The complexity of prokaryotic organelle construction; examples from *Rhodospirillum rubrum* and *Clostridium autoethanogenum*

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

2019

Maria Stanley

Declaration

Name: Maria Stanley

Degree: PhD–Biochemistry

Title: The complexity of prokaryotic organelle construction; examples from *Rhodospirillum rubrum* and *Clostridium autoethanogenum*

No part of this thesis has been submitted in support of an application for any degree or other qualification of the University of Kent, or any other University or Institution of learning.

Abstract

Bacterial microcompartments (BMCs) are proteinaceous organelles that encapsulate specific metabolic pathways to enhance key catalytic processes and protect the cell from toxic intermediates. Carboxysomes and the 1,2-propanediol utilising and ethanolamine utilising metabolosomes have been studied extensively. However, bioinformatic analyses have revealed a host of BMCs that are yet to be experimentally investigated. This study examined the production by *E. coli* of previously uncharacterised shell proteins from two organisms: *Rhodospirillum rubrum* and *Clostridium autoethanogenum*.

The flexible nature of microcompartment formation was revealed by the different structures formed by the *R. rubrum* shell proteins both *in vivo* in *E. coli* and following their isolation. *In vivo*, shell proteins formed swirled sheets that were unable to form compartments. However, upon isolation the components of the protein structures were able to re-assemble to form apparently closed "empty" compartments. This highlighted the complex nature of the protein interactions involved in microcompartment assembly and the effect of a changing environment upon these interactions.

The production of *C. autoethanogenum* shell proteins in *E. coli* revealed a previously unobserved interaction of protein sheets with ribosomes within the cytoplasm. This phenotype was shown to involve a C-terminally extended hexameric shell protein, Caethg_3286. The function of C-terminally extended shell proteins, which are encoded in many BMC operons, is unknown at present although the structure of one, EutK, was resolved and shows a high degree of similarity to nucleic acid binding domains. The crystal structure of the BMC domain of Caethg_3286 revealed that the C-terminus is on the concave surface of the hexamer allowing the C-terminal extension to extend into the cytoplasm and interact with ribosomes. The ribosomal interaction revealed in this study and the potential nucleic acid binding capacity of another C-terminal extension may indicate a role of the C-terminal extension in the regulation of BMC formation at the level of translation.

Figures

Tables

Abbreviations

1,2-PD	1,2-propanediol
2D	Two dimensional
3-PGA	3-phosphglycerate
Å	Angstrom
ADP	Adenosine diphosphate
AE	Activating enzyme
AFM	Atomic force microscopy
APS	Ammonium persulphate
ATP	Adenosine triphosphate
BCA	Bicinchoninic acid
BCIP	5-bromo-4-chloro-3-indolyl phosphate
BLAST	Basic local alignment search tool
BMC	Bacterial microcompartment
BMC-H	Hexameric shell protein
BMC-H_{ex}	Hexameric shell protein with a C-terminal extension
BMC-H_{FeS}	Hexameric shell protein with an iron-sulphur binding centre
BMC-P	Pentameric shell protein
BMC-T	Trimeric shell protein
BMC-T_{DB}	Trimeric shell protein that forms a double layer
BMC-T_{FeS}	Trimeric shell protein with iron-sulphur binding centre
BMC-T_p	Circularly permuted trimeric shell protein
CA	Carbonic anhydrase
CoA	Coenzyme A
Da	Dalton
ddH_2O	Double distilled H_2O
dH_2O	Distilled H_2O
DHAP	Dihydroxyacetone phosphate
DNA	Deoxyribonucleic acid
dNTP	Deoxyribonucleotide triphosphate
DTT	Dithiothreitol
EDTA	Ethylene diamine tetra-acetic acid
EHEC	Enterohaemorrhagic *E. coli*
EP	Encapsulation peptide
Eut	Ethanolamine utilisation
Fe-S	Iron- sulphur
FPLC	Fast protein liquid chromatography
GRE	Glycyl radical enzyme
GRM	Glycyl radical microcompartment
IAA	Iodoacetamide
IEF	Isoelectric focusing
IMAC	Immobilised metal affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
Kb	Kilobase
LB	Luria-Bertani
LVR	Low viscosity resin
MALDI-TOF	Matrix assisted laser desorption/ionisation – time of flight
NAD^+	Nicotinamide adenine dinucleotide (oxidised form)
NADH	Nicotinamide adenine dinucleotide (reduced form)
NBT	Nitro blue tetrazolium
nm	Nanometres
OD_{600}	Optical density at a wavelength of 600 nm

PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDB	Protein data bank
Pdu	1,2-propanediol utilising
PEG	Polyethylene glycol
PHA	Poly beta-hydroxyalkanoates
PISA	Proteins, interfaces, structures and assemblies
Rmsd	Root mean square deviation
RuBisCO	Ribulose bisphosphate carboxylase/oxygenase
RuBP	Ribulose-1,5-bisphosphate
SAM	S-adenosylmethionine
SDS-PAGE	Sodium dodecyl sulphate – polyacrylamide gel electrophoresis
SOC	Super optimal broth and catabolite repression
SSLDs	Small subunit-like domains
TAE	Tris acetate EDTA
TCA	Trichloroacetic acid
TEM	Transmission electron microscopy
TEMED	Tetramethylenediamine
Tris	2-amino-2-hydroxymethyl-1,3-propanediol
tRNA	Transfer ribonucleic acid
UV	Ultra violet

Acknowledgements

I would like to thank the BBSRC for providing the funding to allow me to undertake this PhD as part of a LOLA grant entitled ‘Development of supramolecular assemblies for enhancing cellular productivity and the synthesis of fine chemicals and biotherapeutics’. I would also like to thank the University of Kent for the remission of fees which made undertaking this PhD financially viable.

There are many individuals who I need to thank for their help and support throughout my time here at the University of Kent.

I would first like to thank Professor Martin Warren for giving me the opportunity to complete this PhD. For having the faith in my ability to achieve this goal, perhaps a bit more than I did at times.

A very heartfelt thank you goes to Dr. David Palmer, for taking me under his wing at the very beginning and helping me with patience and kindness at all times. I will be forever grateful. Special thanks also to Dr. Evelyne Deery who provided me with, not only so much help and advice about cloning, but with warmth, laughter and friendship - I would have been completely lost without you. Thanks also goes to Dr. Matt Lee and Dr. Mingzhi Liang for support, help and advice throughout my time here.

I would like to say an enormous thank you to Ian Brown for his infinite patience and support when teaching me the fine art of thin-sectioning. I loved spending time in his lab - great company and great music at all times (except for those few occasions when I was forced to listen to test cricket!).

Thanks also to Prof. Dave Brown for his help and support with analysis of crystallography data, to Judith Mantell at the university of Bristol for her electron tomography expertise and all the LoLa team for their help and feedback.

A big thank you to all of my family. They have supported me throughout the last few years, helping me to de-stress and focus on things other than work.

Lastly, a huge thanks goes to everyone who has been in the Warren Lab over the course of my research. You have all made it all such an enjoyable experience; I have made many friends and had many laughs along the way - and the cakes have been phenomenal!

Chapter 1

Introduction

1.1 Background and history of bacterial microcompartments

It was believed for many years that subcellular organisation was found only within eukaryotic cells. However, more recently it has been recognised that some bacteria also house organelles within the cytoplasm to segregate specific functions from the rest of the cell. These include magnetosomes in magnetotactic bacteria (Balkwill *et al.*, 1980), membrane bound anammoxosomes in selected planctomycetes (van Niftrik and Jetten, 2012) and bacterial microcompartments in a diverse array of bacteria (Axen *et al.*, 2014).

Bacterial microcompartments (BMCs) are proteinaceous structures found within the cytoplasm of many bacteria. Bioinformatic studies have shown that about 20% of all fully sequenced bacterial genomes contain genes encoding BMC domain proteins (AbdulRahman, 2013; Jorda *et al.*, 2013; Axen *et al.,* 2014). Interestingly in the human gut microbiome the number of organisms encoding BMCs is even higher at 23% (Ravcheev *et al.*, 2019). Enzyme pathways are sequestered to the interior of the microcompartment to: (i) enhance enzyme efficiency by increasing local concentrations of enzymes and substrates in close proximity to each other (Jakobson *et al.*, 2017), and (ii) prevent toxic or volatile intermediates of the enzyme pathway entering the cytoplasm (Penrod and Roth, 2006; Sampson and Bobik, 2008; Chowdhury *et al.*, 2015).

First observations of BMCs in the 1950s were in electron microscopy images which revealed polyhedral inclusion bodies in the cyanobacterium *Phormidium uncinartum* (Drews and Niklowitz, 1956). There were further observations in *Synechococcus elongatus* (formerly *Anacystis nidulans*) (Gantt and Conti, 1969) and *Thiobacillis neaploitanus* (Shively *et al.*,1970). It was initially speculated that these polyhedral bodies were phages but in 1973 studies with *Thiobacillus neapolitanus* identified them as proteinaceous organelles containing an enzyme involved in carbon fixation, ribulose bisphosphate carboxylase/oxygenase (RuBisCO), and were subsequently named ‘carboxysomes’ (Shively *et al.*, 1973). BMCs that encapsulate catabolic enzymes have since been identified and termed ‘metabolosomes’ (Brinsmade *et al.*, 2005), for example propanediol utilising (Pdu) and ethanolamine utilising (Eut) microcompartments.

The ever expanding number of fully sequenced bacterial genomes has allowed more recent bioinformatic analyses to identify numerous types of BMC that are only just beginning to be experimentally investigated, including glycyl radical enzyme microcompartments (GRMs) which form the largest group of BMC loci identified (Axen *et al.*, 2014; Zarzycki *et al.*, 2015).

1.2 Carboxysomes

1.2.1 Characteristics of carboxysomes

Carboxysomes were the first type of BMC to be observed and subsequently characterised. Carboxysomes are found in all cyanobacteria and many chemoautotrophs (Rae *et al.*, 2013). There are two types of carboxysome, α and β, which are each associated with distinct forms of RuBisCO – type IA and IB (Badger, Hanson and Price, 2002). α-carboxysomes are encoded by *cso* genes, clustered in operons, and are found in oceanic cyanobacteria and some chemoautotrophs. β-carboxysomes, however, are encoded by *ccm* genes spread throughout the genome and are found mainly in freshwater cyanobacteria (Badger *et al.*, 2002). Whilst both α- and β-carboxysomes are polyhedral they differ in size with α-carboxysomes being 80 – 130 nm in diameter and β-carboxysomes 200 – 300 nm (Rae *et al.*, 2013) (Fig.1.1).

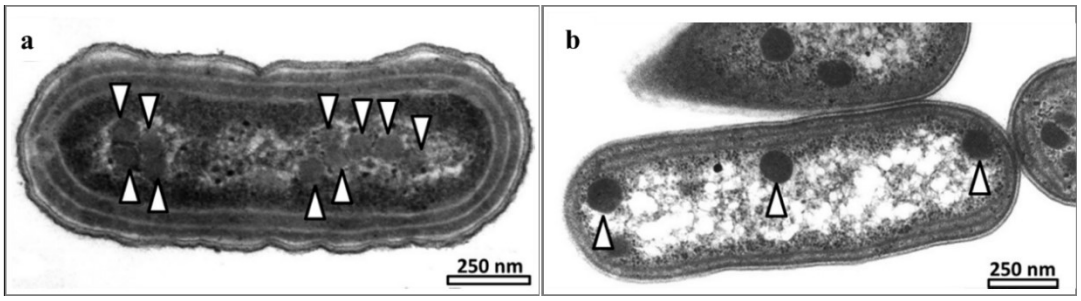


Figure 1.1. TEM images of α- and β-carboxysomes in cyanobacteria. a) α-carboxysomes in *Cyanobium* PCC7001; b) β-carboxysomes in *Synechococcus elongatus* PCC7942. Carboxysomes are indicated (white arrows). Images reproduced from Rae et al., 2013.

1.2.2 The functional role of carboxysomes

The protein shell of the carboxysome harbours both RuBisCO and carbonic anhydrase (Shively *et al.*, 1973; Price and Badger, 1989a; Price *et al.*, 1992), two key enzymes involved in CO_2 fixation. The principal function of the carboxysome is the enhancement of CO_2 fixation by the Calvin-Benson-Bassham cycle as part of the carbon concentrating mechanism (Price and Badger, 1991). The evolution of the carboxysome is proposed to be in response to an increase in atmospheric oxygen millions of years ago as a means to increase the carboxylation reaction over the oxygenation reaction that RuBisCO can both catalyse (Badger and Price, 2003). Inorganic carbon, in the form of HCO_3^-, is actively transported into the cytoplasm and is able to diffuse across the outer shell into the lumen of the carboxysome. Here, carboxysomal carbonic anhydrase converts it to CO_2. CO_2 is unable to diffuse out of the carboxysome and so is concentrated in the region of the RuBisCO resulting in increased activity of the enzyme. The 3-phosphoglycerate produced is able to diffuse out of the carboxysome into the cytoplasm

where it can then enter central metabolism. Disruption of the carboxysome shell results in cells that are unable to grow at normal atmospheric CO_2 levels (Price and Badger, 1989b).

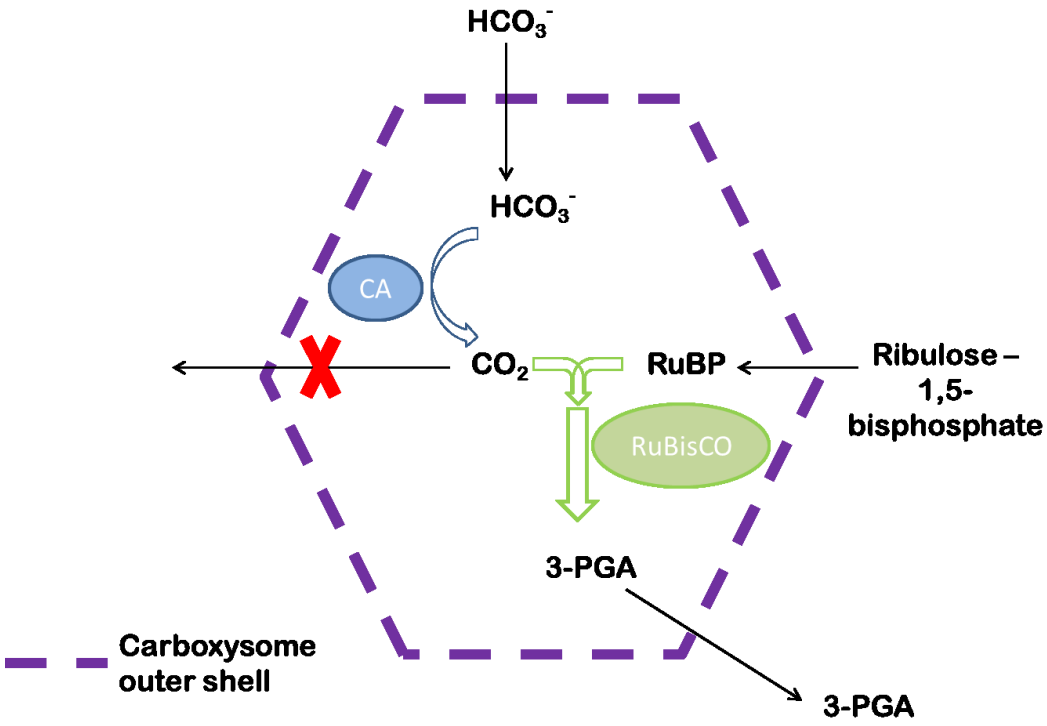


Figure 1.2. Schematic of the model of carboxysome function. Bicarbonate (HCO_3^-) diffuses into the carboxysome where it is converted to CO_2 by carboxysomal carbonic anhydrase (CA). CO_2 is unable to diffuse through the carboxysome shell and is concentrated in the region of RuBisCO which catalyses the carboxylation of ribulose-1,5-bisphosphate (RuBP) to form 3-phosphoglycerate (3-PGA). 3-PGA is able to diffuse out of the carboxysome into the cytoplasm where it enters central metabolism.

1.3 1,2-Propanediol utilisation microcompartments

1.3.1 Characteristics of Pdu BMCs

1,2-Propanediol utilising (Pdu) microcompartments were first observed in *Salmonella enterica* and *Klebsiella oxytoca* when grown anaerobically on 1,2-propanediol (1,2-PD) as a sole carbon source (Shively *et al.*, 1998). This followed genetic analysis revealing the *pdu* operon of *S. enterica* contains genes homologous to carboxysome shell protein genes (Chen *et al.*, 1994). Subsequent studies using immunogold labelling showed enzymes involved in 1,2-PD degradation were localised to these polyhedral bodies formed in *Salmonella* grown on 1,2-PD (Bobik *et al.*, 1999; Leal *et al.*, 2003). The Pdu BMCs have a more irregular shape than carboxysomes and are approximately 100–150 nm in diameter (Fig. 1.3). Proteomic analysis of isolated compartments revealed that, as well as the enzymes for 1,2-PD degradation,

enzymes for the recycling of coenzyme-B_{12} were also associated with the microcompartments (Havemann and Bobik, 2003).

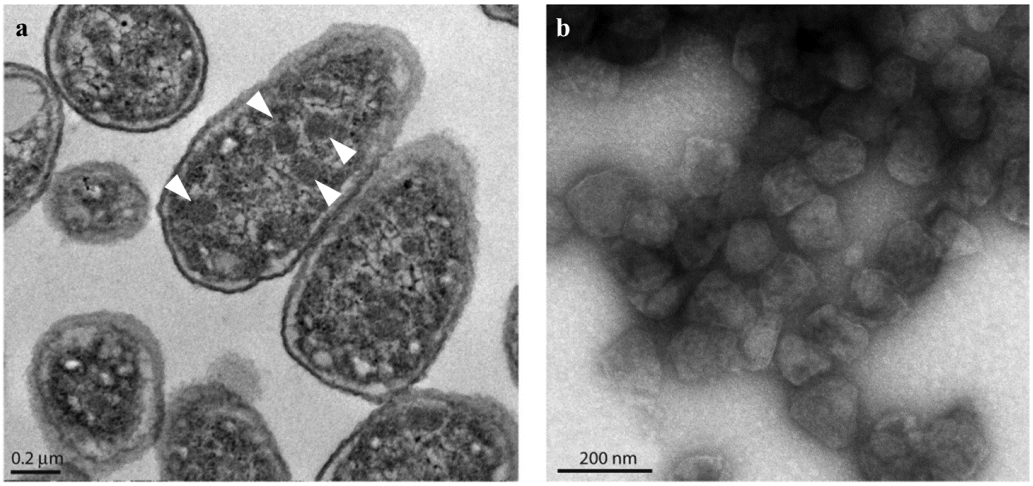


Figure 1.3. TEM images of microcompartments in *S. enterica* grown on 1,2-PD. a) Thin section of *S. enterica* with microcompartments indicated (white arrows). b) Negatively stained isolated *S. enterica* Pdu BMCs. Images reproduced from Chowdhury et al., 2014.

1.3.2 The 1,2-PD degradation pathway

The metabolic pathway for 1,2-PD degradation is summarised in Figure 1.4. The first step in the catabolism of 1,2-PD is dehydration by a B_{12}-dependent diol dehydratase to form propionaldehyde. The proteins necessary for the recycling of co-enzyme B_{12}, cobalamin reductase (PduS) (Cheng and Bobik, 2010; Parsons *et al.*, 2010) and ATP:cob(I)alamin adenosyltransferase (PduO) (Johnson *et al.*, 2001), are also present within the lumen of the BMC. PduGH is a diol dehydratase reactivase enzyme (Bobik *et al.*, 1999) and again is located inside the BMC providing the means to recycle the enzyme and co-factor required for this first step.

Propionaldehyde, produced in step 1, is then converted to propionyl-CoA or 1-propanol by propionaldehyde dehydrogenase or 1-propanol dehydrogenase respectively. The aldehyde dehydrogenase uses NAD^+ and coenzyme A as a co-factors producing propionyl-CoA and NADH. The alcohol dehydrogenase uses NADH as a cofactor and produces 1-propanol and NAD^+. This dual pathway allows the recycling of NAD^+/NADH within the BMC therefore negating the need for the transport of either cofactor across the outer shell if the cofactors are sequestered during microcompartment assembly (Cheng *et al.*, 2012).

1-Propanol is able to diffuse out of the BMC into the cytoplasm of the cell where it can be used as a carbon source for growth. Propionyl-CoA remains within the BMC lumen and is

metabolised to propionyl-PO_4^{2-} by phosphotransacylase, PduL (Liu *et al.*, 2007), recycling coenzyme A for use by propionaldehyde in the process (Liu *et al.*, 2015).

Propionyl-PO_4^{2-} is able to diffuse out of the BMC into the cytoplasm of the cell where it is converted to propionate and ATP by propionate kinase and can then enter the methylcitrate pathway of central metabolism (Horswill and Escalante-Semerena, 1999).

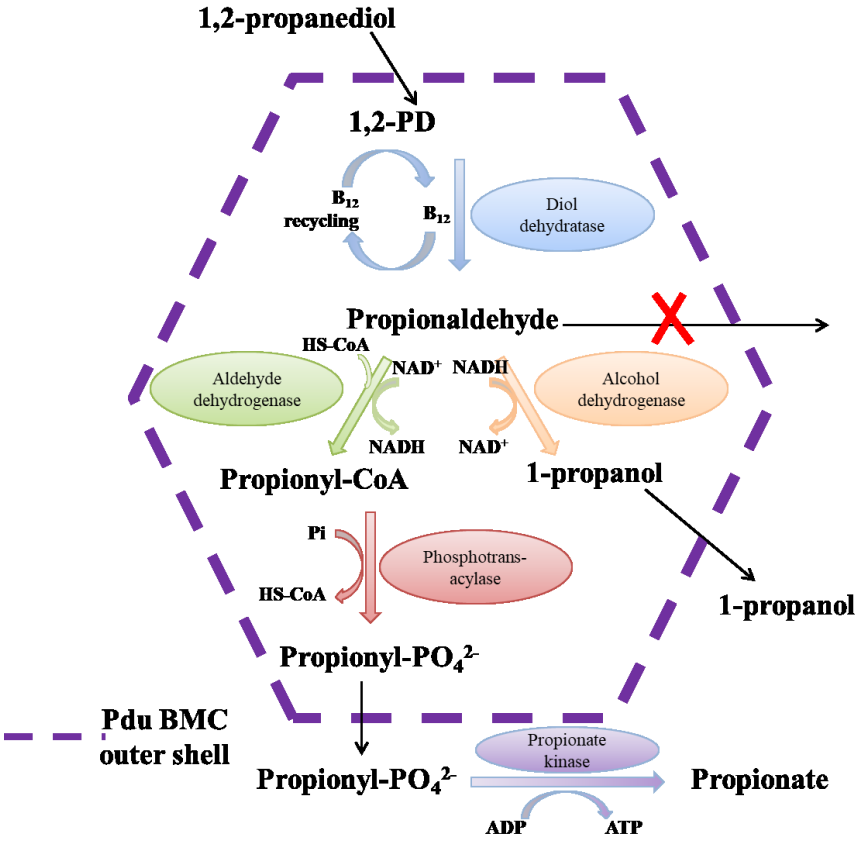


Figure 1.4. Schematic of 1,2-PD metabolism in the Pdu BMC. 1,2-PD diffuses across the outer shell of the Pdu BMC from the cytoplasm where it is dehydrated by a B_{12}-dependent diol dehydratase to form propionaldehyde. B_{12} is recycled by enzymes within the BMC. Propionaldehyde is unable to diffuse out through the shell of the BMC but instead is converted to either propionyl-CoA by propionaldehyde dehydrogenase or to 1-propanol by propanol dehydrogenase. 1-Propanol is able to diffuse out of the BMC into the cytoplasm of the cell. Propionyl-CoA is further converted to propionyl-PO_4^{2-}, which is able to diffuse through the shell of the BMC into the cytoplasm where it is de-phosphorylated by propionate kinase to form propionate whilst also generating ATP.

1.3.3 The functional role of the Pdu BMC

The aldehyde intermediate, formed by the action of propanediol dehydratase on 1,2-PD, is a volatile compound susceptible to loss through the cell membrane. Aldehydes are also highly cytotoxic and can cause both protein and DNA damage. It was therefore proposed that the Pdu

BMC provides a diffusion barrier to the propionaldehyde that is sequestered within it to (i) prevent carbon loss from the diffusion of aldehyde out of the cell, and (ii) protect the rest of the cell from toxic, DNA damaging concentrations of propionaldehyde (Stojiljkovic *et al.*, 1995a).

The Pdu BMCs role in protection from aldehyde toxicity is supported by evidence that *S. enterica* Δ*pduA* mutants, unable to form microcompartments, underwent a period of growth arrest attributed to propionaldehyde toxicity (Havemann *et al.*, 2002). Later evidence showed *S. enterica* mutants unable to form BMCs accumulated concentrations of propionaldehyde that were toxic to wild-type *S. enterica* and had an increased occurrence of DNA mutation (Sampson and Bobik, 2008).

The role of the Pdu BMC in preventing carbon loss from the cell is supported by evidence that increased concentrations of propionaldehyde were lost from *S. enterica* mutants that were unable to form microcompartments when grown on 1,2-PD compared to wild-type cells (Penrod and Roth, 2006). More recently, a role for the Pdu BMC in substrate channelling has been proposed. Modelling has demonstrated that the localisation of sequential pathway enzymes within the confines of the BMC lumen would enhance flux through the pathway so improving efficiency (Jakobson *et al.*, 2017).

1,2-PD is a product of the fermentation of L-rhamnose and L-fucose, sugars found in plant cell walls, bacterial capsules and the glycoconjugates of eukaryotic cells (Chowdhury *et al.*, 2014). It is therefore thought to be an important carbon source in anaerobic environments such as the mammalian large intestine, sediments and soils where 1,2-PD is readily available (Obradors *et al.*, 1988). The genes encoding the shell proteins of the Pdu BMC and its associated pathway enzymes have been identified in the genomes of widespread bacterial phyla that inhabit enteric and sedimentous environments (AbdulRahman, 2013; Jorda *et al.*, 2013; Axen *et al.*, 2014). The ability to form BMCs and metabolise 1,2-PD allows these bacteria to thrive in such niche environments.

1.4 Ethanolamine utilisation microcompartments

1.4.1 Characteristics of Eut microcompartments

Ethanolamine utilising (Eut) microcompartments were first described at the same time as the first report of Pdu BMCs. Polyhedral bodies were observed in TEM images of *E. coli* grown anaerobically on ethanolamine (Shively *et al.*, 1998) and in *S. enterica* grown aerobically on ethanolamine (Brinsmade *et al.*, 2005) (Fig. 1.5). This followed reports of genes homologous to the carboxysome shell protein genes being present in the Eut operon of *S. typhimurium* (Stojiljkovic *et al.*, 1995b). Deletion mutants of *S. enterica* that were missing the Eut shell protein genes were unable to grow on ethanolamine and polyhedral bodies were not formed

(Brinsmade *et al.*, 2005). Eut microcompartments from *S. enterica* have been successfully isolated (Choudhary *et al.*, 2012; Held *et al.*, 2016) but the complete proteomic analysis of associated pathway enzymes has yet to be reported.

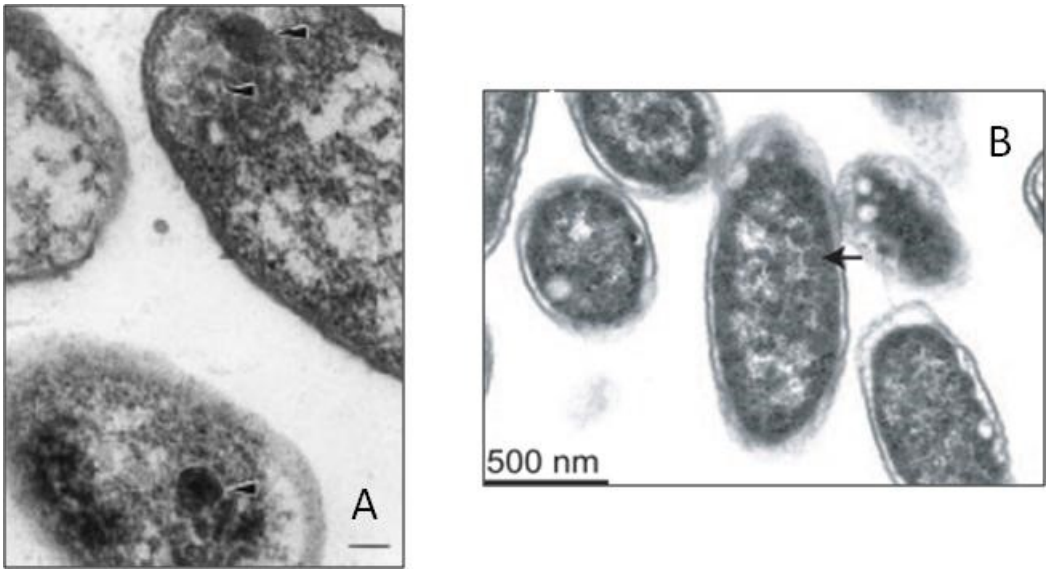


Figure 1.5. TEM image of Eut BMC in: A) *E. coli* K12 cells grown anaerobically on ethanolamine. Image reproduced from Shively et al., 1998; B) *S. enterica* cells grown aerobically on ethanolamine. Image reproduced from Brinsmade *et al.*, 2005. BMCs are indicated (black arrows).

1.4.2 The ethanolamine degradation pathway

Ethanolamine is formed from the breakdown of phosphatidylethanolamine, a major component of bacterial and mammalian cell membranes, by phosphodiesterases (Garsin, 2010). The pathway for ethanolamine metabolism is summarised in Figure 1.6. The pathway is comparable to that for 1,2-PD degradation.

The first step in the pathway is the conversion of ethanolamine to acetaldehyde and ammonia by the B_{12}-dependent ethanolamine ammonia lyase, EutBC (Roof and Roth, 1989). As seen in the Pdu BMC, the Eut BMC operon includes genes encoding an ATP:cob(I)alamin adenosyltransferase for the recycling of coenzyme-B_{12}, EutT (Buan *et al.*, 2004; Sheppard *et al.*, 2004; Buan and Escalante-Semerena, 2006) and an ethanolamine ammonia lyase reactivating factor, EutA (Mori *et al.*, 2004). These activities are necessary as the ethanolamine ammonia lyase is inactivated at each round of catalysis by the cleavage of the C-Co bond in the B_{12} co-factor. This inactive B_{12} must be removed to allow binding of a new molecule of active coenzyme B_{12}, a function that is performed by the reactivase enzyme (Mori *et al.*, 2004).

The degradative pathway proceeds in a similar manner as within the Pdu BMC, with acetaldehyde being converted to either acetyl-CoA or ethanol by acetaldehyde dehydrogenase, EutE, or alcohol dehydrogenase, Eut G, respectively (Kofoid *et al.*, 1999; Brinsmade *et al.,* 2005). NAD^+ and coenzyme A are required as cofactors for the acetaldehyde dehydrogenase and NADH as a cofactor for the alcohol dehydrogenase. The disproportionation of acetaldehyde allows the NAD^+/NADH to be recycled within the BMC therefore an external source is not required assuming some is encapsulated upon compartment assembly.

Ethanol can escape the BMC and enter the cytoplasm of the cell but acetyl-CoA is further metabolised by an acetyl phosphotransferase, EutD, releasing coenzyme A for use by aldehyde dehydrogenase so again promoting the recycling of cofactors within the confines of the BMC (Huseby and Roth, 2013). The final step in the ethanolamine degradation pathway is the conversion of acetyl-PO_4^{2-} to acetate with the concomitant production of ATP by the action of acetate kinases, EutP and EutQ (Moore and Escalante-Semerena, 2016).

1.4.3 **The functional role of the Eut BMC**

The model proposed for the Eut BMC is analogous to that of the Pdu BMC. The degradative enzymes for ethanolamine are contained within a selectively permeable protein shell in order to prevent the aldehyde intermediate from entering the cytoplasm where it may have cytotoxic effects and to prevent it from diffusing out of the cell and being lost to the surrounding environment.

Support for the Eut BMCs role in protecting the cell from aldehyde toxicity is supported by evidence that gsh (glutathione biosynthesis) and polA (DNA repair polymerase) mutants were unable to grow on ethanolamine (Rondon *et al.*, 1995; Rondon *et al.*, 1995). The authors proposed GSH was required to quench reactive aldehyde species generated during ethanolamine and 1,2 –PD catabolism and DNA repair polymerase was required to repair mutations induced by reactive aldehydes.

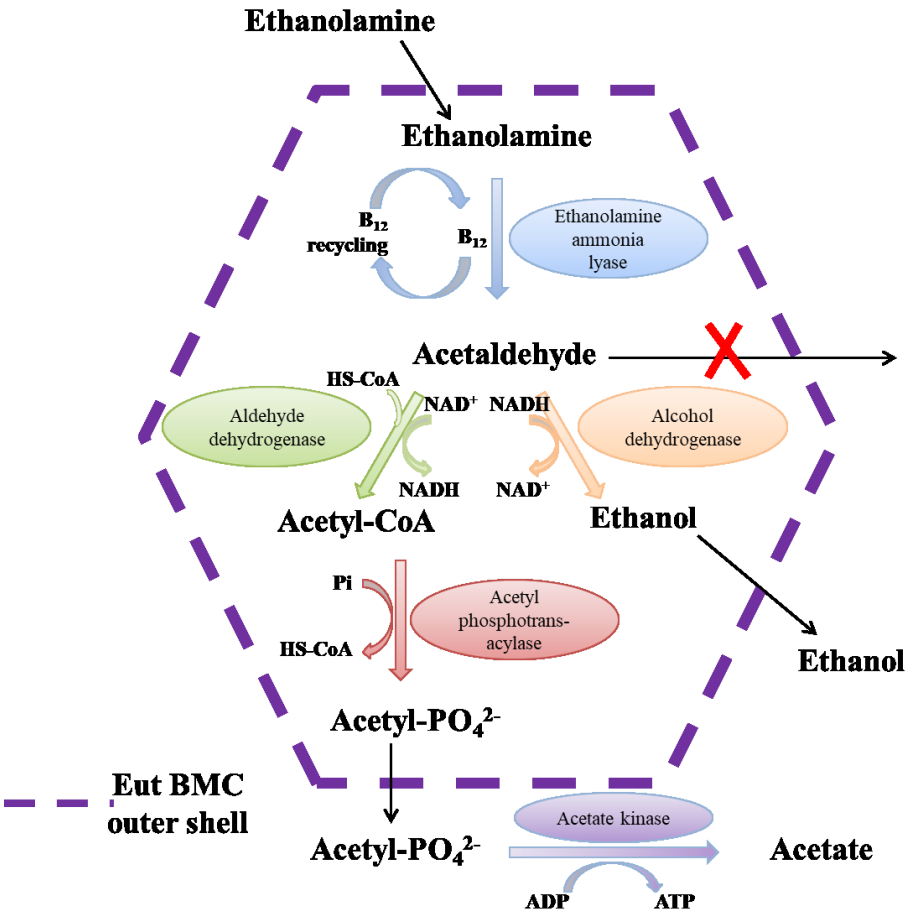


Fig. 1.6. Schematic of the ethanolamine degradation pathway in the Eut BMC. Ethanolamine diffuses across the outer shell of the Pdu BMC from the cytoplasm where B_{12}-dependent ethanolamine ammonia lyase converts it to acetaldehyde. B_{12} is recycled by enzymes within the BMC. Acetaldehyde is unable to diffuse out through the shell of the BMC but instead is converted to either acetyl-CoA by aldehyde dehydrogenase or to ethanol by alcohol dehydrogenase. Ethanol is able to diffuse out of the BMC into the cytoplasm of the cell. Acetyl-CoA is further converted to acetyl-PO_4^{2-}, which is able to diffuse through the shell of the BMC into the cytoplasm where it is de-phosphorylated by acetate kinase to form acetate whilst also generating ATP.

Loss of acetaldehyde to the surrounding medium was shown to be increased in *S. enterica* shell protein mutants grown on ethanolamine (Penrod and Roth, 2006). The authors suggested that the resultant reduction in growth was due to the loss of carbon from the cell and not due to cytotoxicity as the concentration of acetaldehyde in the cytoplasm does not reach cytotoxic levels. A role in the conservation of carbon rather than protection from a toxic aldehyde was proposed. Finally a role for Eut BMCs increasing the local concentrations of enzymes, substrates and co-factors has been proposed (Brinsmade *et al.*, 2005). It was suggested that mutant stains lacking 4 of the BMC shell proteins were unable to grow on ethanolamine because the ethanolamine catabolic enzymes were diluted to the point of inefficiency.

1.5 Glycyl radical enzyme microcompartments (GRMs)

Glycyl radical enzyme microcompartments (GRMs) were identified as the most abundant class of BMCs in a recent bioinformatic study (Axen *et al.*, 2014). GRM loci were identified in diverse bacterial phyla including actinobacteria, firmicutes and α-, γ– and δ-proteobacteria. A recent review identified 536 GRM loci in the UniProt database (Ferlez *et al.*, 2019). Despite their prevalence, there is very little experimental data available.

1.5.1 Glycyl radical enzymes

GRM BMC loci contain genes encoding a glycyl radical enzyme (GRE) together with its associated activating enzyme (AE) as the signature enzyme. Glycyl radical enzymes are homodimeric proteins with a subunit size of 80 – 100 kDa (Selmer, Pierik and Heider, 2005). They are synthesised in an inactive form and require activation by an iron-sulphur cluster containing enzyme, known as its activating enzyme (AE), which belongs to the radical S-adenosylmethionine (SAM) family. SAM binds to an iron atom of the AE active site [4Fe-4S]$^{2+}$ cluster which in turn binds the GRE close to the glycyl loop of its active site which moves into an open conformation. The [4Fe-4S]$^{2+}$ cluster is reduced to [4Fe-4S]$^{1+}$ by an electron donor. The reductive cleavage of SAM to methionine and a transient reactive Ado$^{\cdot}$ radical is achieved through direct electron transfer from the [4Fe-S]$^{1+}$. Ado$^{\cdot}$ abstracts a hydrogen atom from the glycine loop of the GRE forming a glycyl radical (Gly$^{\cdot}$) in the active site. Only one subunit of the GRE dimer is in the active state at any time (Selmer *et al.*, 2005; Backman *et al.*, 2017; Ferlez *et al.*, 2019).

The activated form of the enzyme is very oxygen sensitive. In the presence of oxygen, the GRE is irreversibly inactivated by cleavage of the polypeptide bond at the site of Gly$^{\cdot}$. The encapsulation of a GRE within a BMC could be a mechanism to provide protection of the enzyme from the harmful effect of oxygen.

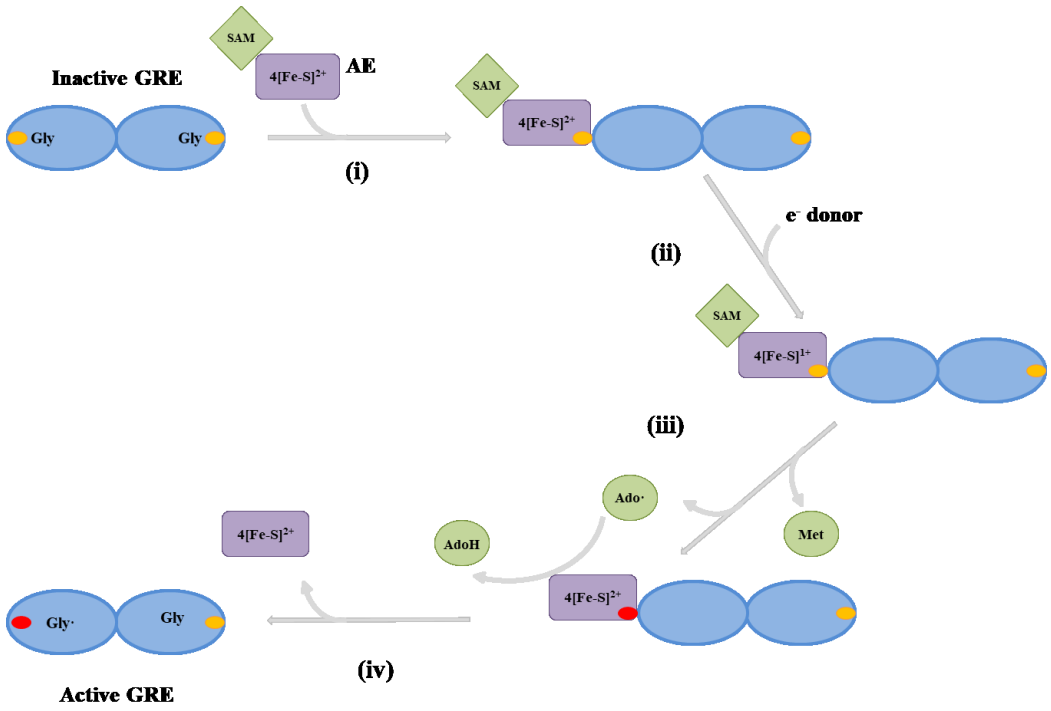


Figure 1.7. Schematic depicting the mechanism of activation of GREs. (i) SAM (green) binds to the active site Fe-S cluster of the AE (purple) and this complex binds close to the glycine loop (yellow) of the GRE (blue) so it is in the open conformation; (ii) The Fe-S of the AE is reduced by an electron donor; (iii) Direct electron transfer from the reduced Fe-S results in the cleavage of SAM to methionine and a reactive Ado˙ radical. Ado˙ abstracts a hydrogen atom from the glycine loop forming AdoH and an active site gly˙ radical (red) in the GRE; (iv) The AE dissociates leaving the GRE in its active state. Figure adapted from Fig. 1 in Ferlez, Sutter and Kerfeld, 2019.

1.5.2 GRM subtypes

Axen (Axen *et al.*, 2014) identified six sub-groups of GRM, GRM1 – 5 and GUF (GRM of unknown function). Another intensive bioinformatic study of the GRMs, examining sequence, structural and phylogenetic data of the signature and other pathway enzymes, confirmed the categorisation of the earlier study and went on to predict the function of these sub-groups (Zarzycki *et al.*, 2015). GRM1 and GRM2 are predicted to metabolise choline as their substrate, GRM3 and GRM4 and GRM5 are predicted to have a 1,2 PD metabolising signature enzyme with GRM5 also containing upstream enzymes for the production of 1,2-PD from plant sugars.

1.5.2.1 Choline metabolising GRMs

GRM1 and GRM2 microcompartments are both predicted to use choline as the substrate for their signature enzyme, a choline trimethylamine lyase (Fig. 1.8). The GRM1 choline trimethylamine lyase has an N-terminal extension of ~60 amino acids distinguishing it from

the GRM2 enzyme which has a ~350 amino acid N terminal extension (Zarzycki *et al.*, 2015). Another differentiation between GRM1 and GRM2 loci is that whilst GRM1 BMC loci also encode Fe-S containing proteins, GRM2 BMC loci do not encode any. The GRM1 locus has been subdivided further into GRM1 and GRM1b based on locus arrangement and gene sequence (Ferlez, Sutter and Kerfeld, 2019).

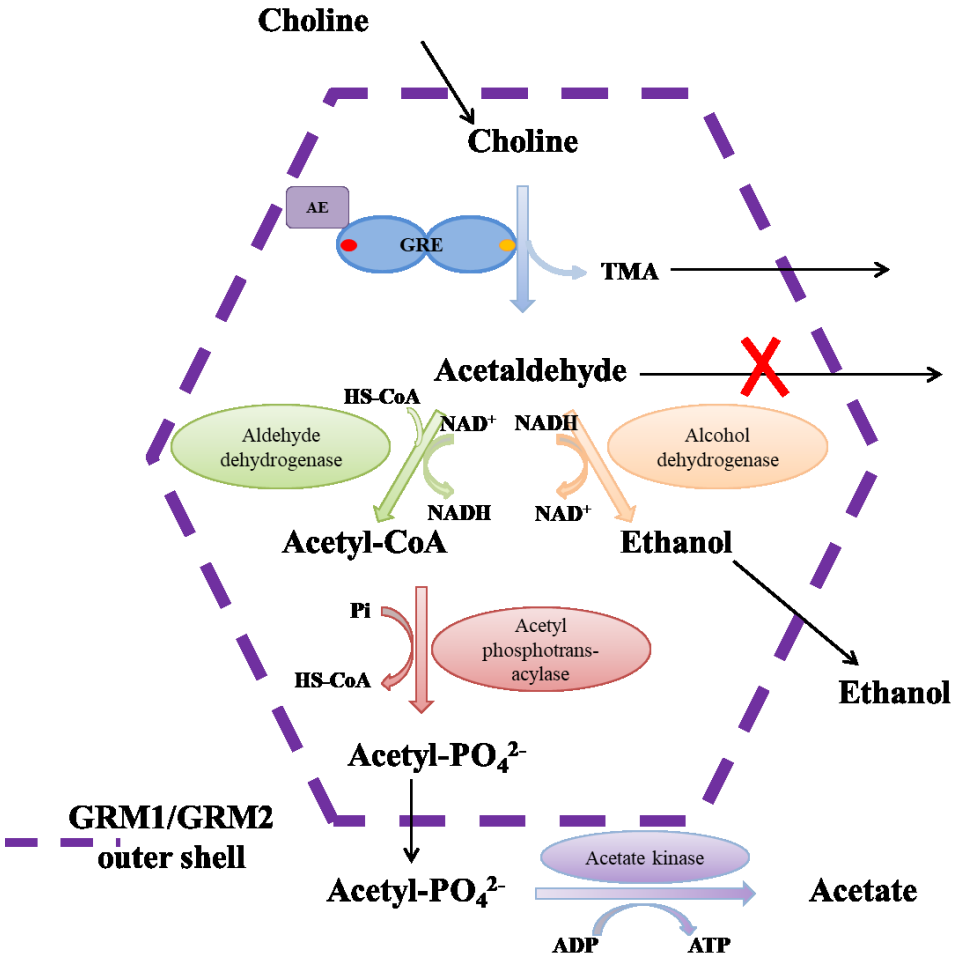


Figure 1.8. Schematic depicting the choline degradation pathway encapsulated in a GRM1/GRM2 BMC. Choline diffuses into the BMC where it is metabolised to acetaldehyde and TMA. Acetaldehyde is unable to diffuse through the outer protein shell and is metabolised to ethanol or acetyl-PO_4^{2-} as seen in the Eut BMC pathway.

The degradative pathway for anaerobic choline metabolism was for some time poorly understood. It was known that the first step was the cleavage of a C-N bond to form acetaldehyde and trimethylamine (Hayward and Stadtman, 1959), although the identity or mechanism of the enzyme catalysing this reaction were not known. It was hypothesised that the subsequent metabolism of acetaldehyde generated in this first step may follow a similar pathway as for the metabolism of acetaldehyde generated by the breakdown of ethanolamine

in *E. coli*. Database searches for genes similar to those of the *E. coli eut* operon not only identified genes for acetaldehyde degrading enzymes but also genes encoding choline trimethylamine lyase and its activating enzyme alongside homologues of the genes encoding the outer shell of the Eut BMC (Craciun and Balskus, 2012; Craciun *et al.*, 2014). This was the first evidence of a GRE associated with BMC proteins. The choline-metabolising GRMs are therefore predicted have a similar role to the Eut microcompartment, protecting the rest of the cell from the harmful effects of the acetaldehyde produced in the initial step of metabolism, preventing carbon loss through volatility and concentrating substrates in a confined environment to increase enzyme efficiency.

In separate studies, *Desulfovibrio desulfuricans* and *D. alaskensis* G20 when grown on choline were shown to upregulate the expression of choline trimethylamine lyase and its AE genes, but the genes with the highest upregulation were those encoding microcompartment shell proteins (Kuehl *et al.*, 2014; Martínez-del Campo *et al.*, 2015), further evidence for the involvement of BMCs in choline metabolism. In another study, *Proteus mirabilis*, a human gut-dwelling bacteria, was also shown to up-regulate expression of shell protein genes when grown on choline and the formation of microcompartments was observed in TEM images (Jameson *et al.*, 2016). In a similar study, TEM images demonstrated that BMCs were formed when the uropathogen *E. coli* 536 was grown on choline (Herring *et al.*, 2018). Together these studies provide compelling evidence to confirm the predicted substrate of the GRM1 and GRM2 microcompartments is indeed choline.

1.5.2.2 1,2-PD utilising GRMs

GRM3 and GRM4 loci encode a 1,2-PD dehydratase GRE alongside homologues of the enzymes for the metabolism of the propionaldehyde product to 1-propanol and propionate, similarly to the Pdu BMC. GRM5 loci also encode these same proteins but additionally encode upstream enzymes predicted to breakdown L-fuculose-phospahte, or L-rhamnulose-phosphate to 1,2-PD and lactaldehyde (Zarzycki *et al.*, 2015) (Fig. 1.9).

Experimental evidence for the involvement of BMCs in the breakdown of 1,2-PD using a GRE is minimal at present. Microarray analysis demonstrated that the genes for microcompartment shell proteins as well as a B_{12}-independent dehydratase were upregulated when *Roseburia inulinivorans*, an enteric bacterium with a GRM5 locus, was grown on fucose. The concomitant production of the 1,2-PD breakdown products, propionate and 1-propanol was also demonstrated (Scott *et al.*, 2006). The dehydratase GRE was confirmed as a selective 1,2-propanediol dehydratase in a later study (LaMattina *et al.*, 2016). Another GRM5 locus was identified in *Clostridium phytofermentans*, a soil dwelling bacteria able to ferment several plant sugars. When grown on fucose or rhamnose, microarray analysis revealed that genes

encoding microcompartment shell proteins were some of the most upregulated. TEM analysis revealed the presence of polyhedral cytoplasmic structures, presumed to be microcompartments, and HPLC analysis revealed the production of propanol and propionate in *C. phytofermentans* cultures grown on fucose or rhamnose (Petit *et al.*, 2013) supporting the prediction that the GRM5 type BMC is involved in the breakdown of fuculose/rhamnulose and 1,2-PD.

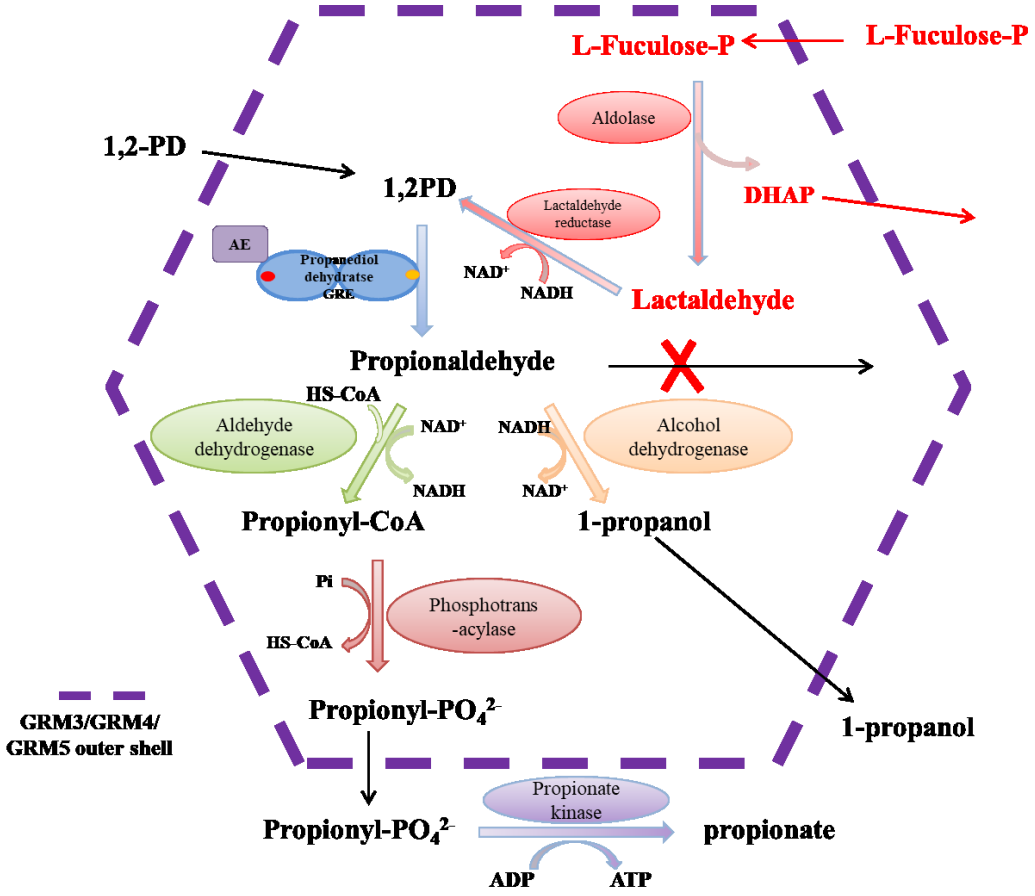


Figure 1.9. Schematic of the enzyme pathways encapsulated in GRM3, GRM4 and GRM5 microcompartments. GRM3 and GRM4 BMCs encapsulate a 1,2-PD dehydratase GRE and its AE alongside downstream enzymes for the subsequent breakdown of propionaldehyde into propionate or 1-propanol (Black text). GRM5 BMCs also encapsulate the upstream enzymes fucose-P-aldolase and lactaldehyde reductase for the conversion of L-fuculose-P or L-rhamnulose-P into 1,2-PD (red text).

At present there is little experimental data on BMCs of the GRM3 and 4 loci. The GRE of the *Rhodopseudomonas palustris* GRM3 loci was shown recently to be a specific 1,2-PD dehydratase (Zarzycki *et al.*, 2017). *In vitro* assays using the heterologously expressed GRE, its AE and the aldehyde dehydrogenase of the GRM3 locus demonstrated the enzymes to breakdown 1,2-PD to propionyl CoA as predicted for this type of BMC.

As with the Pdu BMC utilising B_{12} dependent diol dehydratase, the GRM3, 4 and 5 microcompartments would be predicted to function as a selectively permeable organelle for the containment of propionaldehyde, and lactaldehyde in the case of GRM5, so as to protect the rest of the cell from the toxic effects of these metabolites, prevent the volatile loss of carbon and increase enzyme efficiency by concentrating substrates in the vicinity of the pathway enzymes.

1.6 The role of BMCs in pathogenesis

Evidence for the role of BMCs in the pathogenicity of some bacterial species has only recently emerged. Many pathogens have genes for the formation of BMCs including *Salmonella, Escherichia, Yersinia, Clostridia* and *Shigella* (Jakobson and Tullman-Ercek, 2016) and several studies have demonstrated their up-regulation during infection. In *Salmonella*, the expression of *pdu* genes was shown to be induced in host tissues and *pdu* mutants exhibited a virulence defect in a mouse model of infection (Conner *et al.*, 1998; Heithoff *et al.*, 1999). Both *pdu* and *eut* gene expression is up-regulated in the intestines of chickens infected with *Salmonella* and colonisation of the gut by a *pduA* mutant was inhibited (Harvey *et al.*, 2011). Eut BMC genes are up-regulated during *Listeria monocytogenes* infection of Caco-2 cells, and intracellular replication is attenuated when infected with *eutB* mutant strains, indicating ethanolamine utilisation is essential for replication in mammalian cells (Joseph *et al.*, 2006). *Pdu* gene mutant strains of *Enterococcus faecalis* have attenuated growth in a *C. elegans* model of infection (Maadani *et al.*, 2007) and enterohaemorrhagic *E. coli* (EHEC) was shown to gain a competitive advantage over the endogenous bacteria of the bovine gut because of its ability to utilise ethanolamine (Bertin *et al.*, 2011). In addition, ethanolamine was found to activate virulence gene expression in EHEC cells (Kendall *et al.*, 2012). A link between BMCs and pathogenicity has clearly been established, however, whether this knowledge could be exploited for the treatment of pathogenic bacterial infections remains to be investigated. Significantly, the presence of BMCs within pathogenic bacteria during infection has not been demonstrated.

1.7 The BMC outer shell

The unifying feature of all BMCs is that they all have an outer shell that is composed entirely of protein. There are three main types of protein forming the shell: BMC-H, BMC-T and BMC-P, although BMC loci often encode multiple paralogues of these proteins, for example the Pdu BMC locus encodes PduA/B/J/K/N/T/U shell proteins.

1.7.1 BMC-H proteins

BMC-H, or hexameric, shell proteins are one of the major components of the outer shell of the BMC. They all have a conserved Pfam PF00936 domain known as the BMC domain. The first

crystal structures of BMC shell proteins were of two BMC-H proteins of the *Synechocystis* sp. PCC6803 carboxysome, CcmK2 and CcmK4 (Kerfeld, 2005). The BMC domain of each protein were seen to form an α/β fold that were almost identical to each other and subsequent crystal structures of a range of BMC-H proteins have revealed this tertiary structure to be conserved amongst all BMC domain proteins (Tsai *et al.*, 2007; Tanaka *et al.*, 2009; Crowley *et al.*, 2010; Takenoya *et al.*, 2010; Tanaka *et al.*, 2010). The BMC domain consists of four antiparallel beta-sheets surrounded by alpha helices that oligomerise to form cyclic hexamers (Fig 1.10), which are slightly curved to form a concave and a convex face. A pore is formed at the centre of the six subunits, the diameter ranging from 4 – 7 Å. This central pore was proposed to be a route for movement of metabolites, substrate or co-factors across the outer shell (Kerfeld, 2005).

The crystal structures of several BMC-H proteins also revealed tightly packed, side-to-side tiling of the hexamers to form extended layers of protein (Kerfeld, 2005; Tsai *et al.*, 2007; Tanaka *et al.*, 2009; Takenoya *et al.*, 2010; Tanaka *et al.*, 2010). Conserved residues involved in these crystal contacts suggested that these extended layers reflected the tiling of hexamers to form the facets of the BMC structure.

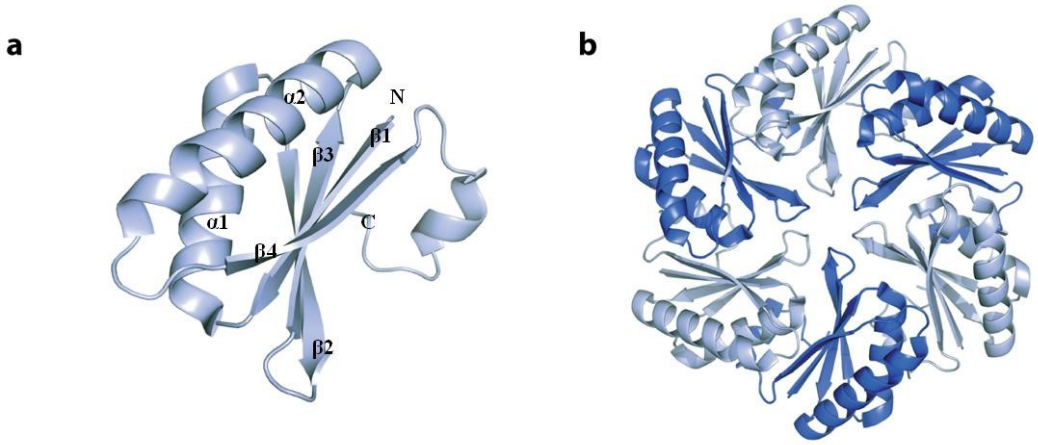


Figure 1.10. a) Typical structure of a BMC domain protein fold. Four beta sheets (labelled β1 – β4) surrounded by three alpha helices (labelled α1 – α3). b) Six shell proteins oligomerise to form a cyclic hexamer. Monomers are alternately coloured light and dark blue to highlight packing in the hexamer. Image adapted from Yeates, Crowley and Tanaka, 2010.

1.7.2 **BMC-T shell proteins**

BMC-T, or trimeric, shell proteins have two fused BMC domains and are encoded in most BMC loci but usually in smaller numbers than the BMC-H proteins (Axen, Erbilgin and Kerfeld, 2014).

Crystal structures have revealed that tandem BMC domain proteins oligomerise as trimers forming a pseudohexamer similar to the BMC-H hexamer (Takenoya *et al.*, 2010; Tanaka *et al.*, 2010; Pang *et al.*, 2012; Cai *et al.*, 2013) (Fig. 1.11). The central pore of the trimers is larger than hexamer pore, approximately 10 – 14 Å. EutL has been crystallised in two conformations, one with an open and one with a closed pore (Takenoya *et al.,* 2010; Tanaka, *et al.* 2010). Alternate side chain conformations of residues within the pore resulted in changes to the pore size, suggesting the possibility of a gating mechanism to allow passage of larger metabolites under certain conditions. BMC-T trimers are able to assemble into extended sheets similar to BMC-H hexamers (Sagermann *et al.* 2009; Tanaka *et al.* 2010; Cai *et al.*, 2013) suggesting they also contribute to the formation of the facets of the BMC.

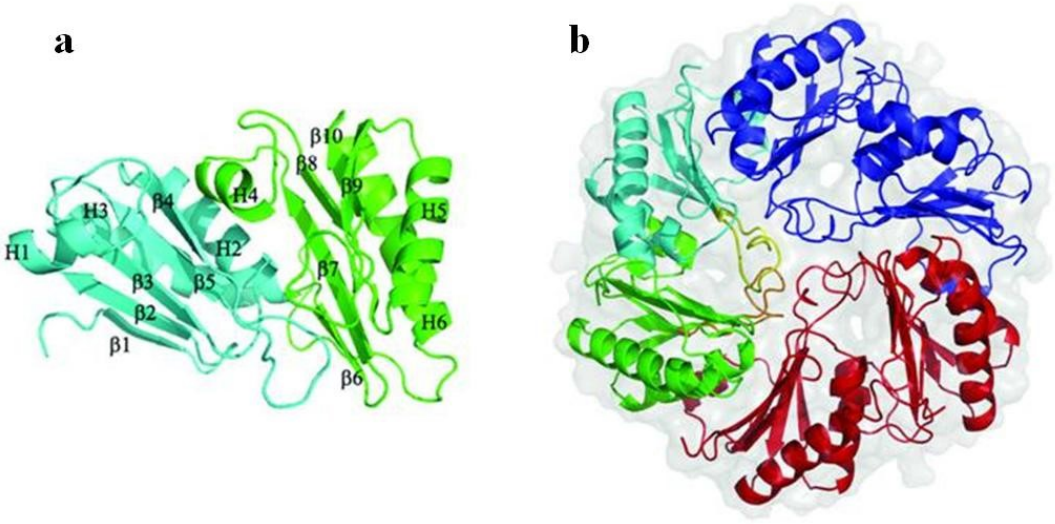


Figure 1.11. a) Example of a BMC-T protein structure, PduB from *Lactobacillus reuteri*. Two BMC domains, coloured cyan and green, are fused together. b) Three BMC-T proteins oligomerise to form a trimer or pseudohexamer. Figure adapted from Pang et al., 2012.

1.7.3 **BMC-P shell proteins**

BMC-P shell proteins contain the Pfam PF03319 domain. BMC-P proteins are found in all BMC loci but usually in smaller numbers than the PF00936 domain-containing shell proteins. How the hexameric and trimeric shell proteins were able to form icosahedral compartments

was unclear until the first crystal structures of BMC-P proteins were determined (Tanaka *et al.*, 2008). The PF03319 domain forms a 5 stranded beta-barrel with one alpha helix that then assemble into a symmetric pentamer with a concave and convex face giving a structure that resembles a pyramid with a pentamer base (Tanaka *et al.*, 2008; Sutter *et al.*, 2013; Wheatley *et al.*, 2013) (Fig. 1.12). Pentameric proteins at the vertices of an icosahedron allowed the modelling of a proposed BMC structure with the facets formed by tiled hexagons and the vertices formed by pentamers (Tanaka *et al.*, 2008) (Fig. 1.13). The importance of BMC-P vertex proteins in the formation of BMCs was demonstrated by the formation of aberrant, elongated structures in the absence of BMC-P protein (PduN) (Parsons *et al.*, 2010) and the impaired function of *Halothiobacillus neapolitanus* carboxysomes of BMC-P mutants (Cai *et al.*, 2009).

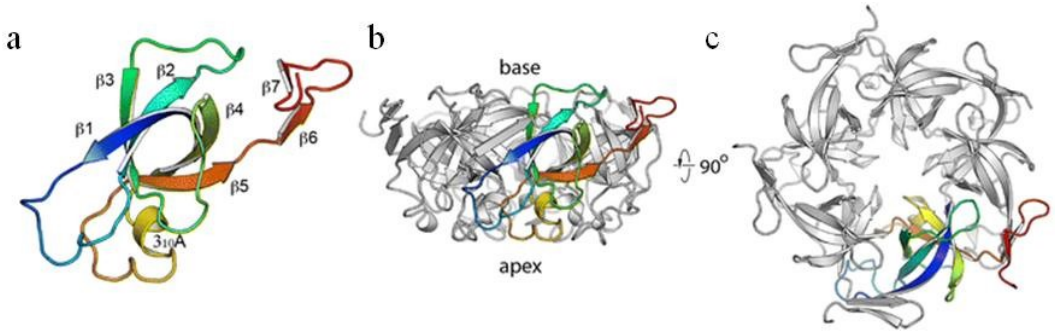


Figure 1.12. a) BMC-P monomer. b and c) 5 monomers assemble to form a pentamer with a concave and a convex surface forming a structure resembling a cone. Image adapted from Keeling *et al.*, 2014.

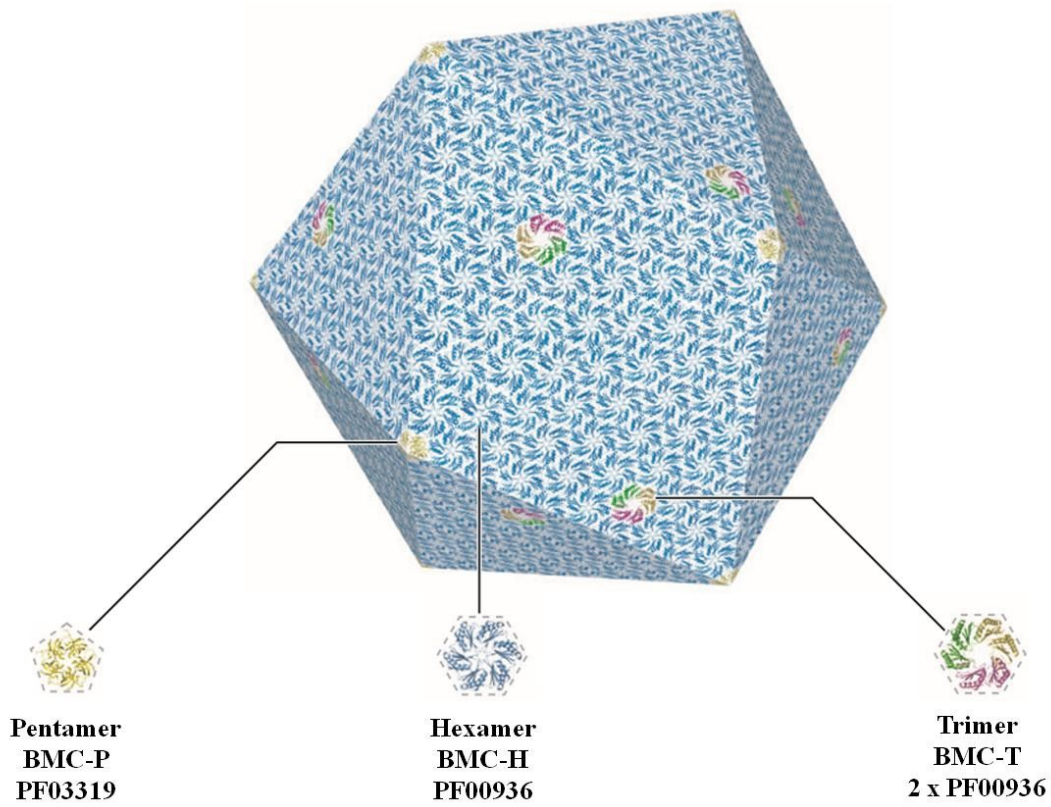


Figure 1.13. Structural model of a BMC. Hexameric and trimeric proteins tessellate to form the flat facets of the microcompartment whilst pentameric proteins cap the vertices to seal the microcompartment. Image from Kerfeld, Heinhorst and Cannon, 2010.

1.7.4 Variations in shell protein subtypes

Whilst most of the shell proteins of the BMC fall into the BMC-H or BMC-T classes there are several variations within these classes

1.7.4.1 Circularly permuted shell proteins

Several shell proteins have been identified that have their secondary structural elements in a different order to the canonical BMC domain although their tertiary structures are very similar. The first such circularly permuted structure identified was PduU from *S. enterica* (Crowley *et al.*, 2008) (BMC-H_P). The different order of the secondary structure elements results in the N and C termini being in different positions. The N terminus extends into the middle of the hexamer with residues from each monomer forming a six-stranded beta-barrel occluding the pore. EutS is also a circularly permuted hexamer with a beta-barrel occluding the pore (Tanaka *et al.*, 2010; Pitts *et al.*, 2012). Hexameric EutS from *E. coli* was also shown to be bent at an angle of approximately 40 º (Tanaka *et al.*, 2010). This was postulated as a possible means of generating the bend required at the edges of the facets in a polyhedral assembly.

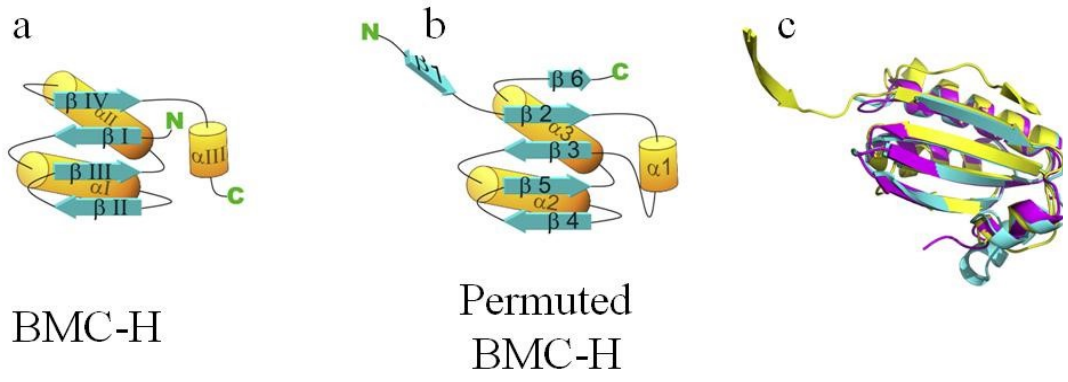


Figure 1.14. Comparison of the tertiary structures of typical BMC-H proteins and PduU (circularly permuted BMC-H). a) The classic fold of the BMC domain with the N and C termini on the same side of the structure. b) The circularly permuted BMC-H protein has the N and C termini on different sides of the structure, with the N – terminus forming an extended beta-sheet. c) Overlaying the BMC-H monomers (cyan and magenta) and the circularly permuted BMC-H protein (yellow) shows the similarity in the tertiary structure. Figure adapted from Crowley *et al.*, 2008.

Some BMC-T proteins are similarly circularly permuted. Examples include CsoS1D and CcmP (Klein *et al.*, 2009; Cai *et al.*, 2013; Larsson *et al.*, 2017).

1.7.4.2 Stacked or double layered BMC-T shell proteins

Another conformation that has been observed is the stacking of BMC-T trimers into double layers (BMC-T_{DB}). This phenomenon has been observed with *Synechococcus elongatus* CcmP and *Prochlorococcus* CsoS1D (Klein *et al.*, 2009; Cai *et al.*, 2013; Larsson *et al.*, 2017). In both cases the pseudohexameric trimers stacked in pairs to form dimers of trimers with the concave sides facing each other. Of additional interest is that the trimers were crystallised in two alternate conformations – one with an open pore, one with a closed pore. The open or closed state was caused by different conformations of the side chains of the pore residues. For both CcmP and CsoS1D, crystals forms had stacked trimers with one pore open and one pore closed (Klein *et al.*, 2009; Larsson *et al.*, 2017). Analysis of the conserved residues contributing to this stacked trimer interaction and the stability of the bonds formed suggest that the double layered trimer is biologically relevant, perhaps providing an 'airlock' arrangement for the passage of metabolites through the shell (Cai *et al.*, 2013).

1.7.4.3 Shell proteins with iron-sulphur cluster binding centres

Both hexameric and trimeric shell proteins have been identified that have iron-sulphur cluster binding sites within their central pore (denoted BMC-H_{FeS} and BMC-T_{FeS}).

Purified PduT, a trimeric shell protein, was shown by EPR analysis to bind a [4Fe-4S] centre. Mutagenesis of PduT cysteine residues identified Cys38 as the residue contributing to the Fe-

S binding (Parsons *et al.*, 2008). Crystal structures of PduT showed Cys38 to be located in the central pore of the trimer with one Cys38 from each subunit contributing to the [4Fe-4S] binding (Crowley *et al.*, 2010; Pang *et al.*, 2011).

GrpU is a circularly permuted hexameric shell protein from a glycyl radical microcompartment. The crystal structures of GrpU homologues from two organisms revealed the presence of an Fe-S binding site within the central pore (Thompson *et al.*, 2014). Modelling of [4Fe-4S] binding proposed that a cysteine residue from three of the six subunits of the hexamer contribute to the binding. Bioinformatic analysis showed GrpU homologues, with a conserved cysteine residue at the central pore position, to be present in many, although not all, GRM loci (Thompson *et al.*, 2014).

Several functions for the presence of an iron-sulphur cluster binding site within a shell protein pore have been proposed, including the transport of electrons through the shell, transport of intact Fe-S cluster through the shell to supply Fe-S cluster requiring proteins within the microcompartment, or as an electrochemical-sensing or oxygen level-sensing mechanism (Parsons *et al.*, 2008; Crowley *et al.*, 2010; Thompson *et al.*, 2014).

1.7.4.4 BMC-H shell proteins with C-terminal extensions

There are some BMC-H proteins that as well as a BMC domain also contain a C-terminal extension of variable length and unknown function (BMC-H_{ex}). There is limited data on the structure of these proteins with no full-length example having been crystallised. The C-terminal extension of EutK has, however, been crystallised and the helix-turn-helix structure has been shown to be very similar to known nucleic acid binding proteins (Tanaka *et al.*, 2010), perhaps indicating a role in the transcriptional control of BMC production. The structure and function of any other C-terminal extensions, which are hugely variable in length, remain to be elucidated.

1.7.5 Shell protein summary

Whilst the basic building blocks of the BMC outer shell, BMC-H, BMC-T and BMC-P, appear to be conserved across all the BMC types, there are numerous variations that could result in a vast array of different properties of the BMC shell, dependent upon the shell protein composition. Differences in pore size and charge will allow different sized and charged metabolites to pass through, potential gating mechanisms could regulate metabolite movement, redox balance could be controlled through Fe-S cluster binding pores and slight differences in hexamer shapes could influence the overall size and shape of the BMC.

1.8 BMC assembly

The mechanisms for assembly of BMCs have been elucidated for the α- and β- carboxysomes, however the mechanism for assembly of metabolosomes is yet to be determined.

1.8.1 β-carboxysome assembly

The mechanism for β-carboxysome assembly began to be elucidated when it was observed that two proteins whose functions were unknown, CcmM and CcmN, were essential for carboxysome formation. Deletion mutants of either CcmM or CcmN resulted in high CO_2 requiring phenotypes and the formation of large polar aggregates instead of carboxysomes (Ludwig *et al.*, 2000; Kinney *et al.*, 2012). Live cell imaging studies with fluorescently labelled RuBisCO showed the initial formation of a RuBisCO aggregate (pro-carboxysome) that subsequently budded off to form carboxysomes (Cameron *et al.*, 2013; Chen *et al.*, 2013). This was correlated with TEM images over the same time course showing the same process. The same protocol was followed with deletion mutants of CcmK, CcmM and CcmN to determine the protein-protein interactions involved at each stage of the process. CcmM was shown to be essential for the initial aggregation of RuBisCO forming the pro-carboxysome. CcmM has an N-terminal carbonic anhydrase (CA) domain and 3-5 RuBisCO small subunit-like domains (SSLDs). Truncation studies determined it was the SSLDs that interacted with RuBisCO to form aggregation (Cameron *et al.*, 2013). CcmM also interacts with carboxysomal carbonic anhydrase (McGurn *et al.*, 2016) drawing it in to the pro-carboxysome. The N-terminal CA domain of CcmM interacts with the N-terminal domain of CcmN, also bringing CcmN into the pro-carboxysome aggregate. A short C-terminal domain of CcmN interacts with shell proteins (Kinney *et al.*, 2012) forming fully encapsulated carboxysomes which then bud off from the pro-carboxysome.

1.8.2 α-carboxysome assembly

In contrast to the 'inside-out' formation of β-carboxysomes, the interior and shell of the α-carboxysomes assemble simultaneously. Cryo-electron tomography has captured images of partially formed carboxysomes with cargo attached to the interior (Iancu *et al.*, 2010) supporting this mechanism. CsoS2 has been shown to be a key requirement in the assembly of α-carboxysomes. CsoS2 knock-outs do not form carboxysomes (Cai *et al.*, 2015). CsoS2 is a protein formed from three domains, all of which contribute to the assembly of the carboxysome. The N-terminal domain interacts with shell proteins, recruiting them until a concentration sufficient for self-assembly is achieved. Simultaneously, the middle domain interacts with the small subunit of RuBisCO causing it to coalesce in a lattice around the middle. The large subunit of RuBisCO in turn interacts with CsoS1shell proteins (Liu *et al.*,

2018), recruiting them to the assembling carboxysome. The C-terminal domain of CsoS2 anchors the growing shell protein sheet to the enzymatic core.

1.8.3 Metabolosome assembly

The mechanism of metabolosome assembly has yet to be established, however, there are clues as to the likely method. Encapsulation peptides (EPs), small α-helical domains at the N- and C-termini (Fan *et al.*, 2012; Aussignargues *et al.*, 2015), have been identified as features of many metabolosome-located enzymes and have been shown to interact with shell proteins. EPs have also been shown to make enzymes less soluble (Zarzycki *et al.*, 2017) and to form aggregates within the cytoplasm (Lee *et al.*, 2016) similar to pro-carboxysome aggregation. Pdu shell protein deletion mutants are also seen to form aggregates of cargo protein, demonstrating that shell proteins are not necessary for the formation of a pre-metabolosome (Havemann *et al.*, 2002; Cheng *et al.*, 2011). The combined actions of aggregating protein and interacting with shell proteins suggests a mechanism similar to that of β-carboxysomes, with the formation of a pre-metabolosome followed by encapsulation by the shell.

1.9 Recombinant BMCs and Bioengineering

1.9.1 Recombinant BMCs

The presence BMC loci in many diverse bacterial phyla is evidence of multiple horizontal gene transfer events (AbdulRahman, 2013). These natural occurrences were replicated in the first example of BMC bioengineering when the whole Pdu operon of *Citrobacter freundii* was transferred into *E. coli* resulting in a strain capable of metabolising 1,2-PD and exhibiting polyhedral structures within the cytoplasm that resemble BMCs in wild-type *C. freundii* (Parsons *et al.*, 2008). The transfer of genes to generate functional BMCs in bacteria not normally capable of producing them has been demonstrated further with the production of α-carboxysomes from *Halothiobacillus neapolitanus* in *E. coli* (Bonacci *et al.*, 2012) and functional β-carboxysomes produced in *E. coli* expressing a synthetic operon containing all 12 genes from *Synecococcus elongatus* (Fang *et al.*, 2018).

Examples of recombinant BMCs transferred into bacteria other than *E. coli* include the transfer of a β-carboxysome operon into *Corynebacterium glutamicum*, a Gram-positive bacterium used in the industrial production of amino acids and proteins, which produced polyhedral structures within the cytoplasm and produced functional RuBisCO (Baumgart *et al.*, 2017). Also, the transfer of the Pdu operon into a range of Gram-negative bacteria resulted in a number of strains capable of 1,2-PD metabolism and forming polyhedral organelles visualised by TEM following purification (Graf *et al.*, 2018).

Eukaryotic expression of BMC genes was demonstrated in the successful transfer of β-carboxysome shell and core protein genes into *Nicotiana benthamiana*, with the proteins targeted to the chloroplast resulting in the production of carboxysome like assemblies (Lin *et al.*, 2014). This was followed by the production of α-carboxysome shell proteins encapsulating functional *Cyanobium* RuBisCO in *Nicotiana tabacum* chloroplasts (Long *et al.*, 2018). These initial experiments are the first steps in the ultimate goal of improving photosynthetic CO_2 fixation in crop plants through the bioengineering of cyanobacterial carbon concentrating mechanisms into plants (Zarzycki *et al.*, 2013).

1.9.2 **Recombinant empty BMCs**

A major step forward in the bioengineering of BMCs was the demonstration that empty BMCs could be formed following the heterologous expression of the 7 shell protein genes of the *C. freundii* Pdu microcompartment in *E. coli* (Parsons *et al.*, 2010). Varying the shell proteins produced in this study provided information on the importance of individual genes for the successful formation of microcompartment structures resulting in the identification of a minimum 5 shell protein complement for microcompartment formation. The production of recombinant empty Eut microcompartments was achieved with expression of the 5 *S. enterica* Eut shell protein genes in *E. coli* (Choudhary *et al.*, 2012). Interestingly, production of a single shell protein, EutS, in *E. coli* resulted in the formation of apparently fully formed empty shells morphologically similar to those formed with the full shell protein complement (Choudhary *et al.*, 2012). The β-carboxysome shell proteins of *Halothece* sp. PCC 7418 were produced in *E. coli* with the formation of polyhedral structures that could be visualised by TEM following purification, although the diameter of these shells were much smaller (~25 nm) than normal fully packed β-carboxysomes (Cai *et al.*, 2016).

The successful formation of empty microcompartments is not limited to the well-characterised carboxysomes and Pdu and Eut BMCs. *Haliangium ochraceum* is a halophilic myxobacterium harbouring shell protein genes for a BMC of unknown function. The BMC from this organism was chosen for study because of its unusual shell protein complement with 3 BMC-T encoding genes, 2 BMC-P encoding genes and 1 BMC-H encoding gene distributed over 3 chromosomal locations (Lassila *et al.*, 2014). Expression of all 7 shell protein genes in *E. coli* resulted in the formation of empty shells that could be purified and imaged by TEM. The shells were small in diameter (~40 nm) but much more uniform in shape than previously purified metabolosome shells (Lassila *et al.*, 2014). These shells however, could not be visualised *in vivo* by TEM. The authors speculated this could be due to the thickness of the TEM sections being similar to the diameter of the shells precluding clear differentiation of the shells from the background of the cell.

1.9.3 The structure of an empty BMC

In a remarkable experiment, *H. ochraceum* empty shells were crystallised and X-ray diffraction data, together with cryo-EM data, were used to solve the structure of a whole empty microcompartment (Sutter *et al.*, 2017) (Fig. 1.15). BMC-H proteins were the predominant tiling unit forming the facets, stacked BMC-T proteins were inserted in the centre of the facets and BMC-P proteins capped the vertices conforming to the predicted model of BMC structure (Tanaka *et al.*, 2008). The structure revealed 4 protein-protein interactions: 2 hexamer-hexamer interactions, 1hexamer-pentamer interaction and 1 hexamer-trimer interaction. The high structural similarity between all hexamer and pentamer proteins would suggest that these interactions are likely to be conserved across all BMC structures. A crucial finding was that the hexamers are oriented so that the concave surface faces out towards the cytoplasm providing information on the position of the N- and C-termini of the hexameric units (Sutter *et al.*, 2017), which, if this is the case for other BMC assemblies, may provide essential information for the bioengineering of shell components.

A recent study has confirmed this orientation in a synthetic β-carboxysome, composed of *Halothece* sp. PCC 7418 shell proteins, whose structure was determined through cryo-electron microscopy single particle analysis (Sutter *et al.*, 2019) (Fig. 1.15). The shells formed in this study were of a few different sizes (210 – 310 Å) but all appeared to be composed of just 2 BMC-H proteins, CcmK1 and CcmK2 and a BMC-P protein, CcmL with no BMC-T protein observed despite its co-expression and presence during the purification procedure. The absence of the BMC-T protein in the final structure allowed the interactions at the interface where 3 hexamers meet to be observed, which was not possible in the *H. ochraceum* structure. Again residues at this interface are highly conserved having small side chains that prevent steric clashes. This synthetic β-carboxysome is very small (~25 nm in diameter) in comparison to natural carboxysomes. This may be due to β-carboxysomes normally forming around a pre-aggregated cargo which may then result in the formation of larger shells.

These two studies have provided a wealth of information about the structure of BMCs including information on the orientation of each of the subunits - hexamers, trimers and pentamers – and how stacked trimers fit into the structure, so giving a sound basis for the future design of synthetic compartments.

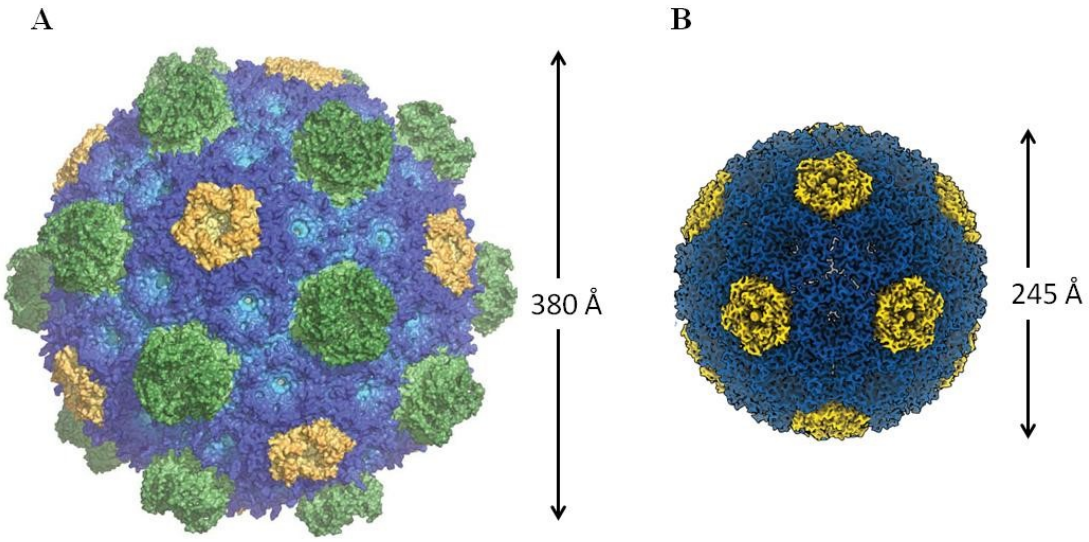


Figure 1.15. The structure of A) an empty *H. ochraceum* BMC and B) an empty *Halothece* β-carboxysome determined by cryo-EM and X-ray crystallography. Hexameric subunits in blue, pentameric subunits in yellow and stacked trimeric subunits in green. Images reproduced from Sutter *et al.*, 2017, 2019.

1.9.4 Targeting to BMCs

1.9.4.1 Encapsulation peptides

The production of BMCs without their native enzyme pathway enclosed allows the possibility of targeting heterologous pathways to the microcompartment interior. Small molecules, such as some substrates and co-factors, are predicted to access the interior through the pores of the shell proteins (Kerfeld, 2005). Larger molecules, unable to gain access through the pores, are incorporated through short, peptide sequences, which form an amphipathic helix, approximately 20 amino acids long and termed encapsulation peptides (EPs). EPs have been identified, through the targeting of fluorescent proteins, at the N-terminus of several core enzymes, e.g. PduP (aldehyde dehydrogenase) (Fan *et al.*, 2010, 2012; Lawrence *et al.*, 2014), PduD (subunit of diol dehydratase) (Fan and Bobik, 2011; Sargent *et al.*, 2013), EutC (ethanolamine ammonia lyase (Choudhary *et al.*, 2012), and Hoch_5813 (aldehyde dehydrogenase) (Lassila *et al.*, 2014). A C-terminal EP was identified for CcmN (Kinney *et al.*, 2012) and internal EP has been postulated for the glycyl radical enzymes of some GRMs (Zarzycki *et al.*, 2015) but this has not been experimentally validated.

1.9.4.2 Alternative encapsulation methods

Targeting of proteins to the BMC lumen has also been achieved without the use of native EPs. Signal sequence were created both through rational design and library screening and linked to fluorescent markers to demonstrate effective encapsulation within BMCs (Jakobson *et*

*al.*2017). Another approach exploited coiled-coil protein-protein interactions to target a fluorescent protein either to the outside or inside of a BMC (Lee *et al.*, 2018). A pair of coiled-coils was designed to interact specifically with each other without self-interaction. One of the coiled-coil pair was linked to a fluorescent marker whilst the other coiled-coil was linked to either the N-terminus of PduA or the N-terminus of a permuted version of PduA which has its N-terminus on the opposite face of the hexamer to the native form. SpyCatcher/SpyTag technology has also been used to target fluorescent proteins to the lumen of *H. ochraceum* compartments (Hagen *et al.*, 2018). SpyCatcher and SpyTag are polypeptides that form a covalent link when they are in close proximity. In this study one peptide was linked to a BMC-T protein so that they presented on the internal face of the BMC when forming a trimer, whilst the other peptide was attached to a fluorescent protein. This technology allowed not only the targeting of proteins to the BMC lumen but two proteins could be loaded in programmed ratios.

1.9.4.3 Targeting heterologous proteins to BMCs

The ability to produce empty microcompartment shells opened up the opportunity to re-purpose BMCs by targeting non-native enzyme pathways to produce commercial chemicals or to protect cells from toxic products and intermediates. An initial study targeted β-galactosidase to the Eut microcompartment shell (Choudhary *et al.*, 2012). The hydrolysis of its substrate X-gal produces an insoluble coloured indole. Discrete accumulation of this product could be observed in cells that were also producing microcompartment shell proteins, compared to the diffuse distribution in control cells so demonstrating a functional enzyme following encapsulation. The first example of re-purposing a BMC to produce a chemical was the targeting of a two-enzyme pathway for the production of ethanol in the Pdu microcompartment (Lawrence *et al.*, 2014). Ethanol production was increased when targeted to the interior of the microcompartment. A further study demonstrated the targeting of a number of different individual enzymes to the Pdu empty shell that retained their activity upon purification and were also protected by the compartment shell from changes to environmental pH (Wagner *et al.*, 2017). More complicated or longer enzyme pathways have yet to be targeted to a BMC interior and the ability to scale up production is also a goal waiting to be achieved.

1.9.5 Bioengineering of shell proteins

The recombinant production of single shell proteins, mainly in *E. coli*, has been shown to produce an array of different architectures, despite their close structural similarity. Some proteins have been shown to form tubes when over-produced (Parsons *et al.*, 2010; Pang *et al.*, 2014; Noël, Cai and Kerfeld, 2016). Other proteins form filaments (Havemann *et al.* 2002; Heldt *et al.*, 2009; Pang *et al.*, 2014) or rolled sheets (Pitts *et al.*, 2012; Sutter *et al.*, 2016) and some form no visible structures at all (Young *et al.*, 2017).

More recently, studies have exploited the structures formed by these single shell proteins as a scaffold for channelling through an enzyme pathway. Initially, targeting of a fluorescent protein to PduA tubes was demonstrated in *C. glutamicum*, using EPs to target the peptide to the scaffold (Huber *et al.*, 2017). Production of PduA*, a variant of PduA with a short C-terminal extension (Parsons *et al.*, 2010), attached to a *de novo* designed coiled-coil results in the formation of a shorter tubes than seen with un-tagged protein which, instead of running parallel to the cell, form a matrix of filaments throughout the cytoplasm (Lee *et al.*, 2018). Targeting two enzymes for the production of ethanol, via another coiled-coil complementary to that attached to PduA*, resulted in a 221% increase in ethanol production, demonstrating the potential for improved enzyme activity when attached to a shell protein scaffold. A similar scaffolding approach was used to test a dual enzyme pathway for chiral amine production (Zhang *et al.* 2018). SpyCatcher/SpyTag technology was used to immobilize enzymes to a EutM scaffold. There did not appear to be any increase in total turnover but the reaction reached completion at a faster rate and the enzymes had improved stability when attached to the scaffold. These results have established a foundation for future studies examining the benefits of tethering enzyme cascades to a protein scaffold utilising BMC shell protein engineering.

1.10 This study

There have been many advances made in the understanding of BMC function and architecture over recent years. There are, however, a plethora of BMC systems that have been identified through bioinformatic analysis but not characterised experimentally. The benefits of exploring a previously uncharacterised system was demonstrated by the study of the *H. ochraceum* BMCs which resulted in the production of a remarkably regular empty BMC, smaller than those previously characterised and incredibly pliable in terms of assembly both *in vivo* and *in vitro* (Lassila *et al.*, 2014; Sutter *et al.*, 2017; Hagen *et al.*, 2018; Greber *et al.*, 2019). This study has therefore set out to investigate two other previously uncharacterised BMCs, both belonging to the vastly under-explored GRM family.

Rhodospirillum rubrum and *Clostidium autoethanogenum* both harbour GRM encoding operons within their genomes. *R. rubrum* is Gram-negative alphaproteobacterium, found in a range of environments including lakes, streams, standing water, mud and sewage (Reslewic *et al.*, 2005). It is a photosynthetic, facultative anaerobe, capable of heterotrophic or autotrophic growth. *R. rubrum* has been the focus of a number of biotechnological studies. It is of interest due to its ability to produce poly beta-hydoxyalkanoates (PHAs) which are biodegradable forms of plastic (Brandl *et al.*, 1989; Ulmer *et al.*, 1994; Revelles *et al.*, 2016). Other studies have focussed on its mechanism of radiation resistance (Mastroleo *et al.*, 2009). *R. rubrum* has

a single BMC operon within its genome classified as GRM3 (Axen, Erbilgin and Kerfeld, 2014).

Clostridium autoethanogenum is a Gram-positive, anaerobic, acetogenic bacterium. *C. autoethanogenum* belongs to a very small group of acetogens that are also capable of producing ethanol from carbon monoxide in syngas fermentation (Abrini, Naveau and Nyns, 1994; Norman *et al.*, 2018). The genome of a closely related species, *C. ljungdhalii*, was identified in a bioinformatic study as having two GRM loci, a GRM1 and a GRM3 (Axen *et al.,* 2014). BLAST searches revealed almost identical operons within the *C. autoethanogeum* genome.

These studies were undertaken with a number of aims. Firstly, to investigate if recombinant empty microcompartments could be formed with the shell proteins of these two organisms in a similar manner to the *C. freundii* Pdu microcompartment. Secondly, as the recombinant empty microcompartments formed by Pdu shell proteins are somewhat irregular in size and shape, the aim was to see if compartments could be formed that were more uniform as this may provide more information about compartments structure, perhaps through crystallographic analysis. Finally, as the project progressed, the role of the C-terminal extension of some BMC-H proteins (BMC-H_{ex}) was explored in order to understand how they interact with other cellular components.

Chapter 2

Materials and methods

2.1 Materials

2.1.1 Chemicals

Most chemicals and antibiotics were purchased from Sigma-Aldrich Ltd. Other materials were purchased from manufacturers as follows: IPTG, ampicillin and agarose from Melford Laboratories Ltd; chelating fast flow sepharose, disposable and empty PD10 columns from GE Healthcare; QIAprep® Spin Miniprep Kit, and QIAquick® Gel Extraction Kit from Qiagen; restriction enzymes were purchased from Promega and from New England Biolabs Inc; B-PER bacterial protein extraction reagent and BCA protein concentration assay kit from Fisher scientific; n-Octyl-beta-D-thioglucopyranoside from Apollo Scientific Ltd.; cloning vectors from Novagen, Invitrogen or Promega; tryptone, yeast extract and bacterial agar from Oxoid Ltd; Roche FastStart High Fidelity PCR System from Roche Diagnostics GmbH; Low viscosity resin, low viscosity hardener VH1, low viscosity hardener VH2, low viscosity accelerator, 400mesh copper grids, 200mesh copper grids and copper slot grids were purchased from Agar Scientific; genomic DNA and freeze dried *Rhodospirillum rubrum* from DSMZ; crystallisation screens were purchased from Molecular Dimensions Ltd.

2.1.2 Bacterial strains

Bacterial strains were purchased from Novagen, Invitrogen or Promega

Table 2.1. Bacterial strains

E. coli strain	Genotype	Reference
DH10ß	F- *mcr*A Δ(*mrr-hsd*RMS-*mcr*BC) φ80*lac*ZΔM15 Δ*lac*X74 *rec*A1 *end*A1 *ara*D139 Δ (*ara*, *leu*)7697 *gal*U *gal*K λ- *rps*L *nup*G /pMON14272 / pMON7124	Invitrogen
JM109	*endA1 glnV44 thi-1 relA1 gyrA96 recA1 mcrB^{+} Δ(lac-proAB) e14- [F' traD36 proAB^{+} lacI^{q} lacZΔM15] hsdR17($r_K^{-}m_K^{+}$)*	Promega
BL21 (DE3)	F- *ompT hsdSB (rB-mB-) gal dcm* (DE3)	Invitrogen
BL21 star (DE3)	F- *ompT hsdS*B (rB-mB-) *gal dcm rne131* (DE3)	Invitrogen

BL21 star (DE3) pLysS	*ompT hsdS*B (rB-mB-) *gal dcm* *rne131* (DE3) pLysS (CamR)	Invitrogen
Rosetta 2 (DE3) pLysS	F^- *ompT* $hsdS_B(r_B^- m_B^-)$ *gal dcm* (DE3) pLysSRARE2 (Cam^R)	Novagen

2.1.3 **Plasmids**

The plasmids used in this study are listed in Table 2.2.

Table 2.2. Plasmids *Caethg*: *C. autoethanogenum*

Plasmid	Description	Reference or source
pET3a	Cloning and expression vector; T7 promotor, Amp^R	Novagen
pET14b	Cloning and expression vector; N terminal His_6 tag; T7 promotor; Amp^R	Novagen
pETcoco-2	Cloning and expression vector; regulation of the copy number with glucose and arabinose: single-copy origin of replication (*oriS*) and medium copy origin of replication (*oriV*); T7*lac* promotor; Amp^R	Novagen
pLysS	Expression vector containing T7 lysozyme gene; Cam^R	Novagen
pET3a *caethg 3278*	*Caethg* shell protein gene *3278* cloned into NdeI SpeI sites of pET3a	This study
pET3a *caethg 3283*	*Caethg* shell protein gene *3283* cloned into NdeI SpeI sites of pET3a	This study
pET3a *caethg 3284*	*Caethg* shell protein gene *3284* cloned into NdeI SpeI sites of pET3a	This study
pET3a *caethg 3286*	*Caethg* shell protein gene *3286* cloned into NdeI SpeI sites of pET3a	This study
pET3a *caethg 3289*	*Caethg* shell protein gene *3289* cloned into NdeI SpeI sites of pET3a	This study
pET3a *caethg 3290*	*Caethg* shell protein gene *3290* cloned into NdeI SpeI sites of pET3a	This study

pET3a *caethg 3278-83-84-86-89-90*	Link-and-lock of *Caethg* shell protein genes *3278-3283-3284-3286-3289-3290*	This study
pET3a *caethg 3278-83-84-86-89*	Link-and-lock of *Caethg* shell protein genes *3278-3283-3284-3286-3289*	This study
pET3a *caethg 3278-83-84-86-90*	Link-and-lock of *Caethg* shell protein genes *3278-3283-3284-3286-3290*	This study
pET3a *caethg 3278-83-84-89-90*	Link-and-lock of *Caethg* shell protein genes *3278-3283-3284-3289-3290*	This study
pET3a *caethg 3278-83-86-89-90*	Link-and-lock of *Caethg* shell protein genes *3278-3283-3286-3289-3290*	This study
pET3a *caethg 3278-84-86-89-90*	Link-and-lock of *Caethg* shell protein genes *3278-3284-3286-3289-3290*	This study
pET3a *caethg 3283-84-86-89-90*	Link-and-lock of *Caethg* shell protein genes *3283-3284-3286-3289-3290*	This study
pET14b *caethg 3286*	*Caethg* shell protein gene *3286* cloned into NdeI SpeI sites of pET14b	This study
pET3a *caethg 3286BMC_80*	Truncation of *Caethg* shell protein gene *3286* encoding just the first 80 amino acids cloned into NdeI SpeI sites of pET3a	This study
pET3a *caethg 3286 BMC_90*	Truncation of *Caethg* shell protein gene *3286* encoding just the first 90 amino acids cloned into NdeI SpeI sites of pET3a	This study
pET3a *caethg 3286 BMC_100*	Truncation of *Caethg* shell protein gene *3286* encoding just the first 100 amino acids cloned into NdeI SpeI sites of pET3a	This study
pET3a *caethg 3286 BMC_110*	Truncation of *Caethg* shell protein gene *3286* encoding just the first 110 amino acids cloned into NdeI SpeI sites of pET3a	This study
pET3a *caethg 3286 BMC_120*	Truncation of *Caethg* shell protein gene *3286* encoding just the first 120 amino acids cloned into NdeI SpeI sites of pET3a	This study
pET3a *caethg 3286 BMC_130*	Truncation of *Caethg* shell protein gene *3286* encoding just the first 130 amino acids cloned into NdeI SpeI sites of pET3a	This study

pET3a *caethg 3286 BMC_140*	Truncation of *Caethg* shell protein gene *3286* encoding just the first 140 amino acids cloned into NdeI SpeI sites of pET3a	This study
pET3a *caethg 3286 BMC_150*	Truncation of *Caethg* shell protein gene *3286* encoding just the first 150 amino acids cloned into NdeI SpeI sites of pET3a	This study
pET3a *caethg 3286 BMC_170*	Truncation of *Caethg* shell protein gene *3286* encoding just the first 170 amino acids cloned into NdeI SpeI sites of pET3a	This study
pET3a *caethg 3286 BMC_190*	Truncation of *Caethg* shell protein gene *3286* encoding just the first 190 amino acids cloned into NdeI SpeI sites of pET3a	This study
pET3a *caethg 3286 BMC_230*	Truncation of *Caethg* shell protein gene *3286* encoding just the first 230 amino acids cloned into NdeI SpeI sites of pET3a	This study
pET3a *caethg 3286 BMC_242*	Truncation of *Caethg* shell protein gene *3286* encoding just the first 242 amino acids cloned into NdeI SpeI sites of pET3a	This study
pET3a *caethg* 3286C	Truncation of *Caethg* shell protein gene *3286* encoding just the last 168 amino acids of the C-terminal extension cloned into NdeI SpeI sites of pET3a	This study
pET3a *caethg 3278-83-84-(86BMC_80)-89-90*	Link-and-lock of *Caethg* shell protein genes *3278-3283-3284-3286BMC80 (truncation)-3289-3290*	This study
pET3a *caethg 3278-83-84-(86BMC_90)-89-90*	Link-and-lock of *Caethg* shell protein genes *3278-3283-3284-3286BMC90 (truncation)-3289-3290*	This study
pET3a *caethg 3278-83-84-(86BMC_100)-89-90*	Link-and-lock of *Caethg* shell protein genes *3278-3283-3284-3286 BMC100 (truncation)-3289-3290*	This study
pET3a *caethg 3278-83-84-(86BMC_110)-89-90*	Link-and-lock of *Caethg* shell protein genes *3278-3283-3284-3286 BMC110 (truncation)-3289-3290*	This study

pET3a *caethg 3278-83-84-(86BMC_120)-89-90*	Link-and-lock of *Caethg* shell protein genes *3278-3283-3284-3286 BMC120 (truncation)-3289-3290*	This study
pET3a *caethg 3278-83-84-(86BMC_130)-89-90*	Link-and-lock of *Caethg* shell protein genes *3278-3283-3284-3286 BMC130 (truncation)-3289-3290*	This study
pET3a *caethg 3278-83-84-86BMC_140-89-90*	Link-and-lock of *Caethg* shell protein genes *3278-3283-3284-3286 BMC140 (truncation)-3289-3290*	This study
pET3a *caethg 3278-83-84-(86BMC_150)-89-90*	Link-and-lock of *Caethg* shell protein genes *3278-3283-3284-3286 BMC150 (truncation)-3289-3290*	This study
pET3a *caethg 3278-83-84-(86BMC_170)-89-90*	Link-and-lock of *Caethg* shell protein genes *3278-3283-3284-3286 BMC170 (truncation)-3289-3290*	This study
pET3a *caethg 3278-83-84-(86BMC_190)-89-90*	Link-and-lock of *Caethg* shell protein genes *3278-3283-3284-3286 BMC190 (truncation)-3289-3290*	This study
pET3a *caethg 3278-83-84-(86BMC_230)-89-90*	Link-and-lock of *Caethg* shell protein genes *3278-3283-3284-3286 BMC230 (truncation)-3289-3290*	This study
pET3a *caethg 3278-83-84-(86BMC_242)-89-90*	Link-and-lock of *Caethg* shell protein genes *3278-3283-3284-3286 BMC242 (truncation)-3289-3290*	This study
pET3a *caethg 3286BMC_C28*	*Caethg* shell protein gene *3286* encoding the BMC domain plus the last 28 amino acids of the C-terminus cloned into the NdeI/SpeI sites of pET3a	This study
pET3a *caethg 3286BMC_C63*	*Caethg* shell protein gene *3286* encoding the BMC domain plus the last 63 amino acids of the C-terminus cloned into the NdeI/SpeI sites of pET3a	This study
pET3a *caethg 3278-83-84-(3286BMC_C28)-89-90*	Link-and-lock of *Caethg* shell protein genes *3278-3283-3284-3286 BMC_C28-3289-3290*	This study
pET3a *caethg 3278-83-84-(3286BMC_C63)-89*	Link-and-lock of *Caethg* shell protein genes *3278-3283-3284-3286 BMC _C63-3289*	This study
pET3a *caethg 1832*	*Caethg* shell protein gene *1832* cloned into NdeI SpeI sites of pET3a	This study

pET3a *caethg 1831*	*Caethg* shell protein gene *1831* cloned into NdeI SpeI sites of pET3a	This study
pET3a *caethg 1825*	*Caethg* shell protein gene *1825* cloned into NdeI SpeI sites of pET3a	This study
pET3a *caethg 1824*	*Caethg* shell protein gene *1824* cloned into NdeI SpeI sites of pET3a	This study
pET3a *caethg 1822*	*Caethg* shell protein gene *1822* cloned into NdeI SpeI sites of pET3a	This study
pET3a *caethg 1820*	*Caethg* shell protein gene *1820* cloned into NdeI SpeI sites of pET3a	This study
pET3a *caethg 1817*	*Caethg* shell protein gene *1817* cloned into NdeI SpeI sites of pET3a	This study
pET3a *caethg 1816*	*Caethg* shell protein gene *1816* cloned into NdeI SpeI sites of pET3a	This study
pETcoco-2 *caethg 1832-31-25-24-22-20-17-16*	Link and lock of *Caethg* shell protein genes *1832-1831-1825-1824-1822-1820-1817-1816* in pETcoco-2	This study
pETcoco-2 *caethg 1831-25-24-22-20-17-16*	Link and lock of *Caethg* shell protein genes *1831-1825-1824-1822-1820-1817-1816* in pETcoco-2	This study
pETcoco-2 *caethg 1832-31-24-22-20-17-16*	Link and lock of *Caethg* shell protein genes *1832-1831-1824-1822-1820-1817-1816* in pETcoco-2	This study
pETcoco-2 *caethg 1832-31-25-22-20-17-16*	Link and lock of *Caethg* shell protein genes *1832-1831-1825-1822-1820-1817-1816* in pETcoco-2	This study
pETcoco-2 *caethg 1832-31-25-24-20-17-16*	Link and lock of *Caethg* shell protein genes *1832-1831-1825-1824-1820-1817-1816* in pETcoco-2	This study
pETcoco-2 *caethg 1832-31-25-24-22-17-16*	Link and lock of *Caethg* shell protein genes *1832-1831-1825-1824-1822-1817-1816* in pETcoco-2	This study
pETcoco-2 *caethg 1832-31-25-24-22-20-16*	Link and lock of *Caethg* shell protein genes *1832-1831-1825-1824-1822-1820-1816* in pETcoco-2	This study
pETcoco-2 *caethg 1832-31-25-24-22-20-17*	Link and lock of *Caethg* shell protein genes *1832-1831-1825-1824-1822-1820-1817* in pETcoco-2	This study
pET3a *pdu(K-86C)*	Hybrid gene of BMC domain of *C. freundii* shell protein gene *pduK* with	This study

	C-terminus of *caethg* shell protein gene *3286* cloned into pET3a	
pET3a *pdu A-B-J-(K-86C)-N-U*	Link and lock of *C. freundii* shell protein genes A, B, J, N, U with hybrid gene *pdu*(K-86C) in pET3a	This study
pET3a *pdu A-B-J-K-N-U*	Link and lock of *C. freundii* shell protein genes *A, B, J,K, N, U* in pET3a	Dr. E. Deery
pET3a *pdu A-B-J*	Link and lock of *C. freundii* shell protein genes *A, B, J* in pET3a	Dr. M. Liang
pET14b pdu K	*C. freundii* shell protein gene *pdu K* cloned into pET14b	Dr. M. Liang
pET3a pdu N	*C. freundii* shell protein gene *pdu N* cloned into pET3a	Dr. M. Liang
pLysS pdu U	*C. freundii* shell protein gene *pdu U* cloned into pLysS	Dr. M. Liang
pET3a *Rru_903*	*R. rubrum* signature enzyme gene *903* cloned into NdeI/SpeI sites of pET3a	Dr. D.Palmer
pET3a *Rru_905*	*R. rubrum* shell protein gene *905* cloned into NdeI/SpeI sites of pET3a	Dr. D.Palmer
pET3a *Rru_906*	*R. rubrum* shell protein gene *906* cloned into NdeI/SpeI sites of pET3a	Dr. D.Palmer
pET3a *Rru_907*	*R. rubrum* shell protein gene *907* cloned into NdeI/SpeI sites of pET3a	Dr. D.Palmer
pET3a *Rru_908*	*R .rubrum* shell protein gene *908* cloned into NdeI/SpeI sites of pET3a	Dr. D.Palmer
pET3a *Rru_912*	*R. rubrum* shell protein gene *912* cloned into NdeI/SpeI sites of pET3a	Dr. D.Palmer
pET3a *Rru_915*	*R. rubrum* shell protein gene *915* cloned into NdeI/SpeI sites of pET3a	Dr. D.Palmer
pET3a *Rru_905-6-7-8-12-15*	Link-and-lock of *R. rubrum* shell protein genes *905-906-907-908-912-915* in pET3a	Dr. D. Palmer
pET3a *Rru_905-7-12*	Link-and-lock of *R. rubrum* shell protein genes *905-907-912* in pET3a	This study
pET3a *Rru_908-7-12*	Link-and-lock of *R. rubrum* shell protein genes *908-907-912* in pET3a	This study

2.1.4 **Primers**

The primers used in this study are listed in Table 2.3. All primers were obtained from Eurofins Scientific. Nucleotide bases shown in **bold** letters highlight restriction enzyme sites and bases shown in red highlight the addition of a stop codon.

Table 2.3. Primers

Caethg: *Clostridium autothanogenum*

Primer name	Sequence	Restriction site
Caethg 3278 forw	gcc**catatg**gatcaagagataatc	**NdeI**
Caethg 3278 rev	cgtgcc**actagt**ttaaatataaggc	**SpeI**
Caethg 3283 forw	gcc**catatg**gaatttcgaactattaaatcg	**NdeI**
Caethg 3283 rev	cgggcc**actagt**ttaacatatatctttctc	**SpeI**
Caethg 3284 forw	gcc**catatg**ttaatagcaag	**NdeI**
Caethg 3284 rev	gcc**actagt**ctattctacattgc	**SpeI**
Caethg 3286 forw	gcc**attaat**atgcaggcacttgg	**AseI**
Caethg 3286 rev	gcc**actagt**ttaattatgccc	**SpeI**
Caethg 3289 forw	gcc**catatg**aaatctgatgcattagg	**NdeI**
Caethg 3289 rev	gcc**actagt**ttagtcttcctttaaacc	**SpeI**
Caethg 3290 forw	gctgcc**catatg**aaatatgatgcattagg	**NdeI**
Caethg 3290 rev	gcc**actagt**ctacttttcatcttcc	**SpeI**
Caethg 3286 *BMC*_80 rev	ggc**actagt**ttaatgtggacggggaataac	**SpeI**
Caethg 3286 *BMC*_90 rev	ggc**actagt**ttatgatacaatta	**SpeI**
Caethg 3286 *BMC*_100 rev	ggc**actagt**ttatttaatatttctatc	**SpeI**
Caethg 3286 *BMC*_110 rev	ggc**actagt**ttattttgattgattttc	**SpeI**
Caethg 3286 *BMC*_120 rev	ggc**actagt**ttactcttcattgccctc	**SpeI**
Caethg 3286 *BMC*_130 rev	ggc**actagt**ttattctatttttaaag	**SpeI**
Caethg 3286 *BMC*_140 rev	ggc**actagt**ttactcttctttagaatcg	**SpeI**

Caethg 3286 *BMC*_150 rev	ggc**actagt**ttaatctattgttggattttc	**SpeI**
Caethg 3286 *BMC*_170 rev	ggc**actagt**ttattctactgttgaattctc	**SpeI**
Caethg 3286 *BMC*_190 rev	ggc**actagt**ttacaaatctatttta	**SpeI**
Caethg 3286 *BMC*_230 rev	ggc**actagt**ttacttatatttacgagcg	**SpeI**
Caethg 3286 *BMC*_242 rev	ggc**actagt**ttatgctttagaaatggaccttcc	**SpeI**
Caethg 3286C forw	gcc**catatg**gcagtaccactgaaagatag	**NdeI**
Caethg 3286BMC_C28 BMC domain forw	gcc**attaat**atgcaggcacttgg	**AseI**
Caethg 3286BMC_C28 BMC domain rev	ccttcctttgataccaaagtttgatacaattatattatctaattc	**None**
Caethg 3286BMC_C28 C-term domain forw	gaattagataatataattgtatcaaactttggtatcaaaggaagg	**None**
Caethg 3286BMC_C28 C-term domain rev	gcc**actagt**ttaattatgccc	**SpeI**
Caethg 3286BMC_C63 BMC domain forw	gcc**attaat**atgcaggcacttgg	**AseI**
Caethg 3286BMC_C63 BMC domain rev	ccatcttatctacagcttttgatacaattatattatctaattc	**None**
Caethg 3286BMC_C63 C-term domain forw	gaattagataatataattgtatcaaaagctgtagataagatgg	**None**
Caethg 3286BMC_C63 C-term domain rev	gcc**actagt**ttaattatgccc	**SpeI**
Caethg 1816 forw	gcc**catatg**aaatatgatgcattagg	**NdeI**
Caethg 1816 rev	gcc**actagt**ctacttttcatcttcc	**SpeI**
Caethg 1817 forw	gcc**catatg**aaatctgatgcattagg	**NdeI**
Caethg 1817 rev	gcc**actagt**ttagtcttcctttaaacc	**SpeI**
Caethg 1820 forw	gcc**attaat**atgcaggcacttggg	**AseI**
Caethg 1820 rev	gcc**actagt**ctaattatgcccat	**SpeI**
Caethg 1822 forw	gcc**catatg**tcacaagcaatagg	**NdeI**
Caethg 1822 rev	gcc**actagt**ttagaacaatgatgcc	**SpeI**
Caethg 1824 forw	gcc**catatg**ttaatagcaag	**NdeI**

Caethg 1824 rev	gcc**actagt**ttattctacattgcag	**SpeI**
Caethg 1825 forw	gcc**catatg**gaatttcgaac	**NdeI**
Caethg 1825 rev	gcc**actagt**ttaacatatatctttctcc	**SpeI**
Caethg 1831 forw	gcc**catatg**agcaaatatgtagc	**NdeI**
Caethg 1831 rev	gcggcc**actagt**ttaaggaagtaaattttc	**SpeI**
Caethg 1832 forw	gcg**attaat**atgaggtattatggcg	**AseI**
Caethg 1832 rev	gcc**actagt**ttatgcatcgatgtc	**SpeI**
PduK-86C BMC domain forw	gcc**catatg**gaagcaatcactggattacttg	**NdeI**
PduK-86C BMC domain rev	ctatctttcagtggtactgcggtgacgctatgtgacagaatgcc	**None**
PduK-86C C-term domain forw	ggcattctgtcacatagcgtcaccgcagtaccactgaaagatag	**None**
PduK-86C C-term domain rev	gcc**actagt**ttaattatgccc	**SpeI**

2.1.5 **Media and solutions used for bacterial work**

Luria-Bertani (LB) broth:

Tryptone	10	g
Yeast Extract	5	g
NaCl	5	g

Made up to IL with distilled H_2O. Autoclaved.

Luria-Bertani (LB) broth + agar:

Tryptone	10	g
Yeast Extract	5	g
NaCl	5	g
Bacterial agar	15	g

Made up to 1L with distilled H_2O. Autoclaved.

Super optimal broth and catabolite repression (SOC):

Tryptone	2	g
Yeast Extract	0.5	g
1M NaCl	1	mL
2M Mg^{2+} stock (see below)	1	mL
20 % (w/v) glucose	1	mL
1M KCl	0.25	mL

Mg^{2+} stock:

$MgCl_2(H_2O)_6$	2	g
$MgSO_4(H_2O)_7$	2	g

Made up Mg^{2+} stock to 10 mL with distilled H_2O. Filter sterilised using 0.2µm syringe filter. Tryptone, yeast extract, NaCl and KCl dissolved in 97 mL dH_2O and autoclaved prior to the addition of the other components. All components that were not autoclaved were filter sterilised (0.2 µm pore size) prior to use.

Rhodospirillum rubrum growth media:

1 M Phosphate buffer, pH 7.0	20	mL
Concentrated base (see below)	20	mL
Growth factors (see below)	2	mL
Sodium succinate	2	g
$(NH_4)_2SO_4$	0.5	g
NaCl	0.5	g
Casein hydrolysate	1	g

Made up to 1L with distilled H_2O. Adjusted to pH 7.2 prior to autoclaving.

Concentrated base:

Nitrilotiacetic acid (diNa salt)	5.94	g
Metal 44 solution (see below)	25	mL
$MgSO_4(H_2O)_7$	14.5	g

$CaCl_2(H_2O)_2$	1.67	g
$FeSO_4(H_2O)_7$	50	mg
Ammonium molybdate$(H_2O)_4$	4.6	mg

Made up to 1L with distilled H_2O. diNa salt first dissolved and pH adjusted with KOH to pH5. Other reagents added and pH adjusted to 6.8. Autoclaved and stored at 4˚C.

Metal 44 solution:

EDTA	2.5	g
$ZnSO_4(H_2O)_7$	11	g
$FeSO_4(H_2O)_7$	5	g
3 M Sulphuric acid	1.5	mL
$MnSO_4(H_2O)_4$	3	g
$CuSO_4(H_2O)_5$	0.39	g
H_3Bo_3	0.12	g
$CoCl_2(H_2O)_6$	0.2	g

Made up to 1L with distilled H_2O. Solution filter sterilised and stored at 4˚C.

Growth Factors:

Biotin	20	mg
$NaHCO_3$	500	mg
Niacin	1	g
Thiamine HCl	500	mg

Made up to 1L with distilled H_2O. Autoclaved and stored at 4°C.

Solutions for preparing competent cells:

$CaCl_2$	0.1	M

$CaCl_2$	0.1	M
glycerol	15	%

Both filter sterilised through 0.2 µM filter.

Antibiotics:

Antibiotic	**Stock concentration**	**Final concentration**
Ampicillin	100 mg mL^{-1} in dH2O	100 µg mL^{-1}
Chloramphenicol	34 mg mL^{-1} in ethanol	34 µg mL^{-1}

Additives:

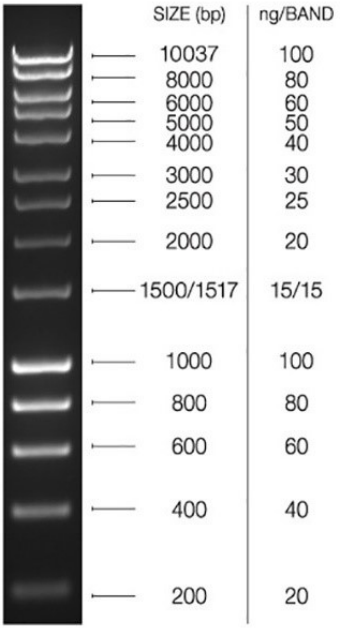
Additive	**Stock concentration**	**Final concentration**
IPTG	1M in dH_2O	100–400µM
L-arabinose	20 % (w/v) in dH_2O	0.01 % (w/v)
Sodium succinate	20 % (w/v) in dH_2O	0.1-0.2 % (w/v)
Glucose	20 % (w/v) in dH_2O	0.2 % (w/v)
1,2-propanediol	20 % (v/v) in dH_2O	0.1-4 % (v/v)

2.1.6 **Solutions for DNA work**

5x DNA loading buffer:

Bromophenol blue (w/v)	0.25	%
Glycerol (v/v)	50	%
TE buffer (v/v)	50	%

Bioline 1kb hyperladder



2.1.7 Solutions for protein work

2.1.7.1 Solutions for immobilised metal ion affinity chromatography (IMAC)

Charge buffer:

$NiSO_4(H_2O)_6$	50	mM

Binding buffer:

Tri-HCl, pH 8.0	20	mM
NaCl	500	mM
Imidazole	5	mM

Wash buffer I:

Tri-HCl, pH 8.0	20	mM
NaCl	500	mM
Imidazole	50	mM

Wash buffer II:

Tri-HCl, pH 8.0	20	mM
NaCl	500	mM
Imidazole	100	mM

Elution buffer:

Tri-HCl, pH 8.0	20	mM
NaCl	500	mM
Imidazole	400	mM

Strip buffer:

Tri-HCl, pH 8.0	20	mM
NaCl	500	mM
EDTA	100	mM

2.1.7.2 Solutions for Sodium Dodecyl Sulphate polyacrylamide gel electrophoresis (SDS-PAGE):

10x Running buffer:

Tri-HCl	30	gL^{-1}
Glycine	144	gL^{-1}

Made up to 2.5 L with distilled H_2O

1x Running buffer:

10x running buffer	100	mL
10 % (w/v) SDS	10	mL

Made up to 1 L with distilled H_2O.

2x Laemmli sample buffer:

0.5M Tris-HCl, pH 6.8	2.5	mL
Glycerol	2	mL
SDS 10 % (w/v)	4	mL
ß-mercaptoethanol	1	mL
Bromophenol Blue 0.08 % (w/v)	0.5	mL

Coomassie blue stain:

Trichloroacetic acid 100 %	250	mL
Coomassie blue R250	0.6	g
SDS 10 % (w/v)	0.1	g
Tris-HCl	0.25	g
Glycine	0.15	g

Made up to 500 mL with distilled H_2O

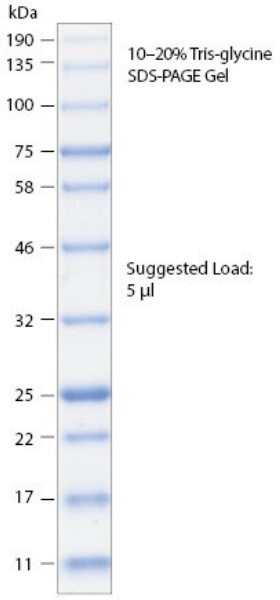
15 % SDS running gel:

dH_2O	2.2	mL
1.5M Tris-HCl, pH 8.8	3.5	mL
30% Acrylamide Acryl/Bis^{TM}29:1	7.8	mL
SDS 10 % (w/v)	1.5	mL
APS 10 % (w/v)	0.15	mL
TEMED	0.01	mL

5% SDS stacking gel:

dH_2O	3.4	mL
0.5M Tris-HCl, pH 6.8	1.9	mL
30% Acrylamide Acryl/Bis^{TM}29:1	1.5	mL
SDS 10 % (w/v)	0.75	mL
APS 10 % (w/v)	0.075	mL
TEMED	0.01	mL

NEB Broad Range prestained Protein Standard (P7706):



2.1.7.3 Solutions for 2D gel electrophoresis

2D Lysis buffer

Urea	30	g
CHAPS	1	g
Dithioreitol (DTT)	0.5	g
Pharmalyte (broad range pH 3 -10)	1	mL

Made up to 50 mL with ddH_2O

Agarose sealing solution

1 x Laemmli sample buffer	100	mL
Agarose	0.5	g
Bromophenol blue	1	grain

SDS equilibration buffer

Urea	180.25	g
1M Tris-HCl, pH8.8	25	mL
Glycerol	172.5	mL
SDS	10	g
Bromophenol blue	3	grains

Made up to 500 mL with ddH_2O

SDS equilibration buffer/DTT solution

SDS equilibration buffer	20	mL
DTT	200	mg

SDS equilibration buffer/ iodoacetamide (IAA) solution

SDS equilibration buffer	20	mL
IAA	500	mg

2.1.7.4 Solutions for size exclusion chromatography (FPLC):

Sample buffer:

Tri-HCl, pH 8.0	20	mM
NaCl	500	mM

Wash buffers:

Ethanol	20	%
NaOH	0.2	M

Solutions filtered prior to use

Molecular weight markers for size exclusion chromatography (Sigma):

Protein	**Approximate molecular weight (Da)**	**Concentration**
Carbonic anhydrase, bovine erythrocytes	29,000	3 mg mL^{-1}
Albumin, bovine serum	66,000	10 mg mL^{-1}
Alcohol dehydrogenase, yeast	15,0000	5 mg mL^{-1}
β-amylase, sweet potato	200,000	4 mg mL^{-1}
Apoferritin, horse spleen	443,000	10 mg mL^{-1}
Thyroglobulin, bovine	669,000	8 mg mL^{-1}
Blue dextran	2,000,000	2 mg mL^{-1}

2.1.7.5 Solutions for Western blotting

Transfer buffer

Tris	25	mM
Glycine	192	mM
Methanol	20	%

PBS

NaCl	140	mM
KCl	3	mM
Na_2HPO_4	10	nM
KH_2PO_4	2	mM

Blocking solution

Non-fat dried milk (in PBS)	5	% (w/v)

Primary antibody

Primary antibody diluted 1:1000 in PBS

Phosphate free buffer

Tris-HCl pH 7.5	50	mM
NaCl	150	mM

Phosphate free blocking solution

Non-fat dried milk	5	% (w/v)
Tris-HCl pH 7.5	50	mM
NaCl	150	mM

Secondary antibody

Alkaline phosphatase-linked secondary antibody diluted 1:5000 in phosphate free buffer

Chromogenic substrate

1 tablet of BCIP/NBT (Sigma) dissolved in 10 mL dH_2O (vortexed to dissolve)

2.1.7.6 Solutions for microcompartment purification

Cell lysis reagent:

n-Octyl-beta-D-thioglucopyranoside	1	g

Made up to 100 mL in 50 mM Tris-HCl, pH 8.

Buffer A:

Tris-HCl, pH 8	50	mM
KCl	500	mM
$MgCl_2$	12.5	mM
1,2 – propanediol	1.5	% (v/v)

Made up to 1 L with ddH_2O

Buffer B:

Tris-HCl ,pH 8	50	mM
KCl	50	mM
$MgCl_2$	5	mM

Made up to 1 L with ddH_2O

2.1.7.7 Solutions for MALDI-TOF in-gel digestion

Reduction and alkylation solutions

NH_4HCO_3	50	mM
Acetonitrile	50	% (v/v)

NH_4HCO_3	50	mM
DTT	10	mM

NH_4HCO_3	50	mM
Iodoacetamide	50	mM

Digestion buffer

NH_4HCO_3	10	mM
Acetonitrile	10	% (v/v)

Trypsin solution

Trypsin	20	µg
resuspension buffer	200	µL

Trypsin working solution

Trypsin solution	20	µL
digestion buffer	180	µL

Matrix solution:

Trifluoroacetic acid	0.1	% (v/v)
Acetonitrile	85	% (v/v)
H_2O	15	% (v/v)
NH_4HPO_4	1	mM
α-cyano-4 Hydroxycinnamic acid	0.7	$mgmL^{-1}$

2.1.7.8 Solutions for embedding samples prior to thin sectioning

Fixing solution:

Glutaraldehyde	2.5	% (v/v)
Sodium cacodylate, pH 7.2	100	mM

Cacodylate wash solution:

Sodium cacodylate, pH 7.2	100	mM

Osmium tetroxide solution:

Osmium tetroxide	1	% (v/v)
Sodium cacodylate, pH 7.2	100	mM

Low viscosity resin:

LV Resin	12	g
VH1	4	g
VH2	9	g
LV accelerator	0.63	g

2.2 Microbiological methods

2.2.1 Sterilisation of reagents

Unless stated otherwise media and buffers were sterilised for 15 min at 121 °C and 1 bar pressure in an autoclave. Temperature sensitive reagents were filter sterilised (0.2 µm pore size).

2.2.2 Liquid cultures

5 mL liquid cultures were inoculated with a single colony from an agar plate culture after transformation of competent cells with plasmid of interest (Section 2.2.4). The medium was supplemented with antibiotics as required and then cultures were shaken at ~160 rpm in 30 mL tubes overnight at 37 °C.

2.2.3 Preparation of *E. coli* competent cells

A bacterial overnight starter culture derived from a single colony was inoculated into 20 mL fresh LB broth and grown to an OD_{600} of 0.3. The cells were cooled on ice for 10 min and centrifuged at 1700 x g at 4 °C for 10 min. The pellets were gently re-suspended in 10 mL of ice-cold 0.1 M $CaCl_2$ and incubated on ice for 20 - 60 min. Cells were collected again by

centrifugation and re-suspended in 1.0 mL of ice-cold 0.1 M $CaCl_2$ containing 15 % (v/v) glycerol. Aliquots of 50 µL were frozen rapidly on dry ice and stored at -80 °C.

2.2.4 Transformation of *E. coli* competent cells

Competent cells were defrosted on ice for 10 min before adding 0.5 µL plasmid DNA or 5 µL of ligation mixture (Section 2.3.4). The mixture was incubated on ice for 15 min and then heat-shocked by incubation at 42 °C for 1 min before rapid transfer to ice and incubation for 2 min. After the addition of 200 µL SOC media, the cells were incubated at 37 °C for 20-60 min to allow antibiotic resistance expression. The mixture was then spread on a LB agar plate containing the required antibiotics and incubated at 37 °C overnight.

2.2.5 Recombinant protein overproduction in *E. coli:*

BL21 Star (DE3) or BL21 Star (DE3) pLysS *E. coli* competent cells were transformed with a vector containing the gene(s) of interest. A 5 or 10 mL starter culture was inoculated with a single colony and incubated overnight at 37 °C. 50 mL (for embedding), 300 mL for microcompartment isolation) or 1 L (for IMAC protein purification) of LB with ampicillin (and chloramphenicol for pLysS strain), inoculated from the starter culture, was grown at 37 °C and shaken at ~160 rpm until the culture reached an OD_{600} of approximately 0.6. Protein production was induced with 400 µM IPTG with overnight incubation at 19 °C. For IMAC, the cells were collected by centrifugation at 4000 rpm for 20 min at 4 °C. The pellet was re-suspended in 30 mL of binding buffer (Section 2.1.7.1) and either directly used for protein purification (Section 2.4.1) or stored at -80 °C. For embedding see section 2.4.9.2 and for microcompartment isolation see section 2.4.8.

2.2.6 Lysis of cells by sonication

Harvested cells were lysed by sonication using a Sonics Vibracell Ultrasonic processor in 30 sec bursts with 30 sec breaks repeated 6 times at 55 % maximum amplitude. The sonicated cells were centrifuged at 18,000 rpm for 20 min at 4 °C to remove cell debris. The soluble cell extract (supernatant) was then purified by IMAC (Section 2.4.1).

2.2.7 Anaerobic Culture of Rhodospirillum rubrum

2.2.7.1 De-gassing of media

Glass bottles were filled with 70 mL of media (Section 2.1.5), stoppered with a rubber bung and capped with a crimping cap. The media was bubbled with argon via a 21G needle and the displaced oxygen allowed to escape via a second 21G needle. The solution was de-gassed for 20 minutes before being autoclaved.

2.2.7.2 Inoculation of cultures

40mM Trimethylamine oxide (TMAO) was added to media, as a terminal electron acceptor, before inoculation with bacteria. A 1M stock, which had been previously de-gassed and autoclaved, was injected using a 21G needle and sterile syringe. Aerobically grown *R. rubrum* cultures were centrifuged to pellet the cells and re-suspended in sterile media. The re-suspended cells were injected using a 21G needle and sterile syringe to a starting OD_{600} ~ 0.1.

2.2.7.3 Anaerobic growth

Anaerobic cultures required a light source for effective growth. Culture bottles were fixed on a shaking platform with two high intensity lights at a fixed distance from the cultures (Fig. 2.1). This was to ensure that the cultures were at ~30°C throughout the growth period.

2.2.8 Growth Curves

2.2.8.1 Aerobic growth curves

Aerobic growth curves were obtained from 0.5 mL cultures in a 24 well culture plate inoculated at an OD_{600} ~0.05. OD_{600} measurements were taken using an OPTIMA microplate reader (BMG LabTech). The plate was incubated at 30°C and shaken at 200 rpm. OD_{600} readings were taken every 30 minutes for up to 60 hours.

2.2.8.2 Anaerobic growth curves

Anaerobic cultures were set up as described in Section 2.2.7.3. 1 mL samples were taken at regular intervals and OD_{600} measured using a Cary 60 UV-Vis spectrophotometer (Agilent Technologies).

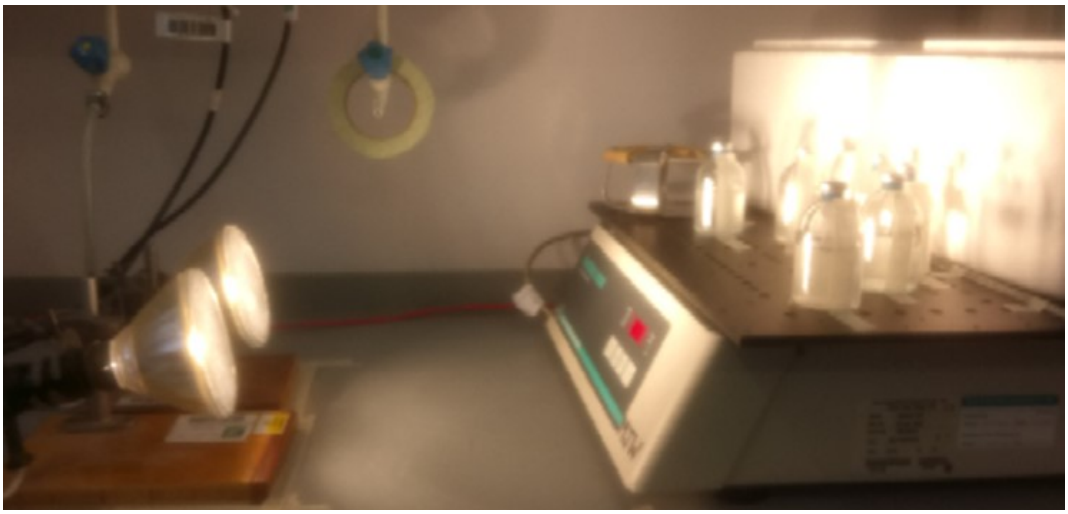


Figure 2.1. Set up of anaerobic cultures: Culture bottles were placed on a shaking platform at a fixed distance from two high intensity lights.

2.3 Molecular biological methods

2.3.1 PCR reactions

All PCR reactions were performed in an Eppendorf Mastercycler 5341 PCR machine using the FastStart High Fidelity PCR System (Roche). The basic PCR reaction and the cycles used are outlined below:

Table 2.4: Basic PCR reaction mixture

Component	µL
milliQ H_2O	32.5 – 34.5
10x PCR buffer (containing 18mM $MgCl_2$)	5
DMSO	0 – 2
5 mM dNTPs	2
10 µM 5' forward primer	2
10 µM 3' reverse primer	2
DNA template	1
Taq polymerase (5 $U\mu L^{-1}$)	0.5

Table 2.5: Temperature protocol for PCR reactions

Step	Temperature	Time	Cycles	Function
1	96 °C	2 mins	1	Initial denaturation of DNA
2	96 °C	30 sec	35	denaturing of amplified DNA
3	50-58 °C	30 sec		Annealing of primers
4	72 °C	1 min per 1000bp		Elongation
5	72 °C	7 min	1	Final elongation
Stop	4 °C	constant		

Once the PCR was complete, the product was separated by agarose gel electrophoresis (Section 2.3.2).

2.3.2 Agarose gel electrophoresis of DNA

DNA fragments were separated by DNA gel electrophoresis. 1 % (w/v) agarose gels were prepared in 1x TAE buffer with the addition of ethidium bromide to a final concentration of 0.5 µg mL^{-1} for visualisation of the DNA on a UV transilluminator. The DNA samples containing 20 % (v/v) DNA loading buffer (section 2.1.6) were loaded into the wells of the agarose gel and electrophoresis was carried out at 80 V for approximately 50 min using a

SubCell GT electrophoresis tank (BioRad) connected to a Power PAC 300 power supply (BioRad).

2.3.3 Gel extraction of a DNA fragment

The DNA band of interest was excised from an agarose gel using a scalpel blade. Purification was carried out using a QIAquick® Gel Extraction Kit (Qiagen) using the microcentrifuge protocol in the handbook.

2.3.4 Ligation of DNA

Vectors and inserts were digested with the relevant restriction enzymes as described in Section 2.3.6 and extracted from the gel (Section 2.3.3). The ligation of DNA fragments into the vector was carried out at room temperature for 2 hours or at 4 °C overnight using the following components:

insert	2.5	µL
vector	1.5	µL
10x T4 DNA ligase buffer (Promega)	1	µL
milliQ H_2O	4.5	µL
T4 DNA Ligase (3 U μL^{-1}, Promega)	0.5	µL

2.3.5 Isolation of plasmid DNA

A QIAprep® Miniprep Kit (Qiagen) was used for the purification of plasmid DNA as described in the handbook using the microcentrifuge protocol.

2.3.6 Restriction digests

Plasmid or PCR product DNA was digested using the relevant enzymes (10 U μL^{-1}) and the optimal buffer chosen according to either the Promega or New England Biolabs information provided. The reactions were incubated for 1-2 hours at the temperature required by the restriction enzyme before being subjected to electrophoresis (Section 2.3.2).

Table 2.6: Typical DNA digest protocol

Component	Single digest	Double digest
milliQ H_2O	3.5 µL	3 µL
Restriction enzyme 1	0.5 µL	0. 5µL
Restriction enzyme 2	--	0.5 µL
10x buffer	1 µL	1 µL
Plasmid/PCR DNA	5 µL	5µL

2.3.7 Cloning into vectors

2.3.7.1 Single gene cloning into vectors

The vectors (generally pET3a or pET14b) and the genes of interest were digested (Section 2.3.6) at the restriction sites introduced into the PCR product (Section 2.3.1) via the 5' and 3' primers. After extraction of the required DNA fragments (Section 2.3.3), vectors and genes were ligated together (Section 2.3.4). Competent *E. coli* (JM109 or DH10ß™) were transformed (Section 2.2.4) with the ligation mixture and plated onto appropriate media. Single colonies were selected, grown and the recombinant plasmids amplified, isolated (Section 2.3.5) and sequenced to verify the DNA sequence (Genewiz LLC).

2.3.7.2 Multiple gene cloning

Multiple gene constructs were formed by the consecutive cloning of genes into pET3a via the 'Link and Lock' cloning protocol (McGoldrick *et al.*, 2005). The 'Link and Lock' method fuses two previously digested DNA fragments which have compatible cohesive ends. The ligation site cannot be cut with either of the original restriction enzymes. Multiple genes can be sequentially added to a plasmid using this method. The 'Link and Lock' process is summarised in Figure 2.2. PCR was use to amplify genes and the genes were individually cloned into pET3a introducing a SpeI site at the 3' end of the genes via the 3' primers (Table 2.3). The ribosome binding site and genes to be subcloned into the 'Link and Lock' construct was excised using the restriction enzymes XbaI and either HindIII or ClaI (Fig. 2.2). The recipient plasmid containing the initial gene(s) was digested using SpeI and either HindIII or ClaI. Ligating the two fragments caused the fusion of SpeI and XbaI, as they have compatible cohesive ends, leading to a non-cleavable site, whereas the HindIII or ClaI sites fused as normal (Fig. 2.2). The removal of the SpeI site means the newly formed construct only contained one SpeI site so the process could be repeated again to add more genes.

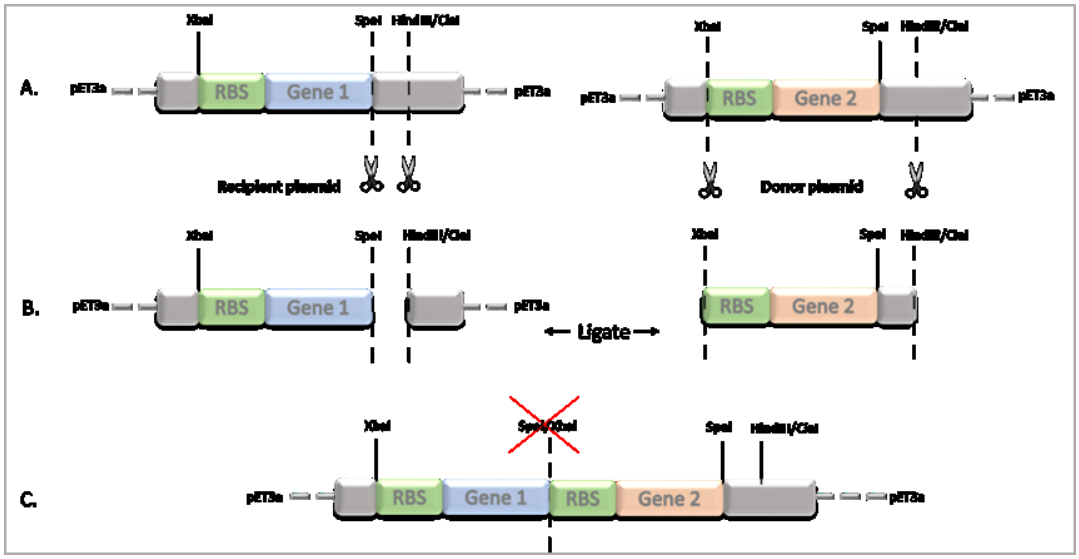


Figure 2.2. Schematic of the 'Link and Lock' cloning technique. A. The required genes were cloned into pET3a vectors and digested with appropriate restriction enzymes: recipient plasmid with SpeI and HindIII or ClaI; plasmid containing gene 2 (donor plasmid) with XbaI and HindIII or ClaI. B. The recipient plasmid and the donor fragment containing the gene and ribosome binding site (RBS) were ligated. C. The construct was formed with compatible cohesive ends (XbaI and SpeI) ligating and forming a non-cleavable site. The new construct was then capable of undergoing the process again to add a new gene of interest.

2.3.8 Construction of hybrid gene using extension overlap PCR

Hybrid genes containing the N-terminal BMC domain of one gene and the C-terminal domain of a second gene were constructed using extension overlap PCR. A summary of extension overlap PCR can be seen in figure 2.3. Initial PCR reactions amplified (i) the desired N-terminal region of gene 1 using a reverse primer containing the first ~20 bases of the C-terminus of the second gene at the 3' end and (ii) the desired C-terminal region of the second gene using a forward primer containing the final ~20 bases of the N-terminal region of the first gene at the 5' end. These gene products have overlapping regions and are used as templates in a second round of PCR with the desired hybrid gene as the product of this reaction.

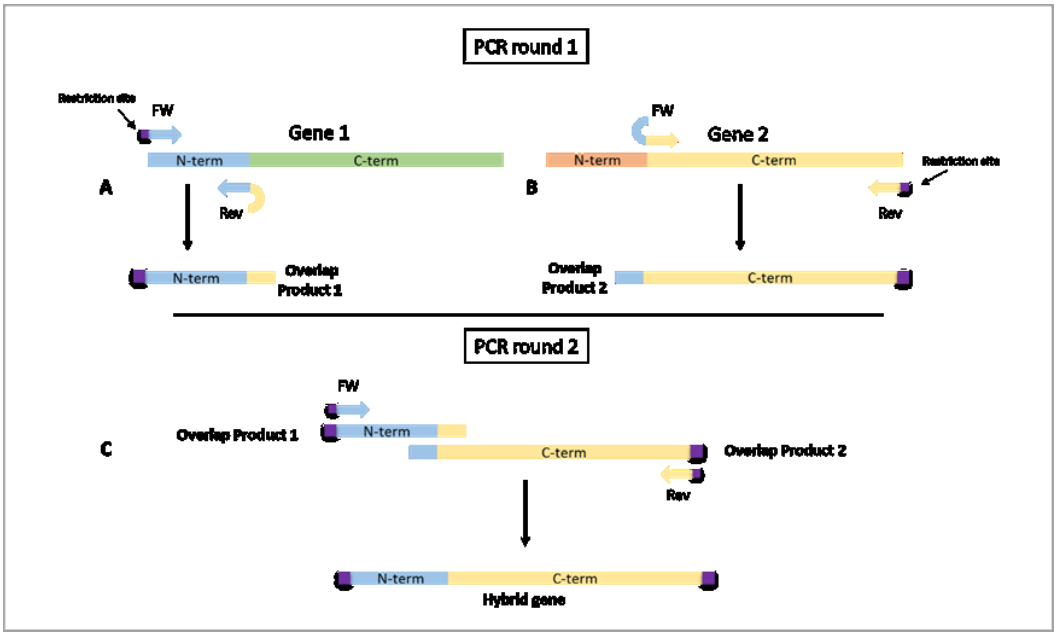


Figure 2.3. Schematic of cloning a hybrid gene by overlap extension PCR. A. In the first round of PCR, the N-terminal domain of Gene 1 is amplified using a forward primer containing a 5'restriction site for cloning into the pET3a vector and a reverse primer consisting of the reverse complement of the last ~20 bases of the N-terminal region of Gene1 and the first ~20 bases of the C-terminal region of Gene 2. B. The C-terminus of Gene 2 was amplified using a forward primer containing the last ~20 bases of the N-terminal domain of Gene 1 and the first ~20 bases of the C-terminus of Gene 2. The reverse primer contained a 3' restriction site for cloning into pET3a. C. In the second round of PCR, the two products from the first round of PCR overlap at the complementary end, acting as primers for each other and are extended by polymerase

in the first cycle of PCR. The extended PCR products are then used as templates in the subsequent rounds of PCR using the FW and Rev primers containing 5' and 3' restriction sites respectively.

2.4 Biochemical methods

2.4.1 Protein purification using immobilised metal ion affinity chromatography

Columns containing approximately 5 mL of Chelating Sepharose Fast Flow Resin (GE Healthcare) were first washed with dH_2O and then charged with 15 mL of $NiSO_4$ (charge buffer). Following equilibration with binding buffer, the soluble cell extract obtained after sonication and centrifugation (Section 2.2.6) was applied to the column and allowed to flow through by gravity. Binding buffer followed by wash buffers I and II were applied to the column to remove unbound protein. The His_6-tagged protein was recovered with elution buffer which was collected in 2 mL fractions. The column was re-generated with strip buffer and washed with dH_2O. Buffer composition can be found in section 2.1.7.1.

2.4.2 Buffer exchange

The eluted protein was desalted by buffer exchange into imidazole-free buffer using pre-packed, disposable PD-10 columns (GE Healthcare). The column, containing Sephadex G25 resin, was equilibrated with 25 mL of imidazole-free buffer followed by application of 2.5 mL of eluted protein. The protein was then eluted with 3.5 mL of imidazole-free buffer.

2.4.3 Size exclusion chromatography

Size exclusion chromatography was carried out using an AKTA PURE FPLC system (GE Healthcare). The protein sample was separated using a Superdex® 200 Increase 10/300 column. After equilibration with imidazole–free buffer, 0.5 mL of sample was loaded onto the column and the protein eluted at a flow rate of 0.75 $mLmin^{-1}$. Column calibration was performed with a size exclusion marker kit, MWGF-1000 (Sigma, see section 2.1.7.4 for kit contents).

2.4.4 Protein concentration determination – Bicinchoninic acid (BCA) assay

The BCA protein assay depends upon two reactions. Firstly, peptide bonds in proteins reduce Cu^{2+} ions to Cu^{+} in an alkaline environment. The amount of Cu^{2+} reduced is directly proportional to the concentration of protein. Two molecules of bicinchoninic acid then bind to each Cu^{+} forming a purple-coloured complex that absorbs light at 562 nm with absorbance increasing linearly with increasing protein. 50 µL of protein sample was incubated at 37°C for 30 minutes with 1 mL of BCA assay kit working solution (Fisher). The solutions were then left to return to room temperature for approximately 15 minutes before measuring absorbance

at 562 nm. A bovine serum albumin standard curve, 0.025 – 2 $mgmL^{-1}$, was run simultaneously.

2.4.5 Polyacrylamide gel electrophoresis

2.4.5.1 Sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE)

Proteins were separated according to their mass using SDS-PAGE. Gels were prepared with a 15% (v/v) running gel and a 5% (v/v) stacking gel. Samples were denatured by the 1:1 addition of Laemmli sample buffer (Section 2.1.7.2) and boiling for 10 minutes. 5 - 10 µL of denatured sample was loaded into each well and 7 µL of molecular mass marker (NEB P7706) run on each gel to estimate the relative molecular mass of the proteins of interest. Electrophoresis was run at a constant voltage of 200V using an Atto Dual Mini-Slab system and an Atto Mini Power electrophoresis power supply. Gels were stained using Coomassie blue stain (Section 2.1.7.2) and de-stained with dH_2O.

2.4.5.2 2D gel electrophoresis

2D gel electrophoresis was used to separate proteins in two dimensions. In the first dimension, proteins were separated by differences in their isoelectric point (isoelectric focusing) and in the second dimension according to their mass. Microcompartment protein samples were purified (Section 2.4.8) and then precipitated by raising the NaCl concentration to 80mM. The sample was re-suspended in 2D lysis buffer (Section 2.1.7.3) to a concentration of 100 μgmL^{-1}. A 125 µL sample was mixed with a small amount of bromophenol blue and vortexed to mix thoroughly. The sample was then applied to an isoelectric focusing (IEF) gel holder and an immobilised pH gradient (IPG) gel strip (pH 3-10; GE Healthcare) applied. After overlaying with silicone oil and placing on the lid, the gel holder was positioned on an IPGphor isoelectric focusing system platform (GE Healthcare) and run at the following settings.

20°C, 50µA/strip

Step 1	Step and hold	30 V	14.00 hrs
Step 2	Gradient	200 V	00.45 hrs
Step 3	Step and hold	500 V	00.45 hrs
Step 4	Step and hold	1000 V	00.45 hrs
Step 5	Gradient	8000 V	01.00 hrs
Step 6	Step and hold	8000 V	09.00 hrs

Following completion of the IEF steps the strip was incubated with equilibration/DTT solution (Section 2.1.7.3) followed by equilibration/IAA solution (Section 2.1.7.3) before separation of the proteins according to their mass by SDS-PAGE.

2.4.6 MALDI-TOF in gel digestion

Following 2D gel electrophoresis, the proteins present in isolated microcompartment samples were identified by Matrix assisted laser desorption/ionisation - time of flight mass spectrometry (MALDI-TOF).

2.4.6.1 Alkylation and reduction

Samples separated by 2D gel electrophoresis were excised from the SDS gel after staining and cut into 1mm pieces. The gels were dehydrated by incubation with acetonitrile for 15 minutes followed by incubation with DTT in NH_4HCO_3 (Section 2.1.7.6) to reduce disulphide bonds. A further dehydration with acetonitrile was followed by alkylation of the free sulfhydryl groups with iodoacetamide in NH_4HCO_3 (Section 2.1.7.6) for 20 minutes in the dark. Two washes in NH_4HCO_3 were followed by a final dehydration step with acetonitrile. The gels were then dried by vacuum centrifuge.

2.4.6.2 In-gel digest

The gel fragments were re-hydrated in trypsin working solution (Section 2.1.7.6) for 15 minutes at 4°C. Following centrifugation, the supernatant was removed, replaced with digestion buffer and incubated overnight at 37°C.

2.4.6.3 Extraction of peptides

5µL of acetonitrile was added and the sample sonicated for 15 minutes. The tubes were centrifuged and the supernatant collected. 50 % acetonitrile/5 % formic acid (v/v) was added to the gels and sonicated again for 15 minutes before centrifugation and collection of the supernatant which was pooled with the supernatant collected earlier. After addition of matrix solution (Section 2.1.7.6) the samples were analysed by MALDI-TOF on a Bruker Ultraflex system.

2.4.7 Western Blot analysis

Samples were prepared for SDS-PAGE as described in section 2.4.5.1. Samples were loaded onto 2 duplicate gels and the samples run at a constant voltage of 200V. One gel was stained with Coomassie blue stain as a reference gel and the second gel used for Western blotting.

The gel was washed in transfer buffer for 5 – 10 min. Nitrocellulose membrane was cut to size and equilibrated in methanol for 10 secs followed by washing in dH_2O for 5 min. The membrane was then equilibrated in transfer buffer along with fibre pads and filter paper. The holding cassette for the Trans-blot transfer tank was loaded by laying in the following order onto the anode side: pre-soaked fibre pad, filter paper, gel, membrane, filter paper, fibre pad. The holding cassette was closed and placed in the tank containing a magnetic stirrer and a

frozen Bio-Ice cooling block. The tank was filled to the top and a constant voltage applied of 100 V for 1 hour for the transfer of proteins from the gel to the membrane.

Binding sites were blocked by incubation of the membrane with 5 % (w/v) non-fat dried milk in PBS overnight at 4°C. The blocking solution was discarded and immediately replaced with a solution of the primary antibody directed against the protein under investigation (1:1000 dilution) and incubated at room temp for 2 hr with gentle agitation. The primary antibody solution was discarded and the membrane washed 3 times with PBS. The membrane was then equilibrated in phosphate free buffer for 10 min before adding the secondary antibody (an alkaline phosphatase conjugate) in phosphate free blocking solution (1:5000 dilution). The membrane was incubated with the secondary antibody for 1 hour and then washed 3 times in 150 mM Tri-HCl pH7.5. The membrane was then incubated with substrate (5-bromo-4-chloro-3-indolyl phosphate (BCIP) and nitro blue tetrazolium (NBT)) and left until colour developed. The membrane was washed in plenty of dH2O to completely remove substrate before drying and storing away from light.

2.4.8 Microcompartment isolation

Microcompartment isolation was performed using the method described by Sinha but replacing BPER II with n-Octyl-beta-D-thioglucopyranoside (Sinha *et al.*, 2012). 300 mL cultures of cells were grown and induced as described in section 2.2.5. The cells were centrifuged at 2700 x g and the pellet washed twice in buffer A (section 2.1.7.5). After centrifuging again at 2700 x g, the cells were lysed by re-suspending in 10 mL buffer A and 15 mL 1 % (w/v) n-Octyl-beta-D-thioglucopyranoside supplemented with 25 mg lysozyme and 5µL benzonase. The suspension was shaken gently at room temperature for 30 minutes then placed on ice for 5 minutes. Cell debris was removed by centrifugation at 12,000 x g for 5 minutes at 4°C. The supernatant was collected and the centrifugation at 12,000 x g repeated. The supernatant was then centrifuged at 20,000 x g for 20 minutes to pellet the microcompartments. The pellet was re-suspended in 0.5 mL buffer B (section 2.1.7.5). Remaining cell debris was removed by centrifuging 3 times at 12,000 x g for 1 minute, the microcompartments remaining in the supernatant with each centrifugation.

2.4.9 Transmission Electron Microscopy

2.4.9.1 Fixing and staining isolated compartments

Formvar-coated copper mesh grids were placed on a parafilm coated wax sheet and secured in place by gently pressing on the side of the grids. 15 µL of the isolated microcompartment sample was placed on each grid and incubated for 2 min. 15 µL of fixative (2.5 % (v/v) glutaraldehyde in 100mM sodium cacodylate buffer, pH 7.2) was added to each grid and incubated for a further 2 min. The grids were then transferred into a fresh 15 µL drops of

fixative. The grids were transferred sequentially through 3 drops of fix followed by 3 drops of ddH_2O. The grid was dried with filter paper and 1 drop of stain (2 % (w/v) uranyl acetate in H_2O) added before immediately suctioning off. The grids were then dried before imaging within 48 hrs.

2.4.9.2 Sample preparation: embedding whole cells

20 mL cultures of BL21 Star (DE3) pLysS cells, transformed with a plasmid harbouring single or multiple genes encoding the proteins of interest, were grown to OD_{600} ~0.6 at 37°C, induced with 100 – 400 µM IPTG and incubated shaking overnight at 19°C. Cells were collected by centrifugation at 1700 xg for 10 minutes and fixed, gently mixing, for 2 hours in 2.5% glutaraldehyde/100mM sodium cacodylate pH 7.2. Cells were centrifuged for 2 minutes at 10,000 rpm before re-suspension in each of the following wash steps. Cells were washed twice in 100mM sodium cacodylate buffer to remove glutaraldehyde before gentle mixing for 1 hour with 1 % (v/v) osmium tetroxide/100mM sodium cacodylate. Osmium tetroxide cross-links phospholipids in the sample whilst also acting as a stain as it is electron opaque. The cells were pelleted and washed twice in ddH_2O. The samples were dehydrated by pelleting and washing for 10 minutes each in a series of increasing ethanol concentrations, finishing with three washes in 100 % ethanol. Samples were then washed twice in propylene oxide to remove any residual ethanol before mixing for 30 minutes in 50 % propylene oxide/50 % LVR resin. Samples were re-suspended in LVR resin and mixed gently for 1.5 hours before centrifuging at 13,000rpm for 10 minutes. Samples were again re-suspended in fresh LVR resin and mixed gently for a further 1.5 hours. After centrifugation at 13,000rpm for 10 minutes, the cell pellets were re-suspended in 0.5 mL LVR resin and transferred to beem capsules. The capsules were centrifuged at 1100 rpm for 5 minutes before polymerising at 60°C for approximately 18 hours.

2.4.9.3 Sample preparation: sectioning

After polymerisation, block faces were first polished with a glass knife before trimming to a trapezoid shape measuring approximately 0.5 x 0.5 mm. 70 nm sections were then cut using a Diatome Ultra 45° diamond knife on a Leica EM UC7 ultramicrotome. Sections were collected on copper 400 mesh grids and allowed to dry before staining.

2.4.9.4 Sample preparation: staining

Grids holding the thin sections were placed in drops of 4.5 % (w/v) uranyl acetate in 1% (v/v) acetic acid and incubated at room temperature for 45 minutes. Each grid was removed to a drop of ddH_2O before being washed in a stream of ddH_2O for approximately 20 seconds and then placed into a second drop of ddH_2O. Grids were gently blotted dry before being placed into a drop of Reynolds lead citrate and incubated at room temperature for 7 minutes. The

grids were then washed by placing into drops of ddH_2O and washing in a stream of ddH_2O as before. Grids were allowed to dry before storage in grid boxes.

2.4.9.5 Imaging

Samples were imaged using a JEOL 1230 transmission electron microscope and a Gatan OneView camera.

2.4.10 Protein crystallisation and X-ray crystallography

2.4.10.1 Protein crystallisation

The hanging drop vapour diffusion method was used for protein crystallisation. Proteins were purified by IMAC (section 2.4.1) followed by size exclusion chromatography (section 2.4.3). Protein was concentrated to concentrations of 5 and 10 $mgmL^{-1}$ using an Amicon Ultra-4 centrifugal filter (Millipore) with a molecular weight cut-off of 10kDa.

Crystallisation screens were set up using Structure Screen 1 and Structure Screen 2 from Molecular Dimensions. 1 mL of reservoir solution was placed in each well of a 24 well XRL plate (Molecular Dimensions). 1 µL of protein was mixed with 1 µL of reservoir solution on a siliconized cover slip (Hamilton Research). Two drops of protein/reservoir solution were mixed on each cover slip to give two final concentrations of 2.5 and 5 $mgmL^{-1}$. The coverslip was inverted over the well of the plate and sealed with a thin layer of high vacuum grease (Dow Corning). The plates were then held at a constant temperature of 19°C and observed regularly for crystal formation.

Any crystals formed were removed from the drop and transferred to a drop of cryo-protectant, 10 % (v/v) glycerol in reservoir solution using a CryoLoop™ (Hamilton Research) and then flash frozen in liquid nitrogen. Crystals were stored in liquid nitrogen until data collection.

2.4.10.2 X-ray diffraction

Data was collected at the Diamond Light Source, Oxfordshire and structure resolved using molecular replacement.

Chapter 3

A study of a *Rhodospirillum rubrum* GRE BMC.

3.1 Introduction

Bacterial microcompartments harbouring biochemical pathways for carbon–fixation, B_{12}-dependent 1,2-propanediol and ethanolamine utilisation have been extensively investigated. However, bioinformatics analyses have identified a diverse array of BMCs that have not yet been experimentally characterised (Axen *et al.*, 2014). BMC loci containing a glycyl radical enzyme (GRE) and its associated activation enzyme (AE) as the signature enzyme represent the largest class of putative BMCs identified, yet very little experimental data about them is available. GRE associated microcompartments (GRMs) have been classified GRM1 – GRM5 based on an analysis of their core enzymes and shell protein properties (Zarzycki *et al.,* 2015). The genome of *Rhodospirillum rubrum* harbours one such GRE associated microcompartment.

R. rubrum is a Gram-negative alphaproteobacterium and is found primarily in brackish water, sewage and mud. Within its genome there is an 18-gene cluster containing 6 genes that encode for shell proteins, genes encoding a glycyl radical diol dehydratase and its associated activating enzyme plus other enzyme pathway and auxiliary proteins.

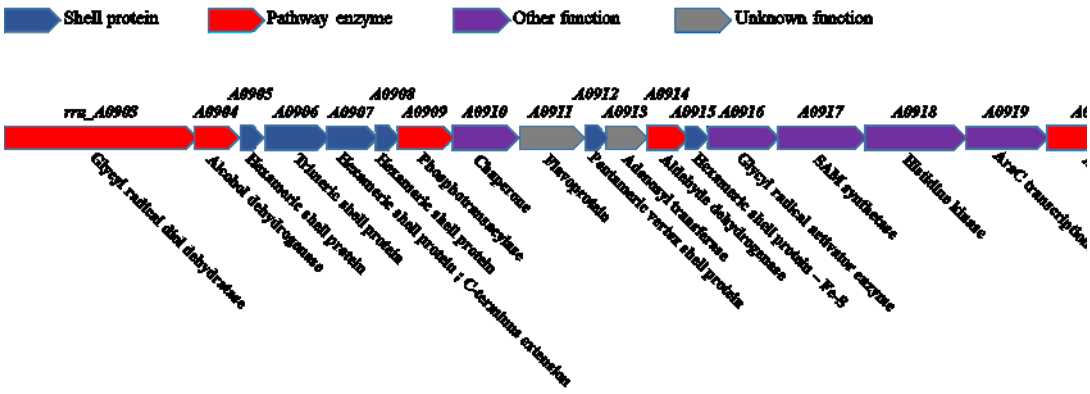


Figure 3.1. Schematic of the *R. rubrum* BMC gene cluster. Shell proteins are indicated by blue arrows, pathway enzymes by red arrows, other functional proteins by purple arrows and proteins of unknown function by grey arrows.

Bioinformatic studies (Axen *et al.*, 2014; Zarzycki *et al.*, 2015) placed the *R. rubrum* BMC in the GRM3 class of GRE associated microcompartments, predicting the substrate to be 1,2-PD. A subsequent study (Zarzycki *et al.*, 2017) confirmed the signature enzyme of the *Rhodopseudomonas palustris* BisB18 BMC to be a selective B_{12} independent 1,2-propanediol dehydydratase. The signature enzyme of the *R. rubrum* BMC shares 87 % identity with the *R. palustris* enzyme strongly suggesting that 1,2-PD is also the substrate of the *R. rubrum* BMC pathway.

This study was designed with two aims. Firstly, to investigate the structures formed when the shell proteins of the *R. rubrum* BMC are recombinantly produced in *E. coli*. Studies with non-GRM shell proteins have demonstrated the ability for empty BMC formation (Joshua B Parsons *et al.*, 2010), therefore this study aimed to investigate whether this was also possible with the *R. rubrum* shell proteins and to observe any difference or similarities in the structures formed. Secondly, the observation of GRE-associated BMCs in native species has been limited so far to the fucose/rhamnose metabolising BMC of *Clostridium phytofermentans* (Petit *et al.*, 2013) and the choline utilising microcompartment of *Proteus mirabilis* (Jameson *et al.*, 2016). This study therefore aimed to see if BMC production could be induced in *R. rubrum* under specific growth conditions.

3.2 Results: Recombinant production of *R. rubrum* shell proteins in *E. coli*

3.2.1 Protein sequence alignments

The BMC operon of *R. rubrum* contains 6 genes that encode for shell proteins representing two hexameric shell proteins (BMC-H; single BMC domain - pfam00936), one trimeric shell protein (BMC-T; two BMC domains), one pentameric vertex protein (BMC-P; single pfam03319 domain), one hexameric shell protein with a C-terminal extension of unknown function (BMC-H_{ex}) and one hexameric shell protein with Fe-S centre within the central pore (BMC-H_{FeS}). The predicted substrate for the enzyme pathway encapsulated within the *R. rubrum* BMC is 1,2-propanediol, therefore the primary sequences of the *R. rubrum* shell proteins were compared with those of the *Citrobacter freundii* 1,2-propanediol utilising (Pdu) BMC shell proteins (Table 3.1).

Table 3.1 Sequence identity between *R. rubrum* shell proteins and the shell proteins of the *C. freundii* Pdu BMC. Percentage identity between *R. rubrum* and *C. freundii* Pdu shell proteins. Pairwise global alignments performed in Emboss Needle (Madeira *et al.*, 2019). Percentage sequence similarity is shown in parentheses. Similarity greater than 50 % is highlighted in yellow.

	Rru_A0905 (BMC-H)	Rru_A0906 (BMC-T)	Rru_A0907 (BMC-H_{ex})	Rru_A0908 (BMC-H)	Rru_A0912 (BMC-P)	Rru_A0915 (BMC-H_{Fe})
PduA	**79** (87) %	**11** (14)%	**37** (46)%	**73** (81) %	**15** (25) %	**16** (25) %
PduB	**8** (15) %	**59** (73) %	**6** (10) %	**8** (14) %	**8** (15) %	**8** (11) %
PduJ	**72** (81) %	**10** (14) %	**22** (27) %	**72** (82) %	**18** (33) %	**16** (26) %
PduK	**22** (34) %	**7** (12) %	**24** (35) %	**22** (34) %	**14** (24) %	**6** (17) %
PduN	**16** (37) %	**8** (12) %	**10** (17) %	**19** (38) %	**46** (61) %	**19** (30) %
PduU	**15** (25) %	**14** (22) %	**8** (13) %	**14** (23) %	**15** (27) %	**18** (33) %
PduT	**18** (28) %	**16** (27) %	**11** (16) %	**17** (26) %	**10** (18) %	**14** (20) %

Comparisons show that there are PduA-, PduB-, PduJ- and PduN-like proteins present in the *R. rubrum* BMC operon. This may reflect need for the transport and encapsulation of similar substrates and/or products to the Pdu microcompartment. Both PduK and Rru_907 are C-terminally extended hexameric shell proteins yet the C-terminal extensions do not appear similar perhaps reflecting a different function within the two systems. The role of the C-terminal extension of PduK in the Pdu BMC is still unclear. Some studies have implicated a role in encapsulation of enzyme pathway components (Unpublished data – Dr. M. Liang) which may explain the differences in the primary sequences as the signature enzymes of the two systems are different. PduU is a circularly permuted BMC-H protein with a distinctive ß-

barrel in the central pore (Crowley *et al.*, 2008). PduU has no homologues within the *R. rubrum* BMC operon, suggesting it plays a role not required by the *R. rubrum* BMC. Similarly, PduT, which is a trimeric shell protein with an iron-sulphur binding site within the central pore, shares no sequence similarity with the *R. rubrum* shell proteins.

3.2.2 Production of single shell proteins in *E. coli*

Overproduction of BMC shell proteins in *E. coli* can help identify their role through their ability to self-assemble into various structures, for example filaments or sheets. To investigate the self-assembly characteristics of the *R. rubrum* shell proteins, competent *E. coli* BL21*(DE3) pLysS cells were transformed separately with pET3a vectors harbouring one of the 6 shell protein genes encoded in the *R. rubrum* BMC operon (vectors provided by Dr. David Palmer). The resulting strains were grown at 37 °C in LB media containing ampicillin and chloramphenicol to an $OD_{600} \approx 0.6$. Protein production was induced with 400 μM isopropyl β-D-1-thiogalactopyranoside (IPTG) and cultures grown overnight at 19 °C before harvesting for fixing and thin sectioning.

3.2.2.1 Rru_A0905

The first shell protein encoded in the BMC operon is Rru_A0905, a hexameric, 98 amino acid, 9.9 kDa shell protein with high sequence similarity to PduA (87 %) and PduJ (81 %) (Table 3.1).

Over-production of Rru_A0905 in *E. coli* resulted in the formation of protein tubes throughout the cytoplasm. TEM images of longitudinal sections show a series of filaments extending through the length of the cell (Fig 3.2). This is similar to the structures reported for PduA and PduA* expressed in *E. coli* (Pang *et al.*, 2014; Lee *et al.*, 2018).

TEM images of transverse sections showed a collection of circular structures as previously observed with PduA expression in *E. coli*, although these structures did not appear to be packed in a regular honeycomb lattice as described for PduA (Parsons *et al.*, 2010; Pang *et al.*, 2014). In addition, there were a number of larger structures that resemble concentric circles or rolled sheets of protein (Fig. 3.3). These ‘tubes-within-tubes’ have not been previously observed with the overproduction of PduA in *E. coli.*

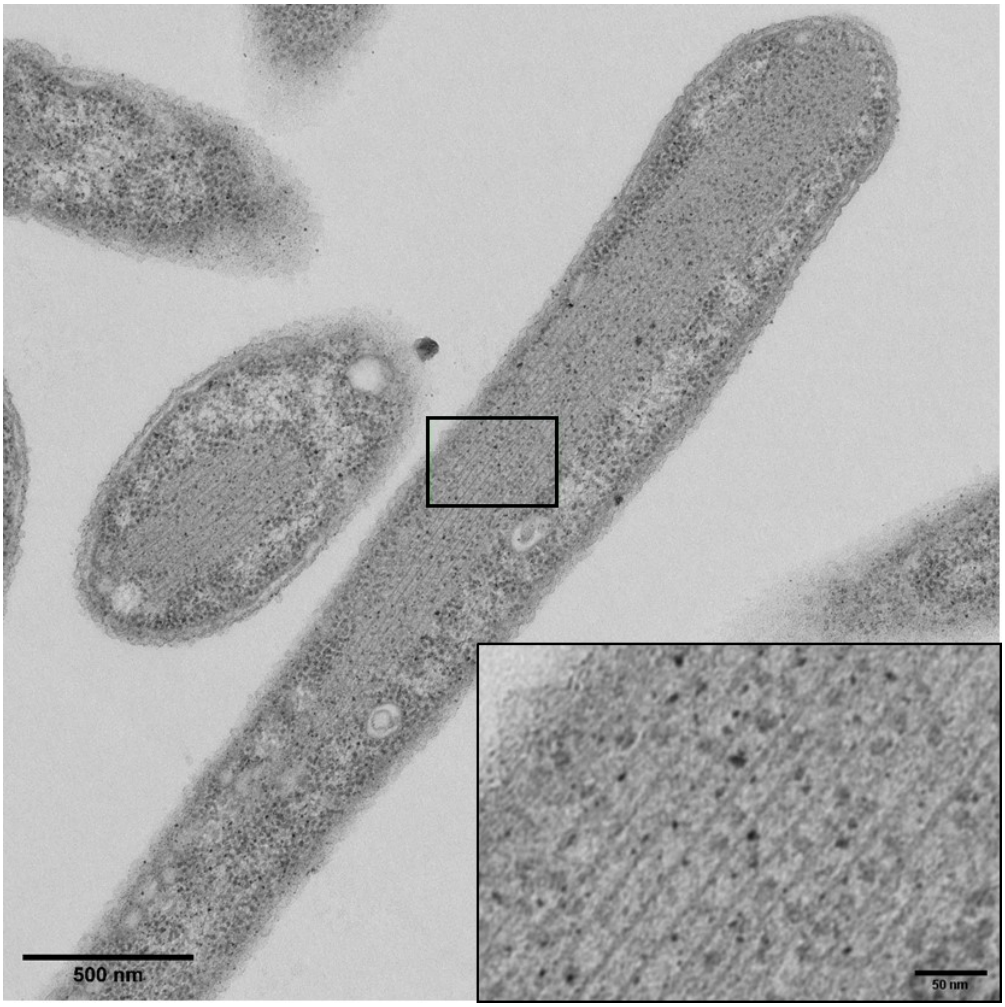


Figure 3.2. TEM image of longitudinal thin section of E. coli over-producing the hexameric shell protein Rru_A0905. Inset shows enlargement of the area outlined in the main image. Scale bar main image 500 nm, inset 50nm.

The diameter of the single layered tubes observed in transverse sections (measured using ImageJ (Schneider *et al.*, 2012)) were 27.8 ± 2.4 nm (n = 76 measurements). The diameter measured from longitudinal sections was 24.8 ± 2.1 nm (n = 72 measurements). The mean diameter measured in longitudinal sections was slightly smaller and the data had a wider distribution than the transverse section diameter measurements (Fig. 3.4). Longitudinal sections may not be sectioned through the widest part of the tube structures resulting in measurements that are more variable.

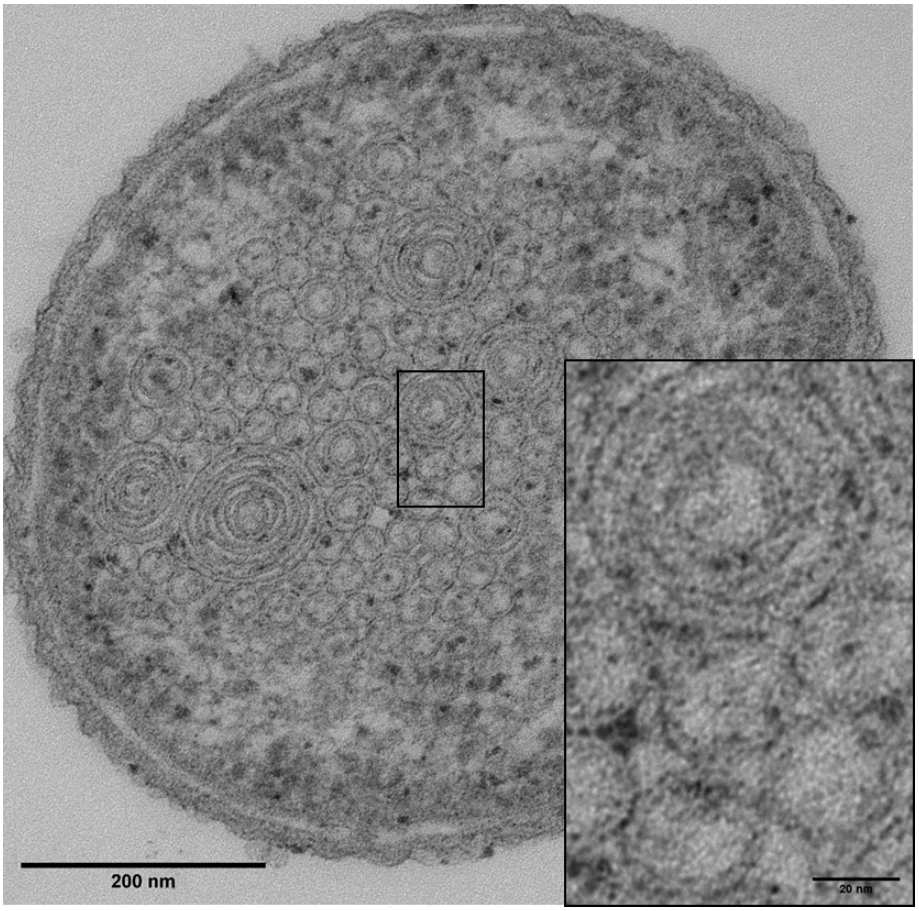


Figure 3.3. TEM image of a transverse section through an *E. coli* cell expressing Rru_A0905. Single circular tube-like structures (black arrows) and multiple concentric tubes or rolled sheets (white arrows) are indicated. Inset shows enlargement of the area outlined in the main image. Scale bar main image 200 nm, inset 20nm.

The protein sheets forming the tubes have a mean thickness of 3.07 ± 0.54 nm. This correlates well with the crystal structure measurement of the thickness of a shell protein hexamer of 20 – 30 Å (Kerfeld, 2005) suggesting that the tube structures are constructed from a single layer of Rru_A0905 hexamers.

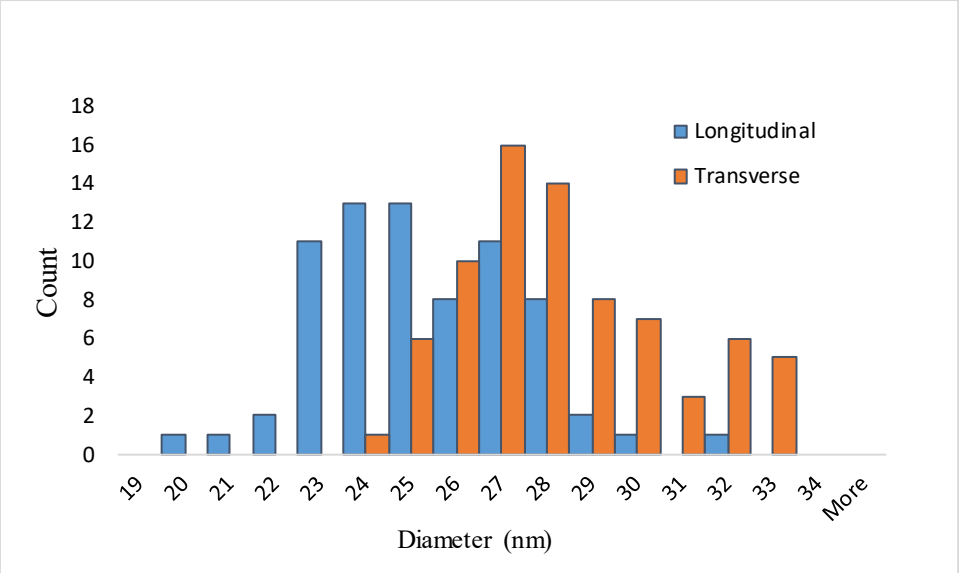


Figure 3.4. Histogram showing the diameter of Rru_A0905 tubes measured in *E. coli*. Longitudinal sections n = 72; transverse sections n = 76

3.2.2.2 Rru_A0906

The second shell protein encoded in the *R. rubrum* BMC operon is Rru_A0906, a trimeric 28.5 kDa shell protein, 280 amino acids in length and containing 2 tandem BMC domains. Its closest homologue of the Pdu shell proteins is PduB, also a trimeric shell protein, with 73 % sequence similarity (Table 3.1).

The *E. coli* BL21*(DE3) pLysS strain producing Rru_A0906 was prepared as described for Rru_A0905 in Section 3.2.2. TEM images showed Rru_A0906 to form insoluble aggregates of protein with large inclusion bodies being observed in most cell sections (Fig. 3.5).

Inclusion bodies were not observed after the overproduction of other tandem BMC domain proteins in *E. coli*. Overproduction of PduB produced long filaments that wound around close to plasma membrane of the cell, whereas cells producing EtuB appeared normal with no additional internal protein structures (Heldt *et al.*, 2009). *In vitro*, PduB has also been shown to form filaments upon lowering of the salt concentration (Uddin *et al.*, 2018).

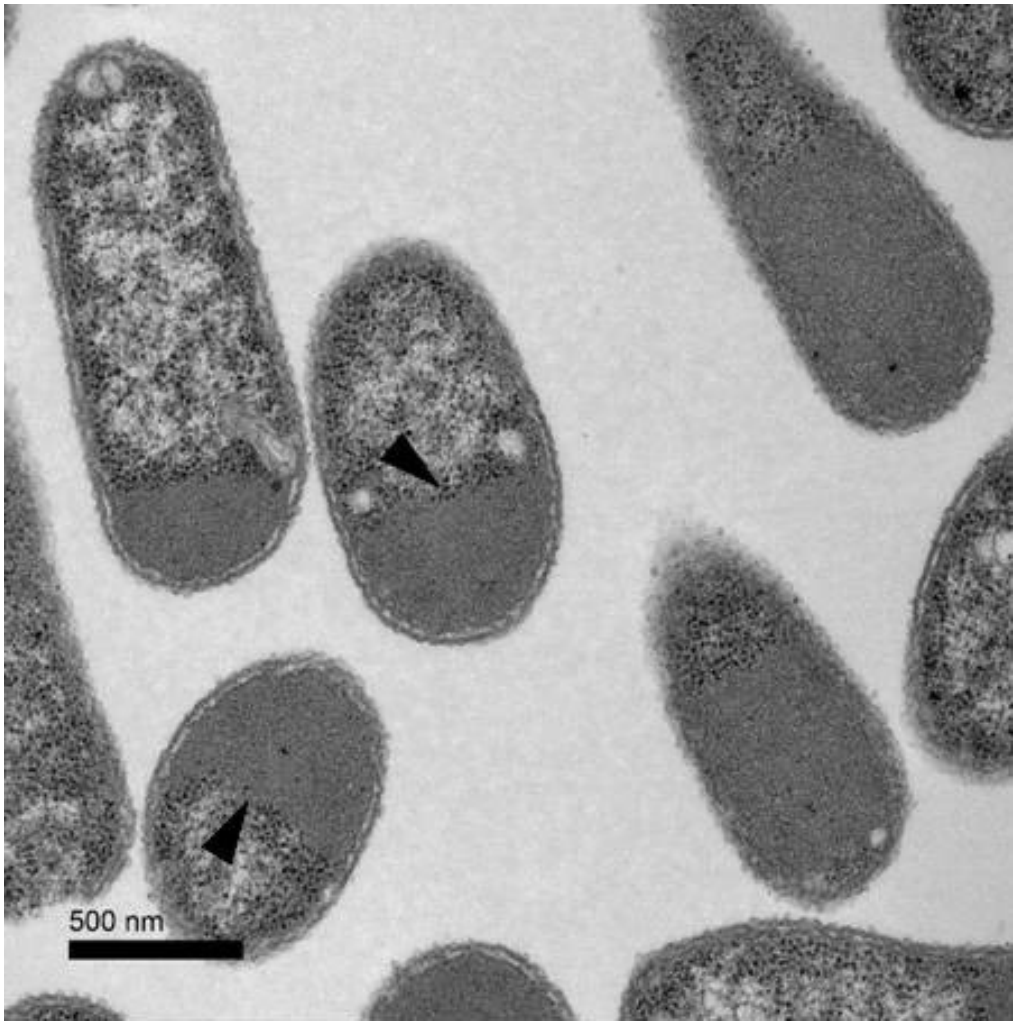


Figure 3.5. TEM image of thin section of *E. coli* overproducing Rru_A0906. Inclusion bodies are indicated (black arrows). Scale bar 500 nm.

3.2.2.3 Rru_A0907

Rru_A0907 is a hexameric 21.8 kDa shell protein, composed of 220 amino acids, with a C-terminal extension of unknown function after the N-terminal BMC domain. PduK is its Pdu homologue although it only shares 35 % similarity (Table 3.1; Fig. 3.6). The C-terminal extension (residues 90 onwards) shares little sequence similarity with the PduK C-terminal extension.

```
           1       10        20        30        40        50        60        70        80        90       100
           |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
RruA0907   MAQLSLGLIETIGLAAAIEAADAAVKSANVALVGYEFARGDGMTVVKVQGDVGAVNAAIAAAAVAAARVGRVVSTRVIARPAAGIDGLIVNRDTVGVPPR
PduK        VKQSLGLLEVSGLALAISCADVMAKAASITLVGLEKTNGSGWMVIKIIGDVASVQAAISTGVSFADQRDGLVAHKVISRPGDGILSHSVT---------
Consensus  ....SLGL.E..GLA.AI..AD...K.A...LVG.E...G.G..V.K..GDV..V.AAI......A......V...VI.RP..GI....V..........

           101     110       120       130       140       150       160       170       180       190       200
           |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
RruA0907   VVSAPAAEPTGDASPSAATSSPAPIDTPLPQPVPTPQAPSAAATVDEVPTPPVMDETPAPPVVNETPAQPAVDETPVKPAVDETQTHGAERTAPSPSRQG
PduK       ----------------------PESESEPAPAPTPVVPHEEIPEDHAAPEAPQDAELISCNLCLDPACPRQKGEPRSLCLHSGKRGEA
Consensus  ......................P.....P.P.PTP..P......D.........D...........PA.P.....P.............A..........

           201     210       220
           |--------+---------|
RruA0907   KTGGRPTAGNNGKPRNRHDA
PduK
Consensus  ....................
```

Figure 3.6. Protein sequence alignment of Rru_A0907 with PduK. Residues coloured red show conserved residues. Sequence alignment was performed using Multalin online sequence alignment tool (Corpet, 1988)

TEM images of *E. coli* overproducing Rru_A0907 show normal cells with no additional intracellular structures visible (Fig. 3.7). There are no images available reporting any structures formed by the recombinant expression of C-terminally extended BMC shell proteins.

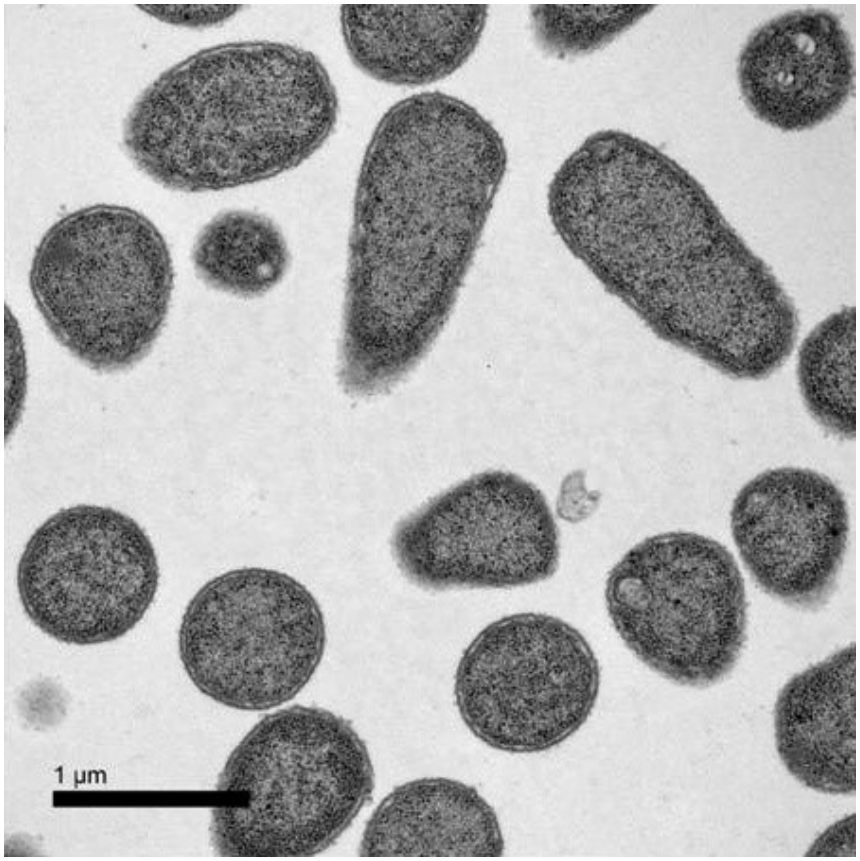


Figure 3.7. TEM image of *E. coli* overproducing Rru_A0907. Scale bar 1 µm.

3.2.2.4 Rru_A0908

Rru_A0908 is a hexameric 9.2 kDa shell protein that is composed of 92 amino acids that shares high sequence similarity to both PduA and PduJ (81 % and 82 % similarity respectively) (Table 3.1). Rru_A0908 also has 81% similarity to Rru_A0905.

When *rru_A0908* was expressed in *E. coli*, filamentous structures were observed in electron micrographs of longitudinal sections (Fig. 3.8). This would perhaps be expected due to the high sequence similarity between Rru_A0908 and PduA and Rru_A0905, which both form tubes that look like filaments in longitudinal sections when recombinantly produced in *E. coli*. However, unlike Rru_A0905 and PduA, only a small number of tubes are observed. Instead, images of transverse sections show the longitudinal filaments are formed from rolled sheets of protein (Fig 3.9). Increasing the magnification show the rolled sheets appear to form layers or stacks of protein sheets (Fig 3.9 inset).

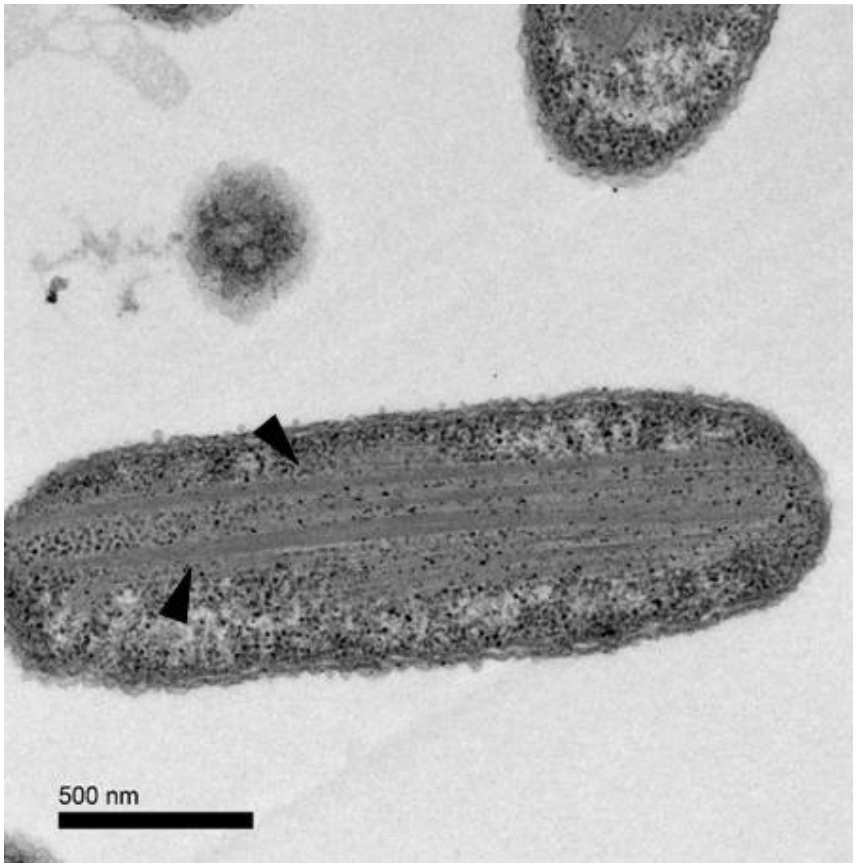


Figure 3.8. TEM image of a longitudinal section through *E. coli* cells producing Rru_A0908. Filamentous protein structures are indicated (black arrows). Scale bar 500 nm

The mean distance between the layers of sheets was 3.09 ± 0.38 nm (n = 16 measurements) which is similar to the measured depth of a hexamer (20 – 30 Å) (Kerfeld, 2005).

There are clear differences in the structures formed by Rru_A0908 and Rru_A0905 despite their sequence similarity. The rolled sheets formed by Rru_A0905 (the concentric circles in Fig. 3.3), appear to be formed from a single layer of hexameric tiles. In contrast, Rru_A0908 appears to form multi-layered sheets. Some studies have proposed that some shell proteins may exist as double layered hexameric tiles (Klein *et al.*, 2009; Samborska and Kimber, 2012; Cai *et al.*, 2013). Rru_A0908 may also form double-layered tiles resulting in the closely packed sheets observed in this study.

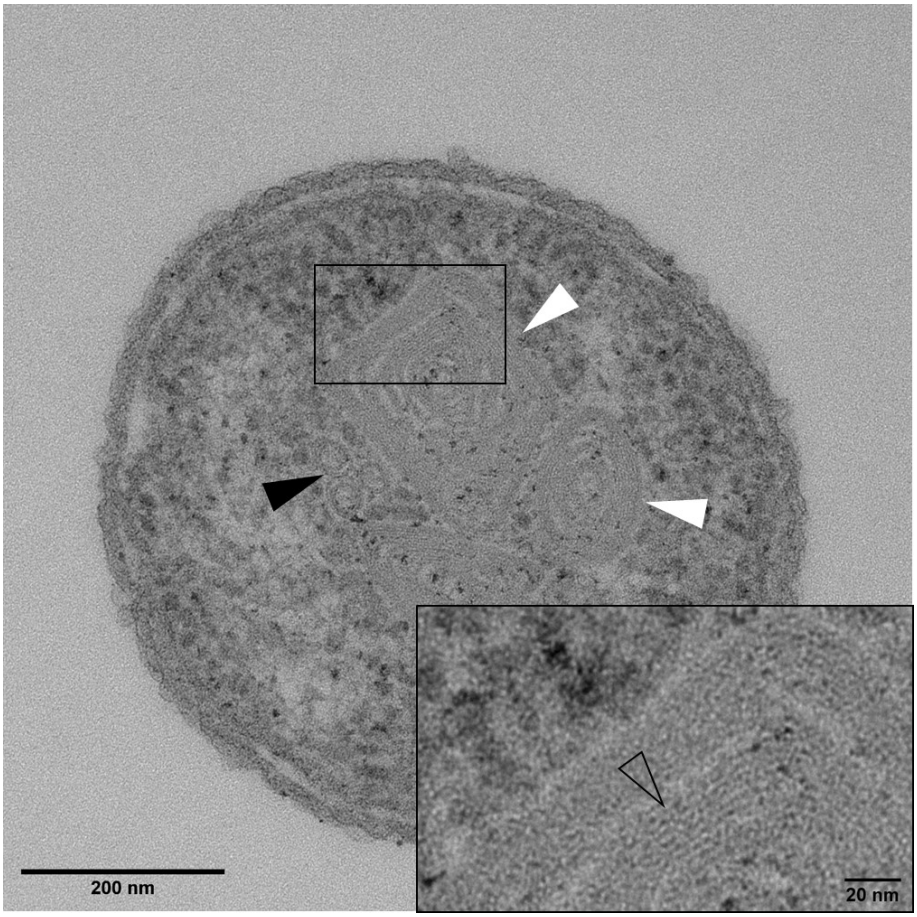


Figure 3.9. TEM image of a transverse section through an *E. coli* cell producing Rru_A0908. Wrapped sheets of protein (white arrows) and tubes (black arrows) are indicated. The inset is an enlarged view of the section outlined in the main image. Layered sheets are indicated (open arrow). Scale bar main image 200 nm, inset 20nm.

3.2.2.5 Rru_A0912

Rru_A0912 is a 9.0 kDa protein composed of 89 amino acids containing the Pfam 03319 domain. Proteins with the PF03319 domain are believed to exist as pentamers and form the vertices of the microcompartment shell. The crystal structure of Rru_A0912 has been resolved and was shown to exist in a pentameric state (Wheatley *et al.*, 2013).

When expressed in *E. coli*, Rru_A0912 does not form any structures and cells appear normal in TEM images (Fig. 3.10).

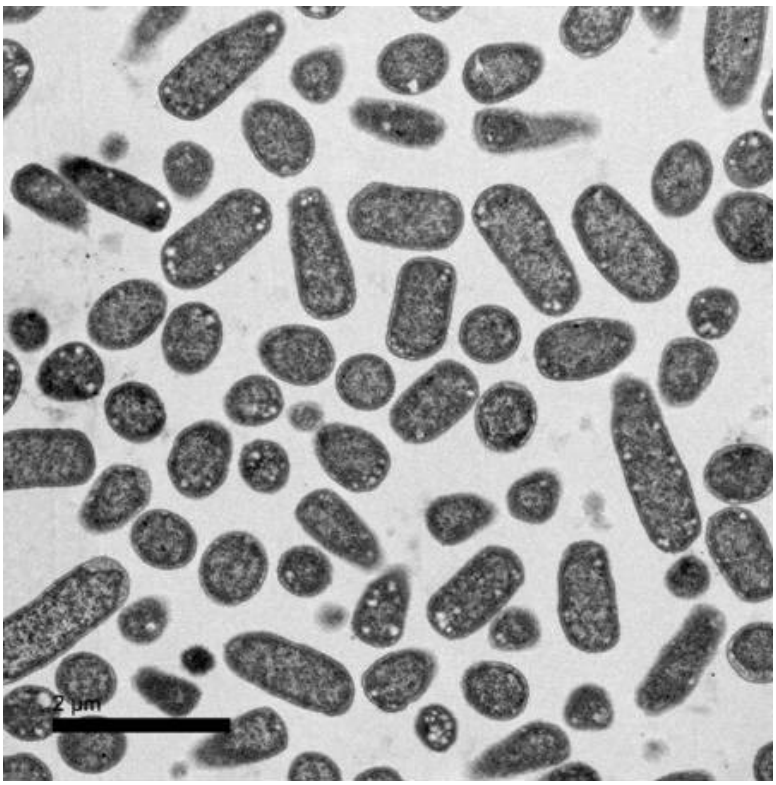


Figure 3.10. TEM image of *E. coli* producing Rru_A0912. Cells all appear normal. Scale bar 2 µm

3.2.2.6 Rru_A0915

Rru_A0915 is an 11.2 kDa, 99 amino acid, hexameric shell protein containing the PF00936 domain. Rru_A0915 does not share a high degree of similarity with any of the Pdu shell proteins (Table 3.1), 33 % similarity with PduU being the greatest. However, comparison with the protein sequence of another glycyl radical microcompartment shell protein, GrpU from *Pectobacterium wasabiae,* showed Rru_A0915 to share 63 % similarity (Fig.3.11). Crystal structures of the GrpU protein from *P. wasabiae* and also *from Clostridiales bacterium* 1_7_47FAA have shown that, despite low sequence similarity to other BMC shell proteins, the structural fold is very similar (Thompson *et al.*, 2014). Spectroscopic analysis showed evidence of an iron-sulphur centre proposed to be involved in electron transport or Fe-S cluster transport. A conserved cysteine-69 residue, shown to be involved in Fe-S binding, is also found in Rru_A0915 suggesting this protein may also house an Fe-S centre.

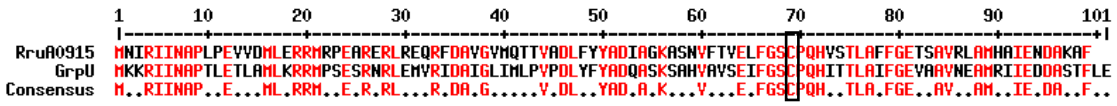


Figure 3.11. Protein sequence alignment of Rru_A0915 with GrpU from *P. wasabiae*. A conserved cysteine at position 69, proposed to be involved in Fe-S cluster binding or electron transport, is indicated (box). Conserved residues are indicated in red. Sequence alignment was performed in Multalin online sequence alignment tool (Corpet, 1988).

TEM images of Rru_A0915 over-produced in *E. coli* show filaments of protein in both longitudinal and transverse sections (Fig. 3.12), suggesting the protein hexamers are forming sheets. There is no evidence for the formation of tubes as with Rru_A0905 or rolled sheets as seen with Rru_A0908. Straight filaments are observed in most sections with filaments running close to the plasma membrane occasionally noted.

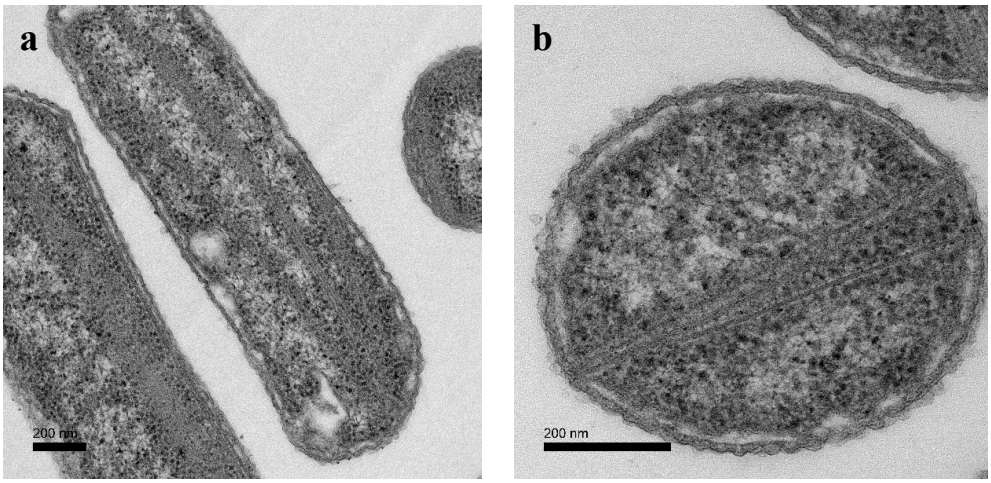


Figure 3.12. TEM images of Longitudinal (a) and transverse (b) sections of *E. coli* producing Rru_A0915. Protein filaments are indicated (arrows). Scale bars 200 nm

3.2.2.7 Summary of phenotypes observed with over-production of single shell protein of *R. rubrum*

The six shell proteins of *R. rubrum* examined in this study show a range of phenotypes when over-produced in *E. coli*. Whilst filaments of protein are observed in longitudinal sections upon the over-production of Rru_A0905, Rru_A0908 and Rru_A0915, the three-dimensional structures formed by each is varied. Rru_A0905 forms tubes, Rru_A0908 forms tightly-layered rolled sheets and Rru_A0915 forms straight sheets. Each of these individual structures may provide different properties needed in the overall formation of a microcompartment shell. The only trimeric shell protein in the *R. rubrum* BMC operon is Rru_A0906 and this was seen to form large aggregates upon over-production. The presence of other shell proteins may be necessary to allow this protein to form higher-level structures. No structures are observed

when Rru_A0907 and Rru_A0912 are over-produced. Again the presence of other shell proteins may be necessary for the formation of structural protein features within the cell.

3.2.3 Over-production of all *R. rubrum* shell proteins

A single construct harbouring all the *R. rubrum* shell protein genes (*rru_A0905*, *rru_A0906, rru_A0907, rru_A0908, rru_A0912, rru_A0915)*, cloned using the "Link-and-Lock" technique (McGoldrick *et al.*, 2005), was provided by Dr. David Palmer. Each of the genes had its own ribosome binding site and were cloned in the order in which they are found in the BMC operon (Fig. 3.13). *E. coli* BL21*(DE3) pLysS were transformed with a pET3a vector containing all six shell protein genes and the resulting strain was grown at 37 °C to an OD600 ≈ 0.6 in LB media containing ampicillin and chloramphenicol. Protein production was induced with 400 µM isopropyl β-D-1-thiogalactopyranoside (IPTG) and cultures were incubated with shaking overnight at 19 °C before harvesting for fixing, embedding and thin sectioning.

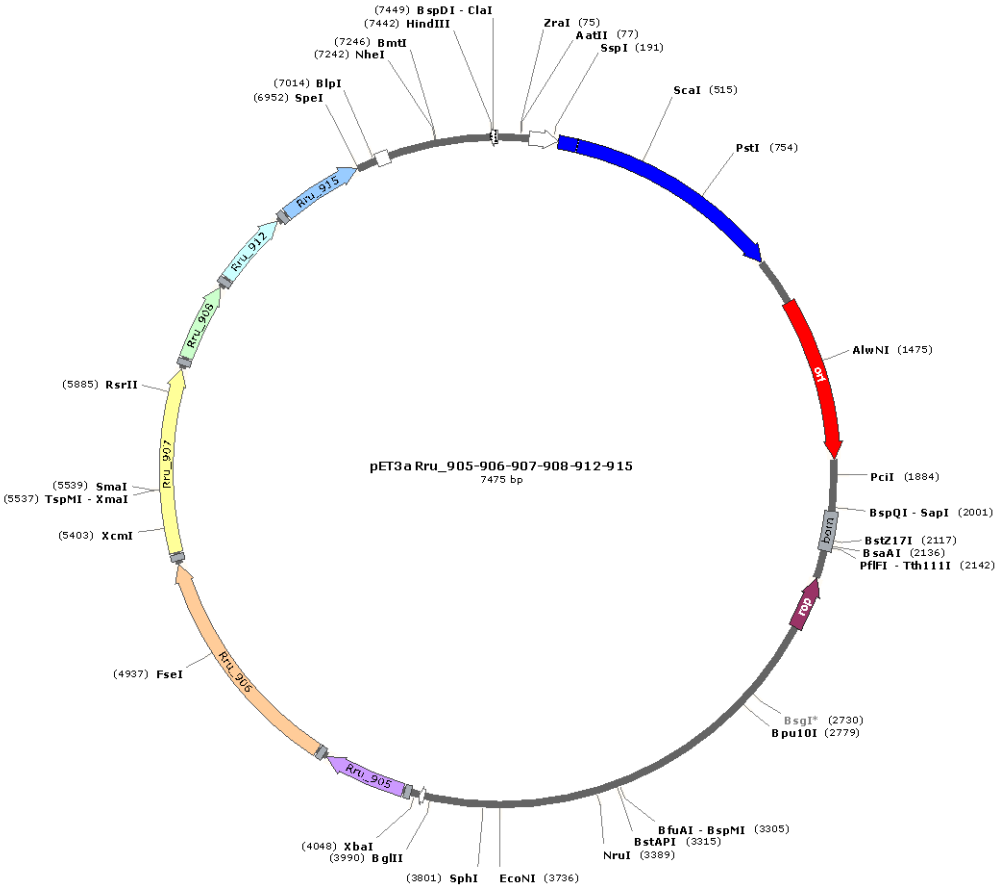


Figure 3.13. Schematic of the pET3a vector harbouring the 6 shell protein genes of the *R. rubrum* BMC operon. Genes coloured as follows: pink *rru_A0905*; orange *rru_A0906*; yellow *rru_A0907*; green *rru_A0908*; pale blue *rru_A0912*; mid blue *rru_A0915*. Ribosome binding site shown as grey box preceding each shell protein gene. Figure produced in SnapGene software (from GSL Biotech; available at snapgene.com)

3.2.3.1 Phenotype of *E. coli* overproducing all shell proteins of *R. rubrum*

Electron microscope images of *E. coli* over-producing all six *R. rubrum* shell proteins showed a range of structures within the cytoplasm. Rosette-like structures were observed in both longitudinal and transverse sections which are possibly rolled sheets of protein. Some longer filaments are also seen with many curving around the outer edges of the cell (Fig. 3.13). There are a small number of "closed" structures although the presence of ribosomes within these assemblies would suggest that these compartments are not completely closed (Fig. 3.14).

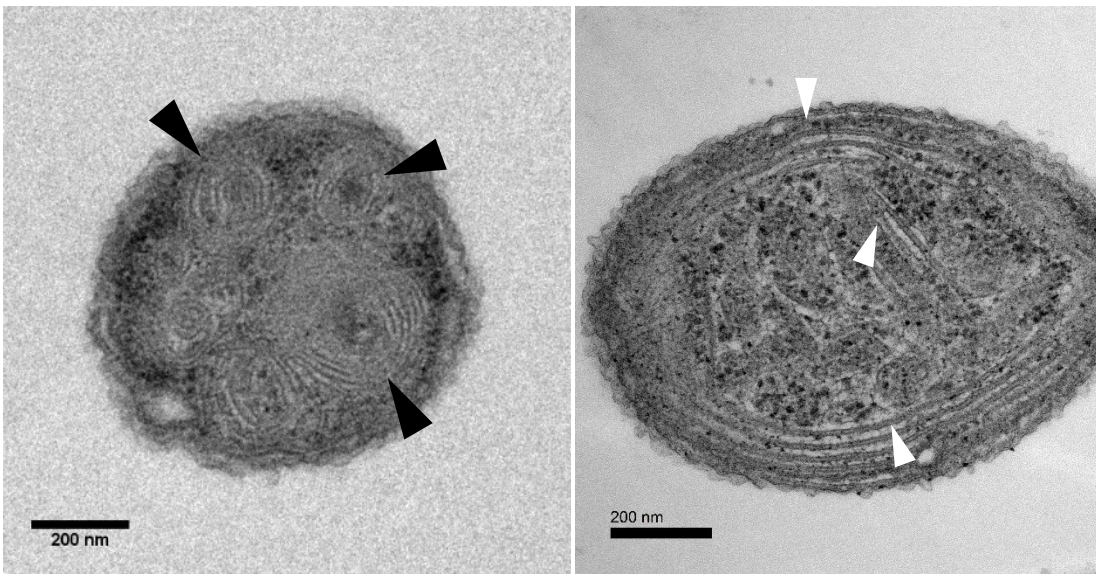


Figure 3.13. TEM images of *E. coli* over-producing all six shell proteins of the *R. rubrum* BMC operon. Rosette-like, swirled sheet structures (black arrows) and filamentous structures in and around the outside edge of the cell (white arrows) are indicated. Scale bars 200 nm

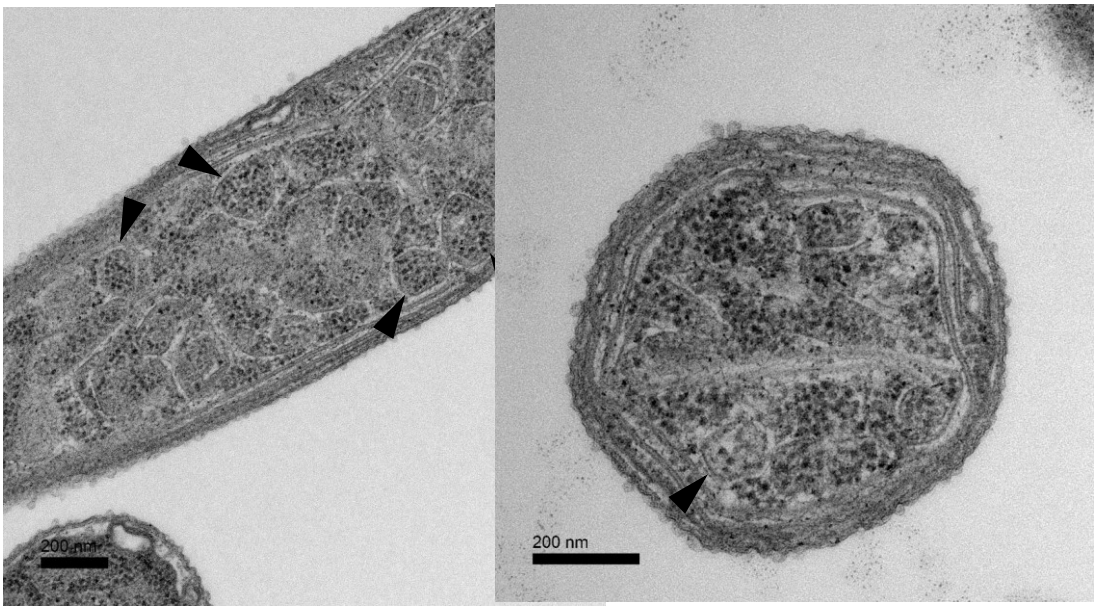


Figure 3.14 TEM images of *E. coli* over-producing all six shell proteins of the *R. rubrum* BMC operon. Closed structures (black arrows). Scale bars 200 nm

3.2.3.2 SDS-PAGE analysis

Shell protein overproduction was confirmed by SDS–PAGE analysis of extracts taken from cultures of the strain. Samples (10 mL) were taken from cultures before embedding. After centrifugation the cell pellet was re-suspended in buffer, mixed with sample disruption buffer and heated to 100 °C for 10 min before loading onto an SDS-PAGE gel. It was difficult to confirm the presence of all the shell proteins as several are of a similar size (Rru_A0905, Rru_A0908 and Rru_A0915) and so ran at the same positon on the gel (Fig. 3.15a). Rru_A0906, Rru_A0907 and Rru_A0912 can be seen clearly on the gel. Rru_A0907, a 21.5 kDa protein, ran high on the SDS gel at ~35 kDa. This had been observed previously upon purification of Rru_A0907 alone by IMAC (Fig 3.15b).

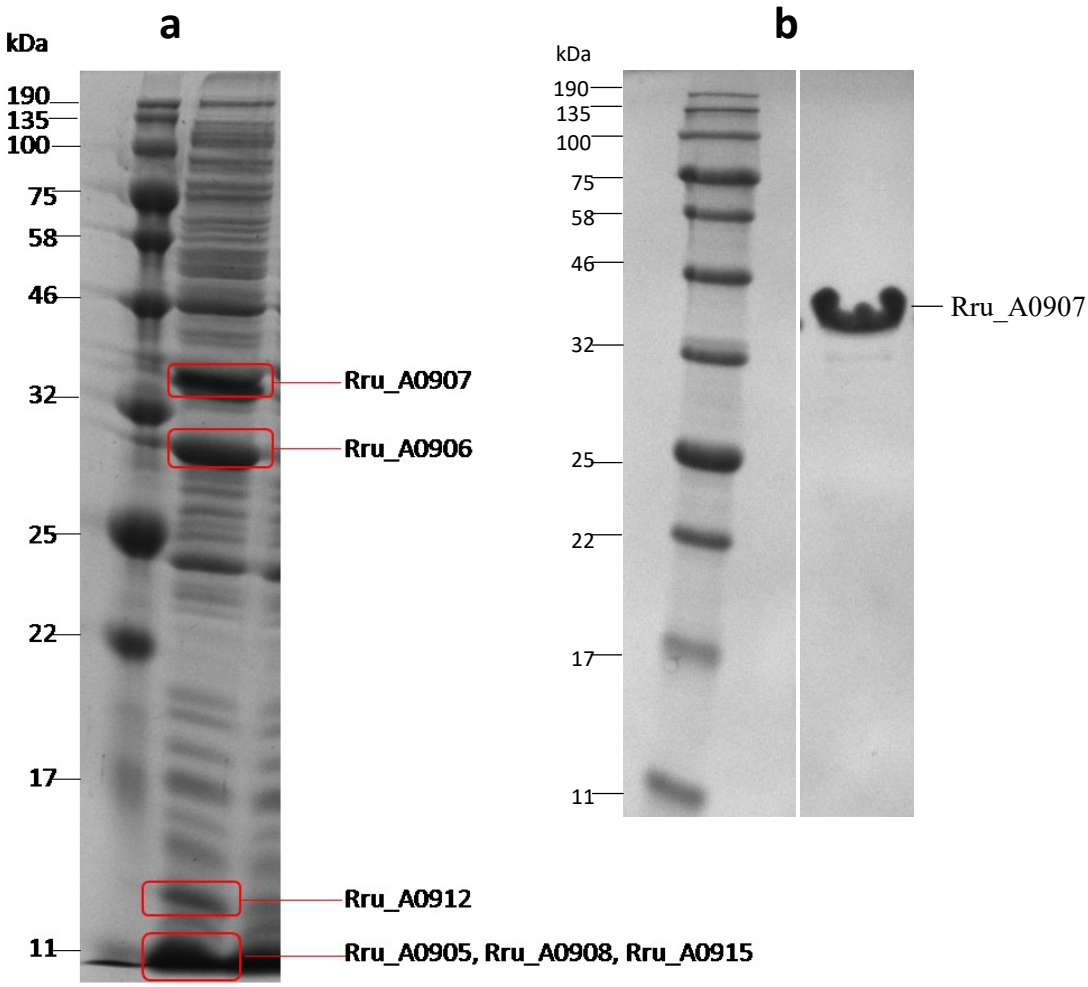


Figure 3.15. a) SDS-PAGE of protein overproduction in *E. coli* transformed with a vector containing all six *R. rubrum* shell protein genes. Red boxes indicate bands of over-produced proteins. b) SDS-PAGE of IMAC-purified Rru_A0907. Although Rru_A0907 has a predicted molecular mass of 21.5 kDa, it migrates slower by SDS-PAGE.

3.2.3.3 2D gel electrophoresis and Maldi-Tof analysis

It was not possible to distinguish between Rru_A0905, Rru_A0908 and Rru_A0915 on an SDS_PAGE gel due to the similarity in size of these proteins. 2D gel electrophoresis was used to separate the proteins by isoelectric focusing before then separating by size. Proteins were first isolated using the method described in Section 2.4.7 to provide a cleaner sample prior to 2D gel electrophoresis.

To confirm the identity of the separated proteins, bands/spots were excised form the gel and prepared for MALDI-TOF analysis as described in Section 2.4.6. Samples were analysed on a Bruker UltraFlex system.

Rru_A0905, Rru_A0906, Rru_A0907, Rru_A0908 and Rru_A0915 were positively identified in the samples, however, the vertex protein, Rru_A0912, was not detected (Fig. 3.16).

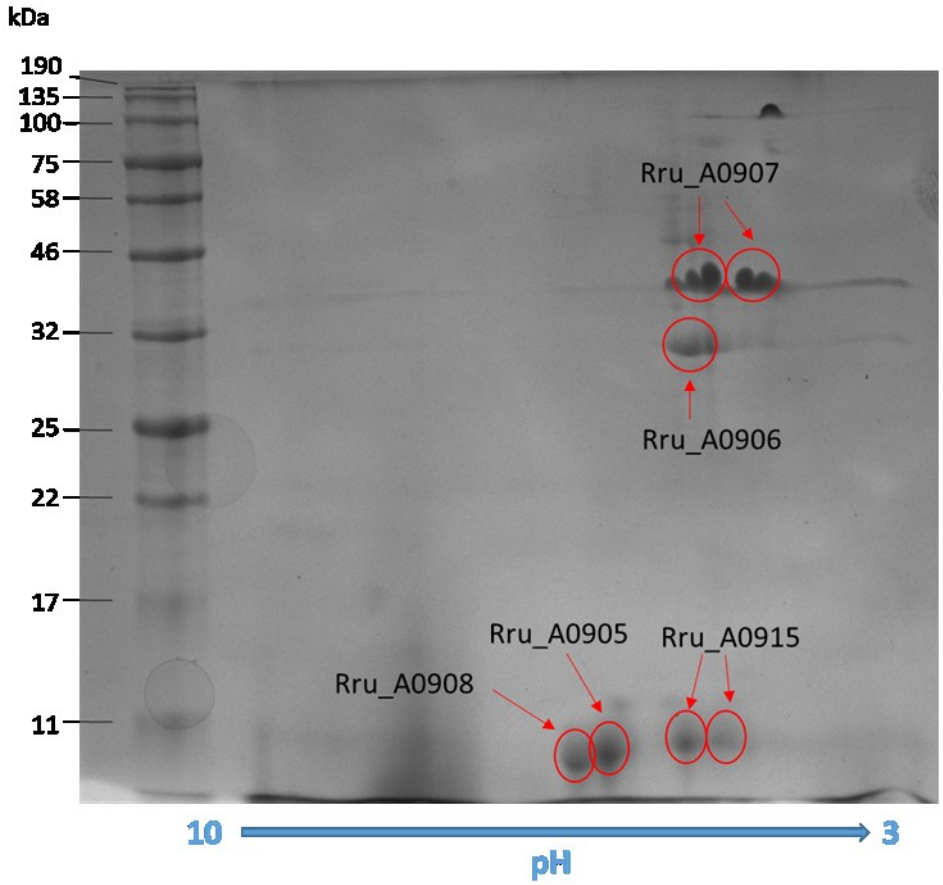


Figure 3.16. 2D gel electrophoresis and identification by MALDI-TOF of proteins overproduced in *E. coli* expressing *R. rubrum* shell proteins. Protein marker is in left hand lane. Proteins identified by MALDI-TOF are circled in red and labelled with the protein identified. The blue arrow below the gel shows the direction of the pH gradient of the 1st separation dimension.

A band corresponding to a protein the size of Rru_A0912 was observed in the SDS-PAGE analysis (Fig. 3.15) but these samples were of whole cell lysate whereas for the 2D-analysis the proteins had been isolated before separation and some protein is lost during the isolation procedure. Additionally, Rru_A0912 is a predicted vertex protein which is present in the microcompartment in much lower numbers than the facet forming shell proteins. A combination of fewer proteins in the sample and loss of protein during the isolation may therefore have resulted in the inability to detect Rru_A0912 by MALDI.

3.2.3.4 Visualisation of purified shell protein samples by TEM

Overproduction of the six shell proteins of the *R. rubrum* BMC operon failed to produce microcompartment-like structures in *E. coli* (Section 3.2.3.1). However, it was possible to isolate the protein structures produced using the protocol described in Section 2.4.7. Isolated

protein structures were imaged by TEM following either (i) fixing and negative staining of samples applied directly to formvar-coated copper grids, or (ii) fixing and resin-embedding samples before thin sectioning and staining.

Interestingly, despite not observing closed compartments *in vivo*, the isolated protein samples did appear to contain closed compartments. These were observed in both negative stained (Fig. 3.17) and thin-sectioned samples (Fig.3.18). Other structures were also observed, including filaments and longer tube-like structures.

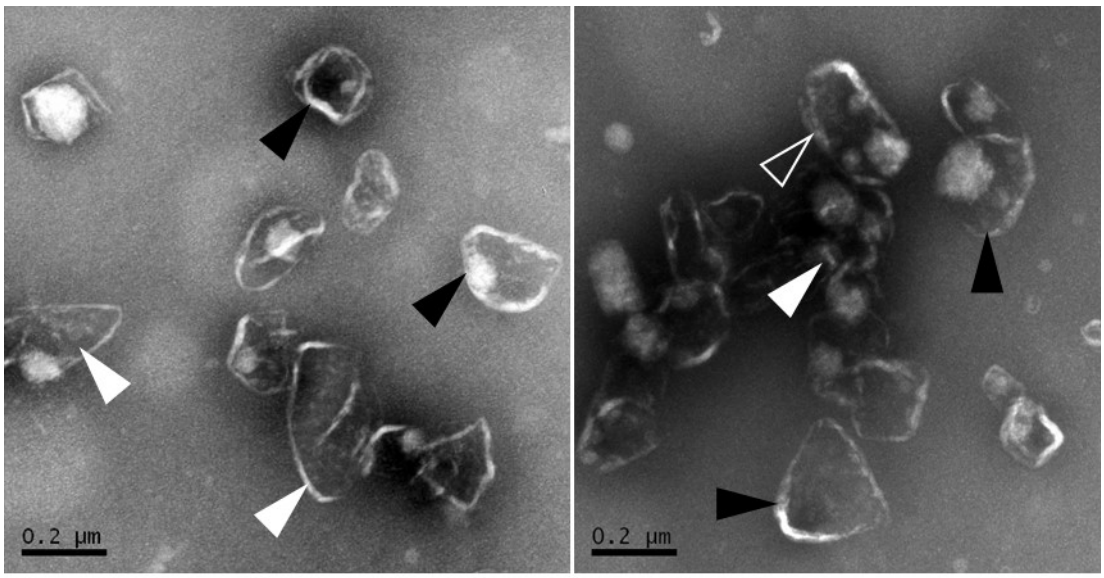


Figure 3.17. TEM of negative-stained isolated proteins produced on expression of the six *R. rubrum* shell protein genes in *E. coli*. Closed compartment like structures (black arrows) mis-formed open structures (white arrows) and longer capsule like structures (open arrows) are indicated. Scale bar 0.2 µm

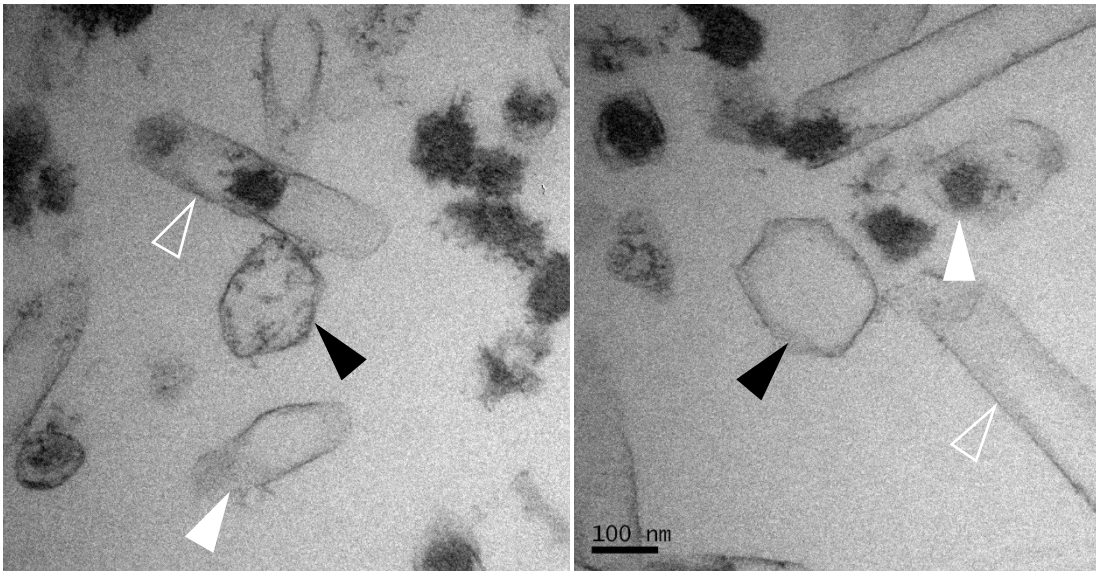


Figure 3.18. TEM images of thin-sectioned isolated proteins produced following expression of six *R. rubrum* shell protein genes in *E. coli*. Closed compartment like structures (black arrows) mis-formed open structures (white arrows) and elongated capsule like structure (open arrows) are indicated. Scale bar 100 nm

The presence of closed compartment–like structures was unexpected as intact microcompartments had not been detected in sections of whole cells. It is possible that the formation of compartments is a dynamic process and the conditions during the isolation procedure were favourable for compartment formation. Alternatively, it may be that the appearance of compartments was an artefact of imaging sections through the sample. A cross-section of wrapped sheets or tubes of protein may also appear to be compartments. Electron tomography could be used to obtain more information on the topography of the compartments.

3.2.3.5 Production of enzyme cargo with *R. rubrum* shell proteins

One model of carboxysome assembly proposes that the shell forms around an aggregate of the main pathway enzyme (Chen *et al.*, 2013; Kerfeld and Melnicki, 2016). To see if this was also necessary for the formation of *R. rubrum* compartments in *E. coli* the signature enzyme of the *R. rubrum* BMC, a diol dehydratase glycyl radical enzyme, was over-produced in conjunction with the shell proteins. The genes were cloned into a single copy plasmid, pETcoco2, as the signature enzyme gene (*rru_A0903*) was unable to be cloned into a pET3a vector containing the six shell proteins. The plasmid was maintained at a single copy per cell, prior to induction with IPTG, by the inclusion of 0.2 % (w/v) glucose in the culture medium. Production of the signature enzyme with the shell proteins resulted in the formation of swirled sheets of protein as previously seen in the absence of this pathway enzyme. However, aggregates of protein appeared to be formed in association with the protein sheets. Some closed structures, possibly compartments, are also visible (Fig. 3.19). Expression of the shell protein

genes from the pETcoco2 plasmid without the pathway enzyme gene *rru_903* was included as a control and showed the same phenotype as seen with expression of the genes from a pET3a plasmid (Fig. 3.19).

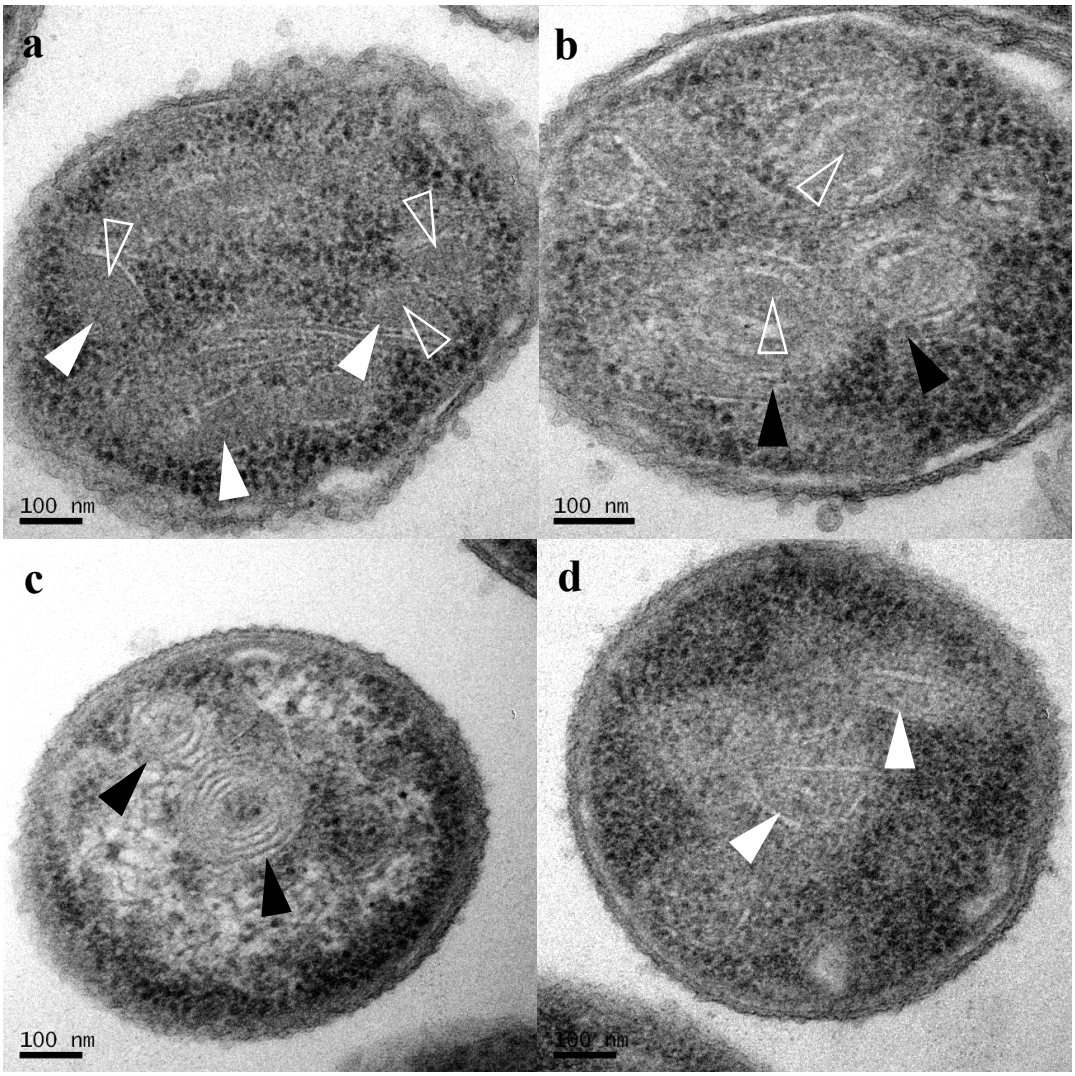


Figure 3.19. TEM images of *E. coli* following expression of the diol dehydratase glycyl radical enzyme gene *rru_A0903* and the *R. rubrum* shell protein genes in pETcoco2 (a & b) and just the six shell protein genes (c and d). Swirled sheets of protein seen in both conditions are indicated (black arrows), closed structures also indicated in both conditions (white arrows). Protein aggregates are indicated in images a & b (open arrows). Scale bars 100 nm

3.2.3.6 Summary of the recombinant production of the six shell proteins of the *R. rubrum* BMC operon

In summary, overproduction of the six shell proteins of the *R. rubrum* operon in *E. coli* gives rise to a range of structures within the cell, including some mis-formed, open compartment-like assemblies. Production of the signature pathway enzyme, diol dehydratase, as a scaffold did not improve microcompartment formation. However, imaging of the isolated protein

structures does appear to show the presence of closed compartment-like structures suggesting that the formation of compartments is a dynamic process with compartments only forming in favourable conditions. The dynamic nature of the protein-protein interactions between shell protein hexamers has been demonstrated by atomic force microscopy (Sutter *et al.*, 2016). Shell protein hexamers both disassociated from and assembled into protein sheets showing the flexible nature of facet assembly. Isolated microcompartments were also visualised by TEM, despite not being visible *in vivo,* in a study examining the expression of shell proteins from the *Haliangium ochraceum* BMC operon (Lassila *et al.*, 2014). In this case, the authors did not exclude the possibility that microcompartments were formed during isolation rather than *in vivo*.

3.2.4 Microcompartment assembly with minimal shell proteins

The production of microcompartments has been demonstrated using just three *H. ochraceum* shell proteins (Lassila *et al.*, 2014), although these compartments were only observed after isolation. Further studies in-house have demonstrated the assembly of microcompartments *in vivo* using a minimal complement of shell proteins from both Pdu and Cut BMC operons. In both instances, compartments formed with just a PduA-like, a PduK-like and a PduN-like shell protein (*Unpublished data – Dr. M. Lee and Dr. M. Liang)*. Constructs were therefore cloned containing a minimal shell protein gene complement – one Pdu A-like, one PduK-like and one PduN-like. Two PduA–like genes are present in the *R. rubrum* BMC operon, *rru_A0905* and *rru_A0908*, therefore two alternative constructs were cloned using the "Link-and-Lock" protocol (McGoldrick *et al.*, 2005): pET3a *rru_A0905-A0907-A0912* and pET3a *rru_A0908-A0907-A0912.* BL21*(DE3) pLysS *E. coli* cells were transformed with the vectors, grown to an $OD_{600} \approx 0.6$ before induction of protein production with IPTG. Cells were incubated at 19 °C overnight before harvesting for fixing, embedding and thin sectioning.

3.2.4.1 Overexpression of rru_A0905, rru_A0907 and rru_A0912

Over-production of Rru_A0905, Rru_A0907 and Rru_A0912 in *E. coli* resulted in the formation of numerous protein structures throughout the cytoplasm. Some structures resembled closed compartments but most structures observed appeared to be swirled sheets of protein (Fig. 3.20).

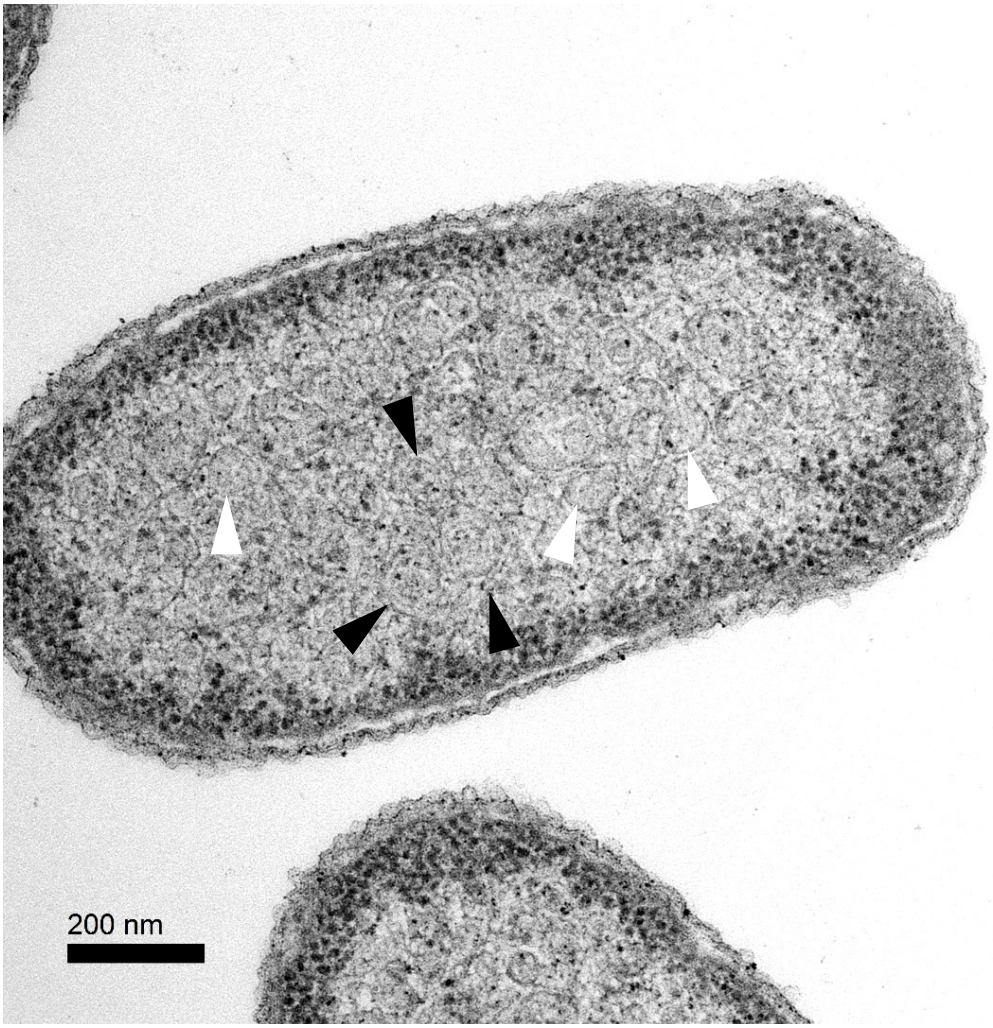


Figure 3.20. TEM image of *E. coli* overproducing Rru_A0905, Rru_A0907 and Rru_A0912. Swirls of protein (black arrows) and closed compartment like structures (white arrows) are indicated. Scale bar 200 nm

3.2.4.2 Overexpression of rru_A0908, rru_A0907 and rru_A0912

Over-production of Rru_A0908, Rru_A0907 and Rru_A0912 in *E. coli* resulted in the formation of numerous protein structures throughout the cytoplasm similar to those observed with Rru_A0905 instead of Rru_A0908. Swirled protein structures and a few closed compartment like structures were again observed (Fig. 3.21).

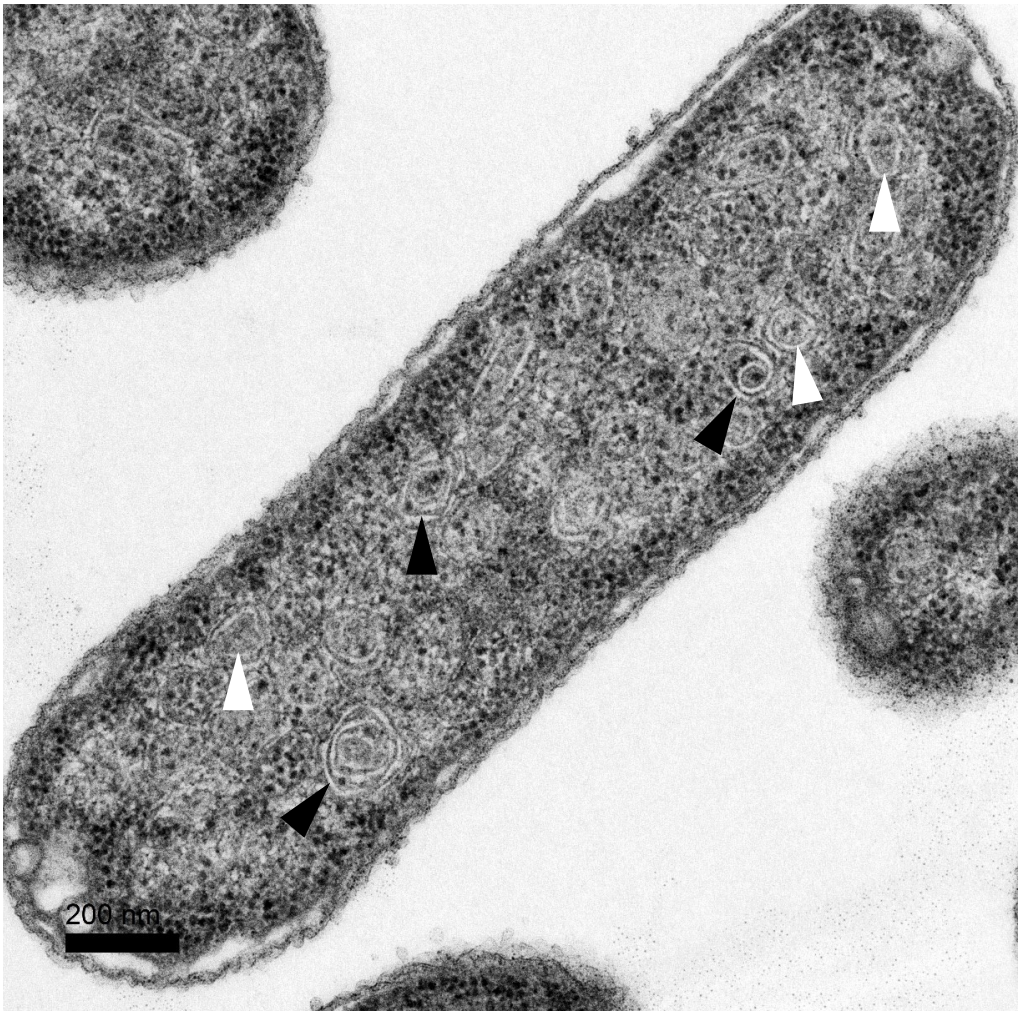


Figure 3.21. TEM image of *E. coli* overproducing Rru_A0908, Rru_A0907 and Rru_A0912. Swirls of protein (black arrows) and closed compartment like structures (white arrows) are indicated. Scale bar 200 nm

3.2.4.3 Purification of minimal shell protein structures

Purified microcompartment-like structures were observed despite not being evident *in vivo* whole cell sections when expressing all the shell protein genes of *R. rubrum*. A small number of closed compartment-like structures were observed in *E. coli* upon expression of a minimal complement of shell protein genes (Sections 3.2.4.1 and 3.2.4.2), although the majority of structures appeared to be swirled sheets of protein. Purification of these proteins was carried out using the method described in Section 2.4.8. Purified protein was fixed and stained on formvar-coated copper grids as described in Section 2.4.9.1 and imaged by TEM.

Purified samples of Rru_A0905-A0907-A0912 and Rru_A0908-A0907-A0912 both contained structures resembling microcompartments (Fig 3.22). The structures appear to be closed compartments despite observing few of these structures *in vivo*. These are similar to the purified structures observed when all the shell proteins were produced and again supports

microcompartment formation being a dynamic process with purification conditions enabling compartment formation. Smaller, circular structures were also observed in these samples. It is unknown what these structures are but may be lipid vesicles formed upon cell lysis.

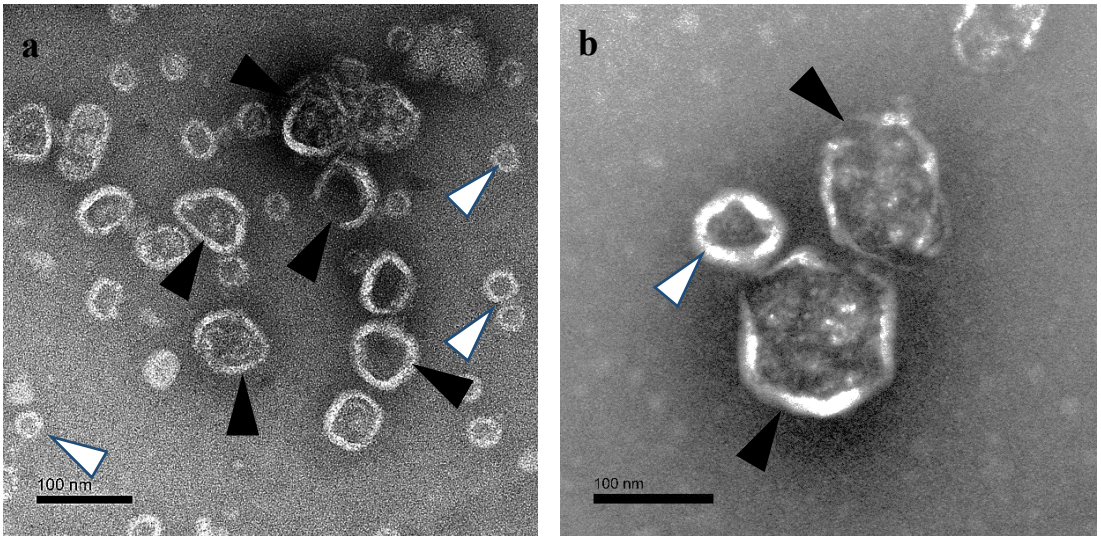


Figure 3.22. TEM images of purified compartments produced in *E. coli* upon expression of a minimal shell protein gene complement. a) Rru_A0905-A0907-A0912; b) Rru_A0908-A0907-A0912. Microcompartment–like structures (black arrows) and smaller structures (white arrows) indicated. Scale bars 100 nm

3.2.5 Summary of recombinant production of a minimal *R. rubrum* shell protein complement in *E. coli*

The formation of microcompartments using a minimal number of shell proteins has previously been demonstrated (Lassila *et al.*, 2014) and was also observed in this study by expressing the genes for 1 hexameric shell protein, 1 pentameric vertex protein and 1 C-terminally extended hexameric shell protein. As was also seen with all six shell proteins, fully closed microcompartments were not clearly evident in whole cells. However, TEM imaging of purified samples showed the presence of closed microcompartment-like structures, suggesting the conditions during purification are favourable for BMC formation. This supports microcompartment formation being a dynamic process with hexamers and pentamers able to associate and dissociate from protein sheets as shown by Sutter *et al.*, 2016.

The formation of microcompartments with a reduced number of shell proteins is an interesting observation. Future studies could examine different permutations of shell proteins to investigate whether compartments are formed with a range of shell protein combinations. This may allow the fine tuning of compartments to give variable flux of substrate and product if proteins with a range of pore sizes and properties are substituted for each other. The formation of microcompartments during purification also highlights the potential for *R. rubrum* BMC

production *in vitro* by the addition of purified individual shell proteins as has been demonstrated with the *H. ochraceum* minimal shell protein microcompartment (Hagen *et al.*, 2018). Again this opens up the possibility of exploring compartment formation when varying the stoichiometry of the shell protein components which may result in compartments of different size or with varying targeting and/or flux properties.

3.3 Results: Growth of *R. rubrum* on sodium succinate and 1,2-propanediol: analysis of resultant phenotypes by TEM

The presence of the BMC operon within the *R. rubrum* genome indicates that this bacterium is capable of forming microcompartments under specific conditions. Pdu BMCs are produced in *S. enterica* when grown on 1,2-PD, the substrate of the encapsulated enzyme pathway (Bobik *et al.*, 1999). As 1,2-PD is the predicted substrate of the *R. rubrum* BMC studies were designed to examine whether microcompartments were formed when *R. rubrum* was grown on 1,2-PD. *R. rubrum* is a facultative anaerobe therefore both aerobic and anaerobic growth on 1,2-PD was examined. Initial growth studies were carried out to identify suitable conditions for growth prior to embedding samples for TEM analysis.

3.3.1 *R. rubrum* growth curves

3.3.1.1 Aerobic growth curves

R. rubrum cultures (0.5 mL) were grown aerobically in 24-well culture plates and the OD at 600 nm measured using a plate reader as described in Section 2.2.10.2. Overnight cultures were centrifuged and re-suspended to an OD_{600} = 1.0 in culture medium without any carbon source. Re-suspended cells (25 µL) were added to each well containing 475 µL of media with the carbon source(s) to be tested to give a starting OD_{600} ~ 0.05. The OD_{600} was measured every 30 minutes for 24+ h.

The standard growth media for *R. rubrum* (see Section 2.1.5) contains 0.2 % (w/v) sodium succinate as a carbon source. However, the predicted substrate of the encapsulated enzyme pathway of the *R. rubrum* BMC is 1,2-propanediol (1,2-PD). Growth of *R. rubrum* was therefore measured using either succinate or 1,2-PD as the sole carbon source. Initial studies showed that after 12 hours *R. rubrum* reached a maximum OD_{600} ~ 0.7 when grown with 0.2 % (w/v) succinate. However, growth on 0.2 % (v/v) 1,2-PD was substantially reduced with a maximum OD_{600} of only 0.1 being reached after 24 h (Fig 3.23).

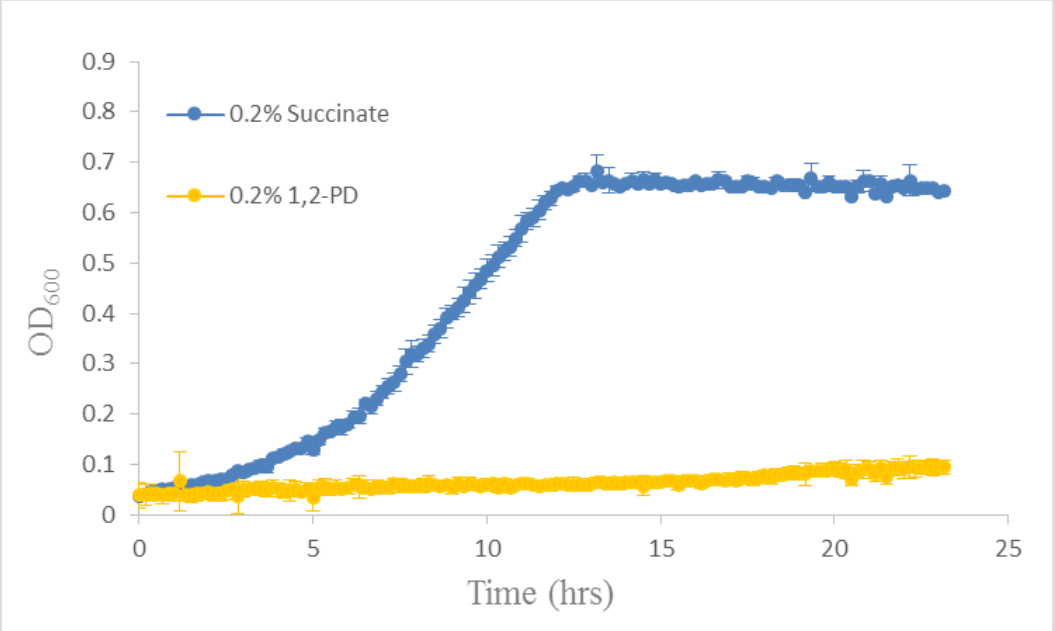


Figure 3.23. Aerobic growth curves of *R. rubrum* grown with either 0.2 % (w/v) succinate or 0.2 % (v/v) 1,2-PD as the sole carbon source. Data represents the mean ± SD of 3 replicates.

It was anticipated that greater growth would be necessary for any future studies when assessing the structural features of *R. rubrum* when grown in the presence of 1,2-PD. Therefore, growth was subsequently measured using a combination of succinate at a lower concentration (0.1% and 0.05% (w/v)) supplemented with 0.2% (v/v) 1,2-PD. The standard media containing 0.2% (w/v) sodium succinate as a sole carbon source was included as a reference and was seen to produce the best growth with cultures reaching an OD_{600} ~ 0.8 after 24 h. Lower succinate concentrations resulted in reduced maximum growth, reaching an OD_{600} of ~ 0.5 and ~ 0.3 for 0.1 % and 0.05 % (w/v) succinate respectively. 0.2 % (v/v) 1,2-PD as the sole carbon source again resulted in a much reduced maximum growth (OD_{600} ~ 0.2). Supplementing 0.2 % (v/v) 1,2-PD with 0.1% and 0.05% (w/v) succinate increased growth with a maximum OD_{600} ~ 0.65 and ~ 0.5 respectively (Fig. 3.24).

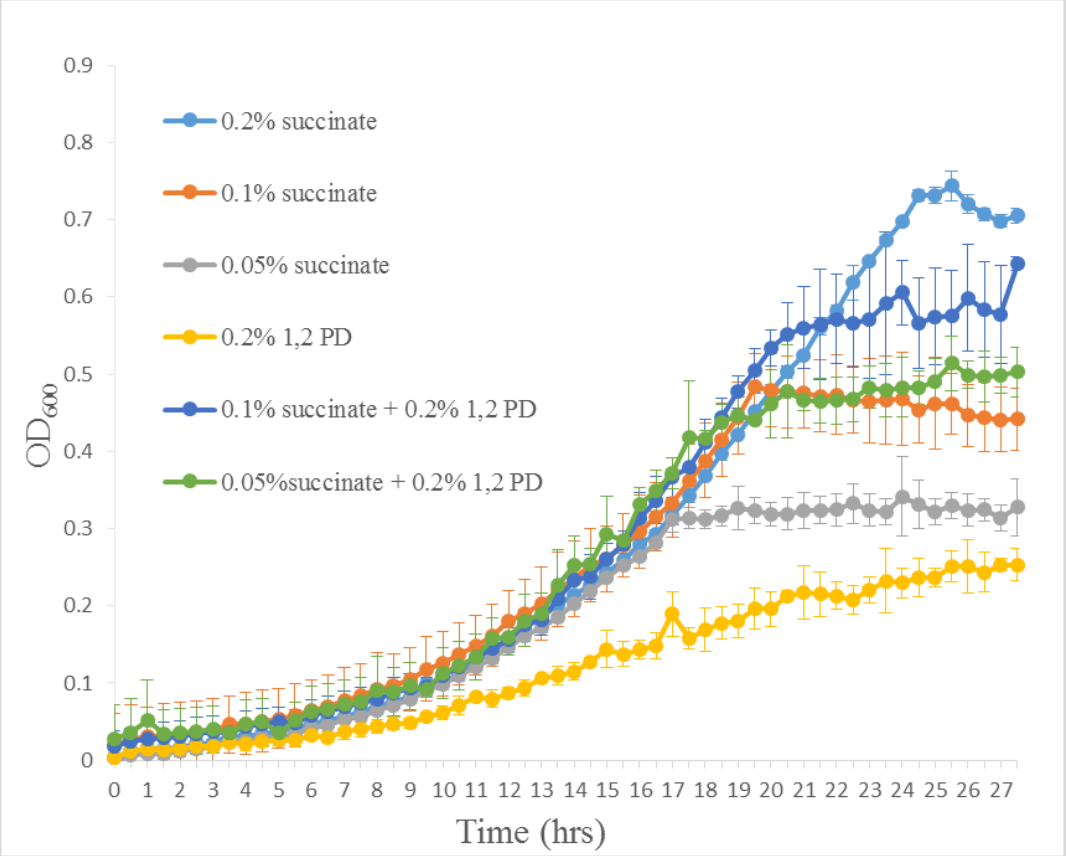


Figure 3.24. Aerobic growth of *R. rubrum* with 0.05%, 0.1 % or 0.2 % (w/v) succinate only; 0.2 % (v/v) 1,2-PD only; and 0.05 % (w/v) succinate + 0.2 % (v/v) 1,2-PD or 0.1 % (w/v) succinate + 0.2 % (v/v) 1,2-PD. Data represents the mean ± SD of 3 replicates and is representative of two experiments.

Further experiments were conducted to measure the effect of higher concentrations of 1,2-PD combined with 0.05 % (w/v) succinate on growth. Growth was measured at concentrations of 0.2 %, 0.4 %, 1 % and 2 % (v/v) 1,2-PD both as the sole carbon source and combined with 0.05% (w/v) sodium succinate (Fig. 3.25). These higher concentrations of 1,2-PD did not appear to improve growth above that seen with 0.2 % (v/v) 1,2-PD with a maximum OD_{600} of approximately 0.1 being measured after 30 h with all concentrations of 1,2-PD tested. Addition of 0.05 % (w/v) sodium succinate to the different 1,2-PD concentrations increased the maximum OD_{600} to 0.32 for 0.05 % (w/v) succinate + 0.2 % (v/v) 1,2-PD; 0.37 for 0.05 % (w/v) succinate + 0.4 % (v/v) 1,2-PD; 0.35 for 0.05 % (w/v) succinate + 1 % (v/v) 1,2-PD and 0.28 for 0.05 % (w/v) succinate + 2 % (v/v) 1,2-PD. Growth was not improved compared with growth on 0.05 % (w/v) succinate only (OD_{600} ~0.33).

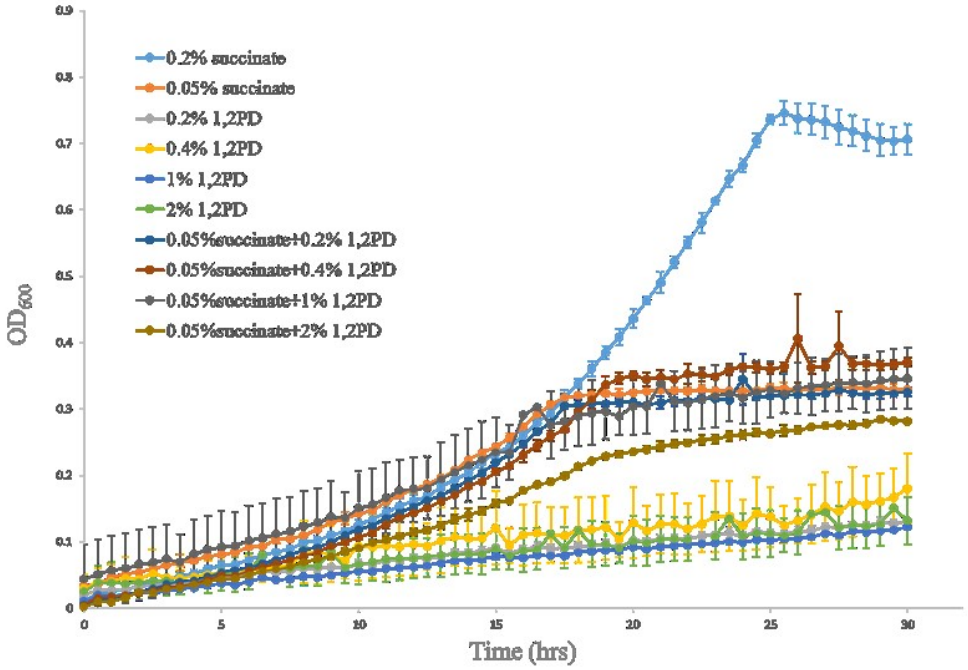


Figure 3.25. Aerobic growth of *R. rubrum* with 0.2 % or 0.05 % (w/v) succinate only; 0.2 %, 0.4 %, 1 % or 2 % (v/v) 1,2-PD only; and 0.05 % (w/v) succinate + 0.2 % (v/v) 1,2-PD, 0.05 % (w/v) succinate + 0.4 % (v/v) 1,2-PD, 0.05 % (w/v) succinate + 1 % (v/v) 1,2-PD, or 0.05 % (w/v) succinate + 2 % (v/v) 1,2-PD. Data represents the mean ± SD of 2 replicates and is representative of two experiments

These aerobic growth studies have shown that *R. rubrum* can grow with both succinate and 1,2-PD as a carbon source although growth is minimal when grown on 1,2-PD alone. Growth with a combination of both succinate and 1,2-PD in some instances appears to be marginally greater than with either carbon source alone but this was not consistently observed. Consequently, future studies examined the resulting phenotypes following growth on either a single carbon source or with a combination of the succinate and 1,2-PD.

3.3.1.2 Anaerobic growth curves

Media for anaerobic growth was prepared in stoppered, glass culture bottles as described in Section 2.2.9. TMAO was added as a terminal electron acceptor to the culture media at a final concentration of 40 mM in addition to 0.2 % (w/v) sodium succinate. Cultures were inoculated at a starting OD_{600} of approximately 0.1 from an aerobically grown starter culture which had been centrifuged and re-suspended in sterile media without any carbon source.

Initial studies compared anaerobic growth in light and dark conditions. It was seen that in the absence of light, growth was negligible but, when grown in the light, cultures reached an OD_{600}

~ 1.0 (Fig. 3.26). All further anaerobic growth studies were conducted in light conditions as described in Section 2.2.9.3.

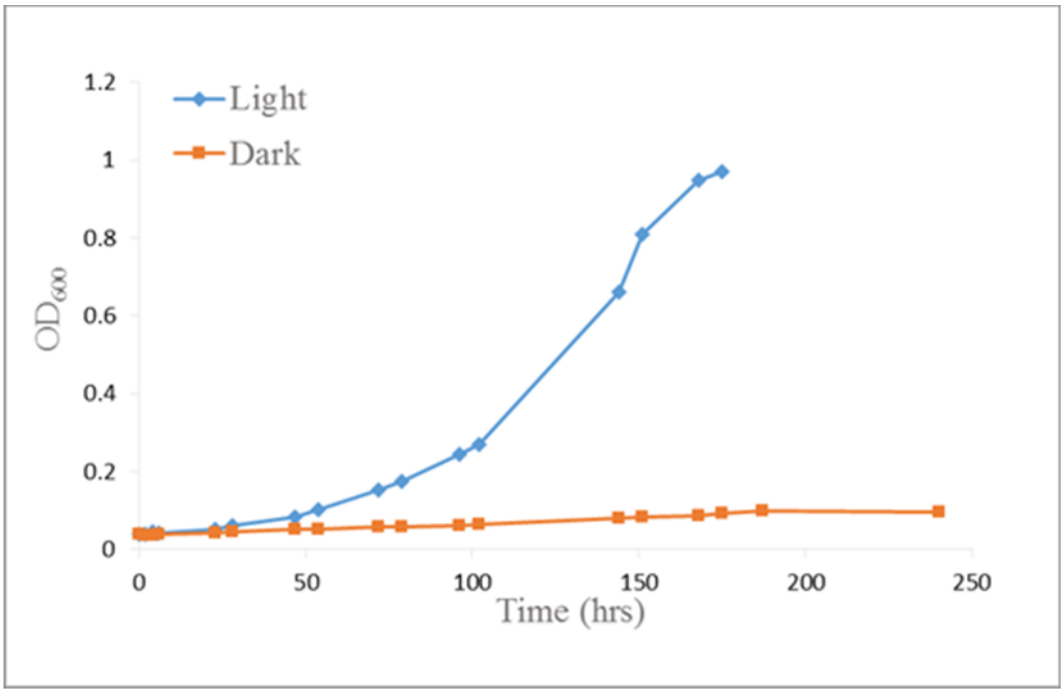


Figure 3.26 Anaerobic growth of *R. rubrum* in light and dark conditions. *R. rubrum* was grown on 0.2 % (w/v) succinate in light and dark conditions. Data is representative of 2 experiements.

Anaerobic growth of *R. rubrum* in light conditions was measured with the following carbon sources: i) 0.2 % (w/v) succinate, ii) 0.1 % (w/v) succinate, iii) 0.05 % (w/v) succinate, iv) 0.2 % (v/v) 1,2-PD, v) 0.1 % (w/v) succinate + 0.2 % (v/v) 1,2-PD, and vi) 0.05 % (w/v) succinate + 0.2 % (v/v) 1,2-PD. Growth with 0.2 % (w/v) succinate reached an OD_{600} of approximately 3.0 after 28 h (Fig. 3.27). Less growth was seen with lower succinate concentrations, a maximum OD_{600} of approximately 2.0 and 1.3 with 0.1 % and 0.05 % (w/v) succinate respectively. The least growth was seen with 0.2 % (v/v) 1,2-PD with an OD_{600} of approximately 1.0 being reached after 48 h. Combining 0.1 % (w/v) succinate or 0.05 % (w/v) succinate with 0.2 % (v/v) 1,2-PD increased the maximum growth seen with an OD_{600} of ~2.3 and ~3.5 after 48 h for 0.2 % (v/v) 1,2-PD with 0.05 % (w/v) or 0.1 % (w/v) succinate respectively. The OD_{600} appeared to plateau after 28 h with succinate alone but when combined with 1,2-PD the OD_{600} continued to increase up to the 48 h final time point.

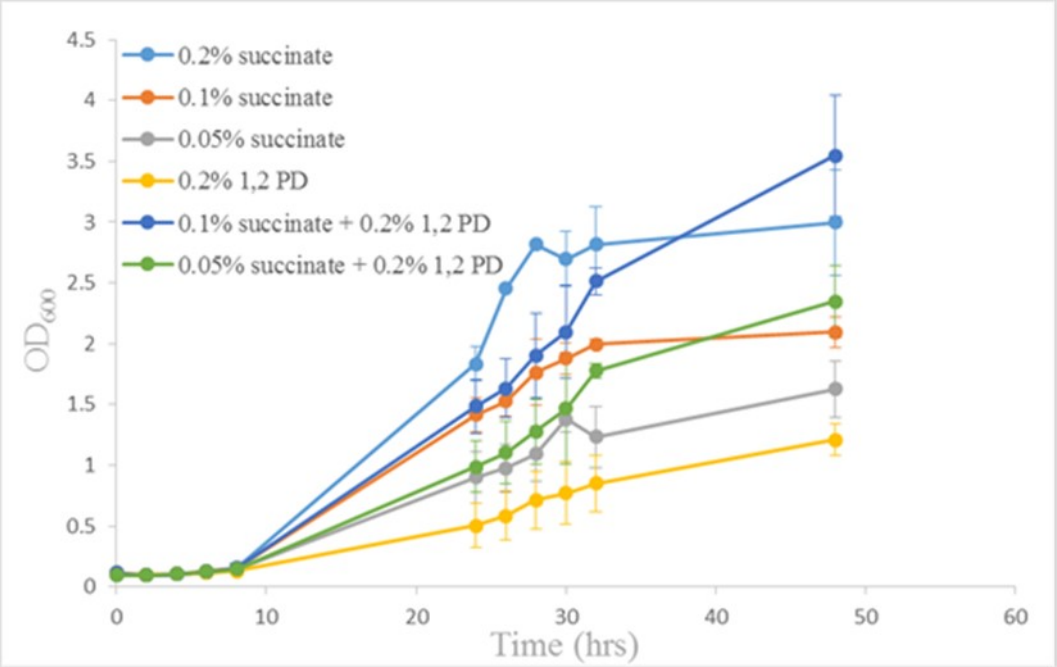


Figure 3.27. Anaerobic growth of *R. rubrum* under light conditions. *R rubrum* cultures were grown with i) 0.2 % (w/v) succinate; ii) 0.1 % (w/v) succinate; iii) 0.05 % (w/v) succinate; iv) 0.2 % (v/v) 1,2-PD; v) 0.1 % (w/v) succinate + 0.2 % (v/v) 1,2-PD; and vi) 0.05 % (w/v) succinate + 0.2 % (v/v) 1,2-PD. Data represent the mean ± SD of two experiments.

R. rubrum was subsequently cultured anaerobically with increased concentrations of 1,2-PD both as the sole carbon source and in combination with 0.05 % (w/v) sodium succinate to see if there was increased growth. Growth curves were measured with 0.2 %, 0.4 %, 1 %, 2 % and 4 % (v/v) 1,2-PD but no improvement in growth above that observed with 0.2 % (v/v) 1,2-PD was seen. The maximum OD_{600} was between 0.85 and 1.0 for all concentrations tested. This was less than the maximum OD_{600} of 1.6 observed when grown with 0.05 % (w/v) succinate (Fig. 3.28). Growth was also slower with 1,2-PD; the maximum growth only being reached after 72 hrs whereas with 0.05% (w/v) succinate maximum growth was seen at 48 hr.

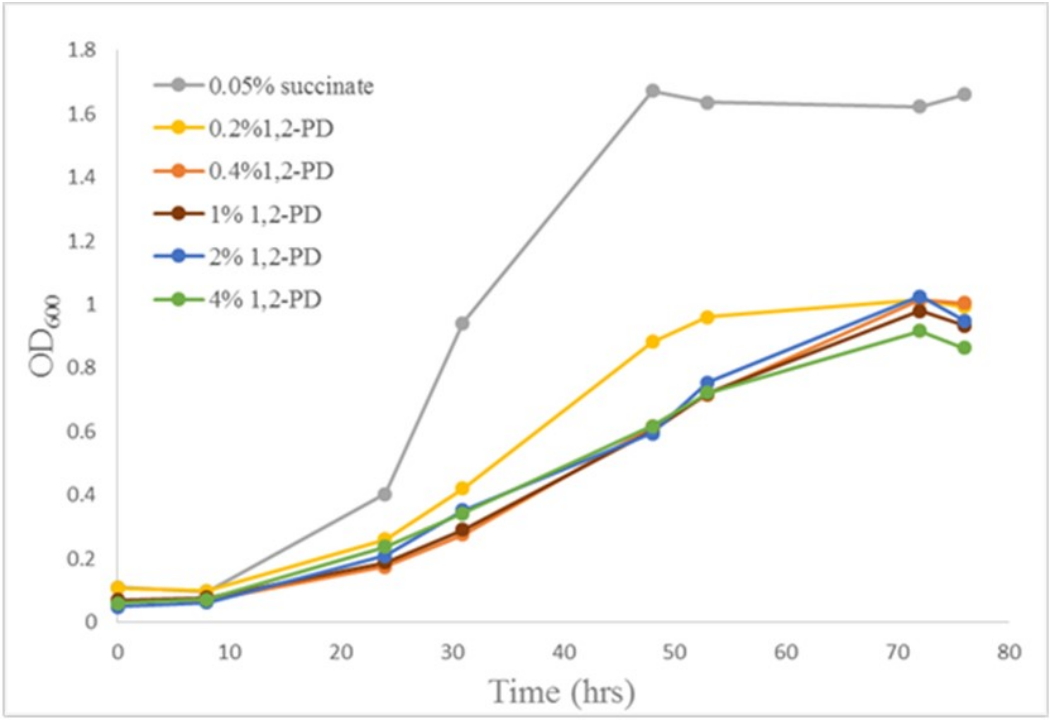


Figure 3.28. Anaerobic growth of *R. rubrum* with the following single carbon source: (i) 0.05% (w/v) sodium succinate; (ii) 0.2% (v/v) 1,2-PD; (iii) 0.4% (v/v) 1,2-PD; (iv) 1% (v/v) 1,2-PD; (v) 2% (v/v) 1,2-PD and (vi) 4% (v/v) 1,2-PD.

A combination of 0.05 % (w/v) succinate with each of the 1,2-PD concentrations increased the maximum OD_{600} to between 2.6 and 3.0. However increased concentrations of 1,2-PD did not result in an increased maximum OD_{600}, with similar growth being observed for all concentrations tested (Fig 3.29). Growth appears slower with 1,2-PD present in the media, the OD_{600} reaching a maximum between 54 and 72 hr growth compared to 48 hr with succinate alone.

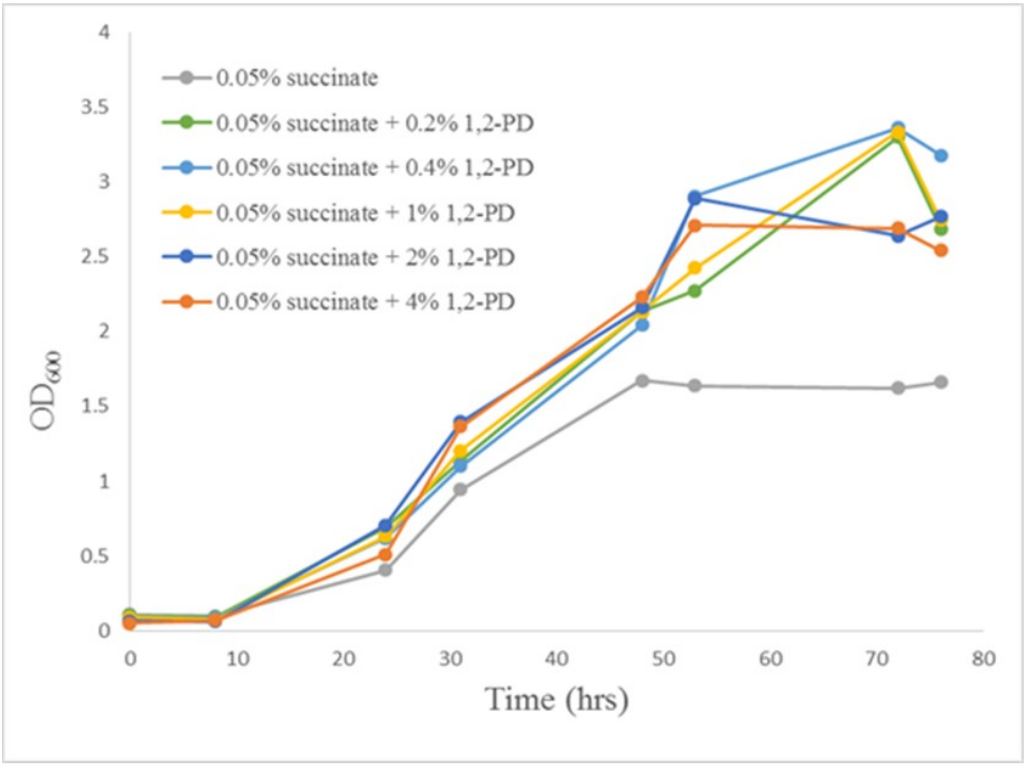


Figure 3.29. Anaerobic growth of *R. rubrum* with the following carbon sources: (i) 0.05 % succinate; (ii) 0.05 % (w/v) succinate + 0.2 % (v/v) 1,2-PD; (iii) 0.05 % (w/v) succinate + 0.4 % (v/v) 1,2-PD; (iv) 0.05 % (w/v) succinate + 1 % (v/v) 1,2-PD (v) 0.05 % (w/v) succinate + 2 % (v/v) 1,2-PD; (vi) 0.05 % (w/v) succinate + 4 % (v/v) 1,2-PD

As with aerobic growth, *R. rubrum* was able to grow anaerobically with either succinate or 1,2-PD as a carbon source. However, growth on 1,2-PD was substantially reduced compared to growth on succinate. A combination of succinate and 1,2-PD resulted in greater growth than with either carbon source alone. Subsequent studies therefore examined the resulting phenotypes following anaerobic growth on either succinate or 1,2-PD as a sole carbon source or with a combination of the two.

3.3.2 Transmission Electron Microscopy

3.3.2.1 Visualisation of *R. rubrum* phenotypes following aerobic growth on sodium succinate and/or 1,2-PD at a range of concentrations.

Cultures of *R. rubrum* (50 mL) were grown aerobically in 250 mL baffled flasks in media containing either (i) 0.2 % (w/v) succinate; (ii) 0.05 % (w/v) succinate; (iii) 0.2 % (v/v) 1,2-PD (iv) 0.4 % (v/v) 1,2-PD; (v) 0.05 % (wv) succinate + 0.2 % (v/v) 1,2-PD and (vi) 0.05 % (w/v) succinate + 0.4 % (v/v) 1,2-PD. Cultures were inoculated to a starting OD_{600} = 0.05 from a starter culture that had been centrifuged and re-suspended in culture media without any carbon source. Cultures were harvested by centrifugation 24 h after inoculation. Cell pellets were fixed and embedded as described in Section 2.4.8.1. Briefly, cell pellets were fixed in

2.5 % (w/v) glutaraldehyde before staining with osmium tetroxide. The pellets were then dehydrated with a series of increasing ethanol concentrations before washing with propylene oxide. Cells were incubated in resin for a total of 3 hours before final resuspension in resin and overnight polymerisation.

Ultrathin sections (70 nm) of embedded samples were collected on copper mesh grids and stained with uranyl acetate and Reynolds lead citrate as described in Section 2.4.8.2 and 2.4.8.3 before visualisation on a Jeol 1230 Transmission Electron Microscope.

3.3.2.1.1 Aerobic growth on 0.2% (w/v) sodium succinate

0.2% (w/v) succinate is the standard carbon source in the *R. rubrum* media used in this study (Section 2.1 5). Under aerobic growth conditions, cells grown with 0.2% (w/v) succinate as the sole carbon source had the highest OD_{600} (~1.0) upon harvest at 24 hrs post-inoculation. Structures resembling microcompartments were not seen in any of the cells. Heavily stained structures, are observed in some cells (Fig. 3.30). It is not known what these bodies are but they may be composed of lipid as the stains used are lipophilic. Glycogen granules, which form as organic carbon reserves when *R. rubrum* is grown on succinate (Cohen-Bazire, 1963), are evident in some cells – these appear as white areas within the cytoplasm that are not bounded by a membrane. Many cells appear to have large periplasmic spaces perhaps indicating that the cells are not healthy. This may be due to the culture being in late stationary phase at harvest.

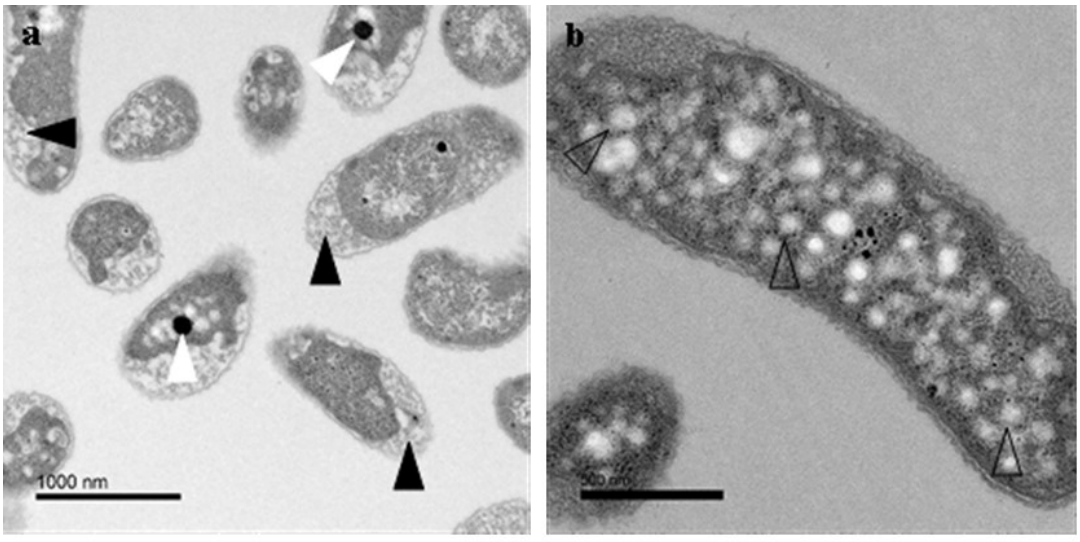


Figure 3.30. TEM images of *R. rubrum* grown aerobically on 0.2 % (w/v) sodium succinate. Heavily stained bodies (white arrows), glycogen granules (open arrows) and large periplasmic spaces (black arrows) are indicated. Scale bars: a) 1000 nm; b) 500 nm

3.3.2.1.2 Aerobic growth on 0.05 % (w/v) sodium succinate

Cultures reached an OD_{600} ~ 0.6 after 24 hr growth on 0.05 % (w/v) sodium succinate. TEM images of *R. rubrum* when grown aerobically on 0.05 % (w/v) sodium succinate appear to have a similar phenotype to those grown at a higher succinate concentration. Glycogen granules are evident in many cells but, as with the cells grown at higher succinate concentrations, many cells appear unhealthy with large periplasmic spaces (Fig. 3.31). No microcompartment-like structures are observed.

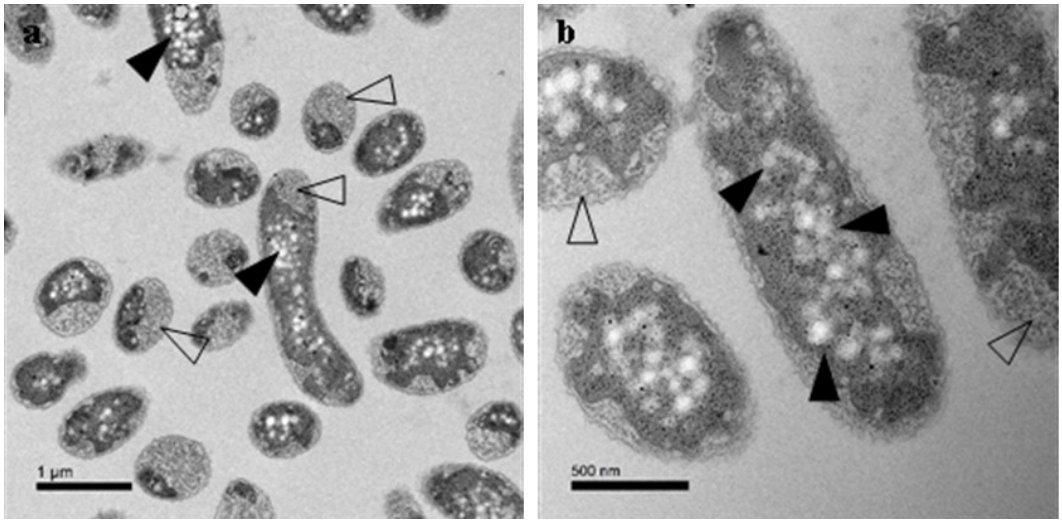


Figure 3.31. TEM images of *R. rubrum* cells grown aerobically on 0.05 % (w/v) sodium succinate. Glycogen granules (black arrows) and large periplasmic spaces (open arrows) are indicated. Scale bars: a) 1 µm; b) 500 nm

3.3.2.1.3 Aerobic growth on 0.2 % (v/v) 1,2-PD

After 24 h growth on 1,2-PD *R. rubrum* cultures reached an OD_{600} ~ 0.4. TEM images show glycogen granules to be evident, as seen in cells grown on succinate. Many cells also have small electron dense inclusions (Fig. 3.32). These inclusions are present in 44 % of cells and measure 57.6 ± 9.3 nm (n = 50) in diameter. They are possibly polyphosphate granules which can form in bacteria under stress conditions (Brown and Kornberg, 2004). Whilst some of these inclusions appear to have a polyhedral shape, most are circular. Both the size and shape of these inclusions make them unlikely to be microcompartments. The size is more consistent with the *R. rubrum* polyphosphate granule measurements of 500 Å previously reported (Cohen-Bazire, 1963).

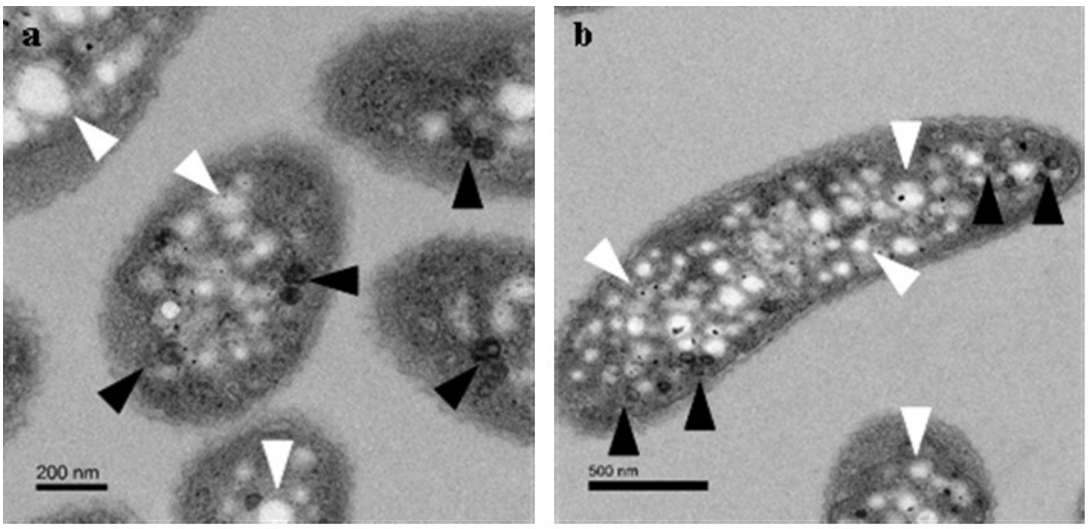


Figure 3.32. TEM images of *R. rubrum* grown aerobically on 0.2 % (v/v) 1,2 PD. Examples of glycogen granules (white arrows) and polyphosphate granules (black arrows) are indicated. Scale bars a) 200 nm; b) 500 nm

3.3.2.1.4 Aerobic growth on 0.4 % (v/v) 1,2-PD

Following 24 hr growth on 0.4 % (v/v) 1,2 PD, *R. rubrum* reached an OD_{600} ~ 0.4, similar to the growth seen with 0.2 % (v/v) 1,2-PD. TEM images show a similar phenotype with cells grown on this higher concentration of 1,2-PD (Fig. 3.33). Electron dense possible polyphosphate bodies are observed in 48 % of cells measuring 61.5 ± 8.7 nm (n=38) in diameter. There is no evidence of microcompartment formation under these growth conditions.

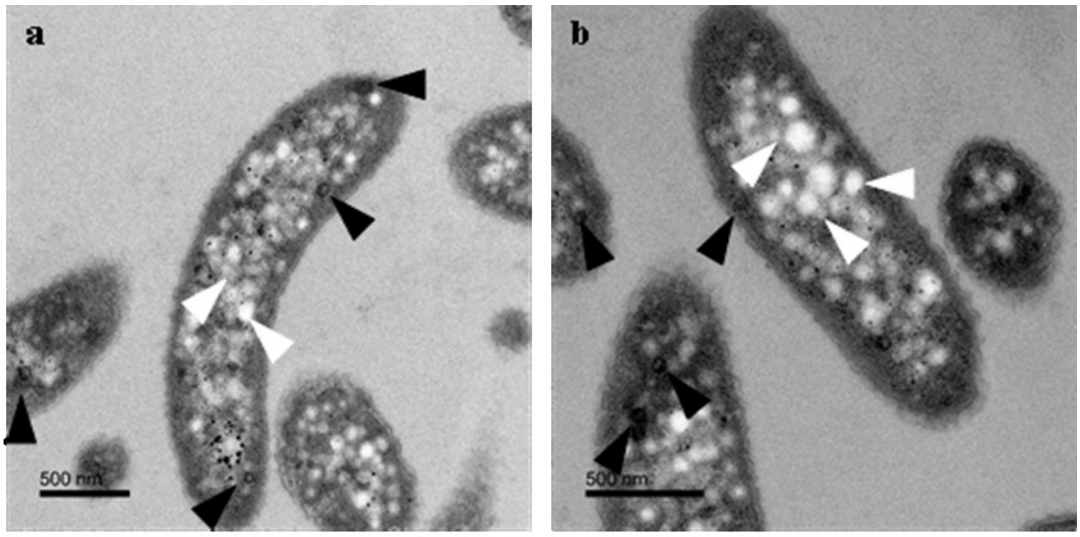


Figure 3.33. TEM images of *R. rubrum* grown aerobically on 0.4 % (v/v) 1,2-PD. Polyphosphate bodies (black arrows) and glycogen granules (white arrows) are indicated. Scale bars 500 nm.

3.3.2.1.5 Aerobic growth on 0.05 % (w/v) sodium succinate and 0.2 % (v/v) 1,2-PD

R. rubrum cultures grown aerobically on 0.05 % (w/v) sodium succinate and 0.2 % (v/v) 1,2-PD reached an OD_{600} ~ 0.7 after 24 hr. TEM images of sectioned cells reveal no evidence of any microcompartment-like structures. Images instead show there to be glycogen and electron dense polyphosphate granules present as has been seen in other aerobically grown *R. rubrum* cultures. Many cells appear have areas of expanded periplasm (Fig. 3.34) perhaps indicating the cells are not healthy. This appears to be a consistent effect of cultures grown with sodium succinate.

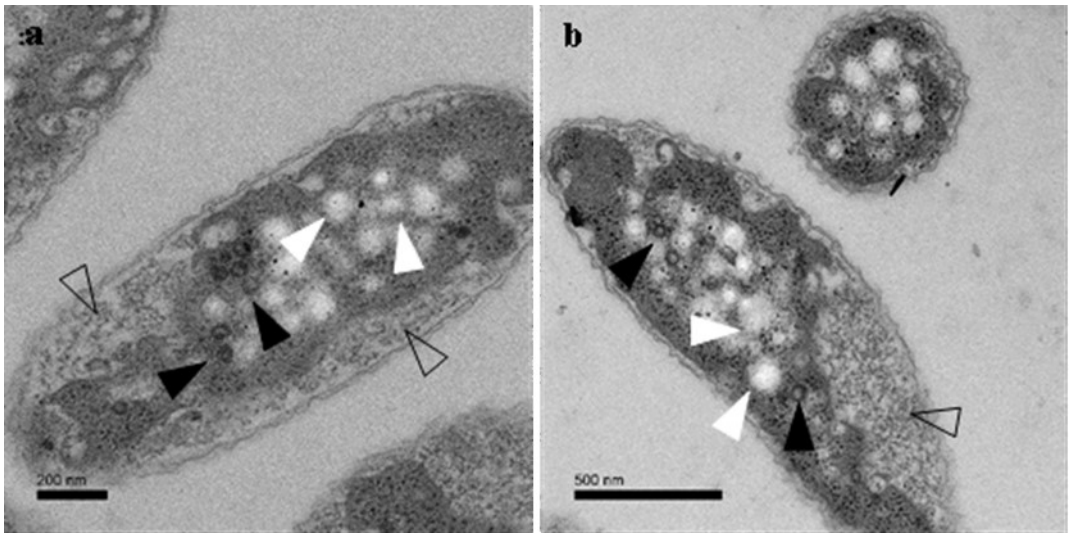


Figure 3.34. TEM images of *R. rubrum* grown aerobically on 0.05 % (w/v) sodium succinate and 0.2% (v/v) 1,2-PD. Polyphosphate bodies (black arrows), glycogen granules (white arrows) and large areas of periplasm (open arrows) are indicated. Scale bars a) 200 nm; b) 500 nm.

3.3.2.1.6 Aerobic growth on 0.05 % (w/v) sodium succinate and 0.4 % (v/v) 1,2-PD

Cultures of *R. rubrum* grown aerobically on 0.05 % (w/v) sodium succinate and 0.4 % (v/v) 1,2-PD reached an OD_{600} ~ 0.7 after 24 hrs. The phenotype observed for these cells is very similar to that observed for cells grown on 0.05 % (w/v) succinate and 0.2 % (v/v) 1,2-PD. The higher concentration of 1,2-PD appears to have no effect on the phenotype observed. As seen previously, glycogen storage granules and polyphosphate granules are evident and many cells appear unhealthy with large areas of periplasm present (Fig. 3.35).

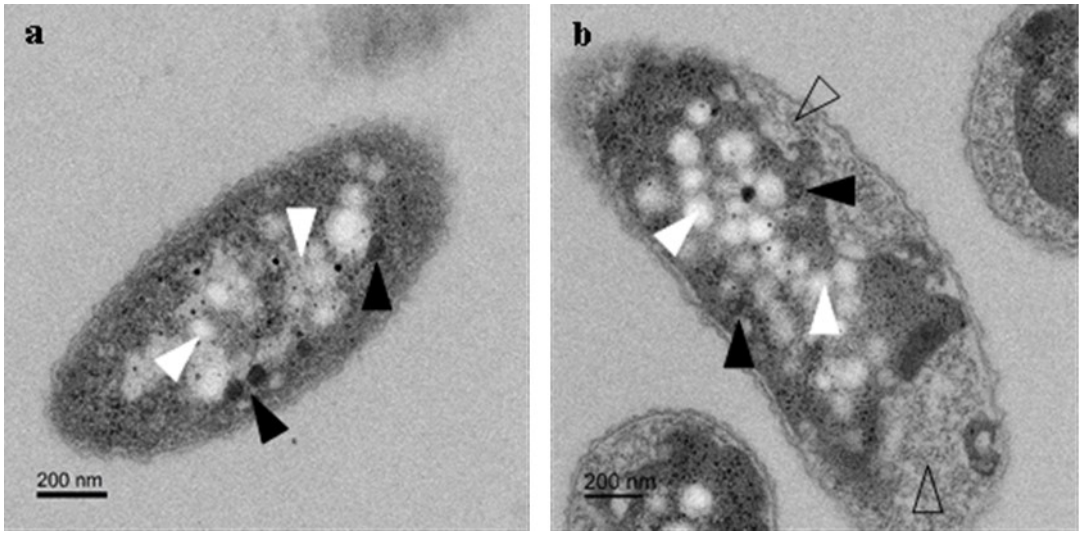


Figure 3.35. TEM images of *R. rubrum* grown aerobically on 0.05 % (w/v) sodium succinate and 0.4 % (v/v) 1,2-PD. Polyphosphate bodies (black arrows), glycogen storage granules (white arrows) and large areas of periplasm (open arrows) are indicated. Scale bars 200 nm.

3.3.2.2 Visualisation of *R. rubrum* phenotypes following anaerobic growth on sodium succinate and/or 1,2-PD at varied concentrations.

Cultures of *R. rubrum* (75 mL) were grown anaerobically in the light in sealed 100 mL anaerobic glass vessels in media containing either (i) 0.2 % (w/v) succinate; (ii) 0.05 % (w/v) succinate; (iii) 0.2 % (v/v) 1,2-PD (iv) 0.4 % (v/v) 1,2-PD; (v) 0.05 % (w/v) succinate + 0.2 % (v/v) 1,2-PD and (vi) 0.05 % (w/v) succinate + 0.4 % (v/v) 1,2-PD. Cultures were inoculated to a starting OD_{600} of ~0.1 from an aerobically grown starter culture that had been centrifuged and re-suspended in culture media without any carbon source. Cultures were harvested 48 h post-inoculation and samples were fixed and embedded as described in Section 2.4.8.1. Ultrathin sections (70 nm) were collected on copper mesh grids and stained for visualisation as described in Sections 2.4.8.2 and 2.4.8.3.

3.3.2.2.1 Anaerobic growth on 0.2 % (w/v) sodium succinate

Sodium succinate (0.2 % (w/v)) is the carbon source in the standard media for *R. rubrum* growth. After 48 hrs of anaerobic growth, cultures reached an OD_{600} ~ 2.2. Chromatophores, membrane bound structures containing the photosynthetic pigments bacteriochlorophyll and carotenoids (Pardee, Schachman and Stanier, 1952) are clearly visible in these cells after anaerobic growth in the light (Fig 3.36). Glycogen granules are visible in many cells. These are distinguishable from the chromatophores as they appear as lighter areas not bounded by a membrane.

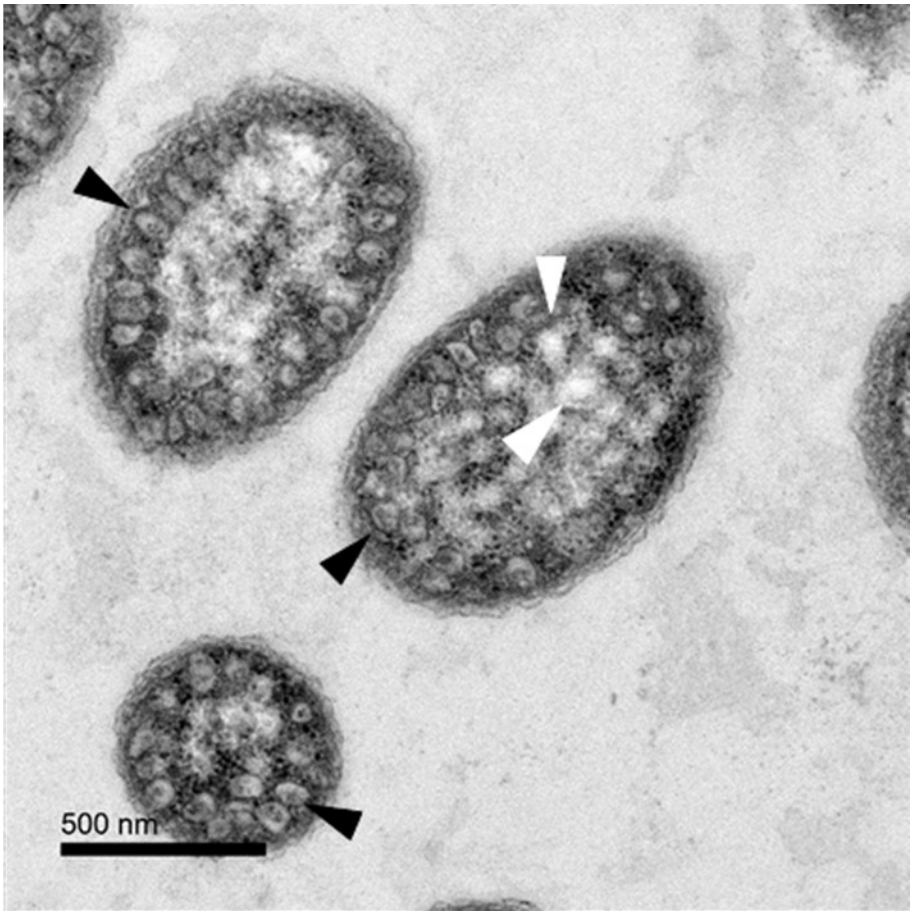


Figure 3.36. TEM image of thin section of *R. rubrum* cells grown anaerobically with 0.2 % (w/v) succinate. Chromatophores (black arrows) and glycogen granules (white arrows) are indicated. Scale bar 500 nm.

3.3.2.2.2 Anaerobic growth on 0.05 % (w/v) sodium succinate

R. rubrum cultures grown on 0.05 % (w/v) sodium succinate reached an OD_{600} ~ 1.4 after 48 hr of anaerobic growth. At this concentration of sodium succinate electron dense structures resembling microcompartments are observed in 10 % of cells. Many of these structures appeared to have a polyhedral form whilst others were less defined (Fig 3.37). These structures ranged from 59 – 111 nm in diameter with an average diameter of 82 ± 13 nm (n=46) (Fig. 3.38). Whilst this is small for a microcompartment, most studies measuring an average diameter of 100 – 200 nm, smaller compartments have been reported (Erbilgin, McDonald and Kerfeld, 2014). These electron dense structures are predominantly observed as single entities within the cells, multiple structures only being observed in a small number of bacteria. This is also different to the electron dense structures putatively identified as polyphosphate granules that are seen when *R. rubrum* is grown aerobically (Section 3.3.2.1.3), as multiple polyphosphate bodies were observed in each cell. The observation of microcompartment-like structures is unexpected as the presence of 1,2-PD is thought to be necessary to induce the production of BMC proteins.

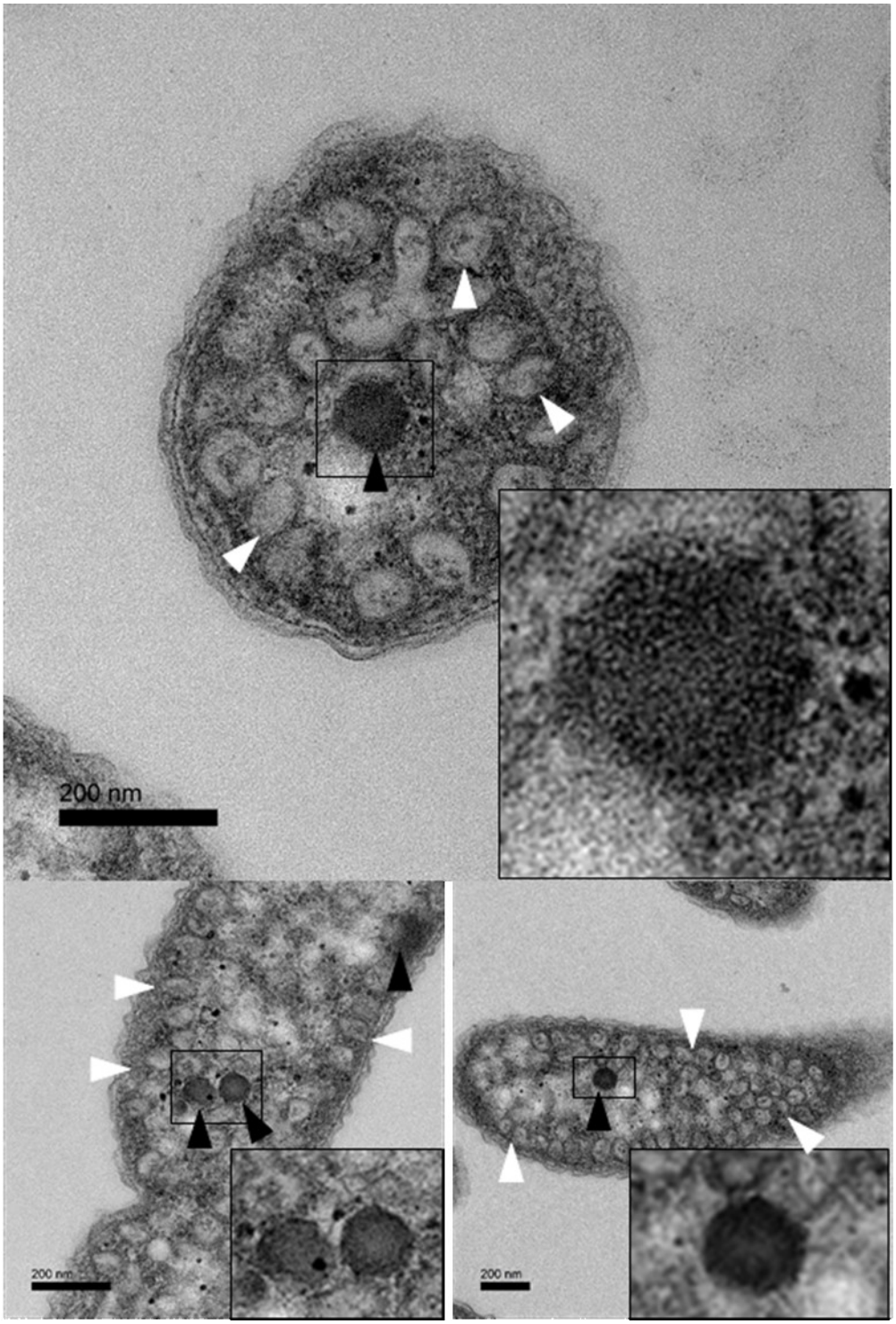


Figure 3.37. TEM images of *R. rubrum* grown anaerobically on 0.05 % (w/v) sodium succinate. Microcompartment-like electron dense structures (black arrows) and chromatophores (white arrows) are indicated. Insets show enlargements of the areas highlighted in the main image. Scale bars 200 nm.

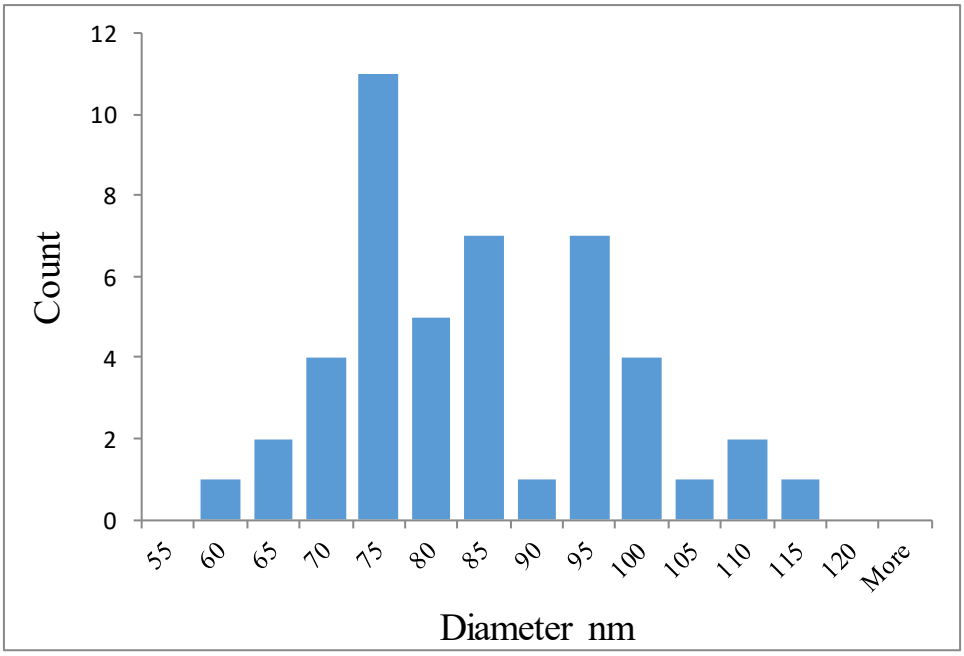


Figure 3.38. Histogram showing the diameters of the microcompartment-like structures observed in *R. rubrum* grown anaerobically on 0.05 % (w/v) sodium succinate, n=46.

If these structures are microcompartments, it is unclear why they would be formed under these conditions. It may be that, as the normal source of carbon becomes depleted, microcompartments are formed as a strategy to try and utilise any other resources that may be within the environment as an alternative source of energy or it may simply be that under stress conditions compartments are formed as a survival strategy.

3.3.2.2.3 Anaerobic growth on 0.2 % (v/v) 1,2-PD

Following 48 hr of anaerobic growth on 0.2 % (v/v) 1,2-PD *R. rubrum* cultures reached an OD_{600} ~ 0.7. Electron dense microcompartment-like structures were observed in 19 % of bacteria with an average diameter of 84.6 ± 16 nm (n= 49) ranging from 50 – 115 nm (Fig. 3.39). The diameter of the compartments is similar to that observed for cells grown on 0.05 % (w/v) sodium succinate but the number of cells observed with microcompartment-like structures was slightly higher. Most compartments were seen as single structures within the bacteria, although some cells did contain multiple structures. Multiple polyhedral in the cytoplasm of bacteria grown on 1,2-PD are generally observed (Shively *et al.*, 1998; Bobik *et al.*, 1999), therefore the predominance of only single structures within each cell is unexpected.

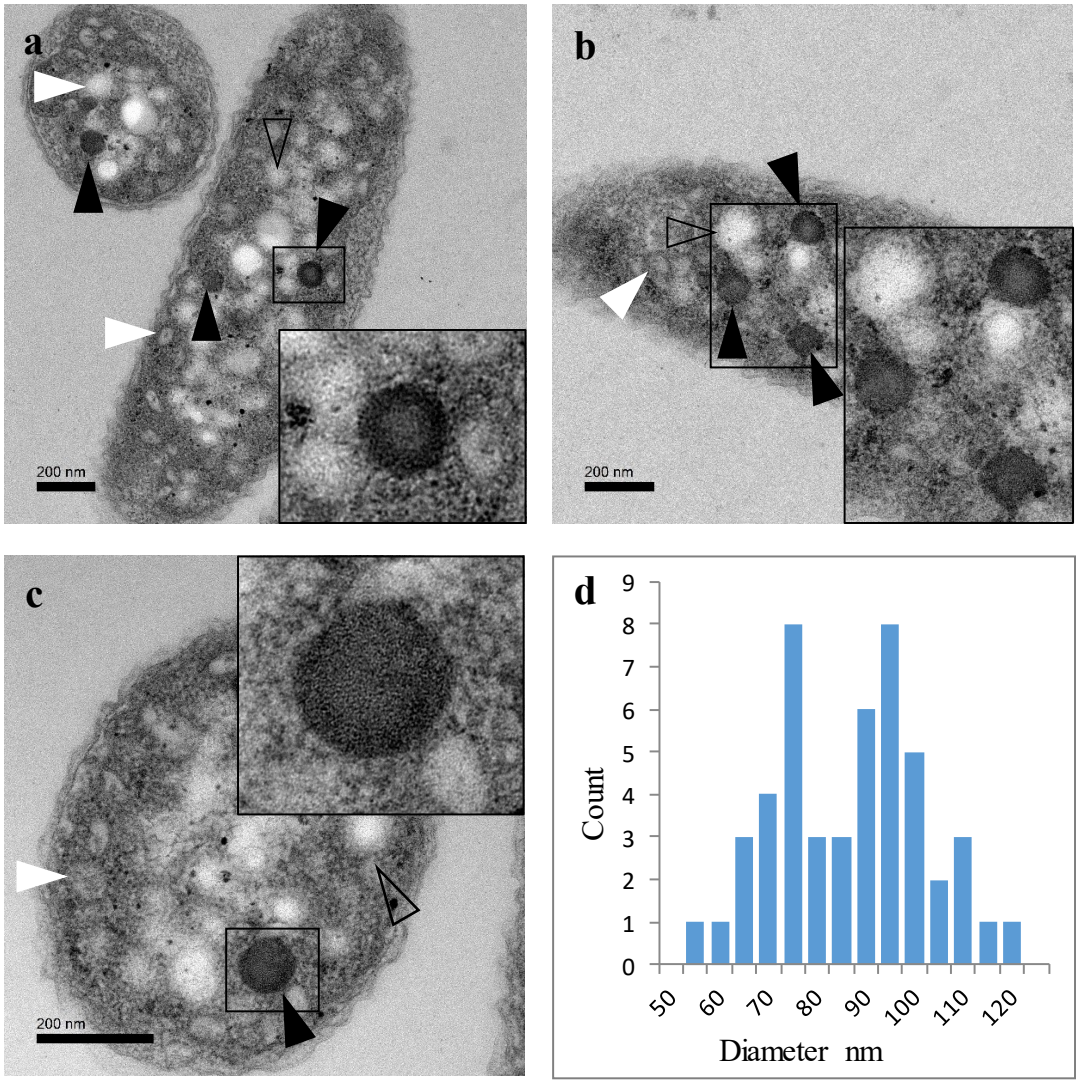


Figure 3.39. (a-c) TEM images of *R. rubrum* grown anaerobically on 0.2 % (v/v) 1,2-PD and d) A histogram showing the diameter of the microcompartment-like structures (n=49). Microcompartment-like electron dense structures (black arrows), chromatophores (white arrows) and glycogen storage granules (open arrows) are indicated. Insets show enlargements of the areas highlighted in the main image. Scale bars 200 nm.

3.3.2.2.4 Anaerobic growth on 0.4 % (v/v) 1,2-PD

Anaerobic cultures of *R. rubrum* grown on 0.4 % (v/v) 1,2-PD reached an OD_{600} ~0.7 after 48 hrs. Electron dense microcompartment-like structures were observed in 22 % of cells (Fig. 3.40). The diameter of these structures ranged from 65 – 112 nm with an average diameter of 89.7 ± 11 nm (n=28). Again, the majority of cells contained just 1 – 2 structures. The diameters and number of cells containing structures are very similar to those observed for *R. rubrum* grown on 0.2 % (v/v) 1, 2-PD.

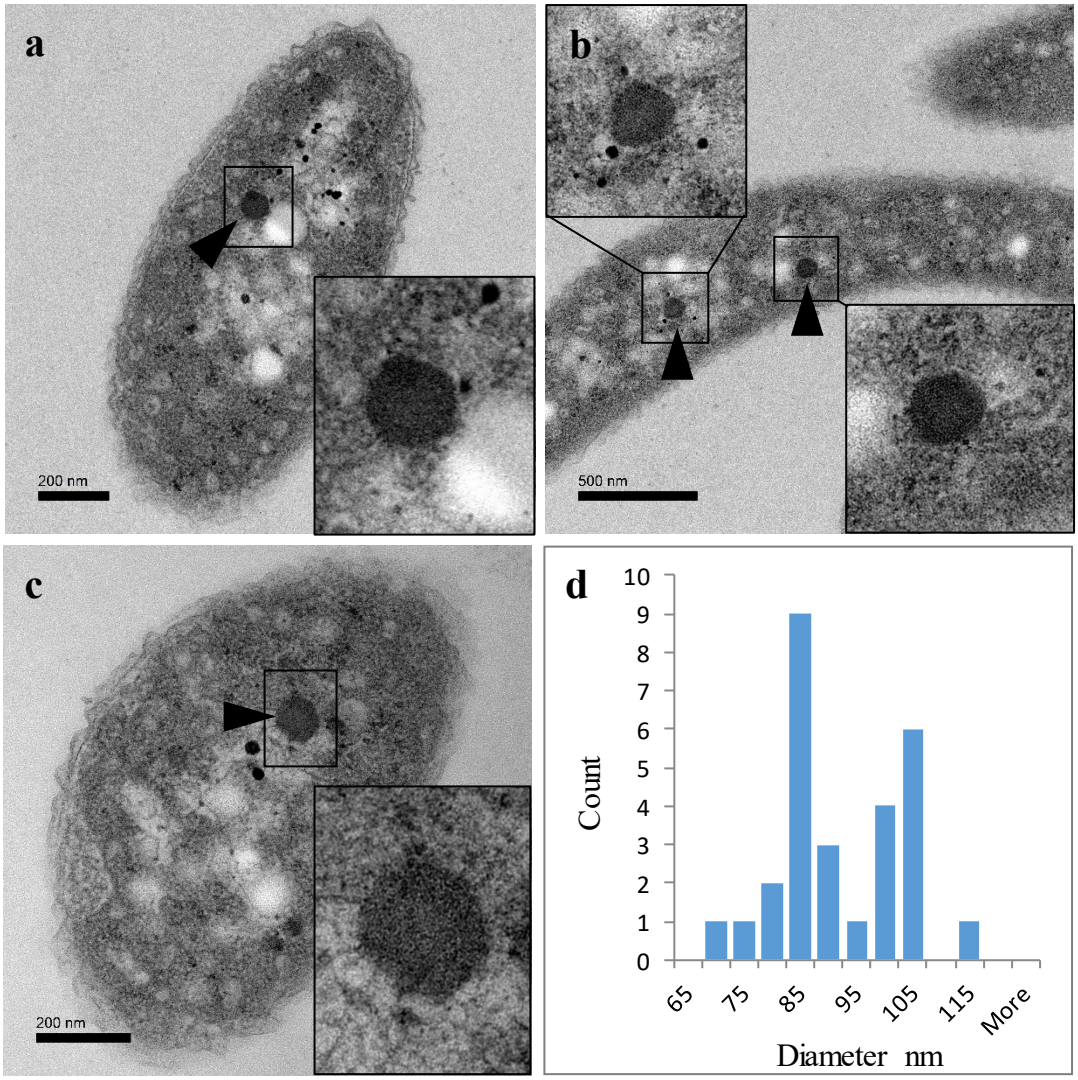


Figure 3.40. (a –c) TEM images of *R. rubrum* grown anaerobically on 0.4 % (v/v) 1,2-PD and (d) A histogram showing the diameter of microcompartment-like structures (n=28). Electron dense microcompartment-like structures are indicated (black arrows). Insets show enlargements of areas indicated on the main image. Scale bars (a and c) 200 nm (b) 500 nm

3.3.2.2.5 Anaerobic growth on 0.05 %(w/v) sodium succinate and 0.2 % (v/v) 1,2-PD

R. rubrum cultures grown on 0.05% (w/v) sodium succinate and 0.2% (v/v) 1,2-PD reached an OD_{600} ~ 0.8 after 48 hr anaerobic growth. TEM images revealed 18.5 % of cells contained electron dense structures resembling microcompartments (Fig. 3.41). The diameter of these structures ranged from 48 to 113 nm with an average diameter = 84.0 ± 15 nm (n= 64).

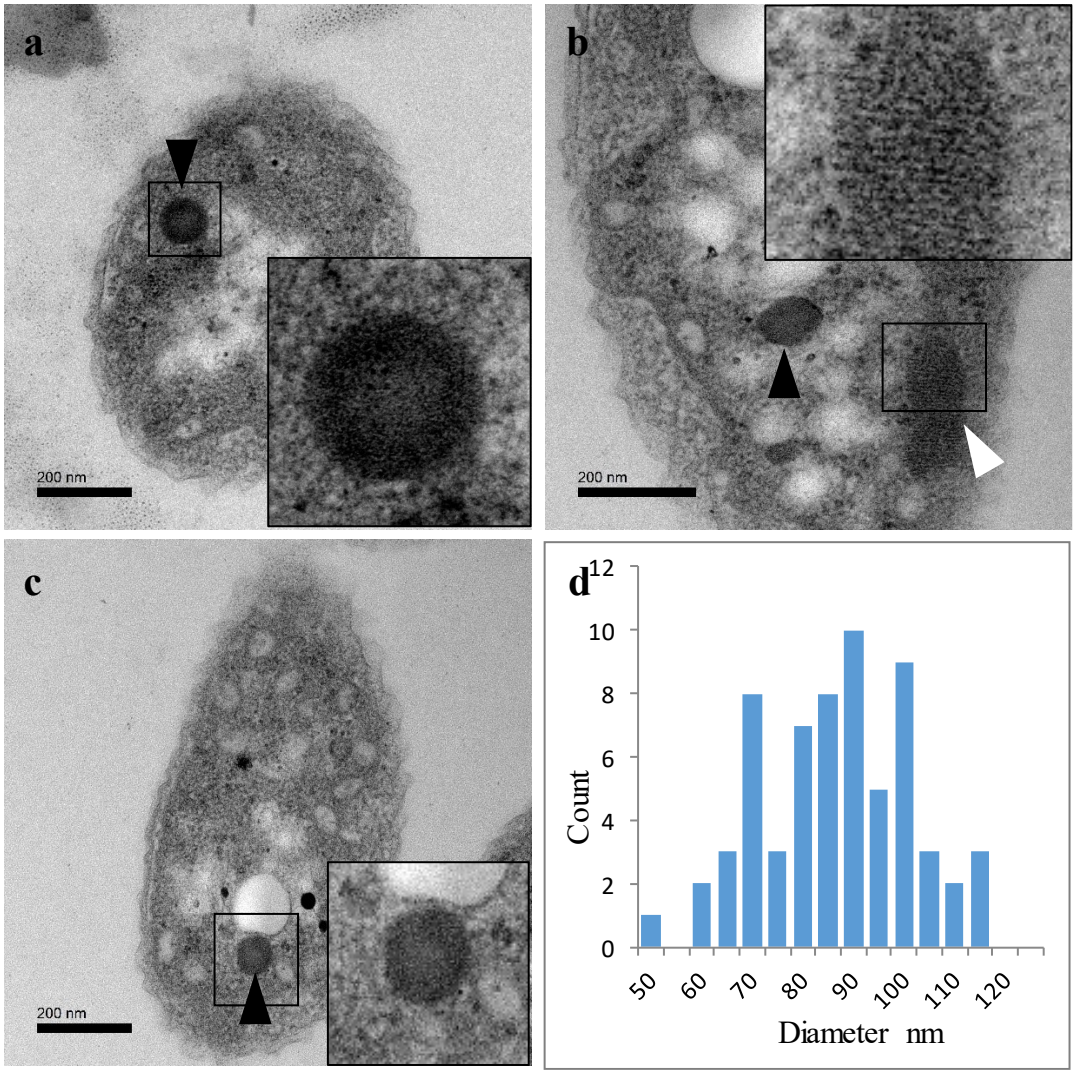


Figure 3.41. (a –c) TEM images of *R. rubrum* grown anaerobically on 0.05 % (w/v) sodium succinate and 0.2 % (v/v) 1,2-PD and (d) A histogram showing the diameter of microcompartment-like structures (n=64). Electron dense microcompartment-like structures (black arrows) and striated inclusion (white arrow) are indicated. Insets show enlargements of areas indicated on the main image. Scale bars 200 nm

The cell shown in Fig. 3.41b also contains an electron dense inclusion that appears to be striated. Inclusions of this type were observed in a small number of cells (<1%) under all growth conditions tested. It is unknown what these structures are but the striations look similar to the thylakoid membranes in plant chloroplasts so these inclusions may be mis-formed photosynthetic membranes.

3.3.2.2.6 Anaerobic growth on 0.05 % (w/v) sodium succinate and 0.4 % (v/v) 1,2-PD

R. rubrum cultures grown anaerobically on 0.05 % (w/v) sodium succinate and 0.4 % (v/v) 1,2-PD reached an OD_{600} ~ 0.5 after 48 hr. 28 % of cells contained electron dense microcompartment like structures measuring 78.7 ± 15 nm (n=54) in diameter (Fig. 3.42).

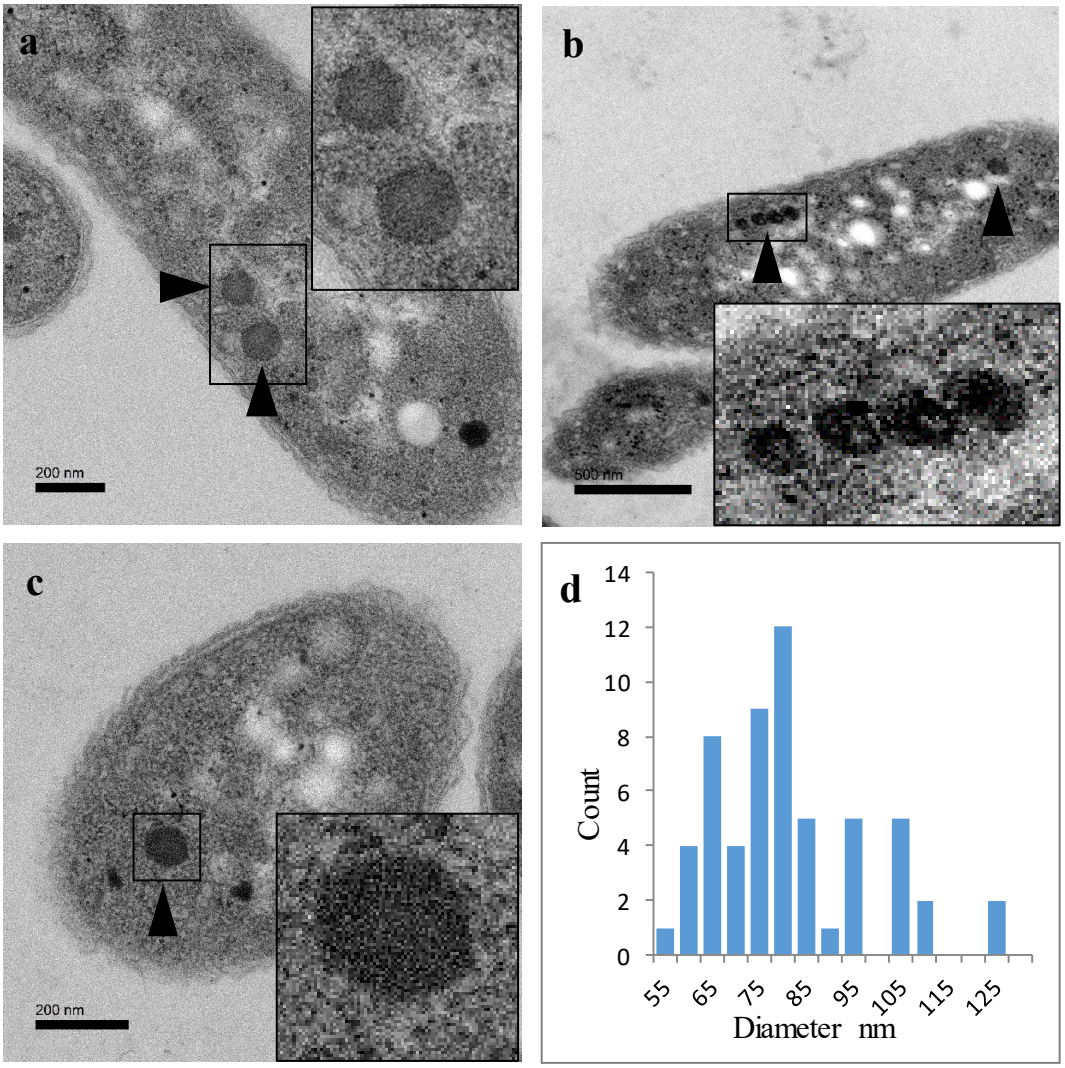


Figure 3.42. (a –c) TEM images of *R. rubrum* grown anaerobically on 0.05 % (w/v) sodium succinate and 0.4% (v/v) 1,2-PD and (d) A histogram showing the diameter of microcompartment-like structures (n=54). Electron dense microcompartment-like structures are indicated (black arrows). Insets show enlargements of areas indicated on the main image. Scale bars (a and c) 200 nm (b) 500 nm

3.3.2.3 Summary of phenotypes observed with aerobic and anaerobic growth on sodium succinate and/or 1,2-PD

When grown aerobically, *R. rubrum* did not form microcompartments with any of the carbon sources tested. Small electron dense inclusions were observed in many cells when 1,2-PD was included in the media but these were generally rounded in appearance and had a small diameter suggesting they were not microcompartments but were maybe polyphosphate granules or other inclusions. However, when grown anaerobically on a lower concentration of sodium succinate (0.05% (w/v)) larger electron dense structures resembling microcompartments were observed in approximately 10 % of *R. rubrum* cells. When 1,2-PD was present in the media, either with or without 0.05 % (w/v) succinate, the number of cells with structures resembling

microcompartments increased to approximately 20 %. The structures are small compared to most previously identified compartments however there are examples of smaller compartments reported. It is also possible that these structures are not compartments but instead are larger polyphosphate granules or other inclusion bodies. Further experiments are needed to determine the nature of these structures within *R. rubrum*.

3.3.3 SDS-PAGE of crude cell lysates

If the structures observed in anaerobically grown cells on 1,2-PD are microcompartments, the shell proteins and enzymes that comprise the BMC may be detectable by SDS-PAGE if they are expressed strongly enough. A sample of each culture was taken at stationary phase and centrifuged to pellet the cells. Cells were re-suspended in dH_2O and lysed by freeze-thawing. Samples were mixed 1:1 with Laemmli sample buffer and heated to 100°C for 10 min before loading on a 15 % (v/v) polyacrylamide gel.

Comparison of the protein bands obtained with the aerobically grown cultures with protein bands from the anaerobically grown cultures do not reveal many obvious differences (Fig. 3.43).

If the BMC proteins were produced in the anaerobically grown cells, extra bands would be expected at the bottom of the gel where the hexameric shell proteins would be seen (~10 kDa) and these are not apparent. The only band that appears to be strongly produced in the anaerobic cultures that is not seen in the aerobic cultures is between the 46 and 58 kDa marker. The aldehyde dehydrogenase (Rru_A0917) is a 55 kDa protein therefore would be expected to be at about this position but if this enzyme is strongly produced it would be anticipated that the other enzymes would also be visible. Additionally, this extra band is also seen in Lane 1 containing an extract from cells grown with 0.2 % (w/v) sodium succinate as a carbon source. These cells did not contain any microcompartment-like structures when imaged by TEM. This would suggest this protein band is not compartment-related but is a protein produced by *R. rubrum* when grown anaerobically, perhaps a membrane protein related to chromatophore production.

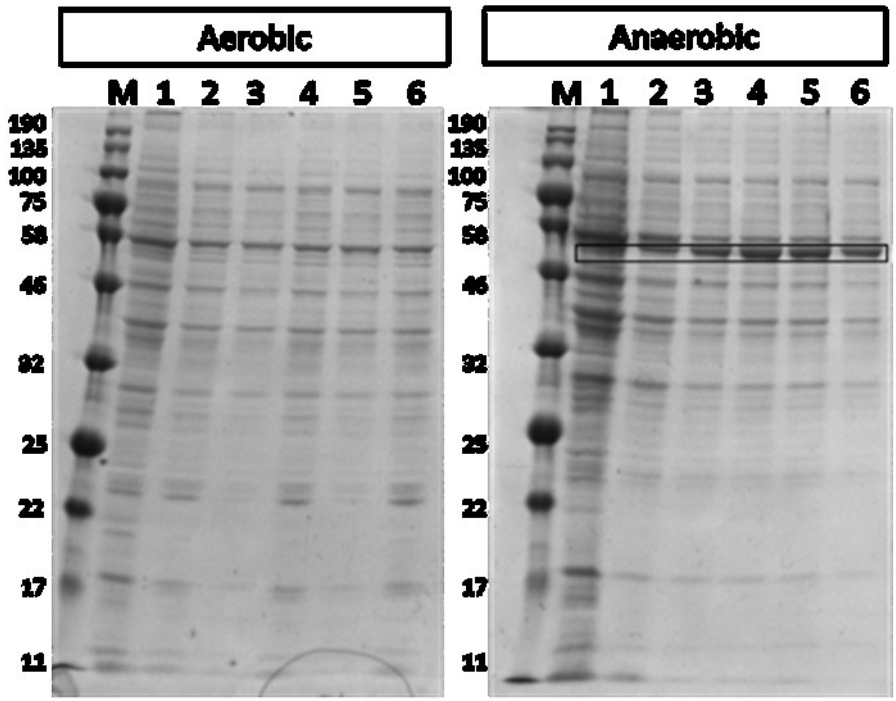


Figure 3.43. SDS-PAGE of crude cell lysates of aerobically and anaerobically grown *R. rubrum*. Lanes: M) MW marker; 1) 0.2 % (w/v) sodium succinate; 2) 0.05 % (w/v) sodium succinate; 3) 0.2 % (v/v) 1,2-PD; 4) 0.4 % (v/v) 1,2-PD; 5) 0.05 % (w/v) sodium succinate + 0.2 % (v/v) 1,2-PD; 6) 0.05 %(w/v) sodium succinate + 0.4 % (v/v) 1,2-PD. Box highlights extra band at ~55 kDa not seen in aerobic samples.

3.3.4 Western Blot Analysis

BMC proteins were not detectable by SDS-PAGE therefore detection by western blot was attempted using an antibody against residues 58 – 74 of *C. freundii* PduA. Rru_A0905 and Rru_A0908 are the two *R. rubrum* shell proteins most closely aligned with PduA. There is 65 % and 53 % identity between the residues of Rru_A0905 and Rru_A0908 and the antibody recognition sites (Fig. 3.44), therefore they would possibly also be recognised by the antibody. Initial studies examined antibody recognition for each of the *R. rubrum* shell proteins. BL21*(DE3) pLysS *E. coli* were transformed with plasmids containing one of the 6 *R. rubrum* shell protein genes. Cultures were induced with IPTG after reaching an OD_{600} ~ 0.6 and grown overnight at 19°C before harvesting. Cell pellets were re-suspended in dH_2O before freezing to aid lysis. Laemmli sample buffer was added to thawed cell suspensions and heated at 100°C before loading on a 15 % (v/v) polyacrylamide gel. A western blot was performed as described in Section 2.4.7. A sample of whole lysate of *E. coli* overproducing PduA was included as a positive control.

The bands for the individual shell proteins were identified by SDS-PAGE (Fig. 3.45a) and a duplicate gel was transferred for Western blot analysis.

```
                                                                58
                                                                ↓
PduA   MQQEALGMVETKGLTAAIEAADAMVKSANVMLVGYEKIGSGLVTVIVRGDVGAVKAATDA   60
Rru905 MQQEALGMVETKGLVGAIEAADAMVKSANVVLMGYEKIGSGLVTVMVRGDVGAVKAATDA   60
Rru908 MSGEALGMVETKGLIGSIEAADAMTKSANVTLIGYEKIGSGLVTTLVRGDVGAVKAAVDA   60
*. ********** .:*******.***** *:**********.:**********.**

                                  74
                                  ↓
PduA   GAAAARNVGEVKAVHVIPRPHTDVEKILPKGIS-----   93
Rru905 GAVAASKIGEVVSVHVIPRPHTEVEKILPRGLSASAAQ        98
Rru908 GAAAAEKVGTLVSKHIIPRPHSDVERILPHLG------   92
     **.** ::* : : *:*****.:**:***:
```

Figure 3.44. Protein sequence alignment of PduA and Rru_A0905 and Rru_A0908. The region recognised by antibody PduA58-74 is underlined and highlighted in yellow with identical residues in Rru_A0905 and A0908 also highlighted in yellow. Sequence alignment was performed with Clustal MUSCLE (Edgar, 2004).

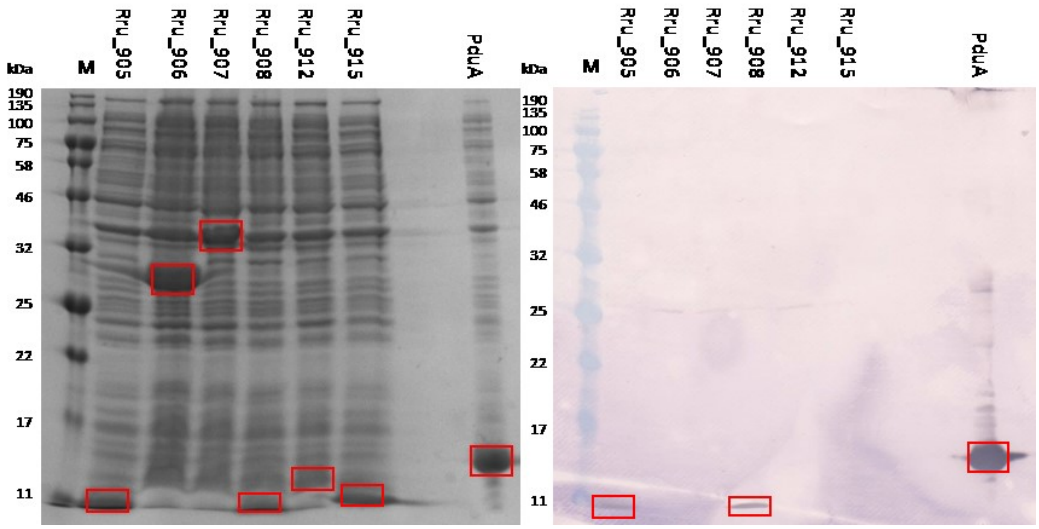


Figure 3.45. a) SDS-PAGE of whole cell lysates of E. coli over-producing R. rubrum shell proteins Rru_A0905, Rru_A0906, Rru_A0907, Rru_A0908, Rru_A0912 and Rru_A0915. E. coli overproducing Pdu A was included as a positive control. Proteins bands are highlighted with a red box. b) Western blot of whole cell lysates using primary antibody against PduA58–74. Recognition by the antibody is highlighted with a red box. Lanes are labelled above (M: molecular mass marker)

Both Rru_A0905 and Rru_A0908 are detectable by Western blot using an anti- PduA58-74 (Fig. 3.45b). The bands detected are not strong compared to the band obtained for the positive control PduA however the SDS-PAGE gel also shows there to be less protein in the *R. rubrum* shell protein samples than in the PduA sample.

75 mL cultures of *R. rubrum* were grown anaerobically on 0.05 % (w/v) sodium succinate or 0.05 % (w/v) sodium succinate + 0.2 % (v/v) 1,2-PD as described in Section 2.2.9.3. After 48 h, cultures were centrifuged at 2700 x g for 10 min and the pellet washed twice in 20 mM Tris-HCl buffer, pH 7.5. Cells were lysed by resuspension in 1 % (v/v) n-Octyl-Beta-D-thioglucopyranoside, supplemented with 25 mg lysozyme and 5 µL benzonase, and incubated at room temperature for 30 min. The lysate was centrifuged at 12000 x g for 5 min and samples of the supernatant and pellet taken for SDS-PAGE and western blot analysis. A whole cell lysate sample of *E. coli* producing PduA was included as a positive control. The PduA was detected by the anti-PduA58-74 antibody but nothing was detected in the supernatant or pellet samples from anaerobically grown *R. rubrum* (Fig. 3.46).

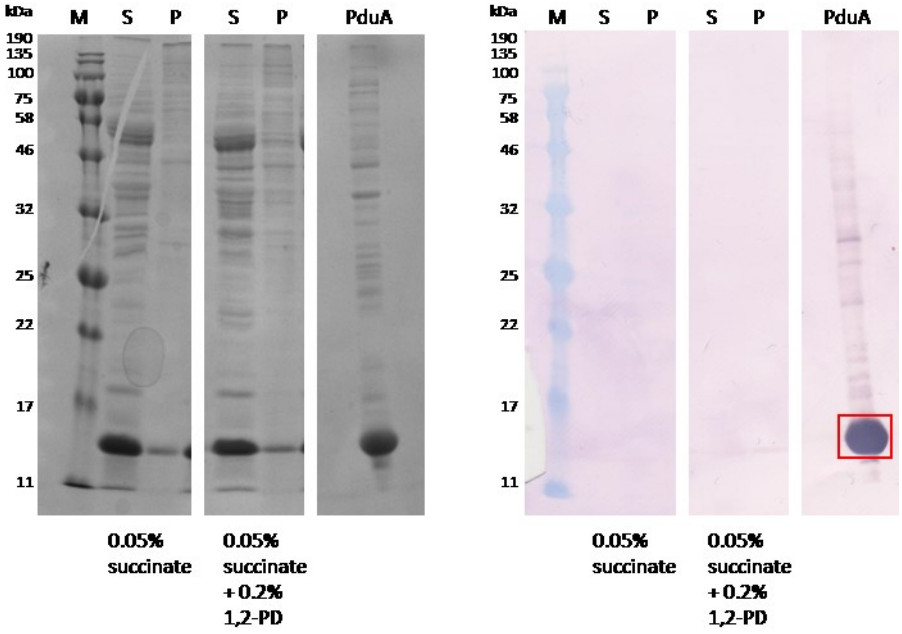


Figure 3.46. a) SDS-PAGE of the supernatant (S) and pellet (P) samples of lysed *R. rubrum* cells grown on 0.05 % (w/v) succinate or 0.05 % (w/v) succinate + 0.2 % (v/v) 1,2-PD. b) Western Blot of the supernatant (S) and pellet (P) samples of lysed *R. rubrum* cells grown on 0.05 % (w/v) succinate or 0.05 % (w/v) succinate + 0.2 % (v/v) 1,2-PD. PduA included as a positive control.

Neither of the shell proteins Rru_A0905 or Rru_A0908 were detectable by western blot in *R. rubrum* cells grown anaerobically on 0.05 % (w/v) sodium succinate or 0.5 % (w/v) sodium succinate + 0.2 % (v/v) 1,2-PD, despite these being the growth conditions under which microcompartment-like structures were observed by TEM. As the proteins are not evident on the SDS-PAGE gel, if they are present they can only be there at a very low concentration which the antibody is not sensitive enough to detect. Future studies could use a larger culture volume re-suspended in smaller volumes to concentrate any proteins present to improve the chances of detection.

These results are unable to verify the structures observed in some of the anaerobically grown cells are microcompartments. Further studies will be needed in order to confirm this, perhaps immuno-EM or microcompartment purification studies.

3.4 Discussion

This study aimed to investigate whether the shell proteins of the *R. rubrum* GRE BMC operon were able to form recombinant empty microcompartments in *E. coli,* as had been achieved with the Pdu BMC of *C. freundii* (Parsons *et al.*, 2010), and whether BMC production could be induced in wild-type *R. rubrum* cells.

The structures observed in whole cells following production of all the six shell proteins did not have the appearance of empty BMCs. Most of the structures observed appeared to be swirled sheets of protein distributed throughout the cells. There were a very small number of structures that resembled closed structures but the presence of ribosomes within them suggested that they were not closed compartments but simply appeared so due the angle through which they were sectioned. Interestingly, upon imaging of the purified proteins, compartment-like structures were observed, both in fixed samples and thin-sectioned samples. This was an unexpected finding but could reflect the dynamic nature of compartment assembly. An elegant study visualising the formation of protein sheets from hexameric shell proteins by high speed AFM (Sutter *et al.*, 2016) demonstrated that hexamers were able to associate with and dissociate from a protein sheet in a continuous fashion over the time course of the recording. This clearly revealed the flexible nature of facet assembly and may explain the appearance of BMCs in purified samples despite their apparent absence *in vivo*. The conditions for purification (50 mM Tris-HCl pH 8.0; 500/50 mM KCl; 12.5/5 mM $MgCl_2$) may provide the right environment for the shell proteins to self-assemble and form empty shells. It is possible that within the confines of the cell, the high concentration of proteins produced overwhelmed the assembly system and disrupted compartment formation. A similar phenomenon was observed when attempting to form minimal shell protein compartments using one BMC-H, one BMC-P and one BMC-H_{ex} – swirled sheets were the predominant structural feature observed *in vivo* but purified samples showed evidence of BMC formation.

The expression of the shell protein genes in this study was under the control of a T7 promotor system. A more controllable system with the differential production of each component at varying stoichiometries may provide a means to *R. rubrum* recombinant BMC production *in vivo*.

In order to determine the conditions necessary for compartment formation, *R. rubrum* was grown both aerobically and anaerobically in the presence and absence of 1,2-PD which was predicted to be the substrate of the GRE encoded in the BMC operon. There was no evidence

of BMC formation under any of the aerobic conditions tested which is not unexpected due to the oxygen sensitive nature of glycyl radical enzymes. However, when grown anaerobically and with a low concentration of sodium succinate electron dense structures were observed in approximately 10 % of cells. The diameter (82 ± 13 nm), although at the smaller end of the range, and the shape of these structures was consistent with appearance of BMCs observed in other organisms. This was a surprising observation as formation of BMCs is usually induced by the presence of the signature enzyme substrate. In this case it is possible that the formation of BMCs is a survival response after the depletion of sodium succinate, perhaps compartments are formed as a strategy to utilise any resources that may be in the environment. 1,2-PD as the sole carbon source or as a supplement to 0.05 % (w/v) sodium succinate was seen to induce the formation of electron dense, polygonal structures in approximately 20 % of cells. They were of a similar size and shape to those observed with sodium succinate alone but were found in twice the number of cells. The increased number of bacteria with these electron dense structures when 1,2-PD was in the media would support the hypothesis that 1,2-PD is the substrate for this compartment. The natural habitat of *R. rubrum* in pond water and sewage would provide a ready source of 1,2-PD from the breakdown of the plant sugars fucose and rhamnose by other organisms. However, neither SDS-PAGE or Western blot analysis was able to identify any of the shell proteins in the lysates of *R. rubrum* grown on 1,2-PD but this may be due to the relatively small number of compartments present. Analysis of larger culture volumes may be more successful in identifying the presence of compartment proteins. Other methods may also be employed to facilitate a positive identification of these structures as microcompartments. Immuno-electron microscopy could perhaps be used with the anti-PduA58-74 as a primary antibody to see if the structures were labelled. The isolation of microcompartments using established methods could also be attempted with any isolated structures being imaged by TEM and possibly electron tomography. Whilst this study was unable to unequivocally identify the structures within *R. rubrum* as BMCs the evidence obtained so far is suggestive of their formation in the presence of 1,2-PD in an anaerobic environment.

Chapter 4

Characterisation of the *Clostridium autoethanogenum* GRE BMC shell proteins and evidence for ribosome interaction

4.1 Introduction

The glycyl radical enzyme associated microcompartments (GRMs) represent the largest class of BMC yet identified by bioinformatics (Axen *et al.*, 2014). These GRM loci have been identified in the genomes of a diverse array of bacterial species. Despite their apparent abundance, there is little experimental data available describing either BMC formation within the native organism or via heterologous expression of the relevant recombinant proteins. Chapter 3 discussed studies with the recombinantly produced shell proteins of the *R. rubrum* BMC. In this chapter studies examining the BMC shell proteins of another organism, *Clostridium autoethanogenum,* are reported and discussed.

Clostridium autoethanogenum is a Gram-positive, acetogenic bacterium. Acetogens are anaerobic bacteria with the ability to grow on carbon monoxide and carbon dioxide as their sole carbon source, producing acetic acid via the reductive acetyl-CoA pathway (also known as the Wood Ljungdahl pathway). A small subset of acetogens, for example *Clostridium autoethanogenum* and *Clostridium ljungdahlii,* are of particular industrial interest due to their ability to also produce ethanol from CO, CO_2 and H_2. This ability to utilise waste industrial greenhouse gases whilst producing a renewable non-petroleum based fuel has focused considerable research on these organisms over recent years (Norman *et al.*, 2018).

The genome of *C. autoethanogenum* harbours two gene clusters containing genes encoding proteins for the formation of BMCs (Fig. 4.1). A smaller 18 gene cluster is predicted to encode a 1,2-PD utilising GRM3-type BMC. Comparison of the signature enzyme with the functionally characterised 1,2-PD dehydratase of the *R. palustris* GRM3 BMC reveals 57 % identity between the protein sequences increasing confidence that this is a 1,2-PD metabolising BMC. The second larger cluster contains between 26 and 32 genes. It is not clear if some of the flanking genes, which encode regulatory proteins, are part of the BMC cluster. This GRM1-type BMC cluster includes genes for 8 shell proteins and a signature enzyme, choline trimethylene-lyase, for the utilisation of choline. The signature enzyme shares 78 % identity with the GRE of the *D. desulfuricans* Cut BMC. There are also two aldehyde dehydrogenases, which is only seen in GRM5-type BMCs (Zarzycki *et al.*, 2015).

This study aimed to examine the structures formed by the recombinant production of the individual shell proteins in *E. coli* and to investigate if any novel assemblies are formed. The ability to form empty BMCs in *E. coli* with the production of all the shell proteins was also investigated.

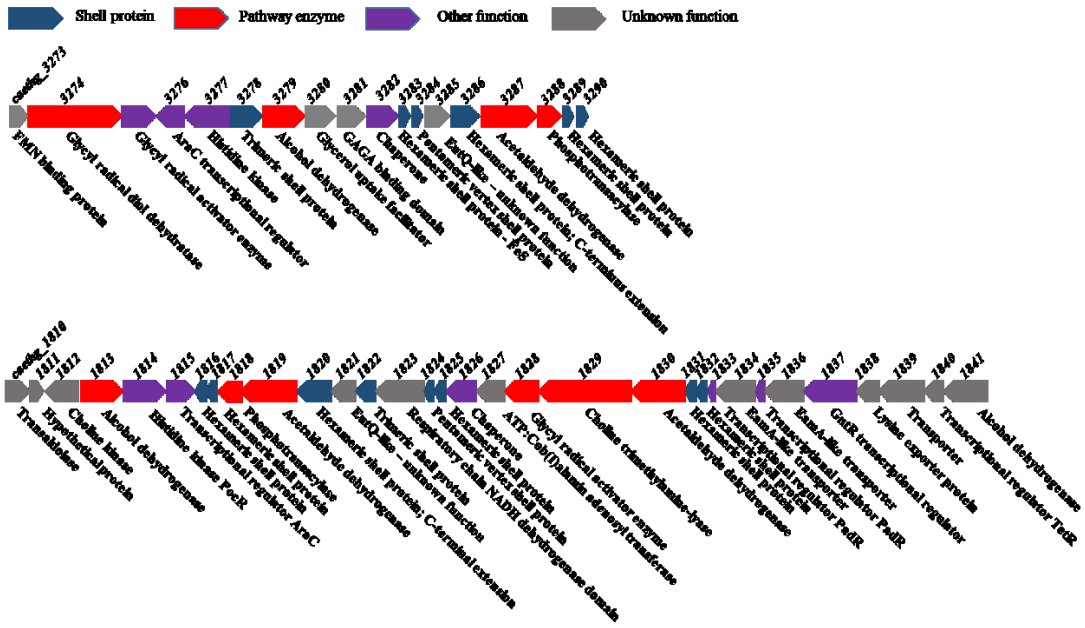


Figure 4.1. Schematic of the *C. autoethanogenum* BMC gene clusters. a) 18 gene cluster; b) 32 gene cluster. Shell proteins are indicated by blue arrows, pathway enzymes by red arrows, other functional proteins by purple arrows and proteins of unknown function by grey arrows.

4.2 Results: Small BMC gene cluster

4.2.1 Protein sequence alignments

The shell proteins of the smaller gene cluster include three hexameric shell proteins (BMC-H: single BMC domain - Pfam00936), one trimeric shell protein (BMC-T: two BMC domains), one pentameric vertex protein (BMC-P: single Pfam03319 domain) and one hexameric shell protein with a C-terminal extension (BMC-H_{ex}) of unknown function. The predicted substrate for the enzyme pathway encapsulated within the BMC, based on shell protein and enzyme complement (Zarzycki, Erbilgin and Kerfeld, 2015), is 1,2-PD. Therefore, the protein sequences of the *C. autoethanogenum* shell proteins were compared with the sequences of the *Citrobacter freundii* 1,2-PD utilising (Pdu) BMC shell proteins (Table 4.1).

Table 4.1. Sequence identity between *C. autoethanogenum* small GRM3 gene cluster shell proteins and the shell proteins of the *C. freundii* Pdu BMC. Percentage identity between *C. autoethanogenum* and *C. freundii* Pdu shell proteins. Pairwise global alignments performed in Emboss Needle (Madeira *et al.*, 2019). Percentage sequence similarity is shown in parentheses. Similarity greater than 50 % is highlighted in yellow.

	Caethg_3278 (BMC-T)	Caethg_3283 (BMC-H_{fe})	Caethg_3284 (BMC-P)	Caethg_3286 (BMC-H_{ex})	Caethg_3289 (BMC-H)	Caethg_3290 (BMC-H)
PduA	8 (13) %	18 (31) %	16 (34) %	19 (26) %	70 (84) %	68 (80) %
PduB	53 (71) %	10 (17) %	9 (15) %	5 (9) %	10 (16) %	10 (16) %
PduJ	10 (14) %	18 (28) %	17 (34) %	18 (29) %	67 (80) %	66 (78) %
PduK	7 (13) %	10 (20) %	12 (20) %	20 (29) %	22 (34) %	22 (32) %
PduN	7 (14) %	23 (35) %	32 (51) %	9 (17) %	21 (37) %	20 (34) %
PduU	14 (25) %	20 (35) %	18 (28) %	8 (14) %	14 (23) %	14 (23)%
PduT	14 (23) %	12 (20) %	10 (18) %	10 (14) %	18 (29) %	18 (28) %

As with the *R. rubrum* GRE BMC, the *C. autoethanogenum* small BMC gene cluster encodes for PduA-, PduB-, PduJ- and PduN-like proteins. Caethg_3286 is a C-terminally extended hexameric protein but shares little similarity with PduK, which also has a C-terminal extension. The C-terminal extension of Caethg_3286 is much longer than the PduK C-terminus (170 amino acids vs. 80 amino acids) suggesting diverse properties and different functional roles. Caethg_3283 shares little similarity with any of the Pdu shell proteins but does share 78 % similarity with GrpU of *Clostridiales bacterium 1_7_47FAA* which has an Fe-S binding centre within the central pore. PduT, a trimeric shell protein with an Fe-S binding centre, has no homologues within the *C. autoethanogenum* GRM3 BMC. Similarly, PduU, a circularly permuted hexameric shell protein, shares little similarity with any of the *C. autoethanogenum* shell proteins.

4.2.2 Production of individual shell proteins in *E. coli*

The structures formed by individual shell proteins, for example sheets or filaments, can inform on their possible contribution to the BMC architecture. Therefore, the phenotypes of *E. coli* over-producing individual recombinant *C. autoethanogenum* shell proteins were investigated. The genes encoding the shell proteins were first amplified from the genome by PCR and cloned individually into a pET3a vector. Competent *E. coli* BL21*(DE3) pLysS cells were transformed with a pET3a vector harbouring one of the shell protein genes and the resulting strain was grown at 37 °C in LB media containing ampicillin and chloramphenicol to an OD_{600} ~ 0.6. Protein production was induced with 400 µM isopropyl β-D-1-thiogalactopyranoside (IPTG) and cultures grown overnight at 19 °C before harvesting for fixing and thin sectioning.

4.2.2.1 Caethg_3278

Caethg_3278 is a 30 kDa shell protein, 287 amino acids in length and containing two BMC domains that is predicted to form pseudohexameric trimers. It shares the greatest identity, 53 %, with PduB which is also a trimeric shell protein. Residues 1-37 of PduB, which precede the BMC domains, have been implicated in the targeting of enzymes to the BMC lumen (Lehman *et al.*, 2017) but these are not conserved in Caethg_3278 suggesting that this region is either not involved in targeting for the *C. autoethanogenum* BMC or that different residues interact with the encapsulated protein.

```
                1       10        20        30        40        50        60        70        80        90       100
                |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
Caethg3278      MDQEIINKVLEELRKKTNCENSSDTTDKTQNTQNSKSELKEEKIIEKEEKREAFMINDTPVITEYVGCAIGDTVGLVIANIDGSIQEKMNLDPKYRSLGI
PduB                                SSNELVDQIMAQVIARVATPEQQAIPENNPPTRETAMAEKSCSLTEFVGTAIGDTVGLVIANVDSALLDAMKLEKRYRSIGI
Consensus       .....................N........Q.....................RE..M.......TE.VG.AIGDTVGLVIAN.D......M.L...YRS.GI

                101     110       120       130       140       150       160       170       180       190       200
                |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
Caethg3278      ISARTGAGPHIMAADEAVKATNTVIAAIELPRDTKGGAGHGSLIIFASEDVSDARRAVEVALKDVERTFGDVYGCDAGHAELHYTARASLALEKAYGAPL
PduB            LGARTGAGPHIMAADEAVKATNTEVVSIELPRDTKGGAGHGSLIILGGNDVSDVKRGIEVALKELDRTFGDVYANEAGHIEMQYTARASYALEKAFGAPI
Consensus       ..ARTGAGPHIMAADEAVKATNT....IELPRDTKGGAGHGSLII....DVSD..R..EVALK...RTFGDVY...AGH.E..YTARAS.ALEKA.GAP.

                201     210       220       230       240       250       260       270       280     287
                |--------+---------+---------+---------+---------+---------+---------+---------+-------|
Caethg3278      GKSFGIVVGAPAGVGLVMADTAMKTANVELVTYASPNDGTSHSNEVIISITGDSGAVKQSVIAARERVGIGILKSMGQNPISTTKPYI
PduB            GRACGVIVGAPASVGVLMADTALKSANVEVVAYSSPAHGTSFSNEAILVISGDSGAVRQAVISARERIGKTVLGTLGSEPKNDRPSYI
Consensus       G...G..VGAPA.VG..MADTA.K.ANVE.V.Y.SP..GTS.SNE.I..I.GDSGAV.Q.VI.ARE.G...L...G..P......YI
```

Figure 4.2. Protein sequence alignment of Caethg_3278 with PduB. Conserved residues shown in red. Protein sequences aligned using the program Multalin (Corpet, 1988).

When over-produced in *E. coli*, Caethg_3278 forms insoluble aggregates mainly at the poles of the cells (Fig. 4.3). This is unlike PduB which forms filaments upon over-production in *E. coli* but similar to *R. rubrum* Rru_A0906 which also formed aggregates.

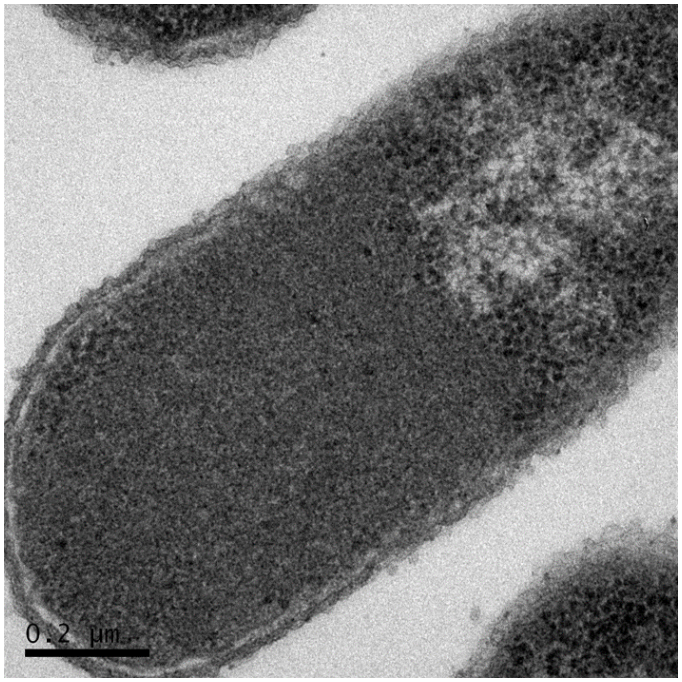


Figure 4.3. TEM image of *E. coli* over-producing Caethg_3278. Insoluble aggregated protein can be seen. Scale bar 0.2 µm

4.2.2.2 Caethg_3283

Caethg_3283 is a 101 amino acid protein with a molecular mass of 11 kDa. It has a single BMC domain predicted to form a cyclic hexamer. Caethg_3283 shares little similarity with any of the Pdu shell proteins but instead shares 59 % identity with GrpU from *Clostridiales bacterium 1_7_47FAA* (Fig. 4.4). GrpU is a circularly permuted hexameric shell protein with an iron-sulphur binding site within the central pore co-ordinated by a conserved cysteine residue. Caethg_3283 has this conserved cysteine residue suggesting that Caethg_3283 also has an Fe-S binding centre within its pore, perhaps for electron transport across the shell.

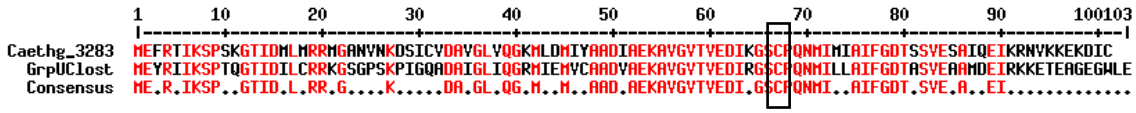


Figure 4.4. Protein sequence alignment of Caethg_3283 with GrpU of *Clostridiales bacterium* 1_7_47FAA. A conserved cysteine at position 68 proposed to be involved in Fe-S binding is indicated (box). Protein sequences aligned using the program Multalin (Corpet, 1988).

No visible structures were formed upon over-production of Caethg_3283 in *E. coli* (Fig. 4.5). Sheets were formed by the *R. rubrum* GrpU-like shell protein, Rru_A0915 (see Section 3.2.2.6), but this was not observed with Caethg_3283.

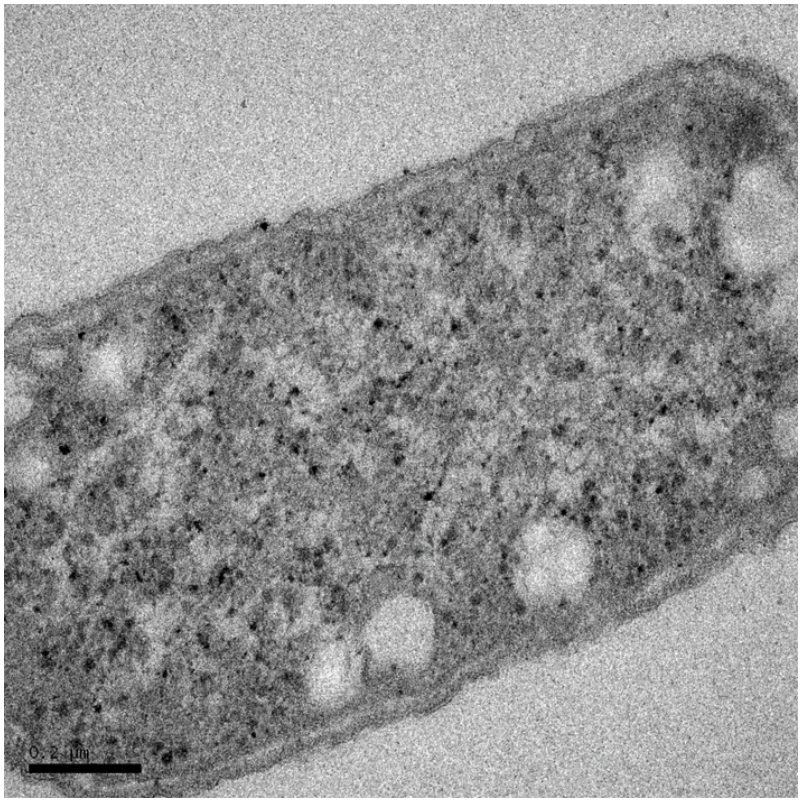


Figure 4.5. TEM image of *E. coli* over-producing Caethg_3283. Cells appear normal. Scale bar 0.2 µm

4.2.2.3 Caethg_3284

Caethg_3284 is an 88 amino acid, 9.3 kDa protein predicted to exist as a pentamer, forming the vertices of the microcompartment. Its closest homologue of the Pdu BMC is PduN, a vertex protein with which it shares 51% similarity.

When over-produced in *E. coli*, Caethg_3284 does not appear to form any structures and cells appear normal (Fig. 4.6).

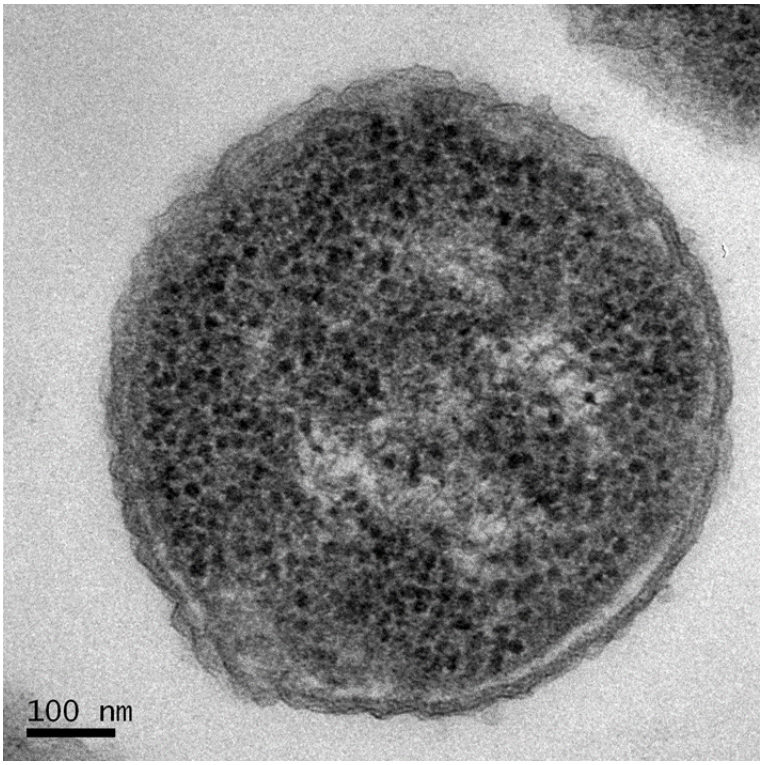


Figure 4.6. TEM image of *E. coli* over-producing Caethg_3284. All cells appear normal. Scale bar 100 nm.

4.2.2.4 Caethg_3286

Caethg_3286 is a 28.5 kDa protein composed of 258 amino acids, containing an N-terminal BMC domain and a C-terminal extension of unknown function. Comparison with the C-terminally extended PduK reveals only a 30 % similarity. Alignment of the protein sequence shows more conserved residues within the BMC domain but the C-terminal extensions have divergent sequences, Caethg_3286 C-terminus being approximately 100 residues longer than PduK (Fig. 4.7). Over-production of Caethg_3286 in *E. coli* does not result in the formation of any observable protein structures within the cytoplasm and cells appear normal (Fig. 4.8).

```
                1       10        20        30        40        50        60        70        80        90       100
                |--------+---------+---------+---------+---------+---------+---------+---------+---------+--------|
Caethg3286       MQALGLIETIGLIAAIESADAMLKAADVNLLEKTYVGGGLVSIVVTGDVGAVKAAVDAGGAAVRNINDTLLVSQHVIPRPHEELDNIIVSAVPLKDRNI
PduK            VKQSLGLLEVSGLALAISCADVMAKAASITLVGLEKTNGSGWMVIKIIGDVASVQAAISTGVSFADQRDGLVAHKVISRPGDGIL---------------
Consensus       ..Q.LGL.E..GL..AI..AD.M.KAA...L.......G.............A...A....G.......D.L..............L................

                101     110       120       130       140       150       160       170       180       190       200
                |--------+---------+---------+---------+---------+---------+---------+---------+---------+--------|
Caethg3286      KTVNPEENQSKVVTMVEGNEEESVEVSLKIEELEYDDSKEEDSMIENPTIDDSIVEDPIEEDSKEENSTVENSENEELLPEKITDPLKIDLTKLNRKAVD
PduK            ---------SHSVTPESESEPAPAPTPVVPHEEIPEDHAAPEAPQDAELISCNLCLDPACPRQKGEPRSLCLHSGKRGEA
Consensus       .........S..VT.....E............E....D.............I......DP.....K.E.................................

                201     210       220       230       240       250       259
                |--------+---------+---------+---------+---------+--------|
Caethg3286      KMVLDYGLEKAIEILSKFKVIELRNLARKYKNFGIKGRSISKADKMLLLVEFRKYYGHN
PduK
Consensus       ...........................................................
```

Figure 4.7. Protein sequence alignment of Caethg_3286 with PduK. More conserved residues (shown in red) are observed within the BMC domain (residues 1-90). Residues within the C-terminal extension share little identity. Protein sequences aligned using the program Multalin (Corpet, 1988).

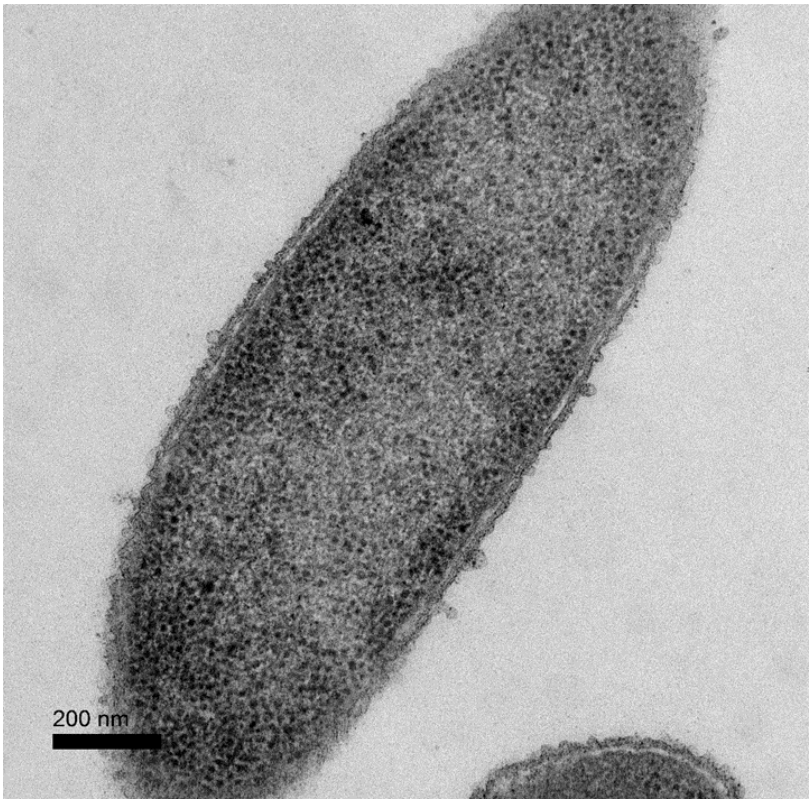


Figure 4.8. TEM image of *E. coli* over-producing Caethg_3286. Scale bar 200 nm.

4.2.2.5 Caethg_3289

Caethg_3289 is a 96 amino acid protein with a molecular mass of 9.6 kDa. It shares 70 % and 67 % identity with PduA and PduJ respectively (Fig. 4.9).

```
                1        10        20        30        40        50        60        70        80        90        96
                |--------+---------+---------+---------+---------+---------+---------+---------+---------+-------|
Caethg_3289     MKSDALGMIETKGLVGSIEAADAMVKAANVNLIGKEHVGGGLVTVMVRGDVGAVKAATDAGAAAAQRVGELISVHVIPRPHVEVETILPKTGLKED
PduA            MQQEALGMVETKGLTAAIEAADAMVKSANVMLVGYEKIGSGLVTVIVRGDVGAVKAATDAGAAAARNVGEVKAVHVIPRPHTDVEKILPKGIS
PduJ             MNNALGLVETKGLVGAIEAADAMVKSANVQLVGYEKIGSGLITVMVRGDVGAVKAAVDAGSAAASAVGEVKSCHVIPRPHSDVEAILPKSA
Consensus       m...ALGmvETKGLvgaIEAADAMVKsANV.LvGyEkiGsGLvTVmVRGDVGAVKAAtDAGaAAA..VGEvksvHVIPRPH.dVE.ILPK......
```

Figure 4.9. Protein sequence alignment of Caethg_3289 with PduA and PduJ. Residues conserved in all three proteins are denoted in red, residues conserved in two of the proteins are denoted in blue. Protein sequences aligned using the program Multalin (Corpet, 1988).

TEM images of *E. coli* over-expressing *caethg_3289* show masses of filamentous structures extending throughout the cytoplasm of almost every cell. The filaments do not appear to be in regular layers but are instead massed tightly together. Swirls in transverse sections and sheets in longitudinal section suggests that the protein forms large sheets that are rolled up and extend longitudinally along the cells often interfering in cell division (Fig. 4.10).

The structures observed in this study appear more disordered than those seen with over-production of PduA and PduJ, despite their close sequence similarity. Similar structures were reported with the over-production of an Eut shell protein from *Clostridium difficile*, CD1918 (Pitts *et al.*, 2012), with which Caethg_3289 shares 87 % similarity (Fig. 4.11).

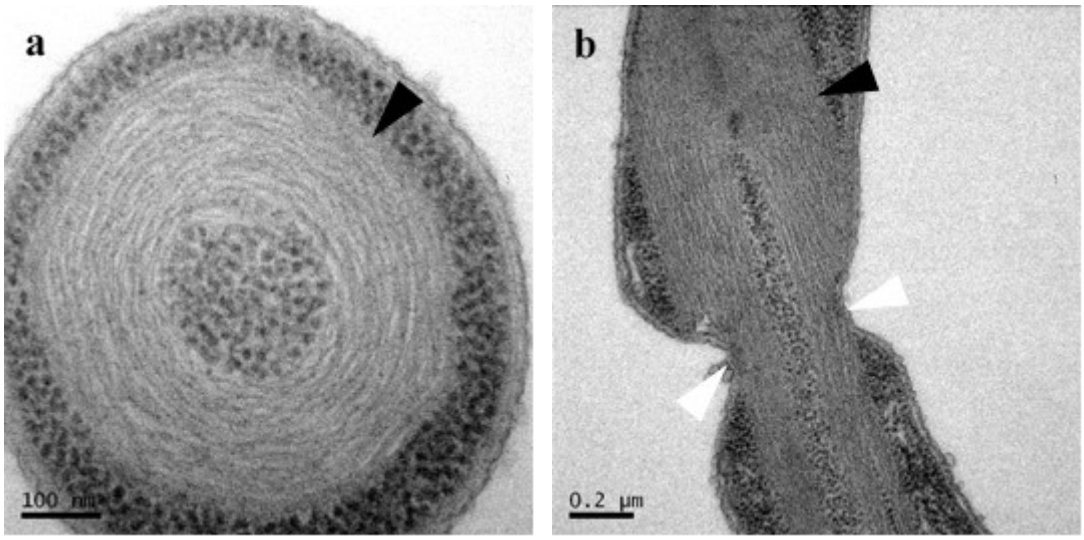


Figure 4.10. TEM image of *E. coli* over-producing Caethg_3289. Sheets of protein (black arrows) and their interference in cell division (white arrows) are indicated. a) transverse section; b) longitudinal section. Scale bars a) 100 nm, b) 0.2 µm

```
                1       10        20        30        40        50        60        70        80        90        98
                |-------+---------+---------+---------+---------+---------+---------+---------+---------+-------|
Caethg_3289       MKSDALGMIETKGLVGSIEAADAMVKAANVNLIGKEHVGGGLVTVMVRGDVGAVKAATDAGAAAAQRVGELISVHVIPRPHVEVETILPKTGLKED
CD1918      MGASANALGMIETKGLVGAIEAADAMVKAANVQLVGKEQVGGGLVTVMVRGDVGAVKAATDAGAAAAERVGELISVHVIPRPHFEVDAILPKVSAELE
Consensus   ......ALGMIETKGLVG.IEAADAMVKAANV.L.GKE.VGGGLVTVMVRGDVGAVKAATDAGAAAA.RVGELISVHVIPRPH.EV..ILPK......
```

Figure 4.11. Protein sequence alignment of Caethg_3289 with CD1918, an Eut shell protein from *Clostridium difficile*. Protein sequences aligned using the program Multalin (Corpet, 1988).

4.2.2.6 Caethg_3290

Caethg_3290 is a 10 kDa, 98 amino acid hexameric protein that shares 93 % identity with Caethg_3289 (Fig. 4.12).

```
                1       10        20        30        40        50        60        70        80        90        98
                |-------+---------+---------+---------+---------+---------+---------+---------+---------+-------|
Caethg3289      MKSDALGMIETKGLVGSIEAADAMVKAANVNLIGKEHVGGGLVTVMVRGDVGAVKAATDAGAAAAQRVGELISVHVIPRPHVEVETILPKTGLKED
Caethg3290      MKYDALGMIETKGLVGSIEAADAMVKAANVYLIGKEHVGGGLVTVMVRGDVGAVKAATDAGAAAAQRVGELVSVHVIPRPHVEVETILPKTNLKEDEK
Consensus       MK.DALGMIETKGLVGSIEAADAMVKAANV.LIGKEHVGGGLVTVMVRGDVGAVKAATDAGAAAAQRVGEL.SVHVIPRPHVEVETILPKT.LKED..
```

Figure 4.12. Protein sequence alignment of Caethg_3289 and Caethg_3290. Protein sequences aligned using the program Multalin (Corpet, 1988).

Over-production of Caethg_3290 in *E. coli* resulted, not surprisingly, in a phenotype like that seen with over-production of Caethg_3289. TEM images show swirled sheets in transverse sections and longer sheets extending through the cell in longitudinal sections (Fig. 4.13).

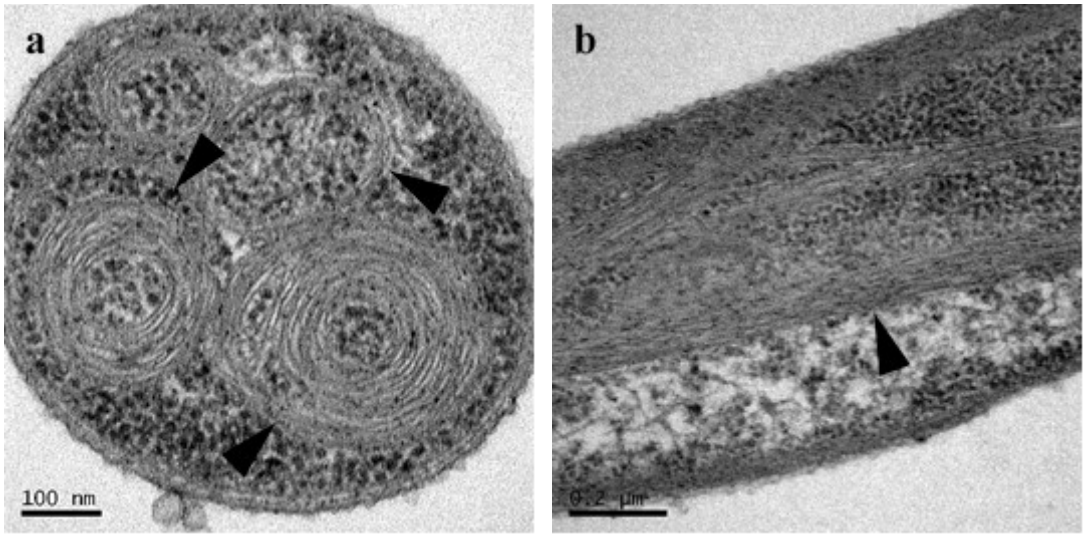


Figure 4.13. TEM images of Caeth_3290 over-produced in *E. coli*. a) Transverse section; b) longitudinal section. Swirled sheets of protein extending through the cell are indicated (black arrows). Scale bars a) 100 nm, b) 0.2 µm

Because Caethg_3289 and Caeth_3290 form such extensive sheets of protein, it would suggest that they make a considerable contribution to the formation of the protein sheets forming the facets of the microcompartment. The presence of two genes encoding almost identical proteins also hints at the importance of these proteins in the formation of the *C. autoethanogenum* BMC.

4.2.2.7 Summary of single shell protein production

Three of the *C. autoethanogenum* shell proteins, Caethg_3283 (BMC-H_{FeS}) Caethg_3284 (BMC-P) and Caethg_3286 (BMC-H_{ex}), did not form any observable structures upon over-production in *E. coli*, with cells appearing normal. The trimeric shell protein, Caethg_3278, formed insoluble aggregates but no higher order structures were observed, unlike the tubes observed upon over-expression of PduB, a trimeric shell protein of the Pdu BMC.

The most extensive protein sheets were observed upon over-production of the BMC-H shell proteins Caethg_3289 and Caethg_3290. Almost every cell contained an abundance of protein sheets that were seen to swirl around and extend throughout the length of the cells. The presence of two genes encoding almost identical proteins and the abundance of protein structures observed would suggest an important role for these proteins in the formation of BMC structures in *C. autoethanogenum*.

4.2.3 Over-production of all *C. autoethanogenum* small gene cluster shell proteins

4.2.3.1 Protein production in BL21* (DE3) pLysS

The ability of *E. coli* to form empty microcompartments through the recombinant production of *C. autoethanogenum* shell proteins was investigated next. A single vector containing all six shell protein genes of the small *C. autoethanogenum* small BMC gene cluster was cloned using the "Link-and-Lock" technique (McGoldrick *et al.*, 2005). Each of the genes had its own ribosome binding site and were cloned in the order in which they are found in the BMC gene cluster (Fig. 4.14).

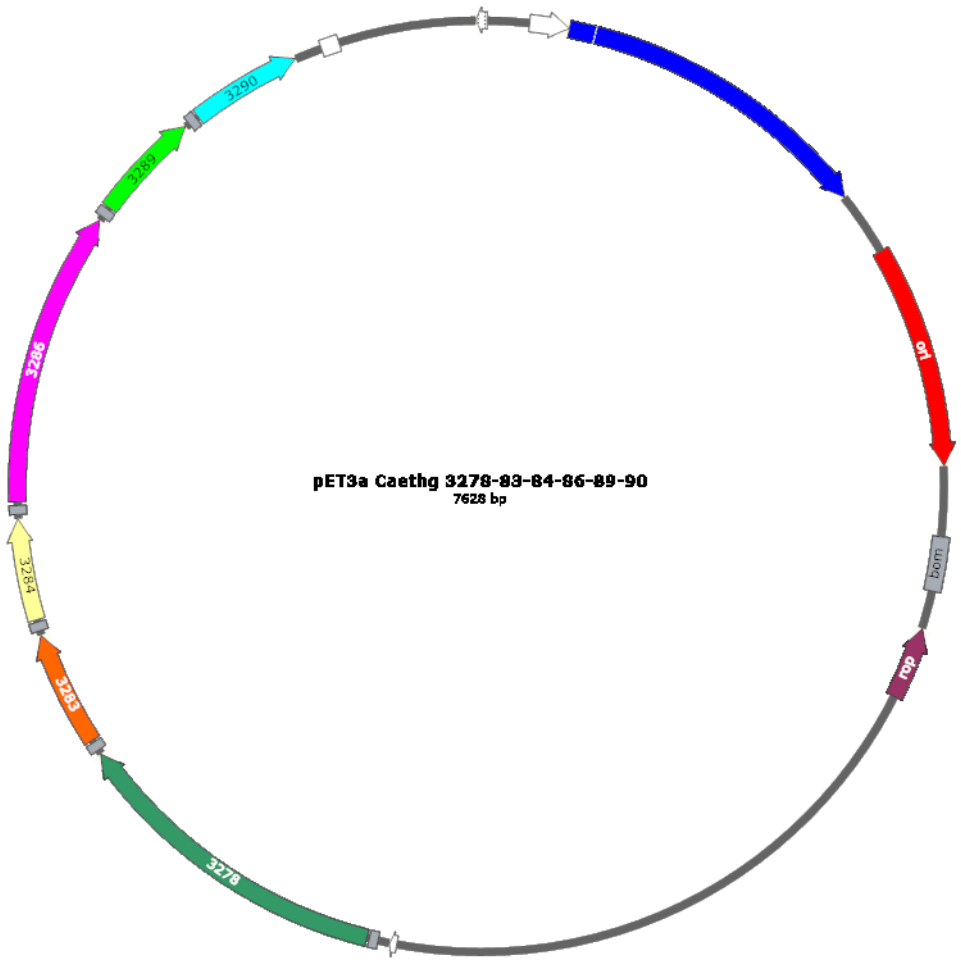


Figure 4.14. Schematic of the pET3a vector harbouring the 6 shell protein genes of the *C. autoethanogenum* small BMC gene cluster. Genes coloured as follows: green *caethg_3278*; orange *caethg_3283*; yellow *caethg_3284*; pink *caethg_3286*; lime green *caethg_3289*; cyan *caethg_3290.* Ribosome binding site shown as grey box preceding each shell protein gene. Figure produced in SnapGene software (from GSL Biotech; available at snapgene.com)

E. coli BL21*(DE3) pLysS cells were transformed with the pET3a vector containing all six shell protein genes and the resulting strain was grown at 37 °C to an OD_{600} ~ 0.6 in LB media

containing ampicillin and chloramphenicol. Protein production was induced with 400 µM isopropyl β-D-1-thiogalactopyranoside (IPTG) and cultures were incubated with shaking overnight at 19 °C before harvesting for fixing, embedding and thin sectioning.

TEM images of *E. coli* over-producing all of the six shell proteins that are encoded in the *C. autoethanogenum* small BMC gene cluster show that no microcompartment structures are formed but instead lines and swirls of protein are observed. Unusually, ribosomes appear to be associated with the protein in many of these cells (Fig. 4.15). The ribosomes appear regularly spaced between the protein sheets.

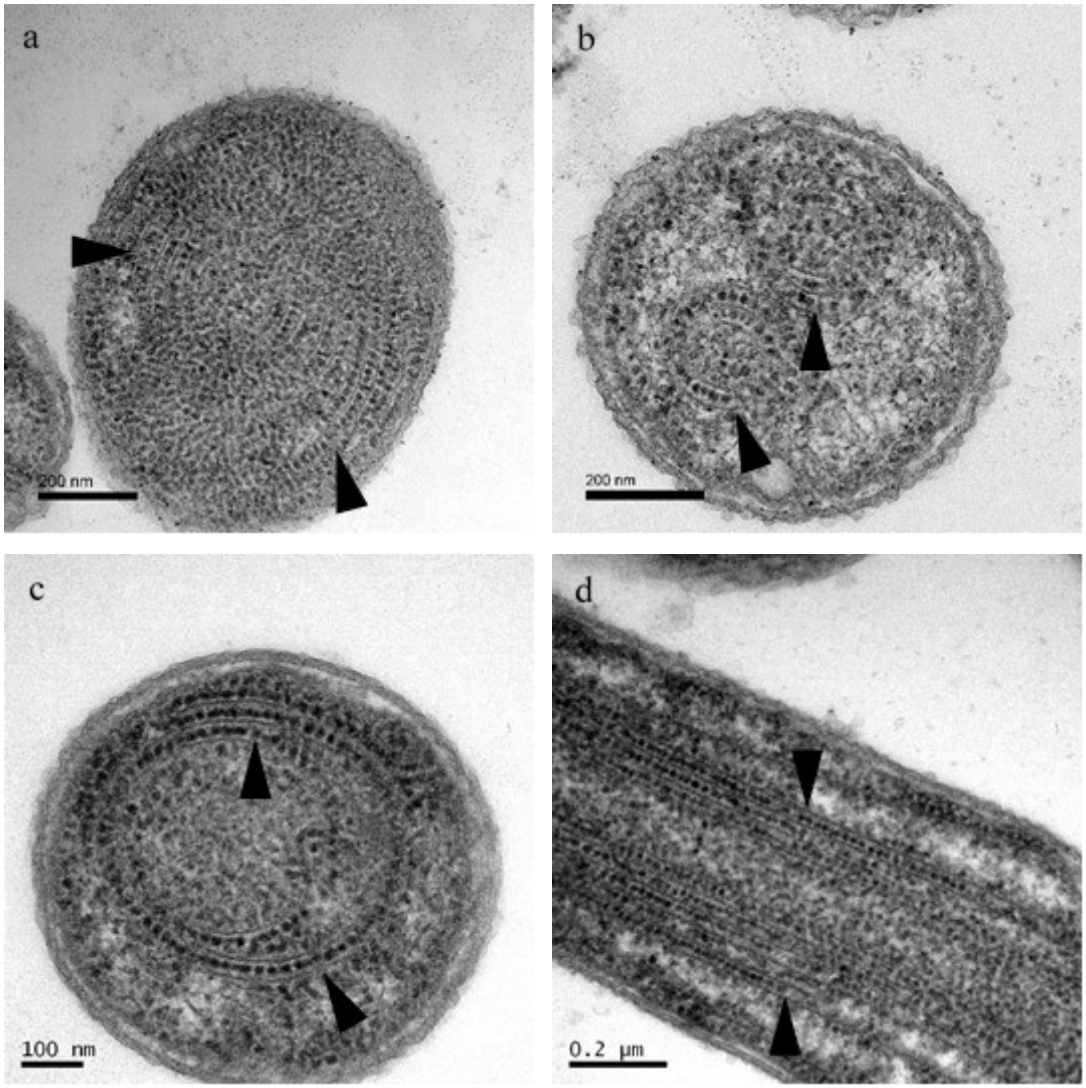


Figure 4.15. TEM images of *E. coli* over-producing *C. autoethanogenum* shell proteins Caethg_3278, Caethg_3283, Caethg_3284, Caethg_3286, Caethg_3289 and Caethg_3290. Wrapped sheets of protein (black arrows) are indicated. Ribosomes can be observed aligned along the sheets. Scale bars: a) 200 nm; b) 200 nm; c) 100 nm; d) 0.2 µm

Varying induction conditions (0 and 100 µM IPTG, 3 hr induction at 100 µM) were tested but all resulted in the same phenotype being observed by TEM (Fig 4.16).

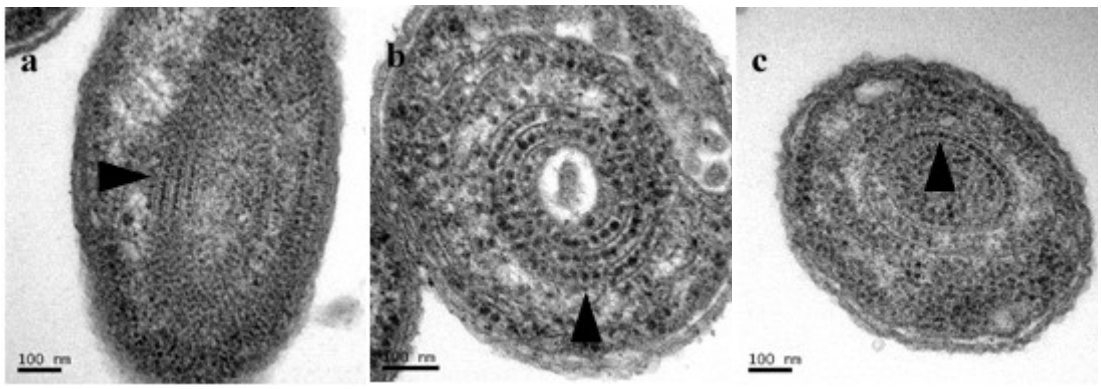


Figure 4.16. TEM images of *E. coli* over-producing the six *C. autoethanogenum* shell proteins Caethg_3278, Caethg_3283, Caethg_3284, Caethg_3286, Caethg_3289 and Caethg_3290 under different induction conditions. *a*) 0 µM IPTG; b) 100 µM IPTG/overnight induction; c) 100 µM IPTG/3 hr induction. Ribosomes associated with protein sheets are indicated (black arrows). Scale bars 100 nm

This phenotype is highly unusual as bacterial ribosomes are not normally found in any form of organised structure but are instead randomly distributed throughout the cytoplasm.

4.2.3.2 Protein production in Rosetta-2 (DE3) pLysS

Ribosome stalling has been shown to induce the formation of polysomes in *E. coli* which, when examined by electron tomography, share some similarities to the lining up of ribosomes seen in this study (Cougot *et al.*, 2014). Ribosome stalling may occur due to the presence of rare codons within the coding region and a low abundance of cognate tRNAs being present in the cytoplasmic pool (Hanson and Coller, 2018). Analysis of the *C. autoethanogenum* shell protein genes revealed the presence of some codons that are less abundant in the *E. coli* genome therefore it is possible that the associated tRNAs required for the translation of these codons are also only present at lower concentrations so resulting in ribosome stalling and producing the phenotype observed in this study. The 6 shell protein genes were subsequently expressed in Rosetta-2 DE3 pLysS cells, a strain of *E. coli* containing additional tRNAs for several rare codons that are present in the shell protein mRNA. TEM images show that ribosomes still appear to be associated with the protein when expressed in Rosetta-2 DE3 pLysS *E. coli* (Fig. 4.17). This would suggest the phenotype observed is not caused by ribosome stalling due to the presence of rare codons.

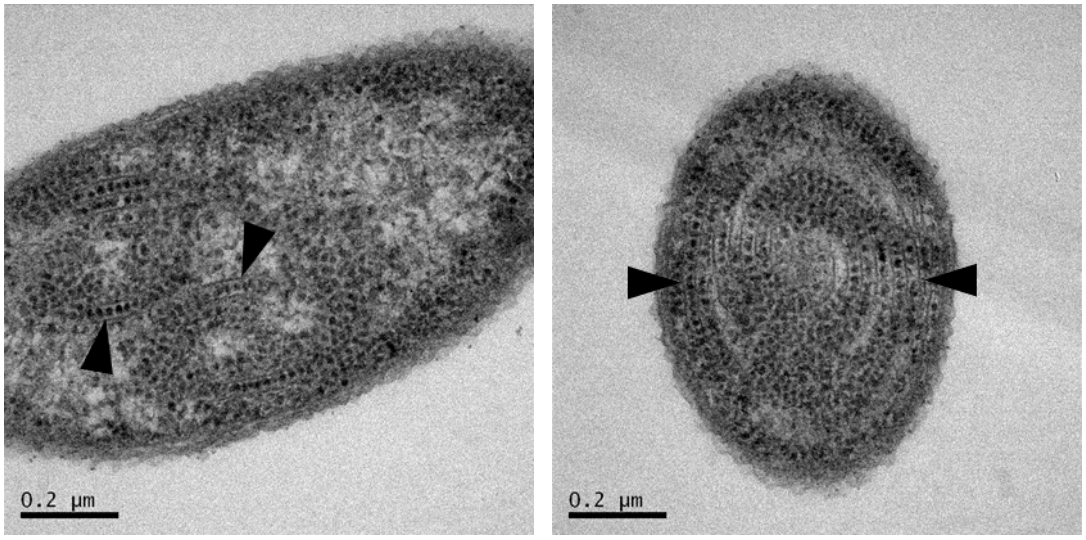


Figure 4.17. TEM images of Rosetta-2 pLysS *E. coli* cells over-producing the six *C. autoethanogenum* shell proteins. Sheets of protein with ribosomes aligned are indicated (black arrows). Scale bars 0.2 µm

4.2.3.3 Imaging serial thin sections

To determine if the protein observed in these sections are forming extended sheets, serial 70 nm sections were cut, mounted on formvar coated 0.5 mm copper slot grids and the same cell located and imaged through several sections. Images are slightly blurry due to the extra depth of the formvar coating on the grids but, the swirls of protein observed in the selected cell can be seen in each sequential section therefore confirming the protein is forming sheets that are rolled up within the cytoplasm (Fig. 4.18). As each section is 70 nm in depth and up to 5 cell sections were located and imaged, the protein sheets extend at least 350 nm through the length of the cell.

Microcompartments were clearly not formed under the conditions of this study, however, extended sheets of protein were able to self-assemble. The observation of ribosome alignment along the length of the protein sheets was an interesting and previously unobserved phenomenon which prompted further investigation.

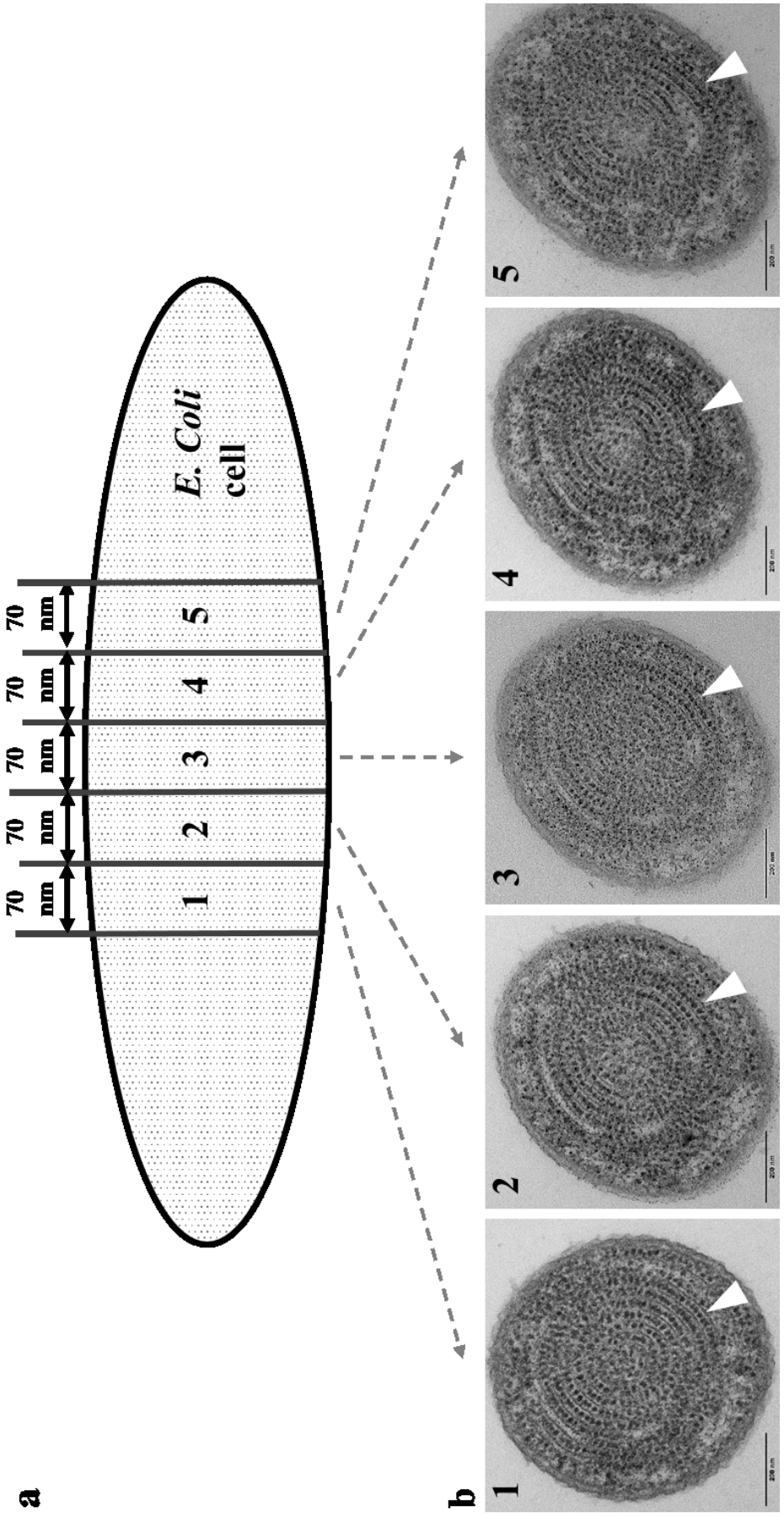


Figure 4.18. Serial thin sections of *E. coli* over-producing the six *C. autoethanogenum* shell proteins. a) Schematic illustrating the serial sectioning of the *E. coli* forming sections 1 – 5. b) TEM images of the same cell same cell located in each of sections 1 – 5. The same swirled protein sheets in each section is indicated (white arrow). Scale bars 200 nm.

4.2.4 Omission of one shell protein gene from multi-gene plasmids

To investigate which shell protein(s) may be responsible for the association of ribosomes along the protein sheet, 6 constructs were made using the "Link-and-Lock" technique, each containing 5 shell protein genes with a single shell protein gene omitted (Table 4.2).

Table 4.2 Table of the constructs prepared to examine the effect of omitting one shell protein

Plasmid	Gene Omitted
pET3a *caethg_3278-83-84-86-89*	*caethg_3290*
pET3a *caethg_3278-83-84-86-90*	*caethg_3289*
pET3a *caethg_3278-83-84-89-90*	*caethg_3286*
pET3a *caethg_3278-83-86-89-90*	*caethg_3284*
pET3a *caethg_3278-84-86-89-90*	*caethg_3283*
pET3a *caethg_3283-84-86-89-90*	*caethg_3278*

Following transformation of BL21*(DE3) pLysS *E. coli*, strains were grown in LB media to an OD_{600} ~ 0.6 before induction of protein production with 400 µM IPTG. Imaging by TEM show ribosomes associated with protein sheets in cells expressing the constructs pET3a *caethg_3278-83-84-86-89,* pET3a *caethg_3278-83-84-86-90,* pET3a *caethg_3278-83-86-89-90,* pET3a *caethg_3278-84-86-89-90* and pET3a *caethg_3283-84-86-89-90* (Fig. 4.19 a – c; e - f). However, *E. coli* expressing pET3a *caethg_3278-83-84-89*-90, missing *caethg_3286*, display a different phenotype. Swirled sheets of protein are observed but ribosomes are not associated with them in any of the cells (Fig. 4.19 d). As ribosomes are not aligned with the protein sheets only when Caethg_3286 is absent, this would suggest that this shell protein is responsible for the ribosomal interaction.

The 'Linked-and-Locked' plasmid containing all the shell protein genes was sequenced to ensure the stop codon had not been lost from the 3' end of *caethg_3286* during cloning which may have resulted in the ribosomes being trapped. However, sequencing confirmed the stop codon was present therefore some other feature of this protein was resulting in the phenotype observed.

Caethg_3286 is the C-terminally extended hexameric shell protein. Earlier data showed that when over-produced in isolation, Caethg_3286 does not form any structures within the cell and no ribosomal interaction is observed (Section 4.2.2.4). Caethg_3286 homomers may be unable to form higher order structures but instead may form heteromers with other shell proteins allowing the self-assembly of structural proteins.

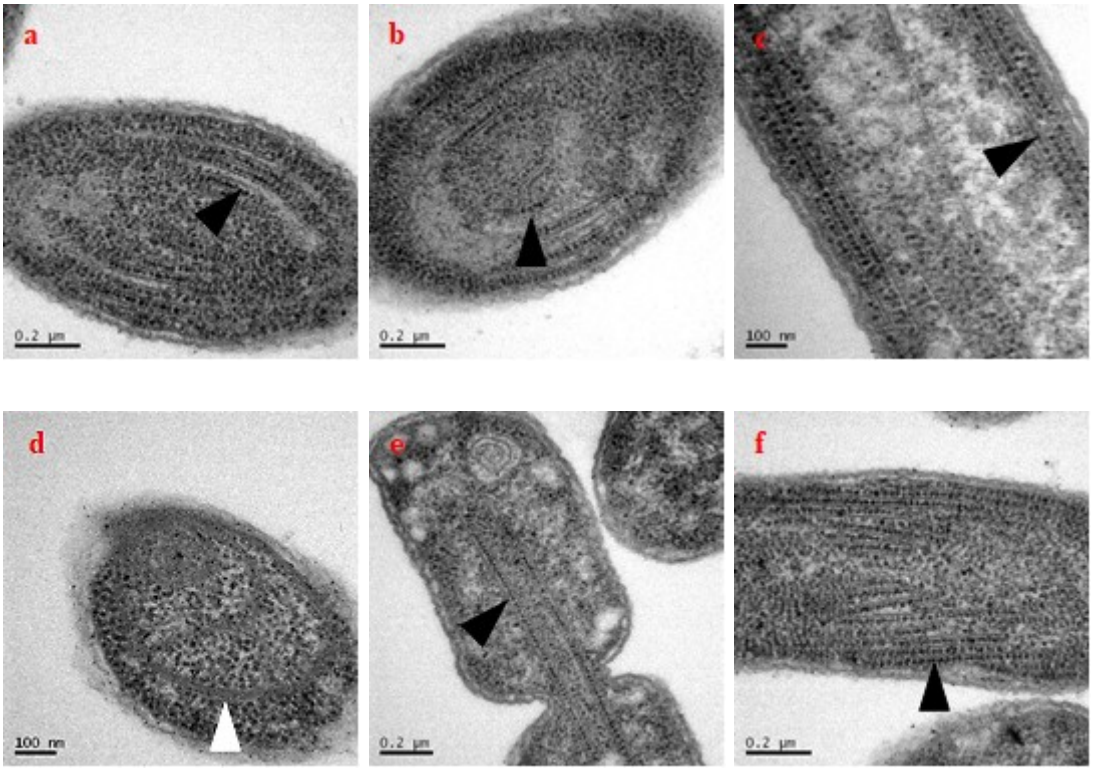


Figure 4.19. TEM images of thin sections of *E. coli* expressing 5 shell protein genes with one gene omitted. a) *caegth_3278* omitted, b) *caegth_3283* omitted, c) *caegth_3284* omitted, d) *caegth_3286* omitted, e) *caegth_3289* omitted, f) *caegth_3290* omitted. Ribosomes are observed associated with protein sheets with each construct tested except when *caethg_3286* was omitted. Sheets of protein with ribosomes aligned (black arrows) and swirled sheets of protein (white arrows) are indicated. Scale bar 0.2 µm (a, b, e & f); 100 nm (c & d)

4.2.5 Summary of small gene cluster shell protein production in *E. coli*

There was no evidence of microcompartment formation on over-production of all the shell proteins encoded by the small BMC gene clusterof *C. autoethanogenum*. Instead, rolled sheets of protein were observed and ribosomes appeared to be associated with these protein sheets. Omission of individual shell proteins determined that the BMC-H_{ex} shell protein, Caethg_3286, appeared to be responsible for this phenomenon. Further studies examining the interaction of this protein with ribosomes are summarised in Chapter 5.

4.3 Results: Production of shell proteins encoded by the GRM1 type large BMC gene cluster of *C. autoethanogenum*

4.3.1 Protein sequence alignments

The larger *C. autoethanogenum* BMC gene cluster has between 26 and 32 genes as it is not clear whether some of the flanking genes involved in BMC formation or not. There are 8 genes encoding for shell proteins: *caethg_1816*, *caethg_1817*, *caethg_1820*, *caethg_1822*, *caethg_1824*, *caethg_1825*, *caethg_1831* and *caethg_1832*. They do not all encode unique shell proteins, however, as some are identical to shell proteins of the GRM3 BMC gene cluster. In fact, 3 of the 6 small gene cluster shell proteins are duplicated in the GRM1 BMC cluster: Caethg_3290 (BMC-H) is identical to Caethg_1816; Caethg_3284 (BMC-P) is identical to Caethg_1824; and Caethg_3283 (BMC-H_{FeS}) is identical to Caethg_1825. Caethg_3289 (BMC-H) also shares 99% identity with Caethg_1817. This means there are PduA-, PduJ- and Pdu N-like shell proteins along with a GrpU-like shell protein (BMC-H_{FeS}) represented. Sequence comparisons of the remainder of the shell proteins with the Pdu BMC shell proteins are summarised in Table 4.3.

Table 4.3 Sequence identity between *C. autoethanogenum* large gene cluster shell proteins Caethg_1820, Caethg_1822, Caethg_1831 and Caethg_1832 and the shell proteins of the *C. freundii* Pdu BMC. Pairwise global alignments performed in Emboss Needle (Madeira *et al.*, 2019). Percentage sequence similarity is shown in parentheses. Similarity greater than 50 % is highlighted in yellow.

	1820 (BMC-H_{ex})	**1822 (BMC-T_{FeS})**	**1831 (BMC-H)**	**1832 (BMC-H)**
PduA (BMC-H)	20 (27) %	16 (26) %	46 (62) %	44 (63) %
PduB (BMC-T)	5 (9) %	17 (26) %	10 (19) %	9 (15) %
PduJ (BMC-H)	19 (26) %	18 (26) %	51 (64) %	50 (74) %
PduK (BMC-H_{ex})	21 (31) %	20 (35) %	17 (29) %	21 (32) %
PduN (BMC-P)	9 (18) %	14 (23) %	6 (11) %	19 (36) %
PduU (BMC-H_{perm})	8 (14) %	16 (26) %	15 (29) %	15 (28) %
PduT (BMC-T_{FeS})	16 (28) %	39 (64) %	14 (25) %	14 (25) %

There are no PduB- or PduU-like shell proteins encoded within the large BMC gene cluster.

Caethg_1820 is a C-terminally extended single BMC domain protein (BMC-H_{ex}) although it only shares 21 % identity with PduK. Like the BMC-H_{ex} protein Caethg_3286, it is much larger than PduK with almost 100 more residues in the C-terminal extension. It shares 86 % identity with Caethg_3286.

Caethg_1831 and Caethg_1832 both share some identity with PduA and PduJ but only to a level of approximately 50 % identity. Within the small BMC gene cluster, Caethg_3289 and Caethg_3290 are the most closely aligned, again sharing approximately 50 % identity. Caeth_1831 and Caeth_1832 have 51 % identity with each other.

Caethg_1822 does not have a homologue within the small BMC gene cluster but does share 39 % identity with PduT, which is a tandem BMC domain protein. PduT was shown to bind a [4Fe-4S] cluster via a conserved Cys38 residue (Parsons *et al.*, 2008), which is also present in Caethg_1822 (Fig. 4.20). Structural studies showed PduT Cys38 to line the pore of a homotrimer (Crowley *et al.*, 2010; Pang, Warren and Pickersgill, 2011). These data would suggest that Caethg_1822 is a trimeric shell protein with an Fe-S binding centre within the central pore.

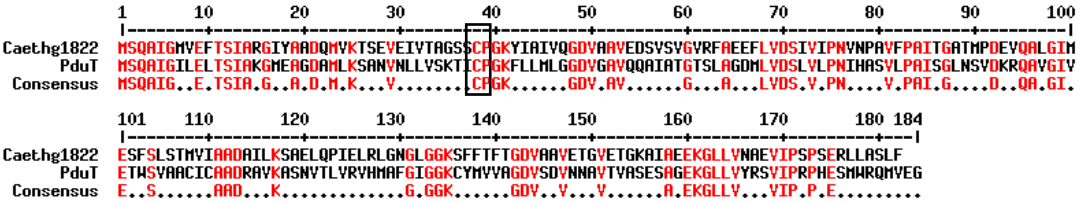


Figure 4.20. Protein sequence alignment of Caethg_1822 with PduT from *C. freundii*. A conserved cysteine residue at position 38, proposed to be involved in FeS binding is highlighted (box). Conserved residues denoted in red. Protein sequences aligned using the program Multalin (Corpet, 1988).

There are two shell proteins in the large gene cluster with FeS binding sites; a BMC-H_{FeS} and a BMC-T_{FeS}. Several functions have been proposed for these FeS-binding pores including electron transport through the shell, the movement of intact FeS clusters through the shell, and as a sensor of intracellular oxygen levels (Crowley *et al.*, 2010).

In summary, whilst there are some small BMC shell proteins that are duplicated in the BMC encoded by the larger gene cluster there are also proteins that have no orthologues within the small cluster. It was considered that the differences in shell protein composition may lead to different phenotypes when the proteins were recombinantly produced in *E. coli*. The

additional shell proteins within the large gene cluster may provide the necessary components for microcompartment assembly.

4.3.2 Over-production of individual shell proteins in *E. coli*

BL21* (DE3) pLysS *E. coli* were transformed with a vector, each containing the gene encoding for an individual shell protein. The resulting strains were grown in LB media at 37 °C until an OD_{600} ~ 0.6 was reached. Protein production was induced with 400 µM IPTG and the cultures were incubated in shaking flasks overnight at 19 °C before harvesting and processing for embedding and thin-sectioning.

4.3.2.1 Caethg_1816, Caethg_1817, Caethg_1824 and Caethg_1825

These proteins are identical to proteins encoded within the small BMC gene cluster and the phenotype of *E. coli* over-expressing these proteins are described in Section 4.2.2.6 Caethg_3290 (Caethg_1816), Section 4.2.2.5 caethg_3289 (Caethg_1817), Section 4.2.2.3 Caethg_3284 (Caethg_1824) and Section 4.2.2.2 Caethg_3283 (Caethg_1825).

4.3.2.2 Caethg_1820

Caethg_1820 is a 27 kDa protein that consists of 248 amino acids and which contains an N-terminal BMC domain with a C-terminal extension of unknown function. It shares 86 % identity with Caethg_3286, the BMC-H_{ex} of the small gene cluster (Fig. 4.21).

```
                1       10        20        30        40        50        60        70        80        90       100
                |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
Caethg1820      MQALGLIETIGLIAAIESADAMLKAADVNLLEKTYVGGGLVSIAVTGDVGAVKAAVDAGGAAVRNINDTLLVSQHVIPRPHEELDNIIVSAVPLKDRNIP
Caethg3286      MQALGLIETIGLIAAIESADAMLKAADVNLLEKTYVGGGLVSIVVTGDVGAVKAAVDAGGAAVRNINDTLLVSQHVIPRPHEELDNIIVSAVPLKDRNIK
Consensus       MQALGLIETIGLIAAIESADAMLKAADVNLLEKTYVGGGLVSI.VTGDVGAVKAAVDAGGAAVRNINDTLLVSQHVIPRPHEELDNIIVSAVPLKDRNI.

                101     110       120       130       140       150       160       170       180       190       200
                |--------+---------+---------+---------+---------+---------+---------+---------+---------+---------|
Caethg1820      TVAPAKNQPEDETTVEVIEEDTVEVPVKIEDSIVEDSIEENST----------VEDPAVEDSKEENSTVENSENEELLPEKITDPLKMDLTKLNRKAVDK
Caethg3286      TVNPEENQSKVVTMVEGNEEESVEVSLKIEELEYDDSKEEDSMIENPTIDDSIVEDPIEEDSKEENSTVENSENEELLPEKITDPLKIDLTKLNRKAVDK
Consensus       TV.P..NQ....T.VE..EE..VEV..KIE.....DS.EE.S..........VEDP..EDSKEENSTVENSENEELLPEKITDPLK.DLTKLNRKAVDK

                201     210       220       230       240       250     258
                |--------+---------+---------+---------+---------+-------|
Caethg1820      MVLDYGLEKAIEILSKFKVIELRNLARKYKNFGIKGRSISKADKKLLLIEFKKYYGHN
Caethg3286      MVLDYGLEKAIEILSKFKVIELRNLARKYKNFGIKGRSISKADKMLLLVEFRKYYGHN
Consensus       MVLDYGLEKAIEILSKFKVIELRNLARKYKNFGIKGRSISKADK.LLL.EF.KYYGHN
```

Figure 4.21. Protein sequence alignment of Caethg_1820 with Caethg_3286, the BMC-H_{ex} proteins of the large and small BMC gene clusters. Conserved residues denoted in red. Protein sequences aligned using the program Multalin (Corpet, 1988).

The BMC domains are highly conserved (99 % identity), as are the last 100 amino acids of the C-terminal extension. Most of the variation arises between residues 100 and 150, with Caethg_3286 having an extra 10 amino acids between residues 130 and 150.

When over-produced in *E. coli*, Caethg_1820 does not appear to form any structures but aggregation was observed in some cells (Fig. 4.22a). This aggregation was not concentrated at the poles as has been seen with other shell protein aggregates but was in the centre of the cells. The most striking observation, however, was that the cells appeared very unhealthy (Fig.

4.22b). The growth following induction was also greatly reduced compared to the growth of other strains post-induction (*E. coli* expressing *caethg_1820* reached an OD_{600} ~1.2 after overnight incubation, whereas cultures of other strains grown in the same batch reached an OD_{600} ~ 4.0). Over-production of Caethg_3286 in *E. coli* did not result in the same phenotype. *Caethg_3286* expressing cells appeared normal and growth was comparable with other strains. This is very surprising considering the relatively small difference in the primary sequence of these two proteins.

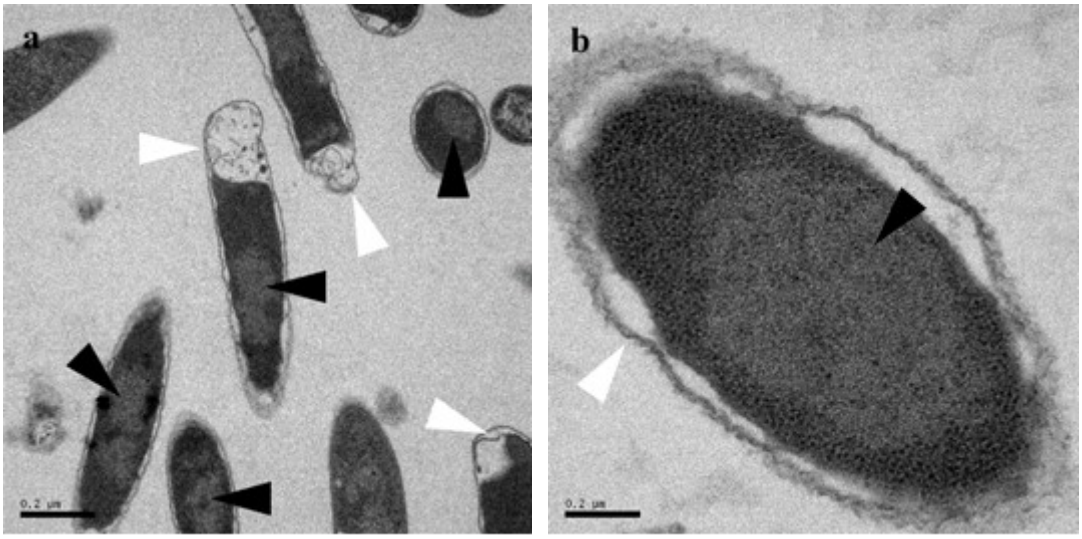


Figure 4.22. TEM images of thin sections of *E. coli* over-producing Caethg_1820. Aggregates (black arrows) and unhealthy looking cells (white arrows) are highlighted. Scale bars 0.2 µm

4.3.2.3 Caethg_1822

Caethg_1822 is an 18 kDa, 182 amino acid, trimeric shell protein that is predicted to have an iron-sulphur binding centre in the central pore (BMC-T_{FeS}). It has no homologue within the small BMC gene cluster of *C. autoethanogenum* but shares 39 % similarity with PduT.

In TEM images, *E. coli* cells over-producing Caethg_1822 appear normal with no protein structures visible unlike *C. freundii* PduT which forms aggregates when over-produced in *E. coli* (unpublished data – Dr. M.Lee).

4.3.2.4 Caethg_1831

Caethg_1831 is a 10 kDa, 100 amino acid, single BMC domain containing shell protein (BMC-H), sharing 46 % and 51 % identity with PduA and PduJ respectively. When over-produced in *E. coli*, Caethg_1831 forms insoluble aggregates mostly at the poles but sometimes distributed in patches throughout the cell (Fig. 4.23a).

4.3.2.5 Caethg_1832

Caethg_1832 is a 10 kDa, 98 amino acid, single BMC domain containing shell protein (BMC-H). It shares 44 % and 50 % identity with PduA and PduJ respectively. Caethg_1831 also shares 51 % identity with Caethg_1831. TEM images of *E. coli* over-producing Caethg_1832 show the formation of aggregates at the poles of the cells (Fig. 4.23b). No higher order structures are formed.

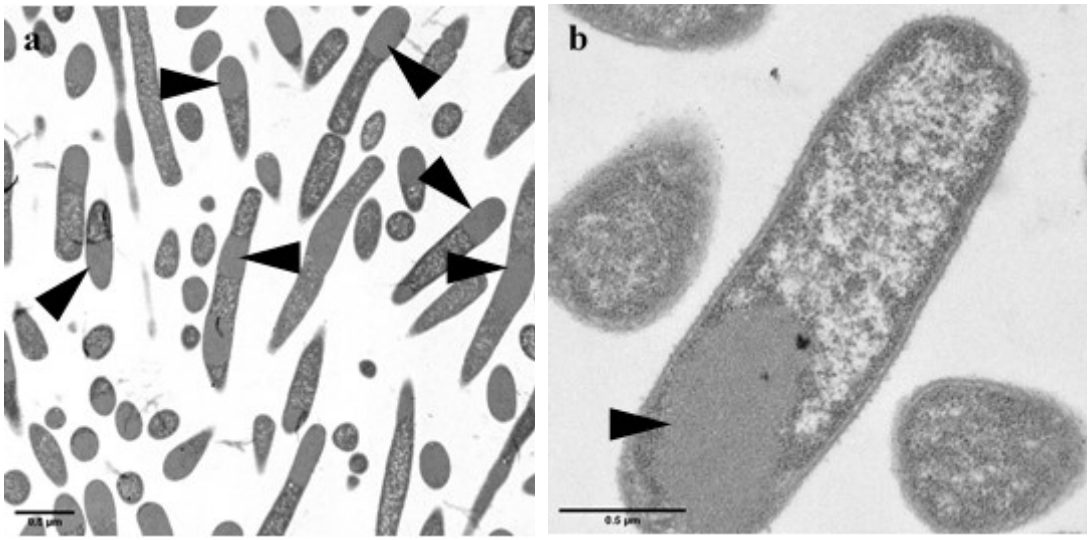


Figure 4.22. TEM images of *E. coli* over-producing a) Caethg_1831 and b) Caethg_1832. Aggregation is indicated (black arrows). Scale Bars 0.5 µm

4.3.2.6 Summary of single shell protein production

Caethg_1816 and Caethg_1817, both BMC-H proteins, were the only shell proteins of the large BMC gene cluster to form higher order structures when produced in *E. coli.* Masses of swirled sheets were observed extending throughout most cells. Insoluble aggregates were formed by three of the shell proteins, Caethg_1822 (BMC-T_{FeS}), Caethg_1831 (BMC-H) and Caethg_1832 (BMC-H). There were no visible structures formed by Caethg_1824 (BMC-P) and Caethg_1825 (BMC-H_{FeS}). Caethg_1820 (BMC-H_{ex}) also formed aggregates but in addition the cells appeared very unhealthy and growth after protein production was significantly impaired.

The phenotypes observed on production of the large BMC gene cluster shell proteins in *E. coli* were somewhat different to the phenotypes observed with the small gene cluster shell proteins. Production of all 8 shell proteins together was subsequently planned to see if this different complement of shell proteins resulted in alternative structures to those observed with the shell proteins of the small gene cluster.

4.3.3 Over-production of all *C. autoethanogenum* large GRM1 BMC shell proteins

All the individual shell protein genes of the large BMC gene cluster had previously been cloned into a pET3a vector. The 'Link-and-Lock' protocol was used in order to form a plasmid harbouring all 8 shell protein genes, each with their own ribosome binding site. However, the cloning of this plasmid proved problematic. Transformation of competent DH10ß *E. coli* repeatedly failed after ligation of plasmid fragments. This may have been due to the combinations of genes being toxic to *E. coli,* the individual genes were therefore sub-cloned into pETcoco-2, an expression vector that can be maintained in a single copy state to minimise toxicity to the host. The 'Link-and–Lock' protocol was then followed to produce a vector containing all 8 shell protein genes.

Competent BL21(DE3) *E. coli* cells were transformed with the pETcoco-2 plasmid containing all 8 shell protein genes and plated on LB agar containing ampicillin (50 μgmL^{-1}) and glucose (0.2 % w/v) to select for cells containing the plasmid and to keep the plasmid as a single copy. Starter cultures, again containing glucose, were inoculated from a single colony and grown overnight before inoculating a 50 mL culture containing ampicillin and arabinose (0.01 % w/v) to induce activation of the multi-copy origin of replication. The cultures were grown at 37 °C to an $OD_{600} \sim 0.6$. Protein production was induced with 400 µM IPTG and cultures grown overnight at 19 °C before harvesting for fixing and thin sectioning.

The colonies obtained after transformation of BL21 (DE3) cells were very small indicating the genes were possibly toxic to *E. coli*. Despite this, cultures grew at a similar rate to control cells transformed with an empty pETcoco-2 plasmid both before and after induction.

No empty microcompartment structures were formed on over-production of the 8 shell proteins of the large BMC gene cluster. Instead, as with the small BMC shell proteins, sheets of protein were formed with ribosomes aligned along many of these sheets (Fig. 4.24).

The extra proteins of the large BMC gene cluster did not appear to have an impact on the type structures formed or stop the apparent interaction with ribosomes.

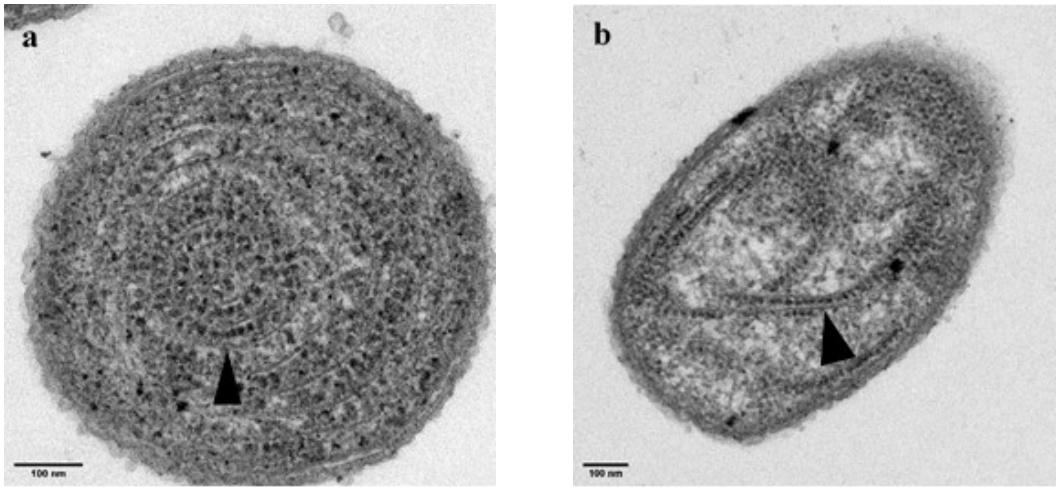


Figure 4.24. TEM images of *E. coli* over-producing all 8 shell protein of the large BMC gene cluster. Sheets of protein with ribosomes aligned are highlighted (black arrows). Scale bars 100 nm

4.3.4 Omission of one shell protein gene from multi-gene constructs

As in Section 4.2.4 constructs were cloned that each had one of the genes encoding shell protein in the large BMC gene cluster missing to investigate if the BMC-H_{ex} shell protein was necessary for the ribosome interaction to occur or if any other shell protein was involved. The constructs were cloned using the 'Link-and-Lock' protocol in a pETcoco-2 vector.

Despite using a pETcoco-2 vector, cloning of the constructs proved difficult with no colonies forming following transformation with ligated plasmids on many occasions. Most plasmids were successfully cloned after several attempts but one plasmid could not be cloned successfully therefore had to be omitted from the study (Table 4.4)

Table 4.4 Table of the constructs prepared to examine the effect of omitting one shell protein. pET3a *caethg_1832-25-24-22-20-17-16* could not be cloned therefore was omitted from further studies

Plasmid	Gene Omitted	Cloned
pET3a *caethg_1831-25-24-22-20-17-16*	*caethg_1832*	✓
pET3a *caethg_1832-25-24-22-20-17-16*	*caethg_1831*	✗
pET3a *caethg_1832-31-24-22-20-17-16*	*caethg_1825*	✓
pET3a *caethg_1832-31-25-22-20-17-16*	*caethg_1824*	✓
pET3a *caethg_1832-31-25-24-20-17-16*	*caethg_1822*	✓
pET3a *caethg_1832-31-25-24-22-17-16*	*caethg_1820*	✓
pET3a *caethg_1832-31-25-24-22-20-16*	*caethg_1817*	✓
pET3a *caethg_1832-31-25-24-22-20-17*	*caethg_1816*	✓

BL21 (DE3) *E. coli* cells, transformed with the appropriate plasmid, were processed as described in Section 4.3.3.

Colonies following transformation were very small and in some cases had to be repeated as no colonies at all were obtained on the first attempt. Again this suggests the plasmids are toxic to the *E. coli.* All cultures grew at a comparable rate to empty vector controls both pre and post induction.

Ribosomes could be seen to associate with protein sheets with all the combinations of shell proteins tested except when Caethg_1820 (BMC-H_{ex}) was missing. In this case, proteins sheets were observed but no ribosomes were aligned (Fig, 4.25). This is the same pattern that was observed with the small BMC shell proteins. Inclusion of the C-terminally extended protein results in ribosomes being associated with the protein sheets.

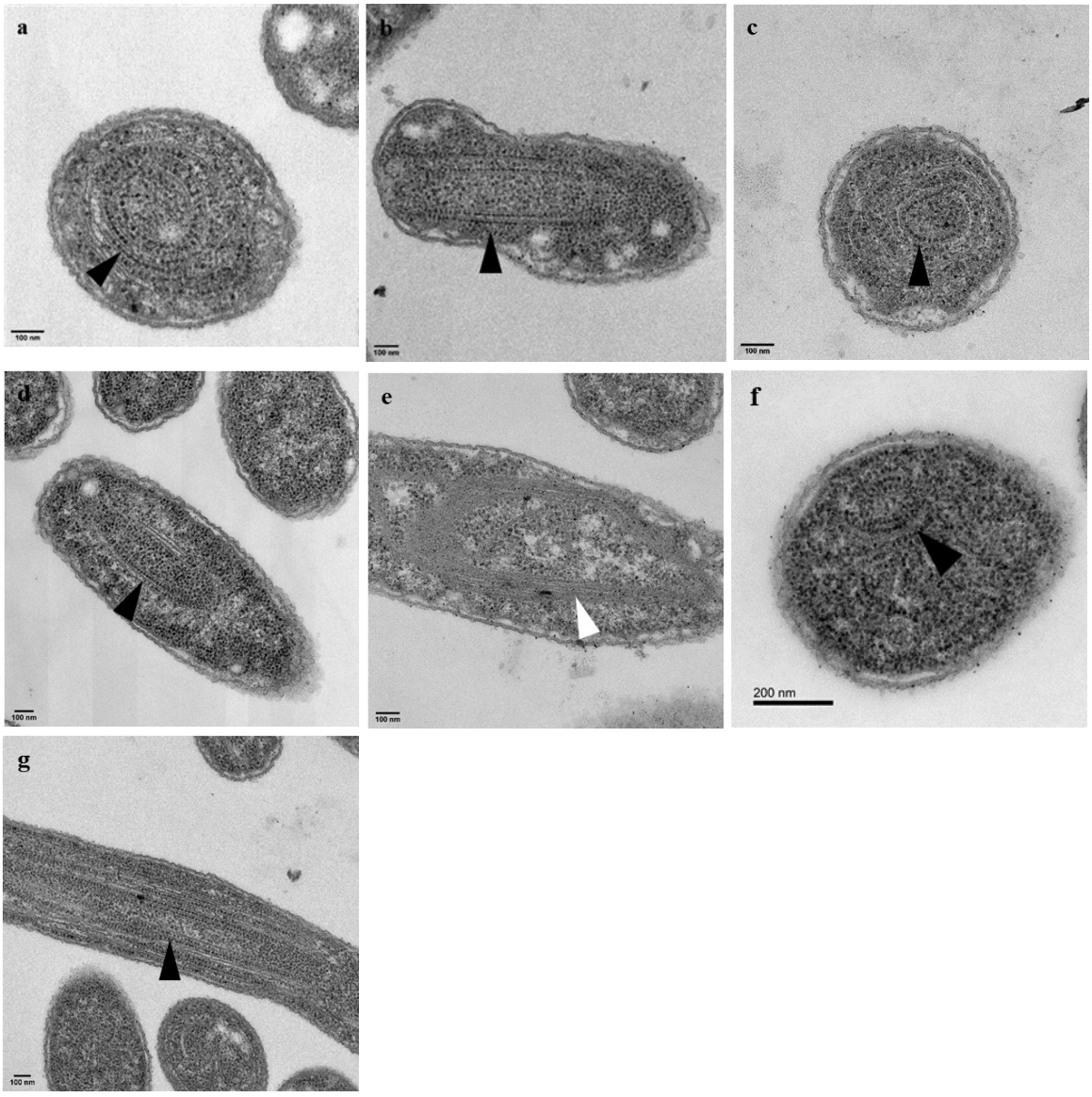


Figure 4.25. TEM images of thin sections of *E. coli* expressing 7 shell protein genes with one gene omitted. a) *caegth_1832* omitted, b) *caegth_1825* omitted, c) *caegth_1824* omitted, d) *caegth_1822* omitted, e) *caegth_1820* omitted, f) *caegth_1817* omitted, g) *caethg_1816* omitted. Ribosomes are observed associated with protein sheets with each construct tested except when *caethg_1820* was omitted. Sheets of protein with ribosomes aligned (black arrows) and swirled sheets of protein (white arrow) are indicated. Scale bars 100 nm

4.3.5 Summary of large gene cluster shell protein production in *E. coli*

Whilst some of the shell proteins are identical to those of the small gene cluster, there are also proteins with no small gene cluster orthologues that, when combined with the other shell proteins, may have resulted in the formation of empty BMCs. This was not the case. As was observed with the small gene cluster, sheets of protein with ribosomes aligned were observed.

However, some differences were noted. The BMC-H_{ex} proteins of each cluster had different characteristics when over-produced. Caethg_3286 produced no visible structures with *E. coli* and cells appeared normal. Caethg_1820 of the larger gene cluster has 87 % identity with Caethg_3286 but over-production in *E. coli* resulted in the formation of aggregates within the cells. The cells also appeared very unhealthy as evidenced by a vastly reduced growth rate following induction. A perhaps subtler difference was the difficulty experienced when cloning the multi-gene constructs of the larger GRM1 type gene cluster. Transformations following ligations were regularly unsuccessful, even after transferring to a pETcoco-2 plasmid. Tellingly, the only construct that did not prove difficult to clone was the plasmid not containing the gene for the BMC-H_{ex} shell protein Caethg_1820, which is also the protein involved in ribosome interaction.

4.4 Discussion

The aim of this study was to examine the potential for shell proteins of *C. autoethanogenum* BMCs to form empty microcompartments when produced in *E. coli*. This has been successfully demonstrated with the shell proteins of the Pdu BMC from *C. freundii* (Parsons *et al.*, 2010) but the lack of literature reporting the successful recombinant production of other empty BMC types is possibly indicative of how difficult this is to achieve. Indeed, this was highlighted in Chapter 3 with the failure to produce empty BMCs in *E. coli* upon cloning of the *R. rubrum* genes encoding the BMC shell proteins. The genome of *C. autoethanogenum* contains two gene clusters encoding BMC shell proteins, however, neither set of shell proteins were able to form microcompartments in *E. coli*.

When expressed in *E. coli*, the shell proteins of both the GRM3 and GRM1 type gene clusters formed sheets of protein extending through the cytoplasm. The outstanding feature in both cases was the observation of ribosomes that appeared to be aligned along the length of some protein sheets.

It was only when the BMC-H_{ex} genes were not included in the constructs that ribosomes were no longer associated with the protein sheets. The C-terminal extension of these proteins is much longer than that of the Pdu BMC-H_{ex} shell protein PduK and is poorly conserved, whereas the BMC domain is similar to that of other shell proteins (66 % similarity to PduA). It would seem more likely therefore that the effect on ribosome distribution observed in this study is a property of the C-terminal extension. This is explored further in Chapter 5.

The structures formed by the individual shell proteins in this study were also quite different to those formed by the Pdu shell proteins. Very few of the shell proteins formed higher order structures, most either formed aggregates or formed no visible structures at all. Only 2 shell proteins of each cluster formed structures within *E. coli* and these were extensive swirls of protein sheets that filled the cytoplasm of most cells. There was no evidence of tube or filament formation as is seen with PduA, PduJ and PduB which have been shown to be major components of the Pdu BMC. It may be the absence of proteins able to form these more ordered structures that lead to the inability to form BMCs.

The differences between different shell proteins clearly has an impact on the ability of the structures to self-assemble when these proteins are produced in combination in *E. coli*. Further studies, perhaps combining the shell proteins of different systems, to explore the impact on empty BMC formation would help build on our current understanding of BMC architecture.

Chapter 5

Characterisation of Caethg_3286, a C-terminally extended hexameric shell protein from *Clostridium autoethanogenum* that interacts with the *E. coli* ribosome

5.1 Introduction

The shell proteins of the *C. autoethanogenum* BMC have been shown to exhibit an unusual phenotype when produced in *E. coli*. Sheets of protein swirl through the cytoplasm with ribosomes aligned along the sheets (see Chapter 4). This unprecedented organisation of ribosomes, which are ordinarily randomly distributed throughout the cytoplasm, was an unexpected observation. Examination of a range of shell protein combinations revealed the proteins that caused the ribosome alignment were the C-terminally extended hexameric shell proteins (BMC-H_{ex}) from both of the BMCs encoded within the *C. autoethanogenum* genome.

BMC-H_{ex} proteins are encoded in many BMC operons including carboxysomes and the Pdu, Eut and Cut BMCs. Nothing is currently known about the function of the C-terminal extension despite the ubiquitous presence in BMCs. Whilst there is some conservation of residues amongst a number of BMC-H_{ex} shell proteins there is also a lot of variation in both the primary sequence and the length of the C-terminal extension.

This chapter describes the characterisation of Caethg_3286, the BMC-H_{ex} shell protein of the small *C. autoethanogenum* BMC, including its structure as determined by X-ray crystallography and the identity of the region within the C-terminus of the protein responsible for ribosome interaction.

5.2 Results

5.2.1 Sequence analysis of Caethg_3286

Caethg_3286 is a 258 amino acid protein consisting of a BMC domain followed by a long C-terminal extension of unknown function. BMC-H_{ex} proteins are also found in the well characterised carboxysomes, Pdu and Eut BMCs but the C-terminal extension is much shorter and there are few conserved residues. A BLAST search reveals, however, that this protein is well conserved within many clostridia species including some extremely virulent pathogens, for example *C. botulinum* and *C. tetani* (Fig. 5.1).

```
C.autoethanogenum      MQALGLIETIGLIAAIESADAMLKAADVNLLEKTYVGGGLVSIVVTGDVGAVKAAVDAGG
C.coskatii             MQALGLIETIGLIAAIESADAMLKAADVNLLEKTYVGGGLVSIVVTGDVGAVKAAVDAGG
C.ragsdalei            MQALGLIETIGLIAAIESADAMLKAADVNLLEKTYVGGGLVSIAVTGDVAAVKAAVDAGG
C.ljungdahlii          MQALGLIETIGLIAAIESADAMLKAADVNLLEKTYVGGGLVSIAVTGDVAAVKAAVDAGG
C.botulinum            MQALGLIETKGLLAAVEAADTMVKSADVSIIEKTYVGGGLVTISVTGDVGAVKASIEAGV
C.tetani               MQALGLIETKGLLAAVEAADVMVKSADVSIIEKTYVGGGLVTISVTGDVGAVKASIEAGA

C.autoethanogenum      AAVRNINDTLLVSQHVIPRPHEELDNIIVSAVPLKDRNIKTVNPEENQSKVVTMVEGNEE
C.coskatii             AAVRNINDTLLVSQHVIPRPHEELDNIIVSAVPLKDRNIKTVNPEENQSKVVTMVEGNEE
C.ragsdalei            AAVRNINDTLLVSQHVIPRPHEELDNIIVSAVPLKDRNIPTVAPAKNQSEDETTVEVIEE
C.ljungdahlii          AAVRNINDTLLVSQHVIPRPHEELDNIIVSAVPLKDRNIPTVAPAKNQPEDETTVEVIEE
C.botulinum            AAVKKLDEGFLVSEHVIPRPHEELESIIGSNTPPEDPSSNDNTENVEDTEDTEAVEKAED
C.tetani               AAVKNLNEEFLVSEHVIPRPHEELESIIGSNNLQEDSSSNEDTSSVENVEKAEVVETTEA

C.autoethanogenum      ESV------------EVSLKIEELEYDDSKEEDSMIENPTIDDSIVEDPIEEDSKEENST
C.coskatii             ESV------------EVSLKIEELEYEDSKEEDSMIENPTKDDSIVEDPIEEDSKEENST
C.ragsdalei            DTV------------EVPVKIEDSIVEDSKEENSTVENTTVE----------DSKEENST
C.ljungdahlii          DTV------------EVPVKIEDSIVEDSIEENSTVENPTVE----------DSKEENST
C.botulinum            TEKVENTKAVNSVEDTESVEVTKDIVDAKNVENTKKAETVKDVKAVDNTEHVDVDTKNNI
C.tetani               LEE---------------EQIVEDIYDSKNLK--------------------DVDTENKI

C.autoethanogenum      VENSENEELLPEKITDPLKIDLTKLNRKAVDKMVLDYGLEKAIEILSKFKVIELRNLARK
C.coskatii             VENSENEEILPEKITDPLKMDLTKLNRKAVDKMVLDYGLEKAIEILSKFKVIELRNLARK
C.ragsdalei            VENSENEELLPEKITDPLKMDLTKLNRKAVDKMVLDYGLEKAIEILSKFKVIELRNLARK
C.ljungdahlii          VENSENEELLPEKITDPLKMDLTKLNRRAVDKMVLDYGLEKAIEILSKFKVIELRNLARK
C.botulinum            ---KENQHGLDGDLDKLHKLNLENLNKEDVDNLIRQSGIEKTILILAKLKVVKLRNLARE
C.tetani               ---KENQEKLDEDLD---KVNLEKLHKNNVDNLVSKNGLEKTISILNKLKVAKLRSLAGE

C.autoethanogenum      YKNFGIKGRSISKADKMLLLVEFRKYYGHN
C.coskatii             YKNFGIKGRSISKADKMLLLVEFRKYYGHN
C.ragsdalei            YKNFGIKGRSISKADKKLLLIEFKKYYGHN
C.ljungdahlii          YKNFGIKGRSISKADKKLLLIEFKKYYGHN
C.botulinum            YKDFGIAGRTISKAGKNLLINKFKLYYEKN
C.tetani               YKDFGIKVSEISKADKNLLIKKFKLYYEKN
```

Figure 5.1. Multiple sequence alignment of BMC-H_{ex} proteins from Clostridia species identified from a BLASTp search. Residues highlighted in yellow show identity to Caethg_3286. Sequence alignment was performed with Clustal MUSCLE (Edgar, 2004).

A secondary structure analysis was performed in Phyre2 (Kelley *et al.*, 2015), which predicted several features. Caethg_3286 has a canonical BMC domain with 3 α-helices and 4 ß-sheets arranged like those of PduA and EutM (Fig. 5.2). These secondary structure elements were predicted to form the classic alpha-beta fold of the BMC domain with 100% confidence.

Figure 5.2. Secondary structure prediction of Caethg_3286. The BMC domain (residues 1 – 90) contains 4 ß-sheets and 3α-helices in the same order as the canonical BMC domain. The C-terminus extension is predicted to have a long region of disorder (residues 90 – 180) followed by a series of α-helices. Secondary structure prediction was performed using the Phyre2 webserver (Kelley *et al.*, 2015).

The studies in this chapter examine the effects of expressing just the BMC domain or just the C-terminal extension alone and in combination with the other shell proteins to see if the protein domain interacting with ribosomes could be identified. For the purposes of this study, the BMC domain was defined as residues 1-90 and the C-terminal extension as residues 91-258. Additionally, crystallisation experiments were performed with the aim of elucidating the tertiary structure of this protein.

5.2.2 Crystallisation of Caethg_3286

Then gene encoding Caethg_3286 was cloned into a pET14b vector to allow the protein to be produced with an N-terminal hexa-histidine tag. The resultant protein was purified by IMAC as described in Section 2.4.1. Most of the protein was found in the soluble fraction in the supernatant with only a small amount lost in the pellet. The elution fractions from the nickel column were tested with Bradford reagent to assess protein concentration and the two fractions containing the highest concentration were subjected to buffer exchange on a PD-10 column (GE healthcare) to remove imidazole before further purification by size exclusion chromatography. Samples of each fraction were denatured by the 1:1 addition of Laemmli sample buffer and heating to 100 °C for 10 min before separation of proteins by SDS-PAGE (Section 2.4.5.1).

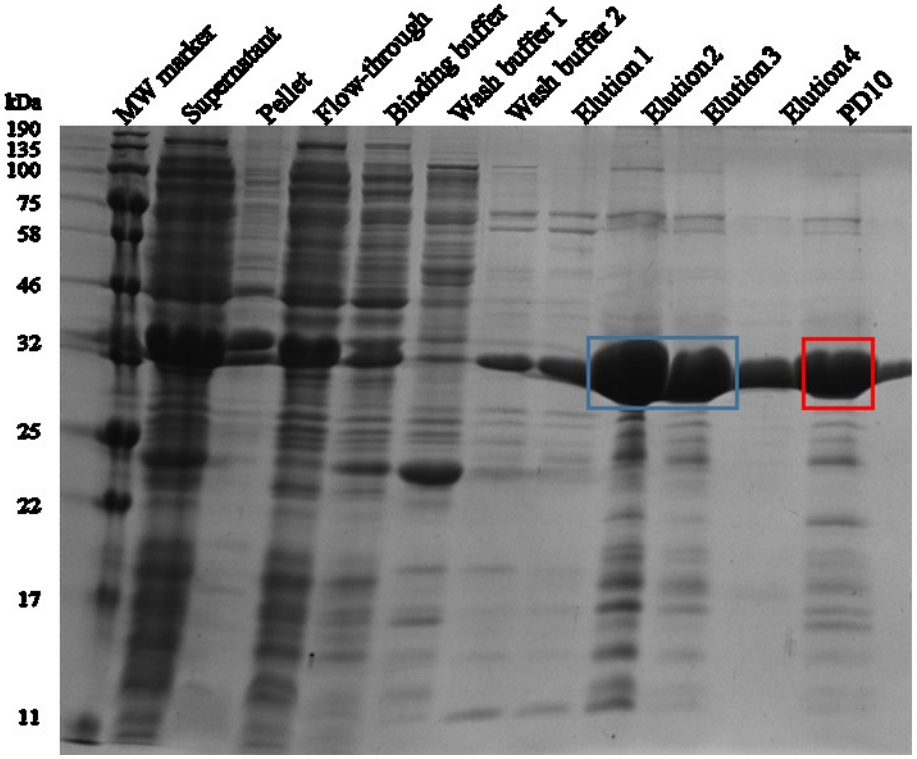


Figure 5.3. SDS-PAGE gel of the IMAC purification of Caethg_3286. The elution fractions 2 and 3, highlighted by the blue box, were applied to a PD-10 buffer exchange column to remove imidazole. The eluent from the PD-10 column (red box) was subjected to further purification by size exclusion chromatography.

In addition to the strong bands observed for Caethg_3286, there were also some smaller contaminating bands of protein in the eluted samples. Further purification was therefore attempted using size exclusion chromatography. The protein sample was concentrated to a 0.5

mL volume and applied to a Superdex® 200 Increase 10/300 column that was attached to an AKTA PURE FPLC system (GE Healthcare).

The protein was seen to elute in one large peak at 9.24 mL followed by two much smaller peaks at 13.7 and 14.9 mL (Fig. 5.4). The first peak at 9.24 mL represents the void volume of the column indicating that the protein is forming a large aggregate rather than any structured oligomer. However, the very small second peak that eluted at 13.7 mL equates to a protein of 178 kDa. This is 6.24 times the size of Caethg_3286 indicating that hexamers may be forming but only in very small numbers. A 105 kDa protein would elute at the third peak at 14.9 mL, suggesting a small amount of tetramer is formed. SDS-PAGE analysis of the eluted fractions show protein is only detectable in the large peak fractions. Further studies aimed at reducing aggregation by the inclusion of DTT (4 mM) or Triton-X (0.1 %) in buffers had no clear effect on the oligomerisation state of the protein.

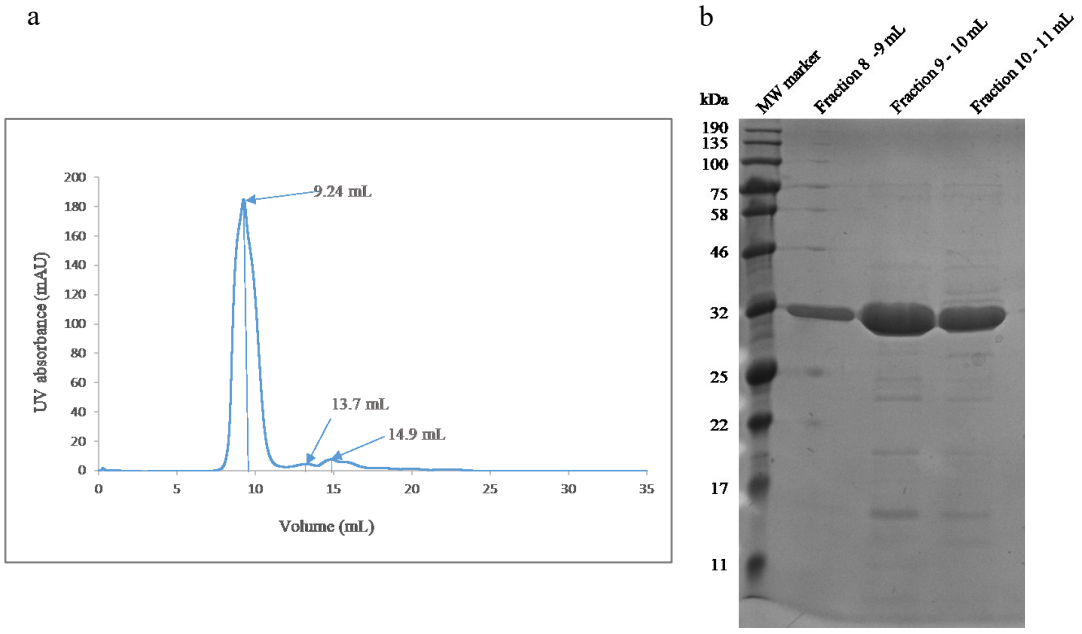


Figure 5.4. a) Size exclusion elution profile of Caethg_3286. A large peak at 9.24 mL is aggregated protein eluting at the column void volume. The small peaks at 13.7 mL and 14.9 mL are proteins forming hexamers and tetramers respectively. b) SDS-PAGE analysis of size exclusion elution fractions. Only protein from the large peak was detectable by SDS-PAGE.

Protein concentration was measured using the Bicinchoninic acid assay (Section 2.44) and fractions were concentrated using Amicon® centrifugal filters to final concentrations of 5 and 10 mg mL^{-1}. Crystallisation screens were set up using the hanging drop vapour diffusion

method as described in Section 2.4.10.1 with Structure Screens 1 and 2 from Molecular Dimensions. Crystal trays were left at 19 °C and examined regularly for crystal formation.

At day 5 small needle-like crystals were seen with two conditions: (a) 0.2 M Magnesium acetate tetrahydrate in 0.1 M Na cacodylate buffer, pH 6.5, containing 20 % (v/v) PEG 8000 and (b) 0.2 M sodium acetate trihydrate in 0.1 M Tris buffer, pH 8.5, containing 30 % (v/v) PEG4000 (Fig. 5.5).

a b

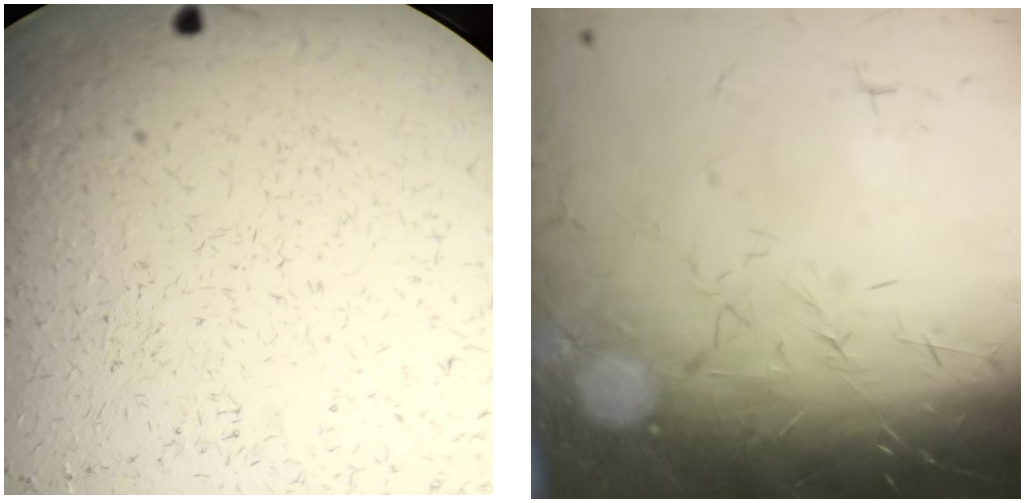


Figure 5.5. Small needle–like crystals were observed after 5 days with a) 0.2M magnesium acetate tetrahydrate/0.1M sodium cacodylate pH 6.5/ 20 % (v/v) PEG 8000; b) 0.2M sodium acetate trihydrate/ 0.1M Tris pH8.5/ 30 % (v/v) PEG4000

Screening around these conditions (buffer pH and salt concentration were varied) was unsuccessful and no crystals, even small needles, could be reproduced.

Crystal trays were observed on a daily basis for one week and then weekly for up to a month and subsequently on an ad hoc basis. It was 4 months after set-up that crystals were observed in another condition containing 2 M ammonium sulphate. Small, rectangular crystals were seen in the drop containing the lower concentration of protein and larger rectangular crystals observed at the higher protein concentration (Fig. 5.6). Crystals were collected in 10 % (v/v) glycerol in 2 M ammonium sulphate for cryo-protection before flash freezing in liquid nitrogen.

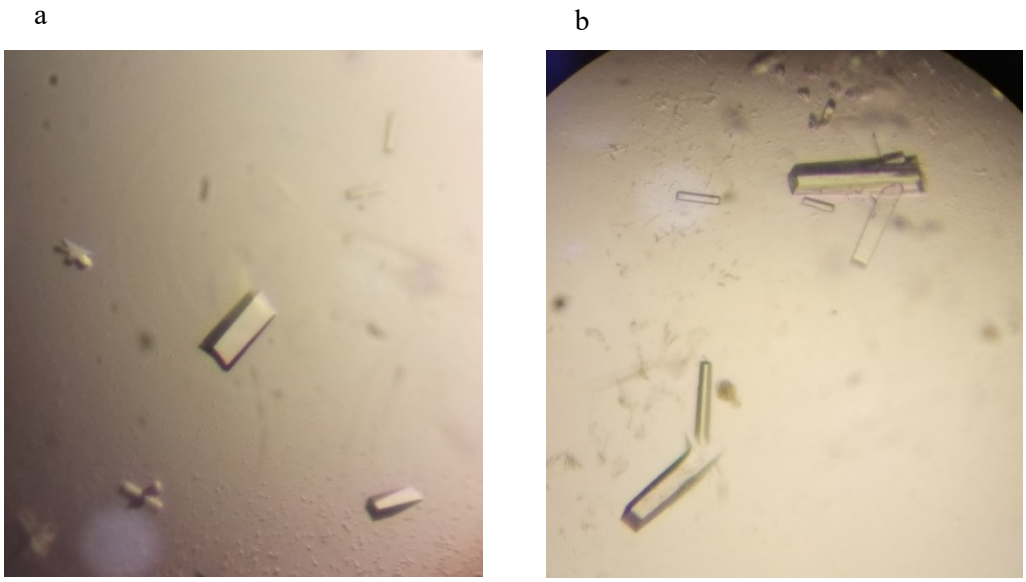


Figure 5.6. Crystallisation of Caethg_3286. Rectangular crystals were obtained in 0.2 M ammonium sulphate with protein added to drops at concentrations of a) 5 mg mL^{-1} and b) 10 mg mL^{-1}.

5.2.3 Crystallographic data collection and structure refinement

Crystallographic data was collected with the support of Professor David Brown (University of Kent/ Charles River Laboratories) at the Diamond Light source in Oxfordshire.

X-ray diffraction data was collected at a resolution of 1.88 Å. The space group of the data was P 4_2 2 2, and the unit cell dimensions were a = 66.736, b = 66.736, c = 114.200 Å. The structure was solved, with the help of Professor D. Brown, by molecular replacement with residues 2-78 of PduA from *Salmonella enterica* (PDB entry 3NGK). Data was analysed through MolRep and 3 copies of the same molecule (Fig. 5.7b) were seen to fit the unit cell. The C-terminus of the protein was missing, but residues 2 – 86 of the N-terminal BMC domain could be mapped. The 3 chains are virtually identical, superimposing with an average Rmsd of 0.38 Å. Coot and RefMac were used to refine the model reaching R = 0.21 and R free = 0.26 indicating a good fit to the electron density data. The initial maps showed 3 copies forming half of a hexamer (Fig. 5.7a). Adding the reflection of these 3 molecules gave the hexameric structure (Fig. 5.7c).

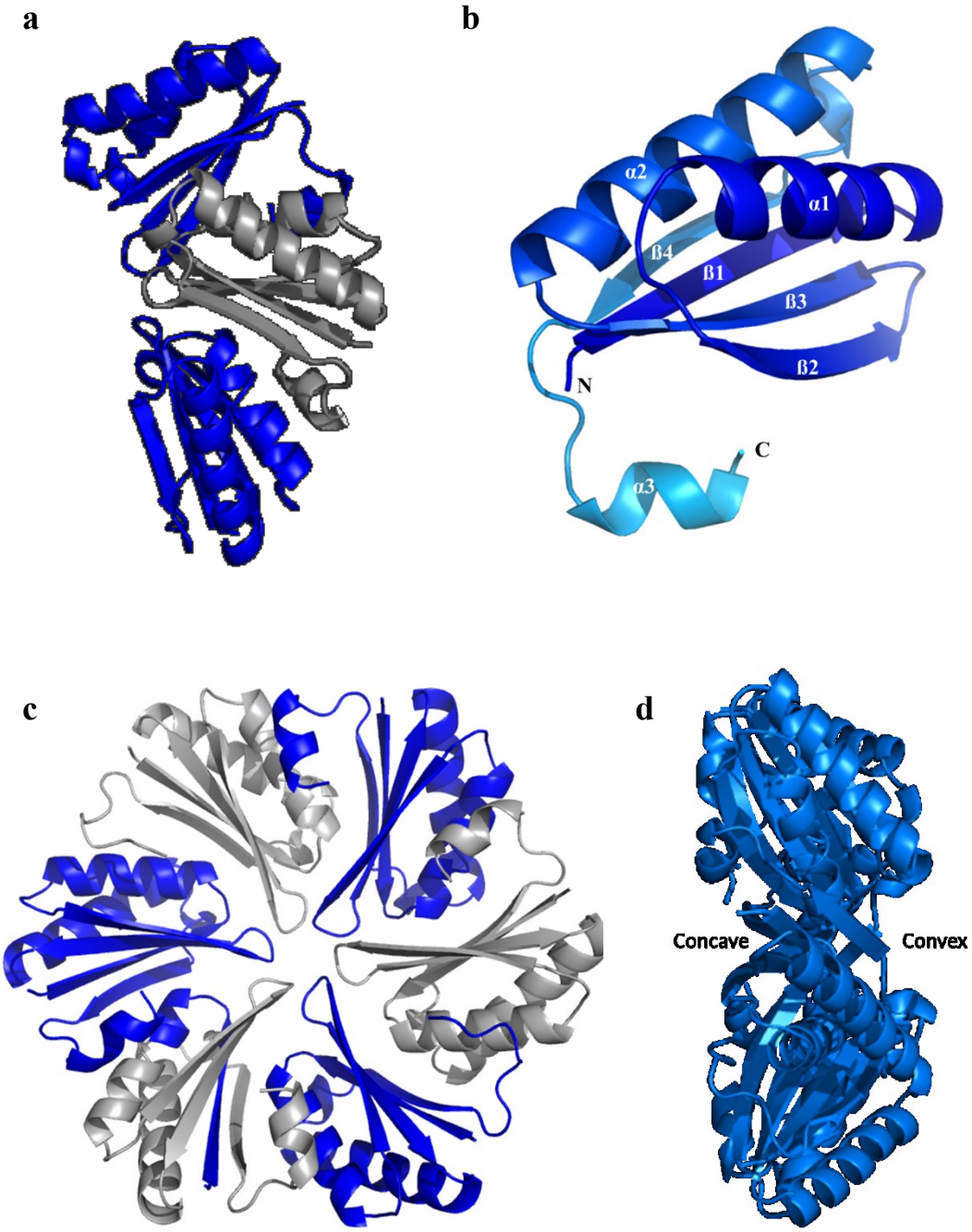


Figure 5.7. Crystal structure of the BMC domain of Caethg_3286. a) The unit cell contains three identical copies of the same protein. Alternate copies are coloured blue and grey to aid visualisation of the individual monomers. b) Single copy of the Caethg_3286 BMC domain. There are 4 ß strands (labelled ß1-ß4) and 3 α helices (labelled α1-α3). c) A hexameric structure is obtained when a reflection of the unit cell is added. d) Side view of the hexamer reveals a concave and a convex surface. The N- and C- termini are found on the concave surface.

5.2.4 **Caethg_3286 structure**

Only the N-terminal BMC domain was visible in the crystal structure of Caethg_3286. The C-terminal extension was likely to be have lost during the long incubation period before crystal formation.

The BMC domain of Caethg_3286 (Fig 5.7b) is typical of other shell protein structures with 4 antiparallel ß-sheets bordered by 3 α-helices. Analysis of the interfaces between the monomers forming the hexamer using PISA (Protein interfaces, Surfaces and Assemblies) gave low hydrophobic p-values (0.114 - 0.184) indicating that the interfaces are likely to be specific and have a high likelihood of being relevant to the protein function. This therefore suggests that a hexamer is likely to be formed in nature. The interface is formed from 8 hydrogen bonds and 2 salt bridges.

The N- and C- termini are both seen to be on the concave surface of the hexamer. The recently solved crystal structure of an intact recombinant BMC shell (Sutter *et al.*, 2017) demonstrated the concave surface to face outwards towards the cytoplasm. This would suggest that the C–terminus of Caethg_3286 would be exposed to the contents of the cytoplasm allowing an interaction with ribosomes.

Three glycine residues at positions 37-39 give flexibility to form loops which can be seen to give rise to the pore at the centre of the hexamer (Fig. 5.8a). The pore measured between opposing amide nitrogen atoms of Gly-38 is estimated at 8.1 Å in diameter, however this is probably an overestimate. Measurement of the pore in a space fill model using MOLE*online* (Pravda *et al*., 2018) gave an estimate of pore diameter to be 4.8 Å. This is within the range of pore sizes observed in other shell protein hexamers (4 -7 Å) which is proposed to be large enough to allow the passage of smaller substrates and metabolites (Chowdhury *et al.*, 2015). Mapping of the electrostatic surface charge shows most of the convex and concave surface to be negatively charged however the central pore region has a much stronger positive charge (Fig. 5.8b). The positive charge of the pore region would support the passage of negatively charges molecules.

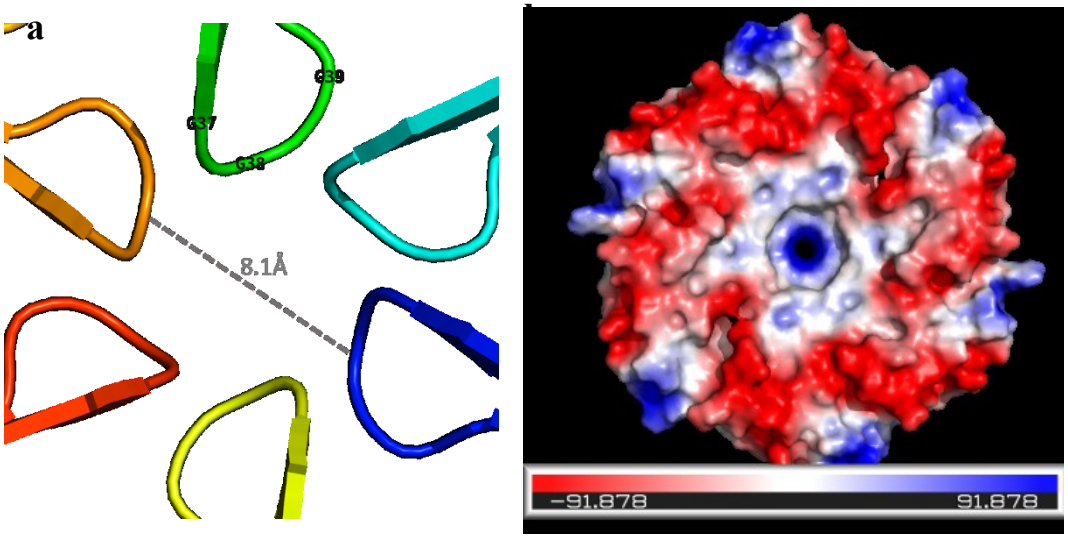


Figure 5.8. The central pore of the Caethg_3286 hexamer. a) 3 Glycine residues form the loops of the central pore with a diameter of approximately 8.1 Å. b) The surface electrostatic potential of the hexamer surface is mostly negative (red) but the central pore region is much more positive (blue).

Studies on the hexamer-hexamer interface of PduA showed Lys-25 and Arg-79 residues were important for the stability of the interaction between PduA hexamers (Crowley *et al.*, 2010). These residues are conserved within the Caethg_3286 structure (Lys-24 and Arg-79). The orientation of these residues is similar to that seen with PduA. The amino side chain of one Lys-24 interacts with the carbonyl oxygen of the Lys-24 of its symmetry partner in the adjoining hexamer (Fig. 5.9). The Arg-79 side chain of one monomer extends across and interacts with Leu-23 of the adjoining hexamer (Fig. 5.9). Residues from four monomers can be seen to contribute to the binding between two adjoining hexamers (Fig. 5.9).

In summary, Caethg_3286 exists as a hexamer in cystalline form and analysis of the interfaces between monomers suggest that these are specific interactions and are likely to exist in their natural form. No information is available about the C-terminus as this was not present in the ctrystal structure. The full crystal structure of any C-terminally extended BMC shell protein has yet to be published. However, the crystal structure of the C-terminal extension of *E. coli* EutK was resolved and seen to form a helix-turn-helix with similarity to nucleic acid binding domains (Tanaka, Sawaya and Yeates, 2010). Whilst there is only a small degree of similarity between the EutK and caethg_3286 C-terminal extensions (20 % similarity) most of the conserved regions are within the predicted helical region of Caethg_3286 (See Fig. 5.2) perhaps indicating a nucleic acid binding function for the C-terminus of Caethg_3286.

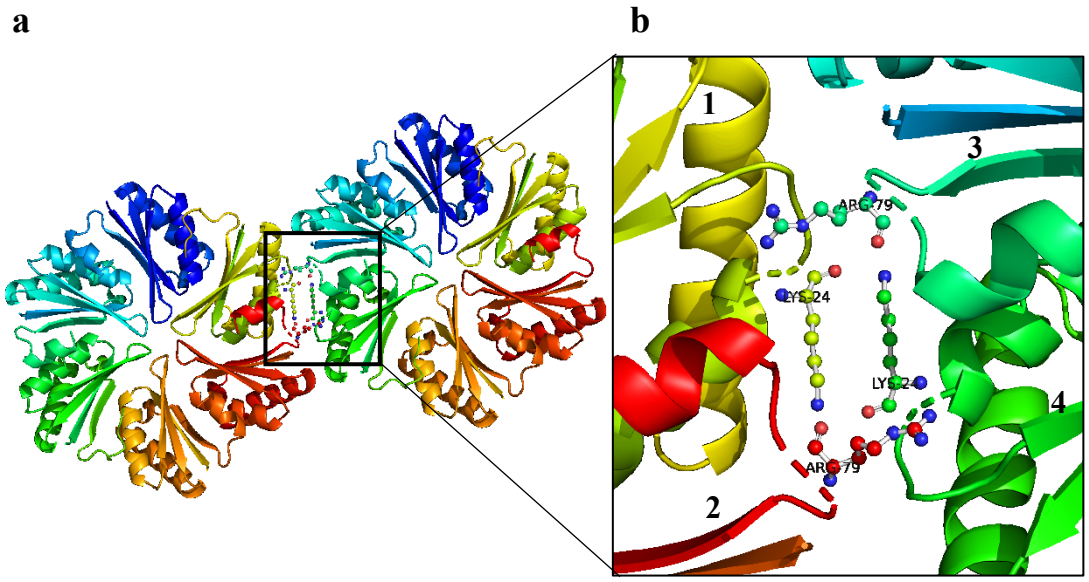


Figure 5.9. The hexamer-hexamer interface. a) Residues from 4 monomers contribute to the interaction between two adjoining hexamers. Each monomer is coloured differently to aid differentiation of monomers. b) Magnified image of the area outlined by the black box in a. Monomers involved in the interactions are labelled 1 - 4. Lys-24 from monomer 1 interacts with the anti-parallel Lys-24 of monomer 4. The side chain of Arg-79 from monomers 2 and 3 extend out and interact with Leu-23 of monomers 4 and 1 respectively.

5.2.5 Caethg_3286 truncations

5.2.5.1 BMC domain and C-terminal domain proteins

When produced with the other shell proteins of the *C. autoethanogenum* small gene cluster, Caethg_3286 interacted with ribosomes of *E. coli* resulting in sheets of protein with ribosomes aligned along the length. Studies were conducted to examine whether production of truncated versions of Caethg_3286, the N –terminal BMC domain (Caethg_3286N) or the C-terminal extension alone (Caethg_3286C), in combination with the other shell proteins were able to reproduce this phenotype.

Truncated genes encoding just the BMC domain or just the C-terminal extension of Caethg_3286 were cloned into pET3a vectors and subsequently cloned into larger constructs with the other shell protein genes using the 'Link-and-Lock' method (Fig. 5.10).

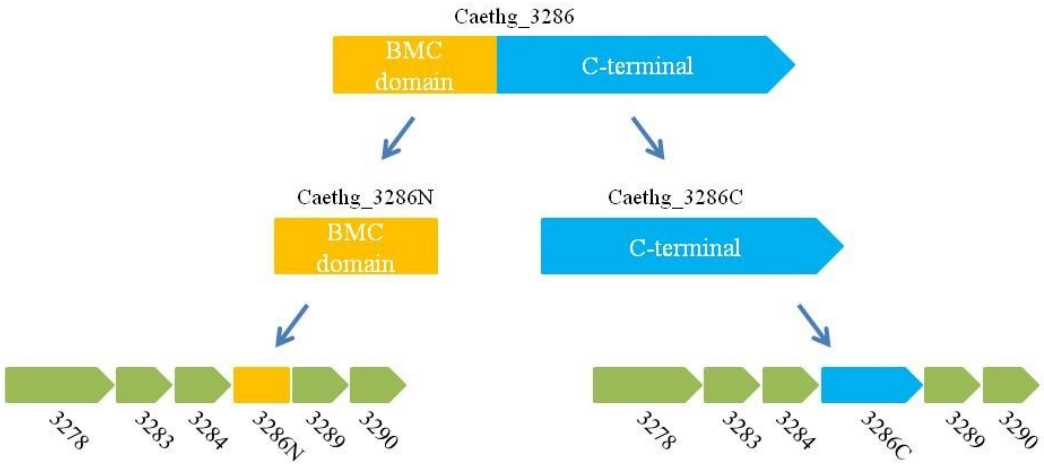


Figure 5.10. Schematic illustrating the truncated forms of Caethg_3286. Genes encoding just the BMC domain and just the C-terminal extension were cloned. The truncated genes were then 'Link-and-Locked' into multi-gene constructs with the other shell protein genes of *C. autoethanogenum*.

Plasmids containing the single genes or 'Link-and-Locked' genes were transformed into BL21*(DE3) pLysS *E. coli*, protein production induced with IPTG and samples processed for embedding and imaging by TEM as described in Section 2.4.9.2.

Caethg_3286N when over-produced in *E. coli* formed many swirled sheets of protein throughout the cytoplasm of the cells (Fig. 5.11a). Caethg_3286C did not form any visible structures when over-produced in *E. coli*. All cells appeared normal (Fig. 5.11b). The full length Caethg_3286 also did not form any visible structures within *E. coli* (See Section 4.2.2.4).

Overproduction of Caethg_3286N or Caethg_3286C in *E. coli*, with all the other shell proteins, formed swirled sheets of protein in both cases. Ribosomes were not associated with the protein sheets in either case (Fig. 5.11 c & d). Neither the N-terminal BMC domain nor the C-terminal extension alone appear to be responsible for the ribosome interaction observed with the full-length protein. The C-terminal extension may need to be attached to the BMC domain of the protein in order for it to interact with ribosomes, perhaps keeping it in a particular conformation. Studies therefore went on to explore whether the ribosomal interaction could be achieved with only part of the C-terminal extension.

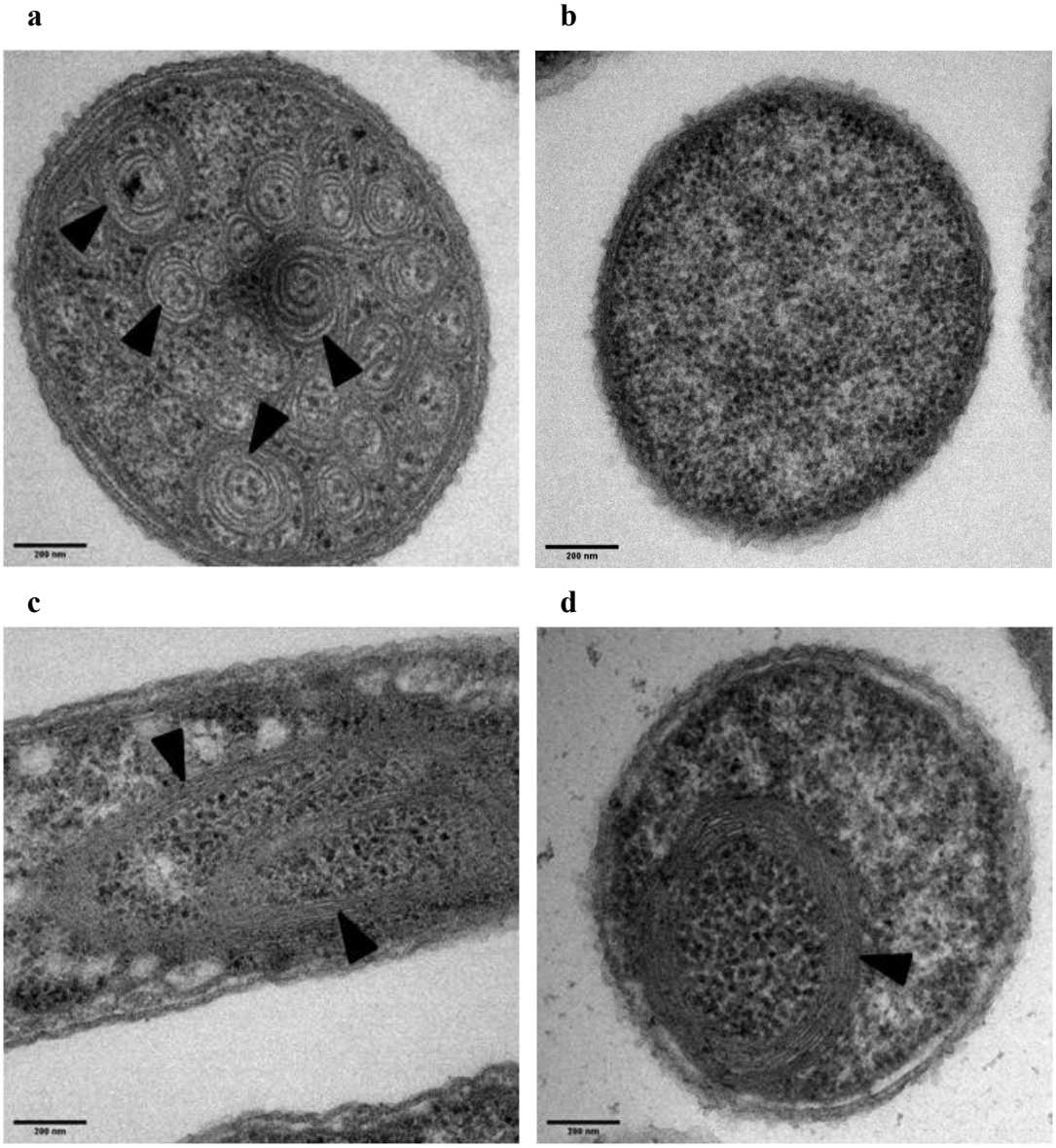


Figure 5.11. TEM images of *E. coli* expressing a) pET3a *caethg_3286N*; b) pET3a *caethg_3286C*; c) pET3a *caethg_3278-83-84-86N-89-90*; d) pET3a *caethg_3278-83-84-86C-89-90*. Swirled sheets of protein are indicated (black arrows). Scale bars 200 nm.

5.2.5.2 Caethg_3286 truncations: BMC domain with truncated C-terminal domains of varying length

Truncated forms of *caethg_3286* were amplified by PCR from pET3a *caethg_3286* using the primers described in Table 2.3 and cloned into a pET3a vector. The encoded proteins ranged in size from 80 to 242 amino acids in length (Fig.5.12). The smallest 80 amino acid protein contains the BMC domain without its final α-helix, the 90 amino acid protein is the whole of the BMC domain and all longer proteins include the whole BMC domain plus some of the C-

terminal extension. Truncated Caethg_3286 proteins were named 3286BMC_'X'; with 'X' denoting the total number of amino acids.

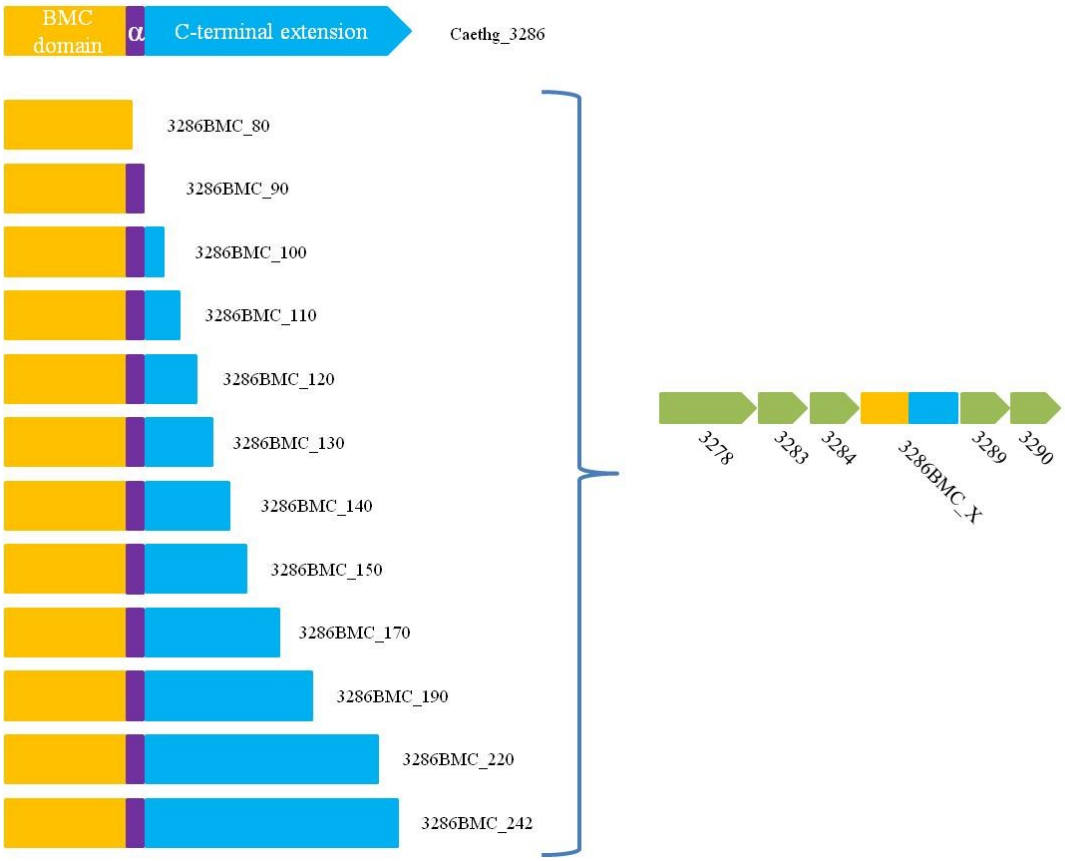


Figure 5.12. Schematic depicting the truncations of Caethg_3286 encoded by genes cloned into pET3a vectors. Each truncated protein contains the first 80 residues of the BMC domain and C-terminal extensions of varying length. The name assigned to each protein includes the total number of amino acids in the protein. Each of the genes encoding these proteins was 'Link-and-Locked' with the other shell proteins of the small BMC gene cluster.

Constructs were then made containing the truncated genes 'Link-and-Locked' with the other shell protein genes. BL21*(DE3) pLysS competent *E. coli* were transformed with the plasmid of interest and protein production induced with IPTG. Cells were processed for embedding and imaging by TEM as described in Section 2.4.9.2.

5.2.5.2.1 Overproduction of individual truncated proteins

Over-production of the individual proteins resulted in a range of structures forming within *E. coli* (Fig. 5.13). One very interesting difference was the contrast between proteins formed by the BMC domain with and without the final α-helix (80 and 90 amino acids respectively). Without the helix, the 80 amino acid peptide formed large aggregates at the poles of the cells

(Fig. 5.13a). However, when the final α-helix was present swirls of protein sheets were formed (Fig. 5.13b). This highlights the importance of the final alpha helix in the assembly of higher order structures by BMC shell proteins. Four residues in this C-terminal region were identified as contributing to the formation hydrogen bonds at the interface between monomers (PISA): His-81, Glu-83, Asn-86 and Ile-87.

Higher order structures were formed by truncated Caethg_3286 proteins up to 120 amino acids in length (Fig. 5.13 b – e). Aggregates were formed by the 130 residues long protein but *E. coli* producing proteins between 140 and 170 amino acids in length appeared normal with no visible structures within the cells. Aggregation was again observed with proteins between 190 and 242 amino acids long. Aggregated 230 amino acid long protein was the most striking. Aggregates appeared denser than with other truncations (130, 190 and 242 amino acids) and formed bands throughout the cells in many cases (Fig. 5.13k). The full-length protein did not form any structures when overproduced in *E. coli*.

The differences observed when overproducing Caethg_3286 truncations at various lengths highlights the significant influence the C-terminal extension has on the ability of the protein to form higher order structures and whether soluble or insoluble protein is produced. How these changes affect the tiling of multiple shell proteins was investigated next by expressing constructs containing the other *C. autoethanogenum* shell proteins together with each of the truncated forms of Caethg_3286.

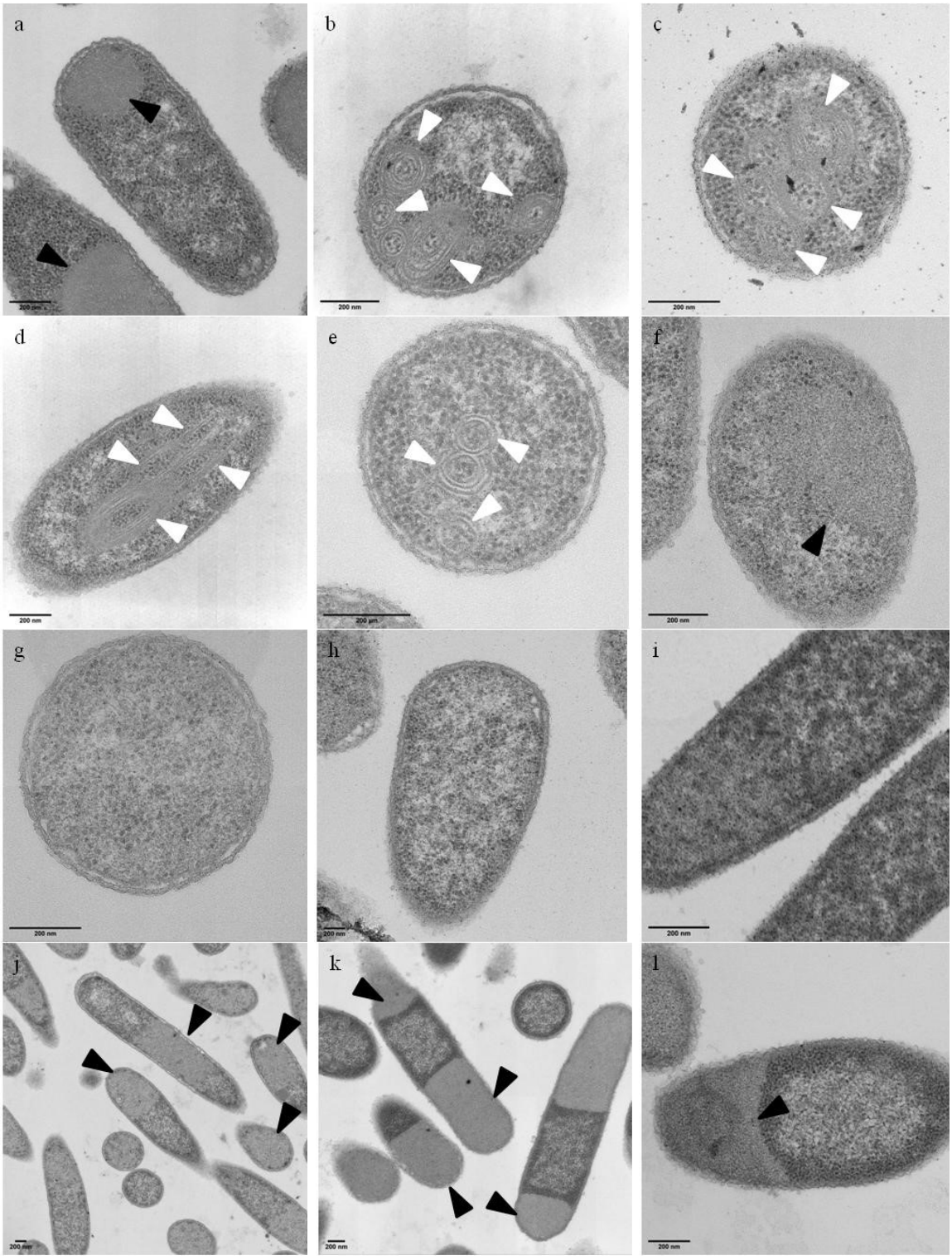


Figure 5.13. TEM images of *E. coli* overproducing of truncated forms of Caethg_3286. All proteins produced included the BMC domain, 80 amino acids, followed by a variable length C-terminus. a) 80 a.a. b) 90 a.a. c) 100 a.a. d) 110 a.a. e) 120 a.a. f) 130 a.a. g)140 a.a. h) 150 a.a. i) 170 a.a. j) 190 a.a. k) 230 a.a. l) 242 a.a. Protein sheets (white arrows) and aggregated protein (black arrows) are indicated. Scale bars 200 nm.

5.2.5.2.2 Production of truncated Caethg_3286 with all shell proteins

Caethg_3286, a BMC-H_{ex} shell protein, was previously shown to interact with ribosomes when produced in *E. coli* in conjunction with the other shell proteins of the *C. autoethanogenum* BMC. This was not seen when either the just BMC domain or just the C-terminal extension were produced with the other shell proteins suggesting the C-terminal domain needs to be attached to the BMC domain in order for the interaction to occur. A range of truncated forms of Caethg_3286 were produced with the other shell proteins to examine if a particular region of the protein was responsible for interacting with ribosomes.

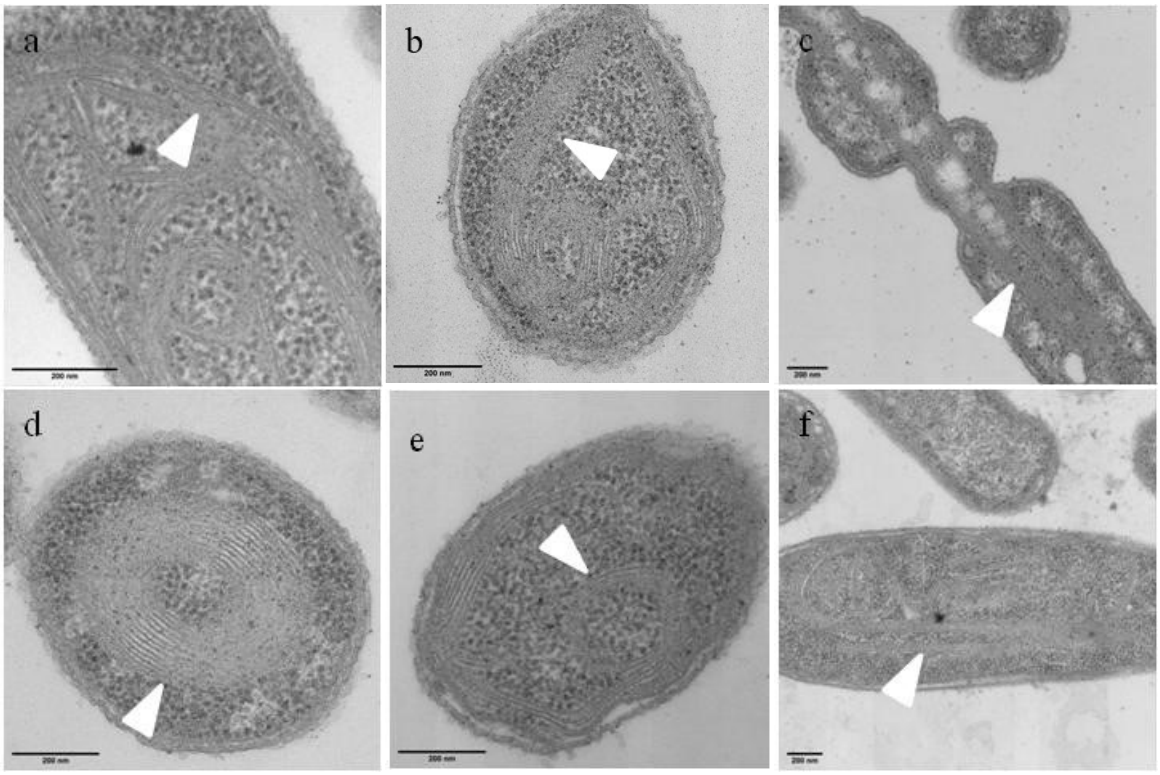


Figure 5.14. Examples of *E. coli* producing truncated BMC-H_{ex} proteins together with all the other shell proteins of the *C. autoethanogenum* BMC. Swirled sheets of protein were observed in all instances (white arrows). Images shown are *E. coli* transformed with a) pET3a *caethg_3278-83-84-86BMC_80-89-90*; b) pET3a *caethg_3278-83-84-86BMC_120-89-90*; c) pET3a *caethg_3278-83-84-86BMC140-89-90*; d) pET3a *caethg_3278-83-84-86BMC_170-89-90*; e) pET3a *caethg_3278-83-84-86BMC_230-89-90*; f) pET3a *caethg_3278-83-84-86BMC_242-89-90.* Scale bar 200 nm.

Swirled sheets of protein were produced in *E. coli* when all of the truncated forms of the protein were produced with the other *C. autoethanogenum* shell proteins. There were no interactions observed between protein sheets and ribosomes with any of the truncations tested (Fig. 5.14 a – f for examples). The phenotypes observed were indistinguishable from each other for all Caethg_3286 truncations tested, regardless of the structures formed by the

individual proteins. *E. coli* transformed with a plasmid containing all the full-length shell proteins was included in these experiments as a positive control and ribosomes were seen aligned along protein sheets. These results indicate either that the whole length of the C-terminal extension is required for protein-ribosome interaction with or that the protein domain not included in the longest truncation of 242 amino acids is the crucial domain for ribosome association.

5.2.5.2.3 Caethg_3286 BMC domain plus the end of the C-terminal extension

To explore this second possibility, proteins were designed that included the full BMC domain of Caethg_3286 (the first 90 amino acids) followed by the last 28 amino acids (Caethg_3286BMC-C28) or the last 63 amino acids (Caethg_3286BMC-C63). These regions of the C-terminal extension include 2 and 5 predicted alpha helices respectively (See Fig. 5.2). All of the disordered region at the beginning of the C-terminal extension was excluded.

Constructs were made by splicing the two regions using overlap extension PCR as described in Section 2.3.8. Briefly, the BMC domain was amplified from pET3a *caethg-3286N* (section 5.2.5.1) by PCR with a 3' extension, introduced via the reverse primer, of the beginning of the C-terminal region to be added. The C-terminal region was amplified from pET3a *caethg_3286C* (Section 5.2.5.1) by PCR with a 5' extension, introduced via the forward primer, of the end of the BMC domain. The products of these PCR reactions were used as templates for a second round of PCR. The 1st round products when annealed act as primers for each other to be extended by polymerase producing the final desired PCR product (Fig.5.15).

The successful cloning of these products into a pET3a vector was confirmed by sequencing. The 'Link –and Lock' protocol was then used to make constructs containing these spliced genes with the other shell protein genes of the BMC. This was successful for the construct containing *caethg_3286BMC-C28* (pET3a *caethg_3278-83-84-(86BMC-C28)-89-90*), however, the final gene (caethg_3290) failed to ligate into the vector so, therefore, only a 5 gene construct could be formed: pET3a *caethg_3278-83-84-(86CBMC-C63)-89.*

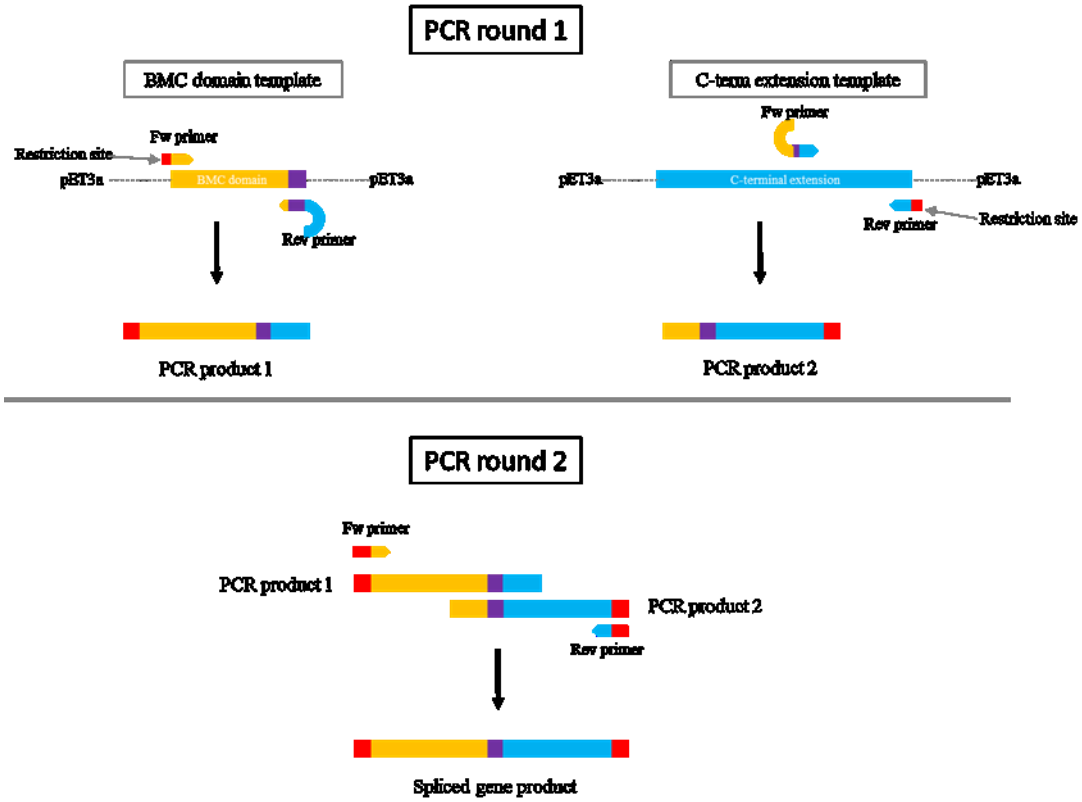


Figure 5.15. Schematic of the overlap extension PCR protocol. In the 1st round of PCR, primers are used that modify the ends of the products so that they have matching sequences. In the 2nd round of PCR, these products are denatured and then re-anneal so that the top strand of product 1 overlaps the bottom strand of product 2. These act as primers and are extended by polymerase to form the spliced product. The forward and reverse primers then anneal to the recombinant product in the following PCR cycles so that this spliced gene product is amplified.

BL21* (DE3) pLysS competent *E. coli* were transformed with the vectors (single gene and multi-gene constructs), grown to an OD_{600}~0.6, induced with IPTG and processed for embedding and thin sectioning.

When overproduced in *E. coli*, Caethg_3286BMC-C28 formed dense aggregates, mainly at the poles of the cells (Fig 5.16a). Caethg_3286BMC-C63 also formed aggregates but these were not located at the poles and also appeared to be groups of smaller patches of aggregated protein (Fig. 5.16b). Many cells also looked unhealthy with extended areas of periplasm visible.

E. coli expressing pET3a *caethg_3278-83-84-(86BMC-C28)-89-90* formed sheets of protein that swirled throughout the cytoplasm. There was no interaction observed between the sheets of protein and ribosomes within the cells. Expression of pET3a *caethg_3278-83-84-(86BMC-*

C63)-89 resulted in the formation of protein sheets in the cytoplasm (Fig. 5.16c). There were no clear instances of ribosomes associating with the protein sheets (Fig. 5.16d).

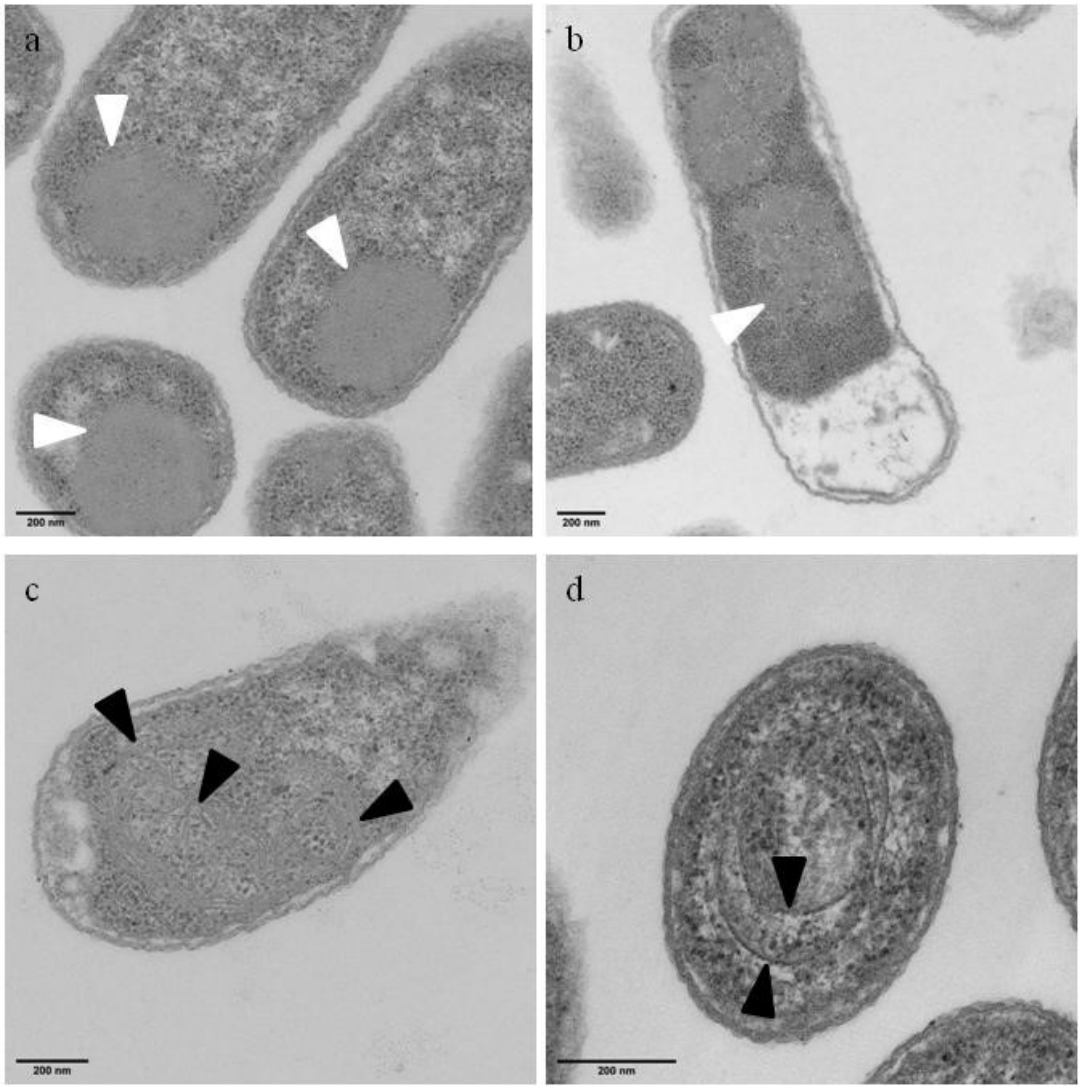


Figure 5.16. TEM images of *E. coli* expressing a) pET3a *caethg_3286BMC-C28*, b) pET3a *caethg_3286BMC-C63,* c) pET3a *caethg_3278-83-84-86BMC-C28-89-90, d)* pET3a *caethg_3278-83-84-86BMC-C63-89.* Aggregation (white arrows) and protein sheets (black arrows) are indicated. Scale bars 200 nm.

The failure to observe ribosomes aligning along the sheets of protein in this experiment suggests that the full-length protein is necessary for the interaction to occur. The long region of predicted disorder may provide a degree of flexibility within the protein that is required for the incorporation of the BMC domain into protein sheets whilst allowing the C-terminus to interact with ribosomes within the cytoplasm.

5.2.5.3 Splicing C-terminus of Caethg_3286 onto PduK BMC domain

Studies examining truncations of the C-terminal extension were unable to identify any particular domain that contributed to the ribosome binding observed in studies with the full-length protein. It was concluded that the full-length protein was necessary for an interaction with ribosomes to occur. In order to see if this ribosome binding could be observed if the C-terminal extension was added to a different shell protein, overlap extension PCR was again used to splice the C-terminal coding domain of *caethg_3286* onto the BMC domain of *pduK* (see Section 2.3.8), denoted Pdu(K-86C). This spliced gene was expressed with the remaining Pdu shell protein genes PduA, B, J, N, U (Fig. 5.17).

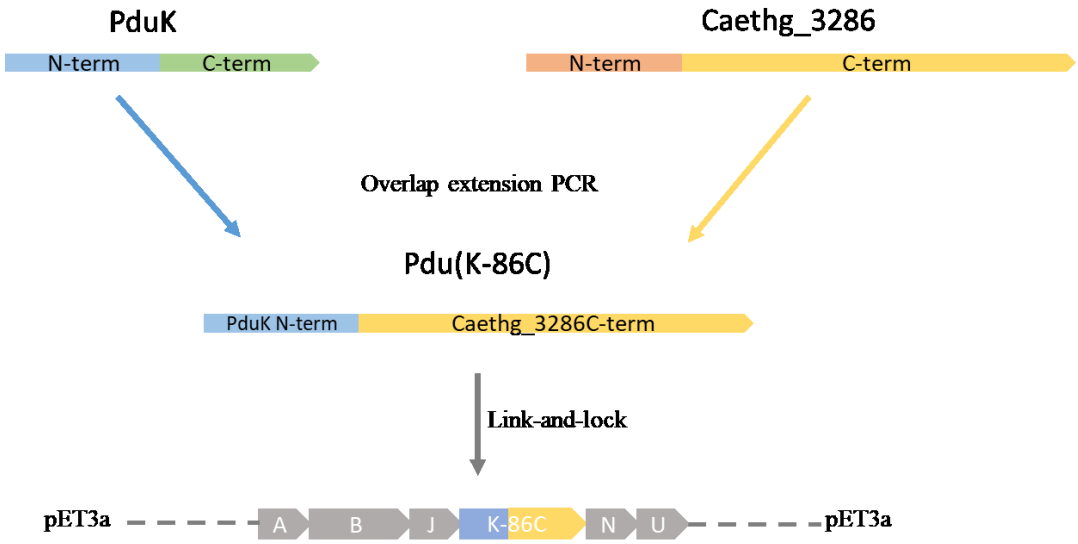


Figure 5.17. Schematic depicting the production of a spliced gene containing the *pduK* BMC domain and the *caethg_3286* C-terminal extension. Overlap extension PCR was used to produce a spliced gene product, *pdu(K-86C)*. *Pdu(K-86C)* was 'link-and-locked' with other Pdu shell protein genes to create pET3a *pduABJ(K-86)NU*.

Previous studies have shown the formation of empty microcompartments upon expression of pET3a *pduABJKNU.* This experiment aimed to see the effect of adding the C-terminal extension to PduK on microcompartment formation. The plasmids used in this study were: pET3a *PduABJKNU,* pET3a *pduABJ(K-86C)NU* and pET3a *pduABJNU*. BL21* (DE3) pLysS competent *E. coli* were transformed with the plasmid of interest and processed for protein production, embedding and thin-sectioning as previously described in Section 2.4.9.

As previously observed, PduABJKNU formed empty compartments when produced in *E. coli* (Fig. 5.18a). When PduK was replaced by Pdu(K-86C) compartment-like structure could be observed, although many mis-formed compartments were also seen. In a very small number

of cells (<1%) ribosomes were seen aligning around and along the structures formed (Fig. 5.18b). When PduK was absent only swirled sheets of protein were observed and no compartments were visible (Fig. 5.18c).

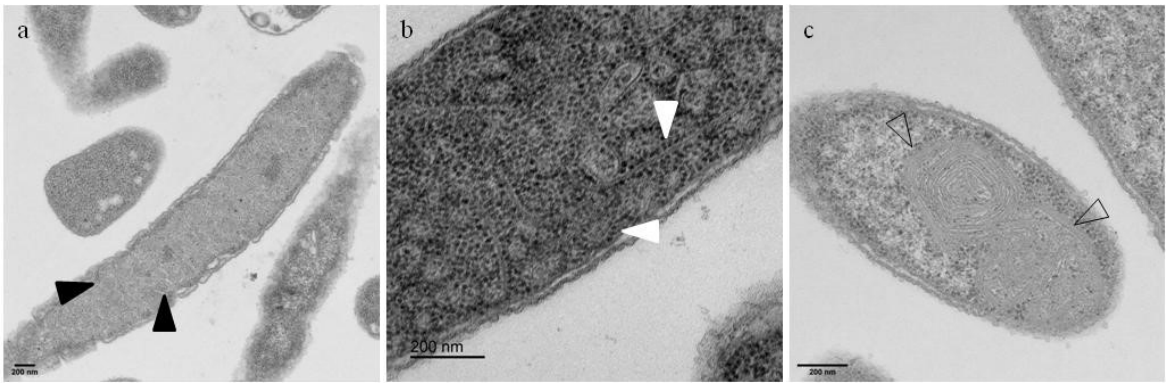


Figure 5.18 TEM images of *E. coli* expressing a) pET3a *pduABJKNU*, b) pET3a *pduABJ(K-86C)NU*, c) pET3a *pduABJNU*. a) Empty microcompartments are formed in large numbers in many cells (black arrows.) b) Ribosomes appear to be aligned along sheets of protein and around the outside of microcompartment-like structures in some cells (white arrows). c) When no BMC-H_{ex} protein is present, microcompartments are not formed. Instead, sheets of protein are seen swirled in the cells (open arrows). Scale bars 200 nm.

PduK is clearly essential for Pdu empty microcompartment formation in *E. coli*., as its absence had a profound affect with no compartments forming without its presence (Fig. 5.18c), although in *S. enterica* a *ΔpduK* mutant was shown to form aggregated BMCs (Cheng *et al.*, 2011). Changing the C-terminal extension of PduK had a less pronounced effect. Compartment like structures were still observed but there were many more mis-formed compartments observed than with the standard PduABJKNU compartments. The apparent binding of ribosomes was seen only in a very small number of cells, less than 1%, suggesting that this effect is limited. It would be interesting to examine the effect of Pdu(K-86C) on the formation of a minimal microcompartment formed with just PduA, K and N. Whilst a BMC-H_{ex} protein appears necessary for the formation of microcompartments, the exact role it plays and the role and position of ribosome binding site of Caethg_3286 in particular is still unclear from these results.

5.3 Discussion

Caethg_3286 was shown in Chapter 4 to cause the association of ribosomes with protein sheets when produced together with the other shell proteins of the BMC. This chapter described the experiments to characterise this protein and attempt to identify the domain that was responsible for this phenomenon.

Crystallography studies were successful in the formation of crystals that diffracted. Unfortunately, the structure of the whole protein could not be resolved as the C-terminal extension was cleaved during the period of crystal formation. Examination of the predicted secondary structure reveals a long series of approximately 100 residues that are disordered. It is possible that the presence of this disordered region provided too much flexibility to allow crystal formation therefore it was only after loss of this region that crystals were able to form. Determination of the structure of the BMC domain was successful however and this revealed Caethg_3286 to form the canonical BMC domain fold consisting of 4 Beta strands and 3 alpha helices. Residues in the final alpha helix contribute to hydrogen bonding between monomers to form the hexamer. This was supported by TEM evidence. Production of the BMC domain without this final alpha helix formed aggregates in *E. coli*, whereas when the alpha helix was present sheets of protein were formed indicating that hexamers assembled and could tile to form extended protein sheets.

Other features of the BMC domain were consistent with other previously described shell protein structural features. Residues shown in PduA to be essential for the formation of the hexamer-hexamer interface were conserved in Caethg_3286 and the position in the crystal structure would suggest the same role.

The similarities between the BMC domain of Caethg_3286 and other shell proteins would suggest that it would be unlikely that it is this domain that is providing the properties leading to the interaction with ribosomes. Indeed, this was confirmed by the absence of ribosome binding observed when Caethg_3286N was produced with the other BMC shell proteins. However, production of Caethg_3286C with the other shell proteins also failed to induce the binding of ribosomes to the protein sheets. This was presumed to be because, as Caethg_3286C is not attached to any of the shell proteins when produced in this way, it was not incorporated into the protein sheet and therefore unable to bring the ribosomes into the vicinity of the protein sheet.

It was hoped that the production of the BMC domain with varying lengths of the C-terminal extension would allow identification of the domain interacting with ribosomes. This was not

the case, however, as only the full-length protein was seen to cause the interaction. Production of proteins excluding the intrinsically disordered central region also failed to identify the necessary region for ribosome binding. Futures studies could examine the incorporation of some of this disordered region to see if this impacts ribosome binding. In summary, Caethg_3286 was identified as the protein interacting with ribosomes in *E. coli*, and the C-terminal extension appears to be responsible for this. Studies with truncated C-terminal extensions indicate that it is the terminal region of the extension that binds to the ribosome but at least some portion of the initial disordered region is required for the phenotype to be observed.

It is interesting that this protein is highly conserved among the *Clostridia* species, in particular in some extremely virulent pathogens, for example *C. botulinum* and *C. tetani*, organisms that give rise to botulism and tetanus, respectively. It has been suggested that the ability to form microcompartments gives some bacteria a metabolic advantage in particular environments, for example enteric pathogens such as *Salmonella* (Jakobson and Tullman-Ercek, 2016). It may be that this unique property of this BMC-H_{ex} shell protein provides some advantage to bacteria producing shells with these proteins incorporated. Why the protein interacts with ribosomes in this way remains to be established but it is possibly a translational control mechanism limiting the production of BMCs within cells. The structure of the C-terminal extension of EutK strongly supports it being a nucleic binding protein (Tanaka *et al.*, 2010) perhaps again mediating control of BMC production but through binding of DNA. Determining the function of these BMC-H_{ex} proteins, which appear crucial to BMC formation in some systems, would perhaps provide valuable knowledge for the future bio-engineering of recombinant BMC systems.

Chapter 6

Discussion

Subcellular compartmentalisation within bacteria has only been recognised relatively recently. BMCs were first recorded in the 1950s (Drews and Niklowitz, 1956) but their function as an organelle housing enzymes was not appreciated until almost 20 years later (Shively *et al.*, 1973). Carboxysomes were the first class of BMC to be investigated (Shively *et al.*, 1973) but it was soon recognised that BMCs existed that housed enzymes for other functions and were termed metabolosomes (Brinsmade *et al.*, 2005). Since then, great progress has been made in understanding them, both functionally and structurally. The rapid advances in genome sequencing allowed bioinformatic studies to identify the widespread presence of BMC encoding regions in bacterial genomes and identify a new class of microcompartment that had not been previously recognised, namely glycyl radical enzyme microcompartments (GRMs) (Jorda *et al.*, 2013). Despite being the most prevalent class of BMC loci, there is limited experimental data on GRMs. This study focussed on the GRMs identified in two bacteria, *Rhodospirillum rubrum* and *Clostridium autoethanogenum*. The shell proteins of these organisms were recombinantly produced and the structures formed were investigated. Later studies examined the role of the C-terminal extension of a BMC-H_{ex} protein in the formation of novel intracellular structures. Additionally, growth conditions for wild–type *R. rubrum* were modified to examine whether induction of microcompartment formation could be achieved.

6.1 Characterisation of the *Rhodospirillum rubrum* GRM

The first BMC investigated was that of *R. rubrum*, a Gram-negative facultative anaerobe. The BMC encoding gene cluster of *R. rubrum* was classified as a GRM3 (Axen, Erbilgin and Kerfeld, 2014) which is predicted to encapsulate the pathway for 1,2-PD metabolism. The gene cluster encodes 6 shell proteins: 2 BMC-H, 1BMC-T, 1 BMC-H_{ex} and 1BMC-H_{FeS}.

6.1.1 Production of single shell proteins

Overproduction of the individual shell proteins in *E. coli* revealed a number of protein architectures. Rru_A0905 (BMC-H), Rru_A0908 (BMC-H) and Rru_A0915 (BMC-H_{FeS}) all appeared to form filamentous structures when seen in longitudinal sections, however, their 3D structure varied. Rru_A0905 forming tubes and rolled sheets, Rru_A0908 formed tightly layered rolled sheets and Rru_A0915 forming flat sheets.

The tubes formed by Rru_A0905 did not appear to form a honeycomb lattice when viewed in cross section as was seen for PduA (Parsons *et al.*, 2010; Pang *et al.*, 2014) but this may just be due to the tubes not packing as tightly as PduA. The diameter of these tubes were slightly wider (27.8 ± 2.4 nm) than PduA tubes (20.4 ± 1.1 nm) (Pang *et al.*, 2014). The successful targeting of proteins to an intracellular scaffold of tubes has been demonstrated to increase production of ethanol in a two enzyme pathway by 221 %. It may also be possible to exploit the Rru_A0905 tubes in a similar manner, the increased diameter of these tubes perhaps

allowing the packaging of more enzymes into the interior and so the possibility of longer enzyme pathway or more enzyme molecules/tube. Targeting to any of the *R. rubrum* shell proteins has not been demonstrated in this study therefore future studies would need to investigate this to enable progress in bioengineering of Rru_A0905 tubes.

The structures formed by Rru_A0908 were also interesting as high magnification images of transverse sections revealed sheets that folded or rolled around but that also seemed to be packed in tight layers. CcmK2, a carboxysomal BMC-H shell protein, has been shown to form a dodecameric unit with two hexamers interacting between their concave faces (Samborska and Kimber, 2012). However, the residues forming the α-helical extension through which the CcmK2 hexamers interact are not conserved in Rru_A0908. Therefore, if there is a similar stacking of hexamers for Rru_A0908 so forming the observed layers, interactions must be mediated through alternative residues. Attempts to crystallise Rru_A0908 were unsuccessful (data not shown), therefore, how the hexamers interact could not be determined.

Rru_A0905 and Rru_A0908 both form sheets that curve or roll whereas Rru_A0915 forms flat sheets, as evidenced by straight filaments in both longitudinal and transverse sections. It is possible that Rru_A0915 is only incorporated into the centre of a facet where no curvature is required, whereas Rru_A0905 and Rru_A0908 may be incorporated towards the edges of facets where curvature is necessary. The curvature may result from a number of factors: (i) the tight packing of the over-produced protein sheets within the confines of the cell may just force the protein to fold around itself; (ii) the contacts between the hexagonal building blocks may not form flat interfaces but may join at an angle; and (iii) the shell protein hexamer may not form a flat disc but may have a bend. The crystal structure of the recombinant *H. ochraceum* BMC (Sutter *et al.*, 2017) revealed contacts between hexameric tiles could be at a 30º angle providing curvature to the structure. Hexamer-hexamer contacts at angles such as this could provide the curvature to form rolled sheets or tubes as seen with Rru_A0905 and Rru_A0908. The crystal structure of several hexameric shell proteins have been determined and the majority, whilst having a concave and convex surface, are essentially flat and have been shown to tile into flat sheets (Kerfeld, 2005; Tsai *et al.*, 2007; Takenoya *et al.*, 2010). However, the crystal structure of EutS revealed the hexameric assembly to have a bend of approximately 40º (Tanaka *et al.*, 2010) which was suggested to contribute to the formation of the edges of the facets where a bend is necessary to form an icosahedron. EutS is, however, a circularly permuted BMC-H protein and no other hexamers have been shown to exhibit such bent structures therefore it is unlikely that either Rru_A0905 or Rru_A0908 hexamers bend in such a way. Only determination of their crystal structures could verify this conclusively, however, neither protein could be successfully crystallised (data not shown).

6.1.2 **Production of multiple shell proteins**

This study attempted to form empty microcompartments through the production of all of the shell proteins present in the *R. rubrum* BMC gene cluster in *E. coli* as demonstrated with the Pdu BMC shell proteins of *C. freundii* (Parsons *et al.*, 2010). However, empty BMCs were not evident *in vivo*. Instead, straight sheets, rosettes and rolled sheets were observed. What was surprising, however, was that upon purification, closed microcompartment-like structures could be observed both in negatively stained samples and in thin sectioned samples. There were still mis-formed and open structures present but there were numerous apparently closed compartments. This result suggests that the formation of protein structures using the hexameric building blocks is not static with structures remaining intact once formed but, instead, may be a dynamic process whereby hexagonal building blocks can disassemble and re-form alternative structures under changing conditions. The flexible nature of facet assembly has been demonstrated using high speed atomic force microscopy (HS-AFM) (Sutter *et al.*, 2016). The formation of protein sheets was visualised by imaging the same area of mica surface after the addition of purified BMC-H in solution, every 17 seconds. Distinct hexamers could be distinguished and these were seen to both associate with and dissociate from the extended sheet over the time course of the recording. This flexibility in the formation of structures could be why compartments were seen in *in vitro* purified samples but not *in vivo*.

Several studies have examined the effect of changing conditions upon the stability of shell protein structures. Changing pH and salt concentration was seen to affect both the size of the protein sheet formed by BMC-H hexamers and the number of dynamic shell protein interactions when examined by HS-AFM (Faulkner *et al.*, 2019). The formation of nanotubes by RmmH (a BMC-H protein from *Mycobacterium smegmatis*) was similarly affected by both pH and salt concentration (Noël *et al*., 2016). Additionally, the stability of isolated microcompartments was also seen to be influenced by pH (Kim *et al*., 2014). The difference both in pH and salt concentration between the *in vivo* and *in vitro* conditions of this study may have allowed the formation of closed compartments from the disassociated hexameric components of the *in vivo* assemblies when exposed to the *in vitro* conditions. These structural differences between *in vivo* and *in vitro* assemblies were observed for both the full shell and the minimal shell constructs tested in this study. Studies examining the *in vitro* assembly of BMCs and other protein assemblies have advanced recently with the development of techniques to prevent self-assembly of individual components during purification (Hagen *et al.*, 2018). The removal of tags preventing self-assembly allowed the formation of 3 different microcompartments *in vitro*. The greater ability to control the *in vitro* environment therefore makes microcompartment assembly *in vitro* an attractive option.

6.1.3 The formation of microcompartments in *R. rubrum*

The BMC gene cluster of *R. rubrum* was identified through bioinformatics. However, microcompartments have not been observed in any functional studies with *R. rubrum.* In all previously observed metabolosomes, the pathway substrate needs to be present to induce microcompartment formation. The BMC gene cluster in *R. rubrum* encodes a GRM3 class of microcompartment, with 1,2-PD as its predicted substrate. The signature GRE of another GRM3 cluster in *Rhodopseudomonas palustris* BisB18 was shown to selectively metabolise 1,2-PD in *in vitro* studies, confirming its activity as a B_{12}-independent 1,2-PD dehydratase (Zarzycki *et al.*, 2017). The *R. rubrum* GRE, encoded by *rru_A0903*, shares 87% identity (93 % similarity) with the *R. palustris* GRE, strongly supporting the prediction of the *R. rubrum* BMC substrate being 1,2-PD. Therefore, this study examined the structures observed by TEM following both anaerobic and aerobic growth of *R. rubrum* with 1,2-PD as a carbon source.

When grown aerobically, compartment-like structures were not observed in any cells, regardless of the carbon source tested. However, anaerobic growth resulted in the formation of microcompartment-like structures under a number of different conditions. Surprisingly, polyhedral inclusions were observed in approximately 10 % of cells when grown on 0.05 % (w/v) sodium succinate as the sole carbon source. No compartments were observed when a higher concentration of sodium succinate was used. *Acetonema longum* is reported to produce BMCs when starved or grown on 1,2-PD (Tocheva *et al.*, 2014). It is possible that the low succinate concentration employed here put the cells into a 'starvation' state where the cells responded with the formation of compartments. When 1,2-PD (0.2% or 0.4 % (v/v)) was included in the media, either with or without 0.05 % (w/v) succinate, microcompartment–like structures were observed in approximately 20 % of cells. This doubling of production would support the prediction that 1,2-PD is the substrate for the signature enzyme of this microcompartment. The natural habitat of *R. rubrum* includes aquatic environments such as lakes, streams and standing water, which could provide a ready source of 1,2-PD formed from the breakdown of plant sugars by other bacteria.

Whilst it is tempting to conclude that the structures observed in these cells are microcompartments, there are a number of factors that would be prudent to consider before reaching such a conclusion. Firstly, the number of cells exhibiting compartment-like structures is low. Whilst the percentage of cells exhibiting structures has not generally been reported, examination of published images shows many of the cells within the frame of view to contain compartments – certainly more than the 20 % seen here. Secondly, in most instances, only 1–to-2 compartments are visible in any one cell. Again this seems a small number when compared to number of compartments observed in other organisms. Finally, attempts to detect

microcompartment shell proteins by SDS-PAGE and Western blot were unsuccessful. This final point may be a consequence of the small number of compartment-like structures produced. Increasing the concentration of 1,2-PD in the media may stimulate greater production of compartments, therefore this would be a good starting point for future studies.

Future studies could also examine whether transcription of the BMC genes are upregulated when *R. rubrum* is grown on 1,2-PD. Measurement of 1,2-PD and its metabolism products by HPLC would also provide additional evidence of 1,2-PD utilisation.

6.2 Characterisation of the *Clostridium autoethanogenum* GRM shell proteins

The shell proteins encoded by the GRM gene clusters of *Clostridium autoethanogenum* were studied next. There are two BMC encoding regions within the genome of *C. autoethanogenum*, one classified as GRM1 (choline utilising) and one GRM3 (1,2-PD utilising). The shell proteins of both gene clusters were studied but the greater focus was on the shell proteins of the smaller GRM3 cluster.

6.2.1 Production of single shell proteins

There was a significant amount of duplication in the shell proteins encoded by the two gene clusters. In fact, 3 of the shell proteins in each cluster were identical and a fourth shared 99 % identity with its homologue in the second cluster. Two of these BMC-H shell proteins, Caethg_3289/Caethg_1816 and Caethg_3290/Caethg_1817, were the only ones observed to form any higher order structures when over-produced in *E. coli*. The two shell proteins forming structures also shared 93 % identity with each other. The proteins appeared to form an abundance of swirled sheets that extended throughout the whole cell, often interfering with cell division. Any order to the folding or position of the protein sheets relative to each other was not apparent. This is possibly because there was such an abundance of protein that it was compacted together within the confines of the cell. No tubes or layered sheets were observed in any cell. All other shell proteins either formed insoluble aggregates or formed no protein structures at all. This would suggest that these two proteins are key components of the BMC as Pdu shell proteins that form higher order structures when produced alone, for example PduA and PduB, are major components of the Pdu BMC.

6.2.2 Production of multiple shell proteins

No microcompartments were formed upon production of all the BMC shell proteins from either gene cluster. Instead rolled sheets of protein were observed with ribosomes seemingly aligned along or trapped between the sheets. This was observed with shell proteins of both the large and small gene clusters. This is a very unusual phenotype as *E. coli* ribosomes are

normally randomly distribute throughout the cytoplasm. There are examples of ribosomes appearing to form paired lines, or polysomes, during ribosome stalling (Cougot *et al.*, 2014), but they still have a very different appearance to the alignment of ribosomes along the protein sheets as seen in this study (Fig. 6.1). Nevertheless, the possibility of ribosome stalling due to rare codons within the *C. autoethanogenum* shell protein sequences was explored by expressing the genes in an *E. coli* strain with a plasmid producing tRNAs for some of these rare codons, Rosetta-2 pLysS. The same phenotype was observed in Rosetta-2 pLysS *E. coli* cells, indicating the ribosome alignment was not a result of ribosome stalling due to rare codons.

Examination of the literature highlighted a similar phenotype in *E. coli* cells that were depleted of elements of the membrane targeting machinery. This resulted in the formation of endoplasmic membrane networks with ribosomes bound (Herskovits *et al.*, 2002) (Fig. 6.1). In this case, the ribosomes were trapped at their primary membrane docking site as the downstream mechanisms were blocked. In the case of protein synthesis, the ribosome dissociates from the peptide chain when the stop codon is recognised by a release factor. Absence of a stop codon would prevent release of the newly formed protein. The construct containing the 'link-and locked' shell protein genes was sequenced to double check that stop codons had not been lost or mutated during cloning. However, sequencing confirmed that the stop codons were present.

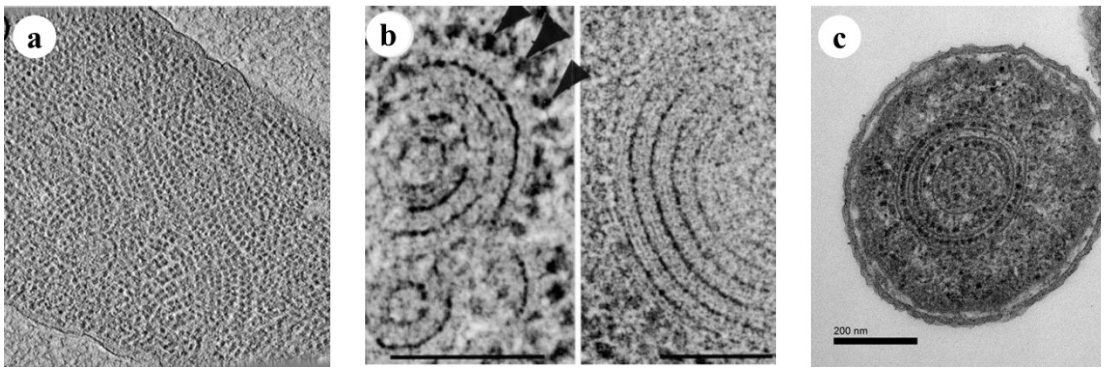


Figure 6.1. Examples of *E. coli* strains with unusual ribosome phenotypes. a) *E. coli* forming polysomes in double rows (Image reproduced from Cougot *et al.*, 2014). b) *E. coli* forming a membranous network with ribosomes attached (Image reproduced from Herskovits *et al.*, 2002.) c) Image of *E. coli* producing *R. rubrum* shell proteins with ribosomes associated with the protein sheets (this study).

Constructs each missing a single shell protein gene were cloned in order to identify if one particular protein was responsible for the observed phenotype. Only absence of the BMC-H_{ex} proteins resulted in no interaction between the ribosomes and protein sheets. Obtaining the

same result in both systems was strong evidence that the BMC-H_{ex} protein was responsible for causing ribosome interaction with the protein structure.

The BMC-H_{ex} proteins from each gene cluster share 87 % identity; however, despite this, when they were over-produced in *E. coli* they exhibited different phenotypes. Caethg_3286 did not form any structures and cells all appeared normal. Caethg_1820, the large GRM1 cluster BMC-H_{ex} protein, formed aggregates when over-produced. In addition to this, cell growth was significantly reduced following induction of protein production in comparison to other strains grown in the same experiments. TEM images revealed unhealthy looking cells. Caethg_1820 appears to be toxic to *E. coli*, which was also evidenced in the difficulty of cloning multi-gene constructs containing the *caethg_1820* gene. Alignment of the 2 proteins revealed the main differences were in a region just after the BMC domain suggesting this may be region responsible for the toxicity. Further work was carried out with the shell proteins of the small GRM3 cluster because of the difficulties encountered working with constructs of the GRM1 gene cluster.

6.2.3 Characterisation of the BMC-H_{ex} protein Caethg_3286

The final part of this study examined the protein identified as mediating the association of ribosomes with the protein sheets, Caethg_3286, a BMC-H_{ex} shell protein. X-ray crystallography was used to determine the structure of this protein and various clones were constructed to explore whether a particular region of the BMC-H_{ex} protein was interacting with ribosomes.

Unfortunately, only the structure of the BMC domain of Caethg_3286 was resolved. The C-terminal extension was missing, probably cleaved during the long incubation before crystal formation. The crystal structure of the BMC domain revealed it to be similar to the structure determined for other BMC-H proteins with 4 anti-parallel beta-sheets surrounded by 3 alpha-helices. Like other shell proteins with a single PF00936 domain, six monomers form a cyclic hexamer. However, what effect the C-terminal extension has on the hexamer is still to be determined. The C-terminus of Caethg_3286 was seen to be on the concave face of the hexamer. The crystal structure of a BMC and cryoEM structure of a carboxysome have both revealed the concave faces to face the cytoplasm (Sutter *et al.*, 2017, 2019). This suggests the C-terminal extension would be exposed to the cytoplasm allowing interaction with ribosomes or other components of the cell.

Numerous truncated forms of Caethg_3286 were produced but the ribosome associated phenotype was only evident when the full length protein was produced with the other shell proteins. The structural data showing the BMC domain to be typical of other shell proteins

together with the TEM data with truncated forms of the protein led to the conclusion that the C-terminus is crucial for the interaction between protein sheets and ribosomes. The final part of the extension, from residues 230 – 258, appears to be key but these studies were unable to identify the specific residues involved in the interaction with the ribosome.

This interaction with ribosomes is very interesting and may point to a role for this protein in the translational control of BMC formation. The only other C-terminal extension that has been investigated is the EutK C-terminus. The EutK C-terminal extension was crystallised and was seen to form a helix-turn-helix structure (Tanaka, Sawaya and Yeates, 2010). 90% of the hits from similarity searches were of nucleic acid binding domains, strongly suggesting that the C-terminus of EutK has a role in nucleic acid binding. A consistent feature of the nucleic acid binding protein hits and the EutK C-tail was a prominent positively charged surface. The final portion of the C-terminus of Caethg_3286, which appears to be important for the ribosome interaction, has a large proportion of positively charged residues (25 % residues are Arg and Lys) which may be an important feature for nucleic acid binding. Interestingly, the surface of the ribosome has a negative charge and has been implicated in binding cytoplasmic proteins and slowing their diffusion (Schavemaker, Śmigiel and Poolman, 2017). This may be a possible mechanism for the interaction between ribosomes and proteins in this study. Success in the crystallisation of EutK C-terminus is encouraging and it may be possible to crystallise the C-terminus of Caethg_3286. Determination of the structure of the C-terminal extension of Caethg_3286, and other extensions, may help towards understanding the role of these proteins in BMC biogenesis.

6.3 Conclusion

These studies have highlighted the complex nature of the interactions between shell proteins in the formation of microcompartments. The ability of *R. rubrum* shell proteins to form empty BMCs *in vitro* despite formation *in vivo* not being observed has highlighted the flexible nature of shell protein assembly and the influence of environmental conditions on this process. Interactions with other cellular components has also been highlighted in this study. The interaction between shell proteins and ribosomes may provide a clue as to how BMC formation in a cell is controlled.

References

AbdulRahman, F. (2013) 'The distribution of polyhedral bacterial microcompartments suggests frequent horizontal transfer and operon reassembly', *Journal of Phylogenetics & Evolutionary Biology*. 01(04), p. 118. doi: 10.4172/2329-9002.1000118.

Abrini, J., Naveau, H. and Nyns, E. J. (1994) '*Clostridium autoethanogenum, sp. nov*., an anaerobic bacterium that produces ethanol from carbon monoxide', *Archives of Microbiology*. 161(4), pp. 345–351. doi: 10.1007/BF00303591.

Aussignargues, C., Paasch, B. C., Gonzalez-Esquer, R., Erbilgin, O. and Kerfeld, C. A. (2015) 'Bacterial microcompartment assembly: The key role of encapsulation peptides', *Communicative & Integrative Biology*, 8(3), p. e1039755. doi: 10.1080/19420889.2015.1039755.

Axen, S. D., Erbilgin, O. and Kerfeld, C. A. (2014) 'A Taxonomy of Bacterial Microcompartment Loci Constructed by a Novel Scoring Method', *PLoS Computational Biology*. 10(10), p. e1003898. doi: 10.1371/journal.pcbi.1003898.

Backman, L. R. F., Funk, M. A., Dawson, C. D. and Drennan, C. L. (2017) 'New tricks for the glycyl radical enzyme family.', *Critical reviews in biochemistry and molecular biology*. 52(6), pp. 674–695. doi: 10.1080/10409238.2017.1373741.

Badger, M. R., Hanson, D. and Price, G. D. (2002) 'Evolution and diversity of CO_2 concentrating mechanisms in cyanobacteria', *Functional Plant Biology*. 29(3), p. 161. doi: 10.1071/PP01213.

Badger, M. R. and Price, G. D. (2003) 'CO_2 concentrating mechanisms in cyanobacteria: molecular components, their diversity and evolution', *Journal of Experimental Botany*. 54(383), pp. 609–622. doi: 10.1093/jxb/erg076.

Balkwill, D. L., Maratea, D. and Blakemore, R. P. (1980) 'Ultrastructure of a magnetotactic spirillum', *Journal of Bacteriology*, 141(3), pp. 1399–1408.

Baumgart, M., Huber, I., Abdollahzadeh, I., Gensch, T. and Frunzke, J. (2017) 'Heterologous expression of the *Halothiobacillus neapolitanus* carboxysomal gene cluster in *Corynebacterium glutamicum*', *Journal of Biotechnology*. 258, pp. 126–135. doi: 10.1016/j.jbiotec.2017.03.019.

Bertin, Y., Girardeau, J. P., Chaucheyras-Durand, F., Lyan, B., Pujos-Guillot, E., Harel, J. and Martin, C. (2011) 'Enterohaemorrhagic *Escherichia coli* gains a competitive advantage by using ethanolamine as a nitrogen source in the bovine intestinal content', *Environmental Microbiology*. 13(2), pp. 365–377. doi: 10.1111/j.1462-2920.2010.02334.x.

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 B12-dependent 1,2-propanediol degradation', *Journal of Bacteriology*, 181(19), pp. 5967–5975.

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', *Proceedings of the National Academy of Sciences*, 109(2), pp. 478–483. doi: 10.1073/pnas.1108557109.

Brandl, H., Knee, E. J., Fuller, R. C., Gross, R. A. and Lenz, R. W. (1989) 'Ability of the phototrophic bacterium *Rhodospirillum rubrum* to produce various poly (β-hydroxyalkanoates): Potential sources for biodegradable polyesters', *International Journal of Biological Macromolecules*, 11(1), pp. 49–55. doi: 10.1016/0141-8130(89)90040-8.

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', *Journal of Bacteriology*, 187(23), pp. 8039–8046. doi: 10.1128/JB.187.23.8039-8046.2005.

Brown, M. R. W. and Kornberg, A. (2004) 'Inorganic polyphosphate in the origin and survival of species', *Proceedings of the National Academy of Sciences*. 101(46), pp. 16085–16087. doi: 10.1073/pnas.0406909101.

Buan, N. R. and Escalante-Semerena, J. C. (2006) 'Purification and initial biochemical characterization of ATP:Cob(I)alamin adenosyltransferase (EutT) enzyme of *Salmonella enterica*', *Journal of Biological Chemistry*, 281(25), pp. 16971–16977. doi: 10.1074/jbc.M603069200.

Buan, N. R., Suh, S. J. and Escalante-Semerena, J. C. (2004) 'The *eutT* gene of *Salmonella enterica* encodes an oxygen-labile, metal-containing ATP: Corrinoid adenosyltransferase enzyme', *Journal of Bacteriology*, 186(17), pp. 5708–5714. doi: 10.1128/JB.186.17.5708-5714.2004.

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(10).

Cai, F., Sutter, M., Cameron, J. C., Stanley, D. N., Kinney, J. N. and Kerfeld, C. A. (2013) 'The structure of CcmP, a tandem bacterial microcompartment domain protein from the β-carboxysome, forms a subcompartment within a microcompartment.', *The Journal of biological chemistry*. 288(22), pp. 16055–63. doi: 10.1074/jbc.M113.456897.

Cai, F., Dou, Z., Bernstein, S., Leverenz, R., Williams, E., Heinhorst, S., Shively, J., Cannon, G. and Kerfeld, C. (2015) 'Advances in understanding carboxysome assembly in *Prochlorococcus* and *Synechococcus* implicate CsoS2 as a critical component', *Life*, 5(2), pp. 1141–1171. doi: 10.3390/life5021141.

Cai, F., Bernstein, S. L., Wilson, S. C. and Kerfeld, C. A. (2016) 'Production and characterization ofsynthetic carboxysome shells with incorporated luminal proteins.', *Plant physiology*, 170(3), pp. 1868–77. doi: 10.1104/pp.15.01822.

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, pp. 1131–1140. doi: 10.1016/j.cell.2013.10.044.

Chen, A. H., Robinson-Mosher, A., Savage, D. F., Silver, P. A. and Polka, J. K. (2013) 'The bacterial carbon-fixing organelle is formed by shell envelopment of preassembled cargo', *PLoS ONE*, 8(9). doi: 10.1371/journal.pone.0076127.

Chen, P., Andersson, D. I. and Roth, J. R. (1994) 'The control region of the *pdu/cob* regulon in *Salmonella typhimurium*', *Journal of Bacteriology*, 176(17), pp. 5474–5482. doi: 10.1128/jb.176.17.5474-5482.1994.

Cheng, S., Sinha, S., Fan, C., Liu, Y. and Bobik, T. A. (2011) 'Genetic analysis of the protein shell of the microcompartments iInvolved in coenzyme B12-dependent 1,2-propanediol degradation by *Salmonella*', *Journal of Bacteriology*, 193(6), pp. 1385–1392. doi: 10.1128/JB.01473-10.

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(10). doi: 10.1371/journal.pone.0047144.

Cheng, S. and Bobik, T. A. (2010) 'Characterization of the PduS cobalamin reductase of

Salmonella enterica and its role in the Pdu microcompartment', *Journal of Bacteriology*, 192(19), pp. 5071–5080. doi: 10.1128/JB.00575-10.

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(3), p. e33342. doi: 10.1371/journal.pone.0033342.

Chowdhury, C., Sinha, S., Chun, S., Yeates, T. O. and Bobik, T. A. (2014) 'Diverse bacterial microcompartment organelles.', *Microbiology and molecular biology reviews : MMBR*. American Society for Microbiology (ASM), 78(3), pp. 438–68. doi: 10.1128/MMBR.00009-14.

Chowdhury, C., Chun, S., Pang, A., Sawaya, M. R., Sinha, S., Yeates, T. O. and Bobik, T. A. (2015) 'Selective molecular transport through the protein shell of a bacterial microcompartment organelle', *Proceedings of the National Academy of Sciences*, 112(10), pp. 2990–2995. doi: 10.1073/pnas.1423672112.

Cohen-Bazire, G. (1963) 'The fine structure of *Rhodospirillum rubrum*', *The Journal of Cell Biology*. 16(2), pp. 401–419. doi: 10.1083/jcb.16.2.401.

Conner, C. P., Heithoff, D. M., Julio, S. M., Sinsheimer, R. L. and Mahan, M. J. (1998) 'Differential patterns of acquired virulence genes distinguish *Salmonella* strains', *Proceedings of the National Academy of Sciences*, 95(8), pp. 4641–4645. doi: 10.1073/pnas.95.8.4641.

Corpet, F. (1988) 'Multiple sequence alignment with hierarchical clustering.', *Nucleic acids research*. 16(22), pp. 10881–90. doi: 10.1093/nar/16.22.10881.

Cougot, N., Molza, A. E., Delesques, J., Giudice, E., Cavalier, A., Rolland, J. P., Ermel, G., Blanco, C., Thomas, D. and Gillet, R. (2014) 'Visualizing compaction of polysomes in bacteria', *Journal of Molecular Biology*, 426(2), pp. 377–388. doi: 10.1016/j.jmb.2013.09.035.

Craciun, S. and Balskus, E. P. (2012) 'Microbial conversion of choline to trimethylamine requires a glycyl radical enzyme', *Proceedings of the National Academy of Sciences*. 109(52), pp. 21307–21312. doi: 10.1073/pnas.1215689109.

Craciun, S., Marks, J. A. and Balskus, E. P. (2014) 'Characterization of choline trimethylamine-lyase expands the chemistry of glycyl radical enzymes', *ACS Chemical Biology*. 9(7), pp. 1408–1413. doi: 10.1021/cb500113p.

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(9), pp. 1324–1332. doi: 10.1016/j.str.2008.05.013.

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', *Journal of Biological Chemistry*, 285(48), pp. 37838–37846. doi: 10.1074/jbc.M110.160580.

Drews, G. and Niklowitz, W. (1956) 'Beiträge zur Cytologie der Blaualgen', *Archiv für Mikrobiologie*. 24(2), pp. 147–162. doi: 10.1007/BF00408629.

Edgar, R. C. (2004) 'MUSCLE: a multiple sequence alignment method with reduced time and space complexity', *BMC Bioinformatics*. 5(1), p. 113. doi: 10.1186/1471-2105-5-113.

Erbilgin, O., McDonald, K. L. and Kerfeld, C. A. (2014) 'Characterization of a planctomycetal organelle: A novel bacterial microcompartment for the aerobic degradation of plant saccharides', *Applied and Environmental Microbiology*, 80(7), pp. 2193–2205. doi:

10.1128/AEM.03887-13.

Fan, C., Cheng, S., Liu, Y., Escobar, C. M., Crowley, C. S., Jefferson, R. E., Yeates, T. O. and Bobik, T. A. (2010) 'Short N-terminal sequences package proteins into bacterial microcompartments', *Proceedings of the National Academy of Sciences*. 107(16), pp. 7509–7514. doi: 10.1073/pnas.0913199107.

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', *Proceedings of the National Academy of Sciences*. 109(37), pp. 14995–15000. doi: 10.1073/pnas.1207516109.

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', *Journal of Bacteriology*, 193(20), pp. 5623–5628.

Fang, Y., Huang, F., Faulkner, M., Jiang, Q., Dykes, G. F., Yang, M. and Liu, L.-N. (2018) 'Engineering and Modulating Functional Cyanobacterial CO_2-Fixing Organelles', *Frontiers in Plant Science*. 9, p. 739. doi: 10.3389/fpls.2018.00739.

Faulkner, M., Zhao, L.-S., Barrett, S. and Liu, L.-N. (2019) 'Self-Assembly Stability and Variability of Bacterial Microcompartment Shell Proteins in Response to the Environmental Change', *Nanoscale Research Letters*. 14(1), p. 54. doi: 10.1186/s11671-019-2884-3.

Ferlez, B., Sutter, M. and Kerfeld, C.A. (2019) 'A designed bacterial microcompartment shell with tunable composition and precision cargo loading', *Metabolic Engineering*, 54, pp. 286–291. doi: 10.1016/j.ymben.2019.04.011.

Ferlez, B., Sutter, M. and Kerfeld, C. A (2019) 'Glycyl Radical Enzyme-Associated Microcompartments: Redox-Replete Bacterial Organelles.', *mBio*. 10(1), pp. e02327-18. doi: 10.1128/mBio.02327-18.

Gantt, E. and Conti, S. F. (1969) '*Ultrastructure of Blue-Green Algae'*, *Journal of bacteriology*. 97 (3), pp. 1486-1493.

Garsin, D. A. (2010) 'Ethanolamine utilization in bacterial pathogens: roles and regulation', *Nature Reviews Microbiology*, 8(4), pp. 290–295. doi: 10.1038/nrmicro2334.

Graf, L., Wu, K. and Wilson, J. W. (2018) 'Transfer and analysis of *Salmonella pdu* genes in a range of Gram-negative bacteria demonstrate exogenous microcompartment expression across a variety of species', *Microbial Biotechnology*. 11(1), pp. 199–210. doi: 10.1111/1751-7915.12863.

Greber, B. J., Sutter, M. and Kerfeld, C. A. (2019) 'The Plasticity of Molecular Interactions Governs Bacterial Microcompartment Shell Assembly', *Structure*, 27 (5), pp. 749-763. doi: 10.1016/j.str.2019.01.017.

Hagen, A., Sutter, M., Sloan, N. and Kerfeld, C. A. (2018) 'Programmed loading and rapid purification of engineered bacterial microcompartment shells', *Nature Communications*. 9(1), p. 2881. doi: 10.1038/s41467-018-05162-z.

Hagen, A. R., Plegaria, J. S., Sloan, N., Ferlez, B., Aussignargues, C., Burton, R. and Kerfeld, C. A. (2018) 'In Vitro Assembly of Diverse Bacterial Microcompartment Shell Architectures', *Nano Letters*. 18(11), pp. 7030–7037. doi: 10.1021/acs.nanolett.8b02991.

Hanson, G. and Coller, J. (2018) 'Codon optimality, bias and usage in translation and mRNA decay', *Nature Reviews Molecular Cell Biology*. 19(1), pp. 20–30. doi: 10.1038/nrm.2017.91.

Harvey, P. C., Watson, M., Hulme, S., Jones, M. A., Lovell, M., Berchieri, A., Young, J.,

Bumstead, N. and Barrow, P. (2011) 'S*almonella enterica* Serovar Typhimurium Colonizing the Lumen of the Chicken Intestine Grows Slowly and Upregulates a Unique Set of Virulence and Metabolism Genes', *Infection and Immunity*. 79(10), pp. 4105–21. doi: 10.1128/IAI.01390-10.

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', *Journal of Bacteriology*, 185(17), pp. 5086–5095. doi: 10.1128/JB.185.17.5086-5095.2003.

Hayward, H. R. and Stadtman, T. C. (1959) 'Anaerobic degradation of choline. I. Fermentation of choline by an anaerobic, cytochrome-producing bacterium, *Vibrio cholinicus* n. sp.', *Journal of bacteriology*. 78(4), pp. 557–61.

Heithoff, D. M., Conner, C. P., Hentschel, U., Govantes, F., Hanna, P. C. and Mahan, M. J. (1999) 'Coordinate intracellular expression of *Salmonella* genes induced during infection.', *Journal of bacteriology*, 181(3), pp. 799–807.

Held, M., Kolb, A., Perdue, S., Hsu, S.-Y. Y., Bloch, S. E., Quin, M. B. and Schmidt-Dannert, C. (2016) 'Engineering formation of multiple recombinant Eut protein nanocompartments in *E. coli*', *Scientific Reports*. 6(2), pp. 199–207. doi: 10.1038/srep24359.

Heldt, D., Frank, S., Seyedarabi, A., Ladikis, D., Parsons, J. B., Warren, M. J. and Pickersgill, R. W. (2009) 'Structure of a trimeric bacterial microcompartment shell protein, EtuB, associated with ethanol utilization in *Clostridium kluyveri*', *Biochem. J*, 423, pp. 199–207. doi: 10.1042/BJ20090780.

Herring, T. I., Harris, T. N., Chowdhury, C., Mohanty, S. and Bobik, T. A. (2018) 'A bacterial microcompartment is used for choline fermentation by *Escherichia coli* 536', *Journal of Bacteriology*. doi: 10.1128/JB.00764-17.

Herskovits, A. A., Shimoni, E., Minsky, A. and Bibi, E. (2002) 'Accumulation of endoplasmic membranes and novel membrane-bound ribosome-signal recognition particle receptor complexes in *Escherichia coli*.', *The Journal of cell biology*. 159(3), pp. 403–10. doi: 10.1083/jcb.200204144.

Horswill, A. R. and Escalante-Semerena, J. C. (1999) '*Salmonella typhimurium* LT2 catabolizes propionate via the 2-methylcitric acid cycle.', *Journal of bacteriology*. 181(18), pp. 5615–23.

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 Synthetic Biology*, 6(11), pp. 2145–2156. doi: 10.1021/acssynbio.7b00167.

Huseby, D. L. and Roth, J. R. (2013) 'Evidence that a metabolic microcompartment contains and recycles private cofactor pools.', *Journal of bacteriology*. 195(12), pp. 2864–79. doi: 10.1128/JB.02179-12.

Iancu, C. V., Morris, D. M., Dou, Z., Heinhorst, S., Cannon, G. C. and Jensen, G. J. (2010) 'Organization, Structure, and Assembly of α-Carboxysomes Determined by Electron Cryotomography of Intact Cells', *Journal of Molecular Biology*, 396(1), pp. 105–117. doi: 10.1016/j.jmb.2009.11.019.

Jakobson, C. M., Tullman-Ercek, D., Slininger, M. F. and Mangan, N. M. (2017) 'A systems-level model reveals that 1,2-Propanediol utilization microcompartments enhance pathway flux through intermediate sequestration', *PLOS Computational Biology*. 13(5), p. e1005525. doi: 10.1371/journal.pcbi.1005525.

Jakobson, C. M., Slininger Lee, M. F. and Tullman-Ercek, D. (2017) '*De novo* design of signal sequences to localize cargo to the 1,2-propanediol utilization microcompartment', *Protein Science*. 26(5), pp. 1086–1092. doi: 10.1002/pro.3144.

Jakobson, C. M. and Tullman-Ercek, D. (2016) 'Dumpster Diving in the Gut: Bacterial Microcompartments as Part of a Host-Associated Lifestyle', *PLoS Pathog*, 12(5), p. e1005558. doi: 10.1371/journal.ppat.1005558.

Jameson, E., Fu, T., Brown, I. R., Paszkiewicz, K., Purdy, K. J., Frank, S. and Chen, Y. (2016) 'Anaerobic choline metabolism in microcompartments promotes growth and swarming of *Proteus mirabilis*', *Environmental Microbiology*. 18(9), pp. 2886–2898. doi: 10.1111/1462-2920.13059.

Johnson, C. L. V., 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', *Journal of Bacteriology*, 183(5), pp. 1577–1584. doi: 10.1128/JB.183.5.1577-1584.2001.

Jorda, J., Lopez, D., Wheatley, N. M. and Yeates, T. O. (2013) 'Using comparative genomics to uncover new kinds of protein-based metabolic organelles in bacteria', *Protein Science*. 22(2), pp. 179–195. doi: 10.1002/pro.2196.

Joseph, B., Przybilla, K., Stuhler, C., Schauer, K., Slaghuis, J., Fuchs, T. M. and Goebel, W. (2006) 'Identification of L*isteria monocytogenes* Genes Contributing to Intracellular Replication by Expression Profiling and Mutant Screening', *Journal of Bacteriology*, 188(2), pp. 556–568. doi: 10.1128/JB.188.2.556-568.2006.

Keeling, T. J., Samborska, B., Demers, R. W. and Kimber, M. S. (2014) 'Interactions and structural variability of β-carboxysomal shell protein CcmL', *Photosynthesis Research*, 121(2–3), pp. 125–133. doi: 10.1007/s11120-014-9973-z.

Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. and Sternberg, M. J. E. (2015) 'The Phyre2 web portal for protein modeling, prediction and analysis', *Nature Protocols*. 10(6), pp. 845–858. doi: 10.1038/nprot.2015.053.

Kendall, M. M., Gruber, C. C., Parker, C. T. and Sperandio, V. (2012) 'Ethanolamine Controls Expression of Genes Encoding Components Involved in Interkingdom Signaling and Virulence in Enterohemorrhagic *Escherichia coli* O157:H7', *mBio*. 3(3). doi: 10.1128/mBio.00050-12.

Kerfeld, C. A. (2005) 'Protein Structures Forming the Shell of Primitive Bacterial Organelles', *Science*, 309(5736), pp. 936–938. doi: 10.1126/science.1113397.

Kerfeld, C. A., Heinhorst, S. and Cannon, G. C. (2010) 'Bacterial Microcompartments', *Annual Review of Microbiology*, 64(1), pp. 391–408. doi: 10.1146/annurev.micro.112408.134211.

Kerfeld, C. A. and Melnicki, M. R. (2016) 'Assembly, function and evolution of cyanobacterial carboxysomes', *Current Opinion in Plant Biology*, 31, pp. 66–75. doi: 10.1016/j.pbi.2016.03.009.

Kim, E. Y., Slininger, M. F. and Tullman-Ercek, D. (2014) 'The effects of time, temperature, and pH on the stability of Pdu bacterial microcompartments', *Protein Science*. 23(10), pp. 1434–1441. doi: 10.1002/pro.2527.

Kinney, J. N., Salmeen, A., Cai, F. and Kerfeld, C. A. (2012) 'Elucidating Essential Role of Conserved Carboxysomal Protein CcmN Reveals Common Feature of Bacterial Microcompartment Assembly', *Journal of Biological Chemistry*, 287(21), pp. 17729–17736.

doi: 10.1074/jbc.M112.355305.

Klein, M. G., Zwart, P., Bagby, S. C., Cai, F., Chisholm, S. W., Heinhorst, S., Cannon, G. C. and Kerfeld, C. A. (2009) 'Identification and Structural Analysis of a Novel Carboxysome Shell Protein with Implications for Metabolite Transport', *Journal of Molecular Biology*, 392, pp. 319–333. doi: 10.1016/j.jmb.2009.03.056.

Kofoid, E., Rappleye, C., Stojiljkovic, I. and Roth, J. (1999) 'The 17-gene ethanolamine (*eut*) operon of *Salmonella typhimurium* encodes five homologues of carboxysome shell proteins.', *Journal of bacteriology*, 181(17), pp. 5317–29.

Kuehl, J. V., Price, M. N., Ray, J., Wetmore, K. M., Esquivel, Z., Kazakov, A. E., Nguyen, M., Kuehn, R., Davis, R. W., Hazen, T. C., Arkin, A. P. and Deutschbauer, A. (2014) 'Functional genomics with a comprehensive library of transposon mutants for the sulfate-reducing bacterium *Desulfovibrio alaskensis* G20', *mBio*, 5(3). doi: 10.1128/mBio.01041-14.

LaMattina, J. W., Keul, N. D., Reitzer, P., Kapoor, S., Galzerani, F., Koch, D. J., Gouvea, I. E. and Lanzilotta, W. N. (2016) '1,2-Propanediol dehydration in *Roseburia inulinivorans*: Structural basis for substrate and enantiomer selectivity', *Journal of Biological Chemistry*, 291(30), pp. 15515–15526. doi: 10.1074/jbc.M116.721142.

Larsson, A. M., Hasse, D., Valegård, K. and Andersson, I. (2017) 'Crystal structures of β-carboxysome shell protein CcmP: ligand binding correlates with the closed or open central pore', *Journal of Experimental Botany*. 68(14), pp. 3857–3867. doi: 10.1093/jxb/erx070.

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', *Journal of Molecular Biology*, 426(11), pp. 2217–2228. doi: 10.1016/j.jmb.2014.02.025.

Lawrence, A. D., Frank, S., Newnham, S., Lee, M. J., Brown, I. R., Xue, W.-F., Rowe, M. L., Mulvihill, D. P., Prentice, M. B., Howard, M. J. and Warren, M. J. (2014) 'Solution Structure of a Bacterial Microcompartment Targeting Peptide and Its Application in the Construction of an Ethanol Bioreactor', *ACS Synthetic Biology*, 3(7), pp. 454–465. doi: 10.1021/sb4001118.

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 B_{12}-dependent 1,2-propanediol degradation by *Salmonella enterica* serovar Typhimurium LT2', *Archives of Microbiology*, 180(5), pp. 353–361. doi: 10.1007/s00203-003-0601-0.

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*.', *Metabolic engineering*, 36, pp. 48–56. doi: 10.1016/j.ymben.2016.02.007.

Lee, M. J., Mantell, J., Brown, I. R., Fletcher, J. M., Verkade, P., Pickersgill, R. W., Woolfson, D. N., Frank, S. and Warren, M. J. (2018) 'De novo targeting to the cytoplasmic and luminal side of bacterial microcompartments', *Nature Communications*. 9(1), p. 3413. doi: 10.1038/s41467-018-05922-x.

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. and Warren, M. J. (2018) 'Engineered synthetic scaffolds for organizing proteins within the bacterial cytoplasm', *Nature Chemical Biology*. 14(2), pp. 142–147. doi: 10.1038/nchembio.2535.

Lehman, B. P., Chowdhury, C. and Bobik, T. A. (2017) 'The N Terminus of the PduB Protein Binds the Protein Shell of the Pdu Microcompartment to Its Enzymatic Core', *Journal*

of Bacteriology. 199(8), pp. e00785-16. doi: 10.1128/JB.00785-16.

Lin, M. T., Occhialini, A., Andralojc, P. J., Devonshire, J., Hines, K. M., Parry, M. A. J. and Hanson, M. R. (2014) 'β-Carboxysomal proteins assemble into highly organized structures in *Nicotiana* chloroplasts', *The Plant Journal*. 79(1), pp. 1–12. doi: 10.1111/tpj.12536.

Liu, Y., Leal, N. A., Sampson, E. M., Johnson, C. L. V., Havemann, G. D. and Bobik, T. A. (2007) 'PduL Is an Evolutionarily Distinct Phosphotransacylase Involved in B_{12}-Dependent 1,2-Propanediol Degradation by *Salmonella enterica* Serovar Typhimurium LT2', *Journal of Bacteriology*. 189(5), p. 1589. doi: 10.1128/JB.01151-06.

Liu, Y., Jorda, J., Yeates, T. and Bobik, T. (2015) 'The PduL phosphotransacylase is used to recycle coenzyme A within the Pdu microcompartment.', *Journal of Bacteriology*, 197(14), pp. 2392–2399. doi: 10.1128/JB.00056-15.

Liu, Y., He, X., Lim, W., Mueller, J., Lawrie, J., Kramer, L., Guo, J. and Niu, W. (2018) 'Deciphering molecular details in the assembly of alpha-type carboxysome', *Scientific Reports*, 8(1), p. 15062. doi: 10.1038/s41598-018-33074-x.

Long, B. M., Hee, W. Y., Sharwood, R. E., Rae, B. D., Kaines, S., Lim, Y.-L., Nguyen, N. D., Massey, B., Bala, S., von Caemmerer, S., Badger, M. R. and Price, G. D. (2018) 'Carboxysome encapsulation of the CO_2-fixing enzyme Rubisco in tobacco chloroplasts', *Nature Communications*. 9(1), p. 3570. doi: 10.1038/s41467-018-06044-0.

Ludwig, M., Sültemeyer, D. and Price, G. D. (2000) 'Isolation of *ccmKLMN* genes from the marine cyanobacterium, *Synechococcus* sp. Pcc7002 (Cyanophyceae), and evidence that CcmM is essential for carboxysome assembly', *Journal of Phycology*, 36(6), pp. 1109–1119. doi: 10.1046/j.1529-8817.2000.00028.x.

Maadani, A., Fox, K. A., Mylonakis, E. and Garsin, D. A. (2007) '*Enterococcus faecalis* Mutations Affecting Virulence in the *Caenorhabditis elegans* Model Host', *Infection and Immunity*, 75(5), pp. 2634–2637. doi: 10.1128/IAI.01372-06.

Madeira, F., Park, Y., Lee, J., Buso, N., Gur, T., Madhusoodanan, N., Basutkar, P., Tivey, A., Potter, S., Finn, R. and Lopez, R. (2019) 'The EMBL-EBI search and sequence analysis tools APIs in 2019.', *Nucleic acids research*. doi: 10.1093/NAR/GKZ268.

Martínez-del Campo, A., Bodea, S., Hamer, H. A., Marks, J. A., Haiser, H. J., Turnbaugh, P. J. and Balskus, E. P. (2015) 'Characterization and detection of a widely distributed gene cluster that predicts anaerobic choline utilization by human gut bacteria.', *mBio*. 6(2), pp. e00042-15. doi: 10.1128/mBio.00042-15.

Mastroleo, F., Van Houdt, R., Leroy, B., Benotmane, M. A., Janssen, A., Mergeay, M., Vanhavere, F., Hendrickx, L., Wattiez, R. and Leys, N. (2009) 'Experimental design and environmental parameters affect *Rhodospirillum rubrum* S1H response to space flight', *The ISME Journal*, 3(12), pp. 1402–1419. doi: 10.1038/ismej.2009.74.

McGoldrick, H. M., Roessner, C. A., Raux, E., Lawrence, A. D., McLean, K. J., Munro, A. W., Santabarbara, S., Rigby, S. E. J., Heathcote, P., Scott, A. I. and Warren, M. J. (2005) 'Identification and Characterization of a Novel Vitamin B_{12} (Cobalamin) Biosynthetic Enzyme (CobZ) from *Rhodobacter capsulatus*, Containing Flavin, Heme, and Fe-S Cofactors', *Journal of Biological Chemistry*. 280(2), pp. 1086–1094. doi: 10.1074/jbc.M411884200.

McGurn, L. D., Moazami-Goudarzi, M., White, S. A., Suwal, T., Brar, B., Tang, J. Q., Espie, G. S. and Kimber, M. S. (2016) 'The structure, kinetics and interactions of the -carboxysomal -carbonic anhydrase, CcaA', *Biochemical Journal*, 473(24), pp. 4559–4572.

doi: 10.1042/BCJ20160773.

Moore, T. C. and Escalante-Semerena, J. C. (2016) 'The EutQ and EutP proteins are novel acetate kinases involved in ethanolamine catabolism: physiological implications for the function of the ethanolamine metabolosome in *Salmonella enterica*', *Molecular Microbiology*, 99(3), pp. 497–511. doi: 10.1111/mmi.13243.

Mori, K., Bando, R., Hieda, N. and Toraya, T. (2004) 'Identification of a Reactivating Factor for Adenosylcobalamin-Dependent Ethanolamine Ammonia Lyase', *Journal of Bacteriology*, 186(20), pp. 6845–6854. doi: 10.1128/JB.186.20.6845-6854.2004.

van Niftrik, L. and Jetten, M. S. M. (2012) 'Anaerobic ammonium-oxidizing bacteria: unique microorganisms with exceptional properties.', *Microbiology and molecular biology reviews : MMBR*. 76(3), pp. 585–96. doi: 10.1128/MMBR.05025-11.

Noël, C. R., Cai, F. and Kerfeld, C. A. (2016) 'Purification and Characterization of Protein Nanotubes Assembled from a Single Bacterial Microcompartment Shell Subunit', *Advanced Materials Interfaces*, 3(1), p. 1500295. doi: 10.1002/admi.201500295.

Norman, R. O. J., Millat, T., Winzer, K., Minton, N. P. and Hodgman, C. (2018) 'Progress towards platform chemical production using *Clostridium autoethanogenum*', *Biochemical Society Transactions*, 46(3), pp. 523–535. doi: 10.1042/BST20170259.

Obradors, N., Badia, J., Baldoma, L. and Aguilar, J. (1988) 'Anaerobic metaoblism of the L-rhamnose fermentation product 1,2-propanediol in *Salmonella typhimurium*', *Journal of Bacteriology*, 170(5), pp. 2159–2162. doi: 10.1128/jb.170.5.2159-2162.1988.

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 Crystallographica Section D Biological Crystallography*. 68(12), pp. 1642–1652. doi: 10.1107/S0907444912039315.

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.', *The Journal of biological chemistry*, 289(32), pp. 22377–22384. doi: 10.1074/jbc.M114.569285.

Pang, A., Warren, M. J. and Pickersgill, R. W. (2011) 'Structure of PduT, a trimeric bacterial microcompartment protein with a 4Fe–4S cluster-binding site', *Acta Crystallographica Section D Biological Crystallography*. 67(2), pp. 91–96. doi: 10.1107/S0907444910050201.

Pardee, A. B., Schachman, H. K. and Stanier, R. Y. (1952) 'Chromatophores of *Rhodospirillum rubrum*', *Nature*, pp. 282–283. doi: 10.1038/169282a0.

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. J., Warren, M. J. and Prentice, M. B. (2008) 'Biochemical and structural insights into bacterial organelle form and biogenesis', *Journal of Biological Chemistry*, 283(21), pp. 14366–14375. doi: 10.1074/jbc.M709214200.

Parsons, J. B., Lawrence, A. D., McLean, K. J., Munro, A. W., Rigby, S. E. J. and Warren, M. J. (2010) 'Characterisation of PduS, the pdu Metabolosome Corrin Reductase, and Evidence of Substructural Organisation within the Bacterial Microcompartment', *PLoS ONE*. 5(11), p. e14009. doi: 10.1371/journal.pone.0014009.

Parsons, J. B., Frank, S., Bhella, D., Liang, M., Prentice, M. B., Mulvihill, D. P. and Warren, M. J. (2010) 'Synthesis of Empty Bacterial Microcompartments, Directed Organelle

Protein Incorporation, and Evidence of Filament-Associated Organelle Movement', *Molecular Cell*, 38(2), pp. 305–315. doi: 10.1016/j.molcel.2010.04.008.

Penrod, J. T. and Roth, J. R. (2006) 'Conserving a volatile metabolite: a role for carboxysome-like organelles in *Salmonella enterica*.', *Journal of bacteriology*. 188(8), pp. 2865–74. doi: 10.1128/JB.188.8.2865-2874.2006.

Petit, E., LaTouf, W. G., Coppi, M. V., Warnick, T. A., Currie, D., Romashko, I., Deshpande, S., Haas, K., Alvelo-Maurosa, J. G., Wardman, C., Schnell, D. J., Leschine, S. B. and Blanchard, J. L. (2013) 'Involvement of a Bacterial Microcompartment in the Metabolism of Fucose and Rhamnose by *Clostridium phytofermentans*', *PLoS ONE*. 8(1), p. e54337. doi: 10.1371/journal.pone.0054337.

Pitts, A. C., Tuck, L. R., Faulds-Pain, A., Lewis, R. J. and Marles-Wright, J. (2012) 'Structural Insight into the *Clostridium difficile* Ethanolamine Utilisation Microcompartment', *PLoS ONE*. 7(10), p. e48360. doi: 10.1371/journal.pone.0048360.

Pravda, L., Sehnal, D., Toušek, D., Navrátilová, V., Bazgier, V., Berka, K., Svobodová Vareková, R., Koca, J., & Otyepka, M. (2018) 'MOLEonline: a web-based tool for analyzing channels, tunnels and pores (2018 update)' *Nucleic acids research*, 46(W1), W368–W373. doi.org/10.1093/nar/gky309

Price, G. D. and Badger, M. R. (1989a) 'Expression of Human Carbonic Anhydrase in the *Cyanobacterium Synechococcus* PCC7942 Creates a High CO_2-Requiring Phenotype', *Plant Physiology*. 91(2), pp. 505–513. doi: 10.1104/PP.91.2.505.

Price, G. D. and Badger, M. R. (1989b) 'Isolation and Characterization of High CO_2-Requiring-Mutants of the Cyanobacterium *Synechococcus* PCC7942', *Plant Physiology*. 91(2), pp. 514–525. doi: 10.1104/PP.91.2.514.

Price, G. D. and Badger, M. R. (1991) 'Evidence for the role of carboxysomes in the cyanobacterial CO_2-concentrating mechanism', *Canadian Journal of Botany*. 69(5), pp. 963–973. doi: 10.1139/b91-124.

Price, G. D., Coleman, J. R. and Badger, M. R. (1992) 'Association of Carbonic Anhydrase Activity with Carboxysomes Isolated from the Cyanobacterium *Synechococcus* PCC7942', *Plant Physiology*. 100(2), pp. 784–793. doi: 10.1104/PP.100.2.784.

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 CO_2 fixation in cyanobacteria and some proteobacteria.', *Microbiology and molecular biology reviews : MMBR*. 77(3), pp. 357–79. doi: 10.1128/MMBR.00061-12.

Ravcheev, D. A., Moussu, L., Smajic, S. and Thiele, I. (2019) 'Comparative Genomic Analysis Reveals Novel Microcompartment-Associated Metabolic Pathways in the Human Gut Microbiome', *Frontiers in Genetics*, 10, p. 636. doi: 10.3389/fgene.2019.00636.

Reslewic, S., Zhou, S., Place, M., Zhang, Y., Briska, A., Goldstein, S., Churas, C., Runnheim, R., Forrest, D., Lim, A., Lapidus, A., Han, C. S., Roberts, G. P. and Schwartz, D. C. (2005) 'Whole-Genome Shotgun Optical Mapping of *Rhodospirillum rubrum', Applied and Environmental Microbiology, 71(9), pp. 5511–5522. doi: 10.112*8/AEM.71.9.5511-5522.2005.

Revelles, O., Tarazona, N., García, J. L. and Prieto, M. A. (2016) 'Carbon roadmap from syngas to polyhydroxyalkanoates in *Rhodospirillum rubrum*', *Environmental Microbiology*, 18(2), pp. 708–720. doi: 10.1111/1462-2920.13087.

Rondon, M. R., Horswill, A. R. and Escalante-Semerena, J. C. (1995) 'DNA polymerase I

function is required for the utilization of ethanolamine, 1,2-propanediol, and propionate by *Salmonella Typhimurium* LT2', *Journal of Bacteriology*, 177(24), pp. 7119–7124. doi: 10.1128/jb.177.24.7119-7124.1995.

Rondon, M. R., Kazmierczak, R. and Escalante-Semerena, J. C. (1995) 'Glutathione is required for maximal transcription of the cobalamin biosynthetic and 1,2-propanediol utilization (*cob/pdu*) regulon and for the catabolism of ethanolamine, 1,2-propanediol, and propionate in *Salmonella Typhimurium* LT2', *Journal of Bacteriology*, 177(19), pp. 5434–5439. doi: 10.1128/jb.177.19.5434-5439.1995.

Roof, D. M. and Roth, J. R. (1989) 'Functions required for vitamin B12-dependent ethanolamine utilization in *Salmonella Typhimurium*', *Journal of Bacteriology*, 171(6), pp. 3316–3323. doi: 10.1128/jb.171.6.3316-3323.1989.

Sagermann, M., Ohtaki, A. and Nikolakakis, K. (2009) 'Crystal structure of the EutL shell protein of the ethanolamine ammonia lyase microcompartment', *Proceedings of the National Academy of Sciences*, 106(22), pp. 8883–8887. doi: 10.1073/pnas.0902324106.

Samborska, B. and Kimber, M. S. (2012) 'A Dodecameric CcmK2 Structure Suggests β-Carboxysomal Shell Facets Have a Double-Layered Organization', *Structure*. 20(8), pp. 1353–1362. doi: 10.1016/j.str.2012.05.013.

Sampson, E. M. and Bobik, T. A. (2008) 'Microcompartments for B_{12}-dependent 1,2-propanediol degradation provide protection from DNA and cellular damage by a reactive metabolic intermediate', *Journal of Bacteriology*, 190(8), pp. 2966–2971. doi: 10.1128/JB.01925-07.

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. and Palmer, T. (2013) 'A synthetic system for expression of components of a bacterial microcompartment', *Microbiology*, 159(Pt_11), pp. 2427–2436. doi: 10.1099/mic.0.069922-0.

Schavemaker, P. E., Śmigiel, W. M. and Poolman, B. (2017) 'Ribosome surface properties may impose limits on the nature of the cytoplasmic proteome', *eLife*, 6. doi: 10.7554/eLife.30084.

Schneider, C. A., Rasband, W. S. and Eliceiri, K. W. (2012) 'NIH Image to ImageJ: 25 years of image analysis.', *Nature methods*. 9(7), pp. 671–5. doi: 10.1038/NMETH.2089.

Scott, K. P., Martin, J. C., Campbell, G., Mayer, C.-D. and Flint, H. J. (2006) 'Whole-Genome Transcription Profiling Reveals Genes Up-Regulated by Growth on Fucose in the Human Gut Bacterium *Roseburia inulinivorans*', *Journal of Bacteriology*, 188(12), pp. 4340–4349. doi: 10.1128/JB.00137-06.

Selmer, T., Pierik, A. J. and Heider, J. (2005) 'New glycyl radical enzymes catalysing key metabolic steps in anaerobic bacteria', *Biological Chemistry*, 386(10), pp. 981–988. doi: 10.1515/BC.2005.114.

Sheppard, D. E., Penrod, J. T., Bobik, T., Kofoid, E. and Roth, J. R. (2004) 'Evidence that a B_{12}-adenosyl transferase is encoded within the ethanolamine operon of *Salmonella enterica*', *Journal of Bacteriology*, 186(22), pp. 7635–7644. doi: 10.1128/JB.186.22.7635-7644.2004.

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(4112), pp. 584–586. doi: 10.1126/science.182.4112.584.

Shively, J. M., Bradburne, C. E., Aldrich, H. C., Bobik, T. A., Mehlman, J. L., Jin, S. and

Baker, S. H. (1998) 'Sequence homologs of the carboxysomal polypeptide CsoS1 of the thiobacilli are present in cyanobacteria and enteric bacteria that form carboxysomes - polyhedral bodies', *Canadian Journal of Botany*. 76(6), pp. 906–916. doi: 10.1139/b98-088.

Shively, J. M., Decker, G. L. and Greenawalt, J. W. (1970) 'Comparative ultrastructure of the thiobacilli.', *Journal of Bacteriology*. 101(2), pp. 618-627

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*.', *Journal of bacteriology*. 194(8), pp. 1912–8. doi: 10.1128/JB.06529-11.

Stojiljkovic, I., Baumler, A. J. and Heffron, F. (1995a) 'Ethanolamine utilization in *Salmonella Typhimurium*: Nucleotide sequence, protein expression, and mutational analysis of the *cchA cchB eutE eutJ eutG eutH* gene cluster', *Journal of Bacteriology*, 177(5), pp. 1357–1366. doi: 10.1128/jb.177.5.1357-1366.1995.

Sutter, M., Wilson, S. C., Deutsch, S. and Kerfeld, C. A. (2013) 'Two new high-resolution crystal structures of carboxysome pentamer proteins reveal high structural conservation of CcmL orthologs among distantly related cyanobacterial species', *Photosynthesis Research*. 118(1–2), pp. 9–16. doi: 10.1007/s11120-013-9909-z.

Sutter, M., Faulkner, M., Aussignargues, C., Paasch, B. C., Barrett, S., Kerfeld, C. A. and Liu, L.-N. (2016) 'Visualization of Bacterial Microcompartment Facet Assembly Using High-Speed Atomic Force Microscopy', *Nano Letters*, 16(3), pp. 1590–1595. doi: 10.1021/acs.nanolett.5b04259.

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(6344), pp. 1293–1297. doi: 10.1126/science.aan3289.

Sutter, M., Laughlin, T. G., Sloan, N. B., Serwas, D., Davies, K. M. and Kerfeld, C. A. (2019) 'Structure of a synthetic beta-carboxysome shell.', *Plant physiology*. 181(3) pp.1050-1058. doi: 10.1104/pp.19.00885.

Takenoya, M., Nikolakakis, K. and Sagermann, M. (2010) 'Crystallographic insights into the pore structures and mechanisms of the EutL and EutM shell proteins of the ethanolamine-utilizing microcompartment of *Escherichia coli*.', *Journal of bacteriology*. 192(22), pp. 6056–63. doi: 10.1128/JB.00652-10.

Tanaka, S., Kerfeld, C. A., Sawaya, M. R., Cai, F., Heinhorst, S., Cannon, G. C. and Yeates, T. O. (2008) 'Atomic-level models of the bacterial carboxysome shell', *Science*, 319(5866), pp. 1083–1086.

Tanaka, S., Sawaya, M. R., Phillips, M. and Yeates, T. O. (2009) 'Insights from multiple structures of the shell proteins from the beta-carboxysome.', *Protein science.* 18(1), pp. 108–20. doi: 10.1002/pro.14.

Tanaka, S., Sawaya, M. R. and Yeates, T. O. (2010) 'Structure and Mechanisms of a Protein-Based Organelle in *Escherichia coli'*, *Science*. 327(5961), pp. 81–84. doi: 10.1126/science.1179513.

Thompson, M. C., Wheatley, N. M., Jorda, J., Sawaya, M. R., Gidaniyan, S. D., Ahmed, H., Yang, Z., McCarty, K. N., Whitelegge, J. P. and Yeates, T. O. (2014) 'Identification of a Unique Fe-S Cluster Binding Site in a Glycyl-Radical Type Microcompartment Shell Protein', *Journal of Molecular Biology*, 426(19), pp. 3287–3304. doi: 10.1016/j.jmb.2014.07.018.

Tocheva, E. I., Matson, E. G., Cheng, S. N., Chen, W. G., Leadbetter, J. R. and Jensen, G. J. (2014) 'Structure and expression of propanediol utilization microcompartments in *Acetonema longum*.', *Journal of bacteriology*. 196(9), pp. 1651–8. doi: 10.1128/JB.00049-14.

Tsai, Y., Sawaya, M. R., Cannon, G. C., Cai, F., Williams, E. B., Heinhorst, S., Kerfeld, C. A. and Yeates, T. O. (2007) 'Structural analysis of CsoS1A and the protein shell of the *Halothiobacillus neapolitanus* carboxysome.', *PLoS biology*. 5(6), p. e144. doi: 10.1371/journal.pbio.0050144.

Uddin, I., Frank, S., Warren, M. J. and Pickersgill, R. W. (2018) 'A Generic Self-Assembly Process in Microcompartments and Synthetic Protein Nanotubes', *Small*, 14(19). doi: 10.1002/smll.201704020.

Ulmer, H. W., Gross, R. A., Posada, M., Weisbach, P., Fuller, R. C. and Lenz, R. W. (1994) 'Bacterial production of poly(.beta.-hydroxyalkanoates) containing unsaturated repeating units by *Rhodospirillum rubrum*', *Macromolecules*, 27(7), pp. 1675–1679. doi: 10.1021/ma00085a002.

Wagner, H. J., Capitain, C. C., Richter, K., Nessling, M. and Mampel, J. (2017) 'Engineering bacterial microcompartments with heterologous enzyme cargos', *Engineering in Life Sciences*, 17(1), pp. 36–46. doi: 10.1002/elsc.201600107.

Wheatley, N. M., Gidaniyan, S. D., Liu, Y., Cascio, D. and Yeates, T. O. (2013) 'Bacterial microcompartment shells of diverse functional types possess pentameric vertex proteins', *Protein Science*. 22(5), pp. 660–665. doi: 10.1002/pro.2246.

Yeates, T. O., Crowley, C. S. and Tanaka, S. (2010) 'Bacterial Microcompartment Organelles: Protein Shell Structure and Evolution', *Annual Review of Biophysics*, 39(1), pp. 185–205. doi: 10.1146/annurev.biophys.093008.131418.

Young, E. J., Burton, R., Mahalik, J. P., Sumpter, B. G., Fuentes-Cabrera, M., Kerfeld, C. A. and Ducat, D. C. (2017) 'Engineering the Bacterial Microcompartment Domain for Molecular Scaffolding Applications', *Frontiers in Microbiology*, 8(JUL). doi: 10.3389/fmicb.2017.01441.

Zarzycki, J., Axen, S. D., Kinney, J. N. and Kerfeld, C. A. (2013) 'Cyanobacterial-based approaches to improving photosynthesis in plants', *Journal of Experimental Botany*. 64(3), pp. 787–798. doi: 10.1093/jxb/ers294.

Zarzycki, J., Sutter, M., Cortina, N. S., Erb, T. J. and Kerfeld, C. A. (2017) 'In Vitro Characterization and Concerted Function of Three Core Enzymes of a Glycyl Radical Enzyme - Associated Bacterial Microcompartment', *Scientific Reports*, 7. doi: 10.1038/srep42757.

Zarzycki, J., Erbilgin, O. and Kerfeld, C. A. (2015) 'Bioinformatic characterization of glycyl radical enzyme-associated bacterial microcompartments.', *Applied and environmental microbiology*. 81(24), pp. 8315–29. doi: 10.1128/AEM.02587-15.

Zhang, G., Quin, M. B. and Schmidt-Dannert, C. (2018) 'Self-Assembling Protein Scaffold System for Easy *in Vitro* Coimmobilization of Biocatalytic Cascade Enzymes', *ACS Catalysis*, 8(6), pp. 5611–5620. doi: 10.1021/acscatal.8b00986.