Utilisation of bacterial microcompartment technology to enhance spatial organisation in *Escherichia coli*

A thesis submitted to the University of Kent for the degree of PhD in the Faculty of Sciences

2017

Matthew John Lee

Abstract

Enhancing the catalytic efficiency of enzymes has been an aspiration for many years and protein engineering has been quite successful in this area. However, researchers are now turning their attention to the organisation of enzymes within cells to enhance whole pathway productivity. As such, bacterial microcompartments (BMCs) have attracted significant interest due to their ability to not only concentrate enzymes and metabolites but also prevent the release of toxic intermediates into the cell owing to their semi-permeable protein shell. The aim of the research described in this thesis was to develop a number of mechanisms to enhance the spatial organisation of proteins within *E. coli* and thereby increase metabolic productivity.

The ability of BMC encapsulation peptides to target foreign proteins to a BMC was initially investigated. However, we found that the fusion of BMC encapsulation peptides to the enzymes of a 1,2-propanediol (1,2-PD) synthesis pathway resulted in the aggregation of the tagged enzymes rather than their localisation to a BMC. Nonetheless, the tagged enzymes were found to produce significantly more 1,2-PD than strains producing untagged enzymes, irrespective of the presence of BMCs, suggesting that it is the enzyme aggregate that results in enhanced substrate channelling. Furthermore, we showed that a highly conserved hydrophobic motif is conserved across a wide range of BMC targeting peptides and suggest that it is the amphipathic nature of these peptides that results in aggregation of tagged components.

The ability of PduA, a component of the BMC shell, to form filamentous structures was investigated for its potential as a cytoplasmic scaffold. Targeting proteins to the cytoscaffold was explored through the use of *de novo* designed coiled-coil peptides. Localising the enzymes pyruvate decarboxylase and alcohol dehydrogenase on this scaffold by tagging them with the cognate coiled-coil peptide resulted in a significant enhancement in ethanol production in comparison to a strain lacking the scaffold. Additionally, we showed that it was possible to locate this intracellular scaffold to the inner membrane of the cell, further demonstrating the flexibility of this system.

Finally, we utilised the same coiled-coil peptide technology to target fluorescent proteins to BMCs, overcoming the aggregation behaviour of the native targeting peptides. Through the re-design of a BMC shell protein we showed targeting occurs to both the external and luminal faces of these BMCs. Taken together, the evidence presented in this thesis provides a number of alternative strategies to not only enhance spatial organisation within a cell but also to increase productivity of engineered metabolic pathways.

Acknowledgements

Over the past three years I have received help and guidance from a number of people to whom I am most grateful.

First I would like to thank my supervisors Professor Martin Warren and Dr Stefanie Frank for not only giving me the opportunity to complete this PhD but also guiding me through the past three years of study. Your help has been invaluable; the work in this thesis would not have been possible without your supervision.

I am grateful to the whole of the Warren lab for their help and support particularly to Rokas and Joe who have made the whole experience so enjoyable. I would like to thank collaborators that have contributed to the work within this thesis both at Kent and elsewhere. I would like to particularly thank Ian Brown for all his work helping me with TEM and confocal imaging.

Thanks must go to Nat for the endless love and support that you have given me, not forgetting the endless supply of baked goods, you have made the past two years the best years of my life.

Finally to my parents, who have always supported me, thank you.

Table of contents

Abbreviations

1,2-PD	1,2-Propanediol
3-PGA	3-Phosphoglycerate
Ado-B12	Adenosylcobalamin
ADP	Adenosine diphosphate
ATP	Adenosine triphosphate
BMC	Bacterial microcompartment
C. freundii	*Citrobacter freundii*
CA	Carbonic anhydrase
CC-Di-A	Acidic coiled-coil peptide
CC-Di-B	Basic coiled-coil peptide
CCM	CO_2 concentrating mechanism
D18	Targeting peptide from PduD
E. coli	*Escherichia coli*
Eut	Ethanolamine utilisation
Fe-S	Iron-sulfur cluster
GFP	Green fluorescent protein
H. neapolitanus	*Halothiobacillus neapolitanus*
H. ochraceum	*Haliangium ochraceum*
HS-CoA	Coenzyme A
M. smegmatis	*Mycobacterium smegmatis*
N. benthamiana	*Nicotiana benthamiana*
NAD+	Nicotinamide adenine dinucleotide
NADH	Nicotinamide adenine dinucleotide (NAD) + hydrogen (H)
OH-B12	Hydroxocobalamin
P18	Targeting peptide from PduP
Pdu	Propanediol utilisation
Pi	Inorganic phosphate
RuBisCO	Ribulose-1,5-bisphosphate carboxylase/oxygenase
RuBP	Ribulose 1,5-bisphosphate
S. enterica	*Salmonella enterica*
TEM	Transmission electron microscopy
YFP	Yellow fluorescent protein
YPER	Yeast protein extraction reagent

Chapter 1

Introduction

The greater complexity of eukaryotes is reflected in a higher level of sub-cellular organisation with specialised membrane bound organelles and a well-structured cytoskeleton in comparison to prokaryotic organisms. With this in mind bacteria were thought of as comparatively simple organisms, whilst the cellular compartmentalisation and organisation of eukaryotes permitted more complex multicellular life. The traditional view that prokaryotes represent a bag of enzymes that carry out metabolic processes by simple diffusion of metabolites between enzymes has persisted for many years. However, significant advances in prokaryotic cell biology, genome sequencing and improved imaging techniques have revealed that prokaryotes are significantly more organised than previously thought.

1.1. Metabolic organisation in prokaryotes

Several lines of evidence have led to the idea that metabolism, at least, is more organised in prokaryotes than was appreciated from early biochemical studies. Organisation of enzymes within metabolic pathways can lead to enhanced flux. This can be achieved by enzymes forming complexes, which has the effect of increasing the local concentration of substrates, reducing toxicity of pathway intermediates and minimizing the loss of intermediates to side reactions. These processes have given rise to terms such as substrate channelling – the transfer of a substrate from one active site to another without first diffusing into the bulk cytoplasmic environment (Miles *et al.*, 1999; Spivey and Ovádi, 1999). Substrate channelling can occur by two mechanisms; the direct transfer of an intermediate from the catalytic site of one enzyme to that of another or by an increased local concentration of enzymes by specific localization (Spivey and Merz, 1989).

1.2. Multienzyme complexes

Complexes between multiple enzymes are seen frequently in nature. Such complexes allow for the direct channelling of pathway intermediates from one active site to another without loss to side reactions, such complexes can also protect against toxicity. For example, tryptophan synthase, catalysing the last step in tryptophan synthesis is a α_2 β_2 complex. The crystal structure of this complex reveals a hydrophobic tunnel that likely allows for the channelling of the toxic indole intermediate from the active site of the α subunit to the active site of the β subunit,

thereby preventing loss of this intermediate (Hyde *et al.*, 1988; Hyde and Miles, 1990). Additionally, multienzyme complexes may be specifically localised within the cell resulting in a further enhancement to catalytic activity. Such complexes between multiple sequential enzymes and structural elements i.e. the cytoskeleton or membranes have been referred to as metabolons (Srere, 1985).

1.3. Bacterial microcompartments

Prokaryotic cells were long thought to have no internal cellular compartments as observed in eukaryotic cells. However, with the introduction of electron microscopy and subsequent improvements to this technology it has been revealed that prokaryotes do indeed contain intracellular compartments (Drews and Niklowitz, 1956). These compartments, named bacterial microcompartments (BMCs), are proteinaceous organelles that consist of a semi-permeable outer shell encasing the enzymes associated with a specific metabolic process (Yeates *et al.*, 2008; Kerfeld *et al.*, 2010; Chowdhury *et al.*, 2014). It has been shown that BMC shells are composed from several thousand individual polypeptides with an overall mass of between 100 and 600 MDa, therefore representing the largest protein complexes found in bacteria (Yeates *et al.*, 2008; Cheng *et al.*, 2008).

The requirement for compartmentalisation of metabolic pathways has been extensively discussed since their first visualisation; to date two main theories have been put forward to explain their existence (Cheng *et al.*, 2008; Kerfeld *et al.*, 2010).

One proposed advantage to compartmentalization is the increased local concentration of enzymes and substrates. The volume of a single *E. coli* cell is known to be approximately 1 μm^3 resulting in the concentration of a single molecule in a cell being ~ 1 nM. The predicted volume of a 120 nm diameter BMC is 8.9×10^{-4} μm^3 approximately 1000 times smaller than the volume of a cell (Sargent *et al.*, 2013). This decreased volume results in a high local concentration of molecules within the BMC lumen, with a single molecule within a BMC being approximately 1 µM. The increased local concentration may allow for increased substrate channelling though encapsulated multi-enzyme pathways in comparison to soluble enzymes. Indeed a recent computational study has revealed that the main advantage of encapsulation of a pathway is the increased luminal intermediate concentrations in comparison to cytosolic concentrations therefore leading to enhance metabolic flux (Jakobson *et al.*, 2017b).

The second proposed hypothesis is that the shell of the BMC provides a barrier that prevents the release of damaging toxic intermediates and the loss of intermediates to side reactions. Genetic analysis reveals that a common feature of BMCs is that they encapsulate a pathway involving the production and consumption of a toxic or volatile intermediate further supporting this hypothesis (Penrod and Roth, 2006; Dou *et al.*, 2008; Sampson and Bobik, 2008; Yeates *et al.*, 2010).

1.3.1. BMCs in nature

A recent genomics study has revealed that approximately 17% of bacteria contain a BMC locus (Axen *et al.*, 2014). The widespread distribution of bacterial microcompartments across 23 phyla suggests that they are transferred by lateral transfer of BMC operons, thus providing an immediate advantage to the host rather than the transfer of individual genes (Lawrence and Roth, 1996; Bobik *et al.*, 1999; Hess *et al.*, 2001; Axen *et al.*, 2014).

It is understood that two main classes of BMCs exist; anabolic carboxysomes that fix carbon and catabolic metabolosomes that contain a signature aldehyde-generating enzyme (Axen *et al.*, 2014).

1.3.2. Anabolic BMCs, the α and β carboxysome

Carboxysomes were the first intracellular BMC to be visualised and remain the best characterised of the BMC systems in terms of both structural information and knowledge of the encapsulated pathway (Drews and Niklowitz, 1956). These macromolecular structures contain the enzymes required to fix CO_2, ribulose-1,5-bisphosphate carboxylase oxygenase (RuBisCO) and carbonic anhydrase (CA) (Shively *et al.*, 1973; Cannon and Shively, 1983; Fukuzawa *et al.*, 1992). The enzymes are encased within the semi-permeable protein shell of the BMC (Kerfeld *et al.*, 2005; Iancu *et al.*, 2007). It has been estimated that carboxysomes account for approximately 25% of global carbon fixation (Rae *et al.*, 2013). Genetic analysis has revealed that carboxysome genes are present in all cyanobacteria and some chemoautotrophic bacteria (Yeates *et al.*, 2010).

The evolution of carboxysomes as part of the CO_2 concentrating mechanism (CCM) is thought to have been driven by changes in atmospheric conditions. Approximately 350 million years ago there was a large drop in atmospheric CO_2 levels and an increase in O_2 levels (Bergman *et al.*, 2004; Berner, 2006). This change in atmospheric conditions is proposed to be the driving force for the development of the CO_2 concentrating mechanism (CCM) to overcome the inefficient activity of RuBisCO (Badger and Price, 2003).

Carboxysomes can be split into two classes; α, encoded by the *cso* operon and β encoded by the *ccm* operon. The two classes are distinguished by their genetic makeup, shell composition and the enzymes that they encapsulate (Rae *et al.*, 2013). Phylogenetically, α-carboxysomes are found in cyanobacteria (termed α-cyanobacteria) and some chemoautotrophic bacteria whereas β-carboxysomes are found solely in cyanobacteria (termed β-cyanobacteria) (Rae *et al.*, 2013; Axen *et al.*, 2014). The genetic information encoding α-carboxysomes is organised into an operon whereas β-carboxysome genes are dispersed throughout the genome (Cannon *et al.*, 2001). Both α- and β-carboxysomes form polyhedral structures *in-vivo*, however, β carboxysomes are known to be larger and more varied in size (175 – 500 nm in diameter) compared to α-carboxysomes, which have an average size of 120 nm (Figure 1) (Shively *et al.*, 1977; van Eykelenburg, 1980; Long *et al.*, 2007; Roberts *et al.*, 2012; Rae *et al.*, 2013).

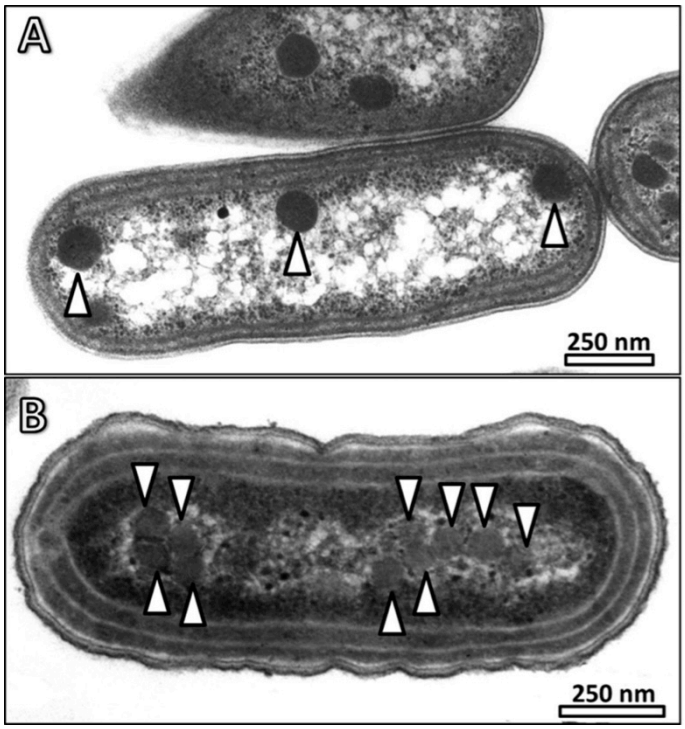


Figure 1. TEM analysis of strains producing carboxysomes. (A) β-Carboxysomes present in *Synechococcus elongatus* PCC 7942. (B) α-Carboxysomes present in *Cyanobium* PCC 7001. Arrows indicate carboxysome structures (Rae *et al.*, 2013).

The process of carbon fixation begins with the uptake of dissolved inorganic carbon in the form of bicarbonate and CO_2 into the cytoplasm as bicarbonate by a number of membrane transporters (Price, 2011). It is thought that bicarbonate diffuses throughout the cytoplasm and into the carboxysome through pores in the BMC shell (Kerfeld *et al.*, 2005). The positive charge of central pore of hexameric shell proteins has been suggested to enhance transport of bicarbonate across the shell of the BMC whereas uncharged O_2 and CO_2 do not confer this advantage (Kerfeld *et al.*, 2005; Tsai *et al.*, 2007; Yeates *et al.*, 2008).

Carbonic anhydrase then dehydrates bicarbonate to CO_2 within the lumen of the BMC, which is converted with ribulose 1,5-bisphosphate (RuBP) by RuBisCO into two molecules of 3-phosphoglyceric acid (3-PGA), which diffuses out of the carboxysome and enters central metabolism. The production of CO_2 within the lumen of the BMC is thought to overcome the slow catalytic activity of RuBisCO (k_{cat} 11.6 sec^{-1}) by increasing the local concentration of its substrate (Morell *et al.*, 1994; Tcherkez *et al.*, 2006). This high local concentration of CO_2 is also thought to inhibit the reaction of RuBisCO with O_2 which leads to inefficient photorespiration (Moroney *et al.*, 2013). The shell of the carboxysome is known to prevent the leakage of CO_2

into the cytoplasm, and therefore its release into the environment due to the poor retention of CO_2 by lipid bilayers (Price and Badger, 1989; Dou *et al.*, 2008). It has also been shown that RuBisCO is less inhibited by O_2 within a carboxysome shell in comparison to a cytoplasmic RuBisCO suggesting that the carboxysome provides a diffusion barrier to O_2, although how a shell differentiates between CO_2 and O_2 is unclear (Marcus *et al.*, 1992).

1.3.3. Catabolic BMCs

Catabolic BMCs, also termed metabolosomes, are involved in a variety of metabolic processes (Brinsmade *et al.*, 2005; Axen *et al.*, 2014). The BMC associated with 1,2-propanediol (1,2-PD) utilisation is, to date, the best characterized of the catabolic BMCs with both wild type and recombinant systems developed to understand the functional characteristics of these compartments (Bobik *et al.*, 1997; Bobik *et al.*, 1999; Havemann and Bobik, 2003; Parsons *et al.*, 2008; Parsons *et al.*, 2010).

1.3.3.1. 1,2-Propanediol utilization

In 1994 sequence analysis of *S. enterica* revealed a protein with a high degree of sequence similarity to the carboxysome shell protein CcmK (Chen *et al.*, 1994). This gene, *pduA*, is located in an operon involved in 1,2-PD degradation and therefore it was concluded that *S. enterica* may produce a similar structure to the carboxysome for the degradation of 1,2-PD (Chen *et al.*, 1994). Subsequently, during growth in the presence of 1,2-PD, microcompartments were visualised that were similar in size and appearance to the carboxysome (Figure 2) (Bobik *et al.*, 1999). These 100 – 200 nm diameter BMCs contain the enzymes required to convert 1,2-PD to 1-propanol or propionate via the toxic intermediate propionaldehyde (Bobik *et al.*, 1999). It has been shown that this metabolosome has an approximate mass of 600 MDa and includes approximately 18,000 polypeptides (Cheng *et al.*, 2008).

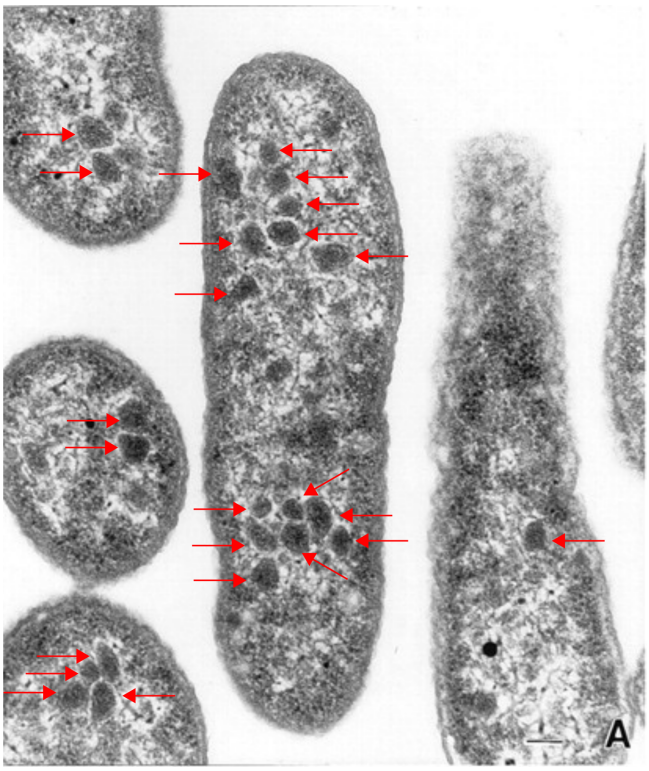


Figure 2. Electron micrograph of *S. enterica* grown aerobically on minimal, 1,2-propanediol, vitamin B12 medium. Arrows indicate BMCs. Scale bar shows 100 nm. (Bobik *et al.*, 1999).

Metabolism of 1,2-PD is thought to be of significant importance in anaerobic environments such as the intestines of higher animals where it is produced as a fermentation product of two plant sugars, fucose and rhamnose (Badia *et al.*, 1985; Obradors *et al.*, 1988). Genetic analysis has revealed that the genes required for 1,2-PD degradation are found in a number of enteric and soil dwelling bacteria including *Klebsiella*, *Lactobacillus*, *Lactococcus*, *Listeria*, *Salmonella*, *Shigella*, *Yersinia* and at least one species of *Escherichia coli* (Cheng *et al.*, 2008). It has also been suggested that 1,2-PD utilising BMCs play a role in the growth and transmission of enteric pathogens (Chowdhury *et al.*, 2014).

1.3.3.2. The *pdu* operon

The genes for the production of BMCs for the utilisation of 1,2-PD are housed on a 21-gene operon (Bobik *et al.*, 1999). *pocR* and *pduF* are located adjacent to the main *pdu* operon and encode a transcriptional regulatory protein and a 1,2-PD diffusion facilitator, respectively (Chen *et al.*, 1994). Seven of the 21 main operon genes encode 1,2-PD degrading enzymes (PduCDELPQW), nine encode BMC shell

proteins (PduABB'JKMNTU), five are associated with vitamin B_{12} recycling and reactivation of the PduCDE complex (PduGHOSX) and one (PduV) is thought to play a role in the spatial segregation of BMCs (Cheng *et al.*, 2008; Parsons *et al.*, 2010).

1.3.3.3. The pathway of 1,2-propanediol utilisation

The conversion of 1,2-PD to propionaldehyde is catalysed by the coenzyme B_{12} dependent diol dehydratase PduCDE complex (Bobik *et al.*, 1997). Propionaldehyde is then converted to either 1-propanol by 1-propanol dehydrogenase (PduQ) or to propionyl-CoA by propionaldehyde dehydrogenase (PduP) (Figure 3) (Leal *et al.*, 2003; Cheng *et al.*, 2012). Propionyl-CoA is converted by PduL, a phosphotranscyclase to propionyl-phosphate, which is subsequently converted to propionate by the propionate kinase (PduW) (Palacios *et al.*, 2003; Liu *et al.*, 2007). Under aerobic conditions some propionyl-CoA is fed into the methyl citrate pathway, a pathway that does not function under anaerobic conditions (Jeter, 1990; Horswill and Escalante-Semerena, 1997; Horswill and Escalante-Semerena, 1999).

1.3.3.4. Reactivation of the PduCDE diol dehydratase

During the first step of 1,2-PD degradation requiring Ado-B_{12} a side reaction can occur leading to the exchange of the upper Ado ligand to an OH which remains bound to the diol dehydratase complex (Honda *et al.*, 1980; Ushio *et al.*, 1982). The OH-B_{12} is removed from the complex by the PduGH reactivase and subsequently converted back to the Ado-B_{12} form by PduS and PduO (Mori *et al.*, 1997; Johnson *et al.*, 2001; Cheng and Bobik, 2010).

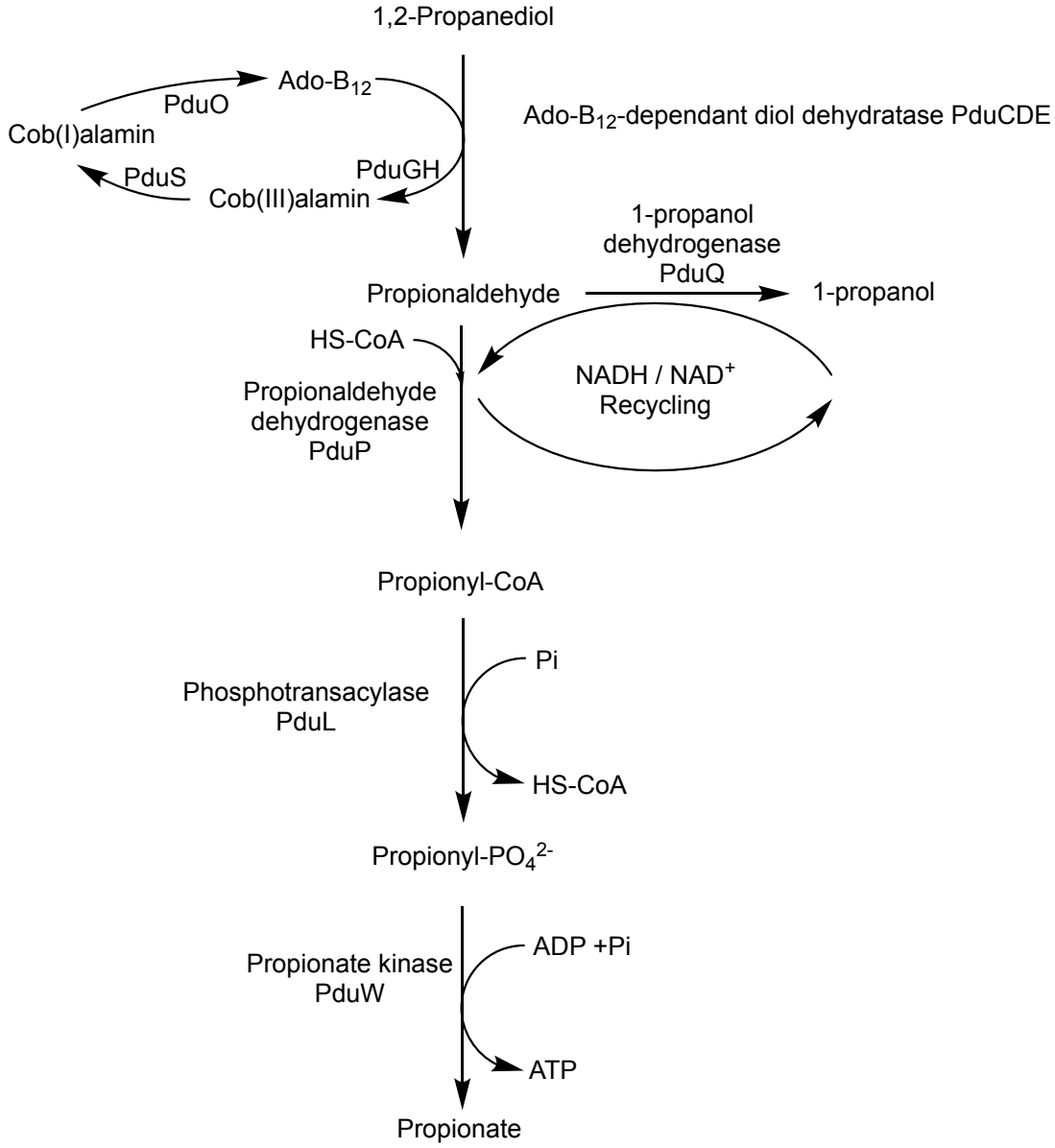


Figure 3. The pathway of 1,2-propanediol degradation.

1.3.4. Structural characteristics of BMCs

Despite the diverse functionality offered by different BMCs the proteins that form the facets of these structures are homologous. BMC shell proteins can be divided into 3 categories; hexamers (BMC-H), trimers (BMC-T) and pentamers (BMC-P). Both hexamers and trimers form the major component of the BMC shell (Havemann and Bobik, 2003). These proteins contain either a single (BMC-H) or double (BMC-T) BMC domain (pfam00936). This domain, approximately 90 amino acids in length forms a characteristic structure consisting of four antiparallel β-sheets surrounded by α-helices (Kerfeld *et al.*, 2005). Interestingly some trimeric BMC shell proteins are

known to form stacked dimers containing a central cavity (Klein *et al.*, 2009). A number of variations on this canonical structure are observed in nature; domains can be linked to form a tandem domain or permuted to create a new N and C terminus while retaining the overall BMC domain architecture (Crowley *et al.*, 2008). Numerous structural studies reveal a central pore within the hexameric disk, thought to allow for the transport of metabolites into and out of the BMC (Havemann *et al.*, 2002; Kerfeld *et al.*, 2005).

Pentameric proteins, containing a structural domain (pfam03319) are proposed to cap the vertices of the compartment. Although it must be noted that deletion of two pentameric CsoS4 proteins in *H. neapolitanus* resulted in the formation of BMCs that appear to be structurally similar to wild type however; some elongated structures were also formed (Cai *et al.*, 2009). Similarly, recombinant expression of just *H. ochraceum* hexameric and trimeric shell proteins resulted in the appearance of regularly shaped BMCs (Lassila *et al.*, 2014).

Structural and sequence analysis has revealed a strong conservation of both the hexameric and pentameric BMC domains between BMC subtypes (Figure 4) (Frank *et al.*, 2013).

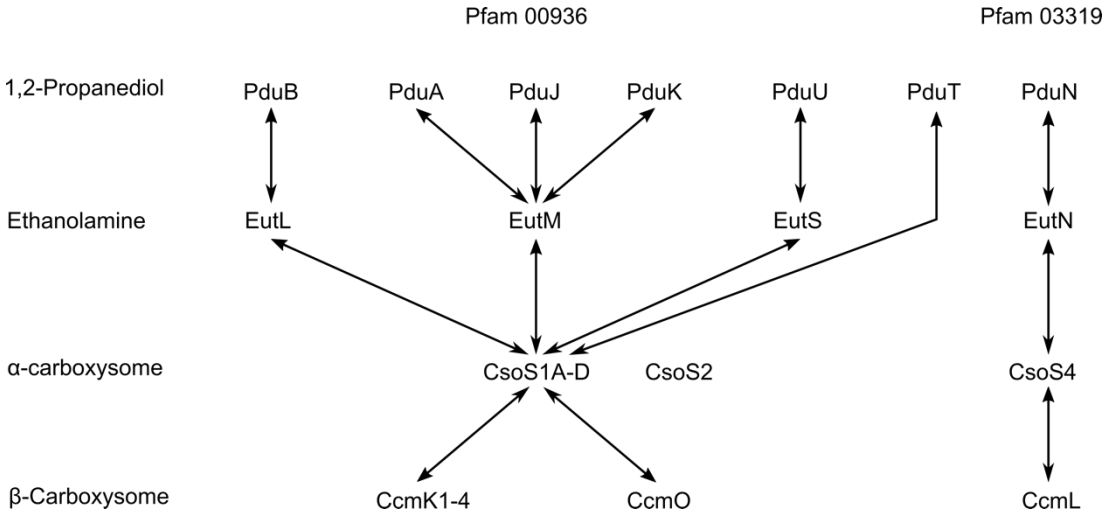


Figure 4. Relationships between shell proteins from anabolic and catabolic BMCs. Adapted from (Frank *et al.*, 2013).

Despite this wealth of knowledge regarding the individual components of the BMC shell the organisation of each of the shell proteins within the shell was not clear. Recently a previously investigated BMC of unknown function has been crystallized

revealing the organisation of shell proteins within the BMC shell (Lassila *et al.*, 2014; Sutter *et al.*, 2017). The hexamer, three trimers and one of the three pentamers were recombinantly expressed, purified and studied by cryo-electron microscopy and crystallography (Sutter *et al.*, 2017). The shell proteins assemble into a 40 nm diameter icosahedron composed of 60 hexamers, 20 trimers and 12 pentamers located at the vertices (Figure 5).

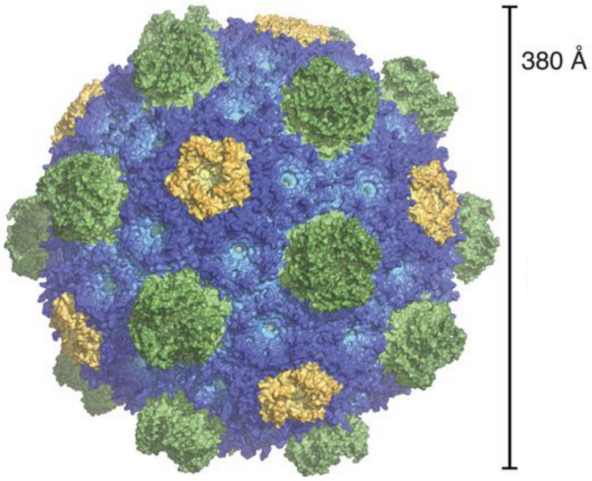


Figure 5. Crystal structure of a BMC shell of unknown function from *Haliangium ochraceum.* Hexamers – blue, trimers – green and pentamers - yellow (Sutter *et al.*, 2017).

The structure reveals four well-defined protein-protein interfaces, two hexamer-hexamer interactions, the hexamer-trimer interaction and the hexamer-pentamer interaction. Hexamers are arranged in both a coplanar orientation, in the case of those connecting pentamers or bent at 30°, when surrounding a pentamer. The hexamer-pentamer interaction is similarly bent 30°, the hexamer-trimer interaction is bent at 25°. Due to the high degree of sequence conservation between BMC types these interactions are likely to be conserved.

Both hexamers and trimers have a distinct concave and convex surface, the structure reveals that the concave side faces out, for stacked trimers the lower trimer that is embedded within the shell is orientated in the same direction. Similarly the pentameric vertex proteins are orientated with the convex side facing the lumen (Sutter *et al.*, 2017).

Given the high degree of conservation among shell proteins of all BMC types this organisation is likely to be conserved (Kerfeld and Erbilgin, 2015).

1.3.5. The shell of the 1,2-Propanediol utilizing BMC

The shell of the BMC associated with 1,2-PD utilization is composed of 9 shell proteins (Bobik *et al.*, 1999; Sinha *et al.*, 2012). The functionality of these shell proteins has been determined by mutagenesis, crystallography and sequence alignments.

Numerous studies have analyzed the size of BMCs associated with 1,2-PD utilisation; an initial study showed that BMCs in *S. enterica* were 100 – 200 nm in diameter (Bobik *et al.*, 1999). A more detailed recent study has found that purified metabolosomes from *C. freundii* are approximately 125 nm in diameter (Mayer *et al.*, 2016).

The absence of a BMC shell in the presence of 1,2-PD results in arrested growth, supporting the hypothesis that this metabolosome sequesters the toxic propionaldehyde intermediate (Havemann *et al.*, 2002; Sampson and Bobik, 2008; Cheng *et al.*, 2011).

1.3.5.1. PduA

PduA is a 93 amino acid protein containing a single BMC domain that forms a hexameric disk with a central 9 Å pore proposed to allow for the transport of the substrate, 1,2-PD into the compartment (Figure 6) (Crowley *et al.*, 2010; Pang *et al.*, 2014). Sequence analysis shows that it is related to major carboxysome shell proteins. It was initially suggested that a *pduA* deletion mutant was not able to form BMCs, however this was further shown to be a double deletion including *pduB* (Havemann *et al.*, 2002; Cheng *et al.*, 2011). Analysis of a correct *pduA* deletion mutant showed that it is not required for BMC formation although the BMCs formed are enlarged compared to wild type (Cheng *et al.*, 2011).

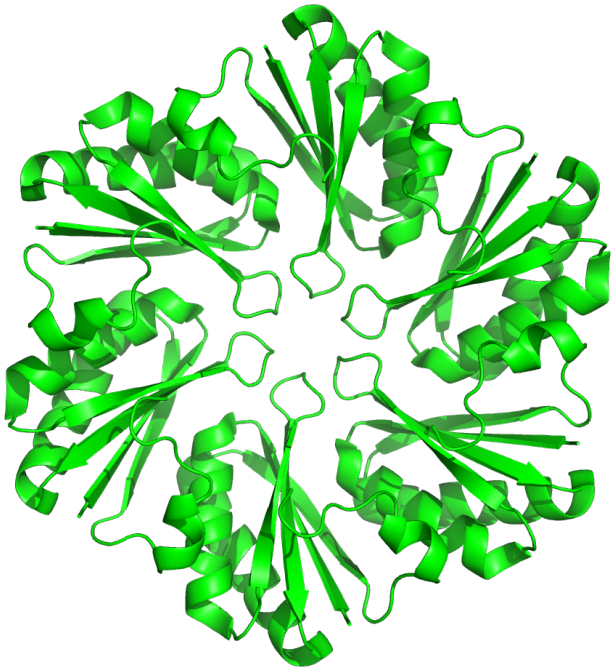


Figure 6. Crystal structure of the *S. enterica* PduA hexamer (Crowley *et al.*, 2010).

The central pore of the hexamer is thought to allow for the transport of 1,2-PD into the lumen of the BMC (Pang *et al.*, 2014). Transport is proposed to be facilitated by the hydrogen bond donors and acceptors lining the pore whereas propionaldehyde, the toxic intermediate of the 1,2-PD utilization pathway is more hydrophobic than 1,2-PD and only posses a single hydrogen bond donor, potentially limiting its escape from the BMC (Crowley *et al.*, 2010).

1.3.5.2. PduB and PduB'

PduB is a tandem BMC domain protein and has been suggested to form a gated pore (Klein *et al.*, 2009; Tanaka *et al.*, 2010; Pang *et al.*, 2012; Sutter *et al.*, 2017). Two forms of PduB are produced; PduB and PduB' the latter resulting from a second translational start site and therefore lacking the 37 N-terminal residues of the full length PduB protein (Havemann and Bobik, 2003; Parsons *et al.*, 2010). The large central pore of PduB/PduB' is thought to permit for the transport of large cofactors into the BMC (Tanaka *et al.*, 2010). The presence of glycerol in the smaller pores, a molecule that is highly similar to 1,2-PD suggests that these pores may facilitate transport of this substrate (Sagermann *et al.*, 2009; Tanaka *et al.*, 2010; Pang *et al.*, 2012). A *pduBB'* deletion mutant inhibits the formation of BMCs and results in the formation of inclusion bodies at the poles of 20% of the observed

cells (Cheng *et al.*, 2011). Additionally, aldehyde toxicity is observed during growth on 1,2-PD in *pduB* mutant strains suggesting that BMCs are not functional (Cheng *et al.*, 2011).

1.3.5.3. PduJ

PduJ shares a high level of sequence similarity to PduA (86%) with pore lining residues also conserved (Crowley *et al.*, 2010). It is therefore assumed, like PduA, to form a hexameric disk, and indeed the crystal structure reveals that the structures of PduA and PduJ are almost identical (Chowdhury *et al.*, 2016). However unlike PduA, deletion of *pduJ* inhibits the formation of BMCs although elongated structures are seen in the cytoplasm (Parsons *et al.*, 2010; Cheng *et al.*, 2011).

1.3.5.4. PduK

PduK is a single BMC domain protein. Unlike PduA and PduJ it contains a ~70 amino acid C-terminal extension (Chowdhury *et al.*, 2014). This C-terminal region contains a Fe-S binding motif (CNLCLDPKCPRQKGEPRTLC). The purified protein is shown to bind iron further supporting the hypothesis of an Fe-S center, although no firm experimental evidence for an Fe-S center has been presented (Crowley *et al.*, 2010). It has been shown that a *pduK* deletion mutant results in protein aggregation showing that it is an essential component of the BMC shell (Cheng *et al.*, 2011).

1.3.5.5. PduM

The role of PduM, a 163 amino acid protein, is uncertain. It does not contain a BMC domain and lacks sequence similarity to any protein of known function (Sinha *et al.*, 2012). The requirement for PduM in shell formation is unclear; deletion mutants resulted in the appearance of abnormally shaped BMCs and shaped cytoplasmic aggregates suggesting its essential role in the formation of wild type BMCs (Sinha *et al.*, 2012). It must however be noted that empty BMCs can be formed in the absence of PduM (Parsons *et al.*, 2010). Therefore further structural and biochemical analyses are required to determine the functionality of this protein.

1.3.5.6. PduN

PduN is similar in amino acid sequence to other pentameric BMC shell proteins such as EutN (Eut BMC), CsoS4 (α-carboxysome) and CcmL (β-carboxysome) (Frank *et al.*, 2013). PduN is predicted to form the vertices of the BMC and therefore is expected to be of low abundance in comparison to other BMC shell proteins (Parsons *et al.*, 2010). Indeed a recent study has shown that it makes up approximately 2% of the metabolosome shell (Mayer *et al*., 2016) Deletion of *pduN* results in impaired growth on 1,2-propanediol and the appearance of abnormal BMC structures (Cheng *et al.*, 2011; Chowdhury *et al.*, 2014).

1.3.5.7. PduU

PduU, a single BMC domain protein with a pore capped by a β-barrel is circularly permuted (Crowley *et al.*, 2008). It contains a unique six-stranded β-barrel that appears to block the central pore the role of which is currently unclear, however it has been proposed to be involved in binding to specific enzymes rather than for metabolite transport (Crowley *et al.*, 2008). Structurally characterized transport proteins containing a β-barrel are composed of twelve or more strands resulting in a much larger diameter pore. Another hypothesis is that the β-barrel may undergo conformational change under certain conditions to open up the channel to transport molecules into and out of the BMC (Crowley *et al.*, 2008).

Deletion of *pduU* does not impair BMC formation and no significant difference in the size of BMCs formed is observed in comparison to wild type BMCs (Cheng *et al.*, 2011).

1.3.5.8. PduT

PduT like PduB contains two BMC domains and forms a pseudohexameric homotrimer (Crowley *et al.*, 2010). The central pore of the PduT trimer contains a [4Fe-4S] centre, with three cysteine residues (C38) providing ligands although it must be noted that the fourth ligand has to date not been identified (Crowley *et al.*, 2010). The [4Fe-4S] centre is accessible from both sides of the hexamer suggesting the protein may play a role in electron transfer across the BMC shell. Alternatively it has been proposed that its function is to rebuild [Fe-S] centres in the lumen of the BMC (Crowley *et al.*, 2010). Deletion of *pduT* does not prevent formation of BMCs

suggesting that its role is non-essential for BMC formation. Similarly the growth rate of mutants lacking PduT is not impaired (Cheng and Bobik, 2010; Cheng *et al.*, 2011).

1.3.5.9. Shell protein composition of the 1,2-propanediol utilising BMC

The relative ratios between different shell proteins are key to understanding their function with regard to the structural characteristics of the BMC shell. A study utilising 2D gel electrophoresis, peptide mass fingerprinting and densitometry has allowed for accurate determination of the relative ratios between shell proteins in *S. enterica* (Havemann and Bobik, 2003). Shell proteins make up 48.1 % of the total protein content of purified BMCs, the remainder being encapsulated and associated proteins.

The majority of the BMC shell is composed of PduB and its shorter isoform PduB', 25.6 % and 25.2 % of the total protein, respectively. PduA and PduJ share a high level of sequence similarity (86 %) and make up 15.6 % and 22.9 % of the shell protein, respectively. PduK, PduT and PduU make up minor components of the shell, 3.3 %, 3.3 % and 3.1 %, respectively, as determined by densitometry (Table 1) (Havemann and Bobik, 2003). Protein spots for PduN were not detected in this study, presumably because of its low abundance (Parsons *et al.*, 2010).

Shell protein	% Total shell protein
PduA	15.6
PduB	26.6
PduB′	25.2
PduJ	22.9
PduK	3.3
PduT	3.3
PduU	3.1

Table 1. Shell protein composition of purified 1,2-propanediol utilizing BMCs. Modified from (Havemann and Bobik, 2003).

A recent study examining the protein composition of 1,2-PD utilizing BMCs from *C. freundii* has revealed a similar protein composition with PduB and PduB' making up 18 % and 31 % of the total shell protein, respectively. In this study PduN was

detected and as expected was found to make up a minor component of the shell (2%) (Mayer *et al.*, 2016).

1.3.6. Principles for BMC assembly

The essential structural components of BMCs have been known for some time for example PduB, PduJ and PduK in the BMC associated with propanediol utilization (Parsons *et al.*, 2010; Cheng *et al.*, 2011). However the nature by which BMC components come together has been relatively poorly understood. Early TEM analysis of *Anabaena variabilis* suggested that the shell of the carboxysome formed first and was subsequently packed with encapsulated protein, although it should be noted the structures observed do not resemble BMC facets (Price and Badger, 1991). However, this leaves the obvious problem: how does cargo protein get encapsulated into a closed structure?

A recent study utilizing TEM and confocal analyses has shown that the formation of β-carboxysomes in the organism *Synechococcus elongatus* do in fact form from the inside out, i.e. the core of enzymes forms first and this protein aggregate is then encapsulated within the shell. The ordered assembly of carboxysomes proceeds as follows:

RuBisCO first aggregates, in a process dependent on full (M58) and short (M35) length CcmM, forming a procarboxysome aggregate. CcmN is then recruited to this procarboxysome through interactions with CcmM. The C-terminal targeting peptide on CcmN then recruits the shell components CcmK2 and CcmO, with the encapsulation of this aggregate by incorporation of the pentameric vertex protein the functional carboxysome then buds off the procarboxysome body (Cameron *et al.*, 2013).

Conversely, electron tomograms of *Halothiobacillus neapolitanus* reveals the presence of partially formed α-carboxysomes suggesting the co-assembly of the shell with encapsulated enzymes (Iancu *et al.*, 2010). Deletion of RuBisCO results in the presence of regularly shaped carboxysomes suggesting that it is not key to the formation of organelles (Menon *et al.*, 2008).

To date, the formation of metabolosomes has not been revealed however it should be noted that the formation of these BMCs is not dependent on the presence of a

protein to be encapsulated as evidence has shown that it is possible to recombinantly produce empty BMCs (Parsons *et al.*, 2010; Sutter *et al.*, 2017).

1.3.7. Targeting to BMCs

The functionality of BMCs is only provided when specific enzymes are encapsulated within the shell. Sequence alignments of the PduCDE diol dehydratase revealed that two of the subunits contain N-terminal extensions that are not present in homologues that are not associated with BMCs, with further analysis revealing that these extensions are not required for enzymatic activity (Daniel *et al.*, 1998; Tobimatsu *et al.*, 2005). Similarly sequence alignments showed that the N-terminal region of PduP is not present in homologues and is similarly not required for enzyme function (Fan *et al.*, 2010). Fan *et al* showed it was possible to localize GFP and other proteins to BMCs by transplanting the N-terminal 18 amino acids of PduD and PduP onto the N-terminus of the reporter protein, thereby opening up the possibility of re-functionalizing these nanoreactors (Fan *et al.*, 2010; Fan and Bobik, 2011). The solution structure of the PduP targeting peptide has recently been solved by NMR spectroscopy, revealing as previously predicted an amphipathic alpha helix (Lawrence *et al.*, 2014).

It has been shown that it is possible to utilize targeting sequences from different BMC systems to target proteins to the Pdu BMC, for example fusion of the first 20 amino acids of EutC to GFP showed localization to the Pdu BMC (Jakobson *et al.*, 2015). This promiscuity for targeting sequences suggests a relatively simple targeting mechanism. Sequence analysis reveals a common hydrophobic motif facilitates protein encapsulation into BMCs (Jakobson *et al.*, 2015). This revelation has allowed for the design of novel targeting sequences based upon this hydrophobic motif that allow for targeting to BMCs (Jakobson *et al.*, 2017a). However these ‘novel’ targeting peptides likely bind to the same site on the BMC shell as the native peptides, therefore limiting their functionality. The development of a truly novel system for the targeting of proteins to BMCs would allow for much tighter control over the ratios of encapsulated components.

It must also be noted that the majority of encapsulated proteins do not contain an obvious amphipathic targeting sequence and therefore other protein-protein interactions must be key to the encapsulation of such components and the formation of functional BMCs.

Previous experimental evidence has shown that the PduP targeting peptide interacts with the shell proteins PduA, PduJ and PduK (Fan *et al.*, 2012; Lawrence *et al.*, 2014). Alanine mutagenesis of 5 conserved C-terminal residues of PduA showed that H81A, V84A and L88A resulted in reduced targeting of PduP to BMCs suggesting that these residues likely interact with the PduP targeting peptide, importantly it should be noted that these residues are located on the concave surface of PduA (Havemann *et al.*, 2002; Fan *et al.*, 2012). Targeting was likely not completely eliminated presumably because of the sequence similarity between PduA and PduJ, including conservation of the three critical C-terminal residues (Fan *et al.*, 2012).

Intriguingly the crystal structure of the *H. ochraceum* BMC reveals that the concave face of the hexameric protein, containing the C-terminus faces out (Sutter *et al.*, 2017). If this organisation is conserved among BMCs, which is likely due to the high level of conservation this suggests either that PduP binds solely to the exterior surface of the BMC shell or that there are multiple binding sites for this protein.

1.3.8. Recombinant production of BMCs

The first BMC to be recombinantly produced was the metabolosome associated with 1,2-propanediol utilization. Through the transfer and expression of the 21 gene *pdu* operon from *C. freundii* into *E. coli* functional BMCs were produced and visualised (Parsons *et al.*, 2008). Further work has shown that it is possible to recombinantly produce empty Pdu BMCs through the expression of a minimal set of 6 shell proteins, PduA, B, B', J, K and N, although slightly larger structures are formed with the addition of PduU, opening up the possibility of re-functionalising these subcellular compartments (Figure 7) (Parsons *et al.*, 2010).

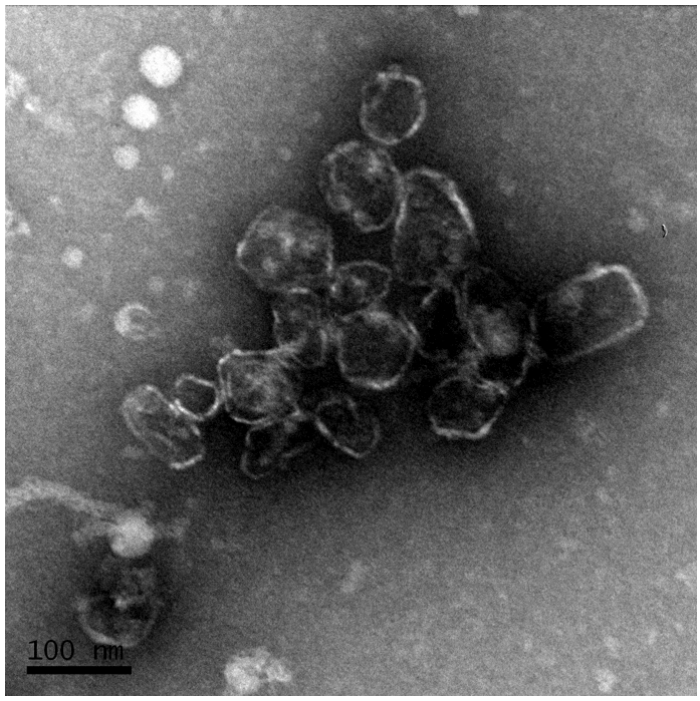


Figure 7. TEM analysis of purified empty BMCs produced through the expression of PduABB'JKNU in BL21 * (DE3) cells and subsequent purification. Scale bar shows 100 nm (Lee, M.J. unpublished).

Several attempts have been made to recombinantly produce Eut BMCs, however to date, all such attempts have resulted *in-vivo* aggregates, it has also been claimed that a single hexameric shell protein is ample for the formation of BMCs, however these structures appear in micrographs to resemble membranous vesicles (McBroom *et al.*, 2006; Choudhary *et al.*, 2012; Held *et al.*, 2016; Quin *et al.*, 2016).

More recently functional carboxysomes have been successfully produced in *E. coli* through the expression of the 10-gene operon encoding α-carboxysomes from *H. neapolitanus*. Upon induction, icosahedral structures were visualised by TEM that appear structurally similar to those of the host, the functionality of these CO_2 fixing organelles was also shown (Bonacci *et al.*, 2012). However under high protein induction, inclusion bodies were seen as well as filamentous structures, suggesting an imbalance in protein stoichiometry (Bonacci *et al.*, 2012).

A BMC of unknown function has recently been recombinantly expressed in *E. coli*. A synthetic operon containing the seven-shell protein encoding genes from 3 chromosomally separated loci in *H. ochraceum* was constructed and recombinantly expressed in *E. coli* (Lassila *et al.*, 2014). BMCs could be readily purified and formed uniform 40 nm structures, however these could not be visualized *in-vivo*

suggesting that they form during the purification process (Lassila *et al.*, 2014; Sutter *et al.*, 2017). Intriguingly, the pentameric vertex protein was not required for the formation of BMCs (Lassila *et al.*, 2014). The functionality of this BMC remains unclear however the presence of an aldehyde dehydrogenase suggests that this BMC facilitates the sequestration of a toxic aldehyde intermediate like many other BMCs (Yeates *et al.*, 2010; Axen *et al.*, 2014).

1.3.8.1. Protein nanotubes

It has been shown that recombinant expression of the *C. freundii* hexameric protein PduA in *E. coli* results in the formation of proteinaceous filaments (Parsons *et al.*, 2010). Further analysis of these structures reveals that they have an external diameter of approximately 20 nm (Figure 8) (Pang *et al.*, 2014). Mutation of residues predicted to be involved in nanotube assembly, lysine 26 and arginine 79 to alanine resulted in the formation of sheet like structures in the cytoplasm of transformed cells (Pang *et al.*, 2014). A homologous protein to PduA, RmmH from *Mycobacterium smegmatis* similarly forms nanotubes when over expressed in *E. coli*, however these nanotubes have a smaller external diameter (14 nm), however, this diameter increased upon purification to 18 nm (Noël *et al.*, 2016). The increase in diameter may be due to the collapsing of the nanotubes as they are dried onto TEM grids.

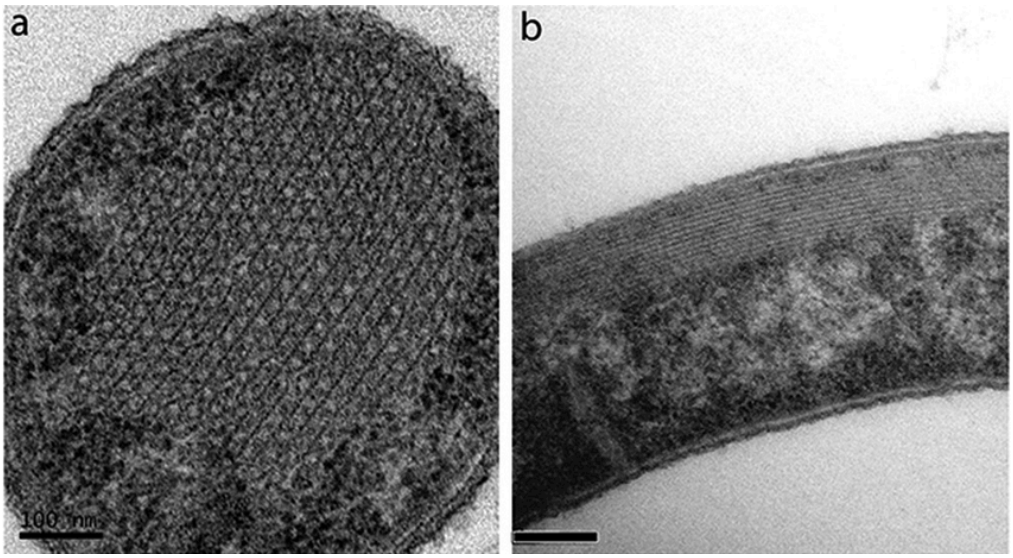


Figure 8. TEM analysis of *E. coli* overexpressing PduA* from *C. freundii* in transverse (a) and longitudinal (b) thin sections. Scale bar in b shows 0.5 μm (Pang *et al.*, 2014).

The functionalisation of these protein nanotubes and sheet like structures may allow for increased localised intermediate concentrations and therefore greater flux through an engineered metabolic pathway without the great cost on the cell of producing a BMC.

1.4. Spatial organisation in synthetic biology

Synthetic biologists have long looked to nature to develop strategies to enhance synthetic metabolic pathways. This has brought BMCs into the fore as they present the potential to both increase local concentrations of enzymes and intermediates and also prevent the release of toxic intermediates from engineered metabolic pathways.

To date the utilisation of BMCs for industrial applications has been limited primarily by the difficulties in recombinantly producing empty BMCs *in-vivo* (Choudhary *et al.*, 2012; Lassila *et al.*, 2014; Sutter *et al.*, 2017). Despite this, significant progress has been made towards the industrial use of BMCs.

1.4.1. BMC technology for fine and commodity chemical production

The first, and at present, only example of the repurposing of a BMC for a biotechnological application was performed in 2014. An ethanol bioreactor was formed through the co-expression of the genes required to form an empty BMC with pyruvate decarboxylase and alcohol dehydrogenase tagged with targeting peptides (Lawrence *et al.*, 2014). These recombinant bioreactors produced more ethanol in comparison to strains producing untagged enzymes with BMCs, however, as subsequently revealed the fusion of targeting peptides onto recombinant proteins results in protein aggregation and increased pathway flux (Lee *et al.*, 2016). It is therefore unclear if the increase in ethanol production was as a result of the encapsulation of the pathway or the aggregation of the pathway.

The utilisation of cellular aggregates for the enhancement of recombinant metabolic pathways also shows promise; by simply tagging the 4 enzymes for the synthesis of 1,2-PD from glycerol in the absence of BMCs a 245% increase in product formation was observed in comparison to a strain expressing untagged soluble enzymes (Lee

et al., 2016). This increase in production is likely due to substrate channelling within the polar inclusion bodies.

The development of novel targeting mechanisms that avoid the problem of amphipathic targeting peptides resulting in protein aggregation would allow for the efficient utilisation of this technology. To date the development of novel targeting mechanisms has been limited to re-engineering existing targeting peptides that share the same hydrophobic motif, and will therefore likely bind to the same region on the BMC shell although the interaction domain for targeting peptides on the shell has not yet been experimentally shown (Jakobson *et al.*, 2015; Jakobson *et al.*, 2017a). Recently, researchers have utilised the non-native protein interaction domains to target fluorescent proteins to recombinant BMCs in *Corynebacterium glutamicum,* however, such fusions also appear to result in protein aggregation within the cell (Huber *et al.*, 2017).

1.4.2. Bioremediation

The removal of polyphosphate from wastewater is a key step in wastewater treatment (Grady Jr *et al.*, 2011). Currently wastewater treatment plants utilize a complex bacterial community to remove polyphosphate from wastewater (Mino, 2000; Gebremariam *et al.*, 2011). Due to their ability to sequester molecules away from the cytoplasmic machinery, BMCs were proposed as a potential technology to facilitate the removal and storage of polyphosphate.

Recently researchers have attempted to engineer a compartment to remove polyphosphate. This was achieved by encapsulating polyphosphate kinase with a targeting peptide resulting in its encapsulation and the build up of polyphosphate within the BMC. Expression of an exopolyphosphatase did not result in the degradation of polyphosphate suggesting that it is sequestered within BMCs (Liang *et al.*, 2017).

1.4.3. Carbon/ nitrogen fixation in C3 plants

As discussed previously cyanobacteria utilize a carbon concentrating mechanism (CCM) to enhance carbon fixation (Kerfeld and Melnicki, 2016). The transplantation of this functionality to plants has the potential to increase global carbon fixation as well as to increase productivity (Zarzycki *et al.*, 2013). The inefficient enzyme

RuBisCO is known to account for approximately 50% of the soluble protein in C3 plants, therefore reducing the amount of RuBisCO required by increasing the efficiency has the potential to enhance yields (Price *et al.*, 2008).

Towards this goal, researchers have been able to form β-carboxysome like structures in *Nicotiana benthamiana* chloroplasts by expression of 3 proteins (CcmO-YFP, CcmK2 and CcmL-YFP or CcmM58-YFP), however similar structures are not seen in the absence of YFP fusions and as such further work is required to probe these *in-vivo* structures (Lin *et al.*, 2014).

1.5. Aims and objectives

Herein, we aim to utilise BMC technology to enhance the subcellular organisation of proteins through a number of strategies.

First, through the utilisation of the BMC targeting peptides derived from PduD and PduP we show that recombinant proteins fused to these peptides aggregate *in-vivo*, and that these protein aggregates enhance production of the commodity chemical 1,2-PD.

Furthermore, we analysed the sequence of a diverse set of targeting peptides and found a common hydrophobic motif that is highly conserved, this motif and its application for the production of commodity chemicals form the basis of a patent.

Third we utilise a three-component system to functionalize PduA nanotubes to form an *in-vivo* filamentous network that proteins can be targeted to.

Finally we aimed to develop a BMC system that would allow for the targeting to both the luminal face as well as the exterior of surface of the BMC through the application of coiled-coil technology.

Chapter 2

Publications

Publications listed in this chapter are contained within the Appendix; contributions to each paper are listed in Chapter 5

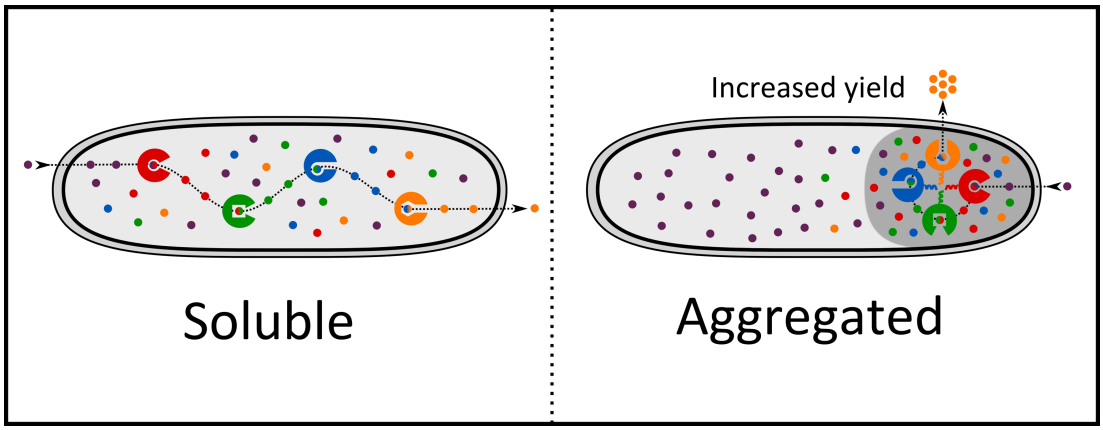
2.1. Employing bacterial microcompartment technology to engineer a shell-free enzyme-aggregate for enhanced 1,2-propanediol production in *Escherichia coli*.

Lee, M.J., Brown, I.R., Juodeikis, R., Frank, S., and Warren, M.J. (2016) Employing bacterial microcompartment technology to engineer a shell-free enzyme-aggregate for enhanced 1,2-propanediol production in *Escherichia coli*. Metab Eng 36: 48-56.

2.1.1. Abstract

Bacterial microcompartments (BMCs) enhance the breakdown of metabolites such as 1,2-propanediol (1,2-PD) to propionic acid. The encapsulation of proteins within the BMC is mediated by the presence of targeting sequences. In an attempt to redesign the Pdu BMC into a 1,2-PD synthesising factory using glycerol as the starting material we added N-terminal targeting peptides to glycerol dehydrogenase, dihydroxyacetone kinase, methylglyoxal synthase and 1,2-propanediol oxidoreductase to allow their inclusion into an empty BMC. 1,2-PD producing strains containing the fused enzymes exhibit a 245% increase in product formation in comparison to un-tagged enzymes, irrespective of the presence of BMCs. Tagging of enzymes with targeting peptides results in the formation of dense protein aggregates within the cell that are shown by immuno-labelling to contain the vast majority of tagged proteins. It can therefore be concluded that these protein inclusions are metabolically active and facilitate the significant increase in product formation.

Abstract Figure:



2.1.2. Context of research

Substrate channelling can occur by either the direct transfer of a substrate between active sites of enzymes or through an increased local concentration of enzymes and substrates (Spivey and Merz, 1989). Bacterial microcompartments are thought to increase local concentrations of enzymes and intermediates. Additionally the presence of a semi permeable protein shell may act to protect the cell from toxic pathway intermediates and prevent the loss of intermediates to side reactions (Yeates *et al.*, 2010).

Synthetic biologists are evermore looking to BMCs as a chassis to enhance production of both fine and commodity chemicals. Industrial applications of BMCs have become apparent through the recombinant expression of BMCs and the generation of empty BMCs by expression of a minimal set of shell proteins (Parsons *et al.*, 2008; Parsons *et al.*, 2010). The encapsulation of proteins within BMCs is mediated by short targeting peptides; such targeting peptides have been previously utilized to target fluorescent proteins to BMCs (Fan *et al.*, 2010; Fan and Bobik, 2011).

Previously it has been shown that co-expression of pyruvate decarboxylase and alcohol dehydrogenase tagged with targeting peptides with the genes required to form empty BMCs resulted in the production of an ethanol bioreactor (Lawrence *et al.*, 2014). This initial study showed the industrial potential of BMC technology.

Here, we aimed to construct a bioreactor for the production of 1,2-PD utilizing the feedstock glycerol, a compound that is produced as a by-product of biodiesel production (Marchetti *et al.*, 2007). 1,2-propanediol, a commodity chemical currently used in the production of antifreeze, thermoset plastics and cosmetics, is currently produced from non-renewable resources. Annual demand for this compound was estimated in 2011 to be 1.36 million tonnes and it is therefore of great interest to seek a renewable production platform (Shelley, 2007; Clomburg and Gonzalez, 2011).

2.1.3. Aims of research

This research aimed to target a longer metabolic pathway to BMCs than has previously been achieved. We chose a 5-step pathway for the production of 1,2-PD from glycerol. The selection of this pathway was based on knowledge that 1,2-PD should be able to get into the BMC as it is the substrate for the wild type organelle and we presume it can therefore also diffuse out of the structure.

The aim of this work was to create an *in-vivo* propanediol bioreactor, which we sought to achieve by tagging four enzymes: glycerol dehydrogenase, dihydroxyacetone kinase, methylglyoxal synthase and 1,2-propanediol oxidoreductase. These enzymes form a pathway that is able to convert glycerol to 1,2-propanediol. The pathway involves the toxic intermediates methylglyoxal and lactaldehyde and therefore we propose that encapsulation of this pathway within a BMC will limit the toxic effects of these molecules.

2.2. Patent application: Genetically modified microorganisms.

Lee, M.J., Frank, S. Warren, M.J. 2015. Genetically Modified Microorganisms. Filed November 2015. International application number: PCT/GB2016/053435.

2.2.1. Abstract

The present invention relates to genetically modified microorganisms and their use in the production of desired products of metabolic pathways, and particularly improving the levels of production of said products. In particular the microorganisms of the present invention are modified to comprise enzymes of a metabolic pathway for a product of interest, wherein the enzymes are each tagged with a bacterial microcompartment (BMC)-targeting signal peptide, and wherein the microorganism lacks bacterial microcompartments. The present invention also relates to cell-free systems comprising said (BMC)-targeting signal peptide tagged enzymes in the absence of BMCs.

2.2.2. Context of research

This patent application concerns the use of bacterial microcompartment targeting peptides for the production of both commodity and fine chemicals. The requirement for fossil fuel derived products is ever increasing; therefore many industrial sectors are investigating methods to overcome this reliance of fossil fuels through the development of more efficient pathways.

2.2.3. Aims of research

Here we aimed to protect our finding that BMC targeting peptides fused to enzymes of a synthetic metabolic pathway enhance productivity of pathways by aggregation of enzymes into an active inclusion body.

The experimental data presented in this patent application is the same as in the publication 'Employing bacterial microcompartment technology to engineer a shell-free enzyme-aggregate for enhanced 1,2-propanediol production in *Escherichia coli*.'

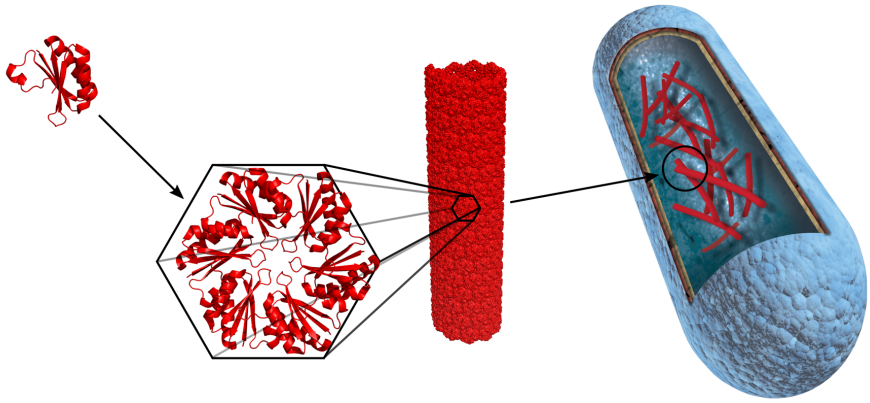
2.3. Engineered synthetic scaffolds for organising proteins within bacterial cytoplasms

Lee, M.J., Mantell, J., Hodgson, L., Alibhai, D., Fletcher, J.M., Brown, I.R., Frank, S., Xue, W-F., Verkade, P., Woolfson, D.N., Warren, M.J. (2018) Engineered synthetic scaffolds for organizing proteins within the bacterial cytoplasm. Nat Chem Biol 14(2): 142-147

2.3.1. Abstract

We have developed a system for producing a supramolecular scaffold that permeates the entire *Escherichia coli* cytoplasm. This cytoscaffold is constructed from a three-component system comprising a bacterial microcompartment shell protein and two complementary *de novo* coiled-coil peptides. We show that other proteins can be targeted to this intracellular filamentous arrangement. Specifically, the enzymes pyruvate decarboxylase and alcohol dehydrogenase have been directed to the filaments, leading to enhanced ethanol production in these engineered bacterial cells compared to those that do not produce the scaffold. This is consistent with improved metabolic efficiency through enzyme colocation. Finally, the shell-protein scaffold can be directed to the inner membrane of the cell, demonstrating how synthetic cellular organization can be coupled with spatial optimization through in-cell protein design. The cytoscaffold has potential in the development of next-generation cell factories, wherein it could be used to organize enzyme pathways and metabolite transporters to enhance metabolic flux.

Abstract Figure:



2.3.2. Context of research

The localisation of proteins on scaffolds to enhance metabolic pathways has gained significant attention in recent years. Researchers have investigated bacterial microcompartments, protein-based scaffolds, lipids and DNA scaffolds as facilitators of enhanced enzyme localisation (Dueber *et al.*, 2009; Grinkova *et al.*, 2010; Delebecque *et al.*, 2012; Lawrence *et al.*, 2014). Despite advances in the field, and with the exception of bacterial microcompartments, to date it has not been possible to produce a visible protein scaffold within the bacterial cytoplasm that allows for the tethering of enzymes. The overexpression of the BMC shell protein PduA has been shown to form proteinaceous filaments *in-vivo* (Parsons *et al.*, 2010). These filaments, approximately 20 nm in diameter and several microns in length have been previously characterised and modified to form a variety of *in-vivo* structures (Pang *et al.*, 2014). Recent work has shown that it is possible to target fluorescent proteins to PduA* nanotubes by fusion to BMC targeting peptides in *Corynebacterium glutamicum* (Huber *et al.*, 2017). The authors also show that it is possible to target to these structures via C-terminally fused protein interaction domains. However the structures visualised by fluorescence microscopy appear to be bundles of filaments as formed by untagged PduA. This bundling limits the accessible surface area and thus limits the biotechnological applications of such structures.

Here we aim to enhance subcellular organisation by targeting recombinant proteins to engineered cytoplasmic scaffolds or cytoscaffolds. By targeting enzymes of a metabolic pathway to these protein filaments we propose that local concentrations of both enzymes and substrates would be increased thereby leading to increased pathway productivity.

2.3.3. Aims of research

The aim of the work presented in this publication was to engineer a bacterial scaffold based on the BMC shell protein PduA. We set out to utilize a pair of complementary coiled-coil peptides to target recombinant proteins to these protein filaments by attaching one peptide to the BMC shell protein PduA and the cognate peptide to a protein of interest.

We aimed to functionalise these nanotubes with pyruvate decarboxylase and alcohol dehydrogenase, enzymes that facilitate the conversion of pyruvate into

ethanol to investigate if localisation of enzymes on a nanotube provides an enhancement to pathway productivity.

Furthermore utilising this coiled-coil technology and a C-terminal membrane targeting peptide we aim to target this scaffold to the inner membrane of the cell. This specific localisation may improve pathway flux due to the increased proximity to the extracellular environment and thus the substrate pool.

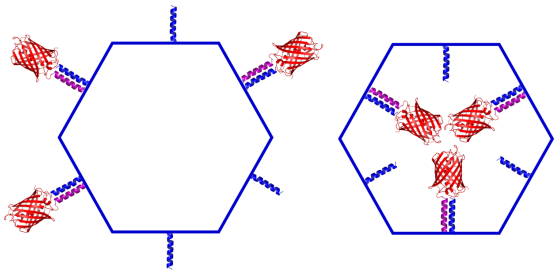
2.4. *De novo* targeting to the cytoplasmic and luminal side of bacterial microcompartments

Lee, M.J., Brown, I.R., Fletcher, J.M., Pickersgill, R.W., Woolfson, D.N., Frank, S., Warren, M.J. *De novo* targeting to the cytoplasmic and luminal side of bacterial microcompartments. Nat. Commun (Under review)

2.4.1. Abstract

Bacterial microcompartments, BMCs, are proteinaceous organelles that encase a specific metabolic pathway within a semi-permeable protein shell. Short encapsulation peptides direct cargo proteins to the lumen of the compartments. However, the fusion of such peptides to non-native proteins does not guarantee encapsulation and often causes aggregation. Here, we report a different approach for targeting recombinant proteins to BMCs that utilizes specific *de novo* coiled-coil protein-protein interactions. Attachment of one coiled-coil module to a component of the bacterial microcompartment shell, PduA, allows targeting of a fluorescent protein marker fused to a cognate coiled-coil partner. This interaction takes place on the outer surface of the BMC. The redesign of PduA to generate a new N-terminus on the luminal side of the BMC results in intact compartments to which fluorescent proteins can still be targeted via the designed coiled-coil system. This study demonstrate that de novo coiled-coil peptide technology can be used to target to the luminal or cytoplasmic faces of BMCs, allowing proteins to be displayed on the surface or internalized within the shell of the BMC.

Abstract Figure:



2.4.2. Context of research

A major obstacle to the employment of BMCs in industrial applications is the problem of aggregation as a result of fusions between enzymes and BMC targeting peptides (Lee *et al.*, 2016). Despite the fact that we have shown that aggregation of an engineered metabolic pathway is advantageous there may be applications that require a further sequestered environment and thus an effective alternative targeting system would be required. For example encapsulation of recombinantly produced toxic proteins may enhance their production.

Previous work has shown that fusion of the protein interaction domains to the C-terminus of PduA and co-expression of a fluorescent protein tagged with the interaction partner results in the fluorescent protein being localised to the poles of the cell, suggesting that the fluorescent protein is aggregating (Huber *et al.*, 2017). Here we aim to develop an altogether different strategy utilising a pair of *de novo* designed coiled-coil peptides to target proteins to recombinantly produced BMCs.

2.4.3. Aims of research

The aim of this work was to overcome problematic protein aggregation by designing a novel mechanism for targeting proteins to BMCs. As with previous work we aimed to utilise a pair of complementary coiled-coil peptides to facilitate targeting. One half of the heterodimeric pair would be fused to a major component of the Pdu BMC shell, PduA, and the cognate peptide would be fused to a protein of interest. The coiled-coil interaction would then facilitate targeting of the protein of interest to the BMC.

Furthermore we aimed to re-engineer a BMC shell protein to enable targeting to both the internal and external face of the structure. We aimed to achieve this through circular permutation of PduA.

Chapter 3

Discussion

3.1. Employing bacterial microcompartment technology to engineer a shell-free enzyme-aggregate for enhanced 1,2-propanediol production in *Escherichia coli*.

The work presented in this paper shows a broad range of results with a number of further applications. First, we show that encapsulation of GFP with a C-terminal proteolysis tag within a BMC protects it from proteolysis by ClpAP and ClpXP whereas the same GFP in the absence of BMCs is effectively degraded. This result reinforces previous evidence that shows the shell of the BMC acts as a semi-permeable membrane and provides an effective barrier between encapsulated and cytoplasmic proteins.

This research shows how the fusion of the targeting peptides D18 and P18 to the enzymes required for the conversion of glycerol to 1,2-PD results in protein aggregation. The amphipathic nature of BMC targeting peptides is thought to be the cause of this aggregation. Furthermore expression of all of the enzymes of the engineered pathway resulted in the appearance of large polar enzyme aggregates. Unexpectedly the data shows that an increase in product formation was not dependent on the presence of BMCs and the tagging of enzymes with targeting peptides was sufficient to significantly enhance 1,2-PD production, suggesting that the localisation of proteins to the inclusion facilitates enhanced product formation.

The work presented in this publication shows how bacterial microcompartment targeting peptides can be utilized to form intracellular enzyme aggregates that contain the enzymes required to convert glycerol to 1,2-PD. Due to the simple nature of these peptide fusions this technology has the potential to be applicable to a wide range of engineered metabolic pathways.

3.2. Patent application: Genetically modified microorganisms.

Sequence alignments of enzymes known to be associated with BMCs have revealed a highly conserved hydrophobic motif that is absent from non-encapsulated enzymes and it is this hydrophobic motif that facilitates protein encapsulation into BMCs. We show experimentally that fusions between proteins and such peptides result in aggregation of such proteins and that these enzyme aggregates confer enhanced pathway productivity. This patent seeks to protect the use of such peptides containing this conserved hydrophobic motif in pathways consisting of 3 or more recombinant proteins forming an enzymatic pathway.

3.3. Engineered synthetic scaffolds for organising proteins within bacterial cytoplasms

Here we demonstrate how an intracellular scaffold can enhance spatial organisation in *E. coli*. This scaffold is constructed from a BMC shell protein, PduA, and a pair of complementary coiled-coil peptides. We show that fusion of CC-Di-B to PduA* results in the formation of a dispersed network of filaments that permeates the entire cytoplasm. This network of filaments that we observe with CC-Di-B-PduA* provides a high surface area exposed to cytoplasmic substrates therefore providing enhanced potential over the bundled PduA* structures. We show that co-expression of CC-Di-A tagged fluorescent proteins with the CC-Di-B-PduA* scaffold result in their localisation to the filaments. We show functionality of this scaffold by appending the enzymes for ethanol production to it resulting in a significant enhancement in ethanol production. We also show that by lining the inner membrane with the CC-Di-A peptide the CC-Di-B-PduA* nanotubes can be localised to the membrane potentially allowing for faster exchange of metabolites and products into and out of the cell. This research represents a new era of in-cell protein design.

3.4. *De novo* targeting to the cytoplasmic and luminal side of bacterial microcompartments

Our results show that the fusion of the coiled-coil peptide CC-Di-B to PduA does not impair BMC formation and the resulting BMCs appear to be of similar dimensions to unmodified empty BMCs. Furthermore we show that by expressing fluorescent proteins tagged with a complementary peptide, CC-Di-A, such proteins can be recruited to BMCs. Structural evidence suggests that the convex side of BMC shell proteins faces the lumen and therefore the concave side containing the N and C-termini of PduA face out suggesting that the fusions that we have constructed label the outer surface of the structure (Sutter *et al.*, 2017). The accessibility of coiled coil peptides on the outer surface of BMCs could allow for a metabolic pathway to be split between the outer surface and luminal space. Or alternatively, more significantly, it could be used to direct BMCs to specific areas of the cell for example by targeting bioreactors to the membrane of the cell thereby increasing channelling of substrates from the extracellular environment into the bioreactor. We also show that by circularly permuting the BMC shell protein PduA it is possible to target fluorescent proteins to the luminal face of the BMC using the same coiled-coil peptides.

3.5. Concluding remarks

The aim of this research was to enhance spatial organisation in *E. coli*. Through deconstruction of the BMC associated with 1,2-PD utilisation, and the utilisation of the component parts, novel architectures have been produced that facilitate enhanced metabolic channelling. First we show how fusions between enzymes and BMC targeting peptides results in the formation of active enzyme aggregates that enhance the production of the commodity chemical 1,2-PD. Second we show how a three component system consisting of a bacterial microcompartment shell protein and a pair of complementary coiled-coil peptides can be utilised to form a cytoscaffold that can be functionalised. Finally, utilising the same coiled-coil technology we show targeting to recombinantly produced BMCs, furthermore we re-engineer a BMC shell protein to enable targeting to both the inner and outer surface of BMCs. In conclusion, the data presented in this thesis offers a number of strategies going forward to enhance protein organisation *in-vivo*, allowing for higher order cellular design in prokaryotes.

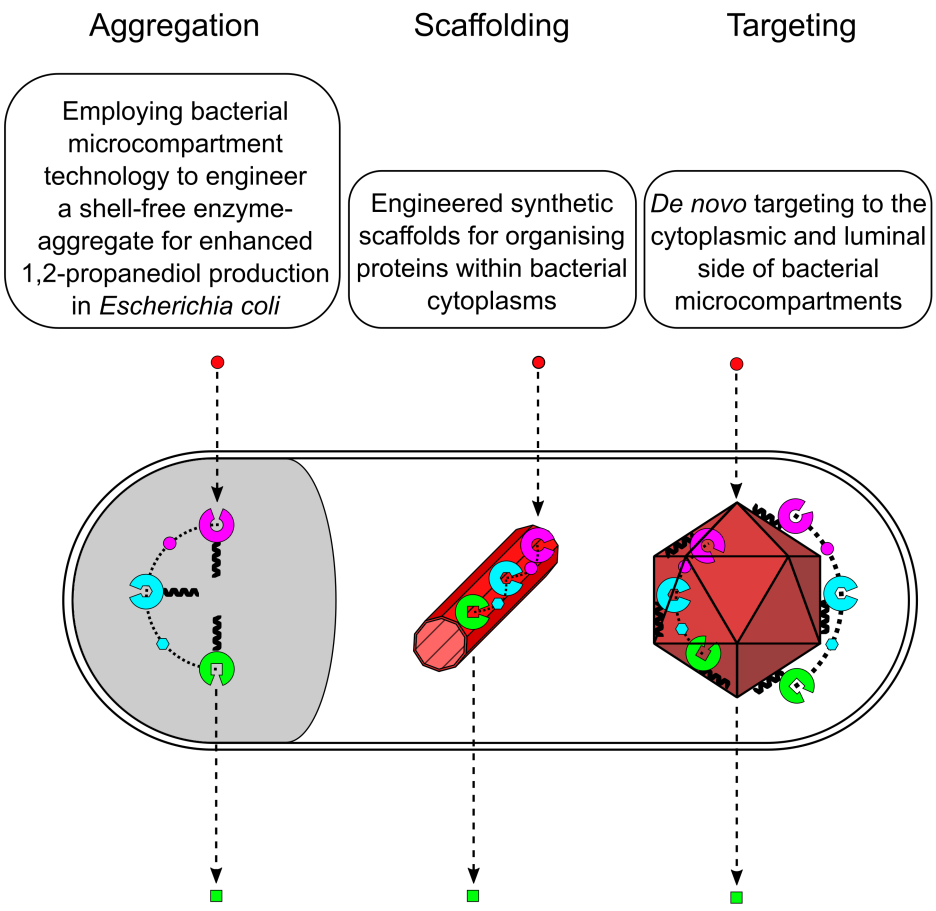


Figure 9. Summary of the BMC based architectures presented in this thesis.

Chapter 4

References

Axen, S.D., Erbilgin, O., and Kerfeld, C.A. (2014) A taxonomy of bacterial microcompartment loci constructed by a novel scoring method. *PLoS Comput Biol* **10:** e1003898.

Badger, M.R., and Price, G.D. (2003) CO2 concentrating mechanisms in cyanobacteria: molecular components, their diversity and evolution. *J Exp Bot* **54:** 609-622.

Badia, J., Ros, J., and Aguilar, J. (1985) Fermentation mechanism of fucose and rhamnose in *Salmonella typhimurium* and *Klebsiella pneumoniae. J Bacteriol* **161:** 435-437.

Bergman, N.M., Lenton, T.M., and Watson, A.J. (2004) COPSE: A new model of biogeochemical cycling over Phanerozoic time. *American Journal of Science* **304:** 397-437.

Berner, R.A. (2006) GEOCARBSULF: A combined model for Phanerozoic atmospheric O_2 and CO_2. Geochim. Cosmochim. Acta **70:** 5653-5664.

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. *J Bacteriol* **181:** 5967-5975.

Bobik, T.A., Xu, Y., Jeter, R.M., Otto, K.E., and Roth, J.R. (1997) Propanediol utilization genes (*pdu*) of *Salmonella typhimurium*: three genes for the propanediol dehydratase. *J Bacteriol* **179:** 6633-6639.

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. *Proc Natl Acad Sci U S A* **109:** 478-483.

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.

Cai, F., Menon, B.B., Cannon, G.C., Curry, K.J., Shively, J.M., and Heinhorst, S. (2009) The pentameric vertex proteins are necessary for the icosahedral carboxysome shell to function as a CO_2 leakage barrier. *PLoS One* **4:** e7521.

Cameron, J.C., Wilson, S.C., Bernstein, S.L., and Kerfeld, C.A. (2013) Biogenesis of a bacterial organelle: the carboxysome assembly pathway. *Cell* **155:** 1131-1140.

Cannon, G.C., and Shively, J.M. (1983) Characterization of a homogenous preparation of carboxysomes from *Thiobacillus neapolitanus. Arch Microbiol* **134:** 52-59.

Cannon, G.C., Bradburne, C.E., Aldrich, H.C., Baker, S.H., Heinhorst, S., and Shively, J.M. (2001) Microcompartments in Prokaryotes: Carboxysomes and Related Polyhedra. *Applied and Environmental Microbiology* **67:** 5351-5361.

Chen, P., Andersson, D.I., and Roth, J.R. (1994) The control region of the *pdu/cob* regulon in Salmonella typhimurium. *J Bacteriol* **176:** 5474-5482.

Cheng, S., and Bobik, T.A. (2010) Characterization of the PduS cobalamin reductase of *Salmonella enterica* and its role in the Pdu microcompartment. *J Bacteriol* **192:** 5071-5080.

Cheng, S., Fan, C., Sinha, S., and Bobik, T.A. (2012) The PduQ enzyme is an alcohol dehydrogenase used to recycle NAD+ internally within the Pdu microcompartment of *Salmonella enterica. PLoS One* **7:** e47144.

Cheng, S., Liu, Y., Crowley, C.S., Yeates, T.O., and Bobik, T.A. (2008) Bacterial microcompartments: their properties and paradoxes. *Bioessays* **30:** 1084-1095.

Cheng, S., Sinha, S., Fan, C., Liu, Y., and Bobik, T.A. (2011) Genetic analysis of the protein shell of the microcompartments involved in coenzyme B12-dependent 1,2-propanediol degradation by *Salmonella. J Bacteriol* **193:** 1385-1392.

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.

Chowdhury, C., Chun, S., Sawaya, M.R., Yeates, T.O., and Bobik, T.A. (2016) The function of the PduJ microcompartment shell protein is determined by the genomic position of its encoding gene. *Mol Microbiol* **101:** 770-783.

Chowdhury, C., Sinha, S., Chun, S., Yeates, T.O., and Bobik, T.A. (2014) Diverse bacterial microcompartment organelles. *Microbiol Mol Biol Rev* **78:** 438-468.

Clomburg, J.M., and Gonzalez, R. (2011) Metabolic engineering of Escherichia coli for the production of 1,2-propanediol from glycerol. *Biotechnol Bioeng* **108:** 867-879.

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.

Crowley, C.S., Sawaya, M.R., Bobik, T.A., and Yeates, T.O. (2008) Structure of the PduU shell protein from the Pdu microcompartment of *Salmonella. Structure* **16:** 1324-1332.

Daniel, R., Bobik, T.A., and Gottschalk, G. (1998) Biochemistry of coenzyme B12-dependent glycerol and diol dehydratases and organization of the encoding genes. *FEMS Microbiol Rev* **22:** 553-566.

Delebecque, C.J., Silver, P.A., and Lindner, A.B. (2012) Designing and using RNA scaffolds to assemble proteins in vivo. *Nat Protoc* **7:** 1797-1807.

Dou, Z., Heinhorst, S., Williams, E.B., Murin, C.D., Shively, J.M., and Cannon, G.C. (2008) CO_2 fixation kinetics of *Halothiobacillus neapolitanus* mutant carboxysomes lacking carbonic anhydrase suggest the shell acts as a diffusional barrier for CO_2. *J Biol Chem* **283:** 10377-10384.

Drews, G., and Niklowitz, W. (1956) Cytology of Cyanophycea. II. Centroplasm and granular inclusions of *Phormidium uncinatum. Arch Mikrobiol* **24:** 147-162.

Dueber, J.E., Wu, G.C., Malmirchegini, G.R., Moon, T.S., Petzold, C.J., Ullal, A.V., *et al.* (2009) Synthetic protein scaffolds provide modular control over metabolic flux. *Nat Biotechnol* **27:** 753-759.

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

Fan, C., Cheng, S., Liu, Y., Escobar, C.M., Crowley, C.S., Jefferson, R.E., *et al.* (2010) Short N-terminal sequences package proteins into bacterial microcompartments. *Proc Natl Acad Sci U S A* **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. *Proc Natl Acad Sci U S A* **109:** 14995-15000.

Frank, S., Lawrence, A.D., Prentice, M.B., and Warren, M.J. (2013) Bacterial microcompartments moving into a synthetic biological world. *J Biotechnol* **163:** 273-279.

Fukuzawa, H., Suzuki, E., Komukai, Y., and Miyachi, S. (1992) A gene homologous to chloroplast carbonic anhydrase (icfA) is essential to photosynthetic carbon dioxide fixation by *Synechococcus* PCC7942. *Proc Natl Acad Sci U S A* **89:** 4437-4441.

Gebremariam, S.Y., Beutel, M.W., Christian, D., and Hess, T.F. (2011) Research advances and challenges in the microbiology of enhanced biological phosphorus removal-a critical review. *Water Environ Res* **83:** 195-219.

Grady Jr, C.L., Daigger, G.T., Love, N.G., and Filipe, C.D. (2011) *Biological wastewater treatment*: CRC press.

Grinkova, Y.V., Denisov, I.G., and Sligar, S.G. (2010) Engineering extended membrane scaffold proteins for self-assembly of soluble nanoscale lipid bilayers. *Protein Eng Des Sel* **23:** 843-848.

Havemann, G.D., and Bobik, T.A. (2003) Protein content of polyhedral organelles involved in coenzyme B12-dependent degradation of 1,2-propanediol in *Salmonella enterica serovar Typhimurium* LT2. *J Bacteriol* **185:** 5086-5095.

Havemann, G.D., Sampson, E.M., and Bobik, T.A. (2002) PduA is a shell protein of polyhedral organelles involved in coenzyme B(12)-dependent degradation of 1,2-propanediol in *Salmonella enterica serovar typhimurium* LT2. *J Bacteriol* **184:** 1253-1261.

Held, M., Kolb, A., Perdue, S., Hsu, S.Y., Bloch, S.E., Quin, M.B., and Schmidt-Dannert, C. (2016) Engineering formation of multiple recombinant Eut protein nanocompartments in *E. coli. Sci Rep* **6:** 24359.

Hess, W.R., Rocap, G., Ting, C.S., Larimer, F., Stilwagen, S., Lamerdin, J., and Chisholm, S.W. (2001) The photosynthetic apparatus of *Prochlorococcus*: Insights through comparative genomics. *Photosynth Res* **70:** 53-71.

Honda, S., Toraya, T., and Fukui, S. (1980) In situ reactivation of glycerol-inactivated coenzyme B12-dependent enzymes, glycerol dehydratase and diol dehydratase. *J Bacteriol* **143:** 1458-1465.

Horswill, A.R., and Escalante-Semerena, J.C. (1997) Propionate catabolism in *Salmonella typhimurium* LT2: two divergently transcribed units comprise the prp locus at 8.5 centisomes, *prpR* encodes a member of the sigma-54 family of activators, and the *prpBCDE* genes constitute an operon. *J Bacteriol* **179:** 928-940.

Horswill, A.R., and Escalante-Semerena, J.C. (1999) *Salmonella typhimurium* LT2 catabolizes propionate via the 2-methylcitric acid cycle. *J Bacteriol* **181:** 5615-5623.

Huber, I., Palmer, D.J., Ludwig, K.N., Brown, I.R., Warren, M.J., and Frunzke, J. (2017) Construction of Recombinant Pdu Metabolosome Shells for Small Molecule Production in *Corynebacterium glutamicum. ACS Synth Biol*.

Hyde, C.C., Ahmed, S.A., Padlan, E.A., Miles, E.W., and Davies, D.R. (1988) Three-dimensional structure of the tryptophan synthase alpha 2 beta 2 multienzyme complex from *Salmonella typhimurium. J Biol Chem* **263:** 17857-17871.

Hyde, C.C., and Miles, E.W. (1990) The tryptophan synthase multienzyme complex: exploring structure-function relationships with X-ray crystallography and mutagenesis. *Biotechnology (N Y)* **8:** 27-32.

Iancu, C.V., Ding, H.J., Morris, D.M., Dias, D.P., Gonzales, A.D., Martino, A., and Jensen, G.J. (2007) The structure of isolated *Synechococcus strain* WH8102 carboxysomes as revealed by electron cryotomography. *J Mol Biol* **372:** 764-773.

Iancu, C.V., Morris, D.M., Dou, Z., Heinhorst, S., Cannon, G.C., and Jensen, G.J. (2010) Organization, structure, and assembly of alpha-carboxysomes determined by electron cryotomography of intact cells. *J Mol Biol* **396:** 105-117.

Jakobson, C.M., Kim, E.Y., Slininger, M.F., Chien, A., and Tullman-Ercek, D. (2015) Localization of proteins to the 1,2-propanediol utilization microcompartment by non-native signal sequences is mediated by a common hydrophobic motif. *J Biol Chem* **290:** 24519-24533.

Jakobson, C.M., Slininger Lee, M.F., and Tullman-Ercek, D. (2017a) *De novo* design of signal sequences to localize cargo to the 1,2-propanediol utilization microcompartment. *Protein Sci* **26:** 1086-1092.

Jakobson, C.M., Tullman-Ercek, D., Slininger, M.F., and Mangan, N.M. (2017b) A systems-level model reveals that 1,2-Propanediol utilization microcompartments enhance pathway flux through intermediate sequestration. *PLoS Comput Biol* **13:** e1005525.

Jeter, R.M. (1990) Cobalamin-dependent 1,2-propanediol utilization by *Salmonella typhimurium. J Gen Microbiol* **136:** 887-896.

Johnson, C.L., Pechonick, E., Park, S.D., Havemann, G.D., Leal, N.A., and Bobik, T.A. (2001) Functional genomic, biochemical, and genetic characterization of the *Salmonella pduO* gene, an ATP:cob(I)alamin adenosyltransferase gene. *J Bacteriol* **183:** 1577-1584.

Kerfeld, C.A., and Erbilgin, O. (2015) Bacterial microcompartments and the modular construction of microbial metabolism. *Trends Microbiol* **23:** 22-34.

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

Kerfeld, C.A., and Melnicki, M.R. (2016) Assembly, function and evolution of cyanobacterial carboxysomes. *Curr Opin Plant Biol* **31:** 66-75.

Kerfeld, C.A., Sawaya, M.R., Tanaka, S., Nguyen, C.V., Phillips, M., Beeby, M., and Yeates, T.O. (2005) Protein structures forming the shell of primitive bacterial organelles. *Science* **309:** 936-938.

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

Lassila, J.K., Bernstein, S.L., Kinney, J.N., Axen, S.D., and Kerfeld, C.A. (2014) Assembly of robust bacterial microcompartment shells using building blocks from an organelle of unknown function. *J Mol Biol* **426:** 2217-2228.

Lawrence, A.D., Frank, S., Newnham, S., Lee, M.J., Brown, I.R., Xue, W.F., *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.

Lawrence, J.G., and Roth, J.R. (1996) Selfish operons: horizontal transfer may drive the evolution of gene clusters. *Genetics* **143:** 1843-1860.

Leal, N.A., Havemann, G.D., and Bobik, T.A. (2003) PduP is a coenzyme-a-acylating propionaldehyde dehydrogenase associated with the polyhedral bodies involved in B12-dependent 1,2-propanediol degradation by *Salmonella enterica serovar Typhimurium* LT2. *Arch Microbiol* **180:** 353-361.

Lee, M.J., Brown, I.R., Juodeikis, R., Frank, S., and Warren, M.J. (2016) Employing bacterial microcompartment technology to engineer a shell-free enzyme-aggregate for enhanced 1,2-propanediol production in *Escherichia coli. Metab Eng* **36:** 48-56.

Liang, M., Frank, S., Lunsdorf, H., Warren, M.J., and Prentice, M.B. (2017) Bacterial microcompartment-directed polyphosphate kinase promotes stable polyphosphate accumulation in *E. coli. Biotechnol J* **12:** 10.1002/biot.201600415. Epub 2017 Feb 10.

Lin, M.T., Occhialini, A., Andralojc, P.J., Devonshire, J., Hines, K.M., Parry, M.A., and Hanson, M.R. (2014) beta-Carboxysomal proteins assemble into highly organized structures in *Nicotiana* chloroplasts. *Plant J* **79:** 1-12.

Liu, Y., Leal, N.A., Sampson, E.M., Johnson, C.L., Havemann, G.D., and Bobik, T.A. (2007) PduL is an evolutionarily distinct phosphotransacylase involved in B12-

dependent 1,2-propanediol degradation by *Salmonella enterica serovar typhimurium* LT2*. J Bacteriol* **189:** 1589-1596.

Long, B.M., Badger, M.R., Whitney, S.M., and Price, G.D. (2007) Analysis of carboxysomes from *Synechococcus* PCC7942 reveals multiple Rubisco complexes with carboxysomal proteins CcmM and CcaA*. J Biol Chem* **282:** 29323-29335.

Marchetti, J.M., Miguel, V.U., and Errazu, A.F. (2007) Possible methods for biodiesel production. *Renew. Sustain. Energy Rev*. **11:** 1300-1311.

Marcus, Y., Berry, J.A., and Pierce, J. (1992) Photosynthesis and photorespiration in a mutant of the cyanobacterium *Synechocystis* PCC 6803 lacking carboxysomes*. Planta* **187:** 511-516.

Mayer, M.J., Juodeikis, R., Brown, I.R., Frank, S., Palmer, D.J., Deery, E.*, et al.* (2016) Effect of bio-engineering on size, shape, composition and rigidity of bacterial microcompartments*. Sci Rep* **6:** 36899.

McBroom, A.J., Johnson, A.P., Vemulapalli, S., and Kuehn, M.J. (2006) Outer membrane vesicle production by *Escherichia coli* is independent of membrane instability*. J Bacteriol* **188:** 5385-5392.

Menon, B.B., Dou, Z., Heinhorst, S., Shively, J.M., and Cannon, G.C. (2008) *Halothiobacillus neapolitanus* carboxysomes sequester heterologous and chimeric RubisCO species*. PLoS One* **3:** e3570.

Miles, E.W., Rhee, S., and Davies, D.R. (1999) The Molecular Basis of Substrate Channeling*. Journal of Biological Chemistry* **274:** 12193-12196.

Mino, T. (2000) Microbial selection of polyphosphate-accumulating bacteria in activated sludge wastewater treatment processes for enhanced biological phosphate removal*. Biochemistry (Mosc)* **65:** 341-348.

Morell, M.K., Paul, K., O'Shea, N.J., Kane, H.J., and Andrews, T.J. (1994) Mutations of an active site threonyl residue promote beta elimination and other side reactions of the enediol intermediate of the ribulosebisphosphate carboxylase reaction*. J Biol Chem* **269:** 8091-8098.

Mori, K., Tobimatsu, T., and Toraya, T. (1997) A Protein Factor Is Essential for in Situ Reactivation of Glycerol-inactivated Adenosylcobalamin-dependent Diol Dehydratase*. Biosci Biotechnol Biochem* **61:** 1729-1733.

Moroney, J.V., Jungnick, N., Dimario, R.J., and Longstreth, D.J. (2013) Photorespiration and carbon concentrating mechanisms: two adaptations to high O_2, low CO_2 conditions*. Photosynth Res* **117:** 121-131.

Noël, C.R., Cai, F., and Kerfeld, C.A. (2016) Protein Nanotubes: Purification and Characterization of Protein Nanotubes Assembled from a Single Bacterial Microcompartment Shell Subunit. *Adv. Mater. Interfaces* **3**: 1500295

Obradors, N., Badia, J., Baldoma, L., and Aguilar, J. (1988) Anaerobic metabolism of the L-rhamnose fermentation product 1,2-propanediol in *Salmonella typhimurium. J Bacteriol* **170:** 2159-2162.

Palacios, S., Starai, V.J., and Escalante-Semerena, J.C. (2003) Propionyl Coenzyme A Is a Common Intermediate in the 1,2-Propanediol and Propionate Catabolic Pathways Needed for Expression of the *prpBCDE* Operon during Growth of *Salmonella enterica* on 1,2-Propanediol. *Journal of Bacteriology* **185:** 2802-2810.

Pang, A., Frank, S., Brown, I., Warren, M.J., and Pickersgill, R.W. (2014) Structural insights into higher order assembly and function of the bacterial microcompartment protein PduA. *J Biol Chem* **289:** 22377-22384.

Pang, A., Liang, M., Prentice, M.B., and Pickersgill, R.W. (2012) Substrate channels revealed in the trimeric *Lactobacillus reuteri* bacterial microcompartment shell protein PduB. *Acta Crystallogr D Biol Crystallogr* **68:** 1642-1652.

Parsons, J.B., Dinesh, S.D., Deery, E., Leech, H.K., Brindley, A.A., Heldt, D.*, et al.* (2008) Biochemical and structural insights into bacterial organelle form and biogenesis. *J Biol Chem* **283:** 14366-14375.

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

Price, G.D., and Badger, M.R. (1991) Evidence for the role of carboxysomes in the cyanobacterial CO2-concentrating mechanism. *Can. J. Bot.* **69:** 963-973.

Price, G.D. (2011) Inorganic carbon transporters of the cyanobacterial CO2 concentrating mechanism. *Photosynth Res* **109:** 47-57.

Price, G.D., and Badger, M.R. (1989) Expression of Human Carbonic Anhydrase in the *Cyanobacterium Synechococcus* PCC7942 Creates a High CO_2-Requiring Phenotype : Evidence for a Central Role for Carboxysomes in the CO_2 Concentrating Mechanism. *Plant Physiol* **91:** 505-513.

Price, G.D., Badger, M.R., Woodger, F.J., and Long, B.M. (2008) Advances in understanding the cyanobacterial CO2-concentrating-mechanism (CCM): functional components, Ci transporters, diversity, genetic regulation and prospects for engineering into plants. *J Exp Bot* **59:** 1441-1461.

Quin, M.B., Perdue, S.A., Hsu, S.Y., and Schmidt-Dannert, C. (2016) Encapsulation of multiple cargo proteins within recombinant Eut nanocompartments. *Appl Microbiol Biotechnol* **100:** 9187-9200.

Rae, B.D., Long, B.M., Badger, M.R., and Price, G.D. (2013) Functions, compositions, and evolution of the two types of carboxysomes: polyhedral microcompartments that facilitate CO2 fixation in cyanobacteria and some proteobacteria. *Microbiol Mol Biol Rev* **77:** 357-379.

Roberts, E.W., Cai, F., Kerfeld, C.A., Cannon, G.C., and Heinhorst, S. (2012) Isolation and characterization of the *Prochlorococcus* carboxysome reveal the presence of the novel shell protein CsoS1D. *J Bacteriol* **194:** 787-795.

Sagermann, M., Ohtaki, A., and Nikolakakis, K. (2009) Crystal structure of the EutL shell protein of the ethanolamine ammonia lyase microcompartment. *Proc Natl Acad Sci U S A* **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.

Sargent, F., Davidson, F.A., Kelly, C.L., Binny, R., Christodoulides, N., Gibson, D., *et al.* (2013) A synthetic system for expression of components of a bacterial microcompartment. *Microbiology* **159:** 2427-2436.

Shelley, S. (2007) A renewable route to propylene glycol. *Chem. Eng. Prog.* **103:** 6-9.

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.

Shively, J.M., Bock, E., Westphal, K., and Cannon, G.C. (1977) Icosahedral inclusions (carboxysomes) of *Nitrobacter agilis. J Bacteriol* **132:** 673-675.

Sinha, S., Cheng, S., Fan, C., and Bobik, T.A. (2012) The PduM protein is a structural component of the microcompartments involved in coenzyme B(12)-dependent 1,2-propanediol degradation by *Salmonella enterica. J Bacteriol* **194:** 1912-1918.

Spivey, H.O., and Merz, J.M. (1989) Metabolic Compartmentation. *Bioessays* **10:** 127-129.

Spivey, H.O., and Ovádi, J. (1999) Substrate Channeling. *Methods*. **19:** 306-321.

Srere, P.A. (1985) The metabolon. *Trends Biochem. Sci.***10:** 109-110.

Sutter, M., Greber, B., Aussignargues, C., and Kerfeld, C.A. (2017) Assembly principles and structure of a 6.5-MDa bacterial microcompartment shell. *Science* **356:** 1293-1297.

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.

Tcherkez, G.G., Farquhar, G.D., and Andrews, T.J. (2006) Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. *Proc Natl Acad Sci U S A* **103:** 7246-7251.

Tobimatsu, T., Kawata, M., and Toraya, T. (2005) The N-terminal regions of beta and gamma subunits lower the solubility of adenosylcobalamin-dependent diol dehydratase. *Biosci Biotechnol Biochem* **69:** 455-462.

Tsai, Y., Sawaya, M.R., Cannon, G.C., Cai, F., Williams, E.B., Heinhorst, S., *et al.* (2007) Structural analysis of CsoS1A and the protein shell of the *Halothiobacillus neapolitanus* carboxysome. *PLoS Biol* **5:** e144.

Ushio, K., Honda, S., Toraya, T., and Fukui, S. (1982) The mechanism of in situ reactivation of glycerol-inactivated coenzyme B12-dependent enzymes, glycerol dehydratase and diol dehydratase. *J Nutr Sci Vitaminol (Tokyo)* **28:** 225-236.

van Eykelenburg, C. (1980) Ecophysiological studies on *Spirulina platensis*. Effect of temperature, light intensity and nitrate concentration on growth and ultrastructure. *Antonie Van Leeuwenhoek* **46:** 113-127.

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., Kerfeld, C.A., Heinhorst, S., Cannon, G.C., and Shively, J.M. (2008) Protein-based organelles in bacteria: carboxysomes and related microcompartments. *Nat Rev Microbiol* **6:** 681-691.

Zarzycki, J., Axen, S.D., Kinney, J.N., and Kerfeld, C.A. (2013) Cyanobacterial-based approaches to improving photosynthesis in plants. *J Exp Bot* **64:** 787-798.

Chapter 5

Contributions to publications

5.1. Employing bacterial microcompartment technology to engineer a shell-free enzyme-aggregate for enhanced 1,2-propanediol production in *Escherichia coli*

Lee, M.J., Brown, I.R., Juodeikis, R., Frank, S., and Warren, M.J. (2016) Employing bacterial microcompartment technology to engineer a shell-free enzyme-aggregate for enhanced 1,2-propanediol production in *Escherichia coli*. *Metab Eng* 36: 48-56.

For this publication I contributed to experimental design alongside S. Frank and M. J. Warren. I wrote the manuscript, which was then reviewed by all authors. I made all figures apart from the abstract figure that was made in collaboration with R. Juodeikis.

Experimentally, I conducted all experimental work excluding thin sectioning of TEM samples, which was carried out by I. R. Brown.

<u>The following sections of the publication have been examined previously for an MSc by research:</u>

3.1. Effect of targeting peptides on the activities of GldA, DhaK, MgsA and FucO.

3.3. Proteins with targeting peptides are recruited to BMCs

The first paragraph of this section concerning targeting of enzymes to recombinant BMCs and associated data (Supplementary Fig. S6) has previously been examined.

5.2. Patent application: Genetically Modified Microorganisms

Lee, M.J., Frank, S. Warren, M.J. 2015. Genetically Modified Microorganisms. Filed November 2015. International application number: PCT/GB2016/053435.

The patent was written between a patent attorney and myself. All inventors reviewed the patent alongside the university's senior commercialisation manager Dr Gary Robinson. The contributions to experimental work are set out in section 5.1.

The experimental data in the following sections have been previously examined:

"Effect of fusing enzymes to BMC-targeting peptides on enzyme specific activities"

"Enzymes fused to BMC-targeting peptides are recruited to BMCs"

The second paragraph of this section has been previously examined.

5.3. Engineered synthetic scaffolds for organising proteins within bacterial cytoplasms

Lee, M.J., Mantell, J., Hodgson, L., Alibhai, D., Fletcher, J.M., Brown, I.R., Frank, S., Xue, W-F., Verkade, P., Woolfson, D.N., Warren, M.J. (2018) Engineered synthetic scaffolds for organizing proteins within the bacterial cytoplasm. *Nat Chem Biol* 14(2): 142-147

The project was conceived between myself, D. N. Woolfson and M. J. Warren.

Experimentally I made constructs, prepared samples for TEM and confocal analysis, imaged samples by TEM, purified nanotubes and analysed them by TEM, AFM and conducted the ethanol production experiments and analyses. J. Mantell undertook tomography and 3D reconstructions. L. Hodgson undertook CLEM sample preparation and imaging. D. Alibhai undertook confocal imaging. I. R. Brown sectioned samples for TEM analysis. W. F. Xue assisted with AFM and statistical analysis. M. J. Lee, J. Mantell, L. Hodgson, J. M. Fletcher, S. Frank, P. Verkade, D. N. Woolfson and M. J. Warren designed the experiments.

The manuscript was written by myself and edited by all authors.

5.4. *De novo* targeting to the cytoplasmic and luminal side of bacterial microcompartments

Lee, M.J., Brown, I.R., Fletcher, J.M., Pickersgill, R.W., Woolfson, D.N., Frank, S., Warren, M.J. *De novo* targeting to the cytoplasmic and luminal side of bacterial microcompartments. Nat. Commun (Under review)

The project was conceptualised between myself, M. J. Warren and D.N. Woolfson.

I conducted all experimental work, excluding the thin sectioning of samples for TEM analysis (I. R. Brown). M. J. Lee, J. M. Fletcher, R. W. Pickersgill, D. N. Woolfson, S. Frank and M. J. Warren designed the experiments. The manuscript was written by myself and edited by all authors.

Chapter 6

Appendix

6.1. Employing bacterial microcompartment technology to engineer a shell-free enzyme-aggregate for enhanced 1,2-propanediol production in *Escherichia coli*.

Metabolic Engineering 36 (2016) 48–56

Contents lists available at ScienceDirect

Metabolic Engineering

journal homepage: www.elsevier.com/locate/ymben

Employing bacterial microcompartment technology to engineer a shell-free enzyme-aggregate for enhanced 1,2-propanediol production in *Escherichia coli*

Matthew J. Lee [a], Ian R. Brown [a], Rokas Juodeikis [a], Stefanie Frank [a,*], Martin J. Warren [a,*]

[a] School of Biosciences, University of Kent, Giles Lane, Canterbury, Kent CT2 7NJ, UK

ARTICLE INFO

Article history:
Received 18 December 2015
Received in revised form
5 February 2016
Accepted 22 February 2016
Available online 8 March 2016

Keywords:
Synthetic biology
Metabolic engineering
Compartmentalisation
Protein aggregation
Biotechnology

ABSTRACT

Bacterial microcompartments (BMCs) enhance the breakdown of metabolites such as 1,2-propanediol (1,2-PD) to propionic acid. The encapsulation of proteins within the BMC is mediated by the presence of targeting sequences. In an attempt to redesign the Pdu BMC into a 1,2-PD synthesising factory using glycerol as the starting material we added N-terminal targeting peptides to glycerol dehydrogenase, dihydroxyacetone kinase, methylglyoxal synthase and 1,2-propanediol oxidoreductase to allow their inclusion into an empty BMC. 1,2-PD producing strains containing the fused enzymes exhibit a 245% increase in product formation in comparison to un-tagged enzymes, irrespective of the presence of BMCs. Tagging of enzymes with targeting peptides results in the formation of dense protein aggregates within the cell that are shown by immuno-labelling to contain the vast majority of tagged proteins. It can therefore be concluded that these protein inclusions are metabolically active and facilitate the significant increase in product formation.

1. Introduction

Metabolic engineering involves the design and redesign of pathways and their deployment in organisms in which they do not naturally exist. This approach allows pathway fluxes, together with substrate and intermediate concentrations, to be manipulated by variation of the network parameters, which can be quantified by metabolic control analysis (Woolston et al., 2013). However, for pathways involving particularly volatile, unstable or toxic intermediates this tactic is likely to prove problematic. To overcome the problem of capricious metabolites nature has evolved a variety of solutions to ensure pathways operate efficiently without a significant build up of pernicious intermediates. In this respect substrate channelling, multienzyme complexes, metabolons and compartmentalisation are all ways in which pathway flux is naturally controlled (Lee et al., 2012).

In bacteria, compartmentalisation is mediated through the deployment of bacterial microcompartments (BMCs), which are used to address the problem of unstable or reactive intermediates (Dou et al., 2008; Havemann et al., 2002; Penrod and Roth, 2006; Sampson and Bobik, 2008). BMCs are proteinaceous complexes that are composed of a semi-porous capsid shell that encases a specific metabolic process (Cheng et al., 2008; Frank et al., 2013; Tanaka et al., 2008; Yeates et al., 2008). The widespread dispersal of BMCs in 23 bacterial phyla mediated through horizontal gene transfer suggests that pathway enhancement through employment of these structures provides a strong evolutionary benefit (Axen et al., 2014). There are two broad classes of BMCs, carboxysomes and metabolosomes, which are associated with either anabolic carbon fixation or catabolic carbon utilisation, respectively. Metabolosomes, in particular, appear to operate pathways that involve aldehydes as intermediates. Indeed, a recent bioinformatic analysis of BMC-associated operons revealed that the vast majority of these operons encode for aldehyde and alcohol dehydrogenases (Axen et al., 2014). The best characterised of the metabolosomes are those associated with 1,2-propanediol utilisation (Pdu) and ethanolamine utilisation (Eut), both of which house cobalamin-dependent enzymes and encase pathways that proceed via propanaldehyde and acetaldehyde respectively (Chowdhury et al., 2014). Compelling evidence has been presented that the compartments help protect the cell from toxicity associated with a high aldehyde concentration (Brinsmade et al., 2005; Sampson and Bobik, 2008; Cheng et al., 2011).

The ability to concentrate a specific metabolic pathway into what is essentially a nano-bioreactor, coupled with the capacity to

Abbreviations: BMC, bacterial microcompartment; Pdu, 1,2-propanediol utilisation; 1,2-PD, 1,2-propanediol; D18, First 18 amino acids of PduD; P18, First 18 amino acids of PduP

* Corresponding authors.
E-mail addresses: S.Frank@kent.ac.uk (S. Frank), m.j.warren@kent.ac.uk (M.J. Warren).

http://dx.doi.org/10.1016/j.ymben.2016.02.007

sequester toxic pathway intermediates, has brought BMCs to the attention of synthetic biologists who view this as a tractable system that can be redesigned to accommodate new pathways (Chowdhury et al., 2014; Lawrence et al., 2014). Such a system has the potential to be used to enhance the yield of commodity chemicals produced via bacterial fermentation. Significantly, the *Citrobacter freundii* Pdu BMC can be produced as an empty compartment through the coordinated production of only the shell proteins (PduA, B, B′, J, K, N, U) (Parsons et al., 2010). Enzymes can then be targeted so that they are incorporated into the BMC through the fusion of peptide sequences that are found at the N-terminus of proteins such as PduD (D18) and PduP (P18) (Fan et al., 2010; Fan and Bobik, 2011). Furthermore, targeting of the *Zymomonas mobilis* pyruvate decarboxylase and alcohol dehydrogenase resulted in the conversion of the Pdu BMC into an ethanol bioreactor (Lawrence et al., 2014).

Bio-product commodities that have successfully transitioned into the market through biotechnological approaches include 1,3-propanediol, polylactic acid (PLA) and polyhydroxyalkanoate (PHA), which have been used for personal care products, antifreeze and biodegradable plastics (Adkesson et al., 2011; Jung et al., 2010; Suriyamongkol et al., 2007). Near-term bio-based products, such as 1,4-butanediol, isobutanol and succinic acid are in progress whilst systems are under development for the production of terpenes and itaconic acid (Burk, et al., 2011; Lee et al., 2005; Peralta-Yahyan et al., 2011, 2012; Steiger et al., 2013; Yim et al., 2011). Approaches such as metabolic engineering and synthetic biology are routinely applied in order to make these processes more efficient and cost competitive. In this paper we outline a method that offers the potential for a significant step-change in bio-commodity production through the development of BMC technology.

The production of 1,2-propanediol from glycerol is similarly recognised as a commercially relevant pathway. 1,2-Propanediol is a commodity chemical that is currently used in the production of plasticisers, antifreeze, thermoset plastics and cosmetics with an annual global demand estimated at around 1.36 million tonnes per year with demand expected to increase over the next few years (Clomburg and Gonzalez, 2011). It is, therefore, of great interest to develop a production method that does not rely on a non-renewable resource (Altaras and Cameron, 1999, 2000; Clomburg and Gonzalez, 2011). Glycerol, on the other hand, is readily available as it is produced as a by-product of the biodiesel production process (Marchetti et al., 2007). It has been reported that for every 100 kg of biodiesel produced 10 kg of glycerol is generated (Yazdani and Gonzalez, 2007). The biochemical pathway for the synthesis of 1,2-propanediol from glycerol (Fig. 1) involves the intermediate methylglyoxal, a compound that is highly toxic to cells in sub-millimolar concentrations (Ferguson et al., 1996). The pathway involves four enzymes, glycerol dehydrogenase (GldA), dihydroxyacetone kinase (DhaK), methylglyoxal synthase (MgsA) and 1,2-propanediol oxidoreductase (FucO). Previously it has been shown that a DNA scaffold enhances 1,2-propanediol production in *Escherichia coli* from glucose with 3 enzymes including MgsA and GldA (Conrado et al., 2012). Here we set out an alternative approach to determine if the proposed pathway for 1,2-propanediol production could be enhanced through compartmentalisation into a BMC. Moreover we present an alternative approach to compartmentalisation that is the aggregation of enzymes into a supramolecular conglomerate.

The aim of the investigation was therefore to set about creating fusion proteins between known Pdu targeting peptides (D18 and P18) and the four 1,2-propanediol producing enzymes in order to allow their targeting to a recombinant empty Pdu BMC system. The effect of the targeting peptides on the activity of the different enzymes and their solubility was investigated. The ability of the targeted enzymes to promote 1,2-propanediol synthesis was determined. The strains were analysed by TEM and protein aggregation was found to play an unexpected but key role in enhancing pathway productivity.

2. Materials and methods

2.1. Plasmid construction

Plasmids were constructed to provide each of the genes of interest with a N-terminal hexa-histidine tag with an optional D18 or P18 targeting peptide.

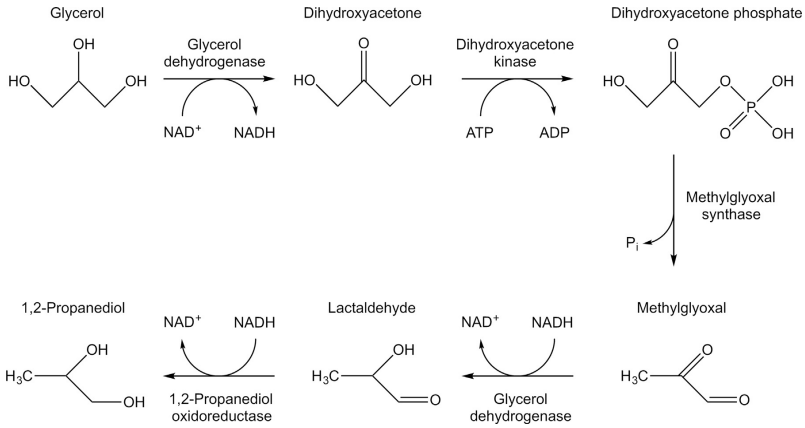


Fig. 1. A pathway for the synthesis of 1,2-propanediol from glycerol. Glycerol dehydrogenase and dihydroxyacetone kinase catalyse the conversion of glycerol to dihydroxyacetone phosphate. Methylglyoxal synthase catalyses the conversion to methylglyoxal. Glycerol dehydrogenase and 1,2-propanediol oxidoreductase catalyse the conversion of methylglyoxal to 1,2-propanediol via the intermediate lactaldehyde.

All primers used in this study are listed in Supplementary Table 3. All genes were amplified with flanking *Nde*I and *Spe*I restriction sites and each was ligated into pET14b, pET14b-D18 and pET14b-P18 vectors using *Nde*I and *Spe*I restriction sites. Plasmids pML-1 to pML-6 as outlined in Supplementary Table 2, were constructed by a 'Link and Lock' approach utilising the compatible sticky ends formed by digestion with *Xba*I and *Spe*I (McGoldrick et al., 2005).

2.2. Overexpression and purification of recombinant protein

BL21 (DE3) pLysS competent cells were transformed with a plasmid containing the gene(s) of interest. 1 L of LB supplemented with ampicillin (100 mg/L) and chloramphenicol (34 mg/L) in baffled flasks was inoculated from an overnight starter culture. The cultures were grown at 37 °C with shaking for 7 h; protein production was induced by the addition of IPTG to a final concentration of 400 μM. The cultures were then incubated overnight at 19 °C with shaking. Cells were harvested by centrifugation at 3320 × *g* for 15 minutes at 4 °C, pellets were resuspended in 20 mM Tris–HCl pH 8.0, 500 mM NaCl, 5 mM imidazole. Cells were lysed by sonication and cell debris removed by centrifugation. Recombinant protein was then purified from the soluble fraction by immobilized metal ion affinity chromatography.

2.3. Activity assays

2.3.1. Glycerol dehydrogenase

The activity of GldA for the oxidation of glycerol to dihydroxyacetone was measured by following the initial rate of reduction of NAD^+ to NADH at 340 nm. Activity assays were carried out in 1 mL reactions containing 0.1 M potassium phosphate buffer pH 8.0, 500 μM NAD^+, 2 mM $MgCl_2$ and 200 nM GldA. The activity of GldA for the reduction methylglyoxal to lactaldehyde, was measured by following the initial rate of the oxidation of NADH to NAD^+ at 340 nm. Activity assays were carried out in 1 mL reaction containing 0.1 M potassium phosphate pH 8.0, 0.1 mM NADH, 2 mM $MgCl_2$ and 200 nM GldA.

2.3.2. Dihydroxyacetone kinase

The activity of DhaK for the conversion of dihydroxyacetone to dihydroxyacetone phosphate was measured in a coupled reaction with Glyceraldehyde 3-phosphate dehydrogenase (G3PDH) by following the oxidation of NADH to NAD^+ at 340 nm. Activity assays were carried out in 1 mL reactions containing 50 mM Tris–HCl pH 7.5, 100 mM NaCl, 1 mM ATP, 0.1 mM NADH, 2.5 mM $MgCl_2$, 7.2 U G3PDH and 125 nM DhaK.

2.3.3. Methylglyoxal synthase

The activity of MgsA was monitored in a colorimetric assay over a time course. 25 μL of 0.5 mM MgsA was incubated in a reaction mixture containing 400 μL 50 mM imidazole pH 7.0, 25 μL 15 mM dihydroxyacetone phosphate and 50 μL dH_2O, the reaction mixture was incubated at 30 °C with shaking. At time intervals 50 μL of the reaction mixture was removed and added to a detection mixture containing 450 μL dH_2O, 165 μL 0.1% 2,4-dinitrophenylhydrazine hydrochloric acid solution. The detection mixture was incubated at 30 °C with shaking for 15 min. 835 μL of 10% (w/v) NaOH was added to the detection mixture which was incubated at room temperature for 15 minutes. Absorbance was then measured at 550 nm.

2.3.4. 1,2-Propanediol oxidoreductase

The activity of FucO was determined for the NADH dependant reduction of glycolaldehyde to elthylene glycol was measured by following the initial rate of the oxidation of NADH to NAD^+ at 340 nm. Activity assays were carried out in 1 mL reactions containing 100 mM Hepes pH 7.5, 10 μM NADH, 100 μM $MnCl_2$ and 200 nM FucO.

2.4. Culture medium and conditions for 1,2-propanediol production

The culture medium designed by Neidhardt et al. (1974) was supplemented with 30 g/L glycerol, 10 g/L tryptone and 5 g/L yeast extract. Strains were cultured in sealed serum bottles with a working volume of 100 mL at 28 °C with shaking. Cultures were inoculated from starter cultures to a starting OD_{600} of 0.05. During growth 1 mL samples were removed at 0, 6, 12, 24, 48, 72 and 96 h for analysis of 1,2-propanediol content.

2.5. Western blot analysis

Nitrocellulose membranes following transfer and blocking were incubated in primary antibody (mouse anti-His (Sigma Aldrich) 1:3000 or mouse anti-GFP (Sigma Aldrich) 1:1000) followed by incubation in a secondary coupled antibody coupled to alkaline phosphatase (Anti-Mouse IgG (H + L), AP Conjugate (Promega) 1:5000). Bands were visualised by incubation in substrate 5-bromo-4-chloro-3-indolyl phosphate/nitro blue tetrazolium (BCIP/NBT).

2.6. Analysis of 1,2-propanediol production

In-vivo 1,2-propanediol production was determined by GC/MS analysis of the growth medium at time intervals (0, 6, 12, 24, 48, 72 and 96 h). The supernatant after centrifugation, was boiled for 10 min at 100 °C followed by centrifugation at 19,750 × *g*. The sample was then acidified with trifluoroacetic acid to a final concentration of 0.01% followed by a second centrifugation at 19,750 × *g*. The supernatant following centrifugation was diluted 1:4 in acetonitrile for GC/MS analysis.

2.7. Visualisation of engineered strains

2.7.1. Embedding of strains for TEM analysis

Strains were embedded, sectioned and stained as described in supplementary information.

2.7.2. Embedding of strains for immunolabelling

Strains were cultured as described previously (Section 2.4) overnight, cells were harvested by centrifugation for 10 min at 3000 × *g*. The cell pellet was resuspended in 2% formaldehyde and 0.5% gluteraldehyde in 100 mM sodium cacodolate pH 7.2 and incubated for 2 h with gentle rotating. Cells were pelleted by centrifugation at 6000 × *g* for 2 min and were washed twice for 10 min with 100 mM sodium cacodylate pH 7.2. This was followed by dehydration of the samples in an ethanol gradient, 50% EtOH for 10 min, 70% EtOH for 10 min, 90% EtOH for 10 min, followed by three 15 min washes in 100% EtOH. Cell pellets were then resuspended in 2 mL LR white resin and incubated overnight with rotation at room temperature after which the resin was changed and incubated for a further 6 h. Cell pellets were resuspended in fresh resin and transferred to 1 mL gelatine capsules and centrifuged at 4000 × *g* to pellet the cells at the tip. Samples were polymerised at 60 °C for 24 h. Samples were ultra-thin sectioned on a RMC MT-XL ultramicrotome with a diamond knife (diatome 45°) sections (60–70 nm thick) were collected on 300 mesh gold grids.

2.7.3. Immunolabelling of sections

Grids were equilibrated in one drop of TBST (20 mM Tris–HCl pH 7.2, 500 mM NaCl, 0.05% Tween 20, 0.1% BSA) before being

transferred into a drop of 2% BSA in TBST and incubated at room temperature for 30 min. Grids were then immediately transferred into a 20 µL drop of primary antibody (mouse anti-his (Sigma Aldrich 1:10)) and incubated for 1 h. Grids were washed in a fresh drop of TBST followed by washing for 10 s in a stream of TBST. Grids were equilibrated in a drop of secondary antibody (Goat anti-mouse IgG 10 nm gold (Agar Scientific 1:50)) then incubated for 30 min in a fresh drop. Excess antibody was removed by washing in two drops of TBST before washing in a stream of ddH_2O and dried.

2.7.4. Staining of immunolabelled sections

Grids were stained for 15 minutes in 4.5% uranyl acetate in 1% acetic acid solution followed by 2 washes in dH_2O. Grids were then stained with Reynolds lead citrate for 3 min followed by a wash in ddH_2O. Electron microscopy was performed using a JEOL-1230 transmission electron microscope equipped with a Gatan multi-scan digital camera at an accelerating voltage of 80 kV.

3. Results

3.1. Effect of targeting peptides on the activities of GldA, DhaK, MgsA and FucO.

Recently, it has been shown that fusing the targeting peptides D18 and P18 to heterologous enzymes, such as the *Z. mobilis* pyruvate decarboxylase and alcohol dehydrogenase, allowed their targeting to recombinantly produced empty BMCs, resulting in the formation of a functional ethanol bioreactor (Lawrence et al., 2014). However, the general effect of such fusions on functionality of individual enzymes had not been investigated in detail. In this study the enzymes involved in the microbial synthesis of 1,2-propanediol from glycerol, namely glycerol dehydrogenase (GldA), dihydroxyacetone kinase (DhaK), methylglyoxal synthase (MgsA) and 1,2-propanediol oxidoreductase (FucO) were cloned separately with both the D18 and P18 N-terminal targeting peptides followed by a hexa-histidine tag. The resulting proteins were purified by IMAC and the kinetic parameters of each of the protein fusions were subsequently determined and compared to enzymes containing only the N-terminal hexa-histidine tag. Herein, we refer to the D18-His and P18-His containing proteins as 'tagged' enzymes and the His-only containing proteins as 'untagged'.

The targeting peptides were found to have a highly variable effect on the specific activities of the enzymes (Fig. 2). The first of the enzymes, GldA, is involved in two distinct steps in the transformation of glycerol to 1,2-propanediol, being responsible for the dehydrogenation of glycerol to dihydroxyacetone as well as the reduction of methylglyoxal to lactaldehyde. With respect to the activity of GldA in the dehydrogenation of glycerol, the D18 targeting peptide resulted in a decrease of 90% in the enzyme's specific activity in comparison to the un-tagged control. Tagging GldA with the P18 targeting peptide had a less dramatic effect although the specific activity was still reduced by 55%. The tags had a similar effect on the activity of GldA to catalyse the reduction of methylglyoxal to lactaldehyde, with the presence of D18 decreasing the specific activity by 83% whilst P18 reduced activity by 53% (Fig. 2a). In contrast, the activity of DhaK, which catalyses

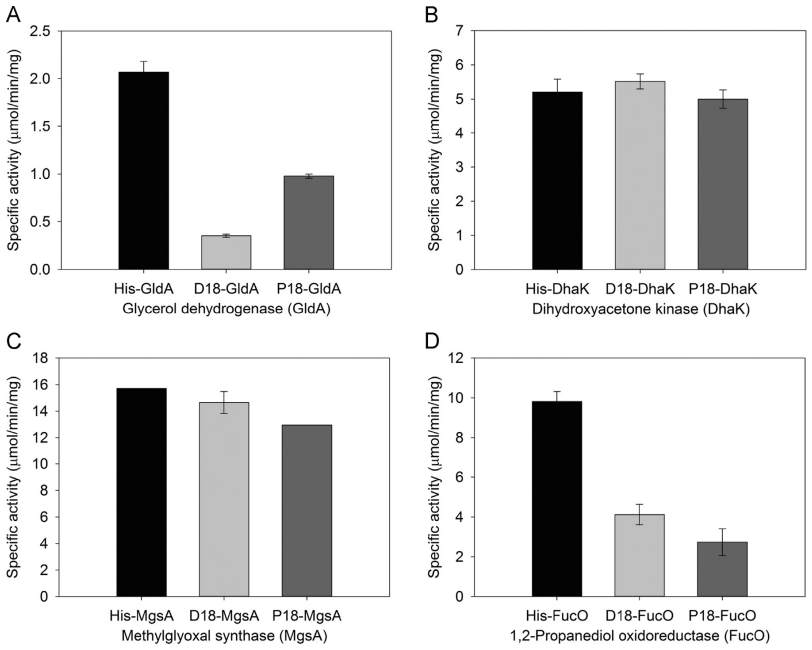


Fig. 2. The effect of targeting peptides on the specific activities of the enzymes involved in the microbial synthesis of 1,2-propanediol. (a) glycerol dehydrogenase (for the reduction of methylglyoxal) (b) dihydroxyacetone kinase (c) methylglyoxal synthase (d) 1,2-propanediol oxidoreductase.

the ATP dependent phosphorylation of dihydroxyacetone, was not significantly affected by the presence of either targeting peptide (Fig. 2b). Fusing MgsA, the methylglyoxal synthase, with either the D18 or P18 targeting peptide had only a slight detrimental effect on enzyme activity, decreasing the activity by 15% and 18% respectively (Fig. 2c). The activity of FucO, the 1,2-propanediol oxidoreductase, was found to decrease by 58% when fused to D18 but was more significantly affected by the fusion of P18, which resulted in a 76% decrease in specific activity (Fig. 2d). It is, therefore, interesting to note that although the D18 targeting peptide is predicted to be structurally similar to P18, the two tags were found to have a differential influence on the activities of the same enzymes. From the data presented in Fig. 2 the most active forms of the enzymes to be taken forward for inclusion into a BMC were identified as P18-GldA, P18-DhaK, D18-MgsA and D18-FucO.

3.2. Targeting peptides cause a degree of protein aggregation

The production levels and solubility of GldA, DhaK, MgsA and FucO, with and without targeting peptides, were investigated by comparing samples obtained during their purification by denaturing polyacrylamide gel electrophoresis (Supplementary Figs. S1–S4). DhaK and MgsA were both found to be well produced and soluble irrespective of the presence of the D18 or P18 targeting peptide. The solubility of FucO was, similarly, not affected by the presence of the targeting peptides, although the yield of un-tagged FucO appeared to be slightly lower. In contrast, even though both P18 and D18-tagged GldA appeared to be produced and could be purified, more of the protein was detected in the insoluble fractions (Supplementary Fig. S1, lane 3). This suggests that the fusion proteins D18-GldA and P18-GldA had a greater tendency to aggregate in comparison to untagged GldA. Moreover, P18-GldA was also found to elute from the IMAC column with an additional band of lower molecular mass, indicative of partial protein degradation.

Previous research has shown that the solubility of some of the enzymes of both the Pdu and Eut metabolosomes is increased by the removal of the N-terminal targeting peptides (Fan and Bobik, 2011; Shibata et al., 2010). Therefore the N-terminal peptides are known to affect solubility. It is clear that the addition of both D18 and P18 to GldA results in a decrease in solubility, most likely through protein aggregation. In order to investigate the aggregation behaviour of the tagged proteins in vivo, the most active protein fusions (P18-GldA, P18-DhaK, D18-MgsA and D18-FucO) were selected for TEM analysis of sections through whole cells. In this respect strains encoding each of the tagged proteins were cultured overnight without induction. Subsequently, the cells were harvested, embedded in low viscosity resin, thin sectioned and visualised using TEM. These were then compared to control strains that produced the untagged protein. For each strain 100 cells were examined and the statistical analysis of each of the strains is shown in Fig. 3. Representative TEM micrographs were compiled and are shown in Supplementary Fig. S5.

Control strains producing un-tagged proteins (GldA, DhaK, MgsA, FucO) displayed a 'normal' phenotype, with only a maximum of 1% of the observed cells containing electron dense areas that could be considered indicative of aggregated proteins. In contrast, half of all observed cells (52%) producing P18-GldA showed protein aggregates, which were located mainly at the poles of the cell (Supplementary Fig. S5 B). The addition of the P18 targeting peptide to the N-terminus of DhaK resulted in aggregate formation in 8% of the observed cells. Fusion of the D18 targeting peptide to MgsA and FucO resulted in the presence of protein aggregates in 12% and 4% of cells respectively. These results confirm that the fusions between the enzymes of the 1,2-propanediol production pathway and the D18 and P18 targeting peptides cause protein aggregation to various extents.

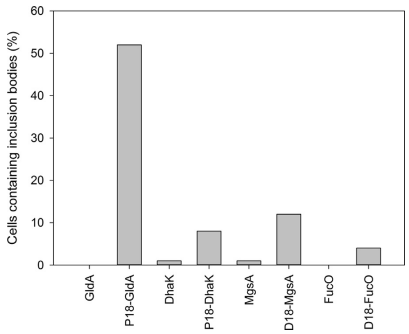


Fig. 3. Statistical analysis showing the percentage of cells expressing tagged/untagged proteins containing inclusion bodies. For each strain 100 cells were observed by TEM.

3.3. Proteins with targeting peptides are recruited to BMCs

The D18 and P18 fusion proteins were investigated for their ability to be targeted to an empty recombinant BMC. This was achieved by generating strains with the ability to co-express, individually, *P18-gldA*, *D18-dhaK*, *P18-mgsA* and *D18-fucO* together with the construct housing the genes for empty shell formation (pLysS-*pduABJKNU*). After growth, the recombinant BMCs were purified using a combination of centrifugation and differential salt precipitation as described previously (Lawrence et al., 2014). The isolated BMCs were then analysed by SDS PAGE for the presence of the characteristic BMC shell–protein profile. This revealed that all the tagged proteins co-purified with the BMC fraction to a greater extent than untagged protein, consistent with the proteins being encased within the BMC (Supplementary Fig. S6). This was further confirmed by kinetic assays of the final purified BMC fraction for the respective tagged enzymes.

To provide further evidence that the D18 and P18 fusions result in BMC encasement we chose to use a protease protection assay that was previously reported by Sargent et al. (2013). In this assay we used GFP as a marker to demonstrate that the protein is protected within the confines of the BMC. To this end plasmids were constructed containing GFP fused to an N-terminal D18 or P18 tag as well as a C-terminal SsrA proteolysis tag (AANDENYALAA*). The C-terminal SsrA tag targets proteins for degradation by the *E. coli* proteases ClpAP and ClpXP (Farrell et al., 2005). *E. coli* was transformed with plasmids encoding the protein fusions with and without shell proteins and the resulting strains were cultured for 24 h. Samples were taken and analysed by SDS-PAGE and subject to western blotting using an anti-GFP primary antibody. The results show that the co-expression of GFP-SsrA fused to targeting peptides, when produced with shell proteins, have the highest levels of GFP (Fig. 4; lanes 7+8). In the absence of a targeting peptide GFP is effectively degraded as observed by the presence of only a faint band present (Fig. 4; lane 6). In the absence of shell proteins all GFP fusion proteins are present to a much lesser extent than fusion proteins in the presence of microcompartments. These results are consistent with the theory that BMCs provide protection for internalised proteins.

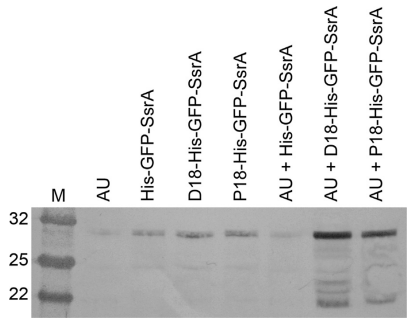


Fig. 4. Protease protection assay of GFP fused to a C-terminal proteolysis tag and an N-terminal targeting peptide in the presence and absence of BMCs. Total lysates were analysed by SDS-PAGE and subsequently western blotted with an anti-GFP primary antibody. Cell densities were normalised to an $OD_{600}=2.5$ for loading of samples. The faint bands seen in lanes 2-6 may, in part, be a result of unspecific binding, as a faint band is also seen in lane 1 (shell proteins only, no GFP).

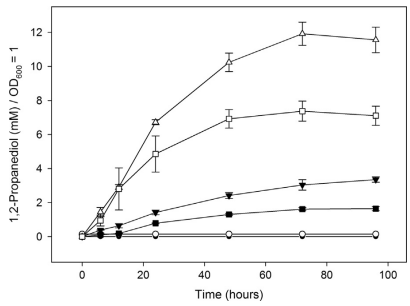


Fig. 5. In vivo 1,2-propanediol production. The graph shows the 1,2-propanediol content (normalised to $OD_{600}=1$) over 96 h in the growth medium of strains that lack shell proteins and 1,2-propanediol producing enzymes (control strain) (●), shell proteins only (control strain) (○), untagged 1,2-propanediol producing enzymes (▾), 1,2-propanediol producing enzymes tagged with targeting peptides (▵),untagged 1,2-propanediol producing enzymes and shell proteins (■),1,2-propanediol producing enzymes tagged with targeting peptides and shell proteins (□). Data points represent an average of three independent experiments; standard deviations are represented by error bars.

3.4. Construction of 1,2-propanediol producing strains

The untagged genes for the four pathway enzymes were cloned consecutively using the Link and Lock procedure to give pML5 (containing *gldA, mgsA, dhaK, fucO*) whereas the tagged versions were cloned in a similar fashion to give pML6 (containing *P18-gldA, D18-mgsA, P18-dhaK, D18-fucO*). Both plasmids were transformed into *E. coli* strain BL21*(DE3). The strains were further engineered to co-express the 1,2-propanediol production plasmids by transforming them with a compatible plasmid housing the shell protein genes (pLysS-*pduABJKNU*). Additionally, a control strain containing empty versions of pET14b and pLysS was generated as was a strain transformed with pET14b and pLysS-*pduABJKNU*. All strains were compared for the production of 1,2-propanediol.

The strains were grown in 100 mL cultures at 28 °C and samples were collected at 0, 6, 12, 24, 48, 72 and 96 h. The resulting growth curves (Supplementary Fig. S7) show that strains encoding proteins with targeting peptides (either in the presence of absence of shell proteins) grow slower and reach a lower final optical density in comparison to strains expressing un-tagged proteins and control strains. When analysing the protein profiles of whole cell samples at the various time points we found marginally higher production levels of some tagged proteins compared to the non-tagged proteins (Supplementary Fig. S8 c, d, e and f). Increased protein production and protein aggregation may lead to cells undergoing senescence mediated by asymmetric segregation of protein aggregates (Baig et al., 2014), which potentially explains why these strains did not grow as well as strains producing the same proteins without targeting peptides. Whole cell samples of strains housing the shell protein construct (pLysS-*pduABJKNU*) showed the protein profile expected for the shell components as observed by SDS-PAGE (Supplementary Fig. S8 b, e and f). The strain encoding only empty microcompartments appears to have a slightly higher production of shell proteins as evidenced by increased band intensity on the SDS-PAGE gels (Supplementary Fig. S8 b).

3.5. In vivo 1,2-propanediol production is elevated in strains producing enzymes with targeting sequences

The 1,2-propanediol content in the growth media of the various strains was quantified by GC–MS. Samples were collected at 0, 6, 12, 24, 48, 72 and 96 h and, following centrifugation, the resultant supernatant was analysed by GC–MS as described in materials and methods. The measured 1,2-propanediol content, as shown in Fig. 5, is expressed for a cell density of $OD_{600}=1$ (for the non-adjusted data see Supplementary Fig. S9). Strains encoding the untagged enzymes, His-GldA, His-DhaK, His-MgsA and His-FucO, whether in the presence of absence of BMCs, were found to produce low levels of 1,2-propanediol despite growing well and reaching the highest cell densities after 96 h.

Intriguingly, the highest levels of 1,2-propanediol were detected in the growth media of strains producing enzymes containing the fused targeting peptides. Thus strains producing P18-GldA, P18-DhaK, D18-MgsA and D18-FucO, whether in the presence or absence of BMCs, grew to a lower cell density than the strains harbouring un-tagged proteins, and despite the negative effect the targeting peptide has on the specific activities of the individual enzymes they produced significantly more 1,2-propanediol (Fig. 2, Fig. 5, and Supplementary Fig. S9) than the strains lacking the targeting peptides. The highest final yield of 1,2-propanediol was 11.56 mM/OD_{600} unit, which was observed when the shell proteins were not present. No 1,2-propanediol was detected in control stains (wild type *E. coli* and a strain producing shell proteins only).

3.6. TEM analysis of 1,2-propanediol producing strains

The higher product yield exhibited by the strain producing tagged enzymes in the absence of shell proteins was unexpected. To investigate if aggregation of the enzymes was causing this effect electron microscopy and immuno-gold labelling were used to visualise the subcellular organisation and location of the recombinant proteins in the various strains after the cells had been thin sectioned. To achieve this the thin sections of the strains were labelled with anti-histidine primary antibody, which is designed to bind to the hexa-histidine tag on the N-terminus of the various proteins. A secondary antibody conjugated to 10 nm gold particles was used to bind to the primary antibody, thereby revealing the intracellular location of 1,2-propanediol producing enzymes.

Control stains housing either empty vectors (pET14b + pLysS) or producing only the BMCs (pLysS-*pduABJKNU*) showed a small amount of antibody binding around the membranes of the cells (Fig. 6A and B); this is likely due to nonspecific binding. Aggregates, or large protein inclusions, are visible in approximately 100%

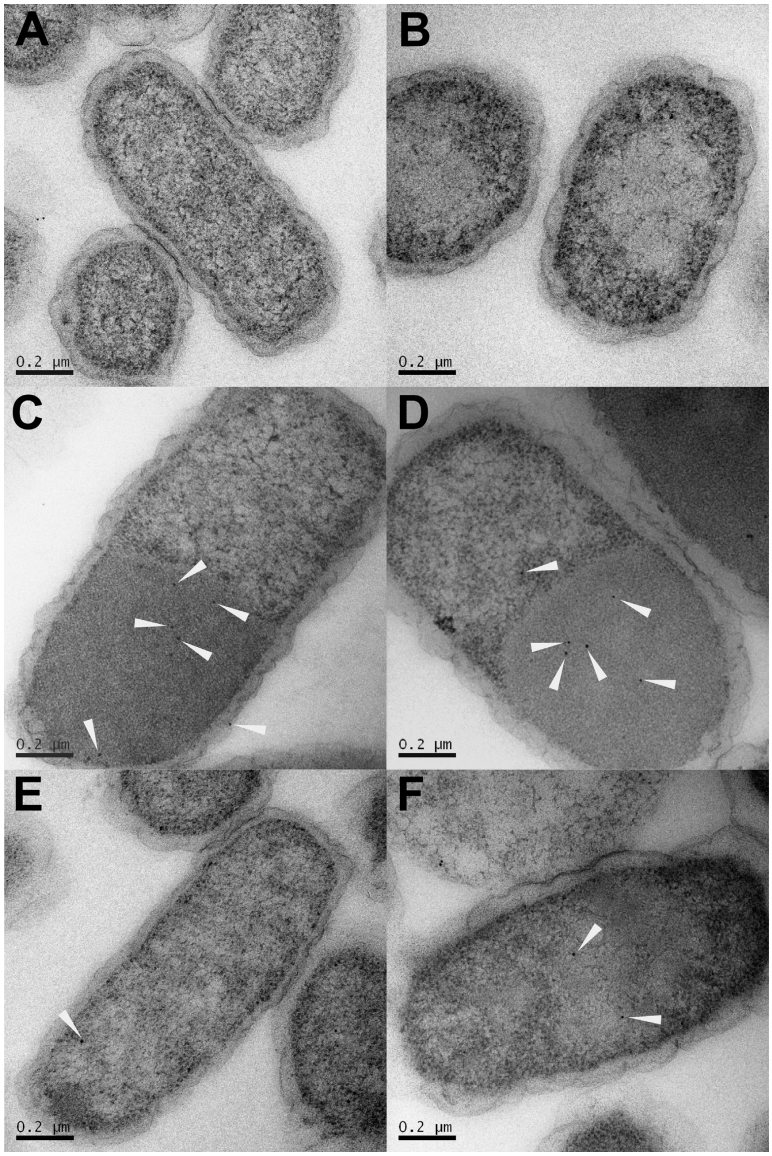


Fig. 6. Thin sections of *E. coli* strains labelled with an anti-his antibody and then with a secondary antibody conjugated to 10 nm gold particles viewed under TEM (A) strain that lacks shell proteins and 1,2-propanediol producing enzymes (control strain) (B) strain producing shell proteins only (control strain) (C) 1,2-propanediol producing enzymes tagged with targeting peptides (D) 1,2-propanediol producing enzymes tagged with targeting peptides and shell proteins (E) untagged 1,2-propanediol producing enzymes (F) untagged 1,2-propanediol producing enzymes and shell proteins. White pointers indicate gold particles.

of observed cells expressing the P18/D18-tagged proteins, regardless of whether BMCs are present. It is in these aggregate areas that the vast majority of antibody binding occurs (Fig. 6C and D). Such protein inclusions are not seen in cells expressing un-tagged enzymes (Fig. 6E and F), suggesting that it is the presence of the targeting peptides that facilitates the aggregation of proteins. In strains producing shell proteins, BMCs can be clearly seen grouped together within the cytoplasm of the cells when viewed by TEM (Supplementary Fig. S10). In the micrographs shown here the BMCs have low contrast because of the nature of sample preparation for immunolabelling (Fig. 6B, D and F). Only very few gold particles can be seen in the strain producing tagged enzymes and microcompartments in the region of the BMCs.

The combined data from both the 1,2-propanediol production studies and the TEM analysis of thin sections of bacteria from the various strains all suggest that maximum 1,2-propanediol production is due to the formation of a large active enzyme-inclusion body within the cell. The presence of the D18 and P18 fusion tags on the four enzymes (GldA, DhaK, MgsA and FucO) appears to result in the formation of a large protein aggregate when the four-tagged enzymes are co-produced. This large aggregate forms both in the presence and absence of BMCs.

4. Conclusions

Previously, we had shown that targeting of a two-enzyme heterologous pathway for ethanol production to engineered BMCs significantly increased the in vivo product yield in *E. coli*. Here, we have chosen to expand and apply our earlier findings to a pathway with industrial significance by engineering recombinant BMCs to house a four-enzyme pathway for the production of 1,2-propanediol from glycerol (GldA, DhaK, MgsA and FucO) in *E. coli*. Glycerol is a cheap and widely available source for the production of 1,2-propanediol, a valuable commodity chemical that is currently obtained from fossil fuel derivatives.

The first stage of the investigation into 1,2-propanediol synthesis in BMCs was to determine the effect of the BMC-targeting peptides, D18 and P18, on the pathway enzymes. The in vitro data presented in this study has revealed that the fusion of targeting peptides to the individual pathway enzymes has a variable effect on the properties of each protein. Differences can be seen in the levels of the protein, stability and in enzymatic activity. Three of the four tagged enzymes (GldA, MgsA and FucO) had lowered specific activities, caused by the addition of the 18 amino acid targeting sequences onto the N-terminus of the proteins. Moreover, the presence of the different peptide fusions, D18 and P18, had altered effects on the activity of the same enzyme. The solubility of each protein was also affected by the attachment of the peptide fusion, with GldA forming large inclusion bodies in the majority of cells observed by TEM when fused with a targeting peptide compared to un-tagged GldA.

To demonstrate that the D18 or P18 peptide was able to localise, individually, the 1,2-propanediol enzymes to the BMC, the four tagged enzymes were separately co-produced with the 'empty' Pdu BMC. In each case the tagged enzyme was found to co-purify with the BMC, suggesting that that the enzyme had been correctly targeted to the BMC. This in itself does not prove that the tagged enzyme has been encased within the recombinant BMC, merely that it associates with the BMC. Evidence that the D18 or P18 tagged proteins are internalised was provided by a protease protection assay using GFP as the cargo for the BMC. The presence of either the P18 or D18 peptide on the GFP protected the GFP from proteolysis but only when the tagged protein was co-produced with BMCs.

In an attempt to target the whole 1,2-propanediol pathway to a BMC, the genes encoding the most active of the individually tagged enzymes (P18-GldA, P18-DhaK, D18-MgsA and D18-FucO) were cloned consecutively on a single plasmid so that the whole pathway would be targeted to empty BMCs. In vivo analysis of 1,2-propanediol production of the strain showed that fusion of targeting peptides to all of the proteins involved in the synthesis of 1,2-propanediol resulted in an increase in product formation.

Rather unexpectedly, the presence of the microcompartment shell was not required for the increased product formation. In fact, the strain generating the most 1,2-propanediol produced tagged 1,2-propanediol pathway enzymes, but no shell proteins. This strain showed an increase in product formation of 245% OD-adjusted and 157% not OD-adjusted in comparison to the strain producing un-tagged enzymes; despite the lower in vitro activity of the individual tagged proteins compared to the un-tagged proteins. TEM analysis showed that co-production of all four proteins resulted in protein aggregation and deposition at the poles of nearly all cells and it is this aggregation that appears to provide a significant benefit to the efficiency of the pathway. Aggregation of our proteins of interest may be due to coiled coil interactions facilitated by the targeting peptides as previously shown for the P18 sequence, which appears to form a coiled coil dimer in solution (Lawrence et al., 2014). Pathway productivity could be further enhanced by introducing a cofactor-recycling enzyme; this has been previously implemented to enhance a related pathway for dihydroxyacetone production (Zhou et al., 2013). Through the use of a tightly regulated fermentation system it would reasonably be expected that pathway productivity could be further enhanced.

Hence, it could be assumed that the increased product yield is a result of concentrating enzymes into aggregates. As such this scaffold might result in increased channelling of substrates and products between proteins due to proximity effects. The formation of this aggregate may be a mimic of what happens in nature during BMC formation. For instance, in the case of the carboxysome, the cargo enzymes initially condense and are then encased by the shell of the BMC. The reason inclusion bodies are observed in this study even when shell proteins are co-produced is that the expression of the pathway genes is under the control of a strong T7 promoter which leads to disproportional ratios of cargo enzyme to shell proteins. Surplus protein, which is not encapsulated, is aggregated. In wild type operons gene expression and thus protein levels are controlled by regulatory elements, promoters and ribosome binding sites of various strengths.

In summary, the work described in this report demonstrates that the presence of short targeting peptides can not only convert individual fusion proteins but also whole pathways into active aggregates that allow for increased product yield in vivo. The aggregations of multiple enzymes allows for increased localised concentrations of enzymes and intermediates within the cell resulting in a higher product yield. In essence, these metabolic aggregates that are formed by the attachment of BMC-targeting peptides represent BMCs without shells. Hence, engineered protein aggregation for instance through the design of coiled coil interactions may hold an important role in the future of metabolic engineering for the production of commodity products. Active protein aggregates may have prolonged extracellular activity and therefore be more robust than BMCs. The use of BMCs for pathway localisation still remains desirable in the field of metabolic engineering. Pathways involved in the microbial production of biofuels often contain toxic intermediates and therefore compartmentalisation may reduce the negative effect of these toxic intermediates on the cell and thus increase fitness and productivity. Here we have shown that the formation of recombinant BMCs requires the

coordinated production of protein cargo in a controlled manner for the construction of bioreactors of predictable functionality.

Acknowledgement

We recognise the help of Usha Dura, Mirabel Inyang, Mathilda Oxandaburu and Tulshi Patel for the initial cloning of glycerol dehydrogenase, methylglyoxal synthase and 1,2-propanediol oxidoreductase. We thank Kevin Howland for assistance with GC–MS. Financial support was provided by the Biotechnology and Biological Sciences Research Council (BB/M002969/1) and the Leverhulme Trust (ECF-213-341).

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.ymben.2016.02.007.

References

Adkesson, D.M., Disney, J.S., Dravis, B.C., Gaddy, J.M., Lehnhardt, W.F., Lievense, J.C., Wenndt, G.M., Ames, T.T., Fitzgibbon, P., Gallagher, F.G., Luyben, M.L., Seapan, M., Trotter, R.E., Alsop, A.W., Yu, E.K., Chu, L.A., 2011. Purification Of Biologically-Produced 1, 3-Propanediol. U.S. Patent no. 7,919,658.

Altaras, N.E., Cameron, D.C., 1999. Metabolic engineering of a 1,2-propanediol pathway in *Escherichia coli*. Appl. Environ. Microbiol. 65, 1180–1185.

Altaras, N.E., Cameron, D.C., 2000. Enhanced production of (R)-1,2-propanediol by metabolically engineered *Escherichia coli*. Biotechnol. Prog. 16, 940–946.

Axen, S.D., Erbilgin, O., Kerfeld, C.A., 2014. A taxonomy of bacterial microcompartment loci constructed by a novel scoring method. PLoS Comput. Biol. 10, e1003898.

Baig, U.I., Bhadbhade, B.J., Mariyam, D., Watve, M.G., 2014. Protein aggregation in *E. coli*: short term and long term effects of nutrient density. PLoS One 9 (9), e107445.

Brinsmade, S.R., Paldon, T., 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.

Burk, M.J., Van Dien, S.J., Burgard, A.P., Niu, W., Genomatica inc., 2011. Compositions and methods for the biosynthesis of 1, 4-butanediol and its precursors. U.S. Patent no. 8,067,214.

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

Cheng, S., Liu, Y., Crowley, C.S., Yeates, T.O., Bobik, T.A., 2008. Bacterial microcompartments: their properties and paradoxes. Bioessays 30, 1084–1095.

Chowdhury, C., Sinha, S., Chun, S., Yeates, T.O., Bobik, T.A., 2014. Diverse bacterial microcompartment organelles. Microbiol. Mol. Biol. Rev. 78, 438–468.

Clomburg, J.M., Gonzalez, R., 2011. Metabolic engineering of *Escherichia coli* for the production of 1,2-propanediol from glycerol. Biotechnol. Bioeng. 108, 867–879.

Conrado, R.J., Wu, G.C., Boock, J.T., Xu, H., Chen, S.Y., Lebar, T., Turnšek, J., Tomšič, N., Avbelj, M., Gaber, R., Koprivnjak, T., Mori, J., Glavnik, V., Vovk, I., Benčina, M., Hodnik, V., Anderluh, G., Dueber, J.E., Jerala, R., DeLisa, M.P., 2012. DNA-guided assembly of biosynthetic pathways promotes improved catalytic efficiency. Nucl. Acids Res. 40, 1879–1889.

Dou, Z., Heinhorst, S., Williams, E.B., Murin, C.D., Shively, J.M., Cannon, G.C., 2008. CO_2 fixation kinetics of *Halothiobacillus neapolitanus* mutant carboxysomes lacking carbonic anhydrase suggest the shell acts as a diffusional barrier for CO_2. J. Biol. Chem. 283, 10377–10384.

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

Fan, C., Cheng, S., Liu, Y., Escobar, C.M., Crowley, C.S., Jefferson, R.E., Yeates, T.O., Bobik, T.A., 2010. Short N-terminal sequences package proteins into bacterial microcompartments. Proc. Natl. Acad. Sci. USA 107, 7509–7514.

Farrell, C.M., Grossman, A.D., Sauer, R.T., 2005. Cytoplasmic degradation of ssrA-tagged proteins. Mol. Microbiol. 57, 1750–1761.

Ferguson, G.P., Chacko, A.D., Lee, C.H., Booth, I.R., 1996. The activity of the high-affinity K+ uptake system Kdp sensitizes cells of *Escherichia coli* to methylglyoxal. J. Bacteriol. 178, 3957–3961.

Frank, S., Lawrence, A.D., Prentice, M.B., Warren, M.J., 2013. Bacterial microcompartments moving into a synthetic biological world. J. Biotechnol. 163, 273–279.

Havemann, G.D., Sampson, E.M., Bobik, T.A., 2002. PduA is a shell protein of polyhedral organelles involved in coenzyme B(12)-dependent degradation of 1,2-propanediol in *Salmonella enterica* serovar typhimurium LT2. J. Bacteriol. 184, 1253–1261.

Jung, Y.K., Kim, T.Y., Park, S.J., Lee, S.Y., 2010. Metabolic engineering of *Escherichia coli* for the production of polylactic acid and its copolymers. Biotechnol. Bioeng. 105, 161–171.

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., Warren, M.J., 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., Dueber, J.E., 2012. Spatial organization of enzymes for metabolic engineering. Metab. Eng. 14, 242–251.

Lee, S.J., Lee, D.Y., Kim, T.Y., Kim, B.H., Lee, J., Lee, S.Y., 2005. Metabolic engineering of *Escherichia coli* for enhanced production of succinic acid, based on genome comparison and in silico gene knockout simulation. Appl. Environ. Microbiol. 71, 7880–7887.

Marchetti, J.M., Miguel, V.U., Errazu, A.F., 2007. Possible methods for biodiesel production. Renew. Sustain. Energy Rev. 11, 1300–1311.

McGoldrick, H.M., Roessner, C.A., Raux, E., Lawrence, A.D., McLean, K.J., Munro, A.W., Santabarbara, S., Rigby, S.E., Heathcote, P., Scott, A.I., Warren, M.J., 2005. Identification and characterization of a novel vitamin B12 (cobalamin) biosynthetic enzyme (CobZ) from *Rhodobacter capsulatus*, containing flavin, heme, and Fe-S cofactors. J. Biol. Chem. 14, 1086–1094.

Neidhardt, F.C., Bloch, P.L., Smith, D.F., 1974. Culture medium for enterobacteria. J. Bacteriol. 119, 736–747.

Parsons, J.B., Frank, S., Bhella, D., Liang, M., Prentice, M.B., Mulvihill, D.P., 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., Roth, J.R., 2006. Conserving a volatile metabolite: a role for carboxysome-like organelles in *Salmonella enterica*. J. Bacteriol. 188, 2865–2874.

Peralta-Yahyan, P.P., Ouellet, M., Chan, R., Mukhopadhyay, A., Keasling, J.D., Lee, T.S., 2011. Identification and microbial production of a terpene-based advanced biofuel. Nat. Commun. 2, 483.

Peralta-Yahya, P.P., Zhang, F., del Cardayre, S.B., Keasling, J.D., 2012. Microbial engineering for the production of advanced biofuels. Nature 488, 320–328.

Sampson, E.M., 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.

Sargent, F., Davidson, F.A., Kelly, C.L., Binny, R., Christodoulides, N., Gibson, D., Johansson, E., Kozyrska, K., Lado, L.L., Maccallum, J., Montague, R., Ortmann, B., Owen, R., Coulthurst, S.J., Dupuy, L., Prescott, A.R., Palmer, T., 2013. A synthetic system for expression of components of a bacterial microcompartment. Microbiology 159, 2427–2436.

Shibata, N., Tamagaki, H., Hieda, N., Akita, K., Komori, H., Shomura, Y., Teriyaki, S., Mori, K., Yasuoka, N., Higuchi, Y., Toraya, T., 2010. Crystal structures of ethanolamine ammonia-lyase complexed with coenzyme B12 analogs and substrates. J Biol Chem. 285, 26484–26493.

Steiger, M.G., Blumhoff, M.L., Mattanovich, D., Sauer, M., 2013. Biochemistry of microbial itaconic acid production. Front. Microbiol. 4, 23.

Suriyamongkol, P., Weselake, R., Narine, S., Moloney, M., Shah, S., 2007. Biotechnological approaches for the production of polyhydroxyalkanoates in microorganisms and plants - a review. Biotechnol. Adv. 25, 148–175.

Tanaka, S., Kerfeld, C.A., Sawaya, M.R., Cai, F., Heinhorst, S., Cannon, G.C., Yeates, T.O., 2008. Atomic-level models of the bacterial carboxysome shell. Science 319, 1083–1086.

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

Yeates, T.O., Kerfeld, C.A., Heinhorst, S., Cannon, G.C., Shively, J.M., 2008. Protein-based organelles in bacteria: carboxysomes and related microcompartments. Nat. Rev. Microbiol. 6, 681–691.

Yazdani, S.S., Gonzalez, R., 2007. Anaerobic fermentation of glycerol: a path to economic viability for the biofuels industry. Curr. Opin. Biotechnol. 18, 213–219.

Yim, H., Haselbeck, R., Niu, W., Pujol-Baxley, C., Burgard, A., Boldt, J., Khandurina, J., Trawick, J.D., Osterhout, R.E., Stephen, R., Estadilla, J., Teisan, S., Schreyer, H.B., Andrae, S., Yang, T.H., Lee, S.Y., Burk, M.J., Van Dien, S., 2011. Metabolic engineering of *Escherichia coli* for direct production of 1,4-butanediol. Nat. Chem. Biol. 7, 445–452.

Zhou, Y.J., Yang, W., Wang, L., Zhu, Z., Zhang, S., Zhao, Z.K., 2013. Engineering NAD^+ availability for *Escherichia coli* whole-cell biocatalysis: a case study for dihydroxyacetone production. Microb. Cell. Fact. 12, 103.

Supplementary information for:

Employing bacterial microcompartment technology to engineer a shell-free enzyme-aggregate for enhanced 1,2-propanediol production in *Escherichia coli*

Matthew J. Lee[1], Ian R. Brown[1], Rokas Juodeikis[1], Stefanie Frank[1] and Martin J. Warren[1],

[1] School of Biosciences, University of Kent, Giles Lane, Canterbury, Kent, CT2 7NJ, U.K.

Corresponding authors:

Martin J. Warren,

Phone: 00 44 (0)1227 824690. E-mail: m.j.warren@kent.ac.uk.

School of Biosciences, University of Kent, Giles Lane, Canterbury, Kent, CT2 7NJ, U.K.

Stefanie Frank

Phone: 00 44 (0)1227 824692. E-mail: S.Frank@kent.ac.uk

School of Biosciences, University of Kent, Giles Lane, Canterbury, Kent, CT2 7NJ, U.K.

Supplementary Table 1: Strains used in this study

Strain	Genotype	Source
JM109	endA1, recA1, gyrA96, thi, hsdR17 (rk–, mk+), relA1, supE44, Δ(lac-proAB), [F′, traD36, proAB, laqIqZΔM15]	Promega
BL21 (DE3)	F– ompT hsdSB (rB– mB–) gal dcm (DE3)	Novagen
BL21 (DE3) pLysS	F- ompT hsdSB(rB- mB-) gal dcm (DE3) pLysS (CamR)	Novagen

Supplementary Table 2: Plasmids used in this study

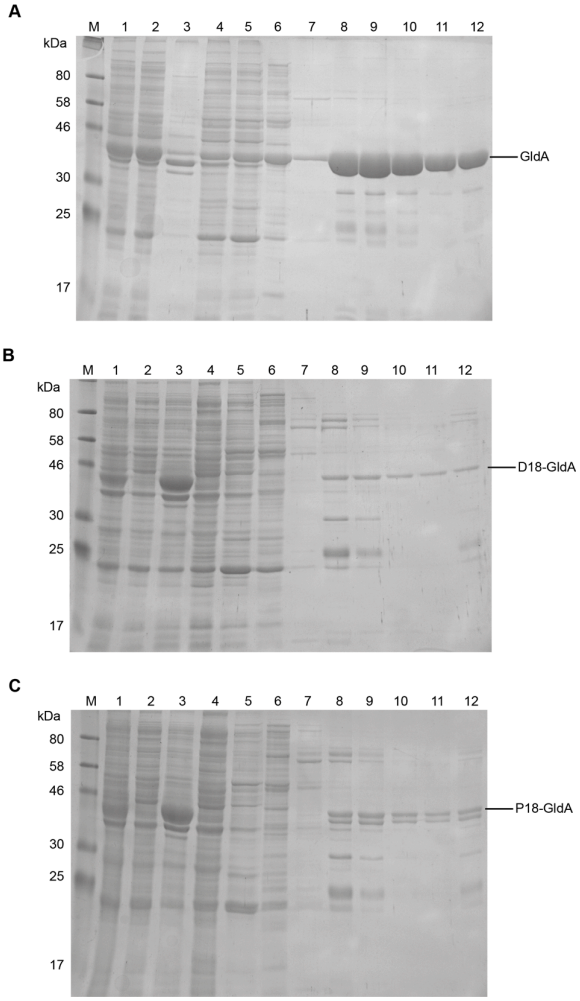
Plasmid name	Description	Source
pET14b	Overexpression vector containing N-terminal hexahistidine-tag, modified to include an *Spe*I site 5' of *Bam*HI	Novagen
pET14b-D18	Overexpression vector containing an N-terminal D18 targeting tag followed by a short amino acid linker (AMGSS) then a hexahistidine-tag	This study
pET14b-P18	Overexpression vector containing an N-terminal P18 targeting tag followed by a short amino acid linker (PMGSS) then a hexahistidine-tag	This study
pLysS	Basal expression suppressor	Novagen
pLysS-2	Basal expression suppressor containing the T7 promoter-MCS-T7 terminator cassette from pET14b	Parsons *et al*., 2008
pLysS-PduABJKNU	pLysS-2 containing genes required for the formation of empty BMCs	Parsons *et al*., 2010
pET14b-gldA	PCR product of gldA ligated into *Nde*I/*Spe*I sites of pET14b	This study
pET14b-dhaK	PCR product of dhaK ligated into *Nde*I/*Spe*I sites of pET14b	This study
pET14b-mgsA	PCR product of mgsA ligated into *Nde*I/*Spe*I sites of pET14b	This study

pET14b-fucO	PCR product of fucO ligated into *Nde*I/*Spe*I sites of pET14b	This study
pET14b-GFP-SsrA	PCR product of gfp-ssrA ligated into *Nde*I/*Spe*I sites of pET14b	This study
pET14b-D18-gldA	*Nde*I/*Spe*I fragment of pET14b-gldA ligated into *Nde*I/*Spe*I sites of pET14b-D18	This study
pET14b-D18-dhaK	*Nde*I/*Spe*I fragment of pET14b-dhaK ligated into *Nde*I/*Spe*I sites of pET14b-D18	This study
pET14b-D18-mgsA	*Nde*I/*Spe*I fragment of pET14b-mgsA ligated into *Nde*I/*Spe*I sites of pET14b-D18	This study
pET14b-D18-fucO	*Nde*I/*Spe*I fragment of pET14b-fucO ligated into *Nde*I/*Spe*I sites of pET14b-D18	This study
pET14b-D18-GFP-SsrA	*Nde*I/*Spe*I fragment of pET14b-GFP-SsrA ligated into *Nde*I/*Spe*I sites of pET14b-D18	This study
pET14b-P18-gldA	*Nde*I/*Spe*I fragment of pET14b-gldA ligated into *Nde*I/*Spe*I sites of pET14b-P18	This study
pET14b-P18-dhaK	*Nde*I/*Spe*I fragment of pET14b-dhaK ligated into *Nde*I/*Spe*I sites of pET14b-P18	This study
pET14b-P18-mgsA	*Nde*I/*Spe*I fragment of pET14b-mgsA ligated into *Nde*I/*Spe*I sites of pET14b-P18	This study
pET14b-P18-fucO	*Nde*I/*Spe*I fragment of pET14b-fucO ligated into *Nde*I/*Spe*I sites of pET14b-P18	This study
pET14b-P18-GFP-SsrA	*Nde*I/*Spe*I fragment of pET14b-GFP-SsrA ligated into *Nde*I/*Spe*I sites of pET14b-P18	This study
pML-1	*Xba*I/*Eco*RI fragment from pET14b-fucO ligated into *Xba*I/*Eco*RI sites of pET14b-gldA	This study
pML-2	*Xba*I/*Eco*RI fragment from pET14b-D18-fucO ligated into *Spe*I/*Eco*RI sites of pET14b-P18-gldA	This study
pML-3	*Xba*I/*Hind*III fragment from pET14b-mgsA ligated into *Spe*I/*Hind*III sites of pET14b-dhaK	This study
pML-4	*Xba*I/*Hind*III fragment from pET14b-D18-mgsA ligated into *Spe*I/*Hind*III sites of pET14b-P18-dhaK	This study

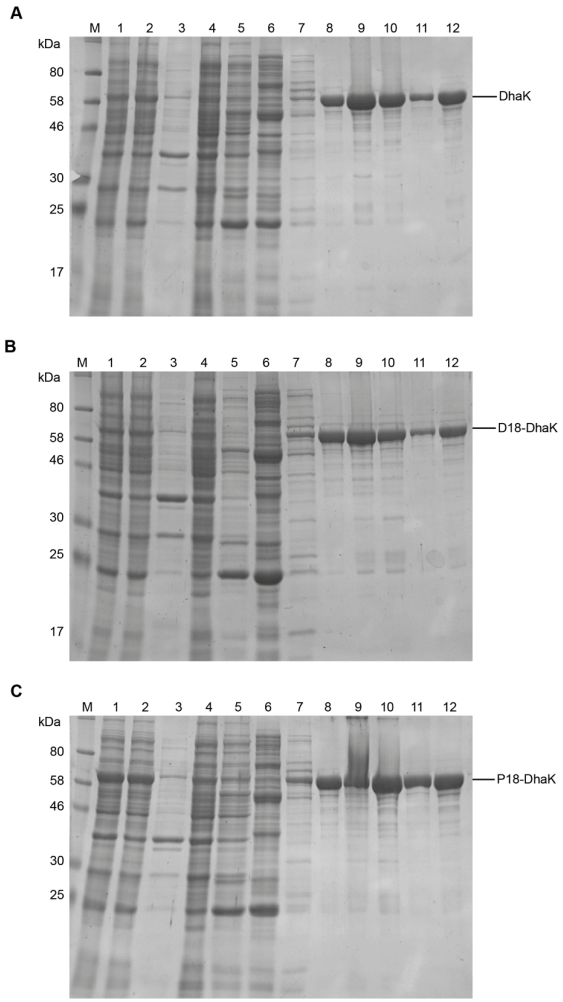
pML-5	*Xba*I/*Cla*I fragment from pML-3 ligated into *Spe*I/*Cla*I sites of pML-1	This study
pML-6	*Xba*I/*Cla*I fragment from pML-4 ligated into SpeI/ClaI sites of pML-2	This study

Supplementary Table 3: Oligonucleotides used in this study, restriction sites are underlined

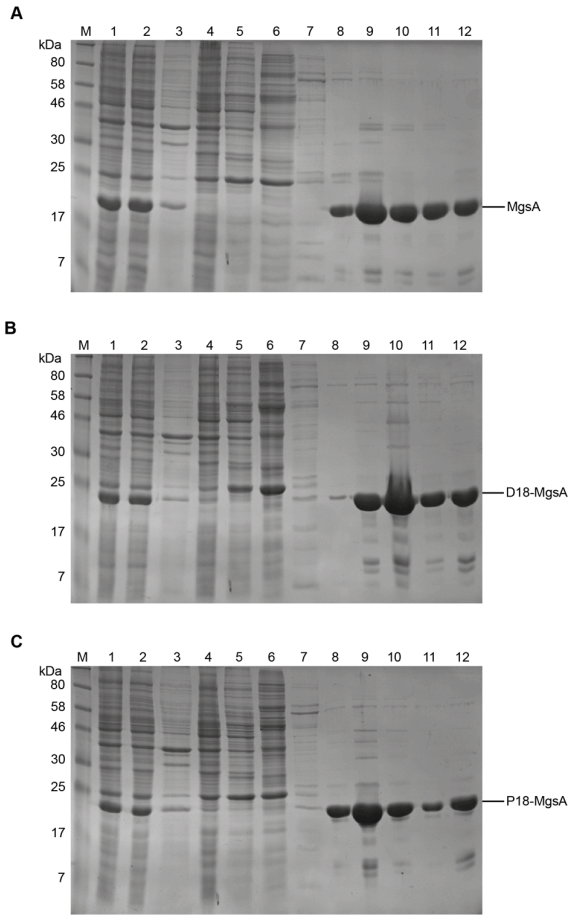
Name	Sequence 5' – 3'
GldA_NdeI_FW	CATCATATGGACCGCATTATTCAATCACC
GldA_SpeI_RV	CATACTAGTTTATTCCCACTCTTGCAGG
dhaK_NdeI_FW	CGCCATATGTCTCAATTCTTTTTTAACCAACGCACC
dhaK_SpeI_RV	CATACTAGTTTAGCCCAGCTCACTCTCCGC
mgsA_NdeI_FW	CATCATATGGAACTGACGACTCGCACTTTACC
mgsA_SpeI_RV	CATACTAGTTTACTTCAGACGGTCCGCGAG
fucO_NdeI_FW	CCGCATATGGCTAACAGAATGATTCTG
fucO_SpeI_RV	CCTACTAGTTTACCAGGCGGTATGG
GFP_NdeI_FW	GTACATATGAGCAAAGGAGAAGAACTTTTC
GFP-SsrA_SpeI_RV	GACACTAGTTTAAGCTGCTAAAGCGTAGTTTTCGTCGTT TGCTGCTTTGTACAGCTCATCCATGCC



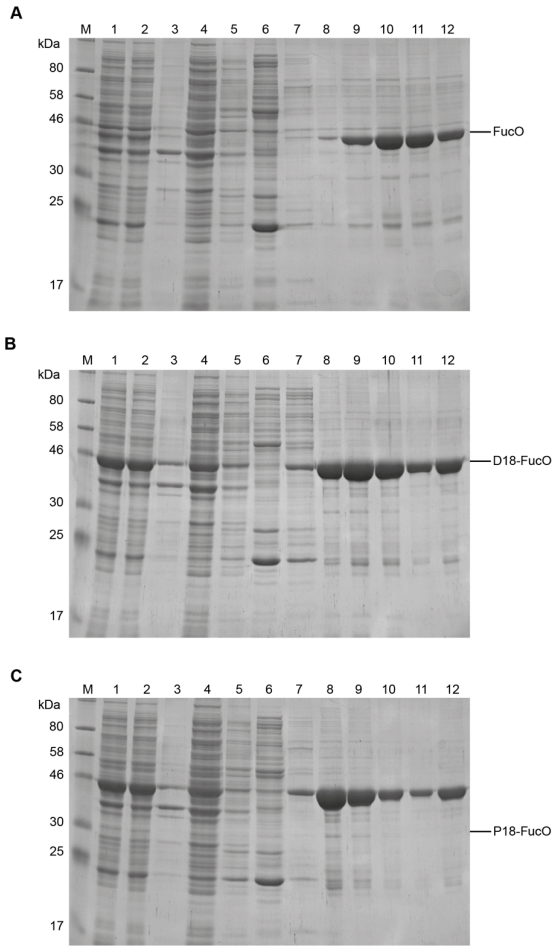
Supplementary Figure S1: SDS-PAGE of GldA purified by IMAC (a) GldA (b) D18-GldA (c) P18-GldA; In comparison to a molecular weight marker. Lane 1 – lysate (2 µl), lane 2 – supernatant after centrifugation (2 µl), lane 3 – pellet after centrifugation (10 µl), lane 4 – supernatant flow through (8 µl), lane 5 – binding buffer flow through (5 mM imidazole) (10 µl), lane 6 – wash buffer 1 flow through (50 mM imidazole) (10 µl), lane 7 – wash buffer 2 flow through (100 mM imidazole) (10 µl), lanes 8– 11 – elution fractions 4-7 (400 mM imidazole) (5 µl), lane 12 – sample after buffer exchange (5 µl).



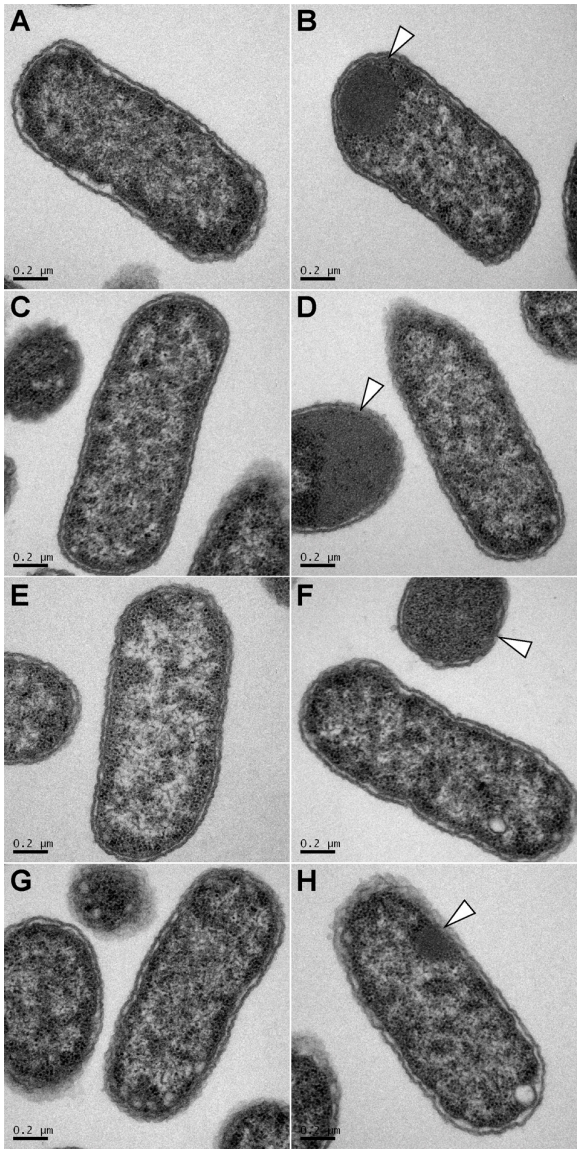
Supplementary Figure S2: SDS-PAGE of DhaK purified by IMAC (a) DhaK (b) D18-DhaK (c) P18-DhaK; In comparison to a molecular weight marker. Lane 1 – lysate (3 µl), lane 2 – supernatant after centrifugation (3 µl), lane 3 – pellet after centrifugation (10 µl), lane 4 – supernatant flow through (10 µl), lane 5 – binding buffer flow through (5 mM imidazole) (10 µl), lane 6 – wash buffer 1 flow through (50 mM imidazole) (10 µl), lane 7 – wash buffer 2 flow through (100 mM imidazole) (10 µl), lanes 8– 11 – elution fractions 2-5 (400 mM imidazole) (3 µl), lane 12 – sample after buffer exchange (3 µl).



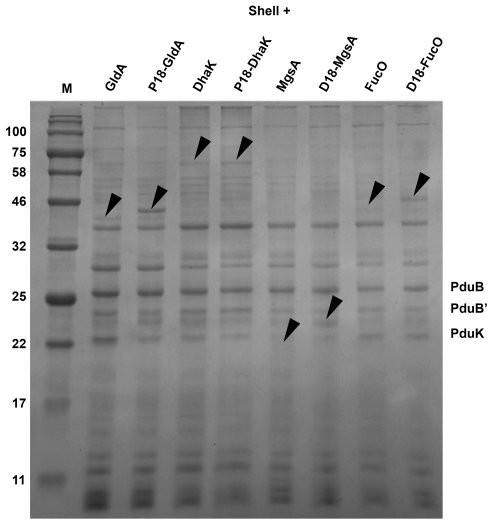
Supplementary Figure S3: SDS-PAGE of MgsA purified by IMAC. (a) MgsA (b) D18-MgsA (c) P18-MgsA; In comparison to a molecular weight marker. Lane 1 – lysate (2 µl), lane 2 – supernatant after centrifugation (2 µl), lane 3 – pellet after centrifugation (10 µl), lane 4 – supernatant flow through (3 µl), lane 5 – binding buffer flow through (5 mM imidazole) (10 µl), lane 6 – wash buffer 1 flow through (50 mM imidazole) (10 µl), lane 7 – wash buffer 2 flow through (50 mM imidazole) (10 µl), lanes 8– 11 – elution fractions 2-5 (400 mM imidazole) (2 µl), lane 12 – sample after buffer exchange (2 µl).



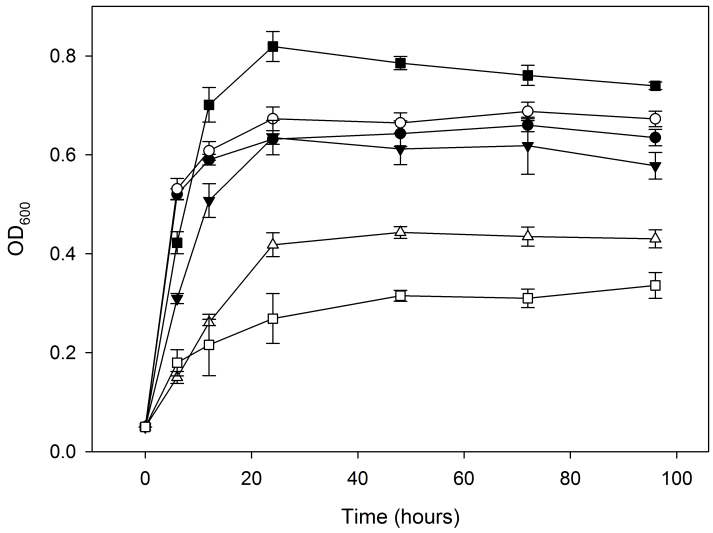
Supplementary Figure S4: SDS-PAGE of FucO purified by IMAC (a) FucO (b) D18-FucO (c) P18-FucO; In comparison to a molecular weight marker. Lane 1 – lysate (4 µl), lane 2 – supernatant after centrifugation (4 µl), lane 3 – pellet after centrifugation (10 µl), lane 4 – supernatant flow through (10 µl), lane 5 – binding buffer flow through (5 mM imidazole) (10 µl), lane 6 – wash buffer 1 flow through (50 mM imidazole) (10 µl), lane 7 – wash buffer 2 flow through (100 mM imidazole) (10 µl), lanes 8– 11 – elution fractions 3-6 (400 mM imidazole) (2 µl), lane 12 – sample after buffer exchange (2 µl).



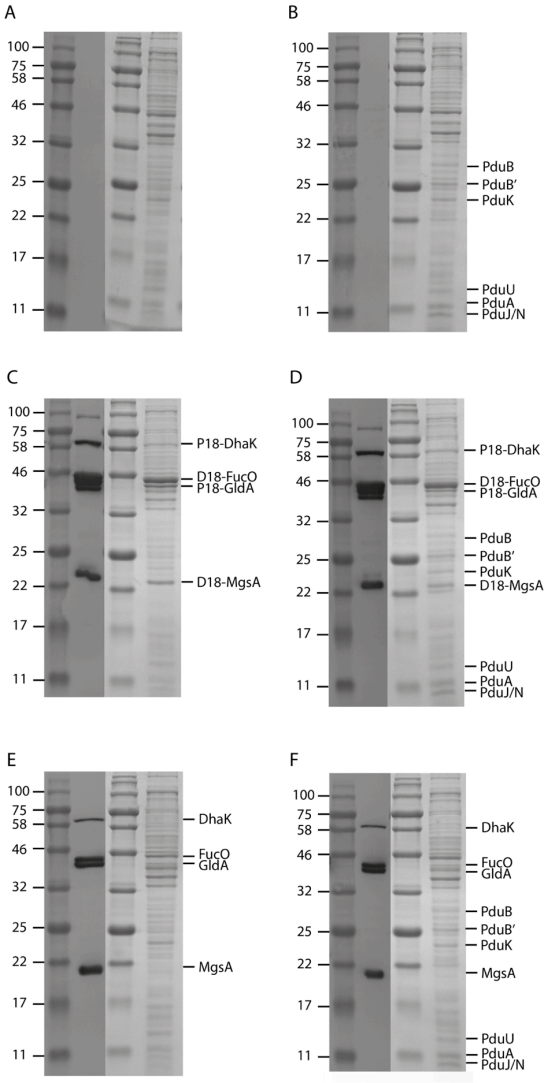
Supplementary Figure S5: TEM analysis of strains expressing (A) GldA (B) P18-GldA (C) DhaK (D) P18-DhaK (E) MgsA (F) D18-MgsA (G) FucO (H) D18-FucO. Arrows indicate protein aggregates. Scale bar shows 0.2 µm



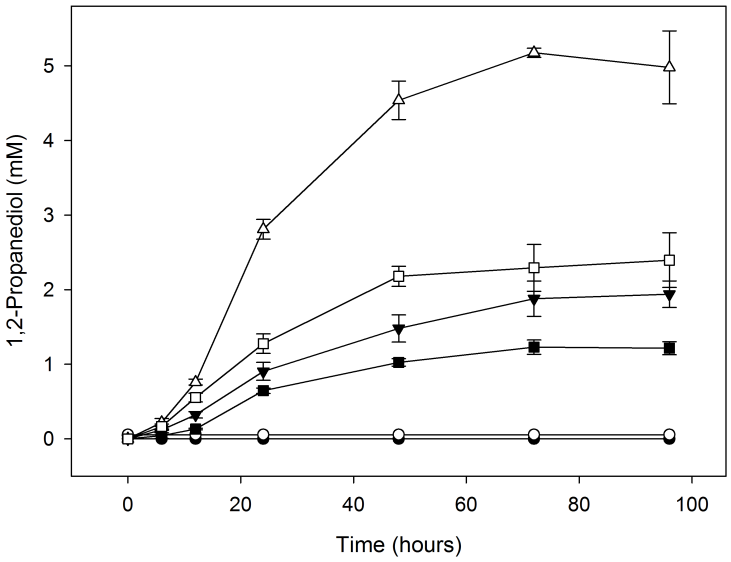
Supplementary Figure S6 SDS-PAGE analysis of purified BMCs co-produced with tagged or untagged GldA, DhaK, MgsA or FucO.



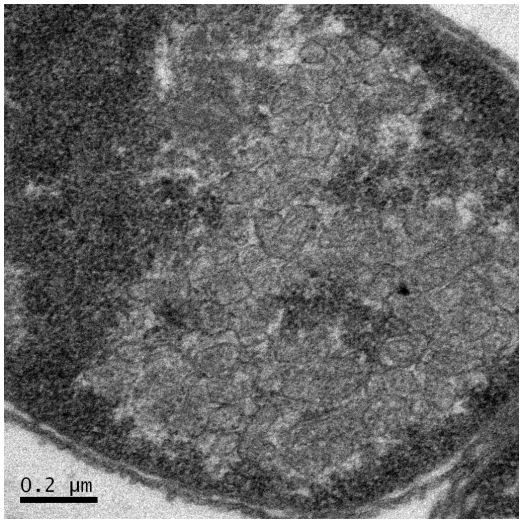
Supplementary Figure S7: Growth curves of strains producing 1,2-propanediol and control strains shown as OD_{600} against time (hours). *E. coli* strain that lacks shell proteins and 1,2-propanediol producing enzymes (control strain) (●), Shell proteins only (control strain) (○), untagged 1,2-propanediol producing enzymes (▼), 1,2-propanediol producing enzymes tagged with targeting peptides (△), untagged 1,2-propanediol producing enzymes and shell proteins (■),1,2-propanediol producing enzymes tagged with targeting peptides and shell proteins (□). Data points represent an average of three independent experiments; standard deviations are represented by error bars.



Supplementary Figure S8: SDS-PAGE and western blot analysis of final whole cell samples of (A) strain that lacks shell proteins and 1,2-propanediol producing enzymes (control strain) (B) shell proteins only (PduABB'JKNU) (control strain) (C) P18/D18-tagged-DhaK, MgsA, GldA and FucO (D) P18/D18-tagged-DhaK, MgsA, GldA and FucO + PduABB'JKNU (E) untagged-DhaK, MgsA, GldA and FucO (F) untagged-DhaK, MgsA, GldA and FucO + PduABB'JKNU adjusted to the same OD_{600} (2.5)



Supplementary Figure S9 *In vivo* 1,2-propanediol production. The graph shows the 1,2-propanediol content over 96 h in the growth medium in of strains that lack shell proteins and 1,2-propanediol producing enzymes (control) (●), shell proteins only (control strain) (○), untagged 1,2-propanediol producing enzymes (▼), 1,2-propanediol producing enzymes tagged with targeting peptides (△),untagged 1,2-propanediol producing enzymes and shell proteins (■),1,2-propanediol producing enzymes tagged with targeting peptides and shell proteins (□). Data points represent an average of three independent experiments; standard deviations are represented by error bars.



Supplementary Figure S10: TEM micrograph showing thin section of *E. coli* strain BL21*(DE3) transformed with pLysS-PduABJKNU.

Supplementary methods

Embedding of strains for TEM analysis

50 ml of LB was inoculated with one colony and grown at 37 °C with shaking to an OD600 of ~ 0.4, cells were harvested by centrifugation at 3000 x g for 10 minutes. The cell pellet was resuspended in 2 ml 2.5% Glutaraldehyde in 100 mM sodium cacodylate buffer pH 7.2 (CAB) and fixed for 2 hours with gentle rotating (20 rpm). Cells were pelleted by centrifugation at 6000 x g for 2 minutes and were washed twice for 10 minutes with 100 mM CAB. Cells were post-fixed with 1% osmium tetroxide in 100 mM CAB for 2 hours and subsequently washed twice with dH_2O. Cells were dehydrated by incubation in an ethanol gradient, 50% EtOH for 10 minutes, 70% EtOH overnight, 90% EtOH for 10 minutes followed by three 10 minute washes in 100% dry EtOH. Cells were then washed twice with propylene oxide for 15 minutes. Cell pellets were embedded by resuspension in 1 ml of a 1:1 mix of propylene oxide and Agar LV Resin and incubated for 30 minutes with rotation. Cell pellets were infiltrated twice in 100% Agar LV resin. The cell pellet was re-suspended in fresh resin and transferred to a 1ml Beem embedding capsule, centrifuged for 5 minutes at 3000 x g to concentrate the cells to the tip of the mould and incubated for 20 hours at 60 °C to polymerise.

Sectioning and visualisation of samples

Samples were ultra-thin sectioned on a RMC MT-XL ultra-microtome with a diamond knife (diatome 45°) sections (60 – 70 nm) were collected on un-coated 300 mesh copper grids. Grids were stained by incubation in 4.5% uranyl acetate in 1% acetic acid solution for 45 minutes followed by washing in a stream of dH_2O. Grids were then stained with Reynolds lead citrate for 7 minutes followed by washing in a stream of dH_2O

Electron microscopy was performed using a JEOL-1230 transmission electron microscope equipped with a Gatan multiscan digital camera operated at an accelerating voltage of 80 kV

Supplementary references

Parsons, J.B., Dinesh, S.D., Deery, E., Leech, H.K., Brindley, A.A., Heldt, D., Frank, S., Smales, C.M., Lünsdorf, H., Rambach, A., Gass, M.H., Bleloch, A., McClean, K.J., Munro, A.W., Rigby, S.E., Warren, M.J., Prentice, M.B. (2008) Biochemical and structural insights into bacterial organelle form and biogenesis. J Biol Chem. 283:14366-14375.

Parsons, J.B., Frank, S., Bhella, D., Liang, M., Prentice, M.B., Mulvihill, D.P., 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.

6.2. Patent application: Genetically modified microorganisms.

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
11 May 2017 (11.05.2017)

WIPO | PCT

(10) International Publication Number
WO 2017/077320 A1

(51) International Patent Classification:
C07K 14/195 (2006.01) *C12N 1/20* (2006.01)

(21) International Application Number:
PCT/GB2016/053435

(22) International Filing Date:
4 November 2016 (04.11.2016)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
1519554.8 5 November 2015 (05.11.2015) GB

(71) Applicant: **UNIVERSITY OF KENT** [GB/GB]; School of Briosciences, The Registry, Canterbury Kent CT2 7NZ (GB).

(72) Inventors: **LEE, Matthew John**; School of Briosciences, University of Kent, Canterbury Kent CT2 7NJ (GB). **FRANK, Stefanie**; School of Briosciences, University of Kent, Canterbury Kent CT2 7NJ (GB). **WARREN, Martin James**; School of Briosciences, University of Kent, Canterbury Kent CT2 7NJ (GB).

(74) Agent: **DEHNS**; St Bride's House, 10 Salisbury Square, London Greater London EC4Y 8JD (GB).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

— *with sequence listing part of description (Rule 5.2(a))*

WO 2017/077320 A1

(54) **Title**: GENETICALLY MODIFIED MICROORGANISMS

(57) **Abstract**: The present invention relates to genetically modified microorganisms comprising one or more heterologous nucleic acid molecules together encoding at least three different proteins, each protein comprising an enzymatic domain and a bacterial microcompartment-targeting signal polypeptide, wherein said enzymatic domains each catalyse a different substrate to product conversion in the same metabolic pathway, and wherein said microorganisms are essentially free of bacterial microcompartments (BMCs); and to cell free systems comprising aggregates comprising at least three different proteins, each protein comprising an enzymatic domain and a bacterial microcompartment-targeting signal polypeptide, wherein said enzymatic domains each catalyse a different substrate to product conversion in the same metabolic pathway, and wherein said system does not comprise bacterial microcompartments; and to methods for the production of said microorganisms and cell free systems and their use in methods of producing a product of interest.

GENETICALLY MODIFIED MICROORGANISMS

The present invention relates to genetically modified microorganisms and their use in the production of desired products of metabolic pathways, and particularly in improving the levels of production of said products. In particular the microorganisms of the present invention are modified to comprise enzymes of a metabolic pathway for a product of interest, wherein the enzymes are each tagged with a bacterial microcompartment (BMC)-targeting signal peptide, and wherein the microorganism lacks bacterial microcompartments. The present invention also relates to cell-free systems comprising said (BMC)-targeting signal peptide tagged enzymes in the absence of BMCs.

BACKGROUND

Bacterial microcompartments (BMCs) are metabolosomes, i.e. discrete protein-based organelles in which steps of a particular metabolic pathway occur. BMCs are typically 40 to 200 nm in diameter and consist of a semipermeable proteinaceous outer layer that encases the enzymes that catalyse steps of a particular metabolic process. Thus, within a BMC, enzymatic activity of a particular metabolic pathway or part thereof is concentrated. The encapsulated environment is ideal for the channelling of toxic/volatile intermediates.

BMCs are thought to be involved in eight or more metabolic processes. BMCs are widely distributed (in approximately 17% of bacteria in 23 different phyla). Multiple BMC types can be found in a single genome. A particular BMC type is associated with a particular metabolic process. These include anabolic processes such as carbon dioxide fixation (in the carboxysome) and catabolic processes such as 1,2-propanediol utilisation (in the Pdu BMC) and ethanolamine utilisation (in the Eut BMC) and choline degradation (in the Cut BMC).

The first characterized BMC was the carboxysome, which is found in cyanobacteria and some chemoautotrophs. In the carboxysome, the enzymes carbonic anhydrase and RuBisCo are retained within the confines of the macromolecular complex to provide an environment for enhanced carbon dioxide fixation.

The propanediol utilization (pdu) operon is composed of 23 genes and encodes largely for proteins that form a BMC with a diameter of between 100 and 150 nm. Six of the genes (pduABJKUT) encode for shell proteins that comprise BMC domains as

their structural core and form hexameric tiles, which align together to form the facets and edges of the outer casing of the capsule structure. The vertices of the BMCs are thought to be formed from the pentameric PduN.

The Pdu BMC shell proteins encapsulate the enzymes for 1,2-propanediol metabolism, including the diol dehydratase (PduCDE), and the alcohol and aldehyde dehydrogenases (PduP and Q). The metabolosome also houses enzymes for the repair and reactivation of the diol dehydratase (PduG, H) and its coenzyme adenosylcobalamin (PduO, S), The shell of the BMC allows the passage of its substrates, cofactors, and coenzymes into the BMC as well as the exit of the metabolic products. This is likely mediated through the central pores that are formed within the tiles of the shell structure. Other proteins are thought to interact with the shell proteins on the external surface of the structure, including PduV, which may help to localize the BMC within the cell.

Recent studies have also revealed targeting sequences that mediate protein encapsulation within BMCs. Enzymes located within BMCs comprise such BMC-targeting signal sequences. Previous studies have demonstrated that tagging proteins not found naturally within BMCs, e.g. GFP, with a BMC-targeting signal sequence results in the localisation of those proteins within BMCs.

In the field of metabolic engineering, a number of attempts have been made to target metabolic pathways of interest to BMCs. The reasoning for such attempts is that pathway encapsulation within a BMC would permit increased flux through the pathway, control of molecules that enter and exit the BMCs, sequestration of intermediates, concentration of reagents, optimisation of reaction environment, etc..

Through heterologous expression, empty Pdu BMCs have been successfully expressed in cells that lack them in wild-type form. It was found that six proteins PduA, B, B', J, K and N were necessary and sufficient for BMC formation in *E. coli*. The absence of PduU and PduT did not prevent shell formation, even though they are known to be shell proteins. In *Salmonella*, it has been found that PduM is required, but PduA is dispensable.

Other groups have successfully targeted non-native proteins to BMCs. The present inventors have previously demonstrated the usefulness of BMCs and targeting sequences in the formation of functional bioreactors for the production of ethanol.

The BMC-targeting sequence of PduP or PduD was fused to the enzymes of the ethanol production pathway: pyruvate decarboxylase and alcohol dehydrogenase, and the tagged enzymes were co-expressed with BMC shell proteins in a bacterial cell that does not naturally produce BMCs. Co-production of tagged-enzymes for ethanol formation and the BMC shell proteins resulted in significantly more ethanol in comparison to strains with cytoplasmic (untagged) enzymes.

US 2012/0210459 discloses various means for designing and implementing BMCs for customizing metabolism in various organisms. Various sequences comprising BMCs are disclosed and the application teaches the expression of said sequences in organisms that do not naturally comprise BMCs. This document teaches co-expression of said BMCs with enzymes of interest, optionally with a BMC targeting signal peptide linked thereto. US 2013/0133102 discloses a variety of known and predicted BMC-targeting sequences.

SUMMARY

Surprisingly, the present inventors have now found that increased levels of a product of interest can be obtained using microorganisms that comprise polypeptides having enzymatic domains and BMC-targeting signal sequences, wherein the enzymatic domains catalyse steps of the same metabolic pathway for the production of said product of interest, but wherein the cell lacks the ability to produce BMCs.

In one aspect, the present invention provides a genetically modified microorganism comprising one or more heterologous nucleic acid molecules together encoding at least three different proteins, each protein comprising an enzymatic domain and a bacterial microcompartment-targeting signal polypeptide, wherein said enzymatic domains each catalyse a different substrate to product conversion in the same metabolic pathway, and wherein said microorganism is essentially free of bacterial microcompartments.

Alternatively viewed, the present invention provides a genetically modified microorganism comprising at least three different recombinant proteins, each protein comprising an enzymatic domain and a bacterial microcompartment-targeting signal polypeptide, wherein said enzymatic domains each catalyse a different substrate to product conversion in the same metabolic pathway, and wherein said microorganism is essentially free of bacterial microcompartments.

Without wishing to be bound by theory, the inventors believe that the BMC-targeting signal polypeptides mediate aggregation of said proteins. The result of the proteins of the invention each comprising an enzymatic domain and a BMC-targeting signal polypeptide is the aggregation of proteins comprising said enzymatic domains. Although aggregation is typically an undesirable occurrence in protein expression systems, the inventors have determined that in the context of multi-step metabolic pathways, it is surprisingly advantageous. As a result of the aggregation, enzymatic activity is spatially concentrated and there can be rapid channelling of the product of one enzymatically catalysed metabolic step to the active site of a second enzymatic domain that catalyses a subsequent step in which said product is a necessary substrate.

The present inventors have determined that the advantages of performing a multi-step metabolic pathway in a BMC (concentration of enzymatic activity and reaction substrates) can in fact be achieved in the absence of a BMC. Furthermore, the present inventors have demonstrated that in cells expressing multiple recombinant proteins comprising an enzymatic domain and a BMC targeting signal polypeptide, *increased* product yield is achieved in cells lacking BMCs as compared to those comprising BMCs. This finding was very surprising, and counterintuitive, given that the consensus in the field prior to the present invention was that product yields would be increased by linking enzymes to BMC-targeting sequences with the specific purpose of recruiting said enzymes into co-expressed BMCs.

Furthermore, as shown in the present Examples, the addition of BMC-targeting sequences to most enzymes reduces the specific activity of the enzymes. In addition, aggregation is usually considered to have a detrimental effect on protein function, and is therefore considered undesirable. Therefore, the present inventors' finding that despite decreased enzyme activity as a result of fusing BMC-targeting signal polypeptides to said enzymes, and despite the absence of the BMCs, an increase in product yield is observed in the microorganisms of the invention, was very surprising.

The present invention is particularly advantageous in multi-step pathways, since a greater number of steps requires a greater number of enzymatic domains and a greater number of interactions between the product(s) of one step and the active site of a subsequent enzymatic domain. In such systems requiring numerous complex

interactions, the spatial concentration of enzymatic domains and reaction substrates is particularly advantageous

Preferably, the microorganism of the invention comprises at least four, five, six, seven, eight or nine of said different recombinant proteins. Alternatively viewed, preferably, the microorganism of the invention comprises one or more heterologous nucleic acid molecules together encoding at least four, five, six, seven, eight or nine of said different proteins. Throughout this application, disclosures relating to one or more polypeptides or proteins are to be considered as disclosures relating to one or more nucleic acid molecules encoding said polypeptides or proteins, and *vice versa*.

The enzymatic domains as referred to herein each catalyse a different substrate to product conversion in the same metabolic pathway. A “metabolic pathway” is a series of substrate to product conversions, each of which is catalysed by an enzyme, wherein the product of one enzyme acts as the substrate for the next enzyme. The enzymatic domains as referred to herein each catalyse a different substrate to product conversion in the same metabolic pathway for the production of a product of interest.

The microorganisms of the invention, also termed herein “microbes”, “microbial host cells” or simply “host cells” or “cells”, may be any microorganism in which recombinant proteins can be expressed. By "microorganism" is meant any unicellular prokaryotic or eukaryotic organism. Preferred microorganisms are bacteria, cyanobacteria, microalgae, filamentous fungi and yeasts. Most preferably, the microorganism is a bacterium.

As explained in more detail below, the invention provides methods of producing a product of interest comprising growing the microorganism of the invention in a culture medium and under conditions wherein the product is produced and optionally recovering the product. Depending on the product of interest, the product may be secreted by the microorganism and recovered from the culture medium, or the product may be sequestered by the microorganism, necessitating extraction therefrom. In either case, to maximise production of the product of interest, the microorganisms are preferably tolerant to the product of interest.

In some embodiments, the microorganisms have the ability to utilize carbohydrates. Optionally, the microorganisms of the invention are photosynthetic, preferably photosynthetic bacteria.

The microorganisms of the present invention comprise at least three recombinant proteins and in some embodiments the microorganisms are further genetically modified to remove the ability of the cell to form BMCs. The ability to genetically modify the microorganism is essential for the production of any recombinant microorganism. Thus, preferably the microorganisms are competent. "Competence" is the ability of a cell to take up extracellular nucleic acid molecules from its environment. The competence may be naturally occurring or induced, i.e. artificial competence, in which the microorganisms in culture are treated to make them transiently permeable to DNA.

Preferably, the microorganisms of the present invention have the ability to grow to high cell densities. It will be within the competencies of the person of ordinary skill in the art to determine the optimal cell density for a particular microorganism and pathway of interest. Preferably, the microorganisms are thermophilic. Preferably, the microorganisms are able to grow under anaerobic conditions. Alternatively preferably, the microorganisms are able to grow under aerobic conditions.

The above characteristics of the microorganism of the invention can be conferred by mutagenesis and selection, genetic engineering, or can be natural.

Preferably, the microorganism is selected from the group consisting of *Clostridium*, *Zymomonas*, *Escherichia*, *Salmonella*, *Serratia, Erwinia, Klebsiella, Shigella, Rhodococcus*, *Pseudomonas*, *Bacillus*, *Lactobacillus*, *Enterococcus*, *Alcaligenes*, *Klebsiella*, *Paenibacillus*, *Arthrobacter*, *Corynebacterium*, *Brevibacterium*, *Schizosaccharomyces, Kluyveromyces, Yarrowia, Pichia*, *Candida*, *Hansenula*, or *Saccharomyces*.

Preferably, the microorganism is a bacterium, more preferably of the genus *Escherichia*, most preferably *E. coli*. *E. coli* is well established as an industrial microorganism used in the production of a variety of products (chemical compounds, amino acids, vitamins, recombinant proteins). The entire *E. coli* genome has also been sequenced, and the genetic systems are highly developed.

The preferred yeast organism is *Saccharomyces cerevisiae*. This organism has a long history of use in industrial processes and can be manipulated by both classical microbiological and genetic engineering techniques. It is well-characterized genetically; the entire genome of *S. cerevisiae* has been sequenced. The organism grows to high cell densities.

Preferred microalgae for use in the present invention include *Chlorella* and *Prototheca*.

As mentioned above, the microorganism of the invention is essentially free of BMCs, i.e. expresses essentially no BMCs. Preferably, the microorganism of the invention is free of BMCs, i.e. does not express BMCs. Preferably, the microorganism of the invention does not naturally express BMCs, i.e. does not naturally comprise the genes necessary for the expression of BMCs. By "naturally" in this context is meant "natively", i.e. prior to any modification according to the invention. In such microorganisms, no genetic modification is required to prevent the microorganism from expressing BMCs. Throughout the application, the terms "express" and "expresses" are interchangeable with the term "having the ability to express", i.e. comprising the genes necessary for expression.

Alternatively, preferably, the microorganism of the invention naturally expresses BMCs but has been modified to reduce essentially all, preferably all, of the ability to express BMCs. In other words, the microorganism of the invention is preferably of a species or strain that natively expresses BMCs but has been modified to reduce essentially all, preferably all, of the cell's ability to express BMCs. Suitable modifications to achieve this reduction are discussed in more detail elsewhere herein.

In nature, microorganisms that naturally express BMCs typically do so only under certain conditions, namely in the presence of inducer molecules, which for any given BMC is the substrate for the pathway that comprises steps catalysed by enzymes located within the BMCs. For instance, Pdu BMCs are only expressed by microorganisms comprising the necessary genes when said microorganisms are exposed to 1,2-propanediol. Similarly, Eut BMCs are only expressed by microorganisms comprising the necessary genes when said microorganisms are exposed to ethanolamine.

Thus, in an alternative embodiment, the microorganism of the invention is of a species or strain that naturally expresses BMCs, i.e. that comprises the genes necessary for the expression of BMCs, and wherein expression of said BMCs is inducible by the presence of one or more inducer molecules, but wherein said microorganism is in an environment, e.g. a culture medium, that does not permit expression of BMCs. In other words, preferably the genetically modified microorganism of the invention is present in a culture medium in which the level of said inducer molecule(s) is too low to induce the expression of said BMCs. Preferably the culture medium lacks said inducer molecule(s).

If the microorganism naturally expresses Pdu BMCs, then said molecule is 1,2-propanediol. If the microorganism expresses Eut BMCs, then said molecule is ethanolamine. Expression of other BMCs is known to be induced by the presence of choline, fucose or rhamnose.

Thus, the present invention provides a microorganism of the invention present in a culture medium in which the level of said inducer molecule(s) is too low to induce the expression of said BMCs. Preferably the culture medium lacks said inducer molecule(s). Alternatively viewed, the present invention provides a culture medium comprising a microorganism of the invention that naturally expresses BMCs, i.e. that comprises the genes necessary for the expression of BMCs,, wherein in said culture medium the level of said inducer molecule(s) is too low to induce the expression of said BMCs. Preferably the culture medium lacks said inducer molecule(s). Preferably, the culture media lacks one or more, preferably all of propanediol, ethanolamine, choline, fucose and rhamnose.

Microorganisms that possess the ability to produce BMCs, i.e. naturally express, i.e. natively express one or more of Pdu, Eut and carboxysome BMCs, or less common BMCs, are known in the art, for instance from Axen *et al.,* (2014) *PLOS Computational Biology* 10(10):e1003898*,* US 2012/0210459, and Jorda J, *et al.*, (2013) *Protein Science : A Publication of the Protein Society.* 22(2):179-195. Dataset S1 of Axen *et al.* comprehensively lists bacterial strains that possess the ability to produce BMCs, i.e. naturally comprise the genes necessary for production of BMCs. The skilled person would be able to determine whether or not a particular microorganism possesses the ability to produce BMCs, i.e. comprises the genes necessary for the expression of BMCs.

The microorganisms of the present invention are genetically modified to increase the yield of the product of interest. As used herein, the term “wild type microorganism” or “wild type cell” encompasses the typical, i.e. most prevalent microorganism of a species or strain as it occurs in nature. Existing strains of a particular species are not necessarily “wild type” strains, however, existing strains lack the modifications of the invention described herein. The term “wild type” is used herein as shorthand to refer to microorganisms of the same strain as the microorganism of the invention but lacking the genetic modifications of the invention as described herein, even though such microorganisms may not be the most prevalent strain. This is how the term is typically used in the field.

The microorganisms of the invention are recombinant microorganisms, i.e. they comprise one or more recombinant nucleic acid molecules. Cells and/or microorganisms may be genetically modified by genetic engineering techniques (e.g., recombinant technology), classical microbiological techniques, or a combination of such techniques. Such techniques are generally disclosed, for example, in Sambrook *et al.*, 1989, *Molecular Cloning: A Laboratory Manual,* Cold Spring Harbor Labs Press.

The genetically modified microorganisms of the invention can include a microorganism in which nucleic acid molecules have been inserted, deleted or modified (i.e., mutated; e.g., by insertion, deletion, substitution, and/or inversion of nucleotides), in such a manner that such modifications provide the desired effect of increased yields of the product of interest within the microorganism or in the culture medium.

As used herein, genetic modifications which result in a decrease in gene expression, in the function of the gene, or in the function of the gene product (i.e. the protein encoded by the gene) can be referred to as inactivation (complete or partial), deletion, interruption, blockage or down-regulation of a gene. They can be referred to as null mutations or loss of function mutations. For example, a genetic modification in a gene which results in a decrease in the function of the protein encoded by such gene, can be the result of a complete deletion of the gene (i.e., the gene does not exist, and therefore the protein is not produced), a mutation in the gene which results in incomplete or no translation of the protein (e.g., the protein is not expressed), or a mutation in the gene which decreases or abolishes the natural function of the protein (e.g., a protein is expressed which has decreased or no

enzymatic activity). Genetic modifications which result in an increase in gene expression or function can be referred to as amplification, overproduction, overexpression, activation, enhancement, addition, or up-regulation of a gene. The terms “gene expression” and “protein expression” are used interchangeably herein. Methods and types of mutation are well-known in the art and any suitable method or type can be present in the microorganisms of the present invention. Mutations include, for instance, missense mutations, nonsense mutations, insertions, deletions, duplications, frameshift mutations and repeat expansions, and any combination thereof.

Addition of recombinant genes to increase gene expression can include maintaining the recombinant gene(s) on replicating plasmids or integrating the recombinant gene(s) into the genome of the production organism. Furthermore, increasing the expression of desired recombinant genes can include operatively linking the recombinant gene(s) to native or heterologous transcriptional control elements.

The microorganisms of the invention comprise one or more heterologous nucleic acid molecules together encoding at least three different proteins (that are by definition recombinant proteins). Alternatively viewed, the microorganisms of the present invention comprise at least three different recombinant proteins. Preferably, each of said proteins is encoded by a different recombinant nucleic acid molecule. Alternatively viewed, preferably each of said heterologous nucleic acid molecules encodes only one of said proteins. Alternatively, two, three, or more of said proteins are encoded by the same recombinant nucleic acid molecule. Preferably, the microorganism of the invention comprises 3, 4, 5, 6, 7, 8 or 9 recombinant proteins as defined herein.

As used herein, the term “protein” means a polymer of amino acid residues. The terms “polypeptide” and “protein” are used interchangeably herein. The recombinant proteins of the invention each comprise a region having enzymatic activity and a BMC-targeting signal polypeptide. A mere oligopeptide comprising 2 (a dipeptide), 3 (a tripeptide) or up to about 25 amino acids is not sufficiently long to comprise a region with enzymatic activity and a BMC-targeting sequence. The proteins of the invention are preferably each a polypeptide comprising at least 75 amino acids, more preferably at least 100 amino acids, still more preferably at least 120 amino acids.

Both full length proteins and fragments thereof are contemplated by the term "protein" as used herein. "Fragments" in the context of the present invention are functional fragments, i.e. a fragment comprises the same enzymatic activity as the full length protein of which it is a fragment. The term "protein" also includes post-expression modifications to the protein, including, but not limited to, glycosylation, acetylation and phosphorylation. The term "protein" also applies to amino acid polymers in which one or more amino acid residue is an artificial chemical mimetic of a corresponding naturally occurring amino acid, as well as to naturally occurring amino acid polymers and non-naturally occurring amino acid polymers. Amino acid polymers may comprise entirely L-amino acids, entirely D-amino acids, or mixture of L- and D-amino acids.

The terms "nucleic acid" and "polynucleotide" are used interchangeably herein to refer to deoxyribonucleotides or ribonucleotides and polymers thereof in either single- or double-stranded form. The term encompasses nucleic acids containing known nucleotide analogs or modified backbone residues or linkages, which are synthetic, naturally occurring, and non-naturally occurring, which have similar binding properties as the reference nucleic acid, and which are metabolized in a manner similar to the reference nucleotides. Examples of such analogs include, without limitation, phosphorothioates, phosphoramidates, methyl phosphonates, chiral-methyl phosphonates, 2-O-methyl ribonucleotides, polypeptide-nucleic acids (PNAs). Unless otherwise indicated, a particular nucleic acid sequence also encompasses "conservatively modified variants" thereof (e.g., degenerate codon substitutions) and complementary sequences, as well as the sequence explicitly indicated.

As used herein, the term "heterologous" when applied to a nucleic acid molecule or protein means a nucleic acid molecule or protein that is not naturally, i.e. natively, present or encoded in the genome of that strain of microorganism. Thus, heterologous nucleic acid molecules are those that are introduced into the microorganism by recombinant techniques. The terms "non-native" and "heterologous" are used interchangeably. In the context of the present invention, the terms "heterologous" and "recombinant" are used interchangeably. As used herein, the term "native" when applied to a nucleic acid molecule or protein means a nucleic acid molecule or protein that is present or encoded in the genome of that strain of microorganism.

The microorganism of the present invention comprises one or more heterologous nucleic acid molecules that encode at least three proteins, each protein comprising an enzymatic domain and a bacterial microcompartment-targeting signal polypeptide. In such a heterologous nucleic acid molecule, the coding sequence for the enzymatic domain may be native to the microorganism and the coding sequence for the BMC-targeting signal polypeptide may be non-native, or *vice versa*. Alternatively, both the coding sequence for the enzymatic domain and the coding sequence for the BMC-targeting signal polypeptide may be non-native to the microorganism, and may be from the same or different non-native sources. Alternatively, the coding sequence for the enzymatic domain and the coding sequence for the BMC-targeting signal polypeptide may both be native to the microorganism but the nucleic acid molecule comprises one or more additional sequences that are non-native to the microorganism. These additional sequences may encode for other sequences within the encoded protein such as linker sequences between the enzymatic domain and the BMC-targeting signal polypeptide, or they may be regulatory sequences such as promoters. Alternatively, the coding sequence for the enzymatic domain and the coding sequence for the BMC-targeting signal polypeptide may be native to the microorganism but are not found natively in the same nucleic acid molecule, such that overall the nucleic acid molecule of the invention is heterologous to the microorganism of the invention.

The heterologous nucleic acid molecules of the present invention are recombinant nucleic acid molecules. Recombinant nucleic acid molecules, also known as "chimeric nucleic acid molecules" are nucleic acid molecules formed by laboratory methods of genetic recombination (such as molecular cloning) to combine nucleic acid sequences from two or more sources.

A "recombinant protein" is a protein that is encoded by a recombinant nucleic acid molecule, preferably by recombinant DNA (also termed "chimeric DNA"). A recombinant protein is encoded by a recombinant gene, i.e. by a chimeric gene, specifically by the coding region(s) of the chimeric gene. Thus, the recombinant nucleic acid molecules of the present invention comprise at least three different chimeric genes encoding the at least three different proteins defined herein. Recombinant protein expression is the expression of proteins within a cell from recombinant DNA. The at least three different proteins that the microorganisms of the invention comprise are recombinant proteins, and the terms "protein" and

"recombinant protein" are used interchangeably in this context throughout the application.

"Chimeric gene" refers to any gene that is not a native gene. A chimeric gene as used herein comprises regulatory and coding sequences that are not found together in nature and/or a coding sequence comprising two or more sequence regions not found together in nature. Accordingly, a chimeric gene may comprise regulatory sequences and coding sequences that are derived from different sources, or regulatory sequences and coding sequences derived from the same source, but arranged in a manner different than that found in nature. A "foreign gene", "non-native" or "heterologous gene" refers to a gene not normally found in the host organism, but that is introduced into the host organism by gene transfer. Foreign genes can comprise genes native to one organism inserted into a different, i.e. non-native, organism, or they can comprise chimeric genes. A "transgene" is a gene that has been introduced into the genome by a transformation procedure.

As explained in more detail below, the enzymatic domain (also termed "the region with enzymatic activity" herein) and the BMC-targeting signal polypeptide of a recombinant protein of the invention preferably originate from different sources, i.e. from different organisms or from different proteins within a single organism. In these embodiments, the coding sequence within the recombinant nucleic acid molecule is a chimeric sequence.

In other embodiments, the enzymatic domain and the BMC-targeting signal polypeptide of a recombinant protein of the invention originate from the same source, preferably from a microorganism that naturally expresses a protein comprising said enzymatic domain and said BMC-targeting signal polypeptide. Optionally, this microorganism is a wild type organism of the same species as the microorganism of the invention, however, typically this is not the case. In these embodiments, the coding sequence within the recombinant nucleic acid molecule is not a chimeric sequence, however, the recombinant nucleic acid molecule as a whole is a chimeric sequence due to the coding sequence being operably linked to one or more regulatory sequences, wherein the coding sequence and the one or more regulatory sequences are from different sources. If two or more regulatory sequences are present, they are preferably from different sources from each other. Preferably, the regulatory sequence(s) is/are non-native to the microorganism of the invention.

As used herein, the term “gene” refers to a nucleic acid that is capable of being expressed as a specific polypeptide, optionally including regulatory sequences preceding (5' non-coding sequences) and following (3' non-coding sequences) the coding sequence. “Native gene” or “wild type gene” refers to a gene as found in nature with its own regulatory sequences. “Endogenous gene” refers to a native gene in its natural location in the genome of an organism.

As used herein, the term “coding sequence” refers to a nucleic acid sequence that codes for a specific amino acid sequence. “Regulatory sequences” refer to nucleotide sequences located upstream (5' non-coding sequences), within, or downstream (3' non-coding sequences) of a coding sequence, and which influence the transcription, RNA processing or stability, or translation of the associated coding sequence. Regulatory sequences may include promoters, translation leader sequences, introns, polyadenylation recognition sequences, RNA processing site, effector binding site and stem-loop structure.

The term “promoter” refers to a nucleic acid sequence capable of controlling the expression of a coding sequence or functional RNA. In general, a coding sequence is located 3' to a promoter sequence. Promoters may be derived in their entirety from a native gene, or be composed of different elements derived from different promoters found in nature, or even comprise synthetic DNA segments. It is understood by those skilled in the art that different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental or physiological conditions. Promoters which cause a gene to be expressed in most cell types at most times are commonly referred to as “constitutive promoters”. It is further recognized that since in most cases the exact boundaries of regulatory sequences have not been completely defined, DNA fragments of different lengths may have identical promoter activity.

The term “operably linked” refers to the association of nucleic acid sequences on a single nucleic acid fragment so that the function of one is affected by the other. For example, a promoter is operably linked with a coding sequence when it is capable of affecting the expression of that coding sequence (i.e., that the coding sequence is under the transcriptional control of the promoter). Coding sequences can be operably linked to regulatory sequences in sense or antisense orientation.

The term “expression”, as used herein, refers to the transcription and stable accumulation of sense (mRNA) or antisense RNA derived from the nucleic acid fragment of the invention. Expression may also refer to translation of mRNA into a polypeptide. The terms “gene expression” and “protein expression” are used interchangeably herein.

The microorganism of the invention is prepared by the transformation of a microorganism with the one or more heterologous nucleic acid molecules together encoding at least three different proteins, each protein comprising an enzymatic domain and a bacterial microcompartment-targeting signal polypeptide, wherein said enzymatic domains each catalyse a different substrate to product conversion in the same metabolic pathway.

As used herein the term “transformation” refers to the transfer of a nucleic acid fragment into a host organism, resulting in genetically stable inheritance. Host organisms containing the transformed nucleic acid fragments are referred to as “transgenic” or “recombinant” or “transformed” organisms.

The one or more heterologous nucleic acid molecules of the invention that encode the at least three different proteins are each preferably comprised within a plasmid, vector or cassette. The terms “plasmid”, “vector” and “cassette” refer to an extra chromosomal element often carrying genes which are not part of the central metabolism of the cell, and usually in the form of circular double-stranded DNA fragments. Such elements may be autonomously replicating sequences, genome integrating sequences, phage or nucleotide sequences, linear or circular, of a single- or double-stranded DNA or RNA, derived from any source, in which a number of nucleotide sequences have been joined or recombined into a unique construction which is capable of introducing a promoter fragment and DNA sequence for a selected gene product along with appropriate 3' untranslated sequence into a cell. “Transformation cassette” refers to a specific vector containing a foreign gene and having elements in addition to the foreign gene that facilitates transformation of a particular host cell.

An “expression vector” is a nucleic acid construct, generated recombinantly or synthetically, with a series of specified nucleic acid elements that permit transcription of a particular nucleic acid in a host cell. The expression vector can be part of a plasmid, virus, or nucleic acid fragment. Typically, the expression vector includes a

nucleic acid to be transcribed operably linked to a promoter. The one or more heterologous nucleic acids of the invention are typically present in an expression vector, preferably a vector comprising a strong promoter.

Preferably, the at least three recombinant proteins are each over-expressed in the microorganism of the invention. As referred to herein, "over-expressed" means that expression of the gene encoding the protein is increased as compared to, i.e. relative to, the level of expression in a control microorganism, i.e. in a microorganism in the same strain which lacks said one or more heterologous nucleic acid molecules. The skilled person would appreciate that the comparison must be made between the level of expression of a protein in the microorganism of the invention and the level of expression of the same protein (i.e. a protein having the same amino acid sequence) occurring in the control microorganism.

A “control microorganism” is a microorganism of the same strain as the microorganism of the invention which has not been modified according to the invention. In this context of over-expression, a “control organism” has not been modified to over-express the gene in question. A control organism for instance may have been transformed with an "empty" vector or a vector with a control sequence. Preferably, the control microorganism is one that does not comprise the one or more heterologous nucleic acid molecules of the invention that encode a protein comprising an enzymatic domain and a bacterial microcompartment-targeting signal polypeptide. The skilled person will readily be able to determine the appropriate control with which to make the comparison of expression levels.

Gene expression is to be considered in terms of the amount of protein product produced, which may be determined by any convenient method known in the art. The terms “gene expression” and “protein expression” are used interchangeably herein. Methods for the direct and indirect determination of protein expression levels are well-known in the art and any such technique could be used by the skilled person in this regard. For example, expression can be determined by measuring protein activity. Alternatively, the amount of protein produced can be measured to determine the level of expression, for example by Western Blotting or other antibody detection systems, or indeed by any method of assessing or quantifying protein. The assay may be an *in vivo* or *in vitro* assay.

In some embodiments, the microorganism of the invention comprises a recombinant protein that is not also expressed by the control microorganism, in which case the level of expression of said protein in the control microorganism is zero. This is the situation, for instance, when the control microorganism is the same strain as the microorganism of the invention and wherein said strain does not naturally express said protein. In other embodiments, the microorganism of the invention comprises a recombinant protein that is also expressed by the control microorganism, in which case the level of expression of the protein in the microorganism of the invention nevertheless exceeds that in the control microorganism.

Thus, the microorganisms of the present invention preferably comprise at least three different recombinant proteins that are each over-expressed relative to the expression level of said protein in a control microorganism as defined above.

Preferably, the one or more heterologous nucleic acid molecules of the present invention together comprise at least three different chimeric genes, each comprising a coding sequence that encodes a different protein comprising an enzymatic domain and a BMC-targeting signal polypeptide as defined anywhere herein operably linked to one or more regulatory elements for the over-expression of said protein. Preferably, said regulatory element is a promoter, more preferably a strong promoter. Thus, preferably each of the heterologous nucleic acid molecules comprises a region encoding one or more of said proteins operably linked to a strong promoter.

Alternatively viewed, the present invention provides a genetically modified microorganism comprising one or more heterologous nucleic acid molecules together encoding at least three different proteins, each protein comprising an enzymatic domain and a bacterial microcompartment-targeting signal polypeptide, wherein said enzymatic domains each catalyse a different substrate to product conversion in the same metabolic pathway, wherein each of said proteins is over-expressed, and wherein said microorganism is essentially free of bacterial microcompartments.

Alternatively viewed, the present invention provides a genetically modified microorganism comprising at least three different proteins, each protein comprising an enzymatic domain and a bacterial microcompartment-targeting signal polypeptide, wherein said enzymatic domains each catalyse a different substrate to product conversion in the same metabolic pathway, wherein each of said proteins is over-

expressed and wherein said microorganism is essentially free of bacterial microcompartments.

Preferably, “over-expressed” or “over-expression” means that the level of expression is at least 10%, preferably at least 20%, even more preferably at least 50%, yet more preferably at least 75%, still more preferably at least 90%, and most preferably at least 100% greater in the microorganism of the invention as compared to the level of expression in the control microorganism as defined above. Alternatively, expression may be 2-, 3- or 4-fold or more higher in the microorganism of the invention as compared to the level in the control microorganism defined above.

Over-expression of a protein of interest can be achieved by any technique known in the art, and such techniques would be well known to one of ordinary skill in the art. According to the present invention "overexpressing" may mean simply that an additional gene is expressed in the microorganism beyond the native gene endogenously present in that microorganism but is not limited to such a mechanism. It may include expressing a gene in a microorganism which does not naturally contain such a gene.

Preferably, the microorganism of the invention comprises aggregates comprising said at least three different proteins.

Over-expression is preferably achieved by introducing into the microorganism a recombinant nucleic acid molecule encoding the protein, for example expressed from a stronger or unregulated promoter relative to the gene in the control microorganism, and/or by introducing multiple copies of a protein-encoding nucleic acid molecule.

Preferably, the introduced nucleic acid molecule is modified as compared to a naturally occurring gene encoding the same protein to render it relieved of transcriptional repression, e.g. by mutating or deleting recognition elements for transcriptional repressors or by using expression control elements (e.g. promoters) which are not subject to transcriptional regulation by the transcriptional regulator(s) which normally control expression of the gene. The endogenous gene may alternatively or additionally be modified in this way, or by addition of a stronger promoter. Thus, mutagenesis (including both random and targeted) may for example be used to mutate the endogenous control or regulatory elements so as to increase expression of the endogenous gene (e.g. increase transcription and/or translation).

Alternatively, the organism may be engineered to introduce additional or alternative regulatory elements.

In a particular embodiment, a gene may be expressed from a non-native or heterologous promoter (that is a promoter which is heterologous to the encoding gene, i.e. is not the native gene promoter) and particularly a strong, non-native or heterologous promoter. Thus, in this embodiment the gene is not used with its native promoter. A gene may be introduced which is under the control of a non-native promoter. As referred to herein, a strong promoter is one which expresses a gene at a high level, or at least at a higher level than effected by its native promoter. The term "strong promoter" is a term well known and widely used in the art and many strong promoters are known in the art, or can be identified by routine experimentation. The use of a non-native promoter may advantageously have the effect of relieving the gene of transcriptional repression, as at least some of any repressive elements will be located in the native promoter region. By replacing the native promoter with a non-native promoter devoid of repressive elements responsive to the effects of pathway products, the gene will be at least partly relieved of transcriptional repression.

Suitable promoters and expression vectors comprising such promoters for achieving over-expression of a protein of interest are well-known in the art, and the one or more nucleic acid molecules of the invention may comprise any such promoter or may be comprised within any such expression vector to achieve over-expression of the at least three proteins of the invention. Examples include the *E. coli* expression vector pGEX in which protein expression is under the control of the tac promoter, and the pET series of vectors which uses a T7 promoter.

As used herein, the term “enzymatic domain” means the region of the protein that performs and is necessary for the catalysis of a substrate to product conversion. The terms “region having enzymatic activity” and “enzymatic domain” are used interchangeably herein. The enzymatic domain may be a complete enzyme or a function fragment thereof, i.e. a fragment that catalyses the same substrate to product conversion as the complete enzyme. The enzymatic domain comprises a catalytic domain, i.e. an active site, and any other amino acids necessary for the domain to have a conformation that permits said active site to be functional, i.e. to perform catalysis in the presence of the relevant substrates.

In one or more of the recombinant proteins of the invention, the BMC-targeting signal polypeptide may be fully comprised within or may be partially comprised within, i.e. may overlap with, the enzymatic domain. Alternatively viewed, the enzymatic domain may comprise all or part of the BMC-targeting signal polypeptide. In such embodiments, the presence of part or all of the BMC-targeting signal polypeptide is necessary for the function of the enzymatic domain. In such embodiments, the sequences of the BMC-targeting signal polypeptide and the enzymatic domain overlap, at least partially.

Preferably, however, the enzymatic domain and the BMC-targeting signal polypeptide are distinct domains. By distinct in this context is meant that the sequences of the BMC-targeting signal polypeptide and the enzymatic domain do not overlap, i.e. the domains are structurally distinct, i.e. have distinct amino acid sequences. In such embodiments, the BMC-targeting signal polypeptide is not comprised fully or partially within the enzymatic domain, i.e. the enzymatic domain does not comprise all or part of the BMC-targeting signal polypeptide. In these embodiments, the sequence of the enzymatic domain and the BMC-targeting signal polypeptide may be directly adjacent, i.e. the C-terminal amino acid of the BMC-targeting signal polypeptide may be adjacent in sequence to the N-terminal amino acid of the enzymatic domain, or the C-terminal amino acid of the enzymatic domain may be adjacent in sequence to the N-terminal amino acid of the BMC-targeting signal polypeptide.

Optionally, the enzymatic domain and the BMC-targeting signal polypeptide are linked by an amino acid linker. The amino acid linker may comprise any number of amino acids. Preferably, the amino acid linker is 1 to 60 amino acids in length, more preferably 2 to 40 amino acids in length, most preferably 4 to 30 amino acids in length. The amino acid linker sequence may itself be a protein or polypeptide with stable secondary structure, i.e. a rigid linker, such as an alpha helical or beta sheet structure, and optionally with tertiary structure. Preferably, however, the amino acid linker sequence lacks stable secondary structure. Instead, it is preferably a random coil. A random coil is a sequence of amino acids with a conformation in which the amino acids are oriented randomly while still being bonded to adjacent amino acids.

Optionally, the linker comprises one or more sequences that assist with protein purification. In this regard, preferably, the amino acid linker comprises a sequence of 2 to 15, more preferably 2 to 12 or 3 to 12, still more preferably 2 to 6 or 3 to 6

consecutive histidine residues. Optionally, the linker comprises one or more cleavage sites.

Preferably, the BMC-targeting signal polypeptide is N-terminal to the enzymatic domain, optionally separated by a linker sequence as described above.

Preferably, the BMC-targeting signal peptide is located at the N-terminus of the protein. However, it is possible for the protein to comprise amino acids N-terminal to the BMC-targeting signal peptide.

Alternatively preferably, the BMC-targeting signal polypeptide is C-terminal to the enzymatic domain, optionally separated by a linker sequence as described above.

Preferably, the BMC-targeting signal peptide is located at the C-terminus of the protein. However, it is possible for the protein to comprise amino acids C-terminal to the BMC-targeting signal peptide.

As used herein, the term “distinct” also requires the domains to be functionally distinct. “Functionally distinct” as used herein means that the function of one of the domains does not require the presence of the other domain. In the context of the present invention, the two domains are the enzymatic domain and the BMC-targeting signal polypeptide. Thus, in these embodiments, the presence of the BMC-targeting signal polypeptide is not necessary for the function of the enzymatic domain, and *vice versa*. In such embodiments, cleavage of the BMC-targeting signal polypeptide does not remove the catalytic ability of the enzymatic domain. If the BMC-targeting signal polypeptide was necessary for the catalytic function of the enzymatic domain, then the domains would not be considered “distinct” as the term is used herein, since the enzymatic domain is the region of the protein that performs and is necessary for the catalysis of a substrate to product conversion and so would include the BMC-targeting signal polypeptide.

Preferably one or more, more preferably all, of the recombinant proteins of the invention are “bipartite proteins”. A bipartite protein is a protein with at least two functionally distinct domains. A bipartite protein of the present invention may comprise more than two functionally distinct domains but as a minimum it must contain an enzymatic domain and a BMC-targeting signal polypeptide that are functionally distinct.

Optionally, one or more of the recombinant proteins is “native” to the microorganism of the invention. By “native” is meant that the microorganism of the invention is of a strain that naturally expresses a protein with the same amino acid sequence. A recombinant protein is by definition not endogenously expressed by a microorganism because it is expressed from a recombinant nucleic acid molecule, however, the amino acid sequence of a recombinant protein may be identical to a protein expressed endogenously by the microorganism, in which case the recombinant protein is said to be native to the microorganism.

One or more of the recombinant proteins may comprise an enzymatic domain that is native and a BMC-targeting signal polypeptide that is non-native to the microorganism of the invention. Alternatively, one or more of the recombinant proteins may comprise a BMC-targeting signal polypeptide that is native and an enzymatic domain that is non-native to the microorganism of the invention. In these embodiments, the non-native components of the one or more recombinant proteins are preferably native to a different species of microorganism, i.e. a microorganism of a species other than the species of the microorganism of the invention. Alternatively, the non-native components are artificial, i.e. have a sequence not found in nature.

Optionally, the enzymatic domain and the BMC-targeting signal polypeptide of a recombinant protein are both native to the microorganism of the invention but are expressed as parts of different native proteins. In this latter embodiment, the recombinant protein is, overall, non-native to the microorganism of the invention because the strain does not naturally express a protein with the same overall sequence.

If present, amino acid linker(s) between the enzymatic domain and the BMC-targeting signal polypeptide may be native or non-native to the microorganism of the invention. Similarly, any other region of the recombinant protein may be native or non-native to the microorganism of the invention.

Preferably, one or more, more preferably all of the recombinant proteins are non-native to the microorganism of the invention. By “non-native” is meant “heterologous”, i.e. that the microorganism is of a strain that does not naturally express a protein with the same sequence. The equivalent definition of “non-native”

applies in the context of the non-native enzymatic domains, non-native BMC-targeting signal polypeptides and non-native nucleic acid molecules disclosed herein.

Non-native proteins, parts/regions of proteins, polypeptides, domains and nucleic acid molecules as referred to herein may each occur in nature, i.e. within a different organism from that of the invention, or may be artificial, i.e. not found anywhere in nature.

In a preferred embodiment, one or more of the recombinant proteins is non-native to the microorganism of the invention but is native to, i.e. expressed naturally by, another microorganism. Many bipartite proteins that comprise an enzymatic domain and a BMC-targeting signal sequence are known in the art. Examples of such enzymes include the diol dehydratase (PduDE), the propionaldehyde dehydrogenase (PduP), the phosphotransacylase (PduL) and the 1-propanol dehydrogenase (PduQ) of microorganisms naturally expressing Pdu BMCs, the aldehyde dehydrogenase (EutE) and the ethanolamine deaminase (EutC) of microorganisms naturally expressing Eut BMCs, and the gamma-carbonic anhydrase (CcmM) of microorganims naturally expressing carboxysomes.

If the recombinant protein comprises an enzymatic domain and a BMC-targeting polypeptide, both of which are non-native to the microorganism of the invention, then optionally said enzymatic domain and said BMC-targeting domain are expressed within different proteins by a single other microorganism species, or within proteins expressed by two different other microorganism species. Alternatively, one or more of said domains and polypeptides may be artificial.

Optionally, the enzymatic domain and/or the BMC-targeting sequence, any portion of the recombinant protein, or the entire recombinant protein is artificial, i.e. has an amino acid sequence not found in nature.

Preferably, one or more, more preferably all, of the recombinant proteins of the present invention are fusion proteins. As used herein, a fusion protein is a single protein having at least two domains that are not present in the same protein in nature. Naturally occurring proteins are thus not "fusion proteins" as the term is used herein. In the preferred fusion proteins of the present invention, the enzymatic domain and the BMC-targeting signal polypeptide are not present in the same protein in nature. Optionally, the enzymatic domain and the BMC-targeting signal

polypeptide of a fusion protein of the invention are expressed within different naturally occurring proteins in the same organism, which is optionally the wild type organism of the same species as the microorganism of the invention, but preferably an organism of a different species. Alternatively, the two domains are expressed naturally within different organisms. Fusion proteins are preferred bipartite proteins of the invention. The fusion proteins of the invention may also comprise or consist of artificial sequences.

As used herein, a "fusion protein construct" is a nucleic acid construct that is composed of different genes or portions thereof in operable linkage. The components include a nucleic acid molecule encoding at least an enzymatic domain as defined herein and a nucleic acid molecule encoding at least a BMC targeting signal polypeptide as defined herein.

Thus, typically the microorganisms of the invention comprise at least three different recombinant proteins that are not expressed naturally by that strain of microorganism. In some embodiments, however, one or more of the recombinant proteins is identical in amino acid sequence to a protein expressed naturally by that strain of microorganism.

The disclosure herein relating to proteins and parts thereof applies *mutatis mutandis* to nucleic acid molecules encoding said proteins and parts thereof.

The microorganisms of the invention comprise at least three different recombinant proteins that each comprise an enzymatic domain, wherein the enzymatic domains each catalyse a different substrate to product conversion in the same metabolic pathway. In the presence of suitable substrates, the microorganism of the invention produces the product of the metabolic pathway.

The product produced by the microorganism of the present invention, i.e. the product of interest, may be any product of a metabolic pathway that comprises at least three enzyme catalysed reactions. Preferably, the metabolic pathway for the product of interest comprises the formation of one or more toxic intermediates, and preferably one or more of the recombinant proteins of the invention catalyses a step in which said toxic intermediate is a substrate or product. Preferred products are organic compounds, preferably alcohols. Preferred alcohols are C_1 to C_{12} alcohols, more preferably C_2 to C_6 alcohols. Preferred alcohols are selected from the group

consisting of ethanol, 1-propanol, 2-propanol, 1-butanol, 2-butanol, isobutanol, tert-butanol, 1,2-propanediol, 1,3-propanediol, 1,2- butanediol, 1,3-butanediol, 1,4-butanediol and 2,3-butanediol.

If the product of interest is 1,2-propanediol, then the enzymatic domains each catalyse a different substrate to product conversion selected from:

i) glycerol to dihydroxyacetone
ii) dihydroxyacetone to dihydroxyacetone phosphate
iii) dihydroxyacetone phosphate to methylglyoxal
iv) methylglyoxal to lactaldehyde
v) lactaldehyde to 1,2-propanediol

Preferably, the microorganism comprises at least three different recombinant proteins comprising enzymatic domains that each catalyse a different substrate to product conversion selected from those listed above. More preferably, the microorganism comprises four recombinant proteins comprising enzymatic domains that each catalyse a different substrate to product conversion selected from those listed above.

Preferably said enzymatic domains are present in the form of complete enzymes. Preferably, the enzyme that catalyses the conversion of glycerol to dihydroxyacetone is of the class EC 1.1.1.6, and is preferably glycerol dehydrogenase. Preferably the enzyme that catalyses the conversion of dihydroxyacetone to dihydroxyacetone phosphate is of the class EC 2.7.1.29, and is preferably dihydroxyacetone kinase. Preferably, the enzyme that catalyses the conversion of dihydroxyacetone phosphate to methylglyoxal is of the class EC 4.2.3.3, and is preferably methylgloxal synthase. Preferably, the enzyme that catalyses the conversion of methylglyoxal to lactaldehyde is of the class EC 1.1.1.6, and is preferably glycerol dehydrogenase. Preferably, the enzyme that catalyses the conversion of lactaldehyde to 1,2-propanediol is of the class EC 1.1.1.77, and is preferably 1,2-propanediol oxidoreductase.

Thus, preferably, the microorganism of the invention comprises at least three different recombinant proteins, each protein comprising a different enzyme and a bacterial microcompartment-targeting signal polypeptide, wherein each of said enzymes is selected from those listed above, and wherein said microorganism is essentially free of bacterial microcompartments. Such organisms produce 1,2-

propanediol in the presence of a suitable substrate, such as glycerol. Preferably, the microorganism comprises all four of the above enzymes.

Each of the at least three proteins of the invention comprises a BMC-targeting signal polypeptide. A BMC-targeting signal polypeptide is a polypeptide that, when present within a protein, mediates the encapsulation of said protein within a BMC, if present. Such polypeptides are known in the art and the proteins of the invention can comprise any such polypeptide. Two, three or more of the at least three proteins of the invention may comprise the same BMC-targeting signal polypeptide. Alternatively, each protein may comprise a different BMC-targeting signal polypeptide.

In nature, BMC proteins are typically expressed from the same operon as the proteins, e.g. enzymes, that comprise BMC-targeting polypeptides and that localise within the BMCs. Thus, in nature, BMC-targeting signal polypeptides tend to localise within a particular type of BMC, i.e. are directed to a particular type of BMC. For example, in the Pdu operon, the Pdu BMC capsule proteins are expressed from the same operon as the Pdu enzymes, and the BMC-targeting signal sequences within said Pdu enzymes are localise to, i.e. are directed to, the Pdu BMCs. The terms "BMC-targeting signal polypeptide" and "BMC-targeting signal sequence" are used interchangeable herein.

US 2013/0133102 discloses known BMC-targeting signal polypeptides and the types of BMCs to which they are directed. The BMC-targeting signal polypeptides disclosed in this document may be used as BMC-targeting signal polypeptides according to the present invention and US2013/0133102 is incorporated herein by reference. Optionally, the BMC-targeting signal polypeptide of the present invention is a BMC-targeting polypeptide disclosed as such in US 2013/0133102.

The BMC-targeting signal polypeptides of the invention each preferably comprise a region with alpha-helical conformation. Preferably the BMC-targeting signal polypeptides of the invention each comprise an amphipathic alpha helix. Preferably, the BMC-targeting signal polypeptide is adjacent to an N-terminal and/or a C-terminal region without stable secondary structure, i.e. a random coil. Methods of predicting the secondary structure of a given amino acid sequence are well-known in the art, as are methods of designing a synthetic amino acid sequence with a desired secondary structure.

The term “amphipathic alpha helix” or “amphipathic α-helix” refers to a polypeptide sequence that can adopt a secondary structure that is helical with one surface, i.e., face, being polar and the other surface being a nonpolar face. Typically, the polar face comprises primarily polar and/or charged amino acids and the non-polar face comprises primarily hydrophobic amino acids. Methods of predicting the hydrophobicity of peptide sequences and secondary structure conformations such as alpha helices are well-known in the art, for instance Pepfold and Pepwheel.

As used herein, hydrophobic amino acids are considered primarily to include amino acid residues, such as Ile (I), Leu (L), Val (V), Met (M), Phe (F), Tyr (Y), Ala (A), Trp (W). Polar uncharged amino acids are considered primarily to include amino acids such as Gln (Q), Asn (N), Thr (T), Ser (S), and Cys (C). Charged amino acids are considered primarily to include amino acids such as Asp (D), Glu (E), Arg (R), Lys (K), and His (H). Proline and glycine are considered neutral amino acids and are not assigned to a specific group.

Proline tends to break or kink helices because it cannot donate an amide hydrogen bond (having no amide hydrogen), and because its side chain interferes sterically. Its ring structure also restricts its backbone dihedral angle to the vicinity of −70°, which is less common in α-helices. One of skill understands that although proline may be present at certain positions in the BMC-targeting signal polypeptides described herein, the presence of more than three prolines within the sequence would be expected to disrupt the helical structure. Accordingly, the BMC-targeting signal polypeptides of the invention preferably do not comprise more than three prolines, more preferably do not comprise more than two prolines within the alpha-helix forming sequence thereof.

Preferably, the BMC-targeting signal polypeptides are capable of forming coiled coils in solution. Coiled coils are a well-known structural conformation comprising two or more alpha helices coiled together in a manner akin to the strands of a rope. Preferably, the BMC-targeting signal polypeptides are capable of forming coiled coil dimers in solution. The alpha helices within a coiled coil may be arranged in a parallel or anti-parallel conformation. Preferably, the coiled coil has a left-handed conformation, although it may instead have a right-handed conformation. Method of determining the existence of coiled coil conformations are well-known in the art, for instance CCBuilder v1.0 available at http://coiledcoils.chm.bris.ac.uk/app/cc_builder/.

The BMC-targeting polypeptide may be of any length. Preferably, the BMC-targeting signal polypeptide is 5 to 70 amino acids in length, more preferably 10 to 22 amino acids in length, most preferably 14 to 20 amino acids in length. Preferably, the BMC-targeting polypeptide comprises an alpha-helical region that is at least 7 amino acids in length.

As discussed above, naturally occurring BMC-targeting signal polypeptides are found within proteins, typically enzymes, that localise to BMCs. Typically, they are N-terminal sequences; they may however be C-terminal sequences, and in rare instances they are not N- or C-terminal sequences but rather are located within the interior of the naturally occurring protein sequence.

Preferably, the BMC-targeting signal polypeptide of the present invention comprises the N-terminal 70, more preferably 60, more preferably 50, 40, 30, 25, 20, 19, 18, 17, or 16 amino acids of a naturally occurring protein that comprises an N-terminal BMC-targeting signal polypeptide. Alternatively preferably, the BMC-targeting signal polypeptide of the present invention comprises the C-terminal 70, more preferably 60, more preferably 50, 40, 30, 25, 20, 19, 18, 17, or 16 amino acids of a naturally occurring protein that comprises a C-terminal BMC-targeting signal polypeptide.

Preferred naturally occurring proteins that include BMC-targeting signal polypeptides are PduD and PduP preferably from *Citrobacter freundii*, *Propionibacterium acnes, Fusobacterium ulcerans, Escherichia coli, Pectobacterium wasabiae, Listeria monocytogenes, Shewanella* sp, *Tolumonas aurensis*, *Yersinia frederiksenii*, *Klebsiella pneumoniae*, *Salmonella typhimurium*, *Salmonella enterica Paratyphi* B str. and *Citrobacter koseri*, more preferably from *Citrobacter freundii*. In all of these proteins, the BMC-targeting signal polypeptide is an N-terminal sequence.

Preferably, the BMC-targeting signal polypeptide of the present invention comprises the N-terminal 70, more preferably 60, more preferably 50, 40, 30, 25, 20, 19, 18, 17, or 16 amino acids of a naturally occurring PduP or PduD protein from an organism that naturally expresses Pdu BMCs, preferably from those microorganisms listed above, more preferably from *Citrobacter freundii*.

Preferably, the BMC-targeting signal polypeptide of the present invention comprises residues 1 to 16, more preferably residues 1 to 18 of PduP from an organism that

naturally expresses Pdu BMCs, preferably from those microorganisms listed above, more preferably from *Citrobacter freundii*, or residues 1 to 18 of PduD from an organism that naturally expresses Pdu BMCs, preferably from those microorganisms listed above, more preferably from *Citrobacter freundii.*

Preferably, the BMC-targeting signal polypeptide comprises the following sequence:
$X_1X_2X_3X_4X_5X_6X_7X_8X_9$
wherein:
X_1, X_4, X_5, X_8, and X_9, are hydrophobic amino acids;
X_2, X_3 and X_6 are each independently polar or charged amino acids; and
X_7 is any amino acid. (SEQ ID NO: 76)

Preferably, X_1, X_4, X_5, X_8, and X_9 are each independently hydrophobic amino acids selected from the group consisting of I, L, V, M, F, Y, A and W;
X_2, X_3 and X_6 are each independently polar or charged amino acids selected from the group consisting of Q, N, T, S, C, D, E, R, K and H; and
X_7 is any amino acid. (SEQ ID NO: 77)

Preferably, X_1, X_4, X_5, X_8, and X_9 are each independently hydrophobic amino acids selected from the group consisting of I, L, V, M, and A;
X_2, X_3 and X_6 are each independently polar or charged amino acids selected from the group consisting of Q, T, E, R, S, D and K; and
X_7 is any amino acid. (SEQ ID NO: 78)

Preferably, X_1 is a hydrophobic amino acid selected from the group consisting of I, L, V and A;
X_2 is a polar or charged amino acid selected from the group consisting of E, R and Q;
X_3 is a polar or charged amino acid selected from the group consisting of T, Q, E, S, D and K;
X_4 is a hydrophobic amino acid selected from the group consisting of I, L, V and M;
X_5 is a hydrophobic amino acid selected from the group consisting of I, L and V;
X_6 is a polar or charged amino acid selected from the group consisting of R, K, Q and E;
X_7 is any amino acid;
X_8 is a hydrophobic amino acid selected from the group consisting of I, L, V and A; and

X_9 is a hydrophobic amino acid selected from the group consisting of I, V, L and M (SEQ ID NO: 79).

Preferably, the BMC-targeting signal polypeptide comprises a sequence selected from the group consisting of LEQIIRDVL (SEQ ID NO:1), LETLIRTIL (SEQ ID NO:2), LETLIRNIL (SEQ ID NO:3), LRQIIEDVL (SEQ ID NO:4), IEEIVRSVM (SEQ ID NO:5), IEQVVKAVL (SEQ ID NO:6), VEKLVRQAI (SEQ ID NO:7), IQEIVRTLI (SEQ ID NO:8), VEEIVKRIM (SEQ ID NO:9), IESMVRDVL (SEQ ID NO:10), VQDIIKNVV (SEQ ID NO:11), IRQVVQEVL (SEQ ID NO:12), VRSVVEEVV (SEQ ID NO:13) and ARDLLKQIL (SEQ ID NO:14) or a variant thereof, more preferably LEQIIRDVL (SEQ ID NO: 1) or a variant thereof, or LETLIRNIL (SEQ ID NO: 3) or LRQIIEDVL (SEQ ID NO: 4) or a variant thereof, most preferably LETLIRNIL (SEQ ID NO:3) or LRQIIEDVL (SEQ ID NO: 4) or a variant thereof.

Preferably, the BMC-targeting signal polypeptide comprises a sequence selected from the group consisting of the sequences shown in the table below:

SEQ ID NO:	Sequence
15	(V/I)(V/Y)G(Q/K)(V/A/G/E)(Y/S/Q)(I/V/L/F)(N/Q/S/L)(K/Q/R)(M/L)(L/M/R)(V/L/C/Q)(T/S)(L/M)FP(H/D/E)(R/N/Q)
16	(L/F)(S/P/A)(P/V)(E/Q)Q(A/S/Q/W)(Q/E/R)RIY(R/Q)G(S/N)
17	M(D/N)(E/Q)(K/Q)(Q/E)(L/I)(K/R/E)(E/D)(I/M)(V/I)(R/E)(S/Q)(V/I)(L/M)A(E/Q/S)
18	(A/K/S)(E/D)(A/E)L(I/V)(E/D/N)(L/E/S)(I/L)(V/I)(R/K/E/Q)(K/R)VL(E/A)(E/K)L
19	MEI(N/D/T)E(K/E)(L/V)(L/V)(R/E)Q(I/V)(I/V)(E/K/A)(D/E)VL(K/S/R/A)(E/D)
20	(M/I)(N/D)(T/E)(D/K)(A/L)(I/L)E(S/E)(M/I)V(R/K)(D/E/Q)VL(S,N)(M/L)(N/E/G)S
21	M(N/D/E)(T/S/E)(S/L)E(L/V)E(T/Q/K/D)(L/I)(I/V)(R/K)(T/N/K)(I/V)(L/I)(S/L/R/N)E
22	(A/P)(K/G)(S/Q)(S/D)(L/A)(T/N)E(E/Q)(D/Q)(I/V)Y(D/E)AVK(K/R)(V/I)(L/I)(E/G)(Q/E/S)(H/S)G(A/S)LD(P/V)
23	MN(D/T)(I/T)(E/Q)(I/L)(A/E)(Q/N)(A/M)(V/I)(S/R/A)(T/K/N)IL(S/A/E/R)(D/K)(N/F/Y)(T/L/G)K
24	LD(A/E)ES(A/V)(A/G)D(M/I)(T/A)E(M/Q)I(A/L)K(E/G)(L/M)(K/Q)(E/D)AG
25	(D/P)(D/N)(A/E)(D/E/A)L(V/I)A(E/A/S)IT(K/R)(K/R/Q)V(M/L)(A/E)QL(G/K)
26	VNEQ(L/M)VQDIV(Q/R/K)EVVA(K/R)MQI(S/T)
27	DQE(A/Q)LV(K/Q)(A/L)IT(D/E)(Q/R/E)VMA(A/E)L(K/S)K

28	MQ(I/A)(D/T)EE(L/A)IRSVV(A/Q)(Q/E)VL(A/S)(E/Q)(V/L)(G/N)
29	(E/Q/D)(N/E/D)(V/I/L)(E/Q/A)(R/Q/D)(I/L/V)(I/L/V)(K/R/N)(E/Q/K)(V/I/L)(L/I/V)(E/Q/G)(Q/R/A)(L/M)(K/G/S)
30	M(A/D)(K/I/N/L)(R/Y/)(E/N/S/L/F)(T/S)(P/N)(R/K)(V/L/F)(K/A)(E/V/M)(L/A)(A/T)(E/K)(R/N)(L/M)
31	I(E/D/G)ALR(A/E/D)ELR(A/R)L(V/I)(V/A)EEL(A/R)(Q/E)L(I/N/G)(K/R)(R/Q)

More preferably, the BMC-targeting signal polypeptide comprises a sequence selected from the group consisting of SEQ ID NO: 15, 17, 18, 19, 20, 21, 22, 23, 26, 28 and 29.

More preferably, the BMC-targeting signal polypeptide comprises a sequence selected from the group consisting of SEQ ID NO:19, SEQ ID NO:20 and SEQ ID NO:21. Still more preferably, the BMC-targeting signal polypeptide comprises a sequence selected from SEQ ID NO: 19 and SEQ ID NO:21.

Preferably, the BMC-targeting signal polypeptide comprises a sequence selected from the group consisting of the sequences shown in the table below:

SEQ ID NO:	Sequence
32	(V/I)(V/Y)G(Q/K)(V/A/G/E)(Y/S/Q)(I/V/L/F)(N/Q/S)(K/Q/R)(M/L)(L/M)(C/Q)(T/S)(L/M)FP(H/D/E)(R/N/Q)
17	M(D/N)(E/Q)(K/Q)(Q/E)(L/I)(K/R/E)(E/D)(I/M)(V/I)(R/E)(S/Q)(V/I)(L/M)A(E/Q/S)
33	(A/K/S)(E/D)(A/E)L(I/V)(E/D/N)(E/S)(I/L)(V/I)(R/K/E/Q)(K/R)VL(E/A)(E/K)L
34	MEI(N/D/T)E(K/E)(L/V)(L/V)(R/E)Q(I/V)(I/V)(E/K)(D/E)VL(K/S/R/A)(E/D)(M/L)
20	(M/I)(N/D)(T/E)(D/K)(A/L)(I/L)E(S/E)(M/I)V(R/K)(D/E/Q)VL(S,N)(M/L)(N/E/G)S
21	M(N/D/E)(T/S/E)(S/L)E(L/V)E(T/Q/K/D)(L/I)(I/V)(R/K)(T/N/K)(I/V)(L/I)(S/L/R/N)E
22	(A/P)(K/G)(S/Q)(S/D)(L/A)(T/N)E(E/Q)(D/Q)(I/V)Y(D/E)AVK(K/R)(V/I)(L/I)(E/G)(Q/E/S)(H/S)G(A/S)LD(P/V)
35	MN(D/T)(I/T)(E/Q)(I/L)(E)(Q/N)(A/M)(V/I)(S/R)(T/K/N)IL(S/A/E/R)(D/K)(N/F/Y)(T/L/G)K
26	VNEQ(L/M)VQDIV(Q/R/K)EVVA(K/R)MQI(S/T)
36	MQ(I/A)(D/T)EE(L/A)IRSVVQ(Q/E)VL(A/S)(E/Q)(V/L)(G/N)
37	(E/Q/D)(N/E/D)(V/I/L)(E/Q)(R/Q/D)(I/L/V)(I/L/V)(K/R/N)(E/Q/K)(V/I/L)(L/I/V)(E/Q/G)(Q/R/A)(L/M)(K/G/S)

More preferably, the BMC-targeting signal polypeptide comprises a sequence selected from the group consisting of SEQ ID NO: 34, SEQ ID NO:20 and SEQ ID NO:21. Still more preferably, the BMC-targeting signal polypeptide comprises a sequence selected from SEQ ID NO: 34 and SEQ ID NO:21.

Preferably, the BMC-targeting signal polypeptide comprises a sequence selected from the group consisting of the sequences shown in the table below, or a variant thereof:

SEQ ID NO:	Sequence
38	VYGKEQFLRMRQSMFPDR
39	LAPEQQQRIYRGN
40	MDQKQIEEIVRSVMAS
41	MNQQDIEQVVKAVLLKM
42	NTELVEEIVKRIMKQL
43	MEINEKLLRQIIEDVLRDM
44	MEINEKLLRQIIEDVLRD
45	MEINEKLLRQIIEDVLSE
46	MNTDAIESMVRDVLSRMNS
47	MNTSELETLIRTILSE
48	MNTSELETLIRNILSE
49	MNTSELETLIRNILSEQL
50	AGTNYTEEQVFAAVKKVLNSSGSTDV
51	MVAKAIRDHAGTAQPSGNA
52	IDIILAQQITVQIVKELKERG
53	DNADLVASITRKVMEQLG
54	VNEQLVQDIIKNVVASMQLT
55	EPEDNEDVQAIVKAIMAKLNL
56	DTEMLVKMITEQVMAALKK
57	MQATEQAIRQVVQEVLAQLN
58	EVEALVQRLTEEILRQLQ
59	IDETLVRSVVEEVVRAF
60	EDARDLLKQILQALS
61	MDIREFSNKFVEATKNM
62	LDALRAELRALVVEELAQLIKR
63	MALREDRIAEIVERVLARL

Preferably, the BMC-targeting signal polypeptide comprises a sequence selected from the group consisting of SEQ ID Nos: 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 54, 57, 59 and 60 or a variant thereof

Preferably, the BMC-targeting signal polypeptide comprises a sequence selected from the group consisting of SEQ ID NOs: 43, 44, 45, 47, 48 and 49 or a variant thereof. More preferably, the BMC-targeting signal polypeptide comprises a sequence selected from the group consisting of 45, 48 or 49 or a variant thereof.

The microorganisms of the invention may comprise variants of any nucleic acid or polypeptide sequence disclosed herein, e.g. variants of the disclosed BMC-targeting signal polypeptides. By “variant” is meant a sequence with at least 75% identity (sequence identity) to the sequence disclosed herein. e.g. to the sequence of a disclosed BMC-targeting signal polypeptide. As used herein, variants retain the same function as the nucleic acid or polypeptide of which they are a variant. Variant BMC-targeting signal polypeptides have the function of directing the polypeptide to which they are attached to a BMC, if present.

The term "percent identity", as known in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the sequences. In the art, "identity" also means the degree of sequence relatedness between polypeptide or polynucleotide sequences, as the case may be, as determined by the match between strings of such sequences. To determine the percent sequence identity of two amino acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in the sequence of one polypeptide for optimal alignment with the other polypeptide). The amino acid residues at corresponding amino acid positions are then compared. When a position in one sequence is occupied by the same amino acid residue as the corresponding position in the other sequence, then the molecules are identical at that position. "Identity" and "similarity" can be readily calculated by known methods, such as but not limited to Clustal and BLAST.

Preferably, a variant sequence has at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% identity to the sequence to which it is compared, e.g. to the BMC-targeting signal polypeptides disclosed herein.

Preferably, a variant amino acid sequence as referred to herein comprises no more than 6, more preferably no more than 5, more preferably no more than 4, more preferably no more than 3, more preferably no more than 2, most preferably no more than 1 mismatch(es) with the sequence to which it is compared, e.g. to a BMC-targeting signal polypeptide disclosed herein. A mismatch is a non-identical amino acid in the same position.

Preferably the variant sequences comprise only conservative substitutions as compared to the original amino acid sequence. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. In the context of the present invention, a conservative amino acid substitution is the replacement of a hydrophobic amino acid with another hydrophobic amino acid, the replacement of a polar amino acid with another polar amino acid or the replacement of a charged amino acid with another charged amino acid.

The microorganism of the present invention is essentially free of BMCs. As used herein, the term “essentially free of BMCs” means essentially free of functional BMCs. A functional BMC is a BMC with the morphology as seen in a microorganism that naturally expresses the BMC. Such organisms are well-known in the art and the morphology of naturally occurring functional BMCs is well-characterised.

BMCs are functional when they have an intact capsule, i.e. a closed shell. Such a morphology is termed a closed capsule morphology herein. Disruptions in BMC formation can lead to improperly formed BMCs that are not fully closed. Functional and non-functional BMCs, i.e. those with and without the correct closed capsule morphology can be readily identified by a number of techniques known in the art, e.g. by transmission electron microscopy. Preferably, the microorganism of the invention is essentially free of closed BMCs. Preferably the microorganism of the invention is essentially free of functional and improperly formed BMCs, i.e. those with and without a closed capsule morphology, i.e. closed and un-closed shells.

Preferably, the microorganism is essentially free of one or more, preferably all of Pdu, Eut and carboxysome BMCs. As mentioned above, preferably, the microorganism of the invention does not naturally express, i.e. comprise, BMCs. Preferably, the microorganism of the invention does not naturally express any BMCs.

Preferably, the microorganism is of a species or strain that does not naturally express BMCs. Alternatively viewed, the microorganism of the invention is preferably of a species or strain that lacks the genes necessary for the formation of BMCs. Such microorganisms are entirely free of BMCs, i.e. they lack BMCs. In the case of microorganism species or strains that do not produce BMCs naturally, the microorganisms of the present invention of the same species or strains are entirely free of said BMCs, and this is achieved simply by not genetically engineering the microorganisms to provide BMC production capability. Clearly, microorganisms that are not bacteria lack any BMCs. Furthermore, the vast majority of sequenced bacterial species lack BMCs.

In an alternative preferred embodiment, the microorganism of the invention is of a species or strain that naturally expresses BMCs but has been modified to reduce, i.e. it comprises modifications that reduce, essentially all of the cell's ability to express BMCs. In other words, the microorganism of the invention is preferably of a species or strain that naturally expresses BMCs but has been modified to reduce essentially all of the ability to express BMCs. Thus, in such microorganisms, the native ability of the cell to express (i.e. produce) BMCs is interrupted, inhibited, or deleted, i.e. such that it expresses no more than *de minimis* level of BMCs.

Alternatively viewed, preferably the microorganism of the invention is of a species or strain that naturally comprises BMCs but has been genetically modified to inhibit essentially all BMC formation. In other words, the microorganism is of a species or strain that possesses the genes necessary for the formation of BMCs but has been modified such that essentially all BMC formation is inhibited, interrupted or deleted, i.e. such that it produces no more than *de minimis* level of BMCs.

In the embodiments in which the microorganism of the invention is of a strain that naturally comprises the genes necessary for the expression of BMCs, the microorganism of the invention preferably comprises a loss of function mutation in one or more of said genes. Preferably, the microorganism comprises loss of function mutations in at least one gene encoding a protein comprising a BMC domain and at least one gene encoding a protein comprising a bacterial microcompartment vertex domain. Preferably, the microorganism comprises a loss of function mutation in the regulatory region of the BMC operon(s) that are present in the naturally occurring microorganism strain, preferably in the operon's promoter.

In the embodiments in which the microorganism of the invention is of a strain that naturally comprises the genes necessary for the expression of BMCs, the microorganism of the invention preferably comprises a disruption in one or more of said genes. Preferably, the microorganism comprises a disruption in at least one gene encoding a protein comprising a BMC domain and at least one gene encoding a protein comprising a bacterial microcompartment vertex domain. Preferably, the microorganism comprises a disruption in the regulatory region of the BMC operon(s) that are present in the naturally occurring microorganism strain, preferably in the operon's promoter.

As mentioned above, a loss of function mutation is a disruption of a gene. A loss of function mutation, i.e. disruption, may comprise complete or partial inactivation of the gene, for instance by missense mutations, nonsense mutations, insertions, deletions, duplications, frameshift mutations, repeat expansions and any combination thereof.

Thorough analyses of bacterial genomes have been performed to identify species that comprise BMC genes (US 2012/0210459, Axen et al., (2014) *PLOS Computational Biology* 10(10):e1003898 and Jorda J, *et al.*, (2013) *Protein Science : A Publication of the Protein Society.* 22(2):179-195). It was determined that 23 different types of BMCs were encoded in 30 distinct locus (sub)types found in 23 bacterial phyla. Dataset S1 of Axen *et al.* comprehensively lists sequenced bacterial species that naturally comprise BMCs.

By "essentially free of BMCs" is meant that the microorganism is in essence free of BMCs but it does not mean that there is a strict requirement for the microorganism to lack BMCs entirely. There is potential for a *de minimis* level of BMC production even after steps have been taken to inhibit the production of BMCs. This is because microorganisms are biological systems that cannot be as precisely controlled as, for instance, mechanical or chemical systems. For instance, random genetic mutation could, in rare instances, lead to the expression of BMC capsule proteins. While a detailed inspection may reveal that some BMCs are present, they are present in such small quantities that for the purposes intended they can be considered absent. The microorganism of the invention is essentially free of BMCs if it has been genetically modified to inhibit essentially all BMC formation. Preferably, the microorganism is free, i.e. entirely free of BMCs.

In any case, the level of BMC production in the microorganisms of the present invention is essentially zero. In embodiments in which the microorganism is of a strain that naturally produces BMCs, the level of BMC production in the microorganism of the present invention is preferably at most 10%, more preferably at most 5%, still more preferably at most 1% of the level of BMC production in the same strain that has not been modified according to the invention. The person of ordinary skill in the art will readily be able to determine the extent of BMC production in a cell or population of cells, for instance by transmission electron microscopy of sectioned cells, and quantifying relative protein levels by SDS gel electrophoresis.

Genetically engineering a microorganism to be essentially free of one or more types of BMC would be within the competencies of one of ordinary skill in the art and any suitable approach may be used. The genes required for BMC formation are well characterised, for instance in US 2012/0210459, Chowdhury *et al.*, (2014) *Microbiol. Mol. Biol. Rev.* 78(3): 438, and Axen et al., (2014) *PLOS Computational Biology* 10(10):e1003898.

Preferably, one or more of the proteins required for BMC formation are down-regulated in the microorganism of the present invention. By down-regulated is meant that the level of expression of said protein is in the microorganism of the invention is at most 10%, preferably at most 5%, more preferably at most 1% of the level of expression of said protein in a microorganism of the same strain that has not been modified according to the invention. As used herein, down-regulation of a protein is equivalent to down-regulation of a gene encoding that protein, and *vice versa*. Most preferably, the microorganism of the invention lacks the one or more proteins required for BMC formation, i.e. the microorganism is one in which the protein or the gene encoding said protein has been eliminated. Most preferably, the microorganism of the invention lacks the one or more genes required for BMC formation, i.e. the microorganism is one in which the one or more genes required for BMC formation have been eliminated.

Genetic engineering techniques for down-regulating and eliminating the expression of a gene/protein of interest are well-known in the art and any such technique may be used in the context of the present invention. Preferably, the microorganism of the invention comprises a deletion, interruption or deleterious mutation in one or more of the genes encoding a protein required for BMC formation that results in the reduction or elimination of expression of said protein. Alternatively, the relevant gene may be

silenced using a short DNA or RNA oligonucleotide that has a sequence complementary to either gene or an mRNA transcript, e.g. antisense oligonucleotides.

The down-regulated protein required for formation of functional BMCs (i.e. BMCs having an intact, closed capsule) is preferably either:

i) a protein comprising a BMC-domain. A BMC domain is a domain common to the majority of known BMC shell proteins. The BMC domains are typically flat hexamers that tile edge to edge to form extended protein sheets; or

ii) a pentameric bacterial microcompartment vertex (BMV) protein. These proteins are non-BMC-domain shell proteins that form the vertices of the BMC capsule.

Both BMC-domain containing proteins and BMV proteins are required for the formation of intact, closed BMCs. Preferably, the microorganism of the present invention is essentially free of closed BMCs.

Preferably, the expression of one or more proteins comprising a BMC domain is down-regulated in the microorganism of the present invention. Preferably the expression of one or more BMV proteins is down-regulated in the microorganism of the present invention. Preferably, the expression of at least one protein comprising a BMC domain and at least one protein comprising a BMV domain is down-regulated in the microorganism of the invention.

As mentioned above, a number of types of BMCs are known in the art, including Pdu BMCs, Eut BMCs, and carboxysomes. Throughout, “Pdu” stands for “propanediol utilization” and Eut” stands for “ethanolamine utilization”. In nature, BMC proteins are typically expressed from the same operon as the proteins, e.g. enzymes, that comprise BMC-targeting polypeptides and that localise within the BMCs. For example, in the Pdu operon, the Pdu BMC proteins are expressed from the same operon as the Pdu enzymes. Some naturally occurring microorganisms express more than one type of BMC. For instance, *Salmonella* possesses the genes necessary for expression of both Pdu and Eut BMCs.

In bacteria that naturally express Pdu BMCs, it is known that the formation of functional BMCs with closed capsules requires at least the expression of the BMC shell proteins PduA, B, B’, J, K, M, N, T and U. Preferably, the expression of PduA, B, B’, J, K, M, N, T or U or any combination thereof is down-regulated in the

microorganisms of the present invention. Preferably, the expression of PduN is down-regulated. PduN is known to be the BMV protein of the closed Pdu BMC and it has been shown previously that PduN deletion mutants form grossly abnormal, non-functional BMCs. Preferably the expression of PduB and PduB' are down-regulated. PduBB' deletion mutants have been shown previously to be unable to form BMCs. Preferably the expression of PduJ is down-regulated. PduJ deletion mutants have been shown previously to be unable to form functional BMCs.

Preferably, the expression of PduN, B, B', J, M or A or any combination thereof is down-regulated. Preferably, the expression of PduN, B, B', J or A or any combination thereof is down-regulated, particularly preferably when the microorganism is *Salmonella*. Preferably, the expression of PduN and any one or more of Pdu B, B', J, M and A is down-regulated. Preferably, the expression of PduN, B, B', J, M and A is down-regulated. Preferably, the expression of PduA, B, B', J, K, N, T and U is down-regulated. Preferably, the expression of PduN, B, B', J and A is down-regulated, particularly preferably when the microorganism is *Salmonella*. Preferably, the expression of PduA, B, B', J, K, M, N, T and U is down-regulated.

Preferably, the expression of pduB, pduB′, pduJ, or pduN or any combination thereof is down-regulated, most preferably the expression of pduB, pduB′, pduJ, and pduN is down-regulated.

In bacteria that naturally express Eut BMCs, it is known that the formation of functional, closed BMCs requires at least the expression of the BMC shell proteins EutK, M, S, L and N. Preferably, the expression of EutK, M, S, L or N or any combination thereof is down-regulated in the microorganisms of the present invention. Preferably, the expression of EutN is down-regulated. EutN is known to be the BMV protein of the closed Eut BMC. Preferably, the expression of EutN and any one or more of EutK, M, S and L is down-regulated. Preferably, the expression of EutN, EutM, Eut S and EutL are down-regulated. Preferably, the expression of EutK, M, S, L and N are down-regulated.

In bacteria that naturally express the alpha-carboxysome, it is known that the formation of functional, closed BMCs requires at least the expression of the proteins CsoS1 A-D, CsoS2 and CsoS4. Preferably, the expression of CsoS1 A-D, CsoS2 or CsoS4 is down-regulated. Preferably, the expression of CsoS1 A-D and CsoS2 is

down-regulated. Preferably, the expression of CsoS1 A-D and CsoS4 is down-regulated. Preferably, the expression of CsoS2 and CsoS4 is down-regulated. Preferably, the expression of CsoS1 A-D, CsoS2 and CsoS4 is down-regulated.

In bacteria that naturally express the beta-carboxysome, it is known that the formation of functional, closed BMCs requires at least the expression of the proteins CcmK2, CcmO and CcmL. Preferably, the expression of CcmK2, CcmO or CcmL is down-regulated. Preferably the expression of CcmK2 and CcmO is down-regulated. Preferably the expression of CcmK2 and CcmL is down-regulated. Preferably the expression of CcmO and CcmL is down-regulated. Preferably the expression of CcmK2, CcmO and CcmL is down-regulated. Preferably the expression of CcmK1, 3 and 4 are also down-regulated.

The glycyl radical enzyme(GRM)-associated bacterial microcompartments can vary in the number and type of shell proteins in the operon. These shell proteins are homologues to the shell proteins of the other BMC systems and belong to the same protein families. Like the Pdu BMC, the GRM BMC comprise s hexamers and pentamers, and the pentamers form the vertices of the BMC capsule. In bacteria that naturally express the GRM BMC, preferably the pentameric protein (Pfam03319) is down-regulated.

Preferably, if a microorganism of the invention is of a species that naturally expresses a particular BMCs type or types, then the microorganism of the invention comprises a deletion or deleterious mutation in a regulatory region of the operon for the production of said BMC type or types, preferably in the promoter. Preferably, if a microorganism of the invention is of a species that naturally expresses a particular BMCs type or types, then the microorganism of the invention comprises a deletion of the operon for the production of said BMC type or types. In other words, if a microorganism of the invention is of a species that naturally expresses a particular BMCs type or types, then the microorganism of the invention is preferably a BMC null mutant. A BMC null mutant is a microorganism that has been modified such that it is devoid of any endogenous genes for the production of BMCs.

As mentioned above, in nature, microorganisms that naturally express BMCs typically do so only under certain conditions, namely in the presence of the substrate for the pathway that comprises steps catalysed by enzymes located within the BMCs. For instance, Pdu BMCs are only expressed by microorganisms comprising the

necessary genes when said microorganisms are exposed to 1,2-propanediol. Similarly, Eut BMCs are only expressed by microorganisms comprising the necessary genes when said microorganisms are exposed to ethanolamine.

Thus, in an alternative embodiment, the microorganism of the invention is of a species or strain that naturally expresses BMCs, i.e. that comprises the genes necessary for the expression of BMCs, but wherein said microorganism is in an environment, e.g. a culture medium, that does not permit expression of BMCs. In other words, preferably the microorganism is present in a culture medium that lacks the molecule(s) that induce(s) the expression of BMCs in the microorganism. If the microorganism naturally expresses Pdu BMCs, then said molecule is 1,2-propanediol. If the microorganism expresses Eut BMCs, then said molecule is ethanolamine. The use of such limited culture media may remove the need for genetic modification of a microorganism to ensure lack of BMC expression.

In another aspect, the present invention provides a method of producing a genetically modified microorganism as described herein, said method comprising transforming a microorganism with one or more heterologous nucleic acid molecules together encoding at least three different proteins, each protein comprising an enzymatic domain and a bacterial microcompartment-targeting signal polypeptide, wherein said enzymatic domains catalyse different substrate to product conversions in the same metabolic pathway, and wherein said microorganism is essentially free of BMCs

The features and embodiments described above in relation to the genetically modified microorganisms of the invention apply *mutatis mutandis* to the methods of producing a genetically modified microorganism disclosed herein. The microorganisms, one or more heterologous nucleic acid molecules, at least three proteins, enzymatic domains, BMC-targeting signal polypeptides, substrate to product conversions, products of interest, and BMCs are as defined above.

In said method, preferably the microorganism is transformed with one or more plasmids, vectors or transformation cassettes comprising said one or more nucleic acid molecules. Preferably, said vector is an expression vector, preferably also comprising a strong heterologous promoter operatively linked to said one or more nucleic acid molecules.

If the microorganism is of a species or strain that naturally expresses BMCs, then the method comprises the step of genetically modifying the microorganism to inhibit essentially all BMC formation. Such steps are as described above. Alternatively, the method comprises culturing the microorganism only in an environment, e.g. a culture medium, that does not permit expression of BMCs, i.e. that lacks the molecule(s) that induce(s) the expression of BMCs in the microorganism.

Alternatively viewed, the present invention provides a method of producing a genetically modified microorganism as described herein, said method comprising over-expressing in a microorganism one or more heterologous nucleic acid molecules together encoding at least three different proteins, each protein comprising an enzymatic domain and a bacterial microcompartment-targeting signal polypeptide, wherein said enzymatic domains catalyse different substrate to product conversions in the same metabolic pathway, and wherein said microorganism is essentially free of BMCs

As explained above, preferably said microorganism is naturally free of BMCs, i.e. is of a strain that does not naturally express BMCs. In an alternative preferred embodiment however, the microorganism of the invention is of a strain that naturally expresses BMCs, therefore, the above methods preferably comprise a step of modifying the microorganism or its environment to inhibit its ability to produce said BMCs, as described above. Said inhibitions are preferably complete inhibitions.

Preferably, the step of modifying the microorganism to inhibit its ability to produce said BMCs comprises the step of down-regulating one or more of the proteins required for BMC formation. Preferably, the step of modifying the microorganism to inhibit its ability to produce said BMCs comprises the step of eliminating one or more of the proteins required for BMC formation. Preferably, said steps comprise deleting, interruption or deleteriously mutating one or more of the genes encoding a protein required for BMC formation that results in the reduction or elimination of expression of said protein. Alternatively, the method comprises silencing the one or more genes using a short DNA or RNA oligonucleotide that has a sequence complementary to either gene or an mRNA transcript, e.g. antisense oligonucleotides.

Preferably, if a microorganism of the invention is of a species that naturally expresses a particular BMCs type or types, then the microorganism of the invention comprises a deletion or deleterious mutation in a regulatory region of the operon for

the production of said BMC type or types, preferably in the promoter. Preferably, if a microorganism of the invention is of a species that naturally expresses a particular BMCs type or types, then the microorganism of the invention comprises a deletion of the operon for the production of said BMC type or types. In other words, if a microorganism of the invention is of a species that naturally expresses a particular BMCs type or types, then the microorganism of the invention is preferably a BMC null mutant. A BMC null mutant is a microorganism that has been modified such that it is devoid of any endogenous genes for the production of BMCs.

The discussion of preferred down-regulated proteins in the context of the microorganisms of the invention applies *mutatis mutandis* to the methods of producing the microorganisms of the invention.

In another aspect, the present invention provides a method of producing a product of interest, said method comprising growing a genetically modified microorganism described herein under conditions wherein the product is produced and optionally recovering the product. Said methods comprise growing the genetically modified microorganism under conditions in which said at least three different proteins are expressed, and preferably wherein said proteins together form aggregates.

The methods of the present invention comprise growing, i.e. culturing, a microorganism of the present invention under conditions that produce the product of interest. Any such conditions can be used and the person of ordinary skill in the art will readily be able to select and optimise the conditions for their specific purposes. Typically, the microorganism will be grown in a culture medium. If the microorganism of the invention is a strain that naturally expresses BMCs, then preferably the method does not comprise the step of applying to the culture medium any inducer molecule(s) that induce the expression of said BMCs in said microorganism. Preferably, the method does not comprise the step of applying propanediol, ethanolamine, choline, fucose or rhamnose.

The methods of the present invention thus comprise fermentation. “Fermentation” as used herein is the bulk growth of microorganisms on a growth medium, with the aim of producing a specific product by a metabolic process. Although fermentation is optionally a process that converts sugar to acids, gases or alcohols, it is not limited to these substrates or products as used herein.

Typically, the microorganisms of the present invention are grown in fermentation media for production of a product of interest. Defined or synthetic growth media may also be used and the appropriate medium for growth of a particular microorganism will be known by one skilled in the art of microbiology or fermentation science.

Fermentation media for production of the products of interest are well-known in the art and the skilled person will readily be able to determine a suitable fermentation media for their specific purpose. An appropriate, or effective, fermentation medium refers to any medium in which a genetically modified microorganism of the present invention, when cultured, is capable of producing the product of interest. Such a medium is typically an aqueous medium comprising assimilable carbon, nitrogen and phosphate sources. Such a medium can also include appropriate salts, minerals, metals and other nutrients.

Carbon sources are well-known in the art and the skilled person will readily be able to determine the appropriate carbon source for the microorganism species being used and the product of interest. It is contemplated that the source of carbon utilized can encompass a wide variety of carbon containing substrates and will only be limited by the choice of organism. Sources of assimilable carbon which can be used in a suitable fermentation medium include, but are not limited to, sugars and their polymers, including, dextrin, sucrose, maltose, lactose, glucose, fructose, mannose, sorbose, arabinose and xylose; fatty acids; organic acids such as acetate; primary alcohols such as ethanol and n-propanol; and polyalcohols such as glycerol. Preferred carbon sources include monosaccharides, disaccharides, and trisaccharides. The most preferred carbon source is glucose or glycerol. The concentration of a carbon source in the fermentation medium should promote cell growth, but not be so high as to repress growth of the microorganism used.

Sources of assimilable nitrogen which can be used in a suitable fermentation medium include, but are not limited to, simple nitrogen sources, organic nitrogen sources and complex nitrogen sources. Such nitrogen sources include anhydrous ammonia, ammonium salts and substances of animal, vegetable and/or microbial origin. Suitable nitrogen sources include, but are not limited to, protein hydrolysates, microbial biomass hydrolysates, peptone, yeast extract, ammonium sulfate, urea, and amino acids.

The fermentation medium can contain other compounds such as inorganic salts, vitamins, trace metals or growth promoters. Such other compounds can also be present in carbon, nitrogen or mineral sources in the effective medium or can be added specifically to the medium. The fermentation medium can also contain a suitable phosphate source. Such phosphate sources include both inorganic and organic phosphate sources. Preferred phosphate sources include, but are not limited to, phosphate salts such as mono or dibasic sodium and potassium phosphates, ammonium phosphate and mixtures thereof. A suitable fermentation medium can also include a source of magnesium, preferably in the form of a physiologically acceptable salt, such as magnesium sulfate heptahydrate, although other magnesium sources in concentrations which contribute similar amounts of magnesium can be used.

The fermentation medium can also include a biologically acceptable chelating agent, such as the dihydrate of trisodium citrate. The fermentation medium can also include a biologically acceptable calcium source, including, but not limited to, calcium chloride. The fermentation medium can also include sodium chloride. Preferably, the culture medium lacks the molecule(s) that induce(s) the expression of BMCs in the microorganism. Alternatively viewed, the present invention provides a culture medium comprising a microorganism of the invention, wherein said culture medium lacks the molecule(s) that induce(s) the expression of BMCs in the microorganism. Such molecules are described above. Preferably, the culture media lacks one or more, preferably all of propanediol, ethanolamine, choline, fucose and rhamnose.

The microorganisms of the invention described herein can be cultured using standard laboratory or industrial techniques known in the art. The growth of the microorganisms described herein can be measured by methods known in the art, for instance by measuring the optical density (OD) of cell cultures over time.

The temperature of the fermentation medium can be any temperature suitable for growth and production of the product of interest. For example, prior to inoculation of the fermentation medium with an inoculum, the fermentation medium can be brought to and maintained at a temperature in the range of from about 20°C to about 45°C, preferably to a temperature in the range of from about 25°C to about 40°C

The pH of the fermentation medium can be controlled by the addition of acid or base to the fermentation medium. In such cases when ammonia is used to control pH, it

also conveniently serves as a nitrogen source in the fermentation medium. Preferably, the pH is maintained from about 3.0 to about 9.0, more preferably from about 4 to about 8.0, still more preferably from about 6.5 to about 7.5, most preferably about 7.4.

Fermentations can be performed under aerobic or anaerobic conditions. The fermentation medium can also be maintained to have a dissolved oxygen content during the course of fermentation to maintain cell growth and to maintain cell metabolism for production of the product of interest. The oxygen concentration of the fermentation medium can be monitored using known methods, such as through the use of an oxygen electrode. Oxygen can be added to the fermentation medium using methods known in the art, for example through agitation and aeration of the medium by stirring or shaking. Preferably, the oxygen concentration in the fermentation medium is in the range of from about 20% to about 100% of the saturation value of oxygen in the medium based upon the solubility of oxygen in the fermentation medium at atmospheric pressure and at a temperature in the range of from about 20°C to about 40°C. Periodic drops in the oxygen concentration below this range may occur during fermentation, however, without adversely affecting the fermentation.

Although aeration of the medium has been described herein in relation to the use of air, other sources of oxygen can be used. Particularly useful is the use of an aerating gas which contains a volume fraction of oxygen greater than the volume fraction of oxygen in ambient air. In addition, such aerating gases can include other gases which do not negatively affect the fermentation.

Although the carbon source concentration can be maintained within desired levels by addition of, for example, a substantially pure glucose solution, it is acceptable, and may be preferred, to maintain the carbon source concentration of the fermentation medium by addition of aliquots of the original fermentation medium. The use of aliquots of the original fermentation medium may be desirable because the concentrations of other nutrients in the medium (e.g. the nitrogen and phosphate sources) can be maintained simultaneously.

The amount of product in the fermentation medium can be determined using a number of methods known in the art, for example, high performance liquid chromatography (HPLC) or gas chromatography (GC).

A batch method of fermentation can be used with the microorganisms described herein. A classical batch fermentation is a closed system where the composition of the medium is set at the beginning of the fermentation and not subject to artificial alterations during the fermentation. Thus, at the beginning of the fermentation the medium is inoculated with the desired organism or organisms, and fermentation is permitted to occur without adding anything to the system. Typically, however, a "batch" fermentation is batch with respect to the addition of carbon source and attempts are often made at controlling factors such as pH and oxygen concentration. In batch systems the metabolite and biomass compositions of the system change constantly up to the time the fermentation is stopped. Within batch cultures cells progress through a static lag phase to a high growth log phase and finally to a stationary phase where growth rate is diminished or halted. If untreated, cells in the stationary phase will eventually die. Cells in log phase generally are responsible for the bulk of production of end product or intermediate.

A Fed Batch system can also be used with the microorganisms described herein. A Fed Batch system is similar to a typical batch system with the exception that the carbon source substrate is added in increments as the fermentation progresses. Fed Batch systems are useful when catabolite repression (e.g. glucose repression) is apt to inhibit the metabolism of the cells and where it is desirable to have limited amounts of substrate in the media. Measurement of the actual substrate concentration in Fed Batch systems is difficult and is therefore estimated on the basis of the changes of measurable factors such as pH, dissolved oxygen and the partial pressure of waste gases such as CO_2. Batch and Fed Batch fermentations are common and well known in the art.

Although a batch mode can be performed, it is also contemplated that continuous fermentation methods could also be performed with the microorganisms described herein. Continuous fermentation is an open system where a defined fermentation medium is added continuously to a bioreactor and an equal amount of conditioned media is removed simultaneously for processing. Continuous fermentation generally maintains the cultures at a constant high density where cells are primarily in log phase growth.

Continuous fermentation allows for the modulation of one factor or any number of factors that affect cell growth or end product concentration. For example, one

method will maintain a limiting nutrient such as the carbon source or nitrogen level at a fixed rate and allow all other parameters to vary. In other systems a number of factors affecting growth can be altered continuously while the cell concentration, measured by media turbidity, is kept constant. Continuous systems strive to maintain steady state growth conditions and thus the cell loss due to the medium being drawn off must be balanced against the cell growth rate in the fermentation. Methods of modulating nutrients and growth factors for continuous fermentation processes as well as techniques for maximizing the rate of product formation are well known in the art of industrial microbiology.

It is contemplated that the present invention can be practiced using either batch, fed batch or continuous processes and that any known mode of fermentation would be suitable. Additionally, it is contemplated that cells can be immobilized on a substrate as whole cell catalysts and subjected to fermentation conditions for production.

Products can be isolated from the fermentation medium by methods known to one skilled in the art. For instance, solids may be removed from the fermentation medium by centrifugation, filtration, decantation, or the like. Products of interest in solution may be isolated from the fermentation medium using methods such as distillation, azeotropic distillation, liquid-liquid extraction, adsorption, gas stripping, membrane evaporation, pervaporation or vacuum flash fermentation.

In another aspect, the present invention comprises a cell free system, said system comprising aggregates comprising at least three different proteins, each protein comprising an enzymatic domain and a bacterial microcompartment-targeting signal polypeptide, wherein said enzymatic domains each catalyse a different substrate to product conversion in the same metabolic pathway, and wherein said system does not comprise bacterial microcompartments.

The system is suitable for the production of a product of interest, wherein said enzymatic domains each catalyse a different substrate to product conversion in the same metabolic pathway for the production of said product of interest.

The features and embodiments described above in relation to the genetically modified microorganisms of the invention apply *mutatis mutandis* to the cell free systems disclosed herein. The microorganisms, one or more heterologous nucleic acid molecules, at least three proteins, enzymatic domains, BMC-targeting signal

polypeptides, substrate to product conversions, products of interest, and BMCs are as defined above.

The conditions of the cell free system preferably mimic *in vivo* conditions. It would be within the competencies of one of ordinary skill in the art to modify the conditions in the cell free system to suit their particular purposes and the particular proteins employed. Preferably the pH of the system is one at which the enzymatic domains can function and the proteins remain aggregated. Preferably, the system comprises a buffer. Preferably, the pH of the system is between 5 and 9. Suitable buffers are well-known in the art. For example, a potassium phosphate buffer (pH 8) may be used, for instance at a concentration of about 100 mM.

Depending on the proteins, particularly the enzymatic domains present in the system, the system may or may not comprise a salt solution. Preferably, the cell free system comprises said aggregates and an aqueous solution, preferably a salt solution, for instance comprising NaCl and/or $MgCl_2$. Optionally, the NaCl is present at a concentration of about 100 mM. Optionally the $MgCl_2$ is present at a concentration of about 2.5 mM. If present, the solution is preferably an aqueous solution.

Preferably, the system comprises the co-factors necessary for the catalytic activity of the enzymatic domains of the at least three proteins in the system, for instance NADH, for instance at a concentration of about 0.1 mM. The skilled person will be aware of the appropriate cofactors for any particular enzymatic domain. Preferably, the system comprises ATP, for instance at a concentration of about 1mM.

Preferably, the system comprises potassium phosphate buffer (pH 8), NaCl, NADH, MgCl2, McCl2 and ATP. Preferably, the system comprises about 100mM potassium phosphate buffer (pH 8), about 100 mM NaCl, about 0.1 mM NADH, about 2.5 mM MgCl2, about 0.1 mM McCl2 and about 1 mM ATP.

Preferably, the system comprises cell lysate obtained from a population of microorganisms of the present invention. Preferably, the cell free system is prepared by culturing the microorganism of the invention, suspending the cultured cells in an aqueous solution and lysing them to result in a cell lysate. The cell lysate will comprise the desired protein aggregates since the aggregates will have formed in the microorganism during the culturing step.

Optionally, one or more purification steps can be performed to remove unwanted cellular fractions and/or to selectively isolate the aggregates. For instance, exclusion fractionation/chromatography may be used and the aggregates isolated by centrifugation. Various methods of cellular fractionation are well known in the art, as are methods for the isolation of proteins of interest including those in aggregated form. Suitable methods are disclosed, for instance, in Principles and Techniques of Practical Biochemistry (Wilson, K. & Walker, J) Cambridge University Press 5th Ed. Purification methods are disclosed, for instance, in Rodríguez-Carmona *et al.* (2010) *Microbial Cell Factories* 9:71.

It will be within the competencies of one of ordinary skill in the art to adjust the composition and conditions of the cell free system for his/her intended purpose and depending on the nature of the at least three proteins therein. For instance, the skilled person will be aware of how the pH of a cell free system can be adjusted to ensure that the aggregated proteins remain aggregated and the enzymatic domains remain functional. Similar considerations will be given to the temperature and salt concentration of the system.

The enzymatic domains present within the aggregated proteins in the cell free system each catalyse a different substrate to product conversion in the same metabolic pathway, i.e. in a pathway for the production of a product of interest. Preferably, the cell free system therefore comprises a suitable level of substrate(s) for the production of the product of interest.

In another aspect, the present invention provides a method of producing a product of interest, said method comprising:

i) providing a cell free system comprising aggregates comprising at least three different proteins, each protein comprising an enzymatic domain and a bacterial microcompartment-targeting signal polypeptide, wherein said enzymatic domains each catalyse a different substrate to product conversion in the same metabolic pathway for the production of the product of interest, and wherein said system does not comprise bacterial microcompartments;

ii) applying to said system the substrate of the first substrate to product conversion in the metabolic pathway that is catalysed by one of said enzymatic domains; and

iii) optionally recovering the product of interest.

The present invention provides a method of producing a product of interest, said method comprising:

i) growing a genetically modified microorganism described herein under conditions wherein said at least three proteins are expressed, preferably over-expressed, and wherein said proteins together form aggregates,

ii) obtaining said aggregates from said microorganisms, and

iii) using said aggregates in a cell free system for the production of said product of interest, under conditions wherein the product is produced;

and optionally recovering the product.

The cell free system is as described herein. Preferably, the step of obtaining said aggregates comprises lysing the microorganisms and obtaining the cell lysate, as described above.

As used herein, the singular forms "a," "an" and "the" include plural references unless the content clearly dictates otherwise. Thus, for example, reference to "a cell" includes a mixture of two or more cells.

The invention will now be further described in the following non-limiting Examples and the Figures in which:

Fig. 1 illustrates a pathway for the synthesis of 1,2-propanediol from glycerol. Glycerol dehydrogenase and dihydroxyacetone kinase catalyse the conversion of glycerol to dihydroxyacetone phosphate, via the intermedia dihydroxyacetone. Methylglyoxal synthase catalyses the conversion of dihydroxyacetone phosphate to methylglyoxal. Glycerol dehydrogenase and 1,2-propanediol oxidoreductase catalyse the conversion of methylglyoxal to 1,2-propanediol, via the intermediate lactaldehyde.

Fig. 2 shows the specific activity of the enzymes involved in the microbial synthesis of 1,2-propanediol: (a) glycerol dehydrogenase (b) dihydroxyacetone kinase (c) methylglyoxal synthase (d) 1,2-propanediol oxidoreductase, when untagged (i.e. not linked to a BMC-targeting signal sequence), when tagged with the BMC-targeting signal sequence D18 and when tagged with the BMC-targeting signal sequence P18.

Fig. 3 provides a statistical analysis showing the percentage of cells that contan inclusion bodies when said cells express either untagged GldA, DhaK, MgsA or FucO, or P18- or D18-tagged versions of said proteins.

Fig. 4 shows the results of a protease protection assay of GFP fused to a C-terminal proteolysis tag (SsrA) and an N-terminal tag being either a BMC targeting signal peptide (P18 or D18) or a non-targeting His-tag, in the presence and absence of BMCs (AU). *E. coli* competent cells were transformed with plasmids encoding the protein fusions with and without shell proteins and the resulting strains were cultured for 24 hours, samples were taken and run on a 15% denaturing polyacrylamide gel. Gels were subsequently submitted to Western blotting using an anti-GFP primary antibody. Total lysates were analysed by SDS-PAGE and subsequently western blotted with an anti-GFP primary antibody. Cell densities were normalised to an OD_{600} = 2.5 for loading of samples.

Fig. 5 shows *in vivo* 1,2-propanediol production. The graph shows the 1,2-propanediol content (normalised to OD_{600} = 1) over 96 h in the growth medium of strains that lack shell proteins and 1,2-propanediol producing enzymes (●),Shell proteins only (○), un-tagged 1,2-propanediol producing enzymes (▼), 1,2-propanediol producing enzymes tagged with targeting sequences (△), un-tagged 1,2-propanediol producing enzymes and shell proteins (■),1,2-propanediol producing enzymes tagged with targeting sequences and shell proteins (□). Data points represent an average of three independent experiments; standard deviations are represented by error bars.

Fig. 6 shows *in vivo* 1,2-propanediol production. The graph shows the 1,2-propanediol content (not normalised to OD_{600} = 1) over 96 h in the growth medium in of strains that lack shell proteins and 1,2-propanediol producing enzymes (●),Shell proteins only (○), un-tagged 1,2-propanediol producing enzymes (▼), 1,2-propanediol producing enzymes tagged with targeting sequences (△),un-tagged 1,2-propanediol producing enzymes and shell proteins (■),1,2-propanediol producing enzymes tagged with targeting sequences and shell proteins (□). Data points represent an average of three independent experiments; standard deviations are represented by error bars.

Fig. 7 shows thin sections of *E. coli* strains labelled with an anti-his antibody and then with a secondary antibody conjugated to 10 nm gold particles viewed under TEM (a) Wild type (b) Shell proteins only (c) 1,2-propanediol producing enzymes tagged with targeting sequences (d) 1,2-propanediol producing enzymes tagged with targeting

sequences and shell proteins (e) un-tagged 1,2-propanediol producing enzymes (f) un-tagged 1,2-propanediol producing enzymes and shell proteins.

Fig. 8 shows TEM analysis of strains expressing (A) His-tagged GldA (B) P18-tagged GldA (C) His-tagged DhaK (D) P18-tagged DhaK (E) His-tagged MgsA (F) D18-tagged MgsA (G) His-tagged FucO (H) D18-tagged FucO. Scale bar shows 0.2 μM.

Examples

In this study we are creating fusion proteins between Pdu targeting peptides and the four 1,2-propanediol producing enzymes to target the enzymes to recombinant Pdu microcompartment shells. We explore how the targeting peptides affect the activity of the different enzymes and their properties, particularly solubility. Strains are engineered for the targeting of all enzymes to microcompartments and compared for their 1,2-propanediol production to strains containing the native enzymes and also to a strain containing enzymes with targeting peptides but no shell proteins. The protein solubility of these strains is investigated by TEM analysis and protein aggregation is found to play an unexpected but important role in the efficiency of our pathway. Finally, we propose an alternative pathway engineering approach alongside compartmentalisation in protein shells.

Materials and methods

Strains

The strains used in this study are shown in Table A below:

Table A: Strains used in this study

Strain	Genotype	Source
BL21*(DE3)	F– ompT hsdSB (rB– mB–) gal dcm (DE3)	Novagen
BL21*(DE3) pLysS	F- ompT hsdSB(rB- mB-) gal dcm (DE3) pLysS (CamR)	Novagen

The BL21*(DE3) strain comprises genes encoding Eut BMCs. To ensure that Eut BMCs were not produced by the microorganisms during this study, ethanolamine was not included in any fermentation media. The absence of BMCs was confirmed by TEM.

Plasmid construction

Plasmids were constructed to include each of the genes of interest with an N-terminal tag comprising a BMC-targeting signal polypeptide ("P18" or "D18") and/or a hexa-histidine tag.

The genes of interest were the four enzymes of the metabolic pathway for the production of 1,2-propanediol from glycerol, as outlined in Fig. 1, i.e. glycerol dehydrogenase (GldA), dihydroxyacetone kinase (DhaK), methylglyoxal synthase (MgsA) and 1,2-propanediol oxidoreductase (FucO).

The BMC-targeting signal polypeptide-containing tags used in the study were as follows:

"D18": MEINEKLLRQIIEDVLSEPMGSSHHHHHHSSGLVPRGSH (SEQ ID NO: 64) (N-terminal 18 amino acids of PduD from *Citrobacter freundii* (the BMC-targeting signal polypeptide) followed by flexible linker PMGSS, 6-his linker, flexible linker SSGL, thrombin cleavage site LVPRGS and amino acid linker H)

"P18": MNTSELETLIRNILSEQLAMGSSHHHHHHSSGLVPRGSH (SEQ ID NO: 65) (N-terminal 18 amino acids of PduP from *Citrobacter freundii* (the BMC-targeting signal polypeptide) followed by flexible linker AMGSS, 6-his linker, flexible linker SSGL, thrombin cleavage site LVPRGS and amino acid linker H)

All primers used in this study are listed in Table B below.

Table B: Oligonucleotide primers used in this study, restriction sites are underlined

Name	Sequence 5' – 3'
GldA_NdeI_FW	CATCATATGGACCGCATTATTCAATCACC (SEQ ID NO: 66)
GldA_SpeI_RV	CATACTAGTTTATTCCCACTCTTGCAGG (SEQ ID NO: 67)
dhaK_NdeI_FW	CGCCATATGTCTCAATTCTTTTTTAACCAACGCACC (SEQ ID NO: 68)

dhaK_SpeI_RV	CATACTAGTTTAGCCCAGCTCACTCTCCGC (SEQ ID NO: 69)
mgsA_NdeI_FW	CATCATATGGAACTGACGACTCGCACTTTACC (SEQ ID NO: 70)
mgsA_SpeI_RV	CATACTAGTTTACTTCAGACGGTCCGCGAG (SEQ ID NO: 71)
fucO_NdeI_FW	CCGCATATGGCTAACAGAATGATTCTG (SEQ ID NO: 72)
fucO_SpeI_RV	CCTACTAGTTTACCAGGCGGTATGG (SEQ ID NO: 73)
GFP_NdeI_FW	GTACATATGAGCAAAGGAGAAGAACTTTTC (SEQ ID NO: 74)
GFP-SsrA_SpeI_RV	GACACTAGTTTAAGCTGCTAAAGCGTAGTTTTCGTCGTT TGCTGCTTTGTACAGCTCATCCATGCC (SEQ ID NO: 75)

All genes were amplified with flanking *Nde*I and *Spe*I restriction sites and each was ligated into pET14b, pET14b-D18 and pET14b-P18 vectors using *Nde*I and *Spe*I restriction sites.

Plasmids pML-1 to pML-6 as outlined in Table C, were constructed by a 'Link and Lock' approach utilizing the compatible sticky ends formed by digestion with *Xba*I and *Spe*I (McGoldrick *et al.*, (2005) *J Biol Chem* 14:1086-1094).

Table C: Plasmids used in this study

Plasmid name	**Genotype**	**Description**	**Source**
pET14b	pET14b	Overexpression vector containing N-terminal polyhistidine-tag	Novagen
pET14b-D18	pET14b-D18	Overexpression vector containing an N-terminal D18 targeting tag and an N-terminal polyhistidine-tag	This study
pET14b-P18	pET14b-P18	Overexpression vector containing an N-terminal	This study

		P18 targeting tag and an N-terminal polyhistidine-tag	
pLysS	PlysS	Overexpression vector	Novagen
pLysS-PduABB'JKNU	pLysS-PduABB'JKNU	Construct for expression of empty Pdu BMC	Parsons *et al.*, 2010
pET14b-gldA	pET14b-gldA	PCR product of gldA ligated into NdeI/SpeI sites of pET14b	This study
pET14b-dhaK	pET14b-dhaK	PCR product of dhaK ligated into NdeI/SpeI sites of pET14b	This study
pET14b-mgsA	pET14b-mgsA	PCR product of mgsA ligated into NdeI/SpeI sites of pET14b	This study
pET14b-fucO	pET14b-fucO	PCR product of fucO ligated into NdeI/SpeI sites of pET14b	This study
pET14b-GFP-SsrA	pET14b-GFP-SsrA	PCR product of gfp-ssrA ligated into NdeI/SpeI sites of pET14b	This study
pET14b-D18-gldA	pET14b-D18-His-gldA	NdeI/SpeI fragment of pET14b-gldA ligated into NdeI/SpeI sites of pET14b-D18	This study
pET14b-D18-dhaK	pET14b-D18-His-dhaK	NdeI/SpeI fragment of pET14b-dhaK ligated into NdeI/SpeI sites of pET14b-D18	This study
pET14b-D18-mgsA	pET14b-D18-His-mgsA	NdeI/SpeI fragment of pET14b-mgsA ligated into NdeI/SpeI sites of pET14b-D18	This study
pET14b-D18-fucO	pET14b-D18-His-fucO	NdeI/SpeI fragment of pET14b-fucO ligated into NdeI/SpeI sites of	This study

		pET14b-D18	
pET14b-D18-GFP-SsrA	pET14b-D18-His-GFP-SsrA	NdeI/SpeI fragment of pET14b-GFP-SsrA ligated into NdeI/SpeI sites of pET14b-D18	This study
pET14b-P18-gldA	pET14b-P18-His-gldA	NdeI/SpeI fragment of pET14b-gldA ligated into NdeI/SpeI sites of pET14b-P18	This study
pET14b-P18-dhaK	pET14b-P18-His-dhaK	NdeI/SpeI fragment of pET14b-dhaK ligated into NdeI/SpeI sites of pET14b-P18	This study
pET14b-P18-mgsA	pET14b-P18-His-mgsA	NdeI/SpeI fragment of pET14b-mgsA ligated into NdeI/SpeI sites of pET14b-P18	This study
pET14b-P18-fucO	pET14b-P18-His-fucO	NdeI/SpeI fragment of pET14b-fucO ligated into NdeI/SpeI sites of pET14b-P18	This study
pET14b-P18-GFP-SsrA	pET14b-P18-His-GFP-SsrA	NdeI/SpeI fragment of pET14b-GFP-SsrA ligated into NdeI/SpeI sites of pET14b-P18	This study
pML-1	pET14b-His-gldA-His-fucO	XbaI/EcoRI fragment from pET14b-His-fucO ligated into XbaI/EcoRI sites of pET14b-His-gldA	This study
pML-2	pET14b-P18-gldA-D18-fucO	XbaI/EcoRI fragment from pET14b-D18-His-fucO ligated into SpeI/EcoRI sites of pET14b-P18-His-gldA	This study
pML-3	pET14b-His-dhaK-	XbaI/HindIII fragment	This study

	His-mgsA	from pET14b-His-mgsA ligated into SpeI/HindIII sites of pET14b-His-dhaK	
pML-4	pET14b-P18-dhaK-D18-mgsA	XbaI/HindIII fragment from pET14b-D18-His-mgsA ligated into SpeI/HindIII sites of pET14b-P18-His-dhaK	This study
pML-5	pET14b-His-dhaK-His-mgsA-His-gldA-His-fucO	XbaI/ClaI fragment from pML-3 ligated into SpeI/ClaI sites of pML-1	This study
pML-6	pET14b-P18-dhaK-D18-mgsA-P18-gldA-D18-fucO	XbaI/ClaI fragment from pML-4 ligated into SpeI/ClaI sites of pML-2	This study

Overexpression and purification of recombinant protein

BL21 * (DE3) pLysS competent cells were transformed with a plasmid containing the gene(s) of interest. 1L of LB supplemented with ampicillin (100mg/L) in baffled flasks was inoculated from an overnight starter culture. The cultures were grown at 37 °C with shaking for 7 hours; protein production was induced by the addition of IPTG to a final concentration of 400 µM. The cultures were then incubated overnight at 19 °C with shaking. Cells were harvested by centrifugation at 3320 xg for 15 minutes at 4 °C, pellets were resuspended in 20 mM Tris-HCl, pH 8.0, 500 mM NaCl, 5 mM Imidazole. Cells were lysed by sonication and cell debris removed by centrifugation. Recombinant protein was then purified from the soluble fraction by immobilized metal ion affinity chromatography.

Activity assays

Glycerol dehydrogenase

The activity of GldA for the oxidation of glycerol to dihydroxyacetone was measured by following the initial rate at 340 nm for the reduction of NAD+ to NADH. Activity assays were carried out in 1 ml reactions containing 0.1 M potassium phosphate buffer pH 8.0, 500 µM NAD+, 2 mM $MgCl_2$ and 200 nM GldA. The activity of GldA for

the reduction methylglyoxal to lactaldehyde, was measured by following the initial rate of the oxidation of NADH to NAD+ at 340 nm. Activity assays were carried out in 1 ml reactions containing 0.1 M potassium phosphate buffer pH 8.0, 0.1 mM NADH, 2 mM MgCl2, 200 nM GldA.

Dihydroxyacetone kinase

The activity of DhaK for the conversion of dihydroxyacetone to dihydroxyacetone phosphate was measured in a coupled reaction with Glyceraldehyde 3-phosphate dehydrogenase (G3PDH) by following the oxidation of NADH to NAD+ at 340 nm. Activity assays were carried out in 1ml reactions containing 50 mM Tris-HCl, 100 mM NaCl,1 mM ATP, 0.1 mM NADH, 2.5 mM $MgCl_2$, 7.2 U G3PDH, 125 nM DhaK.

Methylglyoxal synthase

The activity of MgsA was monitored in a colorimetric assay over a time course. 25 µl 0.5 mM MgsA was incubated in a reaction mixture containing 400 µl 50 mM imidazole pH 7.0, 25 µl 15 mM dihydroxyacetone phosphate, 50 µl dH2O, the reaction mixture was incubated at 30 °C with shaking. At time intervals 50 µl of the reaction mixture was removed and added to a detection mixture containing 450 µl dH2O, 165 µl 0.1% 2,4-Dinitrophenylhydrazine hydrochloric acid solution. The detection mixture was incubated at 30 °C with shaking for 15 minutes. 835 µl of 10% (w/v) NaOH was added to the detection mixture which was incubated at room temperature for 15 minutes. Absorbances were then measured at 550 nm.

1,2-Proanediol oxidoreductase

The activity of FucO was determined for the NADH dependant reduction of glycolaldehyde to elthylene glycol was measured by following the initial rate of the oxidation of NADH to NAD+ at 340 nm. Activity assays were carried out in 1 ml reactions containing 100 mM Hepes, 10 µM NADH, 100 µM MnCl2, 200 nM FucO.

Embedding of strains for TEM analysis

50 ml of LB was inoculated with one colony and grown at 37 °C with shaking to an OD600 of ~ 0.4, cells were harvested by centrifugation at 3000 x g for 10 minutes. The cell pellet was resuspended in 2 ml 2.5% Glutaraldehyde in 100 mM cacodylate pH 7.2 and incubated for 2 hours with gentle spinning. Cells were pelleted by centrifugation at 6000 x g for 2 minutes and were washed twice with 100 mM cacodylate pH 7.2. Cells were stained with 1% osmium tetroxide in 100 mM cacodylate pH 7.2 for 2 hours and subsequently washed twice with dH2O. Cells were

dehydrated by incubation in an ethanol gradient, 50% EtOH for 10 minutes, 70% EtOH overnight followed by two 10 minute washes in 100% EtOH. Cells were then washed twice with propylene oxide for 15 minutes. Cell pellets were embedded by resuspention in 1 ml of a 1:1 mix of propylene oxide and Agar LV Resin and incubated for 30 minutes with spinning. Cell pellets were washed twice in 100% Agar LV resin. The cell pellet was resuspended in fresh resin and transferred to a 0.5 ml mould, centrifuged for 5 minutes at 3000 x g to concentrate the cells to the tip of the mould and incubated for 16 hours at 60 °C to polymerise.

Sectioning and visualisation of samples

Samples were thin sectioned on a RMC MT-XL ultramicrotome with a diamond knife (diatome 45°) sections were placed on 300 mesh copper grids. Grids were stained by incubation in 4.5% uranyl acetate in 1% acetic acid solution for 1 hour followed by 2 washes in dH_2O. Grids were then stained with 0.1% lead citrate for 8 minutes followed by a wash in ddH_2O

Electron microscopy was performed using a JEOL-1230 transmission electron microscope.

Culture medium and conditions for 1,2-propanediol production

The culture medium designed by Neidhardt *et al.*, 1974 was supplemented with 30 g/L glycerol, 10 g/L tryptone, 5 g/L yeast extract and appropriate antibiotics. Strains were cultured in sealed serum bottles with a working volume of 100 ml at 28 °C with shaking. Cultures were inoculated from starter cultures to starting OD_{600} of 0.05. During growth 1 ml samples were removed at 0, 6, 12, 24, 48, 72 and 96 hours.

Western blot analysis

Nitrocellulose membranes following transfer and blocking were incubated in primary antibody (mouse anti-GFP) followed by incubation in a secondary coupled antibody (Anti-mouse IgG AP). Bands were visualised by incubation in substrate 5-Bromo-4-chloro-3-indolyl phosphate/Nitro blue tetrazolium (BCIP/NBT).

Analysis of 1,2-propanediol production

In-vivo 1,2-propanediol production was determined by GC/MS analysis of the growth medium at time intervals (0, 6, 12, 24, 48, 72 and 96 hours). The supernatant after centrifugation, was boiled for 10 minutes at 100 °C followed by centrifugation at 19,750 x g. The sample was then acidified with trifluoroacetic acid to a final concentration of 0.01% followed by a second centrifugation at 19,750 x g. The

supernatant following centrifugation was diluted 1:4 in acetonitrile for GC/MS analysis.

Visualisation of engineered strains

Embedding of strains for immunolabelling

Strains were grown as described previously ("Culture medium and conditions for 1,2-propanediol production", above) overnight, cells were harvested by centrifugation for 10 minutes at 3000 x g. The cell pellet was resuspended in 2% formaldehyde, 0.5% gluteraldehyde in 100 mM sodium cacodolate buffer pH 7.2 and incubated for 2 hours with gentle spinning. Cells were pelleted by centrifugation at 6000 x g for 2 minutes and were washed twice with 100 mM sodium cacodylate pH 7.2. Cells were dehydrated by incubation in an ethanol gradient, 50% EtOH for 10 minutes, 70% EtOH for 10 minutes, 90% EtOH for 10 minutes, followed by three 15 minute washes in 100% EtOH. Cell pellets were then resuspended in 2 ml LR white resin and incubated overnight with spinning at room temperature after which the resin was changed and incubated for a further 6 hours. Cell pellets were resuspended in fresh resin and transferred to 1 ml embedding tubes and centrifuged at 4000 x g to pellet the cells at the tip and incubated for 24 hours at 60° C to polymerize.

Samples were thin sectioned on a RMC MT-XL ultramicrotome with a diamond knife (diatome 45°) sections were placed on 300 mesh gold grids.

Immunolabeling of sections

Grids were equilibrated in one drop of TBST (20 mM Tris-HCl pH 7.2, 500 mM NaCl, 0.05% Tween (RTM) 20, 0.1% BSA) before being transferred into a drop of 2% BSA in TBST and incubated at room temperature for 30 minutes. Grids were then immediately transferred into primary antibody (Anti-His) and incubated for 1 hour. Grids were washed in a fresh drop of TBST followed by washing in a stream of TBST. Grids were equilibrated in a drop of secondary antibody (Goat anti-mouse IgG 10nm) then incubated for 30 minutes in a fresh drop. Excess antibody was removed by washing in two drops of TBST before washing in a stream of ddH_2O and dried.

Staining

Grids were stained for 15 minutes in 4.5% uranyl acetate in 1% acetic acid solution followed by 2 washes in dH_2O. Grids were then stained with 0.1% lead citrate for 3 minutes followed by a wash in ddH_2O.

Electron microscopy was performed using a JEOL-1230 transmission electron microscope.

Results

Effect of fusing enzymes to BMC-targeting peptides on enzyme specific activities

The effect of fusing either of the two targeting peptides P18 and D18 to heterologous enzymes on the functionality of those enzyme had not previously been investigated in detail. In this study the enzymes involved in the microbial synthesis of 1,2-propanediol from glycerol, namely glycerol dehydrogenase (GldA), dihydroxy acetone kinase (DhaK), methylglyoxal synthase (MgsA) and 1,2-propanediol oxidoreductase (FucO) were cloned with both N-terminal targeting peptides (P18 or D18) followed by a hexa-histidine tag. Proteins of interest were purified by IMAC and the kinetic parameters of each of the protein fusions were subsequently determined and compared to enzymes containing only the N-terminal hexa-histidine tag. In this Example, the term “tagged proteins” refers to P18-his and D18-his containing proteins, while the his-only containing proteins as referred to as “untagged” proteins.

It was found that the targeting peptides effect the specific activities of some of the proteins studied (Fig. 2). For instance, tagging GldA (the enzyme that catalyses the oxidation of glycerol) with the D18 targeting peptide resulted in a reduction of its specific activity by 90% compared to un-tagged GldA. Tagging GldA with the P18 targeting peptide reduced the activity of the protein by approximately half (55% reduction) compared to un-tagged GldA (Fig. 2A). GldA’s ability to reduce methylglyoxal to lactaldehyde was similarly affected, with D18 having the greatest negative effect (83% reduction compared to un-tagged GldA) and P18 causing a loss of 53% of specific activity compared to untagged GldA. The activity of DhaK for the ATP dependant phosphorylation of dihydroxyacetone phosphate was determined by a coupled reaction involving a second enzyme, glyceraldehyde 3-phosphate dehydrogenase, in excess. In contrast to GldA, kinetic analysis of DhaK fused with either a P18 or D18 targeting peptide had no significant effect on the enzyme’s activity (Fig. 2B).

Tagging MgsA with either a P18 or D18 targeting peptide had a negative effect on enzyme activity, reducing the activity by 18% and 15% respectively in comparison to untagged MgsA, as shown in Fig. 2C. When fused to D18, FucO’s specific activity

decreased by 58% in comparison to untagged FucO and when fused to P18, FucO's specific activity decreased by 76% in comparison to untagged FucO (Fig. 2D).

It is concluded that the fusion of targeting peptides to the N-termini of proteins is likely to have an effect on the specific activities of a significant proportion of said proteins. Without wishing to be bound by theory, the inventors consider this is most likely due to changes in structural and chemical properties and potential changes in protein folding as a result of the fusion.

Targeting peptides cause protein aggregation that can be visualised by TEM

The production levels and solubility of GldA, DhaK, MgsA and FucO with and without targeting peptides fused thereto were investigated by subjecting samples of the purification process, including the soluble and insoluble fractions after clarification of the crude cell lysate as well as the final purified protein samples, to denaturing polyacrylamide gel analysis (data not shown). DhaK and MgsA were well produced and soluble irrespective of the presence of a P18 or D18 tag. The solubility of FucO was not affected by targeting peptides, but the yield of un-tagged FucO appeared slightly lower compared to FucO containing the targeting peptides. In contrast, although, both P18 and D18-tagged GldA appeared to be produced, the protein bands were predominantly detected in the insoluble fractions of the SDS gels.

This suggests that the fusion proteins D18-GldA and P18-GldA were aggregating compared to GldA. P18-GldA was also found to be eluted from the IMAC column with an additional band of smaller molecular weight, indicative of protein degradation.

In order to investigate the aggregation behaviour of the tagged proteins further, the most active protein fusions (P18-GldA, P18-DhaK, D18-MgsA and D18-FucO), the candidates for the construction of the1,2-propanediol production pathway targeted to microcompartments, were chosen to be visualised by TEM.

Strains encoding each of the tagged proteins and strains encoding the un-tagged proteins were cultured overnight without induction. Subsequently, the cells were harvested, embedded in low viscosity resin, thin sectioned and visualized using TEM. For each strain 100 cells were examined for protein aggregation, statistical analysis of each of the strains is shown in Fig. 3 and representative TEM micrographs were compiled (Fig. 8).

Control strains producing un-tagged proteins (GldA, DhaK, MgsA, FucO) displayed a 'normal' phenotype, with only 1% of observed cells containing electron dense areas indicative of aggregated proteins. In contrast, half of all observed cells (52%) producing P18-GldA showed protein aggregates located at the pole of the cells (Fig. 8). The addition of the P18 targeting peptide to the N-terminus of DhaK resulted in aggregate formation in 8% of the observed cells. Fusion of the D18 targeting peptide to MgsA and FucO resulted in the presence of protein aggregates in 12% and 4% of cells respectively.

These results confirm that the fusions between the enzymes of the 1,2-propanediol production pathway and targeting peptides cause protein aggregation.

Enzymes fused to BMC-targeting peptides are recruited to BMCs

It was investigated whether fusion proteins comprising an enzyme of interest fused to either the P18 or D18 targeting peptide were targeted to bacterial microcompartments.

Strains co-expressing the individual genes of the 1,2-propanediol pathway (*P18-gldA*, P18-mgsA, D18-dhaK, D18-fucO) and the construct for empty shell formation (pLysS-PduABB'JKNU) were cultured and the recombinant microcompartments were purified as described previosly (Lawrence *et al*., (2014) ACS Synth. Biol. 3: 454-465). Samples were taken throughout the purification and analysed on 15% denaturing polyacrylamide gels for the protein profile. Analysis of the resulting SDS-PAGE gels reveals that tagging each of the proteins with a targeting peptide facilitates their co-purification with the microcompartment proteins. This was further confirmed by kinetic assays of the final purified BMC fraction.

Further evidence of protein targeting to microcompartments was provided by a protease protection assay that was previously reported by Sargent *et al*., (2013) Microbiology 159: 2427-2436.

Plasmids were constructed containing GFP fused to an N-terminal P18 or D18 tag and a C-terminal SsrA proteolysis tag (AANDENYALAA*). The C-terminal SsrA tag targets proteins for degradation by the *E. coli* proteases ClpAP and ClpXP. *E. coli* competent cells were transformed with plasmids encoding the protein fusions with and without shell proteins and the resulting strains were cultured for 24 hours, samples were taken and run on a 15% denaturing polyacrylamide gel, adjusted to cell number as determined by OD_{600} measurements. Gels were subsequently submitted to Western blotting using an anti-GFP primary antibody.

The results show that the co-expression of GFP-SsrA fused to targeting peptides and produced with shell proteins have the highest amount of GFP (Fig. 4 lane 7 + 8). In the absence of a targeting peptide GFP is effectively degraded as represented by only a faint band present in lane 6 of Fig. 4. In the absence of shell proteins, all GFP fusion proteins are present to a much lesser extent than fusion proteins in the presence of microcompartments. The faint bands seen could be a result of protein aggregation, which would protect the GFP fusions from proteolytic cleavage. The band seen in lane 1 of Fig. 4 (shell only) most likely represents unspecific binding of the antibody. The difference seen in band intensity between lanes 3 and 6 of Fig. 4 is likely due to differences in expression levels as a result of the co-expression of shell proteins.

These results are consistent with microcompartments providing protection from cytosolic proteases for proteins internalised therein.

<u>Construction of 1,2-propanediol producing strains and comparative analysis of bacterial growth</u>

For the *in vivo* production of 1,2-propanediol, single plasmids were engineered by using link and lock cloning combining firstly the genes coding for the most active protein fusions (pML-6 containing *P18-his-gldA, P18-his-dhaK, D18-his-mgsA, D18-his-fucO*) and secondly the same genes but without targeting sequences (pML-5 containing *his-gldA, his-dhaK, his-mgsA, his-fucO).* Both plasmids were used to transform the *E. coli* strain BL21*(DE3).

With the aim of targeting the 1,2-propanediol producing enzymes to recombinant microcompartments, strains were engineered to co-express the 1,2-propanediol production plasmids with the genes coding for the protein shell (pLysS-

PduABB'JKNU). The shell protein construct allows for the formation of a microcompartment shell to which the fusion enzymes are recruited by virtue of their BMC-targeting peptide. Additionally, the following control strains were set up: firstly BL21*(DE3) transformed with pET14b and pLysS; and secondly, a shell only strain transformed with pET14b and pLysS-PduABB'JKNU. All strains were compared for the production of 1,2-propanediol.

The culture medium designed by Neidhardt *et al.*, 1974 (Neidhardt FC, Bloch PL, Smith DF. Culture Medium for Enterobacteria. Journal of Bacteriology. 1974;119(3):736-747) was supplemented with 30 g/L glycerol, 10 g/L tryptone and 5 g/L yeast extract and appropriate antibiotics. Strains were cultured in sealed serum bottles with a working volume of 100 ml at 28 °C with shaking The cultures were started with an initial OD_{600} of 0.05 by inoculation from 5 ml starter cultures. During growth, 1 ml samples were collected at 0, 6, 12, 24, 48, 72 and 96 hours and optical densities at 600 nm were measured. The resulting growth curves (not shown) indicate that strains encoding proteins with targeting peptides (either with shell proteins or without) grow slower and reach a lower final optical density in comparison to strains expressing un-tagged proteins and control strains. Furthermore, cell densities declined from 24 hours in strains producing un-tagged enzymes whereas the cell densities of the strains with tagged enzymes remained constant, which indicates that cells with tagged enzymes cells are being protected from a toxic intermediate.

In vivo 1,2-propanediol production is elevated in strains producing enzymes with targeting sequences

The 1,2-propanediol content in the growth media of the various strains was quantified by gas chromatography-mass spectrometry (GC-MS). Whole cell samples were collected at 0, 6, 12, 24, 48, 72 and 96 hours and the supernatant following centrifugation was prepared for GC-MS analysis as described in materials and methods. The measured 1,2-propanediol content as shown in Fig. 5 is expressed for a cell density of OD_{600} = 1. Fig. 6 shows the measured 1,2-propanediol content not adjusted for a cell density of OD_{600} = 1.

Strains encoding un-tagged 1,2-propanediol producing enzymes with and without shell proteins showed low 1,2-propanediol production despite growing well and reaching the highest cell densities at 96 hours. Both strains reached the maximum

product concentration (at 96 hours) of 3.59 mM / OD_{600} = 1 in the absence of shell proteins and 1.95 mM / OD_{600} = 1 in the presence of shell proteins.

The highest product concentrations were detected in the growth media of strains producing proteins tagged with targeting peptides. Although both of these strains (with and without shell proteins) grew to lower density than the strains harbouring un-tagged proteins, and despite the negative effect the targeting peptide has on the specific activities of the individual enzymes, they produced significantly more 1,2-propanediol (Fig. 5) than the strains lacking the targeting peptides. The strain containing tagged 1,2-propanediol enzymes and shell proteins reached a final yield of 7.10 mM / OD_{600} = 1. However, the highest final yield of 11.56 mM/ OD_{600} = 1 was observed when the shell proteins were not present. 1,2-propanediol was not detected in control stains (wild type *E. coli* and a strain producing shell proteins only).

A comparison of Fig.5 and Fig.6 shows that the increase in 1,2-propanediol production levels is not simply an effect of differences in cell densities i.e. even though the strains with tagged enzymes don't grow as well as the strains without tagged enzymes, they still produce more 1,2-propanediol.

The higher product yield exhibited by the strain producing tagged enzymes in the absence of shell proteins was unexpected.

To investigate if aggregation of the proteins were causing this effect, electron microscopy and immunolabelling were used to visualise the subcellular organisation and location of the recombinant proteins in the various strains. Sections of the strains were labelled with anti-histidine primary antibody designed to bind to the hexa-histidine tag on the N-terminus of proteins in our pathway thereby revealing the intracellular location. A secondary antibody conjugated to 10 nm gold particles was used to bind to the primary antibody thereby revealing the intracellular location of 1,2-propanediol producing enzymes.

Control strains (wild type and shell only) showed a small amount of antibody binding around the membrane of the cells (Fig. 7A and 7B); this is likely due to unspecific binding. Aggregates are visible in approximately 100 % of observed cells expressing P18/D18-tagged proteins, regardless of the presence or absence of shell proteins, and it is in these areas that the vast majority of antibody binding occurs (Fig. 6C and D). Such structures cannot be seen in cells expressing un-tagged enzymes (Fig. 6E

and F), suggesting that it is the presence of the targeting peptides that facilitates the aggregation of proteins.

It is concluded that aggregation occurs due to tagging proteins of interest with targeting peptides and it is this aggregation that results in a significant increase in product yield despite the reduction in specific activities of the individual tagged pathway enzymes.

It has been determined that the fusion of BMC-targeting peptides to the individual enzymes in the pathway for the production of 1,2-propanediol lowers the specific activities of the enzymes in some (most) cases. The solubility of each enzyme was also affected to varying degrees, with GldA forming large inclusion bodies in the majority of cells observed by TEM when fused with a targeting peptide compared to un-tagged GldA. It has also been demonstrated that the addition of a BMC-targeting tag recruited the enzymes to BMCs and that purified samples thereof remained metabolically active.

Despite the significant decrease of enzyme activity seen with the addition of targeting peptides to both GldA and FucO, expression of the complete tagged pathway enzymes led *to* an increase in product formation as compared to strains in which untagged enzymes were expressed. Rather unexpectedly, the presence of the microcompartment shell was not required for the increased product formation and, furthermore, the strain generating the most 1,2-propanediol produced tagged 1,2-propanediol pathway enzymes, but no shell proteins. This strain showed an increase in product formation of 245% OD-adjusted in comparison to the strain producing un-tagged enzymes; despite the lower *in vitro* activity of the individual tagged proteins compared to the un-tagged proteins.

TEM analysis showed that co-production of all four tagged enzymes resulted in protein aggregation and deposition at the poles of nearly all cells observed and it is this aggregation that appears to provide a significant benefit to the efficiency of the pathway. Aggregation of our proteins of interest is likely due to the amphipathic helical nature of the BMC-targeting sequences and/or by their coiled coil structure. Without wishing to be bound by theory, it can be considered that the aggregation creates a scaffolding effect that result in increased channelling of substrates and products between enzymes, similarly to the environment inside a microcompartment.

This study is the first to demonstrate that the presence of short targeting peptides can not only convert individual fusion proteins but also whole pathways into active aggregates that allow for increased product yield *in vivo*. These aggregations of multiple enzymes allow for increased localised concentrations of enzymes and intermediates and possibly channelling between them thereby resulting in a higher product yield. This is the first study to demonstrate that increased product yields can result from tagging enzymes in a metabolic pathway for the production of said product with a BMC targeting sequence in a cell lacking BMCs themselves.

CLAIMS

1. A genetically modified microorganism comprising one or more heterologous nucleic acid molecules together encoding at least three different proteins, each protein comprising an enzymatic domain and a bacterial microcompartment-targeting signal polypeptide, wherein said enzymatic domains each catalyse a different substrate to product conversion in the same metabolic pathway, and wherein said microorganism is essentially free of bacterial microcompartments (BMCs).

2. The genetically modified microorganism of claim 1, wherein said microorganism comprises one or more heterologous nucleic acid molecules together encoding at least four different proteins, each protein comprising an enzymatic domain and a bacterial microcompartment-targeting signal polypeptide, wherein said enzymatic domains each catalyse a different substrate to product conversion in the same metabolic pathway.

3. The genetically modified microorganism of claim 1 or claim 2, wherein each of said heterologous nucleic acid molecules encodes only one of said proteins.

4. The genetically modified microorganism of any one of claims 1 to 3, wherein each protein is over-expressed relative to the level of expression of said protein in a microorganism of the same strain which lacks said one or more heterologous nucleic acid molecules.

5. The genetically modified microorganism of any one of claims 1 to 4, wherein said microorganism comprises aggregates comprising said at least three proteins.

6. The genetically modified microorganism of any one of claims 1 to 5, wherein each protein comprises an enzymatic domain and a BMC-targeting signal polypeptide linked by an amino acid linker, preferably wherein said amino acid linker lacks stable secondary structure.

7. The genetically modified microorganism of any one of claims 1 to 6, wherein each BMC-targeting signal polypeptide comprises an amphipathic alpha helix.

8. The genetically modified microorganism of any one of claims 1 to 7, wherein each BMC-targeting signal polypeptide comprises the following amino acid sequence:

$$X_1X_2X_3X_4X_5X_6X_7X_8X_9$$

wherein:

X_1, X_4, X_5, X_8, and X_9 are each independently hydrophobic amino acids selected from the group consisting of I, L, V, M, F, Y, A and W;

X_2, X_3 and X_6 are each independently polar or charged amino acids selected from the group consisting of Q, N, T, S, C, D, E, R, K and H; and

X_7 is any amino acid.

9. The genetically modified microorganism of any one of claims 1 to 8, wherein each BMC-targeting signal polypeptide comprises a sequence selected from the group consisting of LEQIIRDVL (SEQ ID NO:1), LETLIRTIL (SEQ ID NO:2), LETLIRNIL (SEQ ID NO:3), LRQIIEDVL (SEQ ID NO:4), IEEIVRSVM (SEQ ID NO:5), IEQVVKAVL (SEQ ID NO:6), VEKLVRQAI (SEQ ID NO:7), IQEIVRTLI (SEQ ID NO:8), VEEIVKRIM (SEQ ID NO:9), IESMVRDVL (SEQ ID NO:10), VQDIIKNVV (SEQ ID NO:11), IRQVVQEVL (SEQ ID NO:12), VRSVVEEVV (SEQ ID NO:13) and ARDLLKQIL (SEQ ID NO:14) or a variant sequence thereof.

10. The genetically modified microorganism of any one of claims 1 to 9, wherein said microorganism is a bacterium.

11. The genetically modified microorganism of any one of claims 1 to 10, wherein said microorganism does not naturally comprise the genes necessary for the expression of BMCs.

12. The genetically modified microorganism of any one of claims 1 to 10, wherein said microorganism is a strain that naturally comprises the genes necessary for the expression of BMCs, and wherein expression of said BMCs is inducible by the presence of one or more inducer molecules, but wherein said microorganism is present in a culture medium in which the level of said inducer molecule(s) is too low to induce the expression of said BMCs.

13. The genetically modified microorganism of any one of claims 1 to 10, wherein said microorganism is a strain that naturally comprises the genes necessary for the

expression of BMCs but wherein said microorganism comprises a loss of function mutation in one or more of said genes.

14. The genetically modified microorganism of claim 13, wherein said microorganism comprises a loss of function mutation in at least one gene encoding a protein comprising a BMC domain and in at least one gene encoding a protein comprising a bacterial microcompartment vertex domain.

15. The genetically modified microorganism of claim 13, wherein said microorganism comprises a mutation in the regulatory region of the BMC operon(s), preferably in the operon's promoter.

16. A method of producing a product of interest, said method comprising growing the genetically modified microorganism of any one of claims 1 to 15 under conditions wherein the product is produced and optionally recovering the product.

17. A cell free system comprising aggregates comprising at least three different proteins, each protein comprising an enzymatic domain and a bacterial microcompartment-targeting signal polypeptide, wherein said enzymatic domains each catalyse a different substrate to product conversion in the same metabolic pathway, and wherein said system does not comprise bacterial microcompartments.

18. A method of producing a product of interest, said method comprising:

iv) providing a cell free system comprising aggregates comprising at least three different proteins, each protein comprising an enzymatic domain and a bacterial microcompartment-targeting signal polypeptide, wherein said enzymatic domains each catalyse a different substrate to product conversion in the same metabolic pathway for the production of the product of interest, and wherein said system does not comprise bacterial microcompartments;

v) applying to said system the substrate of the first substrate to product conversion in the metabolic pathway that is catalysed by one of said enzymatic domains; and

vi) optionally recovering the product of interest.

19. The method of claim 18, wherein said aggregates are obtained from a cell lysate of the microorganism of any one of claims 1 to 15.

Fig. 1

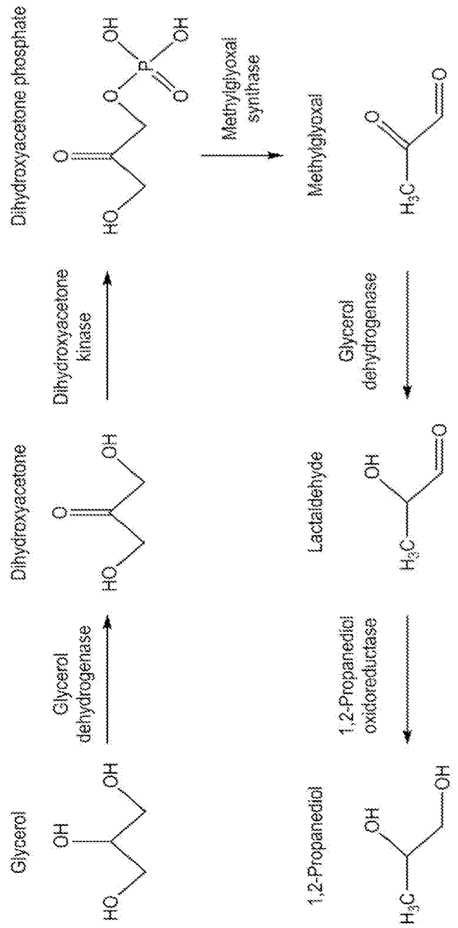

Glycerol
Glycerol dehydrogenase
Dihydroxyacetone
Dihydroxyacetone kinase
Dihydroxyacetone phosphate
Methylglyoxal synthase
Methylglyoxal
Glycerol dehydrogenase
Lactaldehyde
1,2-Propanediol oxidoreductase
1,2-Propanediol

Fig. 2

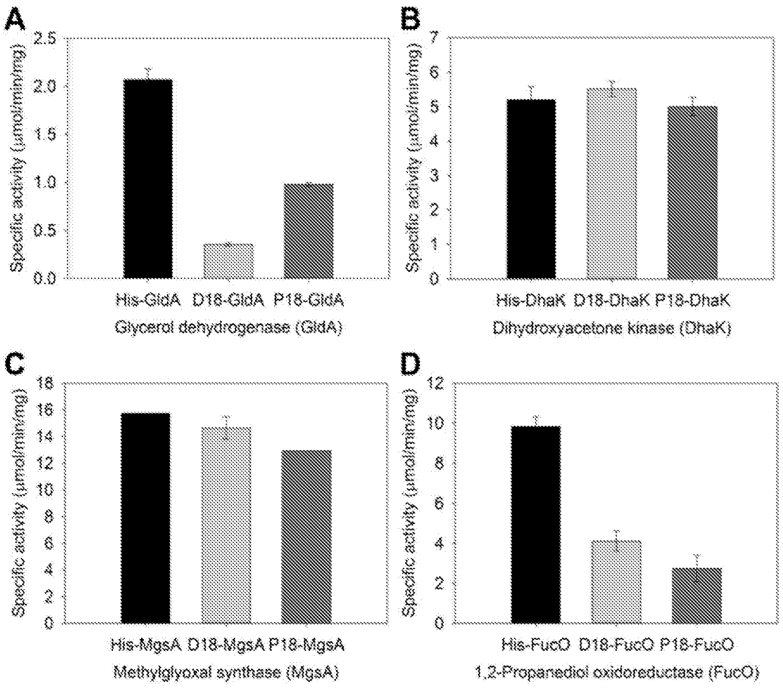

A
Specific activity (μmol/min/mg)
2.5
2.0
1.5
1.0
0.5
0.0
His-GldA D18-GldA P18-GldA
Glycerol dehydrogenase (GldA)
B
Specific activity (μmol/min/mg)
7
6
5
4
3
2
1
0
His-DhaK D18-DhaK P18-DhaK
Dihydroxyacetone kinase (DhaK)
C
Specific activity (μmol/min/mg)
18
16
14
12
10
8
6
4
2
0
His-MgsA D18-MgsA P18-MgsA
Methylglyoxal synthase (MgsA)
D
Specific activity (μmol/min/mg)
12
10
8
6
4
2
0
His-FucO D18-FucO P18-FucO
1,2-Propanediol oxidoreductase (FucO)

Fig. 3

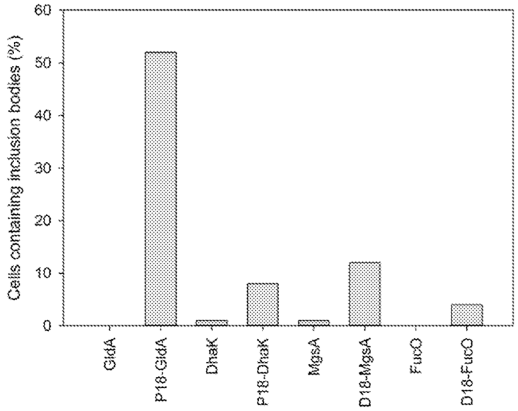

Cells containing inclusion bodies (%)
60
50
40
30
20
10
0
GldA
P18-GldA
DhaK
P18-DhaK
MgsA
D18-MgsA
FucO
D18-FucO

Fig. 4

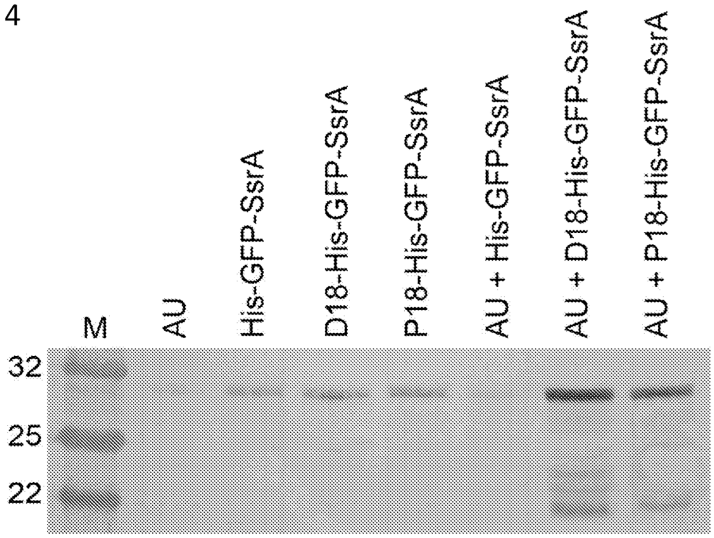

M
AU
His-GFP-SsrA
D18-His-GFP-SsrA
P18-His-GFP-SsrA
AU + His-GFP-SsrA
AU + D18-His-GFP-SsrA
AU + P18-His-GFP-SsrA
32
25
22

Fig. 5

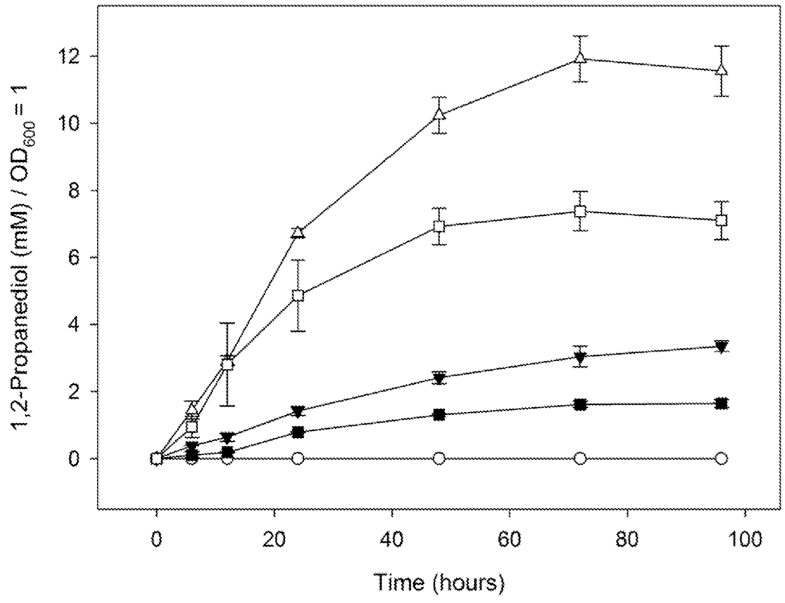

1,2-Propanediol (mM) / OD_{600} = 1
0
2
4
6
8
10
12
0
20
40
60
80
100
Time (hours)

Fig. 6

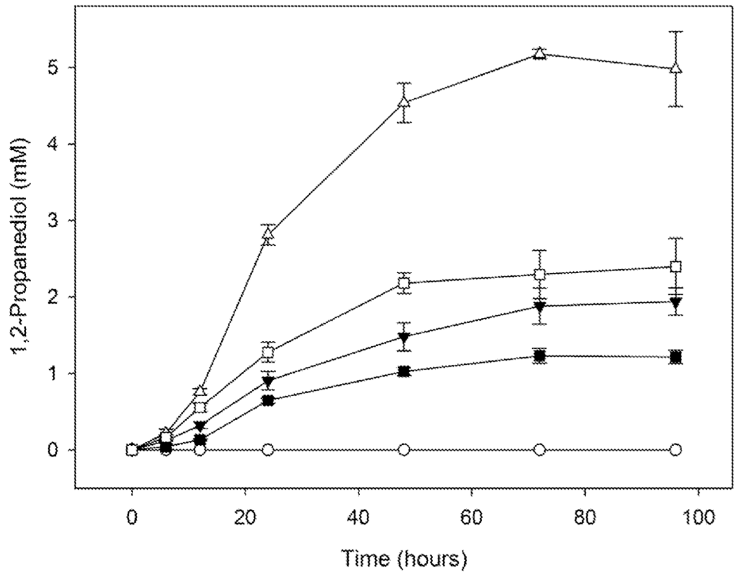


Fig. 7

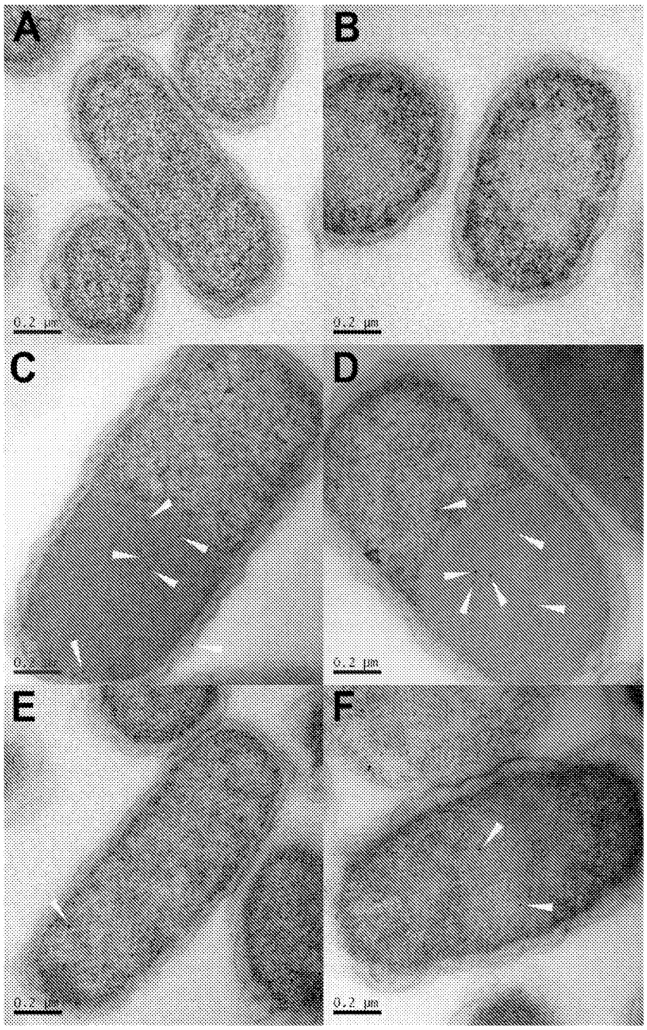
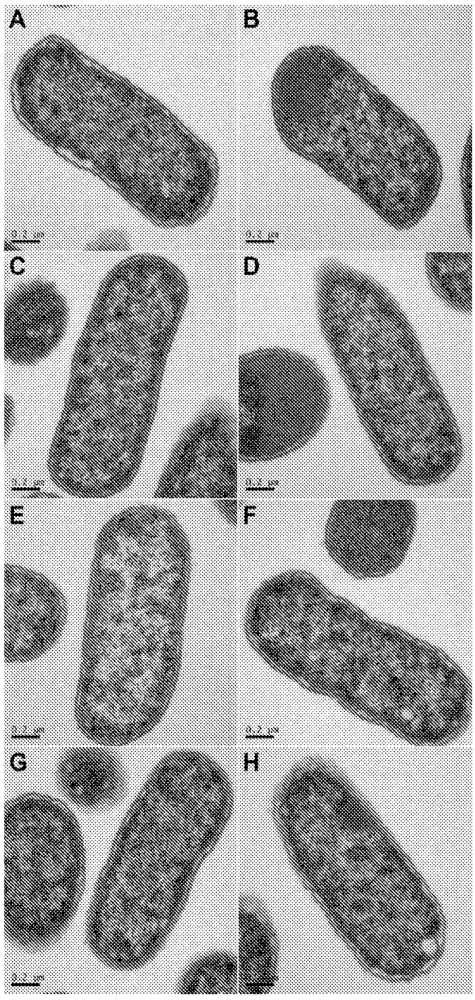


Fig. 8



6.3. Engineered synthetic scaffolds for organising proteins within bacterial cytoplasms

ARTICLE
PUBLISHED ONLINE: 11 DECEMBER 2017 | DOI: 10.1038/NCHEMBIO.2535

Engineered synthetic scaffolds for organizing proteins within the bacterial cytoplasm

Matthew J Lee[1], Judith Mantell[2,3], Lorna Hodgson[2], Dominic Alibhai[3], Jordan M Fletcher[4], Ian R Brown[1], Stefanie Frank[5], Wei-Feng Xue[1], Paul Verkade[2,3,6], Derek N Woolfson[2,4,6]* & Martin J Warren[1]*

We have developed a system for producing a supramolecular scaffold that permeates the entire *Escherichia coli* cytoplasm. This cytoscaffold is constructed from a three-component system comprising a bacterial microcompartment shell protein and two complementary *de novo* coiled-coil peptides. We show that other proteins can be targeted to this intracellular filamentous arrangement. Specifically, the enzymes pyruvate decarboxylase and alcohol dehydrogenase have been directed to the filaments, leading to enhanced ethanol production in these engineered bacterial cells compared to those that do not produce the scaffold. This is consistent with improved metabolic efficiency through enzyme colocation. Finally, the shell-protein scaffold can be directed to the inner membrane of the cell, demonstrating how synthetic cellular organization can be coupled with spatial optimization through in-cell protein design. The cytoscaffold has potential in the development of next-generation cell factories, wherein it could be used to organize enzyme pathways and metabolite transporters to enhance metabolic flux.

In industrial biotechnology and synthetic biology there is a growing need to generate internal supramolecular scaffolds in bacteria *en route* to delivering so-called cell factories[1]. To this end, researchers have investigated protein-based linkers[2], lipids[3] and nucleic acids[4,5] as modulators to attain high-level biomolecular organization. However, none of these approaches have delivered a uniform matrix throughout the bacterial cytoplasm. The advantage of such scaffolding systems is that they can be used to direct and align biosynthetic pathway enzymes to orchestrate greater production of commodity and specialty chemicals, especially in pathways that proceed through unstable or toxic intermediates[6,7]. This is because the close proximity of enzymes on a scaffold allows greater channeling of intermediates through improved flux, stabilization of intermediates and protection from other reactions[8,9].

A number of natural scaffolds are found in bacterial cells. For instance, bacterial microcompartments (BMCs) are organelles with an outer semipermeable scaffold in the form of a protein shell, which encases a specific metabolic pathway[10–12]. BMCs have a diameter of approximately 150 nm and possess high concentrations of internalized enzymes. This is most apparent in carboxysomes, which are anabolic BMCs wherein the high concentrations of carbonic anhydrase and ribulose-1,2-bisphosphate carboxylase/oxygenase (RuBisCO) ensure enhanced carbon fixation[13,14]. In catabolic BMCs, such as the metabolosome associated with propanediol utilization (the so-called Pdu system), internalized enzymes necessitate that propionaldehyde be rapidly transformed into either an alcohol or a CoA thioester, thereby protecting the cell from the potentially toxic aldehyde intermediate[15–17]. In all cases, the enzymes are targeted to the interior of the BMCs by small encapsulation peptides, which interact with a component of the outer-shell scaffold[18–21]. Modeling studies indicate that BMCs enhance flux through intermediate sequestration[22]. Recently, a detailed structure of a recombinant BMC shell has been reported showing the precise orientation of the different shell proteins that tile together to form the outer casing, providing molecular detail on how the shell proteins scaffold together to act as a semipermeable membrane[23].

Apart from BMCs, the other major scaffold within prokaryotic cells is the cytoskeleton[24], which is generally distributed around the inner membrane. This filamentous structure, which includes proteins such as FtsZ, MreB, ParM and MinD[25,26], has roles in cell division, cell morphology and structural polarity. However, the essential nature of these proteins precludes them from being developed as major cellular matrices. For these reasons, we sought to construct a simple and modular bacterial cytoskeleton, which we call a cytoscaffold, using components that we understand and can manipulate predictably.

Previously, we have shown that a single shell protein from the *Citrobacter freundii* Pdu BMC, with a minor modification to its C terminus to improve solubility (PduA*), forms filaments in *E. coli*[27]. PduA itself hexamerizes to form a tile that assembles to make the facets of the BMC casing[28,29]. However, when overproduced recombinantly in *E. coli*, PduA* forms hollow filaments approximately 20 nm in diameter that span the length of the cell[30]. Moreover, these structures often interfere with septation during cell division. Nonetheless, we reasoned that PduA*-based filamentous structures may present tractable scaffolds for tethering other proteins.

Here, we describe a three-component system comprising PduA* and two complementary *de novo*–designed coiled-coil peptides[31], which form an interactive intracellular filamentous arrangement that gives the appearance of a matrix permeating the entire *E. coli* cytoplasm (**Supplementary Results**, **Supplementary Fig. 1**). We show that other proteins can be specifically targeted to these cytoscaffolds. Building on this, we demonstrate that tethering metabolic enzymes for ethanol production to the PduA* scaffold increases their effective local and relative concentrations and results in improved ethanol production. Finally, we show that the scaffold can be directed

[1]Industrial Biotechnology Centre, School of Biosciences, University of Kent, Canterbury, UK. [2]School of Biochemistry, University of Bristol, Medical Sciences Building, University Walk, Bristol, UK. [3]Wolfson Bioimaging Facility, Medical Sciences Building, University Walk, Bristol, UK. [4]School of Chemistry, University of Bristol, Cantock's Close, Bristol, UK. [5]Department of Biochemical Engineering, University College London, Bernard Katz Building, Gordon Street, London, UK. [6]BrisSynBio, Life Sciences Building, Tyndall Avenue, Bristol, UK. *e-mail: M.J.Warren@kent.ac.uk or D.N.Woolfson@bristol.ac.uk

to the inner membrane of the cell, further illustrating its modularity, flexibility, and utility, and demonstrating how synthetic cellular organization can be coupled with spatial optimization.

RESULTS

Construction of a filamentous scaffold

Initially, we tested whether different proteins could be recruited to PduA* filaments *in vivo* in an analogous way to how encapsulation peptides are thought to work in natural BMCs (see above). Attempts to use the natural encapsulation peptides themselves were not very successful, mainly owing to aggregation within the cell[32]. Therefore, we turned to a better-characterized *de novo*–designed heterodimeric coiled-coil system, CC-Di-AB[31], which has been used successfully in the construction of self-assembling peptide cages[33]. The heterodimer comprises two peptides (acidic (A) and basic (B)) that do not self-associate, but interact specifically and tightly when mixed. The concept was to fuse either CC-Di-A or CC-Di-B to PduA* and then test whether a reporter protein with the cognate peptide can be targeted to the filaments. Plasmids encoding fusion proteins of the following type were made: CC-Di-A/B–Gly/Ser linker–HexaHisTag–PduA*, referred to as CC-Di-A–PduA* and CC-Di-B–PduA*. A control plasmid harboring the fusion without the CC-Di-A/B module, i.e., containing only the Gly/Ser linker–HexaHisTag (C–PduA*), was also made. Plasmids were transformed individually into *E. coli* cells, and the resulting strains were grown, induced and analyzed by transmission electron microscopy (TEM) after fixation, embedding, thin sectioning and staining.

Strains expressing PduA* alone generated parallel filaments spanning the length of the cell (**Fig. 1a**) and appear to interfere with septation (**Supplementary Fig. 2**). Unexpectedly, the strains producing the control C–PduA* and the CC-Di-A–PduA* did not form any filamentous structures (**Fig. 1b** and **Supplementary Fig. 3**). C–PduA* expression led to deposits of material at the poles of the cell, suggesting that the Gly/Ser linker and/or the hexahistidine tag alone affects solubility of the fusion protein. This was not seen with CC-Di-A–PduA*, but it is not clear why filaments do not form with this construct. In both cases, western blot analysis revealed only low levels of CC-Di-A–PduA* and C–PduA* relative to untagged PduA*, suggesting potential cytotoxicity of these proteins (**Supplementary Fig. 4**). In contrast, large amounts of CC-Di-B–PduA* were detected (**Supplementary Fig. 4**), and this led to numerous filaments throughout the cytoplasm (**Fig. 1c** and **Supplementary Fig. 5**). These filaments had a diameter of 23.6 ± 2.78 nm (n = 100), similar to those of untagged PduA* filaments[30], but the former were considerably shorter (**Fig. 1d**,**e**). Consequently, CC-Di-B–PduA* filaments were not aligned within cells, and they did not appear to disrupt cells. It is not clear why the CC-Di-B–PduA* filaments are shortened, but it is possible that the highly charged CC-Di-B peptide may limit filament growth in some way.

Characterization of the cytoscaffold

To probe the spatial localization and organization of the shorter CC-Di-B–PduA* filaments, thicker thin sections (250 nm) were cut and prepared for TEM tomography. Analysis of the resulting tomogram confirmed the presence of the shorter filaments throughout the cytoplasm, except in a central region that is largely occupied by genomic DNA (**Fig. 1f**). Using methods developed to track microtubule assemblies in cells[34], we rendered these structures and visualized them in three dimensions (**Fig. 1f** and **Supplementary Videos 1** and **2**). From this, it was clear that the filaments were not aligned, but arrayed with multiple orientations, resulting in the appearance of an internal matrix. Analysis of these filaments revealed an average length of 161.2 ± 102.4 nm (n = 739); though because of the limitation of a 250 nm thin section, the true length is likely longer than this (**Fig. 1e**).

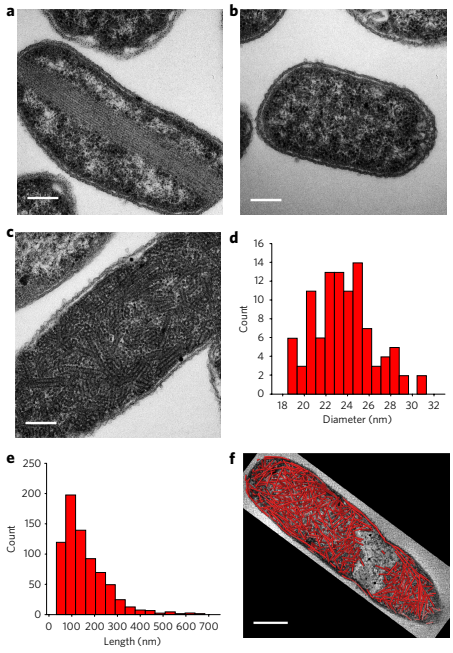


Figure 1 | Transmission electron micrographs and analysis of PduA*-based constructs and filaments in *E. coli*. (**a**) Untagged PduA* filaments. (**b**,**c**) CC-Di-A–PduA* (**b**) and CC-Di-B–PduA* (**c**) filaments. Scale bars in **a**–**c**, 200 nm. (**d**) Histogram showing the diameter of CC-Di-B-PduA* filaments (n = 100 measurements). (**e**) Histogram showing lengths of CC-Di-B-PduA* filaments based on a 250 nm tomogram rendering shown in **f** (n = 739 filaments). (**f**) 3D rendering of CC-Di-B–PduA* filaments in *E. coli* based on a 250 nm tomogram. Scale bar in **f**, 500 nm. See also **Supplementary Video 1**.

To test the robustness of the CC-Di-B–PduA* filaments, and to interrogate their structure in more detail, we purified the filaments from cells using protocols developed for BMC isolation[21]. Cells were lysed, and the filaments were purified by combining centrifugation and differential salt precipitation (**Supplementary Fig. 6**). Purified filaments were analyzed by TEM and atomic force microscopy (AFM). Both confirmed intact filaments, and these approaches provided the opportunity to gain greater insight into the molecular organization of these structures (**Supplementary Fig. 7**). These *ex vivo* filaments tended to cluster together on the TEM grids and AFM substrates. This clustering was also seen, though to a lesser extent, in some thin sections of whole cells visualized by TEM.

Targeting to and functionalization of the cytoscaffold

Next, we tested whether the CC-Di-B peptides of the CC-Di-B–PduA* filaments were accessible for targeting by other proteins labeled with CC-Di-A using the fluorescent protein Citrine. To do this, we made CC-Di-A–Citrine and C–Citrine constructs similar in design to the fusion proteins described above. By cloning these constructs in compatible plasmids, we could either transform these alone or co-transform them with the plasmid producing CC-Di-B–PduA*.

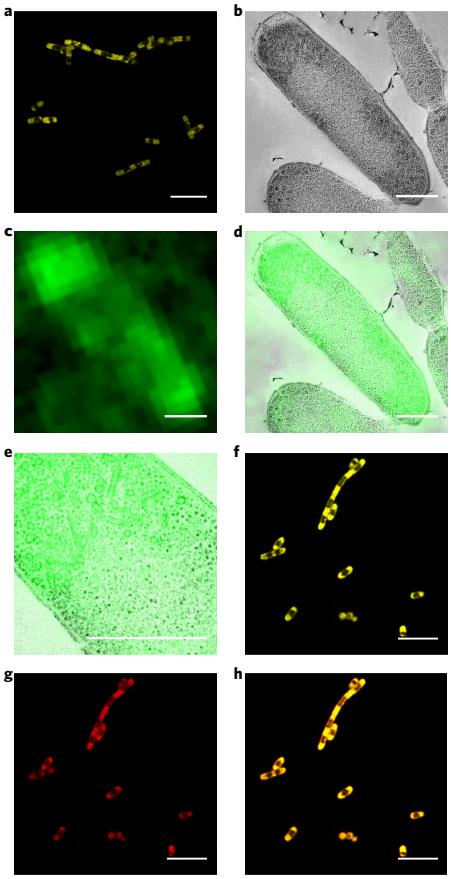


Figure 2 | Localization of fluorescent proteins to a bacterial cytoscaffold. (**a**) Co-expression of CC-Di-B-PduA* with CC-Di-A-Citrine, indicated by Citrine fluorescence. (**b**–**e**) Correlative light electron microscopy of a strain expressing CC-Di-B-PduA* and CC-Di-A-GFP, encompassing TEM (**b**), GFP fluorescence (**c**), overlay (**d**), and detail of the overlaid image (**e**). (**f**–**h**) Fluorescence of cells expressing CC-Di-B-PduA*, CC-Di-A-Citrine, and CC-Di-A-mCherry, including Citrine signal (**f**), mCherry signal (**g**), and overlay of the two (**h**). Scale bars in **a**,**f**–**h**, 5 µm; scale bars in **b**–**e**, 500 nm.

On their own, CC-Di-A–Citrine and C–Citrine each produced uniform fluorescence throughout the cells, consistent with soluble, cytoplasmic proteins (**Supplementary Fig. 8**). Similarly, when co-expressed with CC-Di-B–PduA*, the C–Citrine control produced fluorescence that was distributed throughout the cell. In contrast, co-expression of CC-Di-A–Citrine and CC-Di-B–PduA* produced more punctate fluorescence, as well as reduced fluorescence around the genomic DNA (**Fig. 2a**). This is consistent with CC-Di-A–Citrine being localized to the filamentous scaffold. Correlative light electron microscopy (CLEM)[35,36] of high-pressure frozen cells co-expressing

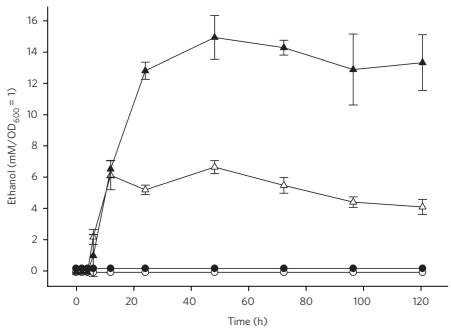


Figure 3 | Ethanol production *in vivo*. Graph showing ethanol content of the growth medium over time normalized to an OD_{600} = 1. Open circles, *E. coli* strain transformed with empty plasmids (pET14b and pLysS); closed circles, strain producing CC-Di-B-PduA* only; open triangles, strain producing CC-Di-A-Pdc and CC-Di-A-Adh; closed triangles, strain producing CC-Di-A-Pdc, CC-Di-A-Adh and CC-Di-B-PduA*. Data points represent an average of three independent experiments; error bars represent s.d.

CC-Di-A–GFP and CC-Di-B–PduA* confirmed the localization of fluorescence to the intracellular filamentous network (**Fig. 2b**–**e**). Control strains expressing the CC-Di-B–PduA* filaments with untagged GFP showed only a cytoplasmic signal (**Supplementary Fig. 9**). Expression of CC-Di-B–Citrine with or without CC-Di-B–PduA* resulted in punctate fluorescence, suggesting self-association of the CC-Di-B peptide (**Supplementary Fig. 8**).

To demonstrate that multiple cargo proteins can be directed to the cytoscaffold, we co-produced CC-Di-A–Citrine and a CC-Di-A–mCherry fusion in cells with CC-Di-B–PduA* filaments. This gave patterns similar to those observed with CC-Di-A–Citrine plus CC-Di-B–PduA*, and the mCherry signals colocalized with the Citrine signals (**Fig. 2f**–**h**).

Although the results clearly demonstrate that fluorescent proteins can be localized to the PduA* filaments through the use of the coiled-coil interaction, we also wanted to investigate whether enzymes could be pegged onto the CC-Di-B-PduA* filaments in a similar fashion. To explore this, we tagged both pyruvate decarboxylase (Pdc) and alcohol dehydrogenase (Adh) with the CC-Di-A peptide and co-expressed these with and without the CC-Di-B–PduA* filaments. Intriguingly, strains expressing Pdc and Adh grew to a substantially higher optical density at 600 nm (OD_{600}) in comparison to both control strains (**Supplementary Fig. 10a**). GC–MS analysis of the growth medium revealed that the introduction of the CC-Di-B–PduA* filamentous network increased ethanol production by 221% per OD unit (t = 120 h) in comparison to a strain expressing Pdc and Adh, but not the cytoscaffold (**Fig. 3** and **Supplementary Fig. 10b**). Western blot analysis showed that this increase in ethanol production was not due to increased protein expression: indeed, the levels of Pdc and Adh were actually reduced by 48 ± 13.3% and 26 ± 5.5%, respectively, in the strain expressing CC-Di-A-tagged enzymes in the presence of the CC-Di-B–PduA* filaments (**Supplementary Fig. 11**). The presence of filaments in these strains was confirmed by TEM analysis (**Supplementary Fig. 12**). These experiments provide strong evidence that the localization of enzymes onto the PduA* scaffold substantially enhances an engineered metabolic pathway.

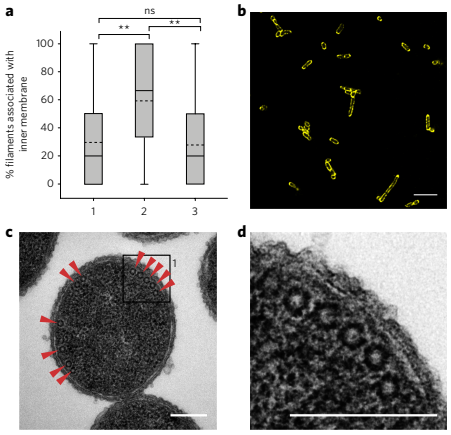


Figure 4 | Targeting the bacterial cytoscaffold to the inner membrane of *E. coli*. (**a**) Box-and-whisker plots showing the number of filaments associated with the inner membrane for three strains expressing variants of the CC-Di-A/B-PduA* system. Plot 1 = CC-Di-B-PduA*; plot 2 = CC-Di-B–PduA* + CC-Di-A–Citrine–MinD; plot 3 = CC-Di-B–PduA* + CC-Di-C–Citrine–MinD. Boxes show first and third quartiles, solid line shows median, dotted lines give the mean and whiskers the minimum and maximum; 250 cells were analyzed for each of the three strains. $^{**}P = 0.01$; ns, not significant. (**b**) Confocal image of strain expressing CC-Di-A–Citrine–MinD. Scale bar, 5 µm. (**c**) TEM micrograph of strains producing CC-Di-B–PduA* plus CC-Di-A–Citrine–MinD. Arrows indicate transverse filaments. (**d**) Zoom in of area 1 in **c** (inset black box). Scale bars in **c** and **d**, 200 nm.

As a final demonstration of the modularity, versatility and potential utility of the new cytoscaffold, we tested whether it could be directed to the cytoplasmic side of the inner membrane of *E. coli* (**Fig. 4**). For this, we added the C-terminal membrane-localizing region of MinD from *Bacillus subtilis* to the CC-Di-A–Citrine fusion to render CC-Di-A–Citrine–MinD[37]. When this construct was expressed in cells and imaged by confocal fluorescence microscopy, halos around the cytoplasm were evident, indicating localization of Citrine to the cell membrane (**Fig. 4b**). This was also the case for the C–Citrine–MinD control (**Supplementary Fig. 13**). When each of these were co-expressed with CC-Di-B–PduA* filaments, we observed differences in location between control and membrane-targeting constructs (**Fig. 4a** and **Supplementary Fig. 14**).

First, in cells expressing CC-Di-B–PduA* alone, an average of 30% of the filaments were associated with the membrane. For CC-Di-B–PduA* plus the C–Citrine–MinD control, this localization was very similar (31%). In contrast, for the CC-Di-B–PduA* plus CC-Di-A–Citrine–MinD combination, 60% of the filaments were localized to the inner membrane, and this difference was statistically significant ($P < 0.01$). One-way analysis of variance (ANOVA) showed no significant difference ($P < 0.01$) in the total number of filaments between the three strains. Collectively, these analyses demonstrated that the cellular spatial location of the CC-Di-B–PduA* filaments can be controlled by interactions with the cognate *de novo*–designed coiled-coil peptide.

DISCUSSION

Previously, we and others have shown that individual shell proteins, which form the hexameric tiles of the BMC casing, generate long, filamentous macromolecular structures when overproduced in host bacterial cells[30,38]. These structures are particularly apparent with PduA* from the Pdu BMC. The filaments formed by PduA* are approximately 20 nm in diameter; they can be several microns long and have a tendency to stack together and align along the length of the cell, to the extent that they interfere with cell septation. We hypothesized that the PduA* filaments can be formed from the self-association of the hexameric tiles into a protein sheet that then rolls into a nanotubule filament. We wondered whether it would be possible to target specific proteins to these filaments to generate higher-order supramolecular organization in the cell by design.

To achieve this, we have employed a heterodimeric coiled-coil system, CC-Di-A and CC-Di-B[31], previously characterized and used, for example, in the *de novo* construction of peptide cages[33]. We find that whereas fusion of the CC-Di-A sequence onto PduA* resulted in low protein production and loss of filament formation, attachment of CC-Di-B to PduA* leads to the formation of much shorter filaments that are dispersed throughout the cytoplasm. The reason for the shorter filaments is not clear, but could be because of a slight frustration of hexamer packing when the positively charged CC-Di-B peptide is appended or from faster nucleation of CC-Di-B–PduA* fusions, resulting in a greater number of shorter filaments. Importantly, given the quantity of filaments that are produced throughout the cell, their formation and presence does not appear to alter cell viability or growth.

By adding the complementary CC-Di-A peptide onto fluorescent proteins, we show through imaging techniques that these tagged proteins can be recruited to the CC-Di-B–PduA* filaments, demonstrating that the filaments can act as a molecular scaffold. A key biotechnological use of scaffolds within a cell would be localization of biosynthetic enzymes in close proximity to one another to facilitate metabolic channeling. This is part of the theory behind multienzyme complexes, although in these cases, direct transfer or channeling of metabolites from one enzyme to the next also takes place. Using simple systems it has been shown that compartmentalization of pyruvate decarboxylase and alcohol dehydrogenase within a recombinant BMC improves production of ethanol from pyruvate[21]. Similarly, the direct fusion of these two enzymes also results in improved flux[39], indicating in both cases that having the second enzyme in close proximity to the first ensures that the unstable acetaldehyde intermediate is more efficiently converted into the alcohol. Therefore, herein we have targeted pyruvate decarboxylase and alcohol dehydrogenase to the CC-Di-B-PduA* filaments using the coiled-coil modules, producing significantly more ethanol in comparison to when the enzymes are expressed without the scaffold. This provides very strong evidence that the cytoscaffold can be used to cluster metabolic enzymes to accelerate the channeling of intermediates from one enzyme to the next.

The CC-Di-B peptide can also be used to control the localization of the PduA* filaments within the cell. This has been achieved by targeting the CC-Di-A-Citrine protein to the inner membrane by fusing on the membrane-targeting region of MinD to the C terminus of the construct. Co-expression of this CC-Di-A–Citrine–MinD protein with CC-Di-B–PduA* directed filaments to the inner-membrane. Such localization strategies can be used to ensure that pathway-enriched filaments have ready access to metabolites that are taken up via transporters or, conversely, to ensure that products are generated near the membrane for export out of the cell.

Furthermore, the fact that the CC-Di-B–PduA* filaments are able to interact easily with either cytosolic proteins or membrane-targeted proteins containing the CC-Di-A peptide suggests that the N-terminal region of PduA must be solvent exposed. The two sides of the hexameric PduA tiles are distinguished by their concave and convex appearance. The N terminus of PduA, to which CC-Di-B is fused, is located on the concave side of the protein. The fact that this CC-Di-B peptide is available to interact with a CC-Di-A-tagged

protein strongly implies that the filament formed from PduA is generated with the concave side exposed to the solvent. This agrees with the recent structure of a recombinant BMC, in which all the shell proteins were found to be oriented with the concave side facing out of the structure[23].

Overall, this work demonstrates a concept for performing and evaluating rational protein design in the cell; specifically, making hybrid scaffolds comprising *de novo*–designed peptides and natural proteins that can be engineered on the micron scale within the *E. coli* cytoplasm. Visualization of filaments with and without appended ancillary proteins, and of those broadly distributed filaments or those localized to the inner membrane, demonstrates the potential of the system as a universal scaffold for the attachment, dispersion or localization of targeted cargo throughout the cell. We believe that these features and properties of the cytoscaffold, coupled with its ease of decoration and remodeling within cells, will enable applications in biotechnology and synthetic biology. More generally, this ability to design and engineer proteins in the cell could usher in a new era of rational protein design and engineering *in vivo*.

Received 31 May 2017; accepted 1 November 2017;
published online 11 December 2017

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

References

1. Polka, J.K., Hays, S.G. & Silver, P.A. Building spatial synthetic biology with compartments, scaffolds, and communities. *Cold Spring Harb. Perspect. Biol.* **8**, a024018 (2016).
2. Zhang, Y. *et al.* Using unnatural protein fusions to engineer resveratrol biosynthesis in yeast and Mammalian cells. *J. Am. Chem. Soc.* **128**, 13030–13031 (2006).
3. Grinkova, Y.V., Denisov, I.G. & Sligar, S.G. Engineering extended membrane scaffold proteins for self-assembly of soluble nanoscale lipid bilayers. *Protein Eng. Des. Sel.* **23**, 843–848 (2010).
4. Delebecque, C.J., Silver, P.A. & Lindner, A.B. Designing and using RNA scaffolds to assemble proteins in vivo. *Nat. Protoc.* **7**, 1797–1807 (2012).
5. Zalatan, J.G. *et al.* Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. *Cell* **160**, 339–350 (2015).
6. Agapakis, C.M., Boyle, P.M. & Silver, P.A. Natural strategies for the spatial optimization of metabolism in synthetic biology. *Nat. Chem. Biol.* **8**, 527–535 (2012).
7. Dueber, J.E. *et al.* Synthetic protein scaffolds provide modular control over metabolic flux. *Nat. Biotechnol.* **27**, 753–759 (2009).
8. Poshyvailo, L., von Lieres, E. & Kondrat, S. Does metabolite channeling accelerate enzyme-catalyzed cascade reactions? *PLoS One* **12**, e0172673 (2017).
9. Wheeldon, I. *et al.* Substrate channelling as an approach to cascade reactions. *Nat. Chem.* **8**, 299–309 (2016).
10. Chowdhury, C., Sinha, S., Chun, S., Yeates, T.O. & Bobik, T.A. Diverse bacterial microcompartment organelles. *Microbiol. Mol. Biol. Rev.* **78**, 438–468 (2014).
11. Frank, S., Lawrence, A.D., Prentice, M.B. & Warren, M.J. Bacterial microcompartments moving into a synthetic biological world. *J. Biotechnol.* **163**, 273–279 (2013).
12. Kerfeld, C.A. & Erbilgin, O. Bacterial microcompartments and the modular construction of microbial metabolism. *Trends Microbiol.* **23**, 22–34 (2015).
13. Cameron, J.C., Wilson, S.C., Bernstein, S.L. & Kerfeld, C.A. Biogenesis of a bacterial organelle: the carboxysome assembly pathway. *Cell* **155**, 1131–1140 (2013).
14. Kerfeld, C.A., Heinhorst, S. & Cannon, G.C. Bacterial microcompartments. *Annu. Rev. Microbiol.* **64**, 391–408 (2010).
15. Bobik, T.A., Havemann, G.D., Busch, R.J., Williams, D.S. & Aldrich, H.C. 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. *J. Bacteriol.* **181**, 5967–5975 (1999).
16. Havemann, G.D. & Bobik, T.A. Protein content of polyhedral organelles involved in coenzyme B12-dependent degradation of 1,2-propanediol in *Salmonella enterica* serovar Typhimurium LT2. *J. Bacteriol.* **185**, 5086–5095 (2003).
17. Sampson, E.M. & Bobik, T.A. 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 (2008).
18. Fan, C. & Bobik, T.A. The N-terminal region of the medium subunit (PduD) packages adenosylcobalamin-dependent diol dehydratase (PduCDE) into the Pdu microcompartment. *J. Bacteriol.* **193**, 5623–5628 (2011).
19. Fan, C. *et al.* Short N-terminal sequences package proteins into bacterial microcompartments. *Proc. Natl. Acad. Sci. USA* **107**, 7509–7514 (2010).
20. Fan, C., Cheng, S., Sinha, S. & Bobik, T.A. Interactions between the termini of lumen enzymes and shell proteins mediate enzyme encapsulation into bacterial microcompartments. *Proc. Natl. Acad. Sci. USA* **109**, 14995–15000 (2012).
21. Lawrence, A.D. *et al.* Solution structure of a bacterial microcompartment targeting peptide and its application in the construction of an ethanol bioreactor. *ACS Synth. Biol.* **3**, 454–465 (2014).
22. Jakobson, C.M., Tullman-Ercek, D., Slininger, M.F. & Mangan, N.M. A systems-level model reveals that 1,2-Propanediol utilization microcompartments enhance pathway flux through intermediate sequestration. *PLOS Comput. Biol.* **13**, e1005525 (2017).
23. Sutter, M., Greber, B., Aussignargues, C. & Kerfeld, C.A. Assembly principles and structure of a 6.5-MDa bacterial microcompartment shell. *Science* **356**, 1293–1297 (2017).
24. Cho, H. The role of cytoskeletal elements in shaping bacterial cells. *J. Microbiol. Biotechnol.* **25**, 307–316 (2015).
25. Cabeen, M.T. & Jacobs-Wagner, C. Bacterial cell shape. *Nat. Rev. Microbiol.* **3**, 601–610 (2005).
26. Cabeen, M.T. & Jacobs-Wagner, C. The bacterial cytoskeleton. *Annu. Rev. Genet.* **44**, 365–392 (2010).
27. Parsons, J.B. *et al.* Synthesis of empty bacterial microcompartments, directed organelle protein incorporation, and evidence of filament-associated organelle movement. *Mol. Cell* **38**, 305–315 (2010).
28. Chowdhury, C. *et al.* Selective molecular transport through the protein shell of a bacterial microcompartment organelle. *Proc. Natl. Acad. Sci. USA* **112**, 2990–2995 (2015).
29. Crowley, C.S. *et al.* Structural insight into the mechanisms of transport across the *Salmonella enterica* Pdu microcompartment shell. *J. Biol. Chem.* **285**, 37838–37846 (2010).
30. Pang, A., Frank, S., Brown, I., Warren, M.J. & Pickersgill, R.W. Structural insights into higher order assembly and function of the bacterial microcompartment protein PduA. *J. Biol. Chem.* **289**, 22377–22384 (2014).
31. Thomas, F., Boyle, A.L., Burton, A.J. & Woolfson, D.N. A set of *de novo* designed parallel heterodimeric coiled coils with quantified dissociation constants in the micromolar to sub-nanomolar regime. *J. Am. Chem. Soc.* **135**, 5161–5166 (2013).
32. Lee, M.J., Brown, I.R., Juodeikis, R., Frank, S. & Warren, M.J. Employing bacterial microcompartment technology to engineer a shell-free enzyme-aggregate for enhanced 1,2-propanediol production in *Escherichia coli*. *Metab. Eng.* **36**, 48–56 (2016).
33. Fletcher, J.M. *et al.* Self-assembling cages from coiled-coil peptide modules. *Science* **340**, 595–599 (2013).
34. Weber, B. *et al.* Automated tracing of microtubules in electron tomograms of plastic embedded samples of *Caenorhabditis elegans* embryos. *J. Struct. Biol.* **178**, 129–138 (2012).
35. Johnson, E. *et al.* Correlative in-resin super-resolution and electron microscopy using standard fluorescent proteins. *Sci. Rep.* **5**, 9583 (2015).
36. Mueller-Reichert, T. & Verkade, P. *Correlative Light and Electron Microscopy II; Methods in Cell Biology* Vol. 124 (Academic Press, 2014).
37. Szeto, T.H., Rowland, S.L., Habrukowich, C.L. & King, G.F. The MinD membrane targeting sequence is a transplantable lipid-binding helix. *J. Biol. Chem.* **278**, 40050–40056 (2003).
38. Noël, C.R., Cai, F. & Kerfeld, C.A. Purification and characterization of protein nanotubes assembled from a single bacterial microcompartment shell subunit. *Adv. Mater. Interfaces* **3**, 1500295 (2015).
39. Lewicka, A.J. *et al.* Fusion of pyruvate decarboxylase and alcohol dehydrogenase increases ethanol production in *Escherichia coli*. *ACS Synth. Biol.* **3**, 976–978 (2014).

Acknowledgments

We are grateful to the Biotechnology and Biological Sciences Research Council of the UK for a strategic LoLa Award to M.J.W., D.N.W., P.V. and W.-F.X. (BB/M002969/1). D.N.W. holds a Royal Society Wolfson Research Merit Award. We thank the Wolfson Bioimaging Facility and BrisSynBio, a BBSRC/EPSRC-funded Synthetic Biology Research Centre (L01386X), for access to confocal and electron microscopes; K. Howland for assistance with GC–MS analysis; R. Sessions and I. Uddin for preparing images used in **Supplementary Figure 1**; L. Harrington and P. Schwille for advice on the MinD system; and the entire BMC-SAGE LoLa group for helpful discussions.

Author contributions

M.J.L. made constructs, prepared samples for TEM and confocal analysis, imaged samples by TEM, purified nanotubes and analyzed them by TEM and AFM and conducted the ethanol production experiments and analyses. J.M. undertook tomography and 3D reconstructions. L.H. undertook CLEM sample preparation and imaging. D.A. undertook confocal imaging. I.R.B. sectioned samples for TEM analysis. W.-F.X. assisted with AFM and statistical analysis. M.J.L., J.M., L.H., J.M.F., S.F., P.V., D.N.W. and M.J.W. designed the experiments. All authors contributed to the manuscript.

Competing financial interests

The authors declare no competing financial interests.

Additional information

Any supplementary information, chemical compound information and source data are available in the online version of the paper. Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to M.J.W. or D.N.W.

ONLINE METHODS

Cloning of coiled-coil fused constructs. DNA encoding CC-Di-A and CC-Di-B embedded within a GS linker followed by a hexahistidine tag and a thrombin cleavage sequence was synthesized and cloned into the *Xba*I/*Nde*I sites of pET14b. A control sequence containing only a GS linker, hexahistidine tag and thrombin cleavage sequence was also synthesized and cloned by the same strategy. Strains and synthesized DNA sequences with corresponding amino acid sequences are shown in **Supplementary Tables 1** and **2**. Plasmids and primers used in this study are outlined in **Supplementary Tables 3** and **4**.

Expression of coiled-coil constructs. *E. coli* BL21*(DE3)-competent cells were transformed with a plasmid(s) containing the gene(s) of interest and plated onto LB agar plates supplemented with appropriate antibiotics (100 mg/L ampicillin and/or 34 mg/L chloramphenicol). For TEM analysis, 50 mL of LB was inoculated1:100 from an overnight starter culture and grown at 37 °C with shaking to an OD_{600} of ~0.4; protein production was induced by the addition of IPTG to a final concentration of 400 μM, and cultures were subsequently incubated overnight at 19 °C with shaking. For time-course analysis, 500 mL of LB was inoculated, grown and induced as described above. At time intervals 50 mL of media was removed for TEM analysis. For purification of nanotubes, 250 mL LB was inoculated 1:100 from an overnight starter culture and grown at 37 °C to an OD_{600} ~0.4. Protein production was induced by addition of IPTG to a final concentration of 400 μM, and cultures were then incubated with shaking at 19 °C overnight. For confocal imaging experiments 50 mL of LB was inoculated 1:100 from an overnight starter culture and grown with shaking at 37 °C to an OD_{600} ~0.4. protein production was induced by addition of IPTG to a final concentration of 400 μM, and cultures were then incubated with shaking at 19 °C for 4 h.

***In vivo* ethanol production.** For *in vivo* ethanol production, 100 mL of LB supplemented with 4% glucose and appropriate antibiotics was inoculated from overnight starter cultures to a starting OD_{600} of 0.05; cultures were grown at 28 °C for 120 h with shaking at 150 rpm. Protein production was induced by addition of IPTG to a final concentration of 400 μM after 4 h of growth. During growth, 1 mL samples were removed at 0, 2, 4, 6, 12, 24, 48, 72, 96 and 120 h for GC/MS analysis of the growth medium. Samples (1 mL) were also taken at each time point for SDS-PAGE analysis. Additional samples (5 mL) were taken after 24 h for TEM analysis.

Western blot analysis. Nitrocellulose membranes following transfer and blocking were incubated in primary antibody (rabbit anti-PduA, 1 μg/mL[27], or mouse anti-His (Sigma-Aldrich; catalog #H1029; diluted 1:3,000)), and then followed by incubation in a secondary antibody coupled to alkaline phosphatase (goat anti-rabbit IgG (H+L) alkaline phosphatase conjugate (Bio-Rad; catalog #170-6518; diluted 1:3,000) or anti-mouse IgG (H+L), AP conjugate (Promega; catalog #S3721; diluted 1:5,000)). Bands were visualized by incubation in substrate 5-bromo-4-chloro-3-indolyl phosphate/nitro blue tetrazolium (BCIP/NBT).

Transmission electron microscopy (TEM) analysis of cells. Cells grown as described previously were harvested by centrifugation at 3,000 × *g* for 10 min. The cell pellet was resuspended in 2 mL 2.5% (w/v) glutaraldehyde in 100 mM sodium cacodylate buffer, pH 7.2 (CAB), and fixed for 2 h with gentle rotating (20 rpm). Cells were pelleted by centrifugation at 6,000 × *g* for 2 min and were washed twice for 10 min with 100 mM CAB. Cells were post-fixed with 1% (w/v) osmium tetroxide in 100 mM CAB for 2 h and subsequently washed twice with ddH_2O. Cells were dehydrated by incubation in an ethanol gradient of 50% EtOH for 10 min, 70% EtOH overnight, and 90% EtOH for 10 min followed by three 10 min washes in 100% dry EtOH. Cells were then washed twice with propylene oxide for 15 min. Cell pellets were embedded by resuspension in 1 mL of a 1:1 mix of propylene oxide and Agar LV Resin and incubated for 30 min with rotation. Cell pellets were infiltrated twice in 100% Agar LV resin. The cell pellet was resuspended in fresh resin and transferred to a 1 mL Beem embedding capsule, centrifuged for 5 min at 3,000 × *g* to concentrate the cells to the tip of the mold and incubated for 20 h at 60 °C to polymerize.

Samples were ultra-thin sectioned on a RMC MT-XL ultra-microtome with a diamond knife (diatome 45°). Sections (60–70 nm) were collected on uncoated 300 mesh copper grids. Grids were stained by incubation in 4.5% (w/v) uranyl acetate in 1% (v/v) acetic acid solution for 45 min followed by washing in a stream of ddH_2O. Grids were then stained with Reynolds lead citrate for 7 min, which was followed by washing in a stream of ddH_2O. Electron microscopy was performed using a JEOL-1230 transmission electron microscope equipped with a Gatan multiscan digital camera operated at an accelerating voltage of 80 kV

Tomography. Sections (250 nm) were cut from the existing blocks, and 15 nm gold fiducials (Aurion, TomoSol solution) were applied to both surfaces of the sections. The sections were imaged at 200 kV in a Tecnai 20 TEM (FEI, the Netherlands) and double tilt series images acquired between −62° to +69.5° (first axis) and −68° to +69.5° (second axis) with 1.5° (above 50°) and 2° increments (below 50°). The pixel size on the 4k by 4k FEI Eagle camera was 0.74 nm. The resulting tomograms were reconstructed and combined using IMOD software[40,41]. The tube-like structures were modeled automatically using the AMIRA XTracing Extension of the AMIRA software suite, developed for automatic tracing of microtubules[34]. A cylinder template is correlated with the data to find and search for the center lines of tubes. A small cropped area was used to refine the fitting parameters as shown in **Supplementary Video 2** and these were then applied to the full data set. AMIRA software was further used for visualizing the data.

Measurements of *in vivo* nanotubes. Diameter measurements of 100 nanotubes from 10 cells were calculated in ImageJ[42]. Length measurements were calculated automatically using the XTracing extension of the AMIRA software suite. Cropping box measurements were removed manually from the data set, leaving a total of 739 tubes.

Purification of CC-Di-B-PduA*. CC-Di-B-tagged PduA* was overproduced as described previously. Cells were harvested by centrifugation at 2,683 × *g*. A 1 g wet cell pellet was resuspended in 20 mL Yeast Protein Extraction Reagent (Thermo Scientific) supplemented with Protease Inhibitor Cocktail Tablets, EDTA-Free (Sigma-Aldrich) and 500 Units Benzonase Nuclease (Merck) and incubated for 3 h at room temperature with gentle shaking. CC-Di-B-PduA* nanotubes were pelleted from the lysate by centrifugation for 5 min at 11,300 × *g*, and the pellet was resuspended in 2 mL of 20 mM Tris–HCl, pH 8, containing 20 mM NaCl. The suspension was centrifuged for 5 min at 11,000 × *g*, and the resulting nanotube containing pellet was resuspended in 20 mM Tris–HCl, pH 8, and centrifuged again as above. The supernatant was removed and adjusted with a solution of 5 M NaCl to give a final concentration of 80 mM. A final centrifugation step as above was performed, and the resulting pellet was analyzed for the presence of PduA* nanotubes.

Analysis of purified nanotubes. *TEM*: Following purification, 20 μL of CC-Di-B-PduA* nanotubes were deposited onto Formvar, carbon-coated 300 mesh copper grids and incubated to 5 min. Glutaraldehyde (20 μL of 2.5% (v/v)) in PBS was then added and incubated for a further 5 min before washing in three drops of 2.5% (v/v) glutaraldehyde in PBS followed by three drops of ddH_2O. Grids were stained with 2% (w/v) aqueous uranyl acetate and subsequently dried. Electron microscopy was performed using a JEOL-1230 transmission electron microscope equipped with a Gatan multiscan digital camera operated at an accelerating voltage of 80 kV.

AFM: Purified CC-Di-B-PduA* nanotubes (20 μL) were deposited onto freshly cleaved mica surfaces and incubated for 5 min, followed by the addition of 20 μL 2.5% (v/v) glutaraldehyde in PBS. Surfaces were washed three times with 1 mL of ddH_2O then dried under a gentle stream of N_2. Images were acquired in air at 20 °C using a Bruker MultiMode 8 Scanning probe microscope operating under Peak-Force tapping mode (ScanAsyst, Bruker) with a ScanAsyst-air probe (Bruker). Areas (10 μm × 10 μm) were scanned at a resolution of 4,096 × 4,096 pixels. Bow and tilt were removed using NanoScope Analysis 1.4 (Bruker).

Confocal imaging. Following growth and induction of protein expression, 1 mL of cells was harvested by centrifugation at 3,000 × *g*. The resulting cell pellet was washed three times in PBS before being incubated for 15 min in

doi:10.1038/nchembio.2535

2% (w/v) formaldehyde in PBS. Cells were then washed three more times in PBS. Cells (10 µL) were pipetted onto a 1.5 thickness coverslip before being inverted onto a drop of ProLong Gold antifade mountant (Life Technologies) on a glass slide. Slides were incubated at room temperature in the dark for 24 h to cure. Images were acquired on a Leica TCS SP8 system attached to a Leica DMi8 inverted microscope (Leica Microsystems). Excitation light (514 nm for Citrine or 594 nm for mCherry) was provided by a white-light laser with a repetition rate of 80 MHz. Images were acquired using a 100 × 1.4 NA oil immersion objective and fluorescence was detected through bandpasses of 520–570 nm (Citrine detection) or 600–650 nm (mCherry detection).

Correlative light electron microscopy. Cells (1 µL) were harvested by centrifugation at 3,000 × *g* for 5 min, and then loaded into a 0.1 mm membrane carrier (Leica) and vitrified by high pressure freezing (EMPACT2 + RTS, Leica). Frozen membrane carriers were transferred into 1 mL of freeze substitution medium (0.2% uranyl acetate, 5% H_2O, in acetone) and held at −90 °C for 5 h in an automated freeze substitution unit (AFS2, Leica) equipped with an attachment for automated reagent exchange (Freeze Substitution Processor, FSP, Leica). Samples were warmed to −45 °C at a rate of 5 °C/h, and held at −45 °C for 2 h before washing in acetone and ethanol for 30 min each. Samples were then infiltrated with 25%, 50% and 75% dilutions of Lowicryl HM20 resin for 3 h each before infiltrating with 100% resin overnight, which was followed by three further changes of resin for 2 h each. UV polymerization was performed over approximately 48 h; Samples were initially kept at −45 °C for 16 h, then warmed to 0 °C at a slope of 5 °C/h, and finally kept at 0 °C for approximately 14 h.

Following polymerization, blocks were removed from flow through containers, and carriers were detached using liquid nitrogen and the specimen carrier detaching tool (Leica) heated to 40 °C. Blocks were trimmed and sectioned with a 45° diamond knife using an EM UC6 microtome (Leica). 70- and 300-nm thick sections were collected on carbon-coated pioloform films on H6 copper finder grids (Agar Scientific). Grids were air dried, mounted in PBS between a glass slide and coverslip and imaged by light microscopy using a Leica DMI4000 B inverted epifluorescence microscope fitted with a 63× oil immersion lens (NA 1.4). After imaging, the grids were washed in H_2O and air dried before imaging in TEM. Image registration of light and electron microscopy images was performed using the eC-CLEM plugin in ICY[43].

Analysis of enzyme levels. Relative amounts of Pdc and Adh were quantified by western blot. Total cell lysate samples, adjusted to cell number were analyzed by SDS–PAGE and subsequently western blot analysis. Peak areas were quantified using the gel analysis tool in ImageJ. Due to the higher molecular weight band close to CC-Di-A-Adh, half of this peak was quantified with the assumption that the peak was symmetrical. Measurements were repeated for each of the cultures.

MinD colocalization. DNA encoding the c-terminal membrane-associating region of MinD was synthesized and cloned into the *Spe*I/*Blp*I sites of pET_CC_Di_A_Citrine_No_Stop and pET_C_Citrine_No_Stop. Cells were transformed as described previously and grown in LB media at 37 °C with shaking to an OD_{600} ~ 0.4, protein production was induced by addition of IPTG to a final concentration of 400 µM. Cultures were then incubated with shaking at 19 °C for 4 h. Cells were harvested, fixed, embedded and sectioned as described previously. A total of 250 cells in the transverse orientation for each strain were analyzed for the presence and location of transverse CC-Di-B-PduA* nanotubes. Statistical analysis was performed in Minitab Software version 17 using a one-way ANOVA (analysis of variance) at the 99% level with post-hoc analysis by Tukey's test.

Life sciences reporting summary. Further information on experimental design and reagents is available in the **Life Sciences Reporting Summary**.

Data availability. All data generated or analyzed during this study are included in this published article (and supplementary information files) or are available from the corresponding authors on reasonable request.

40. Kremer, J.R., Mastronarde, D.N. & McIntosh, J.R. Computer visualization of three-dimensional image data using IMOD. *J. Struct. Biol.* **116**, 71–76 (1996).
41. Mastronarde, D.N. Dual-axis tomography: an approach with alignment methods that preserve resolution. *J. Struct. Biol.* **120**, 343–352 (1997).
42. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
43. Paul-Gilloteaux, P. *et al.* eC-CLEM: flexible multidimensional registration software for correlative microscopies. *Nat. Methods* **14**, 102–103 (2017).

Supplementary Information for:

Engineered synthetic scaffolds for organising proteins within bacterial cytoplasms

Authors: Matthew J. Lee,[1] Judith Mantell,[2,3] Lorna Hodgson,[2] Dominic Alibhai,[2] Jordan M Fletcher,[4] Ian R. Brown,[1] Stefanie Frank,[5] Wei-Feng Xue,[1] Paul Verkade,[2,3,6] Derek N Woolfson,[2,4,6]* Martin J Warren[1]*.

[1]Industrial Biotechnology Centre, School of Biosciences, University of Kent, Canterbury CT2 7NJ, UK

[2]School of Biochemistry, University of Bristol, Medical Sciences Building, University Walk, Bristol BS8 1TD, UK.

[3]Wolfson Bioimaging Facility, Medical Sciences Building, University Walk, Bristol BS8 1TD, UK.

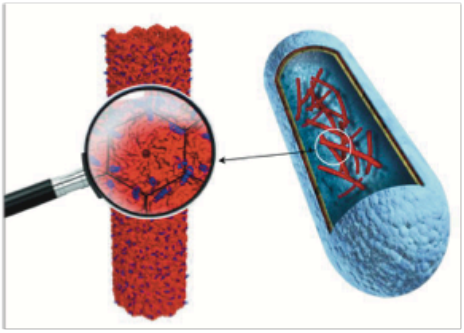
[4]School of Chemistry, University of Bristol, Cantock's Close, Bristol BS8 1TS, UK.

[5]Department of Biochemical Engineering, University College London, Bernard Katz Building, Gordon Street, London WC1E 6BT, UK.

[6]BrisSynBio, Life Sciences Building, Tyndall Avenue, Bristol BS8 1TQ, UK.

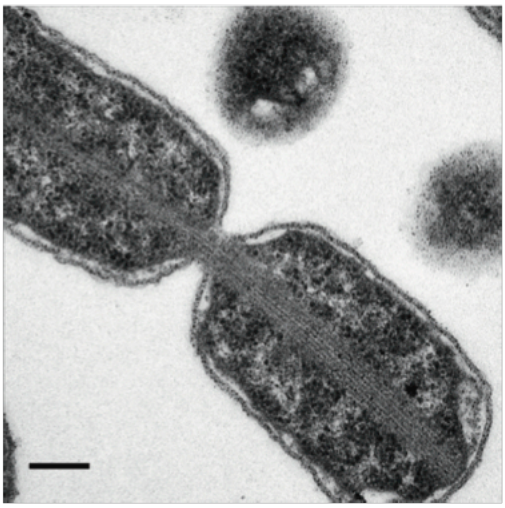
* Correspondence to: Martin J Warren (M.J. Warren@kent.ac.uk) and Derek N Woolfson (D.N.Woolfson@bristol.ac.uk)

Nature Chemical Biology: doi:10.1038/nchembio.2535



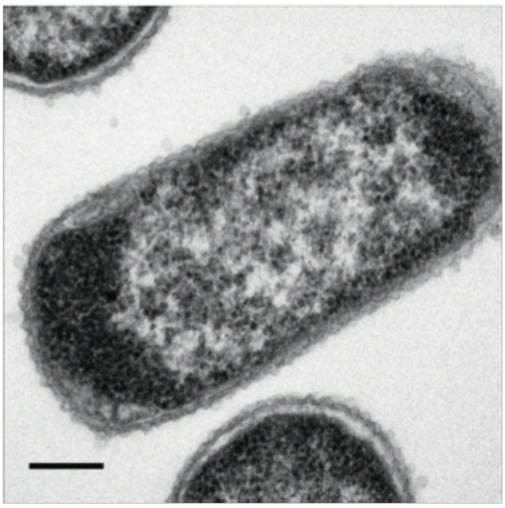
Supplementary Fig. 1. Construction of a bacterial cytoscaffold. Hexagonally arrayed, tubular filaments of the PduA* protein (red) carrying half of a *de novo* designed heterodimeric coiled coil (blue).

Nature Chemical Biology: doi:10.1038/nchembio.2535



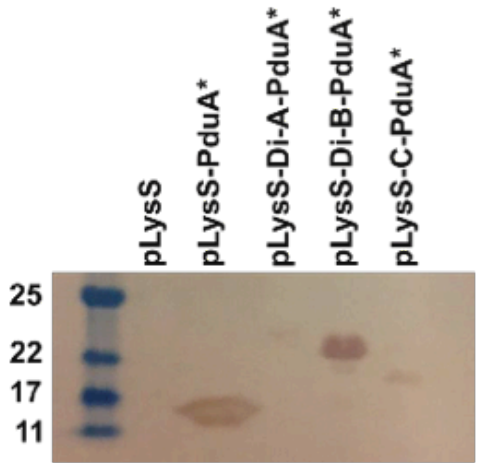
Supplementary Fig. 2. TEM analysis of thin sectioned *E. coli* cells transformed with pLysS-PduA*. Scale bar shows 200 nm

Nature Chemical Biology: doi:10.1038/nchembio.2535



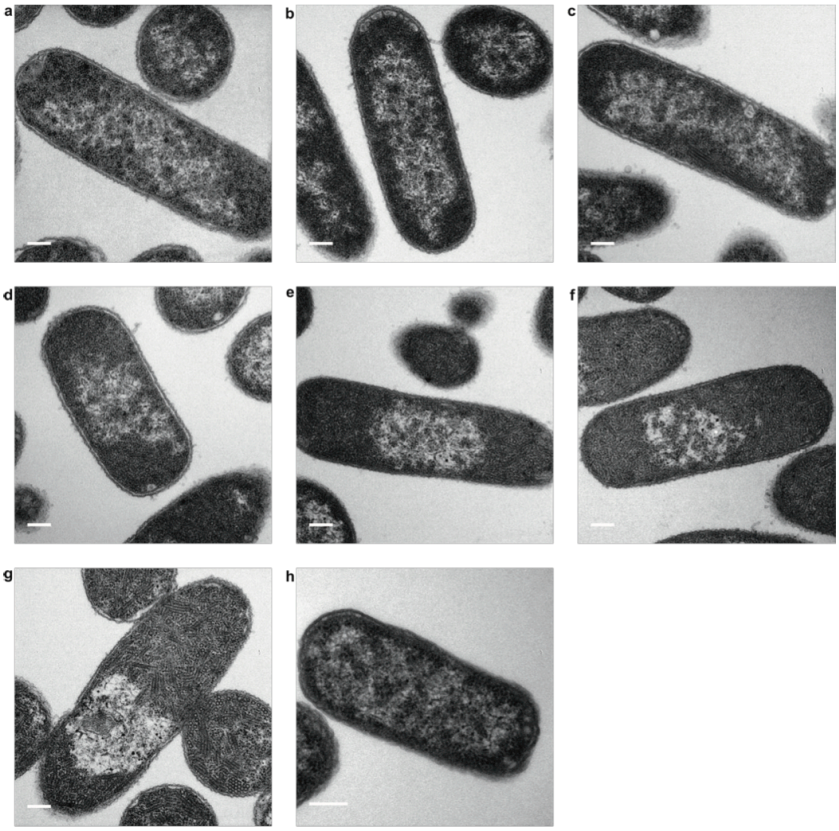
Supplementary Fig. 3. TEM analysis of *E. coli* cells transformed with C-PduA*. Scale bar shows 200 nm.

Nature Chemical Biology: doi:10.1038/nchembio.2535



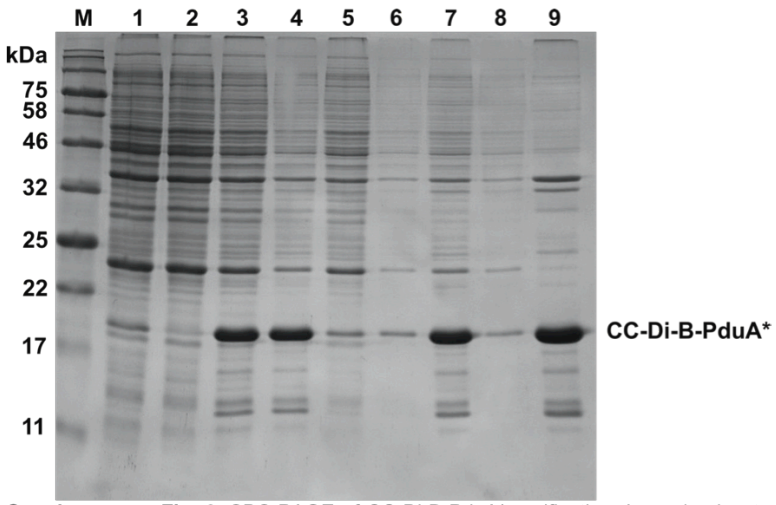
Supplementary Fig. 4. Western blot analysis of *E. coli* cells transformed with varying PduA* constructs. Total lysates were analyzed by SDS-PAGE and subsequently western blotted with an anti-PduA* primary antibody in comparison to a molecular weight standard. Cell densities were normalized to an OD_{600} of 5 for loading.

Nature Chemical Biology: doi:10.1038/nchembio.2535



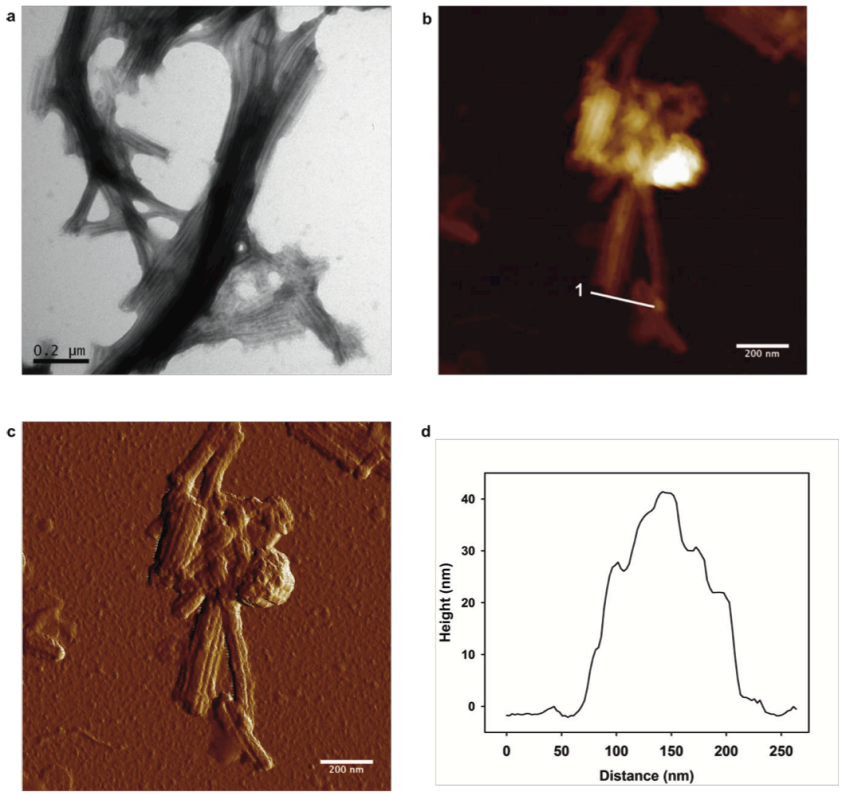
Supplementary Fig. 5. Time course of pLysS-CC-Di-B-PduA* induction with 400 μM IPTG in BL21 * (DE3) cells. (**a**) pre-induction (**b**) 1h post induction (**c**) 2h post induction (**d**) 3h post induction (**e**) 4h post induction (**f**) 6h post induction (**g**) 20h post induction (**h**) Control strain transformed with an empty pLysS vector. Scale bars show 200 nm

Nature Chemical Biology: doi:10.1038/nchembio.2535



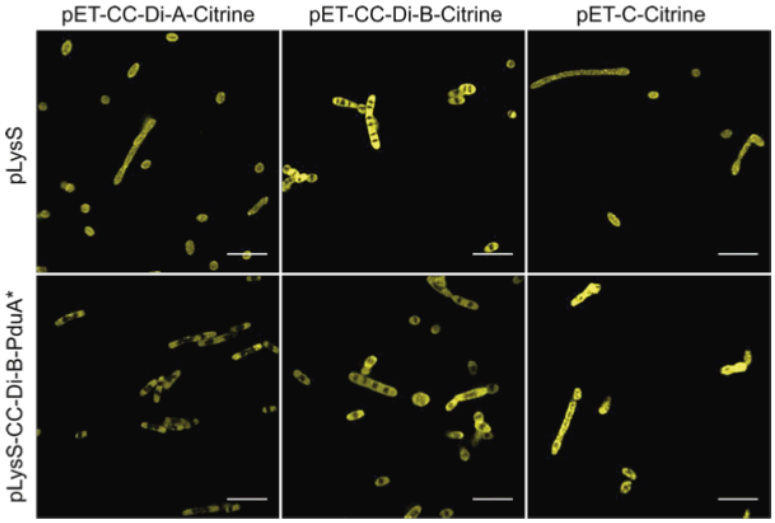
Supplementary Fig. 6. SDS-PAGE of CC-Di-B-PduA* purification. Lane 1 – lysate (2 µL), lane 2 – supernatant after first centrifugation (2 µL), lane 3 – pellet after first centrifugation (5 µL), lane 4 – pellet after second centrifugation (5 µL), lane 5 – supernatant after second centrifugation (5 µL), lane 6 – supernatant after third centrifugation (5 µL), lane 7 – pellet after third centrifugation (5 µL), lane 8 – supernatant after final centrifugation (5 µL), lane 9 – Pellet after final centrifugation (5 µL).

Nature Chemical Biology: doi:10.1038/nchembio.2535



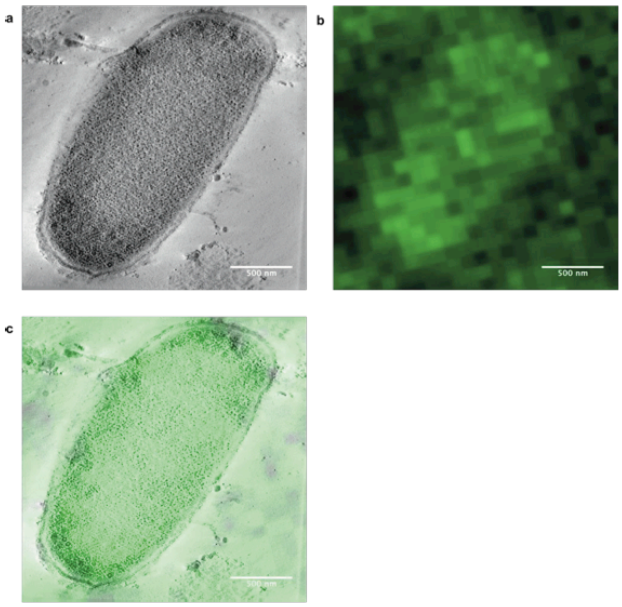
Supplementary Fig. 7. TEM (**a**), AFM height (**b**), PeakForce (**c**) and height profile corresponding to line 1 (panel B) (**d**) analysis of purified CC-Di-B-PduA* nanotubes.

Nature Chemical Biology: doi:10.1038/nchembio.2535



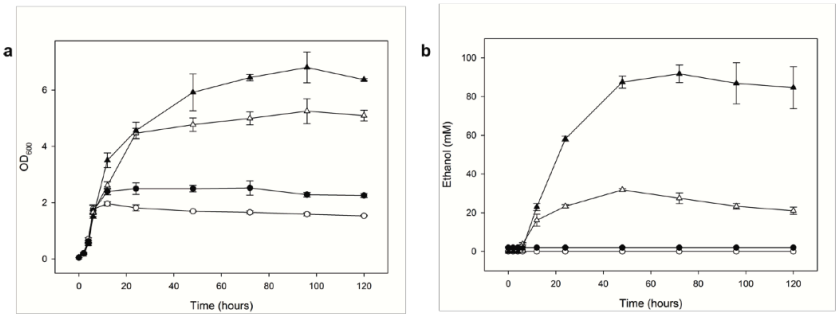
Supplementary Fig. 8. Confocal analysis of BL21 * (DE3) cells expressing CC-Di-B-PduA* or a control empty vector (pLysS) with citrine tagged with CC-Di-A or CC-Di-B or a control C-citrine.

Nature Chemical Biology: doi:10.1038/nchembio.2535



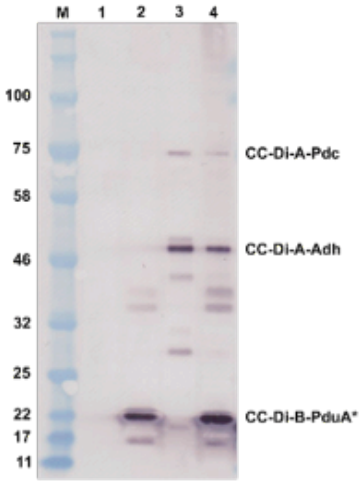
Supplementary Fig. 9. Correlative Light Electron Microscopy of *E. coli* cells expressing CC-Di-B-PduA* and C-GFP.

Nature Chemical Biology: doi:10.1038/nchembio.2535



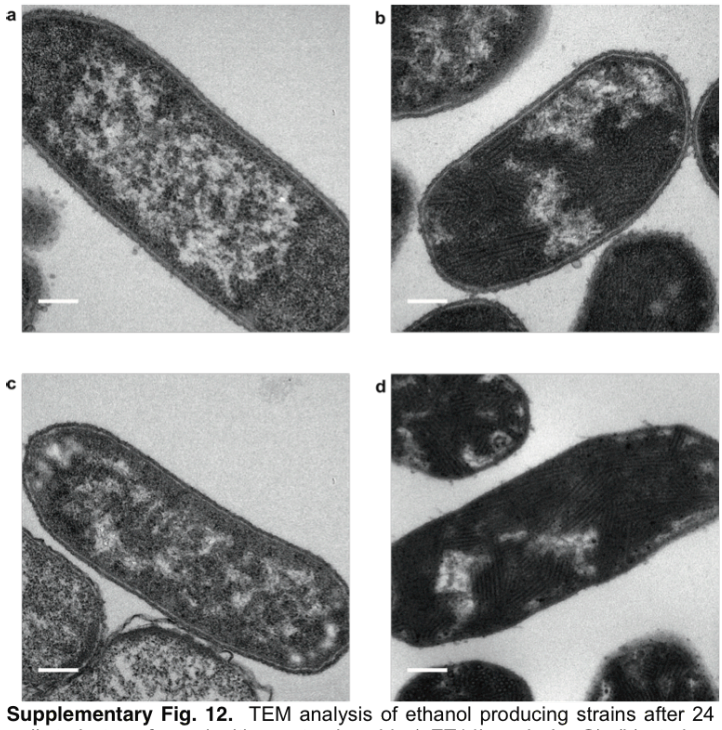
Supplementary Fig. 10. (a) Growth curves of strains producing ethanol and control strains shown as OD_{600} against time (hours). (b) In vivo ethanol production. Graph shows ethanol content of the growth medium over time. *E. coli* strain transformed with empty plasmids (pET14b and pLysS) (○), strain producing CC-Di-B-PduA* only (●), strain producing CC-Di-A-Pdc and CC-Di-A-Adh (△), strain producing CC-Di-A-Pdc, CC-Di-A-Adh and CC-Di-B-PduA* (▲). Data points represent an average of three independent experiments; standard deviations are represented by error bars.

Nature Chemical Biology: doi:10.1038/nchembio.2535



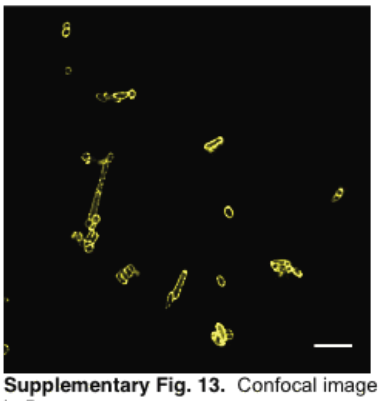
Supplementary Fig. 11. Western blot analysis of strains producing ethanol and control strain in comparison to a molecular weight marker (M). Total lysates at 24h were adjusted to the same OD_{600} (1.6) and analyzed by SDS-PAGE and subsequently western blotted with an anti-polyhistidine primary antibody. (1) *E. coli* strain transformed with empty plasmids (pET14b and pLysS), (2) strain producing CC-Di-B-PduA* only, (3) strain producing CC-Di-A-Pdc and CC-Di-A-Adh, (4) strain producing CC-Di-A-Pdc, CC-Di-A-Adh and CC-Di-B-PduA*.

Nature Chemical Biology: doi:10.1038/nchembio.2535



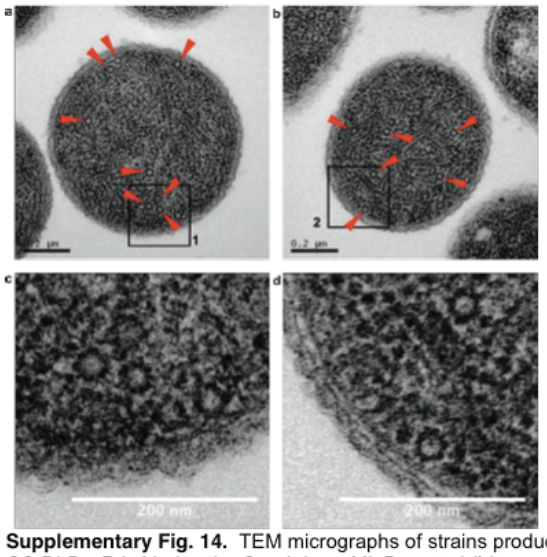
Supplementary Fig. 12. TEM analysis of ethanol producing strains after 24 hours. (a) *E. coli* strain transformed with empty plasmids (pET14b and pLysS), (b) strain producing CC-Di-B-PduA* only, (c) strain producing CC-Di-A-Pdc and CC-Di-A-Adh, (d) strain producing CC-Di-A-Pdc, CC-Di-A-Adh and CC-Di-B-PduA*.

Nature Chemical Biology: doi:10.1038/nchembio.2535



Supplementary Fig. 13. Confocal image of strain expressing C—citrine—MinD; scale bar is 5 μm.

Nature Chemical Biology: doi:10.1038/nchembio.2535



Supplementary Fig. 14. TEM micrographs of strains producing CC-Di-B—PduA* only (**a**); CC-Di-B—PduA* plus the C—citrine—MinD control (**b**), arrows indicate transverse filaments. Zoom in of areas 1 and 2 in panels a and b respectively (**c** and **d**).

Nature Chemical Biology: doi:10.1038/nchembio.2535

Supplementary Tables

Supplementary Table 1. Strains used in this study

Strain	Genotype	Source
JM109	endA1, recA1, gyrA96, thi, hsdR17 (rk–, mk+), relA1, supE44, Δ(lac-proAB), [F', traD36, proAB, laqIqZΔM15]	Promega
BL21 * (DE3)	F– ompT hsdSB (rB– mB–) gal dcm (DE3)	Novagen

Supplementary Table 2. DNA synthesized for this study

Name	DNA / translated sequence (coiled-coil peptides underlined)
CC-Di-A	TCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATATGGGTAGCG AAATTGCCGCGCTGGAAAAAGAAAACGCTGCCCTGGAGCAGGAAAT CGCGGCACTGGAGCAGGGTAGCGGCAGCGGTAGCACCATGGGCAG CAGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGC AGCCATATG
	MGSEIAALEKENAALEQEIAALEQGSGSGSTMGSSHHHHHHSSGLVPR GSHM
CC-Di-B	TCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATATGGGTAGCA AAATCGCCGCACTGAAAAAGAAAAACGCTGCGCTGAAACAGAAAATT GCCGCGCTGAAACAGGGTAGCGGCAGCGGTAGCACCATGGGCAGC AGCCATCATCATCATCATCACAGCAGCGGCCTGGTGCCGCGCGGCA GCCATATG
	MGSKIAALKKKNAALKQKIAALKQGSGSGSTMGSSHHHHHHSSGLVPR GSHM
Control	TCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATATGGGTAGCG GTAGCGGCAGCGGTAGCACCATGGGCAGCAGCCATCATCATCATCA TCACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATG
	MGSGSGSGSTMGSSHHHHHHSSGLVPRGSHM
MinD	ACTAGTGGTTCGGGATCGCCGTTGCAAGTCTTGGAGGAACAAAACA AAGGGATGATGGCTAAAATCAAATCATTTTTCGGTGTACGTTCCTAA CCTAGGTTTGGATCCGGCTGCTAACAAAGCCCGAAAGGAAGCTGAG TTGGCTGCTGCCACCGCTGAGC
	TSGSGSPLQVLEEQNKGMMAKIKSFFGVRS

Supplementary Table 3. Plasmids Used in this study

Plasmid name	Description	Source
pET14b	Overexpression vector containing N-terminal hexahistidine-tag, modified to include an *Spe*I site 5' of *Bam*HI	Novagen
pLysS	Basal expression suppressor	Novagen
pLysS_2	Basal expression suppressor containing the T7 promoter-MCS-T7 terminator cassette from pET14b	Parsons *et al.*, 2008
pET3a_PduA*	PCR product of pduA* ligated into the *Nde*I and *Spe*I sites of pET3a (4 bp removed from 3' end of pduA, this allowed non-coding DNA from the vector to be translated)	Parsons *et al.*, 2010
pLysS_PduA*	*Nde*I/*Spe*I fragment of pET3a_PduA* ligated into *Nde*I/*Spe*I sites of pLysS_2	This study
pET_CC_Di_A	Synthesised CC_Di_A gene fragment cloned into XbaI/NdeI sites of pET14b	This study
pET_CC_Di_B	Synthesised CC_Di_B gene fragment cloned into XbaI/NdeI sites of pET14b	This study
pET_C	Synthesised control (C) gene fragment cloned into XbaI/NdeI sites of pET14b	This study
pET_CC_Di_A_PduA*	*Nde*I/*Spe*I fragment of pLysS_PduA* cloned into *Nde*I/*Spe*I sites of pET_CC_Di_A	This study
pET_CC_Di_B_PduA*	*Nde*I/*Spe*I fragment of pLysS_PduA* cloned into *Nde*I/*Spe*I sites of pET_CC_Di_B	This study
pET_C_PduA*	*Nde*I/*Spe*I fragment of pLysS_PduA* cloned into *Nde*I/*Spe*I sites of pET_C	This study
pET-TBAD-cobA		Gift from E. Deery
pET3a-mCherry-PduABB'JKNU		Parsons *et al.*, 2010
pET-TBAD-mCherry-PduABJKNU.	*Xba*I/*Hind*III fragment of pET3a-mCherry-PduABJKNU cloned into *Xba*I/*Hind*III sites of pET-TBAD-cobA	This study
pLysS-mCherry-PduABB'		Parsons *et al.*, 2010
pLysS-pRha-mCherry-PduAB	Exponential Megapriming PCR (EMP) and splicing by Overlap Extension (SOE) to insert BglII site and pRha promoter at former T7 promoter site of pLysS-mCherry-PduABB'	This study
pLysS_TBAD_mCherry_PduABJKNU	BglII/SpeI fragment of pET-TBAD-mCherry-PduABJKNU cloned into BglII/SpeI sites of pLysS-pRha-mCherry-PduAB	This study
pLysS_CC_Di_A_PduA*	*Bgl*II/*Spe*I fragment of pET_CC_Di_A_PduA* cloned into *Bgl*II/*Spe*I sites of	This study

Nature Chemical Biology: doi:10.1038/nchembio.2535

	pLysS_TBAD_mCherry_PduABJKNU	
pLysS_CC_Di_B_PduA*	*Bgl*II/*Spe*I fragment of pET_CC_Di_B_PduA* cloned into *Bgl*II/*Spe*I sites of pLysS_TBAD_mCherry_PduABJKNU	This study
pLysS_C_PduA*	*Bgl*II/*Spe*I fragment of pET_C_PduA* cloned into *Bgl*II/*Spe*I sites of pLysS_TBAD_mCherry_PduABJKNU	This study
pET_CC_Di_A_Citrine	Citrine PCR product ligated into *Nde*I/*Spe*I sites of pET_CC_Di_A	This study
pET_CC_Di_B_Citrine	Citrine PCR product ligated into *Nde*I/*Spe*I sites of pET_CC_Di_B	This study
pET_C_Citrine	Citrine PCR product ligated into *Nde*I/*Spe*I sites of pET_C	This study
pET_CC_Di_A_mCherry	mCherry PCR product ligated into *Nde*I/*Spe*I sites of pET_CC_Di_A	This study
pET_CC_Di_B_mCherry	mCherry PCR product ligated into *Nde*I/*Spe*I sites of pET_CC_Di_B	This study
pET_C_mCherry	mCherry PCR product ligated into *Nde*I/*Spe*I sites of pET_C	This study
pET_CC_Di_A_Citrine_CC_Di_A_mCherry	*Xba*I/*Eco*RI fragment from pET_CC_Di_A_mCherry ligated into *Spe*I/*Eco*RI sites of pET_CC_Di_A_Citrine	This study
pET_C_Citrine_C_mCherry	*Xba*I/*Eco*RI fragment from pET_C_mCherry ligated into *Spe*I/*Eco*RI sites of pET_C_Citrine	This study
pET23b-Pdc		Lawrence *et al.*, 2014
pET23b-Adh		Lawrence *et al.*, 2014
pET_CC_Di_A_Pdc	NdeI/EcoRI fragment of pET23b-Pdc cloned into NdeI/EcoRI sites of pET_CC_Di_A	This study
pET_CC_Di_A_Adh	NdeI/EcoRI fragment of pET23b-Adh cloned into NdeI/EcoRI sites of pET_CC_Di_A	This study
pET_CC_Di_A_Adh_CC_Di_A_Pdc	XbaI/EcoRI fragment from pET_CC_Di_A_Pdc ligated into SpeI/EcoRI sites of pET_CC_Di_A_Adh	This study

Nature Chemical Biology: doi:10.1038/nchembio.2535

Supplementary Table 4. Oligonucleotides used in this study, restriction sites are underlined

Name	Sequence 5' – 3'
Citrine_NdeI_FW	CACCATATGGTGAGCAAGGGCGAGGAGC
Citrine_SpeI_RV	CACACTAGTTTACTTGTACAGCTCGTCC
mCherry_NdeI_FW	CAGCATATGGTGAGCAAGGGCGAGG
mCherry_SpeI_RV	GACACTAGTTTACTTGTACAGCTCGTCCATGC
Citrine_SpeI_NoStop_RV	CACACTAGTCTTGTACAGCTCGTCC

Nature Chemical Biology: doi:10.1038/nchembio.2535

Supplementary Movies

Supplementary Video 1. Tomography of CC-Di-B-PduA filaments and automated microtubule tracing.

Supplementary Video 2. Refined tracing model on a small area of the tomogram shown in movie S1.

Nature Chemical Biology: doi:10.1038/nchembio.2535

6.4. *De novo* targeting to the cytoplasmic and luminal side of bacterial microcompartments

Title: *De novo* targeting to the cytoplasmic and luminal side of bacterial microcompartments

Authors: Matthew J. Lee[1], Ian R. Brown[1], Jordan M Fletcher[2], Richard W. Pickersgill[3], Derek N Woolfson[2,4,5], Stefanie Frank[6]*, Martin J Warren[1]*

[1]Industrial Biotechnology Centre, School of Biosciences, University of Kent, Canterbury CT2 7NJ, UK
[2]School of Chemistry, University of Bristol, Cantock's Close, Bristol BS8 1TS, UK.
[3]School of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS
[4]School of Biochemistry, University of Bristol, Medical Sciences Building, University Walk, Bristol BS8 1TD, UK.
[5]BrisSynBio, Life Sciences Building, Tyndall Avenue, Bristol BS8 1TQ, UK.
[6]Department of Biochemical Engineering, University College London, Bernard Katz Building, Gordon Street, London WC1E 6BT, UK.

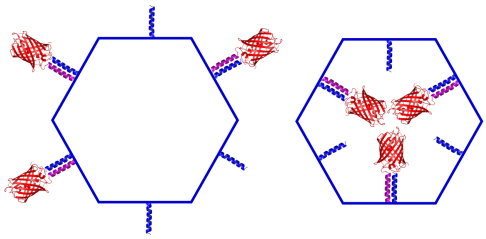
* Correspondence to: Martin J Warren (M.J. Warren@kent.ac.uk) and Stefanie Frank (Stefanie.frank@ucl.ac.uk)

Running title: BMC targeting with coiled-coils

Abstract

Bacterial microcompartments, BMCs, are proteinaceous organelles that encase a specific metabolic pathway within a semi-permeable protein shell. Short encapsulation peptides direct cargo proteins to the lumen of the compartments. However, the fusion of such peptides to non-native proteins does not guarantee encapsulation and often causes aggregation. Here, we report a different approach for targeting recombinant proteins to BMCs that utilizes specific *de novo* coiled-coil protein-protein interactions. Attachment of one coiled-coil module to a component of the bacterial microcompartment shell, PduA, allows targeting of a fluorescent protein marker fused to a cognate coiled-coil partner. This interaction takes place on the outer surface of the BMC. The redesign of PduA to generate a new *N*-terminus on the luminal side of the BMC results in intact compartments to which fluorescent proteins can still be targeted *via* the designed coiled-coil system. This study demonstrate that *de novo* coiled-coil peptide technology can be used to target to the luminal or cytoplasmic faces of BMCs, allowing proteins to be displayed on the surface or internalized within the shell of the BMC.

Graphical abstract



Keywords

Bacterial microcompartments, protein design, synthetic biology, compartmentalization, biotechnology.

Although once thought to lack subcellular organization, prokaryotes utilize a number of mechanisms to systematically arrange cellular enzymes to both enhance metabolic pathways and also to protect cellular components from potentially damaging intermediates. Re-engineering heterologous metabolic pathways with a higher level of organization has become of increasing interest to synthetic biologists looking to maximize yields from engineered pathways[1]. Such mechanisms include protein-based scaffolds, DNA scaffolds, protein aggregation and compartmentalization[2-6]. However, in some eubacteria, nature has evolved small protein-based organelles called bacterial microcompartments (BMCs) that contain a specific metabolic pathway[7]. These BMCs, which are found in around 20% of bacteria, consist of a protein shell that encases the enzymes involved in particular metabolic pathways[8-11]. A recent computational study has revealed that the primary function of BMCs is to enhance luminal intermediate concentrations leading to an enhancement in metabolic flux[12]. In addition, there is convincing evidence that BMCs protect cells from the toxicity associated with high aldehyde concentrations[13, 14].

BMCs can be broadly divided into two subclasses, anabolic carboxysomes and catabolic metabolosomes[7, 15]. Carboxysomes encase the enzymes carbonic anhydrase and ribulose-1,5-bisphosphate carboxylase/oxygenase (RuBisCO) leading to high levels of the substrate CO_2 to overcome the poor catalytic activity of RuBisCO[16, 17]. Of the catabolic metabolosomes, the organelle associated with 1,2-propanediol utilization (Pdu) is , to date, the best characterized[18-20]. This structure, approximately 150 nm in diameter, houses the enzymes required for the metabolism of 1,2-propanediol (1,2-PD) via the toxic intermediate propionaldehyde to form 1-propanol and propionyl-CoA[18, 21]. The genetic information encoding this metabolic module is housed on a 23-gene operon, the so-called *pdu* operon[18]. Seven genes encode for eight proteins that form the shell of the BMC (PduA, B, B', J, K, N, U, T), which either form hexameric (BMC-H), trimeric (BMC-T) or pentameric (BMC-P) disks that assemble to form the shell[22-25]. Indeed a recent structural study has revealed the organization of these proteins in a recombinant BMC structure[26].

The biotechnological potential of BMCs was first realized through the expression of the entire *pdu* operon in *E. coli* resulting in the formation of functional metabolosomes[27]. Further work showed that through the expression of a minimal set of BMC shell proteins (PduA, B, B', J, K, N and U) empty BMCs could be readily produced and isolated[28]. Proteins can be targeted to such empty BMCs by fusion to

short targeting sequences, namely, D18 and P18, that are naturally found on the N-terminus of the proteins PduD and PduP, respectively, which are integral components of the Pdu metabolosome[29-31]. These targeting peptides have been used to show the targeting of fluorescent proteins to BMCs, and, more recently, they have been utilized to construct an *in vivo* ethanol bioreactor[6]. For the latter, the enzymes pyruvate decarboxylase and alcohol dehydrogenase have been targeted to recombinant BMCs by fusion to the D18 and P18 targeting peptides. The resulting strains are able to produce more ethanol than strains producing the enzymes cytoplasmically, demonstrating that compartmentalization is beneficial to this recombinant pathway. Other work has shown that it is possible to utilize targeting peptides from different BMCs to target proteins to 1,2-PD utilization BMCs. This promiscuity suggests a relatively basic mechanism of targeting[32]. Indeed, sequence analysis reveals a conserved hydrophobic motif among BMC systems and subsequent work by Jakobson *et al.* shows that it is possible to design targeting peptides *de novo* based on this hydrophobic motif[32, 33].

Recent work has shown that fusion of the D18 and P18 targeting peptides to proteins of an engineered 1,2-PD synthesis pathway results in poor levels of protein encapsulation and variable levels of protein aggregation[4]. Nonetheless, when co-produced together these large enzyme aggregates produce significantly more product than the untagged cytoplasmic variants. Therefore, such active enzyme aggregates may provide an alternative strategy to enhance metabolic flux. However, aggregation may be problematic if the full potential of BMCs is to be realized and therefore alternative-targeting mechanisms may facilitate the further development of BMC technology. Towards this goal researchers have recently used protein-interaction domains such as PDZ, GBD and SH3 to target fluorescent proteins to recombinant BMCs in the host *Corynebacterium glutamicum*[34].

Here, we describe an altogether different strategy for targeting recombinant proteins to BMCs. This utilizes *de novo* designed coiled coils[35, 36]. By attaching one half of an obligate coiled-coil heterodimer to the *N*-terminal, cytoplasm-facing region of PduA, we demonstrate that it is possible to target it with cargo that has been tagged with the cognate coiled-coil partner. To target cargo to the luminal side of the BMC, PduA was engineered so that its *N* terminus was luminal facing. This permutated version of PduA still forms BMCs when coproduced with the other shell proteins and the use of the heterodimeric coiled-coil peptide system then allows for internalization of tagged cargo.

RESULTS

Construction of coiled-coil labeled BMCs. Previously, it has been shown that heterologous proteins can be targeted to BMCs by fusion to short targeting peptides such as P18 or D18, however such fusions often result in protein aggregation[4, 29, 30]. The *de novo* coiled-coil peptides CC-Di-A and CC-Di-B have been designed to interact specifically with each other and to avoid self-association[35, 36]. Using these peptides, we have demonstrated recently that filaments composed of a CC-Di-B-PduA* fusion can be targeted with fusions of CC-Di-A to fluorescent proteins such as Citrine and mCherry[5]. Therefore, we wondered if the CC-Di-A/CC-Di-B (CC-Di-AB) cognate pairing could also be used to target proteins to intact BMCs. To investigate this, we generated a range of coiled-coil-PduA fusions with the aim of incorporating the fused PduA protein into BMCs within *E. coli*. To achieve this the coiled-coil peptides were fused to the N terminus of PduA *via* a glycine/serine-based linker (GS Linker) but separated by a hexa-histidine tag and a thrombin cleavage site. This led to the construction of synthetic genes encoding for CC-Di-A-PduA and CC-Di-B-PduA. The N terminus of wild-type PduA is located on the concave side of the hexamer and in assembled BMCs faces the cellular cytoplasm[26]. Hence, targeting to the N terminus of PduA should append cargo proteins to the outer surface of the organelle.

In other constructs, the sequence for a fluorescent protein (Citrine) was incorporated between the thrombin cleavage site and PduA to allow for the production of fluorescently labeled, coiled-coil tagged, BMC shells. This gave CC-Di-A-Citrine-PduA and CC-Di-B-Citrine-PduA. A control construct, called C-PduA, containing the GS Linker, hexa-histidine tag and thrombin cleavage site, but without a coiled-coil sequence, was also made. The separate synthetic genes for these modified PduA variants were incorporated into a plasmid containing the remaining genes for the other shell proteins (PduB, B', J, K, N, U) required to form empty BMCs. In this way, a total of six plasmids encoding CC-Di-A-pduA-U, CC-Di-B-pduA-U, C-pduA-U, CC-Di-A-Citrine-pduA-U, CC-Di-B-Citrine-pduA-U and C-Citrine-pduA-U were transformed individually into *E. coli* BL21 * (DE3) cells. The resulting strains were grown, induced and analyzed by transmission electron microscopy (TEM) after fixation, embedding, thin sectioning and staining. TEM analysis revealed the presence of BMC structures within thin sections of cells from all six strains that were grown (Supplementary Figure 1).

The BMCs from the six strains were purified as outlined in methods and were subjected to SDS-PAGE analysis. All the purified BMCs produced banding patterns typical of a BMC, where the major bands visualized included PduB, PduB', PduJ, PduK and PduU. From the strains harbouring CC-Di-B-PduA and CC-Di-B-Citrine-PduA, the modified PduAs were readily discernable (Supplementary Figures 2 and 3). In contrast, purifications of BMCs isolated from strains producing with a CC-Di-A-PduA or C-PduA did not contain a band of the expected size for PduA suggesting that these variants are not incorporated into the BMC shell or that these fusion proteins are not produced by *E. coli* (Supplementary Figures 4 - 7).

To gain a better understanding of the structural characteristics of the modified compartments, the purified BMCs were fixed, embedded, thin sectioned and stained. TEM analysis revealed closed compartments in all samples, which had dimensions similar to native empty Pdu BMCs (Figure 1 and Supplementary Figure 8). However, the micrographs showed a number of structures within the purified samples, which appeared to be broken BMCs. It is possible that these damaged BMCs may have been broken during the purification process, which involves high-speed centrifugation. Lipids and protein aggregates were also visualized in purified samples, a consequence of the relatively crude purification process.

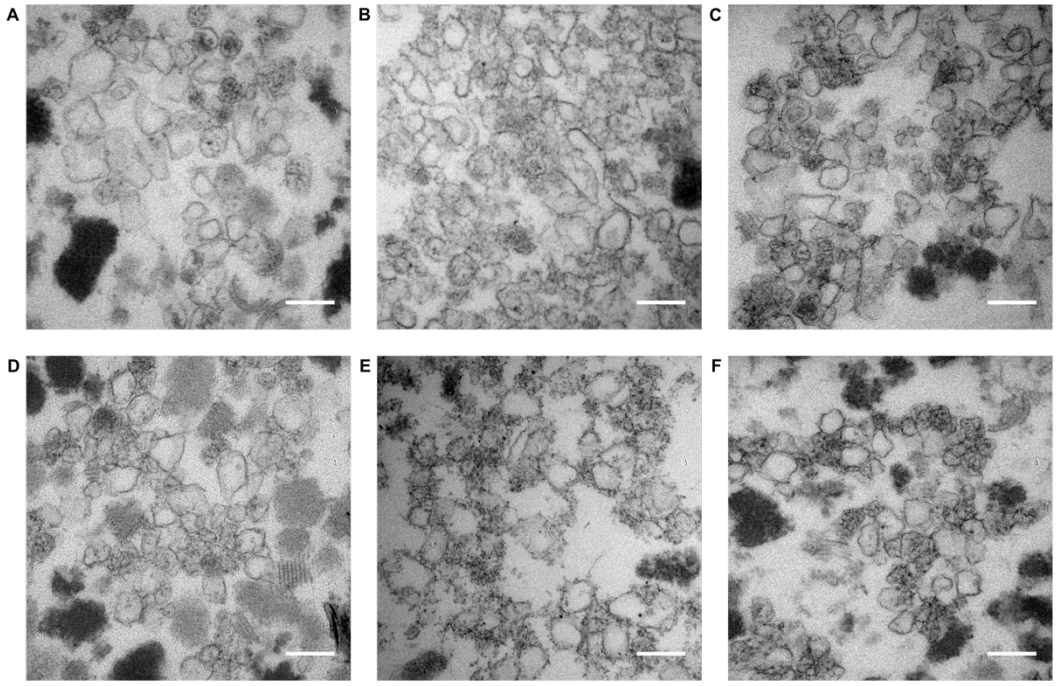


Figure 1 TEM analysis of thin-sectioned purified BMCs formed with the following modified PduA proteins. (**A**) CC-Di-A-PduA-U. (**B**) CC-Di-B-PduA-U. (**C**) C-PduA-U. (**D**) CC-Di-A-Citrine-PduA-U. (**E**) CC-Di-B-Citrine-PduA-U. (**F**) C-Citrine-PduA-U. Scale bars show 0.2 μm.

Next, fluorescently labeled BMC variants were examined by confocal microscopy to confirm that the modified PduAs were being incorporated into the shells of the BMCs. The strain producing the CC-Di-B-Citrine-PduA-U BMCs contained fluorescent puncta, indicating that CC-Di-B-Citrine-PduA was indeed incorporated into the BMC (Supplementary Figure 9 (B)). However, strains producing CC-Di-A-Citrine-PduA-U and C-Citrine-PduA-U did not show any such puncta, suggesting that CC-Di-A-Citrine-PduA and C-Citrine-PduA were not incorporated into BMCs. These results are consistent with the SDS-PAGE data, where the PduA variants were not observed in the latter (Supplementary Figure 9 (A and C)).

Although PduA is a major component of the Pdu BMC it is not essential for BMC formation in *S. enterica*[14]. This explains why it is possible to observe BMCs in samples even though PduA variants are not incorporated into the structure. To confirm this for our recombinant BMC system, we constructed a plasmid containing the genes required for BMC formation but lacking PduA; *i.e.*, with PduB, B', J, K, N and U. TEM analysis of the strain expressing PduB-U revealed correctly formed BMCs within the cytoplasm of *E. coli* (Figure 2).

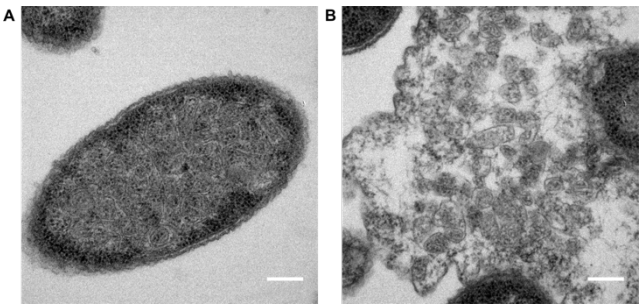


Figure 2 TEM analyses of *E. coli* producing the shell proteins PduB, B', J, K, N and U (PduB-U). (**A**) *In vivo*, and (**B**) from lysed cells. Scale bars show 0.2 µm

Targeting to BMCs *via* coiled-coil interactions. As mentioned previously, one problem with the natural targeting peptides, D18 and P18, is that, when fused to some cargo proteins, they can result in protein aggregation, presumably as a result of the amphipathic nature of these peptides[4, 6, 31]. To investigate if similar protein aggregation occurs with the coiled-coil peptides, CC-Di-A and CC-Di-B, the peptides were fused separately to the fluorescent protein mCherry. Cells expressing the corresponding synthetic genes in *E. coli* were visualized by confocal microscopy. For a control (C-mCherry) and for the CC-Di-A-mCherry fusion, this revealed dispersed cytoplasmic fluorescence indicative of a soluble protein (Supplementary Figures 10A and C). However, expression of CC-Di-B-mCherry resulted in protein aggregation as evidenced by fluorescent puncta in the cells, suggesting that some self-association of the CC-Di-B-based fusion (Supplementary Figure 10B)[5].

The CC-Di-B-PduA fusion protein appears to be incorporated into the empty BMC when co-produced with PduB, B', J, K, N and U. As the CC-Di-B is attached to the N-terminus of PduA, and as the N-terminus of PduA, in the hexamer, is on the concave side of the hexamer, it is likely that within the BMC the CC-Di-B peptide will be facing into the cytoplasm of the cell[26]. To test if the coiled-coil peptides were accessible to their cognate partner peptide, plasmids encoding CC-Di-B-Citrine-tagged BMCs were co-expressed with CC-Di-A-tagged mCherry. Confocal microscopy revealed co-localized fluorescent puncta within the bacterial cells, indicative of an association between the two-tagged components. By contrast, in a control strain producing the same CC-Di-B-Citrine-labeled BMCs and C-mCherry the mCherry signal remained cytoplasmic indicating no interaction with the BMC (Figure 3). Expression of CC-Di-B-tagged BMCs with CC-Di-B-tagged mCherry also resulted in co-localization of the florescence, but in this case the areas of colocalisation were large and more indicative of protein aggregation.

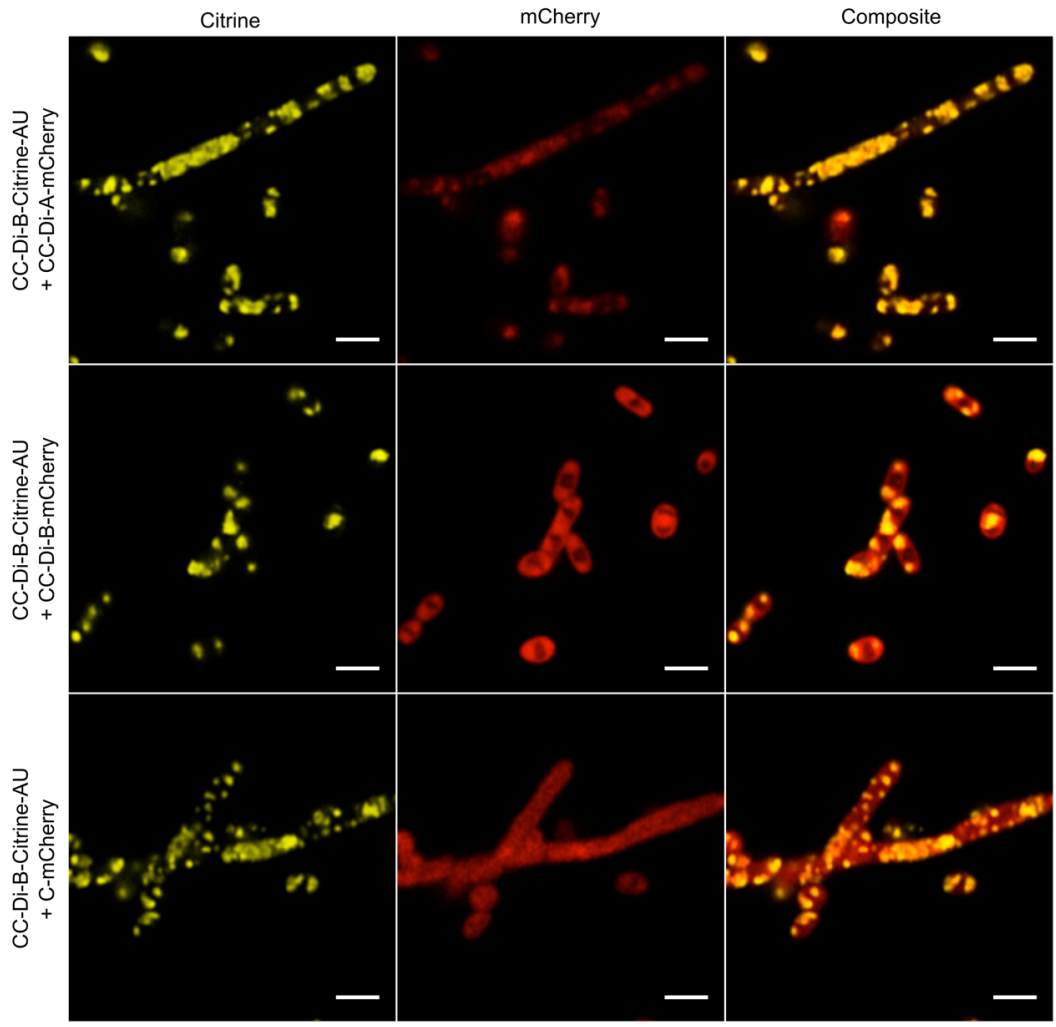


Figure 3 Citrine, mCherry and composite fluorescence signals for *E. coli* BL21 * (DE3) cells expressing CC-Di-B-Citrine-PduA-U with CC-Di-A-mCherry, CC-Di-B-mCherry or C-mCherry. Scale bars show 5 μm.

Redesign of PduA with a lumen-facing *N* terminus.

The BMC shell protein PduU contains the same characteristic BMC domain as PduA, however, it is circularly permuted; *i.e.*, the protein secondary structure elements are identical but are arranged in a different order, which results in the *N* and *C* termini being located on opposing faces of the hexamer[37, 38]. However, the low occupancy of PduU in recombinant BMC shells precludes its effective use in incorporating recombinant proteins into the BMC lumen *via* coiled-coil interactions[21]. Therefore, we aimed to circularly permute the abundant BMC shell protein PduA to make it analogous to PduU; *i.e.*, to relocate the *N* and *C* termini to the convex side of the hexameric disk whilst retaining the characteristic BMC domain architecture[39].

This redesign of PduA involved moving part of the *C* terminus to the *N* terminus. Based on the X-ray crystal structure of PduA[40], Val-68 was chosen as the site for the permutation: it is exposed on the surface of the hexamer, and we hypothesized that permutation of the segment after it should have minimal effect on the structure of the hexamer. To achieve this, the *C*-terminal region of PduA (GEVKAVHVIPRPHTDVEKILP) was grafted onto the *N* terminus, with a flexible linker (GSAGSGASG) used to connect the wildtype *N* and *C* termini, Figure 4.

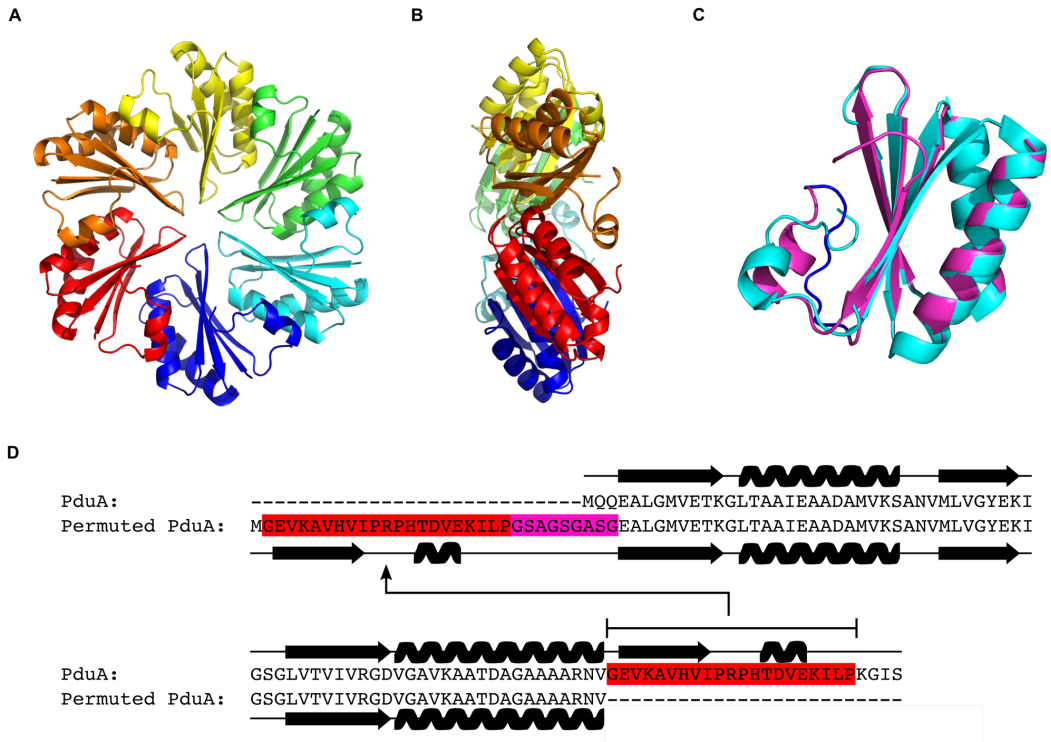


Figure 4 Circular permutation the shell protein PduA. (**A+B**) Architecture of the PduA hexamer (3NGK). (**C**) Computationally generated model of our novel permuted PduA (PduA^{P}) generated with Phyre2[41] (magenta) overlaid with the native structure (cyan) with the flexible linker joining the native N and C termini highlighted in blue. (**D**) Methodology for creating a circularly permuted PduA with the transplanted region highlighted in red and the linker region in cyan.

Synthetic DNA encoding the circularly permuted PduA (PduA^{P}) was inserted into a plasmid that allowed for its expression in combination with other shell proteins required to form an empty BMC shell (PduB, B', J, K, N, U) giving the construct pLysS-PduA^{P}-U. This was transformed into BL21* (DE3) competent cells and expressed by induction with IPTG. After overnight growth the cells were harvested, fixed, dehydrated, embedded, thin sectioned and subsequently analyzed by TEM.

TEM revealed the presence of BMCs *in vivo* that were structurally similar to empty BMCs made from wildtype proteins (PduA-U), showing that the permutation does not have a noticeable effect on the structure (Figure 5A&B). In addition, BMCs observed free in the resin, presumably as a result of some cell lysis, were morphologically similar to their wildtype PduA-containing counterparts (Figure 5C&D), providing further evidence that the circular permutation does not affect the structure of BMCs. However, attempts to purify BMCs containing PduA^{P}, using the previously well-established extraction procedures, failed. Presumably, this is because incorporation of PduA^{P} into BMCs changes their solubility characteristics. To demonstrate that PduA^{P} was incorporated into the BMCs *in vivo*, a fluorescent version of PduA^{P} was constructed incorporating the CC-Di-B-Citrine module onto its N terminus. Confocal microscopy of the strain producing this CC-Di-B-Citrine-PduA^{P}-U revealed florescent puncta, consistent with this fusion being successfully incorporated into the BMC shell (Supplementary Figure 11 (B)). In contrast, strains expressing CC-Di-A-Citrine-PduA^{P}-U or C-Citrine-PduA^{P}-U did not have a punctate fluorescent phenotype suggesting that these variants are not incorporated into the shell of the BMC as seen with the non-permuted variants (Supplementary Figure 11 (A and C)).

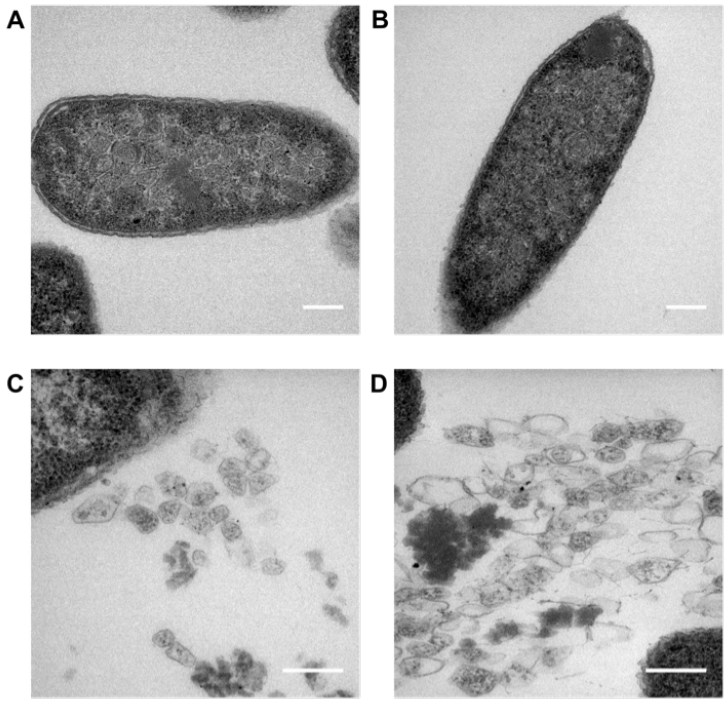


Figure 5 Comparison of wild type PduA BMCs and circularly permuted PduA^{P} BMCs. (**A + C**) PduA-U in-vivo and 'isolated' (**B + D**) PduA^{P}-U. Scale bars show 0.2 µm.

Targeting to BMCs containing PduA^{P}. To investigate targeting to PduA^{P}, we utilized the same CC-Di-AB system used with PduA[35, 36]. Addition of CC-Di-B to PduA^{P}, and in the presence of PduB-U, did not disrupt BMC formation (Supplementary Figure 12). However, we noted that some malformed BMCs were present in these strains, but only to a similar level as observed for the unmodified empty shell construct (PduA-U) (Supplementary Figure 8 and 12). BMCs that appeared to have been released through cell lysis offered better visualization of the structures, and these were morphologically similar to the unmodified control (PduA-U) and to permuted BMCs lacking the coiled-coil sequence (Figures 5 and 6).

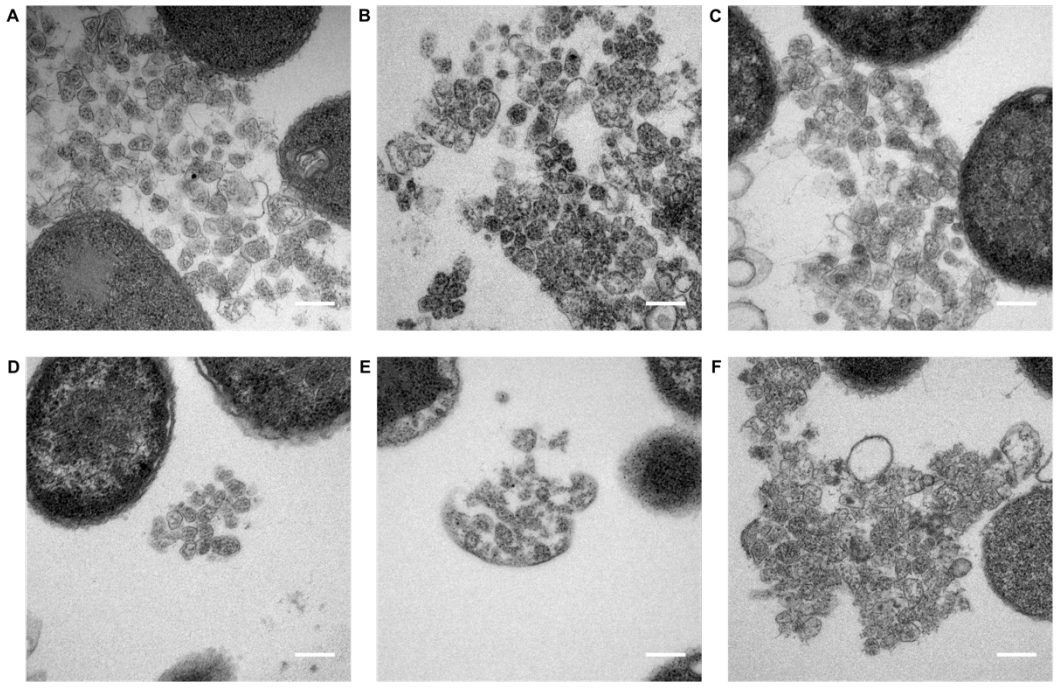


Figure 6 TEM analysis of permuted PduA BMCs tagged with coiled-peptides. (**A**) CC-Di-A-PduA^{P}-U. (**B**) CC-Di-B-PduA^{P}-U. (**C**) C-PduA^{P}-U. (**D**) CC-Di-A-Citrine-PduA^{P}-U. (**E**) CC-Di-B-Citrine-PduA^{P}-U. (**F**) C-Citrine-PduA^{P}-U. Scale bars show 0.2 µm.

Next, we sought to target cargo protein to the coiled-coil tagged PduA^{P} within the BMC. Strains were transformed with a plasmid encoding CC-Di-B-Citrine-PduA^{P}-U and a second plasmid producing the fluorescent protein reporter (mCherry) tagged with either CC-Di-A, CC-Di-B or lacking a coiled-coil sequence (C-mCherry). Strains were grown and analyzed by confocal fluorescence microscopy (Figure 7). Co-

localized fluorescence signals were apparent in the strain expressing CC-Di-B-Citrine-PduA^{P}-U and CC-Di-A-mCherry, although it must be noted that a high amount of cytoplasmic CC-Di-A-mCherry was also present, suggesting that not all of the protein is targeted to the BMC (lumen). A wholly cytoplasmic mCherry signal was observed in the strain producing CC-Di-B-Citrine-PduA^{P}-U together with untagged mCherry (C-mCherry) (Figure 7). Together, these results provide strong evidence that it is possible to utilize *de novo* designed coiled-coil peptides to target to the luminal side of BMCs.

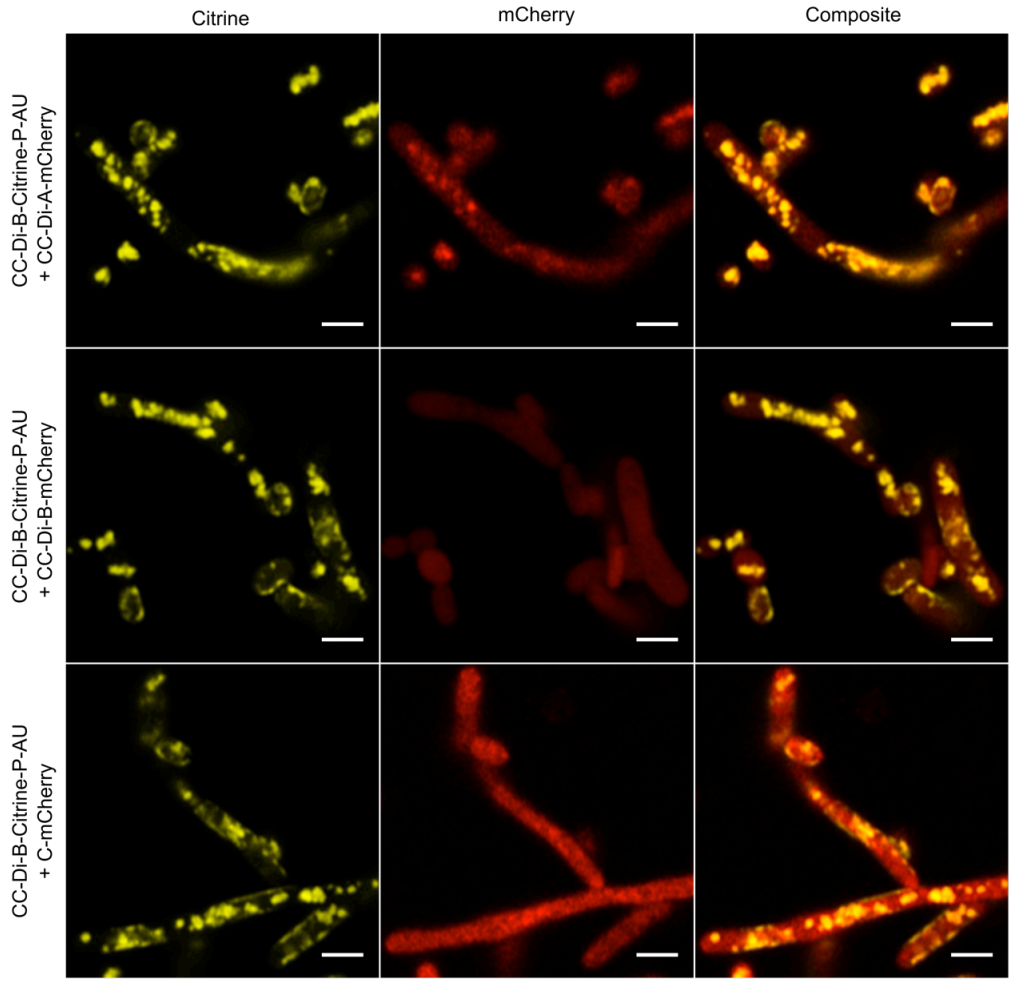


Figure 7 Citrine, mCherry and composite fluorescence signals from *E. coli* BL21 * (DE3) cells expressing CC-Di-B-Citrine-PduA^{P}-U in conjunction with CC-Di-A-mCherry, CC-Di-B-mCherry or C-mCherry. Scale bars show 5 µm.

Conclusion

Previously, we have shown that a *de novo* designed and highly specific coiled-coil peptide pair, termed CC-Di-A and CC-Di-B, can be used to target proteins to filamentous tubules formed in the cytoplasm of *E. coli*. These tubules, which we collectively call a cytoscaffold, are formed by the overproduction of a soluble form of the natural bacterial microcompartment (BMC) protein PduA, termed PduA*[5]. With this system, we find that CC-Di-B-PduA* filaments find protein cargoes tagged with CC-Di-A, and the results are consistent with the N terminus of PduA* facing the cellular cytoplasm.

Therefore, and as detailed herein, we wondered if the same coiled-coil system could be used to direct cargo to an empty recombinant BMC within *E. coli*. Such BMCs are formed from the coordinated production of multiple shell proteins, PduA-U. To achieve our goal, we investigated the effect of adding CC-Di-A or CC-Di-B to the N terminus of wildtype PduA, and co-producing these variants with the remaining shell proteins. We find that only the CC-Di-B-PduA variant incorporates into the recombinant BMCs. It appears that CC-Di-A makes the PduA unstable, which is also what we observe for tagged versions of PduA*[5].

By coproducing CC-Di-B-PduA with PduB-U, intact BMCs are observed by TEM. Similar BMCs are also observed for a CC-Di-B-Citrine-PduA version produced with PduB-U. In this case, the presence of BMCs is confirmed by confocal microscopy, which reveals fluorescent puncta within the cellular cytoplasm. The ability of the CC-Di-A peptide to bind to the CC-Di-B-PduA variants is demonstrated *in vivo* with a CC-Di-A-mCherry fusion protein: co-production of this with the BMC containing the CC-Di-B-Citrine-PduA gives co-localized colored puncta within cells. In these cases, on the basis of the X-ray crystal structure of PduA and its likely incorporation into the BMC, we suggest that the mCherry is being directed to the outer, cytoplasm-facing sides of the BMCs.

Therefore, and following on from this, we speculated the coiled-coil peptides could be used to target to the luminal side of the BMC through the redesign of PduA to relocate the *N* terminus to the convex side of the hexamer. To achieve this, we permutated PduA after residue 68 to maintain its overall structure but to move the *N* and *C* termini to the opposite side of the molecule. This permutated form of PduA, PduA^{P}, forms recombinant BMCs *in vivo* when co-produced with PduB-U, although

these could not be recovered for detailed *in vitro* analysis, suggesting that the PduA^{P} changes the physical properties of the BMC in some way.

Nonetheless, both CC-Di-B and CC-Di-B-Citrine can be fused to PduA^{P} to render competent recombinant BMCs. Indeed, the co-production of CC-Di-B-Citrine-PduA^{P} with PduB-U gives fluorescent puncta within the cells, consistent with the incorporation of PduA^{P} into the BMC. Again, targeting to the CC-Di-B-PduA^{P}-containing BMCs is demonstrated with a CC-Di-A-mCherry fusion and the co-localisation of citrine and mCherry signals. This is consistent with the aim that cargo protein is being incorporated within the lumens of the BMCs.

In summary, we have developed methods for the localization of proteins to both the outside and the inside of BMCs. Targeting to the outside could be used for display of proteins on a comparatively large protein framework, for example, in the generation of antibodies. The internal accommodation of proteins can be used to engineer new metabolic pathways within the BMC; and such endowed BMC could be used as delivery vehicles. Through these approaches we have shown that the field of cellular and metabolic engineering depends not only on the implementation of natural architectures but also on *de novo* design redesign, in order to achieve maximal functionality.

Methods

Cloning of constructs

DNA was synthesized by Eurofins Genomics, as described in Supplementary Information. Plasmids were constructed as outlined in Supplementary Table 2 with primers listed in Supplementary Table 3.

Growth of strains producing BMCs

BL21* (DE3) competent cells were transformed with appropriate plasmids. 50 ml of LB supplemented with ampicillin (100 mg/L) and chloramphenicol (34 mg/L) in baffled flasks was inoculated from an overnight starter culture. Cells were grown at 37 °C with shaking to an OD_{600} ~ 0.4, protein production was induced by addition of IPTG to a final concentration of 400 µM. Cultures were then incubated for 2 hours at 19 °C with shaking for confocal analysis or overnight with shaking for TEM analysis and purification.

Cell lysis

Cells were harvested by centrifugation at 2683 x *g*, a one gram wet cell pellet was resuspended in 20 ml YPER supplemented with Protease Inhibitor Cocktail Tablets, EDTA-Free and 500 Units benzonase nuclease and incubated for 3 hours at room temperature with gentle shaking.

Purification of CC-Di-A tagged and untagged BMCs

BMCs were pelleted from the lysate by centrifugation for 5 minutes at 11,300 X *g*, the pellet was resuspended in 2 ml 20 mM Tris-HCl pH 8, 20 mM NaCl. The suspension was centrifuged at 4 °C for 5 minutes at 11,000 X *g*. The supernatant, containing microcompartments was collected, the NaCl concentration was increased to 80 mM. The suspension was centrifuged at 4 °C for 5 minutes at 11,000 X *g*, the resulting pellet contained BMCs.

Purification of CC-Di-B tagged BMCs

BMCs were pelleted from the lysate by centrifugation for 5 minutes at 11,300 X *g*, the pellet was resuspended in 2 ml 20 mM Tris-HCl pH 8, 20 mM NaCl. The suspension was centrifuged at 4 °C for 5 minutes at 11,000 X *g*. The resulting BMC containing pellet was resuspended in 2 ml 20 mM Tris-HCl pH 8 and centrifuged for 5 minutes at 11,000 X *g*, the supernatant was collected, the NaCl concentration was increased to 80 mM. The suspension was centrifuged as above. The resulting pellet was resuspended in an appropriate volume of 20 mM Tris-HCl pH 8 for analysis.

Embedding of samples for TEM analysis

Following growth and induction of protein expression cells were harvested by centrifugation at 3000 x *g*. For embedding of BMCs the final pellet from the purification was used in place of a cell pellet. The cell/BMC pellet was resuspended in 2 ml 2.5% glutaraldehyde in 100 mM sodium cacodylate buffer pH 7.2 (CAB) and fixed for 2 hours with gentle rotating (20 rpm). Cells were pelleted by centrifugation at 6000 x *g* for 2 minutes and were washed twice for 10 minutes with 100 mM CAB. Cells were post-fixed with 1% osmium tetroxide in 100 mM CAB for 2 hours and subsequently washed twice with ddH_2O. Cells were dehydrated by incubation in an ethanol gradient, 50% EtOH for 10 minutes, 70% EtOH overnight, 90% EtOH for 10 minutes followed by three 10 minute washes in 100% dry EtOH. Cells were then washed twice with propylene oxide for 15 minutes. Cell pellets were embedded by resuspension in 1 ml of a 1:1 mix of propylene oxide and Agar LV Resin and incubated for 30 minutes with rotation. Cell pellets were infiltrated twice in 100% Agar LV resin. The cell pellet was re-suspended in fresh resin and transferred to a 1ml Beem embedding capsule, centrifuged for 5 minutes at 3000 x *g* to concentrate the cells to the tip of the mould and incubated for 20 hours at 60 °C to polymerize.

Sectioning and visualization of samples

Samples were ultra-thin sectioned on a RMC MT-XL ultra-microtome with a diamond knife (diatome 45°) sections (60 – 70 nm) were collected on un-coated 300 mesh copper grids. Grids were stained by incubation in 4.5% uranyl acetate in 1% acetic acid solution for 45 minutes followed by washing in a stream of ddH_2O. Grids were then stained with Reynolds lead citrate for 7 minutes followed by washing in a stream of ddH_2O

Electron microscopy was performed using a JEOL-1230 transmission electron microscope equipped with a Gatan multiscan digital camera operated at an accelerating voltage of 80 kV

Confocal imaging

Following growth and induction of protein expression 1 ml of cells was harvested by centrifugation at 3000 x *g*. The resulting cell pellet was washed 3 times in PBS before incubation for 15 minutes in 2% (w/v) formaldehyde in PBS, cells were then washed a further 3 times in PBS. Cells (10 µl) were pipetted onto a 1.5 thickness coverslip before being inverted onto a drop of ProLong Gold antifade mountant (Life Technologies) on a glass slide slides were incubated at room temperature in the dark for 24 hours to cure.

Images were acquired on a Zeiss LSM 880 with Airyscan system. Excitation light (514 nm for Citrine or 561 nm for mCherry) was provided by an argon lamp (514 nm) or HeNe Laser (561). Images were acquired using a 100x 1.46 NA oil immersion objective lens.

Acknowledgements

We are grateful to the Biotechnology and Biological Sciences Research Council of the UK for a strategic LoLa Award to MJW, DNW and RWP (BB/M002969/1). DNW holds a Royal Society Wolfson Research Merit Award. We thank the entire BMC-SAGE LoLa group for helpful discussions.

Author contributions

MJL conducted all experimental work, excluding the thin sectioning of samples for TEM analysis (IRB). MJL, JMF, RWP, DNW, SF and MJW designed the experiments. MJL wrote the manuscript, which was edited by all authors

Conflict of interest

The authors declare that they have no conflict of interest

Supplementary Information for:

De novo targeting to the cytoplasmic and luminal side of bacterial microcompartments

Authors: Matthew J. Lee[1], Ian R. Brown[1], Jordan M Fletcher[2], Richard W. Pickersgill[3], Derek N Woolfson[2,4,5], Stefanie Frank[6]*, Martin J Warren[1]*

[1]Industrial Biotechnology Centre, School of Biosciences, University of Kent, Canterbury CT2 7NJ, UK

[2]School of Chemistry, University of Bristol, Cantock's Close, Bristol BS8 1TS, UK.

[3]School of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS

[4]School of Biochemistry, University of Bristol, Medical Sciences Building, University Walk, Bristol BS8 1TD, UK.

[5]BrisSynBio, Life Sciences Building, Tyndall Avenue, Bristol BS8 1TQ, UK.

[6]Department of Biochemical Engineering, University College London, Bernard Katz Building, Gordon Street, London WC1E 6BT, UK.

* Correspondence to: Martin J Warren (M.J. Warren@kent.ac.uk) and Stefanie Frank (Stefanie.frank@ucl.ac.uk)

Synthesized DNA – cloned into *Xba*I/*Nde*I sites of pET14b

CC-Di-A-His

TCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATATGGGTAGCGAAATTGC
CGCGCTGGAAAAAGAAAACGCTGCCCTGGAGCAGGAAATCGCGGCACTGGAG
CAGGGTAGCGGCAGCGGTAGCACCATGGGCAGCAGCCATCATCATCATCATC
ACAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATG

CC-Di-B-His

TCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATATGGGTAGCAAAATCGC
CGCACTGAAAAAGAAAAACGCTGCGCTGAAACAGAAAATTGCCGCGCTGAAAC
AGGGTAGCGGCAGCGGTAGCACCATGGGCAGCAGCCATCATCATCATCATCA
CAGCAGCGGCCTGGTGCCGCGCGGCAGCCATATG

C-His

TCTAGAAATAATTTTGTTTAACTTTAAGAAGGAGATATATGGGTAGCGGTAGCG
GCAGCGGTAGCACCATGGGCAGCAGCCATCATCATCATCATCACAGCAGCGG
CCTGGTGCCGCGCGGCAGCCATATG

Synthesized permuted PduA – cloned into *Nde*I/*Bam*HI sites of pET3a

Perm-PduA-GSA-PduB

CATATGGGAGAAGTGAAAGCCGTACACGTCATCCCACGCCCTCACACCGATGT
AGAAAAAATCTTACCGGGTAGCGCAGGCAGTGGGGCAAGCGGTGAAGCGTTA
GGAATGGTAGAAACCAAAGGCTTGACTGCAGCCATAGAGGCCGCAGATGCAAT
GGTGAAGTCAGCCAATGTAATGCTGGTCGGCTACGAAAAAATTGGTTCGGGGC
TGGTAACAGTCATTGTCCGCGGCGATGTTGGCGCAGTCAAAGCAGCAACAGAT
GCAGGTGCCGCCGCAGCACGTAATGTGTAAACTAGTATCTTACCGAAGGGAAT
TAGCTAATGAGCAGCAATGAGCTGGTTGATCAGATCATGGCGCAGGTGATTGC
TCGCGTGGCAACGCCGGAACAGCAGGCTATCCCTGAAAATAATCCTCCAACAC
GAGAAACGGCTATGGCAGAGAAAAGCTGCAGTTTAACGGAGTTTGTCGGTACT
GCGATTGGCGACACCGTCGGTCTGGTAATCGCCAACGTGGACAGCGCCCTAC
TGGACGCAATGAAACTTGAAAAACGGTATCGCTCCATTGGCATCCTTGGCGCG
CGTACTGGTGCAGGCCCGCACATCATGGCCGCAGATGAAGCGGTAAAAGCCA
CCAATACTGAAGTCGTCAGTATTGAGTTGCCACGTGATACCAAAGGCGGCGCG
GGTCACGGTTCGCTGATTATTCTCGGCGGCAACGATGTTTCCGACGTGAAACG
CGGAATTGAGGTTGCGCTGAAAGAACTGGATCGCACCTTTGGCGATGTGTATG

CCAACGAAGCCGGTCACATCGAGATGCAATACACCGCACGCGCCAGCTACGC
GCTGGAAAAAGCCTTTGGTGCACCGATTGGCCGTGCCTGTGGCGTGATCGTC
GGCGCGCCGGATCC

N-Terminal Coiled-coil peptides, coiled-coil regions underlined

CC-Di-A:
MGSEIAALEKENAALEQEIAALEQGSGSGSTMGSSHHHHHHSSGLVPRGSHM

CC-Di-B:
MGSKIAALKKKNAALKQKIAALKQGSGSGSTMGSSHHHHHHSSGLVPRGSHM

C:
MGSGSGSGSTMGSSHHHHHHSSGLVPRGSHM

Supplementary Table 1 Strains used in this study

Strain	Genotype	Source
JM109	*end*A1, *rec*A1, *gyr*A96, *thi*, *hsd*R17 (r_k^-, m_k^+), *rel*A1, *sup*E44, Δ(*lac-pro*AB), [F′, *tra*D36, *pro*AB, *laq*I^qZΔM15]	Promega
BL21 * (DE3)	F^- *omp*T *hsd*SB (r_B^- m_B^-) *gal dcm* (DE3)	Novagen

Suppementary Table 2 Plasmids Used in this study

Plasmid name	Description	Source
pET3a	Overexpression vector, modified to include an *Spe*I site 5' of *Bam*HI	Novagen
pET14b	Overexpression vector containing N-terminal hexahistidine-tag, modified to include an *Spe*I site 5' of *Bam*HI	Novagen
pLysS	Basal expression suppressor	Novagen
pLysS_2	Basal expression suppressor containing the T7 promoter-MCS-T7 terminator cassette from pET14b	Parsons *et al*., 2008
pLysS_PduABJKNU	pLysS-2 containing genes required for the formation of empty BMCs	Parsons *et al*., 2010
pLysS_TBAD_mCherry_PduABJKNU		
pET_CC_Di_A	Synthesised Di-A gene fragment cloned into *Xba*I/*Nde*I sites of pET14b	This study
pET_CC_Di_B	Synthesised Di-B gene fragment cloned into *Xba*I/*Nde*I sites of pET14b	This study
pET_ C	Synthesised Di-C gene fragment cloned into *Xba*I/*Nde*I sites of pET14b	This study
pET3a_PduAB	pET3a overexpression vector containing the genes PduA and PduB	
pET_CC_Di_A_PduAB	*Nde*I/*Spe*I fragment of pET3a_PduAB cloned into *Nde*I/*Spe*I sites of pET_Di_A	This study
pET_CC_Di_B_PduAB	*Nde*I/*Spe*I fragment of pET3a_PduAB cloned into *Nde*I/*Spe*I sites of pET_Di_B	This study

pET_ C_PduAB	*Nde*I/*Spe*I fragment of pET3a_PduAB cloned into *Nde*I/*Spe*I sites of pET_Di_C	This study
pLysS_CC_Di_A_PduABJKNU	*Bgl*II/*Asc*I fragment of pET_Di_A_PduAB cloned into *Bgl*II/*Asc*I sites of pLysS_TBAD_mCherry_PduAB JKNU	This study
pLysS_CC_Di_B_PduABJKNU	*Bgl*II/*Asc*I fragment of pET_Di_B_PduAB cloned into *Bgl*II/*Asc*I sites of pLysS_TBAD_mCherry_PduAB JKNU	This study
pLysS_C_PduABJKNU	*Bgl*II/*Asc*I fragment of pET_Di_C_PduAB cloned into *Bgl*II/*Asc*I sites of pLysS_TBAD_mCherry_PduAB JKNU	This study
pET_CC_Di_A_Citrine_NoStop	Citrine PCR product digested with *Ase*I/*Spe*I and ligated into *Nde*I/*Spe*I sites of pET_Di_A	This study
pET_CC_Di_B_Citrine_NoStop	Citrine PCR product digested with *Ase*I/*Spe*I and ligated into *Nde*I/*Spe*I sites of pET_Di_B	This study
pET_C_Citrine_NoStop	Citrine PCR product digested with *Ase*I/*Spe*I and ligated into *Nde*I/*Spe*I sites of pET_Di_C	This study
pET_CC_Di_A_Citrine_PduAB	*Nde*I/*Spe*I fragment of pET3a_PduAB cloned into *Nde*I/*Spe*I sites of pET_Di_A_Citrine_NoStop	This study
pET_CC_Di_B_Citrine_PduAB	*Nde*I/*Spe*I fragment of pET3a_PduAB cloned into *Nde*I/*Spe*I sites of pET_Di_B_Citrine_NoStop	This study
pET_C_Citrine_PduAB	*Nde*I/*Spe*I fragment of	This study

	pET3a_PduAB cloned into *Nde*I/*Spe*I sites of pET_Di_C_Citrine_NoStop	
pLysS_CC_Di_A_Citrine_PduABJKNU	*Bgl*II/*Asc*I fragment of pET_Di_A_Citrine_PduAB cloned into *Bgl*II/*Asc*I sites of pLysS_TBAD_mCherry_PduAB JKNU	This study
pLysS_CC_Di_B_Citrine_PduABJKNU	*Bgl*II/*Asc*I fragment of pET_Di_B_Citrine_PduAB cloned into *Bgl*II/*Asc*I sites of pLysS_TBAD_mCherry_PduAB JKNU	This study
pLysS_C_Citrine_PduABJKNU	*Bgl*II/*Asc*I fragment of pET_Di_C_Citrine_PduAB cloned into *Bgl*II/*Asc*I sites of pLysS_TBAD_mCherry_PduAB JKNU	This study
pET_CC_Di_A_mCherry	mCherry PCR product ligated into *Nde*I/*Spe*I sites of pET_Di_A	This study
pET_CC_Di_B_mCherry	mCherry PCR product ligated into *Nde*I/*Spe*I sites of pET_Di_B	This study
pET_C_mCherry	mCherry PCR product ligated into *Nde*I/*Spe*I sites of pET_Di_C	This study
pET-PduB	*Spe*I/*Bam*HI fragment of pET_Perm_PduA_GSA_PduB ligated into *XbaI*/*Bam*HI sites of pET_CC_Di_A_PduAB	This study
pLysS_PduBJKNU	*Bgl*II/*Asc*I fragment of pET-PduB cloned into *Bgl*II/*Asc*I sites of pLysS_TBAD_mCherry_PduAB JKNU	This study
pET_Perm_PduA_GSA_PduB	Synthesised Perm-PduA-GSA-	This study

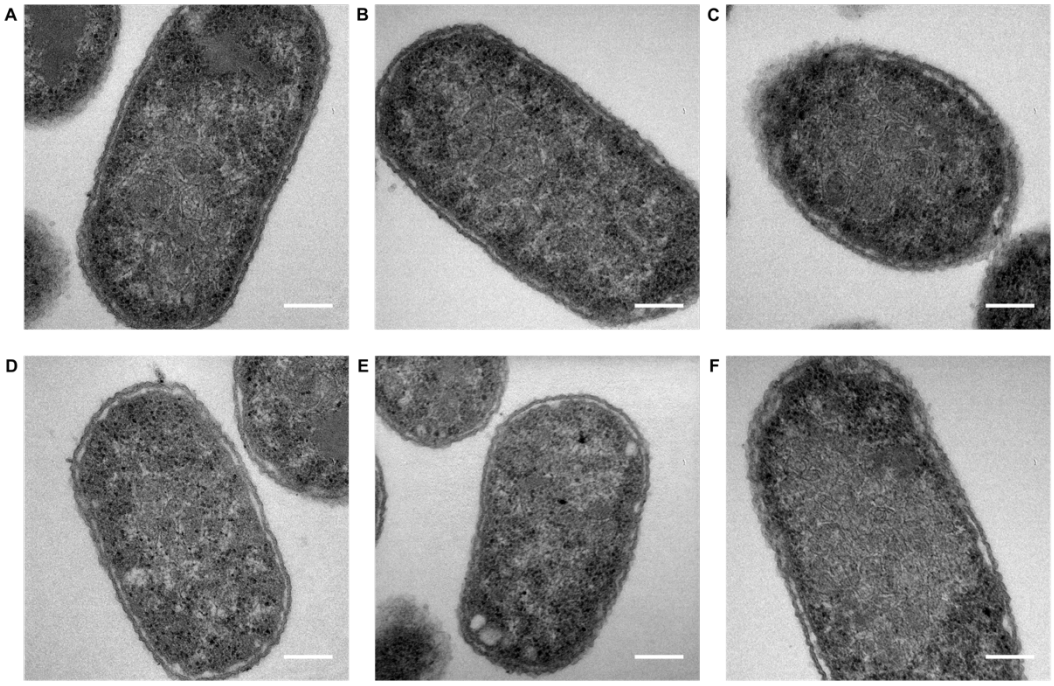
	PduB gene fragment cloned into *Nde*I/*Bam*HI sites of pET3a	
pLysS_Perm_PduA_GSA_PduBJKNU	*Bgl*II/*Asc*I fragment of pET_ Perm_PduA_GSA_PduB cloned into *Bgl*II/*Asc*I sites of pLysS_TBAD_mCherry_PduAB JKNU	This study
pET_CC_Di_A_Perm_PduA_GSA_PduB	*Nde*I/*Bam*HI fragment of pET_Perm_PduA_GSA_PduB ligated into *Nde*I/*Bam*HI sites of pET_CC_Di_A_PduAB	This study
pET_CC_Di_B_Perm_PduA_GSA_PduB	*Nde*I/*Bam*HI fragment of pET_Perm_PduA_GSA_PduB ligated into *Nde*I/*Bam*HI sites of pET_CC_Di_A_PduAB	This study
pET_C_Perm_PduA_GSA_PduB	*Nde*I/*Bam*HI fragment of pET_CC_Di_A_PduAB ligated into *Nde*I/*Bam*HI sites of pET_C_Citrine_PduAB	This study
pET_CC_Di_A_Citrine_Perm_PduA_GSA _PduB	*Nde*I/*Bam*HI fragment of pET_Perm_PduA_GSA_PduB ligated into *Nde*I/*Bam*HI sites of pET_CC_Di_A_Citrine_PduAB	This study
pET_CC_Di_B_Citrine_Perm_PduA_GSA _PduB	*Nde*I/*Bam*HI fragment of pET_Perm_PduA_GSA_PduB ligated into *Nde*I/*Bam*HI sites of pET_CC_Di_B_Citrine_PduAB	This study
pET_C_Citrine_Perm_PduA_GSA_PduB	*Nde*I/*Bam*HI fragment of pET_Perm_PduA_GSA_PduB ligated into *Nde*I/*Bam*HI sites of pET_C_Citrine_PduAB	This study
pLysS_CC_Di_A_Perm_PduA_GSA_Pdu BJKNU	*Bgl*II/*Asc*I fragment of pET_CC_Di_A_Perm_PduA_GS A_PduB cloned into *Bgl*II/*Asc*I sites of pLysS_TBAD_mCherry_PduAB	This study

	JKNU	
pLysS_CC_Di_B_Perm_PduA_GSA_PduBJKNU	*Bgl*II/*Asc*I fragment of pET_CC_Di_B_Perm_PduA_GSA_PduB cloned into *Bgl*II/*Asc*I sites of pLysS_TBAD_mCherry_PduABJKNU	This study
pLysS_C_Perm_PduA_GSA_PduBJKNU	*Bgl*II/*Asc*I fragment of pET_C_Perm_PduA_GSA_PduB cloned into *Bgl*II/*Asc*I sites of pLysS_TBAD_mCherry_PduABJKNU	This study
pLysS_CC_Di_A_Citrine_Perm_PduA_GSA_PduBJKNU	*Bgl*II/*Asc*I fragment of pET_CC_Di_A_Citrine_Perm_PduA_GSA_PduB cloned into *Bgl*II/*Asc*I sites of pLysS_TBAD_mCherry_PduABJKNU	This study
pLysS_CC_Di_B_Citrine_Perm_PduA_GSA_PduBJKNU	*Bgl*II/*Asc*I fragment of pET_CC_Di_B_Citrine_Perm_PduA_GSA_PduB cloned into *Bgl*II/*Asc*I sites of pLysS_TBAD_mCherry_PduABJKNU	This study
pLysS_C_Perm_Citrine_PduA_GSA_PduBJKNU	*Bgl*II/*Asc*I fragment of pET_C_Citrine_Perm_PduA_GSA_PduB cloned into *Bgl*II/*Asc*I sites of pLysS_TBAD_mCherry_PduABJKNU	This study

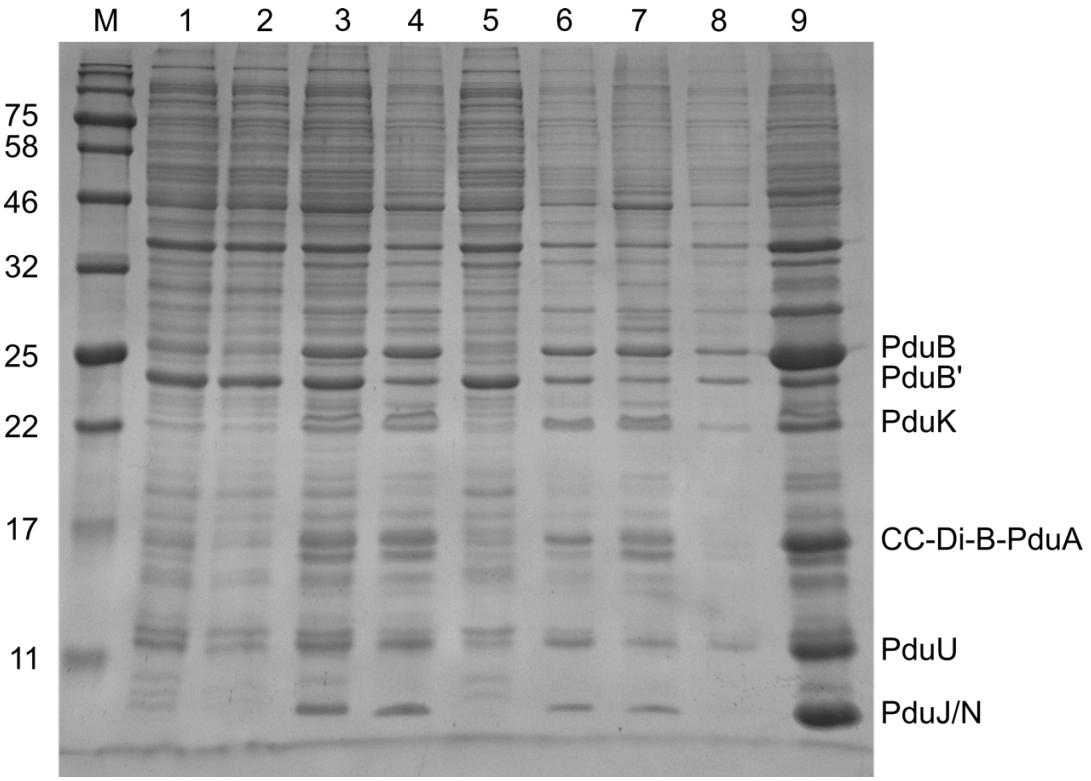
Supplementary Table 3: Oligonucleotides used in this study, restriction sites are underlined

Name	Sequence 5' – 3'
Citrine_*Ase*I_FW	GACATTAATATGGTGAGCAAGGGCGAGGAGCTG
Citrine_*Spe*I_NdeI_RV	GACACTAGTAAACATATGCTTGTACAGCTCGTCCATGCCGAG
Citrine_*Nde*I_FW	CACCATATGGTGAGCAAGGGCGAGGAGC
Citrine_*Spe*I_RV	CACACTAGTTTACTTGTACAGCTCGTCC
mCherry_*Nde*I_FW	CAGCATATGGTGAGCAAGGGCGAGG
mCherry_*Spe*I_RV	GACACTAGTTTACTTGTACAGCTCGTCCATGC

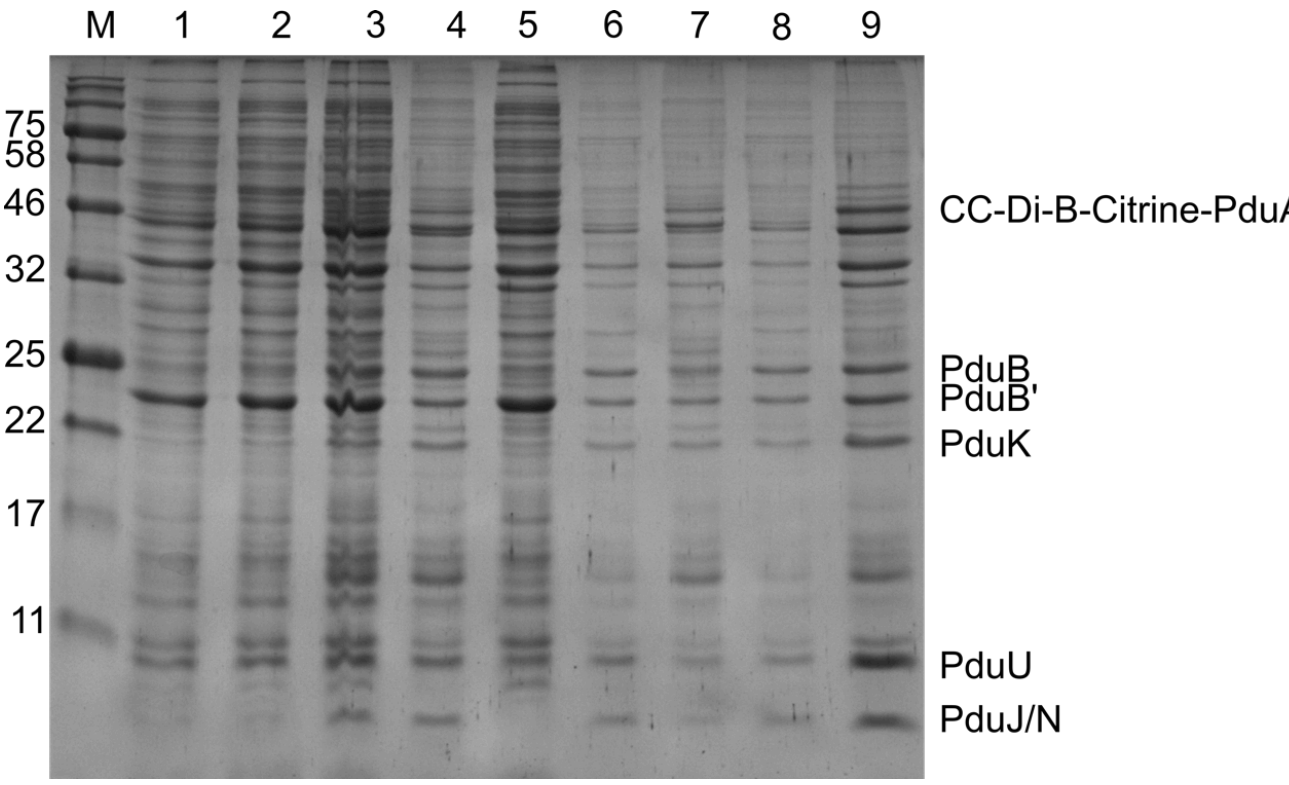
Supplementary Figures



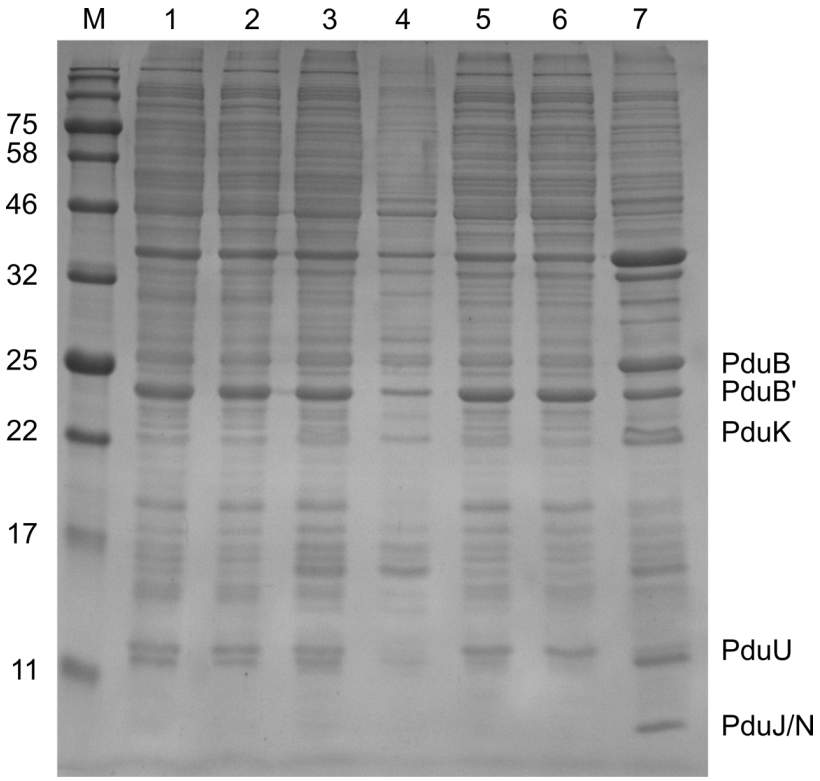
Supplementary Figure 1 TEM analysis of *E. coli* BL21 * (DE3) cells expressing: (**A**) CC-Di-A-PduA-U (**B**) CC-Di-B-PduA-U (**C**) C-PduA-U (**D**) CC-Di-A-Citrine-PduA-U (**E**) CC-Di-B-Citrine-PduA-U (**F**) C-Citrine-PduA-U. Scale bars show 0.2 µm.



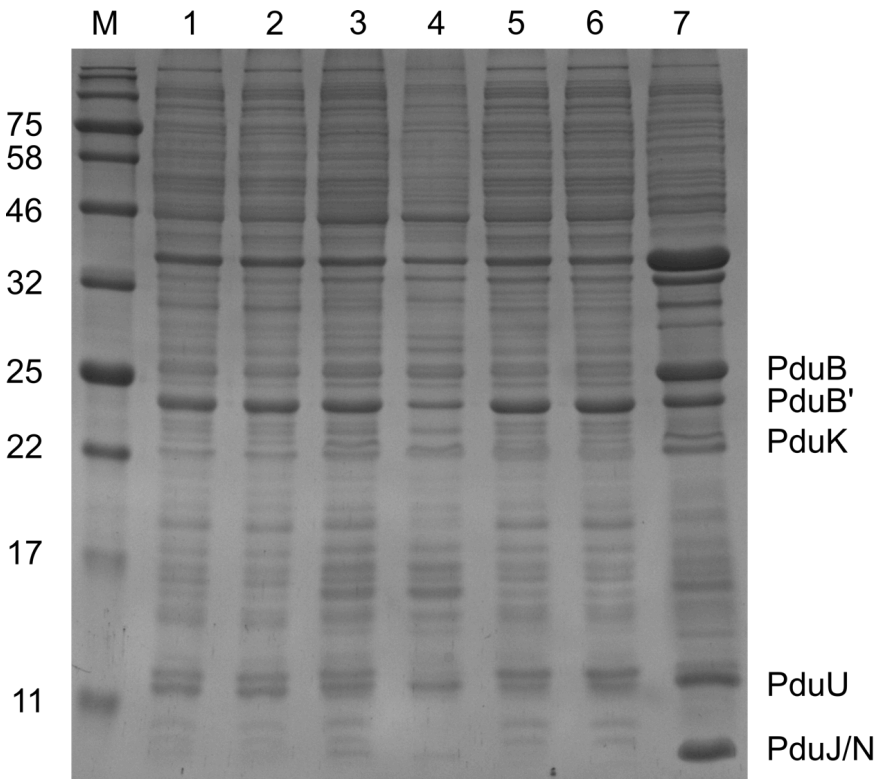
Supplementary Figure 2 SDS-PAGE analysis of CC-Di-B labeled BMC purification in comparison to a molecular weight marker (M). Lane 1 – lysate, lane 2 – supernatant after centrifugation at 11,300 x g, lane 3 – pellet after centrifugation (resuspended in 20 mM Tris – HCl pH 8, 20 mM NaCl), lane 4 – pellet after centrifugation at 11,000 x g (resuspended in 20 mM Tris – HCl pH 8, 20 mM NaCl), lane 5 – supernatant after centrifugation, lane 6 - supernatant after centrifugation of resuspended pellet, lane 7 – pellet after centrifugation of resuspended pellet, lane 8 - supernatant after centrifugation, lane 9 pellet after centrifugation in 80 mM NaCl, resuspended in 20 mM Tris-HCl pH 8. 2 µl of sample was loaded in lanes 1 and 2, lanes 3 to 9 were loaded with 5 µl of sample.



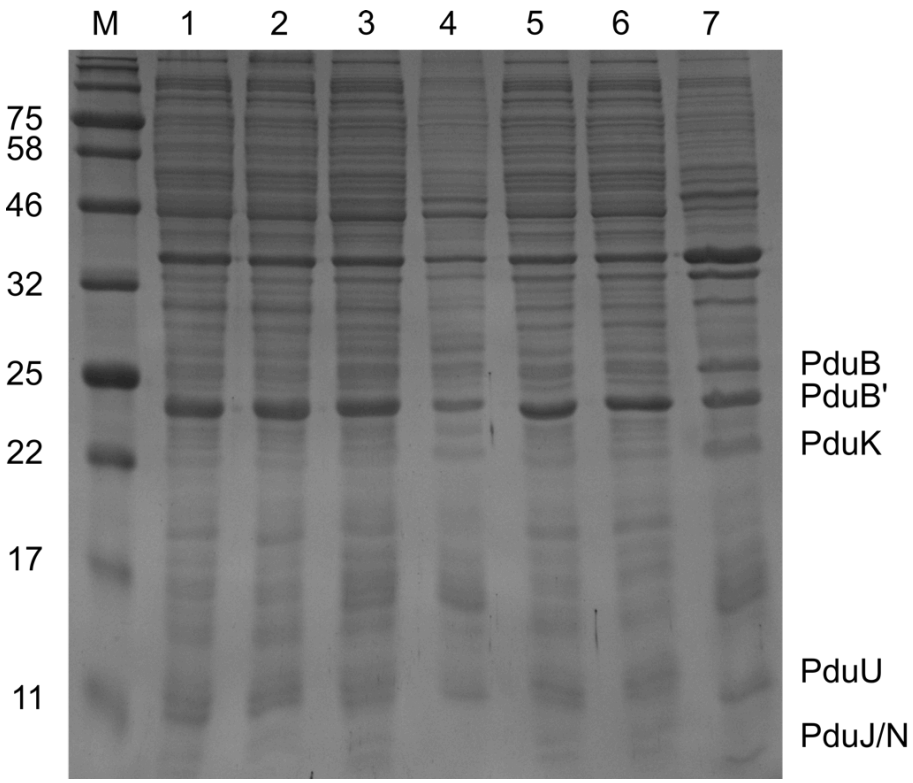
Supplementary Figure 3 SDS-PAGE analysis of CC-Di-B-Citrine labeled BMC purification in comparison to a molecular weight marker (M). Lane 1 – lysate, lane 2 – supernatant after centrifugation at 11,300 x g, lane 3 – pellet after centrifugation (resuspended in 20 mM Tris – HCl pH 8, 20 mM NaCl), lane 4 – pellet after centrifugation at 11,000 x g (resuspended in 20 mM Tris – HCl pH 8, 20 mM NaCl), lane 5 – supernatant after centrifugation, lane 6 - supernatant after centrifugation of resuspended pellet, lane 7 – pellet after centrifugation of resuspended pellet, lane 8 - supernatant after centrifugation, lane 9 pellet after centrifugation in 80 mM NaCl, resuspended in 20 mM Tris-HCl pH 8. 2 µl of sample was loaded in lanes 1 and 2, lanes 3 to 9 were loaded with 5 µl of sample.



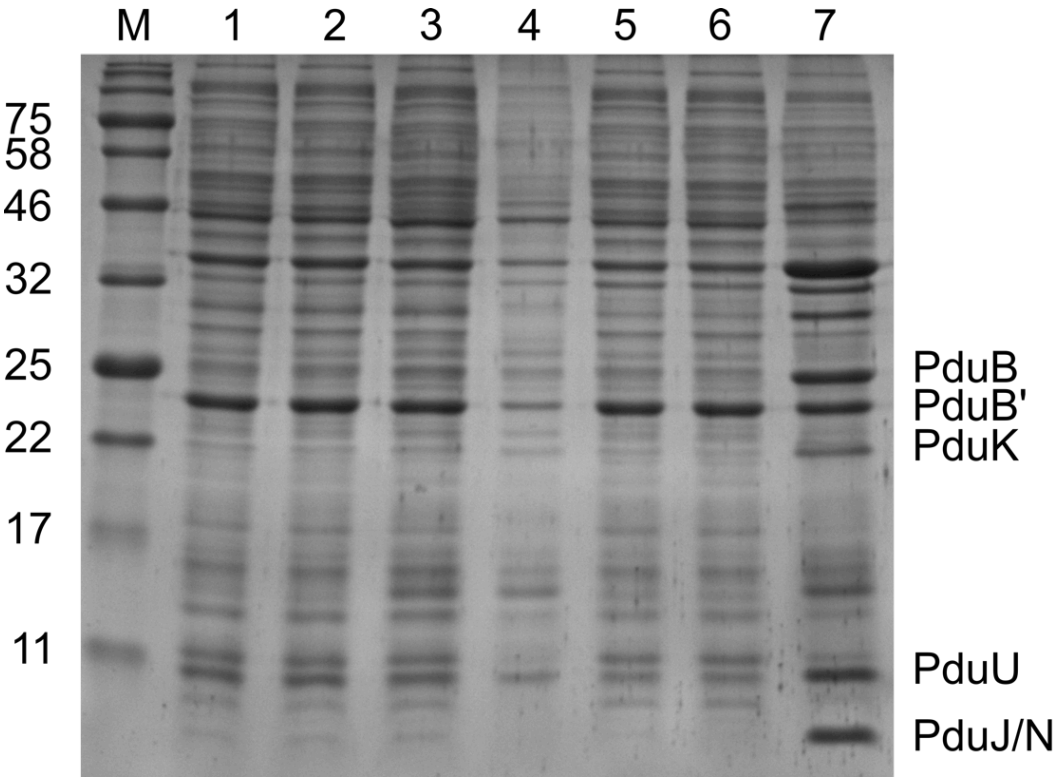
Supplementary Figure 4 SDS-PAGE analysis of CC-Di-A labeled BMC purification in comparison to a molecular weight marker (M). Lane 1 – lysate, lane 2 – supernatant after centrifugation at 11,300 x g, lane 3 – pellet after centrifugation (resuspended in 20 mM Tris – HCl pH 8, 20 mM NaCl), lane 4 – pellet after centrifugation at 11,000 x g (resuspended in 20 mM Tris – HCl pH 8, 20 mM NaCl), lane 5 – supernatant after centrifugation, lane 6 - supernatant after centrifugation in 80 mM NaCl, lane 7 – pellet after centrifugation in 80 mM NaCl, resuspended in 20 mM Tris-HCl pH 8. 2 µl of sample was loaded in lanes 1 and 2, lanes 3 to 7 were loaded with 5 µl of sample.



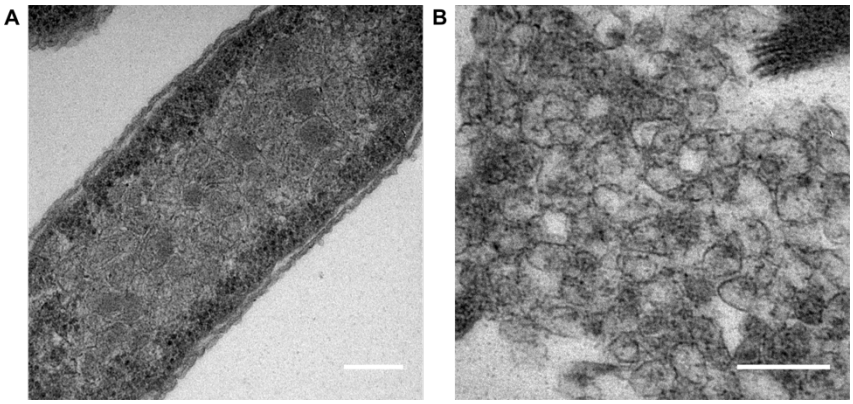
Supplementary Figure 5 SDS-PAGE analysis of C labeled BMC purification in comparison to a molecular weight marker (M). Lane 1 – lysate, lane 2 – supernatant after centrifugation at 11,300 x g, lane 3 – pellet after centrifugation (resuspended in 20 mM Tris – HCl pH 8, 20 mM NaCl), lane 4 – pellet after centrifugation at 11,000 x g (resuspended in 20 mM Tris – HCl pH 8, 20 mM NaCl), lane 5 – supernatant after centrifugation, lane 6 - supernatant after centrifugation in 80 mM NaCl, lane 7 – pellet after centrifugation in 80 mM NaCl, resuspended in 20 mM Tris-HCl pH 8. 2 µl of sample was loaded in lanes 1 and 2, lanes 3 to 7 were loaded with 5 µl of sample.



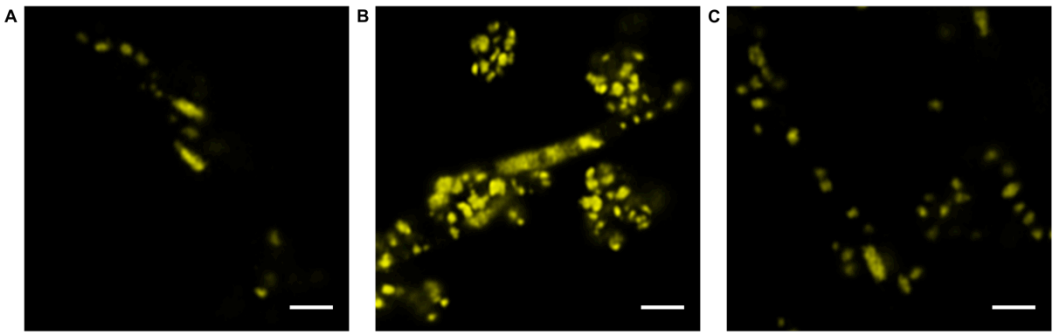
Supplementary Figure 6 SDS-PAGE analysis of CC-Di-A-Citrine labeled BMC purification in comparison to a molecular weight marker (M). Lane 1 – lysate, lane 2 – supernatant after centrifugation at 11,300 x g, lane 3 – pellet after centrifugation (resuspended in 20 mM Tris – HCl pH 8, 20 mM NaCl), lane 4 – pellet after centrifugation at 11,000 x g (resuspended in 20 mM Tris – HCl pH 8, 20 mM NaCl), lane 5 – supernatant after centrifugation, lane 6 - supernatant after centrifugation in 80 mM NaCl, lane 7 – pellet after centrifugation in 80 mM NaCl, resuspended in 20 mM Tris-HCl pH 8. 2 µl of sample was loaded in lanes 1 and 2, lanes 3 to 7 were loaded with 5 µl of sample.



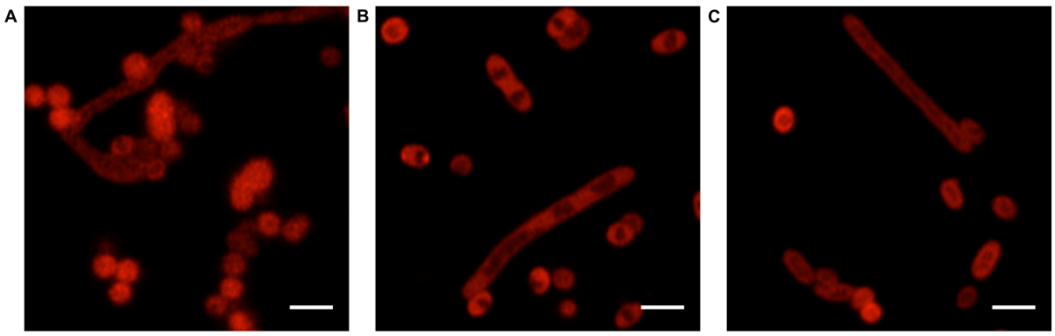
Supplementary Figure 7 SDS-PAGE analysis of C-Citrine labeled BMC purification in comparison to a molecular weight marker (M). Lane 1 – lysate, lane 2 – supernatant after centrifugation at 11,300 x g, lane 3 – pellet after centrifugation (resuspended in 20 mM Tris – HCl pH 8, 20 mM NaCl), lane 4 – pellet after centrifugation at 11,000 x g (resuspended in 20 mM Tris – HCl pH 8, 20 mM NaCl), lane 5 – supernatant after centrifugation, lane 6 - supernatant after centrifugation in 80 mM NaCl, lane 7 – pellet after centrifugation in 80 mM NaCl, resuspended in 20 mM Tris-HCl pH 8. 2 µl of sample was loaded in lanes 1 and 2, lanes 3 to 7 were loaded with 5 µl of sample.



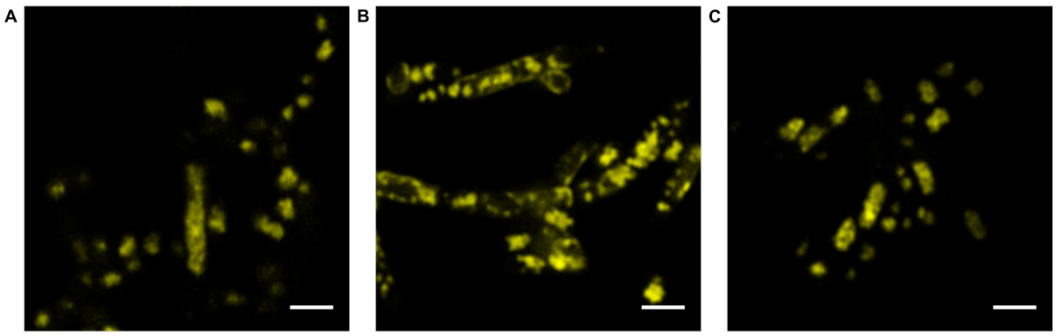
Supplementary Figure 8 TEM micrographs of in-vivo (**A**) and purified (**B**) empty Pdu BMCs. Scale bar shows 0.2 μm.



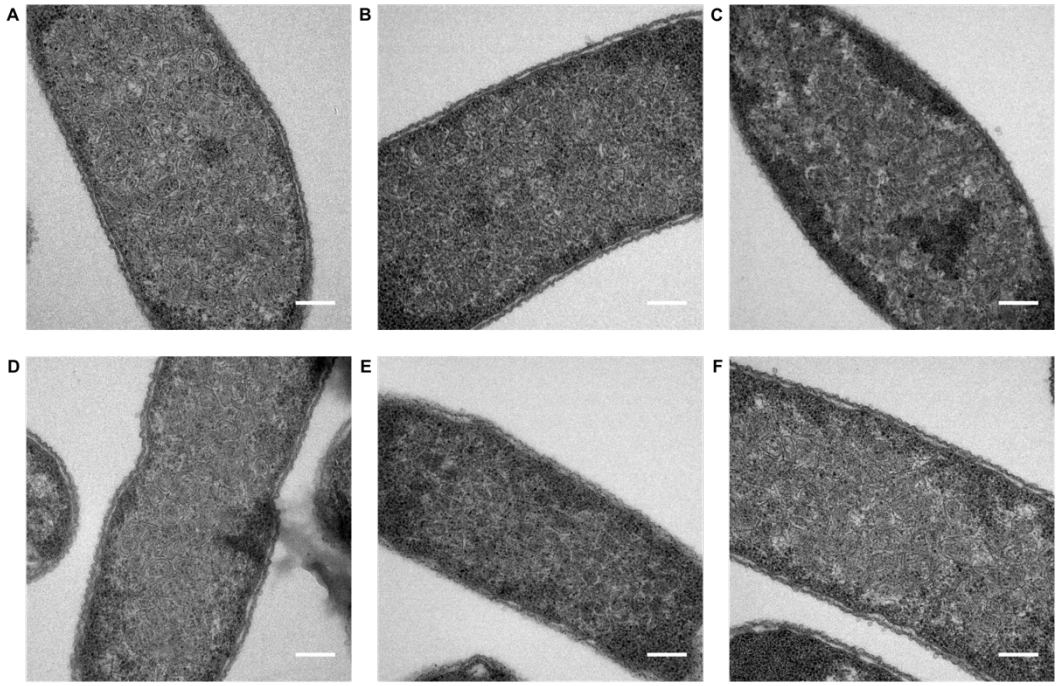
Supplementary Figure 9 Confocal analysis of *E. coli* BL21 * (DE3) cells expressing (**A**) CC-Di-A-Citrine-PduA-U (**B**) CC-Di-B-Citrine-PduA-U (**C**) C-Citrine-PduA-U. Scale bars show 5 μm.



Supplementary Figure 10 Confocal analysis of *E. coli* BL21 * (DE3) cells expressing (**A**) CC-Di-A-mCherry (**B**) CC-Di-B-mCherry (**C**) C-mCherry. Scale bars show 5 μm.



Supplementary Figure 11 Confocal analysis of *E. coli* BL21 * (DE3) cells expressing (**A**) CC-Di-A-Citrine-PduA^{P}-U (**B**) CC-Di-B-Citrine-PduA^{P}-U (**C**) C-Citrine-PduA^{P}-U. Scale bars show 5 µm.



Supplementary Figure 12 T TEM analysis of *E. coli* BL21 * (DE3) cells expressing permuted PduA variants: (**A**) CC-Di-A-PduA^{P}-U (**B**) CC-Di-B-PduA^{P}-U (**C**) C-PduA^{P}-U **D**) CC-Di-A-Citrine-PduA^{P}-U (**E**) CC-Di-B-Citrine-PduA^{P}-U (**F**) C-Citrine-PduA^{P}-U. Scale bars show 0.2 µm.

References

References

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

2. Dueber, J. E. *et al*. Synthetic protein scaffolds provide modular control over metabolic flux. *Nat. Biotechnol.* **27**, 753-759 (2009).

3. Conrado, R. J. *et al*. DNA-guided assembly of biosynthetic pathways promotes improved catalytic efficiency. *Nucleic Acids Res.* **40**, 1879-1889 (2012).

4. Lee, M. J., Brown, I. R., Juodeikis, R., Frank, S. & Warren, M. J. Employing bacterial microcompartment technology to engineer a shell-free enzyme-aggregate for enhanced 1,2-propanediol production in *Escherichia coli*. *Metab. Eng.* **36**, 48-56 (2016).

5. Lee, M.J. *et al*. (2018) Engineered synthetic scaffolds for organizing proteins within the bacterial cytoplasm. Nat Chem Biol 14(2): 142-147

6. Lawrence, A. D. *et al*. Solution structure of a bacterial microcompartment targeting peptide and its application in the construction of an ethanol bioreactor. *ACS Synth. Biol.* **3**, 454-465 (2014).

7. Axen, S. D., Erbilgin, O. & Kerfeld, C. A. A taxonomy of bacterial microcompartment loci constructed by a novel scoring method. *PLoS Comput. Biol.* **10**, e1003898 (2014).

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

9. Yeates, T. O., Kerfeld, C. A., Heinhorst, S., Cannon, G. C. & Shively, J. M. Protein-based organelles in bacteria: carboxysomes and related microcompartments. *Nat. Rev. Microbiol.* **6**, 681-691 (2008).

10. Frank, S., Lawrence, A. D., Prentice, M. B. & Warren, M. J. Bacterial microcompartments moving into a synthetic biological world. *J. Biotechnol.* **163**, 273-279 (2013).

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

12. Jakobson, C. M., Tullman-Ercek, D., Slininger, M. F. & Mangan, N. M. A systems-level model reveals that 1,2-Propanediol utilization microcompartments enhance pathway flux through intermediate sequestration. *PLoS Comput. Biol.* **13**, e1005525 (2017).

13. Sampson, E. M. & Bobik, T. A. 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 (2008).

14. Cheng, S., Sinha, S., Fan, C., Liu, Y. & Bobik, T. A. Genetic analysis of the protein shell of the microcompartments involved in coenzyme B12-dependent 1,2-propanediol degradation by *Salmonella*. *J. Bacteriol.* **193**, 1385-1392 (2011).

15. Cheng, S., Liu, Y., Crowley, C. S., Yeates, T. O. & Bobik, T. A. Bacterial microcompartments: their properties and paradoxes. *Bioessays* **30**, 1084-1095 (2008).

16. Shively, J. M., Ball, F., Brown, D. H. & Saunders, R. E. Functional organelles in prokaryotes: polyhedral inclusions (carboxysomes) of *Thiobacillus neapolitanus*. *Science* **182**, 584-586 (1973).

17. Tcherkez, G. G., Farquhar, G. D. & Andrews, T. J. Despite slow catalysis and confused substrate specificity, all ribulose bisphosphate carboxylases may be nearly perfectly optimized. *Proc. Natl. Acad. Sci. U. S. A.* **103**, 7246-7251 (2006).

18. Bobik, T. A., Havemann, G. D., Busch, R. J., Williams, D. S. & Aldrich, H. C. 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. *J. Bacteriol.* **181**, 5967-5975 (1999).

19. Bobik, T. A., Xu, Y., Jeter, R. M., Otto, K. E. & Roth, J. R. Propanediol utilization genes (*pdu*) of *Salmonella typhimurium*: three genes for the propanediol dehydratase. *J. Bacteriol.* **179**, 6633-6639 (1997).

20. Havemann, G. D., Sampson, E. M. & Bobik, T. A. PduA is a shell protein of polyhedral organelles involved in coenzyme B(12)-dependent degradation of 1,2-propanediol in *Salmonella enterica serovar typhimurium* LT2. *J. Bacteriol.* **184**, 1253-1261 (2002).

21. Mayer, M. J. *et al*. Effect of bio-engineering on size, shape, composition and rigidity of bacterial microcompartments. *Sci. Rep.* **6**, 36899 (2016).

22. Wheatley, N. M., Gidaniyan, S. D., Liu, Y., Cascio, D. & Yeates, T. O. Bacterial microcompartment shells of diverse functional types possess pentameric vertex proteins. *Protein Sci.* **22**, 660-665 (2013).

23. Havemann, G. D. & Bobik, T. A. Protein content of polyhedral organelles involved in coenzyme B12-dependent degradation of 1,2-propanediol in *Salmonella enterica serovar Typhimurium* LT2. *J. Bacteriol.* **185**, 5086-5095 (2003).

24. Pang, A., Liang, M., Prentice, M. B. & Pickersgill, R. W. Substrate channels revealed in the trimeric *Lactobacillus reuteri* bacterial microcompartment shell protein PduB. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 1642-1652 (2012).

25. Pang, A., Frank, S., Brown, I., Warren, M. J. & Pickersgill, R. W. Structural insights into higher order assembly and function of the bacterial microcompartment protein PduA. *J. Biol. Chem.* **289**, 22377-22384 (2014).

26. Sutter, M., Greber, B., Aussignargues, C. & Kerfeld, C. A. Assembly principles and structure of a 6.5-MDa bacterial microcompartment shell. *Science* **356**, 1293-1297 (2017).

27. Parsons, J. B. *et al*. Biochemical and structural insights into bacterial organelle form and biogenesis. *J. Biol. Chem.* **283**, 14366-14375 (2008).

28. Parsons, J. B. *et al*. Synthesis of empty bacterial microcompartments, directed organelle protein incorporation, and evidence of filament-associated organelle movement. *Mol. Cell* **38**, 305-315 (2010).

29. Fan, C. & Bobik, T. A. The N-terminal region of the medium subunit (PduD) packages adenosylcobalamin-dependent diol dehydratase (PduCDE) into the Pdu microcompartment. *J. Bacteriol.* **193**, 5623-5628 (2011).

30. Fan, C. *et al*. Short N-terminal sequences package proteins into bacterial microcompartments. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 7509-7514 (2010).

31. Fan, C., Cheng, S., Sinha, S. & Bobik, T. A. Interactions between the termini of lumen enzymes and shell proteins mediate enzyme encapsulation into bacterial microcompartments. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 14995-15000 (2012).

32. Jakobson, C. M., Kim, E. Y., Slininger, M. F., Chien, A. & Tullman-Ercek, D. Localization of proteins to the 1,2-propanediol utilization microcompartment by non-native signal sequences is mediated by a common hydrophobic motif. *J. Biol. Chem.* **290**, 24519-24533 (2015).

33. Jakobson, C. M., Slininger Lee, M. F. & Tullman-Ercek, D. *De novo* design of signal sequences to localize cargo to the 1,2-propanediol utilization microcompartment. *Protein Sci.* **26**, 1086-1092 (2017).

34. Huber, I. *et al*. Construction of Recombinant Pdu Metabolosome Shells for Small Molecule Production in *Corynebacterium glutamicum*. *ACS Synth. Biol.* (2017).

35. Fletcher, J. M. *et al*. Self-assembling cages from coiled-coil peptide modules. *Science* **340**, 595-599 (2013).

36. Thomas, F., Boyle, A. L., Burton, A. J. & Woolfson, D. N. A set of *de novo* designed parallel heterodimeric coiled coils with quantified dissociation constants in the micromolar to sub-nanomolar regime. *J. Am. Chem. Soc.* **135**, 5161-5166 (2013).

37. Crowley, C. S., Sawaya, M. R., Bobik, T. A. & Yeates, T. O. Structure of the PduU shell protein from the Pdu microcompartment of *Salmonella*. *Structure* **16**, 1324-1332 (2008).

38. Cunningham, B. A., Hemperly, J. J., Hopp, T. P. & Edelman, G. M. Favin versus concanavalin A: Circularly permuted amino acid sequences. *Proc. Natl. Acad. Sci. U. S. A.* **76**, 3218-3222 (1979).

39. Kerfeld, C. A. *et al*. Protein structures forming the shell of primitive bacterial organelles. *Science* **309**, 936-938 (2005).

40. Crowley, C. S. *et al*. Structural insight into the mechanisms of transport across the *Salmonella enterica* Pdu microcompartment shell. *J. Biol. Chem.* **285**, 37838-37846 (2010).

41. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* **10**, 845-858 (2015).