

**Structural studies of the BAM complex, OmpU
outer membrane protein and lipoprotein N-acyl
transferase in Gram-negative bacteria**

by

Huanyu Li, B. Sc.

Thesis submitted for the degree of Doctor of

Philosophy

TO

The University of East Anglia

Norwich Medical School

Faculty of Medicine and Health Sciences

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that use of any information derived there from must be in accordance with current UK Copyright Law. In addition, any quotation or extract must include full attribution.

ABSTRACT

Structural studies of membrane proteins represent a significant challenge in the field owing to their hydrophobic nature, unstable property and resistance to be crystallized. In Gram-negative bacteria, membrane proteins contribute to the characteristic membranous architecture composed of an asymmetric layer of outer membrane (OM) and a symmetric inner cytoplasmic membrane (IM).

Outer membrane proteins (OMPs) play essential roles in nutrient uptake, protein transport, outer membrane assembly, and pathogenesis of Gram-negative bacteria. In *Escherichia coli*, nearly all the outer membrane proteins are inserted into the outer membrane by the β -barrel assembly machinery (BAM), which contains one conserved membrane protein BamA and four lipoproteins BamBCDE. The individual protein structures of the BAM complex have been reported, but the mechanism of OMP assembly by the BAM complex is halted by a lack of structure of the whole complex. During the course of the collaborative BAM complex project, I participated in structural studies of the BAM complex and generated high resolution crystallographic diffraction data that contributes to one of the two determined structures of the BAM complex, and the structural insights have enlightened understanding of the *in vivo* insertion mechanism.

Of diverse types of β -barrel OMPs that are inserted into the OM by the BAM complex, an outer membrane protein called OmpU from *Vibrio cholerae* is a potential virulence factor in addition to its porin identity with undefined atomic structure. I determined the crystal structure of this OMP, in which the long and flexible extracellular loop L4 and a novel N-terminal coil in the pore lumen provide direct structural evidence underlying its particular functions.

The symmetric lipid bilayer of IM accommodates an even more diverse array of IMPs composed of the contrasting dominance of α -helices. Three IMPs are responsible for conducting post-translational modifications of lipoproteins in Gram-negative bacteria, a class of proteins destined to reside on the periplasmic side of either the IM or the OM via acyl chains post-translationally linked to the N-terminal cysteine residues, and they are called phosphatidylglycerol:lipoprotein diacylglyceryl transferase (Lgt), Lipoprotein signal peptidase (Lsp) and lipoprotein N-acyltransferase (Lnt). Structural studies were carried out on Lnt but unsuccessful in determining the atomic structure. Recent structures of Lnt reported during the course of this project are consistent with earlier biochemical studies that piloted the understanding of its function and further elucidate the molecular mechanisms of its primary acyl-transfer function.

ACKNOWLEDGEMENTS

The journey to a doctor is not easy for me being an international student, but it ends delightfully with the help of some amazing people I get to know and work with throughout.

First and foremost, I would like to express my gratitude to my primary supervisor Prof. Changjiang Dong and secondary supervisor Prof. Tom Wileman for their continuous support and guidance in my study. I would also like to thank Dr Andrew Hemmings and Dr Gary Rowley for being the member of transfer panel and their valuable comments and advice during my Ph.D transfer.

My sincere thanks go to my past colleague Bo Peng, a visiting postdoc for one year, the progress of my Lnt project would not been made smooth without the contribution of him, who inspired and encouraged me a lot during those difficult times and assisted me very diligently towards structural characterization of Lnt. My sincere thanks also go to my fellow Ph.D student from Tom Wileman's group and a good friend Weijiao Zhang, who has been making substantial contribution to my current lab work. I would also like to express my gratitude to Yinghong Gu and Yi Zeng for the key roles they played in the BAM complex project.

I am extremely indebted to my parents over the past four years. Their love, care and understanding have taught me how to be a better man, and their financial support throughout enables me to live the life I do, even though they are thousands of miles away from where I am. I do appreciate everything they have done for me, their belief in me, and that they never ask for any return. I am a lucky son of them. Thank you.

TABLE OF CONTENTS

ABSTRACT	2
ACKNOWLEDGEMENTS	3
TABLE OF TABLES	9
ABBREVIATIONS	10
CHAPTER 1	12
Principles of X-ray crystallography	12
1.1 General principles	13
1.2 Geometric principles of diffraction	14
1.3 Unit cell symmetry and space groups	16
1.4 Reflections as three-dimensional waves	19
1.5 Single-crystal X-ray diffraction	20
1.5.1 Data indexing, integration, scaling and merging	22
1.5.2 Obtaining phases	24
1.5.3 Heavy atom method	24
1.5.4 Molecular replacement	27
1.5.5 Phase refinement, model building and structure refinement	28
CHAPTER 2	31
General Introduction	31
2.1 Bacterial membranes	32
2.2 Outer membrane proteins (OMPs)	34
2.3 OMP structures	34
2.4 OMP biogenesis	36
2.5 Identification of the BAM complex	39
2.6 Structures of individual components of the BAM complex	42
2.7 Proposed mechanisms of OMP assembly by the BAM complex	43
2.8 General porins	45
2.9 <i>Vibrio cholerae</i> OmpU	46
2.10 Implications in host-pathogen interactions	46
2.11 Structural information of OmpU	47
2.12 Inner membrane proteins (IMPs)	47
2.13 Lipoproteins	48
2.14 General lipoprotein structural domains	49
2.15 Lipoprotein modifications	51
2.16 Lgt	52
2.17 Lsp	52
2.18 Lnt	53
2.19 Research aims	54
CHAPTER 3	56
Materials and methods used in structural studies of the BAM complex, OmpU outer membrane protein and lipoprotein N-acyl transferase	56
3.1 Plasmid construction of <i>E. coli</i> BamAB+CDE	57
3.2 Expression of <i>E. coli</i> BamAB	58
3.3 Expression of <i>E. coli</i> BamCDE	59

3.4 Purification and reconstitution of the BAM complex from <i>E. coli</i> BamAB + <i>E. coli</i> BamCDE.....	59
3.5 Protein Crystallization	60
3.6 Co-purification and crystallization of <i>E. coli</i> BamAB and <i>E. coli</i> BamCDE	61
3.7 Cloning and construction of pJH114 encoding BamABCDE and pYG120 encoding BamABCDE and an extra BamB copy	61
3.8 Expression of BamACDE (pJH114) (performed by colleagues) and BamABCDE (pYG120)	62
3.9 Purification of BamACDE (performed by colleagues) and BamABCDE	63
3.10 Crystallization, data collection and structure determination of BamACDE (performed by colleagues) and BamABCDE	64
3.11 Plasmid construction of OmpU	68
3.12 Expression of <i>V. cholerae</i> OmpU	68
3.13 Purification of <i>V. cholerae</i> OmpU	69
3.14 Protein crystallization and data collection	69
3.15 Structure determination	70
3.16 Generation of Lnt expression plasmid.....	72
3.17 Expression of Lnt.....	75
3.18 Purification of Lnt	75
3.19 96 detergent screening.....	76
3.20 Crystallization and data collection of <i>V. cholerae</i> Lnt	77
CHAPTER 4	78
Results chapter: structural studies of the five-protein complex of beta-barrel assembly machinery (BAM complex) from <i>Escherichia coli</i>.....	78
4.1 <i>E. coli</i> BamAB+BamCDE separate overexpression and subsequent reconstitution.....	79
4.2 <i>E. coli</i> BamAB+CDE co-purification.....	84
4.3 Expression of all five subunits in a single expression plasmid (pJH114 and pYG120).....	87
4.4 BamACDE structure generated from pJH114 (determined by colleagues).....	91
4.5 BamABCDE structure generated from pYG120	95
4.6 Interactions (defined by proximity within 3.5 Å) between BamA and BamB	101
4.7 Interactions between BamA and BamC	103
4.8 Interactions between BamA and BamD	104
4.9 Interactions between BamA and BamE	105
4.10 Interactions between BamC, BamD, and BamE.....	105
4.11 Conformational changes in BamA	109
CHAPTER 5	112
Results: structural studies of an outer membrane protein OmpU from <i>Vibrio cholerae</i>	112
5.1 OmpU purification and crystallization.....	113
5.2 Overall OmpU fold	115
5.3 Non-canonical N-terminal coil and an additional constriction zone.....	122
CHAPTER 6	129
Results: structural studies of lipoprotein N-acyl transferase in Gram-negative bacteria	129
6.1 <i>Vibrio cholerae</i> Lnt purification and crystallization	130
6.2 96 detergent screening.....	133
6.3 Published Lnt structures.....	136
CHAPTER 7	141
Discussion	141
7.1 BAM complex.....	142

7.2 OmpU.....	147
7.3 Lnt.....	150
REFERENCES.....	154
APPENDIX 1.....	166
APPENDIX 2.....	169
APPENDIX 3.....	170
APPENDIX 4.....	179

TABLE OF FIGURES

FIGURE 1. REPRESENTATION OF ONE UNIT CELL WITH TWO ALANINE MOLECULES WITHIN AND SPATIAL COODINATES X, Y, Z.	15
FIGURE 2. ILLUSTRATION OF SIX UNIT CELLS PACKED IN A CRYSTALLINE LATTICE, WHICH IS A BUILDING BLOCK OF A LARGER CRYSTAL.....	16
FIGURE 3. SEVEN CRYSTAL SYSTEMS.....	17
FIGURE 4. WORKFLOW FOR PROTEIN CRYSTAL STRUCTURE DETERMINATION BY X-RAY CRYSTALLOGRAPHY.	20
FIGURE 5. SINGLE CRYSTAL X-RAY DIFFRACTION BY BEAMS EMITTED FROM AN X-RAY TUBE AS AN EXAMPLE OF X-RAY SOURCE.	22
FIGURE 6. STRUCTURES OF THE CELL WALL IN GRAM-NEGATIVE BACTERIA (A) AND GRAM-POSITIVE BACTERIA (B).	32
FIGURE 7. STRUCTURES OF SELECTED B-BARREL OMPs.	35
FIGURE 8. OMP BIOGENESIS PATHWAY IN GRAM-NEGATIVE BACTERIA.....	39
FIGURE 9. THE STRUCTURES OF THE FIVE COMPONENTS OF THE BAM COMPLEX.....	42
FIGURE 10. TWO PROPOSED MECHANISTIC MODELS OF THE BAM COMPLEX DURING OMP BIOGENESIS.	44
FIGURE 11. LIPOPROTEIN STRUCTURAL DOMAINS.	50
FIGURE 12. LIPOPROTEIN POSTTRANSLATIONAL MODIFICATION IN GRAM-NEGATIVE BACTERIA.....	51
FIGURE 13. CHROMATOGRAM OF E.COLI BAMAB+CDE ON GEL FILTRATION.	80
FIGURE 14. SDS-PAGE GEL PICTURE OF THE PURIFIED <i>E. COLI</i> BAMAB+CDE COMPLEX.....	81
FIGURE 15. A PICTURE SHOWING THE CRYSTALS THAT APPEARED FROM TWO CONDITIONS FROM SEPARATE PURIFICATION OF <i>E. COLI</i> BAMAB AND <i>E. COLI</i> BAMCDE.....	82
FIGURE 16. A PICTURE OF THE WELL-DIFFRACTING CRYSTAL OF <i>E. COLI</i> BAMAB+CDE.....	83
FIGURE 17. CHROMATOGRAM OF <i>E. COLI</i> BAMAB+CDE CO-PURIFICATION ON GEL FILTRATION COLUMN.	85
FIGURE 18. SDS-PAGE GEL PICTURE OF THE CO-PURIFIED <i>E. COLI</i> BAMAB+CDE COMPLEX.	86
FIGURE 19. A PICTURE OF THE CRYSTALS PRODUCED FROM CO-PURIFICATION OF <i>E. COLI</i> BAMAB AND <i>E. COLI</i> BAMCDE.	87
FIGURE 20. CHROMATOGRAM OF <i>E. COLI</i> BAM COMPLEX EXPRESSED FROM PYG120 ON GEL FILTRATION COLUMN.	89
FIGURE 21. (A) SDS-PAGE COMPARISON BETWEEN THE BAM COMPLEX PROTEINS PURIFIED FROM PYG120 AND PJH114 (B) SDS-PAGE RESULT OF THE BAM COMPLEX PROTEINS PURIFIED FROM PYG120.....	90
FIGURE 22. A PICTURE OF THE WELL-DIFFRACTING NATIVE CRYSTALS GENERATED FROM PYG120.....	91
FIGURE 23. CATOON REPRESENTATION OF THE BAMACDE STRUCTURE (HORIZONTAL VIEW FROM THE OM PLANE).	93
FIGURE 24. BAMACDE STRUCTURE MODEL VIEWED FROM THE EXTRACELLULAR SIDE.	94
FIGURE 25. BAMACDE STRUCTURE MODEL VIEWED FROM THE PERIPLASMIC SIDE.....	95
FIGURE 26. BAMABCDE COMPLEX STRUCTURE MODEL VIEWED FROM THE HORIZONTAL OM PLANE.....	97
FIGURE 27. B16C OF THE BARREL DOMAIN OF BAMA IS COILED INTO THE BARREL LUMEN.	98
FIGURE 28. BAMABCDE STRUCTURE MODEL VIEWED FROM THE EXTRACELLULAR SIDE.	99
FIGURE 29. BAMABCDE STRUCTURE MODEL VIEWED FROM THE PERIPLASMIC SIDE.	100
FIGURE 30. IODIDE ANOMALOUS SIGNAL MAP IN A SINGLE UNIT CELL OF BAMABCDE CRYSTAL.....	101
FIGURE 31. INTERACTIONS BETWEEN BAMA AND BAMB.....	102
FIGURE 32. INTERACTIONS BETWEEN BAMA AND BAMC.....	103
FIGURE 33. INTERACTIONS OF BAMD WITH BAMA POTRA 5.....	104
FIGURE 34. INTERACTIONS BETWEEN BAMA AND BAME. EXPANDED VIEW IS PROVIDED AT BAMA POTRA 5 AND BAME RESIDUES WHERE INTERACTIONS TAKE PLACE.	105
FIGURE 35. INTERACTIONS BETWEEN BAMC AND BAMD.....	107
FIGURE 36. INTERACTIONS BETWEEN BAMC AND BAME.	108
FIGURE 37. INTERACTIONS BETWEEN BAMD AND BAME.	109
FIGURE 38. CONFORMATIONAL CHANGES OBSERVED IN BAMA STRUCTURE.....	111
FIGURE 39. SIZE-EXCLUSION CHROMATOGRAM OF <i>V. CHOLERAE</i> OMPU ON GEL FILTRATION COLUMN.	113
FIGURE 40. SDS-PAGE GEL PICTURE OF <i>V. CHOLERAE</i> OMPU.	114
FIGURE 41. A PICTURE OF THE WELL-DIFFRACTIVE OMPU CRYSTAL.....	115
FIGURE 42. CARTOON REPRESENTATION OF OMPU TRIMER, VIEWED FROM THE MEMBRANE PLANE (A), THE EXTRACELLULAR SIDE (B) AND THE PERIPLASM (C).	117
FIGURE 43. CARTOON REPRESENTATION OF OMPU PROTOMER STRUCTURE, VIEWED FOR THE HORIZONTAL MEMBRANE PLANE (A), THE EXTRACELLULAR SIDE (B) AND THE PERIPLASM (C).	118
FIGURE 44. 2Fo-FC ELECTRON DENSITY MAP (CONTOURED AT 1 σ) OF AN ASSIGNED LDAO (A) AND TWO GLYCEROL MOLECULES (B) IN OMPU STRUCTURE.	119

FIGURE 45. OVERLAY OF THE STRUCTURE OF OMP _U PROTOMER ONTO THREE STRUCTURALLY ANALOGOUS PORIN STRUCTURES VIEWED FROM THE EXTRACELLULAR SIDE.	121
FIGURE 46. <i>B</i> FACTOR DIAGRAM OF PROTOMERIC OMP _U REPRESENTED BY THE <i>B</i> FACTOR PUTTY PROGRAMME IN PYMOL.	122
FIGURE 47. THE TWO CONSTRICTION ZONES OF OMP _U	124
FIGURE 48. 2Fo-FC ELECTRON DENSITY MAP (CONTOURED AT 1 σ) OF THE ARGININE CLUSTER AT THE CONSTRICTION LINING.	125
FIGURE 49. PORE RADII PLOT AGAINST COORDINATES IN THE DIRECTION OF THE CHANNEL.	127
FIGURE 50. ELECTROSTATIC POTENTIAL OF OMP _U MODEL WITH (A AND C) AND WITHOUT (B AND D) THE N-TERMINAL COIL.	128
FIGURE 51. SDS-PAGE GEL PICTURE OF <i>V. CHOLERAE</i> LNT FROM SMALL-SCALE EXPRESSION TESTS.	131
FIGURE 52. SIZE EXCLUSION CHROMATOGRAMS OF <i>V. CHOLERAE</i> LNT. SAMPLES WERE INJECTED ONTO A HiLOAD 16/600 SUPERDEX 200 PREP GRADE COLUMN (GE HEALTHCARE).	132
FIGURE 53. SDS-PAGE RESULTS OF THE 96-DETERGENT SCREENING FOR <i>V. CHOLERAE</i> LNT.	135
FIGURE 54. CARTOON REPRESENTATION OF THE STRUCTURE OF <i>E. COLI</i> LNT VIEWED FROM THE HORIZONTAL MEMBRANE PLANE.	138
FIGURE 55. A DIAGRAM OF THE SECONDARY STRUCTURE SEGMENTS IN LNT.	139
FIGURE 56. PROPOSED LNT N-ACYLATION MECHANISM DISSECTED INTO SIX STEPS.	140

TABLE OF TABLES

Table 1. Plasmids used in separate expression of BamAB + BamCDE.....	58
Table 2. Data collection and refinement statistics for BamACDE and BamABCDE complex.....	67
Table 3. Data collection and refinement statistics for OmpU.....	71
Table 4. Choices of phasing models for molecular replacement.....	72
Table 5. PCR primers used for <i>lnt</i> gene cloning in individual selected species.....	74

ABBREVIATIONS

BAM	β -Barrel Assembly Machinery
C ₈ E ₄	Tetraethylene Glycol Monooctyl Ether
CL	Cardiolipin
CMC	Critical Micelle Concentration
DDM	<i>n</i> -dodecyl- β -D-maltopyranoside
IM	Inner Membrane
IMP	Inner Membrane Protein
IPTG	Isopropyl β -D-thiogalactopyranoside
LB	Luria Bertani
LDAO	<i>N,N</i> -Dimethyldodecylamine <i>N</i> -oxide
LMNG	Lauryl Maltose Neopentyl Glycol
LPS	Lipopolysaccharide
MAD	Multiple-wavelength Anomalous Diffraction
mAU	milli-Absorbance Unit
NG	<i>n</i> -Nonyl- β -D-Glucopyranoside
Ni-NTA	Nickel-nitrilotriacetic acid
OD ₆₀₀	Optical Density at 600 nm
OG	<i>n</i> -Octyl- β -D-Glucopyranoside
OM	Outer Membrane
OMP	Outer Membrane Protein
OMPU	Outer Membrane Protein U
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PE	Phosphatidylethanolamine
PG	Phosphatidylglycerol
POTRA	POlypeptide TRAnslocation-Associated
RMSD	Root-Mean-Square Deviation
SAD	Single-wavelength Anomalous Diffraction
SB3-14	3-(<i>N,N</i> -dimethylmyristyl-ammonio-propanesulphonate
SDS-PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
SRP	Signal Recognition Particle
Se-Met	SelenoMethionine

TB

Terrific Broth

TBS

Tris Buffered Saline

TF

Trigger factor

TPR

TetratricoPeptide Repeats

CHAPTER 1

Principles of X-ray crystallography

All the structures solved in this thesis project were obtained using X-ray crystallography as a principal method. As specified by the Protein Data Bank (PDB) Annual Report in 2016, 89.4% of the deposited biomolecular structures in the data bank were determined by X-ray crystallography.

1.1 General principles

There is a general physical law that the wavelength λ of the light needs to be equal or smaller than the object in order for the object to induce diffraction and be visible under magnification in most cases. Protein molecules are therefore not able to diffract visible light, which is electromagnetic radiation with wavelengths of 400–700 nm (nm = 10^{-9} m), and generate expansion of individual atoms, in which bonded atoms are ~ 0.15 nm or 1.5 angstroms ($\text{\AA} = 10^{-10}$ m) apart. X-rays, on the other hand, possesses electromagnetic radiation of this wavelength and can be diffracted by even the smallest molecules. However, X-ray induced structural determination rarely solves the hydrogen atoms in the molecular models, and they are usually not required to be present. Nonetheless, based on the expectation that bond lengths, bond angles, and conformational angles in proteins are similar to those in small organic molecules, the positions of all hydrogen atoms can be confidently interpreted.

Despite the fact that individual atoms can diffract X-rays, producing a focused image of a single molecule as the pattern of a focused image of objects seen in light microscopes are not possible due to two main reasons. First is the nature of X-rays not to be focused by lenses. In X-ray analysis of protein crystals, this can be circumvented by measuring the directions and strengths (intensities) of the diffracted X-rays and then applying a computer to mimic the existence of an image-reconstructing lens. This artificial lens then carries out complicated computations and finally displays the magnified image of the object on a digital screen.

Second, a single molecule is not sufficient to produce detectable X-ray diffracting power and subsequent interpretable data. Protein crystals, composed of countless amount of ordered array of molecules in identical orientations, amplify the diffraction that results from many atoms that diffract identically and produce strong and detectable diffracted X-rays on the detector.

1.2 Geometric principles of diffraction

A crystal is a well-ordered three-dimensional array of molecules held in place by non-covalent interactions. Being a repeated object, the smallest repeating unit in the crystal is the unit cell, which is representative of the whole crystal [1, 2] (Figure 1). The organization of points at the vertices of unit cells is the lattice (Figure 2). Knowing the contents of the unit cell is thus crucial for understanding the electron density produced by the molecule within and location of every single atom. In the unit cell, the location of an atom is defined by a triplet of three-dimensional Cartesian coordinates, x , y , and z , and one of the vertices of the unit cell is usually assigned as the origin of the unit cell's coordinate numbering, which has the values $x = 0$, $y = 0$, and $z = 0$ (0, 0, 0) (Figure 1).

The dimensions of a unit cell are specified by two sets of triplets of letters: a , b , c for the lengths of three edges and α , β , γ for the interaxial angles. There are seven unit cell types, or crystal systems, and they give rise to thirteen lattice types.

In addition to indices of edges and angles of the unit cell, there is one more triplet of letters, hkl , which signifies a particular set of equivalent and parallel planes and is called lattice index or Miller index. h identifies the number of planes in the set per unit cell in the x direction, i.e. the number of sections the set of planes equally divide into at the a edge, k in the y axis and l in the z direction. Indices are presented in parentheses such that the planes with indices hkl are the (hkl) planes. A particular set of planes has its interplanar spacing d . In

one way of perceiving diffraction as reflection of X-ray beam from these sets of planes in the crystal, W.L. Bragg proposed the condition that, once met, produce a diffracted beam, in which an equation is provided as follow:

$$2d_{hkl} \sin \theta = n\lambda$$

Where n is an integer. The smaller the d_{hkl} in a lattice, the higher the resolution of the collected data [3].

Since the incident X-ray is diffracted by particular set of planes, the intensity of the diffracted beam is determined by the amount of electron density the atoms located on this set of planes provide. In fact, the electron density distributed across the set of planes (hkl) produces the reflection hkl of the diffraction pattern in the reciprocal space.

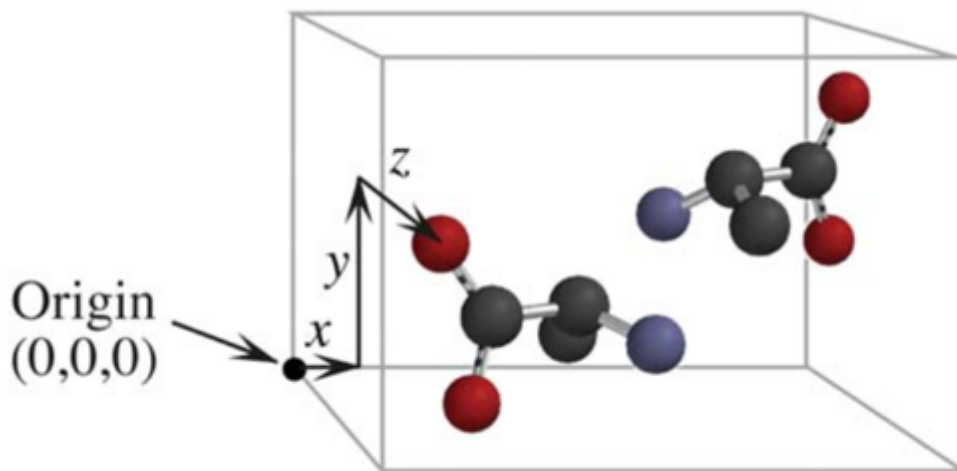


Figure 1. Representation of one unit cell with two alanine molecules within and spatial coordinates x, y, z .

The origin is specified as $(0, 0, 0)$. (Figure taken from [4])

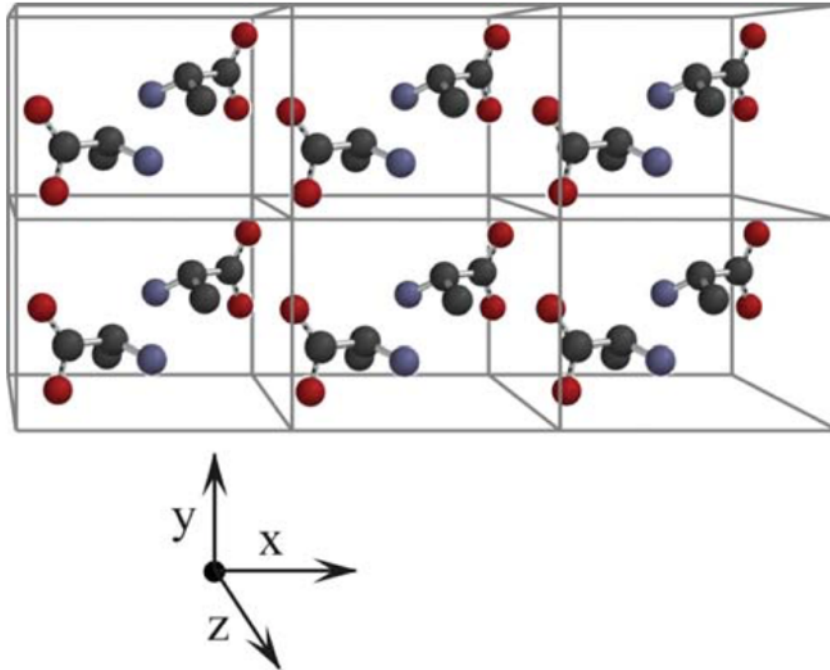


Figure 2. Illustration of six unit cells packed in a crystalline lattice, which is a building block of a larger crystal.

There are two alanine protein molecules (coloured dots and short connecting silver lines) in each unit cell. (Figure taken from [4])

1.3 Unit cell symmetry and space groups

Every unit cell adopts specific three-dimensional structure with defined shape. For example, a cell in which $a \neq b \neq c$ and $\alpha \neq \beta \neq \gamma$ is called triclinic, and is the simplest crystal system. A cell in which $a = b = c$ and $\alpha = \beta = \gamma = 90^\circ$ is called cubic. Figure 3 lists the 7 possible crystal systems.

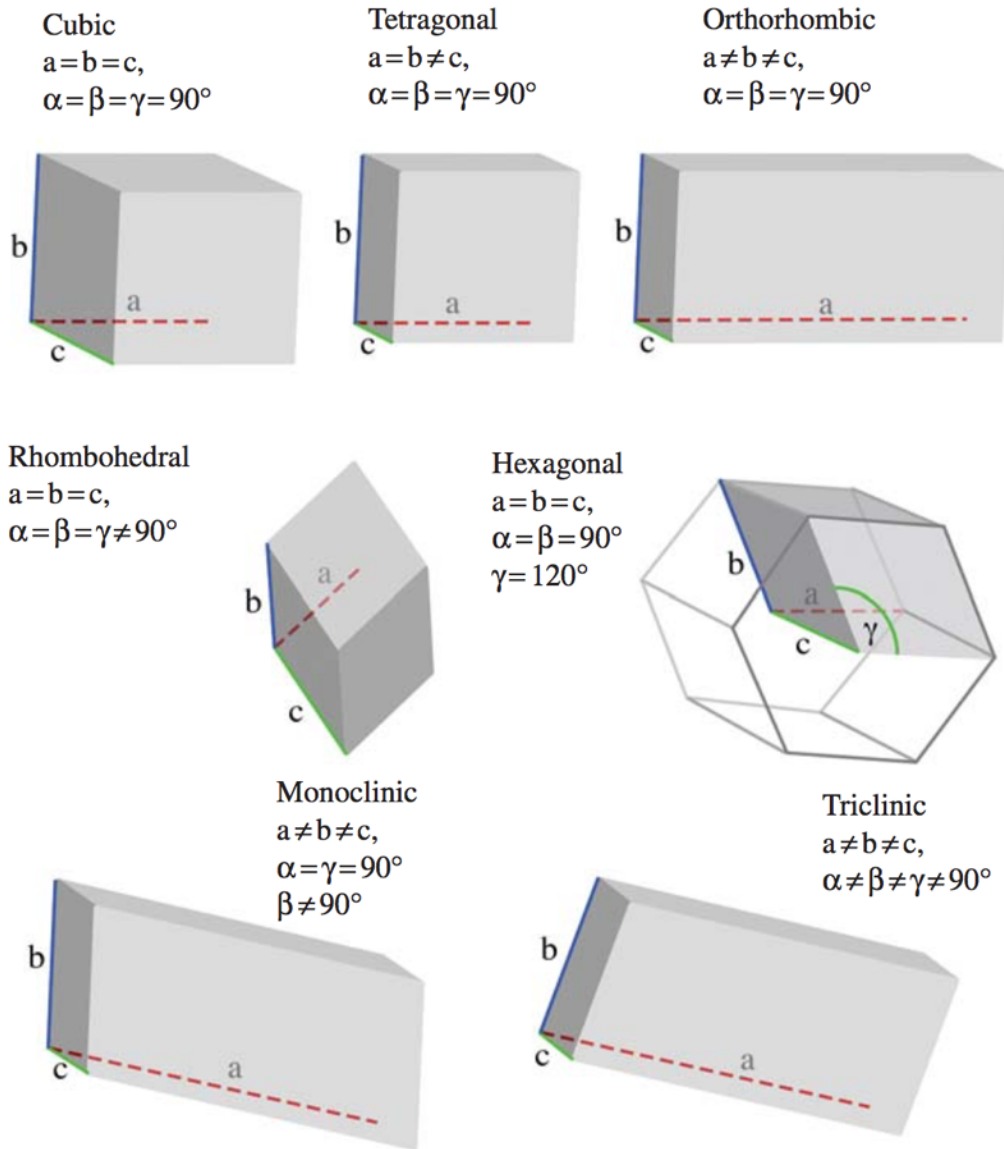


Figure 3. Seven crystal systems.

They are listed in order of symmetry, from the highest (cubic) to the lowest (triclinic).

(Figure taken from [5])

The internal symmetry of a unit cell is not solely reliant on the crystal system it belongs to. If the contents in the unit cell are not symmetry-related, the crystal system is then defined as possessing no internal symmetry. For instance, a structurally defined cubic unit cell can be regarded as not internally symmetric if the contents in the cell are not subject to symmetry.

The symmetry of a unit cell is indicated by the space group of the unit cell, which is specified by a given symbol. $P4_22_12$, for example, is a subtype of space group in which the capital letter P represents the lattice type (a primitive lattice), and the other numbers imply the symmetry operations available in the cell according to its internal symmetry. Space groups are symbolized in the form of XY_z , X represents the lattice type, with P for primitive, I for body-centered and F for face centered, Y indicates the point-group and z designates the required symmetry operations. Among the simplest symmetry operations are translation, rotation and reflection [3, 6]. Translation is equal to movement by a given distance. Rotation means rotating the content with respect to a rotation axis, one of the symmetry elements, by a certain angle. A fourfold rotation axis, for example, is designated by the number 4. Combination of rotation and translation gives rise to the screw axis, represented by the symbol n_m . Meaning an n -fold screw axis with a translation of m/n of the unit translation, $P4_2$ labels a primitive unit cell with a fourfold screw axis parallel to c , with a translation of half of the axis length. Reflection is another type of symmetry operation and entails mirror plane as the symmetry element.

Given the fact that protein molecules are intrinsically asymmetric due to presence of chiral amino acids as building blocks, a unit cell containing one protein molecule will have no symmetry elements and internal symmetry. In addition, there is no symmetry element as mirror plane found in unit cells of proteins and only translations, rotations and screw axes are found in protein crystals, reducing the number of available space groups from 230 to 65 for chiral molecules. What is more usual, however, is the situation in which a unit cell contains at least a few identical protein molecules and arranges them in a way that they are suited for producing certain symmetry elements. A protein molecule in this case that owns no symmetry elements but can be juxtaposed on adjacent copies by symmetry operations is called the asymmetric unit. During data collection, one reflection image cannot determine the

structure on its own, which displays only a minute part of the Fourier transform. As the crystal rotates, hundreds to thousands of reflection images are produced. Different angle ranges of the crystal are required to cover all the unique orientations of the crystal along with the corresponding unique reflections, with higher-symmetry crystals requiring smaller ranges. A full list of symmetry-related references of the 230 space groups can be found in the International Tables for X-ray Crystallography [7].

1.4 Reflections as three-dimensional waves

Each recorded reflection on the film is produced by a beam of electromagnetic radiation that can be regarded as a three-dimensional wave and can be summarized as the total of the diffracting contributions of all atoms in the unit cell, and this sum, termed as a structure-factor equation, can be used to describe a diffracted X-ray in a mathematical manner. The resultant sum for a specific reflection hkl is called the structure factor F_{hkl} , and the Fourier transform of a structure factor gives the electron density function $\rho(x, y, z)$, and the equation of $\rho(x, y, z)$ is provided below:

$$\rho(x, y, z) = \frac{1}{V} \sum \sum \sum F_{hkl} e^{-2\pi i(hx+ky+lz)},$$

Where V is the volume of the unit cell.

Any three-dimensional wave, along with F_{hkl} , possesses three inherent constants: the amplitude F , the frequency h , and the phase $\rho(x, y, z)$, and the equation of $\rho(x, y, z)$ is provided below: summarized as the I_{hkl} , which is readily available from measured reflection intensities when hitting the film. Frequencies of the three-dimensional wave are designated by h, k, l on x, y and z direction respectively, the same indices of the set of planes that yield diffraction, hence the frequency of a structure factor is $1/d_{hkl}$. The phase of F_{hkl} , on the other

hand, is not recorded and cannot be directly determined during data collection, which is referred to as the phase problem.

1.5 Single-crystal X-ray diffraction

Throughout the course of this thesis project, protein crystals were obtained and sent to test crystal diffraction by X-ray beams. Single-crystal X-ray diffraction, employed for all the crystals obtained, consists of steps in the following order: Obtaining a crystal, analysis under X-ray beams, data collection and downstream computational data analysis in order to calculate an electron density map and generate an atomic protein model (Figure 4).

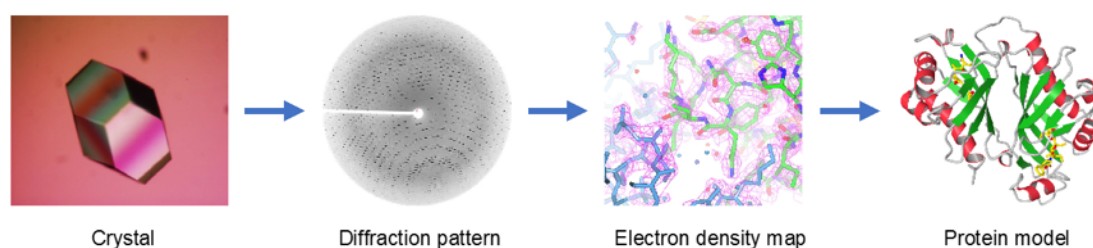


Figure 4. Workflow for protein crystal structure determination by X-ray crystallography.

When collecting X-ray data, a crystal is positioned and fixed between a detector that detects diffracted X-ray and an X-ray source that produces powerful X-ray beams (Figure 5). Upon X-ray emission, the crystal diffracts the beam and produces scattered beams to be received by the detector. As the crystal is continuously rotated (0-360 degree) when mounted in place, different diffraction images are produced, recording different intensities and angles of scattered beams from corresponding orientations of the crystal. Each scattered beam results in a particular spot on the film with certain darkness and position, and these spots are called reflections. Rotating the crystal will yield distinct reflections with respect to the orientation, and each recorded diffraction pattern on the film is the two-dimensional array of reflections positioned in a cross section of a larger hypothetical three-dimensional lattice of reflections

called reciprocal space [4] (Figure 5). The spacing of the unit cells in the real crystal lattice, called the real lattice, and the spacing of reflections on the film, called the reciprocal lattice, are inversely related, thereby allowing relatively straightforward calculation of the dimensions of the unit cell from the macroscopic parameters of the reflections. For the location of each reflection in the reciprocal lattice, three coordinates h, k, l are used. The central reflection (i.e. the centre of the film where undiffracted X-rays pass straight through) is referred to as the origin and assigned the coordinates $(h, k, l) = (0, 0, 0)$ or simply $hkl = 000$. Designation of other reflections follows similar trend with whole-number coordinates.

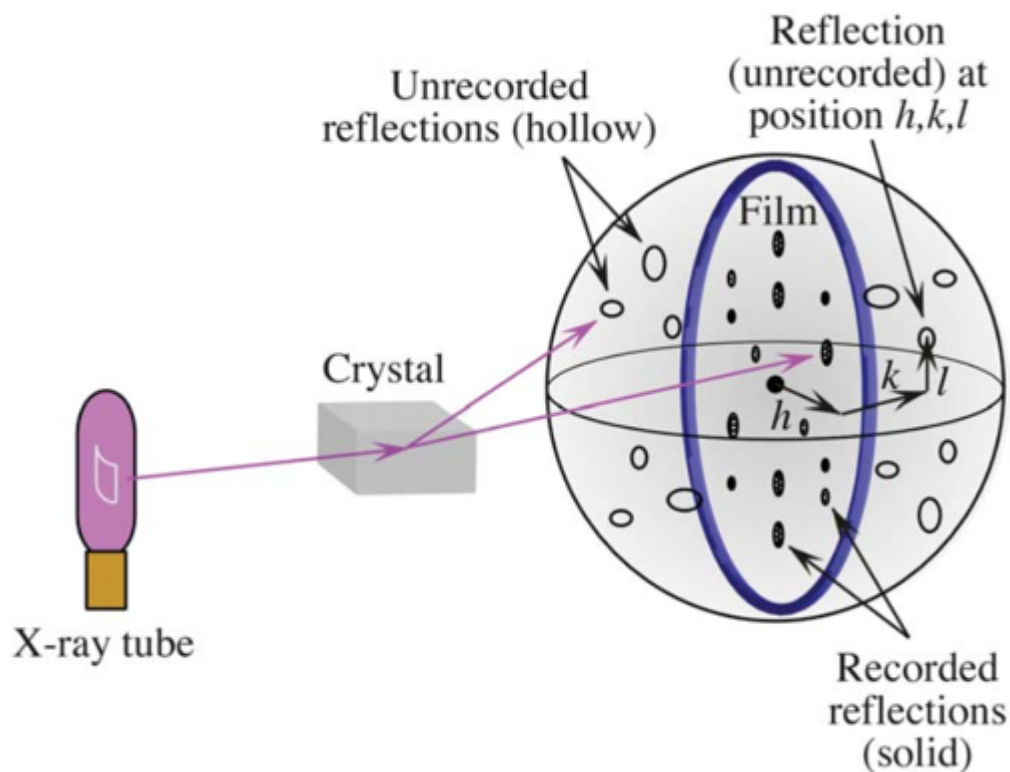


Figure 5. Single crystal X-ray diffraction by beams emitted from an X-ray tube as an example of X-ray source.

The dark blue circle represents a plane in the three-dimensional reciprocal sphere on which the reflections from the specific orientation of the crystal (solid dots) pass through. Unrecorded reflections that are available to be produced by other orientations of the crystal but not recorded are represented by the hollow dots. Each reflection is assigned by the three-dimensional coordinates h, k, l . (Figure taken from [4])

1.5.1 Data indexing, integration, scaling and merging

Once the data set is collected, downstream data analysis is feasible with specific computational programmes. Data indexing is termed as “identifying the dimensions of the unit cell and which image peak corresponds to which position in reciprocal space”. [8] During indexing, each reflection is designated by Miller indices h, k, l to specify individual position of each reflection within the reciprocal space, and the central reflection is regarded

as the origin with the coordinates $hkl = 000$. Along with the intensity of each reflection, I_{hkl} , these data are used to determine the dimensions of the unit cell and the symmetry of the crystal, i.e., its space group.

Once the space group has been assigned, the data is then integrated, a process in which the intensities of these reflections are measured and individually labeled, converting the bundle of collected reflection images into a single file that lists the Miller index and the intensity of each reflection. Representative indexing and integrating computer programmes include iMosflm [9] and HKL-2000 [10].

In the following scaling and merging process, the reflections with the same indexes from more than one data set or more than one frame are compared and set with a common relative scale so that identical reflections are assigned identical intensities. In this way the discrepancy of intensities between these identical reflections, caused by factors including different diffracting ability of crystals, radiation damage of the crystals to varying extent, and difference in the distance X-rays travel with respect to different orientations of crystals, can be appropriately addressed. In practice, the intensity of each frame of reflections is usually multiplied by a scale factor that is determined by a least-squares procedure and the frames are merged into a single data set. In the meantime, a merging R -factor is applied to indicate the degree of consistency between the averaged and scaled intensities of all observations of one reflection and averaged and scaled intensities of individual observations of the identical reflection distributed among different frames. A programme called Scala in CCP4 suite, for example, processes MTZ file (a file format that stores reflection data) of unmerged intensities produced from iMosflm, scales multiple observations of identical reflections and merges the observations into an average intensity. POINTLESS, another programme commonly used in automatic data processing, is used to predict the space group of the crystal by estimating and

choosing the point-group symmetry or Laue group in the lattice with the largest possibility of being correct (smallest difference with ideal lattice) [11].

1.5.2 Obtaining phases

Each reflection possesses its own phase, and in order to calculate $\varrho(x, y, z)$ one has to obtain phase estimate of every single reflection involved in the process. For emphasizing the importance of phase information, the equation of $\varrho(x, y, z)$ can be expressed as the following:

$$\varrho(x, y, z) = \frac{1}{V} \sum \sum \sum F_{hkl} e^{-2\pi i(hx+ky+lz-\alpha'_{hkl})},$$

where α'_{hkl} are the unknown phases of each reflection.

The lost phase information can be retrieved by several popular and efficient methods, including the heavy atom method (isomorphous replacement), anomalous scattering or dispersion, and molecular replacement. The phases obtained by these methods are initially estimates, which need subsequent improvement and extension prior to constructing an accurate and reliable electron density map.

1.5.3 Heavy atom method

Normal routine single-crystal X-ray diffraction analyzes native protein crystals composed of aggregates of amino acids synthesized by endogenous or intracellular sources, and each atom and its corresponding diffracting contribution are regularly repeated in the crystal, which produces protein-specific reflection pattern on the film. Introduction of an additional atom, usually a strong diffractor, to specific and repeating locations in every unit cell in the crystal will constantly alter the reflection pattern because of the extra diffracting contribution of the added atom. The location of the added atom depends on the interactions with certain amino acids in the protein, and typically would not cause interference in protein conformation and existing crystal packing. Common choices of heavy atom include Hg and Pt as well as the halogens including bromine (Br) and Iodine (I). Iodine has the atomic number 53, it produces

significant anomalous signal although its absorption edge is not easily accessible (K absorption edge of 0.3738 Å). Generation of iodide derivatives of the crystals of the BAM complex and subsequent anomalous diffraction greatly assisted in structure determination.

Once the heavy atom is incorporated in the crystal, usually by soaking the crystal in the solution of heavy atom, the locations of those heavy atoms in the unit cell can be revealed by a Fourier sum called the Patterson function $P(u, v, w)$, the letters in the parentheses being the spatial coordinates in the Patterson map. The equation of the Patterson function is provided below:

$$P(u, v, w) = \frac{1}{V} \sum \sum \sum |F_{hkl}|^2 e^{-2\pi i(hu+kv+lw)}$$

It has been shown that in a real unit cell that contains n atoms would result in $n(n-1)$ Patterson atoms in a Patterson unit cell. Each Patterson atom is a peak in the map. After applying trial and error of choosing specific atoms, the correct Patterson atoms that construct the Patterson map matching the calculated map are indicative of the locations of the real heavy atoms in the unit cell.

Despite localization of the heavy atoms, the phases of the reflections can still be ambiguous. Using the Harker diagram, the structure factor of the native reflection, F_p , the structure factor of the heavy atom, F_H , and the structure factor of the heavy atom soaked crystal, F_{pH} , can be illustrated as vectors in a plane. As the diffracting contributions of all atoms are accumulative in a structure factor as well as in amplitude of a structure factor, the following relationship holds for the three structure factors:

$$F_{pH} = F_H + F_p$$

$$|F_{pH}| = |F_H| + |F_p|$$

In case of phase ambiguity, a second heavy atom soaked crystal is adopted and data set is collected. The second heavy atom is required to bind and reside in different locations in the unit cell from the first heavy atom, otherwise the phases of F_H would be identical in both derivatives and this overlap cannot solve the issue. A second suitable and non-overlapping heavy atom can produce a phase solution that satisfies one of the estimates provided by the first heavy atom, thereby narrowing down the favourable choice in the diagram. It is worth noting that in some cases application of three or four heavy atom derivatives may be necessary for appropriate judgment of phase estimates.

Another means to obtain phase estimates using heavy atom derivatives is the anomalous scattering technique, which deploys the unique characteristic of the heavy atom in terms of absorbing X-rays. All chemical elements absorb and discharge X-rays, and, at a wavelength near the element-specific emission wavelength, the absorption decreases dramatically, a phenomenon called an absorption edge. Light atoms in most native protein crystals do not incur anomalous scattering since the wavelengths of X-rays in diffraction analysis are not near the absorption edges of them. Heavy atoms, however, do possess absorption edges within the range and can be used for proper anomalous diffraction with variable wavelengths of X-rays.

Single-wavelength anomalous diffraction (SAD) is an example of anomalous diffraction using a single specific wavelength to maximize anomalous scattering by the heavy atom. When anomalous scattering is triggered by the heavy atom, a small part of X-rays is absorbed by the heavy atom and released with a different phase. This in turn leads to unevenness of intensity of symmetry-related reflections in the reciprocal lattice that would otherwise be identical in non-anomalous diffraction. It is this discrepancy of the intensities in the Friedel pairs caused by anomalous scattering that provides clues of phase information in the native

data set.

There is another type of anomalous diffraction directed at sorting the phase problem, and it is called multi-wavelength anomalous diffraction (MAD). As the name suggests, the technique employs the advantage of variable wavelengths of X-rays and the fact that the measured intensities of Friedel pairs varies greatly with different wavelengths in the absorption spectrum of the heavy atom. Consequently, phase information can be deduced from distinct data sets, resembling the effects of each data set collected from distinct heavy atom derivatives described previously. Furthermore, recorded reflection intensities of MAD also vary to some degree with variable wavelengths, which carries valuable phase information that can also be interpreted for phase solution.

Selenium is a popular choice of anomalous scatterer frequently used to replace sulfur in protein residues and produce SAD and/or MAD. It has the chemical symbol Se and atomic number 34. It has a K absorption edge maximum at 0.9793 Å and an absorption edge inflection point at 0.9794 Å [12]. Incorporation of selenium into the BAM complex via selenomethionine (SeMet) labeling during *in vivo* overexpression was employed to help solve the structure of the BamACDE complex.

1.5.4 Molecular replacement

In certain cases the protein under study is of similar structure and conformation to a known homologue, which can be informed by sequence identity and homology search, and the molecular model of it can be used to calculate initial phases of the unknown protein in the process of structure determination. This method is called molecular replacement. Where applicable, the method is able to solve the structure of the unknown protein with only the native data set and without the heavy atom derivatives. Generally, the known protein model is required to have a sequence identity of at least 30% with the protein under study [13]. In

the simple example of isomorphous phasing, a phasing model of a ligand-free protein can be applied to the molecular replacement process of a crystal of the same protein in the ligand-bound state, thus directly enabling phase calculation and structure determination. The equation of $\rho(x, y, z)$ in the case of molecular replacement is defined as the following:

$$\rho(x, y, z) = \frac{1}{V} \sum \sum \sum |F_{hkl}^{target}| e^{-2\pi i(hx+ky+lz-\alpha_{hkl}^{model})}$$

Where $|F_{hkl}^{target}|$ is the amplitudes of the recorded reflections in the native data set, and α_{hkl}^{model} are the phases of the phasing model.

Non-isomorphous phasing, on the other hand, is a more common but complex situation in which the phasing model is not isomorphous with the unknown structure and hence needs to be superimposed on the unknown structure in order for phasing to take into effect. To superimpose the unknown protein in the unit cell, the phasing model is required to be in the desired location and orientation, which demands separate searches of the two factors. Orientation search can be carried out independently and prior to location search using the Patterson function. The Patterson map can be rotated with respect to the rotation of the phasing model, and the same-oriented unknown and phasing models would result in similar Patterson maps. With the efforts of trial and error in comparing the two Patterson maps, a best orientation of the phasing model can be defined and subsequently the position of the model can be revealed using the structure-factor manner. In practice, structural homology databases including DALI [14] and Phyre2 [15] are representatives of comprehensive online tools that are contributory in searching and identifying candidates of structural homologues to be trialed in molecular replacement.

1.5.5 Phase refinement, model building and structure refinement

The initial electron density map may be preliminary and ambiguous due to the limited accuracy and estimated nature of the phases obtained. The map can be improved by an

iterative process that can be referred to as bootstrapping. In the process, the contents in the current map that can be uniquely assigned and interpreted by modifications including solvent flattening, partial map fitting are retained and incorporated in the next map, which is manipulated in the form of an electron-density function to calculate new structure factors using the following equation:

$$F_{hkl} = \int x \int y \int z p(x, y, z) e^{2\pi i(hx+ky+lz)} dx dy dz$$

The calculated new phases are subsequently used in the equation of electron density in attempt to produce a clear and more interpretable map. In this way, the rounds of iterative improvement will finally produce a high-quality and detailed map in which the molecular model of the protein can be built. An automated model building programme called Buccaneer is a typical model-building tool in which it uses a chain-tracing procedure to generate a partial trace of the polypeptide backbone. In particular, possible locations and orientations C^α are identified by a density likelihood target function employed in the programme in a six-dimensional search against a simulated density map of a known structure. The process can be summarized in a number of steps including identifying C^α locations, building C^α backbones with reference to Ramachandran bond angle restrictions, connecting discontinuous chain fragments into continuous chains, and reducing clashing C^α atoms within the proximity of 2.0 Å. In essence, the overall degree of interpretation is dependent on the accuracy of the initial phasing and phase refinement, and the resolution of the electron density map [16], and Buccaneer is normally compatible with a dataset of more than 3.5 Å resolution. But clearly automated model building is not a perfect substitute for manual model building and it often results in considerable unacceptable errors and misinterpretations. Manual building is needed in conjunction with automated building to compensate for these mistakes. Coot is one of the graphical interactive programmes for manual model building [17].

During refinement process of the phases, monitoring and eliminating biased errors are another essential part in order to help with the process. Introduction of the $(\ln |F_{\text{obs}}| - |F_{\text{calc}}|)$ maps, in which the calculated indexed intensities are deducted from multiples of the observed intensities within each Fourier term, can efficiently minimize the influences and sometimes the dominance of the phases of the model. When $n = 1$, the corresponding map is called a $F_o - F_c$ difference map. The calculated positive or negative values, usually represented by positive or negative peaks in various colours, indicate wrong positions of the model that need to be replaced with and missing portions that need to be fit in. In the final stages of structure determination, a $2F_o - F_c$ map is commonly used to identify minute errors and reach a final agreement between the model and the original data set. REFMAC5 is frequently used to refine models with every building cycle, and the optimization parameters set in the programme include atomic coordinates, atomic displacement, scale factors and twin factors in the case of twinning [18]. As an indication of the level of the agreement, an R -factor is defined as the following equation [19]:

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

It ranges from 0 to ~ 0.6 , and an initial value of below 0.5 would be promising in further improving the model as a starting point. Another similar indicator is called the free R -factor, R_{free} , which is calculated from a subset ($\sim 10\%$) of randomly chosen reflections that are not included in the structure refinement and for cross-validation purposes. It is interpreted as a measure of the degree of prediction the current model reflects on a subset of the excluded intensities. An indicative rule states that R_{free} should be approximately the resolution in angstroms divided by 10, and both R -factor and R_{free} values should converge to a certain extent during the final stages of structure refinement.

CHAPTER 2

General Introduction

2.1 Bacterial membranes

Bacterial cellular contents are enclosed by a cell envelope that defines the shape and volume of the cells, protects them from the hostile and constantly-changing environment, and allows them to import nutrients and export waste. Bacteria can be classified as Gram-positive or Gram-negative, according to the structure of the cell wall (Figure 6).

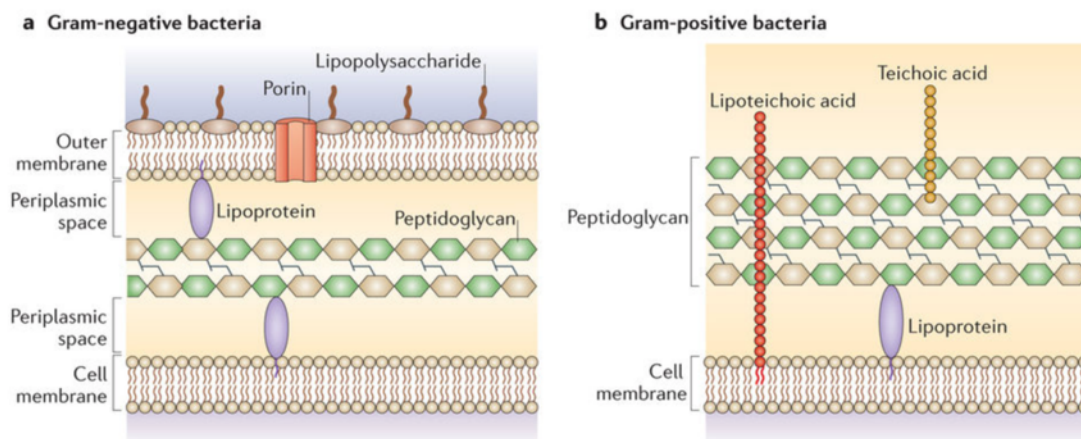


Figure 6. Structures of the cell wall in Gram-negative bacteria (a) and Gram-positive bacteria (b).

The cell wall of Gram-negative bacteria is composed of an outer membrane and an inner membrane; between them a thin layer of peptidoglycan is sandwiched. Gram-positive bacteria are instead surrounded by a symmetrical single lipid membrane coated by extracellular thick layer of peptidoglycan and additional lipoteichoic acid. (Figure adapted from [20])

Gram-positive bacteria are distinguished from Gram-negative bacteria by the presence of a single lipid membrane and a thick layer of peptidoglycan and lipoteichoic acid that is anchored in the membrane by diacylglycerol [20] (Figure 6b). In contrast, Gram-negative bacteria contain two membranous structures that constitute their cell envelopes: an inner cytoplasmic cell membrane and an outer cell membrane, with the inner membrane (IM)

located at the membrane-cytoplasm interface and the outer membrane (OM) faced with the extracellular environment. The parallel organization of the two membranes gives rise to an additional periplasmic space in between, in which a relatively thin layer of peptidoglycan is sandwiched (Figure 6a).

Most of the membranes in bacterial cell wall are composed of phospholipids. Listed in order of abundance, phosphatidylethanolamine (PE), phosphatidylglycerol (PG), and cardiolipin (CL) are the prevalent types [21]. In the IM of Gram-negative bacteria and the single cytoplasmic membrane of Gram-positive bacteria, the two leaflets of lipids are of significant symmetry compared to the highly asymmetric composition of lipid in the OM of Gram-negative bacteria, in which another type of lipid called lipopolysaccharides (LPS) dominantly occupies the outer leaflet of the OM (Figure 6a). LPS consists of three structural domains: lipid A, a core oligosaccharide, and an O-antigen polysaccharide. Among the three domains the lipid A domain is responsible for anchoring LPS to the OM via hydrophobic interactions [22].

Apart from the phospholipids and LPS molecules, there are various types of membrane proteins that are either co-translationally or post-translationally transported and inserted into the membranes and they constitute integral elements in the bacterial envelope. They can be classified based on their positions with respect to the membrane, which leads to the categorization of integral membrane proteins, peripheral membrane proteins and lipid-anchored proteins [23, 24]. Different from Gram-positive bacteria, integral membrane proteins in Gram-negative bacteria can be further divided into two types subject to their locations in the two membranes. The integral membrane proteins of the IM are referred to as inner membrane proteins (IMPs), they form membrane-spanning α -helices embedded in the IM [25, 26]; and the integral membrane proteins of the OM are called outer membrane

proteins (OMPs), and they form β -barrel domains consisting of anti-parallel β -strands to cross the OM [27, 28].

2.2 Outer membrane proteins (OMPs)

The OM is a distinctive cell wall component of Gram-negative bacteria and provides the cell with an extra level of structural protection against external detrimental stimuli. The physical barrier also serves as a permeability regulator that controls the influx and efflux of solutes through the functions of pore-forming integral outer membrane proteins (OMPs). In addition to the OM of Gram-negative bacteria, the OM of chloroplasts and mitochondria in eukaryotic cells also contain β -barrel OMPs [29-31]. More than a passage for solutes, bacterial OMPs are involved in various cellular processes such as nutrient uptake, environmental signal transduction and antimicrobial resistance. Taken together, OMPs are essential for the structural integrity of the OM and their correct biogenesis is vital for cell viability.

2.3 OMP structures

Visual comparison of all the available structures of OMPs in Gram-negative bacteria thus far reveals that most OMPs contain a transmembrane β -barrel domain. Selected structures are presented in Figure 7. The number of observed strands in all structures varies from 8-26, and nearly all β -barrels contain even number of strands [32]. Even numbers of strands are very likely to be the result of divergent evolution, whereby gene duplication of a single β -hairpin element produces alike building blocks [33], and it is thought that four such repeating units are the prerequisites for ultimate completion of barrel construction through inter-strand hydrogen bonding force [32].

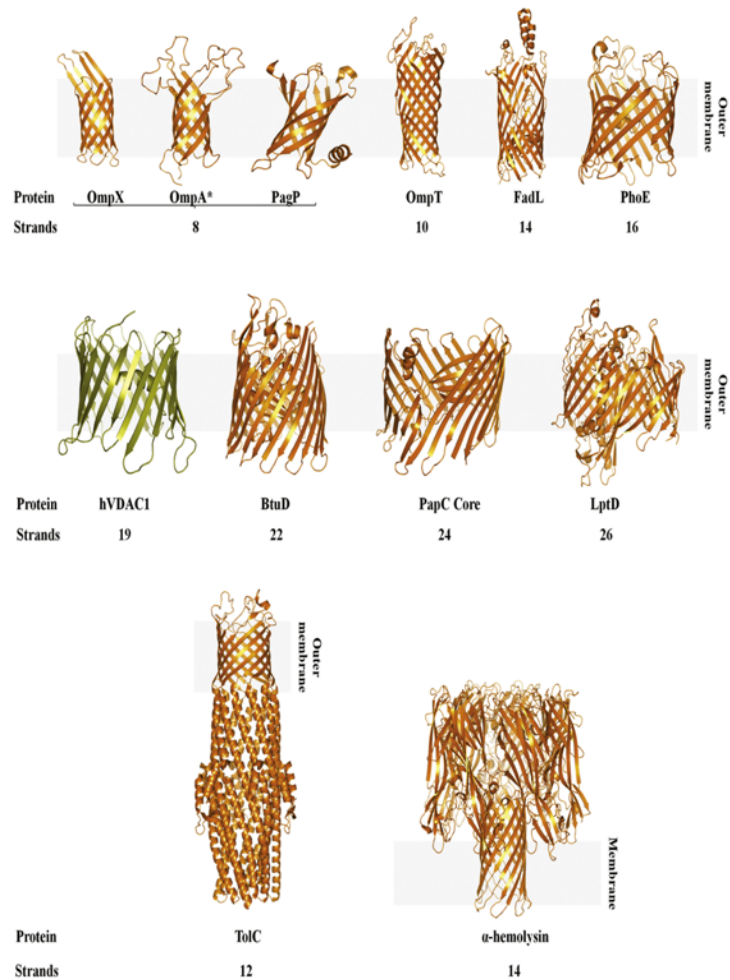


Figure 7. Structures of selected β -barrel OMPs.

Each structure is represented by golden cartoon and with names and number of β -strands, except that the mitochondrial voltage-dependent anion channel (VDAC) is coloured in olive. Example of 18 stranded OMPs is not shown here. All structures are obtained from the Protein Data Bank (PDB) and rendered using PyMOL. (Figure taken from [32])

The β -strands in the barrel adopt antiparallel conformation, and the side chains of the barrel-forming amino acids protrude from both faces of β -sheets. Due to the hydrophobic nature of the transmembrane region, the hydrophilic residues are consequently found mostly lining the barrel lumen, whereas the hydrophobic residues of the barrel face towards the external membrane environment, primarily making contact with the hydrocarbon tails of the lipids in the OM via hydrophobic interactions [34]. At the water-bilayer interface, prevalence of

aromatic residues in OMPs, especially tyrosine and tryptophan, is found to interact with the hydrophilic head of the lipids and likely results in stabilization of the barrel in the lipid bilayer of the OM [35].

The other two notable features in OMP structures are the generally longer loops located on the extracellular side and the shorter turns on the periplasmic side, both of which are required for connecting antiparallel β -strands. Composed of three or four residues, periplasmic turns form a sharp and tight bend at the end of a β -strand and reverse overall direction of the polypeptide chain [36]. Extracellular loops, however, are longer bends that extend the barrel domain further into the extracellular environment, and can be in flexible and dynamic structural conformations compared to the rather rigid and confined structure of turns. Depending on the protein type, the extracellular loops carry diverse functional roles as part of OMP biological functions.

2.4 OMP biogenesis

Biogenesis of all bacterial OMPs takes place in the cytosol, and, due to temporary unfolded state, they must be correctly delivered to their destined locations in either the IM or OM before folding into mature three-dimensional entities with tertiary or quaternary structures and carrying out biological functions. OMPs are firstly synthesized in the cytosol with the co-translated N-terminal signal sequence that is critical for recognition towards the IM and cleavable upon IM translocation [37, 38]. When emerging from the ribosome, the nascent polypeptide chains are appropriately recognized and coupled with the ribosome-associated trigger factors (TFs) [39, 40]. TF is of broad substrate spectrum and it can stay coupled with the unfolded chains after they leave the ribosome [41]. TF subsequently transfers the OMP to the soluble complex of SecA/SecB protein that possesses chaperoning function and targets precursor proteins to the IM [42]. SecA is identified as a SecYEG-binding ATPase and serves as the motor that initiates OMP translocation. Upon binding to SecYEG complex, transfer of

OMP substrate from SecB to SecA takes place and SecB is finally released [43], paving the way for subsequent translocation with the engagement of SecA that utilizes ATP to power the process [44, 45]. An alternative scenario of IM association is the direct contact of the translating ribosome with the SecYEG translocase, the core complex in the “Sec pathway” [46, 47], inducing co-translational insertion of OMP into the Sec machinery [48, 49] (Figure 8). Having passed through the translocon, the translocated OMPs enter the periplasmic space and the N-terminal signal sequence is cleaved by a type of membrane-bound endopeptidase called signal peptidase I (SPaseI) [50]. Upon cleavage, periplasmic nascent OMPs are further accompanied by either SurA or Skp/DegP, the periplasmic chaperones, in order to maintain current folding state and facilitate subsequent folding events while directing them to the boundary of the OM [51, 52] (Figure 8). It has been proposed that most OMPs are escorted to the OM by SurA, and Skp/DegP provide supportive function to either deliver any remaining off-pathway OMPs back to the SurA pathway or bring them directly to the OM [51]. It is worth noting that DegP exhibits both protease and chaperone function [53], thus also capable of degrading off-pathway OMPs.

In the final context of biogenesis, OMPs fold and insert themselves into the OM either spontaneously or with the help of other OMPs [54]. Spontaneous self-folding occurs mostly *in vitro*, and, in the example of OmpA, the individual transmembrane β -strands are inserted with considerable coordination and it was found that the ending phase of secondary structure building and closure of the β -barrel are coincidental [55]. However, *in vitro* spontaneous folding is time consuming and not kinetically favourable even though it does not require external metabolic energy [56]. In fact, *in vivo* cellular context has evolved to deploy specific proteinaceous machinery to accelerate the folding and insertion, which, together with the aforementioned chaperons, dramatically shortens the timescale from up to hours down to minutes and even seconds [54, 57]. The core protein complex embedded in the OM

recognizes the well-protected and competent OMP substrates, accommodate the substrates within the entity, and finally release them into the OM [58]. This multi-subunit complex is called the β -barrel assembly machinery (BAM) (Figure 8). It has been found that the BAM complex in *E. coli* recognizes the substrates via their C-terminal signature sequences with a consensus sequence of X-Z-X-Z-X-Z-Tyr-Z-Phe/Trp, where X is hydrophobic amino acid and Z is a random residue [59-62], and that the BAM complex displays species-specific recognition while excluding OMPs from a different strain [61].

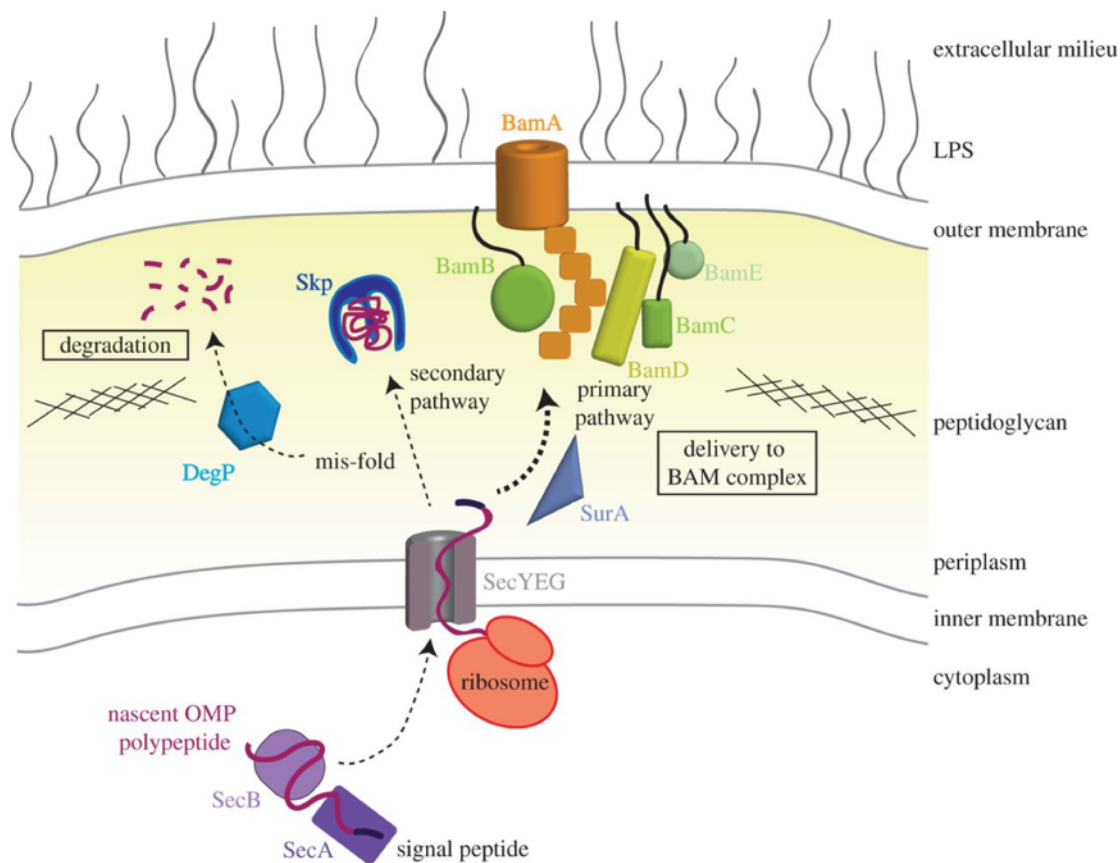


Figure 8. OMP biogenesis pathway in Gram-negative bacteria.

Nascent OMP polypeptides (maroon) are synthesized in the bacterial cytoplasm and directed to the Sec translocase (grey) by the SecA/SecB chaperones (purple and magenta) or the ribosome (red) for translocation across the IM. Then the OMP, in the periplasmic space, is delivered to the BAM complex by the chaperones SurA (slate) or Skp (blue). Incorrectly and accidentally misfolded OMPs during the path are targeted and degraded by the protease DegP (light blue) to eliminate aggregation. (Figure taken from [63])

2.5 Identification of the BAM complex

Recent years have seen research advancements in the knowledge of OMP assembly during bacterial cell growth and membrane biogenesis, and a more comprehensive understanding of *in vivo* folding and insertion of OMPs by the BAM complex is emerging since the discovery of this essential assembly factor. The first protein subunit identified is BamA in 2003 [64],

and it was initially named YaeT or Omp85. The original species from which BamA was identified is *Neisseria meningitidis*, but it was later realized the homologues of this protein are universally spread across Gram-negative bacteria as well as in the double membrane-bound eukaryotic organelles of mitochondria and chloroplasts [65, 66]. The presence of Omp85 in the study was found to be essential for both viability and folding and assembly of all the chosen protein substrates examined [67], and intervention of Omp85 expression through conditional regulation led to abnormal aggregation of unfolded OMPs stalled in the periplasmic space [67, 68]. Later on, Omp85 homologues of *E. coli* [69, 70] and *P. aeruginosa* [71] were identified and termed BamA and Opr86 respectively, and evidence of direct participation of BamA in OMP biogenesis was provided in a study that demonstrated BamA binds unfolded OMPs *in vitro* [72]. Following the characterization of BamA without well-known functional roles, co-immunoprecipitation experiments further revealed other four interacting proteins named YfgL (BamB), NlpB (BamC), YfiO (BamD), and SmpA (BamE) [73, 74]. All of the four components are lipoproteins, have various molecular weights, and are assembled to form an intact BAM complex *in vivo*. In summary, the BAM complex consists of five protein components: BamA, an integral OMP itself, and four other lipoproteins named BamB, BamC, BamD and BamE, all of which are located primarily in the periplasm and anchored to the outer membrane via post-translational lipid modifications [75-79].

Each of the five protein components is found to have different distributions across Gram-negative bacteria species. Distinct from BamA, the homologues of which are found in all bacteria with outer membranes, homologues of BamB, BamC, and BamE are not present in δ - or ϵ - proteobacteria. BamC is found in all species of β - and γ -proteobacteria but completely absent in α -, δ - and ϵ -proteobacteria, proving it to be the least conserved subunit of the BAM complex [80]. BamD is another well-conserved component across

proteobacteria, and it was proposed that the most ancestral BAM complex was made up of only BamA and BamD, and the remaining three lipoproteins were subsequently acquired during the course of evolution [80]. In line with the conservation profile, only BamA and BamD are essential for cell viability, and BamD is crucial for both structural and functional integrity of the BAM complex [81, 82]. It was proposed that BamD is implicated in proper recognition of OMP substrates by the complex [83], and related studies of BamD homologue in mitochondria supported this claim [84]. Further to the involvement in substrate recognition, BamD was also thought to mediate BamA conformation via its direct interactions with the periplasmic domain of BamA during OMP assembly [85, 86], thus allowing coordinated movements of these two essential components. BamA and BamD depleted strains are abundant in unfolded and accumulated OMPs in the periplasmic space and display severe growth defects [64, 87]. Depletion of *yfgL* gene encoding BamB leads to various cellular deficiencies including significantly reduced levels of major OMPs in *E.coli*, sensitivity to a number of membrane-impermeable agents and a bactericidal permeability-increasing peptide, which indicates increased membrane permeability profile [87, 88]. Strains lacking BamC and BamE exhibit compromised OM function and increased permeability albeit with modest degree and no apparent growth defect [74, 87], and, in the absence of BamE, BamA exhibits conformational alteration resembling that caused by an activating mutation in *bamD* [74, 89]. There is some evidence suggesting that although BamB, BamC and BamE are not essential, they are required for assembly of specific OMP substrates, as suggested by single mutations in these lipoproteins [90, 91]. Taken together, it is plausible that BamA and BamD are indispensable for OMP biogenesis and essential for maintaining a functional BAM complex, and BamB, BamC and BamE are secondary members that further preserve the integrity and enhance the efficiency of the BamAD core subcomplex.

2.6 Structures of individual components of the BAM complex

Structures of all five proteins have been solved individually [76, 92-102] (Figure 9). BamA is composed of a large C-terminal β -barrel transmembrane domain and five periplasmic polypeptide translocation-associated (POTRA) domains that extend from the barrel and into the periplasm [103, 104]. BamB is an eight-bladed β -propeller, and positioned in the two loops on one side of the propeller are residues that have been shown to interact and cross-link to BamA [105]. BamC possesses three domains: an unstructured domain at the N-terminus, a N-terminal domain and a C-terminal domain. BamD contains ten α -helices that arrange into five tetratricopeptide repeats (TPRs) [93, 94]. BamE consists of two α -helices and a three-stranded β -sheet. The exact roles of the four lipoprotein components in the process of OMP biogenesis remain to be identified, yet, from structural perspective, it has been proposed that the four lipoproteins interact with and assemble onto the POTRA domains to form a stable and established copy of the heteropentamer[77].

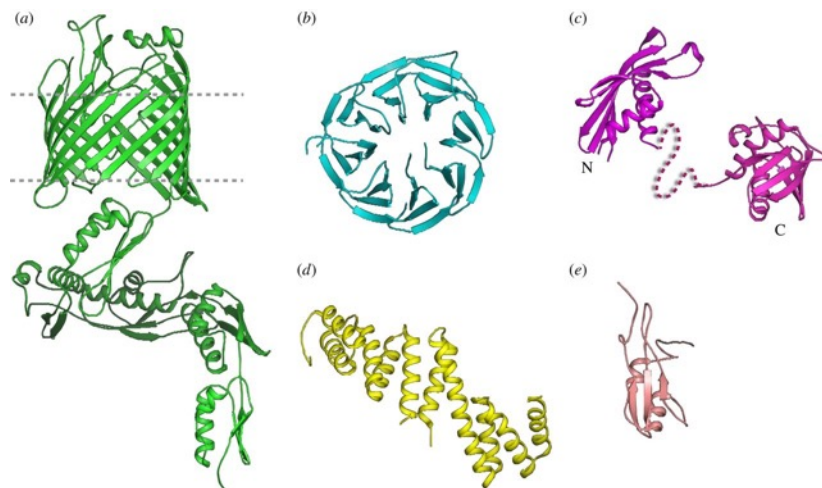


Figure 9. The structures of the five components of the BAM complex.

(a) The crystal structure of full-length BamA from *Neisseria gonorrhoeae*. (b) The crystal structure of BamB from *E. coli*. (c) The crystal structure of the two globular domains of BamC from *E. coli*. The dotted line between N and C domains indicates an ~ 18 -residue linker. (d) The crystal structure of BamD from *E. coli*. (e) The NMR structure of BamE from *E. coli*. (Figure taken from [63])

2.7 Proposed mechanisms of OMP assembly by the BAM complex

Based on both the structural features of BamA revealed from the crystal structures that the first β -strand and the last β -strand in the barrel domain are weakly associated by few hydrogen bonds and the molecular dynamics simulations that the barrel domain of BamA distorts the local membrane environment, two types of insertion mechanism by the BAM complex were proposed: the BamA-assisted model and the BamA-budding model (Figure 10) [63, 76]. The BamA-assisted scenario derived from the aforementioned capability of small and simple nascent OMPs to spontaneously arrange and fold themselves into the OM without external assistance, the structural observation that the width of the hydrophobic belt of the BamA barrel is much thinner along the C-terminal strand and the reported ability of BamA to destabilize the local lipid layer and reduce the thickness of the OM concluded by molecular dynamics simulations [76] (Figure 10a). In this model, nascent OMPs are accompanied by chaperones and escorted to the OM boundary, where they are recognized by the BAM complex. The machinery distorts the local membrane layer and hence reduces the kinetic barrier for protein folding [106-108], allowing folding and insertion to occur. Consistent with the model, studies showed that reconstitution of BamA into phosphoethanolamine-based liposomes could speed up OMP folding rate and the thickness of the phospholipid bilayer was lessened in proteoliposomes that contain BamA [107, 109]. This approach was thought to be dependent on the local interactions between the β -barrel and the OM but was challenged recently by the structural hint and molecular dynamic simulations of the complex that the POTRA domains and the other four lipoproteins may also participate in the interactions of the complex with the OM [110, 111].

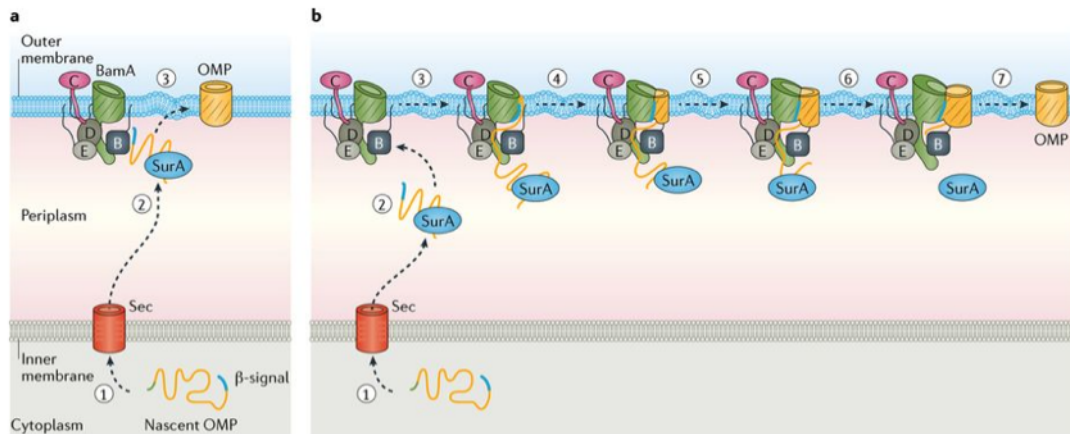


Figure 10. Two proposed mechanistic models of the BAM complex during OMP biogenesis.

(a) The BamA-assisted model. Nascent OMPs are synthesized in the cytoplasm and then transported across the inner membrane and into the periplasm (step 1). SurA or Skp escorts the nascent OMPs and direct them to the BAM complex (step 2). The complex distorts the local membrane bilayer and directs the OMP substrates to close proximity to the primed membrane for insertion into the outer membrane (step 3). (b) The BamA budding model. As with the BamA-assisted model, this model shares step 1 and step 2. Nascent OMPs are next delivered to the periplasmic subunits of the BAM complex and the interactions initiate folding and insertion processes (step 3). Lateral opening of the barrel gate of BamA possibly enables binding of a β -hairpin with the exposed N-terminal strand of BamA and the barrel of the OMP substrate then fully integrate into the barrel of BamA as the barrel of the substrate continues to grow (step 4-6). Eventually the OMP substrate buds away from the barrel domain of BamA possibly by unpairing of the last strand of OMP from BamA and subsequent pairing with its own first strand (step 7). (Figure taken from [112])

The BamA-budding model, on the other hand, is thought to be specific for OMPs of more complex structures including those possessing larger domains and more strands and therefore unable to undergo spontaneous insertion and folding (Figure 10b). In this scenario, chaperone-escorted nascent OMPs are recognized by POTRA domain of BamA and then

accommodated by the complex in the barrel lumen, forming a BamA-OMP intermediate. The lateral opening of the barrel would serve as both a template for synthesis of new β -strands and a path for insertion of the substrate into the OM. As the folding continues, the OMP substrates would reach a stage in assembly where they are too large to be rendered in the BAM complex and consequently bud from and extend away the BamA lateral opening. This completes the budding model and produces a folded and functional OMP. In line with the model, there is a few well and commonly known facts that the barrel of BamA is among the ones with lowest thermostability when comparing to other known β -barrels [106] and it has a melting temperature of $\sim 37^\circ\text{C}$, implying that it may be unstable in physiological conditions.

2.8 General porins

OMPs in Gram-negative bacteria are faced with the extracellular environment and, in the interest of the bacteria to survive the harsh and hostile contexts, they serve as the molecular sieves in conjunction with the relatively impermeable OM to allow diffusion of small hydrophilic molecules (less than 600 Dalton) across the OM through hydrophilic channels formed by general porins. In the meantime, these hydrophilic channels prevent entry of large molecules and promote selective entry of essential nutrients [34]. This type of diffusion driven by porins is subject to the dimensions of the pore in porins and considered as no specific substrate preference, although some show cation- or anion- selectivity due to the amphipathic properties of the residues lining the boundary of the pore. General porins constitute large portion of total OMPs in *E. coli* and it was estimated that 10^5 exist per cell [113]. The first structural insight into general porins was the crystal structure of a porin from *R. capsulatus* [114]. Later on, the atomic structures of OmpF and PhoE porin were obtained [115]. It was predicted that these general porins assemble into stable trimers solely composed of β -strands [116], and the high-resolution structures from these porins confirmed the typical model in that 16 β -strands constitute a monomer in the homotrimer conformation.

2.9 *Vibrio cholerae* OmpU

Vibrio cholerae is the causal organism for the disease cholera. Serogroups O1 and O139 are responsible for the epidemics. In the small intestine, the major cholera toxin along with other virulence factors is synthesized and under the genetic modulation of the *toxR* regulon. There are two other genes, *ompU* and *ompT*, encoding two major OMPs, that are also regulated by the transcriptional activator ToxR [117, 118].

OMPs in *V. cholerae* are involved in host-pathogen recognition as well as pathogenic invasion of host cells [119, 120]. They mainly behave as adhesion proteins [120, 121] or invasion proteins [122, 123]. Among numerous OMPs in the organism, Outer membrane protein U (OmpU) is a major component that constitutes about 30% of the total outer membrane proteins when *V. cholerae* is grown in medium containing 1% NaCl and nearly 60% when grown in salt-free medium [124]. It is characterized as a porin that forms non-specific β -barrel channels allowing for diffusion of hydrophilic molecules across the OM. Apart from its native function, OmpU has been shown to be implicated in bile resistance and resistance of the organism against antibacterial peptides [125, 126].

2.10 Implications in host-pathogen interactions

OmpU has been shown to be associated with the process of bacterial adhesion during *V. cholerae* infection [127], as other OMPs do mentioned previously. In a study, OmpU from *Vibrio mimicus*, a similar species to *V. cholerae*, has been found to be an immunoprotective antigen with six immunodominant linear B-cell epitopes [128]. In addition, it has been further proposed to possess adhesion property by using residues 90-101 and 173-192 as potential binding motifs during bacterial invasion [120]. Strikingly, OmpU from *Vibrio splendidus* strain LGP32 was also reported to mediate invasion of host hemocyte by serving as both an adhesin and invasin [119].

2.11 Structural information of OmpU

Naturally assembling as trimers, OmpU form triple-barrel channels, and each β -barrel surrounds an individual pore. The pore size was determined to be about 1.1 to 1.6 nm, according to previous studies of the relative diffusion rate and hydration radii of sugars and of partitioning of PEG polymers in the OmpU pore [124, 129]. The protomeric OmpU contains 350 residues including a 31 amino acid signal peptide at the N-terminus. The molecular size of each monomeric subunit of mature OmpU is 35 kDa, which gives 105 kDa for oligomeric trimer form. It was predicted to have 16 transmembrane β -strands, alike in conformation with OmpF [129]. However, a detailed structural model of OmpU underlying its cellular functions has not been available to date, although this protein could be overexpressed and purified to high degree of purity as reported [124, 126, 127, 130, 131]. In the trimer, protomers are presumably held together by hydrophobic interactions between the subunits [124]. Three to four Ca^{2+} ions are associated with each trimeric unit, and these ions are critical for maintaining active state of the trimer, as removal of them upon treatment with EDTA/EGTA was found to irreversibly disrupt the β -sheeted building blocks of the protein [124].

2.12 Inner membrane proteins (IMPs)

The IM of Gram-negative bacteria is a symmetrical phospholipid bilayer in which numerous proteins encoded by about 20% of 4000 open reading frames in *E. coli* genome are located [132]. Fundamentally different in secondary structure elements, IMPs are composed of α -helical membrane spanning stretches rather than antiparallel β -strands found in OMPs. Due to diverse topology of IMPs reflected in number of transmembrane segments and the size and characteristics of both cytoplasmic and periplasmic domains, a common structural feature of IMPs is less clear to be observed and one cannot systematically categorize IMPs into a simple

group of mutual structural similarities. However, the biogenesis of almost all IMPs converges in the same path and the process can be divided into three stages. In the first stage, a hydrophobic target signal of nascent polypeptide is targeted and associated with the signal recognition particle (SRP), a ubiquitous ribonucleoprotein, upon exit from the ribosome [132-134]. The target signal is referred to as the signal anchor sequence, and it is located in the first transmembrane segment. While the N-terminal signal sequences in OMPs are more hydrophilic and cleavable, the signal anchor sequences in IMPs are more hydrophobic and they are non-cleavable segments integrated in the structures. As the determinants of the choice of biogenesis pathway, distinct characteristics of the signal sequences in OMP and IMP nascent chains ensure appropriate dispatch of the precursors in the beginning of biogenesis. The next step entails interactions of the SRP-ribosome-nascent chain complex with its receptor called FtsY and formation of the SRP/FtsY complex that utilizes guanosine triphosphate (GTP) hydrolysis to release and transfer IMPs to the Sec translocon, the same membrane protein complex responsible for OMP translocation. Third, the accommodated hydrophobic segments laterally enters the lipid membrane and fold to become mature and functional conformation [132]. It is worth noting that in addition to the Sec machinery, another accessory membrane protein named YidC [135] is also capable of recruiting and inserting IMPs either on its own or functioning as a chaperone in collaboration with the Sec proteins [136, 137].

2.13 Lipoproteins

On the periplasmic side of either the IM or the OM another class of proteins are anchored through acyl chains post-translationally added to the N-terminal cysteine residues. The added acyl chains are the constituents of bacterial phospholipids before transfer; therefore this class of lipid-fused protein was given the name lipoprotein. Like the assorted three-dimensional conformations of IMPs, lipoproteins are difficult to be structurally sorted in a particular

norm. A collective defining structural feature of all lipoproteins, however, is the N-terminal acyl chains attached to a conserved cysteine (normally three in number in Gram-negative species). In most cases and with the example of BamB-E in the BAM complex, lipoproteins are attached to the membranes by the three acyl chains, which implies that the entire protein is soluble with the exception of these fused lipid moieties. In contrast to the periplasmic space where most lipoproteins are located, recent reports have also provided evidence of surface-exposed lipoproteins positioned on the outer surface of the OM and facing the extracellular environment [[138-140](#)].

2.14 General lipoprotein structural domains

Newly translated lipoproteins, usually termed prelipoproteins, carry specific information with regard to their final location in the membrane boundary which is lied in the most N-terminal domain of ~20 amino acids in length [[141](#)]. This segment forms the signal peptide during localization but differs from exported soluble proteins in that the most C-terminal region of the signal peptide includes a four-amino acid motif referred to as the “lipobox” [[142](#), [143](#)]. It has been much more explicit with the identification of more lipoproteins that the first three residues in the lipobox are of greater diversity than the last well-conserved cysteine residue [[144](#)], the key residue that will be targeted for subsequent modifications and ultimately the first residue at the N-termini with the designated position +1 [[145](#), [146](#)]. Residues following the +1 cysteine are found to be disordered and not classified as any defined secondary structures, indicative of a linker or “tether” domain that connects the lipid anchor to the next properly folded domain (third domain) where the soluble structural entity of lipoproteins is situated [[145](#)]. The length of this amino acid tether varies with each lipoprotein and is between 0 and 170 residues observed in the characterized lipoproteins so far [[147-149](#)]. As illustrated in Figure 11, the residues immediately following the +1 cysteine and upstream of the tether also carry crucial sorting signals for the downstream localization

pathway, which is described below in Figure 11.

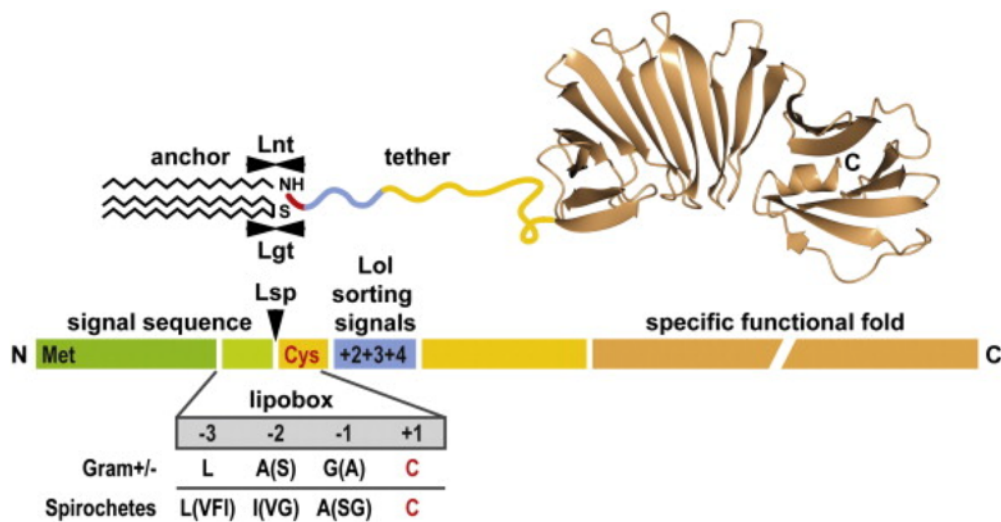


Figure 11. Lipoprotein structural domains.

Translated lipoprotein peptides (prelipoproteins) are initially in the cytoplasm, with an N-terminal signal peptide (green) that is recognized by the Sec translocase. Lgt attaches a diacylglyceryl group to the +1 cysteine (red) to form a thioether-linked S-diacylglyceryl group. Following this, Lsp recognizes the signal peptide that contains the lipobox and cleaves it. This liberates the N-terminal amine group of the cysteine for N-acylation by Lnt. The N-terminal residues following the +1 cysteine contains sorting signals (blue) to be recognized by the Lol machinery, a vital ABC transporter in the IM that determines if mature lipoproteins are sorted to the OM [150]. Downstream of the sorting signal is the disordered tether domain (yellow) and the C-terminal folded domain executing protein-specific functions. The structure of *B. burgdorferi* OspA with PDB accession code 1osp is presented in the C-terminal domain for illustration. (Figure taken from [145])

2.15 Lipoprotein modifications

As a general and universal export and translocation machinery, the Sec translocon also recognizes, accommodates and exports the unfolded prelipoprotein substrates across the IM. Immediately after the translocation in Gram-negative bacteria, these precursors are targeted by three IM enzymes determined by the presence of the lipobox and undergo a three-step sequential reaction for posttranslational modifications. The three IM-embedded enzymes are called phosphatidylglycerol:prolipoprotein diacylglyceryl transferase (Lgt), Prolipoprotein signal peptidase (Lsp) and apolipoprotein N-acyltransferase (Lnt) [151, 152] (Figure 12). Lgt and Lsp are conserved in all bacterial species, whereas Lnt is only present in proteobacteria and actinomycetes.

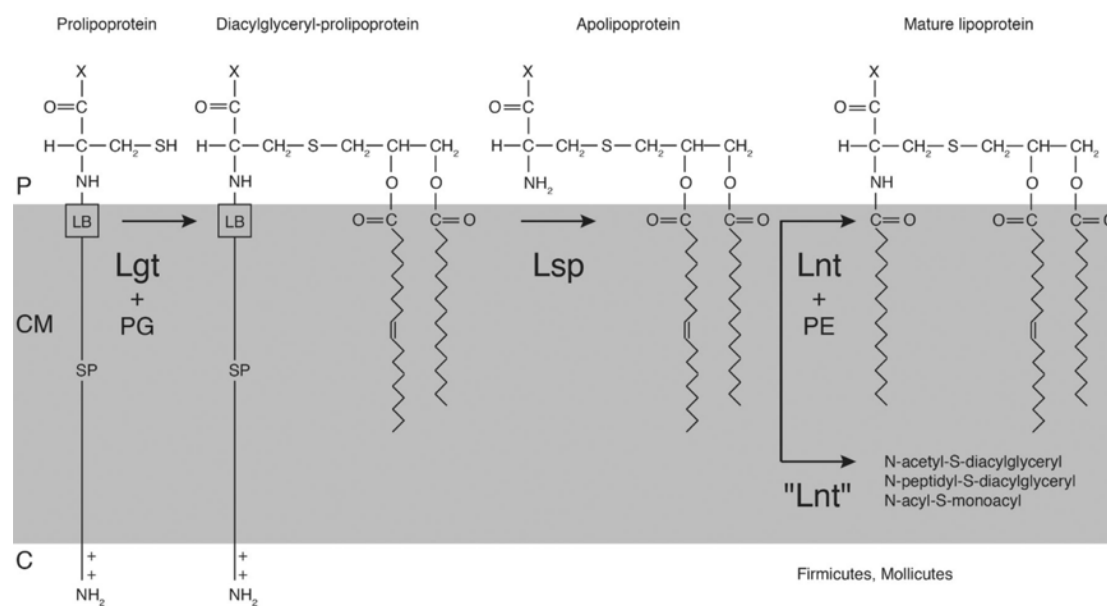


Figure 12. Lipoprotein posttranslational modification in Gram-negative bacteria.

In the IM Lgt, Lsp and Lnt are responsible for diacylglyceryl group transfer, signal peptide cleavage, and N-acylation of the +1 cysteine respectively. Alternative lipoprotein N-acyl modifications by "Lnt" have been reported from firmicutes and mollicutes. C=cytoplasm, P=periplasm, CM=cytoplasmic membrane, LB=lipobox, PE=phosphatidylethanolamine, PG=phosphatidylglycerol, SP=signal peptide. (Figure taken from [153])

2.16 Lgt

Phosphatidylglycerol:prolipoprotein diacylglyceryl transferase (Lgt) is the first enzyme in the pathway responsible for transferring the diacylglyceryl group from phosphatidylglycerol (PG) to the sulfhydryl group of the +1 cysteine via a thioether bond [151] (Figure 12). This IMP was first biochemically characterized in the 1980s and it was reported to be optimally functioning at the pH of 7.8 and temperature of 37°C [154]. Its catalytic activity is specific to PG as an acyl donor because it does not catalyze phosphatidylethanolamine (PE) and cardiolipin (CL) *in vitro*, which suggests that the headgroup of phospholipid is crucial for Lgt recognition [151]. Lgt in *E. coli* was proposed to have five transmembrane helices and a periplasm-exposed C-terminus [155]. However, the issue of how Lgt interacts with PG and lipoproteins remained to be solved until the crystal structure of *E. coli* Lgt was obtained recently [156], in which Lgt possesses a laterally opening central cavity and two PG binding sites and contrasts the former prediction with a novel seven-transmembrane topology. The structure also confirmed the consistency of the hydrophobic central cavity with the specific orientation of PG substrates in which they insert two hydrophobic acyl chains in the central part of the double lipid leaflet and project the more hydrophilic head group towards the periplasmic space. Furthermore, to better understand the molecular basis of enzyme-substrate recognition, it has been suggested that the positively charged residue R143, found in the so-called signature motif of the protein [157], may efficiently attract the negatively charged PG substrate and assist Lgt in specific binding.

2.17 Lsp

Lsp, also known as signal peptidase II, is the next enzyme involved in the triacylation reaction. As the name suggests, its function is to cleave the N-terminal attached signal peptide at the +1 cysteine position [158] (Figure 12). It was predicted to be a small integral

IMP with four TM and both N- and C-termini located in the cytoplasm [159]. Initial biochemical experiments on Lsp were carried out in the early 1980s and demonstrated that it is functionally optimal with a pH of 7.9 and temperature between 37 and 45°C [160]. In proteobacteria, Lsp cannot proceed with signal peptide cleavage without S-diacylglycerol addition, which implies that it specifically recognizes and targets acylated cysteine residues produced by Lgt in the previous step [153]. It was predicted with the assistance of gene fusions of *lsp* to *phoA* and *lacZ* that Lsp in *E. coli* possesses four transmembrane helices with both the N- and C- terminus facing the cytoplasm [159]. Due to a lack of detailed structural information and sequence homology to proteins of known structures, however, the definite substrate binding and catalytic mechanism was unclear until the structural characterization of the enzyme in 2016 from *Pseudomonas aeruginosa* [161]. The structure reveals a transmembrane domain consisting of four TM helices as expected from the topology model, and a second periplasmic domain including a β -cradle subdomain and a periplasmic helix subdomain. The structure is also complexed with a globomycin molecule residing on the periplasmic side of the membrane and indicating the active site of the enzyme containing two catalytic residues, D124 and D143. Consistent with the peptidase activity, the helical signal peptide of the prolipoprotein fits perfectly in the transmembrane domain, and, upon binding to Lsp, the cysteine residue in the lipobox resides further up in the active site, which positions the prolipoprotein in the periplasm and places the scissile bond (between cysteine and the third residue in the lipobox) between the carboxyls of the two catalytic residues for cleavage [161].

2.18 Lnt

In proteobacteria and actinomycetes, diacylated prolipoproteins devoid of signal peptide are now ready for the last catalytic reaction to occur before being converted into mature triacylated lipoproteins (Figure 12). The primary enzyme involved in the catalytic process is

the apolipoprotein N-acyltransferase (Lnt). The first identification of the *lnt* gene was in *S. enterica* by screening the candidates causing accumulation of an abundant lipoprotein called Lpp [162]. It was first identified in the 1990s and defined to function optimally at pH of 6.5-7.5 [163]. It carries out a two-step reaction to transfer an *sn*-1 acyl group from a phospholipid to the amino group of the +1 cysteine via an amide linkage, referred to as ping-pong mechanism [164, 165]. In the first step, the *sn*-1 of PE is attacked by a cysteine residue that is a member of the E-K-C catalytic triad [166], and this forms a thioesteracyl-enzyme intermediate and a lysophospholipid byproduct [165, 167]. In the second step, the produced lysophospholipid is released and the acyl chain is transferred to the subsequently engaged α -amino group of the +1 cysteine of prolipoprotein that was liberated by Lsp previously, resulting in the mature triacylated form [165]. As for the lipid substrate, PE is the favourable substrate with fatty acid chain length of C16-C18, but other phospholipid types that possess smaller head groups can also serve as acyl donors [168]. From a structural perspective, Lnt was predicted to contain a periplasmic domain that is a member of the carbon-nitrogen hydrolase family, and a transmembrane domain consisting of six transmembrane segments [166, 169]. The catalytic triad was proposed to be in the periplasmic domain and, along with other essential residues, constitutes a core reaction site in the enzyme.

2.19 Research aims

Despite the efforts of understanding the roles the BAM complex plays during OMP biogenesis *in vivo*, related information was limited and a detailed insight of how these proteins coordinate recognition, folding and membrane insertion of nascent OMPs has been partially hindered by a lack of structure of the entire complex. To gain structural insights into the overall architecture of the BAM complex and obtain more detailed information regarding the mechanism of OMP assembly, the BAM complex was subject to X-ray crystallographic studies and the crystal structures of BamABCDE and BamACDE were solved successfully.

This was a collaborative project with my colleagues in the group in which I carried out cell culture, purification and crystallization using the different plasmid constructs and obtained high resolution crystallographic diffraction data for structural determination of the BamABCDE complex.

For OmpU project, it aimed at yielding high-purity overexpressed OmpU proteins and well-diffractive crystals for structure determination based on the available structural information in order to gain more comprehensive interpretation of the reported biophysical data and functional claims on OmpU.

For Lnt project, given the recent advance in structural studies of Lgt and LspA, it aimed to carry out structural studies of Lnt in given Gram-negative species using X-ray crystallography in order to establish structural basis for the last step in post-translational modifications of prolipoproteins.

CHAPTER 3

Materials and methods used in structural studies of the BAM complex, OmpU outer membrane protein and lipoprotein N-acyl transferase

3.1 Plasmid construction of *E. coli* BamAB+CDE

The plasmids used for over-expression of *E. coli* BamAB+CDE were a gift from Prof. Daniel Kahne at Harvard University and listed in Table 1. The plasmids had been transformed into *E. coli* BL21 (DE3) prior to the start of the project.

Table 1 Plasmids used in separate expression of BamAB + BamCDE.

Name	Description	Construction
pSK38	pETDuet-bamB- bamA	PCR with primers: bamB-N: ACACCCATGGGACAATTGCGTAAATTACTGCTGC, bamB-C: ACACGCGGCCGCTTAACGTGTAATAGAGTACACGGTTC and bamA- N: GTCCTAGAGCATATGGCGATGAAAAAGTTGC, bamA-C: ACACGACGTCTTACCAGGTTTTACCGATGTAAAC
pSK86	pETDuet-bamB- His ₆ -bamA	pSK38 with inserted His-tag and thrombin site after bamA signal sequence
pSK46	pCDFDuet-bamC- bamD	PCR with primers: bamC-N: ACACCCATGGGAGCTTACTCTGTTCAAAAGTCG, bamC-C: ACACGCGGCCGCTTACTTGCTAAACGCAGC and bamD-N3: ACACCATATGACGCGCATGAAATATCTG, bamD-C3: ACACGACGTCTTATGTATTGCTGCTGTTTGC
pBamE- His	pET22-42-bamE- His ₈	PCR with primers: bamE-N: ATGACATATGCGCTGTAAAACGCTGACTGC and bamE-C: ACGTCTCGAGGTTACCACTCAGCGCAGGTTTGTTATCAATATTG

Adapted from [77]

3.2 Expression of *E. coli* BamAB

E. coli BL21 (DE3) strain carrying pSK86 was cultured in Terrific Broth (TB) [152] media with 100ug/ml ampicillin at 37°C until optical density at 600 nm (OD₆₀₀) reaches ~0.3. The temperature was shifted to 25°C and incubation continued until OD₆₀₀ reached ~0.6. Overexpression of BamA and BamB were induced by addition of IPTG to 0.1 mM final concentration followed by overnight shaking at 20°C for 16-22 h.

3.3 Expression of *E. coli* BamCDE

pSK46 and pBamE-His plasmids were co-transformed into BL21(DE3) strain, which was incubated with 100 ug/ml ampicillin and 50ug/ml streptomycin in LB Broth at 37°C. The Cells were cultured until OD₆₀₀ reached ~0.6 and the two proteins were co-expressed by addition of IPTG to 0.1 mM final concentration for 3-4 hours at 37°C before harvest.

3.4 Purification and reconstitution of the BAM complex from *E. coli* BamAB + *E. coli* BamCDE

E. coli BamAB was detergent-solubilized using the following method.

Cells were suspended in 1xTBS buffer (20 mM Tris-HCL, pH 8.0, 150 mM NaCl) with DNase I (Sigma), lysozyme (Melford) and protease inhibitor tablets (Roche). Cells were lysed by a cell disruptor (Constant Systems Ltd.) at 30 kpsi. The lysate was centrifuged at 5,000 g, 4°C to remove cell debris. The resultant supernatant was ultra-centrifuged at 120,000 g for 1h to pellet whole membranes. Membrane fractions were then solubilized in TBS/1% N-Lauroylsarcosin sodium salt (Sigma) at room temperature for 2h. The sample was Ultra-centrifuged at 120,000 g again for 1h, and the pelleted outer membrane was dissolved in 20 mM Tris-HCL pH 8.0, 300 mM NaCl, 10 mM imidazole with 2% 3-(*N,N*-Dimethylmyristylammonio) propanesulfonate (SB 3-14) (Fluka) at 4°C overnight by gentle stirring.

E. coli BamCDE was detergent-solubilized using the following method.

Cells were lysed and unbroken debris was discarded as previously described. The resultant supernatant was ultra-centrifuged at 120,000 g for 1h to pellet whole membranes. The membrane fractions were then solubilized in 20mM Tris-HCL pH 8.0, 300 mM NaCl, 10 mM imidazole with 2% Triton-X-100 at room temperature for 1h.

The solubilized BamAB subcomplex and BamCDE subcomplex were mixed together and left for 1h to reconstitute the five-protein BAM complex *in vitro*. The mixture was then ultracentrifuged again for 1h before being loaded onto a nickel-nitrilotriacetate affinity resin (Ni-NTA, Qiagen) column. Small scale Ni-NTA affinity columns were initially made prior to large scale purification to test the amount of BamAB and BamCDE solutions added in the mix solution such that equal amounts of BamB and BamC were reflected on SDS-PAGE gel before further size exclusion. The BamABCDE complex sample was loaded onto a Ni-NTA column pre-equilibrated with loading buffer (20 mM Tris-HCL pH 8.0, 300 mM NaCl, 10 mM imidazole, 0.5% SB3-14). The column was washed by a wash buffer (20mM Tris-HCL pH 8.0, 300 mM NaCl, 30 mM imidazole, 0.5% β -OG, 0.023% LDAO) and the BamABCDE complex was eluted with elution buffer (20 mM Tris-HCL, pH 8.0, 300 mM NaCl, 300 mM imidazole, 0.5% β -OG, 0.023% LDAO). The eluted proteins were quantified and loaded onto a HiLoad 16/600 Superdex 200 prep grade column (GE healthcare) pre-equilibrated with gel filtration buffer (20 mM Tris-HCL, pH 8.0, 300 mM NaCl, 0.5% β -OG, 0.023% LDAO) for further purification. Resultant peak fractions were pooled and concentrated. Small volumes of these fractions were run on SDS-PAGE gel to check the complex compositions and the purity of the BamABCDE complex.

3.5 Protein Crystallization

The crystallization was performed using the sitting-drop vapour diffusion method. For the initial screens, purified protein solution was concentrated to a specific concentration (e.g., 10 mg/ml, 15 mg/ml, 20 mg/ml, and 50 mg/ml), and mixed with commercial screening buffers in 1:1 ratio (e.g., 0.15 μ l: 0.15 μ l, 0.3 μ l: 0.3 μ l) as well as dispensed against 65 μ l of reservoir buffer using the Gryphon Crystallization Robot (Art Robbins Instruments). The commercial crystallization screening kits included MemStart, MemSys, MemGold, MemGold2, Memplus, MembFac, MemMeso, and Proplex. Buffer formula that gave a hit was chosen for

subsequent optimizations by varying precipitate concentration and buffer pH. The optimization buffers were either prepared manually with ultrapure water or purchased from Molecular Dimensions. For optimization, protein and buffer solutions were dispensed and mixed in aliquots of 0.5-1 μ l against 70 μ l of reservoir buffer either by hand or by the robot.

3.6 Co-purification and crystallization of *E. coli* BamAB and *E. coli* BamCDE

Cells were cultured and harvested as described above. The two batches of harvested cells were mixed together and suspended in TBS buffer prior to disruption and lysed as described. Whole membrane fractions containing over-expressed BamAB and BamCDE were pelleted and dissolved in TBS/1%-N-Lauroylsarcosin as described. The solution was ultra-centrifuged and the resultant supernatant was stored at 4°C. Resultant pellets were solubilized in 2% SB3-14 and the supernatant from previous ultra-centrifuge was added in the similar manner with an amount that produced equal amounts of BamB and BamC reflected on SDS-PAGE gel. The mixed solution was gently agitated at 4°C for 2 h. Subsequent metal affinity and Size-exclusion chromatography purification were identical to the corresponding methods described previously. Crystallization procedure was identical to that described in 1.2.5. All the crystallization screening plates were stored at 4°C for incubation.

3.7 Cloning and construction of pJH114 encoding BamABCDE and pYG120 encoding BamABCDE and an extra BamB copy

A plasmid named pTRC99a was used to be the original template of pJH114 (Appendix 1). Specifically, a modified version of pTRC99a that lacks the endogenous NdeI site was re-inserted with a NdeI site into the polylinker region using the QuickChange Mutagenesis Kit with primer 5'-CACACAGGAAACAGCATATGGAATTCGAGCTCGG-3' and its complement. The five genes encoding the BAM complex were all amplified by PCR using genomic DNA from *E. coli* AD202 strain as a template. Then BamA gene was first double digested with NdeI and BamHI and ligated into the corresponding plasmid site, BamB gene

was double digested with BamHI and BglII and ligated into the now single digested plasmid with BamHI, and so forth. In the last step an octahistidine tag was added to the C-terminus of BamE.

For construction of pYG120 (created by colleagues) using the Sequence and Ligation Independent Cloning (SLIC) method, pJH114 backbone was amplified by PCR using primers PF_pJH114_SLIC (5'-GTTAATCGACCTGCAGGCATGCAAG-3') and PR_pJH114_SLIC (5'-CTCTAGAGGATCTTAGTGGTGATGATGGTG-3'), and *bamB* gene was amplified using PF_EBB_SLIC (5'-TCATCACCCTAAGATCCTCTAGAGAGGGACCCGATGCAATTGC-3') and PR_EBB_SLIC (5'-CTTGCATGCCTGCAGGTCGATTAACGTGTAATAGAGTACACGGTTCC-3'). Amplified fragments were gel-extracted and digested by T4 DNA polymerase (Fermentas) at 22°C for 35 min and then 70°C for 10 min, before being placed on ice. The digested gene and vector fragments were annealed in 10 mM Tris, pH 8.0, 100 mM NaCl, 1 mM EDTA for 10 min at 75°C and subsequent temperature drop by 0.1°C every 8 s to 20°C. The annealed product was transformed into *E. coli* DH5 α cells for plasmid extraction.

3.8 Expression of BamACDE (pJH114) (performed by colleagues) and BamABCDE (pYG120)

Expression plasmid pJH114 was initially transformed into *E. coli* HDB150 cells [170] (MC4100 *ompT::spc Δ araBAD leuD::kan*) for overexpression. When OD₆₀₀ reached ~0.6-0.8, IPTG was added to final concentration of 0.1 mM, and the cultures continued overnight at 20°C. Overexpression of the selenomethionine-labelled BAM complex was performed in M9 medium with selenomethionine Medium Nutrient Mix (Molecular Dimensions) and 100 mg/L L-(+)-selenomethionine (Molecular Dimensions) according to the preparatory protocol given in Appendix 2. Specifically, O/N culture was washed by PBS solution and distributed into each 1L minimal media. When OD₆₀₀ reached ~0.6, amino acid supplement was added

and culture continued for 30 min. L-(+)-selenomethionine was then added and culture was grown for another 30 min before IPTG induction. Overnight post-induction culture conditions were identical to native protein induction conditions. Overexpression of the BAM complex from pYG120 followed the identical protocol.

3.9 Purification of BamACDE (performed by colleagues) and BamABCDE

For purification of pJH114-generated BamACDE, the cells were pelleted and resuspended in TBS buffer and lysed by passing through a cell disruptor (Constant Systems) at 30 kpsi. The lysate was centrifuged at 18,000 g to separate cell debris and unbroken cells, and the supernatant was ultracentrifuged to pellet the membranes at 100,000 g for 1 h. The cell membranes were resuspended in solubilization buffer containing 20 mM Tris-HCl, pH 8.0, 300 mM NaCl, 10 mM imidazole and 1% DDM and gently agitated for 1h at room temperature. The solution was ultracentrifuged again and the supernatant was loaded onto a 5-ml HisTrap HP column (GE Healthcare) pre-equilibrated with solubilization buffer. The column was washed with wash buffer containing 20 mM Tris-HCl, pH 8.0, 300 mM NaCl and 45 mM imidazole and eluted with elution buffer containing 300 mM imidazole. The eluate was loaded onto a HiLoad 16/600 Superdex 200 prep grade column (GE healthcare) pre-equilibrated with gel filtration buffer containing 20 mM Tris-HCl, pH 7.8, 300 mM NaCl and 2 CMC *N*-nonyl- β -D-glucoside (β -NG) and 1 CMC tetraethylene glycol monoethyl ether (C₈E₄). Purification of the intact BAM complex from pYG120 followed the identical protocol with the exception that wash, elution and gel filtration buffer were supplemented with 1 CMC *N*-octyl- β -D-glucopyranoside (OG) and 1 CMC *N*-dodecyl-*N,N*-dimethylamine-*N*-oxide (LDAO). The eluted peak fractions were pooled and concentrated.

3.10 Crystallization, data collection and structure determination of BamACDE (performed by colleagues) and BamABCDE

The purified proteins were concentrated to 8–12 mg/ml for crystallization. For NaI co-crystallization for the BAM complex from pYG120 (performed by colleagues), NaCl was replaced by 0.2 M NaI in the gel filtration buffer. All crystallization trials were performed using sitting-drop vapour diffusion method and the screening plates were stored at 22°C. Optimization of the crystals were conducted by varying the concentrations of the precipitate and buffer pH as well as adding commercial additives (Molecular Dimensions). The best NaI co-crystallized BamABCDE crystals appeared from 150 mM HEPES, pH 7.5, 30% PEG6000 and CYMAL-4 in MemAdvantage additive screen (Molecular Dimensions) as an additive. The best BamABCDE native crystals appeared from 150 mM HEPES, pH 7.5 and 27.5% PEG6000. The best BamACDE native and selenomethionine-labelled crystals emerged from 100 mM Tris, pH 8.0, 200 mM MgCl₂ · 6H₂O, 24% PEG1000 MME and OGNG (Octyl Glucose Neopentyl Glycol) in MemAdvantage as an additive. The crystals were harvested, cryoprotected by supplementing the crystallization solution with 20% glycerol and stored in liquid nitrogen for data collection.

For BamACDE (performed by colleagues), both native and selenomethionine-substituted data sets were collected on the I03 beamline at Diamond Light Source (DLS) at a wavelength of 0.9173 Å and 0.9795 Å respectively. The crystals belong to $P4_22_12$ space group and cell dimensions of $a = b = c = 254.16$ Å, $c = 179.22$, $\alpha = \beta = \gamma = 90^\circ$. Data sets were indexed, integrated and scaled by XDS [171]. Structure determination was carried out using ShelxD [19, 172] and a diffraction limit of 3.9 Å was determined. Fifty-six selenium sites were found and gave a figure of merit (FOM) of 0.32. BamACDE could be fit into the electron density map following density modification using DM [173]. The BamACDE complex was built using Coot [174] by skeletonizing the electron density map and docking the BAM subunits in

the electron density map with selenomethionine sites used as guides. Rigid body refinement was subsequently performed and NCS refinement was used with TLS refinement against groups automatically determined using PHENIX [175]. Restrained refinement was performed with group B-factors and reference model secondary structure restraints from higher resolution models. Weights were automatically optimised by PHENIX [175].

For BamABCDE expressed from pYG120, the data sets of the BAM complex were collected on the I03 beamline at DLS. I collected and processed data sets of native crystals. Data sets of NaI co-crystallized crystals were collected and processed by my colleagues. The crystals belong to space group $P4_12_12$, with the cell dimensions $a = b = 116.88 \text{ \AA}$, $c = 435.73 \text{ \AA}$, $\alpha = \beta = \gamma = 90^\circ$. There is one complex in the asymmetric unit. Despite molecular replacement attempts with individual component structural models, a resolution of 2.92 \AA was not sufficient for structure determination. BamABCDE complex was crystallized in presence of 0.2 M sodium iodide, and SAD data sets were collected at a wavelength of 1.8233 \AA (performed by colleagues). Four 360° data sets were collected on the same crystal and then combined. The phases were determined by ShelxD [19, 172] at 4 \AA resolution. Eleven iodide sites were assigned and gave a FOM of 0.28 . The phases were refined to 2.9 \AA by DM [173], and the molecular model was built using Coot [174] by skeletonizing the electron density map and docking the individual high-resolution subunits in the electron density map. Rigid body refinement was used to fit this model into the higher resolution native data set while retaining and extending the free R set from the iodide data set. The BamABCDE complex was then refined using PHENIX [175]. TLS groups were automatically determined using PHENIX [175] and restrained refinement was performed with secondary structure restraints and individual B-factors. Weights were automatically optimised. Data collection and structure refinement statistics for selenomethionine-substituted BamACDE complex, native

BamABCDE complex, and sodium iodide co-crystallized BamABCDE complex are listed in Table 2.

Table 2. Data collection and refinement statistics for BamACDE and BamABCDE complex

	BamACDE (created by colleagues)	Se-Met ^{‡a} BamABCDE (created by colleagues)	NaI ^a BamABCDE Native ^{‡a}
Data collection			
Space group	<i>P4</i> ₂ <i>2</i> ₁ <i>2</i>	<i>P4</i> ₁ <i>2</i> ₁ <i>2</i>	<i>P4</i> ₁ <i>2</i> ₁ <i>2</i>
Cell dimensions			
<i>a, b, c</i> (Å)	254.16, 254.16, 179.22	116.72, 116.72, 432.44	116.88, 116.88, 435.73
(°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Wavelength (Å)	0.97951	1.82330	0.97623
Resolution (Å)	29.94–3.90 (4.02– 3.90)*	29.86–4.00 (4.27– 4.00)	103.06–2.92 (3.02– 2.92)
<i>R</i> _{pim} (%)	3.3 (53.8)	10.5 (69.9)	4.7 (65.4)
<i>CC1/2</i> (%)	99.9 (49.4)	100 (99.6)	99.6 (64.8)
<i>I</i> / sigma	11.0 (0.9)	37.0 (11.8)	11.2 (1.0)
Completeness (%)	99.8 (100.0)	98.5 (97.8)	100 (100)
Redundancy	27.1 (27.2)	158.00 (165.1)	12.6 (11.5)
Refinement			
Resolution (Å)	29.92 – 3.90		103.06–2.92
No. reflections	73745		66804
<i>R</i> _{factor} / <i>R</i> _{free}	30.44/31.93		28.13/30.73
No. atoms			
Protein	19796		22815
Ligand/ion	0		0
Water	0		0
<i>B</i> -factors(Å ²)			
Protein	150		118
Ligand/ion	N/A		N/A
Water	N/A		N/A
R.m.s. deviations			
Bond lengths (Å)	0.010		0.003
Bond angles (°)	1.868		0.87
Residues in			
Ramachandran plot			
Favored (%)	90.5		93
Allowed (%)	8.7		6.3
Outliers (%)	0.8		0.7
PDB code	5D0Q		N/A

*Values in parentheses are for highest-resolution shell.

[‡]Highest resolution shell was taken as point where $CC1/2 > 30$ along strongest reciprocal lattice direction.

^aData statistics shown for each wavelength are a combination of two datasets (BamACDE Se-Met) and four datasets (BamABCDE NaI).

^b $R_{\text{factor}} = \sum ||F_o| - |F_c|| / \sum |F_o|$, where F_o and F_c are observed and calculated as structure factors, respectively.

^c R_{free} is calculated using 5% of total reflections, which is randomly selected as a free group and not used in refinement.

3.11 Plasmid construction of OmpU

The gene sequence encoding full-length *ompU* gene (gene ID:2615421) was amplified by PCR using genomic DNA from *Vibrio cholerae* O1 biovar El Tor str. N16961 as a template and forward and reverse primers of atcg **ccatgg**acaataaattaggacttaataagatgaa, and gctact**cgagga**agtcgtaacgtagaccgata, respectively. The PCR product was confirmed on 1% agarose gel electrophoresis. The *ompU* PCR product was recovered from the 1% agarose gel. Both pET28 plasmid and the PCR product were digested by restriction enzymes NcoI and XhoI, and the products were recovered separately. The *ompU* gene was ligated into the pET28 plasmid, and the inserted *ompU* gene was confirmed by sequencing. The generated plasmid contained a hexahistidine tag at the C terminus of OmpU. It was transformed into *E. coli* C43 (DE3) cells for overexpression.

3.12 Expression of *V. cholerae* OmpU

Transformed *E. coli* C43 (DE3) strain was grown in LB Broth media with 50 µg/ml kanamycin at 37°C until OD_{600} reached about 0.6. IPTG (final concentration 0.1 mM) was then added to induce overexpression of OmpU for 7-8 hours before harvesting the cells.

3.13 Purification of *V. cholerae* OmpU

The cells were fully suspended in TBS solution and the cells were disrupted as previously described. The cell lysate was centrifuged at 5,000 g at 4°C for 15 min to remove cell debris. The resultant supernatant was ultra-centrifuged at 120,000 g for 1 h to pellet whole membranes. Membrane fractions were then solubilized in TBS/1%-N-Lauroylsarcosin (sodium salt) at 4°C for 1 h. Ultra-centrifugation was performed at 120,000 g again for 1h and the outer membrane was pelleted. The outer membrane proteins were dissolved in 20 mM Tris-HCL pH 8.0, 300 mM NaCl, and 10 mM imidazole with 1% LDAO at 4°C for 1.5 h and ultra-centrifuged again for 30 min before being loaded onto a Ni-NTA gravity column pre-equilibrated with 20 mM Tris-HCL pH 8.0, 300 mM NaCl, and 10 mM imidazole and 1% LDAO. The column was washed with 20 mM Tris-HCL, pH 8.0, 300 mM NaCl, 30 mM imidazole, 0.5% C₈E₄, 5 mM CaCl₂ and eluted with the same buffer but containing 300 mM imidazole. The eluted proteins were quantified and applied to a HiLoad 16/600 Superdex 200 prep grade column (GE healthcare) pre-equilibrated with gel filtration buffer containing 20 mM Tris-HCl, pH 7.8, 300 mM NaCl, 0.5% C₈E₄, 5 mM CaCl₂. Resultant peak fractions were pooled and concentrated. Small volumes of these fractions were run on SDS-PAGE gel to check the purity of the protein.

3.14 Protein crystallization and data collection

Purified proteins were concentrated to ~10mg/ml and used to set up crystallization trials using the sitting-drop vapour method. The best crystals emerged in a condition of 0.1 M lithium sulfate monohydrate, 0.1 M sodium acetate trihydrate pH 4.6, and 1 M ammonium phosphate monobasic at 16°C after about one week, and they were protected in a cryo-protectant (0.1 M lithium sulfate monohydrate, 0.1 M sodium acetate trihydrate pH 4.6, 1 M ammonium phosphate monobasic, and 20% glycerol), harvested and flash-cooled in liquid nitrogen for data collection. All data sets were collected using a wavelength of 0.9173 Å

(probably used in previous data collection and not adjusted to the default wavelength for native data collection) at Beamline I04-1 at Diamond light resources. 3600 images were recorded for each dataset. The data sets were processed using iMOSFLM[9], and the space group was determined by Pointless[11]. The crystals belong to space group $P2_12_12$ with three protomers in the asymmetric unit and cell dimensions $a = 129.88 \text{ \AA}$, $b = 153.47 \text{ \AA}$, $c = 81.01 \text{ \AA}$, and $\alpha = \beta = \gamma = 90^\circ$. The data was further integrated and scaled by SCALA [176]. The data collection statistics are listed in Table 3.

3.15 Structure determination

The structure was solved by molecular replacement using the structure model of OmpK36 from *K. pneumoniae* (PDB code 1OSM) as the search template (Table 4) and the Phaser program in CCP4 suite [177]. Both R_{factor} and R_{free} were above 0.5 after molecular replacement, but dropped significantly following iterative rigid body refinement. The model was re-built manually in Coot [174]. The refinements were performed using REFMAC 5 [178]. The water molecules were added in the structure automatically using ARP/wARP [179]. The detergent LDAO and C_8E_4 molecules and glycerol molecules were built in the structure manually in Coot. The structure refinement statistics are listed in Table 3.

Table 3. Data collection and refinement statistics for OmpU

Native OmpU	
Data collection	
Space group	<i>P</i> 2 ₁ 2 ₁ 2
Cell dimensions	
<i>a</i> , <i>b</i> , <i>c</i> (Å)	129.88, 153.47, 81.01
α , β , γ (°)	90.0, 90.0, 90.0
Wavelength (Å)	0.91732
Resolution (Å)	64.95–2.22 (2.28 –2.22) ^a
<i>R</i> _{pim} (%)	4.9 (55.6) ^a
<i>CC</i> _{1/2} (%)	100 (60) ^a
<i>I</i> / σ (<i>I</i>)	14.7 (1.7) ^a
Completeness (%)	99.8 (98.8) ^a
Redundancy	8.3 (6.5) ^a
Refinement	
Resolution (Å)	62.73 – 2.22
No. reflections	80495
<i>R</i> _{factor} / <i>R</i> _{free} (%) ^b	20.78/23.57
No. atoms	
Protein	7344
Detergent/glycerol	259
Water	637
Mean <i>B</i> value (Å ²)	
Protein	42.5
Detergent/glycerol	99.4
Solvent	51.7
R.m.s. deviations	
Bond lengths (Å)	0.004
Bond angles (°)	1.000
Ramachandran statistics	
Favoured (%)	93
Outliers (%)	0.84
PDB accession number	5ONU

^aThe values in parentheses correspond to the highest resolution shell.

^b $R_{factor} = \frac{\sum ||F_{obs}| - |F_{cal}||}{\sum |F_{obs}|}$, where F_{obs} and F_{cal} are observed and calculated structure factors, respectively. R_{free} is calculated using 5% of total reflections, which is randomly selected and not used in refinement.

Table 4. Choices of phasing models for molecular replacement.

Templates (Protein Data Bank ID)	Alignment coverage (%)	Confidence	I.D. (%)
2fgq	85	100	20
1osm	88	100	24
2zfg	88	100	23
3nsg	87	100	23
4d65	88	100	25

Confidence represents the possibility of true-homology relationship between query sequence and template (from 0-100). It is not an indicator of the expected accuracy of the model. A match with confidence >90% implies that the predicted model adopts the overall fold shown and that the core of the protein is modeled at high accuracy (2-4Å rmsd from native, true structure) although surface loops are likely to deviate from the native structure. I.D. is the percentage of identity between query sequence and the template. An I.D. of above 30-40% indicates high accuracy model but low I.D. (<15%) can also be considered acceptable as long as the confidence is high.

3.16 Generation of *Lint* expression plasmid

Full-length *lint* gene sequences from various Gram-negative species were amplified by PCR with designated NdeI and XhoI restriction sites. For the thermophilic strains, NdeI and SalI

were chosen as the restriction sites because XhoI site was found in gene sequences. The primers used are listed in Table 5. The amplified sequences were double digested and ligated into digested pET22 vectors treated with the same restriction enzymes. The generated plasmids contained an octa-histidine tag at the C terminus of Lnt. The plasmids were transformed into Top10 chemically competent cells and subsequently extracted before being sent for sequencing. Confirmed constructs were transformed into *E. coli* C43 (DE3) cells for overexpression.

Table 5. PCR primers used for *Int* gene cloning in individual selected species.

Species name	Primer sequences
<i>E. coli</i>	forward: tata catatg gctttgcctcattaattgaacgcc reverse: tata ctcgagt taatgatgatgatgatgatgatggttacgctgacgcagactc
<i>P. aeruginosa</i>	forward: tata catatg cggttgattctcgtcccggc reverse: tata ctcgagt caatgatgatgatgatgatgatggccgaacaggcgtcgtcc
<i>N. gonorrhoeae</i>	forward: tata catatg tttccaaactggacaaatactgg reverse: tata ctcgagt caatgatgatgatgatgatgatggtgttctttgtttcgggaagatg
<i>N. meningitidis</i>	forward: tata catatg ttcagacggtatcttccgaacag reverse: tata ctcgagt caatgatgatgatgatgatgatggtgttctttgtttcgggaagatg
<i>K. pneumoniae</i>	forward: tata catatg gtattgcctctcttgaacgccagc reverse: tata ctcgagt taatgatgatgatgatgatgatgctcggcgacgctggcgaggct
<i>V. cholerae</i>	forward: tata catatg aacagcgtattatctcatcgccaatgc reverse: tata ctcgagt taatgatgatgatgatgatgatgtctagcccggcgctggcgcca
<i>C. freundii</i>	forward: tatata catatg gcattgcctcactaattgaacgccagc reverse: tata ctcgagt taatgatgatgatgatgatgatggttacgacgctggcgaggctc
<i>Y. enterocolitica</i>	forward: tatata catatg cctatcgcttcatacctcaac reverse: tata ctcgagt caatgatgatgatgatgatgatgagttacgggggtattttatca
<i>H. influenzae</i>	forward: tatata catatg aataaatatttacttatttgcgattatc reverse: tata ctcgagt taatgatgatgatgatgatgatgtaagatattcatcttacgacga
<i>T. thermophilus</i>	forward: tatata catatg aagggtccgagaggaccccaaggcgcagtc reverse: tata gtcgac ctaataatgatgatgatgatgatgatgcccgggtccgccaccccgcgaggc
<i>T. scotoductus</i>	forward: tatata catatg cgcccttctcctgggctcctcctggccctt reverse: tata gtcgac ctaataatgatgatgatgatgatgatgcccgggtccgccaccccgcgaggc
<i>T. maritima</i>	forward: tatata catatg aacagaagacgggtgggatcc reverse: tata gtcgac tcaatgatgatgatgatgatgatgtaatcgaattccgataccttcattcctttcac

3.17 Expression of Lnt

For small-scale expression tests, transformed *E. coli* C43 (DE3) strains carrying recombinant *lnt* plasmids from different species were grown in 1L LB Broth media with 100 µg/ml ampicillin at 37°C until OD₆₀₀ reached about 0.6. IPTG (final concentration 0.2mM) was then added to induce overexpression of Lnt for 5-6 hours before harvesting the cells. For large-scale expression, 12L LB was used to culture cells.

3.18 Purification of Lnt

For expression tests, harvested cells were suspended in TBS solution and the cells were disrupted as previously described. The cell lysate was centrifuged at 18,000 g at 4°C to remove cell debris. The resultant supernatant was ultra-centrifuged at 120,000 g for 1 h to pellet whole membranes. Membrane fractions were then solubilized in 20 mM Tris-HCL pH 8.0, 300 mM NaCl, and 10 mM imidazole with 1% DDM at room temperature for 1 h. Ultra-centrifugation was performed at 120,000 g again for 30 min and the supernatant of each species was loaded onto a 1 ml HisTrap FF column. 10 ml balance buffer was used to pre-equilibrate prior to sample loading, then wash buffer (20 mM Tris-HCL pH 8.0, 300 mM NaCl, 30 mM Imidazole, 0.05% DDM) and elution buffer (20 mM Tris-HCL, pH 8.0, 300 mM NaCl, 300 mM Imidazole, 0.05%DDM) were used to wash and elute the column respectively. Fractions of eluted proteins from each column were taken to run SDS-PAGE gels to check the presence of Lnt.

For large-scale purification, Lnt was bound onto a 5 ml HisTrap HP column. Eluted protein was loaded onto a HiLoad 16/600 Superdex 200 prep grade column (GE healthcare) pre-equilibrated with gel filtration buffer containing 20 mM Tris-HCl, pH 7.8, 300 mM NaCl and detergents. DDM, LMNG, LysoFos Choline 12, Fos Choline 12 were used in purification.

3.19 96 detergent screening

The Analytic Selector Kit from anatrace was purchased, it included a detergent screening plate containing 150µl of 94 detergents at 2x working concentration, a blank well and a well for the control detergent currently being used to stabilize Lnt (DDM). The 96 conditions in the plate are listed in appendix 4. A 0.22µm filter plate with receptacle plate for detergent exchange and a receptacle plate for collecting eluted proteins and exchanged detergents were also included in the kit. A large-scale purification was performed as described above using DDM, the final gel filtration buffer contained 20 mM Tris-HCl, pH 7.8, 150 mM NaCl and 0.1% DDM. Peak fractions were pooled and protein concentration was determined. 2.4 ml of 50% Ni-affinity superflow resin (Qiagen) was then added in a 15 ml tube and equilibrated with the gel filtration buffer. A total of 500 µg of purified Lnt was then added to the resin and the total volume was increased to 6 ml using the gel filtration buffer. The tube was put onto a rotator and stored at 4°C overnight. The next day the detergent plates were prepared by mixing 94 detergents at 2x working concentration with 2x washing buffer (20 mM Tris-HCL pH 8.0, 300 mM NaCl, 50 mM Imidazole) and 2x elution buffer (20 mM Tris-HCL, pH 8.0, 300 mM NaCl, 500 mM Imidazole) respectively to create the washing plate (220 µl solution in each well) and the elution plate (80 µl solution in each well). 50 µl of protein bound resin was then added to each well of the 0.2 µm filter plate and pre-washed using the gel filtration buffer in each well. The plate was then centrifuged at 2000 g for 2 minutes to discard the solution. 30 µl of the washing buffer from the washing plate was then added to each well to perform detergent exchange. Following 5 minutes of incubation at room temperature the plate was centrifuged at 2000 g for 2 minutes and the flow-through was discarded. Detergent exchange was repeated 6 times. Next, 70 µl of the elution buffer from the elution plate was added to each well and the plate was centrifuged at 2000 g for 2 minutes. Flow-through was

collected using a new elution receptacle. Each of the 96 elution solutions was then subjected to SDS-PAGE analysis.

3.20 Crystallization and data collection of *V. cholerae* Lnt

Protein was subjected to crystallization trials using the sitting-drop vapour method. For the initial screening, MemStart, MemSys, MemGold, MemGold2, MemMeso crystallization screens were used, and the 96-well plates were stored at 16°C. Crystals emerged from a condition consisting of 0.02 M sodium citrate tribasic dihydrate, 0.08 M sodium phosphate (pH 6.2), 18% w/v PEG2000 within one week. They were flash cooled and stored in liquid nitrogen for data collection. The data sets were collected on the I03 beamline at a wavelength of 0.9763 Å at DLS. The processing XIA2 3dii program defined the resolution to ~9.2 Å with the space group of $C222_1$. The unit cell dimensions are $a = 182.71$ Å, $b = 255.66$ Å, $c = 72.47$ Å, $\alpha = \beta = \gamma = 90^\circ$.

CHAPTER 4

**Results chapter: structural studies of the five-protein complex of
beta-barrel assembly machinery (BAM complex) from
*Escherichia coli***

4.1 *E. coli* BamAB+BamCDE separate overexpression and subsequent reconstitution

As expected, *in vitro* reconstruction of the BAM complex by the solubilized subcomplexes BamAB and BamCDE according to the published protocol [77] was confirmed to be successful as reported. Following nickel-affinity chromatography, the eluted fractions were collected to be further purified by size-exclusion chromatography (Figure 13). The elution profile exhibits three observable peaks and the samples were eluted from around 65ml to 80ml with the peak value of over 250 mAU in the central and most symmetrical peak and significantly lower values in the leading and lagging peaks of irregular shapes. The yield was ~1.5 mg per litre of cell culture. The contents of the eluted fractions were subject to SDS-PAGE analysis and it was found that in the leading peak and the central peak the BAM complex exists in its intact form, while in the lagging peak mainly BamE and minor amounts of BamD exist (Figure 13, Figure 14). The void volume of the gel filtration column is ~43 ml [180], so the molecular weights of the BAM complexes in the leading peak are definitely larger than that of a single BAM complex and indicate higher-order oligomers. The molecules in the lagging peak may be of various oligomeric states including BamE and BamDE subcomplex. Due to the presence of the polyhistidine-tag on BamE, excessive amount of BamE may bind to and be eluted from the affinity column along with those in the complex or subcomplex. It was therefore decided to pool the fractions in the central peak and proceed with subsequent crystallization trials.

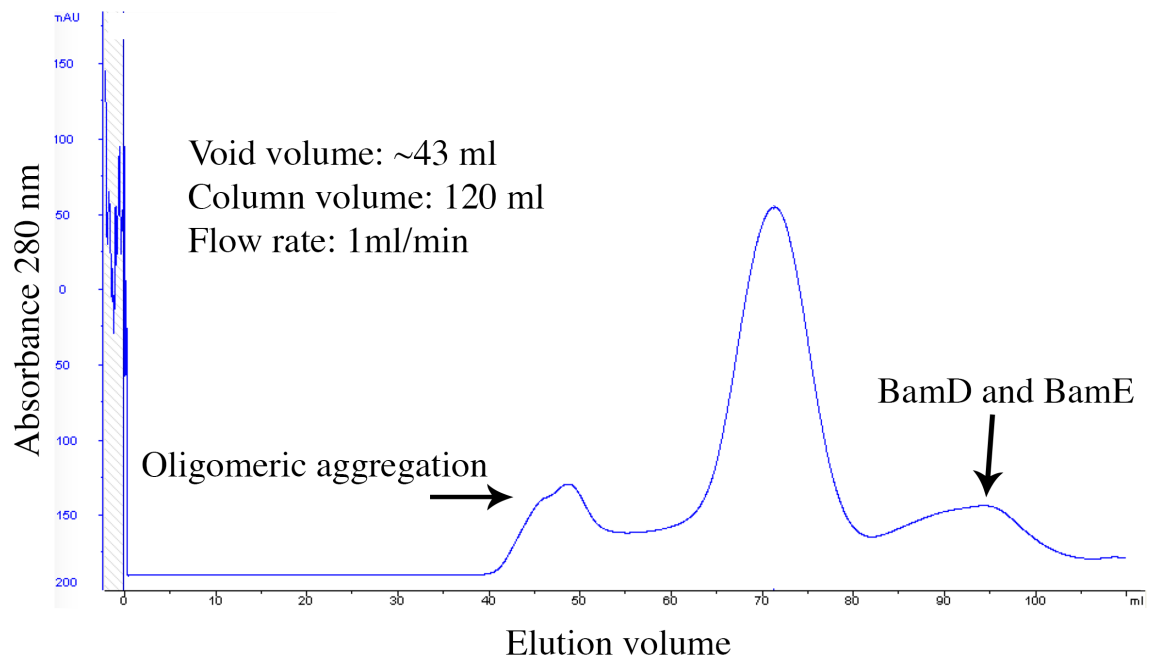


Figure 13. Chromatogram of E.coli BamAB+CDE on gel filtration.

The samples were injected onto a HiLoad 16/600 Superdex 200 prep grade column (GE healthcare) pre-equilibrated with 20mM Tris-HCL (pH 7.8), 300mM NaCl, 1 CMC β -OG, and 1 CMC LDAO. The contents in the leading and lagging peak are individually denoted.

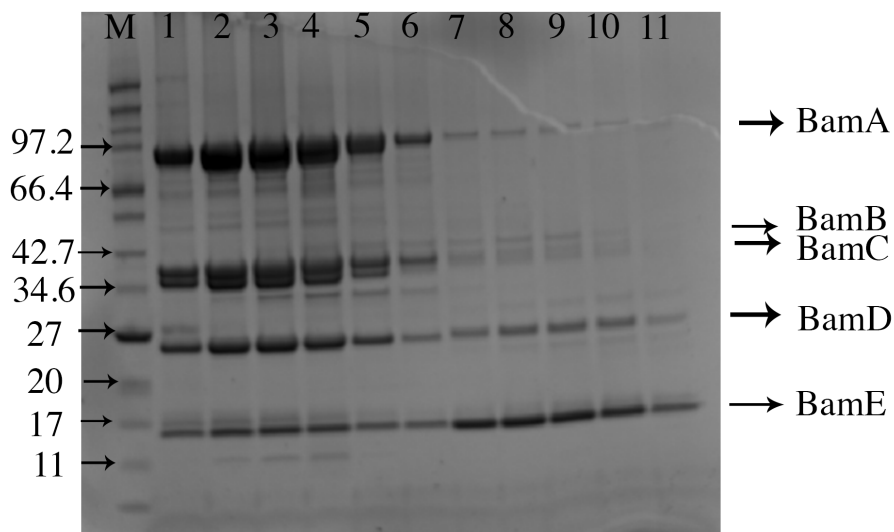
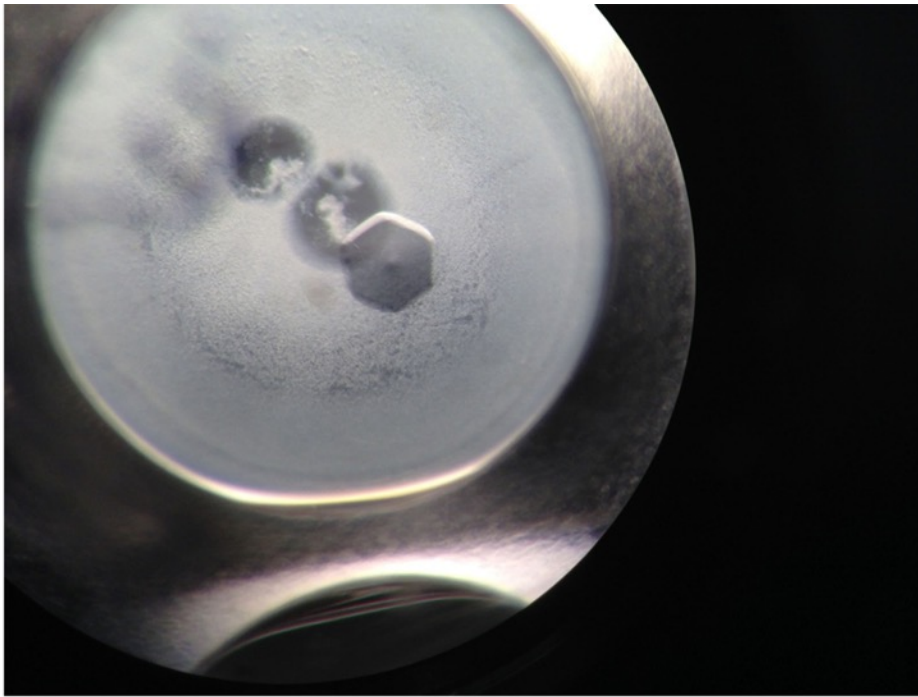
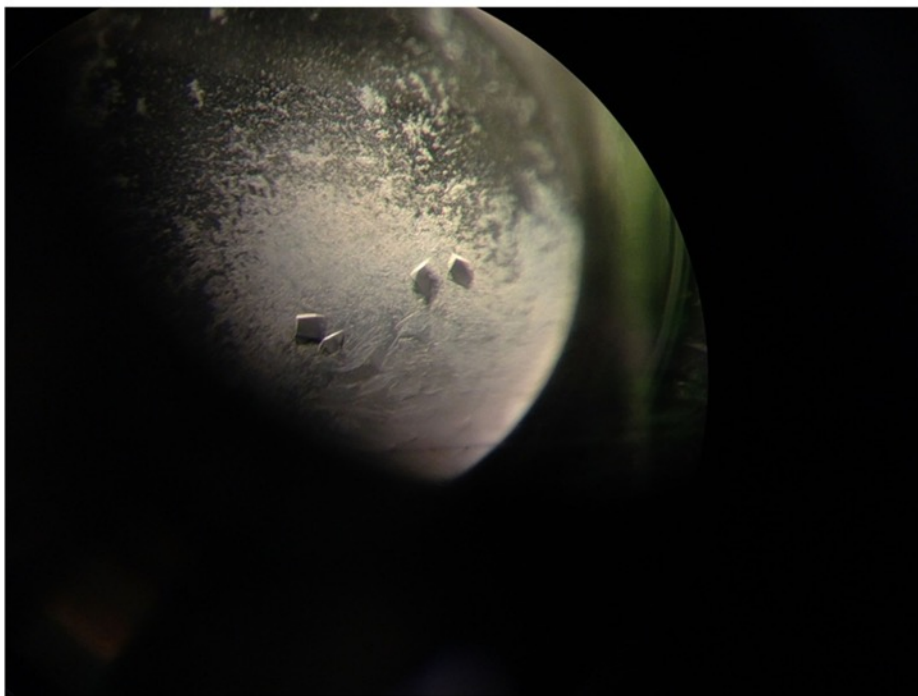


Figure 14. SDS-PAGE gel picture of the purified *E. coli* BamAB+CDE complex.

Protein samples were taken from peak fractions of the peaks in Figure 13. All the five subunit proteins were present in the complex in reasonable stoichiometric amounts. Lane M, protein marker with indicated molecular weights in kDa. Lane 1, fraction from the oligomeric peak. Lane 2-6, fractions from the central peak. Lane 7-11, fractions from the lagging peak.



(a)



(b)

Figure 15. A picture showing the crystals that appeared from two conditions from separate purification of *E. coli* BamAB and *E. coli* BamCDE.

Upon incubation of the screening plates, various screening conditions produced crystals with different types and shapes (Figure 15), and they were cryo-protected before being flash-cooled in liquid nitrogen and sent to DLS for X-ray crystallography analysis. During one visit, there was one crystal with well-diffractive quality and an estimated 4 Å data set was collected (Figure 16).

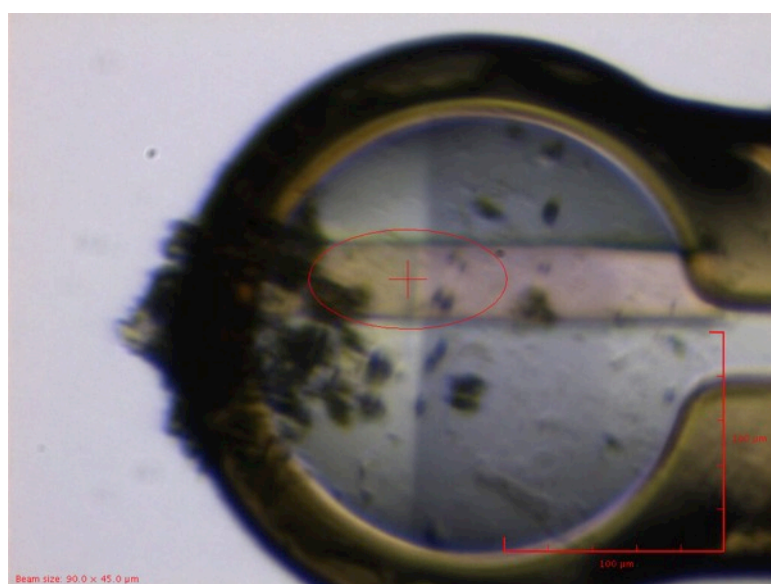


Figure 16. A picture of the well-diffracting crystal of *E. coli* BamAB+CDE.

The crystal was mounted in the rounded cryo-protective loop, the bar-shaped pink crystal is about 200 μm in length and 50 μm in width.

Downstream data processing and structure determination were attempted using individual structures of each component via molecular replacement, but not successful despite extensive and thorough computational analysis using structure-solving programmes. It was later found that this crystal formed from an irrelevant membrane protein that was persistent throughout purification and stayed in the final purified protein solution, because a similar crystal and data set were obtained from another research project aimed at a different outer membrane

protein. Indeed, it can be easily seen that there are some additional bands other than the expected five bands on the SDS-PAGE gel (Figure 14), indicating impurities in the protein solution to be crystallized, therefore it is not surprising that these non-targeted proteins may grow to form crystals in certain conditions, even though their expression levels are significantly lower than the target complex. There were many other BAM crystals with relatively poor diffraction, and also some that did not diffract at all.

The problem of non-diffracting and poorly diffracting crystals is difficult to solve. Despite the formation of crystals from many different conditions indicating that the complexes do interact with each other in a crystal-packing fashion, a large amount of disorder in the crystal lattice is also present, which can result from degradation or dissociation of the protein complex and, not hard to imagine, impurities in the final protein solutions [181].

4.2 *E. coli* BamAB+CDE co-purification

Examining the procedure of the BAM complex reconstitution and purification, a major drawback affecting the formation of the complex and the homogeneity as well as quality of crystals could be the separate extraction steps of BamAB and BamCDE before mixing the two subunits together, allowing only 1h or so for complex formation. The association of the subunits and formation of the complex, therefore, could be much less efficient than they are in the native states intracellularly. In order to improve this less optimal reconstitution and association condition, harvested cells of *E. coli* BamAB and *E. coli* BamCDE were mixed and suspended prior to cell disruption and were then disrupted together in order to promote more prolonged interactions between BamAB and BamCDE by providing more time for interactions from as early as the cell lysis step. The final purified complex solution was of high purity according to SDS-PAGE analysis (Figure 18). The size-exclusion chromatography pattern was improved in that the leading peak shown in separate purification

was eliminated and the magnitude of the lagging peak was reduced, indicative of less heterogeneous compositions and oligomeric aggregations (Figure 17). The yield was ~1.5 mg per litre of cell culture. The fractions in the main peak were pooled and taken for crystallization trials. All the screening plates were stored at 4°C for incubation, and there were a number of wells that produced BAM crystals after about two weeks (Figure 19). Upon X-ray analysis, however, these crystals only diffracted weakly and no further information could be obtained.

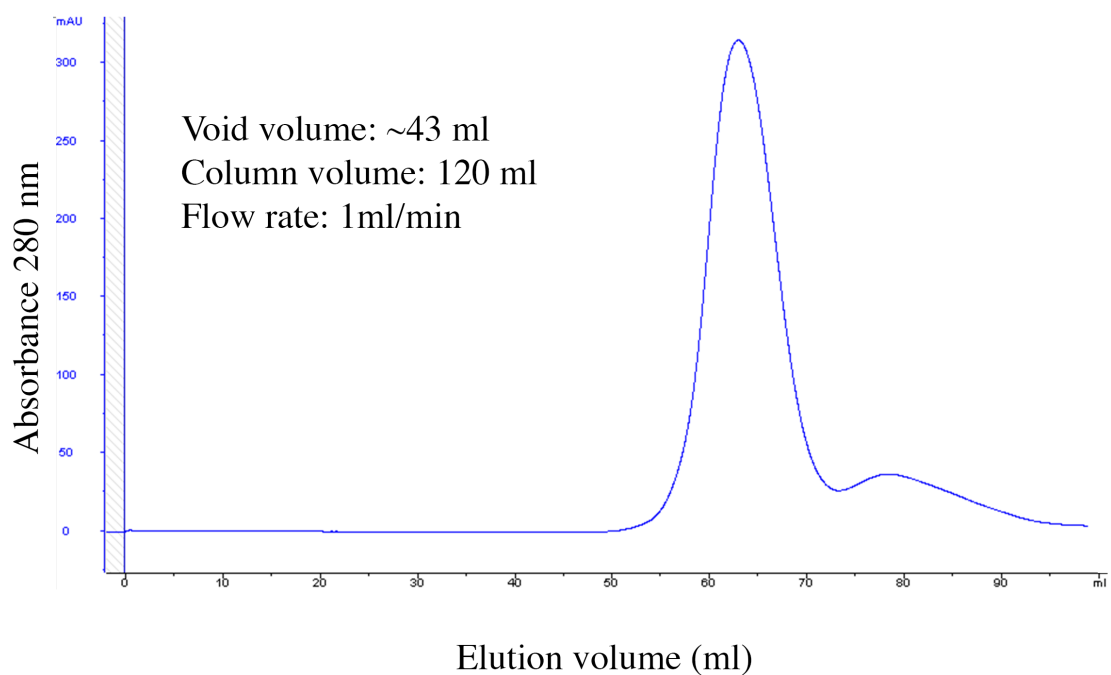


Figure 17. Chromatogram of *E. coli* BamAB+CDE co-purification on gel filtration column. The samples were injected onto a HiLoad 16/600 Superdex 200 prep grade column (GE healthcare) pre-equilibrated with 20mM Tris-HCL (pH 7.8), 300 mM NaCl, 0.5% C₈E₄.

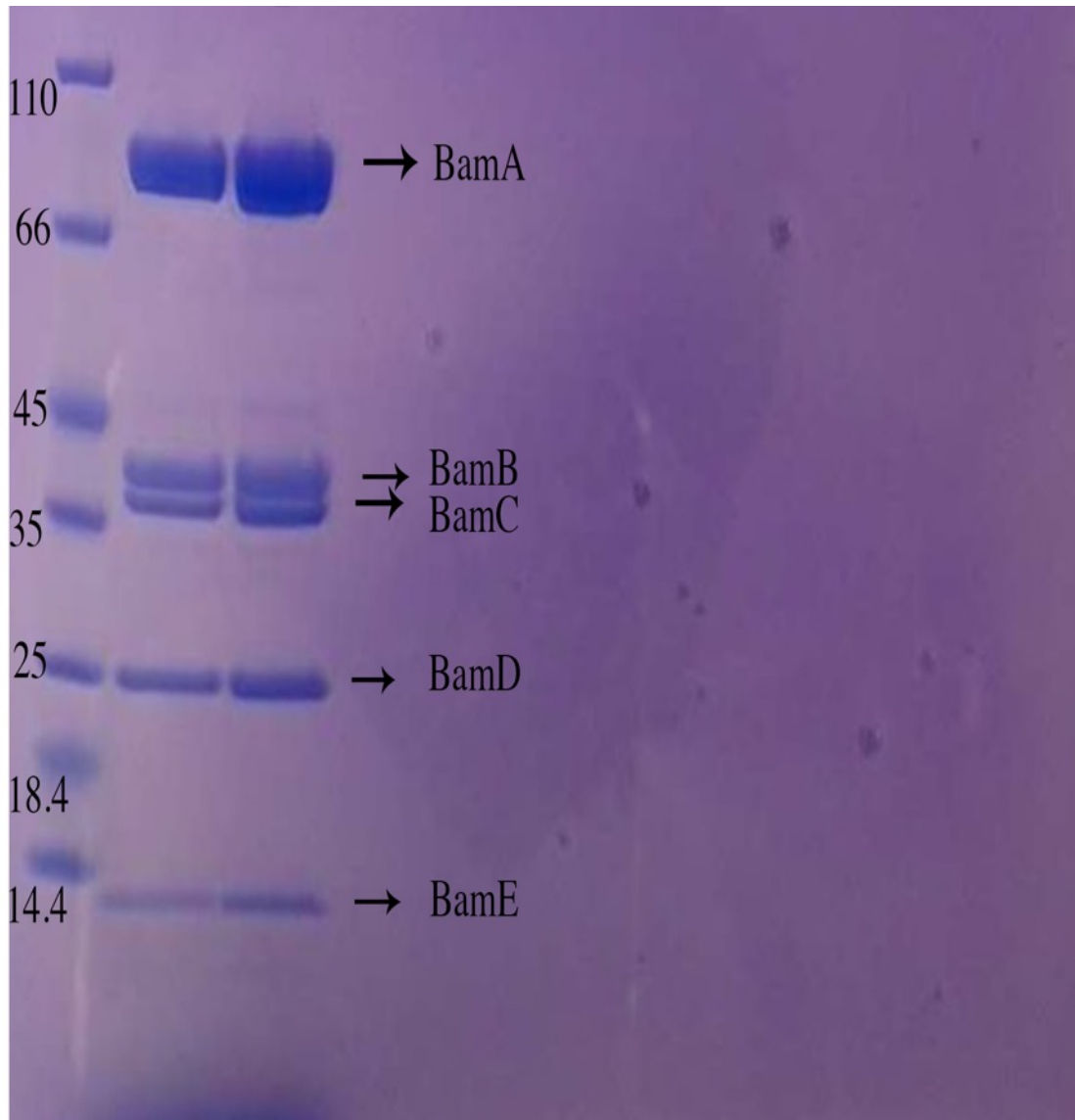


Figure 18. SDS-PAGE gel picture of the co-purified *E. coli* BamAB+CDE complex.

The Bam complex was co-purified from harvested cells of *E. coli* BamAB and *E. coli* BamCDE. Lane M, protein molecular weight marker with indicated molecular sizes in kDa.

Lane 1 and 2, fractions from the main peak in Figure 17.

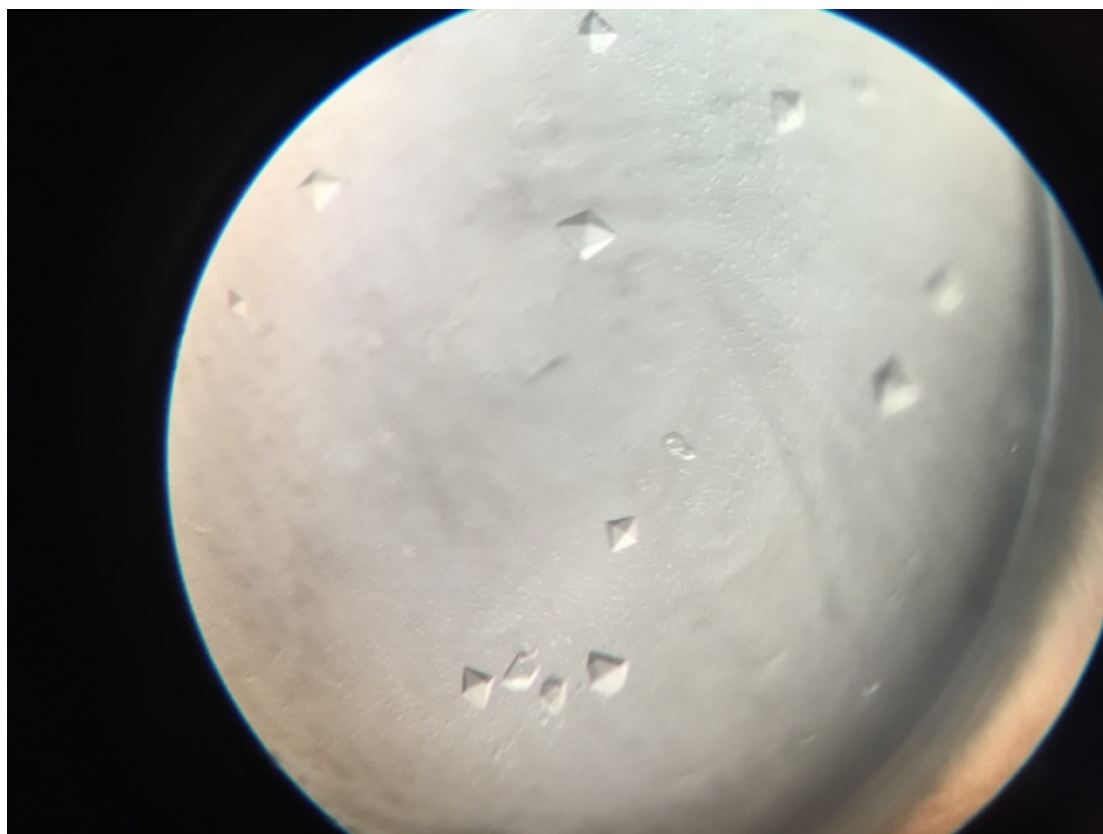


Figure 19. A picture of the crystals produced from co-purification of *E. coli* BamAB and *E. coli* BamCDE.

4.3 Expression of all five subunits in a single expression plasmid (pJH114 and pYG120)

Mainly because of the difficulty in obtaining well-diffracting crystals from the current gene construction, further improvement in cloning and expression was considered. The optimal expression could result from a construct in which all five genes are placed together and are subsequently expressed to form the entire complex intracellularly upon induction. In light of this, pJH114 plasmid was requested from H. D. Bernstein at National Institutes of Health USA and pYG120 was subsequently created (performed by colleagues). I used pYG120 to overexpress and purify the BAM complex. The yield was ~2mg per litre of cell culture. The gel filtration profile shows an almost identical pattern to that of BamAB+CDE co-purification (Figure 20) and SDS-PAGE analysis of eluted peak fractions reveals that the BAM

complex is highly purified and the relative ratios of individual subunits are generally identical to that observed for previous separate overexpression (Figure 21b). The purified complexes using pJH114 (performed by colleagues) and pYG120 were analyzed by SDS-PAGE and compared and it showed a comparable integral ratio of individual subunits after purification of the BAM complex with the exception that BamB seems to be of lower amount comparing to other components and SDS-PAGE result of BamB from pYG120 (Figure 21a). Strikingly, both constructs expressed the BAM complex that was crystallized into high-resolution crystals in a relatively straightforward fashion.

Subsequent crystallization trials for pYG120-overexpressed BAM complex using the sitting drop method yielded diverse types of crystals from the screening conditions. They were all cryoprotected by supplementing the crystallization solution with 20% glycerol and tested for diffraction (Figure 22). The structure of the BamACDE complex was first determined using selenomethionine substituted crystals and the structure of the BamABCDE complex was determined later on using NaI co-crystallization.

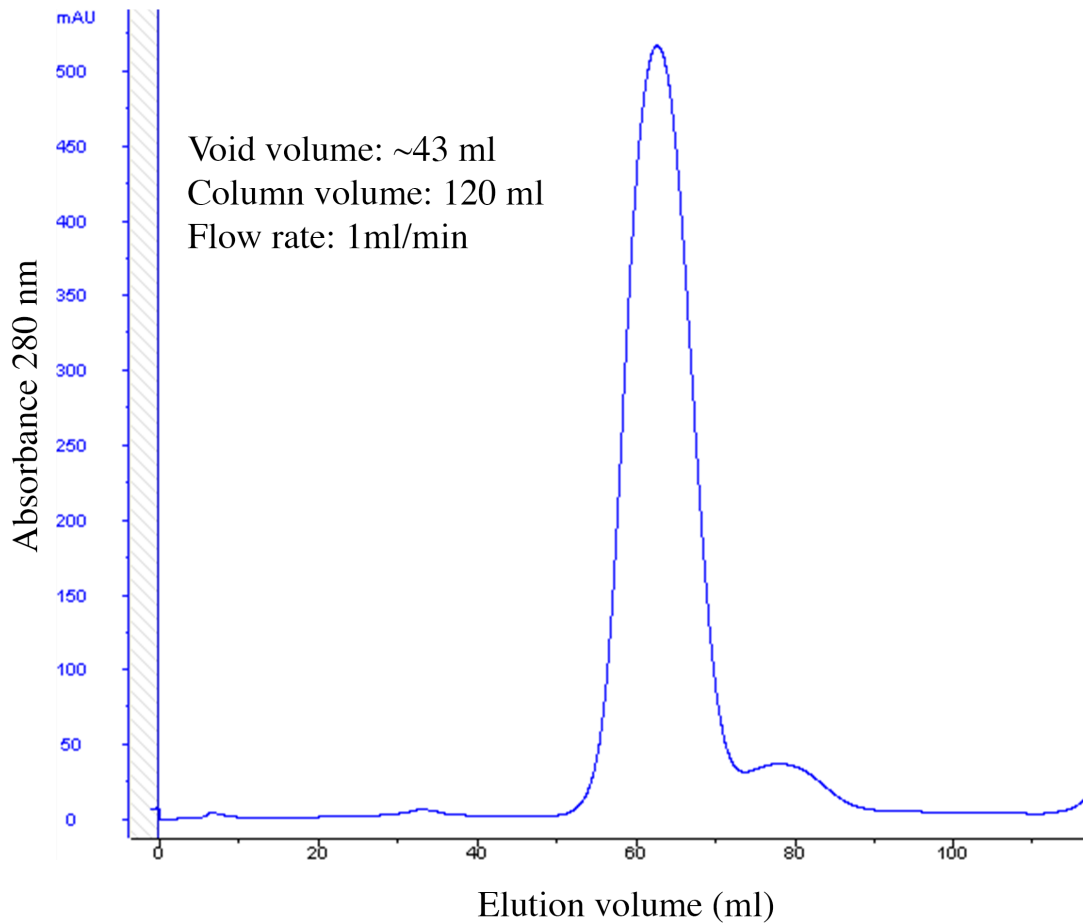
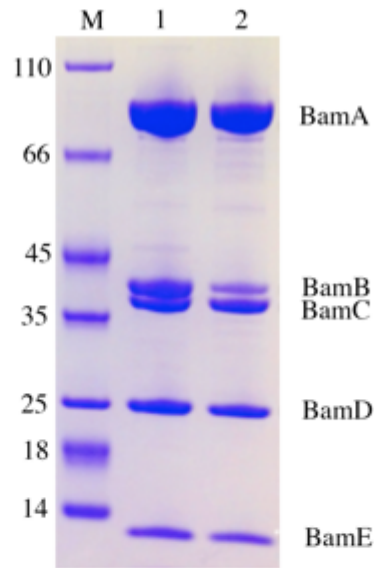
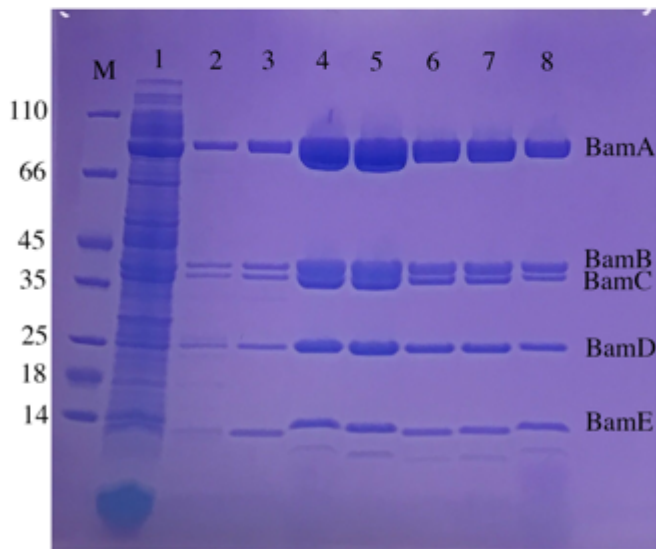


Figure 20. Chromatogram of *E. coli* BAM complex expressed from pYG120 on gel filtration column.

The samples were injected onto a HiLoad 16/600 Superdex 200 prep grade column (GE healthcare) pre-equilibrated with 20 mM Tris-HCL (pH 7.8), 300 mM NaCl, 1 CMC β -OG and 1 CMC LDAO.



(a)



(b)

Figure 21. (a) SDS-PAGE comparison between the BAM complex proteins purified from pYG120 and pJH114 (b) SDS-PAGE result of the BAM complex proteins purified from pYG120.

Lane M in (a), protein marker with indicated molecular sizes in kDa. Lane 1, purified BAM complex proteins from pYG120 overexpression. Lane 2, purified BAM complex proteins from pJH114 overexpression. The relative stoichiometric molecular ratio of BamB in the complex is lower. Lane M in (b), protein marker with indicated molecular sizes in kDa. Lane 1, fraction of flow through sample during IMAC. Lane 2-8, fractions from the main peak in Figure 20.



Figure 22. A picture of the well-diffracting native crystals generated from pYG120.

4.4 BamACDE structure generated from pJH114 (determined by colleagues)

At a resolution of 3.9 Å, the BamACDE structure was first solved by using selenomethionine derivatives. It reveals that BamB is absent in the electron density, leaving BamA (Glu22-Ile806), BamC (Cys25-Lys344), BamD (Glu26-Ser243), BamE (Cys20-Glu110) to form a four-protein complex (Figure 23, Figure 24, Figure 25). The four-protein complex measures 115 Å in length, 84 Å in width and 132 Å in height. For uniformity with reported publications, the β -strands of the barrel domain of BamA are named β 1C- β 16C. BamA follows the general structural feature of OMPs that the C-terminal β -barrel domain is embedded in the OM, while the five soluble POTRA domains extend from the barrel domain and associate with three other subunits. Moreover, the POTRA domains encircle the β -barrel

with BamD, forming a ring in the periplasm. BamC is integrated into the complex by interactions of the C-terminal globular domain with BamD and POTRA 2. As for BamE, it interacts simultaneously with the C-terminal domain of BamD and POTRA 4 and 5 of BamA.

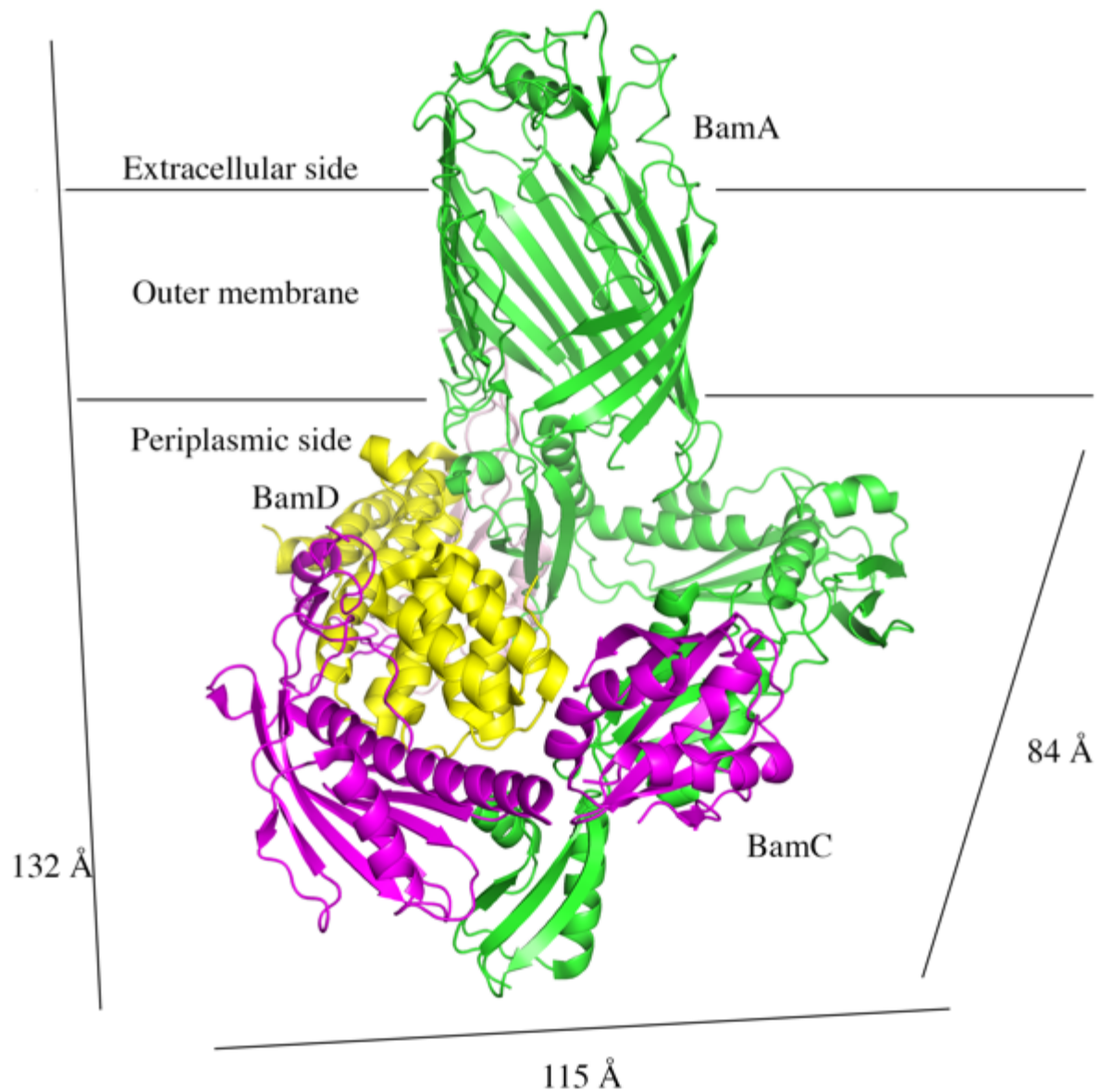


Figure 23. Catoon representation of the BamACDE structure (horizontal view from the OM plane).

BamA is coloured in green, BamC in magenta, BamD in yellow and BamE in light pink. The β -barrel of BamA is embedded in the OM. The dimensions were measured using the widest points of the model in respective axes.

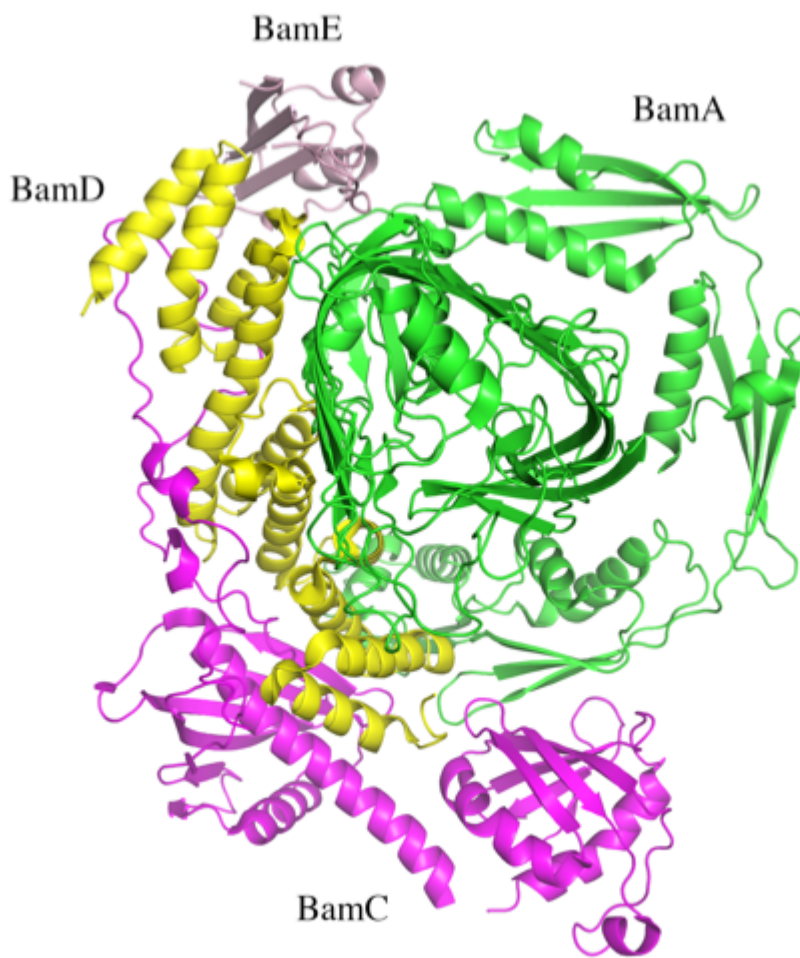


Figure 24. BamACDE structure model viewed from the extracellular side.

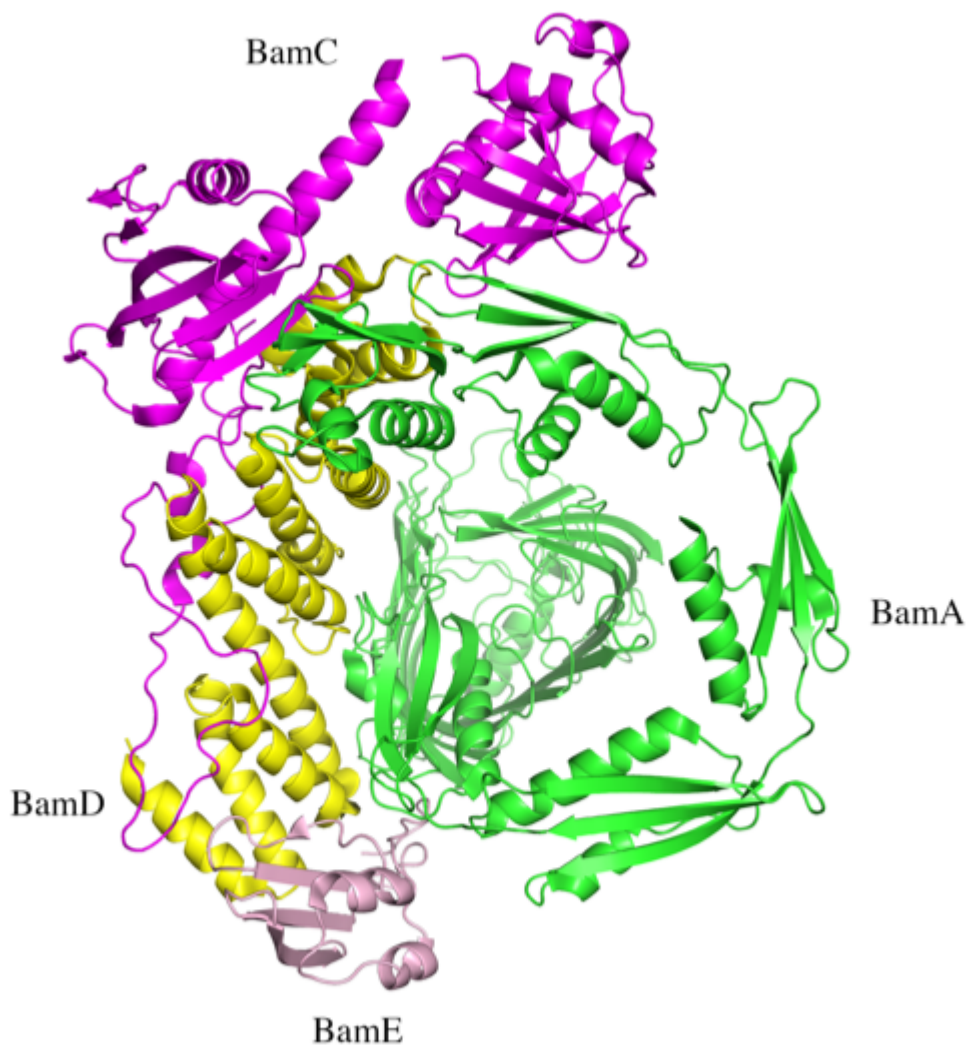


Figure 25. BamACDE structure model viewed from the periplasmic side.

4.5 BamABCDE structure generated from pYG120

The whole BAM complex was successfully crystallized and the structure was solved using single-wavelength anomalous dispersion [20] and molecular replacement phasing methods. The structure was determined to a resolution of 2.9 Å and confirmed to contain all five subunits (Figure 26, Figure 28, Figure 29). Figure 30 presents iodide anomalous signal map contoured at 4 σ in a single unit cell of BamABCDE crystal. The five-protein complex measures ~120 Å in length, 98 Å in width and 140 Å in height (Figure 26). It should be

pointed out that BamC is not visible in the model in full length but contains the N-terminal loop (Val35-Pro88) attached to BamD. Although SDS-PAGE experiment showed the presence of full-length BamC in both the purified complex and the crystals, the remaining portion other than the N-terminal loop of BamC is highly flexible.

The extracellular loops of BamA (L1 to L8) occupy the entry space of the pore in the barrel; therefore the opening in the crown of the hat is closed and inaccessible to the extracellular environment. The periplasmic side of the pore on the other hand is fully open and the periplasmic ring appears stabilized by interactions of BamD and POTRA domains ready for incoming unfolded OMPs.

Compared to the previously reported structure of *Neisseria gonorrhoeae* BamA, BamA structure in the BamABCDE complex shares similar conformations, in which the β 16C of the barrel domain coils towards the interior of the barrel lumen (Figure 27). This coiled β -strand produces a gap between β 1C and β 15C, which may be implicated in forming a gate to allow OMPs to be inserted into the OM.

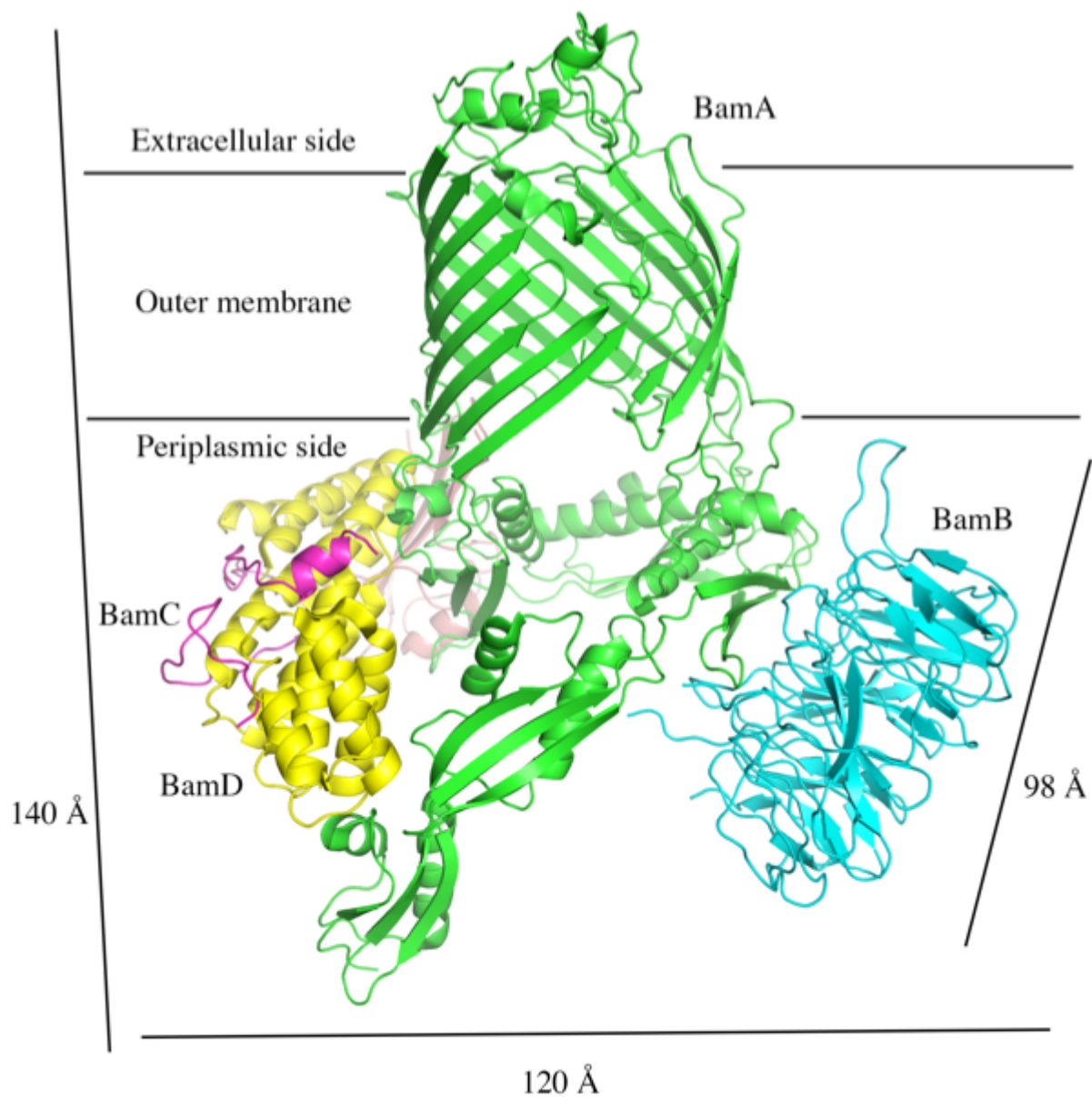


Figure 26. BamABCDE complex structure model viewed from the horizontal OM plane. The colouring of BamA, BamC, BamD and BamE are identical to BamACDE structure figures. BamB is additionally labelled in cyan.

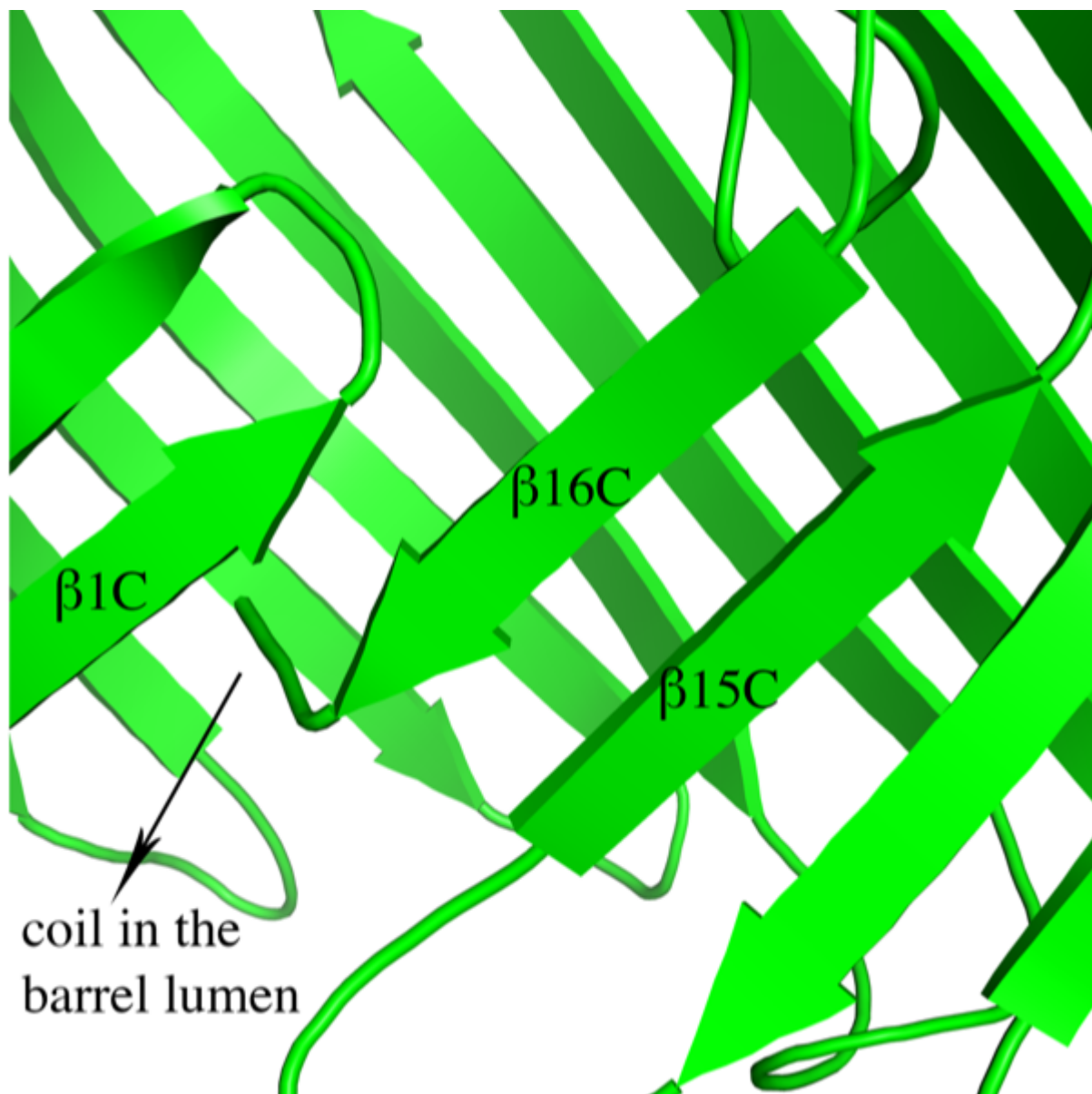


Figure 27. $\beta 16C$ of the barrel domain of BamA is coiled into the barrel lumen.

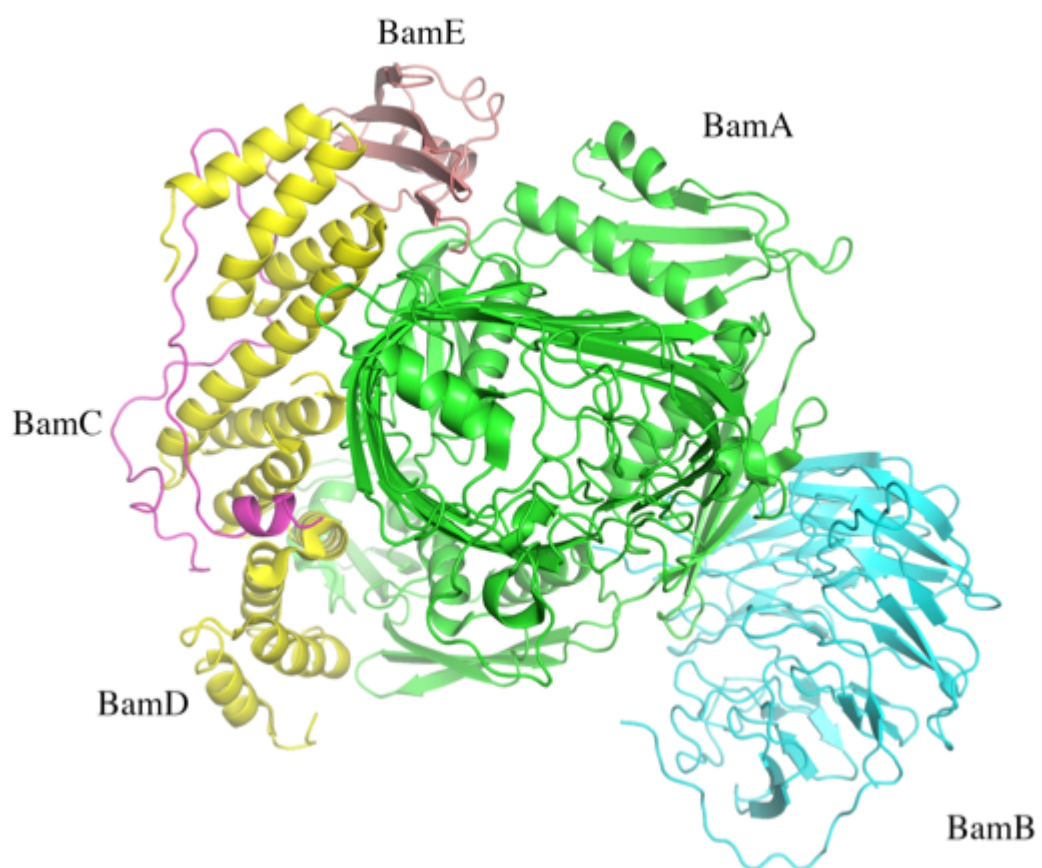


Figure 28. BamABCDE structure model viewed from the extracellular side.

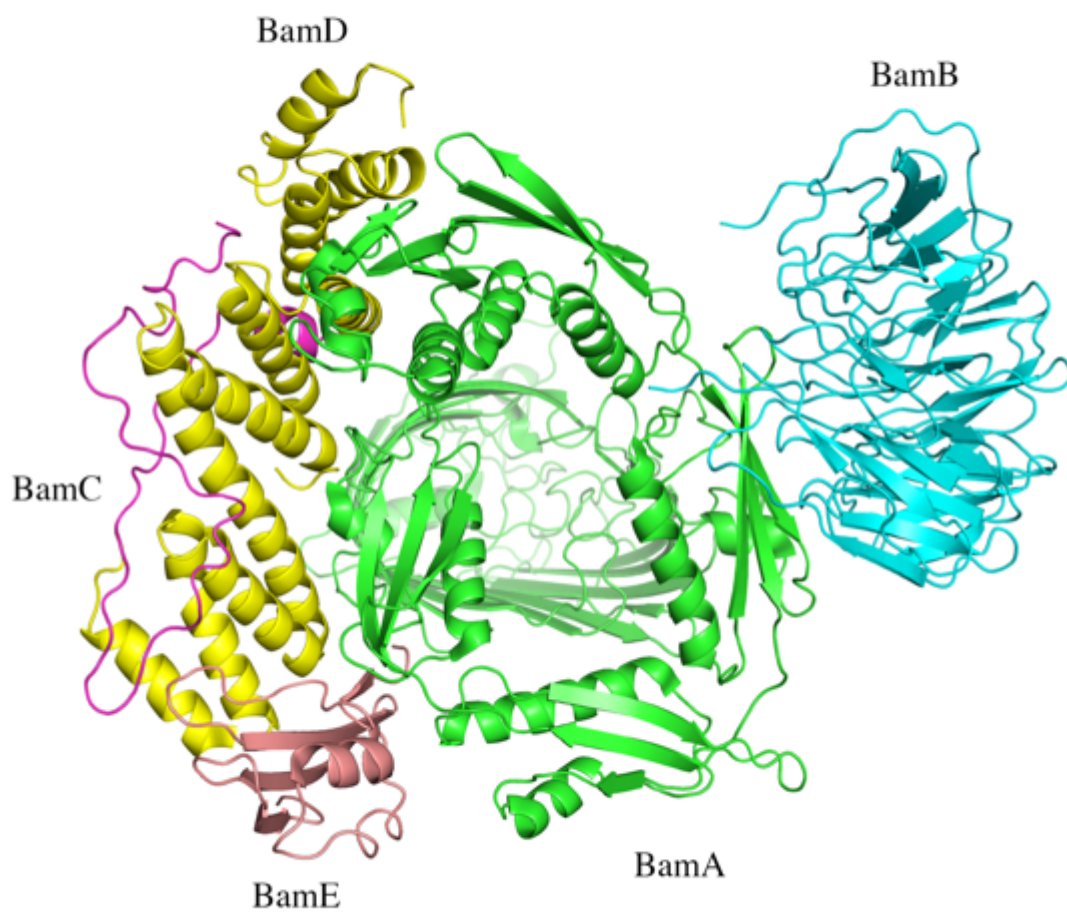


Figure 29. BamABCDE structure model viewed from the periplasmic side.

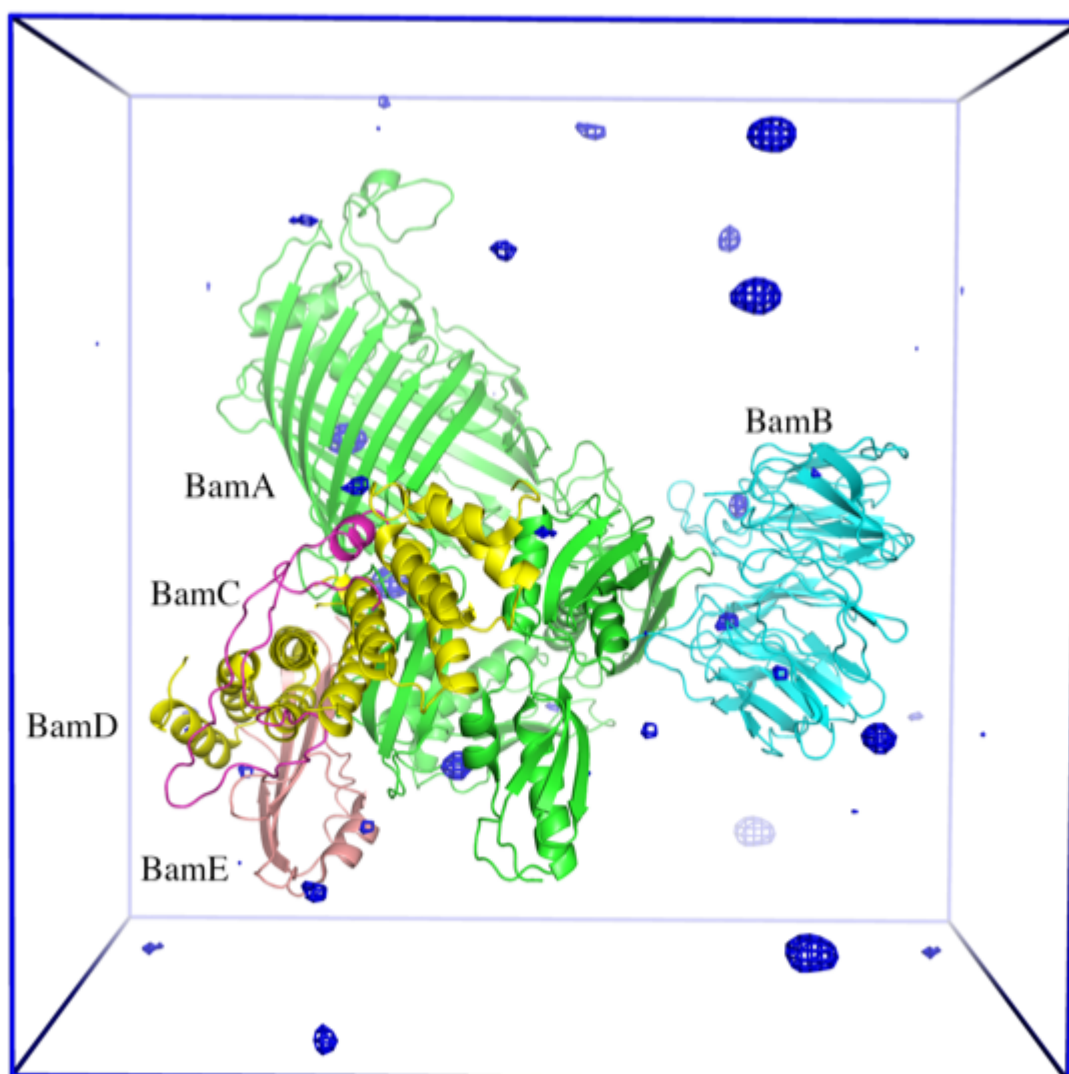


Figure 30. Iodide anomalous signal map in a single unit cell of BamABCDE crystal.

The dark blue meshes depict locations of iodide ions. The dark blue lines represent axes of the unit cell. The contour level is set to 4σ .

4.6 Interactions (defined by proximity within 3.5 \AA) between BamA and BamB

Looking at the two structures, the presence of BamB is the signature feature in the BamABCDE structure. From this structure, the binding of BamB to POTRA 2 and 3 of BamA is clearly seen mainly via salt bridges and van der Waals forces, in which the loops 15,

19, 23 and 27 of BamB at the binding site undergo conformational change to suit the interactions. BamB also interacts with Lys135 and Tyr147 of POTRA 2 (Figure 31).

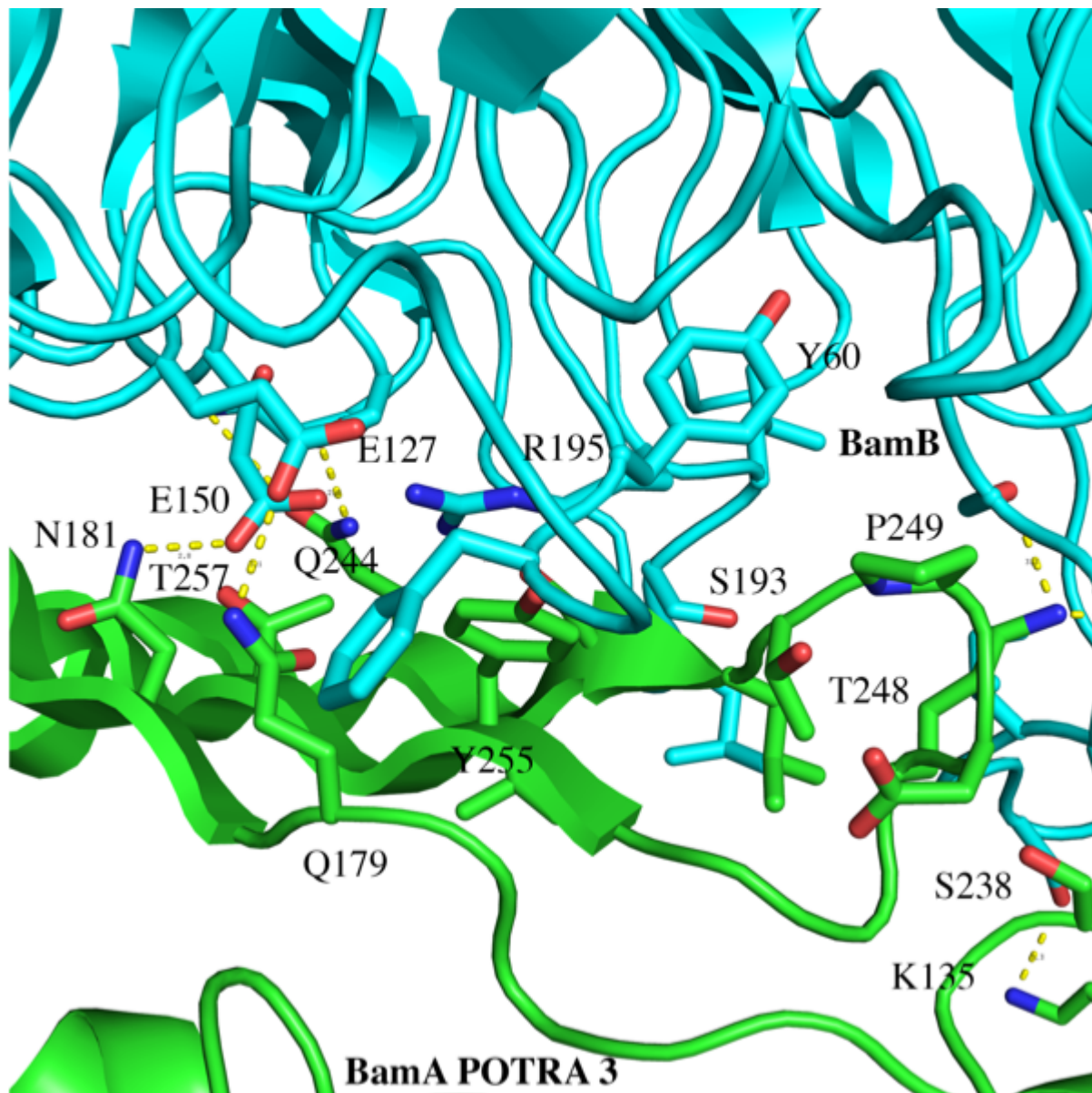


Figure 31. Interactions between BamA and BamB.

Expanded view at POTRA 2 and POTRA 3 in BamA and loops in BamB where interactions take place.

4.7 Interactions between BamA and BamC

Owing to the observation that BamC only exhibits the N-terminal loop in the BamABCDE structure, the interactions between BamA and BamC are therefore best visualized in the BamACDE structure, in which the complete BamC is present. It is clear that BamC contacts with POTRA 1 of BamA through its residues in the N-terminal globular domain, while the C-terminal globular domain associates with POTRA 2 via salt bridges and van der Waals forces with the β -sheets (Figure 32).

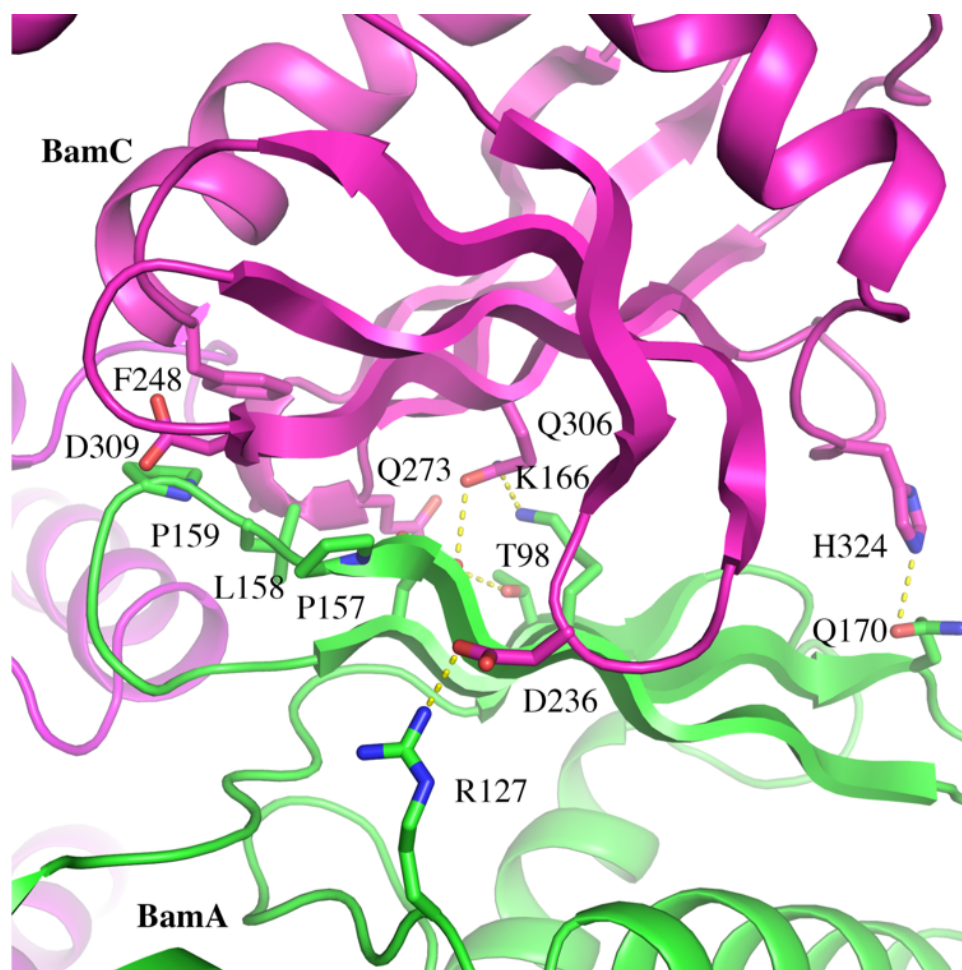


Figure 32. Interactions between BamA and BamC.

In the BamACDE structure, expanded view is provided at the C-terminal globular domain of BamC and POTRA 2 of BamA where salt bridges and van der Waals forces interactions take place.

4.8 Interactions between BamA and BamD

Earlier studies have proposed that BamD interacts only with POTRA 5 of BamA [85], but no direct structural evidence was provided due to a lack of a BamA-BamD structure. In both BAM complex structures, BamD is indeed in contact with POTRA 5 of BamA (Figure 33). Moreover, BamD also interacts with Val480 and Asp481 of the periplasmic turn T2 of BamA as well as residues in POTRA 1 and 2.

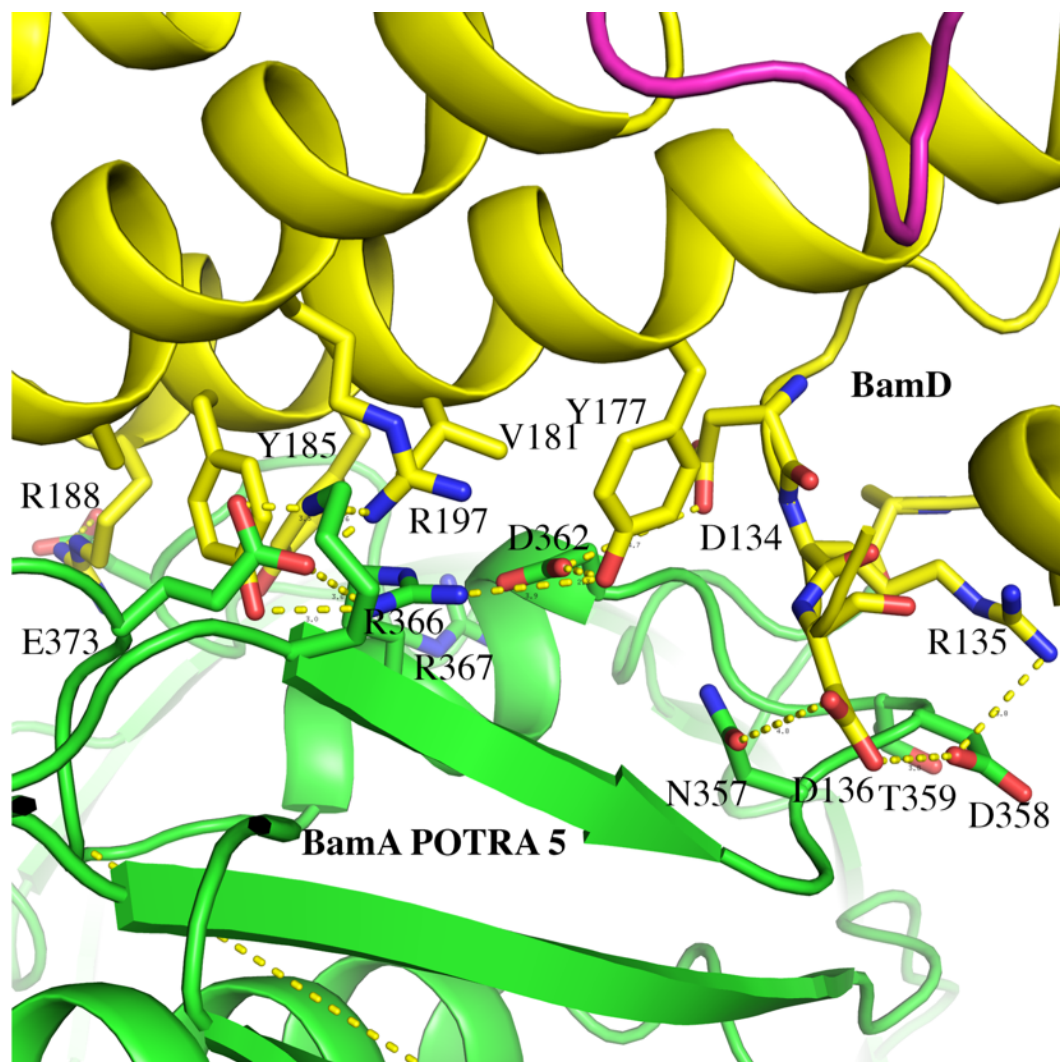


Figure 33. Interactions of BamD with BamA POTRA 5.

Expanded view is provided at the C-terminal domain of BamD and BamA POTRA 5 where interactions take place involving BamA residues Arg366 and Glu373 and BamD residue Arg197.

4.9 Interactions between BamA and BamE

In contrast to a previous report that BamE form contacts only with BamD [89, 182], the periplasmic turns T2 and T3 and POTRA 4 and 5 residues of BamA also interact with BamE residues (Figure 34).

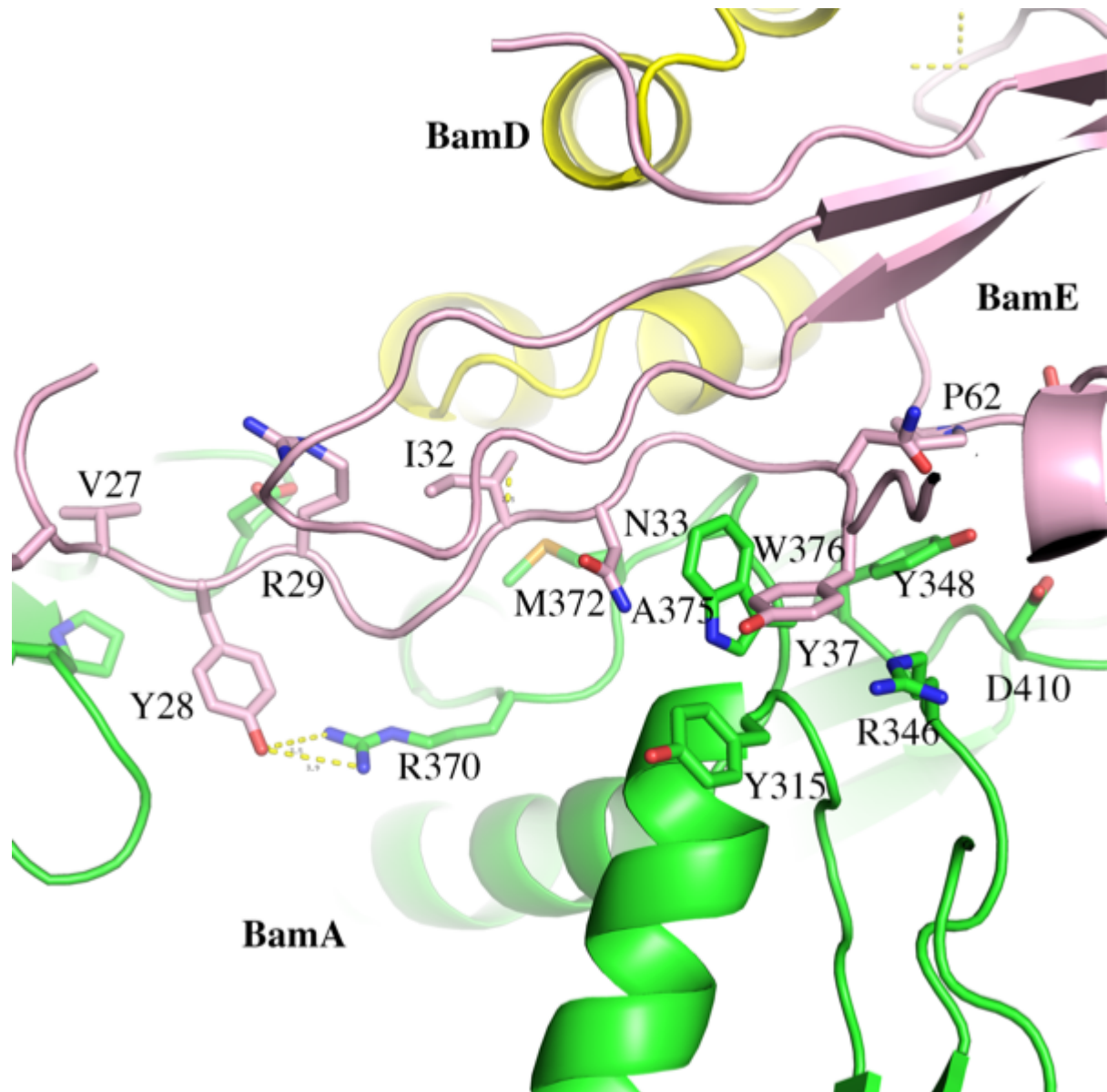


Figure 34. Interactions between BamA and BamE. Expanded view is provided at BamA POTRA 5 and BamE residues where interactions take place.

4.10 Interactions between BamC, BamD, and BamE

In addition to the interactions with BamA, BamC is also closely associated with BamD (Figure 35). Composed of two general structural domains, BamC curves around BamD via the N-terminal loop, inducing substantial interactions, and the loop lies adjacent to BamE, which also promotes extensive contacts. The globular domain of BamC also interacts with the N-terminal domain of BamD (Figure 35). BamE, the smallest subunit in the complex, also forms multiple contacts with BamC and BamD. Specifically, the residues Pro67 and Phe68 in BamE interact with residues Met56 and Ile57 in BamC via van der Waals forces (Figure 36). BamE is in contact with the C-terminal domain residues of BamD (Figure 37).

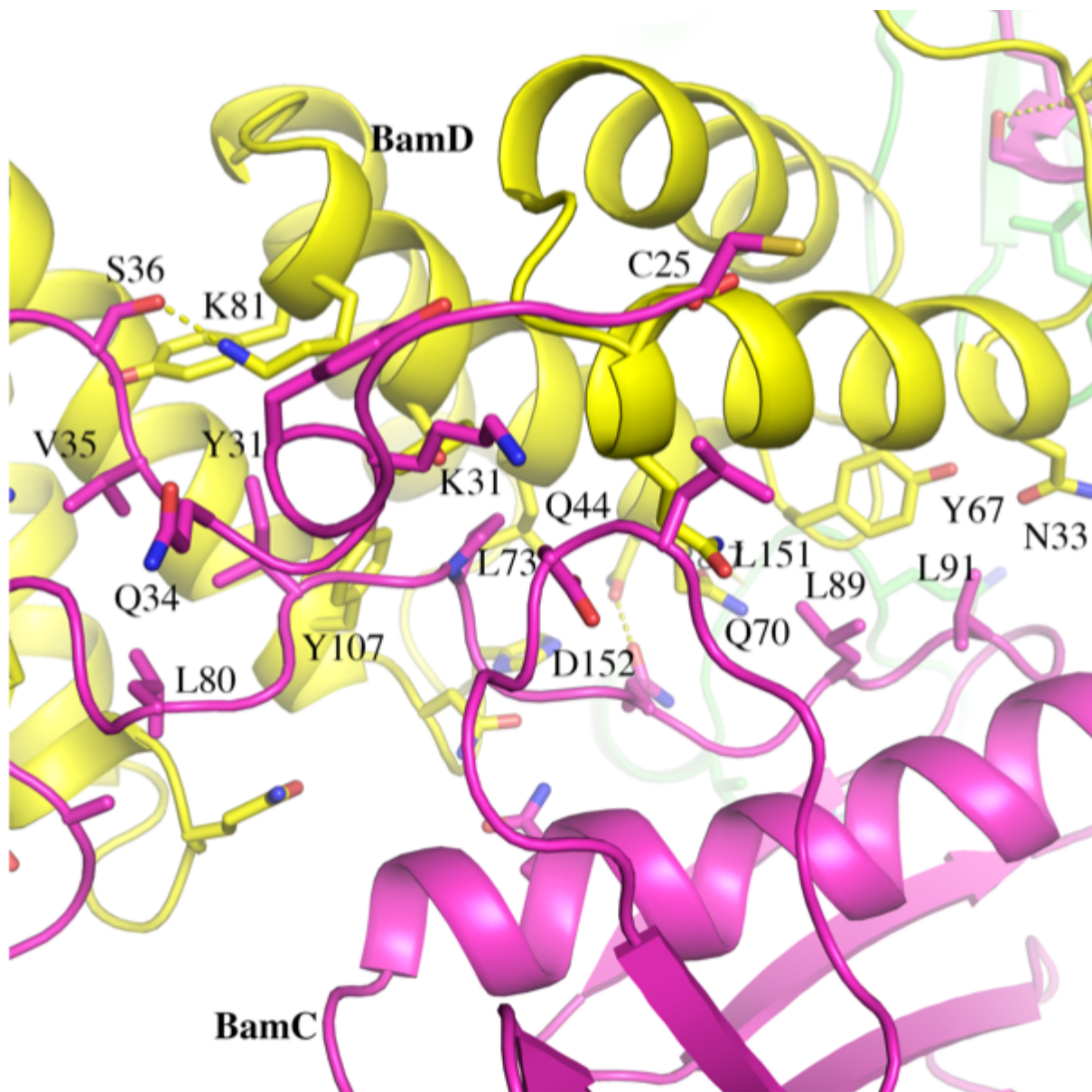


Figure 35. Interactions between BamC and BamD.

Expanded view is provided at the N-terminal domain of BamD and BamC residues where salt bridges and van der Waals forces interactions take place.

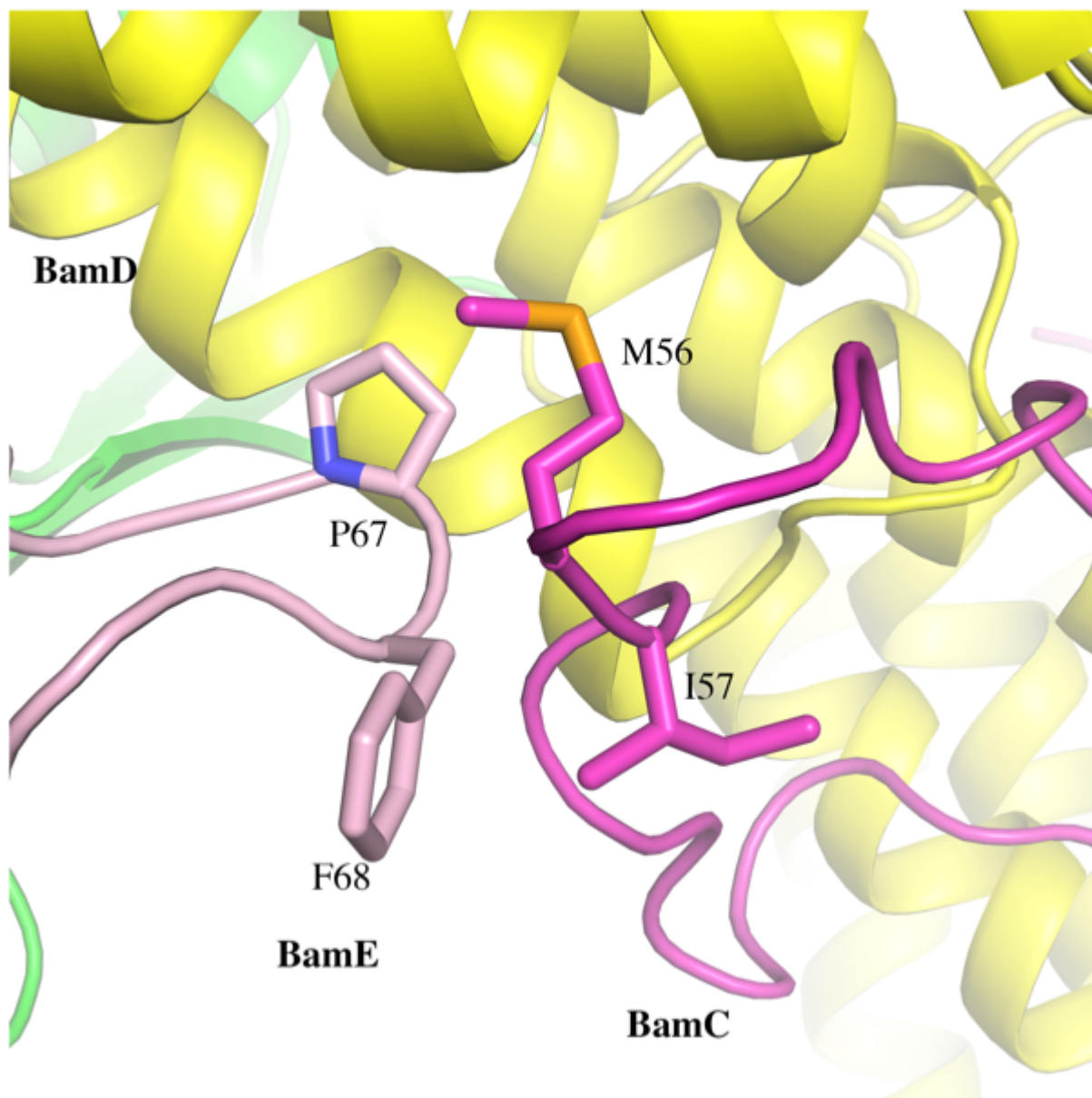


Figure 36. Interactions between BamC and BamE.

Expanded view at BamE residues P67 and F68 and BamC residues M56 and I57 where van der Waals forces interactions take place.

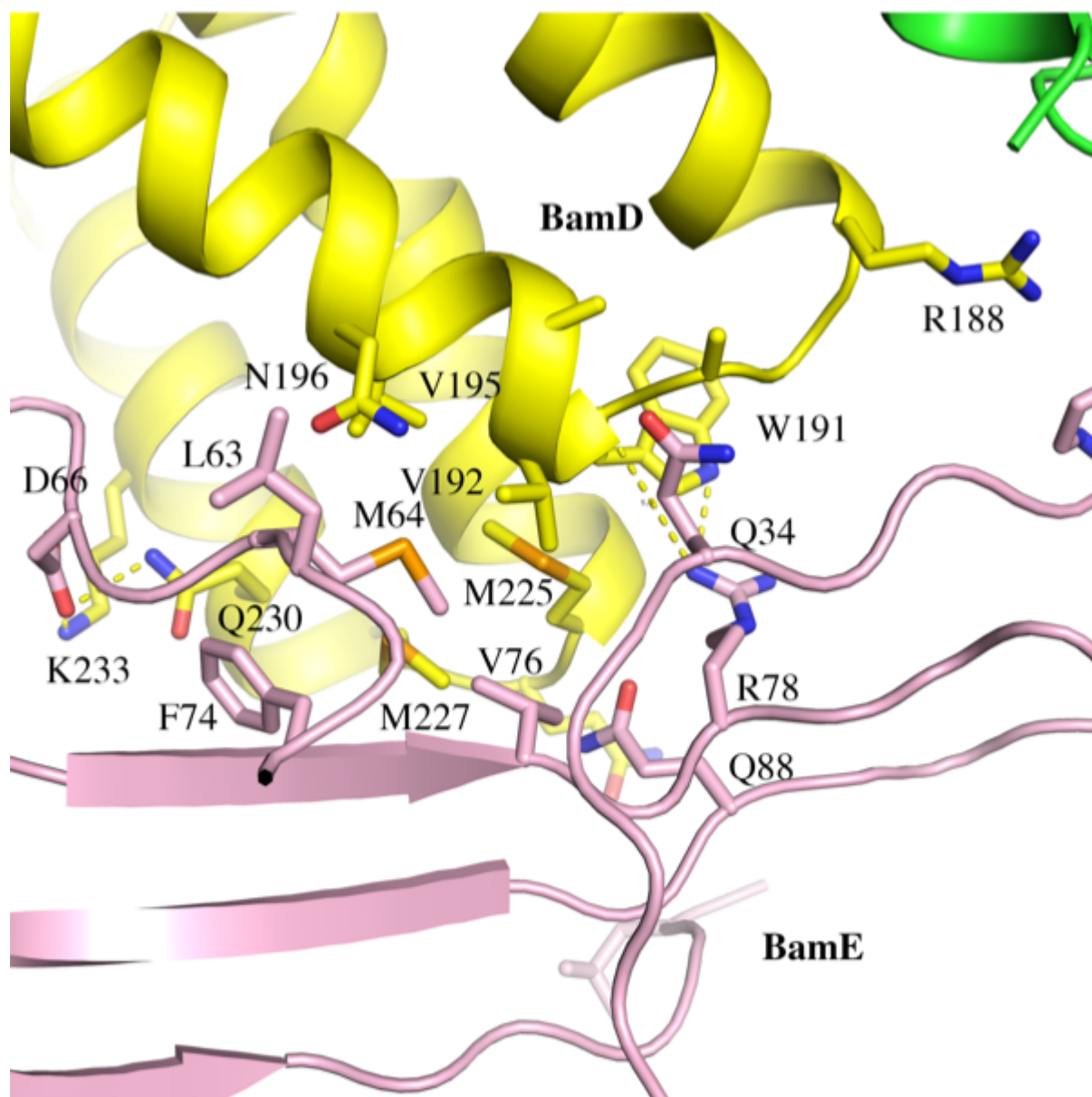


Figure 37. Interactions between BamD and BamE.

In the BamACDE complex, expanded view is provided at the C-terminal domain of BamD and BamE residues where interactions take place.

4.11 Conformational changes in BamA

Unlike the BamABCDE structure, the periplasmic POTRA domain in BamACDE undergoes $\sim 30^\circ$ overall rotation with respect to the location in BamABCDE structure, which leads to POTRA 5 and turn T1 and T4 occluding the periplasmic vestibule of the barrel and preventing entry of OMP substrates (Figure 38a). Another unique and unprecedented feature

observed in the BamACDE structure is that in the β -barrel, the first β -strand, $\beta 1C$, and the last β -strand, $\beta 16C$ are not continuous and instead disrupted by the rotation of the first six β -strands of the barrel away from the barrel pore to an angle of $\sim 65^\circ$ and distance of $\sim 15\text{\AA}$, resulting in a lateral opening of the barrel facing the interior of the pore to the external OM (Figure 38b).

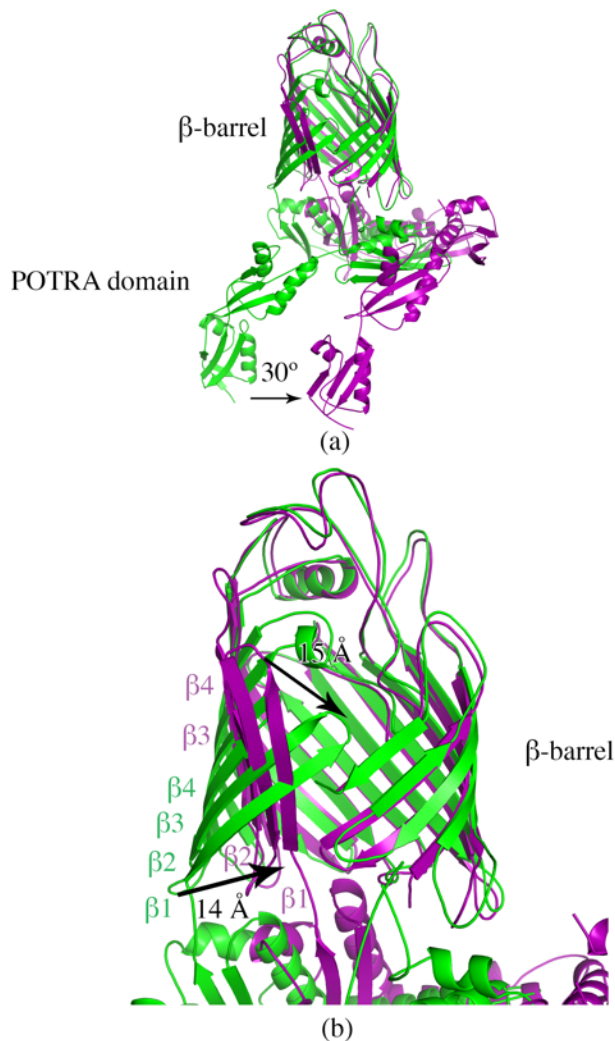


Figure 38. Conformational changes observed in BamA structure.

(a) Superimposition of the BamA structure in BamABCDE structure (green) onto the BamA structure in BamACDE structure (purple) aligned with the barrel domain. The barrel domain exhibits dramatic conformational changes in $\beta 1C - \beta 6C$ and between $\beta 1C$ and $\beta 16C$. Moreover, the periplasmic POTRA domain in BamACDE undergoes a $\sim 30^\circ$ rotation from the BamABCDE structure. (b) Expanded view of the conformational changes in the barrel domain from the inward-open state to the lateral-open state. Rotation of $\beta 1C - \beta 6C$ of BamA of 65° and the distance of $\sim 15 \text{ \AA}$ leads to lateral opening of the barrel.

CHAPTER 5

**Results: structural studies of an outer membrane protein OmpU
from *Vibrio cholerae***

5.1 OmpU purification and crystallization

Initial cell culture started with 4L Terrific Broth media. Cells were harvested and disrupted as described. After being loaded onto gel filtration column, a single peak appeared (Figure 39) and, according to the position of the eluted peak, the sample was estimated to be in a trimeric form [180]. The eluted samples were analyzed by SDS-PAGE and the presence of OmpU protein was confirmed (Figure 40). The protein is eluted from around 60 ml to 80 ml, with the single peak value reaching 400 mAU (~10 mg yield). The protein was concentrated to 10 mg/ml for crystallization, and following incubation at 16°C crystals emerged in a number of conditions. Figure 41 shows one of the crystals producing high-resolution diffraction data during data collection at DLS.

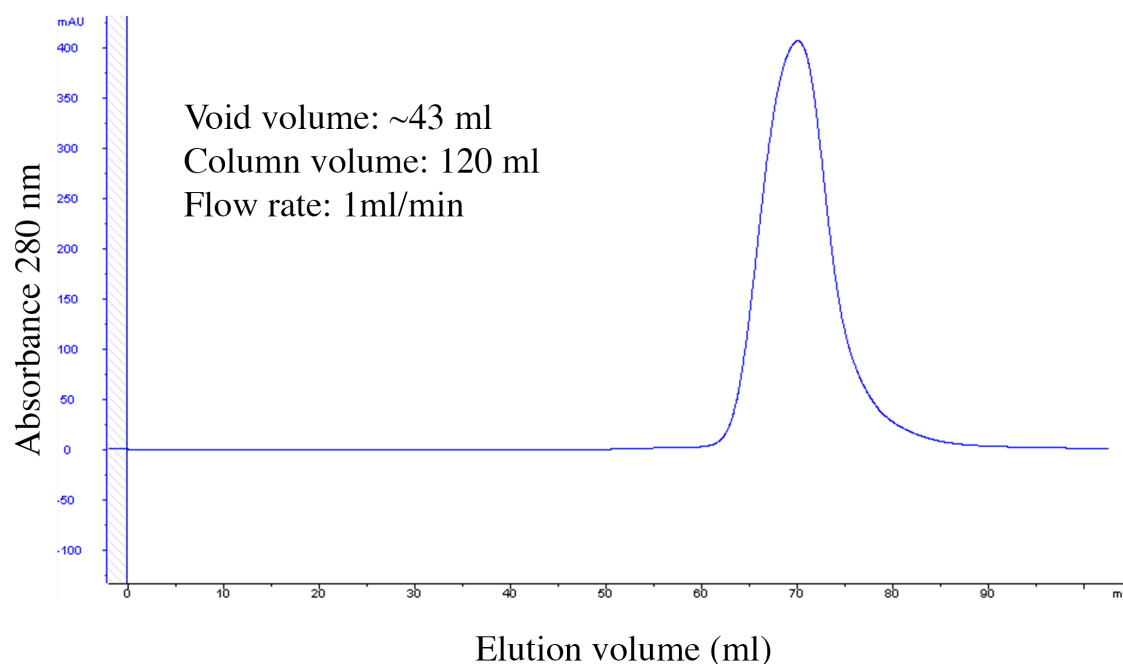


Figure 39. Size-exclusion chromatogram of *V. cholerae* OmpU on gel filtration column. The samples were injected onto a HiLoad 16/600 Superdex 200 prep grade column (GE healthcare) pre-equilibrated with 20 mM Tris-HCL (pH 7.8), 300 mM NaCl, 0.5% C₈E₄.

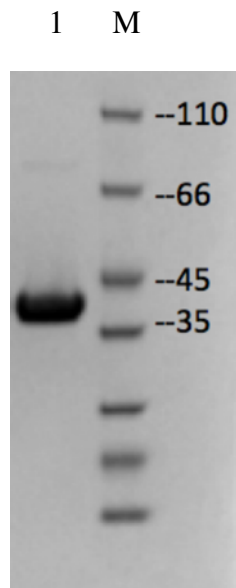


Figure 40. SDS-PAGE gel picture of *V. cholerae* OmpU.

OmpU was highly purified. Lane M, protein marker with indicated molecular sizes in kDa.

Lane 1, fraction from the main peak in Figure 39.

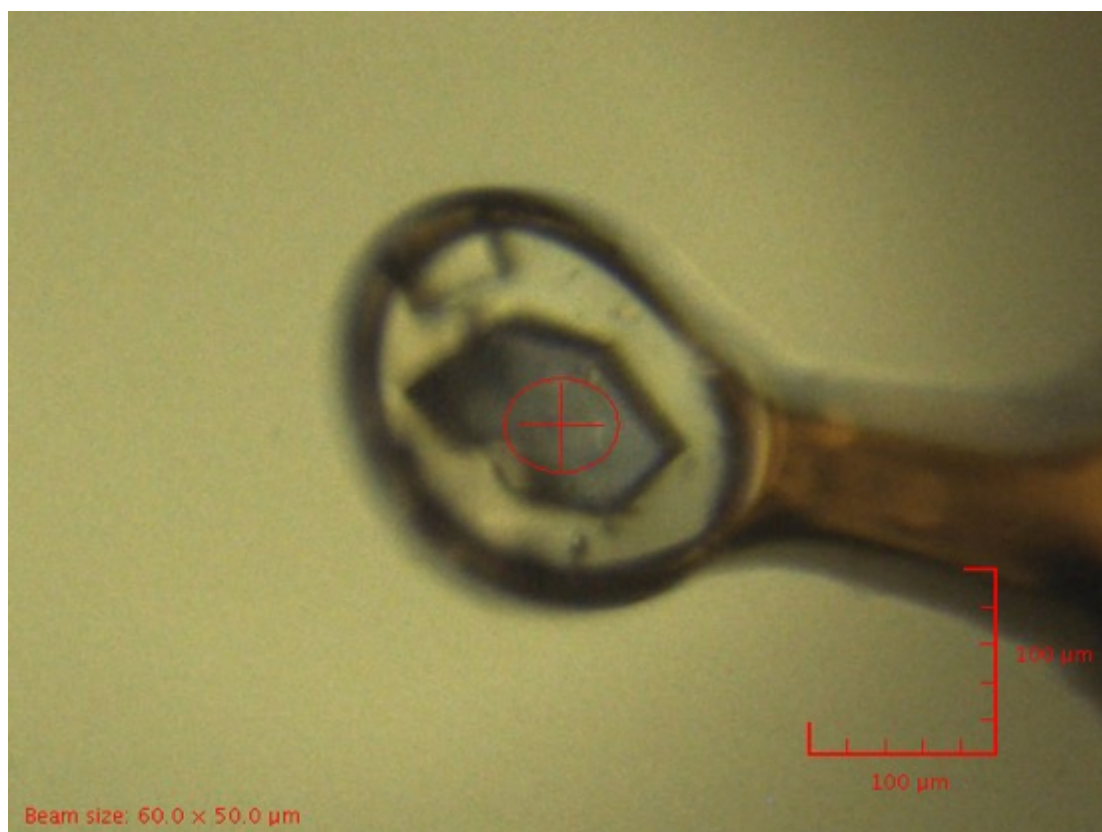


Figure 41. A picture of the well-diffractive OmpU crystal.

5.2 Overall OmpU fold

OmpU was successfully crystallized and the structure was determined to a resolution of ~ 2.2 Å. OmpU forms a homotrimer (Figure 42), and each protomer model contains residues G1-F319 (Figure 43). In addition to the ordered water molecules, 13 detergent molecules and 6 glycerol compounds are assigned, electron densities of one LDAO and two glycerol molecules in OmpU structure are presented in Figure 44. Despite moderate sequence identity to OmpK36 ($\sim 24\%$) [15], the structure of OmpU shares considerable similarity, with a root-mean-square deviation (RMSD) of 1.55 over 268 aligned $C\alpha$ atoms (Figure 45). The fold of the protomer follows general fashion of trimeric porins, in which the 16 antiparallel strands of the β -barrel are connected by 8 short turns at the periplasmic side and 8 loops at the extracellular side (Figure 43). Loop L3 deviates from the wall of the barrel and extends into

the barrel to serve as a constriction loop in the pore lumen. L1, L5, L6 and L7 are turns instead of loops. Protomers are held together to form homotrimers via hydrophobic interactions between barrel surfaces (Figure 42). The protomer-protomer interactions are enhanced by the presence of both the latching loop L2 and L4 that protrudes and makes contacts with the neighbouring protomer. (Figure 42b).

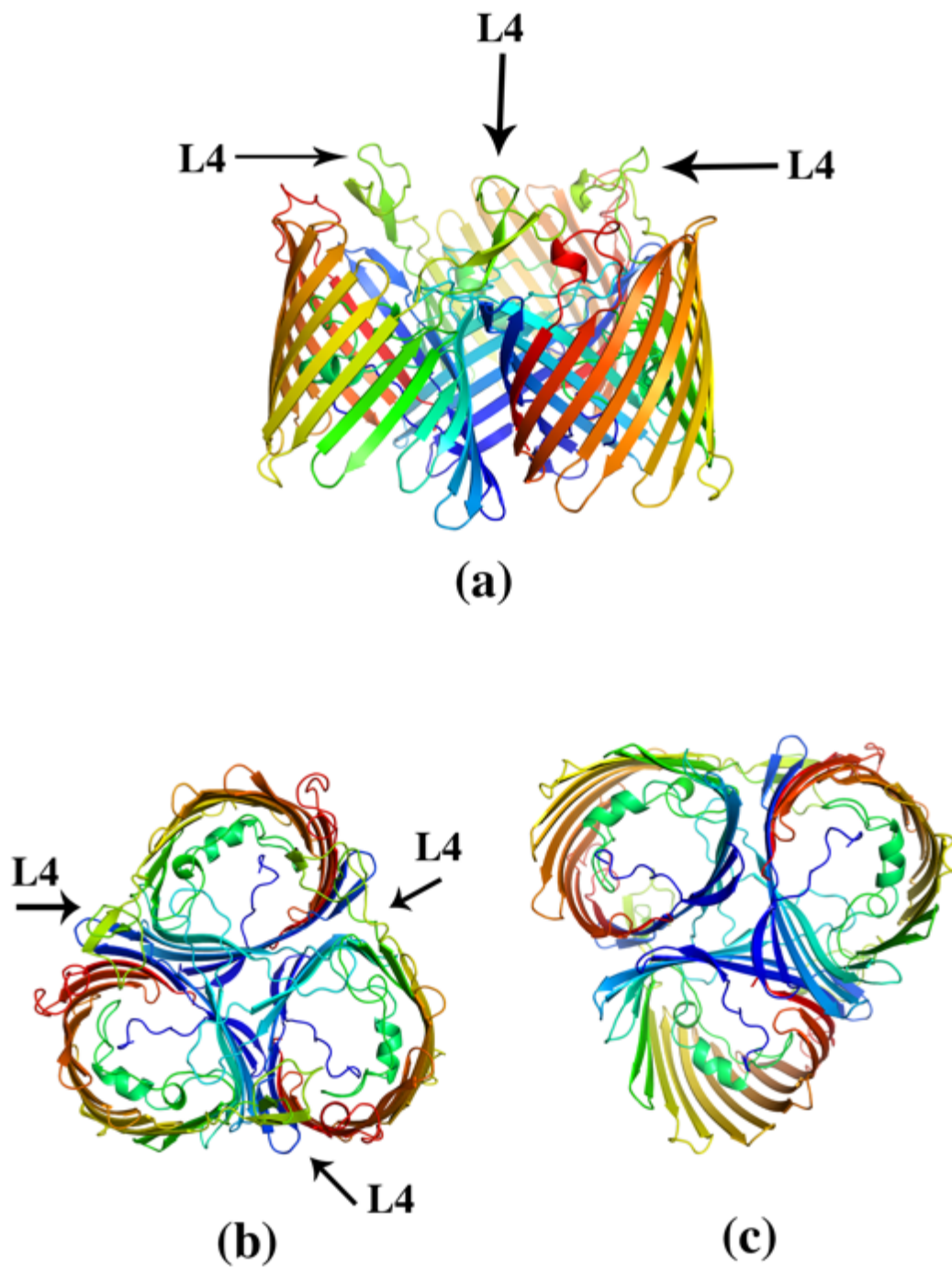


Figure 42. Cartoon representation of OmpU trimer, viewed from the membrane plane (a), the extracellular side (b) and the periplasm (c).

The extracellular loop L4 is labelled in (a) and both L2 and L4 are labelled in (c).

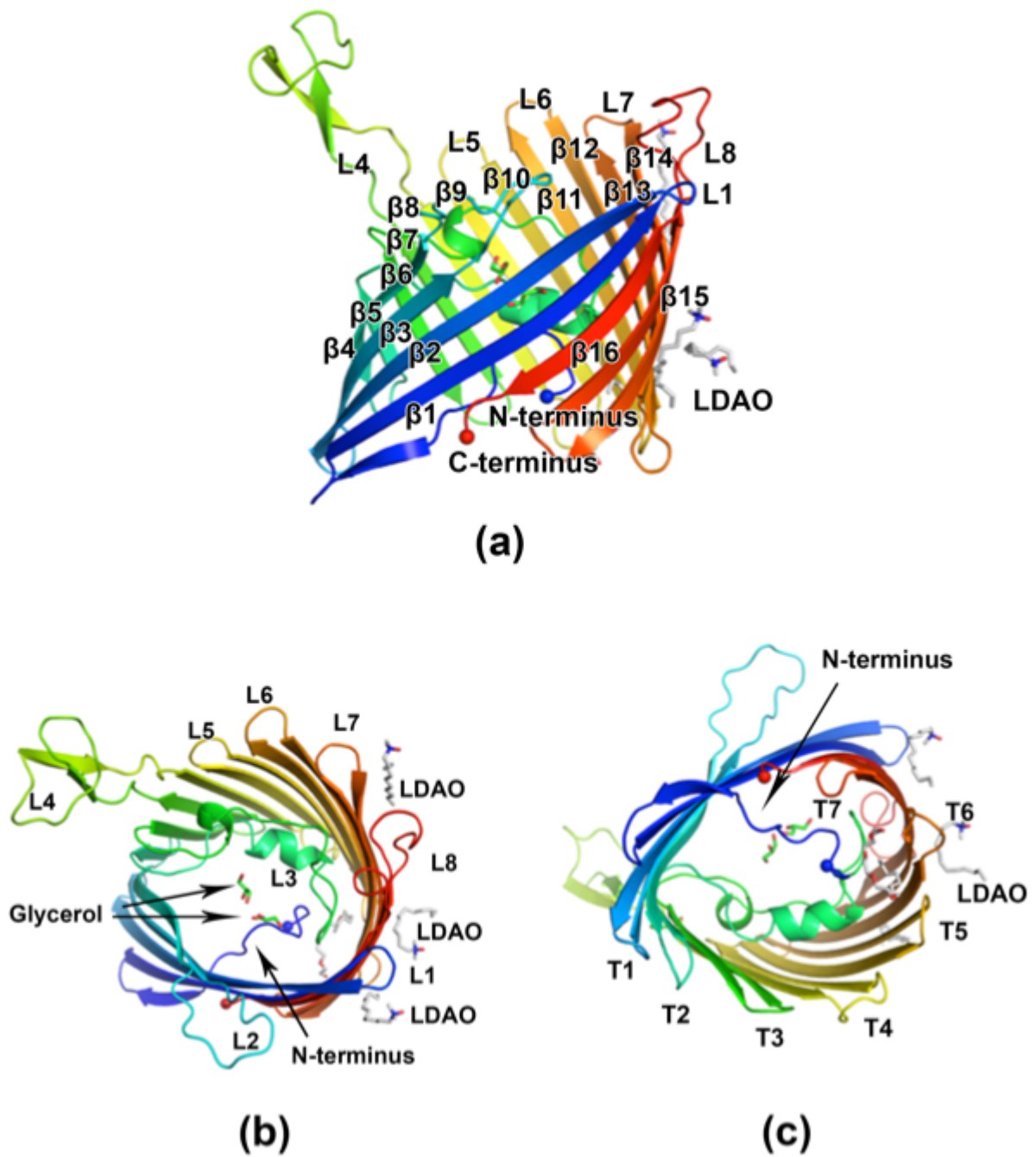


Figure 43. Cartoon representation of OmpU protomer structure, viewed for the horizontal membrane plane (a), the extracellular side (b) and the periplasm (c).

The extracellular loops, the periplasmic turns, the barrel-forming β -strands, detergent and glycerol molecules are individually labeled.

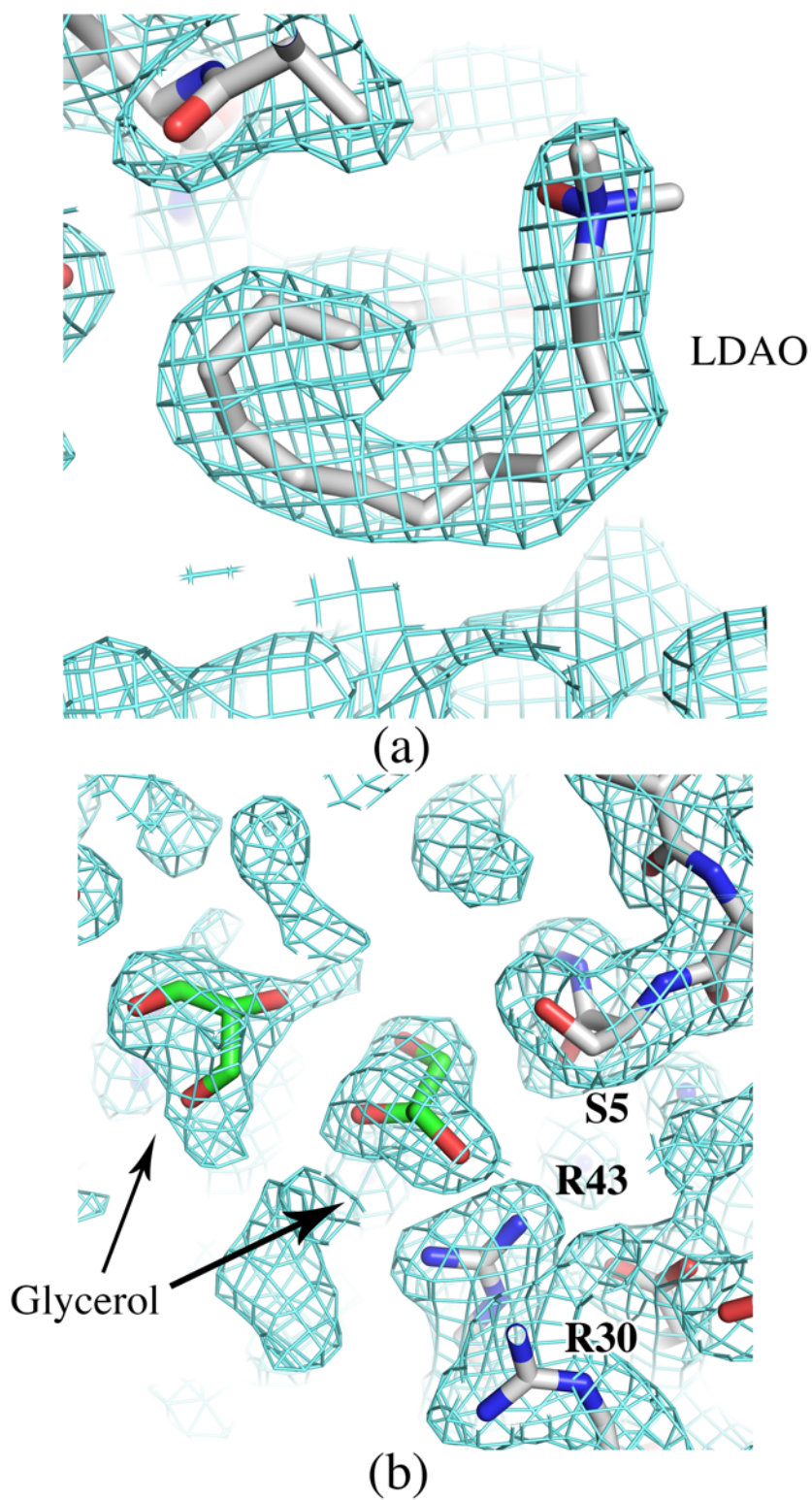


Figure 44. 2Fo-Fc Electron density map (contoured at 1σ) of an assigned LDAO (a) and two glycerol molecules (b) in OmpU structure.

LDAO molecule is shown in full with carbon atoms in grey, oxygen in red and nitrogen in blue. Glycerol molecules are shown in full with carbon atoms in green and oxygen in red.

In further agreement with other bacterial trimeric porins, there are a number of extracellular loops in each protomer that are differentiated by their more mobile nature. These are L4, L6 and L8 (Figure 46). Remarkably, L4 is the longest loop at the extracellular side and, instead of being a single long loop, it contains two short antiparallel β -sheets that resemble a β -hairpin structural motif but are connected by a loop containing more than five residues instead of a short turn. Although L4 of OmpU is shorter than L4 of OmpK36 (Figure 45), L4 of OmpU protrudes further into the extracellular space and projects over L1 of the adjacent subunit to reach proximity to L8. From the side view, L4 has the appearance of a pole that connects transmembrane strand 7 and 8 at the extracellular side. Other loops of OmpU (L5-L8) are all shorter than corresponding loops of OmpK36 to various degrees. The hairpin-like motif of L4 is the highest point of the structure in the extracellular environments and exhibits a high degree of mobility (Figure 46). Structural overlay of OmpU onto the two major *E. coli* porins OmpC and OmpF as well as the MR model OmpK36 shows that L4 in OmpU only partially overlaps with L4 in OmpC and OmpK36 and is minimally overlaid with L4 in OmpF (Figure 45). The edges exposed by both antiparallel β -sheets of L4 may hence serve as an efficient binding target via β -augmentation for foreign receptors at the surface of other organisms.

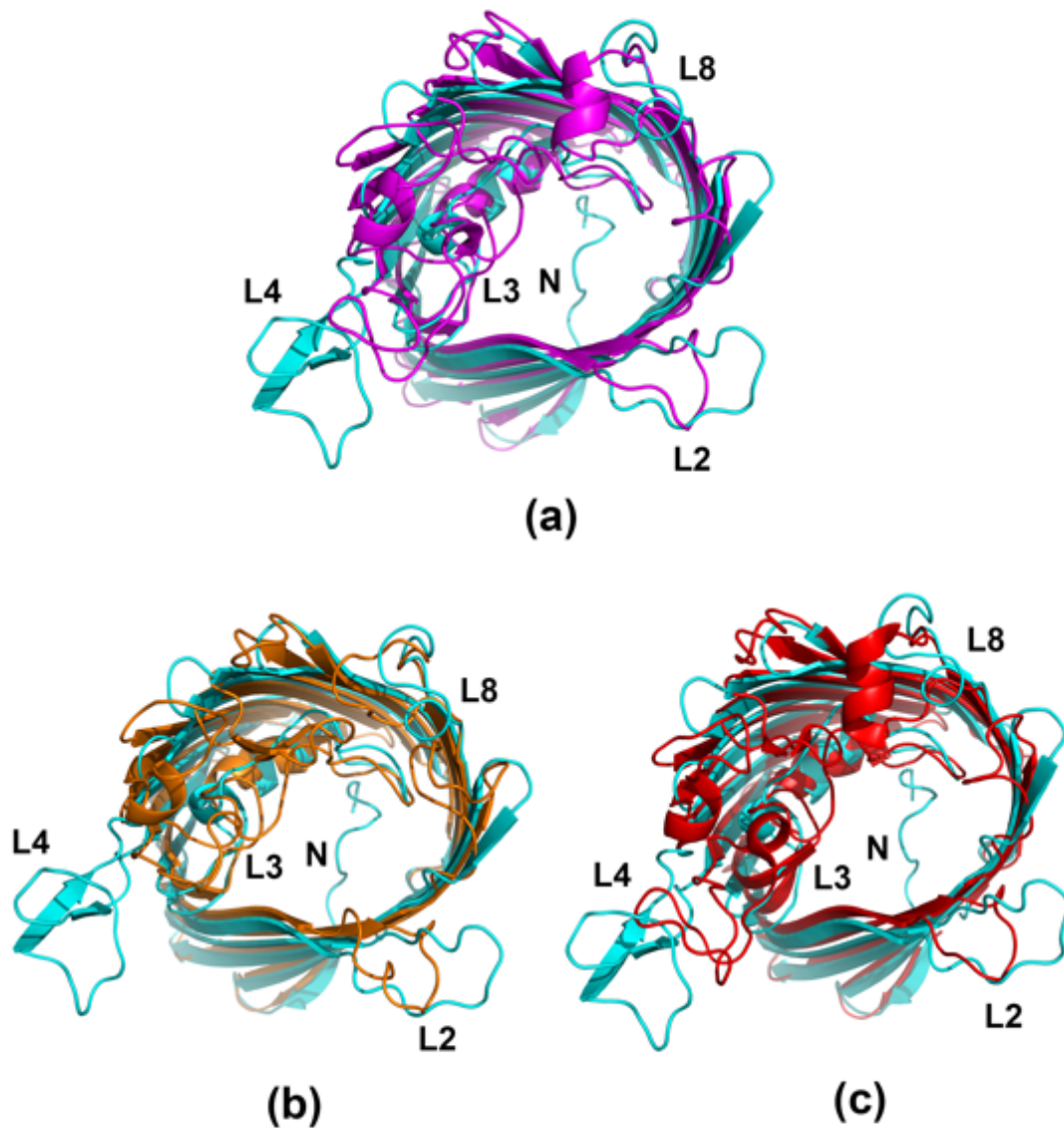


Figure 45. Overlay of the structure of OmpU protomer onto three structurally analogous porin structures viewed from the extracellular side.

In all cases OmpU is shown in cyan. N denotes the N-terminal coil of OmpU in the pore lumen. (a) OmpU superimposed onto protomeric OmpK36 from *K. pneumoniae* (magenta; PDB code 1osm). The two structures superimpose with an r.m.s.d. of 1.55 Å over 268 aligned $C\alpha$ atoms. (b) OmpU superimposed onto protomeric OmpF from *E. coli* (orange; PDB code 2omf). The structures superimpose with an r.m.s.d. of 1.88 Å over 217 aligned $C\alpha$ atoms. (c) OmpU superimposed onto protomeric OmpC from *E. coli* (red; PDB code 2j1n). The structures superimpose with an r.m.s.d. of 1.90 Å over 196 aligned $C\alpha$ atoms. The extracellular loops L2 and L4 and the constriction loop L3 are individually labelled.

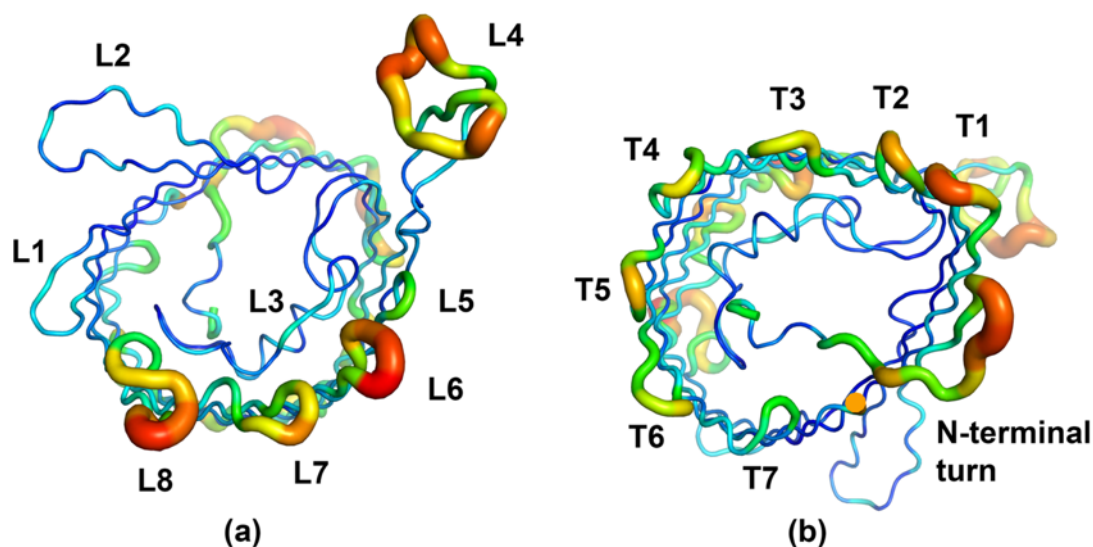


Figure 46. *B* factor diagram of protomeric OmpU represented by the *B* factor putty programme in PyMOL.

(a) Extracellular view showing *B* factors of the extracellular loops. (b) Periplasmic view showing *B* factors of the periplasmic turns. The values of *B* factor are illustrated by colours and line thickness, ranging from low (blue and thin lines) to high (red and thick lines). The external loops (L) and the periplasmic turns (T) are labelled. The overall average *B* factor is 42.5 \AA^2 (data not shown).

5.3 Non-canonical N-terminal coil and an additional constriction zone

As with all other bacterial porins, the β -barrel surrounds an aqueous pore through which cargos diffuse. Interestingly, OmpU forms two constriction regions in the pore (Figure 47). The first L3-formed constriction region is consisted of residues R30, R43, R45, R85, R133, D132, K127, N122 and Y119 on the same horizontal plane (Figure 47). Apart from the L3

constriction loop, the N-terminal short coil (G1 to S11) of OmpU is also located in the pore, which is unprecedented for solved structures of bacterial porins to the best of our knowledge. Its periplasmic-side origin in the pore determines its position below the horizontal plane of the constriction loop L3. In line with its location, the majority of the residues in the coil are hydrophilic. From extracellular top view, the coil is not overshadowed by L3 but instead forms a smaller constriction zone than the L3-formed constriction (Figure 45). The second N-terminus-formed constriction region is formed by residues N3, D7, E65, Y86, D104, K150, G108 and D112 (Figure 47). Surprisingly, the lining of the L3-formed constriction region of OmpU is of different composition compared to OmpK36. The presence of a large cluster of arginine residues (R30, R43, R45, R85, R133) dominates the lining of the constriction region (Figure 48), with additional arginines (R26, R287, R316) buried further down the pore towards periplasmic side. Looking from the extracellular side, the five-arginine cluster takes up about half of the circle lining of the constriction region and are positioned on the opposite side of the constriction-lining residues in L3 that constitute the other half of the circle. Another lone arginine residue lining the constriction region is R219 near periplasmic T6. The lining of these 8 of 11 total arginines in OmpU protein sequence marks the distinct pore properties and may carry crucial function in ion selectivity and channel conductance.

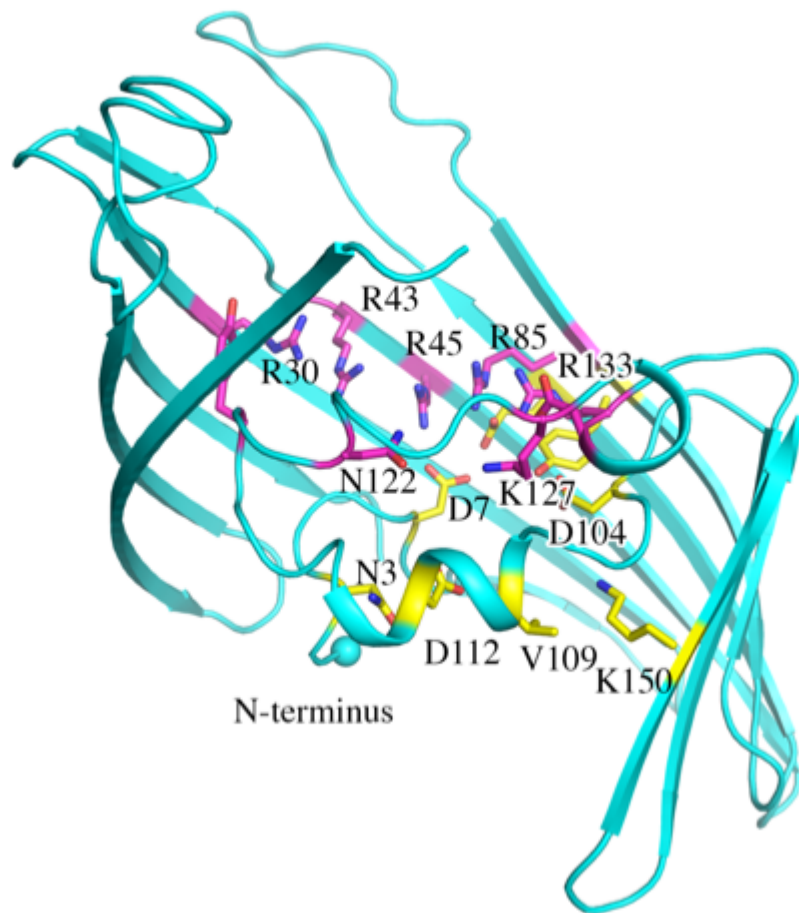


Figure 47. The two constriction zones of OmpU.

The L3-formed constriction is highly positively charged and the N-terminus-formed constriction is highly negatively charged. The OmpU protomer is colored in cyan. The residues in the constriction zone formed by the arginine cluster and L3 are colored in magenta, whereas the residues in the N-terminus-formed constriction are colored in yellow. The sphere denotes the N-terminus.

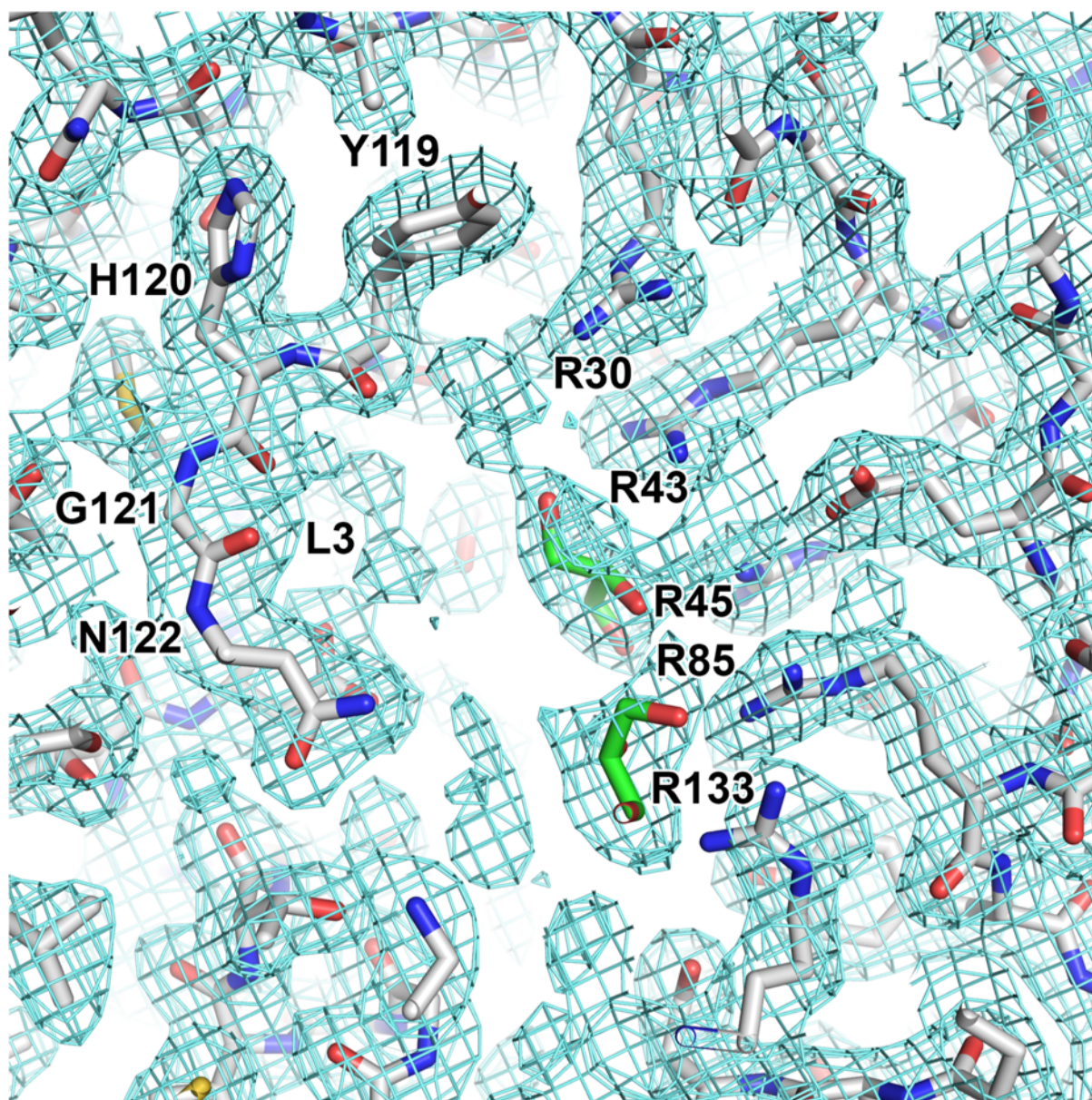


Figure 48. 2Fo-Fc Electron density map (contoured at 1σ) of the arginine cluster at the constriction lining.

Residues are shown in full with carbon atoms in grey, oxygen in red and nitrogen in blue.

Each of the arginine residues is labelled individually. The L3 constriction loop is below the horizontal plane of the figure. Two glycerol molecules in the pore are highlighted with carbon atoms in green.

The exact pore size of OmpU remains controversial, with report defining the effective radius to be 0.55 nm compared to 0.43 nm in OmpT [183], while the other stating OmpU may form smaller pore than OmpT [184]. The dimension of the pore is comparable to that of OmpK36, with ~ 4.7 Å of minimum radius directly measured in PyMOL. We sought to analyze the effect of the N-terminal coil on pore size and to determine more accurate pore dimensions. HOLE programme [185] was used to compute the three-dimensional visualization of the pore as well as to yield a two-dimensional graph of pore radius versus channel coordinates from native OmpU PDB file and N-terminus-deleted OmpU PDB file (Figure 49). The graph indicates that the minimum radius in native OmpU is ~ 3.1 Å, slightly smaller than 3.2 Å found in OmpU with deleted N-terminus. Furthermore, the graph indicates that the pore in native OmpU forms a little bulge (increase in diameter) near the centre region of the pore along the vertical axis, before further shrinking to the narrowest point. Although we have not been able to assign the coordinates in the three-dimensional structure due to the limitation of the programme, it is very likely that the N-terminal coil contributes to an additional narrowing of the pore and the smaller minimum radius. Moreover, the electrostatic maps generated by the two PDB files illustrate that the N-terminal coil indeed reduces the pore size (Figure 50).

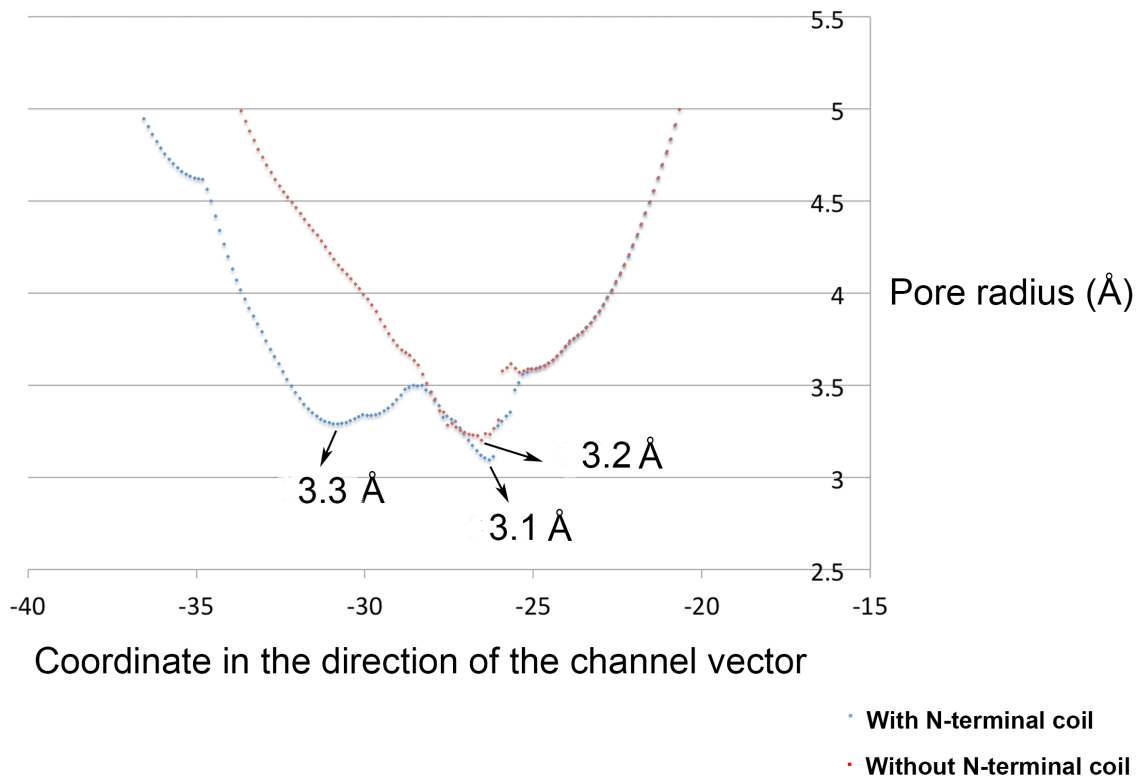


Figure 49. Pore radii plot against coordinates in the direction of the channel.

Blue dots represent the data obtained using protomeric OmpU structure with intact N-terminal coil, while red dots indicate the data from protomeric OmpU with N-terminal coil deleted. The two constriction regions of the pore are symbolized by the two narrowest points (3.3 Å and 3.1 Å). The coordinate of narrowest point in the structure without N-terminal coil (3.2 Å) corresponds well with the 3.1 Å point in the native structure.

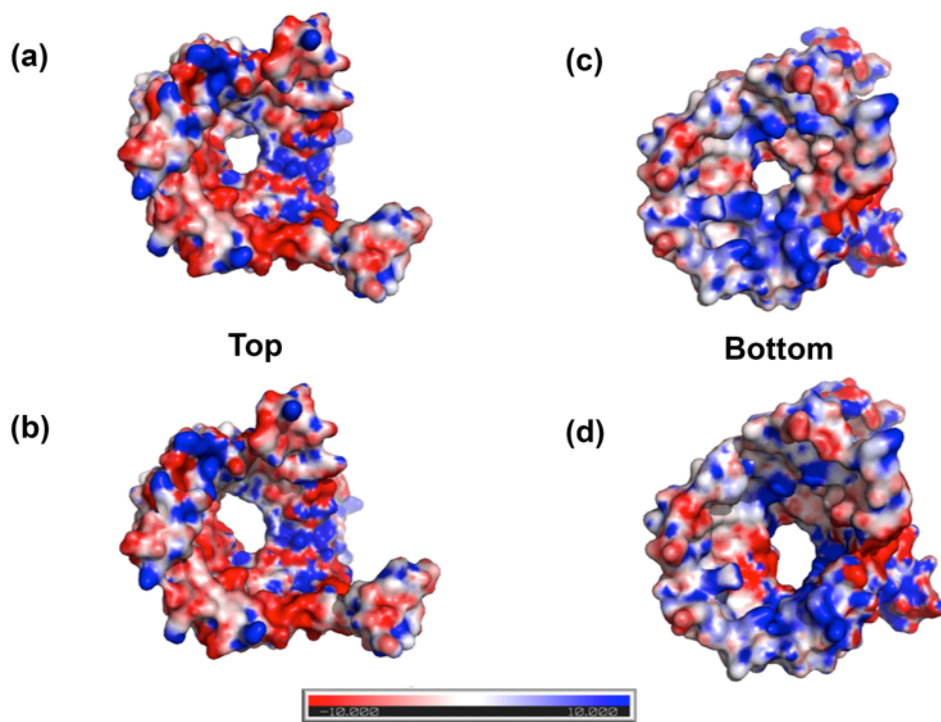


Figure 50. Electrostatic potential of OmpU model with (a and c) and without (b and d) the N-terminal coil.

The model is viewed from top extracellular side (a-b) and from bottom intracellular side (c-d).

The electronegative zone is presented in red (the most negatively charged), the neutral zone in white and the electropositive zone in blue (the most positively charged). All four diagrams share the same electrostatic scale.

CHAPTER 6

**Results: structural studies of lipoprotein N-acyl transferase in
Gram-negative bacteria**

6.1 *Vibrio cholerae* Lnt purification and crystallization

According to the expression tests, Lnt from the chosen species had various levels of overexpression. Furthermore, there was another outstanding issue reflecting on the thermostability of the proteins. Heated protein samples exhibited large area of unclear smear as well as indistinguishable faint bands above the target protein bands. If unheated, the unclear stain would be greatly reduced. It was therefore concluded that Lnt from most of the chosen species were not thermostable and not ideal for subsequent crystallization work. Among these species, however, Lnt from *Vibrio cholerae* O1 biovar El Tor str. N16961 showed least amount of nonspecific stain and was shown to have the highest level of expression (~2mg per litre of cell culture). The eluted protein sample produced a strong band on SDS-PAGE gel at ~ 45 kDa (Figure 51). The molecular weight of *V. cholerae* Lnt is ~ 57 kDa and the faster migration of IMPs like Lnt (gel shifting) is likely to be the result of partial unfolding by SDS [186]. Based on this, Lnt from *V. cholerae* was selected for large-scale purification in an attempt to perform crystallization.

Lnt showed very strict requirement for stabilizing detergent throughout the purification. During the detergent trials using commonly used detergents available in the lab, most were not able to render Lnt soluble and monodisperse in solution, either precipitating very quickly after detergent exchange from DDM or forming oligomeric aggregates as indicated by the early-eluted peaks in gel filtration step. In light of this, DDM was then used as the detergent for purification. Surprisingly, a minimal 10 CMC concentration of DDM was needed to produce a symmetrical peak in SEC at the expected volume, a lot more than usual for most other membrane proteins (Figure 52a).

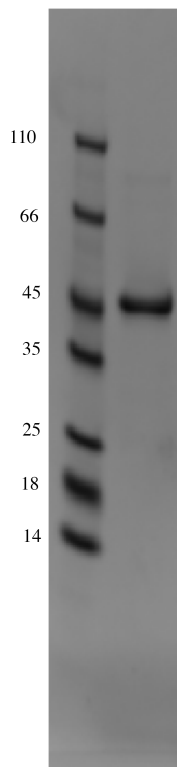
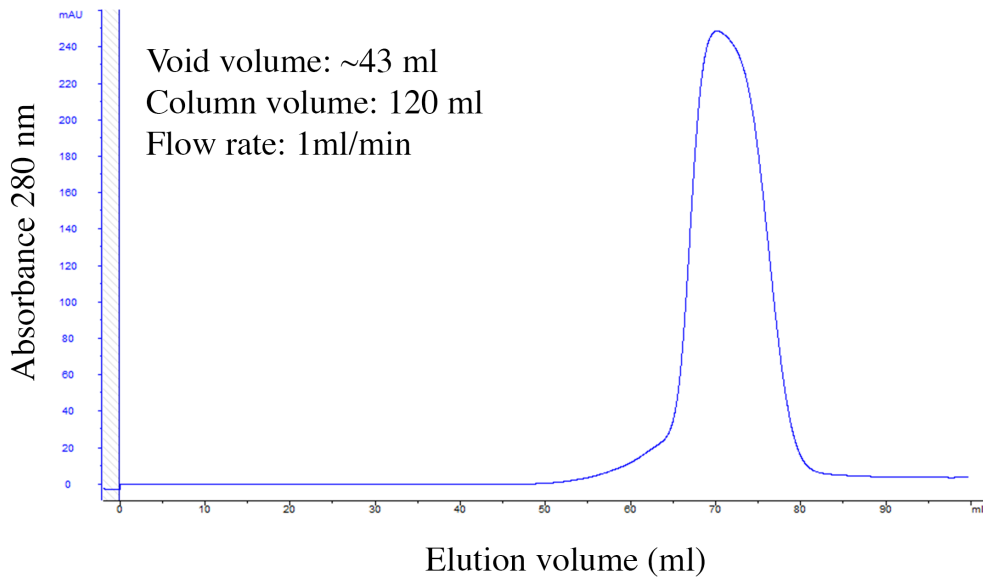
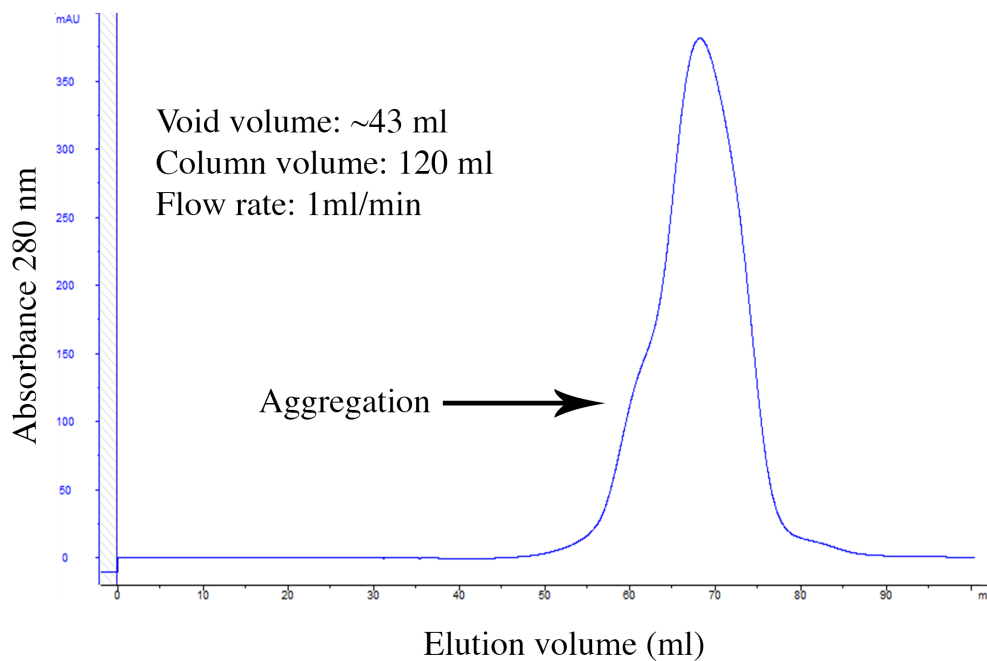


Figure 51. SDS-PAGE gel picture of *V. cholerae* Lnt from small-scale expression tests.

Lnt was highly purified. The lane on the left is the protein molecular weight marker with indicated molecular sizes in kDa.



(a)



(b)

Figure 52. Size exclusion chromatograms of *V. cholerae* Lnt. Samples were injected onto a HiLoad 16/600 Superdex 200 prep grade column (GE healthcare).

(a) Gel filtration chromatogram of *V. cholerae* Lnt purified with 0.1% DDM. (b) GF chromatogram of *V. cholerae* Lnt purified with 0.15% LysoFos Choline 12. A partial leading peak indicating larger aggregates is labelled.

Due to the limited diffraction quality of the crystals, several methods were employed aiming at improving resolution of diffraction of the crystal. Dehydration of the crystal by prolonging the time crystals stay in the mother liquor before harvesting, adding MemAdvantage additives, limited proteolysis, increasing crystallizing drop size of both protein and buffer solution were all attempted. In addition, co-crystallization of Lnt with its preferable phospholipid substrate phosphatidylethanolamine (PE) was also carried out but none of these methods proved to be helpful to further improve crystal diffraction.

6.2 96 detergent screening

Under the circumstances, there was very little one could manipulate the plasmid construct to optimize the current plasmid construct. Having realized the relatively unusual amount of DDM applied to stabilize Lnt led to the speculation that DDM is probably not perfectly suitable for solubilizing Lnt even though it was efficient for extracting Lnt from the whole membranes. Thus, a new round of search for ideal detergent candidates was performed. To facilitate efficient detergent preparation and exchange, the commercially available Analytical Selector kit was purchased from Anatrace. 96 detergent exchanges were then performed from initial stabilizing 0.1% DDM according to the instructions given in the manufacturer's manual, and the 96 exchanged eluents were analyzed by SDS-PAGE gels to eliminate those that are completely incompatible (signified by the absence of a protein band) and others that have limited compatibility (signified by a less intense protein band). Based on the results, five detergents were shortlisted for detergent exchange during large-scale purification: Cyclofos-4, Cyclofos-7, Fos-Choline-Unsat-11-10, Pentaethylene Glycol Monoethyl Ether (C₆E₅), LysoFos Choline 12 (Figure 53). Each of these detergents was further tested in large-scale purifications. LysoFos Choline 12 was found to be the best one as its presence produced a peak at expected retention volume in SEC procedure with an additional minor leading peak (Figure 52b). The purified proteins were crystallized and a number of hexagonal prism-

shaped crystals emerged from one condition within one week at 21°C. Upon diffraction test, however, these crystals showed very weak diffractions. The project was therefore suspended for reconsidering the methods.

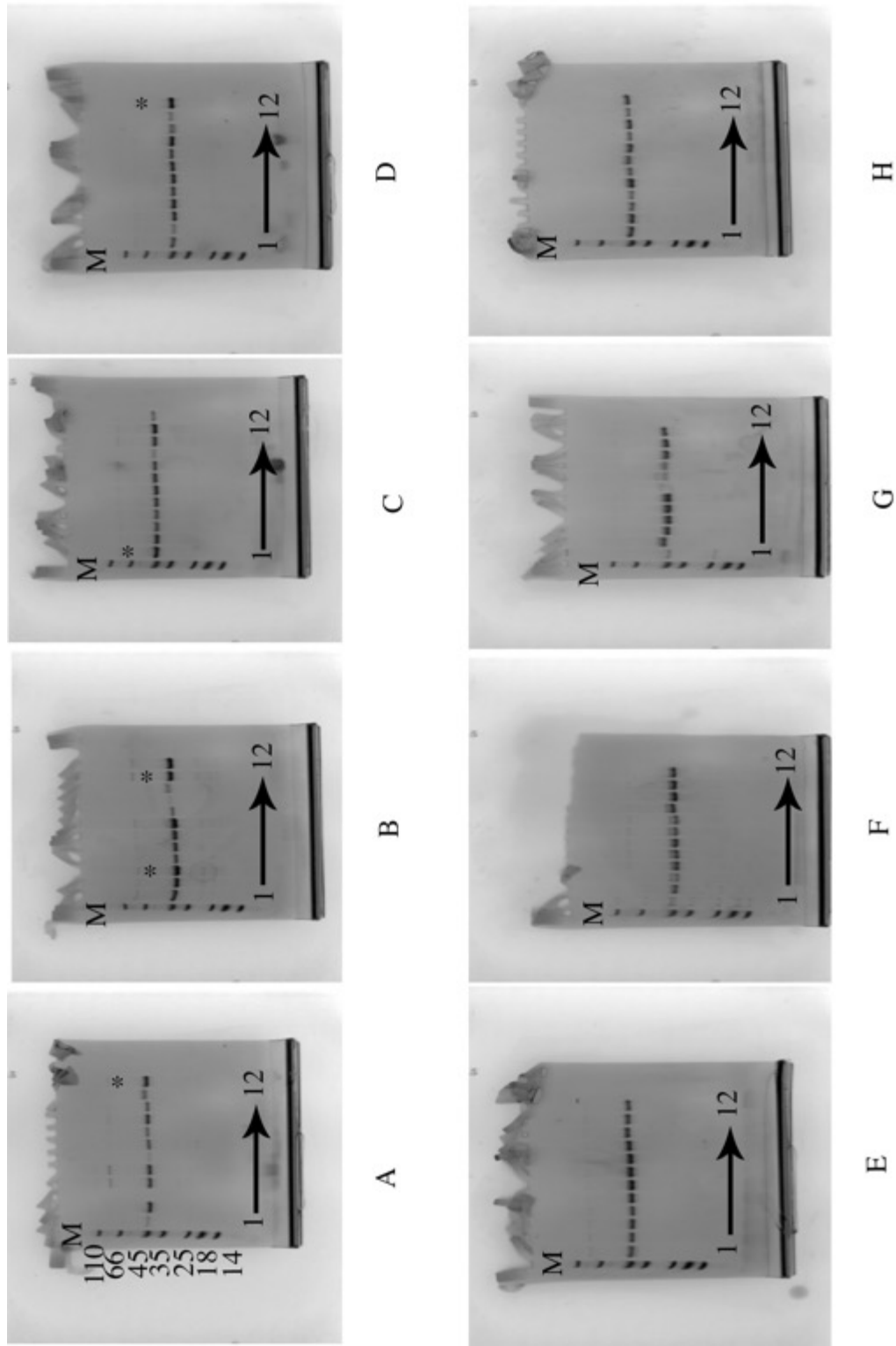


Figure 53. SDS-PAGE results of the 96-detergent screening for *V. cholerae* Lnt.

Gels are labelled with reference to the corresponding rows in the 96-well plate in the kit. Lanes of the five promising detergents are marked with *. The lane of protein marker in individual gel is labelled with 'M'. The molecular weights of protein marker are labelled in gel A, and the other gels share identical marker.

6.3 Published Lnt structures

While the project was being undertaken, the structures of Lnt from *E. coli* and *P. aeruginosa* were reported by other groups [187-189]. As predicted, the reported structures reveal two domains, one periplasmic nitrilase domain and one transmembrane domain (Figure 54). The nitrilase domain is composed of two six-stranded β -sheets and is of the four-layer $\alpha\beta\beta\alpha$ sandwich fold (Figure 54, Figure 55). In the nitrilase domain, the predicted catalytic triad E267-K335-C387 of *E. coli* Lnt resides in a pocket above the membrane plane and their overall conformation is strikingly similar to that of those known soluble nitrilases [188]. There are a number of loops in the front side of domain that extend away in the horizontal direction and form a ring-like structure around the catalytic triad (Figure 54). These loops are part of the scaffolding of the active site pocket and were suggested to be instrumental in guiding the entry and exit of both substrates and products [187]. At the interface between the two domains, interaction between a conserved residue Gly145 in the transmembrane domain and Tyr388 in the nitrilase domain are found to be critical for securing the two domains closely in place. It was further proposed that the strong interactions between Arg438 and Thr478 and Gly479, and between Thr481 and Glu435 primarily stabilize the catalytic triad in the nitrilase domain and, in the meantime, maintain the structural and functional integrity of the whole protein [187].

The transmembrane domain consists of eight transmembrane α -helices and is connected with the nitrilase domain at the boundary of the IM (Figure 54, Figure 55). The two domains of Lnt are connected by long loops L2 and L3, which also connect helix 7 and 8 in the transmembrane domain. Helix 3 and 4 stretch out from the membrane plane and extend into the periplasm and constitute another part of the active site pocket (Figure 54). Notably, helix 4 and 5 are not as closely associated as the other helices and the space between them creates a

cleft that point to the active site. It was therefore considered as the path for the molecules involved in the reaction [187].

Underscored by the structural information of Lnt, the comprehensive molecular mechanism of lipoprotein N-acylation was proposed (Figure 56). Consistent with the ping-pong mechanism, the reaction is carried out in two distinct steps. In the first step, E267, a general base, attacks the γ S of C387 by abstracting a hydrogen, which results in a thiolate in C387. The thiolate attacks the connecting ester bond between the acyl tail in the *sn*-1 acyl chain and the glycerol moiety of the phospholipid to form a thioester acyl-enzyme intermediate, accompanied by a net negative charge on the oxygen of the intermediate. The charge is neutralized and stabilized by the positively charged side chain of K335, a residue that is constantly charged in physiological environment. Elimination reaction disrupts the temporary conformation and releases the phospholipid in its lyso-form. This by-product leaves the active site while the diacylated prolipoprotein substrate enters to initiate the second step of the reaction. The entry of the lipoprotein substrate may be navigated by the S-diacylglyceryl group. Approach of the lipoprotein causes the α -amino group in the +1 cysteine of the lipoprotein to attack the carbonyl carbon in the thiolester group in C387 and simultaneously triggers E267 to abstract a hydrogen from the same amino group of the +1 cysteine. This also leads to a negative net charge on the oxygen attached to the intermediate, which is similarly stabilized by K335. A second disassembly takes place on the intermediate and results from an elimination reaction similar to the previous elimination step, generating matured triacylated lipoprotein and reset of the catalytic triad.

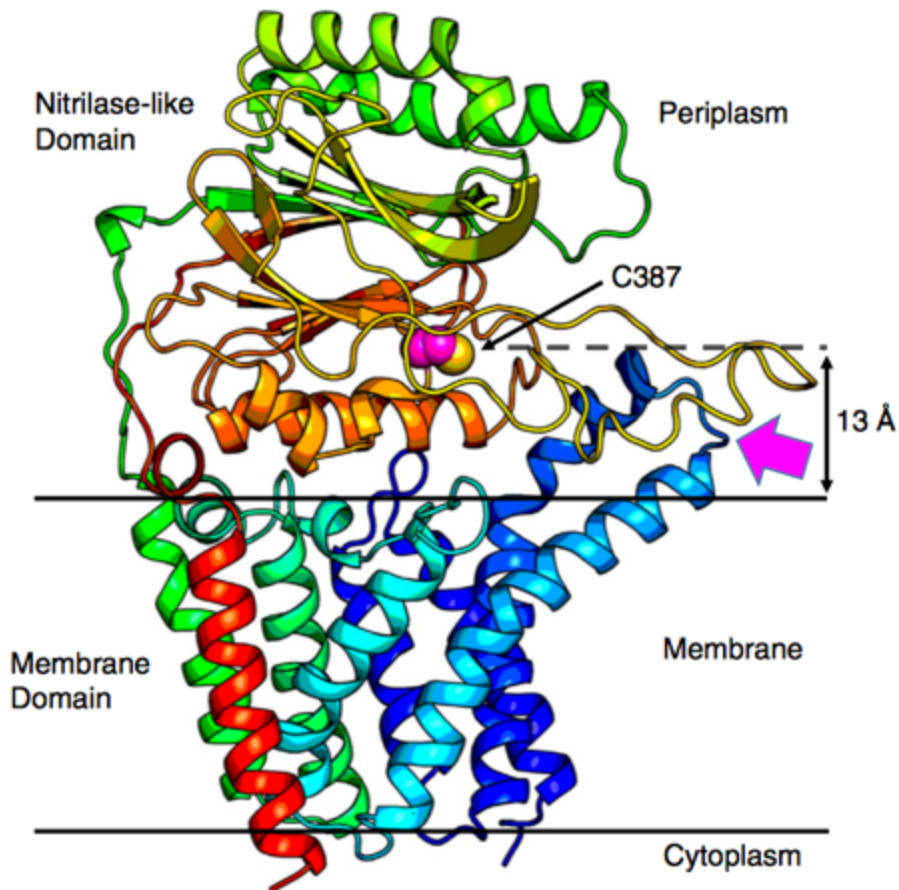


Figure 54. Cartoon representation of the structure of *E. coli* Lnt viewed from the horizontal membrane plane.

The N-terminal helix is coloured in blue and the C-terminal helix is coloured in red. One of the catalytic triad residues C387 is highlighted in sphere representation, with the carbon atoms coloured in magenta and sulfur atom coloured in yellow. The magenta arrow points to the suggested substrate entry gateway. (Figure taken from [187])

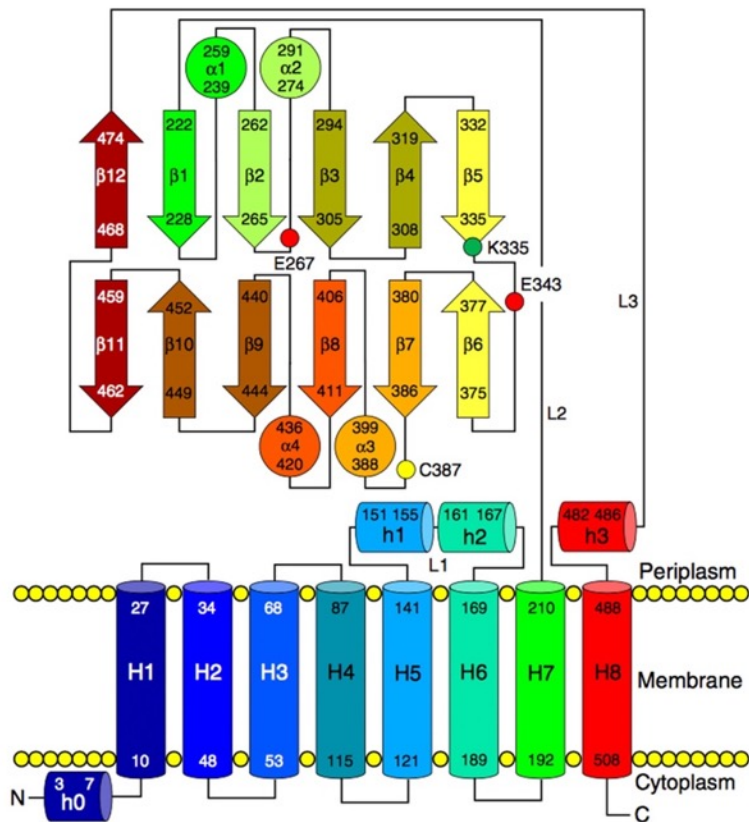


Figure 55. A diagram of the secondary structure segments in Lnt.

The same colours were adopted for corresponding segments as in Figure 51.

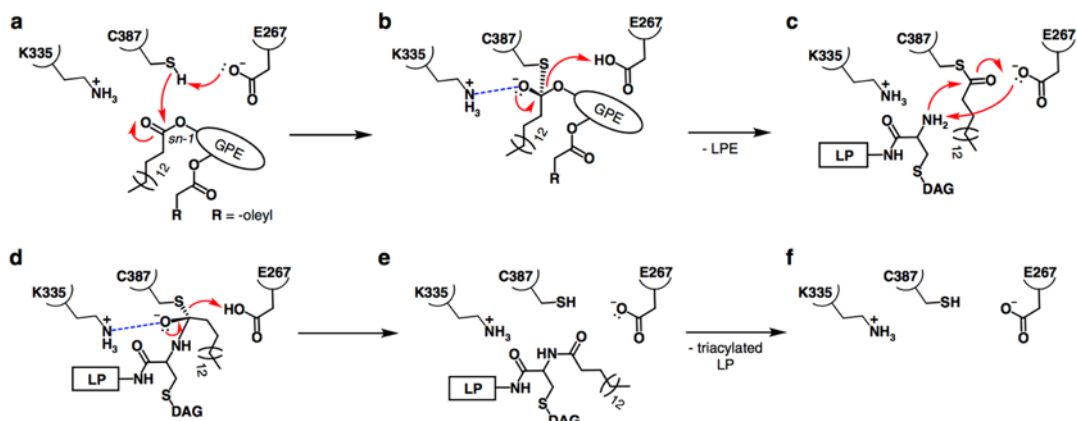


Figure 56. Proposed Lnt N-acylation mechanism dissected into six steps.

(a) The reactions involved in the first auto-acylation step of the ping-pong mechanism. (b) The formation of the thioester acyl-enzyme intermediate and the subsequent elimination reactions. (c) The reactions involved in the second step of N-acylation in the ping-pong mechanism. (d) The formation of the second tetrahedral intermediate and the subsequent elimination reactions. (e) The released triacylated lipoprotein product. (f) The liberated catalytic triad that is reset for another reaction cycle. Electron lone pairs are represented by dot pairs. Electron transfer is represented by red curved arrows. Oxyanion stabilization is illustrated by dashed blue lines. GPE, glyceryl-phosphoethanolamine; LP, apo-lipoprotein; DAG, diacylglycerol; LPE, lyso-PE. (Figure taken from [187])

CHAPTER 7

Discussion

7.1 BAM complex

Since the biochemical characterization of BamA more than a decade ago, subsequent studies have provided molecular evidence of a coordinated multi-subunit protein complex with which BamA is associated in the process of OMP biogenesis. Further structural studies of both single protein components and sub-complexes of the BAM machinery further shed light on the molecular identity and partial organization of the subunits. Until the last few years the atomic structure of a subcomplex containing more than three protein subunits and the entire five-protein BAM complex had been unavailable. As a result, the detailed interactions between each subunit involved in complex formation and molecular mechanism of OMP folding and insertion into the OM was poorly understood. The long-sought crystal structures of the BAM complex from *E. coli* solved in this project, along with the other reported structures in the same year [[111](#), [190](#), [191](#)], promoted a leap towards the complete dissection of this essential apparatus.

In this work, the BAM complex was initially overexpressed separately using BamAB and BamCDE, and subsequently reconstituted according to previous established protocol [[77](#)], but without producing well-diffracting crystals. In light of this, the reconstitution method was further improved by overexpressing all five protein components in a single plasmid [[170](#)], and the resultant purified complex was in stoichiometric quantities of BamA, BamB, BamC, BamD and BamE. This in turn generated good-quality crystals for structure determination. Although the initial separate plasmid constructions in previous reports was able to reconstruct the BAM complex in 1:1:1:1 ratio of BamA:B:C:D *in vitro* and, in a biochemical study, catalyze more than one batch of OMP assembly, indicative of at least partial functional state [[192](#)], the failure of the identically reconstituted complexes in current work to produce high-resolution crystals may be attributed to the heterogeneity of the purified complexes. Indeed, separate purification of the two subcomplexes is likely to give rise to multiple combinations

in the five protein components and different subcomplexes in addition to the desired architecture of the whole BAM complex that may be indistinguishable to size-exclusion chromatography, as evidenced by the nearly identical size-exclusion chromatography profiles between the purified BAM complexes from pJH114 and pYG120. Overexpression of all the components in a single construct was demonstrated in this study to eliminate this heterogeneity and facilitate formation of BAM complexes in homogeneous forms, i.e., each of the purified complexes contains five protein components. But why did stoichiometry of the BAM complex from separate purification look identical to that from pYG120? Firstly, visual inspection of SDS-PAGE gel results by naked eyes can be inaccurate and the results can therefore be misinterpreted. Techniques such as densitometry should be applied to determine the optical densities of individual bands on the gels in order to accurately distinguish the concentration differences between subunits and analyze the stoichiometry of the complex. Secondly, considering mixed compositions of various complexes and subcomplexes, the disparities in composition of individual complex/subcomplex is very likely to be concealed by the presence of other entities in the purified solution. This could in turn produce an averaged pattern of composition upon SDS-PAGE analysis. On account of the structures solved and reported elsewhere [[111](#), [190](#)], the common feature of gene cloning and plasmid construction is the molecular cloning of all five subunits into a single vector for subsequent overexpression. Taken together, the cloning strategy of incorporating all subunit-encoding genes into a single expression vector is the key to success and to be recommended for future structural studies of protein complexes with multiple subunits.

In the BamACDE crystal structure, BamB is absent although it is present in the purified complex. SDS-PAGE analysis of BamACDE crystals confirmed BamB's absence (data not shown), indicating dissociation or degradation during crystallization. In the BamABCDE structure, although BamC is present in both the purified complex and in the crystals via SDS-

PAGE analysis (data not shown), it is not in full length but only the N-terminal unstructured loop (Val35-Pro88). Although the two determined crystal structures provided sound evidence of direct participation of BamC in formation of the BAM complex, *in vivo* native membrane environment is fundamentally different to the detergent solubilized conditions in which the crystals are grown, and the presence of potential OMP substrates, absent in the two structures, may be an additional factor mediating the overall topology of the machinery as well as localization of BamC in physiological environments. The remaining portion of BamC is therefore considered highly flexible, and one of the possible explanations of its absence could be surface exposure of this lipoprotein [193, 194]. Previous protease sensitivity experiments revealed that the N-terminal region containing the unstructured loop that remains attached to the BAM complex is found to be more vulnerable to protease digestion [195, 196], and treatment with trypsin on cells overexpressing BamC resulted in two fragments of BamC corresponding to the two helix-grip domains, providing evidence of surface exposure of these two domains *in vivo*. It was also demonstrated that one of the three conserved segments in the N-terminal loop of BamC is essential for interaction with BamD [193]. It is hence conceivable that the two surface-exposed domains of BamC and the periplasmic N-terminal loop are linked by a portion of BamC that can serve as a putative transmembrane segment in the surface exposure scheme. The absence of C-terminal domain of BamC observed in BamABCDE structure is likely to support the scheme of BamC surface exposure, and future studies such as dynamic protein-protein interactions in the OM, in conjunction with structural and functional studies on surface-exposed lipoproteins are needed to provide detailed mechanistic insights into BamC secretion and its implications in the structure and functions of the BAM complex.

Assuming that both of the determined structures are physiological in native environments, these structures assist understanding of nascent OMP folding and insertion via the contrasting features of the BamACDE structure that the POTRA 5 is positioned directly underneath the β -barrel of BamA, rejecting access of substrates, and that β 1C and β 16C in the barrel are not tightly linked by hydrogen bonds, potentially promoting a lateral opening state. Consequently, the molecular mechanism of OMP insertion can be deduced. A 30° rotation of the periplasmic ring apparatus discovered from the comparison of the two structures gives rise to direct interaction with the lipid head groups of the OM and concomitantly generates the 65° tilting of β 1C- β 6C strands and temporary lateral opening of the BamA barrel between β 1C and β 16C. This subsequently exposes the barrel lumen to the external settings of the OM and promotes membrane instability that paves the way for OMP insertion. The BamABCDE structure resembles an initial substrate-loading state in which the periplasmic mouth is open for substrate approaching and the β -barrel is in the closed and inward-open conformation awaiting laterally opening upon substrate engagement. On the other hand, resembling a substrate-releasing phase in which the nascent OMP has been loaded on the complex and already inserted into the OM through the laterally open barrel, the BamACDE structure may represent the ending phase of the catalytic cycle and give hint on the exact *in vivo* mechanism of OMP assembly. Of the two possible models of insertion mechanism described previously (Figure 10), these structural features of the BAM complex provide evidence in support of the BamA budding model, but with exceptions. In a recent study, LptD/E complex, an essential and large OMP machinery responsible for lipopolysaccharide translocation to the cell surface [197-199], was chosen to study BAM-mediated OMP assembly *in vivo* [200]. LptD contains a large β -barrel in which LptE is plugged. Specific mutations introduced in *lptD* gene slowed barrel assembly of LptD and a late-stage partially folded LptD/E intermediate was found to be trapped on the BAM complex amid folding, with both LptD and LptE interacting with

BamA and BamD. Due to the fact that both BamD and LptE are soluble periplasmic proteins and the possibility that BamA may not spatially accommodate this partially folded yet considerably large substrate for subsequent lateral release into the OM, it was proposed that folding of this late-stage unclosed barrel of LptD initiated in the periplasm in agreement with the BamA-assisted model in which folding of OMP substrates begins outside the OM, and, independent of the β -barrel domain of BamA, the folding would complete when the emerging barrel is closed and LptD/E complex is ultimately dissociated from the BAM complex, which disagrees with the BamA budding model.

Despite the two determined structures and their molecular details of the BAM complex, exactly how the complex functions *in vivo* during OMP biogenesis remains unknown. A number of representative ones include whether the OMP substrate is fully folded or partially folded when escorted to the BAM complex, how the substrate-chaperone complex interacts with the BAM complex in detail, whether the BAM complex carries out insertion only or promote both substrate folding and insertion, and, if so, whether it performs the two functions simultaneously or sequentially. Furthermore, structural studies of the BAM complex in the presence of detergents and in the absence of the substrate may not reflect *in vivo* mechanism and, along with the incomplete understanding of the interactions of each component of the complex with the substrate and concomitant conformational changes in both the substrate and the entire complex, provide rather limited insights into the highly complex assembly process. Future studies, however, can be carried out on the basis of current work and should include structural studies of the BAM/OMP intermediates in distinct stages during OMP assembly and biochemical examinations of these intermediates in whole cells in order to fully dissect the mechanism of the BAM complex.

7.2 OmpU

OmpU porins are increasingly recognized as one of the crucial determinants of *Vibrio* pathogen and host interactions [119, 120, 127]. The crystal structure of OmpU trimer was successfully determined, and showed two defining features that can differentiate OmpU from other structurally related porins. OmpU possesses an additional N-terminal loop consisting of G1 – S11 that extends into the pore at the periplasmic side and forms a second constriction with the constriction loop L3. In addition, the L4 loop at the extracellular side exhibits a signature ‘pole’ standing and protrudes further into the extracellular space.

OmpU has long been proposed as a potential virulence factor involved in pathogen-host interactions during infection, being capable of attracting physical contact and adhering to host cells as well as triggering subsequent invasion by the pathogen. Given its abundance in the outer membrane of *V. cholerae* and implications in adhesion and invasion, experiments that produced $\Delta ompU$ knock-out strains showed reduced ability to express virulence factors and colonize the intestine [118]. From a structural perspective, the extracellular loop L4 may well be considered as a potent binding promoter that interacts with various external proteins with complementary or matching β -sheets, or acting like a hook, attracts binding with receptors on the surface of other cells. One binding scenario was previously reported for OmpX, in which the protruding single-layer β -sheet and connecting loops expose the edge of the β -sheet at L3 and this edge may function as a ‘fishing rod’ to attract potential hydrogen bonding partner [201]. Likewise, the projected β -hairpin like segment of the pole-shaped L4 in OmpU could expose side edges of both strands (β -augmentation) to capture external binding proteins in a possibly more efficient way compared to OmpX. Moreover, a recent study confirmed that OmpU is the receptor of the predating ICP2 species of *V. cholerae*-specific and virulent

podoviruses by interacting with the extracellular loops, leading to cell death [202]. Mutation in residues located in mainly L3, L4 and L8 was reported to neutralize infection and acquire phage resistance. This further emphasizes the essential binding capability of L4 and L8 in initial phase of contact with foreign organisms.

The N-terminus in bacterial porins is normally located in the periplasmic side of the barrel wall and in close contact with C-terminus, in some cases forming salt bridge in between. Both of the termini are located outside of the pore space. However, the non-canonical N-terminus coil found in OmpU structure folds inward to constrict the pore lumen. The actual functional role of this additional coil is unknown, but from the pore dimension analysis done in present study, it is not surprising that the coil may function as a narrowing factor of pore size regulation. Specifically, the pore dimension of OmpU was calculated by the HOLE programme and, in both cases of WT OmpU and OmpU with no N-terminal coil, the minimum radii were comparable to that of other non-specific porins [203]. Nevertheless, the result of the smaller constriction region formed with the N-terminal coil calculated by the HOLE programme together with the same observed trend from electrostatic potential map converges to the conclusion that the N-terminal coil reduces pore size. Moreover, it has been reported that OmpU is capable of excluding the entry of bile salt deoxycholate and thus crucial for cell survival in the intestine [126, 204]. Consistent with the described function, the combination of reported cation selectivity with the observed smaller constriction region of OmpU formed by L3 accompanied by the N-terminal coil is very likely to inhibit the translocation of the bile acid.

Another important function of the coil is its presence in the additional constriction region. Reports in the literature have recognized the constriction loop L3 as a vital filter of pore permeability due to negative net charge and reduction in pore size, and L3 requires molecules to re-orient in order to translocate [205]. Concerning the stable nature (Figure 46) and the residue charges of the coil, it is plausible that it could function as an assistant helper filter in addition to the L3 loop in the additional constriction region throughout the diffusion activity of hydrophilic solutes across the porin. More functional and biochemical studies are required to fully understand the implications of the coil in the function of the trimer and whether it is important for virulence.

It has been reported that each OmpU protomer associates with 3 to 4 calcium ions in order to maintain intact β -sheeted conformation and functional form of the trimer [124]. However, in the protomer, no clear density was found that could be unambiguously assigned to calcium regardless of constant presence of calcium ions during purification and in the final protein solution to be crystallized. Selected potential sites were tested by CheckMyMetal tool [206] but none of them were geometrically acceptable. Therefore, the determined structure of OmpU without obvious metal ions in the structure is very likely to refute the previous statement.

7.3 Lnt

Structural studies were carried out for Lnt from a number of Gram-negative bacterial species but were unsuccessful in determining the crystal structure. Examining the recently published structures and corresponding methods, it is apparent that Lnt is more easily crystallized using the LCP method with DDM or LMNG in final purification as well as the vapour diffusion methods with NTM (n-Nonyl- β -D-Thiomaltopyranoside) in final purification and addition of n-Heptyl- β -D-thioglucopyranoside and CHAPSO in crystallization. Due to the immature development of the LCP method in the lab over the past years and the dominant usage of the vapour diffusion method, crystallization of Lnt in LCP was not applicable. Considering the vapour diffusion method, very limited detergent combinations were chosen and performed, which did not yield crystal formation. Detergent screening of the common detergent choices used in IMP structural studies such as DDM, DM, Fos Choline 12, Lauryl Maltose Neopentyl Glycol (LMNG) and NG should be prioritized to find the optimal solubilizing conditions for other IMPs in future studies.

The genome of *E. coli* encodes at least 90 lipoprotein types [207], and the molecular sizes range from 50 to more than 900 residues. A large portion of the lipoproteins but not all are the substrate of Lnt, suggesting a tolerant and universal recognition mechanism that neglects lipoprotein identity and relies significantly on the Lgt-ligated diacylglyceryl group attached to the +1 cysteine. The determined structural feature of Lnt that there is a cleft that can accommodate lipid molecules and leads to the active site is consistent with the argument. FSL-1-Biotin (FSL-1) in a study, for example, was used to test if Lnt can recognize and recruit it to be a potential substrate [167]. FSL-1 is a synthetic peptide designed on the basis of a bacterial lipoprotein and possesses dipalmitoyl-glyceryl on its N-terminal cysteine. Resembling the +1 cysteine of a prolipoprotein, the peptide was processed by Lnt as a substrate, which further clarifies the notion that the diacylglyceryl group on the +1 cysteine is

a key determinant of substrate recognition of Lnt, regardless of the type of the protein the group is linked to. Nevertheless, the reason behind the evasion of N-acylation for some diacylated mature lipoproteins is not clear. In the Lnt structures, the proposed opening for substrate entry and exit is not spacious enough to accommodate both the diacylglyceryl group and portions of the protein substrate, and if the protein substrate is completely folded from the +1 cysteine to the protein entity, it would be impossible for Lnt to adjust its conformation to provide adequate space. It is therefore plausible to propose that the tether domain in prolipoproteins provides a flexible linkage to the N-terminal cysteine and extensive space between it and the folded protein domains, promoting easier access of the residue to the active site. In addition, the tether domain can serve as an inherent filter for substrate engagement and evasion of reaction in that only the substrates that have unfolded and reasonably long tethers would be N-acylated by Lnt.

E. coli is the most widely studied model organism in the phylum of proteobacteria, and the crystal structures of Lnt in *E. coli* represents a typical protein model that aids in the understanding of the molecular mechanism of N-acylation in lipoprotein post-translational modification pathway in Gram-negative bacteria. It is noteworthy that in other species of actinobacteria in which Lnt is also present, the functional profile of Lnt is different mainly as a result of different compositions of phospholipids and genomic constructions. While *E. coli* Lnt specifies the *sn*-1 position of *E. coli* phospholipids for catalysis, phospholipids of mycobacteria are composed of octadecanoic acid and tuberculostearic acid at the *sn*-1 position and palmitate at the *sn*-2 position [208] and earlier studies found that Lnt in mycobacteria conducts transfer of palmitate or tuberculostearic acid from phospholipids to apolipoprotein [209], suggesting a strikingly different substrate preference of Lnt that can transfer fatty acids from both the *sn*-1 and the *sn*-2 positions of phospholipids in these organisms. Since no high-resolution structural model has been obtained for Lnt in

mycobacteria, the molecular basis of this substrate ambiguity is not clear although the enzyme was predicted to share similar membrane topology with *E. coli* Lnt [210]. Moreover, in *Streptomyces*, the largest genus of Actinobacteria, there are two identified genes each encoding an individual Lnt protein [211]. Mutagenesis of these two enzymes in *S. scabies* demonstrated that completely switching off *lnt1*, the gene encoding one of the two Lnt enzymes, Lnt1, resulted in diacylated form of lipoproteins, which implied the essential requirement of Lnt1 in the pathway. Switching off *lnt2*, on the other hand, caused partial defects in the N-acylation step and produced a mixture of di- and triacylated forms of lipoprotein, and it was suggested that Lnt2 serves as a supportive player enhancing the efficiency of the N-acylation reactions [211]. Further structural and functional studies are required to gain more insights into the last step of lipoprotein modification in these species.

In low GC content Gram-positive bacteria (Firmicutes), Lnt homologs are not found to process apolipoproteins. All the mature lipoproteins had been regarded as diacylated due to a lack of *lnt* orthologue and prevalence of diacylated lipoproteins in former studies [212-215]. However, recent biochemical analysis confirmed the existence of triacylated form in *S. aureus* [216, 217]. The structural evidence of these N-acylated lipoproteins suggested the existence of an unidentified Lnt-like enzyme. The observation that nearly 50% of the triacylated lipoproteins carry the 18 : 0 fatty acids bound to the diacylated conserved cysteine in *S. aureus* led to the indication that the unidentified Lnt candidate favors transfer of the 18 : 0 fatty acids and that the *sn*-1 position in phospholipids can be the primary site of reaction for the putative Lnt since only the *sn*-1 position possesses 16 : 0 to 20 : 0 fatty acids in phospholipids in *S. aureus*, whereas the *sn*-2 position contains a 15 : 0-branched fatty acid [218, 219]. The exact gene location in the chromosome encoding this putative candidate remains to be identified, and, once revealed, it would be intriguing for follow-on structural and functional experiments to fully characterize this functional homolog of *E. coli* Lnt.

REFERENCES

1. Chayen, N.E. and E. Saridakis, *Protein crystallization: from purified protein to diffraction-quality crystal*. Nature methods, 2008. **5**(2): p. 147-153.
2. Blow, D., *Outline of Crystallography for Biologists*. 2002: Oxford University Press.
3. Wlodawer, A., et al., *Protein crystallography for non - crystallographers, or how to get the best (but not more) from published macromolecular structures*. The FEBS journal, 2008. **275**(1): p. 1-21.
4. Rhodes, G., *Crystallography Made Crystal Clear: A Guide for Users of Macromolecular Models* Third Edition ed. 2006: Elsevier Inc.
5. Rhodes, G., *Crystallography Made Crystal Clear: A Guide for Users of Macromolecular Models*. Third Edition ed. 2006: Elsevier Inc.
6. Smyth, M. and J. Martin, *x Ray crystallography*. Molecular Pathology, 2000. **53**(1): p. 8.
7. Crystallography, I.U.o., *International tables for X-ray crystallography*. Vol. 2. 1959: Kynock Press.
8. Wikipedia. *X-ray crystallography*. 2017 [cited 2017 18/10]; Available from: https://en.wikipedia.org/wiki/X-ray_crystallography.
9. Battye, T.G.G., et al., *iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM*. Acta Crystallographica Section D: Biological Crystallography, 2011. **67**(4): p. 271-281.
10. Otwinowski, Z. and W. Minor, *Processing of X-ray diffraction data collected in oscillation mode*. *Macromol Crystallogr Part A* 276: 307–326. 1997.
11. Evans, P.R., *An introduction to data reduction: space-group determination, scaling and intensity statistics*. Acta Crystallographica Section D: Biological Crystallography, 2011. **67**(4): p. 282-292.
12. Carter, C.W. and R.M. Sweet, *Macromolecular crystallography*. 1997: Gulf Professional Publishing.
13. Abergel, C., *Molecular replacement: tricks and treats*. Acta Crystallographica Section D: Biological Crystallography, 2013. **69**(11): p. 2167-2173.
14. Holm, L. and P.i. Rosenstrij ½m, *Dali server: conservation mapping in 3D*. Nucleic acids research, 2010. **38**(suppl_2): p. W545-W549.
15. Kelley, L.A., et al., *The Phyre2 web portal for protein modeling, prediction and analysis*. Nature protocols, 2015. **10**(6): p. 845-858.
16. Cowtan, K., *The Buccaneer software for automated model building. 1. Tracing protein chains*. Acta Crystallographica Section D: Biological Crystallography, 2006. **62**(9): p. 1002-1011.
17. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics*. Acta Crystallographica Section D: Biological Crystallography, 2004. **60**(12): p. 2126-2132.
18. Murshudov, G.N., et al., *REFMAC5 for the refinement of macromolecular crystal structures*. Acta Crystallographica Section D: Biological Crystallography, 2011. **67**(4): p. 355-367.
19. Karplus, P.A. and K. Diederichs, *Linking crystallographic model and data quality*. Science, 2012. **336**(6084): p. 1030-1033.
20. Brown, L., et al., *Through the wall: extracellular vesicles in Gram-positive bacteria, mycobacteria and fungi*. Nature Reviews. Microbiology, 2015. **13**(10): p. 620.

21. Vance, J.E. and D.E. Vance, *Biochemistry of lipids, lipoproteins and membranes*. 2008: Elsevier.
22. Zähringer, U., B. Lindner, and E.T. Rietschel, *Molecular structure of lipid A, the endotoxic center of bacterial lipopolysaccharides*. *Advances in carbohydrate chemistry and biochemistry*, 1993. **50**: p. 211-276.
23. Alenghat, F.J. and D.E. Golan, *Membrane protein dynamics and functional implications in mammalian cells*. *Current topics in membranes*, 2013. **72**: p. 89.
24. Salton, M., *Bacterial membrane proteins*. *Microbiological sciences*, 1987. **4**(4): p. 100-105.
25. Granseth, E., et al., *Experimentally constrained topology models for 51,208 bacterial inner membrane proteins*. *Journal of molecular biology*, 2005. **352**(3): p. 489-494.
26. Rapp, M., et al., *Experimentally based topology models for E. coli inner membrane proteins*. *Protein Science*, 2004. **13**(4): p. 937-945.
27. Fairman, J.W., N. Noinaj, and S.K. Buchanan, *The structural biology of beta-barrel membrane proteins: a summary of recent reports*. *Curr Opin Struct Biol*, 2011. **21**(4): p. 523-31.
28. Nikaido, H., *Molecular basis of bacterial outer membrane permeability revisited*. *Microbiology and molecular biology reviews*, 2003. **67**(4): p. 593-656.
29. Chua, N.-H. and G.W. Schmidt, *Transport of proteins into mitochondria and chloroplasts*. *The Journal of cell biology*, 1979. **81**(3): p. 461-483.
30. Hofmann, N.R. and S.M. Theg, *Chloroplast outer membrane protein targeting and insertion*. *Trends in plant science*, 2005. **10**(9): p. 450-457.
31. Wiedemann, N., et al., *Machinery for protein sorting and assembly in the mitochondrial outer membrane*. *Nature*, 2003. **424**(6948): p. 565-571.
32. Chaturvedi, D. and R. Mahalakshmi, *Transmembrane β -barrels: Evolution, folding and energetics*. *Biochimica et Biophysica Acta (BBA)-Biomembranes*, 2017.
33. Remmert, M., et al., *Evolution of outer membrane β -barrels from an ancestral $\beta\beta$ hairpin*. *Molecular biology and evolution*, 2010. **27**(6): p. 1348-1358.
34. Koebnik, R., K.P. Locher, and P. Van Gelder, *Structure and function of bacterial outer membrane proteins: barrels in a nutshell*. *Molecular microbiology*, 2000. **37**(2): p. 239-253.
35. Killian, J.A. and G. von Heijne, *How proteins adapt to a membrane-water interface*. *Trends in biochemical sciences*, 2000. **25**(9): p. 429-434.
36. Harvey Lodish, D.B., Arnold Berk, S. Lawrence Zipursky, Paul Matsudaira, James Darnell, *Molecular Cell Biology*. Third ed. 1995: Scientific American Books, Inc.
37. Gierasch, L.M., *Signal sequences*. *Biochemistry*, 1989. **28**(3): p. 923-930.
38. Hegde, R.S. and H.D. Bernstein, *The surprising complexity of signal sequences*. *Trends in biochemical sciences*, 2006. **31**(10): p. 563-571.
39. Hoffmann, A., et al., *Concerted action of the ribosome and the associated chaperone trigger factor confines nascent polypeptide folding*. *Molecular cell*, 2012. **48**(1): p. 63-74.
40. Hesterkamp, T., et al., *Escherichia coli trigger factor is a prolyl isomerase that associates with nascent polypeptide chains*. *Proceedings of the National Academy of Sciences*, 1996. **93**(9): p. 4437-4441.
41. Lakshmipathy, S.K., et al., *Identification of nascent chain interaction sites on trigger factor*. *Journal of Biological Chemistry*, 2007. **282**(16): p. 12186-12193.
42. Economou, A. and W. Wickner, *SecA promotes preprotein translocation by undergoing ATP-driven cycles of membrane insertion and deinsertion*. *Cell*, 1994. **78**(5): p. 835-843.

43. Fekkes, P., C. van der Does, and A.J. Driessen, *The molecular chaperone SecB is released from the carboxy - terminus of SecA during initiation of precursor protein translocation*. The EMBO Journal, 1997. **16**(20): p. 6105-6113.
44. Schiebel, E., et al., *$\Delta\mu H^+$ and ATP function at different steps of the catalytic cycle of preprotein translocase*. Cell, 1991. **64**(5): p. 927-939.
45. van der Wolk, J.P., J.G. de Wit, and A.J. Driessen, *The catalytic cycle of the Escherichia coli SecA ATPase comprises two distinct preprotein translocation events*. The EMBO Journal, 1997. **16**(24): p. 7297-7304.
46. De Keyzer, J., C. Van Der Does, and A. Driessen, *The bacterial translocase: a dynamic protein channel complex*. Cellular and Molecular Life Sciences CMLS, 2003. **60**(10): p. 2034-2052.
47. Papanikou, E., S. Karamanou, and A. Economou, *Bacterial protein secretion through the translocase nanomachine*. Nature Reviews Microbiology, 2007. **5**(11): p. 839-851.
48. Veenendaal, A.K., C. van der Does, and A.J. Driessen, *The protein-conducting channel SecYEG*. Biochimica et Biophysica Acta (BBA)-Molecular Cell Research, 2004. **1694**(1): p. 81-95.
49. Prinz, A., et al., *Evolutionarily conserved binding of ribosomes to the translocation channel via the large ribosomal RNA*. The EMBO journal, 2000. **19**(8): p. 1900-1906.
50. Paetzel, M., et al., *Signal peptidases*. Chemical reviews, 2002. **102**(12): p. 4549-4580.
51. Sklar, J.G., et al., *Defining the roles of the periplasmic chaperones SurA, Skp, and DegP in Escherichia coli*. Genes & development, 2007. **21**(19): p. 2473-2484.
52. Krojer, T., et al., *Structural basis for the regulated protease and chaperone function of DegP*. Nature, 2008. **453**(7197): p. 885-890.
53. Spiess, C., A. Beil, and M. Ehrmann, *A temperature-dependent switch from chaperone to protease in a widely conserved heat shock protein*. Cell, 1999. **97**(3): p. 339-347.
54. Tamm, L.K., H. Hong, and B. Liang, *Folding and assembly of β -barrel membrane proteins*. Biochimica et Biophysica Acta (BBA)-Biomembranes, 2004. **1666**(1): p. 250-263.
55. Kleinschmidt, J.H. and L.K. Tamm, *Secondary and tertiary structure formation of the β -barrel membrane protein OmpA is synchronized and depends on membrane thickness*. Journal of molecular biology, 2002. **324**(2): p. 319-330.
56. Surrey, T. and F. Jähnig, *Refolding and oriented insertion of a membrane protein into a lipid bilayer*. Proceedings of the National Academy of Sciences, 1992. **89**(16): p. 7457-7461.
57. Jansen, C., et al., *The assembly pathway of outer membrane protein PhoE of Escherichia coli*. The FEBS Journal, 2000. **267**(12): p. 3792-3800.
58. Ricci, D.P. and T.J. Silhavy, *The Bam machine: a molecular cooper*. Biochimica et Biophysica Acta (BBA)-Biomembranes, 2012. **1818**(4): p. 1067-1084.
59. Kim, K.H., S. Aulakh, and M. Paetzel, *The bacterial outer membrane β - barrel assembly machinery*. Protein Science, 2012. **21**(6): p. 751-768.
60. Struyvé, M., M. Moons, and J. Tommassen, *Carboxy-terminal phenylalanine is essential for the correct assembly of a bacterial outer membrane protein*. Journal of molecular biology, 1991. **218**(1): p. 141-148.
61. Robert, V., et al., *Assembly factor Omp85 recognizes its outer membrane protein substrates by a species-specific C-terminal motif*. PLoS biology, 2006. **4**(11): p. e377.
62. de Cock, H., et al., *Role of the carboxy-terminal phenylalanine in the biogenesis of outer membrane protein PhoE of Escherichia coli K-12*. 1997, Elsevier.

63. Rollauer, S.E., et al., *Outer membrane protein biogenesis in Gram-negative bacteria*. Phil. Trans. R. Soc. B, 2015. **370**(1679): p. 20150023.
64. Voulhoux, R., et al., *Role of a highly conserved bacterial protein in outer membrane protein assembly*. Science, 2003. **299**(5604): p. 262-265.
65. Walther, D.M., D. Rapaport, and J. Tommassen, *Biogenesis of β -barrel membrane proteins in bacteria and eukaryotes: evolutionary conservation and divergence*. Cellular and Molecular Life Sciences, 2009. **66**(17): p. 2789-2804.
66. Wimley, W.C., *The versatile β -barrel membrane protein*. Current opinion in structural biology, 2003. **13**(4): p. 404-411.
67. Genevrois, S., et al., *The Omp85 protein of Neisseria meningitidis is required for lipid export to the outer membrane*. The EMBO journal, 2003. **22**(8): p. 1780-1789.
68. Voulhoux, R. and J. Tommassen, *Omp85, an evolutionarily conserved bacterial protein involved in outer-membrane-protein assembly*. Research in microbiology, 2004. **155**(3): p. 129-135.
69. Doerrler, W.T. and C.R. Raetz, *Loss of outer membrane proteins without inhibition of lipid export in an Escherichia coli YaeT mutant*. Journal of Biological Chemistry, 2005. **280**(30): p. 27679-27687.
70. Werner, J. and R. Misra, *YaeT (Omp85) affects the assembly of lipid - dependent and lipid - independent outer membrane proteins of Escherichia coli*. Molecular microbiology, 2005. **57**(5): p. 1450-1459.
71. Tashiro, Y., et al., *Opr86 is essential for viability and is a potential candidate for a protective antigen against biofilm formation by Pseudomonas aeruginosa*. Journal of bacteriology, 2008. **190**(11): p. 3969-3978.
72. Knowles, T.J., et al., *Fold and function of polypeptide transport - associated domains responsible for delivering unfolded proteins to membranes*. Molecular microbiology, 2008. **68**(5): p. 1216-1227.
73. Wu, T., et al., *Identification of a multicomponent complex required for outer membrane biogenesis in Escherichia coli*. Cell, 2005. **121**(2): p. 235-245.
74. Sklar, J.G., et al., *Lipoprotein SmpA is a component of the YaeT complex that assembles outer membrane proteins in Escherichia coli*. Proceedings of the National Academy of Sciences, 2007. **104**(15): p. 6400-6405.
75. Webb, C.T., et al., *Dynamic association of BAM complex modules includes surface exposure of the lipoprotein BamC*. J Mol Biol, 2012. **422**(4): p. 545-55.
76. Noinaj, N., et al., *Structural insight into the biogenesis of beta-barrel membrane proteins*. Nature, 2013. **501**(7467): p. 385-90.
77. Hagan, C.L., S. Kim, and D. Kahne, *Reconstitution of outer membrane protein assembly from purified components*. Science, 2010. **328**(5980): p. 890-2.
78. Hagan, C.L., T.J. Silhavy, and D. Kahne, *beta-Barrel membrane protein assembly by the Bam complex*. Annu Rev Biochem, 2011. **80**: p. 189-210.
79. Hagan, C.L., D.B. Westwood, and D. Kahne, *bam Lipoproteins Assemble BamA in vitro*. Biochemistry, 2013. **52**(35): p. 6108-13.
80. Anwari, K., et al., *The evolution of new lipoprotein subunits of the bacterial outer membrane BAM complex*. Molecular microbiology, 2012. **84**(5): p. 832-844.
81. Malinverni, J.C., et al., *YfiO stabilizes the YaeT complex and is essential for outer membrane protein assembly in Escherichia coli*. Molecular microbiology, 2006. **61**(1): p. 151-164.
82. Sandoval, C.M., et al., *Crystal structure of BamD: an essential component of the β -barrel assembly machinery of gram-negative bacteria*. Journal of molecular biology, 2011. **409**(3): p. 348-357.

83. Hagan, C.L., J.S. Wzorek, and D. Kahne, *Inhibition of the beta-barrel assembly machine by a peptide that binds BamD*. Proc Natl Acad Sci U S A, 2015. **112**(7): p. 2011-6.
84. Kutik, S., et al., *Dissecting membrane insertion of mitochondrial beta-barrel proteins*. Cell, 2008. **132**(6): p. 1011-24.
85. Ricci, D.P., et al., *Activation of the Escherichia coli β -barrel assembly machine (Bam) is required for essential components to interact properly with substrate*. Proceedings of the National Academy of Sciences, 2012. **109**(9): p. 3487-3491.
86. Rigel, N.W., D.P. Ricci, and T.J. Silhavy, *Conformation-specific labeling of BamA and suppressor analysis suggest a cyclic mechanism for β -barrel assembly in Escherichia coli*. Proceedings of the National Academy of Sciences, 2013. **110**(13): p. 5151-5156.
87. Onufryk, C., et al., *Characterization of six lipoproteins in the σE regulon*. Journal of bacteriology, 2005. **187**(13): p. 4552-4561.
88. Charlson, E.S., J.N. Werner, and R. Misra, *Differential effects of yfgL mutation on Escherichia coli outer membrane proteins and lipopolysaccharide*. Journal of bacteriology, 2006. **188**(20): p. 7186-7194.
89. Rigel, N.W., et al., *BamE modulates the Escherichia coli beta-barrel assembly machine component BamA*. J Bacteriol, 2012. **194**(5): p. 1002-8.
90. Charlson, E.S., J.N. Werner, and R. Misra, *Differential effects of yfgL mutation on Escherichia coli outer membrane proteins and lipopolysaccharide*. J Bacteriol, 2006. **188**(20): p. 7186-94.
91. Mahoney, T.F., D.P. Ricci, and T.J. Silhavy, *Classifying β -barrel assembly substrates by manipulating essential Bam complex members*. Journal of bacteriology, 2016. **198**(14): p. 1984-1992.
92. Noinaj, N., J.W. Fairman, and S.K. Buchanan, *The crystal structure of BamB suggests interactions with BamA and its role within the BAM complex*. J Mol Biol, 2011. **407**(2): p. 248-60.
93. Kim, K.H., S. Aulakh, and M. Paetzel, *Crystal structure of beta-barrel assembly machinery BamCD protein complex*. J Biol Chem, 2011. **286**(45): p. 39116-21.
94. Sandoval, C.M., et al., *Crystal structure of BamD: an essential component of the beta-Barrel assembly machinery of gram-negative bacteria*. J Mol Biol, 2011. **409**(3): p. 348-57.
95. Knowles, T.J., et al., *Structure and function of BamE within the outer membrane and the beta-barrel assembly machine*. EMBO Rep, 2011. **12**(2): p. 123-8.
96. Albrecht, R., et al., *Structure of BamA, an essential factor in outer membrane protein biogenesis*. Acta Crystallogr D Biol Crystallogr, 2014. **70**(Pt 6): p. 1779-89.
97. Ni, D., et al., *Structural and functional analysis of the beta-barrel domain of BamA from Escherichia coli*. FASEB J, 2014. **28**(6): p. 2677-85.
98. Gatzeva-Topalova, P.Z., T.A. Walton, and M.C. Sousa, *Crystal structure of YaeT: conformational flexibility and substrate recognition*. Structure, 2008. **16**(12): p. 1873-81.
99. Heuck, A., A. Schleiffer, and T. Clausen, *Augmenting β -augmentation: structural basis of how BamB binds BamA and may support folding of outer membrane proteins*. Journal of molecular biology, 2011. **406**(5): p. 659-666.
100. Kim, K.H. and M. Paetzel, *Crystal structure of Escherichia coli BamB, a lipoprotein component of the beta-barrel assembly machinery complex*. J Mol Biol, 2011. **406**(5): p. 667-78.

101. Dong, C., et al., *Structure of Escherichia coli BamD and its functional implications in outer membrane protein assembly*. Acta Crystallogr D Biol Crystallogr, 2012. **68**(Pt 2): p. 95-101.
102. Kim, K.H., et al., *Structural characterization of Escherichia coli BamE, a lipoprotein component of the beta-barrel assembly machinery complex*. Biochemistry, 2011. **50**(6): p. 1081-90.
103. Gatzeva-Topalova, P.Z., et al., *Structure and flexibility of the complete periplasmic domain of BamA: the protein insertion machine of the outer membrane*. Structure, 2010. **18**(11): p. 1492-501.
104. Kim, S., et al., *Structure and function of an essential component of the outer membrane protein assembly machine*. Science, 2007. **317**(5840): p. 961-4.
105. Vuong, P., et al., *Analysis of YfgL and YaeT interactions through bioinformatics, mutagenesis, and biochemistry*. J Bacteriol, 2008. **190**(5): p. 1507-17.
106. Burgess, N.K., et al., *β -barrel proteins that reside in the Escherichia coli outer membrane in vivo demonstrate varied folding behavior in vitro*. Journal of Biological Chemistry, 2008. **283**(39): p. 26748-26758.
107. Gessmann, D., et al., *Outer membrane β -barrel protein folding is physically controlled by periplasmic lipid head groups and BamA*. Proceedings of the National Academy of Sciences, 2014. **111**(16): p. 5878-5883.
108. Stanley, A.M. and K.G. Fleming, *The process of folding proteins into membranes: challenges and progress*. Archives of biochemistry and biophysics, 2008. **469**(1): p. 46-66.
109. Sinnige, T., et al., *Solid-state NMR studies of full-length BamA in lipid bilayers suggest limited overall POTRA mobility*. Journal of molecular biology, 2014. **426**(9): p. 2009-2021.
110. Fleming, P.J., et al., *BamA POTRA Domain Interacts with a Native Lipid Membrane Surface*. Biophys J, 2016. **110**(12): p. 2698-2709.
111. Iadanza, M.G., et al., *Lateral opening in the intact [beta]-barrel assembly machinery captured by cryo-EM*. Nature communications, 2016. **7**.
112. Noinaj, N., J.C. Gumbart, and S.K. Buchanan, *The β -barrel assembly machinery in motion*. Nature Reviews Microbiology, 2017. **15**(4): p. 197.
113. Jeanteur, D., J. Lakey, and F. Pattus, *The bacterial porin superfamily: sequence alignment and structure prediction*. Molecular microbiology, 1991. **5**(9): p. 2153-2164.
114. Weiss, M., et al., *Molecular architecture and electrostatic properties of a bacterial porin*. Science(Washington), 1991. **254**(6038): p. 1627-1630.
115. Cowan, S., et al., *Crystal structures explain functional properties of two E. coli porins*. Nature, 1992. **358**(6389): p. 727-733.
116. Jap, B.K. and P.J. Walian, *Biophysics of the structure and function of porins*. Quarterly reviews of biophysics, 1990. **23**(4): p. 367-403.
117. Miller, V.L. and J.J. Mekalanos, *A novel suicide vector and its use in construction of insertion mutations: osmoregulation of outer membrane proteins and virulence determinants in Vibrio cholerae requires toxR*. Journal of bacteriology, 1988. **170**(6): p. 2575-2583.
118. Provenzano, D. and K.E. Klose, *Altered expression of the ToxR-regulated porins OmpU and OmpT diminishes Vibrio cholerae bile resistance, virulence factor expression, and intestinal colonization*. Proceedings of the National Academy of Sciences, 2000. **97**(18): p. 10220-10224.
119. Marylise Duperthuy, P.S., Edwin Garzón, Audrey Caro, Rafael D. Rosa, Frédérique Le Roux, Nicole Lautrédou-Audouy, Patrice Got, Bernard Romestand, Julien de

- Lorgeril, Sylvie Kieffer-Jaquinod, Evelyne Bachère, and Delphine Destoumieux-Garzón, *Use of OmpU porins for attachment and invasion of Crassostrea gigas immune cells by the oyster pathogen Vibrio splendidus*. Proceedings of the National Academy of Sciences, 2011. **108**: p. 2993-2998.
120. Liu, X., et al., *Outer membrane protein U (OmpU) mediates adhesion of Vibrio mimicus to host cells via two novel N-terminal motifs*. PLoS One, 2015. **10**(3): p. e0119026.
 121. Soulas, C., et al., *Outer Membrane Protein A (OmpA) Binds to and Activates Human Macrophages*. The Journal of Immunology, 2000. **165**(5): p. 2335-2340.
 122. Virji, M., K. Makepeace, and E.R. Moxon, *Distinct mechanisms of interactions of Opc -expressing meningococci at apical and basolateral surfaces of human endothelial cells; the role of integrins in apical interactions*. Molecular microbiology, 1994. **14**(1): p. 173-184.
 123. Sa, E.C.C., N.J. Griffiths, and M. Virji, *Neisseria meningitidis Opc invasin binds to the sulphated tyrosines of activated vitronectin to attach to and invade human brain endothelial cells*. PLoS Pathog, 2010. **6**(5): p. e1000911.
 124. Chakrabarti S.R., C.K., Sen K, Das J, *Porins of Vibrio cholerae: purification and characterization of OmpU*. J. Bacteriol., 1996. **178**(no.2): p. 524-530.
 125. Mathur, J. and M.K. Waldor, *The Vibrio cholerae ToxR-regulated porin OmpU confers resistance to antimicrobial peptides*. Infect Immun, 2004. **72**(6): p. 3577-83.
 126. Wibbenmeyer, J.A., *Vibrio cholerae OmpU and OmpT Porins Are Differentially Affected by Bile*. Infection and Immunity, 2002. **70**(1): p. 121-126.
 127. Sperandio V, G.J., Silveira WD, Kaper JB, *The OmpU outer membrane protein, a potential adherence factor of Vibrio cholerae*. Infect Immun, 1995. **63**: p. 4433-8.
 128. Wang, P.-P., et al., *Localization of immunodominant linear B-cell epitopes of Vibrio mimicus outer membrane protein U (OmpU)*. African Journal of Biotechnology, 2012. **11**(35): p. 8751-8757.
 129. Duret, G. and A.H. Delcour, *Size and dynamics of the Vibrio cholerae porins OmpU and OmpT probed by polymer exclusion*. Biophys J, 2010. **98**(9): p. 1820-9.
 130. Cai, S.H., et al., *Cloning, expression of Vibrio alginolyticus outer membrane protein-OmpU gene and its potential application as vaccine in crimson snapper, Lutjanus erythropterus Bloch*. J Fish Dis, 2013. **36**(8): p. 695-702.
 131. Khan, J., et al., *Refolding and functional assembly of the Vibrio cholerae porin OmpU recombinantly expressed in the cytoplasm of Escherichia coli*. Protein Expr Purif, 2012. **85**(2): p. 204-10.
 132. Luirink, J., et al., *Biogenesis of inner membrane proteins in Escherichia coli*. Annu. Rev. Microbiol., 2005. **59**: p. 329-355.
 133. Herskovits, A.A., E.S. Bochkareva, and E. Bibi, *New prospects in studying the bacterial signal recognition particle pathway*. Molecular microbiology, 2000. **38**(5): p. 927-939.
 134. Luirink, J. and I. Sinning, *SRP-mediated protein targeting: structure and function revisited*. Biochimica Et Biophysica Acta (BBA)-Molecular Cell Research, 2004. **1694**(1): p. 17-35.
 135. Samuelson, J.C., et al., *YidC mediates membrane protein insertion in bacteria*. Nature, 2000. **406**(6796): p. 637.
 136. Nagamori, S., I.N. Smirnova, and H.R. Kaback, *Role of YidC in folding of polytopic membrane proteins*. J Cell Biol, 2004. **165**(1): p. 53-62.
 137. Scotti, P.A., et al., *YidC, the Escherichia coli homologue of mitochondrial Oxa1p, is a component of the Sec translocase*. The EMBO journal, 2000. **19**(4): p. 542-549.

138. Michel, L.V., et al., *Dual orientation of the outer membrane lipoprotein P6 of nontypeable Haemophilus influenzae*. Journal of bacteriology, 2013. **195**(14): p. 3252-3259.
139. Cullen, P.A., D.A. Haake, and B. Adler, *Outer membrane proteins of pathogenic spirochetes*. FEMS microbiology reviews, 2004. **28**(3): p. 291-318.
140. Cowles, C.E., et al., *The free and bound forms of Lpp occupy distinct subcellular locations in Escherichia coli*. Molecular microbiology, 2011. **79**(5): p. 1168-1181.
141. von Heijne, G., *The structure of signal peptides from bacterial lipoproteins*. Protein Engineering, Design and Selection, 1989. **2**(7): p. 531-534.
142. Babu, M.M., et al., *A database of bacterial lipoproteins (DOLOP) with functional assignments to predicted lipoproteins*. Journal of bacteriology, 2006. **188**(8): p. 2761-2773.
143. Hayashi, S. and H.C. Wu, *Lipoproteins in bacteria*. Journal of bioenergetics and biomembranes, 1990. **22**(3): p. 451-471.
144. Setubal, J.C., et al., *Lipoprotein computational prediction in spirochaetal genomes*. Microbiology, 2006. **152**(1): p. 113-121.
145. Zückert, W.R., *Secretion of bacterial lipoproteins: through the cytoplasmic membrane, the periplasm and beyond*. Biochimica et Biophysica Acta (BBA)-Molecular Cell Research, 2014. **1843**(8): p. 1509-1516.
146. Madan Babu, M. and K. Sankaran, *DOLOP—database of bacterial lipoproteins*. Bioinformatics, 2002. **18**(4): p. 641-643.
147. Shu, W., et al., *Core structure of the outer membrane lipoprotein from Escherichia coli at 1.9 Å resolution*. Journal of molecular biology, 2000. **299**(4): p. 1101-1112.
148. Brangulis, K., et al., *Crystal structure of the infectious phenotype-associated outer surface protein BBA66 from the Lyme disease agent Borrelia burgdorferi*. Ticks and tick-borne diseases, 2014. **5**(1): p. 63-68.
149. Schulze, R.J., et al., *Translocation of Borrelia burgdorferi surface lipoprotein OspA through the outer membrane requires an unfolded conformation and can initiate at the C - terminus*. Molecular microbiology, 2010. **76**(5): p. 1266-1278.
150. Fukuda, A., et al., *Aminoacylation of the N-terminal cysteine is essential for Lol-dependent release of lipoproteins from membranes but does not depend on lipoprotein sorting signals*. Journal of Biological Chemistry, 2002. **277**(45): p. 43512-43518.
151. Sankaran, K. and H.C. Wu, *Lipid modification of bacterial prolipoprotein. Transfer of diacylglyceryl moiety from phosphatidylglycerol*. Journal of Biological Chemistry, 1994. **269**(31): p. 19701-19706.
152. Kovacs-Simon, A., R. Titball, and S.L. Michell, *Lipoproteins of bacterial pathogens*. Infection and immunity, 2011. **79**(2): p. 548-561.
153. Buddelmeijer, N., *The molecular mechanism of bacterial lipoprotein modification—How, when and why?* FEMS microbiology reviews, 2015. **39**(2): p. 246-261.
154. Tokunaga, M., J.M. Loranger, and H.C. Wu, *A distinct signal peptidase for prolipoprotein in Escherichia coli*. Journal of cellular biochemistry, 1984. **24**(2): p. 113-120.
155. Daley, D.O., et al., *Global topology analysis of the Escherichia coli inner membrane proteome*. Science, 2005. **308**(5726): p. 1321-1323.
156. Mao, G., et al., *Crystal structure of E. coli lipoprotein diacylglyceryl transferase*. Nature communications, 2016. **7**: p. 10198.
157. Pailler, J., et al., *Phosphatidylglycerol:: prolipoprotein diacylglyceryl transferase (Lgt) of Escherichia coli has seven transmembrane segments, and its essential residues are embedded in the membrane*. Journal of bacteriology, 2012. **194**(9): p. 2142-2151.

158. Tokunaga, M., H. Tokunaga, and H.C. Wu, *Post-translational modification and processing of Escherichia coli prolipoprotein in vitro*. Proceedings of the National Academy of Sciences, 1982. **79**(7): p. 2255-2259.
159. Munoa, F., et al., *Membrane topology of Escherichia coli prolipoprotein signal peptidase (signal peptidase II)*. Journal of Biological Chemistry, 1991. **266**(26): p. 17667-17672.
160. Tokunaga, M., J.M. Loranger, and H. Wu, *Prolipoprotein modification and processing enzymes in Escherichia coli*. Journal of Biological Chemistry, 1984. **259**(6): p. 3825-3830.
161. Vogeley, L., et al., *Structural basis of lipoprotein signal peptidase II action and inhibition by the antibiotic globomycin*. Science, 2016. **351**(6275): p. 876-880.
162. Gupta, S.D., et al., *Characterization of a temperature-sensitive mutant of Salmonella typhimurium defective in apolipoprotein N-acyltransferase*. Journal of Biological Chemistry, 1993. **268**(22): p. 16551-16556.
163. Gupta, S.D. and H.C. Wu, *Identification and subcellular localization of apolipoprotein N-acyltransferase in Escherichia coli*. FEMS microbiology letters, 1991. **78**(1): p. 37-41.
164. Hillmann, F., M. Argentini, and N. Buddelmeijer, *Kinetics and phospholipid specificity of apolipoprotein N-acyltransferase*. Journal of Biological Chemistry, 2011: p. jbc.M111.243519.
165. Buddelmeijer, N. and R. Young, *The essential Escherichia coli apolipoprotein N-acyltransferase (Lnt) exists as an extracytoplasmic thioester acyl-enzyme intermediate*. Biochemistry, 2009. **49**(2): p. 341-346.
166. Vidal-Ingigliardi, D., S. Lewenza, and N. Buddelmeijer, *Identification of essential residues in apolipoprotein N-acyl transferase, a member of the CN hydrolase family*. Journal of bacteriology, 2007. **189**(12): p. 4456-4464.
167. Hillmann, F., M. Argentini, and N. Buddelmeijer, *Kinetics and phospholipid specificity of apolipoprotein N-acyltransferase*. Journal of Biological Chemistry, 2011. **286**(32): p. 27936-27946.
168. Gupta, S., W. Dowhan, and H. Wu, *Phosphatidylethanolamine is not essential for the N-acylation of apolipoprotein in Escherichia coli*. Journal of Biological Chemistry, 1991. **266**(15): p. 9983-9986.
169. Gélis - Jeanvoine, S., et al., *Residues located on membrane - embedded flexible loops are essential for the second step of the apolipoprotein N - acyltransferase reaction*. Molecular microbiology, 2015. **95**(4): p. 692-705.
170. Roman-Hernandez, G., J.H. Peterson, and H.D. Bernstein, *Reconstitution of bacterial autotransporter assembly using purified components*. Elife, 2014. **3**: p. e04234.
171. Kabsch, W., *Xds*. Acta Crystallographica Section D: Biological Crystallography, 2010. **66**(2): p. 125-132.
172. Sheldrick, G.M., *Experimental phasing with SHELXC/D/E: combining chain tracing with density modification*. Acta Crystallographica Section D: Biological Crystallography, 2010. **66**(4): p. 479-485.
173. Cowtan, K., *'DM': An automated procedure for phase improvement by density modification*. Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallogr., 1994. **31**: p. 34-38.
174. Emsley, P., et al., *Features and development of Coot*. Acta Crystallographica Section D: Biological Crystallography, 2010. **66**(4): p. 486-501.
175. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution*. Acta Crystallographica Section D: Biological Crystallography, 2010. **66**(2): p. 213-221.

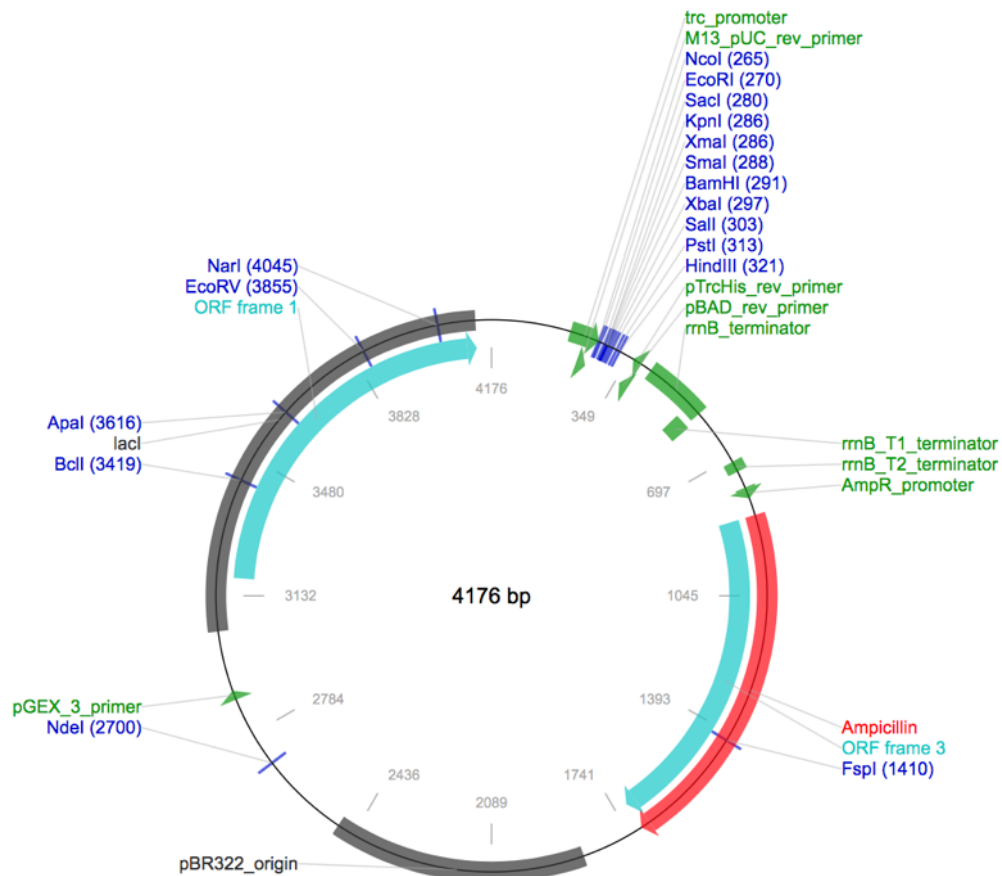
176. Evans, P., *Scaling and assessment of data quality*. Acta Crystallographica Section D: Biological Crystallography, 2006. **62**(1): p. 72-82.
177. Winn, M.D., et al., *Overview of the CCP4 suite and current developments*. Acta Crystallographica Section D: Biological Crystallography, 2011. **67**(4): p. 235-242.
178. Vagin, A.A., et al., *REFMAC5 dictionary: organization of prior chemical knowledge and guidelines for its use*. Acta Crystallographica Section D: Biological Crystallography, 2004. **60**(12): p. 2184-2195.
179. Lamzin, V.S. and K.S. Wilson, *Automated refinement of protein models*. Acta Crystallographica Section D: Biological Crystallography, 1993. **49**(1): p. 129-147.
180. GEHealthcare, *Superdex™ prep grade and prepacked HiLoad™ columns*. 2013, GE Healthcare Life Sciences.
181. Hagan, C.L., *Dissertation: Reconstitution of the E. coli Membrane β -Barrel Assembly Machine from Purified Components*. 2012.
182. Tellez, R., Jr. and R. Misra, *Substitutions in the BamA beta-barrel domain overcome the conditional lethal phenotype of a DeltabamB DeltabamE strain of Escherichia coli*. J Bacteriol, 2012. **194**(2): p. 317-24.
183. Duret, G. and A.H. Delcour, *Size and dynamics of the Vibrio cholerae porins OmpU and OmpT probed by polymer exclusion*. Biophysical journal, 2010. **98**(9): p. 1820-1829.
184. Wibbenmeyer, J.A., et al., *Vibrio cholerae OmpU and OmpT porins are differentially affected by bile*. Infection and immunity, 2002. **70**(1): p. 121-126.
185. Smart, O.S., et al., *HOLE: a program for the analysis of the pore dimensions of ion channel structural models*. Journal of molecular graphics, 1996. **14**(6): p. 354-360.
186. Rath, A., et al., *Detergent binding explains anomalous SDS-PAGE migration of membrane proteins*. Proceedings of the National Academy of Sciences, 2009. **106**(6): p. 1760-1765.
187. Wiktor, M., et al., *Structural insights into the mechanism of the membrane integral N-acyltransferase step in bacterial lipoprotein synthesis*. Nature Communications, 2017. **8**.
188. Lu, G., et al., *Crystal structure of E. coli apolipoprotein N-acyl transferase*. Nature Communications, 2017. **8**.
189. Noland, C.L., et al., *Structural insights into lipoprotein N-acylation by Escherichia coli apolipoprotein N-acyltransferase*. Proceedings of the National Academy of Sciences, 2017. **114**(30): p. E6044-E6053.
190. Han, L., et al., *Structure of the BAM complex and its implications for biogenesis of outer-membrane proteins*. Nature structural & molecular biology, 2016. **23**(3): p. 192-196.
191. Bakelar, J., S.K. Buchanan, and N. Noinaj, *The structure of the β -barrel assembly machinery complex*. Science, 2016. **351**(6269): p. 180-186.
192. Hagan, C.L. and D. Kahne, *Reconstituted Escherichia coli Bam complex catalyzes multiple rounds of β -barrel assembly*. Biochemistry, 2011. **50**(35): p. 7444.
193. Webb, C.T., et al., *Dynamic association of BAM complex modules includes surface exposure of the lipoprotein BamC*. Journal of molecular biology, 2012. **422**(4): p. 545-555.
194. Wilson, M.M. and H.D. Bernstein, *Surface-exposed lipoproteins: an emerging secretion phenomenon in Gram-negative bacteria*. Trends in microbiology, 2016. **24**(3): p. 198-208.
195. Knowles, T.J., et al., *Secondary structure and (1)H, (13)C and (15)N backbone resonance assignments of BamC, a component of the outer membrane protein assembly machinery in Escherichia coli*. Biomol NMR Assign, 2009. **3**(2): p. 203-6.

196. Warner, L.R., et al., *Structure of the BamC two-domain protein obtained by Rosetta with a limited NMR data set*. Journal of molecular biology, 2011. **411**(1): p. 83-95.
197. Braun, M. and T.J. Silhavy, *Imp/OstA is required for cell envelope biogenesis in Escherichia coli*. Molecular microbiology, 2002. **45**(5): p. 1289-1302.
198. Bos, M.P., et al., *Identification of an outer membrane protein required for the transport of lipopolysaccharide to the bacterial cell surface*. Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(25): p. 9417-9422.
199. Sampson, B.A., R. Misra, and S.A. Benson, *Identification and characterization of a new gene of Escherichia coli K-12 involved in outer membrane permeability*. Genetics, 1989. **122**(3): p. 491-501.
200. Lee, J., et al., *Characterization of a stalled complex on the β -barrel assembly machine*. Proceedings of the National Academy of Sciences, 2016. **113**(31): p. 8717-8722.
201. Vogt, J. and G.E. Schulz, *The structure of the outer membrane protein OmpX from Escherichia coli reveals possible mechanisms of virulence*. Structure, 1999. **7**(10): p. 1301-1309.
202. Seed, K.D., et al., *Evolutionary consequences of intra-patient phage predation on microbial populations*. Elife, 2014. **3**: p. e03497.
203. Vollan, H.S., et al., *In silico structure and sequence analysis of bacterial porins and specific diffusion channels for hydrophilic molecules: conservation, multimericity and multifunctionality*. International journal of molecular sciences, 2016. **17**(4): p. 599.
204. Provenzano, D., et al., *The Virulence Regulatory Protein ToxR Mediates Enhanced Bile Resistance in Vibrio cholerae and Other Pathogenic Vibrio Species*. Infection and immunity, 2000. **68**(3): p. 1491-1497.
205. Acosta-Gutierrez, S., et al., *Filtering with electric field: the case of E. coli porins*. The journal of physical chemistry letters, 2015. **6**(10): p. 1807-1812.
206. Zheng, H., et al., *Validating metal binding sites in macromolecule structures using the CheckMyMetal web server*. Nature protocols, 2014. **9**(1): p. 156.
207. Tokuda, H. and S.-i. Matsuyama, *Sorting of lipoproteins to the outer membrane in E. coli*. Biochimica et Biophysica Acta (BBA)-Molecular Cell Research, 2004. **1693**(1): p. 5-13.
208. Okuyama, H., T. KANKURA, and S. NOJIMA, *Positional distribution of fatty acids in phospholipids from Mycobacteria*. The Journal of Biochemistry, 1967. **61**(6): p. 732-737.
209. Brülle, J.K., A. Tschumi, and P. Sander, *Lipoproteins of slow-growing Mycobacteria carry three fatty acids and are N-acylated by apolipoprotein N-acyltransferase BCG_2070c*. BMC microbiology, 2013. **13**(1): p. 223.
210. Baulard, A.R., et al., *In vivo interaction between the polyprenol phosphate mannose synthase Ppm1 and the integral membrane protein Ppm2 from Mycobacterium smegmatis revealed by a bacterial two-hybrid system*. Journal of Biological Chemistry, 2003. **278**(4): p. 2242-2248.
211. Widdick, D.A., et al., *Dissecting the complete lipoprotein biogenesis pathway in Streptomyces scabies*. Molecular microbiology, 2011. **80**(5): p. 1395-1412.
212. Hutchings, M.I., et al., *Lipoprotein biogenesis in Gram-positive bacteria: knowing when to hold 'em, knowing when to fold 'em*. Trends in microbiology, 2009. **17**(1): p. 13-21.
213. Tschumi, A., et al., *Identification of apolipoprotein N-acyltransferase (Lnt) in mycobacteria*. Journal of Biological Chemistry, 2009. **284**(40): p. 27146-27156.

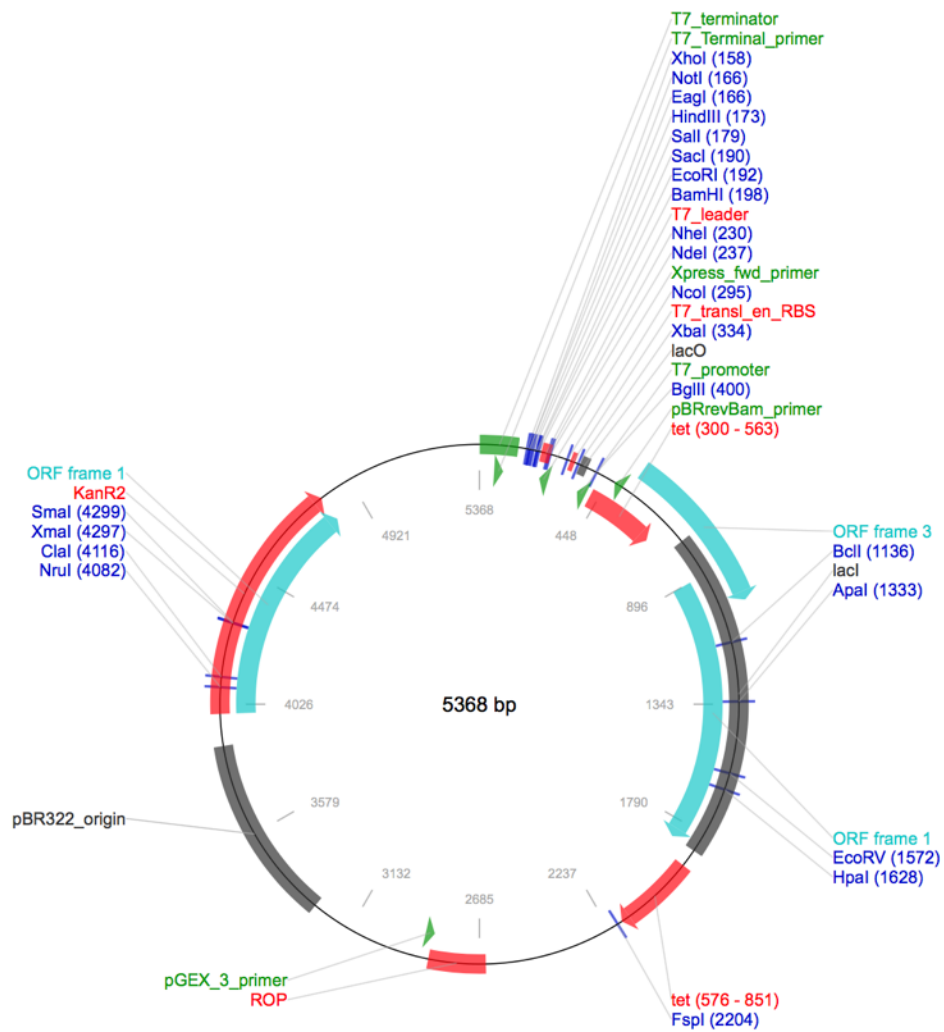
214. Tawaratsumida, K., et al., *Characterization of N-terminal structure of TLR2-activating lipoprotein in Staphylococcus aureus*. Journal of Biological Chemistry, 2009. **284**(14): p. 9147-9152.
215. Shimizu, T., Y. Kida, and K. Kuwano, *A dipalmitoylated lipoprotein from Mycoplasma pneumoniae activates NF- κ B through TLR1, TLR2, and TLR6*. The Journal of Immunology, 2005. **175**(7): p. 4641-4646.
216. Kurokawa, K., et al., *The triacylated ATP binding cluster transporter substrate-binding lipoprotein of Staphylococcus aureus functions as a native ligand for Toll-like receptor 2*. Journal of Biological Chemistry, 2009. **284**(13): p. 8406-8411.
217. Asanuma, M., et al., *Structural evidence of α - aminoacylated lipoproteins of Staphylococcus aureus*. The FEBS journal, 2011. **278**(5): p. 716-728.
218. Nakayama, H., K. Kurokawa, and B.L. Lee, *Lipoproteins in bacteria: structures and biosynthetic pathways*. The FEBS journal, 2012. **279**(23): p. 4247-4268.
219. Fischer, W., *Lipoteichoic acid and lipids in the membrane of Staphylococcus aureus*. Medical microbiology and immunology, 1994. **183**(2): p. 61-76.

APPENDIX 1

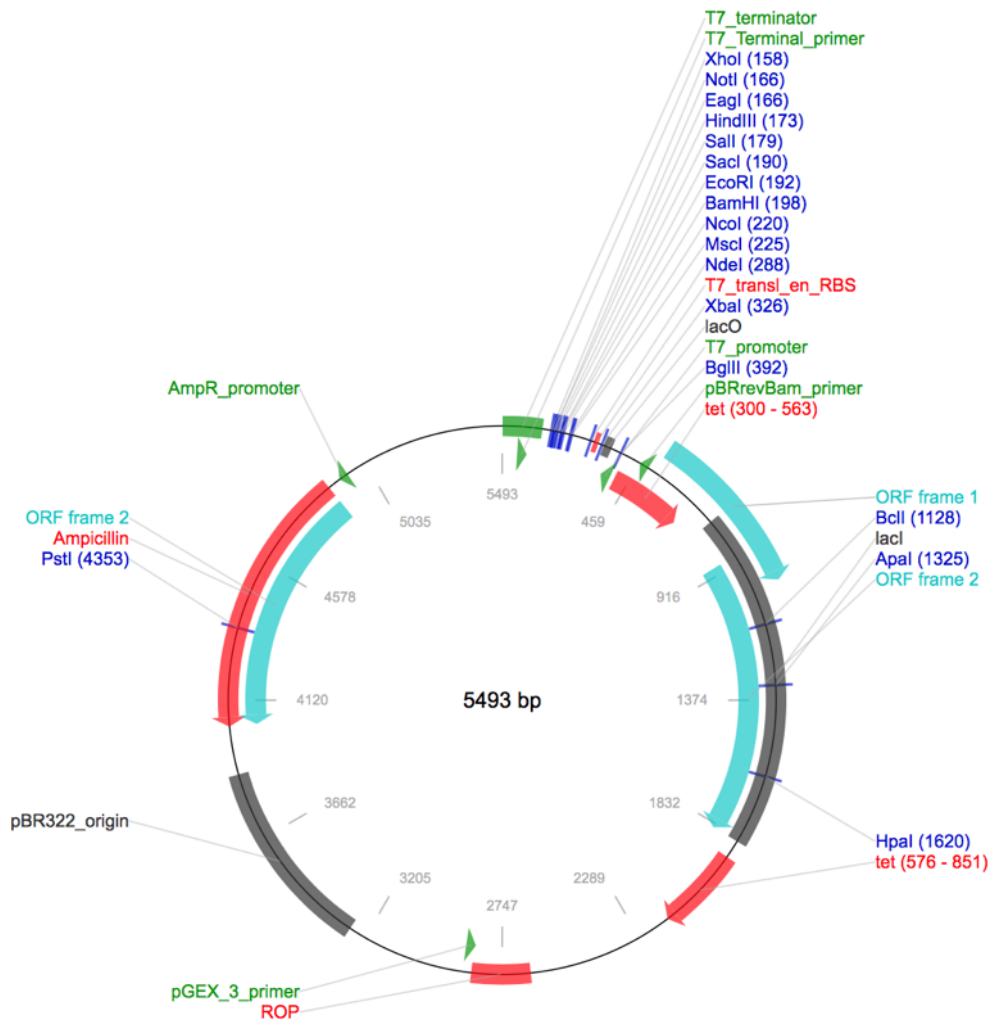
A. pTRC99a plasmid map



B. pET28 b plasmid map



C. pET22 b plasmid map



APPENDIX 2

Preparations of Minimal media stock solution for L-selenomethionine labeling of the BamACDE complex

1. M9 salts solution: 12.8 g $\text{Na}_2\text{HPO}_4 \cdot 7\text{H}_2\text{O}$, 3.0 g KH_2PO_4 (Fisher Scientific) and 1.0 g NH_4Cl (Sigma-Aldrich) dissolved in 900 ml of deionized H_2O and sent to autoclaving.
2. 10x essential media stock: 20% w/v glucose (Sigma-Aldrich), 0.3% w/v magnesium sulphate, 0.01% w/v iron (III) sulphate (Fisher Scientific) and 0.01% w/v thiamine-HCL (Duchefa Biochemie), filter sterilized prior to mixing with M9 salts.
3. 20x Glucose-free selenomethionine nutrient mix (MD12-502-GF, Molecular Dimensions), dissolved in deionized water at 0.022% w/v and filter-sterilized prior to mixing with M9 salts.
4. 1x phosphate buffered saline (PBS) solution: 12.8 g/L $\text{Na}_2\text{HPO}_4 \cdot 7\text{H}_2\text{O}$, 3.0 g/L KH_2PO_4 , pH7.2, 0.5 g/L NaCl and 1.0 g/L NH_4Cl .
5. 100x amino acid supplement: L-lysine, L-phenylalanine, L-threonine at 1% w/v and L-leucine, L-isoleucine and L-valine at 0.5% w/v (Duchefa Biochemie), filter sterilized.
6. 10x Se-Met solution: 10% w/v in filter sterilized deionized H_2O .
7. 5% v/v O/N culture suspension: Overnight culture grown in LB broth was washed using PBS solution to remove remaining organic components. O/N culture was pelleted in 50 ml aliquots at 2,800 rpm for 10 min, gently resuspended in 20 ml PBS and spun down again. Repeated twice and the cells were resuspended in 5 ml of PBS solution ready to be allocated in each 1L of minimal media (5% v/v).

APPENDIX 3

Tables of the interactive residues in the two BAM complex structures

Table 1. Interacting residues between BamA and BamB in the BamABCDE structure defined by proximity within 3.5 Å.

BamA	BamB
Q178	N58
Y255, Q179	F59, R77, L110
T248, P249, K251	Y60, Y263
T187, T189	E107
N181	E127
T190, V183, I182	K128, A129, G149
H186, G184	K146
N259	A148
T257, N181, N259	E150
Q244, S242	S167, N168, M189
V245	S191
L247, S246, P249, T248	L192, S193, L194, V286
Y255, T257	R194, R195
Y142, K135(?)	D242
K216	R243
D250	L326

Table 2. Interacting residues between BamA and BamC in the BamABCDE structure defined by proximity within 3.5 Å.

BamA	BamC
T359	K32

Table 3. Interacting residues between BamA and BamD in the BamABCDE structure defined by proximity within 3.5 Å.

BamA	BamD
V121, R162, A95, R160	N60, R61, Y62
R36, R162, T93	P63
Q35, R36, V37, N71, A41, A38,	F64, L98
R120	P86, L87, A90
G122, E123, A41, N71,	R94, R97, P100, N99,
G40	T101
K361, D362, A363, D162, F354,	D134, Y177
N357, D358, T359, S360, K361, G356, Q355	R135, D136, H139
R366, V480, R367, R366, V480, E373, I352, M372	V181, Y184, Y185, A190, A193
V480, D481	E187, R188
G374	V192, A193
K351, R366	R197

Table 4. Interacting residues between BamA and BamE in the BamABCDE structure defined by proximity within 3.5 Å.

BamA	BamE
D481	R29
R370	P30
M372, W376, A375, E373, G374, Y348, V349, R350	I32, N33, Q34, G35, N36, P62, L63, F77, R78, Q79
G316, Y315, R346, R314, D410	Y37, T39, G60
Q411, P409, S408	T61

Table 5. Interacting residues between BamD and BamE in the BamABCDE structure defined by proximity within 3.5 Å.

DamD	BamE
R188	R29, P30, I32,
V192, A190, A193, W191, N196, V192	Q34, G35, L63,
M227, M225, V195, Q230, K233,	M64, D66, F74, V76, Q88, T90
E199	P67
I237, L202, M218, Y222, V234,	F68
K233	T70
W191, G189	R78
Q226	L110, S111

Table 6. Interacting residues between BamC and BamD in the BamABCDE structure defined by proximity within 3.5 Å.

BamC	BamD
K32	D121, R135, D121, R135
R33, Q34	L119, L85, A83,
S36	A118
G37, D38, E39	K81
L42	F169, S122,
Q43	K165
A44	F169
A45, P46	V168, Q209, D172,
L47	D207, K171, T208,
A48, E49	R212,
L50, H51, L58, P59	N241, L215, I237, L202, P206,
P53	K236, A240,
M56	K233
V60, T61, S62,	R203, D204
D64, Y65, A66	F144, Y205, L174, R141, K171,
I67	L167, T164,
P68	V151
T70	T160, P155
G72, S73, G74	T161, Q158
A75, V76, G77	D162, K165
L80, D81	Y110, Y159,
I82, R83, P84, P85	Y107, Y77, L114, Y80, M111, L73, D74, N104,

Table 7. Interacting residues between BamC and BamE in the BamABCDE structure defined by proximity within 3.5 Å.

BamC	BamE
M56, L58, P59	F68, P67
I57	G69

Table 8. Interacting residues between BamA and BamC in the BamACDE structure defined by proximity within 3.5 Å.

BamA	BamC
Q35	G94
V39, F31	R96
R127	D236
P159	F248
L158	S272
T98	Q273
D164, K166	Q306,
P157, L158	G308, D309
Q170	H324

Table 9. Interactive residues between BamC and BamD in the BamACDE structure defined by proximity of within 3.5 Å.

BamC	BamD
R271	D29
L91	N33, Y67
L151, D152, K32, C25, Y31	Q44
Q87	D45
P84	Q69, H102
L89	L73
I82	Q70
S36	Y77, Y107
Q34	Y80
Q195	K81
Y65	N104
A66	R141, F144, Y205
L80	R152
I67	Q158
S73, G74	L167
A75, Y41, G77, V76, V35	T161
T70	D162, R166, F169
L47, A48	T164
T61	K171, D207
L50, E49	R203
M56	R212
H51, L50	K233, K236
	A240, N241

APPENDIX 4

List of the 96 detergents with individual stock and working concentration in the Analytic Selector Kit (Anatrace)

Part 1

Detergent Preparation and CMC								
Row	Column	Number	Detergent	Abbreviation	Anatrace Prod. No.	CMC (mM)	[Stock Plate] (mM)	Working Conc. (mM)
A	1	1	Water					
A	2	2	Blank (for solubilizing detergent)					
A	3	3	Anzergent® 3-10	Z3-10	AZ310	39	156	78
A	4	4	Anzergent 3-12	Z3-12	AZ312	2.8	16.8	8.4
A	5	5	Anzergent 3-14	Z3-14	AZ314	0.2	20	10
A	6	6	n-Decyl-N,N-Dimethylglycine	DMG	D352	19	76	38
A	7	7	n-Dodecyl-N,N-Dimethylglycine	DOMG	D350	1.5	9	4.5
A	8	8	n-Decyl-N,N-Dimethylamine-N-Oxide	DDAO	D365	10.5	42	21
A	9	9	n-Undecyl-n,n-Dimethylamine-Oxide	UDAO	U360	3.2	19.2	9.6
A	10	10	n-Dodecyl-N,N-Dimethylamine-N-Oxide	LDAO	D360	1	6	3
A	11	11	Cyclofos™-3	CF-3	C510	43	172	86
A	12	12	Cyclofos-4	CF-4	C512	14	56	28
B	1	13	Cyclofos-5	CF-5	C514	4.5	27	13.5
B	2	14	Cyclofos-6	CF-6	C516	2.68	16.08	8.04
B	3	15	Cyclofos-7	CF-7	C518	0.62	12.4	6.2
B	4	16	Fos-Choline®-12	FC-12	F308	1.5	9	4.5
B	5	17	Fos-Choline-13	FC-13	F310	0.75	15	7.5
B	6	18	Fos-Choline-14	FC-14	F312	0.12	12	6
B	7	19	Fos-Choline-15	FC-15	F314	0.07	14	7
B	8	20	Fos-Choline-16	FC-16	F316	0.013	2.6	1.3
B	9	21	Fos-Choline-ISO-9	FC-I9	FCI09	32	128	64
B	10	22	Fos-Choline-ISO-11	FC-I11	FCI11	26.6	106.4	53.2
B	11	23	Fos-Choline-UNSAT-11-10	FC-U10-11	FCU110	6.2	31	15.5
B	12	24	1,2-Diheptanoyl-sn-Glycero-3-Phosphocholine	DHPC	D607	1.4	8.4	4.2
C	1	25	LysopC-12	LPC-12	L212	0.7	14	7
C	2	26	LysopC-14	LPC-14	L214	0.036	7.2	3.6
C	3	27	CHAPS	CHAPS	C316	8	40	20
C	4	28	CHAPSO	CHAPSO	C317	8	40	20
C	5	29	Ph-Tripplu	Ph-Tripplu	T380	3.6	21.6	10.8
C	6	30	Cy-Tripplu	Cy-Tripplu	T385	1.8	10.8	5.4
C	7	31	LAPAO	LAPAO	L3605	1.6	9.6	4.8
C	8	32	Tripao	TRIPAO	T370	4.5	27	13.5
C	9	33	Anapoe®-20 (Tween 20)	T-20	APT020	0.059	11.8	5.9
C	10	34	Anapoe-35 (Brij 35)	Brij-35	APB035	0.091	18.2	9.1
C	11	35	Anapoe-X-100	TX-100	APX100	0.23	23	11.5
C	12	36	Anapoe-X-114	TX-114	APX114	0.2	20	10
D	1	37	Anapoe-X-305	TX-305	APX305	0.65	13	6.5
D	2	38	Anapoe-X-405	TX-405	APX405	0.81	16.2	8.1
D	3	39	[Octylphenoxy]Polyethoxyethanol	NID-P40	APND40	0.3	30	15
D	4	40	Lauryl Maltose Neopentyl Glycol	LMNG	NG310	0.01	2	1
D	5	41	Octyl Glucose Neopentyl Glycol	OGNG	NG311	1.02	6.12	3.06
D	6	42	Decyl Maltose Neopentyl Glycol	DMNG	NG322	0.036	7.2	3.6
D	7	43	CYMAL®-5 Neopentyl Glycol	CYMAL-5-NG	NG325	0.058	11.6	5.8
D	8	44	CYMAL-6 Neopentyl Glycol	CYMAL-6-NG	NG326	0.02	4	2
D	9	45	GDN101 - GDN	GDN	GDN101	0.018	3.6	1.8
D	10	46	Triethylene Glycol Monoethyl Ether	C ₆ E ₁	T330	23	92	46
D	11	47	Tetraethylene Glycol Monoethyl Ether	C ₆ E ₁	T340	30	120	60
D	12	48	Pentaethylene Glycol Monoethyl Ether	C ₆ E ₁	P360	37	148	74

Continued on next page.

Part 2

Row	Column	Number	Detergent	Abbreviation	Anatrace Prod. No.	CMC (mM)	[Stock Plate] (mM)	Working Conc. (mM)
E	1	49	Pentaethylene Glycol Monoheptyl Ether	C ₇ E ₅	P370	21	84	42
E	2	50	Tetraethylene Glycol Mono-octyl Ether	C ₈ E ₄	T350	8	40	20
E	3	51	Pentaethylene Glycol Mono-octyl Ether	C ₈ E ₅	P350	7.1	35.5	17.75
E	4	52	Hexaethylene Glycol Mono-octyl Ether	C ₈ E ₆	H350	10	50	25
E	5	53	Pentaethylene Glycol Monododecyl Ether	C ₁₂ E ₅	P340	0.81	16.2	8.1
E	6	54	Hexaethylene Glycol Monododecyl Ether	C ₁₂ E ₆	H360	0.9	18	9
E	7	55	Polyoxyethylene(9)decyl Ether	C ₁₀ E ₉	AP0109	1.3	7.8	3.9
E	8	56	Heptaethylene Glycol Monododecyl Ether	C ₁₂ E ₇	H370	0.069	13.8	6.9
E	9	57	Octaethylene Glycol Monododecyl Ether	C ₁₂ E ₈	O330	0.09	18	9
E	10	58	Polyoxyethylene(9)dodecyl Ether	C ₁₂ E ₉	AP0129	0.05	10	5
E	11	59	Polyoxyethylene(10)dodecyl Ether	C ₁₂ E ₁₀	AP1210	0.1	20	10
E	12	60	Polyoxyethylene(8)tridecyl Ether	C ₁₃ E ₈	AP0138	0.1	20	10
F	1	61	Big CHAP	CHAP	B300	2.9	17.4	8.7
F	2	62	Big CHAP, Deoxy	CHAP-D	B310	1.4	8.4	4.2
F	3	63	n-Heptyl-β-D-Thioglucofuranoside	HTG	H301	29	116	58
F	4	64	n-Octyl-β-D-Thioglucofuranoside	OTG	O314	9	25	12.5
F	5	65	n-Octyl-β-D-Glucofuranoside	OG	O311	18	72	36
F	6	66	n-Nonyl-β-D-Glucofuranoside	NG	N324	6.5	32.5	16.25
F	7	67	CYGLU* ₃	CYGLU-3	C323G	28	112	56
F	8	68	HECAMEG	Anameg-7	A340	19.5	78	39
F	9	69	Hega-9	HEGA-9	H109	39	156	78
F	10	70	Hega-10	HEGA-10	H110	7	35	17.5
F	11	71	Mega-9	M9	M325	25	100	50
F	12	72	Mega-10	M10	M320	6	12	6
G	1	73	2-Hydroxyethyloctylsulfoxide	OHES	Bachem - P1105	24.3	96.8	48.4
G	2	74	CYMAL-3	CYMAL-3	C323	30	120	60
G	3	75	CYMAL-4	CYMAL-4	C324	7.6	38	19
G	4	76	CYMAL-5	CYMAL-5	C325	2.4	14.4	7.2
G	5	77	CYMAL-6	CYMAL-6	C326	0.56	11.2	5.6
G	6	78	CYMAL-7	CYMAL-7	C327	0.19	19	9.5
G	7	79	2,6-Dimethyl-4-Heptyl-β-D-Maltoside	DMHM	DH325	27.5	110	55
G	8	80	2-Propyl-1-Pentyl-β-D-Maltopyranoside	PPM	P310	42.5	170	85
G	9	81	n-Octyl-β-D-Maltopyranoside	OM	O310	19.5	78	39
G	10	82	n-Nonyl-β-D-Maltopyranoside	NM	N330	6	30	15
G	11	83	n-Decyl-α-D-Maltopyranoside	DαM	D322HA	1.6	9.6	4.8
G	12	84	n-Decyl-β-D-Maltopyranoside	DM	D322	1.8	10.8	5.4
H	1	85	n-Undecyl-α-D-Maltopyranoside	UαDM	U300HA	0.58	11.6	5.8
H	2	86	n-Undecyl-β-D-Maltopyranoside	UDM	U300	0.59	11.8	5.9
H	3	87	ω-Undecylenyl-β-D-Maltopyranoside	ωUDM	U310	1.2	7.2	3.6
H	4	88	n-Dodecyl-α-D-Maltopyranoside	DαDM	D310HA	0.15	15	7.5
H	5	89	n-Dodecyl-β-D-Maltopyranoside	DDM	D310	0.17	17	8.5
H	6	90	n-Tridecyl-β-D-Maltopyranoside	TDM	T323	0.03	3	1.5
H	7	91	n-Octyl-β-D-Thiomaltopyranoside	OTM	O320	8.5	42.5	21.25
H	8	92	n-Nonyl-β-D-Thiomaltopyranoside	NTM	N350	3.2	19.2	9.6
H	9	93	n-Decyl-β-D-Thiomaltopyranoside	DTM	D335	0.9	18	9
H	10	94	n-Undecyl-β-D-Thiomaltopyranoside	UDTM	U342	0.21	21	10.5
H	11	95	n-Dodecyl-β-D-Thiomaltopyranoside	DDTM	D342	0.05	10	5
H	12	96	Sucrose 12	S-12	S350	0.3	30	15