Structural Analysis of the *Escherichia coli* β-barrel Assembly Machinery Complex

by

Kelly H. Kim

B.Sc., Simon Fraser University, 2006

THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

in the
Department of Molecular Biology and Biochemistry
Faculty of Science

SIMON FRASER UNIVERSITY
Summer 2012

Approval

Name: **Kelly H. Kim**

Degree: **Doctor of Philosophy**

Title of Thesis: **Structural Analysis of the *Escherchia coli* β-barrel Assembly Machinery Complex**

Examining Committee:

Chair: **Dr. Sharon M. Gorski**
Associate Professor

Dr. Mark Paetzel
Senior Supervisor
Associate Professor

Dr. Rosemary B. Cornell
Supervisor
Professor

Dr. Edgar C. Young
Supervisor
Associate Professor

Dr. Lisa Craig
Internal Examiner
Associate Professor
Simon Fraser University

Dr. Rachel C. Fernandez
External Examiner
Associate Professor, Department of Microbiology & Immunology
University of British Columbia

Date Defended/Approved: May 31, 2012

Partial Copyright Licence

Abstract

The outer membrane (OM) is a unique structural feature of Gram-negative bacteria. Residing within the outer membrane are β-barrel outer membrane proteins (OMPs) that serve many important cellular functions. As proper folding and assembly of these proteins are crucial for cell viability, Gram-negative bacteria possess a specialized proteinaceous machine, known as the BAM (β-barrel assembly machinery) complex, to catalyze the folding and membrane insertion of OMPs. In *Escherichia coli*, the BAM complex consists of five proteins: one β-barrel membrane protein – BamA, and four lipoproteins – BamB, BamC, BamD, and BamE. The roles of the individual components and how they are arranged into the BAM complex to function together is not yet clearly understood.

During the course of this thesis project, I determined the structures of *E. coli* BamB, BamC, BamE and the BamCD subcomplex. Analysis of the conserved residues and the molecular surface properties of these solved structures helped to identify potential protein-protein interaction sites on each lipoprotein. For example, BamC has two 'helix-grip' domains that are ideally shaped to accommodate α-helices. BamB, on the other hand, has a β-propeller fold that could potentially interact with BamA or substrates via β-augmentation, a mode of interaction in which a pre-existing β-sheet is augmented by an addition of a β-strand of another protein.

Comparing the solved structures with their structural homologs with known functions has also provided important clues about the functional roles of each lipoprotein. BamD structure, for example, closely resembles the binding pocket of a peroxisomal targeting signal receptor PEX5, suggesting a similar substrate recognition function for BamD. Interestingly, our BamCD complex structure shows that the putative substrate binding pocket of BamD is bound and blocked by the conserved unstructured N-terminal region of BamC. This suggests a possibility that BamC may have a regulatory function.The structural and interaction data acquired from this thesis project contributes to a better understanding of the BAM complex structure and provides a platform for future research driven by structure-based hypotheses.

Keywords: outer membrane; outer membrane protein (OMP); β-barrel assembly machinery (BAM) complex; β-barrel; protein folding

To My Family and Tony

Acknowledgements

First of all, I would like to express my deepest gratitude to my senior supervisor, Dr. Mark Paetzel, who has supported me thoughout my thesis with his guidance and knowledge whilst allowing me the room to work in my own way. I appreciate his patience, optimism and contagious enthusiasm, even during the times when experiments were not proceeding the way we hoped they would.

I would also like to thank my supervisory committee members, Dr. Rosemary Cornell and Dr. Edgar Young, for their excellent advice and suggestions throughout my studies. I am grateful to Dr. Lisa Craig and Dr. Sharon Gorski for serving on my thesis committee and Dr. Rachel Fernandez for being my external examiner. I would also like to thank our collaborator, Dr. Lawrence McIntosh from UBC, and also Mark Okon, Hyunseo Kang and Eric Escobar-Cabrera for their help and guidance for my NMR work.

I would like to express my sincere thanks to our lab manager, Deidre de Jong-Wong, for making the Paetzel lab the neatest and the most organized workplace I have ever worked in. Her support throughout my PhD career (and the baked goods she brought for us every Monday) has made this thesis possible. I would also like to thank Dr. Jaeyong Lee, Dr. Anat Feldman, Dr. Sukhbir Kang and Dr. David Oliver for being wonderful mentors and patient teachers at the start of my graduate school career.

I would also like to thank the past and the present members of the Paetzel lab: Alison Li, Chuanyun Luo, Yuliya Shapova, Apollos Kim, Ivy Chung, Charlie Stevens, Sungeun Nam, Daniel Chiang, Minfei Fu, Suraaj Aulakh, Linda Zhang, Wendy Tan, PJ Luczon, Jonathan Tan and Cindy Li. I would like to thank everyone for their encouragement and friendship that made this lab a fun and enjoyable environment. I especially would like to thank Charlie for troubleshooting all the computer problems, and Suraaj for being a great science partner and giving me the opportunity to be involved in her various science communication projects like the Labtricks.com website.

Finally, I would like to thank my parents, my brother, and Tony for their unwavering support throughout my studies and for always believing in me. Thank you.

Table of Contents

List of Tables

List of Figures

Glossary

Å	Ångströms (10^{-10} m)
Anomalous Scattering	A change in a scattered X-ray's phase that is unique from the rest of the atoms in a crystal due to strong X-ray absorbance. Anomalous X-ray scattering results when heavy atoms scatter at a specific wavelength that is near the absorption edge of the heavy atom. For protein crystals, a heavy atom such as selenium must be incorporated into the protein in order to observe accurate anomalous scattering.
Asymmetric Unit	The smallest unit that can be rotated and translated to generate one unit cell using only the symmetry operators. The asymmetric unit may be one molecule, one subunit of a multimeric protein, or more than one molecule.
β-augmentation	A mode of intermolecular protein-protein interaction in which a pre-existing β-sheet is augmented by an addition of a β-strand.
β-barrel	A common protein folding motif. It can be described as a large β-sheet that closes to form a barrel-shaped structure. The hydrogen bonds between the N- and the C-terminal strands stabilize the barrel structure.
BAM Complex	β-barrel Assembly Machinery complex, a protein complex responsible for proper folding and membrane insertion of outer membrane proteins into the outer membrane. It is found in Gram-negative bacteria, and it consists of five different proteins, BamA, B, C, D and E.
B-factor	Also called 'temperature-factor'. A factor that describes the degree to which the electron density is spread out. The B-factor indicates the static or dynamic mobility of an atom. Higher B-factor values indicate higher disorder or mobility.
Bravais Lattice	Describes fourteen distinct three-dimensional arrangements of lattice points.
Chemical Shift	Resonant frequency of a nucleus relative to a standard. It is influenced by the chemical environment around the nucleus. It is expressed in ppm (parts per million).
CL	Cardiolipin, a minor phospholipid component of bacterial and mitochondrial membranes.

Column Volume	The total volume of a chromatography column (the sum of the void volume and the matrix volume)
Completeness	The number of crystallographic reflections measured in an X-ray diffraction data set, expressed as a percentage of the total number of reflections present at the specified resolution. It indicates whether all the reflections in the asymmetric unit have been measured.
Crystal	A regular three-dimensional repeat of molecules, with internal symmetry.
Crystal Lattice	The regular array of points about which molecules composing a crystal are centered.
Crystallographic Refinement	A cyclic process of improving R-factor (i.e. agreement between the molecular model and the crystallographic data) by adjustment of the model to obtain highly precise structural model that matches the measured data.
DDM	n-Dodecyl-β-D-maltoside, a non-ionic detergent most often used for the isolation and stabilization of hydrophobic membrane proteins.
DegP	A periplasmic chaperone/protease that acts in the Skp/DegP pathway to guide OMP-precusors to the BAM complex. Under high-temperature conditions, its protease function becomes active and is able to degrade misfolded or aggregated OMPs in the periplasm.
Fourier Transform	A mathematical operation that allows calculation of electron density map from structure factors, and *vice versa*.
Global Tumbling Time	A measurement of the time the molecule rotates through an angle of one radian. It is dependent on the size, shape, and dynamics of the molecule, as well as the physical characteristics of the solvent.
HSQC	Heteronuclear Single Quantum Correlation, an NMR experiment that results in a 2D spectrum with one axis for the chemical shift of ^{1}H and the other for non-hydrogen nucleus (most often ^{15}N or ^{13}C for protein NMR).
IM	Inner Membrane
IMP	Inner Membrane Protein

in vitro	Latin for “within glass” (i.e. in a test tube or petri dish). It refers to experimentation using components of an organism that have been isolated from their usual biological context.
in vivo	Latin for “within the living”. It refers to experimentation using a whole, living organism.
IPTG	Isopropyl-β-D-thiogalactopyranoside
kDa	Kilodalton, non-SI unit for molecular mass
Lipoprotein	A soluble protein in bacteria anchored to a membrane surface by diacylglycerol linked to its N-terminal cysteine.
LPS	Lipopolysaccharide, a molecule consisting of a lipid and a polysaccharide found exclusively on the outer leaftlet of bacterial outer membranes.
MAD	Multiple-wavelength Anomalous Diffraction, a method for deriving initial phases by measuring diffraction data at several different wavelengths near the absorption edge of a heavy-atom such as selenium. The anomalous signal that results from this can give very accurate phases.
Matthews Coefficient	Also called the specific volume (Vm). It is the crystal volume per unit of protein molecular mass and thus has a unit of $Å^3Da^{-1}$. The average is 2.4, in a range of 1.9-4.2.
Mobile Phase	In chromatography, a ‘mobile phase’ refers to the fluid which percolates through or along the stationary phase, in a definite direction.
Molecular Replacement	A method for deriving initial phases using a known homologous structure. Refer to section 2.5.1.5 for more information.
NMR	Nuclear Magnetic Resonance, a phenomenon in which magnetic nuclei in a magnetic field absorb and re-emit electromagnetic radiation.
NOESY	Nuclear Overhauser Effect Spectroscopy. In a NOESY experiment, nuclei are correlated that are close in space (<5Å). Therefore, peaks in a NOESY contain distance information that can be used to determine the structure of a molecule.
OM	Outer Membrane

OMP	Outer Membrane Protein
PAGE	Polyacrylamide Gel Electrophorosis
PCR	Polyemerase Chain Reaction, a molecular biology technique to amplify a piece of particular DNA sequence across several orders of magnitude.
PDB	Protein Data Bank
PE	Phosphatidylethanolamine
PEG	Polyethylene Glycol, widely used precipitant for promoting crystallization
Peptidoglycan	A polymer consisting of carbohydrates and amino acids that forms a mesh-like layer in the cell walls of bacteria.
Periplasm	Peptidoglycan containing region of the Gram-negative bacteria between the outer and the inner membranes.
PG	Phosphatidylglycerol
POTRA Domains	Polypeptide Transport Associated Domains. Five POTRA domains are found in the periplasmic region of BamA.
Redundancy	The average number of independent measurements of each reflection in a crystallographic data set. Redundancy is calculated as (number of measured reflections)/(number of unique reflections).
Resonance Assignment	A process of identifying which resonances in the NMR spectrum originate from which atom in the protein that is being examined.
R-factor and R_{free}	A measure of agreement between the crystal structure model and the original X-ray diffraction data. See Appendix C for more information.
R_{merge}	A measure of agreement among multiple measurements of the same reflections. See Appendix C for more information.
RMSD	Root-mean-square deviation, a measure of the differences between several measured values. In structural biology, RMSD is used to describe how well two or more structures align with each other. A lower RMSD value indicates higher structural similarity. It also used to

	indicate how well the final crystallographic model is consistent with expected values of bond lengths and bond angles.
S^2	In NMR relaxation experiments, S^2 represents a generalized order parameter representing the degree of spatial restriction.
SAD	Single-wavelength Anomalous Diffraction, a method for deriving initial phases that involves measuring diffraction data at a single wavelength near the absorption edge of a heavy-atom such as selenium.
SAM Complex	Sorting and Assembly Machinery Complex, a system homologous to the BAM complex found in the outer membrane of mitochondria. It is responsible for folding and membrane insertion of β-barrel proteins into the mitochondrial outer membrane.
SDS	Sodium Dodecyl Sulphate
Skp	A periplasmic chaperone acting in the Skp/DegP pathway to prevent aggregation of OMP-precursors in the periplasm.
Space Group	A mathematic description of a crystal lattice with a certain type of symmetry and a unit cell. Symmetry type is defined by a set of crystallographic symmetry operations, which may include rotation, translation, and screw axis that characterize a protein crystal. A Hermann-Mauguin space-group symbol is formed by specifying the Bravais lattice and a list of symmetry directions. For instance, $P2_12_12$, is an orthorhombic space group, with primitive lattice (P) and two-fold screw axes parallel to the x (2_1) and y (2_1) axes, and a normal two-fold rotational axis (2) along the z-axis.
Stokes Radius	The radius of a hypothetical hard sphere that would diffuse through the column medium at the same rate as the molecule being examined.
Structure Factor	A mathematical function describing the amplitude and phase of a wave diffracted from crystal lattice planes characterised by Miller indices h,k,l.
SurA	The main periplasmic chaperone that binds OMP-precursors to prevent their aggregation and to guide them to the BAM complex.
Synchrotron	A particle accelerator that produces very bright light (electromagnetic waves) in the region from infrared through to gamma rays.
T_1	Spin-lattice relaxation time. It is one of two relaxation times governing

	the rate of decay of magnetization. T_1 is characteristic time for restoration of the longitudinal component (which is parallel to the external magnetic field) of magnetization of the nucleus after application of a radio frequency pulse.
T_2	Spin-spin relaxation time. It is one of two relaxation times governing the rate of decay of magnetization. T_2 is characteristic time for the loss of transverse magnetization.
TOC Complex	Translocons at the Outer Envelope of Chloroplasts, a system homologous to the BAM complex found in the outer membrane of chloroplasts. It is responsible for folding and membrane insertion of β-barrel proteins into the outer membrane.
Tris	tris(hydroxymethyl)aminomethane, widely used as a component of buffer solutions. It has a pK_a value of 8.06 at 25 °C.
Unit Cell	The smallest repeating unit that can generate the crystal with only translation operations.
Void Volume	The volume of mobile phase in a chromatography column.
X-ray	A form of electromagnetic radiation that has a wave length ranging between 0.01 to 10 nm.

1. General Introduction

Note regarding contributions:

Portions of this chapter were published as review articles (sections 1.2 and 1.3). The authors and the full references for these articles are listed below.

Kim, K.H., Aulakh, S. & Paetzel, M. (2012) The bacterial outer membrane β-barrel assembly machinery*. Protein Science* **21**, 751-768.

Kim, K.H., Aulakh, S. & Paetzel, M. (2012) Outer Membrane Protein biosynthesis: transport and incorporation of proteins (in)to the OM bilayer. *Bacterial Membranes: Structural and Molecular Biology.* UK: Horizon Scientific Press. [invited book chapter; in press]

1.1. Bacterial Membranes

Bacteria are surrounded by a cell wall that defines the shape of the microorganism and controls the traffic of molecules entering and exiting the cell. In the case of pathogenic bacteria, the cell wall can also play a role in virulence and participate in the disease process (Bos et al., 2007; Silhavy et al., 2010; Tokuda, 2009).

Bacteria can be classified as either Gram-positive or Gram-negative based on the structure of their cell wall (Figure 1-1). Gram-positive bacteria have a cell wall consisting of a single membrane and a thick mesh-like layer of peptidoglycan, a polymer of amino acids and carbohydrates (Silhavy et al., 2010). In comparison, the cell wall of Gram-negative bacteria is made up of two membranes that sandwich a peptidoglycan containing region known as the periplasm. While the membrane facing the cytosol is designated as the inner membrane (IM), the outermost membrane that faces the extracellular environment is known as the outer membrane (OM) (Tokuda, 2009). The peptidoglycan layer in both Gram-positive and Gram-negative bacteria provides structural strength and prevents the cells from bursting by counteracting the osmotic pressure of the cytoplasm. However, the peptidoglycan layer is substantially thicker in Gram-positive bacteria (20-80 nm) than in Gram-negative bacteria (2-8 nm), making the entire cell wall of Gram-positive bacteria much thicker compared to that of the Gram-negative bacteria (Vollmer and Bertsche, 2008; Vollmer et al., 2008).

The bulk of the bacterial cell wall membrane is made up of phospholipids, with phosphatidylethanolamine (PE) being the most abundant, followed by phosphatidyl-glycerol (PG) and cardiolipin (CL) (Vance and Vance, 2008). The IM and the OM of *Escherichia coli* consist of approximately 75% PE, 20% PG and 5% CL (Morein et al., 1996). In both the Gram-positive bacterial membrane and the IM of Gram-negative bacteria, phospholipids are distributed evenly on either leaflet of the membrane. In contrast, the OM of Gram-negative bacteria is asymmetric with respect to the phospholipid distribution. While phospholipids are the major components of the inner leaflet of the OM, the outer leaflet is mostly composed of another class of lipids known as lipopolysaccharides (LPS) (Bos et al., 2007; Diedrich and Cota-Robles, 1974; Silhavy et al., 2010).

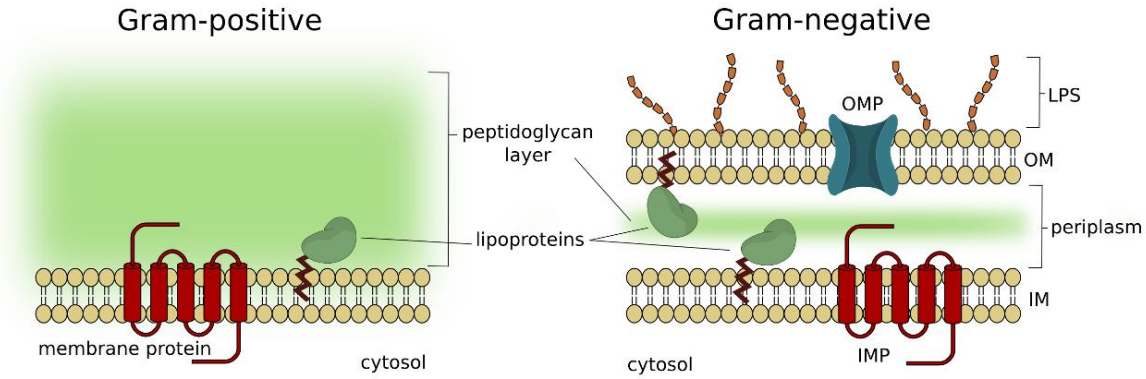


Figure 1-1 Bacterial Cell Walls

Schematic diagrams of Gram-positive (left) and Gram-negative (right) bacterial cell walls are shown for comparison. While Gram-positive bacteria are surrounded by a single phospholipid bilayer membrane, the Gram-negative bacterial envelope consists of two phospholipid bilayer membranes, the inner membrane (IM) and the outer membrane (OM). The OM is an asymmetric lipid bilayer, with lipopolysaccharides (LPS) only found in the outer leaflet of the OM. The space between the IM and the OM is known as the periplasm and it contains a thin layer of peptidoglycan. Gram-positive bacteria also contain a peptidoglycan layer, but it is significantly thicker. α-helical membrane proteins are found in both the Gram-positive bacterial membrane and the IM of Gram-negative bacteria (known as the inner membrane proteins, or IMPs). In contrast, the OM proteins (OMPs) adopt a β-barrel structure. Membrane anchored lipoproteins are also found in the cell walls of both types of bacteria.

In addition to lipids, proteins of various functions are also found in the bacterial membranes. Bacterial membrane proteins can be divided into different categories according to their location relative to the membrane. These include integral membrane (or transmembrane) proteins, peripheral membrane proteins that are temporarily associated with the lipid bilayer, and lipid-anchored proteins (or lipoproteins) (Hayashi and Wu, 1990; Silhavy et al., 2010). In Gram-negative bacteria, the presence of two membranes gives rise to two different classes of integral membrane proteins. The integral membrane proteins of the IM are known as the inner membrane proteins (IMPs), and they span the membrane by forming one or more α-helices. Integral membrane proteins of the OM are correspondingly known as the outer membrane proteins (OMPs), and they differ from the IMPs in that they adopt a β-barrel structure to traverse the membrane (Fairman et al., 2011; Silhavy et al., 2010; Tokuda, 2009).

The synthesis of bacterial membranes is a highly coordinated process that requires a variety of proteinaceous machineries to function together. Different components of the membranes not only need to be properly synthesized, but they must also be correctly transported to and inserted into the pre-existing membranes. Processes involved in cell wall synthesis have proven to be useful targets for antibacterial drugs such as penicillin and vancomycin, and thus are of great research interest. Currently, one of the least understood processes of cell wall synthesis is how OMPs are folded into the OM. It is also the main focus of this thesis. The current understanding of OMPs and their assembly at the OM of Gram-negative bacteria is summarized in the following sections of this chapter.

1.2. Outer Membrane Proteins (OMPs)

The OM is a unique structural feature of Gram-negative bacteria. It is an important physical barrier that acts as a molecular sieve to regulate the traffic of solutes into and out of the bacterium. This selective permeability of the OM is largely determined by the pore-forming proteins found within it known as the outer membrane proteins, or OMPs for short. OMPs play a critical role in the structural and functional integrity of the OM, and their proper biogenesis and functioning are imperative for cell survival (Bos et al., 2007; Silhavy et al., 2010).

1.2.1. OMP Structures

To date, nearly all OMPs with known structures contain a transmembrane β-barrel domain. Two exceptions are the *E. coli* polysaccharide translocon protein Wza and the *Corynebacterium glutamicum* porin PorB, and the TypeIV outer membrane secretion complex found throughout Gram-negative bacteria, all of which span the OM by forming an α-helical barrel (Collins and Derrick, 2007; Dong et al., 2006; Ziegler et al., 2008). In this thesis, the term OMP specifically refers to the integral β-barrel proteins found in the OM. Although OMPs share a common β-barrel architecture, differences in the number of strands, the length and properties of the loops, and oligomeric state add up to a diverse group of proteins with distinct functions. Selected structures of OMPs are illustrated in Figure 1-2 to highlight their structural diversity.

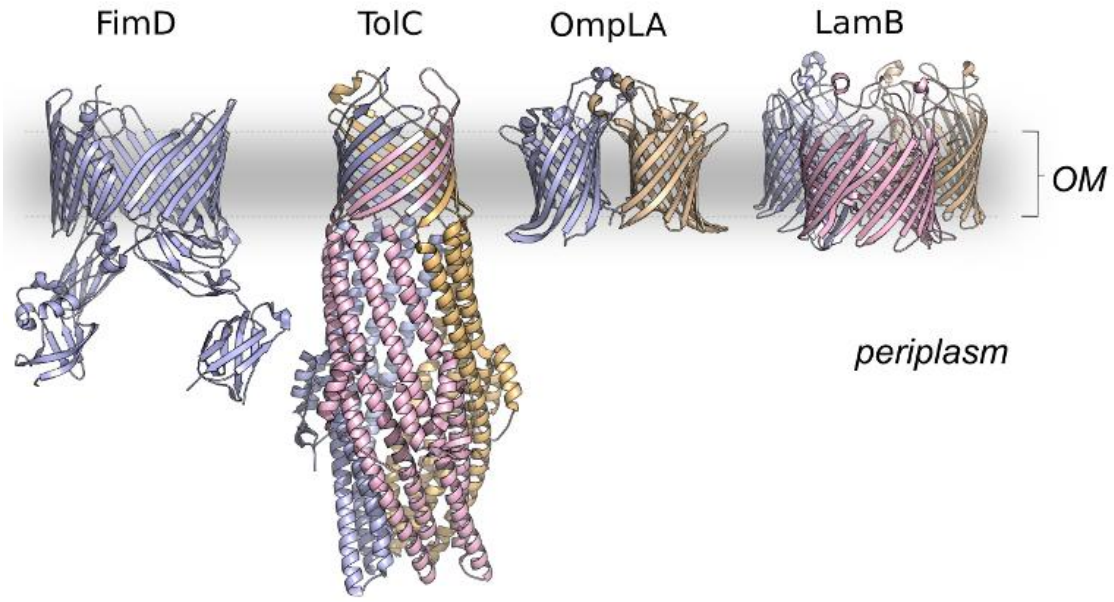


Figure 1-2 Diversity of Bacterial OMP Structure

Some examples of *E. coli* OMPs are shown. OMPs can be found either as a monomer (e.g. FimD; PDB: 3RFZ), an oligomer where multiple subunits come together to form one β-barrel (e.g. TolC; PDB: 1EK9), or an oligomer where each subunit creates its own β-barrel (e.g. OmpLA and LamB; PDB: 1QD6 and 1MAL).

Typically, the transmembrane β-barrels of OMPs contain an even number of β-strands, ranging from 8 to 24, arranged in an antiparallel fashion. In many OMPs, the strands making up the β-barrel domain contain alternating hydrophilic and hydrophobic amino acids. Because sidechains of successive residues on a β-strand point outwards on alternating faces of the β-barrel, the hydrophilic residues are largely found lining the barrel interior whereas the hydrophobic residues are found on the outer surface of the barrel. As a result, the hydrophobic residues of OMPs face towards the membrane, where they interact with the hydrocarbon tails of the OM lipids via hydrophobic interactions (Buchanan, 1999; Koebnik et al., 2000).

The consecutive β-strands of an OMP β-barrel domain are connected by alternating tight turns and longer loops. While the shorter turns are found on the periplasmic face of the β-barrels, the longer and more flexible loops are usually found on the exoplasmic face (Fairman et al., 2011; Koebnik et al., 2000). In many OMPs, the

exoplasmic loops serve functionally important roles. For example, permeability properties, such as the exclusion limit and ion selectivity of many porins (OMPs that form diffusion pores), are determined by an exoplasmic loop that folds back into the β-barrel interior. The exoplasmic loops of OMPs can also be involved in the formation of the substrate binding site, as in the case of the ferrichrome (iron-peptide complex) binding site of an *E. coli* receptor protein FhuA (Buchanan, 1999).

Depending on the protein, OMPs may or may not have soluble domains extending away from the β-barrel. A good example of an OMP that exhibits a large soluble domain is the *E. coli* exporter protein TolC (Figure 1-2). In addition to a β-barrel domain, TolC has a long tunnel shaped periplasmic domain that interacts with other periplasmic and IM components involved in the export process (Koronakis et al., 2000; Zgurskaya et al., 2011). In other OMPs, the soluble domains serve different functions such as substrate binding (e.g. the pilus assembly factor FimD; Figure 1-2) or gating the β-barrel pore activity by acting as a plug (e.g. the iron transporter TbpA) (Oke et al., 2004; Phan et al., 2011).

OMPs exist in various oligomeric states. The β-barrel domains of most OMPs are formed by a single polypeptide chain; in some cases, however, multiple chains can assemble to form one large β-barrel structure. For example, the transmembrane domain of the efflux pump TolC is formed by three monomers that contribute four β-strands each to make a 12-stranded β-barrel (Koronakis et al., 2000). In other OMPs, oligomerization of individually formed β-barrel monomers may be required for function, as is the case for the maltose transporter LamB which forms a homotrimer (Koebnik et al., 2000). Oligomerization can also serve as a means of enzyme activity regulation. In *E. coli*, OmpLA, which hydrolyzes OM phospholipids, becomes active only after calcium induces its dimerization (Dekker et al., 1997).

1.2.2. OMP Functions

The genes encoding OMPs account for approximately 2-3% of Gram-negative bacterial genomes (Wimley, 2003). Despite having similar β-barrel structures, OMPs are functionally very diverse. They carry out many important cellular functions including

nutrient uptake, protein secretion, and antibiotic resistance (Bos et al., 2007; Wimley, 2003). All known and putative *E. coli* OMPs and their functions are listed in Appendix A. Selected OMPs representative of different functional categories are discussed below.

Porins are the most abundant proteins in the OM of Gram-negative bacteria. Typically found as homotrimers of 16-stranded β-barrel subunits, porins function as channels to allow diffusion of molecules between the bacteria and their environment (Delcour, 2003; Nikaido, 1994). There are two classes of porins: non-specific and specific porins. Non-specific porins allow passive diffusion of a variety of small hydrophilic molecules. Although termed 'non-specific', the size of the pore and the polarity of residues lining the pore interior limit the size and charge of the molecule that can pass through (Delcour, 2003; Fairman et al., 2011). Examples of non-specific porins include OmpF (with a preference for large positively charged molecules) and PhoE (with a preference for negatively charged molecules) (Fairman et al., 2011; Koebnik et al., 2000). Specific porins are different from non-specific porins in that they only allow the transport of their specific substrates. For instance, BtuB is specific for the uptake of vitamin B12, FhuA for iron, LamB for maltose and other sugars, and Tsx for nucleosides. While some specific porins allow passive diffusion of solutes upon contact (e.g. LamB and Tsx), others depend on the TonB complex of the IM to provide energy for active transport of the substrate (e.g. BtuB and FhuA) (Nikaido, 1994). Sometimes, specific porins are referred to as channels rather than porins to distinguish them from non-specific porins (Krewulak and Vogel, 2011; Nikaido, 2003; Postle and Kadner, 2003).

In addition to accommodating the import of molecules through porins, the OM allows certain molecules to be exported from the cell as well. This export or secretion process is facilitated by a class of OMPs known as translocons. Various types of translocons exist in the OM. TolC is an OMP involved in the Type I secretion pathway, allowing the export of proteins, small molecules and drugs (thereby contributing to antibiotic resistance) (Zgurskaya et al., 2011). Other instances of OMPs acting as translocons are those involved in the two-partner secretion pathway (a component of the Type V secretion pathway) (Jacob-Dubuisson et al., 2001). In this pathway, the OMP translocon specializes in exporting only one specific substrate. A well-known example is FhaC from *Bordetella pertussis*, which is involved in the export of filamentous hemagglutinin, an adhesin secreted during infection (Clantin et al., 2007; Jacob-

Dubuisson et al., 2001). Some secreted proteins can be transported outside of the cell without an additional OMP channel. In these cases, the proteins contain their own C-terminal β-barrel domain that acts as a transporter for the N-terminal passenger protein that is to be secreted. Because of their ability to transport themselves, these proteins are appropriately named autotransporters. Autotransporters are often associated with pathogenicity, as the passenger proteins usually function as virulence factors. Examples of autotransporters include adhesins such as AIDA-I and Ag43, and proteases such as Hbp and Pet (Desvaux et al., 2004; van Ulsen, 2011).

Aside from the transport of molecules into and out of the cell, there are other functions that take place at the OM, some of which are carried out by enzymatic OMPs. The most well studied cases of OMPs with enzymatic functions are OmpLA, OmpT, and PagP. OmpLA (phospholipase A) hydrolyzes the acyl ester bonds in phospholipids. Its active site is located in the LPS-containing outer leaflet of the OM, where it can detect the presence of phospholipids that disrupt the asymmetry of the OM (Dekker et al., 1997; Snijder et al., 1999). OmpT is a protease that specifically cleaves between two basic residues of a protein, with substrates shown to include antimicrobial peptides released by host immune responses (Stumpe et al., 1998; Sugimura and Nishihara, 1988; Vandeputte-Rutten et al., 2001). Finally, PagP transfers a palmitate chain from a phospholipid in the inner leaflet of the OM to the Lipid A component of a LPS molecule in the outer leaflet (Bishop, 2005).

The final category of OMPs conains proteins that contribute to the formation and integrity of the OM, and thus play a structural role. These OMPs include proteins involved in peptidoglycan formation and OMP assembly, as well as usher proteins that transport and polymerize pili subunits required for cell motility. Examples include Mipa (peptidoglycan synthesis), BamA (OMP folding and membrane insertion), LptD (LPS assembly), OmpX (adhesion and entry into host cells), and FimD (an usher to transport and polymerize subunits of the Type I pili) (Okuda and Tokuda, 2011; Phan et al., 2011; Vogt and Schulz, 1999; Vollmer et al., 1999; Voulhoux et al., 2003).

As described above, OMPs serve a wide range of functions in Gram-negative bacteria. In summary, OMPs assist the OM in fulfilling its role as a protective physical barrier that regulates the traffic of molecules between the bacterium and its

surroundings. Furthermore, some OMPs play direct roles in virulence, thus extending the importance of studying the bacterial OMP structure, function and biogenesis from basic scientific interest to medical significance.

1.2.3. OMP Biogenesis and Degradation

Biogenesis of all prokaryotic proteins begins in the cytosol where they are first synthesized, but all non-cytosolic proteins including OMPs need to be transported to their final destinations before they can fulfill their given functions (Schatz and Dobberstein, 1996). While proper protein targeting and folding is important, protein degradation and quality control is equally critical. Below, and in Figure 1-3, is a summary of the OMP maturation and degradation processes.

Bacterial OMPs are initially synthesized in the cytosol with a cleavable N-terminal signal sequence required to target the protein to the IM (Gierasch, 1989; von Heijne, 1990). As the N-terminal signal sequence emerges from the ribosome, it is first bound by trigger factor (TF), a chaperone that ensures that the nascent OMP polypeptide remains in a stable unfolded state (Hoffmann et al., 2010). The TF then passes the OMP over to the cytoplasmic chaperone SecB, which targets the protein to the Sec translocon (SecY, SecE, SecG, SecD, SecF, YajC) at the IM (Bechtluft et al., 2010). Once at the IM, SecB transfers the pre-OMP to SecA, a homodimeric ATPase that associates with the Sec translocon (Cross et al., 2009; Zimmer et al., 2008). SecA aids the translocation of the unfolded OMP across the IM by threading the OMP through the SecYEG channel in an ATP-dependent manner (Kusters and Driessen, 2011). As the OMP begins to emerge into the periplasmic space, the N-terminal signal sequence is recognized and removed by signal peptidase I (SPaseI), releasing the OMP into the periplasm (Paetzel et al., 2002).

Once released into the periplasm, the OMPs are transported across the periplasmic space to the OM via either the SurA pathway or the Skp/DegP pathway (Rizzitello et al., 2001; Sklar et al., 2007b). SurA, Skp, and DegP are periplasmic chaperones that keep the proteins in a protected unfolded state to prevent misfolding and aggregation (Bitto and McKay, 2003; Patel et al., 2009; Volokhina et al., 2011).

Previous studies have shown that the SurA and the Skp/DegP pathways function in parallel, and that cells are viable when either one of the pathways are missing (Rizzitello et al., 2001). The SurA pathway is believed to play a more important role under normal conditions, while the role of the Skp/DegP pathway appears to become more prominent when the cells are under stress (Sklar et al., 2007b). Regardless of the pathway taken, the journey of all OMPs ends at the OM, which is their final destination. In OMPs, the OM targeting information is intrinsically contained in their C-terminal sequence (i.e. it is not removed like the N-terminal signal sequence) (Robert et al., 2006).

After the OMPs reach the OM, protein folding and membrane insertion take place in a concerted manner (Tamm et al., 2001; Tamm et al., 2004). *In vitro* studies have shown that OMPs are able to fold and insert themselves spontaneously into synthetic phospholipid bilayer membranes without help from any proteinaceous machineries (Surrey and Jahnig, 1992). This suggests that OMP folding does not require an external energy source, and that the information for folding is encoded in their amino acid sequence. However, the folding occurs too slowly *in vitro* to be biologically relevant; hence, *in vivo* OMP folding and membrane insertion require an OMP assembly factor known as the β-barrel assembly machinery (BAM) complex to increase the kinetics of the process (Tamm et al., 2004). The OMP substrates are recognized by the BAM complex via their C-terminal OM targeting signal (Robert et al., 2006). The exact mechanism of OMP folding and insertion by the BAM complex is not well understood yet, but a detailed discussion of current models is provided later in Chapter 8 (Discussion) of this thesis.

As important as the proper OMP synthesis and assembly is the elimination of damaged or misfolded OMPs. In this regard, protein degradation is an essential component of quality control, especially when cells are under stress. Stress on the OMP synthesis pathway such as overproduction can cause OMPs to become misfolded, aggregated, or mislocalized. Therefore, there are systems in place to remove the defective OMPs from the synthesis pathway. When OMPs are mislocalized and not correctly targeted to the BAM complex, their C-termini activate the DegS protease which initiates a cascade of events in the sigmaE pathway resulting in the increased expression of several periplasmic chaperones including SurA, Skp and DegP. This lowers the stress put on the OMP synthesis pathway, preventing further mislocalization

(Merdanovic et al., 2011). Similarly, any damaged or misfolded OMPs are recognized by DegP, which uses its proteolytic function to initiate degradation (Merdanovic et al., 2011; Sklar et al., 2007b).

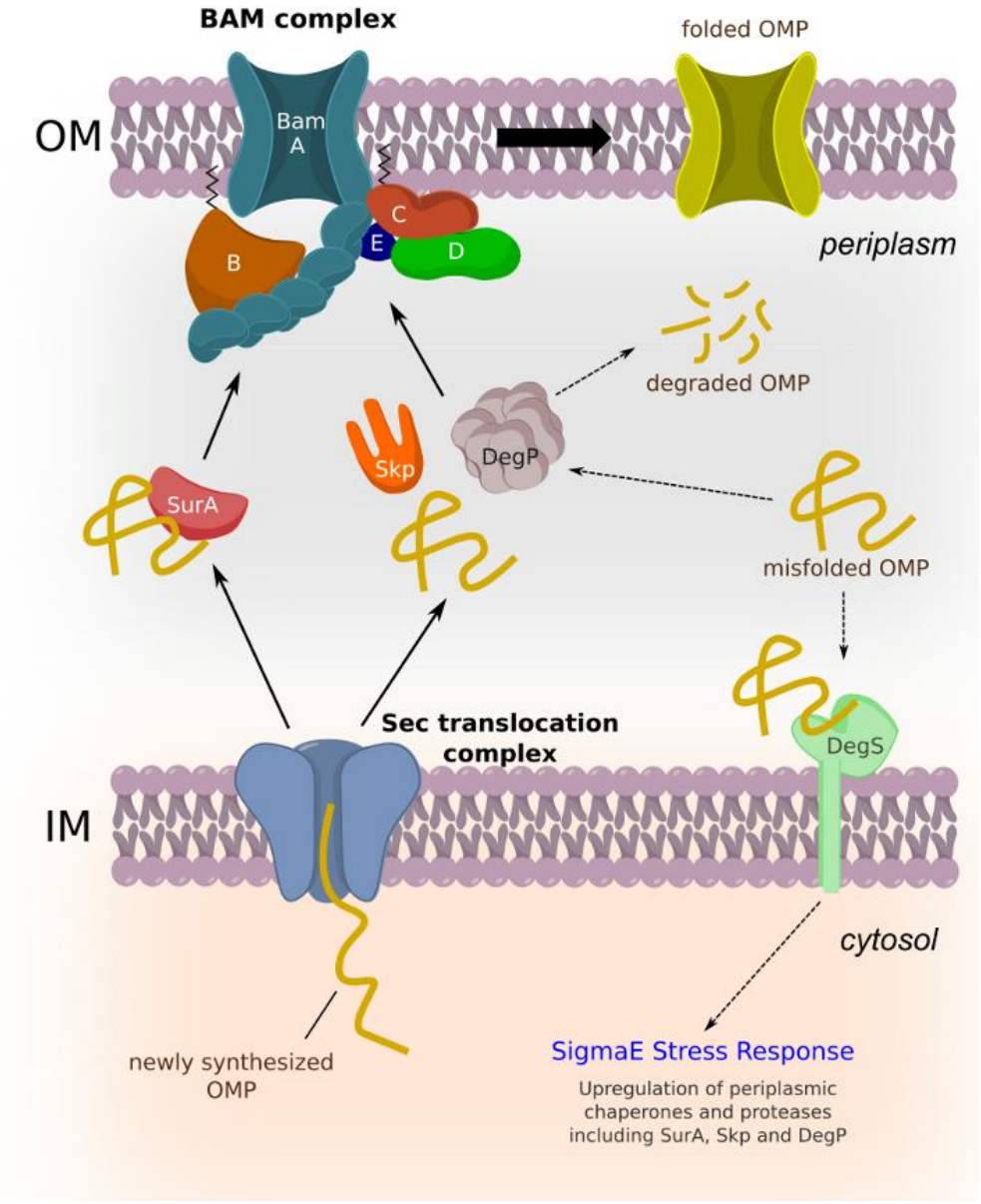


Figure 1-3 OMP Biogenesis and Degradation

Following initial synthesis in the cytosol, an OMP is transported across the IM via the Sec translocation complex. The protein is released into the periplasm where it can take either the SurA or the Skp/DegP pathway to reach the BAM complex of the OM. By an unknown mechanism, the BAM complex mediates the folding and membrane insertion of the OMPs. Misfolded or aggregated OMPs in the periplasm are recognized and degraded by DegP, or bound by DegS which initiates the sigmaE stress response.

1.3. OMP Assembly in Bacterial Outer Membrane

How OMPs are assembled inside living cells has only recently begun to be understood. Prior to the discovery of the BAM complex in 2003, much of the knowledge about OMP folding came from *in vitro* studies. Observed *in vitro* OMP folding behaviours have provided valuable insights into understanding how folding takes place *in vivo*. In the following sections, the main findings from *in vitro* studies and the initial discovery of the BAM complex are summarized.

1.3.1. Early in vitro Studies of OMP Folding

Studying protein folding *in vitro* has the advantage of allowing controlled experiments to be carried out on isolated proteins. This facilitates close examination of factors which influence the folding behaviour of a protein, one at a time. In the case of OMPs, thermodynamic and kinetics studies have provided valuable insights into the biochemical properties and folding mechanisms of OMPs, even before the discovery of the BAM complex.

It has been demonstrated many times that some OMPs can spontaneously fold and insert themselves into a lipid bilayer in the absence of an energy source or a folding factor. For example, when the eight-stranded β-barrel protein OmpA was denatured in 6-8M urea, it subsequently refolded into a lipid bilayer when the urea concentration was reduced via rapid dilution (Surrey and Jahnig, 1992). Since then, OmpA has been frequently used as a model to study the folding mechanism of OMPs and β-barrel proteins in general.

The technique that has made the greatest contribution towards understanding the OMP folding mechanism is the 'time-resolved distance determination by fluorescence quenching (TDFQ)' procedure. Briefly, TDFQ detects the depth of a fluorophore of an OMP in a lipid bilayer consisting of lipids carrying a fluorescence quencher group (Kleinschmidt, 2003; Kleinschmidt, 2006; Tamm et al., 2001). This technique has made it possible to detect OmpA folding intermediates and to monitor the membrane insertion of OmpA over time. The TDFQ technique, combined with kinetics analysis, has revealed that OmpA folding exhibits three distinct kinetic phases (Figure

1-4): 1) the fastest and temperature independent phase attributed to the initial binding of unfolded OmpA to the membrane surface, 2) a slower and strongly temperature dependent phase that corresponds to a deeper but still only partial insertion of OmpA into the membrane, and 3) the slowest phase of β-barrel maturation that was only observed to occur at temperatures greater than 30°C (Tamm et al., 2004). These and other *in vitro* studies suggest that OMP folding (i.e. acquiring inter-strand hydrogen bonds and forming β-hairpins) and membrane insertion are coupled and that both occur in a concerted manner (Tamm et al., 2001; Tamm et al., 2004).

In vitro studies of OMP folding has also enabled direct examination of the way in which the lipid bilayer composition influences the folding behavior and stability of OMPs. Since OMPs are membrane proteins and their folding takes place as they translocate into the OM, their folding kinetics must be significantly influenced by the chemical and physical properties of the lipid bilayer. Bacterial membranes are mostly composed of PE, with smaller amounts of PG and CL. Calculated free energy of urea-induced unfolding of OmpA has been shown to increase as increasing amounts of PE were included in a bilayer composed primarily of PC (i.e. more difficult to unfold). Higher PG and CL content was also observed to increase the OmpA stability in the membrane (Tamm et al., 2004). One may wonder whether LPS, which is exclusively found on the outer leaflet of OM in bacteria, has any effect on OmpA folding. Surprisingly, inclusion of LPS into membranes has been reported to inhibit OmpA folding and insertion (Tamm et al., 2004); however, this result is not conclusive because LPS was incorporated in both leaflets of the membrane in the reported experiment, whereas the OM is asymmetric and contains LPS only on the outer leaflet.

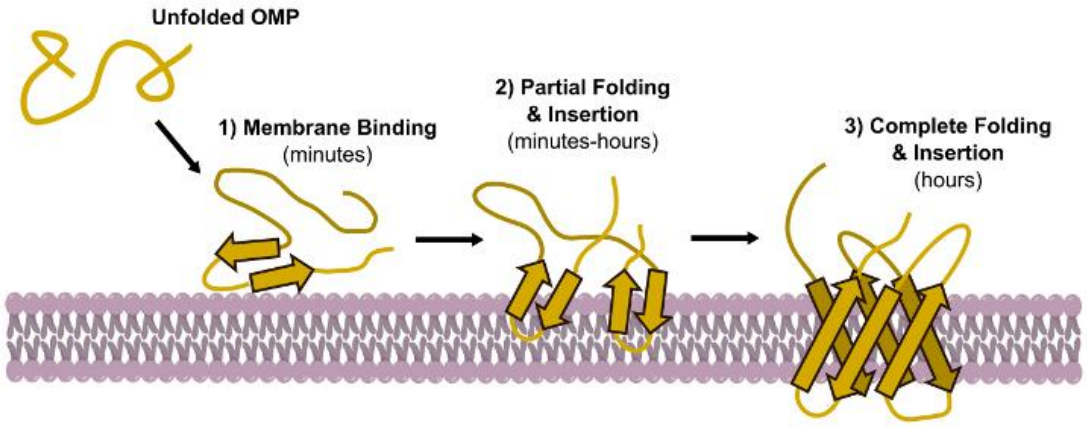


Figure 1-4 OMP Folding Model Based on in vitro Studies

Based on *in vitro* studies (Kleinschmidt, 2003; Kleinschmidt, 2006; Tamm et al., 2001), OMP folding takes place in three distinct kinetic phases: 1) binding to the membrane surface, 2) partial folding and insertion into the membrane, and 3) completion of folding and membrane insertion.

1.3.2. Discovery of the BAM Complex

If OMPs are fully capable of folding themselves into membranes without any help, why do Gram-negative bacteria require the BAM complex for their OMP assembly? There are two major reasons as to why cells would require a protein complex to facilitate the folding and membrane insertion of OMPs. First, the kinetics of folding and insertion of OMPs *in vitro* is too slow to be biologically plausible. While pulse-chase experiments show that newly synthesized OMPs assemble in the OM within about 30 seconds to few minutes, the time for *in vitro* OMP folding completion was in the order of hours (Tamm et al., 2004). Furthermore, the spontaneous insertion of OMPs *in vitro* does not explain the exclusive insertion of these proteins into the OM, and not the IM. OMPs have been shown to fold into membranes of various lipid compositions *in vitro*, but they are only found in the OM in cells (Tamm et al., 2004). For these reasons, a proteinaceous machinery (i.e. the BAM complex) is thought to be required in order to increase the kinetics and specificity of the OMP folding process in living cells.

BamA (formerly known as YaeT or Omp85) of the BAM complex was the first OMP assembly factor to be discovered in 2003 (Voulhoux et al., 2003). BamA was

initially identified in *Neisseria meningitidis*, but its homologues have since been found to exist in all Gram-negative bacteria, as well as in endosymbiotically derived eukaryotic organelles, namely the mitochondria and chloroplasts (Walther et al., 2009). The strong conservation of the gene encoding BamA, together with the observation that the protein is essential for cell viability, reinforces the fundamental importance of BamA and its role in Gram-negative bacteria.

The involvement of BamA in OMP assembly is well supported by various experimental studies. The first clue that suggested BamA could be involved in OMP biogenesis was the location of its gene within the bacterial genome. The *Omp85* gene coding for BamA is located immediately adjacent to the *Skp* gene, which encodes a periplasmic chaperone Skp (Gentle et al., 2004). As well, the *Omp85* gene is found in a gene cluster together with genes that are involved in the biosynthesis of LPS, which is exclusively found on the outer leaflet of the OM (Genevrois et al., 2003). The involvement of BamA in OMP assembly became clearer when a periplasmic accumulation of unfolded OMPs was observed in BamA-depleted strains of Gram-negative bacteria (Voulhoux et al., 2003). A direct involvement of BamA in OMP biogenesis was further supported by a protein interaction study, which demonstrated that BamA can bind unfolded OMPs *in vitro* (Knowles et al., 2008; Voulhoux et al., 2003).

Following the discovery of BamA and its involvement in an unknown OMP assembly pathway, it was later found, via co-immunoprecipitation studies, that *E. coli* BamA exists in a larger protein complex with four other lipoproteins, namely BamB, C, D, and E (Malinverni et al., 2006; Sklar et al., 2007a; Wu et al., 2005). BamA, along with the lipoproteins, are now known as the BAM complex, and its structure and functional mechanism are only now beginning to be understood.

It should be noted that different protein subunits of the BAM complex show different degrees of conservation across different species of Gram-negative bacteria. In proteobacteria, for example, BamA and BamD are found ubiquitously, while BamB and BamE are absent in δ-proteobacteria and ε-proteobacteria. BamC, which is the least conserved component of the BAM complex, is only found in β-proteobacteria and γ-proteobacteria (Anwari et al., 2012). It should also be noted that a recent study has idenfitied yet another lipoprotein, BamF, as a part of the BAM complex. The N-terminal

region of BamF shares some sequence homology to BamC, and interestingly it is only found in α-proteobacteria in which BamC is absent (Anwari et al., 2012). As not much is known about the structure and function BamF at present, the main focus of this review hereafter will be on BamA, B, C, D and E.

1.3.3. BAM Complex Structure and Function

(This is a general introduction to the BAM complex. In each chapter of this thesis, a more detailed and relevant introduction to the structure and function of each BAM complex component will be given.)

The proper functioning of the BAM complex is indispensable not only for the proper OMP assembly, but also for the survival of Gram-negative bacteria (Bos et al., 2007; Gentle et al., 2005). Since its discovery, the BAM complex has attracted much attention from the research community mainly for its potential to serve as a novel antibiotic target, as well as the expectation that it will advance our understanding of the β-barrel membrane protein folding. The fact that systems homologous to the BAM complex exist in mitochondrial and chloroplastic OMs of eukaryotes has further fueled the research effort.

In *E. coli*, five different proteins assemble to form the BAM complex, and together they ensure the proper incorporation of OMPs into the OM. The components of the BAM complex are named alphabetically from BamA to BamE in order of decreasing molecular mass. BamA, the largest and the first discovered component, is an OMP itself that adopts a β-barrel fold to span the OM lipid bilayer (Gentle et al., 2005). In addition to the transmembrane β-barrel domain, BamA also contains a soluble periplasmic domain that consists of five polypeptide-transport-associated (POTRA) motifs (Sanchez-Pulido et al., 2003). BamB, C, D, and E, on the other hand, are lipoproteins, meaning they are soluble proteins anchored to the periplasmic surface of the OM by diacylglycerol linked to the N-terminal cysteine (See Appendix B for more information on lipoprotein biogenesis) (Hagan et al., 2011; Ricci and Silhavy, 2011). The association of the lipoproteins with BamA has been shown by co-immunoprecipitation analysis; however, the exact stoichiometry of a functional unit of the BAM complex is not clear.

Of the five proteins making up the BAM complex, only BamA and BamD are essential for cell viability. Both BamA and BamD depletion strains display severe defects in OMP assembly, as indicated by reduced levels of OMPs in the OM and accumulation of unfolded OMPs in the periplasm (Onufryk et al., 2005; Voulhoux et al., 2003). Deletion of the gene encoding BamB also results in significant defects in OMP assembly, although not lethal. Cells lacking BamB show increased membrane permeability as indicated by higher susceptibility to antibiotics, and they are unable to assemble many OMPs, especially those forming larger β-barrels (Charlson et al., 2006; Ruiz et al., 2006). The absence of BamC or BamE causes mild defects in OMP assembly and increased membrane permeability (Onufryk et al., 2005; Sklar et al., 2007a). Based on the observed knockout phenotypes, it appears that BamA and BamD function in the most critical steps of OMP assembly process, while BamB, C, and E play roles in improving the efficiency of the BAM complex.

When this thesis project was first initiated in 2006, there was only a limited amount of structural and functional information on the BAM complex. At that time, no structural data was available for any of the BAM complex subunits and the lipoprotein components of the BAM complex had just been discovered without any functional roles assigned to them. Six years later, thanks to the efforts from various research groups, we now have a much better understanding of the structural organization of the BAM complex and the functional roles of the individual components. This thesis summarizes a PhD project that was carried out in the Paetzel Lab between the year 2006 and 2012. During the course of this thesis project, I solved the structures of all four lipoprotein components (i.e. BamB, C, D and E). The protein-protein interaction studies complementing the structural investigation also provided valuable insights into how the proteins interact with each other within the BAM complex.

1.4. Research Objectives

The aim of this thesis project was to obtain detailed structural and biochemical information about each protein component of the BAM complex using *E. coli* as the model organism. The three major questions that this thesis attempts to answer are:

i. What are the structures of the individual BAM complex components?

ii. What functional insights can be gained from the structure of each protein?

iii. How are the proteins structurally organized in the BAM complex, and what are the implications?

To answer these questions, the following experiments were carried out in parallel, with emphasis on the structural analysis:

1) Cloning, overexpression, purification and oligomeric state analysis of the individual BAM complex components (Chapter 2-7).

2) Detection and characterization of protein-protein interactions between the components of the BAM complex (Chapter 3).

3) Structure determination of the individual components and subcomplexes of the BAM complex by X-ray crystallography and NMR (Chapter 4-7).

2. General Materials and Methods

This chapter describes the general experimental approach, common procedures used throughout the entire thesis project and background information on structure determination processes by X-ray crystallography and NMR. More detailed methods specific for individual experiments are provided later in the subsequent chapters.

2.1. Overview of Experimental Approaches

As stated earlier, the main goal of this thesis was to determine the structures of the individual proteins and subcomplexes of the *E. coli* BAM complex. For protein structure determination by X-ray crystallography or NMR, a large amount of pure protein sample is required. The initial stage of the thesis project was therefore focused on cloning the five BAM complex components, and subsequently optimizing their purification protocols to obtain high yields of the proteins. Individually cloned, overexpressed and purified BamA, B, C, D and E were then subjected to either X-ray crystallography or NMR studies. Of these five proteins, the crystal structures of BamB and BamC, as well as the NMR structure of BamE, were successfully determined.

While the structural analyses of the individual BAM proteins were in progress, protein-protein interaction studies were carried out concurrently in attempt to better understand how the five proteins associate with each other within the BAM complex. Utilizing protein engineering, various chromatography and dynamic light scattering analysis, the protein network within the BAM complex was dissected and the formation of various subcomplexes were detected and analyzed. This interaction study also helped to identify subcomplexes suitable for crystallization, eventually leading to the successful structure determination of the BamCD heterodimer.

2.2. Cloning

2.2.1. List of the Constructs

For structural and protein-protein interaction studies of the BAM complex, various protein constructs were created during this thesis project. These constructs can be divided into two categories: 1) full length proteins that encompass the complete sequence of the mature proteins as found *in vivo,* and 2) truncated forms of the proteins created for improving their likelihood of crystallization and for testing the effects of the missing domains on their abilities to interact with the other BAM complex subunits.

It should be noted that none of the constructs contain the N-terminal signal sequence to mimic the mature forms of the proteins existing in cells. Also, for the lipoprotein constructs (BamB, C, D and E), the N-terminal cysteine following the N-terminal signal sequence (the site of lipidation) has been removed as well, in order to prevent undesired intermolecular disulfide bond formation. Table 2-1 lists all the constructs created for the work presented in this thesis. The rationales behind the design of truncated constructs are explained later in the relevant chapters.

Table 2-1 List of the Constructs Generated

Construct	Description	Residues (start-end)	Calculated Molecular Mass (kDa)	Affinity Tag
$BamA_{POTRA}$	POTRA domains 1-5	21-433	49	His_{x6} (N-term)
BamB	full length	21-392	42	His_{x6} (N-term)
BamC	full length	26-344	36	His_{x6} (N-term)
$BamC_N$	N-terminal domain	99-217	14	His_{x6} (C-term)
$BamC_C$	C-terminal domain	220-344	15	His_{x6} (C-term)
$BamC_{NC}$	unstructured N-terminal region missing	94-344	29	His_{x6} (N-term)
$BamC_{UN}$	C-terminal domain missing	26-217	22	His_{x6} (C-term)
BamD	full length	21-245	28	His_{x6} (N-term)
BamE	full length	21-113	13	His_{x6} (N-term)

* The molecular masses were calculated by ProtParam (Gasteiger et al., 2005).

2.2.2. General Cloning Procedure

For all the constructs listed in Table 2-1, DNA fragments coding for the desired regions of the proteins were amplified from *E. coli* K12 genomic DNA by PCR using forward and reverse primers. All the primers contained appropriate restriction enzyme sites so that the PCR products could be ligated into the vector of choice. For fusion with the N-terminal or C-terminal hexahistidine tags, the vector pET28a (Novagen) and pET24a (Novagen) were used, respectively. The subsequent DNA sequencing results (Macrogen) confirmed that the gene inserts cloned into the appropriate vectors matched the sequences reported in the Swiss-Prot database. The list of primers, vectors and corresponding UniProt protein sequence IDs for all the constructs created are provided in Appendix C.

2.3. Protein Overexpression

Each expression plasmid coding for the protein of interest was transformed into *E. coli* BL21(λDE3). The transformed cells were grown overnight and used to inoculate (1:100 back dilution) two liters of Luria Bertani (LB) medium containing kanamycin (50 µg/mL) on the following day. Cultures were grown at 37 °C until the OD_{600nm} reached 0.6. The protein overexpression was then induced with 1 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG) for 3 hours. The cells were then harvested by centrifugation (5 minutes at 6000 x g), and the resulting cell pellet was stored in -80 °C freezer until further use.

2.4. Protein Purification

For the purification of the overexpressed proteins, the cell pellet containing the desired overexpressed protein was lysed using an Avestin Emulsiflex-3C cell homogenizer in buffer A (20 mM Tris-HCl, pH 8.0; 100 mM NaCl). The resulting lysate was clarified by centrifugation (29,000 x g) for 30 minutes at 4°C, and the overexpressed proteins were initially purified by nickel affinity chromatography. The protein was eluted from the nickel affinity column with a step gradient method (100-500 mM imidazole in

buffer A in 100 mM increments). Presence of the desired protein in the elution fractions were confirmed by SDS-PAGE. The elution fractions containing the protein were then pooled and concentrated to approximately 10 mg/mL using an Amicon ultra-centrifugal filter device (Millipore).

The concentrated protein sample was further purified by size-exclusion chromatography (Sephacryl S-100 HiPrep 26/60 column) in buffer A on an ÄKTA Prime system (GE Healthcare) at 4°C and a flow rate of 1 mL/min. The purity of the purified protein was once again confirmed by SDS-PAGE, and its concentration was measured by NanoDrop (Thermo Scientific) using the extinction coefficients listed in Appendix C. The purified proteins were stored at 4°C until further use.

2.5. Methods of Protein Structure Determination

The purified proteins were subjected to structural studies by X-ray crystallography and NMR spectroscopy. The details of the crystallization methods, X-ray diffraction and NMR data acquisition, and the structure solving processes for the individual structures solved are provided later in the relevant chapters. Here, instead, are the descriptions of the basic theoretical principles behind these structure determination techniques.

2.5.1. X-ray Crystallography

X-ray crystallography is the primary method for determining protein structures at an atomic resolution. When a beam of X-ray is directed to a protein crystal, the resulting diffraction pattern produced by the crystal contains information that can be used to generate the three-dimensional structure of a protein (Smyth and Martin, 2000). According to the Protein Data Bank (PDB) Annual Report in 2011, 92% of the biomolecule structures deposited in the PDB were determined by X-ray crystallography.

2.5.1.1. What is a Crystal?

A crystal is a regular three-dimensional repeat of molecules. The smallest building block of a crystal is known as the ‘asymmetric unit’, and it can be one molecule,

one subunit of a repeating unit within a multimeric protein, or more than one molecule. Within a crystal, the asymmetric unit can be rotated, and translated by crystal symmetry operations (aka. space group symmetry operation) to build a 'unit cell'. The unit cells are then repeated infinitely (using translation operation only) in three dimensions to generate the whole crystal (Figure 2-1) (Chayen and Saridakis, 2008; Blow, 2002). A protein crystal can be described by its unit cell properties and the type of symmetry existing within it. The dimensions of a unit cell are defined by six numbers: the lengths of the three axes (*a*, *b*, and *c*) and the three interaxial angles (α, β, γ). Each crystal falls into one of the 14 Bravais lattices depending on the shape of the unit cell and the type of crystal lattice. On the other hand, the symmetry type of a crystal is defined by its space group that denotes how the asymmetric units within the crystal are related by symmetry. The symmetry operations found in protein crystals are rotation, translation, and screw axis (combination of rotation and translation) (Smyth and Martin, 2000; Wlodawer et al., 2008). The Bravais lattice and space group of a given crystal are represented by a Hermann-Mauguin notation. For example, $P2_12_12$, crystal has a primitive lattice (P), and two-fold screw axes parallel to the *a* (2_1) and *b* (2_1) axes, and a two-fold rotation axis (2) along the *c* axis.

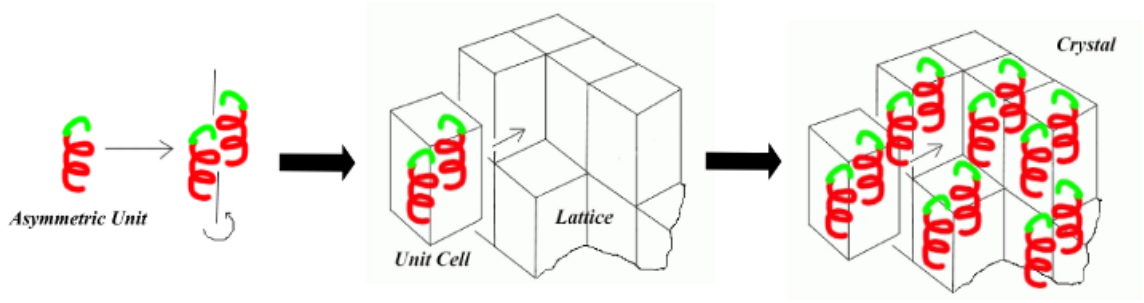


Figure 2-1 Structure of a Crystal

A crystal is made up of asymmetric units that arrange to form a unit cell, which in turn is repeated in all three dimensions to create the entire crystal structure. In this particular example, two asymmetric units that are related to each other by a two-fold rotational symmetry form a single unit cell. This unit cell is then repeatedly duplicated and translated in all directions to form the crystal lattice and generate a crystal. This figure was adapted from an online source (http://www.ruppweb.org/xray/101index.html).

2.5.1.2. Protein Crystallization

The first step of structure determination by X-ray crystallography is protein crystallization. This step is often a major bottleneck of the whole structure determination process, but crystals are necessary because the scattering power of a single protein molecule would be too weak to detect. A huge number of molecules are arranged in the same orientation within a crystal, and this causes scattered X-ray waves to add up in phase and raise the signal to a measurable level. The most common approach of protein crystallization is using vapor diffusion methods to lower the solubility of the protein gradually, by addition of precipitants and controlled evaporation of the solvent that the protein is dissolved in. A crystal must grow to a sufficient size (usually larger than 0.1 mm in all dimensions) before it can be used to obtain a diffraction data (Chayen and Saridakis, 2008). The procedures that follow (from data collection to structure determination) are summarized below and also in Figure 2-2.

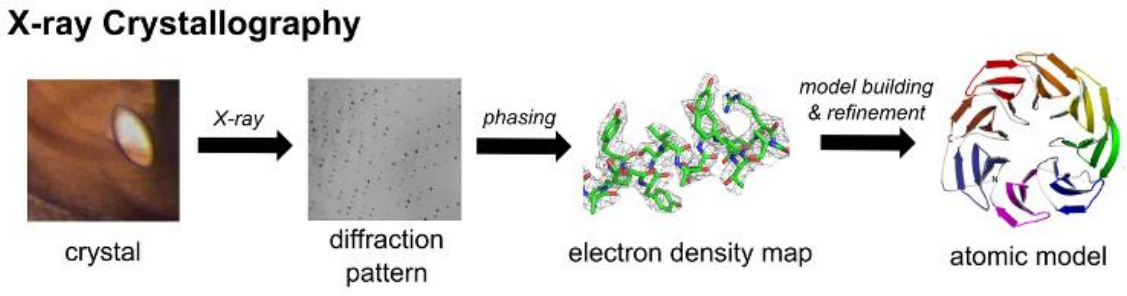


Figure 2-2 Protein Structure Determination by X-ray Crystallography

The workflow of protein structure determination process by X-ray crystallography is shown.

2.5.1.3. Data Collection

Once a diffraction quality crystal is obtained, the crystal is irradiated with a beam of monochromatic (single wavelength) X-ray. A typical wavelength of X-ray used is ~1-1.5 Å, which is on the scale of chemical covalent bond lengths (e.g. C-C bond length = 1.54 Å) and is therefore suitable for determining atomic resolution structures. The incoming beam of X-ray is scattered by electrons of atoms constituting the protein crystal. These scattered X-ray beams are detected by detector such as CCD (charge-coupled device) or image plate detector, resulting in a two-dimensional image containing

the diffraction pattern of regularly spaced spots known as reflections. Diffraction images are recorded at different rotations of the crystal to obtain a complete dataset (Smyth and Martin, 2000; Wlodawer et al., 2008; Blow, 2002). Crystals with higher internal symmetry require less diffraction images to be collected because the internal symmetry of a crystal is reflected in the resulting diffraction pattern. That is, crystals with higher symmetry result in more symmetrically equivalent (redundant) reflections being recorded on the collected images. A data collection strategy is therefore essential so that all reflections are recorded with the least exposure of X-ray dose.

Today, much of the X-ray diffraction data collection is carried out at synchrotrons. The synchrotron X-ray beam is not only more intense and refined than rotating anode X-ray generators used in laboratories, but it can also be adjusted to a specific wavelength via monochromators. This can be useful for certain phasing experiments that require multiple data sets to be collected at various X-ray wavelengths (discussed later in the 'phasing' section).

2.5.1.4. Data Processing

The initial analyses of the diffraction data are carried out in three steps: 1) indexing, 2) integration, and 3) scaling and merging. Indexing refers to extracting unit cell and symmetry (space group) information from the collected diffraction images. The information about unit cell dimensions, Bravais lattice and space group is contained in the geometrical arrangement of the diffraction spots. Also during the indexing step, the reflections on the images are identified and designated (i.e. indexed) with three integers (h, k, l = Miller Indices). In the following integration step, the intensities of these indexed reflections are measured. By this stage, hundreds of images have been converted into a list of Miller Indices (each signifying one particular reflection) and the corresponding intensities (Smyth and Martin, 2000; Wlodawer et al., 2008). After all the reflection intensities are tabulated, addressing and removing the systematic errors inherent in a dataset becomes a necessity. The errors arise, for example, from unavoidable experimental parameters such as radiation damage to the crystal during data collection. As each image is affected differently, each image (and its reflections) needs to be placed on a common relative scale that will compensate for the 'systematic' error. This last stage of data processing is referred to as 'scaling'. During this step, the intensities of

equivalent reflections (same reflections from multiple data sets or symmetrically equivalent reflections) are 'scaled' and then 'merged' (replacing multiple intensity values with the weighted average). The closer the intensity values are between the equivalent reflections (indicated by low R_{merge} value; see Appendix D for how it is calculated), the higher the data quality (Chayen and Saridakis, 2008).

2.5.1.5. Phasing

As described earlier, information about the unit cell and crystal symmetry can be obtained from the relative positions of the recorded reflections. The information about the unit cell content, on the other hand, is mirrored in the measured intensities of the reflections. Accurate estimate of a parameter known as the structure factor allows the calculation of the electron distribution in the unit cell (see Appendix D). The structure factor is a function of the amplitude and the phase of the diffracted X-ray waves. However, unlike the amplitude which can be obtained from the measured intensity, the phase information cannot be easily measured nor calculated. This inherent problem in X-ray crystallography is known as the 'phase problem' (Smyth and Martin, 2000; Wlodawer et al., 2008; Blow, 2002).

There are different ways one can overcome the phase problem. For a protein that has homologues with known structures, the initial phase can be estimated by performing molecular replacement procedures. Molecular replacement utilizes the Patterson function, which is the Fourier transform (a mathematical algorithm; see glossary for definition) of the measured intensities. The Fourier transform of structure factors reveals the electron density and *vice versa* (Blow, 2002). However, without the phase component, structure factor cannot be calculated. Instead, the Fourier transform of reflection intensities (instead of the structure factor) results in a Patterson map (instead of an electron density map). A Patterson map is an interatomic vector map with peaks that are unique to each crystal and dataset, and it can also be generated from a model (e.g. homologue protein with known structure) oriented in an identical unit cell. During molecular replacement, the model Patterson map is calculated from different orientations (rotations and translations) of the model protein until it closely matches the dataset Patterson map. This way, the orientation of protein in question in the unit cell

can be estimated, and one can now move onto the phase refinement step (discussed in the next section) (Smyth and Martin, 2000; Wlodawer et al., 2008; Rhodes, 2006).

When there is no homologue structure available (i.e. when determining a novel structure), different methods must be employed to solve the phase problem. The most common approaches are the MAD (multiple-wavelength anomalous dispersion) and SAD (single-wavelength anomalous dispersion) phasing experiments, both of which involves incorporating anomalous scatterers such as selenium into proteins before crystallization. This is usually done by synthesizing proteins in minimal media with seleno-methionine (instead of methionine). Anomalous X-ray scattering results when heavy atoms scatter at a specific wavelength that is near the absorption edge of the heavy atom. The X-ray diffracted by the anomalous scatterer has a phase that is unique from the rest of the atoms in a crystal due to strong X-ray absorbance, and this phenomenon allows calculation of the amplitude and phase contribution by the selenium atoms in the crystal and to estimate their locations within the asymmetric unit and unit cell. Once their locations are known, the phase for each reflection can be more accurately estimated and an initial density map can be generated. (Smyth and Martin, 2000; Wlodawer et al., 2008; Rhodes, 2006).

2.5.1.6. Model Building and Structure Refinement

Once initial phases are obtained, and thereby an initial electron density map, a protein structure model can be built by fitting polypeptide chains with the known sequence into the electron density map. After the initial model is built, the phases from this model can be applied to the original data to calculate structure factors and generate an improved electron density map. This leads to an improved model, and this improved model can be used for another round of phase refinement to yield a better electron density map and model. This cycle continues until the difference between theoretical structure factors back calculated from the model and the experimentally derived structure factors narrows until acceptable error estimate is reached. (Smyth and Martin, 2000; Wlodawer et al., 2008). In X-ray crystallography, this correlation between the data and the model is usually measured by statistical indicators known as R-factor and R-free (refer to Appendix D for how they are calculated). The R-factor measures how well a model predicts the entire experimental dataset, while the R-free measures how well a

model predicts a subset of experimental reflections that were randomly set aside and not used in refinement. A 'good' model refined to a 2 Å data should have an R-factor and R-free that are approximately 0.2 or below (Wlodawer et al., 2008).

2.5.2. NMR Spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy is a powerful method for studying the structure and dynamics of proteins in solution. It involves acquiring various multidimensional NMR spectra that are analyzed to generate a list of atoms close to each other in space. Based on these distance restraints, a protein structure can then be calculated (Figure 2-3) (Kanelis et al., 2001; Kay, 2005).

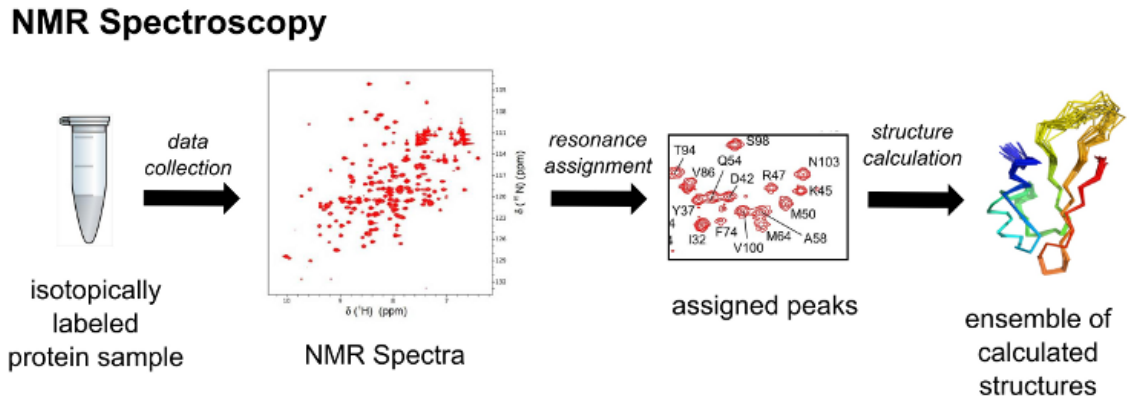


Figure 2-3 Protein Structure Determination by NMR

The workflow of protein structure determination process by NMR is shown.

2.5.2.1. Nuclear Magnetic Resonance (NMR)

NMR spectroscopy relies on the phenomenon of nuclear magnetic resonance. All atoms that contain an odd number of protons and/or neutrons have an intrinsic magnetic moment. When such an atom is placed in a magnetic field and a pulse of electromagnetic radiation is applied, the nucleus of the atom absorbs and re-emits the electromagnetic radiation of a particular frequency known as the resonant frequency. The resonant frequency differs between different atoms due to the chemical environments they are in (e.g. the neighbouring atoms). In NMR spectroscopy, a resonant frequency of a particular nucleus relative to a standard is known as the chemical shift, and it is what is being measured during the experiments (Kanelis et al.,

2001; Wagner et al., 1992). By convention, chemical shifts are expressed in units of ppm (parts per million).

2.5.2.2. Isotopic Labeling

The first step of protein structure determination by NMR spectroscopy is sample preparation. In protein molecules, the only atoms that are 'detectable' by NMR spectroscopy are the hydrogens (i.e. only atom that has an intrinsic magnetic moment). However, one dimensional ^{1}H-NMR spectrum of even a small protein contains overcrowded peaks that are difficult to interpret in any comprehensive manner. Therefore, the second and third dimensions are added to alleviate the spectral crowding and overlap by labeling proteins with isotopes that are detectable by NMR spectroscopy (Kay, 2005). The most common approach is to uniformly label proteins with ^{13}C and ^{15}N. In multidimensional NMR, each proton (H atom) is distinguished not only by its own chemical shift but also by the chemical shifts of its neighbouring atoms. For a typical NMR experiment, approximately ~500 uL pure protein sample at ~1 mM concentration is prepared.

2.5.2.3. NMR Data Collection

Once the protein sample is ready, various multidimensional NMR spectra are recorded. Typically the first step is to record a two-dimensional ^{1}H-^{15}N-HSQC (heteronuclear single quantum correlation) spectrum. This spectrum contains several peaks, each of which corresponds to a hydrogen atom bound to an amide nitrogen atom. In other words, these peaks represent the backbone amide hydrogens of the protein and the sidechain amide hydrogens of amino acids such as Asn and Gln. The number of peaks on a HSQC spectrum is therefore expected to be approximately equal to the number of residues in the protein. Because each protein produces a unique pattern of peaks, the ^{1}H-^{15}N-HSQC spectrum can be thought of as the fingerprint of the protein (Kay, 2005). Following the HSQC experiment, various three-dimensional NMR spectra are recorded that yield information about the chemical shifts of sidechain atoms. (Kanelis et al., 2001).

2.5.2.4. Resonance Assignment

Following the data collection that may take several days, the peaks (representing chemical shifts) on each spectrum are assigned to specific atoms. That is, it must be determined which chemical shift corresponds to which atom of the protein so that different atoms of the protein can be recognized by their distinct chemical shifts. This task is known as the resonance assignment, and is the most time-consuming part of the whole structure determination process by NMR. (Kanelis et al., 2001; Kay, 2005). The backbone amide nitrogens and hydrogens of individual residues are typically assigned first, followed by the backbone carbons and sidechain atoms.

2.5.2.5. Generation of Restraints

Once the resonance assignment is complete, the next step is to generate distance restraints to be used in the structure calculation. To do this, the data from a three-dimensional spectrum obtained by a type of NMR spectroscopy known as NOESY (nuclear overhauser effect spectroscopy) is used. Each peak found on the NOESY spectrum is due to the transfer of magnetization between two nuclei through space. The position of a particular peak (chemical shift) helps identify which two atoms give rise to that peak, while the intensity of the peak contains information about the distance between the two atoms. In addition to the distance restraints, angle restraints can also be generated to aid the structure calculation process. These experimentally determined restraints are then used as input for the structure calculation process (Kanelis et al., 2001; Kay, 2005; Wagner et al., 1992).

2.5.2.6. Structure Calculation

For the final model building step, computer programs are used to calculate a protein structure that satisfies as many of the restraints as possible, in addition to retaining general properties of proteins such as optimal bond lengths and angles. Unlike X-ray crystallography that results in one structure at the end of the experiment, an NMR experiment results in an ensemble of structures. How well the calculated structures converge (i.e. low RMSD values) is an indicator of the quality of the NMR data and the accuracy of resonance assignment (Kanelis et al., 2001; Wagner et al., 1992).

2.6. Structural Analysis

Once the structures of the BAM complex proteins were determined by X-ray crystallography and NMR spectroscopy, various programs and online tools were used to analyze their structures in detail. Below is the list of the programs used:

PROCHECK (Laskowski et al., 1993)
Analysis of the stereochemistry of the determined structures

DALI server (Holm et al., 2008)
Structural homologue search

COOT (Emsley and Cowtan, 2004)
Superposition of two or more different proteins for structural comparison and measurement of atomic distances and angles

PISA server (Krissinel and Henrick)
Identification of protein-protein interaction interfaces

PyMol (DeLano)
Analysis of the surface electrostatics

Consurf (Glaser et al., 2003)
Mapping conservation onto a protein structure

CASTp (Dundas et al., 2006)
Detection and measurement of potential substrate binding cavities

3. Protein-Protein Interaction Analysis of the BAM Complex

Note regarding contributions:

The work presented in this chapter was a collaborative effort with Suraaj Aulakh (M.Sc. candidate) and Jonathan Tan (undergraduate research student, Spring 2012) of the Paetzel Lab. I supervised Jonathan's undergraduate research project, from which data for section 3.3.2 and 3.3.3 were generated. Suraaj co-supervised Jonathan's project with me, and she also contributed by standardizing and generating the standard curve for the size-exclusion chromatography studies summarized in this chapter.

3.1. Introduction

BamA, B, C, D and E of the BAM complex must work together towards the same goal of facilitating the folding and membrane insertion of OMPs. How these five proteins interact with each other to form the higher-order structure and coordinate the OMP assembly in cells still remains unclear.

The first insight into the oligomeric organization of the BAM complex was provided by the Silhavy group at Princeton University. In 2005, two years following the discovery of BamA and its involvement in OMP assembly, the Silhavy group showed that BamA, B and C can be co-immunoprecipitated in *E. coli* (Wu et al., 2005). Shortly after, they identified BamD as another component of the complex, also via co-immunoprecipitation (Malinverni et al., 2006). The BamA/B/C/D interaction was further dissected by performing pull-down experiments using strains defective in one of the lipoproteins. In the BamB or BamC depletion strains, the ability of the rest of the Bam proteins to co-purify was not affected. In the BamD depletion strain, however, only BamA/B (and not BamC) could co-purify. Taken together, these results indicated that BamB and BamD interact directly but independently with BamA. Furthermore, the inability of BamC to co-purify with the rest of the proteins in the absence of BamD suggests that BamC binds to BamD directly but not to BamA (Malinverni et al., 2006). The last component of the BAM complex to be identified in 2007 was BamE. Absence of BamE has been reported to compromise the stability of the BAM complex, resulting in relatively less BamD being co-immunoprecipitated by BamA (Sklar et al., 2007a).

Based on the work by the Silhavy group, BamB and BamC/D/E are believed to interact with BamA independently. The POTRA domains of BamA have been suggested to serve as the docking site for the lipoproteins because of their location in the periplasm (Kim et al., 2007). As mentioned earlier, the periplasmic region of BamA consists of five POTRA domains, numbered 1 to 5 from the N-terminus (Figure 3-1A). Several structures of the POTRA domains have been reported in recent years. They show that the five POTRA domains, despite low sequence similarity, have the same overall structure with each domain consisting of a three-stranded β-sheet and two α-helices (Arnold et al., 2010; Gatzeva-Topalova et al., 2008; Gatzeva-Topalova et al., 2010; Kim et al., 2007; Knowles et al., 2008; Zhang et al., 2011). The protein-protein interaction interfaces

between adjacent POTRA domains keep POTRA1-2 and POTRA3-4-5 rigid; however, a flexible linker between POTRA2 and POTRA3 results in the POTRA domains existing either in a bent or an extended form (Figure 3-1B) (Gatzeva-Topalova et al., 2008; Kim et al., 2007).

Studies involving deletions of individual POTRA domains have shown that BamB requires POTRA 2-5 for association with BamA, while BamC/D/E requires only POTRA 5 (Kim et al., 2007). The POTRA domains have also been shown to interact with the periplasmic chaperone SurA (via POTRA 1) and with unfolded OMPs (via POTRA 1-2) (Figure 3-1C) (Bennion et al., 2010). However, stoichiometry and the mode of these interactions are not yet fully understood. A more detailed understanding of the association between the POTRA domains and its interacting partners could reveal clues about the structural organization and the molecular mechanism of the BAM complex.

This chapter summarizes the *in vitro* protein-protein interaction experiments performed with the aim of better understanding the stoichiometry of the interactions between the POTRA domains and the lipoproteins. These experiments also helped to identify various stable BAM subcomplexes that are suitable for crystallization trials.

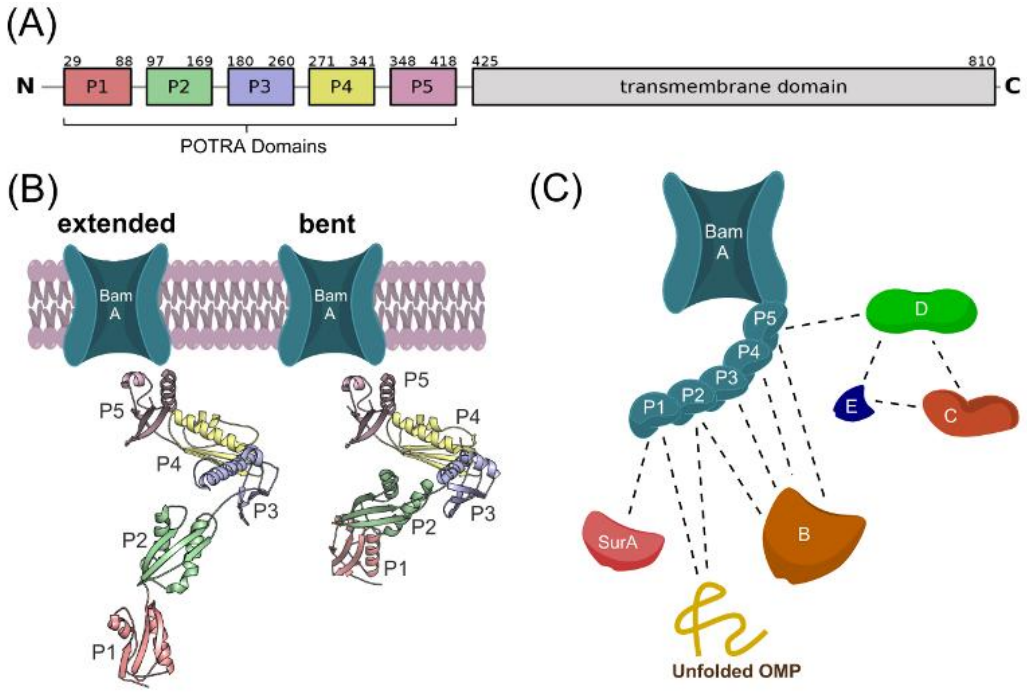


Figure 3-1 (legend on next page)

Figure 3-1 POTRA Domains of BamA and Their Interacting Partners

(A) The domain structure of BamA shows an N-terminal periplasmic region that contains five POTRA domains along with a C-terminal transmembrane domain. (B) Models of full length POTRA domains of BamA in bent and extended conformations are shown. The models were prepared by extending the POTRA1-4 bent (PDB: 2QDF) and extended (PDB: 3EFC) structures by a structure alignment of the overlapping POTRA4 domain of POTRA4-5 structure (PDB: 3Q6B). (C) Experimentally shown interactions of the POTRA domains of BamA with the lipoproteins (BamB, C, D and E), as well as with the periplasmic chaperone SurA and substrate (unfolded OMP) are indicated by the black dotted lines.

3.2. Materials and Methods

3.2.1. Protein Overexpression and Purification

All protein constructs used in this study ($BamA_{POTRA}$, BamB, BamC, BamD, BamE) were cloned, overexpressed and purified as previously described in Chapter 2. For co-purification of two or more proteins, separate cell pellets (each resulting from 2 L culture) overexpressing each protein were combined prior to lysis. The combined cells were then lysed and subsequently purified by nickel affinity chromatography by following the same protocol as described in Chapter 2.

3.2.2. Protein-Protein Interaction Analysis

For determining the oligomeric states of the individual BAM proteins and detecting the formation of complexes between different BAM proteins, size-exclusion chromatography was employed as the main method of choice. In some cases where the size-exclusion chromatography result produced ambiguous data, dynamic light scattering analysis and nickel affinity pull-down experiments were carried out.

3.2.2.1. Analytical Size-exclusion Chromatography

Size-exclusion chromatography separates protein mixtures based on size. To test whether or not a specific mixture of BAM proteins form a complex, various combinations of histidine-tagged proteins co-lysed and purified by nickel affinity chromatography were analyzed by size-exclusion chromatography using the Sephacryl S-100 HiPrep 26/60 column on an ÄKTA Prime system (GE Healthcare). All experiments were carried out using Buffer A (20 mM Tris-HCl, pH 8.0, 100 mM NaCl). For a typical

run, a 4-5 mL protein sample at a concentration of 3-5 mg/mL was injected and resolved at a flow rate of 1 mL/min.

The elution profile on the resulting chromatogram was compared to that of the individually run proteins in order to detect any peak shifts. Accompanying SDS-PAGE gel helped identify the proteins giving rise to each occurring peaks on the chromatogram. To aid comparison, each chromatogram was scaled in the range of 0 to 1. The molecular mass calculation based on the elution volume of a protein was carried out using the standard curve (Appendix E). The size-exclusion column was standardized by eluting a series of marker proteins of known molecular mass. The protein markers used were: ribonuclease A (13.7 kDa), chymotrypsinogen A (25.0 kDa), ovalbumin (43.0 kDa), albumin (67.0 kDa) and blue dextran (2000 kDa).

3.2.2.2. Dynamic Light Scattering Analysis

Dynamic light scattering is a technique that can be used for accurately determining the molecular mass of a molecule. Because the measurement is not affected by the molecule's shape, the dynamic light scattering analysis was used to complement the size-exclusion chromatography analysis for some experiments.

For each run, a protein sample was loaded onto a Superdex 200 column in-line with a multiangle light scattering system (Wyatt Technologies Inc). The molecular mass of the protein was determined by a multiangle light-scattering (MALS) Dawn-EOS instrument with a 684 nm laser (Wyatt Technologies, Inc.) coupled to a refractive index instrument (Optilab Rex; Wyatt Technologies, Inc.). The molecular mass was calculated from the observed light scattering intensity and differential refractive index using ASTRA v5.1 software (Wyatt Technologies, Inc.) based on Zimm fit method using a refractive index increment, dn/dc = 0.185 L g^{-1} (Zimm, 1948).

3.2.2.3. Nickel Affinity Chromatography

Nickel affinity chromatography was used to detect an interaction between $BamA_{POTRA}$ carrying a histidine tag and untagged BamB. Two proteins (both carrying N-terminal histidine tags) were purified separately via nickel affinity and size-exclusion chromatography as described in Chapter 2. The fractions from the size-exclusion chromatography run that contained BamB monomers were pooled and incubated with

thrombin at room temperature overnight. The digested BamB sample was run on a nickel affinity column to ensure the removal of the histidine tag. BamB recovered from the flowthrough fraction was then incubated with approximately equal amount of purified $BamA_{POTRA}$ overnight. The protein mixture was run on nickel affinity column to test whether the untagged BamB can co-elute with $BamA_{POTRA}$, and the elution fractions were analyzed by SDS-PAGE.

3.3. Results

3.3.1. Oligomeric States of the Individual BAM Subunits

Prior to examining the ability of the BAM subunits to form subcomplexes, the oligomeric states of the individual proteins in solution were investigated. The POTRA domains 1 to 5 of BamA (residues 21-433, referred to as $BamA_{POTRA}$ hereafter), BamB, BamC, BamD and BamE were individually purified by nickel affinity chromatography and subsequently analyzed with size-exclusion chromatography. The resulting size-exclusion chromatograms are shown in

Figure 3-2, and the apparent molecular masses of the individual proteins based on their elution volumes are summarized in Table 3-1.

$BamA_{POTRA}$, BamC and BamD showed a single major peak on their size-exclusion chromatograms at elution volumes of 125 mL, 123 mL and 155 mL, respectively (

Figure 3-2A, C and D). Based on their elution volumes and the standard curve (Appendix E), the apparent molecular masses of $BamA_{POTRA}$, BamC and BamD were calculated to be 57 kDa, 55 kDa and 29 kDa, respectively. The theoretical molecular mass of BamD (28 kDa) closely matches the measure value from size-exclusion chromatography (29 kDa), therefore suggesting that BamD exists in solution as monomers. However, there were significant differences between the theoretical and measured molecular masses for $BamA_{POTRA}$ and BamC. In both cases, the measured molecular masses were larger than the theoretical values, but not large enough for them to be in oligomeric states. For $BamA_{POTRA}$, the measured and theoretical molecular

masses are 57 kDa and 49 kDa, respectively. For BamC, the difference is even larger (measured = 55 kDa; theoretical = 36 kDa).

Since the molecular mass calculation based on the elution volume assumes that the molecule adopts a globular structure, it is possible that the larger than expected molecular masses for $BamA_{POTRA}$ and BamC are due to these proteins adopting non-globular conformations. Indeed, the crystal structure of $BamA_{POTRA}$ shows that the POTRA domains have an elongated structure (Figure 3-1). As for BamC, it was subjected to dynamic light scattering analysis that allows determination of molecular mass independent of the molecule's shape. The molecular mass of BamC determined by dynamic light scattering analysis was 35.9 kDa, which is very close to the theoretical value (36 kDa). Taken together, these results show that $BamA_{POTRA}$ and BamC exist as monomers in solution and that they both have non-globular structures.

Unlike $BamA_{POTRA}$, BamC and BamD, the size-exclusion chromatograms of BamB and BamE showed multiple peaks (

Figure 3-2B and E). The chromatogram for BamB shows three major peaks, and SDS-PAGE analysis revealed that they all correspond to the elution of BamB at different times. The first peak eluted at the void volume (98 mL), suggesting it may be aggregated proteins or possibly an oligomeric form of BamB (trimer or larger, since the cut-off molecular mass for the particular column used is 100 kDa and each BamB monomer is 42 kDa). The other two peaks occur at the elution volumes of 106 mL and 135 mL, with the later peak being significantly larger. The molecular masses calculated based on these elution volumes are 82 kDa and 44 kDa, and they closely resemble the theoretical molecular masses of BamB dimer (84 kDa) and monomer (42 kDa). BamB therefore exists in both monomeric and dimeric form in solution, with the former being predominant.

Similar to BamB, the elution profile of BamE suggests that it exists in both monomeric and dimeric forms. The two major peaks observed on the chromatogram of BamE (occurring at 149 mL and 202 mL) correspond to the calculated molecular masses of 32 kDa and 10 kDa. These values are similar to the theoretical masses of BamE dimer (26 kDa) and monomer (13 kDa). The molecular masses of the individual species

of BamE were also determined by dynamic light scattering analysis, and they are in good agreement with the size-exclusion chromatography data (dimer = 28.5 kDa; monomer = 12.4 kDa). Unlike BamB, however, the distribution between the monomeric and dimeric species seems to be similar (estimated from the band intensity on the SDS-PAGE gel and the sizes of the peaks).

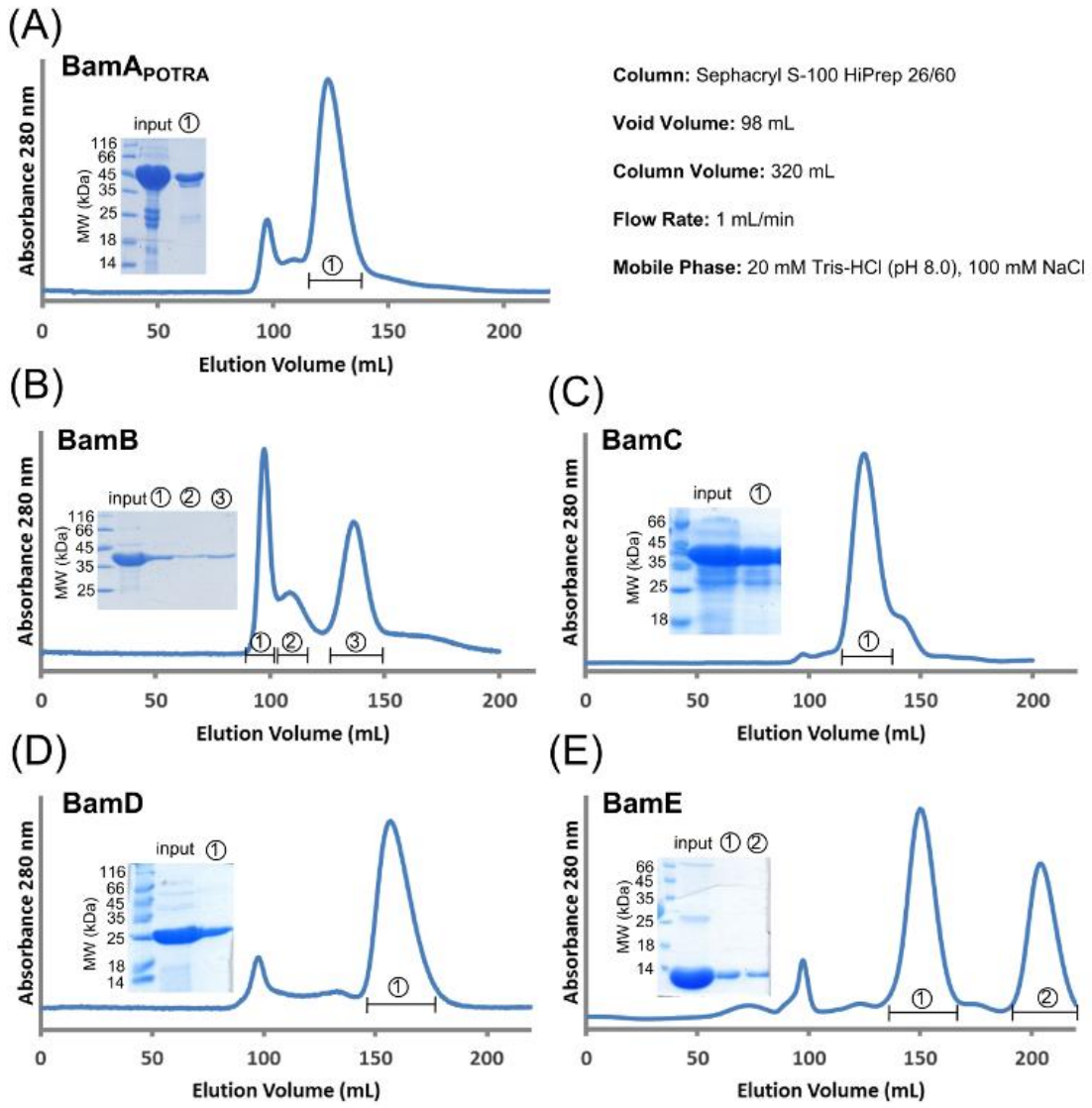


Figure 3-2 Size-Exclusion Chromatograms for the Individual BAM Proteins

Size-exclusion chromatograms and the corresponding SDS-PAGE gels are shown for (A) BamA$_{POTRA}$, (B) BamB, (C) BamC, (D) BamD and (E) BamE. Different peaks on each chromatogram are labeled, and the proteins that elute at the corresponding regions are shown on the SDS-PAGE gels.

Table 3-1 Oligomeric State Analysis of the Individual BAM Proteins

	Elution Volume (mL)	Molecular Mass (kDa)			Probable Oligomeric State
		Size-Exclusion Chromatography	Dynamic Light Scattering	Calculated*	
$BamA_{POTRA}$	125	57	n/a	49	monomer
BamB	106	82	n/a	84	dimer
	135	44	n/a	42	monomer
BamC	123	55	35.9 ± 0.4	36	monomer
BamD	155	29	27.6 ± 0.1	28	monomer
BamE	149	32	28.5 ± 3.0	26	dimer
	202	10	12.4 ± 0.8	13	monomer

* The molecular masses of the proteins were calculated by ProtParam (Gasteiger et al., 2005).

3.3.2. $BamA_{POTRA}$ – BamB Interaction

Previous co-immunoprecipitation studies have suggested that BamB interacts with BamA independently of BamC/D/E (Kim et al., 2007; Malinverni et al., 2006). To test if the POTRA domains of BamA are sufficient for interaction with BamB to form a stable $BamA_{POTRA}$-BamB complex, the cells overexpressing the two proteins (both carrying N-terminal hexahistidine tags) were co-lysed prior to purification by nickel affinity chromatography followed by size-exclusion chromatography. The resulting size-exclusion chromatograms and the corresponding SDS-PAGE gels were analyzed to determine if $BamA_{POTRA}$ and BamB co-eluted as a hetero-oligomeric complex.

The size exclusion chromatogram for $BamA_{POTRA}$-BamB shows the presence of two major peaks with a shoulder preceding the second peak (Figure 3-3A). The elution volumes at which the first, shouldering and second peaks occur are approximately 97 mL, 115 mL, and 125 mL. The molecular masses calculated from the standard curve for these elution volumes are >100 kDa (first peak), ~68 kDa (shouldering peak), and ~55 kDa (second peak). The SDS-PAGE gel shows that the first peak, which elutes at the void volume, contains BamB; however, based on the early elution in the void volume, this region most likely contains aggregated BamB or possibly BamB with higher

oligomeric structure. A similar observation was also made when BamB was purified alone (

Figure 3-2B). From the SDS-PAGE gel, both $BamA_{POTRA}$ and BamB can be seen to be present in regions corresponding to the second peak and the shouldering region preceding it. Although this may suggest possible heterodimer formation, the molecular mass calculated from the standard curve (range of 55-68 kDa for the second peak and its shoulder) are significantly distant from the theoretical molecular mass of 91 kDa for a $BamA_{POTRA}$-BamB heterodimer. Furthermore, although BamB is observed to elute at least 20 mL earlier in the presence of $BamA_{POTRA}$, the elution volume of POTRA was not affected by the presence of BamB. Thus, while the elution volume shift of BamB suggests possible $BamA_{POTRA}$-BamB dimer formation, the unchanged elution profile of $BamA_{POTRA}$ suggests otherwise.

To investigate the $BamA_{POTRA}$-BamB interaction (or the lack thereof) further, the ability of $BamA_{POTRA}$ carrying a hexahistidine tag to pull down untagged BamB via nickel affinity chromatography was tested. Previously purified BamB (in monomeric form) was digested by thrombin in order to remove the N-terminal hexahistidine tag. The removal of the tag was confirmed by passing the digested BamB over a nickel affinity column (Figure 3-3B) and recovering untagged BamB from the flow-through fraction. BamB was then incubated overnight with an approximately equal amount of purified tagged $BamA_{POTRA}$. This mixture was then loaded onto a nickel affinity column, and the resulting SDS-PAGE gel shows that BamB co-eluted with $BamA_{POTRA}$ in the elution fractions containing 100 mM and 300 mM imidazole (Figure 3-3C). Taken together with the size-exclusion chromatography data that shows a significant shift of elution volume for BamB in the presence of $BamA_{POTRA}$, this nickel affinity chromatography experiment provides convincing evidence that the POTRA domains of BamA are sufficient for interaction with BamB, and that they form a stable complex in solution.

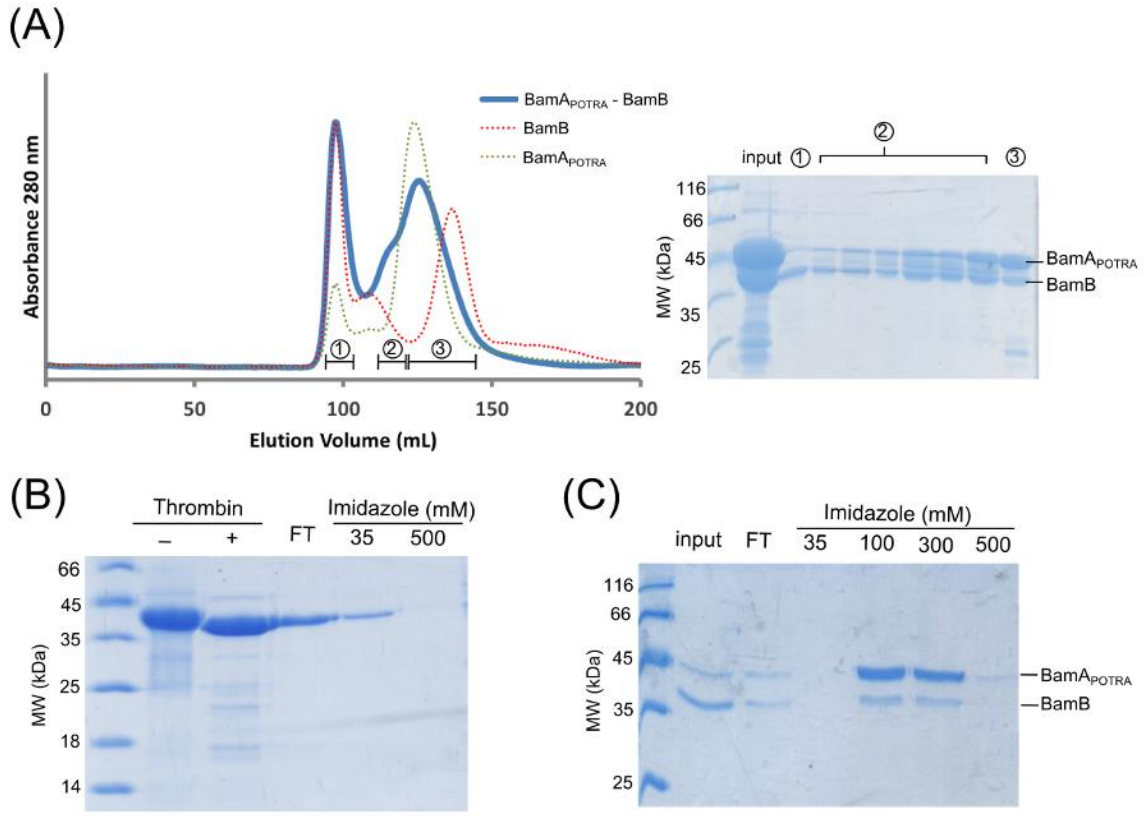


Figure 3-3 Formation of the BamA$_{POTRA}$-BamB Complex

(A) The size-exclusion chromatogram of BamA$_{POTRA}$-BamB (blue) is shown overlaid with those of BamB (red) and BamA$_{POTRA}$ (green) for comparison. The experimental conditions were the same as shown in Figure 3-2. Three different regions on the chromatogram are labeled, and the proteins that elute at the corresponding regions are shown on the SDS-PAGE gel (right). The fractions from region 1 and 3 were pooled before SDS-PAGE analysis, whereas the fractions from region 2 (shouldering peak) were run on the SDS-PAGE gel separately. (B) The SDS-PAGE gel of BamB digested by thrombin. After the removal of the hexahistidine tag, BamB was passed over a nickel affinity column and most of it was recovered from the flowthrough fraction (labeled FT in the figure). (C) The SDS-PAGE gel of nickel affinity chromatography result for untagged BamB incubated with His$_{x6}$-BamA$_{POTRA}$.

3.3.3. BamA$_{POTRA}$ – BamD Interaction

As mentioned earlier, BamC/D/E has been shown by co-immunoprecipitation studies to interact with BamA independently from BamB. More specifically, only BamD among BamC/D/E has been shown to interact directly with BamA (Malinverni et al., 2006). To test if the POTRA domains of BamA are sufficient for a stable BamA$_{POTRA}$-BamD complex formation, the same co-lysis and chromatographic experiments were carried out as for BamA$_{POTRA}$ and BamB.

Following co-lysis and nickel affinity purification, a mixture sample of $BamA_{POTRA}$ and BamD were loaded onto a size-exclusion column. On the resulting chromatogram, three distinct peaks were observed at the elution volumes of 97 mL, 122 mL, and 149 mL (Figure 3-4A). Based on the standard curve, these elution volumes correspond to molecular masses of >100 kDa, ~59 kDa, and ~33 kDa, respectively. From the SDS-PAGE results, the first peak that eluted in the void volume shows the presence of different proteins most likely in an aggregated form. On the other hand, the regions corresponding to the second and the third peaks contain predominantly $BamA_{POTRA}$ and BamD, respectively. These two peaks on the $BamA_{POTRA}$-BamD chromatogram closely match the individual elution profiles of $BamA_{POTRA}$ and BamD. Although BamD elutes slightly earlier in the presence of $BamA_{POTRA}$, the SDS-PAGE gel clearly shows that the third peak contains BamD only. It appears, therefore, that $BamA_{POTRA}$ and BamD are not able to form a stable heterodimeric complex in solution.

As BamE has previously been suggested to play a role in stabilizing the BAM complex structure (Sklar et al., 2007a), cells overexpressing BamE were co-lysed with those overexpressing $BamA_{POTRA}$ and BamD to test if it will result in $BamA_{POTRA}$-BamD-BamE complex formation. The size-exclusion chromatogram obtained by following the same procedure as above displays four peaks occurring at the elution volumes of 97 mL, 123 mL, 149 mL and 198 mL, corresponding to molecular masses (based on the standard curve) of >100 kDa, ~57 kDa, ~33 kDa and 12 kDa (Figure 3-4B). From the SDS-PAGE gel, the first peak on the chromatogram contains various unknown proteins as well as BamE that is possibly in an aggregated form due to its elution at the void volume. The third peak on the chromatogram contains predominantly BamD (and relatively small amount of BamE dimers), while the latest eluting peak contains predominantly BamE monomers. Interestingly, however, the second peak (the largest) contains all three proteins ($BamA_{POTRA}$, BamD and BamE) and therefore suggests formation of the $BamA_{POTRA}$-BamD/E heterotrimeric complex. However, the molecular mass corresponding to the elution volume of this peak (~57 kDa) is distant from the theoretical value of 90 kDa for the heterotrimer. Also, the elution volume of $BamA_{POTRA}$ remained unchanged in the presence of the other two proteins. Nevertheless, clearly there are significant elution volume shifts for both BamD (30 mL earlier) and BamE (79 mL and 26 mL earlier for BamE monomer and dimer, respectively). Although one could

argue that this may be due to the BamD/E heterodimer formation (rather than $BamA_{POTRA}$-BamD/E heterotrimer), BamD/E dimer has been observed in a separate experiment (described in the following section) to produce a peak at a different elution volume (135 mL instead of 123 mL).

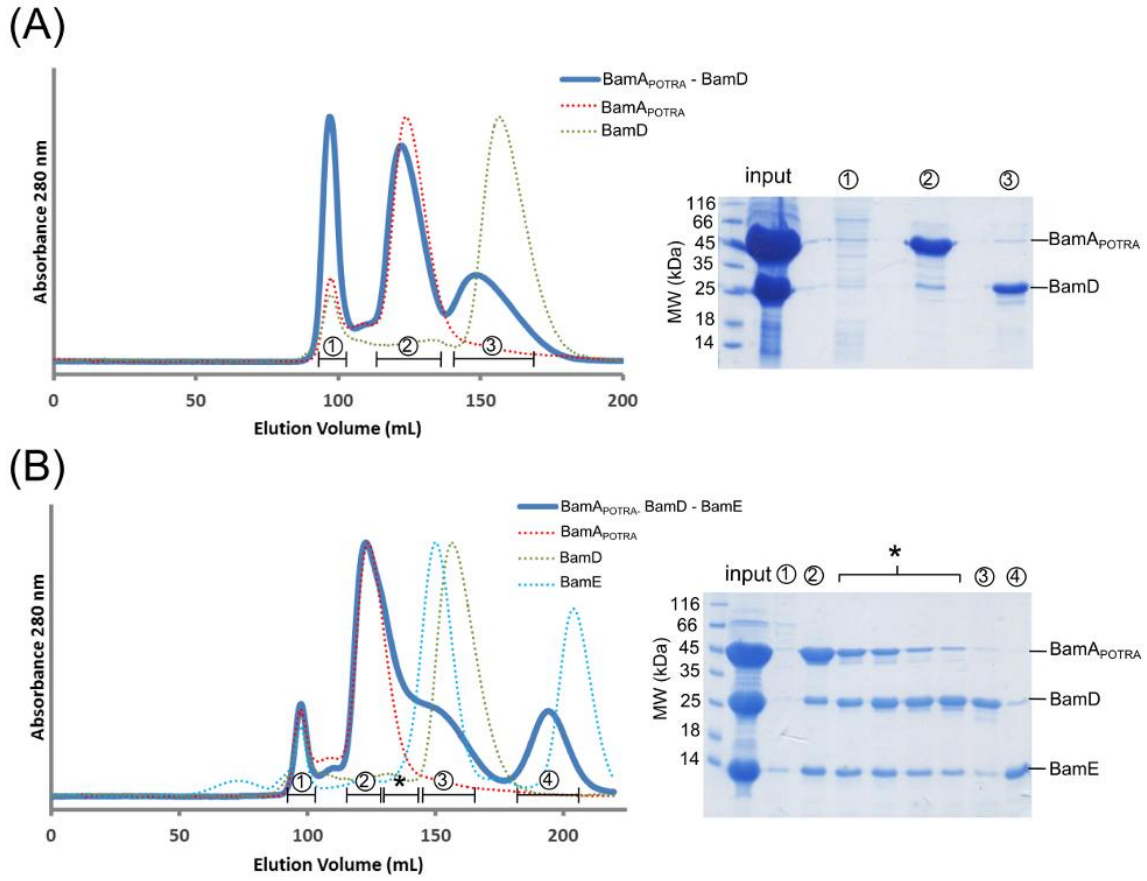


Figure 3-4 Formation of the $BamA_{POTRA}$-BamD-BamE Complex

(A) The size-exclusion chromatogram of $BamA_{POTRA}$-BamD (blue) is shown overlaid with those of BamD (green) and $BamA_{POTRA}$ (red) for comparison. Different regions on the chromatogram are labeled 1, 2 and 3, and the proteins that elute at the corresponding regions are shown on the SDS-PAGE gel (right). (B) The size-exclusion chromatogram of $BamA_{POTRA}$-BamD-BamE (blue) is shown overlaid with those of BamD (green), $BamA_{POTRA}$ (red) and BamE (cyan) for comparison. Different regions on the chromatogram are labeled 1, 2, *, 3 and 4, and the proteins that elute at the corresponding regions are shown on the SDS-PAGE gel (right). Fractions from non-overlapping regions (1, 2, 3 and 4) were pooled before SDS-PAGE analysis, whereas the fractions from the overlapping region labeled '*' were run on the SDS-PAGE gel separately. The experimental conditions were the same as shown in Figure 3-2.

3.3.4. BamC-BamD-BamE Interaction

In addition to the interaction of BamD with $BamA_{POTRA}$, its interaction with BamC and BamE were also investigated. The resulting chromatogram shows one major peak eluting at 110 mL followed by a much smaller peak at 135 mL (Figure 3-5). The molecular mass calculated from the standard curve for the larger peak is 76 kDa, which is very close to the theoretical molecular mass of 77 kDa for the BamC/D/E heterotrimer. Furthermore, the SDS-PAGE gel confirms the presence of all three proteins in the region corresponding to this peak. Lastly, all three proteins (BamC, BamD and BamE) were observed to co-elute earlier, providing further evidence that they must be forming a BamC/D/E heterotrimer.

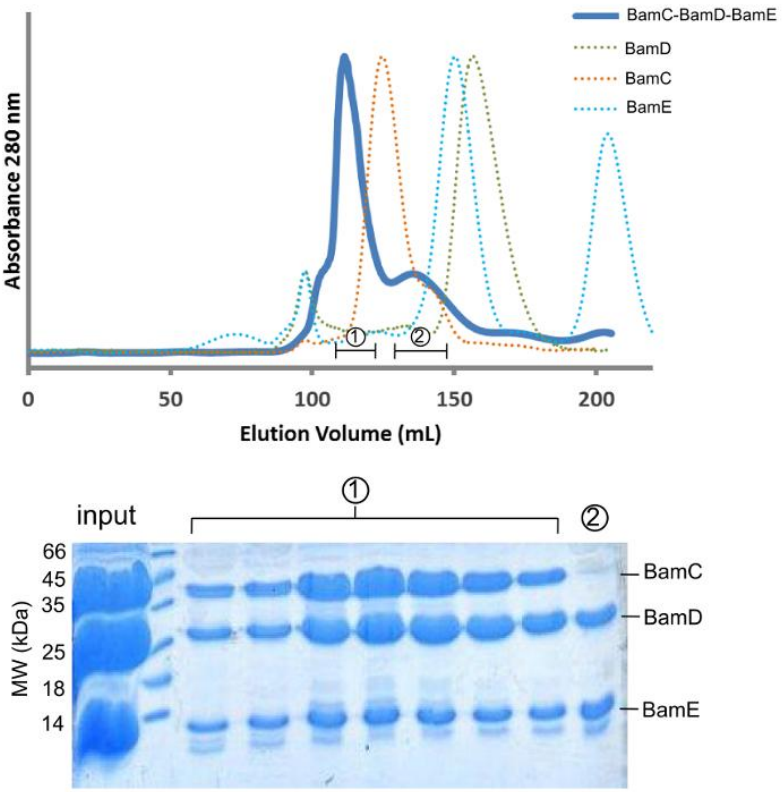


Figure 3-5 Formation of the BamC-BamD- BamE Complex

The size-exclusion chromatogram of BamC-BamD-BamE (blue) is shown overlaid with those of BamC (orange), BamD (green) and BamE (cyan) for comparison. Different regions on the chromatogram are labeled 1 and 2, and the proteins that elute at the corresponding regions are shown on the SDS-PAGE gel (bottom). The fractions from regions 1 were run on the SDS-PAGE gel separately, whereas the fractions from region 2 were pooled before SDS-PAGE analysis.

To further dissect the protein-protein interactions between BamC, BamD and BamE, the abilities of the proteins to form heterodimeric complexes with each other were examined. Three separate experiments were performed using the same procedure as before to determine if co-elution of BamC-BamD, BamC-BamE or BamD-BamE can be detected from size-exclusion chromatography. The resulting chromatograms were analyzed similarly as it was done for the other co-lysis experiments described earlier. The results are summarized in Figure 3-6 and Table 3-2.

Based on the elution volume shifts and the corresponding SDS-PAGE analysis, BamC/D and BamD/E heterodimers appear to form. The BamC/D heterodimer eluted at 120 mL, which corresponds to a calculated molecular mass of 61 kDa. This is close to the sum (64 kDa) of the theoretical molecular masses of BamC (36 kDa) and BamD (28 kDa). Similarly, calculated molecular mass of BamD/E heterodimer based on its elution volume of 135 mL (44 kDa) is similar to the sum (41 kDa) of the theoretical molecular masses of BamD (28 kDa) and BamE (13 kDa). However, the result for the BamC-BamE co-lysis experiment is not clear. Although the unchanged elution volumes of BamC and BamE in presence of each other suggest no dimer formation, the SDS-PAGE gel shows there is a small amount of BamE co-eluting with BamC in the major peak on the chromatogram (Figure 3-5B). Other forms of protein-protein interaction analysis would need to be performed to provide a clearer conclusion about the ability of BamC and BamE to form a heterodimeric complex.

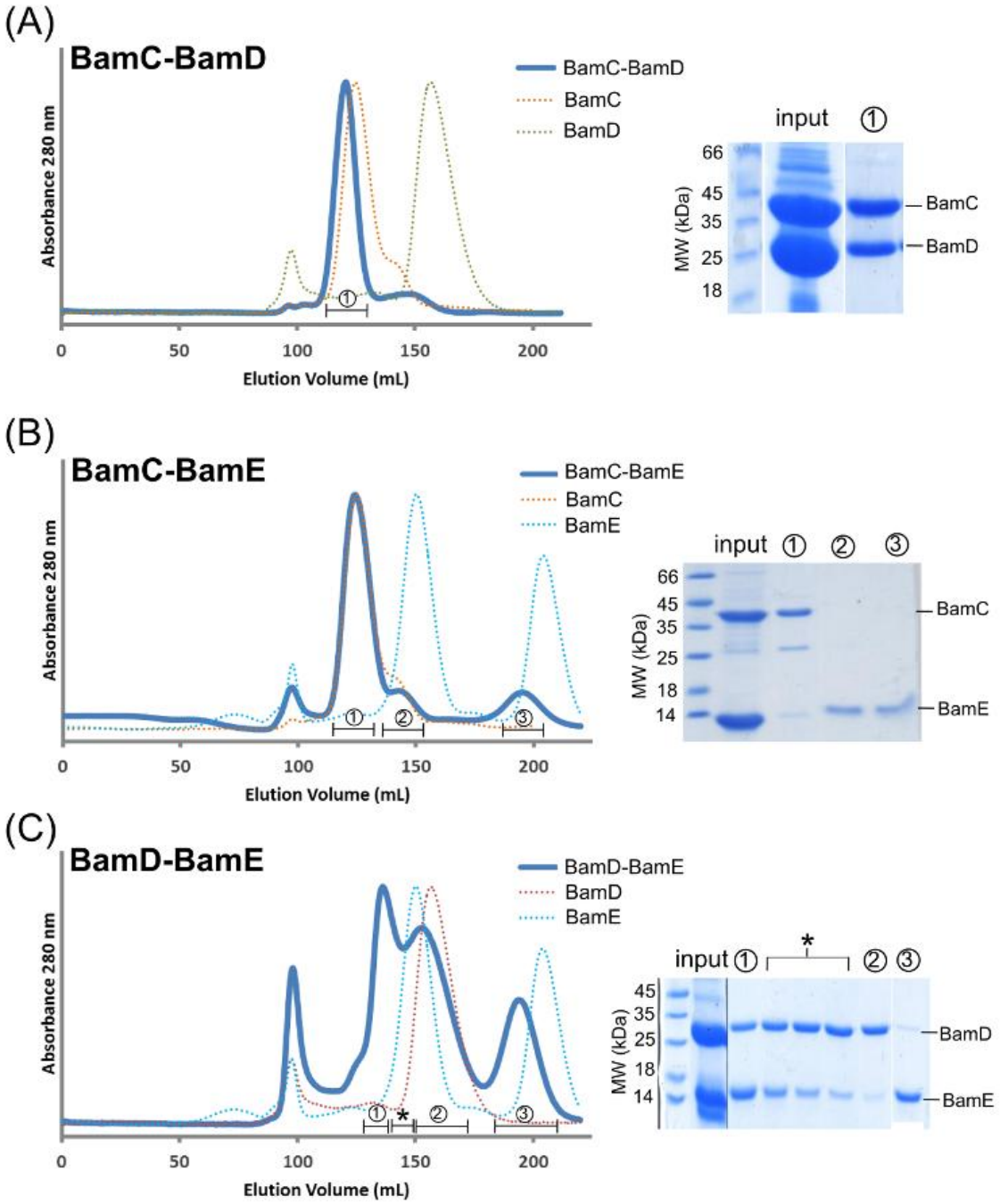


Figure 3-6 Formation of the BamC-BamD and BamD-BamE Complexes

The size-exclusion chromatograms of (A) BamC-BamD, (B) BamC-BamE and (C) BamD-BamE are shown overlaid with those of each protein run individually (dotted lines) for comparison. Different regions on the chromatogram are labeled and the proteins that elute at the corresponding regions are shown on the SDS-PAGE gels (on the right of each chromatogram).

Table 3-2 Summary of Size-Exclusion Chromatography Analysis

Proteins Mixed	Co-elution?	Co-elution Volume (mL)	Molecular Mass (kDa)		Elution Volume Shift (mL)
			Size-Exclusion Chromatography	Calculated*	
$BamA_{POTRA}$ + BamB	yes	115	68	91	$BamA_{POTRA}$ (7 mL) BamB (20 mL)
$BamA_{POTRA}$ + BamD	No	-	-	77	-
$BamA_{POTRA}$ + BamD + BamE	Yes	123	57	90	$BamA_{POTRA}$ (0 mL) BamD (32 mL) $BamE_{Dimer}$ (26 mL) $BamE_{Monomer}$ (79 mL)
BamC + BamD + BamE	Yes	110	76	77	BamC (13 mL) BamD (45 mL) $BamE_{Dimer}$ (39 mL) $BamE_{Monomer}$ (92 mL)
BamC + BamD	Yes	120	61	64	BamC (4 mL) BamD (35 mL)
BamC + BamE	No	-	-	49	-
BamD + BamE	Yes	135	44	41	BamD (20 mL) $BamE_{Dimer}$ (14 mL) $BamE_{Monomer}$ (67 mL)

* The molecular masses of the proteins were calculated by ProtParam (Gasteiger et al., 2005).

3.4. Discussion

It is becoming increasingly recognized that protein complexes are central to many biological processes. An example of such protein complexes is the BAM complex of Gram-negative bacteria that specializes in the folding and membrane insertion of OMPs. The five different subunit proteins of the BAM complex (BamA, B, C, D and E) have been identified, but how they are assembled into the higher-order structure and how many copies of the individual subunits are present in a functional unit of the BAM

complex is not clearly understood. As is the case for other protein complexes, there must be a justifiable reason that the BAM proteins function as a complex rather than individually. This may be for building a permanent larger structure that functions as one unit (as in the case of ion channels), or for co-localizing proteins within a biological pathway to increase functional efficiency (as in the case of cell signalling). Which is the case for the BAM complex is not known; hence, understanding how the individual subunits are arranged with respect to each other, as well as how strong the interactions are between them, could reveal important clues about the molecular mechanism of this important protein complex.

In this chapter, protein-protein interaction experiments involving the periplasmic components of the BAM complex ($BamA_{POTRA}$ and the lipoproteins BamB-E) have been summarized. In addition to the ability of BamC/D/E to form a stable heterotrimer, our results also showed that the POTRA domains of BamA, which exist as a monomer in solution, are sufficient to form a stable heterodimeric complex with BamB and a heterotrimeric complex with BamD and BamE. Interestingly, the $BamA_{POTRA}$-BamD association was shown to occur only in the presence of BamE, confirming the previously suggested role of BamE in stabilizing the BAM complex structure. It is currently not known whether BamE is actually situated between $BamA_{POTRA}$ and BamD to act as a linker, or is causing the conformational change in BamD to allow $BamA_{POTRA}$-BamD interaction to occur.

One interesting observation shared between the $BamA_{POTRA}$-BamB and $BamA_{POTRA}$-BamD/E experiments is that the elution profile of $BamA_{POTRA}$ does not seem to change in the presence of its interacting partner(s), even though the formation of a larger complex typically results in an earlier eluting peak (as in the case of BamC/D/E heterotrimer) (Table 3-2). Size-exclusion chromatography separates proteins by their molecular dimensions or Stokes radius, which is the radius of a hypothetical hard sphere that would diffuse through the column medium at the same rate as the molecule being examined. The unchanged elution profile of $BamA_{POTRA}$, then, suggests that the Stokes radii of the $BamA_{POTRA}$-BamB and $BamA_{POTRA}$-BamD/E complexes remain relatively the same as $BamA_{POTRA}$ alone. Our size-exclusion chromatogram analysis and the crystal structures of the POTRA domains both show that $BamA_{POTRA}$ adopts a non-globular conformation and has the measured radius larger than the lipoproteins. The radii

measured from the solved structures of $BamA_{POTRA}$, BamB, C, D, E are as follows: $BamA_{POTRA}$ (bent = 42 Å; extended = 53 Å), BamB (24 Å), BamC (estimated to be 30-55 Å), BamD (38 Å) and BamE (monomer = 18 Å; dimer = 24 Å). Unless the lipoproteins and $BamA_{POTRA}$ interact in a head-to-tail manner, it is possible that the complex still retains a similar Stokes radius as $BamA_{POTRA}$ alone. Since BamB has been shown to interact with POTRA 2-5 and BamD with POTRA5 which is situated close to the OM (Kim et al., 2007), the head-to-tail interactions between $BamA_{POTRA}$ and the lipoproteins are not likely.

Our experiments have shown that $BamA_{POTRA}$-BamB, $BamA_{POTRA}$-BamD/E, BamC/D/E, BamC/D and BamD/E complexes exist in a 1:1 or 1:1:1 stoichiometric ratio. The fact that various BAM components co-elute from a size-exclusion column as stable complexes suggests high affinity interactions between these proteins. These protein complexes are therefore ideal candidates for high resolution structural studies, and efforts are currently underway in the Paetzel lab to crystallize these protein complexes. So far, the crystal structure of BamC/D has been solved successfully, which is described in Chapter 6 of this thesis.

It should be reminded that the transmembrane β-barrel domain of BamA was not included in this study. It is possible that the periplasmic face of the BamA β-barrel may also participate in the interaction with the lipoproteins. Also, because the lipoproteins are situated close to the inner leaflet of the OM, the OM lipids may also have an influence on the interaction behaviour of the lipoproteins. Future studies should include other methods of analyzing and detecting protein complex formation to complement our size exclusion chromatography results, which sometimes produced ambiguous results (e.g. the BamC/E data). For example, dynamic light scattering analysis of the purified protein complexes can provide a more accurate measurement of the molecular weight of complexes formed, and therefore confirm the stoichiometry of the interactions. Quantification of binding affinities between the BAM components under various conditions using techniques such as surface plasmon resonance (SPR) or isothermal titration calorimetry (ITC) techniques could reveal important functional information. For instance, the experiment could be performed under various conditions (e.g. in the presence of substrate, periplasmic chaperones, etc) to learn more about the dynamics of interactions within the BAM complex.

4. Crystal Structure of BamB

Note regarding contributions:

This chapter was published in the Journal of Molecular Biology. The authors and the full reference for the article are listed below.

Kim, K.H., and Paetzel, M. (2011). Crystal structure of *Escherichia coli* BamB, a lipoprotein component of the β-barrel assembly machinery complex. *J. Mol. Biol.* **406**, 667-678.

All the figures and tables included in this chapter, otherwise noted, were adopted from the article.

4.1. Introduction

BamB is the largest (39.7 kDa) lipoprotein component of the BAM complex. Based on the co-immunoprecipitation studies, BamB is known to interact directly with BamA, independent from BamC, D and E (Malinverni et al., 2006; Sklar et al., 2007a; Wu et al., 2005). However, its exact function within the BAM complex still remains to be elucidated.

Although BamB is not essential for cell viability, OMP assembly is significantly reduced in its absence (Charlson et al., 2006; Onufryk et al., 2005). BamB and SurA (a periplasmic chaperone) deletion mutants are almost indistinguishable from each other in phenotype (Ureta et al., 2007), suggesting that BamB may work in the SurA pathway to facilitate the delivery of β-barrel precursors to BamA. Accordingly, simultaneous absence of BamB and SurA results in a synthetic lethal phenotype (Onufryk et al., 2005). BamB shows synthetic lethality also with deletion of DegP, a periplasmic chaperone/protease that works in the Skp/DegP pathway that is thought to "rescue" proteins falling off the SurA pathway (Ureta et al., 2007). Taken together, these results imply that BamB is involved in the earlier steps of OMP assembly. As the protein is not essential, it is unlikely for BamB to be involved in substrate recognition, and more likely for it to be involved in the delivery of β-barrel precursors to BamA.

We have also shown, in Chapter 3, that BamB is able to form a stable heterodimeric complex with POTRA 1-5 of BamA in solution. Furthermore, previous studies have shown that BamB specifically interacts with BamA via POTRA 2, 3, 4 and 5 domains (Kim et al., 2007). Deletion of any of these four POTRA domains has shown to abolish the observable interaction between BamA and BamB (Kim et al., 2007) by co-immunoprecipitation. The same study also showed through mutagenesis analysis that one of the β-strands in POTRA 3 (Asp241-Leu247) is essential for the BamA /B interaction (Kim et al., 2007). In a separate study, also by mutagenesis analysis, five amino acids of BamB (Leu192, Leu194, Arg195, Asp246, and Asp248) that are essential for interaction with BamA were identified (Vuong et al., 2008). However, due to the lack of structural data, where these residues are located on BamB is not known. Also, the exact mode of the interaction between the BamA and BamB, and its functional implications are not yet clear.

To gain insights into the potential function of BamB and its interaction with BamA, BamB was subjected to X-ray crystallographic study and its structure successfully solved. This chapter summarizes the structural features of BamB and discusses what the structural analysis reveals about the protein and its potential function.

4.2. Materials and Methods

4.2.1. Protein Overexpression and Purification

E. coli BamB (residues 21-392) was cloned, overexpressed and purified as described previously in Chapter 2. Selenomethionine (SeMet)-incorporated BamB to be used for phasing experiments was prepared by growing an overnight culture of BL21(λDE3) transformed with the expression plasmid carrying the BamB gene in M9 minimal medium supplemented with 50 μg/mL kanamycin. 30 mL of overnight culture was used for each liter of media to inoculate a total of three liters of M9 minimal medium (50 μg/mL kanamycin) that was grown at 37 °C until the OD_{600nm} reached 0.6. Each 1-liter culture was then directly supplemented with a mixture of the following amino acids: 100 mg of lysine, phenylalanine, threonine; 50 mg of isoleucine, leucine, valine; 60 mg of selenomethionine. After 15 minutes, protein expression was induced with 1 mM IPTG (final concentration) for 3 hours at 37 °C. The purification procedure of SeMet-incorporated BamB was the same as that used for the native protein.

4.2.2. Crystallization and Data Collection

The crystals used for SAD data collection were grown by the sitting drop vapor diffusion method. The crystallization drops were prepared by mixing 1 μL of protein (30 mg/mL) suspended in buffer A (20 mM Tris-HCl, pH 8.0, 100 mM NaCl) with 1 μL of reservoir solution and then equilibrating the drop against 1 mL of reservoir solution. This yielded crystals in the space group $P4_32_12$ with unit cell dimensions of 101.6, 101.6, 109.8 Å. The crystals have one molecule in the asymmetric unit with a Matthews coefficient of 3.2 $Å^3$ Da^{-1} (63.5% solvent). The optimal crystallization reservoir condition was 0.1 M citric acid and 2 M sodium chloride. Crystallization was performed at room temperature (~22 °C). The cryo-solution condition contained 0.1 M citric acid, 2 M

sodium chloride, and 20% glycerol. Crystals were washed in the cryo-solution before being flash-cooled in liquid nitrogen.

Diffraction data were collected on SeMet-incorporated crystals at beamline 08ID-1 at the Canadian Macromolecular Crystallography Facility (CMCF) of the Canadian Light Source (CLS), using a MarMosaicRayonix MX300 CCD x-ray detector. The crystal-to-detector distance was 320 mm. A total of 120 images were collected with 1° oscillations, and each image was exposed for 1.5 s.The diffraction data were processed with the program iMosflm (Battye et al., 2011). The data collection strategy for the native data was the same as it was for the SAD data set. The data collection statistics are summarized in Table 4-1.

4.2.3. Structure Determination and Refinement

The BamB structure was solved by SAD phasing using a data set collected at the peak wavelength (0.9792 Å), using the programs Autosol and Autobuild within PHENIX version 1.6.4 (Adams et al., 2010). Autosol found six of the possible eight selenium sites. The program Autobuild automatically constructed 90% of the polypeptide chain and performed density modification (histogram matching and solvent flattening). The rest of the model was built using the program Coot. The structure was then refined against the higher resolution native data set (~2.6 Å) by molecular replacement using AutoMR (within PHENIX) followed by restrained refinement using the program Refmac5. The structure includes all but the ten most amino-terminal residues of the construct. The structure also includes 80 ordered water molecules and one sodium ion. The data collection, phasing, and refinement statistics are summarized in Table 4-1.The structural coordinate and structure factors for BamB has been deposited in the RCSB Protein Data Bank (PDB ID: 3P1L).

Table 4-1 Crystallographic Statistics for BamB

The data collection statistics in parentheses are the values for the highest resolution shell. See Appendix D for definitions of statistical values reported.

	Native	**SeMet Incorporated**
Crystal parameters		
Space group	$P4_32_12$	$P4_32_12$
a, b, c (Å)	101.7, 101.7, 108.8	101.6, 101.6, 109.8
Data collection statistics		
Wavelength (Å)	0.9794	0.9792
Resolution (Å)	48.0-2.6 (2.7-2.6)	60.1-2.8 (3.0-2.8)
Total reflections	173,441	210,506
Unique reflections	17,912	14,763
R_{merge}	0.075 (0.399)	0.023 (0.127)
Mean $(I)/\sigma I$	20.1 (6.0)	25.8 (6.7)
Completeness	99.1 (99.7)	100.0 (100.0)
Redundancy	9.7 (9.8)	14.3 (14.6)
Phasing statistics		
No. of sites		6 (out of possible 8)
Overall FOM		0.26
Overall FOM (after density modification)		0.42
Refinement statistics		
Protein molecules in asymmetric units	1	
Residues	361	
Water molecules	80	
Total no. of atoms	2711	
R_{cryst}/R_{free}	0.20/0.23	
Average B-factor (Å^2)		
All atoms	49.2	
Protein atoms	49.5	
Mainchain atoms	48.4	
Waters	39.7	
Sodium (Na^+) ion	31.8	
r.m.s.d. on angles (°)	1.99	
r.m.s.d. on bonds (Å)	0.02	

4.3. Results

4.3.1. General Structural Features

E. coli BamB (residues 21-392) was successfully crystallized, and its structure was solved by single-wavelength anomalous dispersion (SAD) phasing method. The crystal structure, which was refined to 2.6 Å resolution, revealed that the overall shape of BamB resembles that of a short cylinder with a funnel shaped channel running down the cylindrical axis (Figure 4-1). The protein is approximately 48 Å wide in diameter and 28 Å in height, and the opening of the channel in the center of the protein is relatively wider (~18 Å in diameter) at one end than at the opposite end (~10 Å). The solvent exposed cavity created by this channel is approximately 5845 $Å^3$ in volume (Figure 4-1D).

The ring-like structure of BamB arises from the eight bladed β-propeller fold of the protein. Each blade of BamB consists of four antiparallel β-strands, and the blades are arranged around a pseudo-8-fold axis (Figure 4-1A and B; Figure H1A). The last blade (blade 8) is formed by one N-terminal-most β-strand and three C-terminal-most β-strands, and this '1+3 velcro closure' arrangement of the terminal β-strands closes the ring structure. By convention (Fulop and Jones, 1999), each strand in a blade is labeled A through D starting with the strand A as the innermost strand (Figure 4-1B and C). The β-strands within each blade are connected by short turns, while the adjacent blades are connected to each other via a long loop between the outermost β-strand (strand D) of the previous blade and the innermost β-strand (strand A) of the next blade (Figure 4-1B and C). These long connecting loops, referred to as the 'd-a loops' hereafter, are arranged in a radial manner with each loop crossing from the outer edge of each blade towards the common center of the propeller structure. The d-a loops together form a continuous molecular surface on one side of the β-propeller.

Analysis of the electrostatic properties of BamB reveals that the protein has a predominantly negatively charged surface (Figure 4-1E), in agreement with the low theoretical pI value of 4.7 for the region of the BamB construct seen in the crystal structure. The solvent-filled central channel of the protein is especially negatively charged, with several aspartate and glutamate residues (Glu197, Asp246, Asp248, Asp288 and Asp370) lining the channel surface (Figure 4-1F).

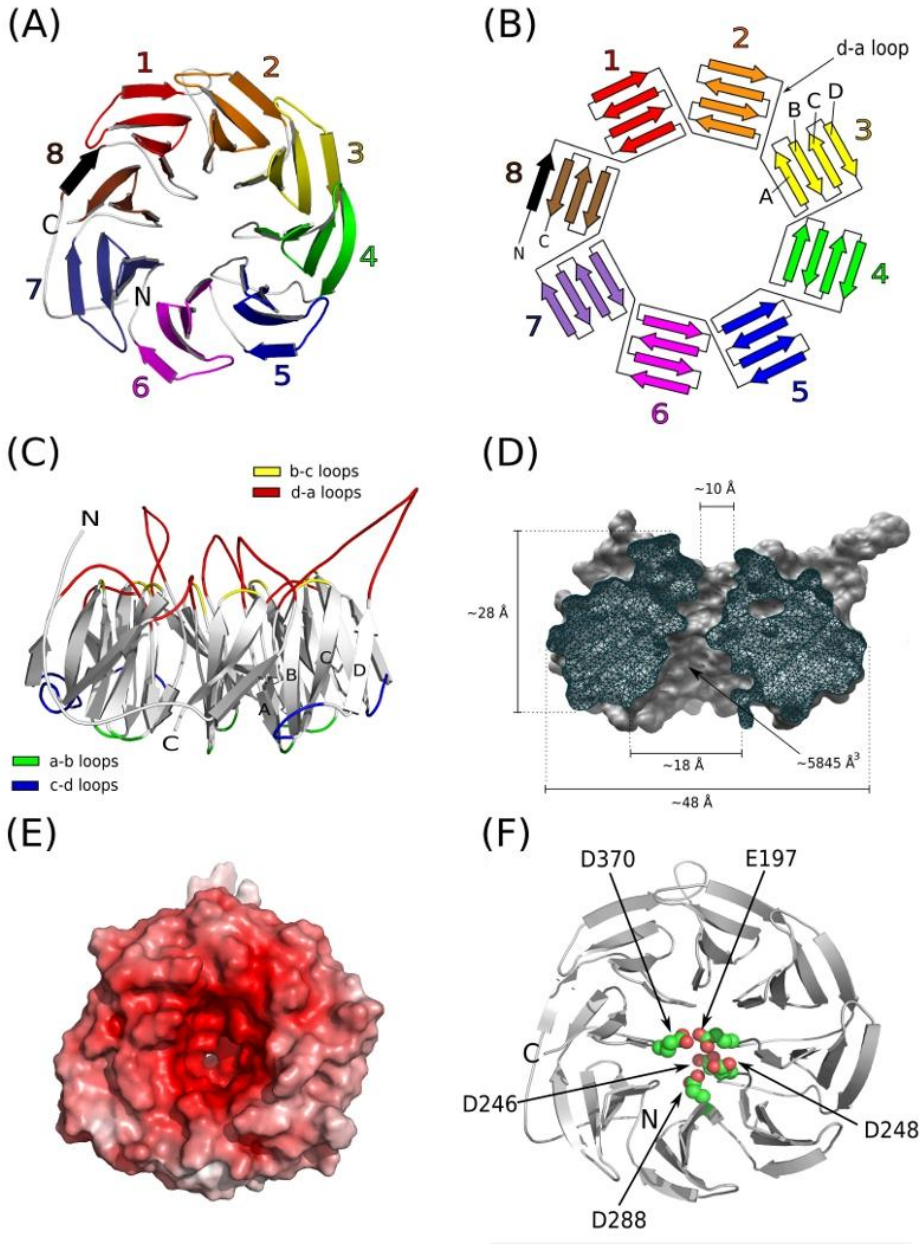


Figure 4-1 General Structural Features of BamB

(A) A ribbon diagram of the BamB structure, with each blade of the β-propeller structure numbered and shown in different colours. The protein is oriented in such a way that the d-a loop end is facing away from the reader. (B) A topology diagram of BamB. By convention, the β-strands of each blade is labeled A through D starting from the innermost strand (strand A) towards the outermost strand (strand D). (C) Different loops found in each blade are colored and labeled. (D) A Cross-section through the middle of the BamB molecule reveals the dimensions of the funnel shaped channel of BamB. (E) Electrostatic properties of the BamB molecular surface. The red and white represent negative and hydrophobic potentials, respectively. (F) Negatively charged amino acids lining the pore of the protein are shown as spheres on a ribbon model of BamB. The protein is shown in the same orientation as the surface diagram in (E).

4.3.2. Conserved Residues

Sequence comparisons of *E. coli* BamB with its functional homologues using ClustalW (Larkin et al., 2007) indicate that a large part of the protein is well conserved throughout different species of Gram-negative bacteria (Figure 4-2). The alignment result shows that there are in total 74 invariant residues, corresponding to 20% of the protein. These invariant residues are spread out throughout the amino acid sequence, but many are found adjacent to each other when the conservation is mapped onto the molecular surface of BamB (Figure 4-3).

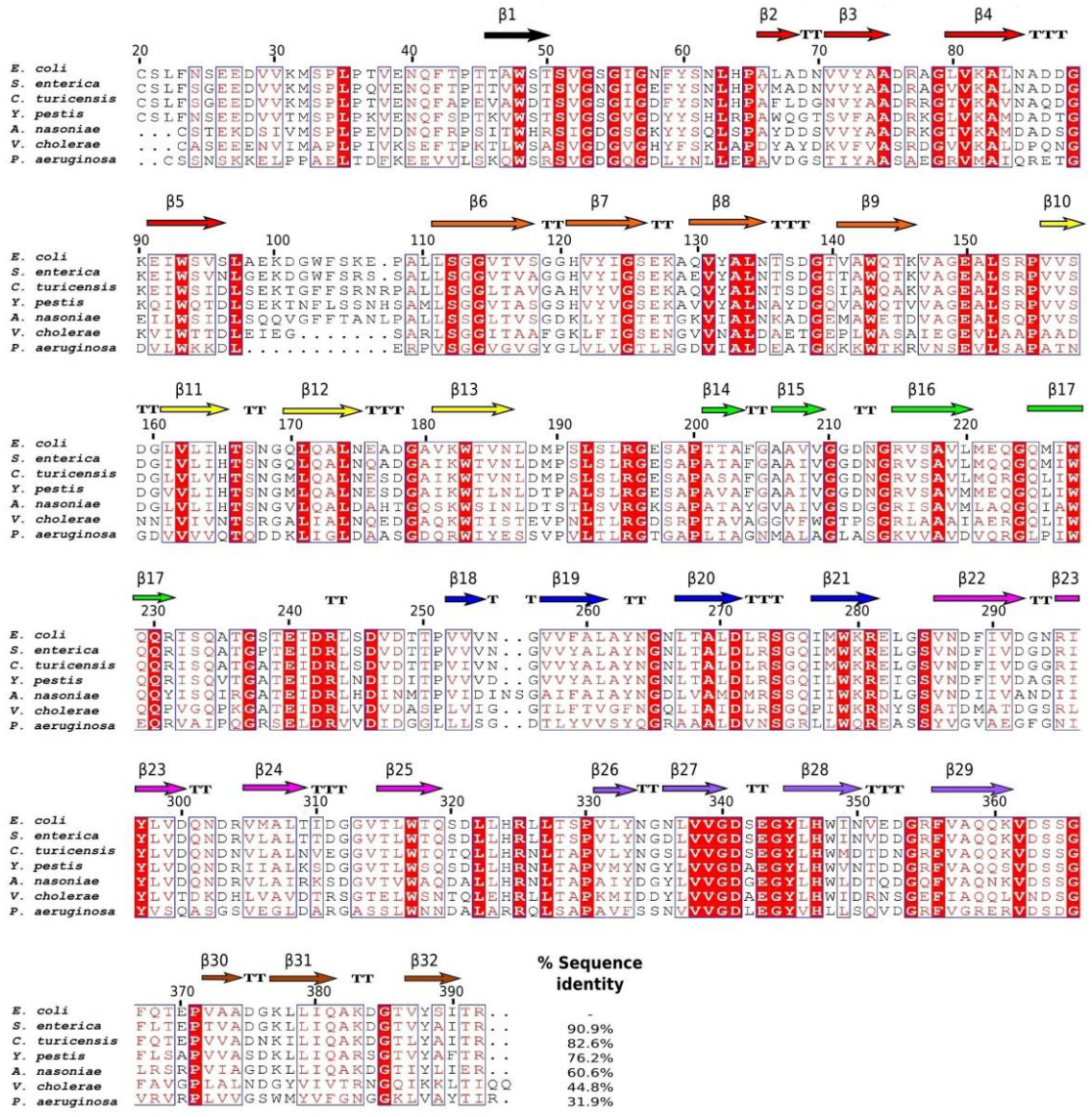


Figure 4-2 (legend on next page)

Figure 4-2 Sequence Alignment of E. coli BamB with Homologous Proteins

The sequences represent those of the mature proteins after their N-terminal signal sequences are removed. The secondary structure of *E. coli* BamB as classified by DSSP (Kabsch and Sander, 1983) is shown above the alignment. Absolutely conserved residues are shown in red boxes, similar residues in red, and stretches of residues that are similar across the group of sequences in blue boxes. The protein sequences were acquired from the Swiss-Prot data base: *Escherichia coli* (P77774); *Salmonella enterica* (B5Q1L2); *Arsenophonus nasoniae* (D2TXC8); *Yersinia pestis* (D1TUC6); *Shigella dysenteriae* (B3WYD7); *Vibrio cholerae* (D7HLH6); *Chronobacter turicensis* (C9XY24); *Pseudomonas aeruginosa* (A6V0W3). The alignment was generated by ClustalW (Larkin et al., 2007) and ESPript (http://espript.ibcp.fr).

Most of the solvent-exposed invariant residues (Arg195, Glu240, Asp242, Arg243, Asp246, Arg325 and Glu343) are localized onto a surface of BamB at one side of the β-propeller formed by the d-a loops (Figure 4-3). Located on the same side of the propeller, and both within and in close proximity of this region, are also the residues that were previously determined by mutagenesis to be important for interaction with BamA (Leu192, Leu194, Arg195, Asp246, and Asp248; discussed more in the next section).

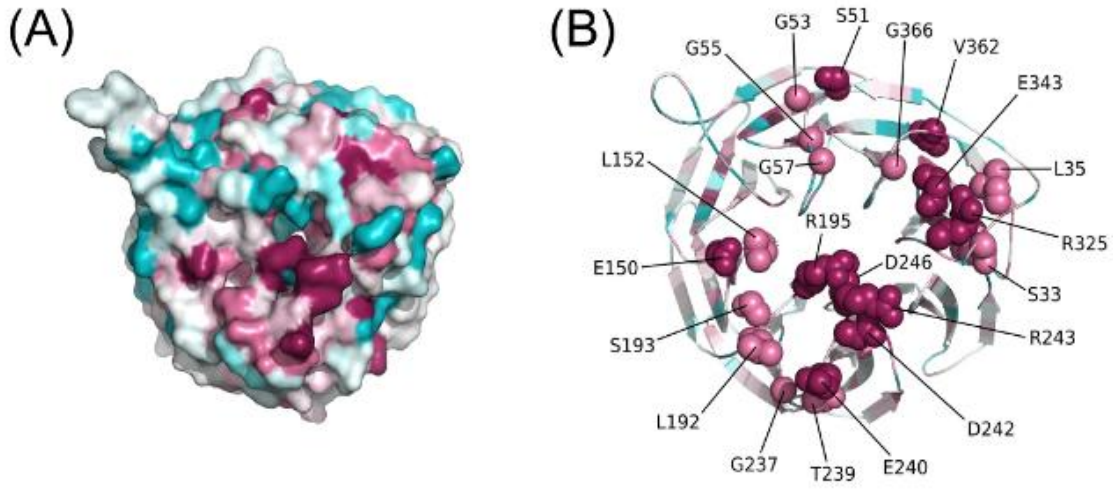


Figure 4-3 Solvent Exposed Conserved Residues of BamB

(A) The solvent exposed conserved residues of BamB are mapped onto the the surface. The protein is shown viewing down the axis with the d-a loops closest to the reader. Individual amino acid residues are coloured according to the degree to which they are conserved. Absolutely conserved residues are shown in maroon, while highly variable residues are shown in cyan. (B) The sidechains of the surface-exposed invariant (maroon) and highly conserved (dark pink) residues are shown as spheres on a ribbon model.

There are also several residues buried inside the protein that are conserved throughout BamB homologues. Interestingly, many of these buried invariant residues are found repeatedly in each blade, suggesting that each of the BamB blades is not only structurally similar but also homologous. Alignment of the sequences from all blades (except blade 8, which is topologically different from the other blades) shows that there are several conserved hydrophobic amino acids as well as an invariant glycine residue (Figure 4-4). When mapped onto the BamB structure, the hydrophobic residues seem to promote inter-blade contacts by forming interlocking hydrophobic interactions, whereas the conserved glycine residues seem to play a structural role in the connecting loop between the strand C and D of each blade.

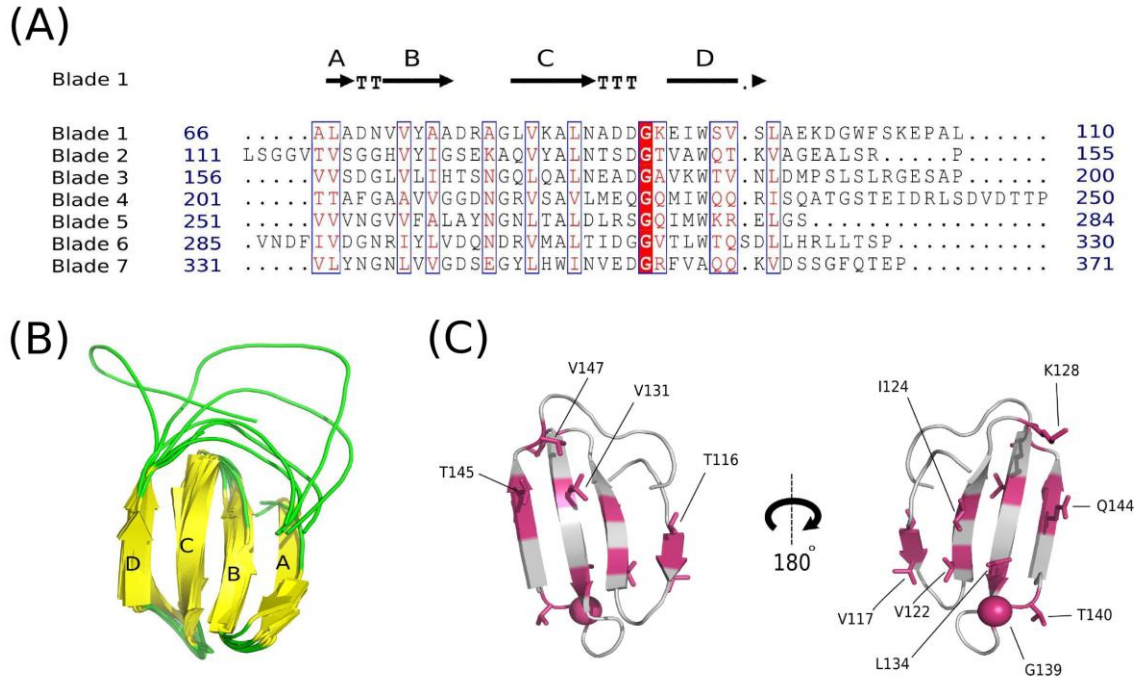


Figure 4-4 Comparison of the Individual Blades of BamB

(A) A sequence alignment of blade 1-7. The secondary structure of blade 1 as calculated by DSSP is shown above the sequence alignment. (B) A superposed ribbon models of blade 1-7. Each strand is labeled A through D starting with the innermost strand of the blade. (C) The boxed residues in sequence alignment in (A) are mapped onto the structure of blade 2. The sidechains of the conserved residues are shown in maroon as sticks on a ribbon model of blade 2. The invariant glycine residue is shown as a sphere.

4.3.3. BamA Interaction Site

As mentioned earlier, some residues found within and near the conserved area on the solvent-exposed surface of BamB have previously been shown to be important for interaction with BamA. Five residues involved in the interaction were identified by Vuong et al (2007) using a series of mutagenesis studies, and they are Leu192, Leu194, Arg195, Asp246, and Asp248 (Vuong et al., 2008). When these residues are mapped onto the BamB crystal structure, all five residues are found in a continuous linear patch on the surface of the protein (Figure 4-5). While Leu192, Leu194, and Arg195 are located on the d-a loop that connects blade 3 and blade 4, Asp246 and Asp248 are located in the d-a loop connecting blade 4 and blade 5. As these two loops are located adjacent to each other, the five residues are brought into close proximity to form a continuous patch on the surface of BamB.

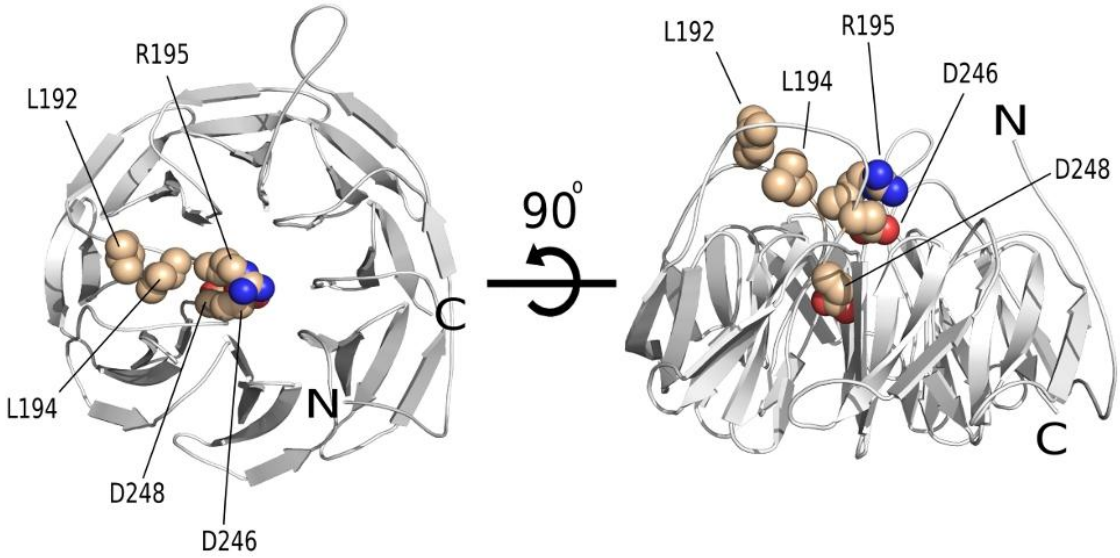


Figure 4-5 BamA Interaction Surface of BamB

The five amino acid residues (L192, L194, R195, D246 and D248) of BamB that have previously been identified to be important for interaction with BamA are shown as spheres on a ribbon model of the BamB backbone structure.

4.3.4. Structural Homologues

A search for structural homologues using the DALI (Holm et al., 2008) server identified several proteins that have a significant degree of similarity in protein topology and architecture with BamB (Figure 4-6). Proteins with a BamB-like fold include an alcohol dehydrogenase from *Pseudomonas putida* known as ADH IIB (Chen et al., 2002) (PDB ID: 1KV9) (Figure 4-6B), *Homo sapiens* Fbw7 (Hao et al., 2007) (part of an ubiquitin ligase complex; PDB ID: 2QVR) (Figure 4-6C), *Paracoccus pantotrophus* cytochrome cd1 nitrite reductase (Sjogren and Hajdu, 2001) (PDB ID: 1HJ5) (Figure 4-6D), and *Saccharomyces cerevisiae* Sif2p (Cerna and Wilson, 2005) (a transcriptional co-repressor of meiotic genes; PDB ID: 1R5M) (Figure 4-6E). All these proteins have an eight-bladed β-propeller structure, and most of them have additional domains positioned on the either side of the β-propeller domain. In all four structurally homologous proteins, protein binding sites (the peptide binding site of Fbw7 and the conserved putative protein binding surface of Sif2p) or ligand binding sites (PQQ binding pocket of AD IIB and the heme binding site of nitrite reductase) are found in the β-propeller domain, more specifically on the surface of the domain formed by the d-a loops that are equivalent to the surface of BamB where the conserved residues and the five residues important for BamA interaction are located.

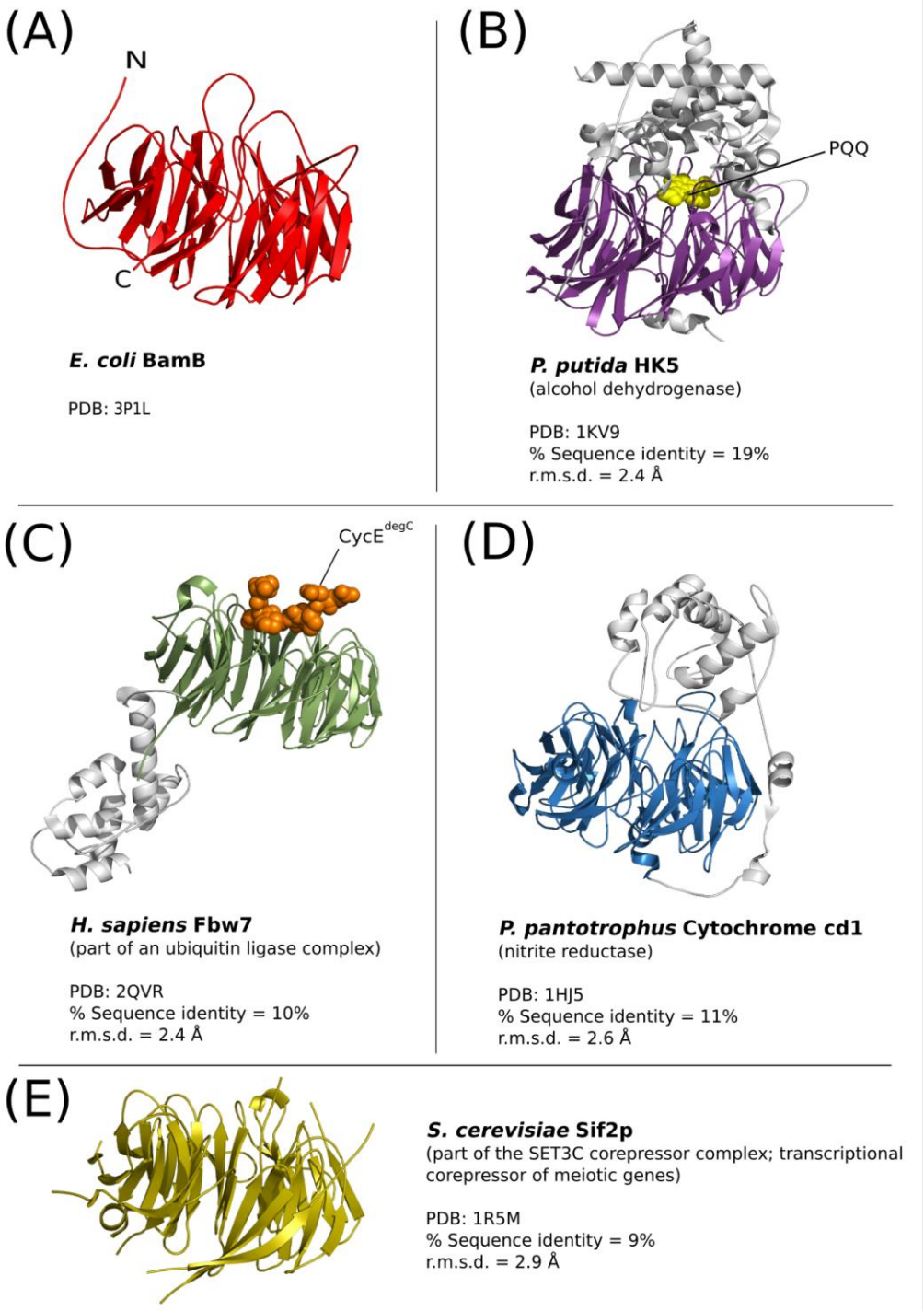


Figure 4-6 Structural Homolgues of BamB

E. coli BamB (A) has a similar protein fold to a variety of other proteins (B-E). Only the regions that share structural homology with BamB are shown in color. Non-homologous regions are shown in grey ribbon. The bound cofactor in (B) and the peptide in (C) are shown as colored spheres and labeled. The percent sequence identity and backbone rmsd for 3D superposition are given for each structure in comparison to BamB.

4.4. Discussion

In *E. coli*, the BAM complex plays an essential role in OMP folding and assembly in the outer membrane (Gentle et al., 2004; Knowles et al., 2009b). The multi-component complex is made up of one integral membrane protein BamA, and four lipoproteins BamB, C, D and E (Anwari et al., 2010; Malinverni et al., 2006; Sklar et al., 2007a). This chapter has described the first crystal structure of BamB, the largest lipoprotein component of the BAM complex.

Analysis of the eight-bladed β-propeller structure of BamB provides us with clues for how BamB may interact with BamA and possibly with protein substrates (i.e. unfolded OMPs). Previous mutagenesis studies have identified residues that are critical for the BamA-BamB association. On BamB, as described earlier, these are the conserved residues (Leu192, Leu194, Arg195, Asp246, and Asp248) found in loops on one face of its β-propeller structure. In BamA, the residues important for BamB association are located on a β-bulge in the second β-strand of POTRA3 (Ile240 and Asp241) (Kim et al., 2007). This bulge forms one edge of the POTRA3 β-sheet, and interestingly it is observed to participate in crystal packing via β-augmentation (a mode of protein interaction in which a strand from one protein is added to an existing β-sheet of another) in two independently solved crystal structures (Figure 4-7A) (Gatzeva-Topalova et al., 2008; Kim et al., 2007). Whether this region of POTRA 3 is where the five conserved residues of BamB bind and also whether the BamA/B interaction is mediated by β-augmentation is yet unclear and requires further experiments.

In addition to the binding surface identified in this study, there may be additional sites of interaction with BamA on BamB molecular surface, as all four POTRA domains (POTRA 2-5) have been shown to interact with BamB (Kim et al., 2007). One possible additional binding surface could be the negatively charged pore, which is located adjacent to the identified binding surface. Another possible site of protein–protein interaction is the propeller rim of the BamB structure. If the POTRA domains, each of which contains a three-stranded β-sheet, interact with BamB indeed via β-augmentation as Kim et al (2007) hypothesized, then it seems plausible that the outermost β-strand (i.e. strand D) in each blade of BamB may serve as a binding surface. The strand D from each blade is exposed to the solvent and available for hydrogen bonding interactions.

Similarly, if the function of BamB involves substrate binding and delivery, as suggested in some studies (Charlson et al., 2006; Ureta et al., 2007), this may be how BamB interacts with the unfolded OMP substrates. As the OMPs that are most affected by BamB deletion are relatively large (16-24 stranded β-barrels), it has been suggested that BamB could aid BamA function by increasing the substrate binding capacity (Heuck et al., 2011).

Following the publication of our BamB structure, other crystal structures of BamB (also from *E. coli*) was published (Albrecht and Zeth, 2011; Heuck et al., 2011; Noinaj et al., 2011). Interestingly, one of these BamB crystal structures supported our hypothesis by revealing that one of the outermost β-strands (strand D) of the β-propeller blades in one BamB molecule interacts with its symmetrically related neighbouring BamB molecule in the crystal via β-augmentation (Heuck et al., 2011). As the BamB structure consists of eight blades that are very close to each other in topology, it seems possible that BamB could provide eight potential protein-protein interaction sites (Figure 4-7B). Future biochemical studies and a co-crystal structure of BamB with the POTRA domains or with an OMP substrate will help confirm this β-augmentation hypothesis.

Even though BamB by itself is not absolutely essential for cell viability, it has been shown that it is essential for proper biogenesis and assembly of of some OMPs, including the type III secretion system in *Salmonella enterica* serovar Enteritidis (Fardini et al., 2007). Fardini *et al.* (2007) showed that the absence of BamB renders the bacteria avirulent due to its inability to secrete various virulence factors via the Type III secretion system (Amy et al., 2004; Fardini et al., 2007). This suggests that BamB has a potential to serve as a novel drug target. Future structural and functional studies of BamB should focus on experimentally validating the hypotheses that were generated based on the structural analysis (e.g. Does BamB interact with the POTRA domains by β-augmentation? Does it interact with substrates? etc.).

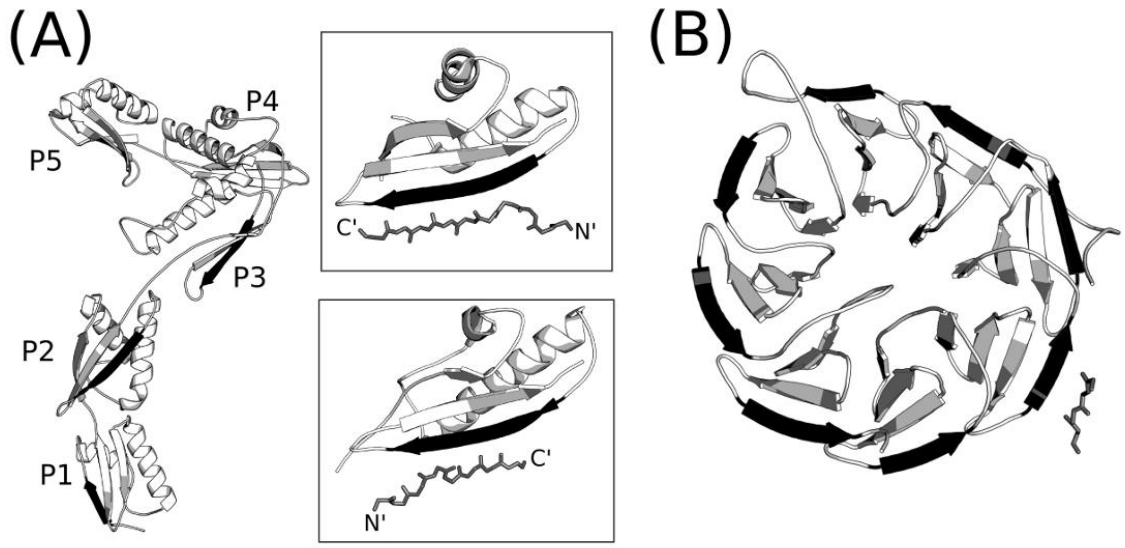


Figure 4-7 Potential BamA-BamB Interaction Sites

(A) A model of POTRA 1-5 was built by superposing the POTRA1-4 structure (PDB: 3EFC) with the POTRA 4-5 structure (PDB: 3Q6B). Colored black are the regions that have been shown experimentally to be the sites of protein-protein interaction. Crystal packing of POTRA domains revealed that at least POTRA 3 can interact with other proteins via β-augmentation (close-up views). One of the β-strands of POTRA 3 that is important for BamB interaction is capable of undergoing β-augmentation in both parallel (top; PDB: 2QDF) and anti-parallel (bottom; PDB: 3EFC) manners. (B) Similar to the POTRA domains, a BamB crystal structure (PDB: 3PRW) shows that the outermost strand in one of the β-propeller blades can interact with another strand (the main chain of which is shown as a stick model) via β-augmentation. All the solvent exposed β-strands that could potentially serve as β-augmentation sites are shown in black. This figure was adopted from Kim *et al*.(2012).

5. Structural Characterization of BamE by NMR

Note regarding contributions:

This chapter was published in Biochemistry. The authors and the full reference for the article are listed below.

Kim, K.H., Kang, H.S., Okon, M., Escobar-Cabrera, E., McIntosh, L.P., and Paetzel, M. (2011). Structural characterization of *Escherichia coli* BamE, a lipoprotein component of the β-barrel assembly machinery complex. *Biochemistry* **50**, 1081-1090.

The work summarized in this chapter was a collaborative effort with Dr.MacIntosh's laboratory at UBC. Mark Okon carried out all of the data collection and initial data processing. Hyunseo Kang and Eric Escobar-Cabrera contributed by training me how to use various NMR softwares for resonance assignment and structure calculation, and also made intellectual contributions throughout the structure solving process. All the figures and tables included in this chapter, otherwise noted, were adopted from the article.

5.1. Introduction

BamE is the smallest (10.4 kDa) and also the most recently discovered component of the BAM complex. Unlike BamA and BamD, it is not an essential member of the complex; however, the loss of BamE has been shown to cause mild OMP assembly defects and an overall increase in the permeability of OM (Lewis et al., 2008; Sklar et al., 2007a).

The previous co-immunoprecipitation studies and our work presented in chapter 3 have shown that the stability of the BAM complex is compromised in the absence of BamE, suggesting that it plays an important structural role (Sklar et al., 2007a). The loss of BamE seems to significantly weaken the interaction between BamD and the POTRA domains of BamA, especially (Sklar et al., 2007a). Interestingly, in a recently published study, the protease sensitivity of BamA has been shown to increase dramatically in the absence of BamE (Rigel et al., 2012). It has therefore been suggested that BamE may modulate the BamA conformation, although it's not yet clear whether BamE does this directly or indirectly through its interaction with BamD.

To obtain clues about the BamE function, a structural investigation by NMR spectroscopy was initiated. As mentioned in Chapter 3, BamE was found to exist in both monomeric and dimeric states in solution. Due to its small size and high expression level, BamE monomer was an ideal candidate for the NMR studies. This chapter presents the structural and dynamic properties of *E. coli* BamE obtained by NMR spectroscopy. For comparison, the NMR spectrum of BamE dimer was collected and the dimerization of BamE was further investigated by size-exclusion chromatographic analysis.

5.2. Materials and Methods

5.2.1. Isotopic Labeling and Protein Purification

BamE was cloned in fusion with an N-terminal histidine tag as described previously in Chapter 2. For NMR studies, uniformly ^{15}N-labeled BamE was expressed in M9 media supplemented with 1 g/L $^{15}NH_4Cl$ (Sigma-Aldrich). Uniformly $^{15}N/^{13}C$-labeled

BamE was expressed in M9 media containing 3 g/L $^{13}C_6$-glucose (Sigma-Aldrich) and 1 g/L $^{15}NH_4Cl$. For both isotopically labeled samples, cultures were grown at 37°C to an OD_{600} of 0.6 and induced with 1 mM IPTG overnight at 25°C. Cells were harvested and purified by nickel affinity chromatography in Buffer A (20 mM Tris-HCl, pH 8.0; 100 mM NaCl) using the protocol described in Chapter 2.

The purified BamE sample was incubated with thrombin (GE Healthcare) overnight for cleavage of the N-terminal histidine tag. The digested protein sample was concentrated to approximately 10 mg/mL using an Amicon ultra-centrifugal filter device (Millipore) with a 3 kDa MW cut-off, and was then further purified by size-exclusion chromatography (Sephacryl S-100 HiPrep 26/60 column) on an AKTA Prime system (GE Health Care). In this last step of protein purification, the buffer was also exchanged to Buffer B (20 mM Na_2HPO_4/NaH_2PO_4, pH 6.8), and the monomeric and dimeric forms of BamE were resolved. The final protein concentrations of the samples used for NMR data acquisition were ~0.5 mM.

5.2.2. Oligomeric State Analysis

Apparent molecular mass of purified BamE was determined by gel filtration chromatography using a calibrated Superdex 75 column (GE Health Care). A sample of 200 µL of BamE (5 mg/mL) was injected, resolved, and analyzed at a flow-rate of 0.5mL/min in buffer A. The oligomeric state of BamE was also monitored using gel filtration chromatography under different conditions of sample pH values (CH_3COONa pH 3.5, MES pH 6.5, Tris-HCl pH 8.0, CAPS pH 10) and salt concentrations (0, 100, 300, 500 mM and 1 M NaCl). For oligomeric state analysis, unlabeled protein sample was used that was grown and purified from LB media. The oligomeric state of purified BamE was also confirmed by dynamic light scattering analysis using the protocol described in Chapter 3.

5.2.3. NMR Data Acquisition

NMR spectra were recorded at 15°C on Varian Unity 500 MHz and 600 MHz NMR spectrometers. The low temperature was used to ensure stability of the protein sample during data collection. All samples consisted of ~0.5 mM protein in 20 mM

Na_2HPO_4/NaH_2PO_4, pH 6.8, and ~10% D_2O. Spectra were processed using NMRPipe (Delaglio et al., 1995) and analyzed using Sparky (Goddard and Kneller). NMR chemical shifts were referenced directly or indirectly to 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS).

5.2.4. Spectral Assignments and Structure Calculation

Using an extensive set of multi-dimensional NMR experiments, the backbone and side chain ^{1}H, ^{13}C and ^{15}N chemical shifts of BamE were assigned by standard methods (Sattler et al., 1999). NOE-derived distance restraints were obtained from simultaneous regular and constant time methyl 3D ^{13}C- and ^{15}N-NOESY-HSQC spectra, all with 100 ms mixing times (Zwahlen et al., 1998). An initial set of NOE crosspeaks were manually assigned and the remaining signals assigned automatically by ARIA. Backbone dihedral angles were determined from $^{13}C^{\alpha}$, $^{13}C^{\beta}$, $^{13}C'$, $^{1}H^{\alpha}$, and $^{1}H^{N}$ chemical shifts using TALOS (Cornilescu et al., 1999). A limited set of hydrogen bond distance restraints were included for selected amides located in β-strands, as determined via manual inspection of NOE patterns and chemical shift information. The BamE structure was then calculated and refined using ARIA 2.2 with CNS 1.2 (Habeck et al., 2004). The chemical shifts and structural coordinates of the BamE ensemble have been deposited in the BioMagResData Bank (accession number: 16926) and RCSB Protein Data Bank (accession number: 2KXX), respectively.

5.2.5. Backbone Amide Relaxation Analysis

Backbone amide relaxation data of ^{15}N-labeled BamE were acquired on a 500 MHz spectrometer at 28 ^{o}C (Farrow et al., 1994). ^{15}N T_1 and T_2 lifetimes and heteronuclear ^{1}H- ^{15}N NOE values were fit using Sparky (Goddard and Kneller), and analyzed according to the model-free formalism with TENSOR2 (Dosset et al., 2000). The predicted global tumbling time was calculated using the program HYDRONMR (Garcia de la Torre et al., 2000).

5.3. Results

5.3.1. Dimerization of BamE

BamE (Ser21-Asn113) that encompasses the entire wildtype sequence immediately following the cleavable N-terminal signal sequence and the conserved lipidation residue Cys20 was produced for a structural study. As described in Chapter 3, the purified BamE was found to exist in both monomeric and dimeric states, as determined by analytical gel filtration chromatography (Figure 5-1A). Two major peaks were observed on the chromatogram, one eluting at an elution volume that is expected for an approximately 10-15 kDa species and the other for a 25-30 kDa species. The fractions corresponding to each peak yielded a single band on SDS-PAGE with the apparent molecular mass expected for monomeric BamE after the removal of the hexahistidine tag (~11 kDa). No other significant proteins of higher molecular weight were observed. Thus the expressed BamE exists in both monomeric and dimeric states in solution.

The dimer and the monomer fractions from the size-exclusion chromatography were collected separately, pooled, and subsequently subjected to a second size-exclusion chromatography run to determine whether there is a concentration-dependent monomer/dimer equilibrium (Figure 5-1B). A single peak was observed in both chromatograms at distinct elution volumes, demonstrating that the monomeric and the dimeric species do not interconvert under the conditions or timescale of this experiment (over the period of approximately one week). After purification by size-exclusion chromatography, the molecular masses of the monomeric and dimeric forms of BamE were verified by multiangle dynamic light scattering analysis (Figure 5-1C). The measured values were 12.4 $\pm$ 0.8 kDa and 28.5 $\pm$ 3.0 kDa for the BamE monomer and the dimer, respectively.

Additional analytical gel filtration chromatography was performed to determine whether dimer formation or dissociation is affected by various conditions. Neither pH (3.5, 6.5, 8.0, and 10), salt concentration (0, 100, 300, 500 mM and 1 M NaCl), nor the presence of a detergent (0.01% n-dodecyl β-D-maltoside) induced dimerization of BamE monomers or dissociation of the dimer (Appendix F). Thus self-association is not due to

simple electrostatic or hydrophobic interactions. Also, since the protein lacks cysteine residues, dimerization of BamE cannot be due to disulfide bond formation.

To investigate further the self-association of BamE, we recorded the ^{15}N-HSQC spectra of the two purified forms of the ^{15}N-labeled protein (Figure 5-2). The spectrum of the monomer shows well dispersed signals, indicative of a stable, folded protein. In contrast, the dimeric form yielded a spectrum with signals of significantly differing intensities, suggestive of both ordered and disordered regions undergoing conformational exchange on a msec-μsec timescale. More importantly, the spectra of the two forms of BamE show remarkably little overlap, suggesting that the monomeric and dimeric forms have substantially different structures. Combined with the lack of observable interconversion, we thus hypothesize that BamE can adopt a kinetically-trapped intertwined or perhaps domain-swapped dimeric conformation (Bennett et al., 1995; Liu and Eisenberg, 2002). It is presently not clear which form exists in within the BAM complex, or if the BamE dimerization holds a functional significance. Accordingly, all subsequent structural and dynamics analyses described in this study were carried out with the monomeric form of BamE.

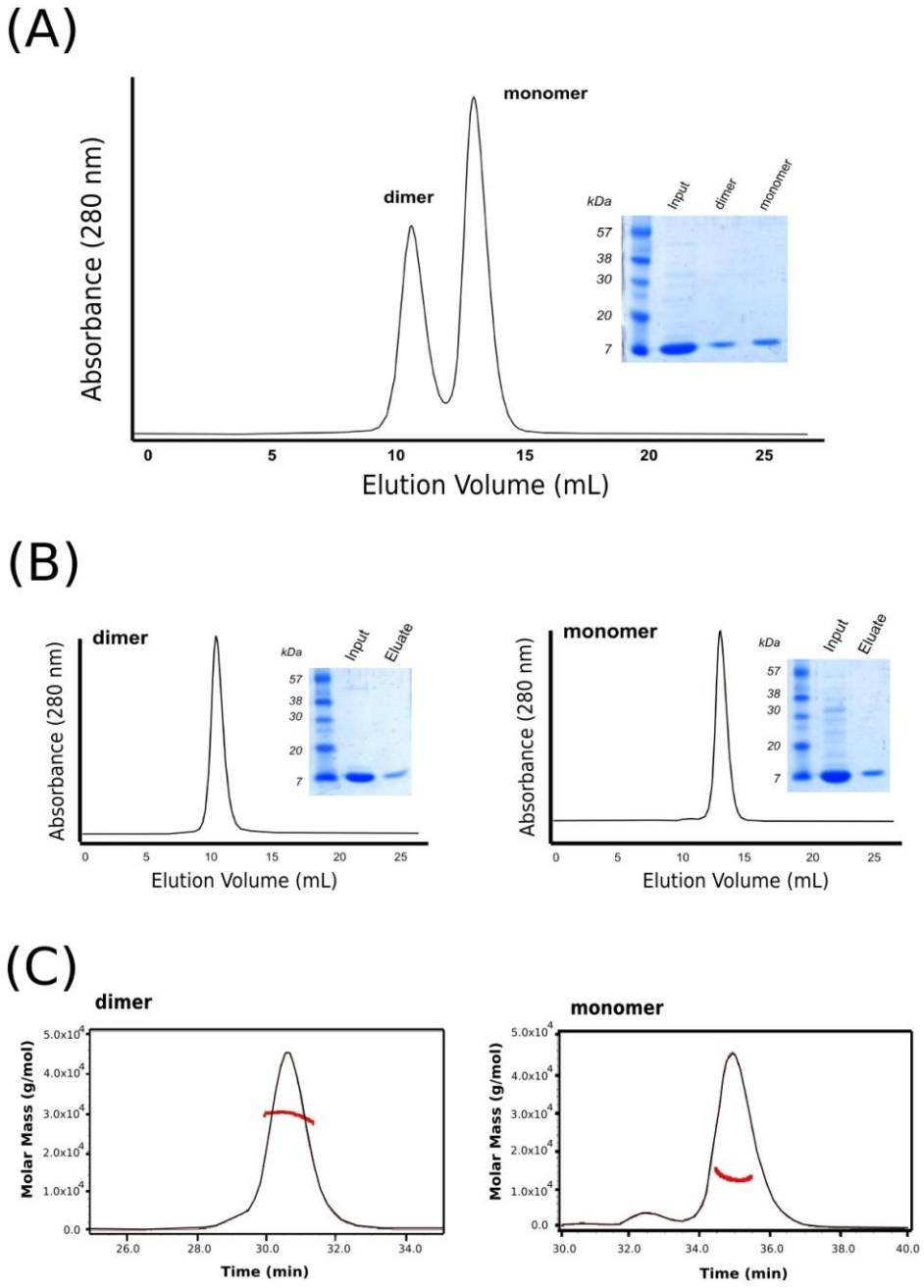


Figure 5-1 Monomeric and Dimeric States of BamE

(A) A size-exclusion chromatogram and the corresponding SDS-PAGE gel of BamE are shown. (B) The dimer and the monomer fractions from (A) were collected separately, and run through the size-exclusion column again to test whether the two states exist in concentration-dependent equilibrium. (C) The molecular masses of BamE monomer and dimer were verified by multiangle dynamic light scattering analysis. The chromatogram from an in-line gel filtration column is shown in black, and the calculated molecular mass in red.

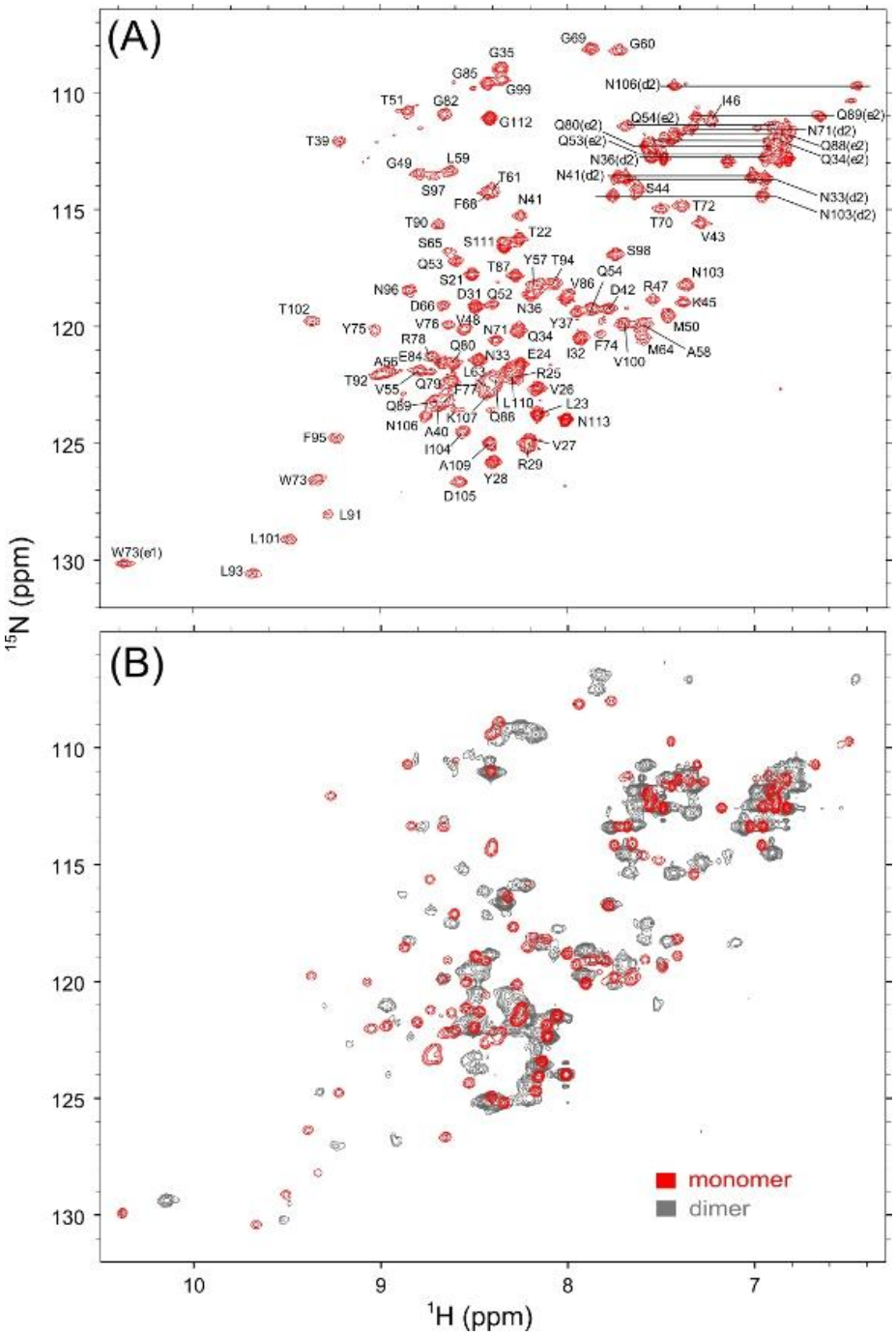


Figure 5-2 NMR spectra of Monomeric and Dimeric BamE

(A) The ^{15}N-HSQC spectrum of BamE monomer is shown with peaks assigned. The well dispersed signals from 1H-15N groups confirms that the monomeric form of the protein is stably folded and a good candidate for further structural analysis. (B) The superimposed ^{15}N-HSQC spectra of the BamE monomer (red) and dimer (grey) show very little peak overlap, indicating distinctly different backbone conformations.

5.3.2. NMR Structure of BamE

Using an extensive set of NOE-derived distance and chemical shift-derived dihedral angle restraints, we calculated the structural ensemble of monomeric BamE with the program ARIA. Table 5-1 shows a summary of the NMR data and structural statistics. The root-mean-square (r.m.s.) deviations between the 20 lowest energy structures for the helical and strand regions of the protein were 0.22 Å (backbone atoms) and 0.52 Å (all atoms).

Table 5-1 NMR Restraints and Structural Statistics for BamE Ensemble

Summary of restraints		
NOEs		
Intraresidue	726	
Sequential	316	
Medium range (1 < | i – j | < 5)	139	
Long range (| i – j | ≥ 5)	299	
Total	1480	
Dihedral angles (□□□□□□1)	55, 55, 0	
Hydrogen bonds	20	
Residues in allowed regions of Ramachandran plot[a]	98.40%	
Mean energies (kcal/mol)		
E_{vdw}	-247.2 ± 19.0	
E_{bonds}	43.1 ± 2.6	
E_{angles}	153.2 ± 9.8	
E_{impr}	73.2 ± 8.5	
E_{NOE}	181.9 ± 15.5	
E_{cdih}	3.3 ± 1.0	
r.m.s. deviation (Å)		
	structured elements[b]	all[c]
Backbone atoms	0.22	0.94
All heavy atoms	0.52	1.38

[a]Calculated with Procheck-NMR (Laskowski et al., 1996), and summed over most favored, allowed, and generously allowed regions.

[b]Core structured region identified from Promotif (Hutchinson and Thornton, 1996), DSSP(Kabsch and Sander, 1983), SSP (Marsh et al., 2006) and MOLMOL (Koradi et al., 1996). α1, 40-43; α2, 52-58; β1, 72-75; β2, 90-95; β3, 101-107.

[c]All the atoms except the flexible N- and C-terminal regions; 40-107.

BamE has a well-structured core that is made up of two N-terminal anti-parallel α-helices (α1: Ala40-Val43; α2: Gln52-Ala58) and a C-terminal twisted anti-parallel β-sheet consisting of three β-strands (β1: Thr72-Tyr75; β2: Thr90-Phe95; β3: Leu101-Lys107) (Figure 5-3). Residues Pro67-Gly69 also forms a helical-like turn. Collectively, these secondary structural elements yield a two-layer sandwich fold with α1 and α2 packing against the β-sheet. Together, the core of BamE (residues 40-107) have the approximate dimensions of 22 Å x 46 Å x 24 Å with a surface area of ~5000 $Å^2$, and volume of ~8200 $Å^3$. In contrast to the well-ordered core, the N (residues 21-39) and C (residues 108-113) terminal segments of BamE, and the 14 residue loop L3 (residues 76-89) joining β1 and β2, appear disordered with high r.m.s. deviations in the calculated ensemble due to a lack of structural restraints (Figure 5-3C). This mobility was confirmed by ^{15}N-relaxation measurements, as discussed below.

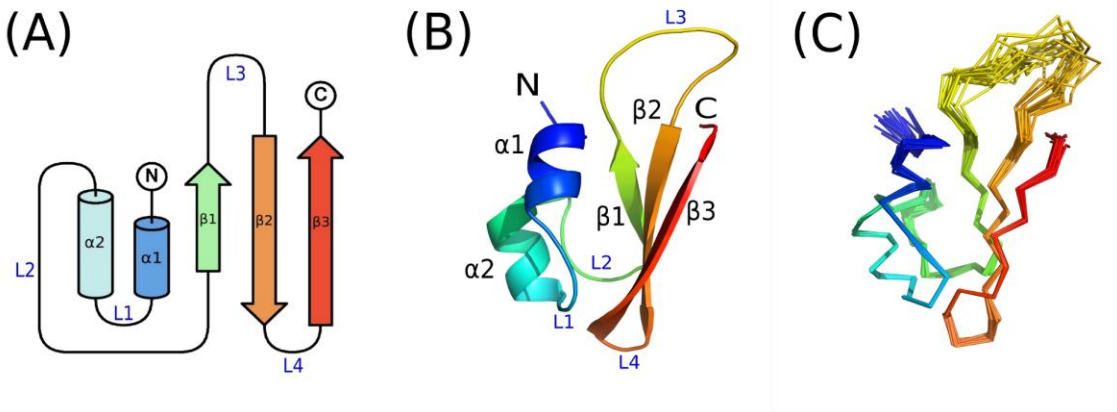


Figure 5-3 The NMR-derived Structural Ensemble of E. coli BamE

(A) A topology diagram of BamE with strands shown as arrows and helices as cylinders. (B) A ribbon diagram of the lowest energy BamE core (residues 40-107) structure with least restraint violations. (C) An ensemble of 20 structures. Colors change progressively from the N-terminus (blue) to the C-terminus (red).

5.3.3. Backbone Dynamics

In parallel with the structural analysis, the dynamic properties of BamE were investigated using ^{15}N T_1, T_2, and heteronuclear NOE relaxation measurements (Figure 5-4). Fitting the T_1 and T_2 data for the ordered mainchain amides (i.e. with heteronuclear NOE values > 0.65) by the model-free formalism yielded a correlation time of approximately 10 nsec for the global isotropic tumbling of BamE. This is somewhat slower than predicted for the lowest energy NMR-derived structure of monomeric BamE using the program HYDRONMR (8.6 nsec), yet faster than expected for a globular 21 kDa dimer (~12 nsec) (Sakakibara et al., 2009). This difference may reflect weak self-association. Alternatively, the disordered termini and large L3 loop may lead to an increased effective hydrodynamic size. This is consistent with the gel filtration studies in which the BamE monomer was observed to elute from the column slightly earlier than expected, at a volume corresponding to a protein species of approximately 15 kDa in size rather than 11 kDa).

In addition to reflecting global motions, amide ^{15}N relaxation provides insights into the local backbone motions of a protein. The residue-specific ^{1}H-^{15}N heternuclear NOE values and fit model-free order parameters S^2 of BamE indicate that indeed both the N and C termini are highly flexible on the nanosecond-picosecond timescale (Figure 5-4). However, the N-terminal residues proceeding $\alpha 1$ may not be entirely unrestricted. Some local order is suggested by the NOE and S^2 values in this region that are intermediate between those of the more distal, highly flexible terminal residues and of those of the ordered helices and strands. The extended loop L3 is also conformationally flexible on this fast timescale, although its motions are dampened relative to those of the terminal regions.

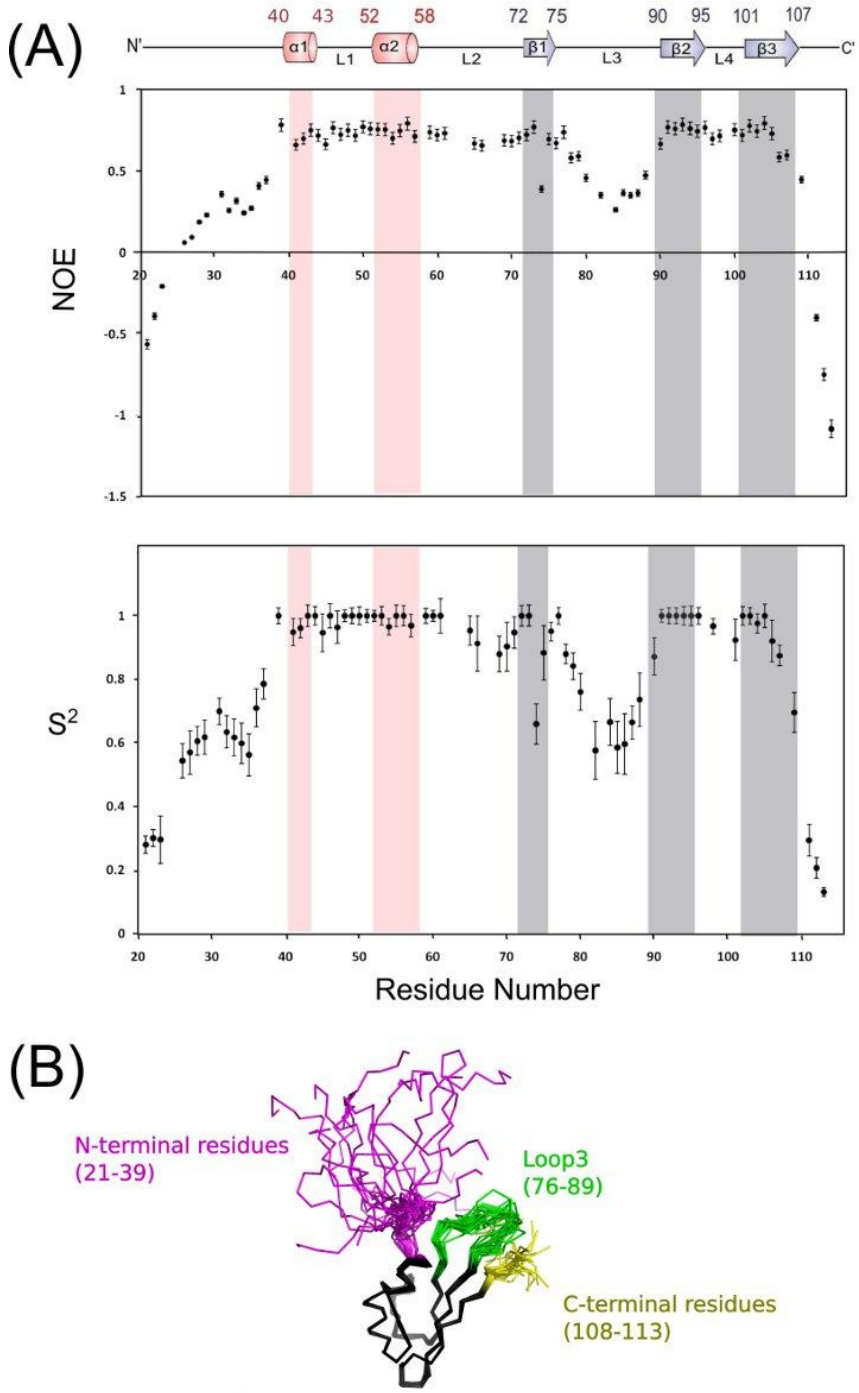


Figure 5-4 Backbone Dynamics of E. coli BamE from Amide ^{15}N Relaxation Analysis

(A) Plots of heteronuclear NOE (upper panel) and fit isotropic model-free S^2 values (lower panel) versus sequence are shown. Smaller NOE and S^2 values, indicative of significant sub-nsec timescale backbone motions, are observed for the both N- and C-termini, as well as the loop L3. (B) These dynamic regions correspond to regions of the BamE structural ensemble with the highest r.m.s. deviations.

5.3.4. Conserved Residues and Molecular Surface Properties

Comparison of the sequence of *E. coli* BamE to those of its homologues from various Gram-negative bacterial species reveals a number of conserved amino acids (Figure 5-5A). The majority of the conserved residues in the core of BamE (Gly49, Gly60, Pro62, Tyr75, and Phe95) reside on the loops or turns (Figure 5-5B). Gly49 is located in L1 (between α1 and α2) where it participates in a type II β-turn, whereas Gly60 and Pro62 are found at turning points of L2 (between α2 and β1). Two conserved aromatic residues, Tyr75 and Phe95, are found as the last residues of β-strands β1 and β2, respectively. The side chains of both these residues point towards the interface between the helices and the β-sheet. Another conserved residue Gln54 is found in α2. When these conserved residues are mapped onto the surface view of the BamE structure (Figure 5-5B), they are seen to be clustered in two separate patches. Analyzing the electrostatic properties of solvent accessible molecular surface of BamE showed that the protein has positively charged residues clustered on the surface formed by the two N-terminal α-helices (Figure 5-6). On the other hand, the V-shaped surface formed by α1 and β3 is hydrophobic (Figure 5-6). Further experiments are needed to verify whether these regions of BamE are involved in interaction with other proteins (e.g. other components of the BAM complex or with substrate proteins), or if they are important mainly for the structural stability and folding of this protein.

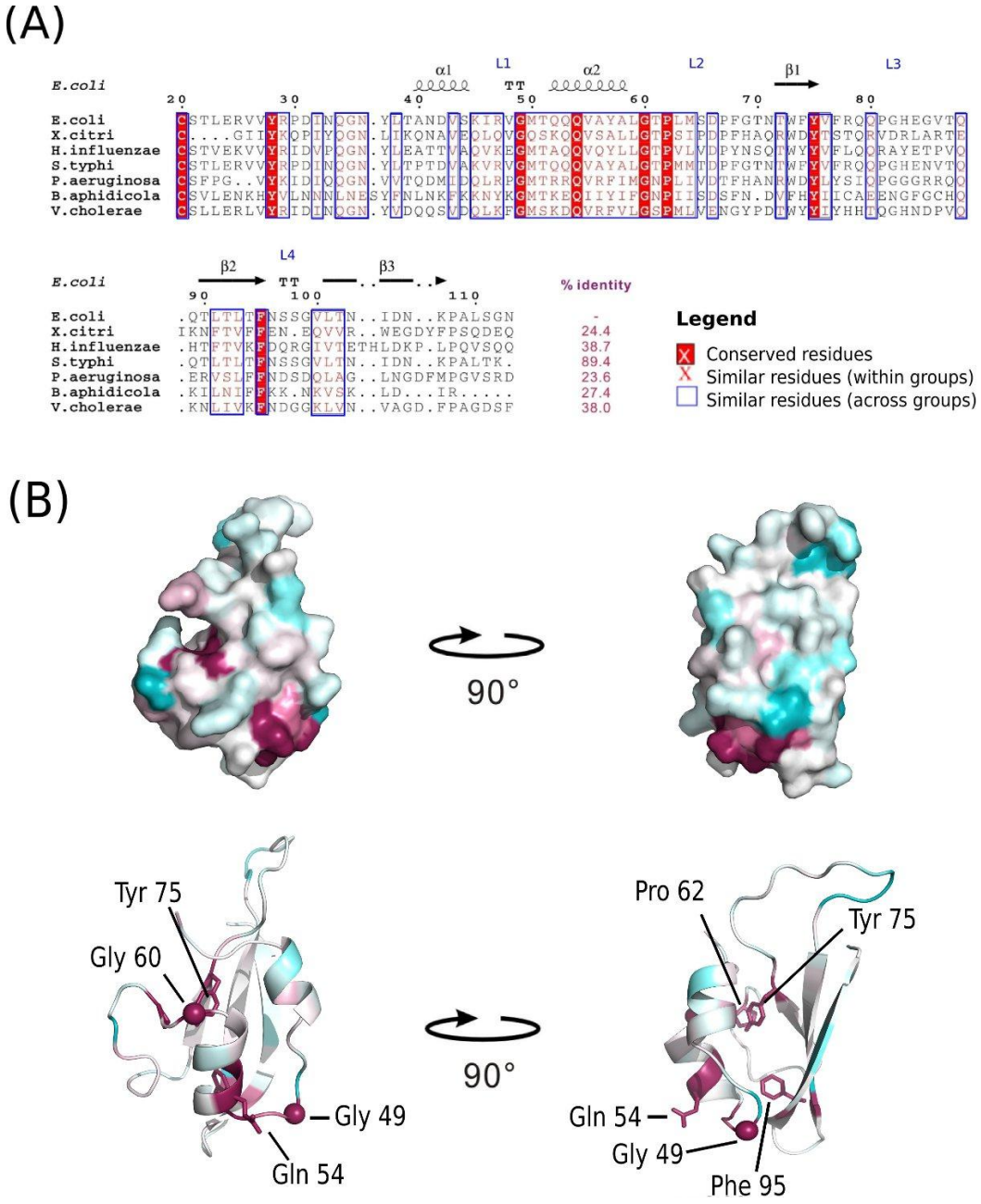


Figure 5-5 Conserved Residues of BamE

(A) Sequence alignment starting from the invariant N-terminal cysteine residue. The NMR-derived secondary structure of *E. coli* BamE as classified by DSSP is shown above the alignment. The protein sequences were acquired from the Swiss-Prot data base: *E. coli* (P0A937); *X. citri* (Q8PMB6); *H. influenza* (P44057); *S. typhi* (Q8XF17); *P. aeruginosa* (O68562); *B. aphidicola* (Q8K9V7); *V. cholerae* (P0C6Q9). (B) A view of BamE sequence conservation mapped onto the BamE surface (top). Individual amino acid residues are colored according to the degree to which they are conserved; absolutely conserved residues are shown in maroon, while highly variable residues are shown in blue. In the ribbon diagram (bottom), the conserved residues are shown in stick representation.

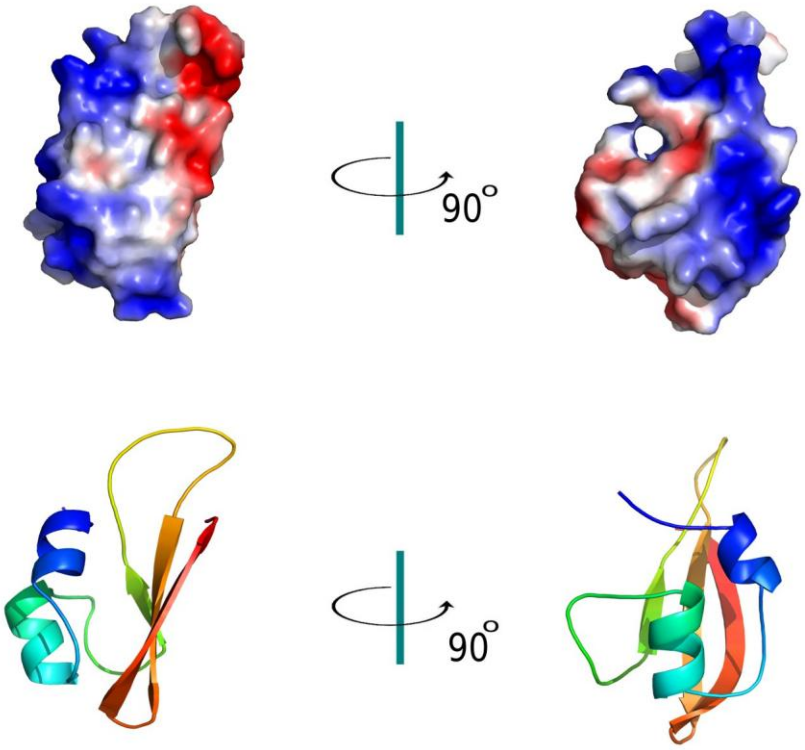


Figure 5-6 Electrostatic Properties of BamE Molecular Surface

The electrostatic potential is mapped onto the solvent accessible surface of BamE (upper panel). The red, blue and white represent negative, positive and neutral potentials, respectively. The protein is also shown in ribbon diagram (lower panel) in the same orientation as the surface diagram.

5.3.5. Structural Homologues

The structure of BamE closely resembles that of OmlA (Vanini et al., 2008) (PDB: 2PXG), a BamE homologue found in *Xanthomonas axonopodis pv. citri* (24.4% sequence identity). Both possess similar secondary structural elements and an overall tertiary fold, and the backbone atoms of the α-helices and the β-sheet can be superimposed with an r.m.s. devation value of 2.66 Å (Figure 5-7). Although quite similar in architecture, three notable differences were observed between the BamE and OmlA structures. 1) Residues corresponding to α1 in BamE are disordered in OmlA. 2) The angle between the α2 helix and the C-terminal β-sheet is more acute in the OmlA structure. 3) The flexible N and C termini of OmlA are significantly longer than in BamE.

A search for structural homologues using the DALI (Holm et al., 2008), CATH (Orengo et al., 1997) and FATCAT servers (Ye and Godzik, 2003) identified several additional proteins that have a significant degree of similarity in topology and architecture with BamE. Proteins with a BamE-like fold include *Streptomyces clavuligerus* BLIP (β-lactamase inhibitor protein) ((Reynolds et al., 2006; Strynadka et al., 1996) PDB: 2G2U) (Figure 5-7A and B), the dimerization domain of an *E. coli* disulfide bond isomerase known as DsbC ((McCarthy et al., 2000). PDB: 1EEJ) (Figure 5-7C), *Thermus thermophilus* TTHA1718, a putative heavy metal binding protein ((Sakakibara et al., 2009) PDB: 2ROE) (Figure 5-7E), and *Hirudo medicinalis* EglinC, an elastase (a serine protease) inhibitor ((Bode et al., 1987). PDB: 1CSE) (Figure 5-7F).

Surprisingly, the search results from all three databases indicate that BamE shares more structural similarity with BLIP, a protein that inhibits a variety of class A β-lactamase enzymes, than with its sequence homologue, OmlA. Structural comparison of BamE and BLIP suggests that BLIP has a tandem repeat of BamE-like folds, as each of the N- and the C-terminal domains of BLIP superimpose well onto the BamE structure with r.m.s. deviation values of 1.91 Å and 3.34 Å, respectively (Figure 5-7A and B). It is interesting to note that the loop L3 of BamE, which was observed to be mobile from our NMR relaxation experiment, is found in a structurally equivalent position as the active site binding loop found in both domains of BLIP (Reynolds et al., 2006). In BLIP, Asp49 found in the active site binding loop serves as a key residue involved in the interaction with the β-lactamase enzymes (Reynolds et al., 2006). Vanini *et al*. (Vanini et al., 2008) observed that Asp62 of OmlA and the functionally important Asp49 of BLIP are found in a structurally equivalent position in both proteins. In our *E. coli* BamE structure, a glutamate (Glu84) residue is found at the equivalent position within L3. L3 is also topologically equivalent to an active site binding loop in EglinC, a protein based inhibitor of the serine protease elastase (Bode et al., 1987).

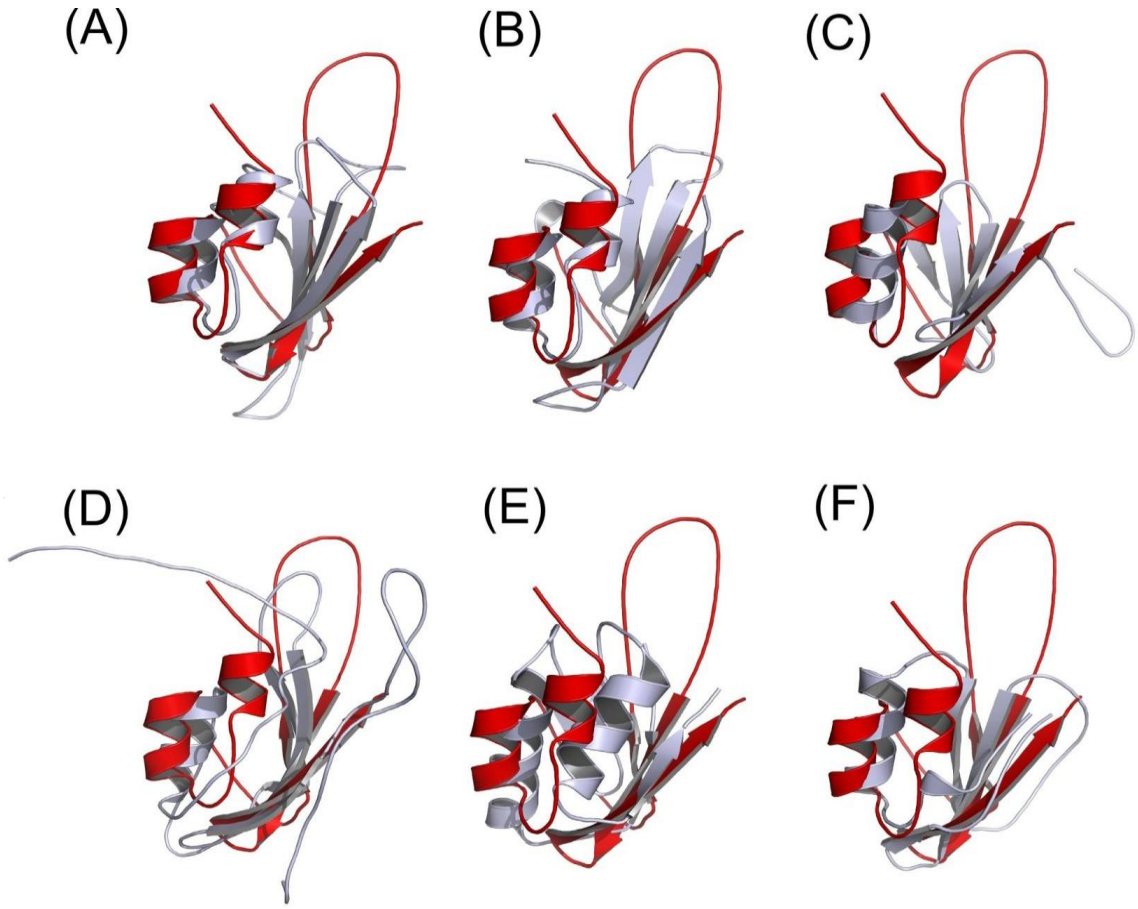


Figure	Superimposed Protein	Organism	PDB ID	RMSD (Å)	Seq ID (%)
(A)	BLIP (domain 2)	*S. clavuligerus*	2g2u	1.91	17.8
(B)	BLIP (domain 1)	*S. clavuligerus*	2g2u	3.34	19.8
(C)	DsbC (N-terminal domain)	*E. coli*	1eej	3.82	8.1
(D)	OmlA	*X. citri*	2pxg	2.66	24.4
(E)	TTHA1718	*T. thermophilus*	2roe	3.15	13.3
(F)	Eglin C	*H. medicinalis*	1cse	2.44	7.5

Figure 5-7 Structural Homologues of BamE

E. coli BamE (red) is superposed on the structures of proteins with similar topology and architecture (white). The r.m.s. deviation values were calculated against the backbone atoms of the α-helices and β-sheets of the lowest energy BamE structure.

5.4. Discussion

This chapter has presented the structural and dynamic properties of monomeric BamE determined by NMR spectroscopy. The NMR structure of BamE revealed that BamE folds as two α-helices packed against a three-stranded anti-parallel β-sheet. In contrast to the globular core of the protein, the N-terminal and C-terminal regions of BamE, as well as the L3 loop, were found to be highly flexible on the sub-nsec timescale.

Structural comparison of BamE to proteins with a similar architecture suggests a possibility of the L3 loop of BamE serving a function as a protein binding motif. The flexible L3 loop shows structural and topological similarity to the protein binding loops of BLIP and EglinC, both of which are BamE structural homologues. The L3 loop of BamE therefore may be the site of protein-protein interaction between BamE and other components of the BAM complex. So far, no evidence has been reported on the ability of BamE to bind substrates (i.e. unfolded OMPs). However, BamE has previously been shown to strengthen the interaction between BamA and BamD (Sklar et al., 2007a); therefore, the POTRA domains of BamA and BamD are the most likely candidates that may interact with the L3 loop of BamE. To verify this hypothesis, deleting or mutating residues of the L3 loop could be carried out to test whether the BamE mutants are still able to co-purify with BamA and BamD as a BamA/D/E subcomplex.

In addition to the structure determination, we have discovered that BamE expressed and purified from *E. coli* also exists in a kinetically-trapped dimeric state that has dramatically different NMR spectra, and hence structural features, relative to its monomeric form. Following the publication of our data, studies from other research groups have confirmed the ability of BamE to dimerize (Albrecht and Zeth, 2011; Knowles et al., 2011). However, there are conflicting data on which oligomeric form of BamE is biologically relevant. While one study reported that the formation of BamE dimer is a result of protein misfolding under temperature stress (Knowles et al., 2011), another study reported that BamE purified from a native outer membrane exhibits a dimeric state (Albrecht and Zeth, 2011). We have also proposed, based on the irreversible nature of BamE dimerization under various laboratory conditions, that BamE may form a domain-swapped dimer. Recently, a crystal structure of *E. coli* BamE in its

dimeric form has been solved, and it revealed that BamE does indeed form a domain-swapped dimer, in which the α-helices of the two monomers are exchanged (Figure 5-8) (Albrecht and Zeth, 2011). However, since the functional significance of the BamE dimer remains to be established, further experiments are required to identify which oligomeric form of BamE is found *in vivo*, or if both forms are functional in the BAM complex.

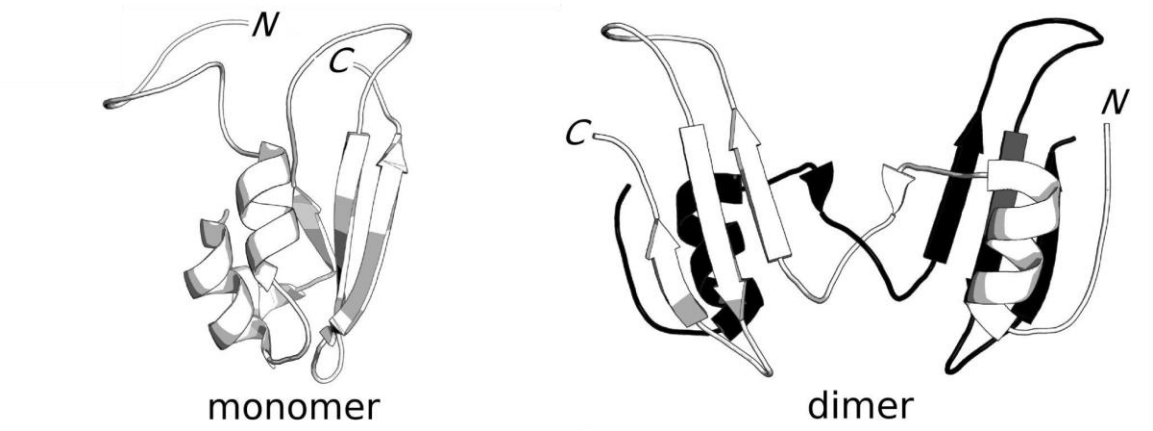


Figure 5-8 The Structure of BamE Dimer

The ribbon diagrams of a BamE monomer (PDB: 2KXX) and a dimer (PDB: 2YH9) are shown. The dimer structure shows how the two monomers (shown in white and black) exchange N-terminal α-helices to form a domain-swapped dimer. This figure was adopted from Kim *et al*. (2012).

6. Crystal Structure of the C-terminal Domain of BamC

Note regarding contributions:

This chapter was published in Acta Cyrstallographica Section F. The authors and the full reference for the article are listed below.

Kim, K.H., Aulakh, S., Tan, W., and Paetzel, M. (2011). Crystallographic analysis of the C-terminal domain of the *Escherichia coli* lipoprotein BamC. *Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun.* **67**, 1350-1358.

The work presented in this chapter was a collaborative effort with Suraaj Aulakh (M.Sc. candidate) and Wendy Tan (undergraduate research student, Summer 2010) of the Paetzel Lab. All the figures and tables included in this chapter, otherwise noted, were adopted from the article.

6.1. Introduction

BamC is the second largest (34.4 kDa) lipoprotein component of the BAM complex. It has been shown to associate with the rest of the BAM complex subunits indirectly via its interaction with BamD (Malinverni et al., 2006). Similar to BamB and BamE, absence of BamC results in non-lethal impairment of OMP assembly. Mutants lacking BamC display outer membrane permeability defects and reduced levels of OMPs in the OM (Onufryk et al., 2005).

The first structural study of BamC was carried out by Knowles and his colleagues who used NMR to determine the secondary structure of the protein (Knowles et al., 2009a). Their NMR data predicted that BamC has a mixture of α- and β-secondary structure elements, and that approximately 70 residues at the N-terminus of BamC is unstructured (Knowles et al., 2009a). It has also been shown that subjecting BamC to a small amount of broad-range protease subtilisin results in the degradation of the full length BamC protein into two protease-resistant fragments (12.2 kDa and 14.5 kDa) corresponding to residues 75-195 and 196-328, respectively (Albrecht and Zeth, 2010; Kim et al., 2011b; Warner et al., 2011). Taken together, these results suggested that BamC has two independently folding domains (the N- and the C-terminal domains) following the unstructured N-terminal region (Figure 6-1).

This chapter presents the crystal structure of the C-terminal domain of *E. coli* BamC. Analysis of the structure and its crystal contacts, as well as comparative study with structural homologues, provides insights into how BamC may interact with other proteins.

6.2. Materials and Methods

6.2.1. Protein Overexpression and Purification

BamC was overexpressed and purified as described previously in Chapter 2 of this thesis.

6.2.2. Limited Proteolysis

For a limited proteolysis experiment, BamC (200 µL at 1 mg/mL) was digested with chymotrypsin (1000:1 BamC:chymotrypsin ratio by mass) and sampled at time points t = 0, 5, 10, 15, 30, 45, 60 and 120 min and overnight. The aliquots were mixed with SDS–PAGE loading dye, boiled for 3 min to stop the reaction, and then run on SDS–PAGE for analysis.

6.2.3. Crystallization and Data Collection

Crystals were grown by the hanging drop vapor diffusion method. The crystallization drops were prepared by mixing 1 µL of protein (30 mg/mL) suspended in buffer A (20 mM Tris-HCl pH 8.0; 100 mM NaCl) with 1 µL of reservoir solution, and then equilibrating the drop against 1 mL of reservoir solution. The proteolytically resistant fragment of the BamC yielded crystals in the space group H3 with unit cell dimensions of a = 78.9, b = 78.9, c = 52.9 Å. The crystals have one molecule in the asymmetric unit with a Matthews coefficient of 2.4 $Å^3$/Da (49.4% solvent). The optimal crystallization reservoir condition was 0.1 M NaCl, 0.1 M HEPES, pH 6.5 and 1.6 M $(NH_4)_2SO_4$. Crystallization was performed at room temperature (~22 °C). The cryo-solution condition contained 0.1 M NaCl, 0.1 M HEPES, pH 6.5 and 1.6 M $(NH_4)_2SO_4$, and 30% glycerol. Crystals were washed in the cryo-solution before being flash-cooled in liquid nitrogen.

Diffraction data was collected on beamline 08ID-1 at the Canadian Macromolecular Crystallography Facility (CMCF) of the Canadian Light Source (CLS), using a MarMosaicRayonix MX300 CCD X-ray detector. The X-ray wavelength used was 0.98058 Å. The crystal-to-detector distance was 180 mm. A total of 180 images were collected with a 0.35° oscillation, and each image was exposed for 0.5 s. The diffraction data was processed with the programs: iMOSFLM (Powell, 1999), POINTLESS (Evans, 2006) and SCALA (Evans, 2006). See Table 6-1 for data collection statistics.

Table 6-1 Crystallographic Statistics for BamC

The data collection statistics in parentheses are the values for the highest resolution shell. See Appendix D for definitions of statistical values reported.

Crystal Parameters	
Space group	H3
a,b,c (Å)	78.9, 78.9, 52.9
Data Collection Statistics	
Wavelength (Å)	0.98058
Resolution (Å)	28.7 – 1.5 (1.6 – 1.5)
Total reflections	110632 (16027)
Unique reflections	19576 (2877)
R_{merge}	0.075 (0.378)
Mean (*I*)/σ (*I*)	10.3 (3.2)
Completeness (%)	99.7 (100.0)
Redundancy	5.7 (5.6)
Refinement Statistics	
Protein chains in A.U.	1
Residues	120
Water molecules	99
Total number of atoms	1029
R_{cryst}/ R_{free} (%)	16.1 / 18.3
Average B-factor (Å^2) (all atoms)	28.0
Rmsd on angles (º)	0.016
Rmsd on bonds (Å)	1.60

6.2.4. Structure Determination and Refinement

Although the full length BamC construct (Ser26-Lys344) was used for crystallization, the protein was cleaved to a smaller fragment in the crystallization drop during incubation. The determined unit dimensions, and symmetry for these crystals are inconsistent with the full length BamC construct fitting into the crystal lattice based on its molecular mass. Molecular replacement trial with the N-terminal domain of BamC (PDB: 2YH6) as a search model failed. However, phases were obtained by molecular replacement using the program PHASER 2.1(McCoy et al., 2007) when the C-terminal

domain structure of *E. coli* BamC recently reported (PDB: 2YH5) was used as a search model. Clear electron density can be seen for residues Ala224-Lys343, which corresponds to the C-terminal domain of BamC. The structure was refined using restrained refinement in REFMAC5 (Murshudov et al., 1997), and further manual adjustments to the atomic coordinates were performed with the program COOT (Emsley & Cowtan, 2004). The final model was obtained by running restrained refinement in REFMAC5 with TLS restraints obtained from the TLS motion determination server (Painter & Merritt, 2006). The refinement statistics are shown in Table 6-1. The atomic coordinates and structure factors have been deposited in the RCSB Protein Data Bank (PDB: 3SNS).

6.3. Results

6.3.1. Purification, Limited Proteolysis and Crystallization

Our limited proteolytic analysis of full-length BamC (residues 26–344) with chymotrypsin revealed a proteolytically resistant fragment of approximately 25 kDa in size (Figure 6-1), which is close to the molecular mass sum of the N-terminal (13 kDa) and C-terminal (13.4 kDa) domains. If the ~25 kDa protease-resistant fragment contains only the globular domains of BamC (and not the unstructured N-terminal region), it must now also lack the N-terminal histidine tag. To determine whether the N-terminal histidine tag was lost owing to proteolysis, the ~25 kDa BamC fragment was run on a nickel affinity affinity column. At this point (48 hours following the initial proteolysis reaction), most of the ~25 kDa species was seen to have cleaved even further to a smaller ~14 kDa species (Figure 6-1C). Both the remaining ~25 kDa and ~14 kDa protein fragments of BamC did not bind to the column, suggesting that proteolysis had occurred at the N-terminus of the protein. This cleavage pattern is consistent with the secondary structure predictions and recently reported NMR data, which suggested that the first ~70 residues of the protein are unstructured (Warner et al., 2011; Knowles et al., 2009) while the rest of the protein folds into two separate globular domains (Figure 6-1; Albrecht & Zeth, 2010, 2011).

Interestingly, when full-length BamC was screened for crystallization, the protein appeared to experience natural degradation in the crystallization drop. Crystals of $BamC_C$ (the C-terminal domain of BamC; residues 224–343) formed in space group H3 and diffracted to beyond 1.5 Å resolution. The rest of this chapter describes the crystal structure and the structural analysis of $BamC_C$.

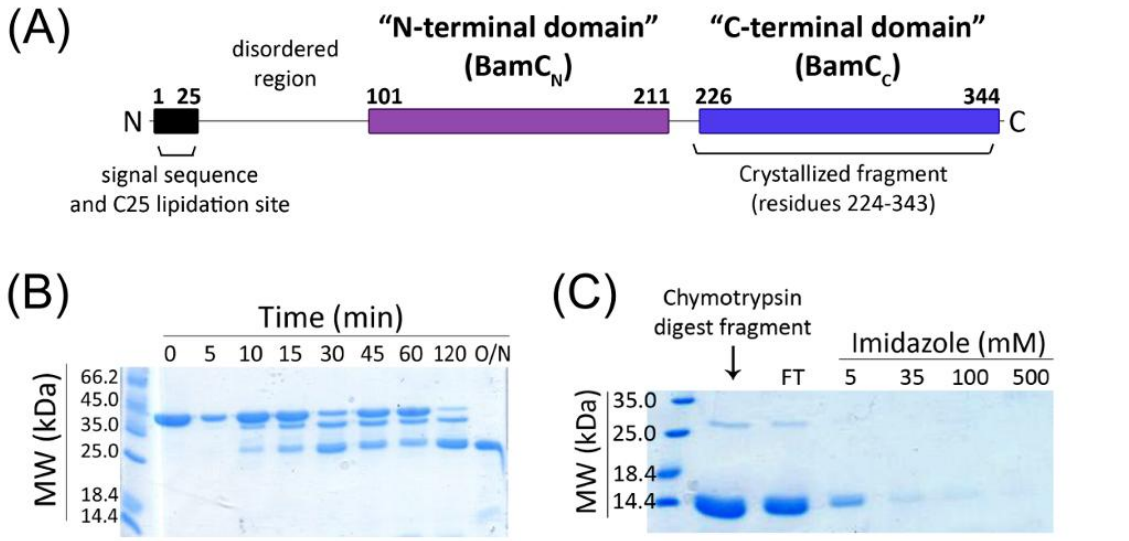


Figure 6-1 Structural Regions of BamC as Determined by Limited Proteolysis

(A) A schematic diagram of the BamC domains. (B) Following purification, the full length BamC (36 kDa) was subjected to limited proteolysis with chymotrypsin. A ~25 kDa fragment started to appear on the gel after 10 minutes of proteolysis reaction. After overnight incubation, it can be seen that all of the full-length BamC has been completely cleaved by the protease and that only the ~25 kDa protease resistant fragment remained. (C) The ~25 kDa BamC fragment was subsequently run on a nickel affinity column to determine whether the N-terminal hexahistidine tag was lost due to proteolysis. At this point, most of the ~25 kDa species can be seen to have cleaved even further to a smaller ~14 kDa species. Both the remaining ~25 kDa and ~14 kDa protein fragments of BamC did not bind to the column, suggesting that the proteolysis occurred predominantly on the N-terminus of the protein.

6.3.2. Overall Protein Fold and Molecular Surface Properties

$BamC_C$ has an 'helix-grip' fold that consists of a six-stranded twisted antiparallel β-sheet, three α-helices, and one 3_{10}-helix (Figure 6-2A and B; Figure H1B). Two α-helices (α1 and α3) pack against one side of the β-sheet via hydrophobic interactions, and a shorter helix α2 is positioned adjacent to the other two helices away from the β-sheet. A short 3_{10}-helix, η1, is part of a loop region between β-strands β3 and β4. The

twisting of the β-sheet creates a concave surface on the solvent-accessible side of the sheet, predominantly formed by β2, β5 and β6. This groove has a surface area of approximately 485 $Å^2$ and is lined with negatively charged residues. Analysis of the overall electrostatic properties of $BamC_C$ reveals a predominantly negatively charged surface (Figure 6-2C) consistent with its theoretical isoelectric point of 4.7.

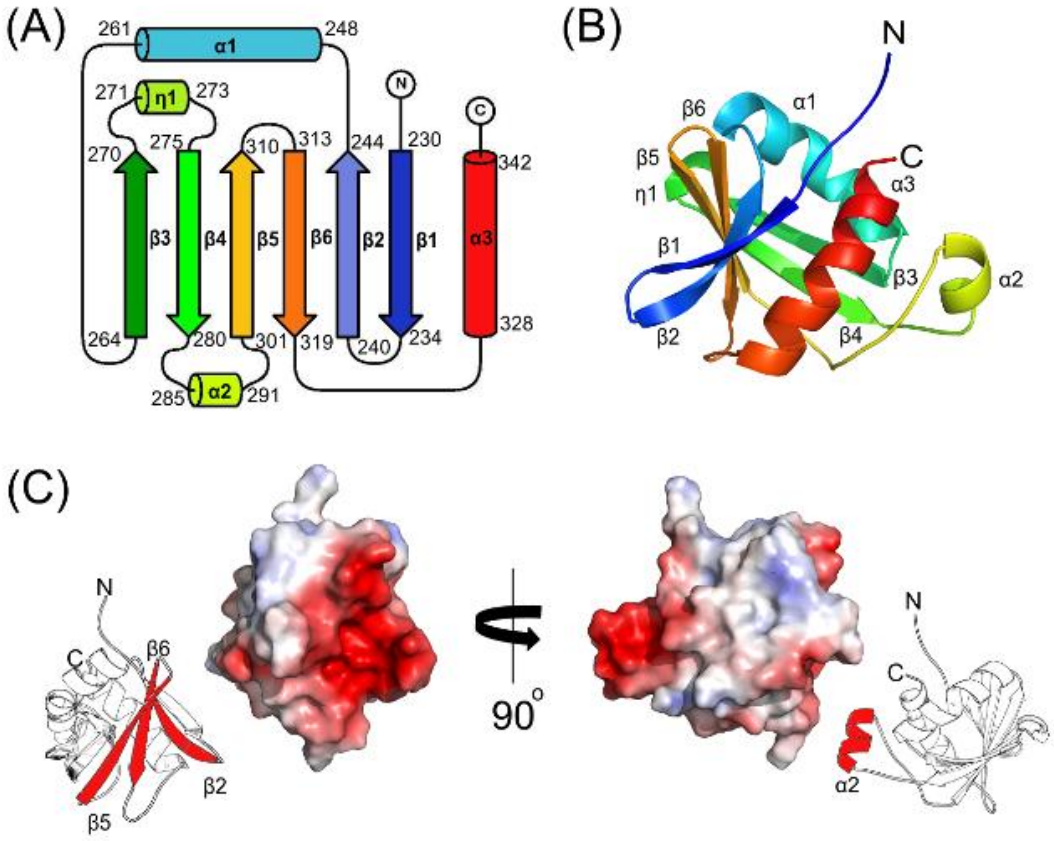


Figure 6-2 $BamC_C$ Fold and Surface Features

(A) A topology diagram of $BamC_C$ is shown with strands represented as arrows and helices as cylinders. (B) A ribbon diagram of $BamC_C$. The color changes progressively from the N-terminus (blue) to the C-terminus (red). (C) The electrostatic potential is mapped onto the surface of $BamC_C$. The red and blue represent negative and positive potentials, respectively.

6.3.3. Conserved Residues

Sequence comparisons of *E. coli* BamC with functional homologues indicates that although there are conserved residues throughout the protein, the majority of the conserved blocks of sequence reside in the unstructured N-terminal region (Ser26-

Thr100) and within $BamC_C$ (Figure 3A). When the conserved residues within $BamC_C$ are mapped onto our structure, all of the invariant residues are found to be solvent exposed, and many of them (Arg245, Asp309, Asn312, Arg313 and Ser315) are located on β5 and β6, and in close proximity to each other such that they form a continuous surface (Figure 3B). This conserved patch on the $BamC_C$ surface forms part of the negatively charged concave surface, mentioned earlier.

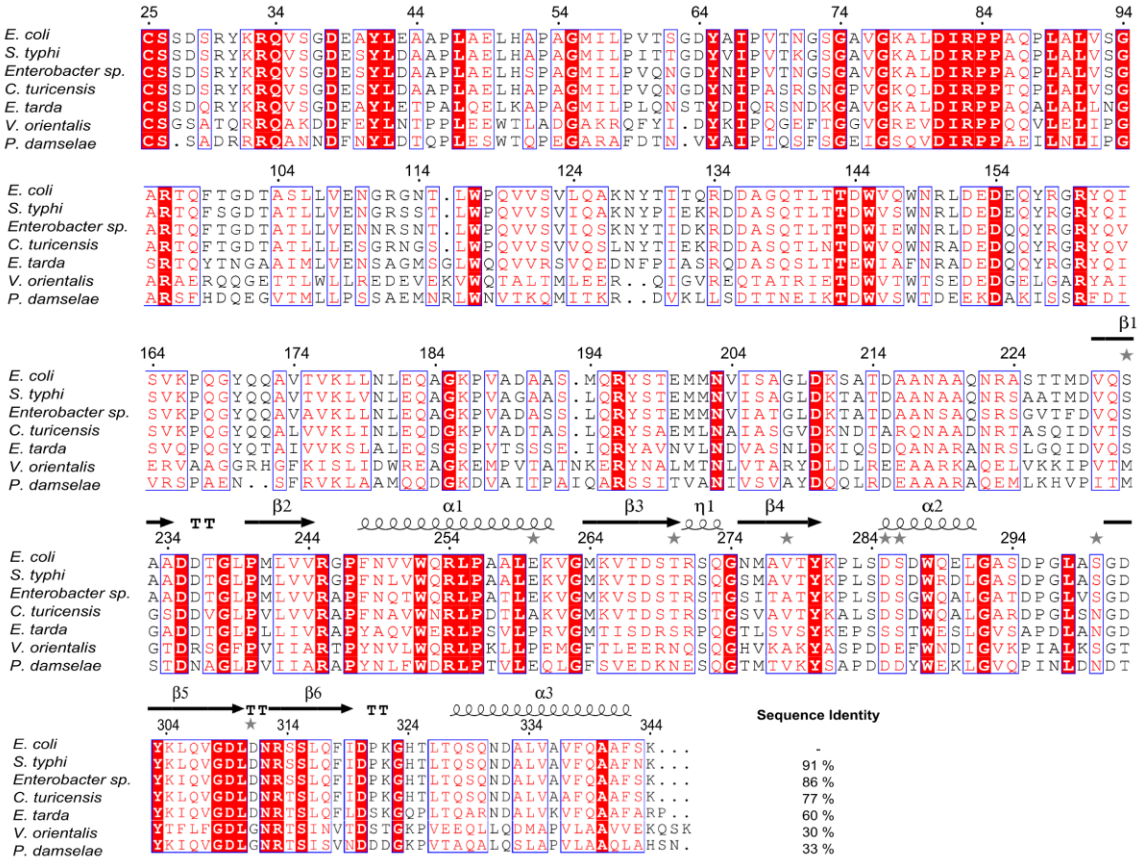


Figure 6-3 Multiple Sequence Alignment of E. coli BamC

A multiple sequence alignment of *E. coli* BamC with homologous proteins is shown. The secondary structure of *E. coli* $BamC_C$ is shown above the alignment. The protein sequences were acquired from the Swiss-Prot data base: *Escherichia coli* (P0A903); *Salmonella typhi* (Q8Z4R9); *Enterobacter sp.* (A4WD57); *Cronobacter turicensis* (C9XXL0); *Edwardsiella tarda* (E0T3S3); *Vibrio orientalis* (C9QLI3); *Photobacterium damselae* (D0YWF8).

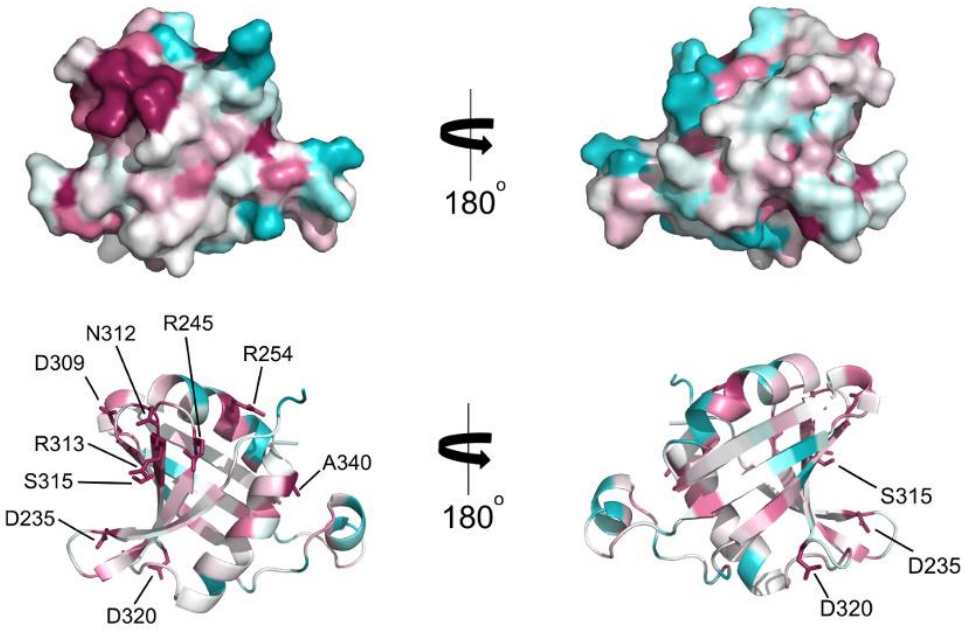


Figure 6-4 BamC$_C$ Conservation Mapped onto the Structure

A view of BamC$_C$ conservation mapped onto the molecular surface generated using the above alignment. Absolutely conserved residues are shown in maroon, highly variable residues are shown in cyan. The ribbon diagram (bottom) shows the conserved residues in stick representation.

6.3.4. Potential Protein-Protein Interaction Site

The BamC$_C$ crystal from which the diffraction data were collected had a space group of H3. Each BamC$_C$ molecule in the crystalline lattice is observed to make three major crystal contacts. The largest interface (875 Å^2) is formed by helix α1 lying within the negatively charged groove of the neighbouring molecule (predominantly involving β5 and β6 of the β-sheet) in the crystalline lattice (Figure 6-5 A and B). The interaction is mostly mediated via hydrogen bonds and electrostatic interactions (Figure 6-5A, inset and Figure 6-7D). The molecular surface involved in this interaction is the previously mentioned negatively charged cavity where many of the conserved residues are located.

To examine how other proteins with BamC$_C$-like folds interact with their binding partners, we performed a search for structural homologues using the DALI server (Holm et al., 2008), and identified several proteins that have similar topology and architecture as BamC$_C$. One such protein is *Homo sapiens* AP2 complex β-appendage domain, which has previously been co-crystallized with a peptide that mimics a region of β-

arrestin, the substrate (PDB: 2IV8) (Schmid et al., 2006). A close examination of the co-crystal structure shows that the binding region of the substrate, which is a short α-helical peptide, binds to the concave surface formed by the twisted β-sheet of the AP2 complex β-appendage domain (Figure 6-5C); this is reminiscent of the major crystal packing interaction observed in the $BamC_C$ crystal. When the structures of the AP2 complex β-appendage domain and the $BamC_C$ are superimposed (backbone r.m.s.d. of 3.0 Å), it can be seen that the binding pockets of the two proteins align closely (Figure 6D).

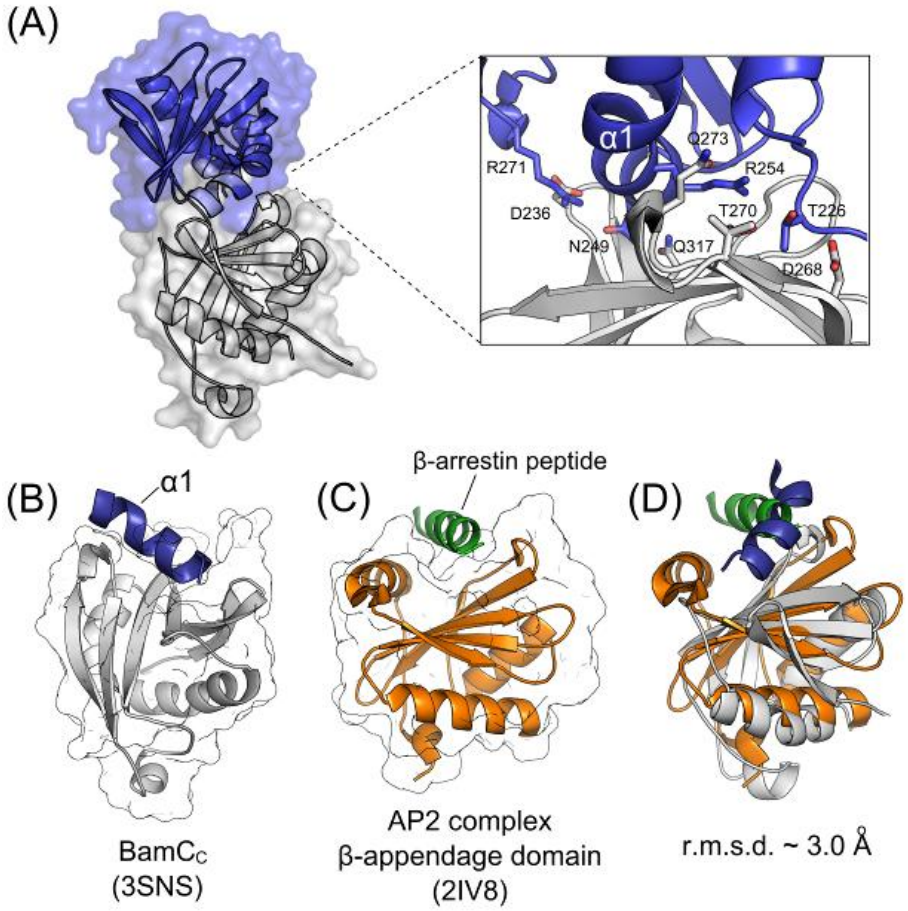


Figure 6-5 A Potential Protein Interaction Surface on $BamC_C$

(A) The most extensive protein-protein interaction within the $BamC_C$ H3 crystalline lattice is shown. A close up view of the residues at the interface is shown (inset). (B) A different view of the interface shows how $BamC_C$ (grey) interacts with α1 of a neighbouring $BamC_C$. Only α1 helix of the interacting $BamC_C$ monomer (blue) is shown for clarity. (C) A structural homologue of $BamC_C$, *Homo sapiens* AP2 complex β-appendage domain (orange) is shown with the bound substrate, a peptide that mimics a region of β-arrestin (PDB: 2IV8) (green). (D) The structure of the AP2 complex β-appendage domain (orange) bound to β-arrestin peptide (green) is superimposed on the structures of $BamC_C$ (grey) and α1 of a neighbouring $BamC_C$.

6.3.5. Comparison with Other BamC Structures

Crystal and NMR structures of both the N-terminal and the C-terminal domains of BamC ($BamC_N$ and $BamC_C$, respectively) were reported while we were in the process of determining our $BamC_C$ structure (Warner et al., 2011; Albrecht & Zeth, 2011).Our $BamC_C$ structure in space group H3 superimposes with the other $BamC_C$ crystal structure in space group $P2_1$ with an overall main-chain r.m.s.d. of 0.7 Å (PDB: 2YH5)(Albrecht & Zeth, 2011) and with the lowest energy NMR solution structure of $BamC_C$ with an overall main-chain r.m.s.d. of 0.9 Å (PDB: 2LAF) (Figure 6-6A)(Warner et al., 2011). Although there is a slight variation in the positions of α1 and α2 helices, the three structures show very close resemblance overall. The structure of the β-sheet is especially well conserved between the compared coordinates (Figure 6-6B and C). Comparing the B-factor distribution of the two crystal structures show well-ordered β-sheet and flexible N- and the C-termini in both structures (Figure 6-6D). Unlike in the $BamC_C$ structure in space group $P2_1$, the structure in space group H3 shows that α2 and the residues immediately following α2 are the regions of the highest flexibility.

The crystal packing protein-protein interaction surfaces observed in the H3 structure are significantly different from those seen in the $P2_1$ structure (Figure 6-7). In the $P2_1$ structure, the largest crystal contact has an interface area of 503 $Å^2$, while the largest interface in our H3 structure has an area of 875 $Å^2$. A detailed comparison of the protein-protein interactions surfaces (crystal contacts) observed in the $BamC_C$ $P2_1$ and H3 crystals is summarized in Figure 6-7.

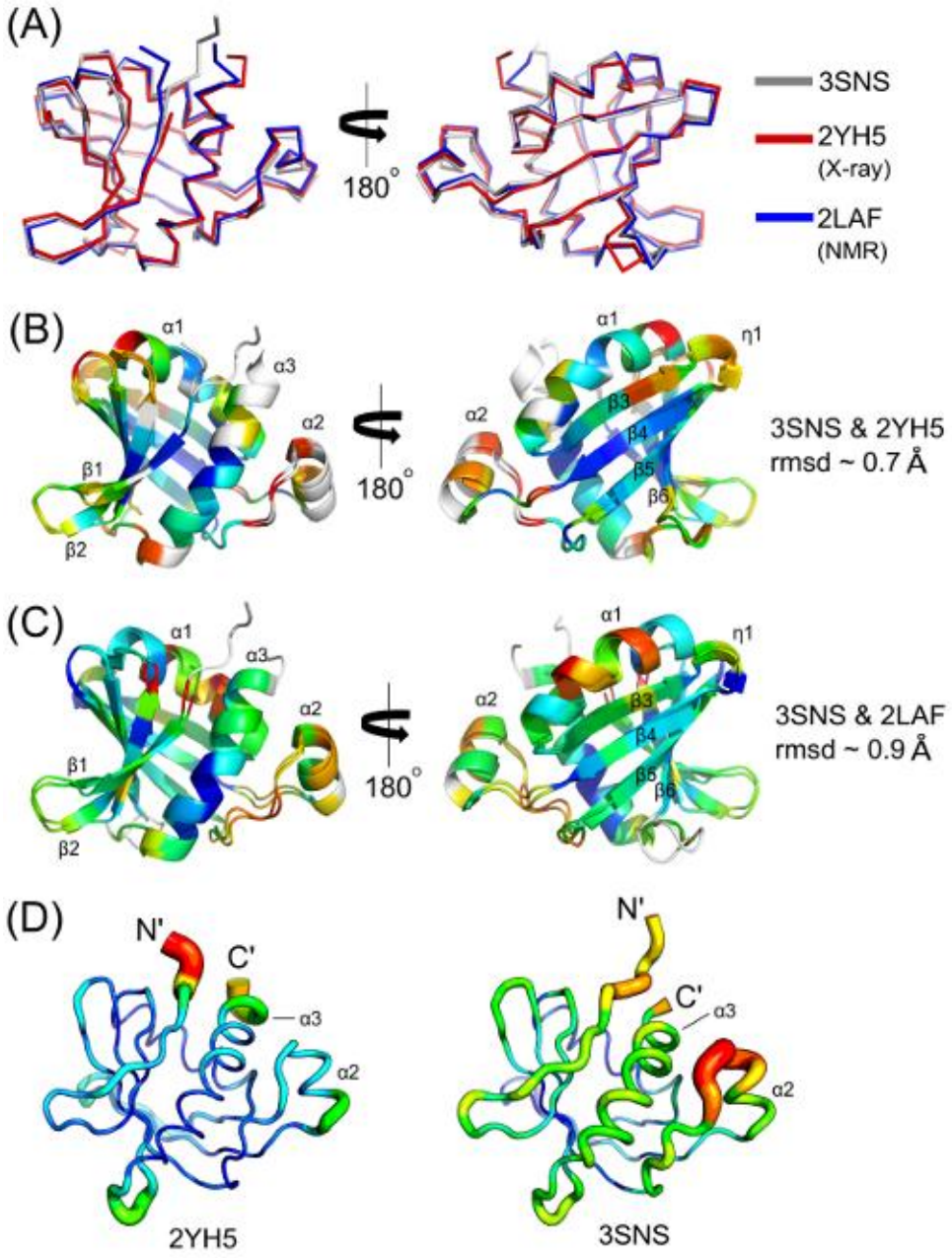
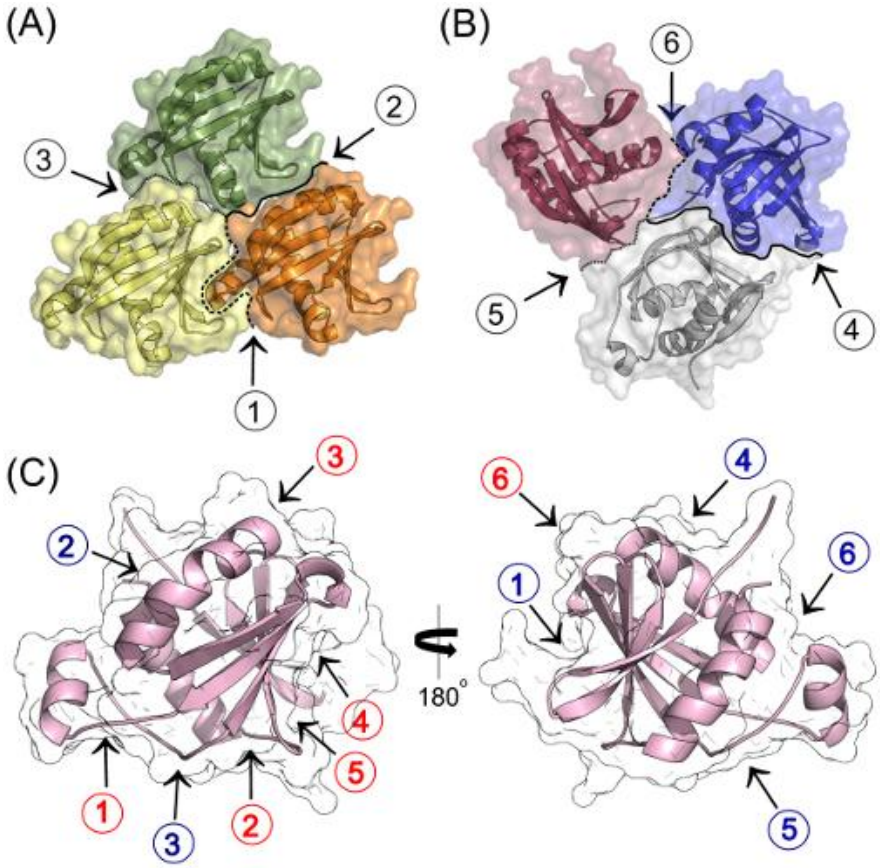


Figure 6-6 Comparison of BamC$_C$ Crystal and Solution Structures

(A) The BamC$_C$ crystal structure (space group H3) (3SNS; grey) is shown superposed with the crystal structure in space group P2$_1$ (2YH5; red) and the NMR solution structure (2LAF; blue). (B) The structures in crystallographic space groups H3 (3SNS) and P2$_1$ (2YH5) are aligned and colored by RMSD. Dark blue represents close alignment, orange/yellow/red represents increasing levels of deviation. Residues not used for alignment are shown in white. (C) The crystal structure in space group H3 (3SNS) and the NMR solution structure (2LAF) are aligned and colored by RMSD, as in (B). (D) `Sausage' representation of the main-chain B-factor distribution of the BamC$_C$ crystal structure in space group H3 (3SNS) and and P21 (2YH5) is shown. Higher B-factor values are indicated by thicker lines and warmer colors.



(D)

Space group	Interface	Interface Area (Å²)	Interfacing Residues	Hydrogen Bonds	Salt Bridges
P 2_1 (PDB: 2YH5)	1	503	31	8	8
	2	278	21	5	0
	3	250	20	2	0
H3 (PDB: 3SNS)	4	875	53	12	4
	5	310	21	4	0
	6	115	12	2	2

Figure 6-7 BamC$_C$ Crystal Contacts

(A) The three largest crystal contacts observed in the $P2_1$ BamC$_C$ crystal structure (2YH5) are shown. Each interface is numbered 1, 2 and 3 in the order of decreasing interface area. (B) The three largest crystal contacts observed in the H3 BamC$_C$ crystal structure (3SNS) are shown. Each interface is numbered 4, 5 and 6 in the order of decreasing interface area. (C) The BamC$_C$ structure is shown in two different views. Crystal contact regions observed in both the H3 and P21 crystal forms of BamC$_C$ are indicated with the same numbering as in (A) and (B). For each crystal contact, one interface surface is labelled in red and its partnering interfacing surface in blue.

6.4. Discussion

BamC is a non-essential lipoprotein of the BAM complex that has a unique modular structure. Our and previously reported limited proteolysis experiments suggest that BamC has two independently folded N- and the C-terminal domains ($BamC_N$ and $BamC_C$), preceded by a long (~70 residues) unstructured N-terminus (Albrecht and Zeth, 2010; Knowles et al., 2009a). The full length BamC protein with the unstructured N-terminal tail has been difficult to crystallize, probably due to disordered conformation of the N-terminal tail (Albrecht and Zeth, 2010; Kim et al., 2011b). However, as described in this chapter, we have been able to successfully determine a high resolution crystal structure of $BamC_C$.

In addition to our $BamC_C$ structure, the structure of $BamC_N$ has also been solved independently by other research groups. The sequence identity of $BamC_C$ and $BamC_N$ is only 12%, yet they superimpose with a backbone r.m.s.d. of 2.7 Å (Albrecht & Zeth, 2011). The two globular domains of BamC have the same 'helix-grip fold' despite sharing low sequence identity with each other (Albrecht and Zeth, 2011; Warner et al., 2011). Both domains consists of a central six-stranded antiparallel β-sheet with two helical units packing tightly against the sheet (Albrecht and Zeth, 2011; Kim et al., 2011b; Warner et al., 2011). The C-terminal domain has an extra 3_{10}-helix in one of the loops connecting the neighboring β-strands. The NMR structure and backbone amide dynamics studies by Warner *et al.* (2011) showed that the two globular domains of BamC are joined by a highly flexible α-helical linker. Thus, BamC forms a modular structure exhibiting conformational flexibility.

Sequence comparisons of BamC from several different species of Gram-negative bacteria indicate that the majority of the conserved blocks of sequence reside in the unstructured N-terminal region and within $BamC_C$ (Albrecht and Zeth, 2011; Kim et al., 2011a; Kim et al., 2011b; Warner et al., 2011). Despite having the same protein fold as $BamC_C$, $BamC_N$ does not comprise many conserved residues (Warner et al., 2011). In $BamC_C$, the conserved residues are found clustered in a groove that could potentially be a site of protein interaction. As described earlier in this chapter, the conserved groove of $BamC_C$ is also the site of a crystal contact where an α-helix from the neighboring BamC molecule binds.

If the conserved groove of $BamC_C$ mentioned above is actually involved in a biologically important protein-protein interaction, one candidate protein that could bind to this region of $BamC_C$ is BamD, which is mostly α-helical and has been proposed to interact with BamC via its C-terminal region (Malinverni et al., 2006). The proposed binding surface of the BamD C-terminal region is lined with several positively charged residues, which would complement the negatively charged $BamC_C$ binding cleft. However, the possibilities of other proteins such as BamE or even substrate proteins binding to this region of $BamC_C$ should not be ruled out.

There are still many aspects of the BamC structure and function that need to be addressed. For example, what is the functional significance of having two domains with the same fold connected by a flexible linker? If each domain of BamC participates in protein binding, a flexible linker in between could perhaps allow the molecule to bind to its interaction partner by wrapping around it or to bind independently to two separate interaction partners. This may be how BamC interacts with BamD, and possibly with substrates. Another unique structural feature of BamC is the unstructured N-terminal region that is ~70 residues long. Interestingly, the N-terminus is the most conserved part of BamC, but currently there is a lack of functional data to help elucidate its role. Recently, however, we successfully solved the structure of $BamC_N$ with the intact unstructured N-terminal region in complex with BamD. The structure revealed that the unstructured region serves as the major interaction surface for association with BamD. More details about the BamCD structure are described in the next chapter (Chapter 7).

7. Crystal Structure of the BamCD Subcomplex

Note regarding contributions:

This chapter was published in the Journal of Biological Chemistry. The authors and the full reference for the article are listed below.

Kim, K.H., Aulakh, S., and Paetzel, M. (2011). Crystal structure of the β-barrel assembly machinery BamCD complex. *J. Biol. Chem.* **286**, 39116-21.

Suraaj Aulakh contributed by generating the protein-protein interaction data involving domain-truncation mutants of BamC. All the figures and tables included in this chapter, otherwise noted, were adopted from the article.

7.1. Introduction

Deciphering the molecular mechanism of the BAM complex (or that of any protein complex for that matter) inevitably requires understanding not only the functional roles of the individual protein subunits, but also how the subunit proteins interact with each other. Within the BAM complex, there are lipoprotein-lipoprotein interactions in addition to the BamA-lipoprotein interactions. The lipoproteins BamC, BamD and BamE have previously been shown to interact with each other via co-immunoprecipitation and also via size-exclusion chromatography studies as described in Chapter 3 (Malinverni et al., 2006; Sklar et al., 2007a), but the exact mode of the interaction is not fully understood yet.

Previous research and also our work have shown that the POTRA domains of BamA interact with the BamB-E lipoproteins (Kim et al., 2007). While BamB and BamD independently interact with the POTRA domains directly, BamC and BamE require BamD to co-purify with BamA (Malinverni et al., 2006; Sklar et al., 2007a). Direct interaction between BamC and BamD has also been shown, with mutagenesis data suggesting that the C-terminus of BamD (residues 227-245) is necessary for the association (Malinverni et al., 2006). However, the region of BamC involved in the interaction with BamD was not identified. We therefore wanted to identify this region to gain more insight into the nature of the BamC-BamD interaction.

As described in the preceding chapter, BamC consists of two independently folding domains (the N- and the C-terminal domains) following the unstructured N-terminal region of approximately ~70 residues in length. This chapter describes a series of experiments in which BamC domain truncation mutations were made and screened for interaction with BamD. Also described in this chapter is the crystal structure of the BamCD subcomplex. The interaction data along with the crystal structure provided us with the first picture of how the two components of the BAM complex interact with each other.

7.2. Materials and Methods

7.2.1. Cloning and Protein Overexpression

Full length BamC and BamD constructs, and the truncation mutants $BamC_{NC}$ (containing both the N- and the C-terminal domains but not the N-terminal unstructured region), $BamC_N$ (only the N-terminal domain), $BamC_C$ (only the C-terminal domain) and $BamC_{UN}$ (the unstructured N-terminal region followed by the N-terminal domain) were cloned and overexpressed as described previously in Chapter 2 of this thesis.

7.2.2. Purification of BamCD Subcomplex

Cells overexpressing BamC and BamD were separately harvested by centrifugation and subsequently combined prior to lysis. The combined cell pellet was lysed and purified by nickel affinity and size-exclusion chromatography using the same procedure as described in Chapter 2 and 3.

7.2.3. Protein-Protein Interaction Studies

To test whether or not different truncation mutant forms of BamC can form a BamCD heterodimer, each truncated mutant was co-lysed with BamD and purified by nickel affinity and size-exclusion chromatography as described in Chapter 2 and 3. The oligomeric state of the purified sample was further confirmed by dynamic multiangle light scattering analysis as described in Chapter 3.

7.2.4. Crystallization and Data Collection

BamCD crystals were grown by the sitting drop vapor diffusion method. A final v/v concentration of 0.03% n-Dodecyl-ß-maltoside (DDM) was added to the protein sample prior to setting up crystallization plates. The crystallization drops were prepared by mixing 1µL of protein (30 mg/mL) suspended in buffer A with 1µL of reservoir solution, and then equilibrating the drop against 1 mL of reservoir solution. The BamCD construct yielded crystals in the space group I 1 2 1 with unit cell dimensions of 73.8, 133.4, 145.0 Å. The optimal crystallization reservoir condition was 0.2 M K_2HPO_4 and 20% PEG3350. Crystallization was performed at room temperature (~22 °C). The cryo-

solution condition contained 0.2 M K_2HPO_4, 20% PEG3350, and 30% glycerol. Crystals were washed in the cryo-solution before being flash-cooled in liquid nitrogen.

Diffraction data were collected on the BamCD crystals at beamline X25 at the National Synchrotron Light Source (NSLS), using an ADSC Q315 CCD x-ray detector. The crystal-to-detector distance was 375 mm. A total of 360 images were collected with 1° oscillations, and each image was exposed for 1s.The diffraction data was processed with the program iMosflm (Battye et al., 2011). See Table 6-1 for data collection statistics.

7.2.5. Structure Determination and Refinement

Although the full length BamCD complex was used for crystallization, BamC was cleaved to a smaller fragment corresponding to $BamC_{UN}$ (D28-A217) in the crystallization drop during incubation. Phases were obtained by molecular replacement using the program PHASER 2.1 (McCoy et al., 2007). Previously solved *E. coli* $BamC_N$ (PDB: 2YH6) and BamD (PDB: 2YHC) structures were used as search models (Albrecht and Zeth, 2011). The The N-terminal unstructured region of BamC was manually fit into the electron density map using the program COOT (Emsley and Cowtan, 2004). The structure was refined using restrained refinement in REFMAC5 (Murshudov et al., 1997), and further manual adjustments to the atomic coordinates were performed with the program COOT. The final model was obtained by running restrained refinement in REFMAC5 with TLS restraints obtained from the TLS motion determination server (Painter and Merritt, 2006). The refinement statistics are shown in Table 6-1. The atomic coordinates and structure factors have been deposited in the RCSB Protein Data Bank (PDB: 3TGO).

Table 7-1 Crystallographic Statistics for BamCD Complex

The data collection statistics in parentheses are the values for the highest resolution shell. See Appendix D for definitions of statistical values reported.

Crystal Parameters	
Space group	I 1 2 1
a,b,c (Å)	73.8, 133.4, 145.0
α, β, γ (º)	90.0, 100.2, 90.0
Data Collection Statistics	
Wavelength (Å)	0.97950
Resolution (Å)	97.5 – 2.9 (3.0 – 2.9)
Total Reflections	218354 (32146)
Unique reflections	30529 (4448)
R_{merge}	0.112 (0.393)
Mean (*I*)/σ (*I*)	14.6 (5.4)
Completeness (%)	99.7 (99.8)
Redundancy	7.2 (7.2)
Refinement Statistics	
Protein molecules (chains) in A.U.	4
Residues	217, 218, 182, 186 (Chain A, B, C and D)
Water molecules	124
Total number of atoms	6505
R_{cryst} / R_{free} (%)	19 / 25
Average B-factor (Å^2) (all atoms)	37.4
Rmsd on angles (º)	1.050
Rmsd on bonds (Å)	0.012

7.3. Results

7.3.1. Formation of the BamCD Heterodimer

We have produced soluble constructs of *E. coli* BamC (Ser26-Lys344) and BamD (Ser21-Thr245), both of which encompass the entire wild type sequences immediately following the cleavable N-terminal signal sequence and the conserved

lipidation residue cysteine (Cys25 and Cys20 in BamC and BamD, respectively). As described in Chapter 3, size exclusion chromatography and multiangle light scattering analysis suggest that both BamC and BamD exist in a monomeric state in solution when purified separately (

Figure 3-2). However, when cells overexpressing each protein were combined prior to lysis, a large population of BamC and BamD was observed to co-purify as a BamCD heterodimeric complex. Both proteins co-eluted from a size exclusion column (Figure 7-1C and D), and the molar mass of the complex verified by multiangle light scattering analysis was 60.1 $\pm$ 1.8 kDa (Figure 7-1E). The calculated molar mass of the purified BamCD complex is approximately equal to the sum of the molecular weights (64 kDa) of BamC (36 kDa) and BamD (28 kDa) constructs used in this study.

To identify BamC-BamD interaction interface, we made a series of domain truncations in BamC and tested whether the mutants can still interact with BamD to form the BamCD complex. The truncated forms of BamC created for this study are as follows: 1) $BamC_N$, the N-terminal domain (Phe99-Ala217), 2) $BamC_C$, the C-terminal domain (Ala220-Lys344), 3) $BamC_{NC}$, both the N- and the C-terminal domain (Thr97-Lys344), and 4) $BamC_{UN}$, the N-terminal unstructured region followed by the N-terminal domain (Ser26-Ala217). The results from the size exclusion chromatography analysis show that any form of BamC that is missing the N-terminal unstructured region (i.e. $BamC_N$, $BamC_C$, and $BamC_{NC}$) is unable to form the BamCD complex (Figure 7-1B & Appendix G). $BamC_{UN}$, on the other hand, was observed to co-purify with BamD throughout the entire purification process despite missing the C-terminal domain (Figure 7-1C and D). Multiangle light scattering analysis later confirmed that $BamC_{UN}$ and BamD form a heterodimeric complex with a calculated molecular weight of 55.6 $\pm$ 1.7 kDa (Figure 7-1F), which is comparable to the sum (50 kDa) of the molecular weights of $BamC_{UN}$ (22 kDa) and BamD (28 kDa). Together, these results suggest that the N-terminal unstructured region of BamC plays an essential role in the formation of BamCD complex.

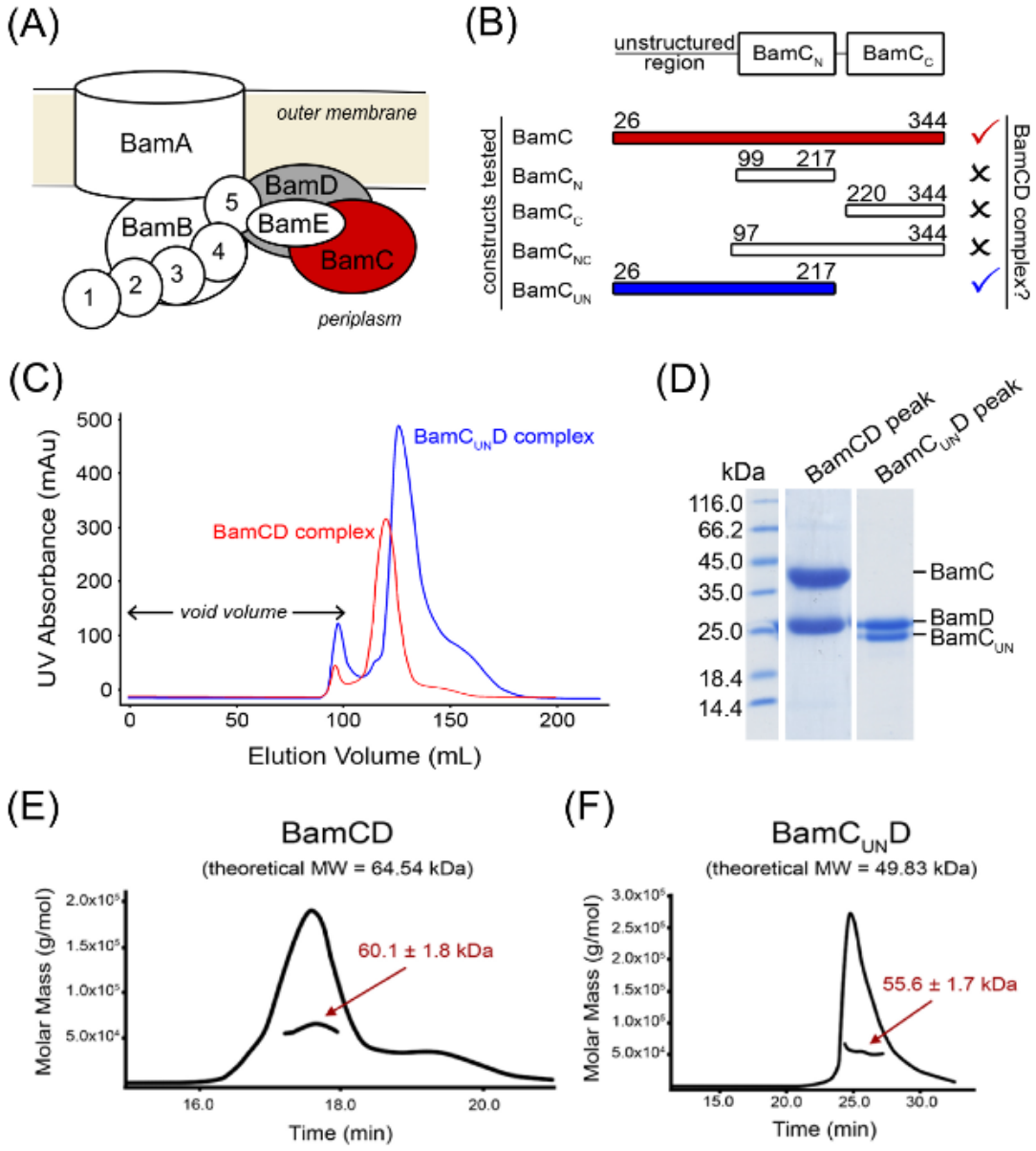


Figure 7-1 Formation of the BamCD Complex

(A) A schematic diagram of the *E. coli* BAM complex. Two of the lipoprotein components BamC and BamD that are the focus of this study are colored in red and grey, respectively. (B) Summary of BamC truncations made and their ability to interact with BamD to form the BamCD complex. (C) Overlayed size-exclusion chromatograms of co-purified BamCD and $BamC_{UN}D$ complexes are shown. (D) SDS-PAGE confirms co-elution of BamC (or $BamC_{UN}$) with BamD from the major chromatogram peaks shown in (C). The molecular masses of the BamCD complex (E) and $BamC_{U}ND$ complex (F) were verified by multiangle dynamic light scattering analysis.

7.3.2. Structure of the BamCD Subcomplex

With the individual structures of BamC and BamD monomers having been previously solved, the BamCD complex was also studied from a structural perspective. BamC forms two globular domains (N- and the C-terminal domains) both with the 'helix-grip' fold that are connected by a flexible linker. The BamD structure, on the other hand, consists of ten α-helices which form five tetratricopeptide repeats (TPRs) (Figure 7-2A) (13, 25). In this study, BamD in a heterodimeric complex with $BamC_{UN}$ has been crystallized, and its structure was solved and refined to 2.9 Å resolution (Table 7-1; Figure H1C and D). There were two copies of the complex in the asymmetric unit, and BamC and BamD superimposes onto their counterpart in the other copy of the complex with backbone r.m.s.d. values of 0.64 Å and 0.31 Å, respectively.

The most outstanding structural feature of the BamCD complex is the 73-residue-long unstructured N-terminus of BamC that has not been observed in previously reported structures; it folds into an elongated lasso-like U-shape, and makes an extensive interaction with BamD by fitting into a trail of crevices that runs along the longitudinal axis of BamD (Table 7-1B). The globular N-terminal domain of BamC that follows this unstructured region lies adjacent to the N-terminal half of the BamD molecule. The N-termini of both lipoproteins that link the proteins to the outer membrane lipid bilayer *in vivo* are found in close proximity of each other. Although a previous study has predicted that BamC interacts with the C-terminal end of BamD (Met227-Thr245) (6), our structure shows that the N-terminal region of BamD is also important site of interaction, as BamC binds along the entire length of the BamD molecule.

To determine whether BamC and BamD molecules underwent any conformational changes upon binding, the structures of both proteins were individually analyzed and compared to the previously solved crystal and NMR structures for the monomeric form of each protein. The N-terminal domain of BamC doesn't seem to have undergone any significant conformational change, as its structure superimposes very closely with the previously reported crystal structure (r.m.s.d. = 0.54 Å) and the NMR structure (r.m.s.d. = 1.24 Å). The structure of BamD in the BamCD complex, on the other hand, shows a significantly different conformation when compared to the monomeric structure (r.m.s.d. = 2.3 Å) (Figure 7-2C). In the BamCD complex structure, the positions

of α-helices in the TPR motifs 3 (α6), 4 (α7 and α8) and 5 (α9 and α10) of BamD are shifted to better accommodate the binding of BamC. The four C-terminal α-helices of the TPR motifs 4 and 5, in particular, show the greatest change in conformation.

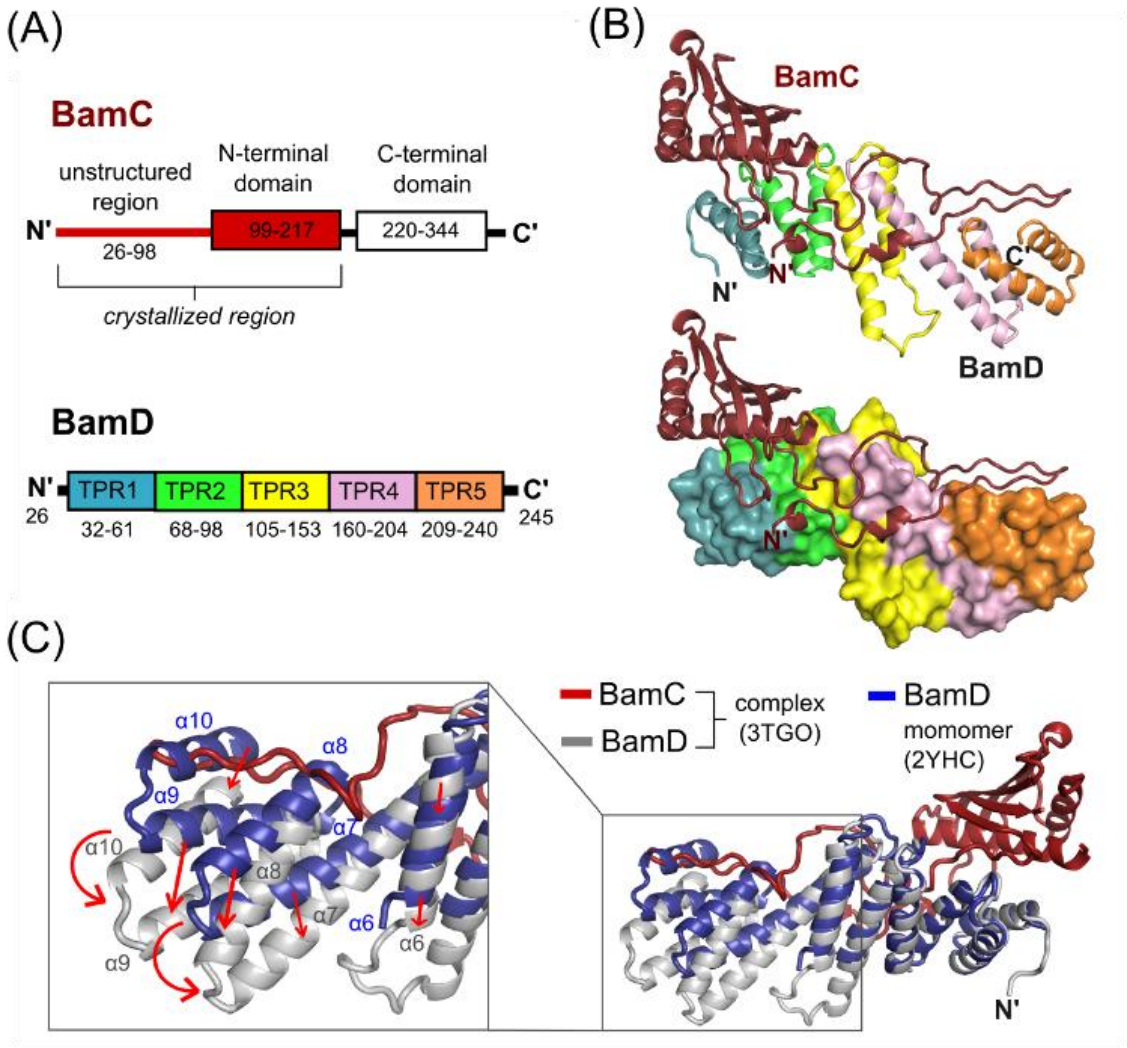


Figure 7-2 Structure of the BamCD Complex and Conformational Changes in BamD upon Binding

(A) Domain organization of BamC and BamD. Each tetratricopeptide repeat (TPR) motif of BamD is shown in a different color. (B) Structure of the BamCD complex. Ribbon (top) and surface (bottom) diagrams of BamD are shown with the TPR motifs colored as in (A). (C) BamCD complex (BamC in red and BamD in grey) superimposed onto monomeric BamD (blue). Boxed inset shows a close-up view of the BamD region that undergoes the greatest conformation change upon binding to BamC, with arrows indicating direction of movement.

7.3.3. BamC and BamD Interaction Interface

The interaction between BamC and BamD is predominantly created by the direct contact between the N-terminal unstructured region of BamC and all five TPR motifs of BamD (Figure 7-2B). In both copies of the BamCD complex present in the asymmetric unit, the unstructured region of BamC folds into a stretched U-shape and binds longitudinally on BamD (Figure 7-3A). This interaction between the two proteins is mediated by numerous hydrogen bonds, salt bridges and van der Waals forces, and has an average interface area of 2249.4 $Å^2$ (Table 7-2). Many conserved residues are found concentrated at the interaction interfaces on both proteins (Figure 7-3B and C), suggesting that the interaction between these two proteins has important biological and functional implications. For BamC, multiple sequence alignment shows that its unstructured N-terminus is the most well conserved region of the protein, which is not surprising considering its essential role in stabilizing the BamCD complex structure. For BamD, about half of the conserved residues are found at the BamCD interaction interface, but the other half are found clustered on the opposite side of the protein and are solvent-exposed in this heterodimeric complex (Figure 7-3C).

7.3.4. Structural Homologues

A pocket present in the N-terminal region of BamD (formed by TPR motifs 1 and 2) has been predicted to bind and recognize the C-terminal targeting sequence of unfolded OMP substrates (13, 25). In fact, this binding pocket of BamD closely resembles that of PEX5, a peroxisomal targeting signal receptor (Figure 7-4A). When the BamD structure is overlaid on the C-terminal domain structure of PEX5 (PDB: 3PVP), a peroxisomal targeting signal receptor, both structures show similar fold with an r.m.s.d. value of 4.2 Å. The binding pockets of PEX5 and its putative counterpart in BamD show especially close resemblance (r.m.s.d. = 1.7 Å). Interestingly, in the BamCD complex structure, this binding pocket of BamD is occupied by part of the unstructured region of BamC (Figure 7-4B). The segment of the unstructured region of BamC that fits into the pocket has no sequence similarity to the C-terminal targeting sequences of OMP substrates, or to PTS1 which is a C-terminal peroxisomal targeting sequence recognized by PEX5.

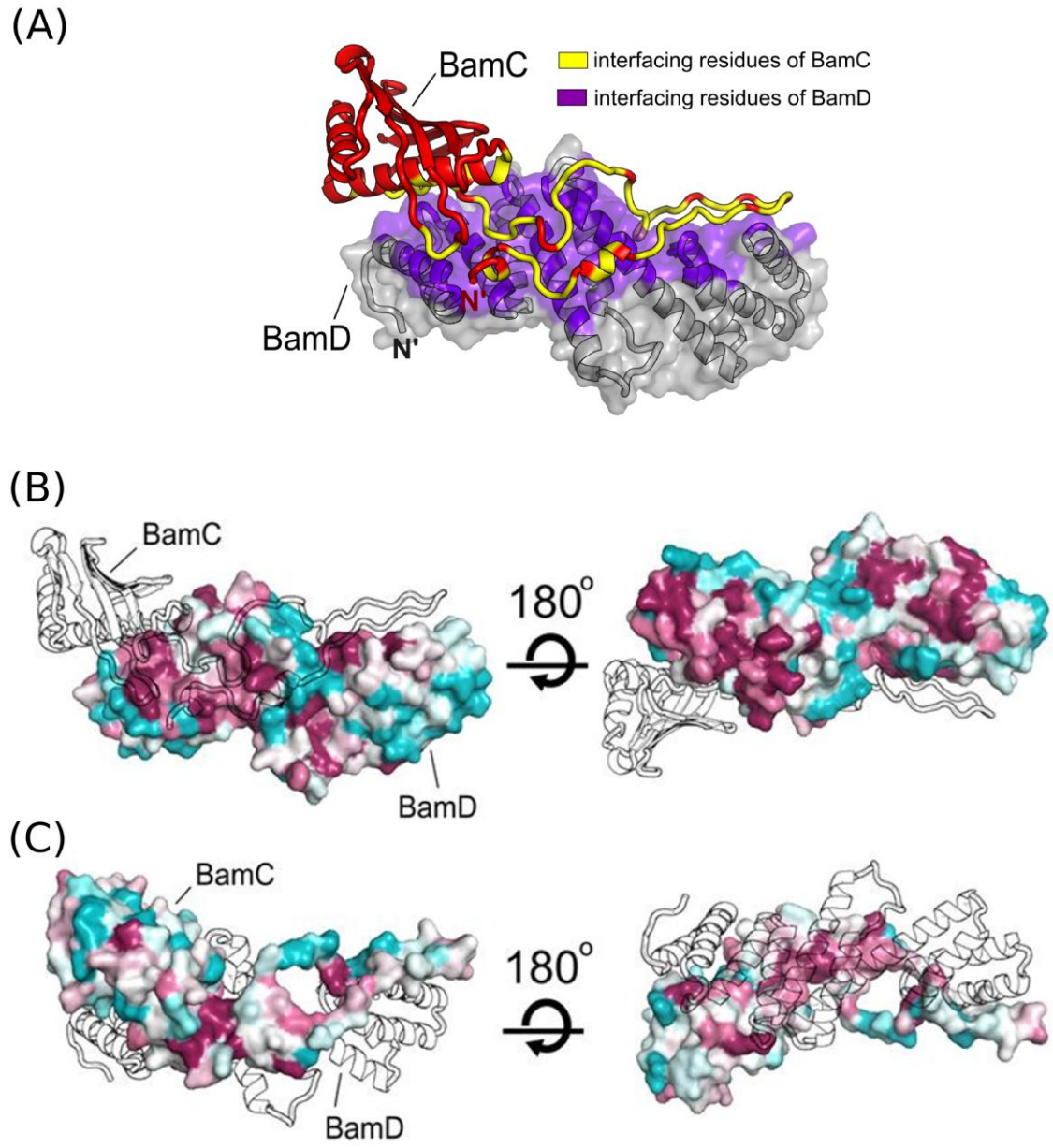


Figure 7-3 The BamC-BamD Interface

(A) Ribbon diagram of BamC (red) with ribbon and semitransparent surface diagram of BamD (grey). The interfacing residues colored yellow (BamC) and purple (BamD). (B) Conserved residues mapped onto BamD with absolutely conserved residues in maroon and highly variable residues in cyan. An outline of BamC is drawn to show its position relative to BamD. (C) Conserved residues mapped onto BamC and colored as in (B). Outline of BamD is also shown.

Table 7-2 List of Interfacing Residues Forming Hydrogen Bonds

BamD (Chain A)	BamC (Chain C)	Distance (Å)
GLN 44[O]	LYS 32[NZ]	2.50
TYR 80[OH]	SER 36[N]	3.16
ASP 162[OD1]	TYR 41[OH]	2.46
ASP 207[O]	ALA 48[N]	3.16
ASN 241[OD1]	HIS 51[N]	2.79
THR 161[OG1]	GLY 74[N]	2.69
ASP 162[OD1]	GLY 77[N]	3.13
ASN 104[OD1]	GLN 87[NE2]	3.20
GLN 44[OE1]	LEU 151[N]	3.74
LYS 81[NZ]	TYR 31[O]	2.84
TYR 77[OH]	LYS 32[O]	2.56
LYS 81[NZ]	GLN 34[O]	2.77
ARG 212[NH2]	ALA 48[O]	2.94
ARG 212[NH2]	GLU 49[O]	3.45
ASN 241[ND2]	HIS 51[O]	2.87
ARG 152[NH2]	GLY 63[O]	2.95
ARG 152[NH2]	ALA 66[O]	3.26
TYR 107[OH]	ILE 82[O]	2.79
GLN 70[NE2]	GLN 87[O]	3.57
GLN 70[NE2]	PRO 88[O]	3.26
ASN 33[ND2]	ASP 210[OD2]	3.80

* Conserved residues are colored pink

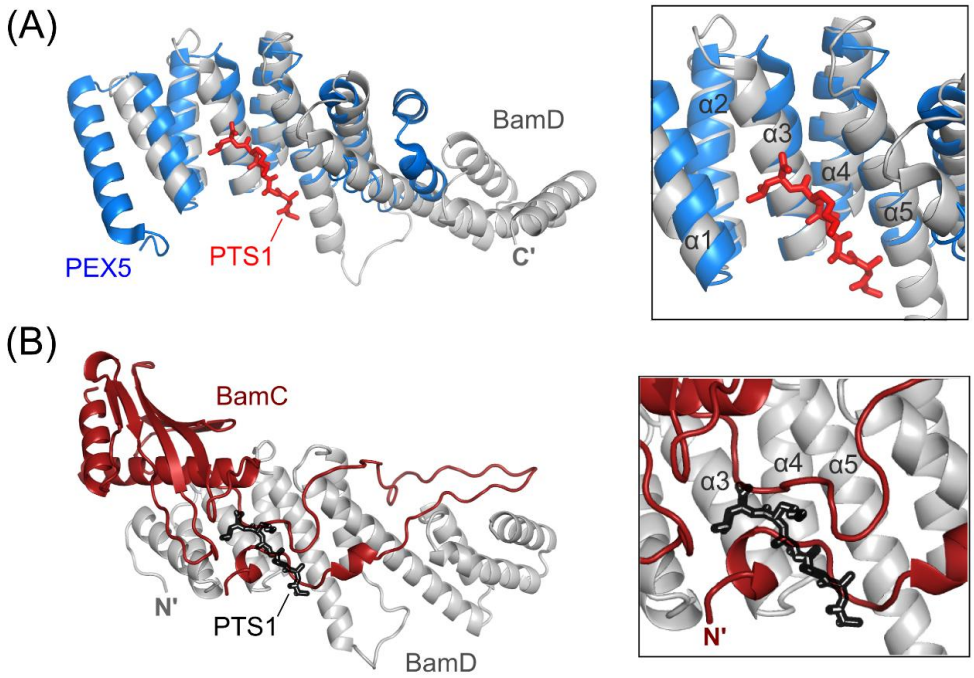


Figure 7-4 Proposed C-terminal Targeting Sequence Binding Pocket of BamD

(A) Superimposed ribbon diagrams of BamD (grey) and the C-terminal domain of PEX5 (blue), a peroxisomal targeting sequence receptor. The PTS1 peptide, the peroxisomal targeting sequence that is recognized and bound by PEX5, is shown in red. (B) In the BamCD complex, part of the unstructured region of BamC (red) blocks the proposed C-terminal targeting sequence binding pocket of BamD (grey). PTS1 (black outline) is shown for reference.

7.4. Discussion

The BAM complex plays an essential role in outer membrane biogenesis by facilitating the assembly of outer membrane proteins. Functional and structural information of the five proteins (BamA, B, C, D and E) that make up the *E. coli* BAM complex has started to emerge in recent years; however, how the individual BAM proteins come together to form the BAM complex largely remains to be elucidated.

In this chapter, we have presented the structure of BamD in complex with $BamC_{UN,}$ the N-terminal domain of BamC preceded by 73-residues long unstructured region. The truncation mutagenesis and size exclusion chromatography results showed that the unstructured N-terminus of BamC is essential for BamCD complex formation, and indeed the crystal structure of the complex clarified why that is the case. The very

N-terminus of BamC, which has been shown to be unstructured but hasn't been seen in the previously reported structures, folds into an extended U-shape that latches onto the BamD molecule. Although a previous study has predicted that BamC interacts with the C-terminal end of BamD (Met227-Thr245) (Malinverni et al., 2006), our structure shows that the N-terminal region of BamD is also important site of interaction, as BamC binds along the entire length of the BamD molecule. Many conserved residues of both BamC and BamD are found at the interaction interface, suggesting that the interaction between these two proteins has important biological and functional implications.

While our structural information, with complementing protein-protein interaction analysis data, has shown how BamC and BamD interact with each other, the insights gained in this study raise many questions regarding the function and the structure of the BAM complex. For instance, a pocket present in the N-terminal region of BamD has been predicted to bind and recognize the C-terminal targeting sequence of unfolded OMP substrates (Sandoval et al., 2011). In our BamCD complex structure, however, the same pocket is occluded by the unstructured N-terminus of BamC. If the function of BamD is indeed recognition of substrate OMPs via their C-terminal targeting sequences, then perhaps one of the roles of BamC (more specifically, the unstructured N-terminus of BamC) is that of a regulatory one where it may block or expose the targeting sequence binding site of BamD depending on the need.

Many other aspects of the BamCD complex structure and function remain to be elucidated. For example, how does the BamCD complex interact with the POTRA 5 domain of BamA? Mapping conserved residues onto the BamD structure reveals that while some conserved residues are found at the BamC binding site, there are other remaining conserved residues found solvent exposed on the other side of the BamD molecule. It is possible that this conserved region of BamD may serve as a binding surface for POTRA 5 of BamA. Another question to be addressed is where the C-terminal domain of BamC is positioned relative to the rest of the BamC and BamD molecules. As mentioned in the previous chapter, the C-terminal domain of BamC has a well conserved pocket of potential protein-protein interaction site. This could be an additional site for interaction with BamD. Alternatively, the pocket may be for association with BamE, as BamC has been shown to interact with BamD and BamE forms a stable heterotrimer (as described in Chapter 3).

8. Discussion

Note regarding contributions:

Portions of this chapter were published as review articles (sections 8.2, 8.3 and 8.4). The authors and the full references for these articles are listed below.

Kim, K.H., Aulakh, S. & Paetzel, M. (2012) The bacterial outer membrane β-barrel assembly machinery*. Protein Science* **21**, 751-768.

Kim, K.H., Aulakh, S. & Paetzel, M. (2012) Outer Membrane Protein biosynthesis: transport and incorporation of proteins (in)to the OM bilayer. *Bacterial Membranes: Structural and Molecular Biology.* UK: Horizon Scientific Press. [invited book chapter; in press]

All the figures included in this chapter, otherwise noted, were adopted from these articles.

8.1. Summary

The main objective of this thesis was to determine the structures of the individual subunit proteins that make up the *E. coli* BAM complex, a multi-protein complex that plays an essential role in OMP folding and assembly. When this thesis project was first initiated in 2006, there were no structural data available for any of the five protein components (BamA, B, C, D and E) of the BAM complex. However, during the time that this thesis was undertaken (Summer 2006 – Spring 2012), the structures of BamB, BamC, BamE and BamCD complex were successfully solved and analyzed. Analysis of these solved structures (e.g. mapping conservation onto the structure, comparison to the structural homologues, backbone dynamics study, etc.) helped to identify the regions of each protein that may be involved in protein-protein interaction or may be functionally important.

Our BamB structure revealed the eight-bladed β-propeller fold and the location of the important residues that have previously been determined by mutagenesis to be important for interaction with BamA. Based on the structure, we have suggested that the outer rim of the β-propeller of BamB may participate in protein-protein interaction via β-augmentation. The structural analysis of BamE by NMR showed that BamE has a globular core structure, with flexible N and C termini, as well as a flexible loop that protrudes out of the structured core. We have hypothesized that these flexible regions of BamE may serve roles in protein-protein interaction. The crystal structure of the C-terminal domain of BamC showed that it exhibits a 'helix-grip' fold that could serve as a protein interaction surface, especially for binding an α-helical element. Finally, the crystal structure of the BamCD complex revealed that the conserved unstructured N-terminal region of BamC binds and occupies the proposed substrate binding pocket of BamD, which has led us to hypothesize that the function of BamC is to regulate the activity of BamD.

In addition to the structural investigation, protein-protein interaction studies were carried out to analyze the stoichiometry of the interactions between the periplasmic components of the BAM complex (i.e. the POTRA domains of BamA and the lipoproteins BamB, C, D and E). We have shown, mainly by size-exclusion chromatography, that a stable $BamA_{POTRA}$-BamB, $BamA_{POTRA}$-BamD/E, BamC/D/E, BamC/D and BamD/E

complexes (all containing single copies of each protein) can be purified *in vitro*. Taken together, the structural and interaction data obtained during the course of this thesis project contribute toward better understanding of the BAM complex structure and provides a platform for future research driven by structure-based hypotheses.

8.2. Proposed Mechanisms of the BAM Complex

As the structural information about the BAM complex is emerging, so is the knowledge about the functional aspects of the complex. Genetics, mutagenesis and *in vitro* analyses, together with the structural data, allow us speculate on how the BAM complex carries out its function. Although the exact mechanism is not yet clearly understood, the process of OMP assembly by the BAM complex could be broken down into three major steps: 1) substrate recognition, 2) substrate binding, and 3) catalysis of the folding and insertion of the substrates into the OM. What is known about each of the three steps based on the current literature and the results obtained during this thesis project are summarized below.

8.2.1. Substrate Specificity

Despite their ability to insert into different lipid bilayers of varying composition *in vitro*, OMPs only assemble at the OM and not at the IM (Patel et al., 2009). This implies that the primary amino acid sequence of OMPs must contain information that targets them to the OM and to the BAM complex, but where does this information reside within the OMP sequences?

In 1991, Struyve and her colleagues first noticed that deletion of the C-terminal segment of PhoE (phosphate porin) prevents the assembly of the protein into the OM (Struyve et al., 1991). Multiple sequence alignment of bacterial OMPs subsequently revealed that the C-termini of the vast majority of OMPs consists of a phenylanine (or tryptophan) at the C-terminal position, and hydrophobic residues at positions 3 (mostly tyrosine), 5, 7 and 9 from the C-terminus (Figure 8-1) (Struyve et al., 1991). The C-terminal phenylalanine is strongly conserved, and its deletion or substitution decreases the level of PhoE detected in the OM significantly, suggesting that the conserved C-

terminal residue is essential for correct assembly of PhoE. A few years later, the same group performed a follow-up study using immunocytochemical labeling, which revealed accumulation of C-terminal phenylalanine mutant forms of PhoE in the periplasm (de Cock et al., 1997). Taken together, these results suggest an important role of the conserved C-terminal residue in OM targeting.

Following the discovery of the BAM complex and its role in OMP assembly, a hypothesis that the BAM complex might recognize its OMP substrates via their C-terminal signature sequence emerged. As BamA is essential and its N-terminal POTRA domain (POTRA 1) has been shown to bind the periplasmic chaperone SurA (Bennion et al., 2010), BamA was initially considered by many to be the most likely candidate for performing the substrate recognition function. The idea that the C-terminal signature sequence of OMPs is recognized by BamA was first investigated by Robert *et al.* (2006), who showed that various OMPs, as well as a peptide mimicking the C-terminal signature sequence of PhoE alone, can bind and modulate BamA channel activity. In the same study, Robert *et al.* (2006) observed that PorA from *N. meningitidis* is unable to induce the channel opening of *E. coli* BamA, despite having the conserved C-terminal phenylalanine and hydrophobic residues at positions 3, 5, 7 and 9 from the C-terminus. Subsequent comparison of the C-terminal sequences of *E. coli* and *N. meningitidis* OMPs show that *N. meningitidis* OMPs differ from the *E. coli* OMPs in that a positively charged residue (predominantly lysine or arginine) is found at position 2 (Robert et al., 2006). In *E. coli* OMPs, the amino acid found at position 2 is usually glutamine. Interestingly, PorA can be assembled correctly into the *E. coli* OM when the lysine residue at position 2 of PorA is mutated to glutamine (Robert et al., 2006). Based on this result, Robert *et al.* (2006) suggested that BamA recognizes its OMP substrates by a species-specific C-terminal motif.

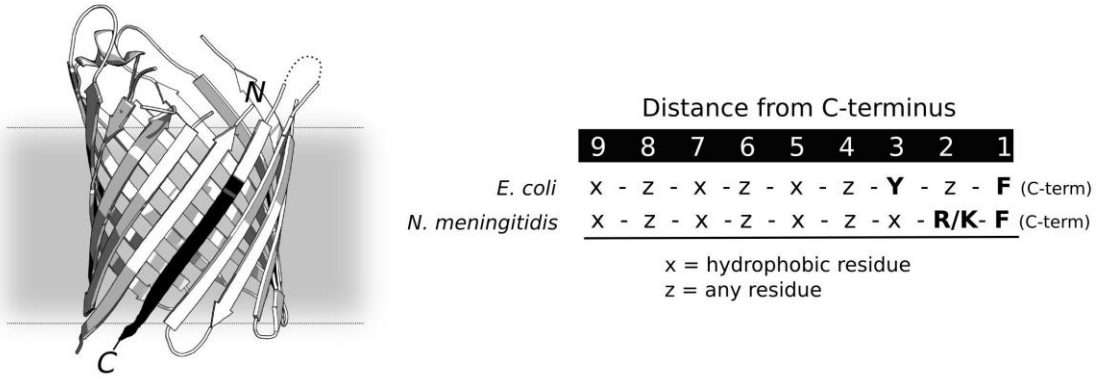


Figure 8-1 Substrate Recognition by the BAM Complex

An OM targeting signal resides at the C-terminus of an OMP sequence. The location of the targeting signal within a folded OmpG β-barrel (PDB: 2F1C) is shown (black). The targeting signal sequences of OMPs seem to be species specific. For comparison, trends in the C-terminal sequences of *E. coli* and *N. meningitidis* OMPs are shown (right).

While it has been established that denatured OMP substrates bind directly and modulate channel conductance of BamA (Robert et al., 2006; Voulhoux et al., 2003), recently emerging experimental data suggests that BamD may also play an important role in the initial substrate selection. When Sandoval *et al*. (2011) solved the first crystal structure of BamD from *R. marinus*, they showed through structure comparison that the N-terminal half (TPR1-3) of BamD forms a pocket that superimposes very closely with the binding pockets of the other proteins where protein substrates in extended conformations bind (Albrecht and Zeth, 2011; Kim et al., 2011a; Sandoval et al., 2011). These proteins include PEX5 (peroxisomal targeting signal receptor) and HOP (Hsp-organizing protein) (Sandoval et al., 2011). Both proteins recognize the C-termini of their substrates, and this has led to the speculation that BamD might serve as OMP targeting signal receptor for the BAM complex by recognizing the C-terminal signature sequences of OMPs (Sandoval et al., 2011). Shortly after the crystal structure of *R. marinus* BamD was published, Albrecht and Zeth reported the crystal structure of *E. coli* BamD, along with the finding that a truncated form of BamD consisting only of TPR1-3 that harbors the proposed binding pocket has been shown to crosslink with synthetic peptides harboring the OMP C-terminal targeting sequence (Albrecht and Zeth, 2011). Interestingly, however, our BamCD subcomplex crystal structure revealed that the proposed binding pocket of BamD is bound to the unstructured N-terminal region of

BamC (Kim et al., 2011a). It is clear from the BamCD structure that the proposed binding pocket of BamD will not be able to bind the C-terminal signature sequence of OMPs, as it is completely occluded by BamC. Nevertheless, the structural similarity between BamD and those of other targeting signal recognition proteins is conspicuous, which gives rise to the question of whether the substrate binding activity of BamD is regulated by BamC interaction.

So which protein, BamA or BamD, initially recognizes the OMP substrates via the C-terminal targeting signal sequence? While the C-terminal fragment of a canonical OMP substrate PhoE has been shown to bind BamA and modulate its channel activity (Robert et al., 2006), the crosslinking data and structural resemblance of BamD to other targeting signal receptor proteins favor the hypothesis that BamD functions in the initial substrate recognition and selection (Albrecht and Zeth, 2011). Future studies of substrate binding affinity to BamA and BamD, as well as further biochemical characterization of the interactions between the OM targeting signal and BamA/BamD, will shed some light on which protein acts as the OM targeting signal receptor.

8.2.2. Substrate Binding

Prior to incorporation into the OM, OMPs must be kept in unfolded forms to prevent misfolding and aggregation. Chaperones such as SurA and Skp play important roles in this process by binding OMPs while they traverse through the dense periplasmic space to reach the OM (Sklar et al., 2007b; Volokhina et al., 2011). Once at the OM, it is currently not clear how the OMPs are passed on to the BAM complex for subsequent assembly process. Does the substrate get progressively transferred from the chaperone to the BAM complex as the folding and the membrane insertion take place? Or does the chaperone unload the substrate all at once, implying that the BAM complex needs to keep the substrate in a non-aggregated form before the assembly process begins? In either case, the proteins of the BAM complex must be able to bind OMP substrates in unfolded forms. The ability of BamA to bind OMPs has already been established, but the location of the binding sites and the mode of interaction are still subjects of further study. In addition to BamA, BamB has also been suggested to be involved in substrate binding (Heuck et al., 2011; Voulhoux et al., 2003).

As described earlier, the POTRA domains of BamA and BamB participate in protein-protein interaction via β-augmentation (Gatzeva-Topalova et al., 2008; Heuck et al., 2011; Kim et al., 2007; Knowles et al., 2008). Both the POTRA and the BamB structures form β-sheets with the edges of the sheets exposed and available for hydrogen bonds (Hagan et al., 2011). In the case of the POTRA domains, Knowles et al. (2008) showed by NMR titration experiment that addition of various nascent β-strand peptides derived from PhoE induced chemical shift changes in residues found on the outer edges of the β-sheets in POTRA 1 and POTRA 2. Furthermore, POTRA 3 was shown to be involved in crystal packing by β-augmentation in two separate crystal structures despite different crystallization conditions (Gatzeva-Topalova et al., 2008; Kim et al., 2007). Since all POTRA domains share the same basic structure, POTRA 4 and POTRA 5 may also be able to participate in protein-protein interaction via β-augmentation. Similarly, β-augmentation has been observed in BamB crystals as well, in which the outermost β-strand of one of the blades in the BamB structure is seen bound to a strand of a neighboring BamB molecule (Heuck et al., 2011). As the BamB structure consists of eight blades that are very close to each other in topology, BamB could provide eight potential substrate binding sites.

If both the POTRA domains of BamA and BamB can bind substrates, how are their roles distinguished? Although OMP assembly is significantly reduced in its absence, BamB is not an essential component of the BAM complex. On the other hand, at least one of the POTRA domains, POTRA 5 (the closest to the membrane), is required for proper function. Considering the essential nature and close proximity to the β-barrel domain of BamA (where the actual catalysis of OMP folding and membrane insertion is thought to take place), the POTRA domains likely serve as the major substrate binding sites and as the passage that leads substrates towards the β-barrel domain. The role of BamB may then be more of a supportive one, since its absence significantly decreases the efficiency but does not halt the proper functioning of the BAM complex. As the OMPs that are most affected by BamB deletion are relatively large (16-24 stranded β-barrels), it has been suggested that BamB could aid BamA function by increasing the substrate binding capacity (Heuck et al., 2011). It is also possible that BamB functions as a reservoir of substrates when the BAM complex function is in high

demand, binding substrates and preventing them from aggregating until they can be delivered to BamA for subsequent assembly process.

8.2.3. Protein Folding and Membrane Insertion

The least understood aspect of the BAM complex function is the process by which OMPS are folded and inserted into the OM to adopt the final β-barrel structure. Although controversial, it has been suggested that OMPs exist in a partially folded state in the periplasm. Aside from the periplasmic chaperones, the conformational change of the POTRA domains from the extended to the bent state (when bound to an unfolded OMP) has been speculated to facilitate formation of β-hairpins in the substrate prior to membrane insertion (Gatzeva-Topalova et al., 2008). *In vitro* β-barrel folding studies suggest the presence of membrane bound folding intermediates, and that completion of folding and membrane insertion of OMPs take place in a concerted manner (Burgess et al., 2008; Kleinschmidt, 2003). As OMPs can spontaneously fold and insert into membranes in vitro and there is no ATP source in the periplasm, the BAM complex function is thought to be mainly associated with increasing the kinetics of the natural folding process of OMPs. Several different models of OMP assembly by the BAM complex have been proposed to date (Figure 8-2), and they are described below.

The earliest model of the BAM complex proposed that a substrate folds within the pore formed by the β-barrel domain of BamA, and would then be laterally released into the OM (Figure 8-2C). If this model is correct, the channel formed by BamA must be large enough to accommodate a folded substrate. Currently available data suggests that a channel formed by a single BamA molecule is not large enough to contain a fully folded OMP. Whereas the pore diameter of BamA estimated from electrophysiological data is 25 Å , the structure of a BamA homologue, FhaC, has a measured diameter of 16 Å if the components that partially block the pore of FhaC are relocated out of the channel (Clantin et al., 2007). However, studies have shown that BamA in monomeric form can properly assemble OmpT (approximately 25 Å in diameter) (Hagan et al., 2011). BamA has been shown to be responsible for the folding of large OMPs such as FimD, which is more than twice as large as OmpT (Palomino et al., 2011). The channel formed by BamA may be large enough if BamA oligomerizes to form a single β-barrel; but, as mentioned above, it seems that monomeric BamA is fully capable of carrying out its

function. Furthermore, lateral release of a folded substrate implies breaking several hydrogen bonds of the BamA β-barrel domain, which would be energetically costly.

The difficulty of biochemically detecting and characterizing folding intermediates, as well as lack of BamA-substrate structure, presents a large challenge to the understanding the folding and insertion mechanism of the BAM complex. The only structure available of a protein that is closely related to BamA is that of a distant homologue, FhaC. The function of FhaC is different from BamA in that it is responsible for translocating one specific substrate across the OM, rather than folding and inserting the substrate into the OM bilayer (Clantin et al., 2007; Delattre et al., 2010). However, its structure still provides valuable insights into the BamA function, as BamA is predicted to have the similar overall structure as FhaC (except that BamA has five and FhaC has only two POTRA domains). The channel formed by the β-barrel of FhaC is approximately 3 Å in diameter, which is too small for accommodating even an unfolded polypeptide (Clantin et al., 2007). However, studies suggest binding of the substrate (FHA) causes a conformational change in FhaC, which could subsequently increase the diameter of the pore from 3 Å to 16 Å (Clantin et al., 2007; Delattre et al., 2010). This increase in the pore size could make enough room for the substrate to enter the channel within the barrel in an unfolded extended form (Clantin et al., 2007). As it has been shown that substrate binding increases the channel conductivity of BamA (Stegmeier and Andersen, 2006), it is possible that BamA also undergoes a similar conformational change as in FhaC. If this is the case, the β-barrel lumen of BamA could serve as a conduit for a substrate to be translocated across the OM in an unfolded form. This model implies that the substrate would emerge into the extracellular space, and that substrate folding/insertion would takes place on the outside of the cell (Figure 8-2A). However, it is difficult to imagine OMPs assembling efficiently without extra folding factors on the extracellular surface of the outer membrane. Perhaps it is possible that LPS, which is only present on the outer leaflet of the OM, promotes substrate folding; but currently there is not sufficient experimental evidence to associate LPS with OMP folding efficiency.

The two models described above assume that the β-barrel lumen of BamA plays a critical role by either providing an isolated environment for OMP folding or providing a passage for substrate to cross the OM. However, a more recently proposed model

predicts that OMP substrates use the outer wall of the BamA β-barrel as a scaffold for folding and membrane insertion (Figure 8-2B) (Hagan et al., 2011). In this model, an unfolded or partially folded substrate will start to insert between the BamA-lipid interfaces as it starts acquiring tertiary structure. Alternatively, if BamA forms a tetramer in vivo, the substrate folding could be contained in the space formed by the four BamA subunits. The limited folding space may facilitate the closing of the β-sheet into a β-barrel, and the β-barrel would be released laterally into the lipid bilayer. This model requires that BamA subunits within the proposed tetramer be able to associate and dissociate with each other to allow substrate release. This model does not require BamA to have a large channel nor does it require breaking inter-strand hydrogen bonds for substrate release.

The mechanism of folding and membrane insertion by the BAM complex remains largely speculative. Here, we propose yet another model that combines and modifies the ideas from each of the models described earlier (Figure 8-2D). As mentioned previously, the structural and functional information now tells us that the channel formed by BamA is not large enough to hold a folded OMP, and breaking inter-strand hydrogen bonds of a β-barrel to release the substrate would be very energetically costly. To attune to the new structural data and to minimize the energy cost, we suggest the model be modified as follows. Instead of the unfolded substrate folding within the β-barrel of BamA, the N- and the C-terminal β-strands of BamA (the two strands that hydrogen bond with each other to close the β-sheet into a β-barrel) could serve as folding templates for the substrate. More specifically, the hydrogen bonds between the two terminal β-strands of BamA would be interrupted and replaced by an incoming substrate, which would form new hydrogen bonds with the terminal strands of BamA. In other words, the β-barrel domain of BamA would be opened and augmented by addition of strands from the substrate to form a larger temporary chimeric β-barrel. As the substrate adopts a β-sheet structure held between the two terminal strands of BamA, it will close into a β-barrel and ‘buds off’ BamA into the lipid bilayer. Although this model requires the β-barrel of BamA to open up, the energy cost of breaking hydrogen bonds is compensated by having the substrate forming new hydrogen bonds with BamA via β-augmentation. Also, this model keeps the hydrophobic residues of a substrate always facing the membrane, and hydrophilic residues always facing away from the OM bilayer. However, as with other models, many

aspects of this model are speculative and need to be experimentally validated in the future.

Do lipoproteins BamB-E play any role in these last steps of OMP assembly? Ieva et al. (2011) used site-specific crosslinking in combination with pulse-chase labeling to show that BamB and BamD remain bound to an OMP substrate longer than BamA, and suggested that BamB and BamD may function at a later stage of assembly such as substrate release (Ieva et al., 2011). BamE, on the other hand, has been shown to bind specifically to phosphatidylglycerol (Knowles et al., 2011), which has previously been shown to enhance the insertion of OMPs into liposomes although the reason for this is unclear. Based on this observation, it has been hypothesized that the function of BamE may be to recruit phosphatidylglycerol to enhance OMP membrane insertion (Endo et al., 2011). How the functions of the lipoproteins are coordinated with that of BamA remains largely enigmatic. Biochemical and kinetics studies of how substrates interact with each of the BAM components may help identify steps in the process of OMP assembly by the BAM complex.

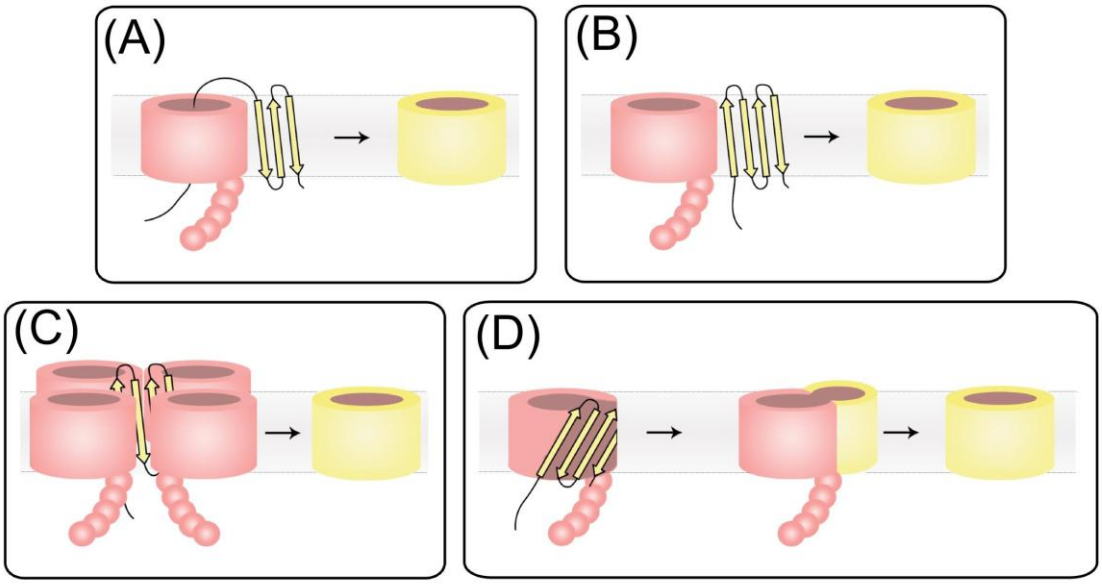


Figure 8-2 (legend on next page)

Figure 8-2 Different Models of the OMP Assembly

Four different models of how the BAM complex may facilitate the folding and insertion of OMPs are shown. BamA is shown in light pink, and the substrate protein in yellow. The lipoproteins BamB/C/D/E are not shown in these models for clarity. The outer membrane is outlined with black lines, with the extracellular space above and the periplasmic space below. (A) In the first model, the substrate protein is first translocated across the outer membrane through the channel formed by the β-barrel domain of BamA. The substrate then inserts and folds into the outer membrane lipid bilayer from outside the cell. (B) In the second model, the substrate inserts into the lipid bilayer from the periplasmic face of the outer membrane. Instead of using the channel of BamA, the insertion and folding of OMPs occur at the BamA-lipid interface. In this model, the outer wall of the β-barrel of BamA provides a scaffold for the substrate folding. (C) This model is similar to the second model, but assumes that BamA forms an oligomeric structure. The coordinated events of substrate folding and membrane insertion are contained within the space formed by the BamA tetramer. The mature, folded OMP substrate is then released laterally into the lipid bilayer of the outer membrane. (D) In the last model, the OMP substrate uses the N- and the C-terminal β-strands of BamA as folding templates. The hydrogen bonds between the two terminal β-strands of BamA are displaced by the incoming substrate that becomes part of the BamA structure as it folds into a β-sheet. The β-sheet of the substrate is then closed to form a β-barrel, and the substrate is released into the lipid bilayer.

8.3. Eukaryotic Homologues

Homologues of the BAM complex also exist in the outer membranes of mitochondria and chloroplasts (Figure 8-3). Similar to the bacterial OMPs, mitochondrial and chloroplastic OMPs are also synthesized in the cytosol prior to being targeted. However, for these eukaryotic proteins, the signal sequence directs the OMP to the organelle membrane (mitochondrion/chloroplast) rather than the plasma membrane of the cell.

8.3.1. The SAM Complex in Mitochondria

In the mitochondrial system, before being inserted into the mitochondrial outer membrane (MOM), the substrate proteins are first imported into the mitochondrion via the Translocase of Outer Mitochondrial membrane (TOM). After entering the intermembrane space (IMS), Tim chaperones transport the OMPs back to the MOM for assembly by the Sorting and Assembly Machinery (SAM) complex (Figure 8-3B). The primary component of this complex is Sam50 (the BamA homologue) which contains only one POTRA domain facing the IMS. It appears that the POTRA domain plays an important role in substrate release as this function is hindered when the domain is

absent (Stroud et al., 2011). Instead of lipoproteins, two cytosolic proteins, Sam35 and Sam37, have been identified as the main accessory proteins, with Sam35 being essential for cell survival (Milenkovic et al., 2004; Paschen et al., 2005). Current research suggests Sam35 to be involved in substrate recognition and Sam37 involved in substrate release (Chacinska et al., 2009; Paschen et al., 2005).

8.3.2. The TOC Complex in Chloroplast

For chloroplasts, protein import from the cytosol into the stroma involves passing the Translocons at the Outer and Inner envelopes of Chloroplasts (the TOC and TIC complexes) (Oreb et al., 2008). In the case of chloroplastic OMPs found in the outer envelope membrane (OEM), it was previously proposed that the OMPs travel into the stroma using the TOC/TIC complexes, and then travel back to the OEM for assembly by Toc75-V (the BamA homologue) (Figure 8-3C). This was based on the assumption that the three POTRA domains of Toc75-V face the IMS similar to Sam50 of the mitochondrial system. However, a recent study has shown the POTRA domains to exist in the opposite orientation, with the POTRA domains facing toward the cytosol (Sommer et al., 2011). This new finding suggests the possibility for OMPs to be imported directly by Toc75-V and immediately inserted into the OEM, without the use of the TOC/TIC pathway. However, with the exact pathway unknown, and essential accessory proteins yet to be identified, the mechanism of chloroplastic OMP assembly is less understood and requires more research (Fairman et al., 2011; Schleiff et al., 2011).

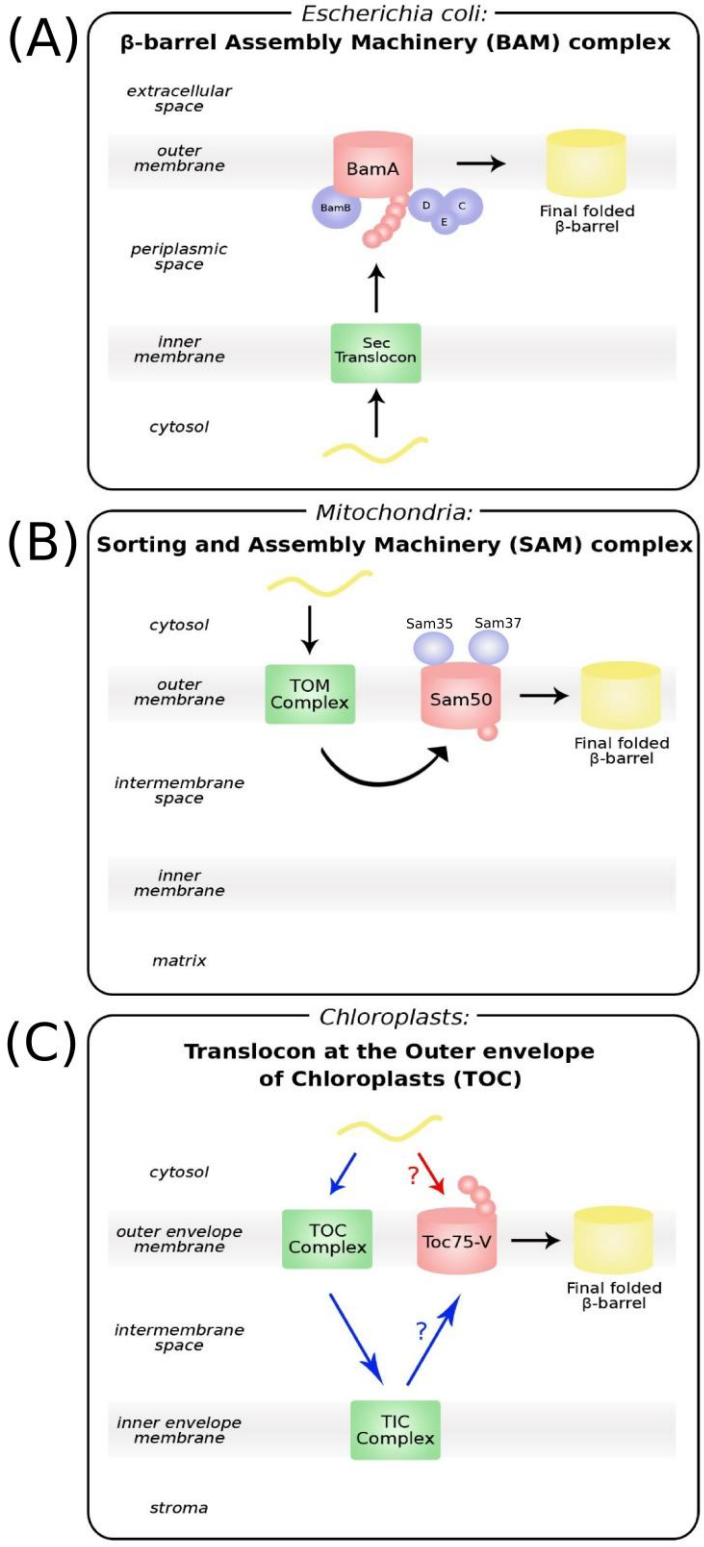


Figure 8-3 (legend on next page)

Figure 8-3 OMP Assembly Systems in Eukaryotes

In both Gram-negative bacteria and eukaryotes, outer membrane β-barrel proteins are first synthesized in the cytosol of the cell and then targeted to either the inner membrane (bacteria) or the proper organelle (mitochondria or chloroplasts). This figure compares the three pathways as the unfolded substrate protein (black curved line) is directed by associated translocons to the assembly complex consisting of the core BamA homologue and accessory proteins, to form the final folded β-barrel (black cylinder). For simplicity, other proteins and chaperones involved in the pathways are not shown. A. The E. coli β-barrel Assembly Machinery (BAM) complex consists of membrane embedded BamA, and four accessory lipoproteins: BamB, C, D, and E. Substrate proteins cross the inner membrane via the Sec translocase, and travel through the periplasmic space before being assembled by the BAM complex at the outer membrane. B. In the mitochondrial system, the substrate proteins enter via the Translocase of Outer Mitochondiral membrane (TOM) and are assembled by the Sorting and Assembly Machinery (SAM) complex. The BamA homologue is Sam50, which works together with cytosolic proteins Sam35 and Sam37 for insertion of OMPs into the outer mitochondrial membrane. C. In chloroplasts, the Translocons at the Outer and Inner envelopes of Chloroplasts (TOC/TIC complexes) are believed to be involved in assembly of OMPs. The BamA homologue is Toc75-V, with accessory proteins yet to be identified. It is unclear if the substrate proteins travel to the stroma prior to being assembled or if they are directly assembled into the outer envelope membrane from the cytosol.

8.4. Conclusion & Future Directions

Structural studies of the BAM complex have not only shown us what each component of the BAM complex looks like, but it has also provided clues to the functional roles of each protein. Future research effort would need to address the questions posed by structural analysis of the BAM proteins. For instance, more experimental evidence is needed to validate the role of BamD as an OMP targeting signal receptor and the ability of BamB to bind substrates. In terms of structural research, the future structural investigation of the BAM complex should focus on determining how the BAM components are arranged within the complex, in addition to determining the structure of the BamA β-barrel domain. Co-crystal structures of a substrate bound to BamA POTRA domains, BamD or BamB would also provide a great deal of insight into the BAM-OMP specificity. These structural studies should be accompanied by binding kinetics studies to characterize how the BAM proteins interact with each other and substrates.

Although great progress has been made in recent years, many aspects of the BAM complex function still remain to be elucidated in order to learn the molecular

mechanism of β-barrel assembly. The major challenge of studying the BAM complex is that currently there is no probe for detecting OMP folding intermediates *in vivo*, which makes it difficult to study exactly what stage of the OMP assembly process is affected when a certain mutation is introduced into the system. Developing a technique that can track the folding status of a substrate would tremendously help dissecting the role of each components of the BAM complex. Time-resolved tryptophan fluorescence quenching techniques have been reported to be useful to track the position of substrate undergoing folding relative to the membrane (i.e. how deep the substrate is inserted into the membrane) (Kleinschmidt, 2003). Perhaps this technique can be incorporated into the *in vitro* BAM complex system reconstituted into a proteoliposome that has proven to be very helpful in studying the BAM complex in isolation and a controlled environment.

Comparing the eukaryotic and bacterial OMP assembly systems yields some interesting similarities and differences between them. For example, unlike BamA, homologues of the lipoproteins BamB-E are not found in eukaryotes (Ricci and Silhavy, 2011; Tommassen, 2010). However, the SAM complex of mitochondria is known to consist of Sam50 (BamA homologue) along with accessory proteins. One of these accessory proteins, Sam35, is essential and has been shown to function in substrate recognition (Chacinska et al., 2009; Milenkovic et al., 2004; Paschen et al., 2005). In bacteria, the essential lipoprotein BamD is predicted to serve a similar role, although Sam35 and BamD do not seem to share any sequence homology. Understanding the roles of the accessory proteins in the eukaryotic OMP assembly systems may therefore shed some light on the exact functions of the BAM lipoproteins, and vice versa.

Could better understanding of the BAM complex be utilized for medical applications? The BAM complex has been suggested to be a suitable drug target for novel antibiotics and vaccine development. The BAM complex is not only essential for the survival of Gram-negative bacteria, but it is functionally non-redundant (i.e. there is no other back-up system in the cell that can perform the same function). Furthermore, the BAM complex is found in the outer membrane, implying uncomplicated drug delivery strategy. In order to realize its potential as a drug target, however, continuing research efforts are required to elucidate the molecular mechanism of the BAM complex.

References

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. 66, 213-221.

Albrecht, R., and Zeth, K. (2011). Structural basis of outer membrane protein biogenesis in bacteria. J. Biol. Chem. 286, 27792-27803.

Albrecht, R., and 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, 1586-1590.

Amy, M., Velge, P., Senocq, D., Bottreau, E., Mompart, F., and Virlogeux-Payant, I. (2004). Identification of a new Salmonella enterica serovar Enteritidis locus involved in cell invasion and in the colonisation of chicks. Res. Microbiol. 155, 543-552.

Anwari,K., Webb,C.T., Poggio,S., Perry,A.J., Belousoff,M., Celik,N., Ramm,G., Lovering,A., Sockett,R.E., Smit,J., Jacobs-Wagner,C., and Lithgow,T. (2012). The evolution of new lipoprotein subunits of the bacterial outer membrane BAM complex. Mol. Microbiol. 84, 832-844.

Anwari, K., Poggio, S., Perry, A., Gatsos, X., Ramarathinam, S.H., Williamson, N.A., Noinaj, N., Buchanan, S., Gabriel, K., Purcell, A.W., Jacobs-Wagner, C., and Lithgow, T. (2010). A modular BAM complex in the outer membrane of the alpha-proteobacterium Caulobacter crescentus. PLoS One 5, e8619.

Arnold, T., Zeth, K., and Linke, D. (2010). Omp85 from the thermophilic cyanobacterium Thermosynechococcus elongatus differs from proteobacterial Omp85 in structure and domain composition. J. Biol. Chem. 285, 18003-18015.

Battye, T.G., Kontogiannis, L., Johnson, O., Powell, H.R., and Leslie, A.G. (2011). iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. Acta Crystallogr. D Biol. Crystallogr. 67, 271-281.

Bechtluft, P., Nouwen, N., Tans, S.J., and Driessen, A.J. (2010). SecB--a chaperone dedicated to protein translocation. Mol. Biosyst 6, 620-627.

Bennett, M.J., Schlunegger, M.P., and Eisenberg, D. (1995). 3D domain swapping: a mechanism for oligomer assembly. Protein Sci. 4, 2455-2468.

Bennion, D., Charlson, E.S., Coon, E., and Misra, R. (2010). Dissection of beta-barrel outer membrane protein assembly pathways through characterizing BamA POTRA 1 mutants of Escherichia coli. Mol. Microbiol. 77, 1153-1171.

Bishop, R.E. (2005). The lipid A palmitoyltransferase PagP: molecular mechanisms and role in bacterial pathogenesis. Mol. Microbiol. 57, 900-912.

Bitto, E., and McKay, D.B. (2003). The periplasmic molecular chaperone protein SurA binds a peptide motif that is characteristic of integral outer membrane proteins. J. Biol. Chem. 278, 49316-49322.

Blow, D. (2002). Outline of Crystallography for Biologists. Oxford University Press, New York.

Bode, W., Papamokos, E., and Musil, D. (1987). The high-resolution X-ray crystal structure of the complex formed between subtilisin Carlsberg and eglin c, an elastase inhibitor from the leech Hirudo medicinalis. Structural analysis, subtilisin structure and interface geometry. Eur. J. Biochem. 166, 673-692.

Bodelon, G., Marin, E., and Fernandez, L.A. (2009). Role of periplasmic chaperones and BamA (YaeT/Omp85) in folding and secretion of intimin from enteropathogenic Escherichia coli strains. J. Bacteriol. 191, 5169-5179.

Bos, M.P., Robert, V., and Tommassen, J. (2007). Biogenesis of the gram-negative bacterial outer membrane. Annu. Rev. Microbiol. 61, 191-214.

Buchanan, S.K. (1999). Beta-barrel proteins from bacterial outer membranes: structure, function and refolding. Curr. Opin. Struct. Biol. 9, 455-461.

Cerna, D., and Wilson, D.K. (2005). The structure of Sif2p, a WD repeat protein functioning in the SET3 corepressor complex. J. Mol. Biol. 351, 923-935.

Chacinska, A., Koehler, C.M., Milenkovic, D., Lithgow, T., and Pfanner, N. (2009). Importing mitochondrial proteins: machineries and mechanisms. Cell 138, 628-644.

Charlson, E.S., Werner, J.N., and Misra, R. (2006). Differential effects of yfgL mutation on Escherichia coli outer membrane proteins and lipopolysaccharide. J. Bacteriol. 188, 7186-7194.

Chayen, N.E., and Saridakis, E. (2008). Protein crystallization: from purified protein to diffraction-quality crystal. Nat. Methods 5, 147-153.

Chen, Z.W., Matsushita, K., Yamashita, T., Fujii, T.A., Toyama, H., Adachi, O., Bellamy, H.D., and Mathews, F.S. (2002). Structure at 1.9 A resolution of a quinohemoprotein alcohol dehydrogenase from Pseudomonas putida HK5. Structure 10, 837-849.

Clantin, B., Delattre, A.S., Rucktooa, P., Saint, N., Meli, A.C., Locht, C., Jacob-Dubuisson, F., and Villeret, V. (2007). Structure of the membrane protein FhaC: a member of the Omp85-TpsB transporter superfamily. Science 317, 957-961.

Collins, R.F., and Derrick, J.P. (2007). Wza: a new structural paradigm for outer membrane secretory proteins? Trends Microbiol. 15, 96-100.

Cornilescu, G., Delaglio, F., and Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J. Biomol. NMR 13, 289-302.

Cross, B.C., Sinning, I., Luirink, J., and High, S. (2009). Delivering proteins for export from the cytosol. Nat. Rev. Mol. Cell Biol. 10, 255-264.

de Cock, H., Struyve, M., Kleerebezem, M., van der Krift, T., and Tommassen, J. (1997). Role of the carboxy-terminal phenylalanine in the biogenesis of outer membrane protein PhoE of Escherichia coli K-12. J. Mol. Biol. 269, 473-478.

Dekker, N., Tommassen, J., Lustig, A., Rosenbusch, J.P., and Verheij, H.M. (1997). Dimerization regulates the enzymatic activity of Escherichia coli outer membrane phospholipase A. J. Biol. Chem. 272, 3179-3184.

Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR 6, 277-293.

DeLano, W. The PyMol (TM) Molecular Graphics System, Version 1.2r2.

Delattre, A.S., Clantin, B., Saint, N., Locht, C., Villeret, V., and Jacob-Dubuisson, F. (2010). Functional importance of a conserved sequence motif in FhaC, a prototypic member of the TpsB/Omp85 superfamily. FEBS J. 277, 4755-4765.

Delcour, A.H. (2003). Solute uptake through general porins. Front. Biosci. 8, d1055-71.

Desvaux, M., Parham, N.J., and Henderson, I.R. (2004). The autotransporter secretion system. Res. Microbiol. 155, 53-60.

Diedrich, D.L., and Cota-Robles, E.H. (1974). Heterogeneity in lipid composition of the outer membrane and cytoplasmic membrane and cytoplasmic membrane of Pseudomonas BAL-31. J. Bacteriol. 119, 1006-1018.

Dong, C., Beis, K., Nesper, J., Brunkan-Lamontagne, A.L., Clarke, B.R., Whitfield, C., and Naismith, J.H. (2006). Wza the translocon for E. coli capsular polysaccharides defines a new class of membrane protein. Nature 444, 226-229.

Dosset, P., Hus, J.C., Blackledge, M., and Marion, D. (2000). Efficient analysis of macromolecular rotational diffusion from heteronuclear relaxation data. J. Biomol. NMR 16, 23-28.

Dundas, J., Ouyang, Z., Tseng, J., Binkowski, A., Turpaz, Y., and Liang, J. (2006). CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. Nucleic Acids Res. 34, W116-8.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. Acta Crystallogr. D Biol. Crystallogr. 60, 2126-2132.

Endo, T., Kawano, S., and Yamano, K. (2011). BamE structure: the assembly of beta-barrel proteins in the outer membranes of bacteria and mitochondria. EMBO Rep. 12, 94-95.

Fairman, J.W., Noinaj, N., and Buchanan, S.K. (2011). The structural biology of beta-barrel membrane proteins: a summary of recent reports. Curr. Opin. Struct. Biol. 21, 523-531.

Fardini, Y., Chettab, K., Grepinet, O., Rochereau, S., Trotereau, J., Harvey, P., Amy, M., Bottreau, E., Bumstead, N., Barrow, P.A., and Virlogeux-Payant, I. (2007). The YfgL lipoprotein is essential for type III secretion system expression and virulence of Salmonella enterica Serovar Enteritidis. Infect. Immun. 75, 358-370.

Farrow, N.A., Muhandiram, R., Singer, A.U., Pascal, S.M., Kay, C.M., Gish, G., Shoelson, S.E., Pawson, T., Forman-Kay, J.D., and Kay, L.E. (1994). Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by 15N NMR relaxation. Biochemistry 33, 5984-6003.

Fulop, V., and Jones, D.T. (1999). Beta propellers: structural rigidity and functional diversity. Curr. Opin. Struct. Biol. 9, 715-721.

Garcia de la Torre, J., Huertas, M.L., and Carrasco, B. (2000). HYDRONMR: prediction of NMR relaxation of globular proteins from atomic-level structures and hydrodynamic calculations. J. Magn. Reson. 147, 138-146.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., and Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server (In) The Proteomics Protocols Handbook. John M. Walker ed., Humana Press) pp. 571-607.

Gatzeva-Topalova, P.Z., Walton, T.A., and Sousa, M.C. (2008). Crystal structure of YaeT: conformational flexibility and substrate recognition. Structure 16, 1873-1881.

Gatzeva-Topalova, P.Z., Warner, L.R., Pardi, A., and Sousa, M.C. (2010). Structure and flexibility of the complete periplasmic domain of BamA: the protein insertion machine of the outer membrane. Structure 18, 1492-1501.

Genevrois, S., Steeghs, L., Roholl, P., Letesson, J.J., and van der Ley, P. (2003). The Omp85 protein of Neisseria meningitidis is required for lipid export to the outer membrane. EMBO J. 22, 1780-1789.

Gentle, I., Gabriel, K., Beech, P., Waller, R., and Lithgow, T. (2004). The Omp85 family of proteins is essential for outer membrane biogenesis in mitochondria and bacteria. J. Cell Biol. 164, 19-24.

Gentle, I.E., Burri, L., and Lithgow, T. (2005). Molecular architecture and function of the Omp85 family of proteins. Mol. Microbiol. 58, 1216-1225.

Gierasch, L.M. (1989). Signal sequences. Biochemistry 28, 923-930.

Goddard, T.D., and Kneller, D.G. SPARKY 3, University of California, San Francisco.

Habeck, M., Rieping, W., Linge, J.P., and Nilges, M. (2004). NOE assignment with ARIA 2.0: the nuts and bolts. Methods Mol. Biol. 278, 379-402.

Hagan, C.L., Silhavy, T.J., and Kahne, D. (2011). beta-Barrel membrane protein assembly by the Bam Complex. Annu. Rev. Biochem. 80, 189-210.

Hao, B., Oehlmann, S., Sowa, M.E., Harper, J.W., and Pavletich, N.P. (2007). Structure of a Fbw7-Skp1-cyclin E complex: multisite-phosphorylated substrate recognition by SCF ubiquitin ligases. Mol. Cell 26, 131-143.

Hayashi, S., and Wu, H.C. (1990). Lipoproteins in bacteria. J. Bioenerg. Biomembr. 22, 451-471.

Heuck, A., Schleiffer, A., and Clausen, T. (2011). Augmenting beta-augmentation: structural basis of how BamB binds BamA and may support folding of outer membrane proteins. J. Mol. Biol. 406, 659-666.

Hoffmann, A., Bukau, B., and Kramer, G. (2010). Structure and function of the molecular chaperone Trigger Factor. Biochim. Biophys. Acta 1803, 650-661.

Holm, L., Kaariainen, S., Rosenstrom, P., and Schenkel, A. (2008). Searching protein structure databases with DaliLite v.3. Bioinformatics 24, 2780-2781.

Ieva, R., Tian, P., Peterson, J.H., and Bernstein, H.D. (2011). Sequential and spatially restricted interactions of assembly factors with an autotransporter beta domain. Proc. Natl. Acad. Sci. U. S. A. 108, E383-91.

Jacob-Dubuisson, F., Locht, C., and Antoine, R. (2001). Two-partner secretion in Gram-negative bacteria: a thrifty, specific pathway for large virulence proteins. Mol. Microbiol. 40, 306-313.

Jain, S., and Goldberg, M.B. (2007). Requirement for YaeT in the outer membrane assembly of autotransporter proteins. J. Bacteriol. 189, 5393-5398.

Kanelis, V., Forman-Kay, J.D., and Kay, L.E. (2001). Multidimensional NMR methods for protein structure determination. IUBMB Life 52, 291-302.

Kay, L.E. (2005). NMR studies of protein structure and dynamics. J. Magn. Reson. 173, 193-207.

Kim, K.H., Aulakh, S., and Paetzel, M. (2011a). Crystal structure of the {beta}-barrel assembly machinery BamCD complex. J. Biol. Chem.

Kim, K.H., Aulakh, S., Tan, W., and Paetzel, M. (2011b). Crystallographic analysis of the C-terminal domain of the Escherichia coli lipoprotein BamC. Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun. 67, 1350-1358.

Kim, S., Malinverni, J.C., Sliz, P., Silhavy, T.J., Harrison, S.C., and Kahne, D. (2007). Structure and function of an essential component of the outer membrane protein assembly machine. Science 317, 961-964.

Kleinschmidt, J.H. (2006). Folding kinetics of the outer membrane proteins OmpA and FomA into phospholipid bilayers. Chem. Phys. Lipids 141, 30-47.

Kleinschmidt, J.H. (2003). Membrane protein folding on the example of outer membrane protein A of Escherichia coli. Cell Mol. Life Sci. 60, 1547-1558.

Knowles, T.J., Browning, D.F., Jeeves, M., Maderbocus, R., Rajesh, S., Sridhar, P., Manoli, E., Emery, D., Sommer, U., Spencer, A., et al. (2011). Structure and function of BamE within the outer membrane and the beta-barrel assembly machine. EMBO Rep. 12, 123-128.

Knowles, T.J., Jeeves, M., Bobat, S., Dancea, F., McClelland, D., Palmer, T., Overduin, M., and Henderson, I.R. (2008). Fold and function of polypeptide transport-associated domains responsible for delivering unfolded proteins to membranes. Mol. Microbiol. 68, 1216-1227.

Knowles, T.J., McClelland, D.M., Rajesh, S., Henderson, I.R., and Overduin, M. (2009a). Secondary structure and (1)H, (13)C and (15)N backbone resonance assignments of BamC, a component of the outer membrane protein assembly machinery in Escherichia coli. Biomol. NMR Assign 3, 203-206.

Knowles, T.J., Scott-Tucker, A., Overduin, M., and Henderson, I.R. (2009b). Membrane protein architects: the role of the BAM complex in outer membrane protein assembly. Nat. Rev. Microbiol. 7, 206-214.

Koebnik, R., Locher, K.P., and Van Gelder, P. (2000). Structure and function of bacterial outer membrane proteins: barrels in a nutshell. Mol. Microbiol. 37, 239-253.

Koronakis, V., Sharff, A., Koronakis, E., Luisi, B., and Hughes, C. (2000). Crystal structure of the bacterial membrane protein TolC central to multidrug efflux and protein export. Nature 405, 914-919.

Krewulak, K.D., and Vogel, H.J. (2011). TonB or not TonB: is that the question? Biochem. Cell Biol. 89, 87-97.

Krissinel, E., and Henrick, K. Protein interfaces, surfaces and assemblies service PISA at European Bioinformatics Institute

Kusters, I., and Driessen, A.J. (2011). SecA, a remarkable nanomachine. Cell Mol. Life Sci. 68, 2053-2066.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., et al. (2007). Clustal W and Clustal X version 2.0. Bioinformatics 23, 2947-2948.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. J. App. Cryst. 26, 283-291.

Lewis, C., Skovierova, H., Rowley, G., Rezuchova, B., Homerova, D., Stevenson, A., Sherry, A., Kormanec, J., and Roberts, M. (2008). Small outer-membrane lipoprotein, SmpA, is regulated by sigmaE and has a role in cell envelope integrity and virulence of Salmonella enterica serovar Typhimurium. Microbiology 154, 979-988.

Liu, Y., and Eisenberg, D. (2002). 3D domain swapping: as domains continue to swap. Protein Sci. 11, 1285-1299.

Malinverni, J.C., Werner, J., Kim, S., Sklar, J.G., Kahne, D., Misra, R., and Silhavy, T.J. (2006). YfiO stabilizes the YaeT complex and is essential for outer membrane protein assembly in Escherichia coli. Mol. Microbiol. 61, 151-164.

McCarthy, A.A., Haebel, P.W., Torronen, A., Rybin, V., Baker, E.N., and Metcalf, P. (2000). Crystal structure of the protein disulfide bond isomerase, DsbC, from Escherichia coli. Nat. Struct. Biol. 7, 196-199.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J. Appl. Crystallogr. 40, 658-674.

Merdanovic, M., Clausen, T., Kaiser, M., Huber, R., and Ehrmann, M. (2011). Protein quality control in the bacterial periplasm. Annu. Rev. Microbiol. 65, 149-168.

Milenkovic, D., Kozjak, V., Wiedemann, N., Lohaus, C., Meyer, H.E., Guiard, B., Pfanner, N., and Meisinger, C. (2004). Sam35 of the mitochondrial protein sorting and assembly machinery is a peripheral outer membrane protein essential for cell viability. J. Biol. Chem. 279, 22781-22785.

Morein, S., Andersson, A., Rilfors, L., and Lindblom, G. (1996). Wild-type Escherichia coli cells regulate the membrane lipid composition in a "window" between gel and non-lamellar structures. J. Biol. Chem. 271, 6801-6809.

Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr. D Biol. Crystallogr. 53, 240-255.

Nikaido, H. (2003). Molecular basis of bacterial outer membrane permeability revisited. Microbiol. Mol. Biol. Rev. 67, 593-656.

Nikaido, H. (1994). Porins and specific diffusion channels in bacterial outer membranes. J. Biol. Chem. 269, 3905-3908.

Noinaj, N., Fairman, J.W., and Buchanan, S.K. (2011). The crystal structure of BamB suggests interactions with BamA and its role within the BAM complex. J. Mol. Biol. 407, 248-260.

Oke, M., Sarra, R., Ghirlando, R., Farnaud, S., Gorringe, A.R., Evans, R.W., and Buchanan, S.K. (2004). The plug domain of a neisserial TonB-dependent transporter retains structural integrity in the absence of its transmembrane beta-barrel. FEBS Lett. 564, 294-300.

Okuda, S., and Tokuda, H. (2011). Lipoprotein sorting in bacteria. Annu. Rev. Microbiol. 65, 239-259.

Onufryk, C., Crouch, M.L., Fang, F.C., and Gross, C.A. (2005). Characterization of six lipoproteins in the sigmaE regulon. J. Bacteriol. 187, 4552-4561.

Oreb, M., Tews, I., and Schleiff, E. (2008). Policing Tic 'n' Toc, the doorway to chloroplasts. Trends Cell Biol. 18, 19-27.

Orengo, C.A., Michie, A.D., Jones, S., Jones, D.T., Swindells, M.B., and Thornton, J.M. (1997). CATH--a hierarchic classification of protein domain structures. Structure 5, 1093-1108.

Paetzel, M., Karla, A., Strynadka, N.C., and Dalbey, R.E. (2002). Signal peptidases. Chem. Rev. 102, 4549-4580.

Painter, J., and Merritt, E.A. (2006). TLSMD web server for the generation of multi-group TLS models. J. App. Cryst. 39, 109-111.

Palomino, C., Marin, E., and Fernandez, L.A. (2011). The Fimbrial Usher FimD Follows the SurA-BamB Pathway for Its Assembly in the Outer Membrane of Escherichia coli. J. Bacteriol. 193, 5222-5230.

Paschen, S.A., Neupert, W., and Rapaport, D. (2005). Biogenesis of beta-barrel membrane proteins of mitochondria. Trends Biochem. Sci. 30, 575-582.

Patel, G.J., Behrens-Kneip, S., Holst, O., and Kleinschmidt, J.H. (2009). The periplasmic chaperone Skp facilitates targeting, insertion, and folding of OmpA into lipid membranes with a negative membrane surface potential. Biochemistry 48, 10235-10245.

Phan, G., Remaut, H., Wang, T., Allen, W.J., Pirker, K.F., Lebedev, A., Henderson, N.S., Geibel, S., Volkan, E., Yan, J., et al. (2011). Crystal structure of the FimD usher bound to its cognate FimC-FimH substrate. Nature 474, 49-53.

Postle, K., and Kadner, R.J. (2003). Touch and go: tying TonB to transport. Mol. Microbiol. 49, 869-882.

Reynolds, K.A., Thomson, J.M., Corbett, K.D., Bethel, C.R., Berger, J.M., Kirsch, J.F., Bonomo, R.A., and Handel, T.M. (2006). Structural and computational characterization of the SHV-1 beta-lactamase-beta-lactamase inhibitor protein interface. J. Biol. Chem. 281, 26745-26753.

Rhodes, G. (2006). Crystallography Made Crystal Clear, Third Edition: A Guide for Users of Macromolecular Models. Burlington, VT. Academic Press.

Ricci, D.P., and Silhavy, T.J. (2011). The Bam machine: A molecular cooper. Biochim. Biophys. Acta

Rigel, N.W., Schwalm, J., Ricci, D.P., and Silhavy, T.J. (2012). BamE modulates the Escherichia coli beta-barrel assembly machine component BamA. J. Bacteriol. 194, 1002-1008.

Rizzitello, A.E., Harper, J.R., and Silhavy, T.J. (2001). Genetic evidence for parallel pathways of chaperone activity in the periplasm of Escherichia coli. J. Bacteriol. 183, 6794-6800.

Robert, V., Volokhina, E.B., Senf, F., Bos, M.P., Van Gelder, P., and Tommassen, J. (2006). Assembly factor Omp85 recognizes its outer membrane protein substrates by a species-specific C-terminal motif. PLoS Biol. 4, e377.

Rossiter, A.E., Leyton, D.L., Tveen-Jensen, K., Browning, D.F., Sevastsyanovich, Y., Knowles, T.J., Nichols, K.B., Cunningham, A.F., Overduin, M., Schembri, M.A., and Henderson, I.R. (2011). The essential beta-barrel assembly machinery complex components BamD and BamA are required for autotransporter biogenesis. J. Bacteriol. 193, 4250-4253.

Ruiz, N., Wu, T., Kahne, D., and Silhavy, T.J. (2006). Probing the barrier function of the outer membrane with chemical conditionality. ACS Chem. Biol. 1, 385-395.

Sakakibara, D., Sasaki, A., Ikeya, T., Hamatsu, J., Hanashima, T., Mishima, M., Yoshimasu, M., Hayashi, N., Mikawa, T., Walchli, M., et al. (2009). Protein structure determination in living cells by in-cell NMR spectroscopy. Nature 458, 102-105.

Sanchez-Pulido, L., Devos, D., Genevrois, S., Vicente, M., and Valencia, A. (2003). POTRA: a conserved domain in the FtsQ family and a class of beta-barrel outer membrane proteins. Trends Biochem. Sci. 28, 523-526.

Sandoval, C.M., Baker, S.L., Jansen, K., Metzner, S.I., and Sousa, M.C. (2011). Crystal structure of BamD: an essential component of the beta-Barrel assembly machinery of gram-negative bacteria. J. Mol. Biol. 409, 348-357.

Sattler, M., Schleucher, J., and Griesinger, C. (1999). Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. Prog. NMR Spectrosc. 34, 93-158.

Sauri, A., Soprova, Z., Wickstrom, D., de Gier, J.W., Van der Schors, R.C., Smit, A.B., Jong, W.S., and Luirink, J. (2009). The Bam (Omp85) complex is involved in secretion of the autotransporter haemoglobin protease. Microbiology 155, 3982-3991.

Schatz, G., and Dobberstein, B. (1996). Common principles of protein translocation across membranes. Science 271, 1519-1526.

Schleiff, E., Maier, U.G., and Becker, T. (2011). Omp85 in eukaryotic systems: one protein family with distinct functions. Biol. Chem. 392, 21-27.

Silhavy, T.J., Kahne, D., and Walker, S. (2010). The bacterial cell envelope. Cold Spring Harb Perspect. Biol. 2, a000414.

Sjogren, T., and Hajdu, J. (2001). Structure of the bound dioxygen species in the cytochrome oxidase reaction of cytochrome cd1 nitrite reductase. J. Biol. Chem. 276, 13072-13076.

Sklar, J.G., Wu, T., Gronenberg, L.S., Malinverni, J.C., Kahne, D., and Silhavy, T.J. (2007a). 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, 6400-6405.

Sklar, J.G., Wu, T., Kahne, D., and Silhavy, T.J. (2007b). Defining the roles of the periplasmic chaperones SurA, Skp, and DegP in Escherichia coli. Genes Dev. 21, 2473-2484.

Smyth, M.S., and Martin, J.H. (2000). X Ray Crystallography. Mol. Pathol. 53, 8-14.

Snijder, H.J., Ubarretxena-Belandia, I., Blaauw, M., Kalk, K.H., Verheij, H.M., Egmond, M.R., Dekker, N., and Dijkstra, B.W. (1999). Structural evidence for dimerization-regulated activation of an integral membrane phospholipase. Nature 401, 717-721.

Sommer, M.S., Daum, B., Gross, L.E., Weis, B.L., Mirus, O., Abram, L., Maier, U.G., Kuhlbrandt, W., and Schleiff, E. (2011). Chloroplast Omp85 proteins change orientation during evolution. Proc. Natl. Acad. Sci. U. S. A. 108, 13841-13846.

Stroud, D.A., Becker, T., Qiu, J., Stojanovski, D., Pfannschmidt, S., Wirth, C., Hunte, C., Guiard, B., Meisinger, C., Pfanner, N., and Wiedemann, N. (2011). Biogenesis of mitochondrial beta-barrel proteins: the POTRA domain is involved in precursor release from the SAM complex. Mol. Biol. Cell 22, 2823-2833.

Struyve, M., Moons, M., and Tommassen, J. (1991). Carboxy-terminal phenylalanine is essential for the correct assembly of a bacterial outer membrane protein. J. Mol. Biol. 218, 141-148.

Strynadka, N.C., Jensen, S.E., Alzari, P.M., and James, M.N. (1996). A potent new mode of beta-lactamase inhibition revealed by the 1.7 A X-ray crystallographic structure of the TEM-1-BLIP complex. Nat. Struct. Biol. 3, 290-297.

Stumpe, S., Schmid, R., Stephens, D.L., Georgiou, G., and Bakker, E.P. (1998). Identification of OmpT as the protease that hydrolyzes the antimicrobial peptide protamine before it enters growing cells of Escherichia coli. J. Bacteriol. 180, 4002-4006.

Sugimura, K., and Nishihara, T. (1988). Purification, characterization, and primary structure of Escherichia coli protease VII with specificity for paired basic residues: identity of protease VII and OmpT. J. Bacteriol. 170, 5625-5632.

Surrey, T., and Jahnig, F. (1992). Refolding and oriented insertion of a membrane protein into a lipid bilayer. Proc. Natl. Acad. Sci. U. S. A. 89, 7457-7461.

Tamm, L.K., Arora, A., and Kleinschmidt, J.H. (2001). Structure and assembly of beta-barrel membrane proteins. J. Biol. Chem. 276, 32399-32402.

Tamm, L.K., Hong, H., and Liang, B. (2004). Folding and assembly of beta-barrel membrane proteins. Biochim. Biophys. Acta 1666, 250-263.

Tokuda, H. (2009). Biogenesis of outer membranes in Gram-negative bacteria. Biosci. Biotechnol. Biochem. 73, 465-473.

Ureta, A.R., Endres, R.G., Wingreen, N.S., and Silhavy, T.J. (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, 446-454.

van Ulsen, P. (2011). Protein folding in bacterial adhesion: secretion and folding of classical monomeric autotransporters. Adv. Exp. Med. Biol. 715, 125-142.

Vance, D.E., and Vance, J.E. (2008). Biochemistry of lipids, lipoproteins and membranes (5th Edn.).

Vandeputte-Rutten, L., Kramer, R.A., Kroon, J., Dekker, N., Egmond, M.R., and Gros, P. (2001). Crystal structure of the outer membrane protease OmpT from Escherichia coli suggests a novel catalytic site. EMBO J. 20, 5033-5039.

Vanini, M.M., Spisni, A., Sforca, M.L., Pertinhez, T.A., and Benedetti, C.E. (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, 2051-2064.

Vogt, J., and Schulz, G.E. (1999). The structure of the outer membrane protein OmpX from Escherichia coli reveals possible mechanisms of virulence. Structure 7, 1301-1309.

Vollmer, W., and Bertsche, U. (2008). Murein (peptidoglycan) structure, architecture and biosynthesis in Escherichia coli. Biochim. Biophys. Acta 1778, 1714-1734.

Vollmer, W., Blanot, D., and de Pedro, M.A. (2008). Peptidoglycan structure and architecture. FEMS Microbiol. Rev. 32, 149-167.

Vollmer, W., von Rechenberg, M., and Holtje, J.V. (1999). Demonstration of molecular interactions between the murein polymerase PBP1B, the lytic transglycosylase MltA, and the scaffolding protein MipA of Escherichia coli. J. Biol. Chem. 274, 6726-6734.

Volokhina, E.B., Grijpstra, J., Stork, M., Schilders, I., Tommassen, J., and Bos, M.P. (2011). Role of the periplasmic chaperones Skp, SurA, and DegQ in outer membrane protein biogenesis in Neisseria meningitidis. J. Bacteriol. 193, 1612-1621.

von Heijne, G. (1990). The signal peptide. J. Membr. Biol. 115, 195-201.

Voulhoux, R., Bos, M.P., Geurtsen, J., Mols, M., and Tommassen, J. (2003). Role of a highly conserved bacterial protein in outer membrane protein assembly. Science 299, 262-265.

Vuong, P., Bennion, D., Mantei, J., Frost, D., and Misra, R. (2008). Analysis of YfgL and YaeT interactions through bioinformatics, mutagenesis, and biochemistry. J. Bacteriol. 190, 1507-1517.

Wagner, G., Hyberts, S.G., and Havel, T.F. (1992). NMR structure determination in solution: a critique and comparison with X-ray crystallography. Annu. Rev. Biophys. Biomol. Struct. 21, 167-198.

Walther, D.M., Rapaport, D., and Tommassen, J. (2009). Biogenesis of beta-barrel membrane proteins in bacteria and eukaryotes: evolutionary conservation and divergence. Cell Mol. Life Sci. 66, 2789-2804.

Warner, L.R., Varga, K., Lange, O.F., Baker, S.L., Baker, D., Sousa, M.C., and Pardi, A. (2011). Structure of the BamC two-domain protein obtained by Rosetta with a limited NMR data set. J. Mol. Biol. 411, 83-95.

Wimley, W.C. (2003). The versatile beta-barrel membrane protein. Curr. Opin. Struct. Biol. 13, 404-411.

Wlodawer, A., Minor, W., Dauter, Z., and Jaskolski, M. (2008). Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. FEBS J. 275, 1-21.

Wu, T., Malinverni, J., Ruiz, N., Kim, S., Silhavy, T.J., and Kahne, D. (2005). Identification of a multicomponent complex required for outer membrane biogenesis in Escherichia coli. Cell 121, 235-245.

Ye, Y., and Godzik, A. (2003). Flexible structure alignment by chaining aligned fragment pairs allowing twists. Bioinformatics 19 Suppl 2, ii246-55.

Zgurskaya, H.I., Krishnamoorthy, G., Ntreh, A., and Lu, S. (2011). Mechanism and Function of the Outer Membrane Channel TolC in Multidrug Resistance and Physiology of Enterobacteria. Front. Microbiol. 2, 189.

Zhang, H., Gao, Z.Q., Hou, H.F., Xu, J.H., Li, L.F., Su, X.D., and Dong, Y.H. (2011). High-resolution structure of a new crystal form of BamA POTRA4-5 from Escherichia coli. Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun. 67, 734-738.

Ziegler, K., Benz, R., and Schulz, G.E. (2008). A putative alpha-helical porin from Corynebacterium glutamicum. J. Mol. Biol. 379, 482-491.

Zimmer, J., Nam, Y., and Rapoport, T.A. (2008). Structure of a complex of the ATPase SecA and the protein-translocation channel. Nature 455, 936-943.

Zwahlen, C., Gardner, K.H., Sarma, S.P., Horita, D.A., Byrd, R.A., and Kay, L.E. (1998). An NMR experiment for measuring methyl-methyl NOEs in 13C labeled proteins with high resolution. J. Am. Chem. Soc. 120, 7617-7625.

Appendices

Appendix A.

List of *E. coli* OMPs

Protein	Length in amino acids	Function	UniProt ID	Evidence as BAM substrate
AfaC	859	Usher	P53517	
Ag43	1039	Autotransporter	P39180	Rossiter et al., 2011
AggC	842	Usher	P46005	
AIDA-I	1286	Autotransporter	Q03155	Jain and Goldberg, 2007
BglH	538	Sugar transport	P26218	
BtuB	614	Vitamin B12 transport	P06129	
CirA	663	Iron transport	P17315	
CS3-2	937	Usher	P15484	
CssD	819	Usher	P53512	
EaE (Intimin)	934	Autotransporter	P43261	Bodelon et al., 2009
EatA	1364	Autotransporter	Q84GK0	
ElfC	866	Usher	P75857	
EspC	1305	Autotransporter	Q9EZE7	
EspP	1300	Autotransporter	Q7BSW5	
FadL	446	Long-chain fatty acid transport	P10384	
FaeD	812	Usher	P06970	
FanD	783	Usher	P12050	
FasD	835	Usher	P46000	
FecA	774	Iron transport	P13036	
FepA	746	Iron transport	P05825	
FhuA	747	Iron transport	P06971	
FhuE	729	Iron transport	P16869	
FimD	878	Usher	P30130	Palomino et al., 2011
Fiu	760	Iron transport	P75780	
FocD	875	Usher	P46009	
Hbp	1377	Autotransporter	O88093	Sauri et al., 2009
HtrE	865	Usher	P33129	
IutA	732	Iron transporter	P14542	
LamB	446	Maltose transport	P02943	Malinverni et al., 2006
LptD	784	LPS assembly protein	P31554	
MipA	248	Peptidoglycan synthesis scaffold protein	P0A908	
NanC	238	N-acetylneuraminic acid transport	P69856	
NfrA	990	N4 bacteriophage receptor	P31600	

NmpC	365	Porin	P21420	
OmpA	346	OM stability, bacterial conjugation	P0A910	Malinverni et al., 2006
OmpC	367	Porin	P06996	Malinverni et al., 2006
OmpF	362	Porin	P02931	Malinverni et al., 2006
OmpG	301	Sugar transport	P76045	
OmpL	230	Sugar transport	P76773	
OmpLA	289	Phospholipase	P0A921	
OmpN	377	Porin	P77747	
OmpP	315	Protease	P34210	
OmpT	317	Protease	P09169	Hagan et al., 2010
OmpW	212	Colicin S4 receptor	P0A915	
OmpX	171	Adhesin	P0A917	
PapC	836	Usher	P07110	
PagP	186	Lipid A palmitoyltransferase	P37001	
PcoB	296	Copper resistance	Q47453	
Pet	1295	Autotransporter	O68900	Rossiter, et al., 2011
PgaA	807	poly-beta-1,6-N-acetyl-D-glucosamine transport	P69434	
PhoE	351	Porin	P02932	Robert et al., 2006
Pic	1371	Autotransporter	Q8CWC7	
Sat	1295	Autotransporter	Q8FDW4	
SfmD	867	Usher	P77468	
TibA	989	Autotransporter	Q9XD84	
TolC	493	OM export protein	P02930	Malinverni et al., 2006
TraN	602	Transfer of F plasmid during conjugation	P24082	
Tsh	1377	Autotransporter	Q47692	
Tsx	294	Nucleoside specific channel	P0A927	
UidC	421	Involved in glucuronide transport	Q47706	
YaeT (BamA)	810	Assembly of OMPs	P0A940	
YbgQ	815	Usher	P75750	
YedS	397	Porin	P76335	
YehB	826	Usher	P33341	
YejO	863	Autotransporter	P33924	
YfcU	881	Usher	P77196	
YhcD	793	Usher	P45420	
YiaT	246	Peptidoglycan synthesis scaffold protein	P37681	
YncD	700	Channel (TonB-dependent)	P76115	
YpjA	1526	Autotransporter	P52143	
YqiG	821	Usher	P76655	
YraJ	838	Usher	P42915	
YuaO	1758	Autotransporter	Q9JMS5	
YuaQ	1371	Autotransporter	Q9JMS3	

Appendix B.

Bacterial Lipoprotein Biogenesis

Lipoproteins are soluble proteins that are found anchored to membranes via covalently linked lipid moieties. In bacteria, lipoproteins are involved in a wide variety of cellular functions, such as biogenesis and maintenance of cell surface structures, transport of substrates and drug efflux. In *E. coli*, at least 90 different lipoproteins are expressed, although the functions of most lipoproteins are unknown (Okuda and Tokuda, 2011).

Lipoproteins of both Gram-positive and Gram-negative bacteria begin in cytosol where they are first synthesized with an N-terminal signal peptide. The signal peptide of lipoproteins contains a consensus sequence, Lue-(Ala/Ser)-(Gly/Ala)-Cys, known as a lipobox (Tokuda, 2009). The signal peptide targets the lipoprotein precursors to the inner membrane, where the Sec translocon facilitates the translocation of the proteins across the membrane. Following the translocation, the N-terminal signal peptide is still retained in the inner membrane, and the processing of the lipoprotein precursors into mature forms (i.e. lipid modification and signal peptide cleavage) takes place on the periplasmic side of the inner membrane (Tokuda, 2009; Okuda and Tokuda, 2011).

The lipoprotein maturation process occurs in three steps and requires three well-conserved enzymes: Lgt (prolipoprotein diacylglyceryl transferase), LspA (prollipoprotein signal peptidase) and Lnt (apolipoprotein N-acyltransferase) (Figure B1). In the first step, Lgt forms thioester linkage between the Cys residue in the lipobox and diaclglycerol. LspA (or signal peptidase II) then removes the signal peptide by cleaving the peptide bond between the (Gly/Ala) and the lipid-modified Cys residues. The Cys residue is thus now at the N-terminus of the cleaved lipoprotein. Finally the last enzyme, Lnt, acylates the N-terminal Cys residue, and the lipoprotein is now in its mature form with its triacylated N-terminal Cys residue (Okuda and Tokuda, 2011).

In Gram-negative bacteria, lipoproteins are found on the periplasmic side of the inner and the outer membranes (Tokuda, 2009). Whether the lipoprotein is retained in the inner membrane following the maturation process mentioned above or is transported to the outer membrane is determined by the N-terminal second residue found in the mature lipoprotein. If the lipoprotein contains an Asp residue immediately following the N-terminal Cys residue, the lipoprotein stays anchored to the inner membrane; however, if the Asp residue is absent, then the lipoprotein is transported to the outer membrane via the Lol (localization of lipoproteins) pathway (Tokuda, 2009).

The Lol system consists of three different components: 1) LolCDE complex, 2) LolA, and 3) LolB. The LolCDE complex, which is an ATPase, functions to release outer membrane-specific lipoproteins from the inner membrane. The relased lipoprotein is then bound by LolA, a chaperone that carries the lipoprotein across the periplasm to the outer membrane. At the outer membrane, the lipoprotein is transferred from LolA to a receptor protein LolB, which then catalyzes the anchoring of the lipoprotein to the outer membrane lipid bilayer. Like the three enzymes that are required for the initial lipid modification and signal peptide removal processes, the Lol proteins are well conserved in Gram-negative bacteria (Okuda and Tokuda, 2011).

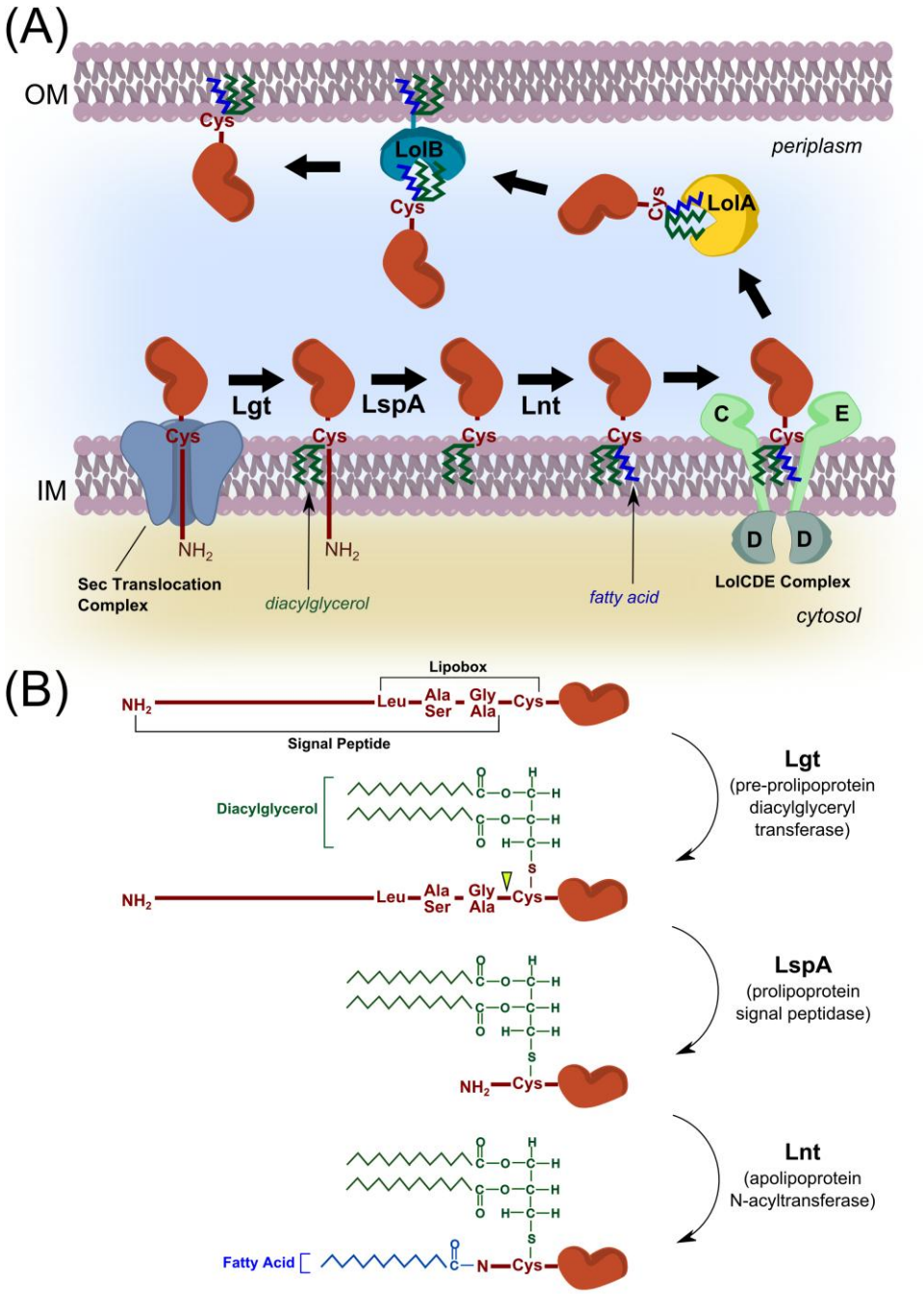


Figure B1. Bacterial Lipoprotein Biogenesis. (A) Following the translocation across the inner membrane, a lipoprotein precursor is modified by three enzymes. First, Lgt adds diacylglycerol to a Cys residue found immediately following the signal sequence. This Cys residue becomes the N-terminal-most residue of the lipoprotein after LspA removes the signal peptide. Following the signal peptide removal, Lnt covalently links the Cys residue to a fatty acid. Lipoproteins destined for the outer membrane are released from the inner membrane by the LolCDE complex, and subsequently transferred to a periplasmic chaperone, LolA. The LolA-lipoprotein complex is then carried to the outer membrane where LolB catalyzes the anchoring of the lipoprotein to the outer membrane lipid bilayer. (B) A more detailed view of the lipid modification steps described in (A) are shown.

Appendix C.

Cloning Details

Construct	Residues	Vector	MWP #	Primers		Restriction Enzymes	UniProt ID	pI‡	Extinction Coefficient ($M^{-1}cm^{-1}$) ‡
				Forward (5' to 3')	Reverse (5' to 3')				
BamA	21-810	pET28a	275	atatacata tggctgaag ggttcg	atatagaat tcttattacc aggtttta	NdeI / EcoRI	P0A940	5.07	1.4×10^5
$BamA_{POTRA}$	21-434	pET28a	849	ggtattggt tacggttaa gaaagtgg cgtgagc	gctcacgcc actttcttaa ccgtaacca atacc	SDM*		5.99	4.0×10^4
BamB	21-392	pET28a	872	atatacata tgtcgctgtt taacagc	atatactcg agttattaa cgtgtaata ga	NdeI / XhoI	P77774	5.01	6.3×10^4
BamC	26-344	pET28a	873	atatacata tgagttctg actcacgc	tataaagct tttattactt gctaaacgc ag	NdeI / XhoI	P0A903	5.74	4.2×10^4
$BamC_N$	99-217	pET24a	858	tatacatat gttcacggg cgatacc	tatactcga gcgcggcgt cagtggc	NdeI / XhoI		6.27	2.4×10^4
$BamC_C$	220-344	pET24a	859	tatacatat ggcgcaaa atcgtgcc	tatactcga gcttgctaa acgcagc	NdeI / XhoI		6.35	1.4×10^4
$BamC_{NC}$	94-344	pET28a	860	tatattcat atgggcgc gcgtaccca	tataaagct tttattactt gctaaacgc ag	NdeI / HindIII		5.95	3.8×10^4
$BamC_{UN}$	26-217	pET24a	857	atatacata tgagttctg actcacgc	tatactcga gcgcggcgt cagtggc	NdeI / XhoI		6.14	2.8×10^4
BamD	21-245	pET28a	874	atatacata tgtcggggt caaaggaa	atatactcg agtgtattg ctgct	NdeI / XhoI	P0AC02	6.44	3.5×10^4
BamE	21-113	pET28a	576	atacatatg tccactctg gag	tatactcga gttattagtt accactc	NdeI / XhoI	P0A937	8.22	1.1×10^4

* $BamA_{POTRA}$ construct was created by site-directed mutagenesis, and therefore restriction enzymes were not used.

‡ Calculated by ProtParam (Gasteiger et al., 2005).

Appendix D.

Statistics in Crystallography

R_{merge}

A measure of agreement among multiple measurements of the same (not symmetry-related) reflections, with the different measurements being in different frames of data or different data sets. R_{merge} is calculated as follows:

$$R_{merge} = \frac{\sum_h \sum_i |I_i - \langle I \rangle|}{\sum_h \sum_i I_i}$$

I_i = 'i'th intensity measurement of reflection 'h'

<I> = the average intensity from multiple observations

R_{symm}

A measure of agreement among the independent measurements of symmetry-related reflections in a crystallographic data set. R_{symm} is calculated as follows:

$$R_{symm} = \frac{\sum_h (I_h - \bar{I}_h)}{\sum_h \bar{I}_h}$$

I and $\bar{I}$ represent intensities of two symmetry-related reflections.

Theoretically, symmetry-related reflections should have identical intensities. If there is a large difference, it suggests some type of measurement error.

Mean(I) / σI

The average intensity of reflections (Mean(I)) divided by the average standard deviation of the reflections (σI). It reports signal over noise.

Completeness

The number of crystallographic reflections measured in a data set, expressed as a percentage of the total number of reflections present at the specified resolution.

Redundancy

The average number of independent measurements of each reflection in a crystallographic data set. Redundancy is calculated as follows:

$$Redundancy = \frac{number\ of\ measured\ reflections}{number\ of\ unique\ reflections}$$

FOM

FOM is figure of merit. Reflections of uncertain phase determination will introduce errors into an electron-density map, and therefore they must be weighted down. Figure of merit is the appropriate weight, calculated from the phase probability distribution. FOM is 1 for a perfectly defined phase angle and becomes closer to 0 as the errors increase. FOM is calculated as follows:

$$FOM = \left| \frac{\sum P(\alpha) e^{i\alpha}}{\sum P(\alpha)} \right|$$

α = phase angle $P(\alpha)$ = phase probability distribution.

R-factor

A measure of agreement between the crystallographic model and the original diffraction data. From the model, the expected intensity of each reflection in the diffraction pattern is calculated, and then compared to the experimental data. The R-factor is calculated as follows:

$$R = \frac{\sum \left| |F_{obs}| - |F_{calc}| \right|}{\sum |F_{obs}|}$$

$|F_{obs}|$ is the structure factor (a function of the amplitude and phase of the diffracted waves) derived from the measured intensity of a reflection in the diffraction pattern, and $|F_{calc}|$ is the structure factor of the same reflection calculated from the current model.

R_{free}

Computed in the same manner as R-factor, but using only a small set of randomly chosen intensities (the "test set") which are set aside from the beginning and not used during refinement. R_{free} gives a better and less-biased measure of refinement progress.

B-factor

A measure of how much an atom oscillates or vibrates around the position specified in the model. Higher B-factor value reflects higher disorder. The B-factor for a given atom 'j' is related to the magnitude of vibration as follows:

$$B_j = 8\pi^2\{u_j^2\} = 79\{u_j^2\}$$

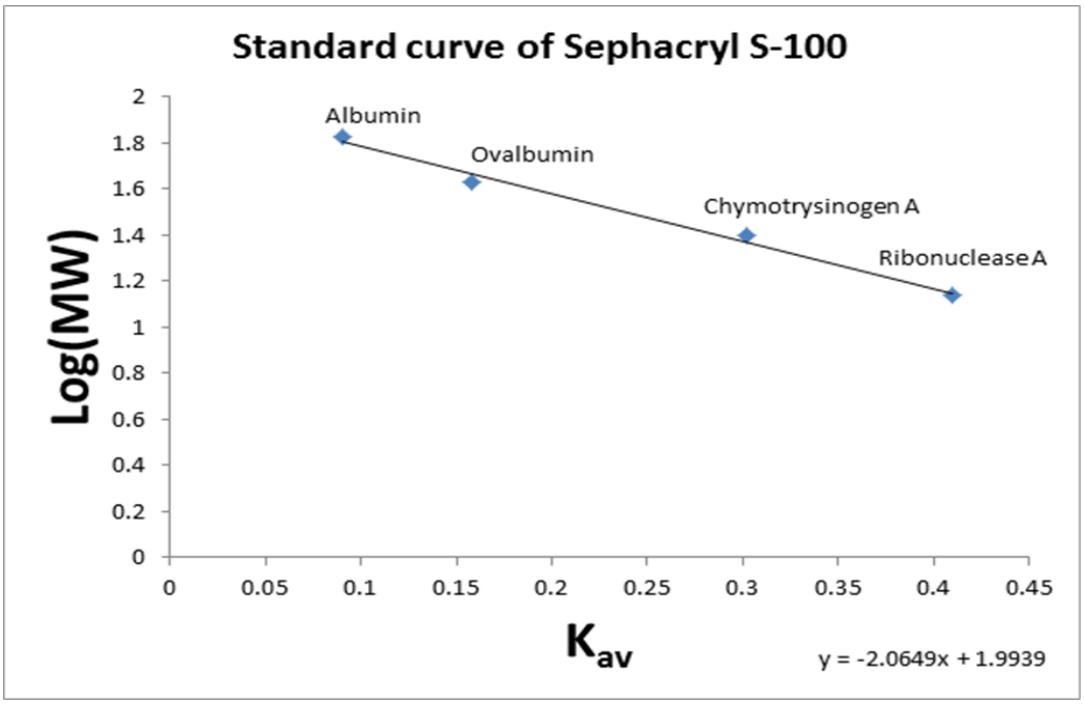
$\{u_j^2\}$ = the mean-square displacement of the atom 'j' from its rest position.

r.m.s.d. from ideal values

A measure of how well the final crystallographic model conforms to expected values of bond lengths and bond angles. A high quality crystallographic model has r.m.s.d. values lower than 0.02 Å for bond lengths and lower than 4 degrees for bond angles.

Appendix E.

Molecular Mass Standard Curve for Size-exclusion Chromatography



Column: HiPrep 26/60 Sephacryl S-100 High Resolution Column attached to the AKTAprime system

Buffer: 20mM Tris-HCl (pH 8.0), 100mM NaCl **Temperature:** 4°C

Flow Rate: 1mL/min

Standards: From Amersham Biosciences' LMW Calibration Kit (product ID: 17-0442-01) (See Table A on the next page)

Concentration: 5mg/mL (400µL final volume) were used except for Blue Dextran for which 1mg/mL (1mL final volume).

Three runs were performed, with Blue Dextran and Chymotrypsinogen A run individually. Chymotrypsinogen A was run separately to prevent digestion of the other samples.

Table A. Protein Standards Used for Generating the Above Standard Curve.

Standard	Molecular Mass (kDa)	Elution Volume on S-100 (mL)
Ribonuclease A	13.7	189
Chymotrysinogen A	25.0	165
Ovalbumin	43.0	133
Albumin	67.0	118
Blue Dextran 2000	2000	98*

* The elution volume of Blue Dextran was used to determine the void volume (V_o = the volume of the mobile phase in the column) as its size is much larger than what this matrixcan retain.

Molecular Mass Calculation Based on the Elution Volume

K_{av} values were calculated as follows:

$$K_{av} = \frac{V_e - V_o}{V_t - V_o} \qquad \textit{Equation \#1}$$

Where $\mathbf{V_o}$ (void volume) was 98mL and $\mathbf{V_t}$ (column volume) was 320mL. $\mathbf{V_e}$ was the elution volume.

The resulting K_{av} vs. Log(MW) plot produced the equation Log(MW) = -2.06(Kav) + 1.99, which can be rewritten as:

$$MW = 10^{(-2.06 \times K_{av} + 1.99)} \qquad \textit{Equation \#2}$$

For each protein sample loaded on size exclusion chromatography, the K_{av} value was calculated based on the elution volume using Equation #1. Then the measured molecular weight was calculated using Equation #2.

Appendix F.

Size-Exclusion Chromatograms of BamE under Various Buffer Conditions

Please refer to Chapter 5 (section 5.2) for materials and methods.

For all following chromatograms,

Column: Superdex 75 column (GE Health Care)

Flow Rate: 0.5 mL/min

Void Volume: 7.7 mL

Buffer: Various buffers used are given in the figure legends.

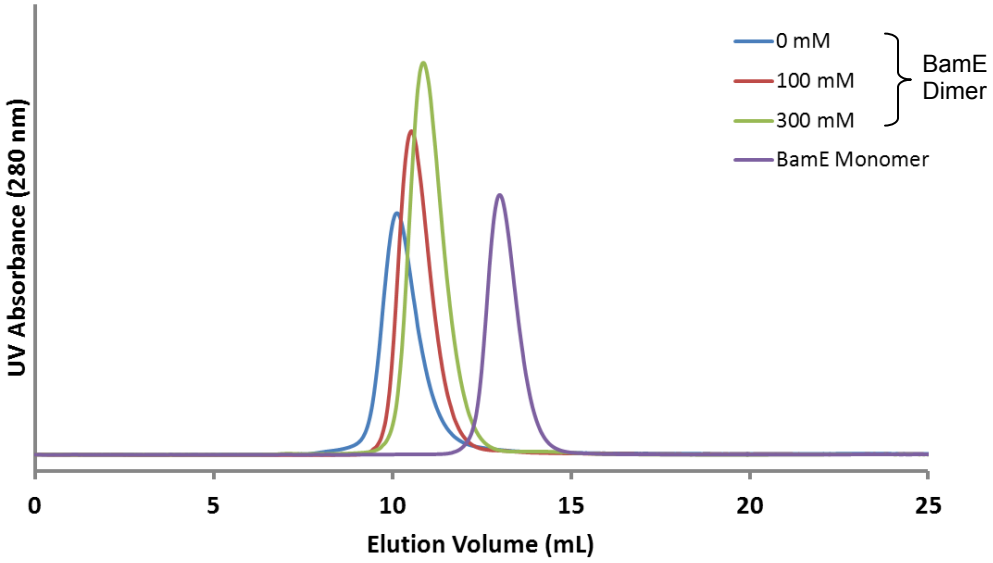


Figure F1. Size-exclusion Chromatography of BamE Dimer under Various Salt Concentrations. Overlaid size-exclusion chromatograms of BamE dimers under various salt concentrations (blue, red and green) and also that of BamE monomer are shown. Previously purified BamE dimer (in 20 mM Tris-HCl pH 8.0, 100 mM NaCl) was separated into three different aliquots, and buffer-exchanged to the same Tris buffer but containing varying salt (NaCl) concentrations as shown on the figure using an Amicon ultra-centrifugal filter device (Millipore). BamE monomer (purple; shown for comparison) was run in buffer containing 20 mM Tris-HCl pH 8.0, 100 mM NaCl). There was no observed dissociation of dimers under the varying salt concentration conditions.

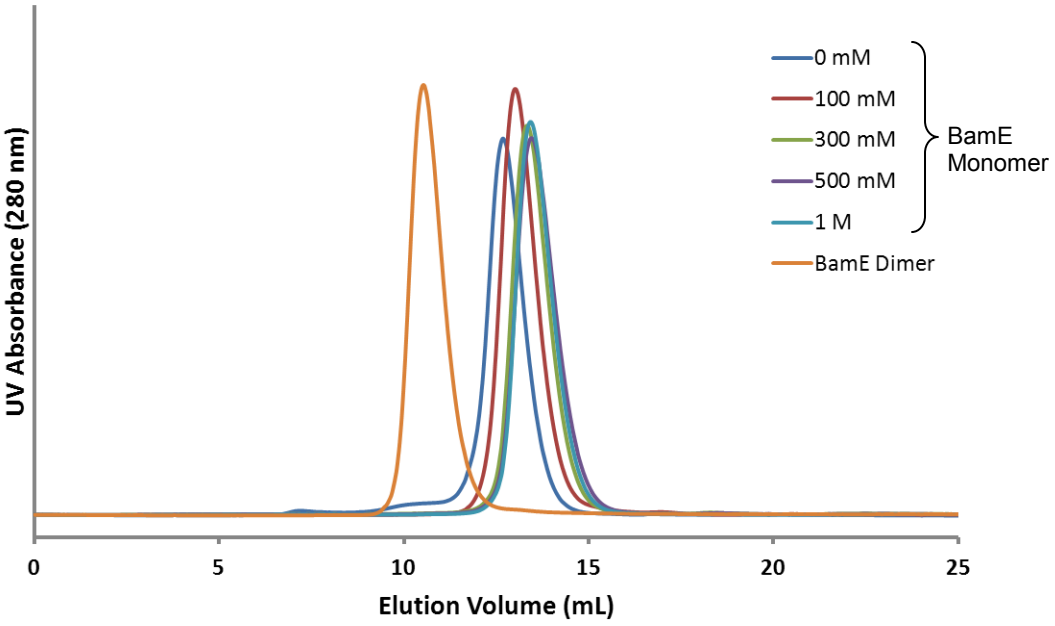


Figure F2. Size-exclusion Chromatography of BamE Monomer under Various Salt Concentrations. Overlaid size-exclusion chromatograms of BamE monomers under various salt concentrations (blue, red, green, purple and light blue) and also that of BamE dimer (orange) are shown. Previously purified BamE monomer (in 20 mM Tris-HCl pH 8.0, 100 mM NaCl) was separated into five different aliquots, and buffer-exchanged to the same Tris buffer but containing varying salt (NaCl) concentrations as shown on the figure using an Amicon ultra-centrifugal filter device (Millipore). BamE dimer (orange; shown for comparison) was run in a buffer containing 20 mM Tris-HCl pH 8.0, 100 mM NaCl). There was no observed dimerization of the monomers under the varying salt concentration conditions.

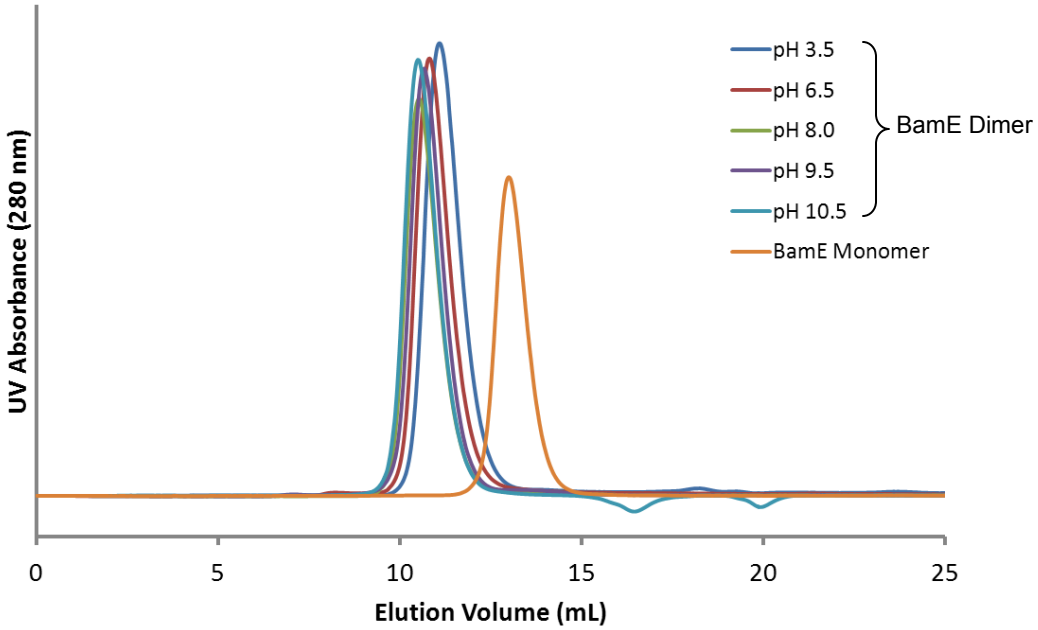


Figure F3. Size-exclusion Chromatography of BamE Dimer under Various pH Conditions. Overlaid size-exclusion chromatograms of BamE dimers under various pH conditions (blue, red, green, purple and light blue) and also that of BamE monomer (orange) are shown. Previously purified BamE dimer (in 20 mM Tris-HCl pH 8.0, 100 mM NaCl) was separated into five different aliquots, and buffer-exchanged to appropriate buffer as shown on the figure using an Amicon ultra-centrifugal filter device (Millipore). All buffers contained 100 mM NaCl, and 20 mM of CH_3COONa, MES, Tris-HCl or CAPS for pH 3.5, 6.5, 8.0, 9.5 and 10, respectively. BamE monomer (orange; shown for comparison) was run in buffer containing 20 mM Tris-HCl pH 8.0, 100 mM NaCl). There was no observed dissociation from dimeric to monomeric form under the varying pH conditions.

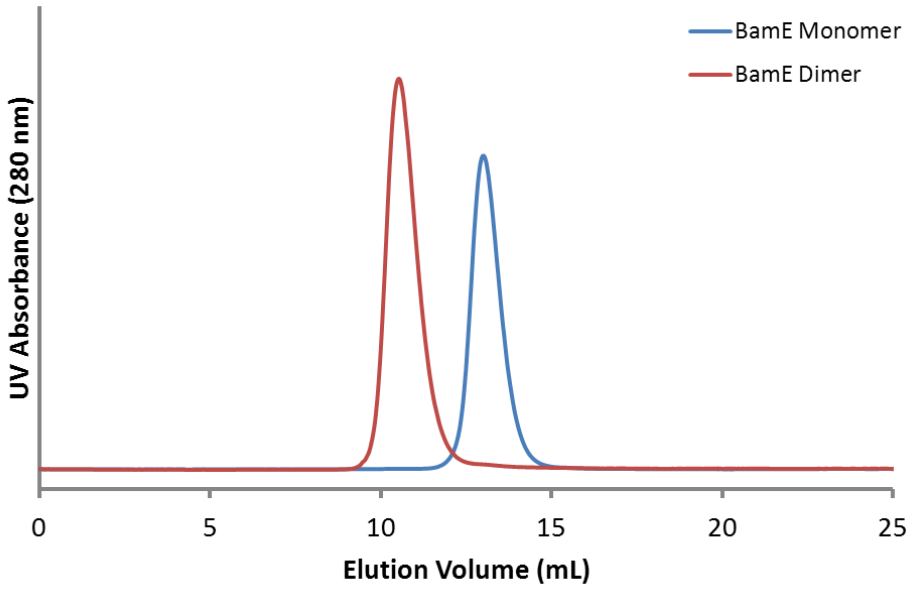


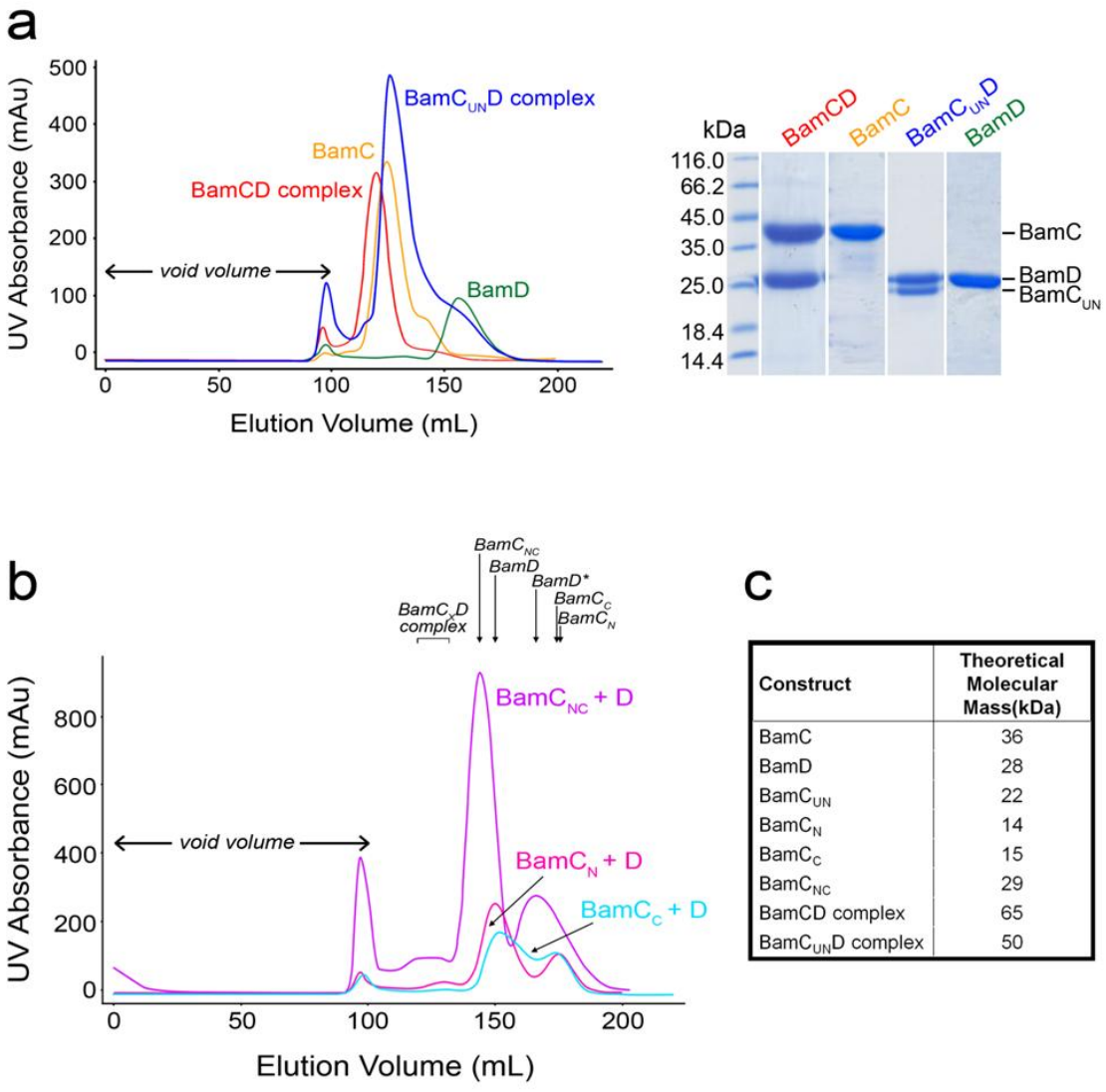
Figure F4. Size-exclusion Chromatography of BamE in Presence of Detergent

Overlaid size-exclusion chromatograms of BamE monomer (blue) and dimer (red) run in buffer containing 20 mM Tris-HCl, pH 8.0, 100 mM NaCl, and 0.01% DDM are shown. There was no observed dissociation from dimeric to monomeric form, or *vice versa*.

Appendix G.

Domain Truncation Mutants of BamC and their Interactions with BamD

Please refer to Chapter 7 (section 7.2) for materials and methods.



Construct	Theoretical Molecular Mass(kDa)
BamC	36
BamD	28
BamC$_{UN}$	22
BamC$_N$	14
BamC$_C$	15
BamC$_{NC}$	29
BamCD complex	65
BamC$_{UN}$D complex	50

Figure G. (legend on next page)

Figure G. Analytical Size-exclusion Chromatography of BamC Constructs.

(a) (Left) Overlaid size-exclusion chromatograms of BamC (orange), BamD (green), BamCD complex (red) and $BamC_{UN}D$ complex (blue) showing that when purified individually, BamC and BamD are monomers. (Right) SDS-PAGE analysis further supports this by showing the presence of one single band in the BamC or BamD peaks, while two bands in the complex peaks. (b) Overlaid size-exclusion chromatograms of samples containing $BamC_{NC}$ +D (purple), $BamC_N$ + D (magenta) and $BamC_C$ + D (cyan). Arrows above the peaks represent the regions where the specified protein is expected to elute. $BamC_XD$ refers to any BamCD complex with either of the BamC constructs (BamC, $BamC_{UN}$, $BamC_{NC}$, $BamC_N$, and $BamC_C$). Since neither of the BamC constructs in this figure had the unstructured N-terminus, they failed to form the BamCD complex. (*Note: the second BamD peak refers to untagged BamD where the N-terminal hexa-histidine tag had been removed for the $BamC_{NC}$ + D trial. Experiments were conducted with and without the tag and gave similar results.) (c) A list of molecular mases of all the BamC and BamD constructs used in this study, as well as those of BamCD and $BamC_{UN}D$ complexes.

Appendix H.

Electron Density Maps of the Solved Crystal Structures

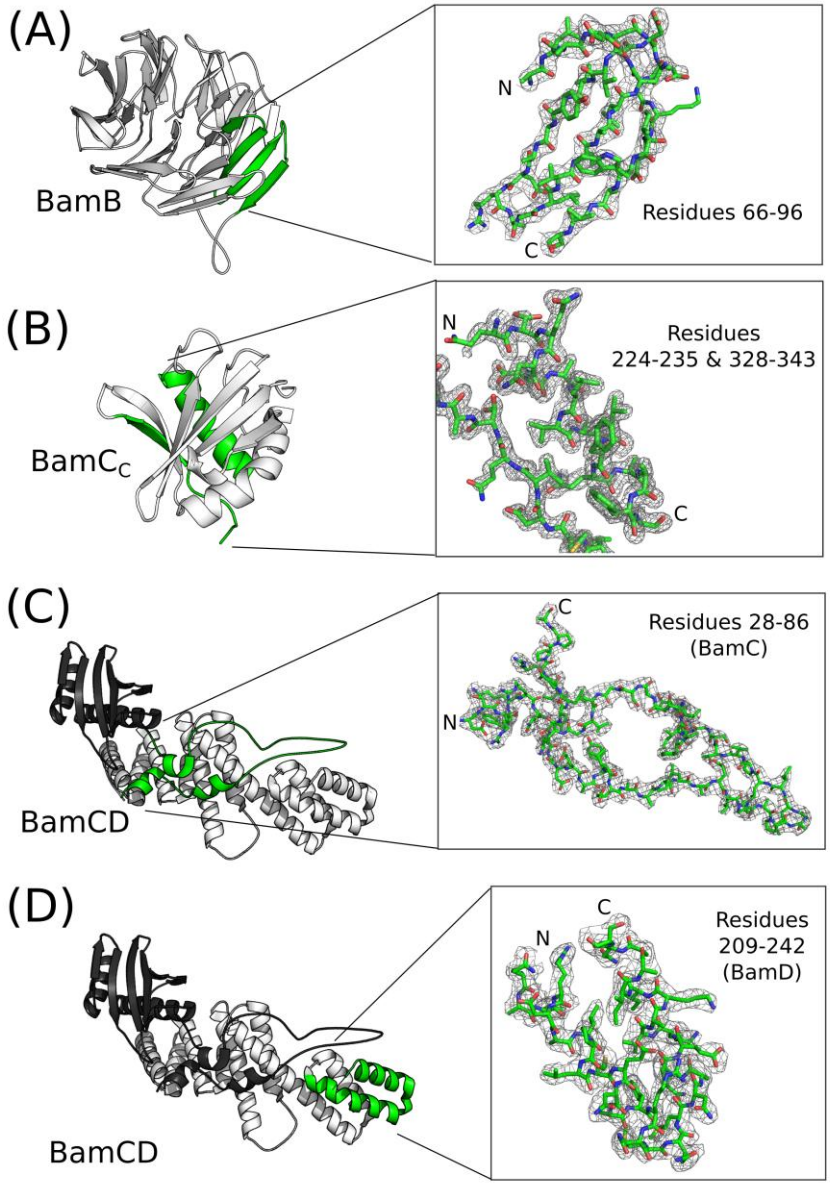


Figure H1. Electron Density Maps of the Solved Crystal Structures

(A-D) Crystal structures solved and the corresponding $2F_o$-F_c electron density maps (contoured at 1 sigma) of selected regions (coloured magenta in A and B; coloured red in C and D) of each protein are shown. The regions for which the electron density maps are shown are represented as a stick model. The PDB IDs for the crystal structures shown in the figure are as follows: BamB (3P1L), $BamC_C$ (3SNS), BamC (3TGO; chain C) and BamD (3TGO; chain A).