Identifying Interactions Between Ethanolamine Utilization Bacterial Microcompartment Cargo and Shell Proteins

Kelsey Dahlgren

Submitted under the supervision of Dr. Claudia Schmidt-Dannert to the University Honors Program at the University of Minnesota-Twin Cities in partial fulfillment of the requirements for the degree of Bachelor of Science, *summa cum laude* in Biochemistry.

April 20, 2015

Abstract

One of the goals of synthetic biology is to engineer metabolic pathways to produce valuable chemical compounds and biofuels. To gain economic advantage over abiotic processes, metabolically engineered processes must attain high fluxes, sustain high yields, and have minimal effect on host growth rates. Major obstacles that must be overcome for the use of synthetic metabolic pathways include diffusion limitations, alternative metabolic routes, toxic intermediates, and inhibitory products. To avoid these problems in their own metabolic pathways, bacteria use bacterial microcompartments (BMCs), organelles composed entirely of protein that contain cargo proteins, functionally related enzymes and auxiliary proteins, within a proteinaceous shell. The ability to group enzymes in a BMC shell that regulates substrate and product transport is a promising tool for synthetic biology. However, knowledge of BMC cargo protein localization mechanisms is unknown, but necessary for the development of BMCs as nanocontainers for biosynthesis or biocatalysis. This study sought to optimize heterologous expression of of ethanolamine utilization (Eut) BMC cargo proteins with the objective of studying interactions between BMC shell proteins and cargo proteins using *in vitro* pull-down experiments. Results from this study will inform future work on engineering BMCs for synthetic biology applications.

Introduction

The development of new, sustainable technologies has been driven by limited resources and the environmental concerns of existing technologies. One sustainable technology that has not yet been fully realized is metabolic engineering. Metabolic engineering is the modification and optimization of the processes performed by an organism that enables the organism to produce pharmaceuticals, fuels, or other valuable resources (Boyle and Silver, 2012). Its ability to replace expensive, inefficient, and unsustainable abiotic or biotic chemical processes with cheaper, more efficient, and/or more sustainable options makes metabolic engineering an attractive option to many different parties (Woolston et al., 2013). One notable success of metabolic engineering is the production of artemisinic acid, a precursor to the antimalarial drug artemisinin (Ro et al., 2006; Westfall et al., 2012; Farhi et al., 2011). The pharmaceutical company Sanofi has produced vast quantities of artemisinic acid via this method, which translates to millions of treatments (Peplow, 2013). Sanofi's success illustrates the power of metabolic engineering and provides insight as to what the future of metabolic engineering may hold.

While Sanofi serves as a prime example of the power of metabolic engineering, the toolkit of metabolic engineering strategies is far from complete. To attain an economic advantage over abiotic processes, engineered pathways must carry high fluxes, sustain high yields, and have a minimal effect on growth rates (Bar-Even and Tawfik, 2013). Persisting obstacles metabolic engineering must overcome include inefficient enzymes, volatile intermediates, inhibitory products, alternate metabolic pathways, and toxic intermediates (Boyle and Silver, 2012; Agapakis et al., 2012; Lee et al., 2012). For example, inefficient enzymes slow the rate product formation when incorporated into metabolic pathways. Furthermore, alternate

metabolic pathways result in the formation of side products, decreasing the product yield. Efficient metabolic engineering requires the development of tools to control metabolic flux, prevent side reactions, and tolerate toxic intermediates to realize its full potential as a sustainable technology.

Nature has developed several strategies to overcome the aforementioned limitations to metabolic engineering. Common strategies in nature are the spatial organization of metabolic pathways through the use of scaffolds or compartments. Scaffolds are often used to reduce side reactions in a metabolic pathway, and are common in signal transduction pathways (Agapakis et al., 2012). Microbial consortia, symbioses of two different types of microbes, are an example of compartmentalization (Agapakis et al., 2012). For example, cyanobacteria separate the incompatible processes of photosynthesis and nitrogen fixation into separate cells (Fay, 1992). Membranous organelles such as lysosomes are a type of compartment used in eukaryotes used to maintain a specific environment for the function of important enzymes (Luzio et al., 2003). More recent studies have demonstrated the existence of bacterial microcompartments (BMCs), proteinaceous organelle-like structures, in some bacteria (Kerfeld et al., 2010). The spatial engineering of metabolic pathways using compartments or scaffolds has the potential to significantly improve the ability of metabolically engineered organisms to produce commodities.

BMCs are compartments consisting of a protein shell enclosing cargo proteins (Kerfeld et al., 2010). The cargo proteins are enzymes and associated proteins that form a metabolic pathway (Kerfeld et al, 2010). The most well studied BMCs are the ethanolamine utilization BMCs and 1,2-propanediol utilization BMCs in *Salmonella* and carbon-fixation BMCs in Cyanobacteria (Kerfeld et al., 2010). The BMC shell is composed of BMC shell proteins that assemble into hexameric arrays with central pores that facilitate small molecule exchange across

the shell (Yeates et al, 2011). The cargo proteins are thought to associate with the shell proteins during BMC formation, and are eventually enclosed by the BMC shell (Choudhary et al, 2012; Fan et al., 2012; Kinney et al., 2012). BMCs are an elegant way to co-localize proteins in metabolic pathways in nature. If metabolic engineering harnesses the power of BMCs for spatial engineering, many previously unfeasible metabolic pathways could be used for the production of valuable compounds.

The aim of this project is to determine the mechanism of cargo protein targeting in ethanolamine utilization (Eut) BMCs from *Salmonella enterica*. Previously, sequences that target proteins to BMC shells have been identified (Choudhary et al., 2012; Fan et al., 2010). In the literature, there are currently two model mechanisms for cargo protein targeting, the cognate pair mechanism and the general binding mechanism; however, both lack strong supporting biochemical evidence. Determining the cargo protein targeting mechanism in Eut BMCs will provide biochemical evidence for a cargo protein targeting mechanism. The first model is the cognate pair mechanism, in which specific interactions between a shell and a cargo protein direct that cargo protein to a BMC; each shell-cargo protein pair interacts using a different binding mechanism with different targeting sequences (Fan et al., 2012). The second model is the general binding mechanism, in which the binding mechanism is conserved between many shell and many cargo proteins (Kinney et al., 2012). Each shell-cargo protein pair interacts in a similar manner through similar targeting sequences in the general conserved mechanism theory (Kinney et al., 2012). To elucidate the Eut BMC binding mechanism, protein interactions between the shell and cargo proteins will be identified using biochemical pull-downs. If many shell proteins interact with many cargo proteins, the general binding mechanism will be

supported. If each shell protein binds to one cargo protein, the specific binding mechanism will be supported.

Intracellular Spatial Engineering of Bacterial Metabolism for Biotechnology

Spatial organization groups enzymes in the same pathway to reduce the issues related to inefficient enzymes, volatile intermediates, inhibitory products, alternate metabolic pathways, and toxic intermediates (Boyle and Silver, 2012; Agapakis et al., 2012; Lee et al., 2012). Scaffolding is a spatial engineering technique that can reduce the negative effects of volatile and toxic intermediates and improve the function of inefficient enzymes to improve the yield of small molecules. Scaffolds group enzymes in the same metabolic pathway together in the cell, therefore increasing the local concentrations of intermediate while the concentrations of the intermediate in the cell remains low. The decrease in the distance between enzymes on a scaffold may lead to metabolic channeling, the direct passing of a metabolite from one enzyme to another without release into the solution (Huang et al., 2001). Scaffolds are often used in nature, and have been effective in many biotechnological applications as well. For example, cellulosomes are large bacterial complexes used to arrange cellulose breakdown enzymes on scaffolds on the cell membrane (Boyle and Silver, 2012). Expressing these proteins heterologously has demonstrated an increase in enzyme activity in a cooperative manner (Boyle and Silver, 2012). Additionally, proteins can be targeted to the cellulosome through the use of dockerin domains, greatly increasing their potential for biotechnological applications (Agapakis et al., 2012). Dueber et al. reported an example of scaffold use for the production of mevalonate, a precursor to a class of chemicals known as isoprenoids, which have important therapeutic and

commercial value (2009). Modifying the enzymes involved in mevalonate production to bind to an introduced scaffold increased artificial mevalonate synthesis 77-fold (Dueber et al., 2009).

Nucleic acid scaffolds have also been used for spatial metabolic engineering. In these scaffolds, the enzymes have been engineered to contain a domain that will bind a specific sequence of nucleic acid. Nucleic acid scaffolds allow for the construction of many interesting geometric shapes, allowing for unique spatial arrangements of enzymes (Boyle and Silver, 2012; Lee et al., 2012). Changes in the length and sequence of the nucleic acid scaffold allow the customization of the number and type of each enzyme on a scaffold. In addition, by programming the nucleic acid assembly, enzymes scaffolded onto those scaffolds can be arranged in custom configurations, possibility allowing for metabolic channeling between enzymes. RNA scaffolds have been used to improve hydrogen production 48-fold (Delebecque et al., 2011). Targeting to nucleic acid scaffolds is simple as well; many options are available for the design of DNA or RNA binding domains, which can be fused to the enzymes. Despite its numerous advantages, scaffolding is an imperfect system; extremely toxic or volatile intermediates can still escape the metabolic enzymes and inhibitory molecules still maintain access to scaffolded enzymes.

Subcellular compartments are another type of spatial engineering tool. These compartments can increase the efficiency of inefficient enzymes, and negate the effects of toxic or volatile intermediates by localizing high concentrations of these metabolites in the compartment and preventing their transport out of the compartment. Compartments are also capable of concentrating proteins to create a crowded environment that prevents the diffusion of substrates and products away from the active sites of the enzymes (Brinsmade et al., 2005). Subcellular compartments can also prevent the loss of product to alternate metabolic routes

through the barrier to molecule transport created by the compartment. Additionally, inhibitory molecules could be actively transported out of the compartment, preventing the negative effects of metabolite production due to inhibition of enzymes.

Many examples of subcellular compartment engineering have been reported. Engineered subcellular compartments can increase catalytic efficiency of certain enzymes. For example, producing methyl iodine in yeast vacuoles increased methyl-iodine yield about 30% over production in the cytosol (Bayer et al., 2009). Though using this method works well for eukaryotes, however, most prokaryotes lack lipid-enclosed organelles, and a different tool must be developed to build subcellular compartments in prokaryotes.

In prokaryotes, several different proteinaceous compartments have been used to spatially engineer metabolic enzymes in nature. Very small protein shells, lumazine synthase and encapsulin, are used to catalyze one step of a metabolic process *in vivo* (Lee et al., 2012). Bacterial microcompartments (BMCs), another type of proteinaceous subcellular compartment, are more complex and larger than lumazine synthase and encapsulin (Lee et al., 2012). BMCs have multiple shell proteins and enclose a metabolic pathway involving several different enzymes when found in nature (Lee et al., 2012). Lumazine synthase, encapsulin, and BMC nanocompartments each have pores in their shells that allow for the transfer of metabolites across the shells (Lee et al., 2012). Recently, a metabolic pathway to produce ethanol from pyruvate has been engineered into a BMC, demonstrating the viability of spatial metabolic engineering within proteinaceous subcellular compartments (Lawrence et al., 2014).

Spatial metabolic engineering has proven to be effective in increasing the yield of products in many different model systems. Often, the increased catalytic efficiency of the proteins due to increases in local concentrations of intermediates results in a greater product

yield. By increasing the efficiency of enzymes, it is possible to decrease the amount of protein expressed whilst maintaining the same output. Thus, less cellular resources will be channeled into producing large amounts of the recombinant proteins, permitting the introduction of additional heterologous proteins allowing more complex metabolic processes. Though scaffolds can reduce some of the issues associated with metabolic engineering, compartments have the potential to solve those issues entirely using controlled transport into and out of the compartment. Further development of spatial engineering techniques has the potential to significantly increase the number of pathways and functions that can be engineered into organisms.

Ethanolamine Utilization BMCs as a Model System for Spatial Engineering

There are several different types of prokaryotic subcellular compartments that could be utilized for spatial metabolic engineering, including lumazine synthase compartments, encapsulins, and BMCs. While their simplicity makes lumazine synthase and encapsulin attractive for biotechnological applications, their size limits their uses; very few enzymes can be enclosed in a single compartment, and getting several different enzymes to make up a metabolic pathway within on compartment would be near impossible (Lee et al., 2012). Though BMCs are more complicated and subsequently more difficult to heterologously express, the larger size of BMCs increases their potential for enclosing an entire engineered metabolic pathway (Boyle and Silver, 2012; Lee et al., 2012). BMCs are a superior model to investigate for spatial engineering of an entire metabolic pathway, because BMCs are larger than lumazine synthase and encapsulin.

BMC shells are generally shaped like polyhedrons and are likely selective for the transport of specific metabolites, and contain domains targeted by the cargo proteins. Though BMC shells resemble viral capsids, no evolutionary relationship between BMCs and viruses has been found (Yeates et al., 2010; Yeates et al., 2007). The conserved structure of BMC shells is due to the evolutionary conservation of an ~100 amino acid sequence, a hallmark of shell proteins (Yeates et al., 2010). Shell proteins generally form disc-shaped hexamers (or trimers if the protein has been duplicated) (**Figure 1**) (Yeates et al., 2010). The disc has a pore in the middle, which has been hypothesized to deliver specific metabolites to the BMC based on several crystal structures containing sulfate in the pore (Yeates et al., 2010; Kerfeld, 2010; Tanaka et al., 2009). In fact, the same proteins have been crystallized in an open and closed pore conformation, further supporting the hypothesis that the pore can transport molecules (Yeates et al., 2010; Klein et al., 2009; Tanaka et al., 2010). Crowley et al., provide evidence for pore transport selectivity in one type of BMC (2010). Another feature of shell proteins is that their N- and C- termini face the inside of the shell and thus can interact with the cargo proteins (Tanaka et al., 2009; Fan et al., 2012). Furthermore, proteins targeted to the interior of one type of BMC bind to C-terminal sequences of a shell protein (Fan et al., 2012).

Though the shells are similar, BMCs encapsulate diverse metabolisms in different bacteria. The function performed by the BMC is determined by the cargo proteins, which carry out a specific metabolic process (Yeates et al., 2011). BMC functions include carbon fixation, degradation of toxic compounds, and the utilization of specific metabolites (Yeates et al., 2010; Kerfeld et al., 2010). Each process overcomes a specific metabolic obstacle. For example, carboxysomes, a type of BMC involved in carbon fixation in Cyanobacteria, increase the efficiency of an inefficient enzyme, RuBisCo. RuBisCo fixes CO_2, but fails to differentiate

between CO_2 and O_2 when performing carbon fixation (Shively, 1973). Carbonic anhydrase is also present in carboxysomes (Yeates et al., 2010). Yeates et al. review the process of carbon fixation in carboxysomes (2010). HCO_3^- is transported into the shell and dehydrated to CO_2 by carbonic anhydrase in the presence of RuBisCo. Using this method, RuBisCo is sequestered in an area of high CO_2 and low O_2, drastically increasing the efficiency of the enzyme. The carboxysome shell prevents the volatile intermediate, CO_2, from diffusing away while preventing an inhibitory molecule, O_2, from replacing CO_2 in RuBisCo.

Other examples of BMCs are the 1,2-propanediol utilization (Pdu) BMCs and the ethanolamine utilization (Eut) BMCs found in the genus *Salmonella.* Pdu BMCs catalyze 1,2 propanediol degradation (Havemann et al., 2003). Coenzyme B_{12}-dependent diol dehydratase in the Pdu pathway produces a propionaldehyde, a mutagen that compromises the integrity of the DNA (Sampson and Bobik, 2008). The Pdu BMC acts to protect the cell against this toxic intermediate. Eut BMCs metabolize ethanolamine to a usable carbon and nitrogen source (Penrod and Roth, 2006). The purpose of the Eut BMC is to retain acetaldehyde, a volatile intermediate (Penrod and Roth, 2006). Another report claims that the BMC serves to concentrate the enzymes and improve their catalytic efficiency (Brinsmade et al., 2005). Without the BMC shell, growth on ethanolamine does not occur; however, when the enzymes are overexpressed, growth is reestablished (Brinsmade et al., 2005). The BMC likely increases the metabolic flux of ethanolamine through the Eut pathway through the retention of a volatile intermediate within the BMC shell; the higher concentration of substrate likely improves the efficiency of the enzyme that converts acetaldehyde to acetyl-CoA.

Pdu BMCs, Eut BMCs, and carboxysomes are potential candidates for spatial metabolic engineering in prokaryotes. The Pdu BMC, Eut BMC, and carboxysome can be expressed

heterologously in *E. coli*, an industrial host (Parsons et al., 2010; Choudhary et al., 2012; Bonacci et al., 2012). However, Pdu BMCs require a minimum of 6 proteins to express heterologously and carboxysomes have only been expressed using 9 proteins (Parsons et al., 2012, Bonacci et al., 2012). In contrast, the Eut BMC has been expressed in *E. coli* using a single gene, EutS (Choudhary et al., 2012). Thus, engineering Eut BMC shells into a host organism in comparison to Pdu BMC shells or carboxysome shells is significantly simpler. If simplicity in the spatial engineering system is desired, Eut BMCs are preferable to other types of BMCs.

Another facet of developing BMCs for use in metabolic engineering is targeting enzymes to the BMC. Though carboxysomes have been expressed heterologously, the only proteins that have been targeted to the carboxysomes are RuBisCo or RuBisCo fusion proteins (Bonacci et al., 2012). In Pdu BMC systems, one signal sequence has been described that targets cargo proteins for encapsulation within the BMC shell (Fan et al., 2010). One signal sequence has also been described for Eut BMCs (Choudhary et al., 2012). If the ability to target enzymes to the BMC is desired in the spatial engineering system, the use of a Pdu or Eut BMC would be preferred over a carboxysome; both Eut and Pdu signal sequences are less than 20 aa long, and carboxysome targeting requires fusion to a full-length protein (Fan et al., 2010; Choudhary et al., 2012; Bonacci et al., 2012). Assuming that researchers want a spatial engineering system that is simple to engineer into a host organism and has a short targeting sequence, Eut BMCs are currently the best candidate for compartment-based spatial engineering in prokaryotes. To utilize Eut BMCs in metabolic engineering, ideally several different target sequences with different target efficiencies would be used to produce an optimal metabolic pathway inside the BMC.

The Eut BMC shell is composed of several proteins. The major shell protein, EutS, forms hexamers (**Figure 1**). While EutS has been shown to be sufficient for heterologous expression of a Eut BMC shell, it is also necessary to produce a Eut BMC shell (Choudhary et al., 2012). Although EutS contains a conserved BMC motif (the ~100 aa sequence present in most BMC shell proteins), the hexamer has a slight curvature instead of the flat hexamers of most shell proteins (Frank et al., 2012). This may explain why EutS can form a shell, while other model systems require several different shell proteins. Another major Eut BMC shell protein is EutM, which also contains a conserved BMC motif (Frank et al., 2013). In addition to EutM and EutS, less abundant proteins such as EutL, EutN, and EutK also contribute to the Eut BMC shell. EutL has a tandem domain, and has been crystallized with the pore in an open and closed conformation, leading researchers to hypothesize that it plays a role in substrate transport across the shell (Sagermann et al., 2009). EutK may also form a pore. EutN is similar in sequence to specific carboxysome proteins that form pentamers at vertexes of the shell (Yeates et al., 2010). However, EutN forms hexamers when crystallized, suggesting that it may play a different role in Eut BMCs than its analog in carboxysomes (Yeates et al., 2010). In summary, EutM and EutS are the major structural proteins of the Eut BMC, although EutK, EutL, and EutN may perform useful functions as well, including substrate transport.

The Eut cargo proteins consist of enzymes that catalyze ethanolamine degradation and utilization (**Figure 2**). EutBC is a coenzyme B_{12}- dependent ethanolamine ammonia lyase, it removes the ammonia from ethanolamine to produce acetaldehyde (Faust et al., 1990). EutA is an ethanolamine ammonia lyase reactivase that prevents EutBC inactivation in the presence of O_2 (Brinsmade et al., 2005). EutA exchanges oxidized vitamin B_{12} bound to EutBC for reduced B_{12} (Brinsmade et al., 2005). EutE converts acetaldehyde produced by EutBC to acetyl-CoA

(Stojiljkovic et al., 1995). The ammonia is used as a nitrogen source for the cell, while acetyl-CoA is used as an energy and carbon source (Starai et al., 2005). EutG is an alcohol dehydrogenase that reduces acetaldehyde to ethanol (Penrod et al., 2004). EutD converts acetyl-CoA to acetyl-phosphate (Starai et al., 2005). The acetyl-phosphate is used by acetate kinase to produce ATP and acetate, which is then excreted (Starai et al., 2005). EutT synthesizes vitamin B_{12} (Buan et al., 2004). Though there are many enzymes catalyzing several processes, the major steps are the conversion of ethanolamine to acetaldehyde to produce ammonia by EutBC and the conversion of acetaldehyde to acetyl-CoA by EutE (**Figure 2**).

Cargo Protein Targeting in Bacterial Microcompartments

The targeting mechanisms of cargo proteins into BMCs are not well understood. However, to realize the potential of BMCs for metabolic engineering in synthetic biology, a method to insert enzymes into BMCs must be developed. One possible approach to developing a targeting method is to elucidate the targeting mechanism of natural cargo proteins to BMCs.

Bacteria often use signal recognition particles (SRPs) to control the traffic of secretory and membrane proteins within the cell (Keenan et al., 2001). However, cargo protein targeting likely does not use this mechanism because SRPs target proteins to the membrane, while the cargo proteins are localized in the BMC. Cargo proteins are targeted to BMC shell with signal sequences (Fan et al., 2010; Choudhary et al., 2012). However other targeting mechanisms, e.g. those use by encapsulin and lumazine synthase, may be used and remain yet to be discovered. In encapsulin, a conserved sequence near the C-terminus of cargo proteins is required for cargo protein sequestration (Sutter et al., 2008). In lumazine synthase shells, negatively charged residues on the lumazine synthase subunits can recruit GFP and HIV protease tagged with ten

positively charged amino acids, which strongly supports electrostatic interactions as the targeting mechanism (Lee et al., 2012).

The discovery of signal sequences on cargo proteins has raised debate about the model mechanism of cargo protein targeting. Fan et al. reported targeting interactions between the N-terminus of cargo protein PduP and the C-termini of the shell proteins PduA and PduJ (2012). Additionally, an 18 amino acid sequence on the N-terminus of PduP is sufficient to target GFP and GST to a Pdu shell (Fan et al., 2010). They hypothesize that specific interactions between cognate pairs of proteins target cargo proteins to BMCs (Fan et al., 2012). In the cognate pair mechanism theory, each pair of enzymes would interact in a different method (**Figure 3A**). For example, one shell-cargo pair may use electrostatic interactions while a different shell-cargo pair would associate using hydrophobic amino acids. In contrast, Kinney et al. claims that a general binding mechanism is conserved between many shell and many cargo proteins based on the analysis of carboxysome protein interactions (2012). In the general conserved mechanism theory, each shell-cargo protein pair interacts in a similar manner (**Figure 3B**). For example, each shell protein may have positive charges on one side an amphipathic alpha-helix that interact with negative charges on one side of an amphipathic alpha-helix in each cargo protein. Because the interaction mechanisms are similar in the general method, most shell proteins could interact with most cargo proteins and vice-versa. Due to slight differences in the targeting domain of the shell or cargo proteins, different shell and cargo proteins might interact at different strengths, resulting in different cargo proteins encapsulated in varying amounts within the BMC. Though there is evidence for protein targeting within BMCs, there is no biochemical evidence favoring either the general mechanism or the conserved mechanism.

As mentioned above, EutS has been shown to build BMCs in *E. coli* (Choudhary et al., 2012). The first 19 amino acids of EutC act as a signal sequence to target a protein such as GFP to the Eut BMC (Choudhary et al., 2012). To identify additional Eut cargo-shell binding partners and determine if Eut BMCs follow the cognate pair or general conserved mechanism, biochemical pull-downs using Eut BMC shell and cargo proteins were utilized. This method was chosen because it is a simple method to test interactions between many pairs of proteins and has been previously utilized to find Pdu BMC shell-cargo binding partners.

The experiments performed for this report have two goals, to identify Eut cargo-shell binding partners and determine if Eut BMCs follow the cognate pair or general conserved mechanisms of binding. Based on structural similarities between Eut BMC proteins, I propose that a general mechanism targets Eut cargo proteins to Eut BMCs. If many cargo proteins bind the same shell protein or vice-versa, the conserved general mechanism theory will be supported. However, if each cargo protein interacts with a specific shell protein, the cognate pair theory will be supported. In addition, the binding pairs identified will be further studied to isolate a signal sequence or discover some other targeting mechanism to Eut BMCs. Following further research, Eut BMCs could be used as a tool for the spatial engineering of metabolic processes using interactions between Eut shell and cargo proteins to target proteins to Eut BMCs.

Materials and Methods

Phusion PCR

PCR was performed to amplify DNA of genes of interest. To perform PCR, a reaction volume of 50µL containing 10 µL 5X Phusion Buffer (NEB), 1 µL DMSO, 1 µL dNTPs, 1 µL template DNA, 1.25 µL of each primer (20 µM), 0.5 µL Phusion polymerase (NEB), and ddH_2O was made. The PCR reaction was carried out with an initial denaturing time of 30 s at 98 ˚C, followed by 30 cycles of a 10 s denaturing time at 98 ˚C, 30 s annealing time at 55 ˚C, and 30 s extension time per kb of amplified gene at 72 ˚C. The final extension time was 10 min at 72 ˚C.

Colony PCR

Colony PCR was performed to screen colonies from ligation plates for gene inserts of the correct size in the vectors. A small sample of a colony, 5 µL GoTaq® Green Master Mix (2X), 0.15 µL pBBinF (20 µM), 0.15 µL pBBinR (20 µM), and 4.7 ddH_2O were added to a PCR tube. Colony PCR was performed with an initial denaturation time at 98 ˚C for 10 min, followed by a 30 cycles of 30 s at 98 ˚C, 30 s at 55 ˚C, and 1 min/kb insert at 72 ˚C. The PCR reaction had a final extension time of 5 min at 72 ˚C.

Ligations

To perform ligations, 50-100 ng vector DNA was combined with insert DNA in a 3:1 molar ratio of insert:vector with 1 µL of T4 ligase buffer (New England Biolabs) per 10 µL reaction and 0.5 µL T4 ligase (New England Biolabs) per 10 µL ligation reaction volume. To allow DNA ligation, the ligation reactions were incubated at room temperature for 2-4 hours or overnight at 16 ˚C. To transform ligations, 2 µL of the ligation reaction was mixed with 25 µL

C2566 electrocompetent cells and placed into a 0.1 mm gap width electroporation cuvette. Electroporation was as in the transformation protocol, except 500 µL of LB media was immediately added to the electroporation cuvette and 250 µL of the electroporation cuvette contents were plated on an agar plate containing the appropriate antibiotic.

Transformations

To transform vector DNA into *E. coli* for expression, 0.5 µL vector miniprep and 25 µL electrocompetent C2566 cells were combined and placed in an 0.1 mm gap electroporation cuvette. The electroporation was performed using the Ec1 setting on a Micropulser Electroporator (BIO-RAD) and 1 mL of LB media was added to the electroporation cuvette. 20 µL of the cuvette contents were plated on an agar plate containing the appropriate antibiotic. Transformation plates were incubated overnight at 37 ˚C.

Cloning

Cloning was performed to make vectors encoding the bait and prey proteins used in the TALON Resin pull-downs of shell and cargo proteins. To amplify the gene of interest, PCR was performed with a Phusion polymerase (New England Biolabs) with *Salmonella enterica* genomic DNA as template and gene specific primers listed in table 1. If a 6xHistidine (H6) tag on the C-terminus of a protein was desired, the stop codon was modified to coding codon using a primer. The PCR products were gel purified using the Illustra GFX PCR DNA and gel band Purification kit (Illustra) according to the manufacturer's protocol. The purified PCR products were then digested using restriction enzymes according to the manufacturer's instructions (New England Biolabs). The vector was also digested according to the manufacturer's instructions (New

England Biolabs). The digested vector and insert were ligated using the ligation protocol. If a H6 tag was required, the PCR product was ligated into a vector encoding the 6xHis-tag followed by a stop codon. The ligation was then transformed via electroporation into *E. coli* C2566 electrocompetent cells and transformed cells were plated on agar plates containing the appropriate antibiotic and incubated at 37˚C overnight. Transformants were screened for the gene of interest in the vector via colony PCR. Positive transformants were selected for inoculation of 4 mL overnight cultures. The plasmids from each 4 mL were miniprepped using the Wizard® Plus Minipreps DNA Purification System according to the manufacturer's protocol (Promega). Sequencing of the vectors using the miniprep as a template was used to confirm the success of the cloning.

Expression Studies

Expression of recombinant proteins were tested using several different methods. To test total expression of proteins, 1 mL of an overnight culture in LB was pelleted, resuspended in about 250 µL of water. These samples were then run on an SDS-PAGE gel. To test total and soluble expression of proteins, 4 mL of an overnight culture in LB was centrifuged for 20 min at 4000 rpm and resuspended in 900 µL Buffer A (TALON resin pull-downs) and 100µL 10 mM PMSF in isopropanol. Then the sample was sonicated at 30% power, for 1 minute, 1 second on, 2 seconds off. The total fraction sample was taken after sonication. The remaining protein sample was centrifuged at 15000 rpm for 10 min at 4 ˚C. The soluble fraction sample was taken at this time.

Sequencing

Sequencing was performed on pUCBB plasmids to confirm that cloning was successful. To perform sequencing, 6.4 uL of 1 μM primer, 500 ng of miniprep, and ddH_2O were combined to make a reaction volume of 6.4 μL. The sequencing reactions were performed by the BioMedical Genomics Center.

TALON Resin Pull-downs

TALON Resin Pull-downs were performed with a 6xHis-tagged bait protein and an untagged prey protein. The plasmid vectors encoding the bait protein and prey protein were each transformed via electroporation, plated on LB containing the appropriate antibiotic, and incubated overnight at 30 ˚C. A culture tube containing 4 mL LB and the appropriate antibiotic was inoculated with a colony on the transformation plate for both the prey and the bait expressing cells and incubated overnight in a shaker at 30 ˚C. 50 μL of the 4 mL culture was then used to inoculate a 50 mL LB culture containing the appropriate antibiotic, which was incubated overnight in a shaker at 30 ˚C. The 50 mL cultures were then spun down at 4,000 rpm for 20 minutes at 4 ˚C to collect the cells expressing the proteins. The excess media was removed and the cells were resuspended in 9 mL Buffer A (50 mM Tris, 250 mM NaCl, 5 mM imidazole, pH 7.5) and 1 mL 10 mM PMSF in isopropanol. The cells were lysed by sonication at 30% power, 1 second on, 2 seconds off, for a total of 4 minutes. The "total fraction" sample of the bait or prey was collected after this step. Centrifugation at 12,000 rpm for 30 minutes at 4 ˚C was used to pellet the insoluble fraction of the cell lysate. A sample of the "soluble fraction" of the protein was collected from the supernatant of the centrifuged total lysate at this time. The remaining soluble bait protein was added to a 15 mL conical tube containing equilibrated TALON Resin. To equilibrate the TALON Resin, 2 mL of TALON Resin and 10 mL Buffer A

was used per the manufacturers equilibration protocol. To bind the bait to the TALON Resin, the TALON resin gently shaken at 4 ˚C for 1 hour or overnight.

After bait binding, the TALON resin was pelleted by centrifugation at 700 g for 5 minutes. A sample of the "unbound bait protein" was collected from the supernatant, and the supernatant was discarded. The TALON resin and bound bait protein were then washed twice. To wash the TALON resin and bound bait protein, the TALON resin and bound protein were resuspended in 10 mL of Buffer A, centrifuged for 2 minutes at 700 g, and the supernatant was discarded. The wash step was repeated and a "bait wash" sample was collected from the supernatant before it was discarded. To facilitate binding between the bait protein bound to the TALON Resin and the prey protein, the soluble fraction of the prey protein was added to the TALON Resin bound to the bait protein and gently shaken for 1 hour at 4 ˚C. The TALON Resin was then centrifuged for 5 minutes at 700 g to pellet the TALON Resin and bound proteins. The "unbound prey protein" sample was collected from the supernatant, which was then discarded. In a manner identical to the TALON Resin with bound bait protein, the TALON resin with bound bait and prey protein was washed twice. Following the second wash, a sample of the supernatant was collected as the "prey wash" sample prior to its disposal.

The TALON Resin and bound proteins were washed twice similarly to the bound bait protein wash, and a sample of the prey wash fraction was collected from the supernatant of the repeat wash step before the supernatant was discarded. Following the prey wash, the TALON resin with bound proteins was resuspended in 5 mL Buffer and applied to a gravity flow column. After the TALON Resin settled to the bottom of the column, the column was opened and the Buffer was allowed to flow through. Protein was eluted off of the TALON Resin using Buffer B (Buffer A with 50 μM imidazole, pH 7.5) and the elution was collected and saved as the "50 μM

elution fraction" sample. The remaining protein on the TALON resin was then eluted with Buffer C (Buffer A with 250 μM imidazole, pH 7.5) and the flowthrough was collected and saved as the "250 μM elution fraction" sample.

SDS-PAGE gels

To visualize the pull-down results or protein expression, SDS-PAGE gels of the protein samples were performed. To prepare the protein samples accumulated during each TALON Resin pull-down, an equal volume of SDS solution was added to each sample. The samples were then boiled for 10-15 minutes to denature the proteins. Fifteen μL of each sample and 5 μL protein ladder was run on 12% acrylamide SDS-PAGE gels at 240 V for 30 min. The gels were then washed in ddH_2O 3 times and then stained with Bio-Safe™ Coomassie (BIO-RAD) for 20 minutes to 1 hour. The gels were destained in ddH_2O for 30 minutes to overnight.

Results

Cargo protein cloning

To overexpress Eut proteins for the pull-downs, the Eut cargo protein genes were cloned into the pUCBB vector using the *Nde*I and *Xho*I restriction sites. For each gene, cloning was carried out using NdeI and XhoI restriction enzymes and the corresponding *Nde*I and *Xho*I primers listed in **Table 1**. *eutA, eutB, eutC, eutD, eutE, eutG, eutH, and eutT* were all successfully cloned into pUCBB using the *Nde*I and *Xho*I restriction sites. After expression studies were performed, several Eut cargo protein genes were cloned into pUCBB using the *Bam*HI and *Xho*I restriction sites. For each gene, cloning was carried out using the *Bam*HI and *Xho*I restriction enzymes and the corresponding *Bam*HI primer and the pBBinR primer listed in **Table 1**. *eutA, eutC, eutE, and eutH* were successfully cloned into pUCBB using the *Bam*HI and *Nde*I restriction sites. **Table 2** displays the plasmids used in this study.

EutC$^{1-19}$eGFP C-terminus 6xHistidine tag addition

To convert EutC$^{1-19}$eGFP to a bait protein for pull-downs, a 6xHistidine (H6) tag was added to the C-terminus of the gene. This was done using a pUCBB encoding a C-terminal H6 tag (CtH6), as the vector. Cloning was performed using *Nde*I and *Xho*I restriction sites and the primers for EutC$^{1-19}$eGFP listed on **Table 1**. pUCBB-EutC$^{1-19}$eGFP-CtH6 was successfully cloned using this method.

EutS-NtH6 does not bind to TALON Resin

To identify Eut cargo proteins that interact with EutS, TALON Resin pull-downs were performed using a EutS protein with an H6 tag on the N-terminus (EutS-NtH6), as bait. EutS-

NtH6 was expressed, but did not bind to the TALON Resin (**Figure 4**). Because EutS-NtH6 contains a H6 tag with high affinity for TALON Resin, EutS-NtH6 should bind to the TALON Resin used in the pull-downs. A band around 12 kDa in the soluble fraction of the cell lysate shows that EutS-NtH6 was expressed (**Figure 4**). A band representing EutS-NtH6 was visible in the unbound protein after binding to the TALON Resin (**Figure 4**). Furthermore, a band representing EutS-NtH6 was absent in the fractions collected during 50 μM and 250 μM imidazole elutions, which elute protein off of the TALON Resin (**Figure 4**). The EutS-NtH6 band in the unbound bait fraction and the absence of a band representing EutS-NtH6 in the elution fractions are consistent with the expected results if EutS-NtH6 did not bind to the TALON Resin. Since EutS-NtH6 does not bind to the TALON Resin in a detectable range, it is unsuitable for use as bait in pull-downs with TALON Resin.

EutS-CtH6 is suitable for TALON Resin pull-downs but does not interact with EutC$^{1-19}$eGFP

To identify Eut cargo proteins that interact with EutS, TALON Resin pull-downs were performed using a EutS protein with an H6 tag on the C-terminus (EutS-CtH6), as bait. EutS-CtH6 expressed and bound to TALON Resin (**Figure 5**). Because EutS-CtH6 contains a H6 tag with high affinity for TALON Resin, EutS-CtH6 should bind to the TALON Resin used in the pull-downs. A band around 12 kDa in the soluble fraction of the cell lysate shows that EutS-CtH6 was expressed (**Figure 5**). A band representing EutS-CtH6 was not visible in the unbound protein after binding to the TALON Resin (**Figure 5**). Furthermore, a band representing EutS-CtH6 was present in the fractions collected during the 250 μM imidazole elution (**Figure 5**).

EutC$^{1-19}$eGFP was used as prey while EutS-CtH6 was used as bait in a TALON Resin pull-down to test for protein interaction. EutC$^{1-19}$eGFP has been shown to localize in BMC

shells consisting of only EutS *in vivo*, so an interaction between these two proteins on a TALON Resin pull-down could act as a positive control (Choudhary et al., 2012). EutC$^{1-19}$eGFP expressed in the prey soluble fraction, and it also appeared in the prey wash fraction (**Figure 5**). EutS-CtH6 bound to the TALON Resin as well (**Figure 5**). However, no band near the size of EutC$^{1-19}$eGFP (26 kDa) was observed in the elution lanes (**Figure 5**). Although an interaction between EutC$^{1-19}$eGFP was expected, it was not observed. It is possible the H6 tag on EutS prevented EutS binding to EutC$^{1-19}$eGFP, since the C-terminal region is known to be important for EutC$^{1-19}$eGFP localization *in vivo* (Choudhary et al., 2012).

EutC$^{1-19}$eGFP-CtH6 does not express

A protein interaction between EutS-CtH6 and EutC$^{1-19}$eGFP was not observed; we still lacked a positive control for the pull-down interaction. A second attempt was made to observe an interaction between EutS and EutC$^{1-19}$eGFP again by reversing their roles; EutC$^{1-19}$eGFP-CtH6 was used as bait and EutS was used as prey. However, the pUCBB-EutC$^{1-19}$eGFP-CtH6 vector did not express EutC$^{1-19}$eGFP-CtH6. EutC$^{1-19}$eGFP was not visible in the bait total, bait soluble, and imidazole elution fractions in a TALON Resin pull-down (**Figure 6**). EutS expressed as soluble protein from the pUCBB-EutS vector (data not shown). In addition to a TALON Resin pull-down run with EutS as prey, a pull-down with EutM as prey and EutC$^{1-19}$eGFP-CtH6 was also performed (**Figure 6**). EutC$^{1-19}$eGFP-CtH6 was not expressed in the pull-down with EutM, confirming the absence of EutC$^{1-19}$eGFP-CtH6 expression observed in the EutC$^{1-19}$eGFP-CtH6 pull-down with EutS. EutM expressed as soluble protein from the pUCBB-EutM vector (**Figure 6**). Since EutC$^{1-19}$eGFP-CtH6 was not expressed, it could not be used as bait in TALON Resin pull-downs.

Cargo Protein Expression

Protein expression assays were performed on each of the constructed vectors encoding a Eut cargo protein to optimize expression for use in TALON Resin pull-down assays. pUCBB-EutB (**Figure 7**), pUCBB-EutT (**Figure 7**), and pUCBB-EutG (**Figure 11**) plasmids using the *Nde*I and *Xho*I restriction sites for cloning solubly expressed the protein of interest. Plasmids failing to express the protein of interest when the gene insert was cloned into the *Nde*I and *Xho*I sites in the pUCBB vector included pUCBB-EutA (**Figure 7**), pUCBB-EutC (**Figure 10**), pUCBB-EutD (**Figure 7**), pUCBB-EutE (**Figure 7**), and pUCBB-EutH (**Figure 9**). A summary of protein expression results can be found in **Table 3**.

To express the cargo proteins that were not expressed from the pUCBB vector in the *Nde*I and *Xho*I sites, several different protocols were explored. Cloning the Eut cargo protein genes into pACBB using the *Nde*I and *Xho*I sites failed to increase expression (data not shown). Cloning the Eut cargo protein genes into the inducible vector pET-21b also did not produce protein expression (data not shown). Growing cultures containing pUCBB encoding a Eut cargo protein cloned into the *Nde*I and *Xho*I sites at different temperatures and for different amounts of time also failed to result in Eut cargo expression (data not shown). The final attempt at expression involved cloning the Eut cargo genes into the *Bam*HI and the *Xho*I site on the pUCBB vector. EutA, EutC, and EutE cargo proteins expressed when cloned into the *Bam*HI and *Xho*I sites in the pUCBB vector (**Figure 10**). However, EutA and EutE cloned into *Nde*I and *Xho*I sites expressed as well in this experiment, in contrast to the results in figure 7.

Discussion

BMCs are a promising option for spatial engineering involving the encapsulation of all related metabolic enzymes in a protein shell that can regulate the transport of substrates and products (Held et al., 2013). To further develop BMCs as a tool, the process of enzyme encapsulation within BMCs, among other things, must be understood. Recently, a BMC shell was expressed heterologously using only one protein, EutS, demonstrating the ability to engineer BMCs into an industrial host, *E. coli* (Choudhary et al., 2012). Previous work has identified one signal sequence that targets a cargo for Eut BMC encapsulation; however, the major mechanism of enzyme targeting to BMC shells has been debated in the literature (Fan et al., 2010; Kinney et al., 2012). The two conflicting hypotheses for BMC targeting include the cognate pair mechanism and the general binding mechanism. The goal of this work was to identify shell-cargo protein interactions within Eut BMCs to elucidate the targeting mechanism. Pull-downs were used to investigate protein interactions. Though no successful protein interactions were observed, insight into pull-down assays using Eut proteins was gained. In addition, many plasmids encoding Eut genes were constructed and added to the Schmidt-Dannert plasmid collection, and expression studies were performed on many Eut cargo proteins for use in pull-downs.

Understanding the BMC targeting mechanism of Eut cargo proteins is an important step for the development of BMCs as a metabolic engineering tool. To explore Eut cargo-shell protein interactions, pull-downs were performed, similar to *in vitro* pull-down interactions between shell and cargo previously performed using Pdu BMC proteins (Fan et al., 2012). EutS-NtH6 did not bind to the TALON resin. This could possibly be a result of EutS protein conformation; the NtH6 may not be exposed, and therefore unable to bind to the column. EutS-

CtH6 was expressed and did bind to the TALON Resin; however, it did not interact with EutC$^{1-19}$eGFP, which has been shown to localize to EutS BMC shells in vivo. There are several possible reasons for this failed interaction. The first is that the CtH6 on EutS interferes with the interaction between the residues on EutS that bind with the EutC$^{1-19}$ residues on GFP (Choudhary et al., 2012). The second is that the encapsulation of cargo proteins within Eut BMCs takes place during a very specific time frame, which is not captured in the in vitro pull-downs.

Expression studies of Eut cargo proteins were important, as the pull-downs assays required proteins that expressed solubly. However, many options for protein expression were explored before a method that consistently expressed Eut proteins was discovered. A method was identified to strongly express each of the following proteins: EutA, EutB, EutC, EutE, EutG, and EutT. However, EutD and EutH expression has not been optimized. EutA, EutC, and EutE are expressed using *Bam*HI and *Xho*I cloning sites on the pUCBB vector. EutB, EutG, and EutT have been expressed using the *Nde*I and *Xho*I sites on pUCBB. The lack of protein expression of several Eut cargo proteins, including EutA, EutC, and EutE, was observed in pUCBB vectors using the *Nde*I and *Xho*I cloning sites multiple times (**Figure 4** and additional data not shown). However, expression of EutA and EutE from the *Nde*I and *Xho*I cloning sites in pUCBB was observed in figure 10, contrasting the results in figure 7. One major difference between the experiment run in figure 10 and previous experiments is the growth time. Cultures expressing EutA and EutE from the *Nde*I and *Xho*I sites on pUCBB were grown for 24 hours, while cultures used for earlier studies had been grown for only 16 hours. The time dependence of EutA and EutE expression from the *Nde*I and *Xho*I sites on pUCBB remains to be explored. The reason for the lack of protein expression from EutC, and possibly EutA and EutE, cloned into the *Nde*I and *Xho*I sites of pUCBB may be due to some type of DNA secondary structure, though that

hypothesis has not yet been tested. In addition of expression studies, solubility studies were also performed on Eut cargo proteins. EutB, EutG, and EutT were found to express solubly. Soluble expression studies remain to be performed on EutA, EutC, and EutE. Understanding the conditions necessary to solubly express Eut proteins is necessary to perform pull-downs. Although not all of the Eut cargo proteins were successfully expressed solubly, many were successfully expressed, with several expressed solubly, and may be used in pull-downs to identify protein interactions.

Though the original goals of this experiment were not achieved, knowledge of Eut protein expression and which proteins cannot be used as bait in pull-downs of Eut BMC proteins will be helpful to anyone attempting to perform Eut shell-cargo protein pull-downs in the future. The Eut protein expression studies may be helpful in a number of other studies, including Eut cargo protein function or structural studies, or other possible Eut shell-cargo protein interaction studies, among other things. Eut BMCs have the potential to greatly improve metabolic engineered pathway capabilities, however the fundamental biology of Eut BMCs must be understood before they can be harnessed by metabolic engineering. Future studies continuing Eut BMC research could result in a powerful new metabolic engineering tool.

References

Agapakis, C.M., Boyle, P.M., and Silver, P.A. (2012). Natural strategies for the spatial optimization of metabolism in synthetic biology. Nat. Chem. Biol. *8*, 527–535.

Bar-Even, A., and Salah Tawfik, D. (2013). Engineering specialized metabolic pathways—is there a room for enzyme improvements? Curr. Opin. Biotechnol. *24*, 310–319.

Bayer, T.S., Widmaier, D.M., Temme, K., Mirsky, E.A., Santi, D.V., and Voigt, C.A. (2009). Synthesis of methyl halides from biomass using engineered microbes. J. Am. Chem. Soc. *131*, 6508–6515.

Bonacci, W., Teng, P.K., Afonso, B., Niederholtmeyer, H., Grob, P., Silver, P.A., and Savage, D.F. (2012). Modularity of a carbon-fixing protein organelle. PNAS *109*, 478–483.

Boyle, P.M., and Silver, P.A. (2012). Parts plus pipes: Synthetic biology approaches to metabolic engineering. Metab. Eng. *14*, 223–232.

Brinsmade, S.R., Paldon, T., and Escalante-Semerena, J.C. (2005). Minimal Functions and Physiological conditions required for growth of Salmonella enterica on ethanolamine in the absence of the metabolosome. J. Bacteriol. *187*, 8039–8046.

Buan, N.R., Suh, S.-J., and Escalante-Semerena, J.C. (2004). The eutT gene of Salmonella enterica encodes an oxygen-labile, metal-containing ATP:corrinoid adenosyltransferase enzyme. J. Bacteriol. *186*, 5708–5714.

Choudhary, S., Quin, M.B., Sanders, M.A., Johnson, E.T., and Schmidt-Dannert, C. (2012). Engineered protein nano-compartments for targeted enzyme localization. PLoS ONE *7*, e33342.

Crowley, C.S., Cascio, D., Sawaya, M.R., Kopstein, J.S., Bobik, T.A., and Yeates, T.O. (2010). Structural insight into the mechanisms of transport across the Salmonella enterica Pdu microcompartment shell. J. Biol. Chem. *285*, 37838–37846.

Delebecque, C.J., Lindner, A.B., Silver, P.A., and Aldaye, F.A. (2011). Organization of intracellular reactions with rationally designed RNA assemblies. Science *333*, 470–474.

Dueber, J.E., Wu, G.C., Malmirchegini, G.R., Moon, T.S., Petzold, C.J., Ullal, A.V., Prather, K.L.J., and Keasling, J.D. (2009). Synthetic protein scaffolds provide modular control over metabolic flux. Nat. Biotech. *27*, 753–759.

Fan, C., Cheng, S., Liu, Y., Escobar, C.M., Crowley, C.S., Jefferson, R.E., Yeates, T.O., and Bobik, T.A. (2010). Short N-terminal sequences package proteins into bacterial microcompartments. PNAS *107*, 7509–7514.

Fan, C., Cheng, S., Sinha, S., and Bobik, T.A. (2012). Interactions between the termini of lumen enzymes and shell proteins mediate enzyme encapsulation into bacterial microcompartments. PNAS *109*, 14995–15000.

Farhi, M., Marhevka, E., Ben-Ari, J., Algamas-Dimantov, A., Liang, Z., Zeevi, V., Edelbaum, O., Spitzer-Rimon, B., Abeliovich, H., Schwartz, B., et al. (2011). Generation of the potent anti-malarial drug artemisinin in tobacco. Nat. Biotech. *29*, 1072–1074.

Faust, L.R., Connor, J.A., Roof, D.M., Hoch, J.A., and Babior, B.M. (1990). Cloning, sequencing, and expression of the genes encoding the adenosylcobalamin-dependent ethanolamine ammonia-lyase of Salmonella typhimurium. J. Biol. Chem. *265*, 12462–12466.

Fay, P. (1992). Oxygen relations of nitrogen fixation in cyanobacteria. Microbiol. Rev. *56*, 340–373.

Frank, S., Lawrence, A.D., Prentice, M.B., and Warren, M.J. (2012). Bacterial microcompartments moving into a synthetic biological world. J. Biotechnol. *163,* 273-279.

Havemann, G.D., Sampson, E.M., and Bobik, T.A. (2002). 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. *184*, 1253–1261.

Held, M., Quin, M.B., and Schmidt-Dannert, C. (2013). Eut bacterial microcompartments: insights into their function, structure, and bioengineering applications. J. Mol. Microbiol. Biotechnol. *23*, 308–320.

Huang, X., Holden, H.M., and Raushel, F.M. (2001). Channeling of substrates and intermediates in enzyme-catalyzed reactions. Annu. Rev. Biochem. *70*, 149–180.

Keenan, R.J., Freymann, D.M., Stroud, R.M., and Walter, P. (2001). The signal recognition particle. Annu. Rev. Biochem. *70*, 755–775.

Kerfeld, C.A., Heinhorst, S., and Cannon, G.C. (2010). Bacterial microcompartments. Annu. Rev. Microbiol. *64*, 391–408.

Kim, H.J., Du, W., and Ismagilov, R.F. (2011). Complex function by design using spatially pre-structured synthetic microbial communities: degradation of pentachlorophenol in the presence of Hg(II). Integr. Biol. *3*, 126–133.

Kinney, J.N., Salmeen, A., Cai, F., and Kerfeld, C.A. (2012). Elucidating essential role of conserved carboxysomal protein CcmN reveals common feature of bacterial microcompartment assembly. J. Biol. Chem. *287*, 17729–17736.

Klein, M.G., Zwart, P., Bagby, S.C., Cai, F., Chisholm, S.W., Heinhorst, S., Cannon, G.C., and Kerfeld, C.A. (2009). Identification and structural analysis of a novel carboxysome shell protein with implications for metabolite transport. J. Mol. Biol. *392*, 319–333.

Lawrence, A.D., Frank, S., Newnham, S., Lee, M.J., Brown, I.R., Xue, W.-F., Rowe, M.L., Mulvihill, D.P., Prentice, M.B., Howard, M.J., et al. (2014). Solution structure of a bacterial microcompartment targeting peptide and its application in the construction of an ethanol bioreactor. ACS Synth. Biol. *3*, 454–465.

Lee, H., DeLoache, W.C., and Dueber, J.E. (2012). Spatial organization of enzymes for metabolic engineering. Metab. Eng. *14*, 242–251.

Luzio, J.P., Poupon, V., Lindsay, M.R., Mullock, B.M., Piper, R.C., and Pryor, P.R. (2003). Membrane dynamics and the biogenesis of lysosomes (Review). Mol. Membr. Biol. *20*, 141–154.

Parsons, J.B., Frank, S., Bhella, D., Liang, M., Prentice, M.B., Mulvihill, D.P., and Warren, M.J. (2010). Synthesis of empty bacterial microcompartments, directed organelle protein incorporation, and evidence of filament-associated organelle movement. Mol. Cell *38*, 305–315.

Penrod, J.T., Mace, C.C., and Roth, J.R. (2004). A pH-sensitive function and phenotype: Evidence that EutH facilitates diffusion of uncharged ethanolamine in Salmonella enterica. J. Bacteriol. *186*, 6885–6890.

Penrod, J.T., and Roth, J.R. (2006). Conserving a volatile metabolite: a role for carboxysome-like organelles in Salmonella enterica. J. Bacteriol. *188*, 2865–2874.

Peplow, M. (2013). Malaria drug made in yeast causes market ferment. Nature *494*, 160–161.

Ro, D.-K., Paradise, E.M., Ouellet, M., Fisher, K.J., Newman, K.L., Ndungu, J.M., Ho, K.A., Eachus, R.A., Ham, T.S., Kirby, J., et al. (2006). Production of the antimalarial drug precursor artemisinic acid in engineered yeast. Nature *440*, 940–943.

Sagermann, M., Ohtaki, A., and Nikolakakis, K. (2009). Crystal structure of the EutL shell protein of the ethanolamine ammonia lyase microcompartment. PNAS *106*, 8883–8887.

Sampson, E.M., and Bobik, T.A. (2008). Microcompartments for B12-dependent 1,2-propanediol degradation provide protection from DNA and cellular damage by a reactive metabolic intermediate. J. Bacteriol. *190*, 2966–2971.

Shively, J.M., Ball, F., Brown, D.H., and Saunders, R.E. (1973). functional organelles in prokaryotes: Polyhedral inclusions (carboxysomes) of Thiobacillus neapolitanus. Science *182*, 584–586.

Starai, V.J., Garrity, J., and Escalante-Semerena, J.C. (2005). Acetate excretion during growth of Salmonella enterica on ethanolamine requires phosphotransacetylase (EutD) activity, and acetate recapture requires acetyl-CoA synthetase (Acs) and phosphotransacetylase (Pta) activities. Microbiology *151*, 3793–3801.

Stojiljkovic, I., Bäumler, A.J., and Heffron, F. (1995). Ethanolamine utilization in Salmonella typhimurium: nucleotide sequence, protein expression, and mutational analysis of the cchA cchB eutE eutJ eutG eutH gene cluster. J. Bacteriol. *177*, 1357–1366.

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

Tanaka, S., Sawaya, M.R., and Yeates, T.O. (2010). Structure and mechanisms of a protein-based organelle in Escherichia coli. Science *327*, 81–84.

Tanaka, S., Sawaya, M.R., Phillips, M., and Yeates, T.O. (2009). Insights from multiple structures of the shell proteins from the β-carboxysome. Prot. Sci. *18*, 108–120.

Westfall, P.J., Pitera, D.J., Lenihan, J.R., Eng, D., Woolard, F.X., Regentin, R., Horning, T., Tsuruta, H., Melis, D.J., Owens, A., et al. (2012). Production of amorphadiene in yeast, and its conversion to dihydroartemisinic acid, precursor to the antimalarial agent artemisinin. PNAS *109*, E111–E118.

Wintermute, E.H., and Silver, P.A. (2010). Dynamics in the mixed microbial concourse. Genes Dev. *24*, 2603–2614.

Woolston, B.M., Edgar, S., and Stephanopoulos, G. (2013). Metabolic engineering: Past and future. Annu. Rev. Chem. Biomol. Eng. *4*, 259–288.

Yeates, T.O., Crowley, C.S., and Tanaka, S. (2010). Bacterial microcompartment organelles: Protein shell structure and evolution. Annu. Rev. Biophys. *39*, 185–205.

Yeates, T.O., Thompson, M.C., and Bobik, T.A. (2011). The protein shells of bacterial microcompartment organelles. Curr. Opin. Struct. Biol. *21*, 223–231.

Yeates, T.O., Tsai, Y., Tanaka, S., Sawaya, M.R., and Kerfeld, C.A. (2007). Self-assembly in the carboxysome: a viral capsid-like protein shell in bacterial cells. Biochem. Soc. Trans. *35*, 508–511.

Figures and Tables

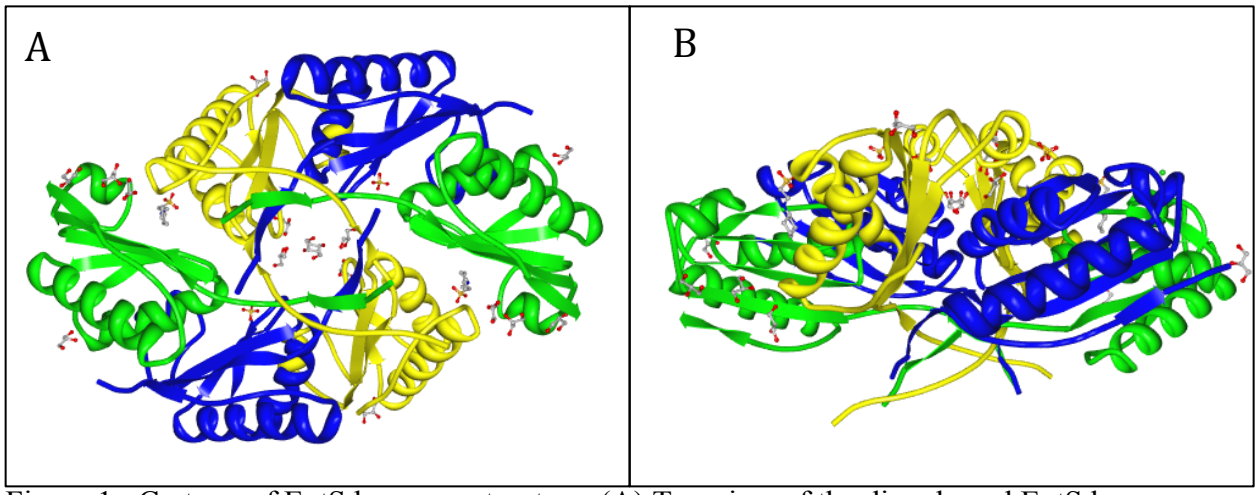

Figure 1: Cartoon of EutS hexamer structure. (A) Top view of the disc shaped EutS hexamer with its central pore (B) Side view of EutS hexamer. The different hexamer EutS monomers are shown in alternating colors. Models were generated with Pymol using the published EutS structure (PDB #3I96).

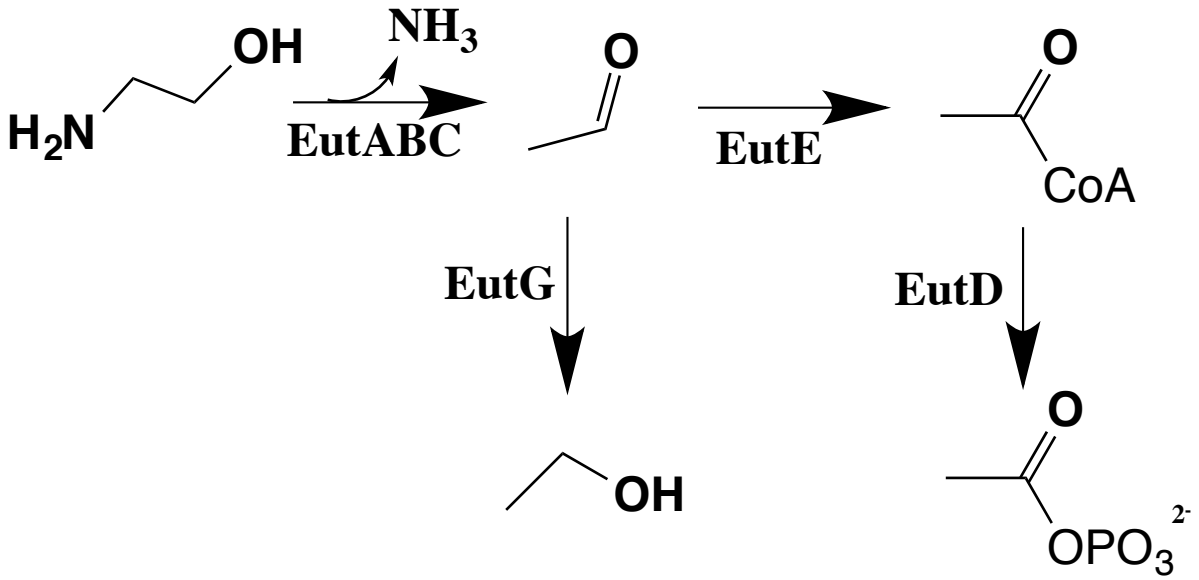

Figure 2. Ethanolamine metabolism in *Salmonella* with the enzyme/enzymes that catalyze the reaction step.

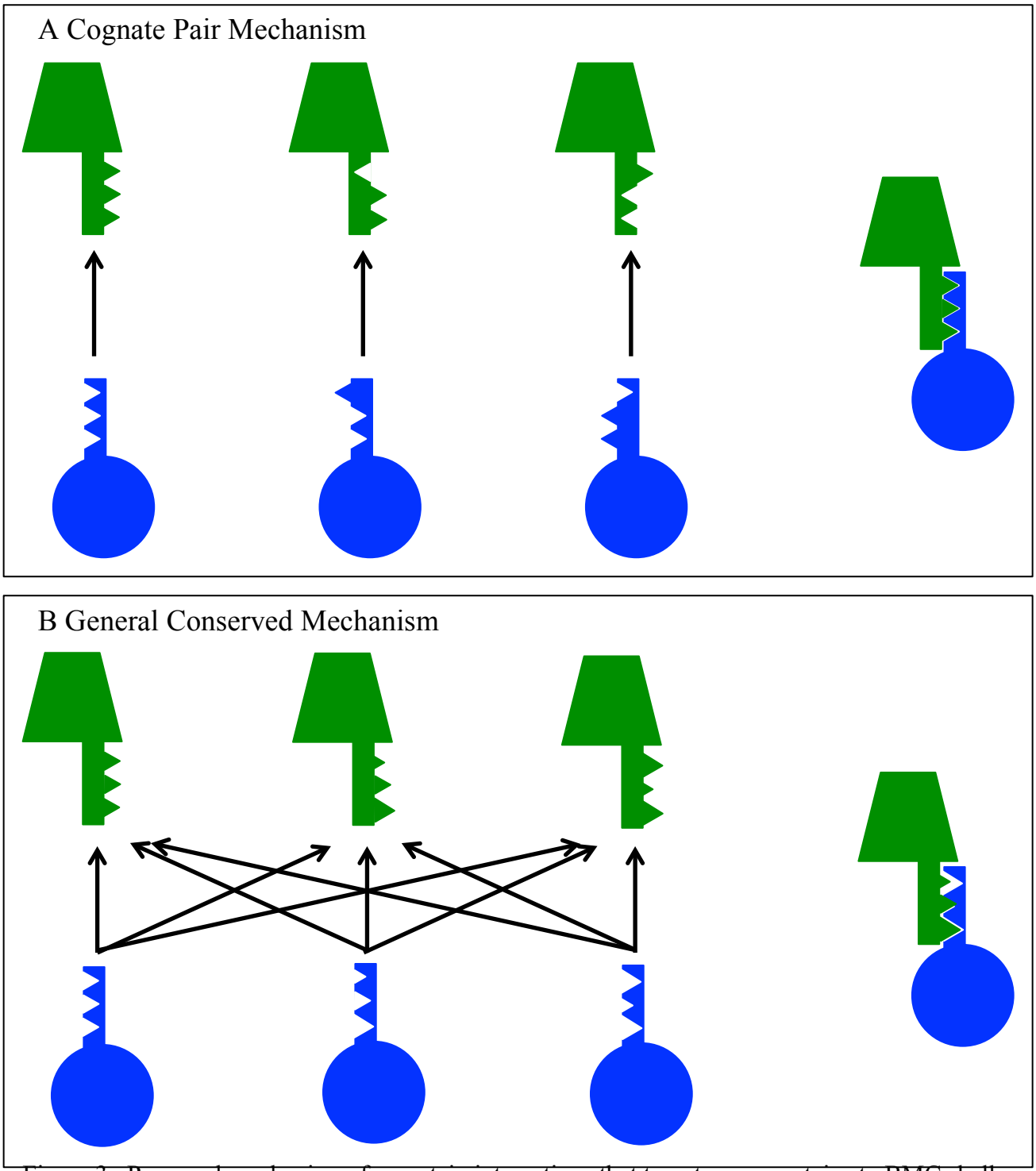

Figure 3. Proposed mechanisms for protein interactions that target cargo proteins to BMC shells. Shell proteins are in green and cargo proteins are in blue. In each figure on the right shows how a shell and cargo protein would interact. The arrows between proteins represent proteins that interact with each other. **A.** A hypothetical cognate pair mechanism between shell and cargo proteins. **B.** A hypothetical general conserved mechanism between shell and cargo proteins.

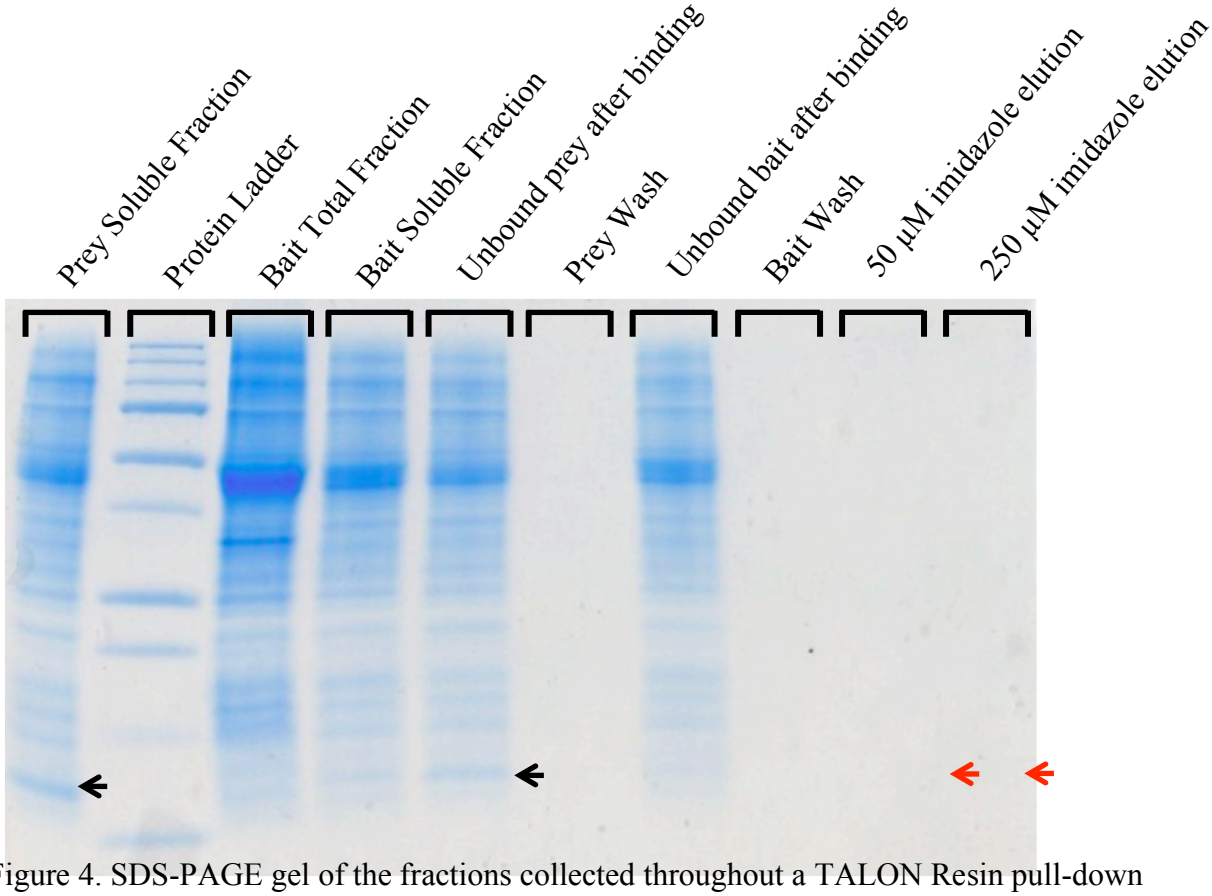

Figure 4. SDS-PAGE gel of the fractions collected throughout a TALON Resin pull-down with EutS-NtH6 as bait and EutC as prey. The fraction names are displayed above each of the lanes. The protein was visualized using BioSafe™ Coomassie dye. Precision Plus Protein™ All Blue Protein Standard from Bio-Rad serves as the protein ladder. The black arrows point to bands representing EutS-NtH6 (13 kDa). The red arrows point to the location where a band would be expected if EutS-NtH6 bound to the TALON Resin. The prey, EutC, is not expressed.

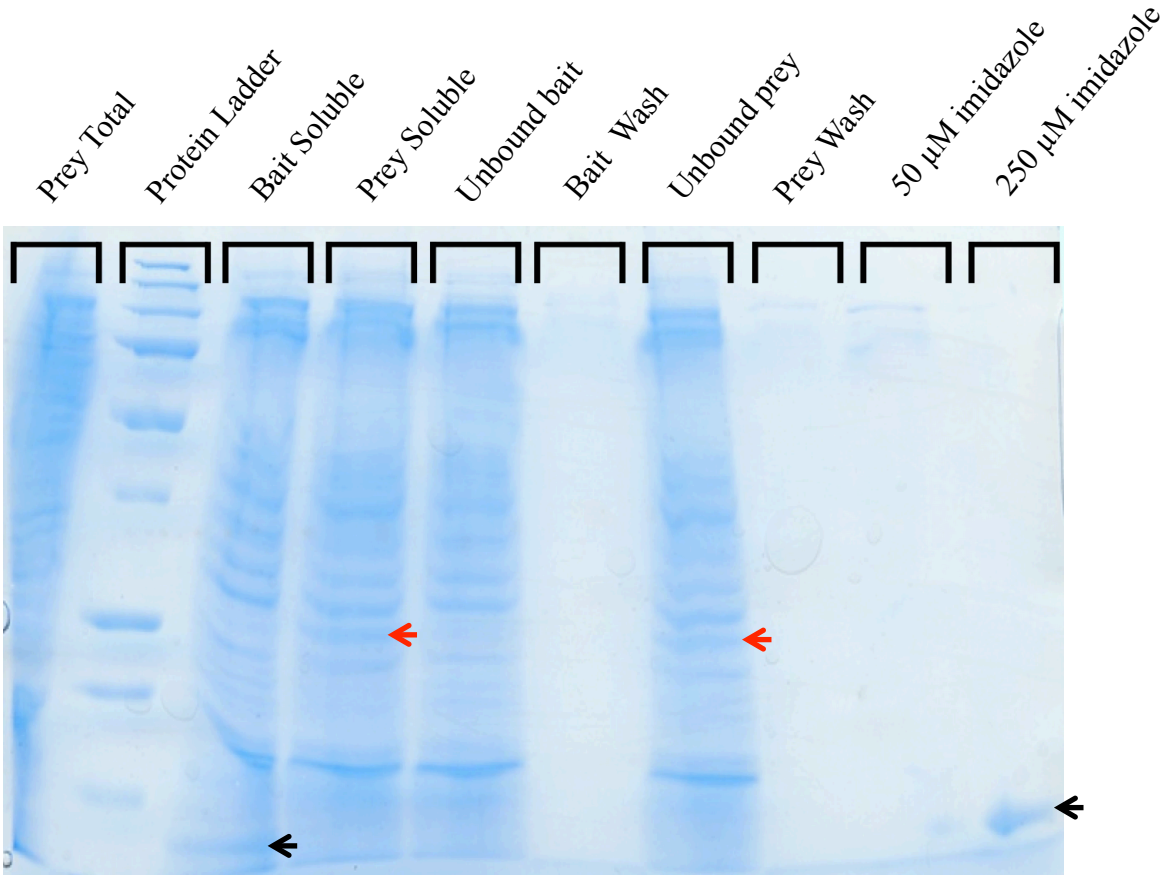

Figure 5. SDS-PAGE gel of the fractions collected throughout a TALON Resin pull-down with EutS-CtH6 as bait and EutC$^{1-19}$eGFP (26 kDa) as prey. The fraction names are displayed above each of the lanes. The protein was visualized using BioSafe™ Coomassie dye. Precision Plus Protein™ All Blue Protein Standard from Bio-Rad serves as the protein ladder. The black arrows point to bands representing EutS-CtH6 (13 kDa). The red arrows point to bands representing EutC$^{1-19}$eGFP.

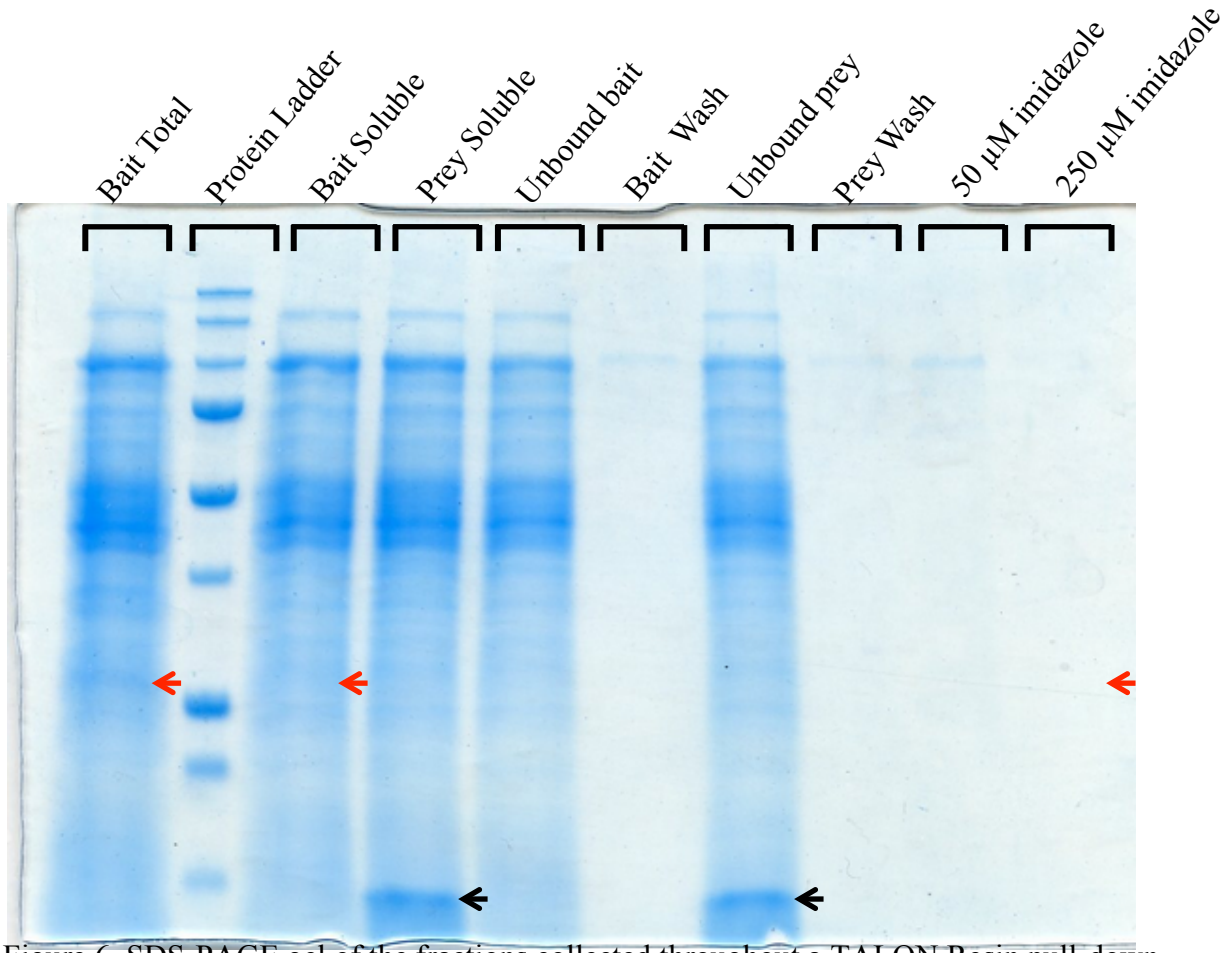

Figure 6. SDS-PAGE gel of the fractions collected throughout a TALON Resin pull-down with EutC$^{1-19}$eGFP-CtH6 as bait and EutM as prey. The fraction names are displayed above each of the lanes. The protein was visualized using BioSafe™ Coomassie dye. Precision Plus Protein™ All Blue Protein Standard from Bio-Rad serves as the protein ladder. The black arrows point to bands representing EutM (10 kDa). The red arrows point to the location where a band would be expected if EutC$^{1-19}$eGFP (29 kDa) expressed (arrows on the left) and bound to the TALON Resin (arrow on the right).

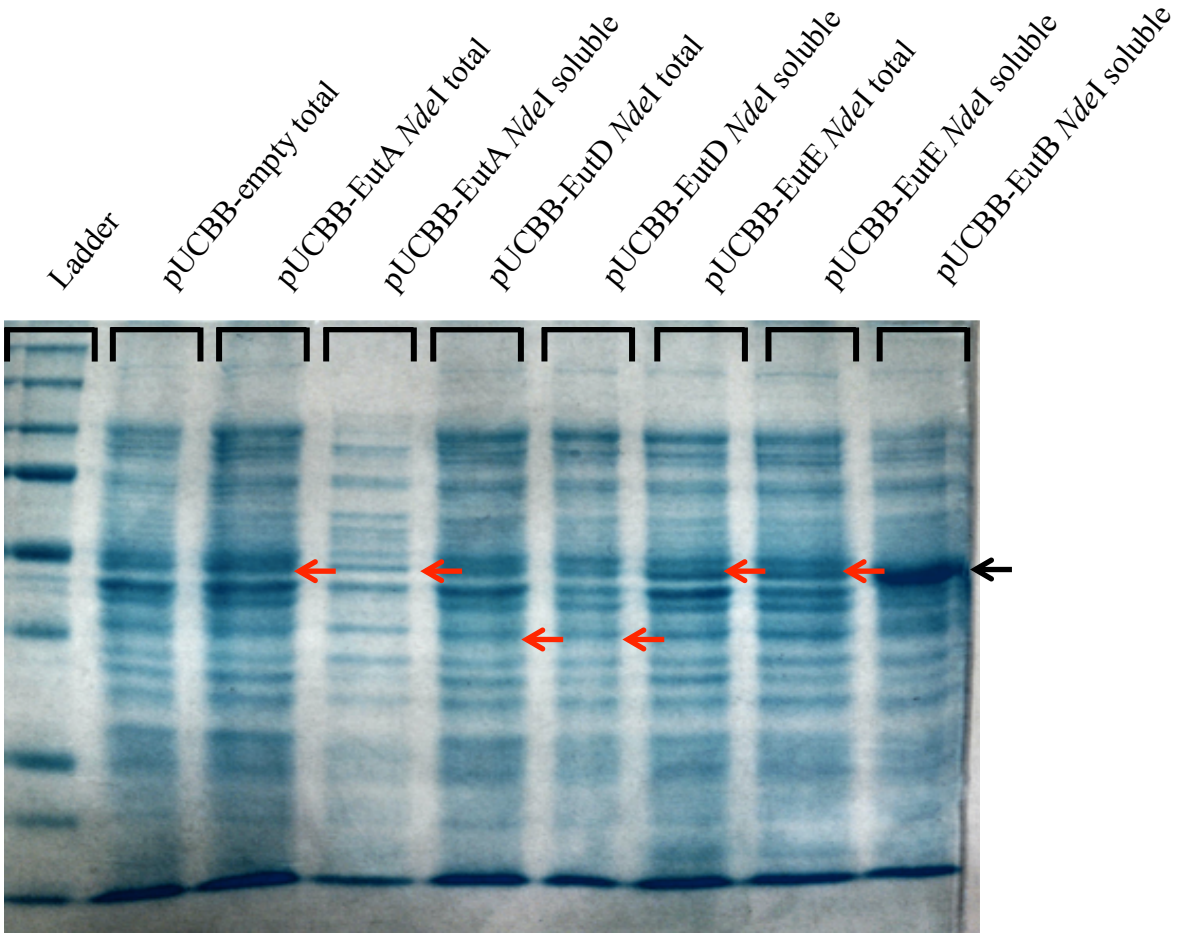

Figure 7. Results of Eut cargo protein expression studies. The protein was visualized using BioSafe™ Coomassie dye. Precision Plus Protein™ All Blue Protein Standard from Bio-Rad acts as the protein ladder. pUCBB-empty was used as a negative control. The black arrow points to the band representing EutB (49 kDa). The red arrows point to where a band would be expected if the Eut cargo protein expressed. EutA (49.5 kDa), EutD (37 kDa), and EutE (49 kDa) were not expressed. The cultures used for this gel were grown at 30 ˚C for 16 hours. pUCBB-empty acts as a negative control.

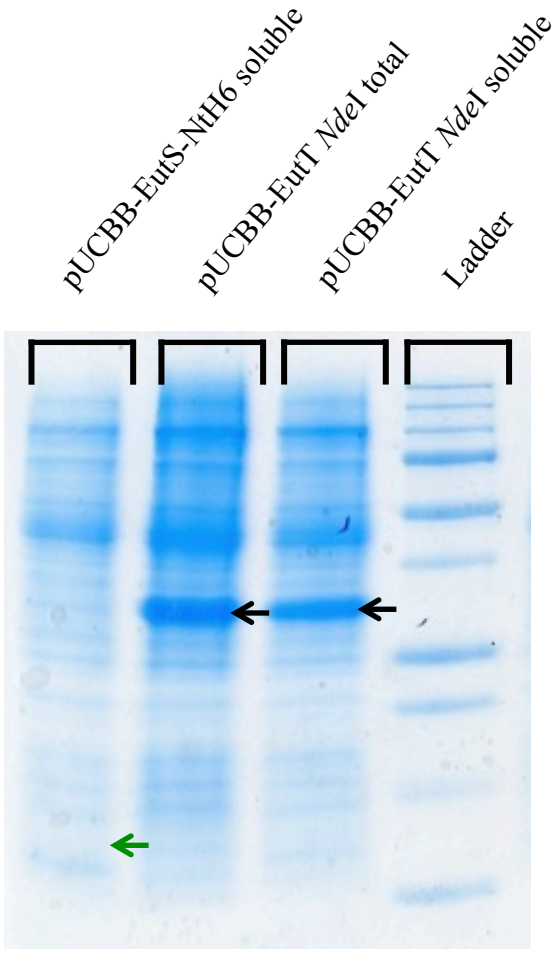

Figure 8. Results of EutT cargo protein expression study. The protein was visualized using BioSafe™ Coomassie dye. Precision Plus Protein™ All Blue Protein Standard from Bio-Rad acts as the protein ladder. The black arrows point to the band representing EutT (30 kDa) in both the total and soluble fractions. pUCBB-EutS-NtH6 transformed into C2566 acts as a negative control demonstrating the lack of a band near the size of EutT. The green arrow points to the band representing EutS-NtH6. The cultures used for this gel were grown at 30 ˚C for 16 hours.

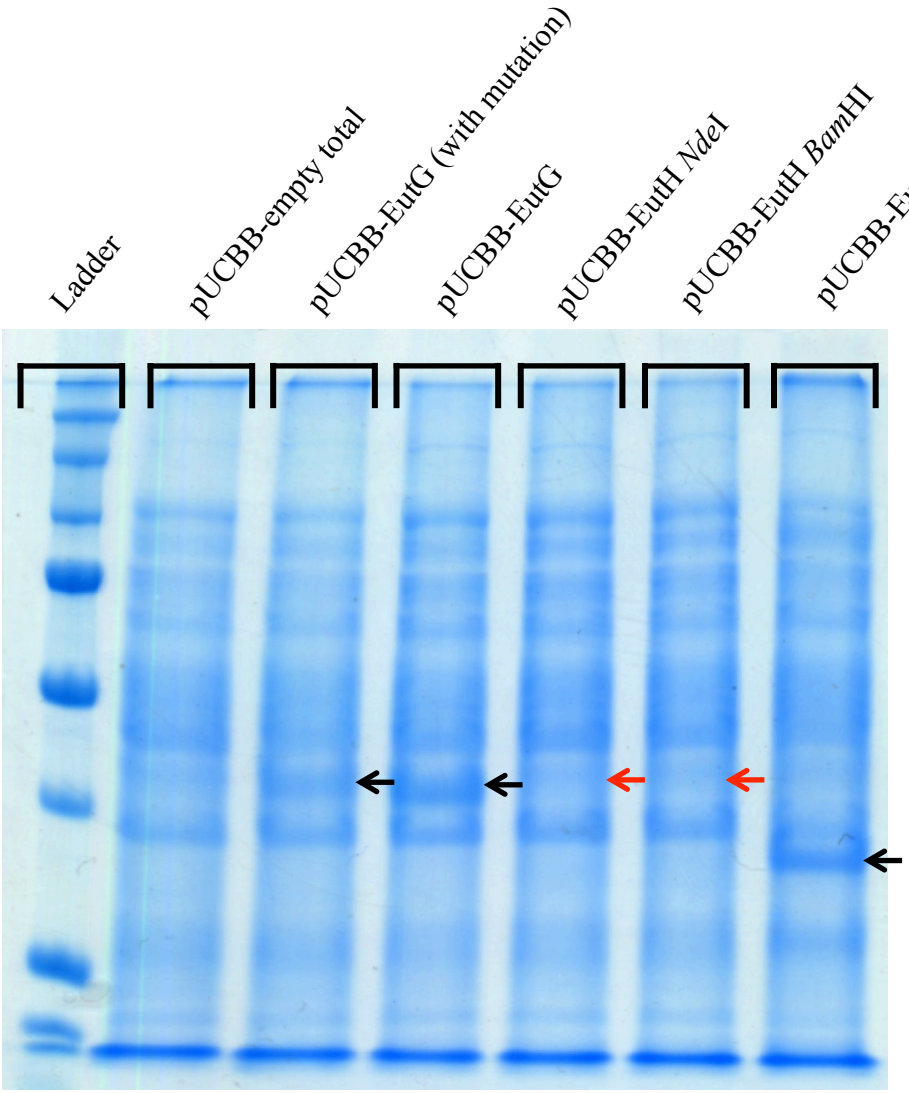

Figure 9. Results of EutG, EutH, and EutT expression studies. The protein was visualized using BioSafe™ Coomassie dye. Precision Plus Protein™ All Blue Protein Standard from Bio-Rad acted as the protein ladder. The EutG and EutT proteins were expressed, as represented by the black arrows near 41 kDa (EutG) and 30 kDa (EutT). EutH did not appear to express in either the *Nde*I or the *Bam*HI site and a red arrow points to where a band would be expected (43 kDa) if EutH expressed. pUCBB-empty acted as a negative control. The samples for this gel were grown at 37 ˚C for 24 hours.

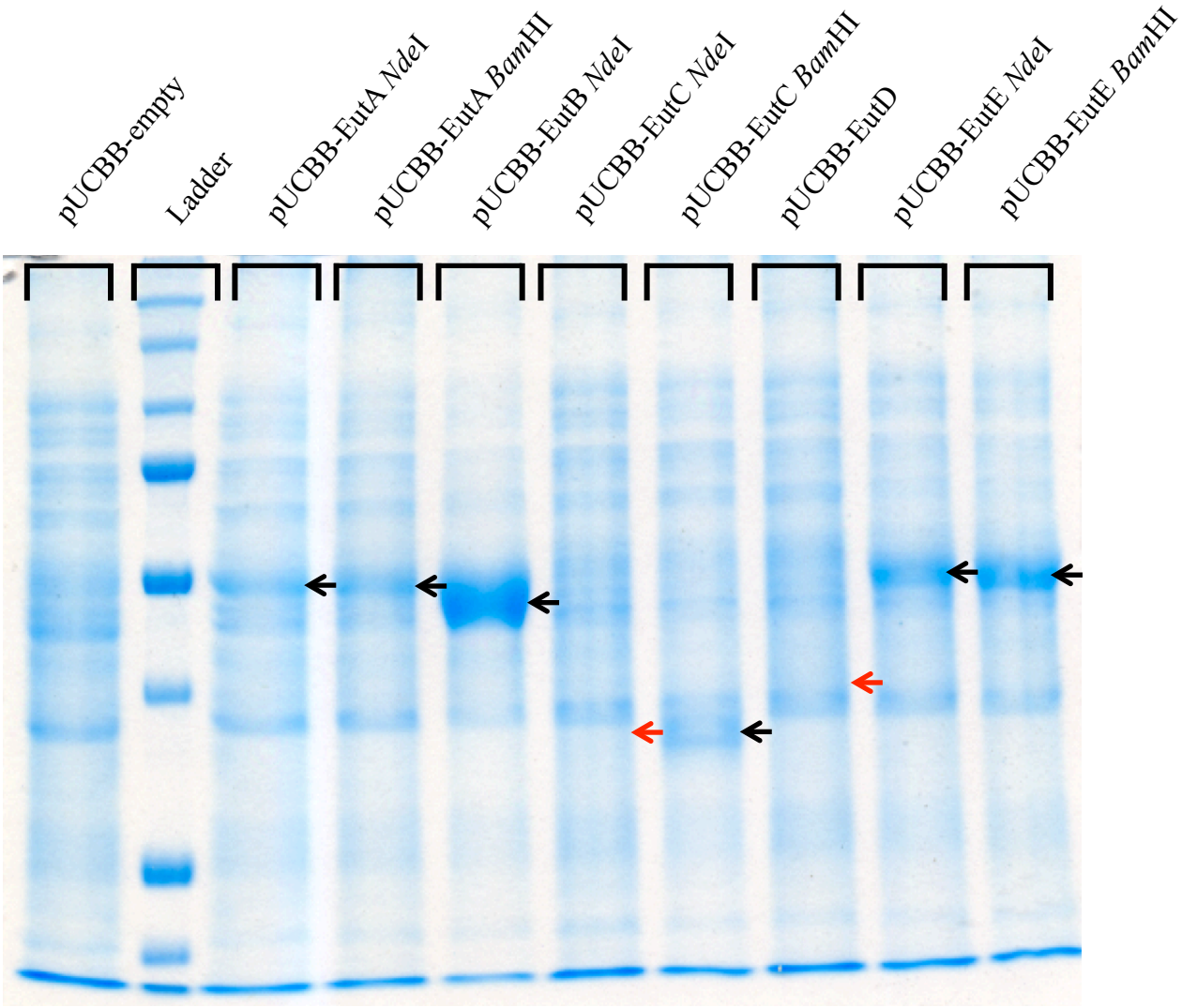

Figure 10. Results of EutA, EutB, EutC, EutD, and EutE expression studies. The protein was visualized using BioSafe™ Coomassie dye. Precision Plus Protein™ All Blue Protein Standard from Bio-Rad acts as the protein ladder. The EutA (49.5 kDa) and EutE (49 kDa) proteins appeared to express in both the *Nde*I and *Bam*HI site, as shown by the black arrows. EutC (32 kDa) expressed in the *Bam*HI site, but not the *Nde*I site, as shown by the black arrow pointing to the EutC band expressed from the *Bam*HI site and the red arrow pointing to where a band would be expected if EutC expressed from the *Nde*I site. A black arrow points to the band representing EutB (49.5 kDa). A red arrow points to where a band would be expected if EutD expressed (37 kDa). pUCBB-empty acts as a negative control. The samples for this gel were grown at 37 ˚C for 24 hours.

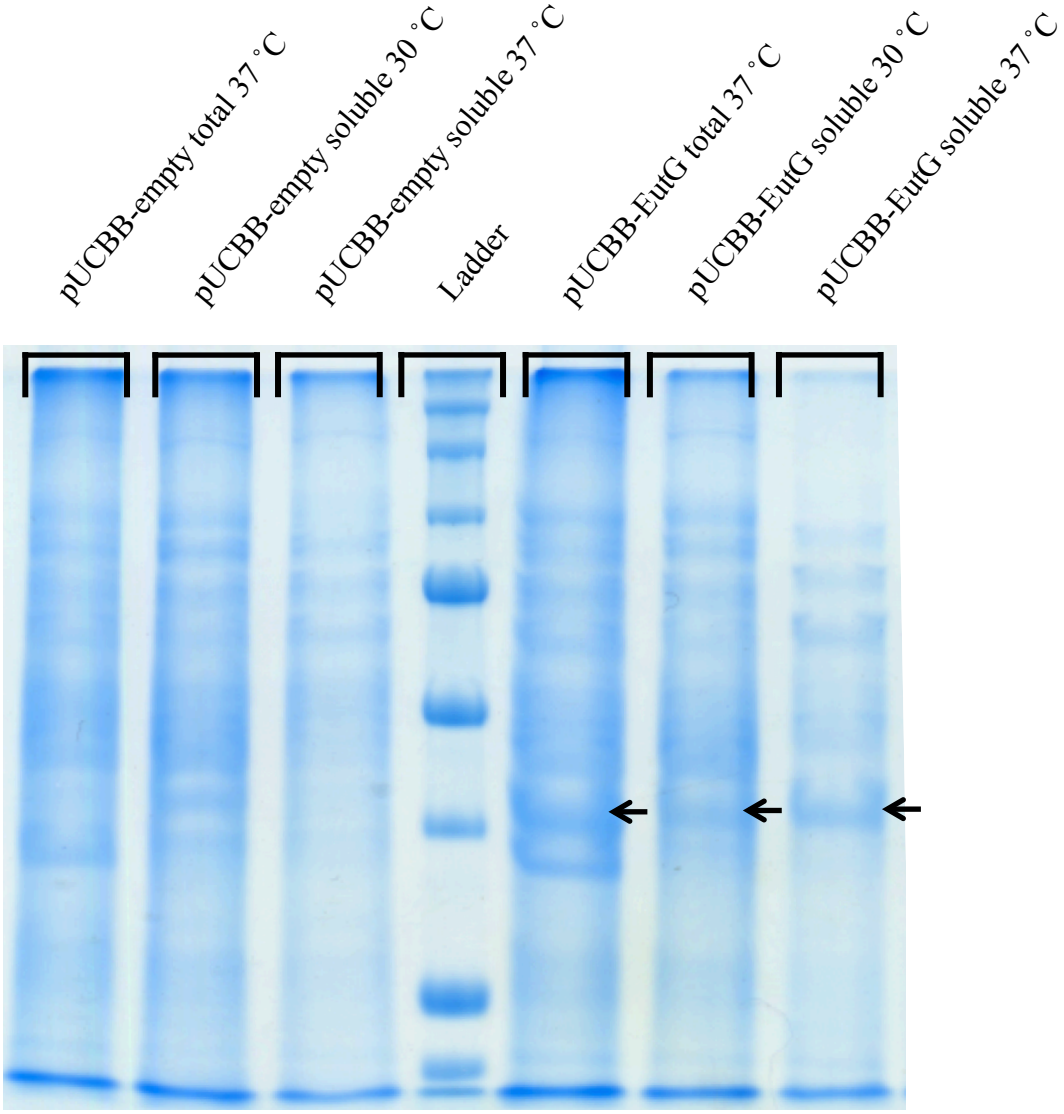

Figure 11. Results of EutG expression studies. The protein was visualized using BioSafe™ Coomassie dye. Precision Plus Protein™ All Blue Protein Standard from Bio-Rad acts as the protein ladder. EutG (41 kDa) appears to express solubly at both 30˚C and 37 ˚C, as shown by the black arrows. The lane containing the total fraction from C2566 containing pUCBB-EutG acts as a positive control, an a black arrow points to the band representing EutG. pUCBB-empty total fraction at 37 ˚C and the soluble fractions at 30 ˚C and 37 ˚C act as negative controls. The samples for this gel were grown at 37 ˚C for 24 hours.

Table 1. Primers used in this study.

Primer (gene direction restriction site)	Sequence
EutA forward *Nde*I	CATCTGCATATGAACACTCGCCAGCTAC
EutA reverse *Xho*I	CTCGAGCTCGAGTCAGGAAGGAAATGCG
EutA forward *Bam*HI	TCTGGATCCATGAACACTCGCCAGCTACTGAGC
EutB forward *Nde*I	CATATGCATATGAAACTAAAGACCACATTG
EutB reverse *Xho*I	TCTGAGCTCGAGTCAGAAGAACAGTGACGG
EutC forward *Nde*I	CATCTGCATATGGATCAAAAACAGATTGAAG
EutC reverse *Xho*I	TTCGAGCTCGAGTTAACGGGTCATGTTG
EutC forward *Bam*HI	TCTGGATCCATGGATCAAAAACAGATTGAAGAAATTG
EutD forward *Nde*I	CATATGCATATGATCATTGAACGCGCCC
EutD reverse *Xho*I	CTCGAGCTCGAGTCATTCAACCAGTGTG
EutD forward *Bam*HI	TCTGGATCCATGATCATTGAACGCGCCCGCG
EutE forward *Nde*I	CATATGCATATGAATCAACAGGATATTG
EutE reverse *Xho*I	CTCGAGCTCGAGTTATACAATGCGAAAC
EutE forward *Bam*HI	TCTGGATCCATGAATCAACAGGATATTGAACAGG
EutG forward *Nde*I (with mutation)	CATATGCATATGAAGCTGAACTACAGAC
EutG reverse *Xho*I	CTCGAGCTCGAGTTACCCGGCAGCCGCG
EutG forward *Bam*HI	TCTGGATCCATATGCAAGCTGAACTACAGACGGC
EutH forward *Nde*I	CATATGCATATGGGAATTAACGAAATCATC
EutH reverse *Xho*I	CTCGAGCTCGAGTCACGATTGCGCCTC
EutH forward *Bam*HI	TCTGGATCCATGGGAATTAACGAAATCATCATG
EutT forward *Nde*I	CATATGCATATGAACGATTTCATCACCG
EutT reverse *Xho*I	CTCGAGCTCGAGTCATGGCTTCTCTCCC
pBBinF	CATCCTGAACTTATCTAGACC
pBBinR	GCAGGTCCTGAAGTTAACTAG

Table 2. Plasmid vectors constructed or used in this study.

Plasmid vectors constructed/used in this study
pUCBB-EutA (*Nde*I site)
pUCBB-EutA (*Bam*HI site)
pUCBB-EutB
pUCBB-EutC (*Nde*I site)
pUCBB-EutC (*Bam*HI site)
pUCBB-EutD (*Nde*I site)
pUCBB-EutE (*Nde*I site)
pUCBB-EutE (*Bam*HI site)
pUCBB-EutG †
pUCBB-EutG
pUCBB-EutH (*Nde*I site)
pUCBB-EutH (*Bam*HI site)
pUCBB-EutT
pUCBB-EutS-NtH6 **
pUCBB-EutS-CtH6 **
pUCBB-EutM **
pUCBB-EutC$^{1-19}$eGFP **
pUCBB-EutC$^{1-19}$eGFP-CtH6

** from the Schimdt-Dannert lab plasmid collection

† with a frameshift mutation

Table 3. Results of expression studies of Eut genes. The "+" symbol represents when a protein was expressed. The "-" symbol represents when the protein was not expressed. The "+/-" symbol represents conflicting protein expression data. NT stands for not tested.

Gene	Expressed in pUCBB in *Nde*I and *Xho*I sites	Expressed in pUCBB in *Bam*HI and *Xho*I sites	Expressed solubly
EutA	+/-	+	NT
EutB	+	NT	+
EutC	-	+	NT
EutD	-	NT	NT
EutE	+/-	+	NT
EutG	+	NT	+
EutH	-	-	NT
EutT	+	NT	+
EutC$^{1-19}$eGFP	+	-	+