Omp85 family proteins and the BAM complex in *Caulobacter crescentus*

Khatira Anwari

A thesis submitted to Monash University in total fulfillment of the requirements for the degree of Doctor of Philosophy

June 2011

Department of Biochemistry and Molecular Biology
Science Technology Research and Innovation Precinct
Monash University

Table of Contents

APPENDICES

List of Figures

Chapter 1:

Chapter 2:

Chapter 3:

Chapter 4:

List of Tables

Declaration

This is to certify that:

i. the thesis comprises only my original work towards the PhD except where indicated in the Preface,

ii. due acknowledgements have been made in the text to all other material used and

iii. the thesis is less than 100,000 words in length, exclusive of tables, maps, bibliographies and appendices.

Abstract

The outer membrane of Gram-negative bacteria is a compartment that houses many proteins involved in basic physiological functions, virulence and multi-drug resistance, and is therefore important for cell survival. Most proteins in the outer membrane adopt a β-barrel conformation, and require the β-barrel assembly machinery (BAM) for integration into the outer membrane. The BAM complex is a multi-subunit protein complex present in the outer membrane of all Gram-negative bacteria. It comprises of a core β-barrel protein, BamA and associated lipoproteins that collectively participate in the folding and insertion of β-barrel proteins. BamA has functional homologues in eukaryotes that form the SAM complex in mitochondria for insertion and assembly of β-barrel proteins into the mitochondrial outer membranes. Work by other groups in the field has used *N. meningitidis* (β-proteobacteria) and *E. coli* (γ-proteobacteria) to study the function of BamA and the BAM complex. In this comparative study, we will use *Caulobacter crescentus* as a model α-proteobacterium to learn about differences and similarities in the BAM complexes across the three different classes of proteobacteria. The combined use of bioinformatics and biochemical experiments in *C. crescentus* have shown some key differences with the absence of the BamC partner lipoprotein and the presence of a novel OmpA family protein as well as two other uncharacterized lipoprotein partners for BamA.

In Chapter 2, we perform a comparative analysis of BamA and the BamA-like protein (Omp68) in *C. crescentus*. The Omp85 superfamily contains BamA-, Omp68- and TpsB-family proteins and using bioinformatics analysis, we proposed that Omp68 is an evolutionary intermediate of BamA and TpsB proteins. We first established a method for preparing outer membranes from *C. crescentus* and demonstrated that like BamA, Omp68 forms an oligomeric complex in the outer membrane although it has characteristics that could classify it as a TpsB protein. The unusual nature of Omp68

is discussed and its possible functions. A method for effective purification of the BAM complex from outer membranes was also established for identification and analysis of the different components of the complex.

In Chapter 3, we demonstrate that the BAM complex possesses modular characteristics and contains BamA and six outer membrane lipoproteins in the α-proteobacterium, *C. crescentus*. In addition to the three known lipoproteins (BamB, BamD and BamE), we identify three other subunits (Pal, BamF and BamG) of which only Pal is essential. We propose that Pal is a protein that anchors the BAM complex to the peptidoglycan layer and promotes proximity to the inner membrane Sec machinery for efficient outer membrane protein assembly. We also show BamF is a genuine component of the BAM complex and a potential homologue of the BamC protein found in all other proteobacteria. Both BamF and BamC contain a conserved motif that is possibly important for docking onto the BAM complex.

In Chapter 4, we focus on biochemical and structural characterisations of BamD from *C. crescentus*. We show BamD is an outer membrane lipoprotein that forms the halo module of the BAM complex. We also demonstrate BamD contains TPR motifs that are essential for functioning of the BAM complex and that protrude into the periplasm for protein-protein interactions. Further experiments involved extensive optimisations of heterologous expression of BamD and purification with the aim of progressing to crystallisation trials. We were able to generate purified folded BamD that will be used for future experiments.

Acknowledgements

Approaching the completion of this thesis, I am indebted to all the people who have supported and helped me, more or less, to get to this point. Although it is impossible to name all, I would like to acknowledge some of them.

First of all, special thanks go to my supervisor, Professor Trevor Lithgow, for his knowledge, guidance and encouragement throughout my candidature. In addition, his devotion and enthusiasm for science that led to his great achievements are a great inspiration for me.

I would like to thank our great collaborators Christine Jacobs-Wagner and Sebastian Poggio for providing us with strains and genetic tools to help us work with *Caulobacter*.

I would also like to thank past and current members of the Lithgow lab for their support: Dejan, Nickie, Joanne, Andrew, Franziska, Michael, Xenia, Abi, Srgjan, Felicity, Marija, Hanim, Miguel, Chaille, Nermin, Rhys, Joel and Victoria. I would also like to give a special thanks to those that made my PhD experience full of laughs: Allen, Dejan, Nickie, Pavel, Nathalie, Jhiang, Janette, Miguel, Tara and Yue. Thank you for supporting me, cheering me up when I was depressed, encouraging me when I was frustrated and advising me when I was confused.

I would like to thank God for strengthening my will to progress to different and difficult phases of life and I am very grateful to have such a wonderful mum for nourishing me with her precious love. I love you with all my heart.

Preface

The contributions of collaborators are outlined below:

i. A/Prof. Christine Jacobs-Wagner (Yale University) for cloning reagents and *Caulobacter* strains. Dr. Sebastian Poggio generated the Pal-mCherry fusion strain and performed microscopy and the FRAP experiment (Chapter 3).

ii. Dr. Susan Buchanan (NIH) and her post-doctoral researcher Dr. Nicholas Noinaj created the structural alignment of Pal and its homologues (Chapter 3).

iii. A/Prof. Tony Purcell and his research assistant Sri H Ramarathinam performed mass spectrometry analysis of samples.

iv. Dr. Andrew Perry performed the MEME analysis on BamF and BamC homologues (Chapter 3).

All other experimentation comprises my original work.

Some of the work undertaken towards this degree has been published in the papers listed below.

Gatsos X, Perry AJ, Anwari K, Dolezal P, Wolynec PP, Likić VA, Purcell AW, Buchanan SK, Lithgow T (2008) Protein secretion and outer membrane assembly in Alphaproteobacteria. *FEMS Microbiol Rev.* 32(6):995-1009

Anwari K, Poggio S, Perry A, Gatsos X, Ramarathinam SH, Williamson NA, Noinaj N, Buchanan S, Gabriel K, Purcell AW, Jacobs-Wagner C, Lithgow T (2010) A modular BAM complex in the outer membrane of the alpha-proteobacterium *Caulobacter crescentus. PLoS One* 85(1):e8619.

Abbreviations

BN	blue native
BSA	bovine serum albumin
CD	circular dichroism
CV	column volume
DDM	*n*-dodecyl-β-maltoside, dodecyl-maltoside
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
DTT	dithiothreitol
ECL	enhanced Chemiluminescence
EDTA	ethylenediaminetetraacetic acid
GST	glutathione S-transferase
HCl	hydrochloride
HEPES	4-(2-Hydroxyethl) piperazine-1-ethanesulfonic acid
HMM	hidden Markov model
IB	inclusion bodies
In	induced
IPTG	isopropyl-β-D-thiogalactoside
LB	Luria broth
MWCO	molecular weight cut-off
NaCl	sodium chloride
NMR	nuclear magnetic resonance
OD	optical density
P	pellet
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction

PDB	room temperature
POTRA	polypeptide transport associated
PMSF	phenylmethanesulfonylfluoride
RmpM	reduction modifiable protein M
S	supernatant
SDS	sodium dodecyl phosphate
TBS	tris buffered saline
Tris	tris (hydroxy methyl) aminomethane
TPR	tetratricopeptide repeat
TpsB	two-partner secretion protein B
Un	uninduced

Chapter 1

Introduction

1.1 Composition of Gram-negative bacterial membranes

Cell membranes are important structures that act as permeability barriers by housing specific proteins to transport metabolites. Gram-negative bacteria have two membrane layers that cordon two major compartments within the cell: the cytoplasm and the periplasm. The inner membrane encapsulates the cytoplasm, whereas the inner and outer membrane are separated by the periplasm (Figure 1.1), which is a viscous compartment with soluble proteins and a thin peptidoglycan layer (1).

The inner and outer membranes differ in both their lipid and protein composition (Figure 1.1). The inner membrane acts as a more complete permeability barrier and contains a large number of transport systems to move molecules into and out of the cell. It is formed from a phospholipid bilayer that contains integral proteins with α-helical transmembrane domains (2). Lipoproteins are also associated with the inner membrane and are bound to the outer leaflet (periplasmic side) of the inner membrane via an N-terminal lipid anchor. Several mechanisms exist to transport proteins across the inner membrane, including the Sec pathway (3, 4), the twin-arginine transport system (5), type I-IV secretion pathways (6), the capsule transport system (7) as well the transport of LPS and lipoproteins via the Lpt (8) and Lol (9) machinery, respectively. Some of these transport systems span the inner and outer membrane, forming proteinaceous contact points between the membranes.

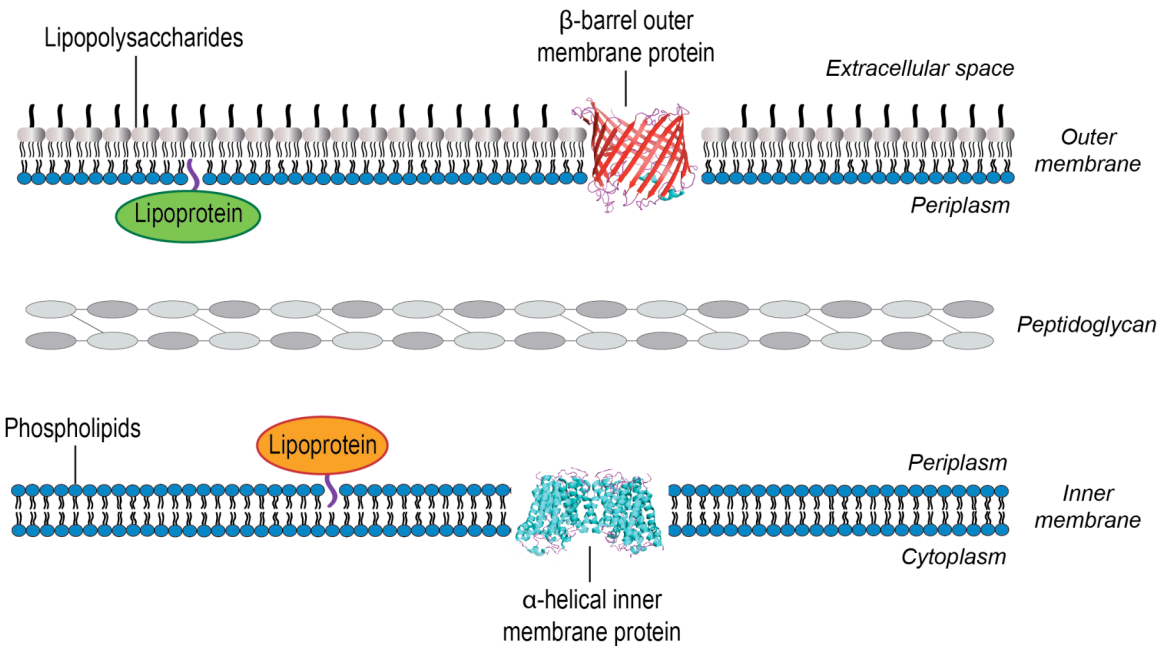


Figure 1.1: Cell envelope of *E. coli.* The inner and outer membrane layers have different properties due to differences relating to lipid and protein composition. Figure is adapted from (2).

The outer membrane acts as a molecular sieve for the passage of hydrophilic solutes. Unlike the inner membrane, it is asymmetric regarding its lipid composition: the surface exposed monolayer consists of glycolipids (lipopolysaccharides, LPS) while the inner leaflet contains phospholipids, predominantly phosphatidylethanolamine for *Escherichia coli* (2) and phosphatidylglycerol for *Caulobacter crescentus* (10). The transport of phospholipids (11) to the membrane remains poorly understood whereas LPS transport has been well characterised (Figure 1.2) (12). The LPS layer assist the barrier function of the outer membrane by greatly reducing the diffusion of hydrophobic compounds that could be noxious to the cell (13). LPS consists of a lipid A moiety that is embedded in the membrane, a short core oligosaccharide, and an O antigen that may be a long polysaccharide (8, 13). The LPS layer is tightly packed and highly ordered to prevent the migration of any phospholipid molecules from the inner leaflet. There are strong lateral

interactions between LPS molecules, mediated by divalent cations that bridge the anionic LPS molecules (13). The significance of the LPS layer is demonstrated when LPS bacterial mutants are studied. For example, the *E. coli lpxA, lpxC and lpxD* mutants with obvious defects in the biosynthesis of the lipid A component are highly susceptible to hydrophobic antibiotics (8, 14, 15). The levels of LPS in these mutants are drastically low and the likely explanation for such sensitivity is lack of a continuous LPS layer in the outer leaflet and the consequential compensatory migration of phospholipids from inner to the outer leaflet of outer membrane (14). This may create regions in the outer leaflet that can allow the diffusion of hydrophobic solutes. Although most Gram-negative bacteria synthesise LPS, there are some bacterial species (e.g. *Deinococcus radiodurans*, the spirochaetes *Treponema* and *Borrelia* and α-proteobacteria such as *Sphingomonas*) that lack the lipid-A system, as defined by the absence of *lpx* genes (8, 13, 16). *Sphingomonas* uses glycosphingolipids instead of lipid A, but which lipids the other species use instead of lipid A has not been determined (8).

Proteins found in the outer membrane are of three major classes: surface organelles such as pili (17), peripherally associated lipoproteins (2, 9) or integral proteins with a β-barrel structure (2, 9, 13, 18). Components of pili are targeted to the outer membrane via the usher pathway. Lipoproteins are proteins with a lipid modified N-terminal cysteine that are generally targeted and anchored to the inner leaflet of the outer membrane via the Lol pathway (9). Some bacterial species also have lipoproteins on the outer leaflet, including *Neisseria* (19-22). Bacterial lipoproteins have diverse functions including maintaining the structural integrity of the bacterial cell envelope (23, 24), cell adhesion, transport, nutrient acquisition, mating, as well as stimulation of inflammatory or immune responses in host cells (25). Deletions or mutations in some genes that encode lipoproteins cause several defects such as periplasmic leakage and hypersusceptibility to antibiotics (26, 27). Furthermore, recent studies have shown that lipoproteins such as BamB, BamC, BamD and BamE play a role in bacterial outer membrane biogenesis (28, 29).

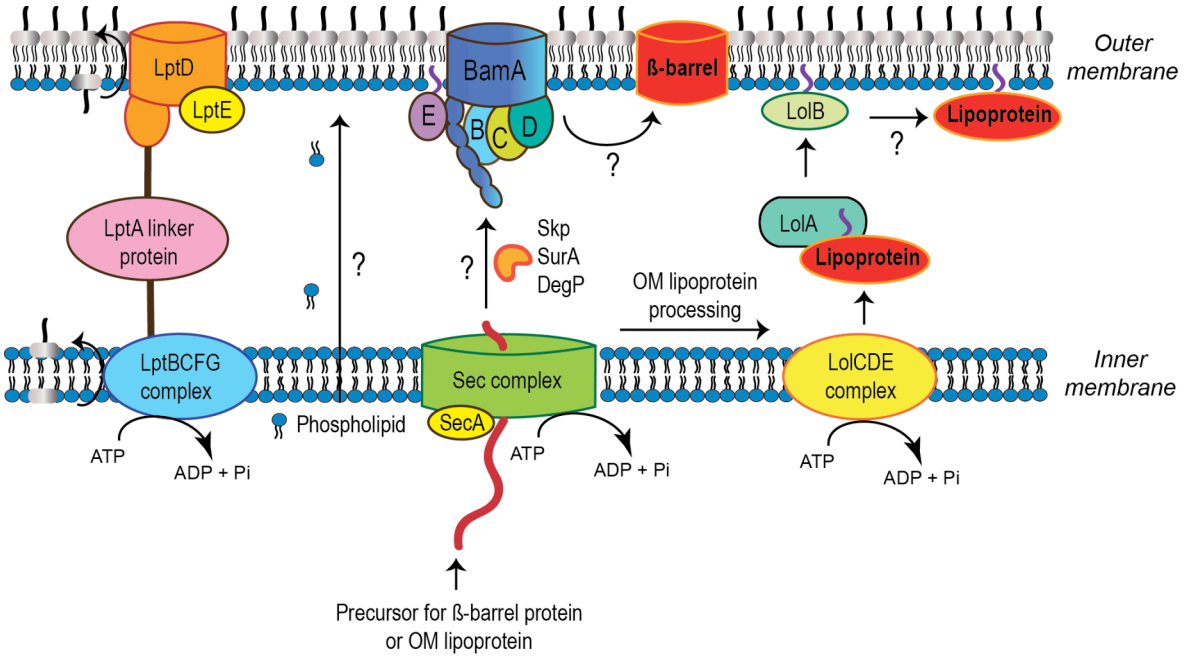


Figure 1.2: Cellular pathways for outer membrane biogenesis in Gram-negative bacteria. There are multiple pathways present in bacteria that maintain the integrity of the outer membrane. All components of the outer membrane are synthesised in the cytoplasm. After synthesis, phospholids and LPS are flipped across the inner membrane. Phospholipid transport is not well characterised but LPS transport occurs via the LptA linker protein in the periplasm, which links the outer membrane LptDE and the inner membrane LptBCFG complexes for transfer of LPS across the periplasm. All ß-barrel proteins and outer membrane (OM) lipoproteins are translocated from the cytoplasm to the periplasm by the Sec complex. Outer membrane proteins that do not have signals for lipidation have their signal sequence cleaved (not shown) and interact with periplasmic chaperones (Skp, SurA and DegP). It is considered these chaperones assist in the folding and transfer of these proteins to the BamABCDE (BAM) complex in the outer membrane. The mechanism of how these proteins interact with the BAM complex and how they adopt a β-barrel conformation is unclear. Outer membrane proteins destined to become lipoproteins are lipidated while their signal sequence is removed in the periplasm. LolA transfers the outer membran lipoprotein across the periplasm for transfer to LolB at the outer membrane for anchorage. Figure is adapted from (2).

Unlike lipoproteins, the β-barrel class of proteins are embedded within the outer membranes of bacteria, mitochondria and chloroplasts. β-barrel proteins consist of an even number of amphipathic anti-parallel β-strands that twist and coil to form β-barrel

cylinders or channels in the outer membrane (18, 30) with long loops on the outside of the outer membrane and short turns of β-hairpin on the periplasmic side (18). Hydrogen bonds between neighbouring β-strands maintain and stabilise the barrel conformation in the non-polar bilayer (31). The segregation of α-helices in the inner membrane and β-barrel proteins in outer membrane occurs at the inner membrane. It is related to the fact that α-helical membrane proteins have long adjacent hydrophobic helices (32) that can move laterally from the Sec machinery whereas β-barrel outer membrane proteins are translocated because they have unpredictable hydrophobic surfaces (30).

Outer membrane proteins perform different functions in the cell. For example, porins act as diffusion channels for the transport of nutrients and waste products across the membrane (13). The major outer membrane proteins in *E. coli* are the trimeric porins (OmpA, OmpC, OmpF, LamB and PhoE) (9), but many additional non-abundant proteins, including some enzymes and adhesins, are also present (13). OmpA and other outer membrane proteins can also provide physical links between the outer and inner membrane (33) whereas others can act as receptors for colicins and bacteriophages (34). Outer membrane proteins are also related to virulence and can help bacteria adapt to hostile conditions in the host during infection or elicit a pro-inflammatory host response (35). Considering many of these proteins form pores in the membrane, the OM integrity is potentially compromised. However, the pore can repel hydrophobic solutes by inward-facing charged residues or be physically blocked by infolding of long loops and/or an existing N-terminal plug found on the periplasmic side (16, 36).

Under certain environmental cues, a layer of polysaccharides can contribute to the protective function of the outer membrane (37). This capsule layer is firmly attached to the outer membrane and can protect the cell from desiccation, phagocytosis or complement-mediated killing (38). In most bacteria such as *C. crescentus*, a proteinaceous S-layer that is anchored to LPS may protect the bacterial cell from environmental stress, promote cell adhesion, and maintain cell shape and rigidity (39). Some bacteria have both the capsule and S-layer with the capsule exterior to and completely covering the S-layer proteins (40).

As with all biological membranes, the bacterial outer membrane is not created *de novo* but grows through a templated process of membrane biogenesis (2). The various components such as phospholipids, LPS, lipoproteins and integral outer membrane proteins are made in the cytoplasm and transported into the periplasm by translocation across the inner membrane (Figure 1.2). They are then sorted and incorporated in the outer membrane. Different protein machines and mechanisms of action take place to target these components to the outer membrane but in the context of this thesis, the transport and assembly of lipoproteins and integral outer membrane proteins will be discussed.

1.2 Lipoproteins synthesis, processing and targeting to the outer membrane

Most lipoproteins synthesized by Gram-negative bacteria are attached to the outer membrane (9). They are synthesized as precursor proteins in the cytoplasm, with a characteristic N-terminal signal sequence followed by a key cysteine residue. The precursor protein is translocated into the periplasm by the Sec protein machinery in the inner membrane (Figure 1.2). Pre-proteins that are destined to become lipidated contain a motif known as a lipobox that directs them to the lipoprotein biogenesis machinery (9, 34). This lipoprotein processing takes place on the periplasmic side of the inner membrane and involves two enzymes and a protease (9). The lipobox motif is typically L_{-3}-$[A/S/T]_{-2}$-$[G/A]_{-1}$-C_{+1} with the +1 cysteine absolutely conserved in all bacterial lipoproteins (34). The cysteine is the first residue of the mature protein while the other residues of the lipobox are part of the signal sequence. The first step of processing involves recognition of the lipobox by the lipoprotein diacylglyceryl transferase (Lgt) and the addition of a lipid group to the sulphydryl group of the key cysteine. Following this crucial step, the signal sequence is cleaved by signal peptidase II to generate a mature protein with a lipidated N-terminal cysteine. Lastly, N-acyl transferase (Lnt) adds a fatty acid chain to the amino group of the cysteine (9, 41).

Lipoprotein sorting within the periplasm and targeting to the outer membrane is handled by the Lol pathway and is based on the nature of the amino acid following the N-terminal cysteine residue (+1 position). The residue Asp (+2 position) immediately following the Cys serves as a typical inner membrane retention signal (42). Gly, Pro or aromatic amino acids (Phe, Trp, Tyr) can also localise the lipoprotein to the inner membrane (43). However, for outer membrane lipoproteins, the presence of Ser or other amino acids at the same position will act as outer membrane targeting signals. Also, amino acids at position +3 can either weaken (Lys) or strengthen (Asp, Glu, Gln, Asn, Arg) inner membrane retention (44, 45). For outer membrane lipoproteins, the lipidated lipoprotein interacts with the energy-dependent LolCDE complex in the inner membrane. This protein complex belongs to the ATP-binding cassette (ABC) transporter family and its disruption affects the release of outer membrane lipoproteins from the inner membrane but does not affect inner membrane lipoprotein processing or localisation (46, 47). In fact, inner membrane lipoproteins remain in the inner membrane because they possess the "Lol avoidance signal" reflected in their sorting signal that prevents their interaction with the LolCDE complex. The release of outer membrane-specific lipoproteins is also dependent on ATP hydrolysis and the periplasmic LolA lipoprotein, which transfers the lipoprotein to the outer membrane lipoprotein LolB for anchorage. The lipoprotein-LolA complex is water soluble, suggesting LolA shields the fatty acid chains of the lipoprotein substrate from the aqueous environment of the periplasm. Upon anchorage by some unknown mechanism involving LolB, lipoproteins then become available for interaction with other outer membrane proteins and can participate in oligomeric complexes. The transfer of an outer membrane lipoprotein between LolCDE-LolA-LolB is thought to be driven by differential affinities for the lipoprotein (9).

1.3 β-barrel protein targeting and assembly

As with lipoproteins, integral β-barrel proteins also need to pass the inner membrane and periplasm since they too are synthesised in the cytoplasm. The N-terminal signal sequence of the pre-protein targets it to the Sec machinery in the inner membrane, which

transports the protein into the periplasm in an electrochemical energy- and ATP-dependent manner (48, 49). The characteristic signal sequence of the pre-protein is removed at the outer face of the inner membrane by signal peptidase, which is associated with the inner membrane (50). Numerous genetic and biochemical experiments have revealed many molecular details of how proteins pass the inner membrane such that there is a sufficient understanding of these initial stages of outer membrane protein biogenesis (51). However, we are only starting to understand what happens to β-barrel proteins beyond their translocation across the inner membrane.

During translocation, or after the release of β-barrel proteins from the inner membrane, several periplasmic proteins are shown to associate with them (52). Of these chaperones, some have been thoroughly studied including Skp (Seventeen kilodalton protein) and SurA (53, 54). These chaperones interact with the nascent protein and apparently transfers it across the periplasmic space to the β-barrel assembly sites in the outer membrane (2). The trimeric Skp protein can physically associate with the plasma membrane (55, 56) and can be cross-linked to some atypical outer membrane proteins (OmpA, LamB and PhoE) as they emerge from the secretion machinery in spheroblasts, suggesting Skp is important in the initial stages of outer membrane protein biogenesis (56, 57). However, null Skp mutants are viable with a moderate reduction in properly folded outer membrane proteins, which implies that other pathways are also acting to facilitate outer membrane biogenesis. An investigation into the action of SurA demonstrates it is the major chaperone for β-barrel outer membrane proteins (53) that can interact with outer membrane pre-proteins at the outer leaflet of inner membrane (58) as well as the β-barrel assembly sites at the outer membrane (53). Like Skp, SurA null mutants are also viable but a combination of Skp and SurA mutations results in a synthetically lethal phenotype (59). Hence, it is proposed that SurA functions in an independent pathway to Skp (53, 59). The basis for a parallel pathway could be that some chaperones have preferred substrates that can be efficiently processed but in the event when this pathway is compromised, other chaperones are available that can perform a similar function but with less efficiency. However, when both pathways are dysfunctional or compromised, a synthetic lethal phenotype is observed.

Several *in vitro* studies have shown that model β-barrel proteins can spontaneously insert into synthetic membranes and adopt native conformations (60-62). However, this spontaneous process takes minutes to hours, which is not physiologically relevant since bacterial doubling times are generally in the order of 30 minutes (54). It is expected that chaperones like Skp and SurA may be involved in preventing aggregation and/or promoting protein folding to create an efficient process of outer membrane protein targeting and assembly. Indeed, there is significant evidence that Skp and SurA binds unstructured and unfolded conformation of outer membrane proteins (54). The core module of SurA has a tunnel-like structural feature that is predicted to be suitable for binding to an extended polypeptide chain (63) and is shown to bind peptide motifs based on integral outer membrane proteins (64). It has also been shown that outer membrane proteins can have some tertiary structure before they are inserted into the outer membrane. In one study, cysteine residues were introduced in PhoE at neighbouring positions of the native form that were expected to embed within the membrane. The periplasmic chaperone DsbA (disulfide bond isomerase) was able to create disulphide bonds within PhoE, creating an element of tertiary structure (65). Since β-barrel proteins rarely contain disulfide bonds in their transmembrane regions (54), it is unlikely disulphide isomerases play a significant role in folding proteins with this architecture.

1.4 The BAM complex: an evolutionarily conserved machine to assemble β-barrel proteins

The β-barrel assembly sites in the outer membrane that were previously mentioned also co-operate with chaperones to fold β-barrel proteins. These sites consist of the multi-subunit protein complex referred to as the β-barrel assembly machinery (BAM) complex, which, in *E. coli*, consists of four lipoproteins (BamB, BamC, BamD and BamE) and the conserved β-barrel protein, BamA (Figure 1.2) (28, 29). There are complications towards gaining a consistent understanding of what is required for an outer membrane protein to assemble since different outer membrane proteins have differential preference for LPS,

phospholipids, periplasmic chaperones or folding factors part of the BAM complex (66). However, research endeavours in this field have recently led to exciting discoveries with the ultimate aim of understanding the mechanism of the BAM complex. Before I discuss the targeting of this class of proteins to the BAM complex, I need to elaborate on the subunits that form this sophisticated piece of cellular machinery.

1.5 BamA: core subunit of the BAM complex

An important breakthrough for understanding outer membrane protein biogenesis was the identification of BamA, which is essential for the assembly of β-barrel outer membrane proteins and cell viability (67). A depletion strain of *N. meningitidis* with BamA expression under the control of an inducible promoter showed outer membrane proteins accumulate in the periplasm in non-native form, suggesting a role for BamA in folding and insertion of proteins in the outer membrane (67). The *bamA* locus in many Gram-negatives includes the gene *yaeL*, *skp* and several other genes related to LPS biosynthesis (68).

BamA is an integral β-barrel outer membrane protein that has a C-terminal transmembrane region with 12-16 predicted β-strands and an N-terminal periplasmic extension consisting of five polypeptide transport-associated (POTRA) domains (36, 67). Each POTRA segment consists of a three-stranded β-sheet overlapped by a pair of α-helices (69). This topology has been confirmed by treatment of *E. coli* cell envelopes with trypsin, resulting in an intact β-barrel domain and proteolysis of the POTRA region (52). Refolded BamA also behaves in a similar manner upon treatment with trypsin (52, 70).

Although the structure of BamA has not yet been solved, progress was made with the recently solved X-ray structures of FhaC (Figure 1.3A), another member of the Omp85 superfamily, and the four N-terminal POTRA domains of BamA (Figure 1.3B, 1.3C) (71-73). A hook-like conformation was adopted by the POTRA domains (72) but recent data

shows an extended conformation is also possible (71, 73). Perhaps both are physiologically relevant. Such an extended conformation could result in spanning of the POTRA domains across the periplasm (Figure 1.3D) (73). Using information from the crystal structure, five BamA mutants with single POTRA deletions were generated and expressed as histidine-tagged proteins in *E. coli* that expresses wild-type BamA. Pull-down studies demonstrated that the POTRA domains of BamA act as docking sites for the lipoproteins in the BAM complex. The fifth and most C-terminal POTRA domain is important for BamC, BamD and BamE interactions whereas the first four N-terminal POTRA domains are all crucial for BamB interaction (72). A similar approach used for BamA in *Neisseria* showed the first four POTRA domains resulted in slight defects in outer membrane protein assembly whereas the fifth POTRA was essential for BamA function (74). The essentiality could be related to disruption of BamA-BamD association within the protein complex. In addition to acting as a scaffold for lipoprotein subunits, the POTRA domains are also important for initial substrate recognition and chaperone-like activity (52, 69, 74).

Some members of the Omp85 superfamily are able to self-associate and form oligomers (75, 76). BamA was identified as part of a high molecular weight oligomeric complex of unknown composition in *N. meningitidis* as demonstrated by semi-native gel electrophoresis (67). Moreover, purified BamA from *E. coli* (52) and HMW1B, a member of the Two-partner secretion (TPS) family in *H. influenzae*, (76), were able to form tetramers using blue native-gel electrophoresis (BN-PAGE) and gel filtration. However, the BamA or HMW1B oligomer is an unstable species that can dissociate to monomeric species. Contradictory to these observations is that the purified BAM complex from *E. coli* is separated by BN-PAGE as a complex with one copy of BamA (77). Since BamA interactions with other BamA proteins within the oligomer were shown to be weak, these may have been disrupted during the purification.

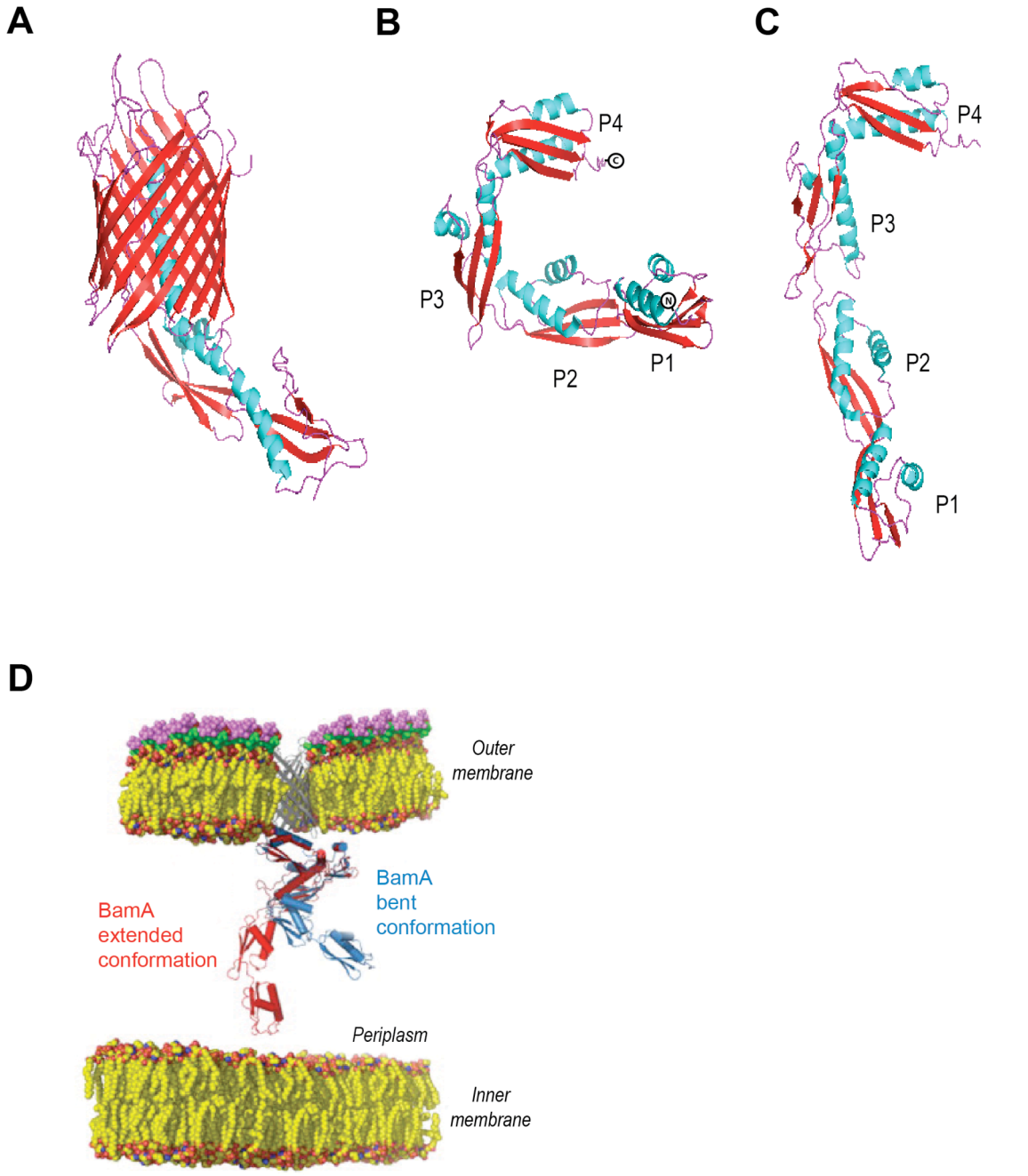


Figure 1.3: Delineating the structure of BamA. (A) Crystal structure of FhaC from *B. pertussis* (PDB code 2QDZ), (B) Crystal structure of the *E. coli* BamA POTRA 1-4 (PDB code 2QDF), (C) Similar to B) but with a different crystal-packing lattice showing a stretched conformation (PDB code 3EFC), (D) Model of BamA with two different conformations of its soluble POTRA 1-4 domain spanning the periplasm. Figure is adapted and taken from (73).

1.6 The complex workings of a protein complex

One of the most interesting questions related to the mechanism of outer membrane assembly is how the BAM complex, in the vast periplasmic pool of proteins, recognises its substrates. It has been suggested that processed outer membrane proteins have structural features that allow them to be recognised by the BAM complex. Incorporation of PhoE into the outer membrane is abolished when the last ten C-terminal amino acids that corresponds to the membrane-spanning segment are deleted (78). Comparative analysis of this segment shows Phe and further hydrophobic residues close to the C-terminus are preferentially found in many outer membrane proteins (52). Through extensive searches, mitochondrial outer membrane proteins also have a targeting sequence in the C-terminal β-strand but with a different signature motif (79). Deletion or substitution of the prominent Phe residue severely affected the assembly of PhoE in *E. coli*. The importance of this residue was further verified by measuring the channel activity or conductivity of BamA reconstituted in planar lipid bilayers after adding denatured wild-type or mutant PhoE. In contrast to native PhoE, the Phe deletion mutant of PhoE and two periplasmic proteins (LamB and Skp) did not affect the conductivity of the BamA channel (52). Furthermore, peptide-binding assays were developed by Knowles *et al.* (71) using peptides based on different β-strands of PhoE and purified POTRA 1-5 domains. Peptides were able to weakly bind to the POTRA domains, suggesting other internal targeting information is present in β-barrel proteins.

β-pairing or β-augmentation is another mechanism that has been proposed to explain how BamA interacts with its substrate (72, 80). This idea arose from the crystal structure of the POTRA 1-4 domains, which showed an artificial dimer with the β-sheet in POTRA 3 paired with the β-strand of the other subunit (72). It is an observation that may provide clues about how β-barrel substrates associate with POTRA domains via its C-terminal β-strand containing the motif that affects assembly of PhoE (52). Such a mechanism would explain the ability of the BAM complex to interact with so many types of outer membrane proteins with different protein sequences but similar hydrophobic periodicity (72).

The mechanism of how the BAM complex assembles proteins is unclear but two models have been proposed (Figure 1.4). One model relates to the oligomerisation of BamA to create a pore within the outer membrane. Proteins can enter the pore, fold and be released into the outer membrane by lateral opening of the oligomer (81). Alternatively, β-strand formation may be provided by the pore formed from the BamA monomer although lateral opening of the pore is thought to be thermo-dynamically unfavourable (82). The second possibility is that the complex formed by BamA provides a scaffold that facilitates the formation of β-strands, possibly through β-augmentation and allows insertion of β-barrel proteins at the protein-lipid interface (83).

A recent breakthrough was achieved by reconstituting the BAM complex into synthetic lipid vesicles and using an enzyme assay to monitor outer membrane protein assembly (77). No source of energy was required for assembly although incubation of the outer membrane protein substrate with SurA was required prior to the assembly assay. Further work needs to be done using this approach to dissect the process into distinct stages and develop a thorough understanding of how the BAM complex carries out its role.

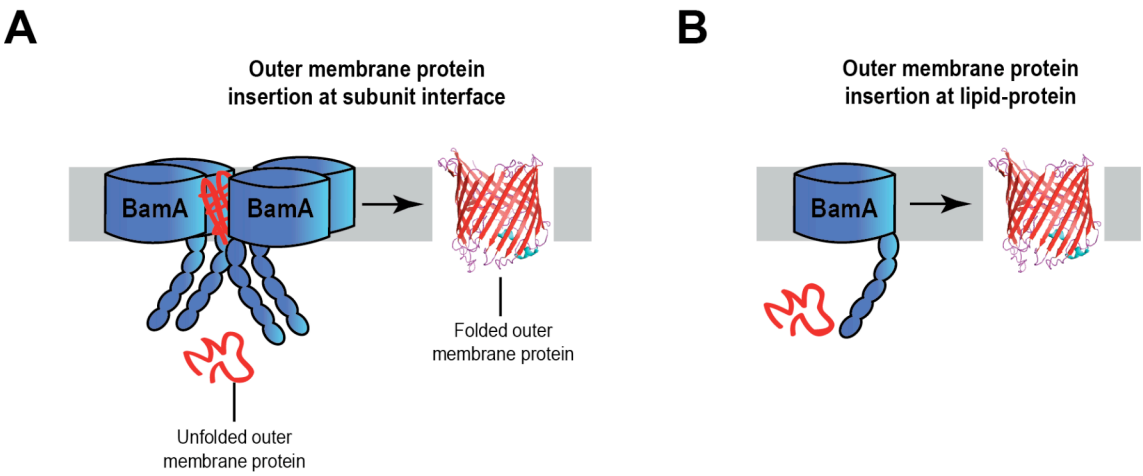


Figure 1.4: Two models outlining how outer membrane proteins are inserted into the outer membrane by BamA. (A) Protein insertion at the subunit interface can occur when BamA oligomerises to form a pore. Release of the protein from the pore is due to lateral opening of the oligomer. (B) Alternatively, BamA acts as a scaffold for protein insertion at the lipid-protein interface. The lipoprotein components of the BAM complex are not shown for simplicity. Figure of models is adapted from (81).

1.7 The Omp85 protein superfamily

A BamA homologue exists in mitochondria (Sam50), which is also important for the assembly of β-barrel proteins (84, 85). Hence, this process is conserved between prokaryotes (at least in Gram-negative bacteria) and eukaryotes. However, the similarity between bacterial and mitochondrial outer membrane protein biogenesis seems to be restricted to the homology of BamA and Sam50 while the sorting signal and partner proteins differ. The lipoprotein subunits of the BAM complex do not have homologues in the SAM complex (16). The SAM complex has metaxin subunits that have no relationship to any bacterial outer membrane proteins. One of the aims of this thesis was to determine whether the bacterial BAM complex has a modular composition, by analogy to the mitochondrial SAM complex.

A common feature of BamA homologues is the soluble extension of POTRA domains attached to the β-barrel domain. These have been identified in the outer membrane proteins BamA and Sam50, in the plastid protein translocase Toc75 and a class of proteins referred to as TpsB proteins (69). TpsB proteins are distantly related to BamA homologues but are mechanistically different from BamA as they are involved in transporting proteins across the outer membrane in the two-partner secretion pathway (Type V) in bacteria. TpsB proteins such as HMW1B in *Haemophilus influenzae*, FhaC in *Bordetella pertussis* and ShlB in *Serratia marcescens* transport specific substrates such as adhesins and hemolysins in an unfolded state across the outer membrane (86). In this sense, TpsB proteins are more related to plastid Toc75 proteins, while BamA proteins are more related to mitochondrial Sam50 proteins. In addition to these proteins, POTRA domains have been predicted in a class of conserved bacterial outer membrane proteins that are distantly related to BamA (69). The *ytfM* gene in *E. coli* encodes the protein YtfM that belongs to this class and its function or the function of its homologues is not clear (70).

1.8 Accessories to assembly: lipoprotein subunits of the BAM complex

BamA and its eukaryotic homologues are highly conserved proteins; however the composition of the complexes formed by this essential protein varies between organisms. Genetic and biochemical analysis in the eukaryote *S. cerevisiae* and the prokaryotes *E. coli*, *N. meningitidis* and *S. enterica* have identified several components that can associate with homologues of BamA.

In *E. coli*, genetic analysis and co-precipitation experiments have shown that BamA interacts with four lipoproteins: BamB, BamC, BamD and BamE (YfgL, YfiO, NlpB and SmpA, respectively). Co-precipitation and other analyses (28, 29, 66) using strains with mutations for various subunits of the complex suggested a basis for the overall architecture of the BAM complex: (*i*) there is direct interaction between BamA and BamB (66), (*ii*) there is a direct interaction between BamA and BamD (66), (*iii*) binding

of BamC occurs through interactions it makes with the C-terminus of BamD (66, 87) and (*iv*) BamE interacts directly with BamA, BamC and BamD, but not with BamB (29). The peripheral subunits of the BAM complex are reviewed below.

1.8.1 BamB: a β-propeller "scaffolding" subunit of the BAM complex

BamB was first identified in *E. coli* using a chemical-genetic approach in which suppressors were identified by adding toxic molecules to bacterial strains that had a leaky outer membrane (88). One such suppressor was a loss of function mutation in *bamB*, which allowed the *E. coli* mutants to grow slowly in the presence of an otherwise toxic compound (88). Deletion of *bamB* results in reduced levels of outer membrane proteins (88, 89). BamB from α-proteobacteria has seven or eight predicted Pyrrolo-quinoline quinone (PQQ) enzyme repeats, a motif representative of β-propeller structures found in some enzymes and in protein domains involved in protein-protein interactions (90). The PQQ enzyme active site residues are not present in BamB, demonstrating that this protein is incapable of enzymatic catalysis. Homology searches suggest that both the α-proteobacterial and the *E. coli* BamB fit the profile of domains with an eight-bladed β-propeller fold (16). Recently, the crystal structure of BamB was solved that displays an eight-bladed β-propeller structure, with four β-strands contributing to each blade (Figure 1.5) (91). Figure 1.5 shows mutations present in one of the β-propeller motifs of BamB that disrupted interactions with the POTRA domains of BamA (16, 87, 91). The β-propeller structure of BamB suggests its interaction with β-strands in the POTRA domains of BamA and/or its ability to assist stabilising nascent β-strands in substrate proteins. While BamB is found in several species of α-proteobacteria including *Caulobacter* and *Rickettsia*, no related sequence was found in *Brucella*, and BamB is also absent from the BAM complex in *Neisseria* (11). It was recently shown that cargo-bound SurA can bypass interaction with BamB and interact directly with BamA (87). It may be that only some protein substrates interact with BamB prior to assembly since SurA and BamB single deletions have a similar phenotype (58).

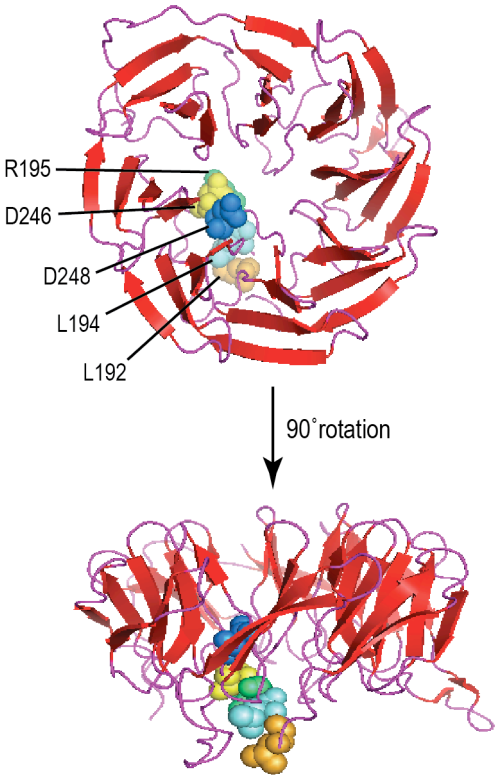


Figure 1.5: **Crystal structure of BamB from *E. coli*.** Five amino acid residues are highlighted as coloured spheres and correspond to the mutations of BamB (L192, L194, R176, D246 and D248) that cause defects for docking to the POTRA domains of BamA (87). Annotations of BamB are adapted from (16, 91). PDB code for the structure is 3P1L.

1.8.2 The BamC subunit of the BAM complex

E. coli mutants lacking BamC have mild defects in outer membrane biogenesis and moderate outer membrane permeability defects including sensitivity to rifampicin (28, 29, 66, 89). A double knockout of *bamC/surA* has a synthetic lethal phenotype, suggesting BamC has an important role for outer membrane protein assembly, possibly redundant to that of SurA (11). A BamC homolog has not been found in any of the α-proteobacteria (Table 1.1) but a distinct (or highly diverged) protein might fulfill the

function of BamC. Alternatively, if the function of BamC is redundant, another component of the BAM complex might compensate.

Table 1.1: Hidden Markov model analysis of the lipoprotein components in the BAM complex[1]

	γ class	β class	α class		
Lipoprotein	***Escherichia***	***Neisseria***	***Caulobacter***	***Rickettsia***	***Brucella***
BamB	P77774	None	Q9A7R7	Q9ZDU2	None
BamC	P0A903	Q9JZR5	None	None	None
BamD	P0AC02	Q9K0B1	Q9A6U9	Q9ZDY1	Q8YI58
BamE	P0A938	Q9K1F0	Q9A8I8	Q9ZCG9	Q8YGH5

[1]Hidden Markov models for each protein family were used to find components of the BAM complex in all bacteria for which complete genome information is available. The species listed in the table are *Caulobacter cresentus* CB15*, Rickettsia prowazekii, Brucella melitensis* 16M*, Escherichia coli* K12 and *Neisseria meningitidis* serogroup B (16).

1.8.3 BamD: a predicted TPR-rich subunit of the BAM complex

BamD is found ubiquitously in Gram-negative bacteria including non-proteobacteria such as *Treponema, Chlorobium* and *Chlamydophila* (16, 28, 66, 89). In *N. gonorrhoeae*, the homologous protein is a peptidoglycan-associated lipoprotein that was called competence lipoprotein (ComL) and, as a consequence, BamD homologs are often annotated as encoding a 'DNA uptake lipoprotein'. The ComL protein is covalently linked to the peptidoglycan in *N. gonorrhoea* (92). It is suggested that ComL might have a role in modulating the composition of the peptidoglycan layer to facilitate the transport of outer membrane proteins across the periplasm (11). A transposon insertion into the middle of ComL resulted in reduced cell size, aberrant cellular morphology and transformation

deficiency (93), presumably a result of the altered outer membrane properties of these mutants. In addition, transposon insertion in the C-terminal region of BamD of *E. coli* showed increased outer membrane permeability and lower amounts of outer membrane proteins. These results suggest that the N-terminal region of BamD is functionally significant. A depletion strain of *bamD* in *E. coli* shows phenotypic changes that are similar to the *bamA* depletion strain in the same species (66). The BamD protein in *Rickettsia* and other α-proteobacteria each strongly predict to have at least three tetratricopeptide repeat (TPR) motifs. However, it is worth noting that the TPR consensus sequence is found primarily in eukaryotic protein sequences and more reasoned analysis using multiple sequence alignments suggest BamD probably contains five TPR motifs. TPR or helix-turn-helix motifs are structural elements that enable protein-protein interactions and have been found operating in a number of protein transport pathways (94, 95). For example, the mitochondrial protein import receptor Tom70 is built from multiple TPR elements (96, 97) and binds β-barrel substrate proteins *en route* to the mitochondrial equivalent of the BAM complex (96). A TPR-rich structure might enable BamD to bind partner proteins (like BamC) and/or substrate proteins.

1.8.4 BamE/SmpA/OmlA: the POTRA-like subunit of the BAM complex

E. coli mutants lacking BamE have defects in outer membrane protein assembly (29). This protein is found ubiquitously in proteobacteria, though no related proteins are obvious in other groups of bacteria with outer membranes. The structure of OmlA, from *Xanthomonas* has been solved (98) and recently, the crystal structure of BamE of *E. coli* was determined (Figure 1.6) (99, 100). As in *E. coli*, OmlA in *Pseudomonas* and *Xanthomonas* is required for outer membrane integrity as *omlA* mutants show increased susceptibility to antibiotics (such as rifampin and chloramphenicol) and detergents (29, 101, 102). The OmlA and BamE structure has a POTRA-like fold although the order of the α-helices and β-strands are different. Its similarity to the POTRA domain suggests that it may interact with subunits within the complex or the POTRA domains of BamA

by the process of β-augmentation (80). The strands of OmlA and BamE are exposed like that of BamB for β-augmentation to occur (80).

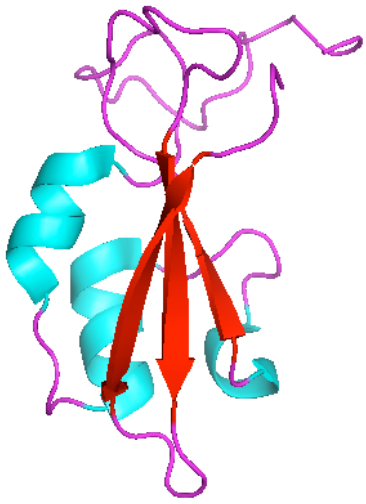


Figure 1.6: Solution structure of BamE from *E. coli*. The structure can be accessed at PDB with the code 2KXX.

1.9 Models for outer membrane protein transport

The mechanism by which proteins pass the periplasm had been a controversial issue for many years and the manner of how they are assembled by the BAM complex is still unclear. The hydrophobicity of outer membrane components does not match the hydrophilic properties of the periplasm and there is no obvious source of energy such as ATP that can drive the transfer of proteins across the periplasm or catalyse the folding and insertion of proteins into the outer membrane. This is in contrast to the ATP-driven function of the Sec machinery in the inner membrane. There are two suggested models

for outer membrane protein transit: *i*) Membrane fusion model and *ii*) Periplasmic model (103). Only the latter model of outer membrane protein biogenesis is repeatedly preferred considering overwhelming evidence for interaction between chaperones and outer membrane proteins, however, the existence of inner and outer membrane links is still controversial (2).

The first hypothesis for protein transfer between the inner and outer membrane related to sites of contact between the two membranes that were first observed by Manfred Bayer in 1968 and termed 'membrane-adhesion sites' or 'Bayer junctions' (104). Bayer also found there were about 200-400 connections per growing cell (104), with similar number of fusions seen by Smit *et al.* (105). At the closest contacts that were observed, both membranes were apart by 5 nm, which is sufficient for at least a single layer of peptidoglycan (106). It was proposed that these adhesion zones might be dynamic structures that form and break to explain the dispersion of an outer membrane protein across the bacterial surface, rather than in discrete regions (105). This was supported by the following observations: *i*) immunolabelling and localisation of bacteriophage MS2 lysis outer membrane protein (107), newly synthesised porins (105) and thioredoxin (108) to the Bayer junctions, and *ii*) existence of an intermediate membranous fraction after membrane fractionation of inner and outer membrane vesicles (108). Indeed, newly outer membrane-incorporated porins were observed in patches near the adhesion zones possibly due to interactions between porins and the peptidoglycan layer (105).

Bayer also argued the presence of these junctions in mitochondria, where gold-labelled translocation intermediates were concentrated at sites where the mitochondrial inner and outer membrane were in close proximity (109, 110). These junctions occurred at 20% of the outer membrane surface of mitochondria (110). In addition, similar structures were observed in chloroplasts (111). However, the existence of these zones of adhesion were challenged by Kellenberger (112). These contact points were dismissed as artefacts of the chemical fixation process since no points of contact were observed in bacterial cells prepared by cryofixation and that the intermediate membranous fraction could be a result of artificial lipid vesicle fusion (112). A rebuttal was published by Bayer (106) but it

remained controversial whether these junctions truly exist.

Another model, which is currently favoured, was proposed for outer membrane protein biogenesis by postulating the transfer of proteins across the periplasm (103). The periplasmic model is based on the existence of a soluble periplasmic intermediate that is released into the medium from spheroblasts (113-116). Spheroplasts have no outer membrane, which is disrupted by EDTA and lysozyme, and secrete various protein species that are destined for the periplasm or outer membrane. In addition, mutant forms of outer membrane proteins such as PhoE accumulated in the periplasm of *E. coli* (117, 118). We do not yet know whether these soluble intermediates truly exist or whether they are by-products of another pathway dealing with misfolded proteins (66). Further validation for this model is offered by findings that the lipoprotein and pilus-subunit pathway involve chaperone-mediated transfer across the periplasm to the assembly sites at the outer membrane (17). This model is detailed by Bos *et al.*, taking into consideration key findings relevant in the field (11).

Although there is no direct evidence for the existence of trans-envelope protein machinery that transfers proteins to the outer membrane, there is a basis for links between the inner membrane and outer membrane. It also has been demonstrated that outer membrane components such as capsular polysaccharides (119) and LPS (120) are exported this way. Polysaccharide export (group 1 and 2) is thought to occur at membrane contact sites in *E. coli*, as observed by electron microscopy (121, 122). In group 2 capsular export, the outer membrane protein KpsD associates with an inner membrane protein KpsE. This association closely positions KpsD with the inner membrane ABC transporter (KpsE) that shuttles polysaccharides across the inner membrane (7). The identification of KpsD in the inner and outer membrane suggests trans-periplasmic export (123). Indeed, mutations of KpsD or KpsDE in *E. coli* affects the transport of polymer to the cell surface since it accumulates in the periplasm (7).

The transport of LPS has also been related to Bayer's bridges, although physical thread-like connections are observed and no close apposition of the two membranes is shown

(120). Further work that combined pulse-chase labelling with membrane fractionation showed that LPS passes the intermediate membrane fraction consisting of inner membrane, outer membrane and peptidoglycan as it is transported from the inner to the outer membrane (124). This intermediate fraction was suggested to contain adhesion zones (125). The most validating data available for the existence of these links is that LPS could be transferred from the inner to the outer membrane in cells that lacked periplasmic contents (126) and that all subunits of the trans-envelope complex can be co-purified from total membranes (12).

The existence of Bayer's junctions is still controversial since it is not proven whether these trans-envelope bridges occur at these junctions or the trans-envelope bridge itself mediates transport, independent of the Bayer's junctions (11). Taking into consideration the peptidoglycan layer, it is unknown whether these sites are dynamically transient or fixed and whether they represent a physical link for continuity between the two membranes or sites where the inner and outer membrane appear to come into close apposition (127). The isolation of an intermediate membrane fraction containing peptidoglycan and enzymes involved in its synthesis (124) suggests that the two membranes are not fused. The peptidoglycan layer and its possible interference regarding protein transport across the periplasm or the formation of trans-envelope complexes warrant further investigation since it has been largely ignored (127). The peptidoglycan layer is porous to an extent with a 50 kDa cut-off value for the diffusion of proteins across the layer during normal growth conditions (128). Hence, penetration of peptidoglycan may be required for larger proteins.

1.10 Why use Caulobacter as a model to study the BAM complex?

The genus *Caulobacter* consists of a collection of Gram-negative, rod-shaped cells that belong to the α-class of proteobacteria. *Rickettsia*, an α-proteobacterium, is often thought to be most closely related to the bacterium that gave rise to mitochondria in eukarytotic cells. The idea of studying the BAM complex in α-proteobacteria was sparked by the

presumption that a simple or primitive form of the protein assembly machinery could reside in this class, which eventually gave rise to the SAM complex in mitochondria. Phylogenetic analysis shows the mitochondrial Sam50 sequences form a sub-clade within the BamA sequences of all α-proteobacteria (129). It is possible that the evolutionary changes may have been present in the α-proteobacterial BAM complex prior to endosymbiosis and perhaps there were functional homologues of Sam35 and Sam37, the peripheral subunits of the SAM complex. In addition, it was important to determine species-related differences in the α-proteobacterial BAM complex compared to its β- and γ-proteobacterial counterparts. Since *Rickettsia* are obligate parasites and therefore difficult experimental organisms, *Caulobacter* was chosen as a model α-proteobacterium since it is non-pathogenic and amenable to lab manipulations.

Chapter 2

Analysis of two members of the Omp85 superfamily: BamA and the YtfM-like protein in *C. crescentus*

2.1 INTRODUCTION

The Omp85 superfamily consists of integral outer membrane proteins that are defined by their phylogenetic relationships, secondary structure predictions and a role in protein translocation (76, 84, 130, 131). In regard to the function, members of the Omp85 superfamily of proteins can be divided into two classes: one class inserts proteins into the outer membrane and the other transports proteins across the outer membrane (132). Members of the BamA/Sam50 class are highly conserved with an essential role in the assembly and insertion of β-barrel proteins (68, 84, 133, 134) for maintaining cell envelope integrity and cell viability in Gram-negative bacteria and eukaryotic cells (84, 85, 133-135). In *E. coli*, BamA forms a complex with four different lipoproteins (BamB, BamC, BamD and BamE) (136, 137). The second class of proteins is involved in protein translocation across the outer membrane and consists of Toc75 from chloroplasts of plant cells (138) and the integral membrane components, called TpsB, of the Two-partner secretion (Tps) in Gram-negative bacteria (131, 139). Toc75 shares its ancestry with BamA/Sam50 homologues but it demonstrates functional similarity to TpsB proteins. Unlike the other characterised proteins of the Omp85 superfamily, TpsB proteins are not known to form complexes with other proteins. Instead, TpsB proteins act alone to translocate large β-helical proteins such as cytolysins, iron-acquisition proteins and adhesins, across the outer membrane and into the extracellular milieu (84, 86, 131). Examples of well-studied TpsB proteins are FhaC from *B. pertussis* (36, 86), ShlB in *S. marcescens* and HMW1B in *H. influenzae* (86). Despite their topological and functional

homology, TpsB proteins demonstrate specificity for their substrates (86). This is in stark contrast to the versatility of BamA and its partner proteins within the complex.

Although these two classes of proteins vary in function, they are structurally similar with a C-terminal integral β-barrel linked to a soluble periplasmic domain (84). There is also a third protein group in the Omp85 superfamily, which possesses such structural similarity but its biological function has remained enigmatic. In *E. coli*, the representative of this third group have been designated YtfM, and the high conservation of YtfM homologues in Gram-negative bacteria testifies its biological importance (70). In this Chapter, a comparative analysis of BamA and Omp68 (68 kilodalton outer membrane protein), the YtfM homologue in *C. crescentus,* is presented. The proposed function of Omp68 and its predicted partner protein (P150) was also investigated using bioinformatics.

2.2 MATERIALS and METHODS

2.2.1 Bioinformatics

Cluster analysis of sequences (CLANS) algorithm (140) was used for phylogenetic analysis of full-length proteins and POTRA domains of Omp85 superfamily proteins (sequences are given in Appendix A1). For clustering, 69 TpsB family proteins (includes the archetypal FhaC, ShlB and HMW1b), 69 Omp68 and 106 BamA homologues were chosen from various species. Duplicate or very similar sequences were discarded from the dataset to lessen the probability of a skewed result. For clustering of POTRA domains, a list of BamA (POTRA 1 to 5), TpsB (1-2) and YtfM-like (1-3) POTRA domains were provided by Arnold *et al.* (141). Omp68 and *E. coli* YtfM POTRA domains were predicted by secondary structure predictions using PSIPRED and added to the given list. A protein sequence with a predicted β-β-α-α-β fold was classified as a POTRA domain (130). Clustering was done to equilibrium in 3D at a *P*-value cut-off as indicated for BLASTp all-against-all searches. Singletons were removed from the final CLANS output.

Operons were provided in an operons prediction database at http://www.microbesonline.org/operons/. The presence of signal peptides was determined using SignalP (142). Sequence alignments were performed using CLUSTAL-W (143) in JalView. The probable secondary structure of proteins was inspected using BetaWrapPro for ß-helix prediction (144) and PSIPRED for α-helix and ß-strands prediction (145). Functional domains and motifs were searched using Pfam (146) and ProDom (147). Building of HMM models and database search were performed with the HMMER package version 2.3.2 (148), and used to search the UNIPROT dataset (Release 12.4, containing SWISS-PROT Release 54.4 and TREMBL Release 37.4) as described previously for mitochondrial protein translocase components (149).

2.2.2 Cloning of omp68 and bamA

The CB15 strain of *Caulobacter crescentus* was used in this study as a source of chromosomal DNA for gene amplification. *C. crescentus* was grown in PYE media (150) at 30°C. Full-length *omp68* was cloned into pET22b with primers *omp68Nde*IFor: 5'-CGGCTTAATGAGAAGCGCTCGTGACGGAGG-3' and *omp68Xho*IRev: 5'-CGGCCTCGAGAAAGCTTTGCCCGATGCTGA-3', followed by *Nde*I-*Xho*I restriction digest and ligation. *bamA* (without the predicted DNA sequence encoding the signal sequence) was cloned into pET22b with primers *bamANde*IFor: 5'-CGGCCATATGCAAGCGGCCCAGACGGGCGTGGT-3' and *bamANot*IRev: 5'-CGGCGCGGCCGCTTGGAACCTTGTGGAGGT-3', followed by *Nde*I-*Not*I restriction digest and ligation. The ligation mixtures were transformed into competent *E. coli* DH5α cells, and the recombinant clones were selected on ampicillin-containing agar plates. The cloned inserts were sequenced. Many unsuccessful attempts were made to clone *omp68* into the pET22b vector, as a result of frame-shift mutations possibly due to leaky expression of the protein in *E. coli* DH5α. However, using 1% (w/v) glucose to suppress leaky expression, this problem was circumvented.

2.2.3 Expression and purification of Omp68 and BamA

Omp68 and BamA were expressed with a hexa-histidine tag to enable immobilised metal affinity chromatography to be used in the purification process. Initially, full-length *bamA* was cloned and expressed but heterologous expression levels were too low using various growth conditions. Removal of the signal sequence improved expression and produced inclusion bodies. Omp68 was also expressed in inclusion bodies (IBs) in BL21 (DE3) cells cultured aerobically at 37°C in Luria-Bertani (LB) and 100 μg/mL ampicillin. Protein expression was induced at OD_{600} of 0.6 using 0.5 mM IPTG for 3 h, followed by centrifugation of the cultures at 10, 000 × g for 10 min. IBs were isolated by lysing cells using the extraction buffer: 0.1% (v/v) Triton X-100, 0.2 mg/mL lysozyme, 25 μg/mL DNAseI, 0.1% (v/v) ß-mercaptoethanol, 10 mM $MgCl_2$, in PBS pH 7.5. IBs were pelleted

at 10, 000 × g for 10 min at 4°C. The pellet was washed with 1% Triton X buffer (1% Triton X-100, 150 mM NaCl, 50 mM Tris-HCl, pH 8.0) followed by two washes with 150 mM NaCl, 50 mM Tris-HCl, pH 8.0. Washed Omp68 IBs were loaded onto a 5 mL HisTrap column (GE Healthcare) connected to AKTA *prime* FPLC system. Unbound material was removed from the column using Buffer A (8 M Urea, 50 mM Tris-HCl, 200 mM NaCl, 30 mM imidazole, pH 8.0). Bound material was eluted using Buffer B (8 M Urea, 50 mM Tris-HCl, 200 mM NaCl, 1 M imidazole, pH 8.0) using a linear gradient (30 mL) of increasing imidazole concentration.

BamA inclusion bodies were washed twice with 2% Triton X-100, 200 mM NaCl, 50mM Tris-HCl, pH 8.0 followed by two washes with 200 mM NaCl, 50 mM Tris-HCl, pH 8.0. Pellet was then washed twice with 1% Triton X buffer followed by two washes with 150 mM NaCl, 50 mM Tris-HCl, pH 8.0. Washed inclusion bodies and analysed by SDS-PAGE to check for level of purity.

2.2.4 Omp68 and BamA anti-sera

Rabbits were injected with 200 µg recombinant Omp68 or BamA protein (immunogens) initially mixed with Complete Freund's Adjuvant followed by Incomplete Freund's Adjuvant for all subsequent injections. The immunogen and adjuvant were mixed thoroughly to form a stable emulsion that is injected into rabbits subcutaneously. Blood was collected every two weeks post-injection and allowed to clot and retract at 37°C for 1 h. The clotted blood was refrigerated for 24 h before the serum was decanted and clarified by centrifugation. Sera were aliquoted and then stored at -80°C

2.2.5 Gel electrophoresis and Western blotting

Proteins were separated by standard denaturing sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE). Alternatively, semi-native SDS-PAGE was used

according to (151). Semi-native gels did not contain SDS and samples were prepared on ice in sample buffers with β-mercaptoethanol and containing only 0.05% (w/v) SDS instead of 2% (w/v) SDS for standard SDS-PAGE. Incubations were performed for 5 min at either 4°C, 56°C or 100°C. Furthermore, semi-native gels were run at constant 50 V at 4ºC to prevent denaturation.

For blue native-PAGE (BN-PAGE), purified outer membrane vesicles were washed (10 mM Tris, 1 mM EDTA, pH 7.5) and recovered by ultracentrifugation. Outer membranes were resuspended in ACA750 buffer (750 mM *n*-aminocaproic acid, 50 mM Bis-Tris, 0.5 mM Na_2EDTA, pH 7.0) as described in (152) and solubilised with DDM (*N*-dodecyl-β-maltoside, Anatrace) by vortexing every 5 min for 20 min. Samples were pre-cleared (450, 000 × g at 4°C for 20 min; TL-100 ultracentrifuge, Beckman). Samples were analysed by BN-PAGE on 5-16% (w/v) polyacrylamide gels (153) using gel buffers described in (152).

For immunoblots, proteins were transferred to nitrocellulose in 25 mM Tris, 190 mM glycine in 10% methanol. The blots were blocked and incubated with antibodies in Tris-buffered saline (TBS) containing 0.1% Tween-20 and 5% non-fat milk powder. Antibody binding was detected by using sheep or goat anti-rabbit immunoglobulin G (IgG) peroxidase-conjugated secondary antibodies (Sigma) and home-made enhanced chemiluminescence detection (ECL). For ECL reaction, equivalent volumes of Solution 1 and Solution 2 were mixed prior to use. Solution 1 was made by separately mixing 200 mg luminol and 35 mg p-coumaric acid in 4.5 mL and 2.5 mL dimethyl sulfoxide (DMSO), respectively. These were combined and added to 250 mL of 100 mM Tris, pH 9.35 to form Solution 1. Solution 2 consisted of 30 µL of hydrogen peroxide (H_2O_2, 30% stock) in 250 mL of 100 mM Tris, pH 9.35. Both solutions were stored at 4ºC.

2.2.6 Sodium carbonate extraction

C. crescentus culture was spun and washed with 1 × PBS. Washed cell pellet was resuspended in 100 µL of 100 mM sodium carbonate (Na_2CO_3), pH 11.3, and incubated on ice for 30 min with intermittent vortexing. Samples were centrifuged at 80,000 × g for 20 min at 4°C to separate non-extractable (membrane pellet) and soluble proteins (supernatant). The membrane pellet was resuspended with an equivalent volume of 100 mM sodium carbonate (Na_2CO_3), pH 11.3. Both insoluble and soluble fractions were precipitated with trichloroacetic acid (TCA) with a final TCA concentration of 15% (w/v). Samples of the total cell homogenate, supernatant and pellet fractions were analysed by SDS-PAGE and Western blotting.

2.2.7 Subcellular fractionation

We developed a method to purify outer membrane vesicles from *C. crescentus.* CB15N was grown in rich PYE media at 30°C as for which the sole carbon source is amino acids (150). For cultures grown in M2 minimal salts medium, 0.2% glucose was the sole carbon source (150). *C. crescentus* from 1 L cultures were harvested at mid-log phase by centrifugation at 10,000 × g for 12 min at 4°C. Cells were resuspended in 50 mL (10 mM Tris, 0.75 M sucrose, pH 7.5) followed by the addition of 100 µg/mL lysozyme, protease inhibitor cocktail (Roche), 25 µg/mL DNAseI and 10 mM $MgCl_2$. Then drop-wise, 1.5 mM EDTA, pH 7.5 was added to a final concentration of 0.5 mM EDTA. Cells were disrupted by two passages through a French pressure cell at 15,000 psi and cellular debris removed by centrifugation. Membranes were collected (42,000 rpm, 1 h, Beckman Ti45 rotor) and resuspended in 25% (w/v) sucrose, 5 mM EDTA, pH 7.5 and fractionated on a six-step sucrose gradient (35:40:45:50:55:60% (w/v) sucrose in 5 mM EDTA, pH 7.5) by centrifugation (Beckman SW41 rotor, 34,000 rpm, 40 h) with slow acceleration and deceleration. Outer membrane vesicles were collected with an ISCO density gradient fractionator and protein concentration of each fraction was measured using the Bradford assay. All steps were performed at 4°C and outer membranes were stored at -80°C.

2.2.8 Protease accessibility

Caulobacter cell culture was harvested at mid-log growth phase and washed twice with 50 mM Tris-HCl, pH 8.0. Cells were either incubated with polymyxin B (final concentration 2 mg/mL) or 50 mM Tris-HCl, pH 8.0 (control) for 10 min. Cells were digested with proteinase K (final concentration 0.05 μg/μL) for 30 min at 4°C. All samples were TCA precipitated and then washed twice with cold acetone. Samples were resuspended with equivalent volumes of 0.2 M NaOH and 2 × SDS sample buffer and analysed by SDS-PAGE.

2.2.9 Co-immunoprecipitation

For immunoprecipitations, purified outer membranes (approximately 800-1000 μg protein) were resuspended in ACA750 buffer and DDM as for BN-PAGE, and centrifuged (25,000 × g, 4°C for 10 min). Cross-linked IgG-Protein A Sepharose beads (Sigma) were added to solubilised outer membranes and incubated for 3 h at 4°C with shaking. After six washes with 1 × PBS containing 0.1% (w/v) DDM (final concentration), a final wash with 1 × PBS was used before eluting the immunoprecipitates with 2 × SDS sample buffer. As a negative control, immunoprecipitations were also performed using pre-immune serum collected from the rabbits before being immunised with the BamA antigen.

2.2.10 Mass spectrometry

Protein-containing bands from SDS gels were excised manually and destained using 100 mM triethylammonium bicarbonate (TEAB), pH 8.5 in 50% (v/v) acetonitrile. Gel pieces were dehydrated using 100% (v/v) acetonitrile and air-dried before adding 20 mM dithiothreitol (DTT) at 60 °C for 1 h and then 100 mM iodoacetamide at RT for 30 min in the dark. Gel pieces were washed with destaining solution and dehydrated as above

followed by rehydration in the presence of 250 ng sequencing-grade trypsin (Sigma) in 50 mM TEAB. Digests were allowed to proceed overnight at 37°C. Tryptic peptides were extracted using 0.1% (v/v) formic acid and analysed by liquid chromatography-tandem mass spectrometry (LC-MS/MS) using a Shimadzu Prominence nano-flow HPLC and an Applied Biosystems Q-STAR ELITE mass spectrometer as previously described (154).

2.3 RESULTS

2.3.1 A comparative bioinformatics analysis of the Omp85 family proteins in C. crescentus

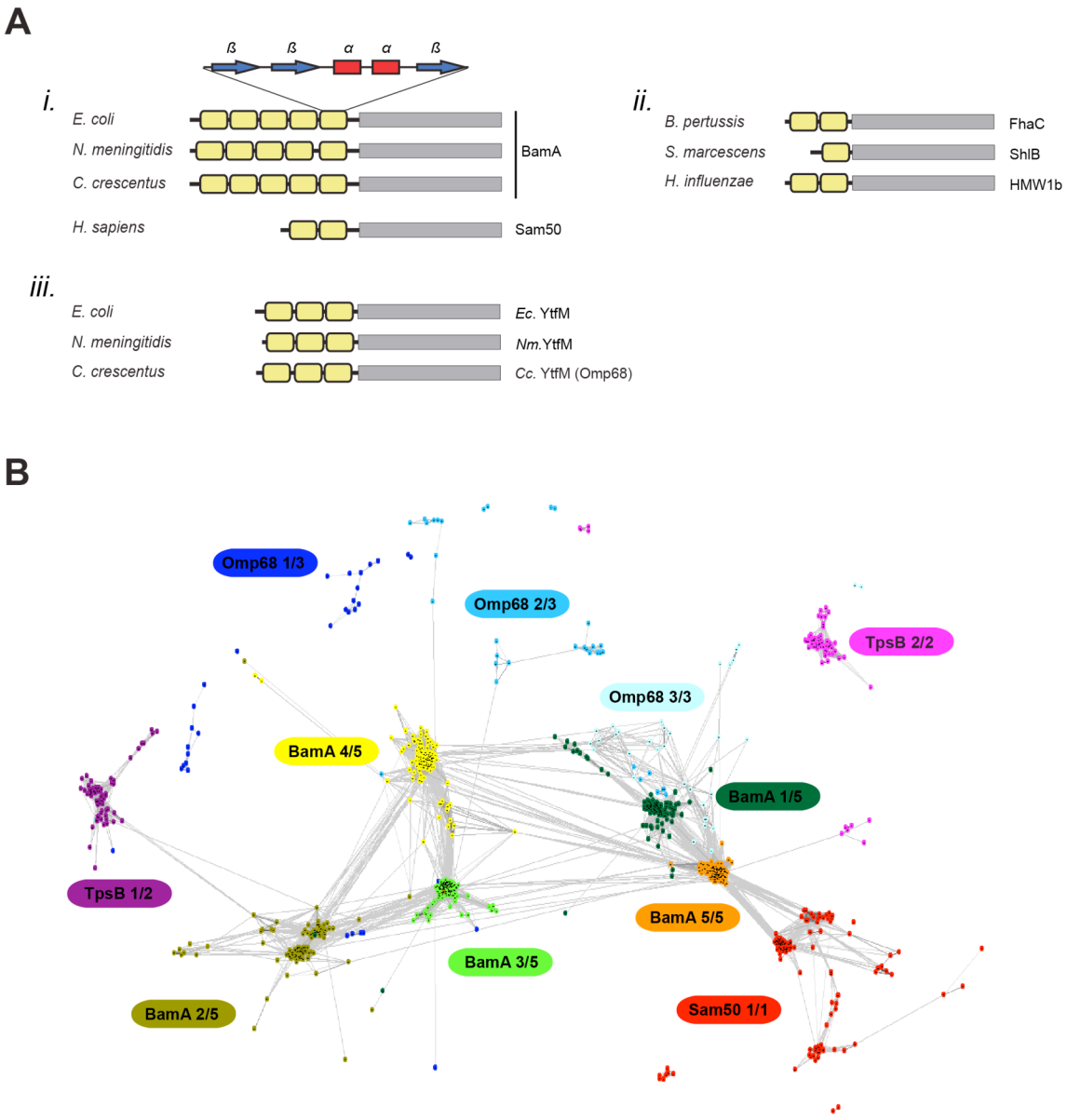
The Omp85 superfamily consists of three sub-groups of outer membrane proteins that share similar domain features; a C-terminal β-barrel domain that forms a membrane pore, linked to a number of polypeptide transport-associated (POTRA) domains. Each POTRA consists of a sequence of two β-sheets, 2 α-helices followed by a single β-sheet (130). The domain architecture of the BamA protein from *C. crescentus* and other α-proteobacteria conform to the basic domain structure seen in most bacterial BamA proteins with five predicted POTRA domains (Figure 2.1A). A similar observation is made for BamA from *E. coli* (γ-proteobacteria) and *N. meningitidis* (β-proteobacteria). The eukaryotic functional homologue of BamA is Sam50, which has only one predicted POTRA domain. The TpsB subgroup of proteins has reduced number of POTRA domains compared to BamA with either one or two POTRA domains. The other group comprising of Omp68 and its homologues have three predicted POTRA domains at the N-terminus (Figure 2.1A). Thus, truncation of the POTRA domains is an important feature of the Omp85 superfamily.

To determine the evolutionary relationship between POTRA domains of Omp68, CLANS analysis was performed with previous data relating to clustering of POTRA domains of BamA and TpsB homologues (Figure 2.1B) (141). The clustering shows that unlike POTRA domains 1 and 2 of TpsB proteins, the C-terminal POTRA-3 domain of Omp68 is closely linked to the C-terminal POTRA-5 of BamA. In comparison, POTRA-2 and POTRA-3 of Omp68 homologues have undergone evolutionary modifications and do not cluster well in relation to their sequence similarities.

The evolutionary and functional link between the mitochondrial Sam50 and the α-proteobacterial BamA is demonstrated by phylogenetic analyses (84). To visually determine the phylogeny of Omp68 within the Omp85 superfamily, CLANS algorithm

was used to cluster selected sequences in 3D space (140). Figure 2.1C shows three distinct clusters of proteins in 2D space; each cluster contains TpsB, BamA or Omp68 homologues. The BamA cluster is contracted compared to the TpsB and Omp68 cluster, suggesting there is more sequence similarities between BamA homologues and more sequence divergence between TpsB and Omp68 homologues. Omp68 is weakly associated with the BamA family cluster with the level of association represented by the number and intensity of lines connecting Omp68 with the cluster.

Figure 2.1: Analysis of the Omp85 superfamily. (A) The domain structure of Omp85 superfamily proteins includes *i*) Omp85 family proteins including BamA and Sam50 homologues, *ii*) TpsB proteins and *iii*) Omp68-like homologues. The predicted POTRA domains are shown in yellow, with the secondary structure β-β-α-α-β indicated above (130), and the predicted β-barrel domain is shown in grey. (B) CLANS result of the POTRA domains of the Omp85 superfamily. Dots, representing protein sequences, are coloured according to their assignment to a cluster. Connections between dots represent the degree of pair-wise sequence similarity, as quantified by BLAST *p*-values (the darker the line, the higher the similarity). (C) CLANS result of Omp85 superfamily proteins. Red (BamA homologues, Omp85 family proteins), Purple (TpsB homologues) and Aqua (Omp68 homologues) dots represent protein sequences from each cluster.


A
i.
ß ß α α ß
E. coli
N. meningitidis
C. crescentus
BamA
H. sapiens
Sam50
ii.
B. pertussis
S. marcescens
H. influenzae
FhaC
ShlB
HMW1b
iii.
E. coli
N. meningitidis
C. crescentus
Ec. YtfM
Nm.YtfM
Cc. YtfM (Omp68)
B
Omp68 1/3
Omp68 2/3
TpsB 2/2
Omp68 3/3
BamA 4/5
BamA 1/5
TpsB 1/2
BamA 5/5
BamA 3/5
BamA 2/5
Sam50 1/1
C
Omp68
BamA
TpsB

2.3.2 Subcellular localisation and topology of the Omp85 family of proteins

To determine whether Omp68 is an integral membrane protein, sodium carbonate extraction was performed. This disrupts protein-protein interactions but does not affect protein-lipid interactions, so that integral proteins are found in the pellet and the soluble proteins and peripherally associated proteins are found in the supernatant. Omp68, like BamA, is found in the membrane pellet whereas the control protein, F1β, is in the supernatant fraction (Figure 2.2A).

The predicted β-barrel structures of BamA and Omp68 strongly suggest that they are outer membrane proteins in *C. crescentus*. To verify this, membrane vesicles were prepared and separated into inner and outer membrane vesicles using a sucrose gradient. The inner membranes fractionated at ~ 50% sucrose whereas outer membranes fractionated at ~ 60% sucrose. The enrichment of Omp68 and BamA in fractions 15-19 shows that it is a component of the outer membrane (Figure 2.2B). The inner membrane protein, TimA, is mainly found in fractions 8-13. There is some apparent BamA in the inner membrane vesicles, but this is unavoidable due to the sensitivity of the BamA anti-serum and inherent experimental limitation of achieving complete separation of inner and outer membranes.

BamA and Omp68 were analysed for heat-modifiable mobility, which is a feature for many outer membrane proteins of β-barrel topology (155, 156). In this experiment, a small amount of SDS is added to extract folded forms of β-barrel proteins. However, after subsequent heat treatment, the β-barrel domains dissociate and the proteins migrate to their expected size on a polyacrylamide gel. Unboiled samples of folded β-barrel proteins retain their conformation and migrate faster than expected on the gel. Immunobloting against BamA shows heat-modifiability for boiled and non-boiled samples in the presence of 0.05% SDS (Figure 2.2C). Although most of the BamA protein migrates at the expected size of 87 kDa, a small proportion of BamA migrates more rapidly when the sample is unboiled (4°C). The same behaviour is observed in the presence of a higher SDS concentration of 2% (data not shown). In comparison, Omp68 does not have such

distinct changes in mobility with or without heat treatment. Instead, heating the sample at 100°C resulted in degradation of Omp68 (Figure 2.2D). This might suggest that BamA has a structurally more compact and stable β-barrel domain than Omp68 as it is able to partially resist the effect of SDS and heat treatment and exhibit heat modifiability.

To determine the topology and whether the periplasmic domain of BamA and Omp68 could be distinguished, a protease accessibility experiment using proteinase K was performed. Addition of polymyxin B and proteinase K does not degrade BamA and Omp68 as might be expected (Figure 2.2E). This suggests the POTRA domains of these proteins are protected from degradation, perhaps due to a compact conformation or the presence of partner proteins.

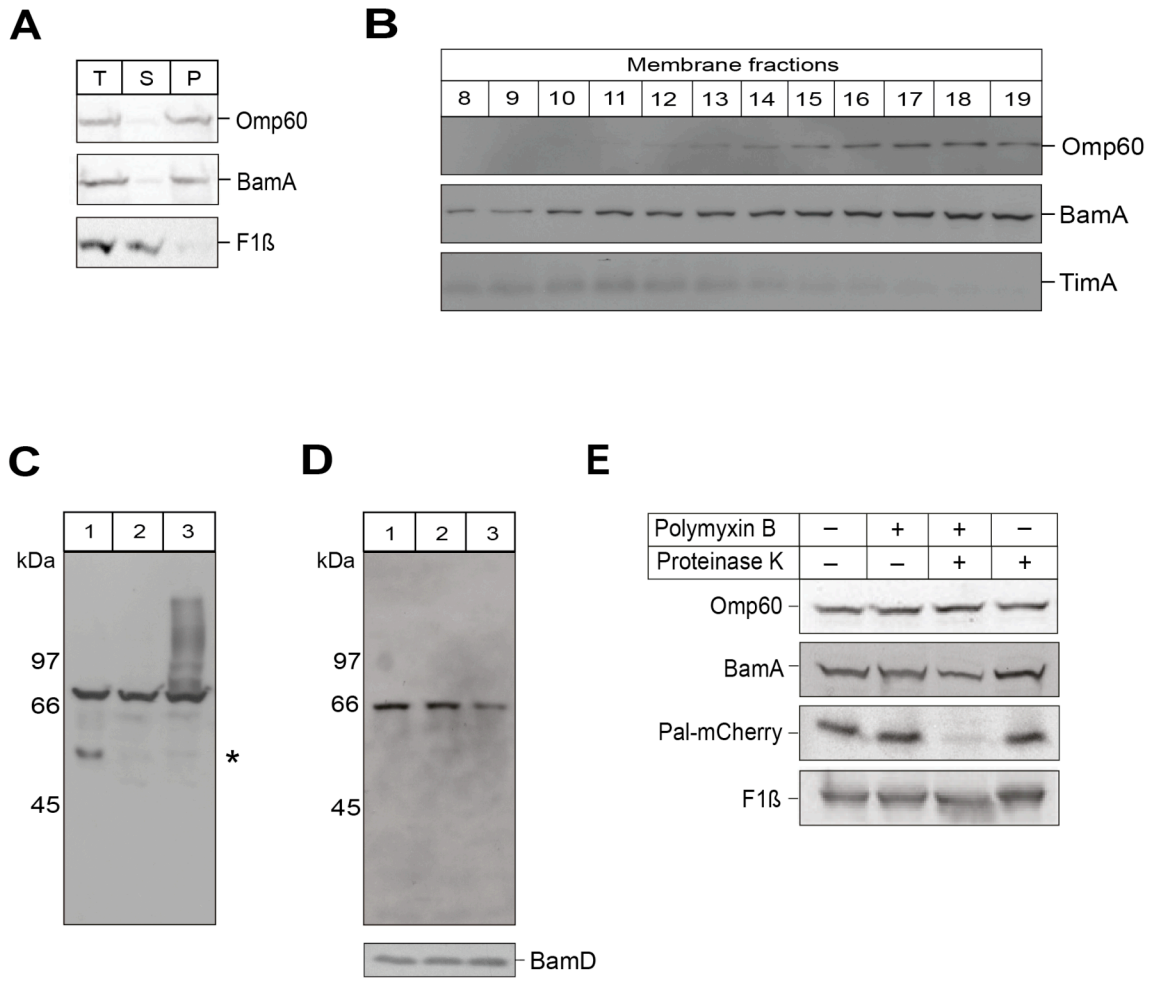


Figure 2.2: Biochemical characterisation of BamA and Omp68. (A) Sodium carbonate extraction of *C. crescentus*. Samples loaded onto gel are T (total cell lysate), S (supernatant) and P (pellet), which were probed with anti-serum against Omp68, BamA and F1ß (cytoplasmic protein). (B) Subcellular localisation of Omp68 showing sucrose gradient-fractionated inner and outer membranes probed with Omp68, BamA and TimA (inner membrane). (C-D) Semi-native PAGE of purified outer membranes from wild-type CB15N. Protein samples were analysed after incubation at 4°C (lane 1), 56°C (lane 2) and 100°C (lane 3) for 5 min. Western blots were probed with anti-BamA sera (C), anti-Omp68 sera (D) and anti-BamD for loading control (D, lower panel). The (*) represents a shift in migration of BamA. Positions of molecular mass markers are shown on the left. (E) Protease shaving of *C. crescentus*. Cells were incubated with (Lane 2, 3) and without (Lane 1, 4) polymyxin B followed by addition of proteinase K (Lane 3, 4). Samples were analysed by SDS-PAGE and immunoblotting. Blots were incubated with antibodies against Omp68, BamA, OmpA-mCherry (membrane-attached lipoprotein) and F1ß (cytoplasmic control protein).

2.3.3 Outer membrane remodelling and protein complexes in C. crescentus

In *C. crescentus*, the outer membrane proteome is remodelled in cells growing on minimal medium compared to complex medium, as evidenced by the identity of each of the major membrane proteins determined by mass spectrometry (Figure 2.3A). The abundance of TonB-dependent receptors is also evident. Levels of BamA remain constant in both growth conditions whereas levels of Omp68 slightly decrease in minimal media.

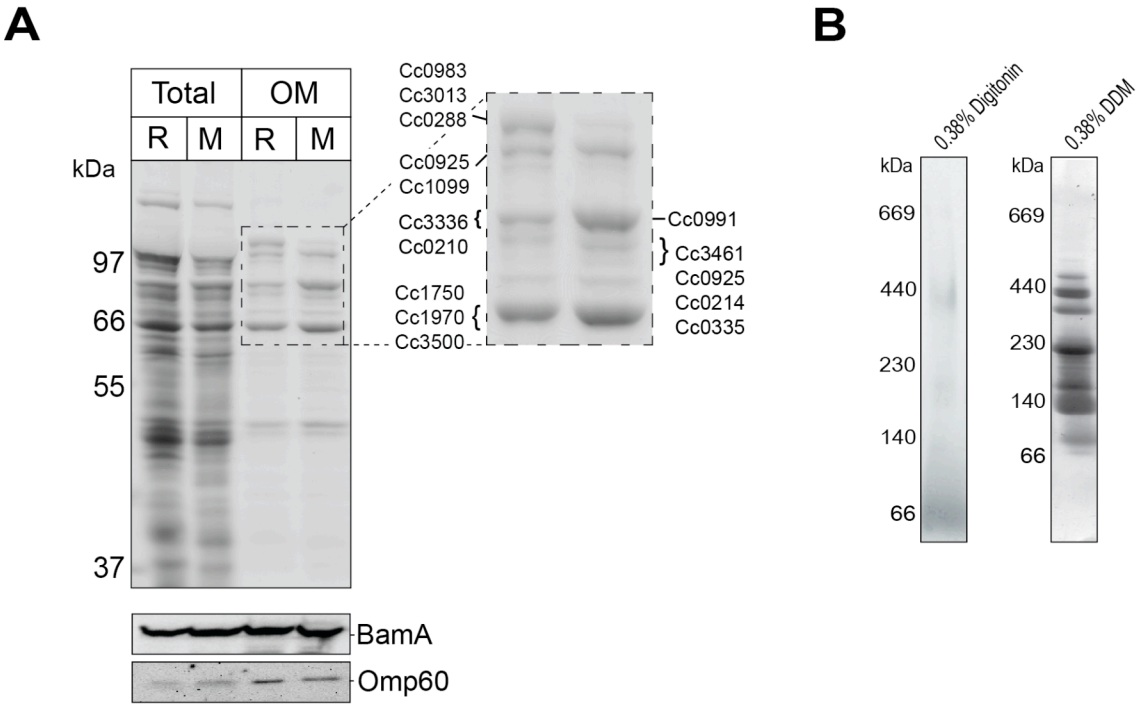


Figure 2.3: Outer membrane proteins of *C. crescentus*. (A) Cultures of *C. crescentus* were grown in either rich (R) or minimal (M) media and cells harvested for fractionation (see Methods). Total cell extracts ("Total", 10 µg protein) and outer membrane vesicles ("OM", 10 µg protein) were loaded for analysis by SDS-PAGE for Coomassie blue staining (upper panel) and immuno-staining with an anti-serum to BamA and Omp60 (lower panels). (B) Outer membrane vesicles (100 µg protein) were solubilised with 0.38% (w/v) digitonin or DDM and loaded for analysis by BN-PAGE. The Coomassie blue staining shows the major outer membrane protein complexes.

In order to make an initial assessment of whether BN-PAGE could be used to study outer membrane proteins in *C. crescentus*, various detergent solubilisation conditions were tested. Figure 2.3B shows a Coomassie-stained gel where the major outer membrane proteins are effectively solubilised and migrate in a clear pattern as well focused complexes by this procedure; dodecyl maltoside (DDM) appears to be superior to digitonin for these experiments.

2.3.4 BamA and Omp68 form protein complexes

To investigate the size of the Omp68 protein machinery in *C. crescentus*, outer membrane vesicles were solubilised with different amounts of DDM and separated by BN-PAGE. Antibodies against Omp68 reveal a complex of ~390 kDa in size that is not disrupted with increasing detergent concentration (Figure 2.4A).

To determine whether or not the BAM complex in *C. crescentus* is modular, outer membrane vesicles were titrated with either digitonin or DDM and analysed by BN-PAGE. The mild digitonin detergent solubilises the BAM complex in a form that is ~500 kDa (Figure 2.4B). Increasing the concentration of digitonin to above 2% (w/v) disrupts the complex. However, given the success in generally solubilising the outer membrane complexes with DDM (Figure 2.4B), its effect on the BAM complex was tested. Titrations of DDM disrupted the ~500 kDa complex into ~300 and ~150 kDa modules (Figure 2.4C).

In *E. coli*, BamA is associated with four lipoprotein partners, and analysis of the *C. crescentus* genome revealed genes that could encode three of these lipoproteins: *bamB* (CC1653), *bamD* (CC1984) and *bamE* (CC1365). There is no *bamC* gene in *C. crescentus* or any other α-proteobacterium (16). In order to determine the subunit composition of the BAM complex in *C. crescentus*, outer membrane vesicles were solubilised in 0.75% (w/v) DDM and incubated with the BamA anti-serum. Immunoprecipitation shows four proteins specifically precipitated together with BamA

(Figure 2.4D). The bioinformatics predictions were validated by the identification of the lipoproteins BamB, BamD and BamE by mass spectrometry (Table 2.1). In addition, three novel lipoproteins were identified of which one was annotated as Pal (peptidoglycan-associated lipoprotein).

To determine the relative affinities of the BAM subunits to BamA, low [0.75% (w/v)] and high [2.25% (w/v)] concentration of DDM was used to solubilise outer membrane proteins prior to immunoprecipitation with the BamA anti-serum (Figure 2.4E). This showed that BamA is weakly associated to Pal and the other two novel lipoproteins as the interaction is disrupted using high concentration of DDM. In comparison, BamB, BamD and BamE strongly interact with BamA as these proteins precipitated with BamA at high concentration of DDM.

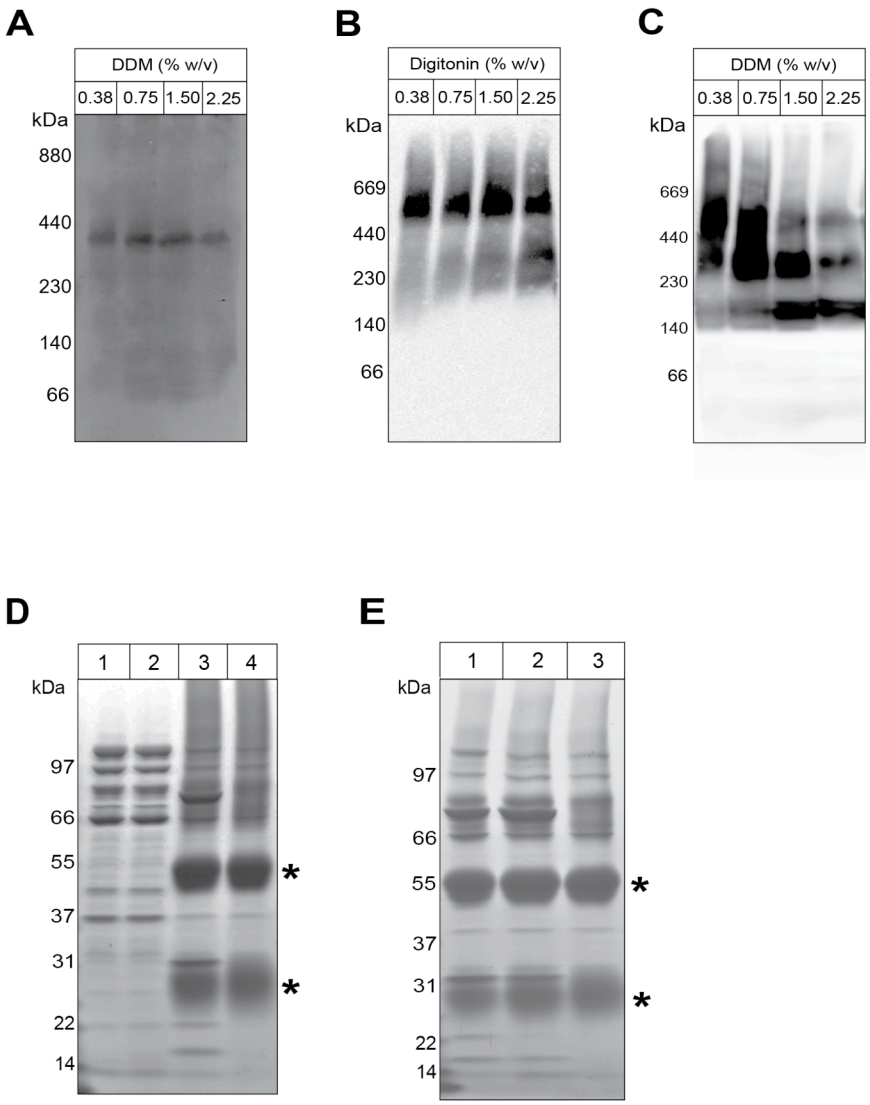


Figure 2.4: Protein complexes of Omp68 and BamA. (A) Outer membrane protein complexes were separated by BN-PAGE using different solubilising concentrations of DDM. Samples were probed using anti-serum against Omp68. The migration positions of the molecular weight markers are shown. (B, C) Outer membrane protein complexes were separated by BN-PAGE using different concentrations of digitonin B) or DDM C) prior to Western blotting using BamA anti-serum. (D) Outer membrane vesicles (800 µg protein) were solubilised with 0.75% (w/v) DDM and the BAM complex immunoprecipitated with an anti-serum recognising the BamA subunit. Samples of the total extract (T), unbound fraction (U) and immunoprecipitate (IP) are shown. (E) BAM complex was immunoprecipitated using BamA anti-serum added to outer membrane vesicles that were solubilised with 0.75% (w/v) (Lane 1) and 2.25% (w/v) (Lane 2) DDM. Immunoprecipitate obtained with pre-immune serum was loaded in Lane 3. Asterisks (*) indicate the position of the IgG heavy and light chains.

Table 2.1: Mass spectrometry data summarising the identification of immunoprecipitated proteins[1]

Protein ID	SwissProt	Predicted size (kDa)	MOWSE score	Peptides matched	Sequence coverage
BamA	Q9A711	86,000	3660	69	69%
BamB	Q9A7R7	50,480	1388	39	57%
BamD	Q9A6U9	34,247	1275	34	68%
Pal	Q9A3H5	20,470	358	11	37%
BamF	Q9AAA7	23,277	417	14	47%
BamG	Q9A7Y3	22,933	154	3	16%
BamE	Q9A8I8	17,509	482	25	55%

[1]Predicted sizes were given by the search output. In addition to the number of high confidence peptides identified by MS/MS data and sequence coverage; a MOWSE score is included, which is a probabilistic score that indicates the match of the experimental peptide precursor masses (peptide mass fingerprint) to the sequence of a candidate parental protein. Typically a MOWSE score .75 is considered as significant (154).

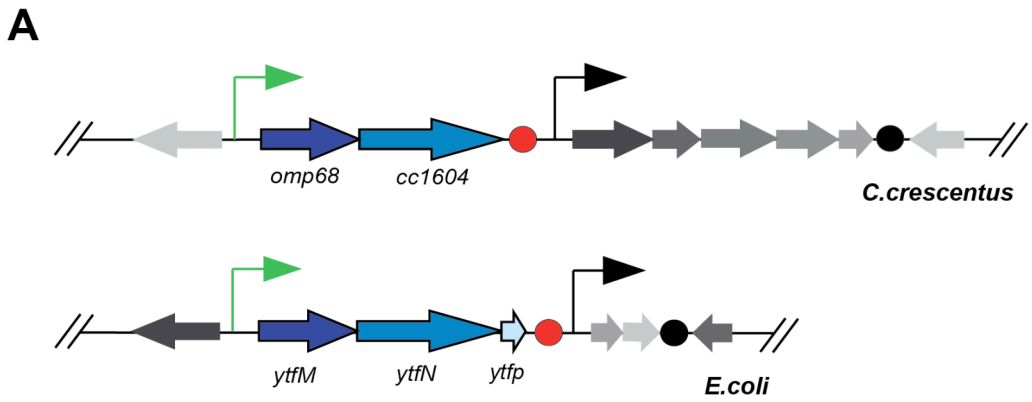
2.3.5 P150 is a possible partner protein of Omp68 but does not have features of TpsA proteins.

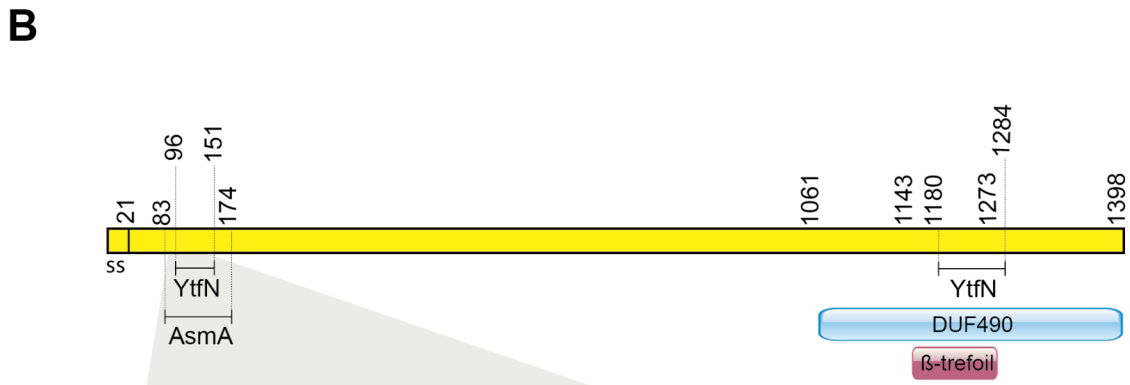
The *omp68* gene is predicted to be in an operon with another gene (annotated as *cc1604*) with a confidence value of 0.905 (Figure 2.5A). For simplicity, we will refer to CC1604 as P150 to account for the size of the protein in kilodaltons. In *E. coli*, and many other γ-proteobacteria, the homologous operon consists of three genes: *ytfM*, *ytfN* and *ytfP* (Figure 2.5A).

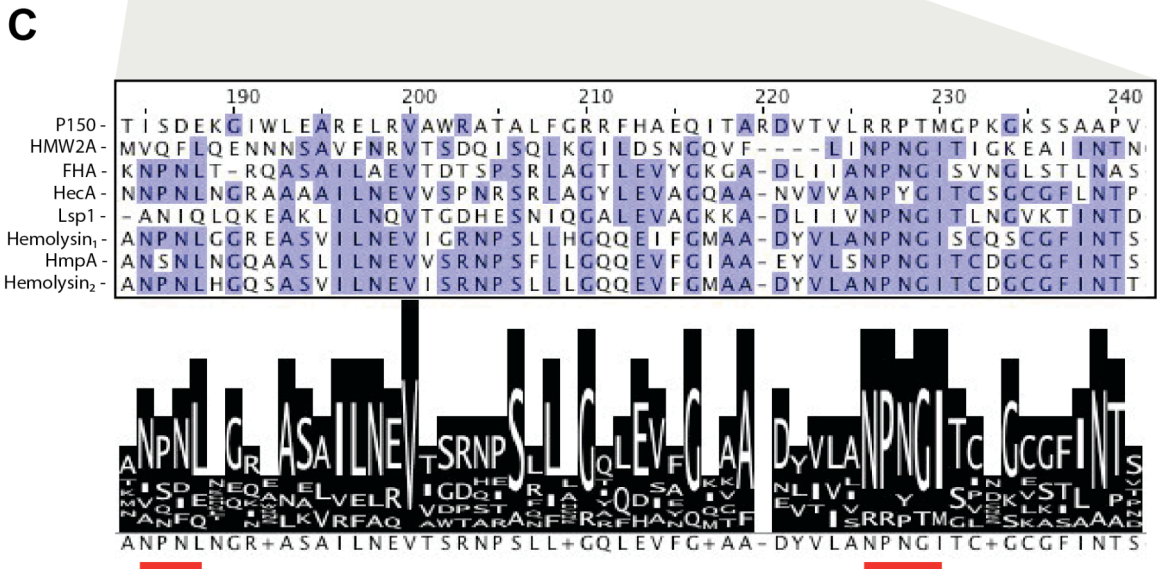
According to SignalP prediction program, a signal sequence is predicted for P150, though with a low score (Figure 2.5B). This suggests it has the signature sequence for translocation through the inner membrane Sec complex. P150 is a large protein (1398 amino acids) and according to Pfam (database of protein families grouped according to multiple sequence alignment and HMMs), it has a DUF490 domain between residues (1061-1398), with the significance of that prediction having an *e*-value of 2.6^{-63}. The DUF490 domain is found in other uncharacterised proteins and belongs to the AsmA-like

OmpF regulator protein superfamily. ProDom predicts YtfN motifs in P150 at position 1180-1284 (PD278822) and at position 96-151 (PD494026). Interestingly, an "AsmA motif" is also detected at position 83-174 (PD126253). Multiple sequence alignment of P150 and representative sequences of TpsA proteins showed that P150 protein lacks the NPNL and NPNGI motifs in the TPS domain (Figure 2.5C), that is important for their translocation or export by TpsB proteins.

Figure 2.5: Protein sequence features of P150. (A) Genetic organisation of the *C. crescentus omp68/p150* operon and its corresponding *ytfM/ytfN/ytfP* operon in *E. coli*. Thick arrows represent chromosomal genes, thin arrows represent transcriptional initiators and circles indicate transcriptional terminators. (B) Sequence features of P150 highlighting the signal sequence, motifs and domains. (C) Sequence alignment of P150 and TPS domains from representative TpsA proteins (top panel) and the resulting consensus logo (bottom panel). Blue shaded amino acids represent highly conserved residues with at least 50% conservation. The height of the single letters in the logo represents the conservation of the amino acid at each position within the region. A red bar below the consensus logo indicates the conserved NPNL and NPNGI motifs. Proteins shown are *Caulobacter crescentus* P150, *Haemophilus influenzae* HMW2A, *Bordetella pertussis* FHA, *Erwinia chrysanthemi* HecA, *Haemophilus ducreyi* Lsp1, *Serratia marcescens* hemolysin 1, *Proteus mirabilis* HpmA and *Edwardsiella tarda* hemolysin 2.


A
omp68
cc1604
C.crescentus
ytfM
ytfN
ytfp
E.coli


B
21
83
96
151
174
1061
1143
1180
1273
1284
1398
ss
YtfN
AsmA
YtfN
DUF490
ß-trefoil


C
190
200
210
220
230
240
P150 -
HMW2A -
FHA -
HecA -
Lsp1 -
Hemolysin1 -
HmpA -
Hemolysin2 -
ANPNLNGR+ASAILNEVTSRNPSLL+GQLEVFG+AA-DYVLANPNGITC+GCGFINTS

2.4 DISCUSSION

A bioinformatics search of Omp85 family proteins revealed two members of the Omp85 family, BamA and a somewhat smaller protein, Omp68 in *C. crescentus*. Both proteins are found ubiquitously in Gram-negative bacteria (70), which is indicative of its conservation throughout evolution and its physiological importance. Omp68 is a 68 kDa protein and shares general domain features with BamA and TpsB proteins (Figure 2.1A). In BamA and TpsB proteins, the role of the C-terminal β-barrel domain is to integrate the protein into the bacterial outer membrane and possibly facilitate oligomerisation to form a pore. The POTRA-containing N-terminus forms the periplasmic portion that can recruit and interact with substrates in a chaperone-like fashion (71, 74, 157, 158), as well as act as a scaffold for binding of other outer membrane proteins as in the case of BamA (141, 158).

The number of POTRA domains for the Omp85 superfamily proteins varies possibly as an adaptive mechanism. These adaptations could be related to substrate selectivity (130) as well as the composition of the complex as each POTRA domain can also be considered a separate functional module to co-ordinate different protein interactions. The BamA homologues from the five different classes (α, β, γ, δ and ε) of proteobacteria generally contain five POTRA domains (159), but exceptional variations are possible. *Myxococcus xanthus* (δ class) has two copies of BamA, one which has seven predicted POTRA domains (141), whereas other bacterial species were shown to have three, four or six (141, 159) with greatest conservation observed for the most C-terminal followed by the most N-terminal POTRA domain (141). In comparison, the structure of the TpsB FhaC protein from *Bordetella pertussis* shows it has two POTRA domains linked to a 16-stranded β-barrel domain (36), similar to HMW1B, whereas ShlB has only one predicted POTRA domain (130). The number of predicted POTRA repeats is three for Omp68 homologues across α, β and γ-proteobacteria (70). It is feasible to deduce the evolutionary lineage of these proteins by considering the number of POTRA domains present and the functional relationship between BamA and Sam50. In this case, the precursor for Omp85 superfamily proteins could be the protein with the maximum

number of POTRA domains such as BamA, which has undergone drastic POTRA truncations to yield proteins with simpler architectures such as Sam50, TpsB or Omp68 homologues as a result of environmental pressures. Arnold *et al.* (141) suggests protein evolution is assumed to generate more complex structures from simple precursors but this may not be the case for these set of proteins (141).

A phylogenetic CLANS analysis of the Omp85 superfamily was performed to visualize pairwise sequence similarities. Distinct clustering has been observed for the β-barrel domains of TpsB and BamA proteins (141). In this analysis, we incorporated full-length sequences of BamA, TpsB and Omp68 homologues to generate three distinct clusters: BamA, TpsB and Omp68-like protein clusters. Omp68 is more closely associated with the BamA subfamily cluster instead of the TpsB cluster. In addition, the far C-terminal POTRA (POTRA-3) domain of Omp68 shares sequence similarity with the most conserved C-terminal POTRA (POTRA-5) of BamA. In contrast, the POTRA domains of the TpsB proteins are highly divergent. This implicates Omp68 as an evolutionary intermediate that gave rise to TpsB-like proteins and may translate to functional similarity between Omp68 and BamA. In *E. coli*, Omp68 is called YtfM, and the *ytfM* knockout strain have only slight growth defects, and no obvious changes in the cell envelope ultrastructure, protein levels of major outer membrane proteins (OmpA, OmpF and TolC) or changes in outer membrane permeability (70). It is possible Omp68 could be a factor for insertion of a subset of proteins into biological membranes; however, there are other features that suggest it could be a protein exporter like TpsB proteins. Hence, at least two distinct functional roles for Omp68 are proposed and considered.

Although Omp68 and BamA are both embedded in the outer membrane of *C. crescentus*, the β-barrel structure of Omp68 is more labile that BamA by semi-native PAGE. BamA seems to be slightly resistant to SDS whereas Omp68 is completely sensitive. A similar finding was observed for YtfM (70). The current understanding is that the β-barrel POTRA domains of Omp68 and BamA are located in the periplasm. The POTRA domains of Omp68 and BamA are highly resistant to degradation by proteinase K, which could suggest they form a highly compact and protease-protected structure in the

experimental condition and/or are protected from proteolysis by interactions with specific partner proteins. I showed by BN-PAGE that Omp68 exists in a ~300 kDa complex and as discussed below, with a potential partner protein is P150.

The outer membrane of *Caulobacter* is enriched in TonB-dependent receptors with an extraordinary 67 different proteins (160) of the 140 predicted outer membrane proteins (160). The major role of these proteins is related to substrate transport across the bacterial outer membrane (161, 162). Growth in complex rich or minimal media results in extensive remodelling of the *Caulobacter* outer membrane (161, 163), with *de novo* synthesis as well as changes in the abundance of existing proteins due to certain environmental cues (161). I demonstrated decreases in the levels of several TonB-dependent receptors as well as Omp68 due to a shift to nutrient-poor medium. It was reported that levels of BamA (ORF06003) apparently increase in minimal growth media to assist with the increased expression and assembly of outer membrane proteins involved in nutrient uptake (161). However, our results indicate the levels of BamA remain constant despite serious remodelling of numerous outer membrane proteins as a result of nutrient availability (Figure 2.3A). This suggests that the activity of BamA or the efficiency of the BAM complex can be adapted to different conditions.

As an initial characterisation of the quaternary structure of Omp68 and BamA, two different non-ionic detergents were used: digitonin and DDM. When we probe for the Omp68 complex in membrane complexes extracted by digitonin and DDM, we detected a complex of ~390 kDa that was extracted with DDM only (Figure 2.4A). One possibility is that this is a hetero-oligomer of Omp68 interacting with the P150 protein or a homo-oligomer of Omp68. It is difficult to ascertain the stoichiometry of either complex due to immeasurable contributions made by bound Coomassie blue-G250 dye and detergent towards the size of the protein complex. It has been suggested that two purified proteins from the Omp85 superfamily, TpsB protein from *H. influenzae*, HMW1B, and BamA from *E. coli,* form oligomers that could be tetramers (52, 76). However, this is the first study on native Omp85 family proteins solubilised from membranes. Many attempts to characterise the subunit composition of the Omp68 complex by co-immunoprecipitation

failed. An alternative method to identify protein complexes is two-dimensional SDS/BN-PAGE (152). However, the abundance of a multitude of TonB-dependent receptors made it difficult to visualise complexes formed by less abundant proteins such as BamA and Omp68.

The motivation to characterise the BAM complex arose from a sophisticated bioinformatics analysis (hidden Markov Models) showing that α-proteobacteria have genes that could encode most, but not all, of the components found in the BAM complex of *E. coli* (Table 1.1) (16). In particular, no sequence encoding a homolog of BamC was found in any species of α-proteobacterium. This suggested two potential scenarios: that the BAM complex in α-proteobacteria is reduced in size, or that yet to be identified components are present in the BAM complex. We characterised the BAM complex from *C. crescentus* by BN-PAGE and immunoprecipitation with an anti-serum directed against the BamA subunit. Antibodies to BamA reveal the BAM complex is ~ 500 kDa in size, but that it dissociates to a complex of ~ 300 kDa, and to smaller core complexes of ~ 150-170 kDa with increasing concentration of detergent. After immunoprecipitation, mass spectrometry revealed that in addition to the BamB, BamD and BamE subunits, the α-proteobacterial BAM complex has three novel subunits (Table 2.1). These novel lipoproteins will be discussed in Chapter 3.

Although we understand the general function of BamA, we do not have a clear understanding of the role of Omp68 in bacteria. Analysis of the *omp68* operon identifies an open-reading frame for *p150* downstream of *omp68*. This is observed for many α-proteobacterial genomes containing homologues of Omp68 and P150. The operon organisation indicates that these two proteins are co-transcribed for a common biochemical function within the cell. In γ-proteobacteria such as *E. coli*, the operon has an additional gene of unknown function called *ytfP*. Omp68 is predicted to be a β-barrel protein that is localised to the outer membrane of *Caulobacter*, whereas P150 is a large β-rich protein that has a signal sequence for translocation through the inner membrane and 1-2 predicted α-helical transmembrane domains that might anchor it in the inner membrane.

Bioinformatics was employed to identify important domains or features of P150 that could imply a function. There are no predicted β-helices in P150 but there is a predicted β-trefoil fold at the C-terminus, after the second transmembrane domain. Canonical TpsA substrates such as FHA have β-helical conformation (164), which are repeating structural segments like β-trefoil. Both are different forms of β-sandwich structures that consist of stacked β-sheets (144). P150 also lacks the "NPNL" and "NPNGI" motifs of TPS domain that are necessary and sufficient for recognition and secretion of ShlA (*S. marcescens*), HMW1 (*H. influenzae*) and FHA (*B. pertussis*) by their respective TpsB exporters (139). The possibility of Omp68 as an ancestral protein to TpsB proteins could therefore be applied to P150 as a potential precursor for TpsA proteins. Unlike TpsA proteins, P150 has a distinguishing and unique C-terminal DUF490 domain (Domain of unknown function 490). In the PFAM database, The DUF490 domain is distantly related to two other domains: AsmA and DUF748. While there is no functional information available for DUF478, there has been some characterisation of an AsmA domain protein. An AsmA motif is also predicted at the N-terminus (position 83-174) of P150. The AsmA protein (379 aa) has a predicted signal sequence that localises it at the inner membrane in *E. coli* (165). AsmA is upregulated when mutant OmpF is expressed and is thought to prevent the assembly of mutant OmpF (165, 166) or OmpC trimers (167). It is possible that P150 may have a similar role in regulating the assembly of Omp68, or other outer membrane proteins, in *Caulobacter*. It could act as a "sensor" of improperly assembled proteins in the outer membrane. The precise roles and interplay of Omp68 and P150 are yet to be determined. Hence, further research into Omp68 and P150 or its homologues will elucidate the important role of these conserved proteins across proteobacteria.

Chapter 3

The BAM complex in *Caulobacter crescentus*

3.1 INTRODUCTION

Outer membrane protein biogenesis is an essential process in Gram-negative bacteria (84). For proteins to be assembled and inserted into the outer membrane, they have to pass the inner membrane and periplasmic space. The Sec machinery mediates protein translocation across the inner membrane whereas periplasmic transport is facilitated by the periplasmic chaperones, Skp and SurA (2). A protein complex in the outer membrane, known as the BAM complex, captures and inserts the outer membrane proteins into the asymmetric lipid bilayer. The BAM complex consists of a central β-barrel subunit (BamA) and lipoprotein partners (84, 168). In *E. coli*, BamA is part of a complex consisting of the outer membrane lipoproteins BamB, BamC, BamD and BamE (137). Knockout studies of each of the components in the complex indicated that BamA and BamD are essential whereas BamB, BamC and BamE are not (136, 137). Aberrant phenotypes are observed for deleting the non-essential components of the BAM complex, which suggests their importance for maintaining the integrity of the cell envelope.

The functional homologue of BamA in mitochondria is Sam50 (75, 84, 85, 169), which forms a complex with two peripheral subunits, Sam35 and Sam37 (170). The modular nature of the SAM complex is crucial for its function, with "modules" such as Mdm10 and Mim1 docking on to assist specific stages of protein assembly (171-176). Modularity can provide a functionally dynamic protein complex to undertake various roles for the recruitment, folding, assembly and insertion of various types of outer membrane proteins. There are no bacterial homologues of these additional subunits as there are no lipoprotein components of the BAM complex present or associated with mitochondrial Sam50 (16).

It has been demonstrated that the SAM complex can assemble both β-barrel and α-helical membrane proteins, and receives its substrates via small TIM chaperones. These chaperones connect the protein translocation machinery (the TOM complex) with the SAM complex (177, 178). Although many aspects of mitochondrial function have evolved from its bacterial ancestor (179-182), there are differences in β-barrel protein targeting in bacteria and mitochondria.

These differences are also present between different species of bacteria. A bioinformatic approach, using hidden Markov Models, shows that α-proteobacteria have genes that could encode most, but not all, of the components found in the BAM complex of *E. coli* (16). While a BamB gene is found in several species of α-proteobacteria, no such sequence was found in *Brucella* (16). A sequence homologue of BamB is also absent from *Neisseria* (16). In particular, no sequence encoding a homologue of BamC was found in any species of α-proteobacterium suggesting that BamC is functionally redundant or that an alternative partner protein functionally complements BamC in α-proteobacteria. We characterised the BAM complex from *C. crescentus* by native gel electrophoresis and immunoprecipitation. After immunoprecipitation, mass spectrometry revealed that in addition to the BamB, BamD and BamE subunits, three novel proteins are present in the α-proteobacterial BAM complex: Pal, BamF and BamG. Pal has a characteristic OmpA-like domain for binding to the peptidoglycan layer of the bacterial cell wall. We demonstrate this anchorage of Pal and suggest it assists the BAM complex in accessing substrate proteins transported through the bacterial cytoplasmic membrane. Also, BamF seems to replace BamC in α-proteobacteria since it has a short conserved sequence at its N-terminus similar to that found in the BamC subunit from β- and γ-proteobacteria. Such a motif may represent the docking sequence to bind the BAM complex.

3.2 MATERIALS and METHODS

3.2.1 Strains and growth

C. crescentus strains were grown in PYE complex media at 30°C. When appropriate, media were supplemented with antibiotics at the following concentrations (liquid/solid media for *C. crescentus*; liquid/solid media for *E. coli*; in µg/ml): spectinomycin (25/50; 50/100), kanamycin (5/25; 30/50) and nalidixic acid (20/20; 0/0).

The construction of the Pal-mCherry-producing strain (CJW2965) is described (183). The Pal depletion strain (CJW3131) was obtained by single crossover of the suicide vector pXCC3229. pXCC3229 was obtained by cloning an internal fragment of Pal-encoding gene (CC3229) into pXMCS2. This internal fragment was obtained by PCR using CC3229*Nde*IFor: 5'-CACATATGAGCTTCGACACCCAGCGC-3' and CC3229*Hin*dIIIRev primers: 5'-CAGGATCCAGGAAGTCGCGCACGGCGTTG-3'.

The *bamF* and *bamG* knockout strains of *C. crescentus* were constructed as follows (see Appendix A2). Regions upstream and downstream of *bamG* were amplified by PCR with the following primers: UpstreamFor*Hin*dIII: 5'- ACGCAAGCTTGGGCGCATGAA-3', UpstreamRev*Eco*RI: 5'-GCACGAATTCGGGTCATGCTG-3', DownstreamFor*Eco*RI: 5'-TATCGAATTCACGCGCGCGAG-3', and DownstreamRev*Nhe*I: 5'-GCTAGCTAGCGATCGGTTGTC-3'. These fragments were ligated into pNPTS138 to form pNPBamG. A Ω cassette carrying spectinomycin resistance was PCR amplified from pBOR using primers: SpectinomycinFor: 5'-CGGCCTGCAGAGTGGATCCCCCGGGCTGCA-3' and SpectinomycinRev: 5'-CGGCGCTAGCGGTATCGATAAGCTTGATAT-3'. The PCR product was digested with *Eco*RI and ligated into pNPBamG. Genomic deletion of *bamG* was further checked by PCR using primers BamGFor*Eco*RI: 5'-TGCGGAATTCATGTGCGAAGATCCG-3', BamGRev*Xho*I 5'- ATATCTCGAGGTCGCCGGCGAA-3' to amplify *bamG* or primers BamGFor*Eco*RI and SpecInternalRev: 5'-GCCAACTTTGTTTTAGGGCGACTGCC-3' to amplify the 5' of *bamG* and an internal sequence of the spectinomycin-resistance

cassette. Serial dilution of wild-type and *bamG* knockout mutant involved growing cells to an OD_{600} of 0.3 and spotting 2 µL of culture on PYE plate. Sterile dilutions (10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}, 10^{-6}) of the culture were performed, from which 2 µL was spotted on plates. Plates were dried and incubated at 30ºC for 3 days.

The *bamF* knockout mutant was generated by single homologous recombination using the suicide vector pBgent (gentamycin^R) with an internal fragment of *bamF*. The internal region of *bamF* was amplified with the following primers: BamFKOFor*Xba*I 5'-GCCGTCTAGAAGTTTCAACCGGGT-3' and BamFKORev*Hin*dIII 5'-ATCGAAGCTTAGGGTCTTCAGGCG-3'. *bamF* deletion mutants were checked for the presence of the BamF protein using the BamF anti-serum.

For heterologous expression of BamF and hence production of BamF antigen for antibody production, BL21 (DE3) cells were transformed to express hexahistidine-tagged BamF without the signal sequence and the N-terminal cysteine residue. The BamF expression construct was generated by amplifying *bamF* using primers BamFpet*Nde*IFor: 5'-GGTACGCATATGCGTATCGAGAGCGTC-3' and BamFpet*Hin*dIIIRev: 5'-ATATAAGTCTCAGGCCCGGGAG-3'. The PCR fragment was cloned into the pET22b vector for expression of BamF inclusion bodies in *E. coli*, protein purification and rabbit immunisation using conditions outlined in *Section 2.2.3* and *2.2.4* for Omp68.

3.2.2 Homology model

A structure-based sequence alignment was performed in Swiss-PdbViewer (http://www.expasy.org/spdbv/) (184) using sequences and coordinates for *E. coli* Pal (PDB code 2K1S), *H. influenzae* Pal (PDB code 2AIZ), *N. meningitidis* RmpM (PDB code 1R1M) and *E. coli* YiaD (PDB code 2K1S, unpublished), and aligning these to the *C. crescentus* Pal sequence. The results were visualized in JalView (185).

3.2.3 Subcellular fractionation

Inner and outer membranes were purified from *C. crescentus* cultures grown in PYE media. See *Section 2.2.7* for further details.

3.2.4 Immunological methods and electrophoresis

The method of generating antibodies recognising BamA is discussed in *Section 2.2.2-2.2.4*. To generate antibodies recognising BamF, the protein was expressed and purified from inclusion bodies according to *Section 2.2.3* for the purification of Omp68.

The method for BN-PAGE and immunoblotting is described in *Section 2.2.5* whereas *Section 2.2.9* describes the method of co-immunoprecipitation that was used to immunoprecipitate proteins with the BamF antibody.

3.2.5 Microscopy of Pal

Microscopy images were obtained using NIKON E80i equipped with a Hamamatsu ORCA ERII or Andor iXon^{EM}+ CCD camera. Cells were grown to an OD_{660} between 0.2 and 0.3 units/mL, washed in minimal (M2G) media and immobilized on an agarose-padded slide. When required, chloramphenicol was added to the culture 30 min before observation and to the agarose pad at a final concentration of 20 μg/ml. Images were taken and processed by Metamorph 7.1.4 software. For photobleaching experiments, a Photonic Instrument Micropoint Laser system was used with a 552-nm Laser Dye. Kymographs were constructed using Metamorph 7.1.4 software.

3.2.6 Mass spectrometry

For analysis of proteins immunoprecipitated with the BamF antibody, gel bands corresponding to the proteins of interest were excised and submitted to the Proteomics Facility at Monash University for electrospray tandem MS.

3.2.7 Sequence motif analysis

HHsenser searches (186) were run to generate two families (BamC/NlpB and BamF) (defaults, nr+enviromental, 1000 sequence cutoff). To remove any bias from targeting motifs common to lipoproteins, signal sequences and the lipoylated cysteine residue were removed from all sequences. The HHsenser searches returned 176 sequences belonging to the BamC/NlpB family and 67 sequences belonging to the BamF family. Each sequence set was aligned with MUSCLE (187), signal sequences removed, as were the few sequences that appeared incomplete and sequences labeled "[marine metagenome]" or uncharacterised/uncultured organisms. All gaps were removed from the alignment. The final cleaned BamC set contained 90 sequences, and final cleaned BamF set contained 51 sequences. To determine common sequence motifs, MEME (188) was run on the union of the BamC and BamF sets using the *oops* ("One Occurrence Per Sequence") and *zoops* ("Zero or One Occurrence Per Sequence") models, with command line options "*-maxsize 1000000 -maxw 80 -nmotifs 5*" and "*-mod oops*" or "*-mod zoops*", for the oops and zoops models respectively. To determine functionally and structurally important conserved residues, 39 BamF sequences with a expect value cut-off set to 0.001 were selected and analysed with the ConSeq server (189). ConSeq uses Bayesian methods to score the rate of evolution at each aligned position

3.2.8 Protease accessibility

A strain of *C. crescentus* expressing Pal-mCherry was grown and harvested at mid-log growth phase and washed twice with 50 mM Tris-HCl, pH 8.0. Cells were either incubated with polymyxin B (final concentration 2 mg/mL) or 50 mM Tris-HCl, pH 8.0 (control) for 10 min. Cells were digested with trypsin (final concentration 0.1 μg/μL) for 30 min at 4°C. Reaction was stopped by adding soybean trypsin inhibitor (STBI) to a final concentration of 0.5 μg/μL. Prior to analysis by SDS-PAGE, 2 × SDS sample buffer was added and samples were heated at 95°C for 5 min.

3.3 RESULTS

3.3.1 Pal is a ubiquitous lipoprotein in Gram-negative bacteria

The BAM complex in *C. crescentus* has the predicted subunit composition (BamA, BamB, BamD and BamE) as well as three other lipoproteins: Pal, BamF and possibly BamG (Figure 3.1A). The molecular interactions of each subunit within the complex are unclear. Since Pal is a peptidoglycan-binding lipoprotein, its interactions with the BAM complex and peptidoglycan potentially creates proximity between the inner and outer membrane and hence, spanning of BamA across the periplasmic space.

Structures of four closely-related Pal homologues have been solved: Pal from *E. coli* (190), Pal from *H. influenzae* (191), RmpM from *N. meningitidis* (192) and YiaD from *E. coli*. Although Pal is ubiquitous in Gram-negative bacteria, the protein sequence is somewhat poorly conserved (193). The sequence identity shared by *C. crescentus* Pal with *E. coli* and *N. meningitidis* Pal lipoprotein is 32% and 28%, respectively. However, there are regions of conserved residues across species that form the peptidoglycan-binding region (191). The *C. crescentus* Pal is sufficiently similar to these proteins that a structural model could be built with confidence (Figure 3.1B).

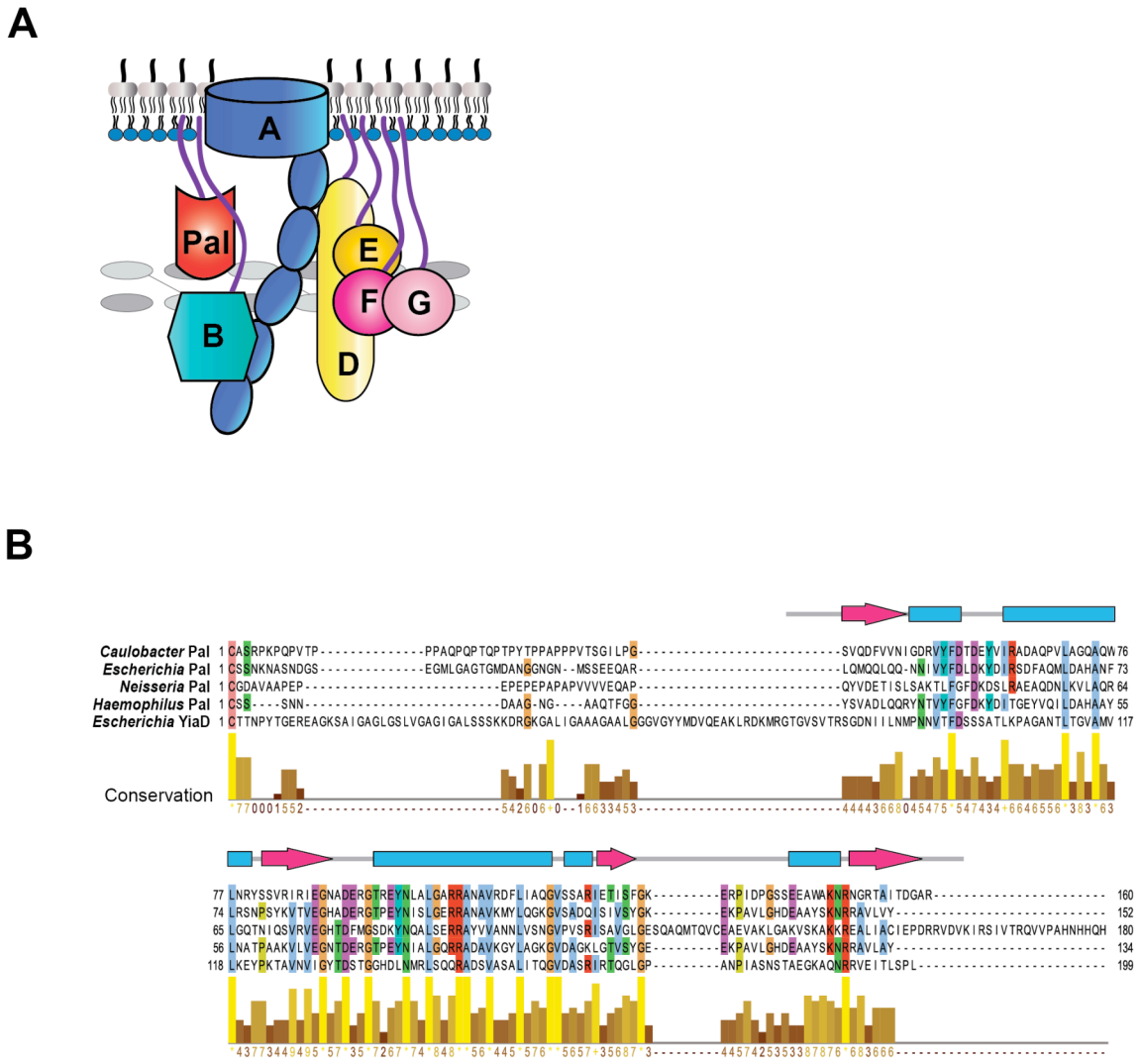


Figure 3.1: The BAM complex in *C. crescentus* and the well-characterised Pal subunit. (A) A model for the BAM complex in *C. crescentus*. No homolog of BamC is encoded and there are three novel lipoproteins identified in this complex (Pal, BamF and BamG). (B) A structural alignment of *C. crescentus* Pal and homologous Pal sequences for which the structure has been solved. Residues that show at least 50% conservation are color coded according to their properties. Above the sequence alignment is the secondary structure for Pal with bars (light blue) and arrows (pink) representing α-helices and β-sheets, respectively. Below the sequence alignment is the conservation plot, which is based on a 1-10 scale with completely conserved residues highlighted with yellow bars and asterisks (*). The numbers at the start and the end of the sequences indicate the first and last amino acid positions of the alignment.

3.3.2 Pal is an outer membrane lipoprotein that associates with the BAM complex

As a subunit of the BAM complex, Pal would be expected in the outer membrane. To test for its subcellular location, we used a strain expressing a Pal-mCherry fusion protein (183)**.** The viability of the strain demonstrates that the Pal-mCherry fusion is functional. Sucrose density gradient fractionation of membrane vesicles revealed that Pal is associated with outer membranes (Figure 3.2A).

To determine whether Pal is bound to outer membrane proteins other than the BAM complex, outer membrane vesicles were prepared from the Pal-mCherry strain and subjected to BN-PAGE (Figure 3.2B). Pal is found in complexes of 70-120 kDa, which do not correspond to complexes containing BamA. In addition, immunoblots using anti-BamA and anti-Pal sera show Pal associated with the ~500 and ~300 kDa form of the BAM complex, but not the 150-170 kDa forms of the BAM complex. Consistent with immunoprecipitation data, at higher detergent concentrations, BN-PAGE shows Pal is dissociated from the core BAM complex (Figure 3.2B).

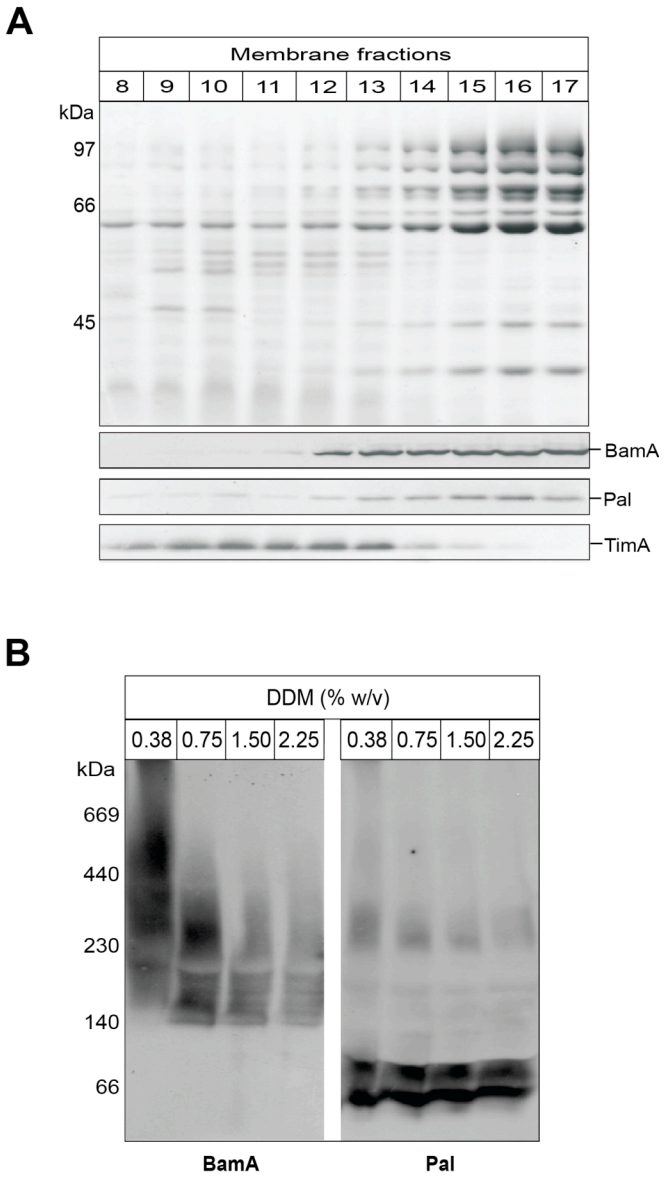


Figure 3.2: Pal is an essential outer membrane protein that is associated with the BAM complex. (A) Membranes from *C. crescentus* that expresses the mCherry-Pal fusion protein were fractionated on a sucrose gradient and analysed by SDS-PAGE. Coomassie blue staining (upper panel) reveals separation of the membrane protein profiles that were immunoprobed for BamA, the inner membrane protein TimA and the mCherry epitope to determine the location of Pal. (B) Outer membrane vesicles (100 µg protein per lane) isolated from *C. crescentus* that expresses the mCherry-Pal fusion protein were solubilised with the indicated concentrations of DDM, separated by BN-PAGE and analysed by immuno-staining with an anti-serum to BamA and mCherry. Arrows indicate three modular forms of the BAM complex.

3.3.3 Pal is essential lipoprotein that interacts with the peptidoglycan layer

Epifluorescence microscopy of *C. crescentus* expressing Pal-mCherry shows a halo of fluorescence consistent with the mCherry fluorescent tag being exposed to the periplasm (183, 194). It was not possible to disrupt the *pal* gene by double homologous recombination without providing the gene on a plasmid, which suggests that it is an essential gene in *C. crescentus* (data not shown). Double recombination was attempted according to the method outlined in Appendix A2. We therefore constructed a strain in which the *pal* gene was under the control of a xylose-inducible promoter. The resultant strain required xylose for growth, but cultures of this strain acquired suppressors very easily. Microscopic examination of a freshly obtained Pal depletion strain growing in the absence of xylose (↓pal) often displayed vesicles and blebs that mainly formed from the poles or division site (Figure 3.3A).

To determine whether the Pal-mCherry fusion is immobilised through an interaction with peptidoglycan, we acquired time-lapse sequences of Pal-mCherry fluorescence before and after photobleaching a small region within cells treated with chloramphenicol to block protein synthesis. Kymographs were constructed from these sequences to display the integrated fluorescence intensity along the long cell axis as a function of time (194). These kymographs showed that no fluorescence recovery was observed after 2 h in all cells photobleached (n=17; Figure 3.3B), indicating that Pal-mCherry does not move laterally in the outer membrane. By contrast, it has been shown that the recovery time of integral membrane proteins such as PBP3 is in the order of tens of seconds (194), and for soluble proteins such as cytoplasmic GFP, the recovery is immediate (data not shown). As further evidence to the relatively static location of Pal, cells expressing Pal-mCherry were completely bleached, and then allowed to grow and synthesise new Pal-mCherry (Figure 3.3C). Time-lapse imaging showed that newly synthesised Pal-mCherry appeared around the cell body during growth (Figure 3.3C, right panel). In contrast, no Pal-mCherry signal appeared in the stalk. This suggests that newly synthesised Pal is added predominantly where new peptidoglycan is incorporated; the stalk is made of old peptidoglycan, with stalk elongation proceeding from the cell body (195, 196).

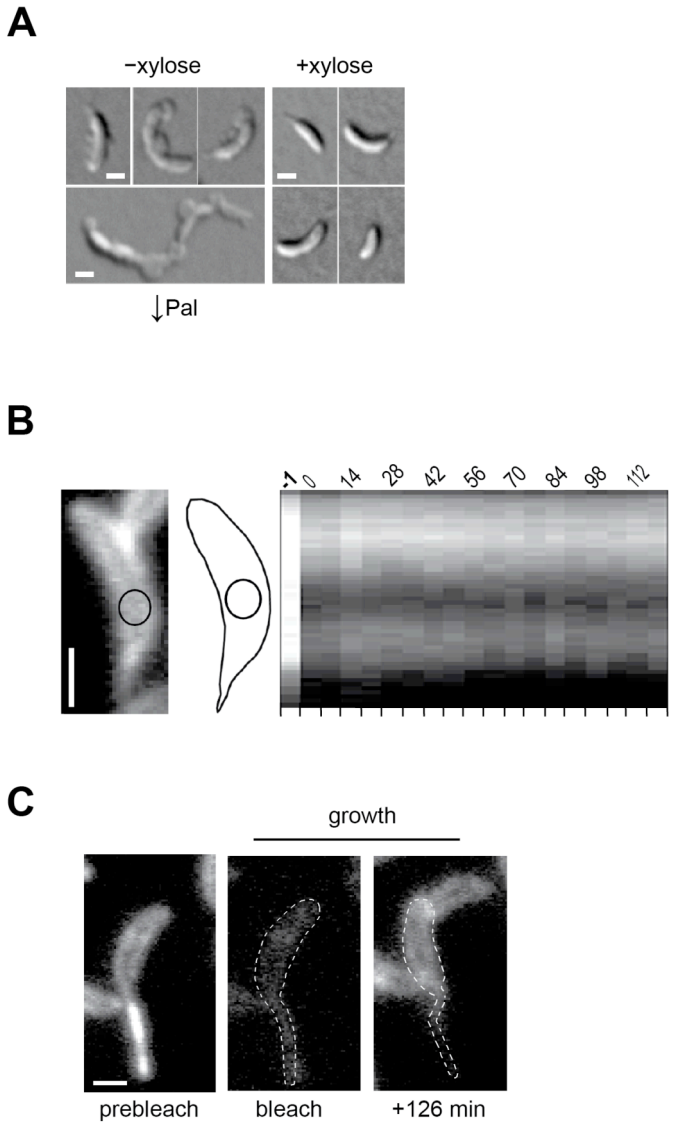


Figure 3.3: Pal is an essential protein that is anchored to the peptidoglycan layer. (A) Cells with the *pal* gene under the control of a xylose-inducible promoter (↓ *pal*) were grown in the presence (right montage) and absence (left montage) of xylose (0.3%) for 10 h. Outer membrane blebs that form predominantly from the division site or cell poles are evident only in the Pal-depleted cells. Scale bars (white) represent 1 micrometer. (B) Photobleaching analysis of Pal-mCherry dynamics required pre-treatment with chloramphenicol to block protein synthesis, and cells expressing Pal-mCherry were immobilised on an agarose-padded slide containing chloramphenicol. The left panel shows a representative cell in which the circle indicates the region that will be photobleached. The right panel shows a kymograph representation of Pal-mCherry fluorescence intensity along the cell axis before (-1 min) and following photobleaching. (C) Fluorescence micrographs showing a cell expressing Pal-mCherry before photobleaching (left), right after being completely photobleached (middle) and after 126 min of growth (right).

3.3.4 BamF is a lipoprotein with a conserved BamC motif

Bacterial lipoproteins have a lipid-modified cysteine at the N-terminus that anchors it to the inner and outer membrane (41). Multiple sequence alignment of BamF and BamC sequences shows the conserved cysteine residue for lipid modification (Figure 3.4A). The presence of threonine at the position next to the N-terminal cysteine (44, 45) suggests BamF has a signal for targeting to the outer membrane.

The lipoprotein BamF has homologues in all species of α-proteobacteria for which genomic sequences are available, but no homologues were detected in other bacterial lineages. However, BLAST searches identified a short region at the N-terminus of BamF with similarity to an equivalent region in BamC from the γ-proteobacterium, *Acinetobacter* albeit with a very low expect value (*e*-value 1.1). To determine whether a conserved motif might be present in this region of BamF proteins, MEME analysis (188) was undertaken allowing for "zero or one motif" in each of the 90 BamC and 51 BamF sequences. The full data for this analysis can be found in the Appendix section (A3). A BamF motif was detected and the same motif was found in some β-proteobacterial BamC homologues (Figure 3.4B). The γ-proteobacterial BamC proteins, such as that found in *E. coli*, contain a different but overlapping motif in the same N-terminal region (Figure 3.4B). After processing of the signal sequence and lipid-modification of the N-terminal cysteine residue, this conserved region of the protein would be the membrane proximal part of BamF and BamC. A 'control' experiment using MEME to find common motifs from within a set of 98 predicted *C. crescentus* outer membrane lipoproteins showed no common motifs, indicating that the motifs detected in BamC and BamF are not simply common signatures at the N-terminal regions of outer membrane lipoproteins.

To determine the functionally conserved regions of BamF, we generated a multiple sequence alignment with the ConSeq server (189), which uses Bayesian methods to evaluate the conservation or evolution of residues at each position. Figure 3.4C shows the BamF protein sequence is more conserved at the N-terminus relative to the C-terminus. Functionally significant residues that are highly conserved and exposed are concentrated

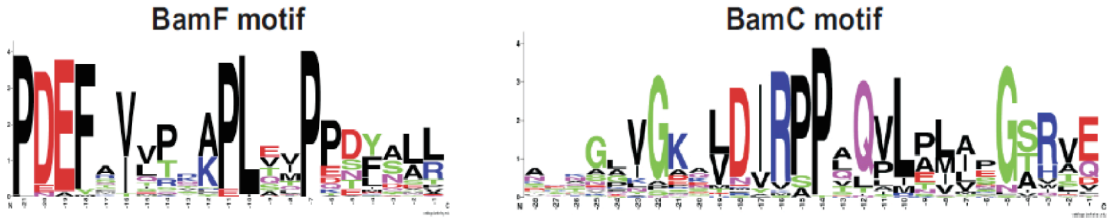
at the region containing the predicted motif. We suggest that this short, membrane proximal sequence motif facilitates interactions with the BAM complex.

Figure 3.4: BamF has a conserved motif that it shares with BamC homologues. (A) CLUSTAL-W alignment of the N-terminal sequences from BamF and its homologues from other α-proteobacteria. The alignment also includes the N-terminal sequences from BamC proteins from *E. coli* and *N. meningitidis*. (B) Sequence motifs were detected using MEME and are plotted here as sequence logos. The related motifs for BamC and BamF are presented separately. (C) Sequence conservation of BamF homologues. Sequence alignment of BamF was analysed with the ConSeq server (189). Conservation scores are normalised so that the average score is zero and the standard deviation is 1. Rapidly evolving residues at particular position are negative (coloured cyan) and slowly evolving residues are positive (coloured purple) while average rate of evolution are coloured white. Yellow residues represent insufficient data i.e. the calculation for this site was performed on less than 10% of the sequences. Letters under the protein sequence are explained as follows: **e**-an exposed residue according to the neural-network algorithm, **b**-a buried residue according to the neural-network algorithm, **f**-a predicted functional residue (high conserved and exposed) and **s**-a predicted structural residue (high conserved and buried).

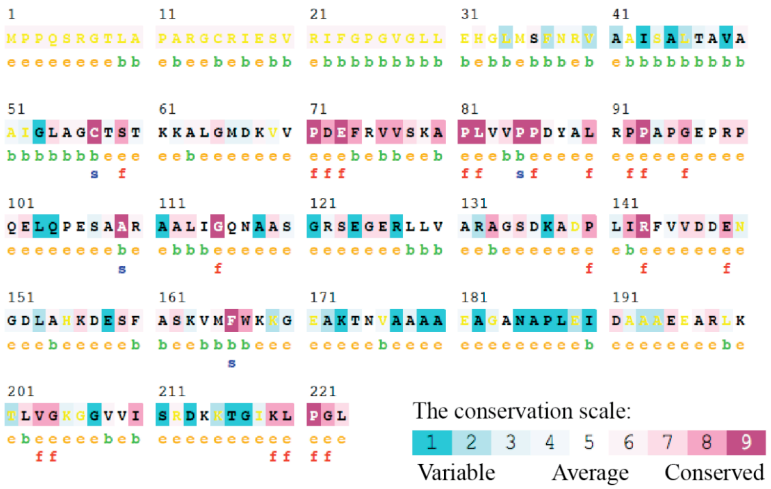
A


signal sequence
conserved segment
Caulobacter MSFNRVAAISALTAVAAIGLAGCTSTK---KALGMDKVVPDEFRVVSKAPLVVPPDYALRPPAPGEPRP 66
Escherichia MAYSVQKSRLAKVAGVSLVLLAACSSDSR-YKRQVSGDEAYLE--AAPLAELHAPAGMILPVTSGDYAIP 68
Neisseria MTHIKPVIAALALIGIAACSGSK-----TEQPKLDYQS-RSHRLIKLEVPPDLNNPDQGNLYRLP 59
Nitrosomonas MLKKRPSRRLRKFLIANITLAILVSGCNLL------PENKKIDYKS--AGKLPPLEVPPDLTSPETNERFAIP 65
Polaromonas MKNLPKRPAFATAQTALTISALVAVGLLAGCSTLR---DVMEGERIDYKS-SATKAPSLDIPPDLTQLNRESRYVVP 73
Ralstonia MKLHQTRRGAVATSAQLLLAALSVALISGCDTVN---EVMQPDRIDYKS-QAKKGPTLEVPPDLTGIQADRRYAAP 72
Burkholderia MKHSAFSSRAIQVSVLALALAALAGCDTLN---DYLAPDRVNYKS-TGS-APPLQVPQDLTAMPLSPSYVAP 67
Acinetobacter MMQLRLGFTFAIASIALAGCSS-------LSLNNHSLDYKKASTLAPLELPADTTMRPFTPLYPAP 59
Psychrobacter MHQTQLSKTSPKPVNAALSTAIQLLIVGASSLAILSGCQATKGVFGKVDDGSLAYQK--AEKLDPIQLPADLEAGPFVPLYATP 82

B


BamF motif
BamC motif

C


The conservation scale:
1 2 3 4 5 6 7 8 9
Variable
Average
Conserved

3.3.5 BamF is an outer membrane lipoprotein of the BAM complex

BamF is not an essential lipoprotein as the gene that encodes BamF can be successfully deleted from the genome (Figure 3.5A and 3.5B). To determine the localisation of BamF, sucrose gradient fractionation was used to separate inner and outer membranes. When we probe for BamF, it is found in the same fractions containing BamA (Figure 3.5C); hence BamF is an outer membrane lipoprotein. Protease accessibility experiment also showed that like the periplasmic Pal-mCherry protein, BamF is a periplasmic protein anchored to the outer membrane (Figure 3.5D).

Analysis of outer membranes using BN-PAGE and subsequent probing for BamF shows it is mainly found in a ~ 500 kDa complex that is reduced to 300 kDa and 20-40 kDa complexes with increasing DDM concentration (Figure 3.5E). The protein complex of BamA behaves similarly except BamA is not found in the 20-40 kDa complexes but in a core 150 kDa complex at high detergent concentration (see *Section 2.3.4*). To determine the interacting partners of BamF, immunoprecipitation was performed on DDM-solubilised outer membranes. The known proteins of the BAM complex are co-purified with BamF as judged by the molecular weight of the immunoprecipitated proteins (Figure 3.5F). Mass spectrometry was performed on these protein bands but lack of sensitivity from the mass spectrometry instrument revealed only three of the predicted subunits of the BAM complex (Table 3.1). To evaluate whether BamF is strongly attached to the BAM complex, low and high DDM concentrations were used to solubilise outer membranes prior to immunoprecipitation with the BamA anti-serum. Compared to BamD, BamF is not tightly associated to the BAM complex at high DDM concentration (Figure 3.5G).

Figure 3.5: BamF is outer membrane lipoprotein that associates with the BAM complex. (A) Schematic representation of the genomic context of *bamF* in wild-type and *bamF* mutant of *C. crescentus* (not drawn to scale). *bamF* was deleted by a single recombination event using the pBgent suicide vector that contained the internal sequence of BamF for recombination to occur. (B) Wild-type and deletion mutant of *bamF* was probed for BamF using BamF anti-serum. BamA and BamD were probed with appropriate anti-serum as controls. (C) Subcellular localisation of BamF and BamA by sucrose gradient fractionations of membranes from *Caulobacter* and Western blotting using BamF and BamA antisera. (D) Protease shaving of *Caulobacter* cells. Cells were incubated with (Lane 2, 3) and without (Lane 1, 4) polymyxin B followed by addition of trypsin (Lane 3, 4). Samples were analysed by SDS-PAGE and immunoblotting. Blots were incubated with antibodies against BamF, Pal-mCherry and TimA (inner membrane protein). (E) Outer membrane protein complexes (5 µg) were separated by BN-PAGE using different solubilising concentrations of DDM. Samples were probed using anti-serum against BamF. The migration positions of the molecular weight markers are shown. (F) Outer membrane vesicles (1000 µg protein) were solubilised with 0.75% (w/v) DDM and the BamF complex immunoprecipitated with an anti-serum recognising the BamF subunit. Samples of total outer membranes (Lane 1), immunoprecipitate using the BamF (Lane 2) or BamA antibody (Lane 4) are shown. Pre-immune serum derived from rabbits that produced the BamF or BamA antibody was used as a control for immunoprecipation (Lane 3 and Lane 5). Asterisks indicate the IgG heavy and light chains. (G) BamA antisera is used to immunoprecipitate the BAM complex from outer membranes that were solubilised with 0.75% (Lane 1) and 2% DDM (Lane 2). Pre-immune serum was used as a control (Lane 3) to immunoprecipitate outer membranes solubilised with 0.75 % DDM. Samples were probed with BamA, BamD and BamF antisera.

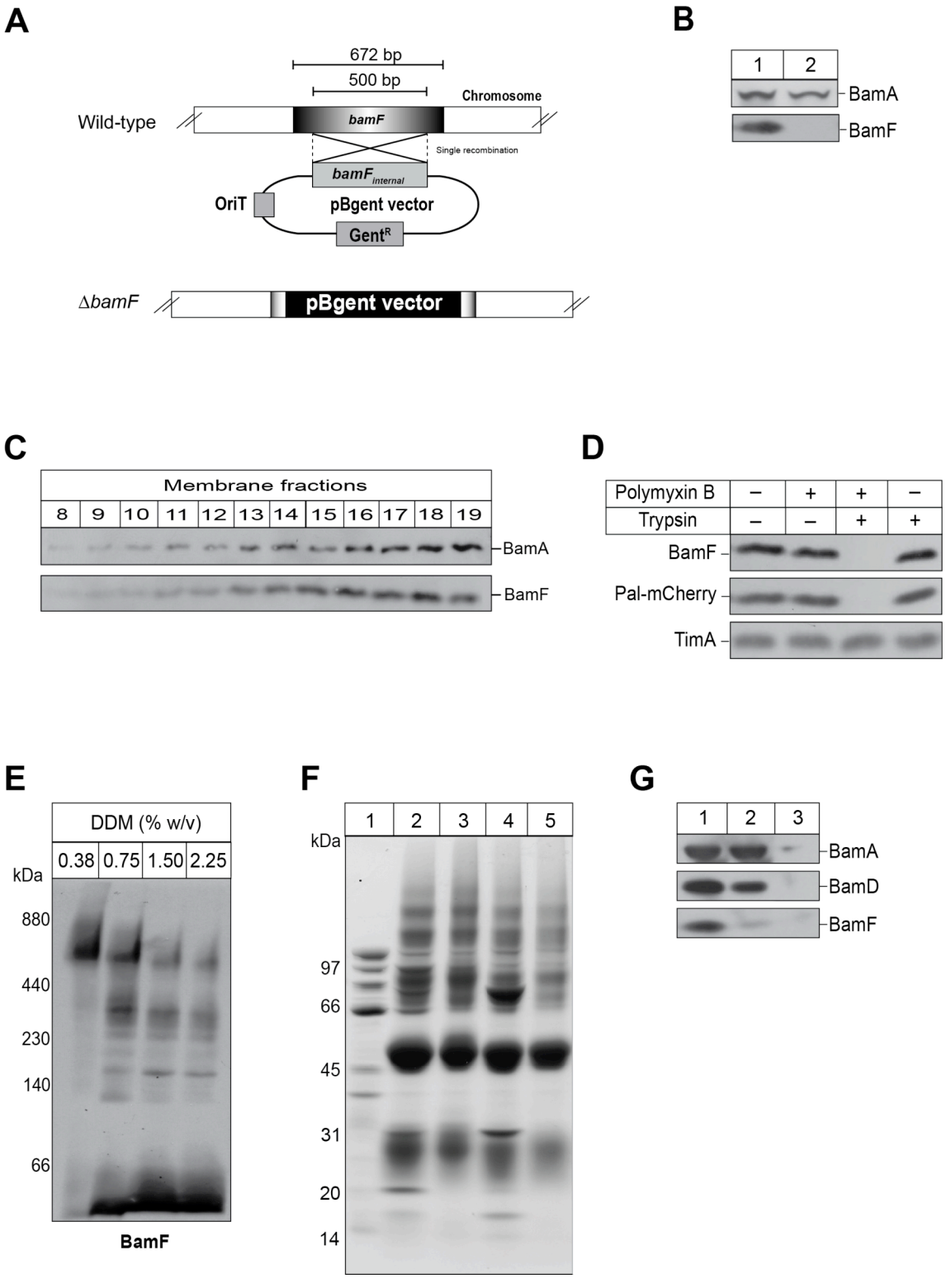

A
672 bp
500 bp
Chromosome
Wild-type
bamF
Single recombination
bamF internal
OriT
pBgent vector
Gent R
ΔbamF
pBgent vector
B
1
2
BamA
BamF
C
Membrane fractions
8
9
10
11
12
13
14
15
16
17
18
19
BamA
BamF
D
Polymyxin B
Trypsin
BamF
Pal-mCherry
TimA
E
DDM (% w/v)
0.38
0.75
1.50
2.25
kDa
880
440
230
140
66
BamF
F
1
2
3
4
5
kDa
97
66
45
31
20
14
G
1
2
3
BamA
BamD
BamF

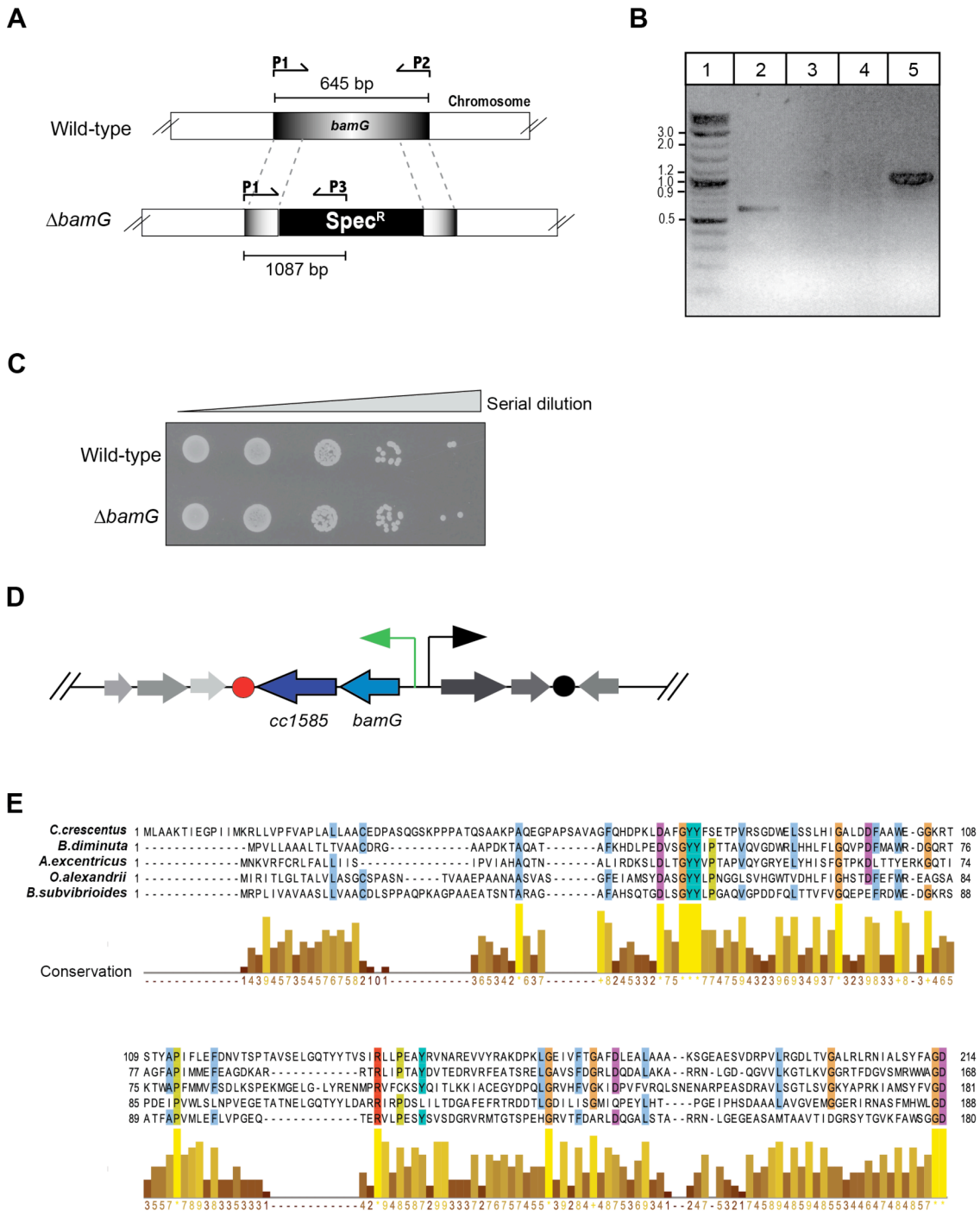
Table 3.1: Mass spectrometry data summarising the identification of immunoprecipitated proteins using BamF antisera

Protein ID	SwissProt	Predicted size (kDa)	MOWSE score	Peptides matched
BamA	Q9A711	86,000	410	26
BamD	Q9A6U9	34,247	1275	34
BamF	Q9AAA7	23,277	84	2

3.3.6 BamG is a non-essential protein in C. crescentus

A knockout plasmid was designed that contained the upstream and downstream flanking regions of *bamG* and a spectinomycin-resistance cassette (Figure 3.6A) that is intended to replace the internal sequence of *bamG*. All *bamG* deletion candidates were screened for the desired genomic structure by PCR using oligonucleotides as described in the Methods section. Compared to wild-type cells, *bamG* cannot be amplified by PCR from the mutant strain (Figure 3.6B, Lane 2 and 3). To verify that the cassette disrupts the gene, a primer based on the internal sequence of the cassette was designed and used for PCR with a forward primer based on the initial sequence of *bamG*. The expected size of the fragment is 1087 bp and is amplified only from the mutant strain (Figure 3.6B, Lane 4 and 5). Hence, *bamG* is non-essential in *C. crescentus* as it can be successfully disrupted using this method. In addition, loss of BamG does not affect viability or result in any growth defect (Figure 3.6C). Analysis of the genomic context of *bamG* reveals it is in an operon with *cc1585*, which is also an uncharacterised protein (Figure 3.6D). Bioinformatic analysis of the BamG protein sequence shows it has a signal sequence and the conserved cysteine residue that is predicted to undergo modification with a lipid anchor. Database searches identified four other sequence homologues of BamG, which are all present in α-proteobacteria (Figure 3.6E).

Figure 3.6: BamG is an unconserved lipoprotein that is not essential in *C. crescentus*. (A) Schematic representation of the genomic context of *bamG* in wild-type and *bamG* mutant of *C. crescentus* (not drawn to scale). *bamG* is 645 bp in length and PCR-amplified with primers P1 and P2. Disruption of the gene with a spectinomycin cassette in a mutant strain can be detected by PCR using primers P1 and P3 to amplify a 1087 bp fragment. (B) PCR screening of wild-type and *bamG* mutant showing 2-Log DNA ladder (NEB, Lane 1), fragments amplified using primers P1 and P2 from wild-type (Lane 2) and mutant (Lane 3), fragments amplified using primers P1 and P3 from wild-type (Lane 4) and mutant (Lane 5). (C) Serial dilution of wild-type and mutant *bamG* cultures grown on PYE at 30°C. (D) Genetic organisation of the *C. crescentus bamG/cc1585* operon. Thick arrows represent chromosomal genes, thin arrows represent transcriptional initiators and circles indicate transcriptional terminators. (E) A CLUSTAL-W sequence alignment of BamG and its homologues. Below the sequence alignment is the conservation plot, which is based on a 1-10 scale with completely conserved residues highlighted with yellow bars and asterisks (*).


A
P1
P2
645 bp
Chromosome
Wild-type
bamG
P1
P3
ΔbamG
Spec^R
1087 bp
B
1
2
3
4
5
3.0
2.0
1.2
1.0
0.9
0.5
C
Serial dilution
Wild-type
ΔbamG
D
cc1585
bamG
E
C.crescentus
B.diminuta
A.excentricus
O.alexandrii
B.subvibrioides
Conservation

3.4 DISCUSSION

In *E. coli*, the lipoprotein Pal is attached to the outer membrane via its N-terminal acylated cysteine residue and interacts with the thin peptidoglycan layer using its conserved OmpA-like domain. This domain can be found in many proteins including the outer membrane protein OmpA, and the inner membrane flagellar motor protein, MotB (197). Mutations in Pal lead to increased cell permeability and the release of periplasmic proteins from the cell. Other phenotypes relate to aberrant cell division and motility as well the release of large amounts of outer membrane vesicles (198-201). Since the role of Pal is related to maintaining the integrity of the cell envelope, pathogenic bacteria also exhibit reduced virulence in mice when Pal is deficient or mutated (202). Similarly, deletion of *bamB* was shown to attenuate some pathogenic bacteria (203) whereas cell permeability defects have been demonstrated with *bamB*, *bamC* (204) and *bamE* (136) knockout strains.

Pal has previously been under scientific scrutiny for its role in the Tol-Pal system, which is a complex that spans the periplasm. The functional basis for this trans-envelope bridge is unknown (197) but it has been strongly linked to imparting cell envelope integrity. In addition to its role in the Tol-Pal bridge, Pal is shown to interact with outer membrane proteins such as OmpA and Lpp (205). In *N. meningitidis*, the Pal homologue is known as RmpM and it forms a complex with several proteins: the two major porins, PorA and PorB (206), and with the TonB-dependent transporters TbpA (transferrin-binding protein A) and LbpA (lactoferrin-binding protein A (207). RmpM was shown to stabilise the complexes formed by these proteins. Recently, it was shown that RmpM associates with the BAM complex in *N. meningiditis* (208). It seems Pal is a general lipoprotein that can interact with various proteins to stabilise their complexes as well as associate with the peptidoglycan. Hence, Pal might provide the BAM hetero-oligomeric complex with stability and anchorage to the peptidoglycan layer in Gram-negative bacteria. Although Pal has not been identified in the BAM complex of *E. coli*, a closely related protein, YiaD, was shown to be a multi-copy suppressor of *bamD* temperature-sensitive mutants ($bamD^{ts}$) (209). Unlike Pal, YiaD is not essential for cell growth but mutations of YiaD

for the conserved residues of the OmpA domain failed to correct the growth of *bamD*^{ts} mutants, indicating that the peptidoglycan-binding region is important for the suppression activity of YiaD. Due to this functional relationship between Pal and BamD, it seems likely that YiaD, not Pal, may be associated with the BAM complex in *E. coli* and that linkage to the peptidoglycan layer becomes essential if the function of the BAM complex is partially compromised as in these *bamD*^{ts} mutants.

While the homologue of Pal (RmpM) is not related to the assembly of some outer membrane proteins in *N. meningitidis* (208), it does have physiological importance for outer membrane protein biogenesis. When outer membrane proteins emerge from the Sec machinery, it is confronted with a meshwork of peptidoglycan in the periplasm. Although the presence of holes in the peptidoglycan has been reported to accommodate globular proteins with a maximal size of 50 kDa (128), penetration of the peptidoglycan gel could be necessary for protein transport and may occur with the aid of a periplasmic bridge connecting the outer and inner membrane (127). Pal may assist in creating these connections by tethering the BAM complex to the peptidoglycan layer. The POTRA domains of BamA can span the periplasmic distance at these connections and create points devoid of peptidoglycan for binding to substrates as they emerge from the Sec machinery. In support of this idea, recent results from NMR (71) and X-ray crystallography (73) demonstrate an extended conformation adopted by the POTRA domains of BamA. Such a conformation was compared to the periplasmic domain of TolC (73), which spans the periplasm and reaches the outer leaflet of the inner membrane (210). Similarly the POTRA domains of BamA can penetrate the peptidoglycan layer and be in close proximity to the inner membrane. The extended POTRA arm could extend as much as 10 nm into the periplasm (73, 211) to capture substrates as they emerge from the Sec machinery. Upon capture, the POTRA arm may flex to reflect the hook-like conformation observed by Kim *et al.* (72). Tethering of the BAM complex to peptidoglycan and cycling of extended and bent conformations of the POTRA arm of BamA may increase the efficiency of protein assembly.

In *C. crescentus*, we observe that BamABDEFG-Pal form a complex at low detergent concentration. However, at higher detergent concentration, Pal dissociates from the complex. This biochemical behaviour of Pal intuitively correlates with the understanding that Pal associates with many different outer membrane proteins, which would require weak interactions to permit promiscuity.

One of the most exciting outcomes of the present analysis is the identification of another lipoprotein referred to as BamF. Its homologues are found in many α-proteobacteria and may represent a functional homologue of BamC. Previous work using sensitive and powerful hidden Markov models (16) could not identify any protein encoding a homologue of BamC in any species of α-proteobacterium. Biochemical analysis of BamF demonstrated the modularity and composition of the BamF protein complex, which validates previous findings that BamA interacts with BamF. It is unclear how BamF is linked to the BAM complex but preliminary *in silico* analysis has identified a short region that is shared by all BamF and many BamC homologues. This region could provide clues about how BamF or BamC docks onto the BAM complex. In line with this observation, we discover that the affinity of BamF for BamA is very weak as most of the protein can be removed from BamA at high detergent concentration. Weak interactions with BamA may be facilitated by the short BamF motif.

Another lipoprotein that was identified in the BAM complex of *C. crescentus* is BamG. No biochemical analysis has been undertaken except that it is a non-essential protein as it can be deleted from the genome without affecting cell growth. Unlike BamF, BamG is not conserved with homologues found in only four other types of aquatic α-proteobacteria. Perhaps BamG is involved in the efficient assembly of a subset of proteins that enrich the outer membrane of these species living in dilute environments with limited availability of nutrients. Further work is required to determine the existence and role of BamG in the BAM complex.

This study provides a platform from which to further characterise the BAM complex in *C. crescentus* for understanding the similarities and differences of the β-barrel assembly

pathway in different classes of proteobacteria. Integrating our experimental data will pave the way for future experiments related to designing mutants based on the BamF motif that was identified. Although the structure of BamC is yet to be divulged, crystals have already been obtained (212). The structure of BamF will provide information about the fold of this protein for comparison with the structure of BamC.

Chapter 4

Characterisation of the TPR-rich lipoprotein subunit of the BAM complex: BamD from *C. crescentus*

4.1 INTRODUCTION

Bacterial lipoproteins are proteins covalently linked to lipids, anchoring the protein to the cytoplasmic or outer membrane (9, 41). The lipid moeities are generally attached to a cysteine located at the amino-terminus of the lipoprotein, serving to position the protein at the membrane-aqueous interface for localisation and function at the membrane surface. One group of outer membrane lipoproteins that recently generated much interest are the lipoproteins associated with the BAM (β-barrel assembly machinery) complex (74). In *E. coli*, the BAM complex contains the core integral subunit BamA and four lipoproteins (BamB, BamC, BamD and BamE) (136, 137). The mechanism of how BamA or its lipoprotein partners are involved in the assembly of β-barrel proteins is yet to be fully explained, although speculations on some of these mechanisms have been made (refer to Introduction) (16, 80, 81). There are direct interactions between BamA and BamB, as well as BamA and BamD. BamC relies on the C-terminus of BamD to associate with BamA (213), whereas BamE interacts with BamA, BamC and BamD in a manner independent of BamB (136). In this set of lipoproteins, crystal structures are currently available for BamB (91) and BamE (99). Recently, the structure of BamC was solved showing a novel fold (214).

Since BamD is so highly conserved across different classes of proteobacteria and is essential for cell viability (16, 137), it was an ideal target for structural characterisation to

delineate regions that may be important for interactions within the complex and with β-barrel substrates of the BAM complex. Structural predictions were made that strongly advocated the presence of tetratricopeptide repeat (TPR) motifs that may mediate protein-protein interactions (16, 215). As part of my PhD project, solving the crystal structure of BamD was an important goal. Very recently in May of this year, two groups reported the structure of BamD from *E. coli* and *R. marinus* (214) (216).

In this chapter, we aimed to characterise BamD in the model α-proteobacterium, *C. crescentus* and optimise an approach to achieve sufficient amounts of BamD for crystallisation and structural analysis.

4.2 MATERIALS and METHODS

4.2.1 Bioinformatics

Basic local alignment search was performed with the protein BLAST software on the NCBI homepage (http://blast.ncbi.nlm.nih.gov/). Multiple sequence alignments were performed with CLUSTAL-W using CLUSTAL-X view (217). The corresponding TPR repeats were calculated with the free TPRpred software of the MPI Tübingen (http://toolkit.tuebingen.mpg.de/tprpred) as well as by manual inspection. PSIPRED was used for α-helix and ß-strands prediction (145).

4.2.2 Cloning and genetics in C. crescentus

Various BamD expression constructs were generated using reagents listed in Table I to achieve the expression of soluble protein in *E. coli.*

To generate the *bamD* mutant of *C. crescentus*, reagents listed in Table 4.1 were used. When appropriate, media were supplemented with antibiotics at the following concentrations (liquid/solid media for *C. crescentus*; liquid/solid media for *E. coli*; in µg/ml): spectinomycin (25/50; 50/100), kanamycin (5/25; 30/50), oxytetracycline (1/1; 12/12) and nalidixic acid (20/20; 0/0). The method of transferring plasmids into *C. crescentus* is detailed in Appendix A2. When double recombination did not occur due to gene essentiality, single recombinants were mated with S17.1 cells harbouring the replicating plasmids (prxmcs). These transformed single recombinants were spread on 3% (w/v) sucrose (and xylose) plates. 0.03% (w/v) xylose was added to cultures to induce gene expression from the prxmcs vector. Correct clones were identified by antibiotic selectivity and measuring relative levels of BamD after growing cells in glucose (0.2%) or xylose (0.03%) to repress or induce the gene promoter, respectively.

Table 4.1: Primers, vectors and cloning reagents used to generate mutants of *bamD* in *C. crescentus*

bamD construct	Primers	Info
bamD knockout plasmid	Upstream*Hin*dIIIFor: 5'-TACGAAGCTTCTGCGCGCCATCGGTCT-3', Upstream*Eco*RIRev: 5'-GCGCGAATTCACTTCGCGGAAATAGTC-3', Downstream*Eco*RIFor: 5'-AGTCGAATTCACTTCCCGGGCG-3', Downstream*Nhe*IRev: 5'-ACTAGCTAGCCGAACGACCGTC-3', SpectinomycinFor: 5'-CGGCCTGCAGAGTGGATCCCCCGGGCTGCA-3', SpectinomycinRev: 5'-CGGCGCTAGCGGTATCGATAAGCTTGATAT-3'	Upstream and downstream flanking regions (600 bp) of BamD were PCR amplified from *C. crescentus* genome and digested and ligated into the pnpts138 vector successively. Spectinomycin resistance cassette was amplified from pBor plasmid and inserted in the Pnpts138 construct using the *Eco*RI site.
bamD complementation plasmid	BamD*Nde*IFor: 5'-GGCCGCCATATGCTTCGTATTTTCCAGG-3', BamD*Sna*BIRev: 5'-TATAACTAGTCTACATGCCCAGCGGCCC-3'	Fragment was amplified using Vent polymerase to generate blunt ends. Fragment was digested with *Nde*I only and vector was digested with *Nde*I/*Sna*BI prior to ligation into replicating prxmcs plasmid.
bamD C-terminal truncation complementation plasmid	BamD*Nde*IFor, Cterm*Nhe*IRev: 5'-GTCAGCTAGCCTAGTTCAGAAGGCGG-3'	Fragment and vector were digested with *Nde*I and *Nhe*I prior to ligation into replicating prxmcs plasmid.

4.2.3 Expression and purification of BamD for antibodies

For antibody production, construct (*i*) in Table 4.2 was used. BamD was expressed and purified from inclusion bodies according to *Section 2.2.3* for the purification of BamA. Rabbits were immunised with purified BamD according *Section 2.2.4.*

4.2.4 Determination of protein concentration

Bradford assays (218) were performed to measure protein concentration. A standard curve constructed against known concentrations of bovine serum albumin (BSA) was used to estimate protein concentration of the sample. Protein samples were diluted with Bradford reagent (1:20 dilution) and incubated at RT for 5 min prior to measurement. The absorbance at 595 nm was measured at an appropriate dilution in a 96-well microplate. To measure the concentration of purified BamD, the Nanodrop instrument (Thermo Scientific) was used.

4.2.5 Gel electrophoresis and Western blotting

SDS-PAGE, BN-PAGE and immunoblotting were performed according to *Section 2.2.5*.

4.2.6 Subcellular fractionation

The method of separating inner and outer membranes from *C. crescentus* has been described in *Section 2.2.7*.

4.2.7 Co-immunoprecipitation

Co-immunoprecipitation of proteins using the anti-BamD antibody was performed according to *Section 2.2.9*.

4.2.8 Microscopy

The *bamD* depletion strain of *C. crescentus* was grown at 30°C for 16 h in the presence of 0.03% (w/v) xylose or 0.2% (w/v) glucose. Cells were prepared for scanning electron microscopy as follows: Cells were grown to an OD_{600} of 0.15 before centrifuging at 10,000 × g for 10 min at 4 °C and rinsing with 1 × PBS. The supernatant was removed and the pellet was resuspended in 3% (v/v) gluteraldehyde solution in 0.1 M phosphate buffer, pH 7.2 and fixed for 24 h. Cells were placed on sterile Thermanox plastic coverslips (Pro SciTech) for 10 min before the coverslips were washed twice with 0.1 M sodium cacodylate. Cells were dehydrated with cold ethanol solutions for 10 min each, using 70%, 90%, 100% and 100% (absolute) ethanol solutions. Cells were then treated with hexamethyldisilazane in 100% (absolute) ethanol for 10 min each, starting with a 25% solution continuing through 50%, 75% and 100% solutions. Cells were then coated with gold-palladium. Observation and photomicrographs were then carried out with a

Hitachi S-3400 N SEM (Hitachi Instrument Inc., Japan). *Image J* was used to analyse the SEM images.

4.2.9 Protease accessibility

To test for the topology of BamD, the method of trypsin shaving and polymyxin B treatment was used as detailed in *Section 3.2.8.*

4.2.10 Production and purification of inclusion bodies

Protein expression and inclusion bodies production was performed as outlined in *Section 4.2.3*. Inclusion bodies (IBs) were washed twice by resuspension in 4 % Triton X buffer (4 % Triton X-100, 150 mM NaCl, 50 mM Tris-HCl, pH 8.0) followed by two washes with 150 mM NaCl, 50 mM Tris-HCl, pH 8.0. Washed IBs were solubilised by addition of a small volume of Buffer A (8 M Urea, 100 mM Tris-HCl, 200 mM NaCl, 5 mM imidazole, pH 7.0) and further purified by affinity chromatography. Sample was cleared using a 0.45 μm syringe filter and loaded onto a 5 mL HisTrap column (GE Healthcare) connected to the AKTA *prime* FPLC system. Unbound material was removed from the column using Buffer A and bound material was eluted using Buffer B (8 M Urea, 100 mM Tris-HCl, 200 mM NaCl, 1 M imidazole, pH 7.0). A linear gradient (30 mL) of increasing imidazole concentration was used for elution. BamD elution was at 75-80% Buffer B or 750-800 mM imidazole. All buffers were cleared by a 0.45 μm filter, degassed under vacuum and stored at 4°C.

4.2.11 Refolding trials

The isolated inclusion bodies were solubilised in Buffer A and desalted by centrifugal concentrators using the buffer: 6 M Guanidine-HCl, 100 mM Tris-HCl, 20 mM DTT, pH

7.5. Small-scale (0.5 mL or 1 mL) refolding trials were done using 590 μg or 1 mg purified BamD for 18 - 20 h at 4°C. Final concentration of Guanidine-HCl in the refolding condition was 88 mM. Inclusion bodies were added to refolding buffer slowly and drop-by-drop with intermittent mixing to avoid localised precipitation. Samples were subsequently dialysed overnight in refolding buffer at 4°C. Samples were centrifuged to isolate soluble protein. Approximately 80% of the protein precipitated after refolding. Refolding at RT resulted in similar yields.

Refolded BamD was soluble at low concentration of approximately 0.2 mg/mL. To optimise conditions for concentrating the sample to at least 7 mg/mL, different buffer types at certain pH values and the addition of various additives known to affect protein solubility were trialed (detailed in Table 4.2). These conditions were used for dialysis and subsequent concentration using Amicon concentrators with a 10 kDa molecular weight cut-off.

For refolded BamD, a 120 mL Superdex 75 preparative grade (GE Healthcare) size exclusion column was equilibrated in 20 mM Tris-Cl, 150 mM NaCl, 2 mM DTT, pH 7.0.

4.2.12 Tryptophan spectroscopy

Fluorescence measurements were performed at RT on a Perkin Elmer LS50B apparatus. The excitation slit was set to 4 nm; the emission slit to 4 nm bandwidth and scan speed was 100. Measurements were taken from 300 to 450 nm using an excitation wavelength of 296 nm and the tryptophan emission spectra were baseline corrected. Folded and unfolded samples of BamD had the same amount of protein. The final Urea concentration in the unfolded sample was 6 M to achieve complete denaturation.

4.2.13 Expression and purification of soluble BamD

Expression was performed as outlined in section *4.2.3* except the cell pellet was resuspended in native buffer (50 mM Tris, 150 mM NaCl, pH 7.0) containing 0.2 mg/mL lysozyme, 25 μg/mL DNAseI, 10 mM $MgCl_2$, pH 7.0 and lysed by two passages through a pre-chilled French pressure cell at 20, 000 psi. The lysate was kept on ice and phenylmethanesulfonylfluoride (PMSF) was added to a final concentration of 0.034% (w/v). Inclusion bodies were pelleted by centrifugation (10,000 × *g*, 10 min, 4°C) and the supernatant was further processed as it contained soluble BamD. The Ni-NTA column was pre-equilibrated with Buffer A (50 mM Tris-HCl, 150 mM NaCl and 30 mM imidazole, pH 7.0) before large volumes of sample were loaded. Unbound material was removed by thorough washing of the column using 10 CV of Buffer A. Bound material was eluted using Buffer B (50 mM Tris-HCl, 150 mM NaCl and 1 M imidazole, pH 7.0).

BamD eluted from the HisTrap affinity column as a soluble protein that was stable for weeks in the elution buffer containing 600-800 mM imidazole. Samples were concentrated using Amicon concentrators with a 10 kDa molecular weight cut-off. Precipitation occurred and samples could not be concentrated more than 1 mg/mL. Subsequent precipitation occurred during storage at 4°C overnight.

For expressed soluble BamD, a Superdex75 10/300 (GE Healthcare) size exclusion column was equilibrated with 1.5 M Urea, 50 mM Tris, 100 mM NaCl, pH 7.0. Fractions were collected and analysed by SDS-PAGE.

4.2.14 Circular dichroism

CD analyses were performed using a Jasco J-815 spectrometer. Far-UV CD spectra from 190 to 260 nm were acquired at 20°C in a 1 mm path-length cuvette, with a 1 nm bandwidth, 1 s response time and 100 nm/min scan rate. The protein concentration of the sample was 0.288 mg/mL of BamD in 1.5 M Urea, 50 mM Tris-HCl pH 7.0. The protein

concentration was determined using the calculated extinction coefficient $\varepsilon_M = 48.82 \times 10^{-3}$ M^{-1} cm^{-1} and the molecular weight 32,853.1 Da of purified BamD. The spectrum, representing the average of three scans, was baseline corrected by subtracting the spectral attributes of the buffer. Spectra Manager (Version 2.08.02) was used to smooth the data using the Savitzky-Golay method for smoothing with a convolution width of 25. Raw data (in millidegrees) was converted to ellipticity (degrees cm^2 $dmol^{-1}$) by calculating the mean residue weight using the molecular weight of BamD as 32,853.1 Da and the number of amino acids of BamD as 289. The CDPro program (219) was utilized to assess the secondary structure contents of the proteins from the spectrum. CDPro fit curves and calculations of secondary structure content are presented in Appendix A4.

4.3 RESULTS

4.3.1 TPR predictions

BamD is a protein of 261 amino acid residues that is predicted to have at least five consecutive TPR units of 34 amino acid residues (Figure 4.1). TPR motifs from a variety of proteins exhibit a high degree of sequence diversity with conservation mainly in terms of size and hydrophobicity of some residues (215). The prediction for TPR motifs in BamD is mainly achieved by the TPRpred program that detects structural repeats based on hidden Markov models or profiles generated from known repeat sequences (220). Sequences closely related to *C. crescentus* BamD were also analysed by this program to define regions of TPR motifs that could not be detected with the *C. crescentus* BamD sequence or those that could be detected albeit with low confidence. Hydrophobic residues are commonly observed at positions 4, 7, 11, 20, 24 and 27 while glycine and proline frequently occupies positions 8 and 32 respectively within the repeat (215). Aromatic residues like tyrosine, tryptophan or phenylalanine are frequent at position 17. Furthermore, the segment between TPR3 and TPR4 and the C-terminus of BamD are predicted to adopt α-helical structure and may be involved in stabilising the TPR repeats.

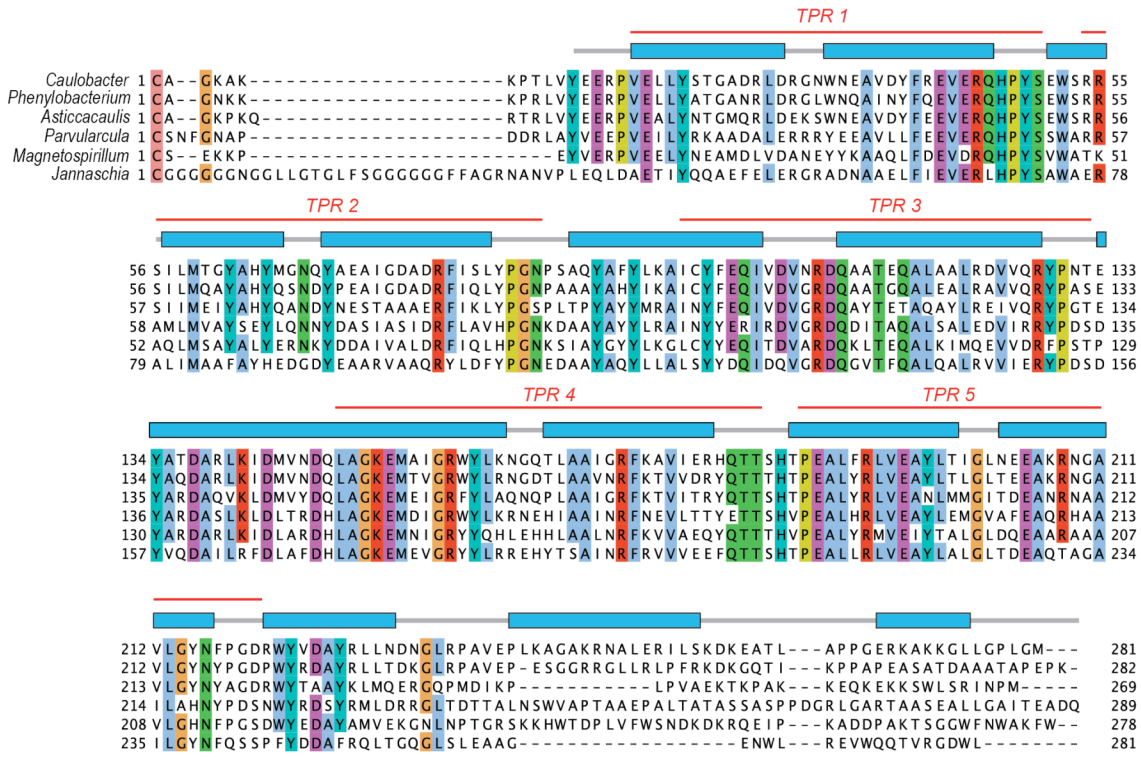


Figure 4.1: CLUSTAL-X view of the multiple sequence alignment of BamD homologues showing TPR and secondary structure predictions. Red lines demarcate TPR motifs and the predicted secondary structure is shown as bars for α-helices (light blue). The numbers at the start and the end of the sequences indicate the first and last amino acid positions of the alignment.

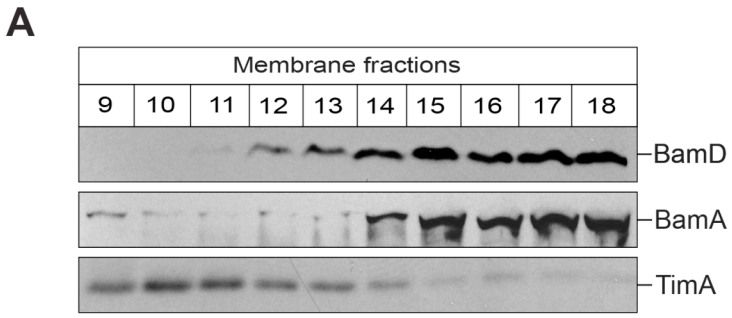
4.3.2 BamD is an outer membrane lipoprotein with a periplasmic domain

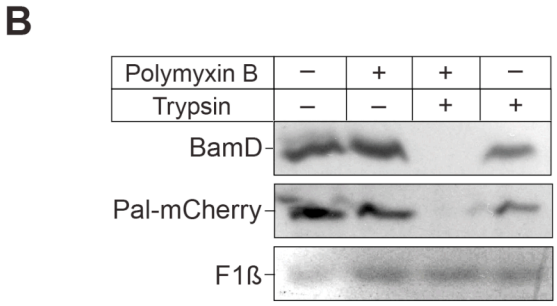
To verify the outer membrane location of BamD in *Caulobacter*, membranes were separated by sucrose gradient fractionation into two distinct populations: inner and outer membranes. The enrichment of BamD in fractions 15-19 suggests that it is a component of the outer membrane (Figure 4.2A). The outer membrane protein, BamA, is found in similar fractions whereas the inner membrane protein, TimA, is mainly found in fractions 8-13.

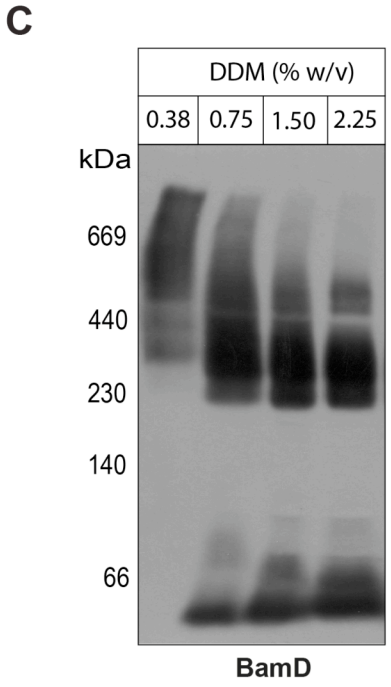
Outer membrane lipoproteins are assumed to sit exposed to the periplasm. To verify the presence of a periplasmic domain of BamD, a protease accessibility experiment was performed using polymyxin B to disrupt the outer membrane layer followed by the addition of trypsin. Addition of polymyxin B and trypsin degrades BamD, suggesting it has a protease-sensitive, periplasmic domain (Figure 4.2B).

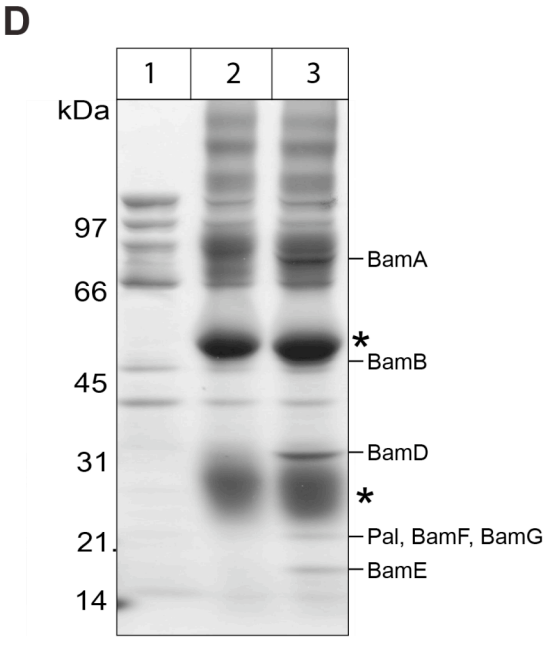
To investigate the quaternary structure of BamD in *C. crescentus*, outer membrane vesicles were solubilised with increasing concentrations of DDM and separated by BN-PAGE. Antibodies against BamD reveal a complex of ~500 kDa in size at 0.38% (w/v) DDM. Increasing the detergent concentration at or above 0.75% DDM disrupts the complex, resulting in a ~300 kDa complex as the major species as well as some monomeric and possibly dimeric forms of BamD (Figure 4.2C). The previously identified ~150 kDa module of BamA is not detected. To identify proteins interacting with BamD, immunoprecipitation was performed by adding ProteinA Sepharose-crosslinked anti-BamD antibodies to outer membranes that were solubilised with 0.75% DDM. As shown, the antibodies precipitated proteins that were previously identified by immunoprecipitation with the BamA antibody (Figure 4.2D). Experiments using the pre-immune serum demonstrate the specificity of the co-precipitation. This shows that BamD is specifically and uniquely associated with the BAM complex.

Figure 4.2: Characterisation of BamD in *C. crescentus*. (A) Protease shaving of *C. crescentus* expressing the periplasmic Pal-mCherry fusion protein. Cells were incubated with (Lane 2, 3) and without (Lane 1, 4) polymyxin B followed by addition of trypsin (Lane 3, 4). Samples were analysed by SDS-PAGE and immunoblotting. Blots were incubated with antibodies against BamD, Pal-mCherry and F1ß (cytoplasmic protein). (B) Western blot analysis using antibodies recognising BamD, BamA and TimA (inner membrane) in fractions containing inner and outer membranes separated on a sucrose gradient. (C) BN-PAGE and immunoblotting using anti-serum against BamD. (D) Immunoprecipitation of outer membrane vesicles solubilised with 0.75% DDM using BamD anti-serum. Total outer membranes used for the experiment is shown in Lane 1, the material that is immunoprecipitated by the pre-immune serum and BamD anti-serum are shown are shown in Lane 2 and Lane 3, respestively. Asterisks (*) represent the area of IgG heavy and light chains.


A
Membrane fractions
9
10
11
12
13
14
15
16
17
18
BamD
BamA
TimA


B
Polymyxin B
Trypsin
BamD
Pal-mCherry
F1ß


C
DDM (% w/v)
0.38
0.75
1.50
2.25
kDa
669
440
230
140
66
BamD


D
1
2
3
kDa
97
66
45
31
21
14
BamA
*
BamB
BamD
*
Pal, BamF, BamG
BamE

4.3.3 BamD is essential for outer membrane integrity in C. crescentus

Unsuccessful attempts to delete the gene encoding BamD suggest that it is an essential gene in *C. crescentus*. Furthermore, double homologous recombination was not possible when a C-terminal truncated version of BamD was expressed, suggesting the C-terminus is essential for viability. We therefore constructed a strain in which the *bamD* gene was under the control of a xylose-inducible promoter. This strain was unable to form colonies on complex media with glucose and lacking xylose, formally confirming that *bamD* is an essential gene. To deplete levels of BamD, cells were grown overnight in media containing 0.03% (w/v) xylose and then inoculated into fresh medium lacking xylose and containing 0.2% (w/v) glucose. The levels of BamD greatly decreased at 6 h after removing xylose from the medium. As controls to the experiment, BamA or TimA levels remained relatively constant (Figure 4.3A).

While the depletion strain required xylose for growth, the presence of many normal cells suggests suppressor mutations were acquired. Microscopic examination of the BamD depletion strain growing in the absence of xylose (↓ BamD) for 16 h displayed vesicles and blebs along the surface (Figure 4.3C). The depletion strain appeared like wild-type *Caulobacter* when grown in the presence of 0.03% xylose (Figure 4.3B). As is the case in *E. coli*, we conclude that in *C. crescentus*, *bamD* is an essential gene (213).

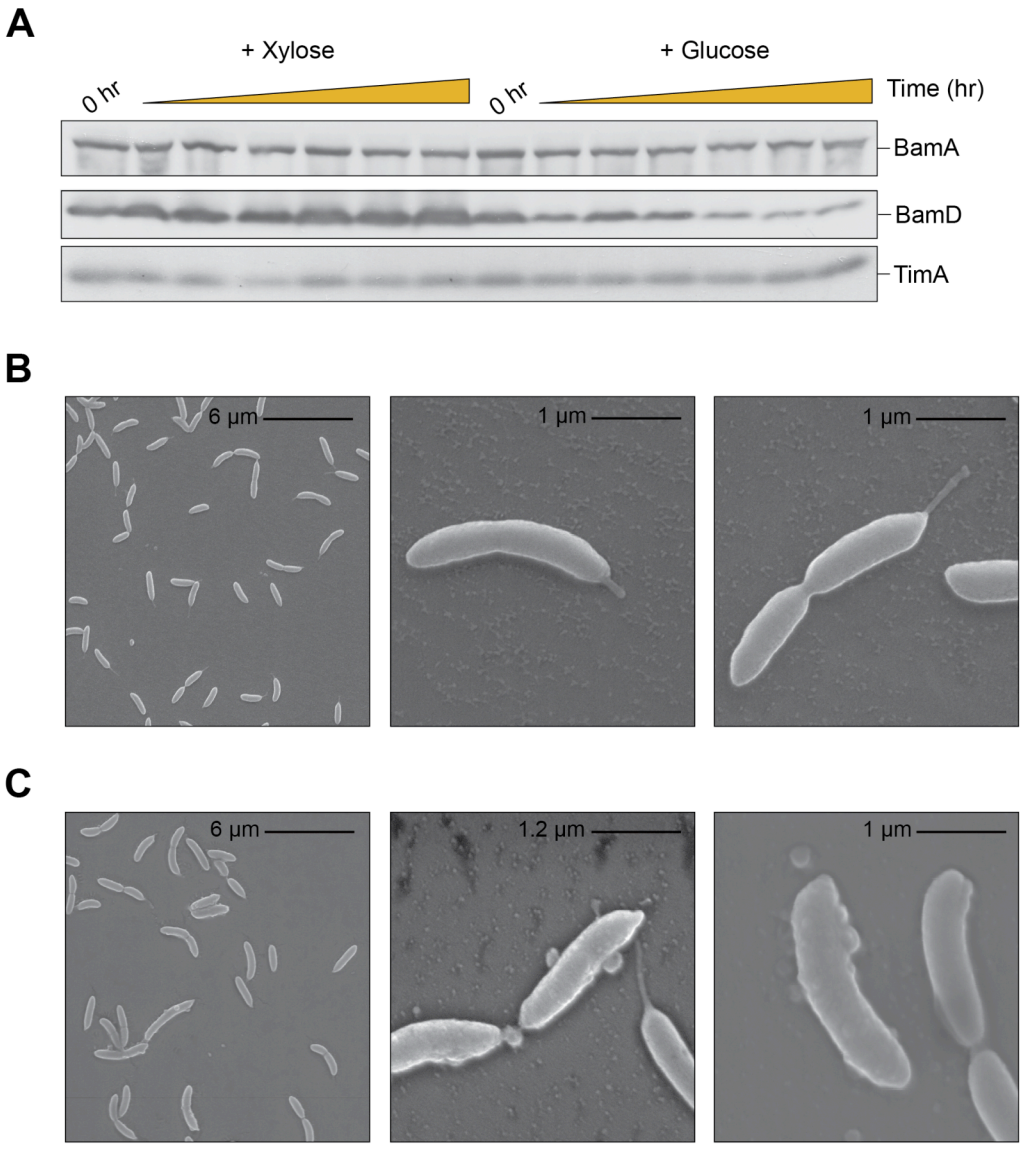


Figure 4.3: Depleted levels of BamD results in cell morphology defects. (A) The depletion mutant of BamD grown for 6 h in the presence of xylose or glucose. Lanes labelled '0 h' correspond to overnight cultures grown in xylose. These cultures were washed and diluted prior to addition of xylose or glucose. (B-C) Scanning electron micrographs of *C. crescentus* BamD depletion strain grown in the presence of xylose (B) or in presence of glucose (C). Bar markers represent length across the micrograph.

4.3.4 Recombinant refolded BamD is a soluble protein at low protein concentration but is prone to aggregation

With the aim of obtaining structural information, a hexa-histidine tag was incorporated at the C-terminus of BamD for expression of the protein in *E. coli* (Figure 4.4A). Expression trials did not produce obvious amounts of soluble protein and so we undertook the task of purifying inclusion bodies (Figure 4.4B) and refolding BamD by using various buffer conditions to achieve soluble protein at sufficient amounts for further analyses (see Table 4.2). The refolding trials produced soluble protein at very low protein concentration, with 1.25 mg/mL as the highest concentration obtained. The presence of DTT increased the amount of soluble protein recovered from refolding trials. Large amounts of precipitation occurred after dialysis and during concentration of the sample. After concentration, further precipitation occurred upon storage at 4°C, suggesting BamD is unstable. While this propensity to aggregation is not unexpected for an essential component of a molecular machine, it required much time in optimisation to express soluble BamD using various conditions. These conditions related the use of different *E. coli* strains for protein expression [BL21 (DE3), BL21 (DE3) GroEL/S, BL21 (DE3) Codon Plus, BL21 (DE3) Rosetta and C41 cells], incubation temperatures (18°C, 25°C and 37°C) as well as the concentrations of IPTG (0.05 mM and 0.5 mM) for inducing protein expression.

Size exclusion chromatography was used to separate refolded BamD from larger aggregates. However, refolded BamD eluted in the void volume of the column (Figure 4.4C). This indicates refolded BamD aggregates.

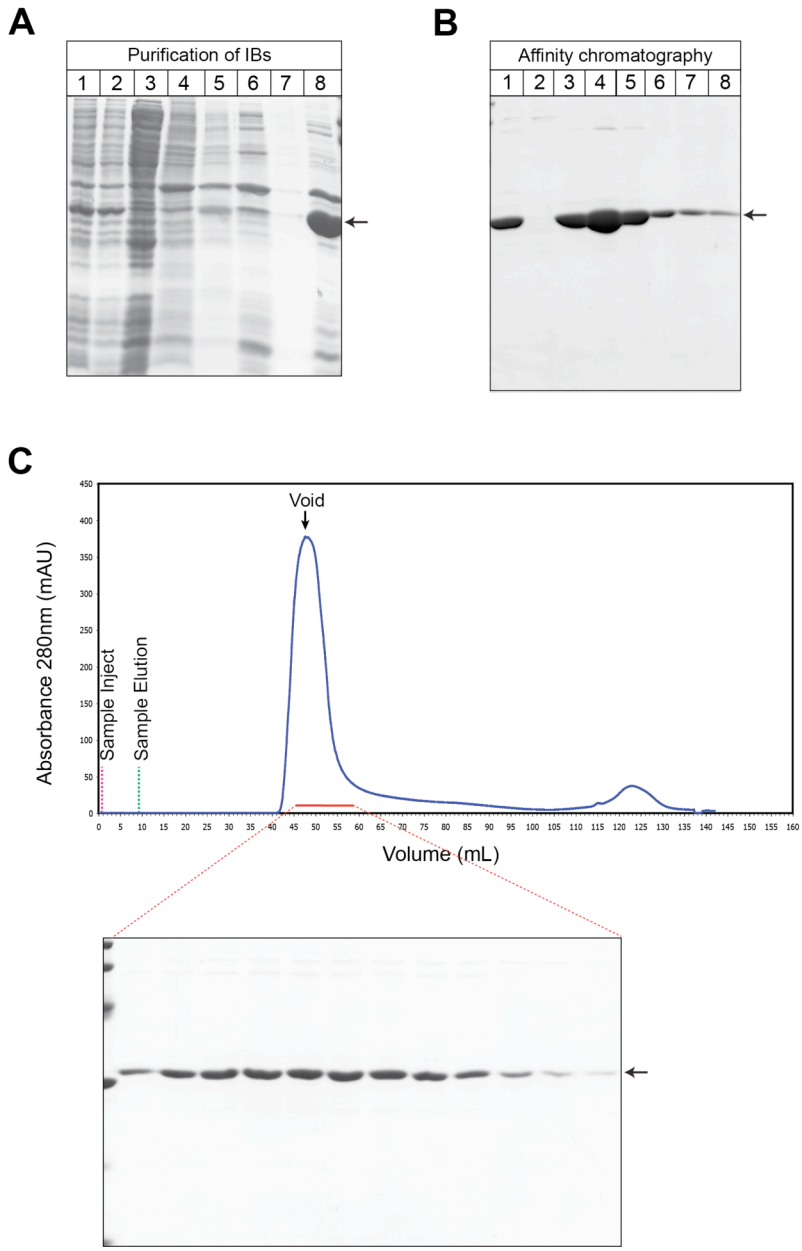


Figure 4.4: Full-length BamD expressed with the N-terminal cysteine residue is fully insoluble. (A) BamD expression was induced with IPTG (Lane 1) and mainly expressed in inclusion bodies (Lane 2), not in the soluble fraction (Lane 3). Inclusion bodies were thoroughly washed to remove contaminants (Lane 4 - 7) and produce purified inclusion bodies of BamD (Lane 8). (B) Affinity purification of BamD expressed in inclusion bodies (Lane 1). Proteins that did not bind the column are shown in Lane 2 and proteins that bound the column and eluted with high concentration of imidazole are shown in Lane 3-8. (C) Gel filtration chromatogram showing refolded BamD eluting at the void volume of the column and SDS-PAGE gel analysis of fractions corresponding to the main peak of the chromatogram. The arrows indicate the position of BamD on the gel.

Table 4.2: Buffer types used for refolding BamD and the yield (in mg/mL) of soluble protein

Condition	Refolding buffer	Protein concentration (mg/mL)
1.0	100 mM Tris-HCl, 5 mM EDTA, 2 mM DTT, pH 7.5	0.20
1.1	As Cond. (1.0) with 0.25 M Arginine	0.17
1.2	As Cond. (1.0) with 0.5 M Arginine	0.13
1.3	As Cond. (1.0) with 33 mM CHAPS	0.17
1.4	As Cond. (1.0) with 10% glycerol	0.10
1.5	As Cond. (1.0) with 0.25 M Arginine, 10% glycerol	0.12
2.0	50 mM Tris-HCl, pH 7.3	0.06
2.1	50 mM Tris-HCl, 2 mM DTT, pH 7.3	0.24
2.2	50 mM Tris-HCl, 6 mM DTT, pH 7.3	0.43
3.0	50 mM Tris-HCl, pH 9.3	0.02
3.1	50 mM Tris-HCl, 2 mM DTT, pH 9.3	0.65
3.3	50 mM Tris-HCl, 4 mM DTT, pH 9.3	1.05
3.4	50 mM Tris-HCl, 6 mM DTT, pH 9.3	1.25*
4.0	100 mM K_3PO_4, pH 6.2	0.03
4.1	100 mM K_3PO_4, 6 mM DTT, pH 6.2	0.23
5.0	100 mM K_3PO_4, pH 7.2	0.04
5.1	100 mM K_3PO_4, 6 mM DTT, pH 7.2	0.24
6.0	100 mM Na_3PO_4, pH 6.2	0.06
6.1	100 mM Na_3PO_4, 6 mM DTT, pH 6.2	0.4
7.0	100 mM Na_3PO_4, pH 7.2	0.06
7.1	100 mM Na_3PO_4, 6 mM DTT, pH 7.2	0.21
8.0	50 mM HEPES, pH 7.3	0.06
8.1	50 mM HEPES, 6 mM DTT, pH 7.3	0.23

4.3.5 BamD can be partially expressed as a soluble protein in E. coli

Production of soluble protein for crystallisation trials is a common hurdle in structural biology. Various constructs of BamD were generated (Figure 4.5 A-F) and different expression conditions were trialed to express soluble protein in *E. coli*. Of the six constructs that were generated (see Table 4.3), only BamD-His_6 yielded soluble protein, albeit at very low levels (Figure 4.5A). With this construct, most of the BamD protein was in inclusion bodies but a small proportion was soluble and could be purified.

Table 4.3: Primers and cloning reagents used for heterologous expression of BamD

***bamD* construct**	**Primers**	**Info**
i. Cys-BamD-His_6	BamDCys*Nde*IFor: 5'-GAATCATATGTGCGCCGGCAAG-3', BamD*Xho*IRev: 5'-TAATCTCGAGCATGCCCAGCGG-3'	*bamD* containing the N-terminal cysteine residue was cloned into pET22b vector to express BamD with a C-terminal hexa-histidine tail for affinity-based purification.
ii. BamD-His_6	BamDNoCys*Nde*IFor: 5'-GAATCATATGGCCGGCAAGGCC-3', BamD*Xho*IRev	Mature BamD (excluding N-terminal cysteine) with a C-terminal hexa-histidine tag expressed from pET22b vector. N-terminal cysteine was removed to avoid potential crosslinking with internal cysteine residue leading to protein aggregation and affecting protein solublity.
iii. BamD	BamDNoCys*Nde*IFor primer, BamDNoHis*Xho*IRev: 5'-TAACCTCGAGCTACATGCCCAG-3'	Untagged mature BamD expressed from pET22b vector. No fusion partner was added due to potential inteference of the tag with the solubility of BamD.
iv. BamD-GST	BamDFor*Bam*HI: 5'-TAATGGATCCGCCGGCAAGGCC-3', BamD*Xho*IRev primer	GST fusion of BamD was generated using PGEX-4T1 vector. GST is commonly known to increase the solubility of many heterologously expressed proteins.
v. His_6(TEV)-BamD	BamD subcloned from PGEX-4T1 construct	Mature BamD with a TEV protease-cleavable his-tag at the N-terminus that is cloned into the ProExHtb vector. Cleavage of the his-tag may increase solubility of the protein.
vi. BamD_truncated-His6	BamDNoCys*Nde*IFor primer, Cterm*Xho*IRev: 5'-GATACTCGAGGTTCAGAAGGCGGTAG-3'	His-tagged BamD with C-terminal truncation of 51 amino acids expressed from pET22b vector. The C-terminal portion that was removed corresponded to the predicted C-terminal capping helix.

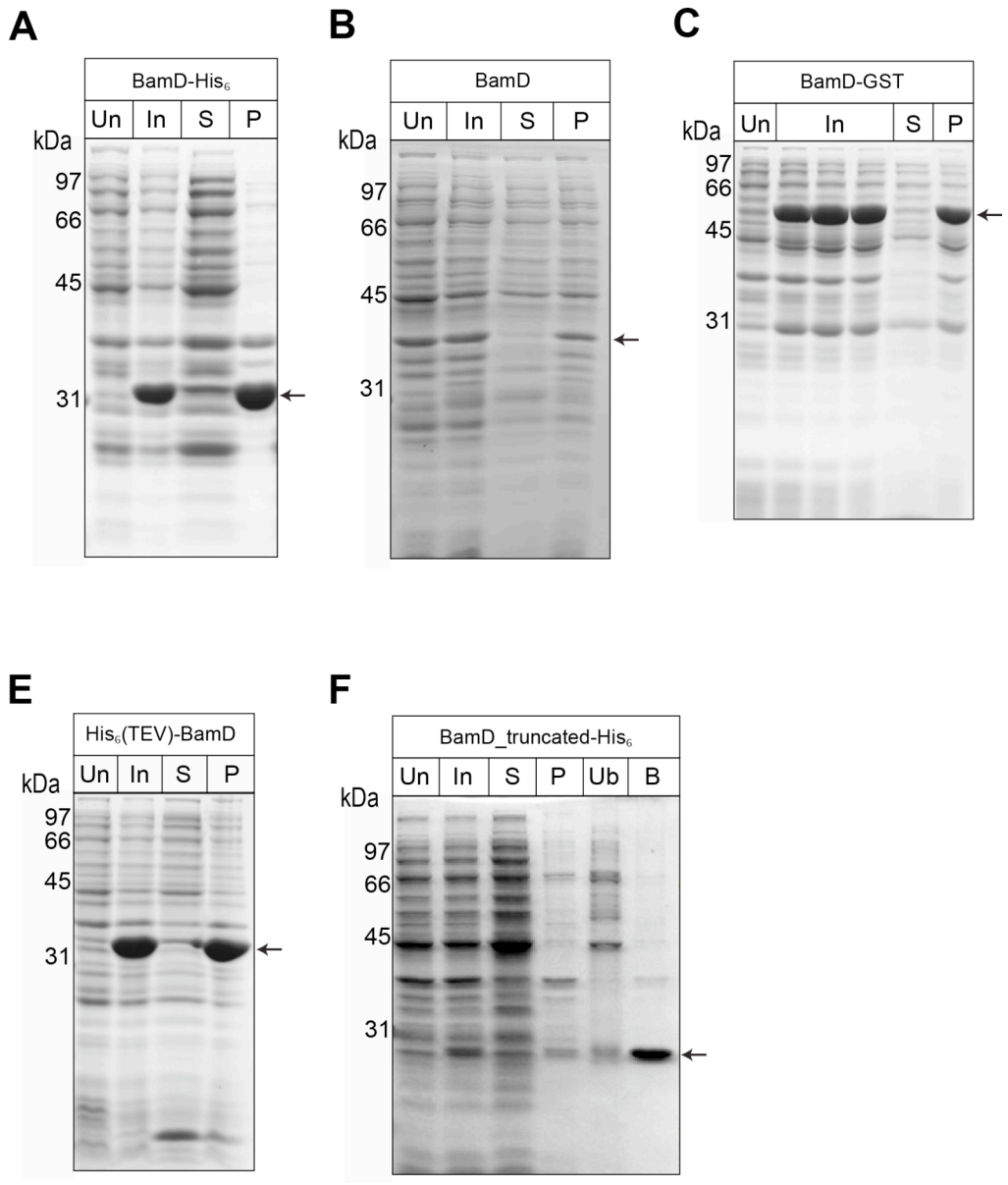


Figure 4.5: Expression of various BamD constructs to achieve solubility. (A-F) Uninduced cell culture ('Un' lane) of BL21 (DE3) codon plus cells transformed with various BamD constructs was induced (In) with IPTG. Following induction, cells were lysed and centrifuged resulting in soluble supernatant and insoluble pellet. Samples for uninduced (Un), Induced (In), supernatant (S) and insoluble pellet (P) fraction were analysed by SDS-PAGE. All constructs lacked the N-terminal signal sequence and cysteine residue. (F) Affinity purification of proteins in the insoluble pellet (P) shows C-terminal truncated BamD is in the bound (B) and not in the unbound (Ub) fraction, indicating truncated BamD is indeed expressed and in the insoluble pellet (P) fraction. The arrows indicate the position of BamD on the gel.

4.3.6 Purified soluble protein is monomeric and stable at low protein concentration

Large-scale cultures (10 L) were used to express and purify approximately 6 mg of BamD by nickle-based affinity purification (Figure 4.6A). Further purification by gel filtration showed the retention volume on a size-exclusion column corresponds to the expected size of BamD (~30 kDa) (Figure 4.6B). However, precipitation was still evident after gel filtration and during protein concentration. The solubility of BamD was dependent on the concentration of imidazole. The limit for concentration was about 0.3 mg/mL whereas in the presence of 600-800 mM imidazole post-nickle-affinity purification, the protein fractions were concentrated to no more than 1 mg/mL.

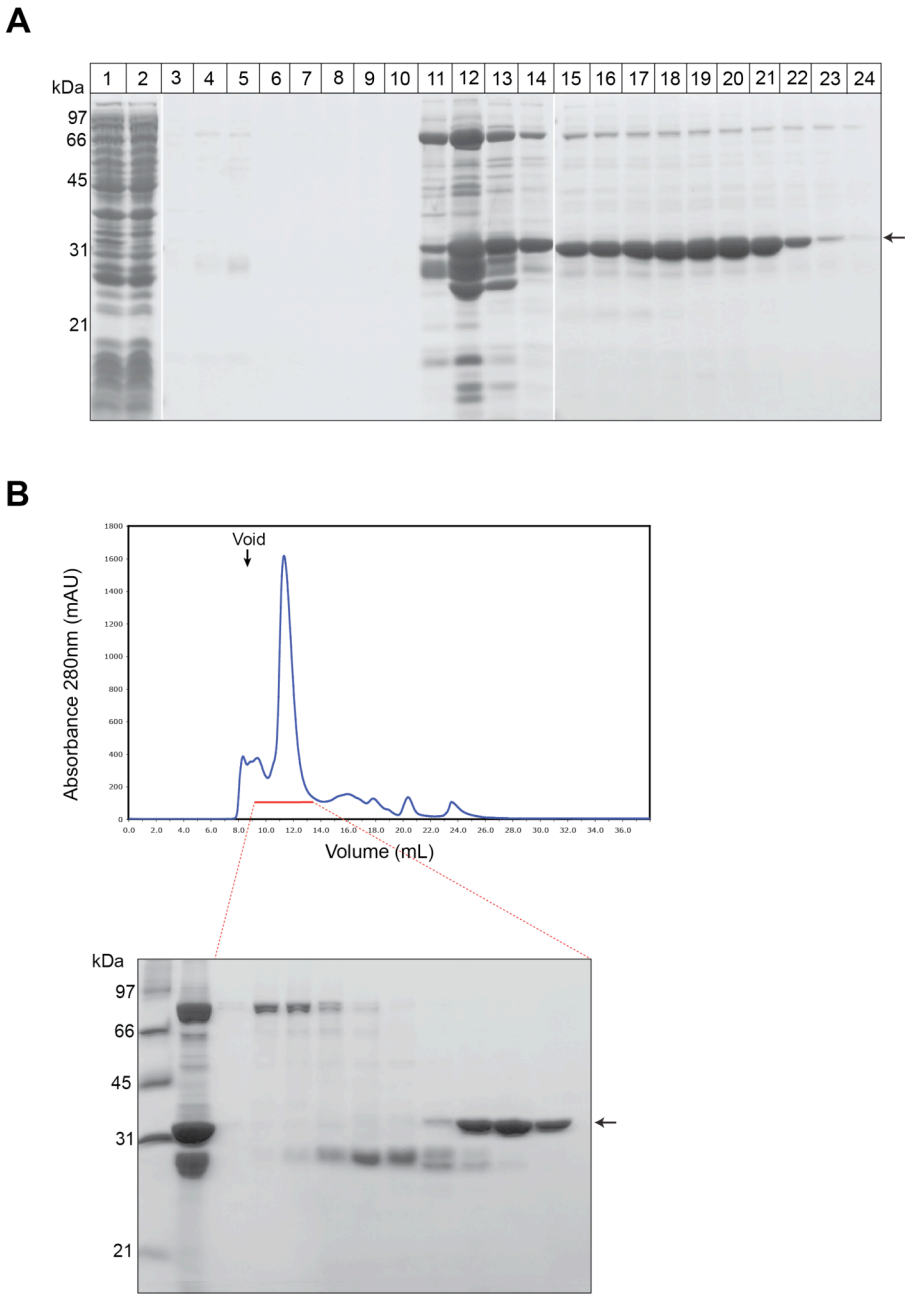


Figure 4.6: Purification of soluble BamD. (A) Nickel-based affinity purification of soluble BamD. Soluble fraction (Lane 1) was passed through a Ni-NTA column to remove unbound (Lane 2) proteins. Bound proteins were eluted using a step-gradient in 5% Buffer B increments (Lane 3 -Lane 24). (B) Gel filtration chromatogram of purified soluble BamD. Fractions from the main peak (red line) were analysed by SDS-PAGE. The arrows indicate the position of BamD on the gel.

4.3.7 BamD can be stabilised by Urea with minimal effect on the folded state

Since BamD proved unstable in several buffer systems and could not be concentrated to sufficient levels for crystallisation trials, we used low levels of Urea to stabilise the protein. The protein was purified using the previous approach except in the presence of 1.5 M Urea in the gel filtration buffer. The protein was then concentrated to approximately 5.6 mg/mL with minimal precipitation although BamD from *E. coli* was concentrated to approximately 40 mg/mL using this approach (214). To determine whether the structure of BamD is affected by the presence of a denaturant, we used two approaches to assess the folding state of BamD: tryptophan spectroscopy and circular dichroism. Tryptophan is an intrinsic fluorophor of proteins and the fluorescence is sensitive to environmental changes. Hence, tryptophan fluorescence emission changes as the Trp indole ring is moved from a polar to an apolar environment. Four Trp in BamD were excited and the emission monitored over a wavelength range. The results obtained were similar to those observed for the porin OmpA (60). Figure 4.7A shows that unfolded BamD had emission maxima at 352 nm compared to the maxima of 344.5 nm for the folded protein. The folded protein also displayed increased emission intensity compared to denatured BamD, which suggests that the Trp fluorescence in the native protein is quenched.

To determine the secondary structural content of the purified protein, circular dichroism (CD) was undertaken. Figure 4.7B shows the troughs at 208 and 222 nm are characteristic of a high content of α-helix and deconvolution of the spectra using CDPro calculates α-helical content of 65-82% (see Appendix A4). This compares well with the known α-helical content of 72% for *R. marinus* BamD (216). Due to the unfavourable presence of optically active reagents (mainly Urea) in the buffer, the dynode voltage increased more than 500 during CD measurements, resulting in low signal-noise ratio. This corresponded to CD measurements below 210 nm (Figure 4.7C). Only data measured between 210- 240 nm was used for CDPro analysis.

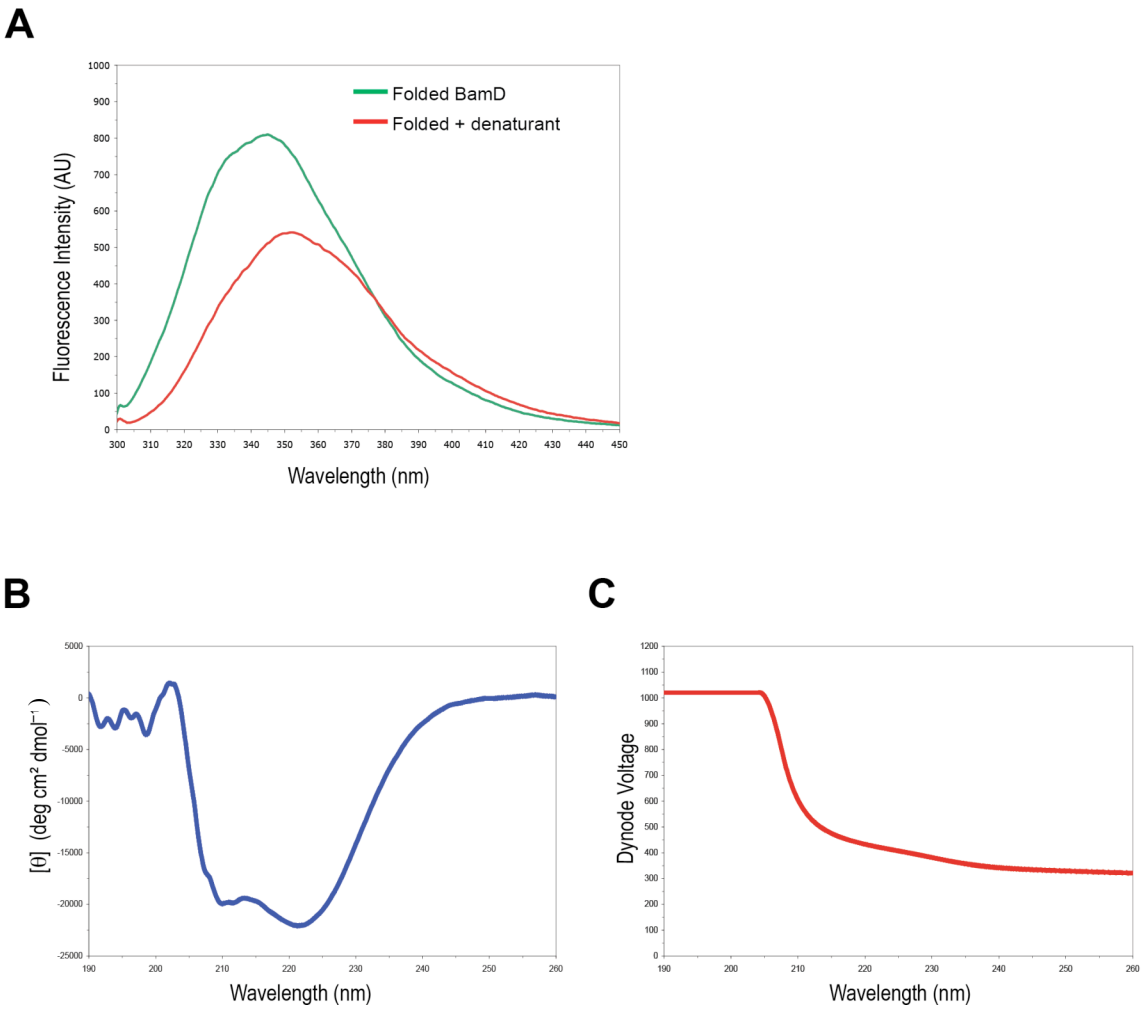


Figure 4.7: Characterisation of purified soluble BamD. (A) Tryptophan spectroscopy of soluble 'folded' BamD (green line) compared to unfolded or denatured protein (red line). (B) The circular dichroism spectrum of BamD. (C) The change in the photomultiplier tube dynode voltage as a function of wavelength for the conditions used.

4.3.8 BamD from R. marinus is similar to C. crescentus BamD

Using X-ray crystallography, the structure of BamD from *R. marinus* was solved in May of this year (Figure 4.8A) (216). The structure shows five TPR motifs in a close-packing arrangement such that the repeats form an elongated super-helix that is 94 Å long (9.4 nm). A C-terminal and an internal capping helix, fused with the first helix of TPR-4, are also present to stabilise the structure. A negatively charged groove is formed between TPR-3 and TPR-4 as a result of this re-arrangement and has been proposed as a binding region. Very recently, Albrecht *et al.* (214) solved the structure of BamD from *E. coli* that showed five TPR repeats with an extended helix of TPR-4 and the absence of the C-terminal capping helix. Sequence alignment shows *E. coli* BamD is shorter and missing the last helix that caps TPR-5 (216). However, a sequence alignment shows *C. crescentus* BamD is more similar to *R. marinus* BamD and contains the C-terminal capping helix (Figure 4.8B).

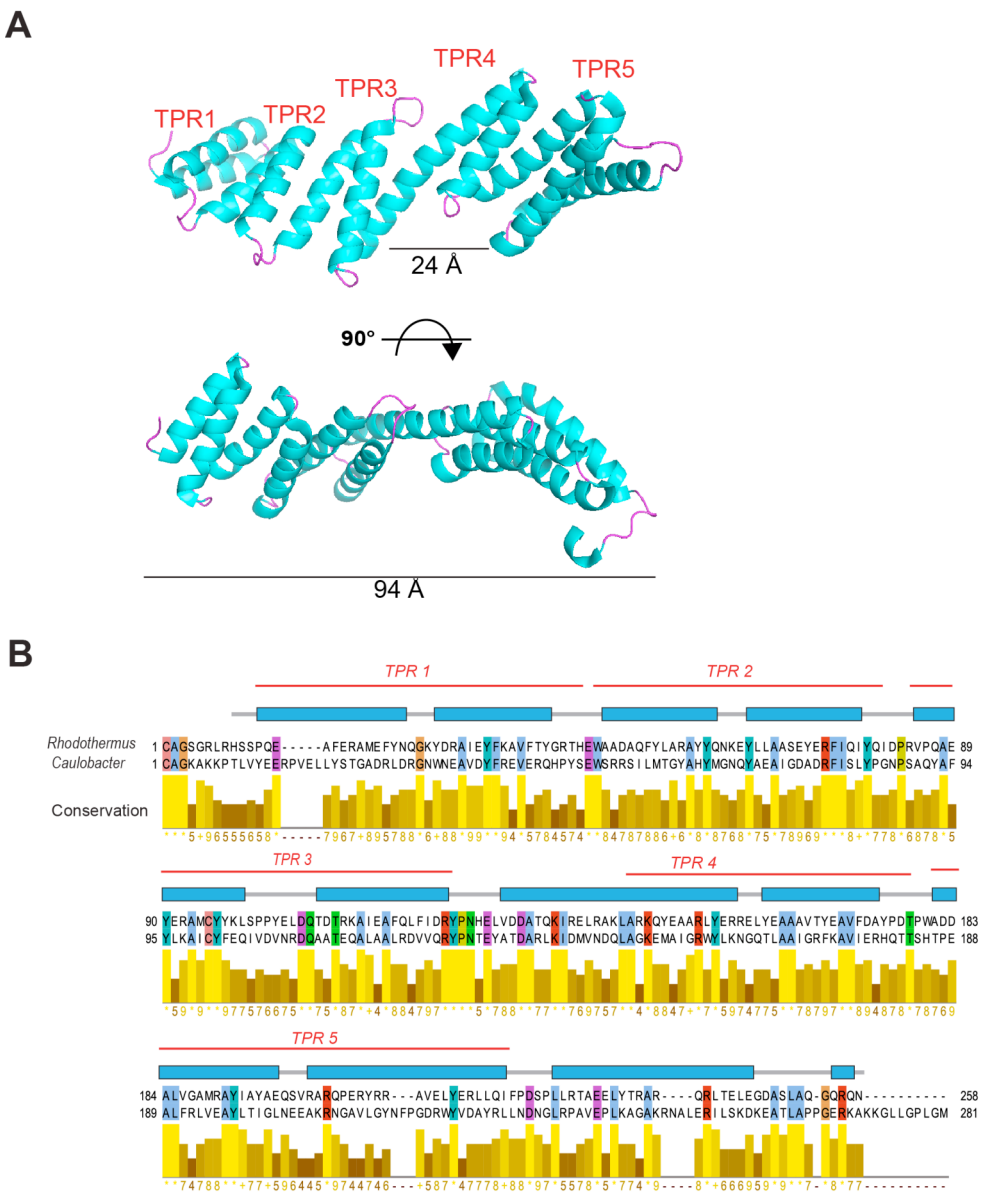


Figure 4.8: **Structure of *R. marinus* BamD and sequence comparison to *C. crescentus* BamD.** (A) Crystal structure of *Rhodothermus* BamD showing two views related by a 90° rotation (PDB code 3QKY). (B) Sequence alignment of BamD from *Rhodothermus* and *Caulobacter* showing known secondary structure based on the structure of *Rhodothermus* BamD. Secondary structure is shown as bars (light blue) for α-helices while the red lines demarcate TPR motifs. Below the sequence alignment is the conservation plot, which is based on a 1-10 scale with completely conserved residues highlighted with yellow bars and asterisks (*).

4.4 DISCUSSION

BamD is considered a ubiquitous protein, generally associated with cell viability and outer membrane integrity (213). In *E. coli* and *N. meningitidis*, BamD is the only essential lipoprotein of the BAM complex (204, 208, 213). However, this essentiality is not seen in *S. enterica* and bioinformatic searches indicate BamD is completely absent from other species of bacteria such as the spirochete *B. burdorferi* (208, 221). In *C. crescentus*, a BamD homologue exists but gene knockout was not possible without a complementing plasmid; suggesting BamD is essential for cell viability. Depletion of BamD resulted in changes to the surface of *C. crescentus*, similar to that observed for the Pal depletion strain (Chapter 3). Formation of blebs or membrane vesicles at the cell surface is a regular occurrence during growth, possibly due to peptidoglycan turnover during growth or overproduction of the outer membrane compared to the peptidoglycan layer (222, 223). However, vesicle formation can be enhanced by disturbances in growth or outer membrane integrity (224, 225). Hence, we can conclude that BamD is important for maintaining the integrity of the outer membrane.

Other bacterial species expressing mutant forms of BamD have major phenotypic defects and diminished levels of assembled and integrated outer membrane proteins relative to wild-type (92, 213). In *N. gonorrhoeae*, a transposon insertion into *bamD* (*comL*) resulted in a truncated form of BamD containing 96 out of the 251 amino acids residues of the mature protein (92, 208). Transposon insertions upstream of this insertion site were attempted but were unsuccessful. Defects as a result of disrupting gonococcal *bamD* related to smaller cell size, growth rate and decreased competence for DNA uptake (92). In *E. coli*, a less drastic truncation encoding 208 of the 226 amino acid residues of the mature protein was achieved and resulted in decreased steady state levels of outer membrane proteins, loss of BamC and weakened binding of BamE to the complex (136, 213). The effects of these partial loss of function implies the N-terminal region of BamD is essential (80) although the C-terminus partially facilitates BamC and BamE interaction with the BAM complex (213). Attempts to generate a truncated *bamD* mutant in *C. crescentus* were not successful. This C-terminal truncation removes the last 51 amino

acid residues of the 285 residues comprising the full-length mature protein and corresponds to the C-terminal extension of BamD.

During my PhD research, there was no structural information of BamD but our prediction, based on multiple sequence profiling of TPR motifs, suggested BamD contained five tandem helix-loop-helix structures capped by a C-terminal extension. These motifs are important in biological scenarios where protein-protein interaction occurs such as in a protein complex or during transient interactions with other proteins (215). This is of relevance to BamD as it is involved in interactions with the most C-terminal POTRA domain of BamA, BamC and possibly with substrates assembled by the BAM complex (72, 80, 213). TPR motifs are also present in the Tom70 protein, the major receptor for mitochondrial proteins in the TOM complex (226). This is a highly relevant precedent since the TPR segments of Tom70 act as docking points for mitochondrial β-barrel precursors prior to their import by the TOM complex and assembly by the SAM complex (226, 227). Since Tom70 also acts as a co-chaperone that co-operates with the cytosolic chaperones Hsp70 and Hsp90 in eukaryotes (228, 229), it could be conceived that BamD may serve a similar function to the major bacterial periplasmic chaperone SurA. There are several co-chaperones that contain TPR motifs, which mediate interactions with chaperones and their associated proteins (229). The ancestry of Tom70 in unknown but it bears no resemblance to BamD other than a common structural fold. BamD also does not share significant similarity with Sam35 and Sam37, the partners of mitochondrial Sam50. Although convergent evolution of TPR repeats is likely to be rare and is difficult to establish (230), the relevance of Tom70 to BamD is likely at the level of convergent evolution toward essential domains for protein interaction (226). An example of convergent evolution of TPR proteins is given by the Tom20 subunit of the TOM complex in animal and plant cells (231).

Considering the essential nature of BamD, obtaining structural information of the protein was one of the prime goals in our field of research. Upon commencement of the project, it was not known if BamD functions by forming a groove via its TPR motifs or uses its TPR motifs as a scaffold for protein interactions. Numerous attempts were made to

express and purify full-length BamD to sufficiently concentrate the protein for structural characterisation. The insolubility experienced in our experiments might be explained by the fact that BamD is natively part of a multi-protein complex and might be stabilised by interactions with BamA or other lipoproteins. Co-expression of the different subunits in *E. coli* was an alternative solution to stabilise and increase solubility of BamD. Recently, Albrecht *et al.* reported achieving X-ray diffraction-quality crystals of BamD from *E. coli* (212). The group mentioned the instability of purified BamD at high protein concentration but the use of 1.5 M Urea was able to circumvent problems with protein instability. Using this same approach, we concentrated BamD from *C. crescentus* to levels that would be appropriate for crystallisation trials. However, once the crystal structure of BamD from *R. marinus* was solved, it impacted on our progress towards setting up crystallisation trials as a similar structural arrangement was proposed for *C. crescentus* BamD; in fact, *C. crescentus* and *R. marinus* BamD are similar in terms of size and sequence that a structural model can be built. The structure also shows that a groove formed between the N- and C-terminus of BamD is a probable site for protein-protein interaction (216). The C-terminal truncation of *C. crescentus* BamD that could not complement the full-length protein (discussed earlier) may be related to disruption of this groove.

An interesting observation from the *R. marinus* BamD structure is the formation of an elongated super-helix that could, like the POTRA domains of BamA, span the periplasmic space in bacteria. A similar mechanism to the POTRA domains of BamA is proposed for BamD, such that it acts as a receptor for incoming β-barrel proteins. The N-terminal domain of BamD is arranged so the first helix of each TPR (TPR-1, -2 and -3) are arranged to form a concave surface, which bears resemblance to proteins that interact with their substrates or chaperones. Sandoval *et al.* (216) compared this arrangement to proteins such as Tom70 that recruits Hsp70/90 chaperones and interacts with β-barrel precursors for translocation into mitochondria. Like Tom70 in the TOM complex, BamD could provide a receptor function within the BAM complex.

As a consequence of this recent revelation into the structure of BamD, the most

interesting endeavour now would involve obtaining a hetero-complex of BamD with other components of the BAM complex such as BamF (Chapter 3) or BamE from *C. crescentus* since BamC and BamE are known partners of BamD in *E. coli* (136, 213). Although structures of individual components of the BAM complex are important, elucidating the mutual protein interactions will further helps us understand the role of BamA and its lipoprotein partners in recognising and inserting β-barrel proteins.

Conclusion

The Omp85 family proteins were classified according to sequence and domain structure. In bacteria alone, there are three groups of proteins (BamA, TpsB and YtfM homologues) that belong to the Omp85 family in which the function of YtfM proteins is unknown. The YtfM homologue in *C. crescentus* is referred to as Omp68. Using bioinformatics, it was demonstrated that relative to TpsB proteins, Omp68 homologues are more closely related to BamA homologues. Furthermore, like BamA and the eukaryotic homologue of BamA (Sam50), the C-terminal POTRA domain that is most connected to the β-barrel domain is also conserved in Omp68. The tight evolutionary constraints placed on this POTRA domain is related to the fact that it is essential for cell survival and its deletion disrupts the BAM complex (72). In contrast, the C-terminal POTRA domain of TpsB proteins are highly divergent. The comparative analysis proposes that Omp68 and its homologues may represent an evolutionary intermediate of BamA that gave rise to TpsB proteins. POTRA truncation was also a lineage-specific feature with BamA having the most number of POTRA domains, possibly advantageous for providing several binding sites for a large number of substrates. Like TpsB and BamA proteins, Omp68 is also able to form a protein complex in the outer membrane but the composition of this complex remains unknown. The identification of the P150 protein, which is encoded by a gene in the same operon as *omp68*, provides a basis for a hetero-protein complex. P150 does not resemble the large TpsA proteins that are secreted by its cognate TpsB proteins. It has a motif and a domain that it shares with AsmA, an inner membrane protein that interacts with OmpF in *E. coli.* By analogy, Omp68 could form a hetero-complex that could span the inner and outer membrane to serve a particular biological function.

The evolution of the BAM complex in terms of its lipoprotein subunit content was also examined by bioinformatics analysis to detect lineage-specific subunits. It is known that

the BAM complex in *E. coli* has four different lipoproteins (BamB, BamC, BamD and BamE) that associate with the core subunit, BamA. Across α-, β- and γ-proteobacteria, it was shown that all α-proteobacteria do not have a gene encoding BamC while *Brucella* from this class also lacks BamB. In *N. meningitidis* from the β-class, there was no sequence homologue of BamB. In contrast, BamD has co-evolved with BamA such that these two proteins are encoded by the same genome from different proteobacteria. Biochemical characterisation of the complex revealed that in the model α-proteobacterium, *C. crescentus*, the predicted subunits (BamB, BamD, BamE) were present in the BAM complex as well as three novel subunits (Pal, BamF and BamG). The BAM complex is also large (500 kDa) and modular. Pal was shown to be an essential subunit that interacts with peptidoglycan. Depletion of Pal resulted in destabilisation of the cell envelope, shown by membrane vesicles that had formed at the cell surface. The physiological relevance of Pal in the BAM complex may be related to spatial organisation of the complex at the outer membrane for efficient uptake and assembly of β-barrel substrates. Pal may reduce the periplasmic distance between the outer membrane and the peptidoglycan layer and tether the BAM complex so that subunits like BamA are able to span the periplasm at sites where proteins emerge from the Sec machinery. Another subunit that was also identified and confirmed to be a genuine subunit of the BAM complex is BamF. BamF is a non-essential lipoprotein that was shown as a possible homologue of BamC by sharing an N-terminal conserved region with many BamC homologues. This motif was stringently conserved amongst BamF homologues in all α-proteobacteria. Conservation of the N-terminal motif and the presence of exposed residues suggest it is a functional site for interaction with the BAM complex and provides a basis for directed mutagenesis to reveal important residues that could interact with an outer membrane component such as BamA. The other novel subunit in the BAM complex of *C. crescentus* is BamG, which is shown to be less conserved as it is exclusively found in four other types of aquatic α-proteobacteria. BamG may have a role in efficiently assembling a high number of outer membrane proteins involved in nutrient uptake in dilute conditions that is the environment for these five different α-proteobacteria.

The general topology of outer membrane lipoproteins has not been experimentally validated before but it is assumed that they are anchored to the inner leaflet of the outer membrane. The topologies of the lipoproteins (BamD, Pal and BamF) were determined by a protease shaving assay to demonstrate that these lipoproteins indeed face the periplasmic space. Curiously, the periplasmic portion of BamA and Omp68 are protease-resistant using the same conditions.

The high conservation and detection of TPR motifs in BamD made it an interesting subject for further experimental analysis. Obtaining the structure of BamD was one of the prime goals and extensive efforts were made to obtain soluble protein for structural analysis. Eventually, BamD was stabilised with low concentration of Urea as it was resistant to aggregation or misfolding using this approach. Tryptophan spectroscopy and circular dichroism showed BamD adopts a folded α-helical structure using this approach. Efforts by other competitive research groups determined the high-resolution structure of BamD from *R. marinus* and *E. coli* having five TPR motifs. The N-terminal TPR motifs form an interaction interface that resembles that of other TPR-rich proteins including Hop, Tom70 and FKBP52 that all interact with the extended C-terminal tail of Hsp70 and Hsp90 (216). This analogy predicts BamD as a docking site for periplasmic chaperones to transfer its bound β-barrel substrate to the BAM complex. Furthermore, the TPR array of BamD is shown as an extended conformation that could be important for interacting with substrates across the periplasmic distance, similar to the protruding POTRA domains of BamA, which can also extend up to 10 nm into the periplasm.

References

1. Van Wielink JE, Duine JA (1990) How big is the periplasmic space? *Trends Biochem. Sci.* 15(4):136-7.
2. Ruiz N, Kahne D, Silhavy TJ (2006) Advances in understanding bacterial outer-membrane biogenesis. *Nat. Rev. Microbiol.* 4(1).
3. Pohlschröder M, Prinz WA, Hartmann E, Beckwith J (1997) Protein translocation in the three domains of life: variations on a theme. *Cell.* 91(5):563-6.
4. Joly JC, Leonard MR, Wickner WT (1994) Subunit dynamics in Escherichia coli preprotein translocase. *Proc. Natl. Acad. Sci. USA.* 91(11):4703-7.
5. Robinson C, Bolhuis A (2004) Tat-dependent protein targeting in prokaryotes and chloroplasts. *Biochim. Biophys. Acta.* 1694(1-3):135-47.
6. Koronakis V et al. (2000) Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export. *Nature.* 405(6789):914-9.
7. Whitfield C (2006) Biosynthesis and assembly of capsular polysaccharides in Escherichia coli. *Annu. Rev. Biochem.* 75(39-68).
8. Raetz CR, Whitfield C (2002) Lipopolysaccharide endotoxins. *Annu Rev. Biochem.* 71:635-700.
9. Tokuda H, Matsuyama S (2004) Sorting of lipoproteins to the outer membrane in E. coli. *Biochim. Biophys. Acta.* 1694(1-3):IN 1-9.
10. Contreras I, Shapiro L, Henry S (1978) Membrane phospholipid composition of Caulobacter crescentus. *J. Bacteriol.* 135(3):1130-6.
11. Bos MP, Robert V, Tommassen J (2007) Biogenesis of the gram-negative bacterial outer membrane. *Annu Rev Microbiol* 61:191-214.
12. Chng SS, Gronenberg LS, Kahne D (2010) Proteins required for lipopolysaccharide assembly in Escherichia coli form a transenvelope complex. *Biochemistry.* 49(22):4565-7.
13. Nikaido H (2003) Molecular basis of bacterial outer membrane permeability revisited. *Microbiol. Mol. Biol. Rev.* 593-656(67):4.
14. Vaara M (1993) Antibiotic-supersusceptible mutants of Escherichia coli and Salmonella typhimurium. *Antimicrob. Agents Chemother.* 37(11):2255-60.
15. Vuorio R, Vaara M (1992) The lipid A biosynthesis mutation lpxA2 of Escherichia coli results in drastic antibiotic supersusceptibility. *Antimicrob Agents Chemother.* 36(4):826-9.
16. Gatsos X et al. (2008) Protein secretion and outer membrane assembly in Alphaproteobacteria. *FEMS Microbiol. Rev.* 32(6):995-1009.
17. Sauer FG, Remaut H, Hultgren SJ, Waksman G (2004) Fiber assembly by the chaperone-usher pathway. *Biochim. Biophys. Acta.* 1694(1-3):259-67.
18. Schulz GE (2002) The structure of bacterial outer membrane proteins. *Biochim. Biophys. Acta.* 1565(2):308-17.

19. Haake DA (2000) Spirochaetal lipoproteins and pathogenesis. *Microbiology.* 146 (7):1491-504.
20. Pugsley AP (1993) The complete general secretory pathway in Gram-negative bacteria. *Microbiol. Rev.* 57(1):50-108.
21. van Ulsen P, Tommassen J (2006) Protein secretion and secreted proteins in pathogenic Neisseriaceae. *FEMS Microbiol. Rev.* 30(2):292-319.
22. Delgado M et al. (2007) Lipoprotein NMB0928 from Neisseria meningitidis serogroup B as a novel vaccine candidate. *Vaccine.* 25(50):8420-31.
23. Suzuki H et al. (1978) Murein-lipoprotein of Escherichia coli: a protein involved in the stabilization of bacterial cell envelope. *Mol. Gen. Genet.* 167(1):1-9.
24. Yem DW, Wu HC (1978) Physiological characterization of an Escherichia coli mutant altered in the structure of murein lipoprotein. *J. Bacteriol.* 133(3):1419–1426.
25. Babu MM et al. (2006) A database of bacterial lipoproteins (DOLOP) with functional assignments to predicted lipoproteins. *J Bacteriol.* 188(8):2761-73.
26. Fadl AA et al. (2005) Murein lipoprotein is a critical outer membrane component involved in Salmonella enterica serovar typhimurium systemic infection. *Infect. Immun.* 73(2):1081-96.
27. Cascales E et al. (2002) Pal lipoprotein of Escherichia coli plays a major role in outer membrane integrity. *J. Bacteriol.* 184(3):754-9.
28. Wu T et al. (2005) Identification of a multicomponent complex required for outer membrane biogenesis in Escherichia coli. *Cell* 121(2):235-45.
29. Sklar JG et al. (2007) Lipoprotein SmpA is a component of the YaeT complex that assembles outer membrane proteins in Escherichia coli. *Proc Natl Acad Sci U S A* 104(15):6400-5.
30. Wimley WC (2003) The versatile beta-barrel membrane protein. *Curr. Opin. Struct. Biol.* 13(4):404-11.
31. Bishop CM, Walkenhorst WF, Wimley WC (2001) Folding of beta-sheets in membranes: specificity and promiscuity in peptide model systems. *J. Mol. Biol.* 309(4):975-88.
32. Mori H, Ito K (2001) The Sec protein-translocation pathway. *Trends Microbiol.* 9(10):494-500.
33. Koebnik R, Locher KP, Van Gelder P (2000) Structure and function of bacterial outer membrane proteins: barrels in a nutshell. *Mol. Microbiol.* 37(2):239-53.
34. Braun V, Wu HC (1994) Lipoproteins: structure, function, biosynthesis and model for protein export. *New Compr. Biochem.* 27:319-341.
35. Torres AG et al. (2006) Outer membrane protein A of Escherichia coli O157:H7 stimulates dendritic cell activation. *Infect. Immun.* 74(5):2676-85.
36. Clantin B et al. (2007) Structure of the membrane protein FhaC: a member of the Omp85-TpsB transporter superfamily. *Science.* 317(5840):957-61.
37. Whitfield C, Roberts IS (1999) Structure, assembly and regulation of expression of capsules in Escherichia coli. *Mol. Microbiol.* 31(5):1307-19.
38. Jann B, Jann K (1997) Structure and biosynthesis of the capsular antigens of Escherichia coli. *Curr. Top Microbiol. Immunol.* 1990:19–42.
39. Sleytr UB, Beveridge TJ (1999) Bacterial S-layers. *Trends Microbiol.* 7(6):253-60.

40. Mesnage S et al. (1998) The capsule and S-layer: two independent and yet compatible macromolecular structures in Bacillus anthracis. *J. Bacteriol.* 180(1):52-8.
41. Narita S, Matsuyama S, Tokuda H (2004) Lipoprotein trafficking in Escherichia coli. *Arch. Microbiol.* 182(1):1-6.
42. Yamaguchi K, Yu F, Inouye M (1988) A single amino acid determinant of the membrane localization of lipoproteins in E. coli. *Cell.* 53(3):423-32.
43. Seydel A, Gounon P, Pugsley AP (1999) Testing the '+2 rule' for lipoprotein sorting in the Escherichia coli cell envelope with a new genetic selection. *Mol. Microbiol.* 34(4):810-21.
44. Terada M, Kuroda T, Matsuyama SI, Tokuda H (2001) Lipoprotein sorting signals evaluated as the LolA-dependent release of lipoproteins from the cytoplasmic membrane of Escherichia coli. *J. Biol. Chem.* 276(50):47690-4.
45. Masuda K, Matsuyama S, Tokuda H (2002) Elucidation of the function of lipoprotein-sorting signals that determine membrane localization. *Proc. Natl. Acad. Sci. USA*. 99(11):7390-5.
46. Yakushi T et al. (2000) A new ABC transporter mediating the detachment of lipid-modified proteins from membranes. *Nat. Cell Biol.* 2(4):212-8.
47. Narita S, Tanaka K, Matsuyama S, Tokuda H (2002) Disruption of LolCDE, encoding an ATP-binding cassette transporter, is lethal for Escherichia coli and prevents release of lipoproteins from the inner membrane. *J. Bacteriol.* 184(5):1417-22.
48. Bernstein HD (2000) The biogenesis and assembly of bacterial membrane proteins. *Curr. Opin. Microbiol.* 3(2):203-9.
49. Veenendaal AK, van der Does C, Driessen AJ (2001) Mapping the sites of interaction between SecY and SecE by cysteine scanning mutagenesis. *J. Biol. Chem.* 276(35):32559-66.
50. Renu Tuteja (2005) Type I signal peptidase. *Arch. Biochem. Biophys.* 441(2):107-111.
51. du Plessis DJ, Nouwen N, Driessen AJ (2010) The Sec translocase. *Biochim. Biophys. Acta.* 1808(3):851-65.
52. Robert V et al. (2006) Assembly factor Omp85 recognizes its outer membrane protein substrates by a species-specific C-terminal motif. *PLoS Biol.* 4(11):e377.
53. Sklar JG, Wu T, Kahne D, Silhavy TJ (2007) Defining the roles of the periplasmic chaperones SurA, Skp, and DegP in Escherichia coli. *Genes Dev.* 21(19):2473-84.
54. Mogensen JE, Otzen DE (2005) Interactions between folding factors and bacterial outer membrane proteins. *Mol. Microbiol.* 57(2):326-46.
55. De Cock H et al. (1999) Affinity of the periplasmic chaperone Skp of Escherichia coli for phospholipids, lipopolysaccharides and non-native outer membrane proteins. Role of Skp in the biogenesis of outer membrane protein. *Eur. J. Biochem.* 259(1-2):96-103.
56. Schäfer U, Beck K, Müller M (1999) Skp, a molecular chaperone of gram-negative bacteria, is required for the formation of soluble periplasmic intermediates of outer membrane proteins. *J. Biol. Chem.* 274(35):24567-74.

57. Harms N et al. (2001) The early interaction of the outer membrane protein phoe with the periplasmic chaperone Skp occurs at the cytoplasmic membrane. *J. Biol. Chem.* 276(22):18804-11.
58. Ureta AR, Endres RG, Wingreen NS, Silhavy TJ (2007) Kinetic analysis of the assembly of the outer membrane protein LamB in Escherichia coli mutants each lacking a secretion or targeting factor in a different cellular compartment. *J. Bacteriol.* 189(2):446-54.
59. Rizzitello AE, Harper JR, Silhavy TJ (2001) Genetic evidence for parallel pathways of chaperone activity in the periplasm of Escherichia coli. *J. Bacteriol.* 183(23):6794-800.
60. Surrey T, Jähnig F (1992) Refolding and oriented insertion of a membrane protein into a lipid bilayer. *Proc. Natl. Acad. Sci. USA.* 89(16):7457-61.
61. Kleinschmidt JH, Wiener MC, Tamm LK (1999) Outer membrane protein A of E. coli folds into detergent micelles, but not in the presence of monomeric detergent. *Protein Sci.* 8(10):2065-71.
62. Conlan S, Bayley H (2003) Folding of a monomeric porin, OmpG, in detergent solution. *Biochemistry.* 42(31):9453-65.
63. Bitto E, McKay DB (2002) Crystallographic structure of SurA, a molecular chaperone that facilitates folding of outer membrane porins. *Structure.* 10(11):1489-98.
64. Bitto E, McKay DB (2003) The periplasmic molecular chaperone protein SurA binds a peptide motif that is characteristic of integral outer membrane proteins. *J. Biol. Chem.* 278(49):49316-22.
65. Eppens EF, Nouwen N, Tommassen J (1997) Folding of a bacterial outer membrane protein during passage through the periplasm. *EMBO J.* 16(14):4295-301.
66. Malinverni JC et al. (2006) YfiO stabilizes the YaeT complex and is essential for outer membrane protein assembly in Escherichia coli. *Mol Microbiol* 61(1):151-64.
67. Voulhoux R et al. (2003) Role of a highly conserved bacterial protein in outer membrane protein assembly. *Science* 299(5604):262-5.
68. Voulhoux R, Tommassen J (2004) Omp85, an evolutionarily conserved bacterial protein involved in outer-membrane-protein assembly. *Res. Microbiol.* 155(3):129-35.
69. Sánchez-Pulido L et al. (2003) POTRA: a conserved domain in the FtsQ family and a class of beta-barrel outer membrane proteins. *Trends Biochem Sci* 28(10):523-6.
70. Stegmeier JF et al. (2007) Characterisation of YtfM, a second member of the Omp85 family in Escherichia coli. *Biol. Chem.* 388(1):37-46.
71. Knowles TJ et al. (2008) Fold and function of polypeptide transport-associated domains responsible for delivering unfolded proteins to membranes. *Mol. Microbiol.* 68(5):1216-27.
72. Kim S et al. (2007) Structure and function of an essential component of the outer membrane protein assembly machine. *Science.* 317(5840):961-4.

73. Gatzeva-Topalova PZ, Warner LR, Pardi A, Sousa MC (2010) Structure and flexibility of the complete periplasmic domain of BamA: the protein insertion machine of the outer membrane. *Structure.* 18(11):1492-501.
74. Bos MP, Robert V, Tommassen J (2007) Functioning of outer membrane protein assembly factor Omp85 requires a single POTRA domain. *EMBO Rep.* 8(12):1149-54.
75. Paschen SA et al. (2003) Evolutionary conservation of biogenesis of b-barrel membrane proteins. *Nature.* 426(6968):862–866.
76. Surana NK et al. (2004) Evidence for conservation of architecture and physical properties of Omp85-like proteins throughout evolution. *Proc. Natl. Acad. Sci. USA.* 101(40):14497-502.
77. Hagan CL, Kim S, Kahne D (2010) Reconstitution of outer membrane protein assembly from purified components. *Science.* 328(5980):890-2.
78. Struyvé M, Moons M, Tommassen J (1991) Carboxy-terminal phenylalanine is essential for the correct assembly of a bacterial outer membrane protein. *J. Mol. Biol.* 218(1):141-8.
79. Kutik S et al. (2008) Dissecting membrane insertion of mitochondrial beta-barrel proteins. *Cell.* 132(6):1011-24.
80. Knowles TJ, Scott-Tucker A, Overduin M, Henderson IR (2009) Membrane protein architects: the role of the BAM complex in outer membrane protein assembly. *Nat. Rev. Microbiol.* 7(3):206-14.
81. Tommassen J (2007) Biochemistry: Getting into and through the outer membrane. *Science.* 317(5840):903-4.
82. Gentle IE, Burri L, Lithgow T (2005) Molecular architecture and function of the Omp85 family of proteins. *Mol. Microbiol.* 58(5):1216-25.
83. Stroud DA, Meisinger C, Pfanner N, Wiedemann N (2010) Assembling the outer membrane. *Science.* 328(5980):831-2.
84. Gentle IE et al. (2004) The Omp85 family of proteins is essential for outer membrane biogenesis in mitochondria and bacteria. *J. Cell Biol.* 164(1):19-24.
85. Kozjak V et al. (2003) An essential role of Sam50 in the protein sorting and assembly machinery of the mitochondrial outer membrane. *J. Biol. Chem.* 278(49):48520-3.
86. Jacob-Dubuisson F, Fernandez R, Coutte L (2004) Protein secretion through autotransporter and two-partner pathways. *Biochim. Biophys. Acta.* 1694(1-3):235-57.
87. Vuong P et al. (2008) Analysis of YfgL and YaeT interactions through bioinformatics, mutagenesis, and biochemistry. *J. Bacteriol.* 190(5):1507-17.
88. Ruiz N, Falcone B, Kahne D, Silhavy TJ (2005) Chemical conditionality: a genetic strategy to probe organelle assembly. *Cell* 121(2):307-17.
89. Onufryk C, Crouch M, Fang F, Gross C (2005) Characterisation of six lipoproteins in the SigmaE regulon. *J Bacteriol* 187(13):4552-4561.
90. Jawad Z, Paoli M (2002) Novel sequences propel familiar folds. *Structure.* 10(4):447-54.
91. Kim KH, Paetzel M (2010) Crystal Structure of Escherichia coli BamB, a Lipoprotein Component of the β-Barrel Assembly Machinery Complex. *J. Mol. Biol.* 406(5):667-78.

92. Fussenegger M, Facius D, Meier J, Meyer TF (1996) A novel peptidoglycan-linked lipoprotein (ComL) that functions in natural transformation competence of Neisseria gonorrhoeae. *Mol. Microbiol.* 19(5):1095-105.
93. Fussenegger M, Facius D, Meier J, Meyer TF (1996) A novel peptidoglycan-linked lipoprotein (ComL) that functions in natural transformation competence of Neisseria gonorrhoeae. *Mol Microbiol* 19(5):1095-105.
94. Blatch GL, Lassle M (1999) The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. *Bioessays.* 21(11):932-9.
95. D'Andrea LD, Regan L (2003) TPR proteins: the versatile helix. *Trends Biochem Sci* 28(12):655-62.
96. Chan NC et al. (2006) The C-terminal TPR domain of Tom70 defines a family of mitochondrial protein import receptors found only in animals and fungi. *J Mol Biol* 358(4):1010-22.
97. Wu Y, Sha B (2006) Crystal structure of yeast mitochondrial outer membrane translocon member Tom70p. *Nat Struct Mol Biol* 13(7):589-93.
98. Vanini MM et al. (2008) The solution structure of the outer membrane lipoprotein OmlA from Xanthomonas axonopodis pv. citri reveals a protein fold implicated in protein-protein interaction. *Proteins.* 71(4):2051-64.
99. Knowles TJ et al. (2011) Structure and function of BamE within the outer membrane and the β-barrel assembly machine. *EMBO Rep.* 12(2):123-8.
100. Kim KH et al. (2011) Structural Characterization of Escherichia coli BamE, a Lipoprotein Component of the β-Barrel Assembly Machinery Complex. *Biochemistry.* 50(6):1081-90.
101. Ochsner UA, Vasil AI, Johnson Z, Vasil ML (1999) Pseudomonas aeruginosa fur overlaps with a gene encoding a novel outer membrane lipoprotein, OmlA. *J. Bacteriol.* 181(4):1099-1109.
102. Fuangthong M et al. (2008) The omlA gene is involved in multidrug resistance and its expression is inhibited by coumarins in Xanthomonas campestris pv. phaseoli. *Arch. Microbiol.* 189(3):211-8.
103. Tommassen J (1986) Fallacies of E. coli cell fractionations and consequences thereof for protein export models. *Microb. Pathog.* 1(3):225-8.
104. Bayer ME (1968) Areas of adhesion between wall and membrane of Escherichia coli. *J. Gen. Microbiol.* 53(3):395-404.
105. Smit J, Nikaido H (1978) Outer membrane of gram-negative bacteria. XVIII. Electron microscopic studies on porin insertion sites and growth of cell surface of Salmonella typhimurium. *J. Bacteriol.* 135(2):687-702.
106. Bayer ME (1991) Zones of membrane adhesion in the cryofixed envelope of Escherichia coli. *J. Struct. Biol.* 107(3):268-80.
107. Walderich B, Höltje JV (1989) Specific localization of the lysis protein of bacteriophage MS2 in membrane adhesion sites of Escherichia coli. *J. Bacteriol.* 171(6):3331-6.
108. Bayer ME, Bayer MH, Lunn CA, Pigiet V (1987) Association of thioredoxin with the inner membrane and adhesion sites in Escherichia coli. *J. Bacteriol.* 169(6):2659-66.
109. Schatz G, Butow RA (1983) How are proteins imported into mitochondria? *Cell.* 32(2):316–318.

110. Schwaiger M, Herzog V, Neupert W (1987) Characterization of translocation contact sites involved in the import of mitochondrial proteins. *J. Cell Biol.* 105(1):235-46.
111. Pain D, Kanwar YS, Blobel G (1988) Identification of a receptor for protein import into chloroplasts and its localization to envelope contact zones. *Nature.* 331(6153):232-7.
112. Kellenberger E (1990) The 'Bayer bridges' confronted with results from improved electron microscopy methods. *Mol. Microbiol.* 4(5):697-705.
113. Metcalfe M, Holland IB (1980) Synthesis of a major outer membrane porin by Escherichia coli spheroplasts. *FEMS Microbial. Lett.* 7(2):111-4.
114. Dalbey RE, Wickner W (1988) Characterization of the internal signal-anchor domain of Escherichia coli leader peptidase. *J. Biol. Chem.* 263(1):404-408.
115. Nikaido H (1992) Porins and specific channels of bacterial outer membranes. *Mol. Microbiol.* 6(4):435-42.
116. Nikaido H, Reid J (1990) Biogenesis of prokaryotic pores. *Experientia.* 46(2):174-80.
117. Bosch D et al. (1986) Periplasmic accumulation of truncated forms of outer-membrane PhoE protein of Escherichia coli K-12. *J. Mol. Biol.* 189(3):449-55.
118. Freudl R et al. (1985) The nature of information, required for export and sorting, present within the outer membrane protein OmpA of Escherichia coli K-12. *EMBO J.* 4(13A):3593-8.
119. Nikaido H. *Biosynthesis and assembly of lipopolysaccharides and the outer membrane layer of Gram-negative cell wall. In Bacterial Membranes and Walls.* Dekker, New York, 1973.
120. Mühlradt PF, Menzel J, Golecki JR, Speth V (1973) Outer membrane of salmonella. Sites of export of newly synthesised lipopolysaccharide on the bacterial surface. *Eur J. Biochem.* 35(3):471-81.
121. Roberts IS (1996) The biochemistry and genetics of capsular polysaccharide production in bacteria. *Annu Rev Microbiol.* 50:285-315.
122. Jann B, Jann K. *Capsules of Escherichia coli.* Cambridge University Press, Cambridge, 1997.
123. Arrecubieta C et al. (2001) The transport of group 2 capsular polysaccharides across the periplasmic space in Escherichia coli. Roles for the KpsE and KpsD proteins. *J. Biol. Chem.* 276(6):4245-50.
124. Ishidate K et al. (1986) Isolation of differentiated membrane domains from Escherichia coli and Salmonella typhimurium, including a fraction containing attachment sites between the inner and outer membranes and the murein skeleton of the cell envelope. *J. Biol. Chem.* 261(1):428-43.
125. Bayer MH, Costello GP, Bayer ME (1982) Isolation and partial characterization of membrane vesicles carrying markers of the membrane adhesion sites. *J. Bacteriol.* 149(2):758-67.
126. Tefsen B et al. (2005) Lipopolysaccharide transport to the bacterial outer membrane in spheroplasts. *J. Biol. Chem.* 280(6):4504-9.
127. Dijkstra AJ, Keck W (1996) Peptidoglycan as a barrier to transenvelope transport. *J. Bacteriol.* 178(19):5555-5562.

128. Demchick P, Koch AL (1996) The permeability of the wall fabric of Escherichia coli and Bacillus subtilis. *J. Bacteriol.* 178(3):768–773.
129. Gentle I et al. (2004) The Omp85 family of proteins is essential for outer membrane biogenesis in mitochondria and bacteria. *J Cell Biol* 164(1):19-24.
130. Sánchez-Pulido L et al. (2003) POTRA: a conserved domain in the FtsQ family and a class of beta-barrel outer membrane proteins. *Trends Biochem. Sci.* 28(10):523-6.
131. Yen MR et al. (2002) Protein-translocating outer membrane porins of Gram-negative bacteria. *Biochim. Biophys. Acta.* 1562(1-2):6-31.
132. Schleiff E, Soll J (2005) Membrane protein insertion: mixing eukaryotic and prokaryotic concepts. *EMBO J. Rep.* 6(11):1023-7.
133. Voulhoux R et al. (2003) Role of a highly conserved bacterial protein in outer membrane protein assembly. *Science.* 299(5604):262-5.
134. Werner J, R. M (2005) YaeT (Omp85) affects the assembly of lipid-dependent and lipid-independent outer membrane proteins of Escherichia coli. *Mol. Microbiol.* 57(5):1450-9.
135. Doerrler WT, Raetz CR (2005) Loss of outer membrane proteins without inhibition of lipid export in an Escherichia coli YaeT mutant. *J. Biol. Chem.* 280(30):27679-87.
136. Sklar JG et al. (2007) Lipoprotein SmpA is a component of the YaeT complex that assembles outer membrane proteins in Escherichia coli. *Proc. Natl. Acad. Sci. USA* 104(15):6400-5.
137. Wu T et al. (2005) Identification of a multicomponent complex required for outer membrane biogenesis in Escherichia coli. *Cell.* 121(2):235-45.
138. Hinnah SC et al. (1997) Reconstitution of a chloroplast protein import channel. *EMBO J.* 16(24):7351-60.
139. Mazar J, Cotter PA (2007) New insight into the molecular mechanisms of two-partner secretion. *Trends Microbiol.* 15(11):508-15.
140. Frickey T, Lupas A (2004) CLANS: a Java application for visualizing protein families based on pairwise similarity. *Bioinformatics.* 20(18):3702-4.
141. Arnold T, Zeth K, Linke D (2010) Omp85 from the thermophilic cyanobacterium Thermosynechococcus elongatus differs from proteobacterial Omp85 in structure and domain composition. *J. Biol. Chem.* 285(23):18003-15.
142. Bendtsen JD, Nielsen H, von Heijne G, Brunak S (2004) Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.* 340:783–795.
143. Higgins D et al. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positionspecific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22:4673– 4680.
144. McDonnell AV et al. (2006) Prediction and comparative modeling of sequences directing beta-sheet proteins by profile wrapping. *Proteins* 63:976-985.
145. Jones DT (1999) Protein secondary prediction based on position-specific scoring matrices. *J. Mol. Biol.* 292:195– 202.
146. Soohammer EL, Eddy SR, Durbin R (1997) Pfam: a comprehensive database of protein domain families based on seed alignments. *Proteins.* 28:405–420.
147. Servant F et al. (2002) ProDom: automated clustering of homologous domains. *Brief Bioinform.* 3:246– 251.

148. Eddy SR (1998) Profile hidden Markov models. *Bioinformatics.* 14(9):755-63.
149. Dolezal P, Likic V, Tachezy J, Lithgow T (2006) Evolution of the molecular machines for protein import into mitochondria. *Science.* 313(5785):314-8.
150. Ely B (1991) Genetics of Caulobacter crescentus. *Methods Enzymol.* 204:372-384.
151. Lugtenberg B et al. (1975) Electrophoretic resolution of the "major outer membrane protein" of Escherichia coli K12 into four bands. *FEBS Lett.* 58(1):254-8.
152. Stenberg F et al. (2005) Protein complexes of the Escherichia coli cell envelope. *J. Biol. Chem.* 280(41):34409-19.
153. Schägger H, von Jagow G (1991) Blue native electrophoresis for isolation of membrane protein complexes in enzymatically active form. *Anal. Biochem.* 199(2):223-231.
154. Purcell AW et al. (2001) Quantitative and qualitative influences of tapasin on the class I peptide repertoire. *J. Immunol.* 166(2):1016-27.
155. Beher MG, Schnaitman CA, Pugsley AP (1980) Major heat-modifiable outer membrane protein in gram-negative bacteria: comparison with the ompA protein of Escherichia coli. *J. Bacteriol.* 143(2):906-13.
156. Tagawa Y, Haritani M, Ishikawa H, Yuasa N (1993) Characterization of a heat-modifiable outer membrane protein of Haemophilus somnus. *Infect. Immun.* 61(5):1750-5.
157. Habib SJ et al. (2007) The N-terminal domain of Tob55 has a receptor-like function in the biogenesis of mitochondrial β-barrel proteins. *J. Cell Biol.* 176(1):77-88.
158. Ertel F et al. (2005) The evolutionarily related beta-barrel polypeptide transporters from Pisum sativum and Nostoc PCC7120 contain two distinct functional domains. *J. Biol. Chem.* 280(31):28281-9.
159. Koenig P et al. (2010) Conserved properties of polypeptide transport-associated (POTRA) domains derived from cyanobacterial Omp85. *J. Biol. Chem.* 285(23):18016-24.
160. Nierman WC et al. (2001) Complete genome sequence of Caulobacter crescentus. *Proc. Natl. Acad. Sci. USA.* 98(7):4136-41.
161. Phadke ND et al. (2001) Analysis of the outer membrane proteome of Caulobacter crescentus by two-dimensional electrophoresis and mass spectrometry. *Proteomics.* 1(5):705-20.
162. Neugebauer H et al. (2005) ExbBD-dependent transport of maltodextrins through the novel MalA protein across the outer membrane of Caulobacter crescentus. *J. Bacteriol.* 187(24):8300-11.
163. Agabian N, Unger B (1978) Caulobacter crescentus cell envelope: effect of growth conditions on murein and outer membrane protein composition. *J. Bacteriol.* 133(2):987-94.
164. Clantin B et al. (2004) The crystal structure of filamentous hemagglutinin secretion domain and its implications for the two-partner secretion pathway. *Proc Natl Acad Sci U S A* 101(16):6194-9.

165. Misra R, Miao Y (1995) Molecular analysis of asmA, a locus identified as the suppressor of OmpF assembly mutants of Escherichia coli K-12. *Mol. Microbiol.* 16(4):779-788.
166. Misra R (1993) OmpF assembly mutants of Escherichia coli K-12: isolation, characterization, and suppressor analysis. . *J. Bacteriol.* 175(16):5049-56.
167. Xiong X, Deeter JN, Misra R (1996) Assembly-defective OmpC mutants of Escherichia coli K-12. *J. Bacteriol.* 178(4):1213-5.
168. Cavalier-Smith T (2006) Rooting the tree of life by transition analyses. *Biol. Direct.* 1:19.
169. Pfanner N, Wiedemann N, Meisinger C, Lithgow T (2004) Assembling the mitochondrial outer membrane. *Nat. Struct. Mol. Biol.* 11(11):1044-8.
170. Chan NC, Lithgow T (2008) The peripheral membrane subunits of the SAM complex function codependently in mitochondrial outer membrane biogenesis. *Mol. Biol. Cell* 19(1):126-36.
171. Becker T et al. (2008) Biogenesis of the mitochondrial TOM complex: Mim1 promotes insertion and assembly of signal-anchored receptors. *J. Biol. Chem.* 283(1):120-127.
172. Hulett JM et al. (2008) The transmembrane segment of Tom20 is recognized by Mim1 for docking to the mitochondrial TOM complex. *J. Mol. Biol.* 376(3):694-704.
173. Lueder F, Lithgow T (2009) The three domains of the mitochondrial outer membrane protein Mim1 have discrete functions in assembly of the TOM complex. *FEBS Lett.* 583(9):1475-80.
174. Meisinger C et al. (2004) The mitochondrial morphology protein Mdm10 functions in assembly of the preprotein translocase of the outer membrane. *Dev. Cell* 7(1):61-71.
175. Meisinger C et al. (2006) Mitochondrial protein sorting: differentiation of beta-barrel assembly by Tom7-mediated segregation of Mdm10. *J. Biol. Chem.* 281(32):22819-26.
176. Popov-Celeketić J, Waizenegger T, Rapaport D (2008) Mim1 functions in an oligomeric form to facilitate the integration of Tom20 into the mitochondrial outer membrane. *J. Mol. Biol.* 376(3):671-680
177. Hoppins SC, Nargang FE (2004) The Tim8-Tim13 complex of Neurospora crassa functions in the assembly of proteins into both mitochondrial membranes. *J. Biol. Chem.* 279(13):12396-405.
178. Wiedemann N et al. (2004) Biogenesis of the protein import channel Tom40 of the mitochondrial outer membrane: intermembrane space components are involved in an early stage of the assembly pathway. *J. Biol. Chem.* 279(18):18188-94.
179. Martin W, Muller M (1998) The hydrogen hypothesis for the first eukaryote. *Nature.* 392(6671):37-41.
180. Kurland CG, Andersson SG (2000) Origin and evolution of the mitochondrial proteome. *Microbiol. Mol. Biol. Rev.* 64(4):786-820.
181. Emelyanov VV (2003) Common evolutionary origin of mitochondrial and rickettsial respiratory chains. *Arch. Biochem. Biophys.* 420(1):130-41.

182. Gray MW, Burger G, Lang BF (1999) Mitochondrial evolution. *Science.* 283(5407):1476-1481.
183. Clements A et al. (2009) The reducible complexity of a mitochondrial molecular machine. *Proc. Natl. Acad. Sci. USA.* 106(37):15791-5.
184. Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. . *Electrophoresis.* 18:2714-2723.
185. Waterhouse AM et al. (2009) Jalview Version 2 - a multiple sequence alignment editor and analysis workbench. *Bioinformatics.* 25(9):1189-1191.
186. Biegert A et al. (2006) The MPI Toolkit for protein sequence analysis. *Nucleic Acids Res.* 34(Web Server issue):W335-339.
187. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32(5):1792-7.
188. Bailey TL, Williams N, Misleh C, Li WW (2006) MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res.* 34:W369-73.
189. Berezin C et al. (2004) ConSeq: the identification of functionally and structurally important residues in protein sequences. *Bioinformatics.* 20(8):1322-4.
190. Bouveret E et al. (1995) Peptidoglycan-associated lipoprotein-TolB interaction. A possible key to explaining the formation of contact sites between the inner and outer membranes of Escherichia coli. *J. Biol. Chem.* 270(19):11071-7.
191. Parsons LM, Lin F, Orban J (2006) Peptidoglycan recognition by Pal, an outer membrane lipoprotein. *Biochemistry.* 45(7):2122-8.
192. Grizot S, Buchanan SK (2004) Structure of the OmpA-like domain of RmpM from Neisseria meningitidis. *Mol. Microbiol.* 51(4):1027-37.
193. Godlewska R, Wiśniewska K, Pietras Z, Jagusztyn-Krynicka EK (2009) Peptidoglycan-associated lipoprotein (Pal) of Gram-negative bacteria: function, structure, role in pathogenesis and potential application in immunoprophylaxis. *FEMS Microbiol. Lett.* 298(1):1-11.
194. Costa T, Priyadarshini R, Jacobs-Wagner C (2008) Localization of PBP3 in Caulobacter crescentus is highly dynamic and largely relies on its functional transpeptidase domain. *Mol. Microbiol.* 70(3):634-51.
195. Schmidt JM, Stanier RY (1966) The development of cellular stalks in bacteria. *J. Cell Biol.* 28(3):423-36.
196. Aaron M et al. (2007) The tubulin homologue FtsZ contributes to cell elongation by guiding cell wall precursor synthesis in Caulobacter crescentus. *Mol. Microbiol.* 64(4):938-52.
197. Cascales E et al. (2007) Colicin biology. *Microbiol. Mol. Biol. Rev.* 71(1):158-229.
198. Webster RE (1991) The tol gene products and the import of macromolecules into Escherichia coli. *Mol. Microbiol.* 5(5):1005-11.
199. Lazzaroni JC et al. (1995) Transmembrane alpha-helix interactions are required for the functional assembly of the Escherichia coli Tol complex. *J. Mol. Biol.* 246(1):1-7.
200. Clavel T et al. (1998) TolB protein of Escherichia coli K-12 interacts with the outer membrane peptidoglycan-associated proteins Pal, Lpp and OmpA. *Mol. Microbiol.* 29(1):359-67.

201. Bernadac A et al. (1998) Escherichia coli tol-pal mutants form outer membrane vesicles. *J. Bacteriol.* 180(18):4872-8.
202. Hellman J et al. (2002) Bacterial peptidoglycan-associated lipoprotein is released into the bloodstream in gram-negative sepsis and causes inflammation and death in mice. *J. Biol. Chem.* 277(16):14274-80.
203. Rolhion N, Barnich N, Claret L, Darfeuille-Michaud A (2005) Strong decrease in invasive ability and outer membrane vesicle release in Crohn's disease-associated adherent-invasive Escherichia coli strain LF82 with the yfgL gene deleted. *J. Bacteriol.* 187(7):2286-2296.
204. Onufryk C, Crouch ML, Fang FC, Gross CA (2005) Characterization of six lipoproteins in the sigmaE regulon. *J. Bacteriol.* 187(13):4552-61.
205. Cascales E, Lloubès R (2004) Deletion analyses of the peptidoglycan-associated lipoprotein Pal reveals three independent binding sequences including a TolA box. *Mol. Microbiol.* 51(3):873-85.
206. Jansen C et al. (2000) Biochemical and biophysical characterization of in vitro folded outer membrane porin PorA of Neisseria meningitidis. *Biochim. Biophys. Acta.* 1464(2):284-98.
207. Prinz T, Tommassen J (2000) Association of iron-regulated outer membrane proteins of Neisseria meningitidis with the RmpM (class 4) protein. *FEMS Microbiol. Lett.* 183(1):49-53.
208. Volokhina EB, Beckers F, Tommassen J, Bos MP (2009) The beta-barrel outer membrane protein assembly complex of Neisseria meningitidis. *J. Bacteriol.* 191(22):7074-85.
209. Tachikawa T, Kato J (2011) Suppression of the Temperature-Sensitive mutation of the bamD gene required for the assembly of outer membrane proteins by multicopy of the yiaD gene in Escherichia coli. *Biosci. Biotechnol. Biochem.* 75(1).
210. Koronakis V, Eswaran J, Hughes C (2004) Structure and function of TolC: the bacterial exit duct for proteins and drugs. *Annu. Rev. Biochem.* 73:467-89.
211. Desrosiers DC et al. (2011) TP0326, a Treponema pallidumβ-Barrel Assembly Machinery A (BamA) Ortholog and Rare Outer Membrane Protein. *Mol. Microbiol.* [Epub ahead of print].
212. Albrecht R, Zeth K (2010) Crystallization and preliminary X-ray data collection of the Escherichia coli lipoproteins BamC, BamD and BamE. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* 66(12):1586-90.
213. Malinverni JC et al. (2006) YfiO stabilizes the YaeT complex and is essential for outer membrane protein assembly in Escherichia coli. *Mol. Microbiol.* 61(1):151-64.
214. Albrecht R, Zeth K (2011) Structural basis of outer membrane protein biogenesis in bacteria. *J. Biol. Chem.*
215. D'Andrea LD, Regan L (2003) TPR proteins: the versatile helix. *Trends Biochem. Sci.* 28(12):655-62.
216. Sandoval CM et al. (2011) Crystal Structure of BamD: An Essential Component of the β-Barrel Assembly Machinery of Gram-Negative Bacteria. *J. Mol. Biol.* 409(3):348-57.

217. Aiyar A (2000) The use of CLUSTAL W and CLUSTAL X for multiple sequence alignment. *Methods Mol. Biol.* 132:221-41.
218. Bradford MM (1976) Rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem.* 72:248-54.
219. Sreerama N, Woody RW (2004) On the analysis of membrane protein circular dichroism spectra. *Protein Sci.* 13(1):100-12.
220. Karpenahalli MR, Lupas AN, Söding J (2007) TPRpred: a tool for prediction of TPR-, PPR- and SEL1-like repeats from protein sequences. *BMC Bioinformatics.* 8(2).
221. Fardini Y et al. (2007) The YfgL lipoprotein is essential for type III secretion system expression and virulence of Salmonella enterica Serovar Enteritidis. *Infect. Immun.* 75(1):358-370.
222. Zhou L, Srisatjaluk R, Justus DE, Doyle RJ (1998) On the origin of membrane vesicles in gram-negative bacteria. *FEMS Microbiol. Lett.* 163(2):223-8.
223. Wensink J, Witholt B (1981) Outer-membrane vesicles released by normally growing Escherichia coli contain very little lipoprotein. *Eur. J. Biochem.* 116(2):331-5.
224. Knox KW, Vesk M, Work E (1966) Relation between excreted lipopolysaccharide complexes and surface structures of a lysine-limited culture of Escherichia coli. *J. Bacteriol.* 92(4):1206-17.
225. Kadurugamuwa JL, Beveridge TJ (1995) Virulence factors are released from Pseudomonas aeruginosa in association with membrane vesicles during normal growth and exposure to gentamicin: a novel mechanism of enzyme secretion. *J. Bacteriol.* 177(14):3998-4008.
226. Chan NC et al. (2006) The C-terminal TPR domain of Tom70 defines a family of mitochondrial protein import receptors found only in animals and fungi. *J Mol Biol.* 358(4):1010-22.
227. Wu Y, Sha B (2006) Crystal structure of yeast mitochondrial outer membrane translocon member Tom70p. *Nat. Struct. Mol. Biol.* 13(7):589-93.
228. Young JC, Hoogenraad NJ, Hartl FU (2003) Molecular chaperones Hsp90 and Hsp70 deliver preproteins to the mitochondrial import receptor Tom70. *Cell.* 112(1):41-50.
229. Caplan AJ (2003) What is a co-chaperone? *Cell Stress Chaperones.* 8(2):105–107.
230. Andrade MA, Perez-Iratxeta C, Ponting CP (2001) Protein repeats: structures, functions, and evolution. *J. Struct. Biol.* 134(2-3):117-31.
231. Perry AJ et al. (2006) Convergent evolution of receptors for protein import into mitochondria. *Curr. Biol.* 16(3):221-9.

Appendix A1 – CLANS analysis

Sequences of full-length proteins from the Omp85 superfamily that were used for CLANS analysis are shown below and grouped as:

A) BamA
B) TpsB
C) YtfM-like
D) YtfM-like POTRA domains (sorted by position of the domains)

A) BamA

>gi|74316811|ref|YP_314551.1| surface antigen (D15) [Thiobacillus denitrificans ATCC 25259]
>gi|16126158|ref|NP_420722.1| outer membrane protein [Caulobacter crescentus CB15]
>gi|197105235|ref|YP_002130612.1| outer membrane protein [Phenylobacterium zucineum HLK1]
>gi|254420176|ref|ZP_05033900.1| outer membrane protein assembly complex, YaeT protein [Brevundimonas sp. BAL3]
>gi|241774459|ref|ZP_04771787.1| outer membrane protein assembly complex, YaeT protein [Asticcacaulis excentricus CB 48]
>gi|114797815|ref|YP_760480.1| OMP85 family outer membrane protein [Hyphomonas neptunium ATCC 15444]
>gi|114569938|ref|YP_756618.1| surface antigen (D15) [Maricaulis maris MCS10]
>gi|83858380|ref|ZP_00951902.1| outer membrane protein [Oceanicaulis alexandrii HTCC2633]
>gi|304320064|ref|YP_003853707.1| outer membrane protein [Parvularcula bermudensis HTCC2503]
>gi|154253627|ref|YP_001414451.1| surface antigen (D15) [Parvibaculum lavamentivorans DS-1]
>gi|254294071|ref|YP_003060094.1| outer membrane protein assembly complex, YaeT protein [Hirschia baltica ATCC 49814]
>gi|288958459|ref|YP_003448800.1| outer membrane protein [Azospirillum sp. B510]
>gi|209964512|ref|YP_002297427.1| outer membrane protein, putative [Rhodospirillum centenum SW]
>gi|144898240|emb|CAM75104.1| Bacterial surface antigen (D15) [Magnetospirillum gryphiswaldense MSR-1]
>gi|300023421|ref|YP_003756032.1| outer membrane protein assembly complex, YaeT protein [Hyphomicrobium denitrificans ATCC 51888]
>gi|298291815|ref|YP_003693754.1| outer membrane protein assembly complex, YaeT protein [Starkeya novella DSM 506]
>gi|158423325|ref|YP_001524617.1| putative outer membrane protein precursor [Azorhizobium caulinodans ORS 571]
>gi|94496934|ref|ZP_01303508.1| surface antigen (D15) [Sphingomonas sp. SKA58]
>gi|154248351|ref|YP_001419309.1| surface antigen (D15) [Xanthobacter autotrophicus Py2]
>gi|307293413|ref|ZP_07573259.1| outer membrane protein assembly complex, YaeT protein [Sphingobium chlorophenolicum L-1]
>gi|296536120|ref|ZP_06898250.1| outer membrane protein assembly complex [Roseomonas cervicalis ATCC 49957]
>gi|254469955|ref|ZP_05083360.1| outer membrane protein assembly complex, YaeT protein [Pseudovibrio sp. JE062]
>gi|118590003|ref|ZP_01547407.1| putative outer membrane protein [Stappia aggregata IAM 12614]

>gi|254502739|ref|ZP_05114890.1| outer membrane protein assembly complex, YaeT protein [Labrenzia alexandrii DFL-11]
>gi|103487436|ref|YP_616997.1| surface antigen (D15) [Sphingopyxis alaskensis RB2256]
>gi|296284734|ref|ZP_06862732.1| outer membrane protein [Citromicrobium bathyomarinum JL354]

>gi|85373590|ref|YP_457652.1| outer membrane protein [Erythrobacter litoralis HTCC2594]
>gi|87199398|ref|YP_496655.1| surface antigen (D15) [Novosphingobium aromaticivorans DSM 12444]
>gi|304391653|ref|ZP_07373595.1| outer membrane protein assembly complex, YaeT protein [Ahrensia sp. R2A130]
>gi|146341061|ref|YP_001206109.1| hypothetical protein BRADO4132 [Bradyrhizobium sp. ORS278]
>gi|115524568|ref|YP_781479.1| surface antigen (D15) [Rhodopseudomonas palustris BisA53]
>gi|114327604|ref|YP_744761.1| outer membrane protein assembly factor yaeT [Granulibacter bethesdensis CGDNIH1]
>gi|182677292|ref|YP_001831438.1| outer membrane protein assembly complex, YaeT protein [Beijerinckia indica subsp. indica ATCC 9039]
>gi|75676042|ref|YP_318463.1| Outer membrane protein [Nitrobacter winogradskyi Nb-255]
>gi|209885092|ref|YP_002288949.1| outer membrane protein assembly complex [Oligotropha carboxidovorans OM5]
>gi|217979934|ref|YP_002364081.1| outer membrane protein assembly complex, YaeT protein [Methylocella silvestris BL2]
>gi|299134993|ref|ZP_07028184.1| outer membrane protein assembly complex, YaeT protein [Afipia sp. 1NLS2]
>gi|260752423|ref|YP_003225316.1| outer membrane protein assembly complex, YaeT protein [Zymomonas mobilis subsp. mobilis NCIB 11163]
>gi|13470835|ref|NP_102404.1| outer membrane protein [Mesorhizobium loti MAFF303099]
>gi|148261429|ref|YP_001235556.1| surface antigen (D15) [Acidiphilium cryptum JF-5]
>gi|153009366|ref|YP_001370581.1| surface antigen (D15) [Ochrobactrum anthropi ATCC 49188]
>gi|239832041|ref|ZP_04680370.1| outer membrane protein assembly complex, YaeT protein [Ochrobactrum intermedium LMG 3301]
>gi|170749839|ref|YP_001756099.1| outer membrane protein assembly complex, YaeT protein [Methylobacterium radiotolerans JCM 2831]
>gi|254693862|ref|ZP_05155690.1| Bacterial surface antigen (D15) [Brucella abortus bv. 3 str. Tulya]
>gi|110633741|ref|YP_673949.1| surface antigen (D15) [Mesorhizobium sp. BNC1]
>gi|162147929|ref|YP_001602390.1| Outer membrane protein assembly factor yaeT precursor [Gluconacetobacter diazotrophicus PAl 5]
>gi|90419599|ref|ZP_01227509.1| outer membrane protein [Aurantimonas manganoxydans SI85-9A1]
>gi|296446138|ref|ZP_06888086.1| outer membrane protein assembly complex, YaeT protein [Methylosinus trichosporium OB3b]
>gi|117925145|ref|YP_865762.1| surface antigen (D15) [Magnetococcus sp. MC-1]
>gi|114704868|ref|ZP_01437776.1| outer membrane protein [Fulvimarina pela>gi HTCC2506]
>gi|239948526|ref|ZP_04700279.1| outer membrane protein assembly complex, YaeT protein [Rickettsia endosymbiont of Ixodes scapularis]
>gi|296115048|ref|ZP_06833690.1| outer membrane protein assembly complex, YaeT protein [Gluconacetobacter hansenii ATCC 23769]
>gi|77463263|ref|YP_352767.1| putative outer membrane protein [Rhodobacter sphaeroides 2.4.1]
>gi|114764265|ref|ZP_01443493.1| putative outer membrane protein [Pela>gibaca bermudensis HTCC2601]

>gi|15888707|ref|NP_354388.1| group 1 outer membrane protein precursor [Agrobacterium tumefaciens str. C58]
>gi|260427307|ref|ZP_05781286.1| outer membrane protein assembly complex, YaeT protein [Citreicella sp. SE45]
>gi|163760890|ref|ZP_02167969.1| putative outer membrane transmembrane protein [Hoeflea phototrophica DFL-43]
>gi|49474290|ref|YP_032332.1| Outer membrane protein [Bartonella quintana str. Toulouse]
>gi|258541752|ref|YP_003187185.1| outer membrane protein [Acetobacter pasteurianus IFO 3283-01]
>gi|294084080|ref|YP_003550838.1| surface antigen D15 [Candidatus Puniceispirillum marinum IMCC1322]
>gi|150396357|ref|YP_001326824.1| surface antigen (D15) [Sinorhizobium medicae WSM419]
>gi|126729712|ref|ZP_01745525.1| outer membrane protein, OMP85 family [Sa>gittula stellata E-37]
>gi|83953535|ref|ZP_00962256.1| outer membrane protein, OMP85 family protein [Sulfitobacter sp. NAS-14.1]
>gi|119386706|ref|YP_917761.1| surface antigen (D15) [Paracoccus denitrificans PD1222]
>gi|15965255|ref|NP_385608.1| putative outer membrane transmembrane protein [Sinorhizobium meliloti 1021]
>gi|240850310|ref|YP_002971703.1| outer membrane protein [Bartonella grahamii as4aup]
>gi|307317020|ref|ZP_07596461.1| outer membrane protein assembly complex, YaeT protein [Sinorhizobium meliloti AK83]
>gi|259416856|ref|ZP_05740776.1| outer membrane protein assembly complex, YaeT protein [Silicibacter sp. TrichCH4B]
>gi|1262291|gb|AAA96788.1| OMP1 precursor [Brucella abortus]
>gi|84500827|ref|ZP_00999062.1| outer membrane protein, OMP85 family protein [Oceanicola batsensis HTCC2597]
>gi|99081248|ref|YP_613402.1| surface antigen (D15) [Ruegeria sp. TM1040]
>gi|114768803|ref|ZP_01446429.1| putative outer membrane protein [alpha proteobacterium HTCC2255]
>gi|163747138|ref|ZP_02154494.1| outer membrane protein, putative [Oceanibulbus indolifex HEL-45]
>gi|255262892|ref|ZP_05342234.1| outer membrane protein assembly complex, YaeT protein [Thalassiobium sp. R2A62]
>gi|86138417|ref|ZP_01056991.1| outer membrane protein, OMP85 family protein [Roseobacter sp. MED193]
>gi|56696552|ref|YP_166909.1| OMP85 family outer membrane protein [Ruegeria pomeroyi DSS-3]
>gi|241204511|ref|YP_002975607.1| outer membrane protein assembly complex, YaeT protein [Rhizobium leguminosarum bv. trifolii WSM1325]
>gi|159044048|ref|YP_001532842.1| hypothetical protein Dshi_1499 [Dinoroseobacter shibae DFL 12]
>gi|116251984|ref|YP_767822.1| putative outer membrane protein [Rhizobium leguminosarum bv. viciae 3841]
>gi|254437680|ref|ZP_05051174.1| outer membrane protein assembly complex, YaeT protein [Octadecabacter antarcticus 307]
>gi|89054944|ref|YP_510395.1| surface antigen (D15) [Jannaschia sp. CCS1]
>gi|308754761|gb|ADO42690.1| putative outer membrane protein [Ketogulonicigenium vulgare Y25]
>gi|84686911|ref|ZP_01014795.1| putative outer membrane protein [Maritimibacter alkaliphilus HTCC2654]
>gi|163736302|ref|ZP_02143721.1| surface antigen (D15) [Phaeobacter gallaeciensis BS107]
>gi|84516077|ref|ZP_01003437.1| putative outer membrane protein [Loktanella vestfoldensis SKA53]

>gi|254475738|ref|ZP_05089124.1| outer membrane protein assembly complex, YaeT protein [Ruegeria sp. R11]
>gi|68171450|ref|ZP_00544838.1| surface antigen (D15):Surface antigen variable number [Ehrlichia chaffeensis str. Sapulpa]
>gi|226941202|ref|YP_002796276.1| probable outer membrane protein [Laribacter hongkongensis HLHK9]
>gi|269958465|ref|YP_003328252.1| Outer membrane protein/protective antigen OMA87 [Anaplasma centrale str. Israel]
>gi|294671228|ref|ZP_06736081.1| hypothetical protein NEIELOOT_02938 [Neisseria elongata subsp. glycolytica ATCC 29315]
>gi|166712740|ref|ZP_02243947.1| outer membrane antigen [Xanthomonas oryzae pv. oryzicola BLS256]
>gi|91775875|ref|YP_545631.1| surface antigen (D15) [Methylobacillus flagellatus KT]
>gi|255059025|ref|ZP_05311185.1| outer membrane protein assembly complex, YaeT protein [Geobacter sp. M18]
>gi|292492496|ref|YP_003527935.1| outer membrane protein assembly complex, YaeT protein [Nitrosococcus halophilus Nc4]
>gi|295676817|ref|YP_003605341.1| outer membrane protein assembly complex, YaeT protein [Burkholderia sp. CCGE1002]
>gi|77918855|ref|YP_356670.1| putative outer membrane protein [Pelobacter carbinolicus DSM 2380]
>gi|190573494|ref|YP_001971339.1| putative outer membrane protein [Stenotrophomonas maltophilia K279a]
>gi|56417133|ref|YP_154207.1| outer membrane protein [Anaplasma mar>ginale str. St. Maries]
>gi|224825021|ref|ZP_03698127.1| outer membrane protein assembly complex, YaeT protein [Lutiella nitroferrum 2002]
>gi|74316811|ref|YP_314551.1| surface antigen (D15) [Thiobacillus denitrificans ATCC 25259]
>gi|77164334|ref|YP_342859.1| Outer membrane protein [Nitrosococcus oceani ATCC 19707]
>gi|92112704|ref|YP_572632.1| surface antigen (D15) [Chromohalobacter salexigens DSM 3043]
>gi|255063971|ref|ZP_05315951.1| outer membrane protein assembly complex, YaeT protein [Nitrosomonas sp. AL212]
>gi|257094436|ref|YP_003168077.1| outer membrane protein assembly complex, YaeT protein [Candidatus Accumulibacter phosphatis clade IIA str. UW-1]
>gi|253996529|ref|YP_003048593.1| outer membrane protein assembly complex, YaeT protein [Methylotenera mobilis JLW8]

B) TpsB

>gi|33592931|ref|NP_880575.1| hemolysin activator-like protein [Bordetella pertussis Tohama I]
>gi|284159974|ref|YP_001063475.2| HlyB family hemolysin activator protein [B
>gi|260913787|ref|ZP_05920262.1| hemolysin activation/secretion family protein LspB [Pasteurella dagmatis ATCC 43325]
>gi|170717741|ref|YP_001784810.1| polypeptide-transport-associated domain-containing protein [Haemophilus somnus 2336]
>gi|238790665|ref|ZP_04634429.1| Hemolysin activation/secretion protein [Yersinia frederiksenii ATCC 33641]
>gi|288550380|ref|ZP_05970214.2| hemolysin secretion/activation protein, ShlB/FhaC/HecB family [Enterobacter cancerogenus ATCC 35316]
>gi|300716753|ref|YP_003741556.1| Filamentous hemagglutinin transporter protein [Erwinia billin>giae Eb661]
>gi|283953750|ref|ZP_06371281.1| hypothetical protein C414_000010117 [Campylobacter jejuni subsp. jejuni 414]

>gi|280957289|ref|ZP_06231943.1| Polypeptide-transport-associated domain protein ShlB-type [Desulfovibrio aespoeensis Aspo-2]
>gi|157372636|ref|YP_001480625.1| polypeptide-transport-associated domain-containing protein [Serratia proteamaculans 568]
>gi|222087780|ref|YP_002546317.1| Hemolysin activator protein hec [Agrobacterium radiobacter K84]
>gi|254491914|ref|ZP_05105093.1| hemolysin secretion/activation protein ShlB/FhaC/HecB [Methylophaga thiooxidans DMS010]
>gi|296104784|ref|YP_003614930.1| TPS family two-partner secretion family member CdiB [Enterobacter cloacae subsp. cloacae ATCC 13047]
>gi|227114201|ref|ZP_03827857.1| putative hemolysin activator protein [Pectobacterium carotovorum subsp. brasiliensis PBR1692]
>gi|191170679|ref|ZP_03032231.1| CdiB [Escherichia coli F11]
>gi|94499197|ref|ZP_01305735.1| putative activator or transporter protein of haemolysin-like protein [Oceanobacter sp. RED65]
>gi|306812067|ref|ZP_07446273.1| putative hemolysin activator HlyB [Escherichia coli NC101]
>gi|242239433|ref|YP_002987614.1| Polypeptide-transport-associated domain protein ShlB-type [Dickeya dadantii Ech703]
>gi|307235932|ref|ZP_07522343.1| putative hemolysin activator HlyB [Escherichia coli TA271]
>gi|300903959|ref|ZP_07121847.1| POTRA domain, ShlB-type [Escherichia coli MS 84-1]

>gi|206576167|ref|YP_002239914.1| haemolysin secretion/activation protein, ShlB/FhaC/HecB family [Klebsiella pneumoniae 342]
>gi|300724496|ref|YP_003713817.1| TpsB protein [Xenorhabdus nematophila ATCC 19061]
>gi|92110262|emb|CAJ87527.1| hypothetical protein [Escherichia coli]
>gi|260424490|ref|YP_003212691.1| hypothetical protein Ctu_1p01150 [Cronobacter turicensis z3032]
>gi|256019173|ref|ZP_05433038.1| putative hemolysin activator HlyB [Shigella sp. D9]
>gi|240850390|ref|YP_002971784.1| hemolysin activator protein Hec [Bartonella grahamii as4aup]
>gi|1772622|gb|AAC31980.1| HecB [Erwinia chrysanthemi]
>gi|194363890|ref|YP_002026500.1| Polypeptide-transport-associated domain-containing protein ShlB-type [Stenotrophomonas maltophilia R551-3]
>gi|116050407|ref|YP_790774.1| hypothetical protein PA14_32780 [Pseudomonas aeru>ginosa UCBPP-PA14]
>gi|291616459|ref|YP_003519201.1| ShlB [Pantoea ananatis LMG 20103]
>gi|303328174|ref|ZP_07358613.1| putative secretion protein, TPS system [Desulfovibrio sp. 3_1_syn3]
>gi|166713668|ref|ZP_02244875.1| outer membrane hemolysin activator protein [Xanthomonas oryzae pv. oryzicola BLS256]
>gi|288937326|ref|YP_003441385.1| Polypeptide-transport-associated domain protein ShlB-type [Klebsiella variicola At-22]
>gi|300724387|ref|YP_003713705.1| TpsB protein [Xenorhabdus nematophila ATCC 19061]
>gi|256829310|ref|YP_003158038.1| Polypeptide-transport-associated domain protein ShlB-type [Desulfomicrobium baculatum DSM 4028]
>gi|303253556|ref|ZP_07339694.1| hypothetical protein APP2_0744 [Actinobacillus pleuropneumoniae serovar 2 str. 4226]
>gi|146309719|ref|YP_001174793.1| polypeptide-transport-associated domain-containing protein [Enterobacter sp. 638]
>gi|300716762|ref|YP_003741565.1| Contact-dependent inhibition of growth factor CdiB [Erwinia billin>giae Eb661]
>gi|157737185|ref|YP_001489868.1| hemolysin activation protein HecB, putative [Arcobacter butzleri RM4018]
>gi|254243152|ref|ZP_04936474.1| conserved hypothetical protein [Pseudomonas aeru>ginosa 2192]

>gi|17546498|ref|NP_519900.1| activation/secretion signal peptide protein [Ralstonia solanacearum GMI1000]
>gi|261393114|emb|CAX50714.1| TpsA3 activation/secretion protein TpsB3 [Neisseria menin>gitidis 8013]
>gi|169633094|ref|YP_001706830.1| putative hemolysin activator protein [Acinetobacter baumannii SDF]
>gi|261821784|ref|YP_003259890.1| Polypeptide-transport-associated domain protein ShlB-type [Pectobacterium wasabiae WPP163]
>gi|308187325|ref|YP_003931456.1| Hemolysin activator protein precursor [Pantoea vagans C9-1]
>gi|37526953|ref|NP_930297.1| hypothetical protein plu3065 [Photorhabdus luminescens subsp. laumondii TTO1]

>gi|153800959|ref|ZP_01955545.1| HecB [Vibrio cholerae MZO-3]
>gi|299530601|ref|ZP_07044018.1| Hemolysin activator HlyB [Comamonas testosteroni S44]
>gi|190573398|ref|YP_001971243.1| putative activation/secretion precursor [Stenotrophomonas maltophilia K279a]
>gi|238023284|ref|ZP_04603710.1| hypothetical protein GCWU000324_03211 [Kingella oralis ATCC 51147]
>gi|209546156|ref|YP_002278046.1| Polypeptide-transport-associated domain protein ShlB-type [Rhizobium leguminosarum bv. trifolii WSM2304]
>gi|258623829|ref|ZP_05718785.1| Hemolysin activator protein [Vibrio mimicus VM603]
>gi|37524246|ref|NP_927590.1| hypothetical protein plu0226 [Photorhabdus luminescens subsp. laumondii TTO1]
>gi|157164569|ref|YP_001466116.1| ATP phosphoribosyltransferase (ATP-PRTase; ATP-PRT) [Campylobacter concisus 13826]
>gi|257464764|ref|ZP_05629135.1| hemolysin activation/secretion protein [Actinobacillus minor 202]
>gi|300991439|ref|ZP_07179563.1| POTRA domain, ShlB-type protein [Escherichia coli MS 45-1]
>gi|254361280|ref|ZP_04977423.1| hemagglutinin secretion protein FhaC [Mannheimia haemolytica PHL213]
>gi|258545221|ref|ZP_05705455.1| ShlB family hemolysin secretion/activation protein [Cardiobacterium hominis ATCC 15826]
>gi|123205|sp|P15321.1|HLYB_SERMA RecName: Full=Hemolysin transporter protein shlB; Flags: Precursor
>gi|37527441|ref|NP_930785.1| hypothetical protein plu3569 [Photorhabdus luminescens subsp. laumondii TTO1]
>gi|170718395|ref|YP_001783618.1| polypeptide-transport-associated domain-containing protein [Haemophilus somnus 2336]
>gi|294635237|ref|ZP_06713739.1| hemolysin activator protein [Edwardsiella tarda ATCC 23685]
>gi|157372714|ref|YP_001480703.1| hemolysin activator HlyB domain-containing protein [Serratia proteamaculans 568]
>gi|253686474|ref|YP_003015664.1| Polypeptide-transport-associated domain protein ShlB-type [Pectobacterium carotovorum subsp. carotovorum PC1]
>gi|258545423|ref|ZP_05705657.1| ShlB family hemolysin secretion/activation protein [Cardiobacterium hominis ATCC 15826]
>gi|293416250|ref|ZP_06658890.1| hypothetical protein ECDG_03853 [Escherichia coli B185]
>gi|207739547|ref|YP_002257940.1| activation/secretion protein [Ralstonia solanacearum IPO1609]
>gi|92115267|ref|YP_575195.1| hemolysin activator HlyB [Chromohalobacter salexigens DSM 3043]

C) YtfM-like

>gi|16125850|ref|NP_420414.1| hypothetical protein CC_1603 OMP60 [Caulobacter crescentus CB15]
>gi|197104873|ref|YP_002130250.1| hypothetical protein PHZ_c1407 [Phenylobacterium zucineum HLK1]
>gi|114571188|ref|YP_757868.1| surface antigen (D15) [Maricaulis maris MCS10]
>gi|254293136|ref|YP_003059159.1| surface antigen (D15) [Hirschia baltica ATCC 49814]
>gi|83858434|ref|ZP_00951956.1| hypothetical protein OA2633_03006 [Oceanicaulis alexandrii HTCC2633]
>gi|298530622|ref|ZP_07018024.1| surface antigen (D15) [Desulfonatronospira thiodismutans ASO3-1]
>gi|170718333|ref|YP_001783561.1| surface antigen (D15) [Haemophilus somnus 2336]
>gi|163795803|ref|ZP_02189767.1| surface antigen [alpha proteobacterium BAL199]
>gi|288958093|ref|YP_003448434.1| surface antigen [Azospirillum sp. B510]
>gi|260464270|ref|ZP_05812462.1| surface antigen (D15) [Mesorhizobium opportunistum WSM2075]
>gi|114798975|ref|YP_760558.1| OMP85 family outer membrane protein [Hyphomonas neptunium ATCC 15444]
>gi|154250840|ref|YP_001411664.1| surface antigen (D15) [Parvibaculum lavamentivorans DS-1]
>gi|90419999|ref|ZP_01227908.1| putative surface antigen (D15) [Aurantimonas manganoxydans SI85-9A1]
>gi|304392587|ref|ZP_07374527.1| surface antigen [Ahrensia sp. R2A130]
>gi|103487047|ref|YP_616608.1| surface antigen (D15) [Sphingopyxis alaskensis RB2256]
>gi|153007398|ref|YP_001368613.1| surface antigen (D15) [Ochrobactrum anthropi ATCC 49188]
>gi|163759508|ref|ZP_02166593.1| hypothetical protein HPDFL43_09152 [Hoeflea phototrophica DFL-43]
>gi|118590308|ref|ZP_01547711.1| hypothetical protein SIAM614_12358 [Stappia aggregata IAM 12614]
>gi|83593434|ref|YP_427186.1| surface antigen [Rhodospirillum rubrum ATCC 11170]
>gi|297537874|ref|YP_003673643.1| surface antigen (D15) [Methylotenera sp. 301]
>gi|227329268|ref|ZP_03833292.1| hypothetical protein PcarcW_18802 [Pectobacterium carotovorum subsp. carotovorum WPP14]
>gi|217978754|ref|YP_002362901.1| surface antigen (D15) [Methylocella silvestris BL2]
>gi|291327125|ref|ZP_06127075.2| outer membrane protein, OMP85 family [Providencia rettgeri DSM 1131]
>gi|226944959|ref|YP_002800032.1| outer membrane surface antigen (D15) [Azotobacter vinelandii DJ]
>gi|284006645|emb|CBA71907.1| cell surface protein [Arsenophonus nasoniae]
>gi|259417957|ref|ZP_05741876.1| surface antigen [Silicibacter sp. TrichCH4B]
>gi|170691991|ref|ZP_02883155.1| surface antigen (D15) [Burkholderia graminis C4D1M]
>gi|260599440|ref|YP_003212011.1| hypothetical protein Ctu_36480 [Cronobacter turicensis z3032]

>gi|289207481|ref|YP_003459547.1| surface antigen (D15) [Thioalkalivibrio sp. K90mix]
>gi|110644585|ref|YP_672315.1| putative outer membrane protein [Escherichia coli 536]
>gi|15597739|ref|NP_251233.1| hypothetical protein PA2543 [Pseudomonas aeru>ginosa PAO1]
>gi|207859556|ref|YP_002246207.1| hypothetical protein SEN4178 [Salmonella enterica subsp. enterica serovar Enteritidis str. P125109]
>gi|254240974|ref|ZP_04934296.1| conserved hypothetical protein [Pseudomonas aeru>ginosa 2192]
>gi|162418364|ref|YP_001608258.1| OMP85 family outer membrane protein [Yersinia pestis Angola]

>gi|283786898|ref|YP_003366763.1| putative outer membrane protein assembly factor [Citrobacter rodentium ICC168]
>gi|227357104|ref|ZP_03841474.1| outer membrane protein [Proteus mirabilis ATCC 29906]
>gi|262045353|ref|ZP_06018377.1| OMP85 family outer membrane protein [Klebsiella pneumoniae subsp. rhinoscleromatis ATCC 13884]
>gi|238897968|ref|YP_002923648.1| putative outer membrane protein [Candidatus Hamiltonella defensa 5AT (Acyrthosiphon pisum)]
>gi|120554929|ref|YP_959280.1| surface antigen (D15) [Marinobacter aquaeolei VT8]
>gi|283477009|emb|CAY72901.1| Uncharacterized protein ytfM precursor [Erwinia pyrifoliae DSM 12163]
>gi|271501872|ref|YP_003334898.1| surface antigen (D15) [Dickeya dadantii Ech586]
>gi|222053457|ref|YP_002535819.1| surface antigen (D15) [Geobacter sp. FRC-32]
>gi|192360745|ref|YP_001983825.1| outer membrane protein, OMP85 family [Cellvibrio japonicus Ueda107]
>gi|218667177|ref|YP_002426365.1| outer membrane protein, OMP85 family [Acidithiobacillus ferrooxidans ATCC 23270]
>gi|270265029|ref|ZP_06193292.1| surface antigen (D15) [Serratia odorifera 4Rx13]
>gi|300310867|ref|YP_003774959.1| outer membrane lipoprotein [Herbaspirillum seropedicae SmR1]
>gi|254489758|ref|ZP_05102953.1| outer membrane protein, OMP85 family, putative [Methylophaga thiooxidans DMS010]
>gi|159045921|ref|YP_001534715.1| hypothetical protein Dshi_3381 [Dinoroseobacter shibae DFL 12]
>gi|288937478|ref|YP_003441537.1| surface antigen (D15) [Klebsiella variicola At-22]
>gi|77163901|ref|YP_342426.1| Outer membrane protein [Nitrosococcus oceani ATCC 19707]
>gi|261254021|ref|ZP_05946594.1| uncharacterized protein YtfM precursor [Vibrio orientalis CIP 102891]
>gi|117620359|ref|YP_856091.1| OMP85 family outer membrane protein [Aeromonas hydrophila subsp. hydrophila ATCC 7966]
>gi|269103570|ref|ZP_06156267.1| uncharacterized protein YtfM precursor [Photobacterium damselae subsp. damselae CIP 102761]

>gi|82702920|ref|YP_412486.1| surface antigen (D15) [Nitrosospira multiformis ATCC 25196]
>gi|85058330|ref|YP_454032.1| hypothetical protein SG0352 [Sodalis glossinidius str. 'morsitans']
>gi|307545166|ref|YP_003897645.1| K07278 outer membrane protein [Halomonas elongata DSM 2581]
>gi|218710694|ref|YP_002418315.1| outer membrane protein [Vibrio splendidus LGP32]
>gi|89072103|ref|ZP_01158699.1| hypothetical outer membrane protein [Photobacterium sp. SKA34]
>gi|152981137|ref|YP_001354293.1| hypothetical protein mma_2603 [Janthinobacterium sp. Marseille]
>gi|285019897|ref|YP_003377608.1| putative outer membrane protein [Xanthomonas albilineans]
>gi|302342274|ref|YP_003806803.1| surface antigen (D15) [Desulfarculus baarsii DSM 2075]
>gi|190576417|ref|YP_001974262.1| putative surface antigen exported protein [Stenotrophomonas maltophilia K279a]
>gi|119945606|ref|YP_943286.1| surface antigen (D15) [Psychromonas ingrahamii 37]
>gi|126174129|ref|YP_001050278.1| surface antigen (D15) [Shewanella baltica OS155]
>gi|294341688|emb|CAZ90107.1| Putative outer membrane protein, putative antigen [Thiomonas sp. 3As]
>gi|88705024|ref|ZP_01102736.1| protein containing bacterial surface antigen (D15) [Congre>gibacter litoralis KT71]
>gi|160871599|ref|ZP_02061731.1| outer membrane protein [Rickettsiella grylli]
>gi|262198434|ref|YP_003269643.1| surface antigen (D15) [Halian>gium ochraceum DSM 14365]
>gi|74316811|ref|YP_314551.1| surface antigen (D15) [Thiobacillus denitrificans ATCC 25259]

D) YtfM-like POTRA domains (sorted by position of the domains)

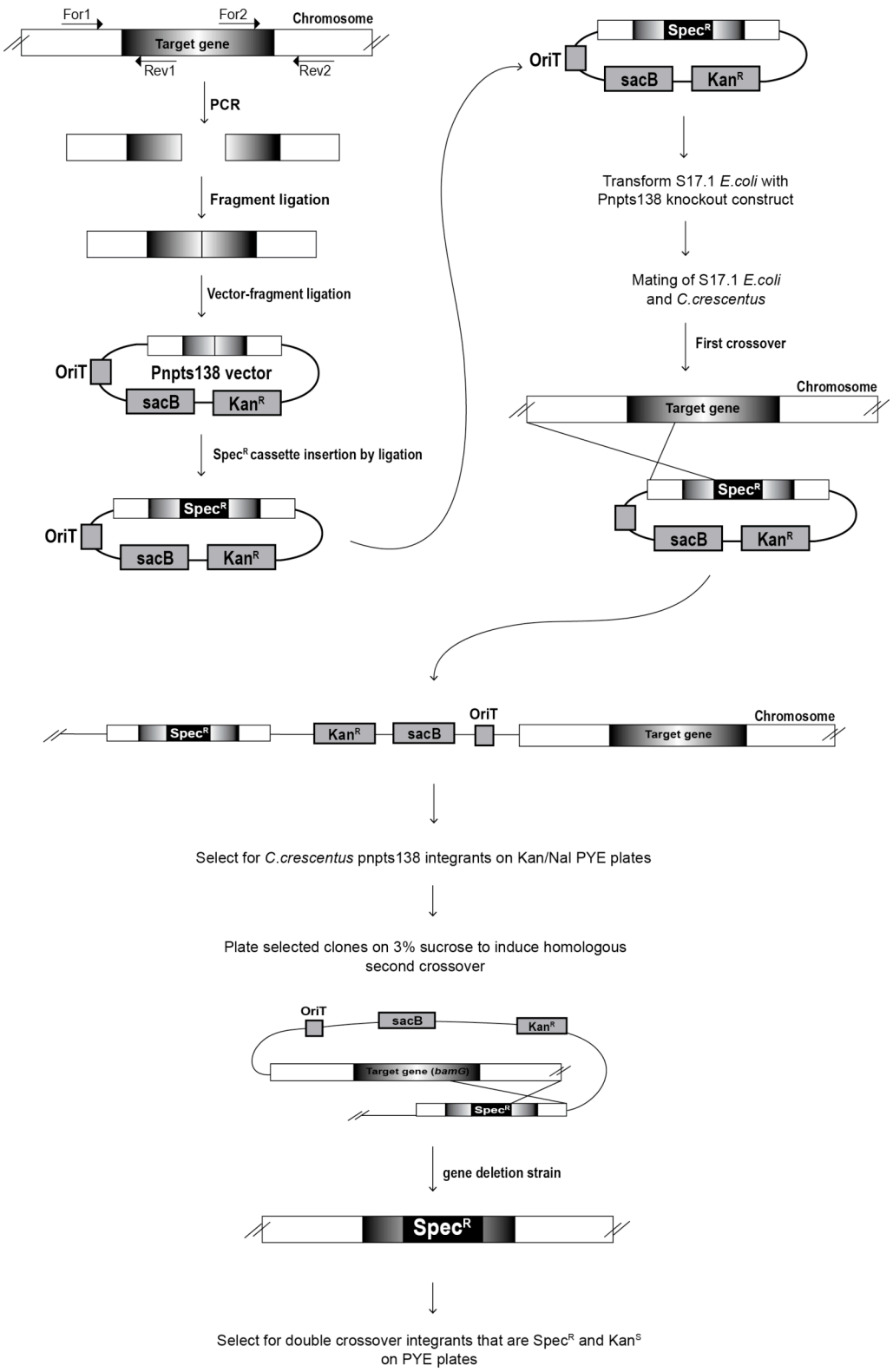
*A POTRA domain is defined as the βααββ secondary structure motif in the PSIPRED secondary structure predictions of the full-length protein. Additional information is given in the sequence headers: >12gi... means POTRA domain 1 of 2 in the protein.

>13>gi|16125850|ref|NP_420414.1|putative outer membrane protein Omp60 [Caulobacter crescentus CB15]
DEPMAQIQGVEDRALRDAIQRALSDSKQPPRSRSEARRRARQAGEDAIAVLRAEGYYAYTVEPD
VTEGDPPRAIVRITP
>23>gi|16125850|ref|NP_420414.1|putative outer membrane protein Omp60 [Caulobacter crescentus CB15]
GPAFLLADPHIDWSGSPPDEGVRQRAVAAMRLTEGEPGRSADVVGAEGRIVAQVAKLGYADVA
AEPREVVVDHADRTVRPTFRIMAG
>33>gi|16125850|ref|NP_420414.1|putative outer membrane protein Omp60[Caulobacter crescentus CB15]
ELVRLNGVDVVTKGRTNPEWVGRLAPWVAGDVYDPEDVAELERRLRDTAVYDSISVSLAGTDK
ASAEGYRPVVVTLSDRR
>13>gi|26251125|ref|NP_757165.1| Hypothetical protein ytfM precursor [Escherichia coli CFT073]
ANVRLQVEGLSGQLEKNVRAQLSTIESDEVTPDRRFRARVDDAIREGLKALGYYQPTIEFDLRPP
PKKGRQVLIAKVTPG
>23>gi|26251125|ref|NP_757165.1| Hypothetical protein ytfM precursor [Escherichia coli CFT073]
VPVLIGGTDVVLRGGARTDKDYLKLLDTRPAIGTVLNQGDYENFKKSLTSIALRKGYFDSEFTKAQ
L>GIALGLHKAFWDIDYNS
>33>gi|26251125|ref|NP_757165.1| Hypothetical protein ytfM precursor [Escherichia coli CFT073]

Appendix A2 – Genetics in *Caulobacter*

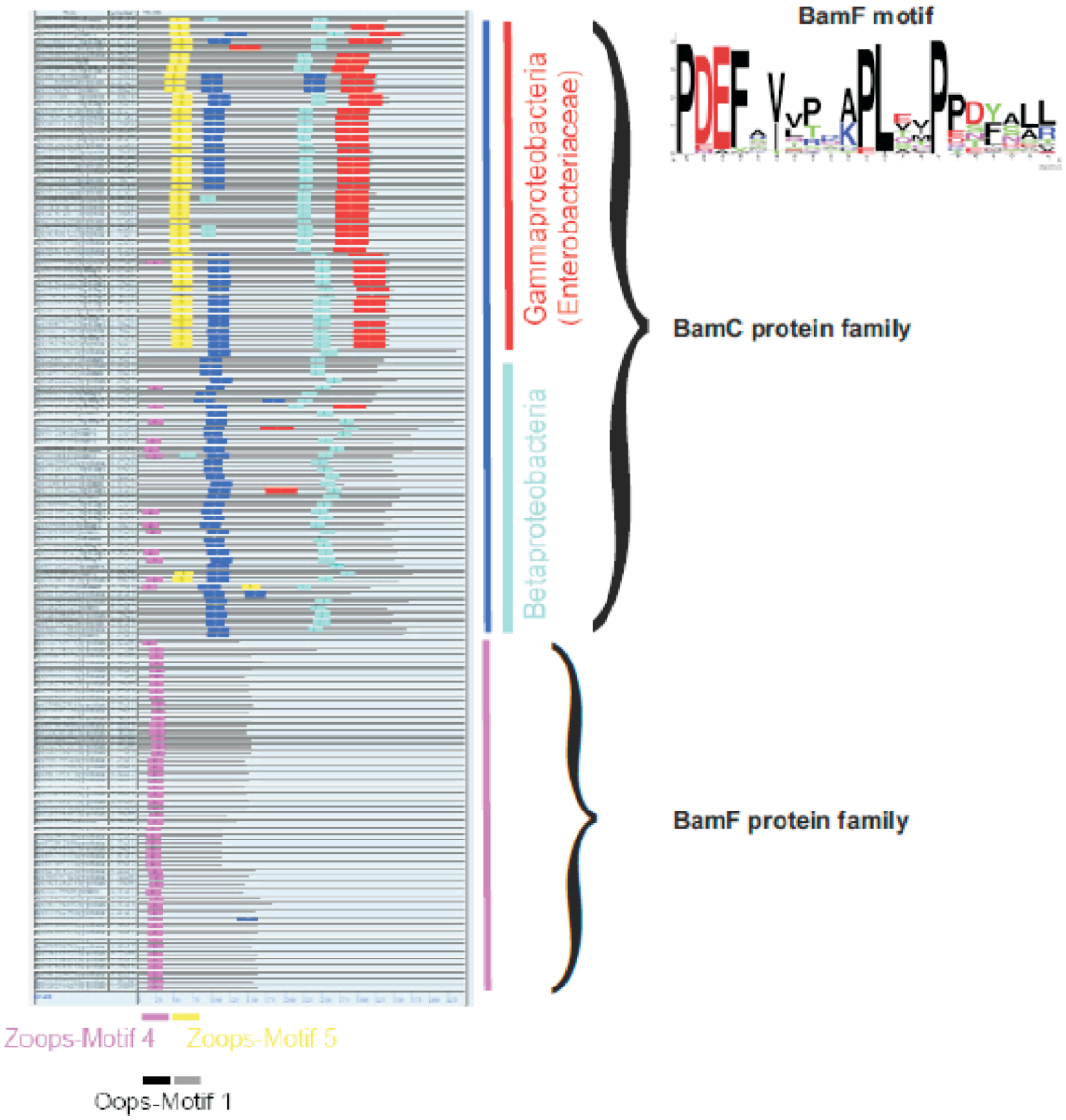
To generate mutants of *C. crescentus* (see Appendix A2), the S17.1 *E. coli* mating strain was transformed with the knockout or complementing plasmids. To introduce plasmids into the *C. crescentus* CB15N chromosome, S17.1 cells harbouring the plasmids were mated for 4 h or overnight with *C. crescentus* at 30C on plain PYE plates. For integration of non-replicating suicide plasmids, transformants with single homologous recombination were selected on plates containing suitable antibiotics and nalidixic acid. For double recombination, clones were grown to stationary phase in PYE media lacking antibiotics at 30ºC and spread on PYE plates containing 3% (w/v) sucrose. Single colonies were picked and transferred in parallel onto plain PYE plates and PYE plates containing kanamycin or spectinomycin. Kanamycin-sensitive clones, which had lost the integrated plasmid due to a second recombination event, were checked for spectinomycin-resistance.

Appendix A2 – Genetics in *Caulobacter*



Appendix A3 – MEME analysis

Motif analysis of the family of proteins closely related to Q9AAA7 (BamF) and NlpB (BamC). Each of the protein sequences analysed is represented as a single line, and these have been sorted into species of γ-proteobacteria (designated by the red stripe), β-proteobacteria (designated by the light blue stripe) and α-proteobacteria (designated by the pink stripe). The common motifs are shown as coloured boxes on each protein sequence: pink "Zoops motif 4" corresponds to the sequence logo shown as the BamF motif. It is found in all α-proteobacterial BamF proteins and also in many β-proteobacterial BamC proteins.



Appendix A4 – CDPro analysis of CD data

Experimental data (Expt_CD) was fitted using CDPro software. Selcon3, Contin and Cdsstr are inbuilt CDPro programs that calculate fractions (%) of secondary structure using different algorithms.

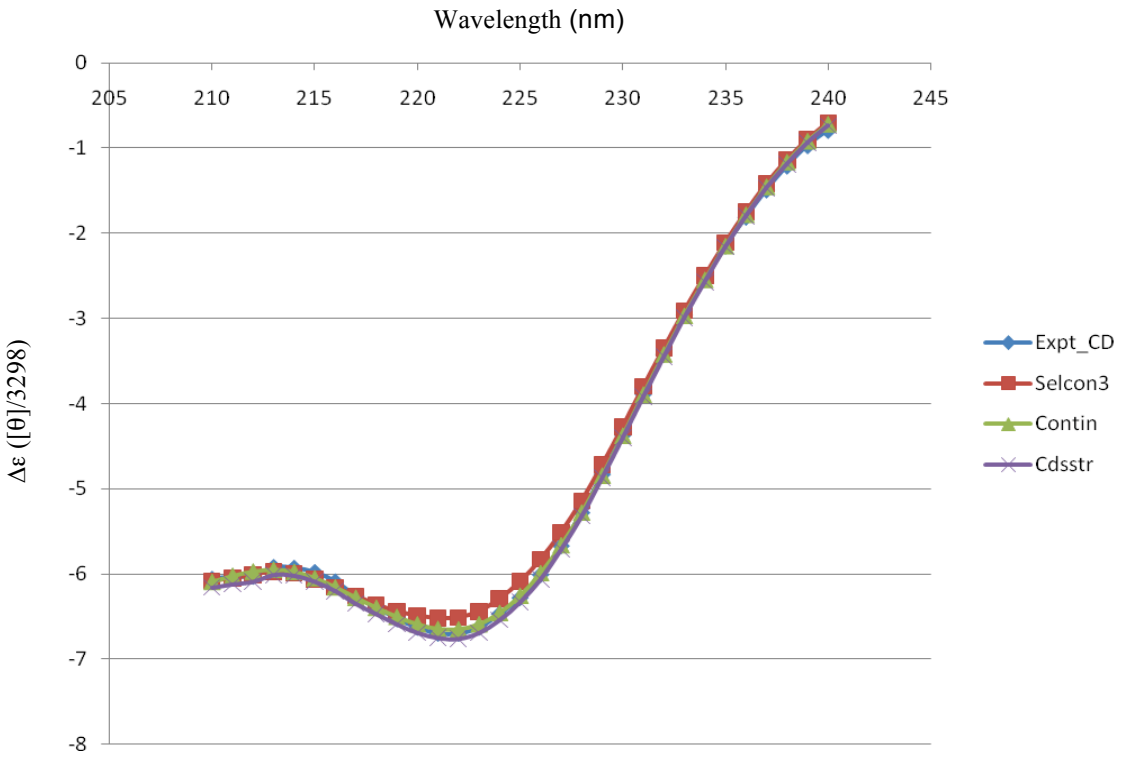


Table: Calculation of secondary structure contents (%) for the CD spectra of BamD using the SP43 reference set.

CDPro Algorithm	α-helix	β-sheet	Turns	Unordered
CONTIN	65.1	3.7	12.2	19.1
CDSSTR	81.7	4.9	5.6	7.2
SELCON3	66.8	3.7	12.3	18.9