* genes. The mutant alleles were expressed in *E. coli* (one allele per bacterial cell), and the resulting naïve pool was subjected to a chromatographic selection for VLP assembly. Since each assembly-competent VLP encapsidated the nucleic acids encoding the peptides in its shell, deep sequencing of the nucleic acids (DNA and RNA) from the naïve library and selected pool was used to reveal *cp* alleles enriched by selection for assembly. We focused on alleles identified after one round of VLP packaging, that is, plasmid-based expression in *E. coli* and purification by precipitation and size exclusion chromatography for VLP-size particles, followed by VLP disruption and nucleic acid isolation [158]. We assessed enrichment at each amino acid position by comparing the Shannon information theoretic entropy at each position and discovered that only residue Ser37 demonstrated an increase in Shannon entropy upon selection (Figure 5.1). Upon closer inspection, we were surprised to find that 70.5% of the reads in the pool selected for assembly encoded a Ser37Pro mutation (as compared to 2.8% of reads in the naïve library). This mutation is observed in the literature only twice, arising from selections in two separate directed evolution studies by Peabody and colleagues, but it was not pursued further in either study and was not characterized for size [159,160]. Given that the mutation arose only when selections did not include an infection step, we speculated that perhaps this variation does not yield infectious particles, precluding its observation in the myriad other evolution studies on MS2 VLPs. We therefore sought to examine the structural and biochemical properties of this highly abundant CP^{S37P} variant.

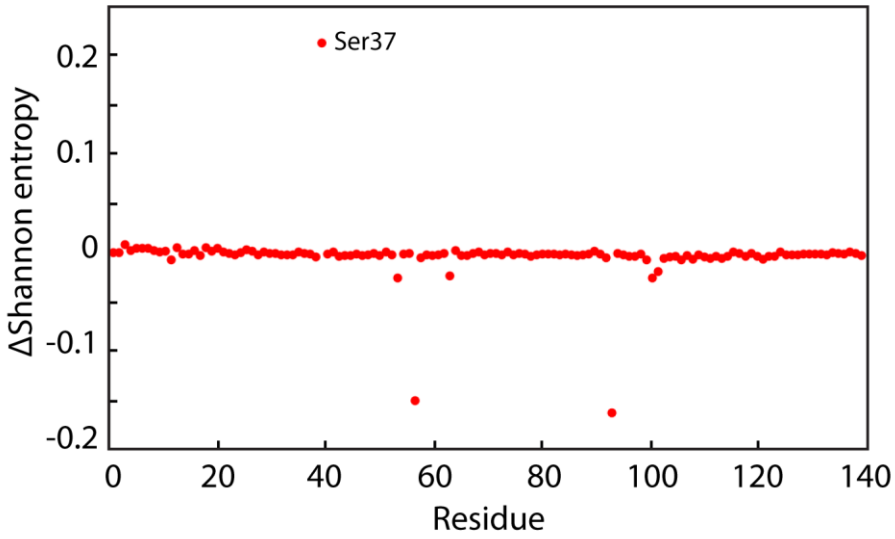


Figure 5.1: Deep sequencing reveals an abundant mutant of the *cp* gene upon selection for VLP assembly. Shown is the difference in Shannon entropy between naïve *cp* library and pool selected for assembly at each amino acid position encoded by the *cp* gene.

The mutant *cp* allele encoding the CP^{S37P} mutant was recapitulated and expressed in *E. coli*, and the resulting VLPs exhibited similar properties to wild-type MS2 capsids throughout purification. Remarkably, transmission electron microscopy (TEM) revealed that the particles formed of CP^{S37P} are smaller than the VLPs formed by the wild-type CP (Figure 5.2A). This observation was confirmed by dynamic light scattering and size-exclusion chromatography (Figure 5.3). We hypothesized that this smaller particle was likewise formed from CP dimers, but assembled into a different geometry. We therefore crystallized the CP^{S37P} VLP and determined its structure by X-ray diffraction. The structures of the pentameric asymmetric unit (ASU) and T=1 icosahedral VLP of CP^{S37P} (PDB ID: 4ZOR) are shown alongside the structures of the trimeric ASU and T=3 icosahedral VLP of CP^{WT} (PDB ID: 2MS2) (Figure 5.2C). Crystallographic statistics can be found in Table 5.2 and a stereoscopic projection of a portion of electron density map can be found in Figure 5.4. The five peptide chains in the CP^{S37P} ASU have nearly identical conformations but differ slightly in the "FG" loop, which surrounds the pore in the center of the pentamer and is within a region which is known to affect assembly [129,150]. Interestingly, the proline residue at position 37 adopts a *cis* conformation in all five chains of the CP^{S37P} ASU, in contrast to the *trans* serine residue present at this position in the CP^{WT} VLP. This results in a slight but clear difference in the conformation of a small loop directly adjacent to the S37P mutation (Figure 5.2B). Notably, the size of the VLP is likely too small to permit encapsidation of the full viral genome, supporting our hypothesis that this mutation was not observed previously because it cannot support infection [161].

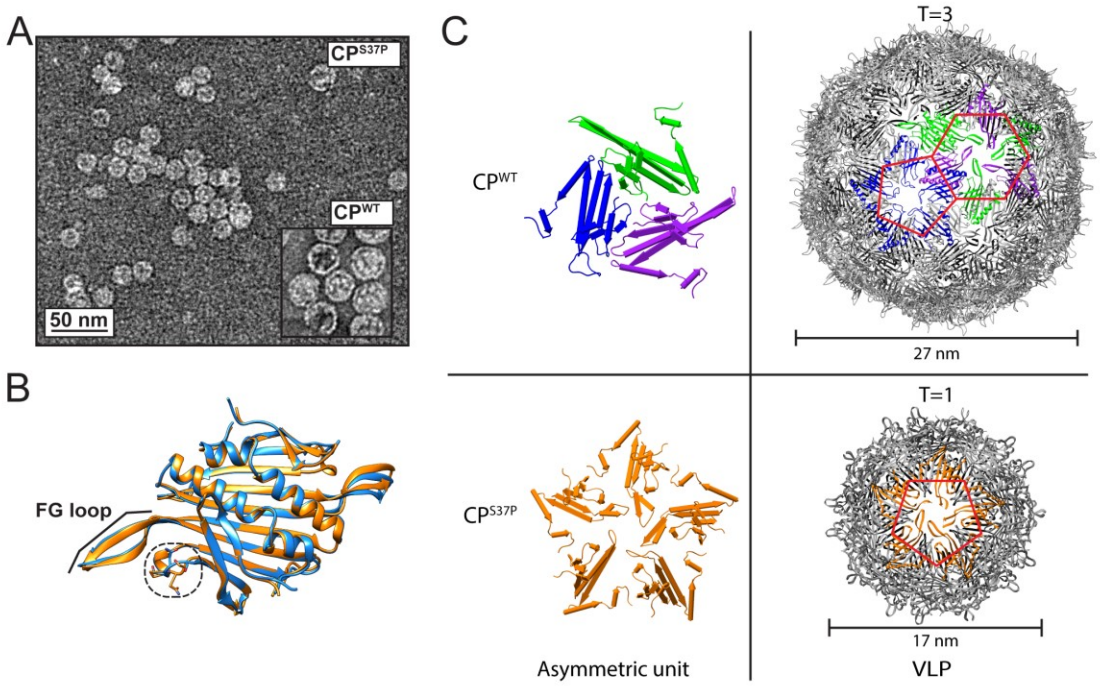


Figure 5.2: CP^{S37P} dimers form a VLP of 17 nm diameter with a T=1 icosahedral structure. (A) Transmission electron micrograph of CP^{S37P} VLPs and CP^{WT} VLPs (inset); (B) structural alignment of CP^{S37P} (orange) and CP^{WT} (blue) dimer structures; (C) crystallographic asymmetric unit structures (left) and complete VLP structures (right) from crystal structures of CP^{WT} (top; 2MS2) and CP^{S37P} (bottom; 4ZOR) VLPs. Hexameric and pentameric capsomeres are outlined in red to highlight the T=1 and T=3 icosahedral geometries.

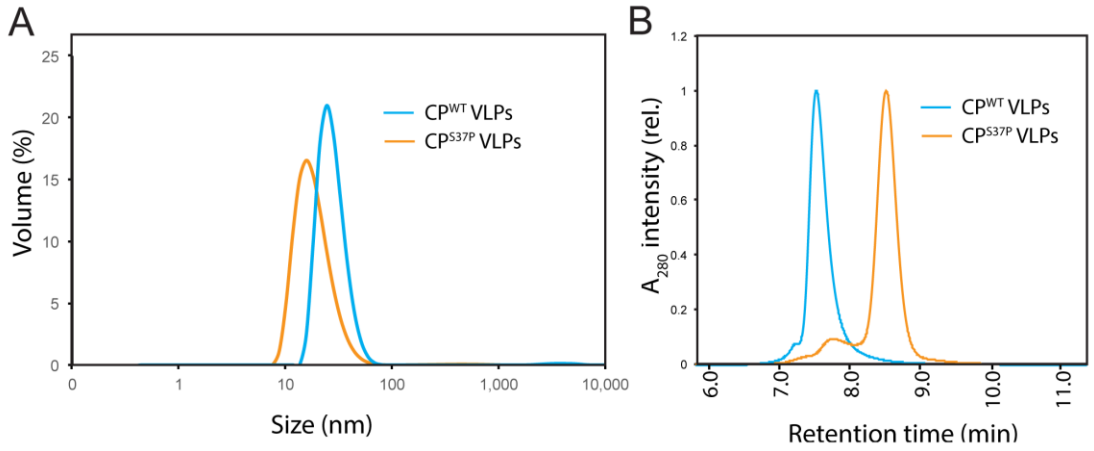


Figure 5.3: Dynamic light scattering and size exclusion chromatography resolve CP^{S37P} VLPs from CP^{WT} VLPs. (A) Histogram of size of VLPs formed of CP^{S37P} (orange) and CP^{WT} (blue) as determined by dynamic light scattering; (B) absorbance at 280 nm (A_{280}) of effluent from size exclusion chromatography of VLPs formed of CP^{S37P} (orange), CP^{WT} (blue).

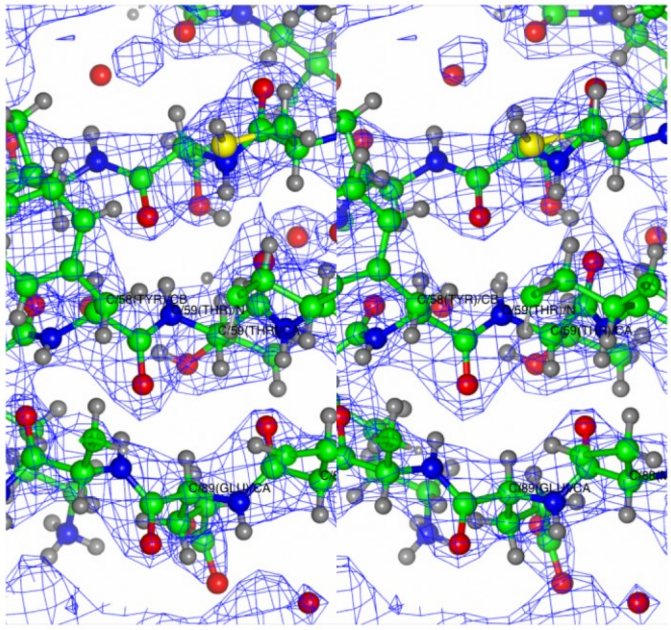


Figure 5.4: Stereoscopic projection of electron density of X-ray crystallographic structure of CP^{S37P} VLP (4ZOR).

Examination of the crystal structures leads to a fascinating question: given the surprisingly strong similarities between the CP^{WT} and CP^{S37P} dimer structures, how did a T=1 structure become spontaneously favored in the case of CP^{S37P}? With this in mind, we observed that key structural changes occur at the dimer-dimer interface. The inter-dimer angle between neighboring dimers was smaller in the CP^{S37P} VLP than in the CP^{WT} VLP, as expected given the formation of a smaller T=1 icosahedral supramolecular assembly. We speculated that this change in inter-dimer interaction was due to differences in hydrogen bonding. Indeed, several side chains near residue 37 are positioned differently in the CP^{S37P} structure as compared to the CP^{WT} structure, resulting in different sets of hydrogen bonding interactions at the dimer-dimer interface. In particular, the *cis* configuration of Pro37 appears to change the conformation of the side chain of Asn36 to interact with Asn98, favoring a reconfiguration of the hydrogen bonding interactions and changing the positioning and angle of the dimer-dimer contact. Two hydrogen bonds were identified in the CP^{S37P} VLP near the Pro37 residue: the first between the carbonyl oxygen of Ile94 and the backbone nitrogen of Ser39, and the second between the side-chain oxygen of Asn98 and the nitrogen of the Asn36 side chain (Figure 5.5). In the wild-type structure, on the other hand, a hydrogen bond forms between the backbone nitrogen of Asn98 and the side-chain oxygen of Asn36.

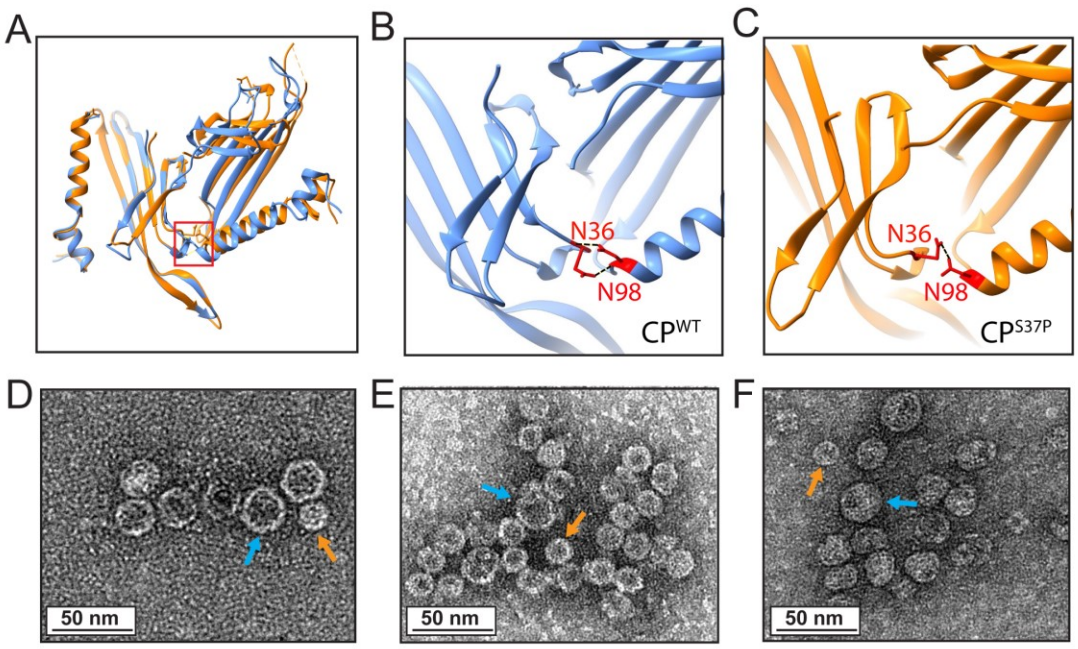


Figure 5.5: Hydrogen bonding contributes to the stability of T=1 icosahedral CP^{S37P} VLPs. (A) Structural alignment of CP^{S37P} (orange) and CP^{WT} (blue) inter-dimer interface structures; residues in (B) CP^{WT} and (C) CP^{S37P} structure altered to probe hydrogen bonding interactions; transmission electron micrographs of VLPs formed of (D) $CP^{N36A/S37P}$, (E) $CP^{S37P/N98A}$, and (F) $CP^{N36A/S37P/N98A}$. Blue arrows indicate VLPs of CP^{WT} size (27 nm); orange arrows indicate VLPs of CP^{S37P} size (17 nm).

To investigate the importance of these inter-dimer interactions in the CP^{S37P} VLP, we mutated the engineered *cp* gene to disrupt the potential hydrogen bonds described above, substituting an alanine at the Asn36 and Asn98 residues to create the double mutants $CP^{S37P/N36A}$ and $CP^{S37P/N98A}$ and the triple mutant $CP^{S37P/N36A/N98A}$. The resulting variants formed a mixed population of VLPs, comprising both 27 nm and 17 nm particles, as judged by TEM (Figure 5.5, Figure 5.6). Surprisingly, both the T=3 and T=1 icosahedral conformations are apparently accessible to the CP^{S37P} mutant when either or both of these hydrogen bonds are disrupted. However, the stability of these particles, as indicated by their melting temperature, was less than that of both CP^{WT} and CP^{S37P} VLPs (Table 5.2). We conclude that the hydrogen bond interaction between residues 36 and 98 stabilize the T=1 configuration of CP^{S37P} VLPs to a larger degree than the T=3 configuration of CP^{S37P} VLPs.

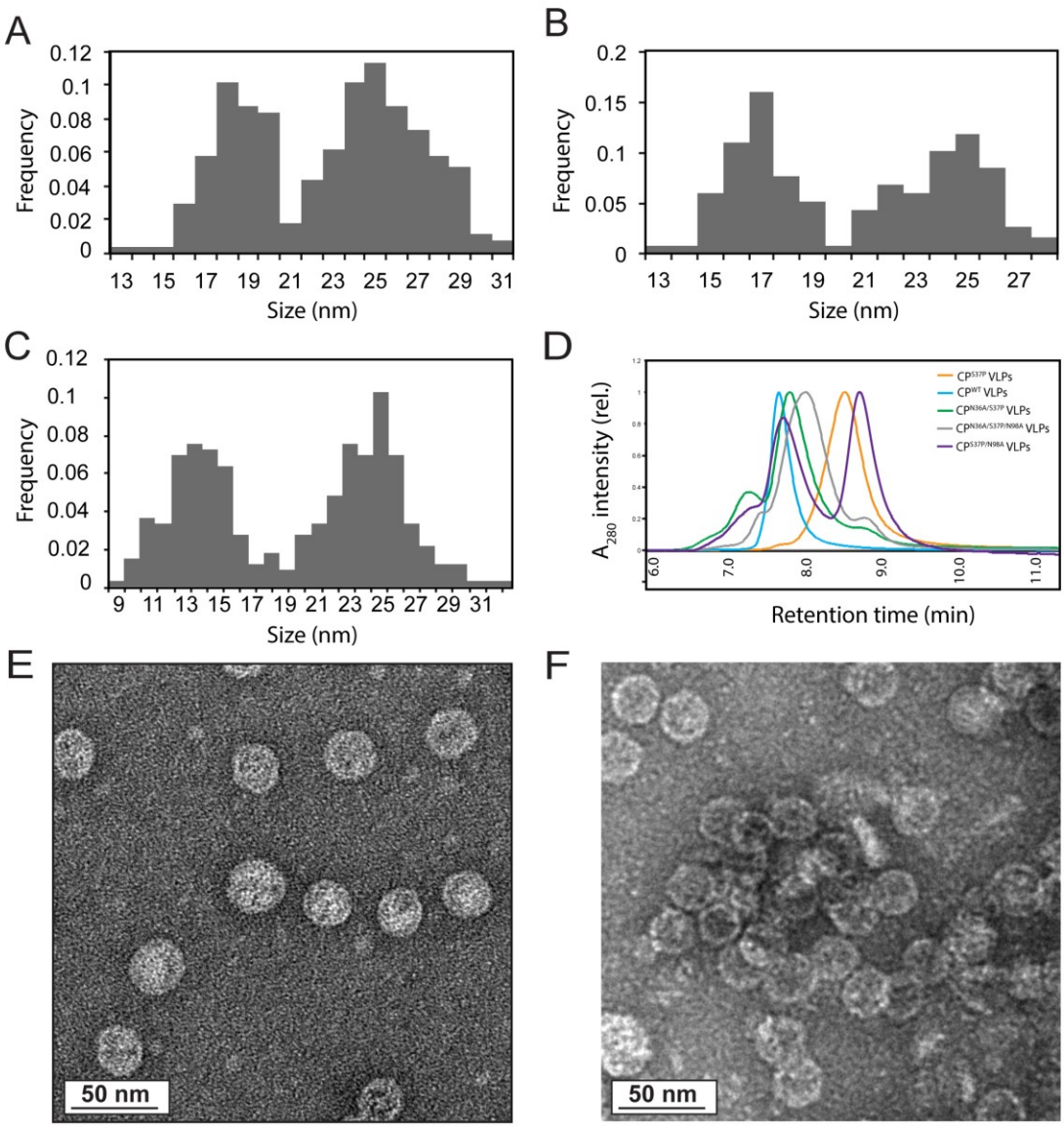


Figure 5.6: Hydrogen bonding contributes to the stability of T=1 icosahedral CP^{S37P} VLPs. Histograms of size of VLPs formed of (A) $CP^{N36A/S37P}$, (B) $CP^{S37P/N98A}$, and (C) $CP^{N36A/S37P/N98A}$ as determined from transmission electron micrographs; (D) absorbance at 280 nm (A_{280}) of effluent from size exclusion chromatography of VLPs formed of CP^{S37P} (orange), CP^{WT} (blue), $CP^{N36A/S37P}$ (green), $CP^{S37P/N98A}$ (purple), and $CP^{N36A/S37P/N98A}$ (grey); transmission electron micrographs of (E) CP^{N36A} and (F) CP^{N98A} VLPs.

Resolution range	43.9 - 2.202 (2.281 - 2.202)
Space group	F 2 3
Unit cell	228.125 228.125 228.125 90 90 90

Total reflections	249217 (23228)
Unique reflections	49435 (4867)
Multiplicity	5.0 (4.8)
Completeness (%)	0.99 (0.99)
Mean I/sigma(I)	7.70 (1.02)
Wilson B-factor	34.36
R-merge	0.1448 (1.577)
R-meas	0.1617 (1.774)
CC1/2	0.995 (0.747)
CC*	0.999 (0.925)
Reflections used in refinement	49102 (4596)
Reflections used for R-free	1994 (189)
R-work	0.2177 (0.3868)
R-free	0.2425 (0.4348)
Number of non-hydrogen atoms	5204
macromolecules	4752
ligands	20
Protein residues	630
RMS(bonds)	0.004
RMS(angles)	1.00
Ramachandran favored (%)	98
Ramachandran allowed (%)	1.5
Ramachandran outliers (%)	0.16

Rotamer outliers (%)	0.96
Clashscore	2.31
Average B-factor	47.23
macromolecules	47.26
ligands	113.58
solvent	43.85
Number of TLS groups	30

Table 5.2: Table of crystallographic data. Statistics for the highest-resolution shell are shown in parentheses.

We also generated the analogous single-point mutations in the wild-type *cp* gene, yielding CP^{N36A} and CP^{N98A} proteins. The corresponding VLPs formed wild-type size particles of 27 nm diameter, as determined by TEM (Figure 5.5, Figure 5.6), but also had lower melting temperatures (Table 5.3). The assembly of these proteins into VLPs of wild-type size indicates that the other inter-dimer interactions that remain are sufficient to result in the formation of wild-type-like VLPs. Moreover, it suggests that energetic considerations beyond hydrogen bonding may be relevant to VLP formation.

VLP	AVG (°C)	STDEV
CP^{WT}	68.3 (n=8)	0.8
CP^{S37P}_Peak1	66.6 (n=8)	0.46
CP^{S37P}_Peak2	77.5 (n=5)	0
CP^{N36A}	55.5 (n=3)	0
CP^{N98A}	64.8 (n=3)	0.58
$CP^{N36AS37P}$	48.7 (n=5)	0.97
$CP^{S37PN98A}$	Could not be determined	
$CP^{N36AS37PN98A}$	52.8 (n=6)	1.17

Table 5.3: Melting temperatures of VLPs formed of CP^{WT}, CP^{S37P}, CP^{N36A}, CP^{N98A}, $CP^{N36A/S37P}$, $CP^{S37P/N98A}$, and $CP^{N36A/S37P/N98A}$ as determined by SYPRO® Orange thermal scanning.

The MS2 VLP can be reassembled *in vitro* to encapsidate RNA, DNA, negatively charged polymers, or proteins bearing a negatively charged peptide tag. As such, it is a useful biotechnological system for applications such as the delivery of a nucleic acid, protein, or small-molecule payload to cells by receptor-mediated endocytosis [155,162,163], as well as applications in nanoscale patterning and studies of enzyme encapsulation. We find that VLPs composed of CP^{S37P} possess many of the same useful

properties as VLPs formed of wild-type CP. For example, CP^{N87C} is a useful variant that provides a cysteine residue in the VLP interior for modification via maleimide conjugation chemistry [154]. We found that $CP^{S37P/N87C}$ formed VLPs of 17 nm diameter amenable to small molecule conjugation using Alexa Flour® 488 maleimide, similarly to wild-type CP VLPs (Figure 5.7). Moreover, VLPs comprising the CP^{S37P} variant could be disassembled and re-assembled around negatively charged cargo *in vitro* in similar conditions to those identified for wild-type MS2 (Figure 5.8) [127]. Interestingly, when disassembled dimers from both CP^{WT} and CP^{S37P} were combined at varying molar ratios, the resulting population of VLPs was comprised of both 17 nm and 27 nm-sized VLPs, as determined by size exclusion chromatography and TEM. LC-MS analysis of SEC fractions show that the smaller size population was comprised primarily of CP^{S37P} protein and the larger of CP^{WT}, suggesting that the two variants assemble separately.

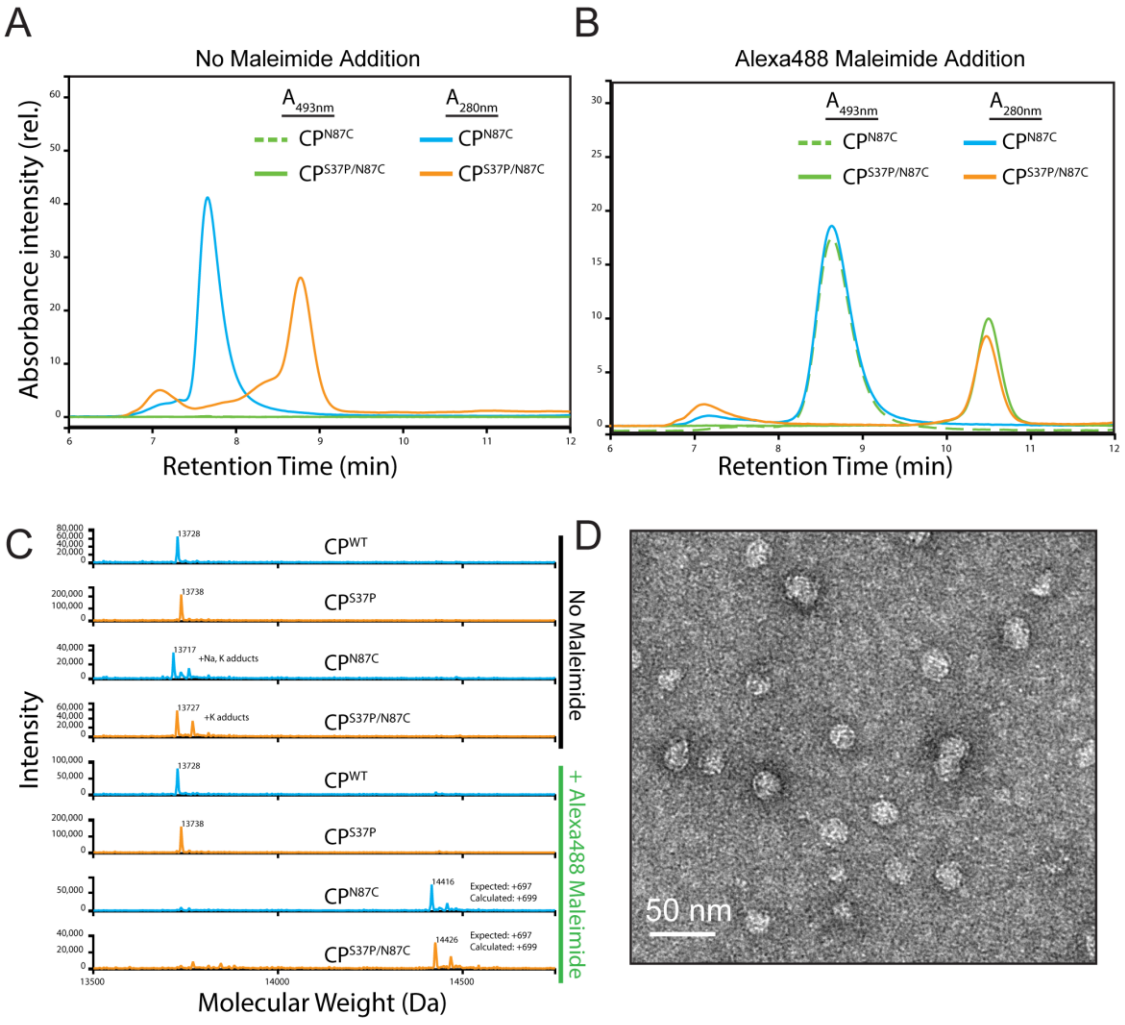


Figure 5.7: $CP^{S37P/N87C}$ VLPs are amenable to specific chemical conjugation with Alexa488 maleimide. Absorbance at 493 nm (A_{493}; green traces) and at 280 nm (A_{280}; blue/orange traces) of effluent from size exclusion chromatography of VLPs formed of CP^{N87C} (green dashed and blue traces) or $CP^{S37P/N87C}$ (green solid and orange traces) without (A) or with (B) maleimide addition; mass spectrometry of SEC effluent of VLPs

formed of CP^{WT} CP^{N87C} (blue traces) or CP^{S37P}/$CP^{S37P/N87C}$ (orange traces) or and without maleimide addition; (D) transmission electron micrograph of VLPs formed of $CP^{S37P/N87C}$.

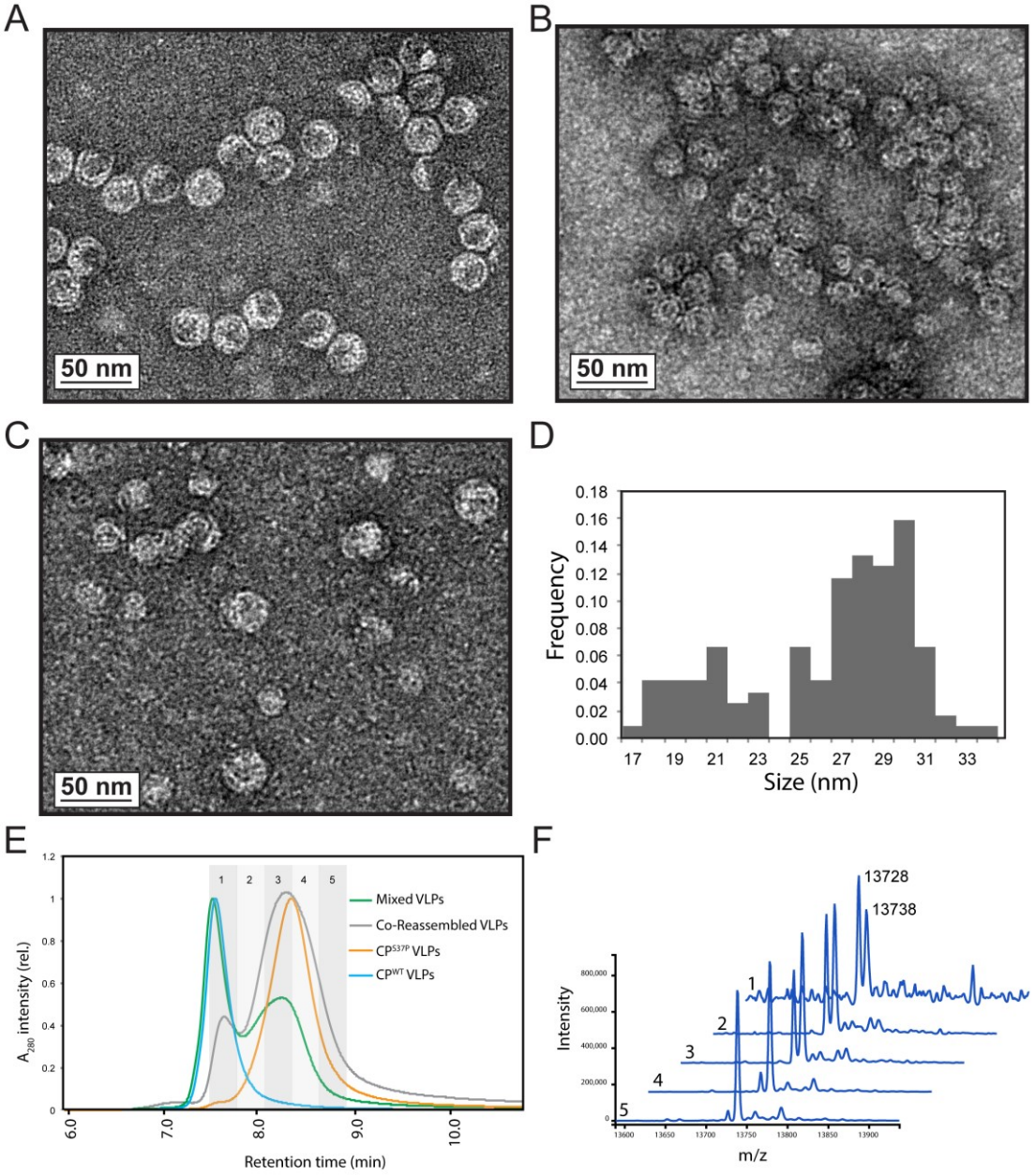


Figure 5.8: CP^{S37P} dimers reassemble *in vitro* to form VLPs. Transmission electron micrographs of VLPs reassembled *in vitro* from dimers of (A) CP^{WT}, (B) CP^{S37P}, and (C) a 4:1 molar mixture of CP^{WT} and CP^{S37P} dimers; (D) histogram of VLP size determined from transmission electron micrograph of VLPs reassembled *in vitro* from a 1:1 molar mixture of CP^{WT} and CP^{S37P} dimers; (E) absorbance at 280 nm (A_{280}) of effluent from size exclusion chromatography of mixed, separately assembled VLPs (green), VLPs reassembled from mixed CP^{WT} and CP^{S37P} dimers (grey), reassembled CP^{S37P} VLPs (orange), and CP^{WT} VLPs (blue); (F) mass spectrometry of SEC effluent fractions 1-5

from VLPs reassembled from mixed CP^{WT} and CP^{S37P} dimers as indicated in the grey A_{280} trace in (E). CP^{WT} *m/z* is approximately 13728; CP^{S37P} *m/z* is approximately 13738.

We expect that the CP^{S37P} mutation occurs frequently during replication of the virus in nature, and wondered whether the resulting capsids are capable of interacting with the *E. coli* host even though the small size of the CP^{S37P} VLP prohibits encapsidation of the full MS2 genome. We therefore examined the ability of CP^{S37P} VLPs to incorporate the MS2 Mat protein, also known as the Assembly (A) protein. The Mat protein is required for infection of F^+ *E. coli* and is thought to interact with wild-type capsid C/C dimer at the 2-fold axis [164]. When the genes encoding both CP^{WT} and Mat are expressed simultaneously, approximately one Mat protein is incorporated into each CP^{WT} VLP [147]. When we isolated CP^{S37P} VLPs produced in this manner and analyzed their protein content by gel electrophoresis, a band was evident at the expected molecular weight of the Mat protein, and its identity was confirmed as the Mat protein by mass spectrometry (Table 5.4). We applied these VLPs to a culture of MS2-susceptible F^+ *E. coli* and observed attachment of CP^{S37P} VLPs to the F-pilus in a manner similar to that observed for the CP^{WT} VLP, as assessed by examination of the pili by TEM (Figure 5.9, Figure 5.10). This indicates that the CP^{S37P} mutant retains the ability to incorporate the maturation protein in the virion, and thus also the ability to interact with the *E. coli* surface like an infectious particle.

	Protein Name	**Mass (Da)**	**% Seq. Coverage**	**Unique Peptide Count**	**% Total Spectra**	**Uniprot Acc.**
WT_A	Bacteriophage MS2 Maturation Protein	43954	29	12	43	C8XPG7
	Bacteriophage MS2 Coat Protein (Fragment)	6630	25.1*	2	7.1	F4ZWL8
S37P_A	Bacteriophage MS2 Assembly (Maturation) Protein	44079	28	10	37	F4ZWJ4
	Bacteriophage MS2 Coat Protein (Fragment)	6630	25.1*	2	7.4	F4ZWL8

Table 5.4: Mass spectrometry analysis of MS2 Mat protein. Peptide threshold of 90%
*Percent coverage of complete protein sequence

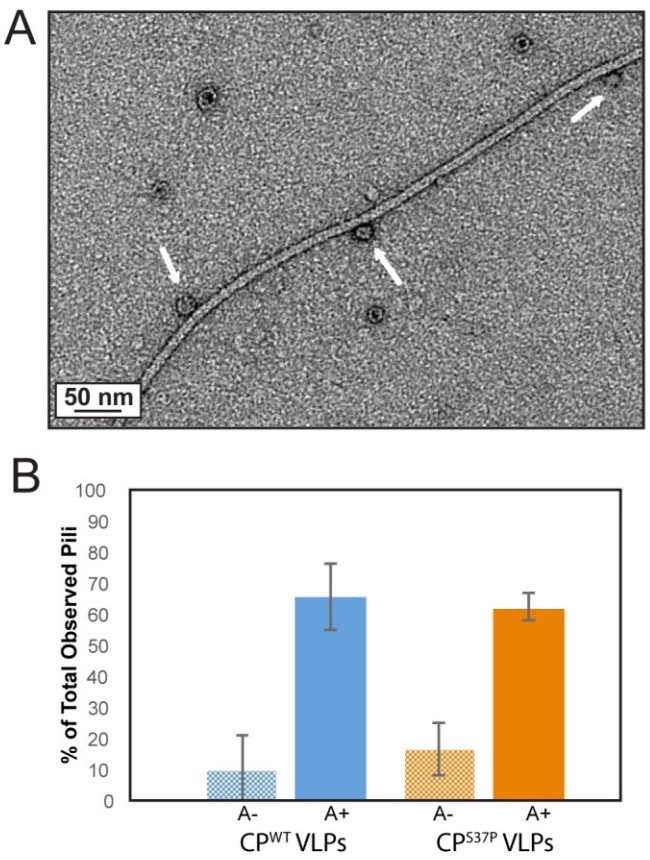


Figure 5.9: CP^{S37P} VLPs incorporate the Mat protein and attach to the F pilus of *E. coli*. (A) Transmission electron micrographs of CP^{S37P} VLPs aligned to *E. coli* F pilus (as indicated by white arrows); (B) histograms of fraction of CP^{WT} (blue) and CP^{S37P} (orange) VLPs aligned to *E. coli* F pilus with (A+) and without (A-) coexpression of the Mat protein.

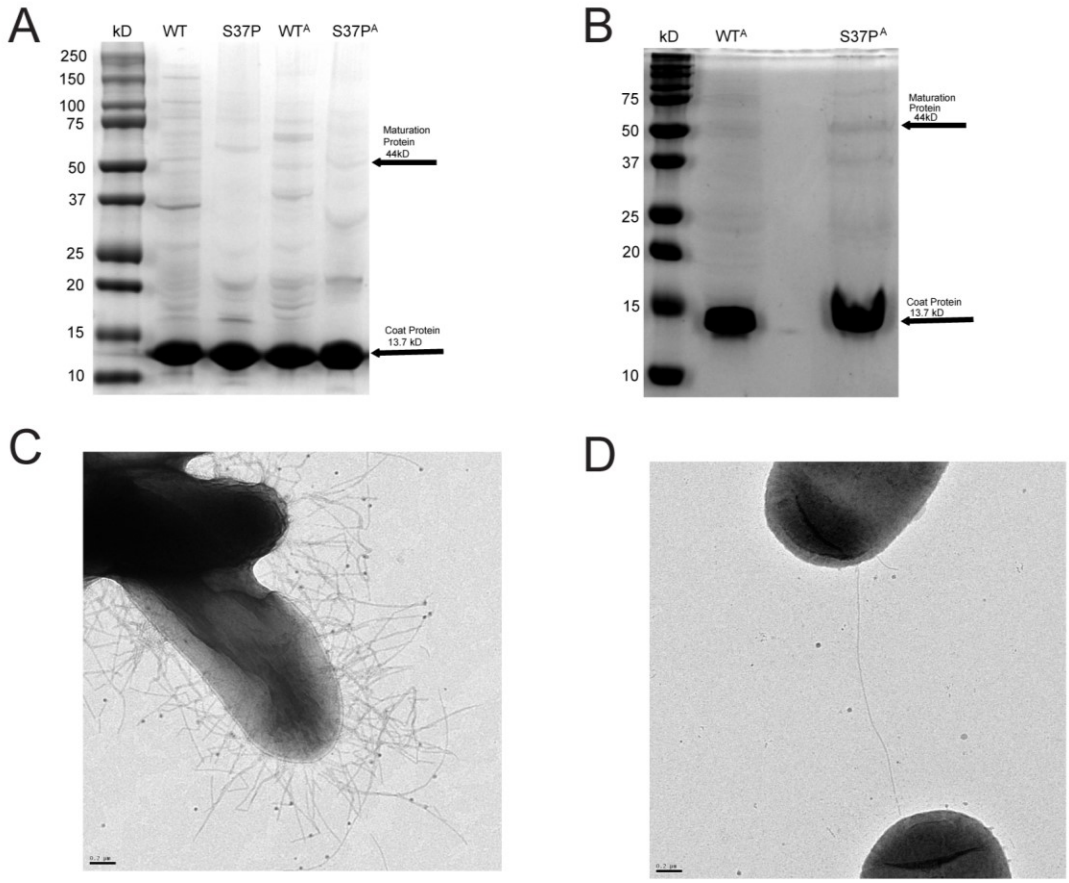


Figure 5.10: The Mat protein (A protein) is incorporated in CP^{S37P} VLPs and mediates attachment to the *E. coli* F pilus. (A) Coomassie-stained SDS-PAGE of VLPs composed of CP^{WT}, CP^{S37P}, CP^{WT+A}, or CP^{S37P+A}; (B) Coomassie-stained SDS-PAGE of VLPs composed of CP^{WT+A} or CP^{S37P+A} from which Mat protein bands were excised and analyzed using LCMS; (C) transmission electron micrograph of *E. coli* showing adhesive fimbriae; (D) transmission electron micrograph of *E. coli* showing the F-pilus.

The observation that a single amino acid change to the MS2 CP protein can mediate a stable switch in VLP structure from a T=3 to a T=1 icosahedron has broad implications for engineering VLP-based nanomaterials, the study of viral capsid evolution, and the *de novo* prediction of protein supramolecular structures. While previous studies demonstrated that altering assembly conditions and VLP cargo can cause perturbations in the final structure of a VLP [145,165–167], the CP^{S37P} mutant described here stably and uniformly adopts the alternate T=1 icosahedral structure in conditions identical to those in which the native CP adopts a T=3 icosahedral structure. The close similarities in primary sequence and dimer conformation of wild-type and CP^{S37P} particles mean that the two VLPs have nearly identical chemical properties, allowing for well-controlled comparisons to be made between protein containers of different sizes (27 nm and 17 nm, respectively) in engineering applications.

Intriguingly, our observation of a transition from a T=3 to T=1 icosahedral structure implies that a T=1 to T=3 transition could also occur as the result of a small sequence perturbation. Several single-stranded DNA and single- and double-stranded RNA viruses (such as the *Parvoviridae*, *Hepeviridae*, and *Partitiviridae*) form T=1 icosahedral capsids; the CP^{S37P} mutant suggests that these viruses might readily sample T=3 icosahedral structures, suggesting a potential route for the evolution of a larger virion with only a single amino acid change.

The ability of CP^{S37P} VLPs to encapsidate nucleic acid molecules present in the cellular milieu is reminiscent of gene transfer agents and polydnaviruses. Gene transfer agents are naturally occurring, genomically encoded phage-like particles that have, over the course of substantial changes to their structural genes, become too small to encapsidate their own genomes and instead encapsidate random segments of host DNA [168], while polydnaviruses are endosymbionts of parasitic wasps which likewise lack the ability to encapsidate their genomes [169,170]. The CP^{S37P} mutation demonstrates that a lysogen may lose the ability to encapsidate its entire genome as the result of a single amino acid change and indicates that such transitions may be more common than previously thought. While this type of behavior, in which a protein lies only a single amino acid change away from adopting a radically different conformation, was previously thought to be confined to proteins that had been extensively modified in the laboratory [171], we have shown that the wild-type MS2 *cp* allele exhibits such behavior with respect to its supramolecular structure.

5.4 Discussion

It is apparent that subtle changes in protein structure, and in particular in intermolecular interactions, can manifest as significant changes in the higher-order structure adopted by a multi-protein complex. Efforts to computationally predict the structure of multi-protein complexes and to design multi-protein structures *de novo* are still underway, and the behavior we observed in the case of the MS2 coat protein will be a valuable test case for methods to predict the structures of multi-protein assemblies. The change in inter-dimer angle seen in the CP^{S37P} T=1 icosahedral assembly as compared to the wild-type VLP occurs despite very little evident change in the conformation of the dimers themselves. While hydrogen bonding seems to contribute to the stability of the smaller T=1 icosahedral structure, other factors also contribute to the accessibility of the T=1 icosahedral structure to the CP^{S37P} mutant. Together, our experiments suggest that fine-grained models of protein-protein interactions will be required to fully describe the effects responsible for the T=3 to T=1 icosahedral transition we observed. Simulating these interactions for large structures made up of many protein subunits, such as viral capsids, is a uniquely challenging problem, with exciting implications in both protein engineering and evolution.

Chapter 6: Using computational approaches to study microcompartment function *in vivo*

The following is in preparation (Jakobson, Slininger, Tullman-Ercek, and Mangan).

6.1 Background

Above we discuss numerous hypotheses as to the precise function of the Pdu MCP and other MCP organelles. These hypotheses are challenging to test *in vivo* due to inherent difficulties in measuring metabolite concentrations in the organelles as compared to the cytosol or the media. For this reason, we pursued a computational approach encompassing a simple reacton-diffusion model in an attempt to quantitatively understand Pdu MCP function and computationally assess the feasibility of the various hypotheses that are consistent with experiment.

Bacterial microcompartments (MCPs) are protein-bound intracellular organelles used by *Salmonella enterica*, *Yersinia pestis*, *Klebsiella spp.*, and other enteric pathogens to spatially organize their metabolism [4,10,77]. MCP metabolons allow the growth of these pathogens on carbon and energy sources, such as 1,2-propanediol [1] and ethanolamine [3], that cannot be used by the host gut microbiota, conferring a competitive advantage upon invasion of the host gut. MCPs are typically approximately 150 nm in diameter, with multiple enzymes localized inside a shell consisting of a porous protein monolayer composed of several distinct shell proteins [1,2,72]; a typical bacterial cell may contain several MCP structures when in the presence of the appropriate substrate. Enzymes are localized to the MCP interior through the interactions of N-terminal signal sequences with the inward-facing C-termini of MCP shell proteins, as well as potentially through other uncharacterized interactions [15,16,50].

Inside the 1,2-propanediol utilization (Pdu) MCP metabolon, 1,2-propanediol metabolism proceeds as follows: 1,2-propanediol is converted to propionaldehyde by the vitamin B_{12}-dependent PduCDE holoenzyme [50], then propionaldehyde is converted to either 1-propanol by the NADH-dependent PduQ enzyme [172] or to propionyl-coA by the NAD^+-dependent PduP enzyme [85]. The PduP and PduQ enzymes are thought to recycle a private pool of NAD^+/NADH inside the MCP lumen [173,172,6]. 1-propanol is not used for cell growth, but propionyl-CoA can be utilized either as a carbon source or for ATP generation through substrate-level phosphorylation [20].

Pdu MCPs are elaborate multi-protein structures subject to exquisite regulation, and much investigation has focused on determining the detailed function of the organelles. Experiments suggest that metabolic pathways are sequestered in the Pdu and ethanolamine utilization (Eut) MCPs in order to protect the cell from toxicity associated with aldehyde intermediates [5], to prevent carbon loss from the metabolic pathway [6], and to provide a private pool of cofactors for the encapsulated pathways [173,6]. These hypotheses are difficult to examine directly due to difficulties associated with measuring the concentrations of small molecules inside the MCPs *in vivo*, so we propose to use a computational approach to assess their feasibility using a coupled reaction-diffusion model of Pdu MCP. For simplicity, we model the Pdu MCP as a spherical compartment in the center of a radially symmetric spherical cell. The model encompasses passive transport of 1,2-PD and propionaldehyde across the cell and MCP membranes, active transport of 1,2-PD into the cell, and the action of the PduCDE and PduP/Q enzymes

(Figure 6.1). We find that, for a broad range of parameter values, an analytical solution that assumes constant metabolite concentrations in the MCP lumen closely approximates the full numerical solution (Figure 6.2; see also Models). This analytical solution is used throughout the following analysis.

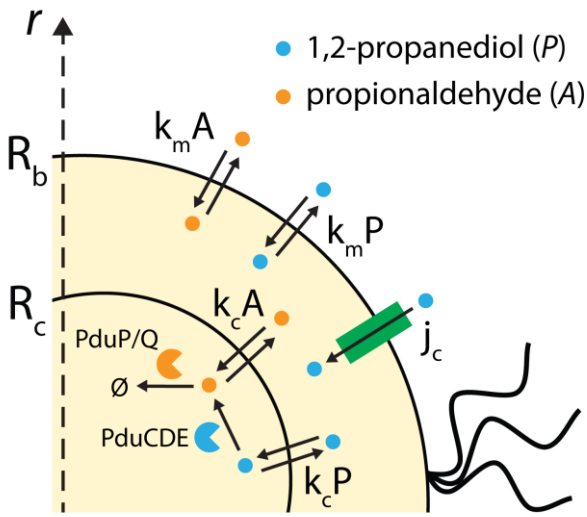


Figure 6.1: Reaction scheme.

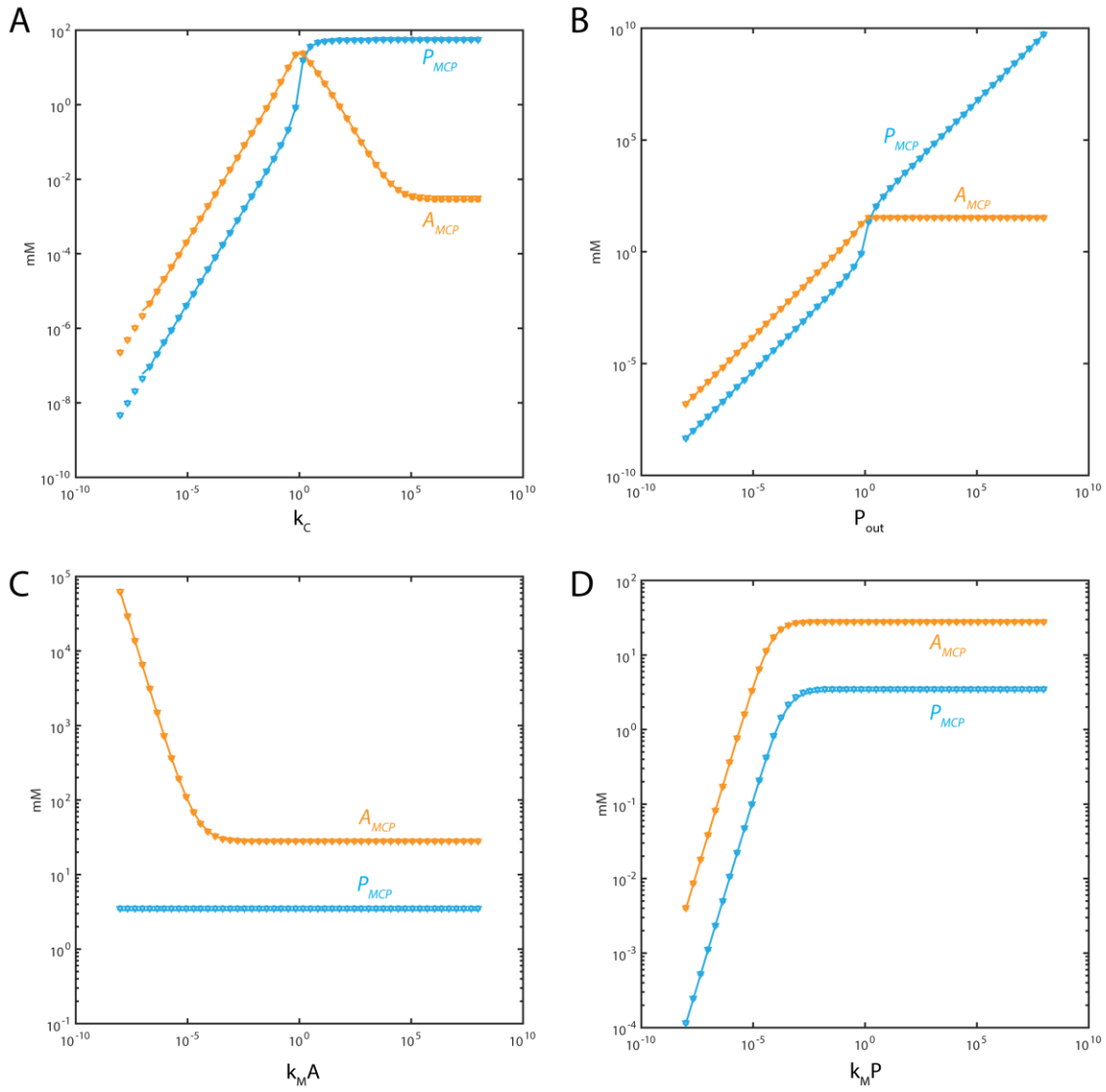


Figure 6.2: Comparison of analytical solution assuming constant concentrations in the MCP (solid lines) and numerical solutions from the edge (circles) and center (triangles) of the MCP for 1,2-PD (blue) and propionaldehyde (orange).

We find that aldehyde sequestration is the key function of the Pdu MCP, and contributes not only to decreasing aldehyde leakage into the cytosol and the growth medium, as is often discussed in the existing literature, but also to greatly increasing flux through the metabolon. Furthermore, we find that active 1,2-PD transport across the cell membrane is dispensable in some situations, but not at low external 1,2-PD concentrations. This transport activity has been proposed previously, but never experimentally observed. Finally, while selective MCP membrane permeability is not always required for encapsulated enzyme saturation, it is often advantageous in this regard. The predictions of our model agree well with existing experimental results and suggest several avenues for continuing computational and experimental investigation.

6.2 Model

See Appendix B for a detailed mathematical description of the model. The various parameter estimates for the model are shown in Table 6.1.

Parameter	Meaning	Estimated value	Units
j_c	Rate of active transport of 1,2-PD across the cell membrane	1	cm/s
k_cA	Permeability of the Pdu MCP to propionaldehyde	10^{-5}	cm/s
k_cP	Permeability of the Pdu MCP to 1,2-PD	10^{-5}	cm/s
R_b	Radius of the bacterial cell	$5x10^{-5}$	cm
R_c	Radius of the Pdu MCP	10^{-5} [1]	cm
D	Diffusivity of metabolites in the cellular milieu	10^{-5} [174]	cm^2/s
k_mA	Permeability of the cell membrane to propionaldehyde	10^{-3} [175]	cm/s
k_mP	Permeability of the cell membrane to 1,2-PD	10^{-3} [175]	cm/s
k_{catCDE}	Maximum reaction rate of a PduCDE active site	$3x10^2$ [176]	1/s
K_{MCDE}	Michaelis-Menten constant of PduCDE	$5x10^2$ [176]	μM
N_{CDE}	Number of PduCDE enzymes per cell	$1.5x10^3$ [2]	1/cell
k_{catPQ}	Maximum reaction rate of a PduP/Q active site	55 [172]	1/s
K_{MPQ}	Michaelis-Menten constant of PduP/Q	$1.5x10^4$ [172]	μM
N_{PQ}	Number of PduP/Q enzymes per cell	$2.5x10^3$ [2]	1/cell
P_{out}	External 1,2-PD concentration	$5.5x10^4$ [5]	μM
A_{out}	External propionaldehyde concentration	0 [5]	μM

Table 6.1: Parameter estimates used in the model.

6.3 Results

MCPs function to reduce toxicity by decoupling cytosolic aldehyde concentration from PduP/Q saturation- In order to assess the function of the Pdu MCP, we first wish to analyze the performance of the Pdu MCP system relative to alternative scenarios in which the encapsulated metabolic enzymes are instead evenly distributed throughout the cytosol or scaffolded together in the absence of a diffusion barrier. We can assess the function of each organization strategy by two criteria: maintenance of the cytosol propionaldehyde concentration below the toxicity limit of 8 mM [5] and saturation of the PduP/Q enzymes with their propionaldehyde substrate. Flux through the Pdu metabolon can only be maximized if these enzymes are saturated. The fulfillment of both these criteria is impossible if the enzymes are distributed throughout the cytosol, since the Michaelis-Menten constant of the PduP/Q enzymes is approximately 15 mM, above the propionaldehyde toxicity limit. In fact, if the PduCDE and PduP/Q enzymes are distributed throughout the cytosol, our model suggests that the steady-state propionaldehyde concentration is maintained several orders of magnitude below the toxicity limit (Figure 6.3, Figure 6.4). In turn, the PduP/Q enzymes are not saturated. In each organizational case, we examine the kinetically relevant propionaldehyde concentration: without MCPs, this is the cytosolic propionaldehyde concentration; with MCPs, this is the propionaldehyde concentration in the MCP.

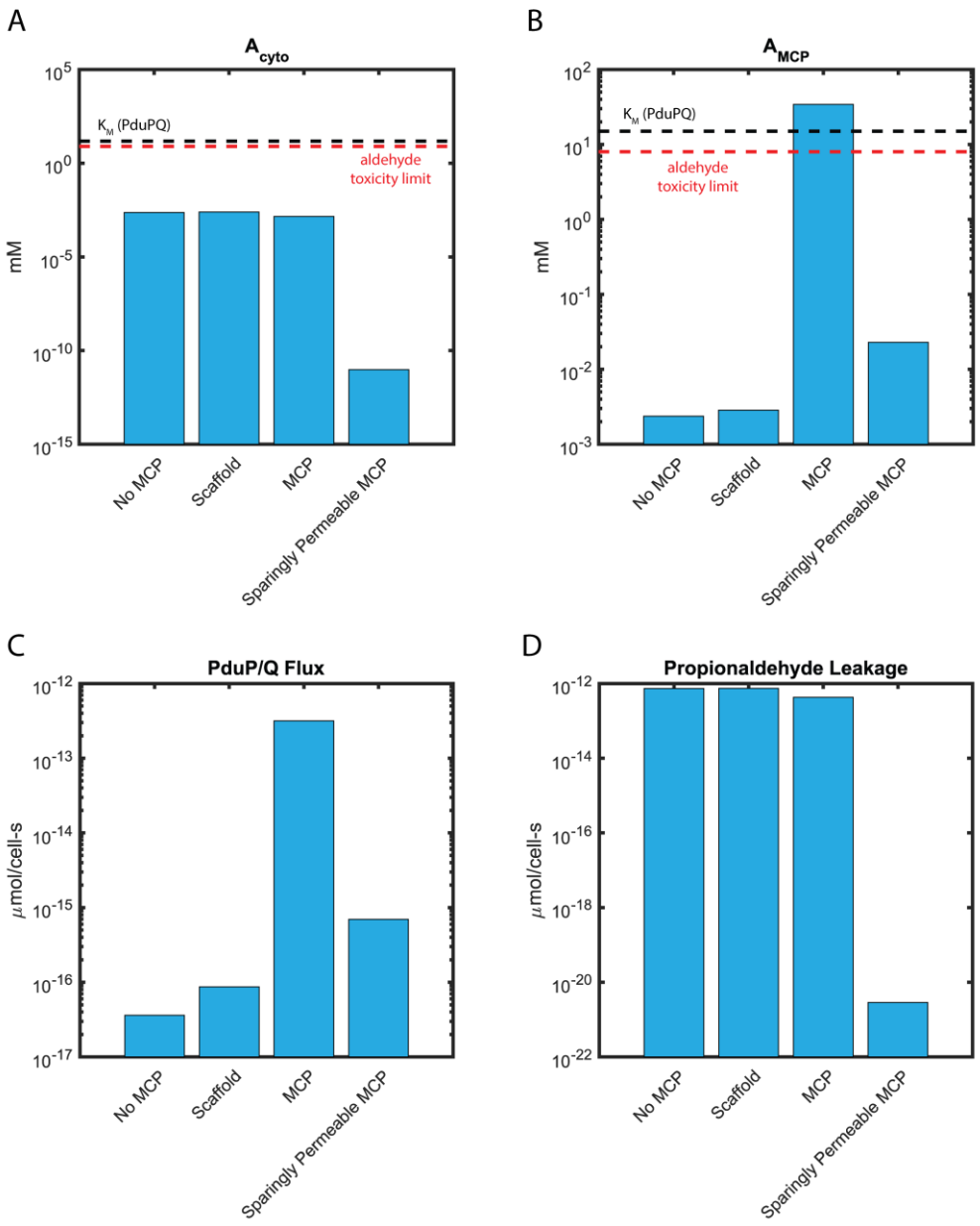


Figure 6.3: (A) Cytosolic propionaldehyde concentration (A_{cyto}), (B) kinetically relevant propionaldehyde concentration (A_{MCP}), (C) carbon flux through PduP/Q, and (D) propionaldehyde leakage across the cell membrane. Shown are the steady state concentrations and fluxes for cases without MCPs, with a scaffold, with MCPs, and with MCPs of extremely low permeability. The 15 mM K_M of PduP/Q is shown as a black dashed line; the 8 mM propionaldehyde cellular toxicity limit is shown as a red dashed line.

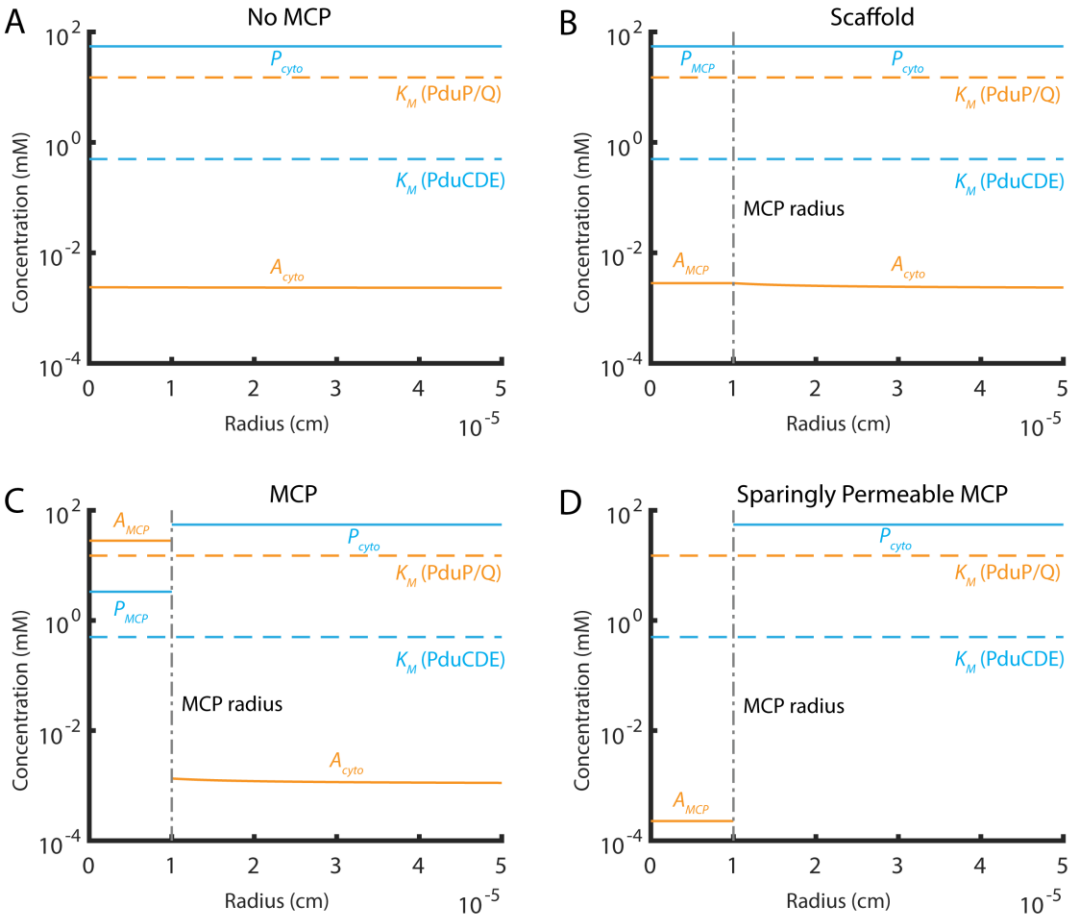


Figure 6.4: Concentration profiles as a function of *r* for a cell with (A) no MCPs; (B) a scaffold with no diffusion limitation ($k_c = 10^{10}$); (C) MCPs ($k_c = 10^{-5}$); and (D) sparingly permeable MCPs ($k_c = 10^{-10}$). 1,2-PD is plotted in blue and propionaldehyde in orange. The K_M of the PduCDE and PduP/Q enzymes are plotted in blue and orange dashed lines, respectively.

Another organizational strategy is that of a scaffold, in which the relevant enzymes are localized together without being surrounded by a diffusion barrier. In this case, the propionaldehyde concentration in the vicinity of the PduP/Q enzymes is somewhat higher than if the enzymes are distributed throughout the cytosol, but still much lower than the saturating concentration (Figure 6.3).

If we instead consider an MCP, in which the enzymes are localized within a diffusion barrier, the PduP/Q enzymes are exposed to a much higher propionaldehyde concentration (above the saturating concentration) while the propionaldehyde concentration in the cytosol remains very low (Figure 6.3).The presence of a diffusion barrier allows the MCP to decouple the cytosolic propionaldehyde concentration (responsible for toxicity) from the kinetically relevant propionaldehyde concentration in the vicinity of the PduP/Q enzymes. Very low nonspecific permeabilities of the diffusion barrier are unfavorable, however: a sparingly permeable MCP maintains a very low cytosolic propionaldehyde concentration, but also a very low concentration of propionaldehyde in the MCP (Figure 6.3).

A very low cell membrane permeability to propionaldehyde allows PduP/Q saturation by causing an accumulation of propionaldehyde in the cytosol and obviates the need for MCPs, but at the cost of cytosolic aldehyde concentrations above the toxicity limit (Figure 6.5). In addition, the membrane permeability to propionaldehyde required for this to occur is dramatically lower than a physiologically reasonable estimate.

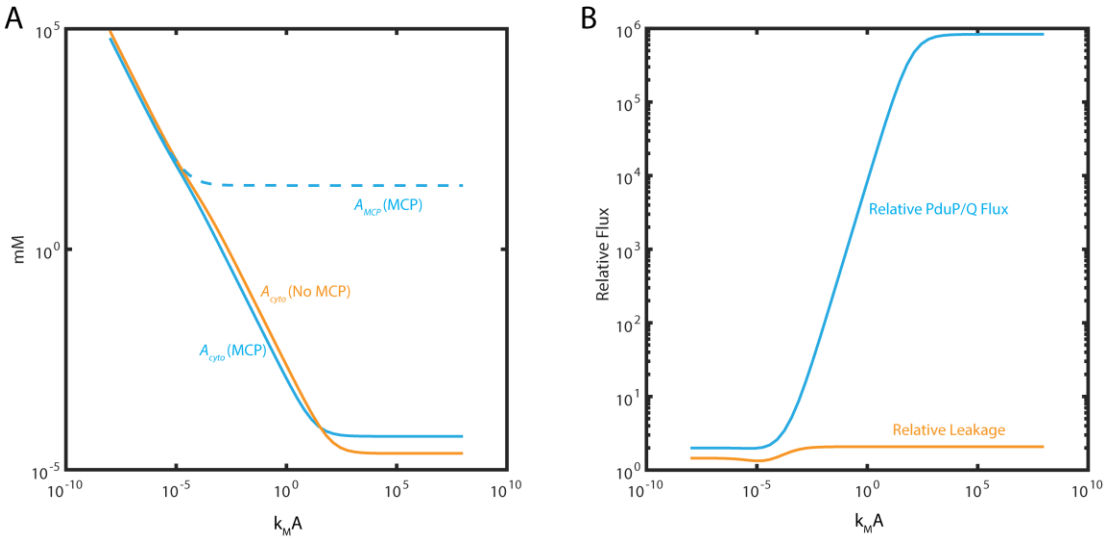


Figure 6.5 (A) Cytosolic aldehyde concentration (A_{cyto}) with and without MCPs and MCP aldehyde concentration (A_{MCP}) with MCPs; (B) relative carbon flux through PduP/Q ($flux_{MCP}/flux_{NoMCP}$) in blue and relative aldehyde leakage rate ($leakage_{NoMCP}/leakage_{MCP}$) in orange as a function of k_mA.

MCPs function to enhance pathway flux- Decoupling PduP/Q saturation from cytosolic propionaldehyde concentration allows significantly greater carbon flux through the MCP metabolon than in the cases of enzyme scaffolding or no organization (Figure 6.3C). Carbon flux per cell through PduP/Q in the MCP case ($2.97x10^{-13}$ µmol/cell-s) is several orders of magnitude higher than in either the scaffold or no MCP cases ($8.61x10^{-17}$ µmol/cell-s and $3.58x10^{-17}$ µmol/cell-s, respectively). Interestingly, this improvement is realized due solely to saturation of the PduP/Q enzymes; the flux through the PduCDE enzyme is similar with ($6.49x10^{-13}$ µmol/cell-s) and without ($7.41x10^{-13}$ µmol/cell-s) MCPs.
We validated our model by estimating the growth rate resulting from the predicted flux through PduP/Q in the MCP case. Our model predicts a flux of $2.97x10^{-13}$ µmol/cell-s for a cell with MCPs, equivalent to $1.74x10^{-5}$ pg/cell-s, when the external 1,2-PD concentration is 55 mM. Approximately one-half of this flux can be used, so assuming that a bacterial cell has a dry weight of approximately 0.3 pg [177], our model predicts a time of approximately 8 hours for a cell with MCPs to metabolize enough biomass through the Pdu MCP metabolon to accumulate the mass of one daughter cell. This figure is in good agreement with experimentally measured doubling times for the growth of *Salmonella enterica* on 55 mM 1,2-PD of approximately 5-10 hours [5].

Another putative MCP function is the prevention of aldehyde loss into the growth medium. We quantify this phenomenon in our model as the net flux of propionaldehyde across the cell membrane into the extracellular space (so-called "propionaldehyde leakage"). This leakage is somewhat lower with MCPs than without, but flux enhancement appears to be the primary MCP function when the external 1,2-PD concentration is 55 mM. These two effects can be quantified as the relative propionaldehyde leakage (defined as the quotient of leakage without MCPs and leakage with MCPs) and PduP/Q flux enhancement (defined as the quotient of flux through PduP/Q with MCPs and flux without MCPs). These quantities, and the concentrations of propionaldehyde in the cytosol and the MCP, are plotted for a range of external 1,2-PD concentrations in Figure 6.6. Interestingly, there exists a transition from primarily flux enhancement to primarily aldehyde loss prevention as the external 1,2-PD concentration decreases.

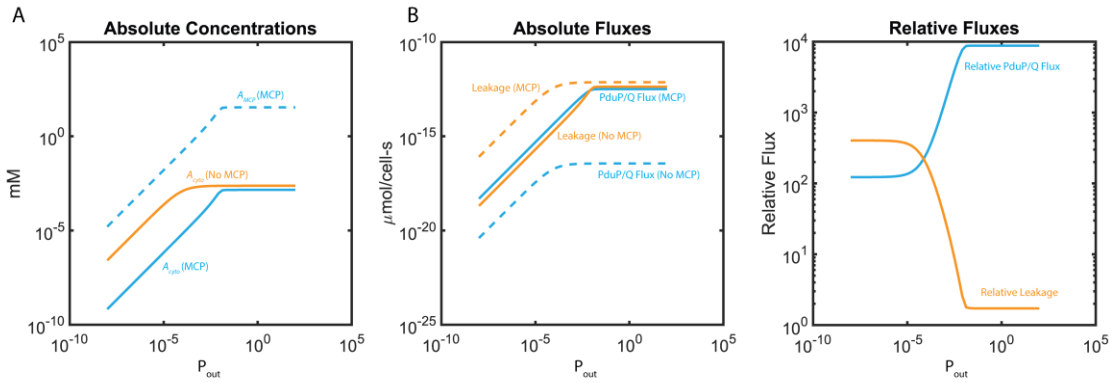


Figure 6.6: (A) Cytosolic propionaldehyde concentrations (A_{cyto}) with and without MCPs and MCP propionaldehyde concentrations (A_{MCP}) with MCPs; (B) steady-state fluxes through PduP/Q and propionaldehyde leakage across the cell membrane with and without MCPs; and (C) relative flux through PduP/Q ($flux_{MCP}/flux_{NoMCP}$) in blue and relative propionaldehyde leakage ($leakage_{NoMCP}/leakage_{MCP}$) in orange as a function of external 1,2-PD concentration.

Our model suggests that there are two distinct regimes of MCP function as external 1,2-PD concentration varies. At high external 1,2-PD concentrations, the cytosolic propionaldehyde concentration is comparable with and without MCPs, but the kinetically relevant propionaldehyde concentration in the vicinity of the PduP/Q enzymes is much higher, and hence flux through the MCP is much higher. On the other hand, at low external 1,2-PD concentrations, the cytosolic propionaldehyde concentration is lower with MCPs, but the kinetically relevant propionaldehyde concentration in the vicinity of the PduP/Q enzymes with MCPs is closer to that observed without MCPs. In the case of low external 1,2-PD concentration, therefore, the difference in propionaldehyde leakage into the extracellular space is large, but the difference in flux through PduP/Q is smaller than at high external 1,2-PD concentrations. It should also be noted that the flux through PduP/Q remains higher at all external 1,2-PD concentrations when the enzymes are localized to the MCP.

Active transport of 1,2-PD is dispensable at high external 1,2-PD concentrations, but not at low external 1,2-PD concentrations- We next set out to determine for what range of external 1,2-PD concentrations the PduP/Q enzymes could be saturated in the MCP system. We find that PduP/Q can be saturated for a range of external *P* concentrations by modulating the nonspecific MCP membrane permeability, k_c; the cell membrane permeability to 1,2-PD, k_mP; and the cell membrane permeability to propionaldehyde, k_mA; but there exists a lower limit of external 1,2-PD concentrations below which PduP/Q cannot be saturated by passive mechanisms (Figure 6.7).

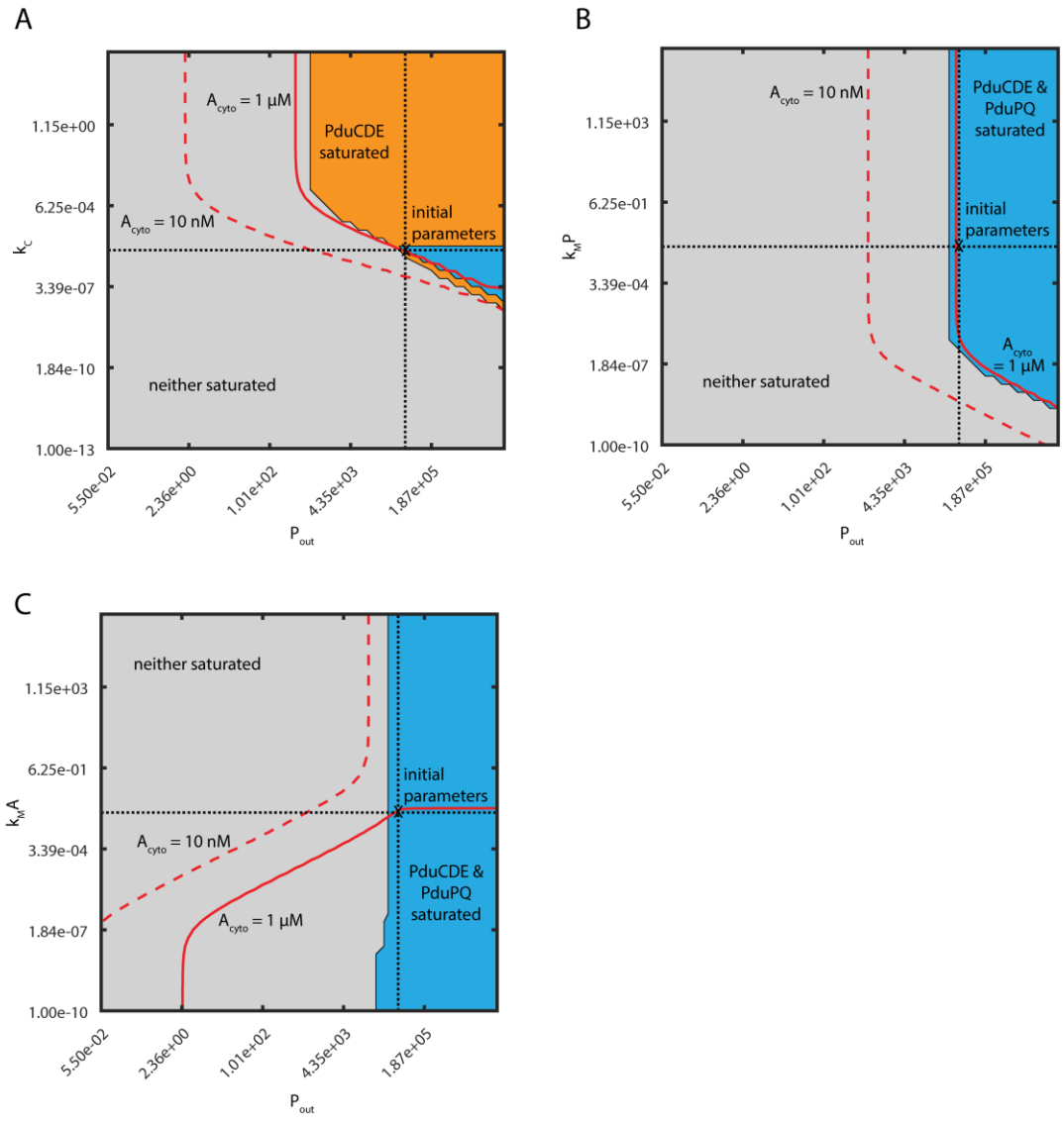


Figure 6.7: Saturation phase spaces of PduCDE and PduP/Q (A) with respect to P_{out} and k_c, (B) with respect to P_{out} and k_mP, and (C) with respect to P_{out} and k_mA. Regions of saturation (concentration of substrate > K_M of the appropriate enzyme) are plotted in blue when PduCDE and PduP/Q are saturated, orange when only PduCDE is saturated, and in grey when neither enzyme is saturated. Red solid lines indicate when

A_{cyto} is 1 μM; red dashed lines indicate when A_{cyto} is 10 nM. Black dashed lines indicate the initial parameter values used in the model of the Pdu MCP.

The *pduF* ORF of the *S. enterica* Pdu operon is a putative membrane protein, and is speculated to encode a 1,2-PD transporter. We therefore sought to explore the possible role of active 1,2-PD transport across the cell membrane in the saturation of the PduP/Q enzymes. We find that active transport of 1,2-PD is dispensable at high external 1,2-PD concentrations (*e.g.* 55 mM), but not at lower external *P* concentrations (*e.g.* 0.5 mM). When the external 1,2-PD concentration is 55 mM, the nonspecific MCP membrane permeability, k_c, can be adjusted to saturate both PduCDE and PduP/Q for any value of the rate of active 1,2-PD transport across the cell membrane, j_c (Figure 6.8). The same is true of the cell membrane permeability to 1,2-PD, k_mP, when the external 1,2-PD concentration is 55 mM. In contrast, when the external 1,2-PD concentration is 0.5 mM, there exists a wide range of values of the rate of active 1,2-PD transport across the cell membrane, j_c, for which PduP/Q cannot be saturated solely by adjusting the passive diffusion parameters k_c, k_mP, and k_mA (Figure 6.9). At this lower external concentration of 1,2-PD, therefore, active transport of 1,2-PD across the cell membrane may play an important role. Indeed, with high rates of active transport, PduP/Q can be saturated for an extremely wide range of external 1,2-PD concentrations (Figure 6.10).

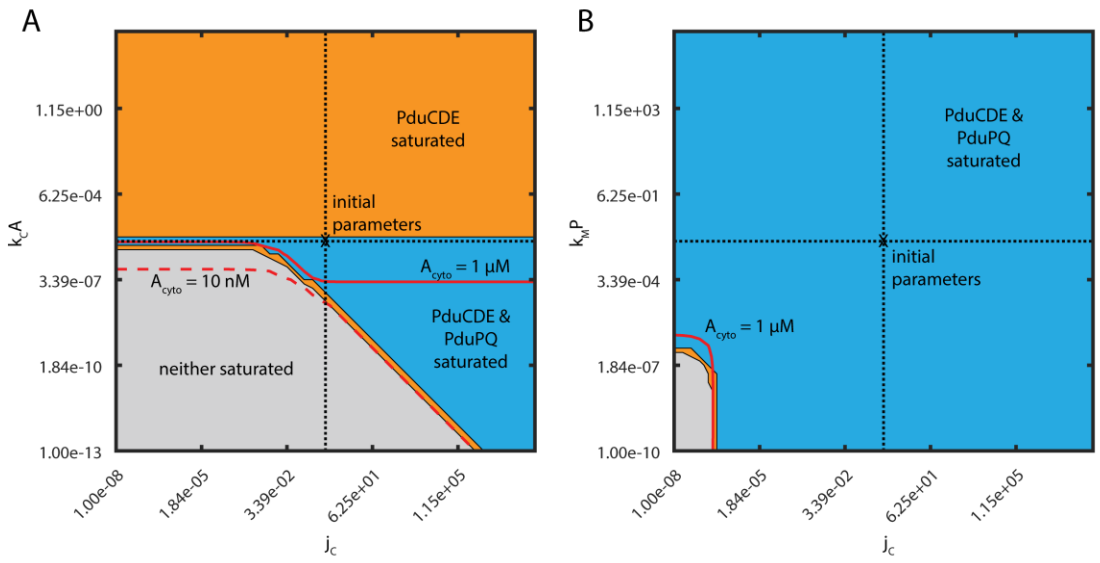


Figure 6.8: Saturation phase spaces of PduCDE and PduP/Q with respect to (A) j_c and k_c and (B) with respect to j_c and k_mP when P_{out} is 55 mM. Regions of saturation (concentration of substrate > K_M of the appropriate enzyme) are plotted in blue when PduCDE and PduP/Q are saturated, orange when only PduCDE is saturated, and in grey when neither enzyme is saturated. Red solid lines indicate when A_{cyto} is 1 μM; red dashed lines indicate when A_{cyto} is 10 nM. Black dashed lines indicate the initial parameter values used in the model of the Pdu MCP.

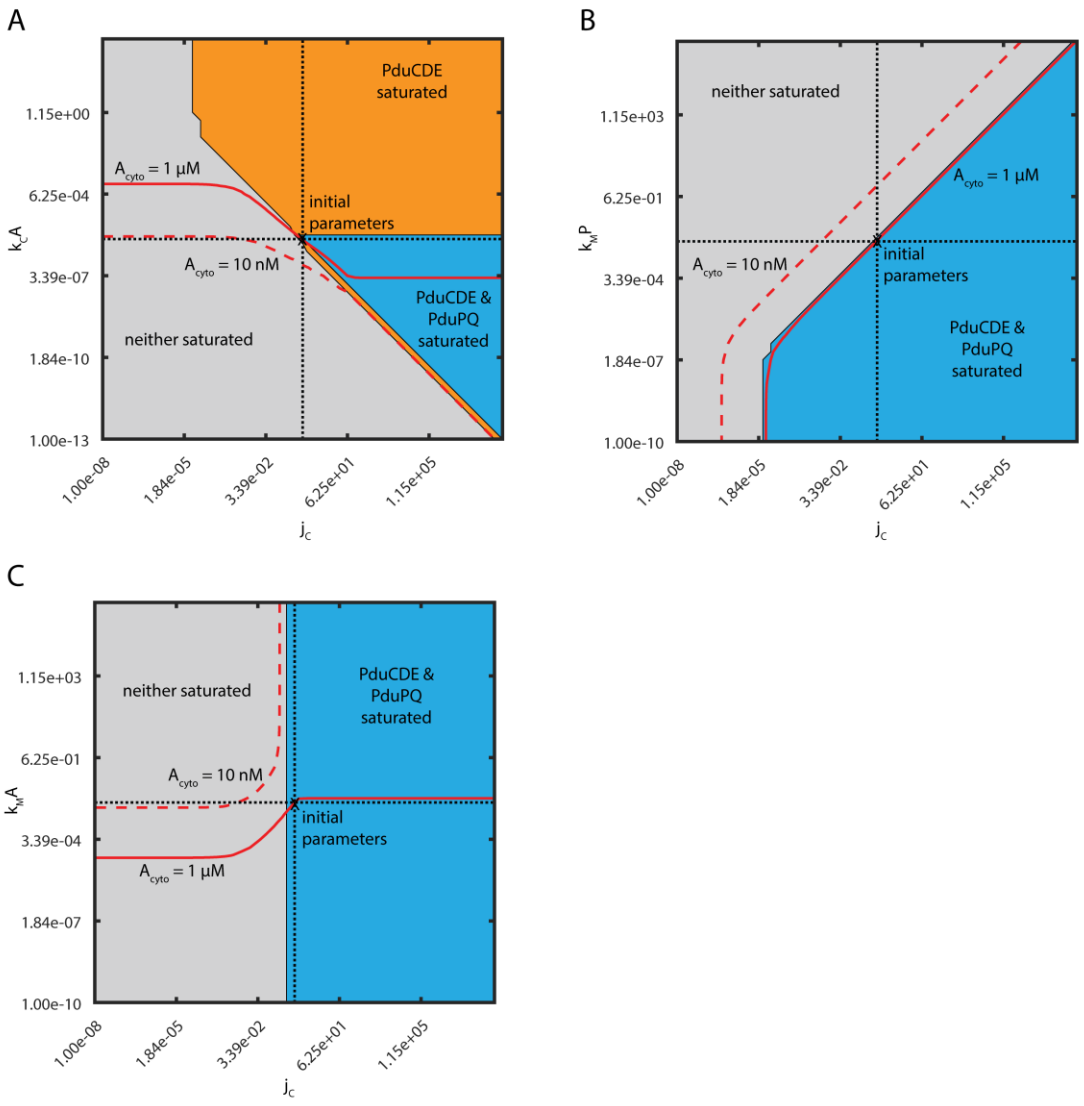


Figure 6.9: Saturation phase spaces of PduCDE and PduP/Q with respect to (A) j_c and k_c, (B) with respect to j_c and k_mP, and (C) with respect to j_c and k_mA when P_{out} is 0.5 mM. Regions of saturation (concentration of substrate > K_M of the appropriate enzyme) are plotted in blue when both enzymes are saturate, orange when only PdyCDE is saturated, and in grey when neither enzyme is saturated. Red solid lines indicate when A_{cyto} is 1 µM; red dashed lines indicate when A_{cyto} is 10 nM. Black dashed lines indicate the initial parameter values used in the model of the Pdu MCP.

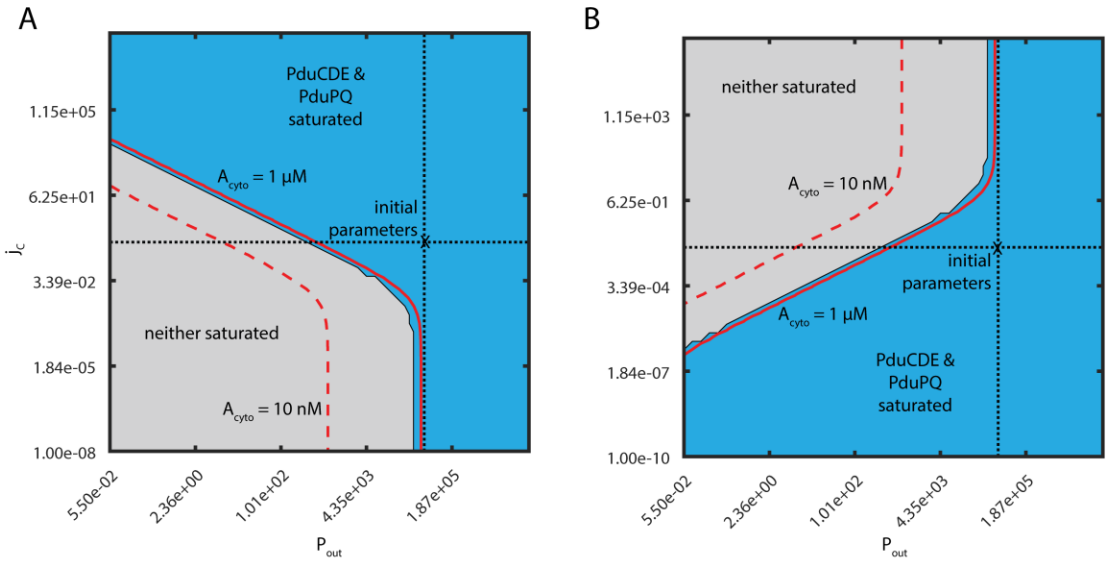


Figure 6.10: Saturation phase spaces of PduCDE and PduP/Q with respect to (A) P_{out} and j_c, and (B) with respect to P_{out} and k_mP. Regions of saturation (concentration of substrate > K_M of the appropriate enzyme) are plotted in blue when both enzymes are saturate, orange when only PdyCDE is saturated, and in grey when neither enzyme is saturated. Red solid lines indicate when A_{cyto} is 1 μM; red dashed lines indicate when A_{cyto} is 10 nM. Black dashed lines indicate the initial parameter values used in the model of the Pdu MCP.

Selective MCP permeability is not absolutely required to saturate PduP/Q- Experimental results indicate that the protein membrane surrounding the Pdu MCP exhibits selective permeability [46]. We therefore sought to explore under what conditions such selective permeability was advantageous for MCP function. We first consider the simple case of nonspecific permeability.

We find that an optimal non-selective MCP shell permeability exists with respect to the kinetically relevant propionaldehyde concentration in the vicinity of the PduP/Q enzymes (Figure 6.11A). This optimum value of k_c reflects a tradeoff between 1,2-PD entry to the MCP and trapping of propionaldehyde within the MCP. The non-selective permeability must be high enough for adequate entry of the PduCDE substrate 1,2-PD, but low enough to contribute to the accumulation of the PduP/Q substrate propionaldehyde within the organelle. This can be seen clearly when the permeabilities of the MCP membrane to 1,2-PD and propionaldehyde are varied separately (Figure 6.11B, C). Higher 1,2-PD permeability k_cP is unambiguously advantageous for increasing propionaldehyde concentration inside the MCP, as is lower permeability to propionaldehyde k_cA.

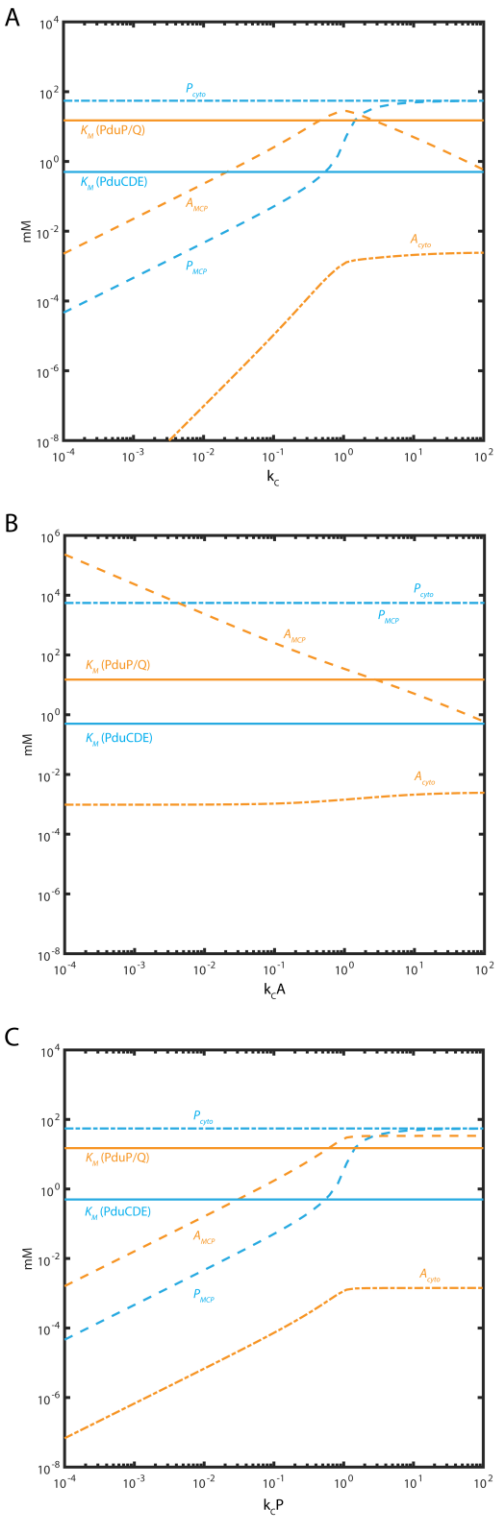


Figure 6.11: Mean concentrations of 1,2-PD and propionaldehyde in the MCP (P_{MCP}; A_{MCP}) and cytosol (P_{cyto}; A_{cyto}) as a function of (A) k_c when $k_cP = k_cA = k_c$; (B) k_cA; and (C) k_cP. K_M of PduCDE and PduP/Q are shown as solid lines.

The presence of an optimal permeability persists even for k_cP fixed at a value somewhat higher or lower than k_cA (Figure 6.12). Moreover, decreasing k_c entails a tradeoff between the two aspects of Pdu MCP function (Figure 6.13A). At low k_c, aldehyde leakage is prevented, but flux is low; near the optimal k_c for flux enhancement, leakage is comparable with and without MCPs.

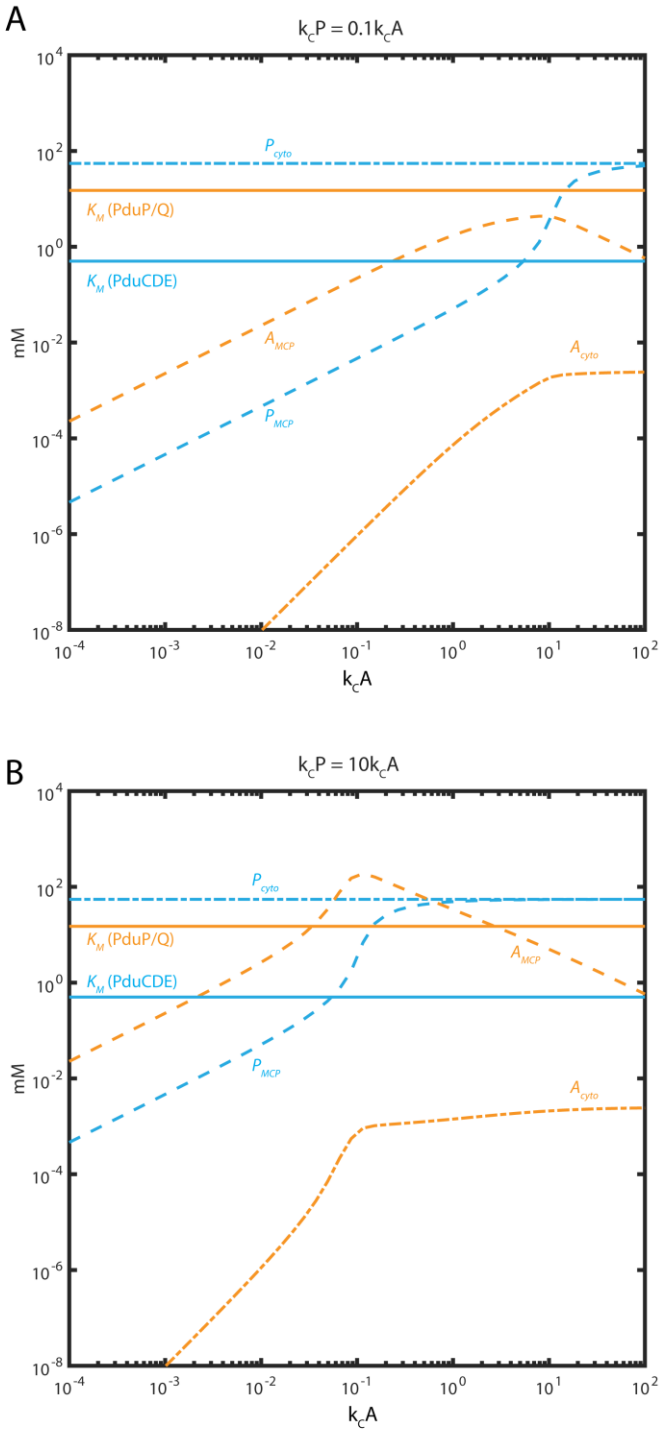


Figure 6.12: Mean concentrations of 1,2-PD and propionaldehyde in the MCP (P_{MCP}; A_{MCP}) and cytosol (P_{cyto}; A_{cyto}) as a function of (A) k_cA when $k_cP = 0.1$ x k_cA and (B) k_cA when $k_cP = 10$ x k_cA. K_M of PduCDE and PduP/Q are shown as solid lines.

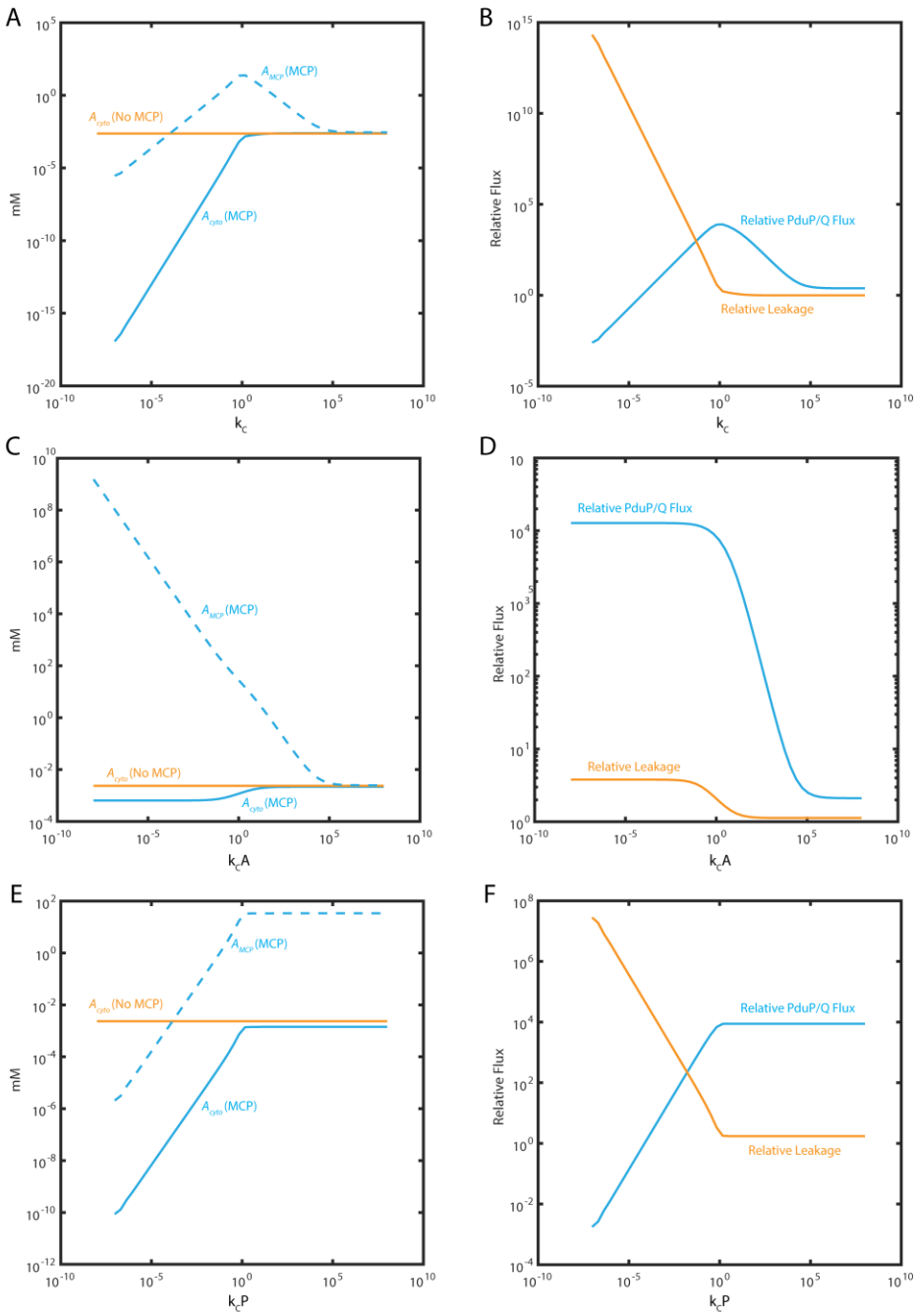


Figure 6.13: (A,C,E) Cytosolic aldehyde concentration (A_{cyto}) with and without MCPs and MCP aldehyde concentration (A_{MCP}) with MCPs; (B,D,F) relative carbon flux through PduP/Q ($flux_{MCP}/flux_{NoMCP}$) in blue and relative aldehyde leakage rate ($leakage_{NoMCP}/leakage_{MCP}$) in orange as a function of (A,B) $k_c = k_cA = k_cP$; (C,D) k_cA; and (E,F) k_cP.

While selective permeability is not absolutely required for PduP/Q saturation, it is required for PduP/Q saturation in some cases (Figure 6.14). It is interesting to note that in all cases, lower relative aldehyde permeability is required (either in the form of lower propionaldehyde permeability or higher permeability to 1,2-PD). As observed above,

decreasing MCP permeability to propionaldehyde is unambiguously beneficial for function, while decreasing MCP permeability to 1,2-PD is always detrimental to relative flux (Figure 6.13).

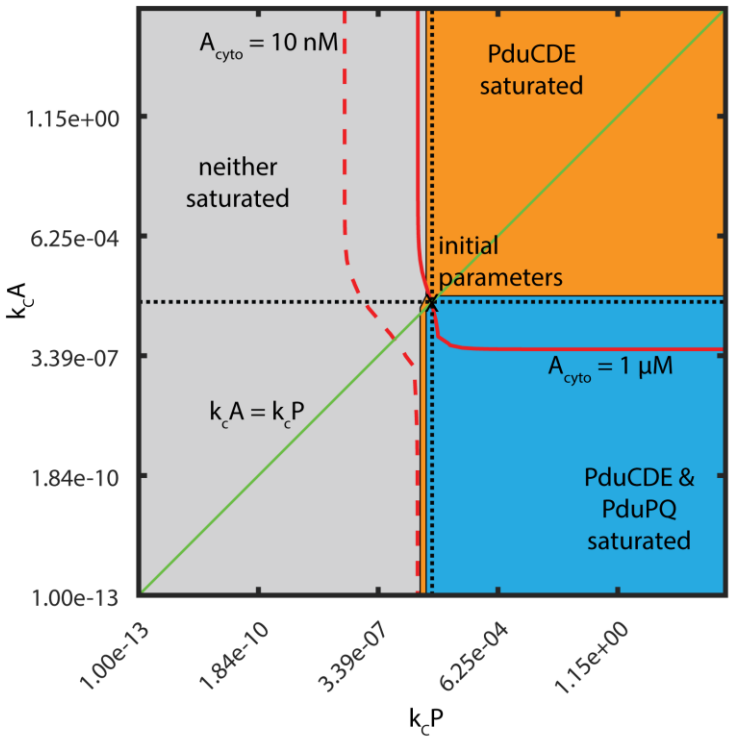


Figure 6.14: Saturation phase space of PduCDE and PduP/Q with respect to k_cA and k_cP. Regions of saturation (concentration of substrate > K_M of the appropriate enzyme) are plotted in blue when PduCDE and PduP/Q are saturated, orange when only PduCDE is saturated, and in grey when neither enzyme is saturated. Red solid lines indicate when A_{cyto} is 1 μM; red dashed lines indicate when A_{cyto} is 10 nM. Black dashed lines indicate the initial parameter values used in the model of the Pdu MCP. Green line indicates when $k_cA = k_cP$.

6.4 Discussion

Metabolite concentration in Pdu MCPs is essentially constant- By comparison of the numerical finite difference solution to our model with an analytical solution assuming constant metabolite concentrations in the Pdu MCP, we find that the metabolite concentrations can be assumed to be approximately constant inside the MCP shell. This is convenient for our calculations, and suggests that future, more complex models of Pdu MCP function may also be analytically tractable. Furthermore, this allows us to neglect the detailed organization of the enzymes inside the MCP, which is as-yet unknown.

Model predictions are in agreement with experimental findings- Our model is suitable to address batch-wise experimental results because a typical experiment measuring the growth of *Salmonella* on 1,2-PD takes place over tens of hours, while our model has characteristic timescales on the order of seconds, at maximum. This allows us to treat a

Salmonella growth experiment as being at a pseudo-steady state relative to our model, and accounts for the good congruence between our model predictions and the experimental measurements of steady-state *Salmonella* growth rates on 1,2-PD. This agreement is observed despite the model parameters being estimated *a priori* or from direct observations of the Pdu MCP, independent of experimental growth results.
It is unknown what external concentration of propionaldehyde is encountered in the host gut by invading *Salmonella*, but the concentration is likely to be low due to the toxicity of aldehyde species to eukarya. With respect to *ex vivo* experiments, the propionaldehyde concentration in the growth medium is observed to increase in batch-wise growth experiments, up to concentrations around 10 mM. The calculations above were performed assuming negligible external propionaldehyde, but we tested our model at a range of external propionaldehyde concentrations (Figure 6.15) and found that the qualitative behavior of the system did not change with increasing external propionaldehyde concentration.

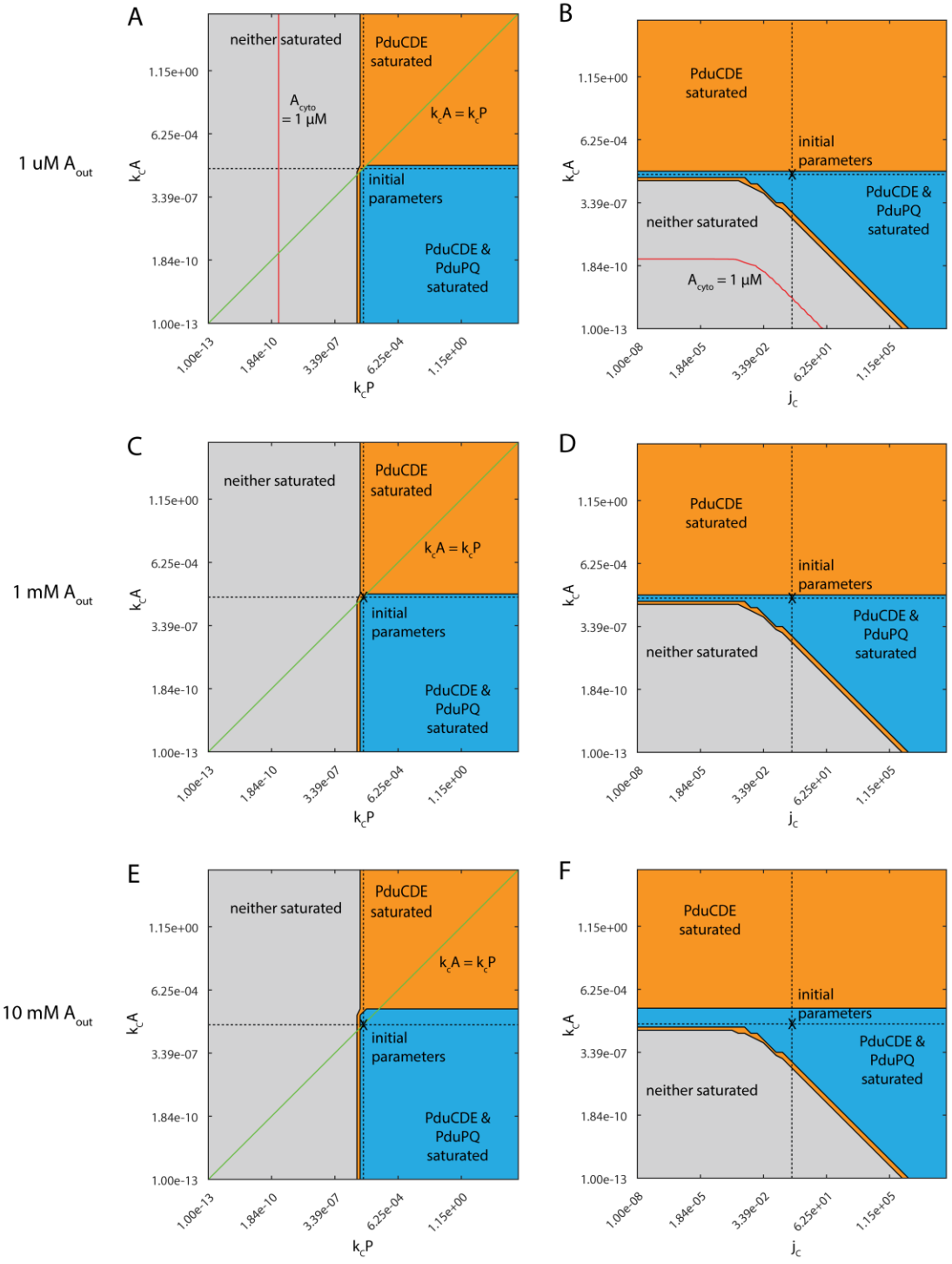


Figure 6.15: (A, C, E) Saturation phase space of PduCDE and PduP/Q with respect to k_cA and k_cP and (B, D, F) saturation phase space of PduCDE and PduP/Q with respect to k_cA and j_c for external propionaldehyde concentrations of (A, B) 1µM, (C, D) 1 mM, and (E, F) 10 mM. Regions of saturation (concentration of substrate > K_M of the appropriate enzyme) are plotted in blue when both enzymes are saturated, orange when only PduCDE is saturated, and in grey when neither enzyme is saturated. Red solid lines indicate when

A_{cyto} is 1 μM; red dashed lines indicate when A_{cyto} is 10 nM. Black dashed lines indicate the initial parameter values used in the model of the Pdu MCP. Green line in (A, C, E) indicates when $k_cA = k_cP$.

Experimental investigations of Pdu MCP function have consistently demonstrated two key phenotypes for strains that express the Pdu enzymes but fail to form MCPs: a slower growth rate and increased propionaldehyde concentration in the media [5]. Slower growth could be attributable to two phenomena: "passive" growth retardation due to lower carbon flux through the Pdu metabolon, and "active" growth retardation due to the toxic effects of propionaldehyde in the cytosol. This second form of growth defect is closely linked to the accompanying observation that strains lacking MCPs have an increased rate of propionaldehyde leakage into the extracellular space. These two phenomena cannot be distinguished by an *in vivo* experiment measuring the growth of cells and the bulk concentrations of the various metabolites. Our simple model supports the idea that both of these phenotypes contribute to changes in growth rate: we observe that cells with MCPs have both higher flux through PduP/Q and lower cytosolic concentrations of propionaldehyde than cells without MCPs, and that cells with MCPs exhibit lower aldehyde flux into the growth medium than cells without MCPs. Additionally, experiments have observed similar 1,2-PD depletion from the growth medium with and without MCPs, a phenotype that is evident in the comparable PduCDE fluxes predicted by our model with and without MCPs [5]. Together these observations indicate that our model captures the important principles of Pdu MCP function. Indeed, we propose that while toxicity reduction likely plays a role in MCP function, flux enhancement is a crucial, and underappreciated, consequence of encapsulation.

It is particularly important to note that many behaviors in our model depend on external 1,2-PD concentration; care must therefore be taken in applying results from *ex vivo* experiments conducted at high 1,2-PD concentration to Pdu MCP function *in vivo*, and in interpreting the results of batch-wise culture experiments in which the external 1,2-PD concentration changes during the experiment.

MCP function depends on extrinsic factors- We find that MCP function, as quantified by the relative PduP/Q flux and the relative propionaldehyde flux into the extracellular space, changes in response to factors extrinsic to the cell, such as external 1,2-PD concentration. Indeed, these may change during an experiment: for example, a low propionaldehyde leakage difference with and without MCPs may be observed at high external 1,2-PD concentrations, but this discrepancy may increase over the course of a batch experiment during which 1,2-PD is depleted from the growth medium. The necessity of mechanisms such as active transport also depends on extrinsic factors; our model suggests that active transport of 1,2-PD across the cell membrane is dispensable at high external 1,2-PD concentrations, but not as the external 1,2-PD concentration decreases. This in turn suggests that experiments to determine the function of the *pduF* gene product, for instance, must be undertaken with great care paid to the culture conditions such that active transport is made relevant to growth.

Selective MCP membrane permeability is not required, but is often advantageous- Selective permeability of the Pdu MCP shell to propionaldehyde as compared to other

metabolites was recently reported [46]. Our model suggests that, while the saturation of PduP/Q enzymes can be achieved without selective permeability in particular cases, it is true in general that selective permeability can enhance Pdu MCP function. Trapping of propionaldehyde in the MCP by decreasing the relative permeability of the MCP shell to propionaldehyde, or by increasing the relative permeability of the MCP shell to 1,2-PD, can serve to saturate the PduP/Q enzymes, all else being equal. This putative trapping mechanism is in congruence with the *in vitro* observation that small molecule efflux from a protein nanoreactor can be affected by the chemical character of the reactor pores [47]. Studies of substrate channeling in enzyme scaffolds also emphasize the importance of creating a high substrate concentration in the vicinity of downstream enzymes [90,92]; in the case of a microcompartment, however, a concentration gradient is developed by the shell diffusion barrier, rather than by a very high local concentration of enzymes. Interestingly, and contrary to previous studies, our results suggest that selective permeability, while often advantageous, is by no means required for significant flux enhancement (Fig. 2).

Outstanding questions for modeling and experiment- Our results have important implications for microbiological studies of MCP function and for the application of encapsulation to synthetic biology. Interestingly, our model indicates that the encapsulation of enzymes in a subcellular compartment can dramatically improve flux through the encapsulated pathway, in addition to reducing the concentrations of intermediates in the cytosol, simply by imposing a non-specific diffusion barrier. This reinforces the notion that encapsulation is a promising strategy to improve the yield and titer of heterologous enzymatic pathways that fail to function in a cytosolic context, although further investigation is required to inform the selection of appropriate pathways for encapsulation.

Many questions also remain to be addressed both computationally and experimentally with respect to native MCP function. First, in our model we neglect NAD^+/NADH cofactor recycling by setting the rate of PduP catalysis equal to that of PduQ. Future models could incorporate cofactor-dependent kinetics of PduP/Q, in addition to including the effect of B_{12} recycling within the MCP shell on the kinetics of PduCDE. Also lacking at this time are detailed simulations of the structure and dynamics of the Pdu membrane pores, particularly with respect to the diffusion of species of varying size and charge.

Studies of this kind would be complemented by direct experimental measurements of the permeability of the MCP membrane to various species, since the values of these permeabilities must at present be estimated. Also of interest are experimental investigations as to the possibility of active 1,2-PD transport across the cell membrane. Our results suggest that such active transport would be advantageous at low external 1,2-PD concentrations but dispensable at high external 1,2-PD concentrations, complicating experiments. The putative membrane-bound Pdu gene product *pduF* is of primary interest in this regard. Lastly, experimental observations of the absolute 1,2-PD and propionaldehyde concentrations encountered by invading *Salmonella* in the host gut and in the environment at large would be valuable in constraining future modeling efforts to pathogenically relevant metabolite concentrations, and in comparing the conditions faced by free-living and host-associated pathogens.

Chapter 7: Conclusion

Since their discovery, bacterial microcompartments and related organelles have puzzled micobiologists and challenged preconceptions regarding bacterial cellular organization. Above, I describe our recent progress in adapting these naturally occurring structures for engineering and synthetic biology applications. We have succeeded in controlling organelle formation and cargo loading, as well as in using an MS2 bacteriophage model system to begin to understand the principles underlying organelle assembly and small molecule transport.

Significant challenges remain before bacterial microcompartments can be flexibly employed as a strategy to enhance the throughput of heterologous synthetic pathways in bacteria. Crucially, we must develop design rules to control small molecule transport in and out of the nanoreactor (a focus of my colleague Mary Slininger's research) as well as heuristics to guide the selection of appropriate pathways for encapsulation (an ongoing thread of our research in collaboration with Niall Mangan).

The future of bacterial organelles for synthetic biology is bright, but progress will require the continued union of microbiology, biochemistry, and engineering research efforts.

References

1. Bobik TA, Havemann GD, Busch RJ, Williams DS, Aldrich HC. The Propanediol Utilization (pdu) Operon ofSalmonella enterica Serovar Typhimurium LT2 Includes Genes Necessary for Formation of Polyhedral Organelles Involved in Coenzyme B12-Dependent 1, 2-Propanediol Degradation. J Bacteriol. 1999;181: 5967–5975.

2. Havemann GD, Bobik TA. Protein content of polyhedral organelles involved in coenzyme B12-dependent degradation of 1, 2-propanediol in Salmonella enterica serovar Typhimurium LT2. J Bacteriol. 2003;185: 5086–5095.

3. Kofoid E, Rappleye C, Stojiljkovic I, Roth J. The 17-Gene Ethanolamine (eut) Operon ofSalmonella typhimurium Encodes Five Homologues of Carboxysome Shell Proteins. J Bacteriol. 1999;181: 5317–5329.

4. Kerfeld CA, Heinhorst S, Cannon GC. Bacterial Microcompartments. Annu Rev Microbiol. 2010;64: 391–408. doi:10.1146/annurev.micro.112408.134211

5. Sampson EM, Bobik TA. Microcompartments for B12-Dependent 1,2-Propanediol Degradation Provide Protection from DNA and Cellular Damage by a Reactive Metabolic Intermediate. J Bacteriol. 2008;190: 2966–2971. doi:10.1128/JB.01925-07

6. Huseby DL, Roth JR. Evidence that a metabolic microcompartment contains and recycles private cofactor pools. J Bacteriol. 2013; doi:10.1128/JB.02179-12

7. Delebecque CJ, Lindner AB, Silver PA, Aldaye FA. Organization of Intracellular Reactions with Rationally Designed RNA Assemblies. Science. 2011;333: 470–474. doi:10.1126/science.1206938

8. Dueber JE, Wu GC, Malmirchegini GR, Moon TS, Petzold CJ, Ullal AV, et al. Synthetic protein scaffolds provide modular control over metabolic flux. Nat Biotechnol. 2009;27: 753–759.

9. Conrado RJ, Wu GC, Boock JT, Xu H, Chen SY, Lebar T, et al. DNA-guided assembly of biosynthetic pathways promotes improved catalytic efficiency. Nucleic Acids Res. 2012;40: 1879–1889. doi:10.1093/nar/gkr888

10. Axen SD, Erbilgin O, Kerfeld CA. A Taxonomy of Bacterial Microcompartment Loci Constructed by a Novel Scoring Method. Tanaka MM, editor. PLoS Comput Biol. 2014;10: e1003898. doi:10.1371/journal.pcbi.1003898

11. Erbilgin O, McDonald KL, Kerfeld CA. Characterization of a Planctomycetal Organelle: a Novel Bacterial Microcompartment for the Aerobic Degradation of Plant Saccharides. Appl Environ Microbiol. 2014;80: 2193–2205. doi:10.1128/AEM.03887-13

12. Srikumar S, Fuchs TM. Ethanolamine Utilization Contributes to Proliferation of Salmonella enterica Serovar Typhimurium in Food and in Nematodes. Appl Environ Microbiol. 2011;77: 281–290. doi:10.1128/AEM.01403-10

13. Thiennimitr P, Winter SE, Winter MG, Xavier MN, Tolstikov V, Huseby DL, et al. Intestinal inflammation allows Salmonella to use ethanolamine to compete with the microbiota. Proc Natl Acad Sci. 2011;108: 17480–17485. doi:10.1073/pnas.1107857108

14. Kerfeld CA, Erbilgin O. Bacterial microcompartments and the modular construction of microbial metabolism. Trends Microbiol. 2014; doi:10.1016/j.tim.2014.10.003

15. Fan C, Cheng S, Liu Y, Escobar CM, Crowley CS, Jefferson RE, et al. Short N-terminal sequences package proteins into bacterial microcompartments. Proc Natl Acad Sci. 2010;107: 7509–7514.

16. Fan C, Cheng S, Sinha S, Bobik TA. Interactions between the termini of lumen enzymes and shell proteins mediate enzyme encapsulation into bacterial microcompartments. Proc Natl Acad Sci. 2012;109: 14995–15000.

17. Bertin Y, Girardeau JP, Chaucheyras-Durand F, Lyan B, Pujos-Guillot E, Harel J, et al. Enterohaemorrhagic Escherichia coli gains a competitive advantage by using ethanolamine as a nitrogen source in the bovine intestinal content. Environ Microbiol. 2011;13: 365–377. doi:10.1111/j.1462-2920.2010.02334.x

18. Rivera-Chávez F, Bäumler AJ. The Pyromaniac Inside You: Salmonella Metabolism in the Host Gut. Annu Rev Microbiol. 2015;69: 31–48. doi:10.1146/annurev-micro-091014-104108

19. Roth JR, Lawrence JG, Bobik TA. Cobalamin (coenzyme B12): synthesis and biological significance. Annu Rev Microbiol. 1996;50: 137–181. doi:10.1146/annurev.micro.50.1.137

20. Price-Carter M, Tingey J, Bobik TA, Roth JR. The Alternative Electron Acceptor Tetrathionate Supports B12-Dependent Anaerobic Growth of Salmonella enterica Serovar Typhimurium on Ethanolamine or 1,2-Propanediol. J Bacteriol. 2001;183: 2463–2475. doi:10.1128/JB.183.8.2463-2475.2001

21. Stecher B, Hardt W-D. The role of microbiota in infectious disease. Trends Microbiol. 2008;16: 107–114. doi:10.1016/j.tim.2007.12.008

22. Keeney KM, Yurist-Doutsch S, Arrieta M-C, Finlay BB. Effects of Antibiotics on Human Microbiota and Subsequent Disease. Annu Rev Microbiol. 2014;68: 217–235. doi:10.1146/annurev-micro-091313-103456

23. Endt K, Stecher B, Chaffron S, Slack E, Tchitchek N, Benecke A, et al. The Microbiota Mediates Pathogen Clearance from the Gut Lumen after Non-Typhoidal

Salmonella Diarrhea. PLoS Pathog. 2010;6: e1001097. doi:10.1371/journal.ppat.1001097

24. Bobik TA, Ailion M, Roth JR. A single regulatory gene integrates control of vitamin B12 synthesis and propanediol degradation. J Bacteriol. 1992;174: 2253–2266.

25. Jeter RM, Olivera BM, Roth JR. Salmonella typhimurium synthesizes cobalamin (vitamin B12) de novo under anaerobic growth conditions. J Bacteriol. 1984;159: 206–213.

26. Winter SE, Thiennimitr P, Winter MG, Butler BP, Huseby DL, Crawford RW, et al. Gut inflammation provides a respiratory electron acceptor for Salmonella. Nature. 2010;467: 426–429. doi:10.1038/nature09415

27. Galán JE, Collmer A. Type III Secretion Machines: Bacterial Devices for Protein Delivery into Host Cells. Science. 1999;284: 1322–1328. doi:10.1126/science.284.5418.1322

28. Galán JE. SALMONELLA INTERACTIONS WITH HOST CELLS: Type III Secretion at Work. Annu Rev Cell Dev Biol. 2001;17: 53–86. doi:10.1146/annurev.cellbio.17.1.53

29. Ackermann M, Stecher B, Freed NE, Songhet P, Hardt W-D, Doebeli M. Self-destructive cooperation mediated by phenotypic noise. Nature. 2008;454: 987–990. doi:10.1038/nature07067

30. Sturm A, Heinemann M, Arnoldini M, Benecke A, Ackermann M, Benz M, et al. The Cost of Virulence: Retarded Growth of Salmonella Typhimurium Cells Expressing Type III Secretion System 1. Ausubel FM, editor. PLoS Pathog. 2011;7: e1002143. doi:10.1371/journal.ppat.1002143

31. Stecher B, Robbiani R, Walker AW, Westendorf AM, Barthel M, Kremer M, et al. Salmonella enterica Serovar Typhimurium Exploits Inflammation to Compete with the Intestinal Microbiota. PLoS Biol. 2007;5: e244. doi:10.1371/journal.pbio.0050244

32. Lupp C, Robertson ML, Wickham ME, Sekirov I, Champion OL, Gaynor EC, et al. Host-Mediated Inflammation Disrupts the Intestinal Microbiota and Promotes the Overgrowth of Enterobacteriaceae. Cell Host Microbe. 2007;2: 119–129. doi:10.1016/j.chom.2007.06.010

33. Nakayama S -i., Watanabe H. Mechanism of hilA Repression by 1,2-Propanediol Consists of Two Distinct Pathways, One Dependent on and the Other Independent of Catabolic Production of Propionate, in Salmonella enterica Serovar Typhimurium. J Bacteriol. 2006;188: 3121–3125. doi:10.1128/JB.188.8.3121-3125.2006

34. Hung C-C, Garner CD, Slauch JM, Dwyer ZW, Lawhon SD, Frye JG, et al. The Intestinal Fatty Acid Propionate Inhibits Salmonella Invasion through the Post-

translational Control of HilD. Mol Microbiol. 2013;87: 1045–1060. doi:10.1111/mmi.12149

35. Skaar EP. The Battle for Iron between Bacterial Pathogens and Their Vertebrate Hosts. PLoS Pathog. 2010;6: e1000949. doi:10.1371/journal.ppat.1000949

36. Hood MI, Skaar EP. Nutritional immunity: transition metals at the pathogen-host interface. Nat Rev Microbiol. 2012;10: 525–537. doi:10.1038/nrmicro2836

37. Gopinath S, Carden S, Monack D. Shedding light on Salmonella carriers. Trends Microbiol. 2012;20: 320–327. doi:10.1016/j.tim.2012.04.004

38. Pang A, Frank S, Brown I, Warren MJ, Pickersgill RW. Structural insights into higher-order assembly and function of the bacterial microcompartment protein PduA. J Biol Chem. 2014; doi:10.1074/jbc.M114.569285

39. Crowley CS, Bobik TA. Structure of the PduU Shell Protein from the Pdu Microcompartment of Salmonella. 2008;

40. Thompson MC, Crowley CS, Kopstein J, Bobik TA, Yeates TO. Structure of a bacterial microcompartment shell protein bound to a cobalamin cofactor. Acta Crystallogr Sect F Struct Biol Commun. 2014;70. doi:10.1107/S2053230X1402158X

41. Thompson MC, Cascio D, Leibly DJ, Yeates TO. An allosteric model for control of pore opening by substrate binding in the eutl microcompartment shell protein. Protein Sci. 2015; n/a–n/a. doi:10.1002/pro.2672

42. Wheatley NM, Gidaniyan SD, Liu Y, Cascio D, Yeates TO. Bacterial microcompartment shells of diverse functional types possess pentameric vertex proteins. Protein Sci. 2013;22: 660–665. doi:10.1002/pro.2246

43. Heldt D, Frank S, Seyedarabi A, Ladikis D, Parsons JB, Warren MJ, et al. Structure of a trimeric bacterial microcompartment shell protein, EtuB, associated with ethanol utilization in *Clostridium kluyveri*. Biochem J. 2009;423: 199–207. doi:10.1042/BJ20090780

44. Tocheva EI, Matson EG, Cheng SN, Chen W, Leadbetter JR, Jensen GJ. Structure and expression of propanediol utilization microcompartments in Acetonema longum. J Bacteriol. 2014; doi:10.1128/JB.00049-14

45. Sinha S, Cheng S, Sung YW, McNamara DE, Sawaya MR, Yeates TO, et al. Alanine Scanning Mutagenesis Identifies an Asparagine-Arginine-Lysine Triad Essential to Assembly of the Shell of the Pdu Microcompartment. J Mol Biol. 2014; doi:10.1016/j.jmb.2014.04.012

46. Chowdhury C, Chun S, Pang A, Sawaya MR, Sinha S, Yeates TO, et al. Selective molecular transport through the protein shell of a bacterial microcompartment organelle. Proc Natl Acad Sci. 2015; 201423672. doi:10.1073/pnas.1423672112

47. Glasgow JE, Asensio MA, Jakobson CM, Francis MB, Tullman-Ercek D. Influence of Electrostatics on Small Molecule Flux through a Protein Nanoreactor. ACS Synth Biol. 2015; doi:10.1021/acssynbio.5b00037

48. Gonzalez-Esquer CR, Shubitowski TB, Kerfeld CA. Streamlined Construction of the Cyanobacterial CO2-Fixing Organelle via Protein Domain Fusions for Use in Plant Synthetic Biology. Plant Cell. 2015; tpc.15.00329. doi:10.1105/tpc.15.00329

49. Lassila JK, Bernstein SL, Kinney JN, Axen SD, Kerfeld CA. Assembly of Robust Bacterial Microcompartment Shells Using Building Blocks from an Organelle of Unknown Function. J Mol Biol. 2014;426: 2217–2228. doi:10.1016/j.jmb.2014.02.025

50. Fan C, Bobik TA. The N-terminal region of the medium subunit (PduD) packages adenosylcobalamin-dependent diol dehydratase (PduCDE) into the Pdu microcompartment. J Bacteriol. 2011;193: 5623–5628.

51. Choudhary S, Quin MB, Sanders MA, Johnson ET, Schmidt-Dannert C. Engineered protein nano-compartments for targeted enzyme localization. PloS One. 2012;7: e33342.

52. Lawrence AD, Frank S, Newnham S, Lee MJ, Brown IR, Xue W-F, et al. Solution Structure of a Bacterial Microcompartment Targeting Peptide and Its Application in the Construction of an Ethanol Bioreactor. ACS Synth Biol. 2014; 140224092559007. doi:10.1021/sb4001118

53. Jakobson CM, Kim EY, Slininger MF, Chien A, Tullman-Ercek D. Localization of Proteins to the 1,2-Propanediol Utilization Microcompartment by Non-native Signal Sequences Is Mediated by a Common Hydrophobic Motif. J Biol Chem. 2015; jbc.M115.651919. doi:10.1074/jbc.M115.651919

54. Aussignargues C, Paasch BC, Gonzalez-Esquer R, Erbilgin O, Kerfeld CA. Bacterial microcompartment assembly: The key role of encapsulation peptides. Commun Integr Biol. 2015;8: e1039755. doi:10.1080/19420889.2015.1039755

55. Roof DM, Roth JR. Autogenous regulation of ethanolamine utilization by a transcriptional activator of the eut operon in Salmonella typhimurium. J Bacteriol. 1992;174: 6634–6643.

56. Ailion M, Bobik TA, Roth JR. Two global regulatory systems (Crp and Arc) control the cobalamin/propanediol regulon of Salmonella typhimurium. J Bacteriol. 1993;175: 7200–7208.

57. Parsons JB, Frank S, Bhella D, Liang M, Prentice MB, Mulvihill DP, et al. Synthesis of Empty Bacterial Microcompartments, Directed Organelle Protein Incorporation, and Evidence of Filament-Associated Organelle Movement. Mol Cell. 2010;38: 305–315. doi:10.1016/j.molcel.2010.04.008

58. Kim EY, Jakobson CM, Tullman-Ercek D. Engineering Transcriptional Regulation to Control Pdu Microcompartment Formation. Cirino PC, editor. PLoS ONE. 2014;9: e113814. doi:10.1371/journal.pone.0113814

59. Doyle L, Hallinan J, Bolduc J, Parmeggiani F, Baker D, Stoddard BL, et al. Rational design of α-helical tandem repeat proteins with closed architectures. Nature. 2015;528: 585–588. doi:10.1038/nature16191

60. Gonen S, DiMaio F, Gonen T, Baker D. Design of ordered two-dimensional arrays mediated by noncovalent protein-protein interfaces. Science. 2015;348: 1365–1368. doi:10.1126/science.aaa9897

61. Engler C, Gruetzner R, Kandzia R, Marillonnet S. Golden Gate Shuffling: A One-Pot DNA Shuffling Method Based on Type IIs Restriction Enzymes. PLoS ONE. 2009;4: e5553. doi:10.1371/journal.pone.0005553

62. Datta S, Costantino N, Court DL. A set of recombineering plasmids for gram-negative bacteria. Gene. 2006;379: 109–115. doi:10.1016/j.gene.2006.04.018

63. Chowdhury C, Sinha S, Chun S, Yeates TO, Bobik TA. Diverse Bacterial Microcompartment Organelles. Microbiol Mol Biol Rev. 2014;78: 438–468. doi:10.1128/MMBR.00009-14

64. Chessher A, Breitling R, Takano E. Bacterial Microcompartments: Biomaterials for Synthetic Biology-Based Compartmentalization Strategies. ACS Biomater Sci Eng. 2015;1: 345–351. doi:10.1021/acsbiomaterials.5b00059

65. Giessen TW, Silver PA. Encapsulation as a Strategy for the Design of Biological Compartmentalization. J Mol Biol. 2015; doi:10.1016/j.jmb.2015.09.009

66. Marx H, Pflügl S, Mattanovich D, Sauer M. Synthetic Biology Assisting Metabolic Pathway Engineering. In: Glieder A, Kubicek CP, Mattanovich D, Wiltschi B, Sauer M, editors. Synthetic Biology. Springer International Publishing; 2016. pp. 255–280. Available: http://link.springer.com/chapter/10.1007/978-3-319-22708-5_7

67. Sampson EM, Bobik TA. Microcompartments for B12-Dependent 1,2-Propanediol Degradation Provide Protection from DNA and Cellular Damage by a Reactive Metabolic Intermediate. J Bacteriol. 2008;190: 2966–2971. doi:10.1128/JB.01925-07

68. Kerfeld CA. Protein Structures Forming the Shell of Primitive Bacterial Organelles. Science. 2005;309: 936–938. doi:10.1126/science.1113397

69. Crowley CS, Cascio D, Sawaya MR, Kopstein JS, Bobik TA, Yeates TO. Structural insight into the mechanisms of transport across the Salmonella enterica Pdu microcompartment shell. J Biol Chem. 2010;285: 37838–37846.

70. Pang A, Warren MJ, Pickersgill RW. Structure of PduT, a trimeric bacterial microcompartment protein with a 4Fe–4S cluster-binding site. Acta Crystallogr D Biol Crystallogr. 2011;67: 91–96. doi:10.1107/S0907444910050201

71. Kim E, Tullman-Ercek D. A rapid flow cytometry assay for the relative quantification of protein encapsulation into bacterial microcompartments. Biotechnol J. 2013;

72. Havemann GD, Sampson EM, Bobik TA. PduA is a shell protein of polyhedral organelles involved in coenzyme B12-dependent degradation of 1, 2-propanediol in Salmonella enterica serovar Typhimurium LT2. J Bacteriol. 2002;184: 1253–1261.

73. Jorda J, Liu Y, Bobik TA, Yeates TO. Exploring Bacterial Organelle Interactomes: A Model of the Protein-Protein Interaction Network in the Pdu Microcompartment. PLoS Comput Biol. 2015;11: e1004067. doi:10.1371/journal.pcbi.1004067

74. Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. Proc Natl Acad Sci U S A. 2000;97: 6640–6645. doi:10.1073/pnas.120163297

75. Cheng S, Sinha S, Fan C, Liu Y, Bobik TA. Genetic analysis of the protein shell of the microcompartments involved in coenzyme B12-dependent 1, 2-propanediol degradation by Salmonella. J Bacteriol. 2011;193: 1385–1392.

76. Sinha S, Cheng S, Fan C, Bobik TA. The PduM Protein Is a Structural Component of the Microcompartments Involved in Coenzyme B12-Dependent 1,2-Propanediol Degradation by Salmonella enterica. J Bacteriol. 2012;194: 1912–1918. doi:10.1128/JB.06529-11

77. Jorda J, Lopez D, Wheatley NM, Yeates TO. Using comparative genomics to uncover new kinds of protein-based metabolic organelles in bacteria. Protein Sci. 2013;22: 179–195. doi:10.1002/pro.2196

78. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011;7: 539. doi:10.1038/msb.2011.75

79. Laemmli UK. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. Nature. 1970;227: 680–685. doi:10.1038/227680a0

80. Temme K, Salis H, Tullman-Ercek D, Levskaya A, Hong S-H, Voigt CA. Induction and Relaxation Dynamics of the Regulatory Network Controlling the Type III Secretion System Encoded within Salmonella Pathogenicity Island 1. J Mol Biol. 2008;377: 47–61. doi:10.1016/j.jmb.2007.12.044

81. Chen P, Ailion M, Bobik T, Stormo G, Roth J. Five promoters integrate control of the cob/pdu regulon in Salmonella typhimurium. J Bacteriol. 1995;177: 5401–5410.

82. Rondon MR, Escalante-Semerena JC. In vitro analysis of the interactions between the PocR regulatory protein and the promoter region of the cobalamin biosynthetic (cob) operon of Salmonella typhimurium LT2. J Bacteriol. 1996;178: 2196–2203.

83. Sheppard DE, Roth JR. A rationale for autoinduction of a transcriptional activator: ethanolamine ammonia-lyase (EutBC) and the operon activator (EutR) compete for adenosyl-cobalamin in Salmonella typhimurium. J Bacteriol. 1994;176: 1287–1296.

84. Kim EY, Tullman-Ercek D. A rapid flow cytometry assay for the relative quantification of protein encapsulation into bacterial microcompartments. Biotechnol J. 2014;9: 348–354. doi:10.1002/biot.201300391

85. Leal NA, Havemann GD, Bobik TA. PduP is a coenzyme-a-acylating propionaldehyde dehydrogenase associated with the polyhedral bodies involved in B 12-dependent 1, 2-propanediol degradation by Salmonella enterica serovar Typhimurium LT2. Arch Microbiol. 2003;180: 353–361.

86. Kerfeld CA, Erbilgin O. Bacterial microcompartments and the modular construction of microbial metabolism. Trends Microbiol. 2014; doi:10.1016/j.tim.2014.10.003

87. Drozdetskiy A, Cole C, Procter J, Barton GJ. JPred4: a protein secondary structure prediction server. Nucleic Acids Res. 2015;43: W389–W394. doi:10.1093/nar/gkv332

88. Srikumar S, Fuchs TM. Ethanolamine Utilization Contributes to Proliferation of Salmonella enterica Serovar Typhimurium in Food and in Nematodes. Appl Environ Microbiol. 2010;77: 281–290. doi:10.1128/AEM.01403-10

89. Pang A, Liang M, Prentice MB, Pickersgill RW. Substrate channels revealed in the trimeric *Lactobacillus reuteri* bacterial microcompartment shell protein PduB. Acta Crystallogr D Biol Crystallogr. 2012;68: 1642–1652. doi:10.1107/S0907444912039315

90. Conrado RJ, Mansell TJ, Varner JD, DeLisa MP. Stochastic reaction–diffusion simulation of enzyme compartmentalization reveals improved catalytic efficiency for a synthetic metabolic pathway. Metab Eng. 2007;9: 355–363.

91. Conrado RJ, Wu GC, Boock JT, Xu H, Chen SY, Lebar T, et al. DNA-guided assembly of biosynthetic pathways promotes improved catalytic efficiency. Nucleic Acids Res. 2012;40: 1879–1889.

92. Lee H, DeLoache WC, Dueber JE. Spatial organization of enzymes for metabolic engineering. Metab Eng. 2012;14: 242–251. doi:10.1016/j.ymben.2011.09.003

93. Sachdeva G, Garg A, Godding D, Way JC, Silver PA. In vivo co-localization of enzymes on RNA scaffolds increases metabolic production in a geometrically dependent manner. Nucleic Acids Res. 2014; doi:10.1093/nar/gku617

94. Huang X, Holden HM, Raushel FM. Channeling of Substrates and Intermediates in Enzyme-Catalyzed Reactions. Annu Rev Biochem. 2001;70: 149–180. doi:10.1146/annurev.biochem.70.1.149

95. Jandt U, You C, Zhang YH-P, Zeng A-P. Compartmentalization and Metabolic Channeling for Multienzymatic Biosynthesis: Practical Strategies and Modeling Approaches. Springer Berlin Heidelberg; 2013. pp. 1–25. Available: http://link.springer.com/chapter/10.1007/10_2013_221

96. Zhang Y-HP. Substrate channeling and enzyme complexes for biotechnological applications. Biotechnol Adv. 2011;29: 715–725. doi:10.1016/j.biotechadv.2011.05.020

97. Agapakis CM, Boyle PM, Silver PA. Natural strategies for the spatial optimization of metabolism in synthetic biology. Nat Chem Biol. 2012;8: 527–535. doi:10.1038/nchembio.975

98. Idan O, Hess H. The Origins of Activity Enhancement in Enzyme Cascades on Scaffolds. ACS Nano. 2013; 130905141001009. doi:10.1021/nn402823k

99. Conrado RJ, Varner JD, DeLisa MP. Engineering the spatial organization of metabolic enzymes: mimicking nature's synergy. Curr Opin Biotechnol. 2008;19: 492–499. doi:10.1016/j.copbio.2008.07.006

100. Ovádi J, Srere PA. Macromolecular Compartmentation and Channeling. In: Harry Walter DEB and PAS, editor. International Review of Cytology. Academic Press; 1999. pp. 255–280. Available: http://www.sciencedirect.com/science/article/pii/S007476960860529X

101. Moon TS, Dueber JE, Shiue E, Prather KLJ. Use of modular, synthetic scaffolds for improved production of glucaric acid in engineered E. coli. Metab Eng. 2010;12: 298–305. doi:10.1016/j.ymben.2010.01.003

102. Avalos JL, Fink GR, Stephanopoulos G. Compartmentalization of metabolic pathways in yeast mitochondria improves the production of branched-chain alcohols. Nat Biotechnol. 2013;31: 335–341. doi:10.1038/nbt.2509

103. Heinig U, Gutensohn M, Dudareva N, Aharoni A. The challenges of cellular compartmentalization in plant metabolic engineering. Curr Opin Biotechnol. 2013;24: 239–246. doi:10.1016/j.copbio.2012.11.006

104. You C, Myung S, Zhang Y-HP. Facilitated Substrate Channeling in a Self-Assembled Trifunctional Enzyme Complex. Angew Chem Int Ed. 2012;51: 8787–8790. doi:10.1002/anie.201202441

105. Torres Pazmiño DE, Riebel A, de Lange J, Rudroff F, Mihovilovic MD, Fraaije MW. Efficient Biooxidations Catalyzed by a New Generation of Self-Sufficient

Baeyer-Villiger Monooxygenases. ChemBioChem. 2009;10: 2595–2598. doi:10.1002/cbic.200900480

106. You C, Zhang Y-HP. Annexation of a High-Activity Enzyme in a Synthetic Three-Enzyme Complex Greatly Decreases the Degree of Substrate Channeling. ACS Synth Biol. 2014;3: 380–386. doi:10.1021/sb4000993

107. Bode SA, Minten IJ, Nolte RJM, Cornelissen JJLM. Reactions inside nanoscale protein cages. Nanoscale. 2011;3: 2376. doi:10.1039/c0nr01013h

108. Minten IJ, Claessen VI, Blank K, Rowan AE, Nolte RJM, Cornelissen JJLM. Catalytic capsids: the art of confinement. Chem Sci. 2011;2: 358. doi:10.1039/c0sc00407c

109. Cardinale D, Carette N, Michon T. Virus scaffolds as enzyme nano-carriers. Trends Biotechnol. 2012;30: 369–376. doi:10.1016/j.tibtech.2012.04.001

110. Kim EY, Tullman-Ercek D. Engineering nanoscale protein compartments for synthetic organelles. Curr Opin Biotechnol. 2013;24: 627–632. doi:10.1016/j.copbio.2012.11.012

111. Patterson DP, Schwarz B, Waters RS, Gedeon T, Douglas T. Encapsulation of an Enzyme Cascade within the Bacteriophage P22 Virus-Like Particle. ACS Chem Biol. 2014;9: 359–365. doi:10.1021/cb4006529

112. Broz P, Driamov S, Ziegler J, Ben-Haim N, Marsch S, Meier W, et al. Toward Intelligent Nanosize Bioreactors: A pH-Switchable, Channel-Equipped, Functional Polymer Nanocontainer. Nano Lett. 2006;6: 2349–2353. doi:10.1021/nl0619305

113. Siti W, de Hoog H-PM, Fischer O, Shan WY, Tomczak N, Nallani M, et al. An intercompartmental enzymatic cascade reaction in channel-equipped polymersome-in-polymersome architectures. J Mater Chem B. 2014;2: 2733. doi:10.1039/c3tb21849j

114. Gaitzsch J, Appelhans D, Wang L, Battaglia G, Voit B. Synthetic Bio-nanoreactor: Mechanical and Chemical Control of Polymersome Membrane Permeability. Angew Chem Int Ed. 2012;51: 4448–4451. doi:10.1002/anie.201108814

115. Schoffelen S, van Hest JCM. Multi-enzyme systems: bringing enzymes together in vitro. Soft Matter. 2012;8: 1736. doi:10.1039/c1sm06452e

116. van Oers M, Rutjes F, van Hest J. Cascade reactions in nanoreactors. Curr Opin Biotechnol. 2014;28: 10–16. doi:10.1016/j.copbio.2013.10.011

117. Louzao I, van Hest JCM. Permeability Effects on the Efficiency of Antioxidant Nanoreactors. Biomacromolecules. 2013;14: 2364–2372. doi:10.1021/bm400493b

118. Dergunov SA, Durbin J, Pattanaik S, Pinkhassik E. pH-Mediated Catch and Release of Charged Molecules with Porous Hollow Nanocapsules. J Am Chem Soc. 2014;136: 2212–2215. doi:10.1021/ja4106946

119. Kim KT, Cornelissen JJLM, Nolte RJM, van Hest JCM. A Polymersome Nanoreactor with Controllable Permeability Induced by Stimuli-Responsive Block Copolymers. Adv Mater. 2009;21: 2787–2791. doi:10.1002/adma.200900300

120. Kim J-K, Lee E, Lim Y, Lee M. Supramolecular Capsules with Gated Pores from an Amphiphilic Rod Assembly. Angew Chem Int Ed. 2008;47: 4662–4666. doi:10.1002/anie.200705863

121. Fornasiero F, Park HG, Holt JK, Stadermann M, Grigoropoulos CP, Noy A, et al. Ion exclusion by sub-2-nm carbon nanotube pores. Proc Natl Acad Sci. 2008;105: 17250–17255.

122. Theil EC. Ferritin protein nanocages use ion channels, catalytic sites, and nucleation channels to manage iron/oxygen chemistry. Curr Opin Chem Biol. 2011;15: 304–311. doi:10.1016/j.cbpa.2011.01.004

123. Roux B, Berneche S, Egwolf B, Lev B, Noskov SY, Rowley CN, et al. Perspectives on: Ion selectivity: Ion selectivity in channels and transporters. J Gen Physiol. 2011;137: 415–426. doi:10.1085/jgp.201010577

124. Yeates TO, Jorda J, Bobik TA. The Shells of BMC-Type Microcompartment Organelles in Bacteria. J Mol Microbiol Biotechnol. 2013;23: 290–299. doi:10.1159/000351347

125. Kinney JN, Axen SD, Kerfeld CA. Comparative analysis of carboxysome shell proteins. Photosynth Res. 2011;109: 21–32. doi:10.1007/s11120-011-9624-6

126. Kerfeld CA, Sawaya MR, Tanaka S, Nguyen CV, Phillips M, Beeby M, et al. Protein structures forming the shell of primitive bacterial organelles. Science. 2005;309: 936–938.

127. Glasgow JE, Capehart SL, Francis MB, Tullman-Ercek D. Osmolyte-Mediated Encapsulation of Proteins inside MS2 Viral Capsids. ACS Nano. 2012;6: 8658–8664. doi:10.1021/nn302183h

128. Capehart SL, Coyle MP, Glasgow JE, Francis MB. Controlled Integration of Gold Nanoparticles and Organic Fluorophores Using Synthetically Modified MS2 Viral Capsids. J Am Chem Soc. 2013;135: 3011–3016. doi:10.1021/ja3078472

129. Stonehouse NJ, Valeg\a ard K, Golmohammadi R, van den Worm S, Walton C, Stockley PG, et al. Crystal structures of MS2 capsids with mutations in the subunit FG loop. J Mol Biol. 1996;256: 330–339.

130. Axblom C, Tars K, Fridborg K, Orna L, Bundule M, Liljas L. Structure of Phage fr Capsids with a Deletion in the FG Loop: Implications for Viral Assembly. Virology. 1998;249: 80–88. doi:10.1006/viro.1998.9279

131. Dayan J, Wilson IB. The phosphorylation of Tris by alkaline phosphatase. Biochim Biophys Acta BBA-Spec Sect Enzymol Subj. 1964;81: 620–623.

132. Schoch R, Han J, Renaud P. Transport phenomena in nanofluidics. Rev Mod Phys. 2008;80: 839–883. doi:10.1103/RevModPhys.80.839

133. Plecis A, Schoch RB, Renaud P. Ionic transport phenomena in nanofluidics: experimental and theoretical study of the exclusion-enrichment effect on a chip. Nano Lett. 2005;5: 1147–1155. doi:10.1021/nl050265h

134. Muller BH, Lamoure C, Le Du M-H, Cattolico L, Lajeunesse E, Lemaître F, et al. Improving Escherichia coli alkaline phosphatase efficacy by additional mutations inside and outside the catalytic pocket. Chembiochem. 2001;2: 517–523.

135. Coleman JE. Structure and Mechanism of Alkaline Phosphatase. Annu Rev Biophys Biomol Struct. 1992;21: 441–483. doi:10.1146/annurev.bb.21.060192.002301

136. Goldstein L. Kinetic behavior of immobilized enzyme systems. Methods Enzymol. 1976;44: 397–443.

137. Patrick WM, Firth AE, Blackburn JM. User-friendly algorithms for estimating completeness and diversity in randomized protein-encoding libraries. Protein Eng. 2003;16: 451–457. doi:10.1093/protein/gzg057

138. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal. 2011;17: pp. 10–12. doi:10.14806/ej.17.1.200

139. Magoč T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinforma Oxf Engl. 2011;27: 2957–2963. doi:10.1093/bioinformatics/btr507

140. Rose GD, Creamer TP. Protein folding: Predicting predicting. Proteins Struct Funct Bioinforma. 1994;19: 1–3. doi:10.1002/prot.340190102

141. Alexander PA, He Y, Chen Y, Orban J, Bryan PN. A minimal sequence code for switching protein structure and function. Proc Natl Acad Sci. 2009;106: 21149–21154. doi:10.1073/pnas.0906408106

142. Gross I, Hohenberg H, Huckhagel C, Kräusslich H-G. N-Terminal Extension of Human Immunodeficiency Virus Capsid Protein Converts the In Vitro Assembly Phenotype from Tubular to Spherical Particles. J Virol. 1998;72: 4798–4810.

143. Cielens I, Ose V, Petrovskis I, Strelnikova A, Renhofa R, Kozlovska T, et al. Mutilation of RNA phage Qbeta virus-like particles: from icosahedrons to rods. FEBS Lett. 2000;482: 261–264.

144. Bruckman MA, Soto CM, McDowell H, Liu JL, Ratna BR, Korpany KV, et al. Role of Hexahistidine in Directed Nanoassemblies of Tobacco Mosaic Virus Coat Protein. ACS Nano. 2011;5: 1606–1616. doi:10.1021/nn1025719

145. Fiedler JD, Higginson C, Hovlid ML, Kislukhin AA, Castillejos A, Manzenrieder F, et al. Engineered mutations change the structure and stability of a virus-like particle. Biomacromolecules. 2012;13: 2339–2348. doi:10.1021/bm300590x

146. Stockley PG, Stonehouse NJ, Valegård K. Molecular mechanism of RNA phage morphogenesis. Int J Biochem. 1994;26: 1249–1260. doi:10.1016/0020-711X(94)90094-9

147. Geraets JA, Dykeman EC, Stockley PG, Ranson NA, Twarock R. Asymmetric Genome Organization in an RNA Virus Revealed via Graph-Theoretical Analysis of Tomographic Data. PLoS Comput Biol. 2015;11: e1004146. doi:10.1371/journal.pcbi.1004146

148. Valegård K, Liljas L, Fridborg K, Unge T. The three-dimensional structure of the bacterial virus MS2. Nature. 1990;345: 36–41. doi:10.1038/345036a0

149. Golmohammadi R, Valegård K, Fridborg K, Liljas L. The Refined Structure of Bacteriophage MS2 at 2·8 Å Resolution. J Mol Biol. 1993;234: 620–639. doi:10.1006/jmbi.1993.1616

150. Axblom C, Tars K, Fridborg K, Orna L, Bundule M, Liljas L. Structure of Phage fr Capsids with a Deletion in the FG Loop: Implications for Viral Assembly. Virology. 1998;249: 80–88. doi:10.1006/viro.1998.9279

151. Caspar DLD, Klug A. Structure and assembly of regular virus particles. Cold Spring Harbor Symp Quant Biol, 27. 1962.

152. Hooker JM, Kovacs EW, Francis MB. Interior Surface Modification of Bacteriophage MS2. J Am Chem Soc. 2004;126: 3718–3719. doi:10.1021/ja031790q

153. Kovacs EW, Hooker JM, Romanini DW, Holder PG, Berry KE, Francis MB. Dual-Surface-Modified Bacteriophage MS2 as an Ideal Scaffold for a Viral Capsid-Based Drug Delivery System. Bioconjug Chem. 2007;18: 1140–1147. doi:10.1021/bc070006e

154. Wu W, Hsiao SC, Carrico ZM, Francis MB. Genome-Free Viral Capsids as Multivalent Carriers for Taxol Delivery. Angew Chem Int Ed Engl. 2009;48: 9493–9497. doi:10.1002/anie.200902426

155. Ashley CE, Carnes EC, Phillips GK, Durfee PN, Buley MD, Lino CA, et al. Cell-Specific Delivery of Diverse Cargos by Bacteriophage MS2 Virus-like Particles. ACS Nano. 2011;5: 5729–5745. doi:10.1021/nn201397z

156. Capehart SL, Coyle MP, Glasgow JE, Francis MB. Controlled Integration of Gold Nanoparticles and Organic Fluorophores Using Synthetically Modified MS2 Viral Capsids. J Am Chem Soc. 2013;135: 3011–3016. doi:10.1021/ja3078472

157. Stonehouse NJ, Stockley PG. Effects of amino acid substitution on the thermal stability of MS2 capsids lacking genomic RNA. FEBS Lett. 1993;334: 355–359. doi:10.1016/0014-5793(93)80711-3

158. Chackerian B, Caldeira J do C, Peabody J, Peabody DS. Peptide epitope identification by affinity selection on bacteriophage MS2 virus-like particles. J Mol Biol. 2011;409: 225–237. doi:10.1016/j.jmb.2011.03.072

159. Peabody DS. The RNA binding site of bacteriophage MS2 coat protein. EMBO J. 1993;12: 595–600.

160. Peabody DS, Al-Bitar L. Isolation of viral coat protein mutants with altered assembly and aggregation properties. Nucleic Acids Res. 2001;29: e113.

161. Fiers W, Contreras R, Duerinck F, Haegeman G, Iserentant D, Merregaert J, et al. Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. Nature. 1976;260: 500–507. doi:10.1038/260500a0

162. Galaway FA, Stockley PG. MS2 viruslike particles: a robust, semisynthetic targeted drug delivery platform. Mol Pharm. 2013;10: 59–68. doi:10.1021/mp3003368

163. Glasgow J, Tullman-Ercek D. Production and applications of engineered viral capsids. Appl Microbiol Biotechnol. 2014;98: 5847–5858. doi:10.1007/s00253-014-5787-3

164. Toropova K, Stockley PG, Ranson NA. Visualising a Viral RNA Genome Poised for Release from Its Receptor Complex. J Mol Biol. 2011;408: 408–419. doi:10.1016/j.jmb.2011.02.040

165. Cadena-Nava RD, Comas-Garcia M, Garmann RF, Rao ALN, Knobler CM, Gelbart WM. Self-assembly of viral capsid protein and RNA molecules of different sizes: requirement for a specific high protein/RNA mass ratio. J Virol. 2012;86: 3318–3326. doi:10.1128/JVI.06566-11

166. Kler S, Wang JC-Y, Dhason M, Oppenheim A, Zlotnick A. Scaffold properties are a key determinant of the size and shape of self-assembled virus-derived particles. ACS Chem Biol. 2013;8: 2753–2761. doi:10.1021/cb4005518

167. Brown AD, Naves L, Wang X, Ghodssi R, Culver JN. Carboxylate-directed in vivo assembly of virus-like nanorods and tubes for the display of functional peptides and residues. Biomacromolecules. 2013;14: 3123–3129. doi:10.1021/bm400747k

168. Lang AS, Zhaxybayeva O, Beatty JT. Gene transfer agents: phage-like elements of genetic exchange. Nat Rev Microbiol. 2012;10: 472–482. doi:10.1038/nrmicro2802

169. Espagne E, Dupuy C, Huguet E, Cattolico L, Provost B, Martins N, et al. Genome Sequence of a Polydnavirus: Insights into Symbiotic Virus Evolution. Science. 2004;306: 286–289. doi:10.1126/science.1103066

170. Bézier A, Annaheim M, Herbinière J, Wetterwald C, Gyapay G, Bernard-Samain S, et al. Polydnaviruses of Braconid Wasps Derive from an Ancestral Nudivirus. Science. 2009;323: 926–930. doi:10.1126/science.1166788

171. Sadowski MI, Jones DT. The sequence–structure relationship and protein function prediction. Curr Opin Struct Biol. 2009;19: 357–362. doi:10.1016/j.sbi.2009.03.008

172. Cheng S, Bobik TA. The PduQ Enzyme Is an Alcohol Dehydrogenase Used to Recycle NAD+ Internally within the Pdu Microcompartment of Salmonella enterica. 2012;

173. Cheng S. Recycling of vitamin B12 and NAD+ within the Pdu microcompartment of Salmonella enterica. 2010; Available: http://lib.dr.iastate.edu/etd/11713/

174. Mastro AM, Babich MA, Taylor WD, Keith AD. Diffusion of a small molecule in the cytoplasm of mammalian cells. Proc Natl Acad Sci. 1984;81: 3414–3418.

175. Robertson RN. Lively Membranes. CUP Archive; 1983.

176. Bachovchin WW, Eagar RG, Moore KW, Richards JH. Mechanism of action of adenosylcobalamin: glycerol and other substrate analogs as substrates and inactivators for propanediol dehydratase - kinetics, stereospecificity, and mechanism. Biochemistry (Mosc). 1977;16: 1082–1092. doi:10.1021/bi00625a009

177. Schaechter M, MaalØe O, Kjeldgaard NO. Dependency on Medium and Temperature of Cell Size and Chemical Composition during Balanced Growth of Salmonella typhimurium. Microbiology. 1958;19: 592–606. doi:10.1099/00221287-19-3-592

178. Smith III FG, Deen WM. Electrostatic double-layer interactions for spherical colloids in cylindrical pores. J Colloid Interface Sci. 1980;78: 444–465. doi:10.1016/0021-9797(80)90584-6

Appendix A

Kinetic Modeling

Species

S^* substrate in assay solution (constant)

S substrate inside capsid

I inhibitor inside capsid

P product total (observable)

Rates

$$\frac{dS^*}{dt} = 0$$

$$\frac{dS}{dt} = -\frac{1}{F_V}\frac{dP}{dt} + k_1(S^* - S)$$

$$\frac{dI}{dt} = \frac{1}{F_V}\frac{dP}{dt} - k_2 I$$

$$\frac{dP}{dt} = \frac{k_{cat} S}{K_M\left(1 + \frac{I}{K_I}\right) + S}$$

Parameters

k_1 substrate transport

k_2 inhibitor transport

k_{cat} catalytic rate constant

K_M Michaelis-Menten constant

K_I competitive inhibition constant

F_V volume fraction of assay solution that is capsid lumen (~10^{-7})

Key assumptions

Michaelis-Menten kinetics with competitive product inhibition by phosphate

$\frac{dS^*}{dt} = 0$ (very large reservoir of substrate in the assay bulk)

$I = 0$ outside capsid (assay bulk is very large relative to capsid phase)

Intrinsic k_{cat}, K_M of enzymes unchanged by encapsulation in $MS2^{T71E}$ relative to $MS2^{WT}$

Rate of formation of product is normalized to enzyme concentration for comparison with experiment

Calculation of apparent diffusivities from fitted parameters

Consider the diffusive flux of substrate S as modeled: $J_M = k_2(S^* - S)[=]\frac{mol}{L-s}$, as compared to the diffusive flux $J_F = D_2\frac{(S^*-S)}{L}[=]\frac{mol}{m^2-s}$, where D_2 is the apparent diffusivity and L is the characteristic length of the flux.

We desire D_2 in terms of the fitted parameter k_2.

First, convert both fluxes to absolute flux in $\frac{mol}{s}$.

$$J_{abs} = J_F * A_{capsid} = \left(D_2\frac{(S^* - S)}{L}\right)A_{capsid}$$

$$J_{abs} = J_M * V_{capsid} = \left(k_2(S^* - S)\right)V_{capsid}$$

$$A_{capsid} = 4\pi\varepsilon R^2$$

Where ε is the fraction of the capsid surface area occupied by the pores. Considering that the diffusing molecule has a hydrodynamic diameter of ~6 Å and the pore has a diameter of ~18 Å, we consider the pore as having a radius as follows: $R_{pore} = \frac{18-6}{2} = 6$ Å.

$$\varepsilon = \frac{32 * \pi {R_{pore}}^2}{4\pi R^2} = 0.016$$

$$V_{capsid} = \frac{4}{3}\pi R^3$$

$$\therefore J_{abs} = \left(D_2\frac{(S^* - S)}{L}\right)4\pi\varepsilon R^2 = \left(k_2(S^* - S)\right)\frac{4}{3}\pi R^3$$

$$\left(\frac{D_2}{L}\right) = (k_2)\frac{1}{3}\frac{R}{\varepsilon}$$

$$D_2 = \frac{1}{3}\frac{RLk_2}{\varepsilon}$$

The characteristic length L is approximately the thickness of the capsid shell (1 nm) and the radius of the capsid R is approximately 13 nm.

Typically, D_2 has units of $\frac{cm^2}{s}$ and in this case k_2 has units of $\frac{1}{s}$. Thus, we can find D_2 as follows.

$$D_2 = \frac{1}{3}\frac{RLk_2}{\varepsilon} = 1.64x10^{-9}\frac{cm^2}{s}$$

For comparison, the diffusivity of phosphate in water at infinite dilution at 25ºC is $D_\infty \approx 5x10^{-6}\frac{cm^2}{s}$ and the diffusivity of atomic carbon in iron at 25ºC is $D \approx 10^{-12}\frac{cm^2}{s}$. We attribute this dramatic decrease in apparent diffusivity to the fact that the Debye length of the system is on the same order of magnitude as the pore radius, indicating that the electrical double layer likely penetrates far into the pore. There is thus a strong electrostatic repulsion between the negatively charged amino acid residues in the vicinity of the pore and the negatively charged substrate and phosphate species, greatly hindering diffusion through the capsid pore. This repulsion is modulated by the amino acid character of the pore residues, as evidenced by the differences in kinetic behavior between the $MS2^{WT}$ and $MS2^{T71E}$ capsid systems.

Smith and Deen[178] develop a theory for a system of charged spheres diffusing in charged pores similar to our system, and find that the Boltzmann factor $\frac{-E(r)}{kT}$ (where

$\frac{D}{D_\infty} \propto \int exp\left[\frac{-E(r)}{kT}\right] rdr$) becomes vanishingly small as the ratio of pore size to Debye length approaches 2, as is the case for our system. Although they considered somewhat larger pores (~10 nm), their findings are in agreement with our observations.

Appendix B

S1 Assumptions underlying the reaction-diffusion model

We model Pdu MCP function using a simple spatially resolved reaction-diffusion model incorporating passive and active transport across the cell membrane, passive transport across the MCP shell, and enzymatic catalysis of two critical steps in 1,2-propanediol metabolism: conversion of 1,2-propanediol to propionaldehyde by the PduCDE holoenzyme, and subsequent conversion of propionaldehyde to downstream products by the PduP and PduQ enzymes (Figure 6.1).

We model the bacterial cell as a spherically symmetrical compartment of radius 500 nm, containing at its center a single spherically symmetrical MCP of radius 100 nm (representing the same fraction of the cell volume as 5 Pdu MCPs in a typical bacterial cell). We consider the system at steady state and with constant external concentrations of 1,2-propanediol (P) and propionaldehyde (A).

The passive transport, active transport, and enzyme parameters were estimated *a priori* or from literature as shown in Table 6.1. In the case of P, in addition to passive transport across the cell membrane, we also include the possibility of active transport across the cell membrane by the putative membrane protein encoded by *pduF*. The passive transport velocities of P and A across the MCP shell can be treated independently or as being equal; we explore the necessity of selective permeability by allowing the velocities to differ.

We assume that the conversion of P to A in the absence of enzymatic catalysis is negligible, so the concentrations of P and A in the cytosol as function of cell radius can be solved analytically as detailed below. Inside the Pdu MCP, we assume Michaelis-Menten kinetic behavior of the PduCDE and PduP/PduQ enzymes. PduP and PduQ are redox-coupled by the recycling of NAD+/NADH, so we assume that their catalytic rates are equal at steady state. We assume that the enzyme-catalyzed reactions are irreversible, and for the purposes of this model neglect reactions downstream of PduP/Q.

S2 Governing equations in the cytosol

We can therefore solve for P and A as a function of cell radius as follows. The equations for diffusion of $1,2-$PD, P, and propanal, A, in the cytosol are

$$\partial_t P = D\nabla^2 P. \tag{1}$$

$$\partial_t A = D\nabla^2 A \tag{2}$$

Where D is the diffusion coefficient of the metabolites in the cytosol.

At steady state, the solutions to $\nabla^2 P = 0$ and $\nabla^2 A = 0$ can be found and have the form

$$P = \frac{B_1}{r} + B_2 \tag{3}$$

$$A = \frac{C_1}{r} + C_2 \tag{4}$$

where B_1, B_2, C_1, and C_2 are constant.

We can then specify the following boundary conditions enforcing continuity of flux of each metabolite at the cell membrane.

$$D\frac{\partial P}{\partial r} = j_c P_{out} + k_m^P \left(P_{out} - P_{cytosol}\right) \tag{5}$$

$$D\frac{\partial A}{\partial r} = k_m^A \left(A_{out} - A_{cytosol}\right) \tag{6}$$

Here active transport of P is set by the transport velocity j_c. The permeabilities of the cell membrane to A and P are set by the passive transport velocities k_m^A and k_m^P.

Similarly, the continuity of flux at the MCP shell sets the linking boundary condition between the concentrations inside the MCP and in the cytosol.

$$D\frac{\partial P}{\partial r} = k_c^P (P_{cytosol} - P_{MCP}) \tag{7}$$

$$D\frac{\partial A}{\partial r} = k_c^A (A_{cytosol} - A_{MCP}) \tag{8}$$

S2.1 Solution in cytosol

Using equation 4 in boundary condition 8 and equation 3 in boundary condition 7, we obtain:

$$B_2 = P_{MCP} - B_1 \left(\frac{D}{k_c^P R_c^2} + \frac{1}{R_c}\right) \tag{9}$$

$$C_2 = A_{MCP} - C_1 \left(\frac{D}{k_c^A R_c^2} + \frac{1}{R_c}\right) \tag{10}$$

so

$$P = B_1 \left(\frac{1}{r} - \frac{D}{k_c^P R_c^2} - \frac{1}{R_c}\right) + P_{MCP} \tag{11}$$

$$A = C_1 \left(\frac{1}{r} - \frac{D}{k_c^A R_c^2} - \frac{1}{R_c}\right) + A_{MCP} \tag{12}$$

Then using 12 in 6 and 11 in 5 we find;

$$B_1 = \frac{k_m^P R_b^2}{D}(P_{cyt}(r = R_b)) - \frac{R_b^2 P_{out}}{D}(j_c + k_m^P) = \frac{k_m^P P_{MCP} - P_{out}(j_c + k_m^P)}{\frac{D}{R_b^2} + k_m^p X} \tag{13}$$

$$C_1 = \frac{k_m^A R_b^2}{D}(A_{cyt}(r = R_b) - A_{out}) = \frac{A_{MCP} - A_{out}}{\frac{D}{k_m^A R_b^2} + X} \tag{14}$$

Where:

$$X = \left(\frac{D}{R_c^2 k_c} + \frac{1}{R_c} - \frac{1}{R_b}\right) \tag{15}$$

Yielding the following complete analytical solutions in the cytosol:

$$P = \frac{k_m^P P_{MCP} - P_{out}(j_c + k_m^P)}{\frac{D}{R_b^2} + k_m^p X} \left(\frac{1}{r} - \frac{D}{k_c^P R_c^2} - \frac{1}{R_c} \right) + P_{MCP} \tag{16}$$

$$A = \frac{A_{MCP} - A_{out}}{\frac{D}{k_m^A R_b^2} + X} \left(\frac{1}{r} - \frac{D}{k_c^A R_c^2} - \frac{1}{R_c} \right) + A_{MCP} \tag{17}$$

S3 Governing quations in microcompartment

In the microcompartment, enzymatic catalysis occurs, so the governing equations are instead:

$$\partial_t P = D\nabla^2 P - R_{CDE} \tag{18}$$

$$\partial_t A = D\nabla^2 A + R_{CDE} - R_{PQ} \tag{19}$$

where the equation for the rate of the PduCDE (diol dehydratase) reaction is

$$R_{CDE}(P) = \frac{V_{CDE} P}{K_{CDE} + P} \tag{20}$$

Here V_{CDE} is the maximum rate of dehydration by PduCDE. K_{CDE} is the half maximum concentration for dehydration.

The equation for the PduP and PduQ reactions is

$$R_{PQ}(A) = \frac{2V_{PQ} A}{K_{PQ} + A} \tag{21}$$

Here V_{PQ} is the maximum rate of aldehyde consumption by PduP, and K_{PQ} is the half maximum concentration. The rate is doubled, assuming that PduQ is present at the same concentration and reacts with the same kinetics at steady state, due to cofactor recycling.

We can use the solution in the cytosol to generate the following boundary condition at the MCP membrane:

$$\frac{\partial P}{\partial r} = -\frac{B_1}{r^2} = -\left(\frac{1}{R_c^2} \right) \frac{k_m^P P_{MCP} - P_{out}(j_c + k_m^P)}{\frac{D}{R_b^2} + k_m^p X} \tag{22}$$

$$\frac{\partial A}{\partial r} = -\frac{C_1}{r^2} = -\left(\frac{1}{R_c^2} \right) \frac{A_{MCP} - A_{out}}{\frac{D}{k_m^A R_b^2} + X} \tag{23}$$

S3.1 Non-dimensionalization

We can then define the following non-dimensional variables:

$$\rho = \frac{r}{R_c} \tag{24}$$

$$a = \frac{A}{K_{PQ}} \tag{25}$$

$$p = \frac{P}{K_{CDE}} \tag{26}$$

The governing equation for A in the MCP at steady state is therefore as follows:

$$\partial_t A = D\nabla^2 A + R_{CDE} - R_{PQ} = 0 \tag{27}$$

Applying the non-dimensionalization and letting $\kappa = \frac{K_{CDE}}{K_{PQ}}$, we have:

$$\frac{K_{PQ}D}{R_c^2}\frac{1}{\rho^2}\frac{\partial}{\partial\rho}\left(\rho^2\frac{\partial a}{\partial\rho}\right) = \frac{2V_{PQ}a}{1+a} - \frac{V_{CDE}p}{1+p} \tag{28}$$

Now let $\gamma = \frac{2V_{PQ}}{V_{CDE}}$ and $\xi = \frac{K_{PQ}D}{V_{CDE}R_c^2}$, yielding:

$$\xi\frac{1}{\rho^2}\frac{\partial}{\partial\rho}\left(\rho^2\frac{\partial a}{\partial\rho}\right) = \xi\nabla^2 a = \gamma\frac{a}{1+a} - \frac{p}{1+p} \tag{29}$$

Similarly for P,

$$\partial_t P = D\nabla^2 P - R_{CDE} = 0 \tag{30}$$

$$\kappa\xi\frac{1}{\rho^2}\frac{\partial}{\partial\rho}\left(\rho^2\frac{\partial p}{\partial\rho}\right) = \kappa\xi\nabla^2 p = \frac{p}{1+p} \tag{31}$$

S3.2 Boundary conditions

We can then nondimensionalize the boundary conditions as follows:

$$\frac{\partial A}{\partial r}|_{r=R_c} = \frac{-C_1}{R_c^2} = \frac{-1}{R_c^2}\left(\frac{A - A_{out}}{\frac{D}{k_m^A R_b^2} + X}\right) \tag{32}$$

$$\frac{K_{PQ}}{R_c}\frac{\partial a}{\partial\rho}|_{\rho=1} = \frac{-1}{R_c}\left(\frac{aK_{PQ} - A_{out}}{\frac{D}{k_m^A R_b^2} + X}\right) \tag{33}$$

$$\frac{\partial a}{\partial\rho}|_{\rho=1} = \frac{-1}{R_c K_{PQ}\left(\frac{D}{k_m^A R_b^2} + X\right)}\left(aK_{PQ} - A_{out}\right) \tag{34}$$

Where $X = \left(\frac{D}{R_c^2 k_c} + \frac{1}{R_c} - \frac{1}{R_b}\right)$ as before, so:

$$\frac{\partial a}{\partial \rho}|_{\rho=1} = \epsilon_a + \beta_a a \tag{35}$$

$$\epsilon_a = \frac{A_{out}}{R_c K_{PQ}\left(\frac{D}{k_m^A R_b^2} + X\right)} \tag{36}$$

$$\beta_a = \frac{-1}{R_c\left(\frac{D}{k_m^A R_b^2} + X\right)} \tag{37}$$

Similarly for P,

$$\frac{\partial P}{\partial r}|_{r=R_c} = \frac{-B_1}{R_c^2} = \frac{-1}{R_c^2}\left(\frac{k_m^P P - P_{out}\left(j_c + k_m^P\right)}{\frac{D}{R_b^2} + k_m^P X}\right) \tag{38}$$

$$\frac{K_{CDE}}{R_c}\frac{\partial p}{\partial \rho}|_{\rho=1} = \frac{-1}{R_c^2}\left(\frac{k_m^P K_{CDE} p - P_{out}\left(j_c + k_m^P\right)}{\frac{D}{R_b^2} + k_m^P X}\right) \tag{39}$$

Where $X = \left(\frac{D}{R_c^2 k_c} + \frac{1}{R_c} - \frac{1}{R_b}\right)$ as before, so:

$$\frac{\partial p}{\partial \rho}|_{\rho=1} = \epsilon_p + \beta_p p \tag{40}$$

$$\epsilon_p = \frac{P_{out}\left(j_c + k_m^P\right)}{K_{CDE} R_c\left(\frac{D}{R_b^2} + k_m^P X\right)} \tag{41}$$

$$\beta_p = \frac{-k_m^P}{R_c\left(\frac{D}{R_b^2} + k_m^P X\right)} \tag{42}$$

These nondimensional equations can then be solved numerically by a finite-difference approach to find the steady-state concentrations in the MCP, and the solutions in the cytosol follow directly.

S3.3 Analytical solution

If we assume that the concentration gradients in the MCP are small, then the concentrations P_{MCP} and A_{MCP} are approximately constant and the full solution to the reaction-diffusion equations in the MCP and cytosol can be found analytically. We therefore only require P_{MCP} and A_{MCP} as a function of P_{cyt} and A_{cyt} to specify the analytical solution.

First, consider the mass balance on A_{MCP} :

$$\frac{dA_{MCP}}{dt} = 0 = SA_{MCP}k_c^A(A_{cyt} - A_{MCP}) + V_{MCP}(R_{CDE} - R_{PQ}) \tag{43}$$

$$0 = 4\pi R_c^2 k_c^A(A_{cyt} - A_{MCP}) + \frac{4}{3}\pi R_c^3(\frac{V_{CDE}P_{MCP}}{K_{CDE} + P_{MCP}} - \frac{2V_{PQ}A_{MCP}}{K_{PQ} + A_{MCP}}) \tag{44}$$

And similarly for P_{MCP}:

$$\frac{dP_{MCP}}{dt} = 0 = SA_{MCP}k_c^P(P_{cyt} - P_{MCP}) - V_{MCP}R_{CDE} \tag{45}$$

$$0 = 4\pi R_c^2 k_c^P(P_{cyt} - P_{MCP}) - \frac{4}{3}\pi R_c^3 \frac{V_{CDE}P_{MCP}}{K_{CDE} + P_{MCP}} \tag{46}$$

This yields two equations which can be solved for A_{MCP} and P_{MCP} when the analytical solutions for $A_{cyt}(r = R_c)$ and $P_{cyt}(r = R_c)$ are substituted.

First solving for P_{MCP}, as this does not depend on A_{MCP} due to the irreversibility of PduCDE:

$$k_c^P(P_{cyt} - P_{MCP}) = \frac{1}{3}R_c\frac{V_{CDE}P_{MCP}}{K_{CDE} + P_{MCP}} \tag{47}$$

$$P_{cyt}(r = R_c) = -\frac{DC_3}{k_c R_c^2} + P_{MCP} \tag{48}$$

$$P_{cyt}(r = R_c) = P_{MCP} - \frac{D}{k_c R_c^2}\Big[\frac{P_{MCP}}{\frac{D}{R_b^2 k_m^P} + X} - \frac{P_{out}(j_c + k_m^P)}{\frac{D}{R_b^2} + k_m^P X}\Big] \tag{49}$$

$$\frac{3D}{R_c^3(\frac{D}{k_m^P R_b^2} + X)}\Big[P_{out}(1 + \frac{j_c}{k_m^P}) - P_{MCP}\Big] = \frac{V_{CDE}P_{MCP}}{K_{CDE} + P_{MCP}} \tag{50}$$

Let $P_{MCP} = pK_{CDE}$.

$$\frac{3DK_{CDE}}{V_{CDE}R_c^3(\frac{D}{k_m^P R_b^2} + X)}\Big[\frac{P_{out}(1 + \frac{j_c}{k_m^P})}{K_{CDE}} - p\Big] = \frac{p}{1 + p} \tag{51}$$

Let $Y = \frac{V_{CDE}R_c^3(\frac{D}{k_m^P R_b^2} + X)}{3DK_{CDE}}$ and let $Z = \frac{P_{out}(1 + \frac{j_c}{k_m^P})}{K_{CDE}}$.

$$\frac{1}{Y}(Z - p) = \frac{p}{1 + p} \tag{52}$$

$$p^2 + p(Y - Z + 1) - Z = 0 \tag{53}$$

$$\tag{54}$$

Let $E = Y - Z + 1$.

$$p = \frac{-E \pm \sqrt{E^2 + 4Z}}{2} \tag{55}$$

Now we can find A_{MCP} similarly, given the solution for P_{MCP}.

$$\frac{3D}{R_c^3(\frac{D}{k_m^A R_b^2} + X)}(A_{out} - A_{MCP}) = \frac{2V_{PQ}A_{MCP}}{K_{PQ} + A_{MCP}} - \frac{V_{CDE}P_{MCP}}{K_{CDE} + P_{MCP}} \tag{56}$$

Let $P_{MCP} = pK_{CDE}$ and $A_{MCP} = aK_{PQ}$.

$$\frac{3DK_{PQ}}{2V_{PQ}R_c^3(\frac{D}{k_m^A R_b^2} + X)}(\frac{A_{out}}{K_P Q} - a) + \frac{1}{2}\frac{V_{CDE}}{V_{PQ}}\frac{p}{1+p} = \frac{a}{1+a} \tag{57}$$

Let $U = \frac{2V_{PQ}R_c^3(\frac{D}{k_m^P R_b^2}+X)}{3DK_{PQ}}$, let $V = \frac{A_{out}}{K_{PQ}}$, and let $W = \frac{1}{2}\frac{V_{CDE}}{V_{PQ}}$.

$$\frac{1}{U}(V - a) + W\frac{p}{1+p} = \frac{a}{1+a} \tag{58}$$

$$a^2 + a(1 + U - V - UW\frac{p}{1+p}) - (UW\frac{p}{1+p} + V) = 0 \tag{59}$$

Let $F = 1 + U - V$.

$$a = \frac{-(F - UW\frac{p}{1+p}) \pm \sqrt{(F - UW\frac{p}{1+p})^2 + 4(UW\frac{p}{1+p} + V)}}{2} \tag{60}$$

Again, the solutions in the cytosol follow directly from these MCP solutions.

S4 Governing equations for no MCP case

In the case when there is no Pdu MCP, we assume that the same number of enzymes are now distributed throughout the cell. The equations in the cell are therefore now as follows:

$$\partial_t A = D\nabla^2 A + R_{CDE} - R_{PQ} \tag{61}$$

$$\partial_t P = D\nabla^2 P - R_{CDE} \tag{62}$$

where the equation for the PduCDE (diol dehydratase) reaction is

$$R_{CDE}(P) = \frac{V_{CDE}P}{K_{CDE} + P} \tag{63}$$

The equation for the PduP and PduQ reactions is

$$R_{PQ}(A) = \frac{2V_{PQ}A}{K_{PQ} + A} \tag{64}$$

These can be non-dimensionalized as follows (*c.f.* with above for MCP case): Define the following non-dimensional variables:

$$\rho = \frac{r}{R_b} \tag{65}$$

$$a = \frac{A}{K_{PQ}} \tag{66}$$

$$p = \frac{P}{K_{CDE}} \tag{67}$$

The governing equation for A in the cell at steady state is as follows:

$$\partial_t A = D\nabla^2 A + R_{CDE} - R_{PQ} = 0 \tag{68}$$

Applying the non-dimensionalization and letting $\kappa = \frac{K_{CDE}}{K_{PQ}}$, we have:

$$\frac{K_{PQ}D}{R_b^2}\frac{1}{\rho^2}\frac{\partial}{\partial\rho}\left(\rho^2\frac{\partial a}{\partial\rho}\right) = \frac{2V_{PQ}a}{1+a} - \frac{V_{CDE}p}{1+p} \tag{69}$$

Now let $\gamma = \frac{2V_{PQ}}{V_{CDE}}$ and $\xi = \frac{K_{PQ}D}{V_{CDE}R_b^2}$, yielding:

$$\xi\frac{1}{\rho^2}\frac{\partial}{\partial\rho}\left(\rho^2\frac{\partial a}{\partial\rho}\right) = \xi\nabla^2 a = \gamma\frac{a}{1+a} - \frac{p}{1+p} \tag{70}$$

Similarly for P,

$$\partial_t P = D\nabla^2 P - R_{CDE} = 0 \tag{71}$$

$$\kappa\xi\frac{1}{\rho^2}\frac{\partial}{\partial\rho}\left(\rho^2\frac{\partial p}{\partial\rho}\right) = \kappa\xi\nabla^2 p = \frac{p}{1+p} \tag{72}$$

S4.1 Boundary conditions

Now considering the boundary conditions,

$$\frac{\partial A}{\partial r}|_{r=R_b} = k_m^A(A_{out} - A) \tag{73}$$

$$\frac{DK_{PQ}}{R_b}\frac{\partial a}{\partial\rho}|_{\rho=1} = k_m^A(A_{out} - K_{PQ}a) \tag{74}$$

$$\frac{\partial a}{\partial \rho}|_{\rho=1} = \epsilon_a + \beta_a a \tag{75}$$

$$\epsilon_a = \frac{R_b A_{out} k_m^A}{D K_{PQ}} \tag{76}$$

$$\beta_a = \frac{-R_b k_m^A}{D} \tag{77}$$

Similarly for P,

$$\frac{\partial p}{\partial \rho}|_{\rho=1} = \epsilon_p + \beta_p p \tag{78}$$

$$\epsilon_p = \frac{R_b P_{out}}{D K_{CDE}}(j_c + k_m^P) \tag{79}$$

$$\beta_p = \frac{-R_b k_m^P}{D} \tag{80}$$

These equations can once again be solved numerically by a finite difference approach.