Engineered shell proteins confer improved encapsulated pathway behavior in a bacterial microcompartment

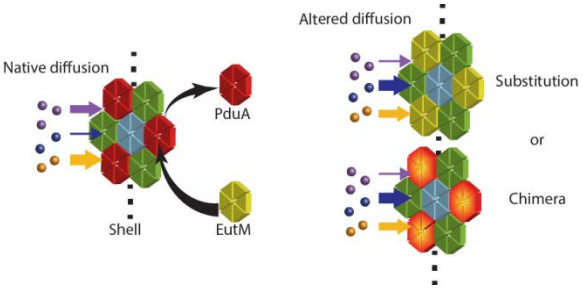
Marilyn F. Slininger Lee[†], Christopher M. Jakobson[†], Danielle Tullman-Ercek[‡]*

† Department of Chemical and Biomolecular Engineering, University of California Berkeley, United States

‡ Department of Chemical and Biological Engineering, Northwestern University, 2145 Sheridan Rd, Technological Institute E136, Evanston, IL 60208-3120, USA; Chemistry of Life Processes Institute, Northwestern University, 2170 Campus Drive, Evanston, IL 60208-3120

*Corresponding author

TOC Abstract Graphic



Abstract: Bacterial microcompartments are a class of proteinaceous organelles comprising a characteristic protein shell enclosing a set of enzymes. Compartmentalization can prevent escape of volatile or toxic intermediates, prevent off-pathway reactions, and create private cofactor pools. Encapsulation in synthetic microcompartment organelles will enhance the function of heterologous pathways, but to do so, it is critical to understand how to control diffusion in and out of the microcompartment organelle. To this end, we explored how small differences in the shell protein structure result in changes in the diffusion of metabolites through the shell. We found that the ethanolamine utilization (Eut) protein EutM properly incorporates into the 1,2-propanediol utilization (Pdu) microcompartment, altering native metabolite accumulation and the resulting growth on 1,2-propanediol as the sole carbon source. Further, we identified a single pore-lining residue mutation that confers the same phenotype as substitution of the full EutM protein, indicating that small molecule diffusion through the shell is the cause of growth enhancement. Finally, we show that the hydropathy index and charge of pore amino acids are important indicators to predict how pore mutations will affect growth on 1,2-

propanediol, likely by controlling diffusion of one or more metabolites. This study highlights the success of two strategies to engineer microcompartment control over metabolite transport: altering the existing shell protein pore via mutation of the pore-lining residues, and generating chimeras using shell proteins with the desired pores.

Keywords: Bacterial microcompartment, Protein shell assembly, Protein pore diffusion, Metabolosome, *Salmonella enterica*, 1,2-propanediol

Bacterial microcompartments (MCPs) are a prokaryotic form of cellular organization consisting of a protein shell used to enhance pathways by encapsulation. MCP-associated pathways include those found in hosts across many bacterial phyla for metabolism of niche carbon sources, such as 1,2-propanediol (1,2-PD), ethanolamine, and ethanol (involving Pdu, Eut, and Etu MCPs, respectively), and those found in cyanobacteria and chemoautotrophs for carbon fixation by RuBisCo in the carboxysome.[1-4] Additional naturally-encapsulated pathways have been recently discovered and a summary of this topic may be found in a number of review articles.[5-8] Some bacterial species have several MCP systems, encoding different organelles tailored to metabolize specific substrates.[9-12] For example, two well-characterized model systems are found in *Salmonella enterica*, which can produce both the Eut and Pdu MCPs from two separate operons. These enable growth on ethanolamine and 1,2-PD, respectively, which are both found in the human gut. The Pdu MCP encapsulates enzymes converting 1,2-PD through a propionaldehyde intermediate into either 1-propanol or propionate (Figure 1a), while the Eut MCP similarly converts ethanolamine through an acetaldehyde intermediate into ethanol and acetyl phosphate. The benefits of encapsulating a pathway include preventing the escape or toxicity of intermediates as well as cycling of cofactors.[13-17]

The *S. enterica* Pdu MCP assembly is an irregular polyhedron 150-200 nm in diameter and is enclosed by a shell composed of a few thousand copies of shell proteins of several types.[18] There are eight Pdu shell protein paralogs that are present in the final assembly in distinct and reproducible ratios.[19] The *S. enterica* Eut MCP is similarly composed of five shell proteins. MCP shell proteins form trimers,

pentamers, or hexamers that come together to assemble the hexagonal or pentagonal tiles of the polyhedron. A pore is formed at the symmetric center of each tile.[18–21]

The central pores of MCP shell proteins are implicated in the transport of small molecule substrates and products across the shell.[22, 23] For instance, Pdu MCP shell protein PduA is abundant in the shell and is suggested to be involved in the selective transport of 1,2-PD.[23] Specifically, occlusion of its pore by PduA^{S40GSG}, PduA^{S40L}, PduA^{S40C}, and PduA^{S40Q} mutations lead to growth defects that are rescued by high 1,2-PD concentrations, while the PduA^{S40A} mutation increased the escape of propionaldehyde.[23] Similarly, the pores of the carboxysome are hypothesized to control flux of bicarbonate and CO_2 across the shell.[24] All characterized carboxysome pore structures are positively charged at the narrowest point, which could enhance the flux of negatively charged bicarbonate relative to the escape of CO_2. Since different pore structures are postulated to lead to altered diffusivity of metabolites, engineering the pores could be the key to tailoring MCP properties for encapsulating non-native systems.

An alternative strategy for engineering MCP properties for non-native pathways is to use novel combinations of natural shell proteins. Shell proteins from different MCP systems have the potential to interact to form chimeric assemblies if expressed together at appropriate levels. Cai et al. demonstrated that an alpha-carboxysome shell protein complements the deletion of a beta-carboxysome shell protein.[25] Sturms et al. observed that co-expression of the Eut and Pdu operons in *S. enterica* from a synthetic promoter system resulted in co-purification of Eut and Pdu components, albeit with a disruption of Pdu MCP function.[26] Such observations are consistent with another finding: among many shell protein homologs, there are essential asparagine, arginine and lysine residues at the edge of each tile that are important for assembly, indicating a common interaction mechanism (For PduA these residues are K26, N29, and R79).[4, 27] In addition, shell assembly does not always depend on cargo loading. The protein shell can form when no cargo is expressed, and foreign proteins may also be encapsulated, or localized to the interior of the shell.[28-31] In fact, peptide sequences that direct native enzymes to the interior of the Eut

MCP also function to encapsulate proteins within the Pdu MCP.[32] Thus, it may be possible to bring Pdu and Eut shell proteins together to form chimeric MCP assemblies with new properties.

In this work, we explore two strategies for altering MCP function: mutations to the pore residues and generation of chimeric MCPs. We study the well-characterized *S. enterica* Pdu MCP, with a focus on its PduA shell protein and its paralog EutM, also from *S. enterica*. We find that EutM behaves as a PduA substitute in the Pdu MCP, and we describe the phenotypic changes resulting from EutM substitution. A Δ*pduA::eutM* chromosomally modified strain incorporates EutM into the Pdu MCP in place of PduA. Interestingly, the Δ*pduA::eutM* strain exhibits improved growth on 1,2-PD over cells expressing the native Pdu MCP, and this new phenotype is recapitulated by a single amino acid mutation in the PduA pore, PduA^{K37Q}. A saturation mutagenesis of position 37 of PduA reveals that the pore is highly amenable to mutation, and that the hydrophobicity and charge of the amino acid at this locus are the most important factors affecting growth on 1,2-PD. This study thus demonstrates the potential for success in altering metabolite transport properties with both MCP engineering strategies, enabling MCP design and optimization for synthetic biology applications.

RESULTS AND DISCUSSION

The pore of PduA is different from the pore of EutM

PduA forms one of three canonical shell protein types: a hexameric tile complex containing a central pore approximately 6 Å in diameter.[22] The residues defining the shape and properties of the pore include those contained in a loop at the hexamer symmetry axis with sequence Lys-Ile-Gly-Ser-Gly-Leu. The importance of the K37 and S40 loci of this sequence was previously demonstrated; certain mutations to either of these residues result in diminished growth of *S. enterica* LT2 on 1,2-PD as a sole carbon source.[23, 27] EutM, a close PduA homolog from the Eut operon, also forms a hexamer with nearly identical predicted structure (RMSD EutM/PduA = 1.004 Å, PDB ID 3MPY (EutM) and 3NGK (PduA)) and a highly similar amino acid sequence (67% identity).[20, 22] Refer to the protein sequence alignment for PduA

and EutM (Figure S1). Key differences between PduA and EutM are found in the pore, particularly at the residues aligning with PduA K37 and S40. The EutM pore loop has the sequence Gln-Ile-Gly-Gly-Gly-Leu (Figure 1b, c). In Figure 1c, the pore-lining lysine residue in PduA and the glutamine in EutM are highlighted in green.

The growth of *S. enterica* on 1,2PD is improved in Δ*pduA::eutM* strain

We substituted *eutM* for *pduA* to test whether the structural similarity of *eutM* is sufficient to complement the deletion of *pduA* and maintain proper Pdu MCP assembly. We also wanted to learn if small differences between the pore structures would manifest in significant differences in phenotypes attributed to metabolite diffusion. In order to minimize disruption of the native regulation of the Pdu operon, we constructed a genomic integration of *eutM* at the *pduA* locus (Figure 1d). To test MCP formation in this Δ*pduA::eutM* strain, the expression of the Pdu operon was induced in minimal NCE media containing 55 mM 1,2-PD as the sole carbon source, and 150 nM coenzyme B_{12}. Growth of the Δ*pduA::eutM* strain in this medium is an indicator of metabolite diffusion through the MCP structure. For instance, a strain harboring a *pduA* deletion without a compensating substitution exhibits decreased cell growth due to increased shell permeability to the toxic intermediate propionaldehyde, despite producing a seemingly closed MCP that may be purified and appears to contain all other necessary shell proteins and enzymes.[21, 27] Furthermore, occlusion or other changes to the architecture of the pores of the MCP shell also lead to growth defects due to restriction of diffusion of the substrate 1,2-PD.[7] Our results show this is not the case for the Δ*pduA::eutM* strain. Not only did the strain harboring the substitution grow without a noticeable defect on 1,2-PD, we observed improved growth over wild type under these growth conditions (Figure 3, Table 1). To control for changes in cell morphology, we also quantified colony forming units at various growth time points to confirm improved growth of the Δ*pduA::eutM* strain (Figure S2). The Δ*pduA::cat/sacb* mutant strain was used as a control for both growth measurements. This strain grew poorly compared to the wild-type (Figure 3, Figure S2). Moreover, no gross changes to cell morphology or physiology were observed in the Δ*pduA::eutM* strain by phase contrast (Figure S3).

123 **EutM is localized to the Pdu MCP in the *S. enterica* Δ*pduA::eutM* strain**

124 We next set out to demonstrate that EutM is incorporated into the shell, and to determine if the
125 formation of the shell was altered by this incorporation. We found that substituting *eutM* for *pduA*
126 allowed the formation of chimeric MCPs that were morphologically similar to the wildtype, as assessed
127 by a number of assays. First, we targeted green fluorescent protein (GFP) for localization within the
128 MCPs produced in each of the variants and analyzed assembly formation *in vivo* using fluorescence
129 microscopy. GFP was expressed as a fusion to the PduD$^{1-20}$ encapsulation tag and an SsrA degradation tag
130 such that only encapsulated GFP remains protected from proteolysis by the ClpXP machinery.[31, 33, 34] This
131 results in bright puncta by fluorescence microscopy. Fluorescent puncta were observed in both
132 Δ*pduA::eutM* and wild-type strains, confirming MCP formation and protein encapsulation *in vivo* (Figure
133 2a, e, b, f, and Figure S4). Automated image analysis by Cell Profiler yielded a count of the number of
134 puncta per cell and indicated that the number of chimeric MCPs per cell is similar to the number of wild-
135 type MCPs per cell (Figure S5).[35] Variability in puncta brightness is often observed between cells as well
136 as between puncta within a cell. While we have not investigated the cause of this feature, it appears
137 consistent across both the wild-type and Δ*pduA::eutM* strains.

138 We next purified MCPs by sedimentation to compare the morphology of chimeric MCPs to wild-
139 type *in vitro*. Transmission electron microscopy (TEM) analysis showed that MCPs purified from the
140 Δ*pduA::eutM* strain are morphologically similar to wild-type MCPs (Figure 2i, j, and Figure S4). An
141 analysis of MCP diameter demonstrated that MCPs purified from the Δ*pduA::eutM* strain have a similar
142 mean diameter to the wild-type but with a somewhat wider distribution of sizes (Figure S6).

143 The protein content of purified MCP samples was analyzed by densitometry measurements of
144 coomassie-stained SDS-PAGE gels. Prominent bands consistently observed in MCP samples were
145 previously identified via mass spectrometry.[19] Those identifications are used here. MCPs purified from
146 the Δ*pduA::eutM* strain exhibited shell protein bands in similar proportions to those found in wild-type
147 MCPs, except that the PduA band was missing and the 13 kDa band expected to be PduU roughly

doubled in density (Figures 2m, S4, S7). We hypothesized that this band represented both PduU and EutM superimposed. To verify the localization of EutM, the purification was repeated with a FLAG epitope tag fused to EutM. Western blotting indicated that EutM-FLAG runs at the apparent size of 13 kDa in a 12.5% acrylamide gel, and we confirmed that the 13 kDa band included EutM-FLAG by mass spectrometry (Figures 2n, S8, Table S1).

Next, we constructed a non-polar Δ*pduA::pduA-FLAG* strain. We purified MCPs from this strain along with the Δ*pduA::eutM-FLAG* strain and compared the levels of FLAG-tagged proteins using SDS-PAGE and western blot. Loading of the purified MCP samples was normalized by total protein concentration. We found that all samples contained similar amounts of PduA- or EutM-FLAG (Figure S9).

To confirm that EutM is not merely co-purifying with Pdu MCPs, anti-FLAG immuno-gold antibodies were used to label Pdu MCPs purified from the Δ*pduA::eutM-FLAG* strain . Immunostained TEM images show that the anti-FLAG immuno-gold antibodies bind well-formed MCP structures. In contrast, the wild-type MCP sample does not bind any gold particles (Figure 2 i, j). Taken together, all evidence supports our hypothesis that EutM properly incorporates into the Pdu MCP shell and causes no morphological defects.

PduA^{K26A} and EutM^{K24A} mutations both abrogate MCP formation

Sinha *et al*. demonstrated that a PduA^{K26A} mutant does not form MCPs due to disruption of hexamer edge binding contacts, resulting in a growth defect on 1,2-PD even though the enzymes required to metabolize 1,2-PD are present.[27] Also, MCP structures cannot be purified from the *pduA*^{K26A} strain via established methods for purifying the Pdu MCP. We replicated these results (Figures 2k, S10, 3), and in addition used fluorescence microscopy with our encapsulation reporter PduD$^{1-20}$-GFP-SsrA to observe this mutant *in vivo*. We did not observe an absence of fluorescence as we expected; instead, we saw cells linked by filaments resembling the nanotube or filament structures formed by overexpression of single

shell proteins or incomplete MCP operons as described previously (Figure 2c, g).[4, 28, 36, 37] These structures were accompanied by bright puncta varying in appearance and predominantly found at the cell poles. These puncta are presumed to be aggregated or poorly formed MCPs since they cannot be purified in the same manner as non-mutated MCPs. Attempts to purify MCPs from this strain gave only dilute aggregate species as seen by SDS-PAGE analysis as well as by TEM (Figure 2k, Figure S10).The same phenotype was observed when a similar mutation was made to *eutM* in the Δ*pduA::eutM* strain (Figure 2d, h, l, and Figure S10). Growth on 1,2-PD is decreased for a Δ*pduA::eutM* ^{K24A} strain, likely due to mis-formed MCPs like those formed in the *pduA*^{K26A} strain (Figure 3). Δ*pduA::eutM*^{K24A} and *pduA*^{K26A} mutants were used as non-MCP forming controls in subsequent experiments. Their behaviors coincide in all tests. This is further evidence that EutM is able to complement PduA and fully integrate into the Pdu MCP. Moreover, our results indicate that the same lysine binding partner is important for assembly of both the wild-type Pdu MCP and the EutM Pdu MCP chimera.

Limiting coenzyme B_{12} condition confirms Δ*pduA::eutM* MCPs form a complete compartment shell

Coenzyme B_{12} is a required cofactor for the first step of 1,2-PD metabolism in the Pdu pathway. This cofactor is supplied in the growth medium because *S. enterica* cannot produce it under our aerobic conditions. Growth at low coenzyme B_{12} concentrations (20 nM) has been used previously as a method to sensitively detect whether mutations to MCP proteins result in non-assembly or defective MCP shells.[21, 23, 27, 38] The limiting coenzyme B_{12} condition is advantageous for non-assembling MCP mutants, such as the *pduA*^{K26A} mutant, and allows improved growth. Conversely, the wild type and mutant strains that assemble well grow poorly at 20 nM coenzyme B_{12} compared to mutants that do not assemble MCPs. We exploited this growth condition as an additional assay to detect if the mutations we made to PduA disrupted MCP assembly. We found that the Δ*pduA::eutM* mutant grows as poorly as the wild-type strain in this condition (Figure S11), indicating that the chimeric MCP shells form a complete diffusion barrier.

The effects of EutM substitution are recapitulated by a point mutation in the PduA pore

196 The above experiments showed that there were no observable differences in shell protein
197 expression or MCP assembly between wild-type and Δ*pduA::eutM* strains. Thus, we set out to investigate
198 whether the behavior of the EutM-Pdu chimera is indeed due to a change in diffusion properties of the
199 shell. We hypothesized that the difference in growth caused by the EutM substitution was the result of a
200 change in diffusion through the central pore of the shell protein. To support this idea, we incorporated a
201 chromosomal point mutation to encode PduA^{K37Q}, to more closely mimic the EutM pore sequence.

202 The analyses performed for the Δ*pduA::eutM* substitution mutant were repeated with *pduA*^{K37Q}.
203 Strikingly, the *pduA*^{K37Q} mutation confers improved growth (Figure 3, S2) over wild type while still
204 maintaining normal MCP morphology and protein content (Figures 2m, S5, S6, S7, S9, S12, S13). A
205 single point mutation to the pore causes the same effect as a complete shell protein substitution. This is
206 strong evidence indicating that diffusion of molecules through the pore of the MCP shell is altered,
207 causing the observed changes in growth and metabolite levels. A modification of lysine to glutamine
208 changes the charge environment in the pore, and may also change the polarity or hydrogen bonding
209 configuration. Both of these factors likely alter the diffusion barrier small molecules experience.[23, 39] It is
210 convenient for MCP engineering purposes that such small mutations to this region can confer improved
211 function of the whole organelle structure.

212 **The EutM substitution and the PduA^{K37Q} pore mutant change the accumulation of downstream**
213 **metabolites**

214 HPLC analysis was performed to measure metabolite levels in the media over the course of
215 growth on 1,2-PD (Figure 4). Fast growth on 1,2-PD has previously been associated with decreased
216 aldehyde levels in the media.[10] Therefore, we expected that the Δ*pduA::eutM* and *pduA*^{K37Q} strains would
217 accumulate less aldehyde in the media. Indeed, we found that the accumulation of downstream products
218 propionaldehyde, 1-propanol, and propionate are decreased in both the Δ*pduA::eutM* and *pduA*^{K37Q}
219 mutants. We also found that the consumption of 1,2-PD occurred at roughly the same rate in wild-type,
220 Δ*pduA::eutM*, and *pduA*^{K37Q} strains (the differences in calculated rate are not statistically significant).

Non-assembling mutants (*pduA*^{K26A} and Δ*pduA::eutM* ^{K24A}) show the highest levels of propionaldehyde, at 8.6 and 10 mM, respectively, compared to 2.1 mM for the wild-type MCP (Figure 5). Propionaldehyde is toxic to *S. enterica* when exogenously added at concentrations of about 8 mM.[14] Also, since propionaldehyde is volatile, accumulation of this molecule in the media results in some unmeasured amount of propionaldehyde evaporation to the head space. The differences in propionaldehyde concentrations among wild type, Δ*pduA::eutM*, and *pduA*^{K37Q} are more subtle. Wild-type *S. enterica* maintains about 2 mM propionaldehyde for much of the exponential phase. We observed no detectable propionaldehyde in Δ*pduA::eutM* cultures apart from a single spike usually occurring between 27 and 30 hours reaching a level averaging 2 mM. The *pduA*^{K37Q} strain behaved similarly to the Δ*pduA::eutM* strain with respect to propionaldehyde accumulation. Therefore, while these strains reach similar maximum concentrations of propionaldehyde in the media, the Δ*pduA::eutM*, and *pduA*^{K37Q} *mutants*sustain this concentration for a shorter period.

In wild-type cultures the 1-propanol concentration level stabilizes at around 25 mM from 36 hours onward. In contrast, Δ*pduA::eutM* and *pduA*^{K37Q} strains reach a maximum of about 20 mM 1-propanol, and the concentration decreases after 36 hours. This indicates that the accumulation rate of 1-propanol is slower in the Δ*pduA::eutM* and *pduA*^{K37Q} mutants. Non-assembling mutants accumulate even less 1-propanol, likely due to the greatly increased escape of the aldehyde before conversion. After the supply of 1,2-PD is exhausted, 1-propanol in the media is likely converted back to propionaldehyde and then consumed.[23] This process in addition to some evaporation from the headspace results in the decline in 1-propanol concentration at the end of each growth period. A surprising trend is seen in the concentration of propionate observed in the media over time. We observed increased propionate accumulation in strains that grew poorly. Non-assembling mutants ranked highest in propionate accumulation, followed by the wild type, then *pduA*^{K37Q} and Δ*pduA::eutM*.

To offer an explanation for these behaviors, we propose that the *ΔpduA::eutM* and *pduA*^{K37Q} mutants restrict the diffusion of intermediate propionaldehyde, the negatively charged propionyl-

246 phosphate product, or both. In the following discussion, we consider how this hypothesis is supported by
247 the observed metabolite concentrations and growth phenotype.

248 The restriction of propionaldehyde diffusion by the *ΔpduA::eutM* and *pduA*^{K37Q} mutants would
249 directly result in the observed decreases in propionaldehyde accumulation compared to the wild-type.
250 Since propionaldehyde is volatile, increased accumulation in the media leads not only to toxicity, but also
251 to a loss of carbon to the gas phase. Preventing aldehyde escape may also result in a higher local
252 propionaldehyde concentration and a kinetic enhancement of pathway throughput. Both of these effects
253 contribute to the differences in final cell density observed between strains, leading to fastest growth in the
254 *ΔpduA::eutM* and *pduA*^{K37Q} mutants.

255 Restriction of propionyl-phosphate diffusion could result in the decreased accumulation of both
256 propionate and 1-propanol by the *ΔpduA::eutM* and *pduA*^{K37Q} mutants. To illustrate, non-assembling
257 mutants allow free diffusion of propionyl-phosphate, facilitating its conversion to propionate by the
258 PduW kinase, which is not localized to the MCP. In contrast, propionyl-phosphate molecules face a
259 moderate diffusion barrier in the wild-type pore, resulting in decreased propionate release compared to
260 non-assembling mutants. Thus, the *ΔpduA::eutM* and *pduA*^{K37Q} mutants may form a greater diffusion
261 barrier since even less propionate is released than in the wild-type strain. The mitigation of 1-propanol
262 accumulation in *ΔpduA::eutM* and *pduA*^{K37Q} mutants might also be due to restricted propionyl-phosphate
263 diffusion. Decreased product escape could cause a pathway back-up reducing the rate of NADH
264 production, leading to slower turnover in the 1-propanol branch of the pathway involved in NADH
265 recycling. Considering pore structure, it is intuitive that a positively charged lysine in the wild-type PduA
266 pore would facilitate the interaction of a negative product such as propionyl-phosphate more than the
267 neutral glutamine, and allow a greater rate of egress.[39, 40] Propionate escape in the wild-type could lead
268 directly to decreased growth rate, or slower propionate uptake into central metabolism could be a result of
269 aldehyde toxicity.

We recognize that there is still much to learn about the fate of propionate and the relationship between the 1,2-PD utilization pathway and the rest of central metabolism. According to our current understanding, the effect of decreased propionaldehyde leakage should result in flux enhancement. However, though strains that appear to restrict aldehyde leakage grow better, they also release less of the downstream products, propionate and 1-propanol. Further study is necessary resolve these two effects.

Shell protein pore residues can be altered to control encapsulated enzyme function

We further explored the relationship between the pore and 1,2-PD metabolism by creating a genomically-encoded library of PduA variants that differ at residue 37. We checked each mutant for MCP formation by fluorescence microscopy using our encapsulation reporter $PduD^{1-20}$-GFP-ssrA (Figure S12). All mutants exhibited fluorescent puncta, indicating the formation of MCPs. Growth on 1,2-PD with a limiting concentration of coenzyme B12 further demonstrated that all mutants did not form "leaky" MCP structures (Figure S11). Finally, we selected a few of the library members for purification. Purified MCPs from these strains are all similar to wild-type MCPs when examined by SDS-PAGE banding pattern and by TEM. (Figure S10 and S13).

The library was then screened for growth in NCE with 1,2-PD and 150 nM coenzyme B12 (Figure S14). The end-point cell density for each mutant is summarized in Figure 6a. To analyze these data we compared average end-point cell density to the characteristics of the amino acid substituted in each mutant. The variables considered were the predicted amino acid pK_a, hydropathy index, and residue molecular weight. We found a moderate negative correlation between hydropathy index and final cell density (correlation coefficient of -0.47), indicating more hydrophobic side chains generally caused a decrease in growth (Figure 6c). Final cell density was also plotted versus pK_a for several charged amino acid side chains. There was a negative correlation (correlation coefficient of -0.6) suggesting that a more positively charged pore is detrimental to growth on 1,2-PD (Figure 6b). There was no clear trend when growth was compared to size of the amino acid side chain alone (Figure S15). This indicates that, of the

three characteristics considered, the hydrophobicity and charge of the amino acids in the pore were most important for retention of small molecules like propionaldehyde within the MCP.

Pore residues of PduA evolve rapidly

Upon analysis of the PduA K37 mutagenesis library, we were surprised to observe that the pore loop of PduA is very tolerant to mutation; indeed, all mutants formed MCPs similar to wild type. It is remarkable that this locus of PduA can accommodate any amino acid without a compensating mutation. For future engineering of the MCP shell, it is desirable to know if other residues may be mutated to change MCP function or to add chemical handles for the conjugation of useful moieties. We therefore used the Rate4Site program (version 2.01) to calculate an evolutionary rate scaling factor for each residue of PduA relative to the overall evolution rate of the protein.[41] The input of this program is a multiple sequence alignment of 193 non-redundant PduA homologs from a variety of MCP systems. The resulting output is mapped over the structure of PduA (Figure 7a). Residues of a darker red color have high evolutionary rates, while white residues evolve slowly relative to the average rate for the entire protein. The evolutionary rate parameter represents the number of non-synonymous replacements expected per unit time for a specific residue. This parameter is related to conservation score but the rate calculation also uses the architecture and branch lengths of the phylogenetic tree in order to account for evolutionary time. If we detect residues that change often over time, but greatly affect the protein function, this could indicate a structural feature important for adapting function to environmental change.

The results indicate that pore positions 37 and 40 have a higher evolutionary rate than the average evolution rate for the entire protein, while the isoleucine and glycine at positions 38 and 39 evolve more slowly than the protein average. This suggests that a hydrophobic motif at the narrowest part of the pore is essential for MCP function in general, whichever pathway is encapsulated within. On the other hand, while the residues forming rings on either side of the hydrophobic motif evolve more quickly, they evidently also affect encapsulated pathway function, most likely by inhibiting or facilitating diffusion of

one or more metabolites. The mutability of certain pore residues may allow the flexibility to encapsulate a number of pathways involving a diversity of small molecules.

Interestingly, the residues K90-I92 near the C-terminus of PduA, as well as several other residues on both concave and convex surfaces, are also very rapidly evolving (labeled on Figure 7a). These areas of the protein structure are less constrained and may serve as good targets for protein modifications, particularly for control of cargo loading or for labeling the outer surface for engineering applications.

PduA pore mutants conferring faster growth on 1,2-PD are not abundant among PduA homologs

Among the strains in the K37 mutation library, the strains with Gln, Glu, Ser, and Trp at position 37 grew to a greater final OD_{600} than the wild-type strain under laboratory conditions. Curiously, this indicates that the *S. enterica* serovar lab strain used in this study is not optimized for flux through the Pdu pathway and other factors may affect the fitness of LT2 in the host gut. We searched among PduA homologs to find out if any of the amino acid variants that resulted in highest growth were among those commonly found in nature. A distribution of the frequencies of each amino acid aligned with position 37 is given in Figure 7b. Together, strains with Gln, Glu, Ser, or Trp aligned with position 37 only represent 9.38% of over 1000 related sequences. The population of PduA homologs in nature is not enriched for these high performers.

Conclusions

In this work, we demonstrate the ability to assemble Pdu MCPs with a foreign shell protein substitution, and the flexibility to use protein engineering to alter the pore structures of PduA without abrogating MCP structure. For the first time, mutations to the Pdu MCP shell have resulted in enhanced function of the enclosed pathway. PduA shell protein mutants, including those with only a single amino acid substitution of a pore-lining residue, have improved growth phenotypes with 1,2-PD as a carbon source. We uncovered significant evidence that these phenotypes are due to modification of small molecule diffusion rates through the pores of the MCP, leading to escape of the intermediate aldehyde

and possibly preventing the escape of the product propionyl-phosphate. These trends are observable through HPLC analysis of metabolites and are supported by our analysis of protein structure and MCP assembly. An investigation of the relationship between amino acid chemistry in the pore and the resulting growth performance indicates that charge and hydrophobicity of the side chain play key roles. Shells with more negatively charged pores tended to grow poorly on 1,2-PD, while an increase in the hydropathy of the residue at position 37 also decreased the fitness of the strain. The PduA K37 mutagenesis library produced a range of growth phenotypes with strains that grew on 1,2-PD faster or slower than the wild type. Generation of chimeras and/or modification of shell protein pore residues are both promising techniques to optimize MCP shell characteristics for an encapsulated pathway of interest and to achieve useful phenotypes not found in nature.

METHODS

Bacterial strains, media, and growth conditions

The bacterial strains used in this study are *Salmonella enterica* serovar Typhimurium LT2 and *Escherichia coli* DH10B. For mutations to *pduA*, the gene was amplified from the *S. enterica* chromosome and inserted into a pBAD33, p15a, cm^R plasmid by golden gate assembly with BsaI restriction endonuclease.[42] This plasmid was transformed into *E. coli*. *pduA* and was mutated as described above using PCR site-directed mutagenesis. EutM K24A was created in the same manner. Primers for each mutation are shown in supplemental Table S2.

Chromosomal modifications were made by recombineering as described by Court.[43] First, the native *pduA* gene was replaced with the *cat/sacb* cassette amplified from the TUC01 genome with primers mfsp105 and either mfsp106 or mfsp195 for *eutM* or *pduA* insertions, respectively (primer sequences listed in Table S2). This amplicon was used to create Δ*pduA::cat/sacb*. The e*utM* gene was amplified with primers mfsp199 and mfsp200 and used to make Δ*pduA::eutM*. To make mutations in

366 *pduA*, the gene was first subcloned into a plasmid vector and mutated by site directed mutagenesis as
367 described above. The inserts were then amplified from plasmids by PCR with primers mfsp197 and
368 mfsp104 and used to make all chromosomal *pduA* mutants. For FLAG tag fusions, the non-polar
369 Δ*pduA::cat/sacb* strain constructed with mfsp106 was used as a starting point to maintain the same RBS
370 region for the subsequent *pduB* gene. The FLAG sequence was added by PCR to *pduA* and *eutM*
371 sequences using primers mfsp304 and mfsp218, respectively before insertion by recombineering. Each
372 strain was confirmed by Sanger sequencing of PCR amplicons from the PduA region of the *S. enterica*
373 chromosome.

374 For growth of *S. enterica*, single colonies were picked from freshly streaked plates and were
375 grown in 5 mL of TB (Terrific Broth) medium at 30°C, 225 rpm, for 24 hours. For growth curves with a
376 1,2-PD carbon source, cultures were diluted to an OD_{600} of 0.05 in 50 mL No-Carbon-E (NCE) minimal
377 medium supplemented with (unless stated otherwise) 55 mM 1,2-PD, 150 nM coenzyme B_{12}, 1 mM
378 magnesium sulfate, and 50 μM ferric citrate.[32] Cultures were grown in 250 mL flasks at 37°C in an orbital
379 shaker at 225 rpm. At every time point, 1.5 mL of cell culture sample was taken for OD_{600} measurement
380 and HPLC analysis. The doubling time of each growth curve was calculated by plotting the log of the
381 OD_{600} vs. time. The linear region of each individual curve was identified (including a minimum of three
382 time points). The slope of the linear regression between time points was determined, and the doubling
383 time was calculated with $t_D = \log(2)\,/slope$.

384 For MCP purifications, flow cytometry, or microscopy experiments, single colonies were grown
385 in 5 mL LB (Lysogeny broth)-Lennox medium at 30°C, 225 rpm, for 24 hours. Cell cultures were diluted
386 from the primary culture 1:1000 into NCE supplemented with 42 mM sodium succinate, 1mM
387 magnesium sulfate, and 50 μM ferric citrate. For cultures containing the plasmid expressing $PduD^{1\text{-}20}$-
388 GFP-ssrA, half the usual amount of antibiotic was used (17 mg/mL chloramphenicol). Culture volumes
389 were 400 mL of media in 2L flasks for MCP purification, and 5 mL of media in 24-well blocks
390 (Analytical Sales and Services, Inc., cat. no. 24108) for flow cytometry and microscopy. Cells were

grown at 37˚C in an orbital shaker at 225 rpm. For cultures with PduD$^{1-20}$-GFP-ssrA expressed from a pBAD33 plasmid, 1.33 mM arabinose was added when subculture reached an OD_{600} of 0.4. After five additional hours of growth, samples were taken for fluorescence microscopy, flow cytometry, or Pdu MCP purification.

High Performance Liquid Chromatography (HPLC) media analysis

At each time point, a 750 µL sample of whole cell culture was taken. This was centrifuged for 5 min at 13,000 g to remove cells. The supernatant was decanted and frozen until time of analysis. Before analysis samples were thawed and filtered. The Shimadzu HPLC detection system was equipped with a Bio-Rad Aminex HPX-87H (300 by 7.8 mm) ion exclusion column, LC-20AD solvent delivery system, SIL-20AC auto sampler, RID-10A refractive index detector, and SPD-M20A diode array detector (Shimadzu, Kyoto, Japan). Mobile phase is 5 mM H_2SO_4 with 0.4 mL min^{-1} flow rate.[14] The column temperature is 35°C.

MCP purification

MCPs were purified by centrifugation of lysate as described previously.[44] They were stored at 4°C until analysis by SDS-PAGE or TEM. Total protein concentration of pure MCP sample was determined by Pierce BCA protein assay kit (Thermo Fisher).

Transmission electron microscopy

Samples were placed on 400 mesh formvar coated copper grids with a carbon film after the grids were treated by glow discharge. 10 μL of purified MCPs at a concentration of about 0.1 mg/mL was placed on each grid for 2 min. The grids were washed three times with deionized water before fixation. Then, 10 μL of 2% glutaraldehyde in water was placed on each grid 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. All solutions were centrifuged directly before use to avoid aggregate or particulate contact with grids.

Samples were imaged with an REI Tecnai T12 transmission electron microscope and a Gatan Ultrascan 1000 camera (Gatan, Pleasanton, CA).

Immunostaining

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 about 0.1 mg/mL was placed on grids for 2 min. The grids were washed three times with deionized water. Blocking buffer was prepared with 25 mL of phosphate buffered salt (PBS) at pH 7.4, 0.2 g bovine serum albumin (BSA), and 25 μL of gelatin from cold water fish skin (Sigma). Then, grids were incubated with blocking buffer for 15 min. Grids were incubated with mouse anti-FLAG primary antibody (Sigma) diluted 1:1000 in the blocking buffer for 2 hours. Grids were washed four times with PBS for 2 min each. Grids were then incubated with anti-mouse IgG –gold secondary antibody produced in goat (Sigma) diluted 1:20 in blocking buffer for 1 hour. Grids were washed again four times with PBS for 2 min each time. Samples were fixed with 2% glutaraldehyde in PBS for 5 min. Samples were washed for 2 min in PBS, then four times in deionized water. Grids were stained with 1% uranyl acetate in water for 3 min before grid was allowed to dry.

Mass spectroscopy

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 dithiothreitol for 20 minutes 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 4C for 3 hours prior to overnight digestion at 37 °C. Liquid was collected following digestion and the gel

pieces were bathed in 100% acetonitrile 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 seconds. 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 *S. enterica* protein database (UniProtKB/SwissProt Protein Knowledge Base).

Protein electrophoresis and western blot

Denaturing protein electrophorese (SDS-PAGE) was performed on purified MCPs using a 12.5% polyacrylamide gel unless otherwise noted. A 130V potential was applied for 70 minutes. Unless otherwise noted, equal volumes (15 µL) of each MCP sample were loaded onto the gel, and densitometry quantifications were normalized by sample protein concentration. Gels were either stained with Coomassie dye or further processed for western blot.[44] Western blotting was done with a PVDF membrane according to standard protocols. Blotting was done with a mouse anti-FLAG primary antibody (Sigma) 1:2000 dilution in 50 mM Tris 150 mM NaCl pH 7.6 with 0.05% Tween-20 (TBST) with 1% w/v dry milk. The secondary antibody was HRP-conjugated goat anti-mouse antibody (Thermo) diluted 1:1000 in TBST. Labeling was visualized with west-pico chemiluminescent substrate (Thermo) using a Bio-Rad ChemiDoc XRS+.

Densitometry analysis of Coomassie stained gels was performed using at least three different MCP purifications for each strain. Each purified sample was run on three different gels. Images were analyzed

using the Image Lab software (Bio-Rad). The absolute signal measurement for each band was normalized by the signal for the band containing proteins PduT and PduE (band 7 in Figure 2). This band had the least variation in signal over all replicates.

Fluorescence microscopy

Bacteria were viewed using a Nikon Ni-U upright microscope with a 100x 1.45 n.a. plan apochromat objective. Images were captured using an Andor Clara-Lite digital camera and Nikon NIS Elements software. Fluorescence images were collected using a C-FL Endow GFP HYQ bandpass filter. All images were taken with 400 ms exposure and adjusted identically for contrast in Adobe Photoshop software.

Flow cytometry

Cultures were grown as described above. Samples were diluted 1:40 in phosphate-buffered saline (PBS) supplemented with 2 g/L kanamycin (to halt translation) in 96-well plates. 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. Error bars represent one standard deviation.

Multiple Sequence Alignment construction and evolutionary rate analysis

Evolutionary rate of each amino acid locus of PduA was calculated using the Rate4Site program version 2.01.[41] The input for this tool was a multiple sequence alignment (MSA) of 193 non-redundant PduA homologs. The homologs were retrieved and the MSA assembled by the jackhmmer web server (https://www.ebi.ac.uk/Tools/hmmer/search/jackhmmer). Rate4Site uses this data to construct a phylogenetic tree. With both MSA and tree information, a rate-scaling factor is calculated that indicates how rapidly each residue evolves relative to the mean protein rate. Default settings (including Bayesian framework) were used here. A larger MSA of about 1000 sequences, also retrieved and aligned with jackhammer, was used to calculate amino acid frequency.

SUPPORTING INFORMATION Supplementary information including tables and figures cited in the text is available free of charge via the Internet at http://pubs.acs.org.

ABBREVIATIONS

MCP: microcompartment
1,2-PD: 1,2-propanediol
GFP: green fluorescent protein
Pdu: 1,2-propanediol utilization
Eut: ethanolamine utilization
NCE medium: No-Carbon-E medium
TEM: transmission electron microscopy
MSA: multiple sequence alignment
PBS: phosphate buffered salt

AUTHOR INFORMATION

Corresponding Author
Danielle Tullman-Ercek
*E-mail: dtercek@berkeley.edu
Department of Chemical and biomolecular Engineering, 201 Gilman Hall, University of California Berkeley 94720, United States

Author Contributions

M.F.S and D.T.E. designed the project. M.F.S. performed the experiments. M.F.S, C.M.J, and D.T.E. wrote the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGEMENTS

The authors wish to thank Chuchu Zhang for development of the Cell profiler pipeline used in this work, as well as the Robert D. Ogg Electron Microscope Laboratory for equipment usage, training, and materials. The authors would also like to thank the Tullman-Ercek lab for thoughtful discussions and comments. The work was funded by the National Science Foundation (award MCB1150567 to D.T.E.).

REFERENCES

(1) Bobik, T. a, Havemann, G. D., Busch, R. J., Williams, D. S., and Aldrich, H. C. (1999) The propanediol utilization (pdu) operon of Salmonella enterica serovar Typhimurium LT2 includes genes necessary for formation of polyhedral organelles involved in coenzyme B(12)-dependent 1, 2-propanediol degradation. *Journal of bacteriology 181*, 5967–75.

(2) Kofoid, E., Rappleye, C., Stojiljkovic, I., and Roth, J. (1999) The 17-gene ethanolamine (eut) operon of Salmonella typhimurium encodes five homologues of carboxysome shell proteins. *Journal of Bacteriology* 181, 5317–5329.

(3) 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.

(4) Heldt, D., Frank, S., Seyedarabi, A., Ladikis, D., Parsons, J. B., Warren, M. J., and Pickersgill, R. W. (2009) Structure of a trimeric bacterial microcompartment shell protein, EtuB, associated with ethanol utilization in Clostridium kluyveri. *The Biochemical journal* 423, 199–207.

(5) Bobik, T. A. (2006) Polyhedral organelles compartmenting bacterial metabolic processes. *Applied Microbiology and Biotechnology 70*, 517–525.

(6) Kerfeld, C. A., Heinhorst, S., and Cannon, G. C. (2010) Bacterial microcompartments. *Annual review of microbiology 64*, 391–408.

(7) Chowdhury, C., Sinha, S., Chun, S., Yeates, T. O., and Bobik, T. A. (2014) Diverse bacterial microcompartment organelles. *Microbiology and Molecular Biology Reviews 78*, 438–468.

(8) Kerfeld, C. A., and Erbilgin, O. (2015) Bacterial microcompartments and the modular construction of microbial metabolism. *Trends in Microbiology 23*, 22–34.

(9) Abdul-Rahman, F. (2013) The distribution of polyhedral bacterial microcompartments suggests frequent horizontal transfer and operon reassembly. *Journal of Phylogenetics & Evolutionary Biology 1*, 118.

(10) Jorda, J., Lopez, D., Wheatley, N. M., and Yeates, T. O. (2013) Using comparative genomics to uncover new kinds of protein-based metabolic organelles in bacteria. *Protein Science 22*, 179–195.

(11) Petit, E., LaTouf, W. G., Coppi, M. V., Warnick, T. a, Currie, D., Romashko, I., Deshpande, S., Haas, K., Alvelo-Maurosa, J. G., Wardman, C., Schnell, D. J., Leschine, S. B., and Blanchard, J. L. (2013) Involvement of a bacterial microcompartment in the metabolism of fucose and rhamnose by Clostridium phytofermentans. *PloS one 8*, e54337.

(12) Axen, S. D., Erbilgin, O., and Kerfeld, C. A. (2014) A taxonomy of bacterial microcompartment loci constructed by a novel scoring method. *PLoS Computational Biology* (Tanaka, M. M., Ed.) *10*, e1003898.

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

555 (14) Sampson, E. M., and Bobik, T. a. (2008) Microcompartments for B12-dependent 1,2-propanediol
556 degradation provide protection from DNA and cellular damage by a reactive metabolic intermediate.
557 *Journal of bacteriology 190*, 2966–71.
558 (15) Huseby, D. L., and Roth, J. R. (2013) Evidence that a metabolic microcompartment contains and
559 recycles private cofactor pools. *Journal of bacteriology 195*, 2864–79.
560 (16) Cheng, S., Fan, C., Sinha, S., and Bobik, T. a. (2012) The PduQ enzyme is an alcohol dehydrogenase
561 used to recycle NAD+ internally within the Pdu microcompartment of Salmonella enterica. *PloS one 7*,
562 e47144.
563 (17) Liu, Y., Jorda, J., Yeates, T. O., and Bobik, T. A. (2015) The PduL phosphotransacylase is used to
564 recycle coenzyme A within the Pdu microcompartment. *Journal of Bacteriology* (Parkinson, J. S., Ed.)
565 *197*, 2392–2399.
566 (18) Yeates, T. O., Crowley, C. S., and Tanaka, S. (2010) Bacterial microcompartment organelles: protein
567 shell structure and evolution. *Annual review of biophysics 39*, 185–205.
568 (19) Havemann, G. D., and Bobik, T. A. (2003) Protein content of polyhedral organelles involved in
569 coenzyme B12 -dependent degradation of 1,2-propanediol in Salmonella enterica Serovar Typhimurium.
570 *Journal of bacteriology 185*, 5086-95.
571 (20) Takenoya, M., Nikolakakis, K., and Sagermann, M. (2010) Crystallographic insights into the pore
572 structures and mechanisms of the EutL and EutM shell proteins of the ethanolamine-utilizing
573 microcompartment of Escherichia coli. *Journal of bacteriology 192*, 6056–63.
574 (21) Cheng, S., Sinha, S., Fan, C., Liu, Y., and Bobik, T. a. (2011) Genetic analysis of the protein shell of
575 the microcompartments involved in coenzyme B12-dependent 1,2-propanediol degradation by
576 Salmonella. *Journal of bacteriology 193*, 1385–92.
577 (22) Crowley, C. S., Cascio, D., Sawaya, M. R., Kopstein, J. S., Bobik, T. a, and Yeates, T. O. (2010)
578 Structural insight into the mechanisms of transport across the Salmonella enterica Pdu microcompartment
579 shell. *The Journal of biological chemistry 285*, 37838–46.

(23) Chowdhury, C., Chun, S., Pang, A., Sawaya, M. R., Sinha, S., Yeates, T. O., and Bobik, T. A. (2015) Selective molecular transport through the protein shell of a bacterial microcompartment organelle. *Proceedings of the National Academy of Sciences 112*, 2990–2995.

(24) Dou, Z., Heinhorst, S., Williams, E. B., Murin, C. D., Shively, J. M., and Cannon, G. C. (2008) CO2 fixation kinetics of Halothiobacillus neapolitanus mutant carboxysomes lacking carbonic anhydrase suggest the shell acts as a diffusional barrier for CO2. *Journal of Biological Chemistry 283*, 10377–10384.

(25) Cai, F., Sutter, M., Bernstein, S. L., Kinney, J. N., and Kerfeld, C. A. (2014) Engineering bacterial microcompartment shells: chimeric shell proteins and chimeric carboxysome shells. *ACS synthetic biology 4*, 444–453.

(26) Sturms, R., Streauslin, N. A., Cheng, S., and Bobik, T. A. (2015) In Salmonella enterica, ethanolamine utilization is repressed by 1,2-propanediol to prevent detrimental mixing of components of two different bacterial microcompartments. *Journal of Bacteriology* (Metcalf, W. W., Ed.) *197*, 2412–2421.

(27) Sinha, S., Cheng, S., Sung, Y. W., McNamara, D. E., Sawaya, M. R., Yeates, T. O., and Bobik, T. A. (2014) Alanine scanning mutagenesis identifies an asparagine-arginine-lysine triad essential to assembly of the shell of the pdu microcompartment. *Journal of molecular biology 426*, 2328–45.

(28) 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. *Molecular Cell 38*, 305–315.

(29) 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.

(30) 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., and Warren, M. J. (2014) Solution structure of a bacterial microcompartment targeting peptide and its application in the construction of an ethanol bioreactor. *ACS Synthetic Biology 3*, 454–465.

606 (31) Kim, E. Y., and Tullman-Ercek, D. (2014) A rapid flow cytometry assay for the relative
607 quantification of protein encapsulation into bacterial microcompartments. Biotechnology Journal 9, 348–
608 354.
609 (32) Jakobson, C. M., Kim, E. Y., Slininger, M. F., Chien, A., and Tullman-Ercek, D. (2015) Localization
610 of proteins to the 1,2-propanediol utilization microcompartment by non-native signal sequences is
611 mediated by a common hydrophobic motif. *Journal of Biological Chemistry 290*, 24519–24533.
612 (33) Fan, C., Cheng, S., Sinha, S., and Bobik, T. A. (2012) Interactions between the termini of lumen
613 enzymes and shell proteins mediate enzyme encapsulation into bacterial microcompartments.
614 *Proceedings of the National Academy of Sciences 109*, 14995–15000.
615 (34) Savage, D. F., Afonso, B., Chen, A. H., and Silver, P. a. (2010) Spatially ordered dynamics of the
616 bacterial carbon fixation machinery. *Science 327*, 1258–61.
617 (35) Carpenter, A. E., Jones, T. R., Lamprecht, M. R., Clarke, C., Kang, I. H., Friman, O., Guertin, D. A.,
618 Chang, J. H., Lindquist, R. A., Moffat, J., Golland, P., and Sabatini, D. M. (2006) CellProfiler: image
619 analysis software for identifying and quantifying cell phenotypes. *Genome Biol. 7*, R100.
620 (36) Pang, A., Frank, S., Brown, I., Warren, M. J., and Pickersgill, R. W. (2014) Structural insights into
621 higher order assembly and function of the bacterial microcompartment protein PduA. *Journal of*
622 *Biological Chemistry 289*, 22377–22384.
623 (37) Noël, C. R., Cai, F., and Kerfeld, C. A. (2016) Purification and characterization of protein nanotubes
624 assembled from a single bacterial microcompartment shell subunit. *Advanced Materials Interfaces 3*,
625 1500295.
626 (38) Havemann, G. D., Sampson, E. M., and Bobik, T. A. (2002) PduA Is a shell protein of polyhedral
627 organelles involved in coenzyme B12-dependent degradation of 1,2-propanediol in Salmonella enterica
628 Serovar Typhimurium LT2. *Journal of Bacteriology 184*, 1253–1261.
629 (39) Glasgow, J. E., Asensio, M. A., Jakobson, C. M., Francis, M. B., and Tullman-Ercek, D. (2015)
630 Influence of electrostatics on small molecule flux through a protein nanoreactor. *ACS Synthetic Biology 4*,
631 1011–1019.

632 (40) Welte, W., Nestel, U., Wacker, T., and Diederichs, K. (1995) Structure and function of the porin
633 channel. *Kidney International 48*, 930–940.
634 (41) Mayrose, I. (2004) Comparison of site-specific rate-inference methods for protein sequences:
635 empirical bayesian methods are superior. *Molecular Biology and Evolution 21*, 1781–1791.
636 (42) Engler, C., Gruetzner, R., Kandzia, R., and Marillonnet, S. (2009) Golden gate shuffling: a one-pot
637 dna shuffling method based on type IIs restriction enzymes. *PLoS ONE* (Peccoud, J., Ed.) *4*, e5553.
638 (43) Datta, S., Costantino, N., and Court, D. L. (2006) A set of recombineering plasmids for gram-
639 negative bacteria. *Gene 379*, 109–115.
640 (44) Sinha, S., Cheng, S., Fan, C., and Bobik, T. a. (2012) The PduM protein is a structural component of
641 the microcompartments involved in coenzyme B(12)-dependent 1,2-propanediol degradation by
642 Salmonella enterica. *Journal of bacteriology 194*, 1912–8.

643

644 **Tables**

645 Table 1. Doubling times of PduA mutants for growth on 1,2-PD. Both *ΔpduA::eutM* and *pduA*^{K37Q} strains
646 are statistically different than the wildtype strain by two tailed T-test. The *ΔpduA::eutM* and *pduA*^{K37Q}
647 strains are not significantly different from each other by two tailed T-test. n=7 for all cases.

Strain	Doubling time (hr)	Standard error
MFSS044, Wildtype	7.85	±0.89
MFSS252, *ΔpduA::eutM*	5.73	±0.29
MFSS238, *pduA*^{K37Q}	5.75	±0.49

648

649 Table 2. Bacterial strains used in this study
650

Strain	Organism	Genotype
MFSS044	*S. enterica* serovar Typhimurium	LT2 wild type
MFSS263	*S. enterica* serovar Typhimurium	LT2 *pduA*^{K26A}
MFSS236	*S. enterica* serovar Typhimurium	LT2 Δ*pduA::cat/sacb*
MFSS252	*S. enterica* serovar Typhimurium	LT2 Δ*pduA::eutM*

MFSS273	*S. enterica* serovar Typhimurium	LT2 Δ*pduA::eutM-FLAG*
MFSS238	*S. enterica* serovar Typhimurium	LT2 *pduA* ^{K37Q}
MFSS393	*S. enterica* serovar Typhimurium	LT2 Δ*pduA:: pduA* ^{K37Q}*-FLAG*
MFSS392	*S. enterica* serovar Typhimurium	LT2 Δ*pduA:: pduA-FLAG*
MFSS337	*S. enterica* serovar Typhimurium	LT2 Δ*pduA::eutM* ^{K24A}
MFSS320	*S. enterica* serovar Typhimurium	LT2 *pduA* ^{K37H}
MFSS321	*S. enterica* serovar Typhimurium	LT2 *pduA* ^{K37C}
MFSS322	*S. enterica* serovar Typhimurium	LT2 *pduA* ^{K37R}
MFSS323	*S. enterica* serovar Typhimurium	LT2 *pduA* ^{K37G}
MFSS324	*S. enterica* serovar Typhimurium	LT2 *pduA* ^{K37P}
MFSS325	*S. enterica* serovar Typhimurium	LT2 *pduA* ^{K37V}
MFSS326	*S. enterica* serovar Typhimurium	LT2 *pduA* ^{K37F}
MFSS327	*S. enterica* serovar Typhimurium	LT2 *pduA* ^{K37T}
MFSS328	*S. enterica* serovar Typhimurium	LT2 *pduA* ^{K37L}
MFSS329	*S. enterica* serovar Typhimurium	LT2 *pduA* ^{K37D}
MFSS330	*S. enterica* serovar Typhimurium	LT2 *pduA* ^{K37E}
MFSS331	*S. enterica* serovar Typhimurium	LT2 *pduA* ^{K37S}
MFSS335	*S. enterica* serovar Typhimurium	LT2 *pduA* ^{K37N}
MFSS336	*S. enterica* serovar Typhimurium	LT2 *pduA* ^{K37W}
TUC01[a]	*E. coli*	DH10B *cat/sacB*

[a]Datta et al.[36]

Figure Captions

Figure 1. Organization and modifications of Pdu MCP

(a) Diagram illustrating the Pdu pathway including enzymes, cofactors, substrates, intermediates and products. The black hexagonal outline represents the boundary of the MCP. Enzymes outside of this boundary are not localized to the interior of the MCP.[39] (b) Schematic demonstrating the hypothesized overall structure of the MCP. Arrows indicate replacement of the PduA shell protein (red) with EutM (yellow). Crystal structures (PDB IDs: 3ngk, 3mpy) to the right illustrate structural similarity.[14,17] (c) Left hand images are ribbon structures of the pores of PduA (red) and EutM (yellow). Residues PduA K37 and the corresponding glutamine of the EutM pore are highlighted in green. To the right are the surface models of the respective pore regions colored by the electrostatic potential. Blue is positive potential, white is neutral, and red is negative potential. (d) A diagram of the Pdu operon and the operon resulting

663 from recombineering to substitute eutM for *pduA* on the *S. enterica* chromosome. Other Pdu shell
664 proteins are blue arrows, enzymes of the pathway and proteins of unknown function are grey arrows.

665

666 Figure 2. EutM incorporates into the Pdu MCP

667 (a, b, c, d) Phase contrast images of *S. enterica* strains LT2 (*wild type*), LT2 Δ*pduA::eutM*, LT2 *pduA*^{K26A},
668 and LT2 Δ*pduA::eutM* ^{K24A}, respectively. All strains are co-expressing Pdu MCPs and PduD$^{1-20}$-GFP-ssrA.
669 (e, f, g, h) GFP fluorescence microscopy images corresponding to phase contrast images a-d. Large
670 filamentous aggregates are indicated with blue arrows in panels g and h. (i, j) Immunostained
671 transmission electron microscopy of MCPs purified from wild type LT2 and LT2 Δ*pduA::eutM-FLAG*,
672 respectively. Bound IgG-conjugated gold particles are indicated with red arrows. (k, l) Transmission
673 electron microscopy images of MCPs purified from the strains depicted in images c, d. For convenience,
674 larger versions of the images in panels (i-l) are provided in Supplemental Figure S16. (m) Coomassie-
675 stained SDS-PAGE of purified MCPs. Lane (i) is a purification from strain LT2 wild type, lane (ii) is
676 from strain LT2 *pduA* ^{K37Q}, lane (iii) is from strain LT2 Δ*pduA::eutM-FLAG*. (n) α-FLAG western blot of
677 purified compartments.

678

679 Figure 3. Strains harboring shell protein variants grow on 55 mM 1,2-PD and 150 nM coenzyme B12

680 MCP variant strains were grown on 55 mM 1,2-PD with 150 nM coenzyme B_{12}. Wild-type strain LT2 is
681 compared to LT2 Δ*pduA::eutM* , *pduA*^{K37Q} , *pduA*^{K26A} , Δ*pduA::eutM* ^{K24A}, and Δ*pduA::cat/sacb* strains.
682 Mean optical density at 600 nm is plotted vs time. Error bars represent 95% confidence interval from six
683 measurements started on four different days.

684

685 Figure 4. HPLC analysis of MCP-forming strains during growth on 1,2-PD and coenzyme B_{12}

686 The concentrations of 1,2-PD and products of the Pdu pathway were monitored during growth on 55 mM
687 1,2-PD and 150 nM coenzyme B_{12}. Mean values of metabolites are plotted vs time. Metabolites measured
688 are 1,2-PD, 1-propanol, propionate, and propionaldehyde, as labeled. In the left hand column (a, c, e, g)
689 the wild-type *S. enterica* LT2 is compared to LT2 Δ*pduA::eutM*. In the right hand column (b, d, f, h) the
690 wild-type strain is compared to LT2 *pduA* ^{K37Q}. Error bars represent standard error from six measurements
691 started on four different days.

692

693 Figure 5. HPLC analysis of non-assembling shell protein mutant strains

694 The concentrations of 1,2-PD and products of the Pdu pathway were monitored during growth on 55 mM
695 1,2-PD and 150 nM coenzyme B_{12}. Mean values of metabolites are plotted vs time. Metabolites measured
696 are 1,2-PD, 1-propanol, propionate, and propionaldehyde, as labeled. The wild-type LT2 strain is
697 compared to LT2 *pduA* ^{K26A}and LT2 Δ*pduA::eutM* ^{K24A} strains. Error bars represent standard error from six
698 measurements started on four different days.

699

700 Figure 6. A library of amino acids at residue 37 of PduA confers a variety of growth effects

701 (a) Mean values of final OD (600nm) are plotted for various PduA mutants from growth on 1,2-PD with
702 150 nm coenzyme B_{12}. Error bars represent standard error from four different measurements started on
703 separate days. The red line highlights wild-type growth for comparison. (b) Plot of final OD vs pK_a of
704 charged amino acids substituted at position 37 of PduA. (c) Plot of final OD vs hydropathy index of
705 amino acids substituted at position 37 of PduA.

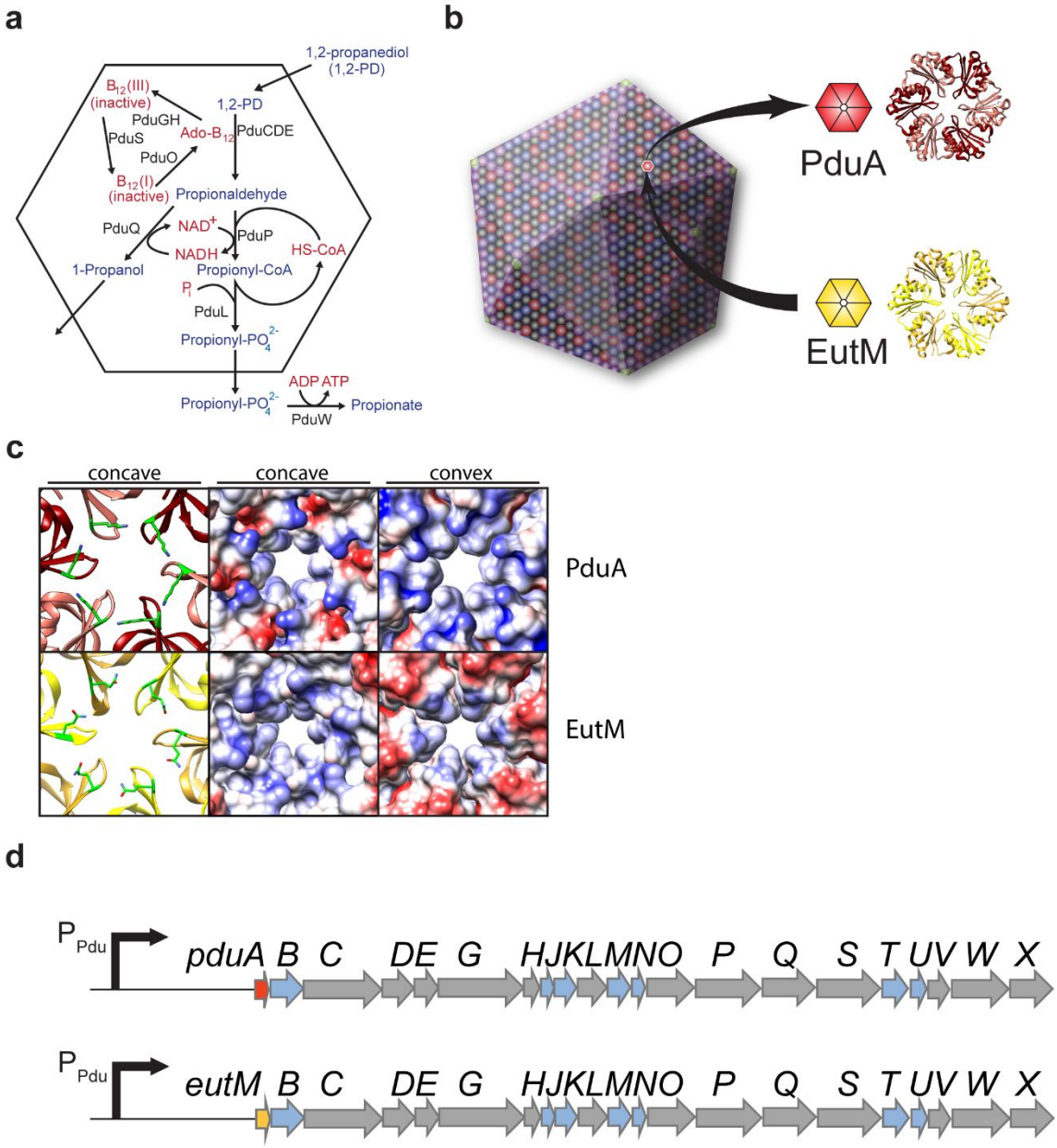
706

707 Figure 7. Phylogenetic analysis of PduA homologs

(a) Rate4Site evolutionary rate scaling factor mapped over the crystal structure of PduA. White represents most conserved residues. Dark red represents residues with high evolution rates. Some residues with highest and lowest evolutionary rates are labeled for reference. (b) Histogram showing the % frequency of each amino acid aligning with PduA K37 in an MSA of n=1000 homolog sequences.

712 **Figures**

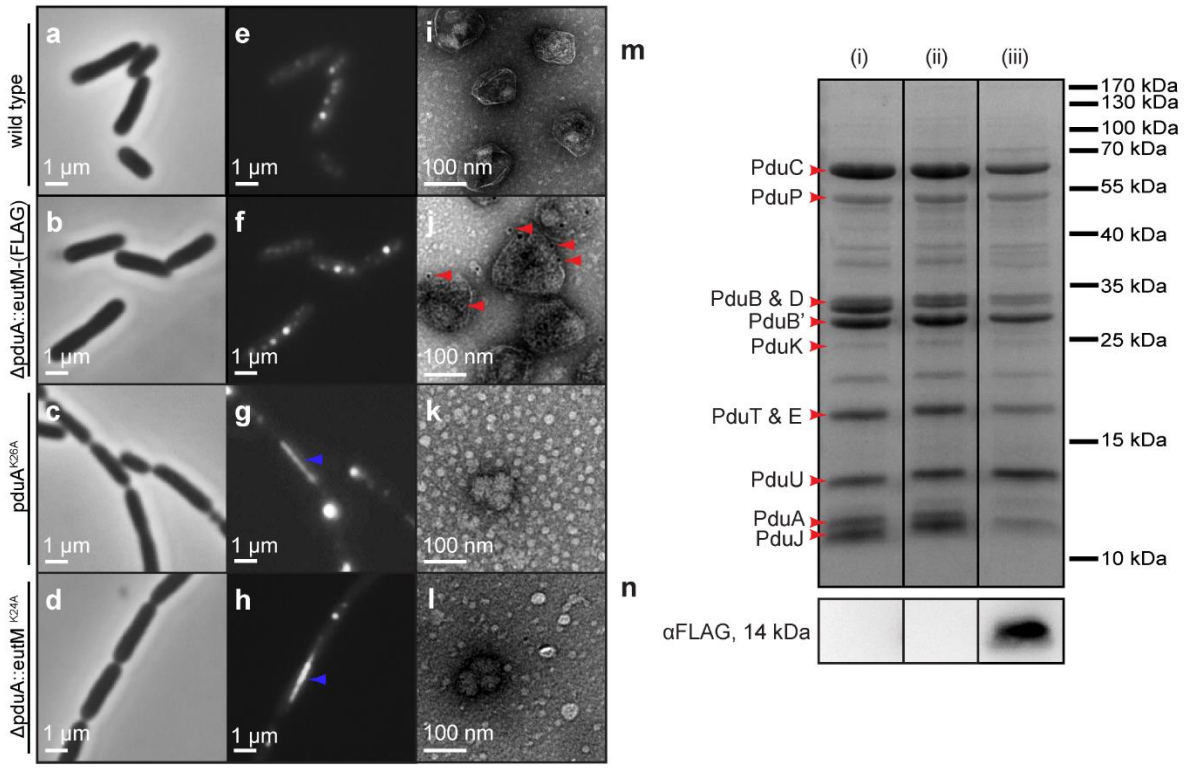
713 Figure 1.



714

715

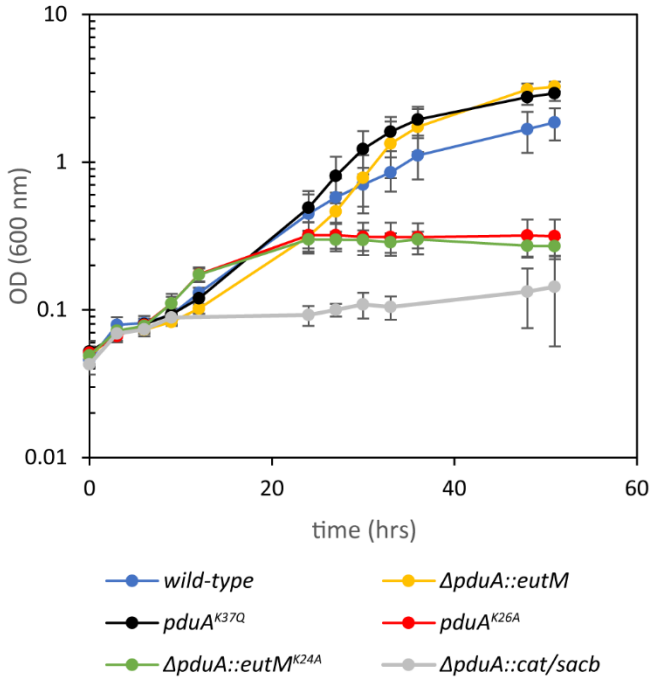
716 Figure 2.



717

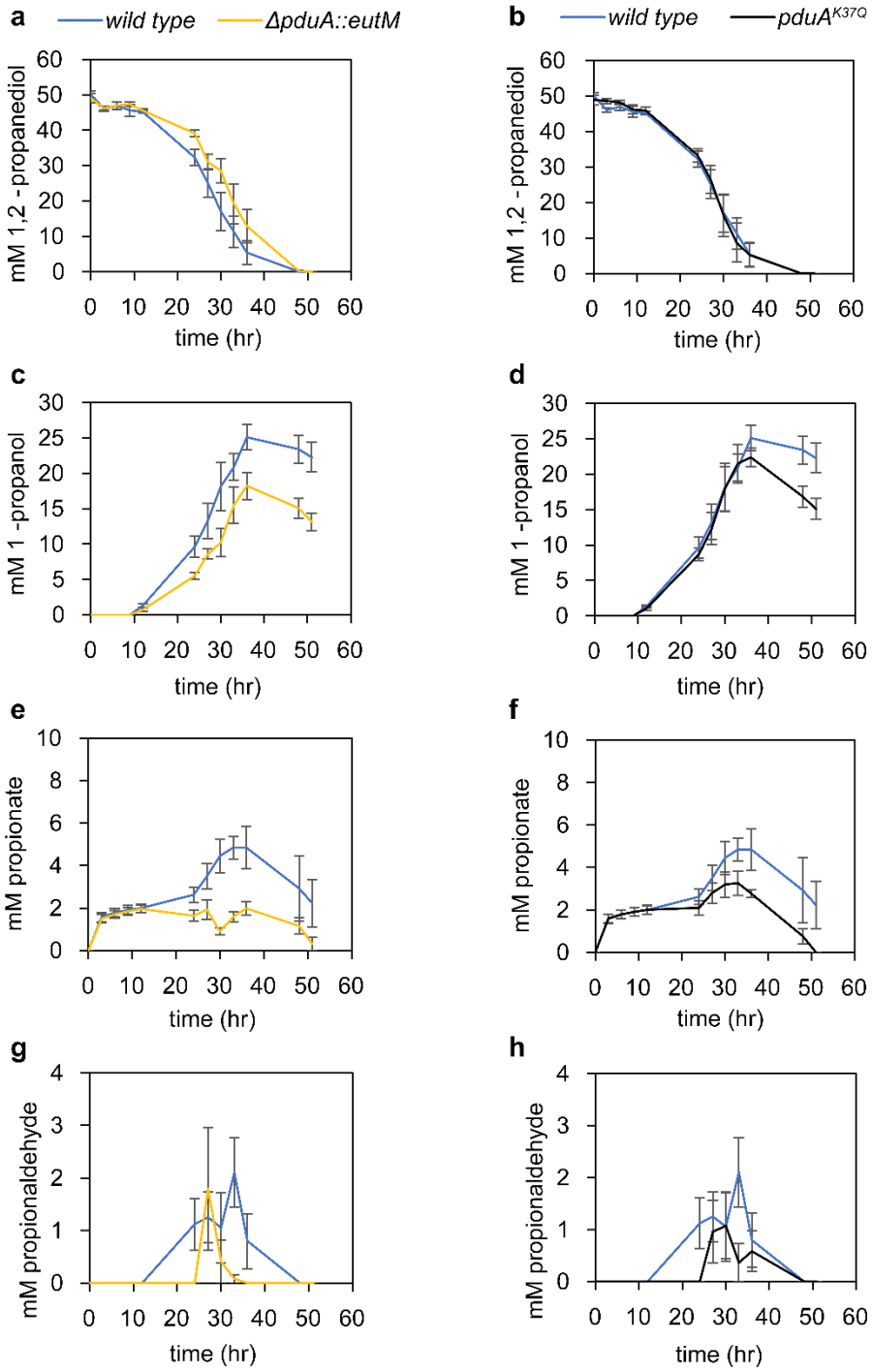
718

719 Figure 3.



720

721 Figure 4.

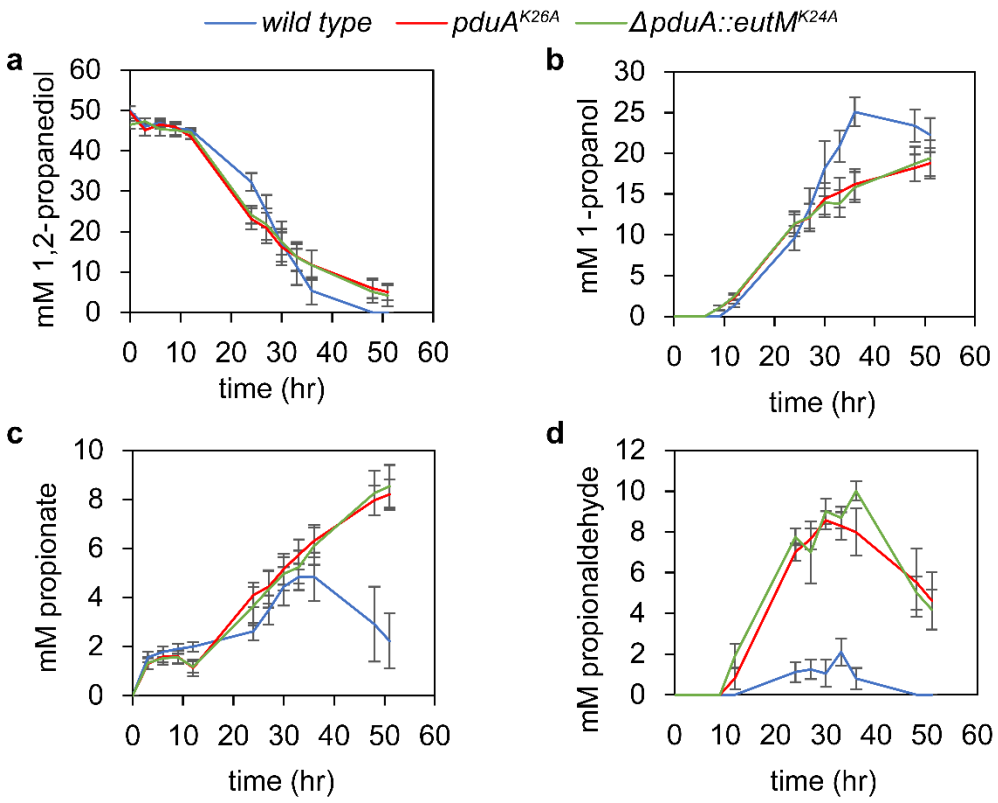


722

723

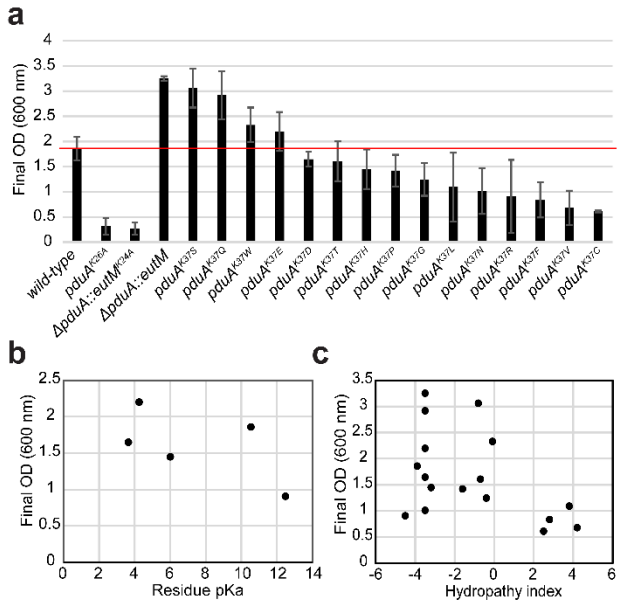
724

725 Figure 5.



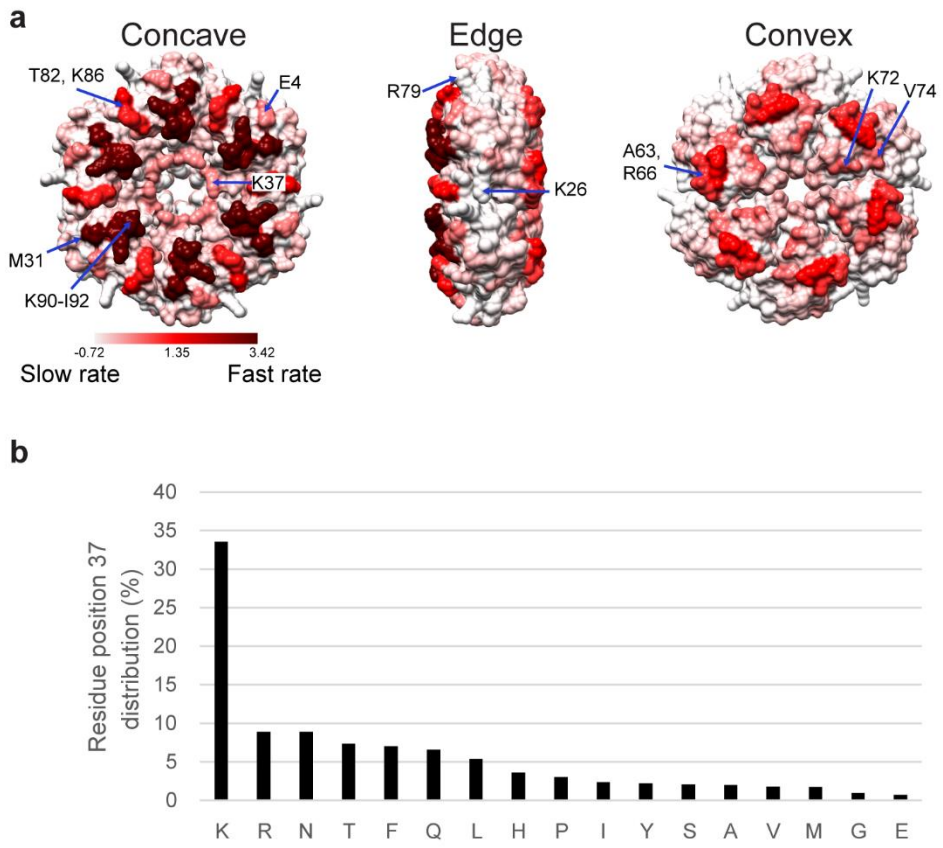
726

727 Figure 6.



728

729 Figure 7.



730