

**Refolding Recombinant *Escherichia coli* BamA:
the integral membrane component of the BAM
(Beta-barrel Assembly Machinery) complex**

by

Minfei Fu

B.Sc., University of British Columbia, 2011

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

© Minfei Fu 2014

SIMON FRASER UNIVERSITY

Summer 2014

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: Minfei Fu
Degree: Master of Science
Title of Thesis: Refolding Recombinant *Escherichia coli* BamA: the integral membrane component of the BAM (Beta-barrel Assembly Machinery) complex
Examining Committee: Chair: Rosemary Cornell
Professor

Mark Paetzel
Senior Supervisor
Professor

Edgar C. Young
Supervisor
Associate Professor

Michel Leroux
Supervisor
Professor

Lisa Craig
Internal Examiner
Associate Professor
Simon Fraser University

Date Defended/Approved: July 23, 2014

Partial Copyright Licence



The author, whose copyright is declared on the title page of this work, has granted to Simon Fraser University the non-exclusive, royalty-free right to include a digital copy of this thesis, project or extended essay[s] and associated supplemental files ("Work") (title[s] below) in Summit, the Institutional Research Repository at SFU. SFU may also make copies of the Work for purposes of a scholarly or research nature; for users of the SFU Library; or in response to a request from another library, or educational institution, on SFU's own behalf or for one of its users. Distribution may be in any form.

The author has further agreed that SFU may keep more than one copy of the Work for purposes of back-up and security; and that SFU may, without changing the content, translate, if technically possible, the Work to any medium or format for the purpose of preserving the Work and facilitating the exercise of SFU's rights under this licence.

It is understood that copying, publication, or public performance of the Work for commercial purposes shall not be allowed without the author's written permission.

While granting the above uses to SFU, the author retains copyright ownership and moral rights in the Work, and may deal with the copyright in the Work in any way consistent with the terms of this licence, including the right to change the Work for subsequent purposes, including editing and publishing the Work in whole or in part, and licensing the content to other parties as the author may desire.

The author represents and warrants that he/she has the right to grant the rights contained in this licence and that the Work does not, to the best of the author's knowledge, infringe upon anyone's copyright. The author has obtained written copyright permission, where required, for the use of any third-party copyrighted material contained in the Work. The author represents and warrants that the Work is his/her own original work and that he/she has not previously assigned or relinquished the rights conferred in this licence.

Simon Fraser University Library
Burnaby, British Columbia, Canada

revised Fall 2013

Abstract

The selective permeability of the outer membrane (OM) of Gram-negative bacteria is mainly determined by the pore-forming proteins within it, which are called outer membrane proteins (OMPs). The proteinaceous apparatus required to fold and assemble the OMPs is known as the beta-barrel assembly machinery (BAM) complex. In *Escherichia coli*, the BAM complex is made up of five proteins. BamA is a beta-barrel protein integrated into the OM, while BamB, BamC, BamD and BamE are lipoproteins attached to the periplasmic side of the OM by a covalently bound lipid. Over-expression of BamA without its signal peptide results in insoluble inclusion body formation. I have investigated refolding strategies for *E. coli* BamA. The refolded BamA has been analysed by size-exclusive chromatography, heat-modifiability, circular dichroism spectroscopy and limited proteolysis. The experimental results are consistent with both a full length BamA and a truncated BamA construct being refolded properly using my discovered conditions. Both refolded constructs are able to crystallize. These initial crystallization conditions will be helpful in the structural analysis of *E. coli* BamA.

Keywords: outer membrane; beta-barrel outer membrane protein; beta-barrel assembly machinery complex; protein refolding

*To my parents, Jenson and all those friends who
help me through*

Acknowledgements

First and foremost, my utmost gratitude and sincere thanks to my senior supervisor, Dr. Mark Paetzel, whose guidance, and support from the beginning to the end of my Masters has been my inspiration as I hurdled over all the obstacles encountered while I conducted my research work.

I would like to thank to my supervisory committee members, Dr. Edgar Young and Dr. Michel Leroux, who contributed and extended their valuable assistance in the preparation and completion of this study. Moreover, much appreciation goes to Dr. Lisa Craig and Dr. Rosemary B. Cornell for agreeing to be my internal examiner and chair, respectively, of my thesis defence.

I would also like to express my special thanks to our Paetzel lab manager, Deidre de Jong-Wong for not only providing us with an organized working environment, but also for being a mentor and guiding me through all the work day difficulties. Without her support, it would be much harder for me to accomplish my graduate studies. Many thanks go to the past and present Paetzel lab members who have been an integral part of my research experience: Dr. Jae Lee, Dr. Kelly Kim, Dr. Charlie Stevens, Dr. Sung-Eun Nam, Dr. Apollos Kim, Dr. Ivy Chung, Chuanyun Luo, Daniel Chiang, Suraaj Aulakh, Linda Zhang, Micheal Ungerer, Zohreh Sharafianardakani and Eugene Shin. The knowledge, encouragement and friendships I gained from everyone have helped me become a better researcher and a more optimistic person.

Thanks to Dr. Jinrong Min and Dr. Nan Zhong for taking me into the Structure Genomic Consortium in University of Toronto when I was an undergraduate student at UBC. This opportunity was the first step in initiating my passion for protein structural research.

Finally, I would like to extend my thanks to my parents. During all these years while I have been so far away from home, you have always been my strongest support and motivation. Also, thanks to Jenson for being such a great friend, soul-mate and life partner. Thank you.

Table of Contents

Approval.....	ii
Partial Copyright Licence	iii
Abstract.....	iv
Dedication.....	v
Acknowledgements.....	vi
Table of Contents.....	vii
List of Tables.....	x
List of Figures.....	xi
List of Acronyms and Glossary.....	xiii

Chapter 1. Introduction.....	1
1.1. Bacteria Membrane.....	1
1.2. Outer Membrane Proteins (OMPs).....	3
1.2.1. OMP Structures.....	3
1.2.2. OMP Functions.....	4
1.2.3. OMP Biogenesis.....	6
1.2.3.1. Post-translational Targeting the Inner Membrane (IM).....	6
1.2.3.2. Translocation Across the IM.....	6
1.2.3.3. Transportation in Periplasm.....	7
1.2.3.4. Folding and Insertion into the OM.....	7
1.2.3.5. OMPs Degradation.....	8
1.3. The β -barrel Assembly Machinery (BAM) Complex.....	10
1.3.1. Discovery and Early Studies of the BAM Complex.....	10
1.3.2. Structure and Function of the <i>Escherichia coli</i> (<i>E.coli</i>) BAM Complex.....	11
1.3.2.1. The Lipoproteins.....	11
1.3.2.2. The Integral Membrane Component: BamA.....	14
1.3.3. Proposed Models of the <i>E. coli</i> BAM Complex.....	18
1.4. Project Objectives.....	20

Chapter 2. Refolding and Purification of <i>E. coli</i> BamA.....	21
2.1. Introduction.....	21
2.1.1. <i>E.coli</i> BamA Constructs.....	21
2.2. Material and methods.....	27
2.2.1. Full length <i>E.coli</i> BamA Δ 1-20 (BamA Δ 1-20).....	27
2.2.1.1. Protein Over-expression.....	27
2.2.1.2. Isolation of Protein Inclusion Bodies.....	27
2.2.1.3. Solubilization and Denaturation of Inclusion Bodies.....	28
2.2.1.4. Denatured Protein Purified by Nickel Affinity Chromatography.....	28
2.2.1.5. Protein Refolding by Rapid Dilution.....	28
2.2.1.6. Purification by Ion Exchange Chromatography.....	28
2.2.1.7. Purification by Analytical Size-exclusion Chromatography.....	29
2.2.2. POTRA Truncated <i>E.coli</i> BamA Δ 1-424 (BamA Δ 1-424).....	29
2.2.2.1. Cloning.....	29

2.2.2.2.	Protein Over-expression	29
2.2.2.3.	Isolation of Protein Inclusion Bodies	30
2.2.2.4.	Solubilization and Denaturation of Inclusion Bodies	30
2.2.2.5.	Denatured Protein Purified by Nickel Affinity Chromatography.....	30
2.2.2.6.	Protein Refolding by Rapid Dilution.....	30
2.2.2.7.	Purification by Analytical Size-exclusion Chromatography	31
2.2.3.	Native <i>E. coli</i> BamA (nBamA Δ 1-20).....	31
2.2.3.1.	Cloning	31
2.2.3.2.	Protein Over-expression	31
2.2.3.3.	Extracting nBamA Δ 1-20 from the OM	32
2.2.3.4.	Protein Purification by Nickel Affinity Chromatography.....	32
2.2.3.5.	Purification by Analytical Size-exclusion Chromatography	33
2.3.	Results	33
2.3.1.	Full length <i>E. coli</i> BamA (BamA Δ 1-20)	33
2.3.2.	Transmembrane <i>E. coli</i> BamA (BamA Δ 1-424)	37
2.3.3.	Native <i>E. coli</i> BamA (nBamA Δ 1-20)	40
2.4.	Discussion.....	44
 Chapter 3. Protein Refolding Status Analysis.....		47
3.1.	Introduction	47
3.2.	Material and Methods.....	48
3.2.1.	Size-exclusion Chromatography	48
3.2.2.	Heat-modifiability Test	48
3.2.3.	Circular Dichroism (CD) Spectroscopy.....	48
3.2.4.	Limited Proteolysis	49
3.3.	Results	49
3.3.1.	Size-exclusion Chromatography	49
3.3.2.	Heat-modifiability Test	51
3.3.3.	Circular Dichroism (CD) Spectroscopy.....	52
3.3.4.	Limited Proteolysis	58
3.4.	Discussion	60
 Chapter 4. <i>E. coli</i> BamA Crystallization.....		63
4.1.	Introduction	63
4.2.	X-ray Crystallography	63
4.2.1.	Protein Crystals	63
4.2.2.	Crystallization	65
4.2.3.	Data Collection and Processing	66
4.2.4.	Phasing	67
4.2.5.	Structure Refinement.....	68
4.3.	Material and Methods.....	68
4.3.1.	<i>E. coli</i> BamA Δ 1-20 Protein Crystallization	68
4.3.2.	<i>E. coli</i> BamA Δ 1-424 Protein Crystallization.....	68
4.3.3.	<i>E. coli</i> BamA Δ 1-424 Protein Optimization	69

4.4. Results	69
4.4.1. <i>E. coli</i> BamA Δ 1-20 and BamA Δ 1-424 Protein Crystals	69
4.5. Discussion	71

Chapter 5. <i>E. coli</i> BamA Homology Modeling Based on <i>Neisseria gonorrhoeae</i> BamA (<i>NgBamA</i>).....	74
5.1. Introduction	74
5.2. BAM Complex Homologues.....	74
5.2.1. Eukaryotic Homologues	75
5.2.2. Bacterial Homologues.....	76
5.3. Material and Methods	78
5.4. Results.....	79
5.4.1. Sequence Identities	79
5.4.2. <i>E.coli</i> BamA Homology Model	79
5.4.2.1. Hydrogen Bonds Close the β -barrel	82
5.4.2.2. Loop 6 (L6) Conformations.....	84
5.4.3. <i>E. coli</i> BamA Homology Model vs. Solved <i>E. coli</i> BamA structures.....	86
5.5. Discussion.....	89

Chapter 6. Summary and Future Directions.....	91
--	-----------

References.....	93
------------------------	-----------

Appendix A. Cloning Details	103
Appendix B. Molecular Mass Standard Curve for Size-exclusion Chromatography	104
Appendix C. CD Spectra of <i>E. coli</i> nBamA Δ 1-20 and BamA Δ 1-20	107
Appendix D. Sequence Alignments of <i>E. coli</i> BamA Homology Model.....	111

List of Tables

Table 1-1 List of Known BamA Structures	16
Table 2-1 List of E.coli BamA Constructs Generated.....	23
Table 2-2 Gel-filtration Elution Volumes for BamA Δ 1-20	34
Table 2-3 Gel-filtration Elution Volumes for BamA Δ 1-424	37
Table 2-4 Gel-filtration elution volumes for nBamA Δ 1-20	42

List of Figures

Figure 1-1 Bacterial Cell Envelopes	2
Figure 1-2 OMP Biogenesis and Degradation	9
Figure 1-3 Lipoprotein Structures	13
Figure 1-4 E. coli BamA Structures	17
Figure 1-5 Four Proposed Models of OMP Assembly.....	19
Figure 2-1 Inclusion Bodies Refolded BamA Δ 1-20	24
Figure 2-2 Inclusion Bodies Refolded BamA Δ 1-424	25
Figure 2-3 Native E. coli BamA Δ 1-20	26
Figure 2-4 E.coli BamA Δ 1-20 Inclusion Bodies Isolation and Protein Purification	35
Figure 2-5 The Size-Exclusion Chromatogram for Refolded BamA Δ 1-20	36
Figure 2-6 E. coli BamA Δ 1-424 Inclusion Bodies Isolation and Protein Purification	38
Figure 2-7 The Size-Exclusion Chromatogram for Refolded BamA Δ 1-424	39
Figure 2-8 E. coli nBamA Δ 1-20 Extraction from the OM and Purification.....	41
Figure 2-9 The Size-Exclusion Chromatogram for nBamA Δ 1-20	43
Figure 3-1 Superposed E. coli nBamA Δ 1-20 and Refolded BamA Δ 1-20 Size-Exclusion Chromatograms	50
Figure 3-2 Heat-modifiability Shift of E. coli nBamA Δ 1-20 and BamA Δ 1-20 as seen on 9% SDS-PAGE Gels.....	52
Figure 3-3 Far UV CD Spectra Characteristic of Common Secondary Structures.....	54
Figure 3-4 CD Spectrum of Buffer Used in the Experiment.....	55
Figure 3-5 CD Spectra of E. coli nBamA Δ 1-20 and BamA Δ 1-20	56
Figure 3-6 CD Spectra of E. coli nBamA Δ 1-20 and BamA Δ 1-20 Converted to Molar Ellipticity	57
Figure 3-7 SDS-PAGE Gels Showing the Results from Protease Digestion of E. coli nBamA Δ 1-20, BamA Δ 1-20, and BamA Δ 1-424	59

Figure 4-1 Protein Structure As Determined by X-ray Crystallography	64
Figure 4-2 Crystal Structure	64
Figure 4-3 Sitting Drop Vapour Diffusion	66
Figure 4-4 E. coli BamA Δ 1-20 and BamA Δ 1-424 Protein Crystals.....	70
Figure 4-5 X-ray Diffraction Patterns of Protein or Salt Crystals	71
Figure 5-1 Structures of B. pertussis FhaC, NgBamA and HdBamA Δ 3.....	77
Figure 5-2 E. coli full length BamA Homology Model.....	78
Figure 5-3 Sequence Alignment of E. coli BamA with N. gonorrhoeae BamA (NgBamA) and H. ducreyi BamA (HdBamA)	81
Figure 5-4 E. coli BamA Homology Model	82
Figure 5-5 The Hydrogen Bonds Between the First and Last β -strands.....	83
Figure 5-6 The L6 Sequence Alignment of BamA.....	85
Figure 5-7 The L6 Conformations in BamA	85
Figure 5-8 E. coli BamA Homology Model Superimposed to E. coli BamA POTRA 1-4	87
Figure 5-9 E. coli BamA Homology Model Superimposed to E. coli BamA β - barrel and E.coli BamA β -barrel with POTRA 5.....	88

List of Acronyms and Glossary

Å	Ångströms (10^{-10} m)
Asymmetric Unit	The smallest unit that generates one unit cell. It can be one molecule, more than one molecule, or one subunit of a multimeric protein.
β -barrel	A protein folding motif with a large β -sheet that forms a barrel-shaped structure. The N- and the C-terminal beta strands are closed by hydrogen bonds.
BAM Complex	β -barrel Assembly Machinery complex in Gram-negative bacteria, the protein complex required for proper folding and insertion of outer membrane proteins into the outer membrane. It is made up of five different proteins, BamA, B, C, D and E.
CL	Cardiolipin, a minor phospholipid component of bacterial and mitochondrial membranes.
Column Volume	The total volume of a chromatography column. It is the sum of the void volume and the matrix volume.
COOT	Crystallographic Object-Oriented Toolkit
DDM	n-Dodecyl- β -D-maltoside, a non-ionic detergent used for isolating and stabilizing membrane proteins.
DegP	A periplasmic chaperone/protease that acts in the Skp/DegP pathway to guide OMP-precursors to the BAM complex. Under high-temperature conditions, its protease function becomes active and it is able to degrade misfolded or aggregated OMPs in the periplasm.
Inclusion Bodies	Aggregates of misfolded recombinant protein over-expressed in the cytoplasm.
IM	Inner Membrane
IMP	Inner Membrane Protein
<i>in vitro</i>	In a test tube or petri dish (“within glass” in Latin). The experiment uses components of an organism that are isolated from their usual biological context.
<i>in vivo</i>	In a whole living organism (“within the living” in Latin).

IPTG	Isopropyl- β -D-thiogalactopyranoside
kDa	Kilodalton
Lipoprotein	A soluble protein in bacteria anchored to a membrane surface by diacylglycerol linked to its N-terminal cysteine.
LPS	Lipopolysaccharide, a molecule composed of a lipid and a polysaccharide found on the outer leaflet of bacterial outer membranes.
Mobile Phase	The fluid, containing the molecules of interest that travels through the stationary phase in chromatography.
OM	Outer Membrane
OMP	Outer Membrane Protein
PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction. It is a biochemical technique used to amplify a piece of DNA by several orders of magnitude.
PDB	Protein Data Bank
PE	Phosphatidylethanolamine
PEG	Polyethylene Glycol, used as precipitant for promoting crystallization
Peptidoglycan	A polymer consisting of carbohydrates and amino acids that forms a mesh-like layer in the cell walls of bacteria.
Periplasm	The space between the outer and the inner membranes of the Gram-negative bacteria, containing peptidoglycan.
PG	Phosphatidylglycerol
POTRA Domains	Polypeptide Transport Associated Domains. Five POTRA domains are found in the periplasmic region of BamA.
RMSD	Root-mean-square deviation, a value of the difference between several measured values. In this thesis, RMSD refers to how well two structures align with each other.
SAM Complex	Sorting and Assembly Machinery Complex, the homologue to the BAM complex found in the outer membrane of mitochondria.

SDS	Sodium Dodecyl Sulphate
Skp	The periplasmic chaperone in the Skp/DegP pathway which prevents aggregation of OMP-precursors in the periplasm.
Stokes Radius	The radius of a hypothetical hard sphere that would diffuse through the column medium at the same rate as the molecule being examined.
SurA	The periplasmic chaperone that binds OMP precursors to prevent aggregation. It is also known as peptidyl-prolyl cis-trans isomerase SurA.
Synchrotron	A cyclic particle accelerator that generates brilliant light (electromagnetic waves) that can be used to gather information about structural and chemical properties at the molecular level.
TOC Complex	Translocons at the Outer Envelope of Chloroplasts, the homologue to the BAM complex found in the outer membrane of chloroplasts.
Tris	Tris(hydroxymethyl)aminomethane, used as a common buffer solution, which has a pK_a of 8.06 at 25°C.
Unit Cell	The simplest repeating unit in a crystal.
Void Volume	The volume of the mobile phase in a chromatography column. It refers to the overall column volume minus the matrix volume.
X-ray	A form of electromagnetic radiation with wavelengths ranging between 0.01 to 10nm.
YASARA	Yet Another Scientific Artificial Reality Application, a molecular visualisation and dynamics program. The YASARA energy minimization server can improve side-chain accuracy in homology modeling.

Chapter 1. Introduction

1.1. Bacterial Membranes

The bacterial cell envelope is a multilayered construct that provides structural integrity to the microorganism. It protects the bacterium from the hostile outside environment and sorts the passenger molecules entering and exiting the cell (Costerton, Ingram, and Cheng 1974; Silhavy, Kahne, and Walker 2010) .

All bacteria are classified into two major groups: Gram-positive and Gram-negative, initially based on Christian Gram's staining results (Silhavy, Kahne, and Walker 2010) . The cell envelope structures of Gram-positive and Gram-negative bacteria are very different (Figure 1-1). In Gram-positive bacteria, such as *Staphylococcus aureus*, the cell envelope is composed of a single membrane with a thick mesh layer of peptidoglycan facing the exterior environment (Silhavy, Kahne, and Walker 2010) . In comparison, Gram-negative bacteria, such as *Escherichia coli*, have an inner membrane (IM) and an outer membrane (OM) with a peptidoglycan containing region between the two, called the periplasm (Silhavy, Kahne, and Walker 2010; Costerton, Ingram, and Cheng 1974) . The structural differences between Gram-positive and Gram-negative bacteria are primarily found in the peptidoglycan layer and in cell membrane composition (Costerton, Ingram, and Cheng 1974; Silhavy, Kahne, and Walker 2010; Vollmer and Bertsche 2008) .

The peptidoglycan region is made up of long polymers consisting of carbohydrates and amino acids that are cross-linked together to form a mesh-like layer. It provides rigidity to the cell and counteracts the osmotic pressure of the cytoplasm. Gram-positive bacteria lack an OM, so their peptidoglycan layer must supply all the structural strength necessary to maintain the cell's shape; as a result, it is substantially thicker than the peptidoglycan layer of Gram-negative bacteria (Vollmer and Bertsche 2008) .

Cell membranes are primarily composed of membrane proteins, lipopolysaccharides (LPS) and phospholipids such as phosphatidylethanolamine (PE), phosphatidyl-glycerol (PG) and cardiolipin (CL) (Ridgway and others 2008) . In Gram-positive bacteria, the phospholipids are evenly distributed on both leaflets with α -helical proteins spanning the distance between the leaflets. The IM of Gram-negative bacteria has same composition as that of Gram-positives. However, the Gram-negative OM has the phospholipids located on the inner leaflet and the LPS on the outer leaflet. Outer membrane proteins (OMPs) are inserted into the outer leaflet (Diedrich and Cota-Robles 1974; Fairman, Noinaj, and Buchanan 2011; Silhavy, Kahne, and Walker 2010) . About 50% of the OM's mass is composed of proteins, which are either integral (transmembrane) proteins or lipoproteins anchored to the OM by lipids (Koebnik, Locher, and Van Gelder 2000) . In this thesis, the term, OMP refers to the integral proteins in the OM. To date, most OMPs whose structures have been solved have transmembrane β -barrel structures, although there are two exceptions which have α -helical barrels instead. These exceptions are the *E. coli* polysaccharide translocon protein Wza and the *Corynebacterium glutamicum* porin PorB (Collins and Derrick 2007; Dong et al. 2006; Ziegler, Benz, and Schulz 2008) .

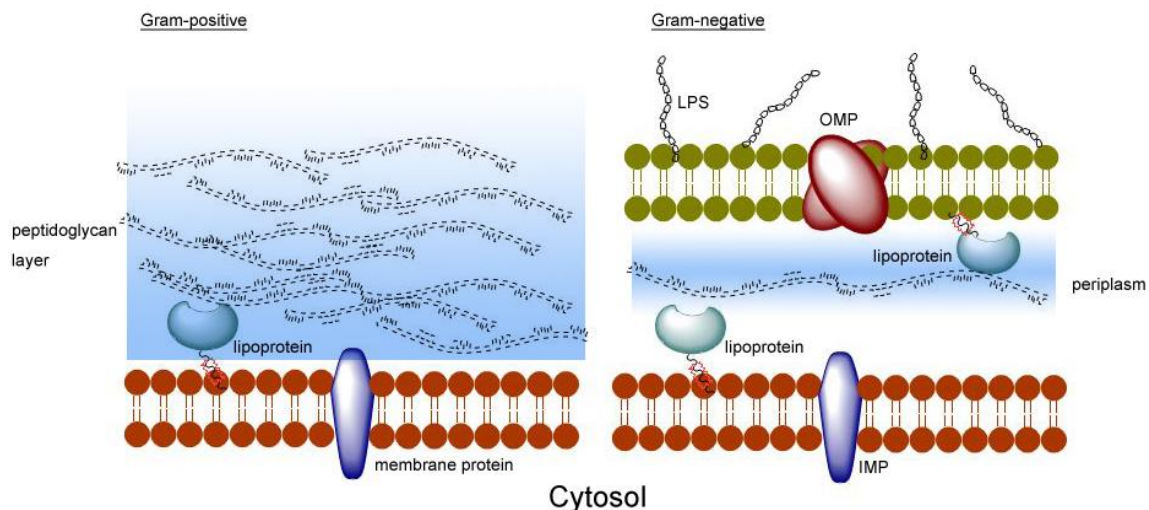


Figure 1-1 Bacterial Cell Envelopes

Gram-positive (left) and Gram-negative (negative) bacterial cell envelopes are shown for comparison. Gram-positive bacteria have a single layer phospholipid bilayer membrane, while Gram-negative bacteria have two layers, an inner-membrane (IM) and an outer-membrane (OM). The proteins inserted in the IM are inner-membrane proteins (IMP) and

the ones inserted in OM are outer-membrane proteins (OMP). Membrane anchored lipoproteins are present on both membranes. In Gram-negative bacteria, the space between the IM and the OM is known as periplasm and it has a peptidoglycan layer. Lacking the Gram-negative's two layers, Gram-positive bacteria instead contain a significantly thicker peptidoglycan layer.

1.2. Outer Membrane Proteins (OMPs)

The OM of Gram-negative bacteria provides an extra layer of protection without compromising the OM's molecule exchange ability (Delcour 2009) . It is a permeable barrier that sorts nutrients and prevents toxic molecules from entering (Rigel and Silhavy 2012) . The selective permeability of the OM is mainly determined by the pore-forming proteins, also called the outer membrane proteins (OMPs), embedded within it. Properly folded OMPs insert into the membrane and play an essential role in cell viability and in maintaining its functional structure (Delcour 2009) .

1.2.1. OMP Structures

Based on existing resolved structures, most OMPs have a transmembrane β -barrel domain, except for previously mentioned *E. coli* Wza and *Corynebacterium glutamicum* PorB which have α -helical barrels (Collins and Derrick 2007; Dong et al. 2006; Ziegler, Benz, and Schulz 2008) . Although most OMPs share a common β -barrel structural feature, they differ in many other characteristics, such as their number of β -strands, their properties of the loop and their oligomeric states.

The typical transmembrane β -barrel domain is composed of an even number of antiparallel β -strands, ranging from eight to 24. The number of amino acids found on each strand varies from six to 25, with an average of 12.3 residues (Koebnik, Locher, and Van Gelder 2000) . These residues are aligned on the β -strands that hydrophobic and hydrophilic amino acids are alternately arranged. The side chains of the hydrophilic residues face towards the interior of β -barrel, while the side chains of the hydrophobic residues point outwards, interacting with the hydrocarbon tails of lipids (Koebnik, Locher, and Van Gelder 2000; Buchanan 1999) .

Each strand of the β -barrel is connected alternately by a short turn found on the periplasmic side of the barrel, and a long loop located on the barrel facing the external space (Pautsch and Schulz 1998; Koebnik, Locher, and Van Gelder 2000; Fairman, Noinaj, and Buchanan 2011). Those exoplasmic loops are highly mobile and serve different essential roles in different OMPs; for example, controlling permeability in many porins and forming substrate binding sites (Buchanan 1999) .

The structural diversity of OMPs is also demonstrated by the various oligomeric states they form. In some cases, the oligomerization of β -barrel OMP single chains are required for proper function, such as the trimeric OmpF which forms diffusion pores for ions and other small molecules to pass through (Schulz 2002; Koebnik, Locher, and Van Gelder 2000) . In other cases, a single β -barrel is composed of multiple chains of OMPs. For examples, the transmembrane domain of TolC is a single barrel containing 12 antiparallel strands, contributed by three TolC monomers (Schulz 2002; Koronakis et al. 2000) .

1.2.2. OMP Functions

The β -barrel OMPs carry out a variety of functions including nutrient uptake, molecule exchange and antibiotic resistance (Bos, Robert, and Tommassen 2007) . Although their functions are very diverse, the β -barrel OMPs can be classified into following major functional categories: porins, translocons, enzymatic OMPs and structural OMPs (Wimley 2003) .

Porins are typically found as homotrimers spanning the OM, with each monomer containing 16 antiparallel β -strands (Delcour 2003; Nikaido 1994) . To transport small and hydrophilic molecules (<600 kDa) across the OM, the channels formed by porins allow the general diffusion of nutrients, requiring either specific or non-specific substrates (Nikaido 2003a) . The non-specific porins select the molecules based on size and polarity, with OmpF preferring large cations and PhoE preferring anions (Delcour 2003; Nikaido 2003b) . The specific porins only allow certain substrates to pass through. The representative examples are BtuB (specific for vitamin B12), LamB (specific for maltose and other sugars) and Tsx (specific for nucleosides) (Nikaido 1994) .

Other OMPs play an essential role in different export pathways in Gram-negative bacteria (Yen et al. 2002) . These OMPs are known as translocons. For example, TolC is an OMP which facilitates the exportation of protein, small molecules and drugs (antibiotic resistance) in the Type I secretion pathway without the use of the SecYEG translocon at the IM (Nikaido 2003a; Zgurskaya et al. 2011) . Translocons are also involved in the two-partner secretion pathway, a component of the Type V secretion pathway. In this pathway, protein exportation requires specific translocons (Jacob - Dubuisson, Locht, and Antoine 2001) . FhaC, from *Bordetella pertussis*, is responsible for exporting filamentous hemagglutinin which is an adhesin secreted during infection (Clantin et al. 2007a) . Another type of translocons, autotransporters are responsible for transporting molecules outside the cell by themselves, using their own N-terminal β -barrel domains to export the C-terminal passenger protein (Desvaux, Parham, and Henderson 2004) . Autotransporteration is usually involved in pathogenicity, with the passengers being virulence factors. Examples include adhesins such as AIDA-I and Ag43, and proteases such as Hbp and Pet (van Ulsen 2011) .

Another group of OMPs, such as OmpLA, OmpT, and PagP in *E. coli*, carry out enzymatic functions that take place at the OM, (Bishop 2005) . OmpLA is also known as Phospholipase A and hydrolyzes OM phospholipids. It is found as either a monomer or a dimer when substrates are present. The active site of OmpLA is located in the LPS-containing outer leaflet, where the presence of phospholipids disrupting the asymmetry of the OM can be detected (Dekker et al. 1997; Snijder et al. 1999) . OmpT is a protease that cleaves the protein with substrates including antimicrobial peptides released by host immune responses. The cleavage occurs between two basic residues of that target protein (Stumpe et al. 1998; Vandeputte - Rutten et al. 2001) . PagP is a potential drug target since it maintains the OM by transferring a palmitate chain from a phospholipid in the inner leaflet to the Lipid A component of LPS molecules in the outer leaflet of the OM (Bishop 2005)

The final group of OMPs play a structural role for the cell wall. The representative OMPs in this category are responsible for peptidoglycan synthesis (Mipa), LPS assembly (LptD), adhesion and entering host cells (OmpX), formation of pili substrates

(FimD) and OMP folding (BamA) (Okuda and Tokuda 2011; Phan et al. 2011; Vogt and Schulz 1999; Vollmer and Bertsche 2008; Voulhoux et al. 2003a) .

1.2.3. OMP Biogenesis

In prokaryotes, all proteins are synthesized in the cytosol, but some of them, such as OMPs, need to be transported to their target membrane or location to be fully functional (Schatz and Dobberstein 1996) . As well, correctly functioning folding and degradation strategies are essential to control protein quality in the entire biogenesis pathway (Figure 1-2).

1.2.3.1. Post-translational Targeting to the inner membrane (IM)

The precursor OMP (pre-OMP) is synthesized post-translationally in the cytosol with a cleavable signal sequence at its N-terminus) . Once the signal sequence emerges from the ribosome, it binds to a ribosome association chaperone, trigger factor (TF), until translation is complete (Hoffmann, Bukau, and Kramer 2010) . Next, the pre-OMP is passed over to another cytoplasmic chaperon, SecB, and then delivered for IM translocation to the SecYEG complex (SecY, SecE, SecG, SecD, SecF, YajC) (Bechtluft et al. 2010; Driessen and Nouwen 2008) . The binding of the chaperons (TF and SecB) to the pre-OMP ensures the pre-OMP in a stable, unfolded form before being translocated across the IM (Hoffmann, Bukau, and Kramer 2010) .

1.2.3.2. Translocation Across the IM

Once SecB guides the pre-OMP to the SecYEG complex, the homodimeric SecA (a known ATPase located on the SecYEG complex), releases SecB and assists the pre-OMP going through the SecYEG complex in a ATP-dependent manner (Cross et al. 2009; Driessen and Nouwen 2008; Zimmer, Nam, and Rapoport 2008) . The pre-OMP then emerges into the periplasmic space and is released after its N-terminal signal sequence is cleaved by signal peptidase I (SPaseI). Once SPaseI has cleaved after the conserved Ala-X-Ala sequence in the C-terminal region of the signal peptide (Paetzel et al. 2002), the pre-OMP becomes a mature OMP.

1.2.3.3. Transportation in Periplasm

As the OMP is translocated across the IM and released into the periplasmic space, it is stabilized in its unfolded form through association with the periplasmic chaperons, SurA or Skp/DegP (Sklar et al. 2007c; Rizzitello, Harper, and Silhavy 2001) . SurA usually recognizes the aromatic residues at the C-terminus of OMPs, whereas Skp recognizes the exposed hydrophobic regions of unfolded OMPs and DegP binds to the misfolded OMPs (Bitto and McKay 2003; Qu et al. 2007). Gene knockout studies have suggested that cells are viable when either the SurA or the Skp/DegP pathway is functional (Rizzitello, Harper, and Silhavy 2001). The SurA pathway is prominent under normal conditions, while the Skp/DegP pathway is essential under stressful conditions (Sklar et al. 2007c). Each pathway plays an essential role in preventing OMPs from misfolding and aggregating during the period where the OMPs are delivered to the OM, depending on the environmental conditions.

1.2.3.4. Folding and Insertion into the OM

After being transported to the OM, OMPs are recognized by the OM targeting information found primarily at the C-terminus with folding and insertion taking place on the OM (Robert et al. 2006). *In vitro* studies show that OMPs can fold and insert into synthetic phospholipid bilayer membranes spontaneously, which suggests the folding and insertion of OMPs do not require any proteinaceous machinery (Surrey and Jahnig 1992a). However, the efficiency of this spontaneously reaction is too low to be biologically relevant. To increase the kinetics of folding and membrane insertion, an OMP assembly factor, the β -barrel assembly machinery (BAM) complex, is required *in vivo*. The presence of the BAM complex in the OM but not in the IM also explains why OMPs are not assembled in the IM of the cells (Tamm, Hong, and Liang 2004a). In *E. coli*, the BAM complex consists of a integral membrane protein BamA and four lipoproteins, BamB, BamC, BamD, and BamE (Ricci and Silhavy 2012; Hagan, Kim, and Kahne 2010). The exact mechanism of the BAM complex is not well understood, but a proposed model will be discussed in Section 1.3.3.

1.2.3.5. OMPs Degradation

An essential part of quality control, the OMPs degradation pathway eliminates any damaged or misfolded OMPs, especially when cells are under stress. When an OMP is mislocalized and not targeted to the BAM complex, the C-terminus of the OMP activates the DegS protease and initiates the sigmaE pathway. When this occurs, the expression of SurA, Skp and DegS are increased to prevent further mislocalization (Miot and Betton 2004, 4) . Meanwhile, the damaged or misfolded OMPs are recognized and degraded by DegP, using its proteolytic function (Sklar et al. 2007c) .

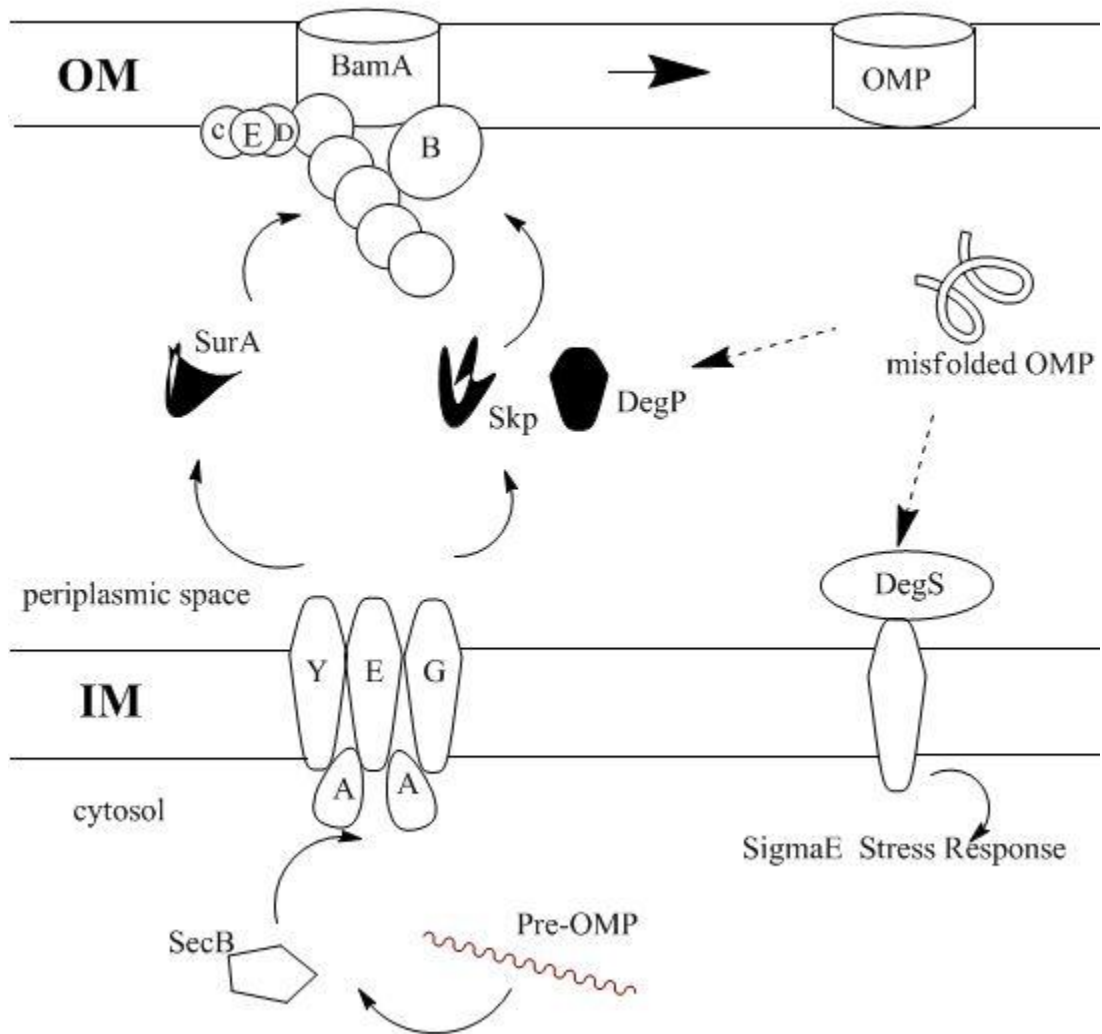


Figure 1-2 OMP Biogenesis and Degradation

Once the pre-OMP is synthesized in the cytoplasm, it is translocated across the IM via the Sec translocation system. Then, the pre-OMP is released to the periplasmic space with the assistance of chaperones (either through the SurA pathway or the Skp/DegP pathway) to prevent the OMP from aggregating and misfolding. The OMP is delivered to the OM and recognized by the BAM complex, which mediates OMP assembly in the OM. Misfolded or aggregated OMPs in the periplasm are recognized and degraded by DegP, or bound by DegS which initiates the sigmaE stress response.

1.3. The β -barrel Assembly Machinery (BAM) Complex

1.3.1. *Discovery and Early Studies of the BAM Complex*

Omp85 is a highly conserved protein in Gram-negative bacteria with homologues also found in eukaryotes. Since the gene encoding Omp85 is very close to other genes involved in the OM biogenesis, it was thought to have a similar function in OMP assembly, rather than in lipid transport and LPS assembly (Voulhoux et al. 2003a; Genevrois et al. 2003) . Experiments suggested that Omp85 depleted strains would cause the accumulation of misfolded OMPs, leading to cell death (Voulhoux et al. 2003a) . Omp85 was the first OMP assembly factor identified in *Neisseria meningitidis* in 2003, and its homologues were also found in mitochondria and chloroplasts (Walther, Rapaport, and Tommassen 2009, 2789-2804; Gentle et al. 2004) . Recently, the full length structure of Omp85 in *Neisseria gonorrhoeae* (NgBamA) has been solved by X-ray crystallography (Noinaj et al. 2013). Ironically, *Escherichia coli* has become a popular model used to study Omp85 (named as BamA), but the structure of the full length *E. coli* BamA remains unknown. The binding partners of BamA were determined by early co-purification experiments. Those components are lipoproteins (BamB, BamC, BamD and BamE) that form a large BAM complex with BamA together (Malinverni et al. 2006; Sklar et al. 2007a; Wu et al. 2005a). These subunits show different degrees of conservation among Gram-negative bacteria. BamA and BamC are most ubiquitous in proteobacteria, while BamB and BamE are absent in δ -proteobacteria and ϵ -proteobacteria. BamC is the least conserved and is only found in β -proteobacteria and γ -proteobacteria (Anwari et al. 2012) . Recent studies suggest another lipoprotein, BamF, exists in α -proteobacteria when BamC is absent. However, the structural and functional aspects of BamA are not well known at present; so, this thesis only focuses on BamA, BamB, BamC, BamD and BamE (Anwari et al. 2012) .

1.3.2. Structure and Function of the *E. coli* BAM Complex

The *E. coli* BAM complex consists of one integral OMP BamA and four lipoproteins: BamB, BamC, BamD and BamE. These five proteins are named in order of decreasing molecular mass. The structural and functional information of each protein will be included in this section.

1.3.2.1. The Lipoproteins

BamB (UniProt ID: P77774) is the largest lipoprotein in the BAM complex with a molecular mass of 39.9 kDa. It is not essential for cell viability, although the deletion of BamB causes significant defects in the assembly of many large β -barrel OMPs (Charlson, Werner, and Misra 2006; Ruiz et al. 2006). Therefore, the function of BamB is believed to enhance the overall activity of BAM complex (Wu et al. 2005b) . As shown in Figure 1-3A, BamB has an eight-bladed β -propeller structure with each blade containing a four-stranded anti-parallel β -sheet (Albrecht and Zeth 2011; Heuck, Schleiffer, and Clausen 2011; Kim and Paetzel 2011). To date, BamB is believed to interact with BamA, but not with other four lipoproteins (Malinverni et al. 2006; Kim et al. 2007; Sklar et al. 2007b) .

The lipoprotein BamC (UniProt ID: P0A903), like BamB, is not essential for cell viability, but the absence of BamC leads to increase membrane permeability and has a mild impact in OMP assembly (Onufryk et al. 2005; Wu et al. 2005b) . BamC is 34.4 kDa and has three independently folding domains; the N-terminal unstructured region, the N-terminal domain and the C-terminal domain, shown in Figure 1-3B (Albrecht and Zeth 2011; Kim, Aulakh, and Paetzel 2011; Kim et al. 2011a) . The structure of BamC has been solved. The structure of the C-terminal domain of BamC has been determined, as has the structure of BamC in its BamCD complex form (Kim et al. 2011a; Kim, Aulakh, and Paetzel 2011).

BamD (UniProt: P0AC02) is highly conserved among Gram-negative species (although absent in eukaryotes) and it is the only essential lipoprotein in the BAM complex. The absence of BamD leads to cell death (Malinverni et al. 2006) . As shown in figure 1-3C, the structure of BamD shows five tetratricopeptide repeat (TPR) motifs, with each TRP containing two α -helices (Kim, Aulakh, and Paetzel 2011; Sandoval et al.

2011) . BamD interacts directly with BamA, BamC and BamE, via those TPR motifs (Wu et al. 2005a; Malinverni et al. 2006; Sklar et al. 2007c; Knowles et al. 2011) . The C-terminal OMP signal of BamD is believed to contribute to substrate recognition and binding, as demonstrated in cross-linking experiments (Albrecht and Zeth 2011).

BamE (UniProt: P0A937) is the smallest lipoprotein in the BAM complex, with a molecular mass of 10.4 kDa. Since BamE is small in size, the monomeric BamE structure was solved using nuclear magnetic resonance (NMR) spectroscopy (Kim et al. 2011b). Homodimeric BamE protein can be expressed and purified; however, its function and biological relevance remain unclear (Albrecht and Zeth 2011; Kim et al. 2011b; Knowles et al. 2011) . As shown in figure 1-3D, the BamE structure contains a three-stranded anti-parallel β -sheet and two α -helices. Both the unstructured N-terminus and C-terminus of BamE show high degrees of flexibility (Kim et al. 2011b) . Previous studies have suggested that BamE is involved in different protein-protein interactions with BamA, BamC and BamD. Therefore, BamE is believed to play an essential role in stabilizing the BAM complex (Kim et al. 2011b; Sklar et al. 2007b) .

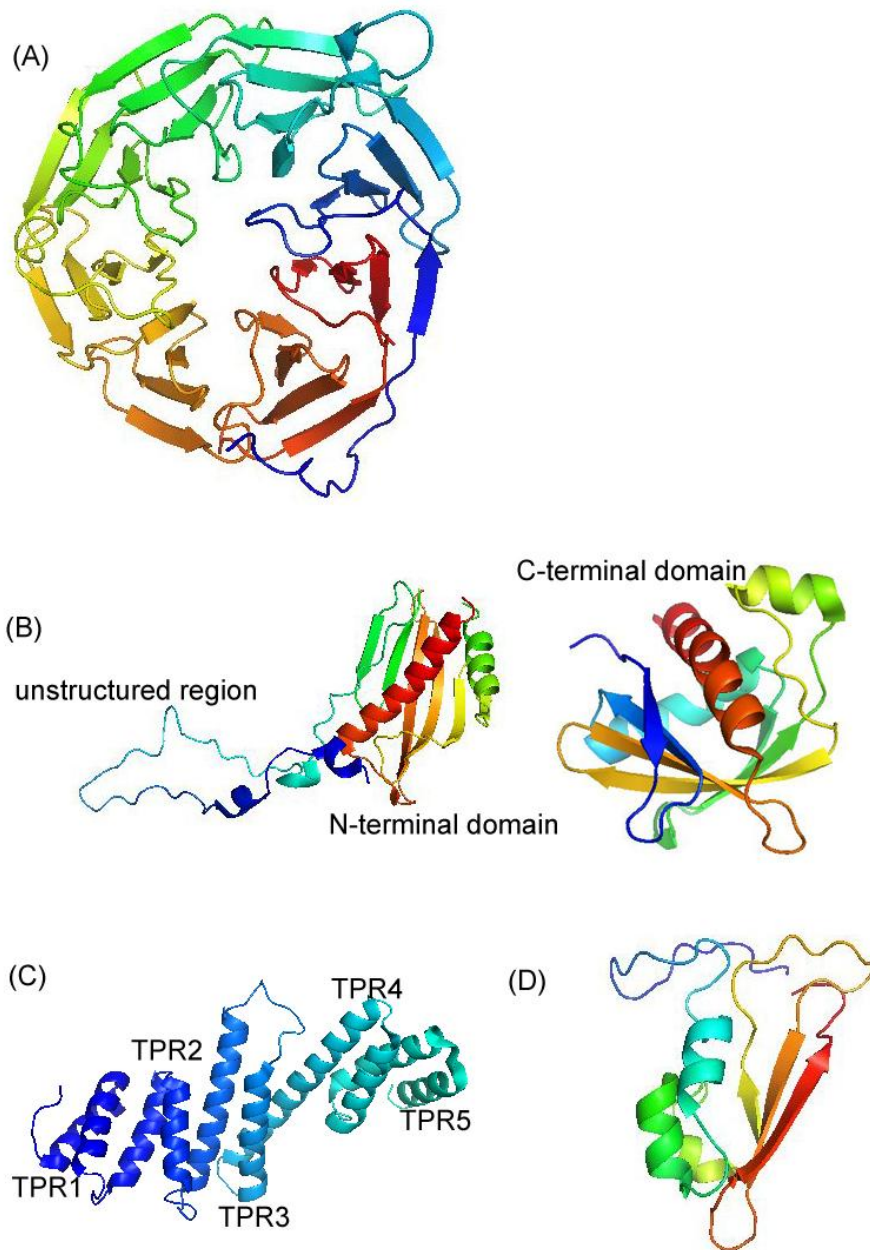


Figure 1-3 Lipoprotein Structures

All structures are displayed using PyMOL. The structures start from N-terminus shown in blue and end at the C-terminus shown in red. (A) The colored ribbon diagram of BamB (PDB: 3P1L) shows eight domains that form an eight-bladed β -propeller structure. (B) The C-terminal domain of BamC was solved separately from the N-terminal domain and the unstructured region. In the left panel, the unstructured region with the N-terminal region is solved as a part of the BamCD dimer structure (PDB: 3TGO). In the right panel, the C-terminal region shows a helix-grip fold (PDG: 2YH5). (C) The colored ribbon diagram shows that BamD (PDB: 3TGO) contains ten α -helices which form the five TPR motifs as

labeled on diagram. (D) The colored ribbon diagram of monomeric BamE (PDB: 2KXX) shows the two α -helices pack against a three-stranded β -sheet.

1.3.2.2. The Integral Membrane Component: BamA

BamA was initially identified in *Neisseria meningitidis*, with homologues discovered in all Gram-negative bacteria, mitochondria and chloroplasts (Walther, Rapaport, and Tommassen 2009). The integral membrane protein BamA is an OMP with a molecular mass of 88 kDa (UniProt ID: P04A940). As the core component in the BAM complex, the gene encoding BamA is highly conserved across Gram-negative bacteria and eukaryotes. The depletion of BamA will lead to cell death.

E. coli BamA consists of two major domains: the N-terminus periplasmic domain and the C-terminus transmembrane domain as shown in figure 1-4. The N-terminal periplasmic region is also known as the polypeptide transport associated (POTRA) domains. There are five POTRA in *E. coli* BamA, which are numbered 1 to 5 from the N- to C- terminus (Sánchez-Pulido et al. 2003). To date, structures for BamA have been determined for the constructs listed in Table 1-1. As of the writing of this thesis, the structures of the transmembrane β -barrel are have just been submitted; while, crystallizing the full length *E. coli* BamA continues to present challenges.

These resolved BamA structures suggest that five POTRA domains of *E. coli* BamA share a similar structure, forming a three-stranded β -sheet with two α -helices in a β - α - α - β - β fold (Kim et al. 2007). The five POTRA domains may alternate between extended and bent forms. This two conformation structure feature is believed to assist in their function such as docking sites for lipoproteins and as part of gating mechanism to isolate the space between periplasmic space and extracellular environment (Noinaj et al. 2013; Gatzeva-Topalova et al. 2010).

Determining the structure of the transmembrane domain of BamA also helped in our understanding of BamA's function. An early structural study of BamA homologue, *Bordetella pertussis Fhc*, suggested that the transmembrane domain of BamA might have a 16-stranded β -barrel (Clantin et al. 2007a). Recently, the full length structure of BamA from *Neisseria gonorrhoeae* and *Haemophilus ducreyi* POTRA 1-3 truncated

BamA from have provided confirmation that the transmembrane β -barrel domain contains a 16-stranded β -barrel with its first and last strands associating together to close the barrel (Noinaj et al. 2013) . The hydrophobic belt of the β -barrel is dramatically reduced in width, which suggests the OM distorting mechanism of BamA as shown in molecular dynamics simulation (Noinaj et al. 2013) . To increase the accuracy of speculated full length *E. coli* BamA structure, a homology model was made based on *NgBamA* structure. Together with a detailed discussion I made is presented in Chapter 5. However, as of writing, two new *E. coli* BamA β -barrel structures have just been submitted. One structure (PDB: 4N75) shows the dimerized transmembrane β -barrel domain, while the other structure (PDB: 4C4V) shows another dimer where each chain contains the transmembrane β -barrel and POTRA 5 (Figure 1-4C). Both structure reinforce that the transmembrane domain is composed of 16-stranded β -barrel. Obtaining clear insight into the transmembrane β -barrel domain is a further step towards solving the full length *E. coli* BamA structure and understanding OMP biogenesis (Albrecht et al. 2014; Ni et al. 2014) .

Co-immunoprecipitation studies suggest that BamA interact directly with BamB and BamD; meanwhile, BamA may allow association of BamC and BamE through BamD (Malinverni et al. 2006) . Deletion analysis shows that POTRA2-5 co-purify with BamA that POTRA5 is responsible for BamC, BamD or BamE co-purification (Kim et al. 2007) . In addition, POTRA1 shows interaction with SurA and POTRA1-2 shows interaction with the OMP substrate, PhoE (Bennion et al. 2010; Knowles et al. 2008) .

Table 1-1 List of Known BamA Structures

Organism	Fragment	POTRA	β-barrel	PDB
<i>Escherichia coli</i> K-12	Residue 266-420	P4,5	N/A	3Q6B
<i>Escherichia coli</i> K-12	Residue 264-424	P4,5	N/A	3OG5
<i>Escherichia coli</i> K-12	Residue 21-351	P1,2,3,4	N/A	2QDF
<i>Escherichia coli</i> K-12	Residue 21-410	P1,2,3,4	N/A	3EFC
<i>Escherichia coli</i> K-12	Residue 427-810	N/A	Yes	4N75
<i>Escherichia coli</i> K-12	Residue 344-808	P5	Yes	4C4V
<i>Escherichia coli</i> K-12	Residue 347-808	P5	Yes	4C4V
<i>Haemophilus ducreyi</i>	Residue 262-793	P4,5	Yes	4K3C
<i>Neisseria gonorrhoeae</i>	Residue 21-792	P1,2,3,4,5	Yes	4K3B

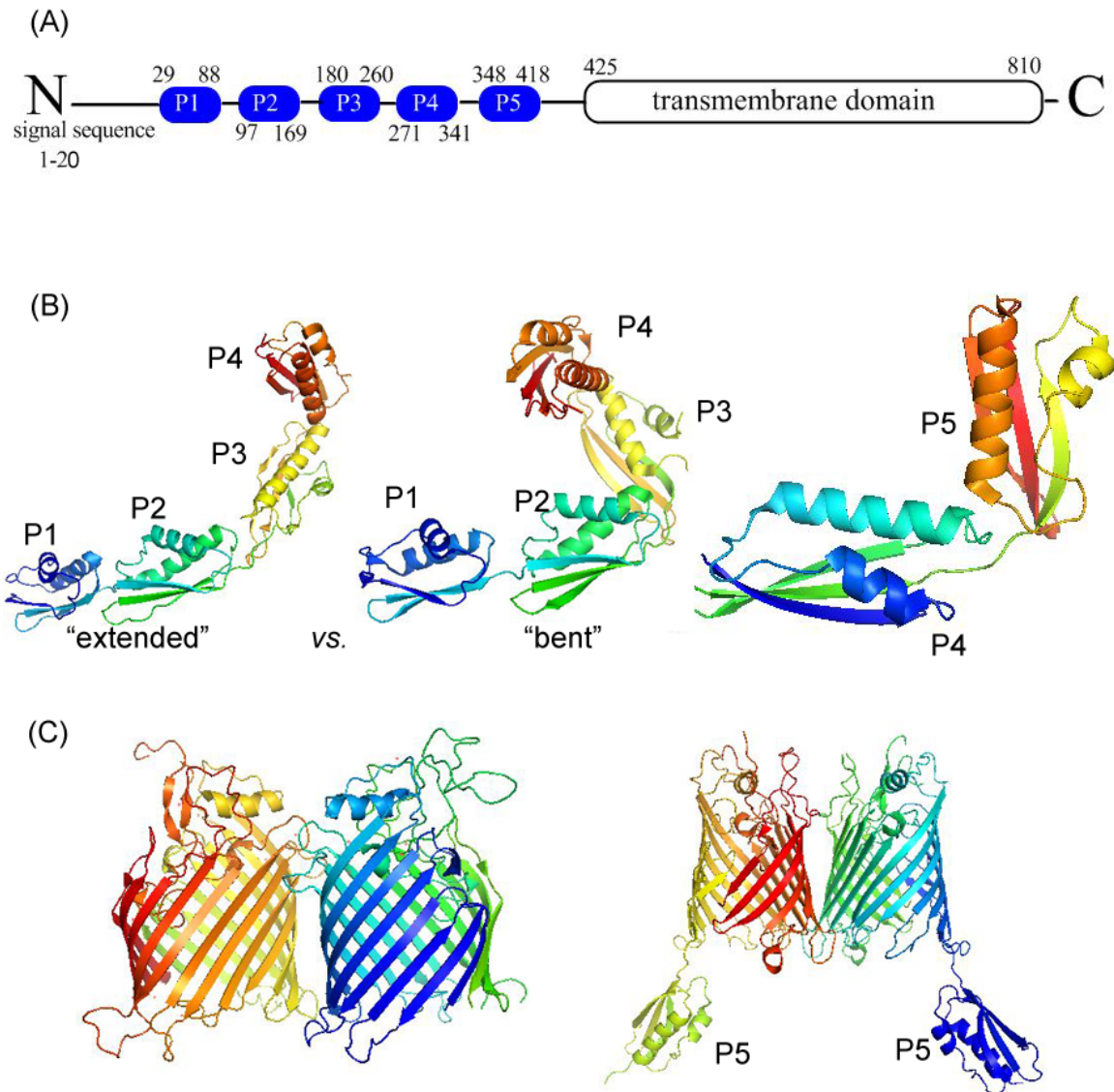


Figure 1-4 *E. coli* BamA Structures

All structures are displayed using PyMOL. The structures start from the N-terminus shown in blue to the C-terminus shown in red. (A) The domain diagram of BamA shows an N-terminal periplasmic region that contains five polypeptide transport associated (POTRA) domains along with a C-terminal transmembrane domain. The first 21 amino acid residues are the signal peptide that has been cleaved by SPase I during transportation across the IM. (B) The POTRA1-4 structures are shown in either extended form or bent form (PDB: 2QDF and 3EFC). The POTRA 5 structure has been solved together with POTRA4 (PDB: 3Q6B). (C) The dimerized transmembrane β -barrel domain of BamA is presented (PDB: 4N75); while, to the right, the BamA dimer structure is together with POTRA 5 (PDB: 4C4V).

1.3.3. Proposed Models of the *E. coli* BAM Complex

The least understood aspect of the BAM complex's function is the detailed process through which OMPs are folded and inserted into the OM lipid bilayer yielding the β -barrel structure. Four models have been proposed to explain this process as shown in Figure 1-5 (Kim, Aulakh, and Paetzel 2012) .

The first model shown in Figure1-5A suggests that the OMP travel through the BamA β -barrel and inserts into the OM from the extracellular side. This assumes that the β -barrel creates a large enough channel for OMP to pass through and that the OMP insertion occurs without additional help (Tomassen 2007). The second model in Figure1-5B shows that the OMP inserting from the periplasmic space and folding between the BamA β -barrel and the lipid bilayer. This model suggests the outer wall of the BamA β -barrel is used as a scaffold for OMP assembly (Hagan, Silhavy, and Kahne 2011) .The third model is based on the observation that tetrameric BamA has a limited space in its center which can facilitate the closing of the β -sheet into a the β -barrel. This model requires the association and dissociation of tetrameric BamA subunits to ensure the release of folded OMP to lipid bilayer (Figure 1-5C). The last model is the one our lab proposed in 2012, where we suggest that the N- and C- terminal β -strands of BamA may serve as folding templates for the OMP substrate (Figure 1-5D). These two terminal β -strands are initially closed by hydrogen bonds. During the process of OMP assembly, those hydrogen bonds would be interrupted by the substrate and then form new hydrogen bonds to ensure the energy cost is compensated. This model keeps the hydrophobic residues of a substrate facing the membrane, while hydrophilic residues face away from the OM (Kim, Aulakh, and Paetzel 2012) . Recently, the structures of the BamA transmembrane β -barrel have been revealed in different species (*E. coli*, *N. gonorrhoeae* and *H.ducrevi*). The electrostatic surface of the transmembrane β -barrel shows a narrowed hydrophobic near the C-terminus which suggests BamA may distort the OM and create a lateral opening between the first and last β -strands(Noinaj et al. 2013; Noinaj et al. 2014; Ni et al. 2014). However, more experiments are required before any one of these models can be considered valid.

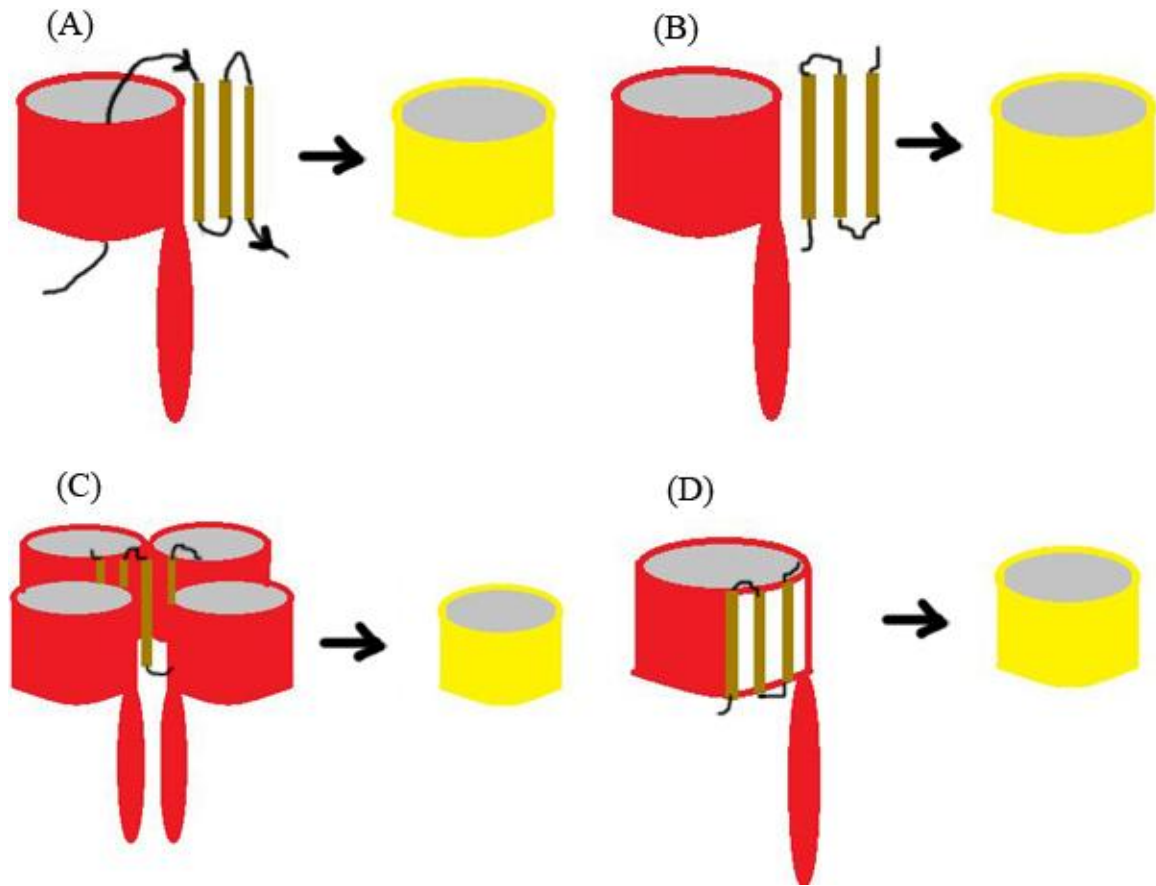


Figure 1-5 Four Proposed Models of OMP Assembly

Four models have been proposed to explain the process by which the BAM complex facilitates the folding and insertion of OMPs. BamA is shown in red; while the protein substrate is shown in yellow. The extracellular space is above the outer membrane, while the periplasmic space is below. (A) The substrate protein is translocated across the outer membrane through the channel (β -barrel domain of BamA), and inserts from the extracellular side. (B) The substrate inserts into the lipid bilayer from the periplasmic face of the outer membrane, using the outer wall of the BamA β -barrel as a scaffold for folding. (C) BamA forms a tetramer to coordinate OMP assembly, similar to the second model. (D) The N- and C-terminal β -strands of BamA are used as folding templates. The initial hydrogen bonds in the β -barrel are interrupted; the incoming substrate will form new hydrogen bonds instead.

1.4. Project Objectives

The goal of this research project was to enhance the understanding of OMP biogenesis by studying *E. coli* BamA which is the core component in BAM complex. Having the ultimate goal of determining the full length *E. coli* BamA structure using X-ray crystallography, recombinant *E. coli* BamA proteins were over-expressed, refolded and purified. The following objectives were carried out, with the emphasis on protein purification and refolding status analysis:

1. Obtain a refolding and purification protocol for *E. coli* full length BamA Δ 1-20 (BamA Δ 1-20) and POTRA truncated *E. coli* BamA Δ 1-424 (BamA Δ 1-424).
2. Extract the native BamA (nBamA Δ 1-20) from *E. coli* OM.
3. Analysis the status of refolded BamA Δ 1-20 and BamA Δ 1-424.
4. Crystallize the refolded *E. coli* BamA.

In order to address these research goals, different experimental methods were carried out including: cloning, over-expression, inclusion bodies purification, membrane protein extraction, size-exclusion chromatography, heat-modifiability shift test, circular dichroism (CD) spectroscopy, limited proteolysis, protein crystallization and homology modeling. Detailed materials and methods will be described in the following chapters.

Chapter 2. Refolding and Purification of *E. coli* BamA

2.1. Introduction

This chapter will discuss the refolding and purification of *E. coli* BamA protein expressed from three different constructs; full length *E. coli* BamA Δ 1-20 (BamA Δ 1-20), POTRA truncated *E. coli* BamA Δ 1-424 (BamA Δ 1-424) and native *E. coli* BamA (nBamA Δ 1-20). The BamA Δ 1-20 and BamA Δ 1-424 inclusion bodies were denatured and refolded following the rapid dilution method. The refolded protein was then further purified by affinity, ion exchange and size-exclusion chromatography. However, nBamA Δ 1-20 is expressed natively and inserted in the outer membrane (OM), so in this case, the protein was simply extracted from the OM and purified. Detailed methods and results for individual experiments are described in subsequent chapters.

2.1.1. *E.coli* BamA Constructs

In this thesis project, two BamA constructs, BamA Δ 1-20 and BamA Δ 1-424, were utilized to determine the correct refolding strategy and search for successful crystallization conditions. The other construct, nBamA Δ 1-20, was designed to express native BamA inserted in the OM of *E. coli*. It served as a reference to test the refolding of BamA Δ 1-20 and BamA Δ 1-424 (Table 2-1).

The *E. coli* BamA Δ 1-20 construct was available in Paetzel lab construct library, having been designed and cloned by Dr. Kelly Kim. The construct contains an N-terminal hexahistidine tag in front of the POTRA domain. Lacking the signal sequence (residues 1-20), BamA Δ 1-20 is expressed in the cytosol, forming inclusion bodies. The theoretical

molecular mass of BamA Δ 1-20 is 90.7 kDa and its isoelectric point (pI) is 5.10 (Figure 2-1 A).

The BamA Δ 1-424 construct is designed to include the transmembrane β -barrel domain with an N-terminal hexahistidine tag located in front of the β -barrel. Without the signal sequence (residues 1-20), BamA Δ 1-424 is expressed in the cytosol, forming inclusion bodies. This POTRA truncated BamA has a theoretical molecular mass of 45.4 kDa and a pI of 4.79 (Figure 2-1 B).

The nBamA Δ 1-20 construct is designed to express full length BamA inserted into the *E. coli* OM. This construct was cloned by myself and Chuanyun. The PCR products were ligated into pET 20b (+) with a C-terminal hexahistidine tag and an N-terminal pelB leader, which serves as a signal sequence. Once the protein is expressed in the cytosol, the pelB sequence directs the protein to the IM for translocation. Following that, the pelB leader is cleaved, releasing the protein into the periplasmic space. Then nBamA Δ 1-20 is assembled and inserted into the OM with an attached C-terminal hexahistidine tag. The nBamA Δ 1-20 protein has a theoretical molecular mass of 90 kDa and a pI of 4.98 (Figure 2-1 C).

The cloning details of BamA Δ 1-424 and nBamA Δ 1-20 are included in Appendix A.

Table 2-1 List of *E.coli* BamA Constructs Generated

MWP#	Constructs	Description	Residues	Molecular Mass(kDa) *	His×6 tag	Signal Sequence
275	BamAΔ1-20	Full length BamA	21-810	91	N- terminal	N/A
853	BamAΔ1-424	POTRA truncated BamA	425-810	45	N- terminal	N/A
1043	nBamAΔ1-20	Native full length BamA	21-810	91	C- terminal	N- terminal(peIB)

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

BamA Δ 1-20

(A)

MGSSHHHHHHSSGLVPRGSHM**A**EGFVVKDIHFEGLRVAVGAALLSMPVRTGDTVNDEDISNT
IRALFATGNFEDVRVLRDGDITLLVQVKERPTIASITFSGNKSVKDDMLKQNLEASGVRVGESLDRT
TIADIEKGLDFYYSVGKYSASVKAVVTPLPRNRVDLKLVFQEGVSAEIQQINIVGNHAFTTDELIS
HFQLRDEVPWWNVVGDRLKYQKQKLAGDLETLRSYYLDRGYARFNIDSTQVSLTPDKKGIYVTVN
ITEGDQYKLSGVEVSGNLAGHSAEIEQLTKIEPGELYNGTKVTKMEDDIKLLGRYGYAYPRVQS
MPEINDADKTVKLRVNDAGNRFYVRKIRFEGNDTSKDAVLRREMRQMEGAWLGSDLVDQGKE
RLNRLGFFETVDTDTQRVPGSPDQVDVVYKVKERNTGSFNFYGIGYGTESGVSFQAGVQQDNWL
GTGYAVGINGTKNDYQTYAELSVTNPFYFTVDGVSLLGGRLFYNDFQADDADLSDYTNKSYGTDVT
LGFPINEYNSLRAGLGYVHNSLSNMQPQVAMWRYLYSMGEHPSTSDQDNSFKTDDFTFNYGWT
YNKLDRGYFPTDGSRVNLTGKVTIPGSDNEYKVTLDATYVPIDDDHKWVVLGRTRWGYGDGL
GGKEMPFYENFYAGGSSTVRGFQSNITIGPKAVYFPHQASNYDPDYDYECATQDGAKDLCKSDD
AVGGNAMAVASLEFITPTPFISDKYANSVRTSFFWDMGTVWDTNWDSSQYSGYPDYSDPSNIRM
SAGIALQWMSPLGPLVFSYAQPFFKKYDGDKAEQFQFNIGKTW-STOP

(B)

Theoretical isoelectric point (pI): 5.10

Theoretical molecular mass: 90.7kDa

(C)

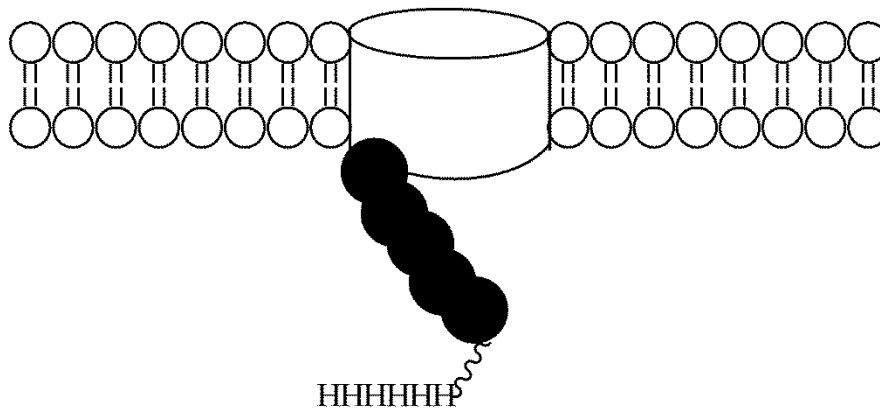


Figure 2-1 Inclusion Bodies Refolded BamA Δ 1-20

(A) The expression sequence of BamA Δ 1-20 (UniProt: P0A940) in vector pET28a(+) has a hexahistidine tag located before the inserted BamA sequence where amino acid residue 21 (Ala, A) is indicated in black and bold front. (B) Theoretical pI and molecular mass of BamA Δ 1-20 are shown*. (C) The diagram shows the protein corresponding to the designed construct inserted into the OM, where the POTRA domain is shown as black circles and the transmembrane β -barrel is in white. The N-terminus hexahistidine tag is shown preceding the POTRA domain. BamA Δ 1-20 is expressed in the cytosol as inclusion bodies since it lacks the signal sequence.

* The theoretical pI and MW were calculated using ProtParam (Gasteiger et al. 2005, 571-607)

BamA Δ 1-424

(A)

MGSSHHHHHHSSGLVPRGSHM**S**FNFGIGYGTESGVSFQAGVQQDNWLGTGYAVGINGTKND
YQTYAELSVTNPYFTVDGVSLGGRLFYNDFFQADDADLSDYTNKSYGTDVTLGFPINEYNSLRAG
LG YVHNSLSNMQPQVAMWRYLYSMGEHPSTSDQDNSFKTDDFTFNYGWTYNKLDRGYFPTDG
SRVNLTKVKTIPGSDNEYKVTLDTATYVPIDDDHKWVVLGRTRWGYGDGLGGKEMPFYENFY
AGGSSTVRGFQSNITIGPKAVYFPHQASNYDPDYDYECATQDGAKDLCKSDDAVGGNAMAVASL
EFITPTPFISDKYANSVRTSFFWDMGTVWDTNWDSSQYSGYPDYSDPSNIRMSAGIALQWMSPL
GPLVFSYAQPFFKKYDGDKAEQFQFNIGKTW-STOP

(B)

Theoretical isoelectric point (pI): 4.79

Theoretical molecular mass: 45.4kDa

(C)

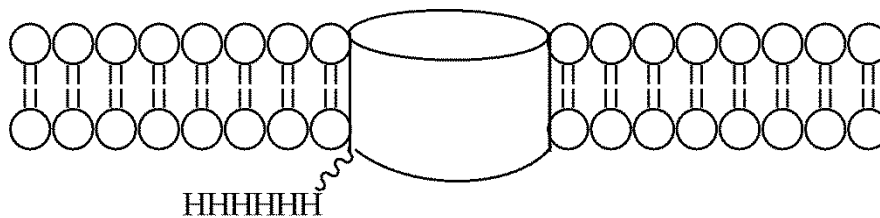


Figure 2-2 Inclusion Bodies Refolded BamA Δ 1-424

(A) The expression sequence of BamA Δ 1-424 (UniProt: P0A940) in vector pET28a(+) showing the inserted BamA Δ 1-424 sequence where amino acid residue 425 (Ser, S) is indicated in bold front. (B) Theoretical pI and molecular mass of BamA Δ 1-424 are shown*. (C) The diagram shows the protein corresponding to the designed construct inserted into OM, where the transmembrane β -barrel is in white and the POTRA domain is missing. The N-terminus hexahistidine tag is shown preceding the N-terminus of the transmembrane β -barrel. BamA Δ 1-424 is expressed in the cytosol as inclusion bodies since lacking signal sequence.

* The theoretical pI and MW were calculated using ProtParam (Gasteiger et al. 2005, 571-607)

nBamA Δ 1-20

(A)

MLYLLPTAAAGLLLLAAQPAMAMDIGINSDPNSSS**A**EGFVVKDIHFEGLRVAVGAALLSMPVRT
 GDTVNDEDISNTIRALFATGNFEDVRVLRDGD TLLVQVKERPTIASITFSGNKSVKDDMLKQNLEA
 SGVRVGESLDRTTIADIEKGLEDFYYSVGKYSASVKAVVTPLPRNRVDLKLVFQEGVSAEIQQINI
 VGNHAFTTDELISHFQLRDEVPWWNVVGD RKYQKQKLAGDLETLSYYLDRGYARFNIDSTQVS
 LTPDKKGIYVTVNITEGDQYKLSGVEVSGNLAGHSAEIEQLTKIEPGELYNGTKVTKMEDDIKLL
 GRYGYAYPRVQSMPEINDADKTVKLRVNV DAGRNFYVRKIRFEGNDTSKDAVLRREMRQMEGA
 WLGS DLVDQGKERLNRLGFFETVDTDTQRVPGSPDQVDVVYKVKERNTGSFNFGIGYGTESGV
 SFQAGVQQDNWLGTYAVGINGTKNDYQTYAELSVTNPYFTVDGVS LGGRLFYND FQADDADL
 SDYT NKSYGTDVTLGFPINEYNSLRAGLGYVHNSLSNMQPQVAMWRYLYSMGEHPSTSDQDN
 SFKTDDFTFNYGWTYNKLD RGYFPTDGSRVNLTGKVTIPGSDNEYKVTLDTATYVPIDDDHKW
 VVLGRTRWGYGDGLGGKEMPFYENFYAGGSSTVRGFQSNTIGPKAVYFPHQASNYDPDYDYE
 CATQDGAKDLCKSDDAVGGNAMAVASLEFITPTPFISDKYANSVRTSFFWDMGTVWDTNWDSS
 QYSGYPDYS DPSNIRMSAGIALQWMSPLGPLVFSYAQPFFKKYDGDKAEQFQFNIGKTWAAALG
 HHHHHH-STOP

(B)

Theoretical isoelectric point (pI): 4.98

Theoretical molecular mass: 90.9kDa

(C)

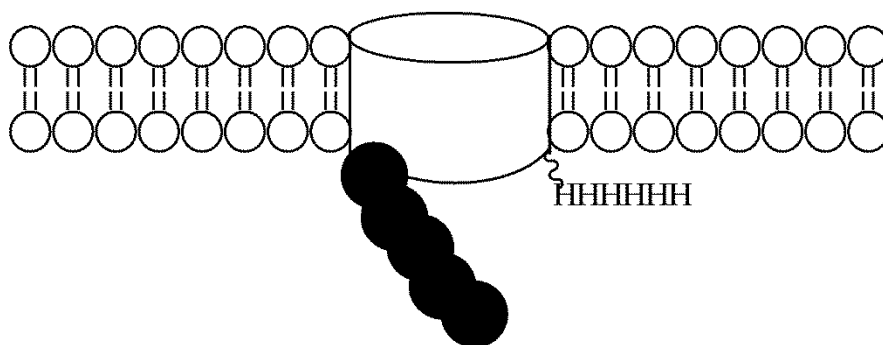


Figure 2-3 Native *E. coli* BamA Δ 1-20

(A) The expression sequence of native BamA Δ 1-20 (UniProt: P0A940) in vector pET20b(+) has inserted BamA Δ 1-20 sequence where amino acid residue 21 (Ala, A) is indicated in bold front. The pelB leader on pET 20b (+) is underlined, and pelB acts as a signal sequence which can be cleaved after the protein has been translocated across the IM. (B) Theoretical pI and molecular mass of nBamA Δ 1-20 are shown*. (C) The native BamA Δ 1-20 is expressed and inserted into the OM. The POTRA domains are in black circles and the transmembrane β -barrel is in white. The C-terminus hexahistidine tag is shown at the end of the β -barrel. At this point, the pelB leader has been cleaved already.

* The theoretical pI and MW were calculated using ProtParam (Gasteiger et al. 2005, 571-607)

2.2. Material and Methods

2.2.1. Full length *E.coli* BamA (*BamA* Δ 1-20)

2.2.1.1. Protein Over-expression

The BamA Δ 1-20 construct was available in Paetzel lab constructs library. The expression plasmid was transformed into *E.coli* BL21 (λ DE3) and cells were grown overnight at 37°C. This small scale overnight culture (15 mL) was used to inoculate two flasks of one liter each Luria Bertani (LB) medium containing kanamycin (50 μ g/ml). The large scale culture was grown until OD_{600nm} reached 0.6 at 37°C and then induced with 1mM isopropyl-1-thio- β -D-galactopyranoside (IPTG) for 3 hours, also at 37°C. The cells were harvested by centrifugation at 6000xg for 5 minutes at 4°C. The resulting cell pellet was stored at -80°C for future purification.

2.2.1.2. Isolation of Protein Inclusion Bodies

The cell pellets were resuspended in buffer A (20mM Tris-HCl pH 8.0, 100mM NaCl) and then sonicated (15"on, 15"off; 3 minutes in total) on ice until the sample was no longer viscous. The sample was further lysed using an Avestin Emulsiflex-3C cell. The BamA Δ 1-20 inclusion bodies were collected by centrifugation, 3444xg for 10 minutes, at 4°C. The white inclusion bodies were then washed twice with buffer A at 4C and recollected following centrifugation.

2.2.1.3. Solubilization and Denaturation of Inclusion Bodies

To be denatured and solubilized, the collected BamA Δ 1-20 inclusion bodies were resuspended in denaturing buffer (20mM Tris-HCl pH8.0, 100mM NaCl, 2mM DTT, 8M Urea) and stirred at room temperature overnight. The denatured sample solution was

then centrifuged, 29000xg for 40 minutes, at 4°C to remove any undissolved inclusion bodies. The supernatant, containing the denatured BamA Δ 1-20 protein, was stored in -80°C for further purification.

2.2.1.4. Denatured Protein Purified by Nickel Affinity Chromatography

The supernatant containing denatured BamA Δ 1-20 was incubated with 5ml nickel-NTA beads for 30 minutes at room temperature. The mixture was then loaded onto a clean column and the flow through was collected. Next, the nickel affinity column was washed with 50ml washing buffer (20mM imidazole in denaturing buffer), followed by elution of the BamA Δ 1-20 in two steps, using elution buffer. The elution buffer was 5ml in each of the two steps, with an increased imidazole concentration (elution1: 200mM imidazole in denaturing buffer, elution2: 300mM imidazole in denaturing buffer). The presence of BamA Δ 1-20 in the resulting elution fractions was confirmed by SDS-PAGE.

2.2.1.5. Protein Refolding by Rapid Dilution

The denatured protein was refolded by a 6x dilution using refolding buffer (20mM Tris-HCl pH8.0, 100mM NaCl, 2mM DTT, 0.2 M arginine, 0.5% TritonX-100). The elution fractions from nickel affinity chromatography were combined and added to the refolding buffer drop by drop, as the refolding solution was stirred rapidly at room temperature. Following the addition of the eluted denatured protein, the refolding solution was stirred for 2 hours. The refolded protein was then concentrated to a smaller volume using a stirred cell device (Millipore).

2.2.1.6. Purification by Ion Exchange Chromatography

The refolded BamA Δ 1-20 was loaded on a column packed with 5ml of Q-Sepharose (GE Healthcare). After the flow through was collected, the column was washed with 20ml washing buffer (20mM Tris-HCl pH8.0, 200mM NaCl, 2mM DTT). Next, the protein was eluted using an increasing concentration gradient of NaCl (400mM, 500mM, 600mM, 800mM, and 1M). The presence and purity of the BamA Δ 1-20 in the resulting elution fractions was confirmed by SDS-PAGE.

2.2.1.7. Purification by Analytical Size-Exclusion Chromatography

In order to further purify the refolded BamA Δ 1-20 and remove the detergent through buffer exchange, the refolded and concentrated BamA Δ 1-20 was loaded on a Superdex 200 column connected to an ÄKTA Prime system (GE Healthcare) and gel filtration buffer A (20mM Tris-HCl pH8.0, 100mM NaCl, 0.01% n-Dodecyl- β -D-maltoside (DDM), 2mM DTT) was pumped through at a flow rate of 1ml/min at room temperature. The resulting elution fractions were tested by 12% SDS-PAGE to confirm the presence and purity of BamA Δ 1-20, and then combined for crystallization or saved in -80°C for future analyses.

2.2.2. POTRA truncated *E. coli* BamA (BamA Δ 1-424)

2.2.2.1. Cloning

The DNA coding for BamA Δ 1-424 (425-810) was amplified by polymerase chain reaction (PCR) using *E. coli* K12 genomic DNA. Both forward and reverse primers were designed to contain corresponding restriction enzyme digestion sites, so that the PCR product could be ligated into pET28a (+) plasmid DNA digested with the same restriction enzymes. The subsequent plasmid was sent for sequencing to confirm the gene sequence was correct.

2.2.2.2. Protein Over-expression

The resulting plasmids were confirmed to contain the DNA sequence coding for BamA Δ 1-424, with an N-terminal hexahistidine tag. They were transformed into *E. coli* BL21 (λ DE3), grown overnight at 37°C. These cells (15 mL) were used to inoculate two flasks of one liter each LB medium containing kanamycin (50 μ g/ml). The large scale cell culture was grown until OD_{600nm} reached 0.6. at 37°C, followed by 1mM IPTG induction for 3 hours at the same temperature. The cells were harvested by centrifugation at 6000xg for 5 minutes at 4°C. The cell pellets were then stored at -80°C for future inclusion bodies isolation.

2.2.2.3. Isolation of Protein Inclusion Bodies

The BamA Δ 1-424 cell pellet was resuspended in 30ml buffer A (20mM Tris-HCl pH 8.0, 100mM NaCl) and then sonicated (15"on, 15"off; 3 minutes in total) on ice until the sample was no longer viscous. The sample was further lysed using an Avestin Emulsiflex-3C cell for 10 minutes. The BamA Δ 1-424 inclusion bodies were collected by centrifugation at 3444xg for 10 minutes at 4°C. The white inclusion bodies were washed twice with buffer A and recollected by centrifugation.

2.2.2.4. Solubilization and Denaturation of Inclusion Bodies

The BamA Δ 1-424 inclusion bodies, collected from 2 liter of LB medium, were resuspended in 30ml denaturing buffer (20mM Tris-HCl pH8.0, 100mM NaCl, 2mM DTT, 8M Urea) and stirred at room temperature overnight to ensure full denaturation. The denatured sample solution was then centrifuged, 29000xg for 40 minutes at 4°C, to remove any undissolved inclusion bodies. The supernatant, containing the denatured BamA Δ 1-20 protein, was saved at room temperature for further purification.

2.2.2.5. Denatured Protein Purified by Nickel Affinity Chromatography

The supernatant containing the denatured BamA Δ 1-424 was incubated with 5ml Ni-NTA beads for 30 minutes at room temperature. The mixture was then loaded onto a clean column. After the flow-through was collected, the nickel affinity column was washed with 50ml washing buffer (20mM imidazole in denaturing buffer) and eluted in two fractions. The elution buffer was 5ml in each of the two fractions, with an increased imidazole concentration (elution1: 300mM imidazole in denaturing buffer, elution2: 500mM Imidazole in denaturing buffer). The presence of BamA Δ 1-424 in the elution fractions was confirmed by 12% SDS-PAGE.

2.2.2.6. Protein Refolding by Rapid Dilution

The denatured protein was refolded by a 10 \times dilution using refolding buffer (20mM Tris-HCl pH8.0, 100mM NaCl, 2mM DTT, 0.2 M Arginine, 0.5% TritonX-100). The elution fractions from nickel affinity chromatography were added to the refolding buffer drop by drop, while the refolding solution was stirred rapidly at room temperature. Following the

addition of the eluted denatured protein, the refolding solution was stirred for 2 hours. The refolded protein was then concentrated to a smaller volume using a stirred cell device (Millipore).

2.2.2.7. Purification by Analytical Size-Exclusion Chromatography

In order to further purify the concentrated, refolded BamA Δ 1-424 protein and remove the detergent prior to crystallization, BamA Δ 1-424 was loaded on to a Superdex200 column connected to an ÄKTA Prime system (GE Healthcare). Gel filtration buffer A (20mM Tris-HCl pH8.0, 100mM NaCl, 0.01% DDM, 2mM DTT) was pumped through the system at a flow rate of 1ml/min at room temperature. The resulting elution fractions were tested by 12% SDS-PAGE to confirm the presence and purity of BamA Δ 1-20. The elution fractions of peak 2 and peak 3 were collected separately and then stored at -80°C for future use.

2.2.3. Native *E. coli* BamA (nBamA Δ 1-20)

2.2.3.1. Cloning

The DNA coding for nBamA Δ 1-20 (21-810) was amplified by polymerase chain reaction (PCR) using *E.coli* K12 genomic DNA. Both forward and reverse primers were designed to contain specific restriction enzyme digestion sites, so that the PCR products could ligate into pET20b (+) plasmid cut with the same restriction enzymes. The nBamA Δ 1-20 insert was inserted between the N-terminal signal sequence (pelB sequence) and the C-terminal hexahistidine tag of pET20b (+). The subsequent plasmids were sent for sequencing to confirm the correct gene sequence.

2.2.3.2. Protein Over-expression

The plasmids containing the DNA sequence coding for nBamA Δ 1-20 were transformed into *E.coli* BL21 (λ DE3) and grown overnight at 37°C. These cells (15 mL) were used to inoculate two flasks of one liter each LB medium containing ampicillin (50 μ g/ml). The large scale cell culture was grown for 4 hours at 20°C, followed by 1mM

IPTG induction overnight at the same temperature. The next day, the cells were harvested by centrifugation at 6000xg for 5 minutes at 4°C. The cell pellets were stored at -80°C for future nBamA Δ 1-20 extraction from the outer membranes of the cells.

2.2.3.3. Extracting nBamA Δ 1-20 from the OM

The nBamA Δ 1-20 cell pellets were resuspended in 30ml buffer B (50mM Tris-HCl pH7.5, 200mM NaCl, 1mM MgCl₂, 1 protease inhibitor cocktail tablet) and then sonicated (15"on, 15"off; 3 minutes in total) on ice. The sample was lysed using an Avestin Emulsiflex-3C cell for 10 minutes and the lysate was centrifuged at 12,000 xg for 10 minutes to remove any remaining unlysed cells.

To separate the OM from the protein, the supernatant was incubated with 2% TritonX-100 for 30 minutes at room temperature, and then centrifuged at 160,000 xg for 90 minutes at 4°C. The resulting pellets were resuspended in buffer C (50mM Tris-HCl pH7.5, 200mM NaCl, 20mM imidazole) and incubated with 1% lauryldimethylamine oxide (LDAO) at 4°C overnight. The following day, the solubilized membrane mixture was centrifuged at 200,000 xg for 60 minutes at 4°C. The supernatant containing soluble nBamA Δ 1-20 was kept at 4°C for future purification.

2.2.3.4. Protein Purification by Nickel Affinity Chromatography

The supernatant containing soluble nBamA Δ 1-20 was diluted 5 times with buffer C and incubated with 5ml Ni-NTA beads overnight at 4°C. The mixture was then loaded onto a clean column. After collecting the flow-through, the nickel affinity column was washed with 50ml washing buffer B (50mM Tris-HCl, pH7.5, 200mM NaCl, 20mM imidazole, 0.01% DDM,) and then eluted into two fractions. Three 5ml fractions of elution buffer were applied to the column sequentially. Each fraction contained an increasing imidazole concentration (Elution1: 50mM Tris-HCl, pH7.5, 200mM NaCl, 0.01% DDM, 2mM DTT, 200mM; Elution2: 50mM Tris-HCl, pH7.5, 200mM NaCl, 0.01% DDM, 2mM DTT, 300mM, Elution3: 50mM Tris-HCl, pH7.5, 200mM NaCl, 0.01% DDM, 2mM DTT, 500mM). The presence of BamA Δ 1-424 was confirmed in two of the elution fractions by 9% SDS-PAGE. These fractions were concentrated to 5ml using an Amicon ultra-centrifugal filter device (Millipore).

2.2.3.5. Purification by Analytical Size-Exclusion Chromatography

To further purify the protein, the concentrated nBamA Δ 1-20 was loaded on to a Superdex200 column connected to an ÄKTA Prime system (GE Healthcare). Gel filtration buffer B (50mM Tris-HCl pH7.5, 200mM NaCl, 0.01% DDM, 2mM DTT) was pumped through the system at a flow rate of 1ml/min at room temperature. The resulting elution fractions were tested by 9% SDS-PAGE to confirm the presence and purity of nBamA Δ 1-20. The elution fractions of peak 2 and peak 3 were collected separately and then stored at -80°C for future use.

2.3. Results

2.3.1. Full length *E. coli* BamA (BamA Δ 1-20)

Cells were lysed, and then subjected to centrifugation at 3444xg for 10 minutes, followed by two washes, to isolated BamA Δ 1-20 IB (inclusion bodies). The resulting IB pellet was denatured in 8M urea and then purified by Ni-NTA affinity chromatography. BamA Δ 1-20 protein was eluted from the column in two fractions. Since the theoretical molecular mass of BamA Δ 1-20 is 88 kDa (refer to Table 2-2), the results from the gel in Figure 2-4A suggested that the two elution fractions contained similar amounts of BamA Δ 1-20, along with other unwanted proteins. Both fractions were combined and refolded by rapid dilution. The results are shown in Figure 2-4B, which indicated that there were still other protein contaminants. Anion exchange chromatography was applied to further purify the refolded BamA Δ 1-20 as shown in Figure 2-4C. The target BamA Δ 1-20 was eluted in the flow through while most of the remaining impurities were eluted in the subsequent elution fractions.

The BamA Δ 1-20 protein, in the flow through, was then loaded on a size-exclusion column to separate the aggregates and exchange the detergent with DDM to remove TritonX-100. The chromatogram and corresponding SDS-PAGE gel are shown in Figure 2-5. The major peaks correspond to fractions 19, 23 and 26, having elution volumes of 45.87ml, 55.48ml and 64.28ml, respectively. Since the void volume is 40-45ml, the first

peak represents protein aggregates. Monomeric BamA Δ 1-20 has a theoretical molecular mass of 91 kDa. According to the standard curve shown in the Appendix B, the experimental molecular mass of the protein-detergent complex (BamA Δ 1-20-DDM) corresponding to peaks 2, 3 were calculated as 440 kDa and 150 kDa, respectively (Table 2-2). The Stoke radius was calculated assuming that the protein has a spherical shape; however, the presence of the POTRA domain which acts as a long tail, may complicate this calculation. The size-exclusion chromatogram of POTRA shows that the elution volume of monomeric POTRA proteins is larger than the volume corresponding to its theoretical molecular mass. Although the experimental molecular mass of the proteins in peak 3 was much larger than their theoretical mass, the BamA Δ 1-20 represented by this peak is believed to be the monomer. The BamA Δ 1-20 protein in peak 2 could be the trimeric form or having certain oligomeric state due to the presence of detergent.

Table 2-2 Gel-filtration Elution Volumes for BamA Δ 1-20

	Peak Elution Volume (mL)	Molecular Mass (kDa)
Elution Peak 1	45.87	void
Elution Peak 2	55.48	440
Elution Peak 3	64.28	150

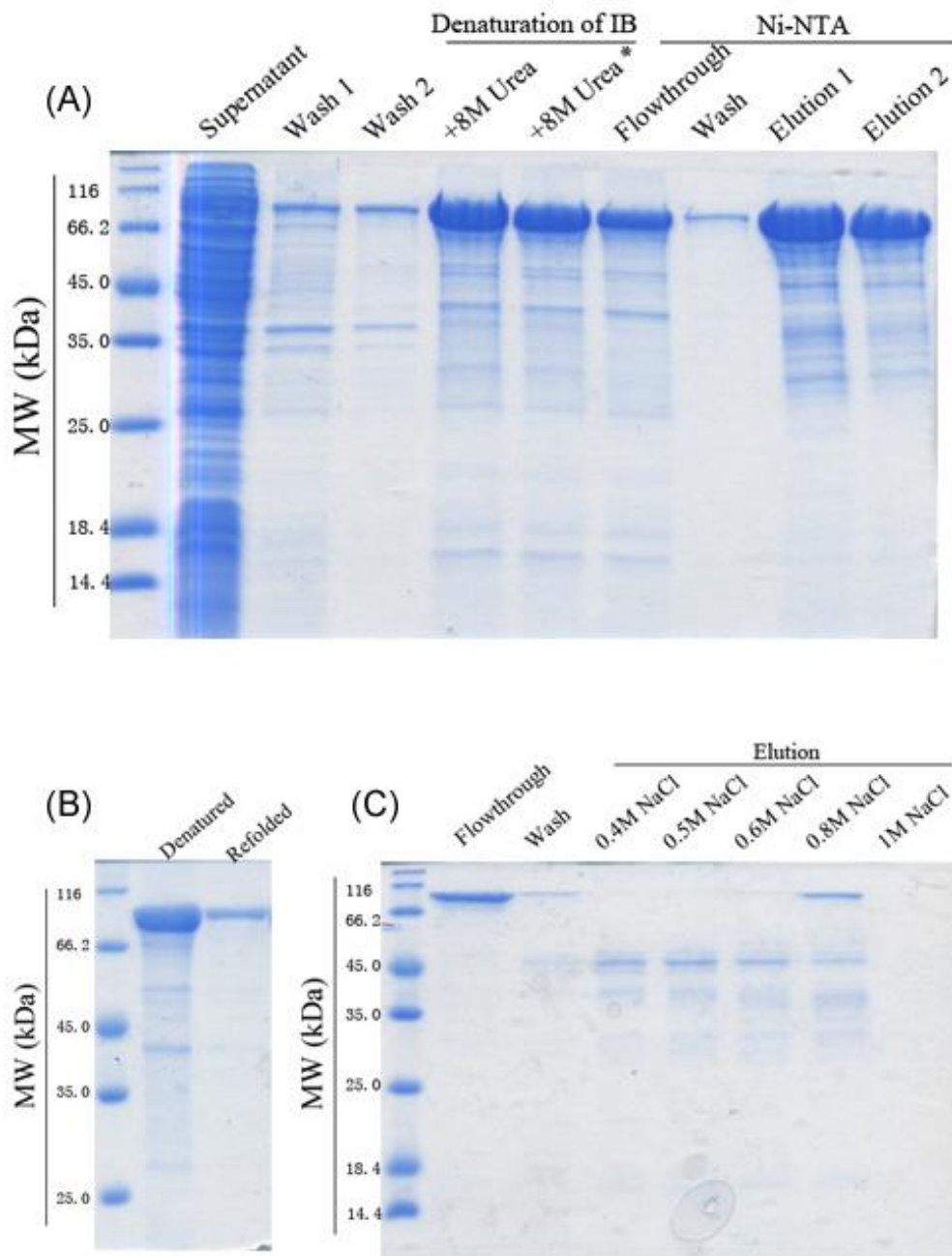


Figure 2-4 *E. coli* BamA Δ 1-20 Inclusion Bodies Isolation and Protein Purification

E. coli BamA Δ 1-20 IB were isolated, denatured, refolded and purified by Nickel affinity chromatography. The corresponding samples are shown on 15% SDS-PAGE. (A) The *E. coli* BamA Δ 1-20 IB were isolated and denatured in 8M urea. Under denaturing conditions, Nickel affinity chromatography was applied to purify the denatured protein. The proteins were eluted in two fractions, which contained 200mM and 300mM imidazole, respectively. (B) The elution fractions from Ni-NTA were combined and refolded by rapid dilution. (C) The refolded BamA Δ 1-20 was purified by anion exchange.

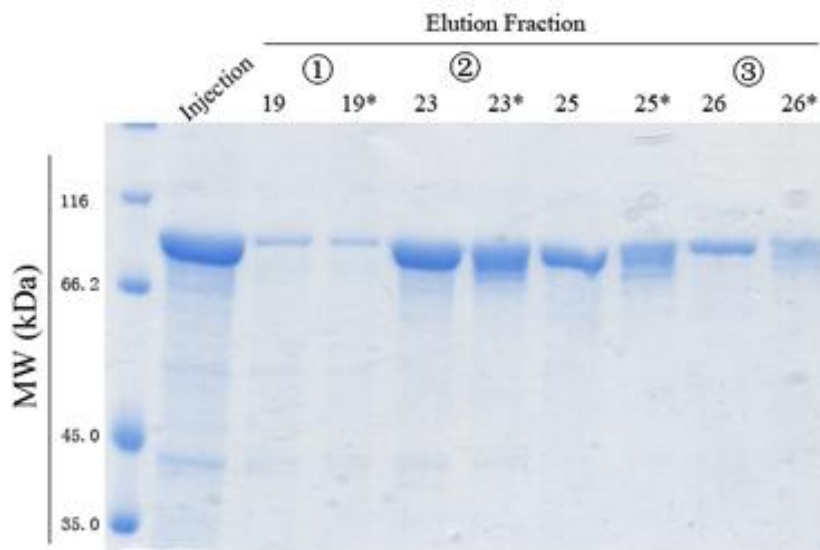
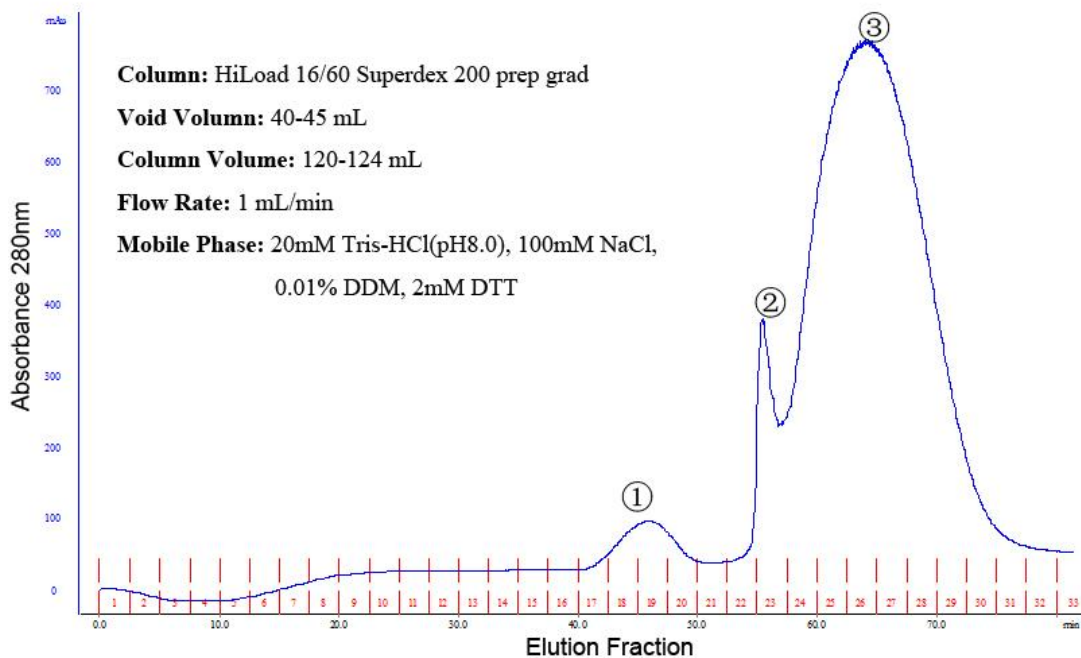


Figure 2-5 The Size-Exclusion Chromatogram for Refolded BamA Δ 1-20

Three BamA Δ 1-20 elution peaks are shown in above chromatogram. The corresponding elution fractions 19, 23, 25 and 26 were boiled at 100°C for 5 minutes before being loaded on a 9% SDS-PAGE gel. Duplicate samples, (19*, 23*, 25* and 26*) which were not boiled (kept at room temperature for 5 minutes) were loaded on the same 12% SDS-PAGE gel, alongside their boiled counterparts for comparison.

2.3.2. Transmembrane *E. coli* BamA (*BamA* Δ 1-424)

The cell lysate was centrifugated at 3444xg for 10 minutes to isolate BamA Δ 1-424 IB, followed by two washes. The resulting IB pellet was denatured in 8M urea. Samples were run on a 15% SDS-PAGE gel which is shown in Figure 2-6A. Under denaturing conditions, the isolated BamA Δ 1-424 protein was purified by Ni-NTA and eluted from the column in two fractions, with concentrations of 300mM and 500mM imidazole, respectively. As seen in the SDS-PAGE gel in Figure 2-3B, the major protein contained in those two fractions had a molecular mass of 45 kDa, close to the theoretical value of BamA Δ 1-424 (see Table 2-3). The protein sample from the first elution fraction contained less impurities than protein from the second elution fraction. Consequently, the protein from the first elution fraction only was refolded by rapid dilution. The gel is shown in Figure 2-6C.

The refolded BamA Δ 1-424 was then loaded on a size-exclusion column to separate the aggregates and exchange the detergent with DDM to get rid of TritonX-100. The chromatogram and corresponding SDS-PAGE gel are shown in Figure 2-4. As no peak appeared in the void volume region at 40ml-45ml, the refolded BamA Δ 1-424 did not contain aggregates. Peaks 1 and 2 had elution volumes of 67.5 mL and 80 mL, respectively and the corresponding molecular mass were calculated as 150 kDa and 48.6 kDa. Since the theoretical molecular mass of monomeric BamA Δ 1-424 is 45 kDa, peak 3 likely represents the monomeric BamA Δ 1-424 while Peak 2 represents the trimeric BamA Δ 1-424.

Table 2-3 Gel-filtration Elution Volumes for *BamA* Δ 1-424

	Peak Elution Volume (mL)	Molecular Mass (kDa)
Elution Peak 1	67.5	150
Elution Peak 2	80	48.6

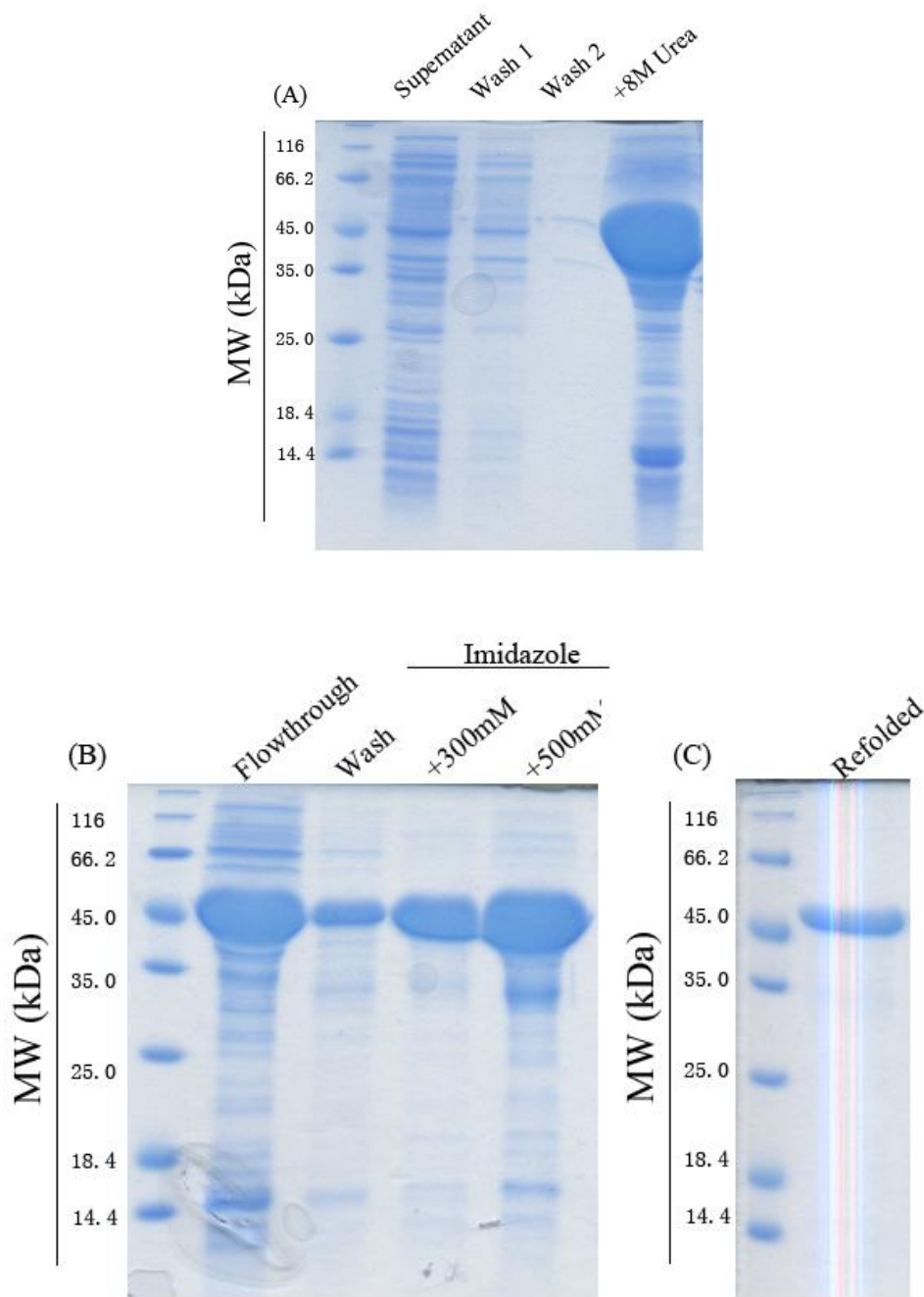


Figure 2-6 *E. coli* BamA Δ 1-424 Inclusion Bodies Isolation and Protein Purification

E. coli BamA Δ 1-424 IB were isolated, denatured, refolded and purified by Nickel affinity chromatography. The corresponding samples are shown on the 15% SDS-PAGE gel. (A) The BamA Δ 1-424 IB were isolated and denatured in 8M urea. (B) By using Ni-NTA under denaturing conditions, BamA Δ 1-424 protein was eluted in two fractions, containing imidazole concentrations of 300mM, and 500mM, respectively. (C) The first elution fraction from the Ni-NTA column was refolded by rapid dilution.

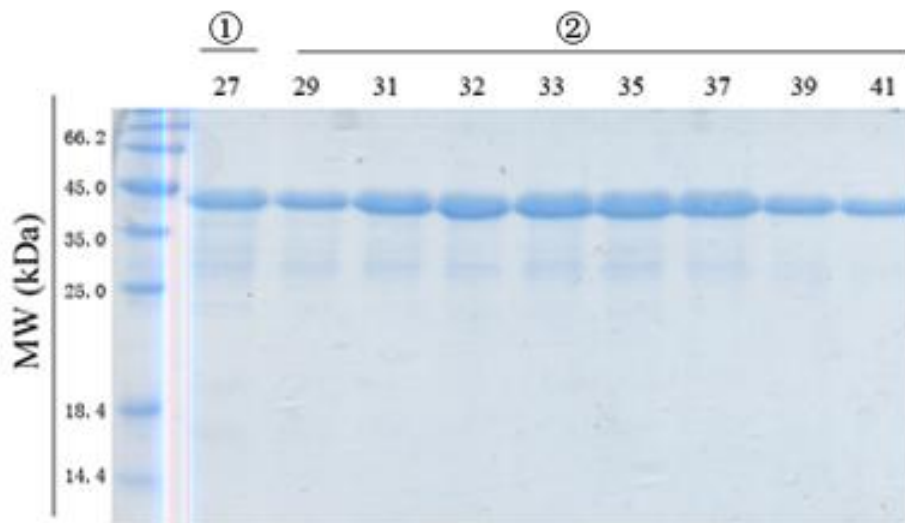
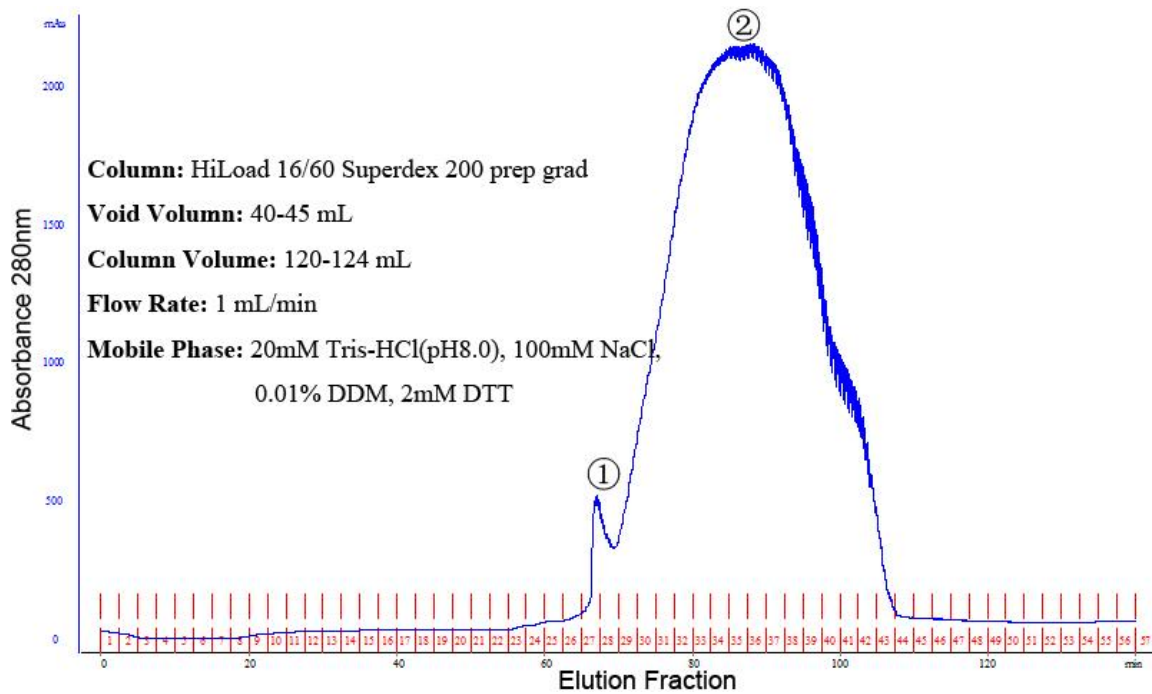


Figure 2-7 The Size-Exclusion Chromatogram for Refolded BamA Δ 1-424

Two BamA Δ 1-424 elution peaks are shown in the above chromatogram. Protein samples from the elution fractions, 27,29,31,32,33,35,37,39 and 41 that are included in those two peaks are labeled on the accompanying 15% SDS-PAGE gel. Fraction 27 corresponds to peak 1, while the remainder of the fractions loaded on gel corresponds to peak2.

2.3.3. Native *E. coli* BamA (nBamA Δ 1-20)

After cell lysis, the lysate was spun down and both soluble and insoluble nBamA Δ 1-20 fractions were obtained. As shown in Figure 2-8A, lane 1, the supernatant contained the soluble fraction of nBamA Δ 1-20 while, as seen in the same figure, lane 2, a significant amount of nBamA Δ 1-20 was also found in the insoluble pellet. To isolate soluble nBamA Δ 1-20 from the membrane, the supernatant was incubated with 2% TritonX-100 for 2 hours at room temperature. Then, the mixture was centrifuged to pellet the membrane fraction and the membrane pellets were solubilized, using either n-octyl- β -D-glucoside (OG) or LDAO. Following centrifugation, the resulting supernatants and pellets were run on SDS-PAGE to verify the presence of nBamA Δ 1-20. In Figure 2-8, most of the isolated nBamA Δ 1-20 protein appeared in the pellets (lanes 4 and 5) which indicated that the membrane was not solubilized enough to extract nBamA Δ 1-20, from the OM using OG. However, LDAO extraction was successful, as indicated by a strong band of nBamA Δ 1-20 in the supernatant (lane 7 in Figure 2-8A). The amount of nBamA Δ 1-20 extracted was sufficient for Ni-NTA purification and further experiments.

To avoid interference with Ni-NTA binding caused by large concentrations of detergent in the buffer, the supernatant fraction from the LDAO extraction was diluted 5 times with buffer C and then incubated with 5 ml of Ni-NTA resin overnight at 4°C. After washing, nBamA Δ 1-20 was eluted in three fractions with very high levels of purity. The results are shown in Figure 2-8B.

The three Ni-NTA elution fractions were combined and further purified by size-exclusion chromatography. The chromatogram and corresponding SDS-PAGE gel are shown in Figure 2-9. Peak 1 suggests that the aggregated nBamA Δ 1-20 was eluted in the void volume. The elution volumes of peaks 2 and 3 were similar to the results of the refolded BamA Δ 1-20 elution profile, suggesting that oligomeric nBamA Δ 1-20 is represented in peak 2 and monomeric nBamA Δ 1-20 in peak 1.

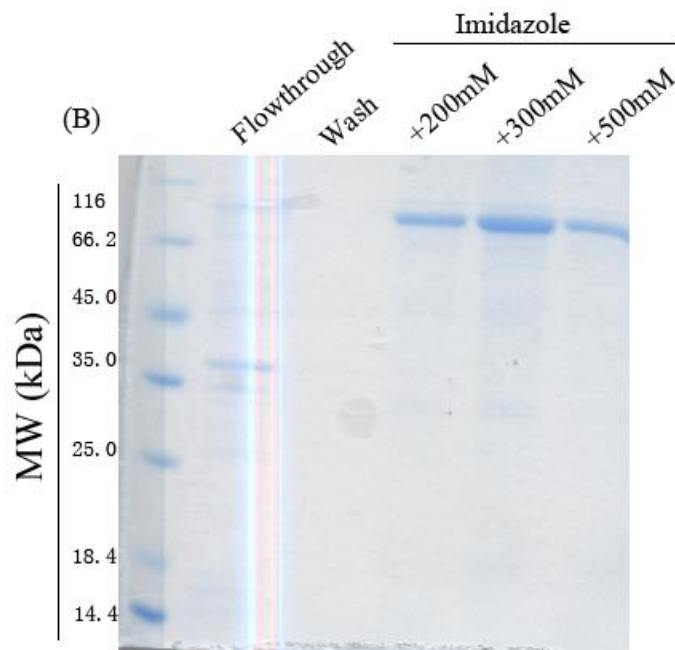
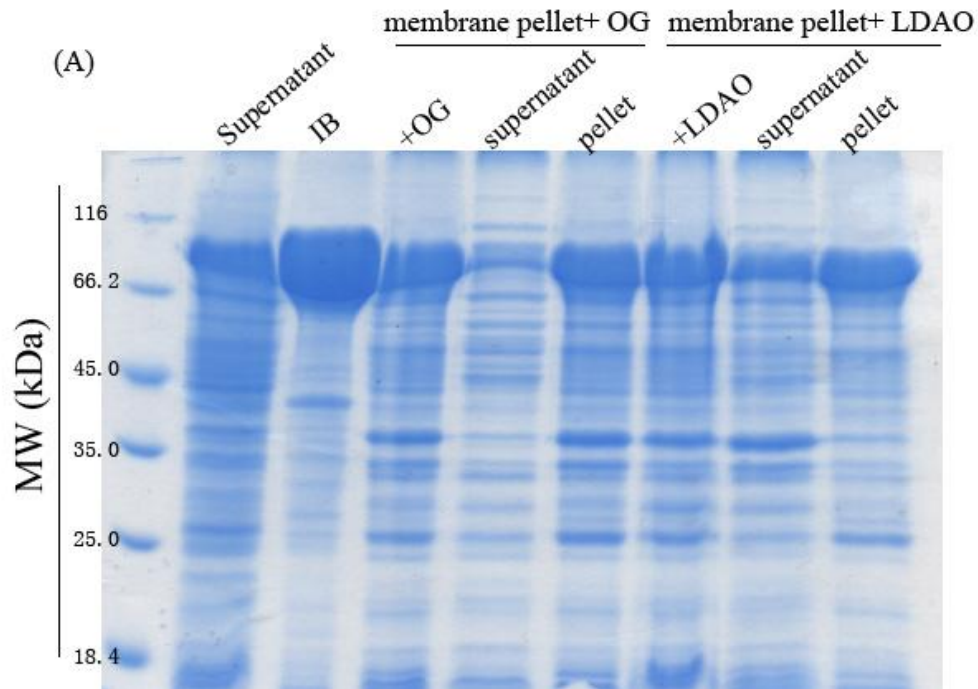


Figure 2-8 *E. coli* nBamA Δ 1-20 Extraction from the OM and Purification

E. coli nBamA Δ 1-20 was extracted from the OM and purified by Nickel affinity chromatography. The corresponding samples are shown on the SDS-PAGE gel above. (A) The cells were lysed and centrifuged to separate the soluble and insoluble (inclusion bodies) nBamA Δ 1-20, as shown in Lanes 1 and 2. Next, supernatant containing the

soluble nBamA Δ 1-20, together with the membranes, was treated with Triton X-100 and centrifuged. The resulting pellet was resuspended and solubilized using either OG (lane 3) or LDAO (lane 6). The protein from the OG and LDAO supernatants are shown in lanes 4, and 7, respectively while the proteins from the pellets are shown in lanes 5 and 8. (B) Protein from the LDAO supernatant was purified using Ni-NTA chromatography. The purified nBamA Δ 1-20 protein was eluted in three fractions, containing 200mM, 300mM and 500mM imidazole, respectively, as shown in Lanes 3,4, and 5.

Table 2-4 Gel-filtration elution volumes for nBamA Δ 1-20

	Peak Elution Volume (mL)	Molecular Mass (kDa)
Elution Peak 1	44.8	void
Elution Peak 2	54.3	440
Elution Peak 3	65	150

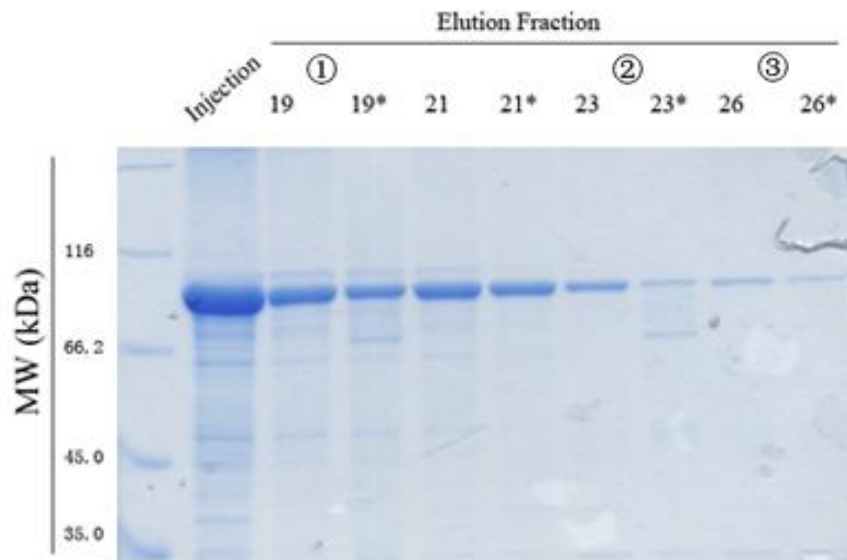
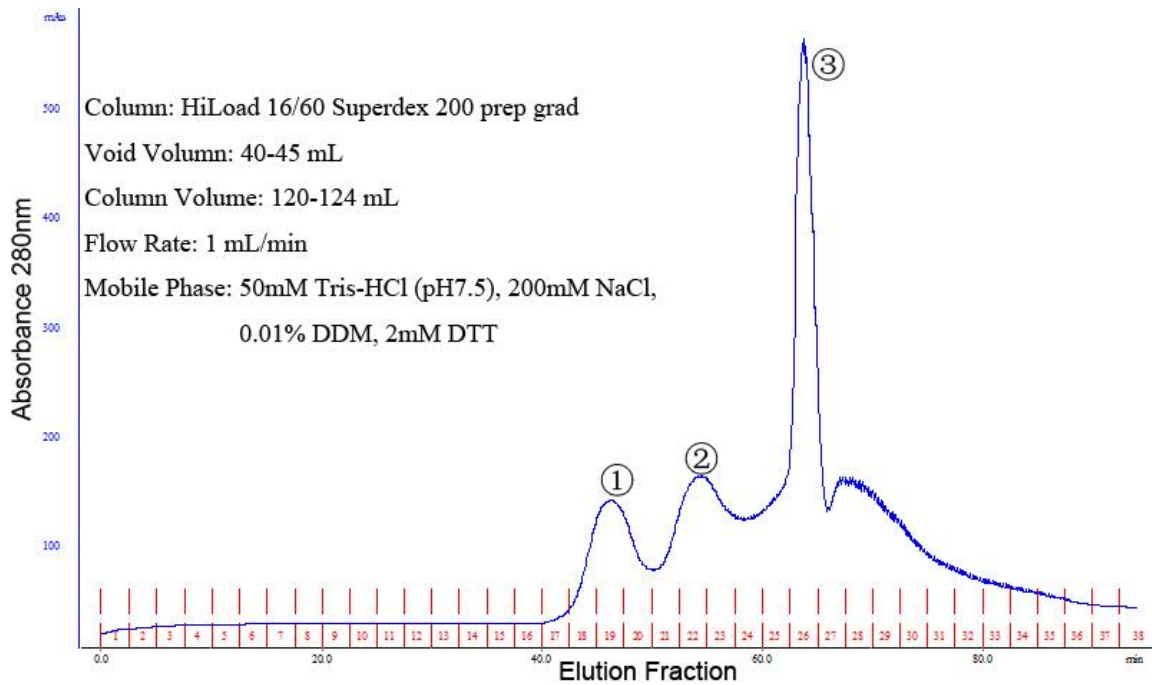


Figure 2-9 The Size-Exclusion Chromatogram for *nBamA* Δ 1-20

Three major *nBamA* Δ 1-20 elution peaks are shown in the chromatogram. Elution fractions 19, 21, 23 and 26 from the size exclusion column were run on 9% SDS-PAGE, after having been boiled at 100°C for 5 minutes. Duplicate samples (19*, 21*, 23* and 26*), were not boiled prior to being loaded on the gel and were kept at room temperature for 5 minutes without boiling. These control samples were run alongside the treated samples for comparison.

2.4. Discussion

This chapter presented the experiments carried out to obtain purified nBamA Δ 1-20, BamA Δ 1-424 and BamA Δ 1-20. While the structures of *E. coli* BamA POTRA domain and transmembrane β -barrel have been determined separately, our ultimate research goal is to gain in depth structural insight of the complete full length structure; therefore, the BamA Δ 1-20 construct was created for that purpose. Together with the full length construct, BamA Δ 1-20, already available in our lab library, *E. coli* BamA Δ 1-424 and BamA Δ 1-20 were over-expressed as inclusion bodies. To design a protocol for inclusion bodies isolation, effective denaturation and protein refolding procedures are essential in this thesis. Furthermore, native BamA Δ 1-20 is designed for the purpose that nBamA Δ 1-20 is assembled in *E. coli* OM; hence native proteins are extracted from the OM.

In vivo, proteins tend to aggregate in non-native oligomeric complexes due to the non-native interactions among structured and kinetically trapped protein during folding and assembly. They are usually insoluble and metabolically stable. In humans, similar non-native complexes have been linked to neuronal degeneration diseases (Kopito 2000, 524-530; Fink 1998) . Early studies suggest that aggregation may depend both on the rate of protein folding and expression. *In vitro*, unfolded polypeptide chains in high concentration tend to aggregate during the over-expression of recombinant proteins in prokaryotic and eukaryotic hosts. The end product of aggregation, a complex of dense, insoluble protein particles, is called inclusion (Kiefhaber et al. 1991) .

BamA is an OMP that is initially synthesized in the cytosol. Its signal peptide (residues 1-20) leads the pre-OMP to the IM where it is recognized and moved across the IM by Sec translocons. After that, the signal peptide is cleaved by SPase I at the Ala-X-Ala site, releasing the pre-OMP into the periplasmic space. BamA is then folded and assembled into the OM as a native BamA. The protein from the construct, nBamA Δ 1-20 can be extracted from the OM in a native form. However, for the BamA Δ 1-424 and BamA Δ 1-20 constructs, whose signal sequences are not included their over-expressed proteins remain in their unfolded forms in the cytosol. This high concentration of unfolded BamA proteins will trigger aggregation formation, and as a result, BamA Δ 1-424 and BamA Δ 1-20 proteins are expressed in inclusion bodies.

Several steps will affect protein refolding after isolation from inclusion bodies, including solubilization of the protein by denaturants, the subsequent removal of those denaturants and detergents to assist refolding (Tsumoto et al. 2003) . In this chapter, BamA Δ 1-424 and BamA Δ 1-20 inclusion bodies are denatured in 8M urea. The denatured proteins are then refolded by rapid dilution. Each drop of denatured protein in urea is added into the refolding buffer which includes TritonX-100, DTT and L-arginine. In this way, urea in the first drop is assumed to be diluted to a point where its presence does not lead to denaturation of the protein but does prevent the protein from aggregation; hence refolding begins. At the same time, TritonX-100 assists the refolding protein by shielding the protein's hydrophobic region from the hydrophobic regions of other protein molecules, also preventing aggregation. Since there are two cysteines in each construct, DTT is utilised to avoid disulfide bond formation which may cause protein aggregation. By the end of the refolding step, the final concentration of urea in the protein solution is ~1.3M. The size-exclusion chromatography results prove that the majority of the refolded proteins are non-aggregates. The proteins present in void peak have molecular mass larger than 6000 kDa and they maybe large oligomers or protein aggregates.

BamA Δ 1-424 and BamA Δ 1-20, nBamA Δ 1-20 extraction and purification is another important component in the quest to better understand the BAM complex. The nBamA Δ 1-20's membrane protein is more likely to be expressed in very low levels or form inclusion bodies, as a result of toxic effects or degradation (Tsumoto et al. 2003) . The SDS-PAGE gel shows the majority of the insoluble nBamA Δ 1-20 protein expressed is likely to be inclusion bodies; while small amount of native nBamA Δ 1-20 protein is inserted into the OM. Membrane protein extraction from the OM requires detergent to solubilise the membrane pellet, thus releasing the inserted protein into the supernatant. Compared to OG, LDAO has been proven to be a more suitable detergent in this case. Once nBamA Δ 1-20 protein has been successfully extracted, purification steps followed. The size-exclusion chromatography results show that nBamA Δ 1-20 and BamA Δ 1-20 proteins share a similar elution profile.

The purified proteins from size exclusive chromatography are collected and then concentrated to 10mg/mL. Before setting crystallization trials, the concentrated BamA Δ 1-424 or BamA Δ 1-20 are tested on SDS-PAGE for purity. Overall, the refolded BamA Δ 1-

424, BamA Δ 1-20 proteins and extracted nBamA Δ 1-20 proteins can be obtained in high purity. The size-exclusion chromatography step effectively separates out most of the protein aggregates or large oligomers that are produced during refolding or purification. The non-aggregated proteins are present in more than one peak, which suggests the proteins may be in some oligomeric states. However, there is one problem. Since the elution peaks containing the putative monomeric proteins are very broad, it is very likely that there are more than one population of proteins present. This might be the result of using detergent, where small protein peptides that have been degraded during purification are trapped together by the detergent. One piece of evidence supporting this conjecture is that a much lower amount of protein is seen on SDS-PAGE for one particular elution fraction, compared to the absorbance of that fraction which is significantly higher than expected. The homogeneity of each protein elution fraction can be further investigated using light scattering spectroscopy or mass spectroscopy to obtain their accurate molecular mass (protein + detergent). In addition, cross-linking and mutational analysis can be performed to address the oligomeric state of BamA Δ 1-20 in solution, which may provide a clue to its physiological oligomerization on the OM.

Chapter 3. Protein Refolding Status Analysis

3.1. Introduction

Large quantities of pure protein homogenates are required before crystallizing each of the *E. coli* BamA constructs. Over-expressing BamA Δ 1-20 and BamA Δ 1-424 yield more crude protein than over-expressing nBamA Δ 1-20. Since denaturing inclusion bodies is one of the steps involved in purifying BamA Δ 1-20 and BamA Δ 1-424, the refolded protein must be properly folded and monodispersed to be a good candidate for crystallization.

This chapter will describe the four different approaches that were used to analyze the refolding status of *E. coli* BamA. The first approach was to separate the refolded protein aggregates by size-exclusion chromatography and estimate the Stokes radius of the refolded BamA from the elution fraction. The second approach was to use a heat-modifiability test to prove the existence of the β -barrel structure. A β -barrel structure requires a high temperature, based on the specific properties of the OMP, to be fully denatured in the presence of SDS and β -mecaptoethanol (Wilson 1991) . Therefore, if a β -barrel is present, a band shift will be observed on an SDS-PAGE gel when compared to the same refolded BamA sample treated at a different temperature (22°C vs. 100°C).The third approach was based on the fact that secondary structures can be characterized by circular dichroism (CD) spectroscopy. And in the fourth approach, digestion of refolded BamA Δ 1-20 and BamA Δ 1-424 with trypsin or chymotrypsin was used to investigate their limited proteolysis digestion patterns.

The above four approaches were used to confirm and characterize properly refolded proteins. The results from refolded BamA Δ 1-20 and BamA Δ 1-424 were then compared to those of nBamA Δ 1-20.

3.2. Material and Methods

3.2.1. Size-exclusion Chromatography

Each of the 5 mL protein samples were injected on to a Superdex 200 column connected to an ÄKTA Prime system (GE Healthcare) at a flow rate of 1mL/min at room temperature. The mobile phase was Tris-HCl buffer containing 0.01% DDM and 2 mM DTT (BamA Δ 1-20 and BamA Δ 1-424: 20 mM Tris-HCl pH8.0, 100 mM NaCl, 0.01% DDM, 2 mM DTT; nBamA Δ 1-20: 50 mM Tris-HCl pH7.5, 200mM NaCl, 0.01% DDM, 2mM DTT). The elution fractions were run on SDS-PAGE to verify the presence of the target proteins.

3.2.2. Heat Modifiability Test

. For the heat modifiability test, each protein sample (0.1 mg/mL) was treated with β -mercaptoethanol (5% final concentration) and boiled at 100°C for 5 minutes. Meanwhile, duplicate samples, also treated with β -mercaptoethanol (5% final concentration), were left at room temperature (22°C) for 5 minutes instead of being boiled. Both sets of samples were then run on a 9% SDS-PAGE gel to observe the differences in band migration.

3.2.3. Circular Dichroism (CD) Spectroscopy

Before carrying out CD spectroscopy, the Tris-HCl buffer in the protein samples (described in Section 3.2.1) was exchanged for CD buffer (potassium phosphate pH8.0, 0.01% DDM). Each 150 μ L protein sample was placed in a quartz cuvette, having an optical path length of 0.05 mm. Each set of readings were taken in triplicate at room temperature (22°C) on a JASCO J-810 CD spectrometer.

3.2.4. Limited Proteolysis

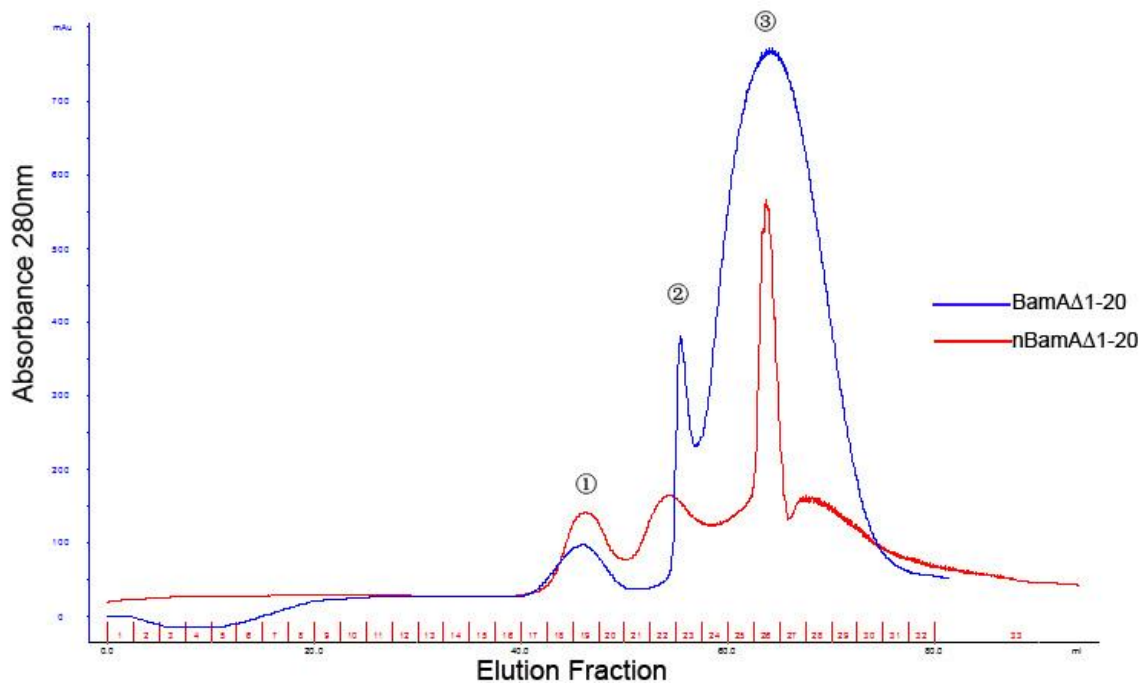
Each protein sample (1 mg/mL) was digested with trypsin or chymotrypsin (0.1 mg/mL) with the mass ratio (protein: protease) of 1000: 1 at room temperature. Each digestion reaction was stopped by adding the protein-protease mixture to loading dye containing β -mecaptoethanol (5% final concentration). Samples were taken at 0, 5, 10, 15, 30 minutes; 1, 2, 3 hours; and overnight. They were run on 9% or 15% SDS-PAGE to observe the digestion patterns.

3.3. Results

3.3.1. Size-exclusion Chromatography

The purified nBamA Δ 1-20 and refolded BamA Δ 1-20 samples were loaded onto a HiLoad 16/60 Superdex 200 preparatory grade column with a flow rate of 1 mL/min. The column volume was between 120 and 124 mL, with a separation range of 10 to 600 kDa. The resulting chromatography curves were overlapped as shown in Figure 3-1. The elution peaks of refolded BamA Δ 1-20 had similar elution volumes, compared to the peaks of nBamA Δ 1-20. Both curves showed three elution peaks with elution volumes around 45, 55 and 64mL, as labelled on the chromatogram. The results suggest that refolded BamA Δ 1-20 behaves the same as nBamA Δ 1-20 in terms of their stokes radii and oligomeric states.

Since the void volume was from 40 to 45mL, Peak 1 in the chromatogram is believed to represent the aggregated proteins, showing that size-exclusion chromatography can separate aggregates from both nBamA Δ 1-20 and refolded BamA Δ 1-20 protein samples. Using the standard curve (Appendix B) and the elution volumes of Peaks 2 and 3, the molecular masses of the protein-detergent complexes were calculated to be ~ 440 and ~150 kDa, respectively.



Column: HiLoad 16/60 Superdex 200 prep grad

Void Volume: 40-45 mL

Column Volume: 120-124 mL

Flow Rate: 1 mL/min

Mobile Phase:

BamA Δ 1-20 : 20mM Tris-HCl (pH8.0, 100mM NaCl, 0.01% DDM, 2mM DTT

nBamA Δ 1-20 : 50mM Tris-HCl (pH7.5, 200mM NaCl, 0.01% DDM, 2mM DTT

Figure 3-1 Superposed *E. coli* nBamA Δ 1-20 and Refolded BamA Δ 1-20 Size-Exclusion Chromatograms

The blue line represents the nBamA Δ 1-20 size-exclusion chromatography curve while the red line represents the refolded BamA Δ 1-20 size-exclusion chromatography curve. Both protein samples were loaded separately on the same column with a flow rate of 1 mL/min. The peaks eluted from the column are labeled peaks 1, 2, 3 and had elution volumes around 45, 55 and 64 mL, respectively.

The stoke radius of a protein based on the elution volume is accurate only when the protein is assumed to be globular in shape. However, BamA Δ 1-20 is known to be composed of a transmembrane β -barrel domain and a POTRA domain which has an elongated shape, as revealed in its crystal structure. This elongated tail may interfere with the migration of the protein in a column so that the elution peak is shifted forward on the chromatogram. In this case, the experimental molecular masses of refolded BamA Δ 1-20 or nBamA Δ 1-20 are larger than their theoretical values. Peak 3 may represent the monomer of either BamA Δ 1-20 or nBamA Δ 1-20, considering BamA's theoretical molecular mass is 88 kDa. Peak 2 may represent the oligomeric states of BamA Δ 1-20 and nBamA Δ 1-20. To address the accurate oligomerization of BamA Δ 1-20 or nBamA Δ 1-20, further investigations are required such as using cross-linking analysis and analytical ultracentrifugation.

3.3.2. Heat-modifiability Test

In Figure 3-2 (A), the nBamA Δ 1-20 sample that was boiled prior to being loaded on a 9% SDS-PAGE gel yielded a protein band located around the 90 kDa position. The protein migrated to the 66.2 kDa position when the nBamA Δ 1-20 protein sample was left at room temperature rather than being boiled prior to loading. The same band shift was observed for refolded BamA Δ 1-20, as shown in Figure 3-2 (B). This suggests that the BamA Δ 1-20 protein formed a β -barrel structure after being refolded.

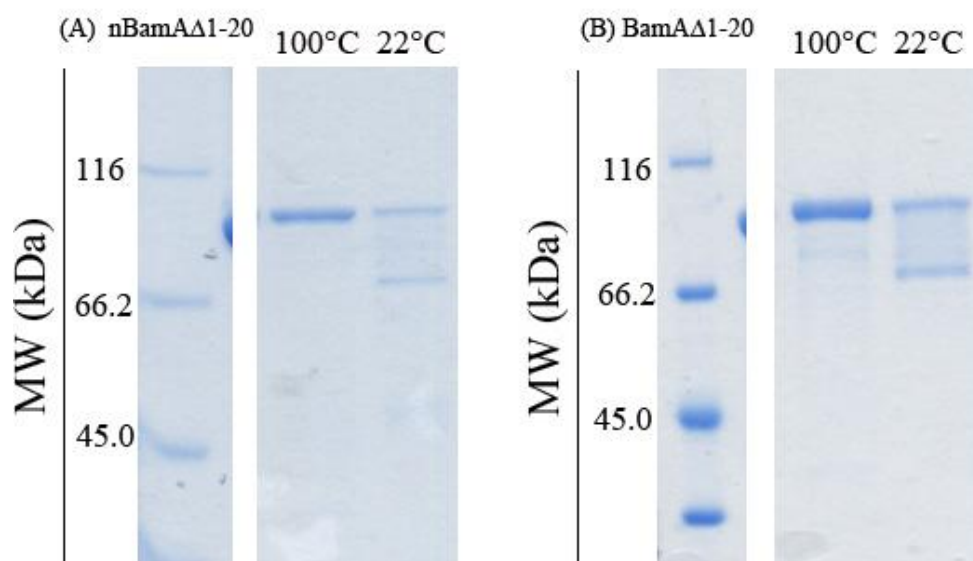


Figure 3-2 Heat-modifiability Shift of *E. coli* nBamA Δ 1-20 and BamA Δ 1-20 as seen on 9% SDS-PAGE Gels

(A) nBamA Δ 1-20 was either boiled (100°C) or left at room temperature (22°C) for 5 minutes prior to being loaded on the gel. (B) BamA Δ 1-20 was either boiled (100°C) or left at room temperature (22°C) for 5 minutes prior to being loaded on the gel.

3.3.3. Circular Dichroism (CD) Spectroscopy

Before studying the secondary structures of nBamA Δ 1-20 and refolded BamA Δ 1-20 using CD spectroscopy, the buffer the protein samples were suspended in was exchanged to avoid strong buffer absorption in the far UV wavelengths (ranging from 190 to 260nm) that CD spectroscopy requires. The purified protein samples were originally in a Tris-HCl buffer (described in Section 3.2.1) which absorbs at the lower wavelengths (~200nm). This buffer was exchanged for a potassium phosphate buffer (potassium phosphate pH8.0, 0.01% DDM) which absorbs around the 185nm wavelength, minimizing the interference in the far UV readings (Greenfield 2007).

The potassium phosphate buffer was measured as a blank and showed low absorbance (< 1 millidegree) between 190nm to 260nm wavelengths (Figure 3-4). The CD spectra were generated by both subtracting the buffer absorbance baseline and

curve smoothing. The far UV measurements were taken between 190 to 260 nm wavelengths and the resulting raw CD spectra were converted to molar ellipticity (Mol. Ellip) in millidegree·cm²dmol⁻¹, as shown in Figure 3-6, to reveal secondary structure. These experiments were done in triplicate to show that those results are reproducible. Two other sets of replicated experimental data are included in Appendix C.

Different types of secondary structures give rise to characteristic CD spectra in the far UV, as shown in Figure 3-3 (Greenfield 2007) . Both the spectra for nBamA Δ 1-20 and BamA Δ 1-20, behaved similarly in terms of Mol. Ellip versus wavelength. In Figures 3-6 (A) and (B), both Mol. Ellip spectra showed broad and weak absorbances (~10000 millidegree·cm²dmol⁻¹) around 210 to 230 nm, with sharp and intense absorbances (~6000 millidegree·cm²dmol⁻¹) in the lower wavelength range (<200 nm). Based on the standard CD spectra shown in Figure 3-3, the spectra of nBamA Δ 1-20 suggests the presence of antiparallel β -strands consistent with BamA structure prediction based on the crystal structure of FhaC, another member of the OMP85 super family (Clantin et al. 2007b). Refolded BamA Δ 1-20 showed similar CD spectra, suggesting it refolds and forms secondary structure much like nBamA Δ 1-20. If refolding had failed, the BamA Δ 1-20 CD spectra was expected to show the same disordered characteristics as seen in the green line in Figure 3-3.

The total absorbance of the sample (protein and buffer) should be within reasonable bounds. This can be monitored by High Tension Voltage (HTV) trace. To avoid excessive noise and demonstrate reliable absorbance data, the HTV trace should be less than 700V (Kelly, Jess, and Price 2005) . In our experiment, the CD absorbances of the refolded BamA Δ 1-20 sample are reliable from wavelengths 190 to 260nm. However, the HTV trace showed higher than 700V for the nBamA Δ 1-20 sample from wavelengths 190 to 195nm, and the readings from 195nm to 260nm are acceptable and show the dominant secondary structure of nBamA Δ 1-20.

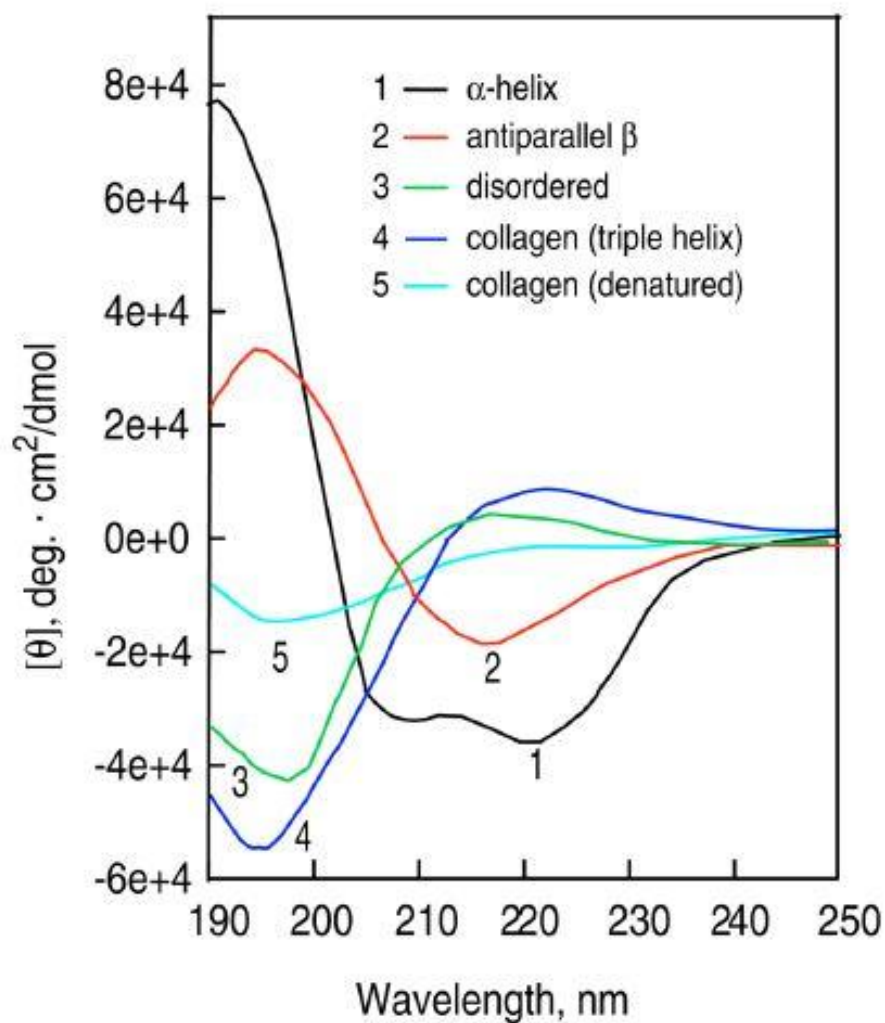


Figure 3-3 Far UV CD Spectra Characteristic of Common Secondary Structures

The black line shows the characteristic shape of a α -helix structure; the red line represents an antiparallel β strand; the green line, a disordered structure; the blue line, collagen in its native form; and the cyan line represents collagen in its denatured form (Greenfield 2007).

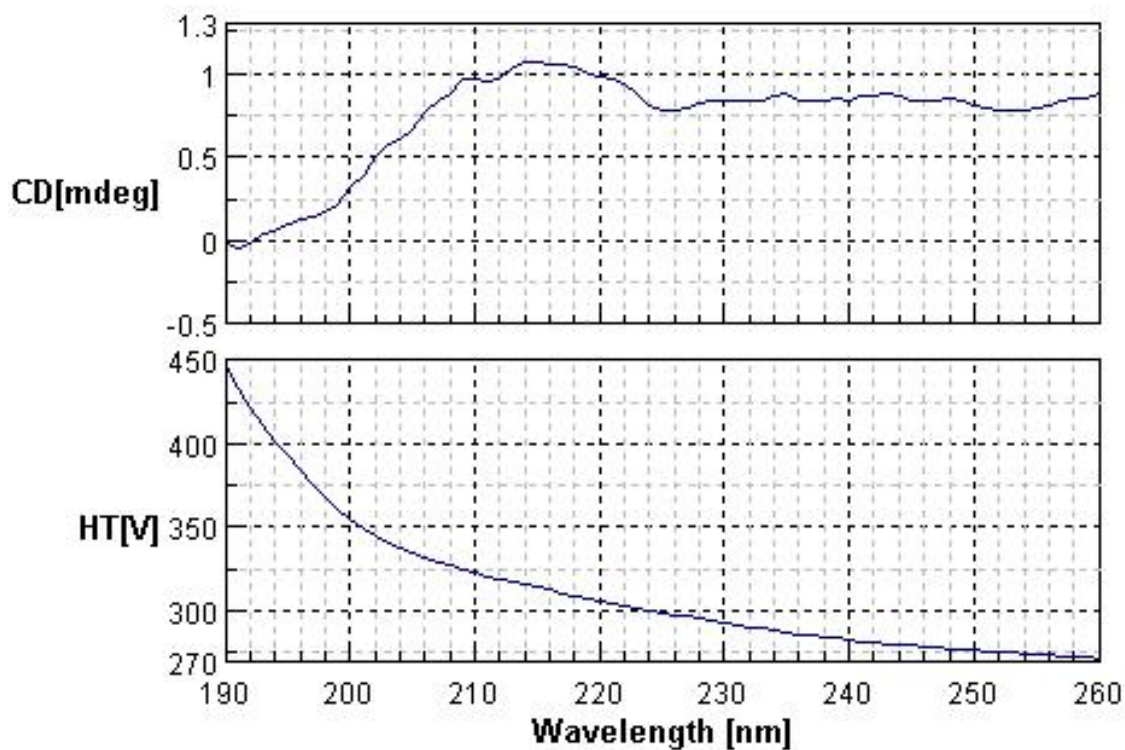


Figure 3-4 CD Spectrum of Buffer Used in the Experiment

Potassium phosphate buffer (potassium phosphate pH8.0, 0.01% DDM) was used in the CD spectroscopy experiments instead of the original Tris-HCl buffer. Very low absorbance, smaller than 1 millidegree, was observed from wavelength of 190 to 260nm was scanned. The high tension voltages are lower than 700 V within this wavelength range of wavelengths.

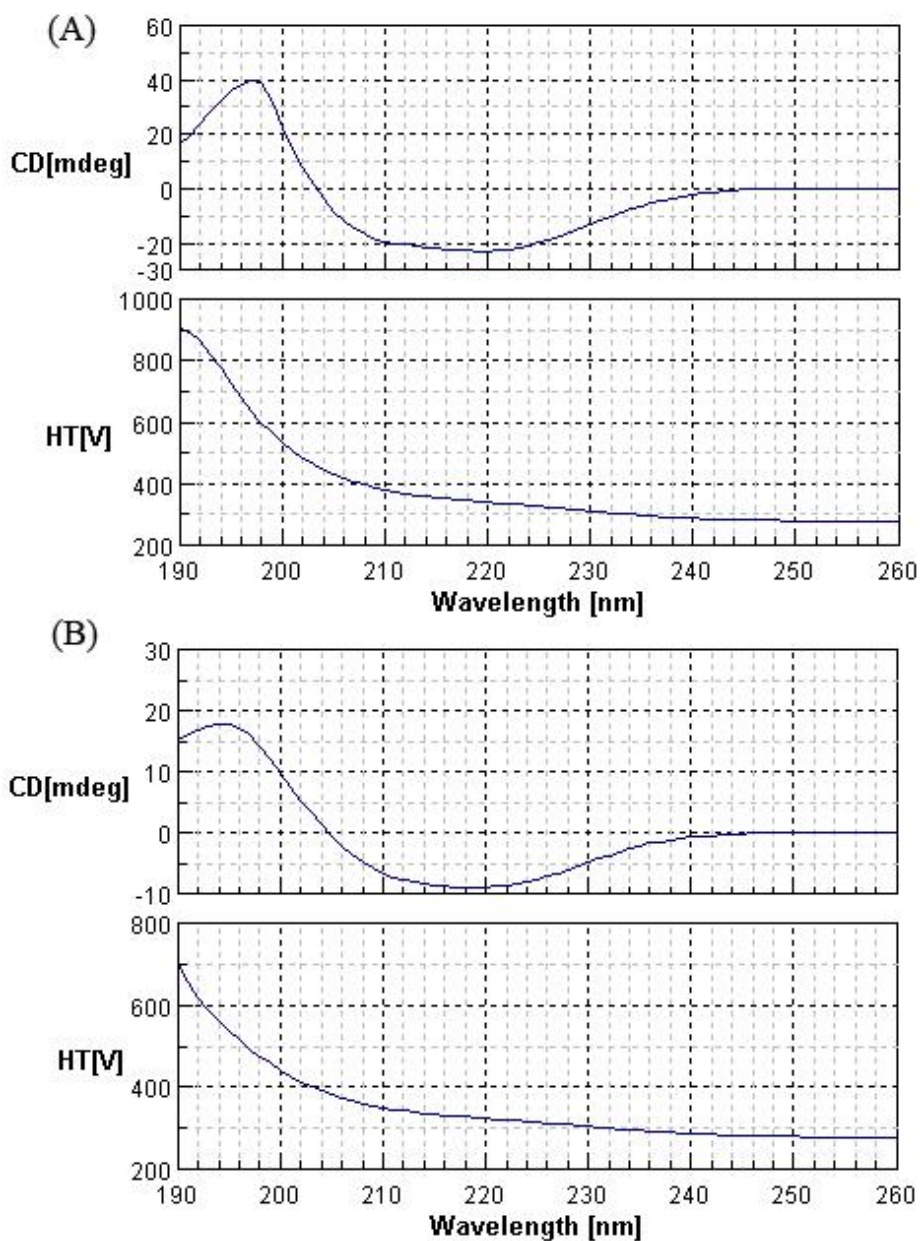


Figure 3-5 CD Spectra of *E. coli* nBamA Δ 1-20 and BamA Δ 1-20

(A) The upper panel shows the far UV CD spectra for *E. coli* nBamA Δ 1-20. The protein concentration is 10.6 μ M. The measurements were taken between wavelengths 190 to 260nm. The lower panel shows the corresponding high tension voltage trace. (B) The upper panel shows the far UV CD spectra for the refolded *E. coli* BamA Δ 1-20. The protein concentration is 5.3 μ M. The measurements were taken between wavelengths 190 to 260nm. The lower panel shows the corresponding high tension voltage trace.

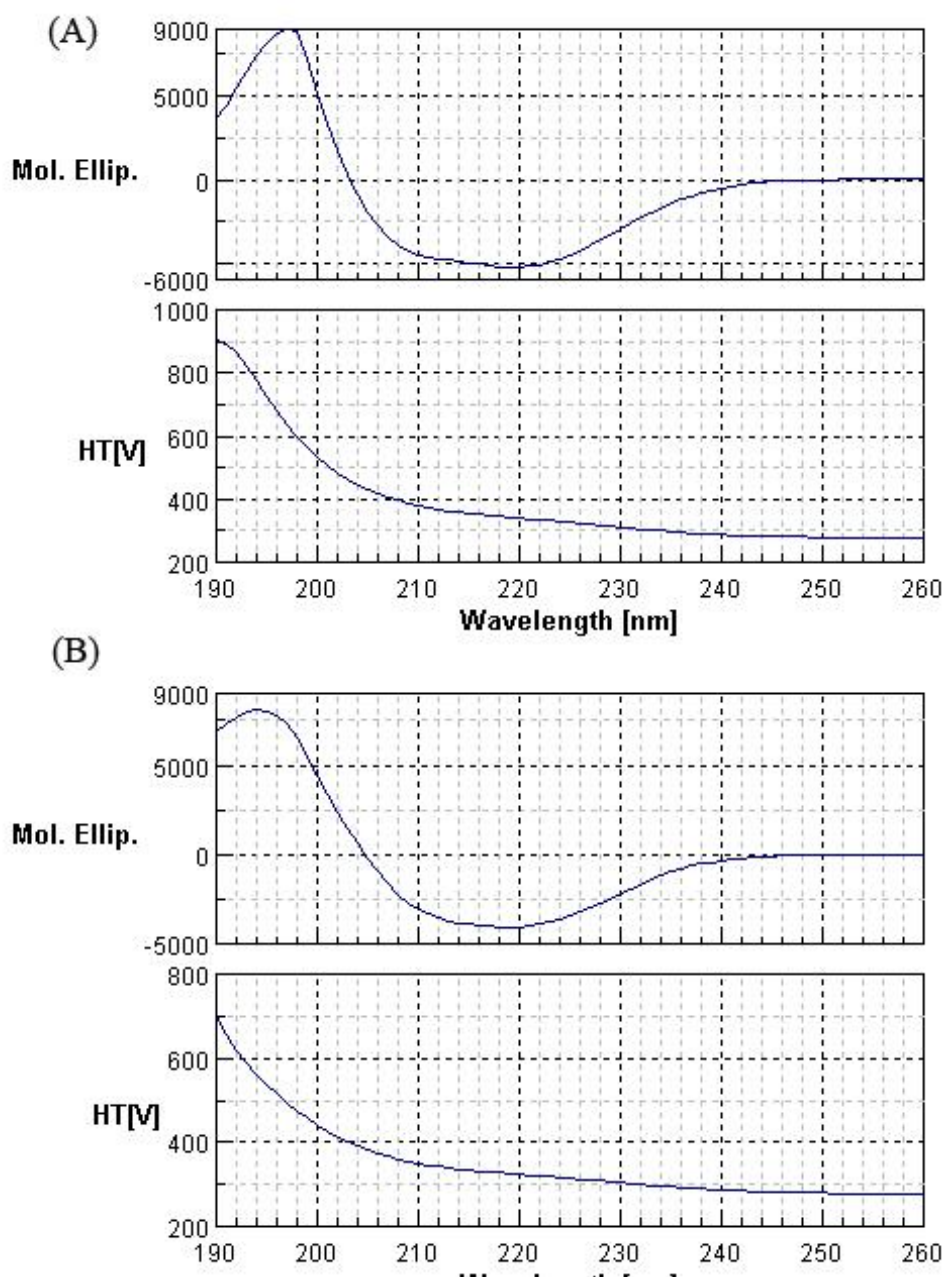


Figure 3-6 CD Spectra of *E. coli* nBamA Δ 1-20 and BamA Δ 1-20 Converted to Molar Ellipticity

(A) The CD absorbance readings of *E. coli* nBamA Δ 1-20 have been converted to molar ellipticity $[\theta]_{\text{molar},\lambda}$ at wavelength λ (190 nm ~ 260 nm). The high tension voltage trace is shown in the lower panel. (B) The CD absorbance readings of refolded *E. coli* BamA Δ 1-20 have been converted to molar ellipticity $[\theta]_{\text{molar},\lambda}$ at wavelength λ (190 nm ~ 260 nm), with unit millidegree $\text{cm}^2\text{dmol}^{-1}$. The high tension voltage trace is shown in the lower panel.

3.3.4. Limited Proteolysis

nBamA Δ 1-20, BamA Δ 1-20 and BamA Δ 1-424 protein samples (1 mg/mL) were digested by trypsin or chymotrypsin (0.1 mg/mL) with the mass ratio (protein: protease) of 1000:1 at room temperature for 0, 5, 10, 15, and 30 minutes; 1, 2, and 3 hours; and overnight.

Based on the SDS-PAGE gel results, nBamA Δ 1-20 and BamA Δ 1-20 had similar digestion patterns. In Figure 3-7 (A) and (C), following overnight digestion with trypsin, the protein fragments represented by the strongest bands on both gels had MWs of 45 kDa. The protein fragments represented by the two fainter bands had MWs of 90 and 60 kDa. Similar results were obtained using chymotrypsin, as seen in Figure 3-7 (B) and (D). The primary protein fragments were 45 kDa in weight and the other proteins fragments yielding faint bands on the gels had MWs of 60 kDa. These digestion patterns suggest that refolded BamA Δ 1-20 has a similar structure to nBamA Δ 1-20; therefore, similar digestion sites were exposed and accessible to the trypsin and chymotrypsin proteases.

In Figure 3-7 (E) and (F), BamA Δ 1-424 was digested by trypsin and chymotrypsin, respectively. BamA Δ 1-424 showed resistance to these two proteases. This result suggests that refolded BamA Δ 1-424 may have a rigid structure, which may also be present in both refolded BamA Δ 1-20 and nBamA Δ 1-20 since 45 kDa protein fragments were also seen in both of their digestions.

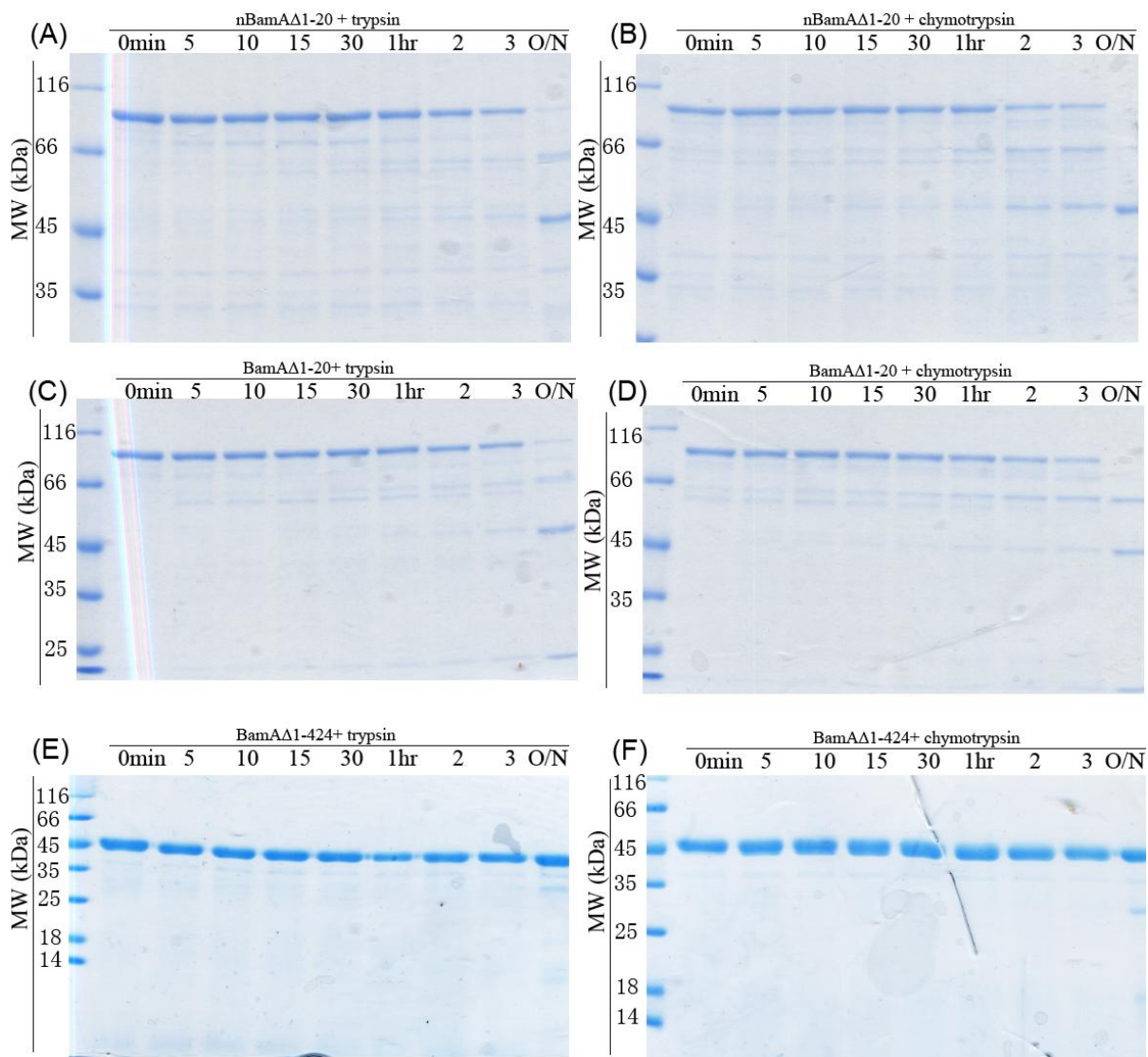


Figure 3-7 SDS-PAGE Gels Showing the Results from Protease Digestion of *E. coli* nBamA Δ 1-20, BamA Δ 1-20, and BamA Δ 1-424

E. coli nBamA Δ 1-20, BamA Δ 1-20 or BamA Δ 1-424 samples were digested with trypsin or chymotrypsin for specific time periods at room temperature. (Mass ratio, protein: protease = 1000: 1) (A) nBamA Δ 1-20 was digested with trypsin for 0, 5, 10, 15, 30 minutes; 1, 2, 3 hours; and overnight. The samples were then run on a 9% SDS-PAGE gel. (B) nBamA Δ 1-20 digested with chymotrypsin and run on a 9% SDS-PAGE gel. (C) BamA Δ 1-20 digested with trypsin and run on a 9% SDS-PAGE gel. (D) BamA Δ 1-20 digested with chymotrypsin and run on a 9% SDS-PAGE gel. (E) The BamA Δ 1-424 digested with trypsin and run on a 12% SDS-PAGE gel. (F) BamA Δ 1-424 digested with chymotrypsin and run on a 12% SDS-PAGE gel.

3.4. Discussion

This chapter presents evidences showing *E. coli* BamA is able to refold using the refolding method proposed in Chapter2. To demonstrate the refolding status of *E. coli* BamA Δ 1-20, as compared to nBamA Δ 1-20, four different approaches are utilized to investigate protein aggregations, heat-modifiability, secondary structure contents and proteolysis features of the refolded proteins.

Size-exclusion chromatography can separate proteins based on their sizes. The resulting protein elution profiles of refolded BamA Δ 1-20 and nBamA Δ 1-20 were superposed for comparison. Both profiles show a small peak in the void volume which suggests proteins included in those corresponding elution fractions form aggregations. These aggregates may be produced during refolding. However, small amounts of aggregations also exist in nBamA Δ 1-20 extracted from the OM, whereas no refolding steps are required following purification. The formation of BamA aggregates might also be caused by detergent effects and some protein features including degradation and hydrophobic interactions between the surfaces of transmembrane β -barrels. Despite the small amount of aggregations, the majority of the proteins tend to be refolded into two different populations in separate peaks. The results of refolded BamA Δ 1-20 and nBamA Δ 1-20 which requires no refolding steps are consistent.

Polyacrylamide gel electrophoresis (PAGE) is a technique widely used to separate biological macromolecules according to their electrophoretic mobility. Mobility is an overall function of the charge, length and conformation of macromolecules including proteins and nucleic acids. Sodium dodecyl sulfate (SDS) is an anionic detergent which linearizes protein and imparts a negative charge in SDS-PAGE. SDS usually binds to proteins in amounts that are proportional to the protein's molecular mass, except for proteins containing large amounts of β structures, where SDS will over-bind (Heller 1978; Weber and Osborn 1969) . However, applying heat (boiling) helps the proteins to be fully unfolded in SDS, whereas SDS-binding will be reduced (Nakamura and Mizushima 1976; Rosenbusch 1974) . The β -barrel proteins are usually resistant in SDS-denaturation condition due to the excessive amounts of hydrogen bonds. The majority of OMPs are rich in β -strands. In previous research, bacterial OMPs, such as OmpA, have shown this heat modifiability feature although the exact working mechanism

is unclear (Surrey and Jahnig 1992b; Tamm, Hong, and Liang 2004b) . Hence, the apparent molecular masses of boiled OMPs on SDS-PAGE gel are expected to be smaller than their theoretical values. In our experiments, boiled nBamA Δ 1-20 samples have an ~20 kDa migration difference compared to unboiled nBamA Δ 1-20 samples. However, unboiled nBamA Δ 1-20 proteins show two bands in SDS-PAGE. Approximately 50% of unboiled nBamA Δ 1-20 protein samples bind to SDS over proportionally to their masses, which causes the shift. The same mobility shift is observed in refolded nBamA Δ 1-20, suggesting β -stranded rich regions exist in nBamA Δ 1-20 and refolded nBamA Δ 1-20. Conversely, the effect of heat-modification visualized in gel electrophoresis is not significant for truncated transmembrane BamA Δ 1-424. This may be due to the high concentration of acrylamide (15%) in gel which is used to run BamA Δ 1-424 (Heller 1978) .

As well as the gel electrophoresis results highlighting the heat-modifiable features of the BamA proteins, the secondary structure of refolded BamA Δ 1-20 was also confirmed by CD spectroscopy, compared to the results from nBamA Δ 1-20. To obtain more reliable results, CD measurements are taken in triplicate to demonstrate the reproducibility of the data. In the resulting CD spectrums, both refolded BamA Δ 1-20 and nBamA Δ 1-20 are observed to have the lowest absorbance (~6000 millidegree \cdot cm²dmol⁻¹) around 217 to 220nm wavelengths. This suggests the majority of the secondary structure components are antiparallel β -strands in both samples. This is consistent with the finding from the heat-modifiability experiments which revealed the existence of β -strands. In future studies, boiled protein samples with or without SDS could be monitored for conversion and denaturation of β -strands by CD spectroscopy. Conversely, the refolded BamA Δ 1-424 did show the expecting strong absorbance in CD spectrums. To optimize this experiment, concentration of refolded BamA Δ 1-424 could be increased to generate stronger signals; meanwhile using other buffers instead of potassium phosphate buffer.

With the experiments conducted so far, the non-aggregated part of the refolded BamA Δ 1-20 proteins has been proven to have substantial antiparallel β -strands. Are these antiparallel β -strands enclosed to form a typical transmembrane β -barrel? This question can be answered by limited proteolysis based on the protease-resistance of the transmembrane β -barrel. The refolded BamA Δ 1-20 and nBamA Δ 1-20 were digested with

either trypsin or chymotrypsin at room temperature. Samples taken from different time points are shown on SDS-PAGE gel in Figure 3-7. Their digestion patterns show a high level of similarity between refolded BamA Δ 1-20 and nBamA Δ 1-20. The primary band remaining after overnight digestion is ~45 kDa, which likely presents the transmembrane β -barrel of BamA. Other than this 45 kDa band, another 60 kDa band presents in Figure 3-7ABCD, which likely indicates POTRA4-5 with transmembrane β -barrel. In order to prove the trypsin-resistance or chymotrypsin-resistance properties of the transmembrane β -barrel, the experiments were repeated using BamA Δ 1-424. The results suggest the transmembrane β -barrel, where contains excessive amount of hydrogen bonds, is relatively stable under this condition. However, not all transmembrane β -barrel proteins survive after overnight digestion. The presence of some degradation or digestion bands imply that two populations may present, protease-resistant BamA and protease-sensitive BamA. Some studies propose these two types of BamA are manipulated by BamD and BamE in such way that the location of Loop 6 is changed in the transmembrane β -barrels (Rigel et al. 2012; Rigel, Ricci, and Silhavy 2013a) . Details will be discussed in Chapter 5.

This chapter provides evidence of successful *E. coli* BamA proteins refolding using different procedures including size-exclusion chromatography to study protein aggregation, heat-modifiability experiments visualized using gel electrophoresis, CD spectroscopy to look at secondary structure and limited proteolysis to determine the protease- resistance of existing transmembrane β -barrels. These results show that a large portion of the purified, denatured BamA can be refolded into its native secondary and tertiary structures. However, accurate structural analysis of the protein requires X-ray crystallography, so crystallizing the refolded BamA protein and solving its three dimensional structure are the ultimate goals of this project.

Chapter 4. *E. coli* BamA Crystallization

4.1. Introduction

This chapter provides a general introduction to X-ray crystallography and presents the images of both full length *E. coli* BamA (BamA Δ 1-20) and truncated transmembrane BamA (BamA Δ 1-424) crystals, together with summaries of their crystallization conditions.

4.2. X-ray Crystallography

Purified proteins studied at the atomic resolution level using X-ray crystallography. After obtaining protein crystals, a beam of X-rays is directed at them to generate a diffraction pattern. The information contained in the resulting diffraction pattern is used to generate a three-dimensional protein structure (Smyth and Martin 2000a). The biomolecular protein structures are then deposited and made available in the Protein Data Bank (PDB) (Figure 4-1).

4.2.1. *Protein Crystals*

A crystal is a regular three-dimensional repeat of molecules, with the smallest building block known as the 'asymmetric unit'. The asymmetric unit is defined as one subunit of a repeating unit as one molecule or more than one molecule. The asymmetric unit can be rotated and translated by space group symmetry operations to make up a 'unit cell', which can be repeated infinitely to generate the entire crystal (Figure 4-1) (Blow 2003; Chayen and Saridakis 2008). A protein crystal can be described by its unit cell properties and the type of symmetry within it, for example, P2₁2₁2. In this particular example, the crystal has a primitive lattice (P), two-fold screw axes parallel to the a (2₁) and b axes (2₁), and a two-fold rotation axis (2) along the c axis in the unit cell.

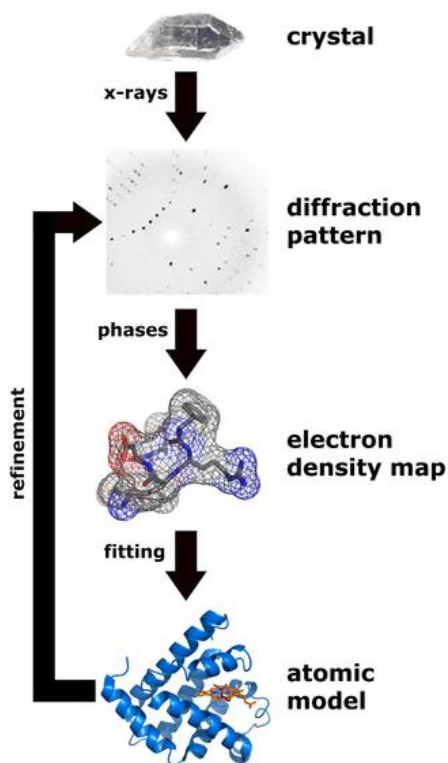


Figure 4-1 Protein Structure As Determined by X-ray Crystallography

The workflow chart describes the process of protein structure determination by X-ray crystallography. The figure was adapted from an online source (<http://www.ebi.ac.uk/training/online/screenshotpages/x-ray-crystallography>)

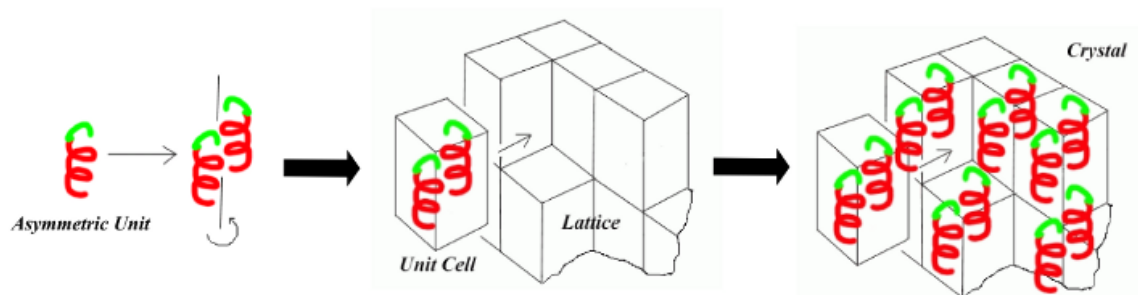


Figure 4-2 Crystal Structure

In this example, two asymmetric units within one unit cell are related to each other by a two-fold rotation. This unit cell is repeated and translated in all directions to form the crystal lattice and make up of an entire crystal. The figure was adapted from an online source (<http://www.ruppweb.org/xray/101index.html>).

4.2.2. Crystallization

For X-ray crystallographic studies, purified protein must be crystallized and this crystal must grow to a sufficient size (larger than 0.1mm in all dimensions) before structure determination can be attempted. To promote protein crystallization, vapour diffusion methods are frequently used, gradually lowering the solubility of the target protein in the presence of precipitants (Chayen and Saridakis 2008) . Large numbers of molecules arranged in an ordered fashion within a crystal will cause scattered X-rays to add up in phase; hence, the signals are strong enough to be detected and diffraction data can be obtained.

The vapour diffusion method can be performed using either hanging drops or sitting drops in a closed system. The sitting-drop vapour diffusion method was the primary system used in this thesis and is described below in Figure 4-3. The crystallization droplet (purified protein sample + reservoir solution) is placed on a pedestal separate from the larger volume reservoir solution. As the drop and reservoir equilibrate, the concentration of purified protein and precipitant within the crystallization drop will gradually increase, leading to the growth of large and well-ordered crystals. In the case of the hanging-drop system, the theory is the same, except that the crystallization droplet is placed on a glass slide flipped upside down so that the droplet is suspended above the reservoir within the vapour diffusion cell.

Growing protein crystals is influenced by many factors including the presence and concentration of salts, precipitants and additives, pH, temperature, protein concentration and protein purity. Crystallization conditions are unique to each protein due to the molecular variations between individual proteins. Therefore, attempting to crystallize a protein for the first time, without a pre-existing protocol, can be very challenging and time consuming.

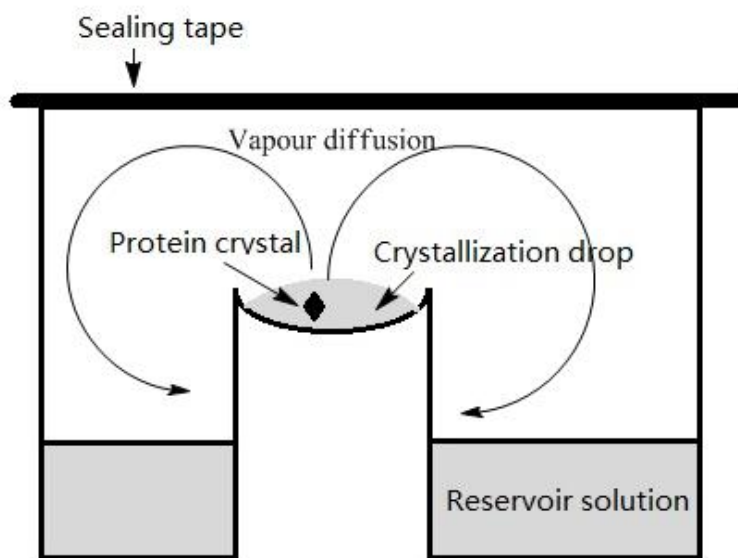


Figure 4-3 Sitting Drop Vapour Diffusion

The sitting-drop vapor diffusion cell is sealed using tape to create an isolation system from the outside environment. The vapour diffuses within this system until the equilibrium is reached.

4.2.3. Data Collection and Processing

The protein crystal is irradiated with a beam of monochromatic X-rays, with a wavelength of $\sim 1\text{-}1.5 \text{ \AA}$. This wavelength is close to the lengths of chemical covalent bonds; hence, it is suitable for determining atomic resolution level structures. The electrons of atoms within the protein crystal can scatter the incoming beam of X-rays, which are then detected by a charge-coupled device (CCD), or an image plate detector. The diffraction pattern is recorded as a two-dimensional image, showing spaced spots (reflections). The protein crystal is rotated to obtain a complete dataset (Blow 2003; Wlodawer et al. 2008; Smyth and Martin 2000b). Diffraction data also can be collected at synchrotrons which produce more intense and refined beams than rotating anode X-ray generators. In this way, the wavelength can be adjusted via monochromator to carry out phasing experiments that require multiple data sets collected at different X-ray wavelengths.

Primary data analysis includes three steps; indexing, integration, and scaling and merging. Indexing refers to the extraction of space group and unit cell information from the diffraction images. The intensities of these indexed reflections are measured in the integration step. In the scaling and merging step, each diffraction image with reflections is placed on a common relative scale to compensate for the errors, and then 'merged' by replacing multiple intensity values with the weighted average. The closer the intensity values are between the equivalent reflections, the higher the data quality (Chayen and Saridakis 2008) .

4.2.4. Phasing

The structure factor is the accurate estimate of a parameter in terms of the amplitude and phase of the diffracted X-ray waves. The amplitude can be obtained from the measured intensity described earlier; however, the phase problem has to be solved by other approaches, depending on whether there is a homologue structure available (Smyth and Martin 2000a; Wlodawer et al. 2008). If homologous structures are available, the initial phase can be estimated with the help of molecular replacement. If no homologues structures can be used, the phase problem in novel structure determination can be overcome using methods such as multiple-wavelength anomalous dispersion (MAD) and single-wavelength anomalous dispersion (SAD) phasing experiments. Both methods require the incorporation of anomalous scatterers. For example, proteins are synthesized by cells grown in minimal media with seleno-methionine rather than methionine. The heavy atoms cause anomalous X-ray scattering at a specific wavelength near the absorption edge of the heavy atom. The resulting phase is unique from the rest of the atoms; therefore, it contributes in calculating amplitude and phase to estimate their location within the asymmetric unit. Following this, an initial electron density map can be generated (Smyth and Martin 2000a; Wlodawer et al. 2008) .

4.2.5. Structure Refinement

Once an initial electron density map is obtained, the polypeptide chains with the known protein sequence can be fitted into the map to generate an initial model. After this step, the phase from this model is applied to the original data to calculate structure

factors and obtain an improved electron density map. Other runs of phase refinement can be performed until the difference between theoretical structure factors and the experimentally derived structure factors is acceptable (Smyth and Martin 2000b; Wlodawer et al. 2008). The correlation is usually measured by statistical indicators known as R-factor and R-free.

4.3. Material and Methods

4.3.1. *E. coli* BamA Δ 1-20 Protein Crystallization

The *E. coli* BamA Δ 1-20 crystals were grown using the sitting drop vapour diffusion method, in a 96-well sitting drop plate. The refolded and purified BamA Δ 1-20 protein was solubilized in size-exclusion buffer (20mM Tris-HCl pH8.0, 100mM NaCl, 0.01% DDM, 2mM DTT). 0.5 μ L of this 10mg/mL concentration protein sample was mixed with 0.5 μ L of the reservoir solution and the resulting drop was then equilibrated with 100 μ L of reservoir solution at room temperature (\sim 22°C). The reservoir solutions used were from Hampton Research Screens: Crystal Screen I, Crystal Screen II, PEG/ION and MembFac.

4.3.2. *E. coli* BamA Δ 1-424 Protein Crystallization

The refolded and purified BamA Δ 1-424 was resuspended in size-exclusion buffer (20mM Tris-HCl pH8.0, 100mM NaCl, 0.01% DDM, 2mM DTT) and concentrated to 10mg/mL. This concentration sample was treated with chymotrypsin (Mass ratio, protein: protease = 1000:1) before being used for crystal plating. Crystals were grown using the sitting drop vapour diffusion method, in a 96-well sitting drop plate. 0.5 μ L of the reservoir solution was added to 0.5 μ L of the protein-protease mixture (\sim 10mg/mL). The resulting drop was equilibrated with 100 μ L reservoir solutions at room temperature (\sim 22°C). The reservoir solutions used were from Hampton Research Screens: Crystal Screen I, Crystal Screen II, PEG/ION and MembFac.

4.3.3. *E.coli* BamA Δ 1-424 Crystals Optimization

The BamA Δ 1-424 crystals were observed in two conditions that are described later in Section 4.4. One initial hit had needle clusters shape BamA Δ 1-424 crystals, and was reproducible in the same condition (0.1M HEPES pH7.5 and 70% MPD (+/-)-2-methyl-2, 4- pentanediol). The optimization of the crystallization condition was performed by varying the pH values of the HEPES and the concentration of the MPD. Crystals were grown using the sitting drop vapour diffusion methods, in a 24-well sitting drop plate. 1 μ L of the reservoir solution was added to 1 μ L of BamA Δ 1-424, and the resulting drop was equilibrated with 1 mL reservoir solutions at room temperature (22°C).

Optimization was also performed using an additive screen, where the same batch of purified BamA Δ 1-424 protein was equilibrated in the original crystallization condition at room temperature (22°C). Crystals were grown using the sitting drop vapour diffusion methods, in a 24-well sitting drop plate.

4.4. Results

4.4.1. *E. coli* BamA Δ 1-20 and BamA Δ 1-424 Protein Crystals

Both *E. coli* BamA Δ 1-20 and BamA Δ 1-424 protein crystals were observed in initial hits for crystallization and are shown in Figure 4-4.

The refolded and purified *E. coli* BamA Δ 1-20 protein samples collected from peak 2 elution fractions were concentrated to 10 mg/mL in size-exclusion chromatography buffer (described above). The reservoir solution was composed of 0.2M Calcium chloride dehydrate, 0.1M Sodium acetate trihydrate pH4.6 and 20% v/v 2-Propanol. As seen in Figure 4-4 (A), the resulting BamA Δ 1-20 crystals had a doughnut-shaped appearance and showed up two weeks after crystal plating. There were no similar crystals observed in the reservoir. Once this initial hint was observed, attempts were made to reproduce the crystals in same condition using a different batch of purified BamA Δ 1-20. However, no positive results were obtained.

The BamA Δ 1-424 protein crystal initial hits were obtained in two conditions. As seen in Figure 4-4 (B), a single crystal was observed after four months, equilibrating in the following reservoir condition; 0.1M ADA pH6.5, 0.1M MgCl₂·6H₂O and 12% PEG6000.

Before being plated, the refolded and purified BamA Δ 1-424 (protein concentration 10 mg/mL) was digested by chymotrypsin overnight at 4°C. The single crystal had a rectangular plate shape. The other initial hit, seen in Figure 4-4 (C), had a reservoir condition of 0.1M HEPES pH7.5 and 70% MPD (+/-)-2-methyl-2, 4- pentanediol. The BamA Δ 1-424 protein formed needle clusters in this condition and showed up after one month. The crystals were replicable using the same batch of purified proteins. After being optimized, the size of the needle clusters increased, while the number of clusters in each drop decreased. The crystals changed to a plate-shape with the addition of additives.

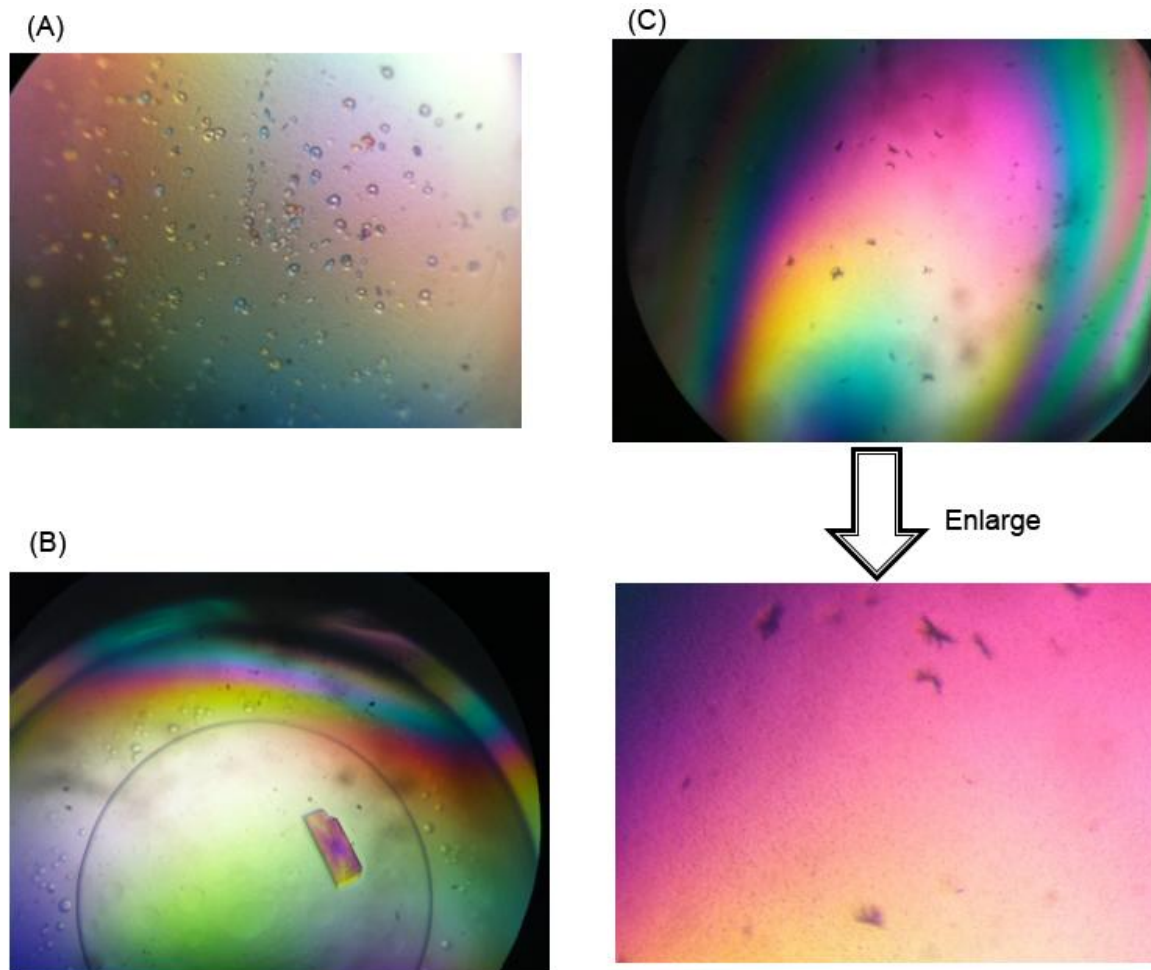


Figure 4-4 *E. coli* BamA Δ 1-20 and BamA Δ 1-424 Protein Crystals

(A) *E. coli* BamA Δ 1-20 crystals were grown using the sitting drop method. (B) *E. coli* BamA Δ 1-424 crystal were grown using the sitting drop method. The BamA Δ 1-424 protein (10mg/ml) was treated with chymotrypsin (Mass ratio, protein: protease = 1000:1)

overnight at 4°C before crystal plating. (C) *E. coli* BamA Δ 1-424 protein (10mg/ml) was treated with chymotrypsin (Mass ratio, protein: protease = 1000:1) for 1 hour at room temperature before crystal plating. An enlargement of the image is included above.

4.5. Discussion

This chapter introduces X-ray crystallography and presents three initial hits during protein crystallization with their crystallization conditions.

Obtaining diffraction quality protein crystals of a sufficient size is often a major bottleneck in the entire process of structure determination. In this thesis project, three different initial crystallization hits for BamA Δ 1-20 and BamA Δ 1-424 were observed. Before being irradiated with a beam of monochromatic X-rays, crystals must be evaluated. The primary concern is to distinguish between inorganic (salt) and organic (protein) crystals.

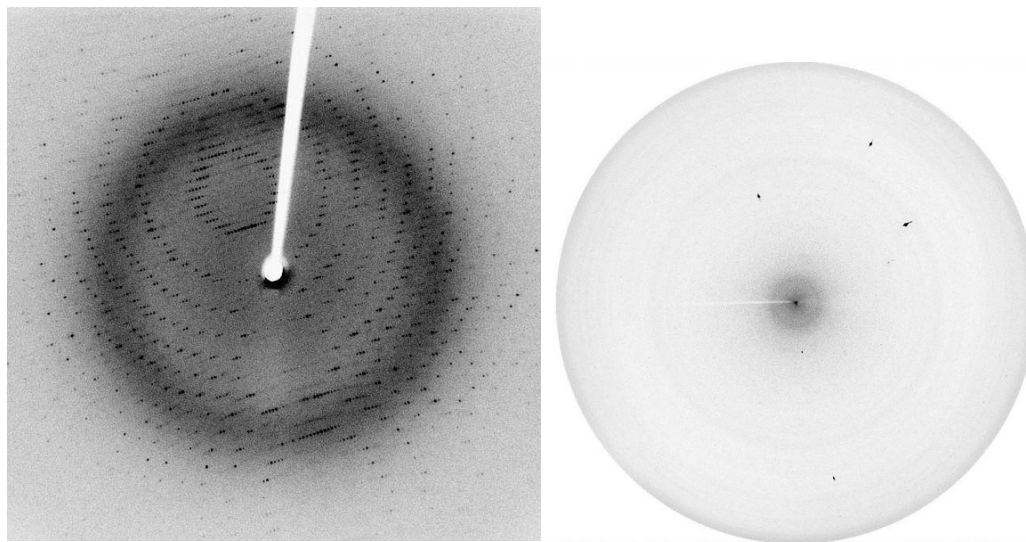


Figure 4-5 X-ray Diffraction Patterns of Protein or Salt Crystals

This is a diffraction pattern comparison between organic (protein) and inorganic (salt) crystals. The left panel image is a diffraction pattern of a SARS protease. The right panel image shows a typical salt crystal diffraction pattern, in this case, a phosphate mineral, Brushite. Both figures were adapted from an online source (<http://benchwise.wordpress.com/2013/04/13/crystallography-for-beginners-part-5-monitoring-and-evaluating-crystallization-experiments-results/>).

A direct and definitive test is to simply mount the crystal in an X-ray beam. Protein or salt crystals can be easily identified by their characteristic reflections in an X-ray diffraction pattern. Salt crystals usually show very few spots (reflections) around middle to high resolution levels due to the small size of their unit cells, while protein crystals have many reflections across all resolution levels (Figure 4-5). However, if crystals are poor quality or they are not large enough to be mounted; other common methods can be used for the same purpose such as: running gels, dye absorption, physical manipulation and carrying out control experiments. In the first method, crystals can be collected, washed and dissolved, then run on SDS-PAGE. If a band indicating the target protein molecular mass is observed, it is very likely to be a protein crystal. Occasionally, protein can degrade during crystallization or a proteolytic reaction can be involved. In this case, the band on SDS-PAGE often yields a molecular size smaller than the theoretical value of the protein. The second method involves using dye to identify whether the crystals are protein or salt. Since proteins are packed very loosely with large solvent channels present within the crystals, small molecules like dyes can travel through the channels and dye the crystal a different color. However, dye-staining can affect crystal diffraction or even crystallize in some conditions, so dye is usually only used if there is only one crystal available. In a salt crystal, which is packed more densely, there is little solvent content inside the crystal, making it very hard to crush. Protein crystals, however, are very fragile and will powder, crumble or break easily because of their loose packing and large solvent channels. Control experiments can be set up in parallel with the original crystal plating. One condition with a specific reservoir solution will include protein and buffer in the suspended or sitting droplet while the other condition with the same reservoir solution will only contain buffer in the droplet. If crystals grow in both conditions, they are most likely salts rather than protein.

In the experiments for this thesis, controls were set up but no crystals were observed in the controls. However, the possibility of the crystals being salt cannot be ruled out since the control and actual experimental droplets were separated and sealed in two different cells. In this case, two different evaporation rates may be involved. Meanwhile, most of crystals obtained from initial hits were too small to be mounted for X-ray diffraction experiments. The needle clusters (Figure 4-4C) are one-dimensional crystals that still need to be optimized. Ongoing optimization experiments varying pH

value or adding additives may change the shape of crystals to a s plate-shaped; however, at the moment, the crystal size is still not large enough. Therefore, more optimization work will be required in future.

In conclusion, this chapter has summarized the initial hits of *E. coli* BamA Δ 1-20 and BamA Δ 1-424 and described their crystallization conditions. These conditions provide a good starting point for obtaining the large and diffractable crystals required to resolve the full length structure of *E. coli* BamA.

Chapter 5. E. coli BamA Homology Model Based on Neisseria gonorrhoeae BamA (NgBamA)

5.1. Introduction

This chapter will demonstrate the *E. coli* BamA Δ 1-20 homology model created by using the *Neisseria gonorrhoeae* full length BamA (NgBamA) crystal structure as a template (Noinaj et al. 2013) . The resulting homology model is then superposed with NgBamA, and the POTRA and β -barrel domains of *E. coli* BamA which has just been solved recently. The purpose of this chapter is to provide insights into the prediction of *E. coli* full length BamA's structure, based on the solved structure of NgBamA.

5.2. BAM Complex Homologues

Homologues of the BAM complex not only exist in all Gram-negative bacteria, but also in the endosymbiotically derived eukaryotic organelles, mitochondria and chloroplasts (Walther, Rapaport, and Tommassen 2009) . The essential component of this machinery, named Omp85 or BamA, is a highly conserved bacterial protein that recognizes a signature sequence at the C-terminus of its substrate OMPs. The homologues of BamA are also found in mitochondria and have similar functions, indicating that the basic mechanism of OMP assembly is crucial and evolutionarily conserved.

5.2.1. Eukaryotic Homologues

Mitochondrial and chloroplastic OMPs are synthesized in the cytosol prior to being targeted. The signal sequence leads the OMP to the organelle (mitochondrion or chloroplast) membrane rather than to the plasma membrane of the cell. BAM complex homologues are named called the Sorting and Assembly Machinery (SAM) complex in

mitochondria, and Translocons at the Outer and Inner envelopes of Chloroplasts (the TOC and TIC complexes) in chloroplasts (Schleiff and Soll 2005) .

In the mitochondrial system, the substrate proteins are first imported into mitochondrion via the translocons of the outer mitochondrial membrane (TOM) before the protein is inserted into the mitochondrial outer membrane (MOM) (Schleiff and Soll 2005). Once the substrate protein enters the intermembrane space (IMS), Tim chaperones deliver the protein back to the MOM for assembly by the SAM complex. The SAM complex consists of Sam50, Sam35 and Sam37. Sam50 is the BamA homologue, containing one POTRA domain facing the IMS. The POTRA domain is essential for substrate release (Schleiff and Soll 2005) . Sam35 and Sam37 are two cytosolic proteins identified as the accessory proteins of the SAM complex. It appears that Sam35 plays an important role in cell survival and is involved in substrate recognition (Milenkovic et al., 2004; Paschen et al., 2005). Meanwhile, Sam37 is involved in substrate release (Chacinska et al., 2009; Paschen et al., 2005).

For chloroplasts, protein imported from the cytosol into the stroma need to pass through the TOC and TIC complexes, and then travel back to the outer envelope membrane (OEM) for assembly by Toc75-V (Oreb et al. 2008) , Toc75-V is the BamA homologue containing three POTRA domains facing the IMS. However, a recent study has shown the POTRA domains face the cytosol (Sommer et al. 2011) . This suggests that OMPs are likely to be imported directly via the Toc75-V pathway. But, limited evidence is currently available so the exact pathway and the essential accessory proteins in the system remain unknown. More research is required before the mechanism of chloroplastic OMP assembly is understood (Schleiff and Soll 2005; Fairman, Noinaj, and Buchanan 2011) .

5.2.2. Bacterial Homologues

BamA proteins have homologues in all Proteobacteria. A well known BamA homologue is *Bordetella pertussis* FhaC, which belongs to the outer membrane protein 85–two-partner secretion B (Omp85-TpsB) superfamily. The Tps B transporter FhaC is involved in the export of filamentous hemagglutinin during infection (Clantin et al. 2007a; Jacob - Dubuisson, Locht, and Antoine 2001) . FhaC has a 16-stranded β -barrel at the

C-terminus, with a periplasmic region at the N-terminus as shown in Figure 5-1 (Clantin et al. 2007c) .

Recent studies have been published to show the BamA homologue structures from two species, *Neisseria gonorrhoeae* and *Haemophilus ducreyi* (Noinaj et al. 2013) . *N.gonorrhoeae* BamA is a full length structure with a transmembrane β -barrel and five POTRA domains (*NgBamA*), while *H. ducreyi* only contains one transmembrane β - barrel and two POTRA domains (*HdBamA Δ 3*) shown in Figure 5-1. The difference in the position of POTRA 5 between the two homologues suggests that two conformations of BamA may be involved in a gating mechanism which can regulate the access of substrate into the β -barrel channel from the periplasmic space. In addition, the hydrophobic belt of the BamA β -barrel is significantly reduced in width along the C-terminus ($\sim 9\text{\AA}$) compared to the other side of the barrel ($\sim 20\text{\AA}$), suggesting the possibility that BamA can destabilize the local membrane environment. Simulation experiments also propose a lateral opening between the $\beta 1$ and $\beta 16$ strands, both of which are originally closed by hydrogen bonds. These structural insights have increased our understanding of the working mechanism of BamA. However, the structure of the full length *E. coli* BamA remains unknown. By using homology modelling to arrange the backbone of the *E. coli* BamA sequence identically to the 3D template (the known structure of *NgBamA*), the hypothesized structure of the full length *E. coli* BamA can be predicted, and OMP biogenesis revealed.

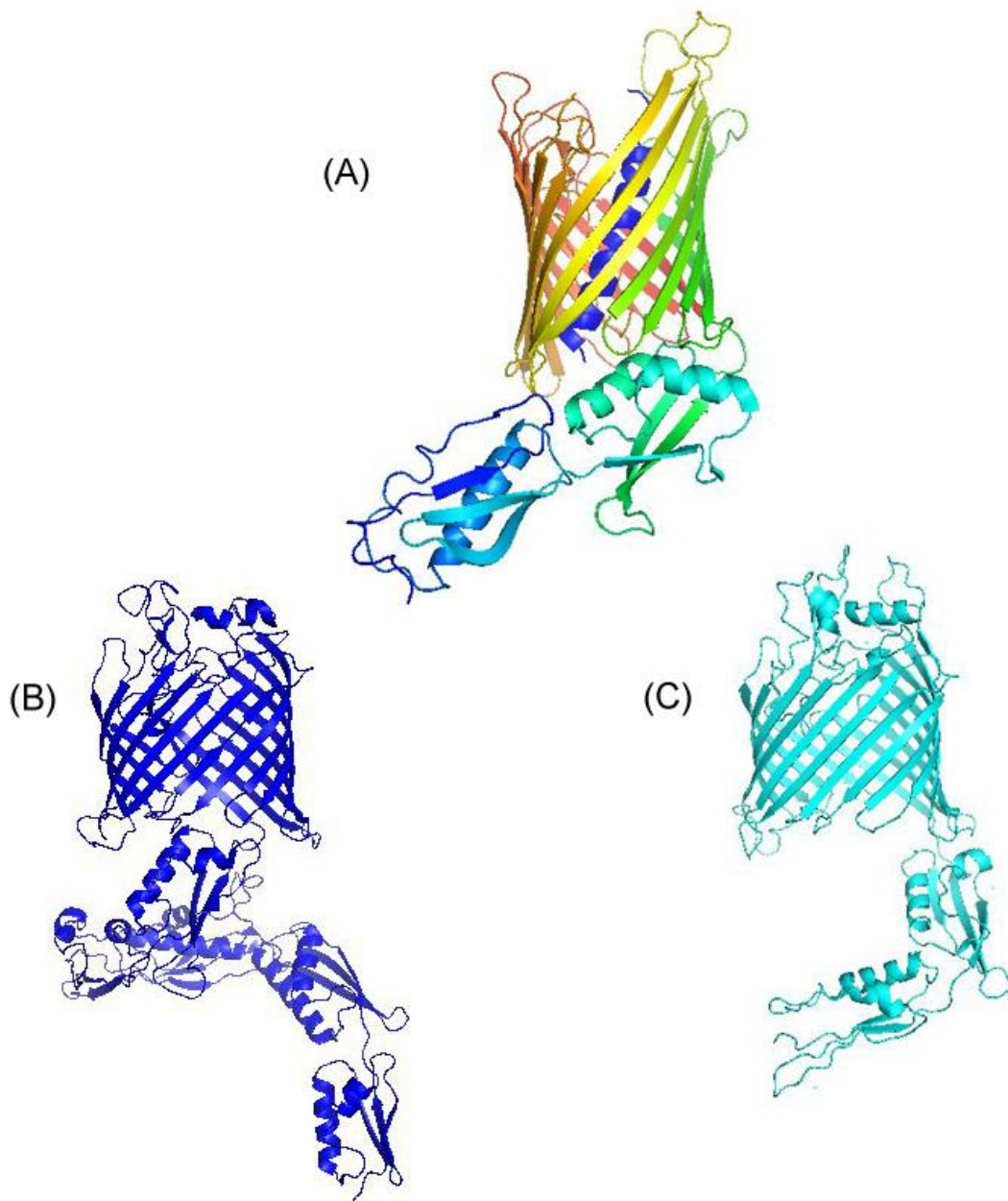


Figure 5-1 Structures of *B. pertussis* FhaC, NgBamA and HdBamA Δ 3

All three structures are displayed in PyMol. (A) The *B. pertussis* FhaC structure (PDB: 2QDZ) shown in colored cartoon contains a transmembrane β -barrel domain and two POTRA domains. (B) NgBamA (PDB: 4K3B) shown in blue cartoon contains a transmembrane β -barrel domain and five POTRA domains. The position of POTRA5 suggests it is in a 'closed' state, which blocks the substrate entrance from periplasmic space into the barrel channel. (C) HdBamA Δ 3 (PDB: 4K3C) shown in cyan cartoon contains a transmembrane β -barrel domain and two POTRA domains. The position of POTRA5 suggests it is in an 'opened' state, which allowing the substrate entrance from the periplasmic space into the barrel channel.

5.3. Material and Methods

The *E. coli* BamA Δ 1-20 homology model was created using the *Neisseria gonorrhoeae* full length BamA (NgBamA) crystal structure (PDB: 4K3B) as a template (Noinaj et al. 2013) . The amino acid sequence of the template (PDB: 4K3B) was mutated at those residues which are different from the *E. coli* BamA Δ 1-20 amino acid sequence, using the program, Crystallographic Object-Oriented Toolkit (Coot) (Emsley and Cowtan 2004) . The mutated structure was then optimized by the program, Yet Another Scientific Artificial Reality Application (YASARA) energy minimization server to adjust side chains directions to avoid collision (Krieger, Vriend, and Spronk 2013) . The workflow of making a homology model is shown in Figure 5-2. Overall differences in the protein backbone structures are quantitated with the root mean square deviation (RMSD) of the positions of the alpha carbons using SuperPose Server (Maiti et al. 2004) .

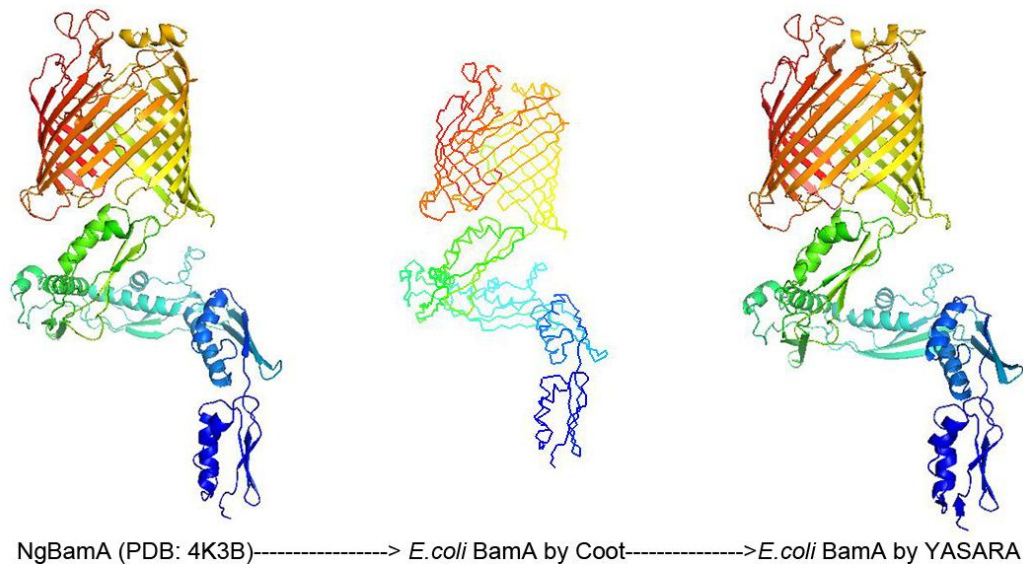


Figure 5-2 *E. coli* full length BamA Homology Model

The process of developing the *E. coli* full length BamA homology model based on the known NgBamA crystal structure. The resulting structures in each step are visualized in PyMOL and shown above.

5.4. Results

5.4.1. Sequence Identities

The sequence of *E. coli* BamA is aligned with the complete sequences of NgBamA and HdBamA to investigate regions of similarity that may be a consequence of functional, structural, or evolutionary relationships between the sequences. Both NgBamA and HdBamA have ~30% sequence identity compared to *E. coli* BamA sequence as shown in Figure 5-3. The identical residues among these three species present across the entire sequence of BamA. There are no significant gaps within these three alignment sequences, except that residues 670-701 of *E. coli* BamA are missing in NgBamA.

5.4.2. *E. coli* BamA Homology Model

Since the sequence identity is only 30%, almost 200 amino acid residues of NgBamA were mutated to be consistent with *E. coli* full length BamA, according to the sequence alignment. This step was done using the program Coot and no secondary structure formation was observed in the output in PyMOL, which may be due to the computational problem. The resulting Coot model was optimized using the YASARA energy minimization server to adjust side chain direction and avoid the side chain collisions. The resulting structure from the *E. coli* BamA homology model is shown in Figure 5-4(A).

Comparing the known structure of NgBamA and the *E. coli* BamA homology model, both structures appear to be very similar. The *E. coli* BamA homology model showed the expected transmembrane β -barrel domain and five POTRA domains. Unlike another prediction in 2003 proposing a 12-stranded β -barrel, our homology model is consistent with the prediction of that the *E. coli* BamA transmembrane domain contains a 16-stranded β -barrel, based on the solved structure of FhaC which is another Omp85 superfamily member (Voulhoux et al. 2003b; Leonard-Rivera and Misra 2012). The homology model RMSD values of the α carbons and backbone are calculated as 1.23 Å and 1.25 Å, respectively.

1 10 20 30 40 50 60 70

EcBamA MAMK**K**L**L**I**A**S**L**L**F**S**S**A**T**V**Y**G**A**E**G****F**V**V****K****D**I**H****F****E**G**L****Q****R**V**A**V**G**A**A**L**L**S**M****P****V****R****T****G****D****T****V****N****D****E****D****I****S****N****T****I****R****A****L****F****A****T****G**

HdBamA . . M**K****K****L****L****S****S****L****L**V**A****N****G** . V**I****A****A****P** . **F****V****V****K****D****I****R****I****D****G****V****Q****T****E****T****G****N****A****I****I****A****S****L****P****V****K****V****G****Q****T****A****T****D****G****D****V****T****N****I****V****K****Q****L****F****S****Q****N**

NgBamA . M**K****L****K****Q****I****A****S****A****L****M****L****M****L****G****I****S****P****L****A****F****A****D****F****T****I****Q****D****I****R****I****V****E****G****L****Q****R****T****E****P****S****T****V****F****N****Y****L****P****V****K****V****G****D****T****Y****N****D****T****H****G****S****A****I****K****S****L****Y****A****T****G**

80 90 100 110 120 130 140

EcBamA N**F****E****D****V****R****V****L****R****D****G****D****T****L****L****V****Q****V****K****E****R****F****T****A****S****I****T****F****S****G****N****K****S****V****K****D****D****M****L****K****O****N****L****E****A****S****G****V****R****V****G****E****S****L****D****R****T****T****I****A****D****I****E****K****G****L****E****D****F**

HdBamA R**F****E****N****V****S****A****A****R****E****G****Q****T****L****V****I****K****V****A****E****R****P****L****I****N****N****L****T****I****K****G****N****N****A****I****P****K****N****A****L****E****Q****N****L****K****A****N****L****I****V****A****G****E****V****Y****D****K****A****K****L****E****A****F****K****Q****A****L****V****D****H**

NgBamA R**F****D****D****V****R****V****E****T****A****D****G****Q****L****L****L****T****V****I****E****R****F****T****I****G****S****L****N****I****T****G****A****K****M****L****Q****N****D****A****I****K****K****N****L****E****S****F****G****L****A****Q****S****Q****Y****F****N****Q****A****T****L****N****Q****A****V****A****G****L****K****E****E**

150 160 170 180 190 200 210

EcBamA Y**Y****S****V****G****K****Y****S****A****S****V****K****A****V****T****P****L****P****R****N****R****V****D****L****K****L****V****F****Q****E****G****V****S****A****E****I****Q****Q****I****N****I****V****G****N****H****A****F****T****T****D****E****L****I****S****H****F****Q****L****R****D****E****V****P****W****N****V****V****G**

HdBamA Y**H****T****M****G****R****Y****Q****A****D****I****Q****T****I****T****P****N****N****N****G****S****I****N****V****E****L****N****I****T****E****G****E****I****A****Y****V****K****K****I****N****F****E****G****N****N****A****F****S****Y****D****E****L****I****K****E****L****E****I****K****P****N****A****P****W****N****I****F****E**

NgBamA Y**L****G****R****G****K****L****N****I****Q****I****T****P****K****V****T****K****L****A****R****N****R****V****D****I****D****I****T****I****D****E****G****K****S****A****K****I****T****D****I****E****F****E****G****N****Q****V****Y****S****D****R****K****L****M****R****Q****M****S****L****T****E****G****G****I****W****T****W****L****T****R**

220 230 240 250 260 270 280

EcBamA D**R****K****Y****Q****K****Q****K****L****A****G****D****L****E****T****L****R****S****Y****L****D****R****G****Y****A****R****F****N****I****D****S****T****Q****V****S****L****T****P****D****K****K****G****I****Y****V****T****V****N****T****E****G****D****Q****Y****K****L****S****G****V****E****V****S****G****N****L****A****G****H**

HdBamA S**S****K****F****Q****Q****Q****E****Y****N****K****D****I****E****I****L****R****D****F****Y****M****D****H****G****Y****A****K****F****T****L****K****D****T****D****V****K****F****N****E****N****K****T****E****V****D****L****T****Y****K****I****N****E****S****Q****Y****N****I****S****E****M****R****I****I****G****D****T****Q****K****L**

NgBamA S**D****R****F****D****R****Q****K****F****A****Q****D****M****E****K****V****T****D****F****Y****Q****N****N****G****Y****F****D****F****R****I****L****D****T****D****I****Q****T****N****E****D****K****T****R****Q****T****I****K****I****T****V****H****E****G****R****F****R****W****G****K****V****S****I****E****G****D****T****N****E****V**

290 300 310 320 330 340

EcBamA S**A****E****I****E****Q** . L**T****K****I****E****P****G****E****L****Y****N****G****T****K****V****T****K****M****E****D****D****I****K****K****L****L****G****R****Y****G****Y****A****P****R****V****Q****S****M****P****E****I****N****D****A****D****K****T****V****K****L****R****V****N****V****D****A****G****N****R****F****Y****V**

HdBamA D**N****E****L****N****Q****L****L****T****H****F****K****A****G****Q****L****F****R****K****T****E****L****S****I****I****E****E****Q****I****K****Q****I****L****G****D****R****G****Y****G****S****A****K****V****D****L****Y****E****K****F****N****E****E****D****H****T****V****Q****I****N****F****I****V****D****A****G****R****R****I****Y****V**

NgBamA P**K****A****E****L****E****K****L****L****T****M****K****P****G****K****W****Y****E****R****Q****M****T****A****V****L****G****E****I****Q****N****R****M****G****S****A****G****Y****A****Y****S****E****I****S****V****Q****E****L****P****N****A****G****T****K****T****V****D****F****V****L****H****I****E****P****G****R****K****I****Y****V**

360 370 380 390 400 410

EcBamA R**K****I****R****F****E****G****N****D****T****S****K****D****A****V****L****R****R****E****M****R****Q****M****E****G****A****W****L****G****S****D****L****V****D****Q****G****K****E****R****L****N****R****L****G****F****F****E****T****V****D****T****D****T****Q****R****V****P****G****S****P****D****Q****V****D****V****V****Y****K****V****K**

HdBamA R**K****I****R****F****E****G****N****D****V****T****A****D****S****T****L****R****R****E****M****R****Q****Q****E****G****A****W****L****S****T****S****A****V****S****L****A****K****S****R****L****E****R****T****G****F****Y****E****T****V****E****M****S****M****P****T****V****K****N****T****D****D****Q****V****D****I****I****Y****K****I****K**

NgBamA N**E****H****I****T****G****N****N****K****T****R****D****E****V****V****R****R****E****R****Q****M****E****S****A****P****Y****D****T****S****K****L****Q****R****S****K****E****R****V****E****L****L****G****Y****F****D****N****V****Q****F****D****A****V****P****L****A****G****T****P****D****K****V****D****L****N****M****S****L****T**

430 440 450 460 470 480

EcBamA E**R****N****T****G****S****F****N****F****G****I****G****T****E****S****G****V****S****F****Q****A****G****V****Q****D****N****W****L****G****T****G****Y****A****V****G****I****N****G****T****K****N****D****Y****Q****T****Y****A****E****L****S****V****T****N****P****Y****F****T****V****D****G****V****S****L****G****R****L**

HdBamA E**R****N****T****G****S****I****N****F****G****V****G****Y****G****S****G****S****L****S****Y****N****A****G****I****T****Q****D****N****F****L****G****M****G****S****L****G****L****N****G****S****R****N****T****D****S****T****N****V****N****L****S****Y****T****E****P****Y****F****T****K****D****G****V****S****L****G****N****I**

NgBamA E**R****S****T****G****S****L****D****L****S****A****G****W****V****Q****D****T****G****L****V****M****S****A****G****V****S****O****D****N****L****F****G****T****G****K****S****A****A****L****R****A****S****R****S****K****T****L****N****G****S****L****S****F****T****P****Y****F****T****A****D****G****V****S****L****G****Y****D****I**

500 510 520 530 540 550

EcBamA F**Y****N****D****F****Q****A****D****D****A****D****L****S** . . D**Y****T****N****K****S****Y****G****T****D****V****T****L****G****F****P****I****N****E****Y****N****S****L****R****A****G****L****G****Y****V****H****N****S****L****S****N****M****Q****P****Q****V****A****M****W****R****Y****L****Y****S****M****G** . . E**H**

HdBamA F**Y****E****D****Y****D****N****S****A****R****K****A****S****A** . A**Y****K****R****K****T****Y****G****A****S****G****T****L****G****F****P****V****D****E****N****N****S****Y****L****G****L****G****Y****T****H****D****K****L****R****N****V****E****R****E****Y****T****R****E****K****Y****V****N****S****M****K****F****P****I****N**

NgBamA Y**G****K****A****F****D****P****R****K****A****S****T****S****V****K****Q****Y****K****T****T****A****G****G****V****R****M****G****I****P****V****T****E****Y****D****R****V****N****F****L****A****A****B****H****L****T****V****N****T****Y****N** . . K**A****P****K****R****Y****A****D****F****I****R** . . K**Y**

560 570 580 590 600 610 620

EcBamA F**S****T****S****D****Q****D****N****S****F****K****T****D****D****F****T****F****N****Y****G****W****T****Y****N****K****L****D****R****G****Y****F****P****T****D****G****S****R****V****N****L****T****G****K****V****T****I****P****G****S****D****N****E****Y****Y****K****V****T****L****D****T****A****T****Y****V****P****I****D****D****D****H**

HdBamA F**Q****N****S****H****Y****D****R** . I**Q****S****A****D****F****D****L****S****F****G****W****N****Y****N****N****L****N****R****G****Y****F****P****T****A****G****S****S****A****N****I****S****G****K****L****T****L****P****G****S****D****N****K****Y****Y****Q****V****G****T****N****F****S****G****Y****I****P****L****N****S****E****H**

NgBamA G**K****T****D****G****A****D****S****F****K****G****L****L****Y****K****G****T****V****G****W****R****N****K****T****D****S****A****S****W****P****T****R****G****Y****L****T****G****V****N****A****E****I****A****L****P****G****S****K****L****Q****Y****Y****S****A****T****H****N****Q****T****W****F****P****L****S****K****T****F**

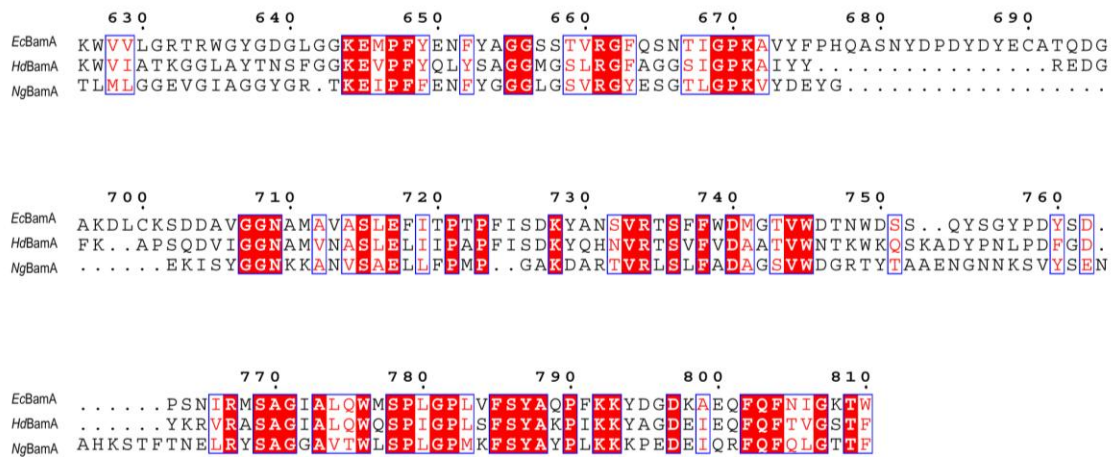


Figure 5-3 Sequence Alignment of *E. coli* BamA with *N. gonorrhoeae* BamA (NgBamA) and *H. ducreyi* BamA (HdBamA)

The amino acid sequence of *E. coli* BamA (UniProt: P0A940) aligned with *NgBamA* (UniProt: Q5F5W8) and *HdBamA* (UniProt: G1UBA5) sequences. The sequence identities are 30%. The common residues among the three species are highlighted in red boxes. The sequence alignment is analysed and generated using ESPript (Gouet et al. 1999) .

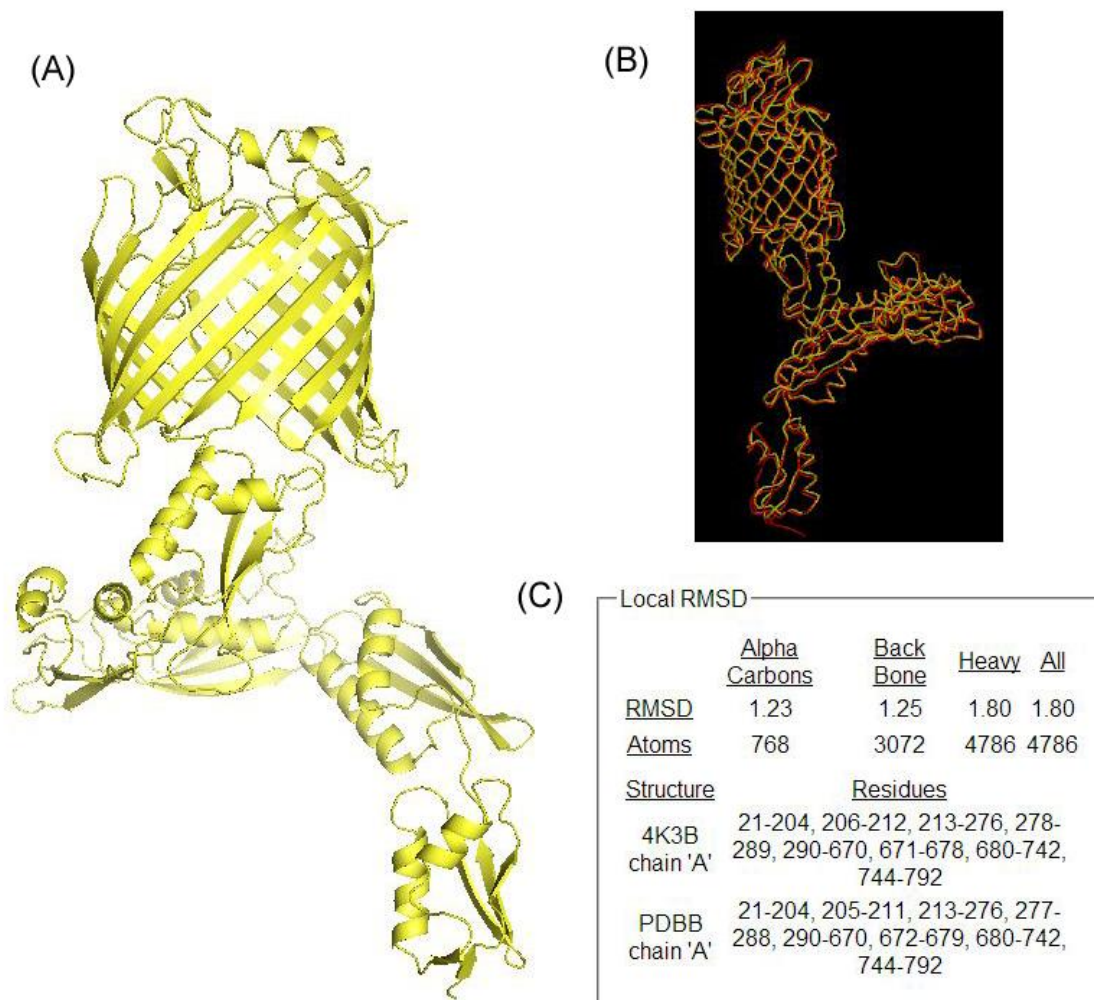


Figure 5-4 *E. coli* BamA Homology Model

(A) *E. coli* BamA homology model based on the known *Ng*BamA structure is shown in yellow cartoon with a transmembrane β -barrel and five POTRA domains. The structure is displayed in PyMOL. (B) Superposed image of the *E. coli* BamA homology model and *Ng*BamA by SuperPose. (C) Local RMSD values are calculated by SuperPose. The *E. coli* BamA homology model is represented as PDBB, while *Ng*BamA is 4K3NB.

5.4.2.1. Hydrogen Bonds Close the β -barrel

In the homology model, the transmembrane β -barrel is composed of 16 antiparallel β -strands, with the first and last strands associating together to close the barrel. The last β -strand (β 16) of the *Ng*BamA structure is bound to the first one (β 1) by

two hydrogen bonds (Noinaj et al. 2013) ; however, more than two hydrogen bonds are seen in the *E. coli* BamA homology model (Figure 5-5).

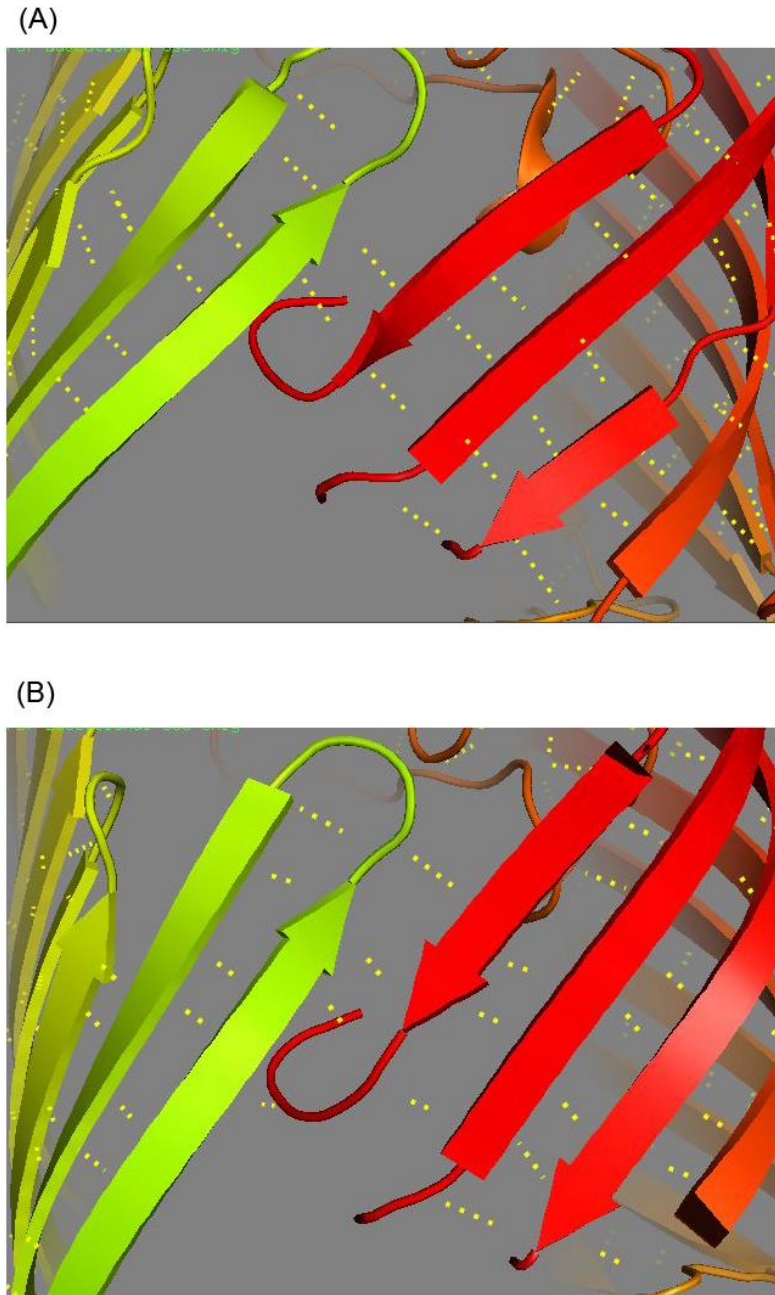


Figure 5-5 The Hydrogen Bonds Between the First and Last β -strands.

(A) The green and red strands located in the middle of the image show the β_1 and β_{16} of *NgBamA*'s β -barrel, respectively. The hydrogen bonds are indicated by yellow dashes. (B) The green and red strands located in the middle of the image show β_1

and β 16 of the *E.coli* BamA homology model's β -barrel, respectively. The hydrogen bonds are indicated by yellow dashes.

5.4.2.2. Loop 6 (L6) Conformations

Based on the solved structures of HdBamA Δ 3 and NgBamA, the interior of the barrel encompasses a volume of $\sim 13,000 \text{ \AA}^3$, with extracellular loops forming a dome to isolate the barrel's interior from the outside space (Noinaj et al. 2013) . Of these extracellular loops, loop 4 (L4), loop 6(L6) and loop 7(L7) make significant contributions to the dome formation, whereas loop 3(L3) and loop 8(L8) provide minor supports (Noinaj et al. 2013). β strands #11 and #12 are connected by an extracellular loop known as loop 6 (L6). Two conformations of L6 have been proposed in the literature to address the position of L6, directly related to either the protease-sensitive or the protease-resistant state, which may be induced by BamD and BamE (Rigel, Ricci, and Silhavy 2013b; Leonard-Rivera and Misra 2012). L6 contains residues which are highly conserved among the Omp85 superfamily and mutagenesis studies have demonstrate that the RGF residues of L6 are critical for biogenesis (Leonard-Rivera and Misra 2012) . The mutation of RGF to alanine may lead to a lethal phenotype by reducing BamA levels and causing folding defects (Leonard-Rivera and Misra 2012). The sequence alignment showed the VRGF motif was conserved among the three species (Figure 5-6), which suggesting the functional similarity is necessary. When L6 from the *E. coli* BamA homology model was aligned with NgBamA, the positions and conformations were consistent, except that one helix present in NgBamA's L6 was missing in the L6 from *E. coli* BamA homology model (Figure 5-7).

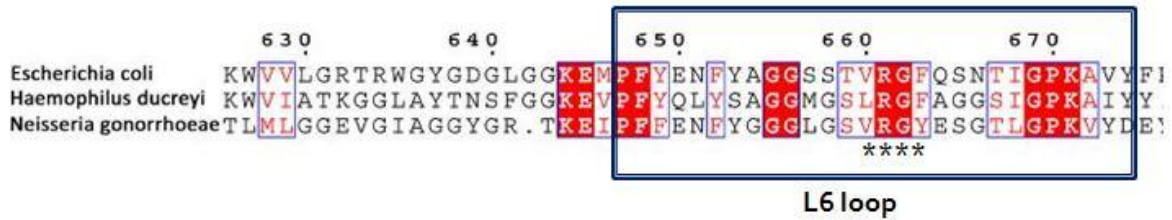


Figure 5-6 The L6 Sequence Alignment of BamA

The L6 loop amino acid sequence is included in the blue box. The conserved VRGF motif is marked with asterisks. The sequence alignment is analysed and generated using ESPript (Gouet et al. 1999). The amino acid sequence of *E. coli* BamA (UniProt: P0A940) aligned with *Ng*BamA (UniProt: Q5F5W8) and *Hd*BamA (UniProt: G1UBA5).

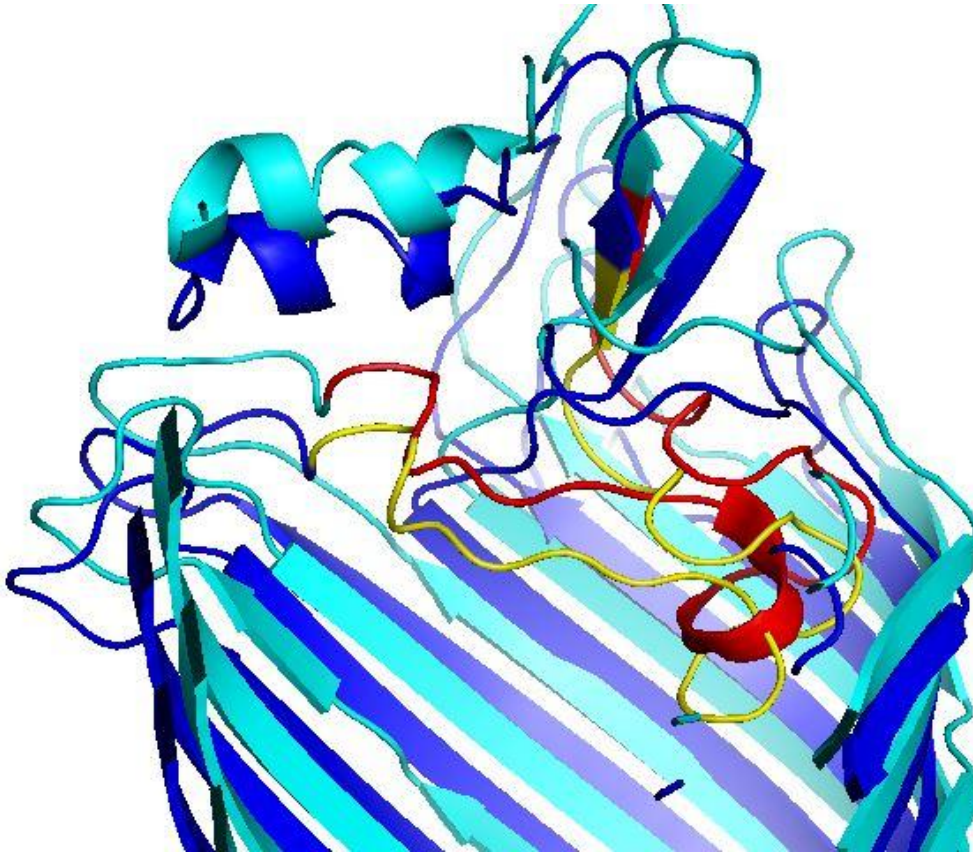


Figure 5-7 The L6 Conformations in BamA

The *E. coli* BamA homology model (shown in blue) is aligned with *Ng*BamA (shown in cyan). The L6 of *E. coli* BamA homology model is in yellow, compared to L6 of *Ng*BamA identified in red. Both L6 loops are partially inserted into the barrels. This cartoon diagram was prepared by PyMOL.

5.4.3. *E. coli* BamA Homology Model vs. Solved *E. coli* BamA structures

The full length *E. coli* BamA homology model based on NgBamA was superimposed on the known *E. coli* BamA structures : POTRA 1-4 in its extended form (PDB: 2QCZ), the β -barrel domain (PDB: 4N75) and the β -barrel domain with POTRA 5 (PDB: 4C4V).

With our full length homology model superimposed on POTRA 1-4 (PDB: 2QCZ), the global RMSD of backbones was calculated as 13.24 Å as shown in Figure 5-8. In addition, the full length homology model was superimposed on either the β -barrel domain structure (PDB: 4N75) or the structure of the β -barrel domain plus POTRA 5 (PDB: 4C4V), with the global RMSD of the backbones as 15.87 Å and 4 Å, respectively. The data shown in Figure 5-9 were comparable and indicated significant structural difference between our homology model and solved *E. coli* BamA structures in the POTRA 5 domain. All the sequence alignments of those superimposed structures are included in Appendix D.



Local RMSD				
	<u>Alpha</u> <u>Carbons</u>	<u>Back</u> <u>Bone</u>	<u>Heavy</u>	<u>All</u>
<u>RMSD</u>	6.10	5.88	6.43	6.43
<u>Atoms</u>	155	620	837	837
	<u>Structure</u>		<u>Residues</u>	
	PDBB chain 'A'		23-177	
	2QCZ chain 'A'		22-176	

Global RMSD				
	<u>Alpha</u> <u>Carbons</u>	<u>Back</u> <u>Bone</u>	<u>Heavy</u>	<u>All</u>
<u>RMSD</u>	13.22	13.24	13.54	13.54
<u>Atoms</u>	309	1236	2403	2403
	<u>Structure</u>		<u>Residues</u>	
	PDBB	22-198, 216-285, 287-289, 291-331, 334-351	chain 'A'	
	2QCZ	23-199, 217-286, 287-289, 291-331, 334-351	chain 'A'	

Figure 5-8 *E. coli* BamA Homology Model Superimposed to *E. coli* BamA POTRA 1-4

The superposition image shows the *E. coli* BamA homology model in red superimposed on the *E. coli* BamA POTRA 1-4 structure in its “extended” form (PDB: 2QCZ) in yellow. Both the superposition image and the calculated RMSD values were processed by SuperPose.

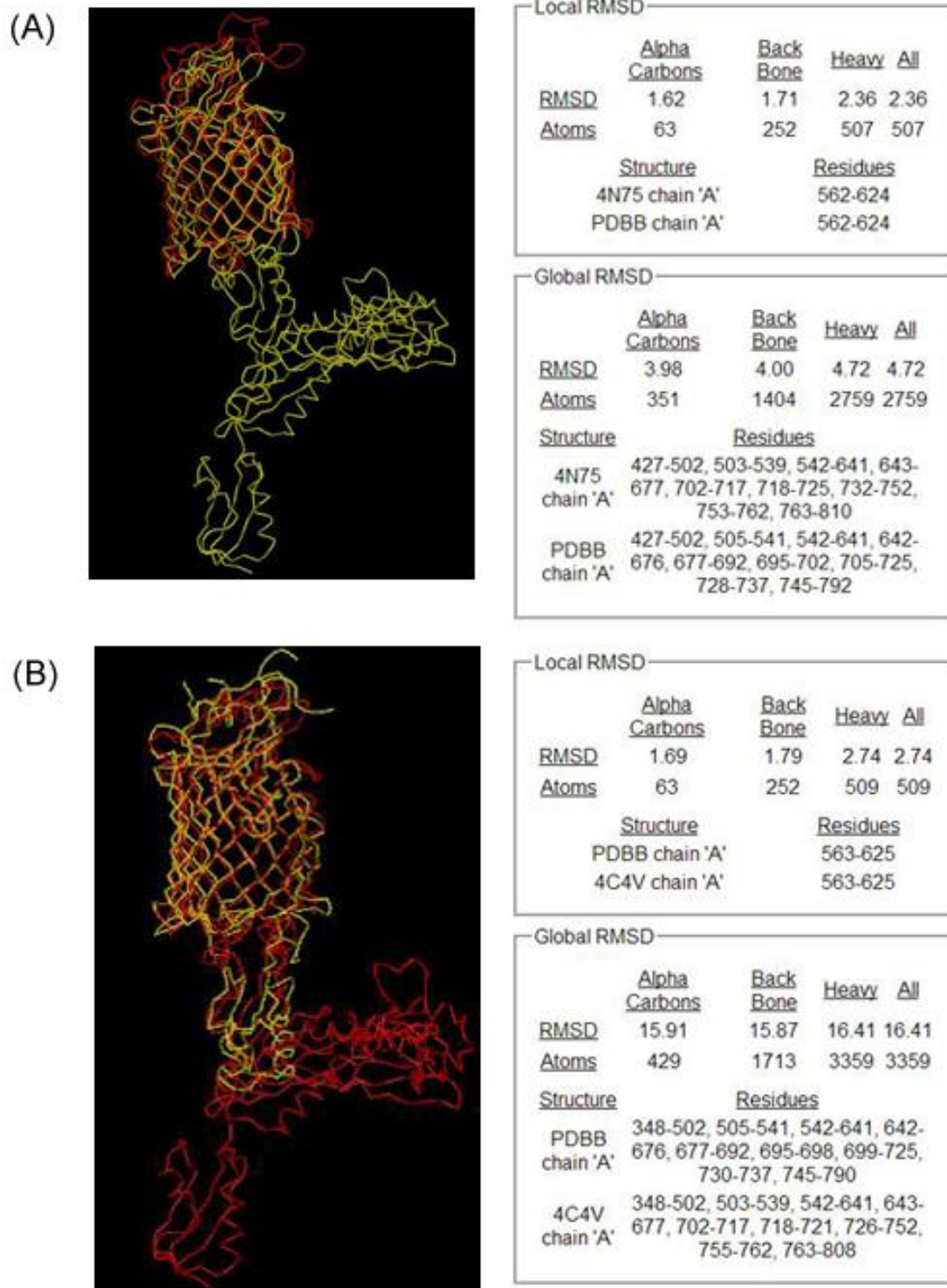


Figure 5-9 E. coli BamA Homology Model Superimposed to E. coli BamA β -barrel and E.coli BamA β -barrel with POTRA 5

Legend is on next page.

(A) The superposition image shows the *E. coli* BamA homology model in yellow superimposed on the *E. coli* BamA transmembrane β -barrel structure (PDB: 4N75) in red. Both the superposition image and the calculated RMSD values were processed by SuperPose. (B) The superposition image shows the *E. coli* BamA homology model in red superimposed on the *E. coli* BamA transmembrane β -barrel structure (PDB: 4C4V) in yellow. Both the superposition image and calculated RMSD values were processed by SuperPose.

5.5. Discussion

Full length *NgBamA* and truncated *H. ducreyi* BamA (*HdBamA* Δ 3 structures have both been recently solved (Noinaj et al. 2013) , providing structural insights to help in the designing of an *E. coli* BamA homology model. Before being used for homology modelling, the *N. gonorrhoeae*, *H. ducreyi* and *E. coli* sequences were aligned to identify the common domains and residues among those three species. Both *NgBamA* and *HdBamA* have 30% sequence identity, compared to *E. coli* BamA. Identical and similar residues were evenly distributed throughout the entire sequence of each species' BamA, suggesting that mutation of the different residues in the *NgBamA* template might not cause significant structural changes, which can lead to the failure of homology modelling.

To obtain full length *E. coli* BamA structure information, the solved structure of *NgBamA* was used as a template to arrange the *E. coli* BamA sequence in the same positions of alpha carbons, the phi and psi angles and secondary structure. This process was done in Coot. Next, the YASARA energy minimization server adjusted the side chain positions to minimize collisions and improve the model. The homology model starts with the same backbone as template structure. The model is then optimized in an arrangement that the net inter-atomic force on each atom is close to zero. Keeping in mind that a homology model is expected to differ from the template structure, even though the two proteins have a high level of sequence identity, and very similar secondary and tertiary structure. Overall differences in the protein backbone structures are quantitated with the root mean square deviation (RMSD) and if the RMSD is within a certain range, the homology model is considered accurate. To gain a frame of reference for RMSD, 1,200 homology models have been made using SWISS-MODEL for the resolved structures, and using a template with the most similar sequence available. This

3DCrunch project results suggest that proteins with 50% sequence identity have on average 1 Å RMSD (Schwede et al. 2003) . After changing those *NgBamA* residues differing from the *E. coli* BamA by Coot, the homology model RMSD values of the α carbon and backbone were calculated as 1.23 Å and 1.25 Å, respectively, using SuperPose (Maiti et al. 2004) . A value of up to 0.5 Å RMSD for the alpha carbons has been obtained in independent determinations of the same protein (Chothia and Lesk 1986) . Compared to the above data, our *E. coli* BamA homology model is considered accurate within the range that the literature suggests.

After being superimposed individually on *E. coli* POTRA and the transmembrane β -barrel domain structures, the transmembrane β -barrel domain of the full length homology model can be seen to have very similar architecture while its POTRA domains show flexibility compared to the real *E. coli* BamA structures. Based on the transmembrane β -barrel with POTRA 5 structure which has just been solved as of April 2014, the linker between POTRA 5 and the transmembrane β -barrel is long and flexible compared to that of the *NgBamA* and *HdBamA Δ 3* structures (Albrecht et al. 2014) . Two hydrogen bonds are found to close the β -barrel in this published structure, whereas more than two hydrogen bonds are seen in our homology model. The structural importance of the L6 loop is visible in both structures, facilitating the dome-forming function and acting as a gating mechanism for the hydrophobic cavity (Albrecht et al. 2014; Noinaj et al. 2013) .

Homology models are useful to provide information on where the alpha carbons of key residues reside within the folded protein. These models can guide future mutagenesis experiments, and suggest or support hypotheses about structure-function relationships. However, an actual full length *E.coli* BamA structure needs to be determined experimentally in order to obtain a more accurate structural and functional understanding of this protein.

Chapter 6. Summary and Future Directions

The research motivation of the BAM complex is often driven by its medical application. It is functionally non-redundant and essential for Gram-negative bacteria. The location of BAM complex on the OM implies a drug delivery strategy, so it has been studied as a potential drug target for novel antibiotics and vaccine development. The complex is found in the OM of all Gram-negative bacteria, with homologous systems in mitochondria and chloroplasts. It is non-redundant; there is no other back-up system in the cell that can perform the same function. Much progress has been made in recent years in understanding the structure and functions of the members of the BAM complex. Structural studies have not only shown us what each component of the BAM complex looks like, but have also provided clues as to the functional roles of each protein. The lipoprotein structures in this complex have been solved in previous studies, giving us insights into how the individual components may interact with each other. BamA, as the core component in the BAM complex, is an integral protein in the OM and plays primary role in OMP assembly. Based on the solved structures so far, researchers believe that the formation of crystal contacts is mediated by the POTRA domain, which are a critical feature in all BamA structures. Although the structures of *E. coli* POTRA domains and transmembrane β -barrel domain have been revealed separately, the full length structure still remains to be elucidated in order to learn the molecular mechanism of β -barrel assembly, the quaternary architecture of BamA and the order of recruiting other lipoproteins to be functional in the BAM complex.

In this thesis project, the *E. coli* BamA proteins were either extracted from the OM or expressed as inclusion bodies which were then refolded. Crystallizing membrane proteins routinely present challenges in designing suitable refolding strategies, since there are no standard refolding protocols. Those refolded BamA were then crystallized and optimized to obtain good quality crystals for solving structures by X-ray crystallography. Meanwhile, an *E.coli* BamA homology model was generated using the

solved full length *Ng*BamA structure as a template, providing suggestions of *E.coli* BamA overall structure and hypotheses about its structure-function relationships.

Future research efforts should include optimization of crystallization conditions to solve the full length *E.coli* BamA, attempt to crystallize nBamA Δ 1-20 and other mutagenesis structures. For instance, L6 is considered to be an essential extracellular loop in the β -barrel, related to the protease-sensitive or protease-inhibition properties of BamA. Resolving the L6 mutant BamA structure may address the functional specificity by revealing the loop position, while resolving the co-crystal structure of L6 mutated BamA with BamD/ BamE may also provide a structural explanation of how BamD/BamE mediates L6's conformational change in BamA. This co-crystallization study can be accompanied by binding kinetics studies to characterize BamD/BamE's interactions with BamA or L6 mutant BamA. X-ray crystallography studies are still needed to address such mechanistic questions, for example, how does BamA cooperate with the other four lipoproteins and how do the lipoproteins assist OMP assembly by BamA?

References

- Albrecht, R, M. Schutz, P. Oberhettinger, M. Faulstich, I. Bermejo, T. Rudel, K. Diederichs, and K. Zeth. 2014. "Structure of BamA, an Essential Factor in Outer Membrane Protein Biogenesis." *Acta Crystallographica Section D: Biological Crystallography* 70 (6): 1779-1789.
- Albrecht, R. and K. Zeth. 2011. "Structural Basis of Outer Membrane Protein Biogenesis in Bacteria." *The Journal of Biological Chemistry* 286 (31): 27792-27803.
- Anwari, K., C. T. Webb, S. Poggio, A. J. Perry, M. Belousoff, N. Celik, G. Ramm, A. Lovering, R. E. Sockett, and J. Smit. 2012. "The Evolution of New Lipoprotein Subunits of the Bacterial Outer Membrane BAM Complex." *Molecular Microbiology* 84 (5): 832-844.
- Bechtluft, P., N. Nouwen, S. J. Tans, and A. JM Driessen. 2010. "SecB—a Chaperone Dedicated to Protein Translocation." *Molecular BioSystems* 6 (4): 620-627.
- Bennion, D, Emily S. Charlson, E. Coon, and R. Misra. 2010. "Dissection of B-barrel Outer Membrane Protein Assembly Pathways through Characterizing BamA POTRA 1 Mutants of Escherichia Coli." *Molecular Microbiology* 77 (5): 1153-1171.
- Bishop, R. E. 2005. "The Lipid A Palmitoyltransferase PagP: Molecular Mechanisms and Role in Bacterial Pathogenesis." *Molecular Microbiology* 57 (4): 900-912.
- Bitto, E. and D. B. McKay. 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 (49): 49316-49322.
- Blow, D. 2003. "Outline of Crystallography for Biologists."
- Bos, M. P., V. Robert, and J. Tommassen. 2007. "Functioning of Outer Membrane Protein Assembly Factor Omp85 Requires a Single POTRA Domain." *EMBO Reports* 8 (12): 1149-1154.
- Buchanan, S. K. 1999. "B-Barrel Proteins from Bacterial Outer Membranes: Structure, Function and Refolding." *Current Opinion in Structural Biology* 9 (4): 455-461.
- Charlson, E. S., J. N. Werner, and R. Misra. 2006. "Differential Effects of yfgL Mutation on Escherichia Coli Outer Membrane Proteins and Lipopolysaccharide." *Journal of Bacteriology* 188 (20): 7186-7194.

- Chayen, N. E. and E. Saridakis. 2008. "Protein Crystallization: From Purified Protein to Diffraction-Quality Crystal." *Nature Methods* 5 (2): 147-153.
- Chothia, C. and A. M. Lesk. 1986. "The Relation between the Divergence of Sequence and Structure in Proteins." *The EMBO Journal* 5 (4): 823-826.
- Clantin, B., A. S. Delattre, P. Rucktooa, N. Saint, A. C. Meli, C. Locht, F. Jacob-Dubuisson, and V. Villeret. 2007a. "Structure of the Membrane Protein FhaC: A Member of the Omp85-TpsB Transporter Superfamily." *Science (New York, N.Y.)* 317 (5840): 957-961.
- Collins, R. F. and J. P. Derrick. 2007. "Wza: A New Structural Paradigm for Outer Membrane Secretory Proteins?" *Trends in Microbiology* 15 (3): 96-100.
- Costerton, J. W., J. M. Ingram, and K. J. Cheng. 1974. "Structure and Function of the Cell Envelope of Gram-Negative Bacteria." *Bacteriological Reviews* 38 (1): 87-110.
- Cross, B. CS, I. Sinning, J. Luirink, and S. High. 2009. "Delivering Proteins for Export from the Cytosol." *Nature Reviews Molecular Cell Biology* 10 (4): 255-264.
- Dekker, N., J. Tommassen, A. Lustig, J. P. Rosenbusch, and H. M. Verheij. 1997. "Dimerization Regulates the Enzymatic Activity of Escherichia Coli Outer Membrane Phospholipase A." *The Journal of Biological Chemistry* 272 (6): 3179-3184.
- Delcour, A. H. 2009. "Outer Membrane Permeability and Antibiotic Resistance." *Biochimica Et Biophysica Acta (BBA)-Proteins and Proteomics* 1794 (5): 808-816.
- Delcour, A. H. 2003. "Solute Uptake through General Porins." *Frontiers in Bioscience : A Journal and Virtual Library* 8: d1055-71.
- Desvaux, M., N. J. Parham, and I. R. Henderson. 2004. "The Autotransporter Secretion System." *Research in Microbiology* 155 (2): 53-60.
- Diedrich, D. L. and E. H. Cota-Robles. 1974. "Heterogeneity in Lipid Composition of the Outer Membrane and Cytoplasmic Membrane and Cytoplasmic Membrane of Pseudomonas BAL-31." *Journal of Bacteriology* 119 (3): 1006-1018.
- Dong, C., K. Beis, J. Nesper, A. L. Brunkan-LaMontagne, B. R. Clarke, C. Whitfield, and James H. Naismith. 2006. "Wza the Translocon for E. Coli Capsular Polysaccharides Defines a New Class of Membrane Protein." *Nature* 444 (7116): 226-229.
- Driessen, A. JM and N. Nouwen. 2008. "Protein Translocation Across the Bacterial Cytoplasmic Membrane." *Annu.Rev.Biochem.* 77: 643-667.
- Emsley, P. and K. Cowtan. 2004. "Coot: Model-Building Tools for Molecular Graphics." *Acta Crystallographica Section D: Biological Crystallography* 60 (12): 2126-2132.

- Fairman, J. W., N. Noinaj, and S. K. Buchanan. 2011. "The Structural Biology of B-Barrel Membrane Proteins: A Summary of Recent Reports." *Current Opinion in Structural Biology* 21 (4): 523-531.
- Fink, A. L. 1998. "Protein Aggregation: Folding Aggregates, Inclusion Bodies and Amyloid." *Folding and Design* 3 (1): R9-R23.
- Gasteiger, E., C. Hoogland, A. Gattiker, M. R. Wilkins, R. D. Appel, and A. Bairoch. 2005. "Protein Identification and Analysis Tools on the ExPASy Server." In *The Proteomics Protocols Handbook*, 571-607: Springer.
- Gatzeva-T., P. Zvezdanova, L. R. Warner, A. Pardi, and M. C. Sousa. 2010. "Structure and Flexibility of the Complete Periplasmic Domain of BamA: The Protein Insertion Machine of the Outer Membrane." *Structure* 18 (11): 1492-1501.
- Genevrois, S., L. Steeghs, P. Roholl, J. Letesson, and P. Ley. 2003. "The Omp85 Protein of *Neisseria Meningitidis* is Required for Lipid Export to the Outer Membrane." *The EMBO Journal* 22 (8): 1780-1789.
- Gentle, I., K. Gabriel, P. Beech, R. Waller, and T. Lithgow. 2004. "The Omp85 Family of Proteins is Essential for Outer Membrane Biogenesis in Mitochondria and Bacteria." *The Journal of Cell Biology* 164 (1): 19-24.
- Gierasch, L. M. 1989. "Signal Sequences." *Biochemistry* 28 (3): 923-930.
- Gouet, P., E. Courcelle, D. I. Stuart, and F. Metoz. 1999. "ESPrpt: Analysis of Multiple Sequence Alignments in PostScript." *Bioinformatics (Oxford, England)* 15 (4): 305-308.
- Greenfield, N. J. 2007. "Using Circular Dichroism Spectra to Estimate Protein Secondary Structure." *Nature Protocols* 1 (6): 2876-2890.
- Hagan, C. L., T. J. Silhavy, and D. Kahne. 2011. "B-Barrel Membrane Protein Assembly by the Bam Complex." *Annual Review of Biochemistry* 80: 189-210.
- Hagan, C. L., S. Kim, and D. Kahne. 2010. "Reconstitution of Outer Membrane Protein Assembly from Purified Components." *Science (New York, N.Y.)* 328 (5980): 890-892.
- Heller, K. B. 1978. "Apparent Molecular Weights of a Heat-Modifiable Protein from the Outer Membrane of *Escherichia Coli* in Gels with Different Acrylamide Concentrations." *Journal of Bacteriology* 134 (3): 1181-1183.
- Heuck, A., A. Schleiffer, and T. Clausen. 2011. "Augmenting B-Augmentation: Structural Basis of how BamB Binds BamA and may Support Folding of Outer Membrane Proteins." *Journal of Molecular Biology* 406 (5): 659-666.
- Hoffmann, A., B. Bukau, and G. Kramer. 2010. "Structure and Function of the Molecular Chaperone Trigger Factor." *Biochimica Et Biophysica Acta (BBA)-Molecular Cell Research* 1803 (6): 650-661.

- Jacob, D. F., C. Locht, and R. Antoine. 2001. "Two-partner Secretion in Gram-negative Bacteria: A Thrifty, Specific Pathway for Large Virulence Proteins." *Molecular Microbiology* 40 (2): 306-313.
- Kelly, S. M., T. J. Jess, and N. C. Price. 2005. "How to Study Proteins by Circular Dichroism." *Biochimica Et Biophysica Acta (BBA)-Proteins and Proteomics* 1751 (2): 119-139.
- Kiefhaber, T., R. Rudolph, H. Kohler, and J. Buchner. 1991. "Protein Aggregation in Vitro and in Vivo: A Quantitative Model of the Kinetic Competition between Folding and Aggregation." *Nature Biotechnology* 9 (9): 825-829.
- Kim, K. H., S. Aulakh, and M. Paetzel. 2012. "The Bacterial Outer Membrane B-barrel Assembly Machinery." *Protein Science* 21 (6): 751-768.
- Kim, K. H., S. Aulakh, W. Tan, and M. Paetzel. 2011a. "Crystallographic Analysis of the C-Terminal Domain of the *Escherichia Coli* Lipoprotein BamC." *Acta Crystallographica Section F: Structural Biology and Crystallization Communications* 67 (11): 1350-1358.
- Kim, H., H. Kang, M. Okon, Eric Escobar-Cabrera, Lawrence P. McIntosh, and Mark Paetzel. 2011b. "Structural Characterization of *Escherichia Coli* BamE, a Lipoprotein Component of the B-Barrel Assembly Machinery Complex." *Biochemistry* 50 (6): 1081-1090.
- Kim, K. H. and M. Paetzel. 2011. "Crystal Structure of *Escherichia Coli* BamB, a Lipoprotein Component of the B-Barrel Assembly Machinery Complex." *Journal of Molecular Biology* 406 (5): 667-678.
- Kim, K. H., S. Aulakh, and M. Paetzel. 2011. "Crystal Structure of Beta-Barrel Assembly Machinery BamCD Protein Complex." *The Journal of Biological Chemistry* 286 (45): 39116-39121.
- Kim, S., J. C. Malinverni, P. Sliz, T. J. Silhavy, S. C. Harrison, and D. Kahne. 2007. "Structure and Function of an Essential Component of the Outer Membrane Protein Assembly Machine." *Science (New York, N.Y.)* 317 (5840): 961-964.
- Knowles, T. J., Douglas F. Browning, M. Jeeves, R. Maderbocus, S. Rajesh, P. Sridhar, E. Manoli, D. Emery, U. Sommer, and A. Spencer. 2011. "Structure and Function of BamE within the Outer Membrane and the B-barrel Assembly Machine." *EMBO Reports* 12 (2): 123-128.
- Knowles, T. J., M. Jeeves, S. Bobat, F. Dancea, D. McClelland, T. Palmer, M. Overduin, and I. R. Henderson. 2008. "Fold and Function of Polypeptide Transport-associated Domains Responsible for Delivering Unfolded Proteins to Membranes." *Molecular Microbiology* 68 (5): 1216-1227.
- Koebnik, R., K. P. Locher, and P. Van Gelder. 2000. "Structure and Function of Bacterial Outer Membrane Proteins: Barrels in a Nutshell." *Molecular Microbiology* 37 (2): 239-253.

- Kopito, R. R. 2000. "Aggresomes, Inclusion Bodies and Protein Aggregation." *Trends in Cell Biology* 10 (12): 524-530.
- Koronakis, V., A. Sharff, E. Koronakis, B. Luisi, and C. Hughes. 2000. "Crystal Structure of the Bacterial Membrane Protein TolC Central to Multidrug Efflux and Protein Export." *Nature* 405 (6789): 914-919.
- Krieger, E., G. Vriend, and C. Spronk. 2013. "YASARA—Yet another Scientific Artificial Reality Application." *YASARA.Org* 993.
- Leonard-Rivera, M. and R. Misra. 2012. "Conserved Residues of the Putative L6 Loop of Escherichia Coli BamA Play a Critical Role in the Assembly of Beta-Barrel Outer Membrane Proteins, Including that of BamA Itself." *Journal of Bacteriology* 194 (17): 4662-4668.
- Maiti, R., G. H. Van Domselaar, H. Zhang, and D. S. Wishart. 2004. "SuperPose: A Simple Server for Sophisticated Structural Superposition." *Nucleic Acids Research* 32 (Web Server issue): W590-4.
- Malinverni, J. C., J. Werner, S. Kim, J. G. Sklar, D. Kahne, R. Misra, and T. J. Silhavy. 2006. "YfiO Stabilizes the YaeT Complex and is Essential for Outer Membrane Protein Assembly in Escherichia Coli." *Molecular Microbiology* 61 (1): 151-164.
- Miot, M. and J. M. Betton. 2004. "Protein Quality Control in the Bacterial Periplasm." *Microbial Cell Factories* 3 (1): 4.
- Nakamura, K. and S. Mizushima. 1976. "Effects of Heating in Dodecyl Sulfate Solution on the Conformation and Electrophoretic Mobility of Isolated Major Outer Membrane Proteins from Escherichia Coli K-12." *Journal of Biochemistry* 80 (6): 1411-1422.
- Ni, D., Y. Wang, X. Yang, H. Zhou, X. Hou, B. Cao, Z. Lu, X. Zhao, K. Yang, and Y. Huang. 2014. "Structural and Functional Analysis of the Beta-Barrel Domain of BamA from Escherichia Coli." *FASEB Journal : Official Publication of the Federation of American Societies for Experimental Biology* 28 (6): 2677-2685.
- Nikaido, H. 2003a. "Molecular Basis of Bacterial Outer Membrane Permeability Revisited." *Microbiology and Molecular Biology Reviews : MMBR* 67 (4): 593-656.
- Nikaido, H. 1994. "Porins and Specific Diffusion Channels in Bacterial Outer Membranes." *The Journal of Biological Chemistry* 269 (6): 3905-3908.
- Noinaj, N., A. J. Kuszak, J. C. Gumbart, P. Lukacik, H. Chang, N. C. Easley, T. Lithgow, and S. K. Buchanan. 2013. "Structural Insight into the Biogenesis of [Bgr]-Barrel Membrane Proteins." *Nature* 501 (7467): 385-390.
- Okuda, S. and H. Tokuda. 2011. "Lipoprotein Sorting in Bacteria." *Annual Review of Microbiology* 65: 239-259.

- Onufryk, C., M. L. Crouch, F. C. Fang, and C. A. Gross. 2005. "Characterization of Six Lipoproteins in the sigmaE Regulon." *Journal of Bacteriology* 187 (13): 4552-4561.
- Oreb, M., A. Hofle, O. Mirus, and E. Schleiff. 2008. "Phosphorylation Regulates the Assembly of Chloroplast Import Machinery." *Journal of Experimental Botany* 59 (9): 2309-2316.
- Paetzl, M., A. Karla, N. CJ Strynadka, and R. E. Dalbey. 2002. "Signal Peptidases." *Chemical Reviews* 102 (12): 4549-4580.
- Pautsch, A. and G. E. Schulz. 1998. "Structure of the Outer Membrane Protein A Transmembrane Domain." *Nature Structural & Molecular Biology* 5 (11): 1013-1017.
- Phan, G., H. Remaut, T. Wang, W.J. Allen, K. F. Pirker, A. Lebedev, N. S. Henderson, S. Geibel, E. Volkan, and J. Yan. 2011. "Crystal Structure of the FimD Usher Bound to its Cognate FimC-FimH Substrate." *Nature* 474 (7349): 49-53.
- Qu, J., C. Mayer, S. Behrens, O. Holst, and J. H. Kleinschmidt. 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 (1): 91-105.
- Ricci, D. P. and T. J. Silhavy. 2012. "The Bam Machine: A Molecular Cooper." *Biochimica Et Biophysica Acta (BBA)-Biomembranes* 1818 (4): 1067-1084.
- Ridgway, N., R. McLeod, JE Vance, and D. E. Vance. 2008. *Biochemistry of Lipids, Lipoproteins and Membranes* Elsevier.
- Rigel, N. W. and T. J. Silhavy. 2012. "Making a Beta-Barrel: Assembly of Outer Membrane Proteins in Gram-Negative Bacteria." *Current Opinion in Microbiology* 15 (2): 189-193.
- Rigel, N. W., D. P. Ricci, and T. J. Silhavy. 2013a. "Conformation-Specific Labeling of BamA and Suppressor Analysis Suggest a Cyclic Mechanism for Beta-Barrel Assembly in *Escherichia Coli*." *Proceedings of the National Academy of Sciences of the United States of America* 110 (13): 5151-5156.
- Rigel, N. W., J. Schwalm, D. P. Ricci, and T. J. Silhavy. 2012. "BamE Modulates the *Escherichia Coli* Beta-Barrel Assembly Machine Component BamA." *Journal of Bacteriology* 194 (5): 1002-1008.
- Rizzitello, A. E., J. R. Harper, and T. J. Silhavy. 2001. "Genetic Evidence for Parallel Pathways of Chaperone Activity in the Periplasm of *Escherichia Coli*." *Journal of Bacteriology* 183 (23): 6794-6800.
- Robert, V., E.B. Volokhina, F. Senf, M. P. Bos, P. V. Gelder, and J. Tommassen. 2006. "Assembly Factor Omp85 Recognizes its Outer Membrane Protein Substrates by a Species-Specific C-Terminal Motif." *PLoS Biology* 4 (11): e377.

- Rosenbusch, J. P. 1974. "Characterization of the Major Envelope Protein from Escherichia Coli. Regular Arrangement on the Peptidoglycan and Unusual Dodecyl Sulfate Binding." *The Journal of Biological Chemistry* 249 (24): 8019-8029.
- Ruiz, N., T. Wu, D.I Kahne, and T. J. Silhavy. 2006. "Probing the Barrier Function of the Outer Membrane with Chemical Conditionality." *ACS Chemical Biology* 1 (6): 385-395.
- Sánchez-Pulido, L., D. Devos, S. Genevrois, M.I Vicente, and A. Valencia. 2003. "POTRA: A Conserved Domain in the FtsQ Family and a Class of B-Barrel Outer Membrane Proteins." *Trends in Biochemical Sciences* 28 (10): 523-526.
- Sandoval, C. M., S. L. Baker, K. Jansen, S.I. Metzner, and M. C. Sousa. 2011. "Crystal Structure of BamD: An Essential Component of the B-Barrel Assembly Machinery of Gram-Negative Bacteria." *Journal of Molecular Biology* 409 (3): 348-357.
- Schatz, G. and B. Dobberstein. 1996. "Common Principles of Protein Translocation Across Membranes." *Science (New York, N. Y.)* 271 (5255): 1519-1526.
- Schleiff, E. and J. Soll. 2005. "Membrane Protein Insertion: Mixing Eukaryotic and Prokaryotic Concepts." *EMBO Reports* 6 (11): 1023-1027.
- Schulz, G. E. 2002. "The Structure of Bacterial Outer Membrane Proteins." *Biochimica Et Biophysica Acta (BBA)-Biomembranes* 1565 (2): 308-317.
- Schwede, T., J. Kopp, N. Guex, and M. C. Peitsch. 2003. "SWISS-MODEL: An Automated Protein Homology-Modeling Server." *Nucleic Acids Research* 31 (13): 3381-3385.
- Silhavy, T. J., D. Kahne, and S. Walker. 2010. "The Bacterial Cell Envelope." *Cold Spring Harbor Perspectives in Biology* 2 (5): a000414.
- Sklar, J. G., T. Wu, L. S. Gronenberg, J. C. Malinverni, D. Kahne, and T. J. Silhavy. 2007a. "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 (15): 6400-6405.
- Sklar, J. G., T. Wu, D. Kahne, and T. J. Silhavy. 2007c. "Defining the Roles of the Periplasmic Chaperones SurA, Skp, and DegP in Escherichia Coli." *Genes & Development* 21 (19): 2473-2484.
- Smyth, M. S. and J. H. Martin. 2000a. "X Ray Crystallography." *Molecular Pathology : MP* 53 (1): 8-14.
- Snijder, HJ, I. Ubarretxena-Belandia, M. Blaauw, KH Kalk, HM Verheij, MR Egmond, N. Dekker, and BW Dijkstra. 1999. "Structural Evidence for Dimerization-Regulated Activation of an Integral Membrane Phospholipase." *Nature* 401 (6754): 717-721.

- Sommer, M. S., B. Daum, L. E. Gross, B. L. Weis, O. Mirus, L. Abram, U. G. Maier, W. Kuhlbrandt, and E. Schleiff. 2011. "Chloroplast Omp85 Proteins Change Orientation during Evolution." *Proceedings of the National Academy of Sciences of the United States of America* 108 (33): 13841-13846.
- Stumpe, S., R. Schmid, D. L. Stephens, G. Georgiou, and E. P. Bakker. 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 (15): 4002-4006.
- Surrey, T. and F. Jahning. 1992a. "Refolding and Oriented Insertion of a Membrane Protein into a Lipid Bilayer." *Proceedings of the National Academy of Sciences of the United States of America* 89 (16): 7457-7461.
- Tamm, L. K., H. Hong, and B. Liang. 2004a. "Folding and Assembly of B-Barrel Membrane Proteins." *Biochimica Et Biophysica Acta (BBA)-Biomembranes* 1666 (1): 250-263.
- Tommassen, J. 2007. "Biochemistry. Getting into and through the Outer Membrane." *Science (New York, N.Y.)* 317 (5840): 903-904.
- Tsumoto, K., D. Ejima, I. Kumagai, and T. Arakawa. 2003. "Practical Considerations in Refolding Proteins from Inclusion Bodies." *Protein Expression and Purification* 28 (1): 1-8.
- Van Ulsen, P. 2011. "Protein Folding in Bacterial Adhesion: Secretion and Folding of Classical Monomeric Autotransporters." In *Bacterial Adhesion*, 125-142: Springer.
- Vandeputte-Rutten, L., R. A. Kramer, J. Kroon, N. Dekker, M. R. Egmond, and P. Gros. 2001. "Crystal Structure of the Outer Membrane Protease OmpT from *Escherichia Coli* Suggests a Novel Catalytic Site." *The EMBO Journal* 20 (18): 5033-5039.
- Vogt, J. and G. E. Schulz. 1999. "The Structure of the Outer Membrane Protein OmpX from *Escherichia Coli* Reveals Possible Mechanisms of Virulence." *Structure* 7 (10): 1301-1309.
- Vollmer, W. and U. Bertsche. 2008. "Murein (Peptidoglycan) Structure, Architecture and Biosynthesis in *Escherichia Coli*." *Biochimica Et Biophysica Acta (BBA)-Biomembranes* 1778 (9): 1714-1734.
- Von Heijne, G. 1985. "Signal Sequences: The Limits of Variation." *Journal of Molecular Biology* 184 (1): 99-105.
- Voulhoux, R., M. P. Bos, J. Geurtsen, M. Mols, and J. Tommassen. 2003a. "Role of a Highly Conserved Bacterial Protein in Outer Membrane Protein Assembly." *Science (New York, N.Y.)* 299 (5604): 262-265.

- Walther, D. M., D. Rapoport, and J. Tommassen. 2009. "Biogenesis of β -Barrel Membrane Proteins in Bacteria and Eukaryotes: Evolutionary Conservation and Divergence." *Cellular and Molecular Life Sciences* 66 (17): 2789-2804.
- Weber, K. and M. Osborn. 1969. "The Reliability of Molecular Weight Determinations by Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis." *The Journal of Biological Chemistry* 244 (16): 4406-4412.
- Wilson, M. E. 1991. "The Heat-Modifiable Outer Membrane Protein of *Actinobacillus Actinomycetemcomitans*: Relationship to OmpA Proteins." *Infection and Immunity* 59 (7): 2505-2507.
- Wimley, W. C. 2003. "The Versatile β -Barrel Membrane Protein." *Current Opinion in Structural Biology* 13 (4): 404-411.
- Wlodawer, A., W. Minor, Z. Dauter, and M. Jaskolski. 2008. "Protein Crystallography for Non-crystallographers, Or how to Get the Best (but Not More) from Published Macromolecular Structures." *Febs Journal* 275 (1): 1-21.
- Wu, T., J. Malinverni, N. Ruiz, S. Kim, T. J. Silhavy, and D. Kahne. 2005a. "Identification of a Multicomponent Complex Required for Outer Membrane Biogenesis in *Escherichia Coli*." *Cell* 121 (2): 235-245.
- Yen, M. R., C. R. Peabody, S.M. Partovi, Y. Zhai, Y. Tseng, and M. H. Saier Jr. 2002. "Protein-Translocating Outer Membrane Porins of Gram-Negative Bacteria." *Biochimica Et Biophysica Acta (BBA)-Biomembranes* 1562 (1): 6-31.
- Zgurskaya, H. I., G. Krishnamoorthy, A. Ntrel, and S. Lu. 2011. "Mechanism and Function of the Outer Membrane Channel TolC in Multidrug Resistance and Physiology of Enterobacteria." *Frontiers in Microbiology* 2: 189.
- Ziegler, K., R. Benz, and G. E. Schulz. 2008. "A Putative α -Helical Porin from *Corynebacterium Glutamicum*." *Journal of Molecular Biology* 379 (3): 482-491.
- Zimmer, J., Y. Nam, and T. A. Rapoport. 2008. "Structure of a Complex of the ATPase SecA and the Protein-Translocation Channel." *Nature* 455 (7215): 936-943.

Appendices

Appendix A. Cloning Details

The cloning strategy for each protein construct used is outlined in Table A. These constructs are variations of the *E. coli* BamA (UniProt ID: P0A940). The subsequent DNA sequencing results (Macrogen) confirmed that the gene inserts cloned into the vectors matched the sequences reported.

Table A Cloning details of the constructs used in this project

Construct	Vector (antibiotic)	Expression Host	Forward Primer (5'-3')	Reverse Primer (5'-3')	Restriction Enzyme
BamA Δ 1-424	pET-28a(+) (Kan ^R)	BL21(DE3)	TATAGGATC CAGCTTCAA CTT TGG T	ATATAGAATTC TTATTACCAG GTTTTA	BamHI/EcoRI
nBamA Δ 1-20	pET-20b(+) (Amp ^R)	BL21(DE3)	TATAGTCGAC GCTGAAGGGT TC	ATATAGAATTC TTATTACCAG GTTTTA	Sall/NotI

Appendix B. Molecular Mass Standard Curve for Size-exclusion Chromatography

Molecular Mass Calculation Based on the Elution Volume

The column used has a total column volume of 124 mL. Based on the elution volume of Blue Dextran, the void volume was determined to be 45 mL. Using the elution volumes of the rest of the standards, the partition coefficient, K_{av} , was calculated as follows:

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

Where V_o (void volume) was 45mL and V_t (column volume) was 124mL. V_e was the elution volume.

MW indicates molecular mass. The resulting K_{av} vs. $\text{Log}(\text{MW})$ plot produced the equation $\text{Log}(\text{MW}) = -3.6177(K_{av}) + 3.1051$ (Figure A1), which can be rewritten as:

$$MW = 10^{(-3.6177 \times K_{av} + 3.1051)} \quad \text{(Equation 2)}$$

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

Table B Protein Standards Used for Standard Curve.

Standard	Molecular Mass (kDa)	Elution Volume (mL)
Ovabumin	43	77.6
Aldolase	158	63.5
Ferritin	440	55.4
Thyroglobulin	669	51.5
Blue Dextran	2000	45*

*The elution volume of Blue Dextran was used to determine the void volume (V_0 = the volume of the mobile phase in the column) as its size is much larger than what this matrix can retain.

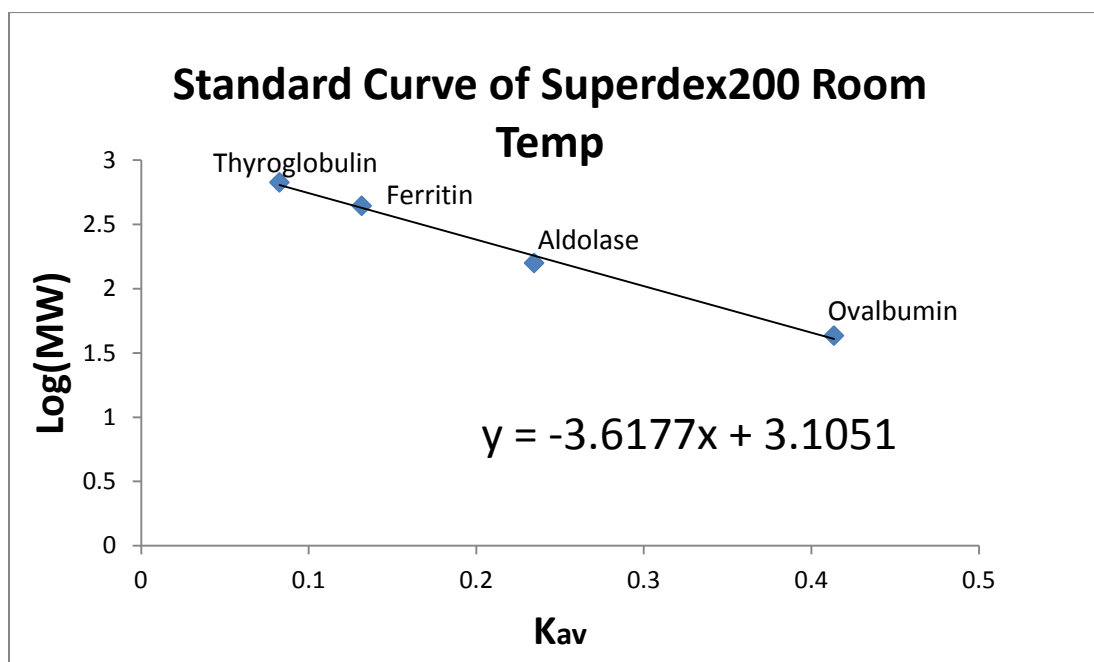


Figure B1 Standard Curve of the Superdex200 Room Temp Gel-filtration Column

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

Column: HiLoad 16/60 Superdex200 Column attached to the AKTA prime system

Buffer:

20mM Tris-HCl pH8.0, 100mM NaCl, 0.01% DDM, 2mM DTT

50mM Tris-HCl pH7.5, 200mM NaCl, 0.01% DDM, 2mM DTT

Temperature: Room temperature (~22°C)

Flow Rate: 1mL/min

Standards: From Amersham Biosciences' LMW Calibration Kit (See Table A)

Concentration: 5mg/mL of standard proteins were used except for Blue Dextran for which 1mg/mL. Each standard was run individually.

Appendix C.

CD Spectrum of *E. coli* nBamA Δ 1-20 and BamA Δ 1-20

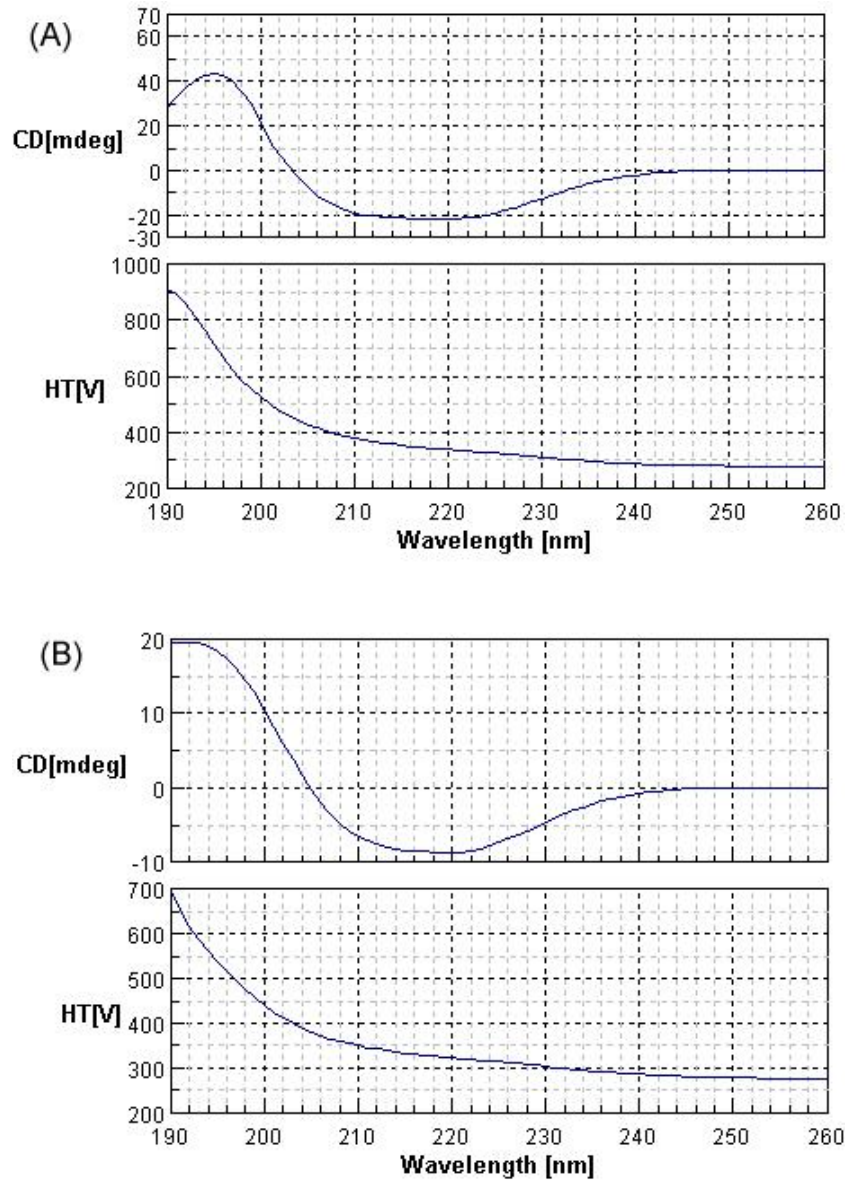


Figure C1 CD Spectra of *E. coli* nBamA Δ 1-20 and BamA Δ 1-20 (Repeat 1)

(A) The upper panel shows the far UV CD spectra for *E. coli* nBamA Δ 1-20. The protein concentration is 10.6 μ M. The measurements were taken between wavelengths 190 to 260nm. The lower panel shows the corresponding high tension voltage trace. (B) The upper panel shows the far UV CD spectra for the refolded *E. coli* BamA Δ 1-20. The

protein concentration is 5.31 μM . The measurements were taken between wavelengths 190 to 260nm. The lower panel shows the corresponding high tension voltage trace.

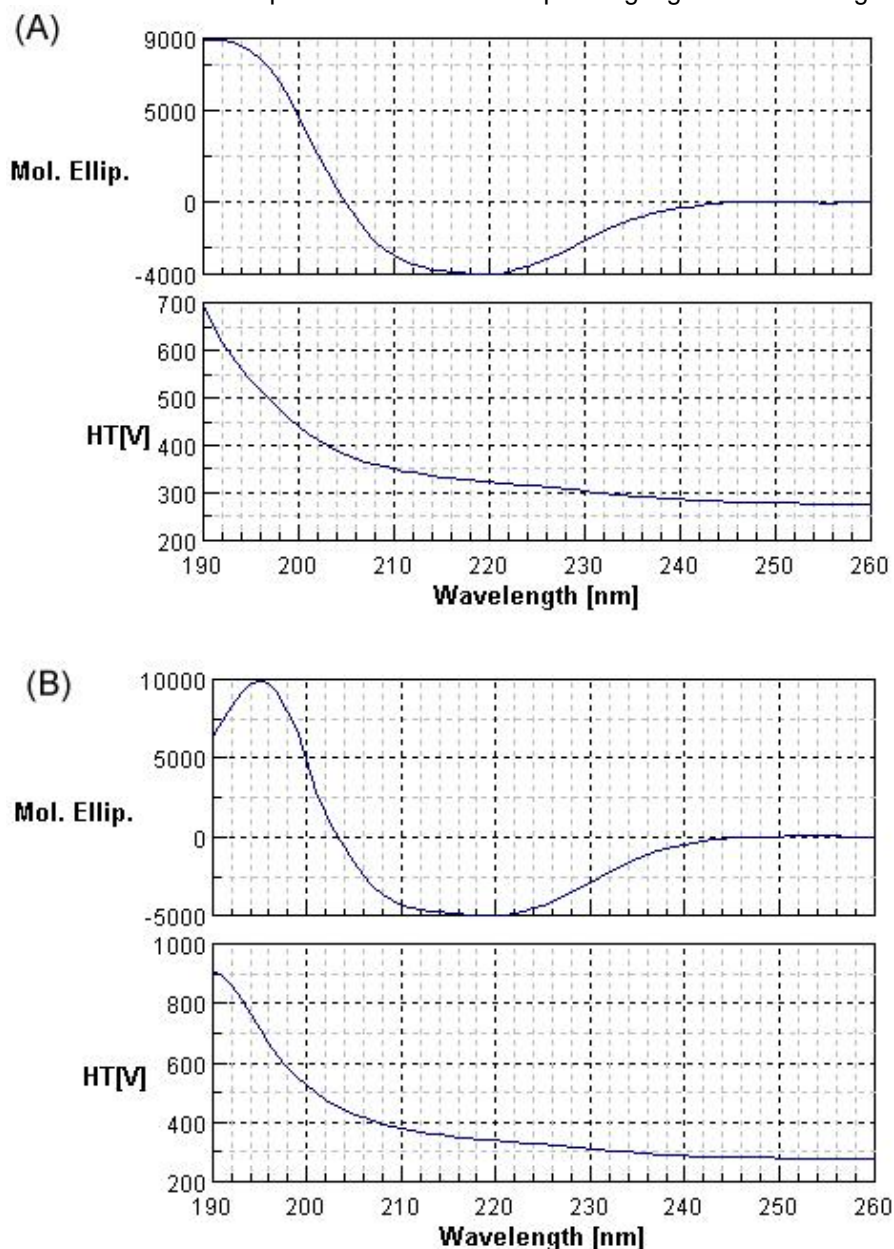


Figure C2 CD Spectra of *E. coli* nBamA Δ 1-20 and BamA Δ 1-20 Converted to Molar Ellipticity (Repeat 1)

(A) The CD absorbance readings of *E. coli* nBamA Δ 1-20 have been converted to molar ellipticity $[\theta]_{\text{molar},\lambda}$ at wavelength λ (190 nm ~ 260 nm). The high tension voltage trace is shown in the lower panel. (B) The CD absorbance readings of refolded *E. coli* BamA Δ 1-20 have been converted to molar ellipticity $[\theta]_{\text{molar},\lambda}$ at wavelength λ (190 nm ~ 260 nm), with unit millidegree $\text{cm}^2\text{dmol}^{-1}$. The high tension voltage trace is shown in the lower panel.

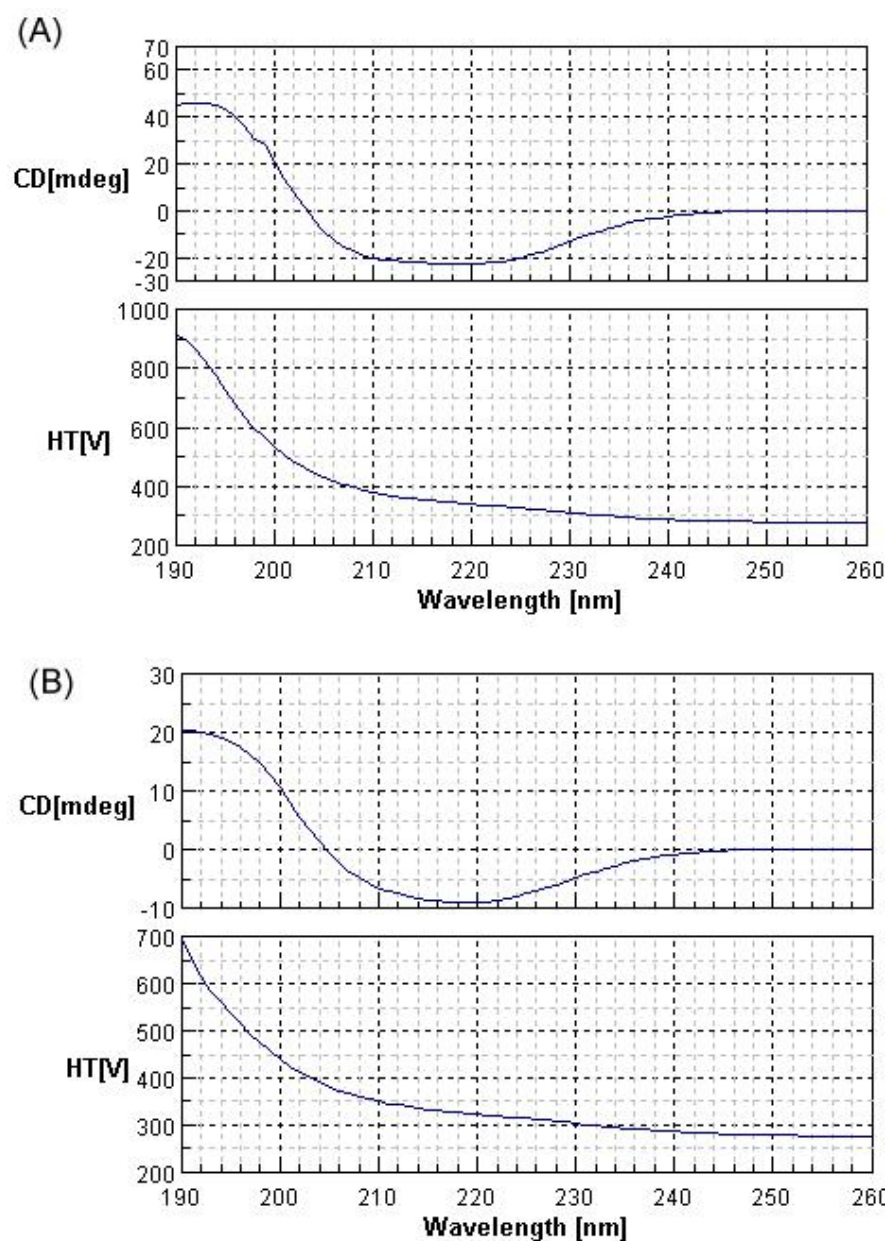


Figure C3 CD Spectra of *E. coli* nBamA Δ 1-20 and BamA Δ 1-20 (Repeat 2)

(A) The upper panel shows the far UV CD spectra for *E. coli* nBamA Δ 1-20. The protein concentration is 10.6 μ M. The measurements were taken between wavelengths 190 to 260nm. The lower panel shows the corresponding high tension voltage trace. (B) The upper panel shows the far UV CD spectra for the refolded *E. coli* BamA Δ 1-20. The protein concentration is 5.31 μ M. The measurements were taken between wavelengths 190 to 260nm. The lower panel shows the corresponding high tension voltage trace.

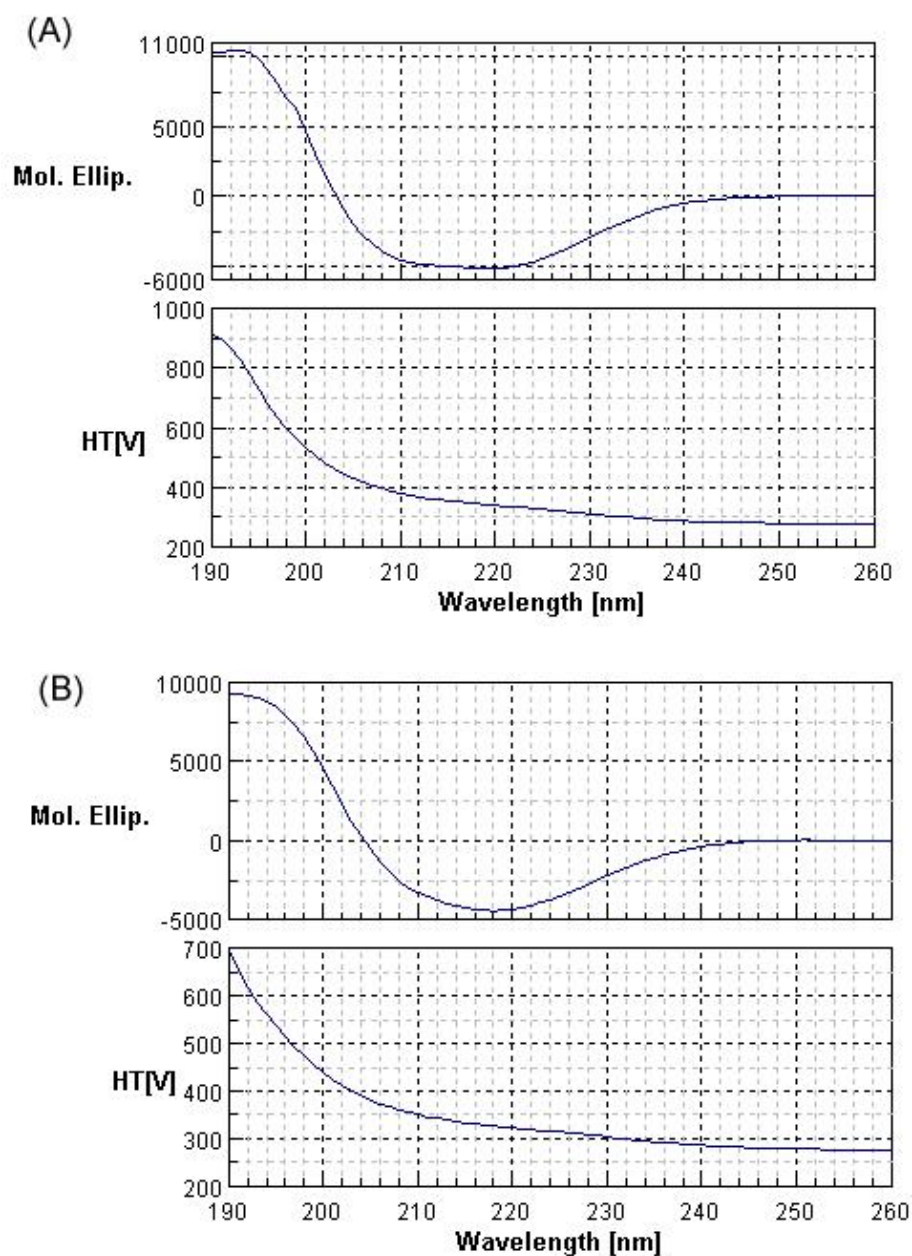


Figure C4 CD Spectra of *E. coli* nBamA Δ 1-20 and BamA Δ 1-20 Converted to Molar Ellipticity (Repeat 2)

(A) The CD absorbance readings of *E. coli* nBamA Δ 1-20 have been converted to molar ellipticity $[\theta]_{\text{molar},\lambda}$ at wavelength λ (190 nm ~ 260 nm). The high tension voltage trace is shown in the lower panel. (B) The CD absorbance readings of refolded *E. coli* BamA Δ 1-20 have been converted to molar ellipticity $[\theta]_{\text{molar},\lambda}$ at wavelength λ (190 nm ~ 260 nm), with unit millidegree $\text{cm}^2\text{dmol}^{-1}$. The high tension voltage trace is shown in the lower panel.

Appendix D.

Sequence Alignment of *E. coli* BamA Homology Model

E. coli BamA Homology Model vs. NgBamA

```
Aligned sequences: 2
# 1: 4K3B_model_default_chain_A
# 2: PDBB_model_default_chain_A
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend penalty: 0.5
#
# Length: 774
# Identity: 287/774 (37.1%)
# Similarity: 429/774 (55.4%)
# Gaps: 6/774 ( 0.8%)
# Score: 1326.0
#
#
#=====
4K3B_model_de 1 ADFTIQDIRVEGLQRTEPSTVFNYLPVKVGDYNDTHGSAI IKSLYATGF 50
PDBB_model_de 1 EGFVVKDIHFEGQLQRVAVGAALLSMPVRTGDTVNDEDISNTIRALFATGN 50
4K3B_model_de 51 FDDVRVETADGQLLLTVIERPTIGSLNITGAKMLQNDAIKKNLESFGLAQ 100
PDBB_model_de 51 FEDVRVLRDGD TLLVQVKERPTIASLTFSGNKS SVKDDMLKQNLEASGVRV 100
4K3B_model_de 101 SQYFNQATLNQAVAGLKEEYLGRGKLN IQITPKVTKLARNRVDIDITIDE 150
PDBB_model_de 101 GESLDRTTIADIEKGLEDFYYSVGVKYSASVKAVVTPLPRNRVDLKL V FQE 150
4K3B_model_de 151 GSAKITDIEFEGNQVYSDRKL MRQMSL TEGGIWTLTRSDR-FDRQKFA 199
PDBB_model_de 151 GVS AEIQQINIVGNHAF TTDKLI SHFQLRDEVPW-WNVVVGDRKYQKQKLA 199
4K3B_model_de 200 QDMEKVTDFYQNGYFDFRILDTDIQT NEDKTRQTIKITVHEGGRFRWGK 249
PDBB_model_de 200 GDLETLSYYLDRGYARFNIDSTQVSLTPDKKGIYVTVNITEGDQYKLSG 249
4K3B_model_de 250 VSI EGD TNEVPKAELEKLLT-MKPGK WYERQQMTAVLGEIQNRMSAGYA 298
PDBB_model_de 250 VEVSGNL-AGHSAEIEQKLT KIEPGELYNGTKVTKMEDDIKLLGRYGYA 298
4K3B_model_de 299 YSEISVQPLPNAGTKTVDFVLHIEPGRKIYVNEI HITGNNKTRDEVVRE 348
PDBB_model_de 299 YPRVQSMFEINDAKTVKLRVNV DAGNRFYVRKIRFEGNDTSKDAVLRRE 348
4K3B_model_de 349 LRQMESAPYDTSKLR SKERV ELLG YFDNVQF DAVPLAGTPDKVDLNM SL 398
PDBB_model_de 349 MRQMEGAWLGS DLVDQK ERLNRLGFFETVDTDTQRVPGSPDQVDVYK V 398
4K3B_model_de 399 TERSTGSLDLSAGWVQDTGLVMSAGVSQDNLFGTGKSAALRASRSK TTLN 448
PDBB_model_de 399 KERNTGSFNFGIGYGTESGV SFQAGVQQDNWLGTGYAVGINGTKNDYQTY 448
4K3B_model_de 449 GSLSF TDPYFTADGVSIGYDIYGKAFDPRKASTSVKQYKTTTAGGGV RMG 498
PDBB_model_de 449 AELSVTNPFYFTVDGVS LGGRLFYND FQADDADLSVKDYTNKSYGTDVTLG 498
4K3B_model_de 499 IPVTEYDRVNFGLAAEHLTVNTYNKAPKRYAD FIRKYGKTDGADGSFKGL 548
PDBB_model_de 499 FPINEYNSLRAGLGYVHNSL SNMQVAMWRYL YSMREHPSTSDQDNSFKTD 548
```



```

PDBB_model_de      301 RVQSMPEINDADKTVKLRVNV DAGNRFYVRKIRFEGNDTSKDAVLRREMR      350
4N75_model_de      1 -----
PDBB_model_de      351 QMEGAWLGSDLVDQGKERLNRLGFFETVDTDTQRVPGSPDQVDVVYKVKE      400
4N75_model_de      1 ----MNF GIGYGTESGVSFQAGVQQDNWLGTYAVGINGTKNDYQTYAE      45
      . |||
PDBB_model_de      401 RNTGSFNF GIGYGTESGVSFQAGVQQDNWLGTYAVGINGTKNDYQTYAE      450
4N75_model_de      46 LSVTNPYFTVDGVSLGGR LFYNDFQADDADLS--DYTNKSYGTDVTLGFP      93
      |||
PDBB_model_de      451 LSVTNPYFTVDGVSLGGR LFYNDFQADDADLSVKDYTNKSYGTDVTLGFP      500
4N75_model_de      94 INEYNSLRAGLGYVHNSLSNMQPQVAMWRYLYSMGHPSTSDQDNSFKTD      143
      |||
PDBB_model_de      501 INEYNSLRAGLGYVHNSLSNM--QVAMWRYLYSMREHPSTSDQDNSFKTD      548
4N75_model_de      144 DFTFNYGWTYNKLD RGYFPTDGSRVNLTGKVTIPGSDNEYKVTLDTATY      193
      |||
PDBB_model_de      549 DFTFNYGWTYNKLD RGYFPKDGSRVNLGAKITIPGSDNEYKVTLDTATY      598
4N75_model_de      194 VPIDDDHKWVVLGRTRWGYGDGLGGKEMPFYENFYAGGSSTVRGFQSN TI      243
      |||
PDBB_model_de      599 VPIDDDHTWVVLGRTRWGYGDGL-GKEMPFYENFYAGGSSTVRGFQSN TI      647
4N75_model_de      244 GPKAVYFPHQASNYDPDYDYECATQDGAKDLCKSDDAVGGNAMAVASLE-      292
      |||
PDBB_model_de      648 GPKAVYFPH-----SDDAVGGNAMAVASLEL      673
4N75_model_de      293 -FITPTPFI--SVRTSFFWDMGT VWDTNWDSS--QYSGYPDYSD-----      331
      |||
PDBB_model_de      674 LFITPSDKYANSVRTSFFWDMGT VWDTNWDSSAEQYSGYPDYSDNAHKSF      723
4N75_model_de      332 PSNIRMSAGIALQWMSPLG PLVFSYAQPFKKYDGDKAEQFQFNIGKTW      379
      |||
PDBB_model_de      724 PSNIRMSAGIALQWMSPLG PLVFSYAQPFKKYDGDKAEQFQFNIGKTW      771

```

***E. coli* BamA Homology Model vs. β -barrel domain of *E. coli* BamA including POTRA 5 (4C4V)**

```

# Aligned_sequences: 2
# 1: 4C4V_model_default_chain_A
# 2: PDBB_model_default_chain_A
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend penalty: 0.5
#
# Length: 787
# Identity:      420/787 (53.4%)
# Similarity:   422/787 (53.6%)
# Gaps:         358/787 (45.5%)
# Score: 2183.0
#
#
#=====

```

```

4C4V_model_de      1 -----
PDBB_model_de      1 EGFVVKDIHFEG LQRVAVGAALLSMPVRTGDTVND EDISNTIRALFATGN      50
4C4V_model_de      1 -----

```


PDBB_model_de	51	FEDVRLRDLRGDTELLVQVKERPTIASLTFSGNKSVKDDMLKQNLEASGVRV	100
4C4V_model_de	1	-----	0
PDBB_model_de	101	GESLDRTTIADIEKGLEDFYYSVVGKYSASVKAVVTPLPRNRVLDLKLVFQE	150
4C4V_model_de	1	-----	0
PDBB_model_de	151	GVSAEIQQINIVGNHAFITDKLISHFQLRDEVFWNVVVGDRKYQKQKLAG	200
4C4V_model_de	1	-----	0
PDBB_model_de	201	DLETLSRYLLDRGYARFNIDSTQVSLTPDKKGIYVTVNITEGDQYKLSGV	250
4C4V_model_de	1	-----	0
PDBB_model_de	251	EVSGNLAGHSAEIEQKLTKEIEPGELYNGTKVTKMEDDIKLLGRYGYAYP	300
4C4V_model_de	1	-----FYVVRKIRFEGNDTSKDAVLRREMR	24
PDBB_model_de	301	RVQSMPEINDADKTVKLRVNVVDAGNRFYVVRKIRFEGNDTSKDAVLRREMR	350
4C4V_model_de	25	QMEGAWLGSDLVDQGERLNRLGFFETVDTDTQRPVGPSPDQVDVVYKVKKE	74
PDBB_model_de	351	QMEGAWLGSDLVDQGERLNRLGFFETVDTDTQRPVGPSPDQVDVVYKVKKE	400
4C4V_model_de	75	RNTGSFNFGIGYGTESGVSFQAGVQQDNWLGTGYAVGINGTKNDYQTYAE	124
PDBB_model_de	401	RNTGSFNFGIGYGTESGVSFQAGVQQDNWLGTGYAVGINGTKNDYQTYAE	450
4C4V_model_de	125	LSVTNPYFTVDGVSLGGRLFYNDQADDADSS--DYTNKSYGTDVTLGFP	172
PDBB_model_de	451	LSVTNPYFTVDGVSLGGRLFYNDQADDADLSVKDYTNKSYGTDVTLGFP	500
4C4V_model_de	173	INEENSLRAGLGYPVHNSLSNMQPQVAMWRYLYSMGEHPSTSDQDNSFKTD	222
PDBB_model_de	501	INEYNSLRAGLGYPVHNSLSNM--QVAMWRYLYSMREHPSTSDQDNSFKTD	548
4C4V_model_de	223	DFTFNYGWTYNKLDRGYFPTDGSRVNLTGKVTIPGSDNEYKVTLDTATY	272
PDBB_model_de	549	DFTFNYGWTYNKLDRGYFPKDGSRVNLGAKITIPGSDNEYKVTLDTATY	598
4C4V_model_de	273	VPIDDDHKWVVLGRTRWGYGDGLGGKEMPFYENFYAGGSSTVRGFQSNTI	322
PDBB_model_de	599	VPIDDDHTWVVLGRTRWGYGDGL-GKEMPFYENFYAGGSSTVRGFQSNTI	647
4C4V_model_de	323	GPKAIYFPHQASAKDLCKSDDAVGGNAMAVASLE--FITPTPFISDKYAN	370
PDBB_model_de	648	GPKAVYFPH-----SDDAVGGNAMAVASLELLFITP----SDKYAN	684
4C4V_model_de	371	SVRTSFFWDMGTVWDTNWDSS----SGYPDYS-----PSNIRMSAGIA	410
PDBB_model_de	685	SVRTSFFWDMGTVWDTNWDSSAEQYSGYPDYSNAHKSFPSNIRMSAGIA	734
4C4V_model_de	411	LQWMSPLGPLVFSYAQPFFKDYGDGKAEQFQFNIGK-- 445	
PDBB_model_de	735	LQWMSPLGPLVFSYAQPFFKDYGDGKAEQFQFNIGK 771	

2QCZ_model_de	311	-----	310
PDBB_model_de	651	AVYFPHSDDAVGGNAMAVASLELLFITPSDKYANSVRTSFFWDMGTVWDT	700
2QCZ_model_de	311	-----	310
PDBB_model_de	701	NWDSSAEQYSGYPDYSDNAHKSFPSNIRMSAGIALQWMSPLGPLVFSYAQ	750
2QCZ_model_de	311	-----	310
PDBB_model_de	751	PFKKYDGDKAEQFQFNIGKTW	771