Structural analysis of the protein shell of the propanediol utilisation metabolosome

Pang, Allan

Structural analysis of the protein shell of the propanediol utilisation metabolosome

Allan Pang

Thesis submitted for the degree of Doctor of Philosophy

School of Biological and Chemical Sciences

Queen Mary, University of London

September 2012

Declaration

I hereby declare that data written in this Thesis is a result of my own original research. Information from other people, be it as a result of collaboration or not, both published and unpublished have been properly cited and acknowledged.

This Thesis, or any equivalent or similar form, has not been submitted or presented to any other university in support for an application for any degree, diploma or any other qualification, other than the university, for which I am now a candidate.

Allan Pang

September 2012

In loving memory of my grandmother

April 12, 1928 – December 26, 2011

Contents at a Glance

Abstract

Propanediol metabolism occurs within a proteinaceous organelle in several bacterial species including *Citrobacter freundii* and *Lactobacillus reuteri*. The propanediol utilisation (Pdu) microcompartment shell is built from thousands of hexagonal shaped protein oligomers made from seven different types of protein subunits. In this Thesis, I investigate and analyse the structure and assembly of the bacterial microcompartment shell proteins.

One of the shell proteins characterised in this work, PduT, has a tandem canonical bacterial microcompartment (BMC) repeat within the subunit and forms trimers with pseudo-hexagonal symmetry. This trimeric assembly forms a flat approximately hexagonally shaped disc with a central pore that is suitable for binding a 4Fe-4S cluster. The essentially cubic shaped 4Fe-4S cluster conforms to the threefold symmetry of the trimer with one free iron, the role of which could be to supply electrons to an associated microcompartment enzyme, PduS.

The major shell protein PduB has a tandem permuted BMC repeat within the subunit and also forms trimers with pseudo-hexagonal symmetry. This shell protein closely resembles its homologous counterpart, EtuB; both possess three small pores formed within the subunits rather than a single pore at the centre of the pseudo-hexameric disc. The crystal structure of PduB provides insights into how substrates such as glycerol are able to use these pores as substrate channels. PduB appears to be able to pack within a sheet of PduA molecules, suggesting how the facet of the shell may be assembled.

The higher order packing of shell proteins was investigated using PduA. Residues important for the packing of molecules into sheets were mutated and the effects on crystal morphology and on the shape of structures formed within the bacterial cell were assessed. PduA appears to assemble into straws in the bacterial cell and mutation of these residues has a profound influence on the structures produced.

Preface

Compartmentalisation offers an important evolutionary advantage to many cells. It enables cells to carry out metabolic activities that require an environment which is different from that of cytoplasm. It was once thought that cellular compartmentalisation only exists in eukaryotic cells. However, the discovery of proteinaceous microcompartments in many bacterial cells revealed that cellular compartmentalisation is much more widespread than it was once thought to be.

This project on bacterial microcompartment has been an exciting and challenging research area for me. I am primarily intrigued by how the numerous different types of protein that form the microcompartment shell come together and function to enclose a metabolic activity. In this Thesis, there are three results Chapters:

The first result Chapter discusses the trimeric shell protein PduT. This Chapter describes how the crystal structure was solved using SAD methods and describes the unusual large central cavity, which is able to accomodate a 4Fe-4S cluster.

The second tackles another trimeric shell protein, PduB. In this Chapter, the crystal structure solved by molecular replacement reveals three subunit pores in which glycerols are trapped revealing their use as substrate channels.

Lastly, the tiling of hexameric shell protein PduA has been explored. This Chapter investigates how a point mutation in the interacting surfaces of the protein can alter the tiling *in vitro* and *in vivo*.

The Thesis finally concludes with a number of future challenges and direction that the research project can take as well as how this protein assembly can be re-engineered to be used for several different biotechnological applications.

This Thesis may not be read by many people and perhaps may end up just one of those required academic exercises in order to get my doctor's title. Well, it is comforting to know that at least one (you) is reading this Thesis and hopefully you will enjoy reading this as much as I enjoyed writing it.

ALLAN PANG

September 2012

Credits

Supervisor	**Professor Richard Pickersgill**
Collaborators	**Professor Martin Warren** (University of Kent)
	Dr. Stefanie Frank (University of Kent)
	Dr. Andrew Lawrence (University of Kent)
	Dr. Joshua Parsons (University of Kent)
	Professor Michael Prentice (University College Cork)
	Dr. Mingzhi Liang (University College Cork)
PhD Panel Members	**Dr. John Viles**
	Dr. Norbert Krauß
Masters of Mutations	**Mr Soschichiro Nagano**
	Dr. Ruth Rose
Pickersgill Group	**Dr. Arefeh Seyederabi**
	Dr. Teng Teng To
	Miss Saima Rehman
	Dr. Shuang Gu
	Miss Yumiko Tashiro-Thatcher
Viles Group	**Dr. Nadine Younan**
	Miss Helen Stanyon
	Mr. Christian Matheou
Funding from	**QMUL SBCS** (Graduating Teaching Scheme)
Twitter Advisers	**Dr. Jon Marles-Wright** (University of Newcastle)
	Dr. David Briggs (University of Manchester)
Synchrotron Resources	**Diamond Light Source**
	ESRF Grenoble
Help with figures	**Mr. Franz Koo**
Love and Support	**Friends and family**

Abbreviations

Alc
Alcohol dehydrogenase

Adh
Acetaldehyde dehydrogenase

AdoCbl
Adenosylcobalamin

BMC
Bacterial microcompartment

CA
Carbonic anhydrase

CCM
Carbon dioxide concentrating mechanism

CCD
Charge-coupled device

CoA
Coenzyme A

CsoS1
Carboxysome shell 1

DEAE
Diethylethanolamine

DNA
Deoxyribonucleic acid

EDTA
Ethylenediaminetetraacetic acid

EPR
Electron paramagnetic resonance

ESRF
European synchrotron radiation facility

Etu
Ethanol utilisation

Eut
Ethanolamine utilisation

Fe-S
Iron-sulfur

FPLC
Fast protein liquid chromatography

HCO_3^-
Bicarbonate

IMAC
Immobilised metal affinity chromatography

IMS
Industrial Methylated Spirit

IPTG
Isopropyl β-D-1-thiogalactopyranoside

LB
Lysogeny broth

MPD
2-Methyl-2,4-pentanediol

NAD
Nicotinamide adenine dinucleotide

NCS
Non-crystallographic symmetry

PBS
Phosphate Buffer Saline

PCR
Polymerase chain reaction

PDB
Protein data bank

PEG
Polyethylene glycol

PGA
3-Phosphoglyceric acid

Pdu
1,2-propanediol utilisation

RuBisCO
Ribulose 1,5-bisphosphate carboxylase

RuBP
Ribulose-1,5-bisphosphate

S200
Superdex 200

SAD
Single anomalous dispersion

SDS-PAGE
Sodium dodecyl sulfate – polyacrylamide gel electrophoresis

Se-Met
Seleno-methionine

TEM
Transmission electron micrograph

TEMED
Tetramethylethylenediamine

UV-Vis
Ultraviolet visible

Glossary of microcompartment

Shell Proteins: proteins that form the microcompartment shell; these enzymes form a polyhedral (usually icosahedral in shape) body that encloses several enzymes, which are involved in a metabolic reactions.

Carboxysomal shell proteins

CsoS1A: a major shell component of alpha-carboxysome shell; six subunits form a hexamer; these hexamers are able to form a tight molecular layer, as seen in the crystal packing and argued to represent the biological structure of the facet of microcompartment shell; a sulfate ion is found in the central pore, which indicates the ability to transport small molecules (Tsai et al., 2007). PDB: 2EWH and 2G13

CsoS1B: shell protein with an extra 12 amino acid in the C-terminal region compared to CsoS1A.

CsoS1C: a shell component of alpha-carboxysome shell, which showed only minor difference (one amino acid residue difference) from CsoS1A; the crystal showed a lattice translocation (Tsai et al., 2009). PDB: 3H8Y

CsoS1D: first crystal structure with fused BMC domains; three subunits form a trimer; the trimer further dimerises to form a two-tiered shell protein, with each dimer of trimers contains one trimer with an open pore while the other with a closed central pore (Klein et al., 2009). PDB: 3F56 and 3FCH

CsoS4A: carboxysomal shell protein that forms as a pentamer that is able to occupy the vertex of the microcompartment shell (Tanaka et al., 2008). PDB: 2RCF

CcmK1, **CcmK2** and **CcmK4**: structurally known shell proteins of beta-carboxysomes; these shell proteins formed as hexamers (Tanaka et al., 2009, Kerfeld et al., 2005). PDB: 3BN4, 3DN9, 2A1B, 2A18, 3CIM, 3DNC, 2A10

CcmL: beta-carboxysome shell protein that forms as pentamer (Tanaka et al., 2008). PDB: 2QW7

Eut Shell Proteins

EutM: subunits adopt a typical BMC fold and form as hexamer, and possibly a major component of eut microcompartment shell (Tanaka et al., 2010); EutM structures suggest formation of tight two-dimensional protein membrane (Takenoya et al., 2010). PDB: 3I6P, 3MPY and 3MPW

EutS: shell protein which has a bend in the hexagonal disk by about 40°; Gly-39 was found to be a crucial residue that allow such permutation; this protein is believed to be responsible for the curvature of the molecular hexagonal layer (Tanaka et al., 2010). PDB: 3I96 and 3IA0

EutL: a trimeric shell protein which possesses three small subunit pores; the central pore can be triggered to open specifically upon exposure of zinc ions (Sagermann et al., 2009, Tanaka et al., 2010). PDB: 3GFH, 3MPV, 3I82, 3I87

EutK: the crystal structure of extended C-terminal region of this shell protein has a positively charged helix-turn-helix motif, suggesting a possible DNA binding site (Tanaka et al., 2010). PDB: 3I71

EutN: this shell protein adopts a Pf00319 domain, similar to CcmL and CsoS4A but unlike the two, this forms as hexamers (Tanaka et al., 2008). PDB: 279H

Etu Shell Proteins

EtuA and **EtuB**: the two known shell proteins to exist in etu microcompartment; the crystal structure of EtuB reveals presence of three small subunit pores (Heldt et al., 2009). PDB: 3IO0

Pdu Shell Proteins

PduA: one of the major shell components of pdu microcompartment; a hexameric shell protein which has a central pore that possibly allow transit of 1,2-propanediol (Crowley et al., 2010). PDB: 3NGK

PduB: a trimeric shell protein which has three small subunit pores, which allow transit of small molecules such as glycerol. PDB: 4FAY

PduJ: shell protein which shares 82% sequence identical to PduA, suggesting similar architecture.

PduK: shell protein which has an extended C-terminal region, which was predicted to have an iron-sulfur cluster binding site.

PduN: protein sequence suggests that the shell protein has Pf00319 domain and possibly form a pentamer that would fit in the vertex of pdu microcompartment shell.

PduT: a trimeric shell protein which has a wide central pore, the presumed binding site for 4Fe-4S cluster (Crowley et al., 2010, Pang et al., 2011). PDB: 3PAC and 3N79

PduU: a hexameric shell protein with long N-terminal regions, which extends away from the hexagonal disk, six of these assemble to form a tightly wound six β-barrel strands, blocking the central pore (Crowley et al., 2008). PDB: 3CGI

Enzymes in Eut microcompartment	Enzymes in Pdu microcompartment
EutBC: ethanolamine ammonia lyase	**PduCDE**: diol dehydratase
EutG: alcohol dehydrogenase	**PduGH**: diol dehydratase reactivation protein
EutE: aldehyde dehydrogenase	**PduS**: cobalamin reductase
	PduO: cobalamin adenosyltransferase
	PduQ: propanol dehydrogenase
	PduP: CoA-dependent propionaldehyde dehydrogenase
	PduL: phosphotransacylase
	PduW: propionate kinase

List of Figures

List of Tables

“You'd see those beautiful spars, those perfect geometric facets, and you could ask, ‘What's controlling this process? How does the crystal end up so perfectly formed - and looking so much like other crystals?’ But it turns out a crystal is just the way molecular forces arrange themselves in solid form. No one controls it. It happens on its own. To ask a lot of questions about a crystal means you don't understand the fundamental nature of the processes that led to its creation. So maybe living forms are a kind of crystallisation. Maybe life just happens. And maybe, like crystals, there's a characteristic order to living things that is generated by their interacting elements. Okay. Well, one of the things that crystals teach us is that order can arise very fast. One minute you have a liquid, with all the molecules moving randomly. The next minute, a crystal forms, and all the molecules are locked in order.”

Dr. Ian Malcolm

Quoted from *The Lost World* by Michael Crichton

Chapter 1

Introduction to protein structure

Proteins are biological macromolecules built from monomeric units called amino acids. Amino acids are comprised of a central carbon, an amino group (NH_2), a carboxylic acid group (COOH), a H atom and an R group (side chain), as illustrated in Figure 1.1A. At neutral pH (around pH 7), the amino acid is in its zwitterionic form NH_3^+ and COO^-. Twenty amino acids are encoded by DNA; each of which carries a different functional R group (Figure 1.1C). Due to their low pKa, the carboxylate side chains of Asp and Glu are deprotonated at neutral pH, and therefore, negatively charged. On the other hand, the basic amino acids Arg and Lys have high pKa, and therefore, their side chains are protonated and positively charged. Histidine bears an imidazole side chain which has nitrogen that can act as weak acid at low pH. At a physiological pH of 7.4, histidine is about 90% in the basic form and 10% in the acid form. Depending on its local environment, histidine can serve as a proton acceptor or proton donor. Four amino acids (Ser, Thr, Asn and Gln) have polar and uncharged side chains. These chemical properties favour their presence at the protein surface. On the other hand, eight amino acids (Ala, Val, Ile, Leu, Met, Phe,

Tyr and Trp) have hydrophobic side chains, which mean they prefer to be buried in the interior of the protein. Glycine is the smallest and only naturally occurring non-chiral amino acid with hydrogen substituent as its side chain and therefore can fit in either hydrophilic or hydrophobic environments. The side chain of cysteine is easily oxidised to form disulfide bond, forming cystine. Proline is unique because the amino nitrogen is bound to two alkyl groups instead of one, making it a secondary amine or an imine.

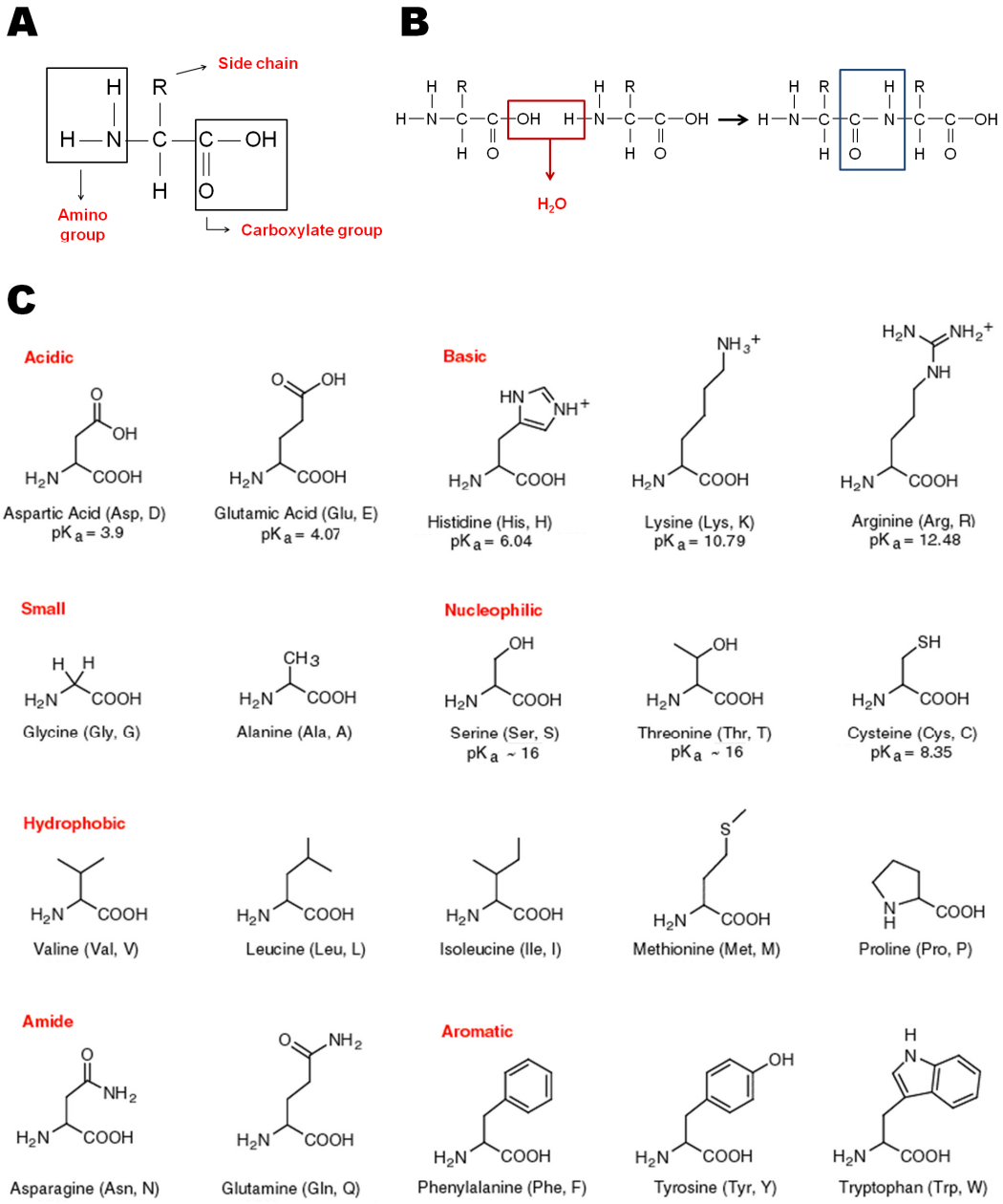


Figure 1.1 Amino acid structures. (A) The generic structure of an amino acid in its unionised form, (B) Amino acids linked together by a peptide bond between the carboxylic acid group of the first amino acid and the the amino group of the next, as a consequence, a water molecule is released; (C) The twenty common amino acids, adapted from www.neb.com.

Levels of protein structure

There are four distinct levels of protein structure; the primary, secondary, tertiary and quaternary structure. The primary structure refers to the linear sequence of the amino acid residues. Individual amino acids are joined together to form a polypeptide chain by a peptide bond (Figure 1.1B). The peptide bond is a covalent chemical bond formed between the carboxyl group of one amino acid and the amino group of the neighbouring amino acid. By convention, the amino group is placed on the left of the central carbon and the carboxylic group on the right. It is also conventional that protein sequences are written starting from the N-terminal region (amino group) to the C-terminal region (carboxylic group).

The common secondary structures are α-helices and β-sheets. In the absence of regular secondary structure, the polypeptide chain is said to be unstructured or irregular in structure. The secondary structure can be predicted with variable accuracy from the primary structure. Amino acids such as methionine, alanine, leucine, glutamate and lysine prefer to adopt helical conformations; while amino acid residues containing large aromatic structures (tryptophan, tyrosine and phenylalanine) and C_β-branched amino acids (isoleucine, valine and threonine) have a tendancy to adopt β-strand conformations. Proline and glycine break the α-helical backbone conformation and are commonly found in chain reversals or turns. The secondary structure elements are stabilised by hydrogen bonds along the main chain. In α-helices, the NH group of an amino acid donates a hydrogen bond to the C=O group of the amino acid four residues earlier (Figure 1.2A). β-sheets are formed from either parallel or anti-parallel polypeptide chains. The side chains of adjacent

residues point alternately up and down along the length of the β-strand. Hydrogen bonding occurs between adjacent β-strands to form a β-sheet (Figure 1.2B). Secondary structure elements often packed together to form a structural motif. The BMC shell proteins, which will be discussed later, are composed of β-α-β structural motifs (Figure 1.2C).

The tertiary structure describes the full three-dimensional fold of a single protein molecule. Proteins tend to fold to a certain conformation, in order to perform a specific biological function. Folded proteins (native state) are commonly more stable and energetically more favourable (lower energy) than their unfolded state. The tertiary structure of a protein is maintained by non-covalent interactions involving both the side chains and the main chain, though clearly it is the side chains that determine the fold of the protein molecule (Figure 1.3). Salt bridges can be formed between positively and negatively charged groups (Figure 1.3A). Hydrogen bonds occur between hydrogen and an electronegative atom such as nitrogen and oxygen (Figure 1.3B). A covalent bond, the disulfide bond, can be formed between two cysteine residues (Figure 1.3C). The driving force for the folding of proteins in aqueous solvent is the hydrophobic effect, due to the burial of hydrophobic side chains from the polar (aqueous) environment (Figure 1.3D). Shell proteins share a distinct conserved protein domain, or protein folding unit, called the BMC domain (Figure 1.4A) and Pf00319 domain.

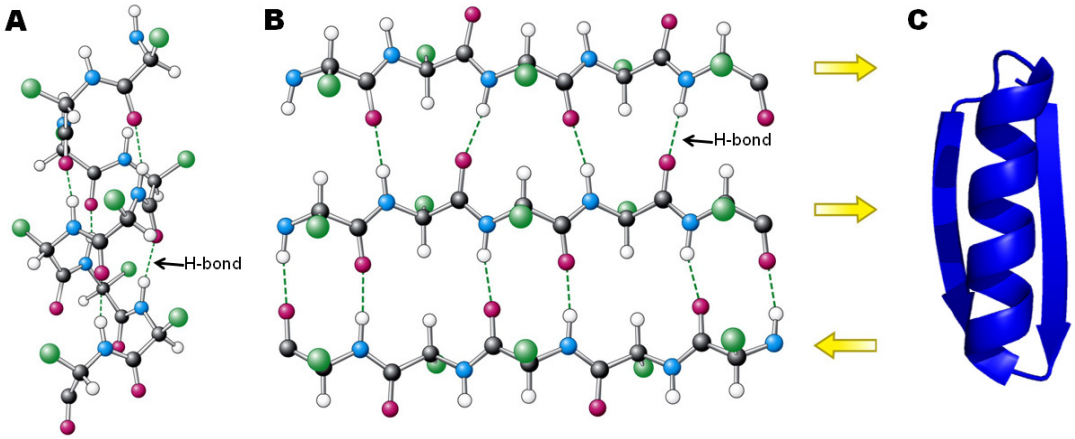


Figure 1.2 Secondary structures. A ball and stick representation of (A) α-helix and (B) β-pleated sheet; carbon atom represented by black circle, nitrogen atom represented by blue circle, hydrogen atom represented by white circle, oxygen atom represented by red circle, side chains represented by green circle, hydrogen bonds are noted (broken green lines); adjacent β-strands can either in parallel (same direction) or anti-parallel (different direction). Figures taken from (Berg et al., 2006), (C) Cartoon representation of a right handed β-α-β motif.

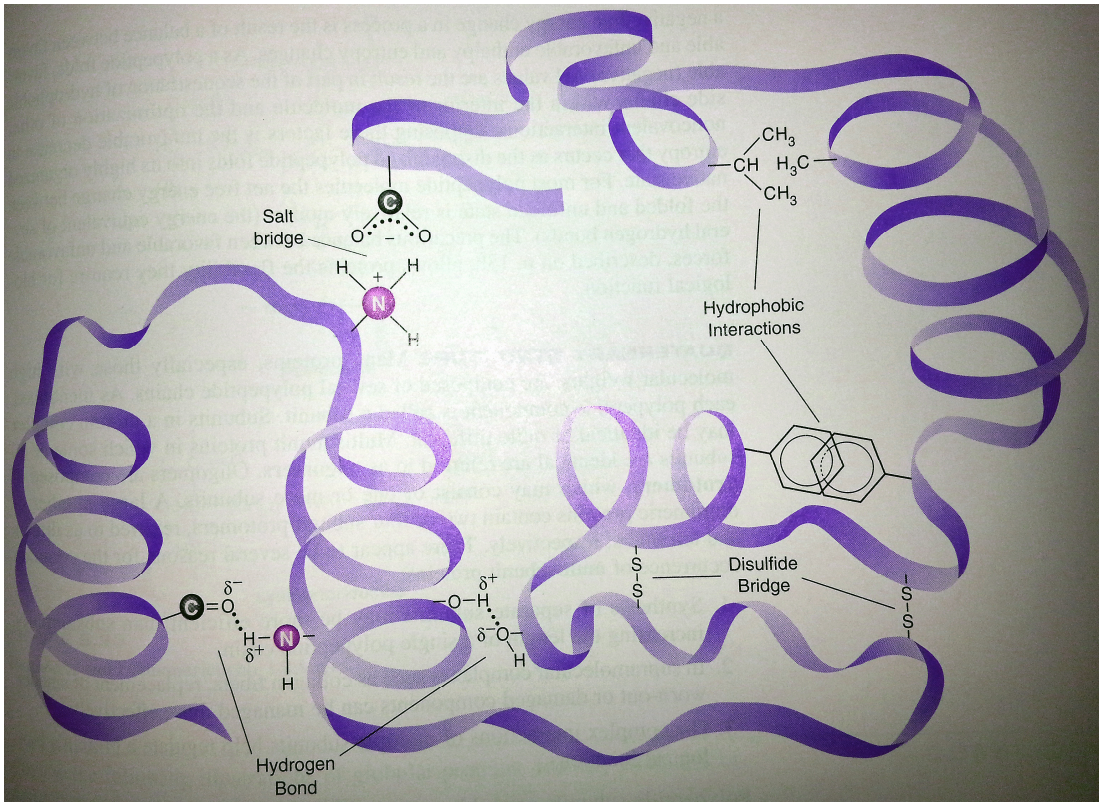


Figure 1.3 Interactions that stabilise the tertiary structure. Salt bridges, hydrophobic interactions, disulfide bridges and hydrogen bonds influence and determine the overall shape of the protein. Figure taken from (McKee and McKee, 2009).

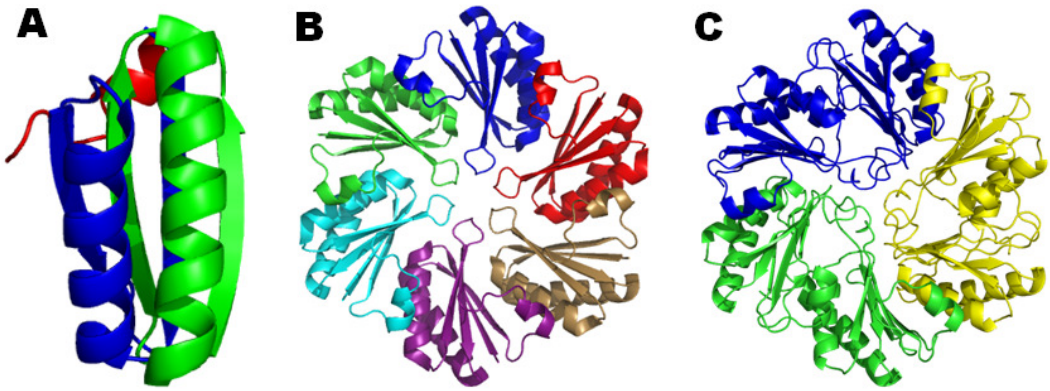


Figure 1.4 Examples of tertiary and quaternary structure. (A) PduA subunit which consist of a conserved protein fold known as BMC domain, (B) PduA subunits further assemble into hexamer, which has a six fold symmetry; PDB: 3NWG, (C) PduB subunits, which consist of two BMC domains, assemble into a trimer, which has a three-fold symmetry; PDB: 4FAY.

Symmetry of oligomeric proteins

When individual protein subunits come together to form a higher level of organisation, this is known as the quaternary structure. The quaternary structure of a protein is held together by the same interactions that stabilise the tertiary structure (Figure 1.3). The quaternary structure may either be composed of two or more different subunits ("hetero-multimers") or of identical subunits ("homo-multimers"). Most protein subunits in oligomers are arranged about one or more axes of rotational symmetry. This arrangement of rotation axes forms a point group of which there are four classes: cyclic, dihedral, cubic and icosahedral. The BMC proteins discussed in this Thesis are examples of oligomers where the subunits are related by 6-fold (Figure 1.4B) or 3-fold rotation axes (Figure 1.4C). Hexameric, trimeric and

pentameric shell proteins further assemble to a higher organisational level to produce a polyhedral shell complex, which will be discussed in depth later in this Thesis.

Ther are three basic types of symmetry operator: (1) mirror plane, (2) rotation axis and (3) inversion. The application of these operators gives rise to the 32 point groups (Table 1.1). However when investigating the symmetry of multi-subunit proteins, only the point groups which contain cyclic symmetry and dihedral symmetry need be considered as inversion and reflection (mirror plane) change the chirality of the molecule, which is not possible for proteins made of chiral "L" amino acid residues. This reduces the possibilities to just 12 point groups.

A multimeric protein can crystallise with the molecular symmetry axes aligned with the crystallographic symmetry axes. An example of this is PduT (discussed in Chapter 3), where the molecular 3-fold axis corresponds to the crystallographic 3-fold axis and there is a single subunit in the asymmetric unit. In other examples, the asymmetric unit may contain more than one molecule. In PduB (discussed in Chapter 4), the three-fold rotational symmetry of the multimer does not correspond to a crystallographic axis and there is a trimer in the asymmetric unit with the subunits related by a non-crystallographic 3-fold axis.

Table 1.1 The 32 crystallographic point groups.

Crystal System	Essential Symmetry	Point Groups
Triclinic	None	1, $\overline{1}$
Monoclinic	2 or *m*	2, *m*, 2/*m*
Orthorhombic	222 or *mm*2	222, *mm*2, *mmm*
Tetragonal	4 or $\overline{4}$	4, 422, $\overline{4}$, 4/*m*, 4*mm*, $\overline{4}$ 2*m*, 4/*mmm*
Trigonal	3 or $\overline{3}$	3, $\overline{3}$, 32, 3*m*, $\overline{3}$ m^2
Hexagonal	6 or $\overline{6}$	6, 622, $\overline{6}$, 6/*m*, 6*mm*, $\overline{6}$ 2*m*, 6/*mmm*
Cubic	23	23, 2/*m* $\overline{3}$, 432, $\overline{4}$ 3*m*, 4/*m* $\overline{3}$ 2/*m*

Introduction to metabolosomes

In 1956, Drews and Niklowitz observed an intracellular structure within the cyanobacterium *Phormidium uncinatum* (Drews and Niklowitz, 1956). The organelles appeared similar to viral capsids which led to the speculation that they might be phages (Bock et al., 1974). In 1969, Gantt and Conti found similar intracellular polyhedral bodies in many other blue green alga. Using electron microscopy, they described the polyhedral structures as crystalline bodies with highly ordered array (Gantt and Conti, 1969). In 1973, these polyhedral inclusions were isolated and discovered to contain an enzyme known as ribulose bisphosphate carboxylase (RuBisCO) (Shively et al., 1973). The investigators proposed to name the bacterial organelle as carboxysome. It was later discovered that these structures exist in many other prokaryotic organisms (Shively et al., 1998, Cannon et al., 2001, Kofoid et al., 1999, Havemann and Bobik, 2003). As opposed to membrane-bound eukaryotic organelles, the polyhedral bodies consist of proteinaceous shell. Thus, the structure is generally known as "bacterial microcompartment" (BMC). Further investigation of these polyhedral bodies revealed that there are at least seven different types of bacterial microcompartment (Bobik, 2007). Figure 1.5 (Kerfeld et

al., 2010) shows some examples of transmission electron micrographs of BMCs. Several BMCs are far more complex than the carboxysome. They consist of various functional metabolic enzymes within the proteinaceous shell, and thus, these microcompartments were proposed to be referred as "metabolosomes" – structural bodies that enclose metabolic activities. Four types of bacterial microcompartments have well-established function and will be discussed in this Thesis: carboxysome, ethanol utilisation (Etu), ethanolamine utilisation (Eut), and 1,2-propanediol utilisation (Pdu) bacterial microcompartments. Recently available crystal structures of shell proteins and associated PDB codes as well as short definition of each were listed in the glossary (pages xiii and xiv).

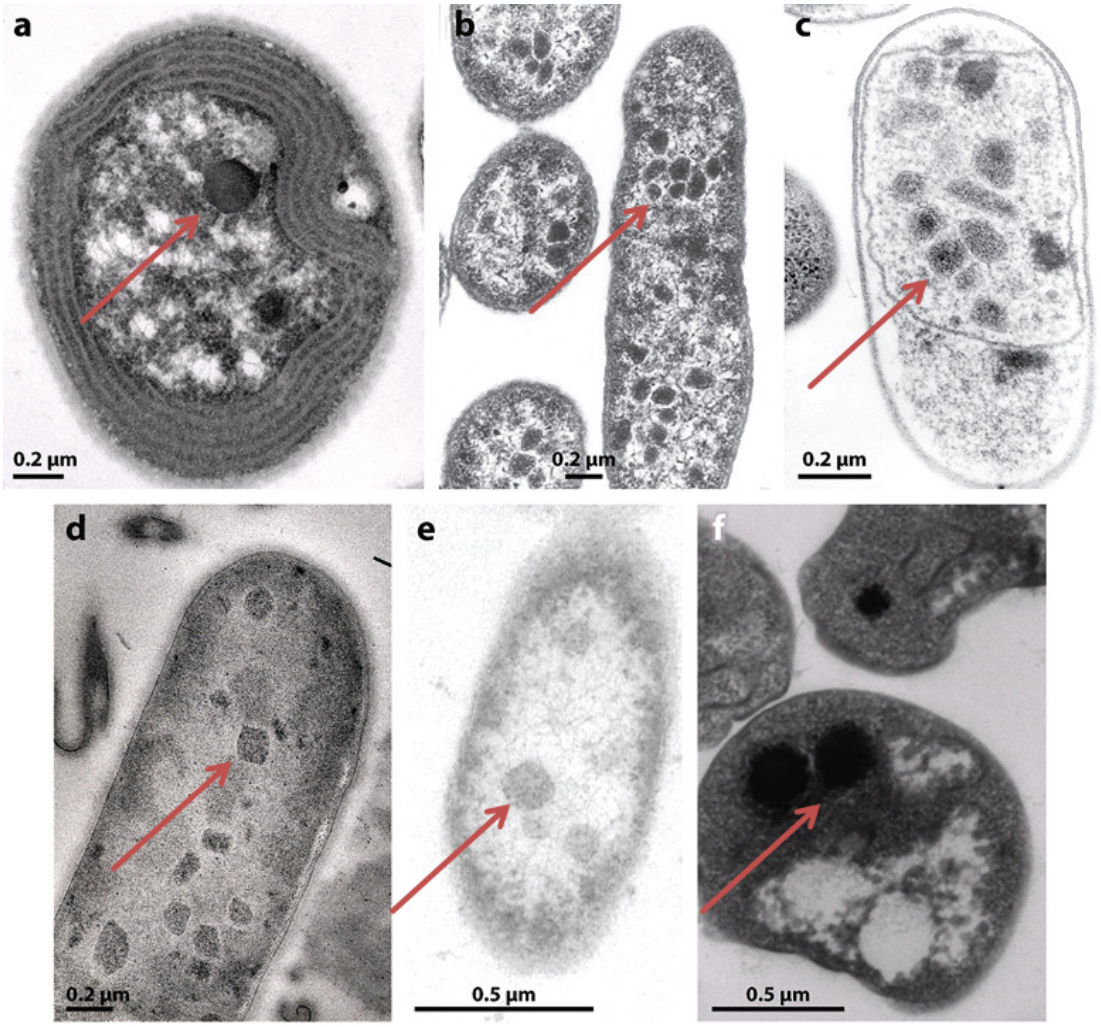


Figure 1.5 Transmission electron micrographs of bacterial microcompartments (BMCs), adapted from Kerfeld et al. (2010). **(a)** *Synechocystis* carboxysome, **(b)** Pdu and **(c)** Eut metabolosome in *Salmonella*, **(d)** Etu metabolosome in *Clostridium kluyveri*, **(e)** BMC in *Clostridium phytofermentans* grown on fucose, **(f)** BMC found in *Pirellula staleyi*. BMCs are marked with red arrows.

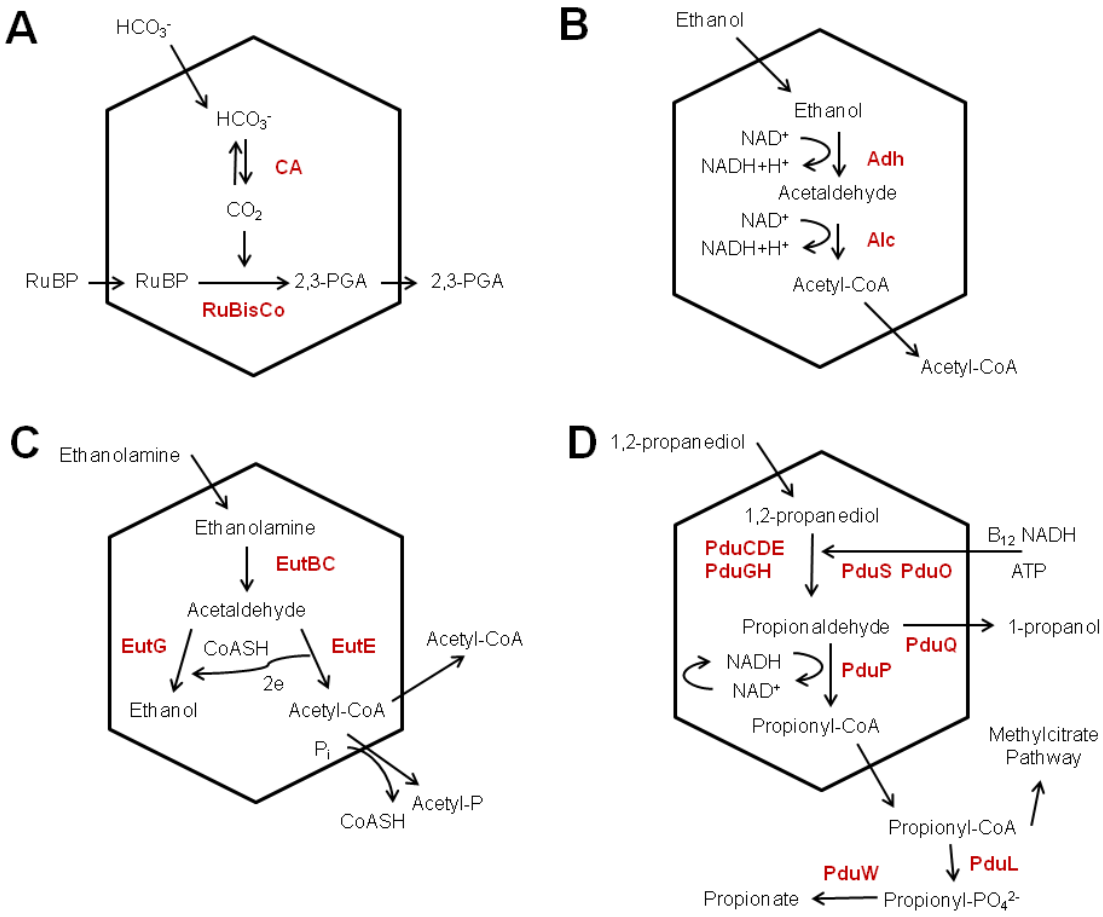


Figure 1.6 Schematic representation of the metabolic pathway of **(A)** carboxysome, **(B)** Etu BMC, **(C)** Eut BMC, and **(D)** Pdu BMC. Enzymes involved in the pathway are highlighted in red. CoA and electron were required for propanediol metabolism but not shown in the figure. HCO_3^- = bicarbonate, CA = carbonic anhydrase, CO_2 = carbon dioxide, RuBP = ribulose-1,5-biphosphate, RuBisCo = ribulose-1,5-bisphosphate carboxylase, 2,3-PGA = 2,3-phosphoglycerate, NAD = nicotinamide adenine dinucleotide, Adh = acetaldehyde dehydrogenase, Alc = alcohol dehydrogenase, EutBC = ethanolamine ammonia lyase, EutE = aldehyde dehydrogenase, EutG = alcohol dehydrogenase, CoASH = coenzyme A, PduCDE = diol dehydratase, PduGH = diol dehydratase reactivation protein, PduS = cobalamin reductase, PduO = cobalamin adenosyltransferase, PduQ = propanol dehydrogenase, PduP = CoA-dependent propionaldehyde dehydrogenase, PduL = phosphotransacylase, PduW = propionate kinase

1.1 Functions

1.1.1 Carboxysome

The first isolation of a carboxysome in 1973 identified the presence of ribulose bisphosphate carboxylase (RuBisCO), an enzyme central to the Calvin cycle (Shively et al., 1973). After its initial discovery, two differing hypotheses were put forward to explain the function of the microcompartment: carboxysomes are either storage bodies for inactive RuBisCO or organelles for CO_2 fixation. Experimental evidences supported the latter theory. Isolated particles were found to contain highly active RuBisCO and there was no difference in kinetic and inhibiting binding properties of isolated cytoplasmic and carboxysomal RuBisCO (Cannon et al., 1991). In addition, when cyanobacteria and chemoautotrophic bacteria were placed in a CO_2 limited media, it was observed that the carboxysomal RuBisCO activity and volume density of carboxysome increased proportionally. These proved that carboxysomes act as an active CO_2 fixing organelle for the bacteria (Beudeker et al., 1980, Miller et al., 1984). The 'organelle hypothesis' was also supported by various mutagenesis studies. Mutagenised cyanobacteria lacking or have deformed carboxysomes required a higher concentration of CO_2, suggesting carboxysomes play an essential role in the carbon concentrating mechanism and actively involved in the Calvin cycle (Price and Badger, 1989, Price and Badger, 1991, Price et al., 1993).

RuBisCO accounts for 60% of carboxysomal proteins with an average of 300 molecules of enzyme per carboxysome (Cannon and Shively, 1983). The compartmentalisation role of carboxysome is believed to compensate the low

enzymatic rate of RuBisCO. RuBisCO typically fixes only 3 – 10 CO_2 molecules each second per molecule of enzyme (Ellis, 2010). Through carboxysomal compartmentalisation, the enzymes were able to operate at or near substrate saturation. The concentration of the substrate CO_2 was increased in the vicinity of the enzyme within the microcompartment. Autotrophs obtain CO_2 through dissolved inorganic carbon (DIC). DIC, which consists largely of bicarbonate, presumably diffuses into the microcompartment. Carbonic anhydrase catalyses the conversion of bicarbonate into CO_2, which in turn is used in carboxylation of ribulose-1,5-bisphosphate (RuBP) via RuBisCO (Figure 1.6A). As the enzyme RuBisCO is also able to catalyse oxygenation of RuBP, it is thought that the carboxysomal compartmentalisation prevents O_2 from entering the microcompartment and therefore prevents the competition against CO_2 fixation (Cannon et al., 2001, Marcus et al., 1992). Such mechanism of carboxysomal function was highly supported by several experimental data and quantitative models of CO_2 assimilation and growth of cyanobacteria (Badger et al., 1985, Reinhold et al., 1991, Reinhold et al., 1987).

1.1.2 Ethanol utilisation (Etu) metabolosome

Clostridium kluyveri are Gram positive anaerobes able to ferment ethanol and acetate to butyrate, caproate and H_2 (Bornstein and Barker, 1948). This process is carried out by two enzymes, the NAD-dependent alcohol dehydrogenase (alc) and acetaldehyde dehydrogenase (adh) to produce acetyl-CoA (Burton and Stadtman, 1953). Genomic analysis of the bacteria revealed that the genes required for the ethanol-acetate fermentation are clustered near microcompartment proteins (Seedorf

et al., 2008), consistent with the electron-micrograph of a thin section of *C. kluyveri* grown on ethanol and acetate (Lurz et al., 1979).

The genes for the ethanol metabolism are encoded within an operon that includes the enzymes and two known shell proteins. Within the microcompartment, the ethanol undergoes two oxidation steps to a final product of acetyl-CoA, which presumably diffuses out to the bacterial cytoplasm (Figure 1.6B). It is strongly believed that the microcompartment protects the cell from the harmful effects of acetaldehyde by encapsulating the intermediate in a proteinaceous shell (Stojiljkovic et al., 1995). To date, the ethanol utilisation microcompartment is poorly understood and so far, only *Clostridium* was reported to possess such a microcompartment (Heldt et al., 2009).

1.1.3 Ethanolamine utilisation (Eut) metabolosome

Phosphotidylethanolamine is an abundant lipid found in mammalian cell membranes, which can be broken down into glycerol and ethanolamine by phosphodiesterase (White, 1973). A variety of bacteria such as *Salmonella*, *Clostridium* and *Klebsiella* are able to use this amino alcohol as carbon and nitrogen source (Chang and Chang, 1975, Garsin, 2010). The ethanolamine metabolism occurs within a microcompartment, known as the Ethanolamine utilisation (Eut) metabolosome. The *eut* operon (of *S. typhimurium*) encodes 17 genes necessary for the formation of the proteinaceous shell and the ethanolamine metabolism (Kofoid et al., 1999). Inside the metabolosome, as indicated schematically in Figure 1.6C, the ethanolamine is degraded into acetaldehyde and ammonia by ammonia lyase EutBC, which requires a co-factor adenosylcobalamin (AdoCbl). The intermediate acetaldehyde can be catabolised by EutG to form alcohol or by EutE to produce acetyl-CoA, which can be

utilised in various metabolic processes such as TCA cycle, the glyoxylate cycle or lipid biosynthesis. Some acetyl-CoA is converted by phosphotransacetylase EutD to acetylphosphate, which is subsequently phosphorylated by housekeeping acetate kinase Ack into acetate. Similar to Etu metabolosome, the Eut microcompartment is thought to play an important role of enclosing acetaldehyde (Stojiljkovic et al., 1995) and essentially EutG helps in conversion of the intermediate to prevent the build-up of toxic levels of acetaldehyde.

1.1.4 1,2-Propanediol (Pdu) utilisation metabolosome

1,2-propanediol is produced after fermentation of two common plant cell wall sugars rhamnose and fucose (Obradors et al., 1988). The organic compound can be used as a source of carbon and energy by enteric bacteria such as *Klebsiella*, *Shigella* and *Yersinia* (Bobik, 2006) and also by bacteria residing in an environment rich in 1,2-propanediol such as aquatic sediments. Electron microscopic studies of various bacteria such as *Salmonella*, *Klebsiella* (Shively et al., 1998) and *Lactobacillus* (Sriramulu et al., 2008) grown in the presence of 1,2-propanediol reveals the appearance of multiple icosahedron-shaped bodies that appear similar to carboxysomes. This is now known as the 1,2-propanediol utilisation metabolosome. Figure 1.6D shows the metabolic pathway of the 1,2-propanediol utilisation metabolosome.

1.2 The proteinaceous shell

1.2.1 BMC domain

A common feature shared by bacterial microcompartments is the proteinaceous shell encapsulating the enzymes, substrates, toxic and labile intermediates. The icosahedron-shaped proteinaceous shell has a size of about 80-140 nanometers in diameter, which is made up of several thousand polypeptides of different types of shell proteins. These shell proteins are easily distinguishable by their conserved protein fold known as the bacterial microcompartment (BMC) domain (InterPro domain IPR000249) (Beeby et al., 2009, Kerfeld et al., 2005). The BMC domain is typically 90 amino acid residues and adopts an α/β fold. When shell proteins are compared to capsomeres, which are protein subunits that constitute icosahedral viral capsids (Vernizzi and Olvera De La Cruz, 2007); the folding motifs found in BMC domain bear no similarity with any of the known structures of capsomeres.

The BMC domain constitutes the major part of the shell protein subunit. A shell protein subunit may either consist of one or two BMC domains, which self-assemble into cyclical disk-shaped hexamers and pseudohexameric trimers, respectively (Figure 1.7). These shell proteins are able to form a molecular layer by tiling, a process where the hexamers fit together to form a sheet of molecules. This assembly constitutes the facets of the shell, while occupying the vertices of the shell are pentameric shell proteins. The subunit of the pentameric shell proteins consists of a different fold compared to the BMC domain of the hexamers and trimers. The fold

belongs to the family Pf03319, which adopts a series of 5 β-sheets that form a circular shape.

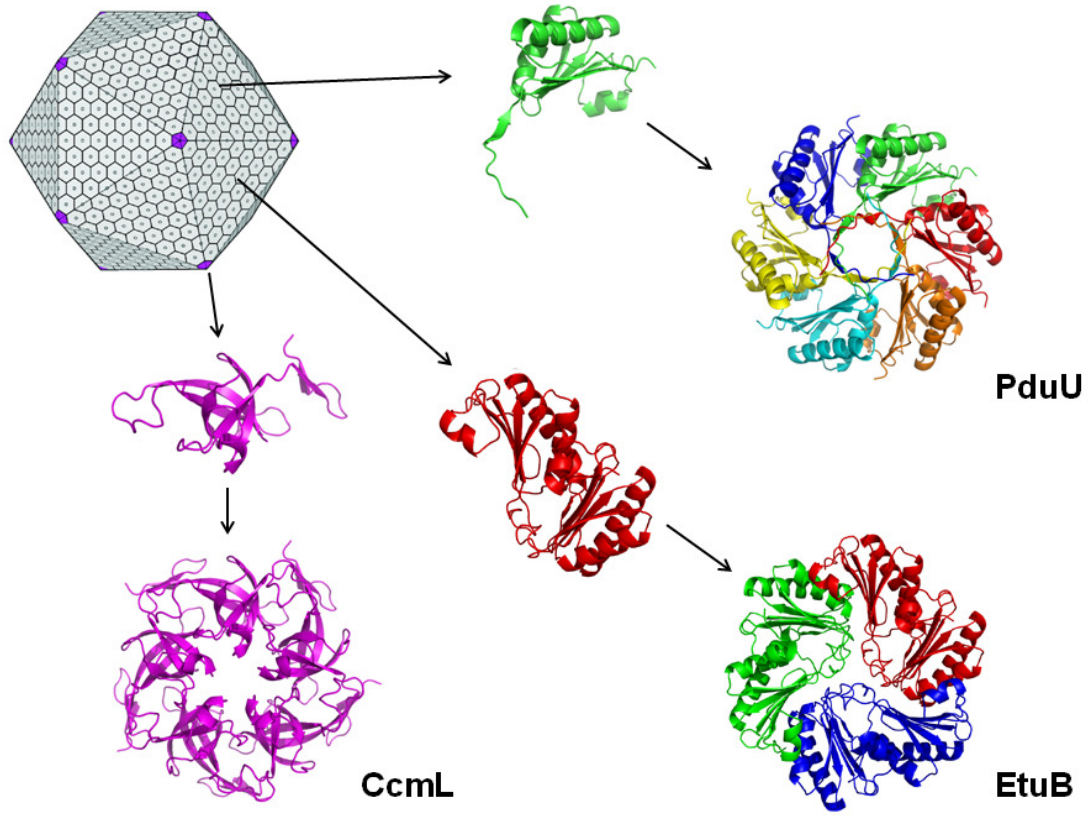


Figure 1.7 Structural assembly of microcompartment shell. BMC shell proteins (top) form hexamers that constitute the main building blocks of the shell, an example here is PduU (PDB: 3CGI) (Crowley et al., 2008). BMC shell proteins with a tandem repeats form trimers, for instance, EtuB (PDB: 3IO0) (Heldt et al., 2009). In the vertices of the BMC shell (coloured in magenta), a pentameric shell protein, CcmL (PDB: 2QW7) (Tanaka et al., 2008), had been established in carboxysomes, but not yet in any other BMCs.

1.2.2 Hexameric shell proteins

To date, there are eight crystal structures of shell proteins containing a single BMC domain (Tsai et al., 2007, Tsai et al., 2009, Tanaka et al., 2009, Takenoya et al., 2010, Crowley et al., 2010, Crowley et al., 2008). A representation of a hexameric single BMC domain shell protein (PduU) is shown in Figure 1.7. CcmK2 and CcmK4, two paralogs that form the β-carboxysome shell, were first described by Kerfeld, *et al*. in 2005. The hexameric structures were measured to be about 70 Å across. Similar architecture was also found in the other single BMC domain containing shell proteins. The common feature observed in these multiple structures is that six wedge-shaped subunits of the shell protein self-assemble into a cyclic hexamer.

1.2.2.1 Molecular tiling of hexamers

Most of these hexameric proteins are the major structural component of the metabolosome shell. The hexamers are arranged side-by-side to form a uniformly oriented molecular layer, which implies how facets of the microcompartment shell are assembled (Figure 1.8). This analysis is supported by the arrangement of hexamers in the crystal lattice, as seen in CcmK2 (Kerfeld et al., 2005) and CsoS1A (Tsai et al., 2007). Depending on the specific type of BMC and tightness of the layer, the interaction between hexamers is stabilised by several surface residues, including the conserved anti-parallel lysine residues (Tsai et al., 2007, Kerfeld et al., 2005, Yeates et al., 2011).

Not all single BMC domain containing shell proteins, however, exhibit a circular disk-shaped hexameric appearance. EutS broke the symmetry by possessing a bend

in the hexagonal disk by about 40° (Figure 1.9). A glycine (residue 39) substitution was found to be crucial for the bent structure, as mutation into valine converted EutS into a flat symmetric configuration (Tanaka et al., 2010). Such permutation is believed to possibly promote curvature in the flat molecular layer formed by the hexamers. So far, no other bent shell proteins have been identified in other bacterial microcompartments.

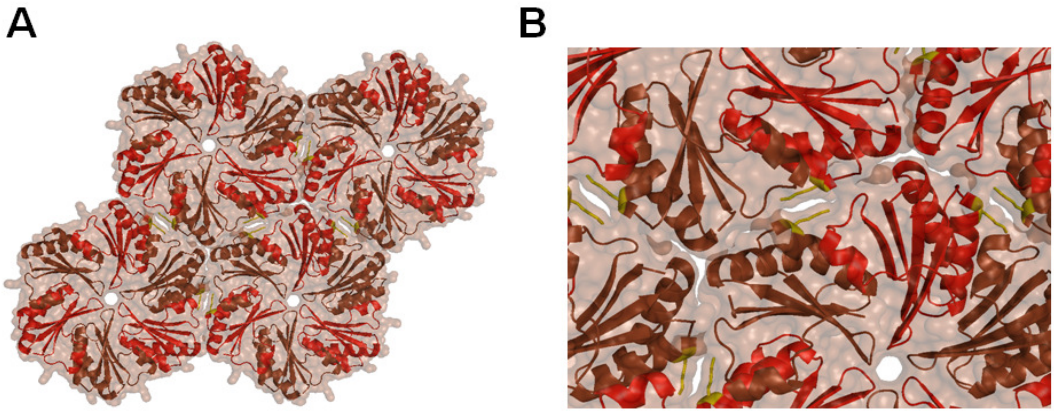


Figure 1.8 Molecular layer of CsoS1A hexamers. **(A)** CsoS1A (PDB: 2EWH) aligned side-by-side to create a molecular layer that represents the shell facet, **(B)** A close-up view of the layer. The interaction between hexamers are stabilised by anti-parallel lysines (yellow) (Tsai et al., 2007).

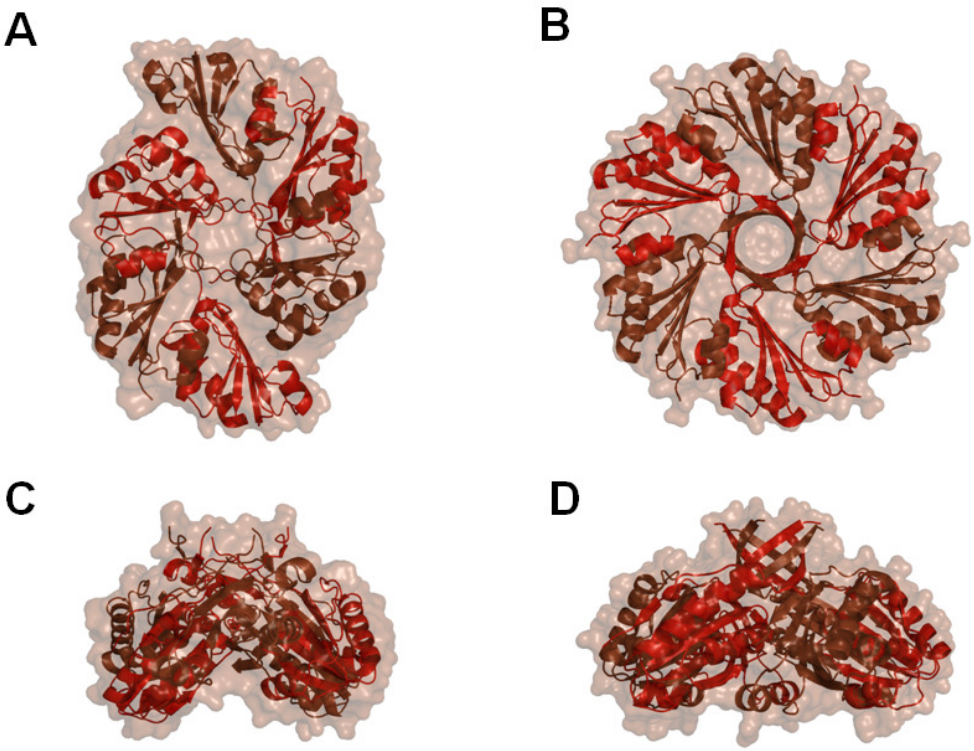


Figure 1.9 Bent EutS does not possess 6-fold symmetry. EutS **(A,C)** wild type and **(B,D)** mutant G39V, shown in top and side view. The wild type hexamer (PDB: 3I96) is curved away from a flat configuration by 40º. Mutation of G39V (PDB: 3IA0) converted into a flat symmetric configuration (Tanaka et al., 2010).

1.2.2.2 Structural variation of hexamers

Sequence and structural comparison of various single BMC domain shell proteins shows differences and variation in the BMC protein folding and the C-terminal region. The majority of the shell proteins exhibit a canonical BMC domain within its subunit while two of the known shell proteins (PduU and EutS) have permuted BMC domains. The two variations of the BMC domain protein folds will be discussed in more detail in Chapter 1.2.3.1 and Figure 1.13.

When the structure of permuted BMC domain (PduU) is aligned with a canonical BMC domain (such as PduA), the BMC core elements overlap but occur in a different order (Figure 1.10). Moreover, PduU has an extra long N-terminal and a short C-terminal region. This extra N-terminal region of PduU extends away from the hexagonal disk and six of these come together to form a tightly wound six β-barrel strands around the hexamer pore (Crowley et al., 2008). This deviation is not only seen between shell proteins of different BMCs, but also between paralogs of the same BMC. Extended 12-amino acid residues are found in CsoS1B but not in CsoS1A (Figure 1.11A). Such differences indicate that the C-terminal site of the subunits of the hexamers play a functional role unique to the specific shell protein; possibly to interact with specific enzymes within the microcompartment. In the Eut microcompartment, the C-terminal region of EutK shell protein is fused with a domain of approximately 60 residues (Figure 1.11B) (Tanaka et al., 2010). The fused domain adopts a positively charged helix-turn-helix motif, which in many instances implicates a binding site for nucleic acid.

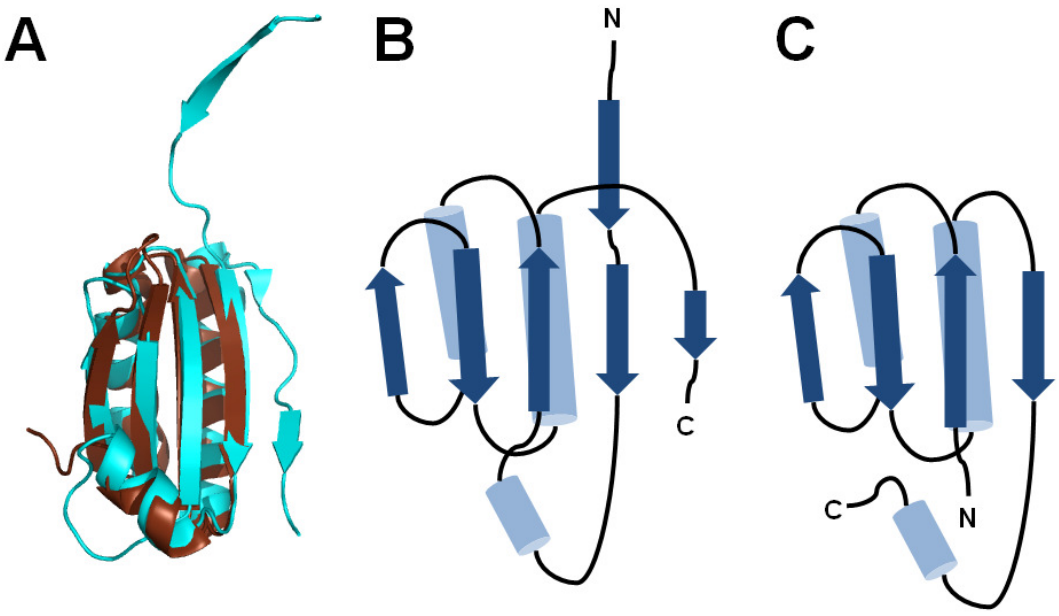


Figure 1.10 Structure comparison of PduA (canonical BMC domain) and PduU (permuted BMC domain) (Crowley et al., 2010, Crowley et al., 2008). **(A)** Superimposition of PduU (cyan) with PduA (brown) and schematic topology of the protein fold of **(B)** PduU and **(C)** PduA. Arrows represent beta sheet and cylinders represent alpha helix.

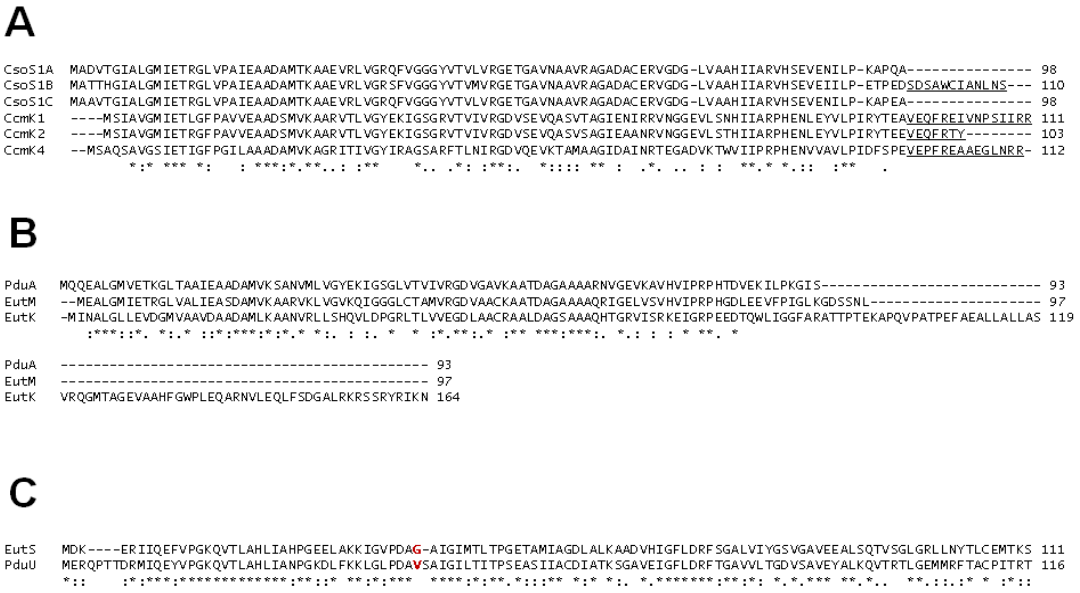


Figure 1.11 Sequence alignment of various hexameric shell proteins focusing on the C-terminal differences and crucial residue in shell protein with permuted BMC domain. **(A)** Comparison of carboxysomal hexameric shell proteins with underlined amino acid residues as the extended C-terminal region. **(B)** Comparison of remaining shell proteins (PduA, EutM, EutK) with single canonical BMC domain. **(C)** Comparison of permuted BMC domain shell proteins (EutS and PduU) with a highlighted (in red) glycine to valine difference.

1.2.3 Trimeric shell proteins

Sequence comparison of various shell proteins identified duplication of the BMC domain (about 180 – 250 amino acid residues) in some of the BMC gene clusters. So far, five crystal structures with tandem BMC domain containing shell proteins have been determined, representing each of the established BMCs (CsoS1D for carboxysomes, EtuB for Etu metabolosome, EutL for Eut metabolosome and, PduB and PduT for Pdu metabolosome). The tandem BMC protein subunits assemble to make a trimer, with pseudo six fold symmetry that structurally mimic the hexamers formed by single BMC domain shell proteins (Figure 1.12).

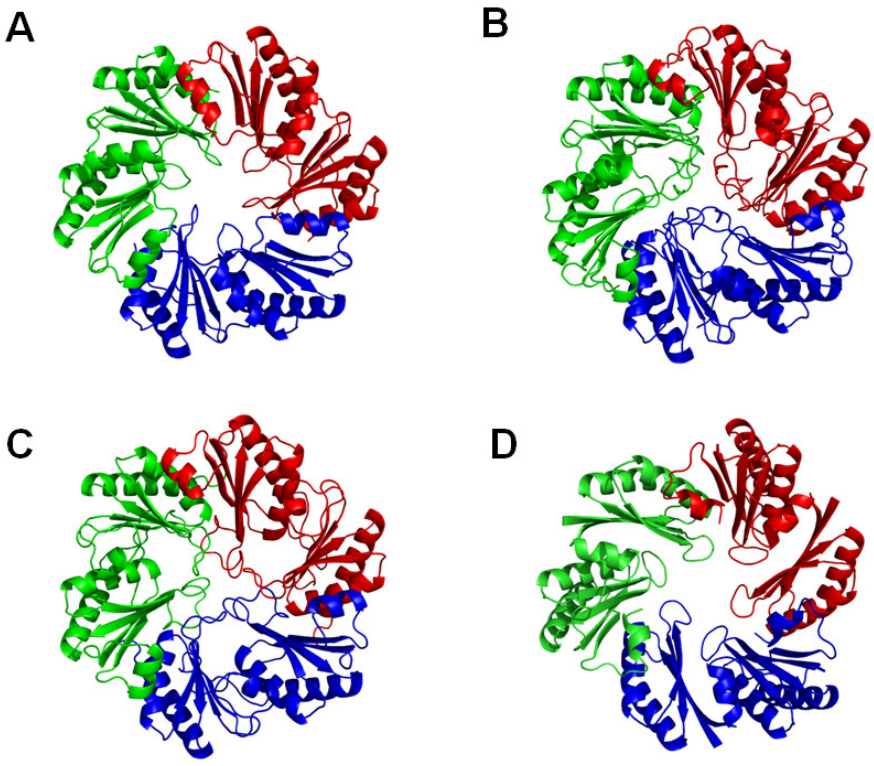


Figure 1.12 Tertiary structure of shell proteins with tandem BMC domain. **(A)** CsoS1D; PDB: 3FCH (Klein et al., 2009), **(B)** EtuB; PDB: 3IO0 (Heldt et al., 2009), **(C)** EutL; PDB: 3I82 (Sagermann et al., 2009), **(D)** PduT; PDB: 3PAC (Pang et al., 2011). Each subunit BMC domain coloured differently.

1.2.3.1 Fold variation within trimers

Figure 1.13 shows variations on the BMC domain. Tandem repeats can be fused together with two canonical BMC domains or two permuted BMC domains. Among the structurally known trimers, only PduT was built from two canonical BMC domains (Pang et al., 2011, Crowley et al., 2010), while the rest (CsoS1D, EutL, EtuB and PduB) comprise circular repeats of permuted BMC domains (Klein et al., 2009, Sagermann et al., 2009, Heldt et al., 2009). The duplicated permuted BMC domains closely resemble hexameric PduU. When EtuB and CsoS1D were superimposed onto EutL; this showed differences in the tertiary arrangement of CsoS1D compared to EutL and EtuB (Figure 1.9). Equivalent subunits of EtuB and EutL are superimposable while CsoS1D fits well only when its subunit occupies the position of two permuted BMC domains in different subunits of EutL (and EtuB).

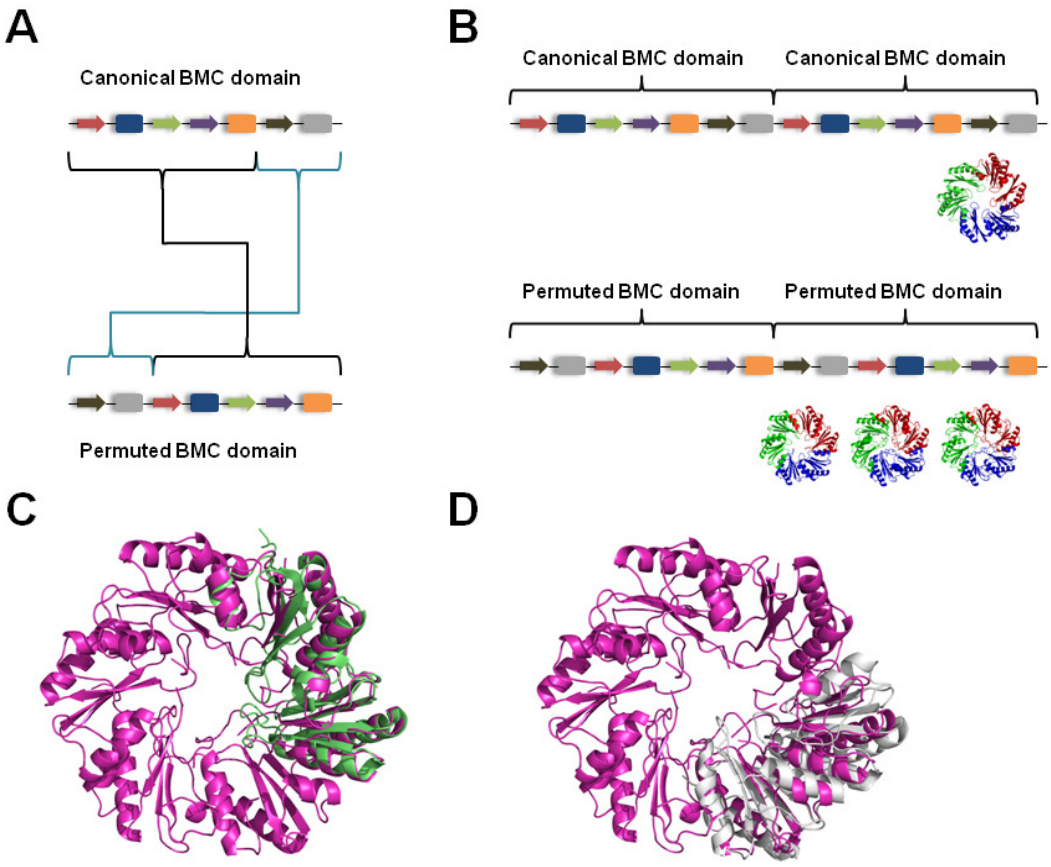


Figure 1.13 Variations on the BMC protein fold of trimeric shell proteins. **(A)** Schematic representation of the BMC protein in its canonical and permuted arrangements. Arrows and rectangles represent beta sheet and alpha-helix, respectively. The C-terminal beta sheet and alpha helix of canonical BMC domain are rearrange to the N-terminal of permuted BMC domain. **(B)** PduT (PDB: 3PAC) (Pang et al., 2011) exhibits double canonical BMC domain in its subunit, while CsoS1D (PDB: 3FCH) (Klein et al., 2009), EtuB (PDB: 3IO0) (Heldt et al., 2009) and EutL (PDB: 3I82) (Tanaka et al., 2010) have sequential permuted BMC domains in their subunits. **(C)** Superimposition of EtuB subunit (green) into EutL (magenta). **(D)** Superimposition of CsoS1D subunit (silver) into EutL (magenta).

1.2.3.2 Molecular fit of trimers

Compared to the hexamers, the relative abundance of trimers in their respective BMC shell is less. The disk-shaped trimers are expected to be dispersed within the molecular layer of hexameric shell proteins. This hypothesis is made possible by the observed packing of the conserved surface lysine residue found in each BMC domain. Shell proteins (CsoS1D, EtuB, EutL and PduB) have the exact fit and six lysines that are able to form interaction with the lysines of the adjacent hexamers (Klein et al., 2009, Heldt et al., 2009, Sagermann et al., 2009). Figure 1.14A shows that a representative trimeric shell protein (CsoS1D) fits in a molecular layer of hexamers (CsoS1A). The trimeric PduT, despite having a good fit into a layer of PduA hexamers, has uneven spacing between anti-parallel of lysines (Figure 1.14B) (Crowley et al., 2010, Pang et al., 2011). It is possible that the PduT trimer may be surrounded by a different shell component of the Pdu metabolosome.

The nature of how tandem domains arise is an evolutionary curiosity. It is suggested that fusion of these domains could aid faster shell assembly by increasing the effective concentration of protein domains with respect to each other (Kuriyan and Eisenberg, 2007). However, in contrast, this could also slow down the assembly of heterohexamers into a molecular layer due to the difference of edges and reducing the number of degrees of freedom of fit (Klein et al., 2009). The fusion of two domains breaks the 6-fold symmetry and accelerates the evolution of the molecule by allowing the sequence of adjacent BMC domains to vary allowing, for instance, the evolution of subunit pores.

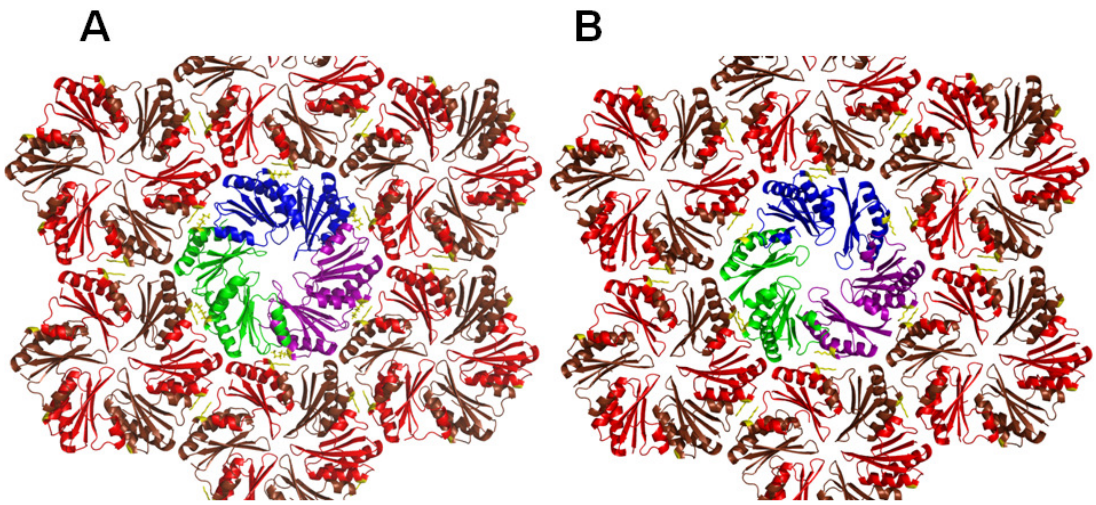


Figure 1.14 Trimeric shell proteins fit into a molecular layer of hexamers. **(A)** Model of interaction of CsoS1D trimer (PDB: 3FCH) (Klein et al., 2009) into a layer of CsoS1A hexamers (PDB: 2EWH) (Tsai et al., 2007). **(B)** Model of interaction of PduT trimer (PDB: 3PAC) (Pang et al., 2011) into a layer of PduA hexamers (PDB: 3NGK) (Crowley et al., 2010). Conserved lysine residues are shown in yellow.

1.2.3.3 Double-disk structures

Further structural variation of the pseudohexameric structures in the form of double disk structures have been observed in CsoS1D (Klein et al., 2009) and in a homologue of PduT (PDB ID: 3NWG). CsoS1D crystals showed two conformational states of the protein: a closed (Figure 1.15A) and open conformation (Figure 1.15B). CsoS1D dimerises with the broader side of the shell proteins facing each other (Figure 1.15C). Such stacking is argued to be a physiologically relevant form of the shell protein, but this has not been confirmed (Klein et al., 2009). Moreover, the interacting residues between two trimers are strongly conserved among orthologs of *Prochlorococcus* and *Synechococcus*. Interestingly, the stacking of the two trimers are offset by a rotation of around 60º across the interface; showing

that a monomer of one of the trimers interacts with two subunits of the other trimer (Figure 1.15D).

The two conformational states were not just found in CsoS1D trimers, but also in another tandem BMC domain containing shell protein, EutL (Sagermann et al., 2009). But unlike CsoS1D, the EutL trimers were not found to be stacked together as dimer of trimers. Also, PduT trimers for both *Salmonella* (Crowley et al., 2010) and *Citrobacter* (Pang et al., 2011) were found to be "single-layered" while the crystals of homologue shell protein of *Desulfitobacterium hafniense* (PDB ID: 3NWG) indicate dimerisation of trimers similar to CsoS1D.

The identification of double-disk structures (CsoS1D and PduT homologue) raises the possibility of the microcompartment shell formed by two layers. But at the same time, such dimerisation seems not to be consistent throughout all microcompartments. At this point, it is unclear whether the observation of trimer dimerisation was due to the crystal packing or perhaps, the dimerisation of the trimers only takes place in selected bacterial microcompartments.

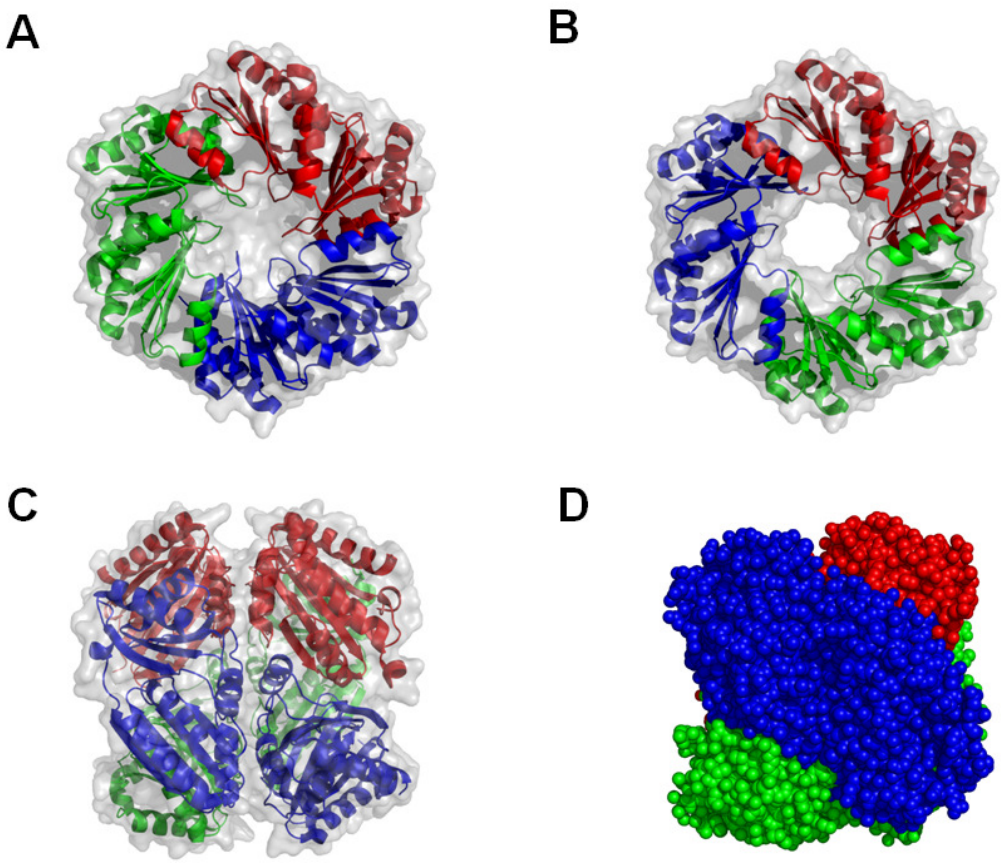


Figure 1.15 Dimers of Csos1D trimers. CsoS1D trimers (PDB: 3FCH) (Klein et al., 2009) exhibit a **(A)** closed conformation and an **(B)** open conformation. **(C)** Stacking of two CsoS1D trimers, as seen in the crystal. **(D)** Space filling representation of dimer of CsoS1D trimers, showing a rotational offset of ~ 60º of the trimer in respect to the opposite trimer.

1.2.4 The pentamers

In many viral capsids, pentameric proteins bend the molecular layer formed by the hexamers and trimers and occupy the vertices of the icosahedron. Since the architecture of viral capsids has a striking similarity with the architecture of carboxysomes, it is plausible that bacterial microcompartments may also have pentameric proteins to fill in the vertices of the shell. In 2008, Tanaka and colleagues reported two crystal structures of pentameric shell proteins from two different types of carboxysomes: CsoS4 from *H. neapolitanus* and CcmL from *Synechocystis*. Both structures have a distinct protein fold known as the Pf00319 domain. This particular domain differs greatly from that of the BMC domain. While the BMC domain is composed of a series of α/β fold, the Pf00319 domain is mainly comprised of beta sheets that form a beta barrel. The individual subunits of the pentamers are held together by the extensive interactions of C-terminal region with the nearby subunit. Structural comparison of two pentamers showed that CsoS4A appears to be smaller in size. The size difference is attributed to the lack of two C-terminal β-strands found in CcmL (Figure 1.16).

The idea that these pentamers occupy the shell vertices is further supported by two key observations. First is that these pentamers has a low relative abundance in the microcompartment (Cai et al., 2008, Tanaka et al., 2008). This means that only 60 copies are present per carboxysome shell, which clarifies why the shell proteins were not easily detected in earlier carboxysome preparations. Later experimental studies confirmed the expression of *csos4* genes (Cai et al., 2008). Second, mutational studies in which the deletion of the genes for the pentameric shell proteins leads to

the "rolling up" of the shell into an elongated rod, indicating the absence of component for the vertex formation (Price et al., 1993). However, a different mutagenesis study revealed that the situation may be more complicated than it seems (Cai et al., 2009). In this experiment, both normal and elongated carboxysomes were detected in the purified mutant carboxysomes, suggesting that microcompartment may reassemble to compensate the loss of the pentameric vertices.

A homologue of CsoS4 and CcmL containing Pf00319 domain, EutN, was shown to assemble as a hexamer, rather than a pentamer (PDB ID: 279H). EutN subunit is superimposable on the subunit of the pentameric shell proteins, indicating that the structural core of the Pf00319 domain is conserved for the three proteins. EutN has a closer resemblance to CcmL due to the presence of extended C-terminal region (Figure 1.16). It is suggested that the difference in the oligomeric state of EutN compared to CsoS4 and CcmL may provide an insight into the Eut microcompartment shell. It is plausible that the eut microcompartment does not have pentameric vertices which leads to an irregular icosahedral shape shell compared to that of carboxysomes. Sequence analysis indicates that PduN may possibly be in the same family of CsoS4/CcmL/EutN.

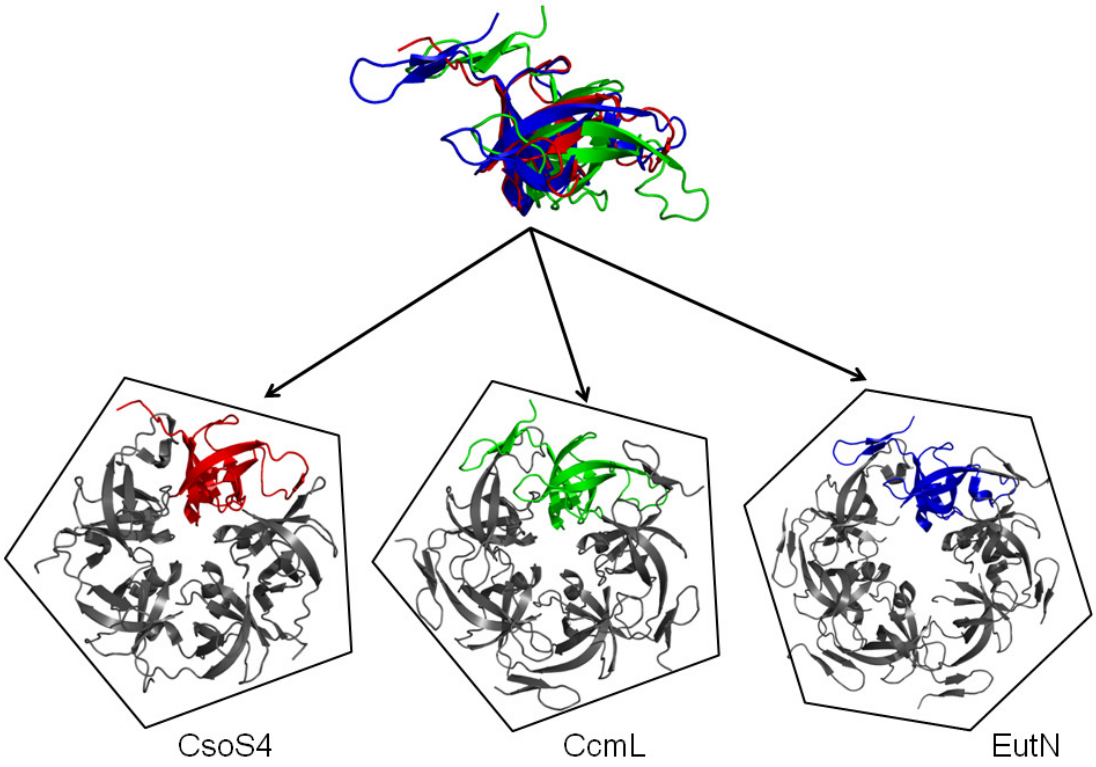


Figure 1.16 Shell proteins containing Pf03319 domain. At the top, superimposition of CsoS4 (PDB: 2RCF) (red), CcmL (PDB: 2QW7) (green) and EutN (PDB: 2Z9H) (blue) subunits. At the bottom, tertiary structures of CsoS4 pentamer (left), CcmL pentamer (middle) and EutN hexamer (right) (Tanaka et al., 2008).

1.2.5 Protein pores and molecular transport

In addition to providing structural integrity to the microcompartment, individual shell proteins are believed to play a crucial role in transporting substrates, products, cofactors as well as preventing labile and toxic metabolic intermediates from leaving the icosahedral shell (Kerfeld et al., 2005, Tsai et al., 2007, Crowley et al., 2010). Several shell proteins exhibit a central pore, as formed by the assembly of individual subunits of the BMC domain (or Pf00319 domains for some shell proteins). Notably,

the size of the central pore differs from one protein to another, as shown in Figure 1.17. Current structures suggest that hexameric single BMC domain containing shell proteins exhibit a fixed central pore while tandem BMC domain containing shell proteins have a more dynamic central pore that can be open or closed. The crystal structure of CsoS1D showed that two trimers with differing pore conformations are stacked together (Klein et al., 2009). This stacking was not evident in EutL. Instead, the two conformations of EutL have been determined separately (Sagermann et al., 2009). The closed form of EutL showed presence of three small pores within the subunit, which were not found in closed form of CsoS1D. This characteristic was also seen in the EtuB trimer (Heldt et al., 2009) and more recently, the PduB trimer. Some of the shell components, such as PduU, have occluded pores. The diversity of the pore size of shell proteins suggests that larger pores allow transport of larger substrates or cofactors while smaller (and more occluded) pores limit the escape of sequestered metabolic intermediates.

In addition to the pore size, the electrostatic potential of the shell (central region) provides further insight into which specific molecules of the bacterial microcompartment the shell proteins allow to transport. The shell proteins have two distinct faces. On one side, one can observe a major depression in the mid region (concave side) while the other side has lesser depression and is referred to as the convex side. These two sides, for most of the shell proteins, have differing electrostatic potential. For instance, the concave side of CsoS1A has a strong positive electrostatic potential while the convex side is slightly negative. It is suggested that the concave side of the shell protein faces into the microcompartment such that major constituents of the carboxysomes (ie RuBisCO and CA), which carry

negatively charge, would be able to interact with the positive charged side of the shell protein (Tsai et al., 2007). In addition, the concave side may able to ward off uncharged molecules (CO_2 and O_2) and therefore, aiding in increasing the concentration of CO_2 within the shell (Tanaka et al., 2009).

The crystal structure of PduT revealed a large central pore region, which is thought to be the site for 4Fe-4S cluster. One cysteine per subunit of the shell protein coordinates with the Fe atom of the iron sulfur cluster. It has been proposed that the shell protein acts as an electron transfer conduit for the metabolic enzyme, PduS, found within the microcompartment. The crystal structure of PduT homolog (PDB ID: 3NWG), showed an alternate pore conformation lacking a binding site for the metal cluster. A detailed structural analysis of PduT will be discussed in Chapter 3.

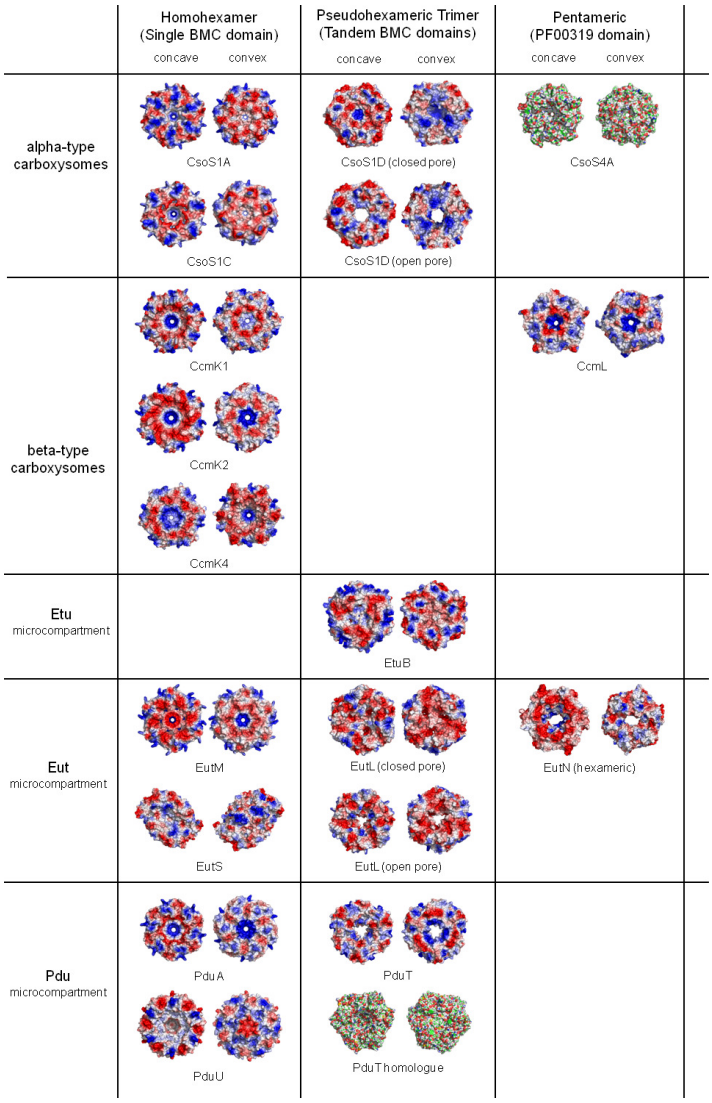


Figure 1.17 BMC electrostatic and pore variations. Diagrams illustrate the structural diversity of of BMC shell proteins from a variety of microcompartments. The two faces (concave and convex) for each shell proteins are represented in each panel. The shell proteins exhibit differences in size, shape and electrostatic properties (blue, white and red are positive, neutral and negative, respectively). Some interesting observations: CsoS1D and EutL exhibit an ability to open and close their pores. PduU consists of a beta barrel that blocks the central negatively charged pore region. The pore region of PduT is a binding site for 4Fe-4S. EutN has a distorted central pore. No central pore is evident in the bent EutS. The diverse characteristics showed are presumably responsible for the ability of shells to transport a variety of products, substrates, and cofactors, while at the same time limit the efflux of labile and toxic intermediates. Adapted from (Yeates et al., 2011).

1.3 Metabolic enzymes within Pdu

The Pdu metabolosome is so far the most complicated known microcompartment shell system. The 21 gene regulon of *Citrobacter freundii* encoding the pdu organelle and propanediol utilisation enzymes (Figure 1.18) has been cloned into *E. coli*, resulting in the production of microcompartments and allowing propanediol utilisation. Seven of these genes form the outer shell while the rest are enzymes responsible for aiding in the metabolic process. Within the Pdu microcompartment, 1,2-propanediol is converted by a diol dehydratase (PduC-D-E) into propionaldehyde. Sequestration of propionaldehyde within the microcompartment may prevent unwanted reactions leading to growth arrest and DNA damage (Sampson and Bobik, 2008). The propionaldehyde is subsequently disproportionated into 1-propanol and propionyl-CoA by the aldehyde dehydrogenase PduQ and CoA-acylating transferase PduP (Leal et al., 2003), respectively. These two products are delivered out to the cytoplasm, where propionyl-CoA is further converted into propionyl-phosphate and propionate by PduL and PduW (Palacios et al., 2003), respectively. Not only does the metabolosome contain the enzymes for 1,2-propanediol breakdown, it also contains reactivation factors for the diol dehydratase PduG-H as well as enzymes for the formation of the coenzyme form of cobalamin, PduO (Johnson et al., 2001) and PduS. The *Lactobacillus reuteri* Pdu metabolosome differs from that of *Salmonella* and *Citrobacter* by an association of the PduL enzyme with the metabolosome structure (Sriramulu et al., 2008), and by the absence of PduT shell protein (Pang et al., 2011, Crowley et al., 2010). This suggests the possibility that propionyl phosphate may be formed inside the *L. reuteri*

metabolosome. Onward respiration of propionate elsewhere in the cell does not occur in *L. reuteri* (Sriramulu et al., 2008), unlike in *Citrobacter* and *Salmonella* (Horswill and Escalante-Semerena, 1999, Price-Carter et al., 2001).

It had been reported that the N-terminal region of various Pdu enzymes play an important role organising the enzymes. For instance, PduD (Fan and Bobik, 2011) and PduP (Fan et al., 2010), were shown to require a short N-terminal region for an efficient packaging of the enzymes into the microcompartment; whereas, the N-terminal of PduV is crucial for targeting of enzyme to the outer surface of the microcompartment (Parsons et al., 2010a). However, the N-terminal extension of PduE seems to play little effect on the packaging, leading the investigators to think that the enzyme could be bound to another enzyme, such as PduD, to ensure efficient packaging of PduD within the microcompartment (Fan and Bobik, 2011).

GENE	FUNCTION	GENE	FUNCTION
pdu A	Shell protein	*pduM*	Unknown
pduB	Shell protein	*pduN*	Shell protein
pduC	Diol dehydratase large subunit	*pduO*	Cobalamin adenosyltransferase
pduD	Diol dehydratase medium subunit	*pduP*	CoA-dependent propionaldehyde dehydrogenase
pduE	Diol dehydratase small subunit	*pduQ*	Propanol dehydrogenase
pduF	Propanediol diffusion facilitator	*pduS*	Cobalamin reductase
pduG	Diol dehydratase reactivation protein	*pduT*	Shell protein
pduH	Diol dehydratase reactivation protein	*pduU*	Shell protein
pduJ	Shell protein	*pduV*	Unknown
pduK	Shell protein	*pduW*	Propionate kinase
pduL	Phosphotransacylase	*pduX*	Unknown

Figure 1.18 A list of genes involved in the propanediol utilisation metabolosome, including both enzymes and shell proteins (highlighted).

1.4 Aims and objectives

The main aim of this project is to deepen the understanding of the proteinaceous shell of the bacterial microcompartment using *Citrobacter* and *Lactobacillus* Pdu microcompartment proteins. The proteins were expressed recombinantly in *E. coli*, purified and used for various structural studies, with X-ray crystallography as the principal method to obtain the three-dimensional structure. Broadly, the objectives of this research are to identify the "form" and "function" of shell proteins. In more specific terms, this Thesis aims:

1. To determine the crystal structures of shell proteins with tandem BMC repeats,
2. To explore how these shell proteins fit into the icosahedral proteinaceous shell,
3. To elucidate the function of the shell proteins and how they contribute to the processing of the metabolic reaction within the microcompartment,
4. To better understand how the shell proteins assemble into a uniform molecular layer.

Chapter 2

Materials and Methods

2.1 Introduction

This chapter presents the general overview of *Materials and Methods* used in the project. The general molecular biology, basic biochemical procedures and laboratory techniques were carried out as described in Sambrook (Sambrook et al., 1989), unless otherwise stated. Changes to the protocols specific to individual proteins are discussed in the respective relevant chapters. Figure 2.1 shows an overview of the experimental procedures that were used. The methods are set out in the order: molecular biology procedures, protein production, and purification. Purified proteins were subsequently used for various experiments, which included crystallisation.

2.2 Constructs

Constructs for the shell proteins and enzymes of *Citrobacter freundii* and *Lactobacillus reuteri pdu* metabolosome were kindly provided by collaborators in University of Kent (Professor Martin Warren) and University College Cork (Professor Michael Prentice), respectively. They cloned the DNA sequences into

plasmid vectors, as described in the Table 2.1. Proteins produced from pET14b plasmid possess an N-terminal hexahistidine tag to allow easy purification and a thrombin as a cleavage site for the tag, while proteins produced from pET23b plasmid consist of a non-cleavable C-terminal hexahistidine tag. Proteins produced from pET3a plasmid have no tag.

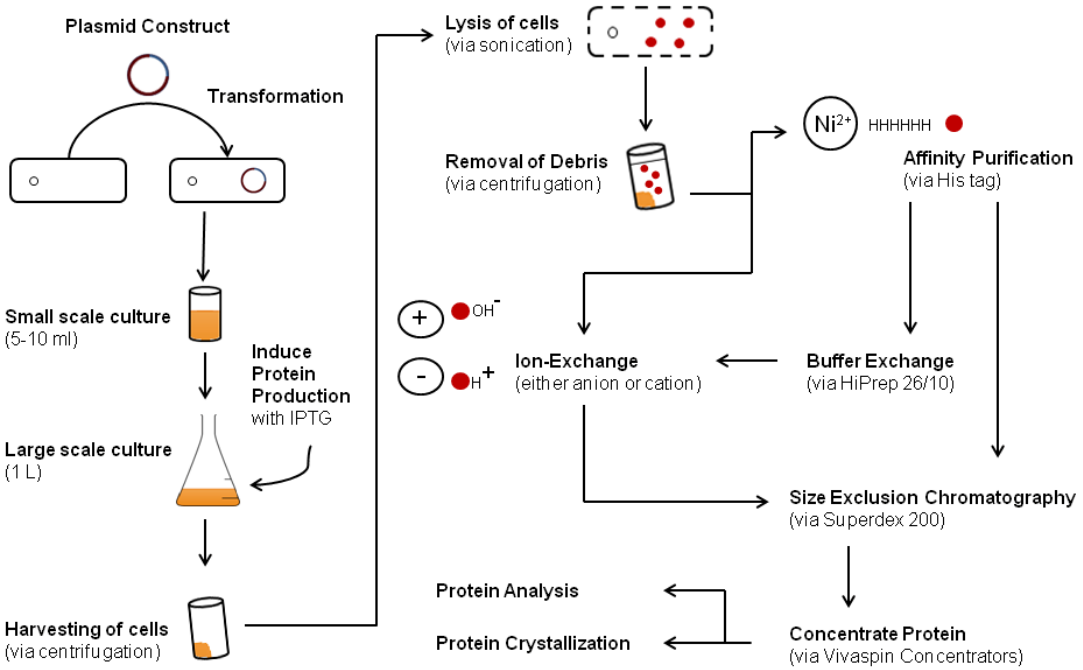


Figure 2.1 Overview of experimental protocols used in this work. The project consists of a series of experiments each of which starts with transformation of a recombinant plasmid into BL21 cells. Cells were grown and protein production was induced by the addition of IPTG. Cells were then harvested by centrifugation and sonicated to release the cytoplasmic contents. The protein of interest can then be captured and purified via nickel affinity purification (for His-tagged proteins), ion exchange and size exclusion chromatography. Multiple purification steps may be required; buffer exchange is sometimes needed to change the buffer, for instance before ion exchange chromatography. Purified concentrated protein is then used for crystallisation trials and various analyses.

Table 2.1 Constructs used in this research.

Proteins	Plasmids	Comments
PduA	pET14b-PduA	PCR product of *Citrobacter freundii pduA* ligated into the *Nde*I and *Spe*I sites of pET14b.
		PduA has a mutated end (S→R) and extra 23 amino acid residues on its C-terminal end
PduA	pET3a-PduA	PCR product of *Citrobacter freundii pduA* ligated into the *Nde*I and *Spe*I sites of pET3a.
PduB	pET14b-PduB	PCR product of *Lactobacillus reuteri pduB* ligated into the *Nde*I and *Spe*I sites of pET14b.
PduJ	pET14b-PduJ	PCR product of *Citrobacter freundii pduJ* ligated into the *Nde*I and *Spe*I sites of pET14b.
PduJ	pET14b-PduJ	PCR product of *Lactobacillus reuteri pduJ* ligated into the *Nde*I and *Spe*I sites of pET14b.
PduK	pET14b-PduK	PCR product of *Citrobacter freundii pduK* ligated into the *Nde*I and *Spe*I sites of pET14b.
PduN	pET14b-PduN	PCR product of *Citrobacter freundii pduN* ligated into the *Nde*I and *Spe*I sites of pET14b.
PduN	pET23b-PduN	PCR product of *Lactobacillus reuteri pduN* ligated into the *Nde*I and *Xho*I sites of pET23b.
PduN	pET14b-PduN	PCR product of *Lactobacillus reuteri pduN* ligated into the *Nde*I and *Spe*I sites of pET14b.
PduP	pET14b-PduP	PCR product of *Citrobacter freundii pduP* ligated into the *Nde*I and *Spe*I sites of pET14b.
PduQ	pET14b-PduQ	PCR product of *Citrobacter freundii pduQ* ligated into the *Nde*I and *Spe*I sites of pET14b.
PduT	pET14b-PduT	PCR product of *Citrobacter freundii pduT* ligated into the *Nde*I and *Spe*I sites of pET14b.
PduT pLysS	pLysS-PduT	PCR product of *Citrobacter freundii pduT* ligated pLysS vector
PduS	pET14b-PduS	PCR product of *Citrobacter freundii pduS* ligated into the *Nde*I and *Spe*I sites of pET14b.
PduS	pET3a-PduS	PCR product of *Citrobacter freundii pduS* ligated into the *Nde*I and *Spe*I sites of pET3a.

2.2.1 Mutagenic Primers

A list of the primers used in the mutagenesis of PduA construct can be found in Table 2.2. All primers were obtained from Eurofins MWG. The nucleotide bases underlined highlights the mutation which is given in the last column of the Table 2.2.

Table 2.2 Primers used to introduce mutation on PduA construct

Primer Name	Sequence	Mutation
PduA*_K26A_F	5' GCCGCAGATGCAATGGTGGCGTCAGCCAATGTAATGCT 3'	AAG → GCG
PduA*_K26A_R	3' AGCATTACATTGGCTGACGCCACCATTGCATCTGCGGC 5'	CTT → CGC
PduA*_K26D_F	5' GCCGCAGATGCAATGGTGGATTCAGCCAATGTAATGCTG 3'	AAG → GAT
PduA*_K26D_R	3' CAGCATTACATTGGCTGAATCCACCATTGCATCTGCGGC 5'	CTT → ATC
PduA*_V51A_F	5' GTCCGCGGCGATGCTGGCGCAGTCAAA 3'	GTT → GCT
PduA*_V51A_R	3' TTTGACTGCGCCAGCATCGCCGCGGAC 5'	ACA → GCA
PduA*_V51D_F	5' GTCCGCGGCGATGATGGCGCAGTCAAA 3'	GTT → GAT
PduA*_V51D_R	3' TTTGACTGCGCCATCATCGCCGCGGAC 5'	ACA → TCA
PduA*_R79A_F	5' CCGTACACGTCATCCCAGCCCCTCACACCG 3'	ACG → AGC
PduA*_R79A_R	3' CGGTGTGAGGGGCTGGGATGACGTGTACGG 5'	CGT → GCT

2.2.2 Generation of Mutagenic Constructs

The mutations generated in this study, PduA (K26A, K26D, V51A, V51D and R79A), were made using the QuikChange® II Site-Directed Mutagenesis Kit from Stratagene. The QuikChange® II site-directed mutagenesis kit can be used to introduce point mutations that switch amino acids as well as delete and insert single or multiple base pairs. The first step of mutagenesis was the synthesis of mutant strand via thermal cycling procedure. The thermal cycling reaction includes a supercoiled double-stranded DNA (dsDNA) vector with an insert of interest, two synthetic oligonucleotide primers both consisting of desired mutations, and PfuUltra

high-fidelity (HF) DNA polymerase. The oligonucleotide primers, each complementary to opposite strands of the vector, were extended during temperature cycling by PfuUltra HF DNA polymerase. Extensions of the oligonucleotide primers generated a mutated plasmid with staggered nicks. Details of the reaction mixture and thermal cycling procedure were noted in Appendix A1.

In the second step of the mutagenesis, the thermal cycling reaction product was purified using peqGOLD MicroSpin Cycle Pure Kit (PEQLAB Ltd.) and eluted out in a NEBuffer II optimised for restriction endonuclease *Dpn*I (target sequence: 5‟-GmATC-3‟). This enzyme (New England Biolabs) specifically digests parental DNA template and selects for newly synthesised mutagenic DNA. Following the *Dpn*I treatment, the reaction mixture was transformed into DH5-alpha competent cells (New England Biolabs) as detailed in Chapter 2.3.2, where the nicks were repaired. The DNA sequences of clones were checked (Eurofins MWG) for successful mutagenesis.

2.3 Molecular biology methods

2.3.1 Preparation of competent cells

Competent *E. coli* strains used in this study were initially brought (suppliers see Table 2.3). The house-keeping strain DH5α and XL1-blue were used for cloning and plasmid propagation, while the strain BL21 (DE3) and BL21 (DE3) pLysS were involved in recombinant overexpression of the metabolosome proteins.

Competent cells can be produce in-house using starter material from the company. 500 µL of overnight culture of competent cells was inoculated into a 200 mL of

sterile LB media without antibiotics. The culture media was incubated and shaken at 37˚C. At 0.6-absorbance reading, the *E. coli* cells were harvested by centrifugation at 13000 rpm, 4˚C for 10 minutes. 0.1 mM $MgCl_2$ was mixed to the pellet and left for 30 minutes at 4˚C. The mixture was centrifuged and the supernatant was discarded. 0.1 mM $CaCl_2$ was added into the pellet and gently resuspended. The mixture was left at 4˚C for 1 hour, then centrifuged for 10 minutes at 13000 rpm, 4˚C. The supernatant was discarded and pellet was resuspended with 0.1 mM of $CaCl_2$ with 50% glycerol. The cell suspension was flash frozen with liquid nitrogen and stored at -80°C.

Table 2.3 The commercial *E. coli* strains used, their antibiotic resistance and suppliers.

Strain and Genotype	Function	Antibiotic Resistance	Supplier
DH5α T1 F- *end*A1 *gln*V44 thi-1 *rel*A1 *gyr*A96 *deo*R *nup*G l*ac*ZdeltaM15 *hsd*R17	house-keeping	Ampicillin	New England Biolabs
XL1-Blue endA1 gyrA96(nal^R) thi-1 recA1 relA1 lac glnV44 F'[::Tn10 proAB$^+$ lacI^q Δ(lacZ)M15] hsdR17(r_K^- m_K^+)	house-keeping	Ampicillin	Stratagene
BL21(DE3)pLysS F– *ompT hsdS*(rB– mB–) *gal dcm* λ(DE3) pLysS (Camr)(λ(DE3): *lac*I, *lac*UV5-T7 gene 1, *ind*1, *sam*7, *nin*5)	protein expression	Ampicillin / Chloramphenicol	Novagen
BL21(DE3) F– *ompT hsdS*(rB– mB–) *gal dcm* λ(DE3): *lac*I, *lac*UV5-T7 gene 1, *ind*1, *sam*7, *nin*5)	protein expression	Ampicillin	Novagen

2.3.2 Transformation of plasmid DNA into host cells

A 100 μL aliquot of competent cells was thawed on ice for 10 minutes. 1 μL of the plasmid construct was incubated with competent cells for 1 hour on ice. The cells were transferred into a 42˚C water bath for 45 seconds and subsequently placed immediately on ice for a 2-minute recovery. To this, 200 μL LB media was added and was shaken for about an hour before spreading out in a LB agar plate with

appropriate antibiotics. Growth of successfully transformed *E. coli* can be seen after incubation at 37˚C overnight. Single colonies were picked with a sterile pipette tip for inoculation of starter cultures.

2.3.3 Bacterial storage

E.coli cells were stored for long term as glycerol stocks. These cells were mixed with glycerol at a final concentration of 15% (v/v). Bacterial glycerol stock was kept at -80ºC.

2.3.4 Broth and media

A nutritionally rich medium, lysogeny broth and its derivative were used for bacterial growth. The media compositions are listed in Table 2.4.

Table 2.4 Composition of LB media used for bacterial growth.

Media	Composition	Manufacturer	Comments
LB media	1% (w/v) tryptone 0.5% (w/v) yeast extract 0.5% (w/v) sodium chloride	Fisher Scientific Oxoid Fisher Scientific	Sterilised by autoclave
LB-G media	LB media + 1% (w/v) glucose	Sigma Life Sciences	Sterilised by autoclave; used for leaky expression
LB media agar	LB media + 1.5% (w/v) agar	Merck	Sterilised by autoclave; once cooled, poured to Petri dish and let it settle

2.4 Protein expression and purification

2.4.1 Antibiotic and IPTG stock

Stock concentrations of 100 mg/mL of ampicillin, 34 mg/mL chloramphenicol and 1M IPTG were prepared and stored at -20°C. Both ampicillin and IPTG were diluted with deionised water to its desired concentration while chloramphenicol was diluted with ethanol. Stock solutions were filter-sterilised through 0.2 μm pore size membrane.

2.4.2 Protein production

Starter cultures were prepared by inoculating bacterial glycerol stock to 5-10 mL LB media containing ampicillin (final concentration – 100 μg/ml). Cells were allowed to grow at 37˚C overnight. 1 L LB media supplemented with ampicillin was

inoculated with the starter culture and shaken at 37˚C until an OD_{600} of 0.6 - 0.8 is reached. IPTG (final concentration - 400 μM) was added to induce protein production. The bacterial culture was left shaking overnight at 16˚C.

2.4.3 Cell density measurement

The optical density at a wavelength of 600 nm (OD_{600}) can be used to determine the number of bacterial cells in the suspension. The Beer-Lambert Law describes the relationship between absorbed light and concentration for dilute solutions (Koch, 1970, Lawrence and Maier, 1977):

$A = \log_{10} I_0/I = \varepsilon cl$

A = absorption
I = intensity of light after passing through the sample
I_0 = intensity of light before passing through the sample
ε = extinction coefficient
c = concentration
l = path length

To measure the cell density, a Hitachi U-3010 UV-Visible spectrophotometer was used. The wavelength of the spectrophotometer was set at 600. A fresh LB media placed on a cuvette was used as a baseline. The bacterial suspension was then placed on cuvette to get the reading of the cell density.

2.4.4 Cell lysis

The bacterial cultures were harvested via centrifugation (Avanti J-30I, Beckman Coulter) with a speed of 8,000 rpm for 20 minutes. Collected cell pellets can be stored at -80ºC.

Collected cell pellets were resuspended with lysis buffer then lysed using sonicator with 20 watts for 20 seconds with a resting time of 1 minute in between. This was performed for 6-8 times. Following lysis, the cell debris was sedimented by centrifugation (Avanti J-30I, Beckman Coulter) at 18,000 rpm for 30 mins. The supernatant, which contains the protein of interest, was collected for purification. It is important that cell lysis procedure is kept on ice to prevent overheating as a result of sonication of bacterial cells.

2.4.5 Immobilised metal affinity chromatography

In this study, a histidine tag attached to the proteins was used for purification with immobilised metal affinity chromatography (IMAC). The aromatic ring of the histidine binds to the Ni^{+2}-ion that is immobilised with a linker to the column matrix, such as iminodiacetic acid. An application of a competitive molecule, imidazole, is applied to displace the His-tag from the nickel column and therefore, eluting the protein of interest.

An empty PD-10 column was filled with Chelating Sepharose Fast Flow resins. These were washed with deionised water to remove ethanol and subsequently charged with 100 mM nickel sulfate. Charged resins were acclimatised to the binding buffer prior to use.

After loading the supernatant onto the column containing charged resins, an increasing concentration of imidazole was applied to remove unspecific binding. Upon application of the highest imidazole concentration, the eluted samples were collected. Buffers and materials used are listed in Table 2.5.

Table 2.5 Buffers and materials used for IMAC purification.

Buffers and Materials	Composition	Comments
PD-10 column	-	from GE Healthcare
Chelating Sepharose Fast Flow	highly cross-linked 6% agarose	from GE Healthcare
Stripping Buffer	20 mM Tris-HCl (pH 8.0) 0.5 M NaCl 100 mM EDTA	used to strip off nickel to reuse resins
Charging Buffer	100 mM $NiSO_4$	used to charged column with nickel ions
General buffer used		
Lysis Buffer / Binding Buffer / Buffer A	50 mM Tris-HCl (pH 8.5) 0.5 M NaCl 10 mM imidazole	initial binding buffer
Buffer B	50 mM Tris-HCl (pH 8.5) 0.5 M NaCl 50 mM imidazole	low imidazole for washing
Buffer C	50 mM Tris-HCl (pH 8.5) 0.5 M NaCl 100 mM imidazole	high imidazole for washing
Elution Buffer	50 mM Tris-HCl (pH 8.5) 0.5 M NaCl 400 mM imidazole	used for elution of protein

2.4.6 Buffer exchange

Proteins eluted out from the IMAC consist of a high NaCl and imidazole concentration; thus, it is necessary to remove this to proceed to the next step of purification. Buffer exchange can usually be done through dialysis; however, this is not ideal as it is time consuming and susceptible to protein aggregation. As a result, it was decided that buffer exchange is carried out via a desalting column (HiPrep

26/10). The column exploits the difference in size of proteins and biomolecules. High molecular weight substances such as proteins elute first while smaller size substances like NaCl and imidazole elute out later.

Protein samples were loaded onto the desalting column pre-equilibrated with the desired buffer. Using AKTA FPLC system, buffer ran through the system, and eluted samples were collected. Peak fractions detected (at 280 nm) were pooled together for succeeding purification steps or experiments.

2.4.7 Ion exchange chromatography

Ion exchange chromatography relies on charge-charge interactions between the proteins in the sample and the charges immobilised on the column resin. There are two types of ion exchange, one is cation exchange chromatography, in which positively charged ions (pH is lower than the isoelectric point of the protein) bind to negatively charged resins; and anion exchange chromatography, in which the binding ions are negative (pH is higher than the isoelectric point of the protein) and the immobilised functional group is positive.

Protein samples were loaded on to the appropriate ion-exchange column equilibrated with a low NaCl buffer. Using AKTA FPLC system, a gradient of increasing NaCl concentration is applied to elute out proteins with a differing ionic strength to the column. The fractions containing the protein of interest were pooled together and concentrated for further purification step. Buffers and materials used are listed in Table 2.6.

Table 2.6 Buffers and materials used for ion-exchange chromatography purification.

Buffers and Materials	Description
HiTrap Q Sepharose Fast Flow	Column used for anion exchange; obtained from GE Healthcare
50 mM Tris-HCl pH 8.0	Low salt buffer
50 mM Tris-HCl pH 8.0 0.5 M NaCl	High salt buffer
HiTrap SP Sepharose Fast Flow	Column used for cation exchange; obtained from GE Healthcare
50 mM MES pH 6.0	Low salt buffer
50 mM MES pH 6.0 0.5 M NaCl	High salt buffer

2.4.8 Gel filtration chromatography

Gel filtration chromatography was used as the final protein purification step. Before application to the gel filtration column, the protein solution was concentrated in a Vivaspin concentrator (Sartorius Stedim Biotech, Aubagne, France) until a final volume of 500 μL was reached. The concentrated sample was loaded onto a loop connected to a Superdex 200 10/300 gel filtration column (GE Healthcare, Little Chalfont, UK), which was pre-equilibrated with desired final buffer. An automated AKTA FPLC was used to run the chromatography at a rate of 0.5 ml min^{-1} with 1 column volume of buffer. The target was collected in 1 mL fraction and analysed by SDS-PAGE. Fraction samples containing the protein of interest were pooled and concentrated for use in crystallisation trials and/or succeeding experiments.

Gel filtration chromatography differentiates proteins based on their size and shape. The column consists of spherical and porous Sephadex beads. These beads are made up of insoluble but highly hydrated polymer (hydroxypropylated cross-linked dextran). Small molecules can enter the beads whereas the larger ones are excluded from the pores. As a consequence, large molecules flow through the interstitial spaces faster and will exit out first. Smaller molecules which occasionally enter the bead will emerge next while the tiny molecules, due to its ability to enter the aqueous spaces within the beads, take a longer path and will flow out of the column last. Molecules of the same size can be separated on the basis of their size. Molecules with a more compact shape like a sphere will enter the beads more easily compared to elongated shape such as rod. Hence, a spherical molecule will elute later compared to an elongated shape molecule.

2.4.9 Protein analysis

2.4.9.1 Sodium dodecyl sulfate polyacrylamide gel electrophoresis

The purity of a protein sample was analysed using the sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) based on a method developed by Laemmli (Laemmli, 1970). SDS is an anionic detergent that denatures proteins by disrupting its non-covalent bonds, causing the molecule to lose their native conformation. Due to the nature of SDS, denatured polypeptides became negatively charge. When proteins are denatured, it is necessary to add dithiothreitol as denatured proteins can form disulfide bonds before they adopt the random-coil configuration essential for size separation. By using this system, protein migration via gel electrophoresis was determined by molecular weight alone.

The Mini-Protean 3 system from Bio-Rad is used for setting up and running polyacrylamide gel. Glass plates were cleaned with ethanol and allowed to dry before assembling into casting frame and casting stand. Resolving gel and stacking gel were made according to the Table 2.7 below:

Table 2.7 Composition of single polyacrylamide gel.

Chemicals	Resolving Gel			Stacking Gel
	12%	15%	18%	5%
1.5M Tris-HCl pH 8.8	1.3 ml	1.3 ml	1.3 ml	-
0.5M Tris-HCl pH 6.8	-	-	-	0.25 ml
30% Acrylamide	2 ml	2.5 ml	2.92 ml	0.33 ml
10% ammonium persulfate	50 µl	50 µl	50 µl	20 µl
10% SDS	50 µl	50 µl	50 µl	20 µl
TEMED	2 µl	2 µl	2 µl	2 µl
dH_2O	1.6 ml	1.1 ml	0.73 ml	1.4 ml

TEMED was only added immediately before the gel mixture is poured onto the gel cassette sandwich. Resolving gel was first poured and ethanol was added on top of it. The gel was allowed to polymerise and ethanol was removed prior to pouring of stacking gel. The comb was inserted between the spacers and left at room temperature to polymerise for 15 minutes.

Protein samples were supplemented with loading buffer and incubated at 100ºC for 5 mins prior to loading. Samples were loaded alongside Prestained Protein Marker, Broad Range (New England BioLabs) to estimate the molecular weight of protein bands. The electrophoresis was run under constant voltage (200 V) until protein bands migrate to the bottom. Gels were viewed after staining and destaining. Table 2.8 lists the buffers and solutions used.

Table 2.8 Buffers and solutions used for SDS-PAGE electrophoresis.

Buffers and Solutions	Composition	Description
10X Running Buffer	30.3 g Tris base 144 g Glycine 10 g SDS Make up to 1 L with dH_2O	Diluted to 1X prior to use on the electrophoresis tank
5X Sample Loading Buffer	10% w/v SDS 10 mM DTT 20% v/v glycerol 0.2 M Tris-HCl pH 6.8 0.05% w/v bromophenol blue	Stored at -20ºC; Diluted to 1X with sample
Staining solution	0.1% Coomassie R250 40% methanol 10% acetic acid	Gel placed in this solution for staining for 1 hour
Destaining solution	40% methanol 10% acetic acid	Staining solution replaced by this solution to remove excess Coomassie stain

2.4.9.2 Protein concentration measurement

An estimated protein concentration was determined using the Beer-Lambert law (as stated in Chapter 2.4.2) by measuring the absorption of the protein solution at a wavelength of 280 nm. This absorbance peak is largely due to the amino acids with aromatic rings. The Hitachi U-3010 UV-Visible spectrophotometer was first calibrated with buffer. A test scan was done ensuring that the buffer gave a zero reading. The protein solution, usually at a dilution factor of 50, was measured. As indicated in the Beer-Lambert law, the extinction coefficient must be determined to calculate the protein concentration. This can be obtained by placing the amino acid sequence into PROTPARAM tool of the EXPASY proteomics server

(http://www.expasy.ch), which calculates based on the tyrosine and tryptophan residues (Edelhoch, 1967, Pace et al., 1995).

As proteins without tryptophan, tyrosine and cysteine cannot be detected at wavelength 280 nm, Bio-Rad Protein assay was employed to ascertain the concentration of the protein. This assay is based on the Bradford dye-binding procedure (Bradford, 1976) which relies on the binding of Coomassie Brilliant Blue G-250 dye to basic amino acids such as arginine. A standard curve was plotted using BSA with a known concentration. Protein of interest was diluted, treated with the dye and measure its absorbance at 595 nm. The measurement was located at the standard curve to find the protein concentration.

2.4.10 Anaerobic purification

For purification and crystallisation of proteins in anaerobic condition, an anaerobic glove box (Belle Technology) was used. The glove box maintains a nitrogen environment with less than 3 ppm oxygen. Before being introduced into the glove box, the buffer and sample solutions were degassed by purging with argon. PD-10 columns, DEAE columns and other chemicals were left in the glove box overnight to minimise the amount of oxygen in the reaction mixtures. Procedures for cell harvest and lysis were similar as discussed earlier. Cell lysate was immediately placed in anaerobic glove box and purified as stated earlier inside the glove box.

2.5 Protein crystallisation

The process of crystallisation relies heavily on protein samples. Typically, proteins need to be in high concentration, as pure as possible and homogenous. Protein

crystals form when the sample is placed in the right condition. To achieve the crystallising condition, the solubility of the purified protein solution is gradually lowered through slow precipitation using a precipitant. If this process is done too quickly, proteins would aggregate and form amorphous precipitates. Under certain conditions, individual protein molecules align themselves in a repeating series of unit cells and thus forming crystals. The crystallisation condition relies on the concentration of the protein and the precipitant, as described in Figure 2.2. Figure 2.3 shows an overview of the procedure from crystallisation to solving structure.

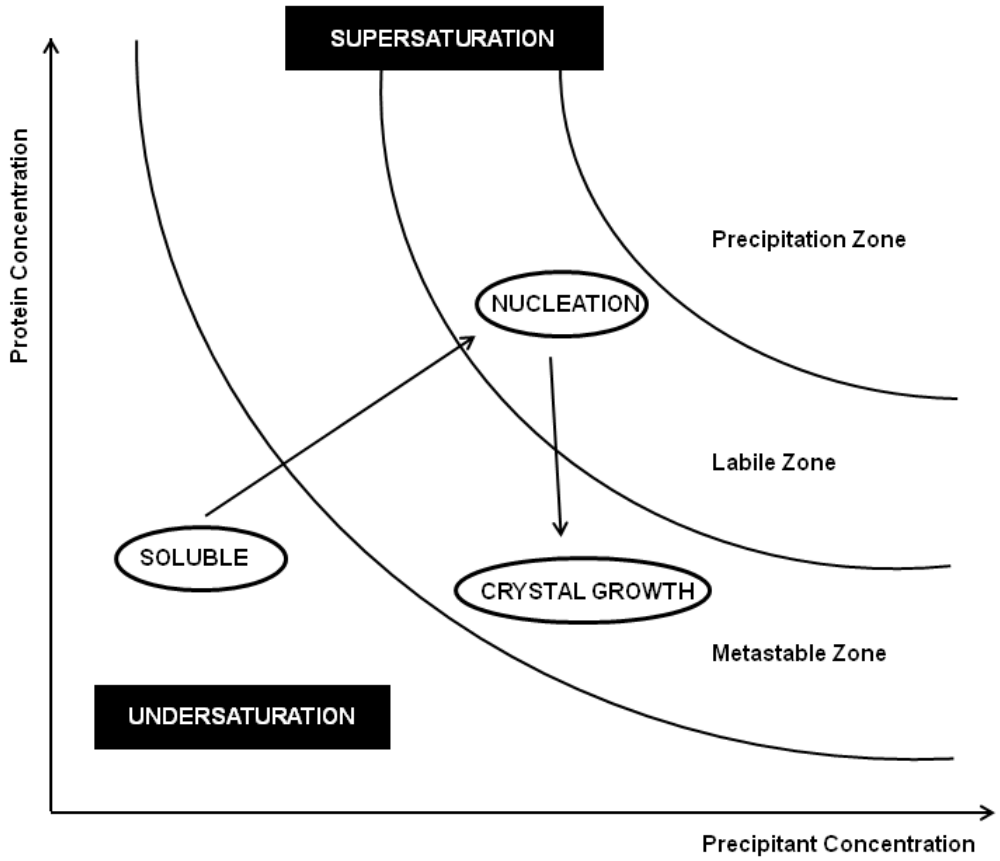


Figure 2.2 Crystallisation phase diagram. As precipitant concentration increases, soluble protein enters the labile zone where nucleation occurs. Once nucleation happens, the concentration of protein lowers and the system enters the metastable zone, on which protein crystal forms. Adapted from (Ducruix and Giege, 1999).

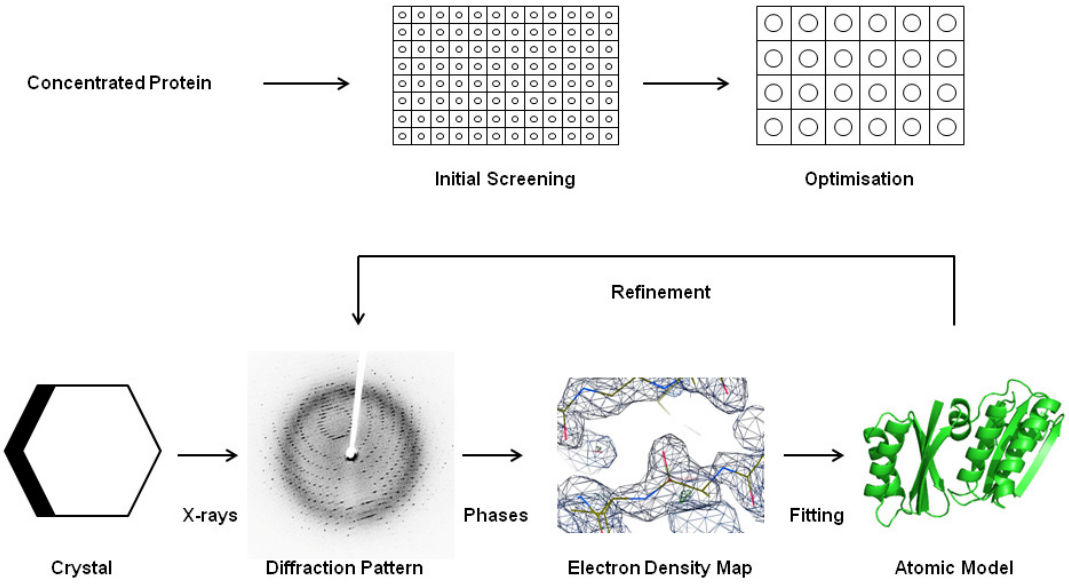


Figure 2.3 Overview of a successful structural biology project. Protein crystals are obtained from a screen. The crystal diffracts and the data are collected. The electron density map is calculated once the phase is determined. A model is produced by fitting the sequence into the electron density map. The model is refined until a satisfactory model is achieved (as judged by its stereochemistry and validation methods).

2.5.1 Initial Screening

Protein samples were used in sparse matrix screens Crystal Screen 1 & 2 (Hampton Research) and Structure Screen I & II HT-96 (Molecular Dimensions). Initial screening was at 18ºC using the hanging drop method and/or sitting drop vapour diffusion method. 80 µL crystal screen reservoirs were dispensed onto the 96 well-plate. The drops of crystallisation experiments were set up automatically with Mosquito (TTP Lab Tech Inc). For the sitting drop method, the tray was sealed with a clear sealing tape. For the hanging drop method, the drops were dispensed onto the

plate seal which was then inverted on top of the crystal plate. The trays were left at 18 ºC and checked regularly for crystal growth.

2.5.2 Optimisation

Conditions, which yielded crystals in the initial screening, were optimised in attempt to get more and bigger crystals. The optimisation procedure involves slight alteration of the reservoir (buffer pH and concentration, precipitant concentration and salt concentration). The concentration of the protein and its ratio of protein and reservoir mix were also varied. For optimisation, a 24-well crystal tray was used. An overview of this technique is provided in Figure 2.4. The rims of the well were sealed with Dow Corning high vacuum silicon grease. The protein sample was mixed with the crystal screen reservoir in the 22x22 mm coverslip (Menzel Glaser). The coverslip was inverted and placed on top of the well. The trays were kept at 18ºC.

2.5.3 Heavy metal soaks

A series of heavy metal salts were dissolved in the screening reservoir condition with appropriate cryoprotectants. Native protein crystals were looped out of the drop and transferred to the heavy metal solutions (augmented with cryoprotectants). The soak time varied and therefore was recorded. The soaked crystals were looped out and frozen immediately using liquid nitrogen.

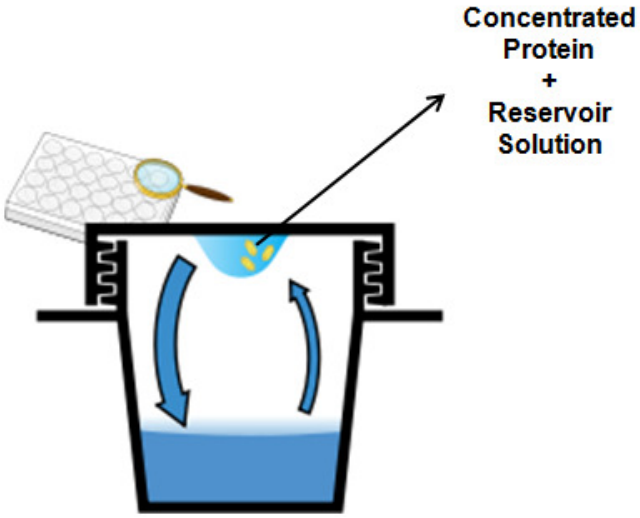


Figure 2.4 An overview of the hanging drop method. 1 µl of concentrated protein solutions mixed with 1 µl of reservoir solution in a drop on a coverslip, inverted over the reservoir solution. The method utilises a process called vapour diffusion. As the vapour leaves the drop to reach equilibrium with its surrounding, the protein and precipitant concentrations in the drop slowly increase to reach crystal growth.

2.6 X-ray analysis

Crystals are needed to be placed in a cryoprotectant solution, to ensure no disruptive ice crystals are formed within the protein crystal. Most commonly used cryoprotectants are glycerol, ethylene glycol and polyethylene glycol (PEG) 400. The recommended amount of each cryoprotectant varies accordingly to the reservoir solution (McFerrin and Snell, 2002). The loop with protein crystal is loaded onto a goniometer head. X-rays from an X-ray source are directed on to the crystal and the diffracted rays collected on a detector (Figure 2.5) (Rhodes, 2006).

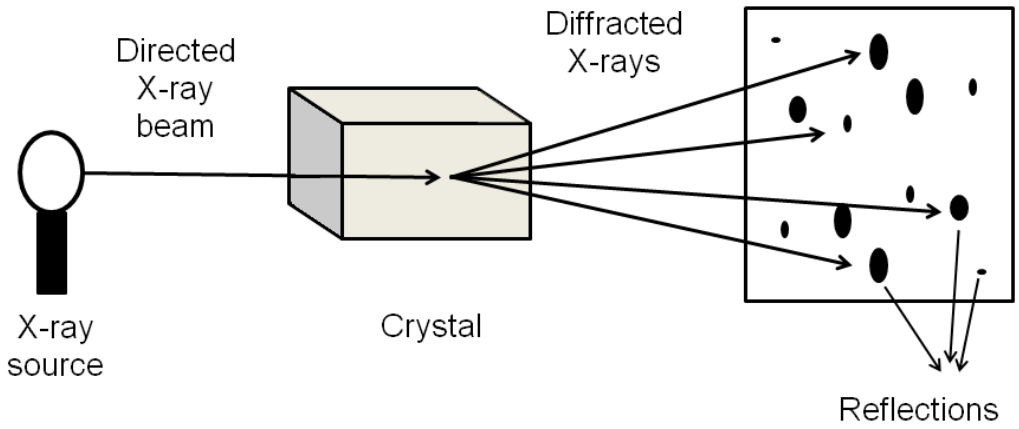


Figure 2.5 Crystallographic data collection. The X-ray beam that hits a protein crystal is diffracted into several discrete beams, each of which produces a distinct spot (reflection). The position of the reflections is due to the packing of the molecules in the crystal. The intensities of these reflections contain the information required to determine the molecular structure.

2.6.1 Data collection

Crystals were initially screened to check for diffraction quality. Two images were typically taken at 90° separation. These images were automatically indexed by xia2 (Winter, 2009, Vagin and Teplyakov, 1997) to predict the point group of the crystal. MOSFLM (Leslie, 1992) is another program that can also be used to do this; auto-indexing gives a list of possible solutions with corresponding penalty. A jump from a low to a distinctly higher penalty as the point group symmetry increases is usually observed; the point group before this "jump" is usually correct, having the highest symmetry with acceptable penalty. These initial images were also used to run the Testgen function in MOSFLM (Leslie, 1992) or automatically analysed by EDNA (Incardona et al., 2009), which utilised the cell information and symmetry of the crystal to calculate a strategy that would give an optimal data collection. The starting angle, rotation angle and number of images to collect were used as suggested.

2.6.2 Data processing

After data collection, xia2 (Winter, 2009, Vagin and Teplyakov, 1997) generates reduced data in a form suitable for phasing and structure solution. MOSFLM (Leslie, 1992) can also be used to do this. In the indexing stage, the program estimates the mosaicity of the crystal, which relates to the number of reflections seen on an individual image, the higher the mosaicity the more reflections observed. Next, the cell was refined using several collected images based on the predicted point group, cell perameters and mis-setting angles. At this point, the chosen space group is a hypothesis and ultimately, successful refinement of the structure would confirm if the space group choice is correct. A program called Pointless (Evans, 2006) is

frequently used to suggest the space group of the crystal based on the relationship of the intensities of the reflections and the presence (or rather absence pattern) of axial reflections to identify systematic absences and corresponding screw-axes.

The reduced data file (mtz output file from MOSFLM or xia2) was processed using SCALA (Evans, 2006), which scales and merges multiple observations of reflections, providing an average intensity. Typically 5% of the experimental data were flagged to be excluded from refinement and used for the calculation of R-free, which is used as an unbiased estimate of the improvement of the model during refinement. The Wilson B-factor of the dataset was also determined at this stage. This parameter, provides an indication of the order of the crystal. The Matthews coefficient (Vm) was also calculated to predict how many protein molecules were present in the asymmetric unit.

2.6.3 Single wavelength anomalous dispersion

For heavy metal derivative crystals, an X-ray fluorescence scan was run. An absorption peak in the X-ray fluorescence scan indicates presence of heavy metal in the protein crystal, but only if the crystal is back-soaked does it indicate that the metal has bound to the protein. After the X-ray fluorescence scan, the values *f'* and *f''* values were processed by CHOOCH (Evans and Pettifer, 2001). These values were used to determine wavelength at which the diffraction data were collected. The diffraction data were automatically processed by xia2 (Winter, 2009, Vagin and Teplyakov, 1997), and the anomalous slope was assessed. The AutoSol (PHENIX) software was used to solve the structures and generate experimental phases with the

SAD method. The program enables identification of heavy-atom sites, phase calculation, density modification and non-crystallographic symmetry (NCS).

2.6.4 Molecular replacement

Molecular replacement often allows rapid solution of the phase problem when the search and target molecules have at least 25% sequence identity. The program used here was Molrep (Vagin and Teplyakov, 1997). Molecular replacement is made up of two steps, first the rotation function and then the translation function. While the rotation function allows for the correct orientation of the search molecule with maximal overlap with the target structure to be determined in the absences of any phases for the unknown structure, the translation function shifts the position of the search molecule correctly relative to the target structure. Therefore, the success of molecular replacement is based on these two functions and the correct solution has the highest correlation value, preferably distinguishable from the first incorrect "solution" by a significant fall in the correlation value of the subsequent "solutions" found.

The agreement between the target model and the solution can be assessed by correlation coefficient. Correlation coefficient (CC) compares P and Q to their mean values within the volume V, and then compares the product of these to the maximum possible value as represented in the equation below, where $\overline{P}$ is the mean value of P, given by (1/V) and similarly for $\overline{Q}$. The denominator scales the result to give values in the range of +1 for prefect correlation, 0 for no correlation and -1 for perfect negative correlation (Blow, 2003).

$$CC = \int v(P(x) - \overline{P}) (Q(x) - \overline{Q}) dV / \int v(P^2(x) - \overline{P}^2)dV \int v (Q^2(x) - \overline{Q}^2)dV$$

The correlation coefficient is most useful in the early stages of strucutre solution, but at the structure improves through refinement and rebuilding the R-free becomes a more sensitive measure of the correctness of the model.

2.6.5 Building an atomic model and refinement

After obtaining the phase, an initial atomic model can be produced using AutoBuild (PHENIX). The program begins with either experimental phases from AutoSol or with an atomic model that can be used to generate calculated phases and carries out cycles of model-building.

The atomic model output from AutoBuild was refined using REFMAC5 and/or phenix.refine. The program utilises the 5% experimental data flagged to assess R-free. Cycles of rebuilding via COOT and refinement were made until an R-factor of less than 20% and R-free of less than 25% were obtained.

The final model is validated and assessed via the Ramachandran plot and MolProbability software, ensuring the stereochemistry of the model is correct while obtaining a low R-factor and R-free.

2.6.6 Brief crystallographic theory

A protein crystal consists of a series of repeated patterns of protein molecules. This repeating pattern is known as the unit cell, which is the smallest building block of a crystal. The unit cell is repeated in three-dimensions to form the crystal. The

coordinates for atoms in a series of unit cells is produced from symmetry operations, such as translation and rotation.

The rules for diffraction are governed by the Bragg's Law, which states that n λ = 2 *d* sin θ, where λ is the X-ray wavelength, *d* is the distance between crystal planes, θ is the angle of the diffracted wave and n is an integer known as the order of the diffracted beam. This is demonstrated in Figure 2.6. The distance between the diffraction spots represents the size of the unit cell while the spot intensity depends on the protein structure.

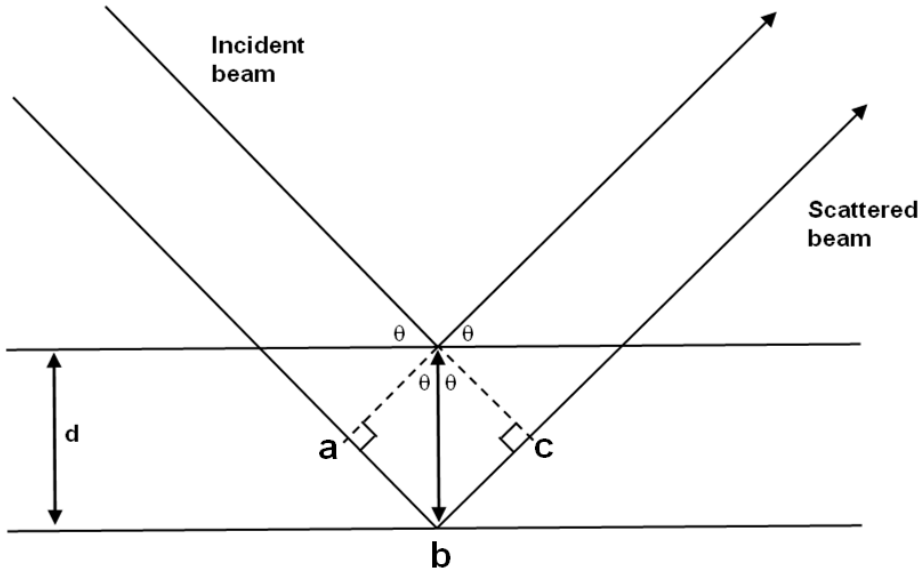


Figure 2.6 Satisfaction of Bragg's Law to obtain diffraction. The X-ray beams are scattered from lattice planes separated by the inter-planar distance *d*. In order to give constructive interference, the two beams have to remain in phase and therefore, the lower beam has to travel the extra distance (*ab* and *bc*). This path difference is given by trigonometry as $2d\sin\theta$, where $2d\sin\theta$ is the distance ab (or bc) and 2θ is the scattering angle.

X-ray diffractometers utilise CCD detectors, which only measures the intensity of the diffraction and thus, poses a problem on phase determination. In a typical diffraction experiment, the intensity of diffraction from planes is measured. In order to calculate the electron density, both the phase and the amplitude of the wave are needed. The amplitude is proportional to the square root of the intensity, and is proportional to the number of electrons in the plane. The phase can be obtained by exploiting the anomalous dispersion and the tuneability of synchrotrons, which enable measurement of the amplitude directly and the phase to be calculated from the anomalous differences in the scattering. To produce anomalous scattering, (seleno-methionine or) metal ions must be incorporated or must be bound specifically to the protein. In this work, Single-wavelength Anomalous Dispersion (SAD) was used but several other techniques, such as isomorphous replacement with anomalous scattering can also be used to solve the phase problem.

The electron density equation is stated below. On the left side of the equation is the electron density (ρ), a three-dimensional function in real space (x, y, z). This can be calculated at each position is space (x,y,z) by summing the structure factor |F(h,k,l)| amplitude of reflection (h,k,l) with the phase angle α(h,k,l). The volume of the unit cell (V) occurs in the scale factor:

$$\rho(x,y,z) = 1/V \sum_{hkl} | F(h,k,l) | \exp [-2\pi i(hx+ky+lz)+i\alpha(h,k,l)]$$

Once the electron density map is generated it is interpreted an atomic model and can be refined subsequently.

2.7 Circular Dichroism

Circular dichroism (CD) is a spectroscopic technique used to gather information about the secondary structure of proteins in solution. As opposed to linearly polarised light, which oscillates only in a single plane, circularly polarised light has a continuously rotating plane of oscillation, forming a helix as it travels. This helix can turn either in a clockwise direction producing a right circularly polarised light or in an anti-clockwise direction to generate a left circularly polarised light. A CD spectrophotometer measures the differential absorption of left and right circularly polarised light. This means that in order to acquire CD information, the molecule in question has to exhibit asymmetric characteristic with preferential absorption of either left or right circularly polarised light.

CD is defined as follows:

$\Delta A = (AL - AR) / A$

AL = absorbance for left circularly polarised light

AR = absorbance for right circularly polarised light

A = absorbance for unpolarised light

The structural characteristic of a sample is usually assessed in the CD spectrum of either near UV (> 250 nm) or far UV (185 – 250 nm) range. The near UV CD spectrum is used to acquire information on the tertiary structure of the protein. This signal is largely due to the absorption, dipole orientation and the nature of the environment around tyrosine, phenylalanine, cysteine and tryptophan. On the other hand, the far UV CD spectrum, which is used in this project, provides an estimate fraction of a molecule that has alpha-helical conformation, beta-sheet conformation

and random coil (Figure 2.7). Although, CD cannot pinpoint where the detected secondary structure is within the molecule, it can be a valuable tool in assessing whether the protein sample is properly folded.

Typically, a 200 µl diluted protein sample with a concentration of about 0.1 mg/ml is placed in a quartz cuvette with a 1 mm path length and subsequently placed onto the Chirascan CD spectrophotometer (Applied Photophysics). Absorbance scans were carried out from 185 to 250 nm at 0.5 nm intervals. For each sample scan, three trials were taken and then averaged. An absorbance scan of the buffer was also taken and used as a baseline. The baseline was used to deduct the absorbance scan of samples. Data acquired from CD were uploaded to an internet-based deconvoluting programme DICHROWEB (http://dichroweb.cryst.bbk.ac.uk), which interprets CD spectra and employs complex algorithm to provide predictions of proportions of each secondary structure characteristic present.

A CD spectroscopy technique can be used to study the conformational stability of a protein under stress. Protein samples can be subjected to thermal stress (by changing the temperature), pH stress (addition of either acid or base) or addition of denaturants (addition of urea, for example). This project utilises the thermal stress to check the stability of the protein sample with a change in the temperature. To do this, the sample was left in a quartz cell inside the CD spectrophotometer. The temperature was increased at five-step incremental starting from 8°C to 88°C, with each step an absorbance scan between 185 to 250 nm was taken. After reaching 88°C, the temperature was lowered down to initial temperature of 8°C to identify the reversibility of thermal unfolding.

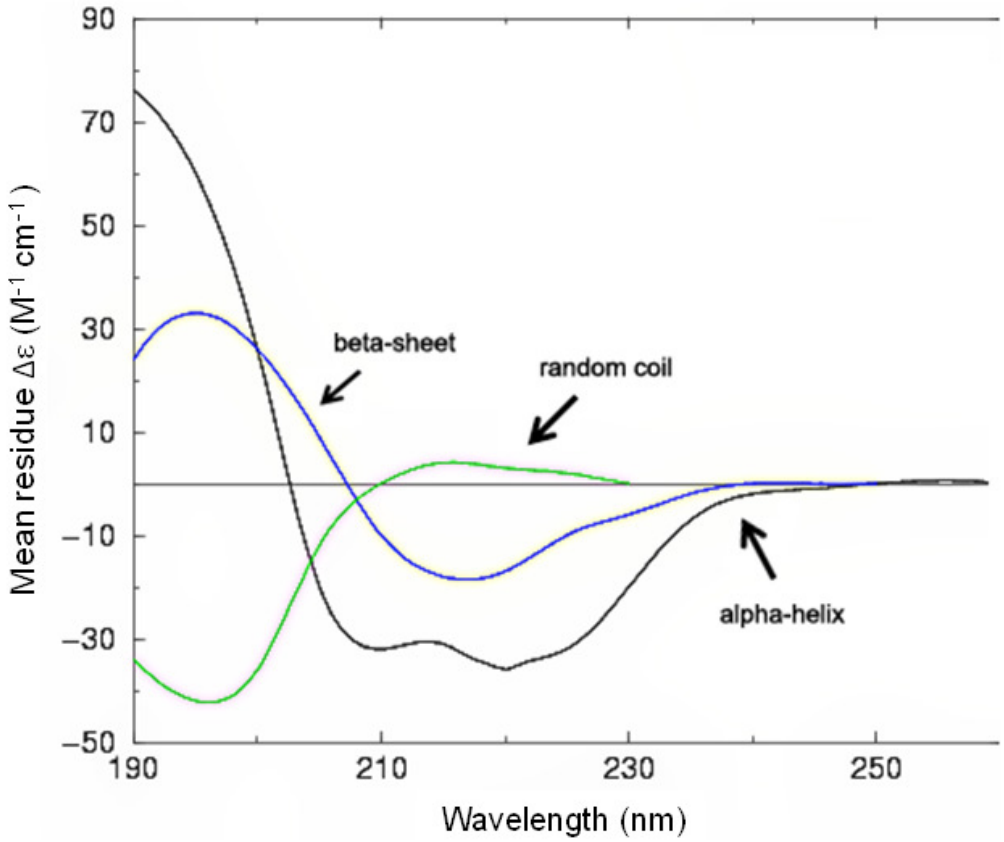


Figure 2.7 Reference CD spectra for various secondary structure conformations. Alpha helix (black) has a characteristic defined with dips in the wavelength 208 and 222. Beta sheet (blue) has a CD spectra fingerprint of a dip in 217 nm. Random coil (green) exhibits no dip in the regions between wavelengths 200 and 230nm; instead there is a dip at 195 nm. Adapted from (Brahms and Brahms, 1980).

2.8 Transmission electron microscopy

Transmission electron microscopy (TEM) is a technique used to visualise intracellular structures by passing through a beam of electrons through ultra thin specimens embedded with bacterial cells in question. An image is created from the interaction of the electrons transmitted through the specimen and focused onto an imaging device.

2.8.1 Dehydration and embedding

The *E. coli* strain BL21*(DE3) pLysS was transformed with the five PduA mutant constructs and PduA wild type. 50 ml cells were cultured at 37 ºC to an OD_{600} of 0.8 and protein production was induced with 400 µM IPTG at 18 °C overnight. 25 ml cells were harvested by centrifugation at 4000 rpm for 10 min. The cells were re-suspended in 2 ml fix consisting of 2.5 % glutaraldehyde in 1x PBS.

Subsequently, the cells were pelleted and washed twice with PBS to remove traces of the fixing solution. The cells were stained for 1 hour in 1% osmium tetroxide and then washed with PBS before dehydration. Dehydration was accomplished by subjecting the samples to a solvent gradient: 60 % IMS over night, 90 % IMS for 15 minutes, 100 % IMS for 15 minutes and twice 100 % dried ethanol for 2 hours. The cells were embedded by first incubating them over night in 30 % Agar Low Viscosity Resin in dried ethanol and then embedding them in 100 % Agar Low Viscosity Resin for 180 minutes (3 changes of LV resin) constituted for a block with medium hardness. The samples were placed in 0.5 ml embedding tubes, centrifuged

for 5 minutes at 4,000 x g to concentrate the cells to the tip and incubated at 60ºC overnight to polymerise.

2.8.2 Sectioning and visualisation

Specimens were thin sectioned with a diamond knife on an RMC MT-6000-XL ultramicrotome, collected on copper grids and post-stained with 5% uranyl acetate for 30 minutes at 60ºC and 0.1% lead citrate for 10 minutes at room temperature. Sections were then observed and photographed with a JEOL-1230 transmission electron microscope.

Chapter 3

Structure of trimeric PduT with a 4Fe-4S cluster binding site

3.1 Introduction

Initial characterisation of PduT revealed that the protein has a novel 4Fe-4S cluster. This suggests that the shell protein may serve as a conduit for electron transfer. Sequence analysis of the shell protein indicates that the subunit has four cysteines. When one of these cysteines (Cys-38) is mutated into alanine, the EPR spectral fingerprint for the 4Fe-4S cluster is lost, indicating that Cys-38 is crucial for coordination of the cluster. The presence of the 4Fe-4S cluster led the authors to suggest PduT is a homotetramer with a cysteine from each subunit binding an Fe of the iron-sulfur cluster (Parsons et al., 2008). In this Chapter, the determination of the structure of PduT is described, which reveals its trimeric assembly and a binding site for 4Fe-4S cluster.

3.2 Production, purification and crystallisation of PduT

The protocol for overexpression of PduT is explained in detail in Chapter 2. As the protein was known to consist of iron-sulfur cluster, freezing of harvested cells was

not advisable; instead, the cells were immediately resuspended in buffer (50 mM Tris-HCl pH 8.0, 0.5 M NaCl, 10 mM imidazole) and lysed by sonication.

3.2.1 Nickel affinity purification

As explained in Chapter 2, the lysate was applied in column filled with nickel-charged resins. A distinct brown colour can be seen on the top layer, which becomes more apparent after a few washes with buffer containing 10 mM, 50 mM and 100 mM imidazole (Figure 3.1A). Upon application of buffer containing 400 mM imidazole, a reddish-brown colour protein was eluted and collected (Figure 3.1B). This distinct colour may be an indication of the presence of iron-sulfur cluster within the protein. A similar observation was made by Josh Parsons (Parsons, 2008).

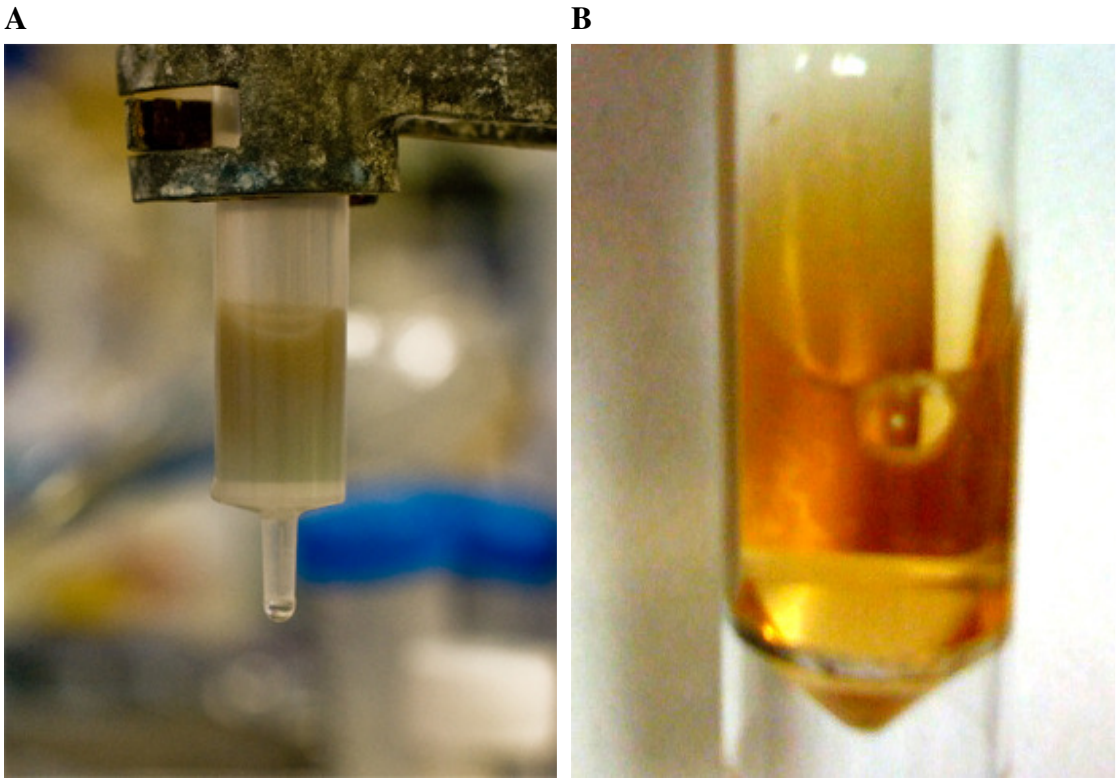


Figure 3.1 Distinct colour of PduT. **(A)** PduT lysate loaded onto the Nickel column, **(B)** PduT eluted from the Nickel column.

3.2.2 Further purification of PduT

The chromatographic profile of PduT on the S200 column showed that the protein eluted at 15.37 ml (Figure 3.2A). The calibration of this column relates native molecular weight and size with the equation: y = -10.53x + 64.85, where y = peak elution, x = log (Da) and the slope being 10.53. The elution volume at 15.37 ml gave a calculated molecular weight of approximately 57 kDa, which corresponds to three times the size of a shell protein subunit (19 kDa), indicating that the protein assembles into trimer, as opposed to the earlier predicted tetramer assembly (Parsons et al., 2008). Polyacrylamide gel analysis of the elution peak (Figure 3.2B) shows the major protein is PduT but there is still some contamination present.

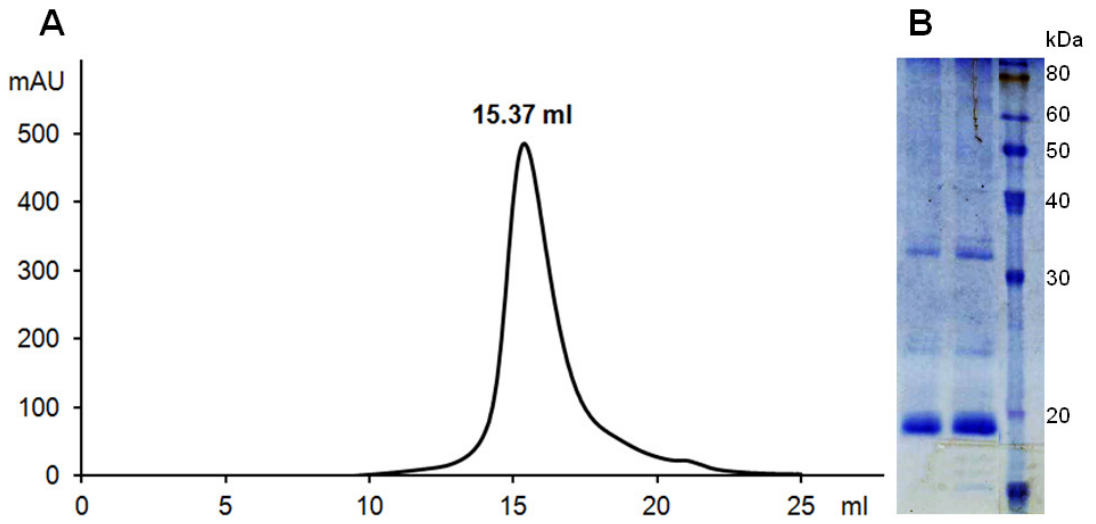


Figure 3.2 Further purification of PduT. **(A)** S200 elution profile of the shell protein, with a peak volume at 15.37 ml. Samples at this region were run in polyacrylamide gel and represented in (**B**). A thick band found at around the size of 20 kDa that corresponds to the subunit size of PduT.

3.2.3 Crystallisation of PduT

1 µl of 20 mg/ml of PduT was mixed with 1 µl of the corresponding reservoir of Hampton Screen I and II using the hanging drop method. Several drops with microcrystals were seen after a week of equilibration. A few of these conditions were selected for optimisation. Hampton Screen II – 46 (0.1 M Bicine pH 9, 0.1 M NaCl, 20% PEG 550) produced hexagonal plate crystals. The crystallisation reservoir condition was refined according to Figure 3.3A, detailing the buffer gradient employed in a 24-well plate. The concentration of PduT was also varied to optimise crystal sizes. PduT crystallised at a low pH values using PEG 550 as precipitant; this particular reservoir condition was further optimised according to Figure 3.3B. A final protein concentration of 7 mg/ml in the drop produces hexagonal plate crystals with a size of about 0.1 mm across (Figure 3.4). The reservoir condition was 0.1 M Bicine pH 9, 0.1 M NaCl and 25% PEG 550.

A

		NaCl concentration			% PEG 550		
		0.08 M	0.12 M	0.14 M	15%	20%	25%
pH of 0.1 M Bicine	7.5	A1	A2	A3	A4	A5	A6
	8.0	B1	B2	B3	B4	B5	B6
	8.5	C1	C2	C3	C4	C5	C6
	9.0	D1	D2	D3	D4	D5	D6

B

		% PEG 550					
		20%	21%	22%	23%	24%	25%
pH of 0.1 M Bicine	7.5	A1	A2	A3	A4	A5	A6
	8.0	B1	B2	B3	B4	B5	B6
		26%	27%	28%	29%	30%	31%
	7.5	C1	C2	C3	C4	C5	C6
	8.0	D1	D2	D3	D4	D5	D6

Figure 3.3 Optimisation of PduT crystallisation conditions. **(A)** Conditions were refined by varying the concentration of NaCl and PEG550 (columns) and the pH of 0.1M Bicine (rows) on a 24-well crystallisation plate. **(B)** After identification of conditions with good size crystals, the crystallisation condition was further refined with smaller step.

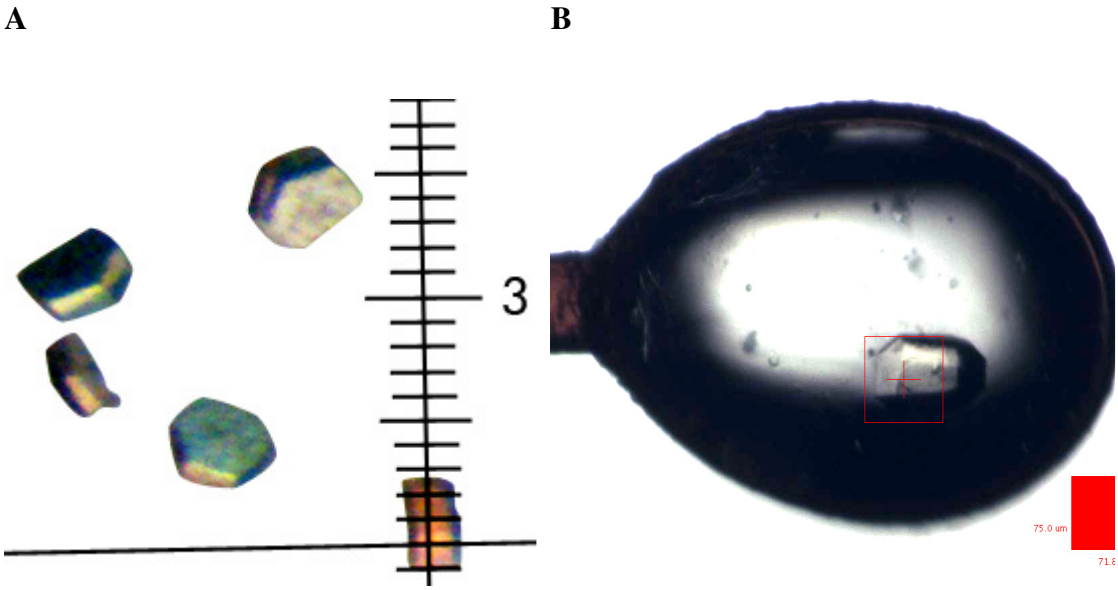


Figure 3.4 PduT crystals. **(A)** Hexagonal PduT crystals grown from 0.1 M Bicine pH 8, 0.1 M NaCl and 20% PEG 550. **(B)** Single PduT crystal in a cryoprotected loop, about to be illuminated with an X-ray beam. The beam size to be used is 75.0 µm x 71.9 µm. Crystal size is approximately 80 µm across.

3.3 Native data collection

A single hexagonal shaped PduT crystal was picked up with a 0.1 mm litholoop and briefly dipped into a cryoprotectant solution. Four different cryoprotectants (glycerol, PEG400, ethylene glycol and MPD) were tried. The reservoir solution was augmented with 15% of the listed cryoprotectants.

Native data were collected at Station ID14-1, ESRF, Grenoble, France. Data collection took place at 100K, at a wavelength of 0.9733 Å, using ADSC Q210 CCD detector to record diffraction images. A total of 135 images were taken, over a

rotation range of 135°, using a 1° rotation angle per image. One of the diffraction images is shown in Figure 3.5.

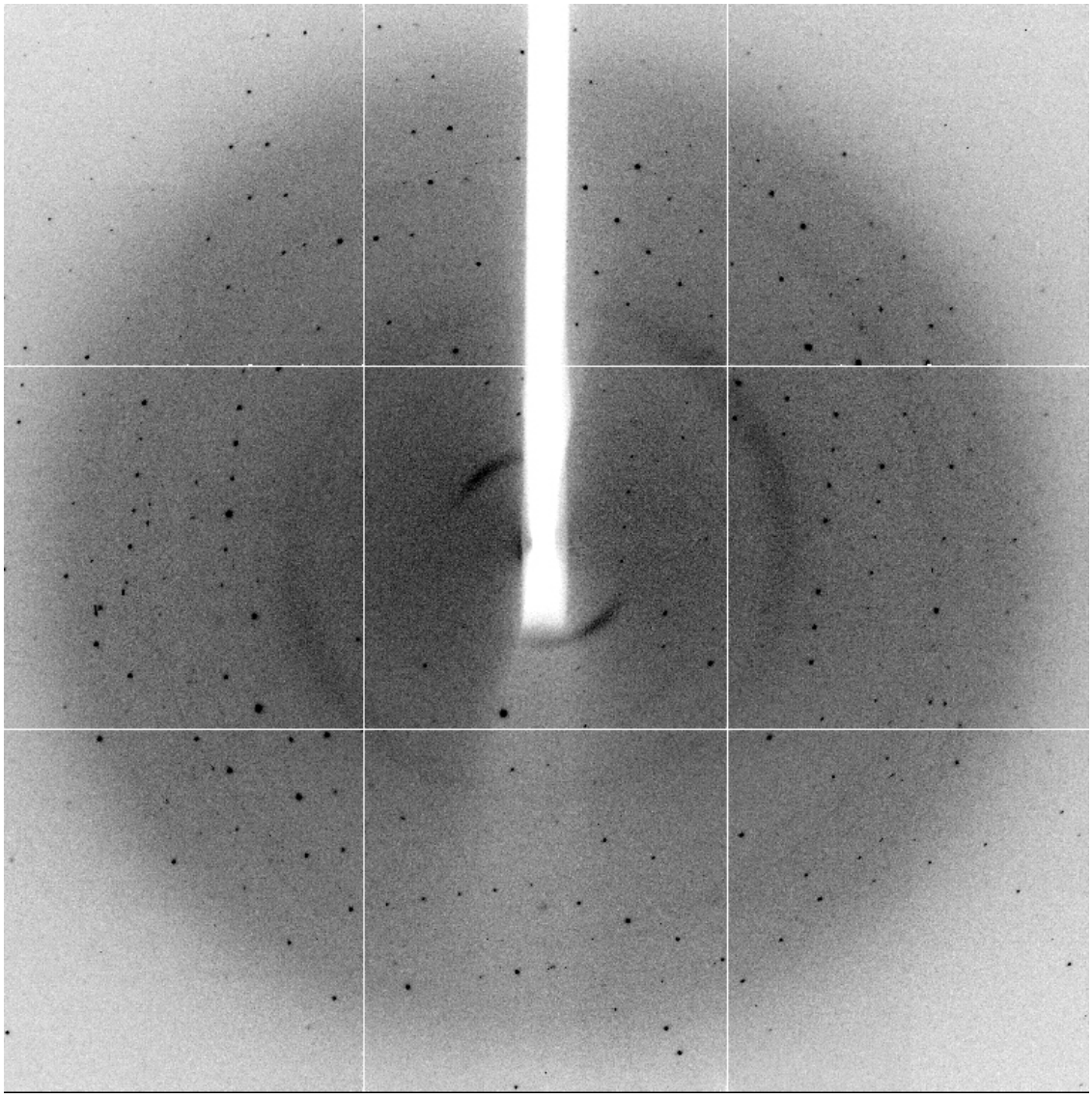


Figure 3.5 Diffraction image from a native PduT crystal. The maximum resolution is 1.86 Å. The image was recorded at station ID14-1, ESRF, Grenoble, France.

Data were indexed and scaled and the data statistics are listed in Table 3.1. The best diffracting crystal used 15% PEG400 as cryoprotectant and diffracted to 1.86 Å resolution. The crystal has a unit cell dimension of a = 67.8 Å, b = 67.8 Å, c = 62.0 Å and the space group was $P6_3$. PduT crystallises with 1 monomer per asymmetric unit, giving a Vm of 2.15 and a solvent content of 43%.

Table 3.1 Crystallographic data statistics for native PduT.

	Overall	Inner Shell	Outer Shell
Low resolution limit	33.84	33.84	1.86
High resolution limit	1.86	1.96	1.86
Multiplicity	8.2	8.2	8.0
Completeness (%)	100	100	100
Number of unique reflections	13647	455	1979
Rmerge ♣	0.095	0.046	0.58
Mean $I/\sigma(I)$	18.5	69	3.7
Rpim ♦	0.035	0.018	0.22
Rmeas ♠	0.101	0.049	0.0621
Wilson B-factor ($Å^2$)	21.2		

♣ $R_{merge} = \Sigma_{hkl} \Sigma_i | I_i (hkl) - (I(hkl)) | / \Sigma_{hkl} \Sigma_i I_i (hkl)$, where $I_i (hkl)$ is the intensity of the *i*th observation and ($I(hkl)$) is the main intensity of the reflection.

♦ R_{pim} = measure of the quality of the data taking account of the multiplicity (Evans, 2006).

♠ R_{meas} = improved version of the traditional R_{merge} and measures how well the different observations agree (Evans, 2006).

3.4 Heaving atom phasing

Crystals were soaked in a number of heavy metal compounds. Crystals which gave a good diffraction pattern were scanned for their specific soaked heavy metal compound. Figure 3.6 shows one of the crystals with X-ray fluorescence indicating the presence of osmium. Some of the protein crystals were twinned and these were disregarded. Table 3.2 presents a list of non-twinned crystals soaked with various heavy metal compounds with the corresponding anomalous slope from SCALA. Among these, the PduT crystal soaked into 5 mM ammonium hexachloroosmate for 10 minutes diffracted to 1.8 Å resolution and yielded the best anomalous data.

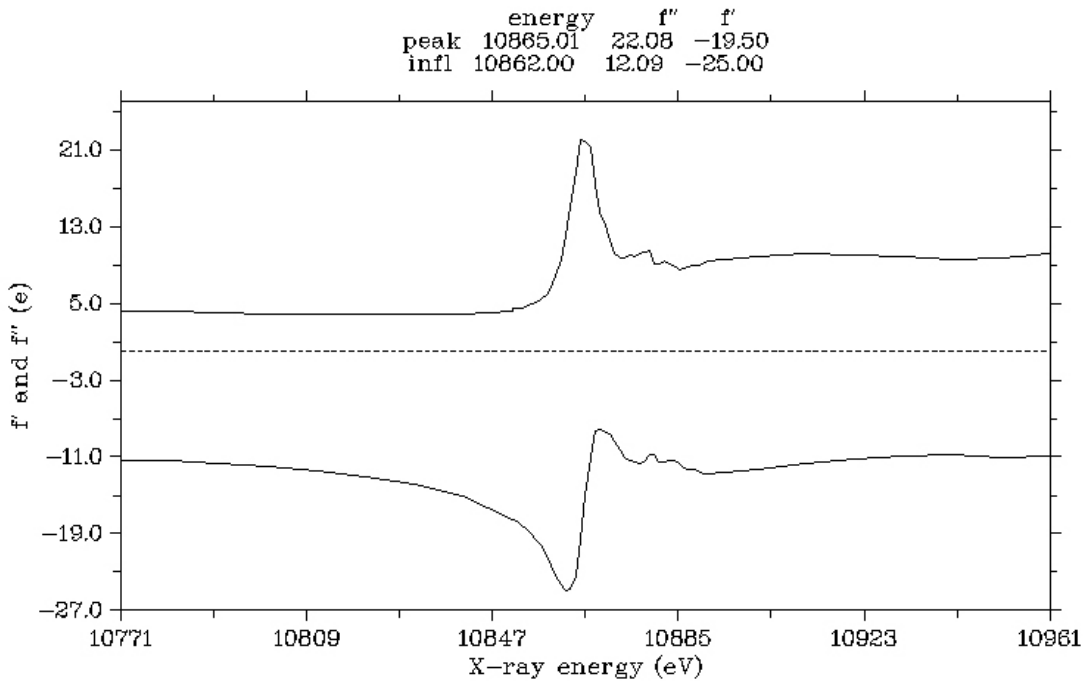


Figure 3.6 X-ray fluorescence plot obtained from ammonium hexachloroosmate soaked crystal. Figure produced using CHOOCH (Evans and Pettifer, 2001). The X-ray energy and its corresponding f' and f'' are labelled accordingly for peak and inflection point.

Table 3.2 A list of heavy metal compounds used to soak PduT crystals. The twinning score ($<I^2>/<I>^2$, where I is the intensity, has an intensity ratio close to 2 for untwined data) (Padilla and Yeates, 2003) and MSAN (midslope of anomalous normal probability; a slope > 1 indicates that the measured anomalous differences are greater than expected from the standard deviation) are given (Evans, 2006).

Heavy metal	Soak Time	Space Group	Twinning Score	MSAN
Ammonium hexachloroosmate	10 minutes	$P6_3$	2.02	1.553
Cis-dichloroammine platinate	10 minutes	$P6_3$	2.18	1.075
Potassium tetrachloroplatinate(II)	7 days	$P6_3$	1.92	0.969
Potassium dicyanoaurate	10 minutes	$P6_3$	2.23	1.06
Potassium dicyanoaurate	20 minutes	$P6_3$	2.03	1.043
Potassium hexachloroosmate	10 minutes	$P6_3$	2.10	1.139
Potassium osmate	10 minutes	$P6_3$	2.24	1.043
Potassium tetracyanoplatinate	10 minutes	$P6_3$	2.19	1.029
Sodium bromide	1 minute	$P6_3$	1.99	1.296
Sodium bromide	1 minute	$P6_3$	2.17	1.196
Sodium bromide	1 minute	$P6_3$	2.05	1.084
Sodium hexachloroiridate	20 minutes	$P6_3$	1.96	1.054
Sodium hexachloroplatinate	10 minutes	$P6_3$	2.08	1.03
Sodium tetrachloroplatinate	10 minutes	$P6_3$	2.16	1.003
Thiomersal	Crystal grown with heavy metal compound	$P2_12_12_1$	2.28	1.338
Uranyl acetate	10 minutes	$P6_3$	2.05	1.126

3.5 Model building and refinement

The data for PduT crystal soaked with ammonium hexachloroosmate were indexed and scaled; the statistics giving the data quality are listed in Table 3.3. The structure of PduT was solved using single-wavelength anomalous dispersion (SAD) phasing, exploiting the anomalous diffraction of the osmate ions, using PHENIX. The AutoSol wizard from PHENIX found eight osmate sites and gave phases with a figure of merit of 0.340. AutoBuild produced a model of 175 residues in six fragments with 85 waters, giving an R-factor, an R-free and a correlation coefficient of 0.223, 0.245 and 0.80, respectively. This model was used in cycles of refinement and rebuilding against the native PduT data to yield the final PduT structure at 1.86 Å resolution. The refinement and validation statistics are presented in Table 3.4. The final model of PduT has clearly defined polypeptide backbone in the electron density map for residues 2–184, with the exception of residues 37–41. An example of the electron density is shown in Figure 3.7. The residues preceding Ser-2 and those in the loop containing residues 37–41 are more flexible and are not clearly defined in the electron density map. Almost all of the side chains have clearly defined electron density, with the notable exception being Phe-130 adjacent to the flexible loop, and some residues exhibit two conformations. The faint reddish colour of the protein sample suggested the retention of the 4Fe–4S iron–sulfur cluster after elution from the nickel column, but the 4Fe-4S cluster is not seen in the final structure; this is presumably because the iron–sulfur cluster is oxygen-labile and is lost during crystallisation.

Table 3.3 Crystallographic data statistics for PduT crystal soaked in ammonium hexachloroosmate.

	Overall (Outer Shell)
Wavelength (Å)	1.1410
No. of osmium in asymmetric unit	5
Low resolution limit	42.53 (1.82)
High resolution limit	1.82 (1.78)
Number of unique reflections	15545 (1089)
Multiplicity	6.6 (5.6 - 3.0)
Completeness (%)	99.3 (95.2)
Rmerge ♣	0.045 (0.429)
Mean $I/\sigma(I)$	21.2 (2.2)
Rpim ♦	0.026 (0.339)
Rmeas ♠	0.07 (0.623)
MSAN ♥	1.55
Wilson B-factor ($Å^2$)	26.4

♣ $R_{merge} = \Sigma_{hkl} \Sigma_i | I_i (hkl) - (I(hkl)) | / \Sigma_{hkl} \Sigma_i I_i (hkl)$, where $I_i (hkl)$ is the intensity of the *i*th observation and ($I(hkl)$) is the main intensity of the reflection.

♦ R_{pim} = measure of the quality of the data taking account of the multiplicity (Evans, 2006).

♠ R_{meas} = improved version of the traditional R_{merge} and measures how well the different observations agree (Evans, 2006).

♥ MSAN = midslope of anomalous normal probability; the difference between measured anomalous differences from expected standard deviation (Evans, 2006).

Table 3.4 Refinement and validation statistics for the final model of PduT.

	Statistics
Resolution (Å)	29.71 – 1.86
Reflections (work / test)	12258 / 1372
R-factor / R-free ♣	0.199 / 0.252 ♦
RMSD bonds (Å) / angles (°)	0.023 / 2.896
Ramachandran plot statistics, residues in	
Most favoured regions	91.2
Additional allowed regions	6.9
Generously allowed regions	1.3
Disallowed regions	0.6

♣ R-factor = Σ_{hkl} $|F_o - F_c|$ / Σ_{hkl} F_o, where F_o and F_c represent the observed and calculated structure factors, respectively. The R-Factor is calculated using the 95% of the data that were included in refinement and R-free is calculated using the excluded 5%.

♦ The higher statistics of final R-free (compare to the initial R-free of 0.245) is attributed to the incorrect amino acid fitting, incomplete build and poor stereochemistry of the initial model output of AutoBuild.

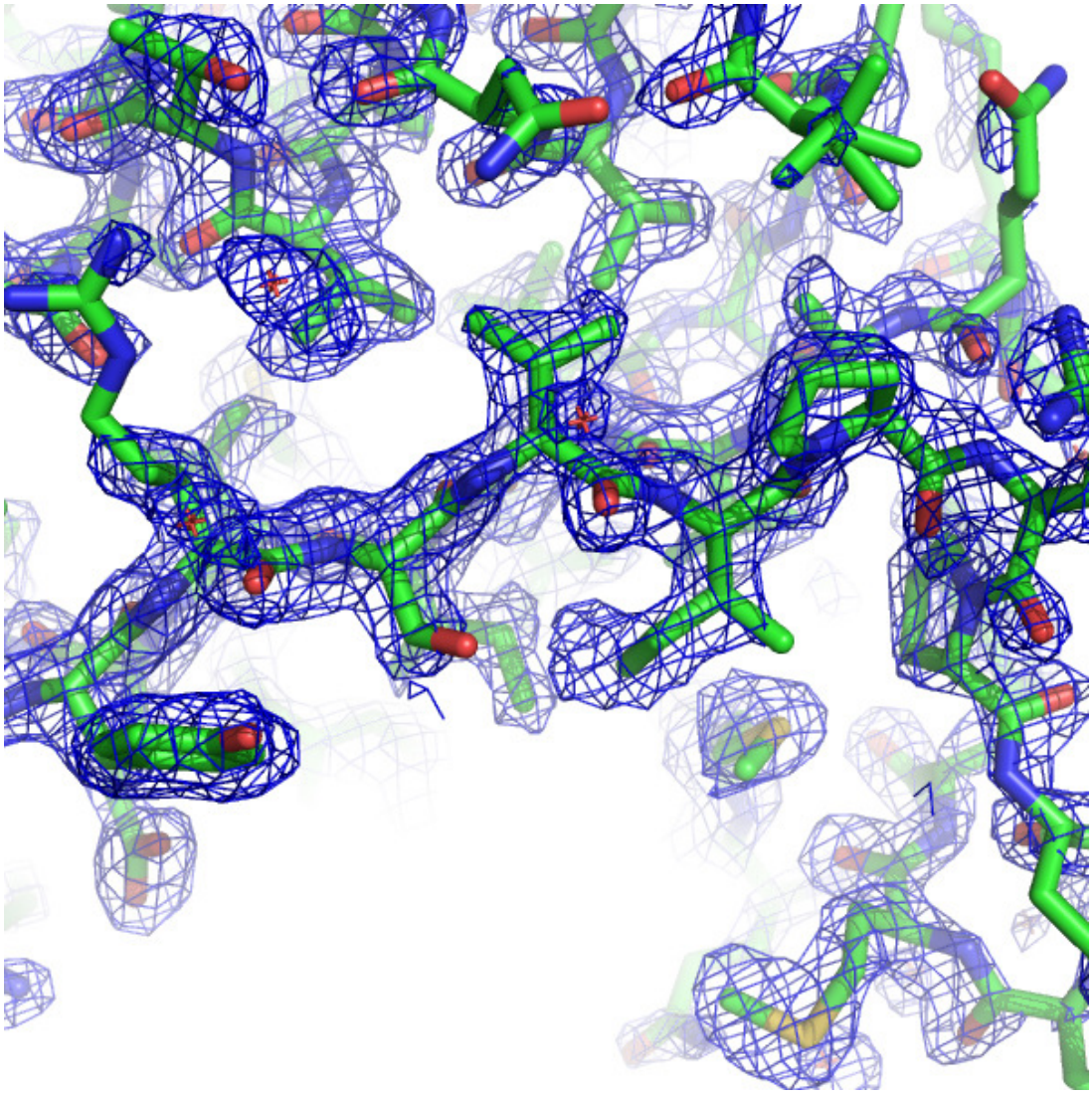


Figure 3.7 Quality of the electron density map of PduT. The σ_A-weighted ($2mF_{obs}$ - DF_{calc}) Fourier synthesis omit map contoured at 1σ showing the quality of the electron density (blue mesh) around Val-169 on β8 of PduT. The structure was solved by SAD phasing exploiting the binding of ammonium hexachloroosmate to the protein. The simple composite omit map was calculated using PHENIX (Adams et al., 2010). This Figure was produced using PyMOL (DeLano and Lam, 2005).

3.6 Subunit analysis

Most shell proteins characterised to date comprise approximately 90 residues and have a single canonical BMC domain. PduT is a 184 residue shell protein with two canonical BMC repeats per subunit. The canonical BMC domains of PduT each consist of two β–α–β motifs connected by a beta hairpin forming an anti-parallel β-sheet (Figure 3.8). The two BMC domains of the PduT subunit are connected by a short α-helix (H3) and a β-turn. The two BMC domains of PduT have 31.5% sequence identity (Figure 3.9A) and superimpose with an RMSD of 1.25 Å for 396 equivalent atoms, this highlights their structural similarity (Figure 3.9B). The plane of the β-sheet of the second BMC domain is however skewed compared to that of the first (Figure 3.9C), a situation that cannot arise in the single BMC repeat proteins since adjacent subunits are related by a pure rotation.

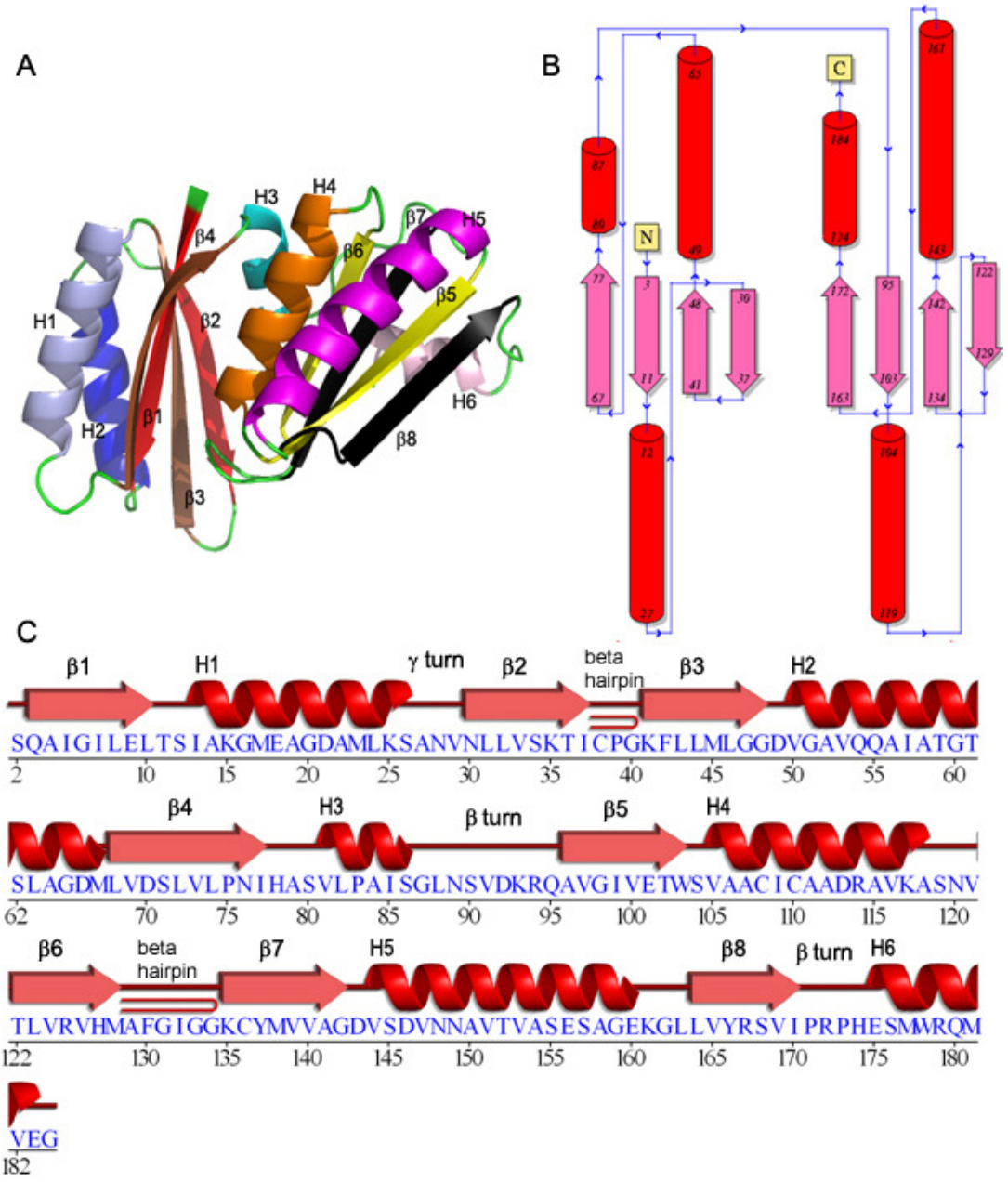


Figure 3.8 The tertiary structure of the PduT subunit. **(A)** Cartoon representation of the tertiary structure of PduT subunit which comprises two BMC repeats. **(B)** Schematic drawing of the topology of the PduT subunit. **(C)** Amino acid sequence of the PduT with secondary structures marked.

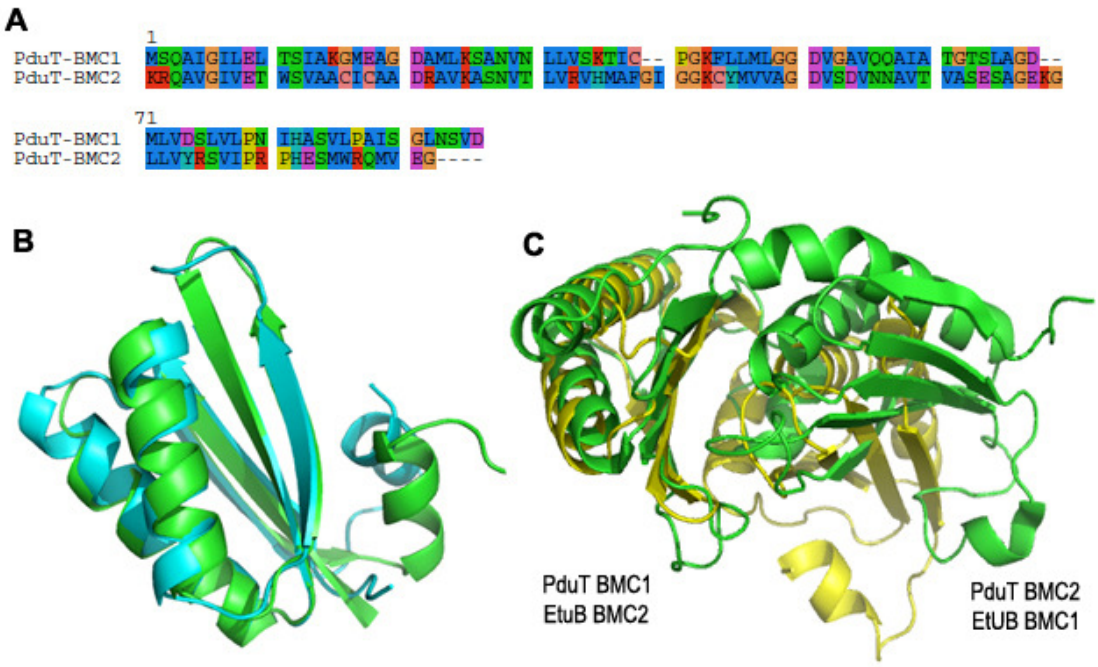


Figure 3.9 Comparison of the first and second BMC repeats of the PduT subunit. **(A)** The sequences of the first and second BMC repeats aligned and **(B)** the BMC domains are superimposed. **(C)** BMC1 of PduT superimposed on BMC2 of EtuB showing optimal superimposition of these two proteins.

3.7 Trimeric structure

Six previously solved shell proteins have a single canonical BMC repeat and form hexamers (Kerfeld et al., 2005, Tsai et al., 2007, Tanaka et al., 2009, Tsai et al., 2009, Tanaka et al., 2010), two others have a single circularly permuted BMC repeat - EutS (Tanaka et al., 2010) and PduU (Crowley et al., 2008), while four shell proteins have two BMC domains per subunit with the circularly permutated fold (Tanaka et al., 2010, Klein et al., 2009, Sagermann et al., 2009, Heldt et al., 2009). PduT is different from these, as it has a duplication of the canonical BMC domain. The PduT trimer forms a flat, approximately hexagonally shaped disk with a large central pore. The similarity with the archetypal carboxysome shell protein CsoS1A is shown in Figure 3.10A–C. In comparison to CsoS1A, the general hexameric appearance of PduT seemingly was perturbed as a result of the skewed plane of β-sheet of the second BMC domain.

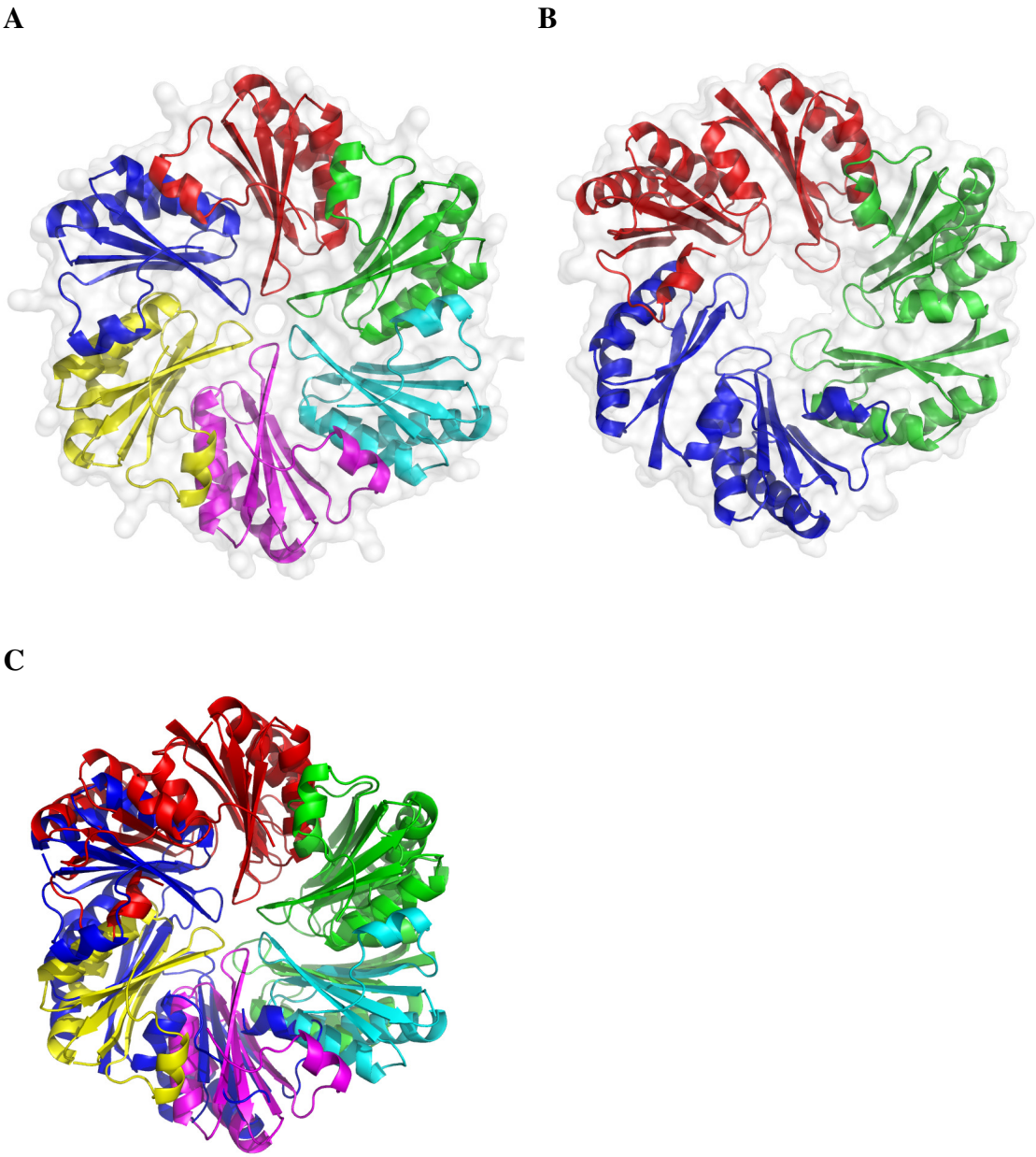


Figure 3.10 Comparison of quaternary structures of PduT and CsoS1A. **(A)** The archetypal carboxysome hexamer (CsoS1A hexamer), **(B)** the PduT trimer and **(C)** the CsoS1A hexamer and PduT trimer superimposed. This Figure was produced using PyMOL (DeLano and Lam, 2005).

3.7.1 Molecular tiling of PduT

The Csos1A hexamers constitute the major component of the alpha carboxysome shell and showed to form a tightly packing to form a molecular layer, representing the facets of the shell (Tsai et al., 2007). The tight packing is stabilised by two anti-parallel lysines (Figure 3.11A). PduT trimer is roughly the correct size to fit into a sheet of CsoS1A molecules, but it is not an exact fit (Figure 3.11B). A similar result is observed when a PduT trimer is tiled in with PduA hexamers (Figure 3.11C), the major component of Pdu metabolosome shell. The conserved anti-parallel lysines are present but the distance between them is too loose for one of the BMC repeat, as shown in Figure 3.12B (6 Å) compared to approximately 3 Å distance between PduA lysines (Figure 3.12A).

In addition to the anti-parallel lysines, the Arg-79 forms a hydrogen bond with Val-24 of the symmetry-related PduA (Figure 3.12A). This also stabilises the molecular tiling. In one of the BMC repeat of PduT, Val-24 is not conserved; instead the backbone carbonyl oxygen atom of Leu-24 can substitute the oxygen atom of Val-24 to form hydrogen bond with the side-chain guanidium group of Arg-79 of PduA (Figure 3.12B). In the other BMC repeat, the corresponding valine (Val-116) is present but not the arginine (Asn-76 instead). In this case, it appears that formation of hydrogen bond with Arg-79 is not possible due to the long distance (4 Å and 8 Å, respectively as shown in Figure 3.12C).

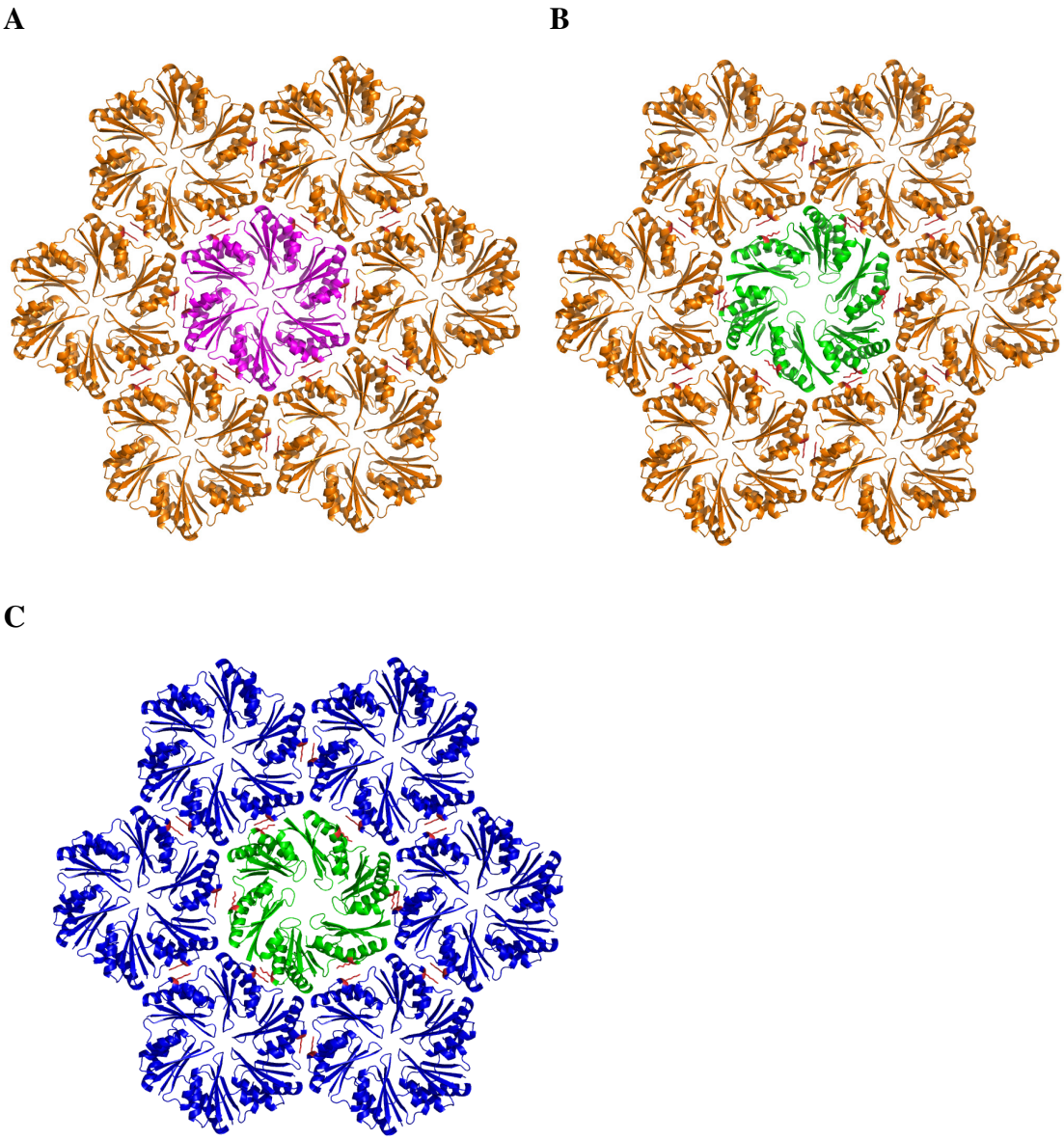


Figure 3.11 Comparison of the higher order structures involving PduT. **(A)** CsoS1A packing in a sheet according to the crystallographic symmetry of the crystal lattice, **(B)** PduT inserted into a sheet of CsoS1A molecules and **(C)** PduT inserted into a sheet of PduA molecules. PduT has approximately the correct dimensions and the lysines highlighted as red sticks necessary to form one of the key signature interactions of sheet mosaics. However, PduT does not fit well, suggesting that it will introduce a distortion into a sheet with CsoS1A and PduA packing. This Figure was produced using PyMOL (DeLano and Lam, 2005).

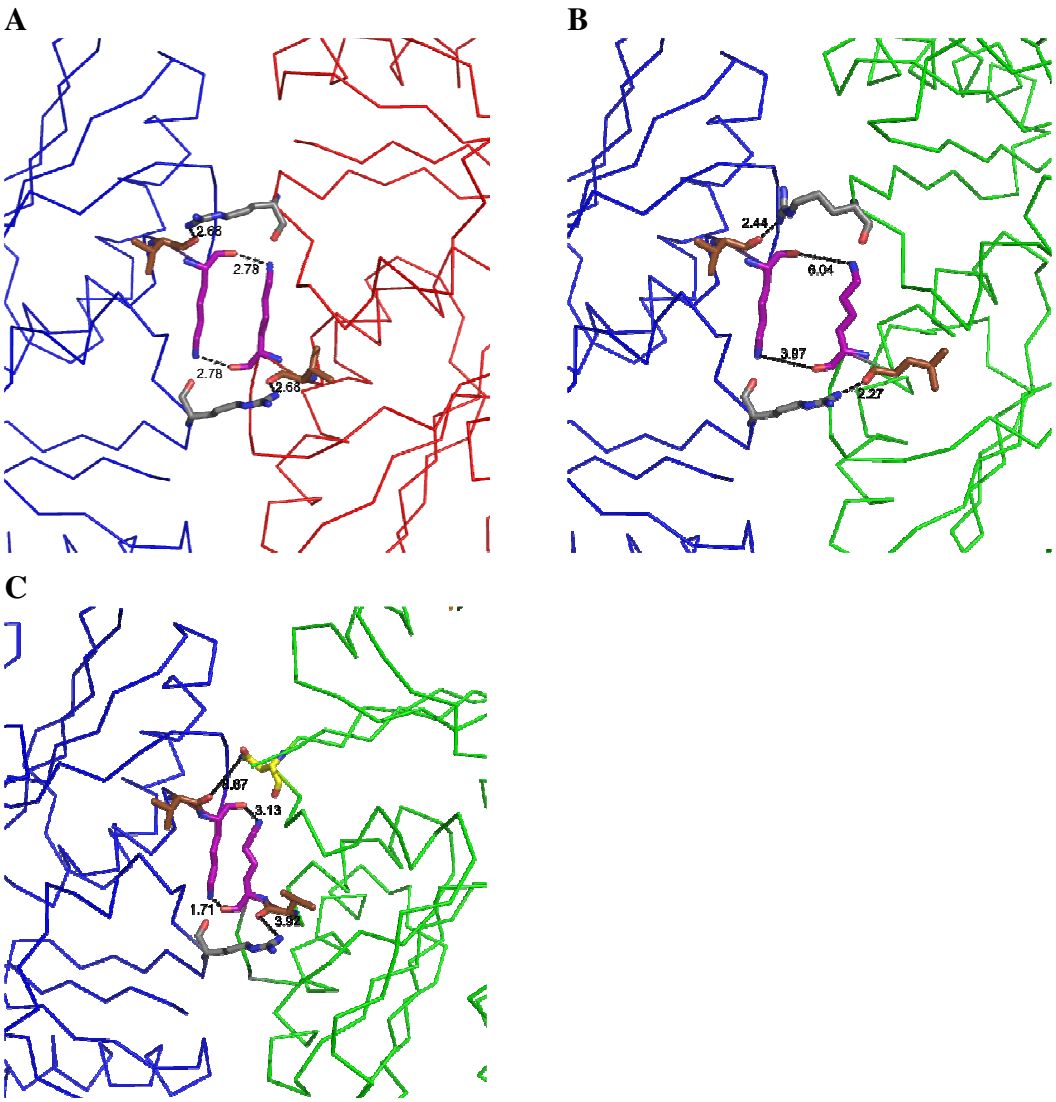


Figure 3.12 A close-up view of the interaction between adjacent structures. **(A)** The two PduA hexamers are coloured separately in red and blue. Lysines are coloured purple. Arg-79 in grey and Val-24 in brown. **(B)** The PduA hexamer is coloured blue. PduT trimer is coloured green. This represents the BMC domain of PduT with a loose lysine-lysine interaction with PduA. Leu-24 of PduT (in brown) substitutes the Val-24 of PduA. **(C)** This represents the BMC domain with PduT lysine close to PduA lysine. In PduT, Asn-76 (coloured yellow) replaces arginine of PduA. Dotted lines with distance further than 3 Å do not illustrate hydrogen bonds but simply to show difference to hydrogen bonding shown in **A**. This Figure was produced using PyMOL (DeLano and Lam, 2005).

3.7.2 Iron sulfur [4Fe-4S] binding site

There are currently five trimeric shell proteins, including PduT with known structures. The three trimeric shell proteins (Csos1D, EutL, and EtuB) were suggested to exist in two different forms. Three small openings are evident in EutL (Sagermann et al., 2009, Takenoya et al., 2010, Tanaka et al., 2010) and EtuB (Heldt et al., 2009) with each opening formed by the subunit. Upon opening of the channel, these small openings transform into a large central pore. PduT exhibits a large triangular-shape opening in the center of the trimer. However, PduT trimer differs in that the central pore is a binding site for an iron-sulfur cluster. The β-hairpin loop from β2 to β3 points towards the pore, producing a three-fold arrangement of cysteines 38 about the molecular three-fold axis (Figure 3.13). This cysteine is implicated in the binding of the 4Fe-4S cluster in PduT, since substitution of this residue caused the characteristic EPR signal to be lost whereas substitution of two other cysteines did not (Parsons et al., 2008). The 4Fe-4S cluster from *E. coli* ferredoxin (PDB code: 2ZVS) can be readily fitted into the central pore such that the essentially cubic cluster conforms to the three-fold symmetry of the trimer with one sulfur and one iron on the molecular three-fold axis (Figure 3.13). The cluster can be rotated so that the other three iron atoms point towards the three cysteine 38 residues. The conformation of the four residue β-hairpin is poorly defined in the electron density map, but it can be readily positioned so as to form S-Fe bonds with the cluster of length approximately 2.3 Å. The on axis iron atom could be either up or down and is potentially available to bind another protein; the

cluster is accessible from both sides and is therefore in a suitable location for single electron transfer across the shell of the bacterial microcompartment.

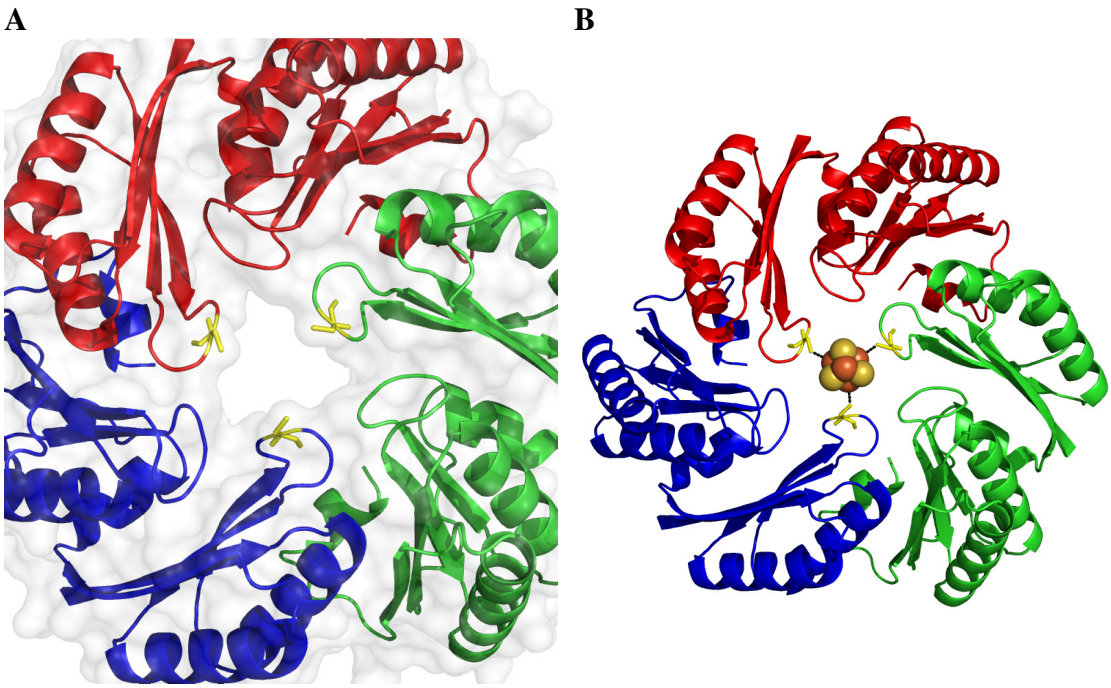


Figure 3.13 PduT and the binding site for 4Fe-4S. **(A)** PduT trimer, showing the three cysteines (coloured yellow) that are pointing towards the central pore. **(B)** Model of PduT trimer fitted in with a 4Fe-4S cluster in its central pore. This Figure was produced using PyMOL (DeLano and Lam, 2005).

3.8 Conclusion and future work

The crystal structure of PduT reveals a trimeric arrangement of subunits each containing a tandem repeat of the canonical BMC domain, contrary to the earlier predicted homotetramer assembly. When fitted into a molecular layer of both CsoS1A and PduA, the trimer has the correct size but the distance of anti-parallel lysines suggests the fit may not be ideal. This implies that the PduT trimer cannot be fitted within a sheet of PduA hexamers without disturbing the packing of the hexamers.

The PduT structure reveals a central pore that none of the known shell proteins have. The cysteine residue previously shown to bind the 4Fe-4S cluster is positioned such that it could bind three of the four iron atoms of a 4Fe-4S cluster, leaving the fourth iron atom free for interaction with another protein such as PduS (Parsons et al., 2008). It has been proposed that the enzyme PduS channels unwanted electrons to the outside of the microcompartment, employing its 4Fe-4S centre to absorb free electrons not utilised in corrin reduction and to pass them onto PduT.

Whilst the structure of PduT provides a wealth of information, the story of PduT is far from complete. We have not been able to experimentally observe the 4Fe-4S cluster *in situ*; and it is not yet known whether the fourth iron atom is facing away or towards the centre of the microcompartment. Many attempts were made to produce the crystal structure of PduT with intact iron-sulfur cluster including the use of the anaerobic glove box and reconstitution of iron-sulfur cluster. These attempts failed to produce crystals with electron density for the iron-sulfur cluster. A modification of the purification step, replacing the Nickel purification step with an ion-exchange

column allows the retention of the distinct colour of the protein even after size exclusion chromatography. This protein sample crystallised in the same condition, as detailed in Chapter 3.2.4 with preference for a higher PEG concentration (30%). The crystal appeared to be cubic in shape (Figure 3.14), in contrast to the hexagonal plate crystals used for the native data collection (Figure 3.4). However, these crystals diffracted poorly with the highest resolution to at best 8 Å. The space group was likely H3 (instead of $P6_3$). Subsequent processing of diffraction data was not successful due to the poor quality of the diffraction spots. In addition to the presence of 4Fe-4S cluster, it will be interesting to see how PduT and PduS interact. Previous study showed that PduS (tagged) and PduT (untagged) co-purified with Nickel affinity purification. If the interaction is deemed to be strong, the two proteins can be crystallised together and provide insight into the electron transfer function of PduT.

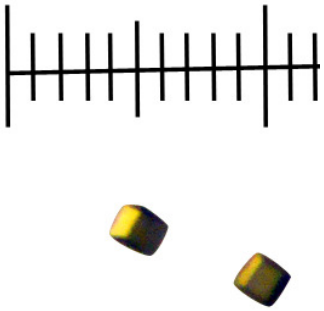


Figure 3.14 Purified PduT (possibly with intact 4Fe-4S cluster; note yellow colour) formed as cubic crystals, an alternative shape to hexagonal crystal. Crystal size is approximately 60 µm x 60 µm.

Chapter 4

Substrate channels in trimeric PduB

4.1 Introduction

Sequence analysis of various Pdu shell proteins indicated that two shell proteins, PduB and PduT, have tandem repeats of BMC domain. In the previous chapter, PduT was shown to have a subunit with two BMC repeats and assemble into a trimer (Crowley et al., 2010, Pang et al., 2011). The crystal structure of EtuB (of *Clostridium kluyveri* Etu metabolosome), a homologue of PduB, was previously solved in the lab. Like PduT, the subunit consists of a tandem BMC repeat and was expected to assemble into a trimeric structure (Heldt et al., 2009). The EtuB trimer has three pores formed within subunits, instead of a central pore formed between symmetry-related subunits. The trimeric shell protein was also reported to fit into sheets of EtuA hexamers modelled on the packing of CsoS1A hexamers, reflecting how the facet of the etu metabolosome may be formed.

The shell protein utilised in this study came from *Lactobacillus reuteri*, a probiotic bacterium able to colonise the gastrointestinal tract of a wide variety of animals

(Casas and Dobrogosz, 2000). The bacterium produces an antimicrobial agent called reuterin, a mixture of monomeric and dimeric forms of beta-hydroxypropionaldehyde (3-HPA) (Talarico and Dobrogosz, 1989), by glycerol cofermentation. The conversion of glycerol to 3-HPA is mitigated by the enzyme PduCDE (GupCDE), found within the Pdu metabolosome (Morita et al., 2008). The cobalamin-dependent diol dehydratase also carries out 1,2-propanediol metabolism. A schematic representation of glycerol and 1,2-propanediol metabolism of *Lactobacillus reuteri* pdu metabolosome is shown in Figure 4.1.

In this Chapter, the crystal structure of PduB was solved using molecular replacement method using EtuB as the search model. Three PduB subunits assemble into a trimer and as seen in the structure of EtuB, the PduB trimer has three subunit pores. However, and this is what makes this structure particularly interesting, three glycerols were found to be trapped in each subunit pore, revealing for the first time that these subunit pores are substrate channels.

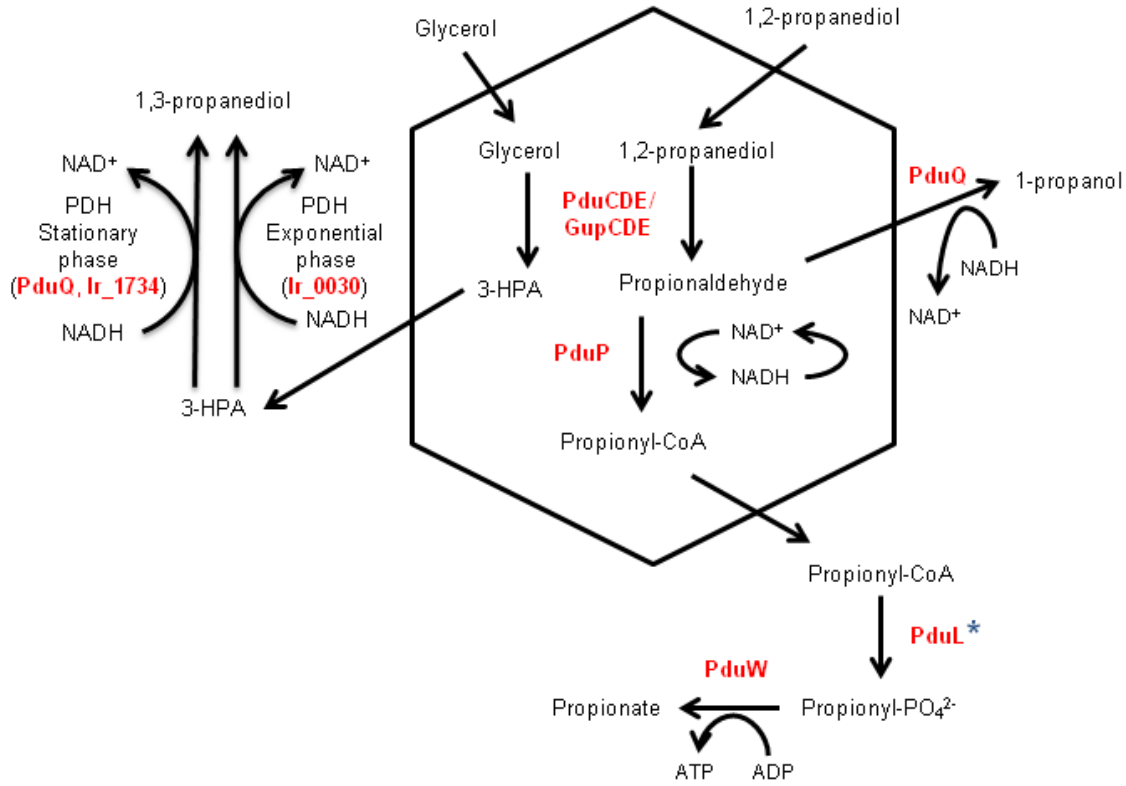


Figure 4.1 Schematic representation of the metabolic pathway of the *Lactobacillus reuteri* propanediol/glycerol utilisation metabolosome. Glycerol substrates are converted to 1.3-propanediol in a two-step reaction; first occurring within the microcompartment and second mediated by either PDH (propanediol dehydrogenase) stationary or exponential phase. 1,2-propanediol is subsequently converted to either 1-propanol or to 1-propionate with the generation of propionaldehyde intermediate occurring inside the microcompartment. PduL may be found within the microcompartment or closely associated to surface of the outer shell. Enzymes responsible for the conversion are highlighted in red.

4.2 Production, purification and crystallisation

The experimental details for the production of PduB were discussed previously in the Materials and Methods section (Chapter 2). Initial production and purification of the shell protein led to the conclusion that PduB was insoluble. Production of the shell protein induced via introduction of IPTG settled in the bacterial pellet rather than the lysate after lysis and centrifugation. The insolubility of the protein was supported by calculation using the Recombinant Protein Solubility Prediction server (Wilkinson and Harrison, 1991) and PROSO (Smialowski et al., 2007). Careful investigation of overexpression of the recombinant shell protein indicated that PduB is not entirely insoluble. A small quantity of protein was recovered from Nickel-charged sepharose beads.

4.2.1 Purification of PduB

In order to achieve a considerate amount of PduB suitable for crystallisation, *E. coli* containing plasmid of His-tagged PduB were grown in 6 L of LB media at 37 °C and induced at a low temperature of 16 °C overnight. As predicted, the majority of the shell protein PduB was found in the insoluble fraction (lane 3) rather than the soluble fraction (lane 4). However, some soluble PduB bound to the Nickel beads. Low concentration of imidazole (10 – 50 mM; lanes 6 and 7) washes out the contaminants. PduB then starts eluting out at 100 mM imidazole (lane 8) together with a number of contaminants but majority of the proteins eluted out at 400 mM imidazole (lane 9). Collected eluted PduB samples contain a couple of high molecular weight contaminants, one of which (~80 kDa) appeared to be dominating

the shell protein of interest (28 kDa). This is represented in the acrylamide gel as shown in Figure 4.2A.

PduB was further purified via size exclusion chromatography. Samples of the peak were analysed in the polyacrylamide gel. As opposed to the eluted sample from Nickel column, the band for the shell protein of interest appears to be more intense compared to the high molecular weight contaminant (Figure 4.2B). Oddly, the contaminant (~80 kDa) is about three times the size of PduB (~28 kDa). Initially, it was thought that this might not be a contaminant, but instead a trimeric form of the shell protein. However, this is unlikely as samples were boiled with reducing agent before running in the polyacrylamide gel containing SDS, which are sufficient to destabilise the majority of proteins. In any case, the sample is 90 – 95% pure, sufficient for crystallisation trials. A final concentration of 7 mg/ml of PduB was obtained from 6L of LB media, enabling to obtain about 1 mg/ml PduB for every liter grown.

It was previously reported that the expression of recombinant PduB shell proteins (*S. enterica* and *C. freundii*) produces two forms, termed as PduB and PduB' (Parsons et al., 2008, Havemann and Bobik, 2003). PduB' has a shorter length lacking a few residues in the N-terminal region. The two protein species arise due to the presence of two translation start sites on the mRNA. The occurrence of two forms was not detected with EtuB, which indicate that this might only exist in PduB. Contrary to the reported two forms, only a single form of PduB (full length) was obtained after purification. However, a careful investigation and analysis of the polyacrylamide gel, a smaller protein fragment can be seen eluting out at lower imidazole range (50 –

100 mM) (Figure 4.2A). It is possible that this is the shorter form of PduB. The N-terminal His-tagged PduB (full length) binds significantly tighter to the Nickel column compared to the shorter form of PduB (lacking N-terminal region of full length, which includes the His-tag).

The second starting codon (methionine) for the shorter PduB form seems to be a conserved residue in many species (*Klebsiella*, *Citrobacter*, *Escherichia*, *Listeria*, and *Salmonella*; available FASTA sequence in NCBI for *Salmonella* is PduB' but has been reported to have the full length form that starts with MSSNEL) (Figure 4.3). For *Lactobacillus*, the presumed second translational start site is located in residue 26, where 'M' aligns with 'L' of other species. PduB shell proteins of *Propionibacterium* and *Rhodobacter* do not share the second conserved methionine residue, indicating that these shell proteins are not able to make two different forms. The significance of the generation of two forms of PduB is unknown.

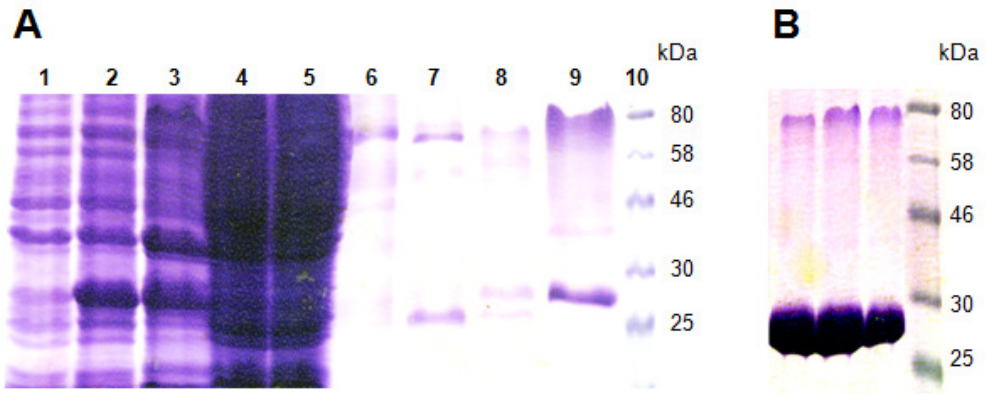


Figure 4.2 Purification of PduB. **(A)** Via nickel column, lane 1: pre-induction, lane 2: post-induction, lane 3: insoluble fraction, 4: clear lysate, 5: flowthrough, 6-8: increasing imidazole concentration (from 10 – 100 mM), 9: elution of His-tagged PduB with 400 mM imidazole. **(B)** Via size exclusion chromatography, purified PduB protein fractions.

```
Klebsiella          -MSSNELVEQIMAQVIARVATPEQQALPDTPHPTGETAMAEKSCSLTEFVGT-AIGDTLG 58
Citrobacter         -MSSNELVDQIMAQVIARVATPEQQAIPENNPPTRETAMAEKSCSLTEFVGT-AIGDTVG 58
Escherichia         -MSSNELVEQIMAQVIARVATPEHEVLPKTTHQIRETAMAEKSCTLTEFVGT-AIGDTVG 58
Salmonella          ------------------------------------MAEKSCSLTEFVGT-AIGDTLG 21
Listeria            ----MSEQNKLIDEILKQVMETVGNDPEKLVEDGGSRQMSEKAVQLTEFVGT-AIGDTIG 55
Propionibacterium   ---MANEQLVDQIMSAVMAKIGNEPASAAPRAAAPSVPASIGRPEATEFVGTSALGDTIG 57
Rhodobacter         ----MKDDLVEKLMDEVMRKMGSTEAPAAAAAEKPAAKAAPKACNLTEFVGT-AIGHTVG 55
Lactobacillus       ----------------------------------MNDFLNSTSTVPEFVGASEIGDTIG 25
                                                              .****:  :*.*:*

Klebsiella          LVIANVDAALLEAMKLEKRYRSIGILGARTGAGPHIMAADEAVKATNTEVVSIELPRDTK 118
Citrobacter         LVIANVDSALLDAMKLEKRYRSIGILGARTGAGPHIMAADEAVKATNTEVVSIELPRDTK 118
Escherichia         LVIANVDSALLDAMKLEKRYRSIGILGARIGAGPHIMAADEAVKATNTEVISIELPRDTK 118
Salmonella          LVIANVDTALLDAMKLEKRYRSIGILAARTGAGPHIMAADEAVKATNTEVVSIELPRDTK 81
Listeria            LVIANVDGQLLEAMKLEKSYRSIGILGARTGAGPHIMAADEAVKATNTEVVKIELPRDTK 115
Propionibacterium   LVIPNVDPQIHALMKIDPKYRSIGIIGDRTGAGPHIFAADEGVKATNTEVVSIELARDTK 117
Rhodobacter         LVIANVDPLLHEKMNIDRKYRSIGIVGARTGAGPHIFAADEAVKATNSEVVLIELPRDTE 115
Lactobacillus       MVIPRVDQQLLDKLHVTKQYKTLGILSDRTGAGPQIMAMDEGIKATNMECIDVEWPRDTK 85
                    :**..**  :   :::   *:::**:. * ****:*:* **.:**** * : :* .***:
```

Figure 4.3 Sequence alignment of the N-terminal region of PduB homologues of different species. The second starting codon (methionine) is highlighted in species *Klebsiella*, *Citrobacter*, *Escherichia*, *Listeria* and *Lactobacillus*. For *Salmonella*, the putative second starting codon aligns with the first residue (M) due to absence of short N-terminal region available in NCBI. Sequence alignment was produced using ClustalW.

4.2.2 Crystallisation of PduB

1 µl of 7 mg/ml of PduB was mixed with 1 µl of the corresponding reservoir of Hampton Screen I and II using the hanging drop method. Five conditions from these yielded crystals after 3 days of equilibration. A list of the screening conditions positive for crystals can be found in Table 4.1. These conditions were further optimised in 24-well plate by varying the ratio of buffer pH, salt and precipitant concentration as well as the ratio of protein to the crystallisation condition.

Three screens (I – 7, 21 and 47) yielded no crystals except for the original condition. For the remaining screening conditions (II – 32 and 33), crystallisation is favoured with higher precipitant concentration (1.4 – 1.8 M Ammonium sulfate and 1.8 – 2.2 M Ammonium formate pH 4 – 4.6, respectively). Like PduT, PduB crystals present as hexagonal plates with almost similar size (60 – 80 µm). A few crystals in Screen II – 32 appear as half hexagons. The two different shapes of PduB crystals are shown in Figure 4.4.

Table 4.1 Crystallisation conditions where crystals were seen in the initial screen.

Screen	Buffer	Salt	Precipitant
I – 7	0.1 M Na cacodylate pH 6.5	-	1.4 M Na acetate
I – 21	0.1 M Na cacodylate pH 6.5	0.2 M Mg acetate	30% MPD
I – 47	0.1 M Na acetate pH 4.6	-	2 M NH_4 sulfate
II – 32	0.1 M HEPES pH 7.5	0.1 M NaCl	1.6 M NH_4 sulfate
II – 33	0.1 M HEPES pH 7.5	-	2 M NH_4 formate

A **B**

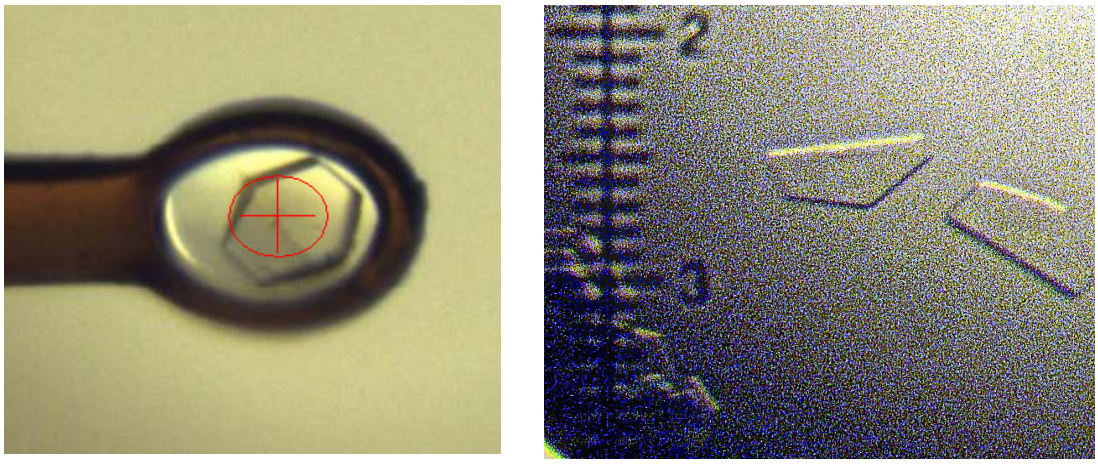


Figure 4.4 PduB crystals in two different shapes. **(A)** Hexagonal plate crystal in a loop ready to be illuminated by X-ray beam. **(B)** Half hexagonal plates. Crystal size is about 80 µm x 80 µm for A and 40 µm x 80 µm for B.

4.3 Crystal screening

Several crystals of PduB were screened to identify the best dataset to work with. PduB crystals suffer from many different problems. These include anisotropic diffraction, lattice disorder and twinning. These problems had been reported on many occasions in the literature involving hexagonal shaped crystals, specifically microcompartment shell proteins (Sagermann et al., 2009, Kerfeld et al., 2005, Tsai et al., 2009, Pang et al., 2011, Tanaka et al., 2009).

4.3.1 Anisotropic diffraction

Many PduB crystals show an anisotropic distribution of diffraction pattern. Diffraction anisotropy occurs when the quality of diffraction depends on the direction of the crystal. In a typical anisotropic diffraction pattern, the resolution of the diffraction is higher in one direction (for instance, horizontal) than in the opposite (vertical) direction. PduB crystals typically suffered from this problem.

4.3.2 Twinning

The collected data for crystals which do not appear to diffract anisotropically were successfully integrated. Twinning was identified after scaling of the integrated MTZ file. This was subsequently confirmed by CTRUNCATE and SFCHECK (Vaguine et al., 1999). Statistics for twinning tests are summarised in Table 4.2. The data were first integrated in the space group of P3, as suggested by MOSFLM, then later in a higher symmetry space group, $P6_3$. In both cases, the L-test (Padilla and Yeates, 2003) and perfect twinning test ($\langle I^2 \rangle / \langle I \rangle^2$) suggest that the crystal is twinned, but not a perfect twin (L-test: untwinned = 0.5; perfect twin = 0.375 and $\langle I^2 \rangle / \langle I \rangle^2$

untwinned = 2; perfect twin = 1.5). The PduB crystal reported here exhibited merohedral twinning. There are different types of twinning according to the number of dimensions that are coincident between the twinned lattice (Yeates and Fam, 1999).

Table 4.2 Statistical data of PduB dataset processed in the space groups P3 and $P6_3$.

Space Group	P3	$P6_3$
Matthews coefficient (Vm)	3.76	3.76
Solvent content	67.3%	67.3%
Asymmetric unit	2	1
L-test statistics	0.413	0.398
Perfect twinning test ($<I^2>/<I>^2$)	1.8293	1.8061
Twinning operators	-h, -k, l h, -h-k, -l -k, -h, -l	h, -h-k, -l

4.3.3 Lattice disorder

While it is possible to produce a reasonable model of PduB in either space groups P3 or $P6_3$, the data cannot be refined to a model with acceptable stereochemistry even with application of twin operators. This raises the suspicion that there could be more severe pathological problems that cannot be solved by simple twinning refinement. To identify this problem, the reflection images were carefully investigated (Figure 4.5). The first few images show numerous intense diffraction spots, indicating mosaicity is high. When the crystal is turned through 90 degrees, it can be noticed that the diffraction spots becomes smeary, indicative of lattice disorder, which could have contributed to the high mosaic pattern seen in the first few images. Lattice disorder can be thought of as an extreme case of twin packing, where the twinning occurs on a microscopic rather than macroscopic level.

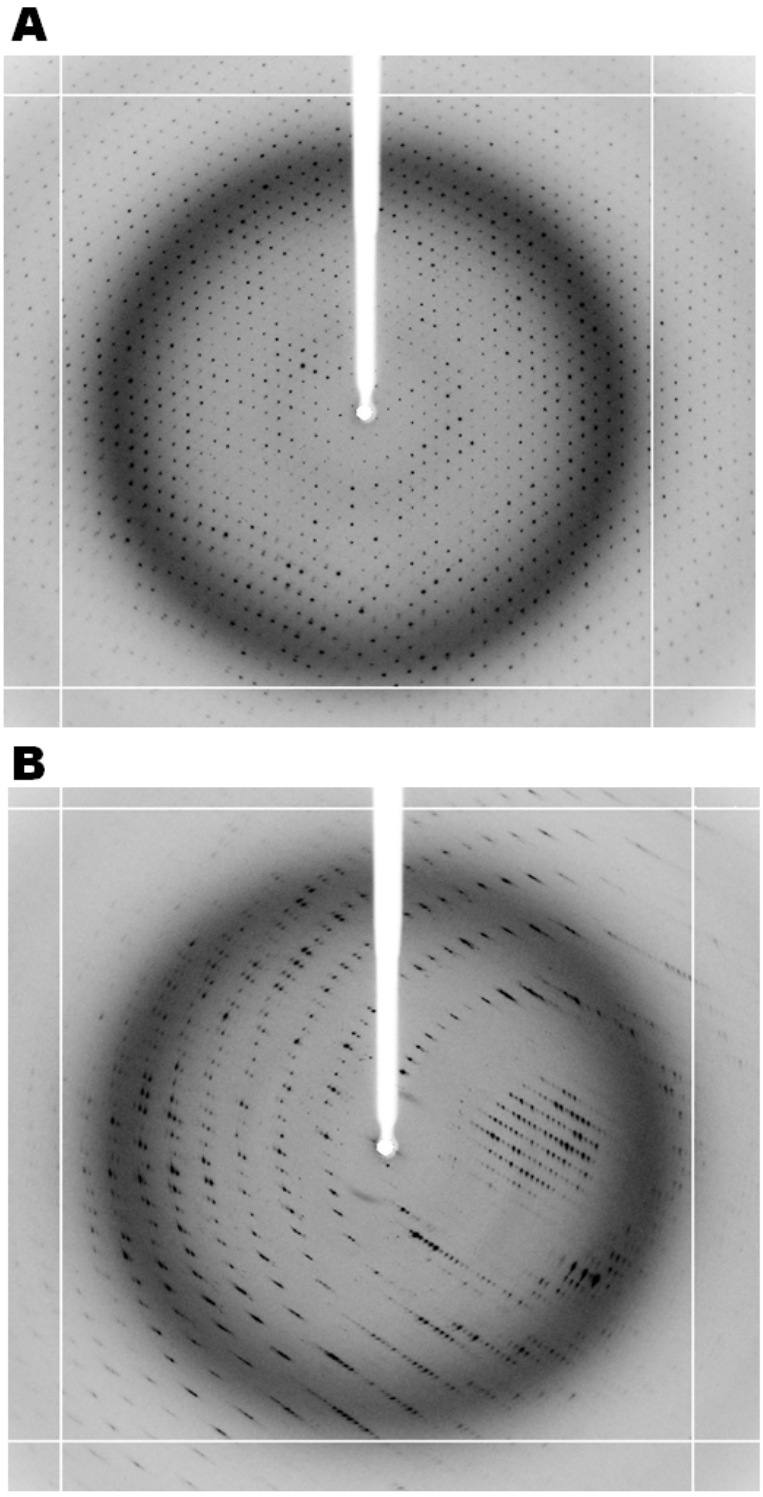


Figure 4.5 Evidence of crystal problems: **(A)** high mosaicity and **(B)** lattice disorder

4.4 Obtaining PduB structure

One of the screened hexagonal crystals diffracted to a high resolution of 1.5Å and showed no evidence of lattice disorder. This data was used in molecular replacement with EtuB (*Clostridium kluyveri*) as the search model. EtuB is an ideal template since EtuB shares 45% sequence identity with PduB (*Lactobacillus reuteri*).

4.4.1 Data collection

A single hexagonal-shaped PduB crystal grown in 0.1 M sodium cacodylate pH 6.5 and 1.4 M sodium acetate was picked up with 0.1 mm litholoop (Figure 4.4A). The crystal was briefly dipped into a reservoir solution supplemented with 20% glycerol prior to flash freeze with liquid nitrogen. This crystal was illuminated with synchrotron X-ray beam at beamline I04-1 in Diamond Light Source (Oxfordshire, UK) and data collection took place at 100K, at a fixed wavelength of 0.9163 Å, using a rapid read-out Marmosaic 300 mm CCD detector to record diffraction images. A total of 1200 images were recorded, over a rotation range of 120°, with a 0.1° rotation angle per image. Diffraction images taken at 0° and 90° are shown in Figure 4.6, notice the significant improvement of diffraction quality compared to Figure 4.5.

The data were indexed using XDSME and scaled using SCALA. The crystal diffracted to 1.55 Å resolution and has a unit cell dimension of a = 69.86 Å, b = 120.86 Å, c = 145.53 Å and a space group of $C222_1$. PduB crystallises with 3 monomers per asymmetric unit, giving a Vm of 1.9 and a solvent content of 36%.

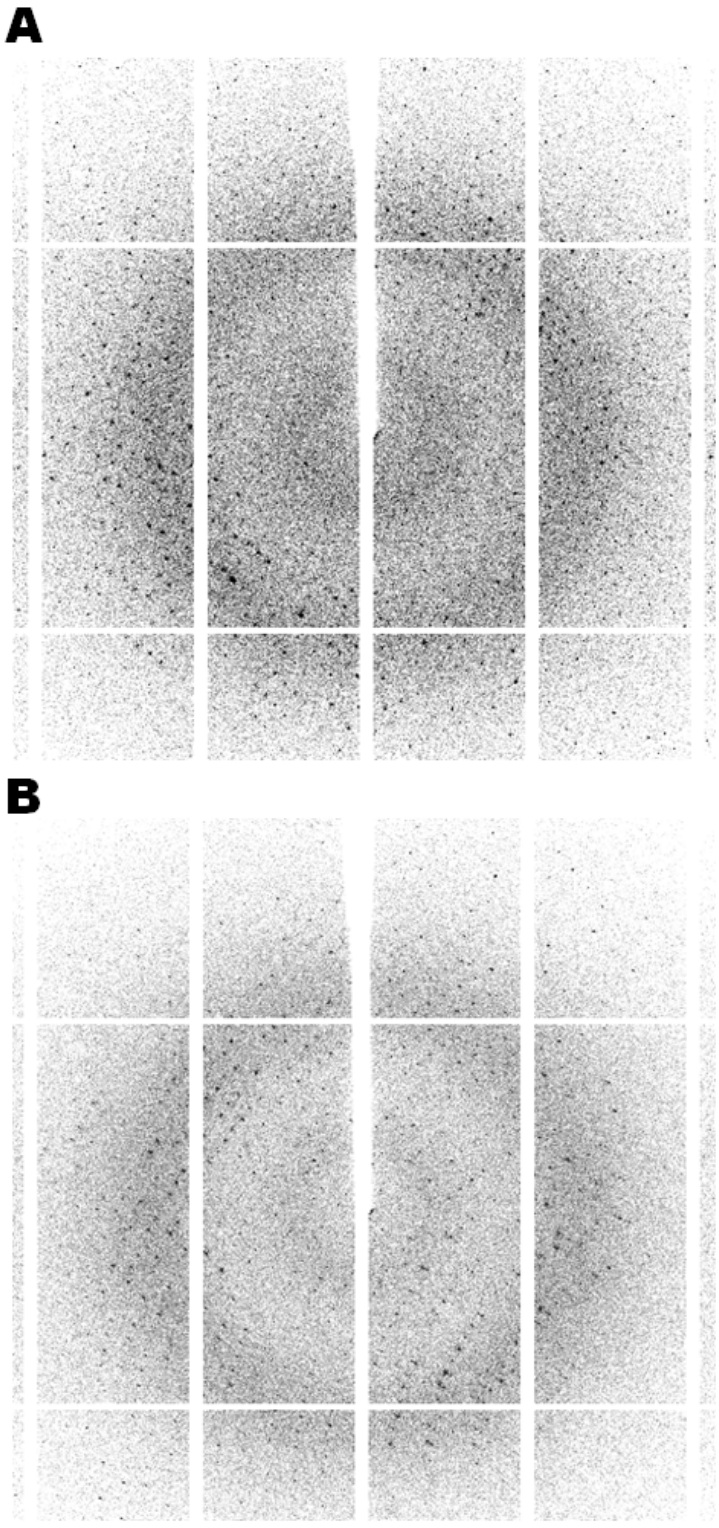


Figure 4.6 Diffraction pattern from PduB crystals at **(A)** 0° and **(B)** 90°. The edge of the image corresponds to 1.5 Å resolution.

4.4.2 Molecular replacement, model building & refinement

The AutoMR and AutoBuild wizard from PHENIX produced a PduB model using EtuB (PDB ID: 3IO0) as the template. The model was composed of 637 residues in six fragments with 497 waters, giving an R-factor, R-free and a correlation coefficient of 0.25, 0.28 and 0.51, respectively. The six fragments were rebuilt and joined manually to produce three fragments with each of them corresponding to one PduB subunit. Seventeen glycerol molecules were added to the model. After rebuilding and refinement, the stereochemistry and validation statistics of the final PduB model is listed in Table 4.3.

The final model of PduB has clearly defined polypeptide backbone in the electron density map for residues 11–237. Residues preceding Val-11 and the C-terminal Lys-238 are poorly defined in the electron density map, presumably because they are flexible or statically disordered in the crystal. The side chains have clearly defined electron density, with the exception of four residues: Glu-19, Arg-30, Lys-171 and Arg-208. Glu-19 and Arg-30 are found in the loops of beta-turn hairpins (β1/β2 and β2/β3) while Lys-171 and Arg-208 are in the alpha helices (α5 and α6). The side chains of these residues are mobile as they are on the surface of the structure. A total of twenty residues exhibit two different conformations. One residue, Asp-83 falls in the disallowed region of the Ramachandran plot, but the electron density map has clear carbonyl-bulges suggesting the main-chain is somewhat distorted for this residue (Figure 4.7A).

Structure solution via molecular replacement method can be susceptible to search model bias. To test that the solution produced by this method is not biased, three

residues of PduB, based on ClustalW sequence alignment with significant difference to the search model were selected and mutated to the residues of EtuB. The three residues of PduB are Cys-75, Ile-76 and Asp-77; the corresponding residues for EtuB are Val-67, Ala-68 and Thr-69. Residues for PduB showed a good fit for electron density map (Figure 4.7B). This is contrary to the electron density map fitted with the corresponding residues of EtuB (Figure 4.7C). The side chain for Val-67 cannot be accommodated by the electron density map, whereas the other two residues, Ala-68 and Thr-69 lack the side chain to fit in the electron density map.

Eighteen ligands were found in the final trimeric structure of PduB. Seventeen of these are glycerols while the remaining ligand is an acetate ion. Chain A binds six glycerol molecules, while chain B binds six glycerols and an acetate ion, and four glycerols can be found in chain C. A further glycerol molecule is found at the centre of the trimeric structure. The differing number of ligands for each chain can be attributed to the different environments in the crystal. However, each subunit pore is occupied by three glycerol molecules binding at the same locations within the pore.

Table 4.3 Crystallographic data and refinement statistics for PduB

Data Collection	
Space group	$C222_1$
a, *b*, *c* (Å)	*a* = 69.865, *b* = 120.864, *c* = 145.531
α, β, γ (°)	$\alpha = \beta = \gamma = 90$
Molecular mass in Da (residues)	24945.3
Monomers per asymmetric unit	3
Wavelength (Å)	0.9163
Resolution (Å)	46.51 – 1.64 (1.64 – 1.55) [a]
Number of unique reflections	87436 (12243) [a]
Multiplicity	4.5 (4.4) [a]
Completeness (%)	99 (96.2) [a]
Rsym ♣	0.088 (0.908) [a]
Mean I / sigma (I)	11.3 (2.0) [a]
Rpim ♦	0.045 (0.483) [a]
Rmeas ♠	0.099 (1.035) [a]
Wilson B-factor	16.004
Refinement	
Resolution (Å)	46.51 – 1.55
Reflections (work / test)	83762 / 4338
R-factor / R-free ♥	0.1762 / 0.2045
RMSD bond (Å) / angle (°)	0.0198 / 1.7151
Ramachandran plot statistics	
Residues in most favoured regions	94.35%
Residues in allowed regions	5.17%
Residues in disallowed regions	0.48%

[a] The parameter values for the range 1.55 – 1.64 Å are given in parentheses

♣ $R_{sym} = \Sigma_{hkl} \Sigma_i | I_i - \langle I \rangle | / \Sigma_{hkl} \Sigma I_i$, where I_i is the intensity of the *i*th observation, $\langle I \rangle$ is the mean intensity of the reflection and the summations extend over all unique reflections (*hkl*) and all equivalents (*i*), respectively.

♦ R_{pim} = measure of the quality of the data after averaging the multiple measurements (Evans, 2006).

♠ R_{meas} = improved version of the traditional R_{merge} (R_{sym}) and measures how well the different observations agree (Evans, 2006).

♥ R-factor = $\Sigma_{hkl} |F_o - F_c| / \Sigma_{hkl} F_o$, where F_o and F_c represent the observed and calculated structure factors, respectively. The R-Factor is calculated using the 95% of the data that were included in refinement and R-free is calculated using the excluded 5%.

The values presented in this table are from SCALA, REFMAC and PROCHECK from the CCP4 suite.

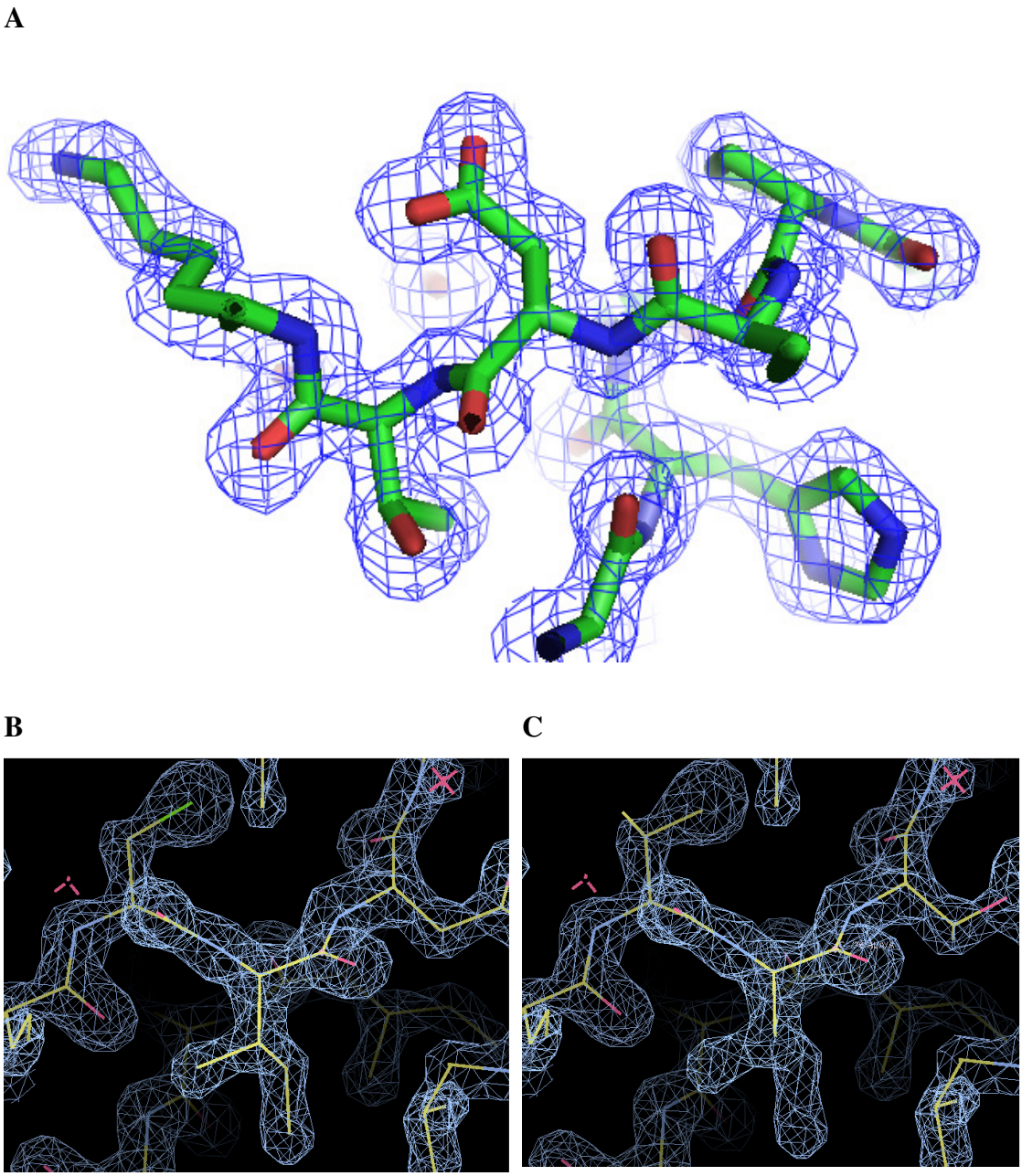


Figure 4.7 Electron density fitting of PduB structure. **(A)** σA-weighted 2Fobs-Fcalc Fourier synthesis contoured around Asp-83 at 1σ showing the quality of the electron density map (blue chicken wire mesh). Model bias test on PduB structure. Electron density map fitted with **(B)** residues of PduB and **(C)** corresponding residues of EtuB.

4.5 Tandem repeat of BMC fold

To date, four shell proteins: CsoS1D (Klein et al., 2009), EtuB (Heldt et al., 2009), EutL (Takenoya et al., 2010) and PduT (Pang et al., 2011, Crowley et al., 2010) are known to contain tandem BMC repeats in each subunit. The crystal structure of PduB confirms that this shell protein also consists of tandem BMC repeat within its subunit (Figure 4.8). When the PduB subunit is superimposed with these shell proteins, this reveals that it is closely similar to EtuB (Z-score: 34.2; 224 aligned residues; RMSD 1.2; 44% identity), and EutL (Z-score 20.8; 193 aligned residues; RMSD 2.3; 19% identity). PduT (Z-score: 8.2; 65 aligned residues; RMSD 1.9; 29% identity) and CsoS1D (Z-score 9.8; 92 aligned residues; RMSD 2.5; 18% identity) were predicted to be probable homologous by DaliLite. This structural difference may reflect the functional difference of CsoS1D and PduT to PduB, EtuB and EutL. The latter three structures have pore in each subunit, while CsoS1D and PduT do not. In addition to this, the arrangement of secondary structure elements in the BMC domain of PduT is different to permuted BMC domains of PduB (Figure 4.8).

The first and second BMC domains of PduB are composed of 109 and 118 residues, respectively; both have 8 secondary structure elements. The two BMC domains share 15% sequence identity (Figure 4.9A) and superimpose with an RMSD of 2.7 Å with 91 aligned residues and Z-score of 10 (Figure 4.9B). The most obvious difference is in the loop regions connecting β4 – β5 and β9 – β10; these loops fill

the central cavity around the three fold axis (Figure 4.9C). These loops sandwich and bind to the central glycerol molecule (which would be discussed later).

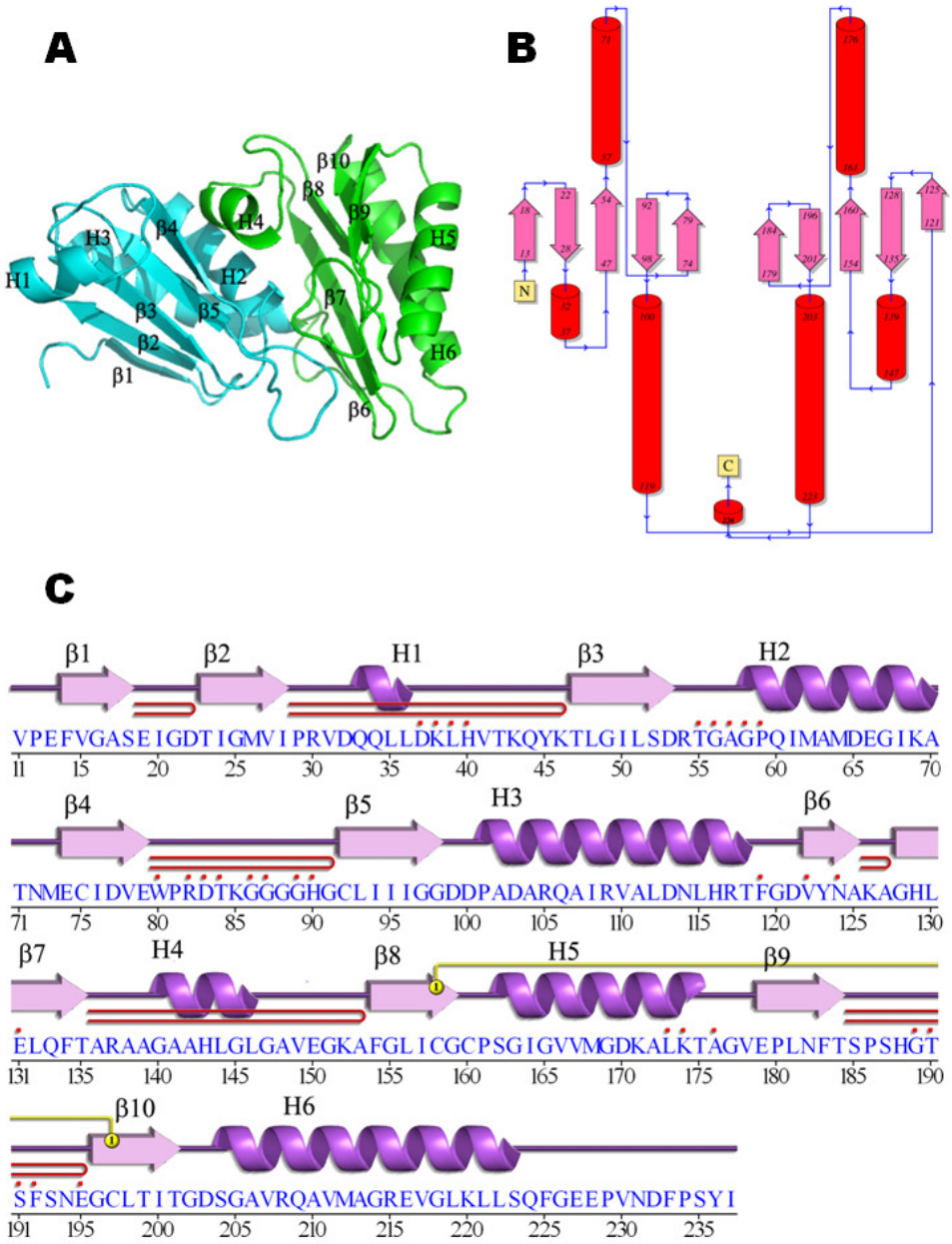


Figure 4.8 The tertiary structure of the PduB subunit. **(A)** Cartoon representation of the tertiary structure of PduB subunit which comprises two BMC repeats. **(B)** Schematic drawing of the topology of the PduB subunit. **(C)** Amino acid sequence of the PduB with secondary structures marked. Red dots represent ligand binding residues. A disulfide bridge can be formed between Cys-158 and Cys-197. This figure was produced using PDBsum (Laskowski et al., 1997).

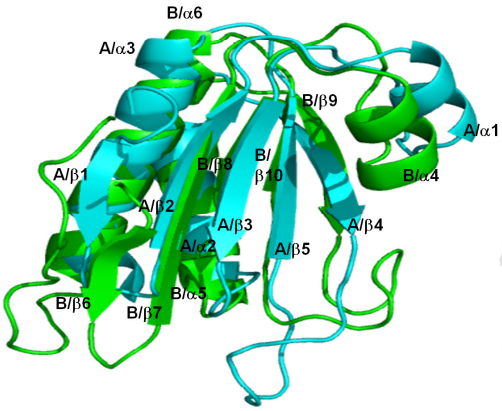
A

```
PduB_BMC1    01 MNDFLNSTSTVPEFVGASEIGDTIGMVIPRVDQQLLDKLHVTKQYKTLGILSDRTGAGPQ 60
PduB_BMC2   120 -GDVYNAKAGHLELQFTARAAGAAHLGLGAVEG-----------KAFGLICGCP-SGIG 165
                 .*. *:.:   *:  ::. ..:  : :  *:            *::*::.. . :*

PduB_BMC1    61 IMAMDEGIKATNMECIDVEWPRDTKGGGGHGCLIIIGGDDPADARQAIRVALDN---LHR 117
PduB_BMC2   166 VVMGDKALKTAGVEPLNFTSPSHGTSFSNEGCLTITG--DSGAVRQAVMAGREVGLKLLS 223
                ::  *:.:*::.:* ::.  * . .. ...*** * *  *.. .***: .. :    *

PduB_BMC1   118 TF------------- 119
PduB_BMC2   224 QFGEEPVNDFPSYIK 238
                 *
```

B



C

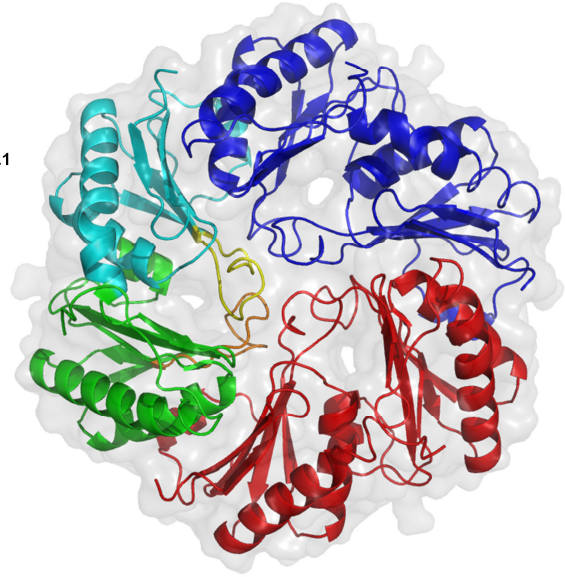


Figure 4.9 Comparison of the first and second BMC repeats of the PduB subunit. **(A)** Sequence alignment of the first and second BMC domain of PduB. Alignment generated by ClustalW. **(B)** Superimposition of the first (cyan) and second BMC domain (green) of PduB. **(C)** Three PduB subunits assemble into a trimer. BMC domains are shown in different colours (blue and red); the tandem BMC domains are coloured in cyan (BMC domain 1) and green (BMC domain 2) for one of the protomers. The distinct loops are coloured yellow (pointing down to the convex side) and orange (pointing away from the convex side).

4.6 Trimeric structure

PduB assembles into a trimer with hexameric appearance, a common theme for shell proteins (CsoS1D, EutL, EtuB and PduT) with two BMC repeats per subunit. One face of the PduB trimer is concave and the other convex. PduB exhibits similar electrostatic potential to EtuB, with minor differences: in the convex face, the highly negative potential section is sandwiched between the positive potential close the three-fold axis and on the periphery and in the concave face the patches of positive potential are much weaker compared to those of EtuB (Fig 4.10). When PduB trimer is viewed from the side, the bowl-shaped cavity appears to be deeper than in PduA and PduB (Figure 4.11A). The cavity of PduU is deeper than PduB due to the presence of six tightly wound β-strands around the hexamer pore (Crowley et al., 2008).

4.6.1 Molecular tiling of PduB

The crystal structure of PduB reveals a higher order packing of trimers into uniform molecular sheets. Similar higher order packing has been observed in other Pdu shell protein crystals including: PduA and PduT (Crowley et al., 2010), but not PduU (Crowley et al., 2008) or EtuB (Heldt et al., 2009). Six conserved lysines are found within the PduB trimer that corresponds to the lysines, which are suggested to maintain the hexameric PduA within a molecular sheet. Although similar lysine residues can be found within the molecular sheets of PduB trimer, the residues do not perfectly align as in sheets of PduA hexamers and the packing is looser (Figure 4.11B). Formation of layers of PduB trimers in the crystal packing closely

resembles to that of PduT trimers; the distance between neighbouring lysines are about 7-8 Å, compare to PduA hexamer, which are tighter (~3 Å). When the PduB trimer is placed into a sheet of PduA hexamers, the trimer fits very well, with the six conserved lysine residues facing anti-parallel to their corresponding adjacent lysine residues of PduA hexamers (Figure 4.11C). This observation is consistent to the biochemical studies suggesting that PduA interacts with PduB (Parsons et al., 2010a), and therefore suggests that the PduB trimer can pack successfully within a layer of PduA molecules.

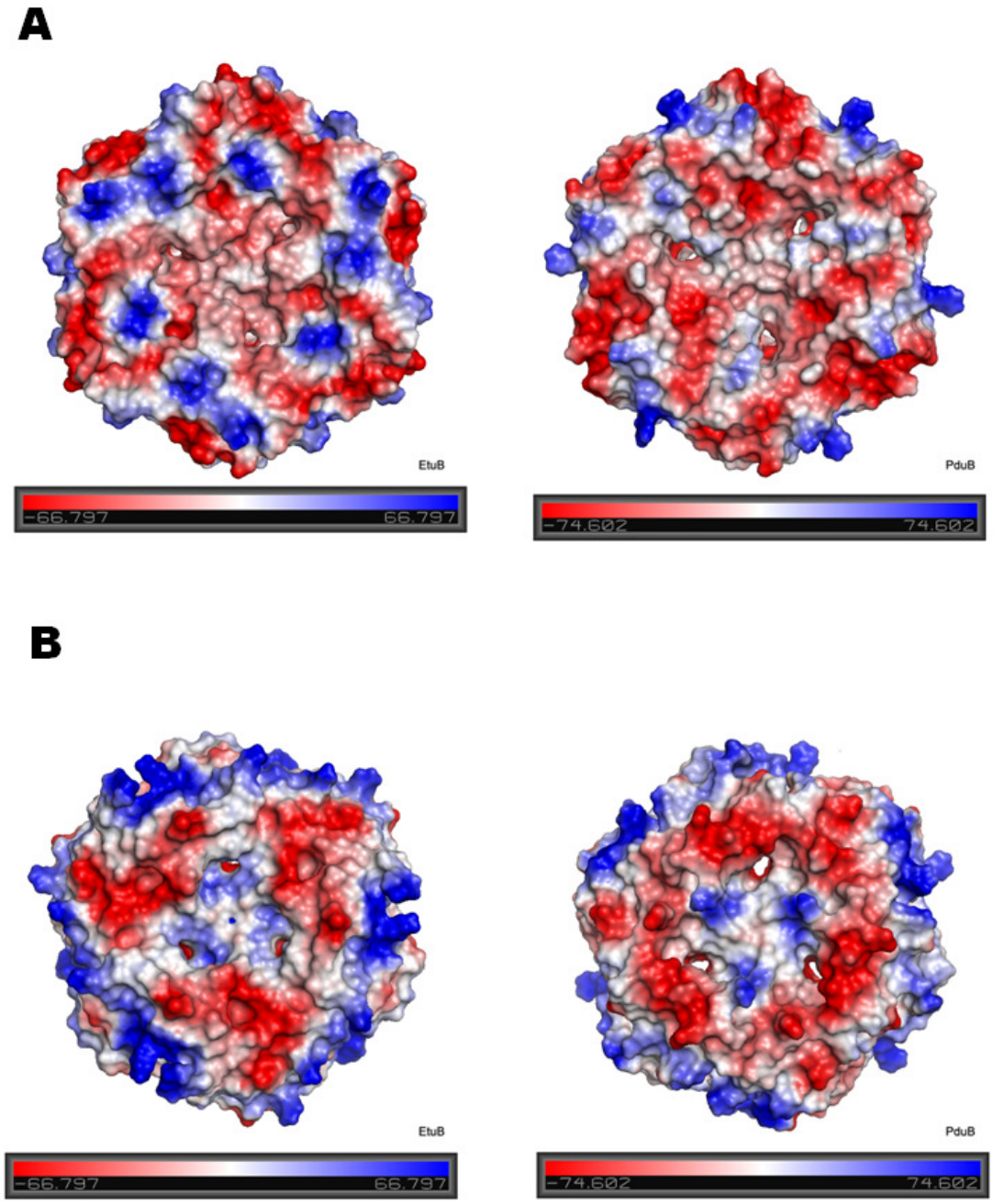


Figure 4.10 Electrostatic potential of EtuB and PduB. **(A)** Concave and **(B)** Convex surfaces of EtuB and PduB. Blue and red patches indicate positive and negative electrostatic potential, respectively. Figure was produced using vacuum electrostatic by PyMOL (DeLano and Lam, 2005).

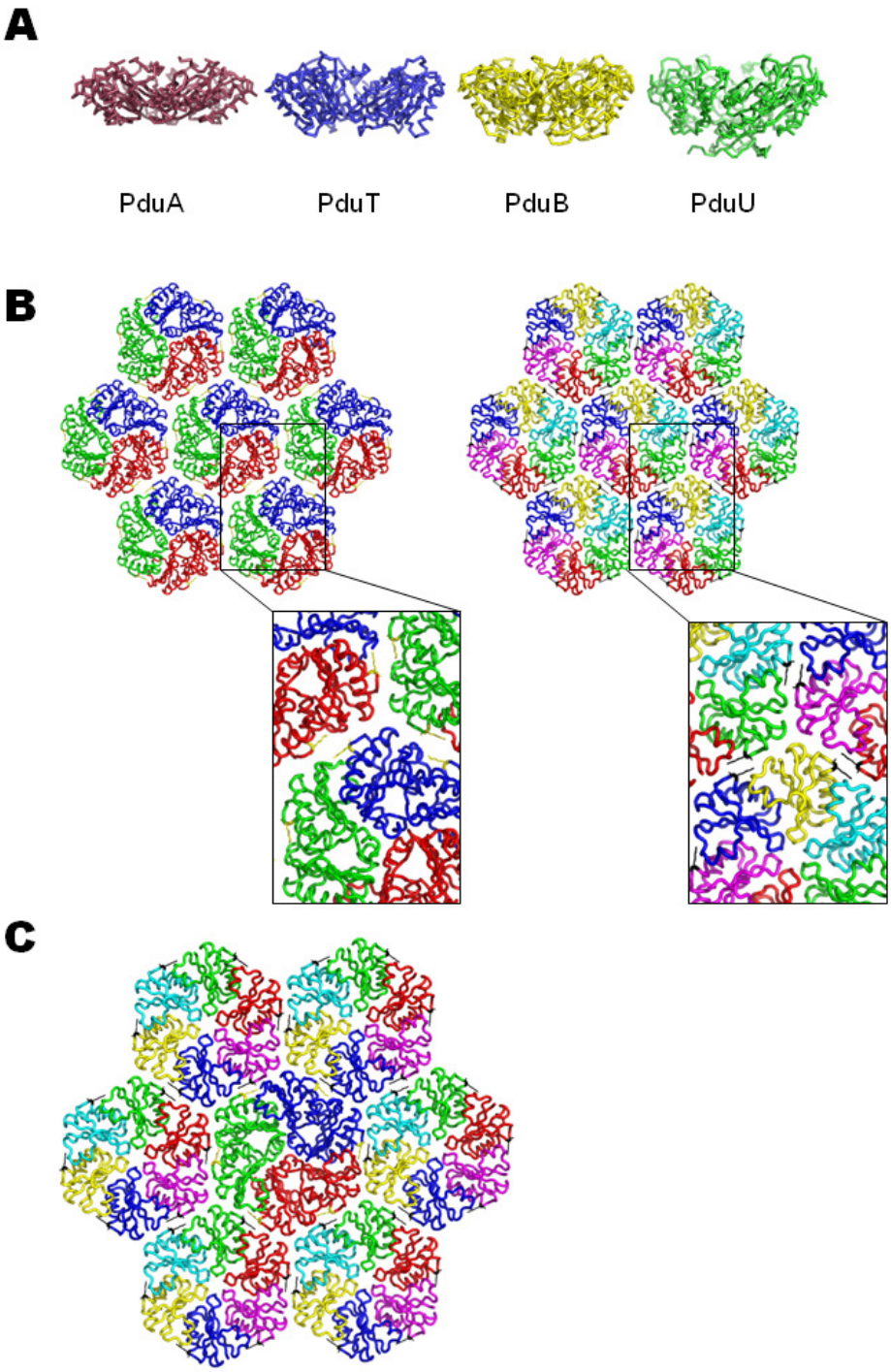


Figure 4.11 Comparison of quaternary and higher order structures involving PduB. **(A)** Side view showing the depth of cavity of various pdu shell proteins, **(B)** Crystal packing of PduB trimers (left) and PduA hexamers (right), **(C)** PduB inserted into a sheet of PduA molecules. Both PduB and PduA showed presence of conserved lysines (highlighted as yellow sticks for PduB and black sticks for PduA). This Figure was produced using PyMOL (DeLano and Lam, 2005).

4.6.2 Subunit pores as glycerol channels

Majority of the shell proteins solved to date have a central pore on the six-fold axis of the hexamer. In contrast, the trimeric shell proteins (EtuB and EutL) have pores within the subunit, a situation facilitated by the gene duplication that has given rise to the tandem repeat of the trimeric shell protein. The crystal structure of PduB reveals that this trimer has similar subunit pore characteristics.

4.6.2.1 Pore Size

Superimposition of PduB, EtuB and EutL allows comparison of subunit pores. It can be surmised that PduB has the widest pore, followed by EtuB and EutL. To validate this, Computed Atlas of Surface Topography of proteins or CASTp server (Dundas et al., 2006) was used to identify the pore size of PduB, EtuB and EutL subunits. The reported surface area of subunit pore was as follows: PduB – 457.9 $Å^2$, EtuB – 376 $Å^2$, and EutL – 319.7 $Å^2$, consistent to the superimposition observation.

The pores of EtuB are lined with three histidine residues (at positions 156, 195 and 224) and two glutamate residues (at positions 197 and 262) (Heldt et al., 2009). Similar residues are found in PduB with two exceptions. His-224 is substituted by Cys-158, which forms a disulphide bridge with Cys-197 (Leu-264 in EtuB) and slightly widens the channel. Interestingly, the two cysteine residues are not conserved in PduB of other species (*Citrobacter* and *Salmonella*), which have valine and isoleucine instead (Figure 12A). In addition to this, Pro-124 (EtuB) is substituted by Gly-58 (PduB) which also slightly widens the pore in PduB. Subtle widening of the channel may allow glycerol access to the channel (Figure 4.12B).

A

```
PduB_Salmonella      RTGAGPHIMAADEAVKATNTEVVSIELPRDTKGGAGHGSLIILGGNDVSDVKRGIEVALK 109
PduB_Citrobacter     RTGAGPHIMAADEAVKATNTEVVSIELPRDTKGGAGHGSLIILGGNDVSDVKRGIEVALK 146
PduB_Lactobacillus   RTGAGPQIMAMDEGIKATNMECIDVEWPRDTKGGGGHGCLIIIGGDDPADARQAIRVALD 113
EtuB_Clostridium     RVGAPAQMMAADEAVKGTNTEVATIELPRDTKGGAGHGIFIVLKAADVSDARRAVEIALK 179
                     *.** .::** **.:*.** *   :* *******.*** :*:: . * :*.::.:.:**.

PduB_Salmonella      ELDRTFGDVYGNEAGHIELQYTARASYALEKAFGAPIGRACGIIVGAPASVGVLMADTAL 169
PduB_Citrobacter     ELDRTFGDVYANEAGHIEMQYTARASYALEKAFGAPIGRACGVIVGAPASVGVLMADTAL 206
PduB_Lactobacillus   NLHRTFGDVYNAKAGHLELQFTARAAGAAHLGLGAVEGKAFGLICGCPSGIGVVMGDKAL 173
EtuB_Clostridium     QTDKYLGNVYLCDAGHLEVQYTARASLIFEKAFGAPSGQAFGIMHAAPAGVGMIVADTAL 239
                     : .: :*:**  .***:*:*:****:   . .:**  *:* *:: ..*:.:*:::.*.**

PduB_Salmonella      KSANVEVVAYSSPAHGT-SFSNEAILVISGDSGAVRQAVTSAREIGKTVLATLGSEPK-N 227
PduB_Citrobacter     KSANVEVVAYSSPAHGT-SFSNEAILVISGDSGAVRQAVISAREIGKTVLGTLGSEPK-N 264
PduB_Lactobacillus   KTAGVEPLNFTSPSHGT-SFSNEGCLTITGDSGAVRQAVMAGREVGLKLLSQFGEEPV-N 231
EtuB_Clostridium     KTADVKLITYGSPTNGVLSYTNEILITISGDSGAVLQSLTAARKAGLSILRSMGQDPVSM 299
                     *:*.*: : : **::*. *::** :.*:****** *:: :.*: * .:*  :*.:*
```

B

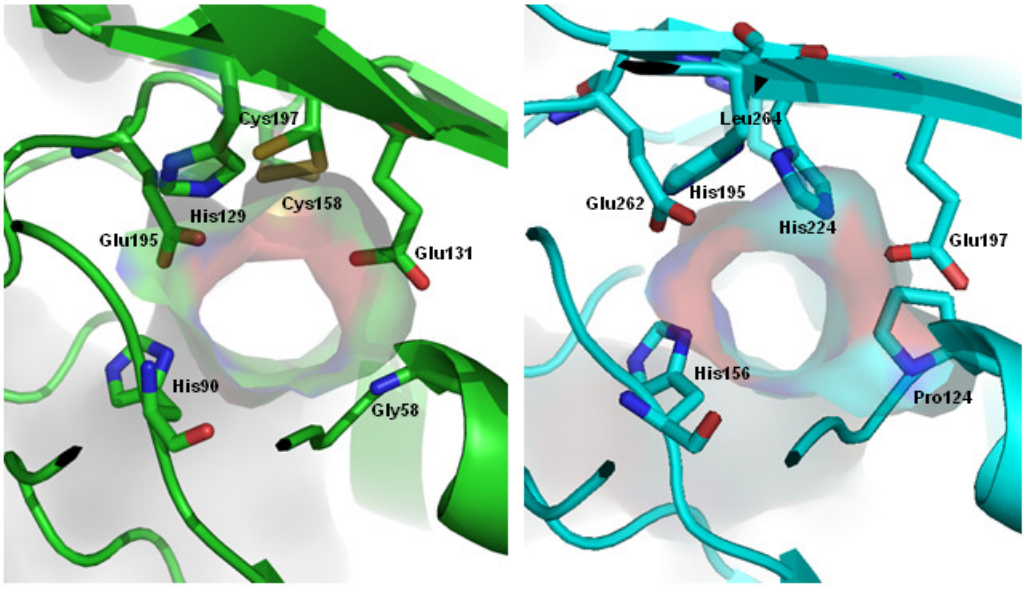


Figure 4.12 Comparison of PduB and EtuB subunit pore. **(A)** Sequence comparison of EtuB and PduB, generated by ClustalW; the pore lining residues are underlined and key residues highlighted in (B) were marked in red, **(B)** Cutaway section looking through the subunit pore of PduB (left) and EtuB (right) showing the positions of residues that line the pore. This figure was produced using PyMOL (DeLano and Lam, 2005).

4.6.2.2 Subunit pores as glycerol channels

2.7 M of glycerol was used as a cryoprotectant for the crystal. The cryoprotectant serendipitously found to be bound within the subunit pores of the shell protein. When viewed in cross section, the channel can be seen to be approximately 22 Å long and 7.5 Å wide. Three glycerol molecules occupy subsites within the channel (Figure 4.13A). A fourth glycerol is also seen located close to the channel opening; this glycerol appears not to be as tightly bound as the chain of three glycerol molecules occupying the channel itself (Figure 4.13A). Starting from the concave side of the channel, the oxygen atom (O1) of glycerol (GOL1) binds to the oxygen (OG1) of the side chain of Ser-191 (Figure 4.14A). The second (GOL2) and third (GOL3) glycerol make more extensive hydrogen bonds, both to each other as well as with residues lining the pore. The key residues for this interaction are Glu-131 (OE1) and Glu-195 (OE1 and OE2) whose carboxylate oxygen atoms hydrogen bond to both glycerols. The main chain amide of Gly-58 and side chain amide (ND1) of His-90 also form hydrogen bond to glycerol 2 (GOL2; Figure 4.14B and 4.14C). These residues are conserved across bacterial PduB sequences as well as EtuB, with an exception of Gly-58, which is a proline in EtuB (Fig 4.12A). The final glycerol (GOL4) completes the chain of 3 glycerol molecules and appears to be more weakly associated with PduB (Figure 4.14D). A table and graphical representation of glycerol interaction with each other and pore-lining residues are shown in Table 4.4 and Figure 4.14, respectively.

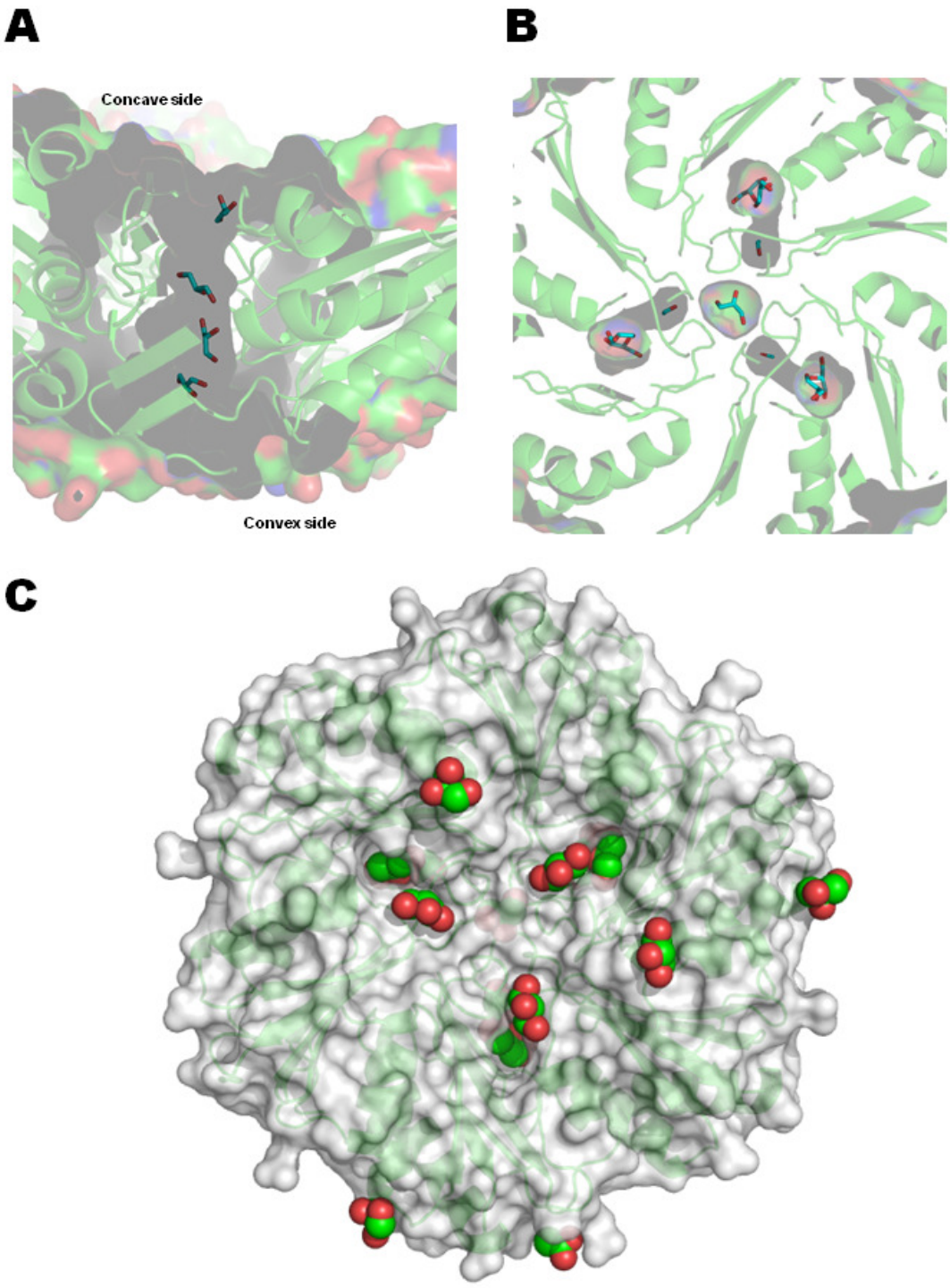


Figure 4.13 PduB and ligands. **(A)** Four glycerols lined up in the subunit pore region of PduB. **(B)** Top view looking down onto the concave side showing glycerols occupying the pores. **(C)** Top view looking at the concave side showing the ligands in this area.

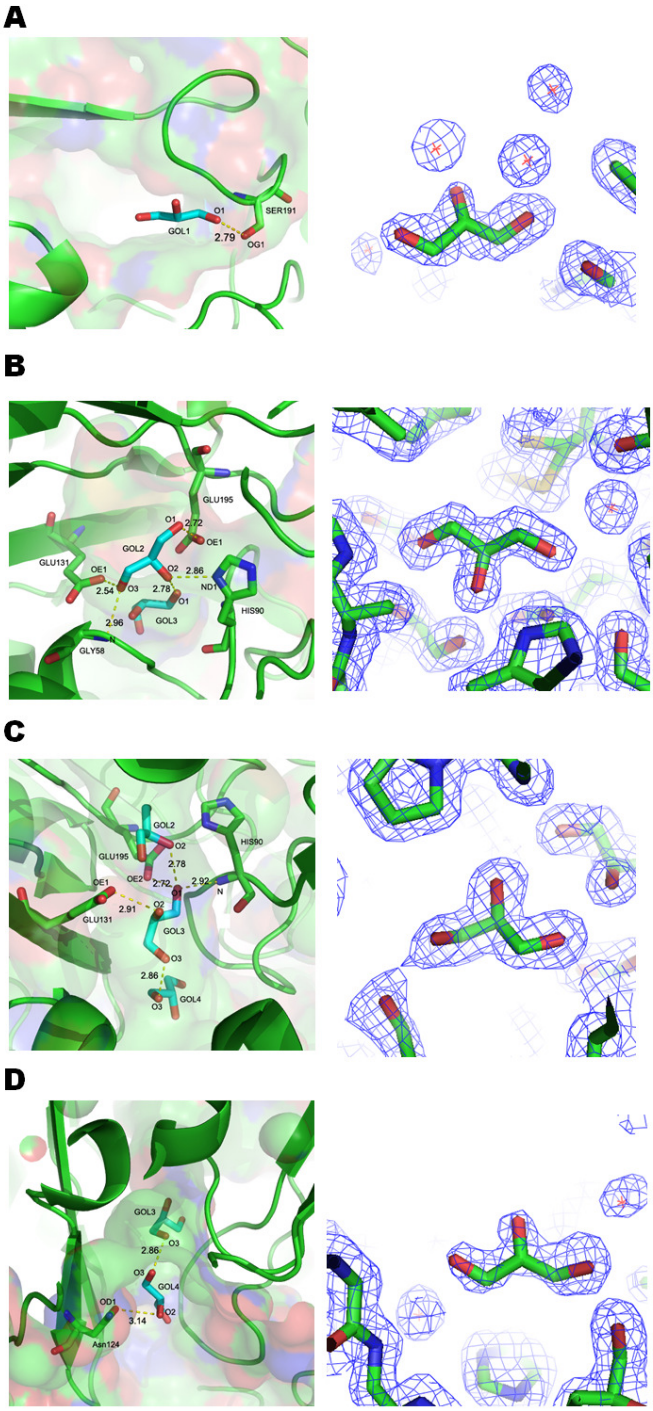


Figure 4.14 Localisation of glycerol in the subunit pores of PduB. **(A)** GOL1 **(B)** GOL2 **(C)** GOL3 **(D)** GOL4 forms hydrogen bonding with the residues lining the pore (left). The σ_A-weighted 2Fobs-Fcalc Fourier synthesis contoured around respective glycerol molecules contoured at 1 σ showing the quality of the electron density map, represented by blue chicken wire mesh (right).

Table 4.4 Interaction of glycerol molecules in the subunit pore with the pore lining residues of PduB and other glycerol molecules

Glycerol in the subunit pore	Glycerol Atom Number			Atom in Contact	Distances (Å)		
	Chain A	Chain B	Chain C		Chain A	Chain B	Chain C
GOL1 / GOL5 / GOL9	O1	O1	O1	Ser-191 (OG)	2.79	2.71	2.55
GOL2 / GOL6 / GOL10	O1	O1	O1	Glu-195 (OE1)	2.72	2.64	2.63
	O2	O2	O2	His-90 (ND1)	2.86	2.85	2.95
	O2	O2	O2	GOL3 (O1) / GOL7 (O3) / GOL11 (O1)	2.78	3.00	2.91
	O3	O3	O3	Gly-58 (N)	2.96	2.91	2.87
	O3	O3	O3	Glu-131 (OE1)	2.54	2.66	2.67
GOL3 / GOL7 / GOL11	O1	O1	O1	His-90 (N)	2.92	2.95	3.01
	O1	O1	O1	Glu-195 (OE2)	2.72	2.73	2.73
	O1	O1	O1	GOL2 (O2) / GOL6 (O2) / GOL10 (O2)	2.78	3.03	2.86
	O2	O2	O3	Glu-131 (OE1)	2.91	2.87	2.63
	O3	O3	O3	GOL4 (O3) / GOL8 (O3) / GOL10 (O3)	2.86	2.66	2.80
GOL4 / GOL8 / GOL12	O2	O2	O2	Asn-124 (OD1)	3.14	4.45	2.82
	O3	O3	O3	GOL3 (O3) / GOL7 (O3) / GOL11 (O2)	2.86	2.66	2.95

4.6.3 Glycerol binding on the 3-fold symmetry axis of PduB

The trimeric shell protein PduB has an occluded central region. This central part has been blocked by the three phenylalanine residues residing in the upper loop (Figs. 4.15A and 4.16A). This structural characteristic has been evident in the closed form of EutL; at one side, the trimer is blocked by a similar six-carbon aromatic ring tyrosine residues (also seen in EtuB) and the other side is by imidazole rings of histidine residues (Figure 4.15B) (Sagermann et al., 2009, Takenoya et al., 2010, Tanaka et al., 2010). Interestingly, the phenylalanine residues seen in CsoS1D (Figure 4.15C) and PduT trimers (Figure 4.15D) do not play a role to block the central region. This might be due to functional importance; for instance, a big central pore has to be formed to accommodate a 4Fe-4S binding cluster (Pang et al., 2011, Crowley et al., 2010). On the other side of PduB, key residues that play a role in blocking the central region are arginine and glycine residues. These arginine residues are seen to be the main key players of blocking the central region of CsoS1D (Figure 4.15C).

A solvated glycerol molecule is trapped on the three-fold axis in the PduB structure (Figs. 4.13B, 4.16B and 4.16C). It is surrounded by three well defined water molecules to which it makes hydrogen bonds. The water molecules in turn form hydrogen bonds with the oxygen atom of the side chain of Asp-83 and oxygen atom of the main chain of Gly-86. The electron density for the central glycerol is remarkably clear given the anticipated rotational averaging of the glycerol occupying this site (Fig 4.16C); this clarity is a consequence of the approximate three-fold symmetry of the glycerol molecule. Two hydroxyls make hydrogen bonds to the

carbonyl oxygen of Asp-83. The site has a loop above (concave side) and a loop below (convex side), both characterised by the presence of residues with small side chains suggesting the possibility of flexibility (breathing motions). The upper loop (189 to 193) has GTSFS (Figure 4.16A) and the lower loop (82 to 89) has sequence RDTKGGGG (Figure 4.16B).

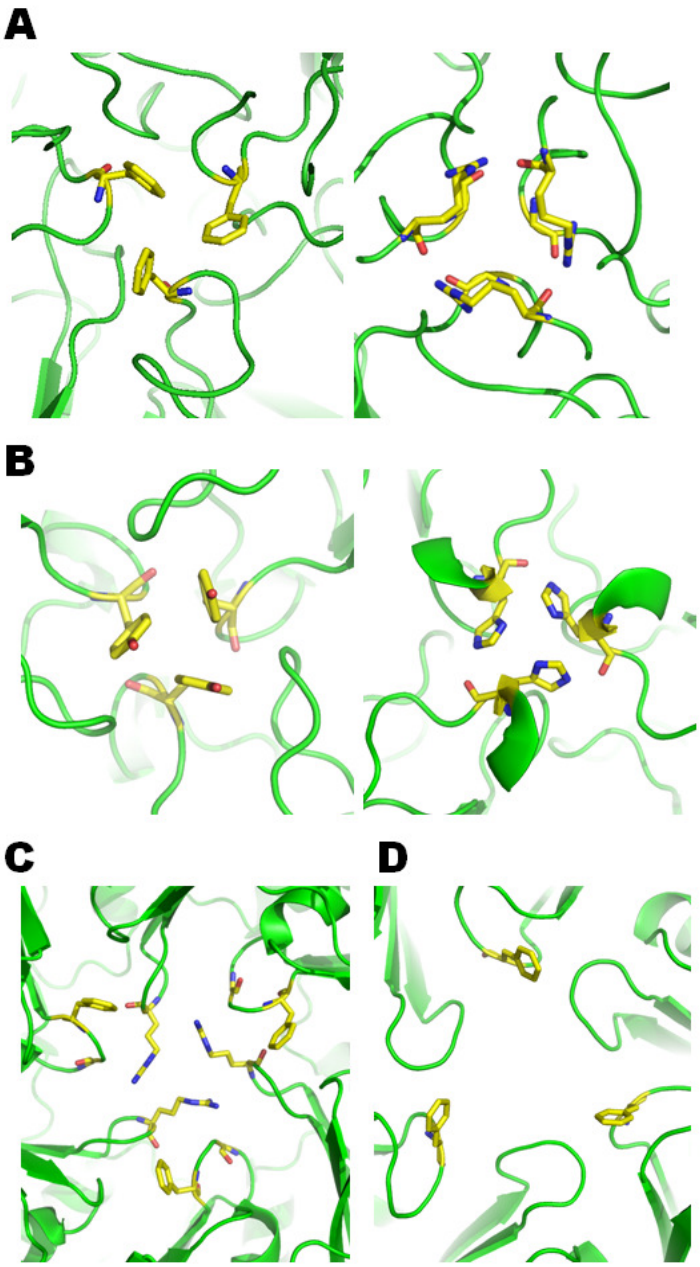


Figure 4.15 Difference in key residues located at the centre of the trimers. **(A)** PduB has Phe-192 in the upper loop (left); Arg-82 and Gly-86 in its lower loop (right), **(B)** Similar three ring residues Tyr-70 (left) and His-75 (right) can be seen in EutL closed form blocking the trimer centre. **(C)** CsoS1D has three Arg-120, the main residues blocking the pore. The shell protein also has Phe-227 and Gly-225 within the centre of the trimer. **(D)** PduT has Phe-130 moved away to give way for 4Fe-4S cluster.

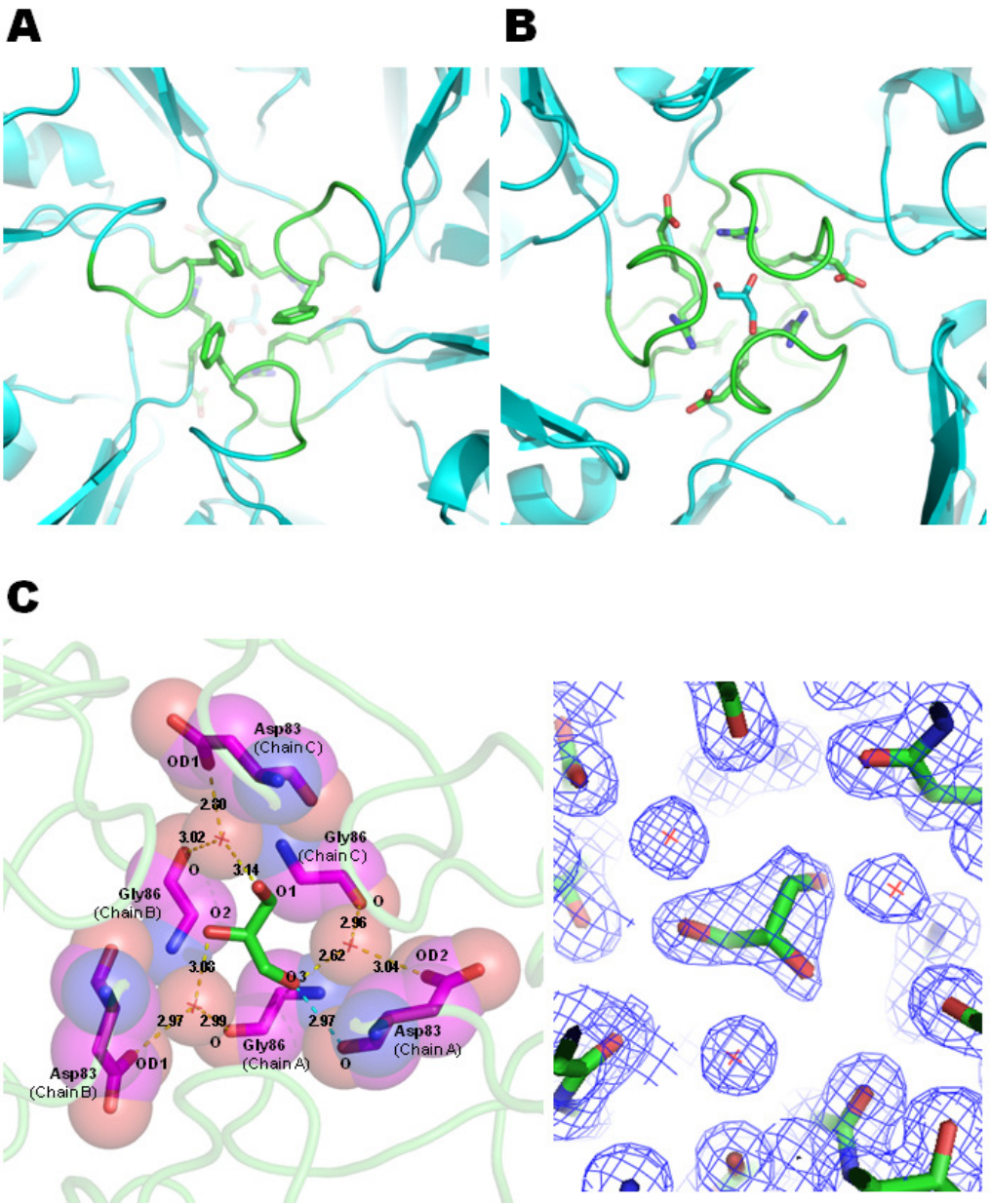


Figure 4.16 Three-fold symmetry axis of PduB. **(A)** View of the upper loop (GTSFS) coloured in green, three phenylalanine residues block this side of the central region, **(B)** View of the lower loop (RDTKGGGG) coloured in green, Asp and Arg residues encircling the glycerol are shown. **(C)** A representative of one orientation of glycerol on the three fold symmetry axis of PduB surrounded by three water molecules, which in turn form hydrogen bonds with Asp-83 (OD1) and Gly-86 (O) (left). Electron density map for this glycerol, surrounded by three water molecules, contoured at 0.5 σ (right).

4.7 Conclusion

The crystal structure of PduB reveals that the shell protein assembles into trimer with three subunit pores formed by the tandem BMC repeat. The first ten residues of the shell protein is a flexible region and therefore not evident in the crystal structure; but suffice to say that this is the full length PduB since majority of the N-terminal region missing in PduB' can be observed in the crystal structure. This N-terminal region is located in the convex side of the trimer (Figure 4.17); further to this, nothing else can be inferred. Whether the existence of two forms PduB and PduB' is important is not clear.

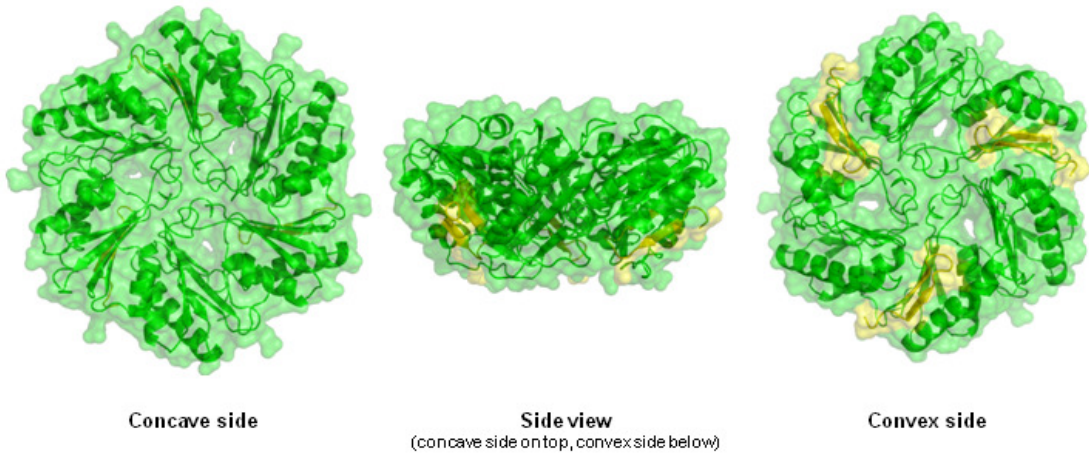


Figure 4.17 N-terminal region of PduB. The concave, side view and convex side of PduB are shown in this figure. The crystal structure reveals part of this N-terminal region (15 amino acid residues), while the other 10 amino acid residues (MNDFLNSTST) were not evident due to flexibility. The N-terminal region (coloured in yellow) has a structural feature of two anti-parallel β-sheets and resides on the convex side of the trimer.

When the symmetry-related molecules of PduB are generated, the shell protein shows ability to tile and form into sheets. The crystal packing of PduB, however, does not suggest how the shell protein would appear in biological condition; this is mainly due to large gaps in between symmetry-related molecules. Compared to PduB, the crystal structure of PduA forms a tightly packed molecular sheet. The PduB trimer is able to fit in well in this molecular sheet, providing an insight into how the facet of the polyhedral shell is formed. This theoretical fitting of the trimer is supported by the experimental observation of the interaction of these two major shell proteins of pdu metabolosome (Parsons et al., 2010a).

Pores or channels are a feature of several bacterial microcompartment shell proteins, these pores can be at the centre of the hexamer or trimer, or within subunits of the trimeric shell proteins. It is anticipated that these pores are functionally important in channelling substrates and metabolites. For years, there has been circumstantial evidence such as ions seen trapped in the pore regions of different shell proteins (Tanaka et al., 2009, Tsai et al., 2007, Takenoya et al., 2010). However, definitive demonstration of specific substrates trapped within the channels of shell proteins has been elusive.

Previous genetic and biochemical studies on the *Lactobacillus reuteri* 1,2-propanediol microcompartment showed that the bacterial species is able to metabolise both glycerol and 1,2-propanediol within this organelle (Sriramulu et al., 2008). The abundance of glycerol ligands binding to sites around the concave face of the shell protein PduB is suggestive of their involvement in glycerol transport (Figure 4.14C), possibly to attract glycerol into the channel though it should be noted

the glycerol concentration is high in this study (2.7 M). Glycerol molecules trapped in the long thin channels of PduB confirm that these trimeric subunit pores may act as channels for substrates. It is to be expected that substrate channels have some affinity for their substrates as a mechanism of channelling and selecting substrates for import into the microcompartment. Several crystal structures of membrane channel proteins with trapped cargo in their pore regions had been characterised; for instance, aquaglyceroporin with glycerol (Newby et al., 2008), maltoporin with sucrose, trehalose and melibiose (Wang et al., 1997), and a mitochondrial channel with ATP (Rostovtseva and Bezrukov, 1998). The permeating molecules do not simply slip through the pore, but experience attraction to the pore-lining residues (Berezhkovskii and Bezrukov, 2005). The structural association of glycerol molecules with the pore allows identification of conserved residues in PduB which are associated with that affinity. Based on the size and hydrogen-bonding potential of 1,2-propanediol compared to glycerol, it is anticipated that the subunit channels may also be conduits for 1,2-propanediol. However, propionaldehyde and β-hydroxypropionaldehyde might not pass through the pores because of the different hydrogen-bonding potential of the aldehyde compared to hydroxyl-group. Aldehydes can accept two hydrogen bonds in a planar arrangement, while alcohols can donate one hydrogen bond and accept two with tetrahedral geometry. Both O1 and O3 of GOL2 are hydrogen-bond donors, because the carboxylate groups of Glu-131 and Glu-195 are expected to be deprotonated (Figure 4.14B). Similarly, O1 of GOL3 is required to donate a proton to Glu-195, and the geometry at this atom is approximately tetrahedral (Figure 4.14C). These two sites therefore appear to act as a specificity filter that allows passage of substrates, but retains reaction intermediates

within the microcompartment. The recently obtained crystal structure of PduB with propionaldehyde in the reservoir mix supports this notion. The reaction intermediates (propionaldehyde) were not seen trapped in the subunit pores; instead, weak density suggests that possibly propionaldehyde can be found in the subunit pore opening (convex side). This suggests that the pores do not have affinity for the aldehyde intermediates, and therefore, do not allow these molecules to go through (data not shown). The pore-lining residues involved in the formation of hydrogen bonds with glycerol molecules in *L. reuteri* PduB are conserved across PduB-like proteins. This is not surprising as the substrates (glycerol, 1,2-propanediol, ethanol and ethanolamine) share similar characteristics which suggests translocation using the same mechanism and closely similar pores.

The trimeric shell protein EutL has been shown to have a central channel that can be open or closed (Sagermann et al., 2009, Takenoya et al., 2010, Tanaka et al., 2010). The presence of the buried glycerol on the three-fold axis of the PduB trimer suggests that the central loops are flexible and glycerol-binding stabilises the conformation seen in the crystal. In addition to this, a glycerol molecule was seen trapped in a central pocket suggesting it locks closed the central loops and raising the possibility of a ligand-gated channel. The crystal structure provides the first evidence of the bacterial microcompartment pore acting as channel for natural substrate.

Chapter 5

Structural analysis of PduA tiling

5.1 Introduction

In a previous study, it was found that PduA, PduB, PduJ, PduK and PduN are the minimum shell components to form a non-aberrant empty pdu microcompartment (Parsons et al., 2010a). Deletion of one of these shell proteins lead to formation of elongated filamentous structures that resembles nothing like bacterial microcompartments (Figure 5.1A). Out of these major shell components, PduA seems to play a crucial role in coordinating with other shell proteins as indicated by the protein-protein binding studies (Parsons et al., 2010a). PduA interacts with PduB, PduJ, PduK, PduN and PduU. This explains perhaps why PduA is the first encoded protein of the *pdu* operon, allowing the newly synthesised shell protein to interact with other subsequently translated shell components. Indeed, when the order of gene expression of *pdu* operon is disrupted, aberrant organelles are formed. This suggests the importance of PduA as a scaffold for the microcompartment assembly.

Interestingly, overproduction of PduA alone in *E. coli* produces large cytoplasmic axial filaments that resembles lattice of hexagonal structures (Figure 5.1B).

Unlike the previouly discussed PduT (Pang et al., 2011) and PduB trimers, the crystal structure of PduA showed six wedge-shaped subunits each containing a single canonical BMC domain assembled into a hexameric structure (Crowley et al., 2010). These PduA hexamers can be packed into uniform molecular sheets. Vital to the understanding of microcompartment assembly is determining how neighbouring shell proteins interact. This will be explored here using a construct known as PduA*, a soluble form of *Citrobacter* PduA, similar construct used in (Parsons et al., 2010a).

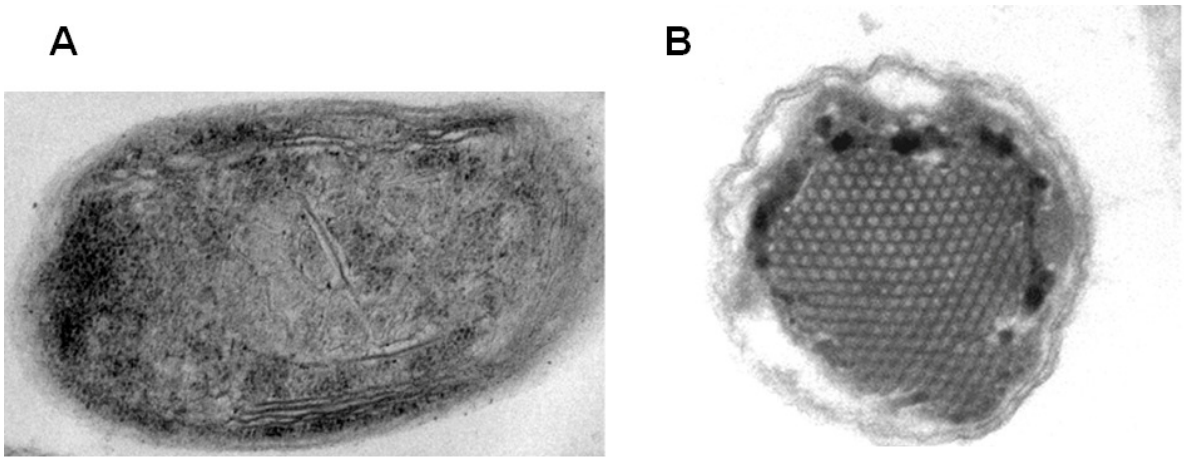


Figure 5.1 Transmission electron micrograph of cells overproducing **(A)** PduBJKN – microcompartments which are improperly formed as filamentous structures (Parsons, 2008), and **(B)** PduA alone which has the appearance of a hexagonal lattice (Parsons et al., 2010a).

5.2 Generation of PduA mutants

5.2.1 Mutational sites

The tiling of PduA shell proteins relies on the interaction between one hexameric structure and its surrounding hexamers. A central PduA hexamer interacts with six adjacent PduA hexamers, in turn each of these six hexamers interacts with six hexamers to form a sheet of hexamers which make the facet of the microcompartment. Crucial to the understanding of the formation of this shell facet is the interaction between one hexamer and another. The interface between the shell protein was analysed using of Knowledge-based Fast Atomic Density Evaluator and Contacts (KFC) Server (Darnell et al., 2007, Zhu and Mitchell, 2011) and Protein-Protein Interface Analysis (ProtorP) Server (Reynolds et al., 2009). Both programs search for possible binding 'hot spots' at the interface. Three regions of interaction were identified (Figure 5.2). These hot spots were confirmed via visualisation of one *Citrobacter* PduA hexamer (as modelled from *Salmonella* PduA) next to another one using PyMOL (DeLano and Lam, 2005).

Within each 'hot spot cluster', one residue was chosen to mutate based on its position in the cluster and conservation within the shell proteins. The residues selected were K26, V51 and R79 (as illustrated in Figure 5.2). The chosen residues were found to be conserved within the major shell components (PduA, PduJ and PduK) and interestingly, also within the trimeric PduT as well (Figure 5.3).

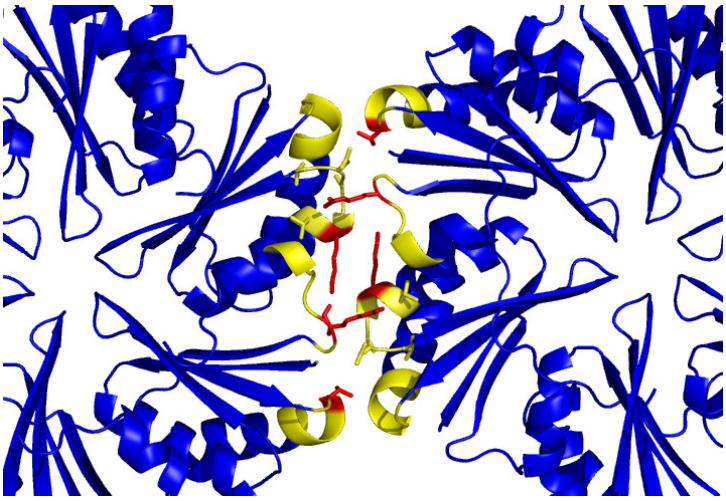


Figure 5.2 Hotspots for PduA-PduA interaction. Three clusters were found to be potential interaction sites, coloured in yellow. Residues chosen to be mutated to either alanine or aspartate are coloured in red.

```
PduA            MQQEALGMVETKGLTAAIEAADAMVKSANVMLVGYEKIG-SGLVTVIVRGDVGAVKAATD 59
PduJ            -MNNALGLVETKGLVGAIEAADAMVKSANVQLVGYEKIG-SGLITVMVRGDVGAVKAAVD 58
PduT            -MSQAIGILELTSIAKGMEAGDAMLKSANVNLLVSKTIC-PGKFLLMLGGDVGAVQQAIA 58
PduK            -MKQSLGLLEVSGLALAISCADVMAKAASITLVGLEKTNGSGWMVIKIIGDVASVQAAIS 59
                  .:::*::* ..:. .:...*.* *:*.: *:  :.   .* . : : ***.:*: *

PduA            AG-------------------------------------------------AAAARNV-- 68
PduJ            AG-------------------------------------------------SAAASAV-- 67
PduT            TGTSLAGDMLVDSLVLPNIHASVLPAISGLNSVDKRQAVGIVETWSVAACICAADRAVKA 118
PduK            TGVSFADQ------------RDGLVAHKVISRPGDGILSHSVTPESESEPAPAPTPVVP 106
                :*                                                ..*   *

PduA            ----------------------GEVKAVH--------------------VIPRPHTDVE 85
PduJ            ----------------------GEVKSCH--------------------VIPRPHSDVE 84
PduT            SNVTLVRVHMAFGIGGKCYMVVAGDVSDVNNAVTVASESAGEKGLLVYRSVIPRPHESMW 178
PduK            ----------------------HEEIPEDHAAPEAPQDAELISCNLCLDPACPRQKGEPR 144
                                       ::   :                   . ** : .

PduA            KILPKGIS---- 93
PduJ            AILPKSA----- 91
PduT            RQMVEG------ 184
PduK            SLCLHSGKRGEA 156
```

Figure 5.3 Sequence alignment of PduA, PduJ, PduK and PduT. Amino acid residues underlined are surface residues available for interaction. Key conserved residues to mutate are coloured in red.

5.2.2 Mutation of PduA construct

To confirm the importance of these selected residues for the ability of PduA to tile, these residues were mutated into a small amino acid with hydrophobic side chain (alanine) and negatively charged aspartate. As described in Materials and Methods, the mutagenesis procedure was adopted from the Quikchange Site-Directed Mutagenesis (Stratagene). PCR reaction for PduA constructs were detailed in Appendix A1. Mutagenic constructs were verified by DNA sequencing (Eurofins).

5.3 Solubility of PduA and mutants

Preliminary experiments by our collaborators suggested that the shell protein PduA is insoluble. With an element of serendipity, the collaborators also obtained a more soluble form of PduA, called PduA*, which has a mutated C-terminal residue (S→R) and an extra C-terminal 23 amino acid residues.

However, after overproduction and purification of PduA*, the shell protein aggregated which led to the suggestion that the protein might be self-associating to form tiles. A closely related shell protein PduJ, with a sequence identity of 82% to PduA, was also found to be insoluble. This prompted an exploration of an alternative way of preventing the shell protein from aggregating; which ties in nicely with the mutation of key residues suggested to be important for tiling.

PduA* as well as the mutants were produced and purified according to the protocol set in Chapter 2. The shell proteins were His-tagged and this tag was exploited for protein purification. After this purification step, the proteins were more than 90% pure, as determined by polyacrylamide gel electrophoresis (Figure 5.4A). The high

concentration of NaCl and imidazole were removed using a buffer exchange column (HiPrep 26|10), leaving the shell protein in a solution of 50 mM Tris pH 8.5 and 50 mM NaCl. The proteins were concentrated to about 10 mg/ml. Proteins were left in solution at 4℃ overnight. Observation of aggregation is done by checking for cloudiness in the solution. The protein solution was also spun down for 1 minute at 16000 rpm. Evidence of pellet at the bottom of the Eppendorf tube is an indication of protein aggregation.

As expected, PduA* was prone to aggregation. Two mutants, K26A and R79A, also shared this characteristic. Addition of urea helped to prevent aggregation of these proteins. However, the mutants K26A and R79A required less urea to solubilise in comparison to the wild type PduA*. K26D and V51D were soluble without the addition of urea. This suggests that mutations K26D and V51D may prevent the interaction between neighbouring PduA molecules. These observations are summarised in Table 5.1.

Table 5.1 Solubility testing of PduA and the mutants. × = pellet at the bottom of the tube = insoluble; ✓ = no pellet = soluble

	Urea concentration					
	0M	1M	2M	3M	4M	5M
PduA	×	×	✓	✓	✓	✓
K26A	×	✓	✓	✓	✓	✓
K26D	✓	✓	✓	✓	✓	✓
V51A	×	✓	✓	✓	✓	✓
V51D	✓	✓	✓	✓	✓	✓
R79A	×	✓	✓	✓	✓	✓

5.4 Characterisation of mutants

5.4.1 Crystals

The molecular packing of proteins often influences the morphology of the crystal. Four of the shell proteins (PduA, PduB, PduT, PduU) that make up the pdu metabolosome form hexagonal plate crystals. These shell proteins were found to pack together to form a sheet with packing similar to that we expect to see in a facet of the metabolosome shell.

K26A, V51A, and V51D crystallised in various crystal morphologies (Figure 5.4B-D). These crystal morphologies deviate from the usual hexagonal plates seen when wild type PduA is crystallised (Crowley et al., 2010). The sea urchin-like crystals of V51D later grew to a bigger size with individual crystal growing from the nucleation site can be snapped off. These crystals were taken to Soleil (Paris, France) and Diamond Light Source (Oxfordshire, UK) for diffraction studies. The crystals diffracted poorly with the best one diffracting to only 8 Å resolution, which is not sufficiently good to yield molecular details. The drastic change in crystal morphology indicates that the crystal packing has been adversely affected by the mutational change. By altering the surface residues of PduA hexamers that are brought together in a molecular tile, it is not surprising that the crystal does not readily pack so well. These observations indicate that residues Lys-26, Val-51 and Arg-79 are important for the association of PduA hexamers.

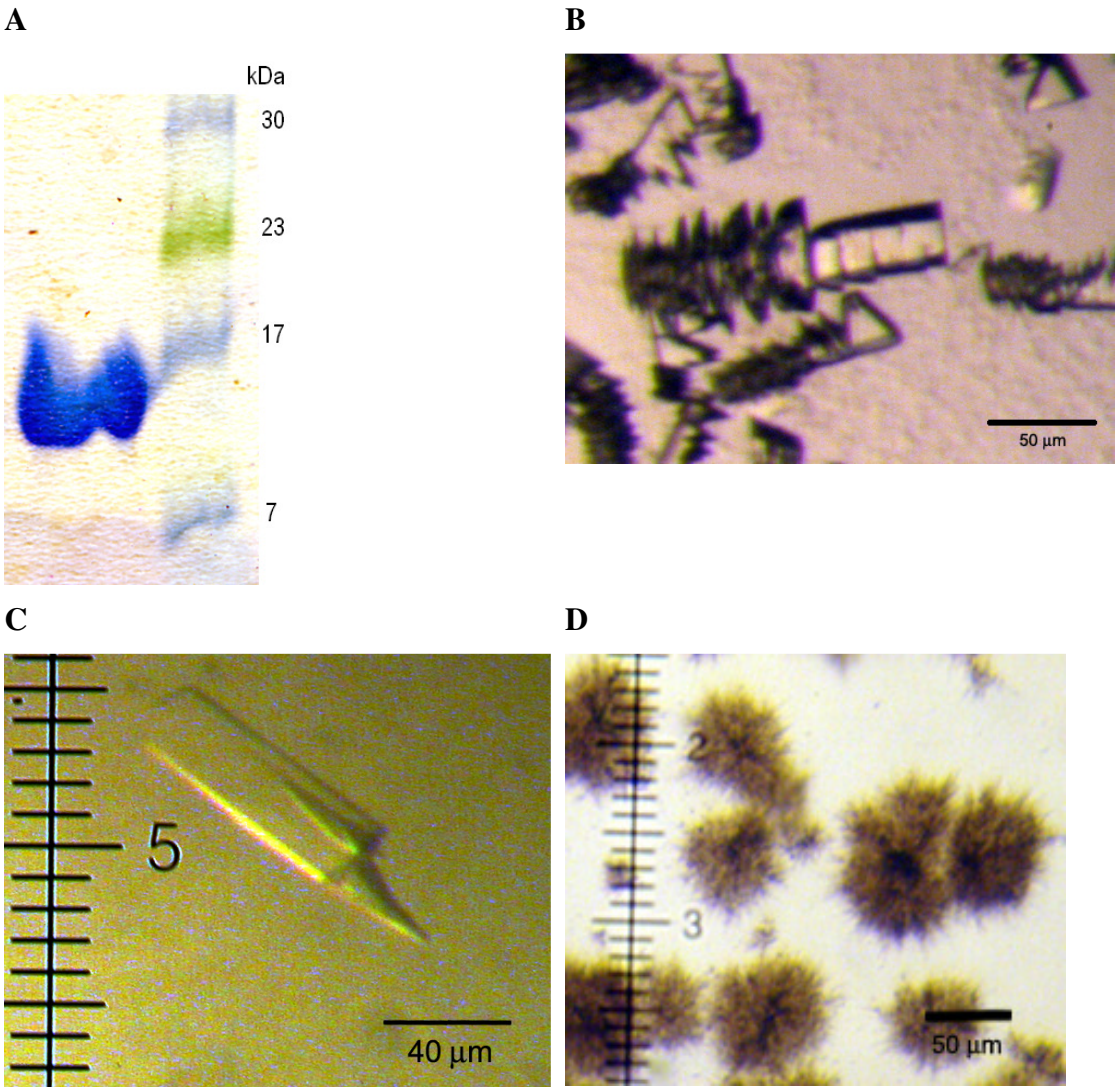


Figure 5.4 Purified PduA protein and PduA mutant crystals. **(A)** SDS-PAGE gel showing purified PduA protein, **(B)** K26A – screw shaped crystals, **(C)** V51A - rectangular shaped crystals, **(D)** V51D – sea urchin shaped crystals.

5.4.2 Transmission electron microscopy

As described in Chapter 2.8, the shell proteins were overproduced in *E. coli* cells. Cells were embedded, sectioned, stained and placed on copper grids for viewing in the electron microscope.

5.4.2.1 PduA forms tube-like structures

The cross-section of *E. coli* cells overproducing PduA shows that the shell protein self-assembles to form an array of hexamers (Figure 5.5). However, in longitudinal sections PduA appears to form thin laminar structures (Figure 5.5). Both hexagon-shaped and the laminar structure have a width of about 250 Å; presumably the laminar structures are bundled together and therefore appears as hexagonal lattice in cross-sectional view. These observations are consistent with earlier reported experimental results (Havemann et al., 2002, Parsons et al., 2010a).

The hexagonal structure seen in the cross-sectional view can be interpreted as PduA hexamers assembled into hexagonal-shape (Fig 5.6B) and these were stacked together on top of one another to appear like filaments (Figure 5.6A). This arrangement, however, is unusual, since it raises numerous questions such as why PduA hexamers are not assembled within this hexagon. In addition to this, the long filaments do not appear to be formed by stacking of hexagonal plates. The structures appear more as continuous tubes. To this end, the second model appears to be more plausible. PduA hexamers have to be arranged side-by-side fashion to form a circular structure. PduA then futher self-assembles longitudinally to form a tube-like structure (Figure 5.6C). When viewed cross-sectionally, this model would appear circular (Figure 5.6D). However, when a number of tube-like structures were

brought together and bundled up, the sides of the tube can easily bent due to frailty of single-layer, which therefore would have a hexameric appearance (Figure 5.5A).

When a sheet of PduA hexamers (12 hexamers across) is rolled into a tube, the diameter is roughly 250Å. This diameter agrees with the TEM images and shows how packing quasi-equivalent to that seen in the crystals could give a thin-walled tube. It also explains why the tubes are empty. There are two ways how PduA could roll up to form a tube, one has the vertices of the hexamers parallel to the axis of the tube, the other has it perpendicular. Further studies are needed to test the veracity of these rolled-sheet models, and to choose between them.

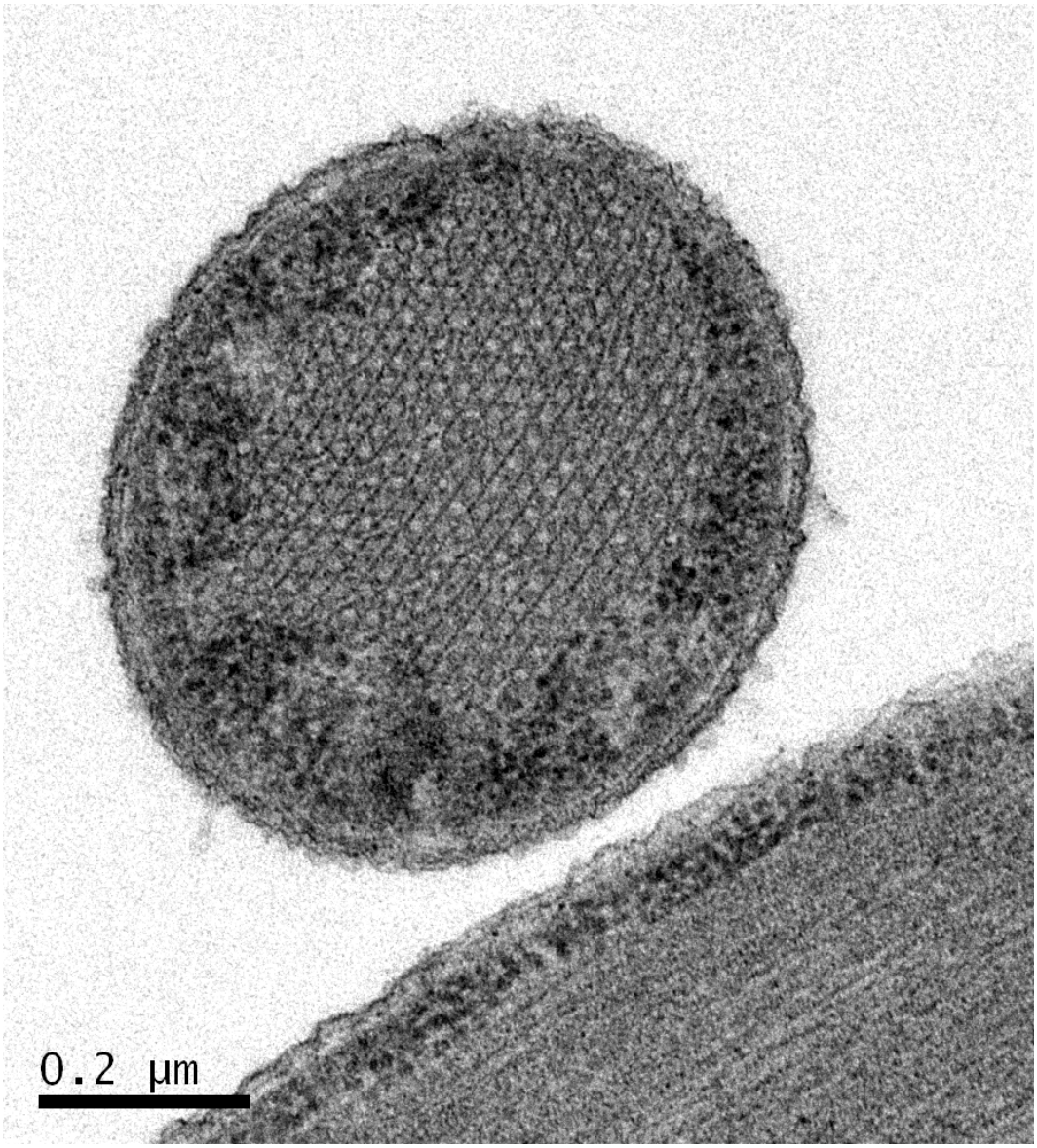


Figure 5.5 Thin section of *E. coli* overproducing PduA. PduA forms a hexagonal lattice within the cell (in cross-section) and appears to be filamentous in transverse section.

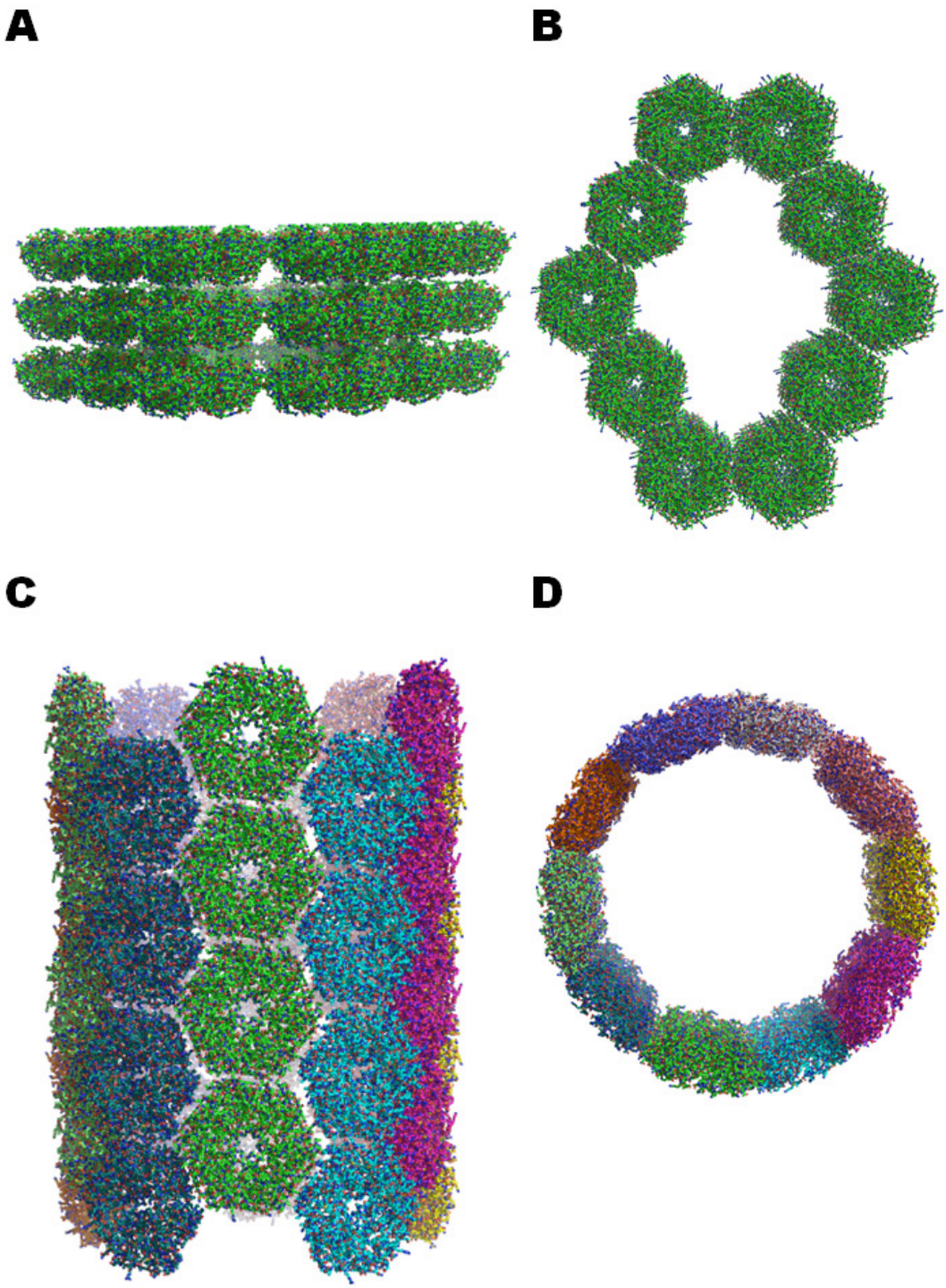


Figure 5.6 Two possible models of the structure within the bacterial cell. The first model represents PduA hexamers **(A)** stacked on top of one another in longitudinal section and **(B)** formed as hexagonal disk in cross section. The second model shows PduA hexamers **(C)** aligned side by side and folded in both ends to appear as **(D)** circular tube.

5.4.2.2 Mutation to aspartate residue disrupts association of hexamers

Overproduction of K26D and V51D PduA in *E. coli* gives no visible structures in the bacterial cell (Figure 5.7). There are two possibilities for this observation. First is the mutant proteins were not expressed within bacterial cells. This, however, is an unlikely explanation since the mutants were successfully extracted, purified and crystallised. The second possibility is that the mutation to aspartate of K26 or of V51 disrupted the self-assembly into a molecular nano-tube.

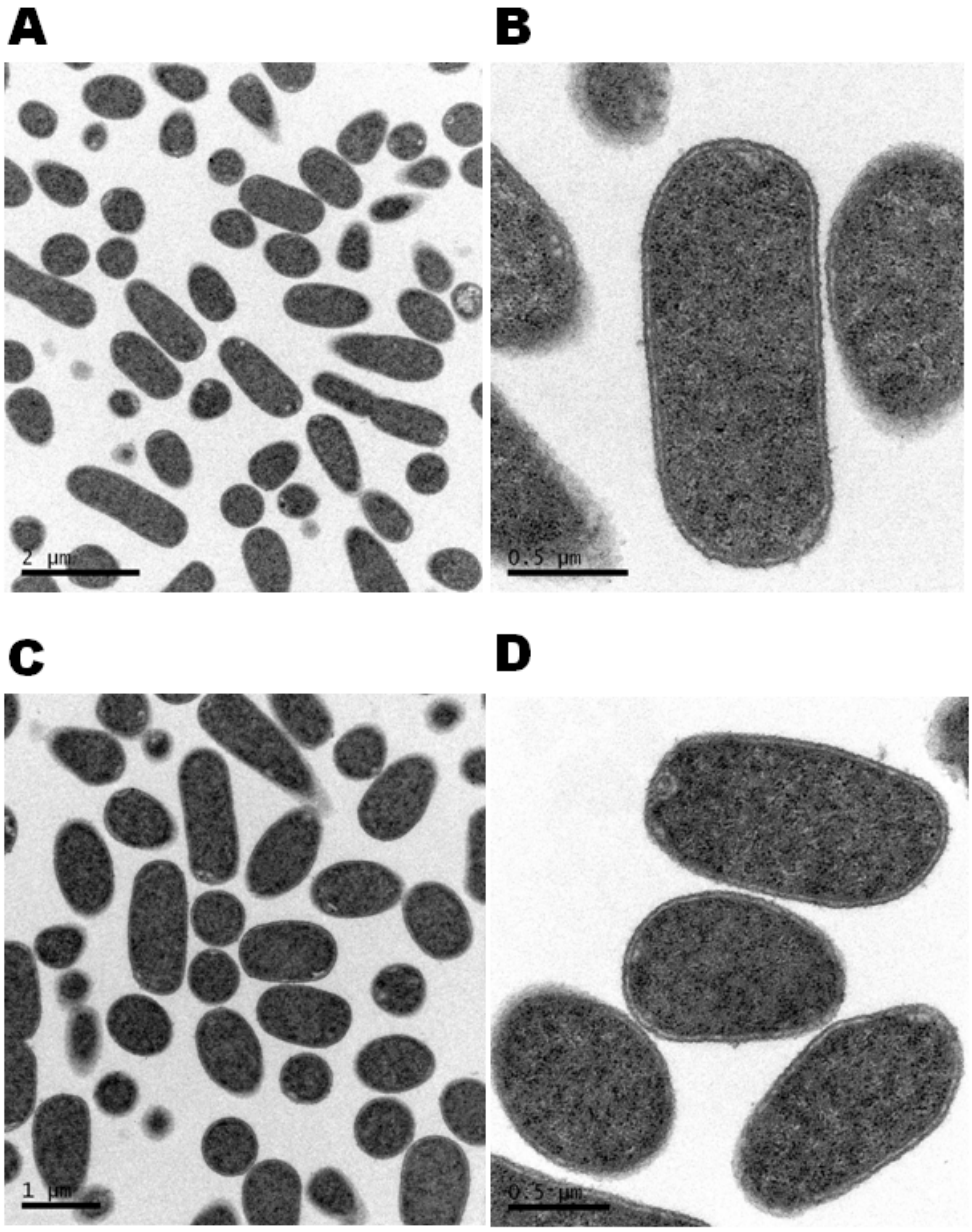


Figure 5.7 Thin sections of *E. coli* overproducing K26D and V51D. Cells appear to be devoid of the shell protein organelle. **A** and **B** are cells showing overproduction of K26D. **C** and **D** are cells showing overproduction of V51D.

5.4.2.3 K26A and R79A assemble into sheets

When alanine is substituted at positions K26 and R79, filamentous structures were observed in both cross section and transverse sections (Figs. 5.8 – 5.11). Cells expressing V51D produces long filaments bundled together and affecting the cell shape (Figs. 5.8 and 5.9). Cells expressing R79A produce similar results but with more adverse effect on the bacterial cell. R79A structures can be found as long as spanning two bacterial cells (Figure 5.10), the long 'filamentous' structures can be seen and bent in edges of the cell (Figure 5.10), as well as filling up the entire cell (Figure 5.11). In addition, the filamentous structure can appear 'indistinct' in transverse section. The "amorphous fur" appearance is presumably due to overlapping of several R79A structures.

Figure 5.12 shows a possible model to explain what is happening in K26A and R79A. Mutant shell proteins K26A and R79A presumably are able to assemble into a molecular sheet but do not "curl up" into a tube-like structure as seen with wild type PduA. The molecular sheets can be thought as rectangular disks, and multiple sheets stack on top of one another. At one angle (cross section), these sheets can be seen in their shorter width while on the other angle (transverse section) these sheets are observed in their longer width.

The mutational change to alanine did not completely disrupt the self-assembly of PduA, in comparison to mutational change into aspartate. Presumably the conversion of the residue into aspartate has a drastic effect, while the anti-parallel lysines are able to form hydrogen bonding; the aspartate residues (if ionised) do not have such capability. Instead, aspartates are negatively charged and repel each other

disrupting the self-association. The alanine residues are small amino acid and usually do not play a functional role. In this case, the amino acid mildy affects the association of hexamers. The filamentous structure is still evident within the bacterial cell; however, the crystal lattice is not. The mutational change to alanine appears to prevent the structures from packing into tubes and they therefore remains as molecular sheets.

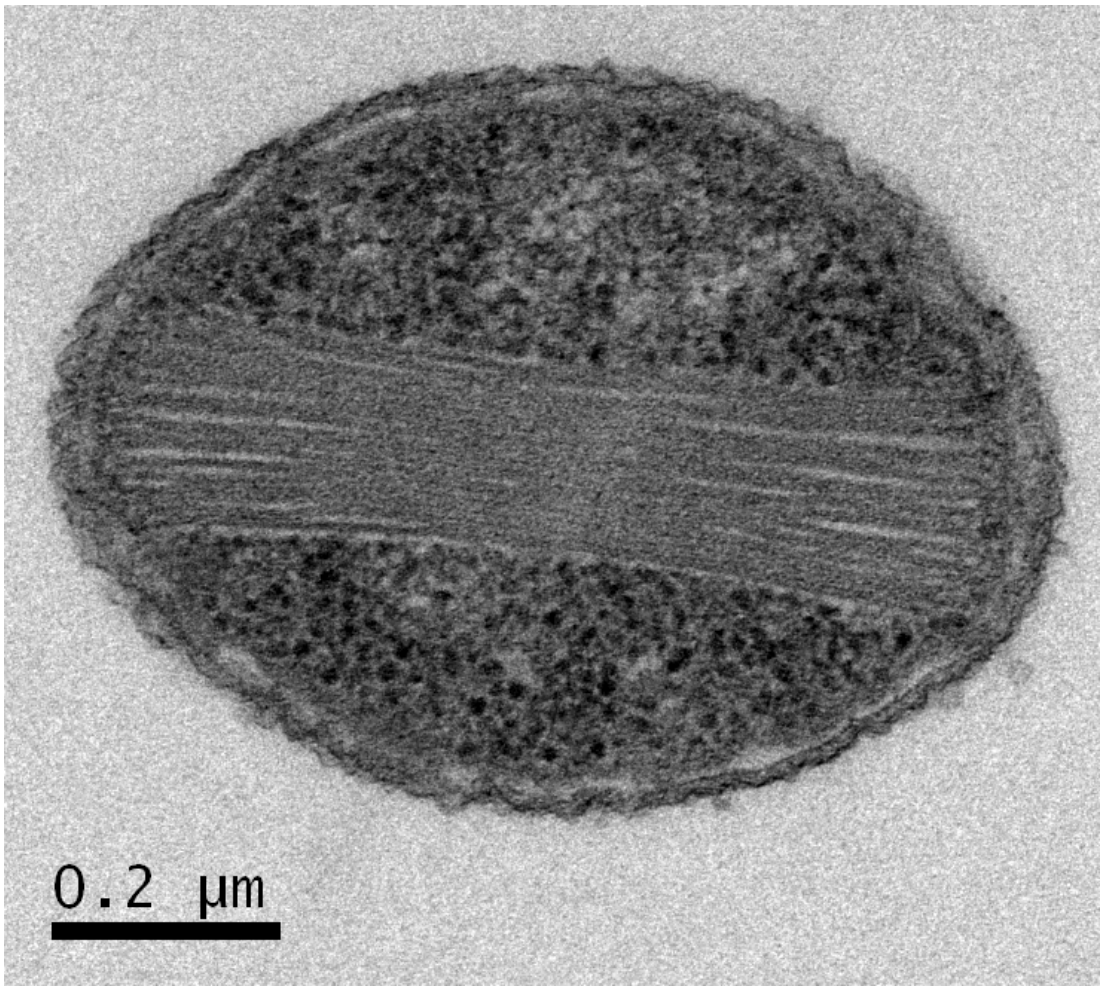


Figure 5.8 Thin section of *E. coli* overproducing K26A in cross sectional view.

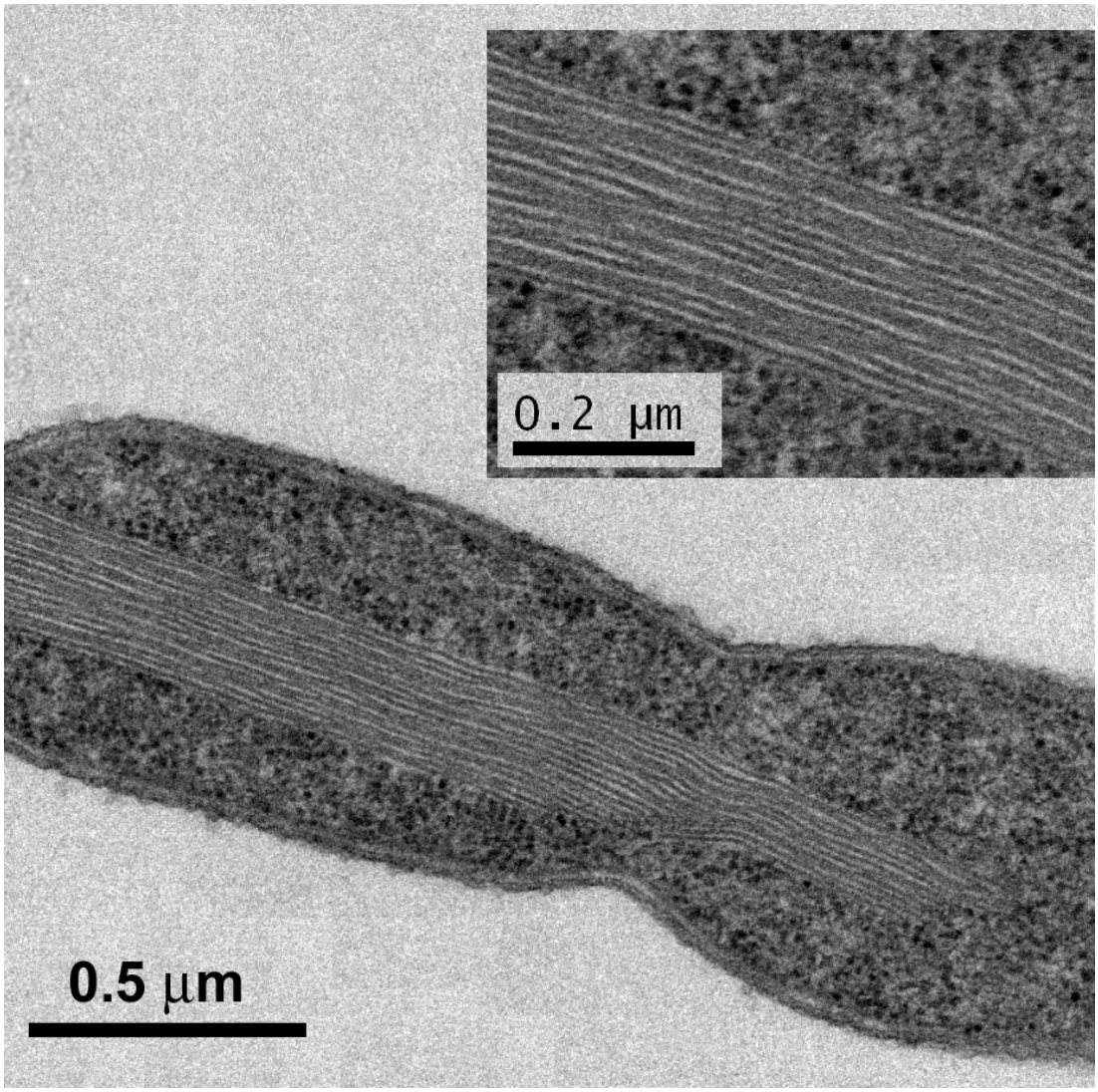


Figure 5.9 Thin section of *E. coli* overproducing K26A in transverse section. Inset shows a higher magnification view of the filamentous strands.

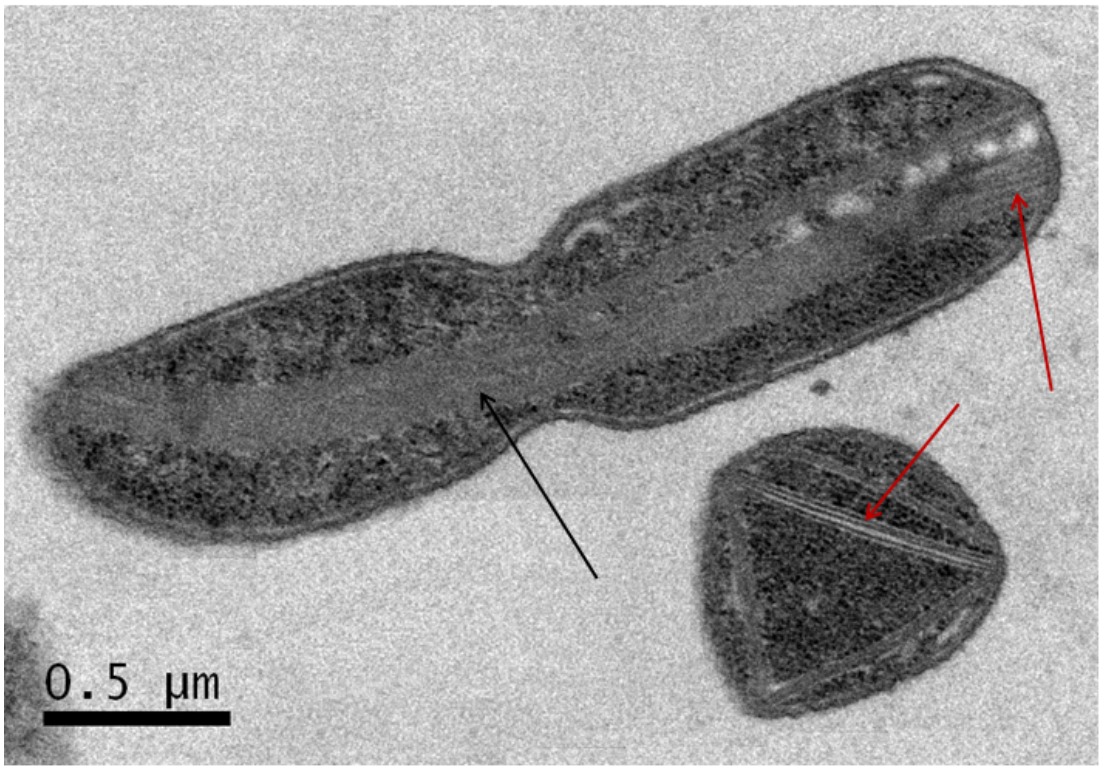


Figure 5.10 Thin section of *E. coli* overproducing R79A in transverse and cross section. The effect of overproducing R79A within *E. coli* in transverse section (top) and cross section (bottom). Red arrows pointing to evident separated filaments. Black arrow pointing to indistinct formation of R79A structures.

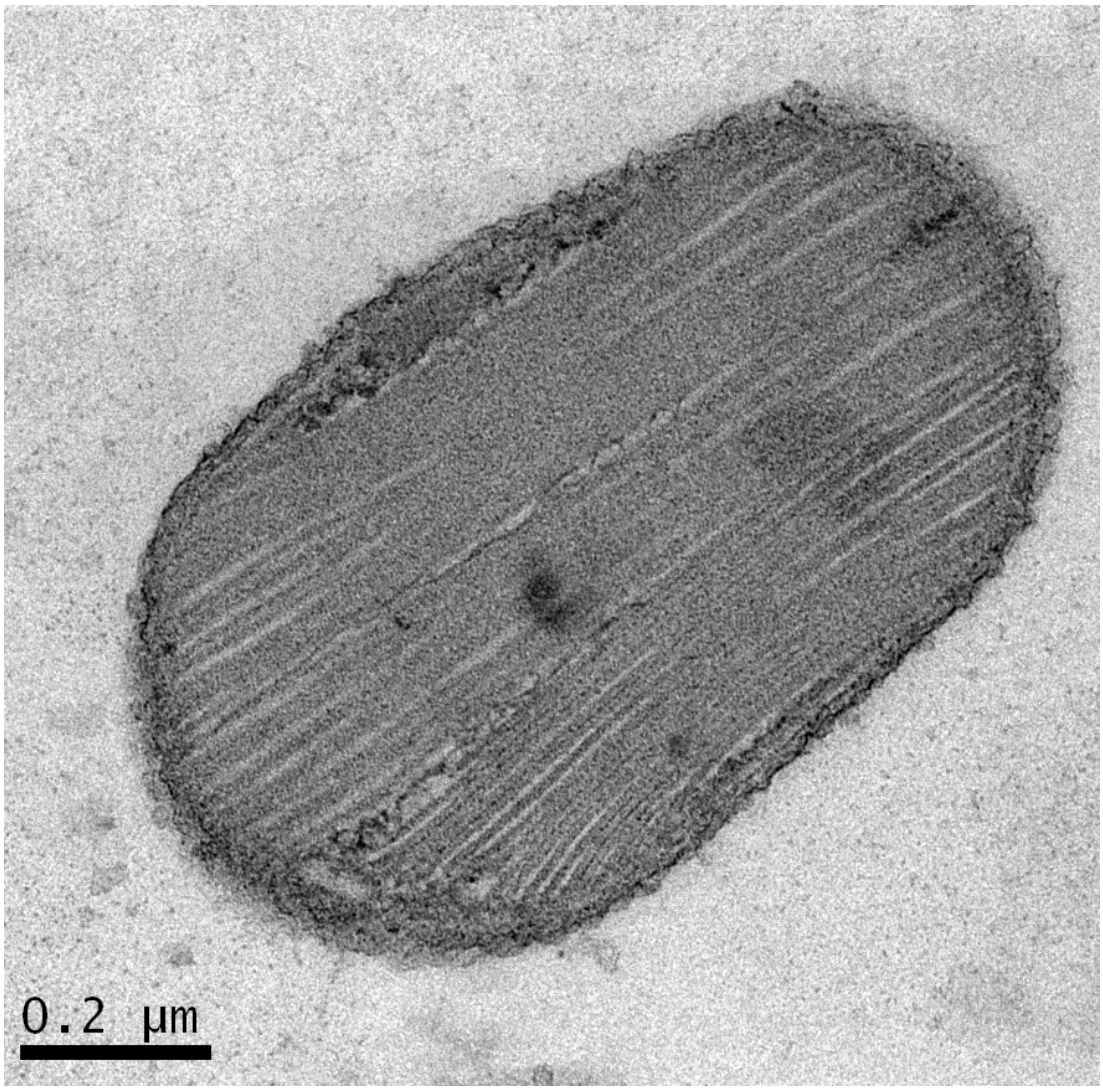


Figure 5.11 Thin section of *E. coli* overproducing R79A in cross section. Filamentous structures of mutant shell proteins occupy the entire cell.

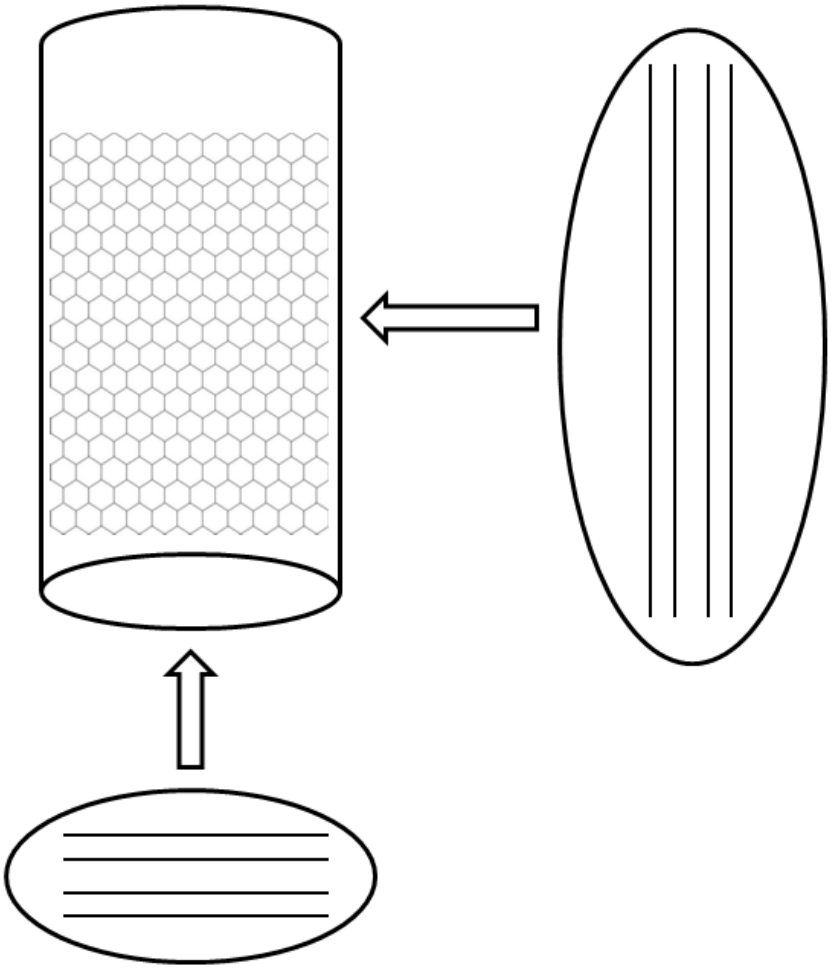


Figure 5.12 Schematic model of K26A and R79A assembly within the cell. *E. coli* cells can be represented by cylinder. The mutant shell proteins are assembled into sheets. Looking through both top and side of the cylinder, the stacking of the sheets can be observed.

5.4.2.4 V51A forms tubes and sheets

The micrographs containing the bacterial cells with expressed V51A mutant shell proteins show an interesting result. In transverse section, the filamentous structures are evident similar to what has been observed in wild type, K26A and R79A (Figure 5.13). But these structures reside closely at the edge of the cell. In the cross sectional view, a few hexagon-like structures are evident near the edge of the bacterial cell, but these disperse away and appear as short filaments (Figure 5.14). It can be surmise that the mutational change to alanine of Val-51 suggests that there is a hint of conservation of the self-assembly of the wild type PduA. Slight perturbation can be observed in the cross-section. It is possible that V51A forms into short sheets but at the edge of the cell, it is able to roll-up to a tube-like formation. Interestingly, V51A preferably found at the edges or corners of the bacterial cell.

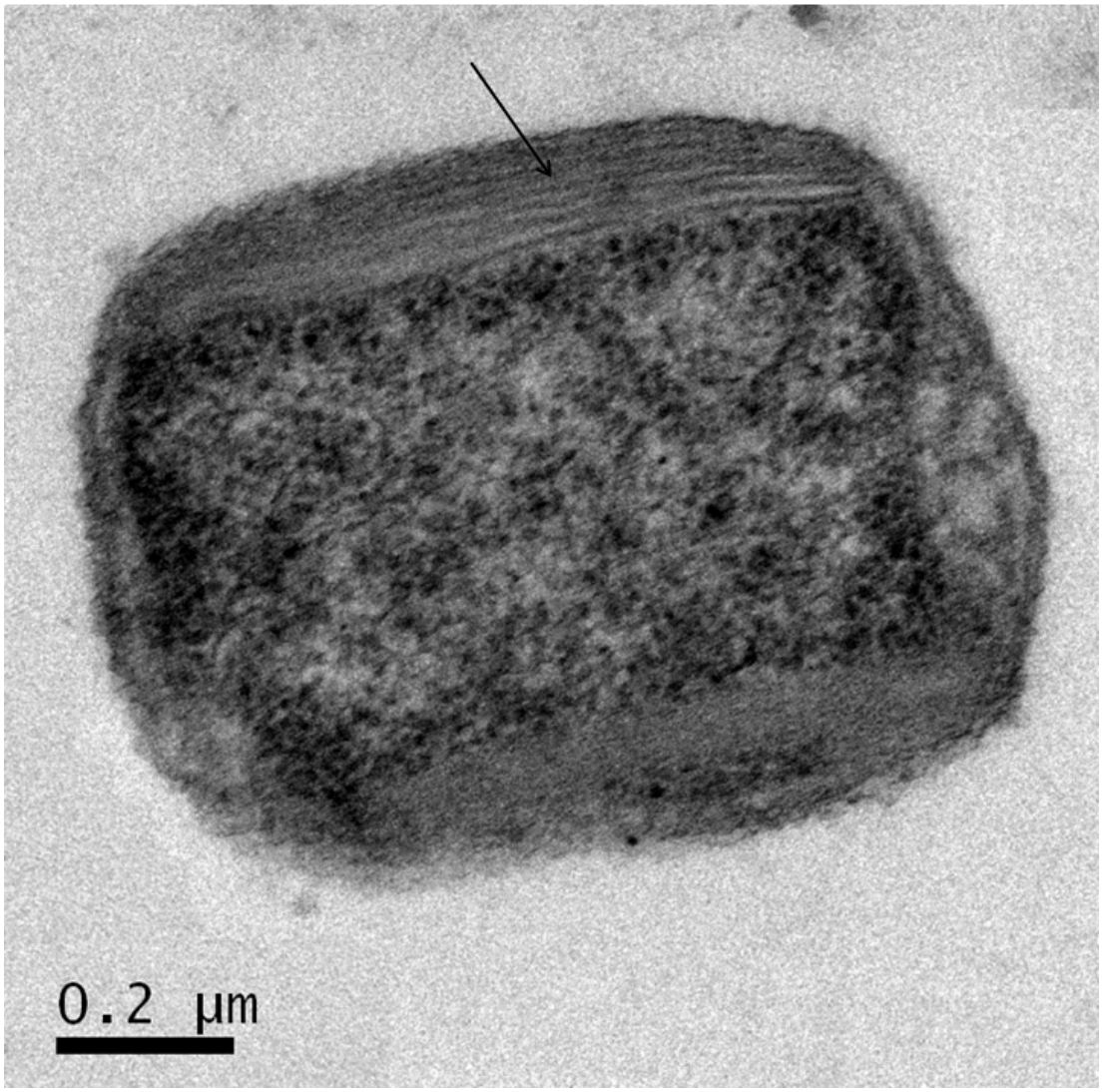


Figure 5.13 Thin section of *E. coli* overproducing V51A in transverse section. Arrow pointing at the filamentous structures.

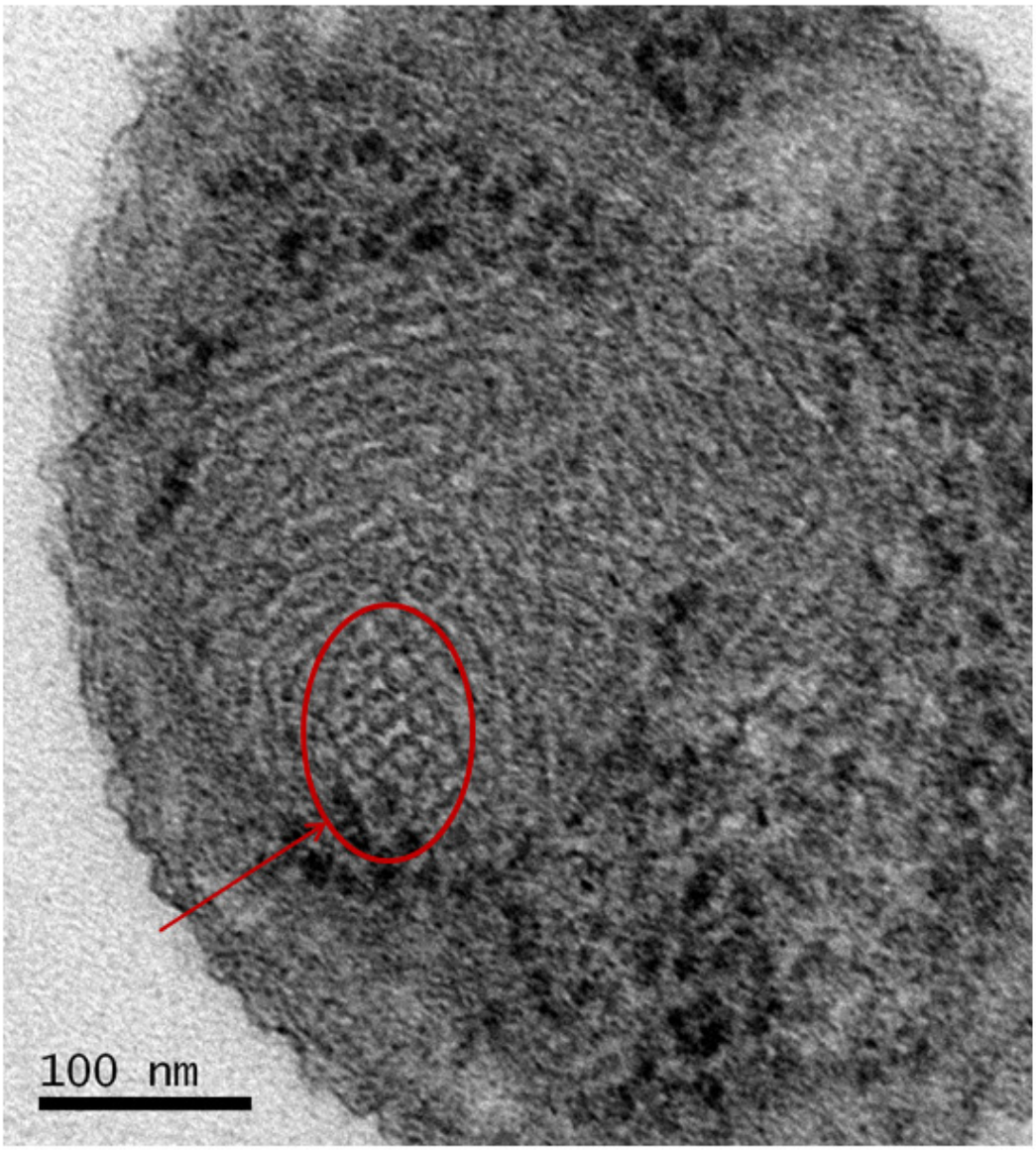


Figure 5.14 Thin section of *E. coli* overproducing V51A in cross section. Hexagon-like structures encircled in red and pointed with red arrow.

5.5 Conclusion and future work

The crystal structures of major shell components of carboxysomes (CsoS1A, CsoS1C, CcmK1, CcmK2) as well as pdu metabolosome (PduA) reveal a conserved lysine residues found in the surface of the hexamers. These surface lysine residues are able to form hydrogen-bonding with adjacent lysines. This interaction is deemed to play a key role in associating neighbouring hexamers to form a tight molecular tile. While the structural information suggests this is convincing, this has not been shown experimentally by mutation. This research paves the way to understanding how shell proteins associate with one another to form the facets of the microcompartment. The careful analysis of the surface of PduA hexamers indicate that the lysine residue is not the sole player in stabilising the molecular tile. There are at least three surface interaction hotspots, of which, in addition to Lys-26, Val-51 and Arg-79 (found in PduA) had been picked as the starting point for mutational analysis. More importantly, these residues were also found to be conserved in other pdu shell proteins, highlighting the importance of these surface residues. Mutational change of Lys-26, Val-51 and Arg-79 into either small alanine residue or negatively charged aspartate residue resulted to drastic changes both in crystal morphology as well as the shape of the structures found within the bacterial cell. These observations suggest that these residues play a role in the formation of molecular tiles. It will be interesting to see as to what extent this mutational change can affect the metabolosome shell. Would the replacement of the native PduA with PduA mutants in the pdu operon prevent the formation of the metabolosome shell? Will this create either subtle or huge distortion in the icosahedral structure?

Our preliminary interpretation of the electron micrographs of PduA suggests that the shell proteins self-assemble into long tubular filaments bundled together to create a hexagon-like lattice when viewed cross-section. This, however, requires more proof and verification. These filamentous structures can be carefully extracted from the bacterial cell and imaged using transmission electron microscopy (TEM), scanning electron microscopy (SEM) and atomic force microscopy (AFM) to identify together PduA hexamers. These techniques can also be applied to mutant shell proteins to identify how the single mutation drastically changes the morphology of the structures *in vivo*. Further to this, fiber diffraction could be used to identify the repeats of the sheets and/or filaments.

If indeed PduA forms tubular filaments within bacterial cell, this opens up several exciting possible applications. The electron microscopy of PduA structures (as well as mutants) suggests that the shell proteins were seemingly sturdy structures, which effectively can affect cell shape. Thermal melt CD experiments using the shell protein indicate that it is able to maintain secondary structure at high temperature, 88℃ (data not shown). These preliminary characteristics seemingly allow PduA shell proteins to be used as an alternative to carbon nanotube, which are applied in several areas of nanotechnology, electronics, optics and material science.

.

Chapter 6

The future of bacterial microcompartments

6.1 Introduction

Bacterial microcompartments are large icosahedral macromolecular complexes found in several bacterial cells. These supramolecular assemblies are surrounded by a proteinaceous shell, which encloses a series of sequentially acting enzymes presumably in order to sequester labile and toxic intermediates of a metabolic process within the cell. Investigation of the individual shell proteins begins to reveal how the microcompartment shell is assembled and how the shell proteins facilitate transport of various molecules into and out of the microcompartment. This Thesis contributes to this growing wealth of knowledge with the recent discovery of the crystal structures of PduT (Pang et al., 2011) (Chapter 3) and PduB (Chapter 4), as well as insight to the tiling property of PduA (Chapter 5). This final Chapter is a summary of this research project as well as an overview of key discoveries in the field of bacterial microcompartment, and research avenues that leads to the eventual

goal of engineering bacterial microcompartments with novel functionalities for applications ranging from metabolic engineering to targeted drug delivery.

6.2 Shell proteins: Structure, Assembly and Function

Shell proteins across different types share a common protein fold known as the BMC domain. The emergence of crystal structures of different shell proteins elucidates key elements underlying the architecture and transport mechanisms. The BMC domain consists of about 90 amino acid residues, with characteristic α/β fold and with central beta sheets surrounded by short α-helices. Typically, six subunits of BMC shell proteins come together and assemble into hexamers with cyclic six-fold symmetry. These hexamers are thought to form the basic building blocks of the microcompartment shells. Crystal structures of shell proteins exhibit two distinct features; on one side, the BMC shell appears to be relatively flat. On the other side, it exhibits a prominent bowl-shaped depression, where the N- and C-termini of the protein are located. Sequence analysis of several shell proteins indicates that N-terminal and more prominently the C-terminal region are varied and diverse. These regions are typically flexible as they are generally not defined in the crystal structures.

A number of crystal structures of BMC shell proteins demonstrate the side-by-side packing of hexamers, forming continuous tightly packed molecular layers in the crystal. Electron microscope studies of shell proteins embedded in grids both *in vitro* (Dryden et al., 2009) and *in vivo* (Parsons et al., 2010a) displayed similar arrangement as observed in three-dimensional crystals. This organization has been postulated to represent the packing arrangement of hexamers within the facets of the

icosahedral microcompartment shells. A close inspection of the packing arrangement of major shell components such as CsoS1A (Tsai et al., 2007) and PduA (Crowley et al., 2010) within crystal lattice revealed a conserved lysine residue located in each BMC domain; this conserved residue faces in anti-parallel fashion to its corresponding residue in its neighbouring hexamer. This is suggested to be responsible for holding these hexamers together in a molecular layer. A careful analysis using a number of surface interaction tools indicated that neighbouring hexamers were seemingly held together, not only by the conserved lysine residue but by a more extensive surface region (Chapter 5). Three residues were selected, representing the three surface regions and mutated to either valine or aspartate. *In vitro* experiment based on crystal morphology and *in vivo* experiment based on electron micrograph of *E. coli* cells showed characteristics that deviate from that of wild type. A number of experiments may be carried out in the future to further support the current experimental results; this may be in the form of successful characterisation of crystal structures of the mutant and/or the packing of two dimensional layers of purified mutant proteins (Dryden et al., 2009).

The first reported crystal structure of carboxysome shell proteins, including major shell component of respective shell types were assembled by BMC folds in what has been accepted as the 'canonical form'. Most canonical BMC domains assemble in such a way a central pore is evident; which has been suggested to play a role in the movement of molecules into and out of the metabolosome. Some shell proteins, however, showed surprising conformational and topological variations within the BMC fold. These shell proteins displayed similar tertiary structure with the secondary structure elements occurring in a circularly 'permuted' order in the protein

chain. PduU was the first known shell protein that exhibits circularly permuted version of the BMC domain (Crowley et al., 2008). This rearrangement makes it feasible to form extended C-terminal tail to create β-barrels that seemingly block the central pore. Further to this, one shell protein, EutS, was bent at approximately 40°, suggesting its unique role and position in the edge of an icosahedral assembly (Tanaka et al., 2010). Another structural variation among shell proteins is the duplication of BMC domain within its subunit. In this case, three copies of the protein assemble to a trimeric structure with pseudo six-fold symmetry. Current crystallographic data revealed that trimers may consist of either duplicated canonical or circularly permuted BMC domains. The duplication of BMC domain provided trimeric shell proteins with unique role in transport. A number of shell proteins with duplicated BMC domains have been reported to have two conformations: an open and a closed form; implying the possibility of gated transport that could give a hint on how large molecules cross the shell (Takenoya et al., 2010, Tanaka et al., 2010, Klein et al., 2009). In addition, the duplication of BMC domain generates three small pores (instead of a central pore); with each pore located within the subunit (Sagermann et al., 2009, Takenoya et al., 2010, Tanaka et al., 2010, Heldt et al., 2009). The subunit pores of the trimer have been shown to play a role as a transit route for metabolites such as glycerol (Chapter 4).

Previous genetic and biochemical studies on *Lactobacillus reuteri* identified that the *pdu* metabolosome of this species is able to metabolise both glycerol and 1,2-propanediol within the microcompartment; and therefore requiring at least one shell protein to mitigate the transit process. This is now confirmed to be the role of PduB, where glycerol metabolites were seen caught in transit passing through the subunit

pores. These experimental data paint a better picture of how these pores are used for metabolite transit. Glycerol is analogous to other metabolites such as 1,2-propanediol and ethanol. The size and shape of the subunit pore is suitable for 1,2-propanediol transport, but whether PduB mitigates such transfer has yet to be proven experimentally. Crystallographic data of PduB from various species may be necessary to identify whether glycerol channelling is unique to *Lactobacillus*. In another case, a crystal structure of trimeric shell protein, PduT (Chapter 3), displayed a large central pore region that is, consistent with the mutagenic study and EPR results (Parsons et al., 2008), suspected to be a binding site for a 4Fe-4S cluster. The shell protein may have the capability of channelling electrons out of the microcompartment via electron transfer from an associated enzyme. The proposed unique function of PduT is illustrated in Figure 6.1. The shell protein apparently accepts one unwanted electron for each PduS corrin reduction. As part of the outer shell of the supramolecular metabolosome complex, PduT may be able to transfer electrons onto other acceptors found in the bacterial cytoplasm. It is still a puzzle why PduT possess a rather unstable 4Fe-4S cluster within its pore region. This is most likely involved with the electron transfer mechanism. In this sense, it can be hypothesised that the free iron atom of the 4Fe-4S cluster may act as an electron acceptor by pointing towards the enzyme PduS; once the electron is accepted, the free iron atom points out to the cytoplasm ready to donate this to unknown acceptor outside of the metabolosome. To prove this functional role of PduT, an extensive effort has to be carried out; this includes obtaining the structure of PduT with an intact 4Fe-4S cluster and a careful biochemical and biophysical study that will show dynamic electron transfer action of PduT with PduS.

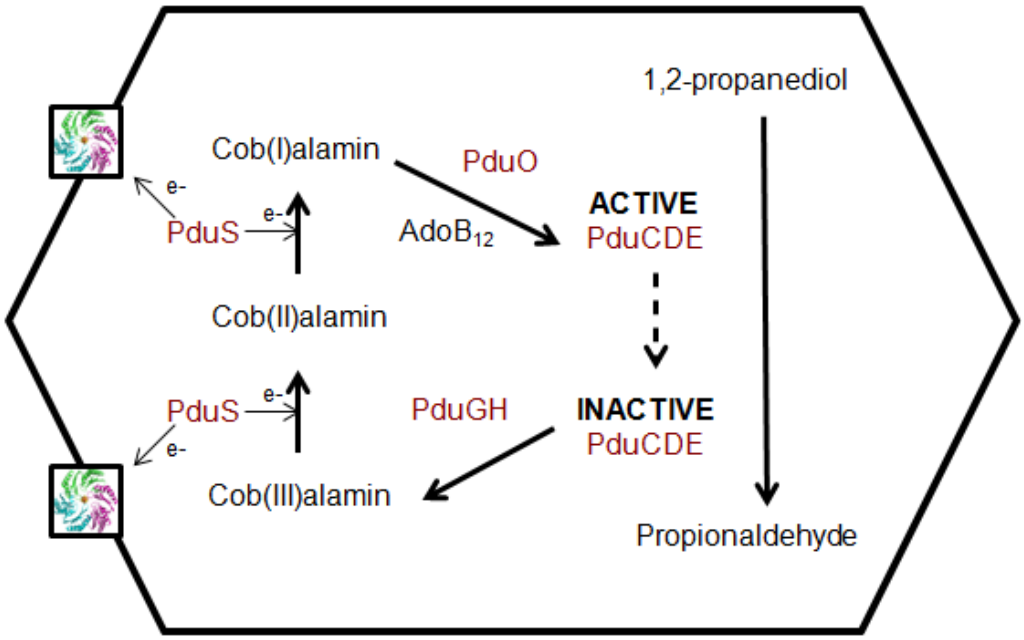


Figure 6.1 Regeneration of $AdoB_{12}$ from cob(III)alamin. In the 1,2-propanediol utilisation metabolosome of *Citrobacter freundii*, the AdoB12 is occasionally converted to cob(III)alamin in a catalytic by-reaction inactivating diol dehydratase (PduCDE). The cob(III)alamin is released from PduCDE by a reactivase PduGH and reduced to cob(I)alamin by two sequential one-electron reductions by corrin reductase PduS. This leaves one electron that may pass out of the metabolosome via the trimeric shell protein PduT . The ATP:cob(I)alamin adenosyltransferase (PduO) converts cob(I)alamin to active cofactor ($AdoB_{12}$), activating the diol dehydratase (PduCDE) to allow conversion of 1,2-propanediol to propionaldehyde.

6.3 Pdu shell: challenges and more research avenues

The 1,2-propanediol utilisation metabolosome is so far the most complicated shell complex known and well-studied. The microcompartment shell is composed of seven different types – PduA, PduB, PduJ, PduK, PduN, PduT and PduU. At the time of writing, four of these have structural information from diverse bacterial

species – PduA (*Salmonella*) (Crowley et al., 2010), PduB (*Lactobacillus*), PduT (*Salmonella* and *Citrobacter*) (Crowley et al., 2010, Pang et al., 2011) and PduU (*Salmonella*) (Crowley et al., 2008). Two of these crystal structures (PduB and PduT) were discussed in detail in this Thesis.

Microcompartment shells share a similar icosahedral assembly to the viral capsid. Three-dimensional crystal structures of a number of viral capsids from different families and genera of viruses have been solved to near-atomic resolution (Natarajan et al., 2005). This of course, provides an enormous amount of information, particularly how individual viral protein interacts to coat the capsid. This approach seems to be transferrable and applicable in exploring the structure of whole microcompartment shell and will be useful as it will confirm how individual shell proteins interact to form the icosahedral assembly. However, so far, this is yet to be achieved. The main hindrance is that pleiomorphic bacterial microcompartments, in particular the *pdu* metabolosome, are produced within the cells. Homogeneity in the sample is one of the main requirements to achieve a good crystal quality. A current goal is to obtain a homogenous sample in reasonable quantity suitable for crystallisation.

While crystallisation of the whole microcompartment shell is yet to be achieved, we have to be contented with the 'divide and conquer' technique. Efforts were placed in achieving crystal structures of the remaining shell proteins – PduJ, PduK and PduN.

PduJ shares about 83% sequence identity to the recently solved crystal structure of PduA (Crowley et al., 2010). It is expected with such high sequence identity that the structure of PduJ will appear closely similar to that of PduA. Obtaining the crystal

structure of PduJ (*Citrobacter*) was attempted but was unsuccessful; primarily because the overexpressed shell proteins reside in the insoluble fraction after cell lysis. In a different construct of PduJ (*Lactobacillus*), a slightly more soluble shell protein was attained; however, the concentration was not thought to be sufficient for crystallisation. In order to progress with this project, it may be wise to develop a methodology to extract PduJ from inclusion bodies or perhaps affix tags that will help in solubilisation. How crucial obtaining the structure of PduJ is not known, but the high sequence similarity of PduA enabled generation of theoretical model of the shell protein (Figure 6.2). Comparison of the theoretical model PduJ and PduA structure (as well as their sequences) reveals five residues that are not conserved and located in the surfaces of both the concave and convex faces. These residues may provide difference in how PduA and PduJ interact or relate to the enzymes located inside and outside of the microcompartment.

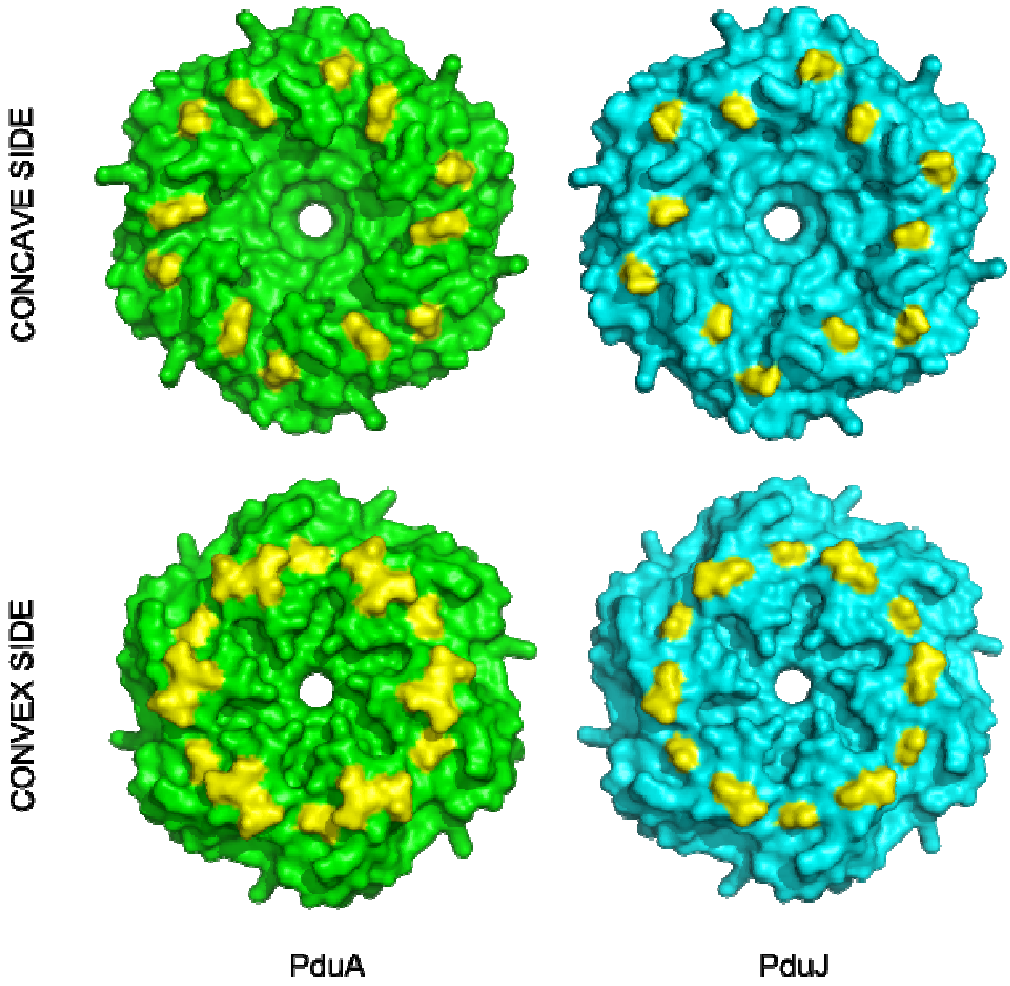


Figure 6.2 Comparison of crystal structure of PduA and theoretical model of PduJ. The theoretical model of PduJ was generated using Swiss-Model (Kiefer et al., 2009). The quality of the model is as follows: QMEAN (Qualitative Model Energy Analysis) score of 0.82 and Z-score of 0.63. QMEAN (Benkert et al., 2008) ranges between 0 and 1 with higher values for better models. Models of low quality are expected to have strongly negative Z-scores (Benkert et al., 2011). PduA and PduJ were coloured green and cyan, respectively. Significantly different residues between PduA and PduJ are coloured yellow. Figure was created using PyMOL (DeLano and Lam, 2005).

Sequence comparison showed that PduK has an extended C-terminal region (80 – 156) that is not seen with other Pdu shell proteins. Sequence analysis of this C-terminal region shows no resemblance to currently available structural data. For this, it was theorised that this may play a unique role for the shell protein. It has been hinted that the C-terminal region has a Fe-S clustering binding motif (Crowley et al., 2010). In another observation, a homologous shell protein EutK also has an extended C-terminal region. The crystal structure of the C-terminal of EutK identified that this region bears a helix-turn-helix motif, a common feature of DNA-binding protein (Tanaka et al., 2010). Whether PduK is able to bind Fe-S cluster or DNA is yet to be resolved. The crystal structure of PduK will answer this ambiguity. The shell protein was successfully produced and purified in the lab; unfortunately, this did not produce any crystals.

Based on sequence analysis, PduN falls on the same family of non-BMC EutN/CcmL family. Previously preliminary investigation suggests that PduN is essential for microcompartment formation and deletion of this shell protein prevents closure of the icosahedral structure indicating that it plays a structural role in the vertices of the shell (Parsons et al., 2010a). Consistent with this, the crystal structures of CcmL and CsoS4 revealed a pentameric assembly that could occupy the vertices of the icosahedral carboxysome (Tanaka et al., 2008). However, the emergence of crystal structure of EutN showed that the homologous shell protein assembles into hexameric structure (Forouhar et al., 2007, Tanaka et al., 2008) instead of the reported pentameric shell proteins of carboxysome. This, now, creates an uncertainty for PduN; is the shell protein a pentamer or hexamer? The crystal structure of PduN will help resolve this uncertainty. I was able to produce, and

purify PduN enough to make a number of small crystals. However, bigger crystals were not successfully obtained. These crystals appear to have hexagonal morphology similar to that which has been observed with other shell proteins. PduN crystals were taken to synchrotrons; but they diffracted to very low resolution. If crystallographic approach unable to produce an atomic model of PduN, a complementary technique such as Small-angle X-ray scattering (SAXS) might give better insight into the oligomeric structure of PduN.

The availability of crystal structures of some of the *pdu* shell proteins provides us a starting step to understand how one shell protein type relates to another in terms of building the icosahedral organelle. Crystal structures show that the individual shell proteins are assembled side-by-side to form either uniform molecular layers or layers of shell proteins in opposite orientation in the crystal lattice. From the multiple crystal structures, one can surmise that this packing is governed by the conserved lysine residues found in the interface of the shell protein. These lysine residues are held together by intermolecular hydrogen bonds; as a result, the shell proteins are tightly packed together, as evident in main shell component PduA (Crowley et al., 2010). The space between the conserved lysine residues in the interface of PduB, PduT and PduU were too distant to form hydrogen bonds and the crystallographic packing of these shell proteins is much looser. It is presumed that the tight packing of shell proteins is a much more close approximation to the shell protein hexamer contacts *in situ*. Using this packing information, the three shell proteins (PduB, PduT and PduU) have the right size and shape to fit in a layer of PduA hexamers (Figure 6.3). However, it has been observed that some of the lysine residues of PduT trimer were too far apart to hydrogen bond in this kind of assembly (Pang et al.,

2011, Crowley et al., 2010). This observation is consistent with the pull down assays; PduB and PduU were shown to interact with PduA while PduT did not (Parsons et al., 2010a). Although the experimental result is consistent with the current model of molecular tiling, a more concrete result is yet to be established. To be conclusive, crystal structures of two or more interacting shell proteins should be obtained.

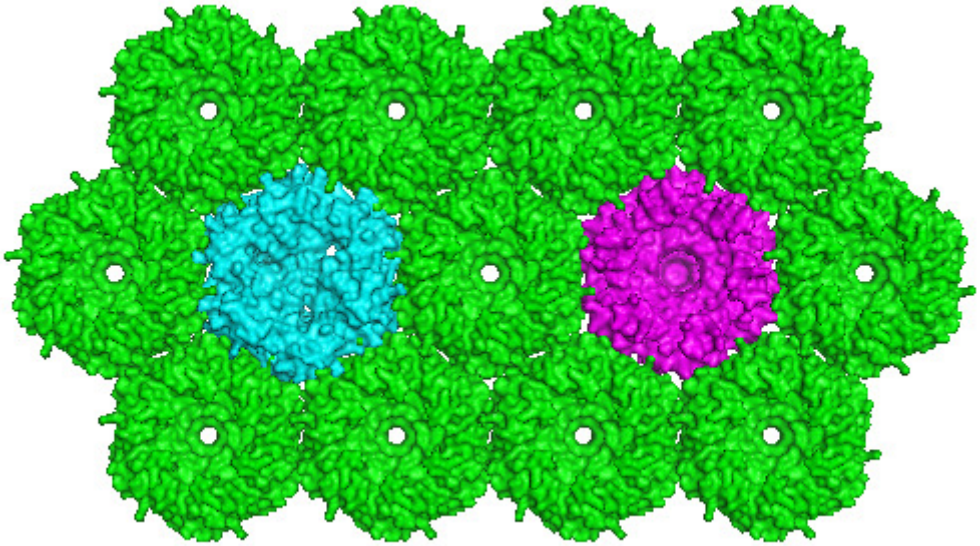


Figure 6.3 Modeling of the facet of the shell based on the crystal structures of PduA, PduB and PduU. PduB trimer (cyan) and PduU hexamer (magenta) were perfectly fitted in a sheet of PduA hexamers (green). The figure was generated using PyMOL (DeLano and Lam, 2005).

Equally important to understanding the shell composition of *pdu* metabolosome is the understanding of the internal organisation of the encapsulated enzymes. The *pdu* shell houses at least nine enzymes that are responsible for conversion of 1,2-propanediol. The encapsulation of these enzymes must rely on interaction and

recognition between the BMC shell components. Recent interaction studies between enzymes and BMC shell components begin to elucidate this research area (Parsons et al., 2010b). For instance, PduT was identified to be an interacting partner of PduS to play a role in the regeneration of $AdoB_{12}$ from cob(III)alamin (Parsons et al., 2010b). In addition, some enzymes such as PduP have an N-terminal amino acid sequence that is crucial in targeting and localising proteins inside the BMC (Sampson et al., 2002, Luo et al., 2010, Leal et al., 2003, Fan et al., 2010). Crystallisation of enzyme/shell complexes is currently in progress to clarify the recognition mechanism as well as identify which side of the shell is facing inward to the BMC.

6.4 Applications of bacterial microcompartment

One of the challenges in the field of metabolic engineering and systems biology is to control reaction pathways occur within the cell. Ironically, the main role and function of bacterial microcompartments is to exactly do this; the organelles allow conversion of novel molecules, improved metabolic reaction efficiency, sequestration of toxic intermediates, and shielding labile intermediates from possible side reactions. As a result, there has been an increased interest to better understand BMCs to attempt to re-engineer the organelle for novel functionalities. Not only bacterial microcompartments are attractive due to their functional significance but more importantly in practical terms, production of useful and important products via bacterial system will be much more efficient and cheaper than traditional chemistry techniques. In theory, BMCs can be re-engineered such that they are able to metabolise non-native molecules in a novel multi-step pathway, without the intermediates being used up in branching reactions. Since the reaction pathway is

independent of the cell, one can even re-engineer BMC to carry out reaction pathways that generate cytotoxic intermediates. There are at least two important aspects in generating synthetic BMC. First aspect is that non-native enzymes need to be targeted inside the microcompartment. It has been identified that the enzymes of BMCs bear an N-terminal region, which allow targeting and localisation of these into the microcompartment (Fan et al., 2010). This has been successfully illustrated by fusing the peptide with fluorescent protein (Fan et al., 2010). The second aspect is to re-engineer the shell protein enclosing the microcompartment to tailor to specific novel function. Crucial to this, of course, is that the pores have to be specific and selective to which molecules are allowed to enter (starting material of the metabolic process) and desirable products that need to be pushed out. In connection to this, the recent crystal structure of PduB provides a preliminary insight into this (Chapter 4). Pore lining residues that are important for metabolite binding can be re-engineered.

BMCs can be re-engineered to be used for a wide range of applications. This Chapter focuses on three important problems that need to be addressed. These are energy production, bioremediation and medical tools.

Many automobiles and vehicles rely on petroleum-based fuels. However, petroleum is not a viable long term fuel since it is not a renewable source and at the same time, contributes to global warming. A greener alternative to this are biodiesels, which consist of combustible molecules and generate a convertible CO_2 that can be processed by other organisms such as bacteria and yeast. However, biodiesels suffer from many disadvantages such as being more corrosive as well as offer lower energy density. To be more efficient and suitable for engine, the fuel has to be composed of

alkanes, the main class compounds in diesel. Alkanes can be produced in a two-step enzymatic reaction using fatty acyl-ACP as the starting material (Schirmer et al., 2010). The pores of the shell might be re-engineered such that it attracts the starting material onto the microcompartment. Fatty acyl-ACP is converted to alkanes by acyl-ACP reductase and aldehyde decarbonylase with an intermediate of fatty aldehyde. CO_2 can be released out of the microcompartment, which can be used in a different microbial process while alkanes are kept inside the microcompartment to prevent further reaction; and these can be isolated out of the bacterial cells to be used or the bacterial cells could be used directly (Figure 6.4).

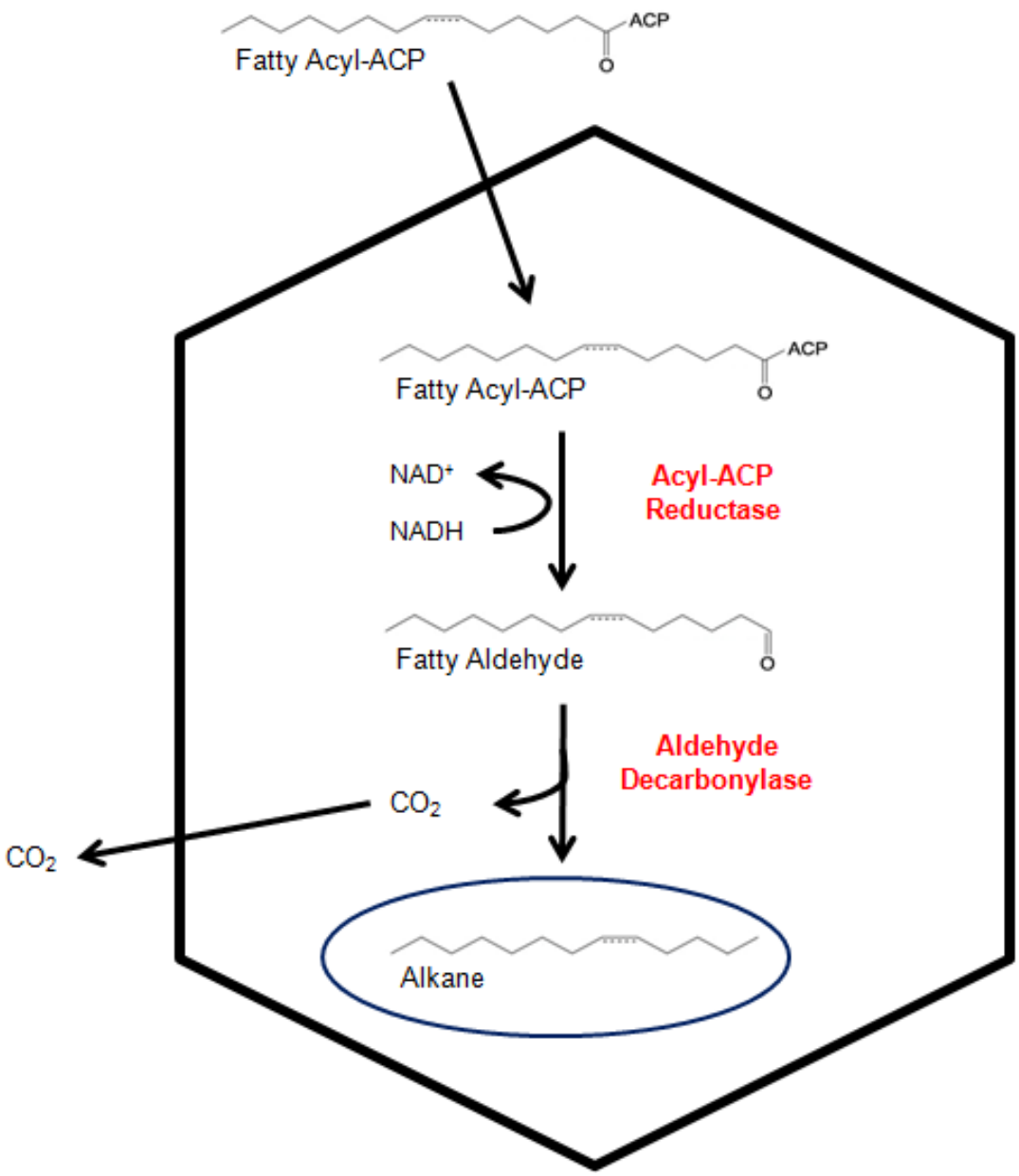


Figure 6.4 Microbial alkane production enclosed in bacterial microcompartment. Enzymes were marked with red. The product of interest, alkane, is encircled in blue.

Hydrogen gas is known to be the greenest energy source, since it releases a lot of energy for its weight with the reaction forming only energy and pure water as its product. However, hydrogen gases are difficult to store and highly reactive to oxygen causing an explosive chain reaction. This is how BMCs can be of used. BMCs can be re-engineered to sequester the reactive hydrogen gas away from environment with oxygen. It was previously reported that a series of steps consisting of 13 enzymes is able to produce high hydrogen yield from starch and water (Zhang et al., 2007). These 13 enzymes can be encapsulated to further increased reaction efficiency as well as control hydrogen reactivity (Figure 6.5).

BMCs could be engineered to trap various enzymes within the microcompartment. These enzymes are able to confer special function; for instance, useful for bioremediation. A group in University of Dundee suggested localising iron-rich proteins such as ferritin and HybA inside the microcompartment, generating a magnetic organelle (iGEM, 2011). Another application is to localise arsenic binding protein (fMT) in the microcompartment, which in principle, could play a role in the purification of water contaminated with heavy metals. The presence of BMCs in several bacterial species hinted that BMCs are susceptible to horizontal gene transfer; therefore, releasing genetically modified bacteria with synthetic BMC may not be appealing. Current research is underway in isolating and/or producing stabilised BMC outside the cell to serve as *ex vivo* reaction chambers.

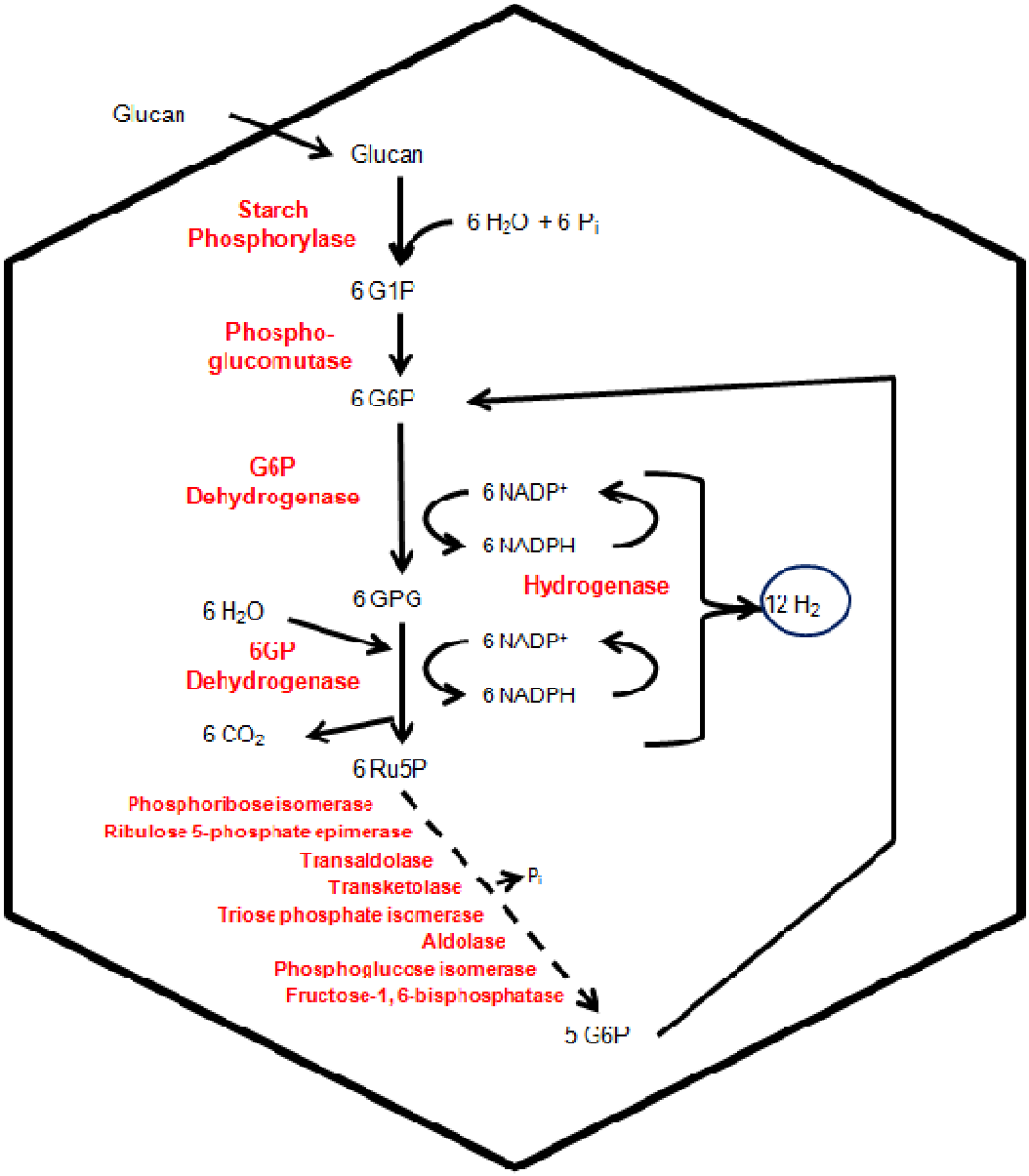


Figure 6.5 Proposed high yield production of hydrogen enclosed in microcompartment. The 13 enzymes were marked with red. Hydrogen usable as energy source is encircled in blue.

Synthetic BMCs can also be generated to be used for medical purposes; similar to how viral particles (Goicochea et al., 2007, Sun et al., 2007, Aniagyei et al., 2008), micelles (McPherson, 2005, Chen et al., 2006, Chen et al., 2005) and vault nanoparticles (Buehler et al., 2011). Some of the applications of this include using BMCs as spectroscopic probes and delivery of cytotoxic drugs to targeted cells. Metal nanoparticle probes can be encapsulated in a protein shell. These probes can be used for optical and magnetic imaging of tumours. To create the synthetic BMC, two things are needed: positively charged tails are added to the self-assembling protein and the metal nanoparticle probes are coated with negatively charged carboxylated PEG molecules. Due to the electrostatic interaction, the positively charged tails of BMC protein will self-assemble around the negatively charged nanoparticle probes (Figure 6.6A).

Antineoplastic drugs usually lead to several unintended side effects, especially since many of these drugs are cytotoxic, harming even normal cells whilst being transported to the target site. BMCs can be re-engineered to carry these drugs and be sequestered from normal cells. To do this, the interior surface of the BMC shell is re-engineered with cysteine thiol groups, which attracts and serve as an attachment to the reactive analogs of the drugs (Figure 6.6B). Since the cysteine thiol groups are artificially introduced to the surface of BMC interior, we have the ability to control the dosage; in addition, we can potentially attach multiple drugs in combination for additive or synergistic effects of drugs useful for cancer therapy. These synthetic BMCs can be targeted to cancer cells by modifying the surface of the outer protein shell layer with cysteine thiols, which small molecule such as folate can affix into. Folates are able to bind specific receptors which were shown to be upregulated in

rapidly dividing tumour cells (Coney et al., 1991, Sega and Low, 2008, Campbell et al., 1991). This, as such, is a good detecting mechanism and could specifically target cancer cells, where the cytotoxic drugs the BMC is carrying are released into (Figure 6.6B).

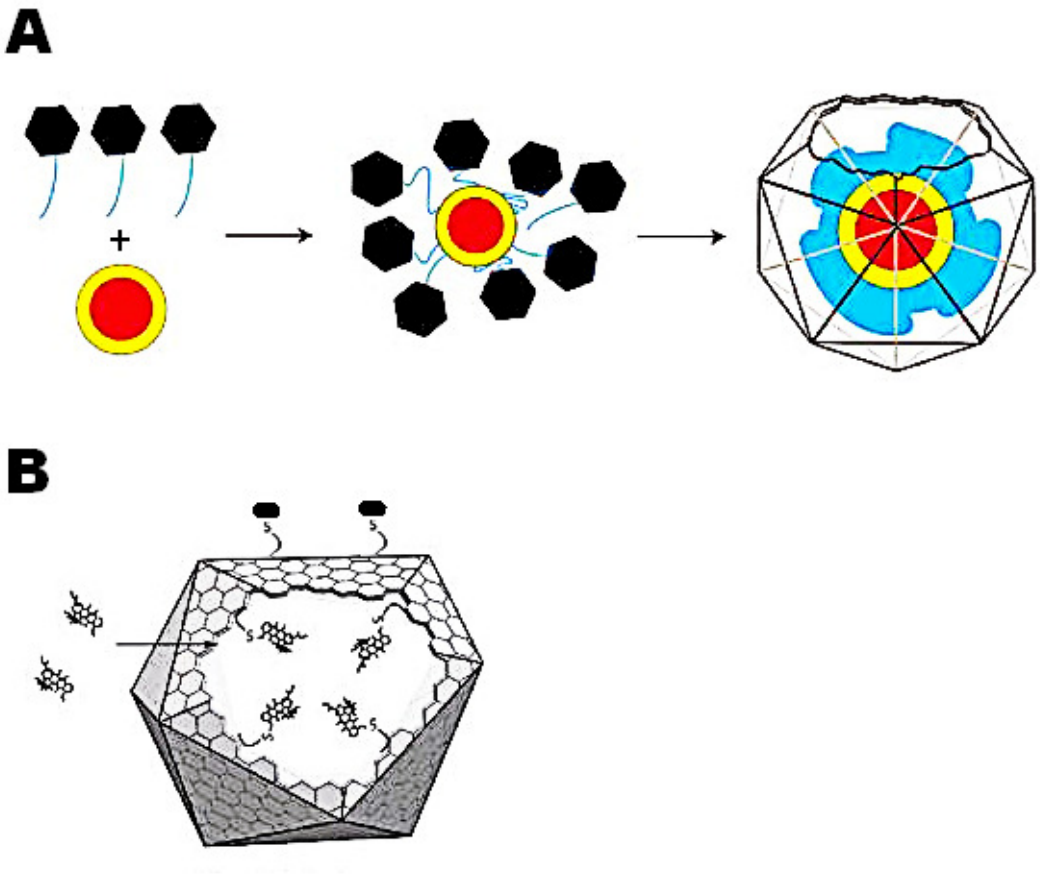


Figure 6.6 Application of BMCs for medical purposes. **(A)** The metal nanoparticle, shown as red circle, is coated by negatively charged polyethylene glycol (yellow circle). The hexamer subunits (black hexagons) are engineered with positively charged light blue tails. The positively charged tails would be drawn to the negatively charged PEG coat to form a BMC shell around the nanoparticle. **(B)** Cysteine thiols are shown engineered on surfaces of both interior and exterior of the BMC shell. A microcompartment is shown in light gray with a cutaway showing the inside of the BMC. Doxorubicin molecules, an example of anti-cancer drugs, are shown in black, attached to the surface of the interior of microcompartment through reaction with thiol groups. On the surface of the BMC exterior, the thiol groups can be attached to cancer cell detectors such as folate. This figure was modified and adapted from (Tsai and Yeates, 2011).

6.5 Final remarks

The availability of atomic resolution structures of various shell proteins has accelerated our understanding of shell assembly. Using this structural information, we can now begin to do mutational studies in attempt to further understand the shell assembly. Despite of the increasing research output on the area of metabolosomes, we are yet to be in the position to validate and generate synthetic BMCs useful for different applications. We are still in the infancy stage, so much had been discovered, but there is much more to learn.

Appendices

A1 Polymerase Chain Reaction

PCR was carried out in a machine (BioRad C1000 thermal cycler), which controls the heating and cooling of the PCR reaction mixture in a precise temperature and time. In the generation of mutant shell protein constructs, PCR reaction was made as stated in Table A1 and carried out as shown in Table A2.

Table A1 PCR reaction mixture. For a 50 µl reaction mixture, the following reagents are added. Taq polymerase enzyme was placed last.

5x HF Buffer	10 µl
10 mM dNTP mix	1 µl
Forward primer	1 µl
Reverse primer	1 µl
Template	1 µl
DMSO	1.5 µl
Phusion Taq polymerase	0.5 µl
ddH_2O	34 µl

Table A2 PCR programme used to synthesise PduA mutant shell protein constructs. PCR ran for 30 cycles. One complete cycle is equivalent to completed denaturation, annealing and extension steps.

Steps	Temperature	Duration
Initial denaturation	98℃	120 seconds
Denaturation	98℃	10 seconds
Annealing	55℃	60 seconds
Extension	72℃	150 seconds
Final extension	72℃	600 seconds
Finished PCR	12℃	∞

A2 Peer-reviewed PduT paper

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structure of PduT, a trimeric bacterial microcompartment protein with a 4Fe–4S cluster-binding site

Allan Pang,[a] Martin J. Warren[b] and Richard W. Pickersgill[a]*

[a]School of Biological and Chemical Sciences, Queen Mary University of London, Mile End Road, London E1 4NS, England, and [b]Centre for Molecular Processing, School of Biosciences, University of Kent, Giles Lane, Canterbury, Kent CT2 7NJ, England

Correspondence e-mail: r.w.pickersgill@qmul.ac.uk

Received 19 October 2010
Accepted 1 December 2010

PDB Reference: PduT, 3pac.

Propanediol metabolism in *Citrobacter freundii* occurs within a metabolosome, a subcellular proteinaceous bacterial microcompartment. The propanediol-utilization (Pdu) microcompartment shell is constructed from thousands of hexagonal-shaped protein complexes made from seven different types of protein subunit. Here, the structure of the bacterial microcompartment protein PduT, which has a tandem structural repeat within the subunit and forms trimers with pseudo-hexagonal symmetry, is reported. This trimeric assembly forms a flat approximately hexagonally shaped disc with a central pore that is suitable for a 4Fe–4S cluster. The essentially cubic shaped 4Fe–4S cluster conforms to the threefold symmetry of the trimer with one free iron, the role of which could be to supply electrons to an associated microcompartment enzyme, PduS.

1. Introduction

Bacterial microcompartments are polyhedral cellular inclusions that consist of a protein shell that encloses a specific metabolic process. The best characterized of these is the carboxysome (Tanaka *et al.*, 2008), which houses the enzymes ribulose bisphosphate carboxylase/oxygenase and carbonic anhydrase. It is thought that the carboxysome accelerates the rate of carbon fixation by increasing the local concentration of carbon dioxide. More recently, similar sequences to the bacterial microcompartment proteins of the carboxysome have been discovered in metabolic operons associated with propanediol utilization (*pdu* genes; Bobik *et al.*, 1999), ethanolamine utilization (Stojiljkovic *et al.*, 1995) and ethanol utilization (Seedorf *et al.*, 2008). In growth conditions that induce these metabolic operons, microcompartments can be seen in the cytoplasm of these bacteria. These metabolic microcompartments are known as metabolosomes (Brinsmade *et al.*, 2005; Parsons *et al.*, 2008).

The 21-gene regulon of *Citrobacter freundii* encoding the *pdu* organelle and propanediol-utilization enzymes (Fig. 1) has been cloned into *Escherichia coli*, resulting in the production of microcompartments and allowing propanediol utilization. Inside the microcompartment, 1,2-propanediol is converted into propionaldehyde by a diol dehydratase composed of PduCDE. Sequestration of propionaldehyde within the microcompartment may prevent unwanted reactions leading to growth arrest and DNA damage (Sampson & Bobik, 2008). The propionaldehyde is subsequently disproportionated into 1-propanol and propionyl-CoA by the aldehyde dehydrogenase PduQ and the CoA transferase PduP, respectively. These two products are delivered out to the cytoplasm, where propionyl-CoA is further converted into propionyl phosphate and propionate by PduL and PduW,

For full paper:

PANG, A., WARREN, M. J. & 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**,** 91-96.

A3 Peer-reviewed PduB paper

Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Substrate channels revealed in the trimeric *Lactobacillus reuteri* bacterial microcompartment shell protein PduB

Allan Pang,[a] Mingzhi Liang,[b,c] Michael B. Prentice[b,c,d] and Richard W. Pickersgill[a]*

[a]School of Biological and Chemical Sciences, Queen Mary University of London, 327 Mile End Road, London E1 4NS, England, [b]Department of Microbiology, University College Cork, Cork, Ireland, [c]Alimentary Pharmabiotic Centre, University College Cork, Cork, Ireland, and [d]Department of Pathology, University College Cork, Cork, Ireland

Correspondence e-mail: r.w.pickersgill@qmul.ac.uk

Received 9 August 2012
Accepted 14 September 2012

PDB Reference: PduB, 4fay

Lactobacillus reuteri metabolizes two similar three-carbon molecules, 1,2-propanediol and glycerol, within closed polyhedral subcellular bacterial organelles called bacterial microcompartments (metabolosomes). The outer shell of the propanediol-utilization (Pdu) metabolosome is composed of hundreds of mainly hexagonal protein complexes made from six types of protein subunits that share similar domain structures. The structure of the bacterial microcompartment protein PduB has a tandem structural repeat within the subunit and assembles into a trimer with pseudo-hexagonal symmetry. This trimeric structure forms sheets in the crystal lattice and is able to fit within a polymeric sheet of the major shell component PduA to assemble a facet of the polyhedron. There are three pores within the trimer and these are formed between the tandem repeats within the subunits. The structure shows that each of these pores contains three glycerol molecules that interact with conserved residues, strongly suggesting that these subunit pores channel glycerol substrate into the metabolosome. In addition to the observation of glycerol occupying the subunit channels, the presence of glycerol on the molecular threefold symmetry axis suggests a role in locking closed the central region.

1. Introduction

A wide variety of heterotrophic bacteria, including Enterobacteriaceae and Firmicutes, produce multiple polyhedral cellular inclusions containing enzymes when induced by specific substrates (Sriramulu *et al.*, 2008; Bobik *et al.*, 1999; Shively *et al.*, 1998). The archetypal structure of this type (a bacterial microcompartment) is the carboxysome, which is found in cyanobacteria and some chemoautotrophic bacteria (Kerfeld *et al.*, 2010) and consists of a thin protein shell enclosing the enzymes RuBisCO and carbonic anhydrase. Similar structures in heterotrophic bacteria are termed metabolosomes and are classified according to the substrate that they process (Brinsmade *et al.*, 2005; Parsons *et al.*, 2008; Heldt *et al.*, 2009). These include propanediol-utilization (Pdu), ethanolamine-utilization (Eut) and ethanol-utilization (Etu) metabolosomes. These different types of metabolosomes are presumed to share a common function, which is to bring together the enzymes and metabolites, increasing their effective concentrations (Price & Badger, 1989) and sequestering them from the bacterial cytoplasm. Physical isolation of a metabolic pathway in this way is believed to provide various advantages for the bacterial cell, including protection from

For full paper:

PANG, A., LIANG, M., PRENTICE, M. B. & PICKERSGILL, R. W. (2012) Substrate channels revealed in the trimeric Lactobacillus reuteri bacterial microcompartment shell protein PduB. *Acta Crystallographica Section D: Biological Crystallography,* 68**,** 1642-1652.

References

ADAMS, P. D., AFONINE, P. V., BUNKOCZI, G., CHEN, V. B., DAVIS, I. W., ECHOLS, N., HEADD, J. J., HUNG, L. W., KAPRAL, G. J., GROSSE-KUNSTLEVE, R. W., MCCOY, A. J., MORIARTY, N. W., OEFFNER, R., READ, R. J., RICHARDSON, D. C., RICHARDSON, J. S., TERWILLIGER, T. C. & ZWART, P. H. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D-Biological Crystallography,* 66**,** 213-221.

ANIAGYEI, S. E., DUFORT, C., KAO, C. C. & DRAGNEA, B. (2008) Self-assembly approaches to nanomaterial encapsulation in viral protein cages. *Journal of materials chemistry,* 18**,** 3763-3774.

BADGER, M. R., BASSETT, M. & COMINS, H. N. (1985) A Model for HCO3– Accumulation and Photosynthesis in the Cyanobacterium Synechococcus sp: Theoretical Predictions and Experimental Observations. *Plant Physiology,* 77**,** 465.

BEEBY, M., BOBIK, T. A. & YEATES, T. O. (2009) Exploiting genomic patterns to discover new supramolecular protein assemblies. *Protein Science,* 18**,** 69-79.

BENKERT, P., BIASINI, M. & SCHWEDE, T. (2011) Toward the estimation of the absolute quality of individual protein structure models. *Bioinformatics,* 27**,** 343-350.

BENKERT, P., TOSATTO, S. C. E. & SCHOMBURG, D. (2008) QMEAN: A comprehensive scoring function for model quality assessment. *Proteins: Structure, Function, and Bioinformatics,* 71**,** 261-277.

BEREZHKOVSKII, A. M. & BEZRUKOV, S. M. (2005) Optimizing transport of metabolites through large channels: molecular sieves with and without binding. *Biophysical journal,* 88**,** L17-L19.

BERG, J. M., TYMOCZKO, J. L. & STRYER, L. (2006) Biochemistry: international edition. WH Freeman & Co. New York.

BEUDEKER, R. F., CANNON, G. C., KUENEN, J. G. & SHIVELY, J. M. (1980) Relations between d-ribulose-1,5-bisphosphate carboxylase, carboxysomes and CO2 fixing capacity in the obligate chemolithotroph *Thiobacillus neapolitanus* grown under different limitations in the chemostat. *Archives of Microbiology,* 124**,** 185-189.

BLOW, D. (2003) Outline of crystallography for biologists.

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

BOBIK, T. A. (2007) Bacterial Microcompartments. *Microbe (Am Soc Microbiol),* 2**,** 25-31.

BOCK, E., DUVEL, D. & PETERS, K. R. (1974) Characterization of a phage-like particle from cells of Nitrobacter.1. Host-particle correlation and particle isolation. *Archives of Microbiology,* 97**,** 115-127.

BORNSTEIN, B. T. & BARKER, H. A. (1948) The energy metabolism of *Clostridium kluyveri* and the synthesis of fatty acids. *J Biol Chem,* 55**,** 223-230.

BRADFORD, M. M. (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical biochemistry,* 72**,** 248-254.

BRAHMS, S. & BRAHMS, J. (1980) Determination of protein secondary structure in solution by vacuum ultraviolet circular dichroism. *Journal of Molecular Biology,* 138**,** 149-178.

BUEHLER, D. C., TOSO, D. B., KICKHOEFER, V. A., ZHOU, Z. H. & ROME, L. H. (2011) Vaults Engineered for Hydrophobic Drug Delivery. *Small*.

BURTON, R. M. & STADTMAN, E. R. (1953) The oxidation of acetaldehyde to acetylcoenzyme. *J Biol Chem,* 202**,** 873-890.

CAI, F., HEINHORST, S., SHIVELY, J. M. & CANNON, G. C. (2008) Transcript analysis of the Halothiobacillus neapolitanus cso operon. *Archives of Microbiology,* 189**,** 141-150.

CAI, F., MENON, B. B., CANNON, G. C., CURRY, K. J., SHIVELY, J. M. & HEINHORST, S. (2009) The pentameric vertex proteins are necessary for the icosahedral carboxysome shell to function as a CO2 leakage barrier. *PLoS One,* 4**,** e7521.

CAMPBELL, I. G., JONES, T. A., FOULKES, W. D. & TROWSDALE, J. (1991) Folate-binding protein is a marker for ovarian cancer. *Cancer research,* 51**,** 5329.

CANNON, G. C., BRADBURNE, C. E., ALDRICH, H. C., BAKER, S. H., HEINHORST, S. & SHIVELY, J. M. (2001) Microcompartments in prokaryotes: Carboxysomes and related polyhedra. *Applied and Environmental Microbiology,* 67**,** 5351-5361.

CANNON, G. C., ENGLISH, R. S. & SHIVELY, J. M. (1991) In situ assay of ribulose-1,5-bisphosphate carboxylase/oxygenase in *Thiobacillus neapolitanus*. *Journal of Bacteriology,* 173**,** 1565-1568.

CANNON, G. C. & SHIVELY, J. M. (1983) Characterization of a homegenous preparation of carboxysomes from *Thiobacillus neapolitanus*. *Archives of Microbiology,* 134**,** 52-59.

CASAS, I. A. & DOBROGOSZ, W. J. (2000) Validation of the probiotic concept: Lactobacillus reuteri confers broad-spectrum protection against disease in humans and animals. *Microbial ecology in health and disease,* 12**,** 247-285.

CHANG, G. W. & CHANG, J. T. (1975) Evidence for the B12-dependent enzyme ethanolamine deaminase in *Salmonella*. *Nature,* 254**,** 150-151.

CHEN, C., DANIEL, M. C., QUINKERT, Z. T., DE, M., STEIN, B., BOWMAN, V. D., CHIPMAN, P. R., ROTELLO, V. M., KAO, C. C. & DRAGNEA, B. (2006) Nanoparticle-templated assembly of viral protein cages. *Nano letters,* 6**,** 611-615.

CHEN, C., KWAK, E. S., STEIN, B., KAO, C. C. & DRAGNEA, B. (2005) Packaging of gold particles in viral capsids. *Journal of nanoscience and nanotechnology,* 5**,** 2029-2033.

CONEY, L. R., TOMASSETTI, A., CARAYANNOPOULOS, L., FRASCA, V., KAMEN, B. A., COLNAGHI, M. I. & ZURAWSKI, V. R. (1991) Cloning of a tumor-associated antigen: MOv18 and MOv19 antibodies recognize a folate-binding protein. *Cancer research,* 51**,** 6125.

CROWLEY, C. S., CASCIO, D., SAWAYA, M. R., KOPSTEIN, J. S., BOBIK, T. A. & YEATES, T. O. (2010) Structural insights into the mechanisms of transport across the Salmonella enterica Pdu microcompartment shell. *J Biol Chem*.

CROWLEY, C. S., SAWAYA, M. R., BOBIK, T. A. & YEATES, T. O. (2008) Structure of the PduU shell protein from the Pdu microcompartment of Salmonella. *Structure,* 16**,** 1324-1332.

DARNELL, S. J., PAGE, D. & MITCHELL, J. C. (2007) An automated decisionâ€ tree approach to predicting protein interaction hot spots. *Proteins: Structure, Function, and Bioinformatics,* 68**,** 813-823.

DELANO, W. L. & LAM, J. W. (2005) PyMOL: A communications tool for computational models. *Abstracts of Papers of the American Chemical Society,* 230**,** 254-COMP.

DREWS, G. & NIKLOWITZ, W. (1956) [Cytology of Cyanophycea. II. Centroplasm and granular inclusions of Phormidium uncinatum.]. *Arch Mikrobiol,* 24**,** 147-62.

DRYDEN, K. A., CROWLEY, C. S., TANAKA, S., YEATES, T. O. & YEAGER, M. (2009) Two-dimensional crystals of carboxysome shell proteins recapitulate the hexagonal packing of threeâ€ dimensional crystals. *Protein Science,* 18**,** 2629-2635.

DUCRUIX, A. & GIEGE, R. (1999) *Crystallization of nucleic acids and proteins: A practical approach*, Oxford University Press, USA.

DUNDAS, J., OUYANG, Z., TSENG, J., BINKOWSKI, A., TURPAZ, Y. & LIANG, J. (2006) CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Research,* 34**,** W116-W118.

EDELHOCH, H. (1967) Spectroscopic Determination of Tryptophan and Tyrosine in Proteins. *Biochemistry,* 6**,** 1948-1954.

ELLIS, R. J. (2010) Tackling unintelligent design. *Nature,* 463**,** 164-165.

EVANS, G. & PETTIFER, R. F. (2001) CHOOCH: a program for deriving anomalous-scattering factors from X-ray fluorescence spectra. *Journal of applied crystallography,* 34**,** 82-86.

EVANS, P. (2006) Scaling and assessment of data quality. *Acta Crystallographica Section D-Biological Crystallography,* 62**,** 72-82.

FAN, C. & BOBIK, T. A. (2011) The N-Terminal Region of the Medium Subunit (PduD) Packages Adenosylcobalamin-Dependent Diol Dehydratase (PduCDE) into the Pdu Microcompartment. *Journal of Bacteriology,* 193**,** 5623.

FAN, C. G., CHENG, S. Q., LIU, Y., ESCOBAR, C. M., CROWLEY, C. S., JEFFERSON, R. E., YEATES, T. O. & BOBIK, T. A. (2010) Short N-terminal sequences package proteins into bacterial microcompartments. *Proceedings of the National Academy of Sciences of the United States of America,* 107**,** 7509-7514.

FOROUHAR, F., KUZIN, A., SEETHARAMAN, J., LEE, I., ZHOU, W., ABASHIDZE, M., CHEN, Y., YONG, W., JANJUA, H. & FANG, Y. (2007) Functional insights from structural genomics. *Journal of structural and functional genomics,* 8**,** 37-44.

GANTT, E. & CONTI, S. F. (1969) Ultrastructure of blue-green algae. *Journal of Bacteriology,* 97**,** 1486-1493.

GARSIN, D. A. (2010) Ethanolamine utilization in bacterial pathogens: roles and regulation. *Nature Reviews Microbiology,* 8**,** 290-295.

GOICOCHEA, N. L., DE, M., ROTELLO, V. M., MUKHOPADHYAY, S. & DRAGNEA, B. (2007) Core-like particles of an enveloped animal virus can self-assemble efficiently on artificial templates. *Nano Lett,* 7**,** 2281-2290.

HAVEMANN, G. D. & BOBIK, T. A. (2003) Protein content of polyhedral organelles involved in coenzyme B-12-dependent degradation of 1,2-propanediol in Salmonella enterica serovar typhimurium LT2. *Journal of Bacteriology,* 185**,** 5086-5095.

HAVEMANN, G. D., SAMPSON, E. M. & BOBIK, T. A. (2002) PduA is a shell protein of polyhedral organelles involved in coenzyme B-12-dependent degradation of 1,2-propanediol in Salmonella enterica serovar typhimurium LT2. *Journal of Bacteriology,* 184**,** 1253-1261.

HELDT, D., FRANK, S., SEYEDARABI, A., LADIKIS, D., PARSONS, J. B., WARREN, M. J. & PICKERSGILL, R. W. (2009) Structure of a trimeric bacterial microcompartment shell protein, EtuB, associated with ethanol utilization in Clostridium kluyveri. *Biochemical Journal,* 423**,** 199-207.

HORSWILL, A. R. & ESCALANTE-SEMERENA, J. C. (1999) Salmonella typhimurium LT2 catabolizes propionate via the 2-methylcitric acid cycle. *Journal of Bacteriology,* 181**,** 5615-5623.

INCARDONA, M. F., BOURENKOV, G. P., LEVIK, K., PIERITZ, R. A., POPOV, A. N. & SVENSSON, O. (2009) EDNA: a framework for plugin-based applications applied to X-ray experiment online data analysis. *Journal of synchrotron radiation,* 16**,** 872-879.

JOHNSON, C. L. V., PECHONICK, E., PARK, S. D., HAVEMANN, G. D., LEAL, N. A. & 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**,** 1577.

KERFELD, C. A., HEINHORST, S. & CANNON, G. C. (2010) Bacterial Microcompartments. *Annual Review of Microbiology, Vol 64, 2010.* Palo Alto, Annual Reviews.

KERFELD, C. A., SAWAYA, M. R., TANAKA, S., NGUYEN, C. V., PHILLIPS, M., BEEBY, M. & YEATES, T. O. (2005) Protein structures forming the shell of primitive bacterial organelles. *Science,* 309**,** 936-938.

KIEFER, F., ARNOLD, K., KÃ¼NZLI, M., BORDOLI, L. & SCHWEDE, T. (2009) The SWISS-MODEL Repository and associated resources. *Nucleic Acids Research,* 37**,** D387-D392.

KLEIN, M. G., ZWART, P., BAGBY, S. C., CAI, F., CHISHOLM, S. W., HEINHORST, S., CANNON, G. C. & KERFELD, C. A. (2009) Identification and structural analysis of a novel carboxysome shell protein with implications for metabolite transport. *Journal of Molecular Biology,* 392**,** 319-333.

KOCH, A. L. (1970) Turbidity measurements of bacterial cultures in some available commercial instruments. *Analytical biochemistry,* 38**,** 252.

KOFOID, E., RAPPLEYE, C., STOJILJKOVIC, I. & ROTH, J. (1999) The 17-gene ethanolamine (eut) operon of Salmonella typhimurium encodes five homologues of carboxysome shell proteins. *Journal of Bacteriology,* 181**,** 5317-5329.

KURIYAN, J. & EISENBERG, D. (2007) The origin of protein interactions and allostery in colocalization. *Nature,* 450**,** 983-990.

LAEMMLI, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature,* 227**,** 680-685.

LASKOWSKI, R. A., HUTCHINSON, E. G., MICHIE, A. D., WALLACE, A. C., JONES, M. L. & THORNTON, J. M. (1997) PDBsum: a Web-based database of summaries and analyses of all PDB structures. *Trends in Biochemical Sciences,* 22**,** 488-490.

LAWRENCE, J. V. & MAIER, S. (1977) Correction for the inherent error in optical density readings. *Applied and Environmental Microbiology,* 33**,** 482-484.

LEAL, N. A., HAVEMANN, G. D. & 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**,** 353-361.

LESLIE, A. G. W. (1992) Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography. *Acta Crystallographica Section D-Biological Crystallography,* 26.

LUO, L. H., SEO, J. W., BAEK, J. O., OH, B. R., HEO, S. Y., HONG, W. K., KIM, D. H. & KIM, C. H. (2010) Identification and characterization of the propanediol utilization protein PduP of Lactobacillus reuteri for 3-hydroxypropionic acid production from glycerol. *Applied Microbiology and Biotechnology,* 89**,** 697-703.

LURZ, R., MAYER, F. & GOTTSCHALK, G. (1979) Electron microscopic study on the quaternary structure of the isolated particulate alcohol-acetaldehyde dehydrogenase complex and on its identity with the polygonal bodies of *Clostridium kluyveri*. *Archives of Microbiology,* 120**,** 255-262.

MARCUS, Y., BERRY, J. A. & PIERCE, J. (1992) Photosynthesis and photorespiration in a mutant of the cyanobacterium Synechocystis PCC 6803 lacking carboxysomes. *Planta,* 187**,** 511-516.

MCFERRIN, M. B. & SNELL, E. H. (2002) The development and application of a method to quantify the quality of cryoprotectant solutions using standard area-detector X-ray images. *Journal of applied crystallography,* 35**,** 538-545.

MCKEE, T. & MCKEE, J. R. (2009) *Biochemistry: the molecular basis of life*, Oxford University Press.

MCPHERSON, A. (2005) Micelle formation and crystallization as paradigms for virus assembly. *Bioessays,* 27**,** 447-458.

MILLER, A. G., TURPIN, D. H. & CANVIN, D. T. (1984) Growth and Photosynthesis of the Cyanobacterium *Synechococcus leopoliensis* in HCO3−-Limited Chemostats. *Plant Physiology,* 75**,** 1064-1070.

MORITA, H., TOH, H., FUKUDA, S., HORIKAWA, H., OSHIMA, K., SUZUKI, T., MURAKAMI, M., HISAMATSU, S., KATO, Y. & TAKIZAWA, T. (2008) Comparative genome analysis of Lactobacillus reuteri and Lactobacillus fermentum reveal a genomic island for reuterin and cobalamin production. *DNA research,* 15**,** 151-161.

NATARAJAN, P., LANDER, G. C., SHEPHERD, C. M., REDDY, V. S., BROOKS, C. L. & JOHNSON, J. E. (2005) Exploring icosahedral virus structures with VIPER. *Nature Reviews Microbiology,* 3**,** 809-817.

NEWBY, Z. E. R., O'CONNELL III, J., ROBLES-COLMENARES, Y., KHADEMI, S., MIERCKE, L. J. & STROUD, R. M. (2008) Crystal structure of the aquaglyceroporin PfAQP from the malarial parasite Plasmodium falciparum. *Nature structural & molecular biology,* 15**,** 619-625.

OBRADORS, N., BADIA, J., BALDOMA, L. & AGUILAR, J. (1988) Anaerobic metabolism of the L-rhamnose fermentation product 1, 2-propanediol in Salmonella typhimurium. *Journal of Bacteriology,* 170**,** 2159.

PACE, C. N., VAJDOS, F., FEE, L., GRIMSLEY, G. & GRAY, T. (1995) How to measure and predict the molar absorption coefficient of a protein. *Protein Science: A Publication of the Protein Society,* 4**,** 2411.

PADILLA, J. E. & YEATES, T. O. (2003) A statistic for local intensity differences: robustness to anisotropy and pseudo-centering and utility for detecting twinning. *Acta Crystallographica Section D: Biological Crystallography,* 59**,** 1124-1130.

PALACIOS, S., STARAI, V. J. & ESCALANTE-SEMERENA, J. C. (2003) Propionyl coenzyme A is a common intermediate in the 1, 2-propanediol and propionate catabolic pathways needed for expression of the prpBCDE operon during growth of Salmonella enterica on 1, 2-propanediol. *Journal of Bacteriology,* 185**,** 2802.

PANG, A., WARREN, M. J. & 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**,** 91-96.

PARSONS, J. B. (2008) Characterisation of the Citrobacter freundii metabolosome. Canterbury, University of Kent.

PARSONS, J. B., DINESH, S. D., DEERY, E., LEECH, H. K., BRINDLEY, A. A., HELDT, D., FRANK, S., SMALES, C. M., LUNSDORF, H., RAMBACH, A., GASS, M. H., BLELOCH, A., MCCLEAN, K. J., MUNRO, A. W., RIGBY, S. E. J., WARREN, M. J. & PRENTICE, M. B. (2008) Biochemical and structural insights into bacterial organelle form and biogenesis. *Journal of Biological Chemistry,* 283**,** 14366-14375.

PARSONS, J. B., FRANK, S., BHELLA, D., LIANG, M. Z., PRENTICE, M. B., MULVIHILL, D. P. & WARREN, M. J. (2010a) Synthesis of Empty Bacterial Microcompartments, Directed Organelle Protein Incorporation, and Evidence of Filament-Associated Organelle Movement. *Molecular Cell,* 38**,** 305-315.

PARSONS, J. B., LAWRENCE, A. D., MCLEAN, K. J., MUNRO, A. W., RIGBY, S. E. J. & WARREN, M. J. (2010b) Characterisation of PduS, the pdu metabolosome corrin

reductase, and evidence of substructural organisation within the bacterial microcompartment. *PLoS One,* 5**,** e14009.

PRICE-CARTER, M., TINGEY, J., BOBIK, T. A. & ROTH, J. R. (2001) The Alternative Electron Acceptor Tetrathionate Supports B12-Dependent Anaerobic Growth ofSalmonella enterica Serovar Typhimurium on Ethanolamine or 1, 2-Propanediol. *Journal of Bacteriology,* 183**,** 2463-2475.

PRICE, G. D. & BADGER, M. R. (1989) Isolation and characterization of high CO2-requiring-mutants of the cyanobacterium *Synechococcus* PCC7942: two phenotypes that accumulate inorganic carbon but are apparently unable to generate CO2 within the carboxysome. *Plant Physiology,* 91**,** 514-525.

PRICE, G. D. & BADGER, M. R. (1991) Evidence for the role of carboxysomes in the cyanobacterial CO2-concentrating mechanism. . *Canadian Journal of Botany-Revue Canadienne De Botanique,* 69**,** 963-973.

PRICE, G. D., HOWITT, S. M., HARRISON, K. & BADGER, M. R. (1993) Analysis of a genomic DNA region from the cyanobacterium *Synechococcus sp* strain PCC7942 involved in carboxysome assembly and function. . *Journal of Bacteriology,* 175**,** 2871-2879.

REINHOLD, L., KOSLOFF, R. & KAPLAN, A. (1991) A model for inorganic carbon fluxes and photosynthesis in cyanobacterial carboxysomes. *Canadian journal of botany,* 69**,** 984-988.

REINHOLD, L., ZVIMAN, M. & KAPLAN, A. (1987) Inorganic carbon fluxes and photosynthesis in cyanobacteria-a quantitative model. *Progress in photosynthesis,* 4**,** 6.289-6.296.

REYNOLDS, C., DAMERELL, D. & JONES, S. (2009) ProtorP: a protein-protein interaction analysis server. *Bioinformatics,* 25**,** 413.

RHODES, G. (2006) *Crystallography made crystal clear: a guide for users of macromolecular models*, Academic press.

ROSTOVTSEVA, T. K. & BEZRUKOV, S. M. (1998) ATP transport through a single mitochondrial channel, VDAC, studied by current fluctuation analysis. *Biophysical journal,* 74**,** 2365-2373.

SAGERMANN, M., OHTAKI, A. & NIKOLAKAKIS, K. (2009) Crystal structure of the EutL shell protein of the ethanolamine ammonia lyase microcompartment. *Proceedings of the National Academy of Sciences,* 106**,** 8883.

SAMBROOK, J., FRITSCH, E. F. & MANIATIS, T. (1989) *Molecular cloning: a laboratory manual,* Cold Spring Harbor, N.Y., Cold Spring Harbor Press.

SAMPSON, E. M. & 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**,** 2966-2971.

SAMPSON, E. M., HAVEMANN, G. D., ALDRICH, H. C. & BOBIK, T. A. (2002) Genetic analysis of the 1,2-propanediol degradation by Salmonella enterica serovar typhimurium LT2. *Abstracts of the General Meeting of the American Society for Microbiology,* 102**,** 282.

SCHIRMER, A., RUDE, M. A., LI, X., POPOVA, E. & DEL CARDAYRE, S. B. (2010) Microbial biosynthesis of alkanes. *Science,* 329**,** 559-562.

SEEDORF, H., FRICKE, W. F., VEITH, B., BRUGGEMANN, H., LIESEGANG, H., STRITTIMATTER, A., MIETHKE, M., BUCKEL, W., HINDERBERGER, J., LI, F. L., HAGEMEIER, C., THAUER, R. K. & GOTTSCHALK, G. (2008) The genome of Clostridium kluyveri, a strict anaerobe with unique metabolic features. *Proceedings of the National Academy of Sciences of the United States of America,* 105**,** 2128-2133.

SEGA, E. I. & LOW, P. S. (2008) Tumor detection using folate receptor-targeted imaging agents. *Cancer and Metastasis Reviews,* 27**,** 655-664.

SHIVELY, J. M., BALL, F., BROWN, D. H. & SAUNDERS, R. E. (1973) Functional Organelles in Prokaryotes: Polyhedral Inclusions (Carboxysomes) of Thiobacllus neapolitanus. *Science,* 182**,** 584-586.

SHIVELY, J. M., BRADBURNE, C. E., ALDRICH, H. C., BOBIK, T. A., MEHLMAN, J. L., JIN, S. & 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-Revue Canadienne De Botanique,* 76**,** 906-916.

SMIALOWSKI, P., MARTIN-GALIANO, A. J., MIKOLAJKA, A., GIRSCHICK, T., HOLAK, T. A. & FRISHMAN, D. (2007) Protein solubility: sequence based prediction and experimental verification. *Bioinformatics,* 23**,** 2536-2542.

SRIRAMULU, D. D., LIANG, M., HERNANDEZ-ROMERO, D., RAUX-DEERY, E., LUNSDORF, H., PARSONS, J. B., WARREN, M. J. & PRENTICE, M. B. (2008) Lactobacillus reuteri DSM 20016 produces cobalamin-dependent diol dehydratase in metabolosomes and metabolizes 1, 2-propanediol by disproportionation. *Journal of Bacteriology,* 190**,** 4559-4567.

STOJILJKOVIC, I., BAUMLER, A. J. & HEFFRON, F. (1995) Ethanolamine utilization in Salmonella typhimurium: nucleotide sequence, protein expression, and mutational analysis of the cchA cchB eutE eutJ eutG eutH gene cluster. *Journal of Bacteriology,* 177**,** 1357-1366.

SUN, J., DUFORT, C., DANIEL, M. C., MURALI, A., CHEN, C., GOPINATH, K., STEIN, B., DE, M., ROTELLO, V. M. & HOLZENBURG, A. (2007) Core-controlled polymorphism in virus-like particles. *Proceedings of the National Academy of Sciences,* 104**,** 1354.

TAKENOYA, M., NIKOLAKAKIS, K. & 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**,** 6056.

TALARICO, T. L. & DOBROGOSZ, W. J. (1989) Chemical characterization of an antimicrobial substance produced by Lactobacillus reuteri. *Antimicrobial agents and chemotherapy,* 33**,** 674-679.

TANAKA, S., KERFELD, C. A., SAWAYA, M. R., CAI, F., HEINHORST, S., CANNON, G. C. & YEATES, T. O. (2008) Atomic-level models of the bacterial carboxysome shell. *Science,* 319**,** 1083-1086.

TANAKA, S., SAWAYA, M. R., PHILLIPS, M. & YEATES, T. O. (2009) Insights from multiple structures of the shell proteins from the beta-carboxysome. *Protein Science,* 18**,** 108-120.

TANAKA, S., SAWAYA, M. R. & YEATES, T. O. (2010) Structure and mechanisms of a protein-based organelle in *Escherichia coli*. *Science,* 327**,** 81-84.

TSAI, S. J. & YEATES, T. O. (2011) Bacterial microcompartments insights into the structure, mechanism, and engineering applications. *Progress in molecular biology and translational science,* 103**,** 1.

TSAI, Y., SAWAYA, M. R., CANNON, G. C., CAI, F., WILLIAMS, E. B., HEINHORST, S., KERFELD, C. A. & YEATES, T. O. (2007) Structural analysis of CsoS1A and the protein shell of the Halothiobacillus neapolitanus carboxysome. *Plos Biology,* 5**,** 1345-1354.

TSAI, Y. S., SAWAYA, M. R. & YEATES, T. O. (2009) Analysis of lattice-translocation disorder in the layered hexagonal structure of carboxysome shell protein CsoS1C. *Acta Crystallographica Section D-Biological Crystallography,* 65**,** 980-988.

VAGIN, A. & TEPLYAKOV, A. (1997) MOLREP: an automated program for molecular replacement. *Journal of applied crystallography,* 30**,** 1022-1025.

VAGUINE, A. A., RICHELLE, J. & WODAK, S. J. (1999) SFCHECK: a unified set of procedures for evaluating the quality of macromolecular structure-factor data and their agreement with the atomic model. *Acta Crystallographica Section D: Biological Crystallography,* 55**,** 191-205.

VERNIZZI, G. & OLVERA DE LA CRUZ, M. (2007) Faceting ionic shells into icosahedra via electrostatics. *Proceedings of the National Academy of Sciences,* 104**,** 18382.

WANG, Y. F., DUTZLER, R., RIZKALLAH, P. J., ROSENBUSCH, J. P. & SCHIRMER, T. (1997) Channel specificity: structural basis for sugar discrimination and differential flux rates in maltoporin1. *Journal of Molecular Biology,* 272**,** 56-63.

WHITE, D. A. (1973) Phospholipid composition of mammalian tissues. IN B., A. G., N., H. J. & C., D. R. M. (Eds.) *Form and function of phospholipids.* New York, Elsevier Publishing Company.

WILKINSON, D. L. & HARRISON, R. G. (1991) Predicting the solubility of recombinant proteins in Escherichia coli. *Nature Biotechnology,* 9**,** 443-448.

WINTER, G. (2009) xia2: an expert system for macromolecular crystallography data reduction. *Journal of applied crystallography,* 43**,** 186-190.

YEATES, T. O. & FAM, B. C. (1999) Protein crystals and their evil twins. *Structure,* 7**,** 25-29.

YEATES, T. O., THOMPSON, M. C. & BOBIK, T. A. (2011) The protein shells of bacterial microcompartment organelles. *Current Opinion in Structural Biology,* 21**,** 223-231.

ZHANG, Y. H. P., EVANS, B. R., MIELENZ, J. R., HOPKINS, R. C. & ADAMS, M. W. W. (2007) High-yield hydrogen production from starch and water by a synthetic enzymatic pathway. *PLoS One,* 2**,** e456.

ZHU, X. & MITCHELL, J. C. (2011) KFC2: A knowledge based hot spot prediction method based on interface solvation, atomic density, and plasticity features. *Proteins: Structure, Function, and Bioinformatics*.