Structural investigation of putative lipid transport systems in *Bdellovibrio bacteriovorus* HD100.

by

Giedrė Ratkevičiūtė

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

College of Life and Environmental Sciences

School of Biosciences

University of Birmingham

June 2023

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties. The intellectual property rights of the author or third parties in respect of this work are as defined by The Copyright Designs and Patents Act 1988 or as modified by any successor legislation.

Any use made of information contained in this thesis/dissertation must be in accordance with that legislation and must be properly acknowledged. Further distribution or reproduction in any format is prohibited without the permission of the copyright holder.

Abstract

Mammalian Cell Entry (MCE) domain-containing proteins have been implicated in glycerophospholipid transport, a key component of the gram-negative OM. The model organism *Escherichia coli* contains three MCE proteins residing within unique operons; MlaD, PqiB and LetB, each presenting potential novel targets for innovative antibacterial agents. Characterising these proteins and their associated complexes within multiple organisms may provide additional insight into these putative lipid transport systems.

Here, we focus on MCE proteins within *Bdellovibrio bacteriovorus* HD100, a predatory gram-negative bacterium, that alternates between free-living growth and predation upon other gram-negative bacteria. Thus far, we investigated Bd1152, a two-MCE domain containing protein residing at the centre of a tripartite operon.

The inherent structural plasticity of Bd1152 posed a challenge in obtaining its high-resolution structure. Nevertheless, we present a crystal structure of Bd1152 as well as a cryo-EM structure of the Bd1152 top MCE ring at 2.6 Å and 3.4 Å resolution, respectively. Both structures reveal hexameric assemblies (MCE rings) with a central channel that has been implicated in phospholipid transport. Furthermore, we utilised the novel Zernike3D and cryoDRGN algorithms to compute inherent continuous flexibility, subsequently revealing that Bd1152 possesses high structural plasticity, potentially providing insight into the mechanism of lipid transport.

Acknowledgments

Firstly, I would like to thank my PhD supervisor, Dr Tim Knowles, for his guidance and support throughout my doctoral studies. His expertise, patience, and encouragement have been invaluable to me during this journey. I would also like to thank my co-supervisor, Prof Andrew Lovering, for his vital contribution to my PhD project. His scientific advice was instrumental in shaping my research.

The work presented in this thesis has relied on advice, support, and guidance from many researchers. I am particularly grateful to Dr Christos Savva for his willingness to teach me the techniques and principles of cryo-EM. His expertise has significantly contributed to obtaining the high-resolution structures presented in this work. I thank Dr Simon Caulton for his help with X-ray crystallography, Dr Jack Bryant for the guidance in the field of genetics, and Prof Julian Bergeron for his assistance with cryo-EM data processing.

I am enormously grateful to Prof Jose Maria Carazo for granting me the opportunity to work in his laboratory on cryo-EM data processing at the I2PC centre. I am also incredibly thankful to Dr Marcos Gragera for his patience, guidance, and teaching me to see in 39-dimensional space. I am thankful to Mikel Iceta and Patricia Losana for creating such a welcoming and supportive environment to work in. Thank you all for an unforgettable experience. Scipion indeed changed my life.

Thank you to the Knowles group, both present and former members, Dr Peter Wotherspoon, Dr Gareth Hughes, Dr David Hardy, Rachel Holyfield, Hannah Johnson and Pooja Sridhar as well as the rest of 7th floor. It was so much fun to work with all of you!

Thank you to my undergraduate supervisor Prof Nelson Fernandez for motivating me to pursue career in science and supporting me ever since.

Special thanks to my friends Bianca Diaconu, Sasina Ruenphirom, Sara Salvanes Lopez and Greta Gaidytė. I feel incredibly lucky to have met such supportive, understanding and kind people. Your friendship is greatly appreciated.

I would like to express my deepest gratitude and appreciation to my incredible family: my mum Eglė, dad Rokas and my brother Paulius. I am forever grateful for your love, sacrifices, and unwavering support. I could not have done this without you.

Finally, I am immeasurably grateful to my husband (and collaborator) Dr Ben Cooper. Your encouragement and confidence in my abilities have been instrumental in helping me achieve my goals. I cannot thank you enough for all that you have done for me. You indeed own 40% of this thesis.

I dedicate this work to my dad, who I miss every day. Your sense of humour and willingness to help others continues to inspire me.

The work in this thesis was funded by a Midlands Integrative Biosciences Training Programme studentship (grant no. BB/M01116jdjhashfirX/1), jointly funded by the BBSRC and University of Birmingham, BBSRC Research grant (grant no. BB/S017283/1) and Instruct-ERIC (grant no. PID2350 and PID23275).

Table of Contents

Abstract **2**

Acknowledgments **3**

Table of Contents **5**

List of Figures **11**

List of Tables **15**

List of Abbreviations **16**

1. Chapter 1 – General Introduction. **20**

1.1. Antimicrobial resistance. **21**

1.2. Architecture of the gram-negative cell envelope **22**

1.3. Outer membrane biogenesis **24**

1.3.1. Phospholipid biosynthesis 24

1.3.2. Transport across the cytoplasmic membrane. 27

1.3.3. Intermembrane transport 31

1.4. Phospholipid trafficking between the IM and OM **40**

1.4.1. The MCE family of proteins 42

1.4.2. The Mla system 45

1.4.3. PqiABC and LetAB 54

1.4.4. The Asma-like protein family 59

1.4.5. The Tol-Pal complex 61

1.5. *Bdellovibrio bacteriovorus*: a potential predator for combating antibiotic-resistant infections **62**

1.6. Bdellovibrio bacteriovorus life cycle **64**

1.6.1. Bdelloplast formation....65
1.6.2. The growth phase and release of progeny....66
1.7. Lipid transport in *Bdellovibrio bacteriovorus* HD100....67
1.8. Biological significance and aim of the project....68
2. Chapter 2 - Materials and Methods....69
2.1. Culture media and growth conditions....70
2.2. Construct design....71
2.2.1. Q5-site directed mutagenesis....73
2.2.2. Cloning....75
2.2.3. Agarose gel electrophoresis....78
2.2.4. DNA preparation and sequencing....78
2.3. Preparation of *E. coli* chemically competent cells....79
2.4. Bacterial transformation....79
2.5. Expression and purification of recombinant proteins....80
2.5.1. Growth conditions for both soluble and membrane proteins....80
2.5.2. Purification of soluble proteins....80
2.5.3. Purification of membrane associated complexes....81
2.5.4. SDS-PAGE analysis....82
2.5.5. Western blot....82
2.5.6. Thin layer chromatography....83
2.5.7. Formation of apo-proteins....83
2.5.8. The GraFix method....84
2.6. Cryo-EM and negative stain transmission electron microscopy (TEM)....85
2.6.1. Preparation of negative stain grids....85
2.6.2. Preparation of graphene oxide coated grids....85

2.6.3. Preparation of Cryo-EM grids. 86
2.6.4. Cryo-EM sample optimisation. 86
2.6.5. Data collection and processing....87
2.7. X-ray crystallography....89
2.7.1. Crystal growth 89
2.7.2. Data collection and processing....90
2.8. Identification of *Bdellovibrio bacteriovorus* MCE proteins. 91
2.9. Structural prediction 91
2.10. Foldseek analysis 92
2.11. Zernike3D 92
2.12. CryoDRGN 93
2.13. Model building 93
3. Chapter 3 - Identification of MCE domain-containing proteins within *Bdellovibrio bacteriovorus* HD100 94
3.1. Introduction 95
3.2. Chapter Aims....97
3.3. Results 98
3.3.1. Identification of MCE domain-containing proteins within *B. bacteriovorus HD100,* prediction of protein domain architectures and neighbouring gene analysis....98
3.3.2. Structural prediction 101
3.3.3. Abundance of 2-MCE domain-containing proteins 103
3.3.4. Identification of differentially expressed genes within the *B. bacteriovorus* HD100 operons 578, 888, 1612, 1750 and 1958. 105
3.4. Discussion 107

3.4.1. A novel porin-like fold is predicted at the c-terminus of Bd3120.....107
3.4.2. Single MCE domain-containing proteins in *B. bacteriovorus*. 109
3.4.3. The unique characteristics and potential functional roles of the Bd1153-Bd1151 operon. 110
3.4.4. The role of MCE proteins within *B. bacteriovorus*.....111

3.5. Selection of *B. bacteriovorus* MCE domain-containing protein for experimental characterisation.....112

3.6. Conclusion.....113

4. Chapter 4 – Cryo-Electron Microscopy of Bd1152. 114

4.1. Introduction 115

4.1.1. Sample preparation for TEM. 116
4.1.2. The microscope. 119
4.1.3. Basic principles of image formation in TEM 125
4.1.4. Phase contrast and Contrast Transfer Function. 127
4.1.5. Data collection and image processing. 131

4.2. Chapter Aims.....139

4.3. Results 140

4.3.1. Structural investigation of membrane associated Bd1153-Bd1152 protein complex. 140
4.3.2. Structural investigation of the soluble $Bd1152^{37\text{-}283}$ construct 152
4.3.3. Structural investigation of the $Bd1152^{37\text{-}283\ I112C\ L238C}$ construct. 158

4.4. Discussion 182

4.4.1. Challenges in structural characterisation of Bd1153. 182
4.4.2. Generating a more stable Bd1152 construct. 183
4.4.3. Issues in the cryo-EM data processing of the $Bd1152^{2C}$ sample. 184

4.5. Conclusion.....190

5. Chapter 5 - X-ray crystallography of Bd1152.....191

5.1. Introduction 192

5.1.1. Crystallisation 192

5.1.2. Diffraction 195

5.1.3. Phasing and crystal structure determination 197

5.2. Chapter aims. 199

5.3. Results 200

5.2.1. Crystallisation of Bd1153-Bd1152, $Bd1152^{37-283}$ and $Bd1152^{2C}$ 200

5.2.2. The Bd1152 construct optimisation for crystallisation. 202

5.2.3. Expression and purification of the $Bd1152^{2C2A}$ construct. 204

5.2.4. Structural determination of $Bd1152^{2C2A}$ 206

5.2.5. Lipophilicity of Bd1152 and predicted key residues for lipid binding. 214

5.1.3. Comparison between X-ray and cryo-EM structures of Bd1152. 217

5.3. Discussion 218

5.3.1. Structural comparison of the Bd1152 and *E. coli* PqiB MCE rings. 218

5.3.2. Strategies to overcome the challenges of Bd1152 crystallisation. 220

5.3.3. Flexibility and conservation of the Bd1152 PLL regions. 223

5.4. Conclusion 224

6. Chapter 6 - Investigating the Bd1152 conformational landscape from Cryo-EM data 225

6.1. Introduction 226

6.2. Chapter aims. 229

6.3. Results 230

6.3.1. 2D and 3D classification suggest conformational heterogeneity. 230

6.3.2. Multi-body refinement 232

6.3.3. Zernike3D 236

6.3.4. CryoDRGN 245

6.3.5. Variation in interdomain distance 247

6.4. Discussion 249

6.4.1. Zernike3D revealed high global dynamics in Bd1152^{2C}. 249

6.4.2. CryoDRGN analysis revealed the opening and closing of the central Bd1152 channel. 251

6.4.3. Interdomain distance variability. 253

6.4.4. Limitations within the study 254

6.4.5. Future perspectives 256

6.5. Conclusion 257

7. Chapter 7 - Final discussion and future work. 258

7.1. Discussion 259

7.1.1. Bd1152 – a putative lipid transport protein 259

7.1.2. The c-terminus of Bd1152 suggests a unique functional mechanism. 261

7.1.3. The role of Bd1151 and Bd1153 265

7.1.4. Other MCE domain-containing proteins within *Bdellovibrio bacteriovorus* HD100. 270

7.2. Concluding remarks. 271

8. Appendix 272

Bibliography 295

List of Figures

Figure 1.1. General architecture of the gram-negative bacteria cell envelope. 22

Figure 1.2. Phospholipid biosynthesis in E. coli 25

Figure 1.3. An overview of the major protein translocation pathways across the inner membrane of gram-negative bacteria. 28

Figure 1.4. An overview of the intermembrane translocation pathways in gram-negative bacteria. 32

Figure 1.5. The general structure of lipopolysaccharide (LPS). 33

Figure 1.6. The transport of lipopolysaccharide (LPS) across the inner membrane via the MsbA flippase in E.coli. 35

Figure 1.7. The proposed model of β-barrel assembly by the Bam complex. 39

Figure 1.8. The cryo-EM structure of the M. smegmatis Mce1 ABC transport complex (PDB 8FEF). 44

Figure 1.9. The Maintenance of Lipid Asymmetry (Mla) pathway in E. coli. 46

Figure 1.10. The structure of E. coli MlaD (PDB:7CH0) in its respective cellular location within the inner membrane (IM.) 51

Figure 1.11. E. coli let and pqi operon architectures. 54

Figure 1.12.The structures of E. coli LetB and LetA. 56

Figure 1.13. The structures of E. coli PqiB, PqiA and PqiC. 58

Figure 1.14. The AsmA-like protein family. 60

Figure 1.15. Bdellovibrio bacteriovorus life cycle. 64

Figure 3.1. The top four (I-IV) MCE domain architectures across bacterial kingdom. 96

Figure 3.2. The architectures of operons in Bdellovibrio bacteriovorus HD100 that harbouring MCE proteins. 100

Figure 3.3. The Alphafold structures of all MCE-domain proteins within Bdellovibrio bacteriovorus HD100. 102

Figure 3.4. Sankey diagram demonstrating the distribution of 2-MCE domain-containing proteins across all kingdoms of life, as identified by FoldSeek analysis. 104

Figure 3.5. Gene expression levels of the B. bacteriovorus HD100 operons containing mce genes during the attack phase (AP) and growth phase (GP), as reported by Karunker et al., (2013). 106

Figure 3.6. Structural comparison of the c-terminal porin-like feature of Bdelovibrio bacteriovorus HD100 Bd3120 and E. coli YfaZ. 108

Figure 4.1. Standard sample preparation for single particle cryo-EM and negative stain analysis. 118

Figure 4.2. Simplified representation of transmission electron microscope 120

Figure 4.3. Schematic representation of TEM lens aberrations. 123

Figure 4.4. Basic representation of the main electron scattering events within a thin TEM specimen. 126

Figure 4.5. Phase contrast and Contrast Transfer Function in cryo-EM. 129

Figure 4.6. Defocus effect on the CTF function. 130

Figure 4.7. CTF Estimation and Phase Flipping. 133

Figure 4.8. The Fourier Slice Theorem. 137

Figure 4.9. Purification of the Bd1153-Bd1152 membrane-associated complex. 141

Figure 4.10. Simplified workflow of the Bd1153-Bd1152 cryo-EM data processing. 144

Figure 4.11. The final cryo-EM maps obtained from the Bd1153-Bd1152 dataset. 147

Figure 4.12. Comparison of the Bd1152 top and bottom MCE rings obtained from the Bd1153-Bd1152 cryo-EM dataset. 148

Figure 4.13. Simplified workflow of the Bd1153-Bd1152 cryo-EM data processing using cryoSPARC v.2.12.0. 151

Figure 4.14. Purification of $Bd1152^{37-283}$. 154

Figure 4.15. Simplified workflow of the $Bd1152^{37-283}$ cryo-EM data processing using Relion 3.1.3. 157

Figure 4.16. The Alphafold structure of Bd1152 (residues 37-283) 159

Figure 4.17. Purification of $Bd1152^{2C}$ 161

Figure 4.18. Simplified workflow of the $Bd1152^{2C}$ cryo-EM data processing using Relion 4.0. 164

Figure 4.19. The 2D class averages of the $Bd1152^{2C}$ particle set obtained after Xmipp3 picking consensus. 166

Figure 4.20. Simplified workflow of the $Bd1152^{2C}$ cryo-EM data processing using Scipion 3.0.26. 169

Figure 4.21. The final cryo-EM map of the $Bd1152^{2C}$ top MCE ring at 3.4 Å in C6 symmetry. 170

Figure 4.22. The cryo-EM model of the $Bd1152^{2C}$ top MCE ring obtained from ModelAngelo and Phenix real space refinement. 172

Figure 4.23. The cryo-EM data processing approaches utilised to address the symmetry mismatch between the $Bd1152^{2C}$ top and bottom MCE rings. 174

Figure 4.24. Comparison of the $Bd1152^{2C}$ MCE ring densities obtained from non-uniform refinement in C6 and C3 symmetries. 175

Figure 4.25. The impact of fluorinated octyl maltoside (FOM) on the $Bd1152^{2C}$ particle distribution. 177

Figure 4.26. The cryo-EM data processing results of the Bd1152^{2C} sample spiked with 0.5 x CMC FOM (fluorinated octyl maltoside) 178

Figure 4.27. Purification of the Bd1152^{2C} GraFix sample, 180

Figure 4.28. Cryo-EM of the Bd1152^{2C} GraFix sample. 181

Figure 4.29. Geometrical representation of imposing C6 and C3 symmetries during the refinement of Bd1152$^{2C.}$ 187

Figure 5.1. Crystallisation phase diagram. 193

Figure 5.2. The process of Vapour Diffusion in hanging drop and sitting drop arrangements. 194

Figure 5.3. Bragg's law of diffraction. 196

Figure 5.4. Examples of Bd1152^{2C} crystals under different crystallisation conditions. 201

Figure 5.5. Head-to-head dimerisation of Bd1152^{2C}. 203

Figure 5.6. Purification of Bd1152^{2C2A}. 205

Figure 5.7. Crystals that resulted in the Bd1152^{2C2A} structure. 206

Figure 5.8. Crystal structure of the Bd1152^{2C2A} hexamer. 209

Figure 5.9. The electron density in the final refined map for the Bd1152^{2C2A} top (A) and bottom (B) MCE domains. 210

Figure 5.10. Comparison of the Bd1152 Alphafold model and the Bd1152^{2C2A} crystal structure. 213

Figure 5.11. Molecular Lipophilicity Potential of Bd1152^{2C2A}. 215

Figure 5.12. Sequence conservation of Bd1152. 216

Figure 5.13. Structural comparison of the Bd1152 top MCE ring obtained from X-ray crystallography and cryo-EM. 217

Figure 5.14. Structural comparison of E. coli PqiB and Bd1152^{2C2A}. 219

Figure 5.15. Bd1152^{2C2A} arrangement within the crystal lattice showing head-to-tail arrangement. 222

Figure 6.1. The framework of the Zernike3D algorithm. 227

Figure 6.2. 2D and 3D classification of Bd1152^{2C} and Bd1152 indicating conformational heterogeneity. 231

Figure 6.3. Conformational dynamics of Bd1152 determined through multi-body refinement. 233

Figure 6.4. Conformational dynamics of Bd1152^{2C} determined through multi-body refinement. 234

Figure 6.5. The data pre-processing steps to obtain an input particle set and the reference map for the Zernike3D analysis. 237

Figure 6.6. The cryo-EM map of Bd1152^{2C} used as the reference for the Zernike3D analysis. _____ *238*

Figure 6.7. The Zernike3D analysis of Bd1152^{2C}. _____ *240*

*Figure 6.8. Flexible IMODFIT fitting of the Bd1152^{2C2A} crystal structure into each of the Bd1152^{2C} Zernike3D maps.*_____ *242*

Figure 6.9. Structural dynamics of Bd1152^{2C} protein derived from the Zernike3D analysis and the IMODFIT flexible fitting of the Bd1152^{2C2A} crystal structure. _____ *245*

Figure 6.10. Dynamics of Bd1152^{2C} obtained from the CryoDRGN analysis. _____ *246*

Figure 6.11. The variation in the interdomain distances between the MCE rings of Bd1152 and E. coli PqiB._ 248

Figure 6.12. Comparison of the LetB and Bd1152 pore lining loop conformations. _____ *252*

Figure 7.1. An example of the operon architecture comprising 2-MCE domain-containing proteins. _____ *260*

*Figure 7.2. Cartoons representing structures of MCE proteins in their respective cellular locations.*_____ *262*

Figure 7.3. The Alphafold structures of 2-MCE domain-containing proteins from different bacterial species. 263

Figure 7.4. Alphafold predictions of several small membrane proteins. _____ *266*

Figure 7.5. The Alphafold structures of E. coli PqiA (AF-P0AFL9-F1) and LetA (AF-P0AD03-F1) with cysteine residues indicated in blue. _____ *267*

Figure 7.6. Structural comparison of Bd1153 with the E.coli PqiA and LetB proteins. _____ *269*

List of Tables

Table 2.1. E. coli strains used in this study. _____ *71*

*Table 2.2. DNA constructs used in this study.*_____ *71*

*Table 2.3. Primers used in this study.*_____ *72*

*Table 2.4. Q5-site directed mutagenesis reaction constituents (NEB)*_____ *73*

*Table 2.5. Thermocycler conditions for PCR*_____ *74*

*Table 2.6. Kinase, ligase and Dpn1 (KLD) treatment.*_____ *74*

*Table 2.7. PCR reaction constituents for PCR with Phusion High-Fidelity DNA polymerase.*_____ *75*

Table 2.8. Double restriction digest reaction constituents. _____ *77*

Table 2.9. 74 DNA ligation constituents using a molar ration of 1:5 pET26b(+) vector and the Bd1153-Bd1152 insert. _____ *77*

*Table 2.10. Cryo-EM data collection parameters used in this study.*_____ *88*

Table 5.1. Bd1152^{2C2A} structure data collection and refinement. _____ *211*

List of Abbreviations

AF	Alphafold
AP	Attack phase
ATP	Adenosine triphosphate
BALOs	*Bdellovibrio*-and-like organisms
Bam	β-barrel assembly machinery
BPA	p-benzoyl-L-phenylalanine
BTP	Bis-tris-propane
CAMP	Cationic antimicrobial peptides
CL	Cardiolipin
CNN	Convolutional neural networks
Cryo-EM	Cryogenic electron microscopy
CTE	C-terminal extension
CTF	Contrast transfer function
DDM	n-Dodecyl-β-D-maltoside
DM	n-Decyl-β-D-maltoside
EDTA	Ethylenediaminetetraacetate
ERSF	European radiation synchrotron facility
FEG	Field emission gun
FOM	Fluorinated octyl maltoside
FSC	Fourier shell correlation
FT	Fourier transform
G3P	Glycerol-3-phosphate
GlcNAc	*N*-acetylglucosamine
GO	Graphene oxide
GP	Growth phase
GTP	Guanosine-5'-triphosphate

HAs	Heavy atoms
HRP	Horseradish peroxidase.
IM	Inner membrane
IMPs	Inner membrane proteins
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LB	Lysogenic broth
Lgt	Phosphatidylglycerol/prolipoprotein diacylglyceryl transferase
LLG	Log likelihood gain
Lnt	Phospholipid/apolipoprotein transacylase
Lol	Localisation of lipoprotein
LPS	Lipopolysaccharide
Lpt	Liposaccharide transport pathway
LSB	Lauryl-sulfobetaine
Lsp	Lipoprotein signal peptidase
MCE	Mammalian cell entry
MDR	Multi-drug resistant
MurNAc	*N*-acetylmuramic acid
MWCO	Molecular weight cut off
NBD	Nucleotide binding domain
NMA	Normal mode analysis
NTA	Nitrilotriacetic acid
NTHi	Nontypeable haemophilus influenzae
OM	Outer membrane
OMP	Outer membrane protein
OMV	Outer membrane vesicle
PA	Phosphatidic acid

PAGE	Polyacrylamide gel electrophoresis
PC	Principal component
PE	Phosphatidylethanolamine
PEG	Polyethylene glycol
PG	Phosphatidylglycerol
PGP	Phosphatidylglycerol phosphate
Pgs	Phosphate synthase
PL	Phospholipid
PLLs	Pore lining loops
PMA	Phosphomolybdic acid
PMF	Proton motif force
PS	Phosphatidylserine
Psd	Phosphatidylserine decarboxylase
PSF	Point spread function
Pss	Phosphatidylserine synthase
RMSD	Root mean square difference
RPKM	The reads per kilobase of transcript per million mapped reads
SDS	Sodium dodecyl sulphate
Sec	Secretion
SEC	Size exclusion chromatography
SMPs	Small Membrane Proteins
SOC	Super optimal broth with catabolite repression
SPA	Single particle analysis
SRP	Signal recognition particle
TAT	Twin arginine translocation
TFZ	Template-free model quality assessment

TLC	Thin layer chromatography
TMD	Transmembrane domain
TNF-α	Tumour necrosis factor alpha
Tween 20	Polyoxyethylene 20
UMAP	Uniform manifold approximation and projection
WT	Wild type
XDR	Extensively drug-resistant
β-OG	n-Octyl β-D-glucopyranoside

Chapter 1 – General Introduction.

1.1. Antimicrobial resistance.

Multi-drug resistant (MDR) and extensively drug-resistant (XDR) bacteria pose an enormous and growing global health threat, characterised by limited antibiotic options and high mortality rates (Chen *et al.,* 2021; Jean *et al.,* 2022; Wang *et al.,* 2021). Indeed, 1.27 million deaths were attributed to antibiotic resistance in 2019 (Murray *et al.,* 2022), consequently ranking it as the third leading cause of death.

Antibiotics have been developed to target essential bacterial processes, including cell wall synthesis (e.g., β-Lactams and vancomycin), protein synthesis (e.g., Aminoglycosides), and DNA synthesis (e.g., Coumarins) (Hooper, 2001). The efficiency of antibiotics relies on their ability to reach their target destination within the cell. This is particularly challenging for gram-negative bacteria due to their complex cell envelope, consisting of three principal layers: the outer membrane (OM), the periplasm (containing peptidoglycan cell wall), and the inner membrane (IM). In contrast, gram-positive bacteria have a single lipid membrane and a thick peptidoglycan layer. The intrinsic double membrane architecture of gram-negative bacteria confers additional resistance to a variety of antibiotics that are effectively utilised to treat gram-positive infections (Breijyeh *et al.,* 2020).

The cell envelope of gram-negative bacteria has been extensively studied as a potential target for antimicrobial development. Despite this, the biogenesis of the cell envelope remains poorly understood, warranting further investigation to improve our understanding of this process and potentially identify new targets for antimicrobial drug development.

1.2. Architecture of the gram-negative cell envelope

The unique, multi-layer architecture of gram-negative bacteria functions as an effective permeability barrier against a variety of hydrophobic molecules, including antimicrobial agents. The general architecture of the gram-negative bacteria cell envelope is depicted in Figure 1.1.

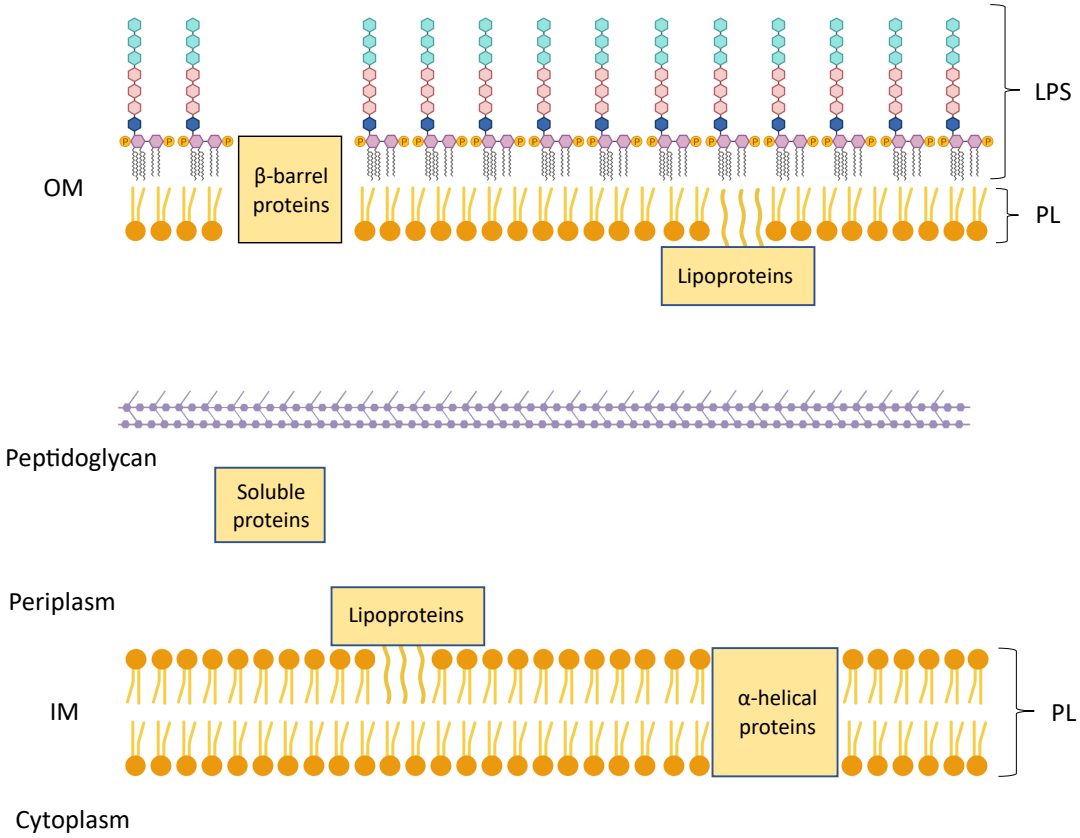


Figure 1.1. General architecture of the gram-negative bacteria cell envelope.

The inner membrane (IM) is composed of a symmetric arrangement of phospholipids (PLs), and accommodates both integral membrane proteins and lipoproteins. The outer membrane (OM) exhibits asymmetry, with lipopolysaccharide (LPS) located in the outer leaflet and PLs situated in the inner leaflet. The OM accommodates lipoproteins and outer membrane proteins. The peptidoglycan is composed of alternating N-acetylglucosamine (NAG) and N-acetylmuramic acid (NAM), cross-linked by short peptides.

The OM exhibits a unique asymmetric arrangement of lipids, with lipopolysaccharide (LPS) located exclusively in the outer leaflet and phospholipids (PLs) present within the inner leaflet (Funahara and Nikaido, 1980). The OM also accommodates both outer membrane proteins (OMPs) and lipoproteins.

OMPs are typically characterised by a β-barrel conformation and often form pores, such as OmpA and OmpC, which allow passive diffusion of small molecules (Singh and Verma, 2008). OMPs also play a crucial role in the folding and insertion of other OMPs (Hagan *et al.,* 2011) as well as contributing to the overall integrity of the OM (Xie *et al.,* 2016). Lipoproteins contain lipid moieties that anchor them to the inner leaflet of the OM *(Koebnik et al.,* 2000). They are involved in variety of cellular functions including cell wall synthesis and secretion systems (Grabowicz and Silhavy, 2017).

The IM and OM are separated by the periplasm (13-25nm) (Seltmann and Holst, 2002), an aqueous compartment that accommodates the peptidoglycan cell wall responsible for maintaining cellular integrity. Peptidoglycan is composed of *N*-acetylglucosamine (GlcNAc) and *N*-acetylmuramic acid (MurNAc), which are cross-linked by short peptides (Wet *et al.,* 2008) and attached to the OM via the lipoprotein Lpp (Seltmann and Holst, 2002).

The IM bilayer is composed entirely of PLs and accommodates both lipoproteins and inner membrane proteins (IMPs), typically exhibiting an α-helical conformation. IMPs are required for essential cellular functions including energy production, lipid biosynthesis and protein secretion (Luirink *et al.,* 2012).

It is noteworthy that PL composition varies among different bacterial species; nevertheless, the model organism *E. coli* predominantly contains phosphatidylethanolamine (PE) (~75%), phosphatidylglycerol (PG) (~20%) and cardiolipin (CL) (~5%) (Raetz and Dowhan, 1990).

1.3. Outer membrane biogenesis

1.3.1. Phospholipid biosynthesis

The process of lipid bilayer formation involves the biogenesis of phospholipids and their subsequent insertion into pre-existing membranes. This process is facilitated by a complex array of biosynthetic pathways, which rely on the activity of a number of enzymes.
The research of Eugene Kennedy (Kennedy *et al.,* 1976) established *E. coli* as the model organism for studying PL synthesis, metabolism and function. Figure 1.2 provides an overview of the pathways involved in phospholipid biosynthesis in *E. coli*.

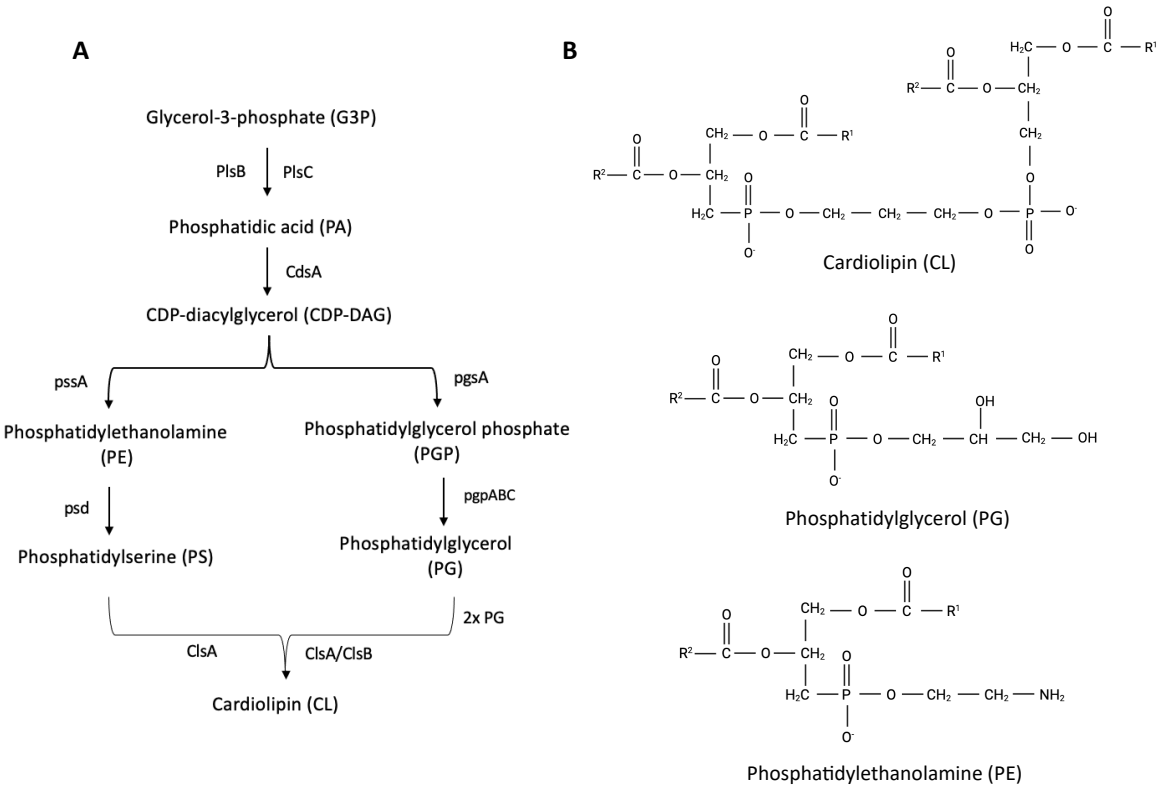


Figure 1.2. Phospholipid biosynthesis in *E. coli*

(A) Kennedy pathway for phospholipid biosynthesis in *E. coli*. PlsB - G3P acyltransferase; PlsC - Acyl-G3P acyltransferase; CdsA - CDP-DAG synthase; Pss - phosphatidylserine synthase; Pgs - phosphatidylglycerophosphate synthase; Psd - phosphatidylglycerophosphate synthase; PgpA-C phosphatidylglycerophosphate phosphatases that are encoded with three genes, ClsA/ClsB – cardiolipin synthase. (Right) Chemical formulas of major phospholipid species in *E. coli*: Cardiolipin (CL), phosphatidylethanolamine (PE), phosphatidylglycerol (PG). Figure reproduced from Cooper (2023).

In *E. coli* and other proteobacteria, phosphatidic acid (PA) is generated by the two-step acylation of glycerol-3-phosphate (G3P), utilising acyl-ACP as the acyl chain donor. Firstly, the acyltransferase PlsB catalyses acylation at the 1-position of G3P, followed by PlsC catalysed acylation at the 2-position of G3P. PA is the simplest of phospholipids and, hence, it is a precursor for the synthesis of more complex phospholipids. Nevertheless, PA is kept at extremely low levels due to the activity of CdsA, which rapidly converts PA into the intermediate molecule CDP-diacylglycerol.

CDP-diacylglycerol is then utilised to generate both phosphatidylserine (PS) and phosphatidylglycerol phosphate (PGP) by serving as phosphatidyl donor moieties to L-serine and G3P, respectively. These reactions are catalysed by phosphatidylserine synthase (Pss) and phosphatidylglycerol phosphate synthase (Pgs) to generate PS and PGP, respectively.

PgpA, PgpB and PgpC catalyse the rapid dephosphorylation of PGP to generate phosphatidylglycerol (PG), whilst phosphatidylserine decarboxylase (Psd) catalyses the decarboxylation of PS to generate phosphatidylethanolamine (PE). Cardiolipin synthases (ClsB and ClsC) catalyse the formation of cardiolipin (CL) from one molecule each of PG and PE, while the ClsA forms CL from two molecules of PG (Dowhan and Bogdanov, 2021).

1.3.2. Transport across the cytoplasmic membrane.

All cell envelope components are synthesised either in the cytoplasm or in close proximity to the cytoplasmic surface of the IM. That being the case, components must be transported to various target sites within the cell envelope following synthesis. The two membranes comprising the gram-negative cell envelope, however, are physically separated by ~220 Å of periplasm. This represents a vast challenge for the intermembrane transport systems as no cellular energy resources exist beyond the confines of the IM. Furthermore, major constituents of the OM are strongly amphipathic, hence necessitating periplasmic transport pathways that shield their hydrophobic portions from the aqueous environment (Shrivastava and Chng, 2019).

The first step in the OM biogenesis is to flip the targeted components across the IM. The main passageway that nascent proteins utilise for their transport into or across the IM is known as Secretion (Sec) pathway. After protein translation at the ribosome, two possible Sec translocation tactics can be employed. The co-translational signal recognition particle (SRP) - dependent pathway specifically targets IMPs, whilst the post-translational pathway (SecB dependent) targets OMPs and lipoproteins.

Furthermore, apart from the Sec pathway, there are two additional systems that facilitate the transport of proteins across the IM. The twin arginine translocation (TAT) pathway transports fully folded proteins, whilst YidC is responsible for the membrane insertion of small IMPs. Figure 1.3. provides an overview of the protein translocation pathways across the IM of gram-negative bacteria.

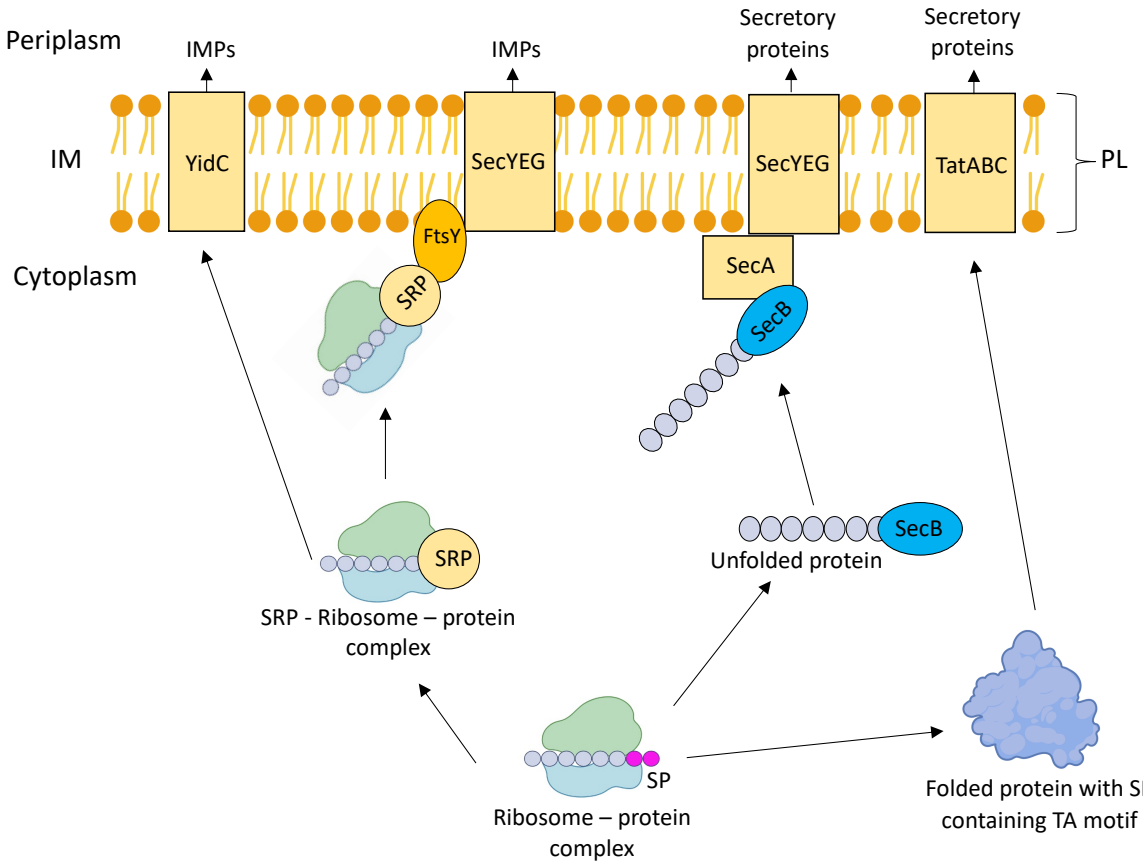


Figure 1.3. An overview of the major protein translocation pathways across the inner membrane of gram-negative bacteria.

Four different pathways are depicted in the figure. (1) The signal recognition particle (SRP) - dependent pathway is responsible for transporting inner membrane proteins (IMPs). This initiates with the recognition of the signal peptide (SP) on the ribosome, forming an SRP-ribosome-protein complex. This complex interacts with the FtsY receptor and SecYEG. (2) The post-translational pathway (SecB depended) targets outer membrane proteins (OMPs) and lipoproteins. SecB binds to the SP of the polypeptide, keeping it in an unfolded state until it is secreted by the SecYEG pathway. (3) YidC is involved in the membrane insertion of small IMPs, which are predicted to be targeted by SRP. (4) The twin arginine translocation (TAT) pathway transports folded proteins across the inner membrane (IM) that possess twin arginine (TA) motifs.

1.3.2.1. The signal recognition particle (SRP) – dependent pathway

The SRP pathway is initiated when SRP binds to the first transmembrane segment of the nascent peptide, subsequently forming an SRP-ribosome-protein complex in a GTP (Guanosine-5'-triphosphate) - dependent manner (Akopian *et al.,* 2013). This complex then moves towards the IM, where it interacts with the FtsY receptor (Angelini *et al.,* 2005), resulting in GTP hydrolysis and dissociation of the SRP. At the IM, the ribosome-protein complex interacts with the largest SecYEG subunit, i.e., SecY (Gumbart *et al.,* 2009), and induces the opening of the SecYEG channel (Knyazev *et al.,* 2013). It was suggested that opening of the SecYEG channel might expose the TM portions to the hydrophobic environment of the lipid bilayer, allowing the IMPs to be inserted (Lycklama a Nijeholt and Driessen, 2012), however, the process remains largely enigmatic.

1.3.2.2. YidC facilitated membrane insertion of small IMPs.

YidC is a second IM translocase that is responsible for the membrane insertion of small inner membrane proteins independently from the Sec pathway (Serek *et al.,* 2004). It was suggested that some YidC substrates are targeted via SRP, as both the SRP-ribosome-protein complex and FtsY were shown to interact with YidC *in vitro,* suggesting a functional relationship (Kedrov *et al.,* 2016; Seitl *et al.,* 2014).

1.3.2.3. The post-translational SecB-dependent pathway

OMPs are synthesised with an n-terminal signal sequence that targets them for translocation across the IM. Such translocation is performed via a post-translational SecB-dependent pathway that transports proteins across the IM independently of the ribosome. The cytoplasmic chaperone SecB binds to the polypeptide chains and keeps the proteins in their unfolded state until there are secreted in a linear fashion through the SecYEG channel. The energy for translocation is provided by the ATPase SecA, which binds to SecB and SecYEG, consequently driving protein export (Young and Duong, 2019). This processes is further enhanced by the proton motif force (PMF) (Shiozuka *et al.,* 1990) and the SecYEG - associated SecDF-YajC membrane complex (Hand *et al.,* 2006; Tsukazaki *et al.,* 2011). During or immediately after translocation, signal peptidase SP1 removes the n-terminal signal peptide, resulting in a mature protein (Dalbey *et al.,* 2012).

1.3.2.4. The twin arginine translocation (TAT) pathway

It is necessary to note that some proteins are folded cytoplasmically; therefore, an alternative method of translocation has to be utilised. The twin arginine translocation (TAT) pathway was shown to transport folded proteins carrying signal peptides that contain a highly conserved twin-arginine motif, into the periplasm (Palmer and Berks, 2012). Despite ongoing research, the reason why certain proteins are targeted to the TAT system remains poorly characterised. Nevertheless, it was proposed that some proteins require cytoplasmic cofactors for proper folding and maturation (Palmer and Berks, 2012). Additionally, certain proteins might face challenges in maintaining their conformation in an unfolded state, potentially increasing the likelihood of misfolding and aggregation (Lee *et al.,* 2006).

1.3.3. Intermembrane transport

Once the components of the OM have been translocated across the IM, they must navigate through the crowded periplasm to reach their target destinations. Early models of intermembrane cargo transport proposed the existence of “Bayer bridges” – fusions between the IM and OM (Bayer, 1968). Several subsequent publications suggested that newly synthesised LPS and OMPs might be transported to the cell surface via these membrane fusion events in *Salmonella typhimurium* (Mühlradt *et al.,* 1973; Smit and Nikaido, 1978). Nonetheless, the “Bayer bridge” model was met with scepticism due to limitations in sample preparation and lack of further genetic or biochemical investigation (Kellenberger, 1990). A more detailed understanding of intermembrane transport emerged with the identification of LolA – a soluble chaperone that shields lipoprotein acyl chains from the aqueous periplasmic environment (Matsuyama *et al.,* 1995). This subsequently led to the identification of the Localisation of lipoprotein (Lol) pathway, as well as other semi-independent highly conserved transport pathways, including the liposaccharide transport pathway (Lpt) and β-barrel assembly machinery (Bam) (Figure 1.4).

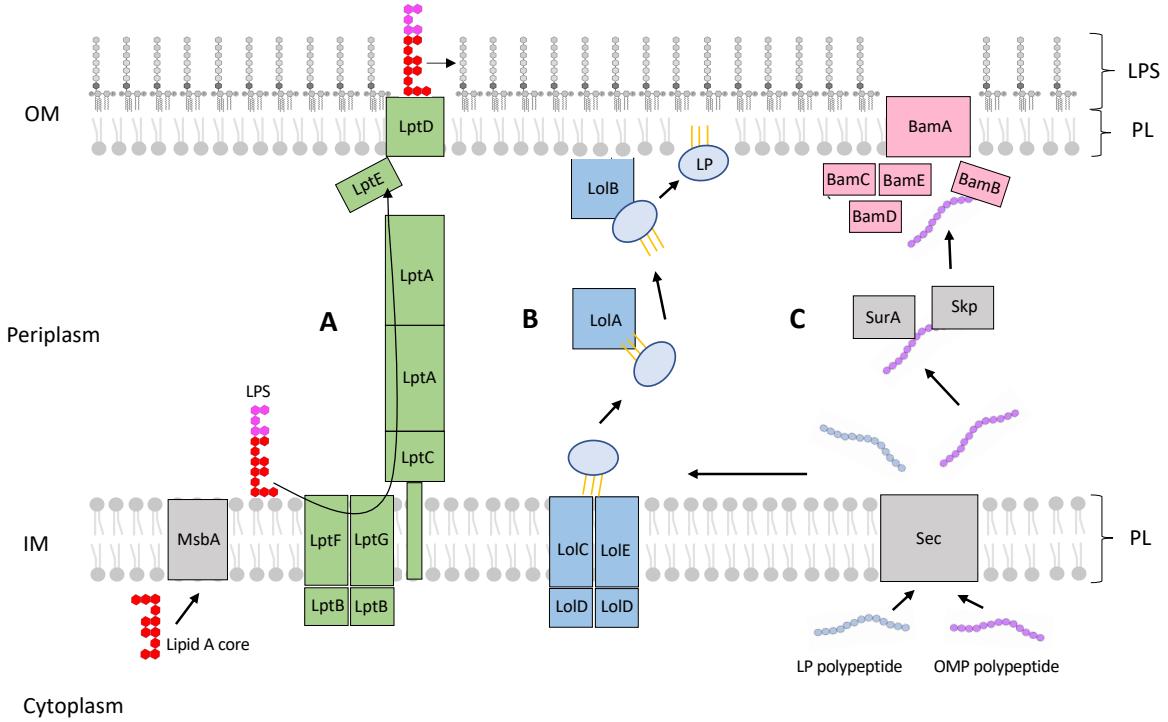


Figure 1.4. An overview of the intermembrane translocation pathways in gram-negative bacteria.

A) MsbA flips the Lipid A core across the inner membrane (IM) followed by O-antigen attachment, and the mature lipopolysaccharide (LPS) molecule is translocated to the outer leaflet of the outer membrane (OM) via the lipopolysaccharide transport (Lpt) pathway. B) Lipoproteins (LPs) are directed to the OM through the localisation of lipoproteins (Lol) system. C) The translocation of outer membrane proteins (OMPs) occurs via the cytoplasmic chaperones SurA/Skp, which carry unfolded OMPs to the β-barrel assembly machinery (Bam) complex. OMPs are inserted through BamA to form a β-barrel structure in the OM.

1.3.3.1. The Lpt machinery

LPS has a tripartite structure consisting of Lipid A, a hydrophobic portion that attaches to the OM, a core oligosaccharide made from non-repeating oligosaccharides and an O-antigen consisting of repeating oligosaccharide units (Figure 1.5).

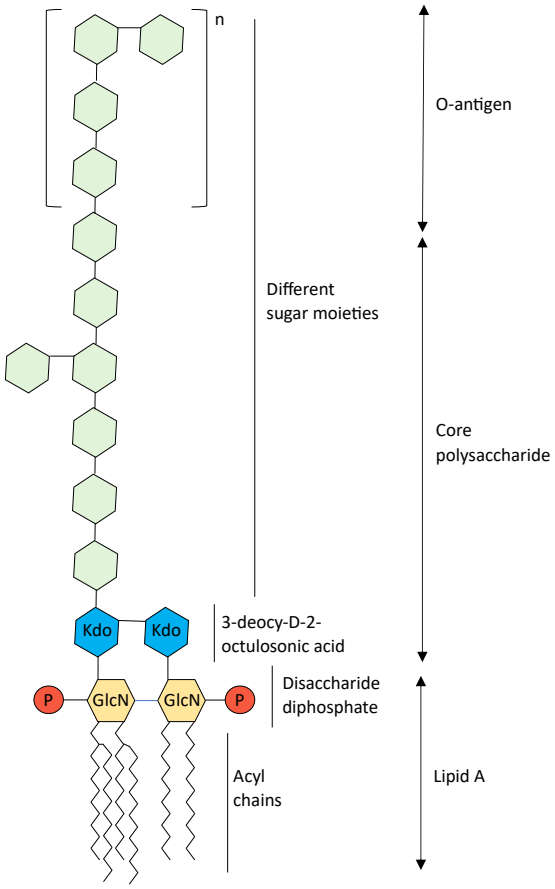


Figure 1.5. The general structure of lipopolysaccharide (LPS).

LPS consists of three main components: lipid A, core polysaccharide and O-antigen (repeats of polysaccharide chain). Abbreviations: GlcN - 2-amino-2-deoxy-glucose (glucosamine); Kdo - 3-deoxy-α-D- manno-oct-2-ulopyranosonic acid; P - phosphate.

All three components of LPS are synthesised on the cytoplasmic face of the IM, where lipid A and core oligosaccharides combine to form rough LPS. The O-antigen is synthesised independently and it is attached to a nascent LPS molecule at the periplasm-facing leaflet of the IM (Bertani and Ruiz, 2018).

Rough LPS is flipped across the IM by the ABC-transporter MsbA (Figure 1.6). The crystal structure of MsbA shows that it forms a homodimer, with each monomer containing a transmembrane domain (TMD) and a nucleotide binding domain (NBD) (Ward *et al.,* 2007). According to the most recent mechanistic model proposed, MsbA functions as a flippase, alternating between inward- and outward- facing conformations (Voss and Trent, 2018). In the apo state (no ATP bound), MsbA is open to the cytoplasm, allowing LPS to enter. The outward- facing conformation is adopted upon ATP binding, resulting in the dimerisation of the NBDs, which initiates conformational changes of TMDs, allowing LPS to be released in the periplasm. Once released by MsbA, LPS is anchored to the IM, where the O-antigen is added. Fully assembled LPS then continues its journey to the OM via the well-defined Lpt (lipopolysaccharide transport) machinery (Okuda *et al.,* 2016).

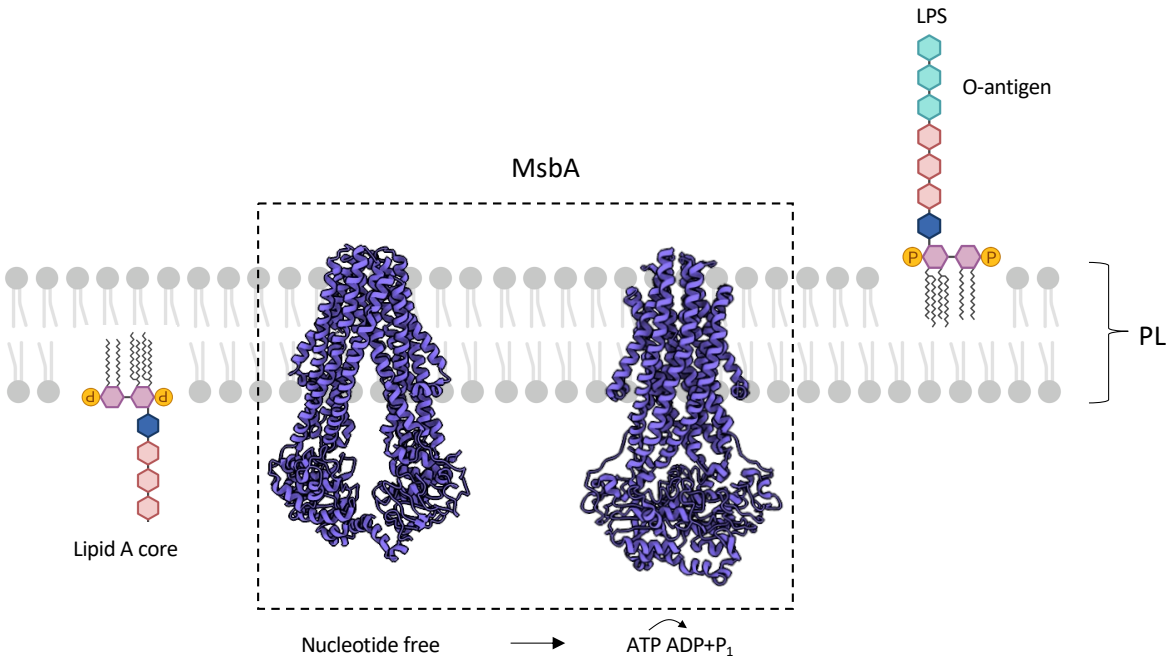


Figure 1.6. The transport of lipopolysaccharide (LPS) across the inner membrane via the MsbA flippase in *E.coli*.

The Lipid A core oligosaccharide is transported to the periplasmic face of the MsbA flippase. The “open” conformation of MsbA (nucleotide free) (PDB:5TV4) allows LPS to enter, whilst the “closed” conformation (ADP–vanadate-bound MsbA) (PDB 5TTP) allows LPS to be released to the periplasm. LPS is then anchored to the inner membrane (IM), where O -antigen is added. The structures are displayed in purple and were visualised with ChimeraX (Pettersen *et al.,* 2021). Figure adapted from Voss and Trent (2018).

In *E. coli,* the LPS transport system is composed of seven proteins organised into two subassemblies: $LptB_2CFG$ and LptDE, located in the IM and OM, respectively, which are connected through the periplasmic protein LptA. The inner membrane complex $LptB_2CFG$ extracts LPS from the IM, where it is picked up by LptA for periplasmic transfer. Several copies of LptA assemble to form an intermembrane channel that allows direct interaction with LptDE, as shown by crosslinking experiments (Freinkman *et al.,* 2012). Nevertheless, it is currently ambiguous how many LptA copies participate in the construction of this periplasmic bridge. The β-barrel protein LptD and the lipoprotein LptE, localised in the OM, receive LPS from the LptA bridge and incorporate it into the outer leaflet of the OM (Figure.1.4, A).

1.3.3.2. The Lol pathway

Lipoproteins are synthesised in the cytoplasm, containing an n-terminal signal sequence (SS) known as lipobox (Hayashi and Wu, 1990), which targets lipoproteins for translocation across the IM by the Sec/Tat translocon. Once a lipoprotein is transported, it remains attached to the outer leaflet of the IM via its SS moiety, allowing the maturation process to take place. Phosphatidylglycerol/prolipoprotein diacylglyceryl transferase (Lgt) adds a diacylglyceryl portion to the sulfhydryl group of the highly conserved cysteine residue located within the n-terminal SS. Lipoprotein signal peptidase (Lsp) cleaves the SS, subsequently resulting in a diacylglyceryl-containing lipoprotein with a new n-terminus referred to as the +1 position. Lastly, phospholipid/apolipoprotein transacylase (Lnt) adds another acyl chain to the amino group of cysteine, resulting in a fully mature triacylated lipoprotein that can be subsequently recognised by the Lol (localisation of lipoproteins) pathway.

In *E. coli*, the Lol pathway uses the residue adjacent to the lipidated cysteine to determine whether the lipoprotein will remain in the IM or be transported to the OM.
If the residue at position +2 is aspartate, the lipoprotein remains in the IM (Gennity and Inouye, 1991; Terada *et al.,* 2001), however, the position of aspartate residue varies across bacterial species (Lewenza *et al.,* 2008; Narita and Tokuda, 2007). When the lipoprotein does not contain the IM retention signal, the Lol pathway translocates it across the periplasm and inserts into the OM (El Rayes *et al.,* 2021) (Figure 1.4, B).

The Lol pathway is composed of five proteins (LolA-E). In the first step, the lipoprotein is extracted from the IM by the LolCDE complex and transferred to the periplasmic chaperone LolA in an ATP dependent manner. LolA delivers lipoproteins to the OM acceptor protein LolB, itself a lipoprotein, which facilitates OM lipoprotein insertion, although this process is poorly characterised (Konovalova and Silhavy, 2015; El Rayes *et al.,* 2021).

1.3.3.3. The Bam pathway.

OMPs are synthesised in the cytoplasm and translocated to the periplasm via the Sec pathway. OMPs interact with periplasmic chaperone proteins SurA or Skp through their hydrophobic moieties. These chaperones facilitate the transport of OMPs across the periplasm while ensuring they remain in an unfolded state, preventing misfolding and aggregation (Allen *et al.,* 2009). SurA appears to play a more prominent role, as evidenced by the significant reduction in OM density upon its deletion. Conversely, the absence of Skp has no discernible impact on OM composition (Sklar *et al.,* 2007).

DegP, another component required for the correct periplasmic transport of OMPs, is believed to function as a protease that degrades aggregated or misfolded OMPs in the periplasm, thereby reducing cellular toxicity (Ge *et al.,* 2014). It is noteworthy that OMPs can independently fold and insert themselves into synthetic lipid bilayers, however, this process is extremely slow and thus not biologically relevant (Burgess *et al.,* 2008). *In vivo*, the β-barrel assembly machinery (BAM) is employed to facilitate assembly and insertion of OMPs into the OM (Figure 1.4, C).

In *E. coli*, the core of the Bam complex is BamA, which is composed of an OM spanning β-barrel domain and five polypeptide translocation-associated (POTRA) domains that extend from the barrel into the periplasm. BamA associates with four lipoproteins: BamB, BamC, BamD and BamE. Although all proteins contribute to the efficient folding of substrates, only BamA and BamD are essential for the assembly of OMPs into the OM in *E.coli* (Hagan *et al.,* 2011). Crystal structures revealed the existence of both “open” and “closed” BamA conformations that are dependent upon the positioning of the POTRA domains relative to the BamA barrel (Noinaj *et al.,* 2013). This suggests the presence of a gated channel regulating access to the interior of the β-barrel domain.

A detailed model of the assembly mechanism by Bam has recently been suggested by Shen *et al.,* (2023). Essentially, the OMP substrate inserts itself into the BamA barrel via its C - terminus (Figure 1.7, A). The C-terminal β-strand of the OMP substrate forms an anti-parallel β-strand with the first β-strand of BamA, proximal to its opening (Figure 1.7, B). The OMP substrate then continues to fold, subsequently closing into a β-barrel (Figure 1.7, C, D). This consequently leads to the BamA β-barrel closure, as well as breaking its interactions with the OMP (Figure 1.7, E).

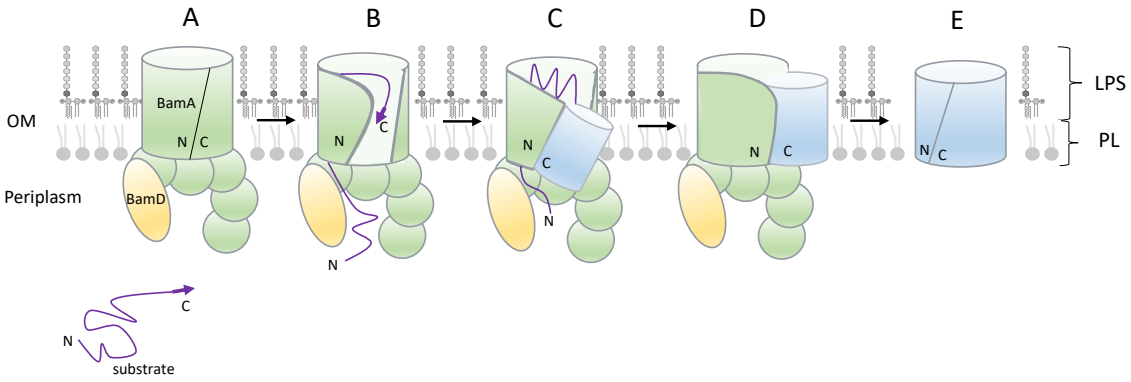


Figure 1.7. The proposed model of β-barrel assembly by the Bam complex.

A) The c-terminus of the substrate interacts with BamA. B) The c-terminus of the substrate interacts with the exposed n-terminus of BamA, consequently forming an anti-parallel β-strand. C) The substrate continuous to fold. D) The BamA β-barrel closes, breaking its interaction with the substrate. E) The OMP is released into the membrane environment, and its n- and c-terminal ends are joined. For simplicity, only BamA, BamD, and the substrate are shown. Figure adapted from (Shen *et al.,* 2023)

1.4. Phospholipid trafficking between the IM and OM

The first evidence of PL trafficking between membranes emerged in the 1970s, with the observation that radiolabelled PLs move between the IM and OM (Jones & Osborn., 1977). Despite ongoing research in this area for almost 50 years and considering that PLs are fundamental constituents of lipid bilayers, the mechanism and structural basis underlying PL trafficking across bacterial membranes remain unresolved.

The mechanism of PL translocation in the anterograde fashion (IM-to-OM) remains poorly understood. Nevertheless, it was suggested that the IM protein PbgA (formerly known as YejM) may play a role in the transfer of CL to the OM. This is supported by the observation that a *pbgA* mutant exhibited normal levels of CL in the IM, but no detectable levels of CL in the OM (Rossi *et al.,* 2017). The crystal structure of PbgA in complex with two CL molecules was determined (PDB:6V8Q) (Fan *et al.,* 2020), although the biological relevance of this complex is unclear. It should be noted that CL was supplemented during the crystallisation process, which is acknowledged as a limitation of the study by the authors (Fan *et al.,* 2020).

PbgA has since been implemented as a LPS sensor based on several functional studies (Fivenson and Bernhardt, 2020; Guest et al., 2020) as well as a crystal structure obtained in complex with a co-purifying LPS molecule (PDB: 6XLP) (Clairfeuille *et al.*, 2020). Nevertheless, the structure also revealed a potential binding cleft for a single phosphatidylethanolamine, which leaves the role of PbgA in lipid transport speculative.

The transport of PL between membranes, however, is not limited to anterograde directionality. The asymmetric organisation of the OM can be disrupted under stress conditions, such as exposure to antibiotics and chelating agents (Malinverni and Silhavy, 2009). As a consequence, LPS may be lost from the cell surface, forcing PLs to migrate from

the inner leaflet to the outer leaflet to fill the gaps. The removal of mispositioned PLs in the OM is essential for maintaining its asymmetry and eliminating perturbations that make gram-negative bacteria more susceptible to the entry of exogenous compounds (Malinverni and Silhavy, 2009). Over the last two decades, several systems have been reported to maintain OM lipid asymmetry, including the PldA and PagP proteins.

PldA maintains the membrane asymmetry by degrading PLs that accumulate in the outer leaflet of the OM. PldA is an OM phospholipase that exists as a monomer in its inactive state. The presence of PLs in the outer leaflet of the OM induces PldA dimerisation, resulting in a catalytically active protein that removes the sn-1 and sn-2 positioned fatty acid side chains from PLs (Dekker, 2000). PagP, another protein accommodated within the OM, transfers a palmitate chain from PLs to the lipid A component of LPS (Bishop *et al.,* 2000). The modification of lipid A with palmitate increases lipid A acylation, resulting in increased resistance to host cationic antimicrobial peptides (CAMPs) (Guo *et al.,* 1998).

An additional family of proteins, containing Mammalian Cell Entry (MCE) domains, has been identified and closely linked to intermembrane phospholipid trafficking in gram-negative bacteria (Malinverni and Silhavy, 2009). Extensive efforts have been invested in uncovering the role of MCE proteins in lipid transport and investigating the underlying mechanisms of bacterial phospholipid trafficking.

1.4.1. The MCE family of proteins

MCE proteins were first discovered when a chromosomal fragment containing the *mce1A* gene in *M. tuberculosis* was introduced into a non-pathogenic *E. coli* strain, allowing it to enter and survive in macrophages (Arruda *et al.,* 1993). This finding led to the designation of these proteins as “Mammalian Cell Entry (MCE)” and spurred a number of studies further investigating their potential role in the entry mechanism and virulence of *M. tuberculosis.* Shimono *et al.,* (2003) demonstrated that an *mce1A* mutant strain of *Mycobacterium bovis bacillus* Calmette-Guérin exhibited a reduced ability to enter HeLa cells, whilst *El-Shazly et al.,* (2007) identified that HeLa cells were able to uptake latex beads coated in the purified *M. tuberculosis* Mce3 protein, further emphasising the paramount importance of MCE proteins in bacterial invasion.

The *mce* genes in *Mycobacterium tuberculosis* are encoded within four operons (*mce*1-*mce*4), each consisting of six *mce* genes (ABCDEF). Each of the operons also contains the *yrbE* genes as well as genes encoding subunits of the ATPase-binding-cassette (ABC) transporter and a variable number of “accessory” proteins. Additionally, the function of the *M. tuberculosis* MCE transporter relies on the presence of the ATPase MceG (Joshi *et al.,* 2006) and the integral membrane protein LucA (Nazarova *et al.,* 2017, 2019), which are located outside the MCE operons.

In addition, studies by Shimono *et al.,* (2003) and Gioffré *et al.,* (2005) reported that *M. tuberculosis* strains lacking the *mce1, mce2 and mce3* operons exhibited an impaired ability to survive within mammalian cells. This finding was further supported by the expression profile of *M. tuberculosis*, which indicated that the *mce* operons are expressed later in infection (Talaat *et al.,* 2004).

Changes in the expression levels of *mce* operons have been also shown to occur in response to stress factors such as hypoxia (Rathor *et al.,* 2016; Rustad *et al.,* 2008), starvation (Betts *et al.,* 2002), oxidative stress (Gample *et al.,* 2019) and surface stress (e.g., SDS) (Manganelli *et al.*, 2001). Furthermore, several studies have suggested that *M. tuberculosis mce* operons encode lipid transport systems (Casali & Riley, 2007; Dunphy et *al.,* 2010; Pandey and Sassetti, 2008*)*. For instance, an altered lipid profile and accumulation of mycolic acids were reported in *M. tuberculosis mce1* mutant strain (Cantrell *et al.*, 2013; Forrellad *et al.*, 2014; Queiroz *et al.,* 2015). Furhtermore, in *M. smegmatis*, the Mce4 proteins have been shown to be essential for cholesterol uptake and utilisation (Klepp *et al.,* 2012). Nevertheless, experimental evidence supporting the role of *M. tuberculosis* MCE proteins in lipid remodelling is scarce, likely due to the complexity of protein expression and purification. Only recently, Chen *et al.,* (2023) published the cryo-EM structure of the Mce1 machinery from *M. smegmatis,* revealing that proteins assemble to form an ABC transporter complex with a c-terminal coiled-coil long enough to span the periplasm (Figure 1.8).

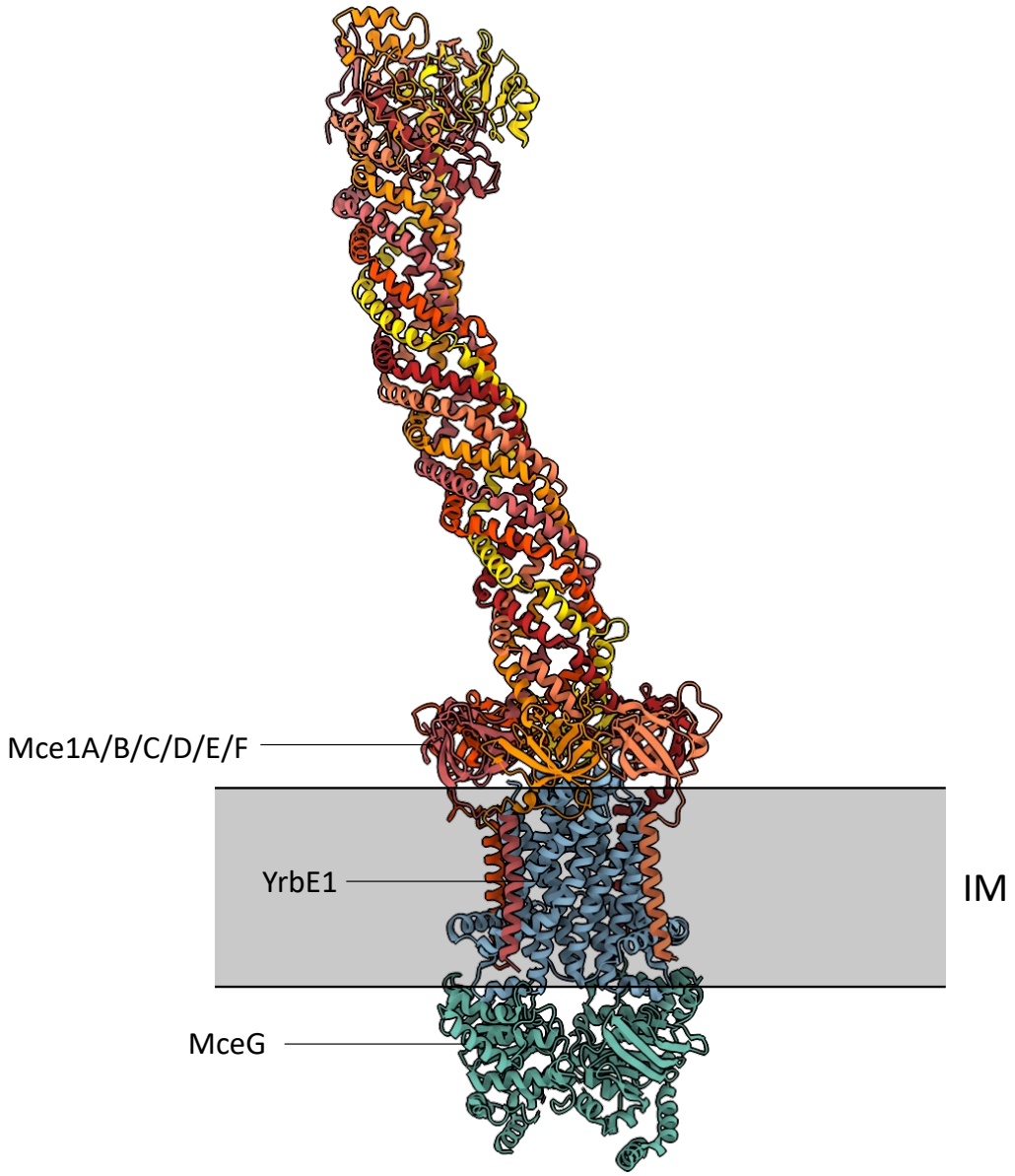


Figure 1.8. The cryo-EM structure of the *M. smegmatis* Mce1 ABC transport complex (PDB 8FEF).

Green colour indicates the ATPase MceG; blue colour indicates the ATP-binding cassette transporter YrbE; pink colour indicates the six MCE proteins Mce1A/B/C/D/E/F. The model was visualised with ChimeraX (Pettersen *et al.,* 2021).

Nonetheless, the role of MCE proteins in lipid transport has been characterised in more detail in Proteobacteria, with a number of recent publications disclosing both the structure and function of these putative lipid transporters (Chi *et al.,* 2020b; Coudray *et al.,* 2020; Hughes *et al.,* 2019; Isom *et al.,* 2020; Tang *et al.,* 2021). The most functionally and structurally studied MCE protein from *E.coli* is MlaD, a key member of the Maintenance of Lipid Asymmetry (Mla) pathway, which is proposed to remove mislocalised PLs from the outer leaflet of the OM (Malinverni and Silhavy, 2009)

1.4.2. The Mla system

A system identified to be crucial for the maintenance of outer membrane leaflet asymmetry in gram-negative bacteria is the Mla-OmpC (Maintenance of Lipid Asymmetry) pathway. In contrast to PldA and PagP, the Mla system is believed to transport mislocalised PLs in the outer leaflet of the OM back to the IM. The components of the Mla pathway were initially identified by Malinverni and Silhavy (2009), who revealed that *E. coli* mutants lacking any of the *mlaFEDCB* genes were hypersensitive to treatment with sodium dodecyl sulphate (SDS)/ethylenediaminetetraacetate (EDTA), a hallmark of OM perturbation.

The Mla system is composed of six members, including the OM lipoprotein MlaA, which is associated with the major porins OmpF and OmpC, the periplasmic chaperone MlaC, and the IM MlaFEDB ATPase complex (Malinverni and Silhavy, 2009) (Figure 1.9).

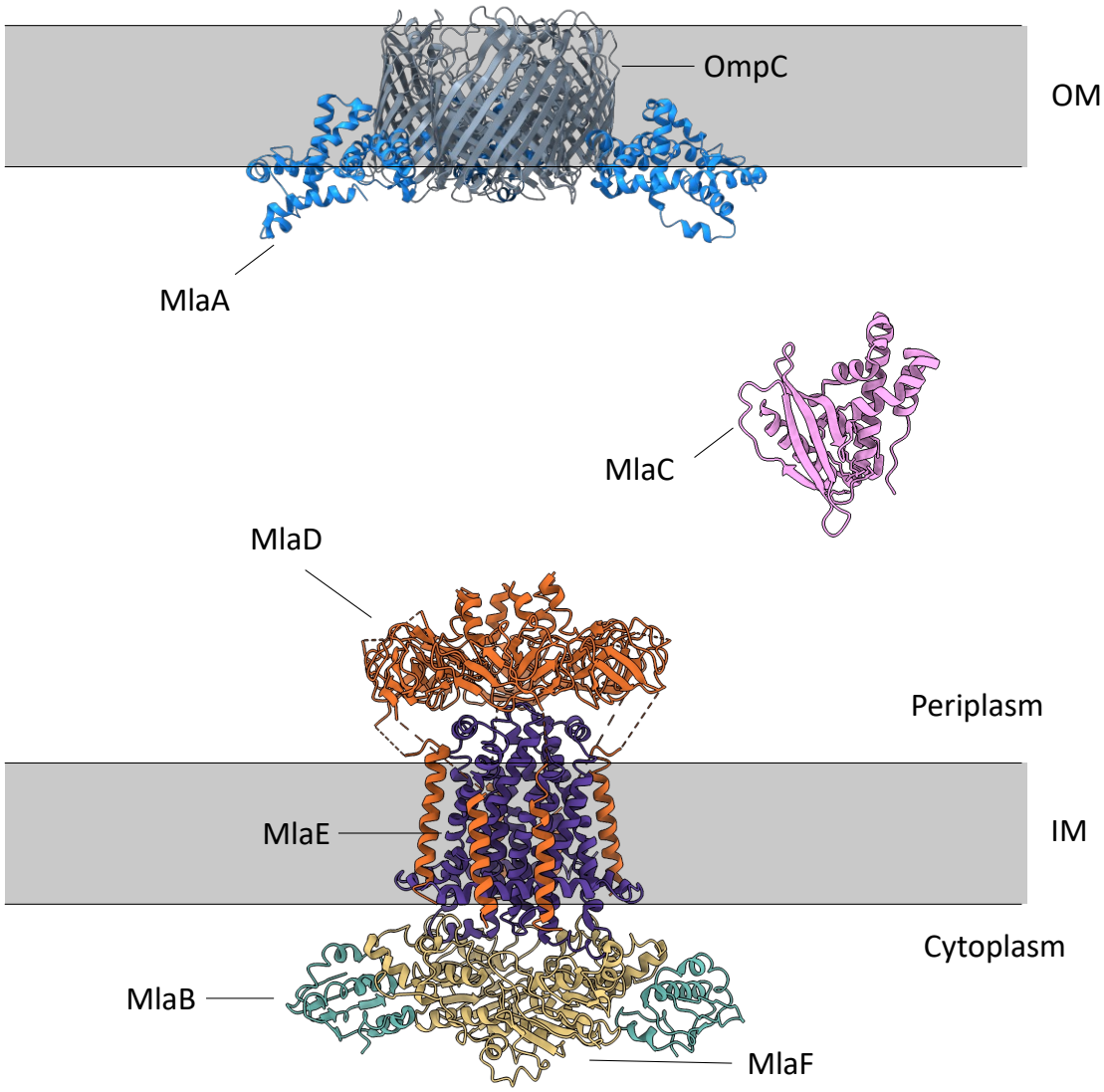


Figure 1.9. The Maintenance of Lipid Asymmetry (Mla) pathway in *E. coli*.

A cartoon representation of the Mla pathway components in their respective cellular locations. MlaA (blue) forms a complex with the β-barrel OMP OmpC (grey) in the OM. The periplasmic chaperone MlaC (pink) transports lipids to the MlaFEDB (MlaF – yellow; MlaE – purple; MlaD – orange; MlaB – mint green) ATPase complex in the IM. (MlaA-OmpF PDB: 5NUP; MlaC PDB: 5UWA; MlaFEDB PDB: 7CGE). The models were visualised with ChimeraX (Pettersen *et al.,* 2021).

1.4.2.1. MlaA

MlaA, the least understood member of the Mla system, is an α-helical, doughnut-shaped monomeric lipoprotein located in the inner leaflet of the OM. It forms an amphipathic pore that is anticipated to provide the path for PL translocation (Abellón-Ruiz *et al.*, 2017). The *mlaA* knockout strain is sensitive to SDS/EDTA in the same manner as the other *mla* gene knockouts. Strikingly, a gain-of-function mutant *MlaA** (deletion of residues N41 and F42) results in considerably increased LPS levels, drug sensitivity, and even cell death during stationary phase (Sutterlin *et al.,* 2016).

MlaA was also identified as a virulence factor in Nontypeable Haemophilus Influenzae (NTHi). Inactivation of the *mlaA* gene resulted in increased fatty acid and phospholipid content and decreased NTHi resistance to a wide range of hydrophobic molecules, consequently leading to decreased epithelial infection (Fernández-Calvet *et al.,* 2018).

MlaA co-purifies with the osmoporins OmpF and OmpC (Chong *et al.,* 2015), however, the removal of OmpC, but not OmpF, results in the accumulation of PLs in the outer leaflet of the OM (Chong *et al.,* 2015). This was further supported by the mutation of OmpC residues believed to interact with MlaA, which resulted in PL accumulation in the OM (Yeow *et al.,* 2018). Indeed, the crystal structure of MlaA was solved in complex with the OmpC trimer (Abellón-Ruiz *et al.,* 2017), further reinforcing that OmpC might have an important role in PL transport via its association with MlaA. Nonetheless, the precise role of OmpC in phospholipid trafficking is currently unknown. Chong *et al.,* (2015) revealed that the *mlaA* mutant lacking its lipobox moiety was still able to localise within the OM, presumably via its association with OmpC. This suggested that OmpC might be important for MlaA localisation and its structural stability.

1.4.2.2. MlaC

MlaC is one of the best studied members of the Mla pathway, likely due to its solubility, stability, and low molecular weight. It functions as an intermembrane transport protein to shuttle PLs between the IM and the OM. MlaC adopts mixed α/β-fold consisting of a twisted four stranded β-sheet and seven α-helices (Hughes *et al.,* 2019). MlaC co-purifies with both PE and PG (Thong *et al.,* 2016), with some groups also observing CL binding (Hughes *et al.,* 2019). The crystal structure of MlaC indeed revealed the presence of a hydrophobic pocket, which has been shown to accommodate acyl chains of PLs (Hughes *et al.,* 2019).

Molecular dynamic simulations conducted by Huang *et al.,* (2016) suggested that the PL binding pocket of MlaC from *R. solanacearum* and *Acinetobacter* would decrease following PL release. This was experimentally confirmed by Hughes *et al.,* (2019), who revealed that the lipid free conformation (MlaC-apo) displayed a closed hydrophobic cavity, suggesting that this unique flexibility allows it to effectively offload PLs upon interacting with either the MlaA-OmpC or MlaFEDB complex.

Indeed, MlaC has been shown to interact with both the MlaA-OmpC complex and the IM complex MlaFEDB, both *in vivo* (Ercan *et al.,* 2019) and *in vitro* (Low *et al.,* 2021; Tang *et al.,* 2021). Nonetheless, structural insight into the MlaC/MlaA interaction was only recently suggested through Alphafold co-folded structures and deep-mutational screening (MacRae *et al.,* 2022). MacRae *et al.,* (2022) proposed that the MlaA c-terminal tail likely plays a significant role in the MlaA/C interaction, however, this observation needs to be furthered supported with structural data of the MlaC/MlaA complex.

1.4.2.3. MlaFEDB

MlaFEDB forms a stable ATPase IM complex in a 2:2:6:2 stoichiometry, as shown by biochemical (Thong *et al.,* 2016) and structural (Ekiert *et al.,* 2017) studies. The MlaFEDB complex is composed of the periplasm-facing MCE protein MlaD, the IM permease MlaE, the cytoplasmic-facing ATP-binding protein MlaF, and the STAS-domain containing protein MlaB.

The MlaFE complex exhibits a typical ABC ATPase fold. MlaF interacts with MlaB, a small ~10 kDa protein, that has been shown to be essential for complex assembly and activity (Ekiert *et al.,* 2017; Kolich *et al.,* 2020; Thong *et al.,* 2016). Each of the MlaF subunits contains a c-terminal extension (CTE). The CTE wraps around the neighbouring MlaF subunit forming a “handshake” motif, which also interacts with MlaB (Kolich *et al.,* 2020). The MlaF mutant with the CTE deletion (MlaFΔCTE) failed to restore the MlaF knockout growth on SDS/EDTA and also failed to interact with both MlaD and MlaE, strongly indicating the essentiality of the CTE motif in the Mla system functionality (Kolich *et al.,* 2020).

MlaFEDB structures determined in the absence of nucleotide consistently revealed an outward-facing V-shaped pocket in the MlaE dimer interface (Chi *et al.,* 2020; Tang *et al.,* 2021; Zhou *et al.,* 2021). Furthermore, multiple MlaFEDB structures demonstrated a distinct density corresponding to PL within the MlaE hydrophobic pocket. (Coudray *et al.,* 2020; Zhou *et al.,* 2021).This suggests that the pocket might provide a hydrophobic passage for PLs.

1.4.2.4. MlaD

The most recognised MCE domain-containing protein in Proteobacteria is MlaD, which contains a single MCE domain and has a high affinity for PLs. Endo *et al.,* (2007) revealed that MlaD is required for the uptake of gamma-hexachlorocyclohexane (gamma-HCH) as the sole source of carbon and energy. Additionally, in *Pseudomonas putida*, MlaD is hypothesised to function by altering PL composition of the OM, making it essential for resistance to toluene (Kim *et al.,* 1998).

Ekiert *et al.,* (2017) first reported the crystal structure of the *E. coli* MCE protein MlaD. Six copies of MlaD assemble into a homo-hexameric ring that associates with the ABC transporter complex in the IM. MlaD is bound to the IM by a single transmembrane n-terminal helix, with its MCE domain residing in the periplasm. The MlaD MCE core domain adopts a seven-stranded β-barrel fold, encompassing just over 100 residues. The n-terminus of each MCE domain is connected by a short linker to the TM helix, which places the n-terminal face of the ring in close proximity to the negatively charged PL head groups. The c-terminal helix from each subunit associates to form a hollow six-helix coiled coil, leaving a channel running through the centre of the ring.

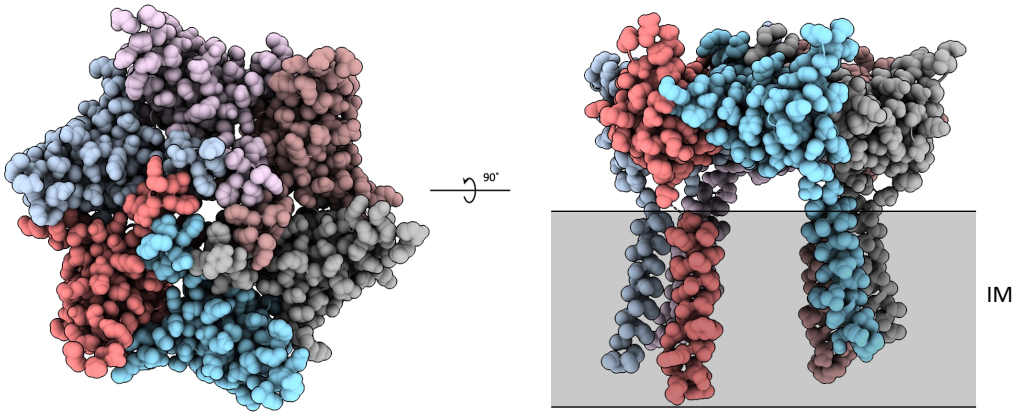


Figure 1.10. The structure of *E. coli* MlaD (PDB:7CH0) in its respective cellular location within the inner membrane (IM.)

Atoms displayed in a sphere style. Chains A, B, C, D, E, and F are coloured in light blue, coral, light steel blue, thistle and rosy brown, respectively. The model was visualised with ChimeraX (Pettersen *et al.,* 2021).

The central pore loop contains hydrophobic residues (Leu153 and Leu154) in the centre of the C6 symmetric complex. At the narrowest point, the channel diameter is ~ 6-8 Å, large enough to allow passage of PLs (Ekiert et al, 2017). Although there is no direct evidence indicating that PLs are trafficked through the central pore, cryo-EM maps of the IM MlaFEDB complex show clear density for six lipid binding sites between the central pore loops in all the nucleotide states (apo- , AppNHP bound and ADP bound) (Mann *et al.,* 2020). The 6x MCE periplasmic “crown” can be rotated against the MlaFEB transmembrane complex; however, the role of the motion in lipid transport remains unclear (Mann *et al.,* 2020).

The Interaction between MlaC and MlaD was only recently characterised through Alphafold, deep mutation screening and a low resolution cryo-EM structure of the MlaFEDB-MlaC complex (MacRae *et al.,* 2022).

The cryo-EM model of the MlaFEDB-MlaC complex revealed that several copies of MlaC are able to bind via the MlaC cleft and β6-β7 loop of MlaD. Further studies should focus on obtaining high-resolution structures of the MlaA-MlaC and MlaFEDB-MlaC complexes, as well as establishing whether ATP binding to MlaFEDB causes any conformational changes to the MlaC/MlaD binding interface.

1.4.2.5. The direction of lipid movement between IM and OM by the Mla pathway.

Significant research effort has been devoted to comprehensively elucidating the intricate mechanism of lipid transport through the Mla pathway and establishing its directionality. Currently, the majority of evidence suggest that PLs are transferred in a retrograde (OM-to-IM) fashion.

It has been shown that upon removal of any of the Mla pathway components, there is an accumulation of PLs in the outer leaflet of the OM (Malinverni and Silhavy, 2009). Furthermore, the structure and molecular dynamics of MlaA suggest that PLs migrate from the outer leaflet of the OM into the central MlaA pore towards the periplasm *(*Abellón-Ruiz *et al.,* 2017). Additionally, the MlaA knockout displays significantly increased OM vesicle formation (OMV) in both *V. cholerae* (Roier *et al.,* 2016) and *N. gonorrhoeae* (Baarda *et al.,* 2019), suggesting that excessive production of OMV is a result of PL accumulation in the outer leaflet of the OM. This implies that the WT strain utilises the Mla pathway to remove mislocalised PLs, further supporting the retrograde transport.

Nevertheless, several contradicting studies have emerged supporting the anterograde (IM-to-OM) transport. Kamischke *et al.,* (2019) revealed that the Mla system in *A. baumannii* is capable of transferring PLs from the IM to the OM *in vivo*, however, this study was heavily criticised by Powers *et al.,* (2020), who emphasised issues with the utilised mutant strains and unreproducible methodology. In addition, *in vitro* studies conducted by Hughes *et al.,* (2019) were the first to observe that MlaD is capable to rapidly transfer PLs to MlaC, highlighting the possibility that the OmpC-Mla system may function in an anterograde (IM-to-OM) fashion. Nonetheless, Low *et al.,* (2021) revealed that although MlaD can transfer PLs to MlaC, such retrograde transfer is eradicated in the presence of ATP. Interestingly, Kolich *et al.,* (2020) showed that there are no conformational changes between apo-MlaF and ADP-MlaF, suggesting that ATP hydrolysis alone is not sufficient to induce lipid export/import, but ATP might function as a modulator of PL transport, responsible for unidirectionality.

Whilst the direction of lipid movement between the IM and OM by the Mla complex has been the subject of debate, it is now widely accepted that PL transport via the Mla pathway occurs in a retrograde fashion.

1.4.3. PqiABC and LetAB

E. coli contains two other MCE-domain containing proteins PqiB and LetB (formerly known as YebT), residing in separate operons (Figure 1.11), and containing three and seven MCE domains, respectively. As for MlaD, PqiB and LetB have been shown to co-purify with PE and PG (Ekiert *et al.,* 2017) and have been strongly implicated in intermembrane PL trafficking.

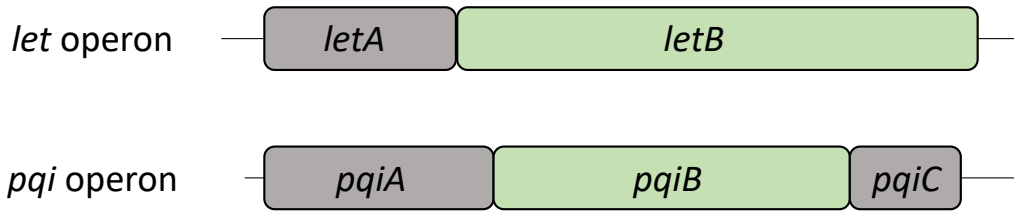


Figure 1.11. *E. coli let* and *pqi* operon architectures.

MCE-domain containing genes are coloured in green. The length of each bar corresponds to the size of each gene.

1.4.3.1. The LetAB complex

LetB is composed of seven stacked MCE rings, forming a tube-like barrel that is long enough to span the periplasm (~230 Å) (Isom *et al.,* 2020) (Figure 1.12, A). The number of MCE rings is crucially important for LetB function, as *letB* mutants with fewer than five rings are unable to complement as established Lauryl-Sulfobetaine (LSB) phenotype (Isom *et al.,* 2020). This is likely due to their insufficient length to provide a direct link between the IM and OM.

A more recent study from Ekiert's lab (Vieni *et al.,* 2021) focused on investigating the relative importance of each of the MCE rings. Interestingly, the deletion of MCE-ring 3 and ring 6 showed even higher fitness than the *E. coli* WT strain on the LSB phenotype. Indeed, structural studies revealed that the ring 6 undergoes the largest conformational changes (Isom *et al.,* 2020). These findings strongly imply that ring 3 and ring 6 might be involved in the regulation of lipid trafficking, hence their deletion eradicates the transport restriction.

Each MCE domain of PqiB, LetB and MlaD possesses a pore lining loop (PLL), located at the center of homo-hexameric assembly, which provides a hydrophobic coating for the central channel through which lipid transport is anticipated. Indeed, the deletion of pore lining loops, as well as single point mutations of the highly hydrophobic and conserved PLL motif ΦxxΦΦ (Φ denotes a hydrophobic residue; x denotes any residue), abolishes LetB function (Ekiert *et al.,* 2017).

LetA (formerly known as YebS) is predicted to be an IM protein and form a stable interaction with LetB, potentially contributing to its role in lipid trafficking. Nevertheless, LetA does not exhibit any sequence homology with known transporters and lacks an ATP-binding domain. According to Alphafold, LetA is composed of eight transmembrane helices divided into two

PqiA domains (Pfam PF04403). LetA is also predicted to possess a cysteine-rich bundle that may extend into the periplasm (Figure 1.12, B). Nonetheless, the structural information and function of LetA, as well as its interaction with LetB, remain highly enigmatic.

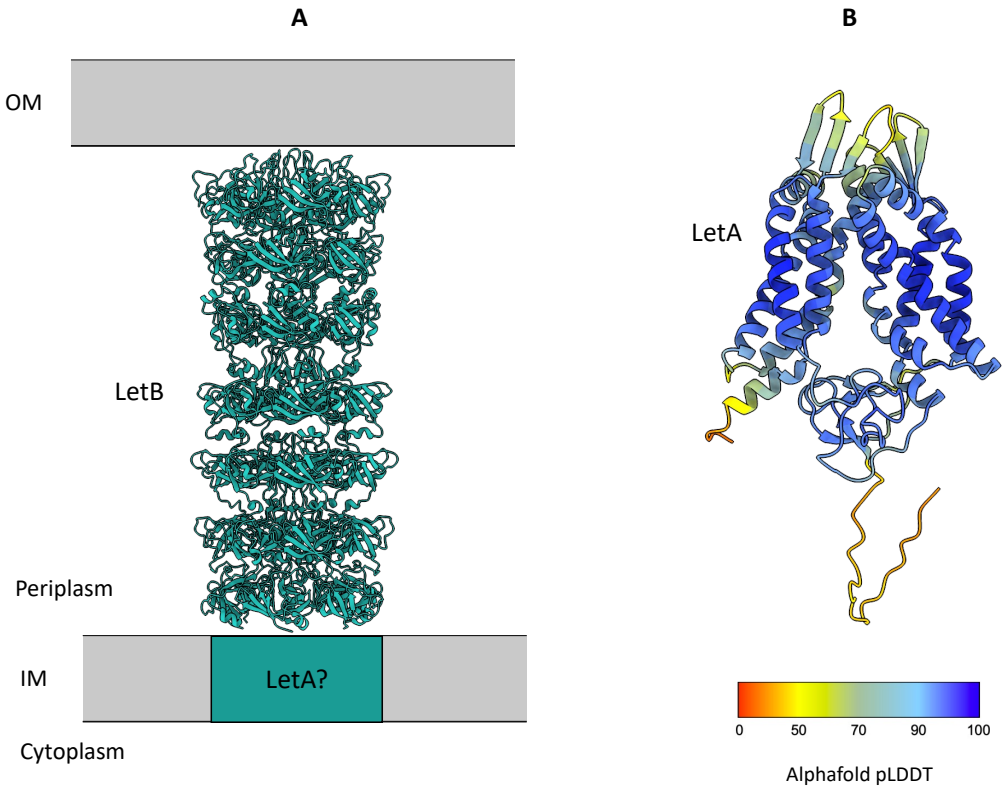


Figure 1.12.The structures of *E. coli* LetB and LetA.

(A) The cryo-EM model of *E. coli* LetB (PDB: 6V0C) in a cartoon representation, shown in its predicted cellular location spanning the periplasm. LetA is represented as a green box in its predicted cellular location within the inner membrane (IM). (B) The Alphafold structure of *E. coli* LetA (AF-P0AD03-F1) in a cartoon representation, coloured by per-residue pLDDT scores. A pLDDT colour key is provided at the bottom of the figure where the spectrum from blue to red corresponds to decreasing pLDDT score (Jumper *et al.,* 2021). The models were visualised with ChimeraX (Pettersen *et al.,* 2021).

1.4.3.2. The PqiABC complex

PqiB forms a homo-hexameric assembly predicted to span the periplasm (~230 Å). Each of the PqiB monomers contains three MCE domains and a c-terminal helix that, when fully assembled, forms a hollow 6-helical coiled coil and a barrel consisting of three homo-hexameric MCE rings (Ekiert *et al.,* 2017) (Figure 1.13). Cryo-EM data processing of the helical coiled coil proved to be challenging due to its high flexibility, most likely arising from the removal of PqiB from its native, membrane-spanning environment (Cooper, 2023).

Cooper (2023) revealed that PqiB constructs with single point mutations in the PLLs failed to restore growth on the 1% LSB phenotype, suggesting that both PqiB and LetB might have similar mechanism of action.

The remaining PqiB operonic components, PqiA and PqiC are predicted to reside within the IM and OM, respectively (Figure 1.13). PqiA shares close structural homology to LetA, as revealed by Alphafold, and results in our lab suggest that it might function in a PMF - dependant manner (unpublished).

In addition, Cooper (2023) uncovered the structure of PqiC, revealing that it adopts an octameric ring assembly. Nevertheless, the structural information provides little insight into the assembly of the PqiABC complex. Whilst a low resolution cryo-EM map of the PqiB-PqiC complex suggested that PqiC might bind directly onto the PqiB needle (Cooper 2023), the cryo-EM model is too poor to allow definitive conclusion and requires further structural investigation.

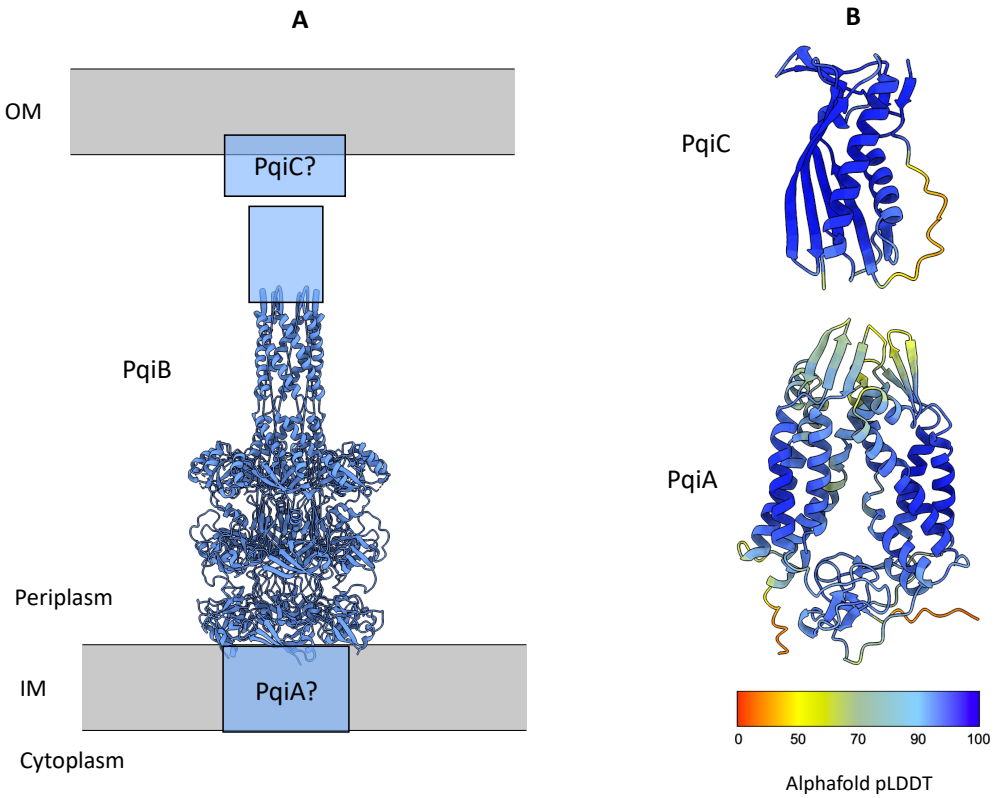


Figure 1.13. The structures of *E. coli* PqiB, PqiA and PqiC.

A) The cryoEM model of *E. coli* PqiB (PDB: 5UVN) in a cartoon representation, shown in its predicted cellular location spanning the periplasm. The blue box suggests the anticipated extent of the PqiB needle. PqiA and PqiC are represented as a blue box in their predicted cellular locations within the IM and OM, respectively. (B) The Alphafold structures of *E. coli* PqiA (AF-P0AFL9-F1) and PqiC (AF-P0AB10-F1) in a cartoon representation, coloured by per-residue pLDDT scores. A pLDDT colour key is provided at the bottom right of the figure, where the spectrum from blue to red corresponds to decreasing pLDDT score (Jumper *et al.,* 2021). The models were visualised with ChimeraX (Pettersen *et al.,* 2021).

1.4.4. The Asma-like protein family

Intermembrane phospholipid transport in gram-negative bacteria may not be limited to the MCE family of proteins and their associated partners. Recently, a novel *E.coli* Asma-like protein family composed of six members was identified: TamB, AsmA, YdbH, YhdP, YhjG, and YicH (Figure 1.14).

Recent bioinformatic analysis identified the common Chorein-N domain in TamB and AsmA, which is also found in eukaryotic PL transporters (Levine, 2019). The Alphaphold structural prediction and the crystal structure of the TamB domain portion (Josts *et al.,* 2017) suggest that AsmA-like proteins adopt a "β-taco" fold with a hydrophobic interior that potentially allows the passage of hydrophobic substrates, such as lipids or membrane proteins. TamB, YhdP, YhdbH are predicted to be of a sufficient length to span the entire periplasm, providing a direct link between the IM and OM (Yeow and Chng, 2022).

A number of knockout strains of AsmA-like proteins were generated to uncover their potential link to PL transport. It was reported that deletion of *asmA* resulted in a reduction of LPS levels (Deng and Misra, 1996) and deletion of *yhdP* displayed vancomycin and detergent sensitivity, a hallmark of OM perturbation (Mitchell *et al.,* 2018). More recently, it has been shown that the triple *yhdP*, *tamB* and *ydbH* deletion is lethal to the cell (Ruiz *et al.,* 2021). In addition, the double deletion of *yhdP* and *tamB* lead to cell lysis and the shedding of excess LPS in the OM vesicles, which is proposed to maintain OM homeostasis when anterograde PL transport is disrupted (Douglass *et al.,* 2022). Collectively, these studies suggest that the AsmA-like protein family might be essential for OM homeostasis by providing additional systems for intermembrane cargo trafficking. Nonetheless, the structural and functional characterisation remains scarce, necessitating further investigation.

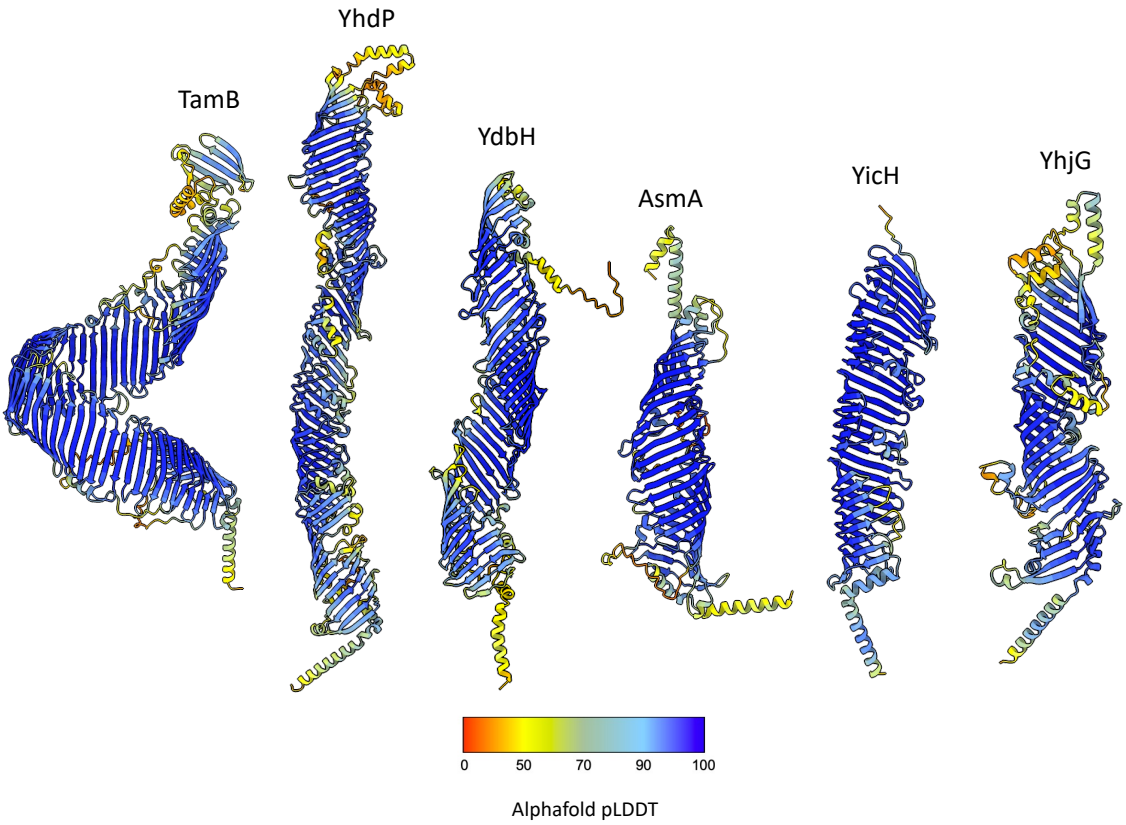


Figure 1.14. The AsmA-like protein family.

Alphafold structures of the six *E. coli* Asma-Like protein family members in a cartoon representation. TamB (AF-P39321-F1), YhdP (AF-P46474-F1), YdbH (AF-P52545-F1), AsmA (AF-P28249-F1), YicH (AF-P31433-F1), YhjG (AF-P37645-F1). The structures are coloured by per-residue pLDDT scores. A pLDDT colour key is provided at the bottom of the figure where the spectrum from blue to red corresponds to decreasing pLDDT score (Jumper *et al.,* 2021). The models were visualised with ChimeraX (Pettersen *et al.,* 2021).

1.4.5. The Tol-Pal complex

The Tol-Pal complex is a highly conserved transmembrane system composed of the IM TolQ, TolR and TolA proteins, the periplasmic protein TolB, and the OM lipoprotein Pal. The Tol-Pal pathway has been observed to have various functions, including cell division (Gerding *et al.,* 2007), polar localisation of chemoreceptors in the IM (Santos *et al.,* 2014), and surface expression O-specific LPS (Gaspar *et al.,* 2000). The Tol-Pal complex knockout strain displayed a “leaky” membrane phenotype, leading to hypersensitivity to drugs and detergents (Llamas *et al.,* 2000; Vinés *et al.,* 2005). More recently, PL accumulation has been observed in Tol-Pal mutant cells in *E. coli* (Shrivastava *et al.,* 2017) and *Salmonella typhimurium* (Masilamani *et al.,* 2018), directly linking Tol-Pal function to PL transport. Nevertheless, Tol-Pal function remains enigmatic and requires further investigation.

1.5. *Bdellovibrio bacteriovorus*: a potential predator for combating antibiotic-resistant infections.

The proliferation of antibiotic resistance has emerged as a critical challenge in modern medicine. Gram-negative bacteria pose a significant threat due to their intricate cell envelope compositions that hinder the penetration of numerous antimicrobial agents. Moreover, the limited availability of antibiotics coupled with the slow and expensive process of developing new ones has exacerbated the issue. Consequently, there exists an urgent requirement to identify alternative strategies for combatting antibiotic-resistant bacteria.

Bdellovibrio bacteriovorus represents a promising target in the fight against antibiotic-resistant bacteria. It is a small, highly motile gram-negative δ-proteobacterium that oscillates between free-living growth and predation on other gram-negative bacteria, including important pathogenic strains such as *E. coli* and *S. enterica* (Atterbury *et al.,* 2011). Bdellovibrio-and-like organisms (BALOs) are widely abundant in nature and have been isolated from several different environments, including soil (Jurkevitch *et al.,* 2000), aqueous environment (Varon and Shilo, 1981) and the gut of healthy human individuals (Iebba *et al.*, 2013).

The unique lifestyle of *Bdellovibrio bacteriovorus*, its predation tactics and the significant number of gram-negative bacteria species that are susceptible to this predator offer a wide range of potential applications in medicine, industry and agriculture. Indeed, *Bdellovibrio bacteriovorus* has received considerable research interest in its utilisation to combat antibiotic-resistant hospital infections. The first study investigating the potential of *Bdellovibrio bacteriovorus* as an antimicrobial agent administered it orally to chickens whose guts were colonised by *Salmonella*. The study revealed that *Bdellovibrio* significantly reduced the number of *Salmonella* with no effect on the health of the infected birds (Atterbury *et al.,*

2011). More recently, it has been shown that injecting *B. bacteriovorus* into zebrafish containing a lethal dose of an antibiotic resistant strain of the human pathogen *Shigella flexneri* promoted pathogen eradication and increased zebrafish survival (Willis *et al.,* 2016). Furthermore, Bonfiglio *et al.,* (2020) revealed that *B. bacteriovorus* can interfere with crucial aspects of adherent invasive *E. coli* strain (AIES) pathogenicity and could control its overgrowth in Crohn's disease patients.

To date, there have been no reported cases of a bacterium that has developed resistance to *B. bacteriovorus,* facilitating its utilisation for antimicrobial drug development. Additionally, the LPS of *B. bacteriovorus* has an unusual chemical structure lacking phosphate groups. As a result, it induces 100-fold less TNF-α secretion from human mononuclear cells when compared to *E. coli* K12 (Schwudke *et al.,* 2003), thereby indicating that this predator does not pose any significant cytotoxicity or inflammatory response.

It is important to note that although *B. bacteriovorus* kills a large percentage of prey *in vitro*, it never results in complete eradication of the attacked population (Varon & Shilo, 1981). Nevertheless, Pasternak *et al.,* (2014) reported that *B. bacteriovorus* is resistant to a surprisingly high number of antibiotic compounds, allowing it to be used in combination with antibiotics. Furthermore, although *B. bacteriovorus* invades only gram-negative bacteria, Im *et al.,* (2018) demonstrated its capability to degrade the *Staphylococcus aureus* biofilm matrix and obtain nutrients, suggesting that *B. bacteriovorus* predation is not limited to gram-negative bacteria alone.

1.6. Bdellovibrio bacteriovorus life cycle

Bdellovibrio bacteriovorus has a biphasic cell cycle consisting of two main phases: (1) The attack phase (AP), during which the predator looks for a prey cell, attaches to its outer membrane and enters the periplasm. (2) The growth phase (GP), during which *B. bacteriovorus* degrades host resources and utilises them to grow and elongate as a filament until multiple fission occurs, resulting in several new *B. bacteriovorus* cells (Figure 1.15).

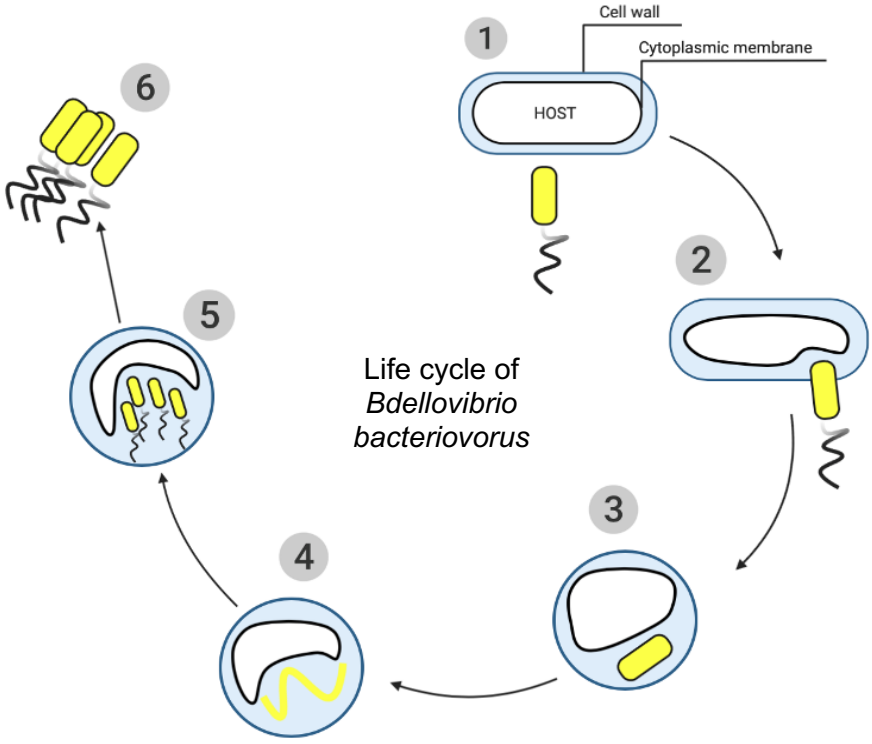


Figure 1.15. *Bdellovibrio bacteriovorus* life cycle.

1 - *B. bacteriovorus* swims at high-speed looking for a prey cell. 2 – It attaches to the prey cell by its non-flagellar end and penetrates the cell wall of the host cell. 3 – the host cell rounds up. Cytoplasmic membrane leaks, allowing components to be digested and absorbed by *B. bacteriovorus*. 4 - *B. bacteriovorus* elongates. 5 – As food supply becomes exhausted, *B. bacteriovorus* filament divides by multiple fission, splitting into several *B. bacteriovorus*, each with its own genome (single circular chromosome). 6 – Host cell swells and new *B. bacteriovorus* are released as the host cell raptures.

Once *B. bacteriovorus* encounters a potential prey cell, it undergoes a brief "recognition period" of reversible attachment with its prey cell. The recognition period is non-specific, as *B. bacteriovorus* is able to attach to non-prey cells, gram-positive bacteria, and even abiotic materials in the medium. Attachment becomes irreversible if the prey cell is suitable (Varon & Shilo, 1968). It was initially believed that *B. bacteriovorus* utilises its flagellum to enter prey cells through a "drilling" mechanism. Nevertheless, Lambert *et al.,* (2006) reported that motility-compromised *B. bacteriovorus* were able to predate. Recently, Kaplan *et al.,* (2022) revealed that during invasion, *B. bacteriovorus* absorbs its flagellum into its own periplasm. Finally, cryo-EM tomography revealed that *B. bacteriovorus* enters the prey cell through a small pore that requires *B. bacteriovorus* to deform and squeeze through. Once *B. bacteriovorus* enters the periplasm, the pore is resealed (Kaplan *et al.,* 2022; Kuru *et al.,* 2017)

1.6.1. Bdelloplast formation

After entering the periplasm, the host cell rounds up forming a bdelloplast. Targeted mutagenesis studies have revealed that the *bd3459* and *bd0816* encoded products, which exhibit homology to peptidoglycan remodelling enzymes, are responsible for rounding up the prey cell. This process reduces self-competitive double prey-entry events (Lerner *et al.,* 2012).

1.6.2. The growth phase and release of progeny.

As early as 1969, Shilo observed that *B. bacteriovorus* has the ability to form elongated filaments (Shilo, M., 1969). When the host resources are depleted, these filaments separate along their length, giving rise to progeny cells. Upon multiple fission of the filament, *B. bacteriovorus* produces either an odd or even number of progeny cells, in sharp contrast to the binary division of most bacteria (Laloux, 2018). Using time-lapse fluorescence microscopy, Fenton *et al.,* (2010) revealed that after the separation and maturation of *B. bacteriovorus* progeny, which involves the development of flagella, the cells exit the exhausted prey though a pore in the membrane rather than the breakdown of the entire OM of the prey. In addition, Harding *et al.,* (2020) revealed that progeny cells secrete peptidoglycan hydrolytic enzymes, resulting in the targeting of the prey cell envelope.

1.7. Lipid transport in *Bdellovibrio bacteriovorus* HD100.

Despite the well documented life cycle of *B. bacteriovorus,* from the 3587 recognised protein-encoding genes*,* over 40% are uncharacterised or hypothetical proteins (Barabote *et al.,* 2007), likely due to its unique lifestyle and distant evolutionary relationship to model bacteria. Barabote *et al.,* (2007) revealed that the genome of *B. bacteriovorus* HD100 encodes the largest percentage of transport systems when compared to any other free-living bacteria.

Bioinformatics analysis by Tajabadi *et al.,* (2017) exposed that 13% of all protein-coding genes have the potential of being ABC type and MSF type transmembrane transporters, with 3.3% and 1.2% of these being candidates for lipid and lipoprotein transporters, respectively (Tajabadi *et al.,* 2017). In addition, Tajabadi and colleagues (2017) also demonstrated that *B. bacteriovorus* possesses three times as many lipid transporters as *B. exovorus,* another parasitic bacteria that grows outside the prey cell but utilises prey cell contents as nutrients. Furthermore, there is evidence showing that *B. bacteriovorus* re-utilises OMPs, lipid A, and fatty acids of the prey bacteria by integrating them into its own membrane system. Additionally, Kuru *et al.,* (2022) disclosed that *B. bacteriovorus* squeezes into the prey's periplasm through a self-degraded pore and reseals it after entry. These studies collectively suggest that cell wall remodelling of both *B. bacteriovorus,* and the prey cell is of paramount importance for the unique life cycle of this predator. However, the involvement of lipid transport systems in the predation and growth mechanisms *of B. bacteriovorus* remains largely enigmatic.

1.8. Biological significance and aim of the project.

A number of publications suggest that MCE proteins are involved in lipid trafficking in gram-negative bacteria. However, research has mostly focused on elucidating intermembrane lipid transport in the model organism *E. coli*, with limited additional studies conducted on *Pseudomonas aeruginosa* (Munguia *et al.,* 2017) and *Acinetobacter baumannii* (Kamischke *et al.,* 2019). While the Mla system has been intensively studied with a number of publications confirming its function as a lipid transport pathway, LetB and PqiB protein involvement in intermembrane lipid trafficking remains only speculative.

MCE proteins are widely distributed across bacterial species and have been found to be located with a variety of neighbouring genes, possessing miscellaneous domain architectures. The diversity of MCE proteins represents a largely unexploited research avenue that could uncover a fundamental knowledge gap in the mechanism behind intermembrane PL transport.

We have decided to expand the bacterial organisms within which the role of MCE proteins is being studied by focusing on *Bdellovibrio bacteriovorus* HD100. Bioinformatic analysis has suggested that membrane transport proteins might play a significant role in the lifecycle of this predatory species (Tajabadi *et al.,* 2017), nevertheless, to date, only the TAT system has been experimentally studied (Chang *et al.,* 2011).

The biological significance of the current project is twofold: analysing putative lipid transport proteins could provide considerable insight into the unique lifecycle of *Bdellovibrio bacteriovorus*, potentially accelerating its therapeutic exploitation. Additionally, diversifying the organisms in which MCE proteins are being studied could provide novel insight into the fundamental processes of OM biogenesis.

Chapter 2 - Materials and Methods

2.1. Culture media and growth conditions

Bacterial cultures were grown in Miller's lysogenic broth (hereafter referred to as LB) (consisting of 10 g/L tryptone, 10 g/L NaCl, 5 g/L yeast extract). Unless otherwise stated, LB was supplemented with the selection antibiotic kanamycin (30 µg/mL). Overnight cultures were prepared in a 20 mL universal tube and incubated at 37˚C in a shaking incubator at 180 rpm. To prepare LB agar, 15 g/L of nutrient agar was added into the LB. Unless otherwise stated, LB agar plates were grown at 37 ˚C overnight.

2.2. Construct design

Table 2.1. *E. coli* strains used in this study.

Strain	Usage	Source
E. coli C43 (DE3)	Protein expression	Miroux and Walker, 1996
E. coli DH5α	Plasmid replication	New England Biolabs
E. coli NEB5α	Plasmid replication	New England Biolabs

Table 2.2. DNA constructs used in this study.

Construct	Plasmid	Restriction sites	His-tag location	Source
Bd1153-Bd1152	pET26b(+)	XhoI-NdeI	c-terminus	GenScript
Bd1153-Bd1151	pET26b(+)	XhoI-NdeI	c-terminus	This study
Bd3410-Bd3413	pET26b(+)	XhoI-NdeI	c-terminus	GenScript
$Bd1770^{34-317}$	pET26b(+)	XhoI-NdeI	c-terminus	GenScript
Bd1153-Bd1152	pET26b(+)	XhoI-NdeI	n-terminus	Twist Bioscience

Table 2.3. Primers used in this study.

Primer	Sequence (5' to 3')	Description
Bd1152s_F	CAGTATCTGACGTCGCTG	Forward primer for the formation of the Bd1152$^{37-283}$ construct from the Bd1153-Bd1152 construct.
Bd1152s_R	CATATGTATATCTCCTTCTTAAAGTTAAAC	Reverse primer for the formation of the Bd1152$^{37-283}$ construct from the Bd1153-Bd1152 construct.
Bd1153-1151_F_Nde1	GGGGGGCATATGAATACACAAGCTCAGCCTCTGACC	Forward primer for amplification of the entire Bd1153-Bd1151 operon from the *B. bacteriovorus* HD100 genome.
Bd1153-1151_R_Xho1	GGGGGGCTCGAGCTTTGTAAAATAGGTGACCAGAAT	Reverse primer for amplification of the entire Bd1153-Bd1151 operon from the *B. bacteriovorus* HD100 genome.
Bd1152s_I112C_F	CAAAGTCAGCTGTCAGGGCATCACC	Forward primer for introducing the I112C point mutation into the Bd1152$^{37-283}$ construct.
Bd1152s_I112C_R	GGGGTGACGATCCAGAAC	Reverse primer for introducing the I112C point mutation into the Bd1152$^{37-283}$ construct.
Bd1152s_L238C_F	AATCAATTCATGTGAAACTCTGCTTCACGG	Forward primer for introducing the L238C point mutation into the Bd1152$^{37-283\ I112C}$ construct.
Bd1152s_L238C_R	TTTAGTTCTGATTTGAACAGAC	Reverse primer for introducing the L238C point mutation into the Bd1152$^{37-283\ I112C}$ construct.
Bd1152$^{37-283\ I112C\ L238C}$F229A_F	ACTGGGTCTGGCTAAATCAGAACTAAAAATCAATTCATGTGAAACTCTGCTTCAC	Forward primer for introducing the F229A point mutation into the Bd1152$^{37-283\ I112C\ L238C}$ construct.
Bd1152$^{37-283\ I112C\ L238C}$F229A_R	TTGGCCTGCACGCCGACC	Reverse primer for introducing the F229A point mutation into the Bd1152$^{37-283\ I112C\ L238C}$ construct.
Bd1152$^{37-283\ I112C\ L238C\ F229A}$L228A_F	CAAACTGGGTGCTTTCAAATCAGAACTAAAAATCAATTCATGTGAAACTCTGCTTCACG	Forward primer for introducing the L228A point mutation into the Bd1152$^{37-283\ I112C\ L238C\ F229A}$ construct.
Bd1152$^{37-283\ I112C\ L238C\ F229A}$_L228A_R	GCCTGCACGCCGACCTTG	Reverse primer for introducing the L228A point mutation into the Bd1152$^{37-283\ I112C\ L238C\ F229A}$ construct.

2.2.1. Q5-site directed mutagenesis.

All deletion and substitution mutations were carried out using a Q5 site-directed mutagenesis kit (NEB) following the manufacturers protocol (E0554). Primers were designed using the NEB primer design tool NEBaseChanger (https://nebasechanger.neb.com) and are listed in Table 2.3. The reaction constituents and thermocycler conditions are summarised in Table 2.4 and Table 2.5, respectively. Annealing temperatures for primers were calculated using the NEBaseChanger tool (https://nebasechanger.neb.com)

Table 2.4. Q5-site directed mutagenesis reaction constituents (NEB)

Reagent	Amount (µL)
10x Hot Start High-Fidelity 2x Master Mix	12.5
10µM Forward primer	1.25
10µM Reverse primer	1.25
Template DNA (1-25ng/µL)	1
Nuclease free water	9

Table 2.5. Thermocycler conditions for PCR

Step	Temperature	Time
Initial denaturation	98°C	30sec
25 Cycles	98°C	10sec
	50-72°C	30sec
	72°C	30sec/kb
Final Extension	72°C	2min
Hold	4°C	

The resulting PCR reaction was treated with KLD reaction buffer and KLD enzyme mix (containing Kinase, Ligase and Dpn1 enzymes) (Please see Table 2.6) and incubated for 20 min at room temperature. NEB5α competent *E. coli* cells (NEB) were transformed with 5 µL of the KLD mix and plated on LB agar supplemented with an appropriate selection antibiotic.

Table 2.6. Kinase, ligase and Dpn1 (KLD) treatment.

Reagent	Amount (µL)
PCR product	1
2 x KLD reaction buffer	5
10 x KLD enzyme mix	1
Nuclease free water	3

2.2.2. Cloning

Primers used for cloning are listed in Table 2.3. DNA fragments required for cloning were amplified using Phusion High-Fidelity DNA polymerase (NEB) supplemented with GC buffer (NEB). The reaction constituents are detailed in Table 2.7. Typically, two 50 µL reactions were run on the thermocycler, utilising the cycling conditions listed in Table 2.5 for 30-35 cycles.

Table 2.7. PCR reaction constituents for PCR with Phusion High-Fidelity DNA polymerase.

Reagent	Amount (µL)
5 x Phusion HF or GC Buffer	20
10mM dNTPs	2
10µM Forward Primer	5
10µM Reverse Primer	5
Template DNA (<250ng)	1
Phusion enzyme	0.5
Nuclease free water	up to 50

The resulting DNA fragments were separated with agarose gel electrophoresis (more detail in section 2.4.3) and extracted using Qiagen QIAquick Gel Extraction kit as per the manufacturer's guidelines. Purified DNA (~1 µg) and pET26b(+) vector (~1 µg) were digested with Nde1 and Xho1 restriction enzymes at 37˚C for 1 h. Reaction constituents are listed in Table 2.8. Immediately following digestion, the plasmid reaction was treated with Quick Cip (NEB) at 37˚C for 1 h to prevent re-ligation. Digested DNA fragments and vector were cleaned up using QIAquick PCR & Gel Cleanup Kit as per manufacturers guidelines and quantified using a DeNovix DS-11 FX spectrophotometer (DeNovix) utilising the 'dsDNA' program. The insert and backbone were ligated at a 5:1 insert:plasmid ratio using T4 ligase (NEB) in an overnight reaction at 16 ˚C. Reaction constituents are listed in Table 2.9. Following incubation, 5 µL of the ligation reaction was transformed into *E.coli* DH5a cells (NEB) and plated on the LB agar supplemented with 30 µg/mL Kanamycin.

Table 2.8. Double restriction digest reaction constituents.

Reagent	Amount (µL)
Nde1 Restriction Enzyme	1
Xho1 Restriction Enzyme	1
10 x CutSmart buffer	5
DNA	x µL = 1µg
Nuclease-free water	to 50

Table 2.9. 74 DNA ligation constituents using a molar ration of 1:5 pET26b(+) vector and the Bd1153-Bd1152 insert.

Reagent	Amount (µL)
10x T4 DNA Ligase Buffer	2
Vector DNA	10 (75ng)
Insert DNA	4 (64ng)
T4 DNA Ligase	1
Nuclease-free water	to 20

2.2.3. Agarose gel electrophoresis

Amplified DNA fragments were separated and visualised with 0.8 – 2 (w/v) % agarose gels. A required mass of agarose (Bioline) was dissolved in 1 x TAE buffer (40 mM Tris, 20 mM Acetic acid, 1 mM EDTA) by heating the solution. SYBR Safe DNA stain (Thermo Fisher Scientific) was added into the solution as per the manufacturers guidelines prior to pouring it into a Bio-Rad horizontal electrophoresis system. DNA samples were mixed with 5x loading buffer Blue (Bioline) at a 1:5 buffer-to-sample ratio. Hyperladder (1 Kb) (Bioline) was used as a marker. Gels were run in 1 x TAE buffer for approximately 30 min at 100 V. Gels were viewed under UV light and imaged using a GelDoc (Bio-Rad) system.

2.2.4. DNA preparation and sequencing

Plasmid DNA was isolated from 5 mL overnight cultures using Qiagen QIAprep Spin Miniprep kit as per the manufacturer's guidelines. Plasmid DNA was quantified using a DeNovix DS-11 FX spectrophotometer (DeNovix) utilising the 'dsDNA' program. When required, it was sent for sequencing with Source Bioscience (www.sourcebioscience.com).

For sequencing, 5 µL of plasmid DNA (50 – 100 ng/µL) was submitted for sequencing using Source Bioscience's T7F or T7R sequencing primers, depending on the mutation position.

2.3. Preparation of *E. coli* chemically competent cells

To prepare *E. coli* competent cells, the desired strain was inoculated in 5 mL LB and grown overnight at 37˚C. This starting culture was then used to inoculate 100 mL of LB at a 1:100 dilution and grown at 37˚C until OD_{600} of 0.4 – 0.5. The cells were placed on ice for 10 min and spun down in 50 mL tubes at 3220 x g for 10 min at 4˚C. The pellet was resuspended in 30 mL of 30 mM potassium acetate, 50 mM manganese chloride, 100 mM potassium chloride, 10 mM calcium chloride, 15% v/v glycerol and incubated on ice for 90 min. The cells were centrifuged again at 3220 x g for 10 min at 4ºC and resuspended in 4 mL of 10 mM $CaCl_2$ and 15 % glycerol. Cells were then aliquoted into sterile Eppendorf tubes (50 µL each) and snap frozen in liquid nitrogen. Aliquots were stored at -80˚C until further use.

2.4. Bacterial transformation.

The plasmid DNA was transformed into the desired *E. coli* strain using heat shock. 1-5 µL of plasmid DNA was incubated with ~50 µL of competent cells on ice for 30 min. The *E. coli* cells were heat shocked at 42˚C for 30 sec and immediately placed on ice for a further 5 min. SOC (Super Optimal broth with Catabolite Repression) solution was added into the cells before incubation at 37˚C for 1 hour. The cells were then plated on LB agar supplemented with the required selection antibiotic and grown at 37˚C overnight.

2.5. Expression and purification of recombinant proteins

2.5.1. Growth conditions for both soluble and membrane proteins.

20 mL LB was supplemented with the required selection antibiotic and inoculated with a single colony of transformed *E. coli* C43 (DE3) cells before growth overnight at 37˚C. The following day, 2 L of LB media was supplemented with the required selection antibiotic and inoculated with the 20 mL overnight culture. Cells were grown at 37˚C until OD_{600} = 0.5-0.6 before protein expression was induced with 1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG). Protein expression continued overnight at 18˚C. Cells were then harvested by centrifugation at 4000 x g (JLA 8.1000, Beckman Coulter) for 15 min at 4˚C. Cell pellets were stored at -80˚C until required.

2.5.2. Purification of soluble proteins

The cell pellets (Bd1152$^{37\text{-}283}$, Bd1152$^{37\text{-}283\ I112CL238C}$, Bd1152$^{37\text{-}283\ I112C\ L238C\ L228A\ F229A}$, Bd1770$^{40\text{-}317}$) were thawed and resuspended in a resuspension buffer (20 mM Bis-Tris-Propane (BTP), 500 mM NaCl, pH 8) supplemented with cOmplete™ EDTA-free protease inhibitor tablets (Roche). Cells were lysed via 3 passes of a C3 Emulsifex cell disruptor (Avestin) at 17000 psi at 4˚C. Cell debris was pelleted by centrifugation at 45000 x g (JA 25.50, Beckman Coulter) for 1 h at 4˚C. The supernatant was then filtered with a 0.1-micron syringe filter (MerckMilipore). A 5mL HisTrap™ HP Ni-NTA column (Cytiva) was equilibrated with 5 column volumes of wash buffer (20 mM BTP, 500 mM NaCl, 50 mM Imidazole, pH 8) before the supernatant was recirculated through the column overnight at 4˚C. The column was subsequently washed with 5 column volumes of wash buffer before the protein was eluted in 12 x 1 mL fractions of elution buffer (20 mM BTP, 500 mM NaCl, 500 mM

Imidazole, pH 8). Collected elution fractions were analysed via SDS-PAGE (see 2.5.4) and loaded onto a Superdex S200 10/300 Increase (GE Healthcare) column for further purification in SEC buffer (20 mM BTP, 150 mM NaCl, pH 8). All buffers were filtered with 0.45-micron filters (MerckMilipore) and degassed.

2.5.3. Purification of membrane associated complexes.

The cell pellets (Bd1153-Bd1151, Bd1153-Bd1152, Bd3413-Bd3410) were thawed and resuspended in a resuspension buffer (20 mM BTP, 500 mM NaCl, pH 8.5) supplemented with cOmplete™ EDTA-free protease inhibitor tablets (Roche). Cells were lysed via 3 passes of a C3 Emulsifex cell disruptor (Avestin) at 17000 psi and 4˚C. Lysed cells were pelleted by centrifugation at 10000 x g (JA 25.50, Beckman Coulter) for 10 min at 4˚C. The supernatant was then transferred into ultracentrifuge tubes and membranes harvested at 165000 x g (Type 70 Ti Fixed Angle Rotor, Beckman Coulter) for 1 h at 4˚C. Membranes were resuspended in solubilisation buffer (20 mM BTP, 500 mM NaCl, 1% DDM (n-Dodecyl β-D-maltoside), pH 8.5) supplemented with cOmplete™ EDTA-free protease inhibitor tablets (Roche). Membranes were solubilised for 3 h at 4˚C before centrifugation at 75000 x g for 25 min at 4˚C to pellet any insoluble material. The resulting supernatant was filtered through a 0.45-micron filter (MerckMilipore) before binding overnight to Ni-NTA beads (Qiagen) pre-equilibrated with 5 column volumes of wash buffer (20 mM BTP, 500 mM NaCl, 50 mM Imidazole, 0.03% DDM, pH 8.5). The Ni-NTA beads were subsequently washed with 5 column volumes of the wash buffer before the protein was eluted with elution buffer (20 mM BTP, 500 mM NaCl, 500 mM Imidazole, 0.03% DDM, pH 8.5). Elution fractions were analysed with SDS-PAGE and loaded onto a Superose 6 10/300 Increase (GE Healthcare) column for further purification in SEC buffer (20 mM BTP, 150 mM NaCl, 0.03% DDM, pH 8.5).

2.5.4. SDS-PAGE analysis

Protein samples were mixed with a 2 x LaemmLi sample buffer (Sigma-Aldrich) (4% SDS, 20% glycerol, 10% 2-mercaptoethanol, 0.004% bromophenol blue and 0.125M Tris-HCl pH 6.8) in a 1:1 ratio. The samples were loaded onto 4-12% bis-tris gels (GenScript) and run in MES buffer (GenScript) at 160 V for 45-50 min. Gels were stained with Instant*Blue*^{TM} Ultrafast Protein Stain (Sigma-Aldrich).

2.5.5. Western blot

Protein samples were separated by SDS-PAGE and subsequently transferred onto a nitrocellulose membrane via a Trans-Blot Turbo transfer system (Bio-Rad). The membranes were initially blocked at 4˚C with TBST buffer (20 mM Tris, 150 mM NaCl, 0.1% w/v Tween®20) supplemented with 5 % (w/v) skimmed milk powder. They were then incubated overnight with a 6x His monoclonal antibody (Takara Bio) diluted 1:5000 in TBST supplemented with 5 % (w/v) skimmed milk powder. Membranes were washed five times with TBST buffer prior to incubation with anti-mouse HRP-conjugated secondary (ThermoFisher Scientific) diluted 1:10000 in TBST for 1 h. Membranes were washed five times with TBST buffer and incubated with ECL (Cytiva) before being imaged using an Amersham 600 series imager (GE Healthcare).

2.5.6. Thin layer chromatography

The lipids were extracted from purified soluble proteins using chloroform-methanol extraction. 2 mL of purified protein at 0.5 mg/mL was mixed with 2 mL of methanol and 1 mL of chloroform in a glass tube. The solution was vortexed for 5 min, incubated at 50˚C for 30 min, and vortexed again for 5 min. The sample was then centrifuged at 2250 x g for 10 min at room temperature, allowing the lower chloroform phase to be extracted and transferred into a new glass tube. The chloroform was evaporated, and the dried lipids were resuspended in 100 µL of fresh chloroform.

5 µL of the sample and lipid standard (*E. coli* polar phospholipids (Avanti)) were loaded onto Silica TLC plates (Sigma-Aldrich). The TLC plate was then run in a 6.5:2.5:1 (chloroform : methanol : acetic acid) solvent mixture until 1 cm was left from the top of the plate. The plates were left to dry before staining with 10% phosphomolybdic acid (PMA) dissolved in ethanol. Plates were developed by heating until the solvent front was clearly visible.

2.5.7. Formation of apo-proteins

The lipid wash procedure was performed according to Hughes *et al.,* (2019). Purified protein was loaded onto a 5mL HisTrap^{TM} HP Ni-NTA column (Cytiva) and washed three times with 5 column volumes of 50 mM BTP, 500 mM NaCl, 10 mM Imidazole, 50 mM n-octyl β-D-glucopyranoside (β-OG), pH 8, with a 1 h incubation period between washes. The column was then washed with 10 column volumes of 50 mM BTP, 150 mM NaCl, 10 mM Imidazole, pH 8, followed by eluting protein with 50 mM BTP, 150 mM NaCl, 500 mM Imidazole, pH 8.

2.5.8. The GraFix method.

GraFix was performed in accordance with Stark *et al.,* (2010). A continuous density gradient was prepared as follows: 5% sucrose and 30% sucrose solutions were prepared. 0.015% of glutaraldehyde was added to the 30% sucrose solution. Using a gradient pourer and peristaltic pump, the high-density and low-density solutions were mixed inside the Open-Top Thinwall Ultra-Clear (14x89mm) tubes (Beckman Coulter) to a final volume of 11 mL. 500 µL of purified Bd1152$^{37\text{-}283\ I112C\ L238C}$ at 2 mg/mL was loaded on top of the column and centrifuged for 17 h at 273000 x g and 4˚C (SW-41 Ti Beckman Coulter). 500 µL fractions were carefully aspirated and analysed via SDS-PAGE. Fractions containing species of the approximate molecular mass of the Bd1152$^{37\text{-}283\ I112C\ L238C}$ hexamer were loaded onto a Superdex S200 10/300 increase column (GE Healthcare) in SEC buffer (20 mM BTP, 150 mM NaCl, pH = 8). The resulting protein was snap frozen in liquid nitrogen and stored at -80˚C until further use.

2.6. Cryo-EM and negative stain transmission electron microscopy (TEM).

2.6.1. Preparation of negative stain grids

Prior to sample application, carbon-coated copper grids (Agar Scientific) were glow discharged for 60 sec at 35 mA using a GloQube® (Quorum) glow discharge unit. 10 µL of 0.03 mg/mL protein sample was applied and incubated for 1 min before the excess liquid was blotted using filter paper. 10 µL of 2.5 % uranyl acetate solution was added onto the grid, incubated for 30 sec, and blotted using filter paper. Images were taken on a JEOL1400 microscope at 50,000x - 150,000x magnification.

2.6.2. Preparation of graphene oxide coated grids.

QUANTIFOIL® 300 mesh R1.2/1.3 AU grids (Quantifoil Micro Tools) were glow discharged for 180 sec at 40mA using a GloQube® (Quorum) glow discharge unit. 3µL of graphene oxide (GO) solution (0.2mg/mL) (Sigma 763705) was applied to the grid and incubated for one minute. Excess liquid was blotted using filter paper. The grid was then washed three times with distilled water drops and left to air-dry.

2.6.3. Preparation of Cryo-EM grids.

3 µL of purified protein was applied to a QUANTIFOIL® AU R1.2/1.3 300 mesh grid (If grids did not contain an additional GO coating, they were glow discharged for 60 sec at 35 mA using GloQube® glow discharge unit (Quorum). A Vitrobot Mark IV (Thermo Fisher Scientific) was used to blot the grids before plunging into liquid ethane. Cryo-EM grids were screened, and data were collected with a Titan Krios transmission electron microscope (Thermo Fisher Scientific) at the Midlands Regional cryo-EM facility operating at 300 kV or with an Arctica cryo-TEM (Thermo Fisher Scientific) at CNB-CSIC (Madrid, Spain) operating at 200 kV. Data collection parameters are listed in Table 2.10.

2.6.4. Cryo-EM sample optimisation.

To optimise the particle distribution on QUANTIFOIL® AU R1.2/1.3 300 mesh grids, we utilised the VitroEase™ Buffer Screening Kit (Thermo Fisher Scientific) as per manufacturers guidelines. Briefly, 10 µL of purified protein sample was spiked with 3 µL of detergent solution (β-OG (octyl β-D-glucopyranoside), Tween 20 (polyoxyethylene 20), DM (n-decyl-ß-D-maltoside) and FOM (fluorinated octyl maltoside) at 0.5 x CMC and immediately snap frozen with Vitrobot Mark IV (Thermo Fisher Scientific).

2.6.5. Data collection and processing

For more details on data processing please refer to Chapter 4. Briefly, data processing was performed in RELION (v3.1.3 and 4.0 - beta) (Scheres, 2012a, 2012b), cryoSPARC (v.2.12.0) (Punjani *et al.,* 2017) and Scipion (v3.0.12) (de la Rosa-Trevín *et al.,* 2016). The major steps of data processing included motion correction (Motioncorr2) (Zheng *et al.,* 2017) and CTF estimation (CTFFIND4; GCTF) (Rohou and Grigorieff, 2015) (Zhang, 2016). Particles were picked using a combination of different picking approaches including: manual picking, Topaz (Bepler *et al.,* 2019) and crYOLO (Wagner *et al.,* 2019). Extracted particles were subjected to several rounds of 2D classification followed by initial model generation. Models were improved by 3D classification and 3D refinement. Cryo-EM maps were visualised in UCSF ChimeraX (Pettersen *et al.,* 2021).

Table 2.10. Cryo-EM data collection parameters used in this study.

	Bd1153-Bd1152	$Bd1152^{37-283}$	$Bd1152^{37-283}$ I112C L238C	$Bd1152^{37-283}$ I112C L238C GraFix	$Bd1152^{37-283}$ I112C L238C + 0.5 x CMC FOM
Detector	K3	K3	K3	K3	Falcon III
Detector mode	Super resolution bin1	Super resolution counting bin 1	Super resolution counting bin 2	Super resolution counting bin 2	Super resolution counting bin 2
Voltage (kV)	300	300	300	300	200
Pixel size (Å)	0.85	0.85	0.835	0.835	0.855
Defocus (μm)	-2.7 to -1.2 in 0.3 μm intervals	-2.7 to -1.2 in 0.3 μm intervals	-2.3 to -0.8 in 0.3 μm intervals	-1 to -2.5 in 0.3 μm intervals	-1.4 to -3 in 0.2 μm intervals
Dose rate on specimen (e^- /pix/sec)	16.5	16.5	18	19.3	31
No. of frames	45	43	50	50	60
Exposure time (s)	2	3	2	2	2
No. of micrographs	12090	533	9667	1533	1586

2.7. X-ray crystallography

2.7.1. Crystal growth

Proteins were purified as described in section 2.5. After purification with size exclusion chromatography, proteins were concentrated using a 10 kDa MWCO Amicon centrifugal filter unit (MerckMilipore). Crystallisation screens were set up using a sitting drop vapour diffusion method at 18˚C. Crystallisation screens were set up by hand in 96 well CombiClover™ crystallisation plates (Molecular dimensions) (2 µL of protein + 2 µL of condition) or using a Mosquito HTS liquid handling robot (SPT Labtech) (400 nL protein + 400 nL reservoir solution per drop). A number of commercial crystallisation screens (Molecular Dimensions) were used for crystallisation, including Morpheus®, Morpheus® II, The BCS Screen, JCSG Plus™, MemGold™, MemGold2™, MemPlus™, MIDASplus™. Crystallisation of $Bd1152^{37\text{-}283\ I112C\ L238C\ L228A\ F229A}$ occurred in a MemGold™ screen condition comprising 0.015 M Tricine, pH 8.5, 24% w/v PEG 4000. Crystals were fished with an appropriately sized circular litholoop™ (Molecular Dimensions) and transferred to a cryoprotectant solution containing the initial condition constituents supplemented with 25% v/v PEG 400. Crystals were incubated for 5 min before being snap freezing in liquid nitrogen. Crystals were then stored in cryo-cooled pucks until data collection.

2.7.2. Data collection and processing.

X-ray diffraction data were collected at the European Radiation Synchrotron Facility (ERSF) (Grenoble, France) beamline ID30A-3 and indexed to the H3 space group. Datasets were collected using standard methods (images = 1800, oscillations = 0.1° Ω, transmission = 100%, exposure = 0.03s, wavelength = 0.9763Å). Collected datasets were automatically reduced and processed by the ISPyB integrated pipelines (xia2 dials (Winter *et al.,* 2022), fast_dp (Winter and McAuley, 2011), xia2 3dii (Winter et al., 2013); autoPROC (Vonrhein *et al.,* 2011) and autoPROC + STARANISO (Vonrhein *et al.,* 2018).

The resolution limit and size of the beam were adjusted depending on the crystal. Crystal data were solved with molecular replacement (MR) using the Phaser program (McCoy *et al.,* 2007) integrated into the PHENIX software (Torices and Muñoz-Pajares, 2015), and manual built in COOT (Emsley *et al.,* 2010). Models were refined via repeated cycles of refinement using phenix.real_space_refine (Afonine *et al.,* 2012) and ISOLDE (Croll, 2018). Model validity was assessed with the MolProbity program, which is integrated within PHENIX (Chen *et al.,* 2010).

2.8. Identification of *Bdellovibrio bacteriovorus* MCE proteins.

A local BLASTP (Madden *et al.,* 1996) search was used to identify MCE domain-containing proteins in *Bdellovibrio bacteriovorus* strain ATCC 15356 / DSM 50701 / NCIB 9529 / HD100 by providing full-length sequences of *E. coli* MlaD (UniProt entry name: P64604), PqiB (UniProt entry name: P43671) and YebT/LetB (UniProt entry name: P76272) as inquiry sequences. In addition, the PFAM database (Mistry *et al.,* 2021) was scanned for the presence of MCE domain-containing sequences in *Bdellovibrio bacteriovorus* HD100. The amino acid sequences were retrieved from the NCBI (National Center for Biotechnology Information) database (http://www.ncbi.nlm.nih.gov) in FASTA format.

2.9. Structural prediction

Alphafold2 (Jumper *et al.,* 2021) was used to predict the structures of all proteins of interest within this study. Amino acid sequences were retrieved from the NCBI database and submitted to Alphafold2 implemented in the ColabFold notebook (https://colab.research.google.com/github/deepmind/alphafold/blob/main/notebooks/AlphaFold.ipynb). Default parameters were used in predictions.

2.10. Foldseek analysis

Foldseek (Kempen *et al.,* 2023) was used to identify 2-MCE domain-containing proteins. We compared the Alphafold model of *B. bacteriovorus* 2-MCE domain-containing protein Bd1152 against all organisms using default parameters (https://search.foldseek.com/search).
For validation of the Foldseek method, we manually inspected Alphafold structures of all identified 2-MCE domain-containing proteins.

2.11. Zernike3D

The Zernike3D algorithm was implemented within *Xmipp* (de la Rosa-Trevín *et al.,* 2013) and is available through Scipion (de la Rosa-Trevín *et al.,* 2016) under a protocol named flexutils – angular align – Zernike3D.

For heterogeneity analysis, a Bd1152$^{37\text{-}283\ I112C\ L238C}$ reference map (C1), its associated particles (883461) (downsampled to box size of 125 voxels) and binary mask were used as inputs for Zernike3D analysis (Herreros *et al.,* 2023, 2021). Zernike3D coefficient space was visualised with UMAP via the *Xmipp* protocol named *Annotate space*. Particles were extracted, and 3D maps were reconstituted with *ab-initio* (cryoSPARC) and *non-uniform* refinement (cryoSPARC). The resulting cryo-EM maps and x-ray structure of Bd1152$^{37\text{-}283\ I112C\ L238C\ L228A\ F229A}$ were subjected to IMODFIT (Lopéz-Blanco and Chacón, 2013) flexible fitting analysis implemented in the Scipion framework.

2.12. CryoDRGN

For heterogeneity analysis, particles were downsampled to the box size of 125 voxels and analysed using cryoDRGN v1.0 with default settings (Zhong *et al.,* 2021).
The particle stack was used to train an 8-dimensional latent variable model in cryoDRGN for 39 epochs. After training, the model was analysed using the “cryoDRGN analyse” pipeline, where k-means clustering was performed with k=20, and the conformation landscape visualised with UMAP.

2.13. Model building

The Bd1152$^{37\text{-}283\ I112C\ L238C}$ cryo-EM map and mutated sequence (residues 1-37 deleted, I112C and L238C substitutions) were used as an input for ModelAngelo (Jamali *et al.,* 2022). The resulting model was refined with Phenix real_space_refine (Afonine *et al.,* 2018).

Chapter 3 - Identification of MCE domain-containing proteins within *Bdellovibrio bacteriovorus* HD100

3.1. Introduction

MCE domain-containing proteins are ubiquitous amongst gram-negative bacteria and have also been identified in plants (Casali and Riley 2007). Nevertheless, research has mainly focused on characterising MCE proteins in Actinobacteria, with a number of studies suggesting their roles in pathogenicity and lipid transport (Cantrell *et al.,* 2013; Gioffré *et al.,* 2005; Klepp *et al.,* 2012). Recently, there has been a growing interest in MCE proteins found in proteobacteria, where they have been proposed to serve as intermembrane phospholipid transport machineries. The best-characterised MCE protein in *E.coli* is MlaD, a member of the Mla pathway (Ekiert *et al.,* 2017; Hughes *et al.,* 2019; Kamischke *et al.,* 2019). Additionally, *E.coli* has two other MCE proteins PqiB and LetB (formerly known as YebT), containing three and seven MCE rings, respectively. These proteins are believed to span the periplasm and provide a direct link between the IM and OM. Both proteins have been partially structurally characterised (Ekiert *et al.,* 2017; Isom *et al.,* 2020), however, their function remains elusive.

Overall, the majority of research on MCE domain-containing proteins has centred upon Actinobacteria and the model organism *E. coli*, with a few studies focusing on other organisms (*Pseudomonas aeruginosa* (Munguia *et al.,* 2017) and *Acinetobacter baumannii* (Kamischke *et al.,* 2019)). Consequently, there remains a large knowledge gap regarding the structure and function of MCE proteins in other bacterial species. Several bioinformatic analyses of MCE proteins have been published. Casali and Riley (2007) focused only on assessing the distribution of *mce* operons in *Actinomycetales,* with a brief insight into genes encoding MCE proteins in gram-negative bacteria. A more detailed bioinformatic analysis of MCE proteins was published by Isom *et al.,* (2017), in which the architectures of all bacterial MCE proteins were identified. 97% of all identified MCE proteins were classified into the four architectures shown in Figure 3.1.

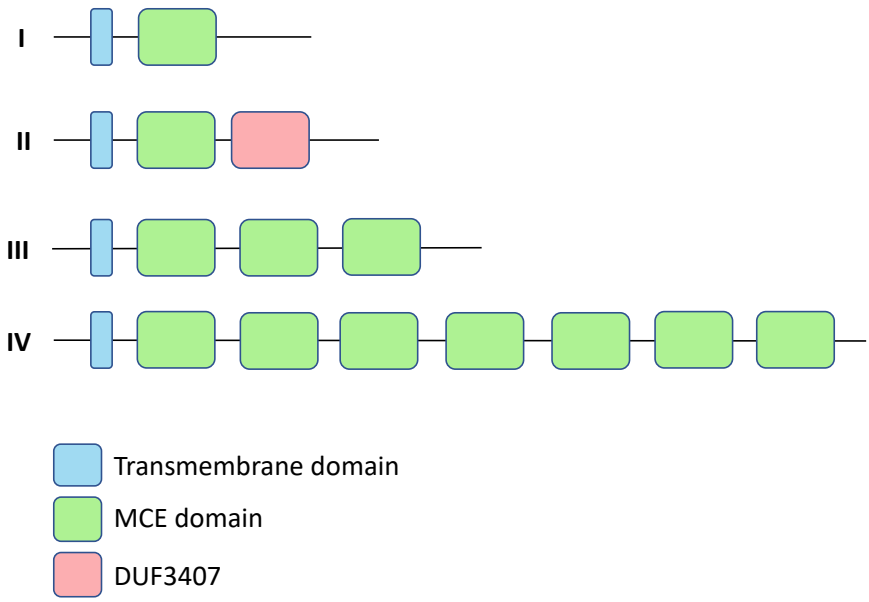


Figure 3.1. The top four (I-IV) MCE domain architectures across bacterial kingdom. Reproduced from Isom *et al.,* (2017).

The most abundant architecture of MCE proteins is type I, which comprises a single MCE domain and includes MlaD, the well-known *E. coli* Mla system component. The type II architecture is the second most abundant, comprising a single MCE domain and a DUF3407 domain (Cholesterol Uptake Porter), which are specific to Actinobacteria. Type III and IV architectures consist of three and seven MCE domains, respectively. Both these architectures appear specific to proteobacteria and are referred to as PqiB (type III) and LetB (type IV) in *E. coli*. Additionally, analysis of neighbouring genes by Isom *et al.,* (2017) revealed that the type I MCE proteins are typically located adjacent to genes encoding permease and ATP-binding domains.

3.2. Chapter Aims

As discussed in section 1.6, bioinformatic analysis conducted by Tajabadi *et al.,* (2017) suggested the potential significance of lipid transport systems in the lifecycle of *Bdellovibrio bacteriovorus.* MCE-domain containing proteins are known to play a role in lipid transport in other bacteria, therefore we anticipate that they provide a similar role in this predatory species. As such, the objective of this chapter was to perform a bioinformatic analysis of MCE domain-containing proteins within the *Bdellovibrio bacteriovorus* HD100 genome, with the aim of identifying potential targets for their subsequent structural and functional characterisation.

3.3. Results

3.3.1. Identification of MCE domain-containing proteins within *B. bacteriovorus HD100,* prediction of protein domain architectures and neighbouring gene analysis.

BLASTP (Madden *et al.,* 1996), PFAM (https://pfam.xfam.org) and Operon Mapper tool (Barabote *et al.,* 2007) were utilised to identify MCE domain-containing proteins within *B. bacteriovorus* and analyse their domain architectures and neighbouring genes.

A total of five MCE proteins were identified within *Bdellovibrio bacteriovorus* HD100 and summarised in Table 3.1. Bd1770, Bd3120, Bd3413, and Bd3803 are predicted to have a similar domain architecture consisting of a single MCE domain, whilst Bd1152 is predicted to harbour two MCE domains. Although an n-terminal transmembrane helix was identified in all MCE-domain proteins via TMHMM - 2.0 (Krogh *et al.,* 2001) in *B. bacteriovorus* HD100, no signal peptide could be identified for any of these proteins via the SignalP 6.0 (Teufel *et al.,* 2022).

The Operon Mapper tool was used to investigate operons containing MCE proteins (Figure 3.2). All genes within the operons were analysed with BLASTP (significant matches were identified as those possessing an E-value < 1.0) and PFAM to identify homologous proteins and determine their domain architectures, respectively. Results revealed that single MCE domain-containing proteins are located in operons with proteins themselves predicted to contain permease and ATP-binding domains, showing homology to the *E. coli* MlaE and MlaF proteins, respectively. Bd1152 is located within an operon together with Bd1151 and Bd1153. Bd1151 is small 53aa protein with a predicted c-terminal transmembrane helix with no homologous proteins, whilst Bd1153 is a 169aa protein showing 37% sequence identity to *E. coli* PqiA.

It is important to note that homologues of the MlaA, MlaC, MlaB and PqiC proteins, members of the *E. coli* Mla and Pqi systems, respectively, were not identified within the *B. bacteriovorus* HD100 genome. All other proteins predicted to be located in operons together with an MCE-domain protein were also investigated, and their domain architectures were determined (Appendix 1).

Table 3.1. MCE-domain containing proteins within Bdellovibrio bacteriovorus HD100.

Gene ID	Location	Description	A number of predicted MCE domains	Homologous proteins	Sequence identity
Bd1152	1093150..1094001	PqiB family protein; hypothetical protein	2	Intermembrane transport protein PqiB (*E. coli*)	31.3%
Bd1770	1703091..1704044	ABC-type multidrug transport system substrate-binding protein	1	No significant similarity found within *E. coli*	n/A
Bd3120	3024468..3025844	ABC-type organic solvent transporter, solute-binding periplasmic protein.	1	No significant similarity found within *E. coli*	n/A
Bd3413	3324478…3325293	ABC-type transport system involved in resistance	1	Outer membrane lipid asymmetry maintenance protein MlaD (*E. coli*)	28.05%
Bd3803	3671782..3672558	ABC-type organic solvent resistance transport system, permease protein	1	Outer membrane lipid asymmetry maintenance protein MlaD (*E. coli*)	26.61%

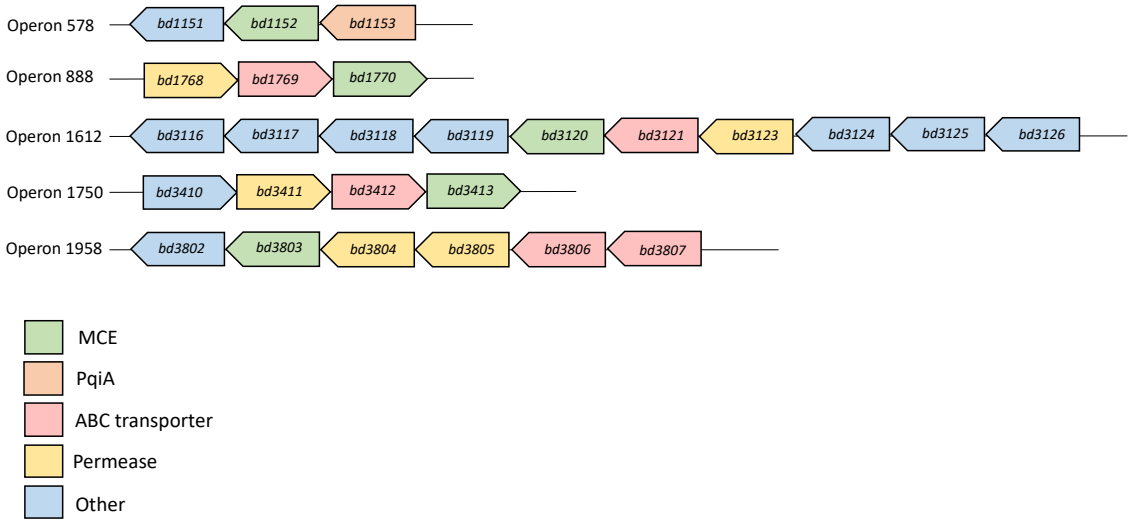


Figure 3.2. The architectures of operons in *Bdellovibrio bacteriovorus* HD100 that harbouring MCE proteins.

Operon components were predicted using the Operon mapper tool (Barabote *et al.,* 2007). Domains were predicted using PFAM (https://pfam.xfam.org). Each operon contains one MCE protein (green). Genes shown in orange are predicted to contain a PqiA domain and exhibit sequence identity to *E. coli pqiA*. Genes shown in pink are predicted to contain ABC transporter-like domains and exhibit homology to *E. coli mlaF,* while yellow indicates genes predicted to contain permease domains and exhibit homology to *E. coli mlaE.* The blue colour represents proteins containing other domains/hypothetical proteins.

3.3.2. Structural prediction

The structures of MCE domain-containing proteins in *B. bacteriovorus* HD100 were predicted using Alphafold2 (Jumper *et al.,* 2021) (Figure 3.3). Each MCE domain adopts a seven stranded β-barrel fold, which is predicted with high model confidence (pLDDT > 90). Bd1152 is predicted to contain two MCE domains, whilst Bd1770, Bd3413, Bd3120 and Bd3803 each contain a single MCE domain and a long c-terminal helix that likely spans the periplasm. Nevertheless, the model confidence for these helices is low (pLDDT<50). Interestingly, Bd3120 is also predicted to contain a porin-like fold at its c-terminus, which has not been previously identified within the Pfam or Hmmer databases.

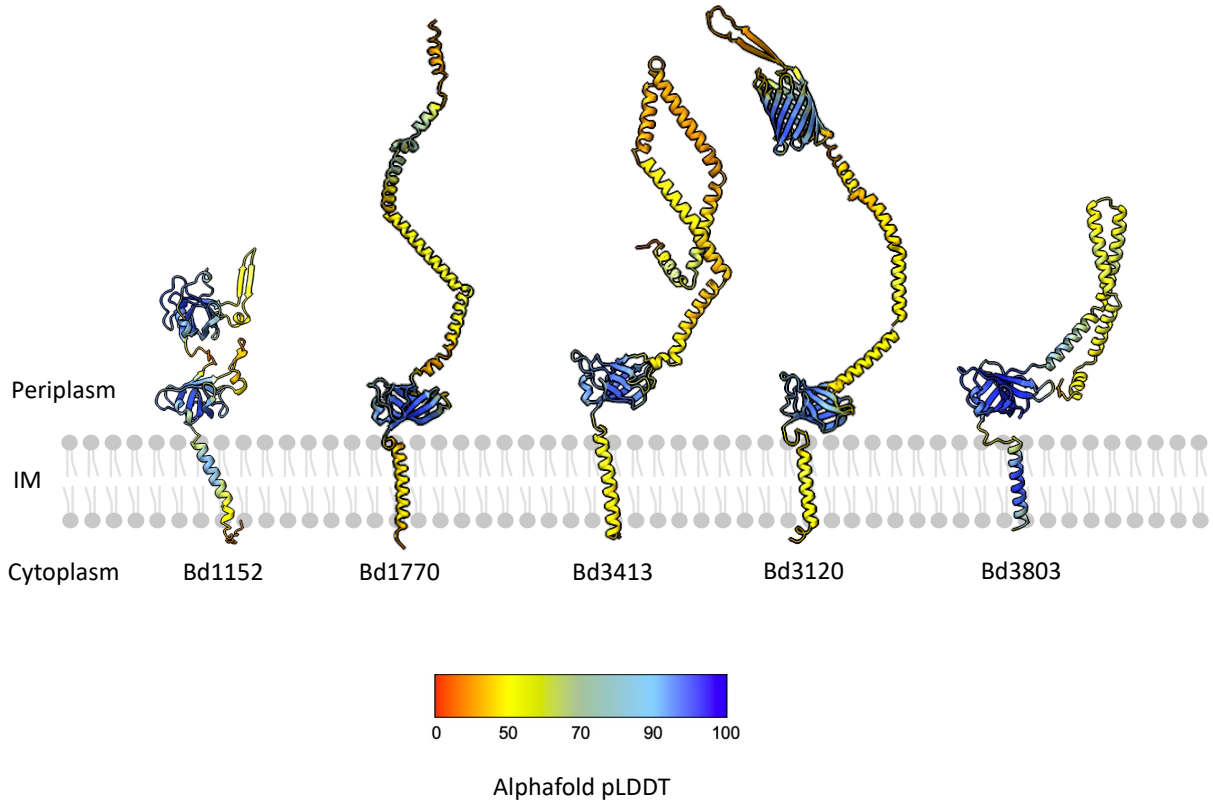


Figure 3.3. The Alphafold structures of all MCE-domain proteins within *Bdellovibrio bacteriovorus* HD100.

The structures are depicted in their predicted cellular location within the inner membrane (IM). The Structures are coloured by per-residue pLDDT scores. A pLDDT colour key is provided at the bottom of the figure where the spectrum from blue to red corresponds to decreasing pLDDT score (Jumper *et al.,* 2021). The models were visualised with ChimeraX (Pettersen *et al.,* 2021).

3.3.3. Abundance of 2-MCE domain-containing proteins

Within the genome of *Bdellovibrio bacteriovorus* HD100, it is predicted that Bd1770, Bd3120, Bd3413, and Bd3803 each contain a single MCE domain, representing the most prevalent Type 1 MCE protein architecture. In contrast, Bd1152 is predicted to contain 2-MCE domains. This finding is of particular interest as Isom *et al.,* (2017) reported that 2-MCE domain-containing proteins comprise only 0.7% of all identified MCE proteins within the bacterial kingdom, thus signifying their notably low abundance.

We utilised the recently developed online Foldseek server (Kempen *et al.,* 2023) to further examine the abundance of 2-MCE domain-containing proteins. The analysis revealed that only 2.5% of all MCE proteins identified by Foldseek correspond to 2-MCE domain-containing proteins. A slightly higher percentage was identified in the current analysis, when compared to Isom *et al.,* (2017), however, this is most likely due to some of the 2- MCE domain containing proteins exhibiting misannotated domain architecture. The development of Alphafold and Foldseek facilitated the identification of 2-MCE domain-containing proteins, more accurately, based on their predicted structures rather than their amino acid sequences alone.

2-MCE domain-containing proteins were found in pathogenic bacteria *Klebsiella pneumoniae* and *Salmonella enterica,* as well as bacteria possessing unusual lifecycles, including *Bdellovibrio bacteriovorus* and *Alcanivorax borkumensis* (Figure 3.3). Moreover, whilst a single MCE-domain-containing protein was previously identified in plants (Casali and Riley 2007), FoldSeek revealed the presence of a 2-MCE domain-containing protein in fungi of the phylum mucoromycota. (Figure 3.4).

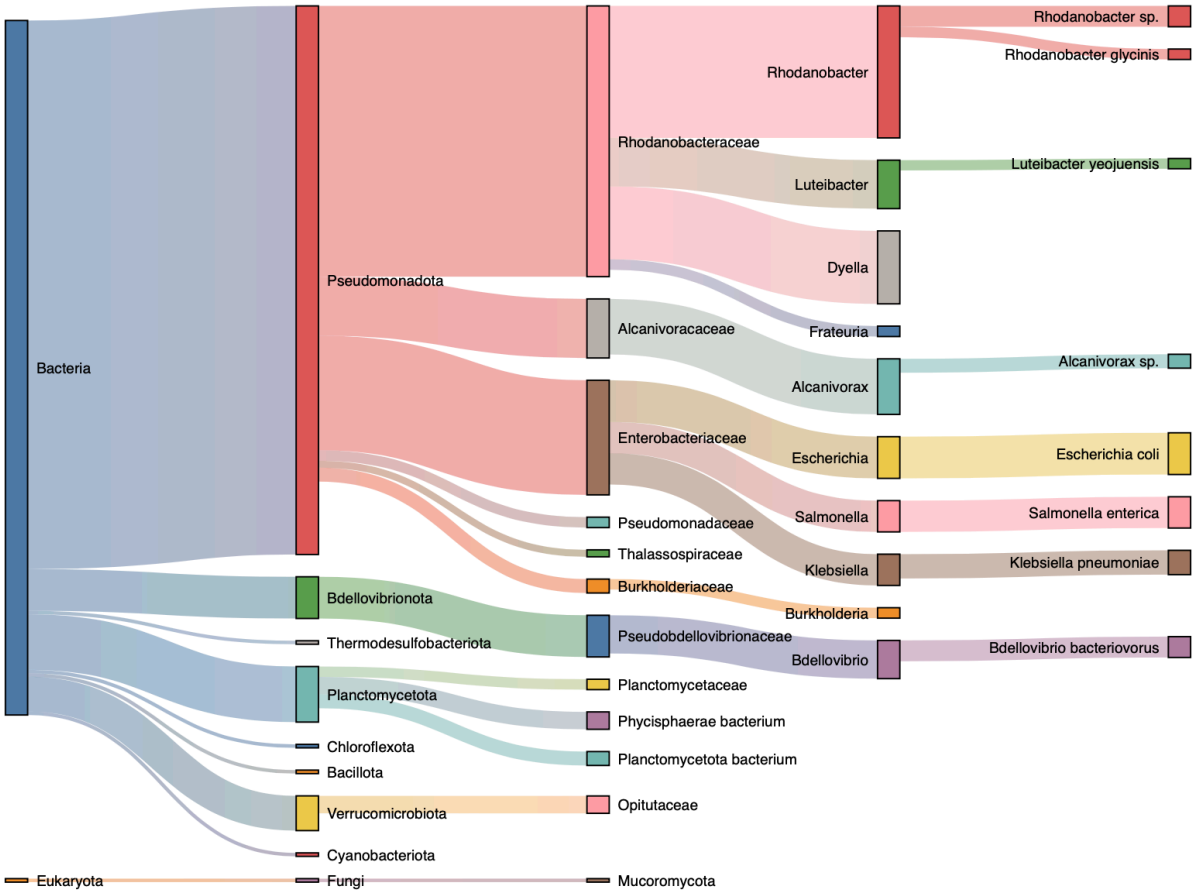


Figure 3.4. Sankey diagram demonstrating the distribution of 2-MCE domain-containing proteins across all kingdoms of life, as identified by FoldSeek analysis. The width of the line is proportional to the number of genes (the raw data is displayed in Appendix 2).

3.3.4. Identification of differentially expressed genes within the *B. bacteriovorus* HD100 operons 578, 888, 1612, 1750 and 1958.

A gene expression profile obtained by Karunker *et al.,* (2013) was used to identify differentially expressed genes within the *B. bacteriovorus HD100* operons 578, 888, 1612, 1750 and 1958 (Figure 3.2). Karunker *et al.,* (2013) used RNA-seq to generate a single-base resolution map of the *B. bacteriovorus* HD100 transcriptome during the attack phase (AP) and growth phase (GP). Samples were collected at 60 min (AP) and 180 min (GP) after mixing the predator and prey.

Karunker *et al.,* (2013) revealed that the *bd1152* and *bd1153* genes present in operon 579, which display homology to *E. coli pqiB* and *pqiA*, respectively, were not detected in the samples (RPKM = 0). Nevertheless, *bd1151* is significantly upregulated during GP (144-fold). As all three genes appear to reside within a single operon, we anticipate similar upregulation of *bd1152* and *bd1153* during GP (Karunker *et al.,* 2013), and the absence of reads might be attributed to technical limitations in the analysis, however, this needs further validation.

Operons 888 and 1750 are upregulated during AP, with MCE-containing genes *bd1770* and *bd3413* showing a 9.7 and 14.5-fold change, respectively. Operons 1612 and 1958 are upregulated during GP, with MCE-containing genes *bd3120* and *bd3803* showing a 9.7 and 14.5-fold change, respectively.

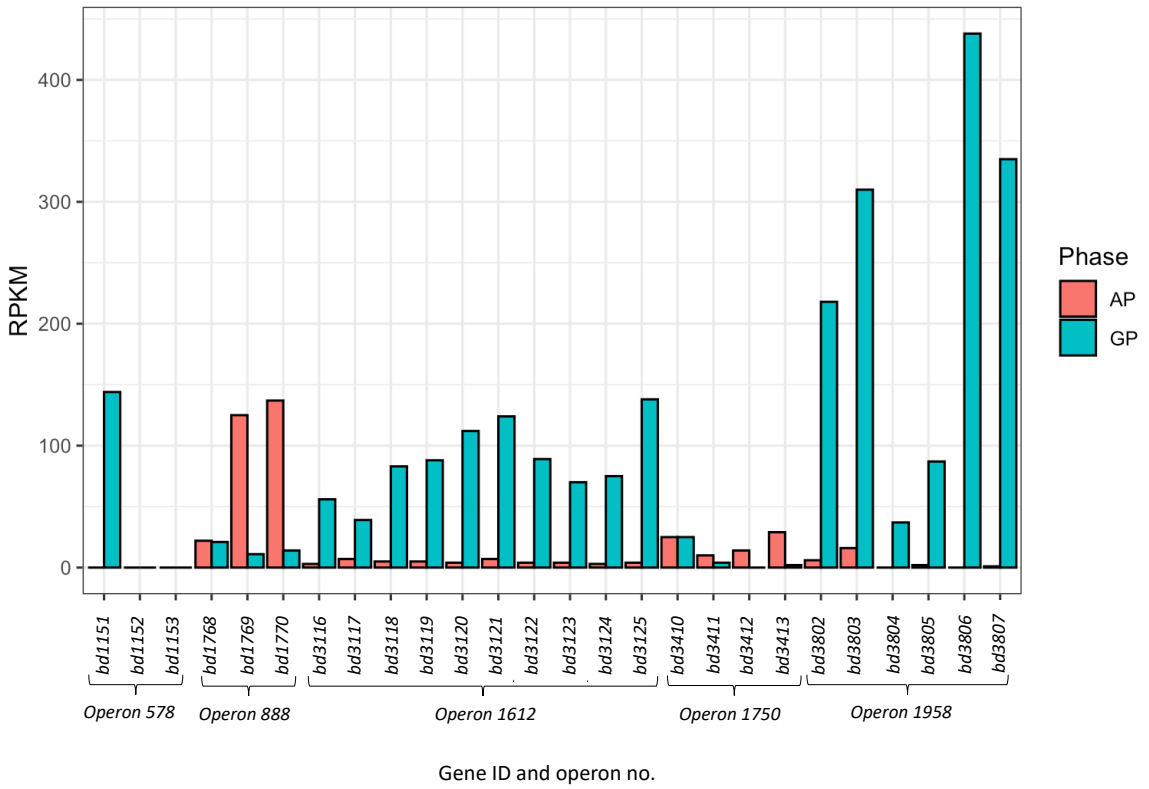


Figure 3.5. Gene expression levels of the *B. bacteriovorus* HD100 operons containing *mce* genes during the attack phase (AP) and growth phase (GP), as reported by Karunker *et al.*, (2013).

The y-axis represents the Reads Per Kilobase of transcript per Million mapped reads (RPKM) values, whilst x-axis indicates the gene ID and operon number. Samples were collected at 60 min (AP) and 180 min (GP) after mixing *B. bacteriovorus* HD100 with a prey cells.

3.4. Discussion

Previous research has strongly implicated the role of MCE proteins in the intermembrane PL transport within gram-negative bacteria (Hughes *et al.,* 2019; Isom *et al.,* 2020; Low *et al.,* 2021; Tang *et al.,* 2021). Nevertheless, these studies have mainly focused on the model organism *E. coli*, highlighting a significant knowledge gap in understanding the functionality of MCE proteins in other bacterial species. In this study, we focused on the predatory bacterium *Bdellovibrio bacteriovorus* and hypothesised that MCE proteins might play an important role in the attack and growth phases of this predator. The aim of this chapter was to probe the MCE family proteins within the *B. bacteriovorus* HD100 genome, subsequently laying a foundation for their further structural and functional investigation.

3.4.1. A novel porin-like fold is predicted at the c-terminus of Bd3120.

Based on bioinformatic analysis, our findings reveal the presence of five MCE proteins in the *B. bacteriovorus* HD100 genome. Among these, Bd1770, Bd3413, and Bd3803 exhibit a shared protein architecture consisting of a TM region and a single MCE domain. Initially, Bd3120 was also predicted to comprise a TM segment and a single MCE domain, but further analysis using Alphafold revealed the presence of a porin-like fold at the c-terminus. Structural homology analysis using DALI (Holm and Sander, 1995) identified close similarity between the porin-like domain at the c-terminus of Bd3120 and the uncharacterised *E.coli* OM protein YfaZ (DALI z score 14.9) (Figure 3.6).

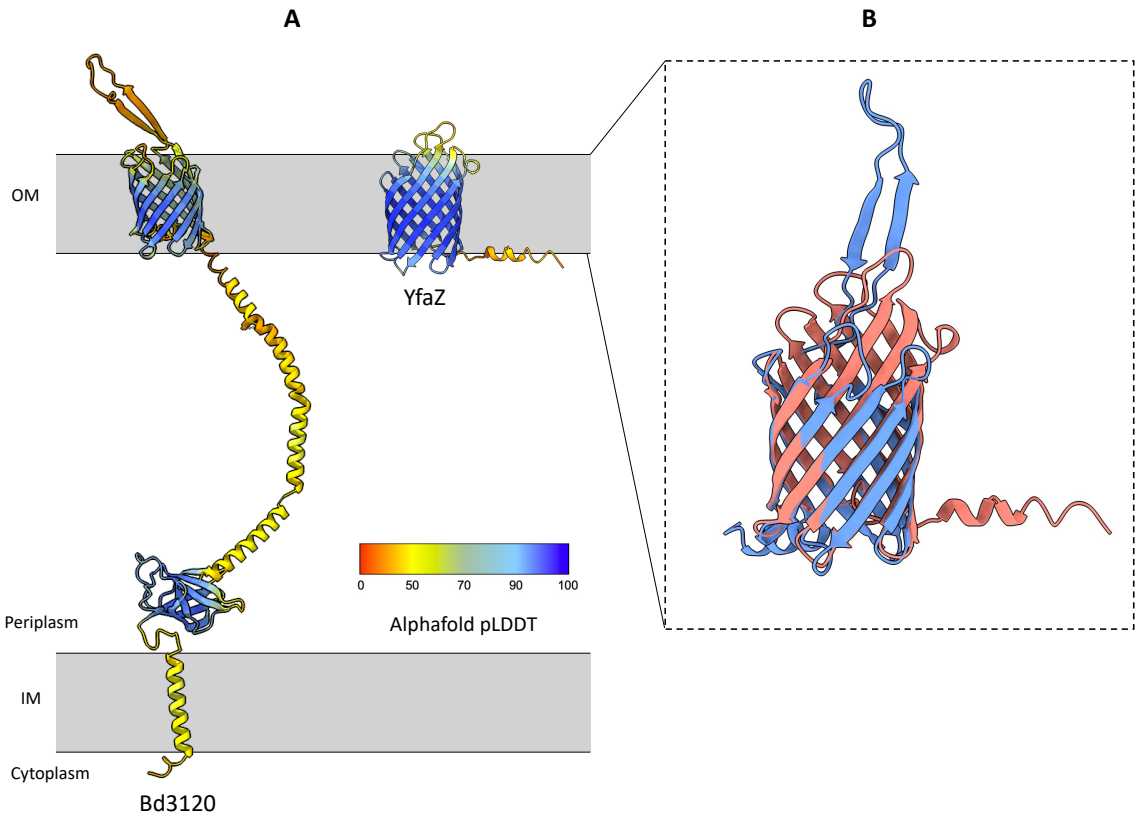


Figure 3.6. Structural comparison of the c-terminal porin-like feature of *Bdelovibrio bacteriovorus* HD100 Bd3120 and *E. coli* YfaZ.

A) The Alphafold structures of Bd3120 (AF-Q6MIN3-F1) and YfaZ (AF-P76471-F1) are depicted in their predicted cellular locations within the IM (Inner membrane) and OM (Outer membrane). The structures are coloured according to their pLDDT scores. A pLDDT colour key is provided in the figure where the spectrum from blue to red corresponds to decreasing pLDDT score (Jumper *et al.,* 2021). B) The Alphafold structures of Bd3120 (residues 280-458) (blue) and *E. coli* YfaZ (pink) were superimposed, with an RMSD of 0.920 Å between 40 pruned atom pairs (across all 145 pairs: 21.011). The models were visualised with ChimeraX (Pettersen *et al.,* 2021).

To date, all elucidated structures of MCE domain-containing proteins demonstrated hexameric multimerisation (Ekiert *et al.,* 2017; Isom *et al.,* 2020). Nonetheless, gram-negative bacteria typically possess porins exhibiting a trimeric arrangement (Arunmanee *et al.,* 2016). This leads us to speculate, whether Bd3120 could form a hexameric arrangement, thus resulting in completely novel hexameric porin at its c-terminus.

3.4.2. Single MCE domain-containing proteins in *B. bacteriovorus*.

Single MCE domain-containing proteins within *B. bacteriovorus* HD100 appear to be localised with proteins predicted to contain permease and ATP-binding domains, exhibiting homology to the *E. coli* MlaE and MlaF proteins, respectively. This observation strongly suggests that they may form an ABC transport complexes, similar to *E.coli* MlaFEDB (Chi *et al.,* 2020).

It is important to note that homologous sequences of the lipoproteins MlaA and PqiC, which are members of the *E. coli* Mla and Pqi systems, respectively, were not identified in the *B. bacteriovorus* HD100 genome. It is unclear whether any functional homologues exist that could execute these functions, or whether *B. bacteriovorus* has a unique method for operating these putative lipid transport pathways.

Moreover, the absence of homologs of *E. coli* MlaC within the *B. bacteriovorus* genome raises an important question regarding the mechanism by which the homologous proteins of the *E. coli* Mla system in *B. bacteriovorus* facilitate the translocation of substrates across the periplasm. Nevertheless, Alphaphold structural prediction has revealed that all single MCE domain-containing proteins within *B. bacteriovorus* possess long helices that likely extend across the periplasm, forming tunnel-like structures. As suggested for the *E.coli* PqiB (Ekiert *et al.,* 2017), LetB (Isom *et al.,* 2020) and *M. smegmatis* MCE1 machinery (Chen *et al.,*

2023), these periplasmic tunnels may provide a direct link between the IM and OM for the cargo transport and eliminate the necessity of periplasmic chaperones.

Additionally, it is worth mentioning that *B. bacteriovorus* has an unexpectedly high numbers of pseudogenes and incomplete systems. For example, Tajabadi *et al.,* (2017) reported that LolB, an OM membrane lipoprotein required for inserting lipoproteins into the OM in *E. coli*, was not identified in the *B*. *bacteriovorus* HD100 genome.

3.4.3. The unique characteristics and potential functional roles of the Bd1153-Bd1151 operon.

The *bd1152* gene is located within an operon harbouring the *bd1151* and *bd1153* genes. The *bd1153 gene* exhibits sequence similarity to *E. coli pqiA*, whilst Bd1151 is a small 53aa protein predicted to possess a c-terminal transmembrane helix. The functional role of Bd1151 is completely enigmatic as it lacks structural features resembling any of the *E. coli* Pqi or Mla system components.

Bd1152 is predicted to contain 2-MCE domains. According to our analysis, 2-MCE domain-containing proteins comprise only 2.5% of all identified MCE proteins highlighting their rare abundance. Additionally, 2-MCE domain-containing proteins appear to be present in bacteria with unique lifestyles (*B. bacteriovorus* and *A. borkumensis*), as well as eukaryotic fungi of the mucoromycota lineage. The projected length of 2-MCE rings is approximately 80 Å), which falls short of spanning the periplasm (~250 Å, (Seltmann and Holst, 2002)). This discrepancy suggests the potential involvement of unique interacting partners that could effectively bridge the periplasmic gap between 2-MCE ring proteins and the OM. Consequently, these observations collectively imply that proteins with 2-MCE domains may

possess distinctive functional roles and mechanisms of action, potentially diverging from previously characterised MCE proteins (Chi *et al*., 2020; Isom *et al.,* 2020; Chen *et al.,* 2023)

3.4.4. The role of MCE proteins within *B. bacteriovorus*

B. bacteriovorus HD100 has a smaller genome compared to *E. coli* (3.8 Mb and 5.5 Mb, respectively), however, it exhibits a substantially greater number of transporters than *E. coli* (Tajabadi *et al.,* 2017). As a predatory bacterium, *B. bacteriovorus* relies on the secretion of proteins to degrade the cytoplasmic contents of its prey and transport them back to the periplasm and ultimately into its own cytoplasm (Bratanis *et al.,* 2020). Additionally, *B. bacteriovorus* is known to uptake and incorporate largely unmodified OM prey cell components, such as LPS (Kuenen and Rittenberg, 1975; Stein *et al.,* 1992) and OMPs (Diedrich *et al.,* 1983; Talley *et al.,* 1987). The unique lifecycle of *B. bacteriovorus* necessitates large-scale inter-membrane and inter-organism cargo transport, for which the five MCE proteins identified in this study might hold a significant importance. Additionally, the variation in gene expression levels among the *B. bacteriovorus* MCE family proteins raise intriguing questions regarding their potential functional divergence or the requirement for specific regulatory mechanisms.

3.5. Selection of *B. bacteriovorus* MCE domain-containing protein for experimental characterisation.

The bioinformatic analysis conducted in this study enabled the identification of five MCE domain-containing proteins within the genome of *B. bacteriovorus* HD100. We initially focused on studying Bd1770 and its associated Bd1768-Bd1770 membrane complex due to its upregulation during attack phase. Whilst the initial expression and purification of this complex appeared promising, we were unable to obtain a homogeneous sample suitable for structural analysis (Appendix 14).

Our initial attempts to purify the Bd1153-Bd1152 membrane complex and a soluble Bd1152 construct, however, yielded more promising results. Although these samples were prone to aggregation, they displayed much higher purity and homogeneity. As a result, we decided to shift our focus to the Bd1153-Bd1152 membrane-associated complex, as expression and purification trials proved to yield promising results.

Based on our bioinformatics analysis, it is reasonable to postulate that the Bd1151-Bd1153 operon may serve as a direct homolog to the PqiABC membrane complex in *E. coli*. However, whilst *B. bacteriovorus* Bd1152-Bd1153 shows sequence identity to *E. coli* PqiAB, they are predicted to possess distinct structural features. For instance, Bd1152 is anticipated to contain two MCE rings, in contrast to PqiB, which comprises three MCE rings. Additionally, a periplasm-spanning coiled coil, identified in PqiB (Ekiert et al., 2017, Cooper 2023), seems to be absent in Bd1152. These observations suggest a potential for a unique functional mechanism in Bd1152. As a result, this system presents an attractive target for further structural investigations as it could potentially provide insight into a novel transport mechanism.

Additionally, to date, only 1, 3, and 7 MCE domain-containing proteins have been experimentally investigated. Therefore, structural information is scarce regarding MCE proteins with an atypical number of MCE domains. Bd1152, a two MCE domain protein, may represent a novel mechanism of action compared to previously characterised systems such as MlaD, PqiB, LetB, and MCE1 (Ekiert *et al.,* 2017; Chi *et al*., 2020; Isom *et al.,* 2020; Chen *et al.,* 2023)

3.6. Conclusion

The current bioinformatic analysis allowed us to identify five MCE domain-containing proteins within the *B. bacteriovorus* HD100 genome and assess their structural and genomic characteristics. Whilst all *B. bacteriovorus* HD100 MCE proteins are predicted to harbour a typical MCE domain fold, Alphafold-based structure determination revealed several unique features, including an atypical number of MCE domains, long periplasmic helices, and a c-terminal porin-like structure. The diversity of MCE domain-containing proteins within *B. bacteriovorus* HD100, accompanied by the variation of expression levels at different life stages, suggest that they may possess distinctive functions and mechanisms of action. The current bioinformatic analysis has provided a starting point for the experimental investigation of these putative transporters. Finally, we have selected to study the Bd1153-Bd1151 operon due to its homology to the *E. coli* PqiABC system and promising initial expression and purification results.

Chapter 4 – Cryo-Electron Microscopy of Bd1152.

4.1. Introduction

3D cryogenic - electron microscopy (cryo-EM) has revolutionised the field of structural biology since its first application to the bacteriophage tail in 1986 (De Rosier and Klug, 1986). The breakneck progress has been facilitated by improvements in electron detection systems, microscope stability and data processing software, which have collectively contributed to the "resolution revolution" (Kühlbrandt, 2014). Cryo-EM currently allows the solving of protein structures to atomic resolution (Nakane *et al.,* 2020; Yip *et al.,* 2020), which makes this method one of the most attractive tools in the structural biology field.

The ongoing "resolution revolution" era of cryo-EM particularly benefits the research of membrane proteins (MPs), which are infamously difficult to crystallise (Ostermeier and Michel, 1997). By side–stepping the necessity of crystals, cryo-EM offers an alternative to gain structural insights into MPs at near-atomic resolution. This includes important drug targets such as transport proteins (Yao et *al.,* 2020) and channels (Baker *et al.,* 2021). In addition to visualising MPs solubilised in detergents, cryo-EM grants an opportunity to study MPs reconstituted into more native environments, such as lipid nanodiscs (Efremov *et al.,* 2017; Kalienkova *et al.,* 2020) or proteoliposomes (Yao *et al.,* 2020), providing valuable insight into protein-lipid interactions.

Furthermore, cryo-EM has emerged as a powerful tool for studying protein dynamics, with the recent development of advanced algorithms such as CryoDRGN (Zhong *et al.,* 2021) and Zernike3D (Herreros *et al.,* 2021, 2023). The rapid progress in the cryo-EM field has significantly contributed to our understanding of previously intransigent proteins and complex systems. This chapter will subsequently discuss the underlying principles of cryo-EM followed by the cryo-EM pipeline leading to the structural characterisation of Bd1152.

4.1.1. Sample preparation for TEM.

Sample preparation for electron microscopy is a bottleneck in protein structure determination, often requiring iterative optimisation by the user. The first crucial step is to obtain a homogenous protein sample under ideal conditions, making it possible to identify monodispersed particles in a variety of orientations. For soluble proteins, this includes finding appropriate buffer conditions and protein concentration, however, for MP sample preparation, the choice of membrane mimetic is an additional variable that requires extensive optimisation (Kampjut *et al.,* 2021; Ratkeviciute *et al.,* 2021).

The most straightforward method for visualising protein samples under electron microscopy is negative staining. A droplet of the sample is applied onto an EM grid, followed by embedding the protein within a thin layer of heavy metal salts (Scarff *et al.,* 2018) (Figure 4.1). Uranyl acetate (UA) is the most commonly used stain due to its ability to provide optimal contrast (personal communication with Dr. Christos Savva), however, other types of stains, such as uranyl formate, UA-zero, and phosphotungstic acid, are also available.

Heavy metals are more resistant to radiation damage and can absorb electrons more readily, however, due to the grain size of the stain, the resolution that is usually achieved is ~18 Å (Fabre *et al.,* 2017; Ha *et al.,* 2016; Scarff *et al.,* 2018). Nevertheless, negative staining is an excellent and quick technique for assessing sample quality and obtaining general information about particle shape, size and symmetry.

To obtain a high-resolution protein structure cryo-EM is employed. Proteins are visualised in a close-to-native aqueous environment by rapidly freezing the sample into a thin layer of vitreous ice. This allows the protein sample to be visualised in an extremely high vacuum and somewhat protects from radiation damage (Taylor and Glaeser, 1976). Essentially, 2-3 µL of

the specimen is applied to the grid in a humidity-controlled chamber. The excess sample is removed by blotting and the grid is rapidly plunged into the cryogen; typically, liquid ethane that has been pre-cooled with liquid nitrogen (Adrian *et al.,* 1984; Passmore and Russo, 2016) (Figure 4.1). Cryo-EM grids are typically prepared utilising semi-automated plunge freezers such as the EM GP2 Automatic plunge freezer (Leica) or Vitrobot IV (ThermoFisher).

In an ideal cryo-EM grid, protein particles should be evenly distributed in a variety of orientations throughout an even ice layer, without interacting with air-water interference. Nevertheless, in reality, protein specimens often exhibit aggregation, adoption of preferential orientations, affinity for support films and non-uniform distribution across the grid (Drulyte *et al.,* 2018). The aforementioned sample behaviour has been one of the main subjects of research in the field of cryo-EM development with a number of recent approaches introduced to the “toolbox” of sample preparation, including the development of fully-automated sample preparation technologies Chameleon (Darrow *et al.,* 2019) and VitroJet (Ravelli *et al.,* 2020).

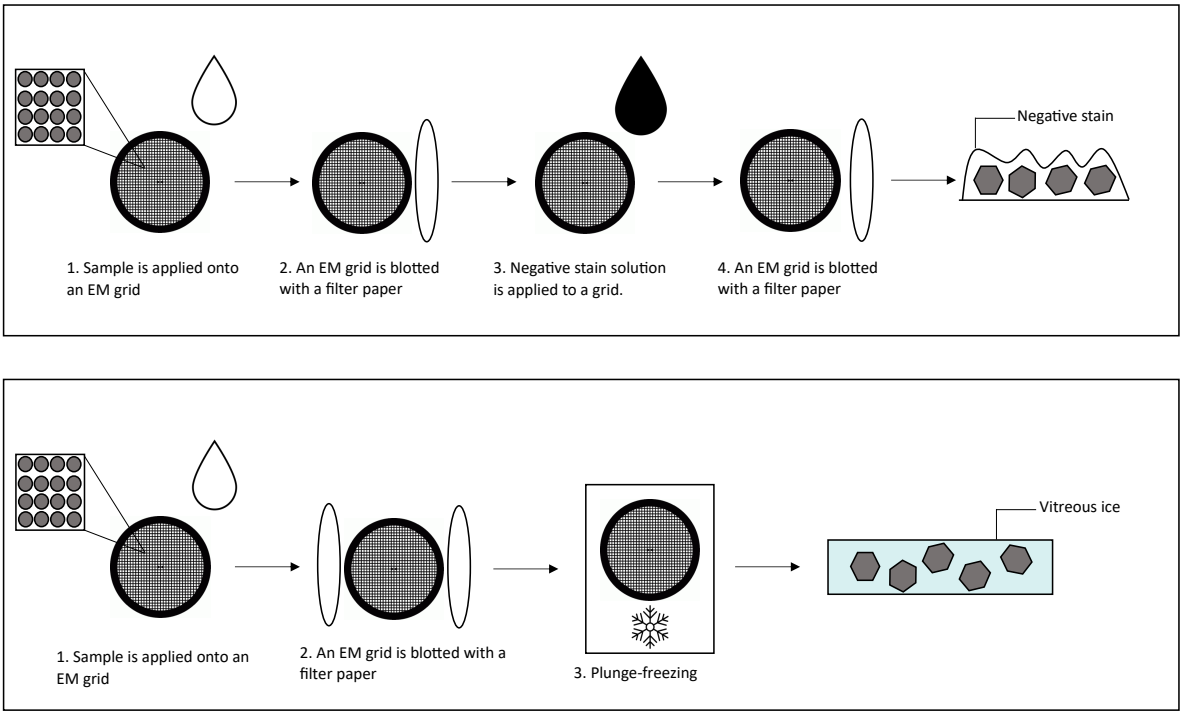


Figure 4.1. Standard sample preparation for single particle cryo-EM and negative stain analysis.

(Top) Sample preparation for negative stain analysis: (1) 5-10 µL of sample is applied onto the grid and incubated (typically 1min). (2) The excess sample is removed by blotting. (3) 10 µL of negative stain solution (typically 2-2.5% uranyl acetate) is applied onto the grid and incubated (typically 30 sec to 1 min). (4) Excess dye is removed by blotting and left to air dry. (Bottom) Sample preparation for cryo-EM: (1) 3 µL of the sample is applied to the grid in the humidity-controlled chamber. (2). The access sample is removed by blotting. (3) The grid is rapidly plunged into the cryogen, typically liquid ethane precooled with liquid nitrogen.

4.1.2. The microscope.

The invention of the electron microscope in 1931 by Max Knoll and Ernst Ruska made it possible to overcome the limitations of visible light. The wavelength of the electron is considerably smaller than that of photons (0.0197Å at 300kv), theoretically providing the potential to image cellular structures and proteins to an exceedingly high resolution. Nevertheless, imperfections and very poor performance of the electromagnetic lenses preclude these theoretical resolution limits from being achieved.

A typical electron microscope consists of three primary components: an electron source, a series of electromagnetic lenses and an electron detector. A simplified representation of electron microscope anatomy is shown in Figure 4.2. The electron microscope column operates under an extremely high vacuum to prevent electrons from interacting with air, allowing the beam to be more coherent and significantly reducing noise in the images (Williams and Carter, 2009). The following sections provide a more detail description of the major components of an electron microscope.

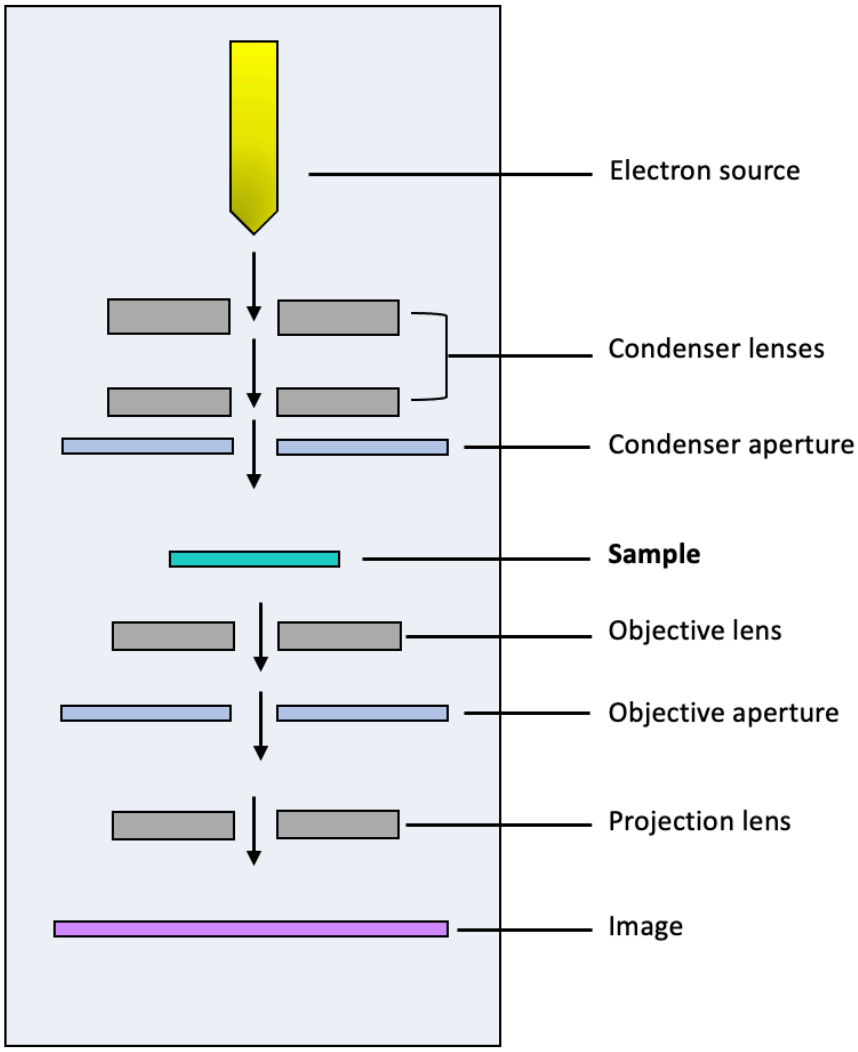


Figure 4.2. Simplified representation of transmission electron microscope

Arrows indicate the direction of the electron beam.

4.1.2.1. Electron sources.

The main electron sources used in electron microscopes include tungsten filament, LaB_6 crystal and field emission gun (FEG). Tungsten filament and LaB_6 crystals are thermionic sources and hence are heated to allow electrons to overcome an energy barrier and be emitted from their natural (bound to atoms) state into the vacuum. The FEG is a field-emission source and emits electrons from a cathode tip by applying voltage.
At present, the FEG is the most widely used electron source in advanced electron microscopes as it produces a beam that is much more coherent compared to both the tungsten filament and LaB_6 crystal. Electrons are emitted from a smaller tip radius (10-25 nm, compared to 50 μm for tungsten filament), significantly improving image quality (Williams and Carter, 2009).

4.1.2.2. Electromagnetic lenses

Once electrons are emitted from the electron source they are accelerated and directed through a series of electromagnetic lenses subsequently interacting with the specimen and hitting the image detection system. Magnetic lenses in the electron microscope consist of copper wires inside an iron shield. When the current is run through the copper wires, it creates an electromagnetic field.

An electron microscope is composed of three sets of lenses: the condenser lens, the objective lens, and the projection lens. The condenser lens system focuses the electron beam onto the sample, while the condenser lens aperture removes electrons with a high spread angle, thereby increasing the coherence of the electron beam. The objective lens forms the image from the electrons that are scattered from the sample. The objective lens aperture enhances image contrast by removing electrons scattered at high angles. The

projection lens is responsible for further magnifying the image before the electrons hit the image detector system (Reimer, 2013). Unfortunately, electromagnetic lenses suffer from poor performance, leading to various aberrations, including spherical aberration, chromatic aberration, and axial astigmatism.

Spherical aberration primary affects the objective lens. It occurs due to the non-uniform deflection of electrons by different regions of the lens, leading to a lack of focus and resulting in blurriness in the image (Williams and Carter, 2009). Spherical aberration can be corrected using specific aberration correctors (Haider *et al.,* 1998).

Chromatic aberration is caused by rays of different energy being focused at different distances from the focal point. This can be corrected by utilising magnetic field correctors, which control the path of the electrons, allowing them to be focused on the same focal point despite their different energies.

Axial astigmatism occurs when rays in two perpendicular planes are not focused equally, resulting in images having an “eclipse” shape. Fortunately, this can be compensated for by using special astigmatism correctors known as stigmator coils (Williams and Carter, 2009).

It is important to note that the aberrations mentioned here affect the final image the most, however, other, higher-level aberrations, also exist.

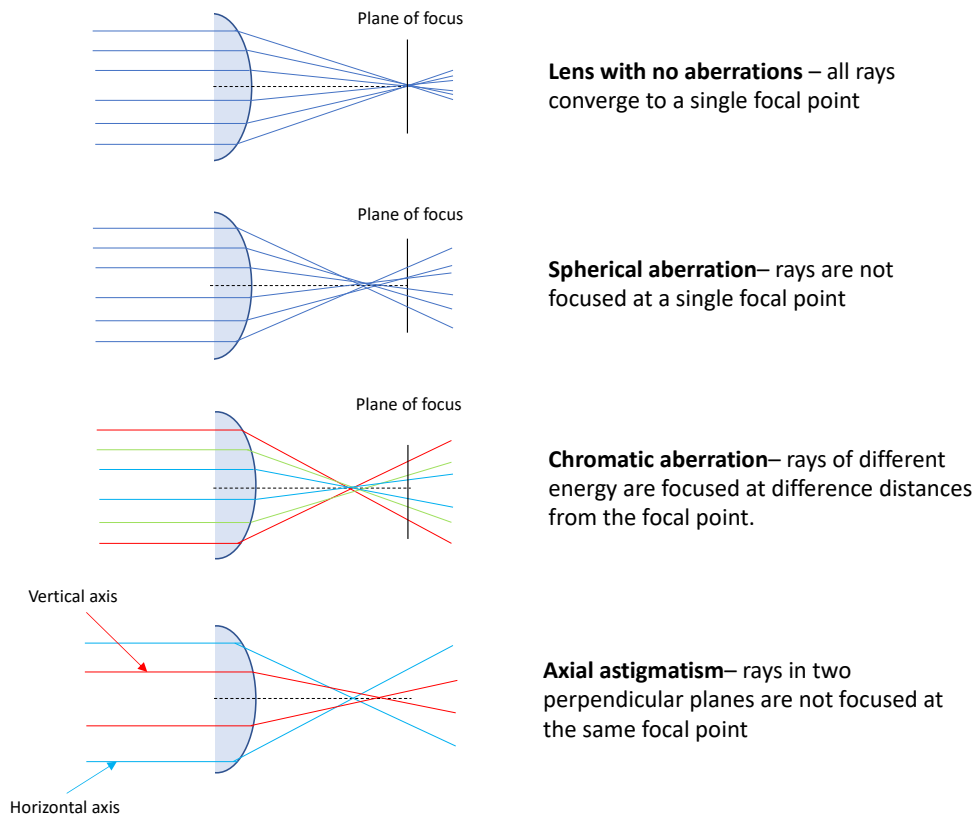


Figure 4.3. Schematic representation of TEM lens aberrations.

The blue-half circle corresponds to the lens and the blue/coloured lines correspond to electron rays. Adapted from Orlova and Saibil (2011)

4.1.2.3. Image detection system

One of the breakthrough developments in the cryo-EM field, and a major contribution of the "resolution revolution", is the development of direct electron detectors, which opened the possibility of electron counting, reducing the signal-to-noise ratio and allowing high-resolution particle reconstitutions (Wu *et al.,* 2016).

In addition, due to incredibly fast detection capabilities, direct electron detectors record movies that are broken down to individual frames. This is beneficial because during electron exposure, the sample experiences mechanical drift and radiation damage. Cryo-EM processing software allows alignment of the images as well as their dose-weighting to compensate the drift and radiation damaged in later frames (further discussed in "Motion Correction"). Routinely used direct electron detectors include the Direct Electron DE-20, FEI Falcon III and IV, Gatan K2, and K3 Summit.

4.1.3. Basic principles of image formation in TEM

An electron is a low-mass, negatively charged particle exhibiting both particle and wave characteristics (De Broglie, 1923). Electron scattering events occur due to electrostatic (Coulomb) interactions when electrons pass in close proximity to other electrons or positively charged atomic nuclei.

Once the electron passes through the sample, three main outcomes occur: (1) unscattered primary electrons – the incident electron is not scattered and passes directly through the magnetic lenses to the image plane (2) inelastically scattered electrons – the incident electron interacts with electrons of the sample, resulting in a loss of energy and a consequent inability to interact with unscattered primary electrons. (3) elastically scattered electrons – the incident electron scatters without losing energy, usually to a higher angle (Williams and Carter, 2009).

The electrons resulting from the interactions within the sample are subsequently directed towards the objective lens, where they are focused and form an image at the image plane. It is the interference between unscattered primary electrons and elastically scattered electrons that produces image contrast (Williams and Carter, 2009). Inelastically scattered electrons are never in phase, usually scattered to a low angle, and have lower energy. Consequently, they can be removed by energy filters, as they tend to be focused above the image plane. It should be noted, that inelastically scattered electrons contribute to radiation damage in the sample, thereby introducing noise into the resulting image. Approximately 15% of electrons are inelastically scattered, while only 5% are elastically scattered. This implies that for every useful event, there are three damaging ones (Williams and Carter, 2009).

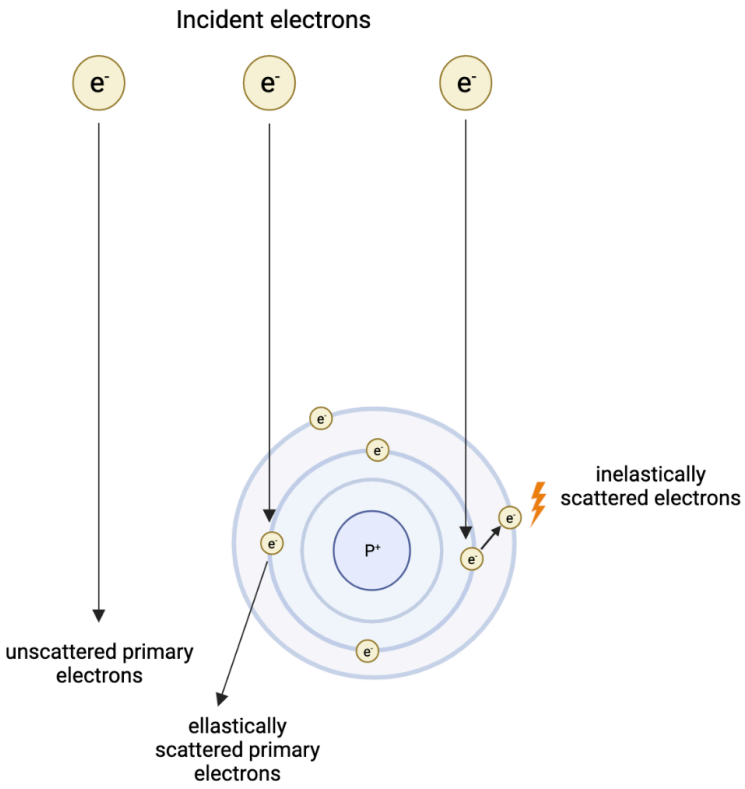


Figure 4.4. Basic representation of the main electron scattering events within a thin TEM specimen.

The direct beam represents the unscattered primary electrons. The event of elastic scattering occurs when electrons are scattered with no loss in energy to a low angle (1-10˚). The event of inelastic scattering occurs when electrons are scattered with a loss of energy usually to a very low angle (<1˚) (Williams and Carter, 2009). Adapted from Orlova and Saibil (2011).

4.1.4. Phase contrast and Contrast Transfer Function.

Proteins are weak phase objects as they are composed of atoms with low atomic numbers (carbon, hydrogen, oxygen, nitrogen, and sometimes sulphur), resulting in low electron scattering factors (z). Therefore, in contrast to negative stain EM, where the thin heavy metal coating of the specimen produces amplitude contrast, in cryo-EM, the most important imaging mode is phase contrast (Wade, 1992).

Roughly speaking, the contrast measured in cryo-EM images is linearly related to the scattering potential of the object. When electrons pass in close proximity to the atomic nucleus, they experience a phase shift of 90˚, resulting in elastically scattered electrons. Additionally, the scattering of each Fourier component at a range of angles introduces a difference in the path length between scattered and unscattered waves, leading to additional phase shifts and subsequent contrast in the image. This can be explained mathematically by the Contrast Transfer Function (CTF), which describes how much contrast is transferred to the image as a function of special frequency (Figure 4.5, C), with value of 1 indicating that 100% of contrast is transferred (Orlova and Saibil, 2011).

Nevertheless, some Fourier components will produce 0 contrast, resulting in a loss of information at these spatial frequencies. To compensated for this, defocus is introduced into the images. When imaging slightly underfocus, the path lengths of scattered waves change, leading to altered phase relationships. As a result, a shift occurs in the sinusoidal CTF function, causing certain spatial frequencies to become visible while others are lost. By collecting data at a range of defocus values, the zero values are covered, and the information of all spatial frequencies is retrieved within the dataset (Figure 4.5.) (Figure 4.6.) (Orlova and Saibil, 2011).

It is important to note that in reality, the CTF is strongly affected by various instabilities, such as incident beam incoherence. These effects are approximated by an envelope function, which describes the signal decay across spatial frequencies. In other words, the amount of signal transferred into the image decreases as the spatial frequency increases (Nakane *et al.,* 2020).

In practice, high defocus leads to more contrast at low spatial frequencies, making it easier to visualise and select particles in the micrographs. Nevertheless, this also leads to increased contrast reversals and the spreading of information (signal delocalisation). On the other hand, using low defocus values reduces the number of zero crossings but decreases the amplitudes (signal transferred) of high-frequency information. Therefore, it is important to collect data at different defocus values to be able to retrieve both high and low-frequency information (Figure 4.6).

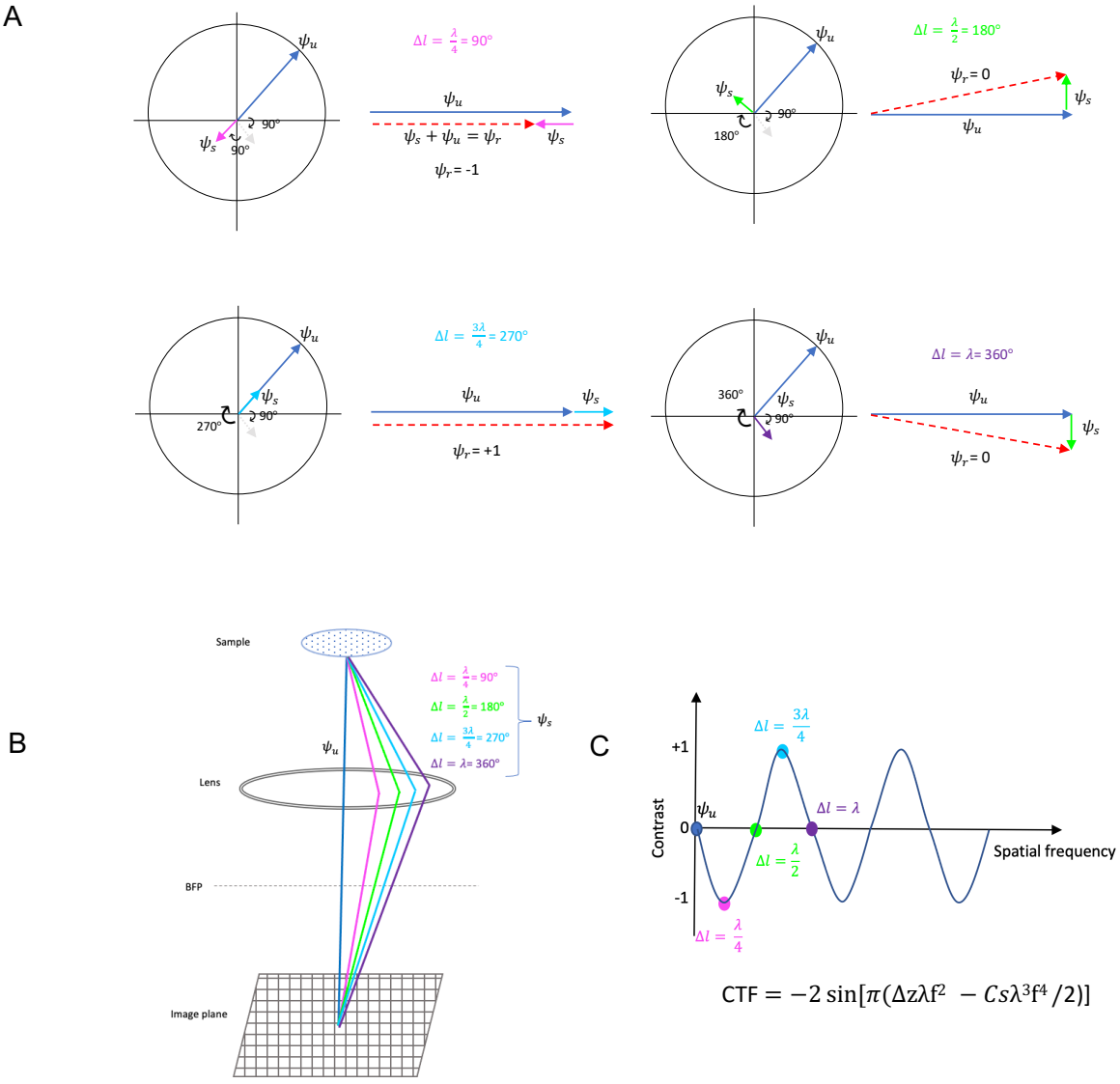


Figure 4.5. Phase contrast and Contrast Transfer Function in cryo-EM.

A) Argand diagrams representing phase shift relationships of scattered and unscattered waves. Four scenarios of scattering events are depicted in the figure: In the elastically scattering event, the electron wave experiences 90˚ shift (grey arrows), (ψ_u – unscattered wave; ψ_s – scattered wave; ψ_r – resulting wave; Δl – path length difference). In addition, each Fourier component is going to scatter differently varying in both scattering angle and the path length which subsequently introduces additional phase shifts. B) Simplified TEM diagram depicting the scattering events illustrated in A. C) CTF function: Δz is defocus, λe is the wavelength of the incident electrons, Cs is spherical aberration, and f is spatial frequency, ϕ - phase shift factor. Adapted from Orlova and Saibil (2011).

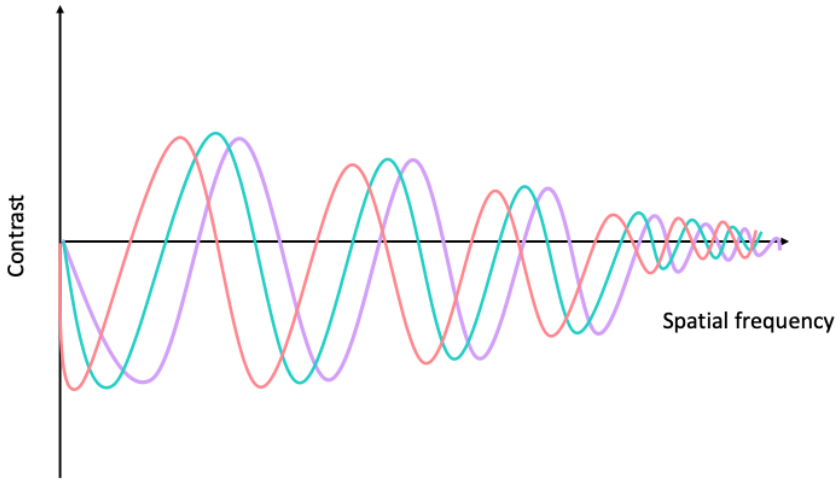


Figure 4.6. Defocus effect on the CTF function.

The CTF is an oscillating function with increasing special frequency and decreasing amplitude (contrast). At certain spatial frequencies, the CTF crosses 0 where no information is transferred. The red line represents the highest defocus, the green line represents an intermediate defocus, and the purple line represents the lowest defocus. Adapted from Orlova and Saibil (2011).

4.1.5. Data collection and image processing.

4.1.5.1. Motion correction

Once the electron wave hits the sample, drift occurs. The development of direct electron detectors has enabled the collection of movies, allowing for the correction of beam-induced motion. Several motion correction algorithms have been developed in recent years (Zheng *et al.,* 2017; Střelák *et al.,* 2021), however, the overall framework is the same across all motion correction techniques. The raw data is acquired as frames over time, and once one or more motion trajectories are identified, they are applied to the frames to create an aligned average (Personal communication with Dr Ali Punjani, September 2020). Micrographs are generally split up into 5x5 patches, as motion is not consistent throughout the entire micrograph. In addition, during motion correction, dose-weighting can be applied. Essentially, low resolution details from each aligned image are encoded into all the collected frames equally, however, due to electron exposure, the high-resolution details are degraded in the later frames, which introduces a considerable amount of noise if not taken into account (Personal communication with Dr Ali Punjani, September 2020).

4.1.5.2. CTF estimation and correction.

Voltage and spherical aberration values are constant for a given microscope, however, the CTF is different for each micrograph due to variations in lens intensity, sample height and particle distribution within the z-axis. Essentially, CTF estimation in cryo-EM processing programmes such as CTFFIND4 (Rohou and Grigorieff, 2015) compares the theoretical CTF curves at different defocus values with experimental image power spectra. A CTF plot in cryo-EM software illustrates Thon rings, which are the Fourier transform of the real space image, showing the constructive and destructive interference through the diffraction of the electron beam (Figure 4.7, A). Once defocus values are estimated, contrast reversals of the CTF can be corrected by phase flipping (Figure 4.7, B) (Cong and Ludtke, 2010).

Essentially, certain spatial frequencies will carry negative CTF values due to phase shifts, and these must be corrected in order to interpret information of spatial frequencies beyond the first 0 crossing of the CTF. It is important to note that at high defocus values, contrast reversals increase, resulting in signal delocalisation (Egerton and Watanabe, 2022). This has to be accounted for when choosing an appropriate box size for particle extraction as a too-small mask diameter or box size can lead to a loss of information (Moriya *et al.,* 2020).

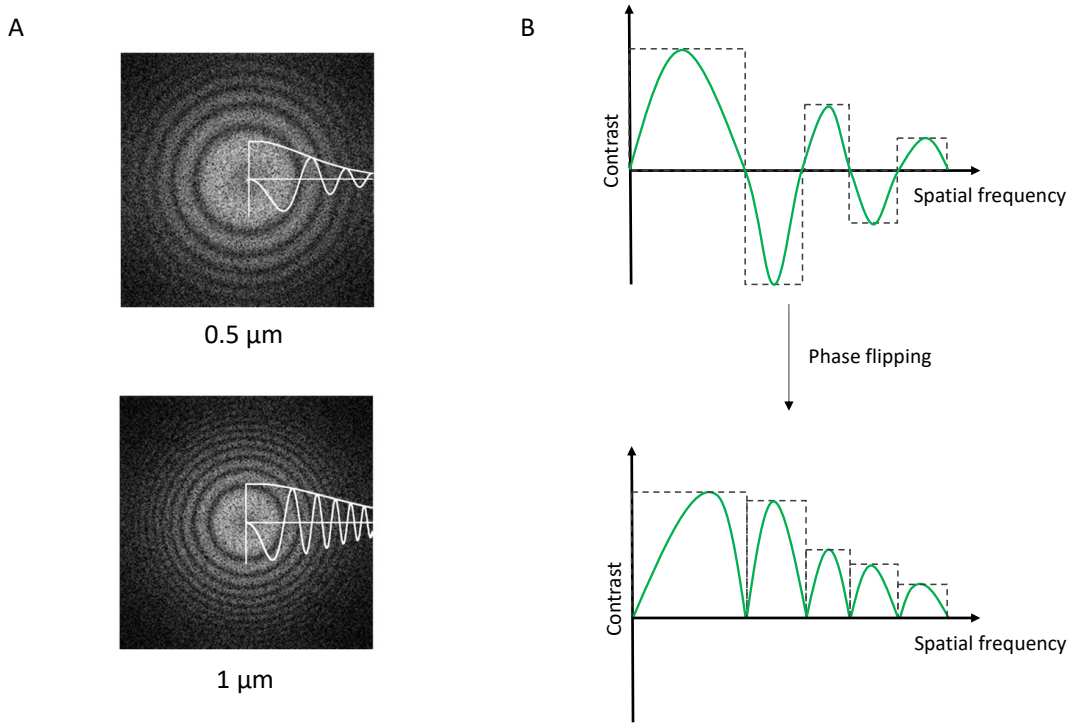


Figure 4.7. CTF Estimation and Phase Flipping.

A) Example of Thon rings with corresponding CTF curves at two different defocus values (0.5 µm and 1 µm). B) Example of phase flipping. Figure adapted from presentation at EMBO cryo-EM images processing course by Dr Dan Clare (2020).

4.1.5.3. Particle picking.

After motion correction and CTF estimation, the next step in the cryo-EM data processing pipeline involves the picking and extraction of individual particles from the micrographs. Manual particle picking can be a time-consuming process that requires significant user input. To address this, various automated and semi-automated particle picking approaches have been developed.

The most common approaches include template-free picking, based on the Laplacian-of-Gaussian (LoG) filter, and methods that utilise prior information, such as 2D class averages, and a cross-correlation function for reference-based picking (Scheres, 2015). However, the latter method introduces reference bias which could lead to false positive picks and the generation of images from noise, also known as “Einstein from noise” (Henderson, 2013; Shatsky *et al.,* 2009).

Recently, a new approach based on deep convolutional neural networks (CNN) has been introduced and employed to develop particle-picking algorithms such as DeepPicker (Wang *et al.,* 2016), FastParticlePicker (Xiao and Yang, 2017), Topaz (Bepler *et al.,* 2019) and crYOLO (Wagner *et al.,* 2019). AI-based particle picking does not require prior information which side-steps the reference-bias issue. Nevertheless, if the algorithm fails to generate satisfactory picks, a training dataset can be generated and used as input.

4.1.5.4. Particle alignment and classification

Once particles in a variety of orientations are selected and extracted from the micrographs, the signal represented must be aligned to average out the noise. However, the orientations of the 2D projections are initially unknown and need to be estimated for their proper alignment. In this regard, a Maximum Likelihood approach is typically utilised to generate the 2D class averages. The method calculates the probability of all the possible orientations of the experimental image relative to the reference image, assuming that noise obeys a gaussian distribution and is independent at each pixel. By calculating the weighted average of all the different orientations of each image, a reference image is obtained for the next iteration (S. Scheres, personal communication, September 2019).

2D classification is usually done within the maximum likelihood framework and the missing parameters are expanded from only missing orientations to a parameter referred to as a classification parameter (Sigworth *et al.,* 2010). The user specifies the number of classes (k), and initially, the dataset is randomly divided into k subsets. The algorithm is robust enough to use the random orientations as a reference for the next iteration, calculating the weighted average probability, which is taken to be the next set of references (S. Scheres, personal communication, September 2020). This procedure is typically repeated for 25 iterations enabling the formation of defined classes.

3D classification mathematically is a much more complex procedure, nevertheless, it is based on the same maximum likelihood framework (Sigworth et *al.,* 2010). In 3D reconstruction, the probability of obtaining a single 3D structure that would best explain the 2D images, given the microscope settings, is calculated. The “Fourier slice theorem” is introduced into the 3D reconstitution (Nogales and Scheres, 2015) (Figure 4.8). The Fourier slice theorem states that the FT of 2D projection equals a central slice of the FT of the 3D object in a direction perpendicular to the slice. The 2D projections are then used to fill in an entire angular space.

However, since the individual particles are present in random orientations, the projection vectors must be determined to insert the projection into the correct plane. This is typically achieved through projection-matching, where a model is projected and compared to the experimental images (Estrozi and Navaza, 2008). In 3D reconstitution, a reliable starting model is required, which introduces reference bias. The optimisation problem is tackled using a “hill climbing” approach, where one starts with an initial guess of an answer and at each iteration, the guess is improved. Nevertheless, it is important to note that the likelihood function of the correct 3D model versus the space of all possible 3D structures is not unimodal but contains local optima. Starting with an incorrect guess can result in iteratively optimising towards a local optima that may produce an incorrect structure that cannot be further improved (Reboul *et al.,* 2018). To address this, a stochastic hill climbing algorithm has been developed (Reboul *et al.,* 2016), which introduces random perturbations to the projection matching process, allowing for unbiased reconstruction from particles only.

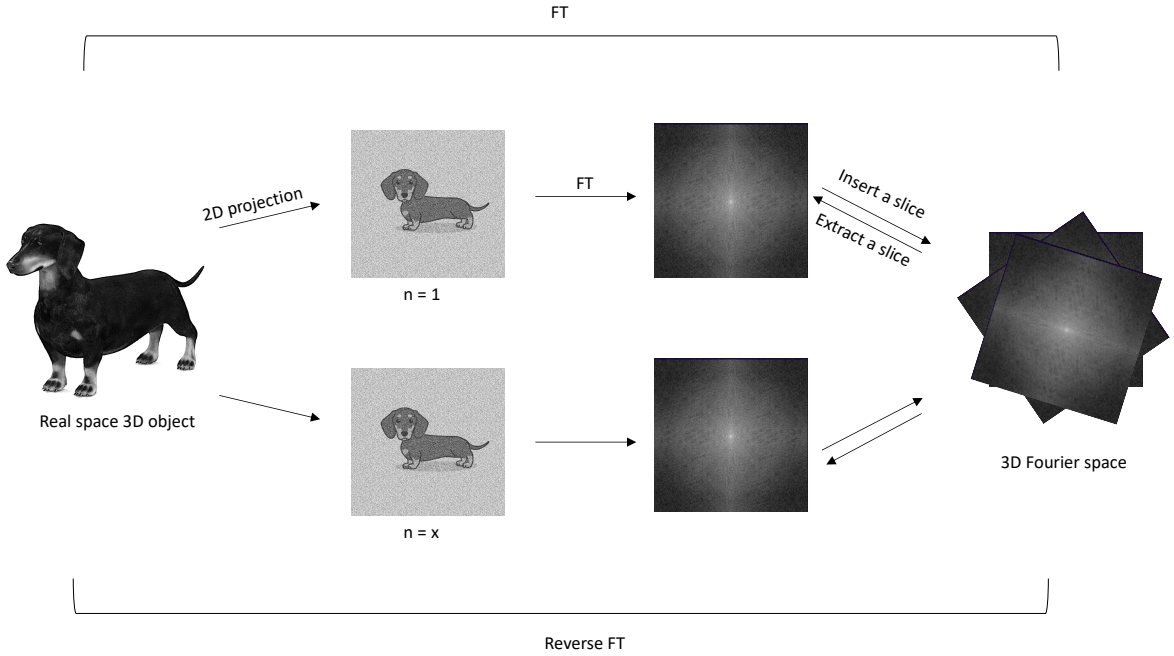


Figure 4.8. The Fourier Slice Theorem.

The Fourier slice theorem states that the Fourier transform (FT) of a 2D projection of an object is equal to a central slice of the Fourier transform of the 3D object along a direction perpendicular to the slice. Figure adapted from Wang *et al.,* (2013).

4.1.5.5. 3D refinement and post-processing.

3D refinement is an iterative process in which the initial 3D model is forward projected to generate 2D templates. These templates are then used to calculate projection vectors for all the particles. The calculated projection vectors are then employed in the 3D reconstruction, resulting in an improved model. This model serves as the new starting point for the next iteration of the refinement process (Sigworth, 2016).

During 3D refinement, the data is randomly split into two halves and refined separately. This allows the calculation of the cross-correlation between the two independent 3D volumes, known as the Fourier Shell Correlation (FSC). The FSC, plotted as a function of spatial frequency, originates at 1 and falls to 0 corresponding to the signal-noise ratio which decreases with increasing resolution (Rosenthal and Henderson, 2003).

The resolution of the structure is most often reported at a 0.143 FSC cut-off value, known as the gold standard FSC. This value is considered to be comparable to measurements used in X-ray crystallography (Harauz and Heel, 1986; Rosenthal and Henderson, 2003). However, if the noise is consistent between the split datasets, a higher resolution might be falsely reported due to partial alignment of noise between the two half-sets. This can result in overfitting and an overestimation of the global resolution (Sousa and Grigorieff, 2007).

There are several other limitations associated with 3D refinement algorithms. One common assumption is that proteins are rigid, which can negatively affect the reconstitution of proteins with disordered groups and result in lower resolution density of flexible regions. In recent years, several algorithms have been introduced to address this issue, such as SIDESPLITTER (Ramlaul *et al.,* 2020), *non-uniform* refinement (Punjani *et al.,* 2020) and 3D flexible alignment (Punjani and Fleet, 2021). These algorithms show promising results in

reconstituting the 3D structures of membrane proteins whilst considering the presence of disorder across the volume map. This will be further discussed in Chapter 6.

4.1.5.6. Model building

Model building is the final step in the cryo-EM data processing pipeline. Once a 3D map is obtained, the model can be fit into the density or built from scratch, depending on the map resolution and availability of existing structures. If the structure is already known, rigid body fitting can be performed obtaining the score for the "goodness of the fit". For high-resolution maps, where side chains are visible, the model can be built into the volume density and refined using software commonly used in X-ray crystallography, such as COOT (Emsley *et al.,* 2010) and Phenix (Torices and Muñoz-Pajares, 2015). It is important to note several automated model-building programmes have recently emerged, such as DeepTracer (Pfab *et al.,* 2021), which can determine a structure based on a cryo-EM map and protein sequence. Recently, a novel algorithm termed ModelAngelo has been introduced (Jamali *et al.,* 2023), demonstrating astonishing results in the model building of cryoEM maps of 3.5 Å and beyond.

4.2. Chapter Aims

Cryo-EM has been previously employed to probe the structures of several MCE proteins including *E.coli* MlaD, PqiB, LetB and *M. smegmatis* MCE1 (Chen *et al.,* 2023; Ekiert *et al.,* 2017; Isom *et al.,* 2020). Therefore, we decided to utilise cryo-EM to investigate the structure of the *B. bacteriovorus* Bd1153-Bd1152 membrane-associated complex. In this chapter, we present the diverse experimental and data processing strategies that were implemented to achieve a high-resolution structure of the Bd1152 top MCE domain.

4.3. Results

4.3.1. Structural investigation of membrane associated Bd1153-Bd1152 protein complex.

4.3.1.1. Expression and purification of the Bd1153-Bd1152 membrane associated complex.

The Bd1153-Bd1152 membrane-associated complex was expressed from an IPTG-inducible pET26b(+) plasmid. The membrane fraction was isolated and solubilised in 1% DDM. Bd1153-Bd1152 was then purified though Immobilised metal affinity chromatography (IMAC). The IMAC flow-through, wash and elution fractions were analysed via SDS-PAGE (Figure 4.9, A), which revealed two prominent bands at ~ 31 kDa and ~18 kDa, corresponding to Bd1152 and Bd1153, respectively. Elution fractions containing Bd1153 and Bd1152 were combined and further purified with a Superose 6 10/300 Increase size exclusion column, resulting in a single major peak (Figure 4.9, B). The resulting elution fractions were analysed via SDS-PAGE, which revealed bands at ~31 kDa and ~18 kDa, corresponding to Bd1152 and Bd1153, respectively. An anti-his western blot confirmed the identity of the major band at ~31 kDa as Bd1152 (Figure 4.9, D).

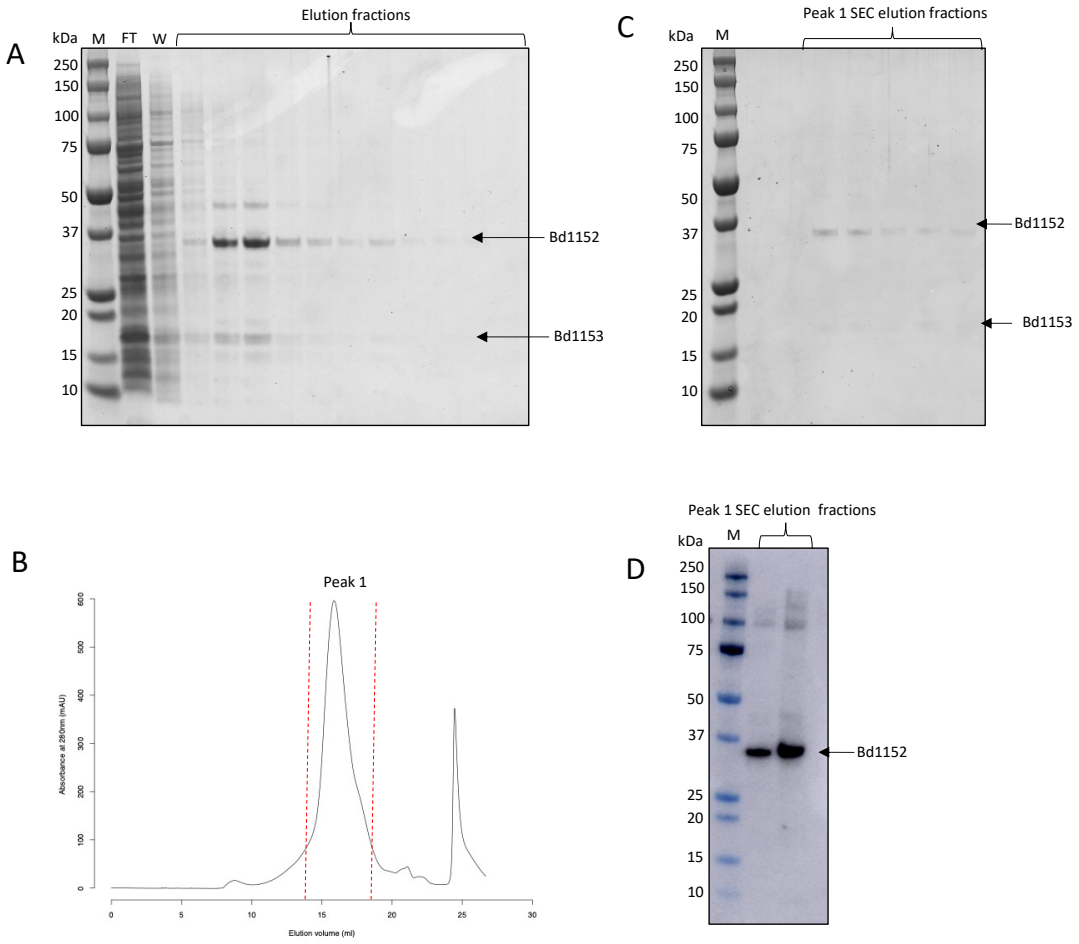


Figure 4.9. Purification of the Bd1153-Bd1152 membrane-associated complex.

A) SDS-PAGE of the Bd1153-Bd1152 IMAC purification displayed predominant bands at ~31 kDa and ~18 kDa, corresponding to Bd1152 and Bd1153, respectively. M – molecular weight marker. FT – flow-through, W – 50 mM imidazole wash. (B) SEC (Superose 6 10/300) revealed a distinctive single peak eluting between ~14 and 18 mL. (C) SDS-PAGE of the Bd1153-Bd1152 SEC elution fractions exhibited two bands at ~31 kDa and ~18 kDa, corresponding to Bd1152 and Bd1153, respectively. (D) An anti-his western blot revealed a prominent band at ~31 kDa, corresponding to the 6x his-tagged Bd1152 protein. SDS-PAGE was performed using 4-12% bis-tris gels.

4.3.1.2. Cryo-EM data collection and processing.

A total of 12090 micrographs were collected using the FEI EPU automated data collection software for the Bd1153-Bd1152 sample. The data collection parameters are provided in Table 2.10. All subsequent processing steps, unless specified otherwise, were performed using Relion 3.1.3. (Scheres, 2012). Motion correction and CTF estimation were carried out using MotionCorr2 (Zheng et al., 2017) and CTFFIND4 (Rohou and Grigorieff, 2015), respectively. Poorly corrected micrographs were removed following visual inspection of the power spectra, resulting in the discarding of 606 micrographs (Figure 4.10)

Particle picking was performed with Topaz (Bepler et al., 2019), followed by two rounds of reference-free 2D classification, resulting in a total of 424303 particles. Top, side, and intermediate views of Bd1153-52 were obtained and selected for initial model generation. To minimise bias, no external reference was provided, nor higher symmetry applied. The initial 3D model exhibited two approximately C6 symmetrical assemblies and the micelle. Following low pass filtering to prevent overfitting, this initial model was further utilised as the reference model for 3D classification (Figure 4.10).

Initially, four distinct models were generated via 3D classification. Among these models, only one exhibited the anticipated features, namely two MCE domain rings and a DDM micelle. However, two of the models, comprising 25% and 42% of the particles (as depicted in Figure 4.10), displayed a 6-fold symmetrical top view and a distorted side view. This suggested that 2D top view projections were assigned to these two classes. Consequently, only 28% of the original particle set (118804 particles) were deemed suitable for subsequent processing.

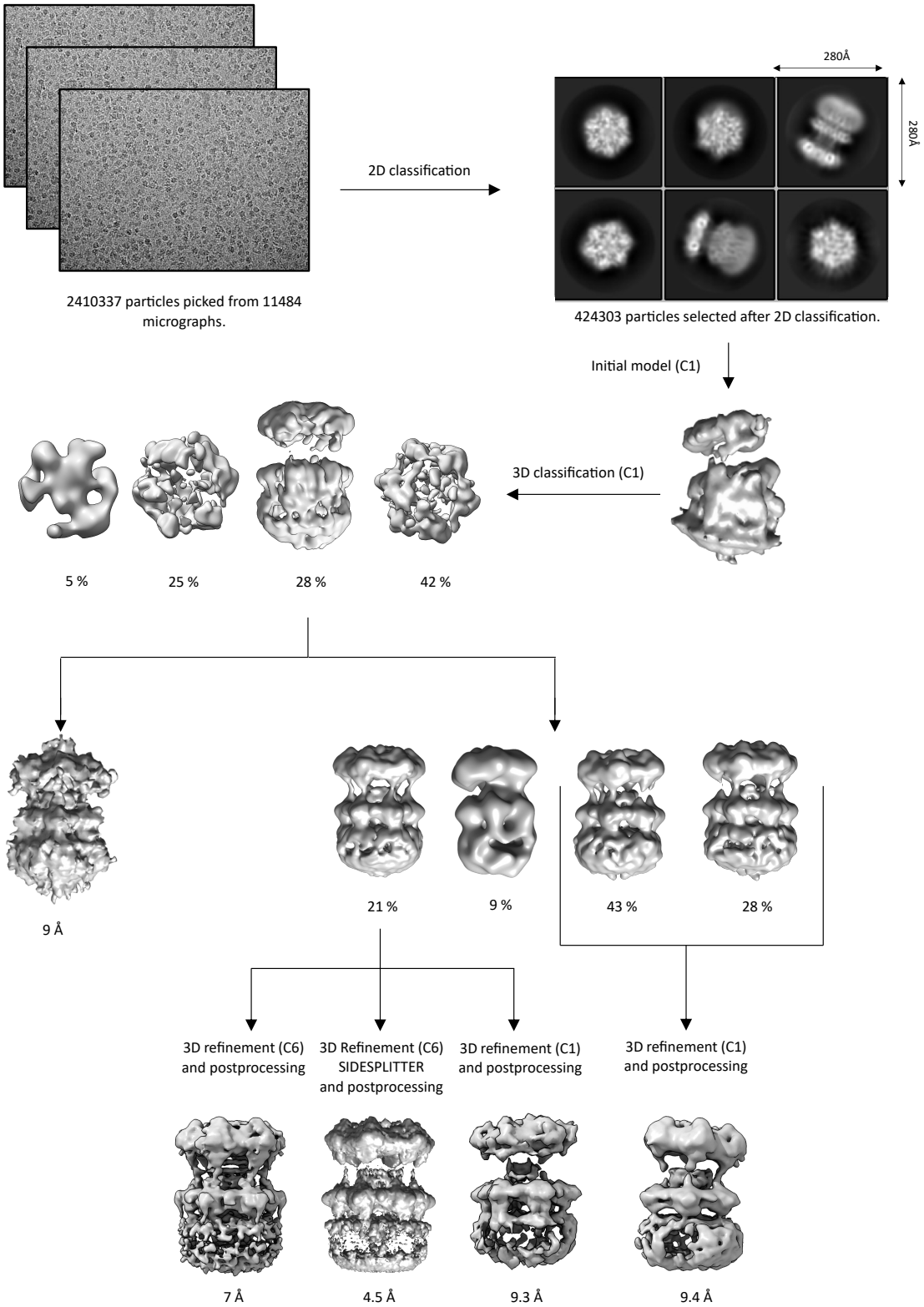

2D classification
280Å
280Å
2410337 particles picked from 11484 micrographs.
424303 particles selected after 2D classification.
Initial model (C1)
3D classification (C1)
5 %
25 %
28 %
42 %
9 Å
21 %
9 %
43 %
28 %
3D refinement (C6) and postprocessing
3D Refinement (C6) SIDESPLITTER and postprocessing
3D refinement (C1) and postprocessing
3D refinement (C1) and postprocessing
7 Å
4.5 Å
9.3 Å
9.4 Å

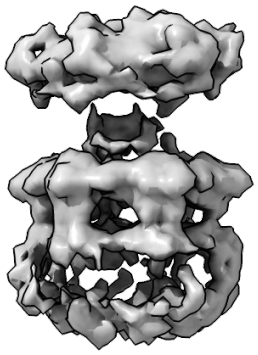
Figure 4.10. Simplified workflow of the Bd1153-Bd1152 cryo-EM data processing. Simplified workflow of cryo-EM data processing for the Bd1153-Bd1152 sample with Relion 3.1.3. Particle picking was performed with Topaz, followed by 2D classification, resulting in a total of 424303 particles that were used for initial model generation. 3D classification (C1) (k=4) produced a single class displaying expected structural features, which was selected for further processing. 3D refinement (C1) and post-processing yielded a 9 Å map. In parallel, 3D classification was performed using a reference map with imposed C6 symmetry (relion_image_handler) (C1) (k=4). The resulting classes were split into two subsets representing the “flat” and “tilted” conformations, which were refined separately, resulting in five final maps. The density maps were visualised with ChimeraX (Pettersen *et al.,* 2021).

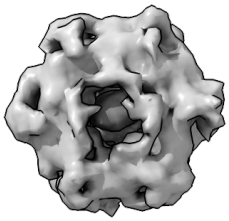
Initially, our objective was to obtain a C1 map to minimise bias. To accomplish this, several rounds of 3D classification in C1 symmetry were performed prior to 3D refinement (C1) and post-processing. This resulted in a 9.3 Å map that exhibited apparent 6-fold symmetry, two MCE domain rings, and a DDM micelle. Despite performing extensive 3D classification and local refinement, utilising 3D masking and signal subtraction, a map of higher resolution was not obtained (Figure 4.10, Figure 4.11).

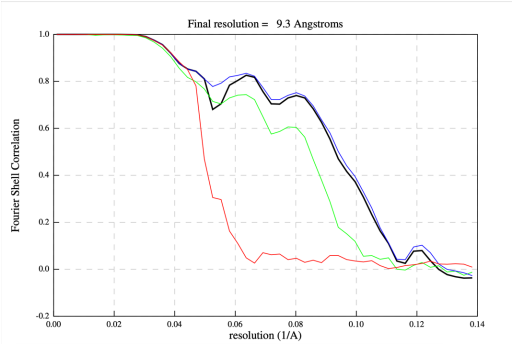
We next attempted to increase the resolution of the map by imposing 6-fold symmetry. Initially, a 6-fold symmetrical map was used as a reference model for 3D classification in C1 symmetry. Four classes were generated revealing the flexibility of the top MCE domain. At this point, the data was split into two subsets: particles belonging to the “flat” (24949 particles) and “tilted” (84350 particles) conformations. For the “tilted” dataset, a 6-fold symmetrical map was used in 3D refinement with C1 symmetry, leading to a resolution of 9.4 Å and a clear lifting of the top MCE ring (Figure 4.10, Figure 4.11, B). For the “flat” dataset, a 6-fold symmetrical map was used in 3D refinement with C6 symmetry, resulting in a 7 Å map (Figure 4.10, Figure 4.11, D).

However, over-fitting is visible with a number of density artefacts resulting from noise-reinforcement in the 3D map. To address this issue, the local signal-to-noise filter termed SIDESPLITTER (Ramlaul *et al.,* 2020) was applied during 3D refinement. This approach successfully reduced over-fitting and minimised the occurrence of symmetry artefacts. Nevertheless, the reported resolution of the final map, stated as 4.5 Å, was deemed over-optimistic and did not reflect the actual resolution attained. The final cryo-EM maps and their corresponding FSC plots are displayed in Figure 4.11.

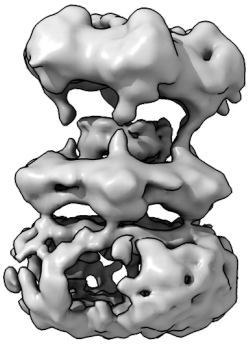
A

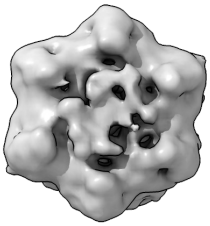


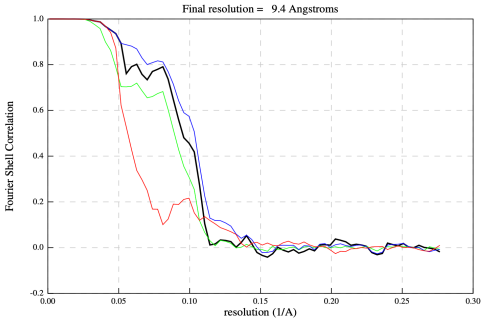



Final resolution = 9.3 Angstroms
Fourier Shell Correlation
resolution (1/A)

B

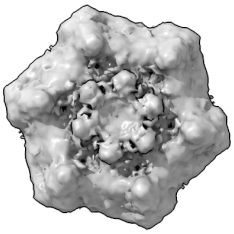


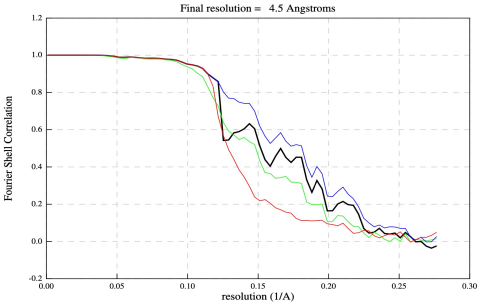



Final resolution = 9.4 Angstroms
Fourier Shell Correlation
resolution (1/A)

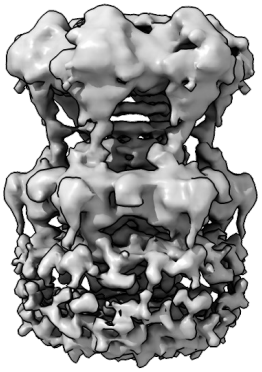
C

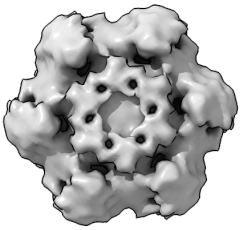





Final resolution = 4.5 Angstroms
Fourier Shell Correlation
resolution (1/A)

D





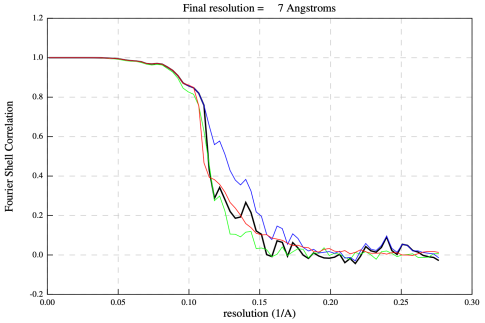

Final resolution = 7 Angstroms
Fourier Shell Correlation
resolution (1/A)

Figure 4.11. The final cryo-EM maps obtained from the Bd1153-Bd1152 dataset.

A) Cryo-EM map from the particles assigned to the “flat” conformation (C1 symmetry). B) cryo-EM map from the particles assigned to the “tilted” conformation (C1 symmetry). C) Cryo-EM map obtained by performing 3D refinement in C6 symmetry supplemented with the Sidesplitter external command. D) Cryo-EM map obtained by performing 3D refinement in C6 symmetry. The FSC plot illustrates the correlation between the two independent refined maps (half-maps) with blue (no mask), loose mask (red), tight mask (cyan) and corrected (purple) curves. A cut-off 0.143 (blue vertical line) was used for resolution estimation. The density maps were visualised with ChimeraX (Pettersen *et al.,* 2021).

Various methods were employed in an attempt to improve the maps, including extensive 3D classification with varying T values and number of classes, local refinement using 3D masking and signal subtraction, symmetry expansion and symmetry relaxation. However, none of these attempts yielded significant improvements (Data not shown).

It is important to note that while Bd1153 was present in the sample throughout the purification (Figure 4.9.), 2D classification and the final maps suggested that the Bd1153-Bd1152 protein complex had dissociated, and only Bd1152 can be observed in the class averages. Additionally, the bottom MCE ring, although low resolution, displayed density resembling the MCE-domain fold, whilst the top MCE ring exhibited overfitting and symmetry artifacts, likely resulting from imposing C6 symmetry (Figure 4.12).

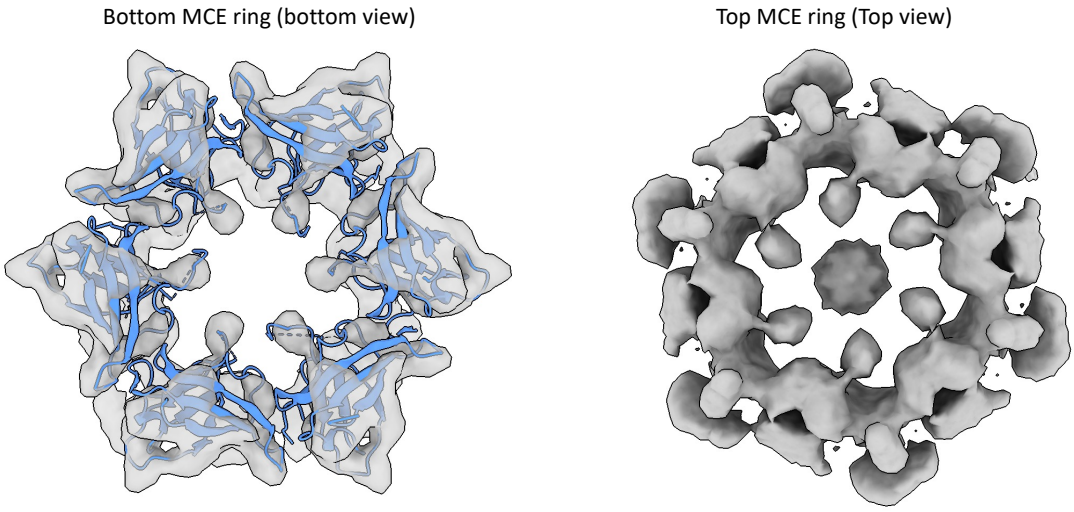


Figure 4.12. Comparison of the Bd1152 top and bottom MCE rings obtained from the Bd1153-Bd1152 cryo-EM dataset.

The blue cartoon represents the Bd1152 Alphafold model of the bottom MCE ring (residues 37-107; 119-157) fitted into the cryo-EM density of the bottom MCE ring. Both cryo-EM maps are displayed at the contour level of 0.03. The density maps and the model were visualised with ChimeraX (Pettersen *et al.,* 2021).

Data was reprocessed with cryoSPARC v.2.12.0 (Punjani *et al.,* 2017), with the assistance of Prof Julian Bergeron and Dr Soi Bui (Kings College, London). Motion correction was performed on all micrographs using cryoSPARC implemented Patch-based motion correction, followed by CTF estimation utilising patch-based CTF. Particle picking was performed using blob picker followed by 2D reference-based picking. A total of 533602 particles were used as an input for *ab initio* model generation in C1 symmetry.

Similar to Relion 3.1.3, an alignment issue was encountered, where side and top views were classified into separate classes (Figure 4.10, Figure 4.13). While model 2 exhibited rough 6-fold symmetrical density, model 3 displayed 6-fold symmetry, two MCE domain rings and a DDM micelle. The latter model was used as a reference for *non-uniform* refinement with C1 and C6 symmetry, resulting in final 6.8 Å and 5.5 Å resolution maps, respectively. However, both C1 and C6 maps exhibited over-fitting.

It is important to note, that as in Relion 3.1.3, combining model 1 and model 2 during *non-uniform* refinement resulted in an incorrect model. Local refinement using 3D masking was tested, however, none of the attempts improved the final maps (data not shown). The processing workflow and the final maps are depicted in Figure 4.13.

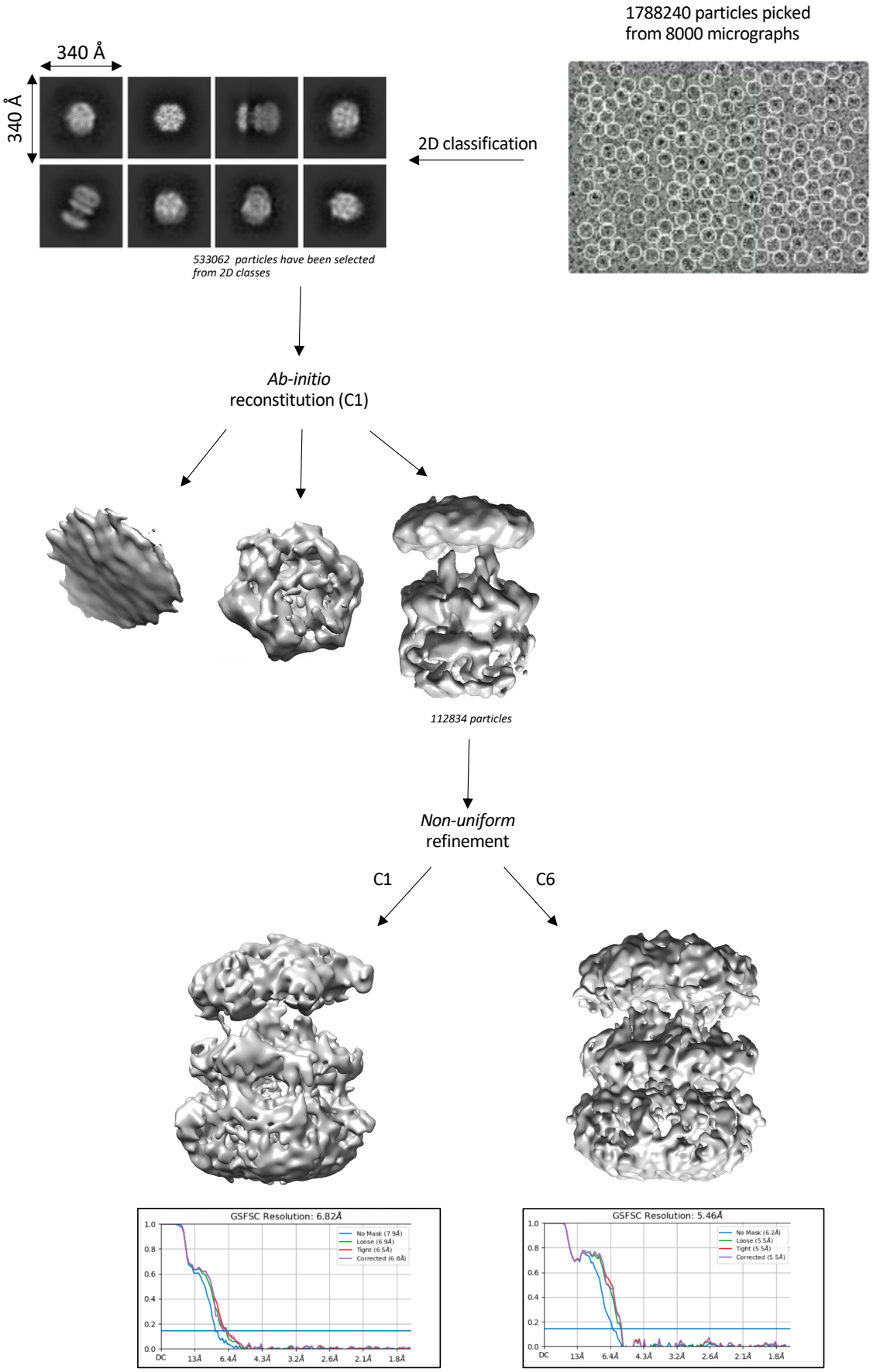

1788240 particles picked
from 8000 micrographs
340 Å
340 Å
2D classification
533062 particles have been selected
from 2D classes
Ab-initio
reconstitution (C1)
112834 particles
Non-uniform
refinement
C1
C6
GSFSC Resolution: 6.82Å
No Mask (7.9Å)
Loose (6.9Å)
Tight (6.5Å)
Corrected (6.8Å)
GSFSC Resolution: 5.46Å
No Mask (6.2Å)
Loose (5.5Å)
Tight (5.5Å)
Corrected (5.5Å)
DC
13Å
6.4Å
4.3Å
3.2Å
2.6Å
2.1Å
1.8Å

Figure 4.13. Simplified workflow of the Bd1153-Bd1152 cryo-EM data processing using cryoSPARC v.2.12.0.

Particle picking was performed using blob picker, followed by 2D reference-based picking. A total of 533602 particles were used as input for *ab initio* model generation in C1 symmetry. A single model displaying the expected structural features was selected for *non-uniform* refinement in both C1 and C6 symmetries. The average resolution derived from Fourier Shell Correlation (FSC) is depicted at the bottom of the figure. The FSC plot illustrates the correlation between the two independent refined maps (half-maps) with blue (no mask), loose mask (red), tight mask (cyan) and corrected (purple) curves. A cut-off 0.143 (blue vertical line) was used for resolution estimation. The density maps were visualised with ChimeraX (Pettersen *et al.,* 2021).

4.3.2. Structural investigation of the soluble Bd1152$^{37-283}$ construct

4.3.2.1. Generation of the soluble Bd1152 construct.

Expression and purification of the Bd1153-Bd1152 complex proved to be challenging. Low yield, aggregation and inherent flexibility collectively contributed to the low quality of the structural data. Next, we decided to simplify the construct and focus only on the periplasmic portion of Bd1152. Consequently, a soluble Bd1152 construct (Bd1152$^{37-283}$) was devised, lacking the residues corresponding to the transmembrane helix that is proposed to anchor Bd1152 to the IM.

4.3.2.2. Expression and purification of the Bd1152$^{37-283}$ construct.

The Bd1152$^{37-283}$ construct was expressed from an IPTG inducible pET26b(+) plasmid. Bd1152$^{37-283}$ was isolated from clarified cell lysate through IMAC purification. The IMAC flow-through, wash and elution fractions were analysed via SDS-PAGE (Figure 4.14, A).
A single elution fraction containing Bd1152$^{37-283}$ was further purified with a Superdex 200 10/300 Increase size exclusion column, resulting in a single peak (Figure 4.14, B). The resulting elution fractions were analysed via SDS-PAGE, which revealed a prominent band at ~31 kDa, corresponding to Bd1152$^{37-283}$ (Figure 4.14, C). A higher molecular weight band (~75 kDa) was also observed, likely corresponding to the oligomerisation of Bd1152$^{37-283}$, a feature previously observed when purifying *E. coli* MCE domain proteins (Cooper, 2023; Wotherspoon, 2022). It should be noted that purified Bd1152$^{37-283}$ was relatively unstable as it would aggregate during spin concentration. TLC analysis indicated that Bd1152$^{37-283}$ co-purified with PG and PE (Figure 4.14, D, Appendix 3), similar to the *E.coli* MlaD, PqiB and LetB proteins (Ekiert e*t al.,* 2017; Isom *et al.,* 2020), suggesting its function as a PL transporter.

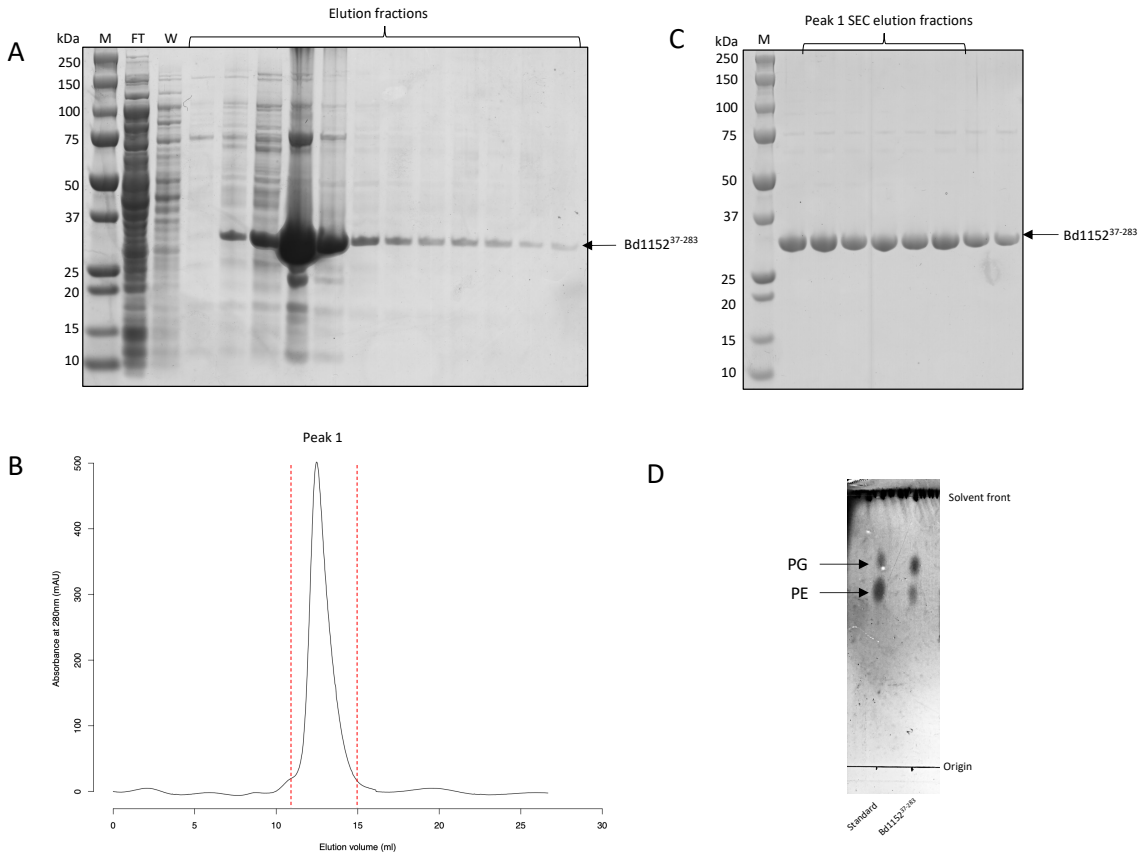


Figure 4.14. Purification of Bd1152$^{37-283}$.

A) SDS-PAGE of the Bd1152$^{37-283}$ IMAC purification showed a predominant band at ~31 kDa, corresponding to Bd1152$^{37-283}$. M - molecular marker, FT – flow-through, W – 50 mM imidazole wash. (B) The SEC (Superdex 200 10/300) trace of a single elution fraction exhibited a single peak eluting between ~11 and 15 mL. (C) SDS-PAGE of the Bd1152$^{37-283}$ SEC elution fractions revealed a prominent band at ~31 kDa, corresponding to Bd1152$^{37-283}$. D) Thin layer chromatography (TLC) of lipids extracted from Bd1152$^{37-283}$. PG – phosphatidylglycerol, PE – phosphatidylethanolamine. PLs are marked on the left according to a *E. coli* polar lipid standard (lane 1). SDS-PAGE was performed using 4-12% bis-tris gels.

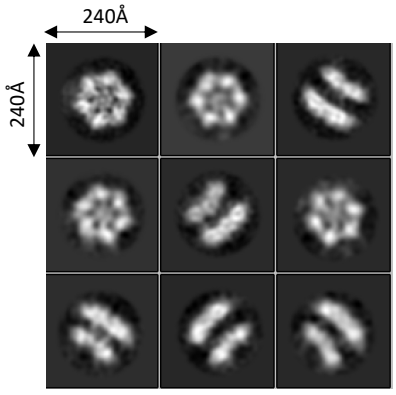
4.3.2.3. Cryo-EM data collection and processing of the Bd1152$^{37-283}$ construct.

A total of 533 micrographs were collected for the Bd1152$^{37-283}$ sample using an FEI EPU automated data collection system, with data collection parameters specified in Table 2.7. Unless stated otherwise, all processing steps were performed using Relion 3.1.3, and the initial processing steps, motion correction and CTF estimation, were carried out as previously described for the Bd1153-Bd1152 dataset.

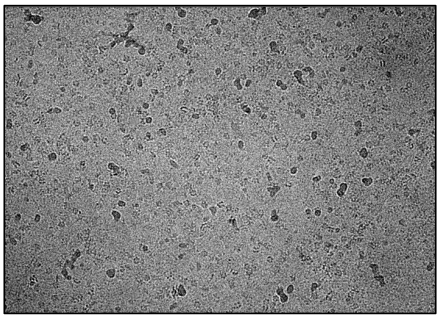
CrYOLO particle picking was employed, and 52583 particles were selected from 463 micrographs. Multiple rounds of 2D classification were performed to assess the data quality and remove “junk”. The 2D classification revealed blurred particles, suggesting the presence of flexibility within the dataset (Figure 4.15)

An initial model could not be obtained from the particles within the dataset. Therefore, the Bd1153-Bd1152 map (Figure 4.11, C), with a masked micelle and a 60 Å low-pass filter applied, was used as a reference model. Subsequent 3D classification in C1 symmetry did not yield a class that accurately represented the overall shape of the particle. However, upon conducting 3D classification in C6 symmetry, a class closely resembling the target particle was obtained and used for further processing (Figure 4.15).

Although 3D classification and 3D refinement (C6) resulted in a Bd1152$^{37-283}$ map that resembled the 2D classes, only an 11 Å resolution map was obtained (Figure 4.15), possibly due to insufficient number of particles and the inherent flexibility of the protein. The data processing workflow is depicted in Figure 4.15.

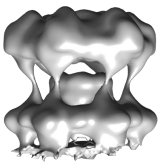


2D classification

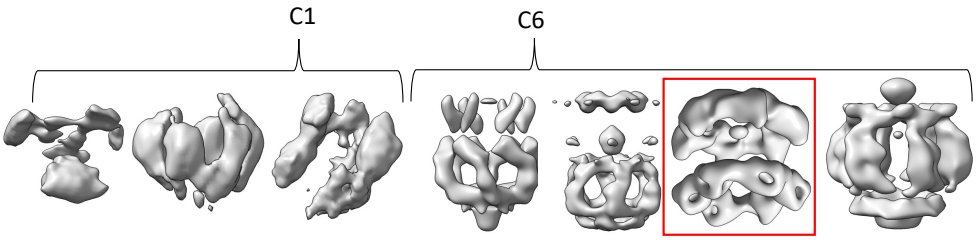


52583 particles picked from 463 micrographs

44698 particles selected after 2D classification.



The Bd1152 cryo-EM map with masked micelle was used as the initial model. Lowpass filter of 60 Å was applied.

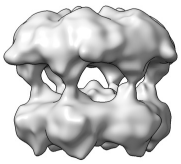


32342 particles

3D refinement (C6)



Post-process



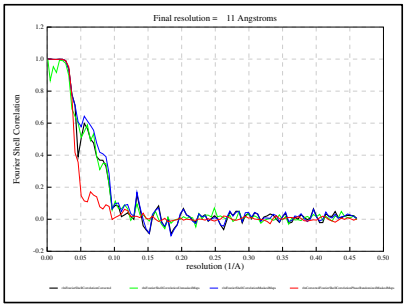


Figure 4.15. Simplified workflow of the Bd1152$^{37-283}$ cryo-EM data processing using Relion 3.1.3.

Particle picking was performed using CrYOLO, followed by several rounds of 2D classification. A 60 Å low-passed filtered map of Bd1153-Bd1152 with a masked micelle was used as the initial model for 3D classification in C6 symmetry. A model that resembled the overall Bd1152$^{37-283}$ architecture was selected for further processing. 3D refinement in C6 symmetry, followed by post-processing, resulted in an 11 Å map. The average resolution derived from Fourier Shell Correlation (FSC) is depicted at the bottom of the figure. FSC plot illustrates the correlation between the two independent refined maps (half-maps) with blue (no mask), loose mask (red), tight mask (cyan) and corrected (purple) curves. A cut-off 0.143 (blue vertical line) was used for resolution estimation. The density maps were visualised with ChimeraX (Pettersen *et al.,* 2021).

4.3.3. Structural investigation of the Bd1152$^{37\text{-}283\ I112C\ L238C}$ construct.

4.3.3.1. Generation of a more rigid Bd1152$^{37\text{-}283}$ for improved particle alignment in cryo-EM data processing.

The cryo-EM map of the *E. coli* 7-MCE domain-containing protein LetB (formerly known as YebT) revealed that local resolution fluctuated considerably between different regions of LetB indicating its heterogeneity across the density map (Isom *et al.,* 2020). Pivoting motions across the MCE rings in the LetB structure are comparable to the flexibility seen in both our Bd1152 and Bd1152$^{37\text{-}283}$ cryo-EM datasets. While investigating the flexibility landscape of a protein is of great interest, 2D and 3D classification of both the Bd1152 and Bd1152$^{37\text{-}283}$ samples suggest that high-angle tilting might prevent high-resolution reconstitution.

To address this, we utilised disulphide bond engineering with the aim of increasing protein rigidity and obtaining a more homogenous sample that would be more amenable for particle alignment and crystallisation. Specifically, we mutated hydrophobic residues Isoleucine 112 and Leucine 238, located in the pore lining loops, to cysteine residues with the intention of forming a disulphide bond between the 1st and 2nd MCE domains (Figure 4.16). It is noteworthy that the location of these mutations is based on the Bd1152 Alphafold model, and the regions where the I112 and L238 residues are located have low pLDDT scores (70 > pLDDT > 50).

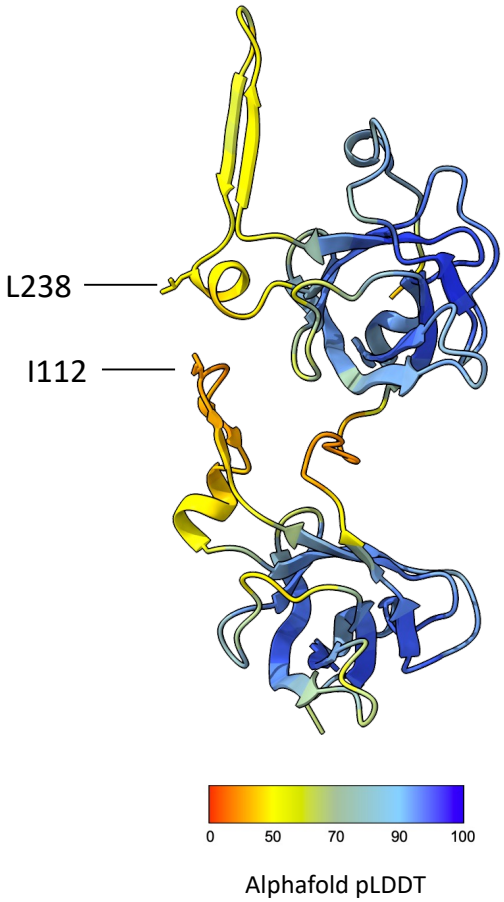


Figure 4.16. The Alphafold structure of Bd1152 (residues 37-283)

The model is coloured by per-residue pLDDT scores. A pLDDT colour key is provided at the bottom of the figure where the spectrum from blue to red corresponds to decreasing pLDDT score (Jumper *et al.,* 2021). The highlighted L238 and I112 residues were mutated to cysteine for cryo-EM investigations. The model was visualised with ChimeraX (Pettersen *et al.,* 2021).

4.3.3.2. Construct design, expression and purification of Bd1152$^{37\text{-}283\ I112C\ L238C}$.

I112C and L238C point mutations were introduced into the Bd1152$^{37\text{-}283}$ construct via utilisation of the Q5-site directed mutagenesis kit (NEB). The final Bd1152$^{37\text{-}283\ I112C\ L238C}$, hereafter referred to as Bd1152^{2C}, construct was expressed from an IPTG inducible pET26b(+) plasmid. Bd1152^{2C} was isolated from clarified cell lysate through IMAC purification. The IMAC flow-through, wash and elution fractions were analysed via SDS-PAGE (Figure 4.17, A), which revealed a prominent band at ~ 31 kDa, corresponding to Bd1152^{2C}. Elution fractions containing Bd1152^{2C} were combined and further purified with a Superdex 200 10/300 Increase size exclusion column, exhibiting a single peak (Figure 4.17, B). The resulting elution fractions were analysed via SDS-PAGE, which revealed a band at ~31 kDa, corresponding to Bd1152^{2C} (Figure 4.17, C). A higher molecular weight band (~75 kDa) was also observed likely corresponding to the oligomerisation of Bd1152^{2C}. The purification of the Bd1152^{2C} construct yielded a high-purity protein sample that exhibited much higher stability compared to the Bd1152$^{37\text{-}283}$ construct. This protein was much more amenable to spin concentration and therefore was further used for both cryo-EM and X-ray crystallography studies (Chapter 5). In addition, TLC analysis revealed that the Bd1152^{2C} protein did not co-purify with PG or PE (Figure 4.17, D, Appendix 3), potentially as a result of the I112C L238C mutations.

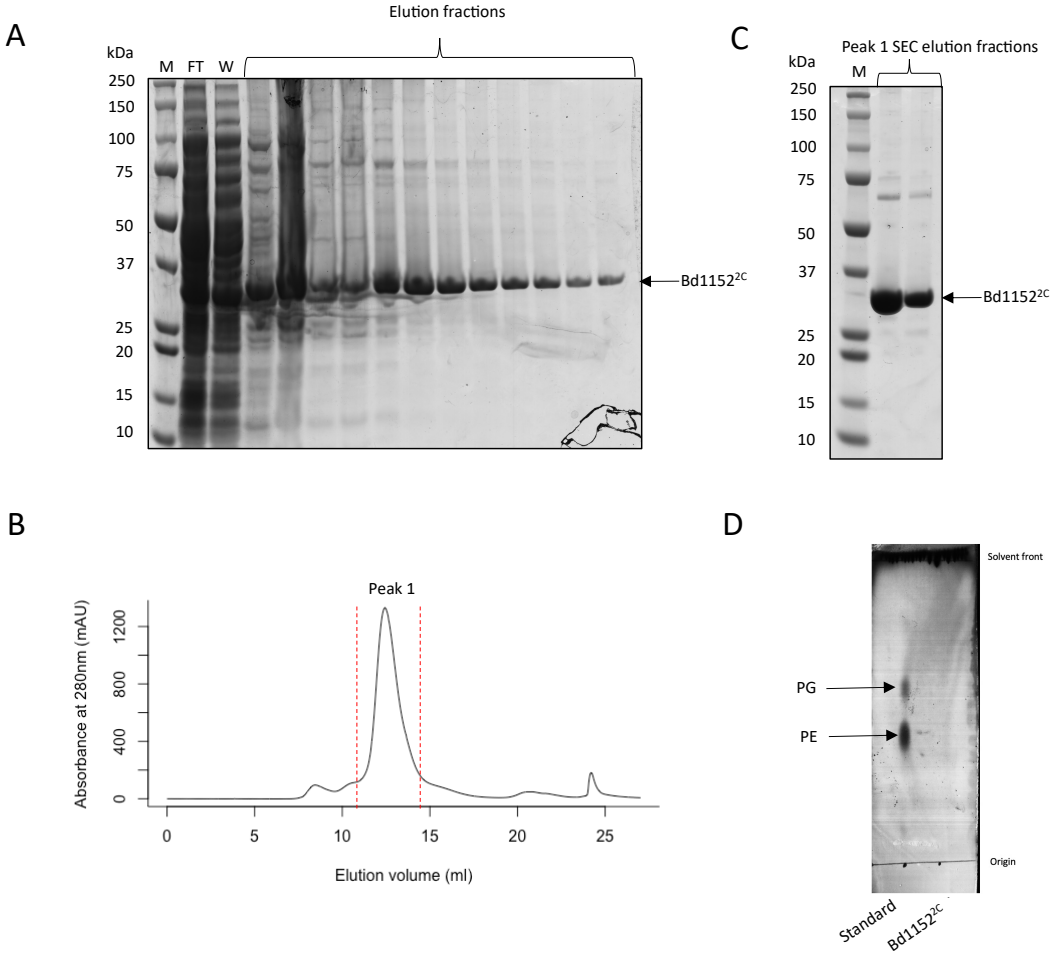


Figure 4.17. Purification of Bd1152^{2C}

A) SDS-PAGE analysis of the Bd1152^{2C} IMAC purification showed a prominent band at ~31 kDa. M - molecular marker, FT –flow-through, W – 50 mM imidazole wash. B) SEC (Superdex 200 10/300) revealed a single peak eluting between ~11 and 14 mL. (C) SDS-PAGE of the Bd1152^{2C} SEC elution fractions showed a prominent band at ~31 kDa, corresponding to Bd1152^{2C}. D) Thin layer chromatography (TLC) of lipids extracted from Bd1152^{2C}. PG – phosphatidylglycerol and PE – phosphatidylethanolamine. PLs are marked on the left according to a *E. coli* polar lipid standard (lane 1). SDS-PAGE was performed using 4-12% bis-tris gels.

4.3.3.3. Cryo-EM data collection and processing of Bd1152^{2C}

A total of 9667 micrographs were collected using the FEI EPU automated data collection software for the Bd1152^{2C} sample, with data collection parameters specified in Table 2.7. Unless stated otherwise, all processing steps were performed using Relion 4.0 (Kimanius *et al.,* 2021), and the initial processing steps, motion correction and CTF estimation, were carried out as previously described for the Bd1153-Bd1152 dataset.

Initially, several micrographs were manually picked to generate 2D classes, which were subsequently used for 2D-reference based picking. A total of 349668 particles were picked and subjected to multiple rounds of 2D classification. An initial model was generated in C1 symmetry using the gradient-driven algorithm implemented in Relion 4.0. A promising starting model, displaying C6 symmetry and two MCE domain rings, was obtained. This model served as the reference for 3D classification. In the first round of 3D classification, four classes were generated, revealing that the same alignment issue persisted in both the Bd1152 and Bd1152^{2C} cryo-EM datasets, as side and top views were classified into separate classes. Only class 2 was selected for further processing. Although the resulting map exhibited a significant improvement when compared to the Bd1152$^{37-283}$ map, the resolution was limited to 6.4 Å. A simplified data processing workflow is depicted in Figure 4.18.

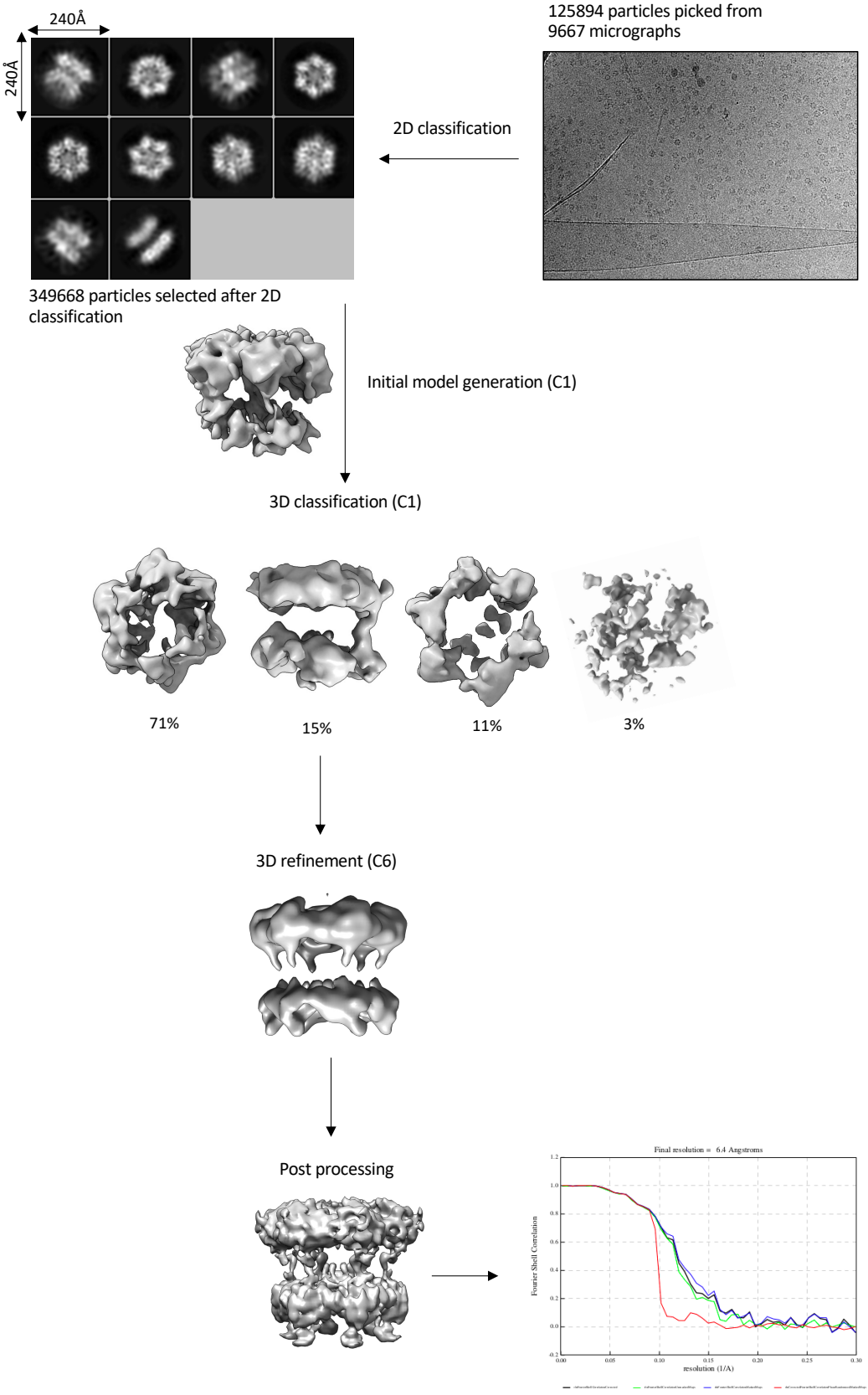

240Å
240Å
125894 particles picked from
9667 micrographs
2D classification
349668 particles selected after 2D
classification
Initial model generation (C1)
3D classification (C1)
71%
15%
11%
3%
3D refinement (C6)
Post processing
Final resolution = 6.4 Angstroms
Fourier Shell Correlation
resolution (1/A)

Figure 4.18. Simplified workflow of the Bd1152^{2C} cryo-EM data processing using Relion 4.0.

Particle picking was performed using 2D-reference based picking, followed by several rounds of 2D classification and generation of initial model. A model resembling the overall Bd1152$^{37-283}$ architecture was selected for further processing. Subsequent steps included 3D refinement in C6 symmetry, followed by postprocessing, resulting in a 6.4 Å resolution map. The average resolution derived from Fourier Shell Correlation (FSC) is depicted at the bottom of the figure. The FSC plot illustrates the correlation between the two independent refined maps (half-maps) with blue (no mask), loose mask (red), tight mask (cyan) and corrected (purple) curves. A cut-off 0.143 (blue vertical line) was used for resolution estimation. The density maps were visualised with ChimeraX (Pettersen *et al.,* 2021).

Subsequently, the Bd1152^{2C} cryo-EM dataset was reprocessed using cryoSPARC v.2.12.0. Motion correction was carried out on all micrographs using cryoSPARC implemented Patch-based motion correction, followed by CTF estimation using patch-based CTF. Particle picking was performed using the blob picker (Punjani *et al.,* 2017), followed by 2D reference-based picking. However, the same issue of alignment persisted in cryoSPARC, where side and top views were classified into separate classes, and an improved model was not obtained (data not shown).

4.3.3.4. Processing of Bd1152^{2C} within the Scipion framework.

While we carried out extensive data processing utilising Relion 4.0 and cryoSPARC v.2.12.0, the resolution was limited to 6-7 Å. Nevertheless, the 2D class averages of Bd1152^{2C} appeared to be of high quality, indicative of a dataset with much higher potential. Consequently, further advanced data processing might be necessary to reconstitute a high-resolution model.

Currently, the University of Birmingham lacks such expertise; therefore, we sought assistance from Prof. Jose Maria Carazo. Cryo-EM data processing was subsequently conducted within the *Scipion* 3.0.26 framework (de la Rosa-Trevín et al., 2016) at the CNB-CSIC (Madrid, Spain), I2PC centre as a part of an Instruct-ERIC internship grant (PID 2350) with the assistance of Dr Marcos Gragera.

Motion correction was performed using Motioncorr2 (Zheng et al., 2017). CTF estimation was carried out with both GCTF (Zhang, 2016) and CTFFIND4 (Rohou and Grigorieff, 2015), followed by metadata averaging using *Xmipp3* CTF consensus (Strelak et al., 2021). Micrographs were discarded based on defocus ratio (cut-off 1.05), resolution (cut-off 4 Å) and fit-quality (cut-off -0.05 and 0.10). At this step, a total of 1099 micrographs were discarded.

Four different particle picking approaches were utilised: *Xmipp*3 (de la Rosa-Trevín et al., 2013), Gautomatch2 (Zheng), crYOLO (Wagner et al., 2019) picking using a general pre-trained model, and crYOLO training using Bd1152^{2C} 2D side views, followed by crYOLO picking. The resulting particle sets from each picking algorithm were "cleaned up' through several rounds of 2D classification (cryoSPARC). The resulting particles were subjected to *Xmipp*3 picking consensus (Sanchez-Garcia et al., 2018) using a consensus parameter of 1, indicating the retention of particles picked by at least one algorithm. This resulted in a total of 967919 particles. Subsequent 2D classification was performed to inspect the output of *Xmipp*3 consensus, revealing a high number of side views and tilt class averages (Figure 4.19, Figure 4.20 and Appendix 4, A).



Figure 4.19. The 2D class averages of the Bd1152^{2C} particle set obtained after *Xmipp3* picking consensus.

Following the *Xmipp3* picking consensus of particles obtained from the *Xmipp3* picker, Gautomatch2, crYOLO (pre-trained model) and crYOLO (model trained with 2D classes of Bd1152^{2C}), the particles were subjected to 2D classification (cryoSPARC).

An initial model was generated using the *ab initio* algorithm (cryoSPARC) in C1 symmetry, resulting in four models (Figure 4.20, Appendix 4, B). Two models, which best resembled the anticipated Bd1152 structure, were selected for further processing (Appendix 4, B, class 2 and 3, 234874 particles). *Ab initio* (cryoSPARC), using particles from the selected model, was performed twice with exact same parameters (k = 4models) (Figure 4.20, Appendix 4, C, D). The output volumes were used for heterogenous refinement twice, using the exact same parameters (k = 4models) (from *ab initio* models from each run) (Figure 4.20, Appendix 4, E, F).

The resulting volumes were subjected to *Xmipp*3 consensus 3D clustering (Strelak et al., 2021), resulting in one subset containing particles with the highest confidence value for belonging to that particular particle cluster. The consensus particles were subjected to *ab-initio* (cryoSPARC) and *non-uniform* refinement in C1 symmetry, resulting in a 6 Å map (Appendix 5). *Non-uniform* refinement (cryoSPARC) in C6 symmetry was performed using a soft mask of the top MCE ring, resulting in the final map of the top MCE ring at 3.4 Å (Figure 4.21).

The local resolution of the map was estimated using the *Xmipp3* DeepRes algorithm (Ramírez-Aportela et al., 2019), revealing resolution fluctuations ranging from 3.4-3.6 Å across the map (Figure 4.21). A simplified processing pipeline is represented in Figure 4.20 and an extended processing figure is depicted in Appendix 4.

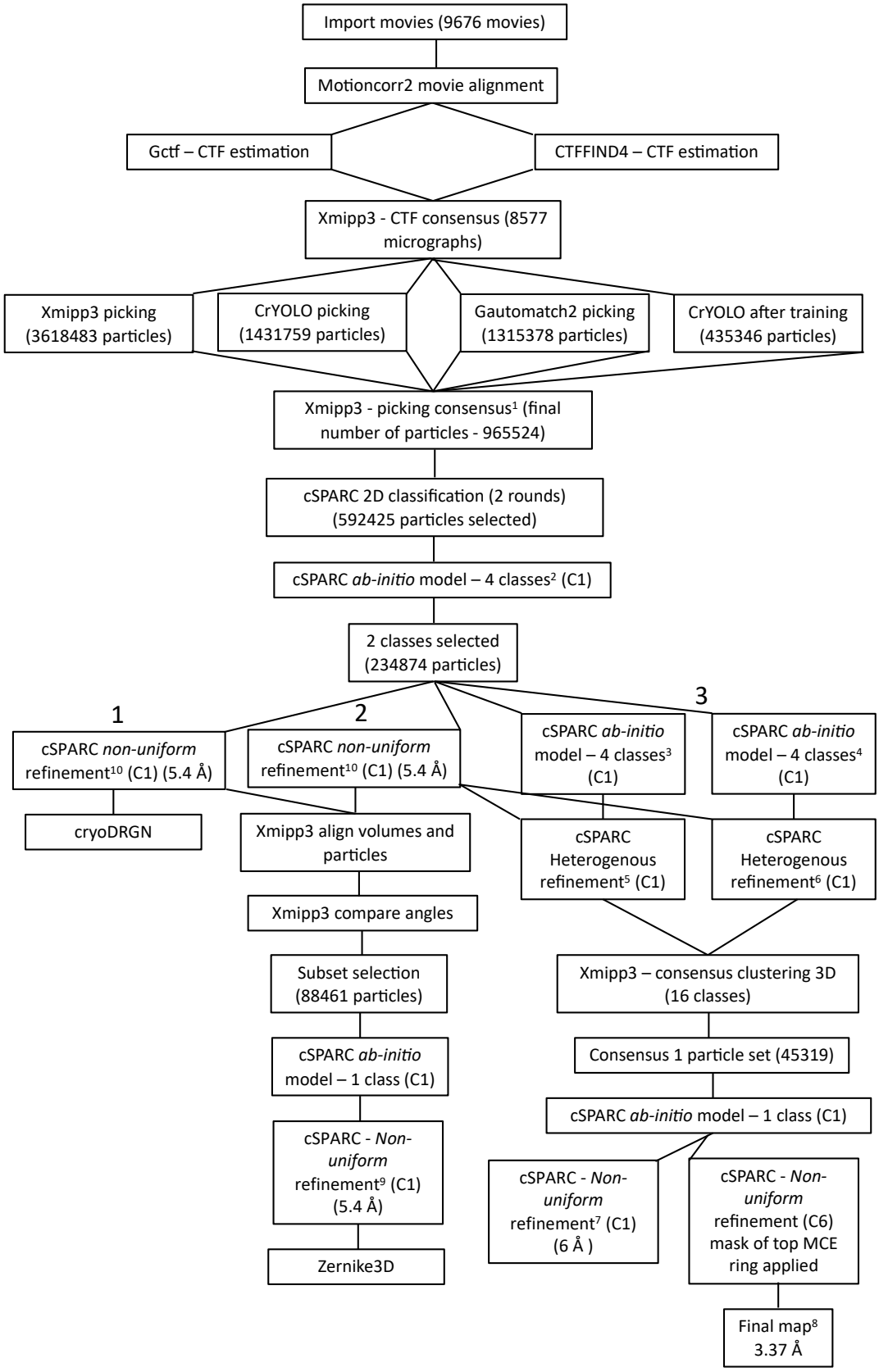

Import movies (9676 movies)
Motioncorr2 movie alignment
Gctf – CTF estimation
CTFFIND4 – CTF estimation
Xmipp3 - CTF consensus (8577 micrographs)
Xmipp3 picking (3618483 particles)
CrYOLO picking (1431759 particles)
Gautomatch2 picking (1315378 particles)
CrYOLO after training (435346 particles)
Xmipp3 - picking consensus[1] (final number of particles - 965524)
cSPARC 2D classification (2 rounds) (592425 particles selected)
cSPARC *ab-initio* model – 4 classes[2] (C1)
2 classes selected (234874 particles)
1
2
3
cSPARC *non-uniform* refinement[10] (C1) (5.4 Å)
cSPARC *non-uniform* refinement[10] (C1) (5.4 Å)
cSPARC *ab-initio* model – 4 classes[3] (C1)
cSPARC *ab-initio* model – 4 classes[4] (C1)
cryoDRGN
Xmipp3 align volumes and particles
cSPARC Heterogenous refinement[5] (C1)
cSPARC Heterogenous refinement[6] (C1)
Xmipp3 compare angles
Subset selection (88461 particles)
Xmipp3 – consensus clustering 3D (16 classes)
cSPARC *ab-initio* model – 1 class (C1)
Consensus 1 particle set (45319)
cSPARC - *Non-uniform* refinement[9] (C1) (5.4 Å)
cSPARC *ab-initio* model – 1 class (C1)
cSPARC - *Non-uniform* refinement[7] (C1) (6 Å)
cSPARC - *Non-uniform* refinement (C6) mask of top MCE ring applied
Zernike3D
Final map[8] 3.37 Å

Figure 4.20. Simplified workflow of the Bd1152^{2C} cryo-EM data processing using Scipion 3.0.26.

CTF estimation was performed using Gctf and CTFIND4, followed by *Xmipp3* CTF consensus. A total of 1099 micrographs were discarded at this stage. Particle picking was carried out using the *Xmipp3*, crYOLO, and Gautomatch picking algorithms, followed by *Xmipp3* picking consensus (for more details, please refer to Appendix 4, A). CryoSPARC *ab-initio* reconstitution generated four maps, from which two classes were selected for further processing (please refer to Appendix 4, B, for more details). This resulted in a total of 234873 particles. Three approaches were employed at this stage: 1) The 234873-particle set was used as an input for cryoSPARC *non-unform* refinement (C1), resulting in a 5.4 Å cryo-EM map. This map was further used for cryoDRGN analysis (please refer to Chapter 6 and Appendix 5). 2) The cryo-EM maps obtained from two *non-unform* refinement jobs (C1) (utilising the 234873-particle set) were subjected to the *Xmipp3* compare angles protocol, and a subset of 88461 particles was selected. CryoSPARC *ab-initio* reconstitution followed by *non-uniform* refinement resulted in a 5.4 Å map, which was further used for Zernike3D analysis (please refer to Chapter 6). 3) CryoSPARC *ab-initio* was performed twice, generating four classes each (Appendix 4, C, D). This was followed by cryoSPARC heterogenous refinement (Appendix 4, E, F). The resulting cryo-EM maps underwent the *Xmipp3* consensus clustering protocol, and a subset of 45319 particles was selected. This particle set was used for cryoSPARC *ab-initio* reconstitution, followed by *non-uniform* refinement in both C1 and C6 symmetries. These resulted in cryo-EM maps with resolutions of 6 Å and 3.37 Å, respectively (Appendix 6 and Figure 4.21, respectively).

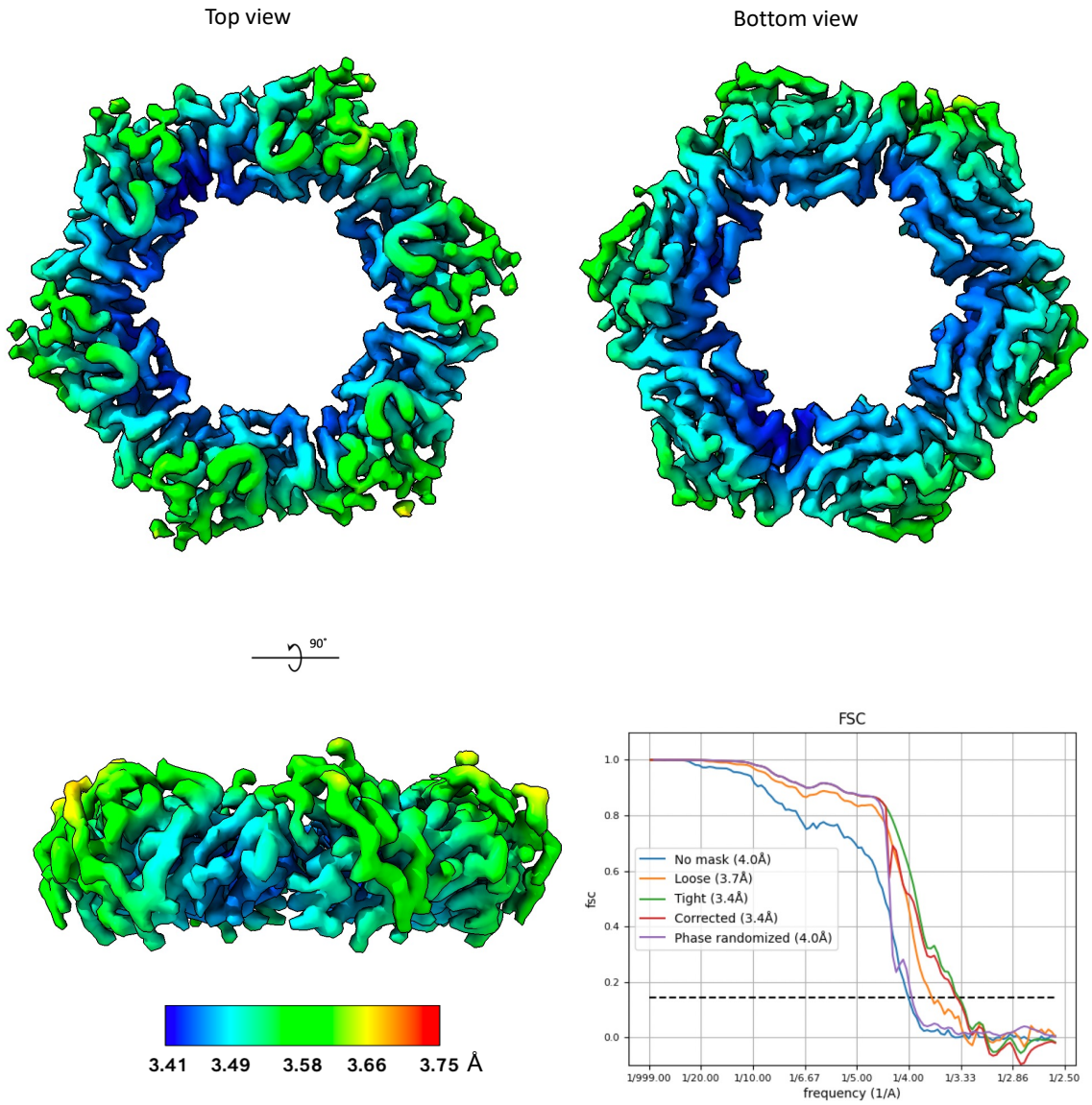


Figure 4.21. The final cryo-EM map of the Bd1152^{2C} top MCE ring at 3.4 Å in C6 symmetry.

Maps are coloured according to the local resolution estimation (Å) with the colour key depicted at the bottom left of the figure. (Bottom right) Global resolution of the map derived from the Fourier Shell Correlation (FSC). The FSC plot showing correlation between the two independently refined maps (half-maps) with blue (no mask), loose mask (red), tight mask (cyan) and corrected (purple) curves. A cut-off 0.143 (black vertical dotted line) was used for resolution estimation. The density map was visualised with ChimeraX (Pettersen *et al.,* 2021).

4.3.3.5. Model building

Next, we utilised the C6 symmetrical map of the Bd1152^{2C} top MCE ring at 3.4 Å resolution to build an atomic model. Initially the cryo-EM map was improved to aid the modelling process by running DeepEMhancer (Sanchez-Garcia *et al.,* 2021). The resulting map had clear MCE domain fold density. We used the sequence of Bd1152^{2C} and the current density map as inputs for the ModelAngelo algorithm (Jamali *et al.,* 2023). Resulting segments were assembled into a continuous protein chain using COOT (Emsley *et al.,* 2010) and subsequently refined using Phenix Real Space Refinement (Afonine *et al.,* 2018). (Figure 4.22). Refinement statistics are provided in Appendix 7.

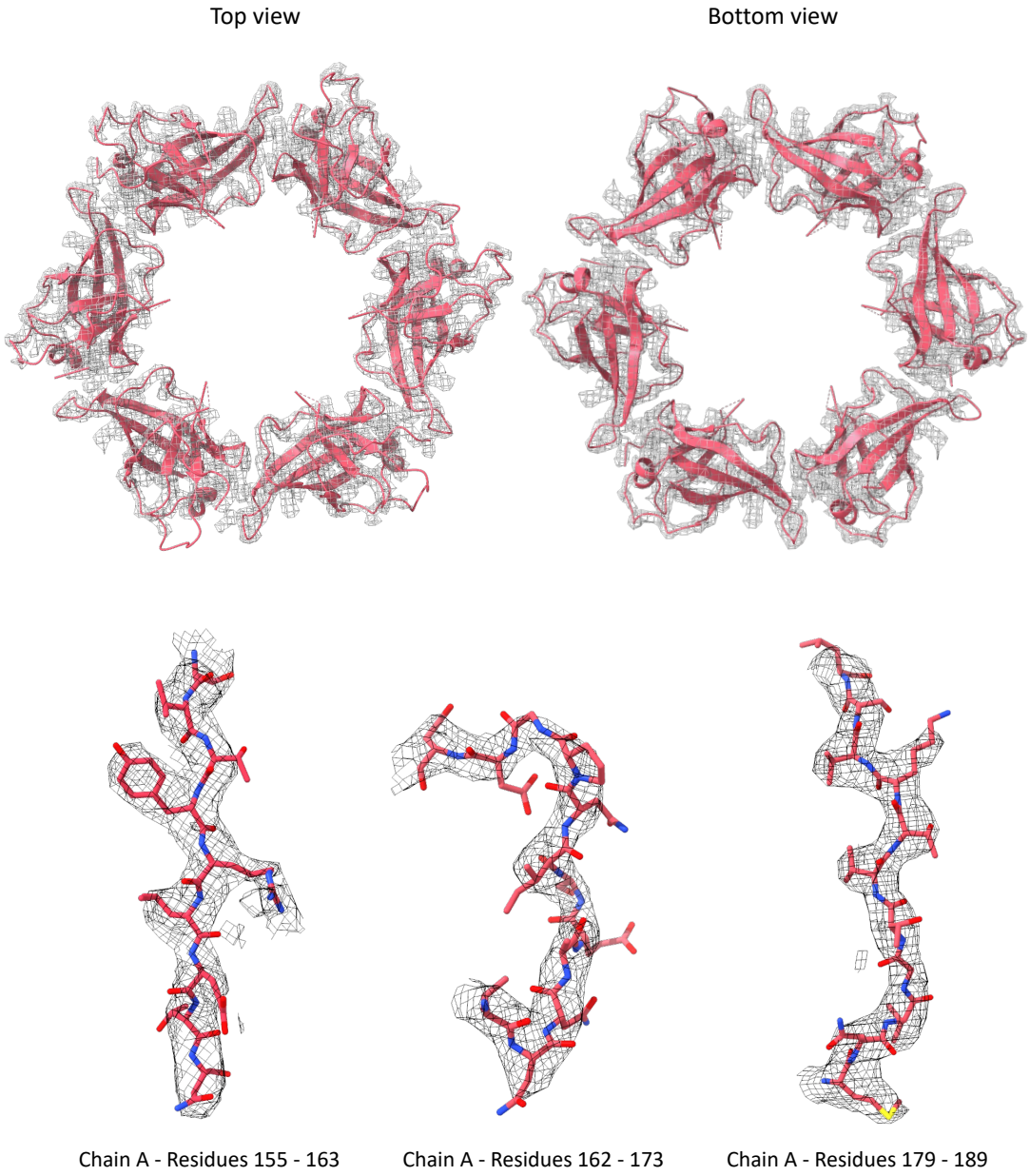


Figure 4.22. The cryo-EM model of the Bd1152^{2C} top MCE ring obtained from ModelAngelo and Phenix real space refinement.

Figure represents the cryo-EM density map (grey) of the Bd1152^{2C} top MCE domain with the refined ModelAngelo Bd1152^{2C} model fitted. (Bottom) Representative regions of the ModelAngelo model built into the cryo-EM of the Bd1152^{2C} top MCE domain. The cryo-EM density map is displayed at a contour level of 0.3. The density map and the model were visualised with ChimeraX (Pettersen *et al.,* 2021)

4.3.3.6. Cryo-EM analysis of the Bd1152^{2C} bottom MCE ring.

Whilst we obtained a high-resolution map of the Bd1152^{2C} top MCE ring in C6 symmetry, the bottom ring exhibited low-resolution and "false" features. We predicted that due to inherent continuous flexibility, the lower and upper MCE rings are not aligned to the same symmetry axis. Therefore, imposing 6-fold symmetry only resolves the top MCE ring and creates symmetry artifacts within the bottom MCE ring.

We utilised *Xmipp*3 signal subtraction and subsequent refinement in C1 and C6 symmetries. However, the map obtained did not resemble the MCE domain fold (Figure 4.23, 1). Next, we attempted to relax the C6 symmetry by employing the *Xmipp3* symmetry break protocol. We performed Relion 3D classification with finer angular sampling (0.9 degrees) and the symmetry relaxation flag, followed by Relion 3D refinement (Figure 4.23, 2). Inspection of the map showed that symmetry was indeed relaxed, and the density corresponding to the lower MCE ring has improved, revealing a low-resolution MCE domain fold density. Nevertheless, the resolution of the overall map was limited to ~ 6 Å.

Assessment of the low-resolution cryo-EM map led us to speculate that the MCE-domain rings may not share the same 6-fold symmetry. We then attempted to break C3 symmetry and subsequently performed Relion 3D classification and refinement. However, the quality of the map did not improve (Figure 4.23, 3). Nevertheless, 3D *non-uniform* refinement of the overall protein in C3 symmetry resulted in a 3.8 Å map, where the lower domain displayed an overall MCE fold resembling the one obtained from the *Xmipp*3 symmetry break protocol (Figure 4.23, Figure 4.24). Local resolution of the map and a corresponding FSC plot are depicted in Appendix 8.

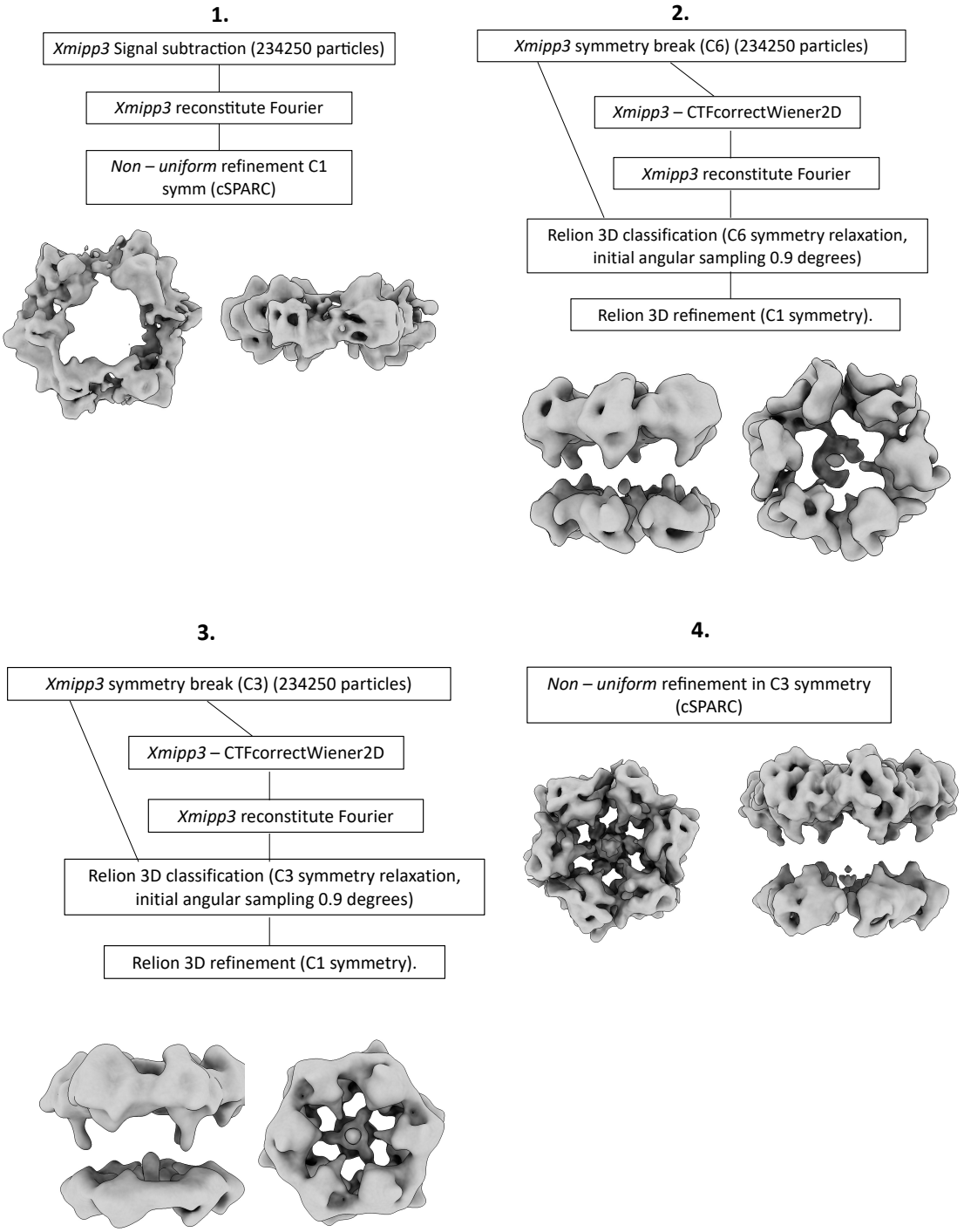


Figure 4.23. The cryo-EM data processing approaches utilised to address the symmetry mismatch between the Bd1152^{2C} top and bottom MCE rings.

The CTF values were applied to the particle set using the *Xmipp3* CTFcorrectWiener2D protocol. The density maps were visualised with ChimeraX (Pettersen *et al.,* 2021).

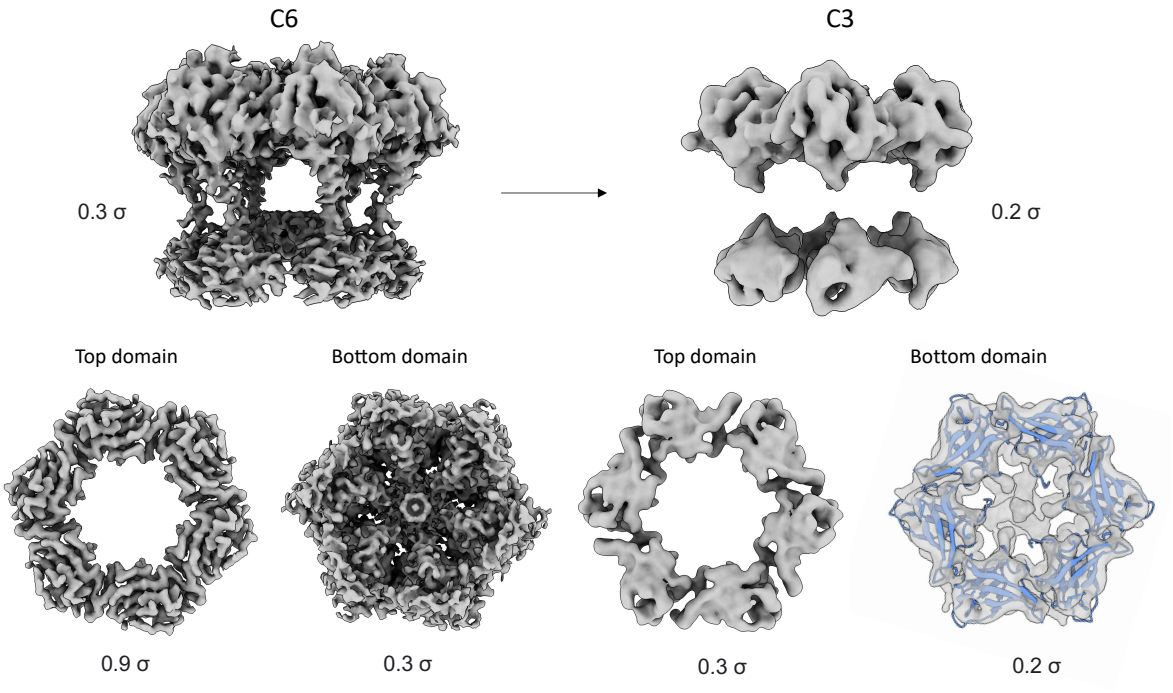


Figure 4.24. Comparison of the Bd1152^{2C} MCE ring densities obtained from *non-uniform* refinement in C6 and C3 symmetries.

Blue cartoon represents the Bd1152 Alphafold model (residues 157-219; 243-283) fit into the C3 cryo-EM map. Maps displayed at contour levels (σ) indicated in the figure. The density maps and the model were visualised with ChimeraX (Pettersen *et al.,* 2021).

4.3.3.7. Affecting the Bd1152^{2C} preferential top orientations with detergent screen

Whilst graphene oxide coated Quantifoil grids (R1.2/1.3, Au mesh 300) facilitated working with very low concentrations of the sample and eliminated the need for protein concentration, our datasets exhibited a persistent preference for top view orientations. This could represent a contributing factor to our inability to reconstitute a high-resolution map of the full Bd1152^{2C} structure. To address this issue, Quantifoil grids (R1.2/1.3, Au mesh 300) were utilised without the addition of graphene oxide. However, this approach proved to be unsuccessful, as protein particles stuck to the carbon film and were excluded from the ice or particles were absorbed at the air-water interface, resulting in predominantly top view orientations (Figure 4.25, left).

To overcome this issue, we plunge froze Quantifoil grids (R1.2/1.3, Au mesh 300) with 2mg/ml of purified Bd1152^{2C} spiked with 0.5 x CMC of different detergents before freezing. We tested 0.5 x CMC of Tween20, βOG, DM and FOM. Among these, only Bd1152^{2C} spiked with FOM displayed a significantly improved particle distribution compared to the control (no detergent addition). The particles were more evenly distributed, and a significantly larger number of side views were present (Figure 4.25. right). Data was collected on Talos Arctica 200 kV microscope at CNB-CSIC (Madrid, Spain) as part of the Instruct-ERIC access proposal grant PID23275.

2D classification revealed the presence of more side views with the lower ring aligned to a higher resolution. The particle distribution with respect to preferential orientations also improved, revealing a variety of different orientations (Figure 4.26). It is important to note that an unusual class was observed in this dataset (Figure 4.26, row 1, column 5). We hypothesised that the top and bottom MCE rings might be dissociating, and the projection observed around the hexamer could correspond to the short linkers (short regions

connecting the top and bottom MCE domains). After extensive processing we were able to obtain a 4.1Å map in C6 symmetry, however, we were not able to improve the resolution of the bottom MCE domain (Figure 4.26).

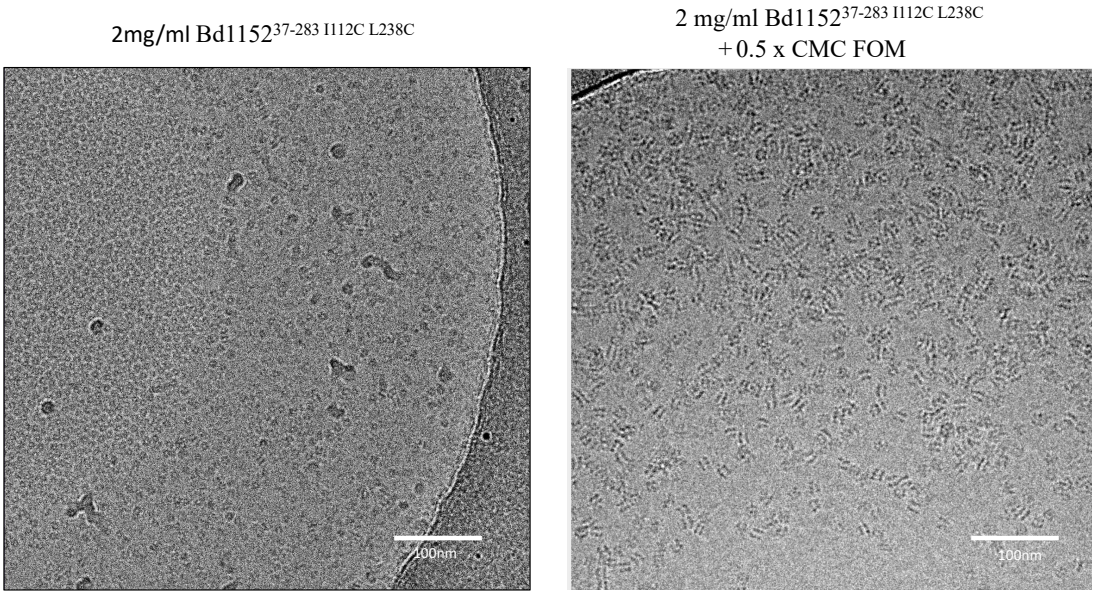


Figure 4.25. The impact of fluorinated octyl maltoside (FOM) on the Bd1152^{2C} particle distribution.

Bd1152^{2C} particle distribution on AU R1.2/1.3 300 mesh holey grids without (left) and with 0.5 x CMC FOM (octyl maltoside, fluorinated) (right).

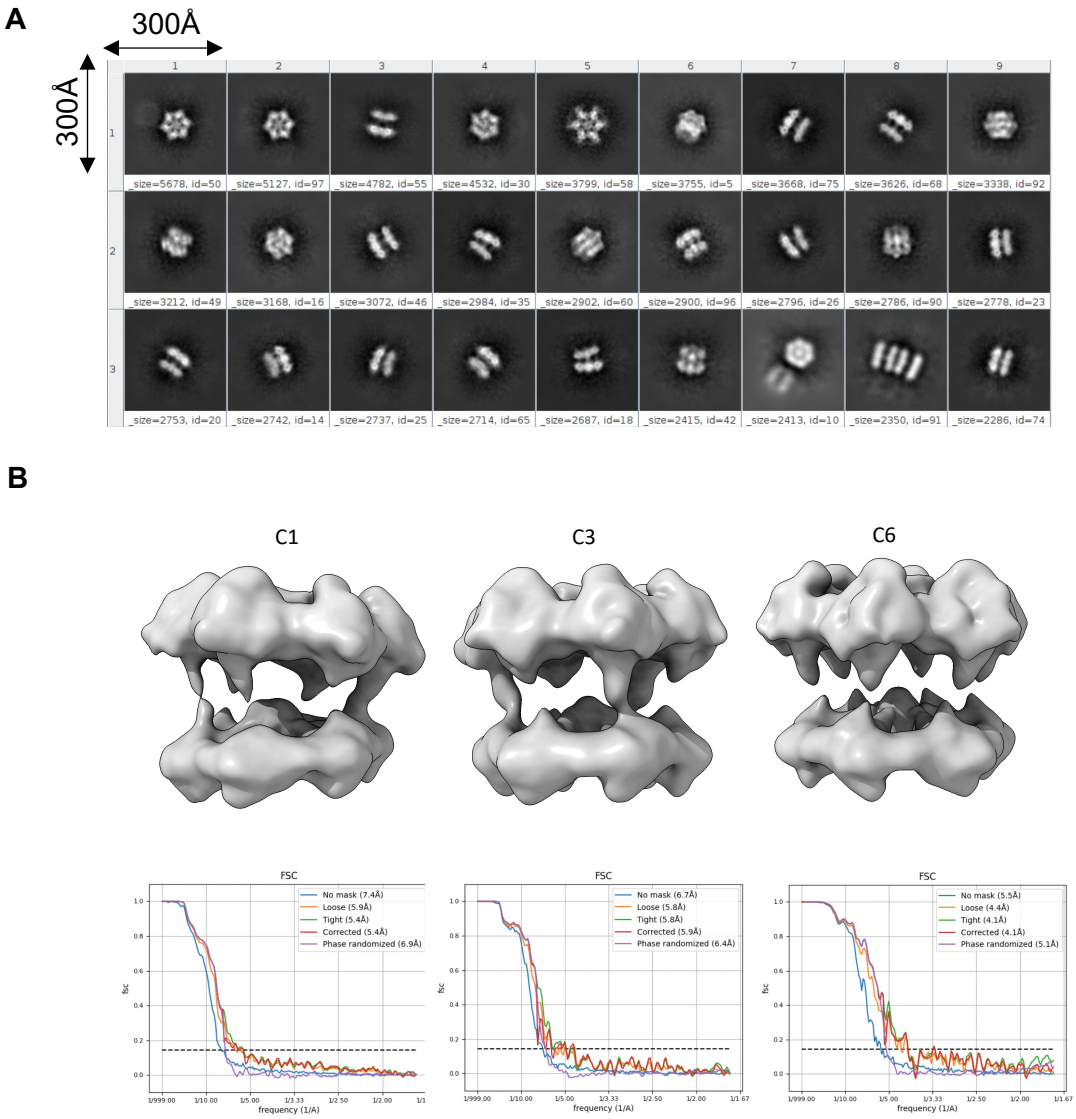


Figure 4.26. The cryo-EM data processing results of the Bd1152^{2C} sample spiked with 0.5 x CMC FOM (fluorinated octyl maltoside)

(A) 2D classification revealing a number of different particle orientations (A total number of particles 147298). (B) Final 3D refinements in C1 (left), C3 (Middle) and C6 (right) symmetries displayed as a side view (middle) with corresponding Fourier Shell Correlation (FSC) plots. The FSC plots show correlation between the two independently refined maps (half-maps) with blue (no mask), loose mask (red), tight mask (cyan) and corrected (purple) curves. A cut-off 0.143 (blue line) was used for resolution estimation. Data was processed within Scipion 3.0.26. Maps are displayed at the contour level of 0.2. The density maps were visualised with ChimeraX (Pettersen *et al.,* 2021).

4.3.3.8. Generation of the Bd1152^{2C} GraFix sample

The inherent continuous flexibility present within the Bd1152^{2C} cryo-EM dataset represents a major bottleneck in high-resolution reconstitution. While disulphide bond engineering increased the stability of the protein to some extent and allowed us to reconstitute the top MCE ring, the flexibility still persisted and precluded us from obtaining a complete high-resolution structure of Bd1152^{2C}. To address this issue, we attempted to stabilise the Bd1152^{2C} through chemical glutaraldehyde crosslinking using the GraFix method (Stark, 2010).

Briefly, purified protein sample was applied to a sucrose gradient (5-30%) (0.015% of glutaraldehyde was added into the 30% sucrose solution), followed by centrifugation. Fractions were retrieved and analysed with SDS-PAGE. The fraction containing a complex corresponding to the Bd1152 hexamer was further assessed with SEC.

Purification of Bd1152^{2C} after GraFix is represented in Figure 4.27. SEC revealed two peaks representing aggregated and correctly folded Bd1152^{2C}, respectively. SDS-PAGE suggested that glutaraldehyde crosslinking occurred, as a major band was seen at ~210 kDa, implying that Bd1152^{2C} remained in a hexameric assembly even under denaturing conditions.

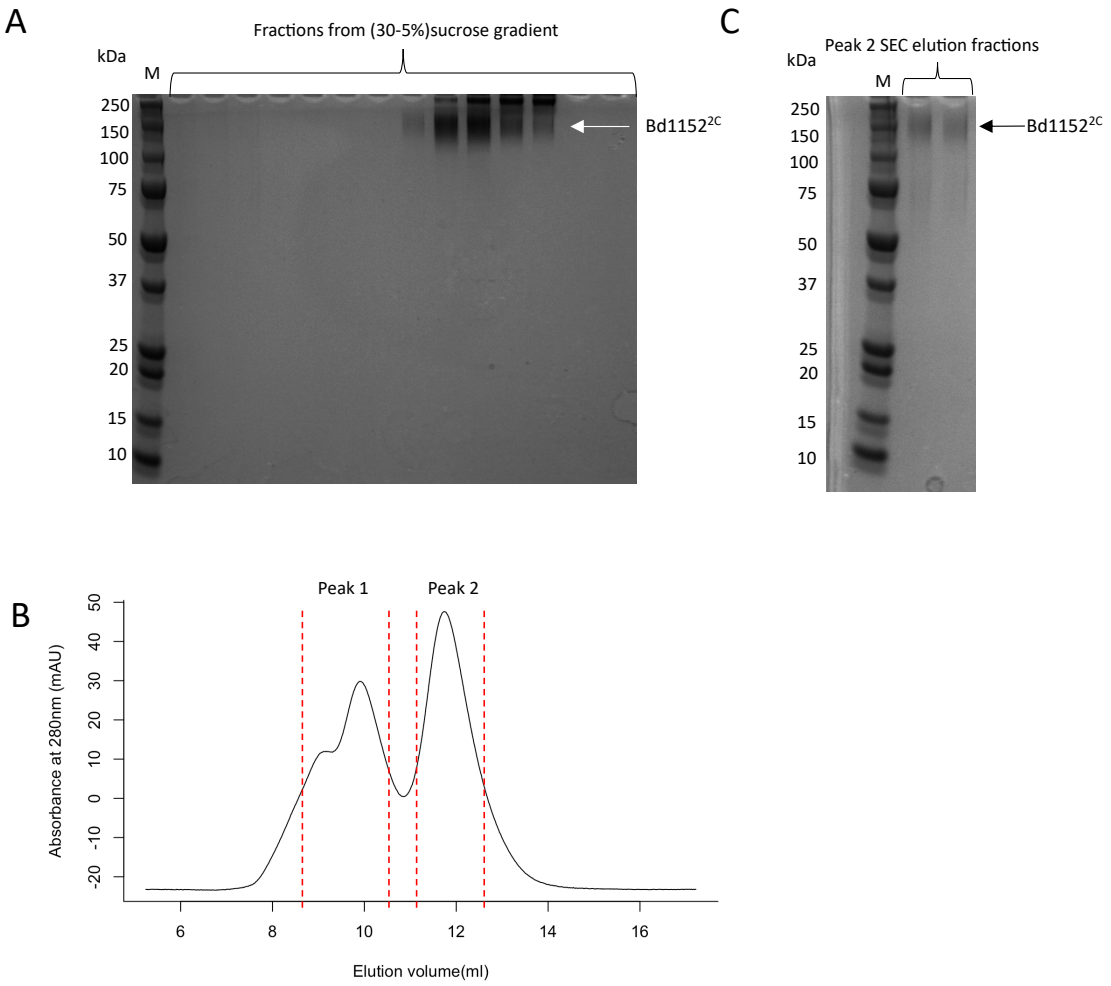


Figure 4.27. Purification of the Bd1152^{2C} GraFix sample,

(A) SDS-PAGE of samples directly taken from the sucrose gradient after ultracentrifugation revealed a prominent band at ~210 kDa, corresponding to the hexameric assembly of Bd1152^{2C} (B) SEC (S200 10/300) revealed two peaks, eluting between 8 -10 mL and 11 - 13 mL, respectively. (C) SDS-PAGE of the elution fractions from the second peak showed a prominent band at ~210 kDa, corresponding to the hexameric assembly of Bd1152^{2C}. SDS-PAGE was performed using 4-12% bis-tris gels.

Approximately 1500 micrographs were collected of the Bd1152^{2C} GraFix sample showing an improved particle distribution and significantly increased number of side views. However, 2D classification revealed numerous particle conformations; some displayed a clear “lift” of the top MCE ring, others showed a more “flat” conformation. In addition, certain 2D classes revealed that there is a considerable length variation between the top and bottom MCE domains (Figure 4.28). While these conformations were classified and refined separately, a high-resolution structure was not obtained, mostly likely due to the high conformational heterogeneity of the sample.

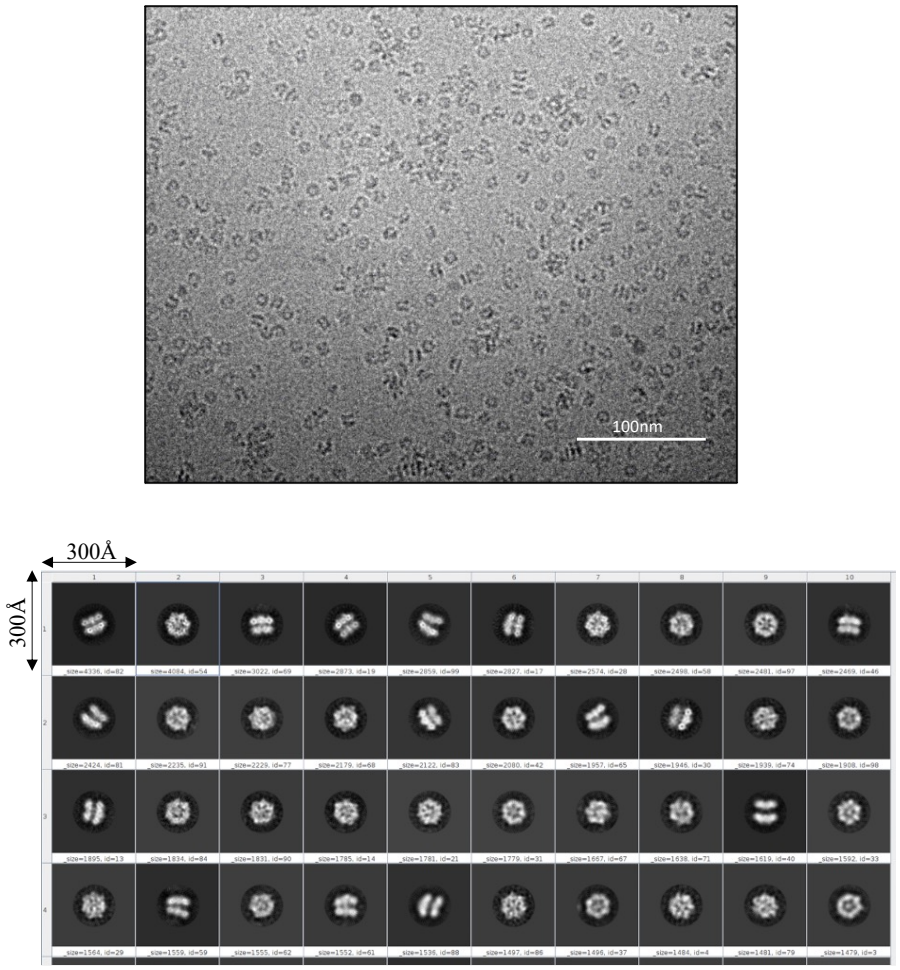


Figure 4.28. Cryo-EM of the Bd1152^{2C} GraFix sample.

(Top) An example micrograph of the Bd1152^{2C} GraFix sample, revealed good particle distribution. (Bottom) 2D classification (cryoSPARC) of the Bd1152^{2C} GraFix sample revealed a highly heterogeneous dataset.

4.4. Discussion

In this chapter, we utilised cryo-EM to probe the structure of the membrane-associated Bd1153-Bd1152 complex. After extensive sample optimisation and cryo-EM data processing, we presented a high-resolution structure of the Bd1152 top MCE ring.

4.4.1. Challenges in structural characterisation of Bd1153.

It is important to note that Bd1153 consistently remained present throughout the purification, as confirmed by mass spectrometry analysis (data not shown). However, the cryo-EM data revealed that only Bd1152 was present in the class averages, as no discernible Bd1153 density was observed within the DDM micelle. Bd1153 exhibits sequence similarity to the *E. coli* inner membrane PqiA protein, a member of the PqiABC pathway, with only Alphafold providing insight into the protein's fold. Results obtained in our lab (data not published), point to a notable challenge in understanding the structure and function of PqiA. This challenge primarily stems from the intricacies involved in the expression and purification of this protein. Notably, when expressed as a single gene, it exerts toxicity on the *E. coli* expression strain. Additionally, as shown by Cooper (2023), PqiA density was not observed in the PqiAB and PqiABC cryo-EM datasets, despite being present throughout the purification (data not published). Interestingly, the mass spectrometry results of the PqiABC sample revealed the presence of FtsH, a metalloprotease which is capable of degrading membrane proteins (Yang *et al.,* 2018). It was hypothesised that FtsH may be responsible for the degradation of PqiA during the period between protein purification and cryo-EM grid vitrification (Cooper, 2023). Following this, further experiments should aim to re-design the Bd1153-Bd1152 construct to contain a strep tag. This will avoid the necessity of Ni-beads hence allowing the EDTA chelating agent to be used throughout the purification to negate the potential FtsH action. Additionally, future avenues should explore the stabilisation of this complex, possibly

though the introduction of disulphide bonds, utilisation of chemical crosslinkers such as glutaraldehyde or the incorporation of the non-native amino acid p-benzoyl-L-phenylalanine (BPA). It is noteworthy, that we re-designed the Bd1153-Bd1152 construct to contain a 6x histidine tag on the n-terminus, with the aim of increasing the yield of purified Bd1153. This modification resulted in an improved purification, as evidenced by the increased yield of Bd1153 observed on SDS-PAGE (Appendix 9) compared to the original construct with the c-terminal tag. However, cryo-EM screening of the purified sample revealed significant heterogeneity (data not shown), preventing us from assessing whether Bd1153 remained associated with Bd1152. Nonetheless, this approach shows promise in elucidating the structure of Bd1153 through further refinement of the sample purification and cryo-EM grid preparation protocols.

4.4.2. Generating a more stable Bd1152 construct.

Probing the structure of Bd1152 via cryo-EM proved to be challenging. During 2D and 3D classification, it was evident that the top MCE ring exhibited a noticeable “tilting”, and the final cryo-EM maps were limited to ~ 9 Å. We next decided to simplify the construct and only focus on elucidating the structure of Bd1152, lacking its transmembrane domain (residues 1-36). Previous studies have shown that the removal of the transmembrane helix and cytoplasmic expression does not affect the folding and assembly of MCE domain-containing proteins in *E. coli*, such as MlaD, LetB, and PqiB (Ekiert *et al.,* 2017; Isom *et al.,* 2020). Nonetheless, the purification of $Bd1152^{37\text{-}283}$ was challenging, as the protein was highly unstable and prone to aggregation, with cryo-EM 2D class averages revealing a number of different conformational states. This provided us with a further indication that Bd1152 may exhibit high structural plasticity. Whilst investigating the flexibility landscape of a protein is of great interest, the 2D and 3D classification of both the Bd1152 and $Bd1152^{37\text{-}283}$ samples

suggest that motion might prevent high resolution reconstitution. Accordingly, a more rigid construct needs to be obtained allowing us to reconstitute a high-resolution map. Introduction of disulphide bonds is an attractive strategy to increase the stability of protein structures (Matsumura *et al.,* 1989; Zavodszky *et al.,* 2001), hence we decided to introduce cysteine mutations into the PLLs of $Bd1152^{37\text{-}283}$ with the aim of reducing inter-domain flexibility and increasing the overall stability of the protein. Purified $Bd1152^{2C}$ was significantly more stable and amenable to spin concentrating, hence was further used for X-ray crystallography (discussed in Chapter 5) and cryo-EM studies. Whilst a significantly improved map was obtained, the resolution was still limited to 6.4 Å.

4.4.3. Issues in the cryo-EM data processing of the $Bd1152^{2C}$ sample.

Several data processing issues were encountered, including particle alignment and flexibility. Therefore, we sought assistance from Prof Jose Maria Carazo and data was subsequently transferred to CNB-CSIC I2PC centre (Madrid, Spain). The dataset was processed by utilising a number of different cryo-EM processing software in parallel within the *Scipion* (de la Rosa-Trevín *et al.,* 2016) framework. Whilst we obtained a final map of the top MCE ring at 3.4 Å in C6 symmetry, bottom MCE ring, however, did not resemble the MCE-fold and showed number of artifacts most likely resulting from imposing C6 symmetry during refinement.

4.4.3.1. Flexibility induced inter-domain symmetry mismatch and particle alignment issues.

Initially we hypothesised, that the flexibility inherent to Bd1152 could cause a break in its dominant symmetry, leading to its categorisation as a quasi-hexameric (qC6) protein (Huiskonen, 2018). This deformation of "perfect" hexameric rings may have caused the quasi-symmetry and significantly reduced resolution when C6 symmetry was applied. According to Huiskonen (2018), it is useful to describe a quasi-symmetrical structure only if relaxing the symmetry improves the final map. Although breaking C6 symmetry did not result in a high-resolution reconstitution of the bottom MCE ring, the map improved significantly with visible MCE domain fold density.

We proceeded to investigate whether both MCE rings of Bd1152^{2C} possess the same six-fold symmetry. We imposed C3 symmetry during *non-uniform* refinement and achieved a 3.8 Å cryo-EM map of the overall Bd1152^{2C} protein. Although high-resolution features for the lower MCE ring could not be resolved, the overall MCE-domain fold was discernible and closely resembled the volume map obtained from the *Xmipp* symmetry break protocol. Similar symmetry mismatch issue has been reported in the *E. coli* LetB structure, where authors showed that rings 1, 2, 3 and 4 are 6-fold symmetrical, whilst rings 5, 6 and 7 are 3-fold symmetrical (Liu *et al.,* 2020). Nevertheless, Isom *et al.,* (2019) succeeded in solving the overall LetB structure in 6-fold symmetry.

In order to summarise and provide further elaboration on the symmetry issues encountered during the data processing of the Bd1152^{2C} sample, we present simplified geometrical representations that are compiled in Figure 4.30. Figure depicts three hypothetical scenarios regarding the symmetry mismatch issues, in which the hexagon and triangle correspond to the top and bottom MCE rings of Bd1152, respectively. In scenario A, the hexagon (C6) and

triangle (C3) are perfectly aligned on the same symmetry axis. If C6 symmetry is applied to both figures (rotating 60 degrees any of the sector on the 6-fold symmetry axis), the hexagon will be reproduced, but not the triangle. In scenario B, both hexagon and triangle are aligned to the same symmetry axis, but this time C3 symmetry is applied resulting in producing both hexagon and triangle. In scenario C, symmetry axis of hexagon and triangle are not aligned. If C3 symmetry is applied, only the hexagon (or triangle) can be reproduced. Scenario A represents the high-resolution reconstitution of the top MCE-ring in C6 symmetry, however, resulting in lower domain with “fake” features. Scenario B represents reconstitution of the cryo-EM map of the overall protein with a high-resolution top MCE ring and a low-resolution bottom MCE ring but displaying clear MCE-fold density. We believe that scenario C best represents the symmetry mismatch within Bd1152, resulting from a combination of flexibility, unalignment to the symmetry axis, and the possibility of the bottom MCE ring possessing C3 symmetry. This makes it challenging to reconstitute both MCE domains to a high-resolution.

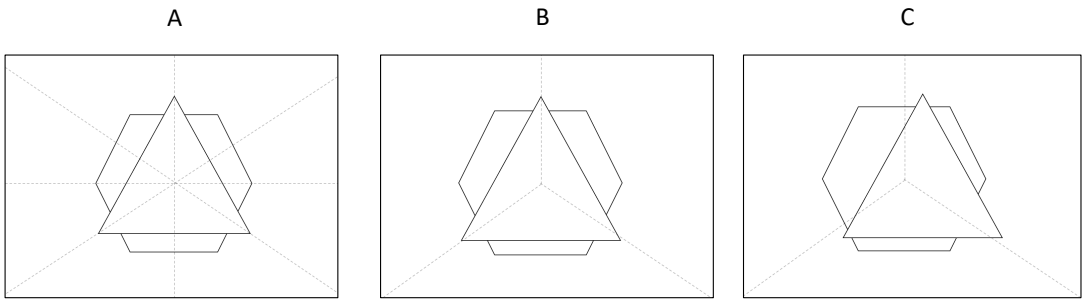


Figure 4.29. Geometrical representation of imposing C6 and C3 symmetries during the refinement of Bd1152^{2C}.

Dotted lines represent the symmetry axes. The hexagon represents the top MCE ring whilst the triangle represents the bottom MCE ring of Bd1152. A) If both the hexagon and triangle are aligned to the same symmetry axis and C6 symmetry is imposed, only the hexagon can be reconstituted. B) If both the hexagon and triangle are aligned to the same symmetry axis and C3 symmetry is imposed, both the hexagon and triangle can be reconstituted. C) If the hexagon and triangle are not aligned to the same symmetry axis and C3 symmetry is imposed, it will only allow the reconstitution of the hexagon or the triangle.

Both MCE rings, however, form homo-hexameric rings making it unclear why imposing three-fold symmetry results in a much higher quality cryo-EM map of the bottom ring compared to six-fold symmetry. Although we were not able to resolve the pore lining loops, we speculate that perhaps three disulphide bonds have formed between C112 of adjacent monomers rather than between C112 and C238 within each protomer. Such a scenario is perhaps sufficient to break the 6-fold symmetry. If this assumption is indeed correct, this potentially could be resolved via the incorporation noncanonical amino acids containing long side-chain thiols (Liu *et al.,* 2016). This could overcome the potential distance constraint between the C112 and C238 residues.

It is noteworthy that in order to address the issue of symmetry mismatch, we extensively utilised signal subtraction and local refinement protocols with the aim to reconstitute both MCE rings separately, which theoretically should resolve the symmetry mismatch issue. Interestingly, the reconstitutions of individual MCE rings revealed highly-overfitted maps (data not shown), somewhat suggesting that additional alignment issues might be present within the dataset.

Indeed, particle misalignment appeared to be a persistent issue throughout all the datasets we collected. For both the Bd1152 and Bd1152^{2C} datasets, top and side views were classified into separate classes. If attempting to "force" the alignment by asking for a single class during 3D classification, 3D refinement or *ab initio* reconstitution, top views would be aligned on the wrong axis, resulting in an incoherent overall model. The reason for this remains unclear, perhaps the overall shape of the protein might be only defined at higher resolution; at low resolution the dimensions make it such that orientation is ambiguous (Personal communication with Prof Jose Maria Carazo, Prof Carlos Oscar, Prof Carolyn Moores, Dr Marcos Gragera). Future avenues should explore the tagging of Bd1152 with additional molecules to facilitate particle alignment. For example, recent development of mega-bodies that bind specifically to membrane scaffold protein (MSP) (Uchański *et al.,* 2021) could be employed to aid the particle alignment of the full-length Bd1152 construct reconstituted into MSP nanodiscs.

4.4.3.2. Improving particle distribution on the EM grid and the rigidity of Bd1152^{2C}.

In addition, we proposed that the number of side views could also limit the final resolution of the bottom MCE ring. Top views are much more abundant than side views within the Bd1152^{2C} dataset. We hypothesised that although there is no clear "missing wedge" issue, the lack of signal at the bottom MCE domain could result from an insufficient number of side view particles. To tackle the issue of orientation we attempted to freeze cryo-EM grids with no graphene oxide coating. Whilst graphene oxide seems to be an excellent choice, as we were able to work with extremely low concentrations of protein and obtain reasonably good particle distribution, graphene oxide coating could have caused preferential orientations (personal communication with Dr Christos Savva). To prepare Quantifoil grids with no graphene oxide coating was tried previously, however, achieving thin ice and even particle distribution was unsuccessful. Usually particles interact with the air-water interface causing them to adopt only top views. Adding detergents below their CMC value has been shown to provide a layer between the air-water interface, hence protecting the particles and improving their distribution on a grid (Kampjut *et al.,* 2021). We tested four different detergents, β-OG, Tween 20, DM and FOM. Addition of the later appeared to have a dramatic influence with 2D classification revealing a significantly improved ratio between the top and side views, with a number of different orientations. However, the flexibility of the particle remained, and we were not able to obtain a high-resolution structure. We also tried to improve the rigidity of Bd1152^{2C} by chemical crosslinking. The GraFix method (Stark, 2010) was selected as it has been previously shown to reduce sample heterogeneity arising from inherent flexibility and improve particle distribution on the grid (Stark, 2010).

Cryo-EM of the Bd1152^{2C} GraFix sample revealed improved particle distribution and quality. However, 2D classification demonstrated that the sample remained heterogeneous with different conformational states visible. Unfortunately, this Bd1152^{2C} GraFix cryoEM dataset was too heterogeneous to reconstitute a high-resolution 3D model.

Finally, it is important to note that we were not able to observe PLL density in the cryo-EM maps. Nevertheless, the low-resolution full-length Bd1152 map displayed a more apparent "tunnel" like density between the bottom and top MCE domains suggesting that the PLLs of Bd1152 might be adopting a similar conformation to the *E. coli* PqiB and LetB proteins (Ekiert et al., 2017; Isom et al., 2020). The flexibility of the Bd1152 PLLs will be further discussed in Chapters 5 and 6.

4.5. Conclusion

The intrinsic flexibility of Bd1152 has posed a significant challenge in our current study. Despite employing various experimental and data processing approaches, as well as the cryo-EM data processing expertise within the lab of Prof Jose Maria Carazo, we were only able to achieve a high-resolution reconstitution of the top MCE ring. Moreover, we encountered several data processing issues, particularly concerning particle alignment and symmetry-related problems. These findings underscore the complexity of investigating the structure of Bd1152 via cryo-EM and highlight the need for further sample optimisation or utilisation of more advanced data processing approaches to uncover its dynamic nature (further discussed in Chapter 6).

Chapter 5 - X-ray crystallography of Bd1152

5.1. Introduction

Since the first protein crystal X-ray diffraction in the early 1930s (Randall *et al.,* 1930), X-ray crystallography has made significant contributions to the understanding of fundamental cellular processes. Despite the emergence of new structural techniques, X-ray crystallography continuous to be the most established and widely used method for determining the structure of macromolecules. The method relies on the properties of electron waves and the ordered, repetitive nature of crystals. The experimental workflow involves three main steps: (1) Crystallisation, (2) Diffraction, (3) and Phasing

5.1.1. Crystallisation

The ordered and repetitive nature of crystals enables X-rays to interact with atoms within the protein, generating a diffraction pattern. This diffraction pattern is then used to determine the positions of individual atoms and ultimately the molecular structure of the protein.

Protein expression and purification are major factors determining successful crystallisation. Typically, a protein needs to be expressed to a high yield (8-20mg/ml) (McPherson and Gavira, 2013) and purified to a high purity (>95%) (Kors *et al.,* 2009). Any contamination present in the sample can interfere with the periodic packing of the macromolecules within the crystal lattice. Generally, the more ordered the crystal lattice, the higher the quality of the crystal and diffraction data.

Crystal growth is usually achieved by adding salts and polymers to a protein mixture, leading to supersaturation and the formation of protein crystals (McPherson and Gavira, 2013). As precipitants compete for hydration, the protein concentration gradually rises, eventually causing the protein to precipitate or crystalise under optimal conditions (Figure 5.1).

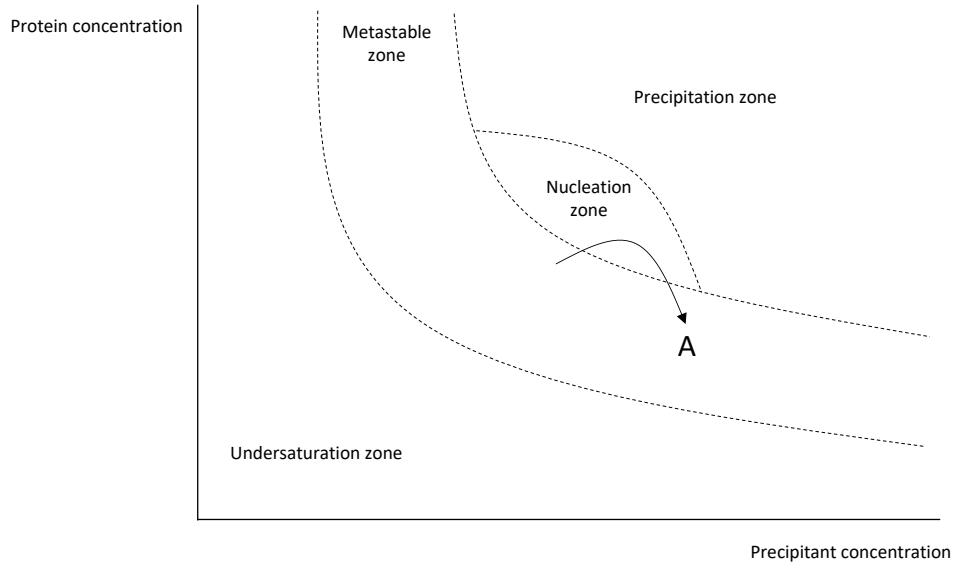


Figure 5.1. Crystallisation phase diagram.

The plot illustrates the phase conversion of a protein solution as a function of its concentration on the y-axis and precipitant concentration on the x-axis. In the vapor diffusion method, water diffuses out of the drop, resulting in the concentration of both protein and precipitant until the system enters the nucleation zone. As nuclei start to form, the soluble protein concentration decreases as some of it transitions into the solid state, thereby shifting the system into the metastable zone where crystal growth occurs (A). Figure adapted from Sherwood & Cooper (2011).

Protein solubility is affected by numerous factors including temperature, buffer conditions, pH, and additives such as ligands (Kramer *et al.,* 2012). Predicting the optimal crystallisation conditions for a protein remains challenging, therefore this has to be assessed through a trial-and-error process. A number of commercial screens are available and routinely used for screening conditions manufactured by Hampton Research, Molecular Dimensions and Qiagen. The most common method of crystal screening is vapour diffusion, which is typically performed in two different arrangements: sitting drop and hanging drop (Delmar *et al.,* 2015) (Figure 5.2)

In the hanging drop method, a small volume of protein (~2-5 µL, 1.2 µL – 25 nL if using an automated Mosquito® crystal (Sptlabtech) pipetting) is dispensed on a coverslip and mixed with the screening condition. The coverslip, containing the screening condition, is then inverted and placed on top of a well. The concentration of the screening condition in the droplet is half that of the well, allowing water to slowly diffuse from the droplet to the well, gradually increasing the protein and precipitant concentration within the droplet. The sitting drop vapour diffusion method is based on the same principle, but the protein-condition drop rests on a "shelf" above the screening condition.

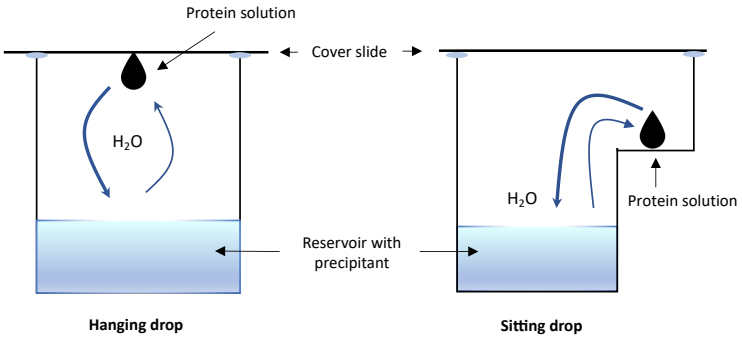


Figure 5.2. The process of Vapour Diffusion in hanging drop and sitting drop arrangements.

5.1.2. Diffraction

Crystalline structures are defined as repetitive and ordered structures. The regular spacing of atoms within the crystal acts as a three-dimensional grating. The significance of this is that X-ray waves scattered by specifically spaced macromolecules will combine in a definite manner and produce a precise diffraction pattern (Jauncey, 1924). X-ray diffraction can be explained by “Bragg’s law”, which is defined by $n\lambda = 2d \sin\Theta$, whereby, n - the order of diffraction, λ – incident wavelength, d – interplanar spacing of the crystal, Θ – incident angle.

The principle of Bragg’s law is demonstrated in Figure 4.3. Incident beams, wave 1 and wave 2, are assumed to be in phase. Once they pass through the crystal, wave 1 is reflected by the upper plane, and wave 2 is reflected from the lower plane. Since wave 2 has to travel a longer distance to be reflected, the emerging waves will have a different phase relationship. The path difference travelled by wave 2 is equal to $2d\sin\Theta$. If this number is an integer number of wavelengths, the waves interfere constructively (phases are equal), hence Bragg’s law is satisfied, and produce a spot on the X-ray detector.

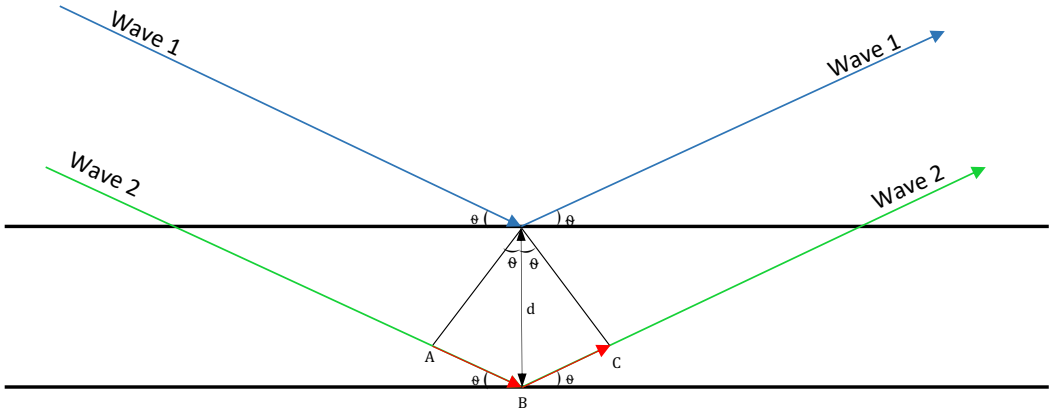


Figure 5.3. Bragg's law of diffraction.

Wave 1 and wave 2 are incident X-ray waves onto a set of crystal lattice planes, upon which their reflection introduces a phase difference. The path difference between wave 1 and wave 2 is AB+BC =2dsinΘ. If the path difference is an integer number, the phase relationship between the two waves will be constructive, subsequently producing a diffraction spot on the detector.

5.1.3. Phasing and crystal structure determination

The intensity of the diffraction spots is related to the features of the electron density in the unit cell by a mathematical relationship known as the Fourier transform. The main objective of X-ray crystallography is to reconstruct an electron density map using the complex Fourier transform synthesis equation. To satisfy the calculations, both magnitudes and corresponding phases need to be determined, however, experimentally, only the spot intensities are measured representing the infamous “Phase problem” (Hauptman, 1991)

The phase problem arises from technological limitations, rather than physical laws, including the coherence of the X-ray source and the speed of X-ray detectors (Sherwood and Cooper, 2011). Currently, the phase problem can be addressed either experimentally utilising the advantages of heavy atoms (HAs) or by molecular replacement.

Experimental phasing using HAs is based on the principle that HAs have higher atomic numbers. Consequently, crystals containing HAs will have a greater change in the diffraction pattern compared to the native crystal. HAs have to be attached to the protein in defined locations. This can be achieved either by soaking the crystal in a HA solution or substituting selenomethionine with methionine during recombinant expression. By comparing the diffraction patterns of crystals containing HAs and native crystals, the initial phase information can be obtained (Taylor, 2010)

The molecular replacement method is based on using a previously solved structure as a search model to determine the orientation and position of the molecule within the unit cell. Successful phase determination relies on high structural and sequence similarity between the search model and the target macromolecule (Evans and McCoy, 2008). Therefore, if studying a structure with no structural similarity, molecular replacement cannot be utilised.

Nevertheless, this limitations has been mitigated with the recent development of Alphafold (Jumper *et al.,* 2021). Alphafold models of the target macromolecule can often be used as search models (Barbarin-Bocahu and Graille, 2022).

Once the initial phases have been estimated, an iterative refinement process can be performed subsequently achieving the optimal final model. X-ray crystal structure refinement aims to obtain the best possible agreement between structural model and experimental data. Two measures, “R-work” and “R-free”, are used to estimate the model quality, or how well it fits the experimental data. R-work quantifies the agreement between calculated and experimental amplitudes. Using R-work alone, however, may lead to a false estimation of the model quality, as it is based on unweighted statistics and can result in overfitting (Shabalin *et al.,* 2018).

The R-free measure was introduced to minimise the effects of overfitting of the atomic model to the experimental data. The method involves randomly selecting a subset of reflections, usually around 5%, which are excluded from the refinement process providing an unbiased evaluation of the correlation between the map and the model (Wang, 2015).

5.2. Chapter aims.

Whilst cryo-EM allowed us to obtain a high-resolution structure of the Bd1152^{2C} top MCE ring, we were unable to resolve the PLLs and the bottom MCE ring, most likely due to inherent continuous flexibility. Considering the crystal structures of the *E.coli* MlaD protein (Ekiert *et al.,* 2017) and the LetB MCE 2-3 rings (Isom *et al.,* 2020), we hypothesised that X-ray crystallography could be a successful approach in probing the complete structure of the Bd1152 protein. This chapter further discusses the workflow of Bd1152 crystallisation, which resulted in a high-resolution structure.

5.3. Results

5.2.1. Crystallisation of Bd1153-Bd1152, Bd1152$^{37-283}$ and Bd1152^{2C}

The Bd1153-Bd1152, Bd1152$^{37-283}$ and Bd1152^{2C} constructs were expressed and purified as described in Chapters 2 and 4. The purified membrane-associated Bd1153-Bd1152 protein complex and Bd1152$^{37-283}$ were prone to aggregation and not amenable to concentrating; therefore, crystal trays were set up at concentrations ranging from 0.03 to 0.3 mg/mL. Whilst several low-quality crystals were obtained, the diffraction data were inadequate for interpretation.

We initially generated the Bd1152^{2C} construct with the aim of improving protein rigidity for cryo-EM investigation (refer to Chapter 4). Nevertheless, this construct proved significantly more stable and amenable to spin concentration. Consequently, we were able to express high yields of Bd1152^{2C} and set up crystal trays at concentrations ranging from 5 to 30 mg/mL. While we obtained numerous crystals in a variety of different conditions (Figure 5.4), none of these diffracted beyond 10 Å.

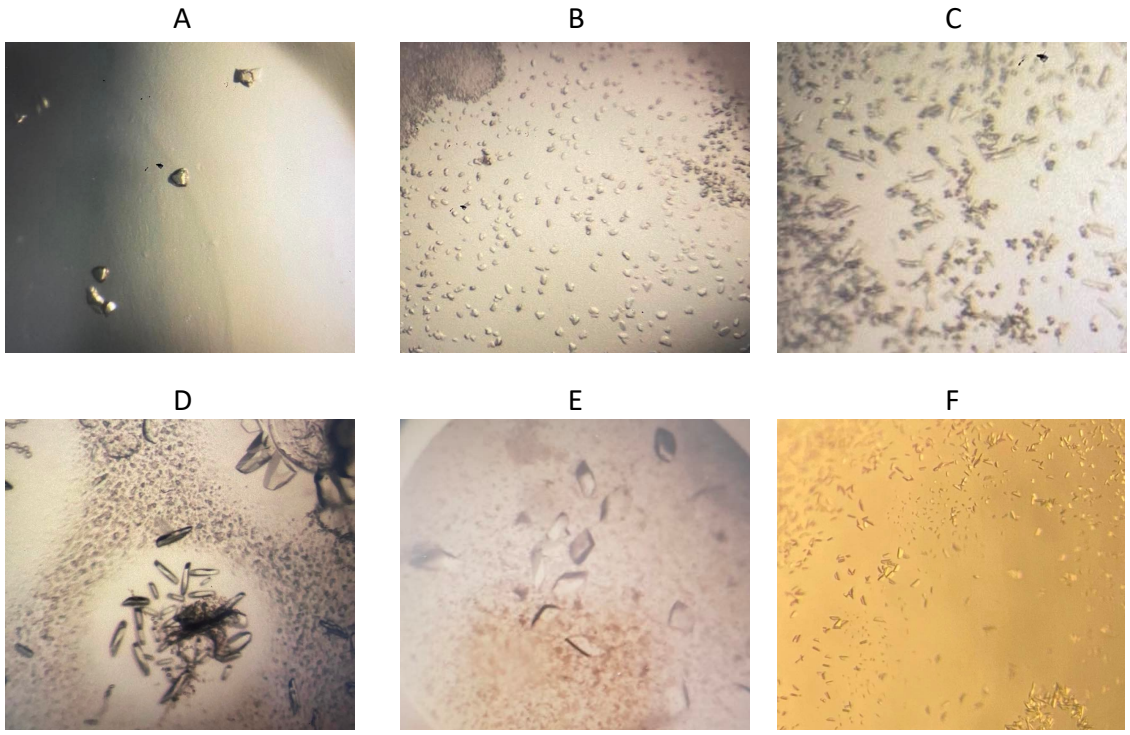


Figure 5.4. Examples of Bd1152^{2C} crystals under different crystallisation conditions. A) 5.2 mg/mL; 0.1 M HEPES pH 6.5, 45% w/v poly (acrylic acid sodium salt) 2100 B) 5.2 mg/mL; 10% w/v polypropylene glycol 400; C) 10 mg/mL; 0.2 M sodium chloride, 0.1 M MOPS pH 7.5 38% v/v PEG400; D) 10 mg/mL; 0.1M MOPS pH 7.0, 9% w/v PEG8000 E) 25 mg/mL 0.05 M sodium sulphate ; 0.05 M lithium chloride, 0.05 M Tris pH 8.5, 32% v/v PEG400) 8 mg/mL; 0.225 M ammonium sulphate, 0.05 M sodium acetate, pH 4.0, 12% v/v PEG4000.

5.2.2. The Bd1152 construct optimisation for crystallisation.

We anticipated that the poor diffraction observed from the $Bd1152^{2C}$ crystals arose due to nonoptimal crystal packing. During the cryo-EM data processing of the $Bd1152^{2C}$ sample, a small proportion of particles were observed to dimerise in a head-to-head fashion (Figure 5.5). Prof Andrew Lovering suggested such behaviour could potentially translate into crystal formation and serve as a bottleneck in obtaining high-resolution diffraction data.

The Alphafold structure of Bd1152 revealed the presence of two hydrophobic residues, Leu228 and Phe229, located at the tip of the pore lining loop of the second MCE-domain. These residues were identified as potential contributors to the anticipated head-to-head crystal packing through hydrophobic interactions, leading to the formation of nonoptimal crystals. Consequently, we substituted Leu228 and Phe229 with alanine residues. This resulted in the generation of the $Bd1152^{37\text{-}283\ I112C\ L238C\ L228A\ F229A}$ construct, hereafter referred to as $Bd1152^{2C2A}$. It is noteworthy, that the location of mutations is based on the Bd1152 Alphafold model, and that the regions where the L228 and F229 residues are located have low pLDDT scores (70 > pLDDT > 50).

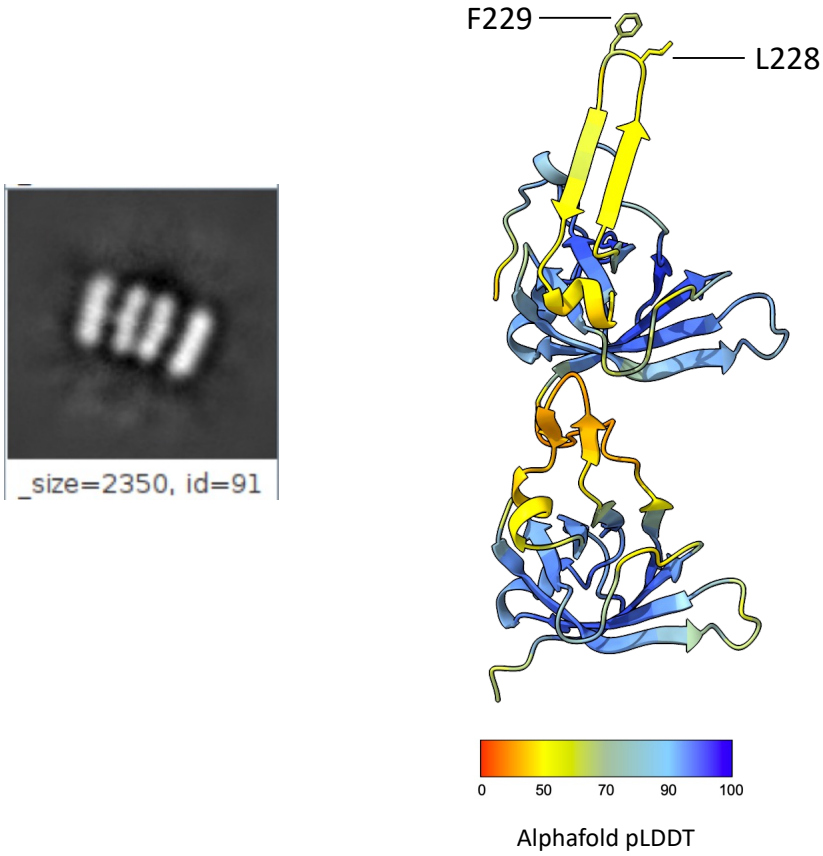


Figure 5.5. Head-to-head dimerisation of Bd1152^{2C}.

(Left) 2D class average obtained from the cryo-EM data processing of Bd1152^{2C}, revealing head-to-head dimerisation. (Right) Alphafold structure of Bd1152 (residues 37-283) highlighting the residues F229 and L228, which are predicted to contribute to the observed head-to-head dimerisation. There residues were subsequently substituted with alanine for crystallography studies. The model is coloured by per-residue pLDDT scores. A pLDDT colour key is provided at the bottom right of the figure where the spectrum from blue to red corresponds to decreasing pLDDT score (Jumper *et al.,* 2021). The model was visualised with ChimeraX (Pettersen *et al.,* 2021).

5.2.3. Expression and purification of the Bd1152^{2C2A} construct.

The Bd1152^{2C2A} construct was expressed from an IPTG inducible pET26b(+) plasmid. Bd1152^{2CA} was isolated from clarified cell lysate through IMAC purification. The IMAC flow-through, wash and elution fractions were analysed via SDS-PAGE (Figure 5.6, A), which revealed a prominent band at ~31 kDa, corresponding to Bd1152^{2C2A}. Elution fractions containing Bd1152^{2CA} were combined and further purified with a Superdex 200 10/300 Increase size exclusion column, resulting in two peaks: the first peak (eluted at ~8 mL) represented aggregated protein and was discarded, while the second peak eluted at ~12 mL. (Figure 5.6, B). The elution fractions from the second peak were analysed via SDS-PAGE and revealed a prominent band at ~31 kDa, correspond to Bd1152^{2C2A} (Figure 5.6, C). A higher molecular weight bands were also present likely corresponding to the oligomerisation of Bd1152^{2C}. All of the resulting fractions were pooled and concentrated for crystallisation.

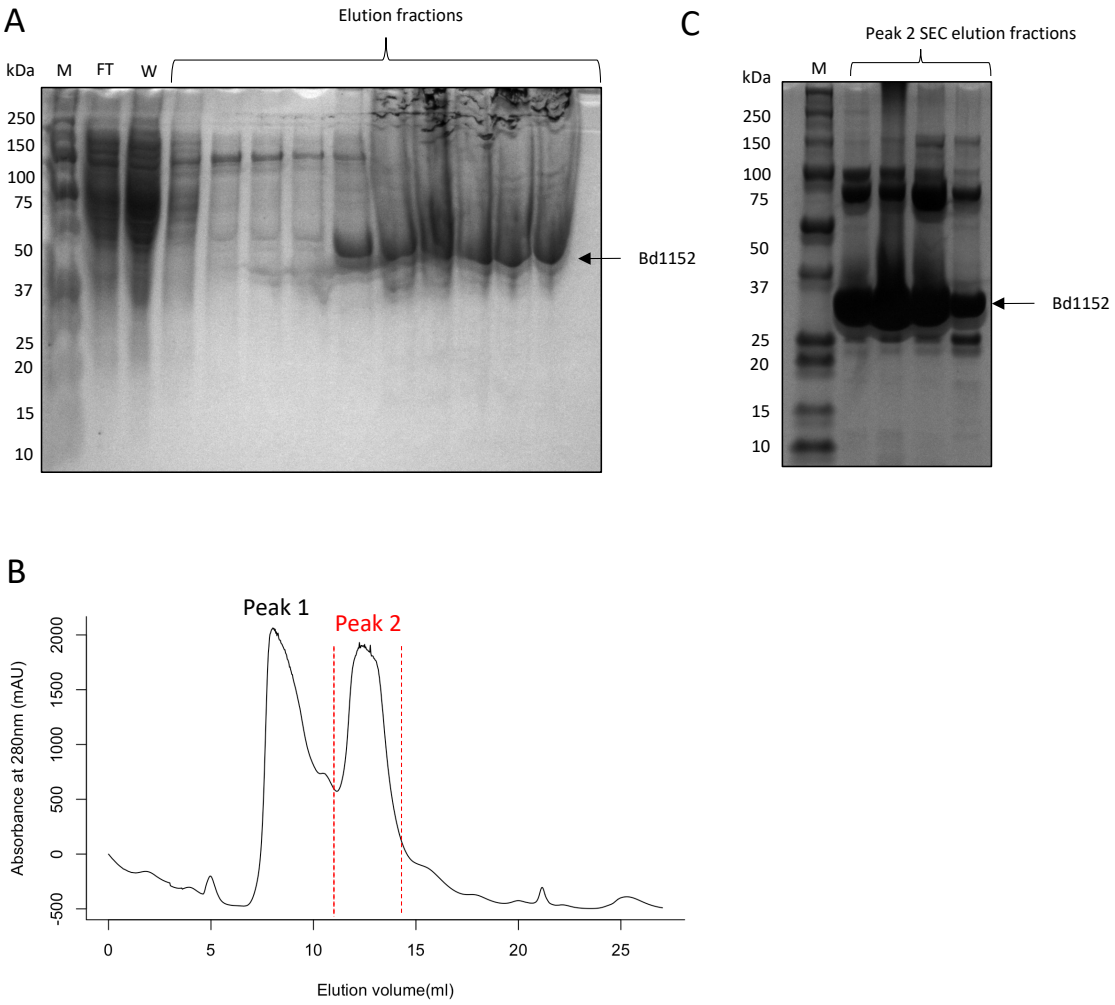


Figure 5.6. Purification of Bd1152^{2C2A} .

A) SDS-PAGE analysis of the Bd1152^{2C2A} IMAC purification showed a prominent band at ~31 kDa. M - molecular marker, FT –flow-through, W – 50 mM imidazole wash. B) SEC (Superdex 200 10/300) revealed two peaks, eluting between 6-10 mL and 11-15 mL, respectively. (C) SDS-PAGE of the elution fractions from the second peak revealed a prominent band at ~31 kDa, corresponding to Bd1152^{2C}. SDS-PAGE was performed using 4-12% bis-tris gels.

5.2.4. Structural determination of Bd1152^{2C2A}

The purification of Bd1152^{2C2A} resulted in a highly pure (Figure 5.7) and stable protein. The protein was amenable to concentrating, allowing us to proceed with screening crystallisation conditions. A hit was observed in a 2 µL sitting drop well comprising a 1:1 ratio of 5 mg/mL Bd1152^{2C2A} and 0.015 M Tricine, pH 8.5, 24% w/v PEG 4000 (Figure 5.7), kindly prepared by Fiona Li during a 3-month master's project under my supervision.

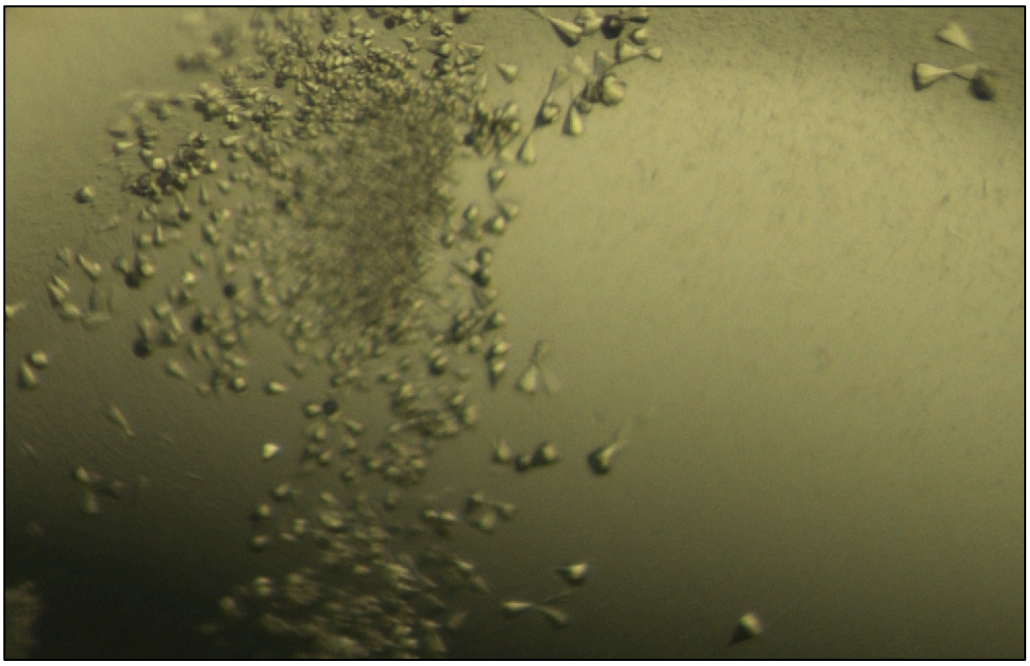


Figure 5.7. Crystals that resulted in the Bd1152^{2C2A} structure.

Crystals obtained via sitting drop vapor diffusion in 0.015 M Tricine, pH 8.5, 24% w/v PEG 4000.

The crystals yielded a 2.6 Å resolution dataset and belonged to the H3-centred trigonal space group. Mathew's coefficient probabilities suggested the presence of two protein copies within the asymmetric unit, therefore, the data were phased using two copies of the Alphafold Bd1152 (residues 37-283) prediction as a molecular replacement model. Nevertheless, the top solution possessed a Log Likelihood Gain (LLG) of 16.6 and Translation Function Z score (TFZ) of 5.7, suggesting that the phasing was not successful.

Next, we trimmed the Bd1152 Alphafold model to only contain the 1st MCE domain (residues 37-157) and used it as the search model for molecular replacement. This resulted in an improved solution with a LLG of 142.51 and TFZ score of 12.9. We were not able to use the Bd1152 Alphafold model to phase the second MCE-domain (residues 157-283); therefore, it was manually built during the refinement process.

Refinement yielded a final model with R_{work}/R_{free} of 0.2754 and 0.3334, respectively. The refined model of *B. bacteriovorus* $Bd1152^{2C2A}$ consists of two molecules (chain A and chain B) in the asymmetric unit. Observing symmetry copies revealed that $Bd1152^{2C2A}$ crystalised in a hexameric state. This oligomeric state was expected, as a hexameric assembly was also observed in our cryo-EM maps (refer to Chapter 4) and other MCE domain-containing proteins (MlaD, LetB, PqiB) (Ekiert *et al.,* 2017; Isom *et al.,* 2020). All data collection and refinement statistics are displayed in Table 5.1.

The core density of both the top and bottom MCE domains was unambiguous, with Cα and side chain information present, allowing for accurate modelling of the canonical seven β-strands of each MCE domain (Figure 5.8) (Figure 5.9). In chain A and B, 243 and 247 (out of 289 residues), respectively, were built. Both chain A and chain B should contain pore lining loops projecting from the first and second MCE domains, located in the central portion of the hexameric assembly as predicted by the Alphafold model. Nonetheless, no density was observed between residues 107–123; 219-243 in chain A and 106-119; 219-243 in chain B.

Unfortunately, all of the introduced mutations (I112C, L238C, I228A, F229A) are located within the PLL regions for which there was no visible density, hence were unable to determine whether the disulphide bond was formed between the 1st and 2nd MCE domains. The absence of electron density at the PLL regions suggest their inherent flexibility and "fluid"- like nature, further supporting the dynamic nature of the PLLs observed during the cryo-EM data processing. The overall architecture of Bd1152^{2C2A} comprises two tandem hexameric assemblies (MCE rings) that feature a central channel. Nevertheless, the precise dimensions of the channel could not be assessed as the PLLs were not modelled.

Both MCE rings display the typical seven-stranded beta barrel fold and are stacked on top of each other on a slight angle connected by a short, disordered linker of 10 residues. While residues 1-37 have been deleted from this construct, *in-vivo* the n-terminus is expected to form a transmembrane helix that anchors Bd1152 to the IM. Finally, the c-terminus shows a disordered loop that folds back over the top MCE domain (discussed further in Chapter 6) (Figure 5.8).

It is worth emphasising that the reported R_{work}/R_{free} values of 0.2754 and 0.3334 appear somewhat high for a 2.6 Å resolution data (personal communication with Prof Stephen Smerdon, 2023). It may be prudent to consider additional data processing, as there is a possibility of a twinning effect in the crystal structure (Yeates, 1997), which, if addressed, could potentially lead to an enhanced model.

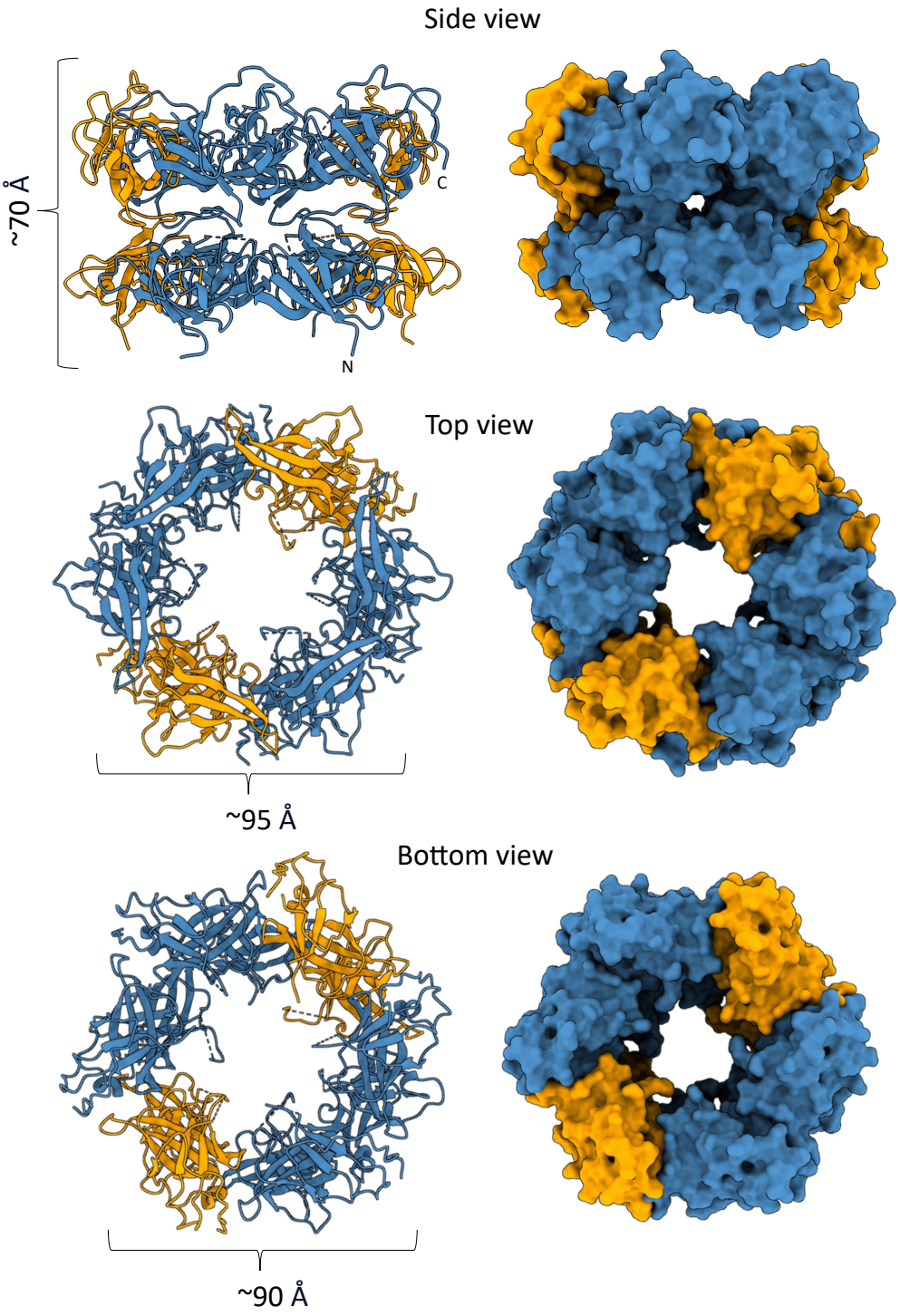


Figure 5.8. Crystal structure of the Bd1152^{2C2A} hexamer.

The structure is displayed in ribbon (left) and surface (right) representations. Orange colour indicates the asymmetric unit, whilst steel blue colour indicates the symmetry copies. N and C represent n- and c- terminus of the Bd1152 chain. The model was visualised with ChimeraX (Pettersen *et al.,* 2021).

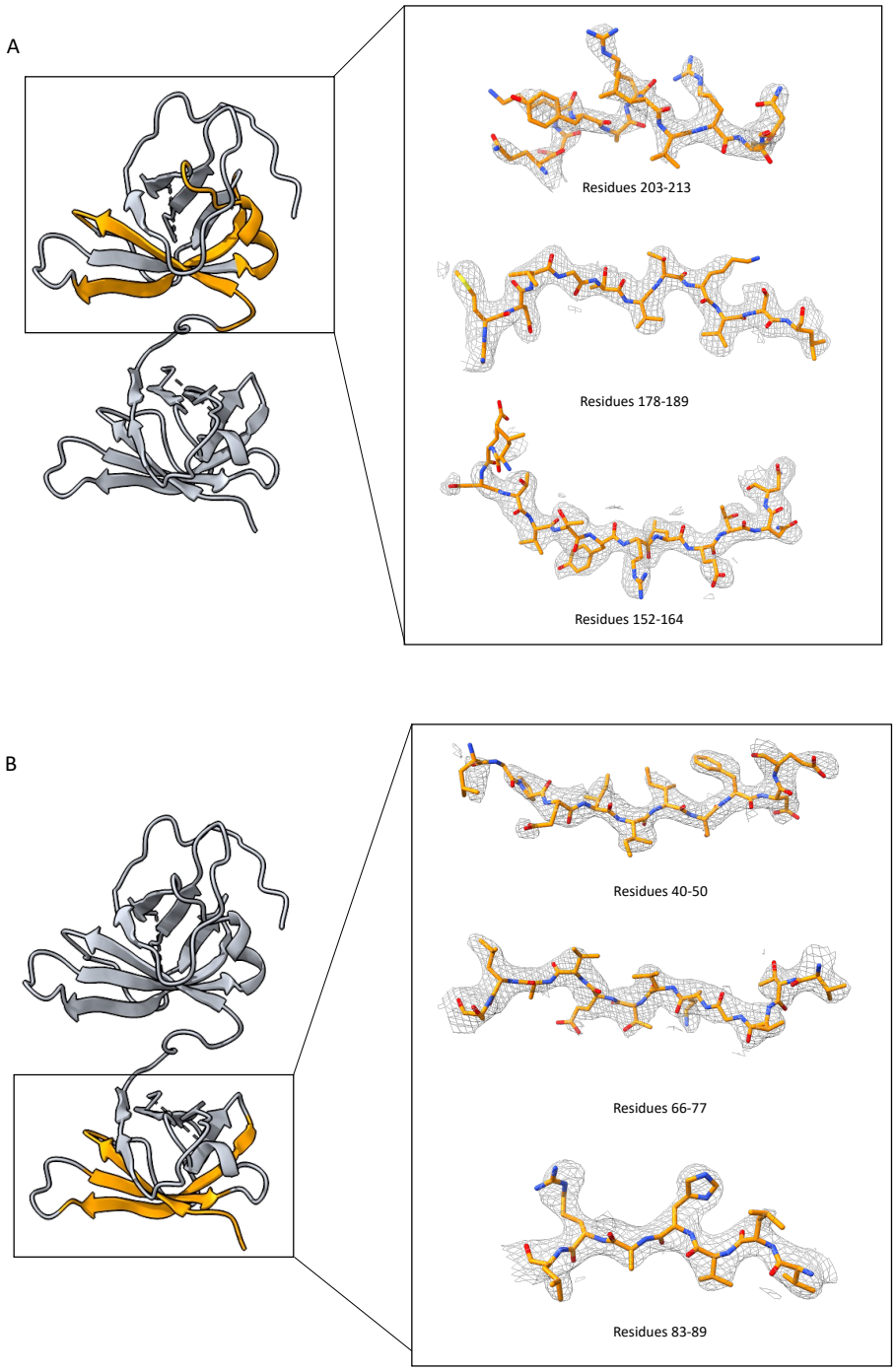


Figure 5.9. The electron density in the final refined map for the Bd1152^{2C2A} top (A) and bottom (B) MCE domains.

The chain A of Bd1152^{2C2A} structure is displayed on the left (grey) with regions, for which the electron density is shown on the right, highlighted in orange. The 2mFo-DFc map (grey) is contoured at 1 sigma and protein is displayed as stick representation. The model and corresponding refined map were visualised with ChimeraX (Pettersen *et al.,* 2021).

Table 5.1. Bd1152^{2C2A} structure data collection and refinement.

	Bd1152^{2C2A}
Wavelength	0.9677
Resolution range	36.8 - 2.632 (2.726 - 2.632)
Nmol in AU	2
Space group	R 3: H
Unit cell	142.139 142.139 60.129 90 90 120
Total reflections	62507 (3210)
Unique reflections	12138 (341)
Multiplicity	5.1 (5.3)
Completeness (%)	90.33 (25.11)
Mean I/sigma(I)	10.3 (1.2)
Wilson B-factor	71.19
R-merge	0.122 (1.476)
R-meas	0.124 (1.641)
R-pim	0.055 (0.713)
CC1/2	0.9978 (0.413)
Reflections used in refinement	12138 (341)
Reflections used for R-free	599 (17)
R-work	0.2653 (0.3908)
R-free	0.3468 (0.4427)
Number of non-hydrogen atoms	3228
Macromolecules	3228
Ligands	0
Solvent	0
Protein residues	413
RMS(bonds)	0.005
RMS(angles)	0.65
Ramachandran favoured (%)	87.22
Ramachandran allowed (%)	8.77
Ramachandran outliers (%)	4.01
Rotamer outliers (%)	2.01
Clashscore	8.47
Average B-factor	61.15
Macromolecules	61.15

Values in parenthesis correspond to the highest resolution shell

We next compared the Bd1152 Alphafold model with the final refined model of $Bd1152^{2C2A}$. The Bd1152 Alphafold model was trimmed to reflect the number of residues present in chain A of the $Bd1152^{2C2A}$ crystal structure. The alignment of these two models yielded an RMSD of 1.1145 Å across 83 pruned atom pairs and 3.590 Å across all 194 pairs. This reflects a large difference in the positioning of the lower MCE domain, with the X-ray model showing a position approximately 2 Å higher than the corresponding domain in the Alphafold model (Figure 5.10). Nevertheless, if the two MCE domains are aligned separately (residues 40 - 144 and 155-283, for the top and bottom MCE domains, respectively) they yield an RMSD of 0.580 Å across 87 pruned atom pairs (2.852 Å across all 96 pairs) and an RMSD of 0.612 Å across 86 pruned atom pairs (0.665 Å across all 87 pairs), for the top and bottom MCE domains, respectively (Figure 5.10).

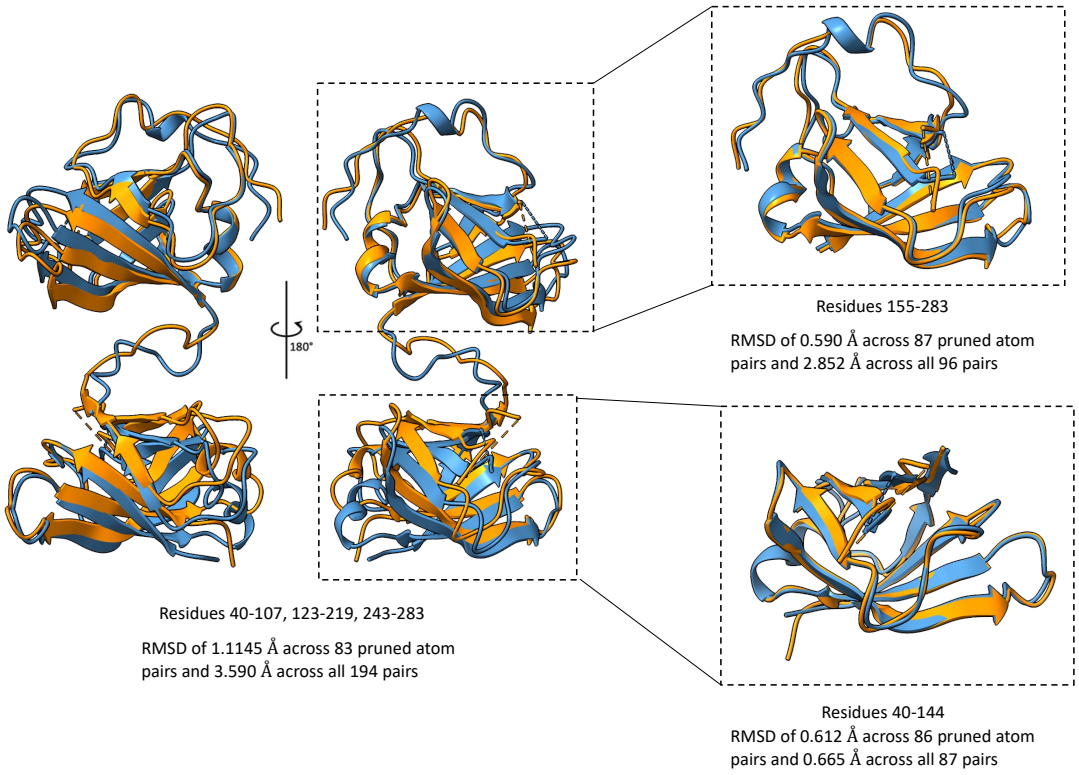


Figure 5.10. Comparison of the Bd1152 Alphafold model and the Bd1152^{2C2A} crystal structure.

Alignment of the Bd1152 Alphafold model (steel blue) and the Bd1152^{2C2A} crystal structure (dark orange) monomers and their individual MCE domains. The Alphafold model was trimmed to reflect the residues present within the Bd1152^{2C2A} crystal structure. The residue numbers and RMSD scores are depicted in the figure. The models were visualised with ChimeraX (Pettersen *et al.,* 2021).

5.2.5. Lipophilicity of Bd1152 and predicted key residues for lipid binding.

We next investigated the molecular lipophilicity potential of Bd1152. As we were not able to resolve the PLLs, we utilised the Bd1152 Alphafold model for the analysis (Figure 5.11, B). The PLLs encompass highly hydrophobic residues, including I112, I115, L118, L121, I122, L226, L228, F229, L233, L238, L241, and L242. These residues are predicted to collectively assemble into a hydrophobic channel that could potentially transport hydrophobic substrates such as lipids through its dynamic nature. Notably, *E. coli* MCE domain-containing proteins LetB and PqiB feature a conserved hydrophobic motif, ΦxxΦΦ, within their PLLs (were Φ - denotes a hydrophobic residue and x-represents any residue) (Ekiert *et al.,* 2017; Isom *et al.,* 2020). The same ΦxxΦΦ hydrophobic motif is observed in the sequence of Bd1152 and its corresponding Alphafold model within the predicted PLL regions (L118 X X L121 I122; L238 X X L241 L242).

To gain further insights, we employed ConSurf (Ashkenazy *et al.,* 2016) to analyse the sequence conservation of the residues of interest. The results revealed that the hydrophobic motif of the first MCE domain is highly conserved (conservation score 8-9, where 9 indicated maximum conservation) (residues L118 and L121, I122), however, the hydrophobic motif of the second MCE domain is not highly conserved (L238, L241, L242) (conservation score 5-6) (Figure 5.12).

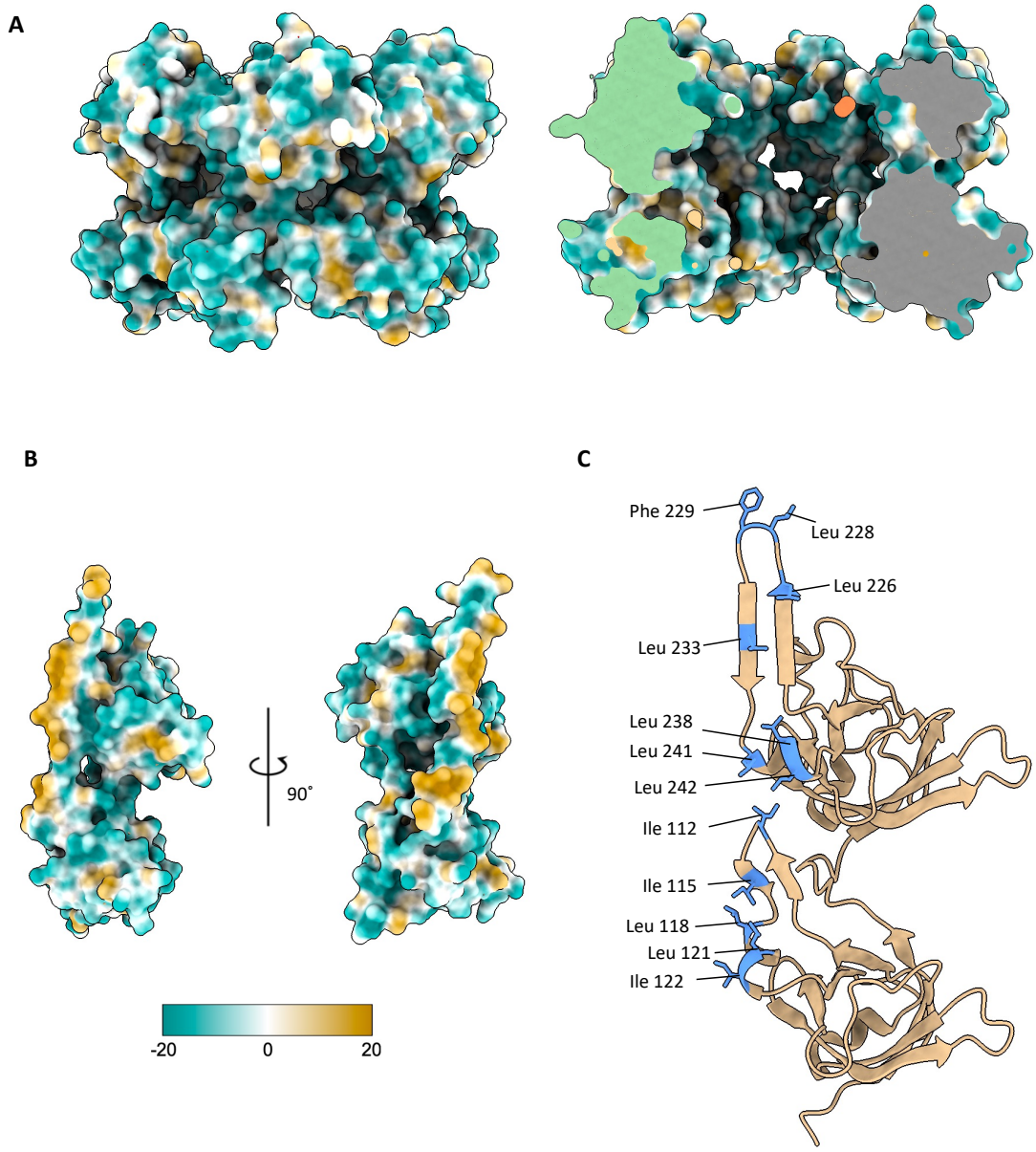


Figure 5.11. Molecular Lipophilicity Potential of Bd1152^{2C2A}.

(A) The X-ray structure of Bd1152^{2C2A} displayed in a surface representation. Views: (top left) outer surface, (top right) inner surface. B) The Alphafold model of the Bd1152 (residues 37-283) monomer displayed in a surface representation. A and B structures are coloured according to the lipophilicity, with hydrophilic and lipophilic residues indicated in dark cyan and yellow, respectively. A colour key is provided at the bottom of the figure. C) Alphafold model of the Bd1152 (residues 37-283) monomer, displayed in a ribbon representation, with the predicted central pore lining hydrophobic residues highlighted in blue and labelled accordingly. The models were visualised with ChimeraX (Pettersen *et al.,* 2021).

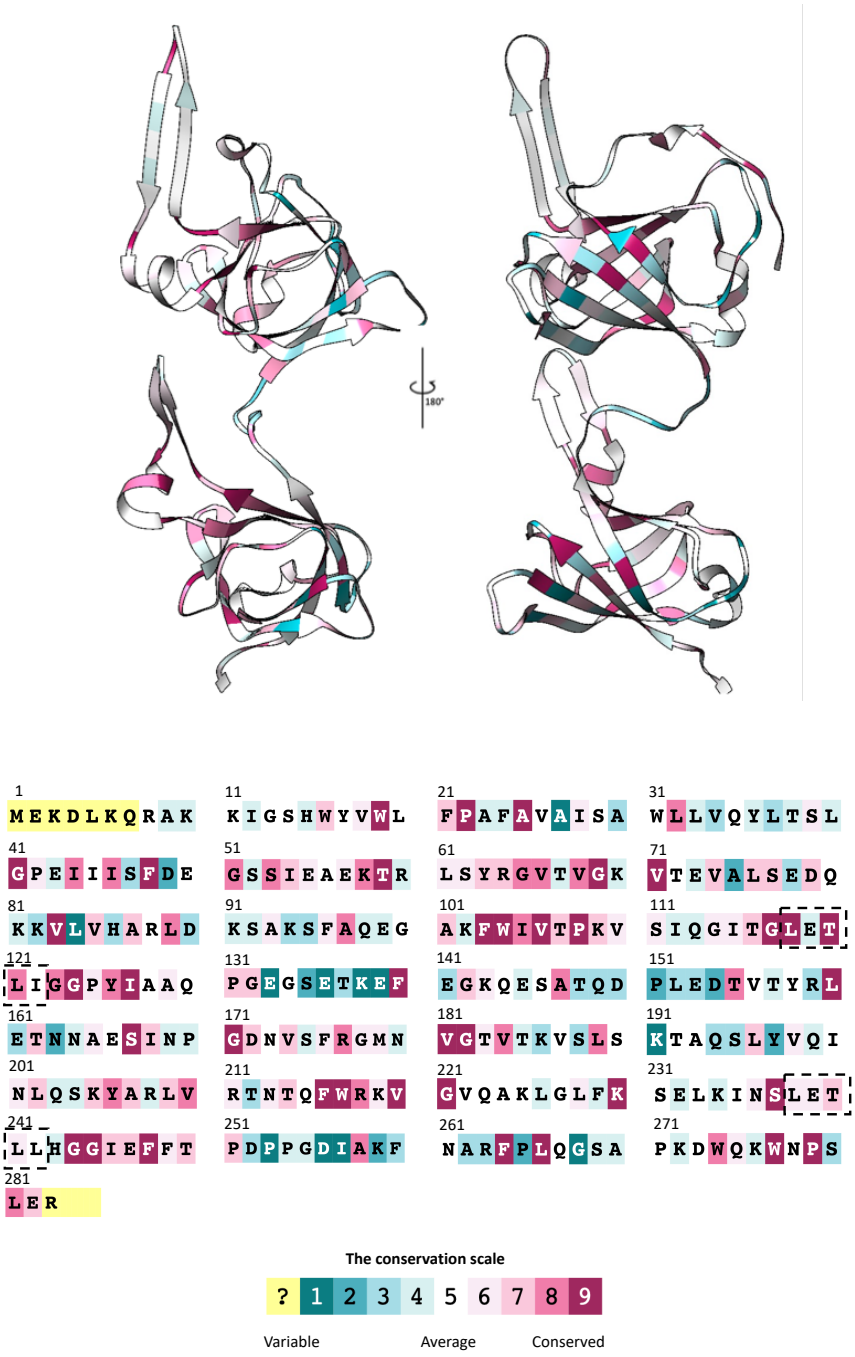


Figure 5.12. Sequence conservation of Bd1152.

(Top) The Alphafold model of Bd1152 (residues 37-283). The Alphafold model was coloured according to ConSurf sequence conservation analysis. (Bottom) The Bd1152 sequence coloured according to the ConSurf sequence conservation analysis. The black dotted box highlights the hydrophobic motif - ΦxxΦΦ (Φ - denotes a hydrophobic residue; x-denotes any residue). The conservation score colour scale is depicted at the bottom of the figure. The model was visualised with ChimeraX (Pettersen *et al.,* 2021).

5.1.3. Comparison between X-ray and cryo-EM structures of Bd1152.

The cryo-EM and X-ray models of the top MCE domain (residues 157-283) are nearly identical with an overall root mean square difference (RMSD) of 0.608Å across 94 pruned atom pairs (across all 95 pairs: 0.677Å). Only the c-terminus shows a larger deviation between the two structures suggesting its inherent flexibility. Both structures revealed that Bd1152 adopts a typical MCE fold with seven-stranded beta barrels and six copies of Bd1152 assemble to form a hexameric ring (Figure 5.13).

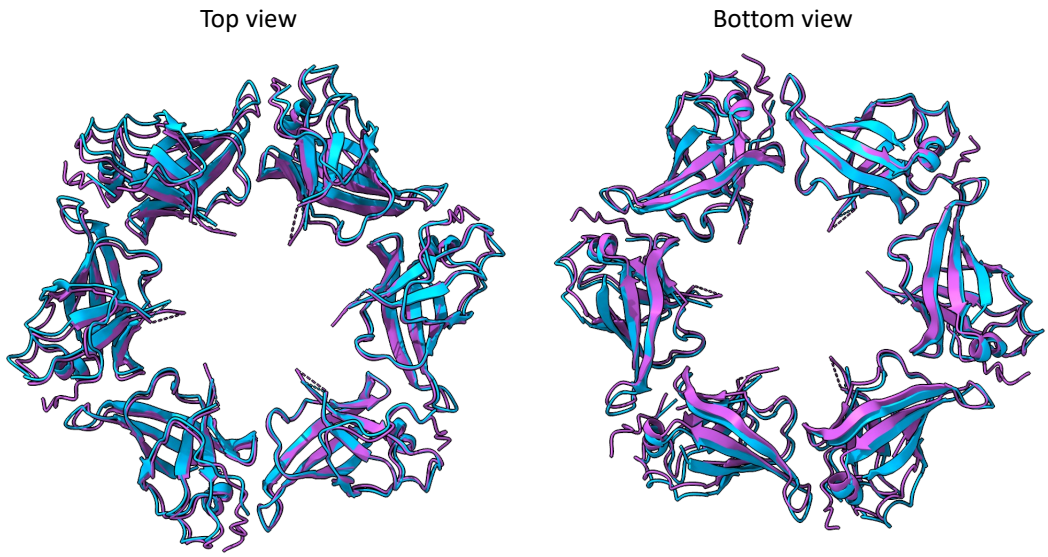


Figure 5.13. Structural comparison of the Bd1152 top MCE ring obtained from X-ray crystallography and cryo-EM.

Alignment of the top MCE domain of Bd1152^{2C} cryo-EM model (blue) and Bd1152^{2C2A} X-ray model (purple) (residues 157-283). RMSD of 0.608Å across 94 pruned atom pairs and 0.677Å across all 194 pairs. The model was visualised with ChimeraX (Pettersen *et al.,* 2021).

5.3. Discussion

Whilst cryo-EM allowed us to obtain a high-resolution map of the Bd1152^{2C} top MCE ring, we were unable to reconstitute a complete Bd1152^{2C} model. Consequently, we utilised X-ray crystallography as an alternative approach. In this chapter, we presented the Bd1152^{2C2A} structure at 2.6 Å resolution.

5.3.1. Structural comparison of the Bd1152 and *E. coli* PqiB MCE rings.

The top and bottom MCE rings of Bd1152^{2C2A} displayed an identical MCE domain architecture, composed of 7-stranded beta barrels. Notably, the MCE rings of Bd1152^{2C2A} displayed high structural similarity to the *E. coli* PqiB protein. However, a noticeable distinction between the two proteins lies in the relative positioning of the MCE domains (Figure 5.14). When aligning the Bd1152^{2C2A} model with the *E. coli* PqiB model, only the bottom MCE rings displayed satisfactory structural alignment, while the second MCE ring of both proteins exhibited misalignment due to Bd1152^{2C2A} adopting a different positioning. This difference appears to be mainly attributed to the shorter inter-domain distance between MCE ring 1 and ring 2 in Bd1152^{2C2A} compared to corresponding MCE rings in the PqiB structure. Nevertheless, when aligning the bottom MCE rings of both proteins separately from the rest of the structure, a reasonable structural alignment was observed. These findings highlight the presence of unique conformational arrangements within Bd1152^{2C2A}, despite its shared structural motifs. Nevertheless, the biological significance of the unique positioning of the Bd1152^{2C2A} MCE domains is currently unknown. It is plausible that this observation may be attributed to crystal packing effects.

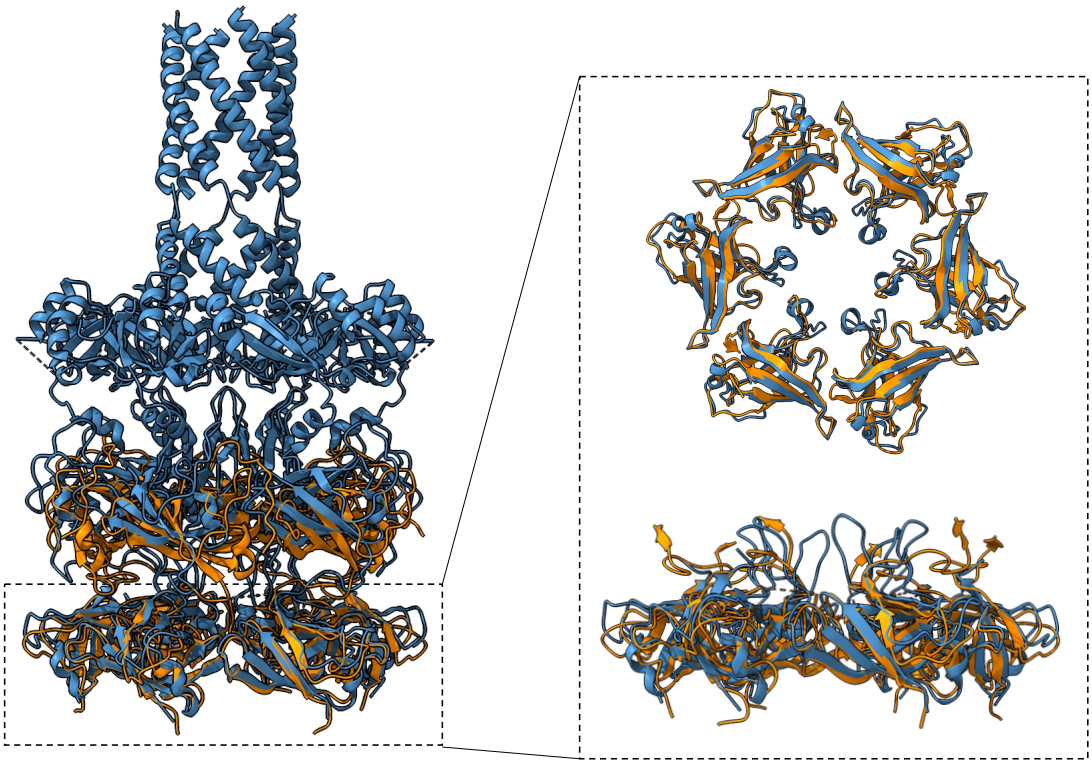


Figure 5.14. Structural comparison of *E. coli* PqiB and Bd1152^{2C2A}.

(Left) Alignment of the Bd1152^{2C2A} (displayed in orange) and *E. coli* PqiB (PDB 5UVN) (displayed in blue) models. RMSD between 198 pruned atom pairs is 1.228 Å; (across all 564 pairs: 37.885 Å). (Right) Alignment of the 1st MCE rings of Bd1152^{2C2A} (residues 37-154) and *E. coli* PqiB (PDB 5UVN) (residues 40 - 155) models. RMSD between 189 pruned atom pairs is 1.214 Å; (across all 1230 pairs: 41.270 Å). The models were visualised with ChimeraX (Pettersen *et al.,* 2021).

5.3.2. Strategies to overcome the challenges of Bd1152 crystallisation.

Crystallisation of Bd1152 proved to be challenging as we failed to obtain high-quality protein crystals for the Bd1153-Bd1152, Bd1152$^{37-283}$ and Bd1152^{2C} constructs. In our efforts to overcome this issue and achieve successful protein crystallisation, we employed two distinct experimental strategies:

5.3.2.1. The impact of PL species on the Bd1152 stability and crystallisation

As demonstrated in chapter 4, Bd1152$^{37-283}$ co-purifies with PG and PE. The presence of mixed PL species within the protein may potentially impede protein crystallisation. The interaction of Bd1152 with the differently charged head groups of the PLs can potentially induce conformational changes in Bd1152, thereby introducing heterogeneity in the sample and posing challenges for protein crystallisation.

The impact of different PL species within the protein on crystallisation has been previously proposed by Dr Peter Wotherspoon, who attempted to remove co-purifying PLs from the *E. coli* MlaD protein, aiming to load it with a PL single species, with the intention of improving crystal diffraction data (Personal communication with Dr. Peter Wotherspoon).

We aimed to employ a similar approach and proceeded with the removal of co-purifying PL from the Bd1152$^{37-283}$ protein using a lipid wash method described by Hughes *et al.,* (2019). However, upon performing the PL wash, we observed immediate protein aggregation (data not shown), preventing us from proceeding with the subsequent loading of a single PL species. This outcome strongly suggests that the presence of PLs within the Bd1152$^{37-283}$ construct is crucial for maintaining the integrity of the protein. Alternatively, it is plausible that the high-concentration detergent lipid wash employed in our study was too harsh for the

Bd1152 and may require further optimisation of the protocol. Further investigations may be necessary to identify the optimal conditions for removing co-purifying PLs without compromising the stability of the $Bd1152^{37-283}$ protein.

Additionally, as demonstrated in Chapter 4, the $Bd1152^{2C}$ construct does not co-purify with lipids, likely due to introduced cysteine mutations. This finding, along with the previous observation, implies that the presence of cysteine mutations and a potential disulphide bridge may play a role in stabilising the lipid-free Bd1152 construct. It is noteworthy, that the lipid-free $Bd1152^{2C}$ produced numerous protein crystals, suggesting that the presence of multiple PL species may indeed hindered the crystallisation process. Nevertheless, despite the promising morphology of the crystals, the diffraction data obtained was of low-resolution.

5.3.2.2. The role of hydrophobic residues in facilitating the crystallisation of Bd1152.

The analysis of cryo-EM data has revealed the presence of a small fraction of particles exhibiting a head-to-head dimerization. This observation might be significant in the context of crystallography, as it suggests an occurrence of potentially non-optimal head-to-head crystal packing. To address this issue, we introduced mutations in the hydrophobic residues L228 and F229, which are predicted to be located at the tip of the PLL projecting outward from the second MCE domain. These mutations successfully facilitated the crystallisation of $Bd1152^{2C2A}$ and led to the determination of a final model at a resolution of 2.6 Å. While it remains uncertain whether $Bd1152^{2C}$ originally packed in a head-to-head manner, an examination of the molecule arrangement within the crystal lattice revealed the head-to-tail packing arrangement of $Bd1152^{2C2A}$ (Figure 5.15). This observation suggests that the mutations introduced in the hydrophobic residues L228 and F229 may have facilitated head-to-tail packing observed in our crystal (Figure 5.15)

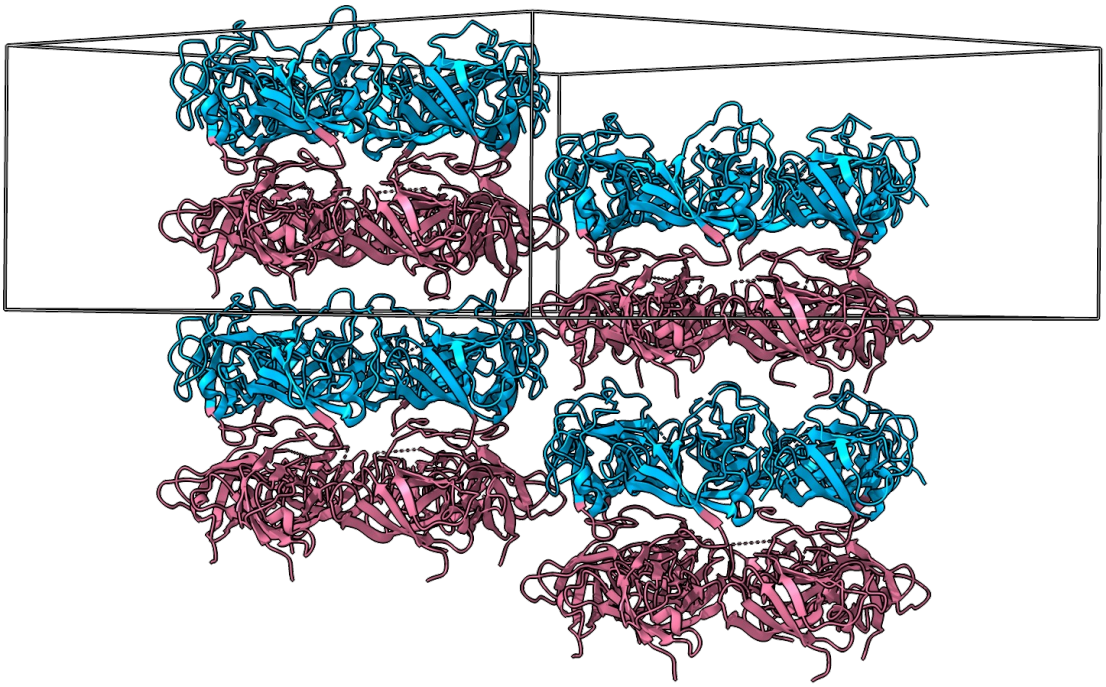


Figure 5.15. Bd1152^{2C2A} arrangement within the crystal lattice showing head-to-tail arrangement.

Bounding black box represents a unit cell. Bd1152^{2C2A} coloured based on the MCE rings (pink colour – 1st MCE ring; blue colour - 2nd MCE ring). The Bd1152^{2C2A} arrangement was visualised with ChimeraX (Pettersen *et al.,* 2021).

The significance of side chains at the protein-protein crystal packing interface has been reported in scientific literature. For instance, it has been shown that polar side chains, such as lysine and glutamate, may exert an unfavourable effect on the crystallisation process (Derewenda, 2004). In contrast, Yamada et al. (2007) demonstrated that the incorporation of leucine residues established a new interface for crystal packing, through hydrophobic interactions, leading to the successful crystallisation of RNase1. Nonetheless, in the case of the Bd1152^{2C} construct, it is plausible that the presence of Leu228/Phe229 residues could have hindered the production of high-quality crystals.

5.3.3. Flexibility and conservation of the Bd1152 PLL regions.

Whilst we were able to solve both MCE rings via X-ray crystallography, unfortunately, no density was observed for the PLL regions likely arising from high degree flexibility. Similar patterns were observed in the cryo-EM map of $Bd1152^{2C}$ and the X-ray structure of LetB MCE2-MCE3 rings (Isom *et al.,* 2020), suggesting that PLLs do indeed adopt a "fluid "- like nature and that this dynamic feature might be highly conserved among MCE domain-containing proteins of various bacterial species.

Finally, the lipophilicity analysis of the Bd1152 Alphafold model revealed that the PLLs consist of a collection of highly hydrophobic residues that may form a hydrophobic channel potentially facilitating the transport of hydrophobic substrates, such as lipids. The importance of PLLs and hydrophobic motif in the function of the *E. coli* PqiB and LetB proteins was demonstrated through *in vivo* complementation assays. Cooper (2023) and Isom *et al.,* (2020) revealed that constructs, possessing single point mutations in the hydrophobic ΦxxΦΦ motif, failed to restore the PqiB and LetB knockout sensitivity on the LSB phenotype.

Although complementation assays with Bd1152 are not currently feasible due to the intricate nature of genetic modifications within *B. bacteriovorus* (personal communication with Prof Andrew Lovering), we pursued an alternative approach to gain insights into the functional importance of hydrophobic residues within the $Bd1152^{37-283}$ ΦxxΦΦ motif. Our objective was to identify critical residues involved in PL binding through TLC analysis. Unfortunately, our attempts to introduce single point mutations in the hydrophobic motif were met with significant challenges. The mutant variants displayed pronounced instability and a strong propensity to aggregate, thus impeding further investigation (data not shown). Nevertheless, these results underscore the significance of hydrophobic interactions in maintaining the integrity of Bd1152.

5.4. Conclusion

Our initial attempts at crystallisation were unsuccessful. Consequently, we introduced mutations to the hydrophobic residues at the predicted apex of the protein, which were believed to influence crystal packing. This resulted in the successful determination of the high-resolution structure of $Bd1152^{2C2A}$, providing the first insight into the structure of a 2-MCE domain protein.

The structure of Bd1152 revealed the adoption of a typical MCE-fold, showing high structural similarity to the *E. coli* PqiB protein. We additionally identified hydrophobic residues within the PLL regions that may play a role in facilitating lipid transport. These findings shed new light on the structural characteristics of Bd1152 and its potential involvement in lipid transport mechanisms.

Chapter 6 - Investigating the Bd1152 conformational landscape from Cryo-EM data

6.1. Introduction

Proteins often undergo conformational changes as a part of their functional cycle. The inherent flexibility is often linked to the function and mechanism behind it. Cryo-EM holds the potential of uncovering the protein motion, subsequently accelerating the understanding of fundamental macromolecular processes. Whilst protein dynamics are of paramount importance, heterogeneity represents an enormous challenge in SPA (Single Particle Analysis). Maximum likelihood classification methods are extremely powerful when it comes to separating discrete particle confirmations; however, they are not well suited for continuous molecular motions. Traditional 3D refinement methods assume complete particle rigidity, resulting in a loss of high spatial frequency information in dynamic regions.

Newly developed 3D refinement algorithms, such as *non-uniform* refinement introduced by cryoSPARC (Punjani *et al.,* 2020), account for inherent particle flexibility. Whilst showing significant improvement in the reconstitution of flexible proteins (Punjani *et al.,* 2020), it does not describe the movement *per se.* Masking-based approaches that focus on selected sub-regions such as multi-body refinement (Nakane *et al.,* 2018), could help improve the resolution of the rigid bodies and better define the individual domains; however, structural assumptions are user-specified, introducing a degree of bias into the analysis. A deep neural network-based algorithm termed cryoDRGN (Zhong *et al.,* 2021) has recently been developed to describe continuous molecular motions. The overall framework of the cryoDRGN includes an encoder-latent space–decoder model. The model is trained to encode 2D images into the continuous vector space (latent space) and estimate per-particle latent encodings that describe heterogeneity Z_i. These encodings are then used by neural networks to reconstitute 3D density maps. CryoDRGN allows inspection of particle distribution that is clustered based on the Z_i and explores different molecular conformations (Zhong et al., 2021).

Another recently introduced algorithm to describe continuous flexibility is termed Zernike3D (Herreros e*t al.,* 2021, 2023), which has been incorporated in Scipion (de la Rosa-Trevín *et al.,* 2016). In general terms, Zernike3D can describe continuous flexibility at the individual particle level, producing a conformational landscape map. The reference volume map, obtained through the "standard" SPA processing pipeline, is used as a reference model to calculate the deformation fields. Deformation fields are then applied to the reference model, back-projected, and compared to the experimental 2D images. The process is iterated until it converges (Herreros *et al.,* 2021, 2023) (Figure 6.1).

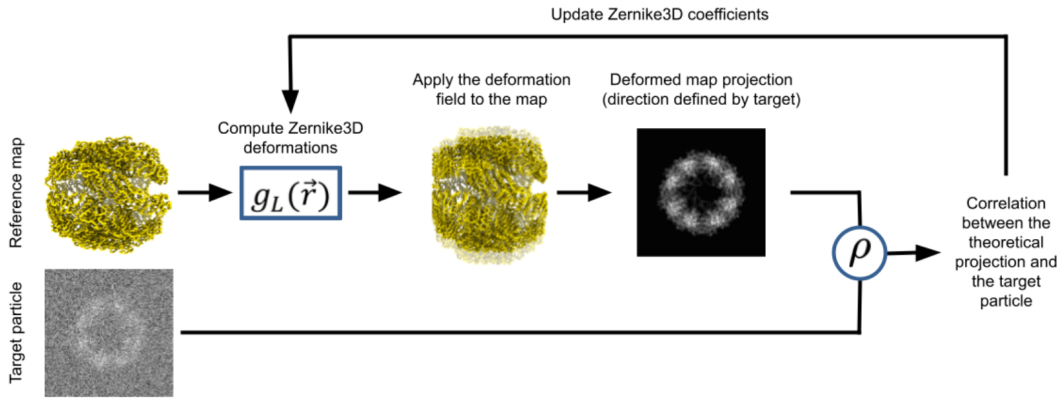


Figure 6.1. The framework of the Zernike3D algorithm.

Zernike3D deformation coefficients are calculated and applied to the reference map. Zernike3D coefficients are estimated by Powell's conjugate direction method, so the Pearson's correlation coefficient is maximised between the experimental data and the reference map. Figure obtained from Herreros *et al.,* (2023).

To define the continuous transitions that macromolecules might adopt between different conformations, the displacement of each region has to be estimated. The displacement that the reference map will adopt based on experimental 2D images is controlled by deformation coefficients. Zernike3D yields a total of 39 components per coefficient set (Herreros *et al.,* 2021, 2023). While clustering of the particles based on their deformation coefficients is performed in 39D space, the visualisation of the conformational landscape requires the reduction of dimensionality to depict clusters representing the heterogeneity in the 2/3D space. This is achieved either through PCA (Principal component analysis) (Ivosev *et al.,* 2008) or UMAP (Uniform Manifold Approximation and Projection) (McInnes *et al.,* 2020)

Results are visualised using a “conformational landscape” where each point (a single particle or a particle cluster) corresponds to its equivalent conformation. According to theoretical assumptions, molecular conformations that are proximal in 2/3D conformational space should exhibit a greater degree of similarity. However, in practical applications, the efficiency of dimensionality reduction techniques necessitates further improvement. Empirical data suggest that molecular conformations that are situated near one another in a 39-dimensional space may appear distant when projected onto 2/3 – dimensional space (personal communication with Dr Marcos Gragera and David Herreros). Consequently, extensive manual inspection is required to investigate conformational variability.

One of the advantages of Zernike3D is that it does not require a minimum number of particles, as the deformation is calculated per particle. This allows the user to produce high-quality reference maps through a number of different cleaning processes.

6.2. Chapter aims.

The cryo-EM data processing results indicated the potential structural plasticity of Bd1152, as observed from the 2D and 3D class averages. However, we hypothesised that the dynamics of Bd1152, as captured through discrete classification methods, might not fully reveal the complete range of motions present. To address this limitation, we employed innovative continuous flexibility algorithms to explore the conformational dynamics of Bd1152. In the following chapter, we present our findings on the dynamic behaviour of Bd1152.

6.3. Results

6.3.1. 2D and 3D classification suggest conformational heterogeneity.

2D and 3D classification revealed conformational heterogeneity in both the soluble Bd1152^{2C} and the membrane - associated Bd1152 constructs. Two primary modes of movement were identified: tilting of the MCE rings and variability in interdomain distances (Figure 6.2). These findings suggest that Bd1152 exhibits structural dynamics which might represent an integral role in its function as a putative lipid transporter. Consequently, we proceeded to gain a more comprehensive understanding of the conformational flexibility of Bd1152.

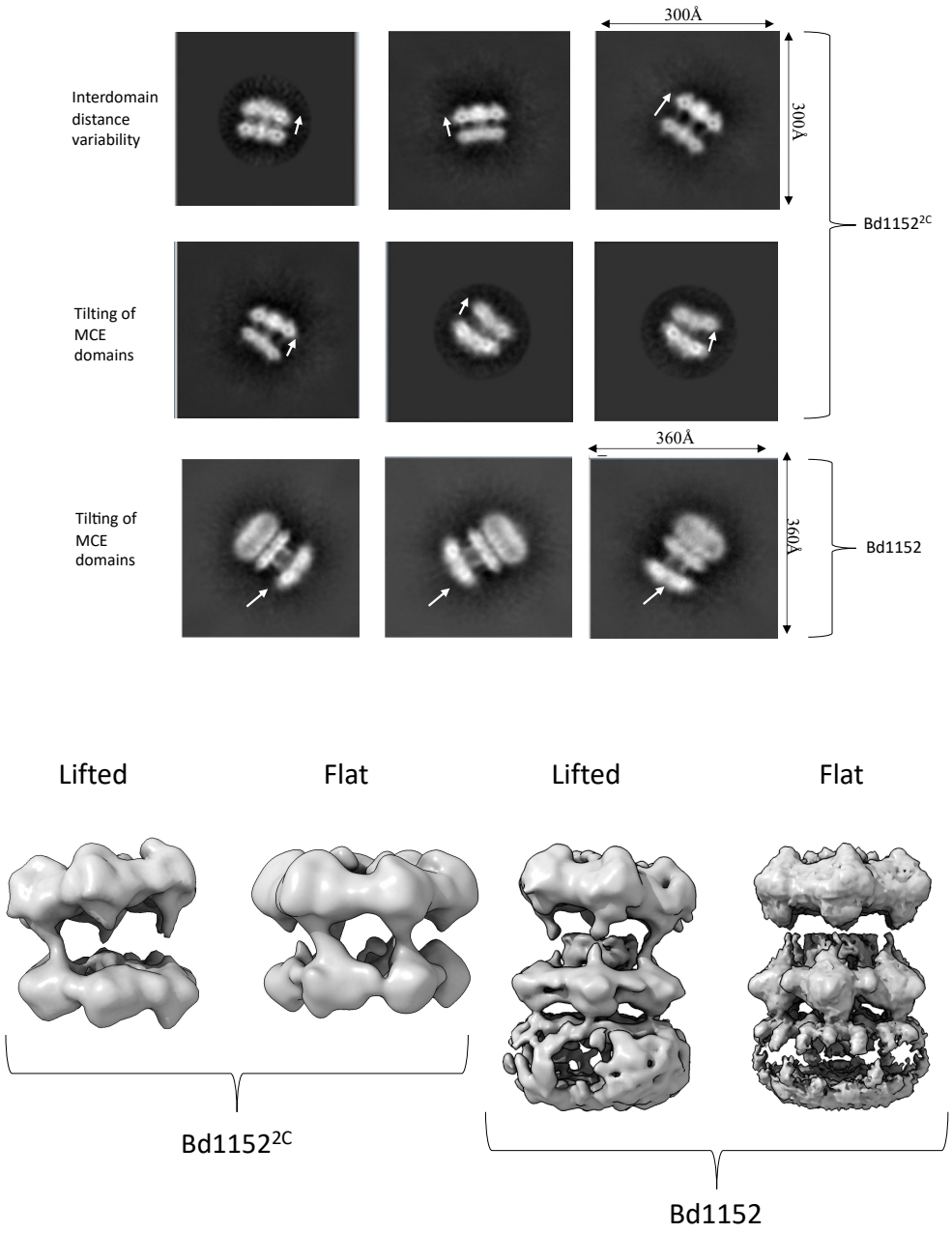


Figure 6.2. 2D and 3D classification of Bd1152^{2C} and Bd1152 indicating conformational heterogeneity.

(Top) 2D class averages of the Bd1152^{2C} and Bd1152 samples demonstrate distinct conformations, with white arrows indicating flexible regions.

(Bottom) 3D class averages of the Bd1152^{2C} and Bd1152 samples demonstrate movement the top MCE ring, resulting in “lifted” and “flat” conformations. The density maps were visualised with ChimeraX (Pettersen *et al.,* 2021).

6.3.2. Multi-body refinement

2D and 3D classification revealed that conformational transitions might exist within the dataset. If that is the case, the discrete classification methods are not suitable for analysing data with continuous motion (Nakane *et al.,* 2018). To investigate the conformational heterogeneity of Bd1152, multi-body refinement was performed, which is based on independent refinements of rigid bodies that maintain their structure but differ in their relative orientations (Nakane *et al.,* 2018). Bd1152 was split into three rigid bodies: the micelle, the 1st MCE ring and the 2nd MCE ring, whilst Bd1152^{2C} was split into 2 rigid bodies: the 1st and 2nd MCE rings. Corresponding masks were created with a 30 Å low-pass filtered map, with soft edges of 8 Å placed on the mask boundaries. (Figure 6.3, A). For Bd1152, the micelle and the 2nd MCE ring were set to rotate with respect to the 1st MCE ring (rlnBodySigmaAngles – 5; rlnBodySigmaOffset – 2). Principal component analysis revealed that PC1, PC2 and PC3 account for ~45% of variance. Eigenvectors 1 and 2 describe an identical 30˚rotational motion of the 2nd MCE ring (Figure 6.3, C). Eigenvector 3 describes a ~15˚ tilt of the 2nd MCE rings, resulting in “flatter” conformation compared to the reference map (Figure 6.3, C).

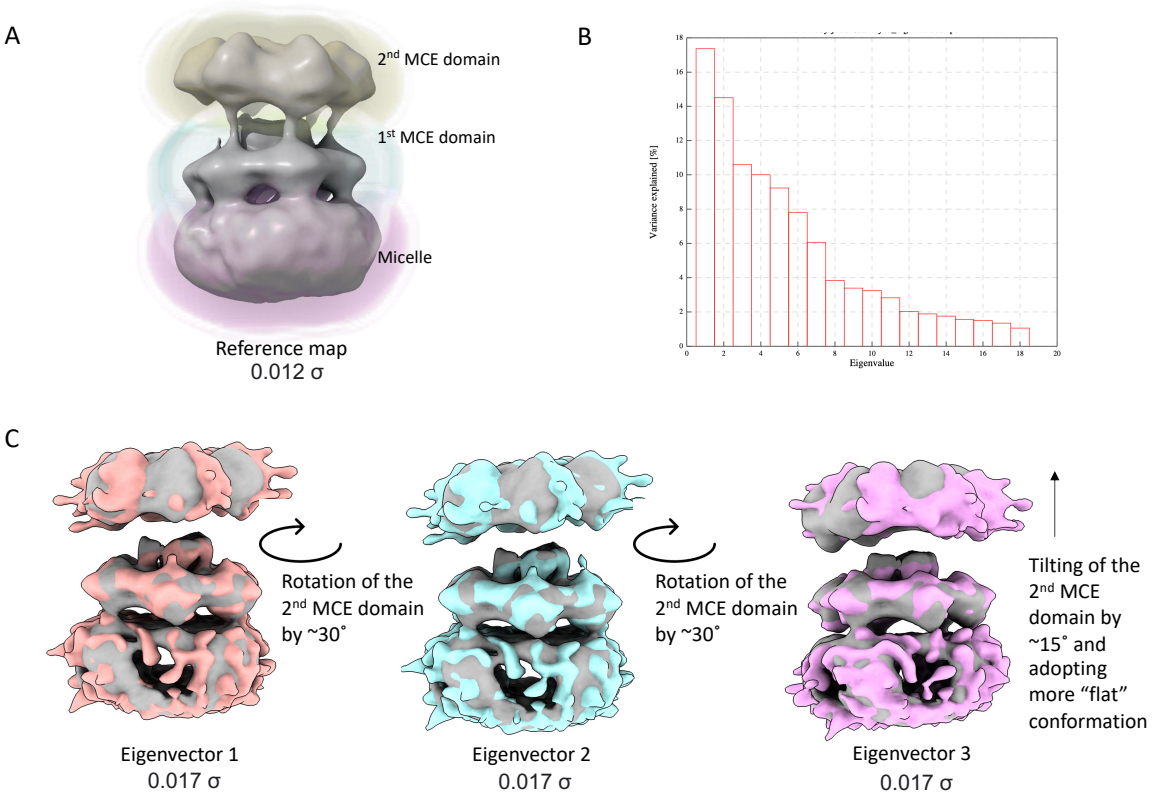


Figure 6.3. Conformational dynamics of Bd1152 determined through multi-body refinement.

(A) The volume map of Bd1152 with three transparent rigid body masks: the micelle, the 1st MCE ring, the 2nd MCE ring). (B) Contribution of all eigenvectors to the variance. The cryo-EM maps (coloured in salmon, blue and purple) corresponding to the first three eigenvectors are visualised in (C), with arrows indicating the observed flexibility. The grey map represents the reference map used as input for multi-body refinement. Contour values (σ) are displayed in the figure. The density maps were visualised with ChimeraX (Pettersen *et al.,* 2021).

For Bd1152^{2C}, the 2nd MCE ring was set to rotate with respect to the 1st MCE ring (rlnBodySigmaAngles – 5; rlnBodySigmaOffset – 2). Principal component analysis revealed that PC1, PC2 and PC3 account for ~45% of variance (Figure 6.4, B). Eigenvector 1 describes a slight tilting of the 2nd MCE ring by ~ 9˚. Eigenvector 2 suggests a slight tilting of the 1st MCE ring by ~5˚ and Eigenvector 3 shows interdomain distance variability (increased by ~10 Å) (Figure 6.4, C).

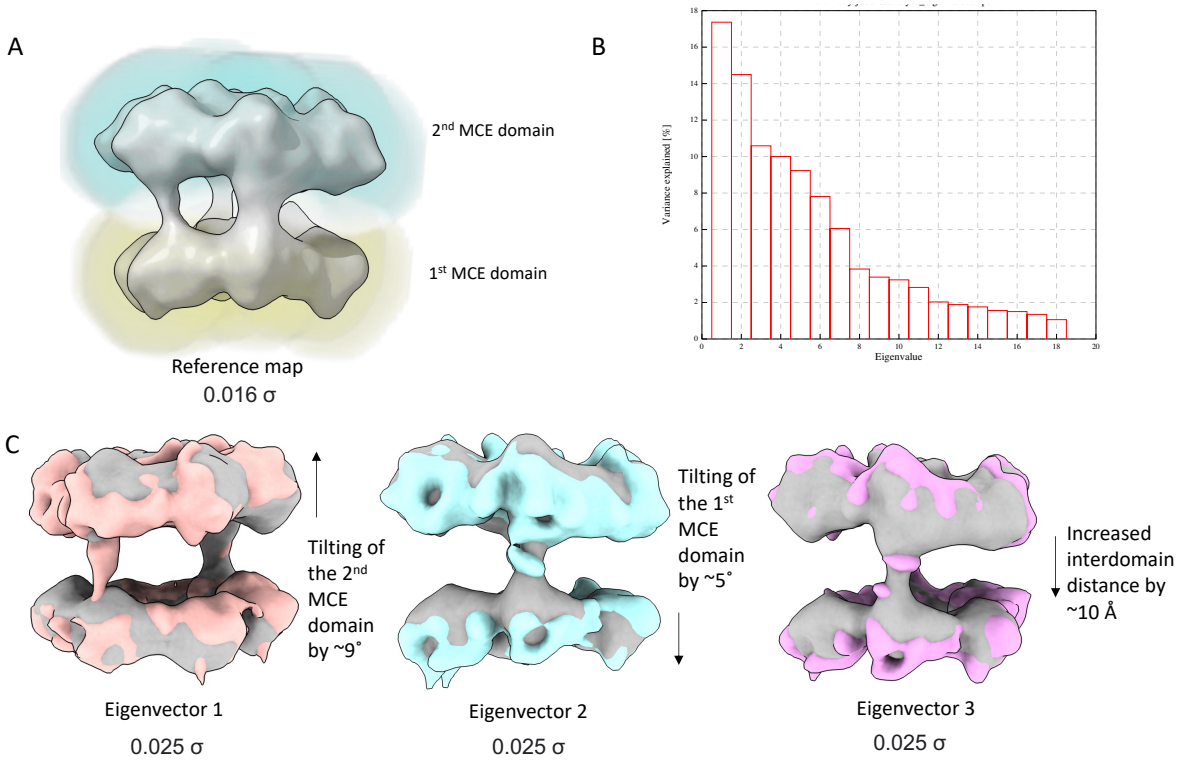


Figure 6.4. Conformational dynamics of Bd1152^{2C} determined through multi-body refinement.

(A) The volume map of Bd1152^{2C} with two transparent rigid body masks: the 1st MCE ring, the 2nd MCE ring). (B) Contribution of all eigenvectors to the variance. The cryo-EM maps (coloured in salmon, blue and purple) corresponding to the first three eigenvectors are visualised in (C), with arrows indicating the observed flexibility. The grey map represents the reference map used as input for multi-body refinement. Contour values (σ) are displayed in the figure. The density maps were visualised with ChimeraX (Pettersen *et al.,* 2021).

Tilting of the 2nd MCE ring could potentially indicate functionally relevant dynamics, as similar tilting of MCE rings was observed in the *E. coli* LetB structure (Isom *et al.,* 2020). The distinct bodies of Bd1152 and Bd1152^{2C} were subjected to further 3D refinement jobs; however, the quality of the maps was not improved (data not shown). Despite the low-resolution of the cryo-EM maps, the implementation of multi-body refinement provided valuable insight in the global dynamics of Bd1152. Nevertheless, the resulting maps exhibited low quality, making the interpretation of results somewhat challenging.

6.3.3. Zernike3D

Zernike3D analysis was carried out at the I2PC center (CNB-CSIC, Madrid) as a part of Instruct-ERIC internship (PID2530) with a help of Dr Marcos Gragera and David Herreros.

Through the use of 2D and 3D classification, as well as multi-body refinement, we observed that Bd1152^{2C} has the ability to adopt both "tilted" and "flat" conformational states. To gain a more comprehensive understanding of the protein's conformational landscape and explore the possibility of continuous flexibility, we employed the novel Zernike3D algorithm developed by Herreros *et al.,* (2023, 2021).

6.3.3.1. Data pre-processing for Zernike3D analysis

Prior to the Zernike3D analysis, we undertook an extensive data pre-processing to ensure the generation of a high-quality map in C1 symmetry (Figure 6.5). This included a selection of a particle set with the highest angular assignment confidence. We performed two *non-uniform* (cryoSPARC) refinements in C1 symmetry, with identical input parameters and a particle set of 234873. The resulting volumes and aligned particles were subjected to the *Xmipp3* compare angles protocol, where the confidence values of angular assignments between the two *non-uniform* refinement jobs were compared. The particle set of 88461 had an angular alignment difference less than 5˚ and therefore was selected for further processing. To obtain a final reference model and its associated particle set for the Zernike3D analysis, we performed an *ab-initio* reconstitution and *non-uniform* refinement (Figure 6.6). Our aim was to optimise the quality of the input data, enhancing the precision and reliability of the subsequent Zernike3D analysis. This was done to ensure that any map differences observed were solely due to particle heterogeneity and not artifacts resulting from poor image alignment.

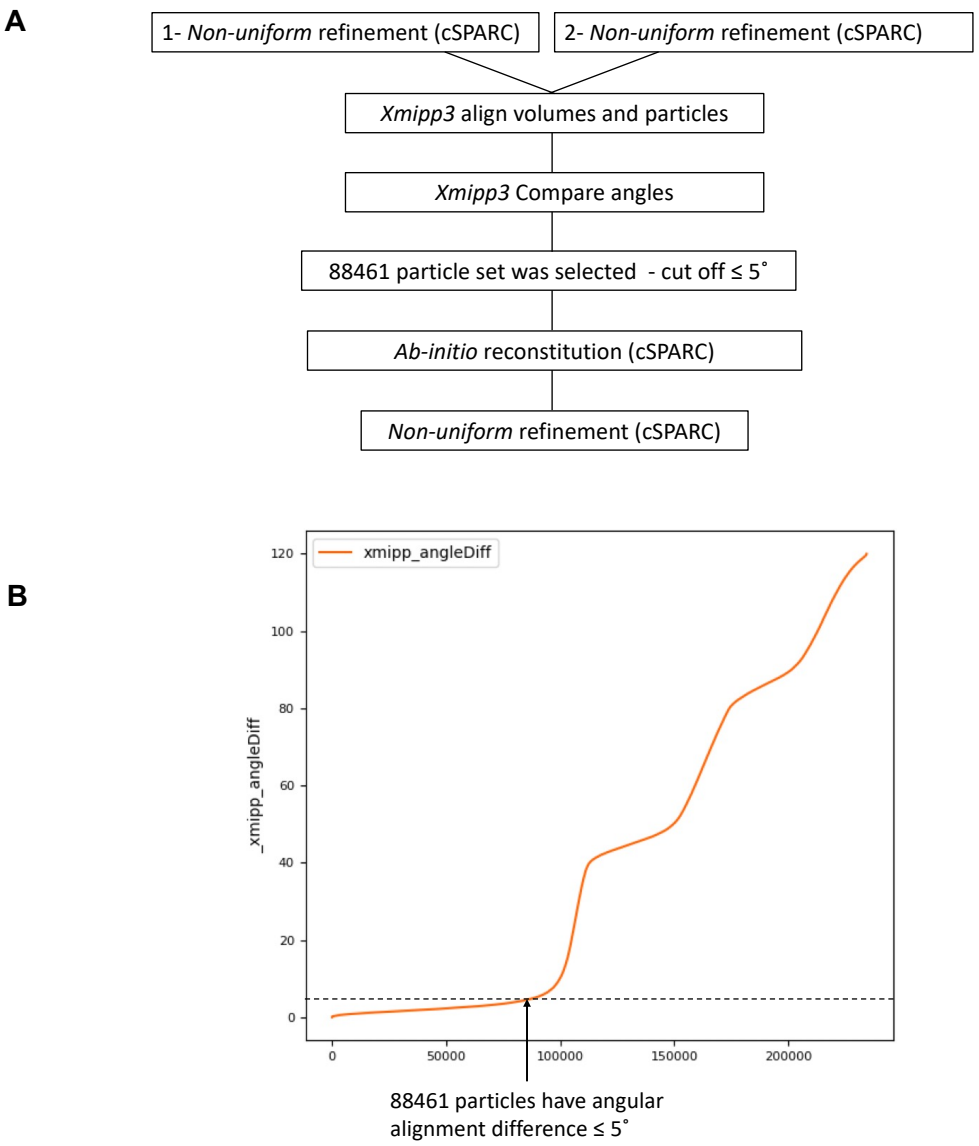


Figure 6.5. The data pre-processing steps to obtain an input particle set and the reference map for the Zernike3D analysis.

A) Data processing workflow: two sets of cryoSPARC *non-uniform* refinements with identical parameters (in C1 symmetry) were performed, followed by the *Xmipp3* protocol of angle comparison. Only particles with an angle alignment difference equal to or less than 5˚ were further used for 3D reconstitution (88461 particles) These particles were subjected to *ab-initio* (C1) reconstitution, followed by *non-uniform* refinement (C1 symmetry). B) The plot obtained from the *Xmipp3* compare angles protocol. The X-axis represents the number of particles, whilst the y-axis represents the angular difference. The black dotted line indicates the cut-off value.

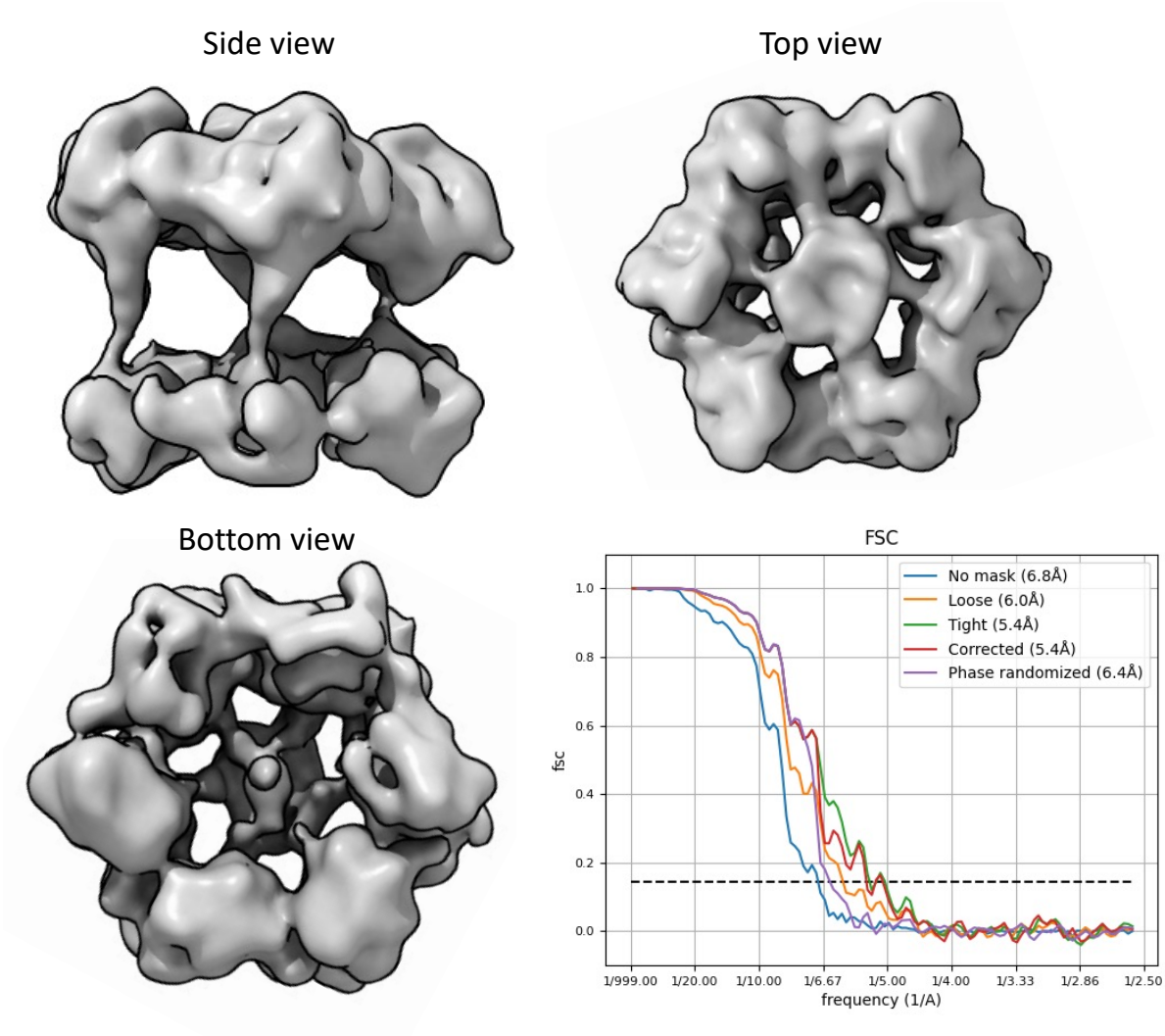


Figure 6.6. The cryo-EM map of Bd1152^{2C} used as the reference for the Zernike3D analysis.

The map was obtained from the data processing pipeline described in Figure 6.5. The average resolution derived from Fourier Shell Correlation (FSC) is shown with blue (no mask), loose mask (red), tight mask (cyan) and corrected (purple) curves. A cut-off 0.143 (black vertical dotted line) was used for resolution estimation. All volumes are displayed at the contour level of 0.28. The density map was visualised with ChimeraX (Pettersen *et al.,* 2021).

6.3.3.2. Zernike3D analysis

The resulting reference model and aligned particle set were subjected to Zernike3D analysis. Zernike3D coefficients are calculated per-particle and represent the “amount” of deformation that needs to be applied to the reference map, so its projection best matches the experimental image. 39 Zernike3D coefficients are calculated per-particle; therefore, representation of deformation field/conformational landscape would require a 39-dimensional space. We utilised UMAP to reduce the dimensionality and represent the conformational landscape in 2D. We visually inspected conformational variability and selected eight clusters containing 5000 particles each to validate the deformed maps. The particle clusters are depicted in the 2D UMAP representation in Figure 6.7 as a coloured dots. Particles belonging to each cluster were used as an input for *ab-initio* reconstitution followed by *non-uniform* refinement, resulting in 8 maps depicted in Figure 6.7.

The reconstitution of particles obtained from Zernike3D analysis of $Bd1152^{2C}$ uncovered a novel degree of flexibility that was not previously detected by maximum likelihood data processing methods. In addition to the anticipated "tilting" of the top and bottom MCE rings, the twisting of these rings in relation to one another and inter-monomer motion were also detected. Moreover, the bottom ring displayed a slight “curving” (Figure 6.7). Overall, Zernike3D revealed that $Bd1152^{2C}$ undergoes continuous conformational changes and experiences high mobility - a feature not previously observed in MCE domain-containing proteins.

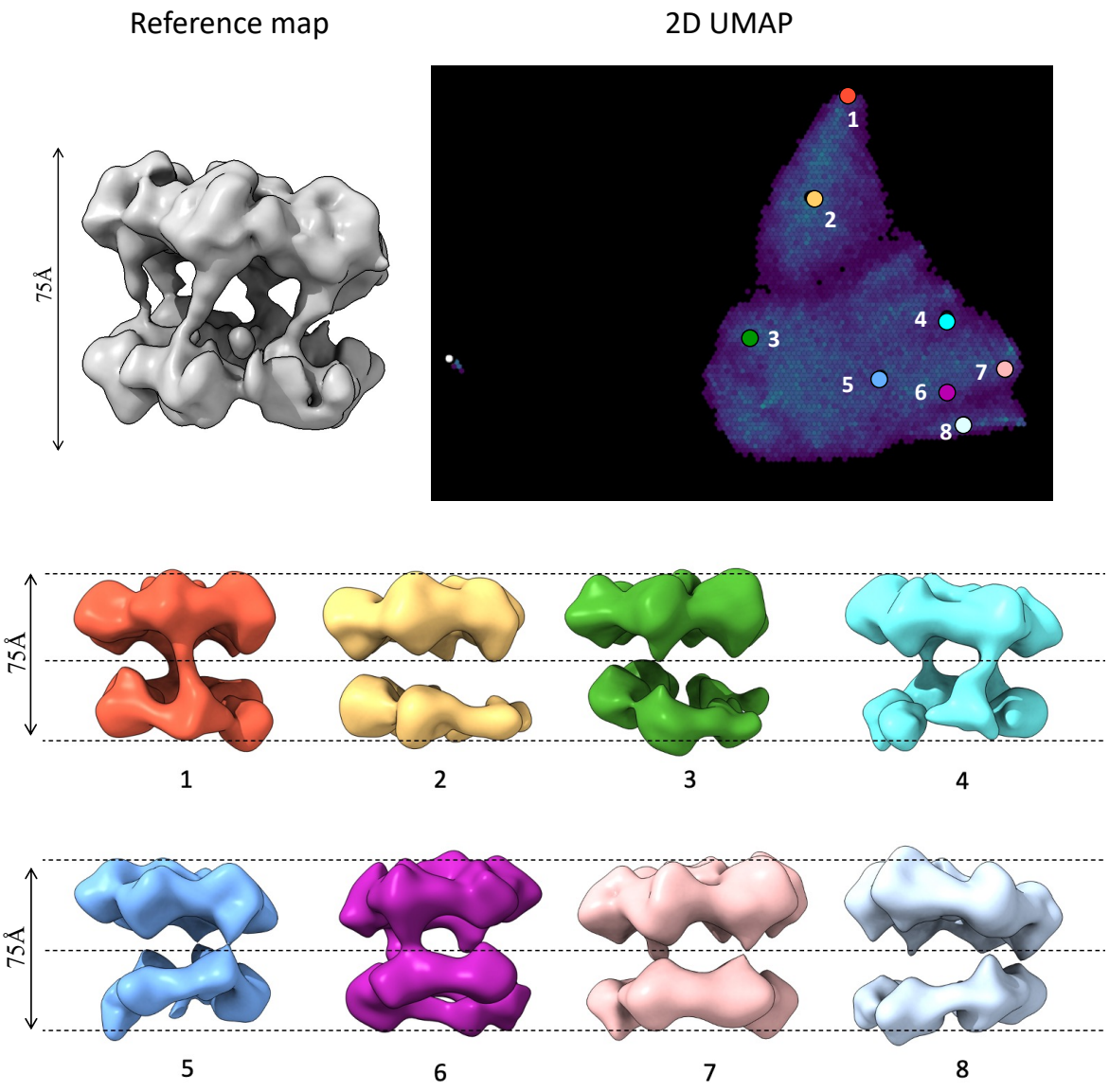


Figure 6.7. The Zernike3D analysis of Bd1152^{2C}.

(Top) The input reference map of Bd1152^{2C} and the 2D UMAP representation of the Zernike3D coefficient space. Each point in the cloud represents cluster of particles sharing similar deformation coefficients. Brighter shades of blue indicate a higher number of particles, while purple shades indicate a lower number of particles. (Bottom) Example of 8 conformations extracted from the conformational landscape (2D UMAP). Different conformations correspond to reconstructions from the 5000 particles (coloured dots in the UMAP). Maps are displayed at a contour level of 0.015. The dotted lines are shown to aid in the visualization of the motions. The density maps were visualised with ChimeraX (Pettersen *et al.,* 2021).

6.3.3.3. Flexible fitting of the Bd1152^{2C2A} crystal structure into the Zernike3D cryo-EM maps.

iMODFIT is based on a normal mode analysis (NMA) in internal coordinates (IC). iMODFIT generates the models by randomly perturbating the initial atomic model and optimising the conformation by utilising various energy minimisation techniques (Lopéz-Blanco and Chacón, 2013). The scoring function used by iMODFIT takes into the account electrostatic potential energy and van der Waal interactions to find the conformation that best fits the cryo-EM map and has the lowest energy therefore ensuring that atomic models are physically meaningful (Lopéz-Blanco and Chacón, 2013). Whilst cryo-EM maps obtained from Zernike3D analysis were limited to 7-9 Å resolution, we proceeded to employ the iMODFIT flexible fitting of the crystal structure of Bd1152^{2C2A} into each of the cryo-EM density maps obtained from the Zernike3D analysis. The Zernike3D maps showed a variety of conformations with some of them displaying high degree monomer displacements from the hexameric rings. Most likely due to physical macromolecular constrains, such as electrostatic energy and van der Waal interactions, iMODFIT was not able to deform the crystal structure to “perfectly” fit the Zernike3D cryo-EM maps. Nevertheless, visual inspection of the fitting showed that initial model was deformed in a manner and direction consistent with the cryo-EM maps.

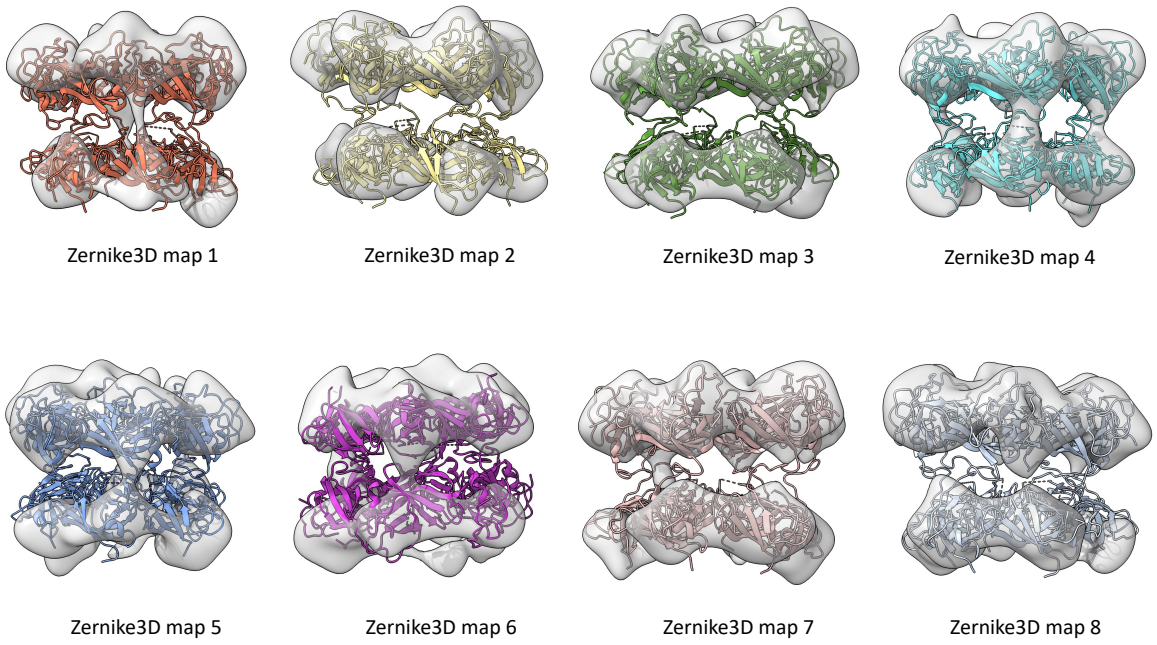
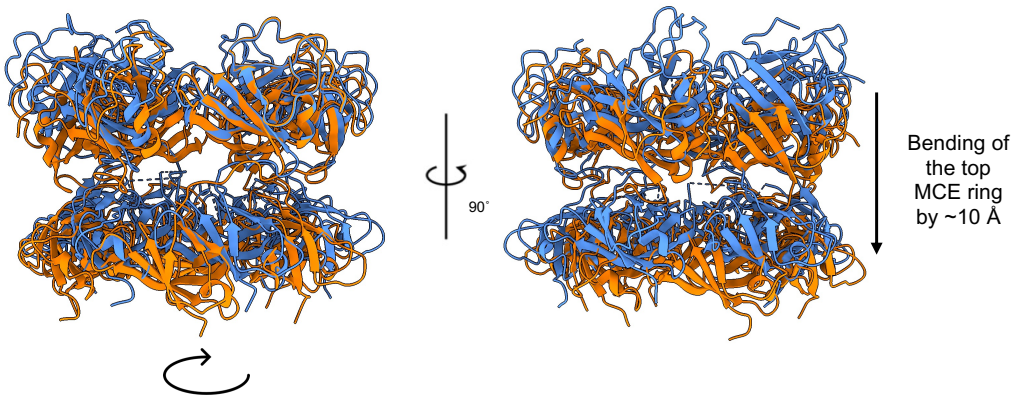


Figure 6.8. Flexible IMODFIT fitting of the Bd1152^{2C2A} crystal structure into each of the Bd1152^{2C} Zernike3D maps.

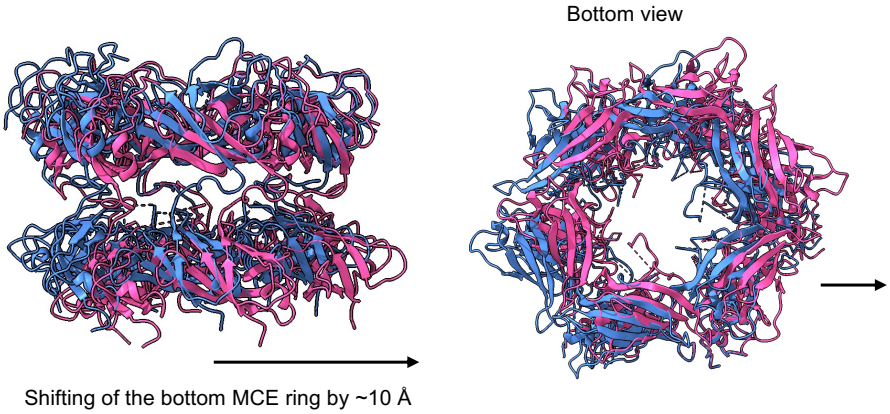
All cryo-EM maps are displayed at the contour level of 0.3. Atomic model colours and numbers correspond to the Zernike3D maps in Figure 6.7. The models and the corresponding density maps were visualised with ChimeraX (Pettersen *et al.,* 2021).

We appreciate that the resolution of the cryo-EM maps is not optimal for drawing definitive conclusions. Nevertheless, it allowed us to explore the overall dynamics of the protein. Analysis with iMODFIT revealed that the bottom MCE ring rotates ~30˚ relative to the top MCE ring. Furthermore, there is bending observed in the lower MCE ring, along with variation in interdomain distances (Figure 6.9). The overall “tilting” of the rings appears to be a result of movement within individual monomers rather than the entire MCE ring. Additionally, we predict that short linkers connecting the two MCE rings provide freedom for the flexibility and may have greater functional importance than previously recognised.

The reference map and the Zernike3D map 1



The reference map and the Zernike3D map 2



The reference map and the Zernike3D map 5

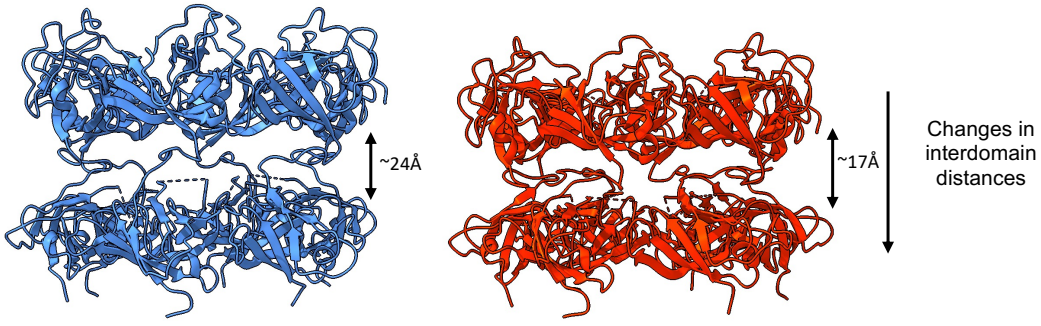


Figure 6.9. Structural dynamics of Bd1152^{2C} protein derived from the Zernike3D analysis and the IMODFIT flexible fitting of the Bd1152^{2C2A} crystal structure. Cornflower blue represents the crystal structure of Bd1152^{2C2A} fitted into the reference map that was used as input for the Zernike3D analysis. The models shown in orange, pink and red represent Bd1152^{2C2A} X-ray structure fitted into the cryo-EM maps emerging from the Zernike3D analysis. The models were visualised with ChimeraX (Pettersen *et al.,* 2021).

6.3.4. CryoDRGN

We next utilised the CryoDRGN algorithm to further explore the characteristics of heterogeneity present within the Bd1152^{2C} dataset. We trained an 8D latent variable model which resulted in 10 distinct 3D volumes obtained directly from cryoDRGN reconstitutions.

The distribution of particles in the latent space is smooth, with no discrete clusters, suggesting that the flexibility Bd1152^{2C} is continuous. Most of the dynamics were observed within the pore lining loops of the first MCE ring, revealing “open” and “closed” conformations. Three models were chosen to display the sequence of the central channel opening and closing (Figure 6.10).

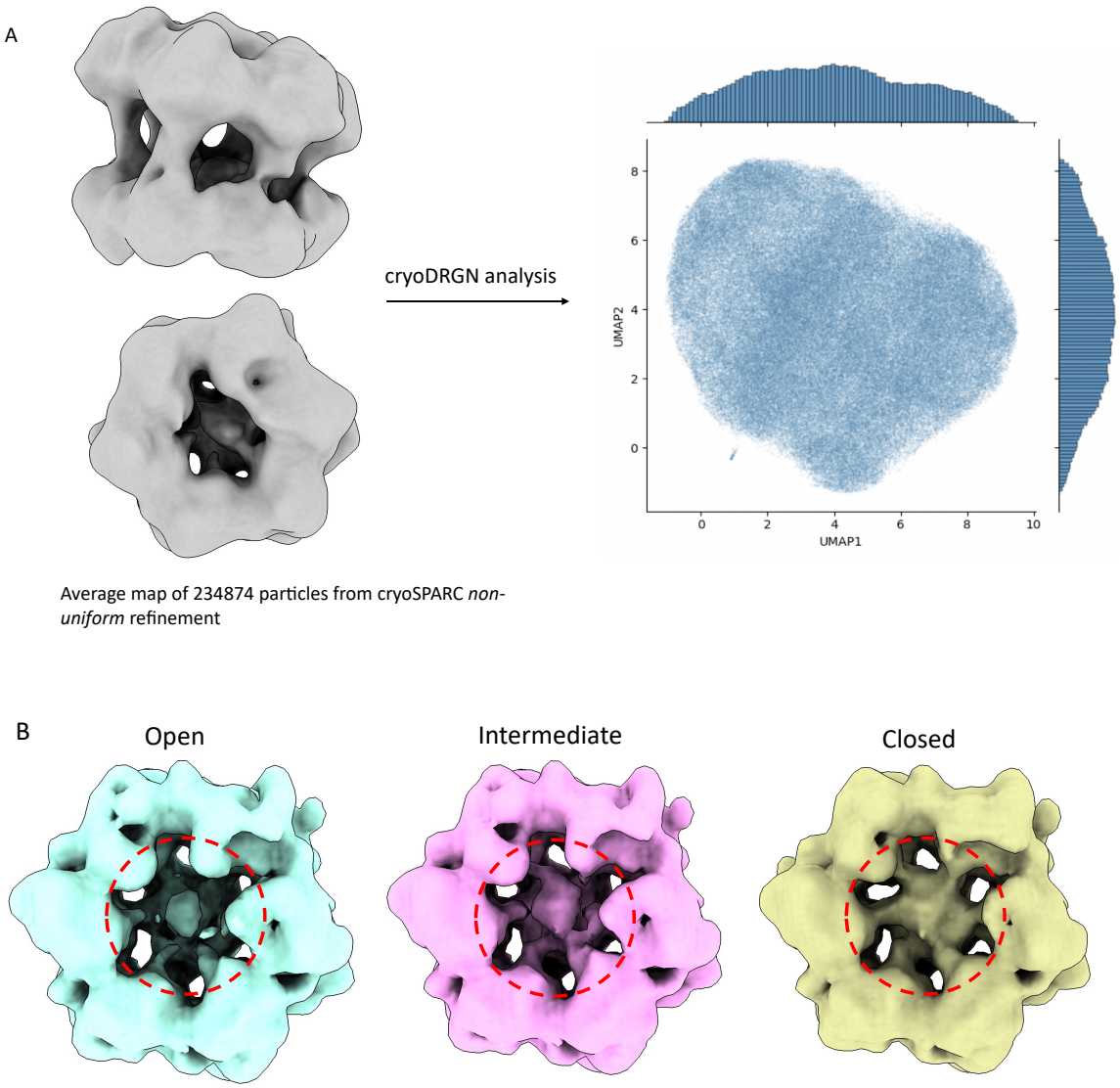


Figure 6.10. Dynamics of Bd1152^{2C} obtained from the CryoDRGN analysis.

A) (Left) The consensus map obtained from cryoSPARC *non-uniform* refinement of 234874 particles, displayed at a contour level of 0.015. (Right) A 2D UMAP visualisation of the latent variable distribution of the 234874 particles after training an 8D latent variable model with cryoDRGN. B) Three cryo-EM maps obtained from the cryoDRGN analysis showing “Open” (blue), “Intermediate” (pink) and “Closed” (yellow) conformations of the bottom MCE ring. All maps are displayed at a contour level of 0.04. The red dashed circles indicate the location of pore lining loops. The density maps were visualised with ChimeraX (Pettersen *et al.,* 2021).

6.3.5. Variation in interdomain distance

The comparison of X-ray and low-resolution cryo-EM maps revealed a significant variation in the interdomain distance. The cryo-EM map of full-length Bd1152 displayed the largest inter-domain distance between the two MCE rings (~39 Å) (Figure 6.11, B), the cryo-EM map of $Bd1152^{2C}$ represents an intermediate state (~29 Å) (Figure 6.11, A), whilst the X-ray structure of $Bd1152^{2C2A}$ and the cryo-EM map of the $Bd1152^{2C}$ GraFix sample exhibited the smallest interdomain distance (~16 Å) (Figure 6.11, C). However, it is unclear whether this could represent a biologically relevant conformation, or it resulted from crystal packing and glutaraldehyde crosslinking, respectively. Nevertheless, a similar feature was reported by Cooper (2023), demonstrating that the cryo-EM structure of the full length PqiB displayed an "extended" conformation between the first and second MCE rings (Figure 6.11, B) when compared to the soluble PqiB model (Figure 6.11, D) (Ekiert *et al.,* 2017).

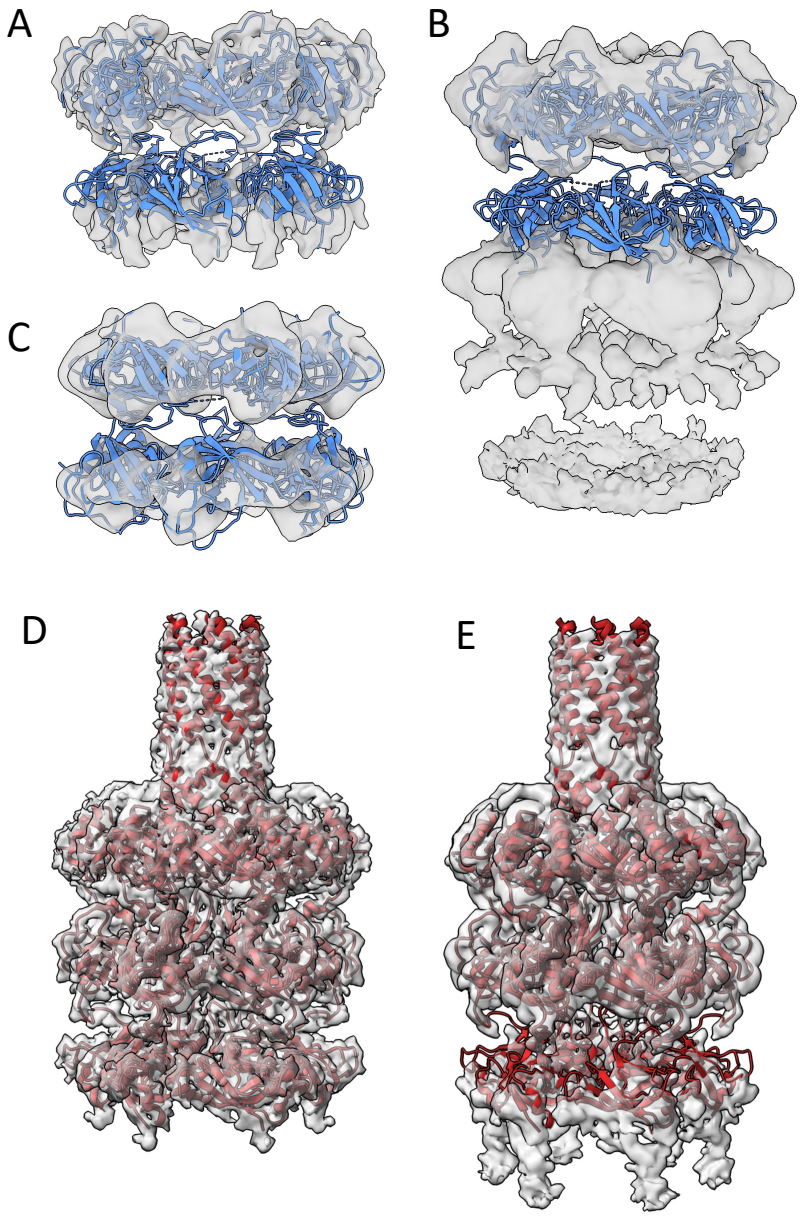


Figure 6.11. The variation in the interdomain distances between the MCE rings of Bd1152 and *E. coli* PqiB.

(A) the cryo-EM map of Bd1152^{2C}; (B) the cryo-EM map of Bd1152; (C) the cryo-EM map of the Bd1152^{2C} GraFix sample. The X-ray structure of Bd1152^{2C2A} was fitted into the A, B, C cryo-EM maps (blue). D) The PqiB$^{39\text{-}564}$ (PBD: 5UVN) fit into the EMD-8608 and E) PqiB FL cryo-EM map (Cooper., 2023). D and E are adapted from Cooper (2023). The models and density maps were visualised with ChimeraX (Pettersen *et al.,* 2021).

6.4. Discussion

In this chapter, we utilised the multi-body refinement, Zernike3D, and cryoDRGN algorithms to investigate the inherent continuous flexibility of $Bd1152^{2C}$. Whilst the flexibility of Bd1152 prevented us from obtaining a high-resolution cryo-EM map, it revealed conformational changes not previously associated with MCE domains.

6.4.1. Zernike3D revealed high global dynamics in $Bd1152^{2C}$.

Classical 2D and 3D classification methods and multi-body refinement allowed us to identify that $Bd1152^{2C}$ might exist in multiple conformational states, as tilting of MCE rings and discrepancies in interdomain distances were observed. However, anticipating that $Bd1152^{2C}$ might possess additional mobility, we employed the recently introduced Zernike3D algorithm to assess its continuous flexibility.

Zernike3D revealed that $Bd1152^{2C}$ is highly dynamic possessing a number of different conformations that were not previously identified with discrete classification approaches. This includes tilting of both the top and bottom MCE rings, rotation of the bottom MCE ring in respect to the top MCE ring and the inter-monomer movement. The latter is the most surprising as we assumed that MCE hexameric rings are highly rigid. The pronounce global dynamics of $Bd1152^{2C}$ may hold significant implications for its putative transporter function. This study suggests that Bd1152 may utilise its inherent flexibility to facilitate the transportation of hydrophobic substrates through the central channel.

While some intra-domain flexibility was observed for the MlaD (Coudray *et al.,* 2020), PqiB (Isom *et al.,* 2020) and LetB (Isom *et al.,* 2020) proteins, for the first time we reveal the inter-MCE domain movement suggesting that MCE proteins might be much more dynamic than previously assumed. Future investigations should prioritise the application of the Zernike3D algorithm to explore the continuous flexibility of the *E. coli* PqiB and LetB proteins. Such studies have the potential to provide comprehensive evidence supporting the notion that the high structural plasticity might be highly conserved feature across MCE proteins with significant implications for their putative transport function.

6.4.1.1. The Bd1152 MCE domain linkers might possess significant structural importance.

The insights gained from the Zernike3D analysis highlight the potential importance of the short linkers that connect the MCE rings of Bd1152 in its putative lipid transport function. These linkers are likely responsible for providing a high degree of freedom for flexibility, suggesting that their functional significance may be even more crucial than previously acknowledged.

This hypothesis could be experimentally tested through *in vivo* complementation assays. Whilst *in-vivo* functional characterisation of Bd1152 within *B. bacteriovorus* comprises significant challenges, due to the complexity of genetic modifications (personal communication with Prof Andrew Lovering), the importance of inter-MCE domain linkers upon the function of multi-MCE domain containing proteins may be assessed in the *E. coli* proteins PqiB and LetB. These proteins contain multiple MCE domains (3 and 7, respectively) separated by analogous linker regions (Ekiert *et al.,* 2017; Isom *et al.,* 2020; Cooper, 2023). Furthermore, both proteins have established phenotypes which have been utilised within previous publications to assess the viability of mutants, thus by generating

constructs of the *E. coli pqiB* and *letB* genes with different linker lengths, we could investigate the importance of these linker regions, and thus protein dynamics upon function (Isom *et al.,* 2020; Vieni *et al.,* 2021).

6.4.2. CryoDRGN analysis revealed the opening and closing of the central Bd1152 channel.

Poor density has been observed at the regions accommodating PLLs. PLLs may be disordered or adopt many different conformations as seen from the X-ray structures of $Bd1152^{2C}$ (refer to Chapter 5) and LetB MCE4-MCE5 domains (Isom *et al.,* 2020). Analysis of continuous flexibility of $Bd1152^{2C}$ with cryoDRGN revealed that PLLs do indeed experience high mobility. While the resolution of the cryo-EM maps is not optimal enough to build a model, the central pore of the first MCE ring exhibits clear density changes at the central channel. PLLs appear to obtain “Open” and “Closed” conformations through a number of intermediate states, potentially representing an insight into the mechanism of transport adopted by Bd1152 .

The conformational changes within the $Bd1152^{2C}$ PLLs closely resemble the PqiB MCE ring 2 and the LetB MCE ring 6 PLL dynamics (Isom *et al.,* 2020) (Figure 6.12). In the LetB structure “open” and “closed” states suggested to be induced by the pivoting motion of the MCE rings in addition to the shifting of the PLLs. We propose that a similar mechanism might be adopted by Bd1152. As seen from the Zernike3D analysis, the pivoting and twisting motions are adopted, hence this movement could cause significant changes within the PLLs, potentially representing biological significance in the transport of substrates.

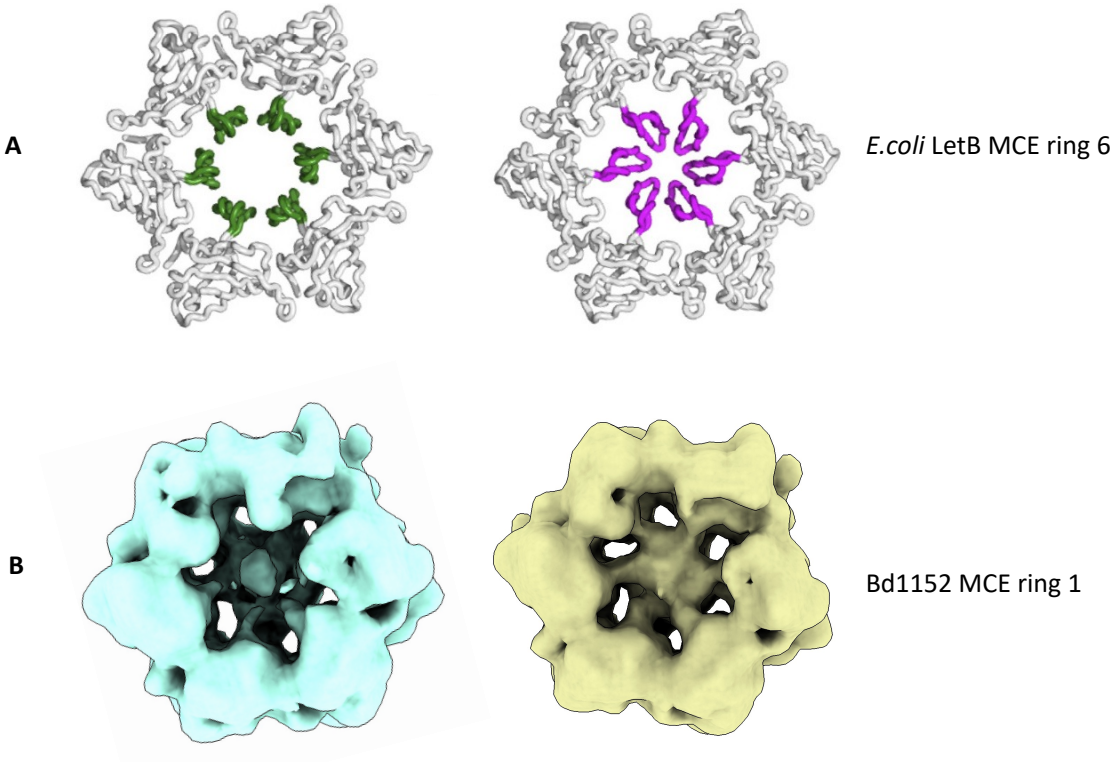


Figure 6.12. Comparison of the LetB and Bd1152 pore lining loop conformations.

A) The structure of the LetB 6th MCE domain (*E. coli)* displaying "opening" and "closing" of the central channel via the pore lining loop movement (Figure obtained from Isom *et al.,* 2020). B) CryoEM maps of Bd1152^{2C} obtained from the cryoDRGN analysis revealing "open" and "closed" conformations of the 1st MCE ring. The density maps were visualised with ChimeraX (Pettersen *et al.,* 2021).

6.4.3. Interdomain distance variability.

Considerable interdomain length discrepancy was observed in the 2D classes, multi-body refinement and Zernike3D analysis of the Bd1152^{2C} construct. Whilst the role of this particular motion is unclear, it could possess biological significance for substrate transport within the central channel.

Additionally, the Bd1152^{2C} cryo-EM map showed more “compact” conformation when compared to the full-length Bd1152 cryo-EM map. Similar observation was reported by Cooper (2023) showing that the full-length *E.coli* PqiB structure adopts more “extended” conformation when compared to the soluble PqiB construct. Whilst at this stage structural data is limited to make assertive conclusions, it does suggest that transmembrane helices might play a role in the final conformation of MCE domain-containing proteins.

6.4.4. Limitations within the study

6.4.4.1. The role of transmembrane region in the flexibility of MCE proteins.

It is crucial to point out that a number of limitations exist within the current study. First, we investigated the flexibility landscape of the soluble Bd1152^{2C} construct, hence the role of transmembrane helices in Bd1152 mobility is currently unclear. We already proposed that the transmembrane region might affect the final conformation of Bd1152, thus there is a high likelihood that it may influence the conformation landscape.

We do predict that membrane-association could restrict some of the Bd1152 movement. Nonetheless, a tilting movement has been observed in the *E.coli* MlaD protein associated with the inner membrane MlaFEDB complex (Coudray *et al.,* 2020) suggesting that tilting movements of Bd1152^{2C} might be functionally relevant. It is important to note that we re-processed the full length Bd1152 cryo-EM dataset within the Scipion framework, with the aim of obtaining higher resolution structure and investigating its conformation landscape. However, despite the extensive data processing, we failed to obtain high-quality cryo-EM model in C1 symmetry that would be suitable for flexibility analysis. Nevertheless, multi-body refinement and 3D classification revealed that flexibility is indeed present within the Bd1152 full-length cryo-EM dataset, suggesting that that structural plasticity might not be limited to the soluble Bd1152 construct. Additionally, multi-body refinement and 3D-flex analysis of the LetB structure (Isom *et al.,* 2020) were also performed on the soluble construct hence introducing a degree of confusion regarding the role of transmembrane helices in the flexibility of the protein. Future studies should focus on investigating the conformational landscape of a full-length constructs ideally reconstituted into a more native environment, such as lipid nanodiscs, to provide insight into their functionally related flexibility.

6.4.5.2. The impact of cysteine mutations and absence of PL on the dynamics of Bd1152.

The cysteine mutations and the absence of the co-purifying PLs could have a significant effect on the flexibility of $Bd1152^{2C}$. However, this is a difficult problem to address as the $Bd1152^{37-283}$ construct is very unstable and prone to aggregation making it difficult to study. As discussed in Chapter 5, we attempted to obtain lipid-free $Bd1152^{37-283}$ through a detergent wash, however, this resulted in the immediate aggregation of the protein.

2D classes emerging from the $Bd1152^{37-283}$ cryo-EM dataset, however, displayed heterogenous conformations, suggesting that global dynamics of the particle might have been preserved in the $Bd1152^{2C}$ construct. Nevertheless, the impact of PL upon Bd1152 flexibility is unknown hence the protocol for obtaining apo-Bd1152 should be optimised prior to further structural investigation.

6.4.5. Future perspectives

It is noteworthy, that conformational landscapes obtained through modern computational approaches purely represent the probabilities of the macromolecule adopting a particular state. It does not provide us with information on the particular order that these conformations are adopted, Additionally, the most energetically stable conformational state will be predominant within the dataset hence resulting in the underrepresentation of the short-lived intermediate states. Time-resolved cryo-EM could represent a potential experimental design that could provide step-by-step structural dynamic information on phospholipid transport. While the Bd1152 protein would not currently be suitable for this approach due to the complexity of the cryo-EM data processing and lack of methods to investigate its function, *E. coli* PqiB could represent a model with a potential to retrieve structure-function relationship of PL transport via time-resolved cryo-EM experiments. Findings in our lab (data not published) suggest that PL transport via the PqiABC pathway is mediated by proton motive force (PMF). Adjusting the pH values of the protein sample followed by grid freezing at different time intervals could allow us to obtain intermediate states of phospholipid transport and structural dynamics associated with it.

6.5. Conclusion

In this chapter, we utilised newly developed continuous flexibility algorithms to explore the conformational landscape of $Bd1152^{2C}$. The Zernike3D analysis revealed that $Bd1152^{2C}$ exhibits a “fluid” like nature with interdomain pivoting and twisting movements. In addition, cryoDRGN revealed that $Bd1152^{2C}$ might adopt “open” and “closed” conformations of the bottom MCE ring channel. The results of this study indicate that structural plasticity might be crucial for the function of Bd1152 as a putative lipid transporter and might be highly conserved among MCE domain-containing proteins in various bacterial species.

Chapter 7 - Final discussion and future work.

7.1. Discussion

The aim of this study was to probe the structure of the *Bdellovibrio bacteriovorus* HD100 inner membrane-associated putative lipid transport protein Bd1152. After extensive optimisation of purification and construct design, we presented the crystal structure of $Bd1152^{2C2A}$ and cryo-EM structure of the $Bd1152^{2C}$ top MCE ring at 2.6 Å and 3.4 Å resolution, respectively, additionally uncovering its inherent structural plasticity through the utilisation of novel continuous flexibility tools. In this chapter, we further discuss the characteristics and potential roles of the Bd1153-Bd1151 operon components. We also address the limitations of our study and outline potential directions for future research.

7.1.1. Bd1152 – a putative lipid transport protein

The experimentally determined structures of MCE proteins are presently limited to those encompassing 1, 3, and 7 MCE domains, constituting the most prevalent MCE domain architectures (Isom *et al.,* 2017). Single MCE domain-containing proteins have been shown to interact with periplasmic chaperones for intermembrane PL transport (Huang *et al.,* 2016; Ekiert *et al.,* 2017; Hughes *et al.,* 2019) or form long periplasmic helices that potentially provide direct link between the IM and the OM (Chen *et al.,* 2023). Similarly, 3 and 7 MCE domain-containing proteins have been demonstrated to be of sufficient length to traverse the periplasm (Ekiert *et al.,* 2017; Isom *et al.,* 2020) providing a hydrophobic tunnel through which the PL transport is anticipated.

Considering the length differences between the *E. coli* LetB and Bd1152 proteins (~230 Å and ~70 Å, respectively), we anticipate that the Bd1153-Bd1151 system potentially interact with a periplasmic chaperone for intermembrane transport as Bd1152 alone is not sufficient to span the periplasm. Nevertheless, *B. bacteriovorus* is infamous for containing a large percentage of cryptic genes, making it challenging to find homologs within its genome. The recent development of Foldseek (Kempen *et al.,* 2023) allowed us to investigate whether any structural homologs to known *E. coli* transport modules exists within *B. bacteriovorus* that could potentially interact with Bd1152. Surprisingly, no structurally homologous proteins to MlaC, PqiC or MlaA were identified. We also examined the operon architectures of 2-MCE domain-containing proteins from various bacterial species, however these sequences only indicated the co-expression of a PqiA domain containing protein, homologous to Bd1153 (Figure 7.1). Collectively, these observations imply that the mechanism of action employed by 2-MCE domain-containing proteins may represent an entirely novel paradigm.

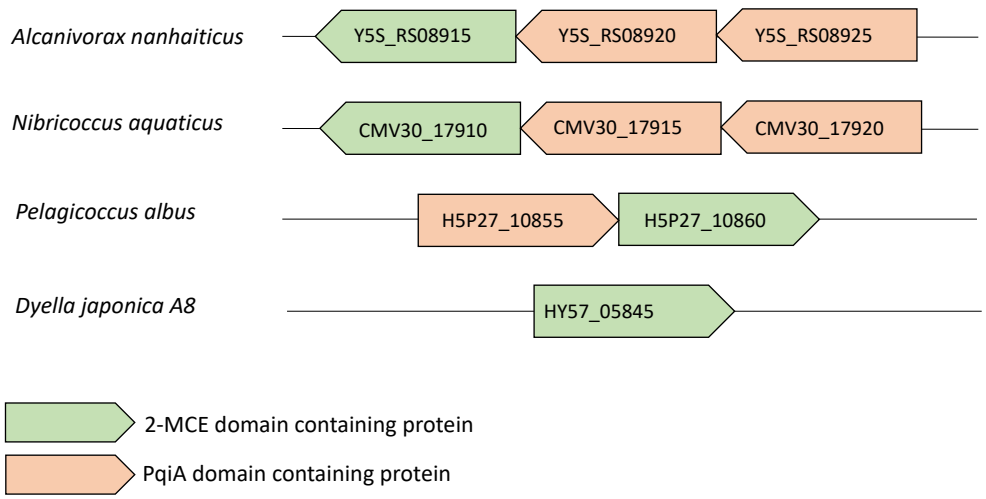


Figure 7.1. An example of the operon architecture comprising 2-MCE domain-containing proteins.

Additionally, the flexibility observed from the cryo-EM dataset of Bd1152^{2C} revealed much greater global dynamics than those reported from previous MCE protein structures (Ekiert *et al.,* 2017, Isom *et al.,* 2020). We propose that this observed structural plasticity may be relevant to the proposed lipid transport function of Bd1152, providing insight into its mechanics.

7.1.2. The c-terminus of Bd1152 suggests a unique functional mechanism.

Whist the structure of Bd1152 revealed the “standard” MCE fold, an intriguing feature of Bd1152 is its c-terminus which represents a 10 aa loop that folds back over the second MCE domain. In *E.coli* MlaD and PqiB, the c-terminus forms a coiled coil providing an interaction interface for MlaC (MacRae *et al.,* 2022) and potentially PqiC (Cooper, 2023), respectively. The LetB structure, however, possesses a similar c-terminus as in Bd1152, suggesting that they might have analogous interacting partners.

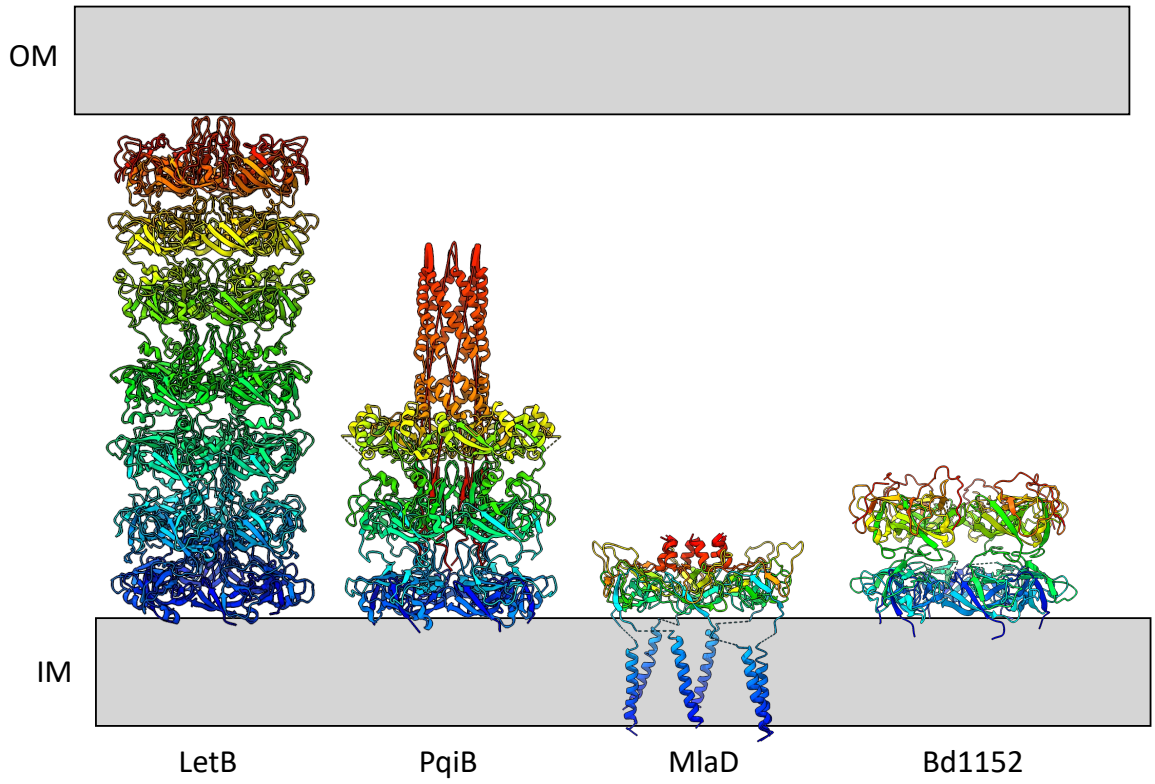


Figure 7.2. Cartoons representing structures of MCE proteins in their respective cellular locations.

Structures depicted in rainbow colouring (blue to red, n- to c-terminus). MlaD (PDB: 7CGE): PqiB (PDB: 5UVN): LetB (PDB: 6V0C). The models were visualised with ChimeraX (Pettersen *et al.,* 2021).

Nonetheless, the c-terminal loop folding back over the second MCE domain appears to be highly conserved among 2-MCE domain-containing proteins, with Alphafold predictions revealing that some species display up to a ~30aa long c-terminus region (Figure 7.3). The Alphafold per residue confidence score for the c-terminal region is very low (pLDDT < 50) (Figure 7.3) implying that this region is poorly modelled potentially suggesting its adoption of a unique architecture not currently captured within PDB entries.

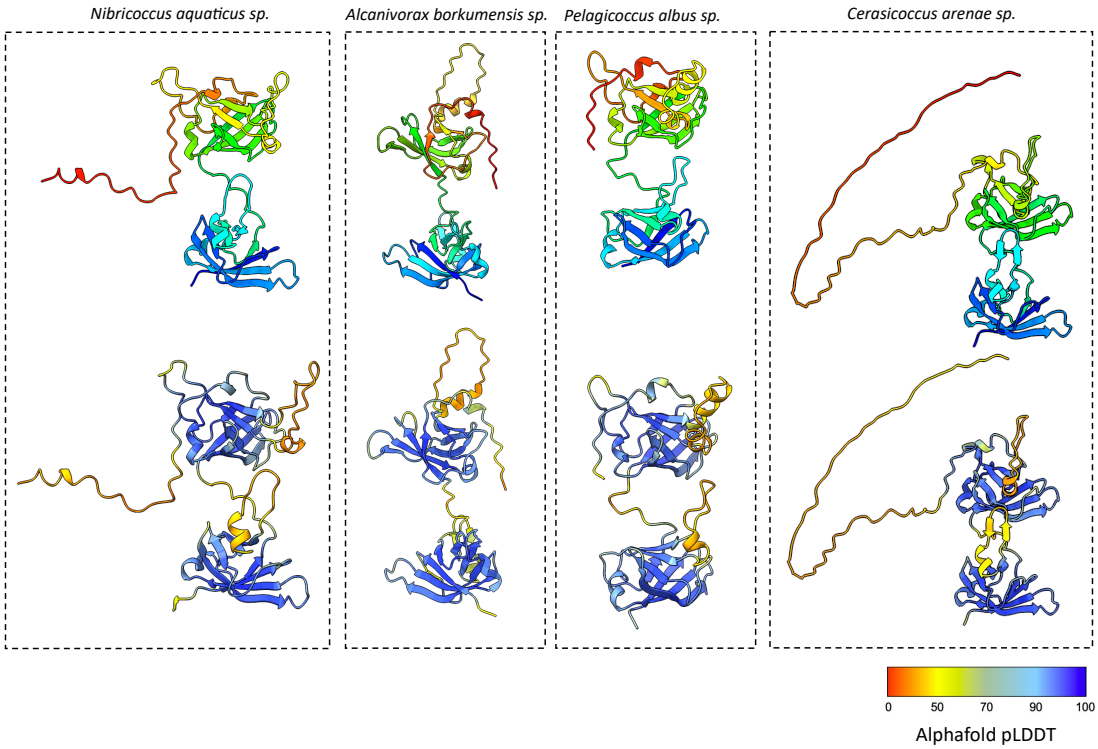


Figure 7.3. The Alphafold structures of 2-MCE domain-containing proteins from different bacterial species.

(Top) The structures depicted in a rainbow colouring (blue to red, c- to c-terminus). Transmembrane helices were deleted to aid figure clarity (*Alcanivorax borkumensis sp*. residues 1-52; *Nibricoccus aquaticus sp.* – residues 1-37; *Pelagicoccus albus sp*. residues 1-37; *Cerasicoccus arenae sp*.- residues 1-44). (Bottom) The structures are coloured by per-residue pLDDT scores. A pLDDT colour key is provided at the bottom right of the figure where the spectrum from blue to red corresponds to decreasing pLDDT score (Jumper *et al.,* 2021). The models were visualised with ChimeraX (Pettersen *et al.,* 2021).

The novel nature of the c-terminus of 2-MCE domain-containing proteins implies its importance in their function. Indeed, further studies relating to the c-terminal region may provide further insight into the mechanics and/or interacting partners. The c-terminal region of Bd1152 contains two Tryptophan residues (W277 and W274) which may provide ideal sites for the incorporation of the non-native amino acid p-benzoyl-L-phenylalanine (BPA). BPA forms photo-inducible crosslinks to nearby C-H bonds and thus may be utilised to identify both substrates and partner proteins (Chin *et al.,* 2002; Krishnamurthy *et al.,* 2011; Isom *et al.,* 2020). Incorporation of BPA within the c-terminus of Bd1152 may provide insight into the identity of the proteins interacting with this region of Bd1152.

7.1.3. The role of Bd1151 and Bd1153

A major consideration of this study is that Bd1152 was expressed and purified either in isolation or as a part of incomplete subcomplex, thus the structure of its operonic components Bd1151 and Bd1153 and their interaction, remain elusive.

7.1.3.1. Bd1151 - small membrane protein.

The structure and function of Bd1151 remains completely enigmatic. Bd1151 is a small 53aa protein with Alphafold showing a single transmembrane helix arrangement (Figure 7.4). Whilst we attempted to express the full length Bd1151 construct solubilised in DDM, purification showed very low levels of the protein, which would immediately aggregate upon elution from the column (Appendix 10).

Bd1151 could be assigned to the novel category of integral membrane proteins, termed small membrane proteins (SMPs). They are defined as being less than 50 amino acids in length and therefore adopt a single transmembrane helix conformation (Storz *et al.,* 2014). The function of SMPs is largely enigmatic. They closely resemble membrane-active antimicrobial peptides (AMPs), however, differently to AMPs, SMPs are believed to be retained within the cell membrane (Alix and Blanc-Potard, 2009; Ghosh *et al.,* 2019). Certain SMPs have been proposed to act as assembly factors for large oligomeric membrane complexes. For instance, in *E. coli*, the KdpABC ATPase complex requires the SMP KdpF (31aa) for its stabilisation, as demonstrated by Gaßel et al. (1999). These observations raise the possibility that Bd1151 may play a role in stabilising the Bd1153-Bd1152 complex. Our investigations revealed that expression and purification of the full Bd1153-Bd1151 complex yielded substantially elevated levels of Bd1152 (Appendix 11) when compared to the Bd1153-Bd1152 construct (Figure 4.9), potentially reinforcing its stabilising function. Additionally, an

anti-his western blot revealed a 31-kDa band (consistent with the size of Bd1152) despite the fact that his tag is located on the c-terminus of Bd1151 (Appendix 11, E). This finding suggests that Bd1151 may indeed interact with Bd1152 and that the complex is able to remain stable even under denaturing conditions. Nevertheless, the cryo-EM screening revealed highly heterogenous sample (Appendix 12). Future studies should aim to increase the purity and stability of the sample subsequently determining the overall Bd1153-Bd1151 structure.

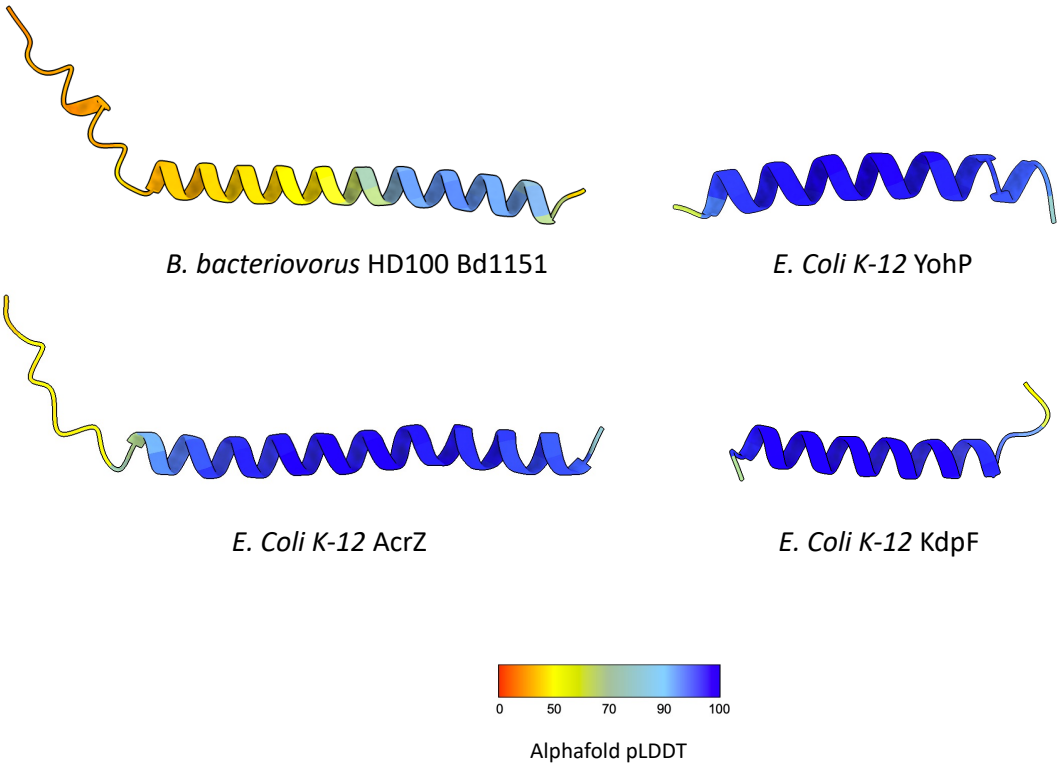


Figure 7.4. Alphafold predictions of several small membrane proteins.

B. bacteriovorus HD100 Bd1151 (AF-Q6MNT5-F1), *E. coli K-12* YhoP (AF-C1P609-F1), *E. coli K-12* AcrZ (AF-P0AAW9-F1); *E. coli K-12* KdpF (AF-P36937-F1). The structures are coloured by per-residue pLDDT scores. A pLDDT colour key is provided at the bottom of the figure where the spectrum from blue to red corresponds to decreasing pLDDT score (Jumper *et al.,* 2021). The models were visualised with ChimeraX (Pettersen *et al.,* 2021).

7.1.3.2. Bd1153 – a functional analogue of E.coli PqiA and LetA?

Most of the 2-MCE domain-containing protein operons, including Bd1152, harbour homologs to the *E. coli* PqiA inner membrane protein which as a member of a PqiABC pathway (Figure 7.1). One of the most striking features of both PqiA and the homologous *E. coli* LetA proteins is that they possess 15 and 11 cysteine residues, respectively.

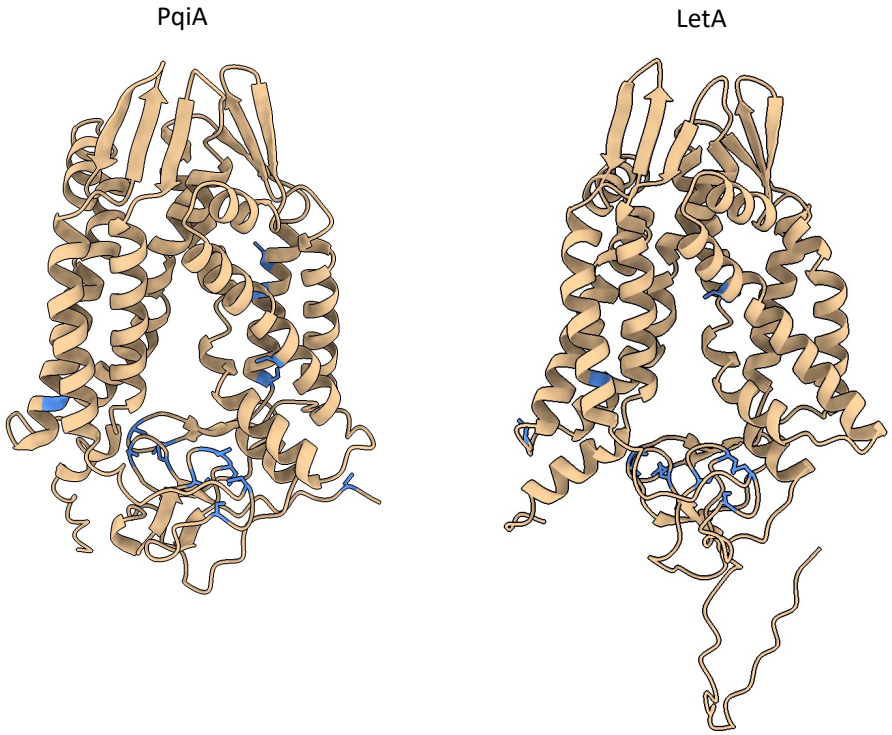


Figure 7.5. The Alphafold structures of *E. coli* PqiA (AF-P0AFL9-F1) and LetA (AF-P0AD03-F1) with cysteine residues indicated in blue.

The models were visualised with ChimeraX (Pettersen *et al.,* 2021).

Such a large number of cysteines in both proteins strongly imply their functional importance. As suggested by Cooper (2023), cysteine residues might have a role in metal ion coordination. Indeed, cysteine residues can display a high affinity to zinc ions forming cysteine-zinc complexes that could be essential for protein structure and regulation (Pace and Weerapana, 2014). Bd1153 itself comprises a single PqiA domain and thus implies functional homology to the PqiA and LetA proteins, however, their wider structural similarities and differences are intriguing. The Bd1153 Alphafold model closely resembles half of the PqiA and LetA Alphafold models with an RMSD of 0.949 Å across 120 pruned atom pairs for PqiA and RMSD of 1.179 Å across 135 pruned atom pairs for LetA (Figure 7.6). However, Bd1153 is composed of only a single lobe (corresponding to its single PqiA domain) and lacks the loop region at the base of the *E. coli* proteins where the majority of their cysteine residues are localised. If the Bd1153-Bd1151 protein complex is indeed a functional homolog of PqiABC/LetAB, then an obvious question arises – does Bd1153 dimerise in *in-vivo*? Furthermore, the reason for the lack of the cysteine bundle region in Bd1153 is unclear. If this region is involved in the regulation of the *E. coli* proteins, then it implies that Bd1153 is regulated via a different manner. Future research should focus on understanding both the structure and function of the PqiA, LetA and Bd1153 proteins consequently providing greater insight into their relationship and mechanisms.

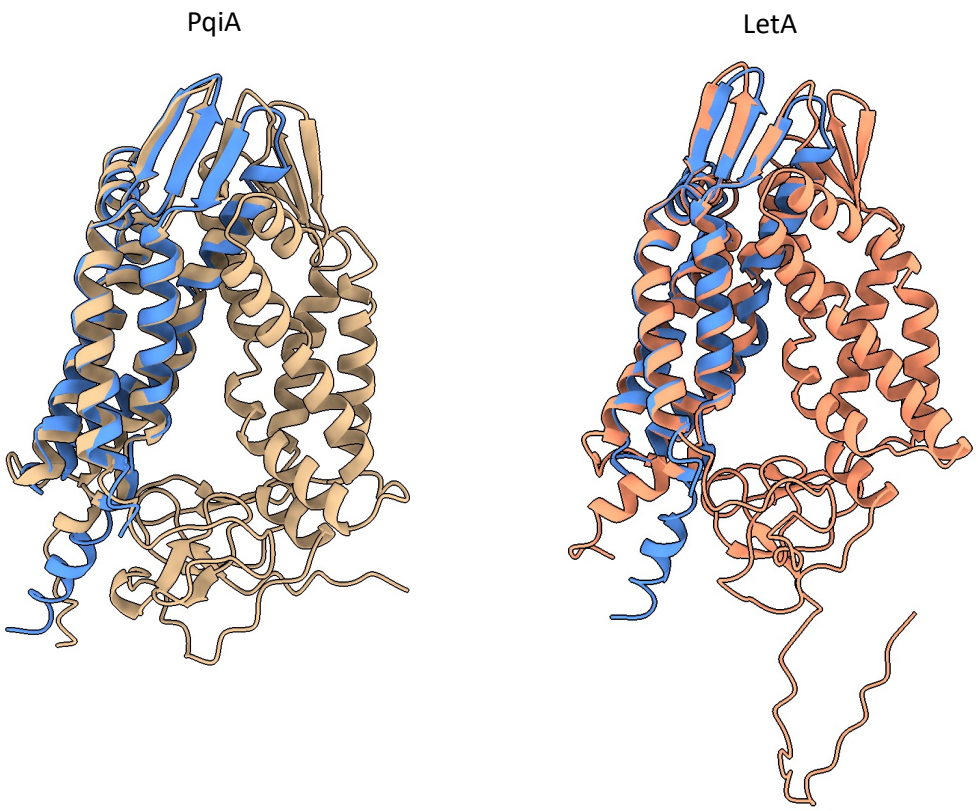


Figure 7.6. Structural comparison of Bd1153 with the *E.coli* PqiA and LetB proteins. The Alphafold structure of Bd1153 (blue) superimposed with the Alphafold structures of PqiA (AF-P0AFL9-F1) (coloured in tan) (RMSD of 0.949 Å across 120 pruned atom pairs (across all 164 pairs: 2.671)) and LetA (AF-P0AD03-F1) (coloured in orange) (RMSD of 1.179 Å across 135 pruned atom pairs (across all 167 pairs: 3.239)). The models were visualised with ChimeraX (Pettersen *et al.,* 2021).

7.1.4. Other MCE domain-containing proteins within *Bdellovibrio bacteriovorus* HD100.

Apart from Bd1152, *Bdellovibrio bacteriovorus* HD100 contains four more MCE domain-containing proteins within its genome that are upregulated and downregulated during different life stages (refer to Chapter 3), implying that MCE-proteins might have versatile functionality. While the initial aim of the current project was to study several *B. bacteriovorus* MCE proteins, experimental complexity has persisted and precluded their structural characterisation. Expression and purification data for the Bd3410-Bd3413 membrane-associated complex and soluble Bd1770 proved to be promising (Appendix 13 and 14, respectively). However, these proteins appeared to be highly unstable with negative stain TEM of a soluble Bd1770 construct revealing a heterogeneous sample (Appendix 15). Future research should aim to investigate other MCE domain-containing proteins and their associated complexes within *B. bacteriovorus* genome potentially revealing novel complex arrangements and interacting partners.

7.2. Concluding remarks.

Overall, this study has provided the first structural and function insights into the MCE proteins of *Bdellovibrio bacteriovorus*, including the first structure of 2-MCE domain-containing protein. Although the investigation primarily focused on Bd1152, the expression, purification and structural determination methods employed may be transferred to investigate other MCE proteins. It is notable, that our results highlight the challenges in crystallography and offer potential strategies to enhance crystal quality. Furthermore, we introduced non-canonical cryo-EM data processing approaches, which could prove beneficial for solving structures of other 2-MCE domain-containing proteins, or proteins exhibiting similar challenges in their biophysical characterisation. Moreover, the utilisation of novel flexibility algorithms provided us with a unique insight into the global dynamics of Bd1152, which may be applicable to other MCE proteins.

Nevertheless, a number of fundamental uncertainties persist, particularly the interaction of MCE domain-containing proteins with their operonic components. Future research should strongly focus upon characterising the entire Bd1153-Bd1151 complex architecture, potentially providing novel insight into intermembrane lipid transport and the understanding of outer membrane biogenesis.

Appendix

Appendix 1. Neighbouring gene analysis of MCE proteins in *B. bacteriovorus* HD100.

Red highlighting indicates the MCE domain-containing proteins. Homologous proteins and sequence identity were identified via BLASTP analysis.

Gene ID	Location	Description	Pfam identified domains	Homologous protein	Sequence identity
Operon 578					
Bd1151	1092944..1093105	Hypothetical protein	No predicted domains	No homologous proteins	n/A
Bd1152	1093150..1094001	PqiB family protein; hypothetical protein	MlaD, MlaD	Intermembrane transport protein PqiB (*E. coli*)	31.30%
Bd1153	1093973..1094482	paraquat-inducible protein A	PqiA	PqiA/YebS family transporter subunit (*E. coli*)	37.95%
Operon 888					
Bd1768	1701560..1702384	ABC transporter permease protein	MlaE	MlaE family lipid ABC transporter permease subunit *(E. coli)*	31.44%
Bd1769	1702384..1703094	ABC transporter ATP-binding protein	ABC Tran	phospholipid ABC transporter ATP-binding protein MlaF (*E. coli*)	26.07%
Bd1770	1703091..1704044	ABC-type multidrug transport system substrate-binding protein	MlaD	No significant similarity found	n/A
Operon 1612					
Bd3116	3020558..3021064	Bacterial transferase hexapeptide repeat; hypothetical protein, cysE	Hexapep, Hexapep	serine O-acetyltransferase (*E. coli*)	55.78%
Bd3117	3021077..3022039	HD-hydrolase domain containing protein	HD, tRNA anti codon	3'-5' exoribonuclease YhaM (*E. coli*	32.86%
Bd3118	3022039..3022695	hypothetical protein predicted by Glimmer/Critica	AhpC-TSA 2	No significant similarity found	n/A
Bd3119	3022692..3024464	hypothetical protein predicted by Glimmer/Critica	TRP19	nitrite reductase small subunit NirD (*E. coli*)	42.55%

Bd3120	3024468..3025844	ABC-type organic solvent transporter, solute-binding periplasmic protein.	MlaD	No significant similarity found	n/A
Bd3121	3025841..3026605	ABC-type organic solvent transport system ATP-binding protein	ABC Tran	Phospholipid ABC transporter ATP-binding protein MlaF (*E. coli*)	36.65%
Bd3123	3026605..3027432	ABC-type organic solvent resistance transport system permease protein	MlaE	Lipid asymmetry maintenance ABC transporter permease subunit MlaE (*E. coli*)	33.61%
Bd3124	3027429..3028637	Alanine racemase; hypothetical protein	Ala racemase N, Ala racemase C.	Alanine racemase (*E. coli*).	32.05%
Bd3125	3028647..3030137	sensor histidine kinase	GGDEF	diguanylate cyclase (*E. coli*).	26.83%
Bd3126	3030138..3032390	sensor histidine kinase	HisKa, HATPase c	Sensory histidine kinase Atos (*E. coli*).	30.41%
Operon 1750					
Bd3410	3322456..3322947	Hypothetical protein	No predicted domains	No significant similarity found	n/A
Bd3411	3322947..3323735	ABC transporter, permease protein	MlaE	lipid asymmetry maintenance ABC transporter permease subunit MlaE (*E. coli*).	30.88%
Bd3412	3323732..3324478	ABC-type organic solvent resistance transport system ATP-binding protein	ABC trans	phospholipid ABC transporter ATP-binding protein MlaF (*E. coli*)	33.74%
Bd3413	3324478… 3325293	ABC-type transport system involved in resistance	MlaD	Outer membrane lipid asymmetry maintenance protein MlaD (*Escherichia coli*)	28.05%
Operon 1958					
Bd3802	3671103..3671753	Hypothetical protein predicted by Glimmer/Critica	No predicted domains	No significant similarity found	n/A

Bd3803	3671782..3672558	ABC-type organic solvent resistance transport system, permease protein	MlaD	Outer membrane lipid asymmetry maintenance protein MlaD (*E. coli*)	26.61%
Bd3804	3672567..3673286	ABC-type organic solvent resistance transport system, permease protein	MlaE	Lipid asymmetry maintenance ABC transporter permease subunit MlaE (*E. coli)*	34.84%
Bd3805	3673283..3674062	ABC-type organic solvent resistance transporter, permease protein.	MlaE	Lipid asymmetry maintenance ABC transporter permease subunit MlaE	27.75%
Bd3806	3674059..3674790	ABC-type organic solvent resistance transport system ATP-binding protein	ABC trans	phospholipid ABC transporter ATP-binding protein MlaF (*E. coli*)	39.24%
Bd3807	3674837..3675595	ABC transporter, ATP-binding protein	ABC trans	phospholipid ABC transporter ATP-binding protein MlaF *(E. coli*)	32.93%

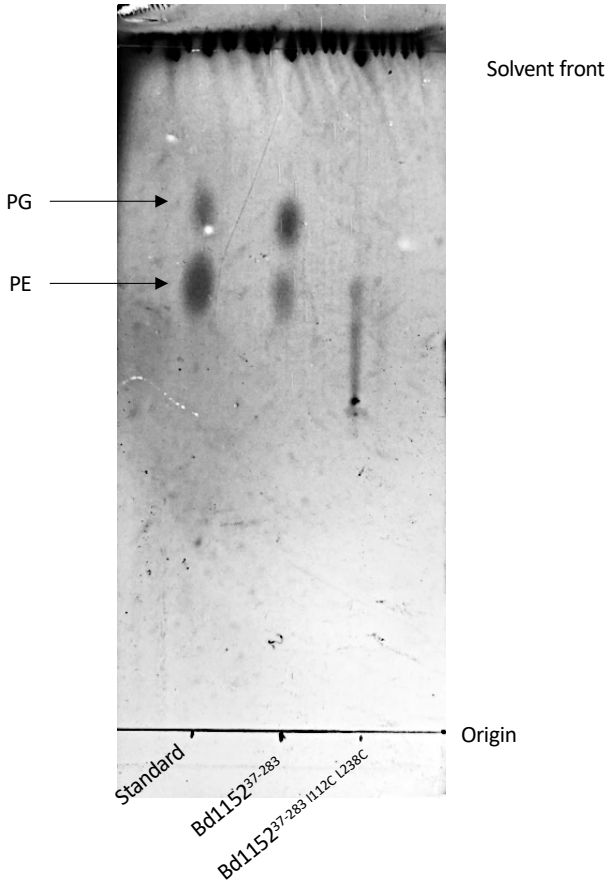
Appendix 2. The data used to generate the Sankey diagram demonstrating the distribution of 2-MCE domain-containing proteins across all kingdoms of life, as identified by FoldSeek analysis.

(Refer to Figure 3.4).

Source	Target	No of Genes
Bacteria	Pseudomonadota	158
Bacteria	Planctomycetota	16
Bacteria	Bdellovibrionota	12
Bacteria	Verrucomicrobiota	10
Bacteria	Thermodesulfobacteriota	1
Bacteria	Chloroflexota	1
Bacteria	Bacillota	1
Bacteria	Cyanobacteriota	1
Eukaryota	Viridiplantae	1
Eukaryota	Fungi	1
Pseudomonadota	Rhodanobacteraceae	78
Pseudomonadota	Enterobacteriaceae	33
Pseudomonadota	Alcanivoracaceae	17
Pseudomonadota	Burkholderiaceae	4
Pseudomonadota	Pseudomonadaceae	3
Pseudomonadota	Thalassospiraceae	2
Planctomycetota	Planctomycetaceae	3
Planctomycetota	Phycisphaerae bacterium	5
Planctomycetota	Planctomycetota bacterium	4
Bdellovibrionota	Pseudobdellovibrionaceae	12
Verrucomicrobiota	Opitutaceae	5
Fungi	Mucoromycota	1
Rhodanobacteraceae	Rhodanobacter	38
Rhodanobacteraceae	Dyella	21
Rhodanobacteraceae	Luteibacter	14
Rhodanobacteraceae	Frateuria	3
Enterobacteriaceae	Escherichia	12
Enterobacteriaceae	Salmonella	9
Enterobacteriaceae	Klebsiella	9
Alcanivoracaceae	Alcanivorax	16
Pseudobdellovibrionaceae	Bdellovibrio	11
Burkholderiaceae	Burkholderia	3
Rhodanobacter	Rhodanobacter sp.	6
Rhodanobacter	Rhodanobacter glycinis	3
Alcanivorax	Alcanivorax sp.	4
Luteibacter	Luteibacter yeojuensis	3
Escherichia	Escherichia coli	12
Bdellovibrio	Bdellovibrio bacteriovorus	6
Salmonella	Salmonella enterica	9
Klebsiella	Klebsiella pneumoniae	7

Appendix 3. Thin layer chromatography of lipids extracted from the Bd1152$^{37\text{-}283}$ and Bd1152^{2C} samples.

PG – phosphatidylglycerol, PE – phosphatidylethanolamine. Lipids are marked on the left according to a *E. coli* polar lipid standard (lane 1).



Appendix 4. Extended cryo-EM data processing workflow (extension of Figure 4.20) of Bd1152^{2C} within *Scipion*3 framework.

A) Particle picking approaches employed in the cryo-EM data processing pipeline:

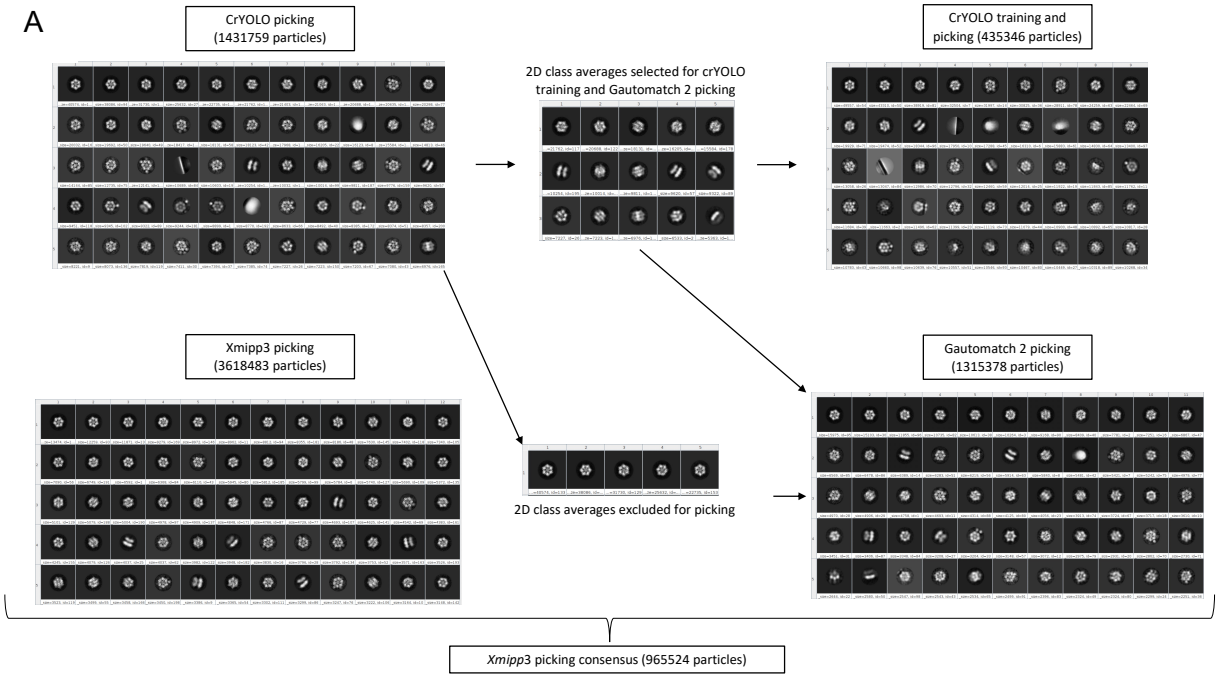
crYOLO picking was performed. The 15 best 2D class averages were selected as 2D templates for crYOLO training (fine-tune previous model set to Yes), followed by crYOLO picking utilising the resulting model. The same 2D class averages obtained from crYOLO picking were used as 2D templates for Gautomatch picking. Moreover, four class averages representing top view orientations were utilised as templates to exclude top view orientations during picking. Finally, *Xmipp*3 picking was performed. The resulting particles were subjected to the *Xmipp*3 particle consensus protocol to join particle sets (duplicates removed). We selected *Xmipp*3 particle consensus value of 1, meaning that if at least one algorithm has picked the particle, it will be kept (refer to Figure 4.20[1])

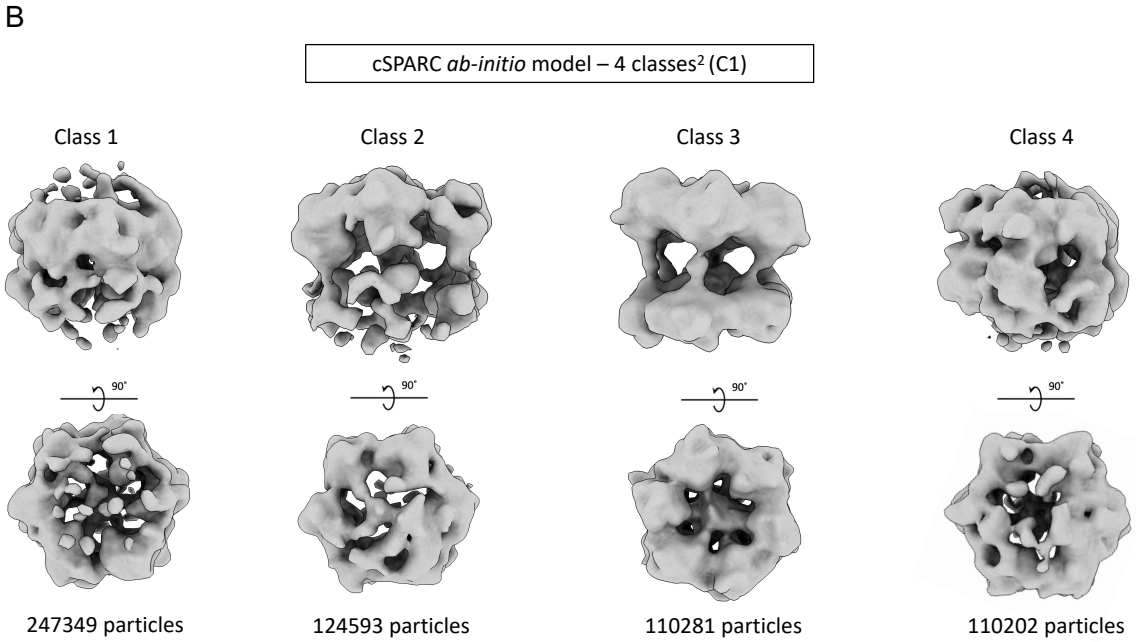
B) Four cryo-EM maps obtained from cryoSPARC *ab-initio* reconstitution in C1 symmetry utilising the 592425-particle set. Class 2 and Class 3 were selected for further processing (refer to Figure 4.20[2]). All maps are displayed at a contour level of 0.05.

C) and D) Cryo-EM maps obtained from cryoSPARC *ab-initio* reconstitution in C1 symmetry utilising the 23474-particle set (refer to Figure 4.21[3] and Figure 4.20[4]). All maps are displayed at the contour level of 0.05.

E) and F) The cryo-EM maps obtained from cryoSPARC heterogenous refinement in C1 symmetry utilising the *ab-initio* models displayed in C and D (refer to Figure 4.20[5;6]). The model displayed in Appendix 5 was used the reference model. All maps are displayed at a contour level of 0.05.

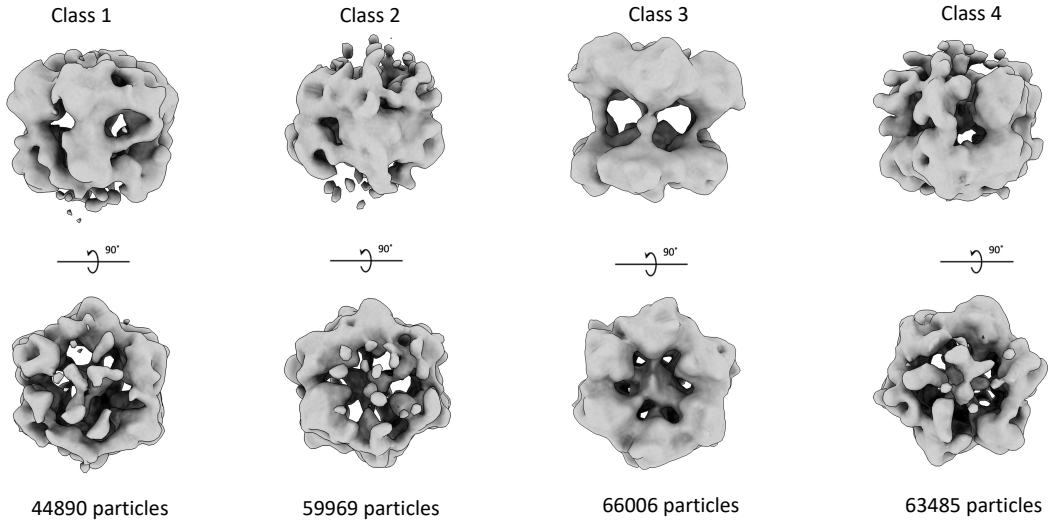
The density maps were visualised with ChimeraX (Pettersen *et al.,* 2021).


A
CrYOLO picking
(1431759 particles)
2D class averages selected for crYOLO
training and Gautomatch 2 picking
CrYOLO training and
picking (435346 particles)
Xmipp3 picking
(3618483 particles)
2D class averages excluded for picking
Gautomatch 2 picking
(1315378 particles)
Xmipp3 picking consensus (965524 particles)


B
cSPARC ab-initio model – 4 classes² (C1)
Class 1
Class 2
Class 3
Class 4
90°
90°
90°
90°
247349 particles
124593 particles
110281 particles
110202 particles

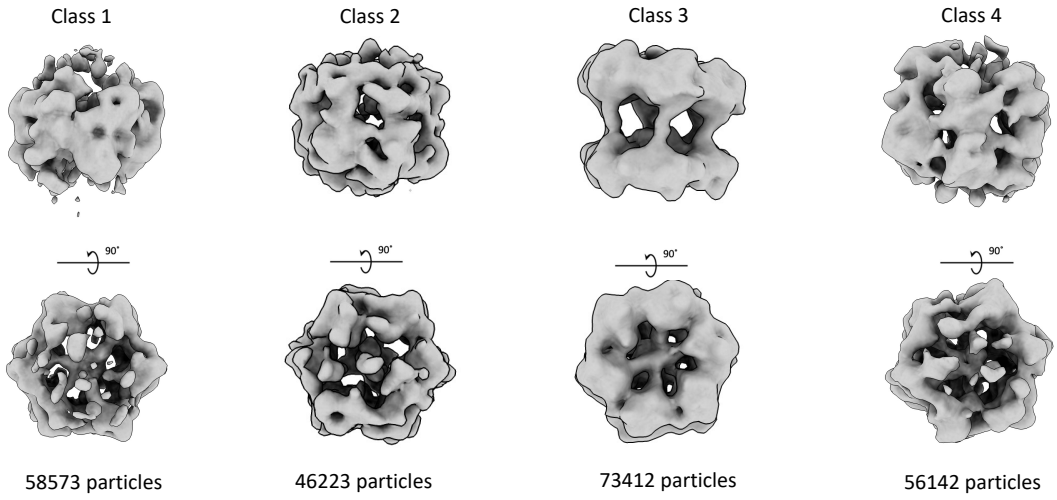
C

cSPARC *ab-initio* model – 4 classes[3] (C1)

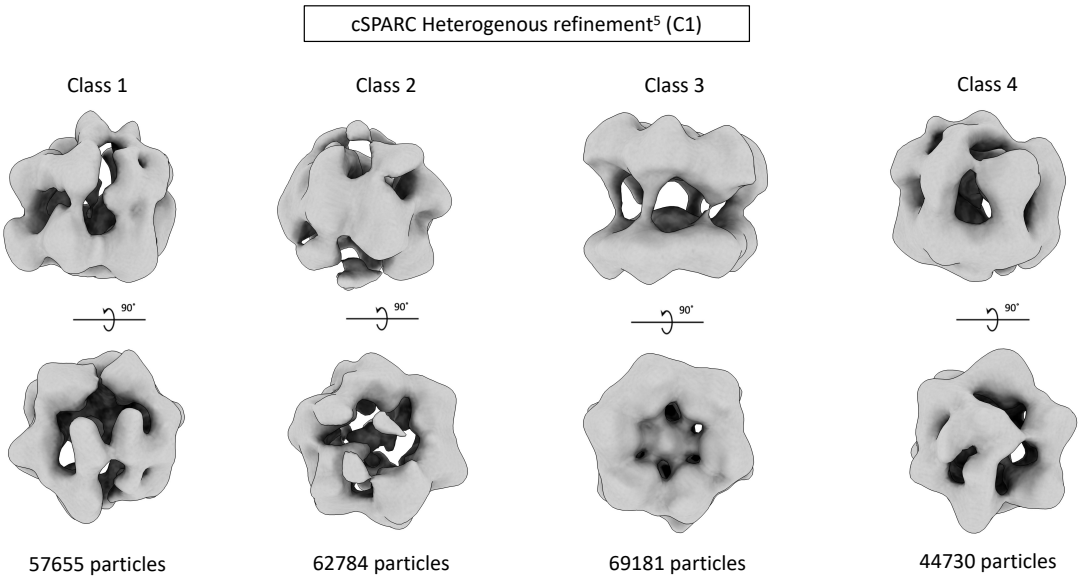


D

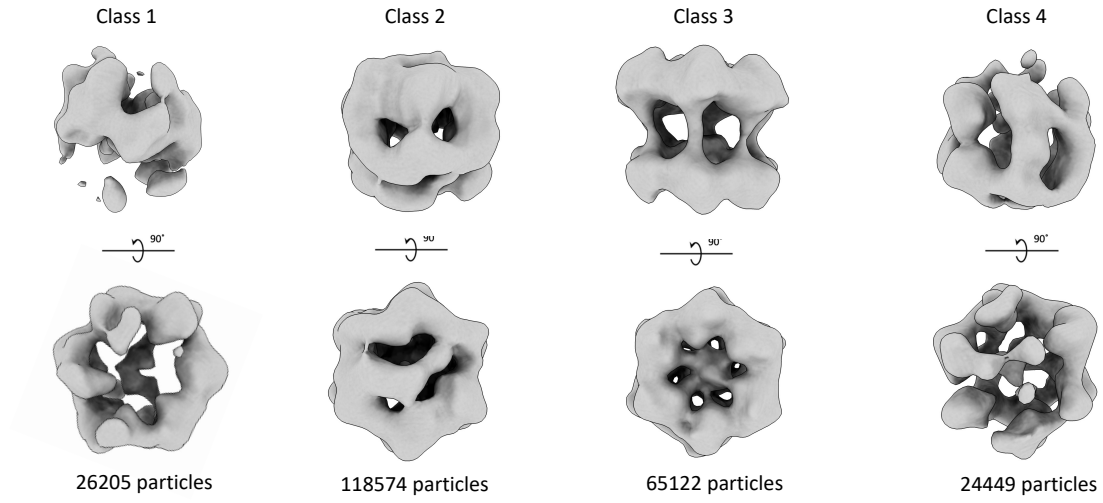
cSPARC *ab-initio* model – 4 classes[4] (C1)



E

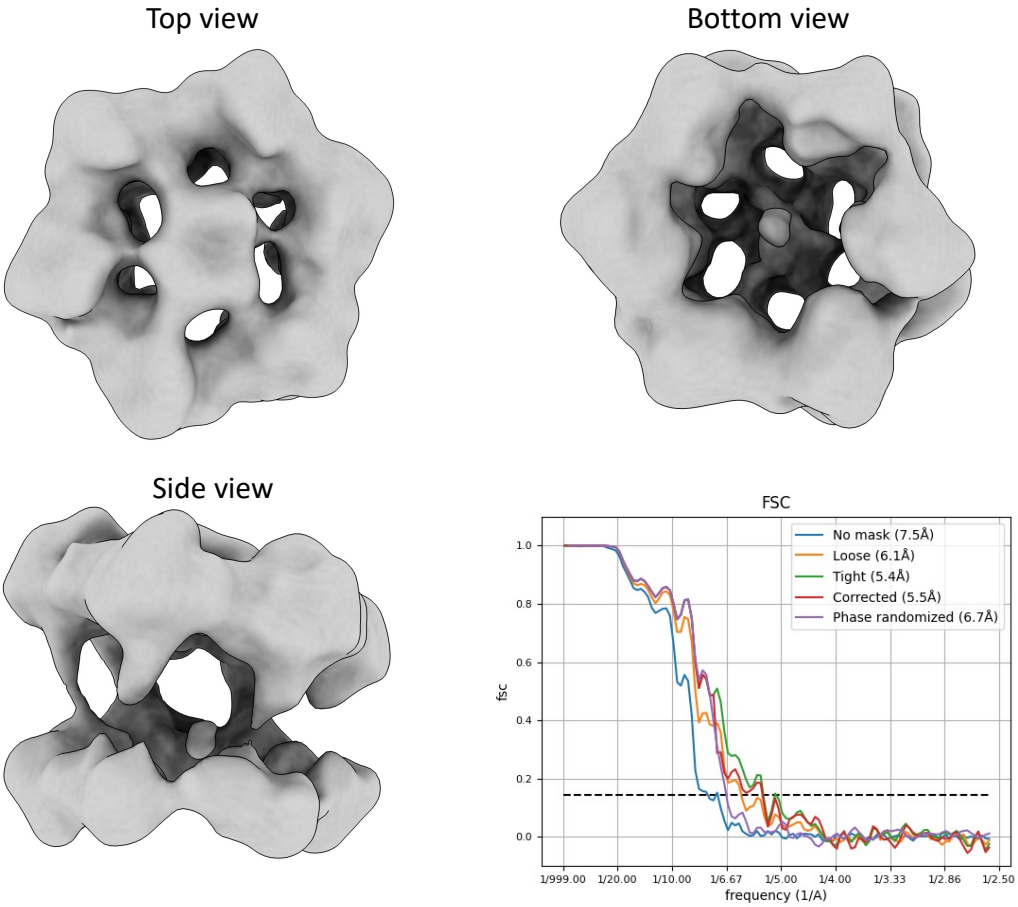

cSPARC Heterogenous refinement5 (C1)
Class 1
Class 2
Class 3
Class 4
90°
90°
90°
90°
57655 particles
62784 particles
69181 particles
44730 particles

F


cSPARC Heterogenous refinement6 (C1)
Class 1
Class 2
Class 3
Class 4
90°
90°
90°
90°
26205 particles
118574 particles
65122 particles
24449 particles

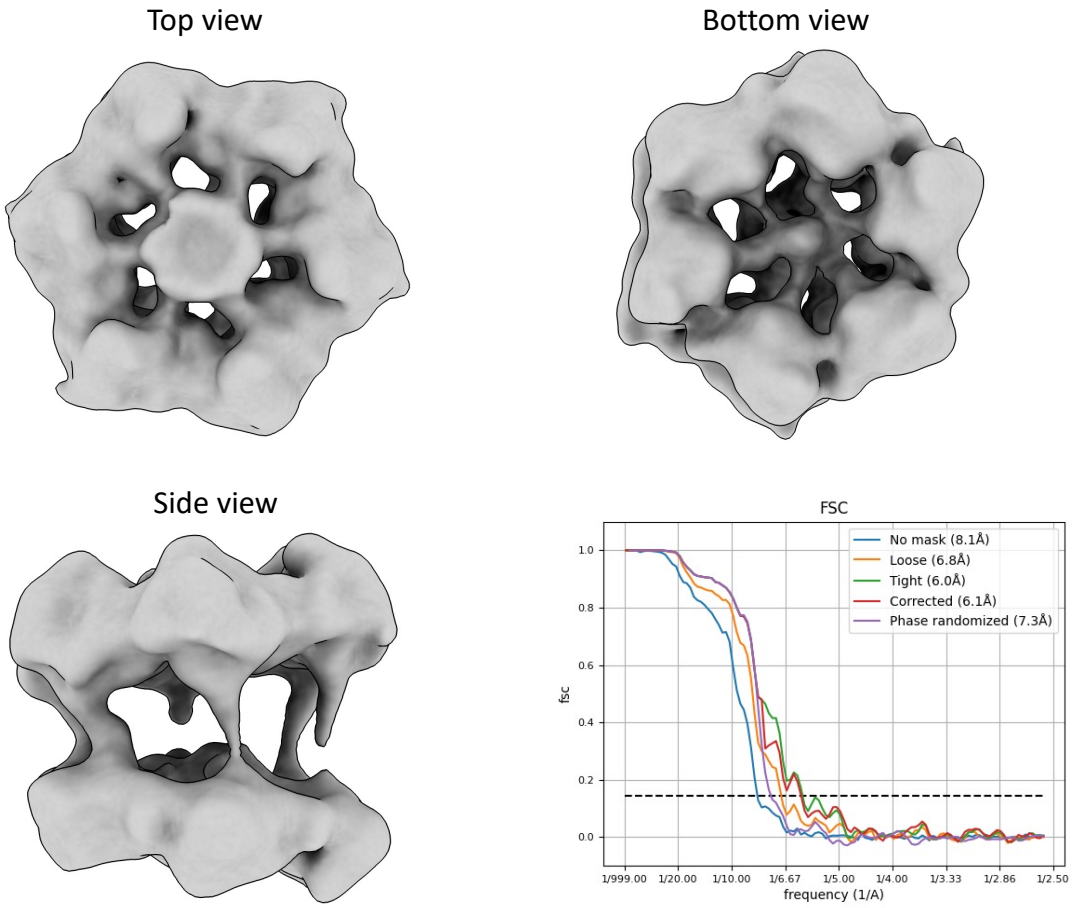
Appendix 5. The cryo-EM map of Bd1152^{2C} used as an input for cryoDRGN analysis (refer to Figure 4.20^{10} and Chapter 6) and as the reference map for cryoSPARC heterogenous refinement (refer to Appendix 3, E, F and Figure 4.20$^{5;6}$).

The average resolution derived from Fourier Shell Correlation (FSC) is shown at the bottom right. FSC plots show correlation between the two independent refined maps (half-maps) with blue (no mask), loose mask (red), tight mask (cyan) and corrected (purple) curves. A cut-off 0.143 (black vertical dotted line) was used for resolution estimation. The map is displayed at the contour level of 0.25. The density map was visualised with ChimeraX (Pettersen *et al.,* 2021).



Appendix 6. The cryo-EM map of Bd1152^{2C} in C1 symmetry (refer to Figure 4.20^{7}).

The average resolution derived from Fourier Shell Correlation (FSC) is shown at the bottom right. FSC plots show correlation between the two independent refined maps (half-maps) with blue (no mask), loose mask (red), tight mask (cyan) and corrected (purple) curves. A cut-off 0.143 (the black vertical dotted line) was used for resolution estimation. The map is displayed at the contour level of 0.25. The density map was visualised with ChimeraX (Pettersen *et al.,* 2021).



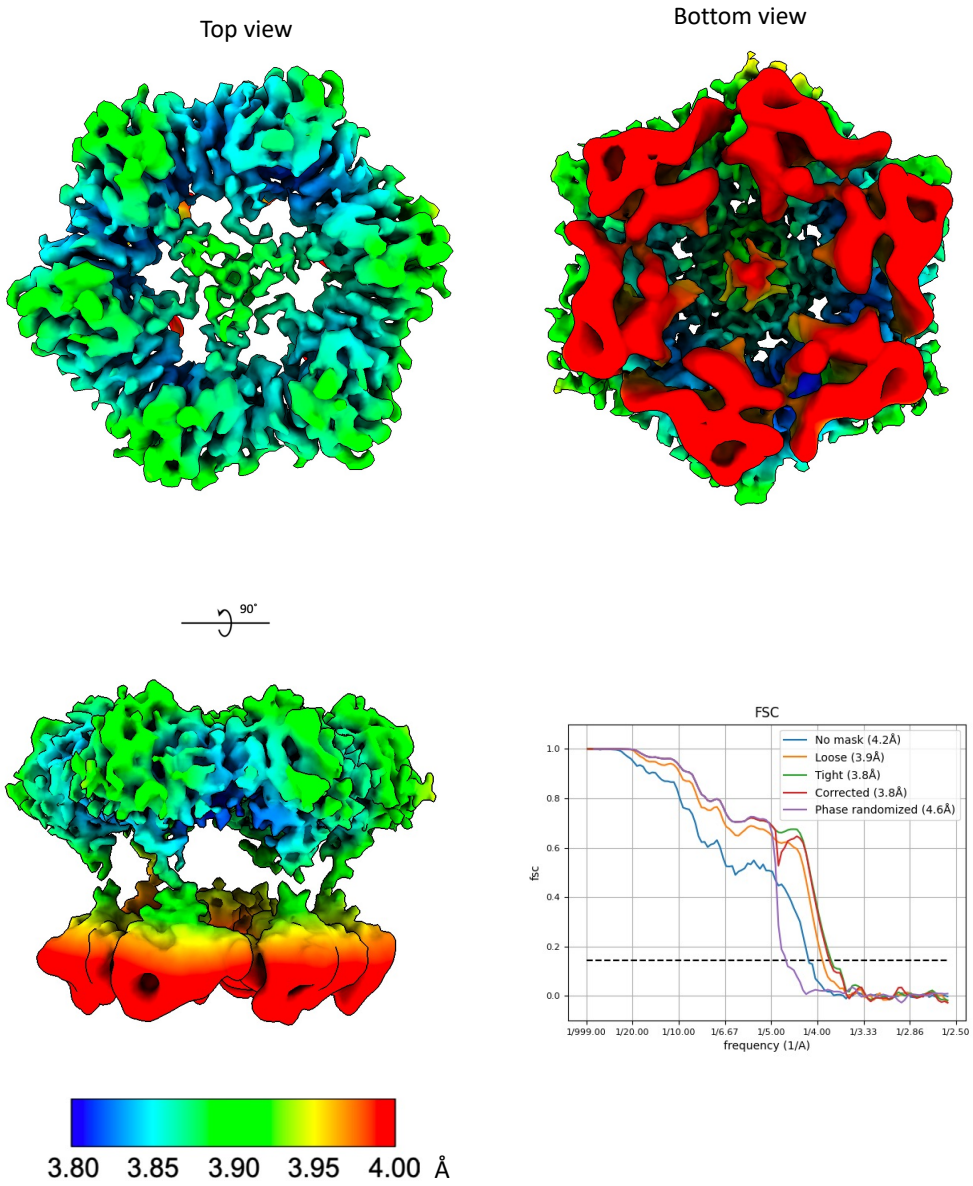
Appendix 7. Refinement statistics for the cryoEM structure of Bd1152^{2C} top MCE ring, related to Figure 4.22.

The model of the top MCE domain of Bd1152^{2C} was built using ModelAngelo algorithm and refined with Phenix real_space_refine algorithm.

Composition (#)	
Chains	6
Atoms	4618 (Hydrogens: 0)
Residues	Protein: 586 Nucleotide: 0
Water	0
Ligands	0
Bonds (RMSD)	
Length (Å) (# > 4σ)	0.002 (0)
Angles (°) (# > 4σ)	0.606 (1)
MolProbity score	1.77
Clash score	6.75
Ramachandran plot (%)	
Outliers	0.18
Allowed	5.71
Favoured	94.11
Rama-Z (Ramachandran plot Z-score, RMSD)	
whole (N = 560)	-0.45 (0.37)
helix (N = 0)	--- (---)
sheet (N = 198)	0.50 (0.38)
loop (N = 362)	-0.77 (0.34)
Rotamer outliers (%)	0.00
Cβ outliers (%)	0.00
Peptide plane (%)	
Cis proline/general	0.0/0.0
Twisted proline/general	0.0/0.0
CaBLAM outliers (%)	4.01
ADP (B-factors)	
Iso/Aniso (#)	4618/0
min/max/mean	
Protein	76.39/106.18/87.17
Nucleotide	---
Ligand	---
Water	---
Occupancy	
Mean	1.00
occ = 1 (%)	100.00
0 < occ < 1 (%)	0.00
occ > 1 (%)	0.00
Box	
Lengths (Å)	103.96, 110.22, 47.60
Angles (°)	90.00, 90.00, 90.00
Supplied Resolution (Å)	3.4
Resolution Estimates (Å). Masked Unmasked	
d FSC (half maps; 0.143)	--- ---
d 99 (full/half1/half2)	2.6/---/--- 2.6/---/---
d model	2.2 2.2
d FSC model (0/0.143/0.5)	---/2.0/3.8 ---/2.0/3.8
Map min/max/mean	-0.02/1.68/0.02
Model vs. Data	
CC (mask)	0.75
CC (box)	0.74
CC (peaks)	0.72
CC (volume)	0.75
Mean CC for ligand	---

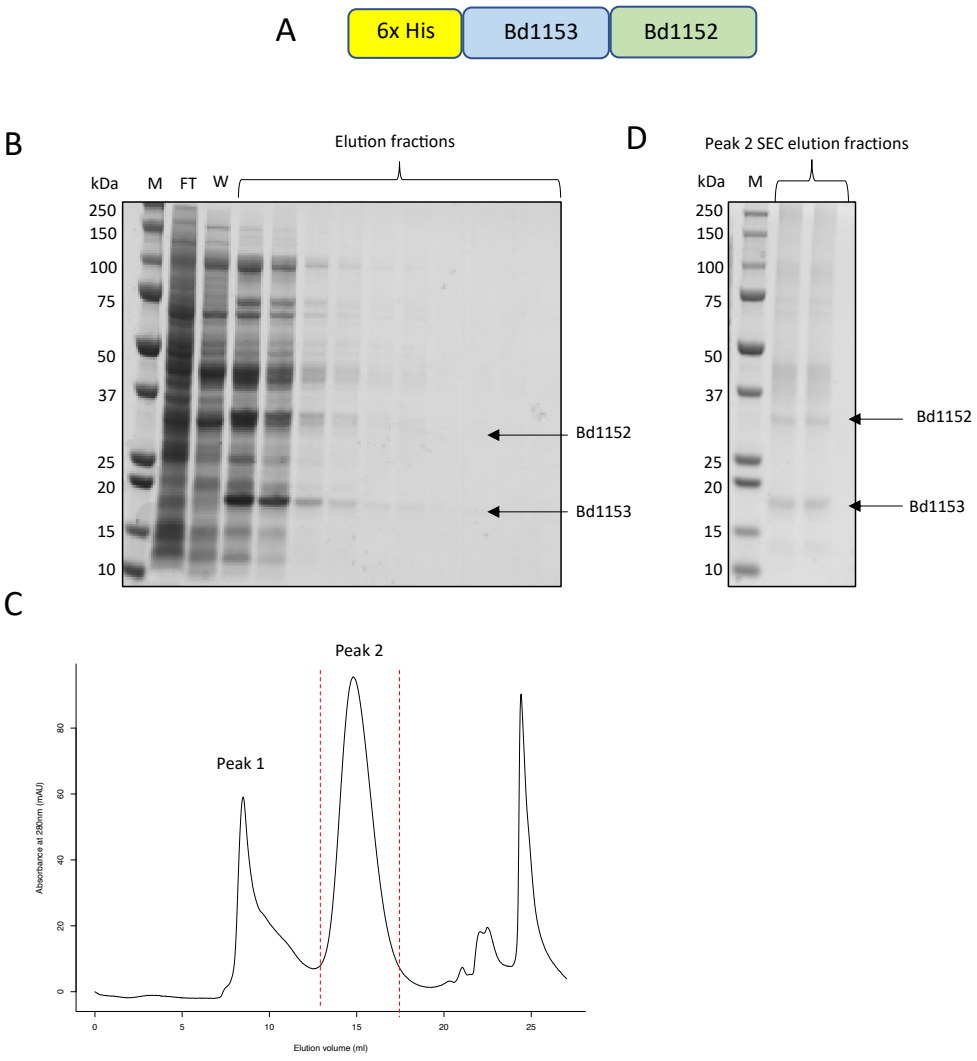
Appendix 8. The cryo-EM map of Bd1152^{2C} at 3.8 Å in C3 symmetry.

The map is coloured according to the local resolution estimation (Å) with the colour key depicted at the bottom left of the figure. (Bottom right) The global resolution derived from Fourier Shell Correlation (FSC). FSC plots show correlation between the two independent refined maps (half-maps) with blue (no mask), loose mask (red), tight mask (cyan) and corrected (purple). A cut-off 0.143 (black vertical dotted line) was used for resolution estimation. The map is displayed at the contour level of 0.16. The density map was visualised with ChimeraX (Pettersen *et al.,* 2021).



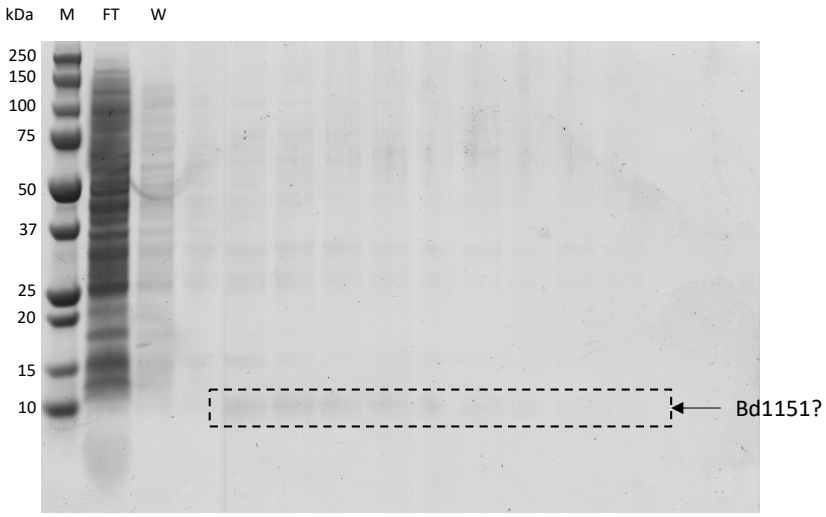
Appendix 9. Purification of the membrane associated Bd1153-Bd1152 complex.

A) The representation of the Bd1153-Bd1152 construct architecture. B) SDS-PAGE of the Bd1153-Bd1152 IMAC purification revealed major bands at ~ ~31 kDa and ~19 kDa, corresponding to Bd1152 and Bd1153, respectively. M- molecular marker, FT –flow-through, W – 50 mM imidazole wash. C) SEC (Superose 6 10/300) exhibited two peaks: the first peak eluted at 6-10 mL, potentially representing aggregated protein. The second peak eluted at ~14-15 mL, likely corresponding to the correctly folded Bd1153-Bd1152 complex. D) SDS-PAGE of the Bd1153-Bd1152 SEC elution fractions showed several bands with ~31 kDa and ~19 kDa, corresponding to Bd1152 and Bd1153, respectively. SDS-PAGE was performed using 4-12% bis-tris gels.



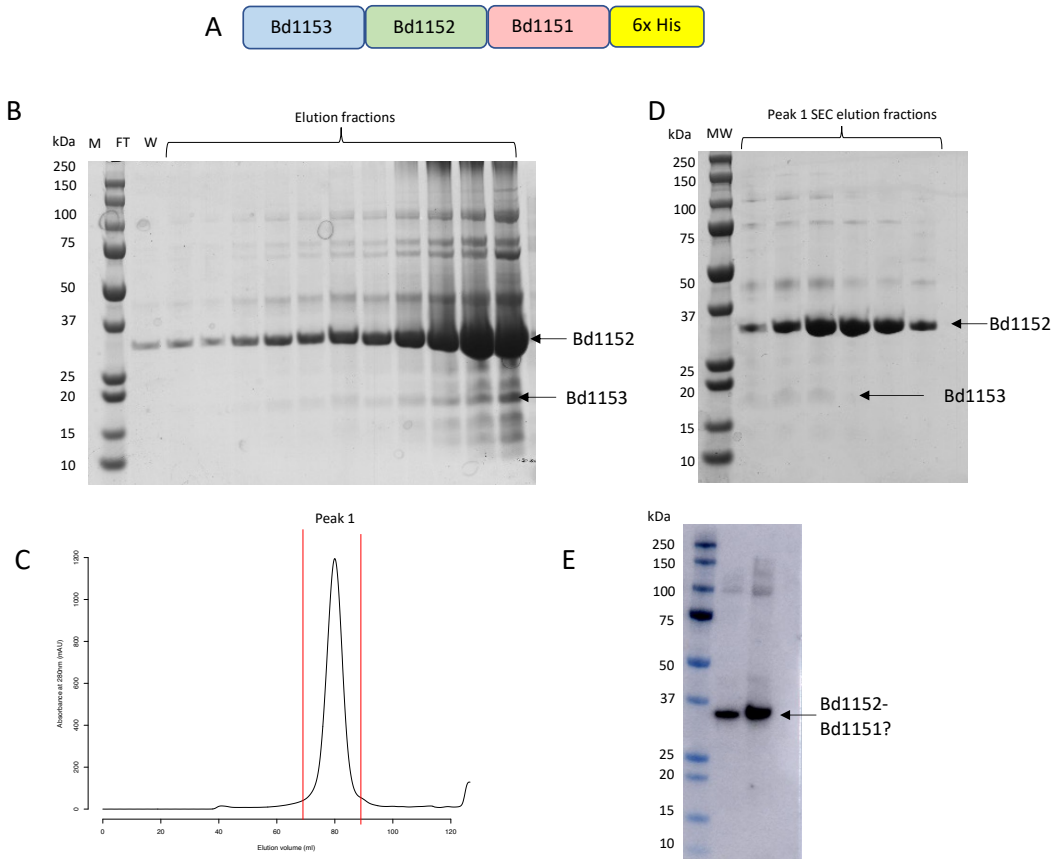
Appendix 10. SDS-PAGE of the Bd1151 IMAC purification.

M- molecular marker. FT – flow-through W – 50mM imidazole wash. A band was observed around ~10 kDa, potentially corresponding to Bd1151. SDS-PAGE was performed using 4-12% Bis-tris gels.

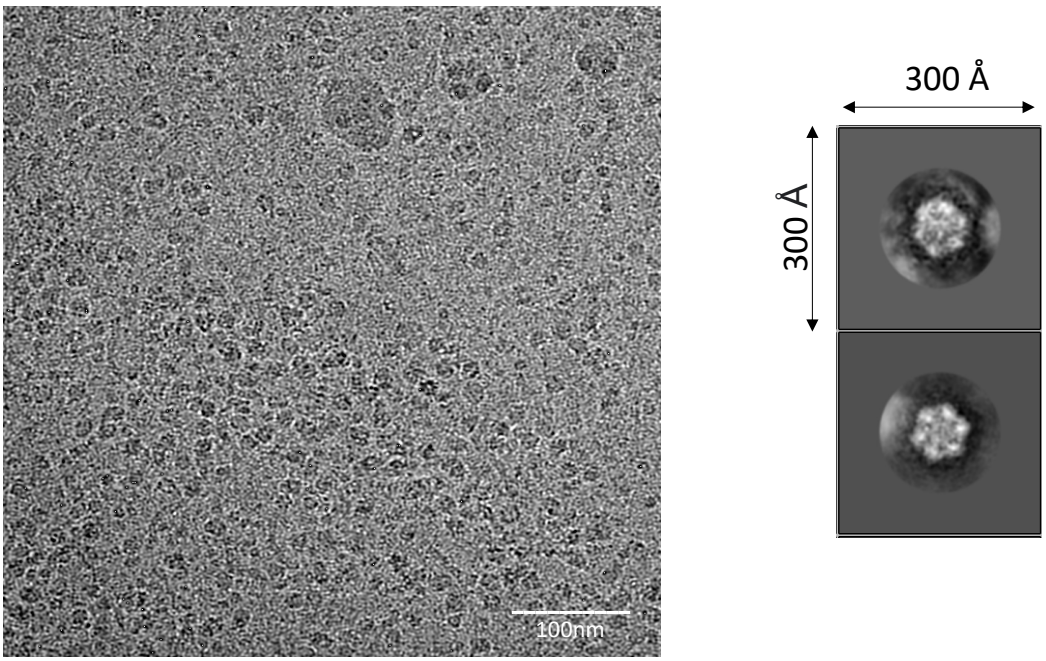


Appendix 11. Purification of the membrane associated Bd1153-Bd1151 complex.

A) Representation of the Bd1153-Bd1151 construct with the c-terminus 6 x histidine tag B) SDS-PAGE of Bd1153-Bd1151 IMAC purification which revealed bands at ~31 kDa and ~19 kDa, corresponding to Bd1152 and Bd1153, respectively. M- molecular marker. FT –flow-through W – 50 mM imidazole wash. C) SEC (Superdex 200 16/600) exhibited a single peak at ~ 80mL. D) SDS-PAGE of the Bd1153-Bd1151 SEC elution fractions revealed several bands, with a band at ~35 kDa and ~19 kDa corresponding to Bd1152 and Bd1153, respectively. E) An anti-his western blot revealed a band of ~35kDa potentially corresponding to the Bd1152-Bd1151 associated complex. SDS-PAGE was performed using 4-12% bis-tris gels.

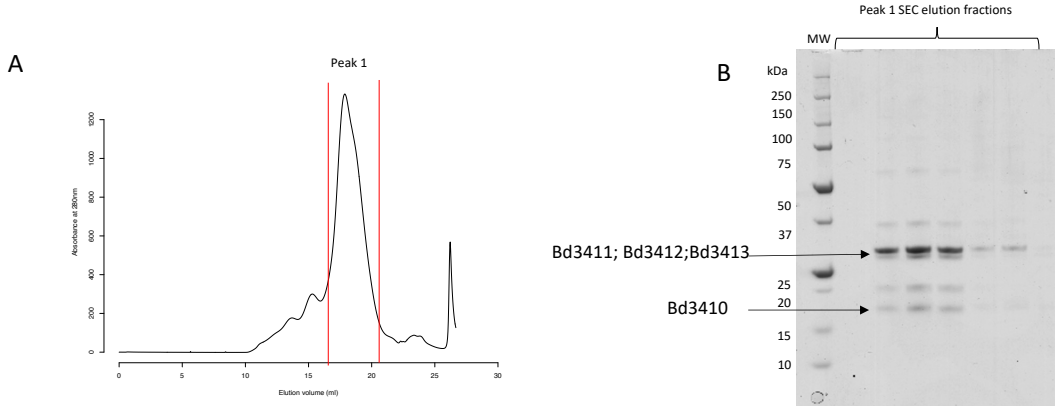


Appendix 12. An example of Cryo-EM micrograph and 2D classification of the Bd1153-Bd1151 membrane associated complex.



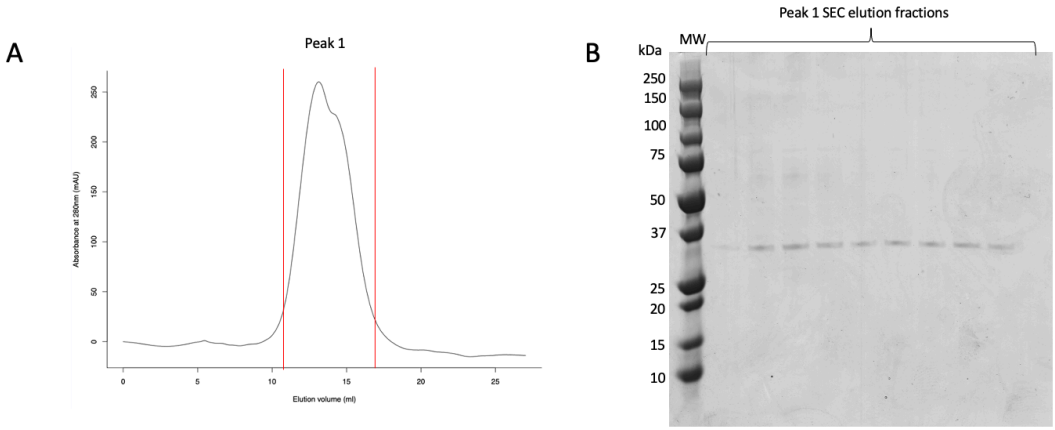
Appendix 13. Purification of the Bd3410-Bd3413 membrane associated complex.

A) SEC (Superose 6 10/300) exhibited a major peak eluting at ~17-21 mL. (B) SDS-PAGE of the Bd3410-Bd3413 SEC elution fractions revealed several bands, with a band of ~30 kDa corresponding to the Bd3412 (27 kDa), Bd3413 (29 kDa), Bd3411 proteins and a band at ~19 kDa corresponding to Bd3410 (17 kDa) (As identified by mass spectrometry results (data not shown)). SDS-PAGE was performed using 4-12% bis-tris gels.



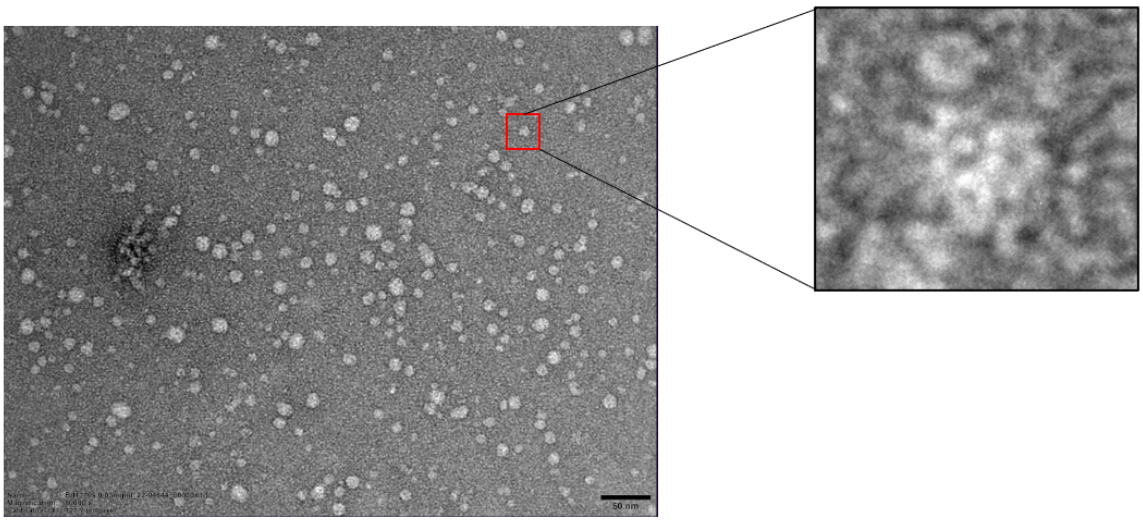
Appendix 14. Purification of Bd1170$^{40-317}$

A) The SEC (Superose 6 10/300) trace of single elution fraction showing a single peak of ~250 mAU. (B) SDS-PAGE of the Bd1770$^{40-317}$ SEC elution fractions showing a single band at ~35 kDa corresponding to Bd1770$^{40-317}$. SDS-PAGE was performed using 4-12% bis-tris gels.



Appendix 15. Negative stain-TEM of Bd1170$^{40-317}$.

Example of a negative stain micrograph of Bd1170$^{40-317}$ showing heterogenous particles. A particle resembling a hexameric assembly and most likely corresponding to Bd1170$^{40-317}$ is depicted on the right. Bd1170$^{40-317}$ sample was stained with 2.5% uranyl acetate. Micrograph was collected at 80k magnification at 120kv on a JEOL1400 TEM.



Bibliography

Abellón-Ruiz, J., Kaptan, S.S., Baslé, A., Claudi, B., Bumann, D., Kleinekathöfer, U., van den Berg, B., 2017. Structural basis for maintenance of bacterial outer membrane lipid asymmetry. Nat Microbiol 2, 1616–1623.

Abrishami, V., Ilca, S.L., Gomez-Blanco, J., Rissanen, I., de la Rosa-Trevín, J.M., Reddy, V.S., Carazo, J.-M., Huiskonen, J.T., 2021. Localized reconstruction in Scipion expedites the analysis of symmetry mismatches in cryo-EM data. Progress in Biophysics and Molecular Biology, CryoEM microscopy developments and their biological applications 160, 43–52.

Adrian, M., Dubochet, J., Lepault, J., McDowall, A.W., 1984. Cryo-electron microscopy of viruses. Nature 308, 32–36.

Akopian, D., Shen, K., Zhang, X., Shan, S., 2013. Signal Recognition Particle: An essential protein targeting machine. Annu Rev Biochem 82, 693–721.

Allen, W.J., Phan, G., Waksman, G., 2009. Structural biology of periplasmic chaperones. Adv Protein Chem Struct Biol 78, 51–97.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W., Lipman, D.J., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25, 3389–3402.

Angelini, S., Deitermann, S., Koch, H.-G., 2005. FtsY, the bacterial signal-recognition particle receptor, interacts functionally and physically with the SecYEG translocon. EMBO Rep 6, 476–481.

Arnold, S.A., Albiez, S., Bieri, A., Syntychaki, A., Adaixo, R., McLeod, R.A., Goldie, K.N., Stahlberg, H., Braun, T., 2017. Blotting-free and lossless cryo-electron microscopy grid preparation from nanoliter-sized protein samples and single-cell extracts. Journal of Structural Biology 197, 220–226.

Arruda, S., Bomfim, G., Knights, R., Huima-Byron, T., Riley, L.W., 1993. Cloning of an M. tuberculosis DNA fragment associated with entry and survival inside cells. Science 261, 1454–1457.

Arunmanee, W., Pathania, M., Solovyova, A.S., Le Brun, A.P., Ridley, H., Baslé, A., van den Berg, B., Lakey, J.H., 2016. Gram-negative trimeric porins have specific LPS binding sites that are essential for porin biogenesis. Proceedings of the National Academy of Sciences 113, E5034–E5043.

Atterbury, R.J., Hobley, L., Till, R., Lambert, C., Capeness, M.J., Lerner, T.R., Fenton, A.K., Barrow, P., Sockett, R.E., 2011. Effects of Orally Administered Bdellovibrio bacteriovorus on the Well-Being and Salmonella Colonization of Young Chicks ▿. Appl Environ Microbiol 77, 5794–5803.

Baarda, B.I., Zielke, R.A., Van, A.L., Jerse, A.E., Sikora, A.E., 2019. Neisseria gonorrhoeae MlaA influences gonococcal virulence and membrane vesicle production. PLOS Pathogens 15, e1007385.

Baker, M.R., Fan, G., Seryshev, A.B., Agosto, M.A., Baker, M.L., Serysheva, I.I., 2021. Cryo-EM structure of type 1 IP3R channel in a lipid bilayer. Commun Biol 4, 1–11.

Barabote, R.D., Rendulic, S., Schuster, S.C., Saier, M.H., 2007. Comprehensive Analysis of Transport Proteins Encoded Within the Genome of Bdellovibrio bacteriovorus. Genomics 90, 424–446.

Barbarin-Bocahu, I., Graille, M., 2022. The X-ray crystallography phase problem solved thanks to AlphaFold and RoseTTAFold models: a case-study report. Acta Crystallographica Section D 78, 517–531.

Bayer, M.E., 1968. Areas of adhesion between wall and membrane of Escherichia coli. J Gen Microbiol 53, 395–404.

Beckwith, J., 2013. The Sec-dependent pathway. Res Microbiol 164, 497–504.
Benjamin F. Cooper, 2023. Structural and functional investigation of the Paraquat Inducible pathway.
Bepler, T., Morin, A., Rapp, M., Brasch, J., Shapiro, L., Noble, A.J., Berger, B., 2019. TOPAZ: A Positive-Unlabeled Convolutional Neural Network CryoEM Particle Picker that can Pick Any Size and Shape Particle. Microsc Microanal 25, 986–987.
Bertani, B., Ruiz, N., 2018. Function and Biogenesis of Lipopolysaccharides. EcoSal Plus 8.
Betts, J.C., Lukey, P.T., Robb, L.C., McAdam, R.A., Duncan, K., 2002. Evaluation of a nutrient starvation model of Mycobacterium tuberculosis persistence by gene and protein expression profiling. Mol Microbiol 43, 717–731.
Bishop, R.E., Gibbons, H.S., Guina, T., Trent, M.S., Miller, S.I., Raetz, C.R., 2000. Transfer of palmitate from phospholipids to lipid A in outer membranes of gram-negative bacteria. EMBO J 19, 5071–5080.
Bonfiglio, G., Neroni, B., Radocchia, G., Marazzato, M., Pantanella, F., Schippa, S., 2020. Insight into the Possible Use of the Predator Bdellovibrio bacteriovorus as a Probiotic. Nutrients 12, 2252.
Breijyeh, Z., Jubeh, B., Karaman, R., 2020. Resistance of Gram-Negative Bacteria to Current Antibacterial Agents and Approaches to Resolve It. Molecules 25, 1340.
Burgess, N.K., Dao, T.P., Stanley, A.M., Fleming, K.G., 2008. β-Barrel Proteins That Reside in the Escherichia coli Outer Membrane in Vivo Demonstrate Varied Folding Behavior in Vitro. J Biol Chem 283, 26748–26758.
Cantrell, S.A., Leavell, M.D., Marjanovic, O., Iavarone, A.T., Leary, J.A., Riley, L.W., 2013. Free mycolic acid accumulation in the cell wall of the mce1 operon mutant strain of Mycobacterium tuberculosis. J Microbiol. 51, 619–626.
Caulton, S.G., Lovering, A.L., 2020. Bacterial invasion and killing by predatory Bdellovibrio primed by predator prey cell recognition and self protection. Curr Opin Microbiol 56, 74–80.
Chang, C.-Y., Hobley, L., Till, R., Capeness, M., Kanna, M., Burtt, W., Jagtap, P., Aizawa, S.-I., Sockett, R.E., 2011. The Bdellovibrio bacteriovorus twin-arginine transport system has roles in predatory and prey-independent growth. Microbiology (Reading) 157, 3079–3093.
Chen, H.-Y., Jean, S.-S., Lee, Y.-L., Lu, M.-C., Ko, W.-C., Liu, P.-Y., Hsueh, P.-R., 2021. Carbapenem-Resistant Enterobacterales in Long-Term Care Facilities: A Global and Narrative Review. Front Cell Infect Microbiol 11, 601968.
Chong, Z.-S., Woo, W.-F., Chng, S.-S., 2015. Osmoporin OmpC forms a complex with MlaA to maintain outer membrane lipid asymmetry in Escherichia coli. Mol Microbiol 98, 1133–1146.
Cong, Y., Ludtke, S.J., 2010. Chapter Eight - Single Particle Analysis at High Resolution. In: Jensen, G.J. (Ed.), Methods in Enzymology, Cryo-EM, Part B: 3-D Reconstruction. Academic Press, pp. 211–235.
Cosate, M.R., Siqueira, G.H., de Souza, G.O., Vasconcellos, S.A., Nascimento, A.L.T.O., 2016. Mammalian cell entry (Mce) protein of Leptospira interrogans binds extracellular matrix components, plasminogen and β2 integrin. Microbiology and Immunology 60, 586–598.
Coudray, N., Isom, G.L., MacRae, M.R., Saiduddin, M.N., Bhabha, G., Ekiert, D.C., 2020. Structure of bacterial phospholipid transporter MlaFEDB with substrate bound. Elife 9, e62518.
Darrow, M.C., Moore, J.P., Walker, R.J., Doering, K., King, R.S., 2019. Chameleon: Next Generation Sample Preparation for CryoEM based on Spotiton. Microsc Microanal 25, 994–995.
De Broglie, L., 1923. Waves and Quanta. Nature 112, 540–540.
de la Rosa-Trevín, J.M., Otón, J., Marabini, R., Zaldívar, A., Vargas, J., Carazo, J.M., Sorzano, C.O.S., 2013. Xmipp 3.0: an improved software suite for image processing in electron microscopy. J Struct Biol 184, 321–328.

De Rosier, D.J., Klug, A., 1968. Reconstruction of Three Dimensional Structures from Electron Micrographs. Nature 217, 130–134.
Dekker, N., 2000. Outer-membrane phospholipase A: known structure, unknown biological function. Molecular Microbiology 35, 711–717.
Delmar, J.A., Bolla, J.R., Su, C.-C., Yu, E.W., 2015. Crystallization of Membrane Proteins by Vapor Diffusion. Methods Enzymol 557, 363–392.
Deng, M., Misra, R., 1996. Examination of AsmA and its effect on the assembly of Escherichia coli outer membrane proteins. Mol Microbiol 21, 605–612.
Derewenda, Z.S., 2004. Rational protein crystallization by mutational surface engineering. Structure 12, 529–535.
Douglass, M.V., McLean, A.B., Trent, M.S., 2022. Absence of YhdP, TamB, and YdbH leads to defects in glycerophospholipid transport and cell morphology in Gram-negative bacteria. PLOS Genetics 18, e1010096.
Dowhan, W., Bogdanov, M., 2021. Eugene P. Kennedy's Legacy: Defining Bacterial Phospholipid Pathways and Function. Front Mol Biosci 8, 666203.
Doyle, M.T., Jimah, J.R., Dowdy, T., Ohlemacher, S.I., Larion, M., Hinshaw, J.E., Bernstein, H.D., 2022. Cryo-EM structures reveal multiple stages of bacterial outer membrane protein folding. Cell 185, 1143-1156.e13.
Driessen, A.J.M., Nouwen, N., 2008. Protein translocation across the bacterial cytoplasmic membrane. Annu Rev Biochem 77, 643–667.
Drulyte, I., Johnson, R.M., Hesketh, E.L., Hurdiss, D.L., Scarff, C.A., Porav, S.A., Ranson, N.A., Muench, S.P., Thompson, R.F., 2018. Approaches to altering particle distributions in cryo-electron microscopy sample preparation. Acta Crystallogr D Struct Biol 74, 560–571.
Efremov, R.G., Gatsogiannis, C., Raunser, S., 2017. Lipid Nanodiscs as a Tool for High-Resolution Structure Determination of Membrane Proteins by Single-Particle Cryo-EM. Methods Enzymol 594, 1–30.
Egerton, R.F., Watanabe, M., 2022. Spatial resolution in transmission electron microscopy. Micron 160, 103304.
El-Shazly, S., Ahmad, S., Mustafa, A.S., Al-Attiyah, R., Krajci, D., 2007. Internalization by HeLa cells of latex beads coated with mammalian cell entry (Mce) proteins encoded by the mce3 operon of Mycobacterium tuberculosis. J Med Microbiol 56, 1145–1151.
Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K., 2010. Features and development of *Coot*. Acta Crystallogr D Biol Crystallogr 66, 486–501.
Endo, R., Ohtsubo, Y., Tsuda, M., Nagata, Y., 2007. Identification and characterization of genes encoding a putative ABC-type transporter essential for utilization of gamma-hexachlorocyclohexane in Sphingobium japonicum UT26. J Bacteriol 189, 3712–3720.
Epand, R.M., Walker, C., Epand, R.F., Magarvey, N.A., 2016. Molecular mechanisms of membrane targeting antibiotics. Biochimica et Biophysica Acta (BBA) - Biomembranes, Antimicrobial peptides, cell membrane and microbial surface interaction 1858, 980–987.
Ercan, B., Low, W.-Y., Liu, X., Chng, S.-S., 2019. Characterization of Interactions and Phospholipid Transfer between Substrate Binding Proteins of the OmpC-Mla System. Biochemistry 58, 114–119.
Estrozi, L.F., Navaza, J., 2008. Fast projection matching for cryo-electron microscopy image reconstruction. Journal of Structural Biology 162, 324–334.
Fenton, A.K., Kanna, M., Woods, R.D., Aizawa, S.-I., Sockett, R.E., 2010. Shadowing the Actions of a Predator: Backlit Fluorescent Microscopy Reveals Synchronous Nonbinary Septation of Predatory Bdellovibrio inside Prey and Exit through Discrete Bdelloplast Pores. J Bacteriol 192, 6329–6335.
Fernández-Calvet, A., Rodríguez-Arce, I., Almagro, G., Moleres, J., Euba, B., Caballero, L., Martí, S., Ramos-Vivas, J., Bartholomew, T.L., Morales, X., Ortíz-de-Solórzano, C., Yuste, J.E.,

Bengoechea, J.A., Conde-Álvarez, R., Garmendia, J., 2018. Modulation of Haemophilus influenzae interaction with hydrophobic molecules by the VacJ/MlaA lipoprotein impacts strongly on its interplay with the airways. Sci Rep 8, 6872.
Flesselles, B., Anand, N.N., Remani, J., Loosmore, S.M., Klein, M.H., 1999. Disruption of the mycobacterial cell entry gene of Mycobacterium bovis BCG results in a mutant that exhibits a reduced invasiveness for epithelial cells. FEMS Microbiology Letters 177, 237–242.
Freinkman, E., Okuda, S., Ruiz, N., Kahne, D., 2012. Regulated assembly of the transenvelope protein complex required for lipopolysaccharide export. Biochemistry 51, 4800–4806.
Gample, S.P., Agrawal, S., Sarkar, D., 2019. Evidence of nitrite acting as a stable and robust inducer of non-cultivability in Mycobacterium tuberculosis with physiological relevance. Sci Rep 9, 9261.
Garces, D.H., Rhodes, W.T., Peña, N.M., 2011. Projection-slice theorem: a compact notation. J. Opt. Soc. Am. A, JOSAA 28, 766–769.
Gioffré, A., Infante, E., Aguilar, D., Santangelo, M.D. la P., Klepp, L., Amadio, A., Meikle, V., Etchechoury, I., Romano, M.I., Cataldi, A., Hernández, R.P., Bigi, F., 2005. Mutation in mce operons attenuates Mycobacterium tuberculosis virulence. Microbes Infect 7, 325–334.
Guo, L., Lim, K.B., Poduje, C.M., Daniel, M., Gunn, J.S., Hackett, M., Miller, S.I., 1998. Lipid A acylation and bacterial resistance against vertebrate antimicrobial peptides. Cell 95, 189–198.
Hagan, C.L., Silhavy, T.J., Kahne, D., 2011. β-Barrel membrane protein assembly by the Bam complex. Annu Rev Biochem 80, 189–210.
Harauz, G., Heel, M., 1986. Exact filters for general geometry three dimensional reconstruction. Optik.
Harding, C.J., Huwiler, S.G., Somers, H., Lambert, C., Ray, L.J., Till, R., Taylor, G., Moynihan, P.J., Sockett, R.E., Lovering, A.L., 2020. A lysozyme with altered substrate specificity facilitates prey cell exit by the periplasmic predator Bdellovibrio bacteriovorus. Nat Commun 11, 4817.
Hayashi, S., Wu, H.C., 1990. Lipoproteins in bacteria. J Bioenerg Biomembr 22, 451–471.
Heidari Tajabadi, F., Medrano-Soto, A., Ahmadzadeh, M., Salehi Jouzani, G., Saier, M.H., 2017. Comparative Analyses of Transport Proteins Encoded within the Genomes of Bdellovibrio bacteriovorus HD100 and Bdellovibrio exovorus JSS. J Mol Microbiol Biotechnol 27, 332–349.
Henderson, R., 2013. Avoiding the pitfalls of single particle cryo-electron microscopy: Einstein from noise. Proceedings of the National Academy of Sciences 110, 18037–18041.
Herreros, D., Lederman, R.R., Krieger, J., Jiménez-Moreno, A., Martínez, M., Myška, D., Strelak, D., Filipovic, J., Bahar, I., Carazo, J.M., Sanchez, C.O.S., 2021. Approximating deformation fields for the analysis of continuous heterogeneity of biological macromolecules by 3D Zernike polynomials. IUCrJ 8, 992–1005.
Herreros, D., Lederman, R.R., Krieger, J.M., Jiménez-Moreno, A., Martínez, M., Myška, D., Strelak, D., Filipovic, J., Sorzano, C.O.S., Carazo, J.M., 2023. Estimating conformational landscapes from Cryo-EM particles by 3D Zernike polynomials. Nat Commun 14, 154.
Hooper, D.C., 2001. Mechanisms of Action of Antimicrobials: Focus on Fluoroquinolones. Clinical Infectious Diseases 32, S9–S15.
Hughes, G.W., Hall, S.C.L., Laxton, C.S., Sridhar, P., Mahadi, A.H., Hatton, C., Piggot, T.J., Wotherspoon, P.J., Leney, A.C., Ward, D.G., Jamshad, M., Spana, V., Cadby, I.T., Harding, C., Isom, G.L., Bryant, J.A., Parr, R.J., Yakub, Y., Jeeves, M., Huber, D., Henderson, I.R., Clifton, L.A., Lovering, A.L., Knowles, T.J., 2019. Evidence for phospholipid export from the bacterial inner membrane by the Mla ABC transport system. Nat Microbiol 4, 1692–1705.
Huiskonen, J.T., 2018. Image processing for cryogenic transmission electron microscopy of symmetry-mismatched complexes. Bioscience Reports 38, BSR20170203.

Iebba, V., Santangelo, F., Totino, V., Nicoletti, M., Gagliardi, A., De Biase, R.V., Cucchiara, S., Nencioni, L., Conte, M.P., Schippa, S., 2013. Higher Prevalence and Abundance of Bdellovibrio bacteriovorus in the Human Gut of Healthy Subjects. PLoS One 8, e61608.
Im, H., Dwidar, M., Mitchell, R.J., 2018. Bdellovibrio bacteriovorus HD100, a predator of Gram-negative bacteria, benefits energetically from Staphylococcus aureus biofilms without predation. ISME J 12, 2090–2095.
Isom, G.L., Coudray, N., MacRae, M.R., McManus, C.T., Ekiert, D.C., Bhabha, G., 2020. LetB Structure Reveals a Tunnel for Lipid Transport across the Bacterial Envelope. Cell 181, 653-664.e19.
Isom, G.L., Davies, N.J., Chong, Z.-S., Bryant, J.A., Jamshad, M., Sharif, M., Cunningham, A.F., Knowles, T.J., Chng, S.-S., Cole, J.A., Henderson, I.R., 2017. MCE domain proteins: conserved inner membrane lipid-binding proteins required for outer membrane homeostasis. Sci Rep 7, 8608.
Jamali, K., Kimanius, D., Scheres, S.H.W., 2023. A Graph Neural Network Approach to Automated Model Building in Cryo-EM Maps.
Jauncey, G.E.M., 1924. The Scattering of X-Rays and Bragg's Law. Proc Natl Acad Sci U S A 10, 57–60.
Jean, S.-S., Harnod, D., Hsueh, P.-R., 2022. Global Threat of Carbapenem-Resistant Gram-Negative Bacteria. Front Cell Infect Microbiol 12, 823684.
Jia, W., Zoeiby, A.E., Petruzziello, T.N., Jayabalasingham, B., Seyedirashti, S., Bishop, R.E., 2004. Lipid Trafficking Controls Endotoxin Acylation in Outer Membranes of Escherichia coli*. Journal of Biological Chemistry 279, 44966–44975.
Jones, N.C., Osborn, M.J., 1977. Translocation of phospholipids between the outer and inner membranes of Salmonella typhimurium. J Biol Chem 252, 7405–7412.
Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P., Hassabis, D., 2021. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589.
Jurkevitch, E., Minz, D., Ramati, B., Barel, G., 2000. Prey Range Characterization, Ribotyping, and Diversity of Soil and Rhizosphere Bdellovibrio spp. Isolated on Phytopathogenic Bacteria. Appl Environ Microbiol 66, 2365–2371.
Kalienkova, V., Alvadia, C., Clerico Mosina, V., Paulino, C., 2020. Single-Particle Cryo-EM of Membrane Proteins in Lipid Nanodiscs. In: Perez, C., Maier, T. (Eds.), Expression, Purification, and Structural Biology of Membrane Proteins, Methods in Molecular Biology. Springer US, New York, NY, pp. 245–273.
Kampjut, D., Steiner, J., Sazanov, L.A., 2021. Cryo-EM grid optimization for membrane proteins. iScience 24, 102139.
Karunker, I., Rotem, O., Dori-Bachash, M., Jurkevitch, E., Sorek, R., 2013. A Global Transcriptional Switch between the Attack and Growth Forms of Bdellovibrio bacteriovorus. PLOS ONE 8, e61850.
Kellenberger, E., 1990. The "Bayer bridges" confronted with results from improved electron microscopy methods. Mol Microbiol 4, 697–705.
Kempen, M. van, Kim, S.S., Tumescheit, C., Mirdita, M., Lee, J., Gilchrist, C.L.M., Söding, J., Steinegger, M., 2023. Fast and accurate protein structure search with Foldseek.
Kennedy, E.P., Rumley, M.K., Schulman, H., Van Golde, L.M., 1976. Identification of sn-glycero-1-phosphate and phosphoethanolamine residues linked to the membrane-derived Oligosaccharides of Escherichia coli. J Biol Chem 251, 4208–4213.

Kim, K., Lee, S., Lee, K., Lim, D., 1998. Isolation and Characterization of Toluene-Sensitive Mutants from the Toluene-Resistant Bacterium Pseudomonas putida GM73. J Bacteriol 180, 3692–3696.
Klepp, L.I., Sabio y Garcia, J., FabianaBigi, 2022. Mycobacterial MCE proteins as transporters that control lipid homeostasis of the cell wall. Tuberculosis 132, 102162.
Koh, Y.S., Roe, J.H., 1996. Dual regulation of the paraquat-inducible gene pqi-5 by SoxS and RpoS in Escherichia coli. Mol Microbiol 22, 53–61.
Kolich, L.R., Chang, Y.-T., Coudray, N., Giacometti, S.I., MacRae, M.R., Isom, G.L., Teran, E.M., Bhabha, G., Ekiert, D.C., 2020. Structure of MlaFB uncovers novel mechanisms of ABC transporter regulation. Elife 9, e60030.
Konovalova, A., Silhavy, T.J., 2015. Outer membrane lipoprotein biogenesis: Lol is not the end. Philos Trans R Soc Lond B Biol Sci 370, 20150030.
Krachler, A.M., Ham, H., Orth, K., 2011. Outer membrane adhesion factor multivalent adhesion molecule 7 initiates host cell binding during infection by gram-negative pathogens. Proc Natl Acad Sci U S A 108, 11614–11619.
Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E.L.L., 2001. Predicting transmembrane protein topology with a hidden markov model: application to complete genomes11Edited by F. Cohen. Journal of Molecular Biology 305, 567–580.
Kühlbrandt, W., 2014. Biochemistry. The resolution revolution. Science 343, 1443–1444.
Kuru, E., Lambert, C., Rittichier, J., Till, R., Ducret, A., Derouaux, A., Gray, J., Biboy, J., Vollmer, W., VanNieuwenhze, M., Brun, Y.V., Sockett, R.E., 2017. Fluorescent D-amino-acids reveal bi-cellular cell wall modifications important for Bdellovibrio bacteriovorus predation. Nat Microbiol 2, 1648–1657.
Lambert, C., Chang, C.-Y., Capeness, M.J., Sockett, R.E., 2010. The first bite--profiling the predatosome in the bacterial pathogen Bdellovibrio. PLoS One 5, e8599.
Lambert, C., Morehouse, K.A., Chang, C.-Y., Sockett, R.E., 2006. Bdellovibrio: growth and development during the predatory cycle. Curr Opin Microbiol 9, 639–644.
Lerner, T.R., Lovering, A.L., Bui, N.K., Uchida, K., Aizawa, S.-I., Vollmer, W., Sockett, R.E., 2012. Specialized Peptidoglycan Hydrolases Sculpt the Intra-bacterial Niche of Predatory Bdellovibrio and Increase Population Fitness. PLOS Pathogens 8, e1002524.
Levine, T.P., 2019. Remote homology searches identify bacterial homologues of eukaryotic lipid transfer proteins, including Chorein-N domains in TamB and AsmA and Mdm31p. BMC Mol Cell Biol 20, 43.
Levitz, T.S., Weckener, M., Fong, I., Naismith, J.H., Drennan, C.L., Brignole, E.J., Clare, D.K., Darrow, M.C., 2022. Approaches to Using the Chameleon: Robust, Automated, Fast-Plunge cryoEM Specimen Preparation. Front Mol Biosci 9, 903148.
Li, J., Chai, Q.-Y., Zhang, Y., Li, B.-X., Wang, J., Qiu, X.-B., Liu, C.H., 2015. Mycobacterium tuberculosis Mce3E Suppresses Host Innate Immune Responses by Targeting ERK1/2 Signaling. The Journal of Immunology 194, 3756–3767.
Liu, C., Ma, J., Wang, J., Wang, H., Zhang, L., 2020. Cryo-EM Structure of a Bacterial Lipid Transporter YebT. Journal of Molecular Biology 432, 1008-1019.
Llamas, M.A., Ramos, J.L., Rodríguez-Herva, J.J., 2000. Mutations in Each of the tol Genes of Pseudomonas putida Reveal that They Are Critical for Maintenance of Outer Membrane Stability. J Bacteriol 182, 4764–4772.
Lopéz-Blanco, J.R., Chacón, P., 2013. iMODFIT: Efficient and robust flexible fitting based on vibrational analysis in internal coordinates. Journal of Structural Biology 184, 261–270.
Low, W.-Y., Thong, S., Chng, S.-S., 2021. ATP disrupts lipid-binding equilibrium to drive retrograde transport critical for bacterial outer membrane asymmetry. Proc. Natl. Acad. Sci. U.S.A. 118, e2110055118.

MacRae, M.R., Puvanendran, D., Haase, M.A.B., Coudray, N., Kolich, L., Lam, C., Baek, M., Bhabha, G., Ekiert, D.C., 2022. Protein-protein interactions in the Mla lipid transport system probed by computational structure prediction and deep mutational scanning.
Malinverni, J.C., Silhavy, T.J., 2009. An ABC transport system that maintains lipid asymmetry in the gram-negative outer membrane. Proc Natl Acad Sci U S A 106, 8009–8014.
Manganelli, R., Voskuil, M.I., Schoolnik, G.K., Smith, I., 2001. The Mycobacterium tuberculosis ECF sigma factor σE: role in global gene expression and survival in macrophages†. Molecular Microbiology 41, 423–437.
Mann, D., Fan, J., Somboon, K., Farrell, D.P., Muenks, A., Tzokov, S.B., DiMaio, F., Khalid, S., Miller, S.I., Bergeron, J.R.C., 2021. Structure and lipid dynamics in the maintenance of lipid asymmetry inner membrane complex of A. baumannii. Commun Biol 4, 1–9.
Masilamani, R., Cian, M.B., Dalebroux, Z.D., 2018. Salmonella Tol-Pal Reduces Outer Membrane Glycerophospholipid Levels for Envelope Homeostasis and Survival during Bacteremia. Infect Immun 86, e00173-18.
Matsuyama, S., Tajima, T., Tokuda, H., 1995. A novel periplasmic carrier protein involved in the sorting and transport of Escherichia coli lipoproteins destined for the outer membrane. EMBO J 14, 3365–3372.
McPherson, A., Gavira, J.A., 2013. Introduction to protein crystallization. Acta Crystallogr F Struct Biol Commun 70, 2–20.
Mi, W., Li, Y., Yoon, S.H., Ernst, R.K., Walz, T., Liao, M., 2017. Structural basis of MsbA-mediated lipopolysaccharide transport. Nature 549, 233–237.
Mindell, J.A., Grigorieff, N., 2003. Accurate determination of local defocus and specimen tilt in electron microscopy. Journal of Structural Biology 142, 334–347.
Mitchell, A.M., Srikumar, T., Silhavy, T.J., 2018. Cyclic Enterobacterial Common Antigen Maintains the Outer Membrane Permeability Barrier of Escherichia coli in a Manner Controlled by YhdP. mBio 9, e01321-18.
Mühlradt, P.F., Menzel, J., Golecki, J.R., Speth, V., 1973. Outer membrane of salmonella. Sites of export of newly synthesised lipopolysaccharide on the bacterial surface. Eur J Biochem 35, 471–481.
Murray, C.J., Ikuta, K.S., Sharara, F., Swetschinski, L., Robles Aguilar, G., Gray, A., Han, C., Bisignano, C., Rao, P., Wool, E., Johnson, S.C., Browne, A.J., Chipeta, M.G., Fell, F., Hackett, S., Haines-Woodhouse, G., Kashef Hamadani, B.H., Kumaran, E.A.P., McManigal, B., Agarwal, R., Akech, S., Albertson, S., Amuasi, J., Andrews, J., Aravkin, A., Ashley, E., Bailey, F., Baker, S., Basnyat, B., Bekker, A., Bender, R., Bethou, A., Bielicki, J., Boonkasidecha, S., Bukosia, J., Carvalheiro, C., Castañeda-Orjuela, C., Chansamouth, V., Chaurasia, S., Chiurchiù, S., Chowdhury, F., Cook, A.J., Cooper, B., Cressey, T.R., Criollo-Mora, E., Cunningham, M., Darboe, S., Day, N.P.J., De Luca, M., Dokova, K., Dramowski, A., Dunachie, S.J., Eckmanns, T., Eibach, D., Emami, A., Feasey, N., Fisher-Pearson, N., Forrest, K., Garrett, D., Gastmeier, P., Giref, A.Z., Greer, R.C., Gupta, V., Haller, S., Haselbeck, A., Hay, S.I., Holm, M., Hopkins, S., Iregbu, K.C., Jacobs, J., Jarovsky, D., Javanmardi, F., Khorana, M., Kissoon, N., Kobeissi, E., Kostyanev, T., Krapp, F., Krumkamp, R., Kumar, A., Kyu, H.H., Lim, C., Limmathurotsakul, D., Loftus, M.J., Lunn, M., Ma, J., Mturi, N., Munera-Huertas, T., Musicha, P., Mussi-Pinhata, M.M., Nakamura, T., Nanavati, R., Nangia, S., Newton, P., Ngoun, C., Novotney, A., Nwakanma, D., Obiero, C.W., Olivas-Martinez, A., Olliaro, P., Ooko, E., Ortiz-Brizuela, E., Peleg, A.Y., Perrone, C., Plakkal, N., Ponce-de-Leon, A., Raad, M., Ramdin, T., Riddell, A., Roberts, T., Robotham, J.V., Roca, A., Rudd, K.E., Russell, N., Schnall, J., Scott, J.A.G., Shivamallappa, M., Sifuentes-Osornio, J., Steenkeste, N., Stewardson, A.J., Stoeva, T., Tasak, N., Thaiprakong, A., Thwaites, G., Turner, C., Turner, P., van Doorn, H.R., Velaphi, S., Vongpradith, A., Vu, H., Walsh, T., Waner, S., Wangrangsimakul, T., Wozniak, T., Zheng, P., Sartorius, B., Lopez,

A.D., Stergachis, A., Moore, C., Dolecek, C., Naghavi, M., 2022. Global burden of bacterial antimicrobial resistance in 2019: a systematic analysis. The Lancet 399, 629–655.
Nakane, T., Kimanius, D., Lindahl, E., Scheres, S.H., 2018. Characterisation of molecular motions in cryo-EM single-particle data by multi-body refinement in RELION. eLife 7, e36861.
Nakane, T., Kotecha, A., Sente, A., McMullan, G., Masiulis, S., Brown, P.M.G.E., Grigoras, I.T., Malinauskaite, L., Malinauskas, T., Miehling, J., Uchański, T., Yu, L., Karia, D., Pechnikova, E.V., de Jong, E., Keizer, J., Bischoff, M., McCormack, J., Tiemeijer, P., Hardwick, S.W., Chirgadze, D.Y., Murshudov, G., Aricescu, A.R., Scheres, S.H.W., 2020. Single-particle cryo-EM at atomic resolution. Nature 587, 152–156.
Nogales, E., Scheres, S.H.W., 2015. Cryo-EM: A Unique Tool for the Visualization of Macromolecular Complexity. Molecular Cell 58, 677–689.
Noinaj, N., Kuszak, A.J., Gumbart, J.C., Lukacik, P., Chang, H., Easley, N.C., Lithgow, T., Buchanan, S.K., 2013. Structural insight into the biogenesis of β-barrel membrane proteins. Nature 501, 385–390.
Okuda, S., Sherman, D.J., Silhavy, T.J., Ruiz, N., Kahne, D., 2016. Lipopolysaccharide transport and assembly at the outer membrane: the PEZ model. Nat Rev Microbiol 14, 337–345.
Orlova, E.V., Saibil, H.R., 2011. Structural Analysis of Macromolecular Assemblies by Electron Microscopy. Chem Rev 111, 7710–7748.
Pace, N.J., Weerapana, E., 2014. Zinc-binding cysteines: diverse functions and structural motifs. Biomolecules 4, 419–434.
Palmer, T., Berks, B.C., 2012. The twin-arginine translocation (Tat) protein export pathway. Nat Rev Microbiol 10, 483–496.
Passmore, L.A., Russo, C.J., 2016. Specimen preparation for high-resolution cryo-EM. Methods Enzymol 579, 51–86.
Pasternak, Z., Njagi, M., Shani, Y., Chanyi, R., Rotem, O., Lurie-Weinberger, M.N., Koval, S., Pietrokovski, S., Gophna, U., Jurkevitch, E., 2014. In and out: an analysis of epibiotic vs periplasmic bacterial predators. ISME J 8, 625–635.
Peter J. Wotherspoon, 2022. An investigation into lipid exchange in the maintainance of lipid assymetry pathway.
Pfab, J., Phan, N.M., Si, D., 2021. DeepTracer for fast de novo cryo-EM protein structure modeling and special studies on CoV-related complexes. Proc Natl Acad Sci U S A 118, e2017525118.
Powers, M.J., Simpson, B.W., Trent, M.S., 2020. The Mla pathway in Acinetobacter baumannii has no demonstrable role in anterograde lipid transport. eLife 9, e56571.
Punjani, A., Fleet, D.J., 2021. 3D Flexible Refinement: Structure and Motion of Flexible Proteins from Cryo-EM.
Punjani, A., Zhang, H., Fleet, D.J., 2020. Non-uniform refinement: adaptive regularization improves single-particle cryo-EM reconstruction. Nat Methods 17, 1214–1221.
Qiang, L., Wang, J., Zhang, Y., Ge, P., Chai, Q., Li, B., Shi, Y., Zhang, L., Gao, G.F., Liu, C.H., 2019. Mycobacterium tuberculosis Mce2E suppresses the macrophage innate immune response and promotes epithelial cell proliferation. Cell Mol Immunol 16, 380–391.
Ramírez-Aportela, E., Mota, J., Conesa, P., Carazo, J.M., Sorzano, C.O.S., 2019. DeepRes: a new deep-learning- and aspect-based local resolution method for electron-microscopy maps. IUCrJ 6, 1054.
Ramlaul, K., Palmer, C.M., Nakane, T., Aylett, C.H.S., 2020. Mitigating local over-fitting during single particle reconstruction with SIDESPLITTER. Journal of Structural Biology 211, 107545.
Rathor, N., Garima, K., Sharma, N.K., Narang, A., Varma-Basil, M., Bose, M., 2016. Expression profile of mce4 operon of Mycobacterium tuberculosis following environmental stress. International Journal of Mycobacteriology 5, 328–332.

Ratkeviciute, G., Cooper, B.F., Knowles, T.J., 2021. Methods for the solubilisation of membrane proteins: the micelle-aneous world of membrane protein solubilisation. Biochem Soc Trans 49, 1763–1777.
Reboul, C.F., Bonnet, F., Elmlund, D., Elmlund, H., 2016. A Stochastic Hill Climbing Approach for Simultaneous 2D Alignment and Clustering of Cryogenic Electron Microscopy Images. Structure 24, 988–996.
Reboul, C.F., Eager, M., Elmlund, D., Elmlund, H., 2018. Single-particle cryo-EM—Improved ab initio 3D reconstruction with SIMPLE/PRIME. Protein Sci 27, 51–61.
Rohou, A., Grigorieff, N., 2015. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. J Struct Biol 192, 216–221.
Roier, S., Zingl, F.G., Cakar, F., Durakovic, S., Kohl, P., Eichmann, T.O., Klug, L., Gadermaier, B., Weinzerl, K., Prassl, R., Lass, A., Daum, G., Reidl, J., Feldman, M.F., Schild, S., 2016. A novel mechanism for the biogenesis of outer membrane vesicles in Gram-negative bacteria. Nat Commun 7, 10515.
Rosas, N.C., Lithgow, T., 2022. Targeting bacterial outer-membrane remodelling to impact antimicrobial drug resistance. Trends in Microbiology 30, 544–552.
Rosenthal, P.B., Henderson, R., 2003. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. J Mol Biol 333, 721–745.
Ruiz, N., Davis, R.M., Kumar, S., 2021. YhdP, TamB, and YdbH Are Redundant but Essential for Growth and Lipid Homeostasis of the Gram-Negative Outer Membrane. mBio 12, e02714-21.
Rustad, T.R., Harrell, M.I., Liao, R., Sherman, D.R., 2008. The enduring hypoxic response of Mycobacterium tuberculosis. PLoS One 3, e1502.
Sanchez-Garcia, R., Gomez-Blanco, J., Cuervo, A., Carazo, J.M., Sorzano, C.O.S., Vargas, J., 2021. DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. Commun Biol 4, 1–8.
Sanchez-Garcia, R., Segura, J., Maluenda, D., Carazo, J.M., Sorzano, C.O.S., 2018. Deep Consensus, a deep learning-based approach for particle pruning in cryo-electron microscopy. IUCrJ 5, 854–865.
Scarff, C.A., Fuller, M.J.G., Thompson, R.F., Iadaza, M.G., 2018. Variations on Negative Stain Electron Microscopy Methods: Tools for Tackling Challenging Systems. J Vis Exp 57199.
Scheres, S.H.W., 2015. Semi-automated selection of cryo-EM particles in RELION-1.3. J Struct Biol 189, 114–122.
Scheres, S.H.W., 2016. Chapter Six - Processing of Structurally Heterogeneous Cryo-EM Data in RELION. In: Crowther, R.A. (Ed.), Methods in Enzymology, The Resolution Revolution: Recent Advances In CryoEM. Academic Press, pp. 125–157.
Schwudke, D., Linscheid, M., Strauch, E., Appel, B., Zahringer, U., Moll, H., Muller, M., Brecker, L., Gronow, S., Lindner, B., 2003. The obligate predatory Bdellovibrio bacteriovorus possesses a neutral lipid A containing alpha-D-Mannoses that replace phosphate residues: similarities and differences between the lipid As and the lipopolysaccharides of the wild type strain B. bacteriovorus HD100 and its host-independent derivative HI100. J Biol Chem 278, 27502–27512.
Serek, J., Bauer-Manz, G., Struhalla, G., van den Berg, L., Kiefer, D., Dalbey, R., Kuhn, A., 2004. Escherichia coli YidC is a membrane insertase for Sec-independent proteins. EMBO J 23, 294–301.
Shatsky, M., Hall, R.J., Brenner, S.E., Glaeser, R.M., 2009. A Method for the Alignment of Heterogeneous Macromolecules from Electron Microscopy. J Struct Biol 166, 67–78.
Shimono, N., Morici, L., Casali, N., Cantrell, S., Sidders, B., Ehrt, S., Riley, L.W., 2003. Hypervirulent mutant of Mycobacterium tuberculosis resulting from disruption of the mce1 operon. Proceedings of the National Academy of Sciences 100, 15918–15923.

Shrivastava, R., Jiang, X., Chng, S.-S., 2017. Outer membrane lipid homeostasis via retrograde phospholipid transport in Escherichia coli. Molecular Microbiology 106, 395–408.
Sigworth, F.J., 2016. Principles of cryo-EM single-particle image processing. Microscopy 65, 57–67.
Sigworth, F.J., Doerschuk, P.C., Carazo, J.-M., Scheres, S.H.W., 2010. Maximum-likelihood methods in cryo-EM. Part I: theoretical basis and overview of existing approaches. Methods Enzymol 482, 263–294.
Smit, J., Nikaido, H., 1978. Outer membrane of gram-negative bacteria. XVIII. Electron microscopic studies on porin insertion sites and growth of cell surface of Salmonella typhimurium. J Bacteriol 135, 687–702.
Sousa, D., Grigorieff, N., 2007. Ab initio resolution measurement for single particle structures. J Struct Biol 157, 201–210.
Stark, H., 2010. GraFix: stabilization of fragile macromolecular complexes for single particle cryo-EM. Methods Enzymol 481, 109–126.
Stolp, H., Starr, M.P., 1963. Bdellovibrio bacteriovorus gen. et sp. n., a predatory, ectoparasitic, and bacteriolytic microorganism. Antonie van Leeuwenhoek 29, 217–248.
Střelák, D., Filipovič, J., Jiménez-Moreno, A., Carazo, J.M., Sánchez Sorzano, C.Ó., 2020. FlexAlign: An Accurate and Fast Algorithm for Movie Alignment in Cryo-Electron Microscopy. Electronics 9, 1040.
Strelak, D., Jiménez-Moreno, A., Vilas, J.L., Ramírez-Aportela, E., Sánchez-García, R., Maluenda, D., Vargas, J., Herreros, D., Fernández-Giménez, E., de Isidro-Gómez, F.P., Horacek, J., Myska, D., Horacek, M., Conesa, P., Fonseca-Reyna, Y.C., Jiménez, J., Martínez, M., Harastani, M., Jonić, S., Filipovic, J., Marabini, R., Carazo, J.M., Sorzano, C.O.S., 2021. Advances in Xmipp for Cryo–Electron Microscopy: From Xmipp to Scipion. Molecules 26, 6224.
Sutterlin, H.A., Shi, H., May, K.L., Miguel, A., Khare, S., Huang, K.C., Silhavy, T.J., 2016. Disruption of lipid homeostasis in the Gram-negative cell envelope activates a novel cell death pathway. Proc Natl Acad Sci U S A 113, E1565-1574.
Talaat, A.M., Lyons, R., Howard, S.T., Johnston, S.A., 2004. The temporal expression profile of Mycobacterium tuberculosis infection in mice. Proceedings of the National Academy of Sciences 101, 4602–4607.
Tang, X., Chang, S., Qiao, W., Luo, Q., Chen, Y., Jia, Z., Coleman, J., Zhang, K., Wang, T., Zhang, Z., Zhang, C., Zhu, X., Wei, X., Dong, C., Zhang, X., Dong, H., 2021. Structural insights into outer membrane asymmetry maintenance in Gram-negative bacteria by MlaFEDB. Nat Struct Mol Biol 28, 81–91.
Teufel, F., Almagro Armenteros, J.J., Johansen, A.R., Gíslason, M.H., Pihl, S.I., Tsirigos, K.D., Winther, O., Brunak, S., von Heijne, G., Nielsen, H., 2022. SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat Biotechnol 40, 1023–1025.
Thong, S., Ercan, B., Torta, F., Fong, Z.Y., Wong, H.Y.A., Wenk, M.R., Chng, S.-S., 2016. Defining key roles for auxiliary proteins in an ABC transporter that maintains bacterial outer membrane lipid asymmetry. eLife 5, e19042.
Torices, R., Muñoz-Pajares, A.J., 2015. PHENIX: An R package to estimate a size-controlled phenotypic integration index1. Appl Plant Sci 3, apps.1400104.
Uchański, T., Masiulis, S., Fischer, B., Kalichuk, V., López-Sánchez, U., Zarkadas, E., Weckener, M., Sente, A., Ward, P., Wohlkönig, A., Zögg, T., Remaut, H., Naismith, J.H., Nury, H., Vranken, W., Aricescu, A.R., Pardon, E., Steyaert, J., 2021. Megabodies expand the nanobody toolkit for protein structure determination by single-particle cryo-EM. Nat Methods 18, 60–68.
Varon, M., Shilo, M., 1981. Inhibition of the predatory activity ofBdellovibrio by various environmental pollutants. Microb Ecol 7, 107–111.

Vieni, C., Coudray, N., Isom, G.L., Bhabha, G., Ekiert, D.C., 2021. Role of Ring6 in the function of the E. coli MCE protein LetB.
Vinés, E.D., Marolda, C.L., Balachandran, A., Valvano, M.A., 2005. Defective O-Antigen Polymerization in *tolA* and *pal* Mutants of *Escherichia coli* in Response to Extracytoplasmic Stress. J Bacteriol 187, 3359–3368.
Wade, R.H., 1992. A brief look at imaging and contrast transfer. Ultramicroscopy 46, 145–156.
Wagner, T., Merino, F., Stabrin, M., Moriya, T., Antoni, C., Apelbaum, A., Hagel, P., Sitsel, O., Raisch, T., Prumbaum, D., Quentin, D., Roderer, D., Tacke, S., Siebolds, B., Schubert, E., Shaikh, T.R., Lill, P., Gatsogiannis, C., Raunser, S., 2019. SPHIRE-crYOLO is a fast and accurate fully automated particle picker for cryo-EM. Commun Biol 2, 1–13.
Wang, C.-H., Ma, L., Huang, L.-Y., Yeh, K.-M., Lin, J.-C., Siu, L.K., Chang, F.-Y., 2021. Molecular epidemiology and resistance patterns of blaOXA-48Klebsiella pneumoniae and Escherichia coli: A nationwide multicenter study in Taiwan. J Microbiol Immunol Infect 54, 665–672.
Wang, F., Gong, H., Liu, G., Li, M., Yan, C., Xia, T., Li, X., Zeng, J., 2016. DeepPicker: A deep learning approach for fully automated particle picking in cryo-EM. J Struct Biol 195, 325–336.
Ward, A., Reyes, C.L., Yu, J., Roth, C.B., Chang, G., 2007. Flexibility in the ABC transporter MsbA: Alternating access with a twist. Proceedings of the National Academy of Sciences 104, 19005–19010.
Williams, D.B., Carter, C.B., 2009a. Transmission Electron Microscopy. Springer US, Boston, MA.
Williams, D.B., Carter, C.B., 2009b. Electron Sources. In: Williams, D.B., Carter, C.B. (Eds.), Transmission Electron Microscopy: A Textbook for Materials Science. Springer US, Boston, MA, pp. 73–89.
Willis, A.R., Moore, C., Mazon-Moya, M., Krokowski, S., Lambert, C., Till, R., Mostowy, S., Sockett, R.E., 2016. Injections of Predatory Bacteria Work Alongside Host Immune Cells to Treat Shigella Infection in Zebrafish Larvae. Curr Biol 26, 3343–3351.
Xiao, Y., Yang, G., 2017. A fast method for particle picking in cryo-electron micrographs based on fast R-CNN. AIP Conference Proceedings 1836, 020080.
Yamada, H., Tamada, T., Kosaka, M., Miyata, K., Fujiki, S., Tano, M., Moriya, M., Yamanishi, M., Honjo, E., Tada, H., Ino, T., Yamaguchi, H., Futami, J., Seno, M., Nomoto, T., Hirata, T., Yoshimura, M., Kuroki, R., 2007. 'Crystal lattice engineering,' an approach to engineer protein crystal contacts by creating intermolecular symmetry: Crystallization and structure determination of a mutant human RNase 1 with a hydrophobic interface of leucines. Protein Sci 16, 1389–1397.
Yao, X., Fan, X., Yan, N., 2020. Cryo-EM analysis of a membrane protein embedded in the liposome. Proc Natl Acad Sci U S A 117, 18497–18503.
Yeates, Detecting and Overcoming Crystal Twinning. Methods in Enzymology 276, 344-358 (1997).
Yeow, J., Chng, S.-S., 2022. Of zones, bridges and chaperones – phospholipid transport in bacterial outer membrane assembly and homeostasis. Microbiology 168, 001177.
Yeow, J., Tan, K.W., Holdbrook, D.A., Chong, Z.-S., Marzinek, J.K., Bond, P.J., Chng, S.-S., 2018. The architecture of the OmpC-MlaA complex sheds light on the maintenance of outer membrane lipid asymmetry in Escherichia coli. J Biol Chem 293, 11325–11340.
Yip, K.M., Fischer, N., Paknia, E., Chari, A., Stark, H., 2020. Atomic-resolution protein structure determination by cryo-EM. Nature 587, 157–161.
Zhang, D.K., n.d. Brief Manual of Gautomatch.
Zhang, F., Xie, J.-P., 2011. Mammalian cell entry gene family of Mycobacterium tuberculosis. Mol Cell Biochem 352, 1–10.

Zhang, K., 2016. Gctf: Real-time CTF determination and correction. J Struct Biol 193, 1–12.
Zhang, L., Zhang, C., Ojcius, D.M., Sun, D., Zhao, J., Lin, X., Li, Liwei, Li, Lanjuan, Yan, J., 2012. The mammalian cell entry (Mce) protein of pathogenic Leptospira species is responsible for RGD motif-dependent infection of cells and animals. Molecular Microbiology 83, 1006–1023.
Zheng, S.Q., Palovcak, E., Armache, J.-P., Verba, K.A., Cheng, Y., Agard, D.A., 2017. MotionCor2 - anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Nat Methods 14, 331–332.
Zhong, E.D., Bepler, T., Berger, B., Davis, J.H., 2021. CryoDRGN: reconstruction of heterogeneous cryo-EM structures using neural networks. Nat Methods 18, 176–185.
Zhou, C., Shi, H., Zhang, M., Zhou, L., Xiao, L., Feng, S., Im, W., Zhou, M., Zhang, X., Huang, Y., 2021. Structural Insight into Phospholipid Transport by the MlaFEBD Complex from P. aeruginosa. Journal of Molecular Biology 433, 166986.
Zivanov, J., Nakane, T., Scheres, S.H.W., 2019. A Bayesian approach to beam-induced motion correction in cryo-EM single-particle analysis. IUCrJ 6, 5–17.