The protein-protein interactions involved in the periplasmic components of the β-barrel assembly machinery (BAM) complex of *Escherichia coli*

by

Suraaj Kaur Aulakh

B.Sc. (Hons.), Simon Fraser University, 2009

Thesis Submitted in Partial Fulfillment
of the Requirements for the Degree of
Master of Science

in the
Department of Molecular Biology and Biochemistry
Faculty of Science

© Suraaj Kaur Aulakh 2012

SIMON FRASER UNIVERSITY

Summer 2012

All rights reserved.
However, in accordance with the *Copyright Act of Canada*, this work may be reproduced, without authorization, under the conditions for "Fair Dealing." Therefore, limited reproduction of this work for the purposes of private study, research, criticism, review and news reporting is likely to be in accordance with the law, particularly if cited appropriately.

Approval

Name: **Suraaj Kaur Aulakh**

Degree: **Master of Science**

Title of Thesis: ***The protein-protein interactions involved in the periplasmic components of the β-barrel assembly machinery (BAM) complex of Escherichia coli***

Examining Committee:

Chair: **Dr. Lisa Craig**
Associate Professor
Department of Molecular Biology and Biochemistry

Dr. Mark Paetzel
Senior Supervisor
Associate Professor
Department of Molecular Biology and Biochemistry

Dr. Christopher T. Beh
Supervisor
Associate Professor
Department of Molecular Biology and Biochemistry

Dr. Jack N. Chen
Supervisor
Associate Professor
Department of Molecular Biology and Biochemistry

Dr. Peter J. Unrau
Internal Examiner
Associate Professor
Department of Molecular Biology and Biochemistry

Date Defended/Approved: June 6, 2012

Partial Copyright Licence

The author, whose copyright is declared on the title page of this work, has granted to Simon Fraser University the right to lend this thesis, project or extended essay to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users.

The author has further granted permission to Simon Fraser University to keep or make a digital copy for use in its circulating collection (currently available to the public at the "Institutional Repository" link of the SFU Library website (www.lib.sfu.ca) at http://summit/sfu.ca and, without changing the content, to translate the thesis/project or extended essays, if technically possible, to any medium or format for the purpose of preservation of the digital work.

The author has further agreed that permission for multiple copying of this work for scholarly purposes may be granted by either the author or the Dean of Graduate Studies.

It is understood that copying or publication of this work for financial gain shall not be allowed without the author's written permission.

Permission for public performance, or limited permission for private scholarly use, of any multimedia materials forming part of this work, may have been granted by the author. This information may be found on the separately catalogued multimedia material and in the signed Partial Copyright Licence.

While licensing SFU to permit the above uses, the author retains copyright in the thesis, project or extended essays, including the right to change the work for subsequent purposes, including editing and publishing the work in whole or in part, and licensing other parties, as the author may desire.

The original Partial Copyright Licence attesting to these terms, and signed by this author, may be found in the original bound copy of this work, retained in the Simon Fraser University Archive.

Simon Fraser University Library
Burnaby, British Columbia, Canada

revised Fall 2011

Abstract

The β-barrel assembly machinery (BAM) complex plays the essential role of folding and inserting outer membrane proteins (OMPs) into the outer membrane of Gram-negative bacteria. In *Escherichia coli*, the BAM complex is comprised of five proteins: BamA, BamB, BamC, BamD, and BamE. This thesis project investigates the interactions between the periplasmic components of the BAM complex by analyzing complex formation using gel-filtration chromatography. Results from the interaction studies have identified the unstructured N-termini of BamC and BamE as requirements for BamCDE subcomplex formation. Furthermore, $BamA_{POTRA}$ was shown to form stable $BamA_{POTRA}$-BamB and $BamA_{POTRA}$-BamD-BamE complexes, but was unable to form the latter complex in the presence of BamC. Together these results provide a model for how the proteins assemble and suggest that the complex is dynamic. By understanding how the BAM components come together brings us one step closer to determining their individual roles and how the BAM complex may function overall.

Keywords: outer membrane; outer membrane protein; lipoprotein; β-barrel assembly machinery; gel-filtration chromatography; protein-protein interactions

To my family

Acknowledgements

I would like to thank Dr. Mark Paetzel for taking me in as a member of the BAM team, which gave me the opportunity to conduct research in a competitive field and led to multiple publications. I am also grateful that he allowed me to bring a camera and tripod to the lab for non-thesis related projects.

I would like to thank Dr. Christopher Beh and Dr. Jack Chen for being on my supervisory committee, and for their valuable suggestions and input over the years. I am also thankful to Dr. Peter Unrau and Dr. Lisa Craig for agreeing to serve as the Internal Examiner and Chair for my defence.

A special thank you goes to Kelly Kim for being a great mentor and BAM partner. Thank you to the past and present members of the Paetzel Lab who have helped develop me into the researcher I am today: Dr. Jae Lee, Dr. Dave Oliver, Yun Luo, Apollos Kim, Ivy Chung, Charlie Stevens, Alison Li, Sung Nam, Dan Chiang, Linda Zhang, Cindy Li, and Minfei Fu. Thank you to Deidre de Jong-Wong for being more than just a lab manager, and for being a great teacher with a lot of patience. I would also like to thank Wendy Tan, Paul Joseph Luczon, and Jonathan Tan who dedicated their undergraduate ISS semester toward studying the BAM complex.

Another thanks to Dan, Charlie, Kelly, Linda, and Cindy for being a part of those projects involving the camera and tripod mentioned above. And thank you to the members of the SFU community who supported and promoted these works in their classroom.

This research was supported by the CIHR Frederick Banting and Charles Best Canada Graduate Scholarship and the Pacific Century Graduate Scholarship. I am thankful to the judging committees for selecting me as a recipient of these awards.

And finally I would like to thank my family for their continuous support, love, and encouragement over the years.

Table of Contents

List of Tables

List of Figures

List of Acronyms

BAM	β-barrel assembly machinery
Da	Dalton (unit for molecular mass; 1 Dalton = 1 gram/mole)
DDM	n-dodecyl-β-maltoside (detergent used here for crystallization studies)
IM	inner membrane
K_{av}	the partition coefficient of a protein on gel-filtration chromatography
kDa	kiloDalton (unit for molecular mass; 1 kiloDalton = 1000 Daltons)
LPS	lipopolysaccharide (lipid found in the outer leaflet of the outer membrane of Gram-negative bacteria)
M	molecular mass (measured in Daltons or kiloDaltons)
mAu	milli-absorbance unit
MBP	maltose binding protein
Ni-NTA	nickel-nitrilotriacetic acid agarose (resin used for nickel affinity chromatography)
NMR	nuclear magnetic resonance
OM	outer membrane
OMP	outer membrane protein
PDB	Protein Data Bank (database of 3D coordinates of known protein structures)
POTRA	polypeptide transport associated (motif found in BamA)
SAM	sorting and assembly machinery (the BAM complex homologue in mitochondria)
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
TOC	translocon at the outer envelope of chloroplasts (the BAM complex homologue found in chloroplasts)
TPR	tetratricopeptide repeat (motif found in BamD)
V_e	elution volume (volume at which a protein emerges from gel-filtration chromatography; proteins resolved on a column suitable for its size should elute after the void volume, but before the total column volume)
V_o	void volume (volume of the mobile phase of the column; proteins too large to enter the gel-filtration matrix of a specific column elute in the void volume)
V_t	total column volume (volume of the packed resin including the mobile phase)

1. Introduction

Portions of this chapter have been published in:

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

Kim, K.H., Aulakh, S. and Paetzel, M. (in press). Outer membrane protein biosynthesis: transport and incorporation of proteins (in)to the OM bilayer. Bacterial Membranes: Structural and Molecular Biology. UK: Horizon Scientific Press.

1.1. Bacterial outer membranes

One of the most recognizable features of Gram-negative bacteria is the outer membrane. Whereas the cell wall of Gram-positive bacteria consists of a plasma membrane and a thick peptidoglycan layer, Gram-negative bacteria have an inner membrane (IM) an outer membrane (OM), and a thin peptidoglycan layer in the periplasmic space between the IM and the OM (Figure 1.1) (Silhavy et al., 2010). The most obvious advantage of the OM is its ability to serve as a physical barrier against the extracellular environment. The asymmetric bilayer of the OM is different from the IM phospholipid bilayer. The OM outer leaflet is composed of glycolipids, mainly lipopolysaccharide (LPS), which is a known endotoxin that triggers an immune response in humans (Raetz and Whitfield, 2002). The inner leaflet of the OM contains the phospholipids, primarily phosphatidylethanolamine and phosphatidylglycerol (Morein et al., 1996).

Aside from lipids, about half of the OM's mass consists of proteins (Koebnik et al., 2000). These proteins generally fall into two categories: lipoproteins, which are

found in the periplasmic space and are covalently anchored by an amino-terminal cysteine residue to the inner leaflet phospholipids; and transmembrane proteins, which span the entire length of the bilayer (Bos and Tommassen, 2004; Koebnik et al., 2000). The latter category is often referred to as outer membrane proteins (OMPs), and current structural information shows most OMPs to have a β-barrel fold (Fairman et al., 2011). The two exceptions found so far in *Escherichia coli* are Wza, which forms an α-helical barrel, and the type IV outer membrane secretion complex, which forms a large tetradecamer (Collins and Derrick, 2007; Dong et al., 2006; Chandran et al., 2009). Thus the term OMP specifically refers to integral β-barrel proteins found in the OM.

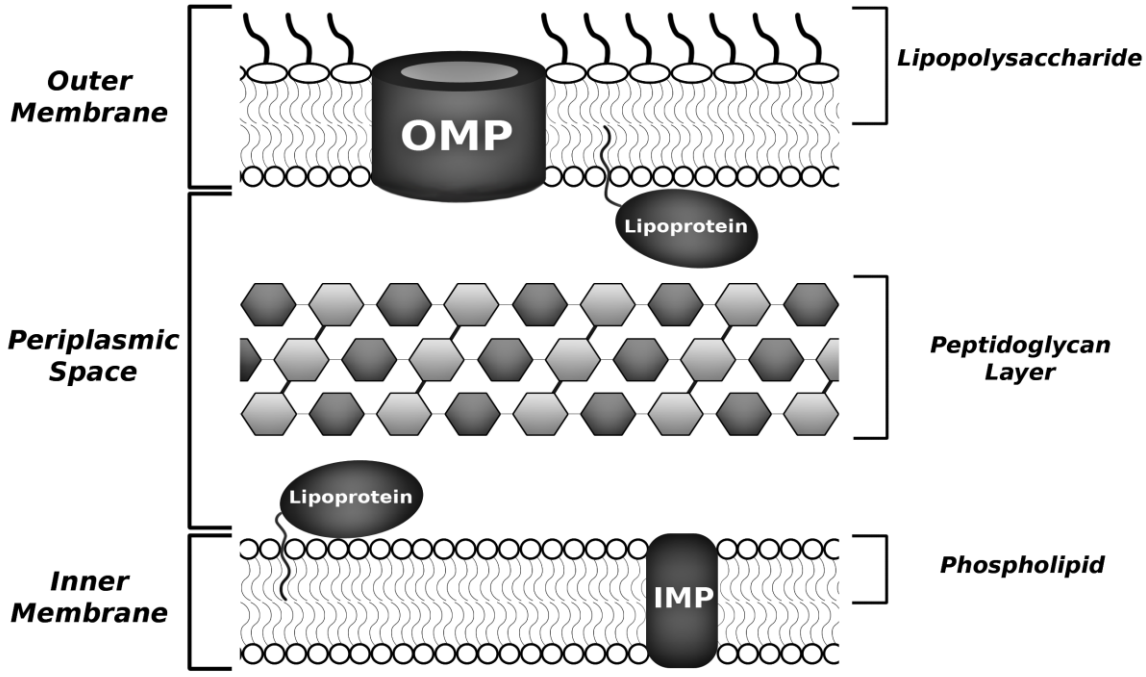

Figure 1.1 The cell wall of a Gram-negative bacterium

The cell envelope of a Gram-negative bacterium consists of the periplasm containing the peptidoglycan layer surrounded by two membranes: the inner membrane (IM) and outer membrane (OM). Unlike the IM, which is mostly made up of phospholipids, the OM is an asymmetric lipid-bilayer containing lipopolysaccharides (LPS) on its outer leaflet. The membrane associated proteins of the Gram-negative bacterial cell envelope can be divided into three categories: outer membrane proteins (OMPs), inner membrane proteins (IMPs), and lipoproteins which are found in the periplasm but are anchored to the periplasmic leaflets of the IM and the OM.

Despite the structural similarities of these β-barrel OMPs, the functions are quite diverse. Some OMPs are involved in the flow of nutrients across the OM, some function as exporters of virulence factors, and some help remove toxins, including antibiotics, out of the cell (Koebnik et al., 2000). The first section of this chapter takes a closer look into the different types of OMPs and their respective roles. However, in order for these proteins to be functional, they must first be exported from the cytosol, across the IM and periplasmic space, and correctly assembled in the OM. The final folding and assembly step is now believed to be carried out by a multi-protein system known as the β-barrel assembly machinery (BAM) complex. Improper assembly by or absence of the BAM complex can lead to defects in the OM, eventually leading to cell death (Gentle et al., 2004). Homologues of the BAM complex can also be found in mitochondria and chloroplasts (Schleiff and Soll, 2005; Walther et al., 2009).

With antibiotic resistance on the rise, researchers are in pursuit of finding novel drug targets, especially in the case of Gram-negative bacteria. Most antibiotics today are directed to cellular processes in the periplasmic space or further inside the cell, for which they must first pass the OM (Delcour, 2009). Thus, understanding the structure and function of the BAM complex not only helps to answer the question of how OMPs get into the OM, but also provides a series of essential protein-protein interactions that could be new targets for antibiotics. As explained later in the chapter, this Master's thesis focuses on dissecting the interactions within the periplasmic components of the BAM complex to study how they assemble into a machine that plays a vital role in cell maintenance and survival.

1.2. Outer membrane proteins

OMPs are β-barrel proteins found in the outer membrane of Gram-negative bacteria, mitochondria and chloroplasts. They usually have an even number of anti-parallel β-strands ranging from 8 to 24, arranged in a barrel shape with a hydrophobic exterior, which allows it to sit in the lipid bilayer (Koebnik et al., 2000; Schulz, 2000; Wimley, 2003). Despite the consistent β-barrel domain, OMPs differ in their number of β-strands, oligomeric state, and in their exoplasmic and periplasmic regions, resulting in a diverse group of proteins with distinct functions (Figure 1.2; Table 1.1). These functions

can be classified into six general categories, which are summarized below with specific examples from *E. coli*.

1.2.1. Non-specific porins

Despite its general use today, the term “porin” was originally coined to describe a class of OMP channels involved in non-specific diffusion of solutes (Nakae, 1976; Nikaido, 1994, 2003). Porins are usually found as homotrimers of 16-stranded β-barrel subunits, and the hydrophilic interior allows the transport of hydrophilic molecules smaller than 600 Da into the cell (Delcour, 2003; Nikaido, 1994). While defined as “non-specific,” porins can be selective in terms of the size and charge of the molecule. Examples of porins include OmpC (preference for small, positive solutes), OmpF (preference for large, positive solutes), and PhoE (preference for negative solutes) (Nikaido, 2003).

1.2.2. Substrate-specific channels

Aside from porins, there are other OMPs that also act as pores for larger solutes to enter through. However, these pores are specific for their substrates, and can be referred to as channels rather porins (Nikaido, 2003). For example, BtuB (Figure 1.2) is specific for the uptake of vitamin B12, FhuA for iron, LamB for maltose and other sugars, and Tsx for nucleosides. Some of these channels (ex. LamB and Tsx) are independent and allow passive diffusion of solutes upon contact, while others (ex. BtuB and FhuA) are TonB-dependent channels as they require assistance for the active transport of substrates (Nikaido, 1994). The TonB complex is found in the IM and spans the periplasmic space to interact with these latter channels to provide the energy required for uptake of the substrate (Krewulak and Vogel, 2011; Noinaj et al., 2010; Postle and Kadner, 2003). LamB is a homotrimer of 16-stranded β-barrel subunits, and thus it has often been classified as a substrate-specific porin and even been given the name maltoporin. However, to avoid confusion with non-specific porins and to stay consistent with the original definition, here it is classified as a substrate-specific channel (Nakae, 1976; Nikaido, 2003).

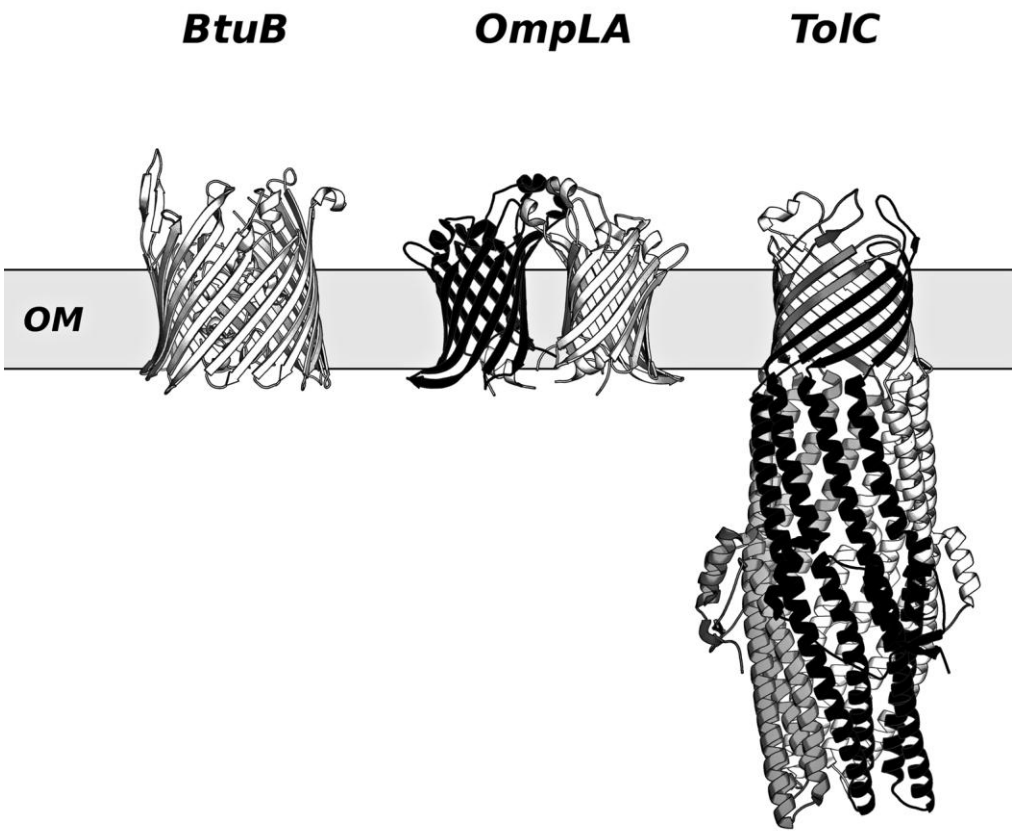

Figure 1.2 Diversity of OMP structures

This figure shows the diversity of structures of outer membrane proteins (OMPs) using examples from *E. coli*. Each OMP prefers a specific oligomeric state such as a monomer (ex. BtuB; PDB: 2GUF), dimer (ex. OmpLA; PDB: 1QD6), or trimer (ex. TolC; PDB: 1EK9). Note that each subunit of OmpLA creates its own β-barrel, while each subunit of TolC contributes to form one β-barrel. Aside from oligomeric state, OMPs can also vary in their exoplasmic and periplasmic regions, as demonstrated by the large periplasmic region of TolC.

Table 1.1 List of currently known E. coli OMPs in the UniProt database

Protein	Length in amino acids	Proposed Function/Category	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	(Bodelón 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	

This table provides a list of currently known OMPs from *E. coli* in the UniProt database. Only those OMPs that were marked as “reviewed” (indicating that their entries have been annotated by an expert) are included here.

1.2.3. Translocons for export of substrates

In addition to the import of molecules, the OM allows certain molecules to travel outside the cell as well. For example, many proteins are synthesized in the cytosol, but need to be secreted outside the cell for proper function. In a Gram-negative bacterium, many OMPs can be found playing a role in the different export pathways (Yen et al., 2002). For example, TolC is an OMP involved in the Type I secretion pathway that allows the export of proteins without the use of the SecYEG translocon at the IM (Nikaido, 2003). TolC forms a unique trimer structure, with three monomeric units contributing 4 strands each to form a single 12-stranded β-barrel (Figure 1.2). In addition to the barrel, a long periplasmic domain is present that forms an α-tunnel, which allows interaction with other components of the pathway found in the periplasm and IM (Koronakis et al., 2000). Aside from protein export, TolC is also known to export small molecules and drugs, contributing to antibiotic resistance (Zgurskaya et al., 2011). Other examples of OMPs acting as translocons are those involved in the two-partner secretion pathway (a component of the Type V secretion pathway). In this pathway, the secreted protein requires a specific OMP translocon for export. A famous example is FhaC from *Bordetella pertussis*, which is involved in the export of filamentous hemagglutinin, an adhesin secreted during infection. FhaC has a 16-stranded β-barrel at the C-terminus, with a periplasmic region at the N-terminus (Clantin et al., 2007). A homologue of FhaC, and another member of the two-partner secretion system, is *E. coli* BamA. However, as BamA plays an important structural role by assembling OMPs in the OM rather than secreting extracellular proteins, it has been categorized under “Structural OMPs” (see Section 1.2.6 below).

1.2.4. Autotransporters

Some proteins destined for transport outside the cell are able to be secreted 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 known as autotransporters. The C-terminal transporter domain is generally a 12-stranded barrel, and in many cases found as a monomer (Oomen et al., 2004; van Ulsen, 2011). However, recent research is beginning to debate whether this domain is

actually able to act as a transporter on its own as opposed to simply serving as a membrane anchor at the OM (Bernstein, 2007). Because many of these autotransporter proteins have been shown to be substrates of the BAM complex, it has been proposed that the BAM complex or another recently identified system known as the translocation and assembly module (TAM) complex may assist in the secretion of the passenger protein (Rossiter et al., 2011; Selkrig et al., 2012). The passenger proteins are usually virulence factors secreted by pathogenic bacterial strains. Examples include adhesins such as AIDA-I and Ag43, and proteases such as Hbp and Pet (van Ulsen, 2011). Intimin is another example, which is an attaching and effacing protein (Touzé et al., 2003). Interestingly, intimin does not fall into the classical definition of autotransporters as its passenger domain is at the C-terminus, while the β-barrel transporter is at the N-terminus. The structure of this β-barrel has yet to be solved, and some experiments suggest possible homodimer formation (Bodelón et al., 2009; Touzé et al., 2003).

1.2.5. Enzymes

Aside from transport, there are other functions required to take place at the OM, some of which are carried out by enzymatic OMPs. To date, only three enzymatic OMPs have been discovered in *E. coli*: OmpLA, OmpT, and PagP (Bishop, 2008). OmpLA (phospholipase A) is a 12-stranded β-barrel that hydrolyzes phospholipids in the OM (Figure 1.2). Its active site is located in the LPS-containing outer leaflet, where it can detect the presence of phospholipids that disrupt the asymmetry of the OM. OmpLA is found as a monomer, and dimerizes upon substrate presentation to become active (Dekker et al., 1997). OmpT is a protease that has a 10-stranded β-barrel fold 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; Vandeputte-Rutten et al., 2001). The oligomeric state of OmpT is unknown, but early gel-filtration studies suggest a possible pentamer formation (Sugimura and Nishihara, 1988). Interestingly, some strains of *E. coli* have another protease in the OM known as OmpP, which is a homologue of OmpT found encoded on the F-plasmid (Bishop, 2008). Finally, PagP is an 8-stranded β-barrel that transfers a palmitate chain from a phospholipid in the OM inner leaflet to the Lipid A component of a LPS molecule in the outer leaflet. Because of its role in maintaining this essential piece of the OM, PagP is being studied as a potential drug target (Bishop, 2005).

1.2.6. Structural OMPs

This final category classifies OMPs that contribute to the formation and integrity of the cell wall structure. For example, Mipa is an OMP that acts as a scaffolding protein to mediate the interaction between specific enzymes involved in peptidoglycan synthesis in the periplasm (Vollmer et al., 1999; Vollmer and Bertsche, 2008). BamA is a component of the BAM complex which catalyses the insertion of OMPs into the OM. It has an N-terminal periplasmic region and a C-terminal β-barrel (more discussion on this protein in Sections 1.3.4 and 1.5). LptD is an OMP involved in the final stages of LPS assembly in the OM (Okuda and Tokuda, 2011). OmpX is a part of a family of OMPs that are involved in adhesion and entry into host cells. Its structure has been solved and shows the presence of an 8-stranded β-barrel (Vogt and Schulz, 1999). And finally, fimbrial usher proteins can also be categorized here as they are involved in the formation of the pili subunits found on the exterior of the bacterium. For example, FimD is a 24-stranded β-barrel that serves as an usher to transport and polymerize subunits of the Type I pili, with the help of the periplasmic chaperone FimC (Phan et al., 2011).

1.3. Outer membrane protein biogenesis

In eukaryotes and prokaryotes, most proteins are synthesized in the cytosol with almost half of them requiring membrane targeting for proper function (Schatz and Dobberstein, 1996). In eukaryotes, these proteins could be destined for an organelle membrane or lumen, the cytoplasmic membrane, or secreted outside the cell. For prokaryotes, the options are limited to the IM, periplasm, OM, or extracellular space. In order for these proteins to function correctly, they must reach their target membrane or compartment efficiently. The information for their delivery is encoded in the primary sequence of the protein, usually found at the N- or C- terminus. The overall OMP biogenesis process is summarized below and also in Figure 1.3.

1.3.1. Post-translational targeting to the inner membrane

In prokaryotes, to ensure proper targeting of proteins destined for the IM or further, an N-terminal signal directs them to the IM. The signal sequence of transmembrane IM proteins (IMPs) is co-translationally recognized and directed to the

IM by the signal recognition particle (SRP). In many cases, this signal is encoded within the N-terminal transmembrane region of the IMP, and thus a separate cleavable signal sequence is not needed (Dalbey et al., 2011).

In contrast, targeting of precursor OMPs (subsequently referred to as pre-OMPs) in Gram-negative bacteria occurs post-translationally, involving a cleavable N-terminal signal sequence with the following features: a positively charged N-terminal region, a hydrophobic region, and a polar C-terminal region containing the cleavage site (Gierasch, 1989; von Heijne, 1990). SRP has a preference for hydrophobic regions as its association with a signal sequence strengthens with increased hydrophobicity (Valent et al., 1995). Thus, SRP preferentially binds to the IMP signal allowing another protein, trigger factor (TF) to bind to the pre-OMP signal sequence. TF is a ribosome associate chaperone, which binds to the pre-OMP as it emerges from the ribosome (Driessen and Nouwen, 2008; Ferbitz et al., 2004; Hoffmann et al., 2010). After translation is complete, TF passes the pre-OMP over to the cytoplasmic chaperone SecB, which associates with regions of the protein that are believed to be buried in the natively folded form (Bechtluft et al., 2010; Knoblauch et al., 1999). This ensures that the pre-OMP is transported to the IM in a stable unfolded state, which is required for IM translocation machinery known as the SecYEG complex (Driessen and Nouwen, 2008).

1.3.2. Translocation across the inner membrane

Once at the IM, SecB transfers the pre-OMP to the homodimeric form of SecA that is bound to the SecYEG translocation complex. SecA is an ATPase, and binding of ATP allows SecB to be released from this complex. At this point, SecA begins to thread the pre-OMP through the SecYEG channel by an unknown mechanism; one proposed model suggests an ATP dependent process in which the two-helix-finger domain of SecA acts as a piston to push the pre-OMP through (Cross et al., 2009; Kusters and Driessen, 2011; Zimmer et al., 2008).

The SecYEG translocon (also referred to as the Sec translocon) is a well conserved complex whose homologues are also found in the endoplasmic reticulum (ER) membrane of eukaryotes, where it is known as the Sec61 translocon (Cross et al., 2009; Pohlschröder et al., 1997). As the name suggests, SecYEG is a heterotrimer of

the IMPs SecY, SecE, and SecG, which come together to form a central pore. The diameter of this pore is too small for the pre-OMP polypeptide chain to enter, and thus it must widen and form an open channel for translocation to occur. This could occur simply by conformational changes within one protomer of SecYEG, or by further assembly of the complex into a dimeric or tetrameric form to create a larger channel (Driessen and Nouwen, 2008; du Plessis et al., 2011). Thus, the oligomeric state required for proper *in vivo* translocation is still under debate. The SecYEG translocon is also found associated with another heterotrimeric complex, SecDF-YajC, which is not essential for translocation, but has shown to enhance the activity of SecYEG through an unknown mechanism (Driessen and Nouwen, 2008; du Plessis et al., 2011).

1.3.3. Transport through the periplasmic space

As the pre-OMP begins to emerge into the periplasmic space, the signal peptide is recognized and cleaved by signal peptidase I (SPaseI). SPaseI specifically cleaves after the conserved Ala-X-Ala sequence in the C-terminal region of the signal peptide, releasing the mature OMP into the periplasmic space (Paetzel et al., 2002). From here, the OMP must travel through the dense periplasm to the OM before being folded. Periplasmic chaperones SurA, Skp, and DegP have been shown to associate with OMPs, and are believed to be the major players involved in stabilizing these unfolded proteins during their transport. SurA recognizes unfolded proteins by aromatic residues, a common feature of OMPs, especially at their C-terminus (Bitto and McKay, 2003). It has been shown that aside from the N-terminal signal sequence for IM targeting, OMPs carry another signal on the C-terminus which is recognized by the BAM complex for assembly and insertion at the OM (Robert et al., 2006). On the other hand, Skp recognizes unfolded proteins by their exposed hydrophobic regions (Qu et al., 2007). The trimeric structure of Skp mimics the shape of a jellyfish with three tentacles, and forms a hydrophobic interior serving as a protective cavity for the OMPs (Walton and Sousa, 2004). Similarly, DegP oligomerizes to form large cages that can also protect the OMPs from aggregation or degradation by proteases (Merdanovic et al., 2011). Unlike SurA and Skp, substrate specificity of DegP is broad as it can recognize unfolded, misfolded and mislocalized proteins in the periplasm. In addition to its function as a chaperone, DegP also shows a proteolytic activity at high temperatures (more detailed discussion in Section 1.3.5). Gene knockout studies suggest that Skp and DegP may

function in a separate pathway than SurA, as individual gene absences are viable, but a double s*urA*⁻ and *degP* ⁻ or a double *surA* ⁻ and *skp* ⁻ mutant is synthetically lethal. This also suggests that both pathways play a redundant role, and that at least one pathway must always be functional for cell survival (Rizzitello et al., 2001).

1.3.4. Folding and insertion into the outer membrane

After being transported to the OM, the OMPs are then folded and inserted into the membrane by the BAM complex. In *E. coli*, this multi-protein machinery consists of an OMP known as BamA (previously known as YaeT) and four OM lipoproteins: BamB, BamC, BamD, and BamE (previously YfgL, NlpB, YfiO, and SmpA) (Hagan et al., 2011; Ricci and Silhavy, 2012). BamA has a C-terminal β-barrel domain spanning the membrane and an N-terminal periplasmic region composed of five polypeptide transport associated (POTRA) domains. The absence of BamA results in an accumulation of unfolded proteins in the periplasm, a deformed OM, as well as cell death; hence it is an essential protein found in all Gram-negative bacteria, with homologues also present in mitochondria and chloroplasts (Gentle et al., 2004; Schleiff and Soll, 2005). Out of the four lipoproteins, only BamD is essential and conserved (Malinverni et al., 2006). When a newly synthesized OMP reaches the OM, it is proposed that its C-terminal targeting sequence (also known as the OMP-signal) is recognized by BamD, directing the OMP towards the BAM complex (Robert et al., 2006; Sandoval et al., 2011; Albrecht and Zeth, 2011; Kim et al., 2011a). The exact mechanism of OMP folding and insertion is not well understood yet, and a detailed discussion of current models is provided in Section 1.5.3 of this chapter.

1.3.5. Degradation of OMPs

Protein degradation is an essential component of quality control. Stress on the OMP synthesis pathway such as overproduction can cause OMPs to become misfolded, aggregated, or mislocalized. Fortunately, 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 that eventually leads to a decrease in the expression level of OMPs (Merdanovic et al., 2011). This lowers the stress put on

the OMP synthesis pathway, preventing further mislocalization. Similarly, any damaged or misfolded OMPs are recognized by DegP, which can use its proteolytic function to initiate degradation (Merdanovic et al., 2011).

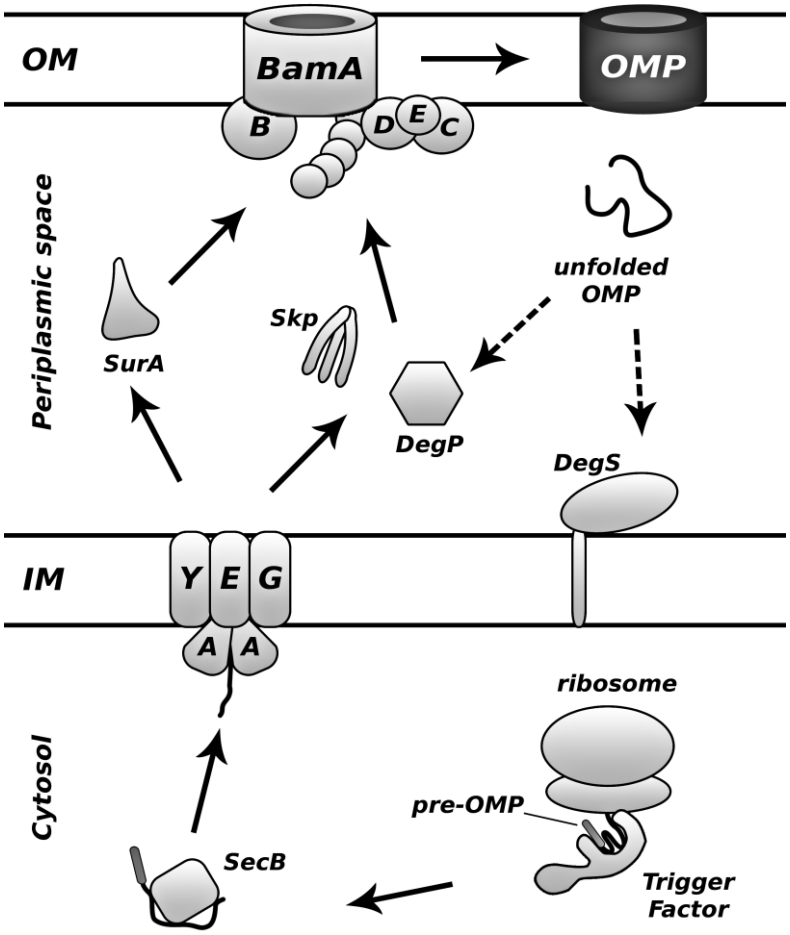

Figure 1.3 OMP biogenesis and degradation

This figure outlines the OMP biogenesis and degradation pathways. The precursor OMP (pre-OMP) contains an N-terminal signal sequence that targets the protein to the IM. This N-terminal signal is bound by Trigger Factor as it emerges from the ribosome during synthesis. Once translation is complete, the pre-OMP is transferred to the cytoplasmic chaperone SecB which keeps it in an unfolded state. The pre-OMP is then released from SecB to an ATP-powered motor protein SecA, which facilitates the translocation of the pre-OMP across the IM via the SecYEG channel. At the IM, the N-terminal signal sequence is removed by signal peptidase I (not shown), and the pre-OMP is released into the periplasm. The protein then takes either the SurA or the Skp/DegP chaperone pathway to travel to the BAM complex of the OM. By an unknown mechanism, the BAM complex mediates the folding and membrane insertion of OMPs. Misfolded or mislocalized OMPs in the periplasm are recognized and degraded by pathways involving DegP or DegS.

1.4. Lipoproteins and their biogenesis

Sections 1.2 and 1.3 provided an introduction to OMPs and their biogenesis as the function of the BAM complex is to assemble this specific category of proteins. However, in *E.coli*, the BAM complex itself is composed of two types of proteins: an OMP (BamA) and four lipoproteins (BamB, BamC, BamD, and BamE). Therefore it is necessary to note the lipoprotein biogenesis pathway since the primary focus of this research project is on the lipoprotein members of the complex (see Section 1.6). Lipoproteins are soluble proteins in the periplasm that are anchored to the inner leaflet of the IM or OM (BamB-E are attached to the OM). This membrane association is made via lipid chains that are covalently attached to the N-terminal cysteine residue of the protein (Okuda and Tokuda, 2011).

Similar to OMPs, lipoproteins are synthesized in the cytosol with an N-terminal signal sequence that directs the pre-lipoprotein to the IM for translocation. One unique feature of this N-terminal signal is the presence of a conserved region known as the "lipobox," which has the sequence Leu-(Ala/Ser)-(Gly-Ala)-Cys (Okuda and Tokuda, 2011; Paetzel et al., 2002). After transport across the IM, the signal sequence of the pre-lipoprotein remains embedded in the membrane, keeping the protein attached to the IM (Figure 1.4). The cysteine residue from the lipobox then becomes covalently linked to a diaclyglycerol group by the prolipoprotein diacylglyceryl transferase (Lgt). The signal sequence is then cleaved by signal peptidase II (also referred to as LspA) to generate the mature lipoprotein with the modified cysteine residue forming the new N-terminus (Paetzel et al., 2002). The final step of the modification is the addition of another fatty acid chain to the amino group of the cysteine, which is done by the apolipoprotein N-acyltransferase (Lnt). If the final lipoprotein is destined for the OM, it is then directed to the localization of lipoprotein (Lol) pathway for transport (Figure 1.4) (Okuda and Tokuda, 2011; Silhavy et al., 2010).

Whether a lipoprotein is to be sent to the OM or be retained at the IM is determined by the residue that immediately follows the N-terminal cysteine. An aspartate at this position is the "Lol avoidance signal" which will keep the protein at the IM, while other residues at this position will be directed to the Lol pathway and reach the OM. Since this residue occupies the +2 position of the mature lipoprotein, this rule is often

referred to as the “+2 rule” (Okuda and Tokuda, 2011; Silhavy et al., 2010). The Lol pathway involves three main steps. The first step is carried out by the LolCDE complex at the IM which recognizes and binds to the lipoproteins without the Asp at position +2. LolCDE then passes the protein to LolA, a periplasmic chaperone that carries the protein to OM. Here, the lipoprotein is transferred to LolB, which assists in its anchoring to the inner leaflet of the OM (Figure 1.4) (Okuda and Tokuda, 2011).

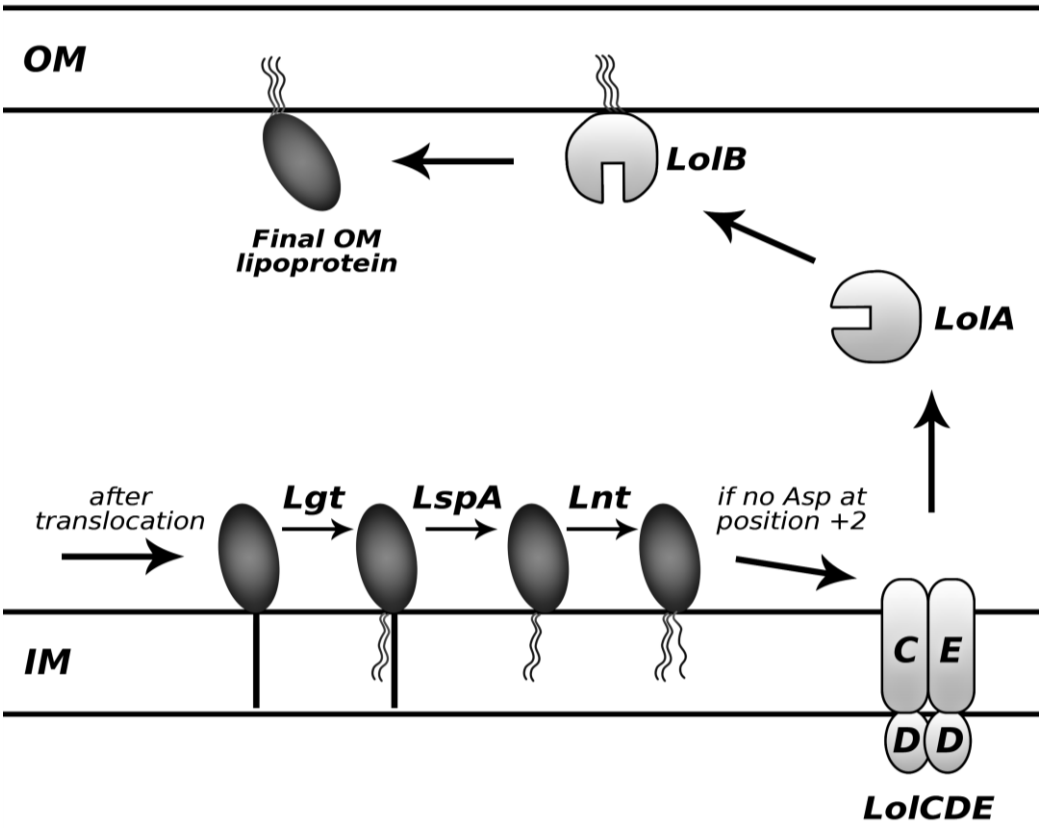

Figure 1.4 Lipoprotein maturation and the Lol pathway

This figure shows the lipoprotein maturation steps and the localization of lipoprotein (Lol) pathway. After translocation across the IM, the pre-lipoprotein (black oval) has its signal sequence (black line) still embedded in the IM. The prolipoprotein diacylglyceryl transferase (Lgt) adds a diacylglycerol group to the cysteine residue which becomes the new N-terminus after cleavage of the signal sequence by LspA (or signal peptidase II). Then, another fatty acid chain is added to this cysteine by the apolipoprotein N-acyltransferase (Lnt). Lipoproteins destined for the IM usually contain an aspartate residue at position +2 (the residue immediately following the cysteine), and avoid entering the Lol pathway. Lipoproteins destined for the OM contain another residue at position +2 and are directed to the Lol pathway, which first involves binding to the LolCDE complex at the IM. Then, the lipoprotein is transferred to the periplasmic chaperone LolA, which carries the protein to LolB at the OM. Finally, the OM lipoprotein becomes anchored to the inner leaflet of the OM.

1.5. The β-barrel assembly machinery (BAM) complex

1.5.1. Discovery and early studies

In 2003, Voulhoux et al. discovered that the Omp85 protein in *Neisseria meningitidis* plays an important role in the assembly and insertion of OMPs into the OM (Voulhoux et al., 2003). Omp85 was known to be a highly conserved protein throughout Gram-negative species, with homologues also existing in eukaryotes. Earlier studies suggested Omp85 to be involved in lipid transport and LPS assembly, however later research has refuted that theory and strongly supports its role in OMP assembly (Genevrois et al., 2003; Voulhoux et al., 2003). Voulhoux et al. demonstrated that in Omp85 depleted strains misfolded OMPs would accumulate because they could not be embedded into the OM, while the absence of Omp85 resulted in cell death (Voulhoux et al., 2003). Similar observations made by other researchers, and identification of homologues in eukaryotic mitochondria and chloroplasts provided further support for Omp85 to play this role (Reumann et al., 1999; Kozjak et al., 2003; Gentle et al., 2004). *Escherichia coli* has now become a popular model for studying Omp85, where it was initially known as YaeT and later renamed to BamA. Early co-purification experiments began to discover other components involved in OMP assembly, and in *E. coli* these accessory proteins were identified to be lipoproteins by the name of YfgL, NlpB, YfiO, and SmpA, which were later renamed to BamB, BamC, BamD, and BamE (Wu et al., 2005; Malinverni et al., 2006; Sklar et al., 2007). Together these proteins form the β-barrel assembly machinery (BAM) complex which is today regarded as the system responsible for catalyzing the folding and OM insertion of OMPs. Section 1.5.2 provides a summary of the structural and functional information that has been gathered in this short decade of BAM research.

1.5.2. Known structural and functional information

1.5.2.1. BamA

BamA (88.4 kDa, UniProt ID: P0A940) is the core of the BAM complex and is itself an OMP. As discussed earlier, BamA is well conserved among Gram-negative bacteria and eukaryotes, and is the essential component of the complex. The domain structure of BamA includes the C-terminal β-barrel embedded in the OM, along with an

N-terminus that extends into the periplasm (Figure 1.5). This latter region consists of five polypeptide transport associated (POTRA) domains, which are numbered 1 to 5 from the N- to C-terminus. Current structural information is available only for the periplasmic region, as the membrane embedded region has been a challenge to crystallize. However, it is predicted that BamA may have a similar 16-stranded β-barrel as its homologue FhaC from *Bordetella pertussis* (Clantin et al., 2007). The POTRA domains of BamA have been shown to share similar structures – a three-stranded β-sheet with two α-helices – despite low sequence identity (Kim et al., 2007). The structural information suggests that the five POTRA domains may alternate between extended and bent forms, which may assist in their proposed function as the docking site for the accessory lipoproteins (Gatzeva-Topalova et al., 2010).

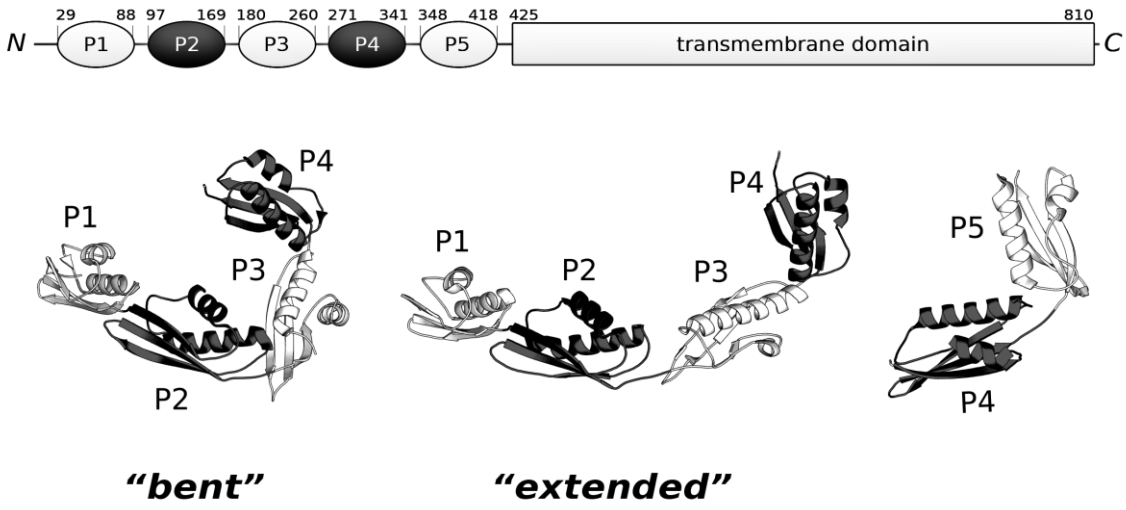

Figure 1.5 Structure of BamA

The domain structure of BamA shows an N-terminal periplasmic region that contains five polypeptide transport associated (POTRA) domains along with a C-terminal transmembrane domain. Only the structure of the POTRA domains are known to date, and the structures have been solved in segments. Ribbon diagrams of the POTRA1-4 domains show the possibility of bent and extended conformations (PDB: 2QDF and 3EFC), while only one structure containing POTRA5 is available (PDB: 3Q6B).

Based on co-immunoprecipitation studies, BamA has been shown to interact directly with BamB and BamD, while the presence of BamD may allow association with BamC and BamE (Malinverni et al., 2006). Looking specifically at the POTRA region, deletion analysis has revealed that the POTRA2-5 domains are required for co-purification with BamB, while POTRA5 is required for co-purification with BamC, BamD, or BamE (Kim et al., 2007). POTRA1 has been shown to interact with the chaperone SurA, while POTRA1-2 have shown to interact with peptides derived from PhoE, an OMP substrate (Bennion et al., 2010; Knowles et al., 2008). These data suggest the POTRA domains to be the region where the lipoproteins associate and substrates may be recruited to for assembly into the OM.

1.5.2.2. BamB

BamB (39.9 kDa, UniProt ID: P77774) is the largest lipoprotein of the BAM complex. It is not a very well conserved component, with its homologue also being absent in *N. meningitides* (Volokhina et al., 2009). Therefore, BamB is not essential for cell survival, but there is a significant decrease in assembly of certain large OMPs in its absence (Charlson et al., 2006). It is therefore believed that BamB's function may be to enhance the overall activity of the BAM complex (Wu et al., 2005). The structure of BamB is an eight-bladed β-propeller, where each blade consists of a four-stranded anti-parallel β-sheet (Figure 1.6) (Kim and Paetzel, 2011; Heuck et al., 2011; Albrecht and Zeth, 2011). As mentioned earlier, BamB is believed to interact with BamA, specifically requiring the POTRA2-5 domains (Kim et al., 2007). And this interaction appears to be independent of the interaction between BamA and the other lipoproteins (Malinverni et al., 2006; Sklar et al., 2007). To date, no direct association has been observed between BamB and the other lipoproteins.

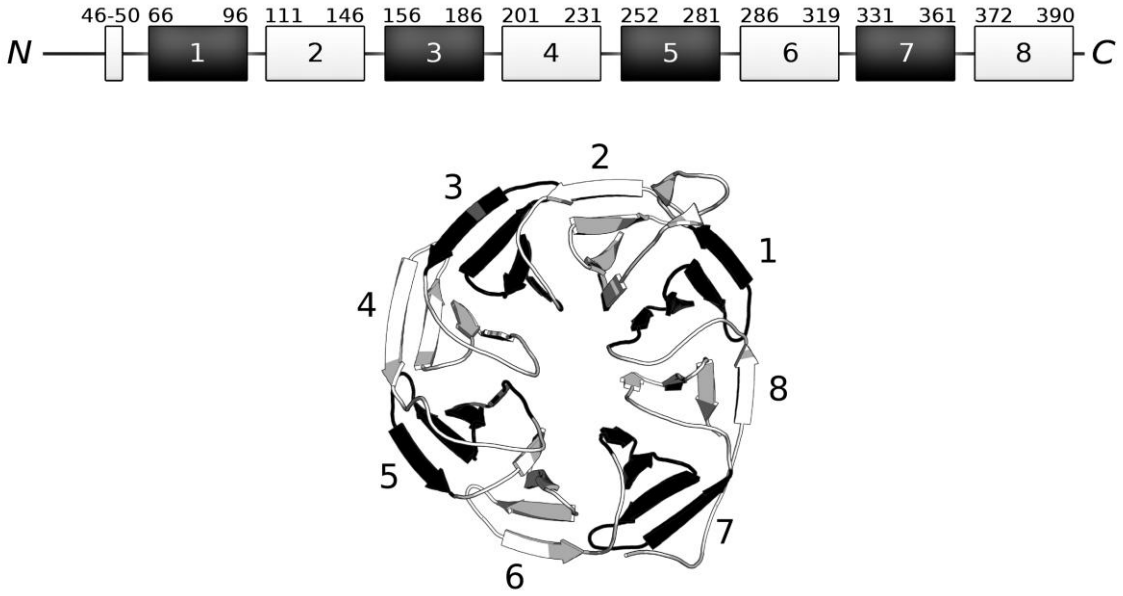

Figure 1.6 Structure of BamB

The domain structure of BamB shows the presence of eight domains. The ribbon diagram of BamB shows these domains to come together to form an eight-bladed β-propeller structure, with residues 46-50 being a part of the eighth blade (PDB: 3P1L). The blades in the ribbon diagram are numbered and coloured to correspond to the domain structure above.

1.5.2.3. BamC

BamC (34.4 kDa, UniProt ID: P0A903) is another lipoprotein that is not essential for cell survival, and its absence makes only a minor impact on OMP assembly (Wu et al., 2005). The domain structure of BamC suggests three independently folding domains, with an unstructured N-terminus, and two similarly folded domains that share low sequence identity (Figure 1.7) (Albrecht and Zeth, 2010, 2011; Kim et al., 2011b, 2011a). The structure of BamC was unknown at the beginning of this project, and its availability played a significant role. Previous studies show direct interaction between BamC and BamD, while it was predicted that the C-terminus of BamD may be the binding site for BamC, but results from this thesis project have shown the N-terminus of BamD to also be important (Malinverni et al., 2006; Kim et al., 2011a). For more discussion on the structure and possible function of BamC, please see Chapter 3.

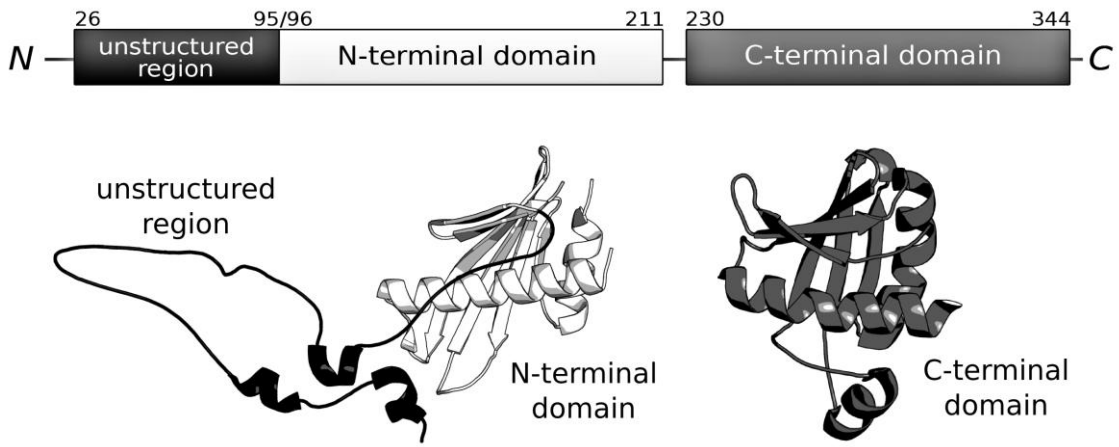

Figure 1.7 Structure of BamC

The domain structure of BamC shows the presence of three domains: an unstructured region at the N-terminus followed by two domains known as the N-terminal domain and the C-terminal domain (also known as $BamC_U$, $BamC_N$, and $BamC_C$). The C-terminal domain was solved separately from the N-terminal domain with the unstructured region, but both globular domains have a similar helix-grip fold (PDG: 2YH5). The unstructured region with the N-terminal region was solved as a part of the BamCD dimer structure (see Chapter 3) (PDB: 3TGO).

1.5.2.4. BamD

BamD (25.8 kDa, UniProt ID: P0AC02) is the second essential component of the BAM complex, and the only required lipoprotein as its absence leads to cell death (Malinverni et al., 2006). BamD is conserved among Gram-negative species, but not in eukaryotes. The structure of BamD reveals the protein to contain ten α-helices forming five tetratricopeptide repeat (TPR) motifs (Figure 1.8) (Sandoval et al., 2011; Albrecht and Zeth, 2011; Kim et al., 2011a). TPR motifs have been commonly seen to be involved in protein-protein interactions, and BamD has been shown to directly interact with BamA, BamC, and BamE (Wu et al., 2005; Malinverni et al., 2006; Sklar et al., 2007; Kim et al., 2011a; Knowles et al., 2011). Comparison of BamD to its structural homologues shows a pocket formed by TPR1 and 2 of BamD to be similarly positioned to the substrate binding pockets found in those homologues (Sandoval et al., 2011). Furthermore, cross-linking experiments using a peptide corresponding to the C-terminal OMP-signal suggests that BamD may indeed be involved in substrate recognition and binding (Albrecht and Zeth, 2011). Further discussion of this role is presented in Chapter 3.

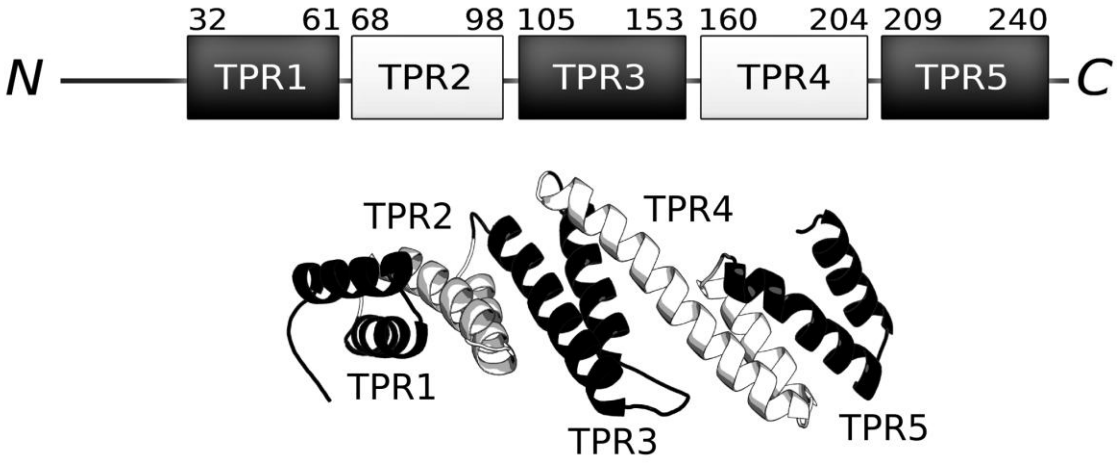

Figure 1.8 Structure of BamD

The domain structure of BamD shows the presence of five tetratricopeptide repeat (TPR) motifs. The ribbon diagram shows that BamD contains ten α-helices which form the five TPR motifs (PDB: 3TGO). Further discussion of BamD structure is provided in Chapter 3.

1.5.2.5. BamE

BamE (10.4 kDa, UniProt ID: P0A937) is a lipoprotein and the smallest member of the BAM complex. Its structure was available at the beginning of this thesis project, and became the basis for many of the experiments. The structure shows a three-stranded anti-parallel β-sheet and two α-helices (Figure 1.9) (Kim et al., 2011c; Albrecht and Zeth, 2011). Interestingly, the N- and C-termini as well as a long loop in between the first and second β-strands are unstructured and show a high degree of flexibility. Furthermore, structural homologues, such as the β-lactamase inhibitor protein (BLIP) share a similar loop which is directly involved in interaction (Kim et al., 2011c). These data provide many possibilities for further investigating the function of BamE. Based on previous experiments, it is suggested that BamE may be involved in stabilizing the interaction between BamA and BamD (Sklar et al., 2007; Knowles et al., 2011). BamE is also the only member of the BAM complex that has shown to form homodimers, however the functional purpose of the dimer is unknown, and there is a debate surrounding its biological relevance (Kim et al., 2011c; Albrecht and Zeth, 2011; Knowles et al., 2011). This thesis project also investigates this topic, and further discussion on BamE and its oligomeric state is provided in Chapter 2.

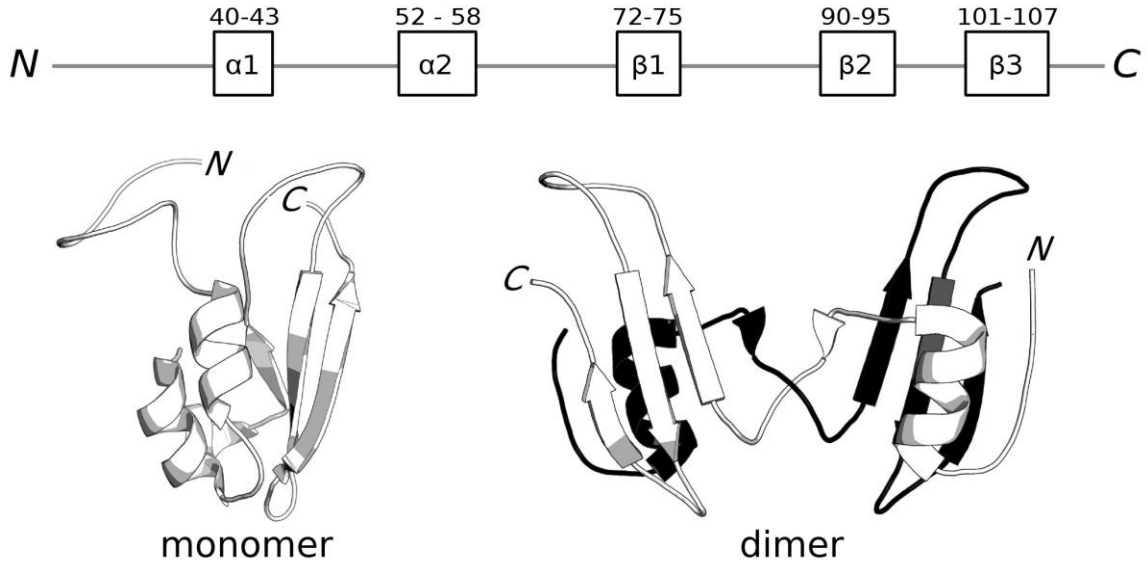

Figure 1.9 Structure of BamE

The domain structure of BamE shows the protein to contain two α-helices and three β-strands. The ribbon diagram of monomeric BamE shows the two α-helices to pack against a three-stranded β-sheet (PDB: 2KXX). A large loop is observed in between the first and second β-strands. In the dimer structure of BamE, the N-terminal helices are exchanged to form a domain-swapped dimer (PDB: 2YH9). Further discussion of the two oligomeric states of BamE is provided in Chapter 2.

1.5.3. Proposed models of mechanism

Once OMPs travel through the periplasm and are brought to the OM, it is unclear how they are folded and inserted into the OM. Based on the structural and interaction data of the BAM complex, it seems that BamD and perhaps the POTRA domains of BamA may be involved in recruiting substrates for assembly. But aside from associating with the various BAM components, one of the mysteries surrounding this process is how the OMP is inserted into OM lipid bilayer. Several models have been proposed including one that suggests that the BamA barrel creates a channel through which the OMP can travel through and insert into the OM from the extracelluar side (Figure 1.10A) (Tommassen, 2007). This model provides challenges such as is the BamA pore large enough to accommodate the OMP, and how would the OMP insert from the extracellular side without additional help. A second model suggests insertion from the periplasmic side, where the OMP folds on the outer surface of BamA, between the BamA barrel and

the lipid bilayer (Figure 1.10B). Similarly, a third proposes the same idea but suggests that an oligomer (ex. a tetramer) of BamA surrounds the OMP as it is folding (Figure 1.10C) (Tommassen, 2007). And finally, we recently proposed a model that the OMP may use BamA as a template, and fold as a part of the BamA barrel, until gradually a large hybrid barrel is formed. Then the newly folded OMP could bud off to form its own independent barrel (Figure 1.10D) (Kim et al., 2012). These models require more investigation in order to understand if any could be valid. Thus, understanding how the BAM complex itself assembles may shed light onto how the complex assembles its substrates.

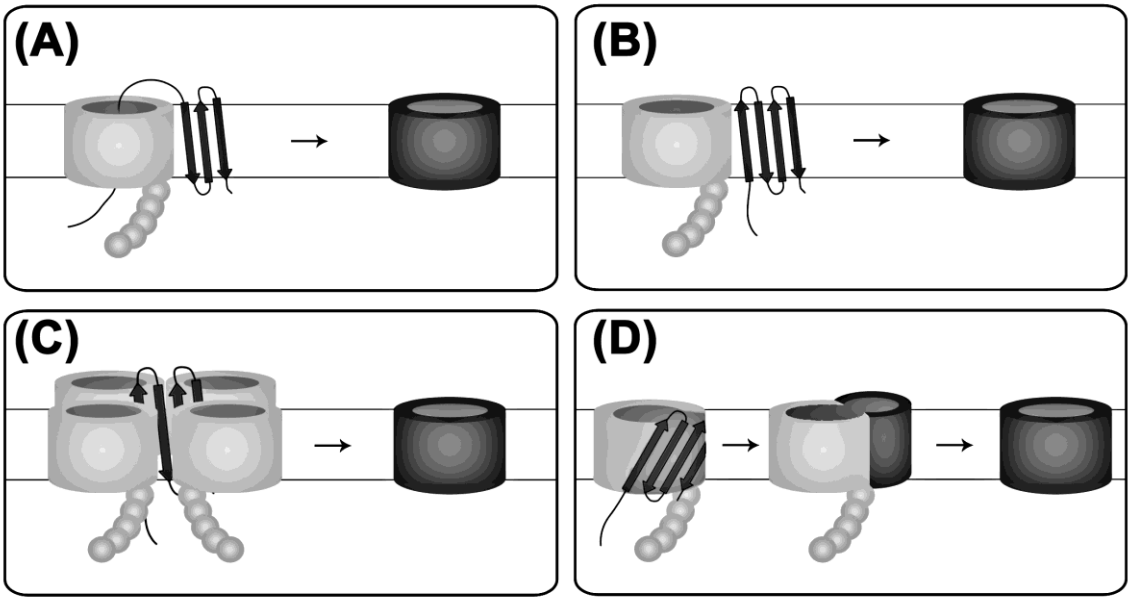

Figure 1.10 Proposed models of OMP folding and insertion by the BAM complex

Four different models of how the BAM complex may facilitate the folding and insertion of OMPs are shown. BamA is shown in light gray, and the substrate protein in black. 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) shows the first model where the substrate protein is first translocated across the outer membrane through the channel formed by the β-barrel domain of BamA, and inserts from the extracellular side. (B) shows the second model where the substrate inserts into the lipid bilayer from the periplasmic face of the outer membrane. (C) shows a model that is similar to the second model, but assumes that BamA forms an oligomeric structure, such as a tetramer. (D) shows the last model where the OMP substrate folds as an extension of the BamA β-barrel, and is then released to become an independent β-barrel.

1.5.4. Homologues in eukaryotes

Homologues of the BAM complex also exist in the OMs of mitochondria and chloroplasts (Schleiff and Soll, 2005). 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. A comparison of the three systems is illustrated in Figure 1.11.

In the mitochondrial system, specifically in *Saccharomyces cerevisiae*, 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) (Paschen et al., 2005). 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. 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 while Sam37 is involved in substrate release (Chacinska et al., 2009; Paschen et al., 2005).

For chloroplasts, as studied in *Arabidopsis thaliana*, 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). 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, facing toward the cytosol (Sommer et al., 2011). This new finding suggests the possibility of OMPs to be imported directly by

Toc75-V and immediately inserted into the OEM, without the use of the traditional 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).

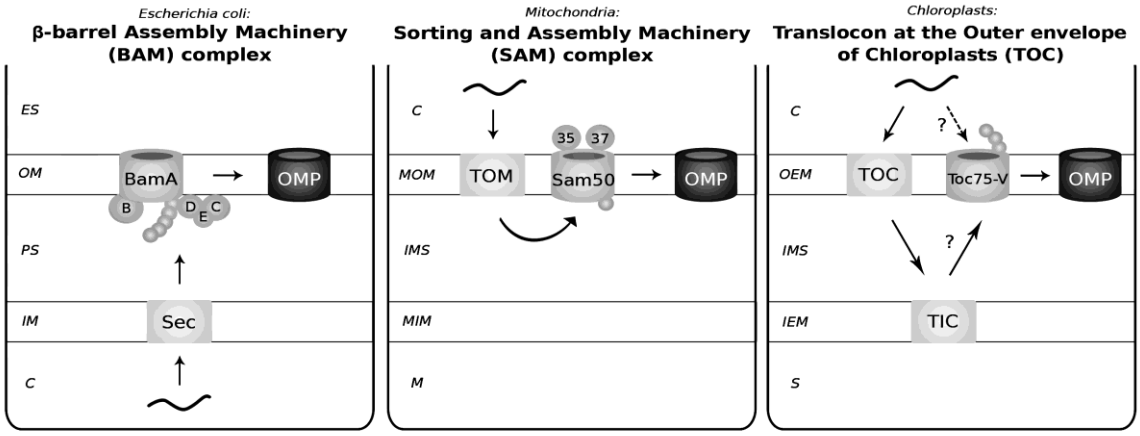

Figure 1.11 The E. coli BAM complex and its homologues in eukaryotes

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 curve) is directed by associated translocons (Sec in *E. coli*, TOM in mitochondria, and TOC/TIC in chloroplasts) to the assembly complex consisting of the core BamA homologue (Sam50 in mitochondria and Toc75-V in chloroplasts) and accessory proteins (BamB-E in *E. coli*, and Sam35 and Sam37 in mitochondria), to form the final folded β-barrel (black barrel labeled OMP). For simplicity, chaperones and other proteins are not shown. Note: the accessory proteins in chloroplasts have not been identified. There are also two possible routes that the OMP may take in chloroplasts as indicated by the solid and dashed arrows. BAM = β-barrel assembly machinery; Sec = SecYEG translocon; ES = extracellular space; OM = outer membrane; PS = periplasmic space; IM = inner membrane; C = cytoplasm; SAM = sorting and assembly machinery; TOM = translocase of outer mitochondrial membrane; MOM = mitochondrial outer membrane; IMS = intermembrane space; MIM = mitochondrial inner membrane; M = matrix; TIC/TOC = translocases at the inner/outer envelope of chloroplasts; OEM = outer envelope membrane; IEM = inner envelope membrane; S = stroma.

1.6. Project Overview

1.6.1. Experimental objectives

The objective of this thesis project is to understand how the periplasmic components of the BAM complex assemble. Previous co-immunoprecipation and co-purification studies conducted by the Silhavy and Kahne groups at Princeton University and Harvard University provide the bulk of what is known about the protein-protein interactions within the BAM complex: the POTRA region of BamA interacts with BamB and BamD independently; BamA forms a complex with BamC, BamD, and BamE, however this complex falls apart in the absence of BamD suggesting that BamA may only be directly interacting with BamD; and, the presence of BamE improves the interaction between BamA and BamD (Figure 1.12) (Wu et al., 2005; Malinverni et al., 2006; Sklar et al., 2007). There are still uncertainties as to which proteins directly interact with each other, and which domains or regions of each protein are involved.

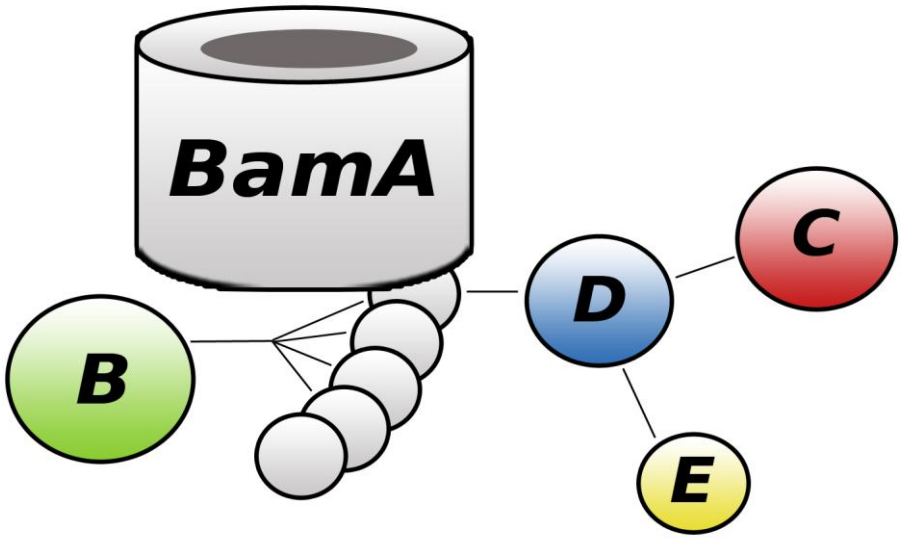

Figure 1.12 Map of known interactions between the BAM components

This figure summarizes the known interactions between the periplasmic components of the BAM complex. Note that the POTRA domains of BamA are numbered 1 to 5 (not labelled in the diagram), with the POTRA1 being furthest from the β-barrel and POTRA5 being the closest to the β-barrel. BamB interacts with BamA independently of BamC, BamD, and BamE. BamB requires the POTRA2-5 domains of BamA for co-purification. BamD requires the POTRA5 domain of BamA for co-purification, and also interacts with BamC and BamE.

The main focus of this project is to study the BamC, BamD, and BamE proteins as previously unpublished data from our lab has shown them to readily form a stable BamCDE subcomplex *in vitro*. Using newly available structural data, the experiments were designed to identify which regions on each protein are involved in the interaction in hopes of constructing better complexes for crystallization studies. Determining the crystal structures of these complexes would finally provide data of how the components come together, which had not been previously available. If successful, similar approaches could be used to further study the interactions involving BamA and BamB, and understand how the entire complex is assembled.

1.6.2. Experimental approach

Using available structural data, variations of the BamC and BamE proteins were constructed to focus on regions of interest. This included truncating the N- or C-termini, and cloning out separate domains of the proteins. Each protein was constructed with a hexahistidine tag to allow easy purification from the cell lysate using nickel affinity chromatography. Interaction was studied by observing complex formation of different combinations of proteins using gel-filtration chromatography. As opposed to affinity-purification techniques, gel-filtration chromatography was selected to study complex formation as this method does not immobilize proteins to the matrix, which could hinder certain complexes from forming.

For each experiment, cell pellets containing the individual proteins were co-lysed in different combinations. The first purification step used nickel affinity chromatography to separate the desired proteins from the other proteins in the cell lysate. Next, the elution fractions were pooled, concentrated, and subjected to gel-filtration chromatography to analyze complex formation. These steps were kept consistent for each combination. Successful complex formation was defined as the elution of intact complexes during gel-filtration chromatography. Note that failure to form a complex does not suggest lack of interaction. Instead, the success of complex formation suggests a strong interaction that can withstand the gel-filtration matrix, yielding a protein complex that can then be used for further interaction and structural studies. Detailed protocols and the theory of each technique used are provided in the Appendices A and B.

2. The N-terminus of BamE is required for the formation of the BamCDE subcomplex

2.1. Strategy

This chapter focuses on identifying which regions of BamE are important for interacting with BamC and BamD. Initial studies using full length versions of these proteins showed that BamE can interact to form BamDE and BamCDE subcomplexes (see sections 2.2.1.1 and 2.2.2.1 below). However, BamE is unable to interact with BamC alone (see Figure C11). Based on available structural data, different versions of BamE were constructed and combined with the other lipoproteins to see the affect on BamDE and BamCDE formation. In addition to these subcomplexes, BamE is also known to form a homodimer, and thus these BamE constructs were also analyzed individually to determine the preferred oligomeric state.

2.1.1. Available structural information

At the beginning of this investigation, the structure of BamE had been recently solved (Kim et al., 2011c). Based on the structure and corresponding NMR backbone dynamics data, BamE was shown to have regions of flexibility in the N- and C-termini, as well as in the large L3 loop (Figure 2.1A). Flexibility could suggest these regions to be involved in interaction as they would be more accessible for protein-protein contact. Sequence alignment of BamE from several Gram-negative species shows that the N-terminus is almost completely conserved, further supporting this region to play an important role (Figure 2.1B). Thus, it was hypothesized that at least one of these regions may be important for the formation of the BamDE or BamCDE subcomplexes, and their absence may prevent the oligomerization process. Alternatively, if oligomerization is not affected, removal of these unstructured and flexible regions would provide a form of BamE that would be better suited for crystallization studies, specifically for solving the

structure of the subcomplexes. This project focuses on the flexible N- and C-termini of BamE as target interaction sites, while the L3 loop has yet to be investigated.

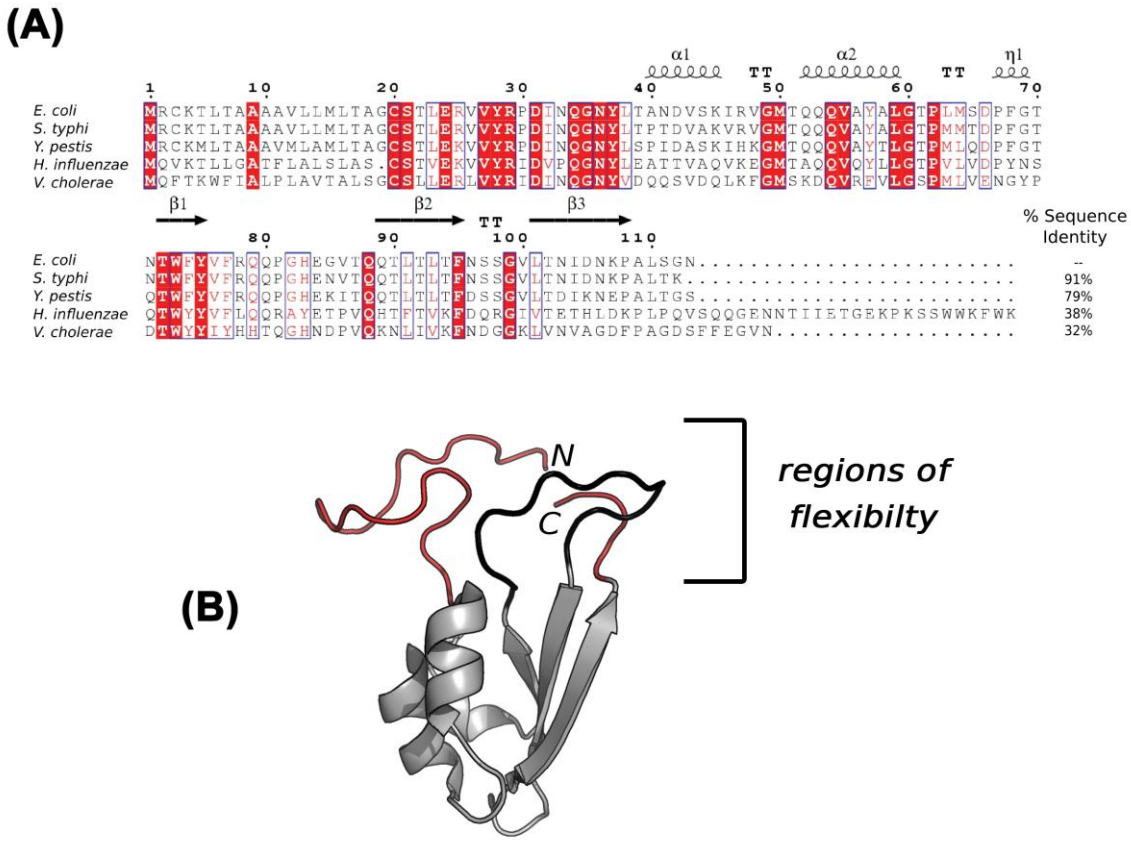

Figure 2.1 Available structural information of BamE

This figure provides the structural information that was available at the beginning of this project, on the basis of which the experiments were designed. (A) shows the sequence alignment is shown of *E. coli* BamE (UniProt ID: P0A937) with homologues from *Salmonella typhi* (Q8XF17), *Yersinia pestis* (G0JGU1), *Haemophilus influenzae* (P44057), and *Vibrio cholerae* (P0C6Q9). Red boxes show absolutely conserved residues, red text shows similar residues, and blue boxes show stretches of similar residues. The secondary structure is shown above the sequence. The alignment shows a region of high conservation between residues 20-38. (B) shows the ribbon diagram of a BamE monomer shows the presence of unstructured N- and C- termini (shown in red) as well as a long L3 loop between strands β1 and β2 (shown in black), all which have previously shown to be flexible. The termini (red) were the focus of this project.

2.1.2. BamE variations constructed

To study if these regions are important for interaction, the termini of BamE were removed individually and together as follows: $BamE_{\Delta N}$, $BamE_{\Delta C}$ and $BamE_{\Delta N \Delta C}$. A list of all the protein constructs used in this experiment with their amino acid boundaries and molecular mass is provided in Table 2.1. The full length proteins (BamC, BamD, and BamE) are missing the N-terminal signal sequence (residues 1-24 for BamC; residues 1-19 for BamD and BamE) and lipidation site (Cys25 for BamC; Cys20 for BamD and BamE) to allow over-expression in the cytoplasm. All constructs have an N-terminal hexahistidine tag. Further cloning details of these constructs are available in Appendix A.

Table 2.1 List of constructs used to study BamE interaction

Construct Name	Residue Boundaries	Molecular Mass (kDa)
BamC	26-344	34.3
BamD	21-245	25.7
BamE	21-113	10.3
$BamE_{\Delta N}$	39-113	8.2
$BamE_{\Delta C}$	21-107	9.7
$BamE_{\Delta N \Delta C}$	39-107	7.6

The molecular mass is calculated based on the amino acid composition of the protein construct.

To study which regions of BamE may be involved in interaction, the various BamE constructs were co-purified with full length BamC and BamD to see the effect on forming the BamDE or BamCDE subcomplexes. Successful interaction was defined by the elution of intact complexes during gel-filtration chromatography.

2.2. Gel-filtration chromatography studies

2.2.1. BamCDE subcomplex formation

2.2.1.1. Full length BamE

When run individually on gel-filtration chromatography, BamC elutes at ~125 mL, BamD at ~155 mL, BamE dimer at ~150 mL and BamE monomer at ~204 mL. It should be noted that based on the standard curve for the column used (for standard curve see Figure A1), the elution volume for BamD corresponds to a mass of ~29 kDa which is close to its calculated mass of 25.7 kDa, and the elution volume for the BamE monomer corresponds to a mass of ~10 kDa which is close to its calculated mass of 10.3 kDa. However, the standard curve estimates the mass of the BamE dimer to be ~32 kDa which is distant from its calculated mass of 20.6 kDa. Although at first glance this may suggest the formation of a BamE homotrimer, this population of BamE is dimeric and elutes earlier due to its extended dimeric structure, causing its peak to overlap with the BamD peak (more discussion on this topic in Sections 2.2.3 and 2.3). Similarly, the elution volume of BamC corresponds to a mass of ~55 kDa, which is distant from the calculated mass of 34.3 kDa. As explained in Chapter 3, this is due to a large unstructured region present at N-terminus of the protein. For the individual chromatograms of these proteins, please see Appendix C.

To see if a complex can form with the three proteins, cell pellets containing full length BamE were co-lysed and purified with cell pellets containing full length BamC and BamD. The resulting gel-filtration chromatogram shows a new peak eluting at ~112 mL, which suggests the formation of a complex (Figure 2.2). Based on the standard curve, this new peak corresponds to a mass of ~73 kDa, which is similar to the calculated mass of 70.3 kDa for the BamCDE subcomplex with a 1:1:1 ratio. Fractions collected from this peak were analyzed on an SDS-PAGE gel, and the results showed the presence of all three proteins, confirming that a BamCDE subcomplex was successfully formed.

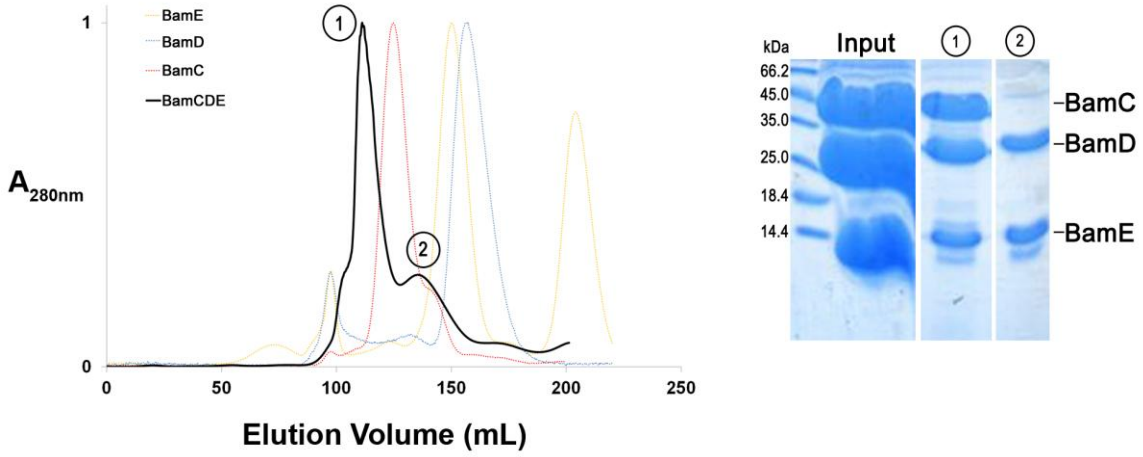

Figure 2.2 Gel-filtration chromatogram and SDS-PAGE showing BamCDE subcomplex formation with full length BamC, BamD, and BamE

The gel-filtration chromatogram (left) shows the elution profile for the sample containing BamC, BamD, and BamE (black curve), which is plotted as the relative UV absorbance at 280nm (A_{280nm}, provided in arbitrary units) for protein eluted at a specific volume (Elution Volume, provided in millilitres (mL)). The results show the presence of two peaks which are labelled as 1 and 2. Peak 1 elutes much earlier than individual BamC (dotted red curve), BamD (dotted blue curve), and BamE (dotted yellow curve) suggesting that they formed a complex. The corresponding SDS-PAGE gel (right) confirms this by showing the presence of BamC, BamD, and BamE in peak 1. Peak 2 contains BamD and BamE, and based on the elution volume suggest a BamDE dimer (see Section 2.2.2 for more details). Overall these results show that a BamCDE subcomplex forms with full length BamC, BamD and BamE. Experiment notes: a HiPrep 26/60 Sephacryl S-100 HR gel-filtration column was used attached to the ÄKTAprime system, with a void volume of 98 mL and a column volume of 320 mL. The experiment was run at 4°C with a buffer containing 20 mM Tris-HCl (pH 8.0) and 100 mM NaCl, where the flow rate was set to 1 mL/min, with 3 mL fractions collected. Fractions eluting in or around the void volume are not shown on the SDS-PAGE gel as they were not in the region of interest and contained aggregated protein.

2.2.1.2. Removal of N-terminus ($BamE_{\Delta N}$)

To study the involvement of the N-terminus of BamE in BamCDE subcomplex formation, amino acid residues 21-38 were removed, which form the unstructured N-terminus as outlined in Section 2.1.2. This $BamE_{\Delta N}$ construct was purified with full length BamC and BamD to see if the BamCDE subcomplex could still form. The resulting gel-filtration chromatogram shows no peak corresponding to the BamCDE subcomplex (Figure 2.3). Instead several peaks are observed at ~121 mL, ~173 mL, and ~208 mL which correspond to masses of ~60 kDa, ~20 kDa, and ~9 kDa. Based on these elution volumes and the SDS-PAGE gel, these peaks indicate the formation of a BamCD dimer

(60.0 kDa), and the individual elution of BamD (25.7 kDa) and $BamE_{\Delta N}$ monomer (8.2 kDa). From this analysis, it appears that the removal of the N-terminus of BamE no longer allows the BamCDE subcomplex to form, suggesting that this region may play an important role in the interaction.

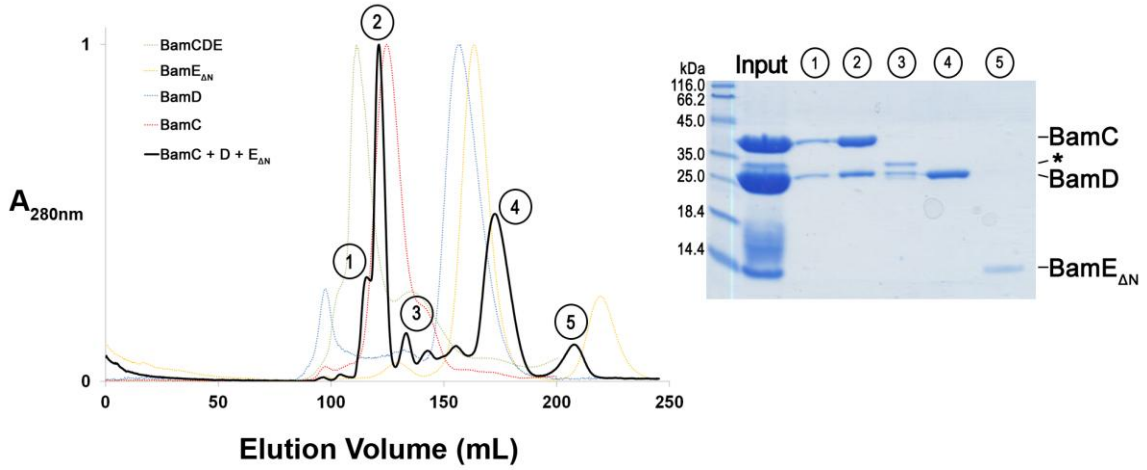

Figure 2.3 Gel-filtration chromatogram and SDS-PAGE showing no BamCDE subcomplex formation with $BamE_{\Delta N}$

The gel-filtration chromatogram (left) shows the elution profile for the sample containing BamC, BamD, and $BamE_{\Delta N}$ (black curve), which is plotted as the relative UV absorbance at 280nm (A_{280nm}, provided in arbitrary units) for protein eluted at a specific volume (Elution Volume, provided in millilitres (mL)). The results show the presence of multiple peaks and regions of interest which are labelled from 1 to 5. Peak 1 and 2 overlap, and elute slightly later than where the BamCDE subcomplex has been observed (dotted green curve), but earlier than individual BamC (dotted red curve), BamD (dotted blue curve), and $BamE_{\Delta N}$ (dotted yellow curve) suggesting that a dimer or complex may have formed. The corresponding SDS-PAGE gel (right) provides more information by showing the presence of only BamC and BamD in regions 1 and 2, suggesting that a BamCD dimer formed but no $BamCDE_{\Delta N}$ subcomplex. The short peak 3 shows the presence of BamD, as well as another band (labelled on the gel with an asterisk, *) which is most likely a degraded form of BamC (see Chapter 3). Peak 4 contains BamD eluting separately, and peak 5 contains monomeric BamE. Overall these results show that a BamCDE subcomplex is unable to form with $BamE_{\Delta N}$. Experiment notes: a HiPrep 26/60 Sephacryl S-100 HR gel-filtration column was used attached to the ÄKTAprime system, with a void volume of 98 mL and a column volume of 32 0mL. The experiment was run at 4°C with a buffer containing 20 mM Tris-HCl (pH 8.0) and 100 mM NaCl, where the flow rate was set to 1 mL/min, with 3 mL fractions collected. Fractions eluting in or around the void volume are not shown on the SDS-PAGE gel as they were not in the region of interest and contained aggregated protein.

2.2.1.3. Removal of C-terminus ($BamE_{\Delta C}$)

To study the involvement of the C-terminus of BamE, amino acids corresponding to this region (residues 108-113) were removed as outlined in Section 2.1.2. This $BamE_{\Delta C}$ construct was co-purified with full length BamC and BamD to see if a BamCDE subcomplex could still form. The resulting gel-filtration chromatogram shows a peak at ~117 mL, corresponding to a molecular mass of ~65 kDa, which is close to the calculated mass of 69.7 kDa for a $BamCDE_{\Delta C}$ subcomplex (Figure 2.4). The SDS-PAGE gel shows the presence of all three proteins in this peak, confirming the formation of a complex. A second peak at ~164 mL corresponds to a mass of ~24 kDa, and indicates the elution of BamD (25.7 kDa) and the $BamE_{\Delta C}$ dimer (19.4 kDa). This experiment suggests that the C-terminus of BamE may not play an important role in forming this subcomplex, as removing this region does not prevent oligomerization.

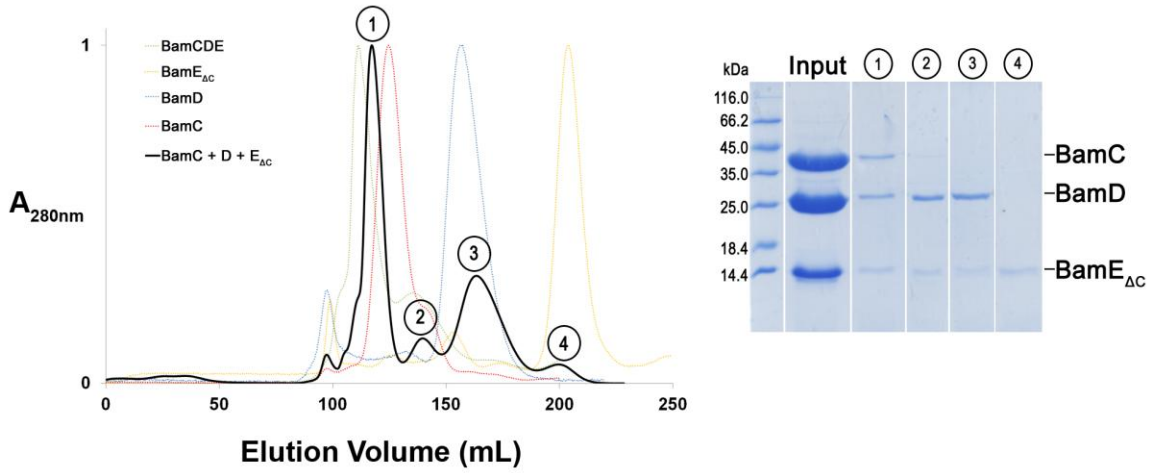

Figure 2.4 Gel-filtration chromatogram and SDS-PAGE showing BamCDE subcomplex formation with $BamE_{\Delta C}$

The gel-filtration chromatogram (left) shows the elution profile for the sample containing BamC, BamD, and $BamE_{\Delta C}$ (black curve), which is plotted as the relative UV absorbance at 280nm (A_{280nm}, provided in arbitrary units) for protein eluted at a specific volume (Elution Volume, provided in millilitres (mL)). The results show the presence of multiple peaks which are labelled from 1 to 4. Peak 1 elutes slightly later than where the BamCDE subcomplex has been observed (dotted green curve), but earlier than individual BamC (dotted red curve), BamD (dotted blue curve), and $BamE_{\Delta C}$ (dotted yellow curve) suggesting that a dimer or complex may have formed. The corresponding SDS-PAGE gel (right) provides more information by showing the presence of BamC, BamD, and $BamE_{\Delta C}$ in peak 1 suggesting that a $BamCDE_{\Delta C}$ subcomplex formed. Peak 3 contains BamD and $BamE_{\Delta C}$, however based on the elution volume, this peak most likely contains BamD and dimeric $BamE_{\Delta C}$ eluting separately. And finally, peak 4 contains monomeric

$BamE_{\Delta C}$. Overall these results show that a BamCDE subcomplex is able to form with $BamE_{\Delta C}$. Note: the short peak 2 shows the presence of BamD and either BamE or $BamE_{\Delta C}$. Based on the elution volume, this peak suggests the formation of a BamDE dimer, but since specifically a $BamDE_{\Delta C}$ was never observed (see Section 2.2.2), this may be the result of endogenous BamE forming a dimer with BamD. Experiment notes: a HiPrep 26/60 Sephacryl S-100 HR gel-filtration column was used attached to the ÄKTAprime system, with a void volume of 98 mL and a column volume of 320 mL. The experiment was run at 4°C with a buffer containing 20 mM Tris-HCl (pH 8.0) and 100 mM NaCl, where the flow rate was set to 1 mL/min, with 3 mL fractions collected. Fractions eluting in or around the void volume are not shown on the SDS-PAGE gel as they were not in the region of interest and contained aggregated protein.

2.2.1.4. Removal of both termini ($BamE_{\Delta N\Delta C}$)

Results from the previous two experiments show that removing the N-terminus of BamE prevents the formation of the BamCDE subcomplex while removing the C-terminus does not. However, based on these studies alone, it cannot be concluded if the N-terminus plays a direct role in BamCDE formation, or if the C-terminus plays an obstructive role in complex formation which the long N-terminus restrains to allow oligomerization. If this latter theory were true, one would expect to see BamCDE subcomplex formation when both termini are removed. To study this, both termini were removed to create the $BamE_{\Delta N\Delta C}$ construct which was co-purified with full length BamC and BamD. The resulting chromatogram shows two main peaks, but none corresponding to a $BamCDE_{\Delta N\Delta C}$ subcomplex (Figure 2.5). Instead, a peak at ~122 mL is observed, which corresponds to a BamCD dimer (60.0 kDa), and a peak at ~175 mL shows the elution of BamD (25.7 kDa) and the $BamE_{\Delta N\Delta C}$ dimer (15.2 kDa). This demonstrates that the absence of the N-terminus of BamE prevents complex formation, independently of the C-terminus.

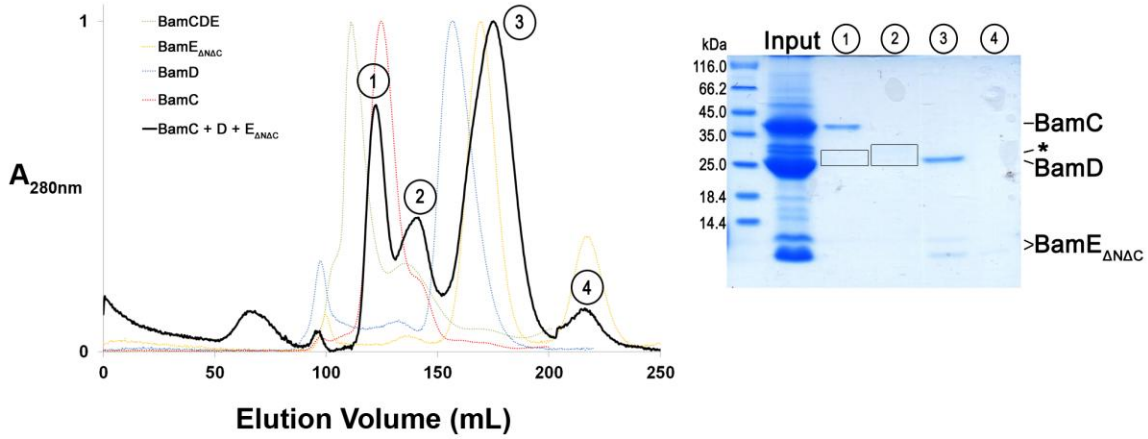

Figure 2.5 Gel-filtration chromatogram and SDS-PAGE showing no BamCDE subcomplex formation with BamE$_{\Delta N\Delta C}$

The gel-filtration chromatogram (left) shows the elution profile for the sample containing BamC, BamD, and BamE$_{\Delta N\Delta C}$ (black curve), which is plotted as the relative UV absorbance at 280nm (A_{280nm}, provided in arbitrary units) for protein eluted at a specific volume (Elution Volume, provided in millilitres (mL)). The results show the presence of multiple peaks which are labelled from 1 to 4. Peak 1 elutes slightly later than where the BamCDE subcomplex has been observed (dotted green curve), slightly earlier than individual BamC (dotted red curve), and much earlier than BamD (dotted blue curve), and BamE$_{\Delta N\Delta C}$ (dotted yellow curve). Based on the chromatogram it is unclear if a dimer formed in peak 1 or if BamC is eluting on its own. The corresponding SDS-PAGE gel (right) provides more information by showing the presence of BamC and faint BamD (boxed) in peak 1 suggesting that a BamCD dimer formed, but not a BamCDE$_{\Delta N\Delta C}$ subcomplex. Peak 2 shows the presence of BamD as well as another band (both are boxed, with the latter band labelled on the gel with an asterisk, *) which is most likely a degraded form of BamC (see Chapter 3). Peak 3 contains BamD and dimeric BamE$_{\Delta N\Delta C}$ eluting individually, while peak 4 contains only BamE$_{\Delta N\Delta C}$. Note that on the gel two bands corresponding to BamE$_{\Delta N\Delta C}$ are observed which represent the tagged (above) and untagged (below) versions of BamE$_{\Delta N\Delta C}$ and none were able to form a BamCDE subcomplex. Overall these results show that a BamCDE subcomplex is unable to form with BamE$_{\Delta N\Delta C}$ (regardless of whether a hexahistidine tag is present or not). Experiment notes: a HiPrep 26/60 Sephacryl S-100 HR gel-filtration column was used attached to the ÄKTAprime system, with a void volume of 98 mL and a column volume of 320 mL. The experiment was run at 4°C with a buffer containing 20 mM Tris-HCl (pH 8.0) and 100 mM NaCl, where the flow rate was set to 1 mL/min, with 3 mL fractions collected. Fractions eluting in or around the void volume are not shown on the SDS-PAGE gel as they were not in the region of interest and contained aggregated protein.

2.2.1.5. Summary of BamE requirements for BamCDE formation

To study the involvement of BamE in the BamCDE subcomplex formation, the N- and C-termini were removed individually and together. Co-purification with full length BamC and BamD using gel-filtration chromatography showed that the N-terminus of BamE is required for successful complex formation, while the C-terminus is not (Figure 2.6; Table 2.2). No BamCDE subcomplex is observed when the N-terminus is missing, regardless of the presence of the C-terminus, suggesting that the involvement of the N-terminus is independent of the C-terminus. While these results suggest the N-terminus of BamE as a requirement for subcomplex formation, this is not enough information to conclude if the N-terminus is directly involved in interacting with either BamC or BamD, or if the absence of this region affects the overall structure of the protein. Thus, further studies, such as determining the structure of this BamCDE subcomplex, would provide insight into understanding how the N-terminus is involved in the interaction.

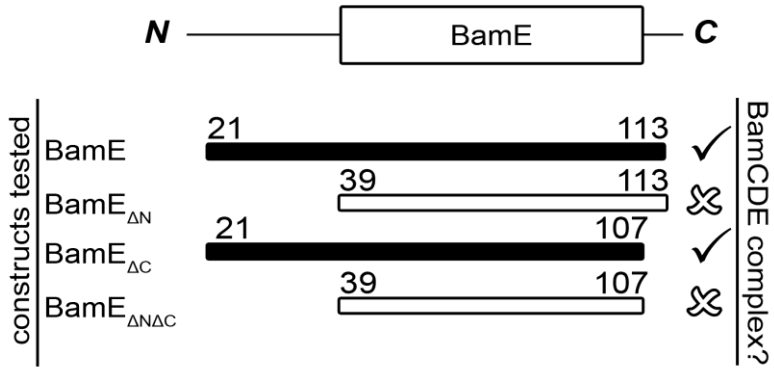

Figure 2.6 Schematic summary of BamE requirements for BamCDE subcomplex formation

This is a schematic summary showing the BamE requirements for BamCDE subcomplex formation. For each construct, the residue boundaries are illustrated and correspond to the domain structure shown above. Successful complex formation is indicated by a black check mark, while no complex formation is indicated by a white X-mark. Based on this diagram, it can be seen that in the absence of the N-terminus, BamE is unable to form the BamCDE subcomplex, suggesting that the N-terminus is required for this association.

Table 2.2 Gel-filtration elution volumes for BamCDE formation

Combination	Peak Elution Volume (mL)	Measured Mass (kDa)	Protein(s) Eluted and their Calculated Mass (kDa)	
BamC	125	55	BamC	34.3
BamD	155	29	BamD	25.7
BamE	150	32	BamE dimer	20.6
	204	10	BamE monomer	10.3
BamC + BamD + BamE	112	73	BamCDE complex	70.3
	136	43	BamDE dimer	36.0
BamC + BamD + $BamE_{\Delta N}$	121	60	BamCD dimer	60.0
	173	20	BamD	25.7
	208	9	$BamE_{\Delta N}$ monomer	8.2
BamC + BamD + $BamE_{\Delta C}$	117	65	$BamCDE_{\Delta C}$ complex	69.7
	164	24	BamD	25.7
			$BamE_{\Delta C}$ dimer	19.4
BamC + BamD + $BamE_{\Delta N\Delta C}$	122	59	BamCD dimer	60.0
	175	19	BamD	25.7
			$BamE_{\Delta N\Delta C}$ dimer	15.2

This table presents a summary of the peaks observed, their elution volumes, and the proteins eluted. The measured molecular mass was estimated from the standard curve of the gel-filtration column using the corresponding elution volume. The calculated molecular mass of the protein eluted is based on the amino acid composition of each protein.

2.2.2. BamDE dimer formation

2.2.2.1. Full length BamE

When run individually on gel-filtration chromatography, BamD elutes at ~155 mL, BamE dimer at ~150 mL and BamE monomer at ~204 mL. To see if a complex can form between BamD and BamE, the cell pellets containing full length versions of both proteins were co-lysed and purified. The resulting gel-filtration chromatogram shows the presence of a new peak eluting at ~136 mL (Figure 2.7). Based on the standard curve for this column, this peak corresponds to a molecular mass of ~43 kDa which is similar to the calculated mass of 36.0 kDa for the BamDE dimer with a 1:1 ratio. The 7 kDa

difference between the estimated and calculated masses can be attributed to the unstructured termini of BamE, which would cause this dimer to elute out a little earlier than expected. When analyzed by SDS-PAGE, this peak shows the presence of both BamD and BamE, confirming that a BamDE dimer had formed. The remaining peaks at ~153 mL and ~195 mL correspond to a molecular mass of ~30 kDa and ~12 kDa which suggest the elution of BamD (25.7 kDa) and dimeric BamE (20.6 kDa) for the second peak, and elution of monomeric BamE (10.3 kDa) for the third peak. Note that while both BamD and dimeric BamE are eluting together in the second peak, no BamDE dimer is occurring here. This is due to the fact that the elution volumes of BamD and BamE dimer (155 mL and 150 mL respectively) are close causing the peaks to overlap.

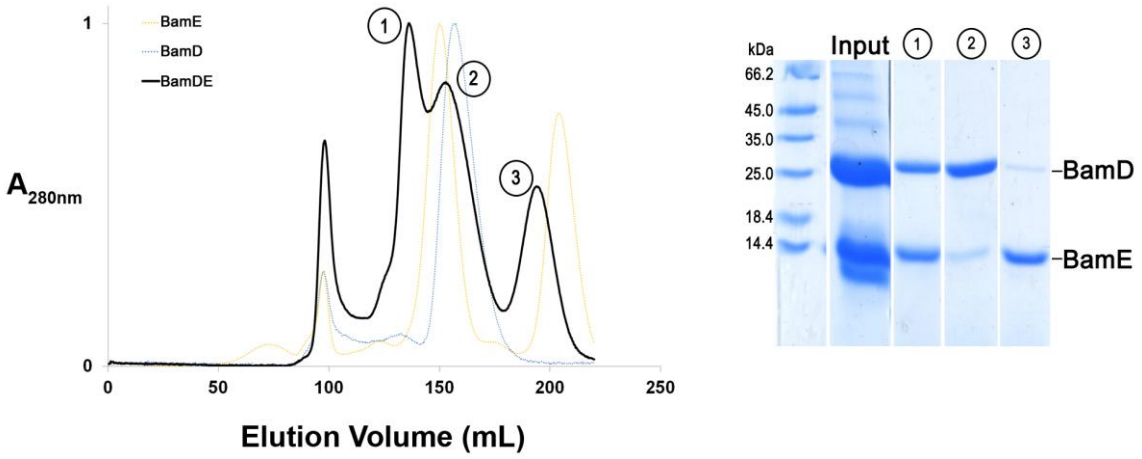

Figure 2.7 Gel-filtration chromatogram and SDS-PAGE showing BamDE dimer formation with full length BamD and BamE

The gel-filtration chromatogram (left) shows the elution profile for the sample containing BamD and BamE (black curve), which is plotted as the relative UV absorbance at 280nm (A_{280nm}, provided in arbitrary units) for protein eluted at a specific volume (Elution Volume, provided in millilitres (mL)). The results show the presence of multiple peaks which are labelled as 1, 2, and 3. Peak 1 elutes earlier than individual BamD (dotted blue curve) and BamE (dotted yellow curve) suggesting that a BamDE dimer may have formed. The corresponding SDS-PAGE gel (right) confirms this by showing the presence of both BamD and BamE in peak 1, supporting that a BamDE dimer formed. Peak 2 also shows the presence of both proteins, but based on elution volume this peak contains BamD and dimeric BamE eluting out separately. Peak 3 shows the elution of monomeric BamE, while traces of BamD still emerge. Overall these results show that a BamDE dimer is able to form with full length BamD and BamE. Experiment notes: a HiPrep 26/60 Sephacryl S-100 HR gel-filtration column was used attached to the ÄKTAprime system, with a void volume of 98 mL and a column volume of 320 mL. The experiment was run at 4°C with a buffer containing 20 mM Tris-HCl (pH 8.0) and 100 mM NaCl, where the flow rate was set to 1 mL/min, with 3 mL fractions collected. Fractions eluting in or around the void volume are not shown on the SDS-PAGE gel as they were not in the region of interest and contained aggregated protein.

2.2.2.2. Removal of N-terminus ($BamE_{\Delta N}$)

To study the involvement of the N-terminus of BamE in BamDE dimer formation, the N-terminal region was removed as outlined in Section 2.1.2. The $BamE_{\Delta N}$ construct was co-lysed and purified with full length BamD. The resulting gel-filtration chromatogram shows the presence of two main peaks eluting at ~158 mL and ~213 mL (Figure 2.8). Based on the standard curve, the molecular mass of these two peaks are estimated to be approximately ~27 kDa and ~8 kDa respectively, which is comparable to the calculated masses of BamD (25.7 kDa) and dimeric $BamE_{\Delta N}$ (16.4k Da, extended) for the first peak, and monomeric $BamE_{\Delta N}$ (8.2kDa) for the second peak. The absence of an earlier peak indicates that a BamDE dimer was unable to form, suggesting that the N-terminus of BamE may be required for the association.

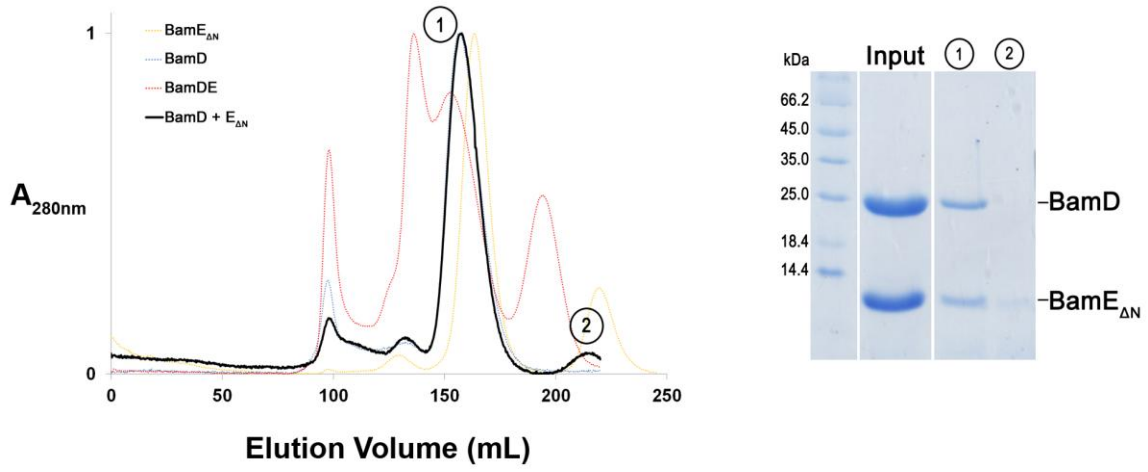

Figure 2.8 Gel-filtration chromatogram and SDS-PAGE showing no BamDE dimer formation with $BamE_{\Delta N}$

The gel-filtration chromatogram (left) shows the elution profile for the sample containing BamD and $BamE_{\Delta N}$ (black curve), which is plotted as the relative UV absorbance at 280nm (A_{280nm}, provided in arbitrary units) for protein eluted at a specific volume (Elution Volume, provided in millilitres (mL)). The results show the presence of two peaks which are labelled as 1 and 2. These peaks elute later than where the BamDE dimer has been observed (dotted red curve), and have similar elution volumes as individual BamD (dotted blue curve, which superimposes tightly with peak 1) and $BamE_{\Delta N}$ (dotted yellow curve) suggesting that a dimer did not form. The corresponding SDS-PAGE gel (right) confirms this by showing the presence of BamD and $BamE_{\Delta N}$ in peak 1, indicating that BamD and dimeric $BamE_{\Delta N}$ are eluting separately, while peak 2 shows the elution of monomeric $BamE_{\Delta N}$. Overall these results show that a BamDE dimer is unable to form with $BamE_{\Delta N}$. Experiment notes: a HiPrep 26/60 Sephacryl S-100 HR gel-filtration column was used attached to the ÄKTAprime system, with a void volume of 98 mL and a column

volume of 320 mL. The experiment was run at 4°C with a buffer containing 20 mM Tris-HCl (pH 8.0) and 100 mM NaCl, where the flow rate was set to 1 mL/min, with 3 mL fractions collected. Fractions eluting in or around the void volume are not shown on the SDS-PAGE gel as they were not in the region of interest and contained aggregated protein.

2.2.2.3. Removal of C-terminus ($BamE_{\Delta C}$)

To study the involvement of the C-terminus of BamE in BamDE dimer formation, the C-terminal region was removed as outlined in Section 2.1.2. This $BamE_{\Delta C}$ construct was co-lysed and purified with full length BamD. The resulting gel-filtration chromatogram shows the presence of only two peaks with elution volumes of ~153 mL and ~195 mL that correspond to molecular mass values of ~30 kDa and ~12 kDa (Figure 2.9). The SDS-PAGE gel suggests the presence of BamD (25.7 kDa) for the first peak, and monomeric $BamE_{\Delta C}$ (9.7 kDa) for the second peak. Similar to the removal of the N-terminus of BamE, the removal of the C-terminus of BamE prevents the formation of a BamDE dimer as no early peak corresponding to a BamDE dimer is observed on the chromatogram. This suggests that the C-terminus of BamE also plays an important role in the interaction.

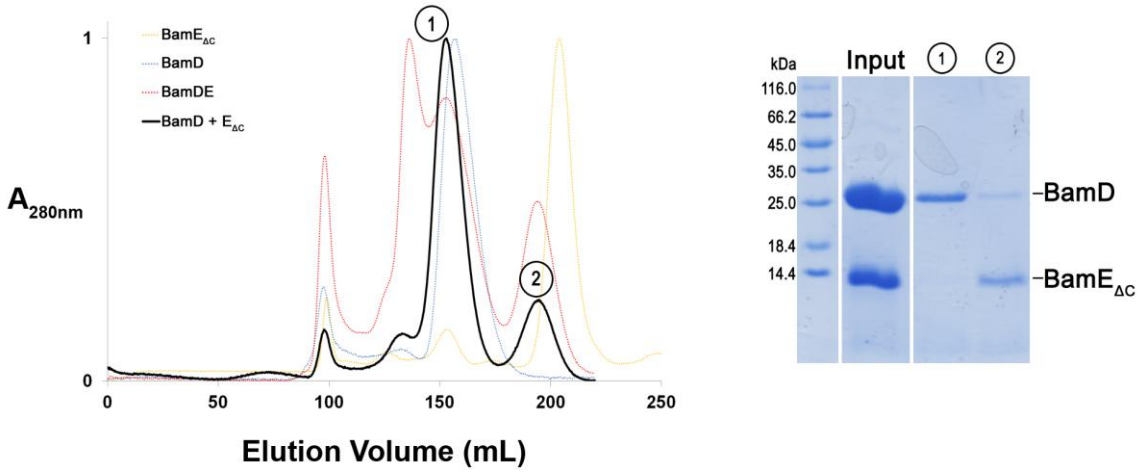

Figure 2.9 Gel-filtration chromatogram and SDS-PAGE showing no BamDE dimer formation with $BamE_{\Delta C}$

The gel-filtration chromatogram (left) shows the elution profile for the sample containing BamD and $BamE_{\Delta C}$ (black curve), which is plotted as the relative UV absorbance at 280nm (A_{280nm}, provided in arbitrary units) for protein eluted at a specific volume (Elution Volume, provided in millilitres (mL)). The results show the presence of two peaks which are labelled as 1 and 2. These peaks elute later than where the BamDE dimer has been observed (dotted red curve), and have similar elution volumes as individual BamD (dotted blue curve) and $BamE_{\Delta C}$ (dotted yellow curve) suggesting that a dimer did not form. The corresponding SDS-PAGE gel (right) confirms this by showing the presence of BamD in peak 1, and peak 2 shows the elution of monomeric $BamE_{\Delta C}$

while traces of BamD still emerge. Overall these results show that a BamDE dimer is unable to form with $BamE_{\Delta C}$. Experiment notes: a HiPrep 26/60 Sephacryl S-100 HR gel-filtration column was used attached to the ÄKTAprime system, with a void volume of 98 mL and a column volume of 320 mL. The experiment was run at 4°C with a buffer containing 20 mM Tris-HCl (pH 8.0) and 100 mM NaCl, where the flow rate was set to 1 mL/min, with 3 mL fractions collected. Fractions eluting in or around the void volume are not shown on the SDS-PAGE gel as they were not in the region of interest and contained aggregated protein.

2.2.2.4. Removal of both termini ($BamE_{\Delta N\Delta C}$)

From the previous results, it appears that the removal of either the N- or C-terminus of BamE prevents the formation of a BamDE dimer. Thus, removal of both termini should produce similar results. The $BamE_{\Delta N\Delta C}$ construct was co-lysed and purified with full length BamD. The resulting gel-filtration chromatogram shows the presence of only two peaks at ~153 mL and ~217 mL, corresponding to a mass of ~30 kDa and ~8 kDa (Figure 2.10). Based on the SDS-PAGE gel it can be seen that BamD (25.7 kDa) is eluting in the first peak while monomeric $BamE_{\Delta N\Delta C}$ (7.6 kDa) is eluting out in the second peak. An earlier peak corresponding to a BamDE dimer was not observed. Results from the individual N- and C-terminal truncations were unable to form the dimer, thus it was expected that removing both termini together would also prevent dimer formation. This suggests that both the N- and C-termini of BamE are required for BamDE association.

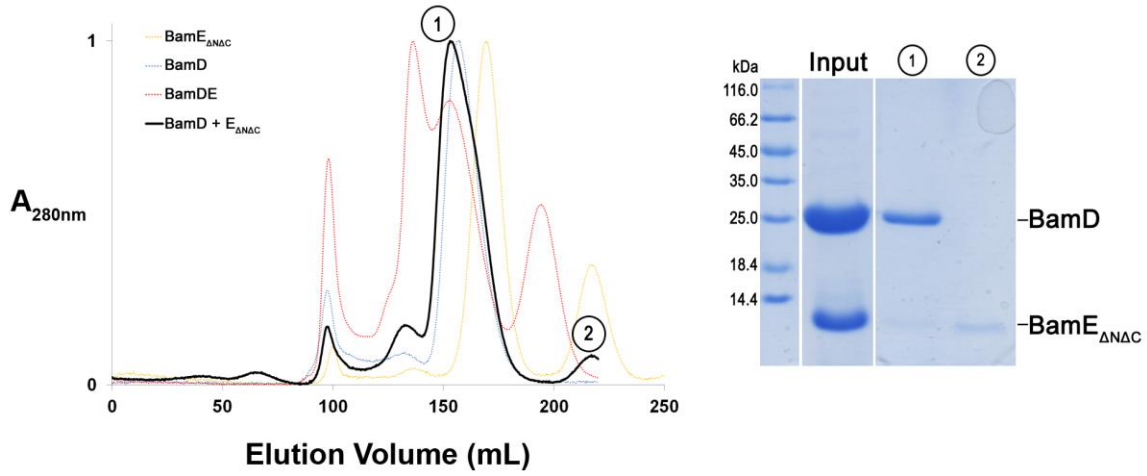

Figure 2.10 Gel-filtration chromatogram and SDS-PAGE showing no BamDE dimer formation with $BamE_{\Delta N\Delta C}$

The gel-filtration chromatogram (left) shows the elution profile for the sample containing BamD and $BamE_{\Delta N\Delta C}$ (black curve), which is plotted as the relative UV absorbance at 280nm (A_{280nm}, provided in arbitrary units) for protein eluted at a specific volume (Elution Volume, provided in

millilitres (mL)). The results show the presence of two peaks which are labelled as 1 and 2. These peaks elute later than where the BamDE dimer has been observed (dotted red curve), and have similar elution volumes as individual BamD (dotted blue curve) and $BamE_{\Delta C}$ (dotted yellow curve) suggesting that a dimer did not form. The corresponding SDS-PAGE gel (right) confirms this by showing the presence of BamD in peak 1, while peak 2 shows the elution of monomeric $BamE_{\Delta N \Delta C}$. Overall these results show that a BamDE dimer is unable to form with $BamE_{\Delta N \Delta C}$. Experiment notes: a HiPrep 26/60 Sephacryl S-100 HR gel-filtration column was used attached to the ÄKTAprime system, with a void volume of 98 mL and a column volume of 320 mL. The experiment was run at 4°C with a buffer containing 20 mM Tris-HCl (pH 8.0) and 100 mM NaCl, where the flow rate was set to 1 mL/min, with 3 mL fractions collected. Fractions eluting in or around the void volume are not shown on the SDS-PAGE gel as they were not in the region of interest and contained aggregated protein.

2.2.2.5. Summary of BamE requirements for BamDE formation

To study the involvement of BamE in the formation of the BamDE dimer, the N- and C-termini were removed individually and together. Co-purification with full length BamD using gel-filtration chromatography showed that both termini are required for successful formation of this dimer (Figure 2.11; Table 2.3). Removal of both or only one terminus prevented the BamDE dimer from forming. However, similar to the summary from Section 2.2.1.5, these data are insufficient for determining how the two regions are involved in the interaction. Although a recent NMR study suggests BamD binds to a region of BamE that includes the disordered N-terminus, it is still too early to clearly state how the two proteins are binding (Knowles et al., 2011). It is possible that the termini contribute by directly associating with areas of BamD; conversely, it is possible that the removal of these regions is disrupting the native structure of BamE, thus preventing proper contacts between BamD and the structured region of BamE. In either case, determining the BamDE dimer structure would provide better insight into how the N- and C-termini are involved.

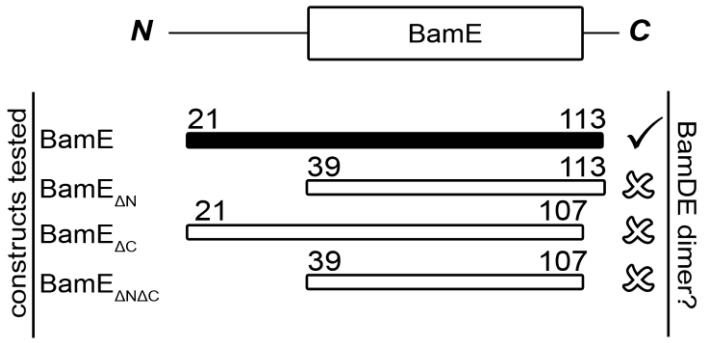

Figure 2.11 Schematic summary of BamE requirements for BamDE dimer formation

This is a schematic summary showing the BamE requirements for BamDE dimer formation. For each construct, the residue boundaries are illustrated and correspond to the domain structure shown above. Successful dimer formation is indicated by a black check mark, while no complex formation is indicated by a white X-mark. Based on this diagram, it can be seen that in the absence of either N- or C-termini, BamE is unable to form the BamDE dimer, suggesting that both termini are required for this association.

Table 2.3 Gel-filtration elution volumes for BamDE formation

Combination	Peak Elution Volume (mL)	Measured Mass (kDa)	Protein(s) Eluted and their Calculated Mass (kDa)	
BamD	155	29	BamD	25.7
BamE	150	32	BamE dimer	20.6
	204	10	BamE monomer	10.3
BamD + BamE	136	43	BamDE dimer	36.0
	153	30	BamD	25.7
			BamE dimer	20.6
	195	12	BamE monomer	10.3
BamD + BamE$_{\Delta N}$	158	27	BamD	25.7
			BamE$_{\Delta N}$ dimer	16.4
	213	8	BamE$_{\Delta N}$ monomer	8.2
BamD + BamE$_{\Delta C}$	153	30	BamD	25.7
	195	12	BamE$_{\Delta C}$ monomer	9.7
BamD + BamE$_{\Delta N \Delta C}$	153	30	BamD	25.7
	217	8	BamE$_{\Delta N \Delta C}$ monomer	7.6

This table presents a summary of the peaks observed, their elution volumes, and the proteins eluted. The measured molecular mass was estimated from the standard curve of the gel-filtration column using the corresponding elution volume. The calculated molecular mass of the protein eluted is based on the amino acid composition of each protein.

2.2.3. BamE homodimer formation

2.2.3.1. Full length BamE

When purified alone using gel-filtration chromatography, full length BamE produces two peaks on the chromatogram (Figure 2.12). The first peak eluting at ~150 mL corresponds to a molecular mass value of ~32 kDa, while the second peak eluting at ~204 mL corresponds to a mass of ~10 kDa. The second peak's molecular mass estimation is very close to the calculated mass of 10.3 kDa for a BamE monomer. However, the first peak's mass estimation is distant to the calculated mass of 20.6 kDa for a BamE dimer, and closer to a mass of 30.3 kDa for a BamE trimer. This would suggest BamE to be present as a mixture of monomers and trimers. However, past literature has shown strong evidence for a BamE dimer to exist rather than a trimer (Kim et al., 2011c). Also, cross-linking experiments with oligomeric BamE samples suggests BamE is forming a dimer (see Figure C12). And recently, the structure of the BamE dimer was solved, which shows two BamE monomers to open up and form a domain-swapped dimer (further discussion of this structure in Section 2.2.3.5) (Albrecht and Zeth, 2011). This extended conformation of the dimer causes the protein to elute out much earlier on the gel-filtration chromatogram as opposed to if the dimer was globular. Research has also shown that while BamE is present as a mixture of dimers and monomers, these two populations do not appear to be in equilibrium as the dimers remain as dimers, while the monomers remain as monomers (Kim et al., 2011c). However, it has not yet been understood why a specific oligomeric state is preferred. In the case of full length BamE, the two populations appear to exist in equal amounts (Figure 2.12).

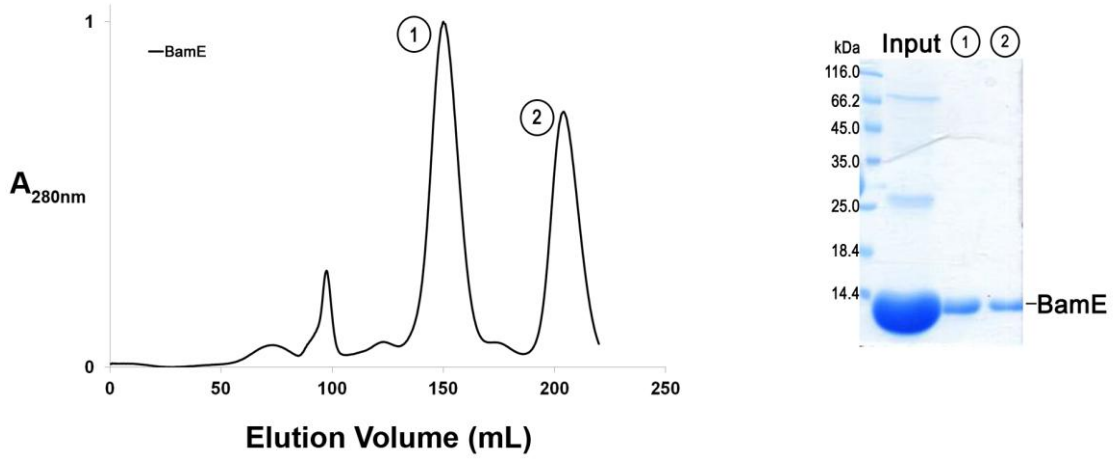

Figure 2.12 Gel-filtration chromatogram and SDS-PAGE showing full length BamE to exist as homodimers and monomers in similar amounts

The gel-filtration chromatogram (left) shows the elution profile for the sample containing full length BamE which is plotted as the relative UV absorbance at 280nm (A_{280nm}, provided in arbitrary units) for protein eluted at a specific volume (Elution Volume, provided in millilitres (mL)). The results show the presence of two peaks which are labelled as 1 and 2. The corresponding SDS-PAGE gel (right) shows both peaks to contain BamE. Based on the elution volume, these peaks suggest BamE to exist as homodimers and monomers. Based on peak distribution, the two populations appear to exist in similar amounts. Experiment notes: a HiPrep 26/60 Sephacryl S-100 HR gel-filtration column was used attached to the ÄKTAprime system, with a void volume of 98 mL and a column volume of 320 mL. The experiment was run at 4°C with a buffer containing 20 mM Tris-HCl (pH 8.0) and 100 mM NaCl, where the flow rate was set to 1 mL/min, with 3 mL fractions collected. Fractions eluting in or around the void volume are not shown on the SDS-PAGE gel as they were not in the region of interest and contained aggregated protein.

2.2.3.2. Removal of N-terminus ($BamE_{\Delta N}$)

To study the involvement of the N-terminus in formation of the BamE homodimer, the N-terminal region of BamE was removed as outlined in Section 2.1.2. This $BamE_{\Delta N}$ construct was purified alone, and the resulting gel-filtration chromatogram shows the presence of two peaks eluting at ~163 mL and ~219 mL which correspond to molecular mass values of ~24 kDa and ~7 kDa (Figure 2.13). These mass estimations are similar to the calculated masses for an extended $BamE_{\Delta N}$ dimer (16.4 kDa) and a $BamE_{\Delta N}$ monomer (8.2 kDa). However, the heights of these peaks are significantly different. By comparing the area of the peaks (see appendix A for method), it can be estimated that ~80% of the $BamE_{\Delta N}$ population formed a homodimer, while only ~20% remained monomeric. This suggests that removal of the N-terminus of BamE promotes homodimerization.

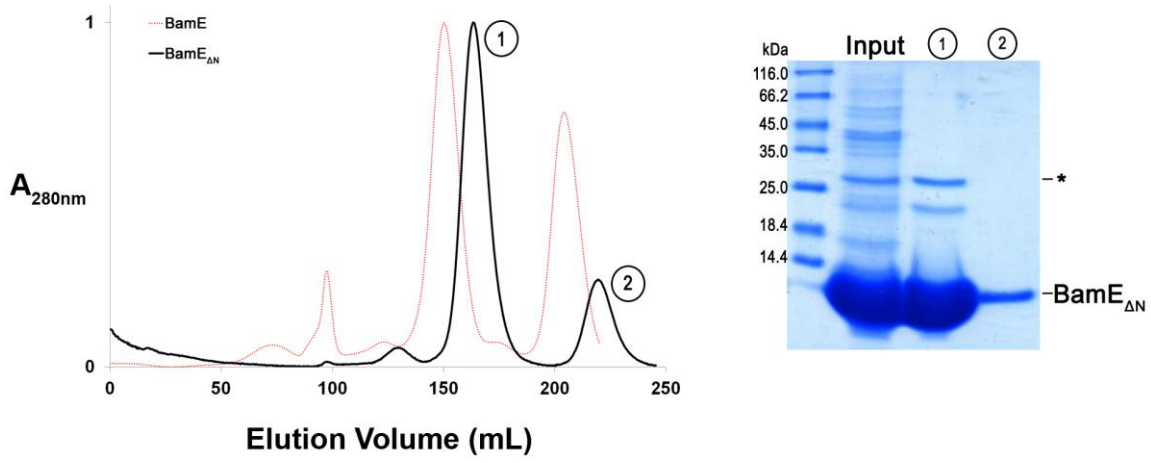

Figure 2.13 Gel-filtration chromatogram and SDS-PAGE showing BamE$_{\Delta N}$ to exist mainly as a homodimer

The gel-filtration chromatogram (left) shows the elution profile for the sample containing BamE$_{\Delta N}$ which is plotted as the relative UV absorbance at 280nm (A_{280nm}, provided in arbitrary units) for protein eluted at a specific volume (Elution Volume, provided in millilitres (mL)). The results show the presence of two peaks which are labelled as 1 and 2. The corresponding SDS-PAGE gel (right) shows both peaks to contain BamE$_{\Delta N}$ (the band labelled with an asterisk, *, shows the presence of endogenous BamD). In comparison to the elution profile of full length BamE (dotted red curve), BamE$_{\Delta N}$ can also be seen to exist as a mixture of homodimers and monomer, with the majority of the population favouring the homodimeric form. Experiment notes: a HiPrep 26/60 Sephacryl S-100 HR gel-filtration column was used attached to the ÄKTAprime system, with a void volume of 98 mL and a column volume of 320 mL. The experiment was run at 4°C with a buffer containing 20 mM Tris-HCl (pH 8.0) and 100 mM NaCl, where the flow rate was set to 1 mL/min, with 3 mL fractions collected. Fractions eluting in or around the void volume are not shown on the SDS-PAGE gel as they were not in the region of interest and contained aggregated protein.

2.2.3.3. Removal of C-terminus (BamE$_{\Delta C}$)

To study the involvement of the C-terminus in formation of the BamE homodimer, the C-terminal region was removed as outlined in Section 2.1.2. This BamE$_{\Delta C}$ construct was purified alone, and the resulting gel-filtration chromatogram shows two peaks eluting at ~152 mL and ~204 mL which correspond to a molecular mass of ~31 kDa and ~10 kDa (Figure 2.14). These mass estimations are similar to the calculated masses for an extended BamE$_{\Delta C}$ dimer (19.4 kDa) and a BamE$_{\Delta C}$ monomer (9.7 kDa). By comparing the area of the peaks, it can be estimated that only ~3% of the population is present as a homodimer, while ~97% is present as a monomer. This suggests that removal of the C-terminus may prevent dimerization of BamE, and that perhaps it may be required for dimerization.

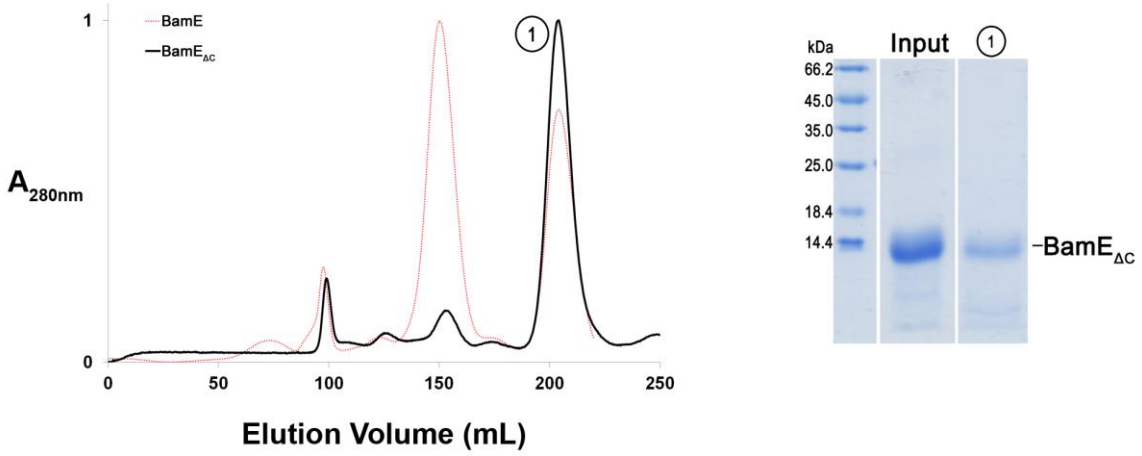

Figure 2.14 Gel-filtration chromatogram and SDS-PAGE showing BamE$_{\Delta C}$ to exist mainly as a monomer

The gel-filtration chromatogram (left) shows the elution profile for the sample containing BamE$_{\Delta C}$ which is plotted as the relative UV absorbance at 280nm (A_{280nm}, provided in arbitrary units) for protein eluted at a specific volume (Elution Volume, provided in millilitres (mL)). The results show the presence of one major peak which is labelled as 1. The corresponding SDS-PAGE gel (right) shows this peak to contain BamE$_{\Delta C}$ (note the unlabelled dimeric peak at ~163 mL was undetectable on the gel). In comparison to the elution profile of full length BamE (dotted red curve), BamE$_{\Delta C}$ can also be seen to exist as a mixture of homodimers and monomer, with almost all of the population favouring the monomeric form. Experiment notes: a HiPrep 26/60 Sephacryl S-100 HR gel-filtration column was used attached to the ÄKTAprime system, with a void volume of 98 mL and a column volume of 320 mL. The experiment was run at 4°C with a buffer containing 20 mM Tris-HCl (pH 8.0) and 100 mM NaCl, where the flow rate was set to 1 mL/min, with 3 mL fractions collected. Fractions eluting in or around the void volume are not shown on the SDS-PAGE gel as they were not in the region of interest and contained aggregated protein.

2.2.3.4. Removal of both termini (BamE$_{\Delta N\Delta C}$)

Results from the previous two experiments in Sections 2.2.3.2 and 2.2.3.3 have shown that removal of the N-terminus of BamE promotes dimerization, while removal of the C-terminus prevents dimerization. Thus, removal of both termini would give insight into which oligomeric state is preferred in the absence of these unstructured regions, and if the C-terminus is required for dimerization. The BamE$_{\Delta N\Delta C}$ was purified alone, and the resulting gel-filtration chromatogram shows the presence of two peaks eluting at ~169 mL and ~217 mL, which correspond to masses of ~21 kDa and ~8 kDa (Figure 2.15). These mass estimations are similar to the calculated masses for an extended BamE$_{\Delta N\Delta C}$ dimer (15.2 kDa) and a BamE$_{\Delta N\Delta C}$ monomer (7.6 kDa). By comparing the area of the peaks, it can be estimated that ~74% of the population of BamE is present in the dimeric form. This again shows that removal of the N-terminus of BamE promotes

dimerization. This result also rejects the possibility of the C-terminus being a requirement for BamE homodimer formation.

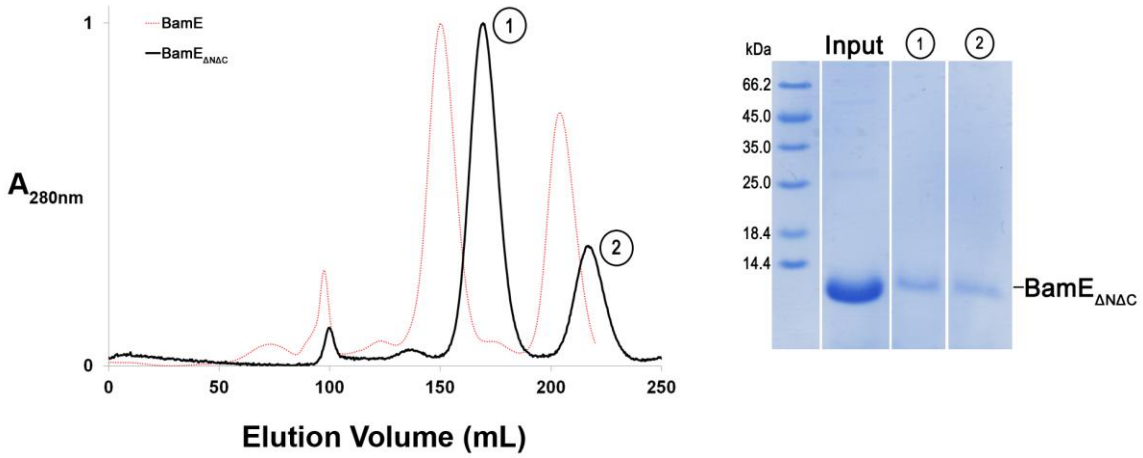

Figure 2.15 Gel-filtration chromatogram and SDS-PAGE showing BamE$_{\Delta N\Delta C}$ to exist mainly as a homodimer

The gel-filtration chromatogram (left) shows the elution profile for the sample containing BamE$_{\Delta N\Delta C}$ which is plotted as the relative UV absorbance at 280nm (A_{280nm}, provided in arbitrary units) for protein eluted at a specific volume (Elution Volume, provided in millilitres (mL)). The results show the presence of two peaks which are labelled as 1 and 2. The corresponding SDS-PAGE gel (right) shows these peaks to contain BamE$_{\Delta N\Delta C}$. In comparison to the elution profile of full length BamE (dotted red curve), BamE$_{\Delta N\Delta C}$ can also be seen to exist as a mixture of homodimers and monomer, with majority of the population favouring the homodimeric form. Experiment notes: a HiPrep 26/60 Sephacryl S-100 HR gel-filtration column was used attached to the ÄKTAprime system, with a void volume of 98 mL and a column volume of 320 mL. The experiment was run at 4°C with a buffer containing 20 mM Tris-HCl (pH 8.0) and 100 mM NaCl, where the flow rate was set to 1 mL/min, with 3 mL fractions collected. Fractions eluting in or around the void volume are not shown on the SDS-PAGE gel as they were not in the region of interest and contained aggregated protein.

2.2.3.5. Summary of requirements for BamE homodimer formation

To study the involvement of the N- and C-termini in BamE dimerization, the termini were removed individually and together, and the resulting BamE constructs were analyzed through gel-filtration chromatography. Results show that in the absence of the N-terminus, BamE is present mainly as a homodimer, while removal of the C-terminus does not provide any definite conclusions (Figure 2.16; Table 2.4). These findings are interesting and contrary to the experiments in Sections 2.2.1 and 2.2.2, where removal of the N-terminus prevented oligomerization with BamC and BamD. In this case, it is possible that the large unstructured N-terminus obstructs BamE from dimerizing, and

thus in its absence, the protein is predominantly found in the dimeric form. However, it is unknown what the significance of the BamE dimer is, and why the N-terminus may play this preventative role.

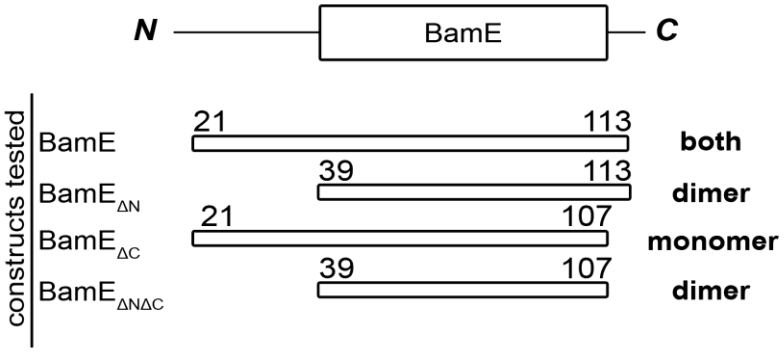

Figure 2.16 Schematic summary of the preferred oligomeric state of each BamE truncation

This is a schematic summary showing which oligomeric state is preferred when either or both termini of BamE are removed. For each construct, the residue boundaries are illustrated and correspond to the domain structure shown above. The preferred oligomeric state is noted on the right, were "both" indicates that the full BamE is present in both forms in similar amounts. Based on this diagram, it can be seen that in the absence of the N-terminus, BamE tends to dimerize, while in the absence of the C-terminus, BamE is mainly present as a monomer.

Table 2.4 Gel-filtration elution volumes for BamE homodimer formation

Combination	Peak Elution Volume (mL)	Measured Mass (kDa)	Protein(s) Eluted and their Calculated Mass (kDa)	
BamE	150	32	BamE dimer	20.6
	204	10	BamE monomer	10.3
BamE$_{\Delta N}$	**163**	**24**	**BamE$_{\Delta N}$ dimer**	**16.4**
	219	7	BamE$_{\Delta N}$	8.2
BamE$_{\Delta C}$	152	31	BamE$_{\Delta C}$ dimer	19.4
	204	**10**	**BamE$_{\Delta C}$ monomer**	**9.7**
BamE$_{\Delta N\Delta C}$	**169**	**21**	**BamE$_{\Delta N\Delta C}$ dimer**	**15.2**
	217	8	BamE$_{\Delta N\Delta C}$ monomer	7.6

This table presents a summary of the peaks observed, their elution volumes, and the proteins eluted. The measured molecular mass was estimated from the standard curve of the gel-filtration column using the corresponding elution volume. The calculated molecular mass of the protein eluted is based on the amino acid composition of each protein. The rows in **bold** represent the majority population observed (monomer or homodimer) for each experiment.

Recently, Albrecht and Zeth solved the crystal structure of the BamE dimer (Figure 2.17) (Albrecht and Zeth, 2011). The structure shows how two BamE monomers open up and form a domain-swapped dimer, where the N-terminal half of one monomer is associated with the C-terminal half of the other monomer. The two monomers are also intertwined in the middle, which reinforces previous observations of the BamE dimer being quite stable and difficult to dissociate (Kim et al., 2011c). Interestingly, when examining the amino acid sequence of the BamE dimer, it should be noted that the researchers also removed the N-terminal region for their crystallization experiment. Their BamE construct is from residues 40-113, similar to the $BamE_{\Delta N}$ discussed in this thesis which has the amino acid boundaries of 39-113. Although the researchers removed the N-terminus specifically for improving the crystallization process, the results above show how this one truncation may have promoted the formation of the BamE dimer in their case as well. As the structure of the full length BamE dimer has yet to be solved, it is difficult to conclude how the N-terminus is involved, and if it is truly interfering with the dimerization process.

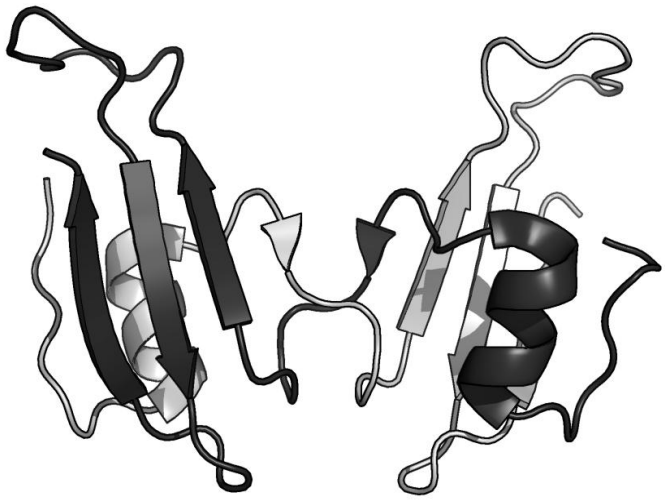

Figure 2.17 Structure of the BamE homodimer

The structure of the BamE homodimer shows two BamE monomers to form a domain-swapped dimer (PDB: 2YH9). Note that in this structure, the N-terminus of BamE was removed to assist with the crystallization process. Based on the gel-filtration studies presented in this chapter, removal of the N-terminus promotes dimerization, and also prevents the formation of the BamDE and BamCDE subcomplexes. This questions what the purpose of the BamE homodimer may be, and if it is biologically significant.

2.3. Conclusions

This chapter presented the experiments that were conducted to investigate the role of BamE in three different protein subcomplexes: BamE-BamE homodimer, BamD-BamE heterodimer, and BamC-BamD-BamE heterotrimer. Based on the available NMR structure of BamE, the N- and C-terminal regions were hypothesized to be important for protein-protein interactions due to their flexibility. BamE was truncated to produce variations that were lacking the N-terminus ($BamE_{\Delta N}$), the C-terminus ($BamE_{\Delta C}$) or both termini ($BamE_{\Delta N\Delta C}$). Using gel-filtration chromatography to detect the presence of complex formation, these constructs were co-purified with the different interacting partners to see if oligomerization still occurred.

The results clearly demonstrate that the N-terminus of BamE is required for the formation of the BamCDE subcomplex, while both the N- and C-termini are required for BamDE formation. However, these results are insufficient to conclude if the required termini are directly involved in the interaction, or if their absence leads to significant disruption of the overall BamE structure, thus preventing complex formation. If these regions were directly involved, then it could be speculated that the C-terminus of BamE interacts with a region of BamD that is independent of the BamC-BamD interface as its absence appears to have no effect on the formation of the BamCDE subcomplex. Meanwhile, the N-terminus of BamE could associate with a region of BamD that is near the BamC-BamD interface, and a tight interaction between the three proteins would therefore allow the BamCDE subcomplex to stay intact when the C-terminus of BamE is removed. A simple model of this could be visualized as the unstructured termini of BamE being two arms that “hang onto” BamD, with both termini supporting each other to form a stable BamDE dimer. In the presence of BamC, the BamC-BamD interaction interface could be overlapping with the N-terminal “arm” of BamE, thus allowing BamE to “hang onto” the complex with one “arm” even if the C-terminal “arm” is absent. Of course, to truly understand the nature of these interactions, having a structure of either the BamDE or BamCDE complexes would accurately illustrate how BamE fits in.

BamE is the only lipoprotein within the BAM complex that has shown to oligomerize on its own. Its ability to form a homodimer has raised several questions such as which oligomeric state is preferred, what causes the dimerization, and what the

significance is of the homodimer. The results presented in this chapter show that removal of the N-terminus of BamE promotes formation of the homodimer. The dimer structure provided by Albrecht and Zeth is also a result of removing the N-terminus (Albrecht and Zeth, 2011). Whether the N-terminus plays an obstructive role or not, the N-terminal region has a significant impact on the oligomeric state. Thus, one would begin to speculate that the ability for BamE to dimerize may be a regulatory factor for BamE's function. However, as presented by previous research, the two forms of BamE are not in equilibrium as the dimeric form cannot be dissociated even under extreme pH and salt conditions (Kim et al., 2011c). Thus, the notion of BamE being regulated through these means is not convincing.

It could also be possible that the BamE homodimer is the active form of BamE that is then able to assemble with the rest of the BAM complex members. However, the formation of the BamE homodimer is more prominent in the absence of the N-terminus, which is necessary for further formation of the BamDE and BamCDE subcomplexes. Thus the BamE homodimer, especially the $BamE_{\Delta N}$ or $BamE_{\Delta N \Delta C}$ homodimers, cannot be the active form. In addition, the molecular mass of the observed BamCDE subcomplex corresponds to 1:1:1 ratio between the three proteins, suggesting that the monomeric form of BamE is the active form.

So if the BamE homodimer is neither the regulatory form nor the active form of BamE, what is its significance? It is possible that the dimer is a by-product of over-expression in the cytoplasm leading to intertwined domain-swapped dimers. This theory is supported by evidence from Knowles et al. whose team showed that extraction of BamE from its native periplasmic compartment yielded the presence of only monomeric BamE, while extraction from the whole cell yielded both monomeric and dimeric species (Knowles et al., 2011). The researchers propose that over-expression in the cytoplasm, and also at high temperatures, could cause BamE to form the dimeric "misfolded aggregate" (Knowles et al., 2011). In contrast, Albrecht and Zeth show BamE isolated from the membrane to also exist as oligomers (Albrecht and Zeth, 2011). The research conducted for this thesis project involved extraction of BamE from whole cells, with results showing the presence of both dimeric and monomeric species for full length BamE. Thus, based on previous research and results presented here, it is highly possible that the BamE homodimer is not biologically significant, but instead just a by-

product of the protein expression method. To further investigate this, different temperatures and growth techniques can be tested for BamE protein expression to see which oligomeric form is preferred under different conditions. Also, denaturing the BamE samples and refolding can give insight into if the BamE dimer (or “misfolded aggregate”) can be refolded into the monomeric form.

In summary, this chapter provides an overview of how protein engineering and gel-filtration chromatography were used to determine the involvement of BamE in various protein-protein interactions. Based on the results it can be proposed that BamE is a monomeric protein which, under certain conditions, can form a homodimeric species. The active monomeric form is involved in forming a BamDE dimer in a 1:1 ratio, as well as a BamCDE subcomplex in a 1:1:1 ratio. While the N-terminus of BamE has been identified as a requirement for BamCDE formation, both termini are needed for BamDE association. This could suggest that the N-terminus of BamE binds near the BamC-BamD interface, resulting in a tighter association that allows the absence of the C-terminal domain of BamE. For a better understanding of the role of the N-terminus in these interactions, a structure of the BamCDE subcomplex would be ideal.

3. The unstructured N-terminus of BamC is required for the formation of the BamCDE subcomplex

Portions of this chapter have been published in:

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 crystallographica. F67, 1350–1358.

Kim, K. H., Aulakh, S., and Paetzel, M. (2011). Crystal Structure of β-Barrel Assembly Machinery BamCD Protein Complex. The Journal of Biological Chemistry 286, 39116–39121.

3.1. Strategy

This chapter focuses on identifying which regions of BamC are important for interacting with BamD and BamE. Initial studies using full length versions of these proteins showed that BamC can interact to form BamCD and BamCDE subcomplexes (see sections 2.2.1.1 and 2.2.2.1 below). However, BamC is unable to form a complex with BamE alone (see Figure C11). Based on available structural data, different versions of BamC were constructed and combined with the other lipoproteins to see the affect on BamCD and BamCDE formation. After identifying a region of BamC involved in interaction, competition analyses were performed in an attempt to confirm these findings.

3.1.1. Available structural information

At the beginning of this project, the structure of BamC was unknown. The only structural information available was based on the secondary structure prediction program PSIPRED, and the findings from Knowles et al. who were able to predict the secondary structure of BamC using NMR assignments of the protein backbone (Jones, 1999; Buchan et al., 2010; Knowles et al., 2009). Results from both suggest BamC to have a long unstructured region at the N-terminus, up until at least the 95th amino acid residue, which would contribute to ~70 residues in the final mature protein lacking the N-terminal signal sequence (Figure 3.1A-B). Sequence alignment of BamC from several Gram-negative species shows the most number of conserved residues to exist in this unstructured region, suggesting the N-terminus to play an important role (Figure 3.1C). This structural information, alongside the results from proteolysis studies (see Section 3.1.2), was used to then create variations of BamC for interaction analysis.

3.1.2. Limited proteolysis studies

Working with BamC posed one challenge – the protein would begin to degrade within a week, even when stored at 4°C. Thus, purification and subsequent experiments would have to be scheduled accordingly. However, rather than observing a degradation ladder of multiple bands on SDS-PAGE, BamC was observed to break down into two distinct fragments (see Figure C4). To investigate this further, a limited proteolysis experiment was set up where chymotrypsin was added to a sample of BamC and digestion was monitored at several time points. Results show that after overnight incubation with chymotrypsin, the 34.3 kDa BamC was fully cleaved into a ~25 kDa product (Figure 3.2A). To study whether this digestion occurred from the N- or C-terminus, the cleaved product was analyzed through nickel affinity chromatography. Since the full length BamC had an N-terminal hexahistidine tag, inability for the cleaved product to be retained on the Ni^{2+}-NTA resin would suggest that the tag had been removed and cleavage occurred from the N-terminus. Interestingly, before nickel affinity chromatography was carried out, the ~25 kDa BamC had digested further into a ~12 kDa product. Nonetheless, results from this analysis showed that the digested BamC was cleaved from the N-terminus, as both fragments were unable to stick to the resin causing them to come out in the flow-through (Figure 3.2B) (Kim et al., 2011b).

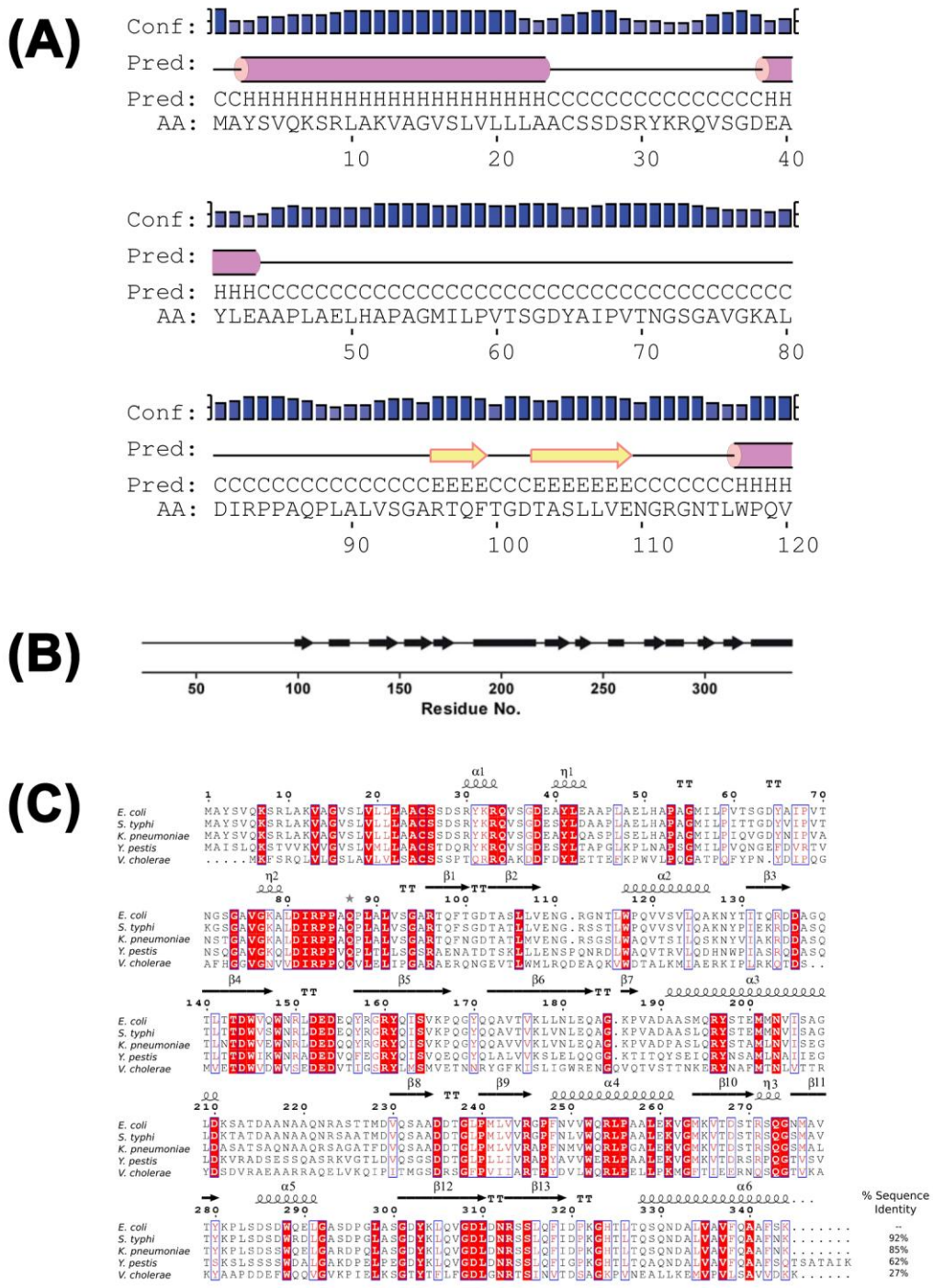

Figure 3.1 Predicted secondary structure and sequence alignment of BamC

This figure shows the predicted secondary structure of BamC that was used to initiate the interaction studies. (A) shows a portion of the results from the PSIPRED server, which predicts a small helix within a region of coiled-coils at the N-terminus of BamC. Note that the first helix from residues 3-23 is within the N-terminal signal sequence which is cleaved and not present in the mature protein. To see the results for the entire length of BamC, please see Figure C3. (B) is an

adapted figure from Knowles et al., which shows their secondary structure prediction also suggesting an unstructured region at the N-terminus (Knowles et al., 2009). (C) shows the sequence alignment of *E. coli* BamC (UniProt ID: P0A903) with homologues from *Salmonella typhi* (Q83T79), *Klebsiella pneumoniae* (B5XVM8), *Yersinia pestis* (D1TUX3), and *Vibrio cholerae* (Q9KQ48). Red boxes show absolutely conserved residues, red text shows similar residues, and blue boxes show stretches of similar residues. Note that the secondary structure displayed above the sequence is based on the current up to date structures which were not available at the beginning of this project. More information about those structures is provided in Sections 3.1.2 and 3.3.

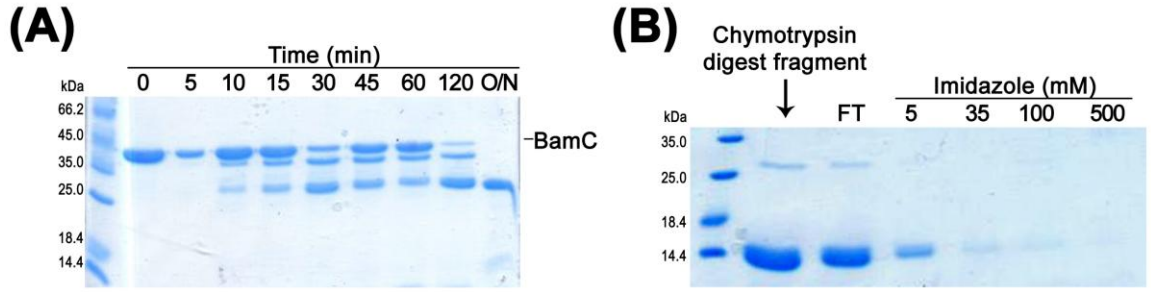

Figure 3.2 Chymotrypsin digest of BamC

This figure shows the cleavage pattern of BamC in the presence of chymotrypsin. (A) shows that after overnight digestion with chymotrypsin, the 34.3 kDa BamC is cleaved into a ~25 kDa fragment. (B) shows that the ~25 kDa fragment was further digested to a ~12 kDa product. When this sample was analyzed through a nickel affinity chromatography column, both fragments were unable to stick to the nickel resin, suggesting that the cleavage occurred from the N-terminus. FT = flow through.

These results proposed BamC to have three independently "folding" domains. One domain could be ~10 kDa and more susceptible to degradation, causing the 34.3 kDa to ~25 kDa shift seen in the overnight chymotrypsin digest. The other two domains could be approximately the same size at ~12 kDa, which would explain the ~25 kDa to ~12 kDa cleavage observed later. Based on the secondary structure information from Section 3.1.1, the ~10 kDa domain would most likely correspond to the unstructured N-terminus. The unstructured nature of this region would also account for why this ~10 kDa fragment was not observed on the SDS-PAGE gel, as it would have been completely obliterated by the protease.

Coincidentally, similar cleavage patterns were observed by Albrecht and Zeth when they subjected BamC (lacking the unstructured N-terminal domain) to subtilisin digest (Albrecht and Zeth, 2010). The resulting two domains were named the N-terminal and C-terminal domains (known hereafter as $BamC_N$ and $BamC_C$, respectively), which the researchers then purified separately for crystallization (Albrecht and Zeth, 2010). Their crystal structures show these two domains to have a similar structure despite difference in sequence, and they are proposed to be linked together by a flexible linker (Albrecht and Zeth, 2011). Our research in the Paetzel Lab also yielded a crystal structure of BamC (PDB: 3SNS) (Kim et al., 2011b). Interestingly, while full length BamC was used for crystallization, what remained in the crystal over time was only $BamC_C$. This confirmed the earlier observations that BamC degrades into its individual domains in the absence of exogenous proteases. This $BamC_C$ crystal structure with the findings from the chymotrypsin digest was published in the journal *Acta Crystallographica* in 2011 (Kim et al., 2011b).

3.1.3. BamC variations constructed

Based on the structural data and sequence alignment, all three domains contain patches of conserved residues with the majority existing in the unstructured N-terminus (known hereafter as $BamC_U$). To study which of these regions are important for interaction, the domains were cloned out and purified individually. In the case of $BamC_U$, due to the lack of secondary structure, cloning and expressing this region by itself proved challenging and thus a $BamC_{UN}$ protein was constructed with contained both the $BamC_U$ and $BamC_N$ regions (see Figure C9). Another construct, $BamC_{NC}$ was created to see the affect of complex formation when the unstructured region was absent. A list of all the protein constructs used in this experiment with their amino acid boundaries and molecular mass is provided in Table 3.1. Note that the full length proteins (BamC, BamD, and BamE) are missing the N-terminal signal sequence (residues 1-24 for BamC; residues 1-19 for BamD and BamE) and lipidation site (Cys25 for BamC; Cys20 for BamD and BamE) to allow over-expression in the cytoplasm. All constructs have an N-terminal hexahistidine tag, except for $BamC_{UN}$, $BamC_N$, and $BamC_C$ which have a C-terminal tag. Further cloning details of these constructs are available in Appendix A.

Table 3.1 List of constructs used to study BamC interaction

Construct Name	Residue Boundaries	Molecular Mass (kDa)
BamC	26-344	34.3
$BamC_{UN}$	26-217	20.7
$BamC_{N}$	99-217	13.2
$BamC_{C}$	220-344	13.4
$BamC_{NC}$	94-344	27.3
BamD	21-245	25.7
BamE	21-113	10.3

The molecular mass is calculated based on the amino acid composition of the protein construct.

To study which regions of BamC may be involved in interaction, the various BamC constructs were co-purified with full length BamD and BamE to see the effect on forming the BamCD or BamCDE subcomplexes. Successful interaction was defined by the elution of intact complexes during gel-filtration chromatography.

3.2. Gel-filtration chromatography studies

3.2.1. BamCD dimer formation

3.2.1.1. Full length BamC

When run individually on gel-filtration chromatography, BamC elutes at ~125 mL and BamD at ~155 mL. Based on the standard curve for the column used (for standard curve see Figure A1), the elution volume for BamD corresponds to a molecular mass of ~29 kDa which is close to its calculated mass of 25.7 kDa. However, the elution volume of BamC corresponds to a mass of ~55kDa which is distant from its calculated mass of 34.3 kDa. As explained in Section 3.1, there is a large unstructured region at the N-terminus ($BamC_{U}$), which is the likely cause of BamC to elute much earlier than expected. For the individual chromatograms of these proteins, please see Appendix C.

To see if a BamCD dimer could form, cell pellets containing full length BamC were co-lysed and purified with a pellet containing BamD. The resulting gel-filtration chromatogram shows a new peak eluting at ~121 mL, which suggests the formation of a

complex (Figure 3.3). Based on the standard curve, this new peak corresponds to a molecular mass value of ~60 kDa which is the same as the calculated mass of 60.0 kDa for the BamCD dimer, with a 1:1 ratio. Fractions collected from this peak were analyzed on an SDS-PAGE gel, and the results showed the presence of both proteins, confirming that a BamCD subcomplex was successfully formed.

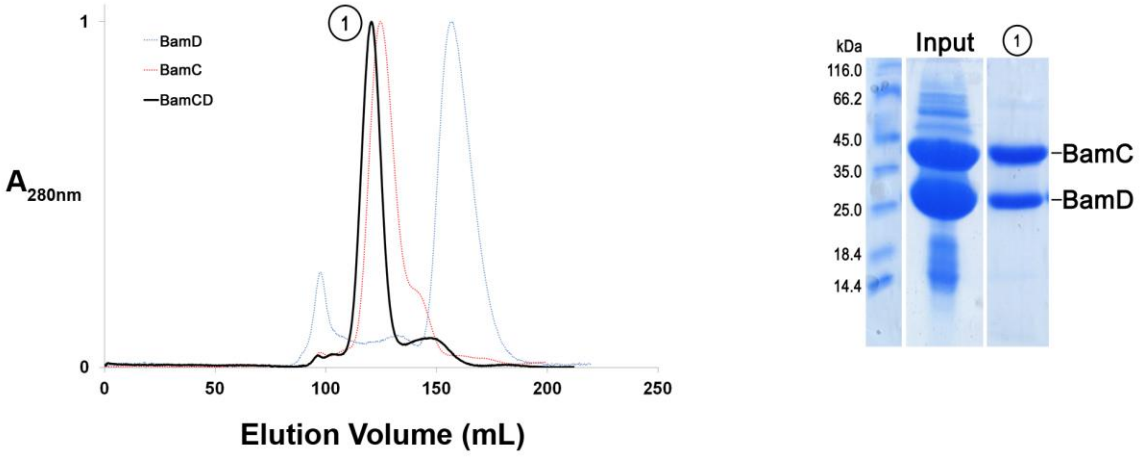

Figure 3.3 Gel-filtration chromatogram and SDS-PAGE showing BamCD dimer formation with full length BamC and BamD

The gel-filtration chromatogram (left) shows the elution profile for the sample containing BamC and BamD (black curve), which is plotted as the relative UV absorbance at 280nm (A_{280nm}, provided in arbitrary units) for protein eluted at a specific volume (Elution Volume, provided in millilitres (mL)). The results show the presence of one peak which is labelled as 1. This peak elutes earlier than individual BamC (dotted red curve) and BamD (dotted blue curve), suggesting that they formed a complex. The corresponding SDS-PAGE gel (right) confirms this by showing the presence of both BamC and BamD in the peak. Overall these results show that a BamCD dimer forms with full length BamC and BamD. Experiment notes: a HiPrep 26/60 Sephacryl S-100 HR gel-filtration column was used attached to the ÄKTAprime system, with a void volume of 98 mL and a column volume of 320 mL. The experiment was run at 4°C with a buffer containing 20 mM Tris-HCl (pH 8.0) and 100 mM NaCl, where the flow rate was set to 1 mL/min, with 3 mL fractions collected. Fractions eluting in or around the void volume are not shown on the SDS-PAGE gel as they were not in the region of interest and contained aggregated protein.

3.2.1.2. N-terminal domain ($BamC_N$)

To study the involvement of the N-terminal domain of BamC, this region was cloned and expressed individually, as outlined in Section 3.1.3. This $BamC_N$ construct was purified with full length BamD to see if the BamCD dimer could still form. The resulting gel-filtration chromatogram shows the presence of two peaks eluting at ~150 mL and ~175 mL (Figure 3.4). Based on the standard curve, these peaks correspond to mass values of ~32 kDa and ~19 kDa, respectively. Comparing this to the calculated mass of BamD (25.7 kDa) and $BamC_N$ (13.2 kDa), it appears that the two proteins eluted out separately. SDS-PAGE analysis further confirms that no BamCD dimer was able to form with the $BamC_N$ domain alone.

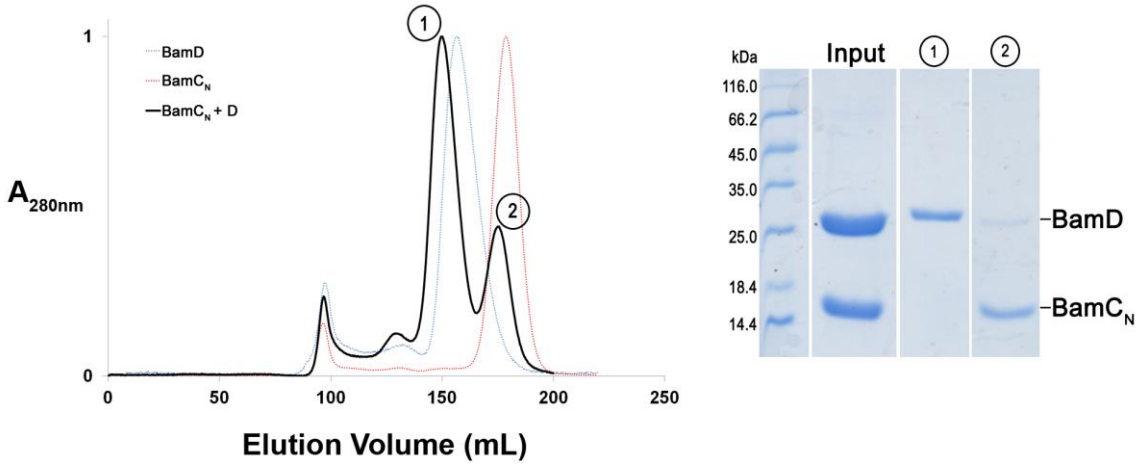

Figure 3.4 Gel-filtration chromatogram and SDS-PAGE showing no BamCD dimer formation with $BamC_N$

The gel-filtration chromatogram (left) shows the elution profile for the sample containing $BamC_N$ and BamD (black curve), which is plotted as the relative UV absorbance at 280nm (A_{280nm}, provided in arbitrary units) for protein eluted at a specific volume (Elution Volume, provided in millilitres (mL)). The results shows the presence of two main peaks which are labelled as 1 and 2. These peaks have similar elution volumes as the individual $BamC_N$ (dotted red curve) and BamD (dotted blue curve) proteins, suggesting that they were unable to form a complex. The corresponding SDS-PAGE gel (right) confirms this by showing the presence of only BamD in peak 1, and mainly $BamC_N$ in peak 2 (the faint BamD band in this lane is most likely due the close proximity of the two peaks, leading to some BamD elution in the second peak). Overall these results suggest that the BamCD dimer is unable to form in the presence of only the $BamC_N$ domain. Experiment notes: a HiPrep 26/60 Sephacryl S-100 HR gel-filtration column was used attached to the ÄKTAprime system, with a void volume of 98 mL and a column volume of 320 mL. The experiment was run at 4°C with a buffer containing 20 mM Tris-HCl (pH 8.0) and 100 mM NaCl, where the flow rate was set to 1 mL/min, with 3 mL fractions collected. Fractions eluting in or around the void volume are not shown on the SDS-PAGE gel as they were not in the region of interest and contained aggregated protein.

3.2.1.3. C-terminal domain ($BamC_C$)

To study the involvement of the C-terminal domain of BamC, this region was cloned and expressed individually, as outlined in Section 3.1.3. This $BamC_C$ construct was purified with full length BamD to see if the BamCD dimer could still form. The resulting gel-filtration chromatogram shows the presence of two peaks eluting at ~151 mL and ~175 mL (Figure 3.5). Based on the standard curve, these peaks correspond to a molecular mass of ~31 kDa and ~19 kDa, respectively. Comparing these estimations to the calculated mass values of BamD (25.7 kDa) and $BamC_C$ (13.4 kDa), it appears that the two proteins eluted out separately. SDS-PAGE analysis further confirms that the BamCD dimer was unable to form with the $BamC_C$ domain alone.

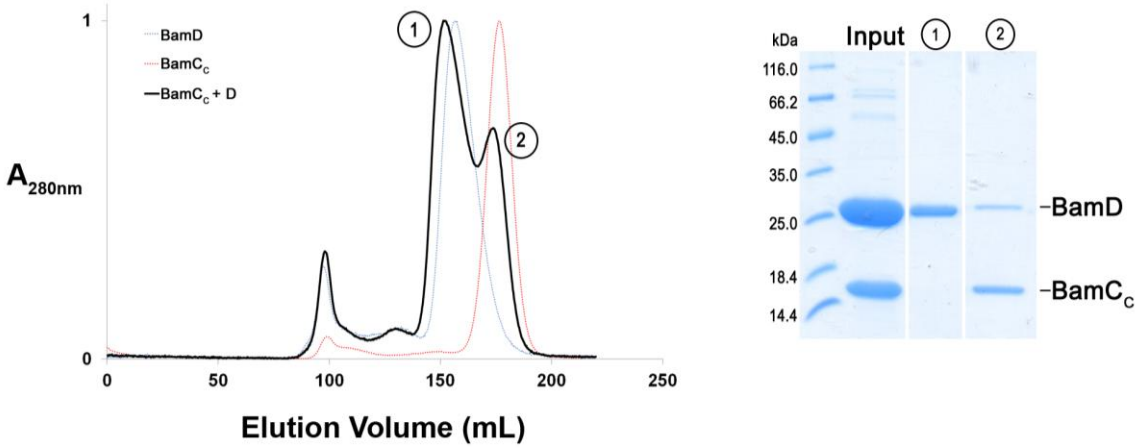

Figure 3.5 Gel-filtration chromatogram and SDS-PAGE showing no BamCD dimer formation with $BamC_C$

The gel-filtration chromatogram (left) shows the elution profile for the sample containing $BamC_C$ and BamD (black curve), which is plotted as the relative UV absorbance at 280nm (A_{280nm}, provided in arbitrary units) for protein eluted at a specific volume (Elution Volume, provided in millilitres (mL)). The results shows the presence of two main peaks which are labelled as 1 and 2. These peaks have similar elution volumes as the individual $BamC_C$ (dotted red curve) and BamD (dotted blue curve) proteins, suggesting that they were unable to form a complex. The corresponding SDS-PAGE gel (right) confirms this by showing the presence of only BamD in peak 1, and mainly $BamC_N$ in peak 2 (the faint BamD band in this lane is due the close proximity of the two peaks, leading to some BamD elution in the second peak). Overall these results suggest that the BamCD dimer is unable to form in the presence of only the $BamC_C$ domain. Experiment notes: a HiPrep 26/60 Sephacryl S-100 HR gel-filtration column was used attached to the ÄKTAprime system, with a void volume of 98 mL and a column volume of 320 mL. The experiment was run at 4°C with a buffer containing 20 mM Tris-HCl (pH 8.0) and 100 mM NaCl, where the flow rate was set to 1 mL/min, with 3 mL fractions collected. Fractions eluting in or around the void volume are not shown on the SDS-PAGE gel as they were not in the region of interest and contained aggregated protein.

3.2.1.4. N- and C-terminal domains ($BamC_{NC}$)

The previous results have shown that the N- and C-terminal domains alone cannot form the BamCD dimer. Thus it was proposed that both domains may be required for the association. To test this hypothesis, the region of BamC corresponding to these two domains was cloned and expressed. This $BamC_{NC}$ construct was purified with full length BamD to see if the BamCD dimer could still form. The resulting gel-filtration chromatogram shows the presence of two peaks eluting at ~142 mL and ~159 mL (Figure 3.6). Based on the standard curve, these peaks correspond to molecular mass values of ~38 kDa and ~27 kDa, respectively. Comparing this to the calculated mass of $BamC_{NC}$ (27.3 kDa) and BamD (25.7 kDa), it appears that the two proteins eluted out separately. Note that there is a flexible linker in between the $BamC_N$ and $BamC_C$ domains, which may be causing the $BamC_{NC}$ protein to elute earlier than expected. SDS-PAGE analysis further confirms that the BamCD dimer was unable to form with only the N- and C-terminal domains of BamC.

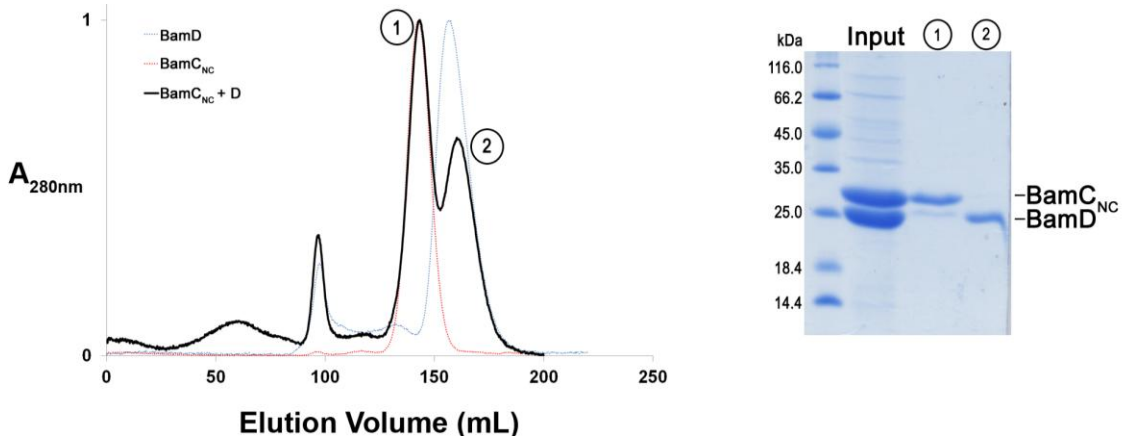

Figure 3.6 Gel-filtration chromatogram and SDS-PAGE showing no BamCD dimer formation with $BamC_{NC}$

The gel-filtration chromatogram (left) shows the elution profile for the sample containing $BamC_{NC}$ and BamD (black curve), which is plotted as the relative UV absorbance at 280nm (A_{280nm}, provided in arbitrary units) for protein eluted at a specific volume (Elution Volume, provided in millilitres (mL)). The results show the presence of two main peaks which are labelled as 1 and 2. These peaks have similar elution volumes as the individual $BamC_{NC}$ (dotted red curve) and BamD (dotted blue curve) proteins, suggesting that they were unable to form a complex. The corresponding SDS-PAGE gel (right) confirms this by showing the presence of mainly $BamC_{NC}$ in peak 1, and only BamD in peak 2 (the faint BamD band in lane for peak 1 is due the close proximity of the two peaks, leading to some BamD elution in the first peak). Overall these results suggest that the BamCD dimer is unable to form in the presence of only the $BamC_{NC}$ domains. Experiment notes: a HiPrep 26/60 Sephacryl S-100 HR gel-filtration column was used attached to the ÄKTAprime system, with a void volume of 98 mL and a column volume of 320 mL. The

experiment was run at 4°C with a buffer containing 20 mM Tris-HCl (pH 8.0) and 100 mM NaCl, where the flow rate was set to 1 mL/min, with 3 mL fractions collected. Fractions eluting in or around the void volume are not shown on the SDS-PAGE gel as they were not in the region of interest and contained aggregated protein.

3.2.1.5. Unstructured N-terminus with the N-terminal domain ($BamC_{UN}$)

The results thus far have shown that the N- and C-terminal domains of BamC are unable to form the BamCD dimer when they are present alone or together. In all of those experiments, the unstructured N-terminus, $BamC_U$, was absent suggesting that this region may be necessary for the interaction. As outlined in Section 3.1.3, the $BamC_{UN}$ construct was designed to study this region's involvement as expression of $BamC_U$ alone was difficult. Consistent with previous experiments, the $BamC_{UN}$ construct was also purified with full length BamD to see if the BamCD dimer could still form. The resulting gel-filtration chromatogram shows the presence of one large peak eluting at ~126 mL (Figure 3.7). Based on the standard curve, this peak corresponds to a molecular mass of ~54 kDa, which is comparable to the calculated mass of 46.4 kDa for the $BamC_{UN}D$ dimer. Note that this dimer elutes earlier indicating a larger molecular mass of ~54 kDa due to the presence of the unstructured $BamC_U$ region. Running the fractions from this peak on SDS-PAGE further confirms the successful formation of the $BamC_{UN}D$ dimer. These results suggest that the unstructured $BamC_U$ region plays an important role in associating with BamD.

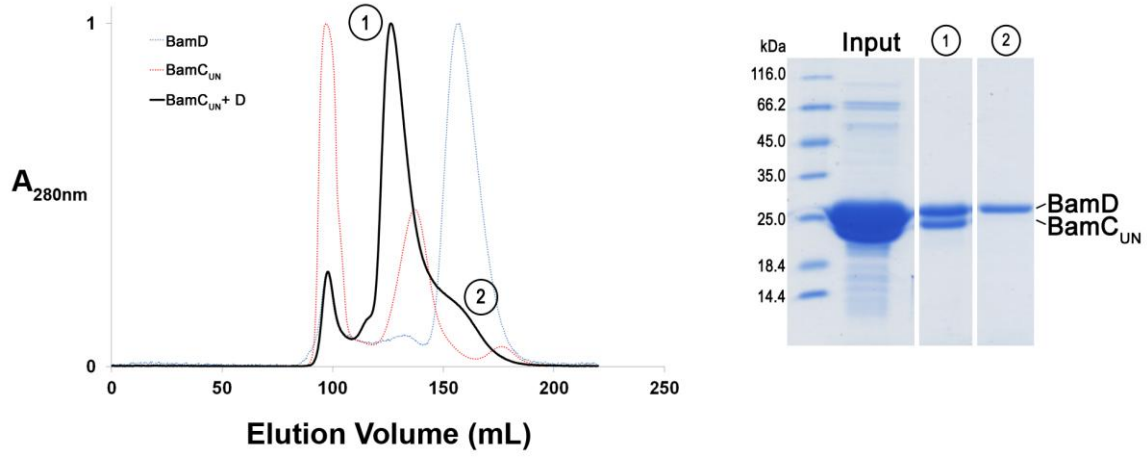

Figure 3.7 Gel-filtration chromatogram and SDS-PAGE showing BamCD dimer formation with $BamC_{UN}$

The gel-filtration chromatogram (left) shows the elution profile for the sample containing $BamC_{UN}$ and BamD (black curve), which is plotted as the relative UV absorbance at 280nm (A_{280nm}, provided in arbitrary units) for protein eluted at a specific volume (Elution Volume, provided in

millilitres (mL)). The results show the presence of one peak and a shoulder which are labelled as 1 and 2. Peak 1 elutes much earlier than individual $BamC_{UN}$ (dotted red curve) and BamD (dotted blue curve), suggesting that they formed a complex. The corresponding SDS-PAGE gel (right) confirms this by showing the presence of both $BamC_{UN}$ and BamD in peak 1, and excess BamD in shoulder 2. Overall these results strongly suggest that the $BamC_U$ region is important for BamCD dimer formation, as previous experiments lacking this region were unable to form the dimer. Note that for the individual $BamC_{UN}$ chromatogram (dotted red curve), a large peak is observed at the void volume, which is due to the presence of other high molecular weight proteins in the input sample. More details about that chromatogram are provided in Figure C8. Experiment notes: a HiPrep 26/60 Sephacryl S-100 HR gel-filtration column was used attached to the ÄKTAprime system, with a void volume of 98 mL and a column volume of 320 mL. The experiment was run at 4°C with a buffer containing 20 mM Tris-HCl (pH 8.0) and 100 mM NaCl, where the flow rate was set to 1 mL/min, with 3 mL fractions collected. Fractions eluting in or around the void volume are not shown on the SDS-PAGE gel as they were not in the region of interest and contained aggregated protein.

3.2.1.6. Summary of BamC requirements for BamCD formation

To study which domains of BamC are involved in the formation of the BamCD dimer, the domains were cloned and expressed individually. Co-purification with full length BamD using gel-filtration chromatography showed that in the absence of the unstructured $BamC_U$ region, no BamCD dimer was observed (Figure 3.8; Table 3.2). This suggested that either $BamC_U$ is required for the association, or BamC must be fully intact for dimerization. To investigate the role of the $BamC_U$ region, the $BamC_{UN}$ construct was co-purified with full length BamD, revealing the successful formation of a $BamC_{UN}D$ dimer. These results support the hypothesis that the $BamC_U$ region plays an important role in the BamCD dimer interaction. However, based on these results alone it cannot be concluded if $BamC_U$ is directly involved in interaction, or if its role is to stabilize the $BamC_N$ domain which could be the main interacting domain. Fortunately, recent structural data provide more insight into this interaction and resulted in a publication, which is discussed in Section 3.3.

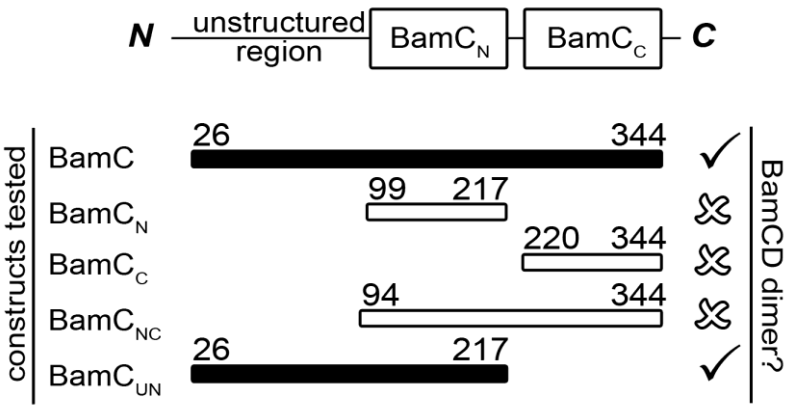

Figure 3.8 Schematic summary of BamC requirements for BamCD dimer formation

This is a schematic summary showing the BamC requirements for BamCD dimer formation. For each construct, the residue boundaries are illustrated and correspond to the domain structure shown above. Successful dimer formation is indicated by a black check mark, while no complex formation is indicated by a white X-mark. Based on this diagram, it can be seen that in the absence of unstructured N-terminus of BamC ($BamC_U$), no BamCD dimer is able to form, suggesting that it is required for the association.

Table 3.2 Gel-filtration elution volumes for BamCD formation

Combination	Peak Elution Volume (mL)	Measured Mass (kDa)	Protein(s) Eluted and their Calculated Mass (kDa)	
BamC	125	55	BamC	34.3
BamD	155	29	BamD	25.7
BamC + BamD	121	60	BamCD dimer	60.0
$BamC_N$ + BamD	150	32	BamD	25.7
	175	19	$BamC_N$	13.2
$BamC_C$ + BamD	151	31	BamD	25.7
	175	19	$BamC_C$	13.4
$BamC_{NC}$ + BamD	142	38	$BamC_{NC}$	27.3
	159	27	BamD	25.7
$BamC_{UN}$ + BamD	126	54	$BamC_{UN}$D dimer	46.4

This table presents a summary of the peaks observed, their elution volumes, and the proteins eluted. The measured molecular mass was estimated from the standard curve of the gel-filtration column using the corresponding elution volume. The calculated molecular mass of the protein eluted is based on the amino acid composition of each protein.

3.2.2. BamCDE subcomplex formation

3.2.2.1. Full length BamC

When run individually on gel-filtration chromatography, BamC elutes at ~125 mL, BamD at ~155 mL, the BamE dimer at ~150 mL, and the BamE monomer at ~204 mL (for more information about the oligomeric state of BamE, see Chapter 2). Based on the standard curve for the column used, the elution volume for BamC corresponds to a molecular mass value of ~55 kDa which is comparable to its calculated mass of 34.3 kDa with consideration of its unstructured N-terminus ($BamC_U$). The elution volume of BamD corresponds to a mass of ~29 kDa which is close to its calculated mass of 25.7 kDa. As discussed in Chapter 2, the extended nature of the BamE dimer causes it to elute much earlier, implying a mass of ~32 kDa when its calculated mass is actually 20.6 kDa. And finally, the elution volume of the BamE monomer corresponds to a mass of ~10 kDa, which is very close to its calculated mass of 10.3 kDa.

As presented in Section 2.2.1.1, to see if a complex would form with all three proteins, individual cell pellets containing full length BamC, BamD, and BamE were co-lysed and purified using gel-filtration chromatography. The resulting chromatogram showed successful formation of the BamCDE subcomplex with a 1:1:1 ratio (Figure 2.2). More discussion of this result is provided in Section 2.2.1.1.

3.2.2.2. N-terminal domain ($BamC_N$)

To study the involvement of the N-terminal domain of BamC, this region was cloned and expressed individually, as outlined in Section 3.1.3. This $BamC_N$ construct was purified with full length BamD and BamE to see if the BamCDE subcomplex could still form. The resulting gel-filtration chromatogram shows the presence of two major peaks with a broad shoulder following the first peak (Figure 3.9). The first peak has an elution volume of ~129 mL which corresponds to a mass of ~50 kDa. Based on the chromatogram alone, it would seem that a $BamC_NDE$ subcomplex had formed, which would have an expected calculated molecular mass of ~49 kDa. However, based on SDS-PAGE, it can be seen that only BamD and BamE are present in this peak, while no band corresponding to $BamC_N$ is visible. Thus, this first peak corresponds to the BamDE dimer (36.0 kDa) which is eluting out slightly earlier due to the unstructured N-terminus of BamE.

The broad shoulder following the first peak (labelled as 2 in the chromatogram) is where BamD and dimeric BamE elute individually. The second peak (labelled as 3 in the chromatogram) is where $BamC_N$ elutes at ~180 mL, corresponding to a molecular mass of ~17 kDa, which is close to the calculated mass of 13.2 kDa for $BamC_N$. From the SDS-PAGE it can be seen that monomeric BamE emerges soon after, creating a slight shoulder on the peak (labelled as 4 in the chromatogram). Interestingly, while the majority of BamD elutes as a part of the BamDE dimer, or individually around ~155 mL, considerable amounts of BamD are still present well past 200 mL. This is most likely caused by excess BamD which may be non-specifically binding to the chromatography resin and eluting much later. Nonetheless, the purpose of this experiment was to see if a complex could form with $BamC_N$. The results presented here show that the BamCDE subcomplex is unable to form, while the BamDE dimer appears to be unaffected.

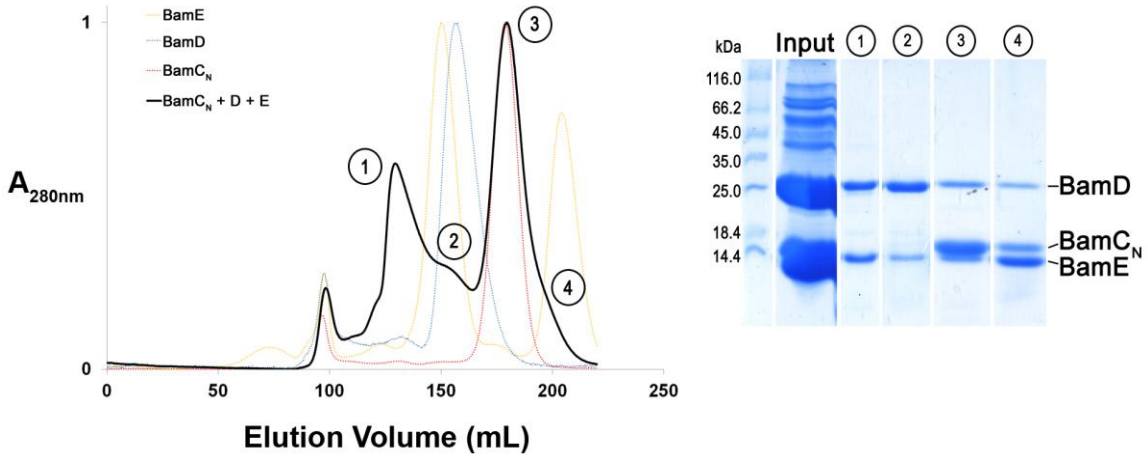

Figure 3.9 Gel-filtration chromatogram and SDS-PAGE showing no BamCDE subcomplex formation with $BamC_N$

The gel-filtration chromatogram (left) shows the elution profile for the sample containing $BamC_N$, BamD, and BamE (black curve), which is plotted as the relative UV absorbance at 280nm (A_{280nm}, provided in arbitrary units) for protein eluted at a specific volume (Elution Volume, provided in millilitres (mL)). The results show the presence of three peaks which are labelled as 1, 2, and 3, as well as a tail following peak 3 which is labelled as region 4. Regions 2-4 have similar elution volumes as individual $BamC_N$ (dotted red curve), BamD (dotted blue curve), and BamE (dotted yellow curve) suggesting that the proteins eluted out separately and were unable to form a complex. In addition, peak 1 elutes much earlier at ~129 mL, suggesting a BamDE dimer (see Chapter 2). The corresponding SDS-PAGE gel (right) confirms this by showing the presence of BamD and BamE in peak1. The lane for peak 2 shows mainly BamD with some BamE, suggesting BamD and dimeric BamE to be eluting separately. In peak 3, BamD is still emerging, along with monomeric $BamC_N$ and some faint BamE. Finally, region 4 shows monomeric BamE eluting, while traces of $BamC_N$ and BamD are still appearing. The presence of multiple bands in each lane is due to overlapping peaks. Overall these results suggest that the BamCDE

subcomplex is unable to form in the presence of only the $BamC_N$ domain, however, the BamDE dimer is still able to form. Experiment notes: a HiPrep 26/60 Sephacryl S-100 HR gel-filtration column was used attached to the ÄKTAprime system, with a void volume of 98 mL and a column volume of 320 mL. The experiment was run at 4°C with a buffer containing 20 mM Tris-HCl (pH 8.0) and 100 mM NaCl, where the flow rate was set to 1 mL/min, with 3 mL fractions collected. Fractions eluting in or around the void volume are not shown on the SDS-PAGE gel as they were not in the region of interest and contained aggregated protein.

3.2.2.3. C-terminal domain ($BamC_C$)

To study the involvement of the C-terminal domain of BamC, this region was cloned and expressed individually, as outlined in Section 3.1.3. This $BamC_C$ construct was purified with full length BamD and BamE to see if the BamCDE subcomplex could still form. Similar to the observations with $BamC_N$ (Section 3.2.2.2), the resulting gel-filtration chromatogram shows the presence of three peaks with a broad shoulder following the first peak (Figure 3.10). The first peak has an elution volume of ~133 mL which corresponds to a molecular mass of ~46 kDa. The SDS-PAGE gel shows the presence of only BamD and BamE, suggesting that the BamDE dimer (36.0 kDa) was able to form. The shoulder following this peak is where BamD and dimeric BamE elute individually. $BamC_C$ elutes in the second peak at ~174 mL, suggesting a mass of ~19 kDa which is close the calculated mass of 13.4 kDa for $BamC_C$. The third peak slightly overlaps with the previous one, and has an elution volume of ~193 mL, corresponding to a mass of ~13 kDa. This is where the monomeric BamE emerges with a calculated mass of 10.3 kDa. Interestingly, again as observed in the $BamC_N$ experiment (Section 3.2.2.2), BamD traces are visible late into the chromatogram, although the exact cause of this is unknown. Based on these results it can be concluded that the BamCDE subcomplex is unable to form with $BamC_C$ alone. Meanwhile, the BamDE dimer is still able to form.

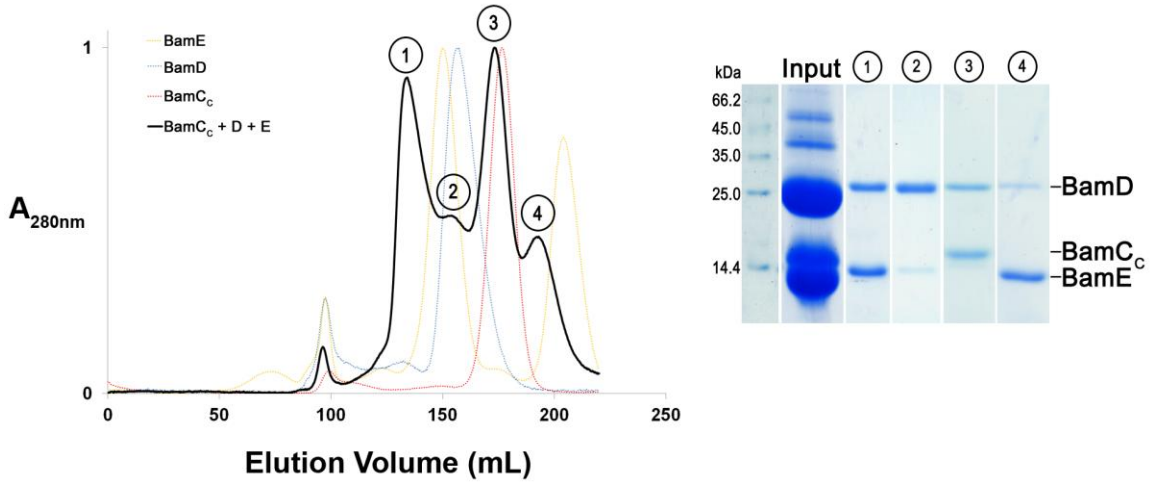

Figure 3.10 Gel-filtration chromatogram and SDS-PAGE showing no BamCDE subcomplex formation with BamC$_C$

The gel-filtration chromatogram (left) shows the elution profile for the sample containing BamC$_C$, BamD, and BamE (black curve), which is plotted as the relative UV absorbance at 280nm (A_{280nm}, provided in arbitrary units) for protein eluted at a specific volume (Elution Volume, provided in millilitres (mL)). The results show the presence of four peaks which are labelled as 1, 2, 3, and 4. Peaks 2-4 have similar elution volumes as individual BamC$_C$ (dotted red curve), BamD (dotted blue curve), and BamE (dotted yellow curve) suggesting that the proteins eluted out separately and were unable to form a complex. In addition, peak 1 elutes much earlier at ~133 mL, suggesting a BamDE dimer (see Chapter 2). The corresponding SDS-PAGE gel (right) confirms this by showing the presence of BamD and BamE in peak1. The lane for peak 2 shows mainly BamD with faint BamE, suggesting the elution of individual BamD. In peak 3, BamC$_C$ elutes separately, with some BamD from the previous overlapping peak. Finally, peak 4 shows monomeric BamE eluting, while faint traces of BamD are still appearing. Overall these results suggest that the BamCDE subcomplex is unable to form in the presence of only the BamC$_C$ domain, however, the BamDE dimer is still able to form. Experiment notes: a HiPrep 26/60 Sephacryl S-100 HR gel-filtration column was used attached to the ÄKTAprime system, with a void volume of 98 mL and a column volume of 320 mL. The experiment was run at 4°C with a buffer containing 20 mM Tris-HCl (pH 8.0) and 100 mM NaCl, where the flow rate was set to 1 mL/min, with 3 mL fractions collected. Fractions eluting in or around the void volume are not shown on the SDS-PAGE gel as they were not in the region of interest and contained aggregated protein.

3.2.2.4. N- and C-terminal domains ($BamC_{NC}$)

The results so far are similar to what was observed with the BamCD studies – in the presence of $BamC_N$ or $BamC_C$ alone, the BamCDE subcomplex does not form. To study if the presence of both domains are sufficient for complex formation, the $BamC_{NC}$ construct was co-purified with full length BamD and BamE. The resulting gel-filtration chromatogram shows the presence of three peaks eluting at ~144 mL, ~158 mL, and ~197 mL (Figure 3.11). These volumes correspond to molecular mass values of ~37 kDa, ~27 kDa, and ~12 kDa, which are similar to the calculated mass values of $BamC_{NC}$ (27.3 kDa), BamD (25.7 kDa) co-eluting with dimeric BamE (20.6 kDa), and monomeric BamE (10.3 kDa). However, there is no earlier peak corresponding to a complex, suggesting that the BamCDE subcomplex was unable to form.

Interestingly in this case, the BamDE dimer was not observed either. This could suggest that the $BamC_{NC}$ may be interacting with BamD in a manner that is preventing BamDE dimer formation. If this is the case, then the $BamC_{NC}$-BamD interaction is not strong enough as the proteins dissociate in the gel-filtration matrix. Furthermore, the interaction might be in a different orientation compared to the full length BamCD dimer and possibly blocking the BamE binding sites on BamD. Although more work needs to be done to understand what is happening here and if $BamC_{NC}$ is able to interact with BamD, it is clear that a BamCDE subcomplex is unable to form with $BamC_{NC}$. This suggests that the $BamC_U$ region, which is absent here, may be required for the complex to form.

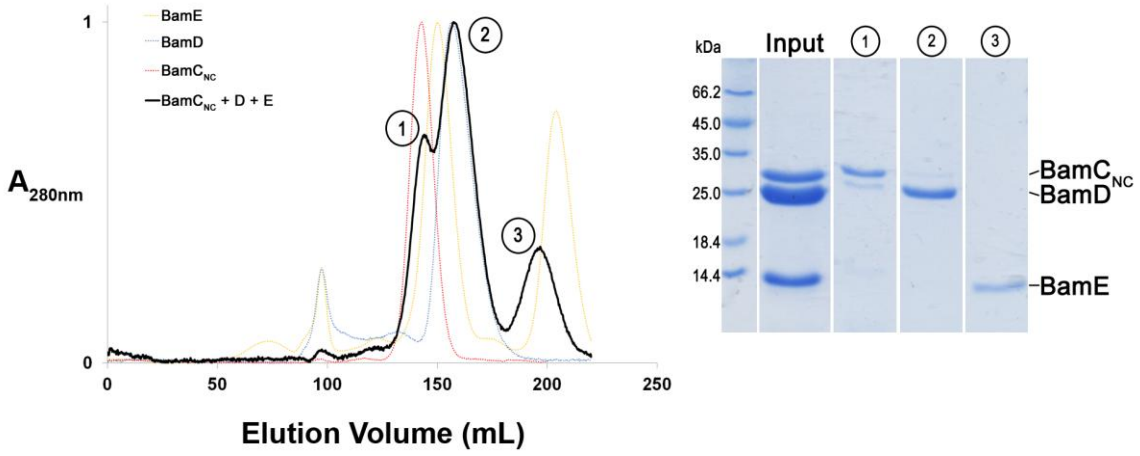

Figure 3.11 Gel-filtration chromatogram and SDS-PAGE showing no BamCDE subcomplex formation with BamC$_{NC}$

The gel-filtration chromatogram (left) shows the elution profile for the sample containing BamC$_{NC}$, BamD, and BamE (black curve), which is plotted as the relative UV absorbance at 280nm (A_{280nm}, provided in arbitrary units) for protein eluted at a specific volume (Elution Volume, provided in millilitres (mL)). The results show the presence of three peaks and which are labelled as 1, 2, and 3. These peaks have similar elution volumes as individual BamC$_{NC}$ (dotted red curve), BamD (dotted blue curve), and BamE (dotted yellow curve) suggesting that the proteins eluted out separately and were unable to form a complex. The corresponding SDS-PAGE gel (right) confirms this by showing the presence of mainly BamC$_{NC}$ in peak 1, mainly BamD in peak 2, and only BamE in peak 3. Faint bands showing presence of BamD in peak 1 and BamC$_{NC}$ in peak 2 are due to overlap of the two peaks. Overall these results suggest that the BamCDE subcomplex is unable to form in the presence of only the BamC$_{NC}$ domains. Interestingly, no BamDE dimer is observed in this case either, suggesting that BamC$_{NC}$ may be interfering with that interaction. Experiment notes: a HiPrep 26/60 Sephacryl S-100 HR gel-filtration column was used attached to the ÄKTAprime system, with a void volume of 98 mL and a column volume of 320 mL. The experiment was run at 4°C with a buffer containing 20 mM Tris-HCl (pH 8.0) and 100 mM NaCl, where the flow rate was set to 1 mL/min, with 3 mL fractions collected. Fractions eluting in or around the void volume are not shown on the SDS-PAGE gel as they were not in the region of interest and contained aggregated protein.

3.2.2.5. Unstructured N-terminus with the N-terminal domain (BamC$_{UN}$)

Based on the results presented so far, the BamC$_N$ and BamC$_C$ domains are unable to form the BamCDE subcomplex when they are present alone or together. This clearly suggests that the absence of the BamC$_U$ region is preventing the complex formation. To study the involvement of the unstructured region, the BamC$_{UN}$ construct was co-purified with full length BamD and BamE to see if a BamCDE subcomplex could still form. The resulting gel-filtration chromatogram shows the presence of a peak at ~125 mL corresponding to a mass of ~55 kDa, which is very close to the calculated

mass of 56.7 kDa for a $BamC_{UN}DE$ subcomplex (Figure 3.12). The SDS-PAGE gel shows the presence of all three proteins in the peak, further suggesting complex formation. A second peak at ~137 mL corresponds to a mass of ~42 kDa, which is close to the calculated mass of 36.0 kDa for the BamDE dimer and can be confirmed by the SDS-PAGE results. Overall, these results show that a $BamC_{UN}DE$ complex is able to form, as well as the BamDE dimer. The formation of the $BamC_{UN}DE$ complex strongly demonstrates that the $BamC_U$ region is important for BamCDE assembly.

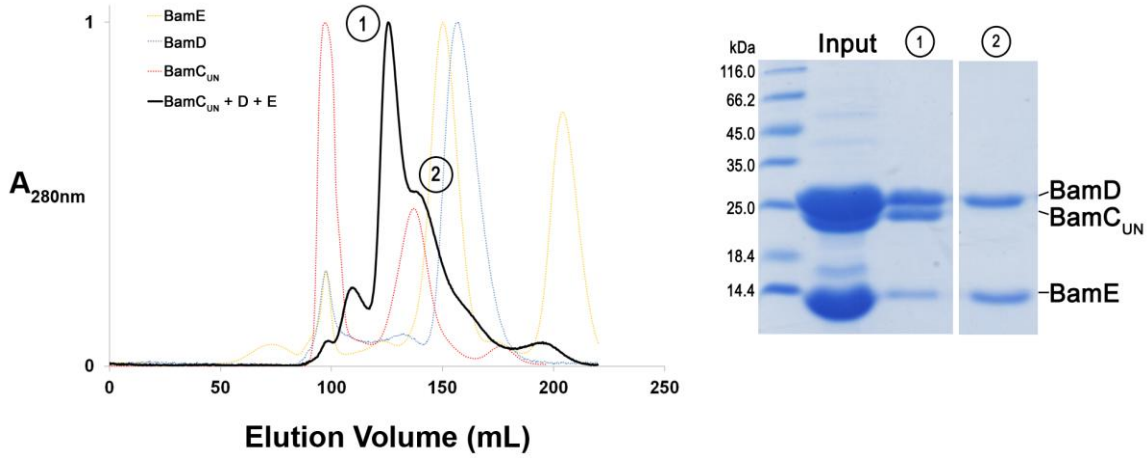

Figure 3.12 Gel-filtration chromatogram and SDS-PAGE showing BamCDE subcomplex formation with $BamC_{UN}$

The gel-filtration chromatogram (left) shows the elution profile for the sample containing $BamC_{UN}$, BamD, and BamE (black curve), which is plotted as the relative UV absorbance at 280nm (A_{280nm}, provided in arbitrary units) for protein eluted at a specific volume (Elution Volume, provided in millilitres (mL)). The results show the presence of one peak and a shoulder which are labelled as 1 and 2. Peak 1 elutes much earlier than individual $BamC_{UN}$ (dotted red curve), BamD (dotted blue curve), and BamE (dotted yellow curve) suggesting that they formed a complex. The corresponding SDS-PAGE gel (right) confirms this by showing the presence of $BamC_{UN}$, BamD, and BamE in peak 1. The shoulder labelled 2 contains BamD and BamE which is eluting at ~137 mL, suggesting a BamDE dimer. Overall these results strongly suggest that the $BamC_U$ region is important for BamCDE subcomplex formation, as previous experiments lacking this region were unable to form the complex. Note that for the individual $BamC_{UN}$ chromatogram (dotted red curve), a large peak is observed at the void volume, which is due to the presence of other high molecular weight proteins in the input sample. More details about that chromatogram are provided in Figure C8. Experiment notes: a HiPrep 26/60 Sephacryl S-100 HR gel-filtration column was used attached to the ÄKTAprime system, with a void volume of 98 mL and a column volume of 320 mL. The experiment was run at 4°C with a buffer containing 20 mM Tris-HCl (pH 8.0) and 100 mM NaCl, where the flow rate was set to 1 mL/min, with 3 mL fractions collected. Fractions eluting in or around the void volume are not shown on the SDS-PAGE gel as they were not in the region of interest and contained aggregated protein.

3.2.2.6. Summary of BamC requirements for BamCDE formation

To study which regions of BamC were required for BamCDE formation, the individual domains were cloned out and purified with full length BamD and BamE. Results from gel-filtration chromatography studies show that in the absence of the unstructured $BamC_U$ region, no BamCDE subcomplex is observed (Figure 3.13; Table 3.3). When the $BamC_{UN}$ construct was co-purified with BamD and BamE, a $BamC_{UN}DE$ subcomplex was able to form, suggesting that the $BamC_{UN}$ region is sufficient for this interaction. However, based on these data alone, it cannot be determined if $BamC_U$ is directly involved in the interaction, or if it plays another role which allows $BamC_N$ to be properly positioned for complex formation. Section 3.3 provides more insight into the role of $BamC_U$ in interaction based on recent structural data.

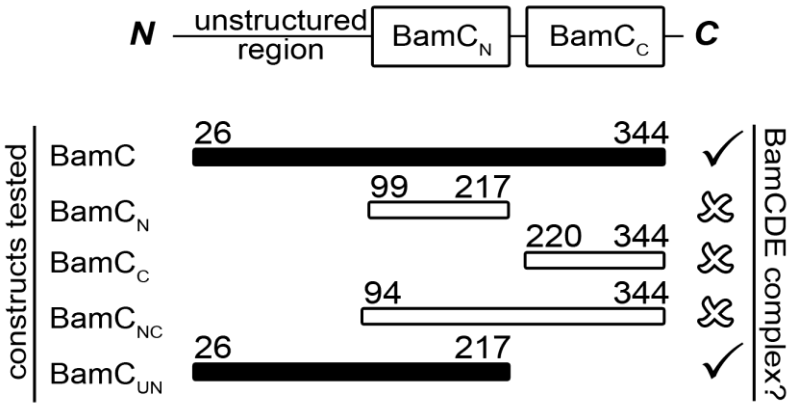

Figure 3.13 Schematic summary of BamC requirements for BamCDE subcomplex formation

This is a schematic summary showing the BamC requirements for BamCDE subcomplex formation. For each construct, the residue boundaries are illustrated and correspond to the domain structure shown above. Successful dimer formation is indicated by a black check mark, while no complex formation is indicated by a white X-mark. Based on this diagram, it can be seen that in the absence of unstructured N-terminus of BamC ($BamC_U$), no BamCDE subcomplex is able to form, suggesting that it is required for the association.

Table 3.3 Gel-filtration elution volumes for BamCDE formation

Combination	Peak Elution Volume (mL)	Measured Mass (kDa)	Protein(s) Eluted and their Calculated Mass (kDa)	
BamC	125	55	BamC	34.3
BamD	155	29	BamD	25.7
BamE	150	32	BamE dimer	20.6
	204	10	BamE monomer	10.3
BamC + BamD + BamE	112	73	BamCDE complex	70.3
	136	43	BamDE dimer	36.0
$BamC_N$ + BamD + BamE	129	50	BamDE dimer	36.0
	180	17	$BamC_N$	13.2
$BamC_C$ + BamD + BamE	133	46	BamDE dimer	36.0
	174	19	$BamC_C$	13.4
	193	13	BamE monomer	10.3
$BamC_{NC}$ + BamD + BamE	144	37	$BamC_{NC}$	27.3
	158	27	BamD	25.7
			BamE dimer	20.6
	197	12	BamE monomer	10.3
$BamC_{UN}$ + BamD + BamE	125	55	$BamC_{UN}$DE complex	56.7
	137	42	BamDE dimer	36.0

This table presents a summary of the peaks observed, their elution volumes, and the proteins eluted. The measured molecular mass was estimated from the standard curve of the gel-filtration column using the corresponding elution volume. The calculated molecular mass of the protein eluted is based on the amino acid composition of each protein.

3.3. BamCD dimer structure

While these interactions studies were being carried out, a crystal structure of the BamCD heterodimer was solved and provided valuable insight into how these two proteins come together (PDB: 3TGO; Figure 3.14) (Kim et al., 2011a). While the full length BamCD dimer was used, it was actually the $BamC_{UN}D$ dimer that eventually crystallized, further supporting the results of the interaction studies in Section 3.2.1 which suggest $BamC_{UN}$ to be sufficient for dimer formation. The inability for full length BamC to crystallize in this case is similar to the observations made with the $BamC_C$

structure discussed in Section 3.1.2, and again demonstrates how BamC is highly susceptible to degradation. The BamCD crystal structure and the results from the BamCD interaction studies in Section 3.2.1 were published in *The Journal of Biological Chemistry* in 2011 (Kim et al., 2011a).

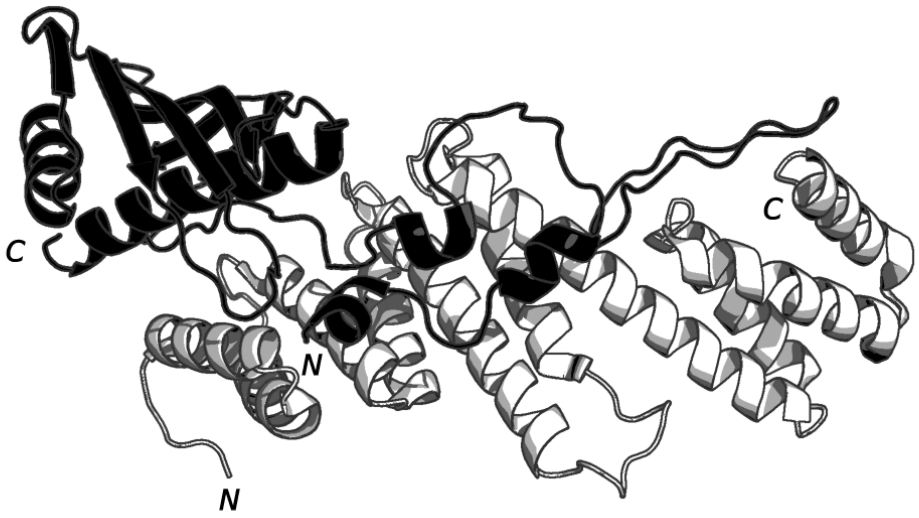

Figure 3.14 Structure of the BamCD dimer

This figure shows the crystal structure of the BamCD dimer, where BamC is shown in black, and BamD in white (PDB: 3TGO). The $BamC_C$ region is missing in this structure, but the unstructured $BamC_U$ region is present along with $BamC_N$ and together corresponds to the $BamC_{UN}$ construct used in the interaction studies in this chapter. From this structure it can be seen that the $BamC_U$ region is important for the BamCD interaction as it makes direct contacts with the entire length of BamD.

3.3.1. Unstructured N-terminus of BamC is directly involved

The structure of the BamCD dimer brought along a major revelation in terms of the potential role of the unstructured region of BamC. Despite $BamC_U$ containing the most number of conserved residues in the entire protein, this unstructured region was often overlooked and the $BamC_N$ and $BamC_C$ domains were predicted to be involved in interaction. The lack of secondary structure in $BamC_U$ made it a nuisance to work with as it would contribute to the early degradation of BamC and also interfered with

crystallization experiments, prompting researchers to remove the region (Albrecht and Zeth, 2010). This made it difficult to clearly understand the role of $BamC_U$, and it was sometimes simply referred to as the N-terminal anchoring region, or the region connecting to the N-terminal acylation site (Albrecht and Zeth, 2011; Knowles et al., 2009). However, the BamCD structure reveals that $BamC_U$ plays a major role in the interaction as it makes direct contacts with BamD (Figure 3.14). The lack of a tight secondary structure allows $BamC_U$ to create a lasso-like arrangement that extends throughout the entire length of BamD. Although $BamC_N$ is also involved in the interaction, the majority of the contacts with BamD are made through $BamC_U$ (Kim et al., 2011a). This structure strongly supports the observation made in the BamC interaction studies and confirms that the unstructured $BamC_U$ region is required for a stable BamCD dimer.

3.3.2. Potential regulatory role for BamC

Previous structural studies of BamD proposed it be the binding site for incoming OMP substrates (Sandoval et al., 2011; Albrecht and Zeth, 2011). This was based on structural homologues such as PEX5 (a peroxisomal targeting signal receptor), whose binding pocket is very similar to a pocket created by the TPR1 and TPR2 motifs of BamD (Figure 3.15A). This theory is further strengthened by experiments showing successful cross-linking between this pocket on BamD and a peptide resembling the C-terminal OMP-signal (Albrecht and Zeth, 2011). However, in the BamCD structure, the BamD binding pocket is occupied by a portion of the $BamC_U$ region (Figure 3.15B) (Kim et al., 2011a). If BamD is indeed involved in substrate binding, and BamC also binds to the same substrate binding pocket, then it is possible that BamC may play a regulatory role in OMP assembly by preventing OMP binding to BamD. Although more work needs to be done to confirm this, the BamCD structure finally provides a potential function for BamC.

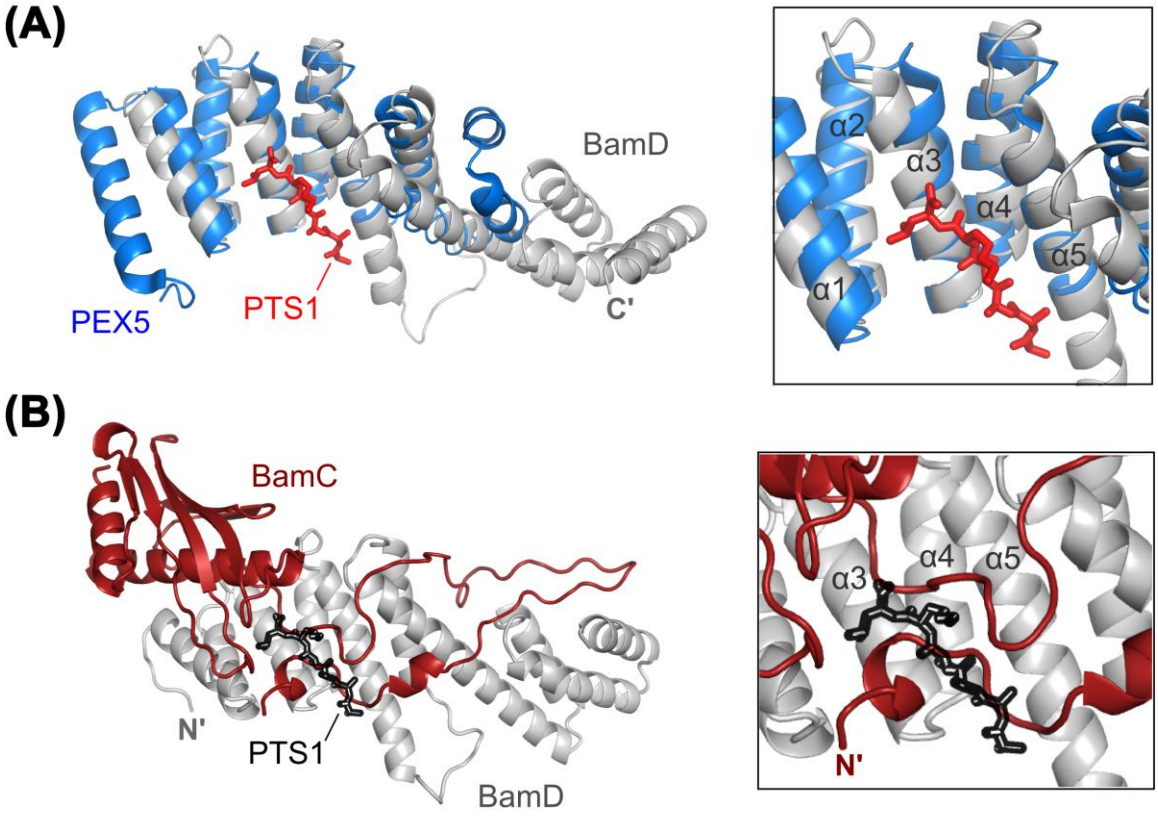

Figure 3.15 Potential regulatory role of BamC

This figure shows the reasoning behind why BamC may function as a regulator for the BAM complex. (A) shows the structure of BamD (white; PDB: 3TGO) superimposed with its structural homologue, PEX5 from *Trypanosoma brucei* (blue; PDB: 3CVP) which is bound to its substrate PTS1 (red). This suggests that BamD may be involved in binding to OMP substrates. A close up view of the pocket is shown on the right. (B) shows BamC (dark red) to bind to the same pocket on BamD that corresponds to the binding pocket of PEX5 in (A). The PTS1 substrate has been superimposed here in black outline to show its overlapping binding site with BamC. Based on this observation, it may be possible for BamC to function as a regulator for OMP substrate binding to BamD. A close up view of the pocket is shown on the right. This figure is adapted from (Kim et al., 2011a).

3.4. Competition Analyses

3.4.1. Strategy

3.4.1.1. Need for "positive results"

The interaction experiments presented in Chapter 2 along with those presented so far in this chapter provide "negative results" where removal of certain regions led to the inability to form the desired subcomplex. If the N-termini of BamE or of BamC are truly required for complex formation, then these regions alone could possibly be sufficient for complex formation. With the structure of the BamCD dimer in Section 3.3 confirming the direct involvement of $BamC_U$ in interaction, the hypothesis that this region alone may be sufficient for association with BamD was investigated. Furthermore, a competition experiment was attempted to see if $BamC_U$ could compete with full length BamC for associating with BamD to form a $BamC_UD$ dimer. This would confirm that $BamC_U$ is sufficient for interaction, as previous interaction experiments used the $BamC_{UN}$ construct which also contains the N-terminal domain of BamC ($BamC_N$). As noted in Section 3.1.3, creating a $BamC_U$ construct was challenging because it resulted in poor protein expression. Thus, for this experiment a fusion protein was designed to ensure that $BamC_U$ would be soluble.

Similarly, now that BamD has been proposed to bind to the OMP-signal with BamC as the regulator, another experiment was designed to test the interactions between these three components. PhoE is an OMP (specifically it is a porin – see Section 1.2.1 for more details) that has been shown to be a substrate for the BAM complex (Robert et al., 2006). The C-terminal OMP-signal from PhoE was added to see if it could compete with BamC for associating with BamD to form a PhoE-BamD complex. This would confirm if BamD is able to bind to OMP substrates, while demonstrating BamC's potential regulatory role. Due to the short length of this OMP-signal, it was fused to another protein to help in solubility, and to make it detectable through gel-filtration chromatography and SDS-PAGE.

3.4.1.2. Competitor fusion proteins constructed

Maltose binding protein (MBP) is very soluble, and when another protein is fused to the C-terminus of MBP using a plasmid such as pMAL-c2x, the expression and

solubility of that target protein can be enhanced (Kapust and Waugh, 1999). Taking this approach, the region corresponding to $BamC_U$ (residues 26-100) was expressed with the N-terminal MBP tag. The resulting MBP-$BamC_U$ construct showed better expression than $BamC_U$ alone and was used for the competition analysis (see Figure C9 in Appendix C for $BamC_U$ expression). Similarly, the C-terminal OMP-signal from PhoE (residues 342-351) was fused to MBP resulting in the MBP-PhoE construct. MBP alone has a mass of 42.5 kDa, thus complex formation between and MBP-fusion protein and BamD would be much easier to detect on a gel-filtration chromatogram due to the large shift in mass. A list of all the protein constructs used in this experiment with their amino acid boundaries and molecular mass is provided in Table 3.4. Note that full length BamC and BamD are the same constructs used in the previous experiments, and thus have N-terminal hexahistidine tags.

Table 3.4 List of constructs used for competition analysis

Construct Name	Residue Boundaries	Molecular Mass (kDa)
BamC	26-344	34.3
BamD	21-245	25.7
MBP-$BamC_U$	26-100	50.2
MBP-PhoE	342-351	43.6

The molecular mass is calculated based on the amino acid composition of the protein construct. For the MBP-fusion constructs, this also includes the mass of MBP (42.5 kDa).

To conduct initial competition studies, a cell pellet of MBP-$BamC_U$ was co-purified with cell pellets of full length BamC and BamD to see which construct BamD would associate with. Similarly, MBP-PhoE was also co-purified with full length BamC and BamD. Competition was defined as the elution of an intact MBP-$BamC_U$-BamD or MBP-PhoE-BamD dimer through gel-filtration chromatography, in the presence of full length BamC.

3.4.2. Gel-filtration chromatography studies

3.4.2.1. Competition between MBP-BamC$_U$ and full length BamC

When MBP-BamC$_U$ was co-lysed and purified with full length BamC and BamD, the resulting gel-filtration chromatogram revealed three peaks (Figure 3.16A). However, instead of observing peaks for both MBP-BamC$_U$-BamD and BamCD dimers, only the BamCD dimer appeared to have formed. The peaks have elution volumes of ~122 mL, ~139 mL, and ~164 mL which correspond to molecular mass values of ~59 kDa, ~41 kDa, and ~24 kDa. These mass estimations and the SDS-PAGE gel suggest the elution of a BamCD dimer (60.0 kDa), MBP-BamC$_U$ (50.2 kDa), and BamD (25.7 kDa). If an MBP-BamC$_U$-BamD complex was to form, it would have a calculated mass of 75.9 kDa, and would elute at ~110 mL. However no peak is observed in this region, suggesting that this complex did not form.

In addition, the SDS-PAGE gel shows the presence of additional protein bands in some peaks. In the second peak, a faint band corresponding to BamD is present. This is most likely due to the close proximity of the second and third peaks, and not due to interaction. Thus, a faint band corresponding to MBP-BamC$_U$ is also visible in the lane for the third peak. However, the SDS-PAGE gel shows the presence of another band larger than BamD in the second peak. This fragment looks similar to degraded BamC observed in Section 3.1.2 and could be BamC$_{NC}$ due to cleavage of the unstructured region. Its expected elution volume is ~142 mL which is close to the volume of ~139 mL observed for this peak.

In summary, it appears that MBP-BamC$_U$ is unable to compete with BamC to form an MBP-BamC$_U$-BamD complex. This is also supported by additional experiments where MBP-BamC$_U$ was purified with BamD alone, as no complex formation was observed through gel-filtration chromatography in that case either (Figure 3.16B). Thus, it is likely that either BamC$_U$ is unable to associate with BamD by itself, or the MBP-BamC$_U$ construct is not ideal for this experiment. Note that the elution volume of the second peak corresponding to MBP-BamC$_U$ gave a mass estimation of ~41kDa which is distant from the calculated mass of 50.2kDa. In addition, on SDS-PAGE this fusion protein migrates faster than expected and appears less than 45kDa. There could be two explanations for this phenomenon: either the fusion protein is dividing apart or the

$BamC_U$ region is degrading resulting in MBP alone; or, the unstructured $BamC_U$ region is wrapping around MBP or trapped inside a crevice of MBP (Figure 3.17). While these hypotheses are based on only a few experiments, they could suggest why the $BamC_U$ region is inaccessible for interaction, and why the addition of the unstructured $BamC_U$ is not contributing to the size of the fusion protein. Therefore it is recommended that a different approach be used, and that $BamC_U$ could be fused to a different protein, or perhaps be positioned to the N-terminus of MBP to mimic its environment within full length BamC.

However, it should also be noted that these trials were performed using approximately equimolar concentration of MBP-$BamC_U$ in relation to BamD and full length BamC, as the cell pellets used for each protein were derived from two litre cultures each. To properly test the competition, MBP-$BamC_U$ should have been included in excess. For example, a four litre cell culture of MBP-$BamC_U$ could have been co-lysed and purified with a two litre cell culture of BamD to overcome the small size of $BamC_U$, and its inability to access BamD as efficiently. Once the amount of excess MBP-$BamC_U$ required to complex with BamD is determined, then the competition experiment including full length BamC can be performed to see which form of BamC would be preferentially bound to by BamD. This would provide better results from which the ability of $BamC_U$ to act as a competitor for full length BamC can be determined, as the results from the experiments presented here are inconclusive.

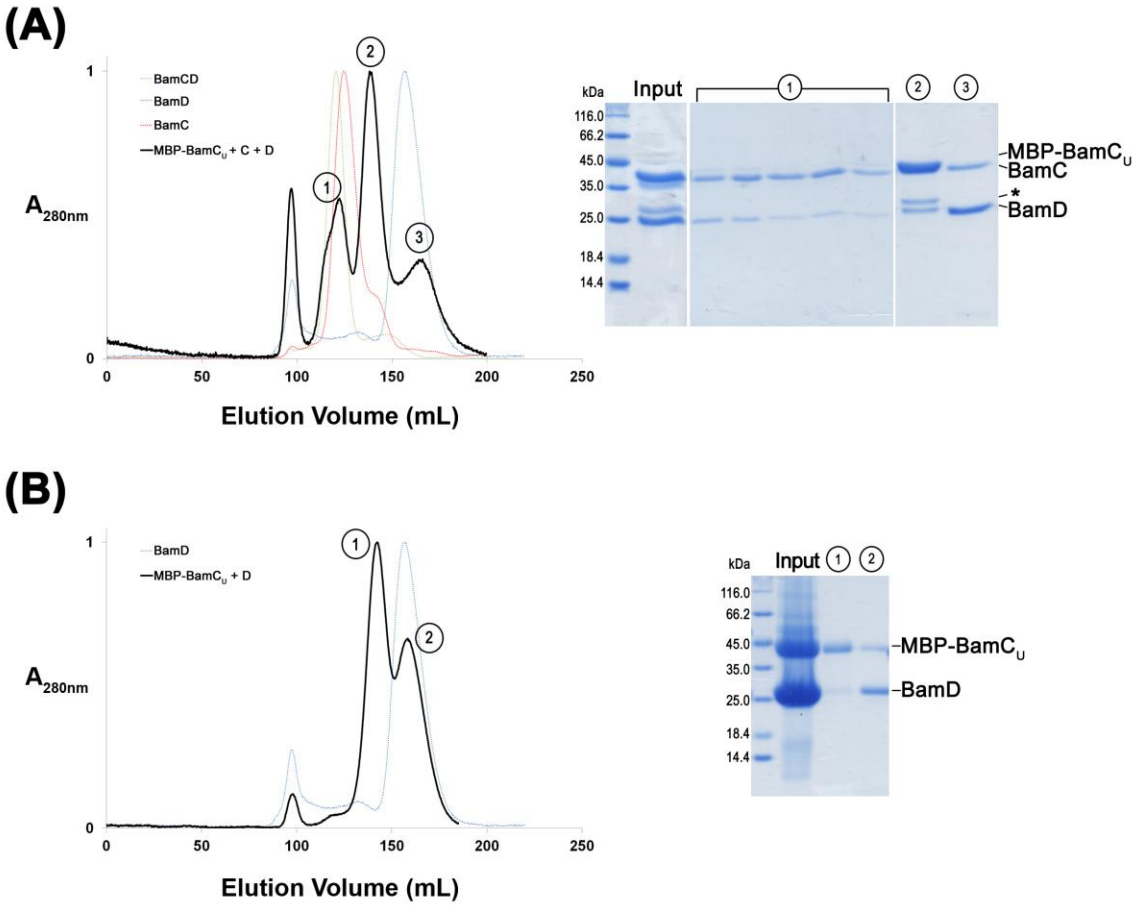

Figure 3.16 Gel-filtration chromatogram and SDS-PAGE showing preliminary competition experiments with MBP-BamC$_U$

The gel-filtration chromatograms show the elution profile for a sample that is plotted as the relative UV absorbance at 280nm (A_{280nm}, provided in arbitrary units) for protein eluted at a specific volume (Elution Volume, provided in millilitres (mL)). (A) shows the chromatogram (left) of a sample containing MBP-BamC$_U$, BamC, and BamD (black curve), and results in the presence of three peaks which are labelled as 1, 2 and 3. Peak 1 elutes where the BamCD dimer has been observed (dotted green curve), and earlier than individual BamC (dotted red curve) and BamD (dotted blue curve) suggesting that a complex had formed. However, based on the chromatogram alone it cannot be determined if this is BamCD dimer or an MBP-BamC$_U$-BamD complex. The corresponding SDS-PAGE gel (right) provides more information by showing the presence of both BamC and BamD in peak 1 suggesting a BamCD dimer, while MBP-BamC$_U$ elutes in peak 2, and excess BamD in peak 3. Due to the close proximity of peaks 2 and 3, some BamD is seen in peak 2 and some MBP-BamC$_U$ is seen in peak 3. Note that peak 2 contains another band (labelled on the gel with an asterisk, *) which is most likely a degraded form of BamC that has been seen to elute in this region. (B) shows the chromatogram (left) of a sample containing MBP-BamC$_U$ and BamD (black curve), and results in two peaks which are labelled as 1 and 2. Peak 2 superimposes closely with the chromatogram of individual BamD (dotted blue curve). The corresponding SDS-PAGE (right) shows that MBP-BamC$_U$ elutes in peak 1, while BamD elutes in peak 2, and no complex is formed. Overall the results from (A) suggest that MBP-BamC$_U$ is unable to compete with BamC, as the BamCD dimer still forms. In addition, (B) shows that MBP-BamC$_U$ is unable to form a complex with BamD in the absence of BamC, explaining why MBP-

$BamC_U$ was unable to compete with BamC in (A). Experiment notes: a HiPrep 26/60 Sephacryl S-100 HR gel-filtration column was used attached to the ÄKTAprime system, with a void volume of 98 mL and a column volume of 320 mL. The experiments were run at 4°C with a buffer containing 20 mM Tris-HCl (pH 8.0) and 100 mM NaCl, where the flow rate was set to 1 mL/min, with 3 mL fractions collected. Fractions eluting in or around the void volume are not shown on the SDS-PAGE gel as they were not in the region of interest and contained aggregated protein.

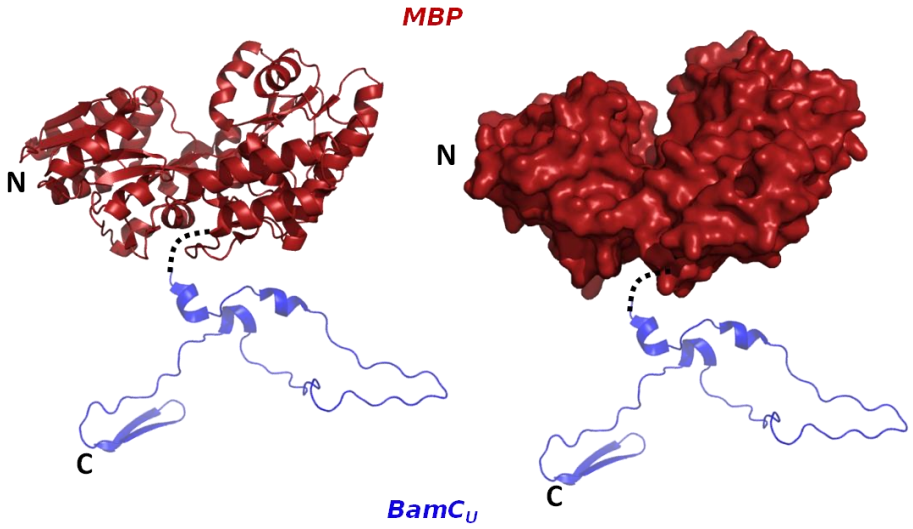

Figure 3.17 Schematic showing the ideal MBP-$BamC_U$ construct

This figure provides the structures of MBP (red; PDB: 1PEB) and $BamC_U$ (blue; PDB: 3TGO) linked together in an ideal situation. It was hoped that after fusing the proteins together, $BamC_U$ would fold similar to this manner (as observed in the BamCD dimer structure), and be accessible for interaction with BamD. Ribbon diagrams are provided on the left, while the surface representation of MBP is provided on the right. As shown by the surface representation of MBP, a crevice is present. It is possible that the unstructured $BamC_U$ may be trapped in this crevice and therefore inaccessible for interaction with BamD.

3.4.2.2. Competition between MBP-PhoE and BamC

MBP-PhoE was co-lysed and purified with full length BamC and BamD to see if MBP-PhoE could compete with BamC for binding to BamD. The resulting gel-filtration chromatogram revealed three main peaks (Figure 3.18A). The peaks eluted at ~116 mL, ~140 mL, and ~157 mL corresponding to molecular mass values of ~67 kDa, ~40 kDa, and ~28 kDa. Based on the SDS-PAGE gel, these peaks are due to the elution of the

BamCD dimer (60.0 kDa), MBP-PhoE (43.6 kDa), and BamD (25.7 kDa). Due to the overlapping of the second and third peaks, faint traces of BamD are observed in the second peak, while traces of MBP-PhoE are observed in the third peak. This experiment shows that MBP-PhoE was unable to compete with BamC to form an MBP-PhoE-BamD complex. This result is supported by additional experiments where MBP-PhoE was purified with BamD alone through gel-filtration chromatography, and no complex formation observed in that case either (Figure 3.18B).

While this was just an initial experiment, it is possible that the MBP-PhoE construct may be at fault similar to the MBP-$BamC_U$ as discussed in the previous section. The OMP-signal used from PhoE is only 10 residues long, and perhaps extending it may improve binding with BamD. Also, while PhoE has a β-barrel structure with the OMP-signal contributing to the final strand of the protein, it is likely that the secondary structure of this peptide is affected by being fused to MBP, especially since BamD is supposed to binding to the unfolded form of the PhoE OMP signal (Figure 3.19). Thus, future experiments may consider fusing the OMP-signal to protein that has β-strands near the C-terminus to mimic its natural environment, as opposed to being fused to MBP which is mainly α-helical. However, it should be noted that similar to Section 3.4.2.1, the experiments presented here involved approximately equimolar concentrations of MBP-PhoE in relation to BamD and BamC. Thus no complex observed between MBP-PhoE may not be due to poor construct design, but instead may be due to low amounts of MBP-PhoE. Thus, MBP-PhoE should be included in excess to test complex formation with BamD, and then followed by competition with BamC to test which protein BamD will preferentially bind to. Finally, it should be reminded that the binding between BamD and PhoE has not been shown previously, so it is possible that this interaction may not exist. However based on previous cross-linking experiments using a peptide with a similar sequence, investigating the interaction between BamD and PhoE is hopeful, and should continued to be pursued (Albrecht and Zeth, 2011).

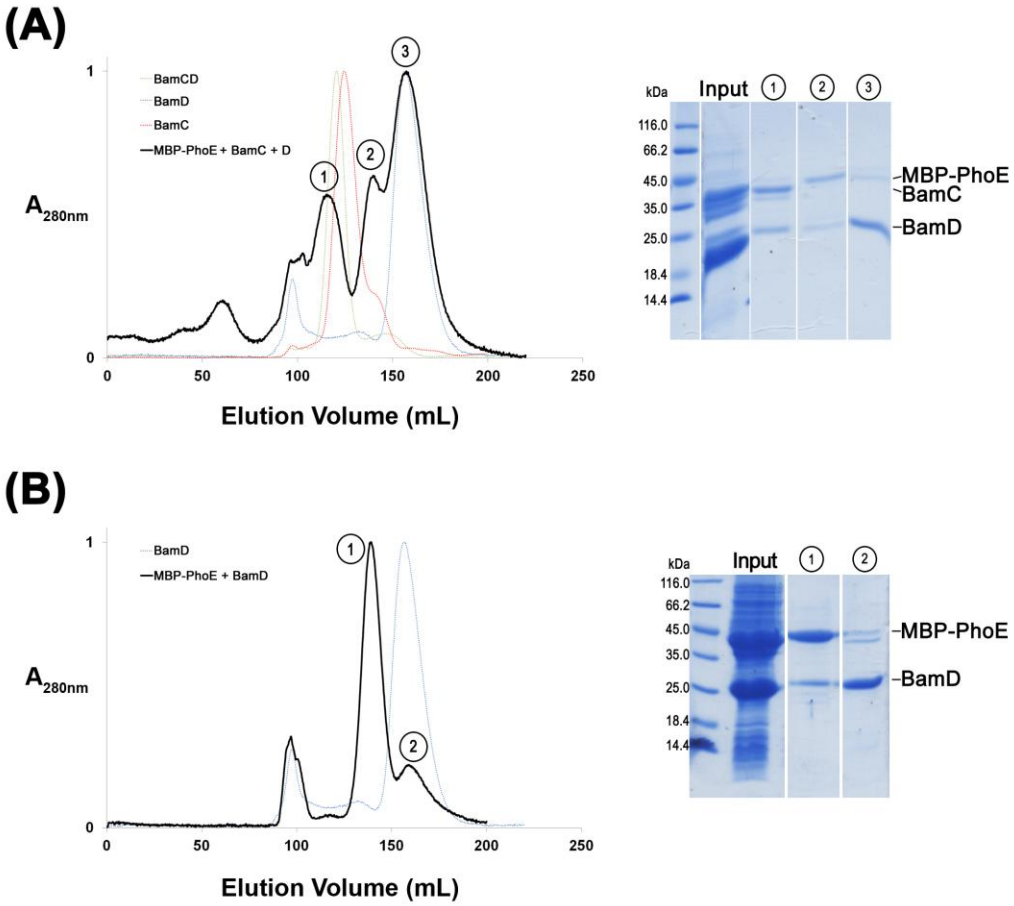

Figure 3.18 Gel-filtration chromatogram and SDS-PAGE showing preliminary competition experiments with MBP-PhoE

The gel-filtration chromatograms show the elution profile for a sample that is plotted as the relative UV absorbance at 280nm (A_{280nm}, provided in arbitrary units) for protein eluted at a specific volume (Elution Volume, provided in millilitres (mL)). (A) shows the chromatogram (left) of a sample containing MBP-PhoE, BamC, and BamD (black curve), and results in the presence of three peaks which are labelled as 1, 2 and 3. Peak 1 elutes where the BamCD dimer has been observed (dotted green curve), and earlier than individual BamC (dotted red curve) and BamD (dotted blue curve) suggesting that a complex had formed. However, based on the chromatogram alone it cannot be determined if this is BamCD dimer or an MBP-PhoE-BamD complex. The corresponding SDS-PAGE gel (right) provides more information by showing the presence of both BamC and BamD in peak 1 suggesting a BamCD dimer. (B) shows the chromatogram (left) of a sample containing MBP-PhoE and BamD (black curve), and results in two peaks which are labelled as 1 and 2. Peak 2 elutes where individual BamD has been observed (dotted blue curve). The corresponding SDS-PAGE (right) shows that MBP-PhoE elutes in peak 1, while BamD elutes in peak 2, and no complex is formed. Overall the results from (A) suggest that MBP-PhoE is unable to compete with BamC, as the BamCD dimer still forms. In addition, (B) shows that MBP-PhoE is unable to form a complex with BamD in the absence of BamC, explaining why MBP-PhoE was unable to compete with BamC in (A). Experiment notes: a HiPrep 26/60 Sephacryl S-100 HR gel-filtration column was used attached to the ÄKTAprime system, with a void volume of 98 mL and a column volume of 320 mL. The experiments were run at 4°C with a buffer containing 20 mM Tris-HCl (pH 8.0) and 100 mM NaCl, where the flow rate was set to 1 mL/min, with 3 mL fractions collected. Fractions eluting in or around the void volume are not shown on the SDS-PAGE gel as they were not in the region of interest and contained aggregated protein.

Figure 3.19 The PhoE OMP signal

This figure shows the structure of a PhoE monomer (PDB: 1PHO). Note: while this figure shows a monomer of PhoE, the protein is actually found as a homotrimer as it is a porin (see Section 1.2.1). The C-terminal OMP signal is shown in black, which makes up the last β-stand of this β-barrel. The OMP signal of PhoE is proposed to be the last 10 residues of the protein (IVAVGMTYQF), and this is the sequence that was used in the MBP-PhoE fusion protein. Note that this sequence is naturally found as a β-strand, but is most likely unfolded when it is recognized by the BAM complex (a role that is proposed to be carried out by BamD). Therefore, fusing this sequence to an α-helical protein such as MBP (see Figure 3.17) may alter its secondary structure, and prevent BamD from binding to it.

3.5. Conclusions

This chapter presented the experiments that were conducted to investigate the role of BamC in the BamCD and BamCDE subcomplexes. Based on proteolysis studies and structure prediction data, BamC was proposed to have three independently "folding" domains: the unstructured N-terminus ($BamC_U$), the N-terminal domain ($BamC_N$), and the C-terminal domain ($BamC_C$). The structural data of BamC that became available later further supported this domain architecture. Sequence alignments showed regions of conserved residues to be present mainly in the $BamC_U$ and $BamC_C$ regions. Based on this information, the individual domains were cloned out as follows: $BamC_{UN}$, $BamC_N$,

$BamC_C$, and $BamC_{NC}$. Note that because of the difficulty in expressing $BamC_U$ alone, it was cloned and expressed as $BamC_{UN}$, while the $BamC_{NC}$ was designed to see the effect of removing this unstructured region. Using gel-filtration chromatography to detect complex formation, these constructs were co-purified with the different interaction partners to identify which region of BamC is required in oligomerization.

The results clearly demonstrate that the unstructured $BamC_U$ region is required for complex formation as the BamCD and BamCDE subcomplexes only form in its presence. In addition, it was shown that $BamC_{UN}$ is sufficient, implying that the $BamC_C$ domain does not play a vital role within these complexes. The BamCD dimer structure further emphasized these results by showing that $BamC_U$ makes direct contacts with BamD. The flexibility of $BamC_U$ allows the formation of a lasso-like arrangement that extends along the entire length of BamD, and allows for the formation of numerous van der Waals, ionic, and hydrogen bond interactions with all five TPR motifs of BamD (Kim et al., 2011a). Furthermore, a portion of $BamC_U$ fits in a pocket on BamD that has been proposed to be the binding pocket for OMP substrates. If BamD is indeed involved in substrate binding, then perhaps BamC may play a regulatory role by also binding to this pocket to block substrates.

Competition analyses were performed to investigate if $BamC_U$ alone would be sufficient for interaction with BamD. However, the MBP-$BamC_U$ fusion protein constructed failed to show any complex formation, suggesting that either $BamC_U$ is not sufficient, or that this fusion protein is not ideal for this experiment. Similarly, to investigate the proposed roles of BamD being the substrate binding protein and BamC as its regulator, the MBP-PhoE fusion protein was designed where the C-terminal OMP-signal from PhoE (an OMP, and known substrate of the BAM complex) was fused to MBP. MBP-PhoE was also unable to form a complex with BamD. Therefore, it is speculated that the presence of MBP in these constructs may have prevented proper interaction with BamD, and another approach would be needed. In addition, as explained in Sections 3.4.2.1 and 3.4.2.2, the small size of the competitor proteins ($BamC_U$ and the PhoE OMP-signal) may have resulted in low binding affinities for BamD, which could have been overcome by including these proteins in excess concentration. This would provide better results for determining whether or not they are competitors of full length BamC for binding to BamD.

In summary, this chapter provides an overview of how protein engineering, gel-filtration chromatography, and protein crystallography were used to determine the involvement of BamC in the BamCD and BamCDE subcomplexes. Together, these results provide evidence that the unstructured $BamC_U$ region is required for the interactions, with $BamC_{UN}$ being sufficient. However, the proposed regulatory role of BamC needs more supporting data. A competition experiment similar to the one presented in this chapter provides one approach of how this could be investigated.

4. $BamC_{UN}$ + BamD + $BamE_{\Delta C}$ are the minimal requirements for the formation of the BamCDE subcomplex

4.1. Strategy

Chapters 2 and 3 revealed that $BamE_{\Delta C}$ is equivalent to full length BamE when forming the BamCDE subcomplex, and similarly $BamC_{UN}$ is equivalent to BamC when forming this complex. The next step would be to combine the different truncations to see which combination is still able to form the BamCDE complex. However, since only $BamC_{UN}$ and $BamE_{\Delta C}$ have shown to be successful in the past, only these truncations were carried forward for further studies. By combining $BamC_{UN}$ and $BamE_{\Delta C}$ with full length BamD, it was hypothesized that the BamCDE subcomplex would still form. This would then confirm that the unstructured N-termini of both BamC and BamE are necessary for complex formation. Furthermore, the resulting $BamC_{UN}DE_{\Delta C}$ provides a good crystallization candidate as the unnecessary components would be removed, allowing the first structure of this subcomplex to be determined. Alternatively, if no $BamC_{UN}DE_{\Delta C}$ can form, then this would suggest that the missing C-termini of BamC and BamE have a more important role than previously shown, and that the full length version of one of these proteins may be required for stabilizing the complex.

For this experiment, only three protein constructs were used, $BamC_{UN}$, BamD, and $BamE_{\Delta C}$, whose amino acid boundaries and molecular mass values are listed in Table 4.1. As described in Chapters 2 and 3, all proteins contain hexahistidine tags: $BamC_{UN}$ has the tag on the C-terminus while BamD and $BamE_{\Delta C}$ have the tag on the N-terminus. Similar to previous experiments, the cell pellets containing the individual proteins were co-lysed and purified, and successful complex formation was defined as the elution of the intact $BamC_{UN}DE_{\Delta C}$ complex during gel-filtration chromatography.

Table 4.1 List of constructs used as minimal requirements for BamCDE

Construct Name	Residue Boundaries	Molecular Mass (kDa)
$BamC_{UN}$	26-217	20.7
BamD	21-245	25.7
$BamE_{\Delta C}$	21-107	9.7

The molecular mass is calculated based on the amino acid composition of the protein construct.

4.2. Summary of gel-filtration studies

As presented in Chapters 2 and 3, when full length BamC, BamD, and BamE are co-lysed and purified together, they form the BamCDE subcomplex with a mass of 70.3 kDa which elutes at ~112 mL on gel-filtration chromatography. When full length BamC and BamD are purified with $BamE_{\Delta C}$, the resulting $BamCDE_{\Delta C}$ subcomplex has a mass of 69.7 kDa which elutes at ~117 mL. When $BamC_{UN}$ is purified with full length BamD and BamE, the resulting $BamC_{UN}DE$ subcomplex has a mass of 56.7 kDa, eluting at ~125 mL. To confirm if $BamC_{UN}$ and $BamE_{\Delta C}$ can truly substitute for their full length versions, these truncated constructs were co-lysed and purified with full length BamD to see if complex formation could still occur. The resulting gel-filtration chromatogram shows a large peak followed by a shorter peak, with elution volumes of ~122 mL and ~149 mL (Figure 4.1). Based on the standard curve, these peaks correspond to molecular mass values of ~59 kDa and ~33 kDa. SDS-PAGE shows all three proteins to co-elute in the first peak, suggesting the formation of a $BamC_{UN}DE_{\Delta C}$ subcomplex with a molecular mass of 56.1 kDa. The second peak is where the excess BamD (25.7 kDa) appears to elute, along with the $BamE_{\Delta C}$ dimer (19.4 kDa). Table 4.2 provides a summary of the different combinations tested for forming the BamCDE subcomplex. Based on these results it can be seen how the N-termini of both BamC and BamE are important for these interactions. Thus, it can be concluded that $BamC_{UN}$, BamD, and $BamE_{\Delta C}$ are the minimal requirements for BamCDE subcomplex formation.

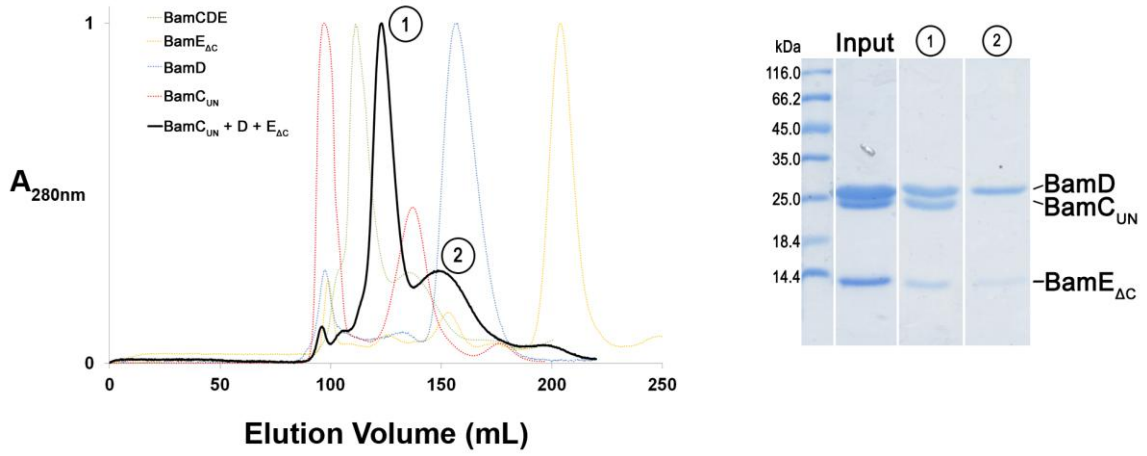

Figure 4.1 Gel-filtration chromatogram and SDS-PAGE showing successful BamCDE subcomplex formation with BamC$_{UN}$, BamD, and BamE$_{\Delta C}$

The gel-filtration chromatogram (left) shows the elution profile for the sample containing BamC$_{UN}$, BamD, and BamE$_{\Delta C}$ (black curve), which is plotted as the relative UV absorbance at 280nm (A_{280nm}, provided in arbitrary units) for protein eluted at a specific volume (Elution Volume, provided in millilitres (mL)). The results show the presence of two peaks which are labelled as 1 and 2. Peak 1 elutes much earlier than individual BamC$_{UN}$ (dotted red curve), BamD (dotted blue curve), and BamE$_{\Delta C}$ (dotted yellow curve) suggesting that they formed a complex. The corresponding SDS-PAGE gel (right) confirms this by showing the presence of BamC$_{UN}$, BamD, and BamE$_{\Delta C}$ in peak 1. Peak 2 contains excess BamD. Overall these results show that a BamC$_{UN}$DE$_{\Delta C}$ subcomplex is able to form. Experiment notes: a HiPrep 26/60 Sephacryl S-100 HR gel-filtration column was used attached to the ÄKTAprime system, with a void volume of 98 mL and a column volume of 320 mL. The experiment was run at 4°C with a buffer containing 20 mM Tris-HCl (pH 8.0) and 100 mM NaCl, where the flow rate was set to 1 mL/min, with 3 mL fractions collected. Fractions eluting in or around the void volume are not shown on the SDS-PAGE gel as they were not in the region of interest and contained aggregated protein.

Table 4.2 Summary of BamCDE interaction studies

BamC	BamD	BamE	BamCDE subcomplex?
Full length	Full length	Full length	Yes
Full length	Full length	$BamE_{\Delta N}$	No
Full length	Full length	$BamE_{\Delta C}$	Yes
Full length	Full length	$BamE_{\Delta N\Delta C}$	No
$BamC_N$	Full length	Full length	No
$BamC_C$	Full length	Full length	No
$BamC_{NC}$	Full length	Full length	No
$BamC_{UN}$	Full length	Full length	Yes
$BamC_{UN}$	Full length	$BamE_{\Delta C}$	Yes

This table shows the outcomes of the BamCDE interaction studies, with successful BamCDE subcomplex formation highlighted in gray.

4.3. $BamC_{UN}$-BamD-$BamE_{\Delta C}$ as a crystallization candidate

Protein crystallization is often challenging, especially if the target protein has unstructured regions. Due to the flexibility of these regions, they can exist in multiple conformations, resulting in a heterogeneous sample that will not favour producing a tightly packed and uniform crystal (International Union of Crystallography, 2012). Thus, the $BamC_{UN}DE_{\Delta C}$ complex provides a theoretically good crystallization candidate as the flexible regions or regions not directly required for interaction have been removed, and the remaining unstructured regions are tightly bound in the complex. This resulting subcomplex is much more compact than the full length BamCDE subcomplex, and therefore should favour crystallization. The purified $BamC_{UN}DE_{\Delta C}$ complex was set up for crystallization experiments using commercial crystal screens (see methods in Appendix A). Initial experiments produced small crystals in several conditions that can be optimized to produce larger crystals. Table 4.3 provides the details of the crystallization conditions that produced these initial hits. Figure 4.2 provides images of these crystals that were observed approximately two weeks after setting up the crystallization drops.

Table 4.3 Initial conditions that produced $BamC_{UN}DE_{\Delta C}$ crystals

Crystallization Technique	Sitting drop vapour diffusion
Protein Sample	$BamC_{UN}DE_{\Delta C}$ (with hexahistidine tags intact)
Concentration	20 mg/mL
Protein Buffer	20 mM Tris (pH 8.0), 100 mM NaCl, 0.03% DDM
Temperature	Room Temperature (~20°C)
Reservoir buffer conditions that produced crystals	*Hampton Research Crystal Screen 2 condition #1 (CS2-1)* 2.0 M sodium chloride 10% w/v polyethylene glycol 6000
	Hampton Research Crystal Screen 2 condition #9 (CS2-9) 0.1 M sodium acetate trihydrate pH 4.6 2.0 M sodium chloride
	Hampton Research PEG/ION Screen condition #13 (PEG-13) 0.2 M sodium thiocyanate 20% w/v polyethylene glycol 3350
	Hampton Research PEG/ION Screen condition #40 (PEG-40) 0.2 M sodium phosphate dibasic dihydrate 20% w/v polyethylene glycol 3350
	Hampton Research PEG/ION Screen condition #42 (PEG-42) 0.2 M potassium phosphate dibasic 20% w/v polyethylene glycol 3350
	Hampton Research PEG/ION Screen condition #44 (PEG-44) 0.2 M ammonium phosphate dibasic 20% w/v polyethylene glycol 3350

This table provides the details of crystallization experiments that were set up which resulted in the initial crystals observed in Figure 4.2.

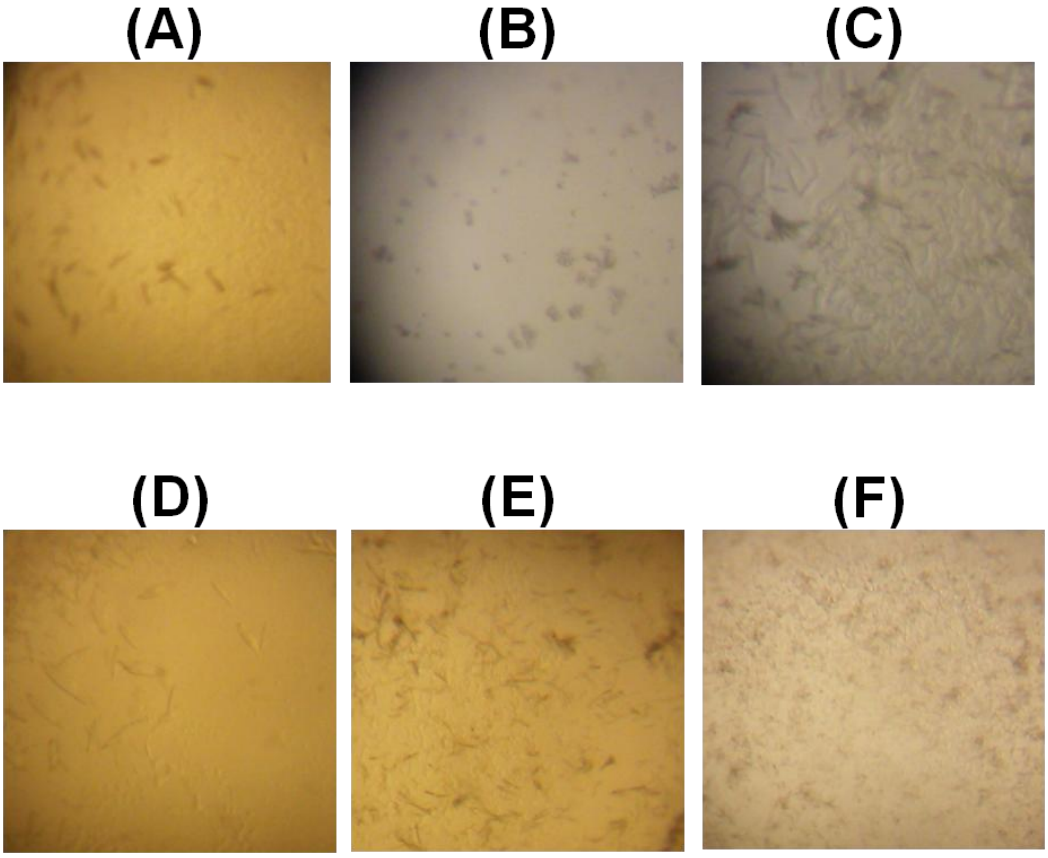

Figure 4.2 Initial $BamC_{UN}DE_{\Delta C}$ crystals

These images of initial $BamC_{UN}DE_{\Delta C}$ crystals were taken two weeks after the crystallization experiments were set up. The images were taken from wells containing the conditions (A) 2.0 M sodium chloride, 10% w/v polyethylene glycol 6000; (B) 0.1 M sodium acetate trihydrate pH 4.6, 2.0 M sodium chloride; (C) 0.2 M sodium thiocyanate 20% w/v polyethylene glycol 3350; (D) 0.2 M sodium phosphate dibasic dihydrate, 20% w/v polyethylene glycol 3350; (E) 0.2 M potassium phosphate dibasic, 20% w/v polyethylene glycol 3350; and (F) 0.2 M ammonium phosphate dibasic, 20% w/v polyethylene glycol 3350.

The results presented here show that $BamC_{UN}DE_{\Delta C}$ is a good candidate for crystallization. Having produced initial crystals provides a promising start for this investigation, and optimization of the conditions may yield crystals suitable for X-ray diffraction studies. Solving the structure of $BamC_{UN}DE_{\Delta C}$ should be a priority as it will bring valuable knowledge in understanding how this subcomplex assembles, which will give insight into how the entire BAM complex may assemble.

5. Assembly of the entire BAM complex

5.1. Strategy

Chapters 2-4 have presented detailed studies into understanding the interactions between BamC, BamD, and BamE. However, the BAM complex also includes BamA and BamB which were not investigated. This chapter provides an overview of the initial studies that have been conducted to understand how BamA and BamB fit into the story presented in the previous chapters. These experiments were conducted in collaboration with Kelly Kim and Jonathan Tan in the Paetzel Lab. It should be reminded that BamA is a membrane protein (it is an OMP), so for these experiments only the periplasmic region containing the five POTRA domains were used, and this construct was termed $BamA_{POTRA}$. Table 5.1 provides a list of all constructs used, their amino acid boundaries and molecular mass. All proteins lack the N-terminal signal sequence while the lipoproteins also have the lipidation site removed. The proteins contain N-terminal hexahistidine tags; additionally, untagged versions of BamC, BamD, and BamE were also used to confirm some of the results.

Table 5.1 List of constructs used to study assembly of the BAM complex

Construct Name	Residue Boundaries	Molecular Mass (kDa)
$BamA_{POTRA}$	21-434	46.4
BamB	21-392	39.7
BamC	26-344	34.3
BamD	21-245	25.7
BamE	21-113	10.3

The molecular mass is calculated based on the amino acid composition of the protein construct.

5.2. Gel-filtration chromatography studies

5.2.1. $BamA_{POTRA}$ forms a dimer with BamB

When $BamA_{POTRA}$ was co-lysed and purified with BamB, the gel-filtration chromatogram was difficult to analyse due to overlapping peaks (Figure 5.1A). A large peak at the void volume was observed (labelled as 1), which is similar to what has been seen when BamB is individually run on the column (see Figure C2A in appendix), and the corresponding SDS-PAGE gel shows only the presence of BamB. However, the main peak (labelled as 2 and 3) contains both $BamA_{POTRA}$ and BamB, suggesting that a complex may have formed. This is supported by the fact that BamB is eluting earlier than what has been observed when it is analyzed individually through gel-filtration. However, as no significant shift was observed for $BamA_{POTRA}$, this data does not provide definitive evidence of complex formation. So another approach was used: purified BamB was digested with thrombin to remove the hexahistidine tag (see Figure C2B in appendix). The resulting untagged BamB was incubated overnight with purified tagged $BamA_{POTRA}$ (with tag intact), and run through nickel affinity chromatography. As BamB no longer had the hexahistidine tag, it could only be retained on the nickel affinity column if it formed a complex with $BamA_{POTRA}$. The resulting SDS-PAGE gel from this experiment shows untagged BamB to co-elute with $BamA_{POTRA}$, strongly demonstrating that a complex had formed (Figure 5.1B). Thus, $BamA_{POTRA}$ is able to form a complex with BamB, and based on the elution profile from gel-filtration chromatography, it is believed to be a dimer.

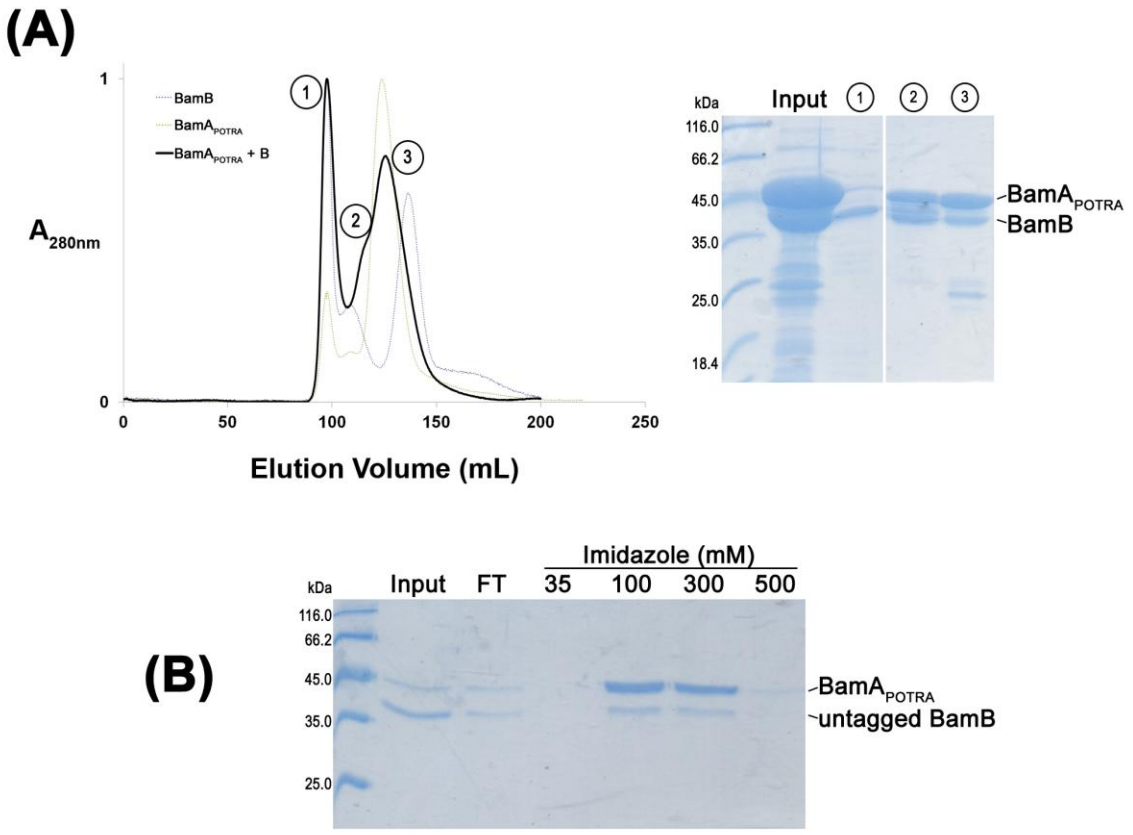

Figure 5.1 Gel-filtration chromatogram and SDS-PAGE showing BamA$_{POTRA}$ to form a dimer with BamB

(A) presents the gel-filtration chromatogram (left) showing the elution profile for the sample containing BamA$_{POTRA}$ and BamB (black curve), which is plotted as the relative UV absorbance at 280nm (A_{280nm}, provided in arbitrary units) for protein eluted at a specific volume (Elution Volume, provided in millilitres (mL)). The results show the presence of two peaks which are labelled as 1 and 3, and a shoulder labelled as 2. Peak 1 elutes much earlier than individual BamB (dotted purple curve), suggesting that a complex may have formed. However, this peak overlaps with the peak for individual BamA$_{POTRA}$ (dotted green curve). The corresponding SDS-PAGE gel (right) suggests that a complex formed as both BamA$_{POTRA}$ and BamB elute together. (B) shows the second experiment that was done to confirm interaction. Untagged BamB was successfully able to be retained on the nickel affinity column, suggesting that it was able to form a complex with BamA$_{POTRA}$, which contained the hexahistidine tag. For an SDS-PAGE gel confirming the removal of the hexahistidine tag from BamB prior to this experiment, please see Figure C2B. Experiment notes: a HiPrep 26/60 Sephacryl S-100 HR gel-filtration column was used attached to the ÄKTAprime system, with a void volume of 98 mL and a column volume of 320 mL. The experiment was run at 4°C with a buffer containing 20 mM Tris-HCl (pH 8.0) and 100 mM NaCl, where the flow rate was set to 1 mL/min, with 3 mL fractions collected. Fractions eluting in or around the void volume are not shown on the SDS-PAGE gel as they were not in the region of interest and contained aggregated protein.

5.2.2. $BamA_{POTRA}$ forms a subcomplex with BamD and BamE

When $BamA_{POTRA}$ was co-lysed and purified with BamD, it was expected that the two would form a dimer as previous research has shown direct interaction between BamA and BamD, the two essential components of the BAM complex (Malinverni et al., 2006). Furthermore, it has been proposed that specifically the fifth POTRA domain is required for this interaction, which is included in the $BamA_{POTRA}$ construct (Kim et al., 2007). Interestingly, our gel-filtration chromatography results showed no complex formation between $BamA_{POTRA}$ and BamD (Figure 5.2A). Thus it was hypothesized that the absence of the membrane embedded β-barrel domain of BamA may be preventing proper association between the proteins, especially since the POTRA5 domain required for interaction is the closest POTRA domain to the membrane. A previous study also showed that in the presence of BamE, the interaction between BamA and BamD is strengthened, suggesting that BamE may play a stabilization role (Sklar et al., 2007). So BamE was introduced into this experiment to see if it could assist in complex formation.

When $BamA_{POTRA}$ was co-lysed and purified with BamD and BamE, the resulting gel-filtration chromatogram and SDS-PAGE gel show that a $BamA_{POTRA}$-BamD-BamE subcomplex formed (Figure 5.2B). Both BamD and BamE elute much earlier than where they are expected to elute individually, and much earlier than where the BamDE dimer is expected to elute. The results presented here appear to support the theory proposed by Sklar et al., and suggest that while $BamA_{POTRA}$ is unable to form a stable dimer with BamD, the addition of BamE forms a stable subcomplex, even in the absence of the membrane embedded β-barrel domain of BamA (Sklar et al., 2007). This demonstrates the importance of BamE in this interaction, and suggests that BamE might play a stabilization role in the BamA-BamD association.

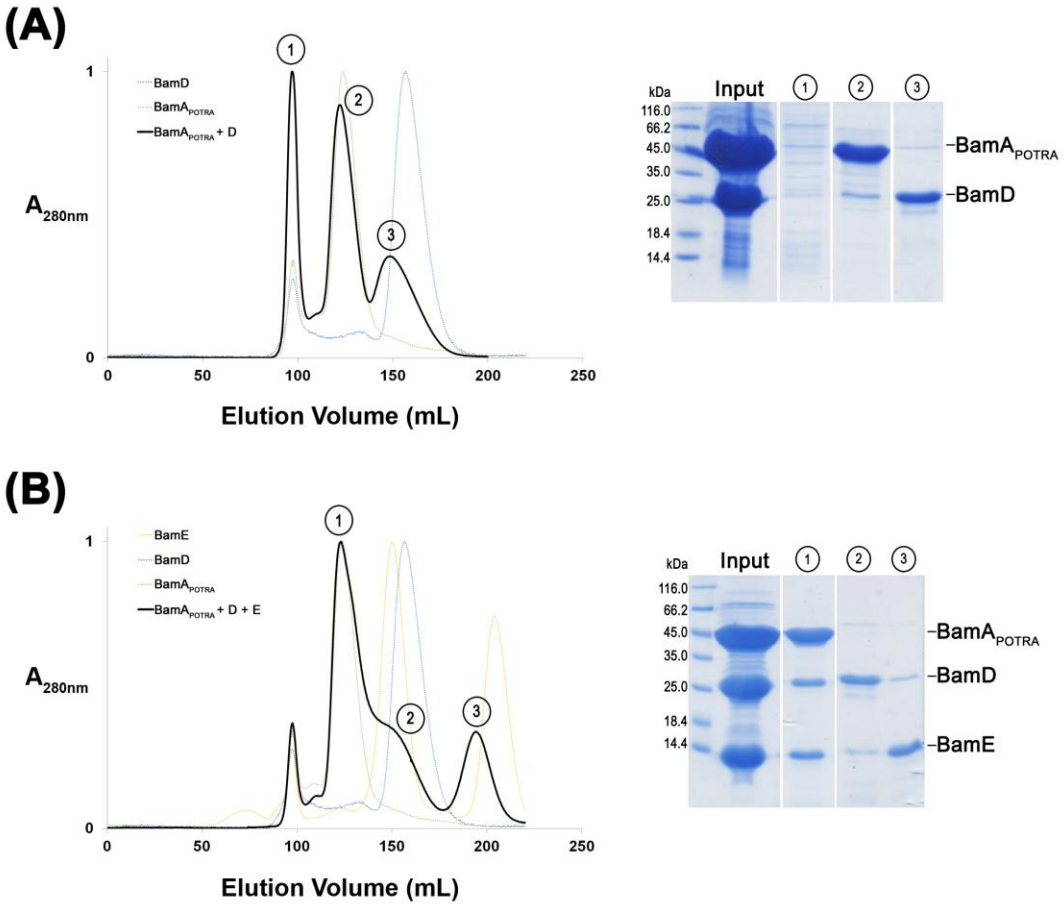

Figure 5.2 Gel-filtration chromatogram and SDS-PAGE showing BamA$_{POTRA}$ to form a subcomplex with BamD and BamE

The gel-filtration chromatogram (left) shows the elution profile for a sample which is plotted as the relative UV absorbance at 280nm (A_{280nm}, provided in arbitrary units) for protein eluted at a specific volume (Elution Volume, provided in millilitres (mL)). The corresponding SDS-PAGE gel is shown on the right. (A) shows the results for a sample containing $BamA_{POTRA}$ and BamD (black curve), and shows the two proteins to elute in their monomeric forms: dotted green curve for monomeric $BamA_{POTRA}$ and dotted blue curve for monomeric BamD. (B) shows the second experiment where the sample contained $BamA_{POTRA}$, BamD, and BamE (black curve). The first peak elutes earlier than individual BamD (dotted blue curve), individual BamE (dotted yellow curve), and the BamDE dimer (not shown in this figure). The SDS-PAGE gel shows the presence of all three proteins in that peak, and indicating that a complex formed. This suggests that BamE may help to stabilize the interaction between BamA and BamD. Experiment notes: a HiPrep 26/60 Sephacryl S-100 HR gel-filtration column was used attached to the ÄKTAprime system, with a void volume of 98 mL and a column volume of 320 mL. The experiment was run at 4°C with a buffer containing 20 mM Tris-HCl (pH 8.0) and 100 mM NaCl, where the flow rate was set to 1 mL/min, with 3 mL fractions collected. Fractions eluting in or around the void volume are not shown on the SDS-PAGE gel as they were not in the region of interest and contained aggregated protein.

5.2.3. *BamA$_{POTRA}$ is unable to compete with BamC*

With data showing BamA$_{POTRA}$-BamD-BamE and BamCDE subcomplexes to form, it was hypothesized that all four proteins should also be able to assemble into a tetramer. However, when BamA$_{POTRA}$ was co-lysed and purified with BamC, BamD, and BamE, no obvious tetramer was observed. Instead, the gel-filtration chromatogram shows a broad peak, which the SDS-PAGE gel indicates to be a combination of possible BamCDE and BamA$_{POTRA}$-BamD-BamE subcomplexes (Figure 5.3A). Based on gel-filtration chromatography alone it is difficult to really understand what complexes are forming in this situation, and if a small population of tetramer did form. Thus, nickel affinity chromatography was performed using hexahistidine tagged BamA$_{POTRA}$, and untagged BamC, BamD, and BamE. The resulting SDS-PAGE gel shows the tagged BamA$_{POTRA}$ to elute separately, while BamC, BamD, and BamE came out in the flow-through, most likely forming a BamCDE subcomplex (Figure 5.3B). This is an interesting result as it shows that a tetramer between BamA$_{POTRA}$, BamC, BamD, and BamE is unable to form. Also interesting is the absence of a BamA$_{POTRA}$-BamD-BamE subcomplex that was observed in Section 5.2.2. Instead, only the BamCDE subcomplex was able to form, which suggests that the BamCDE subcomplex is favoured over the BamA$_{POTRA}$-BamD-BamE subcomplex. Furthermore, it can be proposed that BamA$_{POTRA}$ and BamC share overlapping binding regions on BamD, which would explain why only one of them can be involved in complex formation at one time.

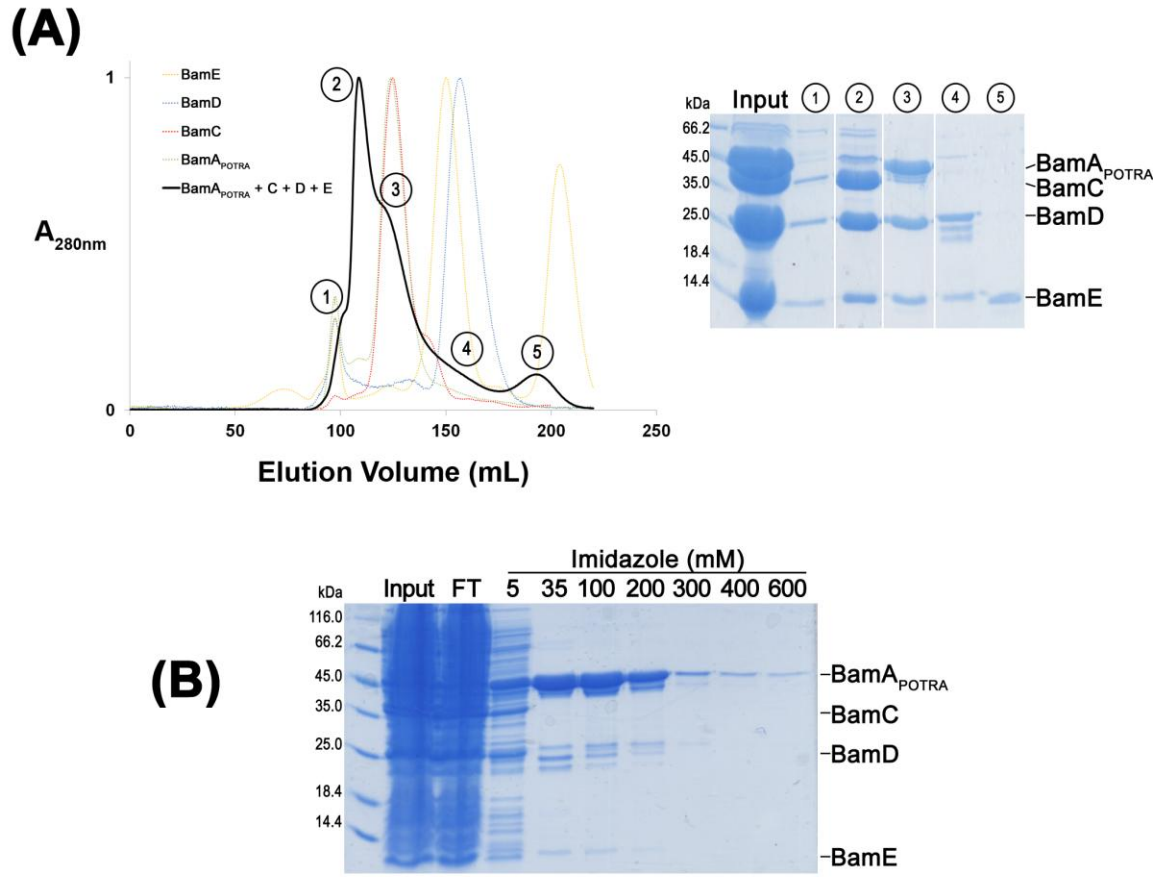

Figure 5.3 Gel-filtration chromatogram and SDS-PAGE showing BamA$_{POTRA}$ unable to compete with BamC

(A) presents the gel-filtration chromatogram (left) showing the elution profile for the sample containing BamA$_{POTRA,}$ BamC, BamD, and BamE (black curve), which is plotted as the relative UV absorbance at 280nm (A$_{280nm}$, provided in arbitrary units) for protein eluted at a specific volume (Elution Volume, provided in millilitres (mL)). The results show the presence of one main peak with a broad shoulder. From the corresponding SDS-PAGE gel it is difficult to determine if a BamA$_{POTRA}$-BamCDE tetramer was able to form, or if only BamCDE and BamA$_{POTRA}$-BamDE subcomplexes formed. (B) shows the result from a nickel affinity chromatography experiment with BamA$_{POTRA}$ contained the hexahistidine tag. Untagged BamC, BamD, and BamE were unable to complex with BamA$_{POTRA}$ in this situation and eluted in the flow through (FT), and most likely formed a BamCDE subcomplex. Interestingly, no BamA$_{POTRA}$-BamDE subcomplex was observed either, suggesting that the BamCDE subcomplex may be favoured. Note that the other bands present in the elution fractions are degradation products of BamA$_{POTRA}$ (see Figure C1 for more details). Experiment notes: a HiPrep 26/60 Sephacryl S-100 HR gel-filtration column was used attached to the ÄKTAprime system, with a void volume of 98 mL and a column volume of 320 mL. The experiment was run at 4°C with a buffer containing 20 mM Tris-HCl (pH 8.0) and 100 mM NaCl, where the flow rate was set to 1 mL/min, with 3 mL fractions collected. Fractions eluting in or around the void volume are not shown on the SDS-PAGE gel as they were not in the region of interest and contained aggregated protein.

5.2.4. Summary of gel-filtration studies with $BamA_{POTRA}$

The data presented in Sections 5.2.1 to 5.2.3 show that $BamA_{POTRA}$ is able to form complexes with BamB, as well as with BamD and BamE. However, in the presence of BamC, $BamA_{POTRA}$ is unable to associate with BamD and BamE, suggesting that there may be a competition between BamC and $BamA_{POTRA}$, and perhaps even overlapping binding pockets. However, it should be noted that these were only preliminary trials and that further investigation should be done to confirm some of these interactions. For example, $BamA_{POTRA}$ and BamC should be co-lysed and purified to see if the two proteins are able to associate with each other. And to investigate possible competition between these two proteins, the experiment with $BamA_{POTRA}$, BamC, BamD, and BamE could be repeated with excess amounts of $BamA_{POTRA}$, in an attempt to see if the $BamA_{POTRA}$-BamD-BamE subcomplex can be formed in the presence of BamC.

Also, it is interesting to observe that despite the presence of various lipoproteins, the elution volume of $BamA_{POTRA}$ did not appear to change, as the protein (either alone or in complex) eluted at ~125 mL on the gel-filtration column (Table 5.2). It is hypothesized that the extended nature of $BamA_{POTRA}$ is causing this phenomenon, as gel-filtration chromatography separates proteins based on its Stokes radius, rather than its molecular mass (Laurent and Killander, 1964; Siegel and Monty, 1966). However, the relationship between elution volume and mass through gel-filtration chromatography is valid for globular proteins, and thus often used. But as described in Section 1.5.2.1, $BamA_{POTRA}$ does not have a globular structure. The “extended” and “bent” conformations that have been observed would cause $BamA_{POTRA}$ to elute earlier than what is expected for its mass. In addition, the association of lipoproteins may not alter the overall Stokes radius (unless there is a head to tail association that elongates the overall structure), and thus a significant shift in elution volume would not be observed. Further discussion of Stokes radius is provided in Section 6.7 and in Appendix B.

Table 5.2 Gel-filtration elution volumes for studies involving $BamA_{POTRA}$

Combination	Peak Elution Volume (mL)	Measured Mass (kDa)	Protein(s) Eluted and their Calculated Mass (kDa)	
$BamA_{POTRA}$	**123**	57	$BamA_{POTRA}$	46.3
BamB	135	44	BamB	39.7
BamC	125	55	BamC	34.3
BamD	155	29	BamD	25.7
BamE	150	32	BamE dimer	20.6
	204	10	BamE monomer	10.3
$BamA_{POTRA}$ + BamB	115	68	$BamA_{POTRA}$-BamB	86.0
	125	55	$BamA_{POTRA}$-BamB	86.0
$BamA_{POTRA}$ + BamD + BamE	**127**	53	$BamA_{POTRA}$-BamDE	82.3
	153	30	BamD	25.7
			BamE dimer	20.6
	198	12	BamE monomer	10.3
$BamA_{POTRA}$ + BamC + BamD + BamE	108	79	BamCDE	70.3
	125	55	$BamA_{POTRA}$-BamDE	82.3
	155	29	BamD	25.7
			BamE dimer	20.6
	196	12	BamE monomer	10.3

This table presents a summary of the peaks observed, their elution volumes, and the proteins eluted. The measured molecular mass was estimated from the standard curve of the gel-filtration column using the corresponding elution volume. The calculated molecular mass of the protein eluted is based on the amino acid composition of each protein. The values in **bold** have been highlighted to emphasize the similarity in elution volume for samples containing $BamA_{POTRA}$.

5.3. Proposed model of complex assembly

The findings presented in this chapter and in Chapters 2-4 provide enough details to propose a model of how the BAM complex may assemble. If BamD is the OMP substrate binding protein, it explains why it is an essential component of the complex. BamA is the other essential component, and somehow the substrate must be recognized and brought to BamA for folding and insertion into the OM. How BamA performs these steps, and where BamD fits in is still under investigation. In *E. coli* and other organisms where BamC is present, this protein may function as a regulator to control OMP assembly. For example, BamC could be bound to BamD in the absence of free BamA to prevent a backlog of substrates entering the periplasm from the cytoplasm, which could have downstream effects such as controlling OMP biosynthesis. BamE joins with BamCD to from BamCDE, which could be involved in stabilizing this complex, as well as stabilizing the later association between BamD and BamA. Then, in the presence of substrate and free BamA, BamC dissociates from the complex (or undergoes a conformational change) allowing the substrate to bind to the proposed OMP-signal binding pocket of BamD. Meanwhile, the periplasmic POTRA domains of BamA also associate to form the BamADE complex with the substrate bound (Figure 5.4). Previous research shows that in the absence of BamB, the OM insertion of mainly larger OMPs is impaired by the BAM complex (Charlson et al., 2006). Thus, BamB's participation at this point of our model may be dependent on the size of the OMP substrate.

It is still unclear whether one monomer or an oligomer of BamA is required for the OMP folding and insertion steps, or what this mechanism involves. Referring back to Section 1.5.3, four possible mechanisms were compared: a) the OMP translocates through a pore formed within BamA, and then inserts and folds into the OM from the extracellular face; b) the OMP inserts directly from the periplasm in between BamA and the OM lipid bilayer and folds; c) the OMP inserts in between an oligomer of BamA proteins, where it is released into the lipid bilayer after folding; or, d) the OMP uses BamA as a template and begins to fold within the BamA barrel, forming a hybrid barrel with BamA, and then buds off into the lipid bilayer to form an independent barrel. Presented here is a model of how the BAM complex itself could assemble, which is necessary in order to determine how the BAM complex is then able to fold its substrates.

Based on data that suggest BamC to not participate with BamA-related subcomplexes, it appears that the BAM complex is a dynamic machine where the lipoproteins are constantly associating and dissociating for OMP folding, rather than staying attached to the core BamA protein.

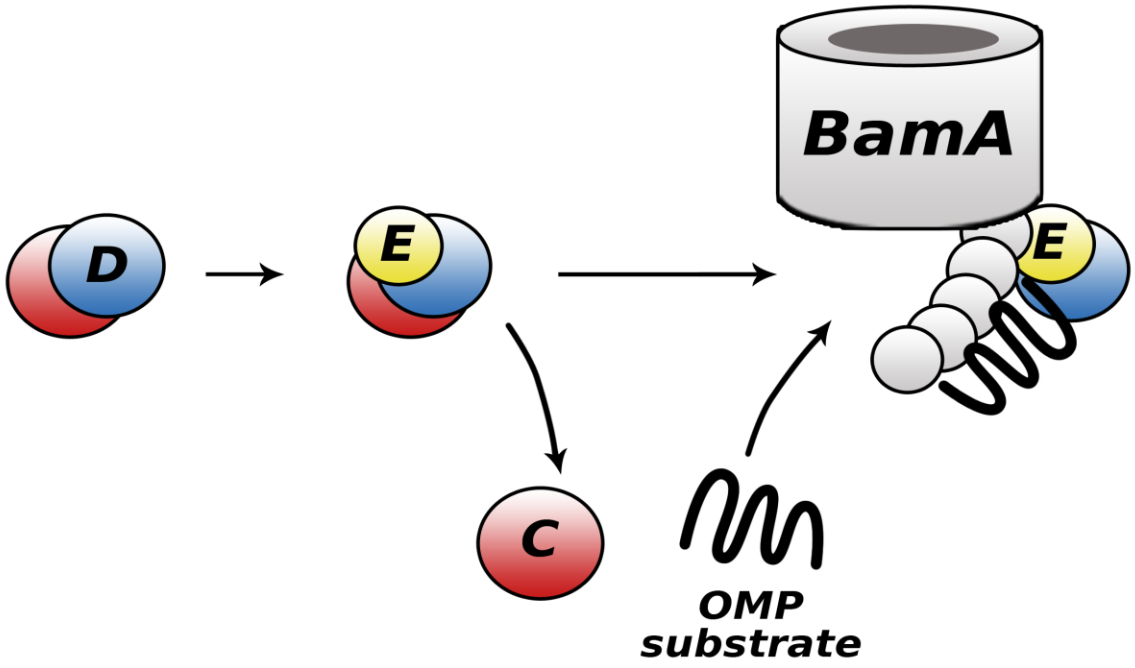

Figure 5.4 Proposed model of assembly of the BAM complex components

This figure provides a model based on the results from this thesis. BamB is not included as it has been shown to interact with BamA independently, and no extensive research was done on BamB in this project. This model suggests that BamC (red) binds to BamD (blue) for regulatory control, by blocking the proposed OMP binding pocket on BamD. Then BamE (yellow) joins to create a BamCDE subcomplex, and prepares to stabilize the later interaction between BamA and BamD. In the presence of OMP substrate, and when free BamA is available, BamC dissociates (or undergoes a conformational change), and a complex is formed between BamA, BamD, BamE, and the OMP substrate.

6. Critique of Experimental Design and Considerations for Future Work

6.1. Proteins removed from native environment

All proteins used in these experiments were expressed without the native N-terminal signal sequence. As a result, the proteins remained in the cytoplasm as opposed to the periplasm where they would be naturally found. Thus, it is possible that the proteins may behave differently and interact differently with their binding partners. However, it should be noted that none of the proteins contain any cysteine residues that would be required for disulfide linkages that would be formed in the oxidizing environment of the periplasm. In addition, the proteins showed good expression, provided good yields, and were soluble. Thus, it can be assumed that the proteins were able to properly fold and associate with their binding partners. However, the other component that is also lacking from the lipoproteins in this study (BamB-E) is the lipidation site that anchors the protein to the inner leaflet of the OM. If BamB-E were in the periplasm, they would be near the lipid bilayer which could alter their interactions. For example, BamE has been shown to bind to phosphatidylglycerol, a lipid that is found in the OM (Knowles et al., 2011). Furthermore, the authors of that research suggest the binding site to be in close proximity of their observed BamD binding site (Knowles et al., 2011). Taking this together with the fact that anchoring of the lipoproteins to the membrane would restrict their geometry raises the question of which protein complexes that are being observed in the cytoplasm are able to naturally occur in the periplasm. The cytoplasm is missing other components that may be interacting with these proteins *in vivo*, thus either allowing or preventing certain complexes from forming. In addition, it is possible that overexpression of large amounts of these proteins may also be driving the oligomerization process (see below). Thus, expressing the periplasmic components of the BAM complex with their signal sequences and directing them to the periplasm may provide a better picture of how they behave *in vivo*.

6.2. Overexpression may promote oligomerization

As discussed above, while overexpression in the cytoplasmic compartment may be a variable to consider, overexpression in general should also be noted. Forcing protein expression in much larger amounts than what is observed *in vivo* could cause certain by-products to form. These could include the formation of protein complexes that may not occur naturally, as well as different oligomeric states. For example, as discussed in Chapter 2, BamE has been observed to exist as a monomer and homodimer, however, the biological relevance of each form is still under debate. There has been research suggesting that the BamE dimer is a result of overexpression, specifically in the cytoplasm where the protein can accumulate and result in the dimeric form (Knowles et al., 2011). In contrast, other research suggests the dimer is not a by-product of overexpression as it also exists in the membrane bound form (Albrecht and Zeth, 2011). The proteins used in this thesis project were overexpressed in the same manner, which resulted in good yields as two litre cell pellets were sufficient for experiments. However, as in the case of BamE, both the dimeric and monomeric forms were observed raising the question if overexpression was the cause for the dimeric BamE. Therefore it is suggested that to better understand the nature of BamE dimerization, as well as the formation of hetero-complexes, different expression conditions should be tested to see if there is an effect, such as slowing the growth over a long time period, and lowering the temperature.

6.3. Co-lysis versus co-expression

Co-lysis was performed for each experiment to allow complex formation before the individual proteins began to degrade. As described in Chapter 3, working with BamC alone resulted in degradation within days of purification, and this could be avoided by co-lysis: co-lysing the cell pellets containing individual proteins, as opposed to separately lysing and purifying the proteins and combining them later. While the use of co-lysis is a strength in these experiments, some results show one drawback of this technique. When expressing truncated versions of proteins that do not form complexes, co-lysis may allow the endogenous full length protein to be co-purified with the potential binding partner. For example, in Section 2.2.2, although no obvious $BamDE_{\Delta N}$, $BamDE_{\Delta C}$, or $BamDE_{\Delta N\Delta C}$

dimers were observed, small peaks were observed in the chromatograms at around 136 mL, where the full length BamDE dimer elutes (Figures 2.8–2.10). It is difficult to conclude if small populations of the truncated dimers did form, or if this was the result of endogenous full length BamE (untagged) associating with the overexpressed BamD (tagged) during co-lysis, and thus allowing it to be co-purified on nickel affinity chromatography prior to gel-filtration. If co-lysis was not performed, only the purified truncated forms of BamE would be present with BamD, and a better interpretation of the chromatogram could be made. Although in this situation the chromatograms provide strong data suggesting no $BamDE_{\Delta N}$, $BamDE_{\Delta C}$, or $BamDE_{\Delta N\Delta C}$ dimers forming, it should be noted that some endogenous proteins may be present when data from co-lysis experiments become ambiguous.

Co-expression is another technique that could be used in place of co-lysis. This method involves multiple proteins to be expressed in the same cell, allowing complex formation *in vivo*, and therefore does not require multiple cell pellets to be lysed together. Also, any protein constructs that might be unstable or insoluble alone could be co-expressed with their binding partners to help stabilize it. For example, the $BamC_U$ construct mentioned in Chapter 3 showed poor expression. However, co-expressing the construct with BamD may stabilize $BamC_U$ as BamD can bind to it and prevent its degradation. While co-expression has many benefits and appears to be easier than co-lysis, it does require proteins to be cloned into compatible plasmids, and cells to retain the numerous plasmids. Alternatively, there are plasmids that contain two multiple cloning sites (MCS) to allow two genes to be expressed from the same plasmid, such as the pETDuet vector from *Novagen*. While this reduces the amount of unique plasmids required, attempts at cloning BAM proteins into this vector yielded expression problems (data not shown). Thus, since co-lysis was successful in earlier experiments, it was the method chosen to continue later studies with.

6.4. Presence of fusion tags

While fusion tags such as histidine or maltose binding protein (MBP) have been widely used for their assistance in purification and solubility (for MBP), there are also some drawbacks (Kapust and Waugh, 1999; Terpe, 2003). One problem with MBP has

been outlined in Section 3.4.2 where it was questioned if the attached passenger proteins in our MBP-fusion constructs were being affected by the presence of MBP, making them inaccessible to interact with their binding partners. However, this could also be caused by the design of the overall fusion construct, where the length of the passenger protein may not be long enough to properly fold or interact as it would in its full length form. While MBP was only a part of a few experiments, the remainder of experiments presented in this thesis involve the use hexahistidine tags. Some researchers have noticed histidine tags to promote oligomerization of their proteins, thus questioning the validity of the experiments conducted here (Wu and Filutowicz, 1999; Amor-Mahjoub et al., 2006). Furthermore, in one of the crystal structures of BamD, the predicted binding pocket for OMP substrates was seen to be occupied by the histidine tag of an adjacent BamD molecule in the crystal (Albrecht and Zeth, 2011). This would cause one to speculate if the protein complexes observed in Chapters 2–5 are true complexes, or a result of histidine tag interference. While some experiments were repeated with the removal of the tag and no differences were observed, the majority were not, and thus it is suggested that all experiments be repeated without the tag to ensure that the complexes observed are not caused by tag interference. In addition, further work can also look into the placement of the tag – if an N-terminal hexahistidine tag does not allow oligomerization, are similar results observed with a C-terminal tag? This would also give insight into whether the terminus with the tag plays an important role in oligomerization.

6.5. Domain truncation versus point mutations

The experiments presented here use domain truncation analysis as a method of identifying which regions of BamC or BamE are involved in complex formation. As noted in Chapters 2 and 3, failure to form complexes could be caused by two factors: either the removed region is needed to directly make the association, or its removal is disrupting the overall secondary structure of the protein. Thus, it could be suggested that other techniques, such as creating point mutations, may provide a better way of locating regions and specific residues required for complex formation without significantly altering the structure. These mutations could be made through site-directed mutagenesis of conserved residues, or through error-prone PCR methods for generating random

mutations. While initial mutagenesis approaches were attempted for this project, the subsequent gel-filtration chromatography experiments showed no significant difference from the wildtype protein for complex formation (data not shown). Therefore, in order to narrow down which region of the proteins may be important, truncations were created and tested for complex formation. From those results, which indicate the N-termini of BamC and BamE to be important, future mutagenesis experiments can be designed to identify specific residues required for complex formation.

6.6. Use of gel-filtration chromatography

To get a better understanding of the requirements of complex formation and the stoichiometry, the role of protein concentration could have been investigated. This would involve measuring the concentration of individual proteins and combining them in different molar ratios. However, as co-lysis was performed, it would not be possible to determine the concentrations of each individual protein sample. So in order to determine the individual concentrations, the proteins would need to be purified separately prior to gel-filtration, which would allow accurate measurement of the concentrations. Furthermore, the proteins could also be combined in different molar ratios to determine the stoichiometry. In addition to protein concentration, the selection of gel-filtration column matrix would also have an effect on the results. In these experiments, a *HiPrep 26/60 Sephacryl S-100 HR* gel-filtration column was used due to its ability to resolve proteins between 1 - 100 kDa, which is suitable for the proteins studied here. However, in terms of resolution, the manufacturers suggest using the *Superdex* columns, which have much smaller resin beads resulting in better resolution that would help to separate the overlapping peaks seen in Chapter 5 with $BamA_{POTRA}$. In hindsight, the *Superdex* columns may have provided better results, however, the *Sephacryl S-100* has a separation range that covers all the BAM proteins and their complexes, whereas multiple *Superdex* columns (*Superdex 75* and *Superdex 200*) would be required to cover this range. Also, based on the larger size of the *Sephacryl* column, and its ability to purify large amounts of protein at an industrial scale, the selection of this column is justified by the desire to use the eluted proteins for further crystallization studies which require high concentrations (Amersham Biosciences, 2002).

6.7. Stokes radius versus molecular mass

In the experiments presented here, gel-filtration chromatography is used to determine complex formation. Based on a standard curve created for the column, the molecular mass of the eluted protein is calculated using the elution volume. However, a more correct way to use this technique is to calculate the Stokes radius from the elution volumes, as molecular mass estimation is only valid for certain proteins, specifically those with a globular structure (Laurent and Killander, 1964; Siegel and Monty, 1966). This explains why certain proteins eluted earlier than expected by their mass: the unstructured regions of BamC and of BamE, along with the extended nature of the BamE dimer and of $BamA_{POTRA}$ all caused these proteins to elute early. In addition, $BamA_{POTRA}$ constantly eluted at a similar volume despite sometimes being in a complex, suggesting that its overall Stokes radius was not significantly affected (see Chapter 5). Thus it may seem like the data provided in this thesis would be better presented if the proteins were compared using their Stokes radius instead of their molecular mass. However, it should be noted that even though the structures of the proteins studied here are now available, determining the Stokes radius is not as simple as measuring the radius of the protein molecule. The Stokes radius is also known as the hydrodynamic radius and it takes into consideration the hydration sphere of the protein molecule, as well as its behavior in the chromatography medium. To calculate it requires more variables including the viscosity of the solution (Siegel and Monty, 1966; Winzor, 2011). Thus in the experiments conducted here, using molecular mass provides a good estimation even for the non-globular proteins, as the purpose of these studies is to generally determine which complexes are able to form. Further discussion of Stokes radius is provided in the Appendix B.

6.8. Additional techniques to confirm interaction

The experiments presented here provide a good foundation from which further work can be done to study the complexes of the periplasmic components of the BAM complex. However, while gel-filtration chromatography was used as the primary method for detecting complex formation, it is not the only one. For a better understanding of the protein-protein interactions between the lipoproteins, as well as $BamA_{POTRA}$, other

techniques can be used to provide more knowledge of the stoichiometry and binding kinetics. For example, dynamic light scattering can provide a better estimation of the molecular mass of the overall complex, as well as the hydrodynamic radius. Protein-protein interactions can also be monitored using fluorescence, isothermal titration calorimetry, or surface plasmon resonance experiments which can provide binding constants to understand how tight the interactions are, and which interactions are more favoured. But while many of these techniques may require additional equipment, other simple methods could also be employed to further confirm these complex formations, such as cross-linking and native PAGE gels. In this thesis project, the secondary method chosen to confirm complex formation was X-ray crystallography in an attempt to visually see the complexes. However, due to the time required to fully optimize conditions for crystallizations of the various complexes, focus was kept on the gel-filtration studies.

7. Conclusion and Future Directions

The BAM complex is an essential protein complex found in the OM of all Gram-negative bacteria, with homologous systems also existing in the mitochondria and chloroplasts of eukaryotes. Because of its essential and non-redundant role, as well as its location near the periphery of the cell, the BAM complex is being studied as a potential drug target. However, before any studies assessing its ability as a drug target can be performed, the structure and function of this complex must first be thoroughly examined. Previous research has provided a basic outline of how the individual components may interact with each other and OMP substrates, and the recent emergence of multiple structures is providing insight into what the complex may look like. The work presented in this thesis tries to further examine the protein-protein interactions within the periplasmic components of the BAM complex by using truncation analysis to identify regions of BamC and BamE that may be required for complex formation.

Using structural data that was available, various truncations were made on BamC and BamE to isolate or remove certain regions of interest. Gel-filtration chromatography was used as the primary method for detecting complex formation. These co-purification studies were designed to determine which combinations of proteins would readily form stable complexes, and stay intact during gel-filtration chromatography. A full summary of these gel-filtration studies is provided in Table 7.1. The resulting complexes would then indicate which regions of the proteins are required for oligomerizaton, and what combinations can be used for further study. The results presented here strongly suggest that the unstructured N-termini of BamC and BamE are required for BamCDE subcomplex formation, which led to the successful formation of the $BamC_{UN}DE_{\Delta C}$ complex that was able to successfully crystallize. The preliminary interaction studies in Chapter 5 also provide a model for how the complex may assemble. The top priority for future research would be to optimize the crystallization

conditions of the $BamC_{UN}DE_{\Delta C}$ complex so that a diffraction quality crystal can be produced, from which the first structure of the BamCDE subcomplex can be determined.

Aside from structural studies and experiments proposed in Chapter 6, there are a number of other approaches that can be taken to quickly proceed with further examining these interactions. For example, Chapter 2 provides a detailed investigation into determining the involvement of the N- and C-termini of BamE, but the long flexible L3 loop of BamE could not be studied. This would be an interesting region to study as structural homologues of BamE share this common feature and have been shown to play important roles in their function (Kim et al., 2011c). Simple mutagenesis studies of the conserved residues, or attempts to remove or shorten the loop could be made to investigate the effects on complex formation.

The studies in Chapter 3 showed the N-terminus of BamC to be required for complex formation, however, attempts to determine if this $BamC_U$ region is sufficient for interaction were unsuccessful as expressing the region alone was difficult, and constructing an MBP-fusion protein did not provide conclusive results. Another method of testing this theory would be by co-expressing the $BamC_U$ region with BamD. Since both constructs are available in our lab, this approach would simply require the two proteins to be placed in complimentary vectors with different antibiotic markers, and transformed into the same cells. Co-expressing BamD may assist in the $BamC_U$ expression and prevent it from being degraded, if a stable $BamC_UD$ dimer is able to form. Alternatively, if the MBP-$BamC_U$ fusion construct is to be further investigated, the $BamC_U$ region could be placed at the N-terminus of MBP, to mimic the environment of the full length BamC protein where $BamC_U$ is located at the N-terminus. And finally, another approach would involve taking the current $BamC_{UN}$ construct, and mutating the conserved residues that have shown to directly contact BamD in the BamCD dimer structure (Kim et al., 2011a). This experiment seems to be the most promising as $BamC_{UN}$ is already known to express well, and the mutagenesis study would identify key residues that may be necessary for the interaction.

From the data presented in this thesis, there is a better understanding of the protein-protein interactions within the BAM complex. Truncation analysis coupled with gel-filtration chromatography provides a good approach for dissecting various

complexes. While further research is needed to fully understand the mechanism of the BAM complex, Chapter 5 provides a reasonable model of how the periplasmic components may come together. By first understanding how the individual BAM proteins fit together, we can then better investigate how the complex may function overall.

Table 7.1 Summary of gel-filtration studies of the periplasmic components of the BAM complex

Combination	Peak Elution Volume (mL)	Measured Mass (kDa)	Protein(s) Eluted and their Calculated Mass (kDa)	
BamC	125	55	BamC	34.3
BamD	155	29	BamD	25.7
BamE	150	32	BamE dimer	20.6
	204	10	BamE monomer	10.3
BamC + BamD + BamE	112	73	BamCDE complex	70.3
	136	43	BamDE dimer	36.0
BamC + BamD + $BamE_{\Delta N}$	121	60	BamCD dimer	60.0
	173	20	BamD	25.7
	208	9	$BamE_{\Delta N}$ monomer	8.2
BamC + BamD + $BamE_{\Delta C}$	117	65	$BamCDE_{\Delta C}$ complex	69.7
	164	24	BamD	25.7
			$BamE_{\Delta C}$ dimer	19.4
BamC + BamD + $BamE_{\Delta N\Delta C}$	122	59	BamCD dimer	60.0
	175	19	BamD	25.7
			$BamE_{\Delta N\Delta C}$ dimer	15.2
BamD + BamE	136	43	BamDE dimer	36.0
	153	30	BamD	25.7
			BamE dimer	20.6
	195	12	BamE monomer	10.3
BamD + $BamE_{\Delta N}$	158	27	BamD	25.7
			$BamE_{\Delta N}$ dimer	16.4
	213	8	$BamE_{\Delta N}$ monomer	8.2
BamD + $BamE_{\Delta C}$	153	30	BamD	25.7
	195	12	$BamE_{\Delta C}$ monomer	9.7

BamD + $BamE_{\Delta N\Delta C}$	153	30	BamD	25.7
	217	8	$BamE_{\Delta N\Delta C}$ monomer	7.6
$BamE_{\Delta N}$	163	24	$BamE_{\Delta N}$ dimer	16.4
	219	7	$BamE_{\Delta N}$ monomer	8.2
$BamE_{\Delta C}$	152	31	$BamE_{\Delta C}$ dimer	19.4
	204	10	$BamE_{\Delta C}$ monomer	9.7
$BamE_{\Delta N\Delta C}$	169	21	$BamE_{\Delta N\Delta C}$ dimer	15.2
	217	8	$BamE_{\Delta N\Delta C}$ monomer	7.6
BamC + BamD	121	60	BamCD dimer	60.0
$BamC_N$ + BamD	150	32	BamD	25.7
	175	19	$BamC_N$	13.2
$BamC_C$ + BamD	151	31	BamD	25.7
	175	19	$BamC_C$	13.4
$BamC_{NC}$ + BamD	142	38	$BamC_{NC}$	27.3
	159	27	BamD	25.7
$BamC_{UN}$ + BamD	126	54	$BamC_{UN}$D dimer	46.4
$BamC_N$ + BamD + BamE	129	50	BamDE dimer	36.0
	180	17	$BamC_N$	13.2
$BamC_C$ + BamD + BamE	133	46	BamDE dimer	36.0
	174	19	$BamC_C$	13.4
	193	13	BamE monomer	10.3
$BamC_{NC}$ + BamD + BamE	144	37	$BamC_{NC}$	27.3
	158	27	BamD BamE dimer	25.7 20.6
	197	12	BamE monomer	10.3
$BamC_{UN}$ + BamD + BamE	125	55	$BamC_{UN}$DE complex	56.7
	137	42	BamDE dimer	36.0
$BamC_{UN}$ + BamD + $BamE_{\Delta C}$	122	59	$BamC_{UN}DE_{\Delta C}$ complex	56.1
	149	33	BamD $BamE_{\Delta C}$ dimer	25.7 19.4
MBP-$BamC_U$ + BamD	142	38	MBP-$BamC_U$	50.2
	158	27	BamD	25.7

MBP-BamC$_U$ + BamC + BamD	122	59	BamCD dimer	60.0
	139	41	MBP-BamC$_U$	50.2
	164	24	BamD	25.7
MBP-PhoE + BamD	139	41	MBP-PhoE	43.6
	158	27	BamD	25.7
MBP-PhoE + BamC + BamD	116	67	BamCD dimer	60.0
	140	40	MBP-PhoE	43.6
	157	28	BamD	25.7
BamA$_{POTRA}$	123	57	BamA$_{POTRA}$	46.3
BamB	135	44	BamB	39.7
BamA$_{POTRA}$ + BamB	115	68	BamA$_{POTRA}$-BamB	86.0
	125	55	BamA$_{POTRA}$-BamB	86.0
BamA$_{POTRA}$ + BamD + BamE	127	53	BamA$_{POTRA}$-BamDE	82.3
	153	30	BamD	25.7
			BamE dimer	20.6
	198	12	BamE monomer	10.3
BamA$_{POTRA}$ + BamC + BamD + BamE	108	79	BamCDE	70.3
	125	55	BamA$_{POTRA}$-BamDE	82.3
	155	29	BamD	25.7
			BamE dimer	20.6
	196	12	BamE monomer	10.3

This table shows the main peaks in which complexes formed, and does not include all minor peaks or shoulders in which some of the individual proteins may have eluted. The measured molecular mass was estimated from the standard curve of the gel-filtration column using the corresponding elution volume. The calculated molecular mass of the protein eluted is based on the amino acid composition of each protein. Note: for the combinations containing BamA$_{POTRA}$, no significant shifts in elution volume are observed due to the hypothesis that the overall Stokes radius is not changing when BamA$_{POTRA}$ is associating with the lipoproteins (see Section 5.2.4).

Appendices

Appendix A: Materials and Methods

Cloning

The cloning strategy for each protein construct used is outlined in Table A1. Each construct was cloned into either a pET vector (*Novagen*) which has a kanamycin resistance marker, or into pMAL-c2x (*New England BioLabs*) which has an ampicillin resistance marker. Standard cloning procedures were followed.

Table A1 Cloning details of the constructs used in this project

Construct	Residues	Vector (antibiotic)	Expression Host	Forward Primer (5' to 3')	Reverse Primer (5' to 3')	Restriction Enzyme	Molecular Mass (kDa)	MWP #
$BamA_{POTRA}$	21-433	pET-28a (Kan^R)	BL21(DE3)	ggtattggttacggttaagaaag tggcgtgagc	gctcacgccactttcttaaccgtaa ccaatacc	Site directed mutagenesis	46.4	849
BamB	21-392	pET-28a (Kan^R)	BL21(DE3)	atatacatatgtcgctgtttaacagc	atatactcgagttattaacgtgtaataga	NdeI / XhoI	39.7	872
BamC	26-344	pET-28a (Kan^R)	BL21(DE3)	atatacatatgagttctgactcacgc	tataaagcttttattacttgctaaacgcag	NdeI / XhoI	34.3	873
untagged BamC	26-344	pET-24a (Kan^R)	BL21(DE3)	atatacatatgagttctgactcacgc	tataaagcttttattacttgctaaacgcag	NdeI / XhoI	34.3	733
MBP-$BamC_U$	26-100	pMAL-c2x (Amp^R)	BL21(DE3)	tatagagctcagttctgactcacgc	tatagaattcttacgtgaactgggtacg	SacI/EcoRI	50.2	856
$BamC_{UN}$	26-217	pET-24a (Kan^R)	Tuner	atatacatatgagttctgactcacgc	tatactcgagcgcggcgtcagtggc	NdeI / XhoI	20.7	857
$BamC_N$	99-217	pET-24a (Kan^R)	BL21(DE3)	tatacatatgttcacgggcgatacc	tatactcgagcgcggcgtcagtggc	NdeI / XhoI	13.2	858
$BamC_C$	220-344	pET-24a (Kan^R)	BL21(DE3)	tatacatatggcgcaaaatcgtgcc	tatactcgagcttgctaaacgcagc	NdeI / XhoI	13.4	859
$BamC_{NC}$	94-344	pET-28a (Kan^R)	BL21(DE3)	tatattcatatgggcgcgcgtaccca	tataaagcttttattacttgctaaacgcag	NdeI / HindIII	27.3	860
BamD	21-245	pET-28a (Kan^R)	BL21(DE3)	atatacatatgtcggggtcaaaggaa	atatactcgagtgtattgctgct	NdeI / XhoI	25.7	874
untagged BamD	21-245	pET-24a (Kan^R)	BL21(DE3)	atatacatatgtcggggtcaaaggaa	atatactcgagtgtattgctgct	NdeI / XhoI	25.7	734
BamE	21-113	pET-28a (Kan^R)	BL21(DE3)	atacatatgtccactctggag	tatactcgagttattagttaccactc	NdeI / XhoI	10.3	577
untagged BamE	21-113	pET-24a (Kan^R)	Tuner	atacatatgtccactctggag	tatactcgagttattagttaccactc	NdeI / XhoI	10.3	735
$BamE_{\Delta N}$	39-113	pET-28a (Kan^R)	BL21(DE3)	tatacatatgaccgctaacgacgtatcc	atatctcgagttattagttaccactcagcg	NdeI / XhoI	8.2	579
$BamE_{\Delta C}$	21-107	pET-28a (Kan^R)	Tuner	tatacatatgtccactctggagcga	atatgcggccgcttatttgttatcaatatt	NdeI / NotI	9.7	861
$BamE_{\Delta N\Delta C}$	39-107	pET-28a (Kan^R)	BL21(DE3)	tatacatatgaccgctaacgacgtatcc	atatgcggccgcttatttgttatcaatatt	NdeI/NotI	7.6	581
MBP-PhoE	342-351	pMAL-c2x (Amp^R)	BL21(DE3)	ctagaattgttgctgttggtatgacttat caatttttaaa	agcttttaaaattgataagtcatacc aacagcaacaatt	XbaI/HindIII	43.6	850

These constructs are variations of the *E. coli* proteins BamA (UniProt ID: P0A940), BamB (P77774), BamC (P0A903), BamD (P0AC02), BamE (P0A937), and PhoE (P02932). MWP# refers to the unique identification number given to each construct when it is stored in the Paetzel Lab database.

Protein Expression

100 mL of autoclaved LB media containing the correct antibiotic (50 µg/mL kanamycin for constructs in the pET vectors or 100 µg/mL ampicillin for constructs in the pMAL-c2x vector) was inoculated with the colony carrying the desired plasmid. The culture was left to grow overnight at 37°C on a 250rpm shaker. The following morning, 2 L of LB media with 50 µg/mL of kanamycin or 100 µg/mL of ampicillin was inoculated with 20 mL of this overnight culture. The cells were grown on a 250rpm shaker at 37°C for 3 hours (until the OD_{600}≈0.6). Then, isopropyl β-D-1-thiogalactopyranoside (IPTG) was added to a final concentration of 1 mM, and induction was carried out for 3 more hours at 37°C on a shaker. The cells were spun down using the JLA-10.5 rotor at 6000xg for 5 minutes at 4°C. The pellet was stored at -80°C.

Cell Lysis

The cell pellet was thawed and resuspended in buffer containing 20 mM Tris-HCl (pH 8.0) and 100 mM NaCl. One tablet of the *Roche Complete, Mini, EDTA-free Protease Inhibitor Cocktail* was added before lysis. Lysis was performed by using the *Fisher Scientific Sonic Dismembrator (Model 500)* to sonicate the cells at 30% amplitude three times for 5 seconds, with 10 second breaks in between. This was then followed by homogenization using the *Avestin EmulsiFlex-C3* high pressure homogenizer. The cell lysate was then spun down using the JA-17 rotor at 14500rpm for 30 minutes at 4°C. Note that the cells/cell lysate were kept on ice whenever possible during this whole process to prevent degradation. Co-lysis was performed by combining individual cell pellets from 2 L cultures of different proteins together before the sonication step.

Nickel Affinity Chromatography

A *Bio-Rad Econo Column* was filled with 5 mL of *Ni-NTA* resin beads (*QIAGEN*). Chromatography was performed using buffers containing various concentrations of imidazole dissolved in 20 mM Tris-HCl (pH 8.0) and 100 mM NaCl. The supernatant from the clarified cell lysate was loaded on the nickel resin that was pre-equilibrated with 35 mM imidazole. After collecting the flow-through, the resin was washed using 25 mL of 5 mM imidazole and 25 mL of 35 mM imidazole with both wash fractions being collected. The protein of interest was then eluted using 5 mL fractions of buffers containing 100 - 500 mM imidazole. A 10 µL aliquot of the input sample, flow-through, wash and elution fractions were taken for SDS-PAGE analysis.

Amylose Affinity Chromatography

As explained in Section 3.4.2, when the MBP-fusion proteins (MBP-$BamC_U$ and MBP-PhoE) were co-purified with hexahistidine tagged BamD through nickel affinity chromatography, no co-elution was observed. Thus, amylose affinity chromatography was required to purify the MBP-fusion proteins from the cell lysate. A *Bio-Rad Econo Column* was filled with 3 mL of amylose resin (*New England BioLabs*). The resin was equilibrated with buffer containing 20 mM Tris-HCl (pH 8.0) and 100 mM NaCl prior to sample loading. After collecting the flow-through, the resin was washed with 50 mL of the same buffer used for equilibration. The MBP-fusion protein was then eluted with 20 mL of buffer containing 20 mM Tris-HCl (pH 8.0), 100 mM NaCl, and 10 mM maltose, which was collected in 5 mL fractions. A 10 µL aliquot of the input sample, flow-through, wash and elution fractions were taken for SDS-PAGE analysis.

SDS-PAGE

15% resolving SDS-PAGE gels were used, and were run at 180 V for about 1 hour (or until dye front reached the end of the gel). 10 µL of each sample was mixed with 10 µL of 2x loading buffer. The samples were boiled for five minutes before being loaded on the gel. 10 µL of each

sample and 5 µL of *Fermentas Unstained Protein Molecular Weight Marker* were loaded. After electrophoresis, the gel was stained using *Fermentas Page Blue* stain.

Concentration of Protein Samples

Before loading samples on the gel-filtration chromatography column, the affinity chromatography fractions containing the desired protein(s) were concentrated down to ~5 mL. Before protein crystallization, purified samples were concentrated down to ~1 mL or until the desired concentration was reached. In both cases, the protein was concentrated using a *Millipore Amicon Ultra* centrifugal filter with the corresponding molecular weight cut off. These filters were centrifuged at 4500rpm until the desired volume was reached.

Gel-filtration Chromatography

Gel filtration chromatography was carried out using the *HiPrep 26/60 Sephacryl S-100 High Resolution Column* attached to the *ÄKTA Prime system* (*GE Healthcare*). ~5 mL samples were loaded onto the column, and had concentrations in the range of ~5-10 mg/mL. The buffer used contained 20 mM Tris-HCl (pH 8.0) and 100 mM NaCl. The resulting chromatogram was compared to a standard curve (Figure A1) for the column. The protein identity from the fractions was then determined by loading the fraction samples on an SDS-PAGE gel.

To create a standard curve, protein standards from the *LMW Calibration Kit* (*Amersham Biosciences*) were used which included ribonuclease A, chymotrypsinogen A, ovalbumin, albumin, and Blue Dextran 2000. 5 mg/mL of each standard was run, except for Blue Dextran which was loaded at a concentration of 1 mg/mL. Three runs were performed: ribonuclease A, ovalbumin, and albumin in the first run; Blue Dextran in the second run; and chymotrypsinogen A in the third run. Chymotrypsinogen A was run separately to prevent digestion of the other standards. A table of the standards, their properties, and observed elution volumes is provided below:

Table A2 Protein standards used to calibrate the gel-filtration column

Standard	Molecular Mass (kDa)	Stokes Radius (Å)	Elution Volume on Sephacryl S-100 (mL)
Ribonuclease A	13.7	16.4	189
Chymotrypsinogen A	25.0	20.9	165
Ovalbumin	43.0	30.5	133
Albumin	67.0	35.5	118
Blue Dextran 2000	2000		98*

The molecular mass and Stokes radius values are presented as provided by the *LMW Calibration Kit*. *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 column can retain.

The column used has a total column volume of 320 mL, and based on the elution volume of Blue Dextran, the void volume was determined to be 98 mL. Using the elution volumes of the rest of the standards, the partition coefficient, K_{av}, was calculated as follows:

$$K_{av} = \frac{V_e - V_o}{V_t - V_o} \quad \textbf{(Equation 1)}$$

where V_e is the elution volume, V_o is the void volume (98 mL), and V_t is the column volume (320 mL).

The standard curve was plotted as a K_{av} vs. *Log(M)* graph, where *M* indicates molecular mass. The equation of the best fit line was determined to be: $\boldsymbol{Log(M) = -2.06(K_{av}) + 1.99}$ (Figure A1). This equation was rewritten as:

$$M = 10^{(-2.06 \times K_{av} + 1.99)} \quad \textbf{(Equation 2)}$$

Then for each gel-filtration chromatography run, the elution volume of each protein was used to calculate its K_{av} value using Equation 1 above. The molecular mass (*M*) was calculated using Equation 2. Throughout the thesis, this mass derived from the standard curve is referred to as the *measured molecular mass*. An example chromatogram visualizing the different volume values required for these calculations is shown in Figure A2.

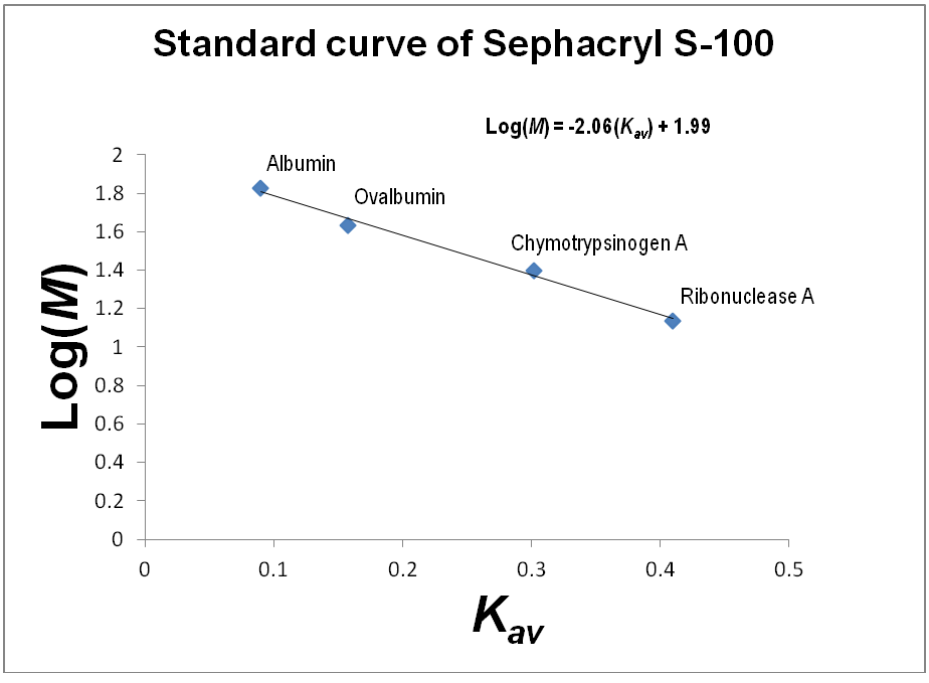

Figure A1 Standard curve of the Sephacryl S-100 gel-filtration column

This is the standard curve used to estimate the molecular masses of eluted complexes.

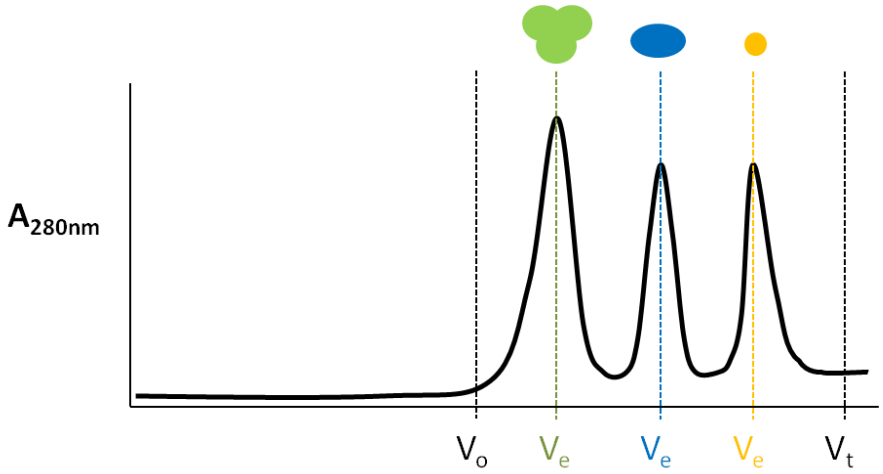

Figure A2 Example gel-filtration chromatogram

This figure shows what a chromatogram would look like, and how three proteins with different sizes and shapes would expect to elute. Proteins that can be resolved by a particular gel-filtration column have elution volumes (V_e) that fall between the range of the void volume (V_o) and the total column volume (V_t). Large proteins or complexes (green) would expect to elute earlier, while smaller proteins (yellow) would elute last.

Analysis of Gel-filtration Chromatograms

The chromatograms were analyzed using the *PrimeView V1.00 Evaluation* software (*Amersham Pharmacia Biotech*), where the elution volumes were determined by looking at the volume which provided the highest point of each peak.

Each chromatogram figure was created using *Microsoft Excel*. The chromatogram curves were first exported from *PrimeView* using File → Export → Curves. Then the desired curves were selected, and "Normalise Retention" was checked if exporting multiple curves together. These curves were exported and saved as an ASCII file (with an *.asc* file extension). This simple text file is then able to be opened in *Excel*. In *Excel*, the individual chromatograms where presented as two columns corresponding to the elution volume (mL) and the absorbance (mAu). To allow comparison between chromatograms, each individual chromatogram was normalized as follows:

1) The minimum absorbance value was determined ("*min*")
2) The *min* value was subtracted from each absorbance value ("*new absorbance value*") so that no value would be less than 0.
3) The maximum value was determined from the *new absorbance values* ("*new max*")
4) The *new absorbance value* was divided by the *new max* ("*final value*") so that no value would be greater than 1.
5) An XY scatter plot was constructed using the original elution volume values as the X-axis and the new *final values* as the absorbance values for the Y-axis.
6) To overlap chromatograms, the X- and Y-axis values from another normalized chromatogram were added as a new series under the "Select Data" option.

To compare the amount of monomer and homodimer populations present of BamE and its truncations in Section 2.2.3, the "Peak Integrate" function was used to determine the area underneath each peak. Then the percentage of each peak's area from the total area (sum of both peak areas) was calculated and reported.

Limited Proteolysis of BamC

A purified BamC sample with a concentration of 1 mg/mL was mixed with chymotrypsin at a ratio of 1000:1 protein to protease. The sample was left at room temperature, with 10 µL aliquots removed at 0, 5, 10, 15, 30, 45, 60, and 120 minutes, plus an aliquot taken after overnight digestion. After each aliquot was removed, it was mixed with 10 µL of 2X SDS loading buffer and then boiled for 3 minutes to stop the reaction. The proteolysis pattern was then observed on an SDS-PAGE gel.

Thrombin Digest of BamB

To confirm if a $BamA_{POTRA}$-BamB dimer can form, the hexahistidine tag was removed from BamB using 1 unit of thrombin (*GE Healthcare*) per 1 mg of protein. Digestion was allowed to occur overnight at room temperature. Removal of the tag was confirmed by running the digested sample through nickel affinity chromatography, from which the flow through fraction was collected and used for a second nickel affinity chromatography experiment with tagged $BamA_{POTRA}$.

Protein Crystallization

A sample containing purified $BamC_{UN}$-BamD-$BamE_{\Delta C}$ (with hexahistidine tags intact) at 20 mg/mL in buffer containing 20 mM Tris-HCl (pH 8.0) and 100 mM NaCl with 0.03% v/v *n*-dodecyl β-D-maltoside (DDM) added after concentration, was set up for crystallization using the sitting drop vapour diffusion method. Crystallization conditions tested were derived from the following commercially available buffer kits from *Hampton Research*: Crystal Screen I, Crystal Screen II, and PEG Ion. 100 µL of these buffers were aliquoted into individual wells of a 96-well sitting drop plate (*Axygen*). In the pedestal above each well, 1 µL of the protein sample was mixed with 1 µL of the corresponding reservoir solution. All experiments were conducted at room temperature.

Appendix B: Theory of Techniques

Affinity Chromatography

Affinity chromatography allows the purification of proteins using an affinity tag. Recombinant proteins are constructed where an affinity tag is fused to its N- or C-terminus the protein of interest. Two commonly used tags are the histidine and maltose-binding protein (MBP) tags (Terpe, 2003).

Nickel affinity chromatography allows easy purification of proteins with a histidine tag. Usually, a hexahistidine tag is used which consists of six consecutive histidine residues that will bind with high affinity to a nickel based resin (Hochuli et al., 1987; Crowe et al., 1995). This allows only the protein of interest to be retained on the resin, while other proteins will elute in the flow through. To elute the protein of interest, the nickel-hexahistidine interaction can be disrupted in two ways: by decreasing the pH or by introducing imidazole. The histidine residues contain an imidazole ring which binds to the nickel. This ring has a pKa value of ~6.0, indicating that at a lower pH, it will become protonated and not bind as strongly to the nickel resin. The second approach of increasing the imidazole concentration creates competition between the imidazole salt and the imidazole of the hexahistidine tag for binding to the nickel resin. At around 100 mM imidazole, the hexahistidine tagged proteins will begin to elute, and elution is usually carried out to ~500 mM imidazole to ensure complete collection of the protein (Crowe et al., 1995; Terpe, 2003).

Amylose affinity chromatography allows easy purification of proteins with the MBP tag. The MBP tag is a popular choice when the protein of interest has problem expressing or is insoluble, as MBP has been shown to greatly increase the solubility of its passenger proteins (Kapust and Waugh, 1999). This system exploits MBP's high affinity for maltose, as well as its ability to bind to amylose. The amylose resin allows only proteins with the MBP tag to be retained, while other proteins will elute in the flow through. Then to elute the MBP-fusion protein, usually 10mM maltose is introduced which MBP will preferentially bind to, causing it to dissociate from the amylose resin (Riggs, 2000; Terpe, 2003).

Gel-filtration Chromatography

Gel-filtration chromatography (also referred to as size-exclusion chromatography) allows purification of proteins, and other molecules, based on their size. The technique emerged in the 1950s where a gel based stationary phase of a chromatography system was shown to separate molecules over a range of sizes (Porath and Flodin, 1959). Over the years, different types of gel matrices have been developed for separation such as cross-linked dextran (commonly known as *Sephadex*), agarose (*Sepharose*), a cross-linked dextran-bisacrylamide matrix (*Sephacryl*), and a cross-linked dextran-agarose matrix (*Superdex*) (Porath and Flodin, 1959; Andrews, 1964; Amersham Biosciences, 2002; Winzor, 2011). Large molecules that cannot enter far into the pores of the gel matrix elute early, while smaller proteins are retained in the matrix for a longer time, and elute later. Although gel-filtration was recognized to separate the molecules base on size, it was not certain what parameter of size. Initially it was proposed that the system separates based on molecular mass, but this was later shown to be limited to globular proteins, and thus Stokes radius was proposed as the correct parameter to be considered (Andrews, 1964; Laurent and Killander, 1964; Siegel and Monty, 1966). This also led to the development of the partition coefficient, K_{av}, to describe how a specific molecule travels in a specific gel-filtration system based on its elution volume (V_e), the total column volume (V_t) and the volume of the mobile phase (also known as the void volume, V_o) (Laurent and Killander, 1964). A logarithmic relationship was shown to be made between the K_{av} value and the Stokes radius (or molecular mass for globular proteins) for plotting a standard curve using known proteins. Then from this standard curve, the

Stokes radius, or molecular mass, of unknown proteins could be determined (Laurent and Killander, 1964; Siegel and Monty, 1966; Amersham Biosciences, 2002; Winzor, 2011). However, as described in Section 6.7, determining the Stokes radius is not a simple calculation that can be derived from gel-filtration alone, and often other techniques such as dynamic light scattering need to be used to accurately determine this value. Overall, it has been concluded that while Stokes radius is the correct parameter to be used, gel-filtration chromatography is an acceptable technique for estimation of molecular mass (Andrews, 1964; Amersham Biosciences, 2002; Winzor, 2011). Thus, in the data presented in this thesis, molecular mass is used as the parameter for identifying possible complexes formed between the BAM complex proteins.

Protein Crystallization

One of the most widely used methods to determine the 3D structure of proteins is X-ray crystallography. This technique requires the protein of interest to be prepared as a crystal for studies using X-ray diffraction. The manner by which the X-ray is diffracted by the crystal is used to determine the three dimensional coordinates of each atom in the molecule (Chayen and Saridakis, 2008; International Union of Crystallography, 2012). However, before the X-ray diffraction experiment can be conducted, the protein must be purified and crystallized, the latter often being the most difficult step. In the work presented in this thesis, complexes that formed were subjected to initial screening for crystallization purposes. The main crystallization experiment is outlined in Chapter 4, where the $BamC_{UN}DE_{\Delta C}$ complex was able to crystallize in an initial set of experiments. Crystallization is often difficult because there are multiple factors to consider when setting up the experiment such as the protein concentration, buffer composition, pH, temperature, as well as the crystal plating method used (Chayen and Saridakis, 2008; Hampton Research, 2010). Another variable to consider is the protein itself – the protein can be modified to improve crystallizability such as removing flexible regions, mutagenesis of highly entropic residues, and construction of fusion proteins, in hopes producing a protein that will form a tight crystal lattice (Dale et al., 2003; Cooper et al., 2007; International Union of Crystallography, 2012). In this project, removing possible flexible regions, and specifically removing regions not necessary for complex formation, successfully yielded in micro-crystals that can be further optimized in hopes of growing larger diffraction quality crystals.

Appendix C: Additional Figures

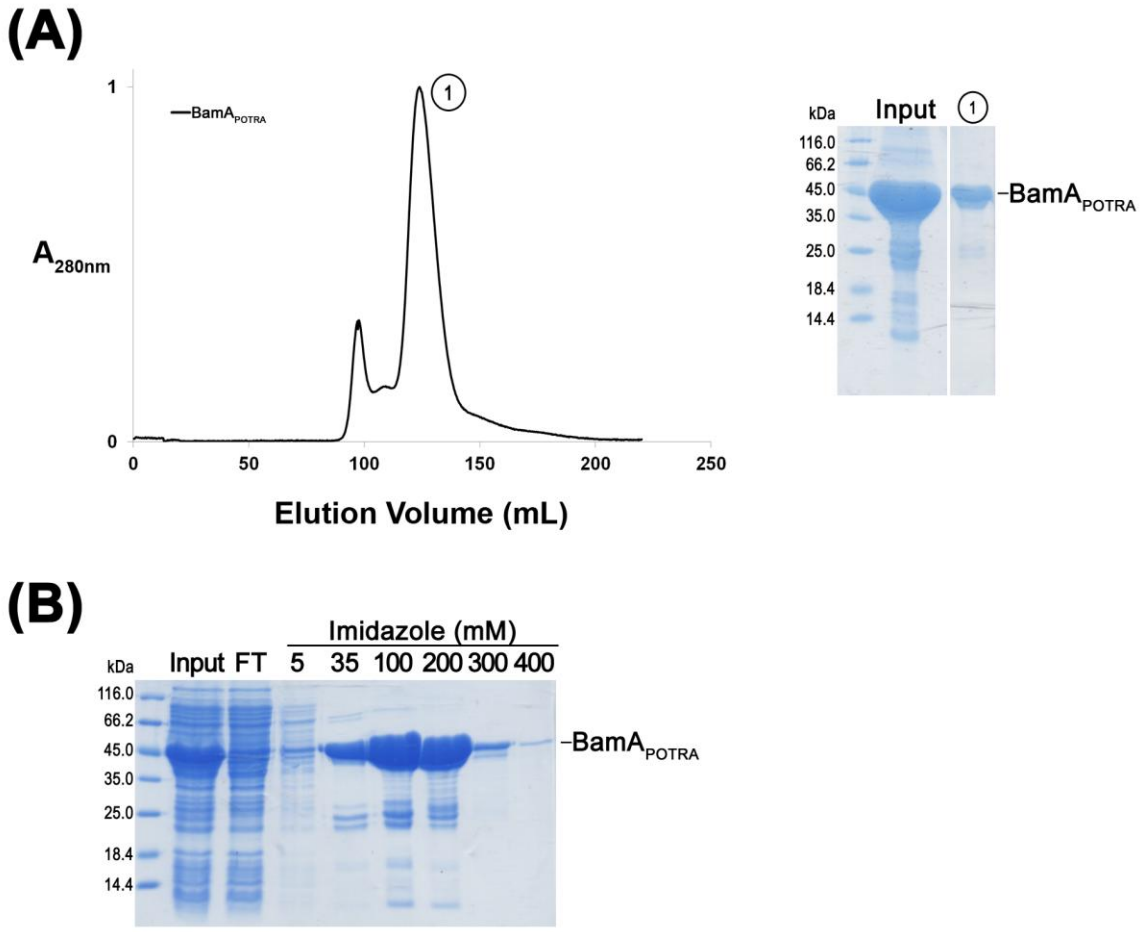

Figure C1 $BamA_{POTRA}$

(A) shows the individual chromatogram and corresponding SDS-PAGE gel of $BamA_{POTRA}$.
(B) presents the SDS-PAGE gel after nickel affinity chromatography of $BamA_{POTRA}$, and shows that this protein is susceptible to degradation.

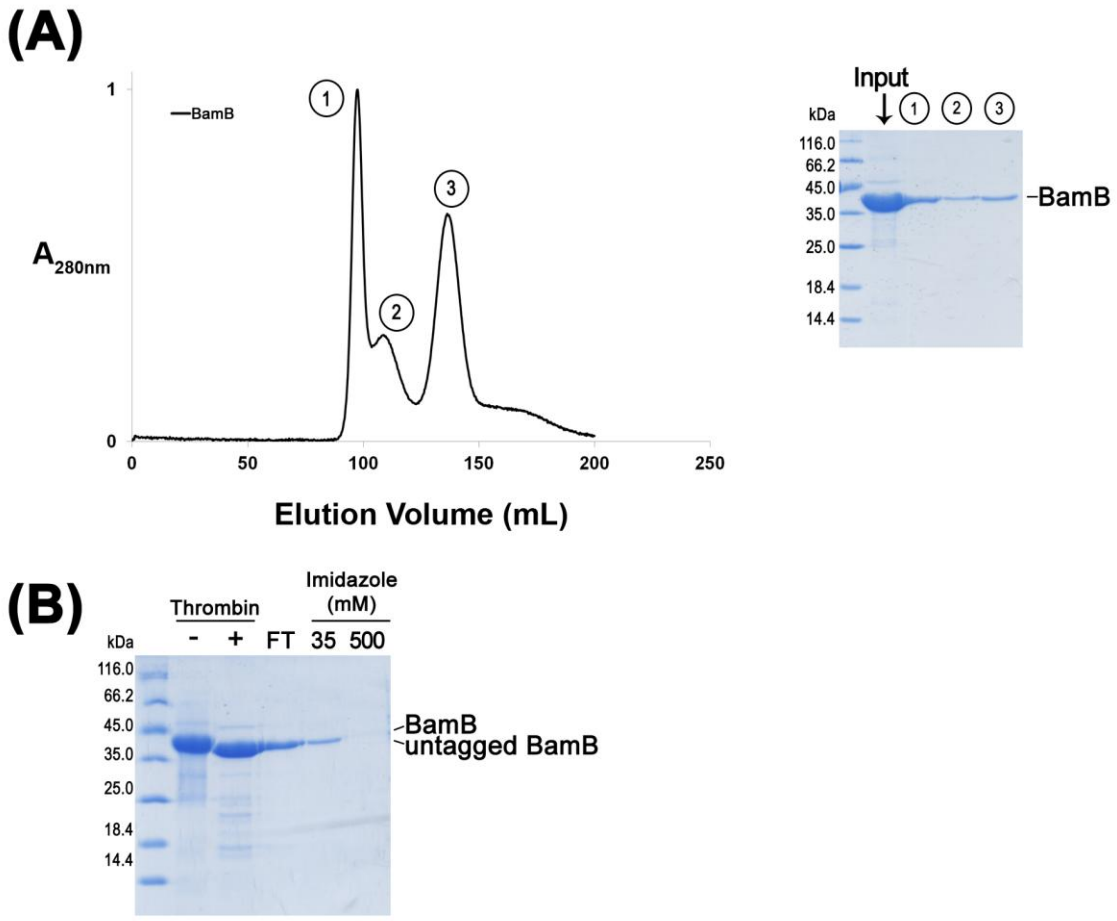

Figure C2 BamB

(A) shows the individual chromatogram and corresponding SDS-PAGE gel of BamB. (B) presents the SDS-PAGE gel that confirms that the hexahistidine tag was successfully removed for the experiment in Figure 5.1B. After thrombin digestion to remove the tag, the protein was run through a nickel affinity chromatography column, where the majority eluted in the flow through (FT). That fraction was then carried forward for the experiment.

Figure C3 Full PSIPRED analysis of BamC

This figure provides the complete results from the PSIPRED secondary structure prediction program, as described in Section 3.1.1.

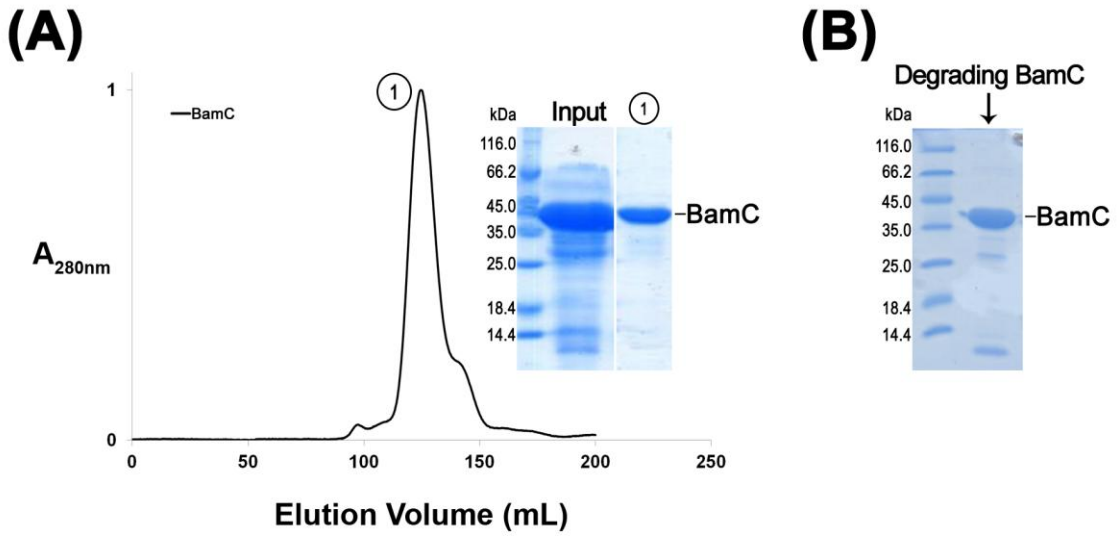

Figure C4 BamC

(A) provides the gel-filtration chromatography results of individual BamC. (B) shows how BamC begins to degrade during storage at 4°C within a few days after purification.

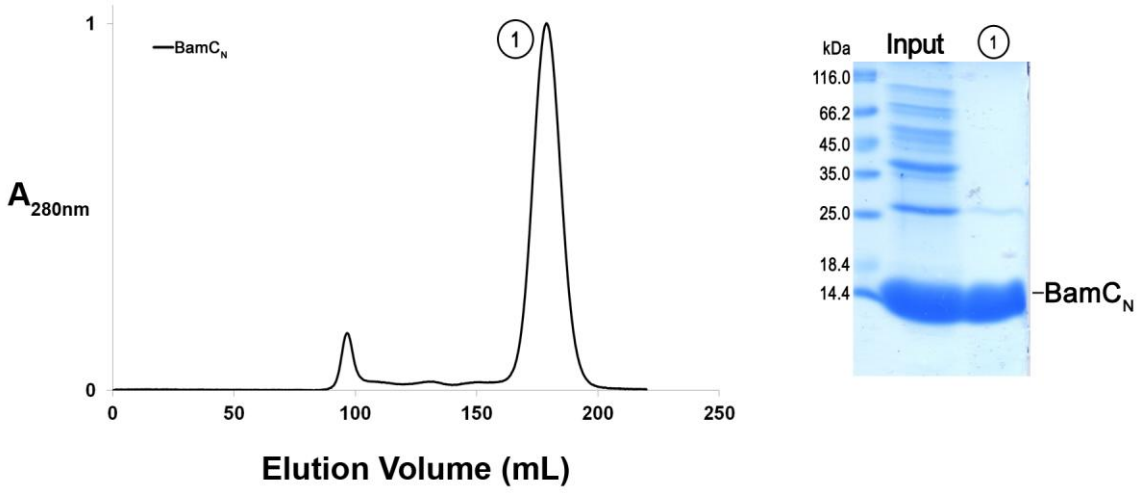

Figure C5 $BamC_N$

This figure provides the gel-filtration chromatography results of individual $BamC_N$.

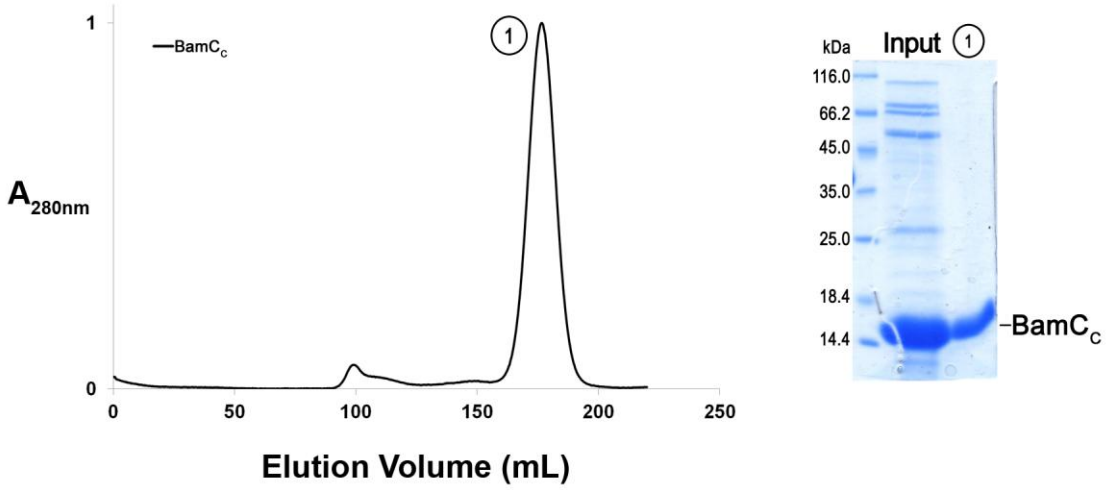

Figure C6 $BamC_C$

This figure provides the gel-filtration chromatography results of individual $BamC_C$.

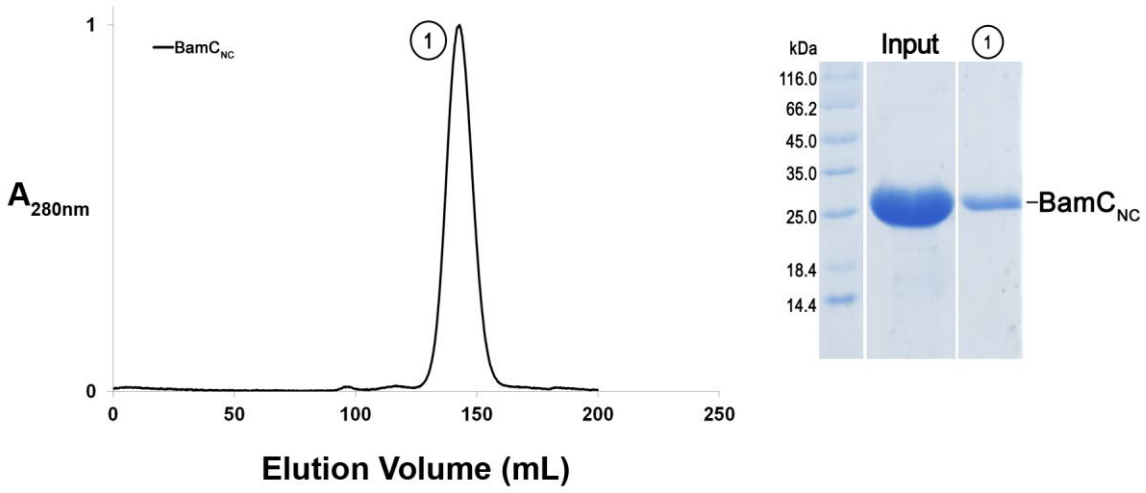

Figure C7 BamC$_{NC}$

This figure provides the gel-filtration chromatography results of individual BamC$_{NC}$.

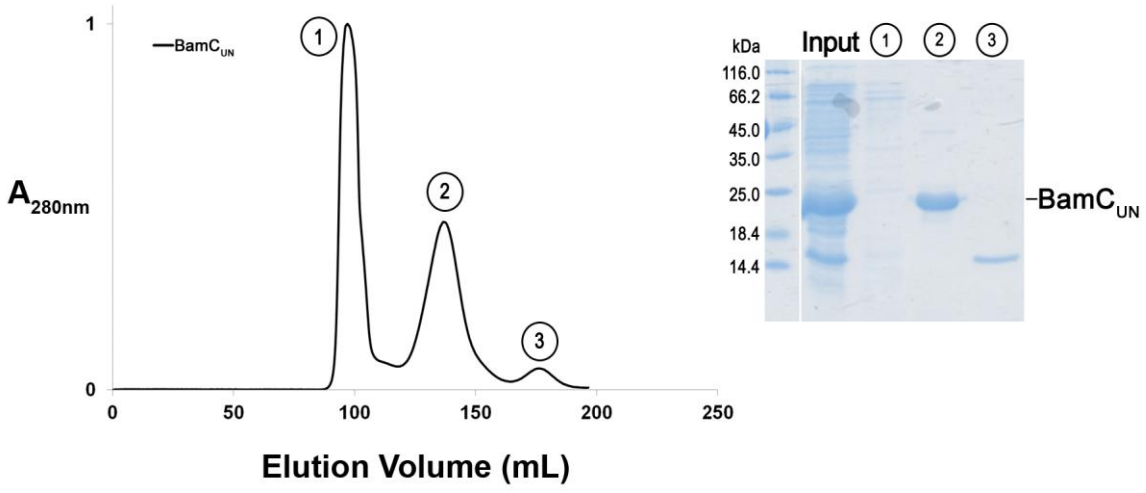

Figure C8 BamC$_{UN}$

This figure provides the gel-filtration chromatography results of individual BamC$_{UN}$. Note, despite a large peak (peak 1) at the void volume, BamC$_{UN}$ elutes in peak 2. Peak 1 corresponds to aggregated proteins and the contaminating proteins with high molecular mass that were present in the input sample. A possible degradation product (most likely BamC$_{N}$) is seen in peak 3.

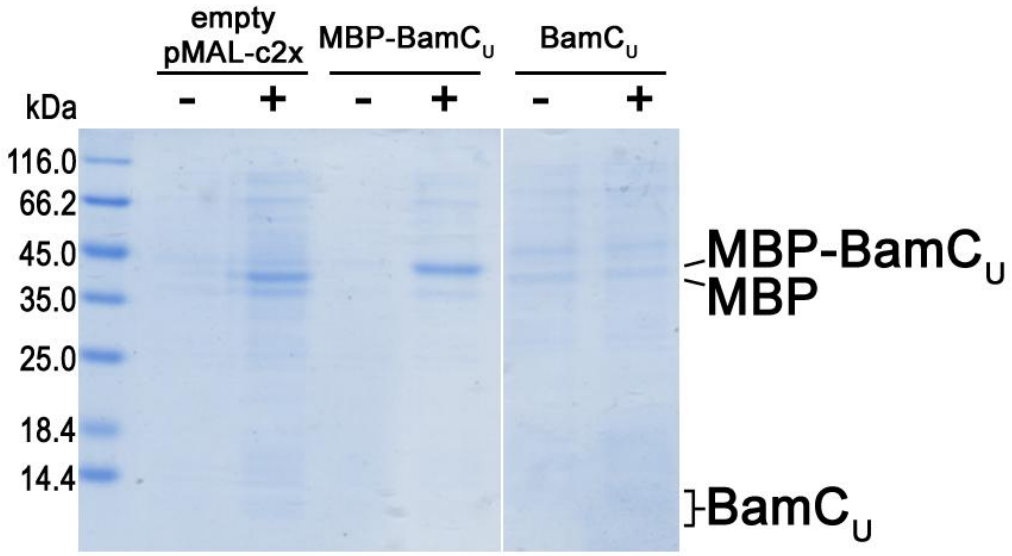

Figure C9 Expression of BamC$_U$ and MBP-BamC$_U$

This figure shows how BamC$_U$ alone was not able to express, but MBP-BamC$_U$ was. For comparison, empty pMAL-c2x vector was expressed, which contains the MBP tag. Minus sign (-) indicates before inducing expression with IPTG, plus sign (+) indicates 3 hours after induction.

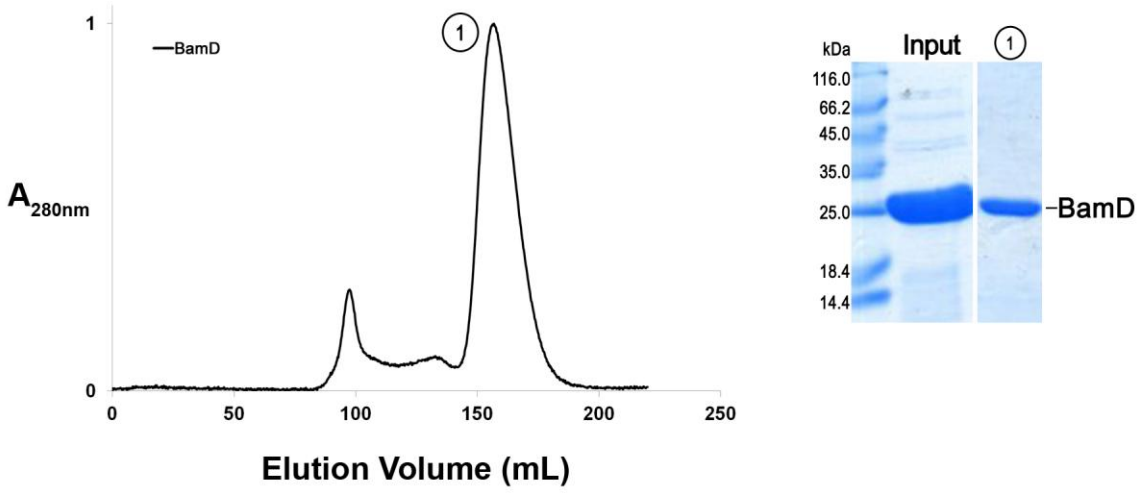

Figure C10 BamD

This figure provides the gel-filtration chromatography results of individual BamD.

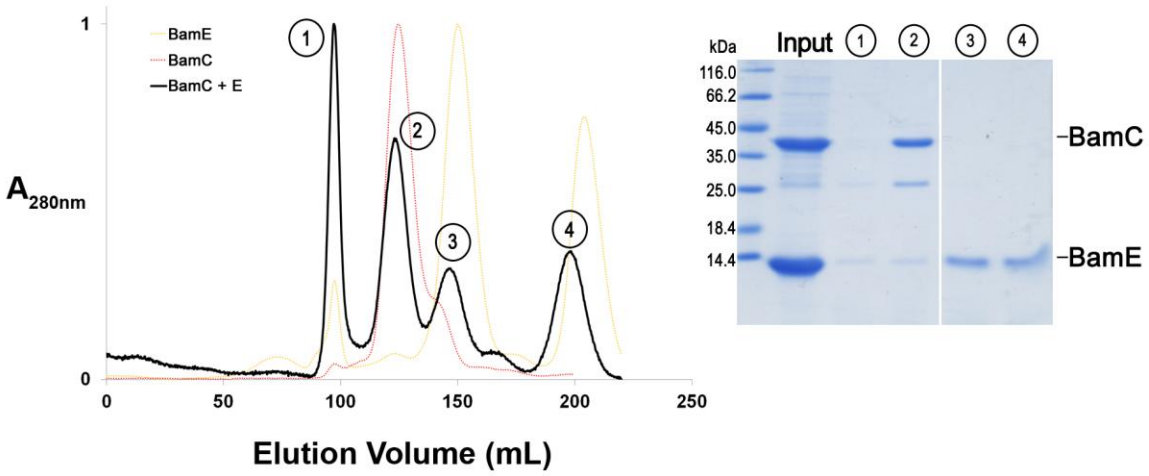

Figure C11 BamC and BamE do not form a complex

This figure provides the gel-filtration chromatography results of a sample containing BamC and BamE. Based on the elution profile and the corresponding SDS-PAGE gel, it can be concluded that the two proteins do not form a BamCE dimer.

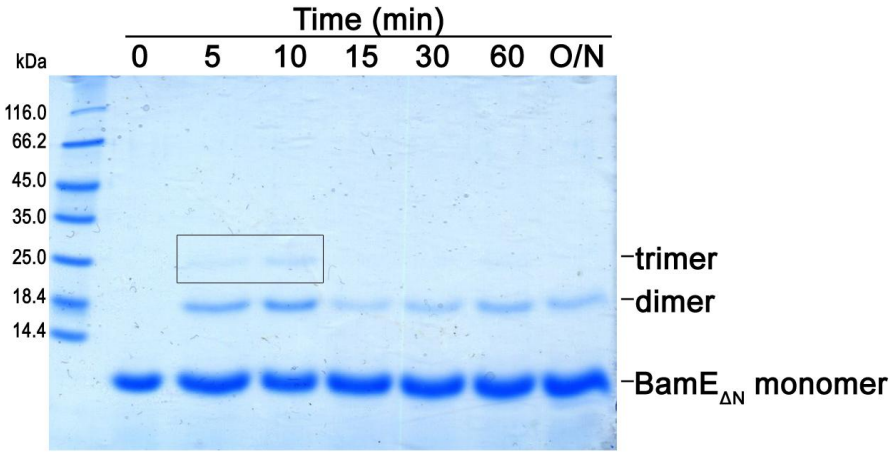

Figure C12 Cross-linking of the $BamE_{\Delta N}$ oligomer

This figure provides the results of a cross-linking experiment set up using 0.9 mg/mL of purified $BamE_{\Delta N}$ eluting in peak 1 of Figure 2.13, and 2% formaldehyde. The reaction was left at room temperature. At each time point, 10 µL aliquots were taken, mixed with 10 µL of 2X SDS loading buffer and boiled for 5 minutes. From here it appears that this population of $BamE_{\Delta N}$ is a dimer and not a higher order oligomer. While there are faint bands indicating possible trimer formation, but as they only appear in two of the aliquots and are faint, it could just be a result of forced cross-linking rather than an actual trimer forming. Note: the term “monomer” on the gel refers to the fact that this is a denaturing gel, and the population of dimer not cross-linked was dissociated into its monomeric form.

References

Albrecht, R., and Zeth, K. (2010). Crystallization and preliminary X-ray data collection of the Escherichia coli lipoproteins BamC, BamD and BamE. Acta crystallographica *F66*, 1586–1590.

Albrecht, R., and Zeth, K. (2011). Structural basis of outer membrane protein biogenesis in bacteria. The Journal of biological chemistry *286*, 27792–27803.

Amersham Biosciences (2002). Gel Filtration: Principles and Methods (Uppsala, Sweden: Amersham Biosciences).

Amor-Mahjoub, M., Suppini, J.-P., Gomez-Vrielyunck, N., and Ladjimi, M. (2006). The effect of the hexahistidine-tag in the oligomerization of HSC70 constructs. Journal of chromatography B *844*, 328–334.

Andrews, P. (1964). Estimation of the molecular weights of proteins by Sephadex gel-filtration. The Biochemical journal *91*, 222–233.

Bechtluft, P., Nouwen, N., Tans, S. J., and Driessen, A. J. M. (2010). SecB--a chaperone dedicated to protein translocation. Molecular BioSystems *6*, 620–627.

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

Bernstein, H. D. (2007). Are bacterial “autotransporters” really transporters? Trends in microbiology *15*, 441–447.

Bishop, R. E. (2008). Structural biology of membrane-intrinsic beta-barrel enzymes: sentinels of the bacterial outer membrane. Biochimica et biophysica acta *1778*, 1881–1896.

Bishop, R. E. (2005). The lipid A palmitoyltransferase PagP: molecular mechanisms and role in bacterial pathogenesis. Molecular microbiology *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. The Journal of biological chemistry *278*, 49316–49322.

Bodelón, G., Marín, E., and Fernández, L. A. (2009). Role of periplasmic chaperones and BamA (YaeT/Omp85) in folding and secretion of intimin from enteropathogenic Escherichia coli strains. Journal of bacteriology *191*, 5169–5179.

Bos, M. P., and Tommassen, J. (2004). Biogenesis of the Gram-negative bacterial outer membrane. Current opinion in microbiology *7*, 610–616.

Buchan, D. W. A., Ward, S. M., Lobley, A. E., Nugent, T. C. O., Bryson, K., and Jones, D. T. (2010). Protein annotation and modelling servers at University College London. Access *38*, 563–568.

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

Chandran, V., Fronzes, R., Duquerroy, S., Cronin, N., Navaza, J., and Waksman, G. (2009). Structure of the outer membrane complex of a type IV secretion system. Nature *462*, 1011–1015.

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

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

Clantin, B., Delattre, A.-S., Rucktooa, P., Saint, N., Méli, 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 in microbiology *15*, 96–100.

Cooper, D. R., Boczek, T., Grelewska, K., Pinkowska, M., Sikorska, M., Zawadzki, M., and Derewenda, Z. (2007). Protein crystallization by surface entropy reduction: optimization of the SER strategy. Acta crystallographica *D63*, 636–645.

Cross, B. C. S., Sinning, I., Luirink, J., and High, S. (2009). Delivering proteins for export from the cytosol. Nature reviews. Molecular cell biology *10*, 255–264.

Crowe, J., Masone, B. S., and Ribbe, J. (1995). One-step purification of recombinant proteins with the 6xHis tag and Ni-NTA resin. Molecular biotechnology *4*, 247–258.

Dalbey, R. E., Wang, P., and Kuhn, A. (2011). Assembly of bacterial inner membrane proteins. Annual review of biochemistry *80*, 161–187.

Dale, G. E., Oefner, C., and D'Arcy, A. (2003). The protein as a variable in protein crystallization. Journal of Structural Biology *142*, 88–97.

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. The Journal of biological chemistry *272*, 3179–3184.

Delcour, A. H. (2009). Outer membrane permeability and antibiotic resistance. Biochimica et biophysica acta *1794*, 808–816.

Delcour, A. H. (2003). Solute uptake through general porins. Frontiers in bioscience *8*, d1055–71.

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.

Driessen, A. J. M., and Nouwen, N. (2008). Protein translocation across the bacterial cytoplasmic membrane. Annual review of biochemistry *77*, 643–667.

Fairman, J. W., Noinaj, N., and Buchanan, S. K. (2011). The structural biology of β-barrel membrane proteins: a summary of recent reports. Current opinion in structural biology *21*, 1–9.

Ferbitz, L., Maier, T., Patzelt, H., Bukau, B., Deuerling, E., and Ban, N. (2004). Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins. Nature *431*, 590–596.

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. The EMBO journal *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. The Journal of cell biology *164*, 19–24.

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

Hagan, C. L., Kim, S., and Kahne, D. (2010). Reconstitution of outer membrane protein assembly from purified components. Science *328*, 890–892.

Hagan, C. L., Silhavy, T. J., and Kahne, D. (2011). β-Barrel Membrane Protein Assembly by the Bam Complex. Annual review of biochemistry *80*, 189–210.

Hampton Research (2010). Crystallization: Research Tools Volume 18. (Hampton Research).

von Heijne, G. (1990). The signal peptide. The Journal of membrane biology *115*, 195–201.

Heuck, A., Schleiffer, A., and Clausen, T. (2011). Augmenting β-augmentation: structural basis of how BamB binds BamA and may support folding of outer membrane proteins. Journal of molecular biology *406*, 659–666.

Hochuli, E., Döbeli, H., and Schacher, A. (1987). New metal chelate adsorbent selective for proteins and peptides containing neighbouring histidine residues. Journal of chromatography *411*, 177–184.

Hoffmann, A., Bukau, B., and Kramer, G. (2010). Structure and function of the molecular chaperone Trigger Factor. Biochimica et biophysica acta *1803*, 650–661.

International Union of Crystallography (2012). International Tables for Crystallography Volume F E. Arnold, D. M. Himmel, and M. G. Rossmann, eds. (Chester, England: International Union of Crystallography).

Jain, S., and Goldberg, M. B. (2007). Requirement for YaeT in the outer membrane assembly of autotransporter proteins. Journal of bacteriology *189*, 5393–5398.

Jones, D. T. (1999). Protein secondary structure prediction based on position-specific scoring matrices. Journal of molecular biology *292*, 195–202.

Kapust, R. B., and Waugh, D. S. (1999). Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused. Protein science : a publication of the Protein Society *8*, 1668–1674.

Kim, K. H., Aulakh, S., and Paetzel, M. (2011a). Crystal Structure of β-Barrel Assembly Machinery BamCD Protein Complex. The Journal of biological chemistry *286*, 39116–39121.

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

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 crystallographica *F67*, 1350–1358.

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

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

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.

Knoblauch, N. T., Rüdiger, S., Schönfeld, H. J., Driessen, A. J., Schneider-Mergener, J., and Bukau, B. (1999). Substrate specificity of the SecB chaperone. The Journal of biological chemistry *274*, 34219–34225.

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 β-barrel assembly machine. EMBO reports *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. Molecular microbiology *68*, 1216–1227.

Knowles, T. J., McClelland, D. M., Rajesh, S., Henderson, I. R., and Overduin, M. (2009). 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. Biomolecular NMR assignments *3*, 203–206.

Koebnik, R., Locher, K. P., and Van Gelder, P. (2000). Structure and function of bacterial outer membrane proteins: barrels in a nutshell. Molecular microbiology *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.

Kozjak, V., Wiedemann, N., Milenkovic, D., Lohaus, C., Meyer, H. E., Guiard, B., Meisinger, C., and Pfanner, N. (2003). An essential role of Sam50 in the protein sorting and assembly machinery of the mitochondrial outer membrane. The Journal of biological chemistry *278*, 48520–48523.

Krewulak, K. D., and Vogel, H. J. (2011). TonB or not TonB: is that the question? Biochemistry and cell biology *89*, 87–97.

Kusters, I., and Driessen, A. J. M. (2011). SecA, a remarkable nanomachine. Cellular and molecular life sciences *68*, 2053–2066.

Laurent, T. C., and Killander, J. (1964). A theory of gel filtration and its experimental verification. Journal of chromatography A *14*, 317–330.

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. Molecular microbiology *61*, 151–164.

Merdanovic, M., Clausen, T., Kaiser, M., Huber, R., and Ehrmann, M. (2011). Protein quality control in the bacterial periplasm. Annual review of microbiology *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. The Journal of biological chemistry *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. The Journal of biological chemistry *271*, 6801–6809.

Nakae, T. (1976). Identification of the outer membrane protein of E. coli that produces transmembrane channels in reconstituted vesicle membranes. Biochemical and biophysical research communications *71*, 877–884.

Nikaido, H. (2003). Molecular basis of bacterial outer membrane permeability revisited. Microbiology and molecular biology reviews *67*, 593–656.

Nikaido, H. (1994). Porins and specific diffusion channels in bacterial outer membranes. The Journal of biological chemistry *269*, 3905–3908.

Noinaj, N., Guillier, M., Barnard, T. J., and Buchanan, S. K. (2010). TonB-dependent transporters: regulation, structure, and function. Annual review of microbiology *64*, 43–60.

Okuda, S., and Tokuda, H. (2011). Lipoprotein Sorting in Bacteria. Annual review of microbiology.

Oomen, C. J., van Ulsen, P., van Gelder, P., Feijen, M., Tommassen, J., and Gros, P. (2004). Structure of the translocator domain of a bacterial autotransporter. The EMBO journal *23*, 1257–1266.

Oreb, M., Tews, I., and Schleiff, E. (2008). Policing Tic "n" Toc, the doorway to chloroplasts. Trends in cell biology *18*, 19–27.

Paetzel, M., Karla, A., Strynadka, N. C. J., and Dalbey, R. E. (2002). Signal peptidases. Chemical reviews *102*, 4549–4580.

Palomino, C., Marín, E., and Fernández, L. Á. (2011). The fimbrial usher FimD follows the SurA-BamB pathway for its assembly in the outer membrane of Escherichia coli. Journal of bacteriology *193*, 5222–5230.

Paschen, S. A., Neupert, W., and Rapaport, D. (2005). Biogenesis of beta-barrel membrane proteins of mitochondria. Trends in biochemical sciences *30*, 575–582.

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.

du Plessis, D. J. F., Nouwen, N., and Driessen, A. J. M. (2011). The Sec translocase. Biochimica et biophysica acta *1808*, 851–865.

Pohlschröder, M., Prinz, W. A., Hartmann, E., and Beckwith, J. (1997). Protein translocation in the three domains of life: variations on a theme. Cell *91*, 563–566.

Porath, J., and Flodin, P. (1959). Gel filtration: a method for desalting and group separation. Nature *183*, 1657–1659.

Postle, K., and Kadner, R. J. (2003). Touch and go: tying TonB to transport. Molecular microbiology *49*, 869–882.

Qu, J., Mayer, C., Behrens, S., Holst, O., and Kleinschmidt, J. H. (2007). The trimeric periplasmic chaperone Skp of Escherichia coli forms 1:1 complexes with outer membrane proteins via hydrophobic and electrostatic interactions. Journal of molecular biology *374*, 91–105.

Raetz, C. R. H., and Whitfield, C. (2002). Lipopolysaccharide endotoxins. Annual review of biochemistry *71*, 635–700.

Reumann, S., Davila-Aponte, J., and Keegstra, K. (1999). The evolutionary origin of the protein-translocating channel of chloroplastic envelope membranes: identification of a cyanobacterial homolog. Proceedings of the National Academy of Sciences of the United States of America *96*, 784–789.

Ricci, D. P., and Silhavy, T. J. (2012). The Bam machine: A molecular cooper. Biochimica et biophysica acta *1818*, 1067–1084.

Riggs, P. (2000). Expression and purification of recombinant proteins by fusion to maltose-binding protein. Molecular biotechnology *15*, 51–63.

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. Journal of bacteriology *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 biology *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., et al. (2011). The Essential beta-Barrel Assembly Machinery Complex Components BamD and BamA Are Required for Autotransporter Biogenesis. Journal of bacteriology *193*, 4250–4253.

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 β-Barrel Assembly Machinery of Gram-Negative Bacteria. Journal of molecular biology *409*, 348–357.

Sauri, A., Soprova, Z., Wickström, D., de Gier, J.-W., Van der Schors, R. C., Smit, A. B., Jong, W. S. P., 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. Biological chemistry *392*, 21–27.

Schleiff, E., and Soll, J. (2005). Membrane protein insertion: mixing eukaryotic and prokaryotic concepts. EMBO reports *6*, 1023–1027.

Schulz, G. E. (2000). beta-Barrel membrane proteins. Current opinion in structural biology *10*, 443–447.

Selkrig, J., Mosbahi, K., Webb, C. T., Belousoff, M. J., Perry, A. J., Wells, T. J., Morris, F., Leyton, D. L., Totsika, M., Phan, M.-D., et al. (2012). Discovery of an archetypal protein transport system in bacterial outer membranes. Nature structural & molecular biology *19*, 506–510.

Siegel, L. M., and Monty, K. J. (1966). Determination of molecular weights and frictional ratios of proteins in impure systems by use of gel filtration and density gradient centrifugation. Application to crude preparations of sulfite and hydroxylamine reductases. Biochimica et biophysica acta *112*, 346–362.

Silhavy, T. J., Kahne, D., and Walker, S. (2010). The bacterial cell envelope. Cold Spring Harbor perspectives in biology *2*, a000414.

Sklar, J. G., Wu, T., Gronenberg, L. S., Malinverni, J. C., Kahne, D., and Silhavy, T. J. (2007). Lipoprotein SmpA is a component of the YaeT complex that assembles outer membrane proteins in Escherichia coli. Proceedings of the National Academy of Sciences of the United States of America *104*, 6400–6405.

Sommer, M. S., Daum, B., Gross, L. E., Weis, B. L. M., Mirus, O., Abram, L., Maier, U.-G., Kühlbrandt, W., and Schleiff, E. (2011). Chloroplast Omp85 proteins change orientation during evolution. Proceedings of the National Academy of Sciences of the United States of America *108*, 13841–13846.

Stroud, D. A., Becker, T., Qiu, J., Stojanovski, D., Pfannschmidt, S., Wirth, C., Hunte, C., Guiard, B., Meisinger, C., Pfanner, N., et al. (2011). Biogenesis of mitochondrial beta-barrel proteins: the POTRA domain is involved in precursor release from the SAM complex. Molecular biology of the cell *22*, 2823–2833.

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. Journal of bacteriology *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. Journal of bacteriology *170*, 5625–5632.

Terpe, K. (2003). Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. Applied microbiology and biotechnology *60*, 523–533.

Tommassen, J. (2007). Biochemistry. Getting into and through the outer membrane. Science *317*, 903–904.

Touzé, T., Hayward, R. D., Eswaran, J., Leong, J. M., and Koronakis, V. (2003). Self-association of EPEC intimin mediated by the β-barrel-containing anchor domain: a role in clustering of the Tir receptor. Molecular Microbiology *51*, 73–87.

van Ulsen, P. (2011). Protein folding in bacterial adhesion: secretion and folding of classical monomeric autotransporters. In Advances in experimental medicine and biology, D. Linke and A. Goldman, eds. (Dordrecht: Springer Netherlands), pp. 125–142.

Valent, Q. A., Kendall, D. A., High, S., Kusters, R., Oudega, B., and Luirink, J. (1995). Early events in preprotein recognition in E. coli: interaction of SRP and trigger factor with nascent polypeptides. The EMBO journal *14*, 5494–5505.

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. The EMBO journal *20*, 5033–5039.

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. Biochimica et biophysica acta *1778*, 1714–1734.

Vollmer, W., von Rechenberg, M., and Höltje, J. V. (1999). Demonstration of molecular interactions between the murein polymerase PBP1B, the lytic transglycosylase MltA, and the scaffolding protein MipA of Escherichia coli. The Journal of biological chemistry *274*, 6726–6734.

Volokhina, E. B., Beckers, F., Tommassen, J., and Bos, M. P. (2009). The beta-barrel outer membrane protein assembly complex of Neisseria meningitidis. Journal of bacteriology *191*, 7074–7085.

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.

Walther, D. M., Rapaport, D., and Tommassen, J. (2009). Biogenesis of beta-barrel membrane proteins in bacteria and eukaryotes: evolutionary conservation and divergence. Cellular and molecular life sciences *66*, 2789–2804.

Walton, T. a, and Sousa, M. C. (2004). Crystal structure of Skp, a prefoldin-like chaperone that protects soluble and membrane proteins from aggregation. Molecular cell *15*, 367–374.

Wimley, W. C. (2003). The versatile beta-barrel membrane protein. Current opinion in structural biology *13*, 404–411.

Winzor, D. J. (2011). Gel filtration: a means of estimating the molecular masses of proteins. Biochemical Journal *2011*, 1–3.

Wu, J., and Filutowicz, M. (1999). Hexahistidine (His6)-tag dependent protein dimerization: a cautionary tale. Acta biochimica Polonica *46*, 591–599.

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.

Yen, M. R., Peabody, C. R., Partovi, S. M., Zhai, Y., Tseng, Y. H., and Saier, M. H. (2002). Protein-translocating outer membrane porins of Gram-negative bacteria. Biochimica et biophysica acta *1562*, 6–31.

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. Frontiers in microbiology *2*, 189.

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.