

UNIVERSITY OF BIRMINGHAM

IMI INSTITUTE OF MICROBIOLOGY AND INFECTION

STRUCTURAL ANALYSIS OF TRANSFERRED OUTER MEMBRANE PROTEIN AND BD1904 FAMILY OF *BDELLOVIBRIO BACTERIOVORUS*

by

REBECCA J. PARR

A thesis submitted to the University of Birmingham for the degree of
DOCTOR OF PHILOSOPHY

Supported by wellcome trust

Institute of Microbiology and Infection
School of Biosciences
College of Life & Environmental Sciences
University of Birmingham
December 2022

143415 Word Count: 39,912

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This unpublished thesis/dissertation is copyright of the author and/or third parties.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

Antimicrobial resistance is an ever-increasing global issue and alongside traditional therapies, alternative approaches are needed. One new biological therapy being investigated is the use of *Bdellovibrio bacteriovorus*, a periplasmic predator of fellow Gram-negative bacteria, including multi-drug resistant pathogens. Despite this intriguing lifestyle, much remains unknown about the molecular mechanisms that enable predation. To further our understanding, X-ray crystallography has been employed to determine the structure and function of membrane and periplasmic proteins that are predicted to play essential roles during invasion. Structural biology is an effective tool to investigate proteins of interest as *B. bacteriovorus* has limited homology to other bacteria.

The primary focus of this thesis is the characterisation of a cryptic outer membrane protein that has previously been shown to transfer from the outer membrane of predator to the inner membrane of prey during predation. The structure was solved using X-ray crystallography of refolded protein from inclusion body material in *E. coli*. Through a large amount of optimisation crystals diffracted to a resolution that could be used to generate a model of the protein, improving the resolution from 10 Å to 2.8 Å. The data collected enabled building of the protein in its entirety, revealing an N-terminally-plugged oval-shaped 16-stranded β-barrel which unlike all other porin structures to date forms a pentamer. The structure and oligomeric state have been validated using cryo-EM of a construct from native membranes. Alongside probing the structure, the porin has been initially characterised using liposome swelling assays, electrophysiology, and the generation of an in-frame deletion in *B. bacteriovorus*.

The secondary focus of this PhD has been to structurally characterise three periplasmic proteins (Bd1645, Bd2126 and Bd2778), which are sequence homologs to Bd1904. *bd1904* is the most upregulated gene upon entry into the bdelloplast, but despite previous investigation the protein function is unknown. The crystal structure of Bd2778 shows that it is a structural homolog to Bd1904, which has enabled the development of two working hypotheses for the wider family.

ACKNOWLEDGMENTS

Firstly, I would like to thank my PhD supervisor Professor Andy Lovering. You have been so supportive during my PhD and have always made time to discuss the project and drive it in a forward direction. Your positivity has been vital in keeping me motivated, particularly during the biggest scientific challenges. It's been a joy to be a part of your lab group and witness your genuine excitement every time one of us gets a new novel structure. The hard work to get a structure was worth it in the end!

Secondly, I would like to thank my PostDoc, Simon Caulton. You've been a great second to Andy, and someone I've always been able to bounce ideas off. Thank you for answering my questions with a question, although frustrating at the time it probably taught me more in the end… You've been a great role model for what a PostDoc can be.

To the rest of the Lovering Lab (Hannah, Amber, Mauricio, Liam, Matt, and Stella) thank you for the simple act of being helpful in the lab, working in an encouraging and supportive environment certainly makes a difference. To those who have not yet submitted hope it all goes well.

To those who I have collaborated with; Giedrė Ratkevičiūtė, Yiling Zhu, Augustinas Silale, Satya Bhamidimarri, Carey Lambert, Paul Radford, and Callum Clark, I would like to thank you for the training or technical support you provided me with.

Outside of science I would like to thank my two Birmingham housemates Lydia and Zoe for ~~always reminding me I could have studied medicine and work in the NHS with all the fun that brings~~ providing distractions from my PhD and being fun to live with. Also, I suppose I should thank my partner Lawrence for going on my PhD journey with me, both the ups and the downs and supporting me through them.

To all the members of 7th Floor (2018-2022) you've made my time doing this PhD very fun; I've enjoyed our social Thursday when we weren't too busy or banned by the small pandemic that took place during this time. Continue being the friendly and welcoming environment I've always known you to be.

Finally, I would like to thank the University of Birmingham for match funding my PhD as part of the Wellcome Trust Antimicrobials and Antimicrobial Resistance DTP.

Contents Listings

Table of Contents

ABSTRACT ..... *iii*

ACKNOWLEDGMENTS ..... *iv*

Contents Listings ..... *v*

Table of Contents ..... **v**

List of Figures ..... **xi**

List of Tables ..... **xx**

List of Abbreviations ..... **xxii**

CHAPTER ONE: INTRODUCTION *1*

1.1 Methods of Bacterial Predation **4**

1.1.1 Epibiotic Predation 4

1.1.2 Endobiotic Predation 5

1.1.3 Phagocytosis-like Cell Engulfment Predation 6

1.2 Predatory Bacteria use in Antibacterial Therapy **6**

1.2.1 Studies Assessing Predator use *in vivo* 7

1.2.2 Moving Towards Studies in Humans 9

1.2.3 Future Considerations 10

1.3 *Bdellovibrio bacteriovorus* **11**

1.3.1 *B. bacteriovorus* Biphasic Life Cycle 11

1.4 Research Aims and Objectives **19**

CHAPTER TWO: MATERIALS AND METHODS ***21***

2.1 Introduction to Methods **21**

2.1.1 Theory of X-ray Crystallography 21

2.2 General Methods **31**

2.2.1 Generation of Competent Cells and Transformation into *E. coli* 31

2.2.2 Overnight Cultures 32

2.2.3 Glycerol Stocks 32

2.2.4 SDS-PAGE 32

2.2.5 Semi-Native PAGE 33

2.2.6 Protein Concentration 33

2.3 Construction of Over-Expression Strains **34**

2.3.1 Restriction-Free Cloning 34

2.2.2 Colony PCR 37

2.3 Site-directed Mutagenesis **38**

2.4 Protein Growth and Expression **38**

2.4.1 Expression Trials 38

2.4.2 Expression of Protein in Inclusion Bodies (Native) 40

2.4.3 Expression of Protein in Inclusion Bodies in Presence of Selenomethionine 40

2.4.4 Expression of Membrane Proteins 40

2.4.5 Expression of Soluble Proteins 41

2.5 Purification of Proteins 41

2.5.1 Purification of Protein from Inclusion Body Material 41

2.5.2 Purification of Membrane Proteins 45

2.5.3 Purification of Soluble Proteins 46

2.6 X-ray Crystallography 48

2.6.1 Protein Crystallisation Experiments 48

2.6.2 Crystal harvesting 51

2.6.3 X-ray Diffraction and Data Collection 52

2.6.4 Structural Determination and Refinement 53

2.7 Cryo-Electron Microscopy 54

2.7.1 Vitrification and Data Collection 54

2.7.2 Data Processing (Collaboration) and Model Building 55

2.8 Liposome Swelling Assay 56

2.9 Single Channel Conductance Experiments (Collaboration) 57

2.10 Construction of Gene *bd0427* Deletion Strain 58

2.10.1 Amplification of DNA Fragments for Gibson Assembly 59

2.10.2 Gibson Assembly 59

2.10.3 Construct Verification 59

2.10.4 Conjugation into *B. bacteriovorus* HD100 (Collaboration) 60

2.10.5 Generation and Verification of Ex-conjugant Strains (Collaboration) 61

CHAPTER THREE: OBTAINING AND OPTIMISING DIFFRACTION-QUALITY CRYSTALS OF BD0427 62

3.1 Introduction 62

3.1.1 Outer Membrane Proteins Biogenesis 68

3.1.2 Porins and Specific Channels 74

3.1.3 OMPs in *B. bacteriovorus,* the Story of Bd0427 80

3.1.4 Foreword 82

3.2 Results 83

3.2.1 Bd0427 ΔSP 84

3.2.2 Refold, Purification and Crystallisation of Bd0427 His-tagged Constructs 102

3.3 Summary 113

CHAPTER FOUR: STRUCTURE OF BD0427 114

4.1 Foreword 114

4.1 Results 114

4.1.1 Data Collection and Model Building 114

4.1.2 Crystal Packing 120

4.1.3 Overall Fold of Bd0427 120

4.1.4 Homopentamer Interface 124

4.1.5 Surface 126

4.1.6 The Transmembrane Channel and N-terminal Plug 127

4.1.7 Density for Detergent Tails in Hydrophobic Pocket 130

4.1.8 Unmodelled Electron Density 131

4.1.9 Conservation of Residues 132

4.2 Discussion 133

4.2.1 Bd0427 is Homopentameric 133

4.2.2 Bd0427 Structural Comparison to Deposited Porin Structures 136

4.2.3 Could the Lysine Ring be Important for Transfer? 137

4.2.4 The N-terminal Plug is a Hatch Domain that Completely Blocks the Channel 139

4.2.5 Hydrophobic Pocket Lipid Binding Hypothesis 141

4.2.6 AlphaFold and Foldseek Enable Identification of Bd0427 as Part of Wider Superfamily of Oligomeric Porins 142

CHAPTER FIVE: VALIDATION OF BD0427 STRUCTURE AND BD0427 CHARACTERISATION. 146

5.1 Foreword 146

5.2 Results 146

5.2.1 Attempts to Refold Bd0427 Δplug 146

5.2.2 Refold of Bd0427 Constructs with His-Tag Insertion 147

5.2.3 Purification of Bd0427 *pelB* A103His for Cryo-EM 148

5.2.4 Sample Preparation and Data Analysis of Bd0427 *pelB* A103His 149

5.2.5 Cryo-EM Density Map of Bd0427 *pelB* A103His (Collaboration) 152

5.2.6 Cryo-EM Structure of Bd0427 *pelB* A103His 153

5.2.7 Cryo-EM Structure Bd0427 *pelB* A103His Analysis 154

5.2.8 Unmodelled Density in Cryo-EM Map 157

5.2.9 Liposome Swelling Assay 158

5.2.10 Preliminary Characterisation by Electrophysiology (Collaboration) 161

5.2.11 Generation of *B. bacteriovorus* In-Frame Deletion Strain (Collaboration) 162

5.3 Discussion 164

5.3.1 Bd0427 Cryo-EM Structure has the same Oligomeric State and Fold as Bd0427 Crystal Structure 164

5.3.2 Potential Ligand Binding Below α-helix from Loop 2 165

5.3.3 Bd0427 Forms a Channel 166

5.3.4 *bd0427* is not an Essential Gene 168

CHAPTER SIX: STRUCTURAL CHARACTERISATION OF BD1904 FAMILY 170

6.1 Foreword 170

6.2 Results 171

6.2.1 PSI-BLAST Result 171

6.2.2 Purification and Crystallisation Trials of Bd1645 172

6.2.3 Purification and Crystallisation Trials of Bd2126 173

6.2.4 Purification and Crystallisation Trials of Bd2778 175

6.2.5 Data Collection and Model Building of Bd2778 178

6.2.6 Structure of Bd2778 182

6.2.7 Dimer Interfaces of Bd2778 185

6.2.8 Formation of the Dodecamer 186

6.2.9 Potential Ligand Binding Sites of Bd2778 190

6.2.10 Conservation of Bd2778 192

6.3 Discussion 193

6.3.1 Bd2778 a Divergent Structural Homolog of Bd1904 193

6.3.2 PEG Binding Mimics Crown Ether Binding Motif 198

6.3.3 Working Hypotheses of Bd1904 and Bd2778 Function 198

6.3.4 ColabFold Models of Bd1645 and Bd2126 Confirm they are Structural Homologs to Bd1904 and Bd278 201

CHAPTER SEVEN: CONCLUDING REMARKS AND FUTURE DIRECTIONS 205

7.1 Identification of a New Superfamily of Oligomeric Outer Membrane Proteins 205

7.2 What has been learnt about the Function of Bd0427 206

7.3 Bd1904, Bd1645, Bd2126 and Bd2778 are Structural Homologs 209

LITERATURE CITED 210

APPENDIX 228

1. Primers used in this Study 228

2. Plasmids used in this Study 229

3. Strains used in this Study 231

4. Optimisation Trays used in this Study 233

List of Figures

Figure 1-1: Mechanisms of Antibiotic Resistance 1

Figure 1-2: Methods of Bacterial Predation 4

Figure 1-3: *B. bacteriovorus* Host-Dependent Life Cycle 12

Figure 1-4: Plaque Formation Upon Attachment of *B. bacteriovorus* to Prey 14

Figure 1-5: Constriction of *B. bacteriovorus* as the Predator Enters Prey 16

Figure 1-6: Hypothesised Mechanism for Bd0427 Insertion into Prey's Inner Membrane 18

Figure 2-1: Vapour Diffusion Method for Protein Crystallisation. 21

Figure 2-2: Phase Diagram for Vapour Diffusion. 22

Figure 2-3: Bragg's Law 23

Figure 2-4: Harker Diagram for Single Isomorphous Replacement and Multiple Isomorphous Replacement 26

Figure 2-5: Symmetry-related Reflections conforming to Friedel's Law (a) and Breaking Friedel's Law (b) 27

Figure 2-6: Harker Diagram of Anomalous Dispersion 28

Figure 2-7: Workflow of CryoEM Map Generation 56

Figure 2-8: Cuvette used in Single Channel Conductance Experiments 58

Figure 3-1: The Gram-negative Cell Envelope 62

Figure 3-2: The LPS Structure from *E. coli* K12 alongside Lipid A from *B. bacteriovorus* ... 63

Figure 3-3: The Pathways for Outer Membrane Biogenesis 64

Figure 3-4: Published Structures of *E. coli* Outer Membrane Proteins from each Class 67

Figure 3-5: Post-Translational and Co-Translational Protein Transport Across the Inner Membrane 68

Figure 3-6: Transport of Outer Membrane Proteins Across the Periplasm. 71

Figure 3-7: Structures of Bam Inserting EspP into the OM 74

Figure 3-8: Porin and Specific Channel Structures 75

Figure 3-9: Ribbon Diagrams of Four Porins Showing Loop Organisation 77

Figure 3-10: Aromatic Residues and Electrostatic Surface of OmpF and PorB 78

Figure 3-11: Cartoon Diagram of Specific Porins 79

Figure 3-12: Cartoon Diagram of MspA. 80

Figure 3-13: Workflow of the Production of High-Quality Protein Crystals. 83

Figure 3-14: Semi-native PAGE of Bd0427 Refold in DDM, βOG or TritonX100 84

Figure 3-15: Analytical SEC of Bd0427 in DDM with corresponding Semi-native PAGE 86

Figure 3-16: Refold and Purification of Bd0427 in DDM 88

Figure 3-17: Initial Crystal Hits of Bd0427 in DDM 89

Figure 3-18: Crystal Hits from Seeding Trays of Bd0427 in DDM 90

Figure 3-19 Refold and Purification of Bd0427 Se-Met in DDM 91

Figure 3-20: Crystal hits of Bd0427 Se-Methionine Pool 1 92

Figure 3-21: Crystal hits of Bd0427 Se-Methionine Pool 2 93

Figure 3-22: Refold and Purification of Bd0427 in βOG 94

Figure 3-23: Initial Crystal Hits of Bd0427 in βOG 95

Figure 3-24: Crystal Hits of Bd0427 in βOG 96

Figure 3-25: Refold and Purification of Bd0427 Se-Met in βOG 98

Figure 3-26: Crystal Hits of Se-Met Bd0427 in βOG sent to Diamond Light Source April 2021 99

Figure 3-27: Crystal Hits of Se-Met Bd0427 in βOG Sent to Diamond Light Source June 2021 100

Figure 3-28: Semi-native PAGE of Refold of Bd0427 ΔSP N-His in DDM 102

Figure 3-29: Refold and Purification of Bd0427 ΔSP C-His in DDM 104

Figure 3-30: Initial Crystal Hits of Bd0427 ΔSP C-His in DDM 105

Figure 3-31: Crystal Hits of Bd0427 ΔSP C-His in DDM 105

Figure 3-32: Refold and Purification on HisTrap of Bd0427 ΔSP C-His in DDM 106

Figure 3-33: SDS-PAGE of Overnight Cleavage of the C-terminal His-tag of Bd0427 ΔSP C-His by Tev 107

Figure 3-34: Crystal Hits of Bd0427 ΔSP C-His in DDM following HisTrap Purification .. 108

Figure 3-35: Refold and Purification on HisTrap of Se-Met Bd0427 ΔSP C-His in DDM... 109

Figure 3-36: Crystal Hits of Se-Met Bd0427 ΔSP C-His in DDM following HisTrap Purification 110

Figure 3-37: Semi-native PAGE of Bd0427 ΔSP C-His Detergent Exchange into C8E4..... 111

Figure 3-38: Refold, On-column Detergent Exchange and Purification of Bd0427 ΔSP C-His in βOG 112

Figure 3-39: Initial Crystal Hits of Bd0427 ΔSP C-His Exchanged into βOG 113

Figure 4-1: Crystal and Data Collection of Bd0427 in β-OG, Crystal Form A 115

Figure 4-2: 2FO-FC Electron Density Map of Bd0427 in β-OG, Crystal Form A 117

Figure 4-3: 2FO-FC Electron Density Map of Bd0427 Crystal Forms B (a), C (b) and D (c), sigma level 1.5. 118

Figure 4-4: Schematic showing the Unit Cell of each Crystal Form. 120

Figure 4-5: Overall Fold of Bd0427 Homopentamer. 121

Figure 4-6: Bd0427 Protein Fold and Topology 122

Figure 4-7: Loops 1, 2 and 8 in Detail 123

Figure 4-8: Hydrogen-bonds and Salt-bridges Involved in the Interface Bd0427 Homopentamer 125

Figure 4-9: Hydrophobic Interactions of the Interface 126

Figure 4-10: Surface of Bd0427 127

Figure 4-11: The N-terminal plug and the Interactions the Plug makes with the Barrel 128

Figure 4-12: Channel of Bd0427 determined using MOLEonline 129

Figure 4-13: Packing of Detergent Tails 130

Figure 4-14: Coot 2mFo-DFc and mFo-DFc Map of Bd0427, Crystal Form A at sigma 1.5 131

Figure 4-15: Conserved residues of Bd0427 using Consurf 132

Figure 4-16: Bd0427 Fold from each Crystal Form 133

Figure 4-17: Comparison of Bd0427 to Homotrimeric Porins and Specific Channels 134

Figure 4-18: Top Results from DALI, using Bd0427 as Search Model 137

Figure 4-19: Lysine Rings of Bd0427 and PorB (3VZT). 138

Figure 4-20: β-barrels with N-Terminal Plugs. a) CymA in open (4V3H) and closed (4D51) conformations. b) FadL with N-terminal plug (1T16). The N-terminal plug is coloured in pink and the β-barrel is coloured in teal. 139

Figure 4-21: Comparison of N-Terminal Plugs of FadL and Bd0427. a) Alignment of N-terminal plugs of FadL (pink) and Bd0427 (blue). b) Bd0427 N-terminal plug coloured per residue RMSD value in ChimeraX. c) Sequence structural alignment, residues coloured in blue are the 15 pruned atom pairs. 141

Figure 4-22: Summary Structural Homologs of Bd0427 found using AlphaFold and Foldseek. 143

Figure 4-23: AlphaFold Predicted Structures of FSU_2404 (a) and TDE_2508 (b). 144

Figure 4-24: AlphaFold Models of Bd0427 Selecting for a Dimer (a), Trimer (b) and Pentamer (c). 144

Figure 4-25: Oligomeric State of Bd0427 Structural Homologs, from Foldseek search, Predicted by AlphaFold and ColabFold. 145

Figure 5-1: Semi-native PAGE of Refold Attempt of Bd0427 Δplug in DDM (a) and β-OG (b) 147

Figure 5-2: Semi-native PAGE of Refold of Bd0427 A103His and Bd0427 K260His in 2 % (w/v) β-OG 148

Figure 5-3: Purification of Bd0427 *pelB* A103His in 0.3 % DDM 149

Figure 5-4: Representative Micrographs (a-c) and Particles Picked from Auto-Picking (d-f). 150

Figure 5-5: Subset Selection from Initial 2D-Classification. 150

Figure 5-6: 2D-Classifications of Bd0427 *pelB* A103His from Second Round of 2D-Classification. 151

Figure 5-7: Summary of Cryo-EM Density Map Generation 152

Figure 5-8: Models of Bd0427 A103His 154

Figure 5-9: Secondary Structure Assignment and Superimposition of each Chain of Bd0427 A103His 155

Figure 5-10: Bd0427 A103His Local Resolution and Fit to Cryo-EM Density Map 156

Figure 5-11: Unmodelled Density in the Cryo-EM Map 158

Figure 5-12: Purification of Bd0427 A103His in 0.03 % DDM 159

Figure 5-13: Permeability of Proteoliposomes Reconstituted with Bd0427 *pelB* A103His .. 160

Figure 5-14: Representative Conductance of Bd0427 ΔSP. All conductance measurements were measured in 10 mM HEPES (pH 7) and 1 M KCl 161

Figure 5-15: Representative Conductance of Bd0427 *pelB* A103His. All conductance measurements were measured in 10 mM HEPES (pH 7) and 1 M KCl 162

Figure 5-16: Position of *bd0427* in *B. bacteriovorus* HD100 Genome, Plasmid Map and PCR Gels 163

Figure 5-17: Superimposition of Bd0427 Crystal Structure (blue) and Cryo-EM Structure (pink). 164

Figure 5-18: Potential Ligand Binding Sites 166

Figure 5-19: Surface Models of Bd0427 Crystal (blue) and Cryo-EM (pink) Structures 166

Figure 5-20: Surface of Bd0427 with N-terminal Plug Deleted (blue) and OmpF 168

Figure 6-1: Bd1904 Structure 170

Figure 6-2: Sequence Alignment using ClustalW of Homologous Proteins to Bd1904, from *B. bacteriovorus* Strains 171

Figure 6-3: Purification of Bd1645 by IMAC and SEC 172

Figure 6-4: Crystal from Bd1645 Crystallisation Screen 173

Figure 6-5: Purification of Bd2126 by IMAC and SEC 174

Figure 6-6: Crystals from Protein Crystallisation Trials with Bd2126. Crystals grown in 0.1 M Tris (pH 8) and 50 % (v/v) Pentaerythritol propoxylate (5/4 PO/OH). 175

Figure 6-7: IMAC Purification of Bd2778 Method 1. 175

Figure 6-8: Bd2778 Crystals 176

Figure 6-9: Purification of Bd2778 by IMAC and SEC 177

Figure 6-10: Crystals of Bd2778 using Method 2 178

Figure 6-11: Crystal Collected on and Diffraction Pattern 178

Figure 6-12: Trimmed ColabFold Bd2778 Model used as Search Model in Molecular Replacement 178

Figure 6-13: Bd2778 Unit Cell. 179

Figure 6-14: 2mFo-DFc Electron Density Map of Bd2778 1.5 σ 181

Figure 6-15: Structure of Bd2778 182

Figure 6-16: Fold of Bd2778 Protomer and Electrostatic Surface 183

Figure 6-17: Different Forms of Bd2778 184

Figure 6-18: The Two Dimer Forms of Bd2778 184

Figure 6-19: Interface of Bd2778 Two Dimer Conformations 186

Figure 6-20: Formation of Bd2778 Dodecamer 187

Figure 6-21: Hydrogen Bonds and Salt Bridges Involved in Forming the Dodecamer 188

Figure 6-22: Hydrophobic Interactions of Bd2778 Dodecamer Interfaces 189

Figure 6-23: Ligand Binding Site of PEG, around Lys66 190

Figure 6-24: Potential Ligand Binding Site next to Ser160 of '*Form 3*' 191

Figure 6-25: Conserved Residues of Bd2778 192

Figure 6-26: Conserved Residues Involved in the Binding of PEG. This shows the PEG binding site from dimer Y. 193

Figure 6-27: Superposition of Bd1904 to Bd2778 Forms 194

Figure 6-28: Bd1904 and Bd2778 Core Structure 194

Figure 6-29: Tryptophan 131 in Bd2778 195

Figure 6-30: Solvent Exposed Tyrosine Residues in Bd1904 compared to Exposed Tyrosine and Histidine Residues Involved in the Binding of PEG in Bd2778 195

Figure 6-31: Bd1904 Dimer, including Interactions involved in Dimer Formation 196

Figure 6-32: Bd1904 Multimer, observed from Crystal Symmetry Copies 197

Figure 6-33: Binding of PEG and Crown Ether around Lysine in Anti-PEG Antibodies 198

Figure 6-34: Results from Foldseek using Bd1904 as the Search Model 199

Figure 6-35: Results from Foldseek using Bd2778 as a Search Model 200

Figure 6-36: Integration Host Factor, with (b) and without (a) DNA 200

Figure 6-37: Proline Residues in β-Hairpin Turn of IHF (a) and Bd1904 (b). 201

Figure 6-38: Superposition of Bd1645 and Bd2126 to Bd2778 (a) and Bd1904 (b) 202

Figure 6-39: Predicted Structures of Bd1645 and Bd2126 by ColabFold 202

Figure 6-40: Solvent Exposed Residues with Aromatic Rings Involved in Binding PEG in Bd2778 compared to Residues in Bd1645 and Bd2126. 203

Figure 6-41: Residues with Aromatic Sidechains within '*Arm 2'* (a) and '*Arm 1*' (a) in Bd1645 (red) and Bd2126 (brown). 204

List of Tables

Table 1-1 Priority Pathogens as listed by WHO (2017) 3

Table 2-1: Antibiotic Final Concentrations 31

Table 2-2: Filter Sterile RF Buffers for Competent Cells 32

Table 2-3: Protein Concentration Method 34

Table 2-4: Expression Construct used in the PhD Thesis 35

Table 2-5: Media Solutions 39

Table 2-6: Media and Induction Protocol for Soluble Proteins 41

Table 2-7: Buffers used for Isolation of Inclusion Material 42

Table 2-8: SEC buffers for Refolded Protein 43

Table 2-9: Buffers used in IMAC of Refolded Protein 44

Table 2-10: IMAC ÄKTA Programs used for Soluble Proteins 46

Table 2-11: Dialysis Buffers of Soluble Proteins 47

Table 2-12: Initial Crystallisation Screens 48

Table 2-13: List of Cryoprotectants 52

Table 2-14: Substrate Solutions use in Liposome Swelling Assay 57

Table 3-1: Outer Membrane Protein Classes 65

Table 3-2: Crystal Resolution and CCanom 101

Table 4-1: Summary of Data Collection and Processing Statistics of Bd0427 Crystal Form A 116

Table 4-2: Summary of Data Collection and Processing Statistics of Bd0427 Crystal Form B, C and D 119

Table 4-3: Interface Statistics 125

Table 4-4: Role of each Loop involved in the Interface of either Homotrimers or a Homopentamer 135

Table 4-5: Homologs to Bd0427 in *Bdellovibrio* species 142

Table 5-1: Correlation Coefficient Model to Data 153

Table 5-2: Initial Rates and Relative Permeability of Substrates through OmpF and Bd0427 *pelB* A103His 160

Table 6-1: Summary of Data Collection and Processing Statistics of Bd2778 180

Table 6-2: Interface Statistics 185

Table 0-1: Primers used in this Study 228

Table 0-2: Plasmids used in this Study 229

Table 0-3: Strains used in this Study..... 231

List of Abbreviations

AMR	Antimicrobial resistance
OMP	Outer membrane protein
IM	Inner membrane
OM	Outer membrane
SAD	Single-wavelength anomalous dispersion
MR	Molecular replacement
ASU	Asymmetric unit
CC	Correlation coefficient
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
EDTA	Ethylene diamine tetra acetic acid
LB	Lysogeny broth
TB	Terrific broth
AI	Auto-induction
IPTG	Isopropyl-β-D-thiogalactoside
OD_{600}	Optical density at a wavelength of 600 nm
PCR	Polymerase chain reaction
dNTPs	Deoxyribonucleotide triphosphate
PSI-BLAST	Position-specific iterative basic local alignment search tool

DDM	n-Dodecyl-β-D-Maltoside
β-OG	Octyl β-D-glucopyranoside
PEG	Polyethylene glycol
PPG	Polypropylene glycol
PPG	Polypropylene glycol
DMSO	Dimethyl sulfoxide
RMSD	Root Mean Square Deviation

CHAPTER ONE: INTRODUCTION

Since the initial discovery of penicillin by Fleming (1929), evidence of antimicrobial resistance (AMR) has never been far behind a new antimicrobial discovery, with the first signs of penicillin resistance noticed in 1940 (Abraham and Chain, 1940). Following the discovery of β-lactams, there have been many new antibiotic classes identified: carbapenems, aminoglycosides, tetracyclines, rifamycins, macrolides, ketolides and glycopeptides (Coates et al., 2011) but this has always been followed by an associated AMR mechanism (Figure 1-1) (reviewed in Munita and Arias, 2016). Whilst the occurrence of resistance mechanisms does not come as a surprise (as microorganisms follow Darwinian selection processes and the

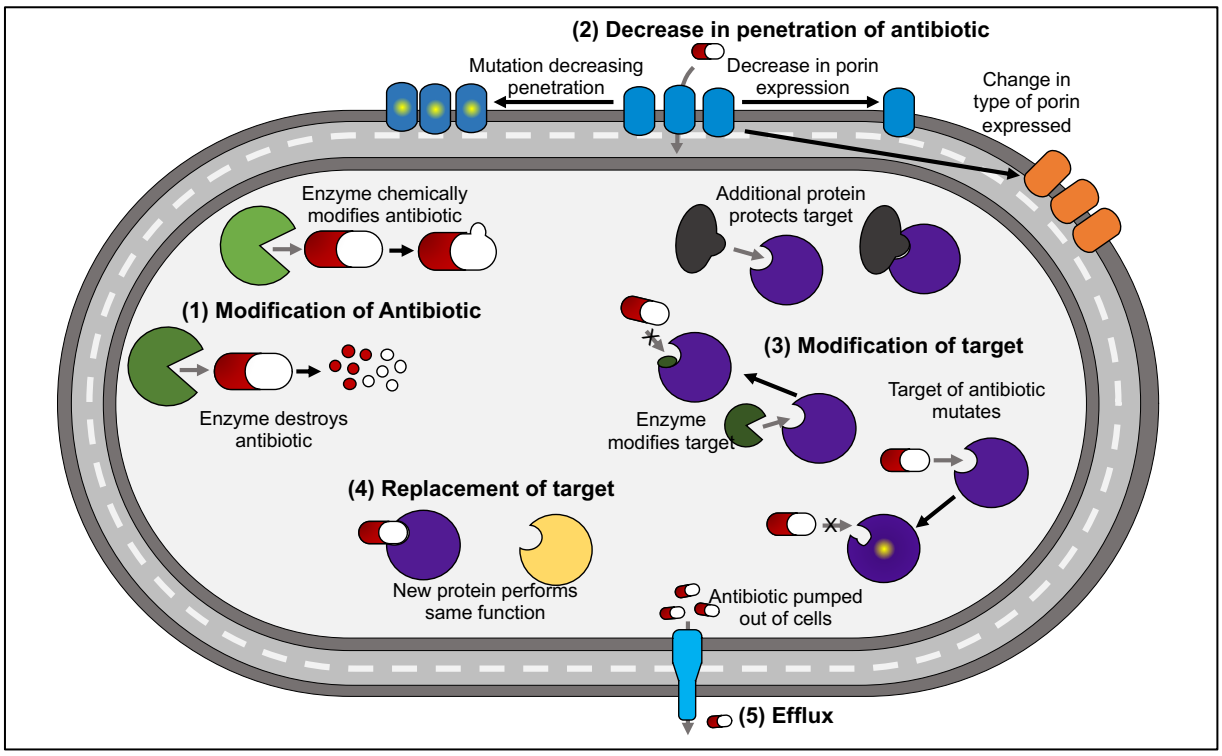


Figure 1-1: Mechanisms of Antibiotic Resistance. In antimicrobial resistance there are five main mechanisms to induce resistance to antibiotics. These are: (1) modification of the antibiotic, either through a chemical modification or degradation by an enzyme; (2) decrease in penetration of the antibiotic, through a mutation in a porin preventing or reducing influx, or a decrease in expression of the porin or changing the main porin expressed; (3) modification of the antibiotic target either through mutation of the target, modification of the target by an enzyme or protection of the target by another protein; (4) replacement of the target, new protein performs the function that cannot be impacted by the antibiotic; (5) efflux, the active pumping out of the antibiotic. These mechanisms are further described by Munita and Arias (2016).

antibiotic/antimicrobial drug provides a pressure), this does present a major scientific challenge.

In 2014, in a review on AMR, it was predicted that 700,000 people could die annually worldwide of antimicrobial related illnesses and that by 2050 this could increase to 10 million deaths per annum, leading to a major economic and medical crisis (O'Neill, 2014). This AMR problem is driven mostly by incorrect and inappropriate use of antimicrobials in healthcare, agriculture and through environmental contamination (reviewed in Holmes et al., 2016) and in 2015 the World Health Organisation (WHO) issued a global action plan outlining five objectives to address this issue. In this plan the WHO identified the AMR problem as a global priority, noting specifically that antibiotic resistance is a major concern (WHO, 2015). This was followed, in the UK, by the issue of "The UK's 20-year vision for antimicrobial resistance" (HM Government, 2019). This documentation outlines that despite active efforts in preventing AMR since 2013, the number of drug resistant infections (in humans) continues to rise. The UK government presents nine ambitions in this document, with ambition two being to drive innovation that includes developing "high quality, new and alternative treatments" for use "in humans, animals, agriculture and the environment".

In the literature there are a number of alternative antimicrobial therapies being actively developed including: bacteriophages, vaccines, peptides, nucleic acids, lysins, probiotics, nanobiotics/enzybiotics, CRISPR-Cas editors, immunotherapies, faecal transplant processing (reviewed in Kumar et al., 2021; Czaplewski et al., 2016) and predatory bacteria (Dashiff et al., 2011). These are at different stages of investigation ranging from the bacteriophage therapy and vaccines already used in clinical settings to therapies that are in preliminary investigation (Jault et al., 2019; Isturiz et al., 2017; Schooley et al., 2017; Wright et al., 2009; Peltola, 2000).

This thesis is focused on developing understanding of the molecular pathways of *B. bacteriovorus,* a Gram-negative predator of other Gram-negative bacteria, that is currently being investigated as an alternative therapy to tackle the ongoing AMR crisis. *B. bacteriovorus* is considered a model organism of the Gram-negative predator world. To that end, this chapter will first introduce methods of bacterial predation, followed by ongoing research into the use of predatory bacteria to combat bacterial pathogens, including those identified in the WHO's priority list (Table 1-1) and concludes with the molecular detail of how this organism lives and replicates in its micro-niche.

Table 1-1 Priority Pathogens as listed by WHO (2017)

Bacteria	Resistance	Priority Level
Acinetobacter baumannii	carbapenem-resistant	Critical
Pseudomonas aeruginosa	carbapenem-resistant	Critical
Enterobacteriaceae	carbapenem-resistant, 3rd generation cephalosporin-resistant	Critical
Enterococcus faecium	vancomycin-resistant	High
Staphylococcus aureus	methicillin-resistant, vancomycin intermediate and resistant	High
Helicobacter pylori	clarithromycin-resistant	High
Campylobacter	fluroquinolone-resistant	High
Salmonella	fluroquinolone-resistant	High
Neisseria gonorrhoeae	3rd generation cephalosporin-resistant, fluroquinolone-resistant	High
Streptococcus pneumoniae	penicillin-non-susceptible	Medium
Haemophilus influenzae	ampicillin-resistant	Medium
Shigella	fluroquinolone-resistant	Medium

1.1 Methods of Bacterial Predation

Bacterial predators are bacteria that kill other bacteria as a source of nutrients for survival and replication. They are disseminated across a range of different environments with diverse methods of predation. Three major methods of predation have been observed, epibiotic, endobiotic and phagocytosis-like cell engulfment (Figure 1-2) (Shiratori et al., 2019; Pérez et al., 2020).

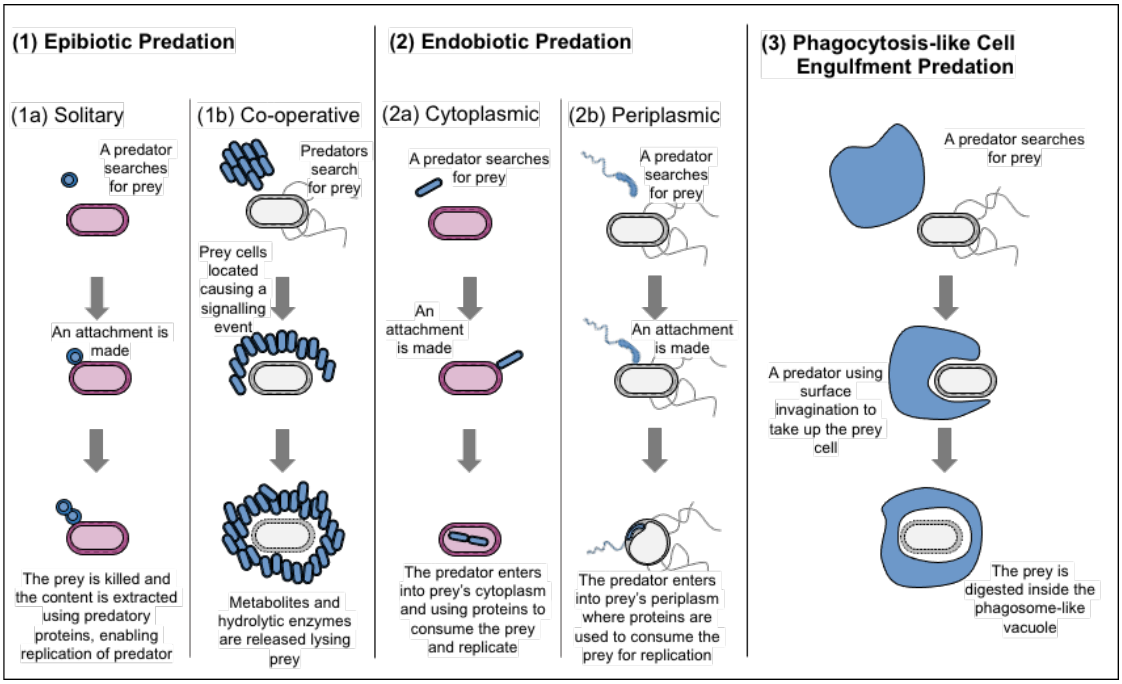


Figure 1-2: Methods of Bacterial Predation. This includes three methods: epibiotic, endobiotic and phagocytosis-like cell engulfment.

1.1.1 Epibiotic Predation

Predators that are categorised as epibiotic predators are defined as predators that kill and source the nutrients of prey from whilst remaining outside of the prey bacterium. This method can be further sub-divided into epibiotic solitary predators and epibiotic co-operative predators, also known as wolf-pack predators (Pérez et al., 2020).

Epibiotic solitary predators (Figure 1-2 1a), such as *Vampirococcus lugosii, Candidatus absconditicoccus, Bradymonadales sediminis, B. exovorus sp*. and *Micavibrio aeruginosavorus*

(Davidov et al., 2006; Koval et al., 2013; Moreira et al., 2021; Mu et al., 2020; Yakimov et al., 2022), predate by forming attachments to a prey cell. The predator can then kill and extract the cellular content using an arsenal of predatory proteins, ultimately leading to replication of the predator whilst attached to the prey membrane (Kadouri et al., 2007). Predators that use this method of predation have a diverse prey type and can predate either a singular species, as with *Ca. absconditicoccus,* or predate a range of prey species, as with *V. lugosii, B. exovorus sp.* and *M. aeruginosavorus.*

Epibiotic co-operative predators (Figure 1-2 1b), such as *Myxococcus xanthus*, *Herpetosiphon* spp., *Lysobacter* spp. and *Streptomyces globisporus* (Berleman et al., 2006; Pan et al., 2017; Seccareccia et al., 2015; Zeng et al., 2021), predate using co-operation with other predatory cells - upon a signalling event the predator cells release hydrolytic enzymes and, in certain examples, secondary metabolites to lyse prey cells (Pérez et al., 2020). Both *M. xanthus* and *Streptomyces globisporus* require direct contact with prey cells to instigate predation, despite initial theories that *Streptomyces* spp. predate using secretion of far-ranging secondary metabolites (Zeng et al., 2021; Muñoz-Dorado et al., 2016; Pérez et al., 2016). This raises the question of whether all epibiotic co-operative predators rely on contact to prey cells to initialise predation. This is an area that needs further investigation into the molecular mechanisms involved to be able to offer a definitive answer.

1.1.2 Endobiotic Predation

Endobiotic predators are bacteria that enter their prey to kill and source nutrients for growth and replication (Pérez et al., 2016). This method of predation can involve invading either the prey cytoplasm or periplasm.

Both types of endobiotic predator follow similar predation mechanisms. Firstly, they form an initial attachment on their surface of its prey before entering into either the cytoplasm or

periplasm (Sockett, 2009; Guerrero et al., 1986). Once inside the relevant bacterial compartment the prey is consumed enabling replication of the predators. Once all the prey's resources are depleted the predator's progeny exit the cell and begin the hunt for new prey cells. Currently, there is one known example of an endobiotic predator that invades the cytoplasm (Figure 1-2 2a), *Daptobacter,* which predates members of the *Chromatiaceae* (Guerrero et al., 1986). However, there are three bacteria known to invade the periplasm, these are: *B. bacteriovorus*, *Halobacterovorax* sp. and *Peredibacteria* sp. (Koval et al., 2015; Sockett, 2009; Davidov and Jurkevitch, 2004).

1.1.3 Phagocytosis-like Cell Engulfment Predation

Phagocytosis-like cell engulfment was first reported by Shiratori in 2019. This is the only example of this type of predation in the literature, conducted by *Candidatus* Uab *amorphum*. *Ca.* Uab *amorphum* is a planctomycete bacterium that has been shown to predate on Gram-negative and Gram-positive bacteria as well as a picoeukaryotic alga (Shiratori et al., 2019). Unlike the previous two methods of predation, a prey cell is taken up by the predator via surface invagination (Figure 1-2 3), forming a phagosome-like vacuole (Shiratori et al., 2019). Initial evidence suggests that whilst inside the phagosome-like vacuole the predator begins to digest the prey cell, though there are no further observations as to how the whole prey cell is consumed and whether the whole process is completed whilst inside the vacuole. As this is a recent discovery, there are still a number of molecular processes needing investigation to begin to have a in depth understanding of how this organism works.

1.2 Predatory Bacteria use in Antibacterial Therapy

Over a decade ago, two predatory bacteria were identified as potential therapies against human bacterial pathogens (Dashiff et al., 2011). The two predators identified were *M. aeruginosavorus* and *B. bacteriovorus* (Sockett, 2009; Kadouri et al., 2007). Following initial

publication of the idea of using predatory bacteria to help reduce the AMR problem, the interest in this alternative therapy has grown with a number of reviews being published in recent years discussing the progress, advantages and concerns involving the use of predatory bacteria to tackle this challenge (Atterbury and Tyson, 2021; Cavallo et al., 2021; Pérez et al., 2020; Waso et al., 2021).

In order to determine if using predatory bacteria is a viable therapeutic approach, a detailed understanding of how these predators predate on a molecular level is required, as well as understanding into whether using bacterial predators to reduce pathogen load is possible in complex biological systems. Both *B. bacteriovorus* and *M. aeruginosavorus* are capable of predating on MDR-pathogens, including species listed in the WHO's priority pathogens list (Table 1-1) shown using plaque predation assays (Dashiff et al., 2011; Kadouri et al., 2013). More recently, *B. bacteriovorus* has been shown to control growth of pathogens associated with Crohn's disease (Bonfiglio et al., 2019). Whilst these studies assess the viability of the predators to target MDR and clinically relevant pathogens they do not address whether this is possible in complex biological systems. For this a number of *in vivo* studies have been conducted.

1.2.1 Studies Assessing Predator use *in vivo*

Atterbury and colleagues (2011) conducted the first *in vivo* study using *B. bacteriovorus* in an attempt to treat chicks infected with *Salmonella* enteritidis p125109. In this study the authors demonstrated that 3-days post oral administration of *B. bacteriovorus* the load of *Salmonella* compared to the control group was reduced.

Shatzkes and colleagues (2016) furthered this work by using bacterial predators to reduce the load of *Klebsiella pneumoniae* in a mammalian system. Here rats infected with *K. pneumoniae* localised to the lungs were administered with either predator by respiratory inoculation; after a period of 24 hours rats administered with either pathogen showed reduced pathogen burden

compared to the control. However, when Shatzkes and colleagues (2017a) attempted to reduce the load of *K. pneumoniae* in a bloodstream infection in rats, where the predator was administered intravenously, a reduction in *K. pneumoniae* was not observed. The authors state that this could be due to incorrect dosing of the predator.

To develop understanding of what happens during infection and treatment with bacterial predators, Willis and colleagues (2016) carried out experiments in zebrafish larvae. Due to their transparent nature, alongside a fully developed immune system, Zebrafish provided a useful model to fully visualise predator and prey interaction within a host as well as any interaction with the immune system (Atterbury and Tyson, 2021; Willis et al., 2016). In this study Zebrafish larvae were injected with a lethal MDR *Shigella flexneri* followed by injection of *B. bacteriovorus*. Successful predation by *B. bacteriovorus* was observed, evident by the presence of bdelloplasts within the Zebrafish larvae. This study also demonstrated that *B. bacteriovorus* can work alongside the immune system, demonstrated through a comparison of infected immune-competent (wild-type) Zebrafish larvae to infected Zebrafish larvae with depleted leukocytes both treated with *B. bacteriovorus.* In this experiment clearance of *S. flexneri* following injection of *B. bacteriovorus* was better in the immune-competent Zebrafish larvae.

Experiments using mouse models have been conducted to test whether predatory bacteria can be used to clear *Yersina pestis* infections (Findlay et al., 2019; Russo et al., 2018). Similarly, to work by Shatzkes and colleagues (2016), Russo and colleagues (2018) administered *B. bacteriovorus* by respiratory inoculation into C57BL/6 mice models with lung infections with *Y. pestis,* again the authors demonstrated a reduction in pathogen load following treatment. Treatment of *Y. pestis* with *B. bacteriovorus* was also assessed by Findlay and colleagues (2019) in SKH-1 mice, this time administered using injections of the predator. Again, the bacterial predator was able to reduce the infection. Though it is important to note that when

Findlay and colleagues (2019) conducted additional preliminary studies using this same method in a different mouse model (Balb/c), there was no reduction in *Y. pestis* infection, with the authors stating further investigations are needed to find out why the treatment was unsuccessful and to determine if this is in the minority or majority when using different mouse models.

In a number of studies, it has been demonstrated that treatment with predatory bacteria does not lead to disease or toxicity and results in few adverse effects following treatment in a number of animal models (chicken (Atterbury et al., 2011), rat (Shatzkes et al., 2017a), mouse (Shatzkes et al., 2015), rabbit ocular tissue (Romanowski et al., 2016) and zebrafish larvae (Willis et al., 2016)). There is also no evidence of inflammation following administration of predatory bacteria (Romanowski et al., 2016; Shatzkes et al., 2016) and despite studies showing increases in cytokine levels there were no adverse health effects from this increase and levels returned to normal shortly after (Shatzkes et al., 2017b, 2016, 2015). Atterbury and colleagues (2011) demonstrated that *B. bacteriovorus* does not remain at detectable levels in chicks and cross contamination between chicks was not seen. All of these studies indicate that administration of predatory bacteria does not cause significant harm to the health of these animal models.

1.2.2 Moving Towards Studies in Humans

Whilst there has been mostly success using animal models, it is unclear whether the use of predators as an alternative therapy in humans will be both safe and effective. To understand how the immune system might interact with the use of *B. bacteriovorus* as a therapy, Raghunathan and colleagues (2019) conducted a study to see if humans already contain antibodies against *B. bacteriovorus* and if they interacted with immune cells. This study demonstrated that humans usually contain low levels of anti-*B. bacteriovorus* antibodies and that *B. bacteriovorus* is engulfed by macrophage-like cells. *B. bacteriovorus* remains within these macrophage-like cells for 24-hours without affecting immune cell viability whilst causing

moderate cytokine responses, before being degraded by the phagolysosome (Raghunathan et al., 2019). The authors of this study also indicated that there is a period where the predator could be used therapeutically before clearance by the immune system. This provided an initial understanding of how the immune system may interact to the use of predators and from this initial study suggests that there should not be a severe reaction to the addition. Further investigation needs to be carried out to understand how human cells, organs and systems will react to the treatment with predators, ultimately leading to clinical trials to test safety and efficacy in humans.

1.2.3 Future Considerations

Although there has been a great progress since the idea of using predatory bacteria to treat bacterial infections, the use of *B. bacteriovorus* as a therapeutic is still a number of studies away from being considered safe to trial on humans. Alongside these studies, other considerations could be looked at. These include whether wild-type predators should be used or with increasing knowledge of the predators and improving diagnostics, a library of predators could be created each that predate on a smaller range of prey. There has also been a study into the use of *B. bacteriovorus* working alongside bacteriophage to increase the degree of killing of *E. coli* prey in liquid culture (Hobley et al., 2020), indicating that *B. bacteriovorus* has the potential to be used alongside other therapies, to improve outcome.

Predatory bacteria offer a great deal of potential to the world of medicine, but to be able to use predators as therapeutics, there is a need for continued development on understanding every aspect of these predators and how they function.

1.3 *Bdellovibrio bacteriovorus*

B. bacteriovorus is an endobiotic Gram-negative bacterial predator that is currently being investigated as an alternative therapy to help combat the AMR problem. The species was originally discovered by Stolp and Petzold in 1962 (Stolp and Petzold, 1962; Stolp and Starr, 1963); identified as a small (0.3 μm wide to 1.6 μm long) vibroid-shaped Gram-negative bacterium, which has recently been reclassified within the *Oligoflexia* (Hahn et al., 2017).The bacterium has a unique biphasic life cycle, which involves the predator invading Gram-negative bacteria to replicate and produce progeny; though the bacterium has also demonstrated an ability to replicate in a host-independent manner, mainly in laboratory conditions and licensed via mutagenesis (Lambert et al., 2006).

Although there have been a number of successful studies to help understand, at a molecular level, how *B. bacteriovorus* recognise, enter, replicate, and exit prey, there are several processes that remain uncharacterised. Over 40 % of *B. bacteriovorus*' 3.8 Mb (HD100) genome consists of uncharacterised or hypothetical genes (Duncan et al., 2019). This thesis sets out to develop understanding of 5 of these hypothetical genes (*bd0427*, *bd1645*, *bd1904*, *bd2126* and *bd2778*), to increase knowledge of how *B. bacteriovorus* predates Gram-negative bacteria on a molecular level.

1.3.1 *B. bacteriovorus* Biphasic Life Cycle

The biphasic life cycle (Figure 1-3) can be separated into two distinct phases: the attack phase, where *B. bacteriovorus* actively search out prey, and the growth phase, where the predator invades, replicates, and exits its prey.

The attack phase is characterised by a free-swimming or free-gliding behaviour, where *B. bacteriovorus* actively search out prey. In high moisture environments *B. bacteriovorus* use a

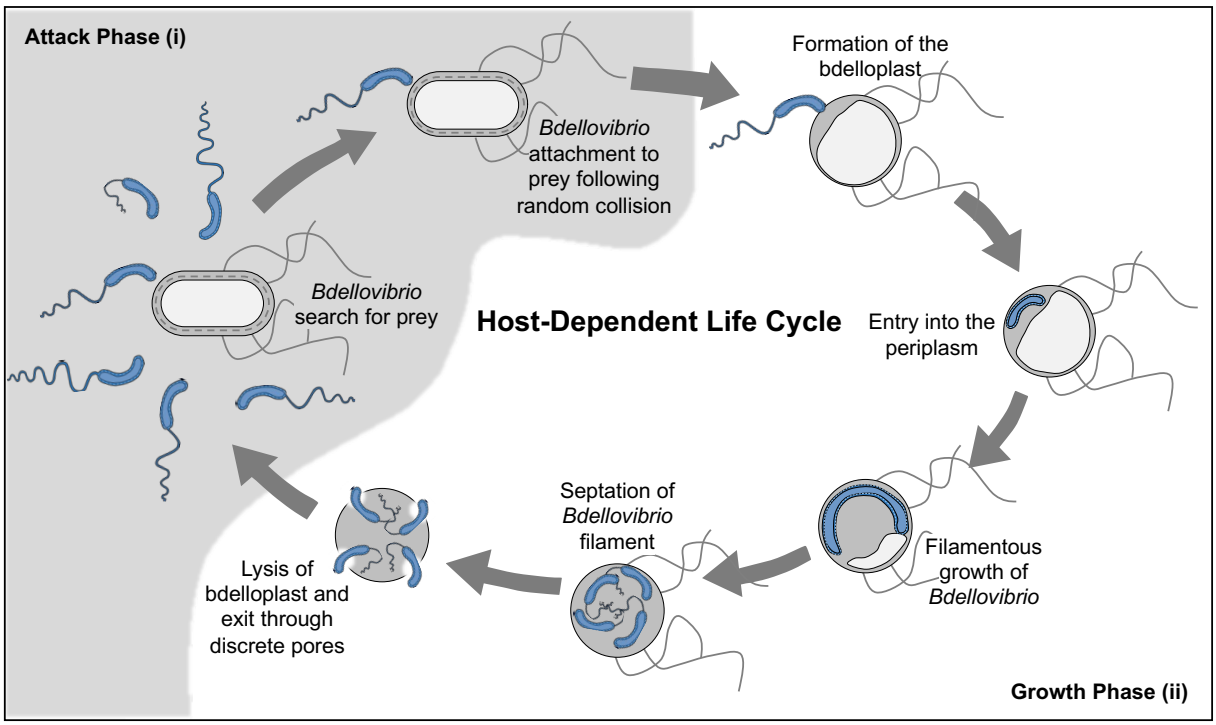


Figure 1-3: *B. bacteriovorus* Host-Dependent Life Cycle. The cycle starts in the attack phase where the predator is actively searching for prey. Through random collision a reversible attachment is made to a prey cell. This is followed by irreversible attachment and entry into the prey's periplasm using enzymes. As the predator is entering, enzymes are released to cause the prey cell to round, forming an osmotically stable bdelloplast. Once inside, the predator uses the prey's digested cell content to grow and replicate as a filament, before septation and exit from the prey cell.

single sheathed flagellum to swim, up to 160 μm.s^{-1}, towards prey-rich regions (Lambert et al., 2006). Early studies concluded that attachment to prey was only driven by random collision and that *B. bacteriovorus* did not use chemotaxis for locating prey-rich regions (Straley and Conti, 1977). However, it is now predicted that *B. bacteriovorus* does use chemotaxis to identify and swim towards prey rich regions, as the predator has previously been shown to be attracted to amino acids, sugars, nucleotides and oxygen (Matilla and Krell, 2018; Parales et al., 2015; Sampedro et al., 2014). Mutations in two chemotaxis proteins resulted in less efficient predation compared to the wild-type (Lambert et al., 2003; Medina et al., 2008). *B. bacteriovorus* can also use gliding, at 15 – 20 μm.h^{-1}, to travel to prey and this is thought to be used most in low-moisture environments, on solid surfaces or in biofilms (Lambert et al., 2011). Once the predator is in prey-rich regions it relies on random collision to make initial contact with prey (Lambert et al., 2011, 2006).

The initial contact between predator and prey cells is reversible, where the predator can detach and attack a different prey cell (Starr and Baigent, 1966). This contact at the pole opposite the flagellum, known as the biting pole, has previously been attributed to type IV pili where mutational studies have shown that they play a critical role in predation (Avidan et al., 2017; Evans et al., 2007; Mahmoud and Koval, 2010). Recently, this contact has been observed using cryo-electron tomography where it can be seen that the tips of type IVa pili is in contact with the prey cell, though they also observe that occasionally contact is seen between thin fimbriae and the outer membrane of prey cell (Kaplan et al., 2022). Following this *B. bacteriovorus* uses flagellum-independent gliding and type IV pili retraction to bring the prey's outer membrane and peptidoglycan and predator close together (Kaplan et al., 2022; Lambert et al., 2011). Once in close contact, empty non-piliated basal bodies can be seen to continue contact between predator and prey, extending through the prey's outer membrane with their extracellular rings in the peptidoglycan layer (Kaplan et al., 2022). Kaplan (2022) also notes the observation of rose-like structures at the predators biting pole in close vicinity to the contact site and suggests that these could be type I secretion systems though more investigation would be needed to confirm this. During this initial "recognition" period there is an unknown signalling event, which leads to the phosphorylation of diguanylate cyclase B. This enzyme produces localised cyclic-di-GMP, which activates CdgA (a cyclic-di-GMP receptor with a function associated with invasion) and other proteins, initiating invasion of prey (Caulton and Lovering, 2020; Meek et al., 2019). Following this an irreversible attachment is formed where predators are deemed committed to invasion (Burnham et al., 1968).

Following commitment to invasion, an unstructured plaque of material is observed at the contact point, known as the "attachment plaque" (Figure 1-4), that extends from the predator's outer membrane to the prey's peptidoglycan layer (Kaplan et al., 2022). Whilst the plaque is

being formed at the biting pole, the predator's flagellum is re-adsorbed into the periplasmic space (Kaplan et al., 2022), it had previously been thought to have been shed (Lambert et al., 2006). During this period there is significant peptidoglycan modification on the prey wall material that causes the formation of an osmotically stable spherical structure, known as a bdelloplast (Stolp and Starr, 1965). The initial formation of this structure is attributed to two genes that encode for proteins from the penicillin-binding protein 4 family (Lerner et al., 2012). A double in-frame deletion strain of these genes resulted in a significant reduction (90 %) of cell rounding of prey, suggesting the involvement of these genes in bdelloplast formation, as well as an increase in observed double infection by *B. bacteriovorus* (Lerner et al., 2012). This led the researchers to conclude that the formation of the bdelloplast also acted as an occupancy signal to attack phase *B. bacteriovorus* (Lerner et al., 2012). This rounding also increases the

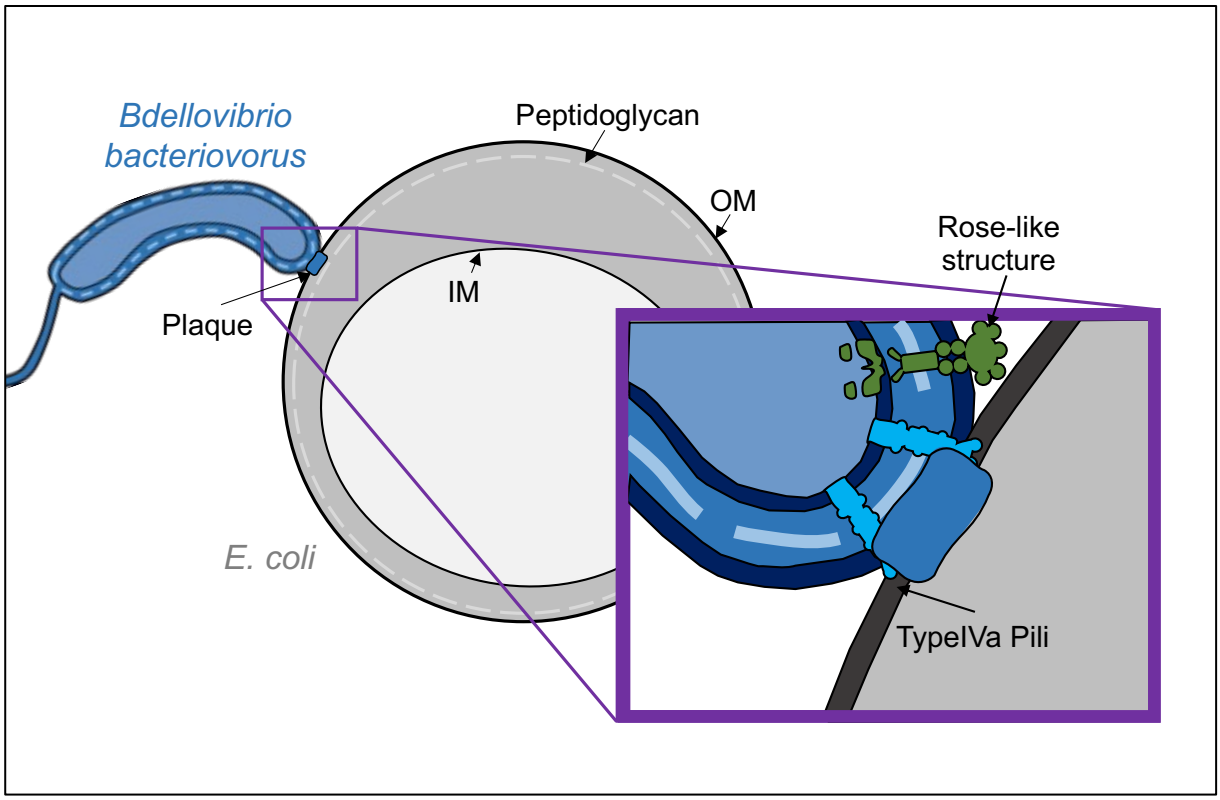


Figure 1-4: Plaque Formation Upon Attachment of *B. bacteriovorus* to Prey. *B. bacteriovorus* forms an "attachment plaque" from its outer membrane (OM) to the prey's peptidoglycan. During this stage increased separation can be seen between the OM and inner membrane (IM) of the bdelloplast. Based on cryo-electron tomography by Kaplan (2022).

size of the periplasmic space, which may explain why the double in-frame deletion also showed an increase in the time taken to enter into the periplasm (Lerner et al., 2012).

Now committed to invasion, the predator needs to enter the periplasmic space. It was originally suggested that a "drilling" mechanism using flagellum-mediated motion was used to enter through prey outer membrane (Burnham et al., 1968; Stolp and Starr, 1965), but this was discarded when mutants affecting flagellar motility could still invade (Lambert et al., 2006). The most recently supported hypothesis, hypothesised that type IV pili, using ratcheting and twitching, pull the predator into the prey's periplasm (Evans et al., 2007) but the recent work by Kaplan (2022) show that the type IV pili are nearly all retracted and or degraded by the formation of the attachment plaque. This leaves an open question to how the predator enters the prey. What is known about entry is that the attachment plaque is replaced by a pore, where it has been observed that a significant density extends from the prey's peptidoglycan to its outer membrane making direct contact with *B. bacteriovorus*'s outer membrane (Burnham et al., 1968; Kaplan et al., 2022; Starr and Baigent, 1966). The proposed mechanism for pore formation involved the use of localised degradation by transglycosylases (Thomashow and Rittenberg, 1978). In 2010, a transglycoylase was identified that was upregulated upon invasion (Lambert et al., 2010a). This gene was studied by generation of a deletion strain, to determine if it was solely responsible for pore formation; this study showed that invasion was still able to occur and that there were other genes involved in pore formation (Lerner et al., 2012). The pore is reinforced with peptidoglycan to provide strengthening (Kuru et al., 2017) providing further evidence that the pore is formed by enzymatic means.

As *B. bacteriovorus* enters through the pore, force is exerted on the predator cell resulting in constriction as the predator passes through the pore (Figure 1-5) which results in the reduction of the distance between the predator's outer and inner membrane by ~50 % (Abram et al., 1974; Kaplan et al., 2022). The ability of the predator to be able to sustain this constriction has been credited to its flexible nature due to having an extensive network of short and long filamentous structures that can be rapidly rearranged as well as having a unique cell envelope, where the outer membrane has been shown to be a similar density to the inner membrane something that is not observed in model organisms such as *E. coli* (in *E. coli* the outer membrane is more dense than the inner membrane) (Borgnia et al., 2008). A study has also indicated that the curved shape of *B. bacteriovorus* is important for entry into a bdelloplast (Banks et al., 2022).

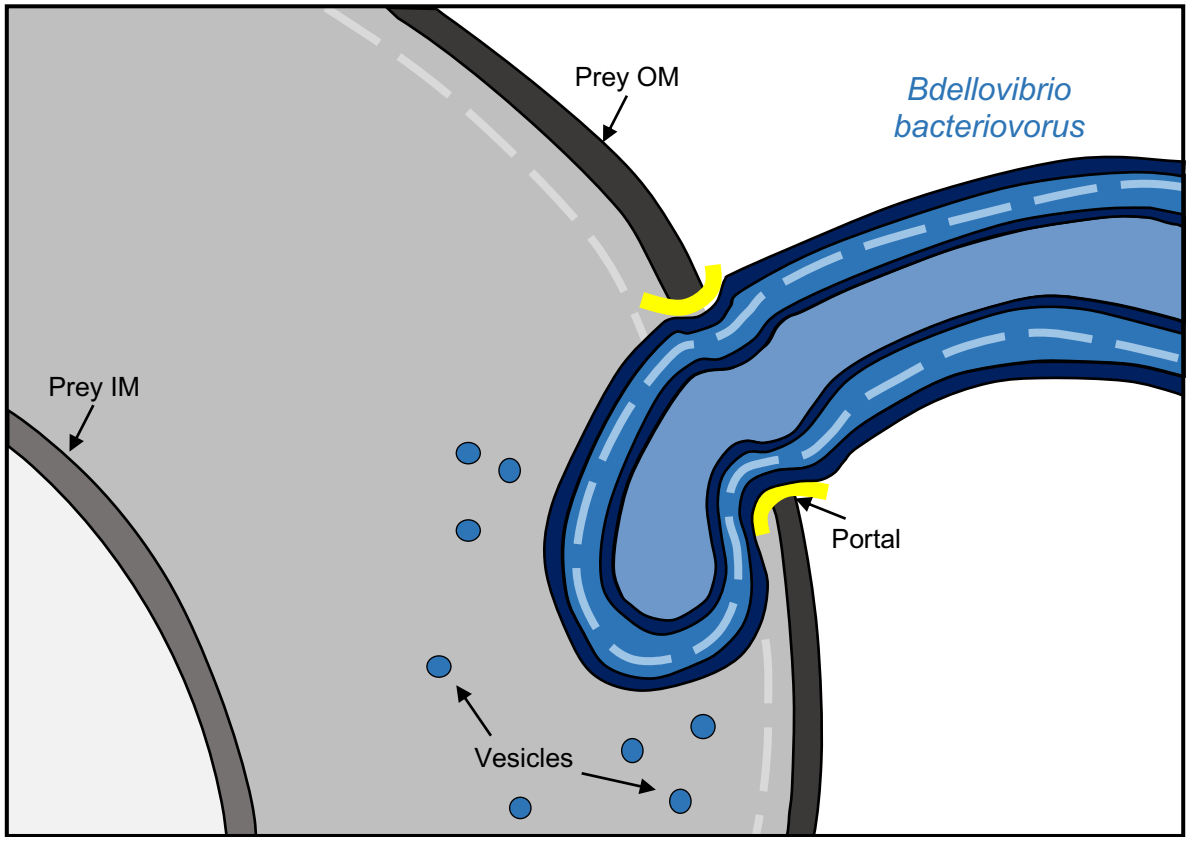


Figure 1-5: Constriction of *B. bacteriovorus* as the Predator Enters Prey. As *B. bacteriovorus* enters into the bdelloplast, constriction of the predator's cell can be seen causing a reduction in the distance between the predators outer and inner membrane. A portal can also be seen around the entry point. Based on cryo-electron tomography by Kaplan (2022).

Once the predator is inside the bdelloplast the pore that the predator entered through is sealed producing an extracellular bubble thought to be membrane (Kaplan et al., 2022; Kuru et al., 2017; Snellen and Starr, 1974). Entry into the bdelloplast signals a number of events to occur including upregulation of specific genes such as *bd1904, bd3769* and *bd0420 (Lambert et al., 2010a)*. Following a short period, the prey cell experiences cell death (Lambert et al., 2010b), though the processes leading to cell death are unknown. The prey is then harvested for nutrients to enable growth and replication of the predator, this includes harvesting of nucleic acids, amino acids, sugars, and peptidoglycan. These are then either used directly, such as nucleotides or the amino acids *B. bacteriovorus* cannot synthesise (Rendulic et al., 2004; Rittenberg and Shilo, 1970), or they can be modified to produce the molecules required (Lambert et al., 2010a; Negus et al., 2017).

The exact method for extraction of nutrients from the prey's cytoplasm is not well understood. There is no obvious method to insert degradation enzymes into the prey's cytoplasm, as *B. bacteriovorus* lack type III secretion systems (Sockett, 2009), though Barabote (2007) has suggested potential alternative transport methods, these have yet to be investigated. Recently, Kaplan (2022) observed that in the periplasm of the bdelloplast there are many uniformly sized vesicles (Figure 1-5), these could be a possible mechanism for the transport of proteins into the bdelloplast's cytoplasm. Whilst this proposes a method for protein translocation into the cytoplasm there is currently no known method for exporting cytoplasmic macromolecules and molecules to the bdelloplast's periplasm where *B. bacteriovorus* may be able to access them. However, there has been the identification of a *B. bacteriovorus* OmpF-like protein (Bd0427) that is present in the prey inner membrane following invasion - the function of this protein is unknown but it may have a role in transport of molecules from the cytoplasm to periplasm (Beck et al., 2004), and it remains possible that this protein could be transported to the prey

inner membrane by the observed vesicles (Figure 1-6). For uptake of nutrients, *B. bacteriovorus* has a number of outer membrane proteins that share homology with porins of known substrate, including porins for uptake of nucleosides, sugars and siderophores/iron (Lambert et al., 2008).

The uptake of nutrients enables growth and replication of the predator. *B. bacteriovorus* grows as a single filament, which can be many times longer than a single predator (Eksztejn and Varon, 1977; Kuru et al., 2017). During this process the 3.8 Mb chromosome is replicated, which initiates at the predatory pole (Makowski, 2019) and relevant proteins are synthesised.

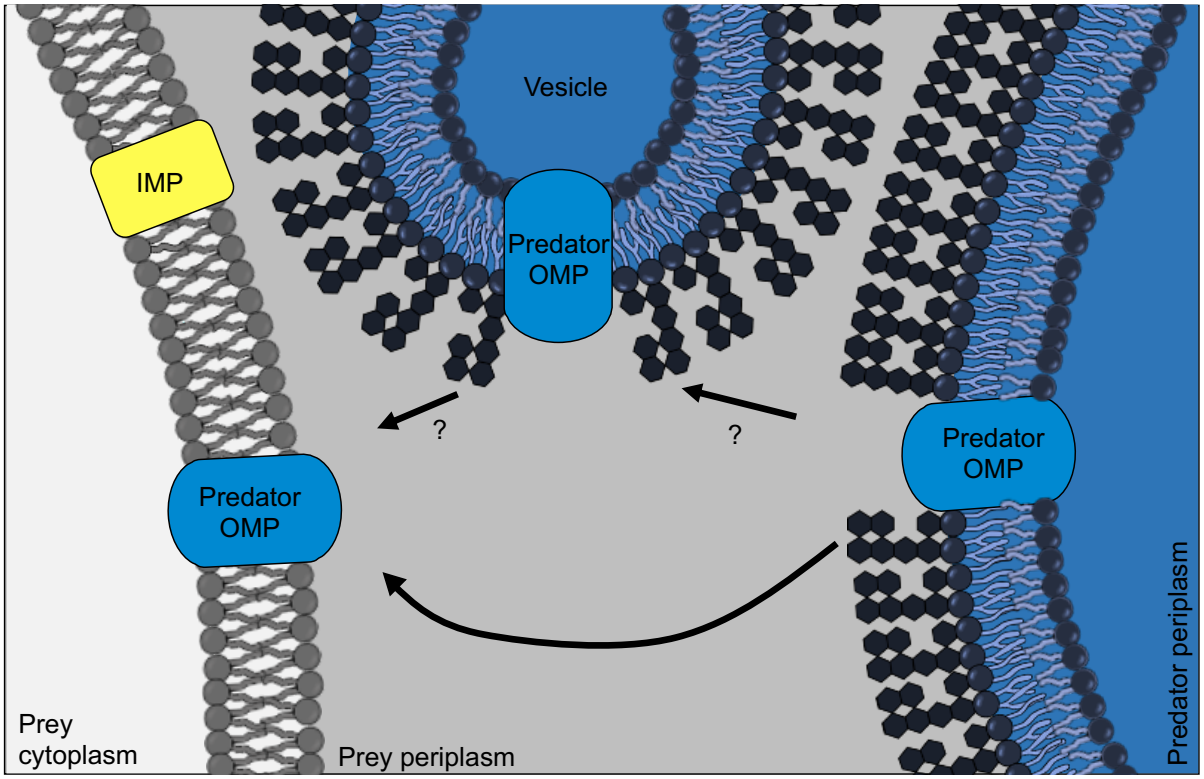


Figure 1-6: Hypothesised Mechanism for Bd0427 Insertion into Prey's Inner Membrane. This shows the inner membrane of the bdelloplast (prey) featuring an inner membrane protein (IMP), which is shown to receive the predators outer membrane protein (OMP), Bd0427, from the predators outer membrane by an unknown method. Could this transfer be mediated by the observed vesicles seen by Kaplan (2022)?

After depletion of resources, an unknown signal initiates septation into even or odd numbers of progeny, this depends on the number of complete chromosomes produced (Makowski, 2019; Fenton, 2010). The new progeny exits through discrete pores, with Harding and colleagues (2020) showing that a diversified lysozyme specific for GlcNAc-deacetylated peptidoglycan, DslA, plays a role in forming these localised pores (Harding, 2020). Once *B. bacteriovorus* has

exited, a final maturation step occurs outside the bdelloplast involving cell elongation (Fenton et al., 2010; Lambert et al., 2016). Following maturation, *B. bacteriovorus* return to the attack phase of the cycle, again searching for prey.

1.4 Research Aims and Objectives

Though much is understood about how the *B. bacteriovorus* life cycle works, including some molecular detail, there is still a large proportion that remains unknown. Before *B. bacteriovorus* can be used therapeutically a clear mechanistic understanding is needed of how the predator recognises prey and how the bacteria predate on the hosts to take up nutrients from the cytoplasm. To address this, this thesis will aim to develop the molecular understanding of five proteins: Bd0427, Bd1645, Bd1904, Bd2126, and Bd2778.

The primary focus of this PhD will be on the aforementioned outer membrane protein, Bd0427, that has been found within the inner membrane of prey following predation. This protein is predicted to form a 16-stranded β-barrel, that shares no sequence homology with proteins with known structures. Therefore, the protein will initially be investigated using a structural approach, X-ray crystallography. The information gained by structural determination will determine the biochemical assays that will be conducted. Alongside the structural characterisation an in-frame deletion strain in *B. bacteriovorus* will be generated to determine if the protein is essential and to see if a phenotype can be observed.

The secondary focus of this PhD will be on Bd1904's wider family: Bd1645, Bd2126 and Bd2778. *bd1904* is the most up regulated protein upon predation but despite structural and molecular characterisation the function remains unknown. A homology search revealed that *B. bacteriovorus* has three sequence homologs to Bd1904, that encode hypothetical proteins. In this project X-ray crystallography will be used to determine the structures of these three proteins, to establish if they are structural homologs to Bd1904. The structures will then be

compared to identify common features that are important for function. This can then be used to help determine the function of the potential protein family.

CHAPTER TWO: MATERIALS AND METHODS

2.1 Introduction to Methods

2.1.1 Theory of X-ray Crystallography

2.1.1.1 Protein Crystallisation and Harvesting

In order to conduct X-ray diffraction experiments on proteins, a protein crystal needs to be produced, as a single protein molecule would be unable to scatter enough X-rays for them to be detected. Currently the most common method used is vapour diffusion. This involves mixing the protein in solution with a precipitant solution such as polyethylene glycol (PEG), in either a sitting (Figure 2-1a) or hanging (Figure 2-1b) drop over a large reservoir of precipitant solution, in a closed system. The system then slowly moves toward equilibrium with water evaporating from the protein/precipitant drop to the reservoir. During this period the protein and precipitant concentration in the drop increases, causing the protein to shift through the

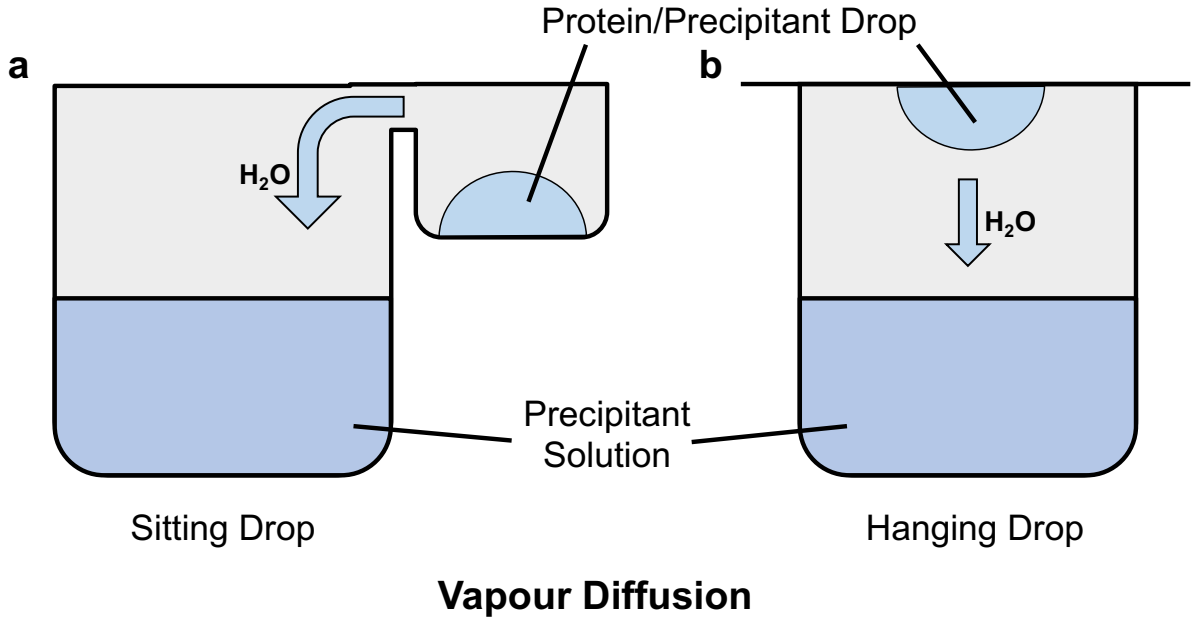


Figure 2-1: Vapour Diffusion Method for Protein Crystallisation. Schematic of sitting-drop (a) and hanging-drop (b) vapour diffusion experiments, the protein/precipitant drop loses water to the reservoir causing an increase in the concentration of protein and precipitant in the drop.

stages of the phase diagram (Figure 2-2). Formation of a crystal/s occurs in two stages of the phase diagram: the nucleation and growth stage. As a protein crystal grows the protein concentration in solution decreases. For ideal crystal formation the crystal must form in the nucleation stage and as the protein concentration in the solution decreases and the system tends towards equilibrium, the crystal growth is ideal when the crystal is in the metastable phase (Figure 2-2d). This will result in a small number of large crystals, which are considered optimal for conventional X-ray diffraction experiments.

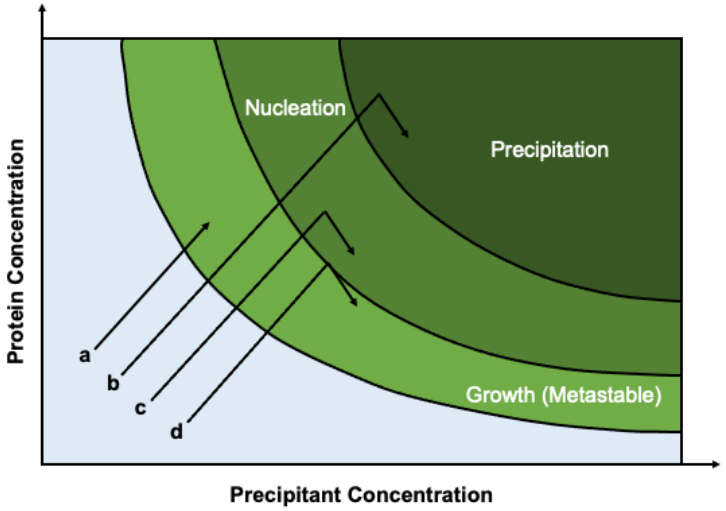


Figure 2-2: Phase Diagram for Vapour Diffusion. The concentration of protein and precipitant starts off low and increases during vapour diffusion. The examples (a-d) represent conditions that would result in: a clear condition (a), protein precipitate (b), microcrystals (c) and a few single crystals (d).

When attempting to find conditions to crystallise a new construct it is common to screen conditions that have previously been successful in crystallising proteins. Precipitant conditions that have been successful in producing crystals have been combined into screens, by companies, either with 24, 48 or 96 conditions, to aid screening in a time-wise manner.

Following crystal growth, the crystal/s are harvested, ready for X-ray diffraction. In recent years most X-ray diffraction experiments are conducted in a stream of nitrogen to keep the crystal frozen. This protects the crystal from radiation damage which is caused by a photon interacting with a protein atom or electron. To prepare protein crystals for X-ray diffraction experiments under a stream of nitrogen the crystals are cryoprotected before flash freezing in liquid nitrogen.

2.1.1.2 X-ray Diffraction and Bragg's Law

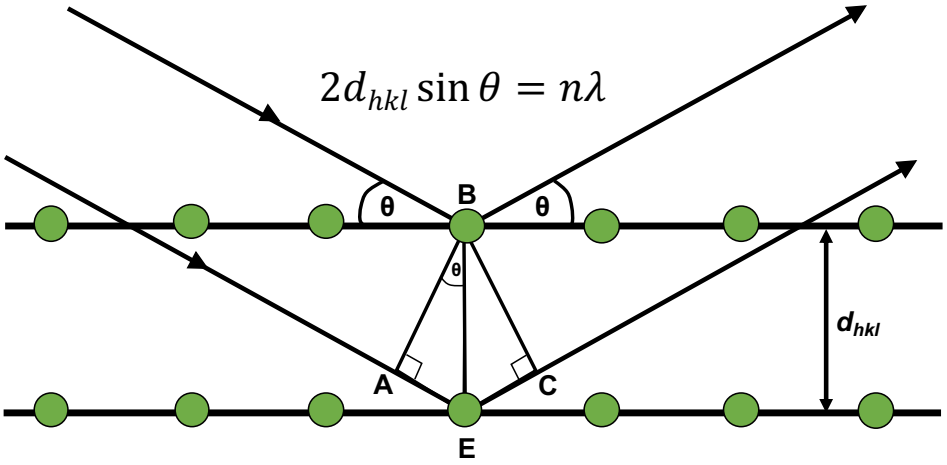


Figure 2-3: Bragg's Law. Bragg proved that two parallel planes (*h, k, l,* represented as black lines with green dots representing atoms) separated by a distance, d_{hkl}, produce diffraction at angle θ when X-rays of wavelength λ are reflected at the same angle and where n is an integer. Here AB is perpendicular to the X-ray wave producing a right angled triangle which enables calculation of the distance AE = $d_{hkl}\sin\theta$ and AEC = $2d_{hkl}\sin\theta$.

A crystal is a lattice made up of translationally repeated units (the unit cell, with edge lengths *a*, *b*, *c*), which results in a series of crystal lattice planes. Each plane is assigned a Miller index, a set of three integers that indicate their position in the unit cell, for example a plane parallel to the *bc* face would have the Miller index (1 0 0). When a crystal is subjected to X-rays of a single wavelength (λ) at a specific incident angle (θ), the atoms in a plane will reflect these X-rays. When two equivalent planes (Figure 2-3), separated by a specific distance (d), are each hit with X-rays in phase, if upon scattering these are separated (the distance EC) by a multiple n, where n is an integer, of the wavelength, the X-rays are still in phase. This results in constructive interference and means that Bragg's Law (Figure 2-3) has been satisfied. A diffraction pattern is made up of spots (reflections) that fulfil Bragg's Law. When the distance EC is not an integer times the wavelength, the X-rays are out of phase and destructive interference results in no diffraction spot. Each reflection can therefore be described as a single

wave produced by constructive interference, with the mathematical representation of this wave called a structure factor (F_{hkl}).

2.1.1.3 Structural Determination and The Phase Problem

The diffraction pattern and the atoms, or the electron density of the atoms ($\rho(x, y, z)$), that produce the diffraction pattern have been shown to share a mathematical relationship, described by a Fourier transform (Equation 2-1). This relationship states that if all the electron density values of a unit cell are known, the X-ray diffraction, in the form of amplitudes and relative phases, can be calculated. This is also true in the reverse direction, if the amplitudes and relative phases of diffracted X-rays are known, the electron density of the unit cell can be calculated. This relationship indicates how X-ray diffraction can be used to understand the atomic structure of a protein, by using it to calculate electron density. However, the diffraction pattern only gives us information on one of these elements; the amplitude, which is known to be the square root of the measured intensity of a spot. The relative phase of a reflection cannot be determined from a diffraction pattern only. This results in a knowledge deficit known as 'The Phase Problem'.

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

Equation 2-1: Fourier Transform Sum. This links electron density map ($\rho(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$) to the diffraction pattern. V is volume of the unit cell, F*hkl* is the structure factor, *h,k,l*, is a reflection.

To solve the phase problem, of proteins greater than 200 non-hydrogen atoms, there are two methods used by crystallographers. The first is known as experimental phasing in which there are two main ways this is achieved: isomorphous replacement and anomalous dispersion, from the introduction of heavier atoms. The second method is molecular replacement. Proteins or peptides smaller than 200 non-hydrogen atoms can be solved using the Patterson method or the Direct method.

2.1.1.3.1 Isomorphic Replacement

Isomorphous replacement involves the comparison of diffraction patterns from a crystal soaked in heavy atoms such as mercury, uranium, and platinum to diffraction data of a crystal with the same unit cell that has not been soaked, known as a native crystal. As heavy atoms have a larger number of electrons, they scatter more strongly than other atoms, causing measurable differences in the diffraction patterns of the heavy atom soak and native crystals. This strong scattering also enables the location of the heavy atoms to be determined using either the Patterson or Direct method. Knowing the location enables calculation of the X-ray scattering of the heavy atoms, therefore giving the amplitude and relative phases of the heavy atoms. This calculation of the heavy atom phases is what enables the solution of the protein structure (Taylor, 2010). This is best shown using a Harker diagram where the structure factors are represented as a vector sum: $F_P+F_H = F_{PH}$, where F_P is the structure factor of the native reflection, F_H is the structure factor of the heavy atom, and F_{PH} is the structure factor of the heavy atom-soaked reflection. As the location of the heavy atom is known both the amplitude and its relative phase (angle from axis) is known, this vector is set within the diagram (Figure 2-4a-c). By plotting the vector diagram, it can be seen that there are two possible solutions for the phase of the protein (the angle between the axis and either F_P^a or F_P^b, Figure 2-4c). If the two intersects are close to one another a mean can be taken as an estimate of the phase, enabling calculation of the electron density. This is known as single isomorphous replacement (SIR). If they do not lie close to one another, the mean may not be a good estimate of the phase, in this case a diffraction pattern of the crystal soaked in a different heavy atom is also used. This results in a different F_H and F_{PH} (Figure 2-4d-f), which when superimposed on to the Harker diagram from the previous heavy atom soak, identifies a common solution of F_P (Figure 2-4g). This

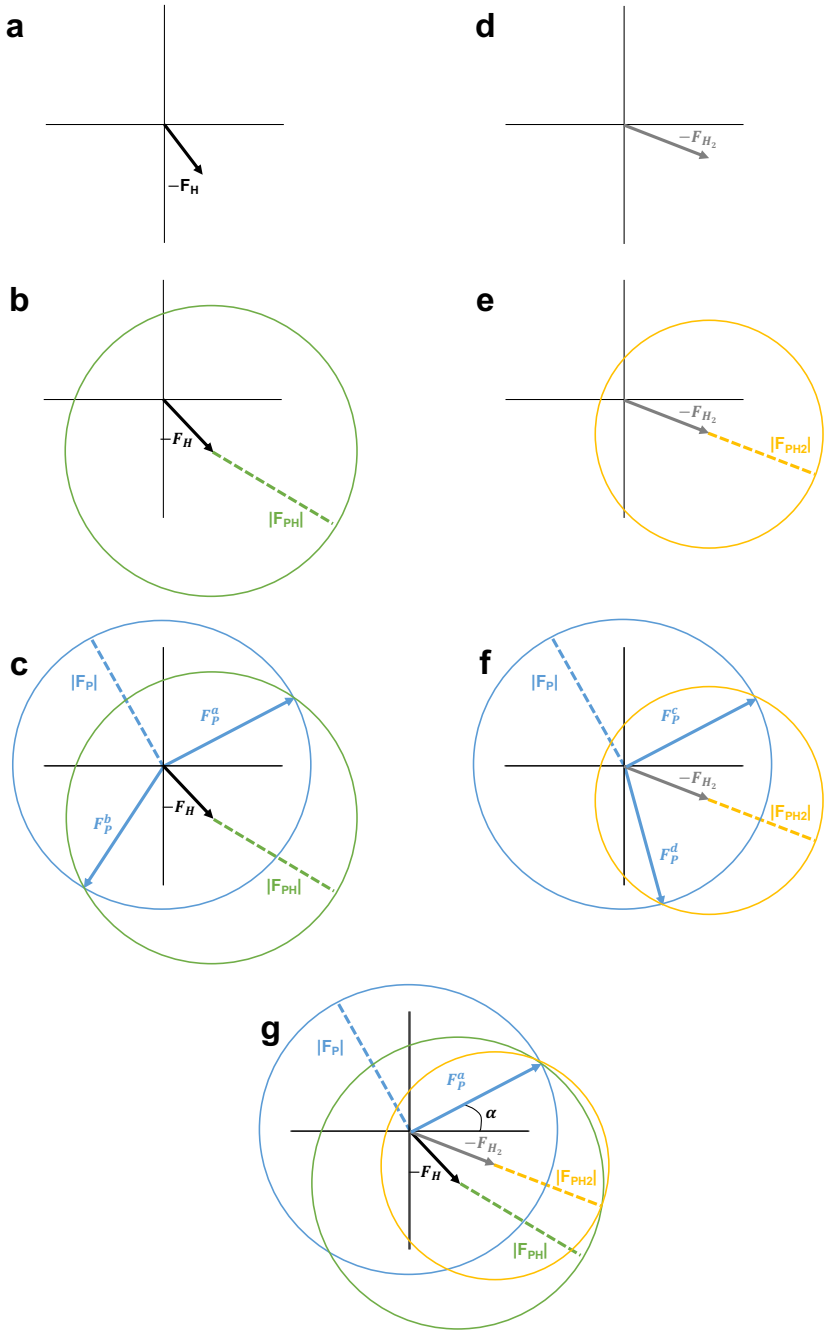


Figure 2-4: Harker Diagram for Single Isomorphous Replacement and Multiple Isomorphous Replacement. The structure factor for each heavy atom (F_H or F_{H2}) is drawn as a vector with the intensity represented as length and phase represented as the angle made with the real and imaginary axis (a&d). The intensity of the heavy atom-soaked reflection ($|F_{PH}|$ or $|F_{PH2}|$) can be drawn onto the vector diagram using equation $|F_{PH}|=|F_P|+|F_{PH}|$ (b&e), alongside this the intensity of the native reflection ($|F_P|$) can be added to the diagram (c&f) giving two intercepts (possible) solutions for the phase, if these are close to one another a mean can be taken as the solution, this is how SIR works. If this is not possible two heavy atom-soaks with different heavy atoms are used, and each vector diagram is combined (g), giving a single solution.

allows the phase (α) of this reflection to be determined. This is known as multiple isomorphous replacement (MIR).

2.1.1.3.2 Anomalous Dispersion

Anomalous dispersion has become more widely used since the introduction of tuneable X-ray beamlines at synchrotrons. The method still requires the presence of heavy atoms which can be soaked into the crystal, be naturally occurring in the protein, or can be incorporated into the protein during expression using either selenomethionine or selenocysteine instead of methionine or cysteine (also common is the bromination of nucleic acids). Though unlike isomorphous replacement this method works via the principle that at specific wavelengths the heavy atoms absorb the energy from the X-ray and re-emit them out of phase. This causes violation of a condition, known as Friedel's law, which states reflections from symmetrically related planes *h k l* and *-h -k -l* will have the same intensity with the negative of the phase (Figure 2-5) (Friedel, 1913). This difference between symmetry-related reflections is what is known as anomalous dispersion.

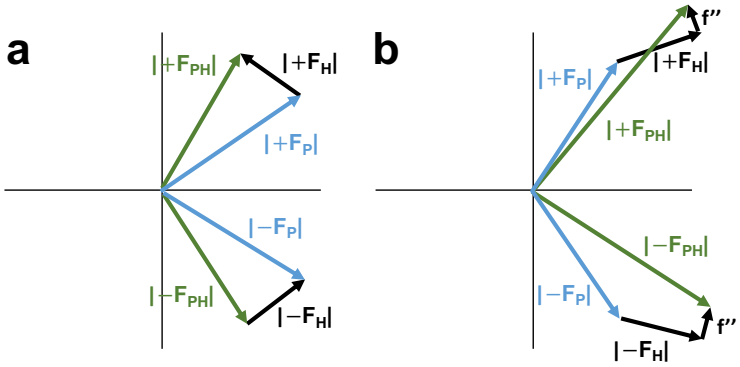


Figure 2-5: Symmetry-related Reflections conforming to Friedel's Law (a) and Breaking Friedel's Law (b). When Friedel's law is met the structure factors F_{hkl} and $F_{\overline{hkl}}$ are mirror images of each other. However, under anomalous scattering this is not the case, meaning F_{hkl} and $F_{\overline{hkl}}$ are not mirror images as shown in b.

The wavelength where anomalous dispersion is at its maximum is known as the absorption edge and represents the energy where there is a sharp change in the absorption spectrum, which

occurs as the energy of the X-ray corresponds to the energy of a shell of an atom. When this occurs the structure factor can be described as a complex number made up of imaginary (f'') and real (f') components (generally f'' is proportional to atomic absorption of the X-ray and the fluorescence and f' is the derivative of this).

To use this information to determine the relative phase of F_P of each reflection, the difference between the two symmetry-related reflections is used. There are three ways to use anomalous dispersion: SIR with anomalous scattering (SIRAS), multiple-wavelength anomalous dispersion (MAD), or single-wavelength anomalous dispersion (SAD). In all these cases the heavy atoms are located using the Patterson or Direct method.

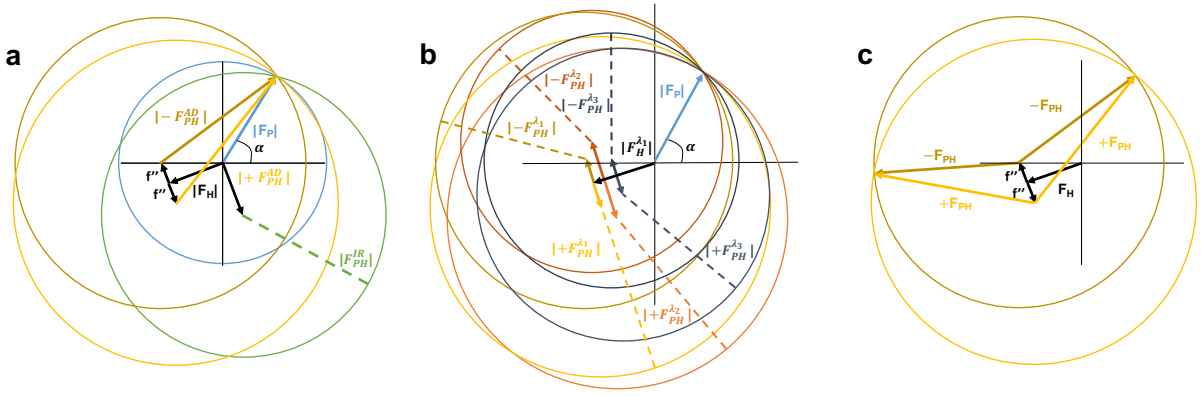


Figure 2-6: Harker Diagram of Anomalous Dispersion. a) SIRAS this uses a combination of SIR and SAD to determine the phases b) MAD involves the collection of data at three wavelengths, the peak (f' is maximum), the inflection point (f' is at its minimum) and a remote point (where anomalous scattering is minimal) each of these are plotted as vector diagrams producing a signal solution c) SAD involves the collection of data at one wavelength, the peak, this can then be used to produce two possible solutions of the phase, as seen by the vector diagram. This along with density modifications can then enable determination of the phase for a reflection.

SIRAS involves the collection of three diffraction patterns of the protein crystal, two (native and heavy atom-soaked crystal) at the same wavelength, and one (heavy atom-soaked crystal) at the wavelength of the maximum absorbance. The two collected at the same wavelength are used exactly as in SIR, and generate two possible solutions for the phase of F_P. The third diffraction pattern contains anomalous dispersion, and the difference in symmetry-related reflections are used to generate two possible solutions (Taylor, 2010). These can be plotted as

vectors on a Harker diagram (Figure 2-6a), enabling superimposition leading to one solution for the phase of F_P.

For MAD, diffraction patterns are collected of the heavy atom containing crystal at three different wavelengths: the peak (where maximum absorption occurs, and f'' is maximum), the inflection point (where f' is at its minimum) and the remote point (where anomalous signal is minimal) (*Crystallography Made Crystal Clear*, 2006; Taylor, 2003). The changes in amplitude of the structure factor of the equivalent reflection at each wavelength can then be used to determine the phase of F_P (Figure 2-6b), similarly to MIR.

SAD involves collecting only one diffraction pattern of a heavy atom containing crystal at the wavelength where maximum absorption occurs. The differences between the two symmetry-related reflections can then be used to generate two possible phases for F_P (Figure 2-6c). This, along with density modification methods, enables good estimates of the phase of F_P (Dodson, 2003), and as such is most commonly used.

2.1.1.3.3 Molecular Replacement

Traditionally molecular replacement involved the protein of interest with an unknown structure having a structural homologue (with >25-30 % sequence identity and the r.m.s. deviation between C_α atoms <2.0 Å) whose structure has previously been determined (Evans and McCoy, 2008). This homologous model can then be used instead of experimental phasing, for initial estimates of the phases of the structure factors of a diffraction pattern of a native crystal. This method works by trialling all the possible orientations and positions (defined by six parameters: three rotation angles and three translations) of the model in the unit cell, to find where the predicted diffraction from the model in a specific orientation and position best matches the diffraction from the native crystal (observed diffraction). These model phases are then used as the initial phases for the reflections from the observed diffraction to produce electron density

into which a model of the protein with an unknown structure can be built. Testing every single orientation and position a model could have within a unit cell results in a very large number of possibilities to trial and calculate; and would thus require a large amount of computing power. As a result, two methods have been developed to reduce the computing power required to produce an electron density map using molecular replacement.

The traditional method relies on the Patterson function to aid determination of the orientation and position, by first searching for the orientation and then the best position of the best orientation (it splits the 6D search into two 3D searches). To do this a Patterson map is calculated of the observed data, using the amplitudes, and another Patterson map is calculated from the model in different orientations (Grosse-Kunstleve and Adams, 2001). The two Patterson maps are compared by superposition, to find the best orientation. This single orientation can then be used to search for the best position of the model by translating the model Patterson map within the unit cell. Again, via superimposition the two Patterson maps are compared, resulting in an orientation and position where the Patterson map of the model best matches the observed Patterson (*Crystallography Made Crystal Clear*, 2006). The phases of the model in this position and orientation can then be used to calculate the electron density. MOLREP was developed in 1997 to provide a fully automated programme to utilise this method of molecular replacement and has continued to be used by crystallographers (Vagin and Teplyakov, 2010, 1997).

The maximum likelihood method is the second and more recent method used in molecular replacement (Evans and McCoy, 2008; McCoy et al., 2007). Here, maximum likelihood statistics are used to select a best model orientation and position, based on which solution explains the observed data with the highest probability. An advantage of this method is it allows modelling of the errors which can aid determination of the correct electron density (Read,

2001). The program Phaser was developed to automate molecular replacement using this method (McCoy et al., 2007).

As ab initio protein structural prediction has improved (Baek et al., 2021; Jumper et al., 2021; Mirdita et al., 2022; Skolnick et al., 2021) it is also now possible to use these de-novo models as an input model for molecular replacement. These are used in the same way as the homology models but instead the structural prediction of the target protein sequence is used (Flower and Hurley, 2021).

2.2 General Methods

2.2.1 Generation of Competent Cells and Transformation into *E. coli*

Competent *E. coli* cells (of the strains listed in Table 0-3, Appendix) were generated by inoculation of 100 mL LB media (Table 2-5) with 1 mL of overnight culture (see below) and relevant antibiotic (Table 2-1). The culture was incubated at 37 ˚C, shaking at 180 r.p.m. until an OD_{600nm} of between 0.4-0.6.

Table 2-1: Antibiotic Final Concentrations

Antibiotic	Final Concentration ($\mu g.mL^{-1}$)
Kanamycin	100
Chloramphenicol	34

The culture was transferred to a centrifuge bottle and left incubating on ice for 30 min, before a centrifugation step at 3,220 g for 5 min at 4 ˚C, to pellet the cells. The cells were gently resuspended in 25 mL of ice cold sterile RF1 buffer (Table 2-2) and incubated on ice for 30 min. The centrifugation step was repeated to pellet the cells for gentle resuspension in 4 mL ice cold sterile RF2 buffer (Table 2-2), before incubation on ice for 30 min. This resuspension was aliquoted (30 µL) into sterile microcentrifuge tubes before flash freezing in liquid nitrogen, for storage at -80 ˚C.

Table 2-2: Filter Sterile RF Buffers for Competent Cells

RF1 (pH 5.8, adjusted with acetic acid)		RF2 (pH 6.8, adjusted with NaOH)	
Rubidium Chloride	100 mM	MOPS	10 mM
Manganese (II) chloride tetrahydrate	50 mM	Rubidium Chloride	10 mM
Potassium acetate	30 mM	Calcium chloride dihydrate	75 mM
Calcium chloride dihydrate	10 mM	Glycerol	15 % (w/v)
Glycerol	15 % (w/v)		

To transform plasmids into chemically competent *E. coli* cells, 50 ng of plasmid was added to 50 µL of competent cells and incubated on ice for 30 min, heat shocked at 42 ˚C and allowed to recover on ice for 5 min. 1 mL of LB media was added and the cells were grown at 37 ˚C for 1 hr. The cells were pelleted and resuspended in 30 µL media for spreading on LB agar plates.

2.2.2 Overnight Cultures

Overnight cultures were generated by inoculating 10 mL of LB media and relevant antibiotic, with a single colony or scrape of glycerol stock. The cell culture was incubated at 37 ˚C, shaking at 180 r.p.m. overnight.

2.2.3 Glycerol Stocks

Glycerol stocks were generated by adding 500 µL of overnight culture to 500 µL of 50 % sterile glycerol, before flash freezing in liquid nitrogen. The glycerol stock was then stored at -80 ˚C.

2.2.4 SDS-PAGE

All gels were made using the Mini-PROTEAN gel casting apparatus (BIO-RAD). Standard gels were 15 % polyacrylamide. The resolving gel was made with 0.375 M tris (pH 8.8), 0.1 % (w/v) SDS, 0.1 % (w/v) ammonium persulphate (APS), 15 % acrylamide and 8 µL tetramethylethylenediamine (TEMED) made up to a final volume of 8 mL with distilled water. The stacking gel was made with 0.125 M tris (pH 6.8), 0.1 % (w/v) SDS, 0.1 % (w/v) APS, 6 % acrylamide and 5 µL TEMED made up to 5 mL with distilled water. The finished gels were

run at 180 V for 1 hr in 1x running buffer (0.25 M tris base (pH 8.3), 2 M glycine, 1 % (w/v) SDS). Gel samples were loaded in 10 µL aliquots in the presence of 1x loading dye (100 mM tris (pH 6.8, 4 % (w/v) SDS, 0.2 % (w/v) bromophenol blue, 20 % (v/v) glycerol, 200 mM β-mercaptoethanol). All gels were stained using Brilliant Blue stain (80 mg/L Coomassie Brilliant Blue, 35 mM HCl) and de-stained in distilled water until the desired level of staining was achieved. The gel was then imaged.

2.2.5 Semi-Native PAGE

This method is based on the method in (Noinaj et al., 2015) to show heat modifiability of β-barrels, used in this study to determine re-folding success.

Precast Invitrogen NuPAGE 10 % Bis-Tris Protein Gels (ThermoFisher Scientific) were used. The gels were run at 150 V for 1 hr in 1x MES running buffer (50 mM MES, 50 mM tris base, 0.1 % SDS, 1 mM EDTA, pH 7.3) at 4 ˚C.

All protein concentrations were adjusted to ~2 $mg.mL^{-1}$, and all gel samples were made in duplicate. To the samples 2x sample loading dye (20 % glycerol, 120 mM tris-HCl (pH 6.8), 2 % (w/v) SDS, 20 mM EDTA) was added, one set of the duplicate was heated to 95 ˚C for 5 min, and the other was left at room temperature. Following this, 10 µL of gel sample was loaded onto the gel. Staining was achieved using Instant Blue protein stain (Expedeon) on a rocking platform for 5 min, before de-staining in distilled water until the desired staining was achieved.

2.2.6 Protein Concentration

Protein samples were concentrated using Vivaspin 20 spin concentrators (Table 2-3), centrifuged at 3220 g at either 4 or 20 ˚C. Alternatively, protein samples were concentrated using PEG dialysis, with the protein placed in dialysis tubing surrounded by PEG 20,000 on the outside of the dialysis tubing. The protein concentration was determined using a DeNovix

DS-11 FX nanodrop, measuring the A280, with the molecular mass and extinction coefficient calculated using ExPASy ProtParam (Gasteiger et al., 2005).

Table 2-3: Protein Concentration Method

Protein	Concentration Method
Bd0427 in DDM	Vivaspin 20 100 KDa spin concentrator at 4 ˚C
Bd0427 in βOG, TritonX100	Vivaspin 20 30 KDa spin concentrator at 4 ˚C
Bd2778 (method 1 & 2)	PEG dialysis (molecular cut off 3 KDa) at room temperature
Bd2778 (method 2)	Vivaspin 20 3/5 KDa spin concentrator at 20 ˚C
Bd1645, Bd1904, Bd2126	Vivaspin 20 3/5 KDa spin concentrator at 4 ˚C

2.3 Construction of Over-Expression Strains

The plasmid constructs were generated using restriction-free cloning (van den Ent and Löwe, 2006), which involves the amplification of the gene of interest with overlapping regions able to bind to the plasmid, followed by linear PCR to generate the plasmid containing the DNA fragment. DpnI digest was used to remove the parent plasmid (van den Ent and Löwe, 2006) before transformation of the PCR product. Colony PCR was used to identify colonies that had taken up the plasmid containing the DNA fragment. Expression construct designs used in this PhD project can be seen in Table 2-4.

2.3.1 Restriction-Free Cloning

The PCR reactions (100 µL) for gene amplification contained: 1 x HiFi buffer, 10 % DMSO, 0.1 mM dNTPs, 0.2 µM forward and reverse primer, 1 µL Velocity DNA Polymerase (BIOLINE), 1 µL HD100 template DNA and 63 µL sterile H_2O. This procedure utilised the following program: initial denaturation step at 98 ˚C for 2 min, followed by 30 cycles of denaturation at 98 ˚C for 30 s, annealing at a gradient between 50 ˚C and 60 ˚C for 30 s and extension at 72 ˚C for 45 s. This was followed by a final extension step at 72 ˚C for 10 min. Each PCR product (4 µL) was run on a 1 % agarose gel alongside 4 µL of 1 kb ladder (NEB),

Table 2-4: Expression Construct used in the PhD Thesis

Protein Name	Identifier	Description	Organism	Expression Construct	Cloning Method	New Plasmid Name
Bd0427	ΔSP	Full length mature protein excluding signal peptide, for inclusion body expression	*B. Bacteriovorus* HD100	pET28a	Restriction free	pET28a_*bd0427*
	ΔSP N-His	Full length mature protein excluding signal peptide with N-terminal His_6-tag, for inclusion body expression	*B. Bacteriovorus* HD100	pET28a	Restriction free	pET28a_*bd0427*_NHis
	ΔSP C-His	Full length mature protein excluding signal peptide with Tev cleavable C-terminal His_8-tag, for inclusion body expression	*B. Bacteriovorus* HD100	pET28a_*bd0427*	SDM	pET28a_*bd0427*_CHis
	A103His	Full length mature protein excluding signal peptide with His_6-tag (SSHHHHHHGSS) after residue A103, for inclusion body expression	*B. Bacteriovorus* HD100	pET28a_bd0427	SDM	pET28a_*bd0427*_A103His
	K260His	Full length mature protein excluding signal peptide with His_6-tag (SSHHHHHHGSS) after residue K260, for inclusion body expression	*B. Bacteriovorus* HD100	pET28a_bd0427	SDM	pET28a_*bd0427*_K260His
	Disulfide-bonded Plug	Full length mature protein excluding signal peptide, with following mutations A29C and S287C, for inclusion body expression	*B. Bacteriovorus* HD100	pET28a_bd0427	SDM	pET28a_*bd0427*_A29C-S287C
	Δplug	Bd0427 with no signal peptide and a deletion from residue 21-52, to remove N-terminal plug	*B. Bacteriovorus* HD100	pET28a_bd0427	SDM	pET28a_*bd0427*_Δplug

	pelB A103His	Full length mature protein excluding signal peptide, with *pelB* leader sequence and a His_6-tag (SSHHHHHHGSS) after residue A103	*B. Bacteriovorus* HD100	pET29b	Twist Biosciences Clonal Gene	pET29b_*bd0427_pelB*_A103His
Bd1645	Bd1645	Full length mature protein excluding signal peptide, with *pelB* leader sequence and C-terminal His_6-tag	*B. Bacteriovorus* HD100	pET26b	Restriction free	pET26b_*bd1645*
Bd2126	Bd2126	Full length mature protein excluding signal peptide, with *pelB* leader sequence and C-terminal His_6-tag	*B. Bacteriovorus* HD100	pET26b	Restriction free	pET26b_*bd2126*
Bd2778	Bd2778	Full length mature protein excluding signal peptide, with *pelB* leader sequence and C-terminal His_6-tag	*B. Bacteriovorus* HD100	pET26b	Restriction free	pET26b_*bd2778*
OmpF	OmpF	Full length mature protein excluding signal peptide, with a *tam*A signal sequence	*E. coli*	pB22	van den Berg, B lab	n/a

at 100 V for 60 min, visualised under UV light. The PCR reactions (100 µL) contained: 1x HiFi buffer, 0.1 mM dNTPs, 1 µL Velocity DNA polymerase, 1 µL plasmid, 4 µL DNA fragment and 23 µL sterile H_2O. This procedure utilised the following program: initial denaturation step at 98 ˚C for 2 min, followed by 30 cycles of denaturation at 98 ˚C for 30 s, annealing at a gradient between 55 ˚C and 65 ˚C for 1 min and extension at 72 ˚C for 10 min. This was followed by a final extension step at 72 ˚C for 10 min. Following this, 2 µL of Dpn1 was added to the PCR product and incubated at 37 ˚C for 2 hr, before a purification step using a QIAquick PCR Purification Kit (QIAGEN).

2.2.2 Colony PCR

To confirm the insert of the gene into the target plasmid, the plasmids were concentrated, and 50 ng was transformed into 50 µL of chemically competent *E. coli* XL10-Gold Ultracompetent cells (Agilent). Following overnight growth on a LB agar kanamycin plate, 10 colonies were selected for colony PCR.

All colonies selected were plated onto a new LB agar kanamycin plate. This PCR reaction (30 µL) contained: 1 colony, 1 x MyTaq Red Mix (BIOLINE) and 0.2 µM pET forward and reverse primers, made up to 30 µL with sterile H_2O. This procedure utilised the following program: initial denaturation step at 98 ˚C for 8 min, followed by 35 cycles of denaturation at 98 ˚C for 24 s, annealing at 55 ˚C for 45 s and extension at 72 ˚C for 2.25 min, with a final extension step at 72 ˚C for 10 min. Each PCR product was run on a 1 % agarose gel alongside 4 µL of 1 kb ladder (NEB), at 100 V for 60 min, visualised under UV light.

Successful colonies were grown in LB media overnight and plasmids were extracted using QIAprep Miniprep Kit (QIAGEN) before sequencing. Plasmids containing the correct DNA fragments were transformed into an *E. coli* expression strain (Figure 0-3, Appendicies) for expression.

2.3 Site-directed Mutagenesis

To introduce substitutions, insertions and deletions into the expression plasmids site-directed mutagenesis was conducted; mutated constructs details are provided in Table 2-4. Blunt ended primers were designed using NEBasechanger tool and the PCR reactions (25 µL) contained: 1x Q5 master mix, 0.5 µM forward and reverse primers, 25 ng template DNA made up to the final volume with sterile H_2O. This utilised a program of an initial denaturing step at 98 ˚C for 30 s, followed by 25 cycles or denaturation at 98 ˚C for 30s, annealing at 50-65 ˚C for 15 s, and extension at 72 ˚C for 4 min with a final extension step at 72 ˚C for 10 min. Successful PCR was determined by running products on an agarose gel, at 100 V for 60 min with visualisation using UV light, where linearised plasmid could be seen. These products underwent kinase, ligase and DpnI (KLD) treatment to phosphorylate the 5' end, re-circularise the plasmid and digest the methylated parent plasmid. The KLD reaction (10 µL) contained: 1x T4 ligase buffer, 1 µL product, 0.75 µL polynucleotide kinase (NEB), 0.75 µL T4 ligase (NEB), 0.5 µL DpnI (NEB) made up to final volume with sterile H_2O. This was left at room temperature for 1 hr, before transformation in 50 µL chemically competent *E. coli* XL10-GOLD cells. Individual colonies were selected for overnights, with plasmids extracted using QIAprep Miniprep Kit (QIAGEN) before sequencing. Plasmids containing the correct DNA fragments were transformed into an *E. coli* expression strain for expression.

2.4 Protein Growth and Expression

2.4.1 Expression Trials

To optimise expression, small batches (20 mL) of LB and 2LB media (for inclusion body and membrane protein expression) or LB, TB and AI media (soluble protein expression) were made (Table 2-5), supplemented with relevant antibiotic (Table 2-1) and inoculated with 200 µL of

overnight culture. AI media was trialled with each of the sugar mixes with and without 0.5 % (w/v) glucose. All small batches were incubated at 37 ˚C, shaking at 180 r.p.m., until an OD_{600} of 0.6. LB, 2LB and TB were induced with 0.1, 0.5 or 1 mM IPTG. Membrane and soluble protein expression trials were grown overnight at 18 ˚C and inclusion body expression trials were grown at 37 ˚C for 4 hrs. 10 mL of culture was pelleted at 3220 g and resuspended in 1 mL buffer (20 mM HEPES (7.5), 300 mM NaCl), before sonication using microtip for 30 s. The lysed samples were centrifuged, 16,873 g for 10 min, and the supernatant and pellet were run on an SDS-PAGE alongside cell pellet before induction.

Table 2-5: Media Solutions

Mixture	Concentrations
2LB Media	20 $g.L^{-1}$ tryptone, 10 $g.L^{-1}$ yeast extract and 10 $g.L^{-1}$ NaCl
Lysogeny broth (LB) Media	10 $g.L^{-1}$ tryptone, 5 $g.L^{-1}$ yeast extract and 10 $g.L^{-1}$ NaCl
Terrific broth (TB) Media	12 $g.L^{-1}$ tryptone, 24 $g.L^{-1}$ yeast extract, 9.4 $g.L^{-1}$ K_2HPO_4, 2.2 $g.L^{-1}$ KH_2PO4 and 4 $g.L^{-1}$ glycerol
Autoinduction (AI) Media	10 $g.L^{-1}$ tryptone, 5 $g.L^{-1}$ yeast extract, 3.55 $g.L^{-1}$ Na_2HPO_4, 3.4 $g.L^{-1}$ KH_2PO_4 and 2.675 $g.L^{-1}$ NH_4Cl, 2.5 % trace metals, 2 mM $MgSO_4$. Plus 20 mL sugar mix
Minimal Media	1x M9 salt solution, 0.4 % glycerol, 1 mM $MgSO_4$, 1 mM $CaCl_2$, 35 µM $FeCl_2$, 1 $\mu g.mL^{-1}$ biotin, 1 $\mu g.mL^{-1}$ thiamine, 1 % trace metals
10x M9 salt solution	75.2 $g.L^{-1}$ Na_2HPO_4:$2H_2O$, 30 $g.L^{-1}$ KH_2PO_4, 5 $g.L^{-1}$ NaCl and 5 $g.L^{-1}$ NH4Cl
Trace metals	130 mM Na_2EDTA, 80 mM $ZnSO_4$, 180 mM H_3BO_3, 30 mM $MnCl_2$, 7 mM $CoCl_2$, 6 mM $CuSO_4$ and 0.8 mM $(NH_4)_6Mo_7O_{24}$
Amino acid mix	1 g lysine, 1 g threonine, 1 g phenylalanine, 0.5 g leucine, 0.5 g isoleucine, 0.5 g valine, 0.6 g selenomethionine
Sugar Mix A	95 $g.L^{-1}$ lactose and 185 $g.L^{-1}$ glycerol
Sugar Mix B	190 $g.L^{-1}$ lactose and 185 $g.L^{-1}$ glycerol
Sugar Mix C	95 $g.L^{-1}$ lactose and 277.5 $g.L^{-1}$ glycerol
Sugar Mix D	190 $g.L^{-1}$ lactose and 277.5 $g.L^{-1}$ glycerol

2.4.2 Expression of Protein in Inclusion Bodies (Native)

For large scale protein expression in inclusion bodies of *bd0427* in *E. coli* BL21(DE3) cells, 1 L LB media was inoculated with 10 mL overnight culture, supplemented with 100 μg.mL^{-1} kanamycin. The culture was incubated at 37 ˚C, shaking at 180 r.p.m. until an OD_{600nm} of 0.6. Protein expression was induced with 1 mM IPTG, with cells harvested after a 4 hrs expression period, shaking at 180 r.p.m. at 37 ˚C.

2.4.3 Expression of Protein in Inclusion Bodies in Presence of Selenomethionine

For large scale protein expression of pET28a-*bd0427* in *E. coli* BL21(DE3) cells in the presence of selenomethionine, cells from a 20 mL overnight culture were pelleted by centrifugation at 3,220 g for 5 min at 4 ˚C and resuspended in 10 mL minimal media (Table 2-5). The resuspended culture was added to 1 L of minimal media, supplemented with 100 μg.mL^{-1} kanamycin, and incubated at 37 ˚C, shaking at 180 r.p.m. until an OD_{600nm} of 0.4. Here 0.51 g of amino acid mix (Table 2-5), was added and the culture was incubated at 37 ˚C for 15 min, before protein expression was induced with 1 mM IPTG. Cells were harvested after 4 hrs expression at 37 ˚C.

2.4.4 Expression of Membrane Proteins

For large scale protein expression of pET26b-*bd0427_pelB*_A103His in *E. coli* BL21(DE3) RIPL cells (NEB), 1 L LB media was inoculated with 10 mL overnight culture, supplemented with 100 μg.mL^{-1} kanamycin and 34 μg.mL^{-1} chloramphenicol. The culture was incubated at 37 ˚C, shaking at 180 r.p.m. until an OD_{600nm} of 0.6. Protein expression was induced with 0.5 mM IPTG, with cells harvested after an overnight expression period, shaking at 180 r.p.m. at 18 ˚C.

2.4.5 Expression of Soluble Proteins

Large scale expression of soluble proteins (Bd1645, Bd2126, Bd2778) were expressed in the media listed in Table 2-6. Two flasks containing 1 L of media were inoculated with 10 mL overnight culture, supplemented with relevant antibiotic/s (Table 2-6). The culture was incubated at 37 ˚C, shaking at 180 r.p.m. until an OD_{600nm} of 0.6. For AI cultures the temperature was reduced to 18 ˚C during protein expression and for other cultures IPTG was used to induce protein expression, followed by a reduction in temperature to 18 ˚C. The cultures were left at 18 ˚C, shaking at 180 r.p.m. overnight before harvesting.

Table 2-6: Media and Induction Protocol for Soluble Proteins

Plasmid	*E. coli* Expression Strain	Media	Sugar Mix/IPTG concentration	Antibiotic resistance
pET26b_*bd1645*	Codon Plus(DE3) RIPL	AI	Sugar Mix C	Chloramphenicol, Kanamycin
pET26b_*bd2126*	Codon Plus(DE3) RIPL	TB	0.5 mM IPTG	Chloramphenicol, Kanamycin
pET26b_*bd2778*	BL21(DE3)	AI	Sugar Mix C	Kanamycin

2.5 Purification of Proteins

2.5.1 Purification of Protein from Inclusion Body Material

This section focuses on the purification of all Bd0427 constructs excluding Bd0427 *pelB* A103His.

2.5.1.1 Isolation of the Inclusion Body Material

Cells were harvested by centrifugation, 6,240 g for 15 min, before being resuspended in resuspension 1 buffer (6 mL per 1 g) (Table 2-7). Resuspended cells were lysed by sonication (Ultrasonic processor 500 w, 20 kHz at 40 % amplitude, (Fisher Scientific)) pulsing for 20 s ‘on’, 20 s ‘off’, for a total of 2 min ‘on’. To separate the soluble and insoluble fractions a

centrifugation step, 48,380 g for 1 hr, was carried out; following this, the soluble fraction (lysate) was discarded.

The insoluble pellet then underwent extensive wash steps to separate the cell debris from the inclusion body material. Insoluble pellet was resuspended in 20 mL of wash A buffer (Table 2-7), using a magnetic flea for 1 hr at room temperature. This was followed by a centrifugation step, 48,380 g for 30 min, with the supernatant being discarded. The wash A step was repeated twice. A final two wash steps involving resuspension in wash B buffer (Table 2-7) using a magnetic flea for 30 min at room temperature. Followed by pelleting step, 48,380 g for 30 min. The inclusion body pellet was then stored at – 20 ˚C ready for re-folding.

Table 2-7: Buffers used for Isolation of Inclusion Material

Buffer name	Components
Resuspension 1	50 mM tris (pH 8), 5 mM EDTA
Wash A buffer	50 mM tris (pH 8), 5 mM EDTA, 2 % (w/v) TritonX-100
Wash B buffer	50 mM tris (pH 8), 5 mM EDTA

2.5.1.2 Refolding Inclusion Body Material

Inclusion body pellet (approximately 200 mg wet weight) was tumbled in 10 mL of 50 mM tris (pH 8), 100 mM NaCl, 8M Urea, to solubilise the unfolded protein, resulting in a final concentration of 0.1 mM protein, before the concentration was measured the insoluble material was pelleted by centrifugation, 48,380 g for 20 min. The soluble protein-containing fraction was added dropwise to 150 mL refold buffer (50 mM tris (pH 8), 100 mM NaCl and detergent (1% (w/v) DDM (Melford), 2 % (w/v) β-OG (Melford) and 0.625 % (v/v) TritonX-100, were all trialled), stirred slowly by a small magnetic flea. Aggregated protein was removed from the re-fold buffer by filtration (0.22 μm). Refolding success was determined using a semi-native PAGE, owing to the heat modifiability of outer membrane proteins, where folded protein (un-boiled sample runs further into the gel than unfolded protein (boiled). This method was used for Bd0427 constructs ΔSP, ΔSP N-His and ΔSP C-His.

2.5.1.3 Size Exclusion Chromatography (SEC) of Refolded Protein

To separate homologous samples following refolding, SEC was used. Initially, protein refolded in 1 % (w/v) DDM was screened with three buffers (DDM 1-3, Table 2-8) on an Analytical Superdex 75 10/300, then on an Analytical Superdex 200 10/300 with DDM 3 buffer. The protein elution was monitored by UV at 280 nm, with protein containing fractions ran on a semi-native PAGE to assess oligomeric state and purity.

Following this initial screen all SEC was performed using a HiLoad 16/600 or 26/60 Superdex 200 prep grade column attached to an ÄKTA pure (Cytiva). The column was pre-equilibrated with 1.2 column volume (c.v.) of filtered and degassed SEC buffer (Table 2-8) before the protein was loaded with SEC buffer 3 times the sample volume (e.g. 5 mL sample volume 15 mL SEC buffer was used to empty the loop). Elution was carried at 1 $mL.min^{-1}$ over 1.2 c.v. protein elution was monitored by UV at 280 nm. Fractions containing protein were ran on a semi-native PAGE to compare oligomeric states, for pooling and concentrating. Following concentration samples were either used for on-going experiments or flash frozen in 100 µL aliquots in liquid nitrogen and stored at -80 ˚C.

Table 2-8: SEC buffers for Refolded Protein

Buffer name	Components
DDM 1 buffer	50 mM tris-HCl (pH 8), 150 mM NaCl, 0.03 % (w/v) DDM
DDM 2 buffer	50 mM tris-HCl (pH 8), 150 mM NaCl, 0.1 % (w/v) DDM
DDM 3 buffer (final)	50 mM tris-HCl (pH 8), 150 mM NaCl, 0.3 % (w/v) DDM
DDM Se-Met buffer	50 mM tris-HCl (pH 8), 150 mM NaCl, 0.3 % (w/v) DDM, 3mM β-mercaptoethanol
β-OG buffer	50 mM tris-HCl (pH 8), 150 mM NaCl, 50 mM β-OG
β-OG Se- Met buffer	50 mM tris-HCl (pH 8), 150 mM NaCl, 50 mM β-OG, 3mM β-mercaptoethanol

2.5.1.4 Immobilised Metal Affinity Chromatography (IMAC) Purification of Refolded Protein

Purification of Bd0427 ΔSP C-His was also trialled with the addition of IMAC before SEC. Refolded protein was cycled over a 1 mL HisTrap HP column, pre-equilibrated with refold buffer, using a peristaltic pump at low speed (0.1 $ml.min^{-1}$) for 1 hr. The column was washed with 20 x c.v. DDM 3 buffer (Table 2-9), before elution of 1 mL fractions with elution 1a buffer (Table 2-9). A semi-native PAGE was ran of eluted fractions before SEC.

2.5.1.5 On-column Detergent Exchange of Refolded Protein

To allow screening of other detergents an on-column detergent exchange method was designed. As with section 2.5.1.4, refolded protein was cycled over a 1 mL HisTrap HP column, pre-equilibrated with refold buffer, using a peristaltic pump at low speed for 1 hr. The resin was then washed with 30 c.v. DDM 3 buffer (Table 2-9), followed by 30 c.v. DDM + new detergent buffer (Table 2-9), then 30 c.v. new detergent buffer (Table 2-9). The protein was then eluted using elution buffer with new detergent (Table 2-9). Fractions were assessed using a semi-native PAGE before SEC.

Table 2-9: Buffers used in IMAC of Refolded Protein

Buffer name	Components
DDM 3 buffer (final)	50 mM tris-HCl (pH 8), 150 mM NaCl, 0.3 % (w/v) DDM
Elution 1a buffer	50 mM tris-HCl (pH 8), 150 mM NaCl, 0.3 % (w/v) DDM, 300 mM Imidazole
DDM + C_8E_4 buffer	50 mM tris-HCl (pH 8), 150 mM NaCl, 0.15 % (w/v) DDM, 0.15 % (w/v) C_8E_4
C_8E_4 buffer	50 mM tris-HCl (pH 8), 150 mM NaCl, 0.3 % (w/v) C_8E_4
Elution 1b buffer	50 mM tris-HCl (pH 8), 150 mM NaCl, 0.3 % (w/v) C_8E_4, 300 mM Imidazole
DDM + β-OG buffer	50 mM tris-HCl (pH 8), 150 mM NaCl, 0.15 % (w/v) DDM, 25 mM β-OG
β-OG buffer	50 mM tris-HCl (pH 8), 150 mM NaCl, 50 mM β-OG
Elution 1c buffer	50 mM tris-HCl (pH 8), 150 mM NaCl, 50 mM β-OG, 300 mM Imidazole

2.5.2 Purification of Membrane Proteins

This section discusses the purification protocol of Bd0427 *pelB* A103His.

2.5.2.1 Isolation of Membrane and Solubilisation of Membrane Protein

The cell pellet, pelleted by centrifugation at 6,240 g for 15 min, was resuspended in 25 mL resuspension buffer 2 (1xPBS, 5 mM EDTA and EDTA-free mini cOmplete protease inhibitor tablet (Merck)) before lysis. The resuspended pellet was lysed using an Emulsiflex C3 disruptor (Avastin), cycled through the machine three times at a pressure of 17 000 psi. To remove cell debris the lysed cells were centrifuged at 10,000 g for 30 min. The soluble fraction was centrifuged again in an ultra-centrifuge, at 150,000 g for 1 hr, to pellet the membranes. The membrane was resuspended in solubilisation buffer (50 mM Tris (pH 8), 150 mM NaCl, 1 % (w/v) DDM, 1 ml per 40 mg membrane) using a hand homogeniser, before rotating at 4 ˚C overnight. Un-solubilised membranes were then pelleted by a final ultra-centrifugation step, 150,000 g for 30 min, with the soluble membrane ready for purification.

2.5.2.2 Immobilised Metal Affinity Chromatography (IMAC) of Membrane Protein

The soluble membrane fraction was added to pre-equilibrated nickel-NTA metal chelate beads (1 mL beads per 100 mL solubilised membrane) and left rotating for 4 hr at 4 ˚C, to allow the protein to bind to the beads. The flow through was collected using a gravity flow column before two washes (5 c.v.) at increasing concentrations of imidazole (50 mM Tris (pH 8), 150 mM NaCl, 0.3 % (w/v) DDM, 10 mM then 25 mM Imidazole). The protein was eluted with buffer containing 200 mM Imidazole. The fractions were run on an SDS-PAGE to assess purity of the

fractions. Fractions containing protein of the correct molecular weight were pooled and concentrated.

2.5.2.3 Size Exclusion Chromatography (SEC) of Membrane Protein

SEC was performed using a Superdex 200 (either 10/300 or 16/60 column) attached to an ÄKTA prime (Cytiva). The column was pre-equilibrated with 1.2 c.v. buffer (50 mM Tris (pH 8), 150 mM NaCl, 0.3 % (w/v) DDM). The protein was loaded onto the column using 3 times the sample volume of buffer and elution was performed at a flow rate of 0.5 $mL.min^{-1}$ (10/300) or 1 $mL.min^{-1}$ (16/60). The protein elution was monitored using UV absorbance at 280 nm. The protein containing fractions were ran on a semi-native PAGE to assess purity and oligomeric state.

2.5.3 Purification of Soluble Proteins

This section discusses the purification protocols of Bd1645, Bd2126, and Bd2778.

2.5.3.1 Isolation of Soluble Protein

Cells expressing soluble protein were pelleted by centrifugation, 6,240 g for 15 min, before resuspension in resuspension buffer 3 (20 mM HEPES (pH 7.5), 300 mM NaCl, 20 mM Imidazole), Bd2778 also had 20 mg of DNase I added at this stage. The resuspended cells were lysed using an Emulsiflex C3 disruptor (Avastin), at a pressure of 17 000 psi, cycled through the machine three times. The cell debris was then pelleted using centrifugation, 48,380 g for 1 hr, leaving the soluble fraction in the supernatant.

2.5.3.2 Immobilised Metal Affinity Chromatography (IMAC) of Soluble Protein

Table 2-10: IMAC ÄKTA Programs used for Soluble Proteins

Protein	Buffer B	Wash Step	Gradient	Elution

Bd1645	20 mM HEPES (pH 7.5), 300 mM NaCl, 500 mM Imidazole	10 c.v.	0 – 100 % Buffer B over 20 c.v.	As part of gradient
Bd2126	20 mM HEPES (pH 7.5), 300 mM NaCl, 500 mM Imidazole	10 c.v.	0 – 30 % Buffer B over 10 c.v. 30 % Buffer B over 10 c.v. 30 – 100 % Buffer B over 10 c.v.	As part of gradient
Bd2778 (method 1)	20 mM HEPES (pH 7.5), 300 mM NaCl, 300 mM Imidazole	10 c.v.	0 – 30 % Buffer B over 10 c.v.	Eluted with 20 mM HEPES (pH 7.5), 300 mM NaCl, 50 mM EDTA over 3 c.v.
Bd2778 (method 2)	20 mM HEPES (pH 7.5), 300 mM NaCl, 500 mM Imidazole	15 c.v.	0 – 15 % Buffer B over 10 c.v. 15% Buffer B over 5 c.v. 15 – 100 % Buffer B over 5 c.v.	As part of gradient

The soluble fraction was filtered using a 0.4 μm PES filter, before cycling the lysate onto a pre-equilibrated 5 mL His-Trap column using a peristaltic pump at low speed for 1 hr, all on ice excluding Bd2778 (method 2) which was purified at room temperature. The protein was washed and eluted on a ÄKTA pure, using a gradient IMAC program at 5 $mL.min^{-1}$ (Table 2-10, shows details for each protein) The protein elution was monitored by UV at 280 nm, with fractions of any peaks ran on an SDS-PAGE to assess purity. Pure protein was pooled and either concentrated for SEC or dialysed (Table 2-11) and then concentrated for protein crystallisation or SEC.

Table 2-11: Dialysis Buffers of Soluble Proteins

Protein	Dialysis
Bd1645	20 mM Citrate (pH 6), 300 mM NaCl
Bd2126	No dialysis
Bd2778 (method 1)	20 mM HEPES (pH 7.5), 300 mM MgCl
Bd2778 (method 2)	20 mM HEPES (pH 7.5), 300 mM NaCl

2.5.3.3 Size Exclusion Chromatography (SEC) of Soluble Protein

All soluble proteins were ran on a Superdex 75 26/60 HiLoad column. The column was equilibrated in 1.2 c.v. of the same buffer as used dialysis (Table 2-11, Bd2126 used same buffer as Bd2778 (method 2)) on an ÄKTA pure. The samples were loaded using three-times the loop volume of buffer and the protein was eluted at 2 $mL.min^{-1}$ in 3 mL fractions. The elution was monitored by UV at 280 nm and an SDS-PAGE was ran of protein fractions. Pure protein was pooled and concentrated for follow on experiments.

2.6 X-ray Crystallography

2.6.1 Protein Crystallisation Experiments

Crystallisation of all proteins followed the same initial process. Initial screening was conducted using commercially available crystal screens (Table 2-12). They were mostly set up manually in CombiClover 96-well sitting-drop crystallisation plates (Molecular Dimensions), with 100 µL of precipitant mix in the reservoirs and 1 µL of protein sample in the well mixed with the reservoir solution at 1:1 ratio. Alternatively, some trays were set up using a Mosquito (SPT Labtech) robot using Swissci 3-drop 96-well plates (Molecular Dimensions), with 35 µL of precipitant mix in the reservoirs and 400 nL of protein sample in the well, again at a ratio of 1:1 (protein: reservoir solution). The proteins were typically screened at concentrations between 4 – 30 $mg.mL^{-1}$, with the crystallisation plates left at 18 ˚C. The crystal trays were then checked manually at 1 day, 1 week, 1 month and before every synchrotron trip. Successful crystallisation resulted in crystals being harvested or optimisation experiments being performed.

Table 2-12: Initial Crystallisation Screens

Protein	Buffer	Screens set up	Concentration	Method Used
Bd0427 ΔSP	50 mM Tris (pH 8), 150 mM	MIDASplus, Morpheus	7.5 $mg.mL^{-1}$	Manual

	NaCl, 0.3 % DDM			
	50 mM Tris (pH 8), 150 mM NaCl, 50 mM β-OG	MemGold, MemPlus, MIDASplus	4.5 mg.mL^{-1}	Manual
Bd0427 ΔSP Se-Met	50 mM Tris (pH 8), 150 mM NaCl, 0.3 % DDM, 2 mM β-mercaptoethanol	MIDASplus, Morpheus	9 mg.mL^{-1} (Pool 1), 4.5 mg.mL^{-1} (Pool 2)	Manual
		MemGold, MemPlus, MIDASplus	5.8 mg.mL-1 (Pool 3), 4 mg.mL^{-1} (Pool 4)	Mosquito
	50 mM Tris (pH 8), 150 mM NaCl, 50 mM β-OG, 3 mM β-mercaptoethanol	MemGold, MIDASplus	4.5 mg.mL^{-1}	Manual
	50 mM Tris (pH 8), 150 mM NaCl, 50 mM β-OG, 3 mM β-mercaptoethanol 3 % 1,2,3-Heptanetriol	MemGold, MemGold2	4.5 mg.mL^{-1}	Manual
Bd0427 ΔSP C-His	50 mM Tris (pH 8), 150 mM NaCl, 0.3 % DDM	MemGold, MIDASplus	7 mg.mL^{-1} (Pool 1), 6.5 mg.mL^{-1} (Pool 2)	Mosquito
	50 mM Tris (pH 8), 150 mM NaCl, 0.3 % DDM (after HisTrap & SEC)	MemGold, MemGold2, MIDASplus	9 mg.mL^{-1}, 4.5 mg.mL^{-1}, 9 mg.ml^{-1} & 0.9 mg.mL^{-1} Tev	Mosquito
		MemGold	9 mg.mL^{-1}	Manual
	50 mM Tris (pH 8), 150 mM NaCl, 50 mM β-OG	MemGold, MemGold2, MIDASplus	9 mg.mL^{-1}, 4.5 mg.mL^{-1}, 9 mg.ml^{-1} & 0.9 mg.mL^{-1} Tev	Mosquito
Bd0427 Disulfide-bonded Plug	50 mM Tris (pH 8), 150 mM NaCl, 0.3 % DDM	MemGold, MIDASplus	5 mg.mL-1	Manual
Bd1645	20 mM Citrate (6), 300 mM NaCl	BCS Screen, JCSGplus, MIDASplus, PACT Premier	28 mg.mL^{-1}, 20 mg.mL^{-1}, 10 mg.mL^{-1}	Mosquito

		BCS Screen, JCSGplus, MIDASplus, Morpheus, PACT Premier	4.5 mg.mL-1	Mosquito
	(Dialysed) 20 mM MES (6), 300 mM NaCl	BCS Screen, JCSGplus, MIDASplus	5.8 $mg.mL^{-1}$	Manual
Bd2126	20 mM HEPES (7.5), 300 mM NaCl	JCSGplus, MIDASplus	9 $mg.mL^{-1}$	Mosquito, hanging drop
		BCS Screen, PACT Premier, MIDASplus, Morpheus	5 mg.mL-1	Mosquito
Bd2778 (method 1)	20 mM HEPES (7.5), 300 mM $MgCl_2$	JCSGplus, MIDASplus, PACT Premier	6 $mg.mL^{-1}$	Manual
		BCS Screen, Morpheus, ProPlex	5 $mg.mL^{-1}$	Manual
		Mini Screen	5 $mg.mL^{-1}$	Manual, 24-well hanging drop
		JCSGplus	5.6 $mg.mL^{-1}$	Manual
		JCSGplus, MIDASplus, Morpheus, PACT Premier, ProPlex	7 $mg.mL^{-1}$, 4.5 $mg.mL^{-1}$	Mosquito
Bd2778 (method 2), PEG Dialysis	20 mM HEPES (7.5), 300 mM $MgCl_2$	JCSGplus, MIDASplus	7 $mg.mL^{-1}$	Manual
	20 mM HEPES (7.5), 300 mM NaCl, 2 mM EDTA	JCSGplus, Morpheus, PACT Premier, ProPlex, JB Classic 1, 2, 3, 4, 6, 7, 8, 10	7.5 $mg.mL^{-1}$, 3.75 $mg.mL^{-1}$	Mosquito
Bd2778 (method 2), spin concentrator	20 mM HEPES (7.5), 300 mM NaCl, 2 mM EDTA	JCSGplus	8.8 $mg.mL^{-1}$	Manual
		BCS Screen, MIDASplus	8.8 $mg.mL^{-1}$	Mosquito

	PACT Premier	5 mg.mL-1	Mosquito
20 mM HEPES (7.5), 300 mM NaCl	BCS Screen, JCSGplus, MIDASplus, Morpheus, PACT Premier	5.9 mg.mL^{-1}	Mosquito

Optimisation was selected when there was only one crystal (and a greater number were wanted), if the crystals were small and not possible to harvest, or if the crystals were of a less than ideal morphology (fine needles, thin plates, multiple nucleation events, etc.). Crystals were optimised using two methods. The first involved creating a 24-well screen based on the precipitant mix in the original condition. These could be set up as either sitting-drop or hanging-drop vapour diffusion screens. The protein:reservoir drop was set up at a 1:1 ratio in these screens unless specified otherwise, with the reservoir containing 500 μL of precipitant mix. The second involved producing a seed stock of the original crystal hit, using a MicroSeed bead (Molecular Dimensions), the crystal and the precipitant buffer at multiple dilutions, and using this alongside protein in either commercially available crystallisation screens or 24-well optimisation plates. During this PhD, multiple optimisation screens were set up (Appendix 4).

2.6.2 Crystal harvesting

Crystals destined for X-ray diffraction need to be harvested from their drops and cryoprotected prior to transport and diffraction. Crystals were harvested using round or mesh Litholoops (Molecular Dimensions) of varying sizes (0.05 – 0.3 mm), the crystal was then either flash frozen in liquid nitrogen, for those in buffers with cryoprotectants of appropriate concentrations (Table 2-13), or soaked in a cryoprotectant solution before freezing. Cryoprotectant solutions were made by either making up the crystallisation buffer with increased concentration of a cryoprotectant within the condition (e.g. 10 % glycerol is increased to 20 %) or when making up the crystallisation buffer a cryoprotectant was added in stages (e.g. 0 % cryoprotectant to 20

– 30 % cryoprotectant). For crystals of Bd0427 constructs when additional cryoprotectants were required either 20 % (v/v) ethylene glycol or 20 % (v/v) PEG 400 was used. For Bd1645, Bd2126 or Bd2778 crystals either 20 % (v/v) glycerol or 20 % (v/v) ethylene glycol was used. For heavy-atom phasing crystals were transferred to cryoprotectant buffer containing either 10 mM tetrachloroplatinate or 20 mM sodium orthovanadate and were left to soak for a period of time (10 s – 1 hr) before snap cooling.

Table 2-13: List of Cryoprotectants

Cryoprotectant	Concentration
Glycerol	≥ 20 % (v/v)
Ethylene Glycol	≥ 20 % (v/v)
Glycerol ethoxylate	≥ 30 % (v/v)
PEG 300, 350 MME, 400, 500 MME, 600	≥ 20 % (v/v)
PPG 400	≥25 % (v/v)
Pentaerythritol ethoxylate (15/4 EO/OH)	≥25 % (v/v)
Pentaerythritol propoxylate (5/4 PO/OH)	≥25 % (v/v)

2.6.3 X-ray Diffraction and Data Collection

All diffraction data were collected at either the I03 or I04 Beamline at Diamond Light Source (Oxford, UK). Crystals under a stream of nitrogen were centred manually, before screening (short data collection every 90˚) to enable estimation of resolution and space group to allow selection of data collection parameters. For crystals of protein expressed in the presence of selenomethionine a fluorescence edge scan was used to determine the experimental values of f'/f' anomalous scattering factors. For heavy atom-soaked crystals the wavelength selected was determined using the Washington Anomalous Scattering website (Cromer and Liberman, 1970; Merritt, 2012). The diffraction data was auto processed by the ISPyB automated pipelines (Xia2 dials, Xia2 3dii, autoPROC (Gildea et al., 2022, p. 2; Vonrhein et al., 2011; Winter et al., 2022)), before use in structural determination.

2.6.4 Structural Determination and Refinement

2.6.4.1 Solving of Bd0427 using Phasing

As Bd0427 shares no significant homology with proteins of known structure, molecular replacement was not possible and heavy atom phasing was used, using single-wavelength anomalous dispersion (SAD). Automated data that had anomalous signal correlation coefficient (CCanom) values above 0.5 were selected for structural determination. To build an initial model the CRANK2 pipeline on CCP4cloud (Skubák, 2022) was used, with the input of diffraction data (xia2 dials mtz), protein sequence and the atom type that provides anomalous scattering (Se). This pipeline involves FA (structure factor amplitude) estimation, substructure determination, substructure refinement and phasing, density modification, hand determination and combined model building (Skubák, 2022; Skubák and Pannu, 2013). The model was completed by iterative cycles of refinement in Phenix Refine and manual building with Coot (Afonine et al., 2012; Emsley et al., 2010).

2.6.4.2 Solving of Bd0427 using Molecular Replacement

Native datasets or datasets with a CCanom below 0.5 were solved using the pentameric model produced in section 2.6.4.1 as the search model for molecular replacement using Phenix PHASER (McCoy et al., 2007). The TFZ scores for all datasets solved using molecular replacement were above 35, indicating the data is solved. The model was then completed by cycles of refinement and manual building in Coot as in section 2.6.4.1.

2.6.4.3 Solving of Bd2778 using Molecular Replacement with AlphaFold2 Model

A native data set was collected of Bd2778. Using the Matthews probability (Matthews, 1968; Weichenberger and Rupp, 2014) it was predicted that there were likely 22 copies of Bd2778 in

the asymmetric unit with a solvent content of 47.53 %. Initially an attempt was made to solve the structure using the structure of sequence homologue Bd1904 (structure unpublished) with 13.45% identity using molecular replacement. This did not yield a solution.

Following this, attempts were made to recrystallise the protein in the same conditions, to enable heavy-atom phasing. This proved difficult and a heavy-atom dataset was not obtained.

Following the public release of ColabFold: AlphaFold (Jumper et al., 2021; Mirdita et al., 2022), this was used to generate a de-novo model of Bd2778. This model was initially used as the search model for molecular replacement in Phenix PHASER, searching for 8-22 copies. This did not yield a solution. This model was trimmed to remove the flexible regions before use as a new search model. Searching for 8 copies in the ASU resulted in a TFZ above 8. The model was then built using Phenix Autobuild, CCP4 MolRep with a density modified map and fixed model, and manual building in Coot (Terwilliger, 2013; Terwilliger et al., 2008; Vagin and Teplyakov, 2010, 1997). The model was finished with cycles of refinement using Phenix Refine and Coot.

2.7 Cryo-Electron Microscopy

2.7.1 Vitrification and Data Collection

Graphene oxide coated grids were made by adding 3 µL of graphene oxide solution (0.2 $mg.mL^{-1}$) (Sigma) to Quantifoil R1.2/1.3 200-mesh grids, which had been glow discharged for 30 s using GloQube glow discharge unit. The grid was then incubated for 1 min, and the excess liquid was blotted using filter paper before the grid was washed three times with H_2O.

The prepared graphene oxide grid was held by tweezers and added to the FEI Vitrobot mark IV chamber set at 4 ˚C with 100 % humidity. 3 µL of Bd0427 A103His (0.02 $mg.mL^{-1}$ in 50 mM

Tris (pH 8), 150 mM NaCl and 0.03 % (w/v) DDM) was applied to the grid, which was blotted for 30 s (blot force 10) before plunging into liquid ethane.

The vitrified grid was screened before data collection at 300 kV on a FEI Titan Krios G3 microscope (Themo Fisher Scientific) at the Midlands Regional CryoEM facility (University of Leicester). The data collection was automated using EPU software at a magnification of 105000 x, with a defocus range of -1 to -2.4 µm and a pixel size of 0.835 Å. The total dose on the specimen was 55.4 $e^-/Å^2$ over 2 s in 50 frames. This resulted in 10,312 micrographs.

2.7.2 Data Processing (Collaboration) and Model Building

The initial 2D-classifications were generated using Relion 4.0 (Kimanius et al., 2021). The motion and contrast transfer function (CTF) correction were conducted on the micrographs using MotionCor2 and CtfFind4, respectively (Rohou and Grigorieff, 2015; Zheng et al., 2017). The poor-quality data, caused by ice crystal formation, was removed manually resulting in 4,736 micrographs used for particle picking. 2,967,865 particles were picked by Laplacian-of-Gaussian picking and extracted with a 180 px mask for 2D-classification. A sub-selection of the resulting 2D-classifications was performed resulting in 300,234 particles, these were extracted with a mask of 140 px and used in the final 2D-classification.

The electron density map and *ab inito* model generation were performed by our collaborator (Giedre Ratkevičiūtė), using CryoSPARC (Jamali et al., 2022; Punjani et al., 2017; Sanchez-Garcia et al., 2021) according to the workflow in Figure 2-7. The final model was built manually in Coot using a combination of the ModelAngelo model generated by our collaborator and the Bd0427 crystal structure fitted into the cryoEM density map using Phenix 'Dock in Map'. The model then underwent cycles of refinement using Phenix Real-space refine (Afonine et al., 2018) and Coot.

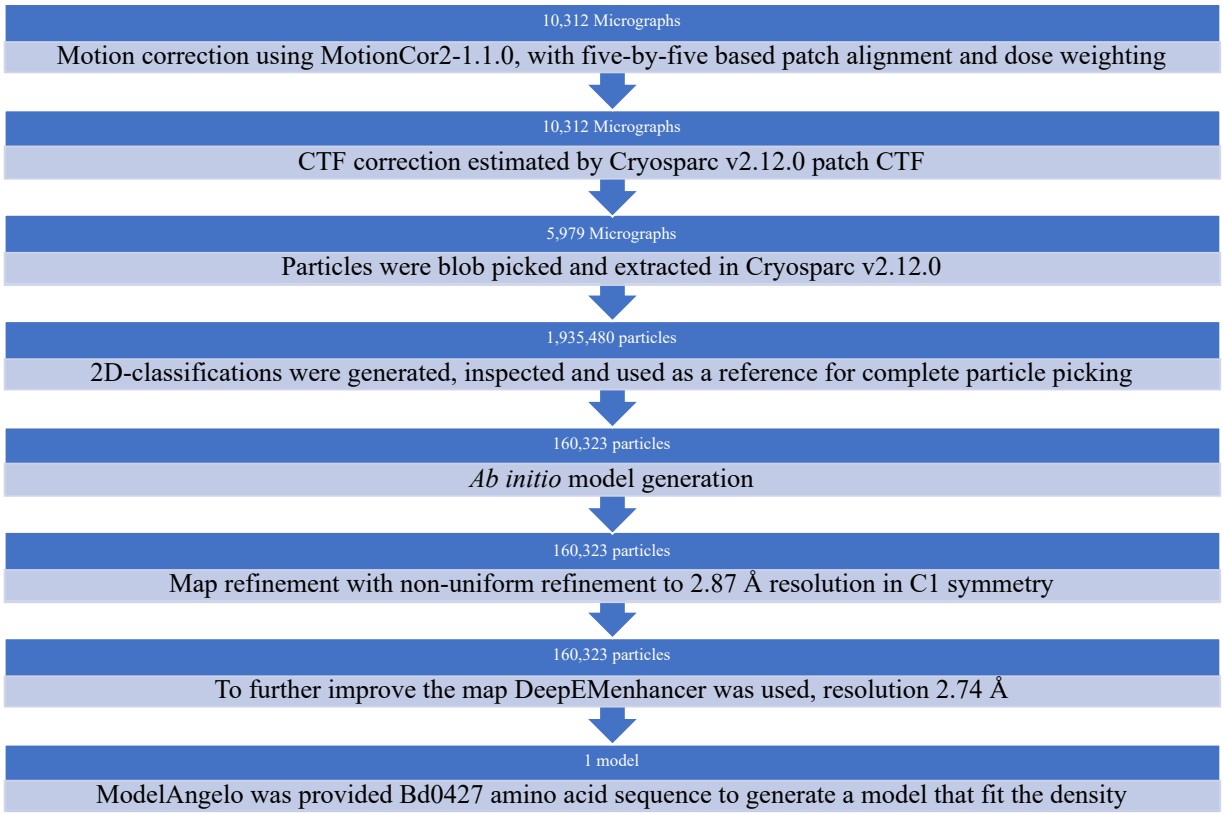


Figure 2-7: Workflow of CryoEM Map Generation

2.8 Liposome Swelling Assay

To make the proteoliposomes 80 µL of liposome stock solution (100 $mg.mL^{-1}$ phosphatidylcholine and 2.3 $mg.mL^{-1}$ dihexadecyl phosphate in chloroform) per sample was dried individually under air flow. The lipid film was left to fully dry under vacuum for over 2 hr. The dried lipids were each resuspended in 100 µL of milliQ water (filter with 0.22 µm filter) before protein (15 µg of OmpF, 15 µg Bd0427, same volume of detergent buffer (50 mM Tris (pH 8), 150 mM NaCl, 0.03 % (w/v) DDM) for the control) was added to each sample immediately followed by a vortex step. The samples were sonicated in water bath for 1.5 min (3 times 30 s cycles) to clarify. The samples were transferred to a 24-well plate to be dried overnight under vacuum. The proteoliposomes were then resuspended in 200 µL buffer (10 mM HEPES (pH 7), 12 mM stachyose), and left to stand for 1-2 hr.

Liposome swelling was monitored following the absorbance at OD400 over a 30 s period with measurements being taken every 1 s. The assay contained 15 µL of liposomes or proteoliposomes and 200 µL of substrate solution (Table 2-14). The assay is repeated for each substrate solution in triplicate.

Table 2-14: Substrate Solutions use in Liposome Swelling Assay

Substrate Solutions
10 mM HEPES (pH 7), 12 mM L-arabinose
10 mM HEPES (pH 7), 15 mM glycine
10 mM HEPES (pH 7), 15 mM alanine
10 mM HEPES (pH 7), 12 mM leucine
10 mM HEPES (pH 7), 8 mM lysine
10 mM HEPES (pH 7), 12 mM D-glucose
10 mM HEPES (pH 7), 15 mM maltose

To analyse the proteoliposomes swelling the rate of swelling is estimated using linear regression of the initial slope. This value has the rate of swelling of the liposome control subtracted from it and then the relative permeability is then compared to OmpF swelling in the presence of L-arabinose as a percentage.

2.9 Single Channel Conductance Experiments (Collaboration)

These experiments used the cuvette in Figure 2-8, which shows two chambers (*cis* and *trans*) separated by a 25 µm thick Teflon septum (Goodfellow Corporation). An ~75 µm aperture in the septum was pre-treated with 1 % (v/v) hexadecane in hexane for 30 min. A planar lipid bilayer membrane was formed across the aperture using 3 µL of 5 $mg.mL^{-1}$ 1,2-diphytanoylphosphocholine (DphPC, Avanti Polar Lipids) in n-pentane added to each chamber containing 1.25 mL of 10 mM HEPES (pH 7) and 1 M KCl. The electrodes of AgCl (World Precision Instraments) were added to each chamber to measure current, one was grounded in the *cis* side and the other was connected to an amplifier (200B Axopatch, Axon Instruments) in the *trans* side. To insert single channels into the lipid bilayer Bd0427 A103His was added at a concentration of 10^{-2} – 10^{-4} fold of 5.6 $mg.mL^{-1}$ protein, in 1 % (v/v) Genapole X100, to the *cis*

side of the cuvette. Voltage (25 –175 mV) was applied to the sample to observe changes in current upon insertion. The current was amplified using an amplifier and digitised using Axon Digidata 1440 digitiser. The data was collected using Clampex and analysed using Clampfit software.

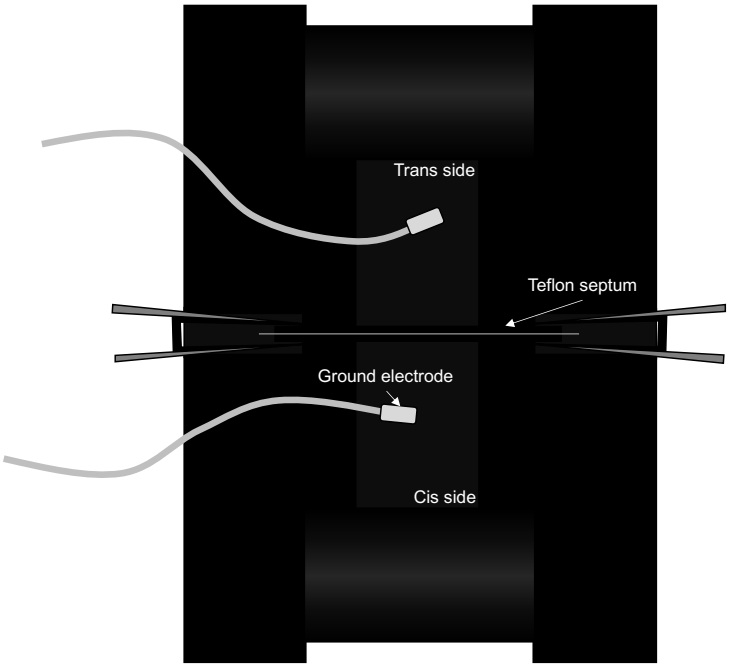


Figure 2-8: Cuvette used in Single Channel Conductance Experiments

2.10 Construction of Gene *bd0427* Deletion Strain

Plasmid constructs were generated using Gibson Assembly, which involves amplifying multiple overlapping DNA fragments for assembly into a vector in a single isothermal step (Gibson et al., 2009). The vector used was pK18*mobsacB*, a suicide vector, which encodes a kanamycin resistance gene (*Tn5 Km^R)* and a levansucrase (*sacB*). The levansucrase results in sucrose sensitivity in both Gram-negatives and Gram-positives due to using sucrose to produce a toxic product (Gay et al., 1983; Schäfer et al., 1994). The plasmid was integrated into the genome using homologous recombination. First the plasmid is conjugated into *B. bacteriovorus* from donor strain *E. coli* S17-1 which works by having a much greater number of *E. coli* cells than *B. bacteriovorus* so that a proportion act as prey and the rest can act as donors, before a

single cross-over event. A selection pressure of 5 % sucrose is applied to the *B. bacteriovorus* cells which results in the formation of a second cross-over event or reversal back to wild-type.

2.10.1 Amplification of DNA Fragments for Gibson Assembly

To generate an in-frame deletion of *bd0427*, upstream and downstream flanking regions of the gene were individually amplified by PCR using the forward and reserve primers named *bd0427*_KO_up and *bd0427*_KO_dw (Table 0-1, Appendix). The PCR reactions (100 µL) contained: 1x Phusion HF buffer, 0.1 mM dNTPs, 0.2 µM forward and reverse primers, 1 µL Phusion DNA polymerase, 1 µL HD100 template DNA and 73 µL H_2O. This procedure utilised the following program: initial denaturation step at 98 ˚C for 2 min, followed by 30 cycles of denaturation at 98 ˚C for 10 s, annealing at a gradient between 56 ˚C and 58 ˚C for 10 s and extension at 72 ˚C for 1 min. This was followed by a final extension step at 72 ˚C for 5 min. To visualise PCR products, 4 µL was ran on a 1 % agarose gel alongside 4 µL of 1 kb ladder (NEB), at 100 V for 60 min, and viewed using UV light box.

2.10.2 Gibson Assembly

The DNA fragments were purified by gel extraction with the QIAquick Gel Extraction Kit (QIAGEN). The DNA fragments were cloned into pK18*mobsacB* (cut with EcoR1 and HindIII), using the NEBuilder Hifi DNA Assembly Kit (BEW), at a ratio of 5:1 and 7:1 (insert:plasmid). The assembly was performed at 50 ˚C for 1 hr.

2.10.3 Construct Verification

To verify the plasmid construction, 2 µL of Gibson Assembly product was transformed into *E. coli* DH5α cells. The cultures were plated onto LB agar plates with 40 $ng.\mu L^{-1}$ X-Gal and 40 $ng.\mu L^{-1}$ IPTG, to select for white colonies. The white colonies were screened for insert using colony PCR, as described in section 2.2.2. Colonies containing insert of 2 kb were grown in LB

media overnight for plasmid preparation using QIAprep Miniprep Kit (QIAGEN). The plasmid insert was verified using Sanger sequencing (Source Biosciences).

2.10.4 Conjugation into *B. bacteriovorus* HD100 (Collaboration)

To conjugate the plasmid into *B. bacteriovorus* HD100, the plasmid was transformed into donor strain *E. coli* S17-1 before incubation with *B. bacteriovorus.* For the conjugation, 10 mL of donor strain overnight culture and 10 mL of wild-type (WT) *B. bacteriovorus* culture (10 mL CaHEPES (295 $mg.mL^{-1}$ $CaCl_2$:$2H_2O$ and 5.94 $g.L^{-1}$ HEPES (pH 7.6)), 600 µL of *E. coli* S17-1 prey, 200 µL of WT *B. bacteriovorus* HD100 culture) were incubated overnight at 37 ˚C and 29 ˚C, respectively, shaking at 180 r.p.m.. The cells were pelleted at 5,525 g for 20 min (*B. bacteriovorus*) or 5 min (*E. coli*) and resuspended in 100 µL CaHEPES. The resuspended *B. bacteriovorus* cells were applied to a Hybond-N nylon membrane on the surface of a PY agar plate (10 $g.L^{-1}$ peptone (Broadbean), 3 $g.L^{-1}$ yeast extract (Difco), 10 g.L-1 select agar (Sigma), pH 6.8) and left to dry. The resuspended donor cells were then added to the membrane and the plate was incubated overnight at 29 ˚C. The cells were removed from the nylon membrane by resuspension in 1 mL CaHEPES. The *B. bacteriovorus* cells plus 150 µL kanamycin resistance prey (*E. coli* S17-1 (pZMR100)) were added to YPSC overlay plates (0.25 $g.L^{-1}$ $MgSO_4$, 0.5 $g.L^{-1}$ NaOAc, 1 $g.L^{-1}$ peptone (broadbean), 1 $g.L^{-1}$ yeast extract, 0.25 $g.L^{-1}$ $CaCl_2.2H_2O$, pH 7.6, containing 10 $g.L^{-1}$ (bottom) or 6 $g.L^{-1}$ (top) agar (Sigma)) with 50 $\mu g.mL^{-1}$ kanamycin in the bottom only incubated at 29 ˚C for 4-5 days. After this period host-dependent merodiploid plaques formed and these were picked and added into 2 mL prey cultures (2 mL CaHEPES, 150 µL of *E. coli* S17-1 prey (pZMR100)) containing 50 $\mu g.mL^{-1}$ kanamycin. The *B. bacteriovorus* HD100 prey cultures were left shaking at 29 ˚C until fully lysed.

2.10.5 Generation and Verification of Ex-conjugant Strains (Collaboration)

To generate ex-conjugant strains, 50 µL of *B. bacteriovorus* HD100 prey lysate culture was aliquoted into 2 mL of sucrose prey culture (2 mL CaHEPES, 5 % sucrose and 150 µL of heat-killed *E. coli* S17-1 prey) left shaking at 29 ˚C overnight. 50 µL of this culture was then added to 2 mL prey culture (2 mL CaHEPES, 150 µL of *E. coli* S17-1 prey), left shaking at 29 ˚C overnight. These two steps were repeated until the *B. bacteriovorus* cells had five successful rounds of prey lysis in sucrose. Serial dilutions of *B. bacteriovorus* were cultured on YPSC overlay plates containing *E. coli* S17-1 prey at 29 ˚C. After 7 days the resulting plaques were picked into 96-well plates containing 200 µL CaHEPES and 15 µL *E. coli* S17-1, left overnight at 29 ˚C to recover. Two parallel 96-well plates were then set up using 20 µL from each well of the previous plate. The first plate contained 200 µL CaHEPES and 15 µL *E. coli* S17-1 and the second plate contained 200 µL CaHEPES, 15 µL *E. coli* S17-1 (pZMR100) and 50 µg.mL^{-1} kanamycin. Both plates were left to incubate at 29 ˚C overnight to select for kanamycin-sensitive clones (gene deletion strains or WT strains). The kanamycin-sensitive strains were screened initially using PCR on the cultures. The cultures that indicated gene deletion, were grown in 10 mL prey lysate cultures, shaking at 29 ˚C overnight before genomic DNA extraction using GenEluteTM Bacterial Genomic DNA Kits (Sigma). PCR was performed using the clone's genomic DNA, using Phusion DNA polymerase following the NEB protocol and primers (bd0427_KO_S_F and bd0427_KO_S_R) designed to anneal approximately 100 bp away from the homologous recombination site. The PCR product was purified using gel extraction with the GenElute Gel Extraction Kit (Sigma) and sequenced to confirm gene deletion construct.

CHAPTER THREE: OBTAINING AND OPTIMISING DIFFRACTION-QUALITY CRYSTALS OF BD0427

3.1 Introduction

As with many Gram-negative bacteria *B. bacteriovorus* has a characteristic cell envelope that consists of three distinct layers: the inner membrane (IM), the peptidoglycan containing periplasm and the outer membrane (OM) (Figure 3-1).

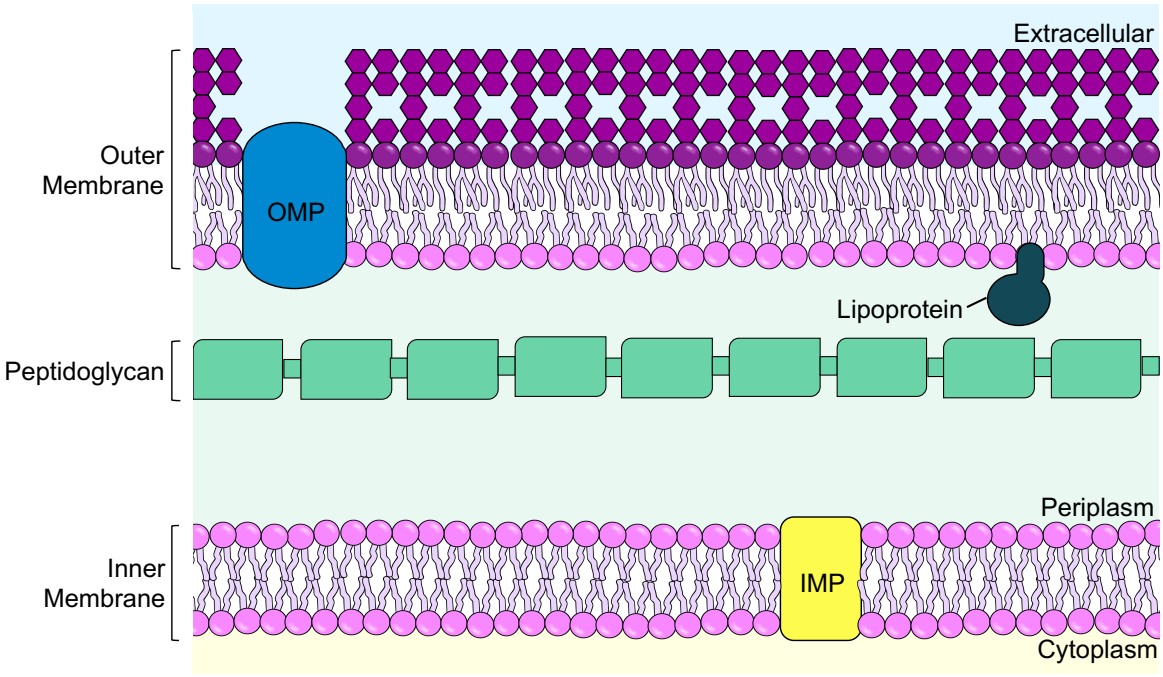


Figure 3-1: The Gram-negative Cell Envelope. This consists of three layers the inner membrane (IM), periplasm, and the outer membrane (OM). The IM is made up of phospholipids (pink), inner membrane proteins (yellow) and lipoproteins. The periplasm sits between the two membranes and is also home to the peptidoglycan layer. The OM is made up of phospholipids on the inner leaflet and lipopolysaccharides (purple) on the outer leaflet, as well as outer membrane proteins (blue) and lipoproteins.

The IM separates the cytoplasm from the periplasm and is a symmetrical phospholipid bilayer that contains inner membrane proteins (IMPs) and lipoproteins. IMPs generally have hydrophobic α-helices that span the membrane and lipoproteins are anchored to the membrane via an N-terminal fatty acid tail via a cysteine residue (Dalbey et al., 2011). As a result of direct contact with the cytoplasm the inner membrane is an energy-rich environment driven by proton

motive force and ATP. Enclosed by both membranes is the periplasmic space, an oxidative environment containing the structurally important peptidoglycan and soluble proteins. The OM, unlike the inner membrane, is an asymmetric bilayer made up of phospholipids on the inner leaflet (periplasmic side) and lipopolysaccharides (LPS), consisting of lipid A, the core oligosaccharide and distal O-antigen (Figure 3-2) on the outer leaflet (extracellular side) (Bertani and Ruiz, 2018). Within the OM are outer membrane proteins (OMPs) and as with the inner membrane, lipoproteins. OMPs found within this membrane are generally found to consist of an antiparallel β-barrel, with the external face in the lipid tail environment being mostly

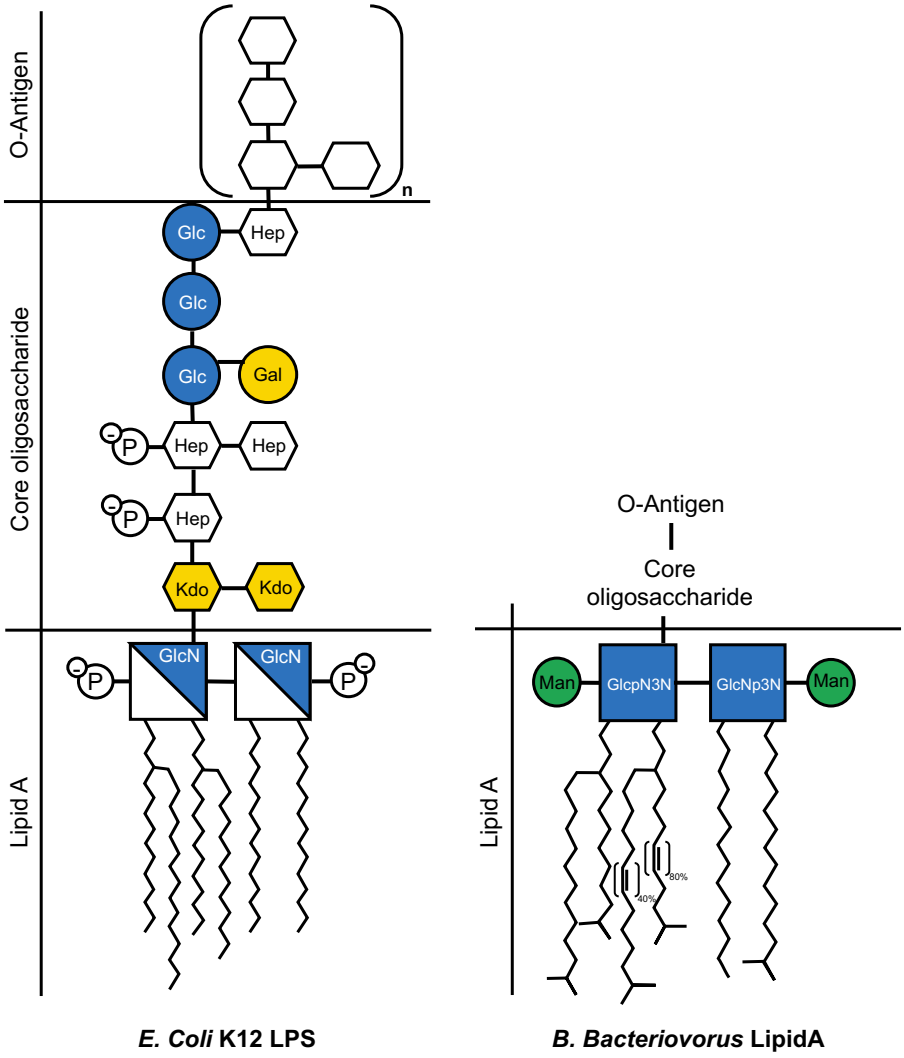


Figure 3-2: The LPS Structure from *E. coli* K12 alongside Lipid A from *B. bacteriovorus*. No information on the core oligosaccharide is available for *B. bacteriovorus*

sidechains that are hydrophobic, and the sidechains exposed in the solvent generally being hydrophillic.

It is the OM that gives Gram-negatives an advantage, because the LPS in the outer leaflet of the bacteria's cell wall is impermeable to large, hydrophilic and hydrophobic molecules, thus protecting the bacterium from harmful substances (environmental dangers, toxins from micro-competitors, drugs) (Bertani and Ruiz, 2018). LPS, as mentioned above is made up of three components: lipid A, the core oligosaccharide and distal O-antigens. In *E. coli* the lipid A is negatively charged due to the presence of phosphate groups which bind divalent cations (Ca^{2+} and Mg^{2+}) and this, along with fatty acid tails that are mostly saturated, enables a stable, tightly packed LPS layer (Bertani and Ruiz, 2018; Caroff and Karibian, 2003; Kotra et al., 1999).

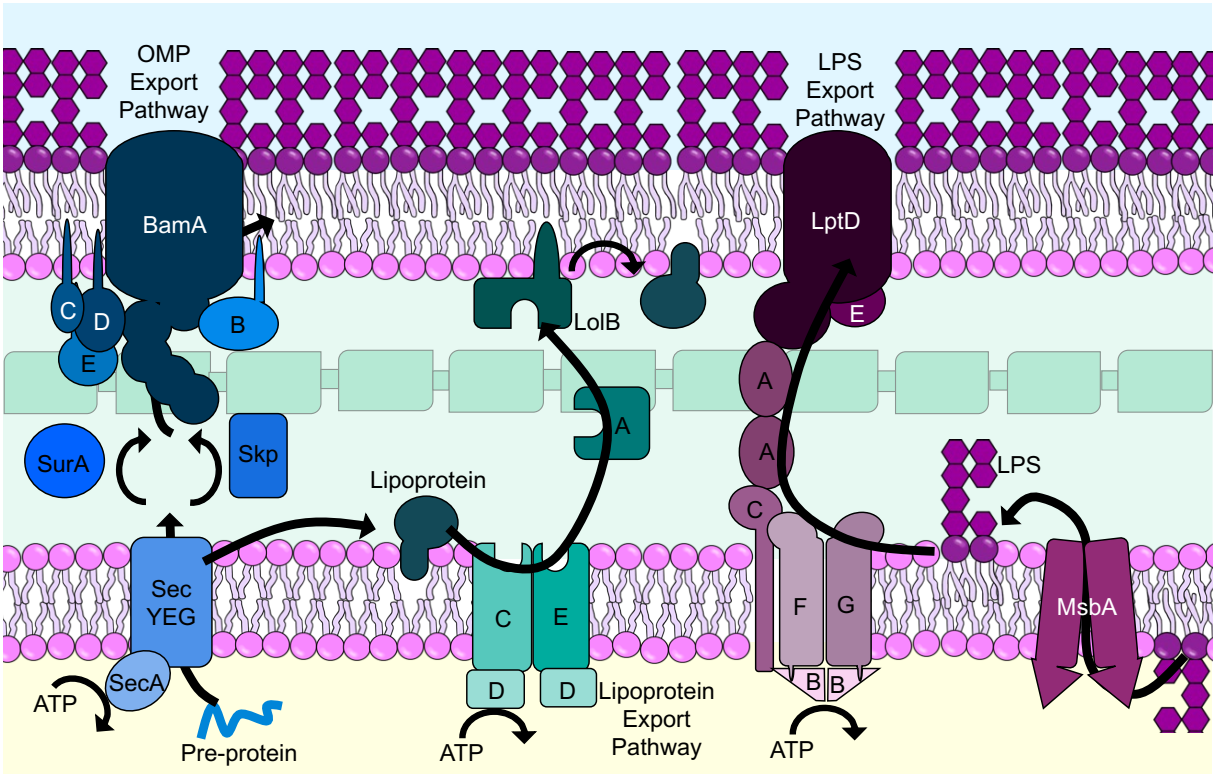


Figure 3-3: The Pathways for Outer Membrane Biogenesis. The pathways for transport of OMPs (blue), Lipoproteins (green) and LPS (purple) are all well characterised. OMPs are transported across the IM by Sec, the periplasm by SurA and Skp where they are inserted into the OM by Bam. The Lol pathway is responsible for the transport and insertion of lipoproteins into the OM. LPS flipped from the leaflet on the cytoplasmic side of the IM to the periplasmic side by MsbA, LPS is then transported to the OM by the Lpt pathway.

However, LPS is very diverse across Gram-negatives (Nikaido, 2003) and in *B. bacteriovorus* the lipid A of its LPS is neutral, due to having a mannose instead of phosphate group (Figure 3-2) and the fatty acid tails are also more saturated (Schwudke et al., 2003). These differences are potentially the reason the *B. bacteriovorus* membrane is more flexible (Borgnia et al., 2008).

To maintain the OM, pathways for trafficking and transporting the individual components that makeup the membrane exist. Those that have been well characterised can be seen in Figure 3-3, although there is still some information lacking regarding the transport of lipids to the OM.

OMPs are used to enable transport of large and/or hydrophilic molecules across the OM, there are several different types of outer membrane protein (OMP) classes (Table 3-1), but despite this most have been observed to form β-barrels. This can be appreciated when looking at the

Table 3-1: Outer Membrane Protein Classes

Class	General Function	Example
Porins	Enable passive diffusion of small hydrophilic molecules <600 Da, nonspecific (Nikaido, 2003)	OmpF, OmpC
Specific channels	Catalyse the diffusion of a specific class of molecules (Nikaido, 2003)	LamB, ScrY
TonB dependent receptors	Uses the energy from the cytoplasm through TonB to transport large molecules such as vitamins (Nikaido, 2003)	BtuB, FecA
Double-membrane-spanning secretion systems	Export proteins with specific C-terminal signal, from the cytoplasm to the extracellular domain. Consists of Type I secretion system and resistance-nodulation-division (Costa et al., 2015; Delepelaire, 2004)	HlyA(TolC), CvaC(TolC)
Type II secretion systems	Sec-dependent export proteins from the periplasm to the extracellular domain, substrates include hydrolysing enzymes and toxins (Costa et al., 2015)	GspD
Type III secretion systems	Found in pathogenic bacteria. These double-membrane spanning proteins export effector proteins into the target cell's cytoplasm (Costa et al., 2015)	EspA
Type IV secretion system	Export of proteins and DNA into target cells and pilus biogenesis (Costa et al., 2015)	VirB7-9

Type V secretion systems	Sec-dependent export of effector proteins using either one or two proteins, this system includes five subtypes (a-e) (Costa et al., 2015; Meuskens et al., 2019; Nikaido, 2003)	IgA, TspA and TspB, YadA, PlpD,
Type VI secretion system	Exports toxic proteins into target cells (Costa et al., 2015)	Tss
Chaperone-ushers pathway	Export of appendages such as pili and fimbriae involved in cell recognition and attachment (Costa et al., 2015; Nikaido, 2003)	Pap, Fim
Outer membrane polysaccharide export proteins	Export of extracellular polysaccharides or capsular polysaccharides (Cuthbertson et al., 2009; Sande et al., 2019)	Wzx and Wzy, KpsD

deposited structures of OMPs in the PDB, and a representation of all deposited OMP structures from *E. coli* is provided in Figure 3-4. The major OMPs in Gram-negatives tend to be porins or specific channels (Nikaido, 2003).

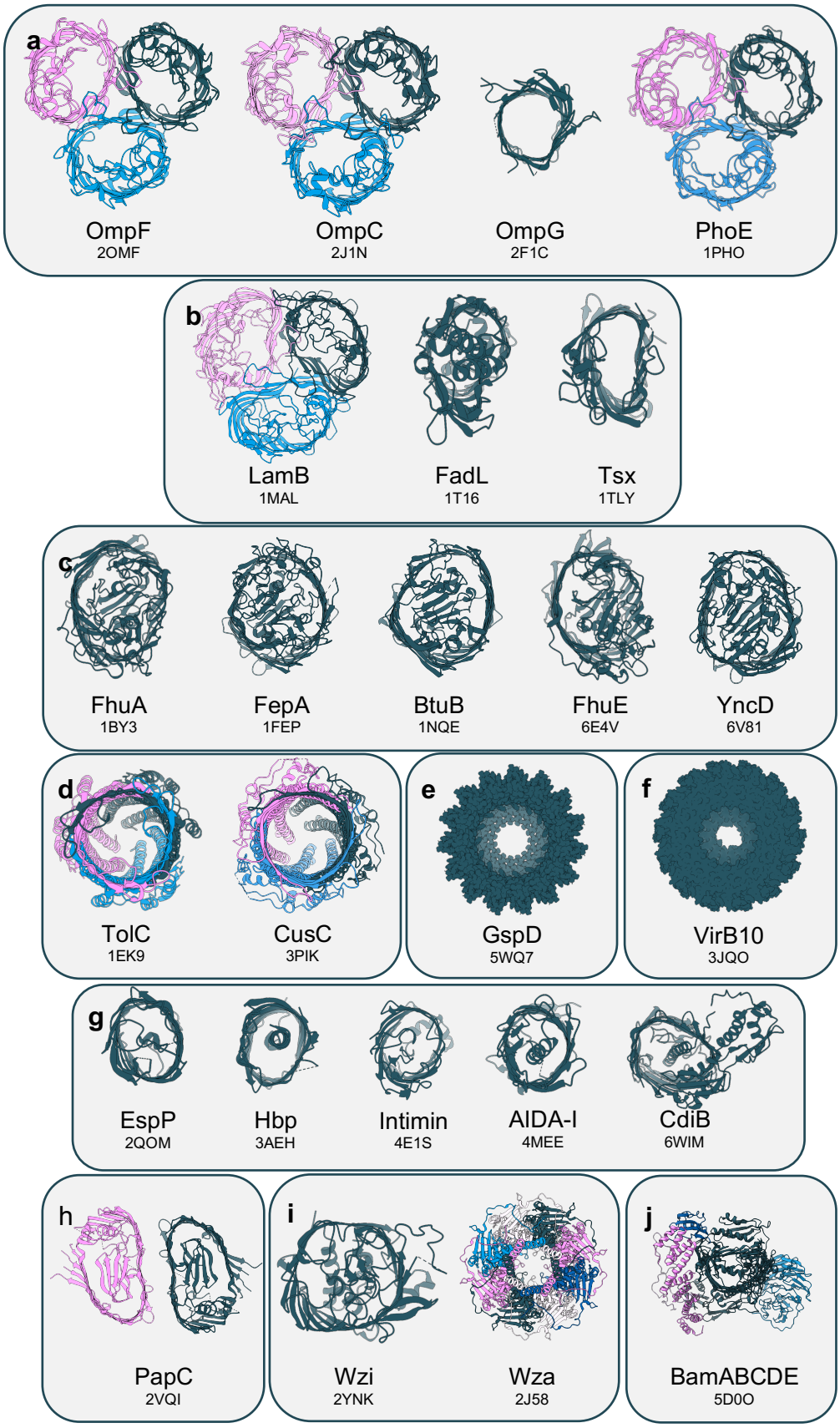


Figure 3-4: Published Structures of *E. coli* Outer Membrane Proteins from each Class. This includes general porins (a), specific channels (b), TonB-dependent receptors (c), double-membrane spanning secretion systems (d), type II secretion systems (e), type IV secretion systems (f), type V secretion systems (g), chaperone-usher pathway (h), outer membrane polysaccharide export pathway (i), other (j). There are currently no structures from *E. coli* of type III or type VI secretion systems, and there are several proteins that have not been classified.

3.1.1 Outer Membrane Proteins Biogenesis

OMPs are synthesised in the cytoplasm and transported to the OM where they function. This is a multi-step processes involving many proteins, some of which are still not fully understood.

3.1.1.2 Crossing the Inner Membrane

To signal for this transport, OMPs are synthesised with a cleavable N-terminal signal sequence that enables trafficking of the protein to the SecYEG complex in the IM for translocation to the periplasm. To transfer OMPs to the SecYEG complex from the ribosome two pathways have been proposed: the post-translational and co-translational pathways (Figure 3-5).

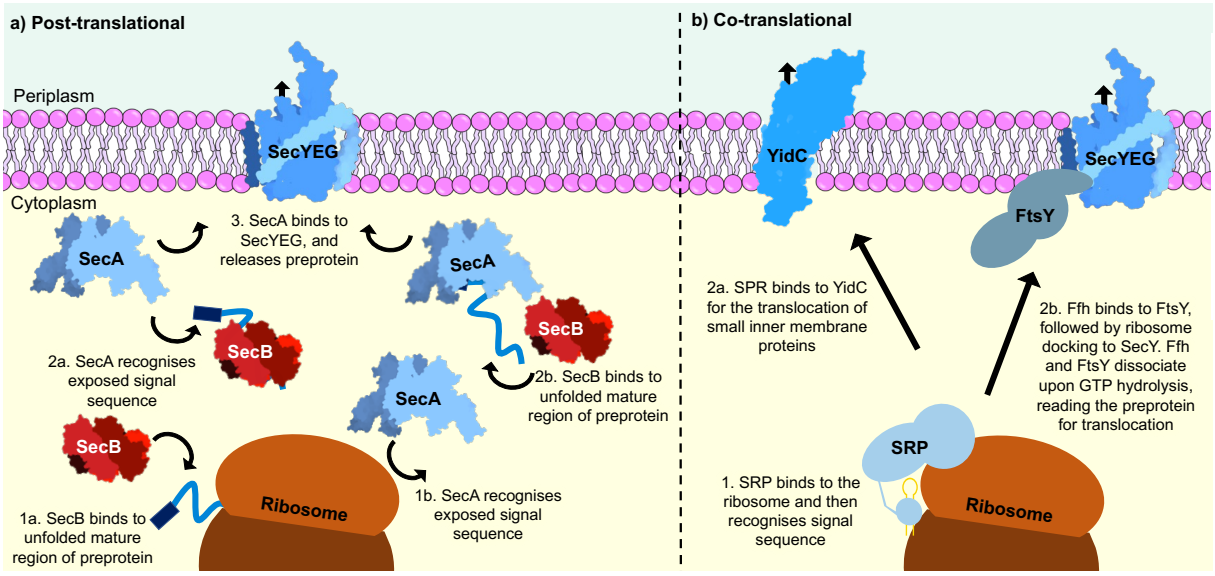


Figure 3-5: Post-Translational and Co-Translational Protein Transport Across the Inner Membrane. In the post translational pathway, the precursor protein is bound by SecA and SecB before transport to the SecYEG translocator. In the co-translational pathway, the signal recognition particle (SRP) binds to the ribosome where it recognises signal sequences, this leads to trafficking to the inner membrane. The complex is then bound to either YidC or SecYEG via FtsY for translocation across the membrane. PDBs used: 1QYN, 2FSF, 6R7L, 6AL2.

The post translation pathway is where transport across the inner membrane occurs after protein translation. This pathway involves SecB, a tetrameric molecular chaperone formed a dimers of dimers that binds to the mature region of precursor proteins to prevent aggregation, and SecA, an ATPase motor that is found in the cytoplasm or associated with the IM via lipid interactions

bound to SecYEG complex (Collier et al., 1988; Dekker et al., 2003; Lecker et al., 1990). In the post-translation pathway the precursor protein can either be first bound by SecB, which leaves the N-terminal signal sequence exposed for SecA to recognise and bind (Breukink et al., 1995; Dekker et al., 2003), or the binding of SecA first to the ribosome close to the location of polypeptide release (Huber et al., 2011) where SecA binds the N-terminal signal sequence; followed by the binding of SecB to SecA and the precursor proteins mature sequence and released from the ribosome (Huber et al., 2017). The binding of SecA to SecB through its C-terminus (Dekker et al., 2003). The SecAB and precursor complex then binds to SecYEG ready for translocation across the membrane (Hartl et al., 1990; Suo et al., 2015).

The co-translational pathway is where protein transport across the IM is coupled to translation and is mostly used for inner membrane proteins (Oswald et al., 2021; Saraogi and Shan, 2014; Valent, 1998). In this pathway a signal recognition particle (Ffh protein bound to 4.5S RNA) binds to the ribosome, where it recognises exposed N-terminal signal sequences (Elvekrog and Walter, 2015; Valent, 1998). Upon recognition the ribosome-SRP complex is trafficked to the inner membrane where it can bind directly to YidC insertase for insertion of small inner membrane proteins (Steudle et al., 2021) or the GTPase domain of Ffh can bind, in a GTP-dependent manner, to GTPase domain of FtsY (an SRP receptor) (Samuelson et al., 2000; Saraogi and Shan, 2014; Serek et al., 2004; Shanmugam and Dalbey, 2019; Steinberg et al., 2018). FtsY is already bound to SecYEG, and once Ffh binds movements by FtsY enable docking of the ribosome to SecY's ribosome binding site. GTP hydrolysis then allows the dissociation of Ffh and FtsY, readying SecYEG for translocation (Oswald et al., 2021).

Trafficking of OMPs to the IM mainly occurs via Sec-dependent pathway, therefore the discussion surrounding translocation will focus on the steps following this pathway.

SecYEG forms a claw-like channel in the IM with a central constriction point, known as the pore ring, and a periplasmic plug which prevents the translocation of proteins (Berg et al., 2004). Upon binding of SecA through exposed cytosolic loops from SecY (Berg et al., 2004; Mori and Ito, 2006), conformational changes occur resulting in widening of the pore ring (Zimmer et al., 2008). This widening allows the insertion of the two-helix finger (2HF) of SecA and the precursor protein as a loop, causing displacement of the periplasmic plug (Li et al., 2016; Zimmer et al., 2008). SecA then mediates translocation through the SecY channel, through ATP hydrolysis. Currently the exact mechanism as to how this occurs is not known but two models have been proposed: the push and slide model and the Brownian-ratchet model. The push and slide model proposes that the 2HF pushes the precursor protein through the channel via its tip, with the 2HF reset by ATP hydrolysis (Bauer et al., 2014). During this ATP hydrolysis a conformation change in SecA results in '*clamping*' action on the precursor protein preventing moving in a backwards direction which upon release of the products (ADP and phosphate), the clamp is released ready for pushing with the reset 2HF (Catipovic et al., 2019; Catipovic and Rapoport, 2020). The second model is known as the Brownian-ratchet model, where ATP-hydrolysis biases the random diffusion of the protein through the SecY channel towards the periplasm (Allen et al., 2020, 2016).

SecYEG has also been shown to form a complex with SecDFYajC and YidC known collectively as the holo-translocon where these additional proteins in the complex aid translocation of the precursor protein (Botte et al., 2016; Oswald et al., 2021; Troman and Collinson, 2021), though more investigation is required to fully understand this complex.

3.1.2.2 Traversing the Periplasmic Space

Following translocation of a precursor protein across the IM the signal sequence is cleaved by signal peptidase I ready for transport to the OM (Paetzel et al., 2002). Two proteins have been

identified that could enable transport across the periplasm: survival protein A (SurA) and seventeen kilodalton protein (Skp) (Figure 3-6a).

SurA was first identified as a gene that when mutated effected the survival of *E. coli* cells in stationary phase (Tormo et al., 1990) and is the primary periplasmic chaperone of OMPs (Sklar et al., 2007) (Behrens, 2001). The crystal structure of SurA shows the protein forms two domains: the 'core domain', made up of the N-terminus, C-terminus and a 100-residue peptidyl-prolyl isomerase (PPIase) domain, and the 'P2 domain', a second 100-residue PPIase domain (Bitto and McKay, 2002). The core domain is responsible for binding of the precursor protein upon release from SecYEG and a preference for binding aromatic residues, prevalent in OMPs, has been shown suggesting a possible selective mechanism (Hennecke et al., 2005; Xu et al., 2007). SurA has been shown to interact with BamA, part of the β-barrel assembly machinery see below, but the exact mechanism for transferring the protein from SurA to the OM remains unknown (Sklar et al., 2007).

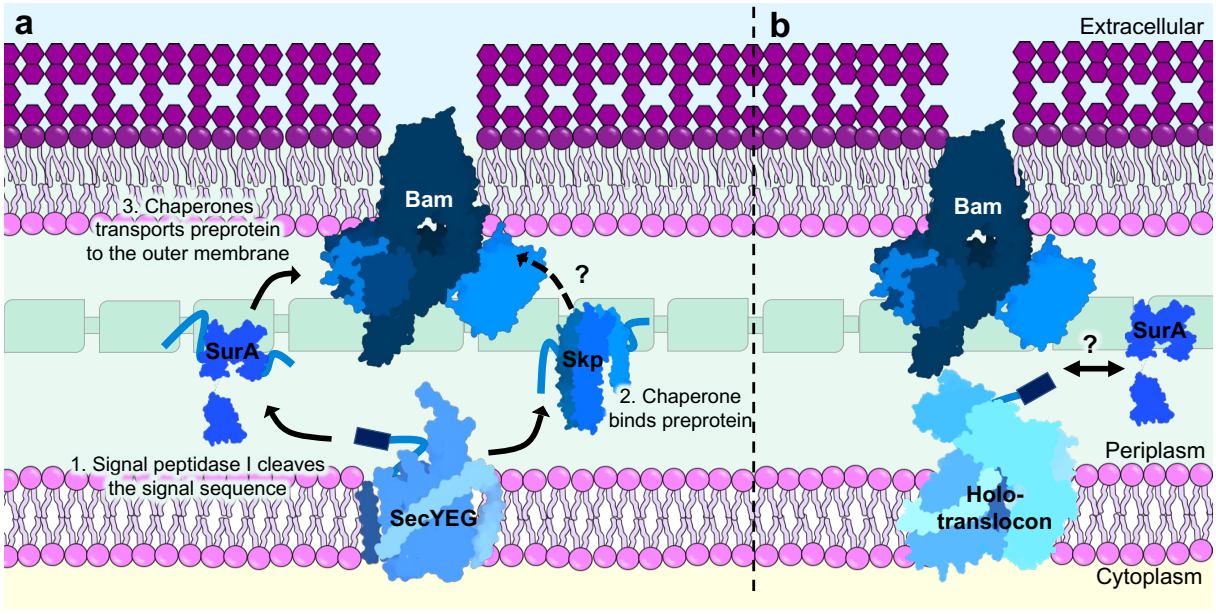


Figure 3-6: Transport of Outer Membrane Proteins Across the Periplasm. a) Hypothesis for transport of OMPs across the periplasm using chaperones. b) Newly proposed hypothesis where SecDF and YidC periplasmic regions interact with BAM, the protein involved in OMP insertion into the OM. This connects the IM and OM proteins involved in translocation and insertion, providing a potential route for transport across the periplasm which may be aided by chaperones. PDBs used: 6R7L, 1M5Y, 1SG2, 5AYW and 5MG3 (PDBs of SecYEG, SecDF and YidC fitted to EM density).

Skp is currently thought to be a periplasmic chaperone as the deletion of Skp results in a reduction of OMPs in the OM (Chen and Henning, 1996; Schäfer et al., 1999). The protein forms a trimeric jelly-fish-like structure where the cavity formed by the α-helical 'legs' have been shown to bind unfolded OMPs (Holdbrook et al., 2017; Korndörfer et al., 2004; Walton et al., 2009; Walton and Sousa, 2004). As with SurA, it is not clear how Skp deposits OMPs to the OM but Skp has been shown to fold OMPs into a membrane *in vitro* (Bulieris et al., 2003; Patel et al., 2009).

There is also emerging evidence that suggests there is a direct connection between the proteins involved in translocation across the IM and the proteins involved in inserting OMPs into the OM (Oswald et al., 2021; Troman et al., 2023; Troman and Collinson, 2021). Direct interactions have been shown between periplasmic domains of SecDF and YidC with the periplasmic domain of the proteins involved in OMP insertion (Alvira et al., 2020), suggesting that this may aid to transport of the precursor protein to the OMP (Figure 3-6b). Currently it is proposed that this would work with the previously discussed periplasmic chaperones (Oswald et al., 2021; Troman et al., 2023; Troman and Collinson, 2021). As this is still a recent discovery more work needs to be done to fully understand the implications of these discoveries in terms of OMP transport across the periplasm.

3.1.2.3 Inserting into the Outer Membrane

Upon delivery of OMPs to the OM, the β-barrel assembly machinery (BAM) aids insertion of OMPs into the OM. This complex varies by species but in *E. coli* is made up of five proteins, BamA-E (Wu et al., 2005). BamA consists of a C-terminal β-barrel domain and multiple periplasmic polypeptide transport associated (POTRA) domains at its N-terminus (Kim et al., 2007; Ni et al., 2014). The remainder of the complex (BamB-E) are all lipoproteins that are

anchored to the inner leaf of the OM (Wu et al., 2005). BamB and BamD interact with the POTRA domains of BamA, whereas BamC and BamE interact with BamD (Gu et al., 2016).

The exact mechanism for how the BAM complex aids insertion of folded OMPs into the membrane is still not well understood. However, recent studies with modified EspP, an autotransporter, have shed some interesting new insights. These studies show that the β1 strand of BamA interacts with the C-terminus of the incoming OMP (iOMP) (Figure 3-7). More specifically, there is a hydrophobic pocket in which the C-terminal phenylalanine from the iOMP interacts with BamA (Doyle et al., 2022; Wu et al., 2021). This interaction is believed to act as a template to enable β-strands to form from the C-terminus of the iOMP (Doyle et al., 2022; Horne and Radford, 2022). EM-structures by Doyle and colleagues (2022) showed that the iOMP enters the membrane as a β-sheet, with the unpaired N-terminal β-strand far away from BamA (Figure 3-7). Interestingly, upon looking at the membrane plane they noticed that the β-sheet caused a deflection and have suggested that the tension caused by this could be used to energise coiling of the β-sheet into a barrel, causing a B-shaped double barrel with BamA (Doyle et al., 2022). The barrel is then closed due to preferential binding of the C-terminus β-strand of the iOMP to the N-terminal β-strand of the iOMP, compared to C-terminus β-strand of iOMP to BamA β1 (Doyle et al., 2022). Though these papers shed new insights into how BAM could aid folding of OMPs into the OM, as they have focused only upon one

autotransporter, it would be important to see if this same/similar mechanism could be observed of a different OMP.

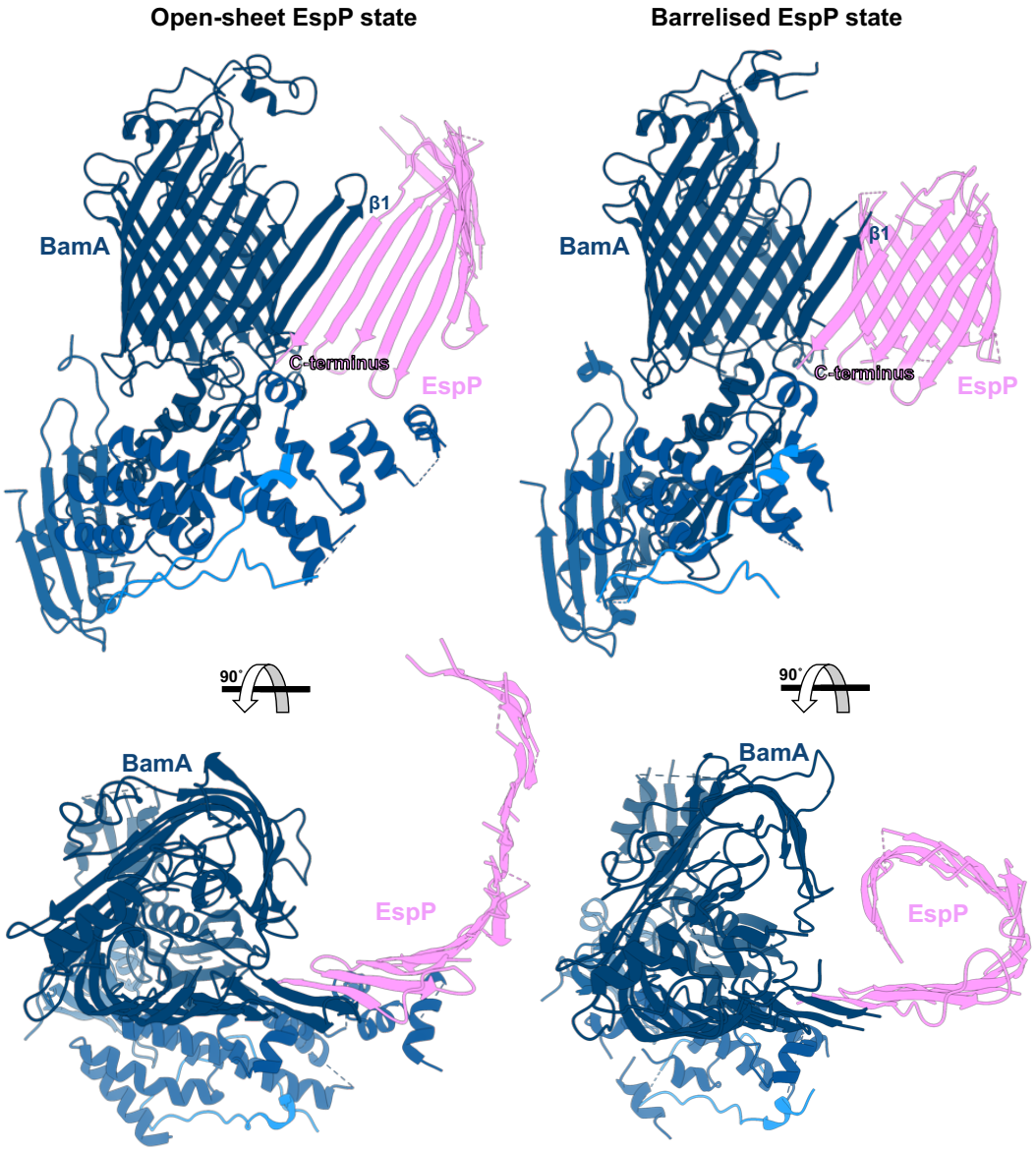


Figure 3-7: Structures of Bam Inserting EspP into the OM. EspP show in open-sheet and barrelised states. PDBs used: 7TT5 and 7TT7.

3.1.2 Porins and Specific Channels

Porins are defined as proteins that form nonspecific diffusion channels, whereas specific channels are defined as proteins that form channels that catalyse the diffusion of a specific class

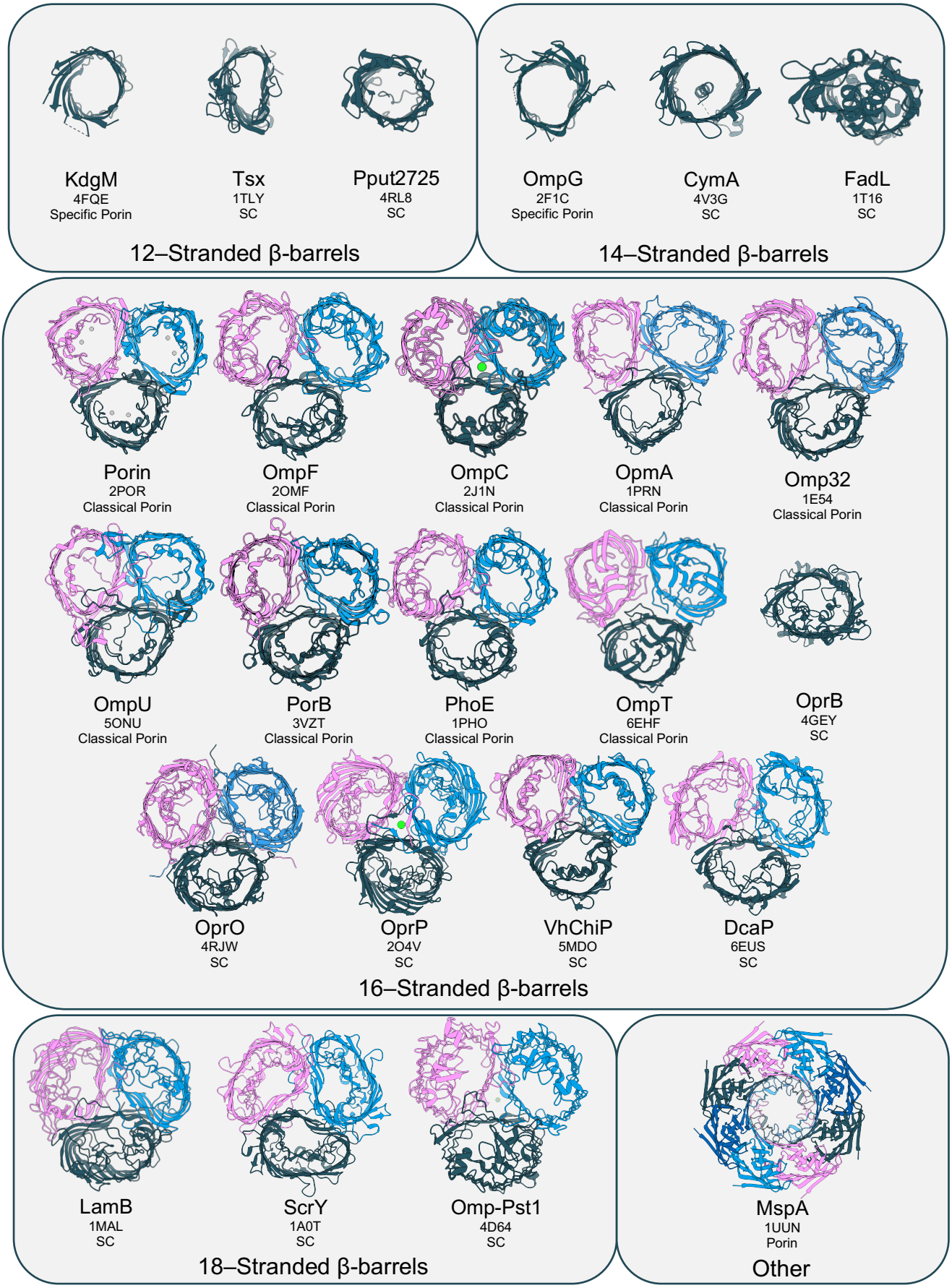


Figure 3-8: Porin and Specific Channel Structures. Porins are classified into classical porins, specific porins and porins. Specific Channels are denoted by SC. All chains are coloured differently with the exception of MspA, which contains 8 chains.

of nutrients (Nikaido, 2003). They are important for the influx of nutrients from the environment into the bacterium. Porins tend to transport molecules below 600 daltons in size, with some observed preferences for either anions or cations. In contrast, specific channels have been demonstrated to transport numerous classes of nutrients including sugars, nucleotides and polyphosphates (Modi et al., 2015; Schirmer et al., 1995; Ye and van den Berg, 2004). Historically, OM channels were classified based on the predicted structure and observed oligomeric states, with 16-stranded β-barrels predicted to be porins and 18-stranded β-barrels predicted to be specific channels (Galdiero et al., 2012). Though in recent years there have been exceptions to these rules: OmpG, CymA, FadL and Tsx to name a few (Subbarao and van den Berg, 2006; van den Berg et al., 2015, 2004; Ye and van den Berg, 2004), making it more difficult to classify OM channels into these two categories. Despite this there remains major structural similarities within these classes.

3.1.2.1 Classical Porin Structure

Classical porins form homotrimers of 16-stranded antiparallel oval β-barrels (Figure 3-8) where the β-strands are tilted with respect to the barrel-axis between 35-50˚, which enables varying barrel size (Nikaido, 2003; Schulz, 2002). The β-strands are connected by loops, which are generally longer on the extracellular side and β-turns on the periplasmic side, which are shorter (Galdiero et al., 2012). Loop 3 is usually the longest loop and often folds back into the barrel causing a narrowing in the channel, known as either the ‘eyelet’ or ‘constriction point’ (Figure 3-9). The constriction point defines the size of molecule that can pass through the channel. The wider the channel, the larger the molecules that can pass through. Loop 2 has a major role in stabilising the oligomer, by extending into the groove of the adjacent protomer (Figure 3-9), which is supported by loops 1 and 4.

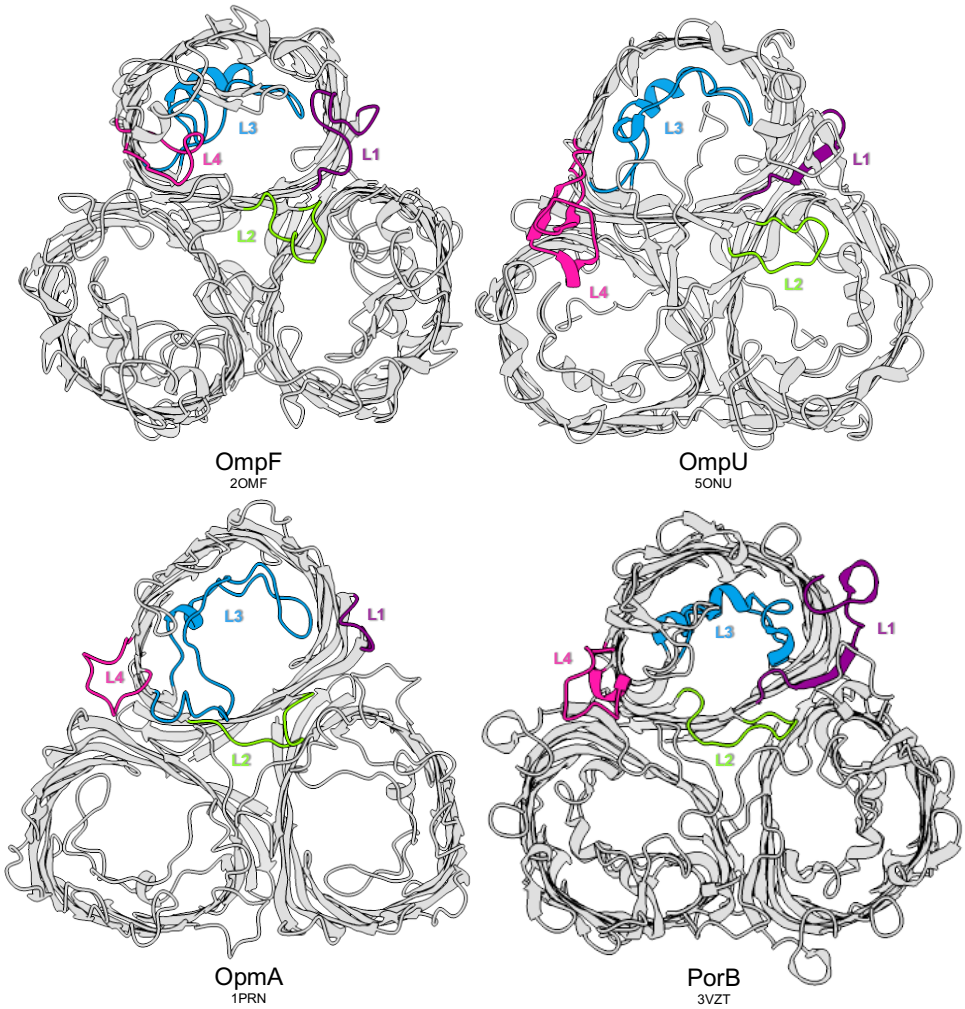


Figure 3-9: Ribbon Diagrams of Four Porins Showing Loop Organisation. Loop 1 in purple, loop 2 in green, loop 3 in blue and loop 4 in pink.

On the external surface of the barrel, there is a band of residues, approximately 25 Å, that are hydrophobic with high Kyte-Doolittle values, for suitability in the bilayer (Cowan et al., 1992; Kyte and Doolittle, 1982). When looking at this in more detail, the barrel contains two rings of aromatic residues at the extracellular side and the periplasmic side (Figure 3-10), known as aromatic girdles (Cowan et al., 1992; Schulz, 2002). These stabilise the OMP in the membrane (Galdiero et al., 2012; Schulz, 2002). The residues lining the channel tend to be polar, and usually indicate what type of ions will be transported by the channel, with Omp32, PhoE and PorB transporting anions and OmpT and OmpU transporting cations (Cowan et al., 1992; Pathania et al., 2018; Tanabe et al., 2010; Zeth et al., 2000).

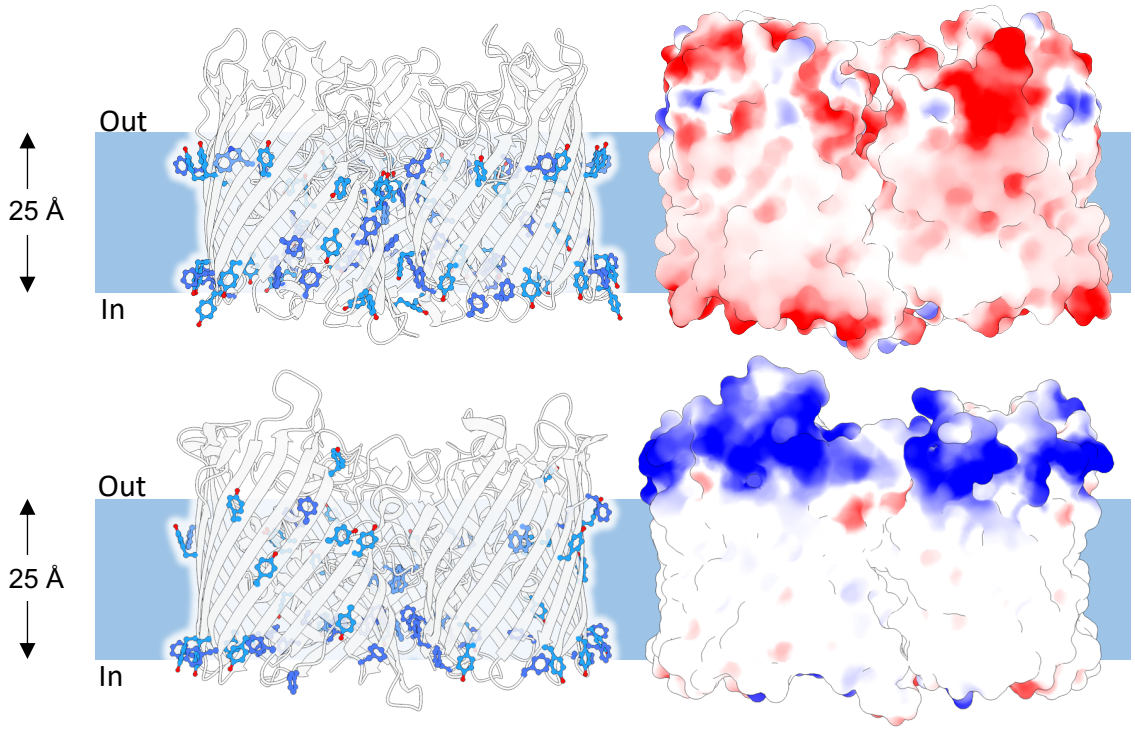


Figure 3-10: Aromatic Residues and Electrostatic Surface of OmpF and PorB. Top membrane contains OmpF, 2OMF, models. Bottom membrane contains PorB, 3VZT, models.

3.1.2.2 Specific and Non-Cassical Porin Structures

Despite classical porins being well defined, KdgM, OmpG and MspA, do not conform to the above description of classical porins. KdgM and OmpG are involved in the transport of a specific class of nutrient, oligogalacturonides and oligosaccharides, respectively, but do not have specific binding pockets for the substrates to catalyse the diffusion differentiating the proteins from specific channels (Blot et al., 2002; Subbarao and van den Berg, 2006). MspA is from *Mycobacterium smegmatis*, which has a unique (to Mycobacteria) cell envelope (Faller et al., 2004).

KdgM and OmpG are both monomeric proteins formed from 12- and 14-stranded β-barrels Figure 3-11). They both have a hydrophobic band surrounded by two aromatic girdles and have longer loops on the extracellular side and shorter loops (turns) on the periplasmic side (Hutter et al., 2014; Subbarao and van den Berg, 2006), the same as in classical porins. Though the

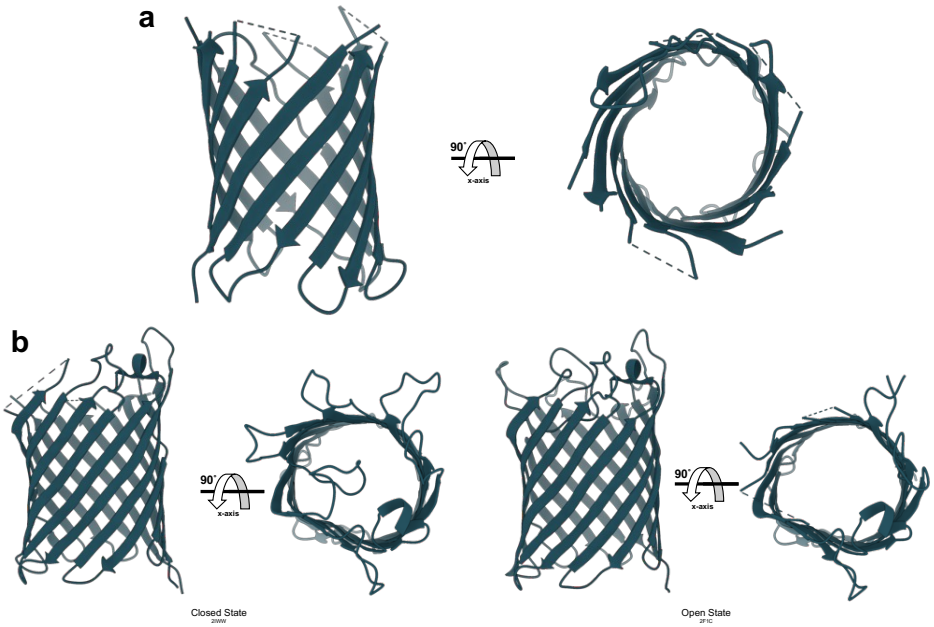


Figure 3-11: Cartoon Diagram of Specific Porins. a) Side view and top view of 12-stranded β-barrel, KdgM (4FQE). b) Side and top view of OmpG in an open and closed state.

constriction region is formed differently to classical porins, instead of a loop 3 folding into the pore, the constriction points in both of these porins are formed from the side chains of the residues that line the pore (Hutter et al., 2014; Subbarao and van den Berg, 2006). OmpG has also been shown to have a closed state where the pore is blocked by loop 6 (Figure 3-11b) at low pH (Yildiz et al., 2006), which has not been demonstrated for classical porins.

MspA forms a unique overall fold resembling a goblet, with a rim, stem, and base (Figure 3-12), that has a total height of 96 Å. The stem and base both form 16-stranded β-barrels of different diameters (40 Å (C_α atoms) and 28 Å (C_α atoms), respectively) and heights and the rim is made up of eight copies of a sandwich of two 4-stranded antiparallel β-sheets (Faller et al., 2004). The constriction region is between the two 16-stranded β-barrels and, like KdgM and OmpG, is formed by the side chains of the residues in the pore (Faller et al., 2004). As with all porins there is a hydrophobic band that encompasses both 16-stranded β-barrels, and the rim's surface is polar. Unlike other porins there are no aromatic girdles, but there is a band of arginine residues between the stem and rim (Faller et al., 2004).

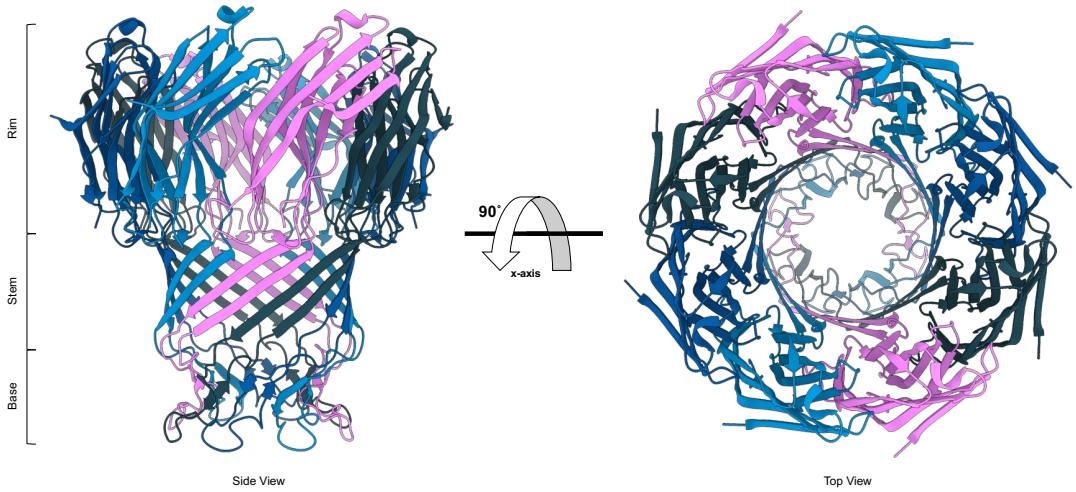


Figure 3-12: Cartoon Diagram of MspA.

3.1.2.3 Specific Channels Structure

Structures of specific channels are more diverse than porins, with the barrel sizes ranging from 12- to 18-stranded β-barrels with multiple examples at each size of barrel (Figure 3-8). As with porins specific channels have been found to form monomers and homotrimers, with those that form homotrimers using loop 2 to latch onto the adjacent monomer just like porins. As with porins they all have a hydrophobic band lined with aromatic girdles and loops that are longer on the extracellular side and shorter on the periplasmic side. The main difference is that they contain binding pockets for the substrate class.

3.1.3 OMPs in *B. bacteriovorus,* the Story of Bd0427

B. bacteriovorus HD100 has 100 proteins predicted to localise to the OM, with some of these predicted to encode proteins with homology to known OMPs, such as Bd1230 and Bd0667 & Bd3258, with homology to LamB and Tsx, respectively. Over 60% of these proteins share no homology to proteins with known functions (Lambert et al., 2008). This means a large proportion of the OMPs in *B. bacteriovorus* are not characterised and with the OM most likely

playing an important part in the predatory life cycle it is important that these proteins are understood.

Early work on OMPs in *B. bacteriovorus* also suggested that during the growth phase, prey OMPs were taken up by the predator unmodified (Diedrich et al., 1984, 1983; Guerrini et al., 1982). This was supported by observations of isolated OMPs from *B. bacteriovorus* with similar protein digests to OMPs from prey, and with OMP samples from prey and predator binding anti-OmpF antibody (Talley et al., 1987). It was not understood as to why *B. bacteriovorus* would need to acquire intact OMPs from prey. This was later disputed, with observations that suggested no OMPs were taken up intact from prey to predator, but that prey cells that have been invaded contained an OmpF-like protein in their inner membrane (Beck et al., 2004; Tudor and Karp, 1994). This provided the first evidence that *B. bacteriovorus* may transfer one of its OMPs to prey cells during predation. It was not until studies by Beck *et al.* (2004) and Barel *et al.* (2005) that the protein was identified using mass spectrometry, identifying an OMP with no homology to any other OMPs of known function (the previously mentioned Bd0427, HD100).

Following the identification that an OMP is inserted into the IM, it is interesting to think about the reason for this transfer. Two currently uncharacterised processes within the life cycle could be explained by this protein's insertion into the IM. There is currently no explanation for the disruption of the proton motive force, which results in prey cell death. This protein could be inserted into the inner membrane and allow passive diffusion through the IM interrupting the proton motive force required for the prey to respire normally. This has previously been observed with PorB, *Neisseria gonorrhoea,* where PorB is inserted into the immune cell's mitochondria's OM. This causes loss of electron potential resulting in apoptosis of the immune cell (Deo et al., 2018). Also, there is currently no understanding for how nutrients are transported from the prey's cytoplasm to the prey's periplasm. This OMP could enable transport of these molecules

across the IM and allow import into the growing predator. An alternative possibility is that as *B. bacteriovorus* releases many vesicles upon entry into the periplasm (Kaplan et al., 2022) and Bd0427 is the major OMP of *B. bacteriovorus* the protein could be inserted into the IM by chance (Beck et al., 2004).

3.1.4 Foreword

As Bd0427 shares no sequence homology with any proteins with known structures, a structural approach has been taken to aid identification of function. To do this, X-ray crystallography has been selected. To be able to determine the structure, high quality crystals that diffract to high resolution are needed. This chapter will focus the steps involved in producing high-quality crystals of Bd0427.

3.2 Results

Constructs for the recombinant expression of Bd0427 in *E. coli* (BL21(DE3), Rosetta, CodonPlus (DE3) RIPL) OM have previously been screened by our lab and no expression of Bd0427 was seen (data not shown). Therefore, three constructs for the expression of Bd0427 in inclusion bodies were designed. Attempts to refold these constructs were made before

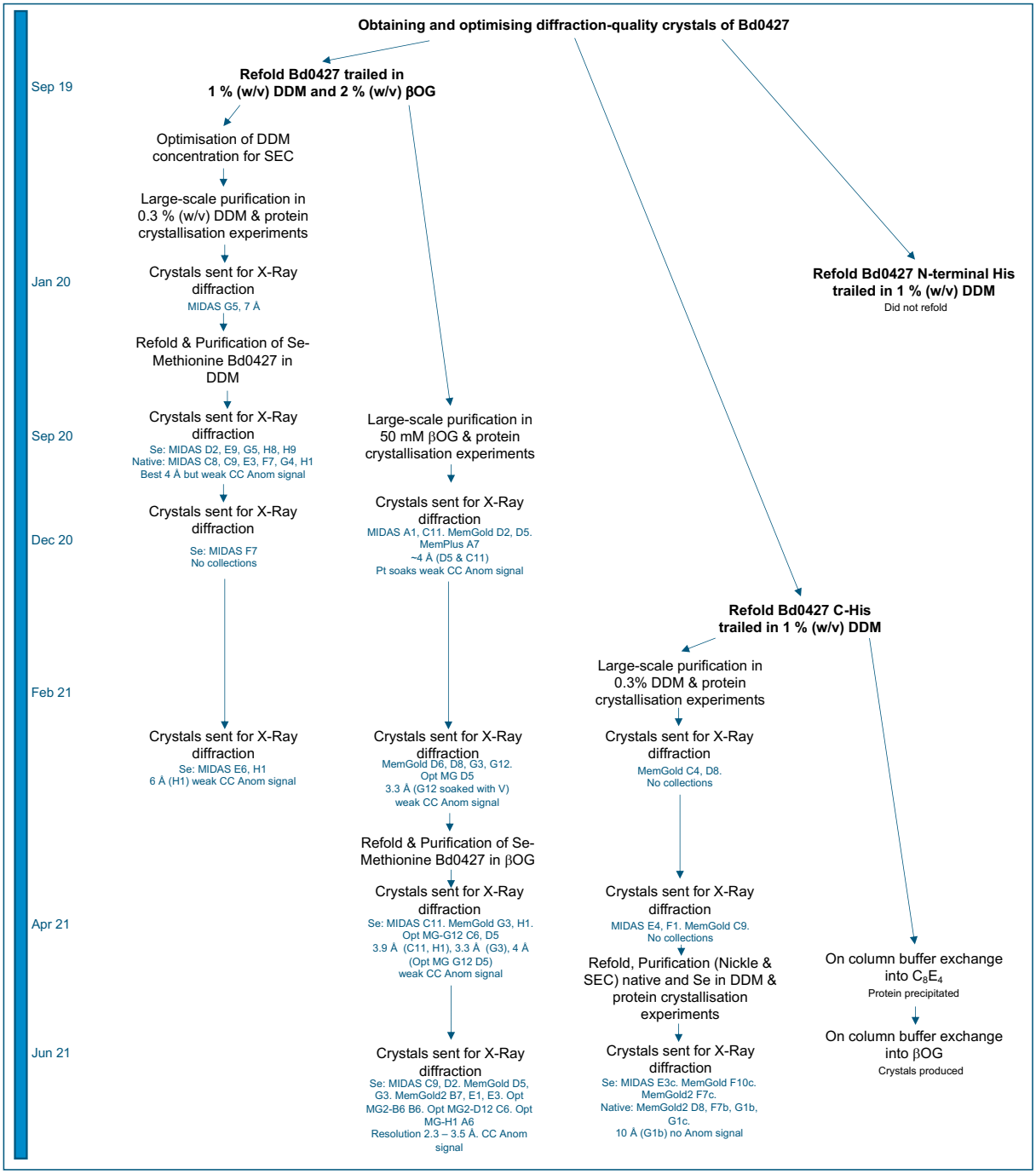


Figure 3-13: Workflow of the Production of High-Quality Protein Crystals.

purification and crystallisation trials were set up. A workflow to produce high-quality crystals of Bd0427 can be seen in Figure 3-13.

3.2.1 Bd0427 ΔSP

Bd0427 ΔSP was designed with no signal peptide for expression in inclusion bodies, with no affinity tags. As affinity tags can interfere with crystallisation, this construct was studied first to see if it could be refolded and to optimise purification and crystallisation.

3.2.1.1 Refold Optimisation of Bd0427 ΔSP from Inclusion Material

Bd0427 ΔSP was expressed without a signal peptide resulting in the protein remaining in the cytoplasm where it was packaged into inclusion bodies. Following purification of the inclusion body material, inclusion material was denatured in 8 M urea to a final concentration of 0.1 mM of protein. The solubilised inclusion material was then rapidly diluted using a dropwise method into detergent-containing buffers (Section 2.5.1.2). Utilising the heat modifiability property of OMPs, a semi-native PAGE, containing samples of the unfolded inclusion material in 8 M urea and samples of refold attempts (either boiled for 10 minutes or unboiled), was run. This enabled

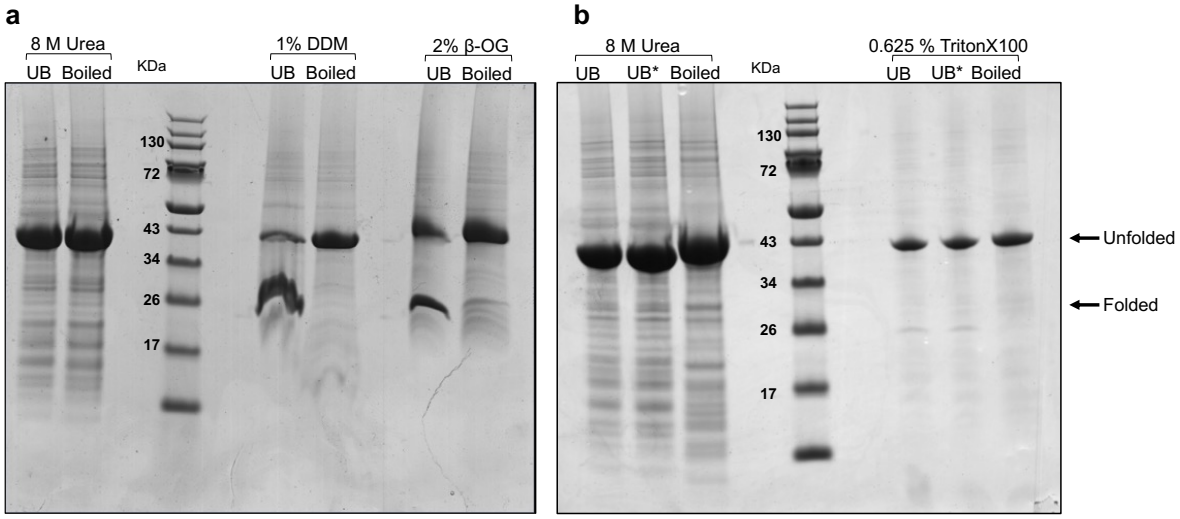


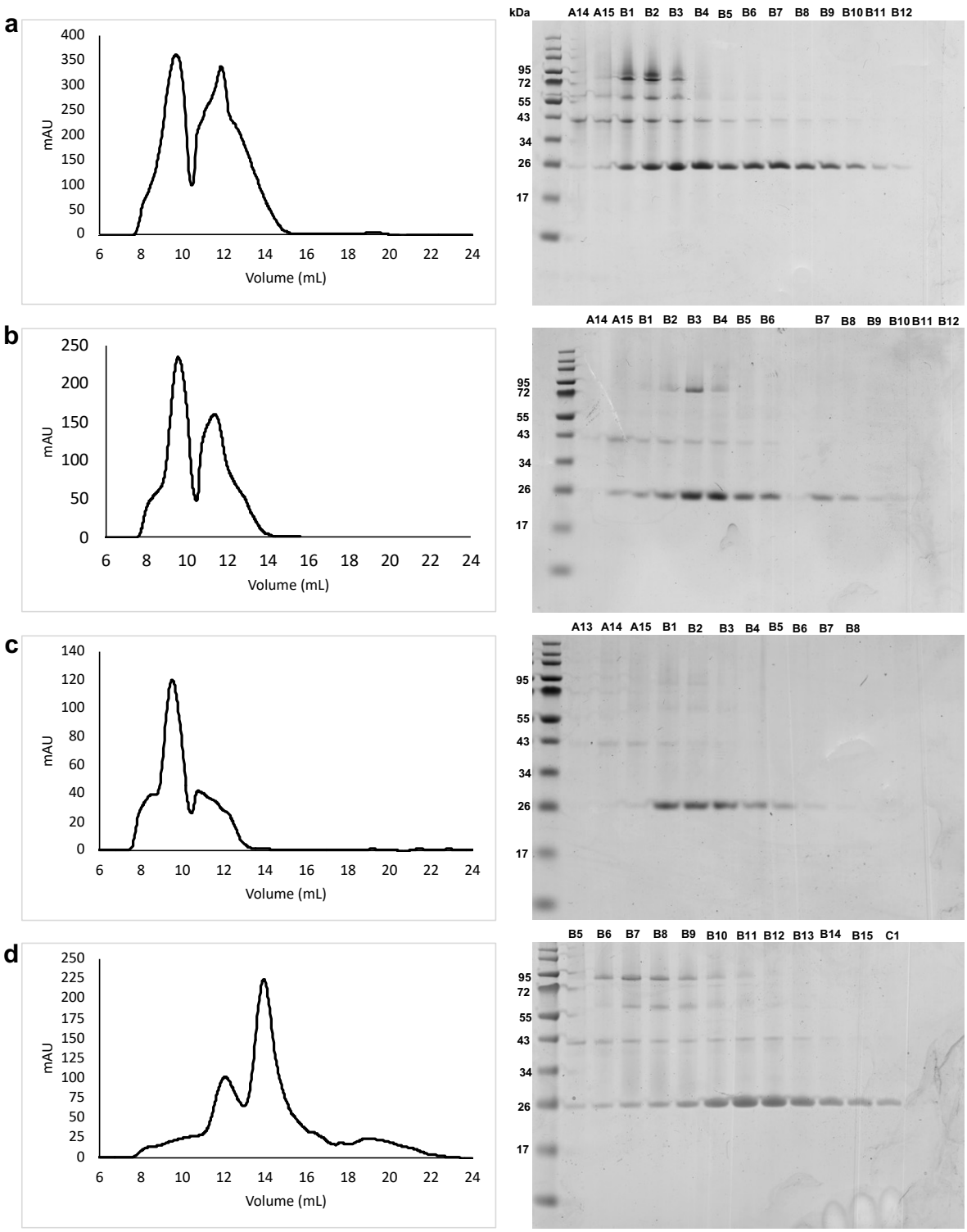
Figure 3-14: Semi-native PAGE of Bd0427 Refold in DDM, βOG or TritonX100. All samples in 1% SDS loading dye with the protein at 1 mg.mL^{-1}; one sample was boiled at 100 °C for 10 min and the other was unboiled (UB) (UB*, in 0.5 % SDS loading dye). Protein Ladder = Color Prestained Protein Standard, Broad Range (10 250 KDa) (NEB, P7719S).

comparison of each sample to identify which buffers produced the highest proportion of folded Bd0427, as folded OMPs run further than unfolded OMPs on a gel. As seen in Figure 3-14a, both buffers containing 1 % (w/v) DDM or 2 % (w/v) β-OG resulted in a band from the room temperature sample running further on the semi-native PAGE compared to the band in each 8 M urea and boiled samples. This was not the case with 1 % (v/v) Triton-X100 (Figure 3-14b) where the band for the room temperature sample ran the same distance as both the bands in the 8 M urea or boiled samples. This indicates that Bd0427 can be refolded in 1 % (w/v) DDM and 2 % (w/v) β-OG but not 1 % (v/v) Triton-X100, with the sample in 1 % DDM producing a more intense band at the migration distance of the folded protein.

3.2.1.2 Purification and Crystallisation of Refolded Bd0427 in DDM

3.2.1.2.1 Optimisation of Purification Method of Bd0427 in DDM

As refolding in 1% (w/v) DDM produced a higher proportion of folded to unfolded protein when compared to refolding in 2% (w/v) β-OG, buffers containing DDM were initially selected to carry forward into crystallisation screens. To find an optimal detergent concentration analytical size exclusion chromatography (SEC), at three concentrations of DDM (0.03 %, 0.1 % and 0.3 % (w/v)), was used to assess the homogeneity. As the oligomeric state of the protein was unknown, the protein was initially screened on a Superdex 75 10/300 GL column. The SEC runs, loaded with 250 µL at 2 $mg.mL^{-1}$ of protein sample, at the three different concentrations of DDM, all resulted in multiple polydisperse and non-baseline resolved peaks (Figure 3-15a-c), including a peak at the void volume (approximately 8 mL). All fractions with a positive mAU at 280 nm, were ran on a semi-native PAGE, to access purity and oligomeric state. The fractions from each SEC resulted in a semi-native PAGE with bands at four different molecular weights (approximately 26, 43, 55, 95 KDa), with the distribution weighted towards the band mAU at 280 nm, were ran on a semi-native PAGE, to access purity and oligomeric state. The



e

Fractions	A13	A14	A15	B1	B2	B3	B4	B5	B6	B7	B8	B9	B10	B11	B12	B13	B14	B15	C1
Volume (mL)	7.51	8.01	8.51	9.01	9.51	10.01	10.51	11.01	11.51	12.01	12.51	13.01	13.51	14.01	14.51	15.01	15.51	16.01	16.51

Figure 3-15: Analytical SEC of Bd0427 in DDM with corresponding Semi-native PAGE. a) S75 10 300 in 0.03 % (w/v) DDM. b) S75 10 300 in 0.1 % (w/v) DDM. c) S75 10 300 in 0.3 % (w/v) DDM. d) S200 10 300 in 0.3 % DDM. Protein Ladder = Color Prestained Protein Standard, Broad Range (10-250 KDa) (NEB, P7719S). e) Conversion of volume to fractions ran on the semi-native PAGE.

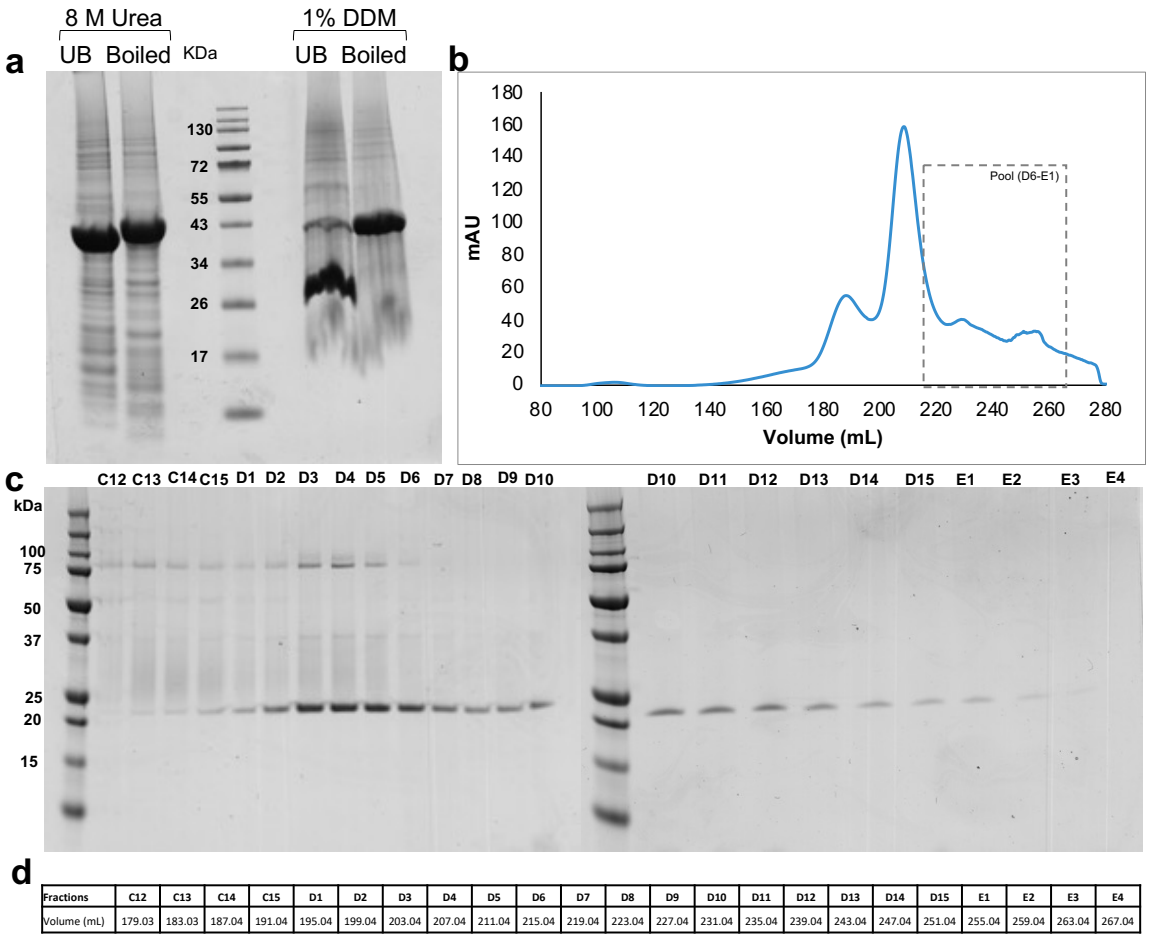
fractions from each SEC resulted in a semi-native PAGE with bands at four different molecular weights (approximately 26, 43, 55, 95 KDa), with the distribution weighted towards the band running at 26 KDa (Figure 3-15a-c). This suggests that the protein is forming oligomers of either differing molecular weights or one oligomer with a molecular weight of approximately 95 KDa is formed which is dissociating during SEC and/or in the semi-native PAGE. To better resolve the peaks from SEC, a Superdex 200 10/300 GL column, was used. SEC on this column was trialled in a buffer with 0.3% (w/v) DDM, as this buffer on a Superdex 75 10/300 GL column, produced fractions with the highest proportion of protein in a single band (26 KDa, Figure 3-15c).

This SEC run also resulted in multiple polydisperse, and non-baseline resolved peaks, with two main peaks (peak 1, 11.5-13 mL, and peak 2, 13.5-16 mL) both asymmetric with a tailing shoulder. The fractions for these peaks were run on a semi-native PAGE. The fractions of peak 1 resulted in bands at four molecular weights as seen previously, but here the band at 95 KDa was the major band, with the band at 26 KDa, becoming more prominent in the later fractions from the peak. The fractions of peak 2 produced four band as well, but the bands at 55 and 95 KDa, became very faint in the later fractions of the peak, here the major band was the band at 26 KDa. Following this the method was scaled up for a higher amount of refolded protein.

3.2.1.2.2 Large-scale Purification of Bd0427 in 0.3 % DDM

Bd0427 was refolded producing similar results to the initial refold trial (Figure 3-16a) and after concentration to 2 $mg.mL^{-1}$ (in 4 mL) Bd0427 was purified using a HiLoad 26/60 Superdex 200 pg column. The results in Figure 3-16b show Bd0427 eluted in multiple polydisperse and non-baseline resolved peaks, as was seen in Figure 3-15d. The fractions were subjected to semi-native PAGE (Figure 3-16c) to enable pooling of fractions. The fractions D6-E1 resulted in a

single band at approximately 23 KDa, therefore these fractions were pooled and the protein was concentrated in a 100 KDa spin concentrator for use in crystal screens.



Fractions	C12	C13	C14	C15	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15	E1	E2	E3	E4
Volume (mL)	179.03	183.03	187.04	191.04	195.04	199.04	203.04	207.04	211.04	215.04	219.04	223.04	227.04	231.04	235.04	239.04	243.04	247.04	251.04	255.04	259.04	263.04	267.04

Figure 3-16: Refold and Purification of Bd0427 in DDM. a) Semi-native PAGE of refold in DDM. All samples in 1% SDS loading dye; one sample was boiled at 100 °C for 10 min and the other was unboiled (UB). Protein Ladder = Color Prestained Protein Standard, Broad Range (10 250 KDa) (NEB, P7719S). b) Size exclusion chromatogram from HiLoad 26/60 Superdex 200 pg column eluted with tris (pH 8), 100 mM NaCl, 0.3 % (w/v) DDM. c) Semi-native PAGE of SEC peak. Protein ladder = Precision Plus Protein All Blue Standard (Bio Rad). d) Conversion of volume to fractions ran on the semi-native PAGE.

3.2.1.2.3 Initial Crystal Hits of Bd0427 in DDM

Bd0427 was initially screened in MIDASplus and Morpheus (Molecular Dimensions) at 7.5 $mg.mL^{-1}$ using vapour diffusion. This resulted in two crystal hits (Figure 3-17a and b), both in MIDASplus. Crystals from these conditions were looped using cryoloops and cryocooled in liquid nitrogen prior to transport to Diamond Light Source. The crystals (Figure 3-17a and b)

were subject to an X-ray beam and diffraction data was collected at Diamond's IO3 beamline. The crystals in Figure 3-17a did not diffract but the crystals in Figure 3-17b diffracted to a resolution of 20 Å. Following these results an optimisation crystal tray was set up (Appendix 4.2). The optimisation screen generated the crystals in Figure 3-17c, which when subjected to X-rays, diffracted to a resolution of 7 Å.

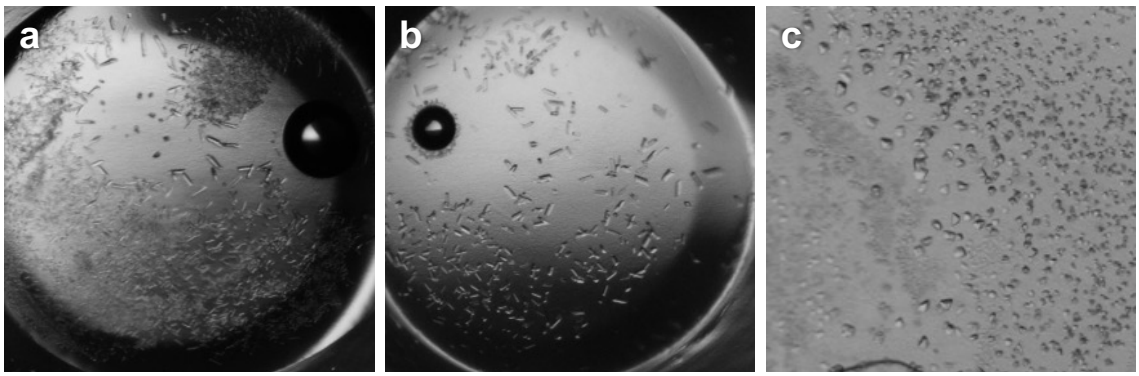


Figure 3-17: Initial Crystal Hits of Bd0427 in DDM. Crystal hit in a) MIDASplus E4 (0.1 M HEPES (pH 7), 0.2 M sodium thiocyanate, 40 % (v/v) Pentaerythritol propoxylate (5/4 PO/OH)), b) MIDASplus G5 (55 % (v/v) PEG400), c) optimisation screen of MIDASplus G5 (55 % PEG400, drop size 4µL:4µL (protein:reservoir)).

3.2.1.2.4 Crystal Hits from Seeding Experiments

To produce crystals that diffracted to higher resolution the crystals from MIDASplus G5, were used to generate a seed stock to provide nucleation points to the protein in solution to increase the likelihood of crystal formation in new conditions. The seed stock and Bd0427, at 5 $mg.mL^{-1}$, were used in MIDASplus and Morpheus (see Appendix 4.1 for ratios). Seeding resulted in four new crystal hits in MIDASplus (Figure 3-18), and the crystals were subjected to X-ray diffraction. The crystals were also derivatised by soaking in cryoprotectant solution supplemented with 10 mM tetrachloroplatinate. All crystals from these conditions produced low resolution diffraction, with a native crystal from MIDASplus H1 (Figure 3-18d), producing the best resolution dataset at 4 Å. None of the tetrachloroplatinate-soaked crystals produced a diffraction pattern with anomalous signal.

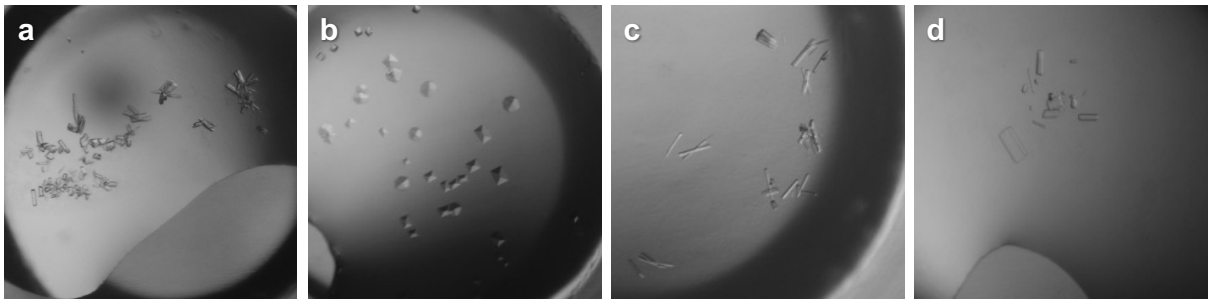


Figure 3-18: Crystal Hits from Seeding Trays of Bd0427 in DDM. Crystal hit in a) MIDASplus C8 (0.2 M NaCl, 0.1 M Tris (8), 12.5 % (w/v) Polyvinylpyrrolidone), b) MIDASplus C9 (0.1 M NaCl, 25 % (v/v) Pentaerythritol porpoxylate (5/4 PO/OH), 10 % (v/v) DMSO), c) MIDASplus E3 (0.3 M KCl, 0.1 M Glycine (9.5), 20 % (v/v) Pentaerythritol ethoxylate (15/4 EO/OH)), d) MIDASplus H1 (0.1 M Bis-Tris (6), 25 % (w/v) SOKALAN CP 42).

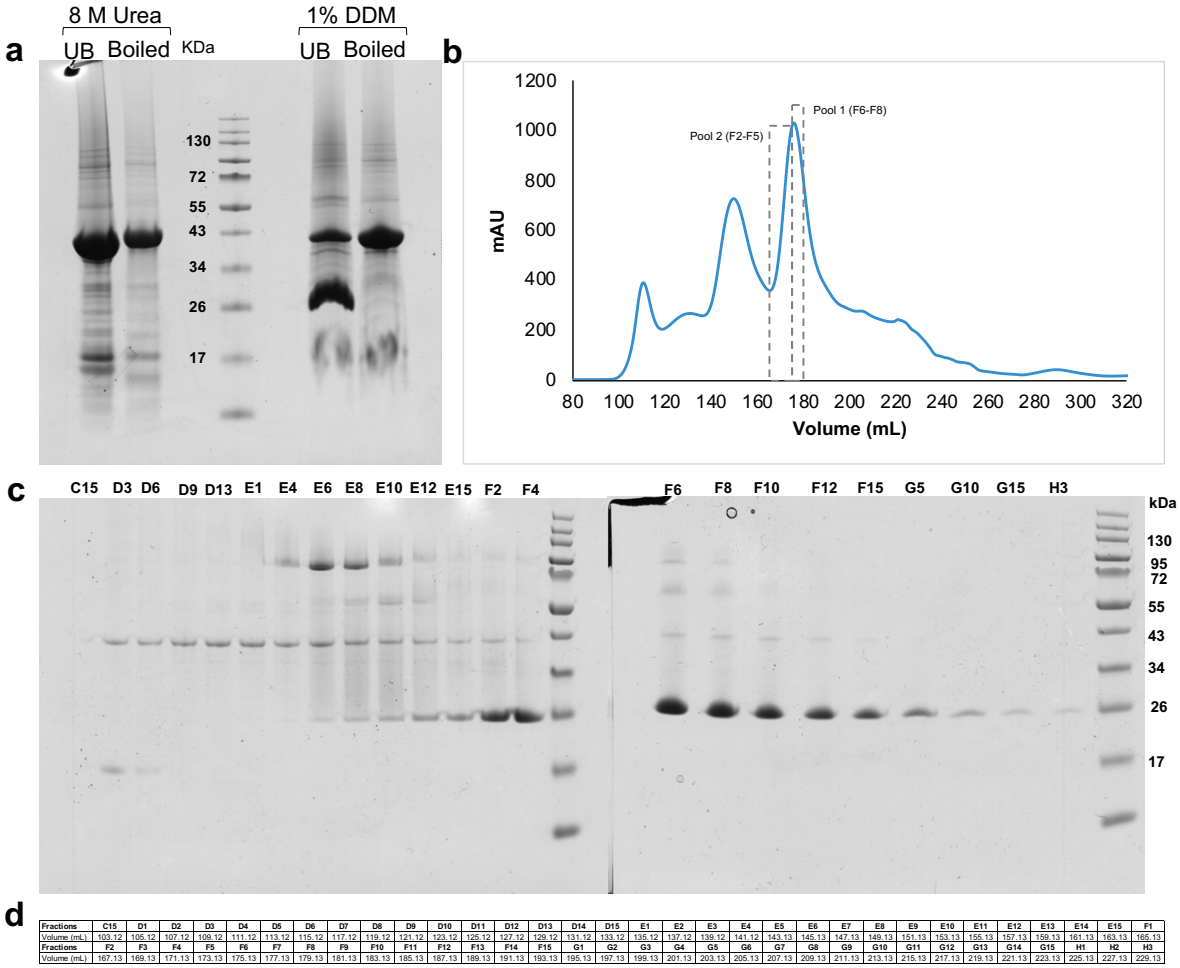
3.2.1.3 Refold, Purification and Crystallisation of Selenomethionine Bd0427 in DDM

As previously mentioned, there are no homologous proteins to Bd0427 with known protein structures, therefore molecular replacement cannot be used to estimate the phases of the structure factors. Therefore, to solve the phase problem, heavy atoms must be used. Tetrachlorplatinate-soaked crystals did not produce a diffraction pattern with anomalous signal, so selenomethionine-labelled Bd0427 was produced. Crystals formed from this protein would be subjected to X-rays at the wavelength of the selenium edge to give anomalous signal in the diffractions pattern.

3.2.1.3.1 Refold and Purification of Se-Met Bd0427 in DDM

Selenomethionine-labelled Bd0427 from inclusion material was successfully refolded in 1 % (w/v) DDM (Figure 3-19a). This sample was concentrated to 8 $mg.mL^{-1}$ (in 4 mL) before loading on a HiLoad 26/60 Superdex 200 pg. Se-methionine Bd0427 eluted in multiple polydisperse, and non-baseline resolved peaks (Figure 3-19b), with new peaks that did not correspond to the peaks from the SEC chromatogram of native Bd0427 (Figure 3-16b). Fractions were run on a semi-native PAGE (Figure 3-19c) to enable pooling of fractions. Fractions with a major band at a molecular weight of approximately 26 KDa were pooled into

two pools (Pool 1, 175-179 mL (F6-F8) and Pool 2, 167-175 mL (F2-F5)) and concentrated for crystallisation.



Fractions	C15	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15	E1	E2	E3	E4	E5	E6	E7	E8	E9	E10	E11	E12	E13	E14	E15	F1
Volume (mL)	103.12	105.12	107.12	109.12	111.12	113.12	115.12	117.12	119.12	121.12	123.12	125.12	127.12	129.12	131.12	133.12	135.12	137.12	139.12	141.12	143.13	145.13	147.13	149.13	151.13	153.13	155.13	157.13	159.13	161.13	163.13	165.13
Fractions	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	G1	G2	G3	G4	G5	G6	G7	G8	G9	G10	G11	G12	G13	G14	G15	H1	H2	H3
Volume (mL)	167.13	169.13	171.13	173.13	175.13	177.13	179.13	181.13	183.13	185.13	187.13	189.13	191.13	193.13	195.13	197.13	199.13	201.13	203.13	205.13	207.13	209.13	211.13	213.13	215.13	217.13	219.13	221.13	223.13	225.13	227.13	229.13

Figure 3-19 Refold and Purification of Bd0427 Se-Met in DDM. a) Semi-native PAGE of refold in DDM. All samples in 1% SDS loading dye; one sample was boiled at 100 °C for 10 min and the other was unboiled (UB). Protein Ladder = Color Prestained Protein Standard, Broad Range (10 250 KDa) (NEB, P7719S). b) Size exclusion chromatogram from HiLoad 26/60 Superdex 200 pg column eluted with tris (pH 8), 100 mM NaCl, 2 mM β-mercaptoethanol, 0.3 % (w/v) DDM. c) Semi-native PAGE of SEC fractions. Protein Ladder = Color Prestained Protein Standard, Broad Range (10 250 KDa) (NEB, P7719S). d) Conversion of volume to fractions ran on the semi-native PAGE.

3.2.1.3.2 Crystal Hits of Se-Met Bd0427 in DDM

Pool 1 was concentrated to 9 mg.mL^{-1} and screened in MIDASplus and Morpheus (Molecular Dimensions). A selection of the crystals produced are shown in Figure 3-20, all but the crystals from MIDASplus A1, D2 and E4 (Figure 3-20a, b and d) were submitted to X-ray diffraction

at IO3 beamline at Diamond Light Source. The crystal from MIDASplus H8 diffracted to the highest resolution (4 Å) but did not produce a diffraction pattern with anomalous signal.

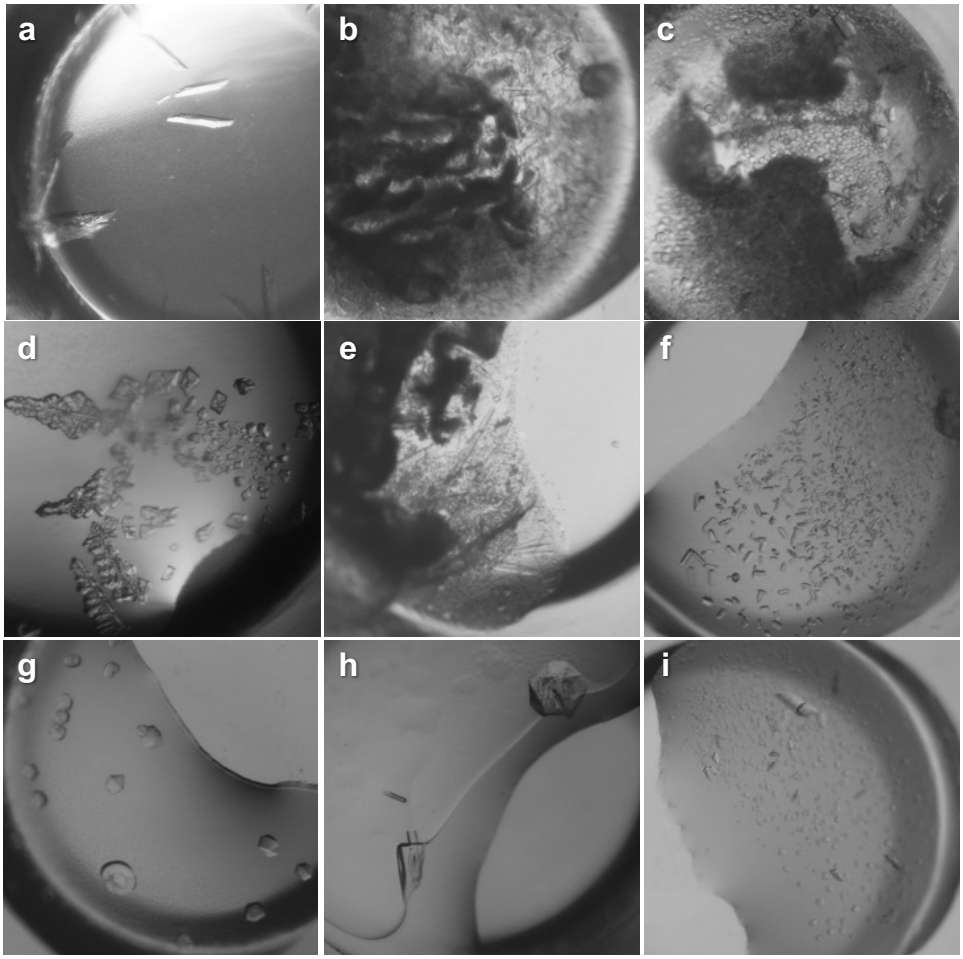


Figure 3-20: Crystal hits of Bd0427 Se-Methionine Pool 1. Crystal hits in a) MIDASplus A1(0.1 M HEPES (6), 50 % (v/v) PPG400, 5 % (v/v) DMSO), b) MIDASplus D2 (0.1 M HEPES (7.5), 30 % (v/v) Pentaerythritol ethoxylate (15/4 EO/OH), 6 % (w/v) Polyvinylpyrrolidone), c) MIDASplus D8 (0.1 M Tris (8), 20 % (v/v) Glycerol ethoxylate, 10 % (v/v) Tetrahydrofurane), d) MIDASplus E4 (0.2 M Sodium thiocyanate, 0.1 M HEPES (7), 40 % (v/v) Pentaerythritol propoxylate (5/4 PO/OH)), e) MIDASplus E9 (0.1 M Lithium sulfate, 0.1 M Tris (8), 25 % (v/v) Jeffamine ED-2003), f) MIDASplus G5 (55 % PEG400), g) MIDASplus G8 (0.1 M Tris (8), 20 % (w/v) SOKALAN CP 42, 5 % (v/v) Methanol), h) MIDASplus H8 (0.1 M Tris (8.5), 30 % (w/v) SOKALAN CP 42), i) MIDASplus H9 (0.2 M Ammonium acetate, 0.1 M HEPES (7), 25 % SOKALAN HP 56).

Pool 2 was also screened in MIDASplus and Morpheus at a concentration of 4.5 $mg.mL^{-1}$. This resulted in the formation of the crystals shown in Figure 3-21. Crystals from MIDASplus F7 (Figure 3-21b) were sent to Diamond Light Source and submitted to X-ray diffraction at IO3 beamline. These crystals were weak diffractors and no full datasets were collected.

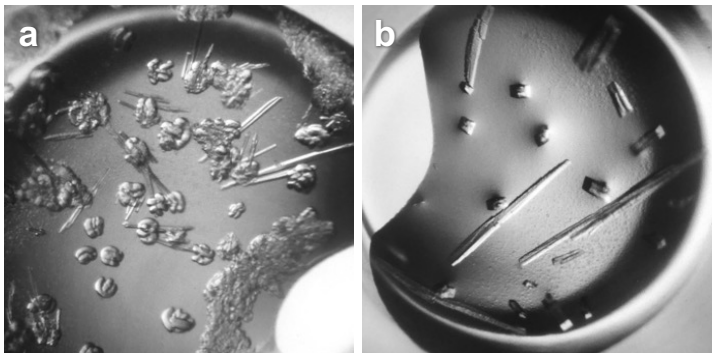


Figure 3-21: Crystal hits of Bd0427 Se-Methionine Pool 2. Crystal hits in a) MIDASplus B2 (0.2 M Sodium thiocyanate, 0.1 M HEPES (7), 15 % (v/v) Pentaerythritol propoxylate (5/4 PO/OH)), b) MIDASplus F7 (25 % (w/v) SOKALAN PA 25 CL).

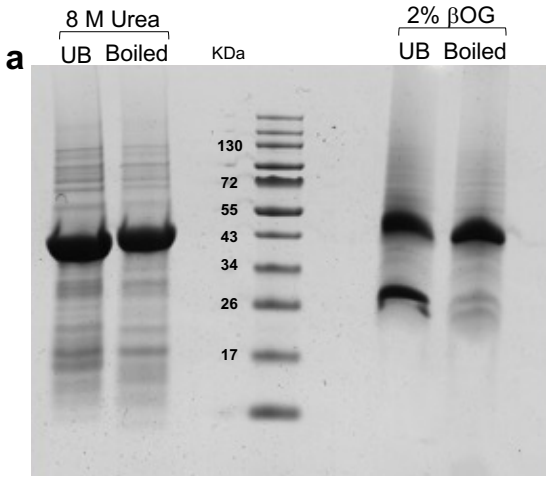
3.2.1.4 Refold, Purification and Crystallisation of Bd0427 in βOG

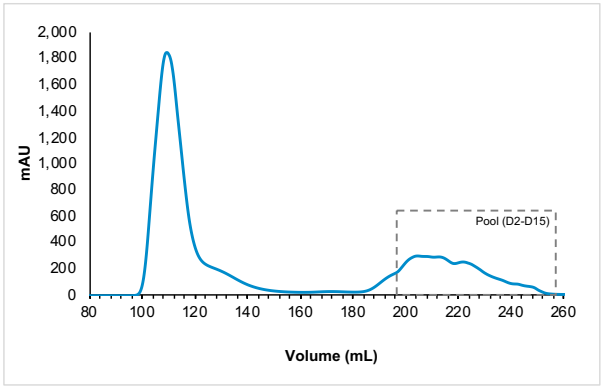
Despite screening several crystals of Bd0427 in DDM, the best resolution of a dataset was 4 Å with no anomalous signal. Therefore, it was decided to return to refolding in βOG to see if the resolution could be improved using a different detergent.

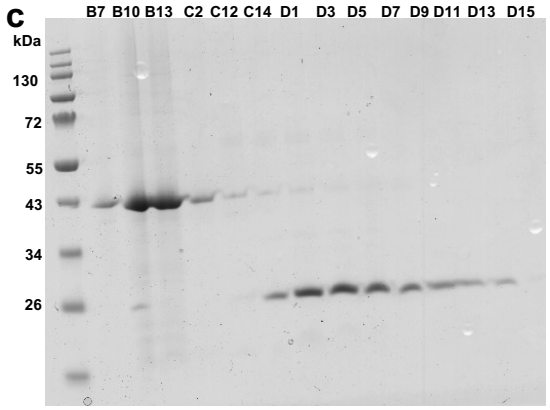
3.2.1.4.1 Refold and Purification of Bd0427 in βOG

Bd0427 was refolded in 2 % (w/v) βOG with similar efficiency to the original refold trial (Figure 3-22a). This was concentrated to 8 $mg.mL^{-1}$ (in 4 mL) for loading on to HiLoad 26/60 Superdex 200 pg column, pre-equilibrated in 50 mM tris (pH 8), 150 mM NaCl and 50 mM βOG. Unlike with DDM, the SEC chromatogram for Bd0427 in βOG had two peaks (Figure 3-22b), the first elutes in void volume of the column, the second is polydisperse with fronting and tailing shoulders and elutes at a similar volume to the pooled fractions of Bd0427 in DDM. Fractions of the peaks were run on a semi-native PAGE, to enable pooling of fractions. The fractions of the peak that eluted at the void volume resulted in a single band on the semi-native PAGE, at a 43 KDa, this corresponds to the band of the unfolded protein in the semi-native PAGE of the refolded protein (Figure 3-22c). The fractions of the second peak also resulted in a single band at approximately 26 KDa, this corresponds to the molecular weight of the folded protein. On the semi-native PAGE, fractions C12-D1 also have a faint band at 55 KDa, due to

this these were excluded when pooling fractions (Figure 3-22c). The pooled fractions (199-251 mL (D2-D15)) were concentrated in a 30 KDa spin concentrator ready for crystallisation.







d

Fractions	B7	B8	B9	B10	B11	B12	B13	B14	B15	C1
Volume (mL)	99.16	103.16	107.16	111.16	115.16	119.16	123.16	127.16	131.16	135.16
Fractions	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11
Volume (mL)	139.16	143.16	147.16	151.16	155.17	159.17	163.17	167.17	171.17	175.17
Fractions	C12	C13	C14	C15	D1	D2	D3	D4	D5	D6
Volume (mL)	179.17	183.17	187.17	191.17	195.17	199.17	203.17	207.17	211.17	215.17
Fractions	D7	D8	D9	D10	D11	D12	D13	D14	D15	
Volume (mL)	219.17	223.17	227.17	231.17	235.17	239.17	243.17	247.17	251.17	

Figure 3-22: Refold and Purification of Bd0427 in βOG. a) Semi-native PAGE of refold in β-OG. All samples in 1% SDS loading dye; one sample was boiled at 100 °C for 10 min and the other was unboiled (UB). Protein Ladder = Color Prestained Protein Standard, Broad Range (10 250 KDa) (NEB, P7719S). b) Size exclusion chromatogram from HiLoad 26/60 Superdex 200 pg column eluted in 50 mM tris (pH 8), 150 mM NaCl and 50 mM β-OG. c) Semi-native PAGE of SEC peaks. Protein Ladder = Color Prestained Protein Standard, Broad Range (10 250 KDa) (NEB, P7719S). d) Conversion of volume to fractions ran on the semi-native PAGE.

3.2.1.4.2 Initial Crystal Hits of Bd0427 in βOG

The pooled protein was concentrated to 4.5 mg.mL^{-1} and screened in MIDASplus, MemGold, and MemPlus (Molecular Dimensions). Crystals initially formed in the four conditions (Figure 3-23), these were looped using cryoloops and cryocooled in liquid nitrogen before transport to Diamond Light Source. Native and tetrachloroplatinate-soaked crystals were subjected to X-ray diffraction at the IO3 beamline and datasets were collected. The crystals from MIDASplus

C11 and MemGold D5 resulted in diffraction data at the best resolution (4 Å), which included a tetrachloroplatinate-soaked crystals from MIDASplus C11. The CC-Anomalous of this data was below 0.3, and the phases could not be determined using this anomalous signal.

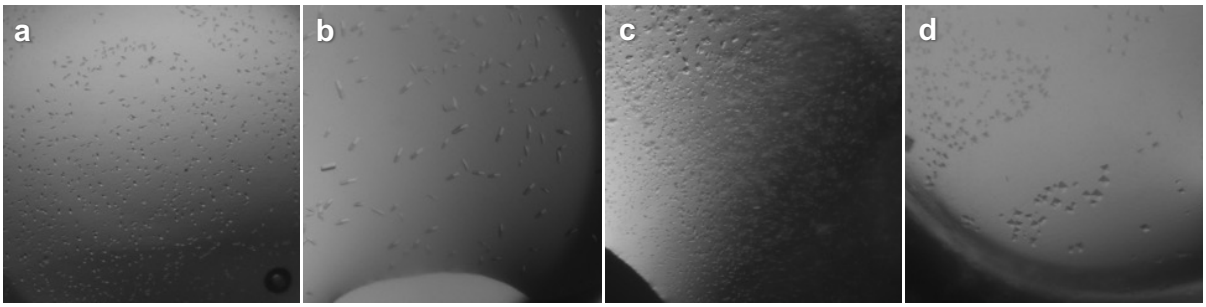


Figure 3-23: Initial Crystal Hits of Bd0427 in βOG. Crystal hits in a) MIDASplus A1 (0.1 M HEPES (6), 50 % (v/v) PPG400, 5 % (v/v) DMSO), b) MIDASplus C11 (0.1 M Magnesium formate dihydrate, 0.1 M Tris (8.5), 30 % (v/v) Pentaerythritol ethoxylate (15/4 EO/OH)), c) MemGold D2 (0.05 M Magnesium acetate tetrahydrate, 0.05 M Sodium acetate (5), 28 % (v/v) PEG400), MemGold D5 (0.05 M Magnesium acetate tetrahydrate, 0.05 M Sodium acetate (5.4), 24 % (v/v) PEG400).

3.2.1.4.3 Crystal Hits of Bd0427 in βOG from Commercial and Optimisation Screens

After a couple more months there were several more crystal hits in the MemGold (Figure 3-24a-i) as well as several hits in the optimisation screen based on the condition from MemGold D5 (Appendix 4.3)(Figure 3-24j-l). A selection of these crystals (Figure 3-24b, c, e, f, j, k, l) were selected for subjection to X-ray diffraction at the IO4 beamline at Diamond Light Source. All crystals were taken native and as a heavy atom soak in either tetrachloroplatinate or sodium orthovanadate. The two best crystal conditions were MemGold G3 and G12 with highest resolutions of 3.9 and 3.3 Å, respectively. Both crystals produced diffraction with weak anomalous signal (CCanom<0.3), meaning the structures could not be solved.

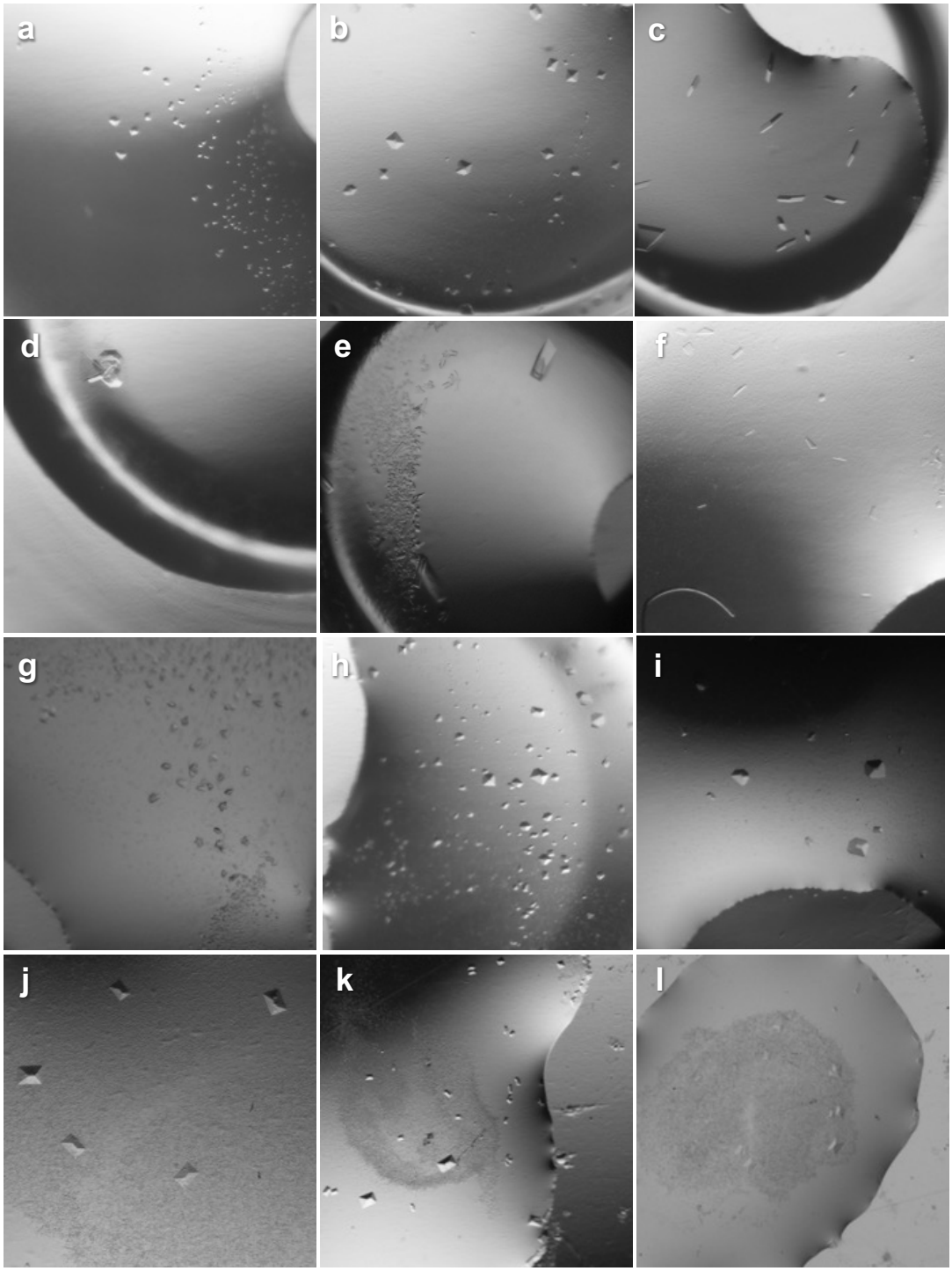


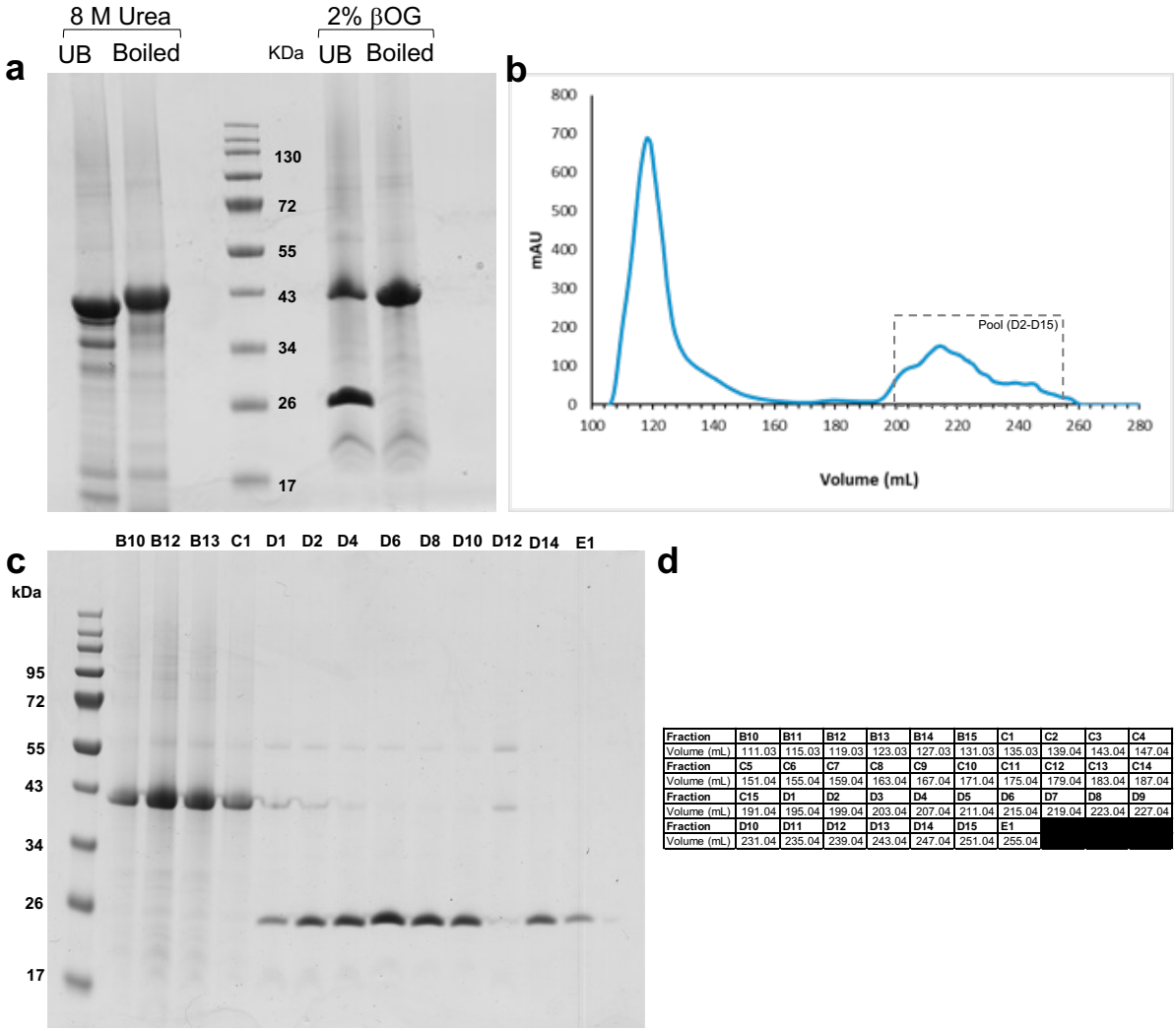
Figure 3-24: Crystal Hits of Bd0427 in βOG. Crystal hits in a) MemGold B3 (0.05 M $CaCl_2$, 0.05 M Barium chloride, 0.1 M Tris (8.2), 32 % (v/v) PEG 400, b) MemGold D8 (0.05 M Magnesium chloride hexahydrate, 0.1 M Glycine (9), 22 % (v/v) PEG400), c) MemGold G3 (0.05 M Magnesium chloride hexahydrate, 0.1 M NaCl, 0.1 M Tris (8.5), 33 % (v/v) PEG 400), d) MemGold G8 (0.1 M NaCl, 0.1 M MES (6.5), 33 % (v/v) PEG400, 4 % (v/v) Ethylene glycol), e) MemGold G12 (1 mM Zinc sulfate heptahydrate, 0.05 M HEPES (7.8), 28 % (v/v) PEG600), j, k, l) Conditions A2, A3, A5 from Optimisation Screen of MemGold D5 (0.05 M Magnesium acetate tetrahydrate, 0.05 M Sodium acetate (5), 28 % (v/v) PEG400).

3.2.1.5 Refold, Purification and Crystallisation of Se-Met Bd0427 in βOG

As refolding Bd0427 in βOG resulted in the crystal that produced the best resolution to date, but still did not enable solving due to a weak anomalous signal to solve the phases, I decided to attempt refolding and crystallisation of selenomethionine-labelled Bd0427 in βOG.

3.2.1.5.1 Refold and Purification of Se-Met Bd0427 in βOG

Selenomethionine-labelled Bd0427 was successfully refolded in 2 % (w/v) βOG (Figure 3-25a). Following refolding, Bd0427 was concentrated to 9 $mg.mL^{-1}$ (in 3.5 mL) and run on a HiLoad 26/60 Superdex 200 pg column. This resulted in the SEC chromatogram in Figure 3-25b which has a similar profile to the SEC chromatogram of native Bd0427 in βOG, an initial peak in this case just after the void volume and a second polydisperse peak from 200-260 mL. The fractions of the peaks were run on a semi-native PAGE (Figure 3-25c) to enable pooling. The fractions from the first peak have a band at the distance of 43 KDa, which corresponds to the band of the unfolded protein. The fractions from the second peak all have a band at approximately 26 KDa, which corresponds to the folded protein, this is also the same as the bands seen for native Bd0427. As well as the bands seen at 26 KDa there are two faint bands at higher molecular weights (approximately 43 and 55 KDa), for the second peak which likely represent an oligomer (55 KDa) and some unfolded protein (43 KDa). The second peak was pooled (199-255 mL), as with native Bd0427 and concentrated for crystal screening.



Fraction	B10	B11	B12	B13	B14	B15	C1	C2	C3	C4
Volume (mL)	111.03	115.03	119.03	123.03	127.03	131.03	135.03	139.04	143.04	147.04
Fraction	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14
Volume (mL)	151.04	155.04	159.04	163.04	167.04	171.04	175.04	179.04	183.04	187.04
Fraction	C15	D1	D2	D3	D4	D5	D6	D7	D8	D9
Volume (mL)	191.04	195.04	199.04	203.04	207.04	211.04	215.04	219.04	223.04	227.04
Fraction	D10	D11	D12	D13	D14	D15	E1			
Volume (mL)	231.04	235.04	239.04	243.04	247.04	251.04	255.04			

Figure 3-25: Refold and Purification of Bd0427 Se-Met in βOG. a) Semi-native PAGE of refold in β-OG. All samples in 1% SDS loading dye; one sample was boiled at 100 °C for 10 min and the other was unboiled (UB). Protein Ladder = Color Prestained Protein Standard, Broad Range (10 250 KDa) (NEB, P7719S). b) Size exclusion chromatogram from HiLoad 26/60 Superdex 200 pg column eluted in 50 mM tris (pH 8), 150 mM NaCl, 3 mM β-mercaptoethanol and 50 mM β-OG. c) Semi-native PAGE of SEC peak. Protein Ladder = Color Prestained Protein Standard, Broad Range (10 250 KDa) (NEB, P7719S). d) Conversion of volume to fractions ran on the semi-native PAGE.

3.2.1.5.2 Crystal Hits of Se-Met Bd0427 in βOG

The pooled fractions were concentrated to 4.5 mg.mL^{-1} for use in MIDASplus and MemGold (Molecular Dimensions) and crystal optimisation trays of MemGold G12 (Appendix 4.4), the best diffracting crystal of Bd0427 in βOG. In the first month there were a few conditions with crystal in them (Figure 3-26), including conditions similar or the same as MemGold G12 (Figure 3-26c,e,f), though these crystals had a different morphology to the crystals of native

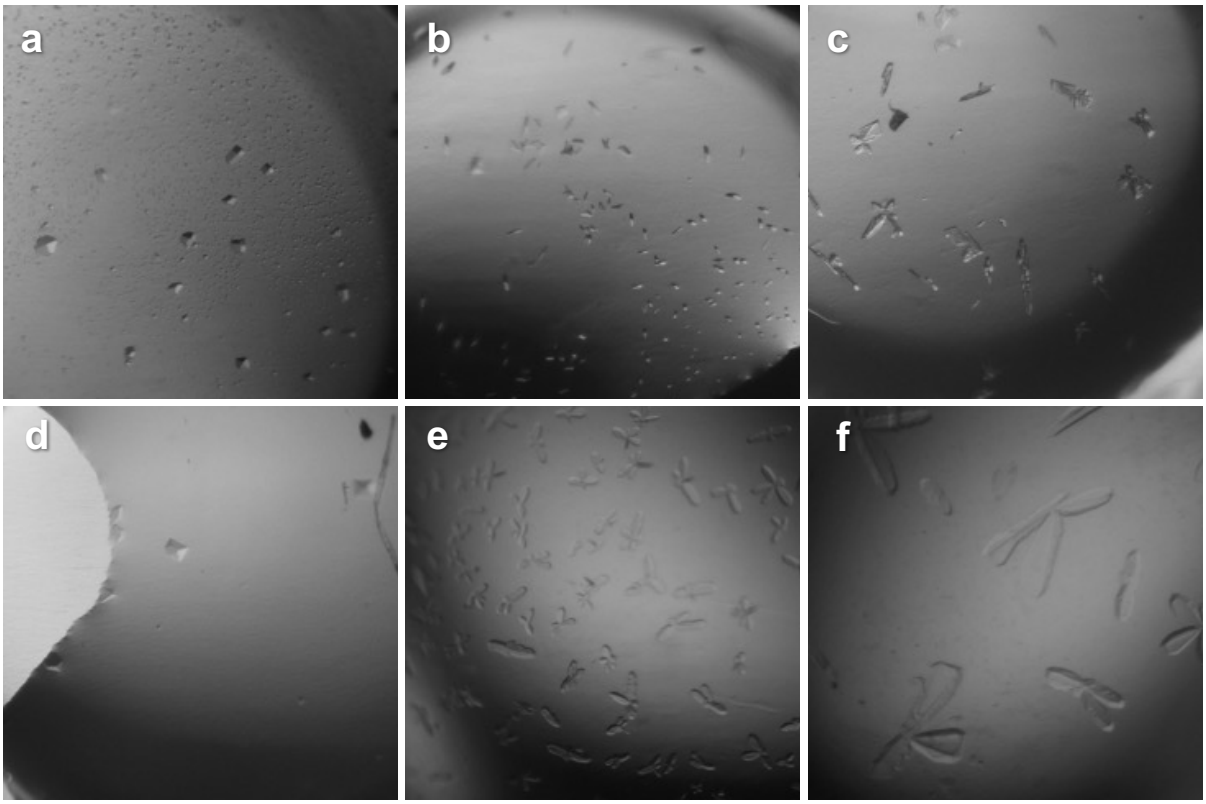


Figure 3-26: Crystal Hits of Se-Met Bd0427 in βOG sent to Diamond Light Source April 2021. Crystal hits in a) MIDASplus C11 (0.1 M Magnesium formate dihydrate, 0.1 M Tris (8.5), 30 % (v/v) Pentaerythritol ethoxylate (15/4 EO/OH)), b) MemGold G3 (0.1 M Magnesium chloride hexahydrate, 0.1 M NaCl, 0.1 M Tris (8.5), 33 % (v/v) PEG400), c) MemGold G12 (1 mM Zinc sulfate heptahydrate, 0.05 M HEPES (7.8), 28 % (v/v) PEG600), d) MemGold H1 (0.1 M NaCl, 0.1 M Sodium phosphate (7), 33 % (v/v) PEG300), e) Optimisation tray 1 of MemGold G12 C6 (1 mM Zinc sulfate heptahydrate, 0.05 M HEPES (7.8), 28 % (v/v) PEG600. Protein:Reservoir solution, 2 µL:1 µL), f) Optimisation tray 2 of MemGold G12 D5 (1 mM Zinc sulfate heptahydrate, 0.05 M HEPES (8.2), 32 % (v/v) PEG600. Protein:Reservoir solution, 3 µL:1 µL).

Bd0427 (Figure 3-24e). Crystals were selected from five of these conditions to be sent to Diamond Light Source (Figure 3-26a,b,d-f). The crystals were screened at beamline IO4, at 0.9794 Å. Datasets were collected for each condition with the highest resolution from crystals from MemGold G3 (Figure 3-26b), with a resolution of 3.3 Å. Despite labelling with selenomethionine the dataset had a low CCanom of less than 0.4, resulting in no structural determination, as the signal could not be used to calculate the phases.

Following this MemGold2 (Molecular Dimensions) as well as MemGold using Bd0427 spiked with 3 % (w/v) 1,2,3 heptanetriol and optimisation screens designed around successful crystallisation conditions (MemGold (G3, H1), MemGold2 (B6, D12, E11), MIDASplus (C11)

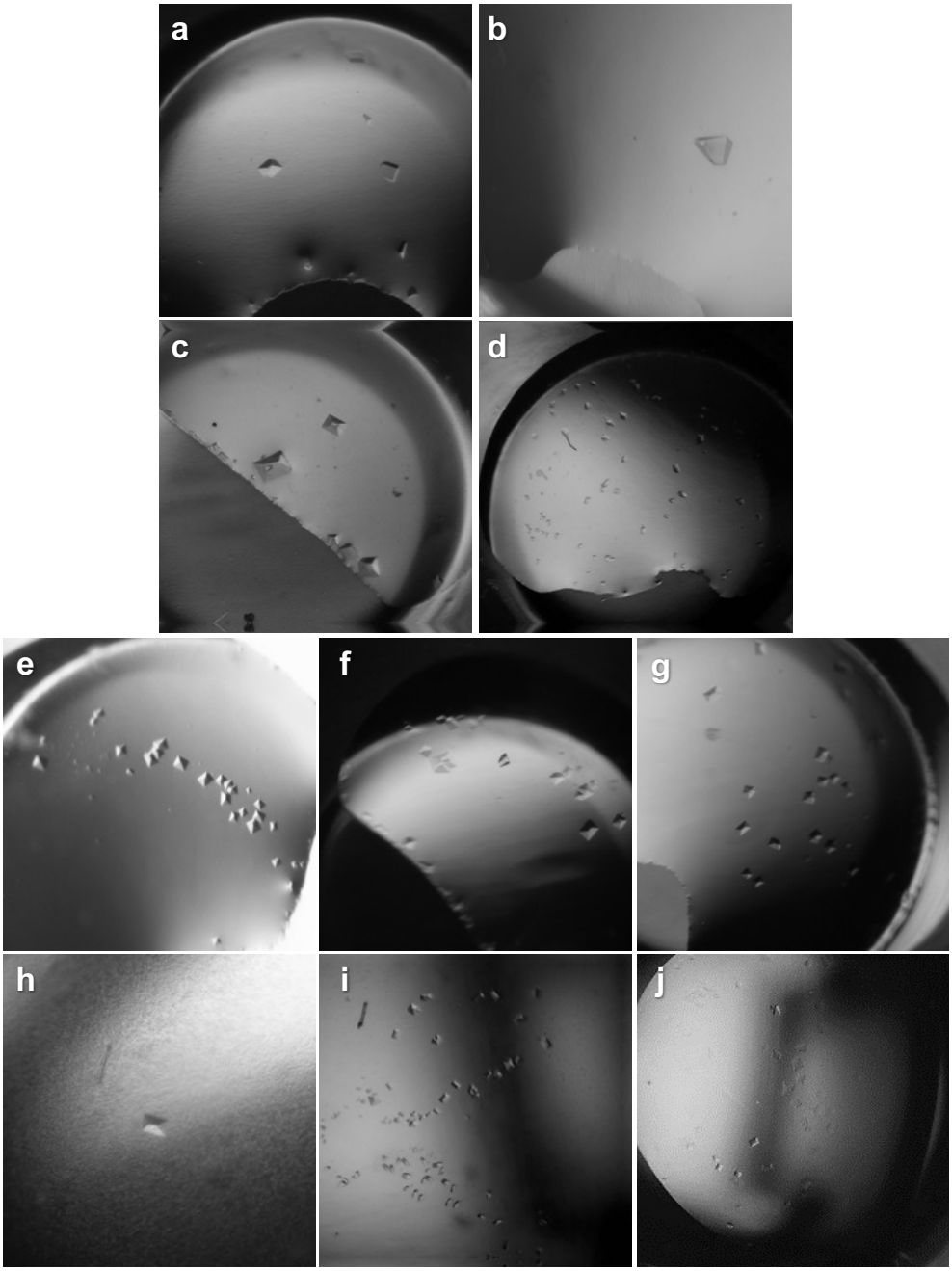


Figure 3-27: Crystal Hits of Se-Met Bd0427 in βOG Sent to Diamond Light Source June 2021. Crystal hits in a) MIDASplus C9 (0.1 M NaCl, 25 % (v/v) Pentaerythritol propoxylate (5/4 PO/OH), 10 % (v/v) DMSO), b) MIDASplus D2 (0.1 M HEPES (7.5), 30 % (v/v) Pentaerythritol ethoxylate (15/4 EO/OH), 6 % (w/v) Polyvinylpyrrolidone), c) MemGold D5 (0.05 M Magnesium acetate tetrahydrate, 0.05 M Sodium acetate (5.4), 24 % (v/v) PEG400), d) MemGold E12 (0.1 M NaCl, 0.05 M Glycine (9.5), 33 % (v/v) PEG300), e) MemGold2 B7 (0.04 M NaCl, 0.04 M Tris (8), 25 % (v/v) PEG350 MME), f) MemGold2 E1 (0.2 M Sodium acetate trihydrate, 0.1 M MES (6.5), 28 % (v/v) PEG400), g) MemGold2 E3 (0.2 M NaCl, 0.1 M HEPES (7), 29 % (v/v) PEG400), h) Optimisation of MemGold2 B6 B6 (0.03 M MgCl2, 0.1 M MES (6.5), 28 % (v/v) PEG400. Protein:Reservoir solution, 2 µL:1 µL), i) Optimisation of MemGold2 D12 C6 (0.2 M Calcium acetate, 0.1 M HEPES (7), 24 % PEG400. Protein:Reservoir solution, 1 µL:2 µL), j) Optimisation of MemGold H11 A6 (0.1 M NaCl, 0.1 M Sodium phosphate (7), 33 % (v/v) PEG300. Protein:Reservoir solution, 1 µL:1 µL)

(Appendix 4.5-4.9)) were set up, using the 4.5 $mg.mL^{-1}$ sample. This resulted in the production of more crystals in different conditions, as well as crystals that took longer to form in the previous crystallisation screens (Figure 3-27). These crystals were looped using cryoloops and cryocooled for X-ray diffraction at Diamond Light Source. Crystals were screened at beamline IO4, with a selenium edge scan being conducted to determine the optimal wavelength to use (0.9793 Å). When collecting the initial datasets, the transmission of the beam was reduced to 30 % to reduce radiation damage and improve anomalous signal. During this data collection there were three crystal conditions with crystals that diffracted to a high resolution with a CCanom greater than 0.5 (Table 3-2). These datasets all enablelled solving of the phases, resulting in the determination of Bd0427 structure, see Chapter 4.

Table 3-2: Crystal Resolution and CCanom

Crystal Condition	Resolution (Å)	CCanom
MIDASplus C9	2.3	1.0
MemGold D5	3.5	0.7
MemGold2 B7	3.4	1.0

3.2.2 Refold, Purification and Crystallisation of Bd0427 His-tagged Constructs

As well as attempting to obtain diffraction-quality crystals of Bd0427 ΔSP, two other constructs with no signal peptide (ΔSP N-His, ΔSP C-His), with N-terminal and Tev-cleavable C-terminal His-tags, respectively, were worked with. This was to maximise the opportunity to produce high-quality protein crystals of Bd0427.

3.2.2.1 Refold of Bd0427 ΔSP N-His from Inclusion Body Material

As with Bd0427 ΔSP, Bd0427 ΔSP N-His was expressed in the cytoplasm where it was packaged into inclusion bodies. The inclusion material was purified and the protein was denatured in 8 M urea to a final concentration of 0.1 mM protein. Then attempts were made,

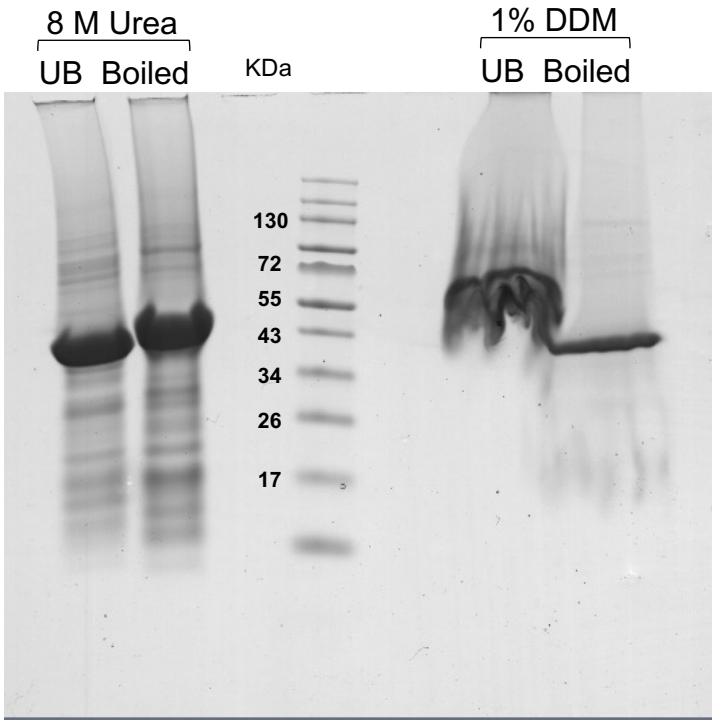


Figure 3-28: Semi-native PAGE of Refold of Bd0427 ΔSP N-His in DDM. All samples in 1% SDS loading dye; one sample was boiled at 100 °C for 10 min and the other was unboiled (UB). Protein Ladder = Color Prestained Protein Standard, Broad Range (10 250 KDa) (NEB, P7719S).

by rapid diluting using a dropwise method, to refold Bd0427 ΔSP N-His in 1 % (w/v) DDM. To determine if the construct was refolded a semi-native PAGE was run. The gel contained

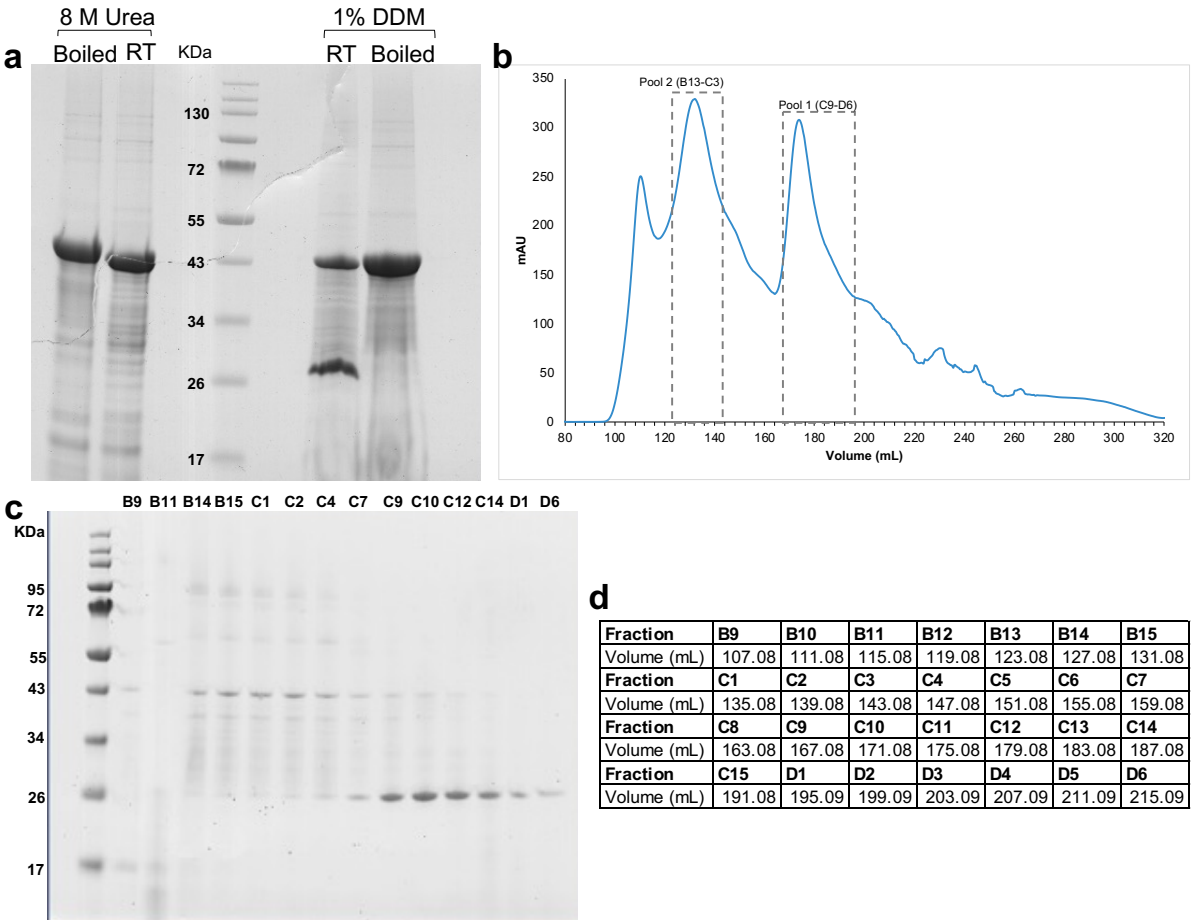
samples of the unfolded inclusion material in 8 M urea and samples of the refold attempt in DDM that had been either boiled for 10 minutes or unboiled .Bd0427 ΔSP N-His was not refolded in 1 % (w/v) DDM (Figure 3-28), as the band for the sample unboiled migrated the same distance as the bands for the boiled sample in 1 % (w/v) DDM and the samples in 8 M urea, indicating the protein was still in an unfolded state. It was concluded that the N-terminal His-tag must prevent folding of Bd0427, so no further attempts were made.

3.2.2.2 Refold, Purification and Crystallisation of Bd0427 ΔSP C-His from Inclusion Body Material

3.2.2.2.1 Refold and Purification of Bd0427 ΔSP C-His

Bd0427 ΔSP C-His was expressed in the cytoplasm for packaging into inclusions bodies. The inclusion body material was purified and added to a buffer containing 8 M urea to denature the protein before refolding in 1 % (w/v) DDM, as with Bd0427 ΔSP. Success of the refold was determined by running a semi-native PAGE (Figure 3-29a), which indicated that Bd0427 ΔSP C-His could be refolded in 1 % (w/v) DDM. As determined by the migration distance of the unboiled sample of Bd0427 in 1 % DDM, in comparison to the samples in either 8 M urea or the refold in 1 % DDM that was boiled. The refolded protein was concentrated to 8 $mg.mL^{-1}$ and purified on a HiLoad 16/60 Superdex 200 pg column (Figure 3-29b). The protein eluted from the SEC column in multiple polydisperse, and non-baseline resolved peaks. The fractions were run on a semi-native PAGE to enable pooling (Figure 3-29c). Pool 1 (C9 to D6, 167 – 199 mL) contained fractions which when ran on a semi-native PAGE produced a single band that had migrated to a molecular weight of approximately 26 KDa, corresponding to the folded protein. Fractions B13 to C3 produced the most intense band at approximately 43 KDa, the same migration distance as the unfolded protein but the fractions also resulted in two faint bands

at approximately 55 and 95 KDa, representative of oligomers so these fractions were pooled (pool 2, B13-C3, 123 – 147 mL) for crystallisation experiments.



d

Fraction	B9	B10	B11	B12	B13	B14	B15
Volume (mL)	107.08	111.08	115.08	119.08	123.08	127.08	131.08
Fraction	**C1**	**C2**	**C3**	**C4**	**C5**	**C6**	**C7**
Volume (mL)	135.08	139.08	143.08	147.08	151.08	155.08	159.08
Fraction	**C8**	**C9**	**C10**	**C11**	**C12**	**C13**	**C14**
Volume (mL)	163.08	167.08	171.08	175.08	179.08	183.08	187.08
Fraction	**C15**	**D1**	**D2**	**D3**	**D4**	**D5**	**D6**
Volume (mL)	191.08	195.09	199.09	203.09	207.09	211.09	215.09

Figure 3-29: Refold and Purification of Bd0427 ΔSP C-His in DDM. a) Semi-native PAGE of refold in DDM. All samples in 1% SDS loading dye; one sample was boiled at 100 °C for 10 min and the other was unboiled (UB). Protein Ladder = Color Prestained Protein Standard, Broad Range (10 250 KDa) (NEB, P7719S). b) Size exclusion chromatogram from HiLoad 26/60 Superdex 200 pg column eluted with tris (pH 8), 100 mM NaCl, 0.3 % (w/v) DDM. c) Semi-native PAGE of SEC peak. Protein Ladder = Color Prestained Protein Standard, Broad Range (10 250 KDa) (NEB, P7719S). d) Conversion of volume to fractions ran on the semi-native PAGE.

3.2.2.2.2 Crystal Hits of Bd0427 ΔSP C-His in DDM

Pools 1 and 2 were concentrated to 7 $mg.mL^{-1}$ and 6.5 $mg.mL^{-1}$ respectively, and initially used in MemGold (Molecular Dimensions). Crystals formed in conditions containing protein from peak 1, these crystals of Bd0427 ΔSP C-His were very small in comparison to the crystals without an affinity-tag (Figure 3-30). The crystals grown in MemGold C4 and D8 (Figure 3-30c

and d) sent for X-ray diffraction. Native and heavy atom-soaked crystals were screened, however none of the crystals diffracted.

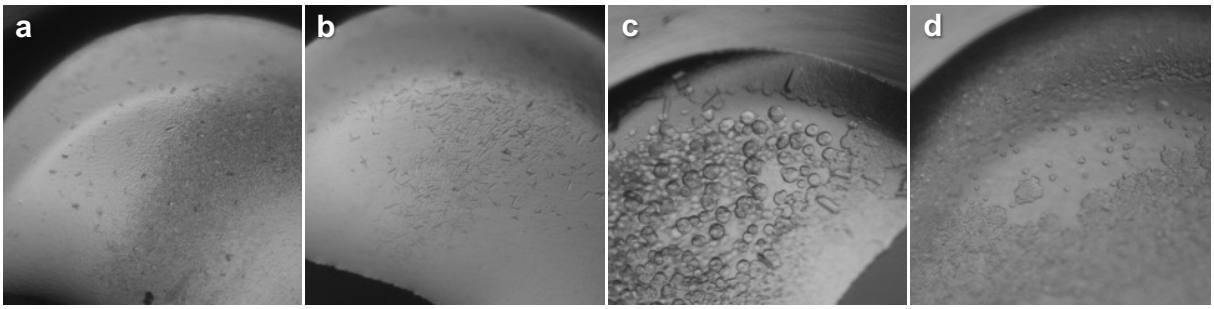


Figure 3-30: Initial Crystal Hits of Bd0427 ΔSP C-His in DDM. Crystal hits a) of Peak 2 MemGold D11 (0.2 M Sodium acetate (6.8), 8.8 % (w/v) PEG6000), b) of Peak 1 MemGold B10 (0.05 M Magnesium chloride hexahydrate, 0.1 M Sodium cacodylate (6.7), 6.6 % (w/v) PEG3350, 0.1 % (w/v) Sodium azide), c) of Peak 1 MemGold C4 (0.05 M NaCl, 0.02 M Magnesium chloride hexahydrate, 0.1 M Sodium citrate (6), 22 % (w/v) PEG400), d) of Peak 1 MemGold D8 (0.05 M Magnesium chloride hexahydrate, 0.1 M Glycine (9), 22 % (v/v) PEG400).

Following this, the pools were used in MIDASplus, which yielded more crystals (Figure 3-31). Two of these crystals, E4 and F7, along with slower forming crystals from MemGold C9 (Figure 3-31b-d), were selected for X-ray diffraction screening at Diamond Light Source. Native crystals were screened, with none diffracting.

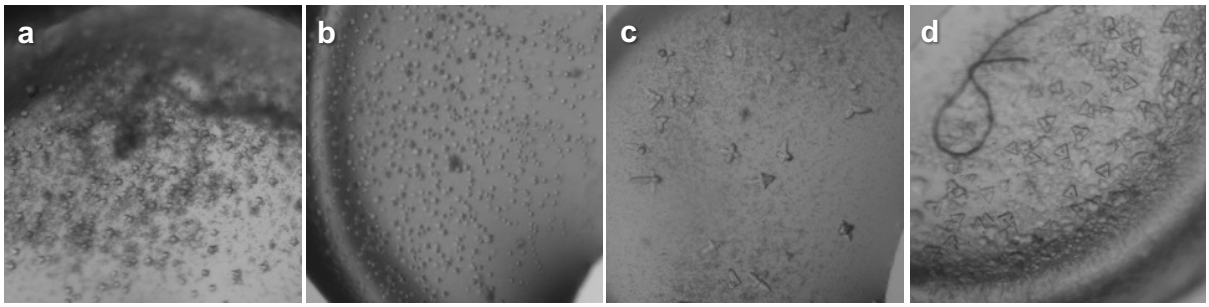
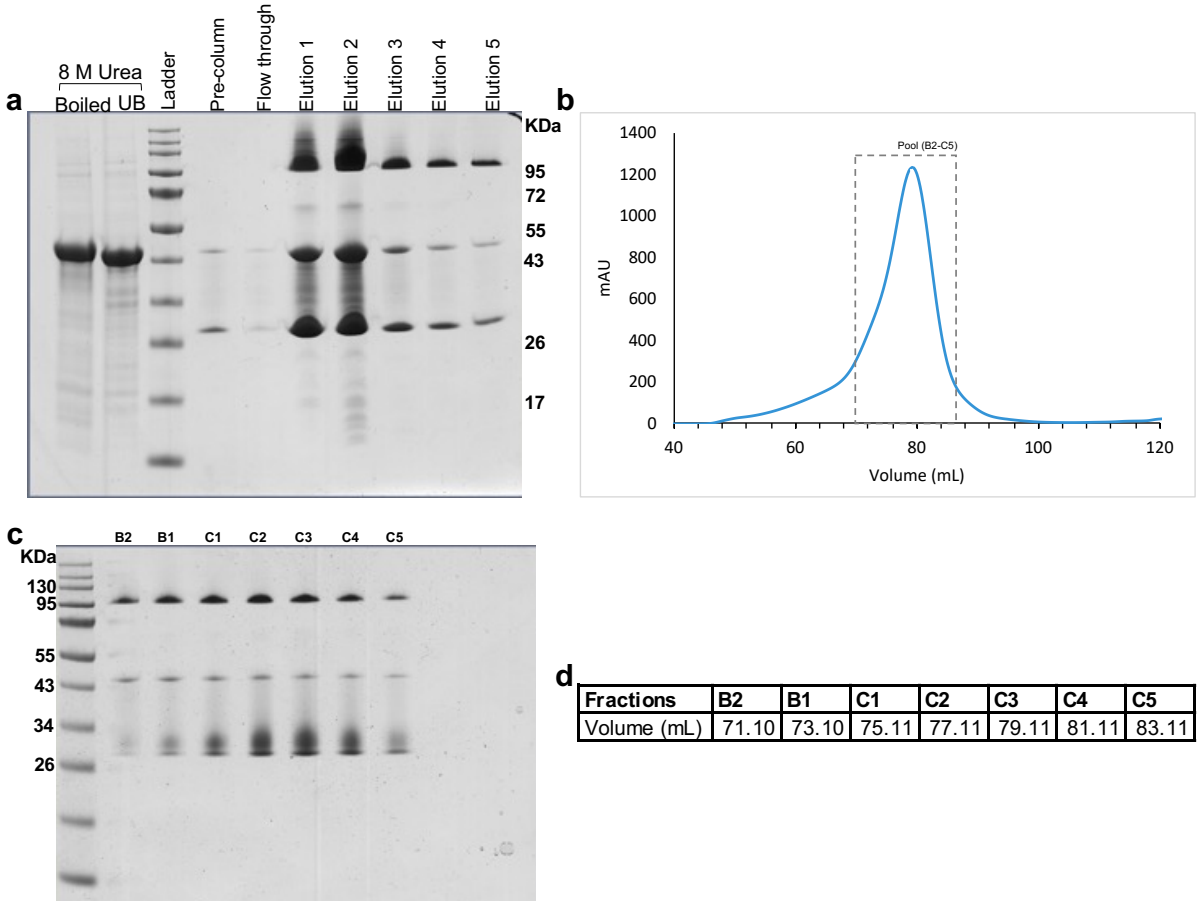


Figure 3-31: Crystal Hits of Bd0427 ΔSP C-His in DDM. Crystal hits of Peak 1 in a) MIDASplus B3 (0.1 M NaCl, 0.1 M HEPES (7), 25 % (w/v) SOKALAN PA 25 CL), b) MIDASplus E4 (0.2 Sodium thiocyanate, 0.1 M HEPES (7), 40 % (v/v) Pentaerythritol propoxylate (5/4 PO/OH)), c) MIDASplus F7 (25 % (w/v) SOKALAN PA 25 CL), d) MemGold C9 (0.1 M NaCl, 0.1 M Magnesium chloride hexahydrate, 0.1 M HEPES (8), 11 % (w/v) PEG1500).

3.2.2.2.3 Refold and Purification of Bd0427 ΔSP C-His with the Addition of Nickel Column Purification

Bd0427 ΔSP C-His was refolded in 1 % DDM as in section 3.2.6.2.1, this was then cycled onto a 1 mL HisTrap FF column with the bound protein eluted with imidazole. A semi-native PAGE was run to compare the elution fractions to the unfolded 8 M urea sample (Figure 3-32a), the elution fractions all produced three bands, one migrated the same distance as folded protein (Figure 3-31a), one the same distance as denatured unfolded protein and one migrated at a high molecular weight (~95 KDa) suggesting the formation of an oligomer, as seen in fractions of



Fractions	B2	B1	C1	C2	C3	C4	C5
Volume (mL)	71.10	73.10	75.11	77.11	79.11	81.11	83.11

Figure 3-32: Refold and Purification on HisTrap of Bd0427 ΔSP C-His in DDM. a) Semi-native PAGE of refold in DDM and HisTrap purification. All samples in 1% SDS loading dye; one urea sample was boiled at 100 °C for 10 min and the other was unboiled (UB). Protein Ladder = Color Prestained Protein Standard, Broad Range (10 250 KDa) (NEB, P7719S). b) Size exclusion chromatogram from HiLoad 16/60 Superdex 200 pg column eluted with tris (pH 8), 150 mM NaCl, 0.3 % (w/v) DDM. c) Semi-native PAGE of SEC peak. Protein Ladder = Color Prestained Protein Standard, Broad Range (10 250 KDa) (NEB, P7719S). d) Conversion of volume to fractions ran on the semi-native PAGE.

Bd0426 ΔSP in DDM. The eluted fractions were pooled and purified on a HiLoad 16/60 Superdex 200 pg column. The protein eluted as a single monodispersed peak (Figure 3-32b), unlike the construct purified only using SEC (Figure 3-29b), suggesting the HisTrap column selects for a single species. Fractions of the peak were run on a semi-native PAGE, with the samples still producing three bands that migrated the same distances as the previous semi-native PAGE (Figure 3-32a and c). These fractions (B2-C5, 71 – 85 mL) were pooled for crystal trays.

3.2.2.2.4 Crystal Hits of Bd0427 ΔSP C-His in DDM Following HisTrap Purification

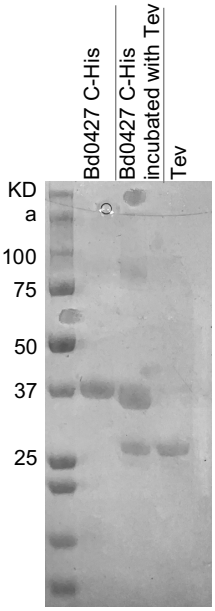


Figure 3-33: SDS-PAGE of Overnight Cleavage of the C-terminal His-tag of Bd0427 ΔSP C-His by Tev. Bd0427 (9 mg.mL^{-1}), Tev (0.9 mg.mL^{-1}), Protein ladder = Precision Plus Protein All Blue Standard (Bio Rad).

The pooled protein was concentrated to 9 mg.mL^{-1} and 4.5 mg.mL^{-1} before Tev was used to determine if the C-terminal His-tag could be cleaved (Figure 3-33). As the sample incubated overnight with Tev runs further down the gel in comparison to the protein sample without Tev, it was concluded Tev could fully cleave the His-tag. Crystal Trays (MemGold, MemGold2, MIDASplus) were set up with and without Tev (0.9 mg.mL^{-1}) at 9 mg.mL^{-1} and without Tev at 4.5 mg.mL^{-1}. The crystal screens produced a range of crystal hits, a selection is shown in Figure 3-34. Crystals from three of these conditions (Figure 3-34b, d, e) were sent to Diamond Light Source for screening as native crystals and with tetrachloroplatinate-soaking. The crystals were screened on the IO4 beamline at 0.9794 Å, and only one of the crystals diffracted from MemGold2 G1b, which diffracted to 10 Å.

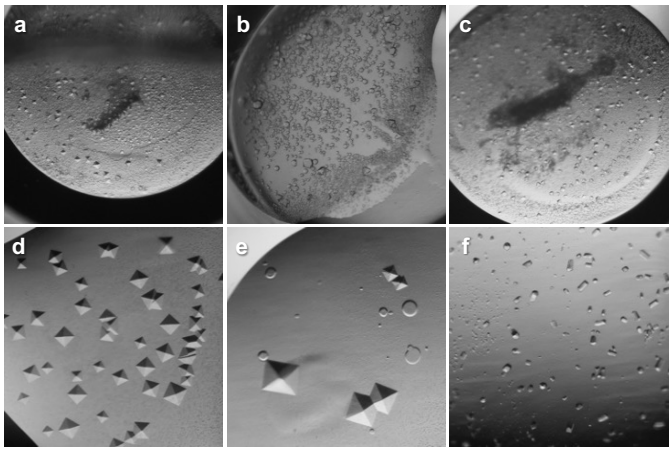


Figure 3-34: Crystal Hits of Bd0427 ΔSP C-His in DDM following HisTrap Purification. Crystals hits in a) MemGold C2b (0.1 M Magnesium chloride hexahydrate, 0.1 M Tris (7.5), 22 % (v/v) PEG400). b) MemGold D8b (0.05 M Magnesium chloride hexahydrate, 0.1 M Glycine (9), 22 % (v/v) PEG400). c) MemGold H4b (0.2 M Lithium sulfate, 0.1 M Sodium citrate (3.5), 28 % (v/v) PEG400). d) MemGold2 F7b (0.37 M Potassium nitrate, 0.1 M MES (6.5), 22 % (v/v) PEG400). e) MemGold2 G1b (0.5 M Potassium chloride, 0.05 M HEPES (6.5), 20 % (v/v) PEG400). f) MIDASplus B10a (30 % (v/v) Jeffamine M-600, 10 % (v/v) DMSO). Drop 'a' 9 $mg.mL^{-1}$, drop 'b' 9 mg.mL-1 with 0.9 $mg.mL^{-1}$ Tev, drop 'c' 4.5 $mg.mL^{-1}$.

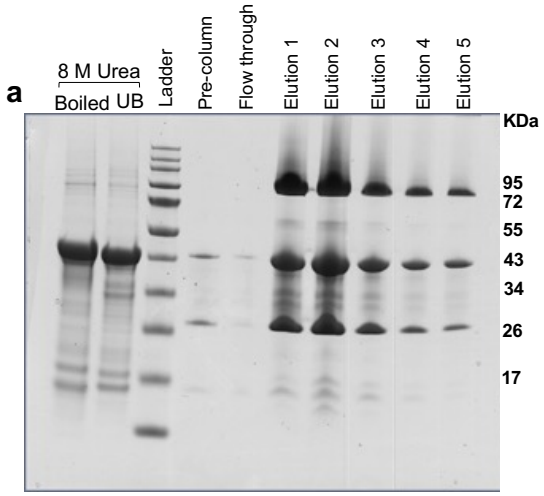
3.2.2.3 Refold, Purification and Crystallisation of Se-Met Bd0427 ΔSP C-His

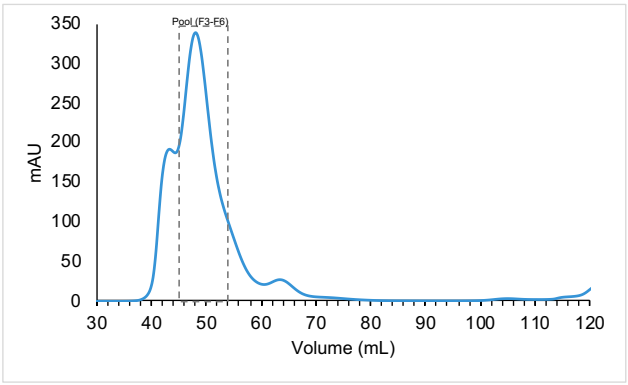
3.2.2.3.1 Refold and Purification of Se-Met Bd0427 ΔSP C-His in DDM including HisTrap Purification

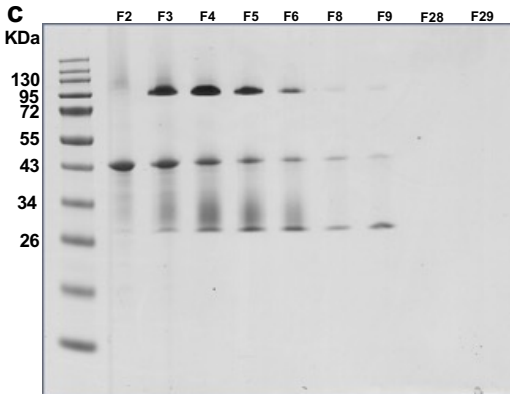
As the phases need to be solved in order to determine the structure of Bd0427, Bd0427 ΔSP C-His was expressed in the presence of selenomethionine. The HisTrap-SEC purification produced diffraction-quality crystal, so this method was chosen.

The protein refolded as previously seen in section 3.2.2.2.3 (Figure 3-32a) before binding a HisTrap FF column. The protein was eluted with imidazole with each fraction resulting in three bands on a semi-native PAGE (Figure 3-35a) which migrated as seen previously with native Bd0427 ΔSP C-His (Figure 3-32a). Bd0427 was then purified on a HiLoad 16/60 Superdex200 pg column. This produced different results to section 3.2.2.2.3 (Figure 3-32b), with multiple non-baseline resolved peaks all eluting from the column earlier than the previously observed single peak (Figure 3-35b). The peak fractions were run on a semi-native PAGE (Figure 3-35c) resulting in observation of three bands for fractions F3 – F6, that migrated the same distance as

seen previously (Figure 3-32c). These fractions were pooled (45 – 54 mL) ready for crystallisation.







d

Fractions	F2	F3	F4	F5	F6
Volume (mL)	42.15	45.15	48.15	51.15	54.15
Fractions	**F7**	**F8**	**F9**	**F28**	**F29**
Volume (mL)	57.15	60.15	63.15	120.14	123.14

Figure 3-35: Refold and Purification on HisTrap of Se-Met Bd0427 ΔSP C-His in DDM. a) Semi-native PAGE of refold in DDM and HisTrap Purification. All samples in 1% SDS loading dye; one urea sample was boiled at 100 °C for 10 min and the other was unboiled (UB). Protein Ladder = Color Prestained Protein Standard, Broad Range (10 250 KDa) (NEB, P7719S). b) Size exclusion chromatogram from HiLoad 16/60 Superdex 200 pg column eluted with tris (pH 8), 150 mM NaCl, 3 mM β-mercaptoethanol and 0.3 % (w/v) DDM. c) Semi-native PAGE of SEC peak. Protein Ladder = Color Prestained Protein Standard, Broad Range (10 250 KDa) (NEB, P7719S). d) Conversion of volume to fractions ran on the semi-native PAGE.

3.2.2.3.2 Crystal Hits of Bd0427 Construct 3 Selenium-methionine in DDM following HisTrap Purification

As in Section 3.2.2.2.4 the pooled protein was concentrated to 9 mg.mL^{-1} and 4.5 mg.mL^{-1} and used in MemGold, MemGold 2 and MIDASplus (Molecular Dimensions). The screens were set up with and without Tev for the protein at 9 mg.mL^{-1} and without Tev for the protein at 4.5 mg.mL^{-1}. These conditions initially produced four conditions with crystals (Figure 3-36), these

were sent to Diamond Light Source and screened at 0.9794 Å at the IO4 beamline. These initial crystals did not diffract.

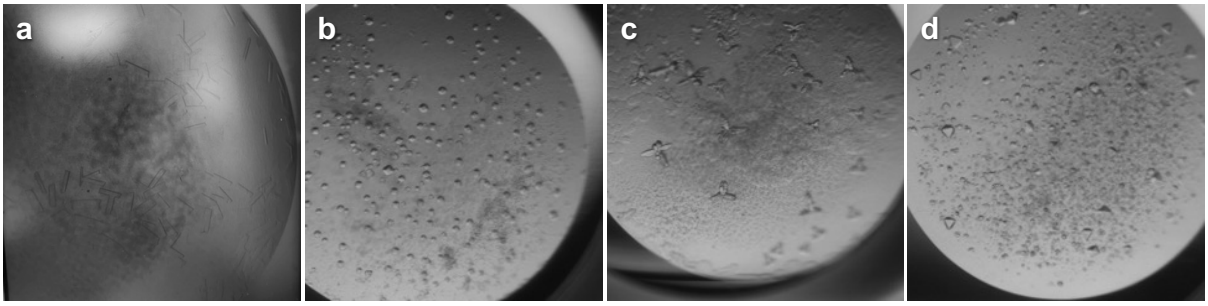


Figure 3-36: Crystal Hits of Se-Met Bd0427 ΔSP C-His in DDM following HisTrap Purification. Crystals Hits in a&b) MemGold F7b & F7c (0.02 M Potassium nitrate, 0.03 M Potassium citrate (6.5), 7.7 % (w/v) PEG4000), c) MemGold F10c (0.1 M Calcium chloride dihydrate, 0.1 M Tris (6.5), 13 % (w/v) PEG 2000 MME), d) MIDASplus E3c (0.3 M KCl, 0.1 M Glycine (9.5), 20 % (v/v) Pentaerythritol ethoxylate (15/4 EO/OH)). Drop 'a' 9 mg.mL^{-1}, drop 'b' 9 mg.mL-1 with 0.9 mg.mL^{-1} Tev, drop 'c' 4.5 mg.mL^{-1}.

3.2.2.4 Bd0427 ΔSP C-His On-column Detergent Exchange

To enable screening of other detergents for purification of Bd0427, an on-column detergent exchange method was designed. This was trialled with C_8E_4 and βOG. If this method worked this was to be used in a high-through put method with the Membrane Protein Laboratory, Oxford, to screen a wide range of detergents. Due to this being initially tested in June 2021 and the structure of Bd0427 construct 1 being determined in June 2021, the detergent screen was never conducted.

3.2.2.4.1 Bd0427 ΔSP C-His On-column Exchange into C_8E_4

Bd0427 ΔSP C-His was refolded as in the previous sections (3.2.2.2 and 3.2.2.4) then bound to a 1 mL HisTrap column. The protein was then washed in buffer containing 0.3 % (w/v) DDM, then 0.15 % (w/v) DDM with 0.15 % (v/v) C_8E_4 and finally with 0.3 % (v/v) C_8E_4, before elution with detergent buffer with 300 mM Imidazole. Samples from each of these wash and elution steps were ran on a semi-native PAGE (Figure 3-37) to determine where the protein had eluted. As seen by the gel there are no protein bands in any of the wash or elution samples,

suggesting the protein had aggregated. The nickel on the HisTrap column was then stripped using 50 mM EDTA, followed by washing with H_2O and 0.5 M NaOH to see if the protein had aggregated on the column. The lanes with samples from the H_2O and NaOH wash both contain protein bands suggesting the protein aggregated and C_8E_4 is not a suitable detergent for Bd0427 ΔSP C-His.

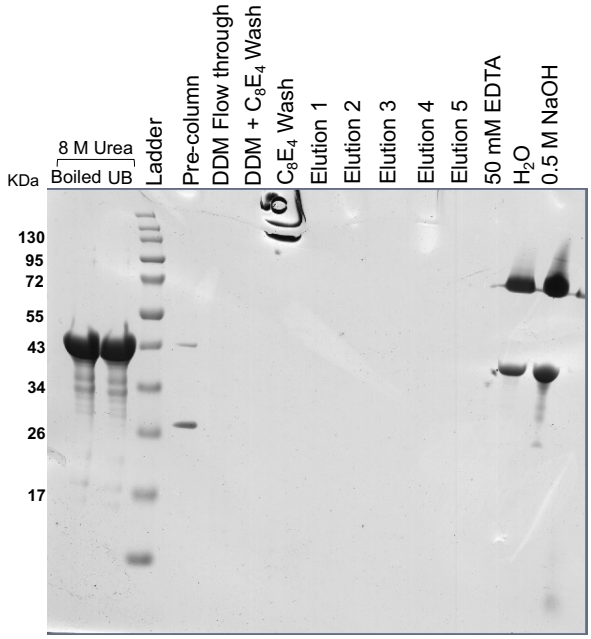
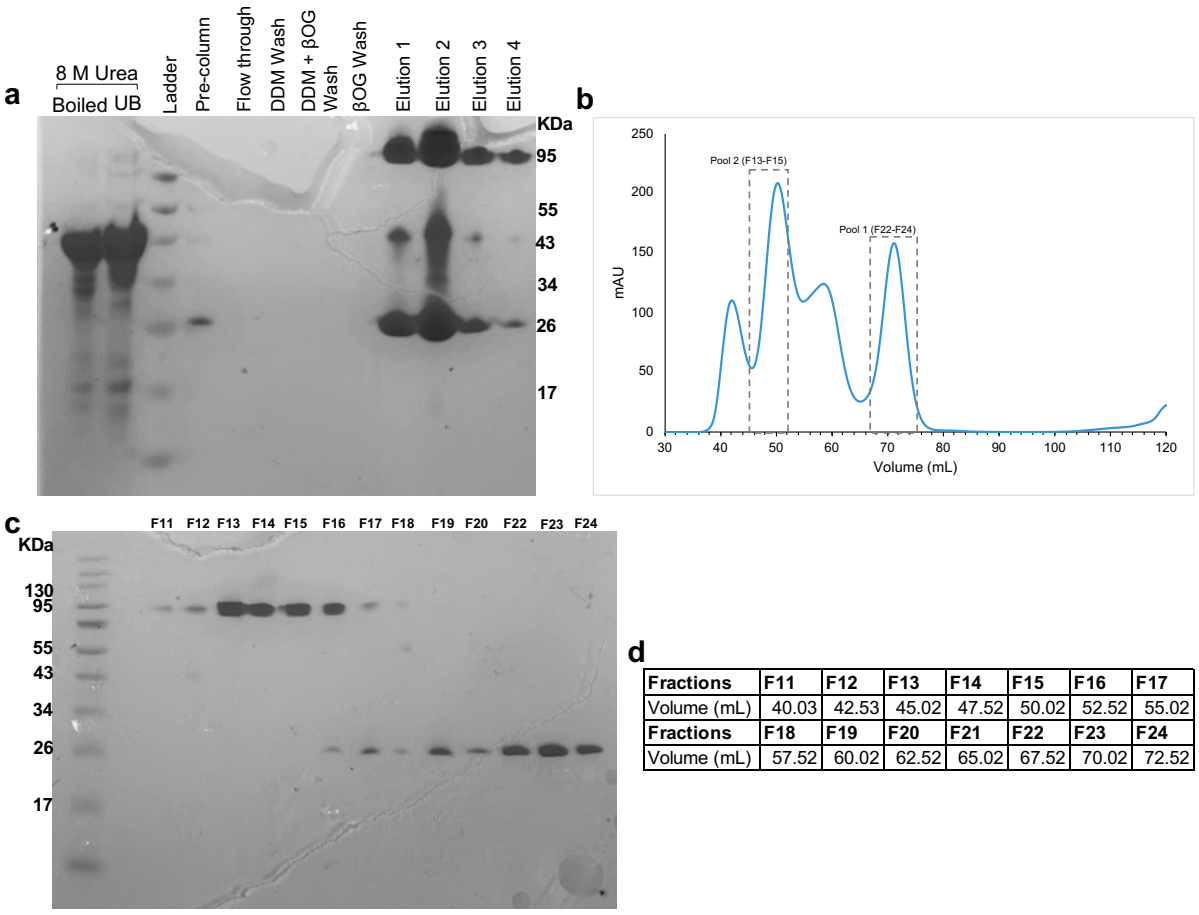


Figure 3-37: Semi-native PAGE of Bd0427 ΔSP C-His Detergent Exchange into C8E4. Samples taken from every step of the detergent exchange, including samples from the cleaning of the column (50 mM EDTA, H_2O & 0.5 M NaOH). A control sample of unfolded inclusion body in 8 M urea were also ran with one urea sample was boiled at 100 °C for 10 min and the other was unboiled (UB). All samples in 1% SDS loading dye. Protein Ladder = Color Prestained Protein Standard, Broad Range (10 250 KDa) (NEB, P7719S).

3.2.2.4.2 Bd0427 ΔSP C-His On-column exchange into βOG

For this exchange into 50 mM βOG, the protein was refolded in 1 % (w/v) DDM as in previous sections, before binding to a 1 mL HisTrap column then washed with 0.3 % (w/v) DDM, 0.15 % DDM with 25 mM βOG and finally with 50 mM βOG before elution of the protein. A semi-native PAGE was run of the samples from each step of the detergent exchange (Figure 3-38a) showing protein bands in the elution samples. The migration of these bands corresponds to the migrations of the bands previously seen for this construct. The elutions were pooled and purified on a HiLoad 16/60 Superdex 200 pg column, to enable pooling. The SEC chromatogram (Figure 3-38b) showed four overlapping non-baseline resolved peaks, the fractions for all but the first peak (~40 mL) were run on a semi-native PAGE (Figure 3-38c).

The protein migrated two distances: one corresponding to the folded monomer (~26 KDa) and one corresponding to the higher molecular weight band (~95 KDa) on the semi-native PAGE of the HisTrap column (Figure 3-38a). From these results two pools were created Pool 1 (67 – 75 mL), and 2 (45 – 52 mL).



Fractions	F11	F12	F13	F14	F15	F16	F17
Volume (mL)	40.03	42.53	45.02	47.52	50.02	52.52	55.02
Fractions	**F18**	**F19**	**F20**	**F21**	**F22**	**F23**	**F24**
Volume (mL)	57.52	60.02	62.52	65.02	67.52	70.02	72.52

Figure 3-38: Refold, On-column Detergent Exchange and Purification of Bd0427 ΔSP C-His in βOG. a) Semi-native PAGE of refold in DDM and HisTrap detergent exchange into βOG. All samples in 1% SDS loading dye; one urea sample was boiled at 100 °C for 10 min and the other was unboiled (UB). Protein Ladder = Color Prestained Protein Standard, Broad Range (10 250 KDa) (NEB, P7719S). b) Size exclusion chromatogram from HiLoad 16/60 Superdex200 pg column eluted with tris (pH 8), 150 mM NaCl, 50 mM βOG. c) Semi-native PAGE of SEC peak. Protein Ladder = Color Prestained Protein Standard, Broad Range (10 250 KDa) (NEB, P7719S). d) Conversion of volume to fractions ran on the semi-native PAGE.

3.2.2.4.3 Crystal Hits of Bd0427 Exchanged into βOG

As with all crystal screens set up of Bd0427 (Figure 3-38b) involving a HisTrap, Bd0427 was concentrated to 9 mg.mL^{-1} and 4.5 mg.mL^{-1} for use in MemGold, MemGold2, MIDASplus

(Molecular Dimensions). Pool 1 was initially used in crystal screens at 9 $mg.mL^{-1}$ with and without Tev and at 4.5 $mg.mL^{-1}$. In these screens there were a range of crystal hits (Figure 3-39), these were not sent to Diamond as they were observed at the week of the X-ray crystallography experiments in Diamond that led to the structure of Bd0427.

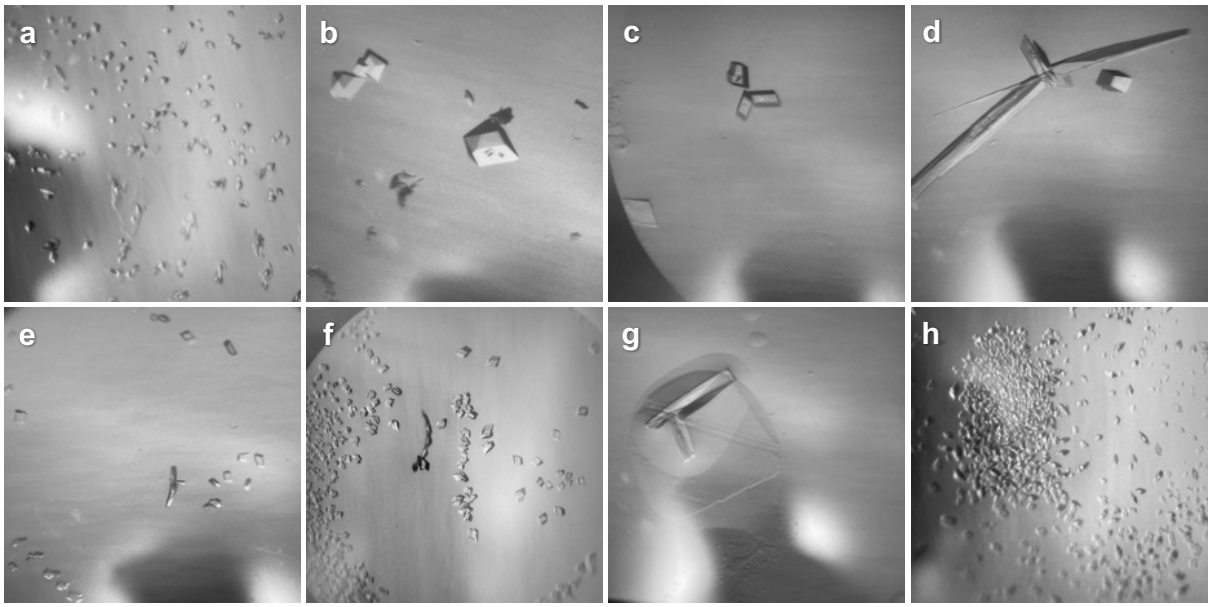


Figure 3-39: Initial Crystal Hits of Bd0427 ΔSP C-His Exchanged into βOG. Crystals in a) MemGold C8a (0.1 M NaCl, 0.1 M Magnesium chloride hexahydrate, 0.02 M Tris (7.5), 11 % PEG (w/v) PEG1500), b) MemGold F4c (0.2 M NaCl, 0.1 M HEPES (7), 22 % (v/v) PEG500 MME), c&d) MemGold G9b & G9c (0.05 M NaCl, 0.1 M Sodium citrate (5.5), 26 % (v/v) PEG400), e) MemGold G10c (0.1 M Lithium sulfate, 0.1 M Glycine (9.3), 30 % (v/v) PEG400), f) MemGold2 B7c (0.04 M NaCl, 0.04 M Tris (8), 25 % (v/v) PEG350 MME), g) MemGold2 D8c (0.2 M Magnesium formate dihydrate, 0.05 M Tris (8), 19 % (w/v) PEG3350), h) MIDASplus D2a (0.1 M HEPES (7.5), 30 % (v/v) Pentaerythritol ethoxylate (15/4 EO/OH), 6 % (w/v) Polyvinylpyrrolidone).

3.3 Summary

In this Chapter three constructs of Bd0427 were investigated. Two of these (Bd0427 ΔSP and Bd0427 ΔSP C-His) were successfully refolded and purified, with crystallisation trials resulting in protein crystals. The Bd0427 ΔSP construct produced the best diffraction-quality crystals, with protein refolded and purified in β-OG producing crystals that diffracted to the highest resolution. To solve the phases of the datasets, the protein was expressed in the presence of selenomethionine, as heavy atom soaking of Bd0427 crystals did not produce a strong enough anomalous signal.

CHAPTER FOUR: STRUCTURE OF BD0427

4.1 Foreword

Multiple diffraction-quality crystals were produced of Bd0427 ΔSP. The crystal of Se-met labelled Bd0427 ΔSP in 50 mM β-OG from MIDASplus C9 (0.1 M NaCl, 25 % (v/v) pentaerythritol propoxylate (5/4 PO/OH), 10 % (v/v) DMSO) diffracted to the highest resolution and enabled solving of the phase problem. This chapter will focus on the data collection, model building and structural analysis of this crystal structure (crystal form A), as well as the data collection and model building of three additional crystal forms (B, C, D). These forms are: Bd0427 ΔSP in 50 mM β-OG from MemGold G12 (1 mM zinc sulfate heptahydrate, 50 mM HEPES (pH 7.8), 28 % (v/v) PEG 600) and Se-met labelled Bd0427 ΔSP in 50 mM β-OG from MemGold D5 (50 mM magnesium acetate tetrahydrate, 50 mM sodium acetate (pH 5.4), 24 % (v/v) PEG 400) and optimisation tray 1 of MemGold-G12 C6 (1 mM zinc sulfate heptahydrate, 50 mM HEPES (pH 7.8), 28 % (v/v) PEG 600), respectively.

4.1 Results

4.1.1 Data Collection and Model Building

Data were collected at Diamond Light Source synchrotron in Oxford, using the remote access ISPyB interface. Diffraction data were processed using ISPyB's automatic pipelines. AIMLESS was used for data that needed to be scaled or merged.

4.1.1.1 Bd0427 in β-OG, Crystal Form A

The dataset used to solve the structure of Bd0427 was collected at I04 Beamline at the selenium edge of 0.9794 Å (E(Peak): 12659.1eV), determined using a fluorescence scan (Figure 4-1)

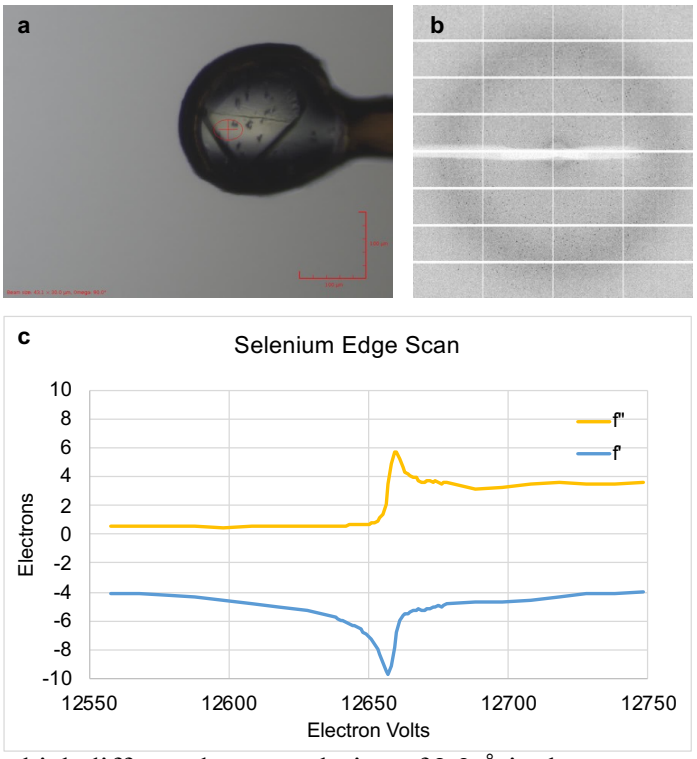


Figure 4-1: Crystal and Data Collection of Bd0427 in β-OG, Crystal Form A a) Looped crystal in β-OG. b) Diffraction pattern of crystal. c) Fluorescence Edge Scan to determine Selenium edge.

which diffracted to a resolution of 2.3 Å in the space group C 2 2 2_1. An adjusted high-resolution cutoff of 2.8 Å was applied to the data, due to incompleteness, using AIMLESS and the dataset was fed into the CRANK2 pipeline in the CCP4 cloud. The Matthews Probability calculator predicted that there would be 9 copies of Bd0427 in the ASU, which would result in a total of 81 Se sites. CRANK2 found 113 sites, with 40 having at least 25 % occupancy, resulting in a phased map with a figure of merit (FOM) of 0.318. These phases were improved using density modification, which resulted in a FOM of 0.667. An initial model was built using the CCP4 cloud automated pipeline (section 2.6.4.1), resulting in a model with 1848 residues, and following macromolecule reciprocal space refinement, using REFMAC5 as part of the pipeline, the updated model contained 2026 residues. This initial model was completed by iterative cycles of refinement in Phenix Refine and manual building in Coot, which resulted in a final model of 1665 residues, 155 water molecules and 10 ligands, with one pentamer in the asymmetric unit. A complete summary of the data collection and processing statistics are summarised in Table 4-1.

Table 4-1: Summary of Data Collection and Processing Statistics of Bd0427 Crystal Form A

Data collection of Bd0427 Crystal Form A					
Diffraction Source			Diamond Light Source		
Beamline/wavelength (Å)			IO4/0.9794		
Total rotation range (˚)			360		
Images collected			3600		
Exposure time per image			0.040s		
Transmission (%)			100		
Space group			C 2 2 2_1		
Cell dimensions					
A (Å)	B (Å)	C (Å)	175.4	191.4	184.3
α (˚)	β (˚)	γ (˚)	90	90	90
Shell			Overall	Outer shell	
Resolution Range (Å)			63.5 – 2.8	2.9 – 2.8	
Total number of reflections			1061967	106386	
Total number of unique reflections			76317	7548	
R_{merge} (within I+/I-)			0.092	0.319	
R_{meas} (within I+/I-)			0.096	0.331	
$R_{p.i.m.}$ (within I+/I-)			0.026	0.088	
I/σ(I)			25.68	4.84	
CC Half			0.993	0.982	
CC*			0.998	0.995	
Multiplicity			13.9	14.1	
Completeness (%)			99.89	99.96	
Wilson B-factor			50.76		
Refinement statistics					
Reflections used in refinement			76240	7545	
Reflections used for R-free			3738	352	
R-work			0.2466	0.3440	
R-free			0.2535	0.3609	
CC(work)			0.895	0.822	
CC(free)			0.831	0.805	
Protein residues			1665		
RMS(bonds) (Å)			0.009		
RMS(angles) (˚)			1.45		
Average B-factors ($Å^2$)			31.57		
Ramachandran favoured (%)			95.11		
Ramachandran allowed (%)			4.47		
Ramachandran outliers (%)			0.42		
Rotamer outliers (%)			0.45		
Clashscore			3.13		

The final electron density is readily interpretable, with well-defined density for the backbone and most sidechains (Figure 4-2). This enabled building of the protein chain in its entirety for each chain of the pentamer.

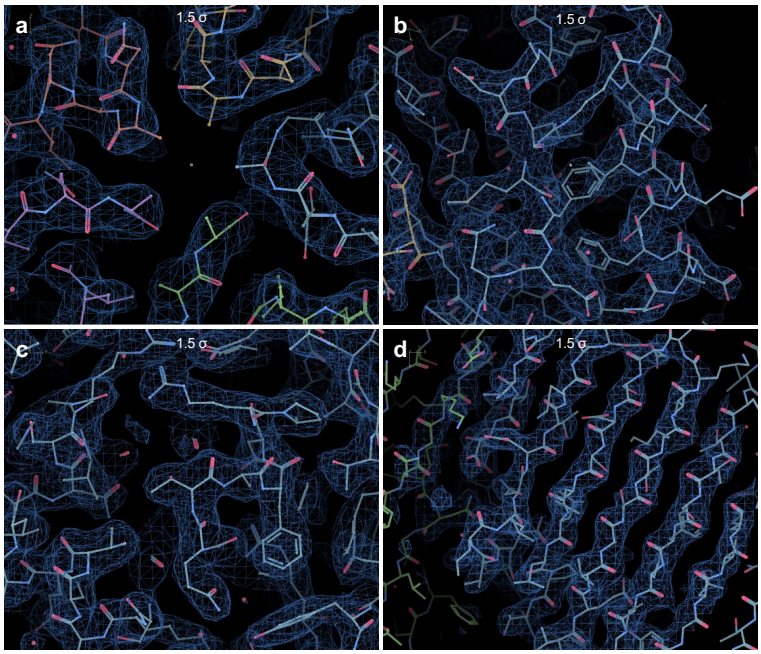


Figure 4-2: 2FO-FC Electron Density Map of Bd0427 in β-OG, Crystal Form A. a) 5-fold axis of pentamer. b) Helix of Loop 2. c) N-terminal plug. d) β-strands of β-barrel. Sigma level 1.5.

4.1.1.2 Bd0427 in β-OG, Crystal Forms B, C & D

Datasets were collected for the three crystal forms B, C and D on Beamline iO4. Data for crystal form B was collected at the zinc edge, 1.25 Å and data for crystal forms C and D were collected near the selenium edge at 0.9793 Å and 0.9792 Å respectively. None of these datasets had a high enough anomalous signal to solve using heavy atom phasing, therefore phases were calculated via molecular replacement, with Phaser in the Phenix program suite. Molecular replacement used the initial model generated from crystal form A, with five copies of Bd0427 in the ASU.

Phases for crystal form B (resolution 3.3 Å) were solved with a TFZ of 71.0 and LLG of 9473. A final model was then built in Coot and PDBredo with cycles of refinement in Phenix Refine. The map enabled building the protein in its entirety with five copies of Bd0427 in the ASU,

including eight Zn atoms, six waters and one SO_4. The electron density is of good quality, see Figure 4-3a where much of the backbone and most sidechains can be modelled with confidence.

Crystal form C (resolution 3.5 Å) was solved in a similar manner to B, resulting in a TFZ of 34.2 and LLG of 2857. Due to the (comparatively) poor resolution, only an initial model was generated, built in Coot and PDBredo with cycles of refinement in Phenix Refine. The model was built with 1655 residues, in five chains in the ASU; residues 260, 261 could not be resolved in chains B, C, D and E and residue 262 could not be resolved in chains C and D due to (presumed) flexibility of the loop. The electron density is of poor quality - the backbone is visible for most of the protein excluding the residues, stated above, and there is almost no density for side chains (Figure 4-3b).

The final crystal form, form D (resolution 3.9 Å), was solved with a TFZ of 47.2 and LLG of 2353. The electron density for this crystal form was poor, and it was hard to build the model using Coot, see Figure 4-3c. The initial model has 1665 residues, with five copies of Bd0427 in the ASU.

The statistics for data collection and refinement for all three crystal forms are provided in Table 4-2.

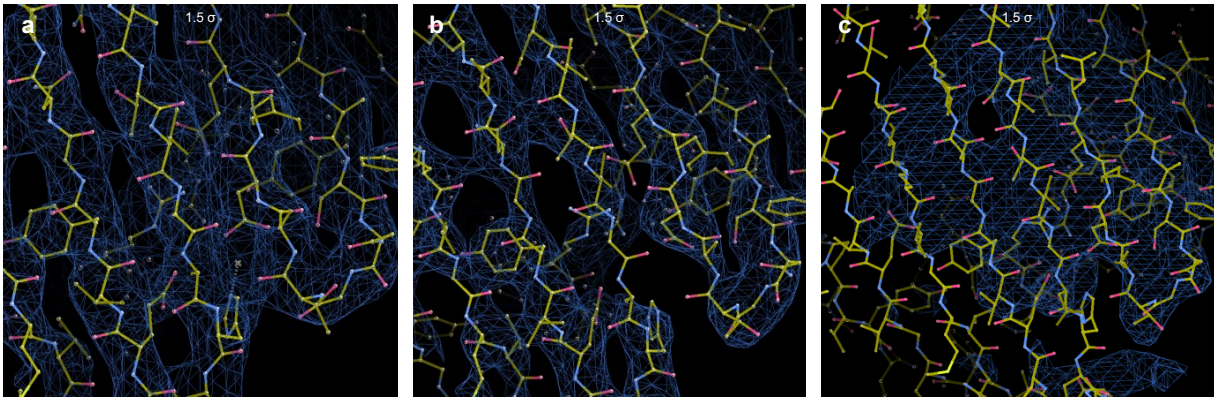


Figure 4-3: 2FO-FC Electron Density Map of Bd0427 Crystal Forms B (a), C (b) and D (c), sigma level 1.5.

Table 4-2: Summary of Data Collection and Processing Statistics of Bd0427 Crystal Form B, C and D

Crystal Form	**B**		**C**		**D**	
Diffraction Source	Diamond Light Source					
Beamline/wavelength (Å)	IO4/1.25		IO4/0.9793		IO4/0.9792	
Total rotation range (˚)	360		360		360	
Images collected	3600		3600		3600	
Exposure time per image	0.040s		0.040s		0.040s	
Transmission (%)	100		24.98		100	
Space group	P 2 2_1 2_1		C 1 2 1		C 2 2 2_1	
Cell dimensions						
A (Å) / B (Å) / C (Å)	117.8	172.6 / 174.0	177.1	189.3 / 182.3	132.6	215.3 / 199.1
α (˚) / β (˚) / γ (˚)	90	90 / 90	90	90 / 90	90	90 / 90
Shell	Overall	Outer shell	Overall	Outer shell	Overall	Outer shell
Resolution Range (Å)	29.4-3.3	3.41-3.3	79.81-3.4	3.6-3.4	56.5-3.8	3.9-3.8
Total number of reflections	692947	51212	563937	54332	390883	37603
Total number of unique reflections	54363	5303	80091	7729	28342	2796
R_{merge} (within I+/I-)	0.399	3.86	0.316	2.28	0.233	4.30
R_{meas} (within I+/I-)	0.416	4.09	0.342	2.461	0.243	4.47
$R_{p.i.m.}$ (within I+/I-)	0.116	1.31	0.129	0.927	0.066	1.21
I/σ(I)	3.74	0.30	3.67	0.31	5.33	0.32
CC Half	0.999	0.395	0.995	0.376	1	0.357
CC*	1	0.752	0.999	0.740	1	0.725
Multiplicity	12.7	9.7	7.0	7.0	13.8	13.4
Completeness (%)	93.98	45.74	97.42	81.46	94.19	46.49
Wilson B-factor	90.36		120.36		171.00	
Refinement statistics						
Reflections used in refinement	51240	2453	78307	6525	26708	1300
Reflections used for R-free	2555	131	3892	323	1368	63
R-work	0.2587	0.4335	0.4168	0.3941	0.3523	0.4576
R-free	0.2998	0.4409	0.4256	0.3953	0.4063	0.5488
CC(work)	0.770	0.750	0.620	0.307	0.701	0.250
CC(free)	0.487	0.752	0.630	0.151	0.582	0.397
Protein residues	1665		1655		1665	
RMS(bonds) (Å)	0.003		0.003		0.004	
RMS(angles) (˚)	0.71		0.79		0.83	
Average B-factors ($Å^2$)	93.99		116.45		190.95	
Ramachandran favoured (%)	95.05		88.58		90.63	
Ramachandran allowed (%)	4.05		9.84		7.55	
Ramachandran outliers (%)	0.91		1.59		1.81	
Rotamer outliers (%)	0.91		0.00		0.00	
Clashscore	4.79		11.25		7.68	

4.1.2 Crystal Packing

In all four crystal forms the homopentamer is packed in the unit cell as a crystallographic dimer of pentamers, which is formed by a bottom (periplasmic side) to bottom interaction, see Figure 4-4. These dimers of pentamers form crystal contacts with each other through the external face of the top loops (extracellular side). In crystal forms A and D there are 4 dimers of pentamers in each unit cell and in crystal forms B and C there are 2 dimers of pentamers in each unit cell.

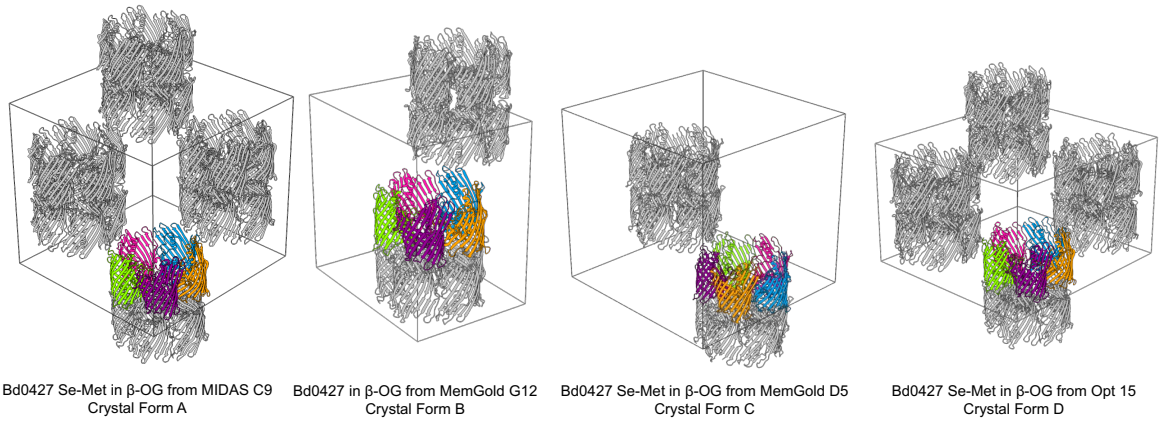


Figure 4-4: Schematic showing the Unit Cell of each Crystal Form.

4.1.3 Overall Fold of Bd0427

The crystal structure is shown in Figure 4-5, covering all 333 residues of the protein construct (this construct excludes the 20-residue signal peptide) forming a homopentamer. Crystal forms A-D all contain homopentamers; the description of the protein fold in this section will thus focus on form A, which diffracted to the highest resolution.

Bd0427 is a 16-stranded anti-parallel β-barrel with an N-terminal plug (residues 21-52), that completely blocks the barrel lumen (Figure 4-6a). The β-barrel itself forms a cylinder lumen that is flattened at the periplasmic ends of β-strands 12 – 14, resulting in the barrel lumen resembling the shape of a tear drop when viewed as a cartoon diagram from bottom (periplasmic

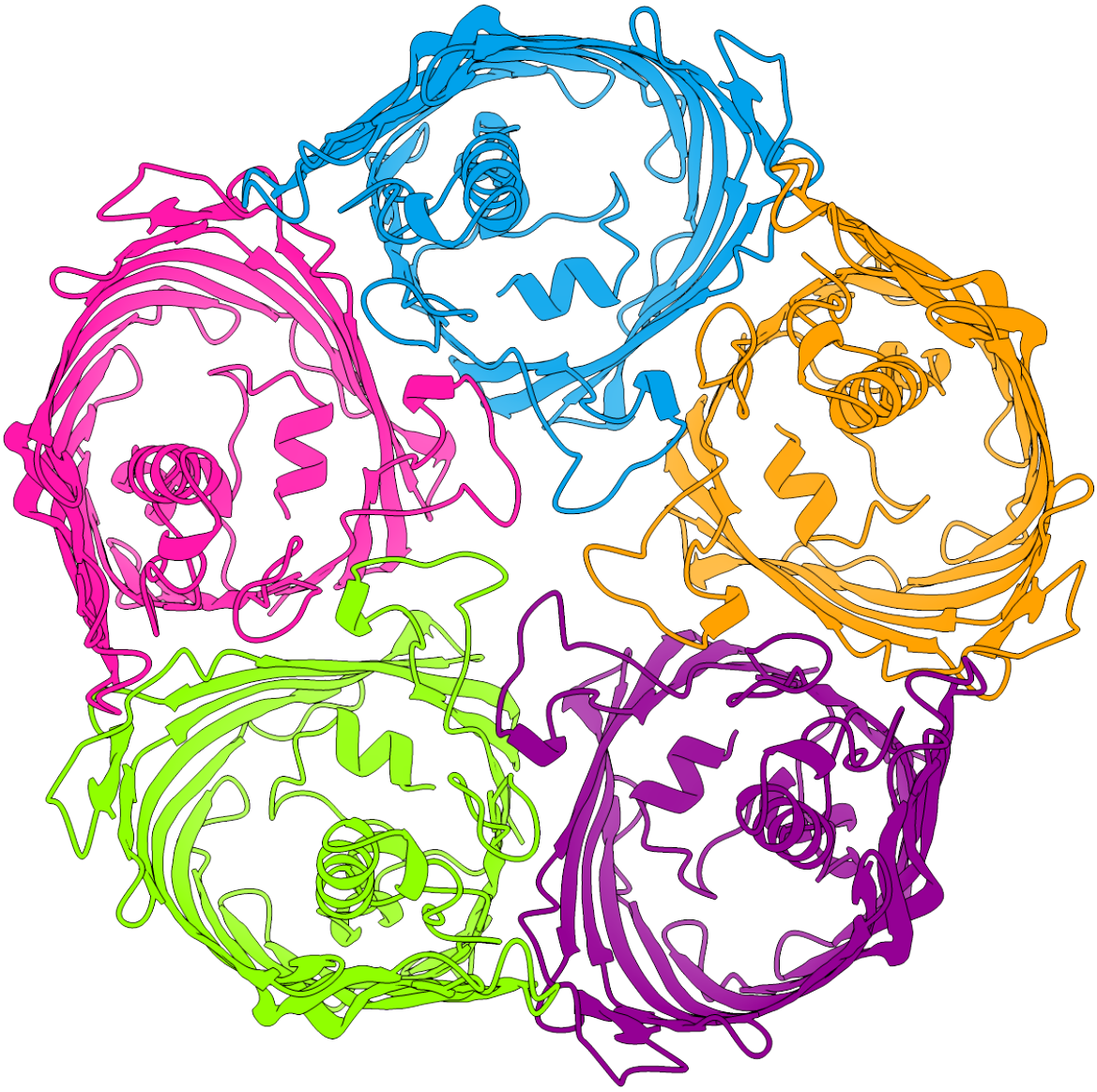


Figure 4-5: Overall Fold of Bd0427 Homopentamer.

side) looking into the pore. The height of the barrel is radically asymmetric, with the barrel wall (strand length) on the outer face of the pentamer (β-strands 5-13) having a greater height than the barrel wall that forms the central region of the pentamer (β-strands 1-4 & 14-16) (Figure 4-6a). The outer edge of the barrel is approximately 55 Å tall, with the shorter side of the barrel in the central region of the pentamer measuring less than half of this height at approximately 20 Å. The diameter of the barrel measures approximately 30 Å (C^{α}-C^{α} distance between opposite strands). Each β-strand in relation to the barrel axis has a tilt of 30-45˚. The shear

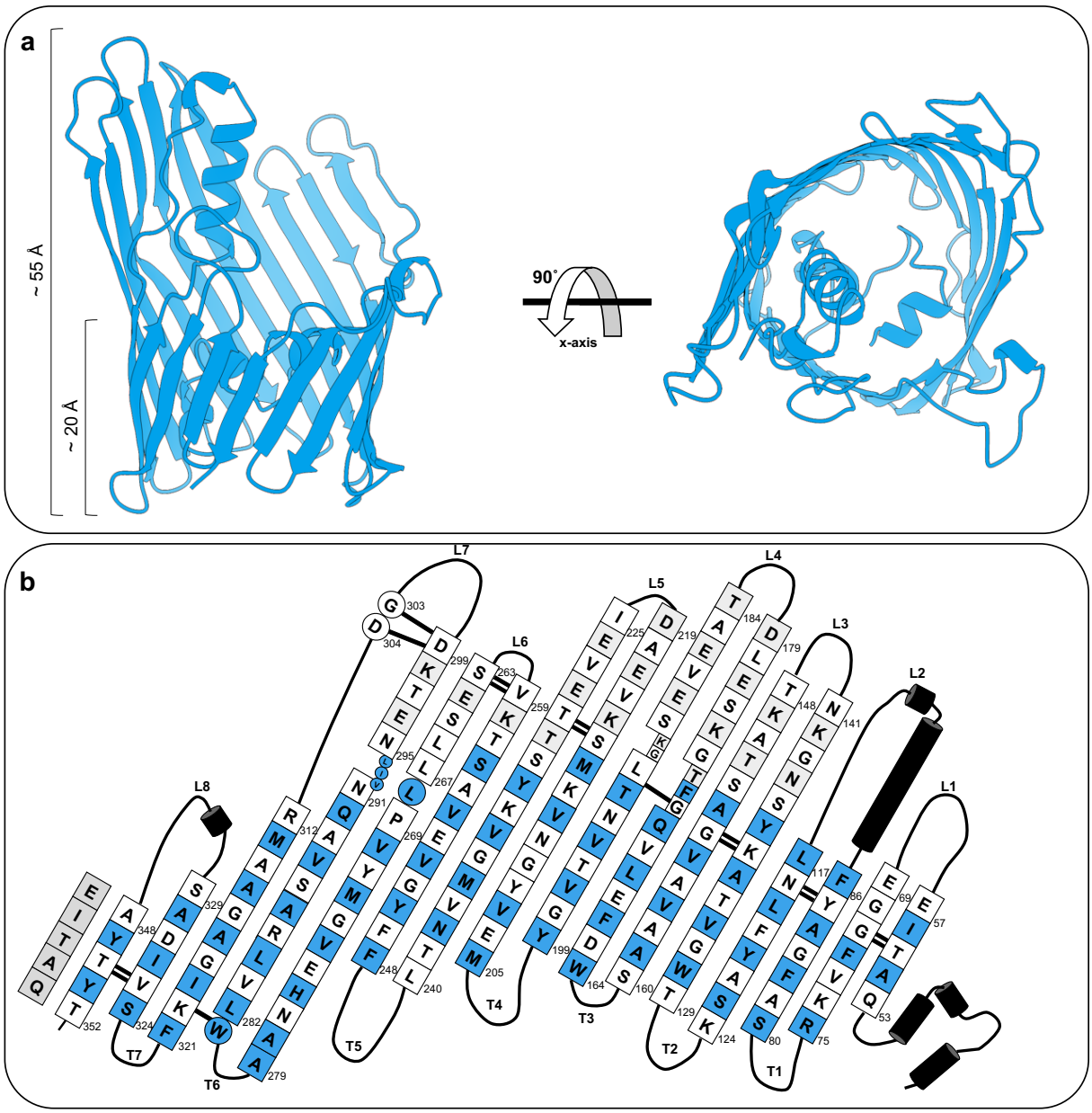


Figure 4-6: Bd0427 Protein Fold and Topology a) Cartoon Ribbon Diagram of Bd0427 Structure. b) 2d Topology Diagram of Bd0427 structure. Squares represent residues forming β-strands, with blue (within hydrophobic band) and grey (in hydrophilic environment) residues facing out towards the membrane or extracellular environment and white facing into the barrel. Circles represent residues involved in loop formation. α-helices are represented by black cylinders. Hydrogen bonds used in calculating the shear number are shown as horizontal black lines.

number is +16, as calculated by running from β13 to the left along the hydrogen bonds once around the barrel and back to the same strand, the residues then counted from the starting residue to the end residue, with residues 292-294 being counted as one residue due to forming a loop and disrupting hydrogen bonding (Figure 4-6b).

The 16 β-strands vary in length by 11 residues with the shortest strand being 5 residues in length and the longest being 16 residues in length, with the mean number of residues per strand at 11 residues. Most adjacent β-strands formed hydrogen bonds between backbone peptide bonds. This was disrupted in β7 and β8, which both contain a twisted β-strand, and in β12 and β13, which contain loops within their β-strands of 1-3 residues. On the internal (periplasmic) side, the β-hairpin turns are 2-7 residues in length (mean 4). The external (extracellular) side consists

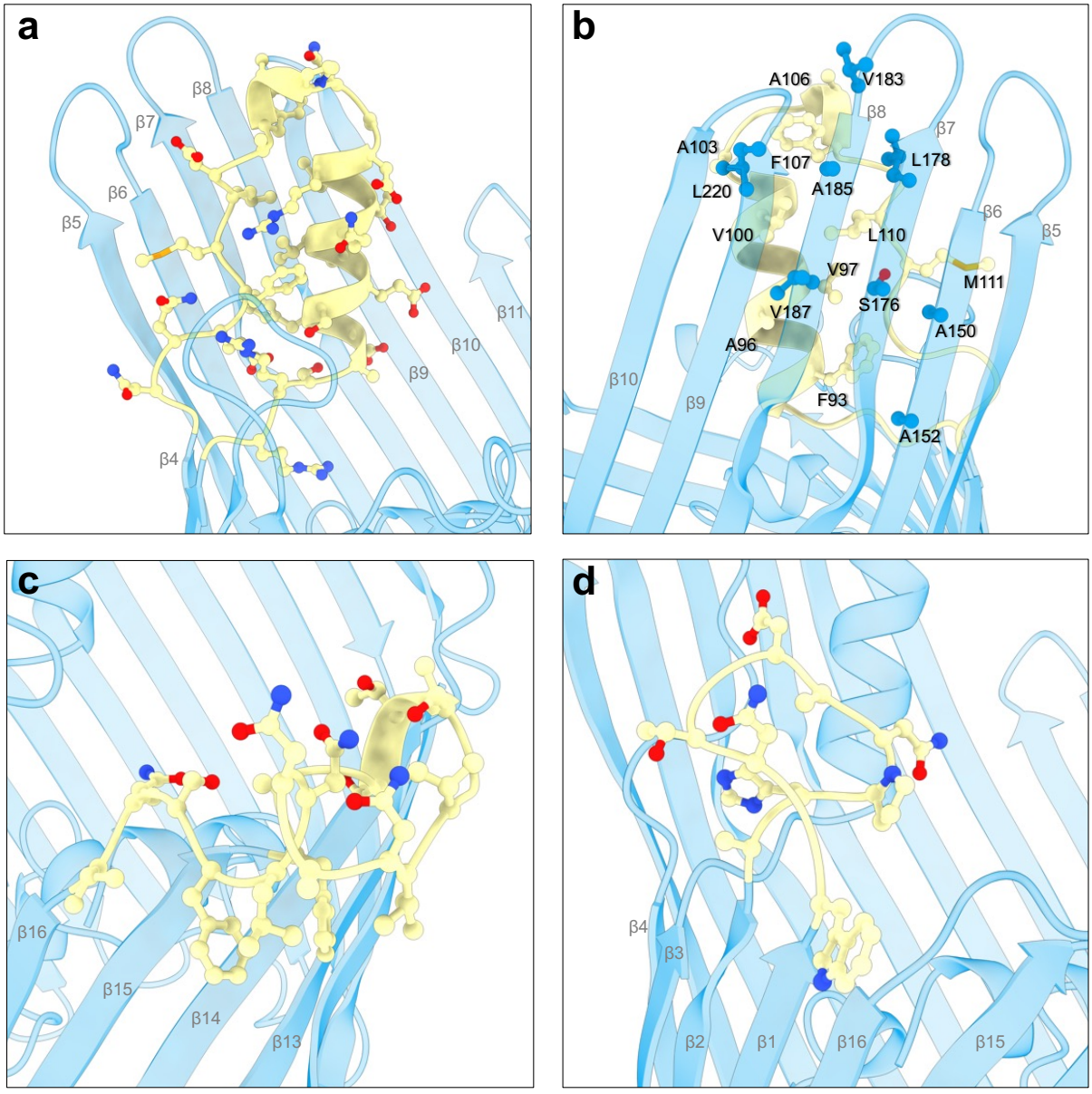


Figure 4-7: Loops 1, 2 and 8 in Detail. a) Cartoon and stick diagram of Loop 2 (yellow). b) Cartoon of the hydrophobic interactions between Loop 2 (yellow) and the barrel wall (blue). c & d) Cartoon and stick diagram of Loops 8 and 1 (yellow), respectively. β-strands labelled in grey.

of loops that varied in length from 3-30 residues (mean 11), with loops 1, 2, 7 and 8 being the longest.

Loop 2, the longest loop (30 residues), is the only loop to fold back into the barrel lumen and forms a 12-residue α-helix followed by a short loop and a second 3-residue α-helix. The 12-residue helix is amphipathic, with the hydrophilic edge directed towards the lumen and the hydrophobic edge facing towards the barrel wall (Figure 4-7a). The helix's hydrophobic residues make several interactions with hydrophobic sidechains from the barrel wall that face into the barrel (Figure 4-7b). Loop 8, the second longest at 19 residues, folds outward from the barrel forming an 'arrow shaped' loop with a 3-residue α-helix. This loop forms interactions with the loop 8 from both adjacent chains in the pentamer. The loop (Figure 4-7c) has hydrophobic sidechains (Ile338, Lue343, Phe344, Val347) that pack against the membrane and polar sidechains (Ser333, Thr334, Thr335, Asn339, Asn341, Asn342, Ser345, Gln346) facing away from the membrane. Loop 1 initially folds outwards from the barrel but owing to a cis-peptide bond, between Asn65 and Pro66, the loops folds into the lumen forming a sharp VI-type turn (Figure 4-7d).

4.1.4 Homopentamer Interface

To document the nature of the homopentamer interactions, protein:protein interfaces were investigated both manually and by use of the PISA server; the result of these are provided in Table 4.3. This analysis revealed that there were five interfaces where each interface involves 98-102 residues consisting of 25-26 hydrogen bonds, 3-7 salt bridges and Van der Waals interactions with a mean interface area of 1719 Å. The interfaces involve five loops (1, 2, 3, 7, 8), two turns (5,7) and sidechains from β1-β4 and β10-β16 (Figure 4-8).

Table 4-3: Interface Statistics

Interface	1		2		3		4		5	
Chain	A	B	B	C	C	D	D	E	E	A
Interface residues	50	49	50	48	52	50	50	49	50	51
Interface atoms	178	183	180	181	185	188	181	183	182	188
Interface area (Å^2)	1717.5		1693.4		1752.9		1692.8		1738.7	
No. Hydrogen bonds	26		25		26		25		26	
No. Salt Bridges	7		5		6		3		5	
ΔG_{int} (kcal/mol)	-23.0		-23.2		-23.3		-25.0		-23.5	
ΔG_{int} P-value	0.216		0.209		0.234		0.153		0.210	

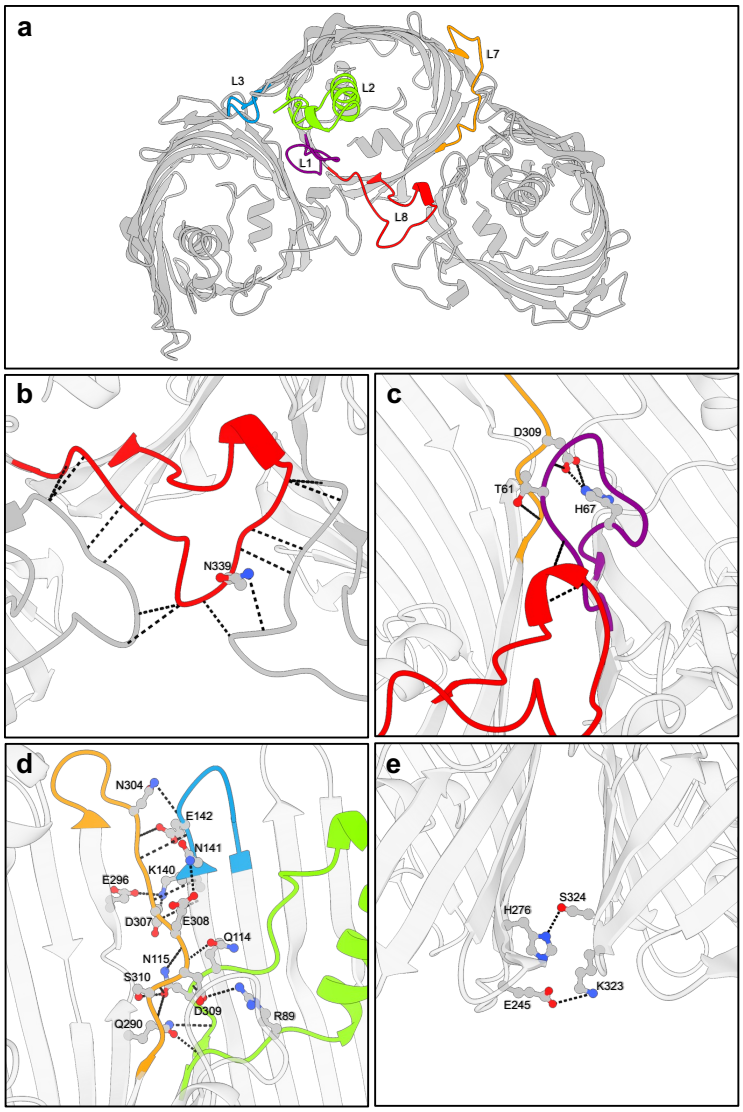


Figure 4-8: Hydrogen-bonds and Salt-bridges Involved in the Interface Bd0427 Homopentamer. Schematic showing chains A, B and C with B being the central monomer. Hydrogen bonds and salt bridges shown with black dashed lines. Loop 1 (purple), 2 (green), 3 (blue), 7 (orange), 8 (red). a) Overall loops involved in the interface b) Hydrogen bonds and salt bridges between loop 7 and β13 in chain A to chain B. c) Hydrogen bonds and salt bridges between loop 1 in chain B and chain A. d) Hydrogen bonds between loop 8 chain B and loop 8 chains A and C. e) Hydrogen bond and salt bridge between β15 and T7 in chain B and chain A.

Loop 8 forms intermolecular hydrogen bonds mostly from the backbone (carbonyls to nitrogens) to the backbone of the two adjacent monomers. This occurs from residue G336 to V347, securing the 5-fold arrangement as seen in Figure 4-8b. Four hydrogen bonds (S333:G59, T334:N60, T311:T61, D309:G62) and two salt bridges (D309:H67) are formed by loop 1 to loop 8 and loop 7 of the adjacent monomer (Figure 4-8c). Loops 2 and 3 form fourteen hydrogen bonds and three salt bridges (E296:K140, D307:K140, D309:R89) to loop 7 and β13 on the adjacent monomer (Figure 4-8a). The final hydrogen bond and salt bridge are found on the periplasmic side of the protein between turn 7 and turn 5 (of adjacent monomer) and β15 and β12 (of adjacent monomer) (Figure 4-8e). Hydrophobic packing at the interface involves the sidechains from β-strands 14, 15, 16, 1, 2, 3, 4 and loop 1 (chain B) and β-strands 10, 11, 12, 13 14 and loop 8 (chain A) (Figure 4-9).

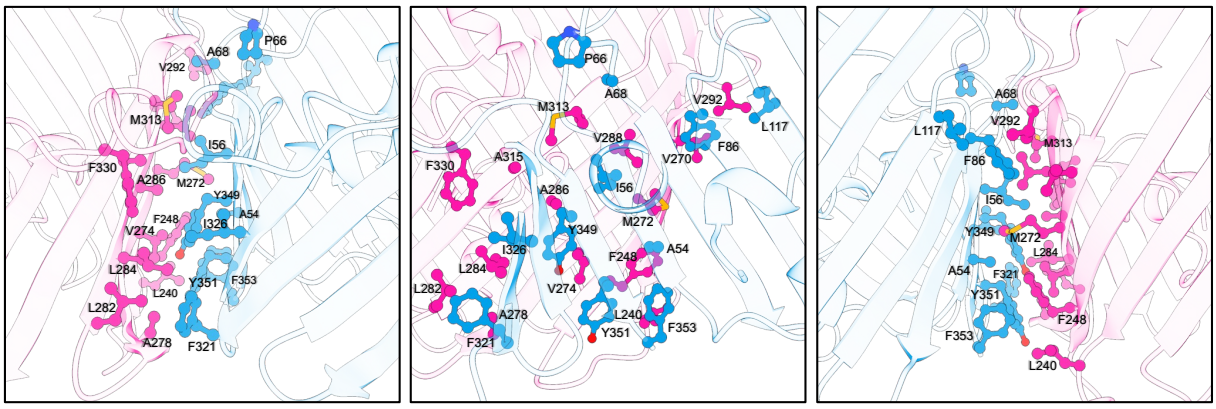


Figure 4-9: Hydrophobic Interactions of the Interface.

4.1.5 Surface

Typical to porins and other OMPs the surface of Bd0427 features a hydrophobic band with a width of approximately 25 Å (Figure 4-10a), that circles the entire pentamer. These residues form hydrophobic interactions with the fatty acid tails within the outer membrane. The boundaries of this hydrophobic band are largely composed of aromatic residues (Figure 4-10b), indicating that Bd0427 has aromatic girdles similar to other porins. The extracellular boundary

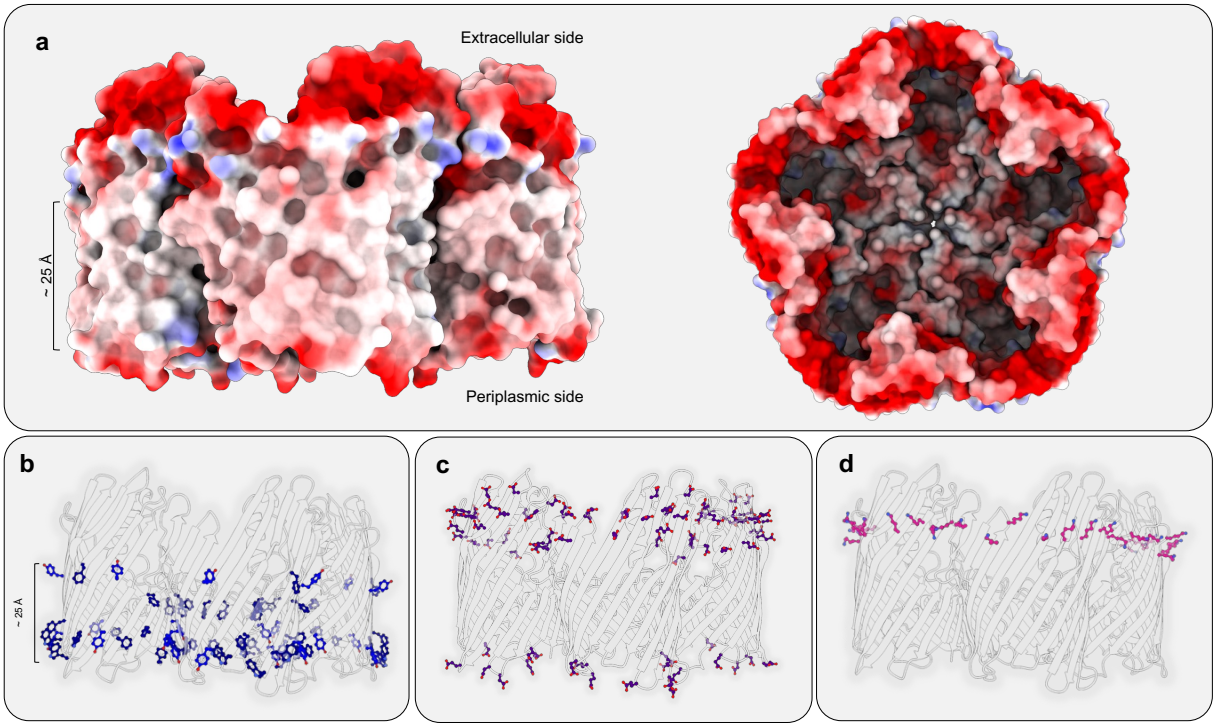


Figure 4-10: Surface of Bd0427. a) Electrostatic surface of Bd0427, side and top view. b) Aromatic girdle of Bd0427. c) Negatively charged residues on the extracellular and periplasmic face. d) Lysine ring of Bd0427.

is composed of tyrosine and phenylalanine, where all tyrosine sidechains point outwards away from the membrane and all the phenylalanine sidechains point into the membrane. The periplasmic boundary is composed of all three aromatic residues, in this boundary all the sidechains of tyrosine point outward away from the membrane, this time towards the periplasm.

Externally facing residues on the extracellular face of the β-barrel are mostly Asp/Glu; this creates a large area of negative charge (Figure 4-10a). An area of negative charge is also observed on the periplasmic-facing side of the barrel, though the area is significantly smaller. As well as this area of negative charge on the extracellular face, there is also a ring of positive charge just above the hydrophobic band formed by a ring of lysine residues (Figure 4-10d).

4.1.6 The Transmembrane Channel and N-terminal Plug

Each chain's transmembrane channel is approximately 25 Å in diameter and the exposed surface is lined mostly with hydrophilic residues. This channel is plugged completely by the N-

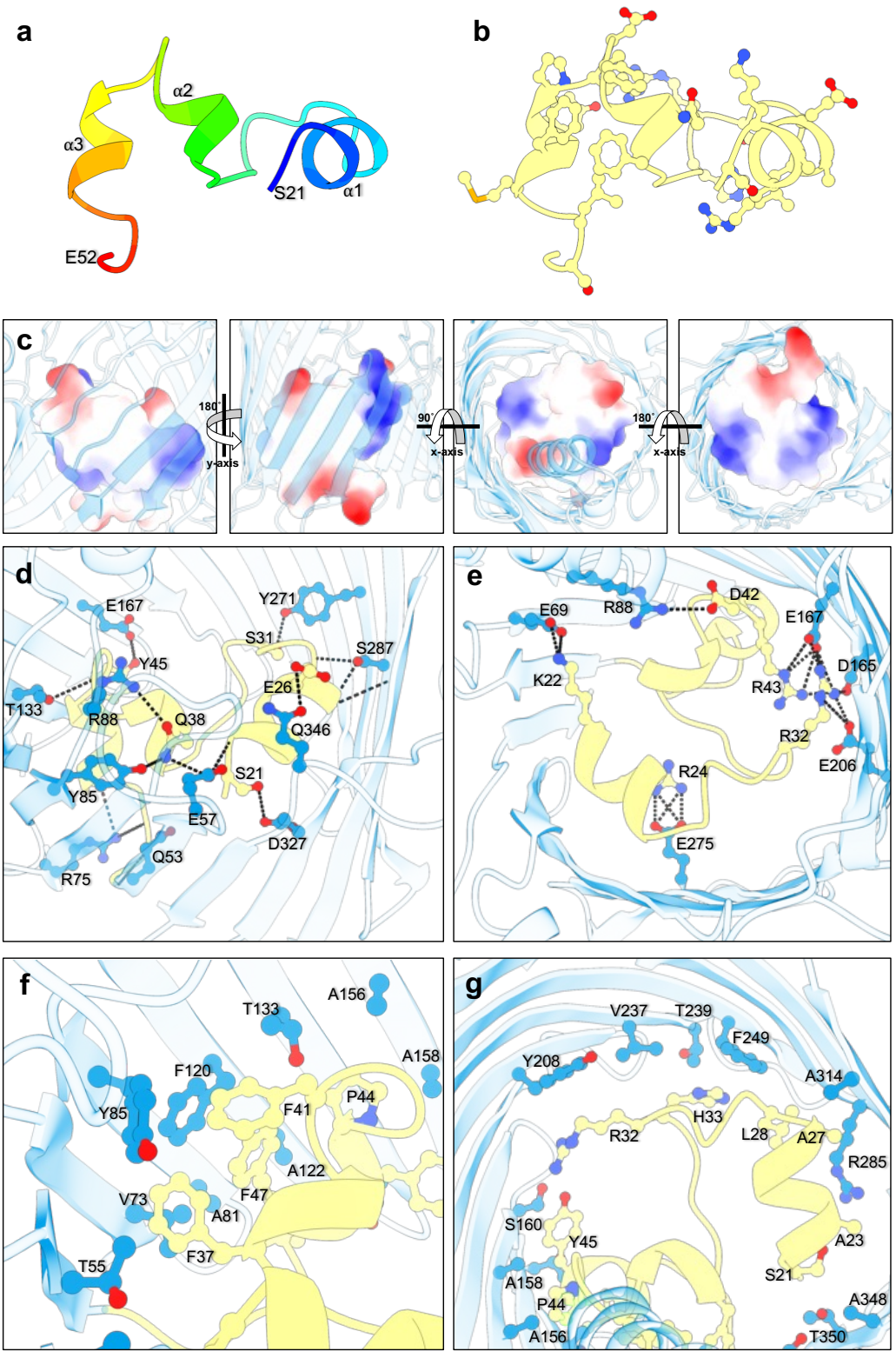


Figure 4-11: The N-terminal plug and the Interactions the Plug makes with the Barrel. a) Cartoon representation of the N-terminal plug of Bd0427, coloured in rainbow (blue – red). b) Cartoon diagram showing the residues involved in the plug barrel interface. c) Schematic showing the electrostatic surface of the plug from four different orientations. d) Cartoon diagram, showing the residues involved in forming hydrogen bonds between plug and barrel wall. e) Cartoon diagram, showing the residues involved in forming salt bridges between plug and barrel wall f&g) Cartoon diagram, showing the residues involved in hydrophobic interactions and van der Waal contacts between plug and barrel wall

terminal 32 residues (21-52) of the protein (Figure 4-11a), which form a compact domain containing three short α-helices (4-6 residues) connected by loops (Figure 4-11a and b). The surface of the plug has two patches of hydrophobic residues and two positively and three negatively charged patches (Figure 4-11c).

The plug forms 15 hydrogen bonds and 17 salt bridges with the residues in the barrel wall and Arg88 from loop 2 (Figure 4-11d and e). Fifteen of the seventeen salt bridges occur between plug residues that are on the opposite side to where the plug transitions into the region that forms the barrel wall (Figure 4-11e). The plug also forms Van der Waals and hydrophobic interactions with the barrel wall (Figure 4-11f and g). In Figure 4-11f the plug can be seen to have a patch of aromatic residues (Phe37, 41 & 47). Two of these residues are involved in Pi-stacking in the form of staggered stacking (parallel aromatic rings are displaced by half the

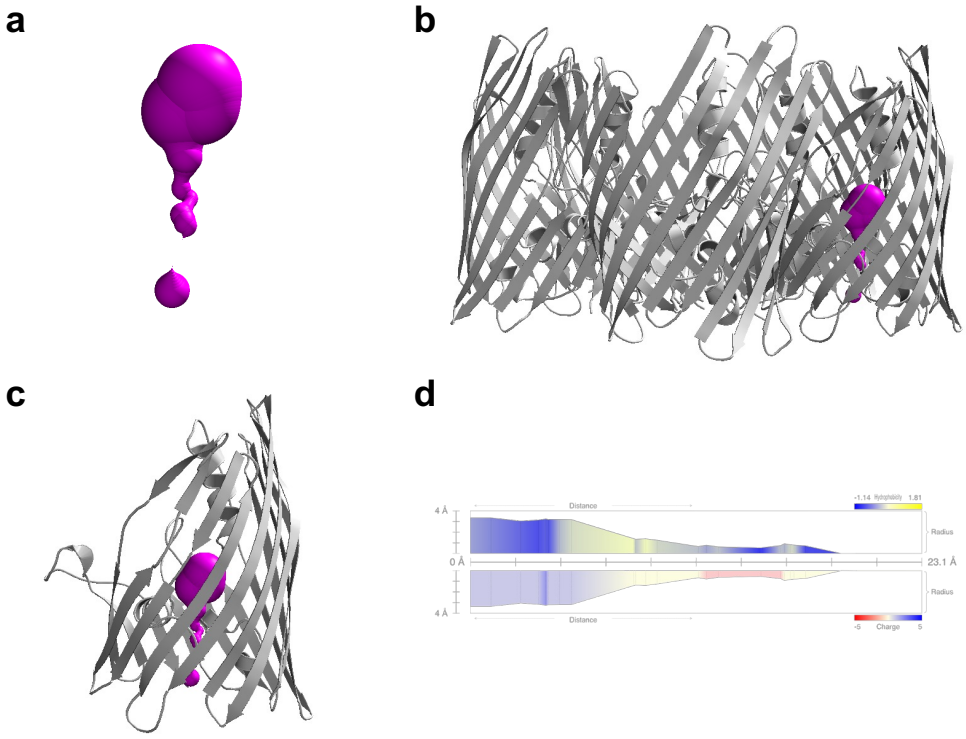


Figure 4-12: Channel of Bd0427 determined using MOLEonline. a) Surface diagram of the channel calculated using MOLEonline. b) Diagram showing calculated channel within Bd0427 crystal structure, channel shown for chain A. c) Diagram showing calculated channel in chain A of Bd0427 d) Graphical representation of channel, including radius, hydropathy and charge from top to bottom.

width of the aromatic ring) and pi-teeing (aromatic rings perpendicular to one another) between Phe47 and Phe120, and Phe41 and Tyr85, respectively.

The blocking of the channel by the N-terminal plug was confirmed by using MOLEonline to search for channels in the pentameric structure. The channel prediction software predicted equivalent channels in each chain (chain A prediction shown Figure 4-12). All these channels had a constriction point of 0 Å (Figure 4-12d), supporting that the plug blocks the pore.

4.1.7 Density for Detergent Tails in Hydrophobic Pocket

The extended Loop 8 sits within the height of the membrane and creates a pocket measuring approximately 15 Å in height. The residues that line this pocket produce a hydrophobic surface,

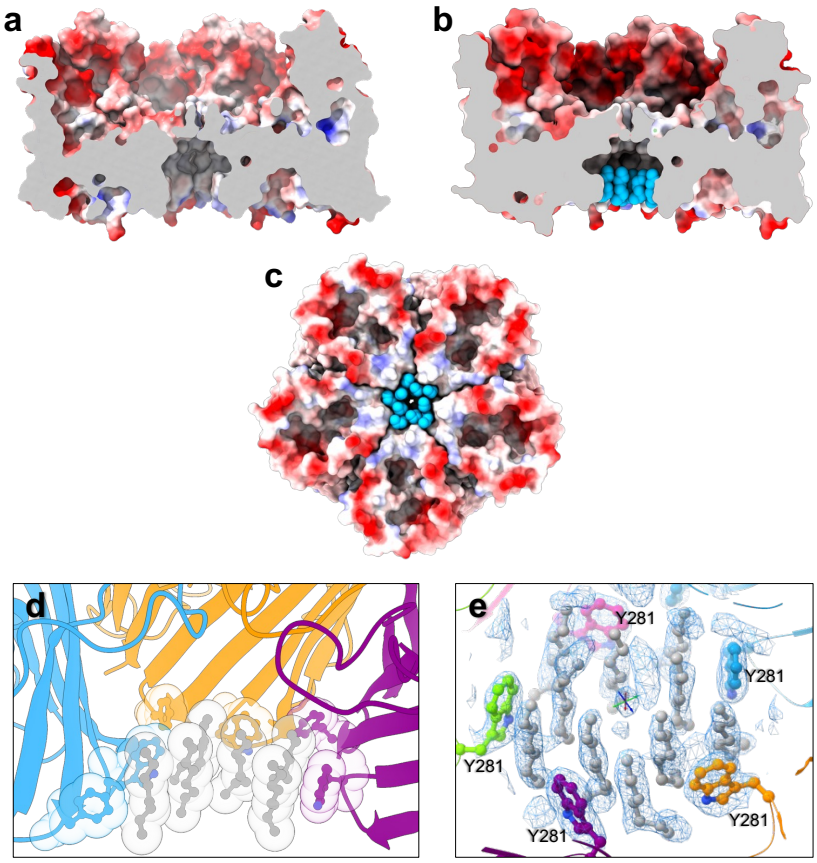


Figure 4-13: Packing of Detergent Tails. a) Cut through surface diagram showing hydrophobic pocket. b) Cut through surface diagram, including octane (in blue) ligands. c) Bottom view of octane (blue) in the pocket of Bd0427. d) Ribbon diagram, showing residues interacting with the ligands (grey). e) Electron density diagram, showing fit of ligand to the density.

which can be seen in the electrostatic surface diagram (Figure 4-13a). Ten vertical rod-like densities can be seen in the map, in which octane, representative of β-OG's fatty acid tail, can be modelled (Figure 4-13e). The ten octane molecules pack via 5-fold symmetry (Figure 4-13c), forming hydrophobic interactions with Trp281 (Figure 4-13d). Despite fitting the entirety of the electron density (Figure 4-13b), it appears the pocket can fit hydrophobic molecules longer than octane, as there is a gap between the octane molecules and the protein surface.

4.1.8 Unmodelled Electron Density

From inspecting the 2mFo-DFc and mFo-DFc maps in Coot, there are two volumes of positive density where a ligand could not be determined unambiguously (Figure 4-14). The first volume (Figure 4-14a) is just above the N-terminal plug above the loop between α1 and α2, surrounded by Glu253 and Arg312. Hydrogen bonds, salt bridges and/or electrostatic interactions could occur between the unknown ligand and Arg312. The second volume (Figure 4-14b) sits towards the bottom of Loop 2's α-helix and forms a long chain-like density. The ligand would be in proximity to form hydrogen bonds, salt bridges and/or electrostatic interactions with Asp42, Arg88 and Lys135, towards the bottom of the density. The top of the density would be able to

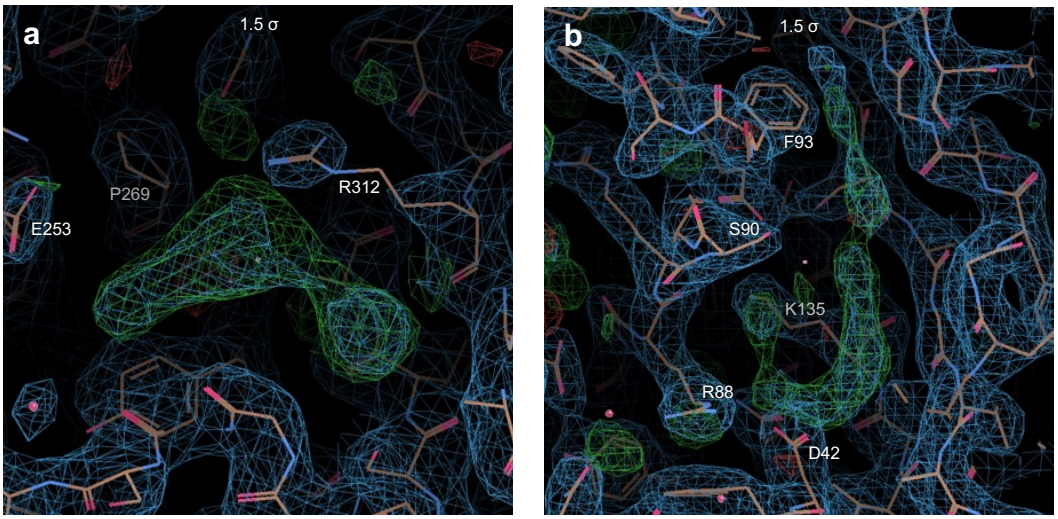


Figure 4-14: Coot 2mFo-DFc and mFo-DFc Map of Bd0427, Crystal Form A at sigma 1.5. a) Electron density map of Bd0427 crystal form A, where a volume of positive density can be seen above the N-terminal plug. b) Electron density map of Bd0427 crystal form A, where a volume of positive density can be seen behind Loop 2's α-helix.

form interactions with the backbone of the protein, as well as hydrophobic interactions with Phe93.

4.1.9 Conservation of Residues

Consurf was used to calculate the conservation of residues in Bd0427 (Figure 4-15). As with other OMPs the loop regions are mostly variable, but all have some conserved residues except loop 4. The β-strands, β8, β9, β10 and the top parts of β11, β12, β13, all have more variable residues than conserved residues; though there is band of adjacent residues on strands β9, β10,

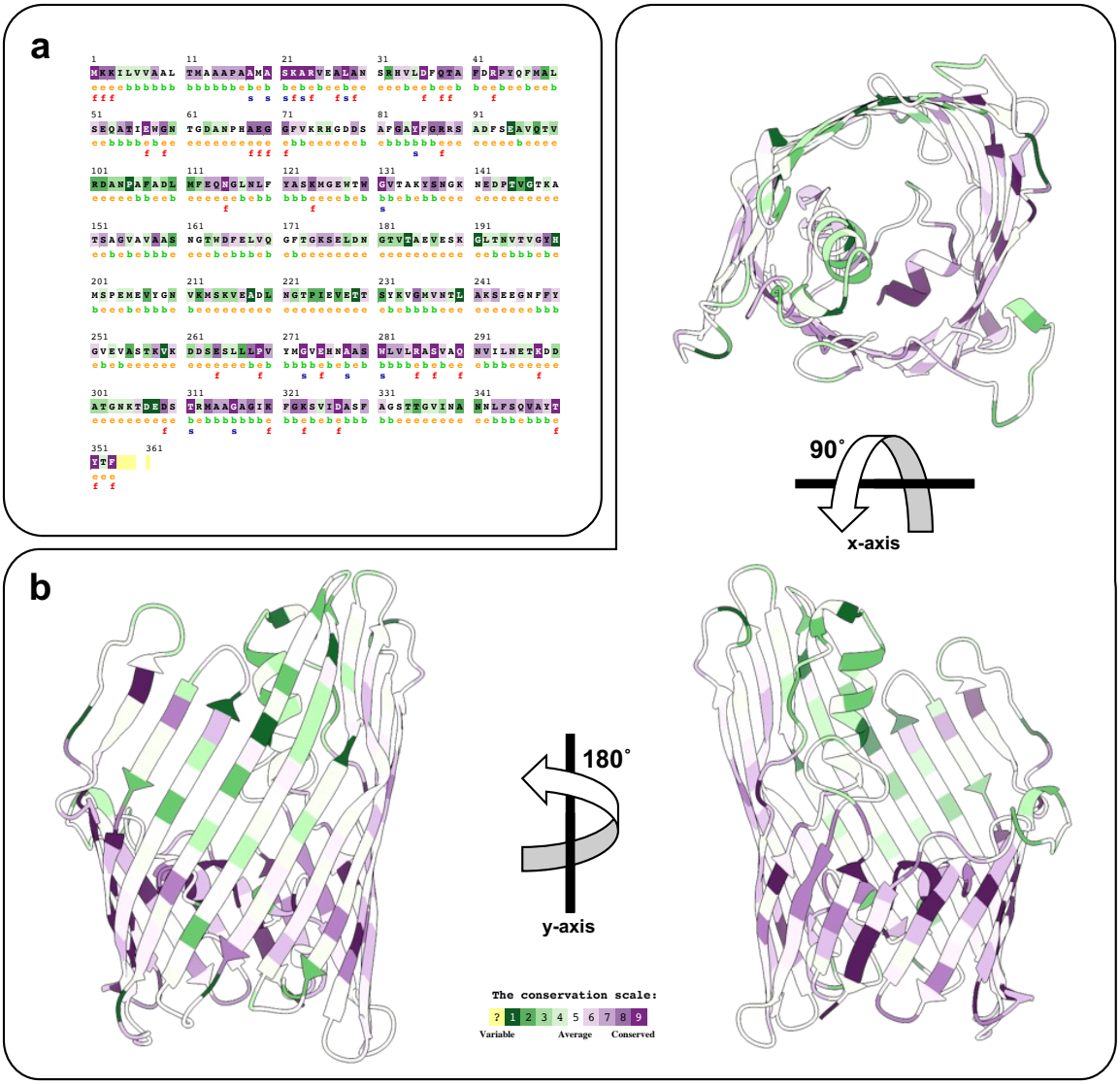


Figure 4-15: Conserved residues of Bd0427 using Consurf. a) Sequence of Bd0427 coloured to show conserved residues, f- predicted function residue, s- predicted structural residue. b) Cartoon of crystallographic structure coloured with ribbon coloured to show conserved residues.

β11, β12, β13 that are conserved (Met213, Thr230, Lys258, Glu264, Lys298). The β-strands, β1, β2, β3, β4, β5, β14, β15, β16 and the bottom parts of β11, β12, β13, all have more conserved residues than variable residues that face both outward and inwards of the barrel. The first 10 residues of the construct (21-30) are highly conserved with most scoring 8 and 9s on the conservation scale (Figure 4-15b).

4.2 Discussion

4.2.1 Bd0427 is Homopentameric

As discussed in Chapter 3, porins and specific channels with known structures have all (excluding MspA from *M. smegmatis* which has a cell envelope which is different to other Gram-negatives) been shown to form either monomers or homotrimers, formed from β-barrels with different numbers of β-strands (12 to 18). Nearly all porins and specific channels with β-

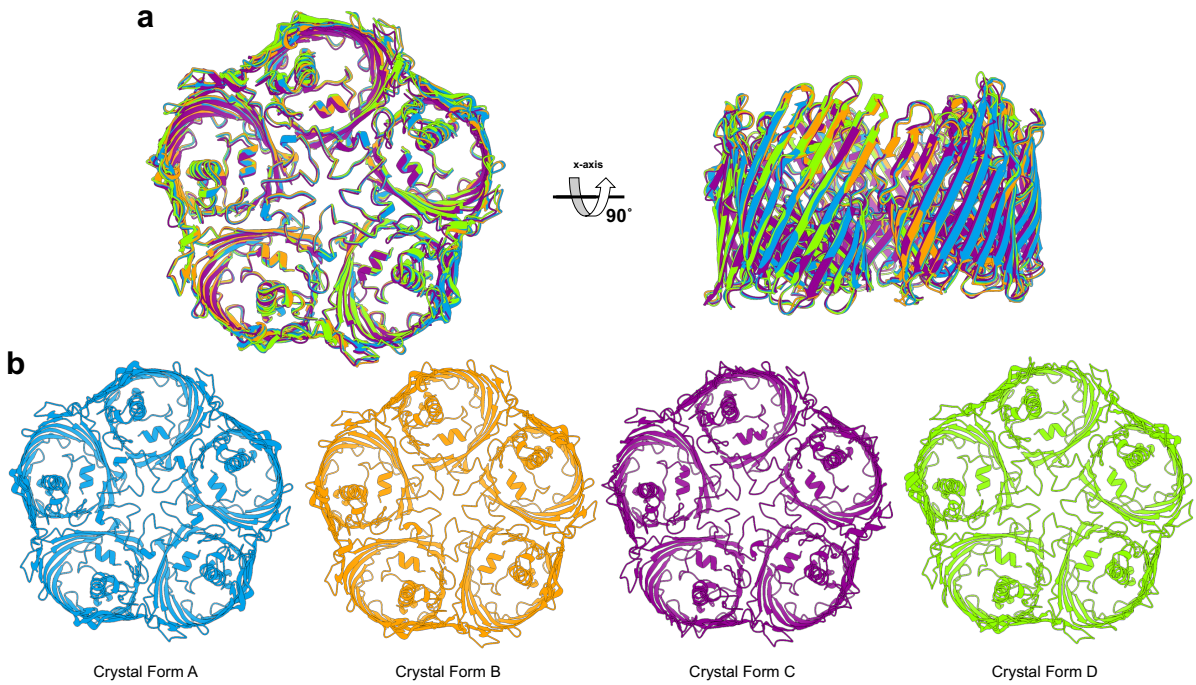


Figure 4-16: Bd0427 Fold from each Crystal Form. a) Superposition of all crystal forms onto crystal form A. b) The homopentameric structure of each crystal form, A-D.

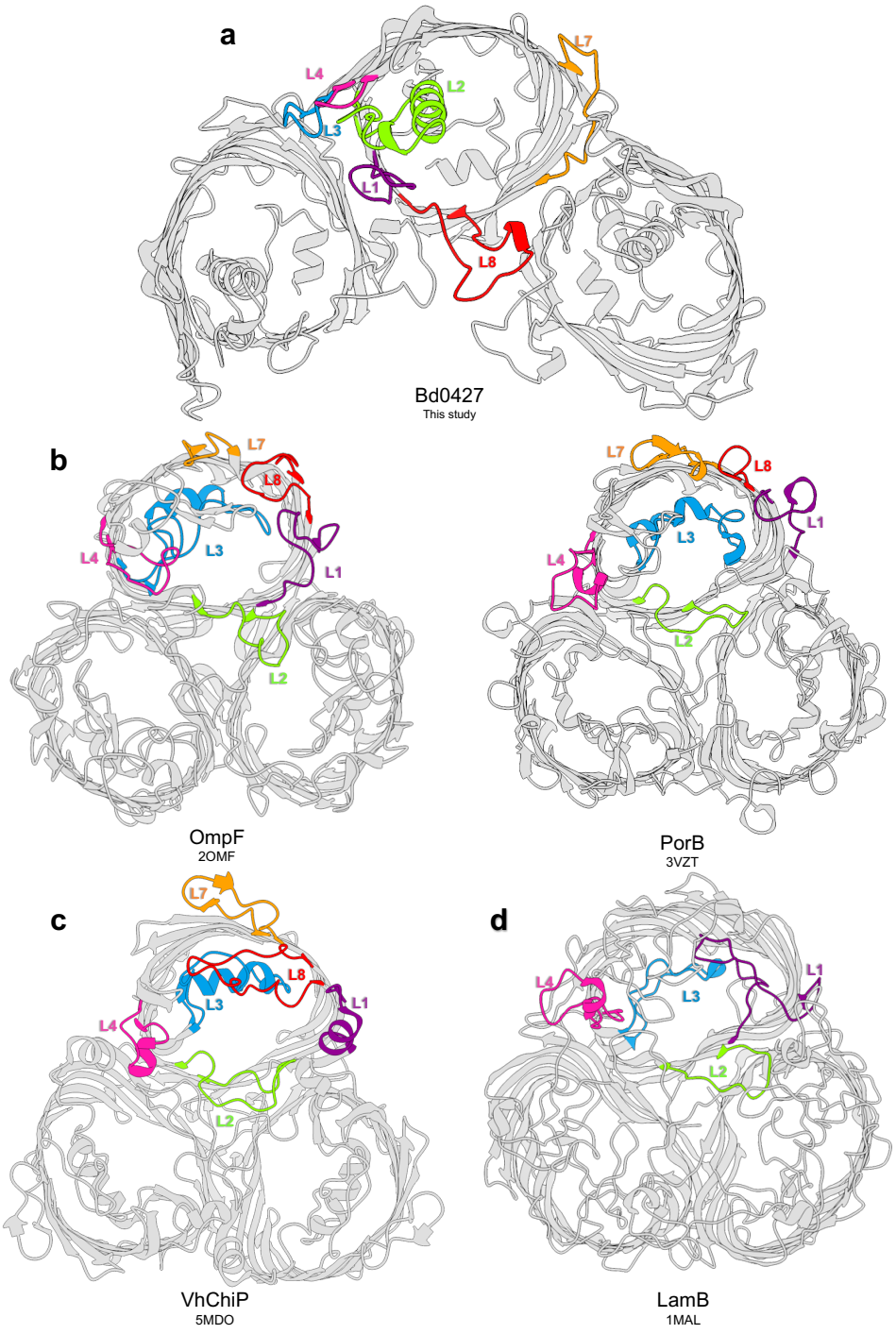


Figure 4-17: Comparison of Bd0427 to Homotrimeric Porins and Specific Channels. a) Bd0427 structure with loops involved in homopentamer interface and those that form homotrimers in other published structures coloured. b&c) Structure of three 16-stranded β-barrel proteins (two porins (b) and a specific channel (c) and the loops involved in forming a homotrimer, and the loops used by Bd0427 to form the homopentamer coloured. d) Structure of 18-stranded specific channel with the loops involved in forming the homotrimer coloured, along with loop 3, loops 7 and 8 are not coloured as these cannot be directly compared due to change in number of strands forming the barrel.

barrels formed of either 16 or 18 strands result in homotrimers. Bd0427 is a 16-stranded β-barrel predicted to function as a porin, and based on the known structures of porins it was hypothesised that Bd0427 would likely form a homotrimer. Determination of the Bd0427 structures showed that all four crystal forms, presented in this chapter, contain a homopentamer in the ASU (Figure 4-16), an oligomeric state not previously observed for 16-stranded β-barrels. These four crystal forms can be structurally aligned to form A (B: RMSD = 0.27 Å, alignment length = 333; C: RMSD = 0.51 Å, alignment length = 326; D = 0.396 Å, alignment length = 332).

In forming a homopentamer, one of the most noticeable differences is the loops used to form intramolecular interactions. Homotrimers use loops 1, 2 and 4, with loop 2 latching onto the adjacent monomer, with loops 1 and 4 forming intramolecular contacts between either loop 4 or 1, respectively, on the adjacent monomer. This pattern is observed in homotrimers of both 16- and 18-stranded β-barrels (Figure 4-17b-d). Bd0427 forms a homopentamer using loops 1, 2, 3, 7, and 8. In the homopentamer it is loop 8, not loop 2, that extends out of the barrel and forms hydrogen bonds with two adjacent monomers (Figure 4-17). Loops 1, 2, and 3 all form intramolecular contacts with loop 7 of the adjacent monomer, instead of loop 1 to loop 4 in homotrimers. The details of these differences are provided in Table 4-4

Table 4-4: Role of each Loop involved in the Interface of either Homotrimers or a Homopentamer

Loop	Homotrimers	Bd0427 (Homopentamer)
1	Forms intramolecular bonds with loop 4 from the adjacent monomer.	Forms intramolecular bonds with loop 7 and 8 from the adjacent monomer.
2	Extends from the barrel and latches onto the adjacent monomer, often within a grove.	Folds back into the barrel and forms one long and one short α-helix that packs against the barrel wall. The start and end of the loop forms interactions with loop 7 from the adjacent monomer.
3	Long loop that folds back into the barrel lumen and forms the constriction point in the pore.	Short loop that forms intramolecular bonds with loop 7.

4	Forms intramolecular bonds with loop 1 from the adjacent monomer.	Not involved in the interface, due to position.
7	Not involved in the interface, due to being on the opposite side of the barrel to the interface.	Forms intramolecular bonds with loops 1, 2, and 3 from the adjacent monomer
8	Not involved in the interface, due to being on the opposite side of the barrel to the interface.	Folds away from the barrel, and forms intramolecular bonds with adjacent monomers.

Hence, we have identified a new oligomeric state for 16-stranded β-barrels that is divergent from 16-stranded β-barrels that form homotrimers. The question resulting from this is whether there is a reason as to why it forms a homopentamer and not a monomer or homotrimer as per "common" porins. It would be interesting to know whether this oligomeric state is essential to function or merely a result of evolution where this is the most stable structure for this protein.

4.2.2 Bd0427 Structural Comparison to Deposited Porin Structures

Bd0427 utilises different loops to form the homopentamer than other porins and specific channels use to form homotrimers. However, the Bd0427 structure also contains characteristics typical to porins and specific channels. All porins and specific channels deposited in the PDB have a hydrophobic band (on average approximately 24 Å) that correlates to the thickness of the hydrophobic region of the OM (Lomize et al., 2011; Wu et al., 2014). Bd0427s hydrophobic bands are ~25 Å, this compares to the hydrophobic band of other OMPs in the PDB. This hydrophobic band is lined by aromatic residues, forming two rings known as aromatic girdles that are characteristic of porins and specific channels. The sidechains of the residues that line the exposed surface of Bd0427 are either polar or charged as per other porins and specific channels.

As this structure was quite diverse from structures of porins and specific channels, the protein structure was compared to all deposited protein structures in the PDB, using DALI (Holm, 2020). This was done to see if there were any similar structures that were not porins or specific

channels which may indicate a potential alternative function. The structures of the proteins with the top Z-scores were manually analysed (top 11 shown in Figure 4-18), from this no structures

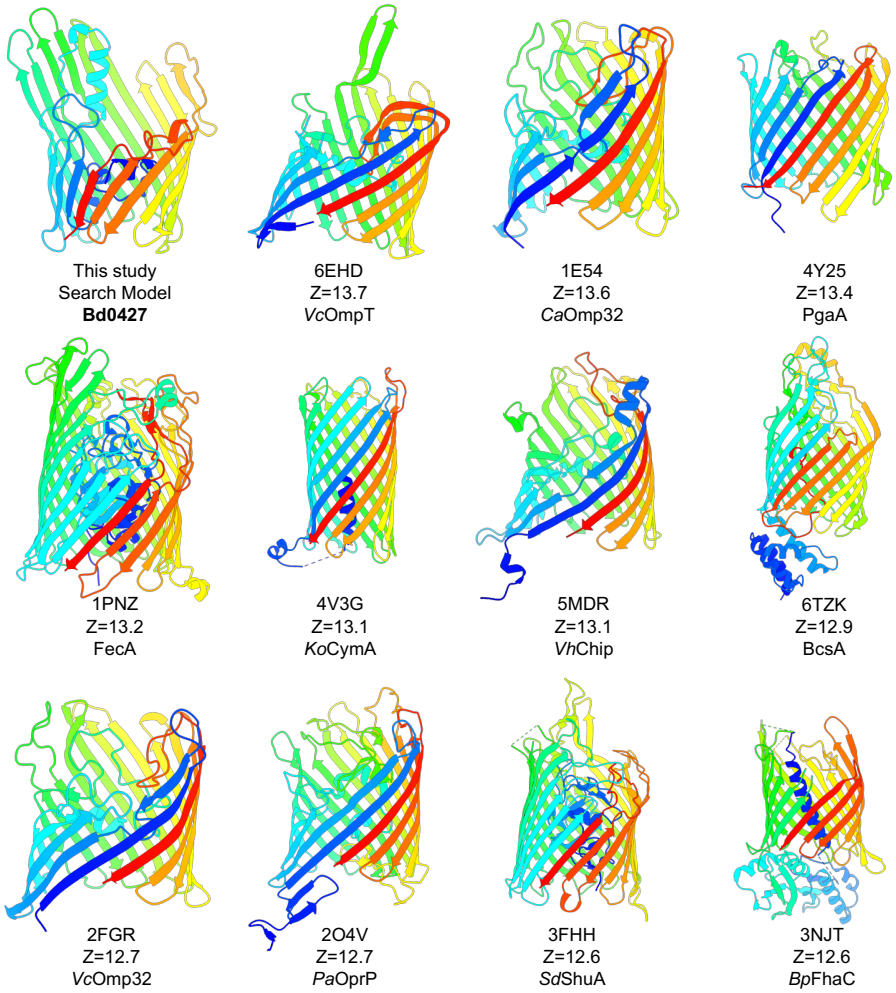


Figure 4-18: Top Results from DALI, using Bd0427 as Search Model. Top results are determined by the hits that had the highest Z-score. Coloured in rainbow (N-terminus blue and C-terminus red).

were identified from the DALI search that had the same β-barrel fold as Bd0427, which further indicates that Bd0427 β-barrel fold is novel.

4.2.3 Could the Lysine Ring be Important for Transfer?

The extracellular face of Bd0427 has a large area of negative charge due to presence of Asp/Glu residues, except for a ring of lysine residues above the hydrophobic band that create a ring of positive charge (Figure 4-19). A lysine ring above the hydrophobic band is also seen in PorB

(Figure 4-19), a porin from *Neisseria gonorrhoea* which gets inserted into the OM of mitochondria within immune cells to instigate cell death (Deo et al., 2018). This ring of lysine residues is not seen in other known porin structures; therefore, is the presence of the lysine ring a requirement of porins that are transferred to a different membrane?

The OM of mitochondria is unlike the OM of Gram-negative bacteria and does not contain LPS; instead, it more closely resembles a Gram-negative IM as it is formed from a phospholipid bilayer. Therefore, both PorB and Bd0427 are transferred into membranes with similar

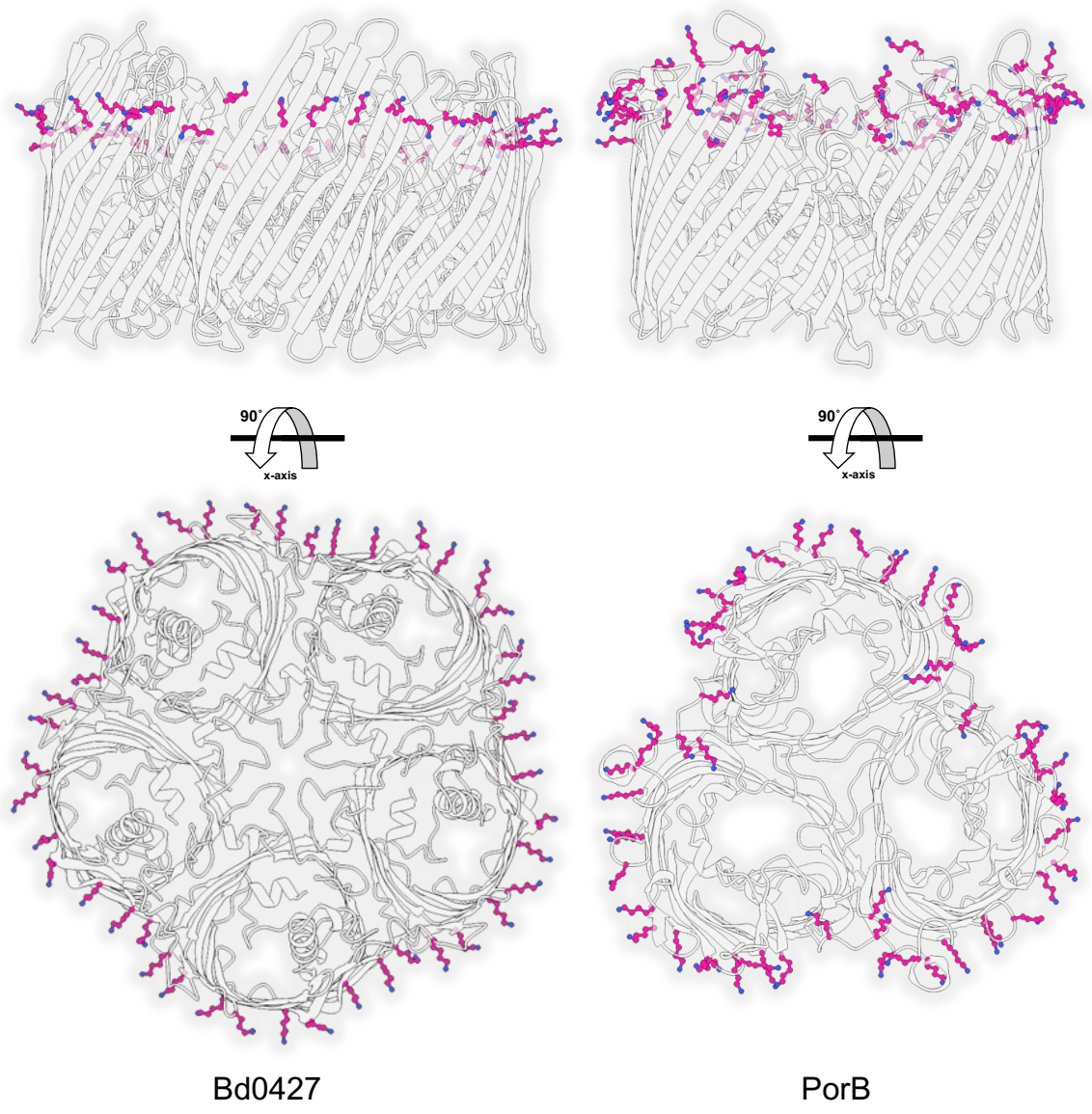


Figure 4-19: Lysine Rings of Bd0427 and PorB (3VZT).

compositions. In transmembrane α-helical proteins, found in the IM in Gram-negatives, lysine residues have previously been shown to act as 'snorkels' with the aliphatic part buried in the hydrophobic part of the bilayer and the amino-group interacting with the negatively charged phosphate in the head group of the lipids (Segrest et al., 1990). As lysine sidechains are long and flexible they can bend and extend, which can allow membrane proteins to adjust the thickness of the hydrophobic band to adjust to different thicknesses of bilayers (Killian and von Heijne, 2000). IMPs are said to have on average a hydrophobic band ~29 Å in size, which on average is 5 Å thicker than the average for OMPs. Therefore, we hypothesise that lysine snorkelling enable Bd0427 adjust as the protein is transferred from an OM environment to IM environment.

4.2.4 The N-terminal Plug is a Hatch Domain that Completely Blocks the Channel

The N-terminal plug of Bd0427 completely blocks the pore in all the crystal structures. Blocking of the pore by an N-terminal plug has previously only been observed in two 14-stranded β-barrel structures: CymA and FadL (Figure 4-20). CymA mediates the passive

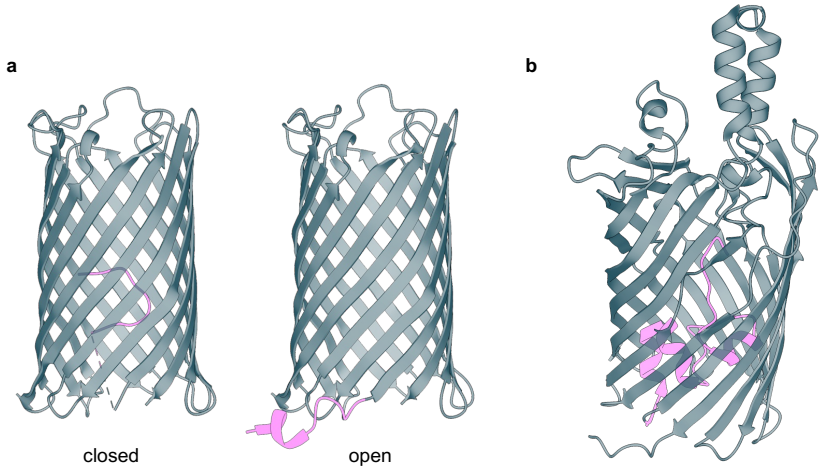


Figure 4-20: β-barrels with N-Terminal Plugs. a) CymA in open (4V3H) and closed (4D51) conformations. b) FadL with N-terminal plug (1T16). The N-terminal plug is coloured in pink and the β-barrel is coloured in teal.

diffusion of cyclodextrins and FadL is a long-chain fatty acid transporter (van den Berg et al., 2015, 2004). The N-terminal plug of CymA is 22 residues in size and does not contain any secondary structure (Figure 4-20a, '*closed*'). The plug is expelled by the ligand, enabling transport of the large (~15Å) substrate through the pore (Figure 4-20a, '*open*'). FadL has an N-terminal plug of 42 residues which consists of 3 short helices (Figure 4-20b), coined a '*hatch*' domain (van den Berg et al., 2004). The plug of FadL has not been shown to be displaced, with the transport of fatty acids observed to pass through an opening in the barrel wall (Hearn et al., 2009). Comparing these N-terminal plugs to Bd0427, the N-terminal plug of FadL and Bd0427 are very similar with an RMSD of 0.88 Å across 15 pruned atom pairs and an RMSD of 4.0 Å across all 32 residues (Figure 4-21). Therefore, the N-terminal plug of Bd0427 is a hatch domain. As the N-terminal plug of FadL is not expelled to enable transport, is this also the case with Bd0427 and instead transport occurs lateral diffusion through an opening in the channel wall? There are two potential places this could occur in Bd427, the loops within β-strands 12

and 13, and the twisted β-strands 7 and 8. Whether Bd0427 forms a pore and if substrates can diffuse through will be further investigated in Chapter 5.

4.2.5 Hydrophobic Pocket Lipid Binding Hypothesis

The hydrophobic pocket under loop 8 contains 10 rod-like densities with 5-fold symmetry that can be modelled with octane, representative of the detergent tail of β-OG. In the proteins native outer membrane environment, we hypothesize that lipids would pack in this pocket. As there are 10 rod-like densities we propose that the pocket could fit the fatty acid tails of five phospholipids. This would form an ordered monolayer on the periplasmic side of the outer membrane with each protomer interacting with two phospholipids. The packing of ordered lipids in membrane proteins has been observed in eukaryotic and Gram-negative inner membrane proteins (Biou, 2023), but we could not find any structures of OMPs containing an ordered lipid monolayer. Whether the ordered lipid monolayer is relevant to function or just structural is currently unknown.

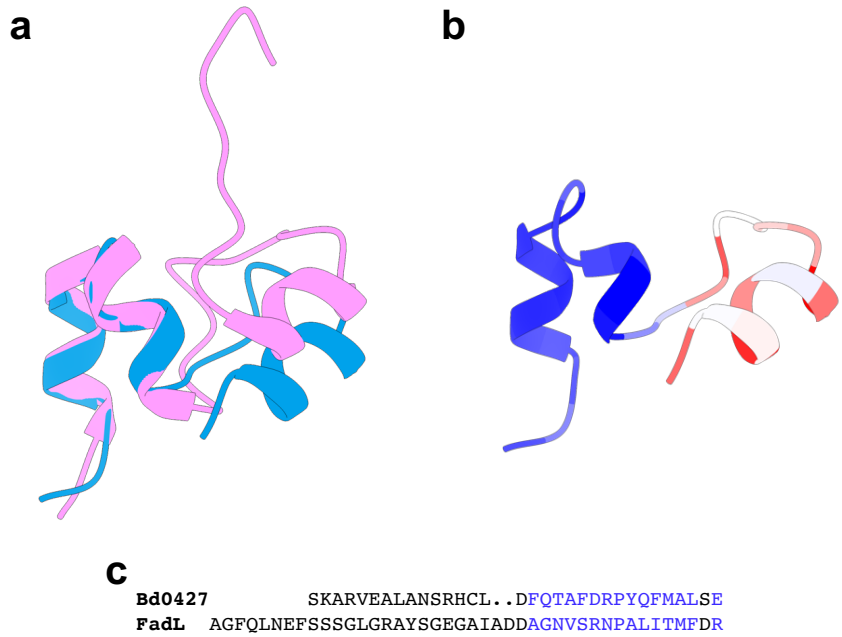


Figure 4-21: Comparison of N-Terminal Plugs of FadL and Bd0427. a) Alignment of N-terminal plugs of FadL (pink) and Bd0427 (blue). b) Bd0427 N-terminal plug coloured per residue RMSD value in ChimeraX. c) Sequence structural alignment, residues coloured in blue are the 15 pruned atom pairs.

4.2.6 AlphaFold and Foldseek Enable Identification of Bd0427 as Part of Wider Superfamily of Oligomeric Porins

Following the release of the AlphaFold Protein Structure Database (Jumper et al., 2021; Varadi et al., 2022) and Foldseek (van Kempen et al., 2022) these tools were used to provide new insights into whether the overall fold of Bd0427 is found in other genus outside of *Bdellovibrio*. It is already known that homologs to Bd0427 are found in all *Bdellovibrio* species and are the major OMP (Beck et al., 2005). There were even some *Bdellovibrio* species that include multiple copies of this protein (

Table 4-5), indicating the importance of this protein.

Table 4-5: Homologs to Bd0427 in *Bdellovibrio* species

Species	Gene	Sequence Identity to Bd0427
B. bacteriovorus HD114	Major OMP	100.0 %
B. bacteriovorus Str. Tiberius	Bdt_0417	80.8 %
B. bacteriovorus W	BDW_01520	43.7 %
B. bacteriovorus W	BDW_01525	43.7 %
B. bacteriovorus NC01	DOE51_02245	40.1 %
B. bacteriovorus qaytius	CIK05_02360	34.5 %
B. bacteriovorus qaytius	CIK05_02355	32.45 %
B. exovorus JSS	A11Q_505	33.1 %
B. exovorus JSS	A11Q_1441	27.1 %
B. exovorus JSS	A11Q_506	24.9 %

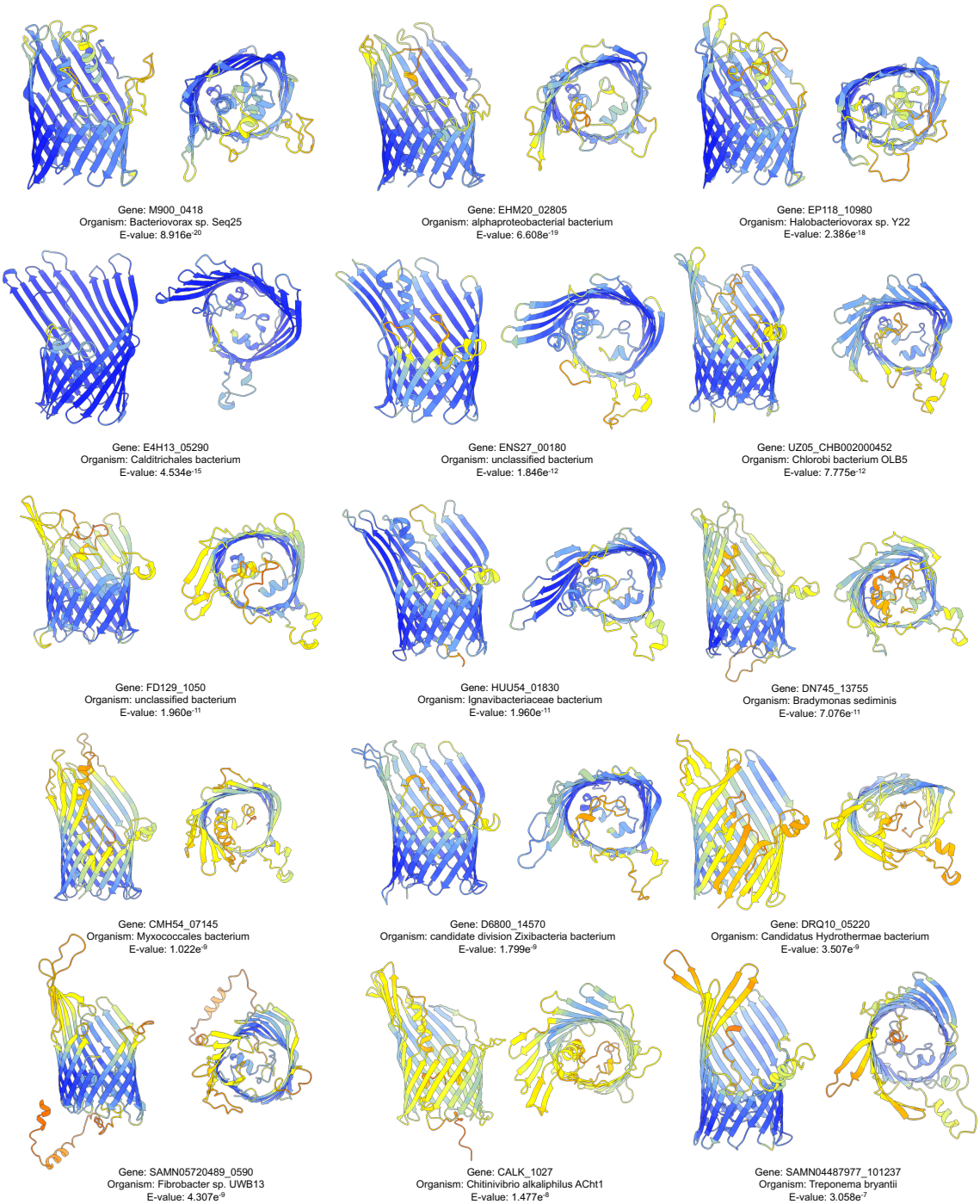


Figure 4-22: Summary Structural Homologs of Bd0427 found using AlphaFold and Foldseek.

From the Foldseek search it was revealed that several different families of bacteria encoded proteins with a similar fold (Figure 4-22), these included commensal bacteria found in rumen, anaerobic bacteria, and other predators. All these proteins, as well as many not included in Figure 4-22, are 16-stranded β-barrels and have an N-terminal plug, asymmetric barrel height,

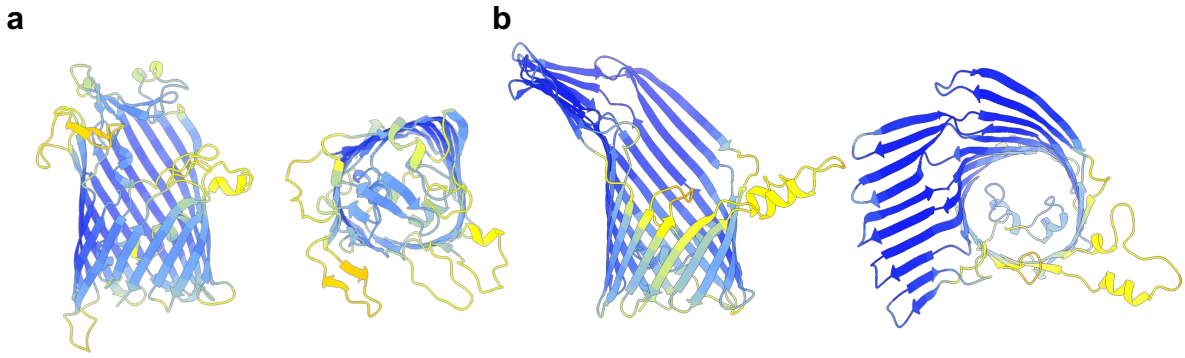


Figure 4-23: AlphaFold Predicted Structures of FSU_2404 (a) and TDE_2508 (b).

and a long loop 8 which folds out of the barrel. This demonstrates that the fold is found in other bacteria, though nearly all have no known function. From these searches two proteins with predicted functions were identified: FSU_2404 from *Fibrobacter succinogenes* and TDE_2508 from *Treponema denticola* (Figure 4-23). FSU_2404 is predicted to be involved in outer membrane vesicle formation and TDE_2508 is predicted to regulate biofilm formation and adhesion to epithelia (Abiko et al., 2014; Yeoman et al., 2021). Therefore, does Bd0427 share these predicted functions? It has previously been observed that vesicles are present in the bdelloplast periplasm (Kaplan et al., 2022), though it could not be determined if these were made by predator or prey, Bd0427 could be involved in this vesicle formation.

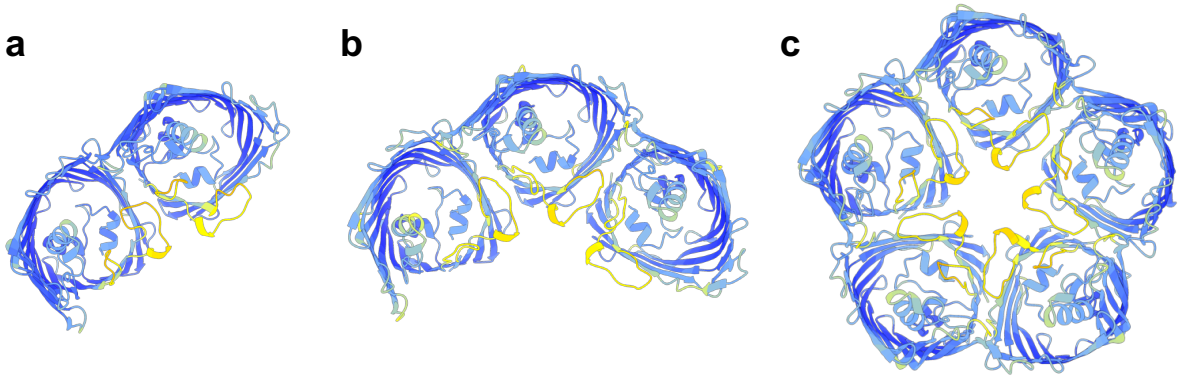


Figure 4-24: AlphaFold Models of Bd0427 Selecting for a Dimer (a), Trimer (b) and Pentamer (c).

In addition to looking at the potential functions of the proteins identified using Foldseek, the oligomeric state of a selection of these proteins were probed using AlphaFold and ColabFold. Initially it was confirmed that the observed oligomeric state of Bd0427 could be determined

using AlphaFold (Figure 4-24) before attempting to fold proteins found using Foldseek. When folding these proteins as oligomers four homologs modelled as pentamers and excitingly it was discovered that several instead modelled as hexamers (Figure 4-25). Some of the proteins would not pack using AlphaFold such as *Treponema* proteins, though TDE_2508 has been shown to form a high molecular weight (~600 KDa) oligomer (Veith et al., 2019). These AlphaFold models alongside the crystal structure suggest that a new class of OMPs have been identified, in which the β-barrels pack to form either pentamers or hexamers.

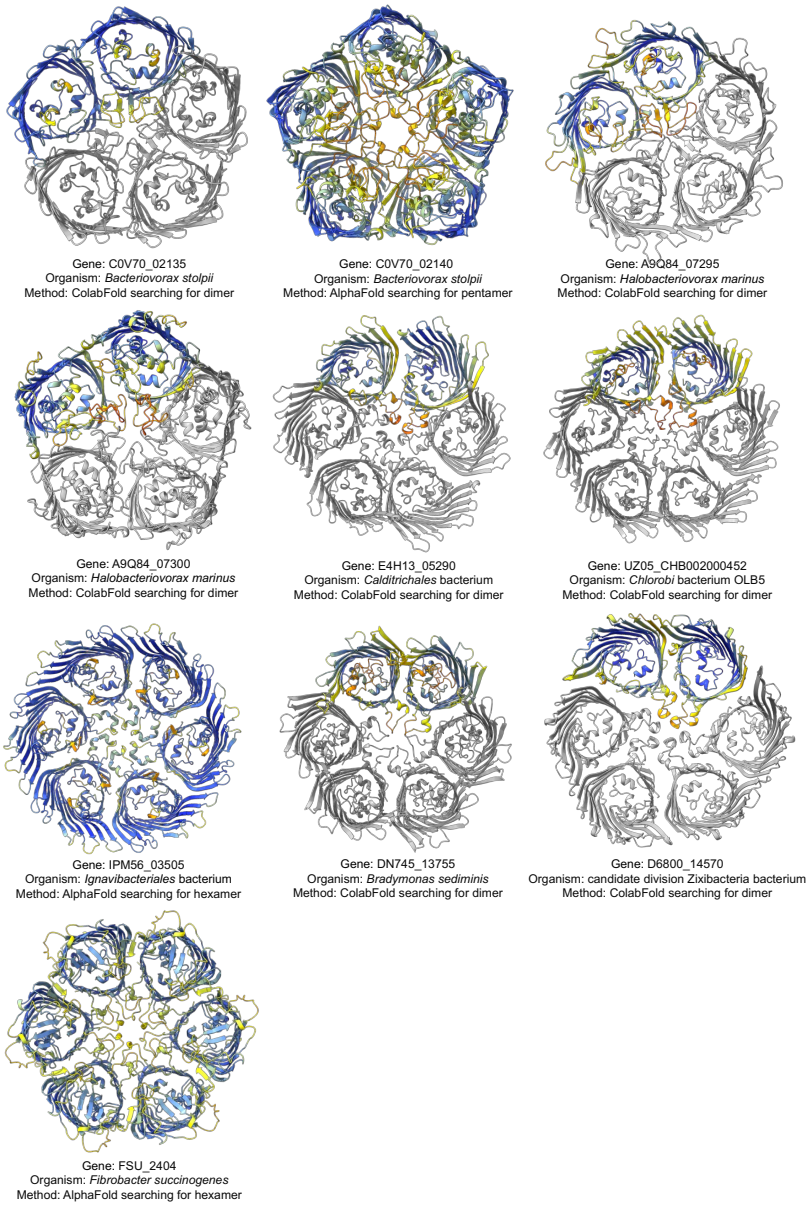


Figure 4-25: Oligomeric State of Bd0427 Structural Homologs, from Foldseek search, Predicted by AlphaFold and ColabFold.

CHAPTER FIVE: VALIDATION OF BD0427 STRUCTURE AND BD0427 CHARACTERISATION

5.1 Foreword

The Bd0427 crystal structure revealed a novel oligomeric state for a 16-stranded β-barrel, which had not previously been observed for other characterised outer membrane proteins. This chapter will focus on validating the observed oligomeric state as well as characterising the protein. To validate the observed oligomeric state a construct with an internal his-tag was designed for expression in the outer membrane of *E. coli* expression strain. The purified product was then used for cryo-EM, to validate the structure. To characterise Bd0427, liposome swelling assays and preliminary electrophysiology experiments were conducted, as well as the generation of an in-frame deletion construct.

5.2 Results

5.2.1 Attempts to Refold Bd0427 Δplug

The Bd0427 Δplug construct was produced by SDM on the Bd0427 ΔSP construct, for expression of Bd0427 Δplug in the cytoplasm where the protein was packaged into inclusion bodies. The inclusion body material was purified and then denatured in 8 M urea, before rapid dilution into detergent-containing buffers, as per the wild-type in Chapter 3. Because Bd0427 ΔSP can be refolded in both 1 % (w/v) DDM and 2 % (w/v) β-OG, this mutated construct was also trialled in these detergents. To determine if the construct had been refolded a semi-native PAGE was run (Figure 5-1). The gel shows that in all the samples the band for Bd0427 Δplug migrated the same distance, and it can be interpreted that the protein did not refold. It was therefore concluded that the protein cannot fold properly without the N-terminal plug.

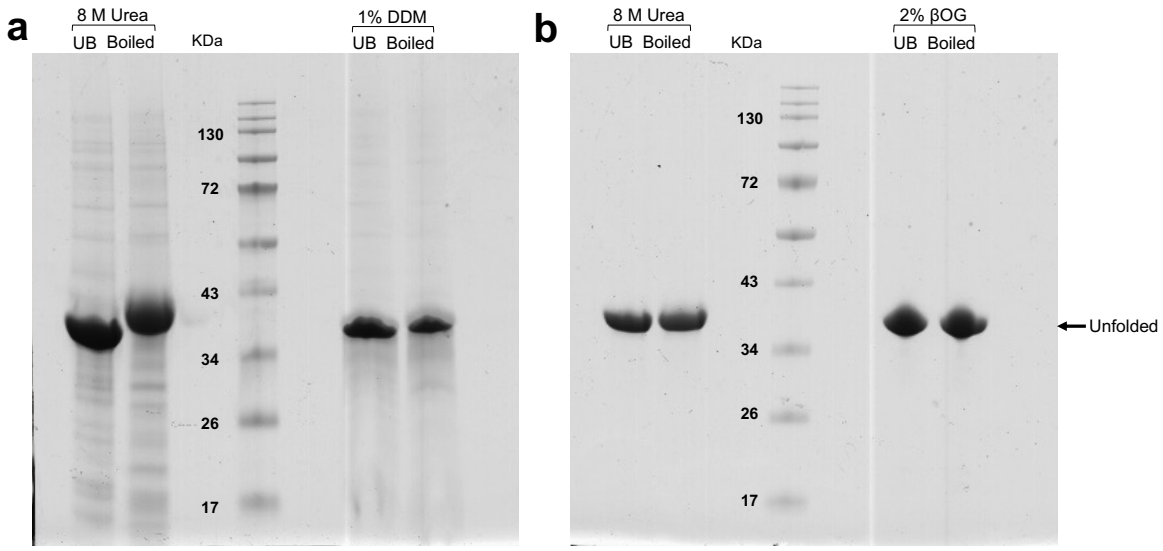


Figure 5-1: Semi-native PAGE of Refold Attempt of Bd0427 Δplug in DDM (a) and β-OG (b). All samples in 1% SDS loading dye; one sample was boiled at 100 °C for 10 min and the other was unboiled (UB). Protein Ladder = Color Prestained Protein Standard, Broad Range (10 250 KDa) (NEB, P7719S).

5.2.2 Refold of Bd0427 Constructs with His-Tag Insertion

To determine if a construct with a His-tag in an internal loop could be used to (eventually) optimise purification of native Bd0427 from the *E. coli* membrane, two constructs (Bd0427 A103His and Bd0427 K260His) were designed. These had His-tags inserted into different loops (loop 2 and 6, respectively) by SDM for initial expression in the cytoplasm for packaging into inclusion bodies. The two constructs would be used in a refold experiment and if either of the constructs could be refolded a second construct for expression in the membrane would be designed.

Following cytoplasmic expression, the resulting inclusion bodies were purified and then denatured in 8 M urea. Refolding was then attempted by rapid dilution into a buffer containing 2 % (w/v) β-OG. The refolding success was observed using a semi-native PAGE. The gel contained bands consistent with folded protein for both constructs (Figure 5-2), indicating both constructs could be refolded in 2 % (w/v) β-OG. The proportion of refolded to unfolded was

better for Bd0427 A103His, therefore this His-tag position was chosen for the native membrane construct design.

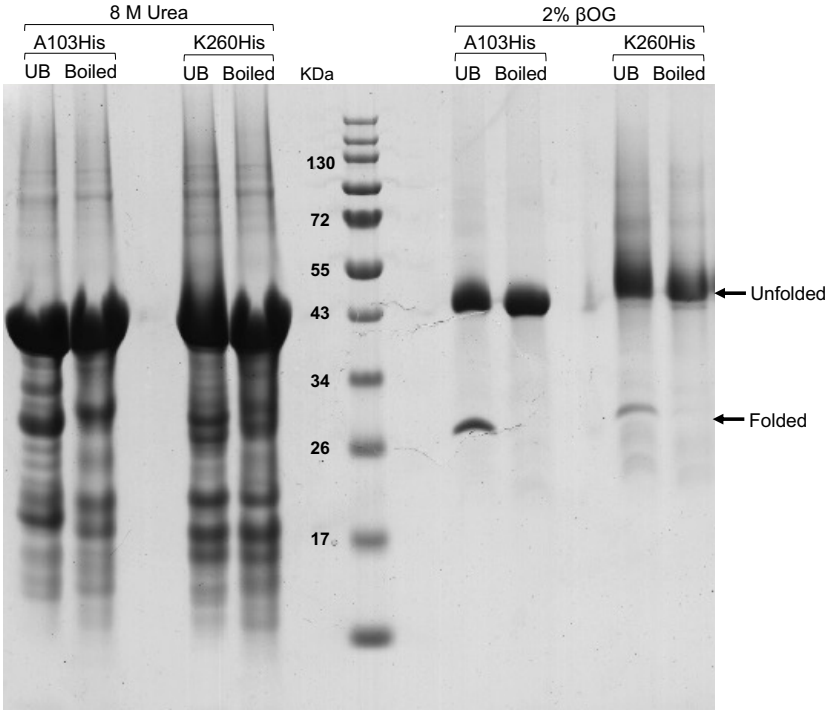
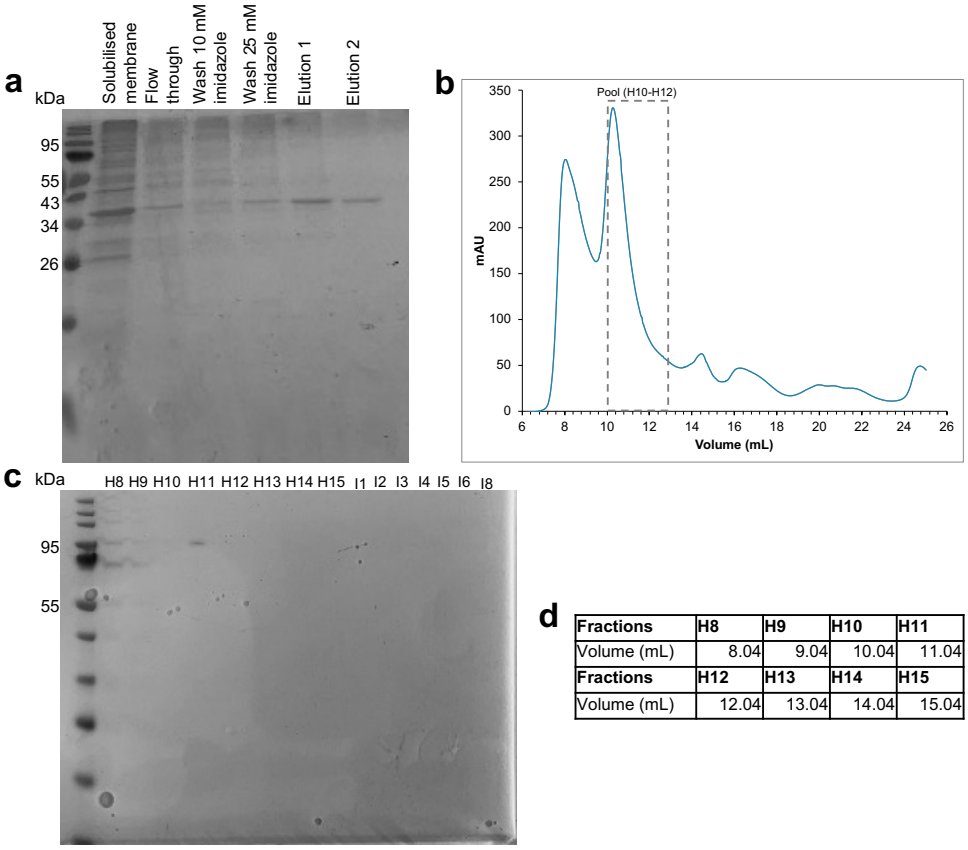


Figure 5-2: Semi-native PAGE of Refold of Bd0427 A103His and Bd0427 K260His in 2 % (w/v) β-OG. All samples in 1% SDS loading dye; one sample was boiled at 100 °C for 10 min and the other was unboiled (UB). Protein Ladder = Color Prestained Protein Standard, Broad Range (10 250 KDa) (NEB, P7719S).

5.2.3 Purification of Bd0427 *pelB* A103His for Cryo-EM

Bd0427 *pelB* A103His was expressed in *E. coli* C43(DE3) membrane and solubilised in 1 % (w/v) DDM. The solubilised protein was initially purified using immobilised affinity chromatography. The purity of the fractions was assessed by SDS-PAGE (Figure 5-3a) with elutions 1 and 2 pooled and concentrated. The protein was then further purified on a Superdex 200 10/300 GL in 0.3 % (w/v) DDM. This resulted in a SEC chromatogram with multiple polydisperse non-baseline resolved peaks (Figure 5-3b). A semi-native PAGE was run of these fractions, which showed that the peak at 10 – 13 mL resulted in a band at the molecular weight of a 95 KDa protein (Figure 5-3c), a band at this distance was also seen for the refolded



Fractions	H8	H9	H10	H11
Volume (mL)	8.04	9.04	10.04	11.04
Fractions	**H12**	**H13**	**H14**	**H15**
Volume (mL)	12.04	13.04	14.04	15.04

Figure 5-3: Purification of Bd0427 *pelB* A103His in 0.3 % DDM. a) SDS-PAGE of IMAC purification, samples boiled at 100 ˚C for 10 min. b) Size exclusion chromatogram from Superdex 200 10/300 GL column eluted in 50 mM tris (pH 8), 150 mM NaCl and 0.3 % (w/v) DDM. c) Semi-native PAGE of SEC peaks. d) Conversion of volume to fractions ran on the semi-native PAGE.

construct. An aliquot (100 µL) of fraction H11 was taken to avoid aggregation and detergent concentration that could occur during concentration for use for cryo-EM.

5.2.4 Sample Preparation and Data Analysis of Bd0427 *pelB* A103His

The sample from H11 was diluted 10-fold in 50 mM tris (pH 8) and 150 mM NaCl to a final concentration of 0.02 mg.mL^{-1} Bd0427 *pelB* A103His in 50 mM tris (pH 8), 150 mM NaCl and 0.03 % (w/v) DDM for addition to a graphene oxide coated grid. The grid was cryo-cooled in liquid ethane ready for screening and data collection.

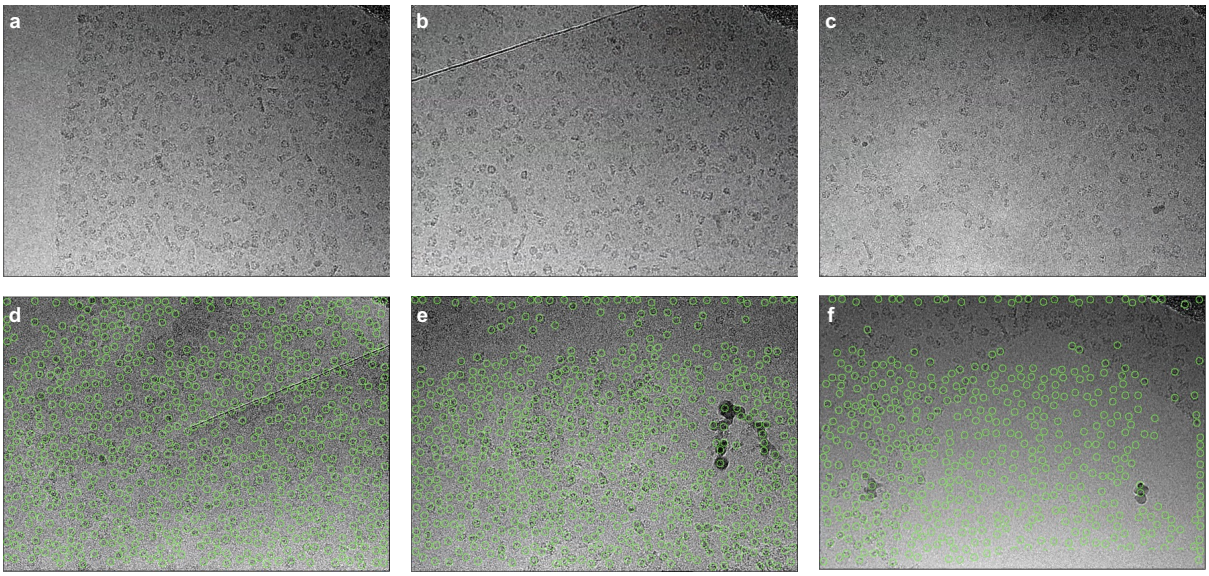


Figure 5-4: Representative Micrographs (a-c) and Particles Picked from Auto-Picking (d-f).

10,312 micrographs were collected and processed with Relion 4.0. After motion-correction and CTF correction, a large number of micrographs were removed as they contained ice crystal formation. This left 4,736 micrographs, representatives of which are shown in Figure 5-4a-c, for use in auto-picking. Auto-picking selected 2,967,865 particles (Figure 5-4d-f), which were extracted for 2D-classification with a mask of 180 Å. From this initial 2D-classification ten subsets were selected (Figure 5-5). The 300,234 particles from this subset selection then underwent a second round of 2D-classification with a 140 Å mask. This produced the 2D-classifications in Figure 5-6. The 2D-classifications show multiple orientations of the protein

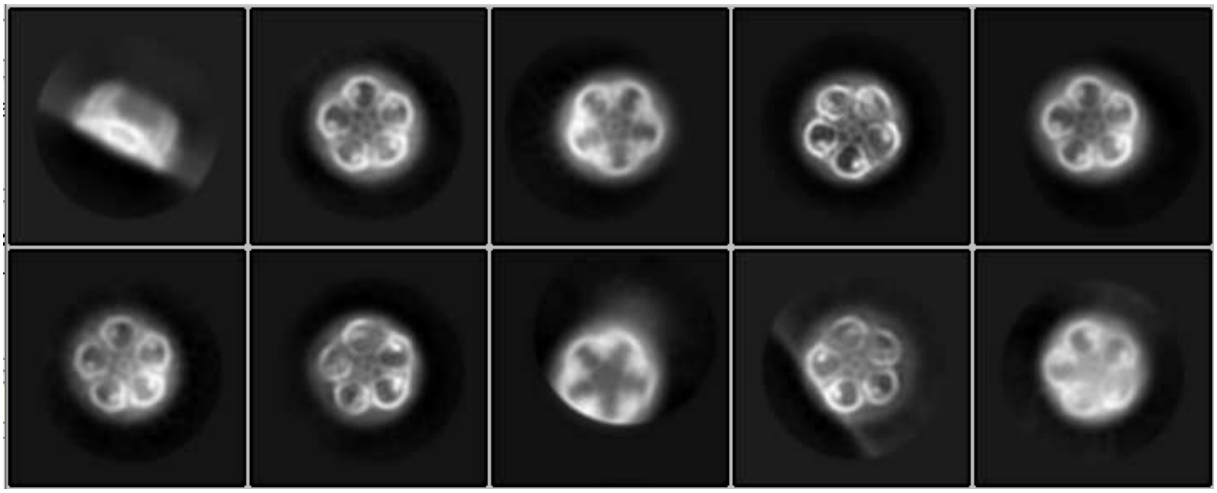


Figure 5-5: Subset Selection from Initial 2D-Classification.

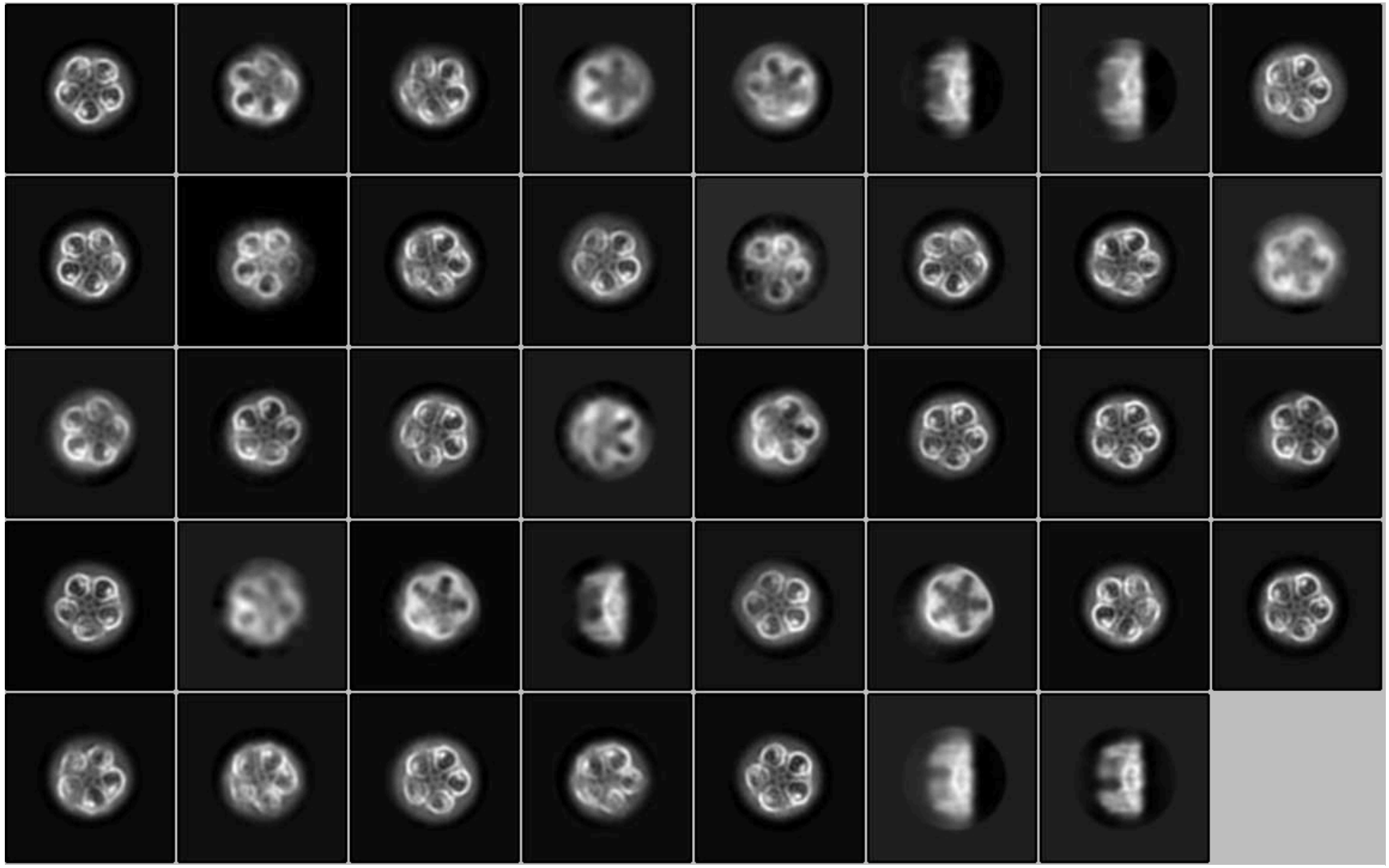


Figure 5-6: 2D-Classifications of Bd0427 *pelB* A103His from Second Round of 2D-Classification.

and contain several classifications showing five rings packed together. This matches the top projection of the homopentameric crystal structure. Density can even be seen for the α-helix of loop 2, within the lumen of the protomers. This provides initial support that the homopentameric state seen in the structure determined by X-ray crystallography is valid.

5.2.5 Cryo-EM Density Map of Bd0427 *pelB* A103His (Collaboration)

The density map and *ab inito* model were generated by our collaborator (GR), using Cryosparc v2.12.0 (Figure 5-7a). In summary 1945480 particles were blob picked from 5,979 micrographs, 2D-classes were generated and inspected (Figure 5-7b) with 160323 particles

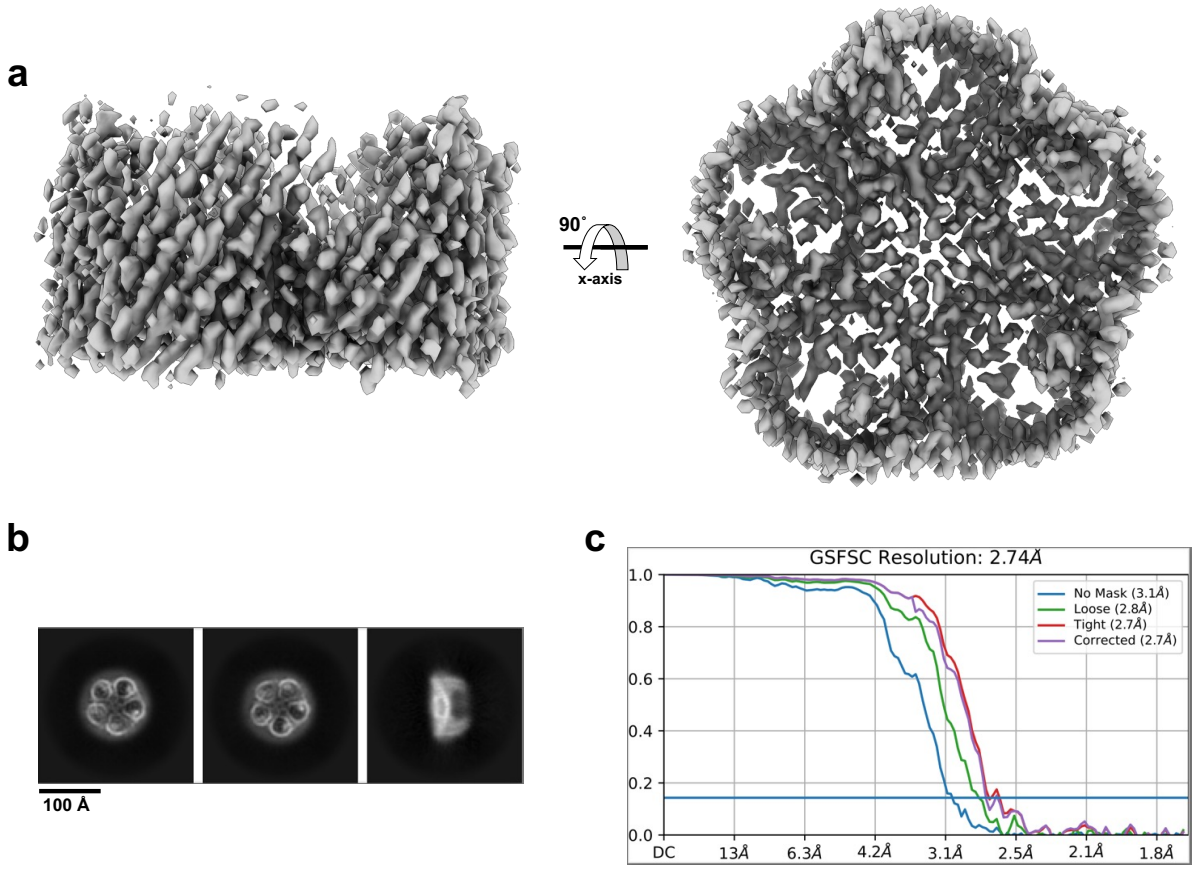


Figure 5-7: Summary of Cryo-EM Density Map Generation. a) The cryo-EM density map of Bd0427 *pelB* A103His (map Lv2). b) Sub selection of 2D-classifications used to generate *ab initio* model. c) The FSC curve for the DeepEMenhancer refinement.

used for *ab initio* model generation. The map was refined using non-uniform refinement imposing C1 symmetry and improved using DeepEMenhancer to a final resolution of 2.74 Å (Figure 5-7c). The map was used along with the sequence to generate an initial model using modelAngelo, which comprised 1735 residues (Figure 5-8a).

5.2.6 Cryo-EM Structure of Bd0427 *pelB* A103His

The final model was manually built into the cryo-EM map using a combination of the initial model from modelAngelo (from collaborator), the Bd0427 crystal form A structure fitted to the map using Phenix 'Dock in Map' and Coot. The model was refined using Phenix Real-space refine (Afonine et al., 2018) resulting in a final model with 1630 residues over five chains (Figure 5-8b), with all correlation coefficients comparing model to data being above 0.65 (Table 5-1). Due to the flexibility of the loop regions, residues 104-114 (chains A, B, C), 104-115 (chain D), 104-116 (chain E), 136-137 (chain A), 135-137 (chain E) and 251-256 (chains A-E) could not be resolved.

Table 5-1: Correlation Coefficient Model to Data

CC_{mask}	0.69	**CC_{peaks}**	0.67
CC_{box}	0.71	**CC_{volume}**	0.69

The overall fold of the cryo-EM structure is a 16-stranded β-barrel, with an N-terminal plug (residues 21-32). The oligomeric state is a homopentamer, as observed in the 2D-classifications. This model is composed of the same overall fold as the crystal structure. Unlike the inclusion body-expressed refolded protein that produced the crystal structure, this protein was expressed in the membrane and solubilised in a different detergent (the crystal structure was from β-OG refolded material). This removes the possibility that the pentameric state of the crystal structure is an artifact of either the refolding protocol or crystal packing and indicates that this is the first OMP that has been identified as a homopentamer.

5.2.7 Cryo-EM Structure Bd0427 *pelB* A103His Analysis

In the cryo-EM model there is variation in the secondary structure assigned to each chain, with the main differences in the N-terminal plug (1, 2 or 3 α-helices), the helix (11-13 residues) in loop 2, with chain D also containing a second 3-residue helix, and β-strands 6, 7, 8 and 14, with variation in length as well as disruption to the β-strand (Figure 5-9a). To probe the similarity between each chain the N-terminal plug alone and whole chain (325 residues) were

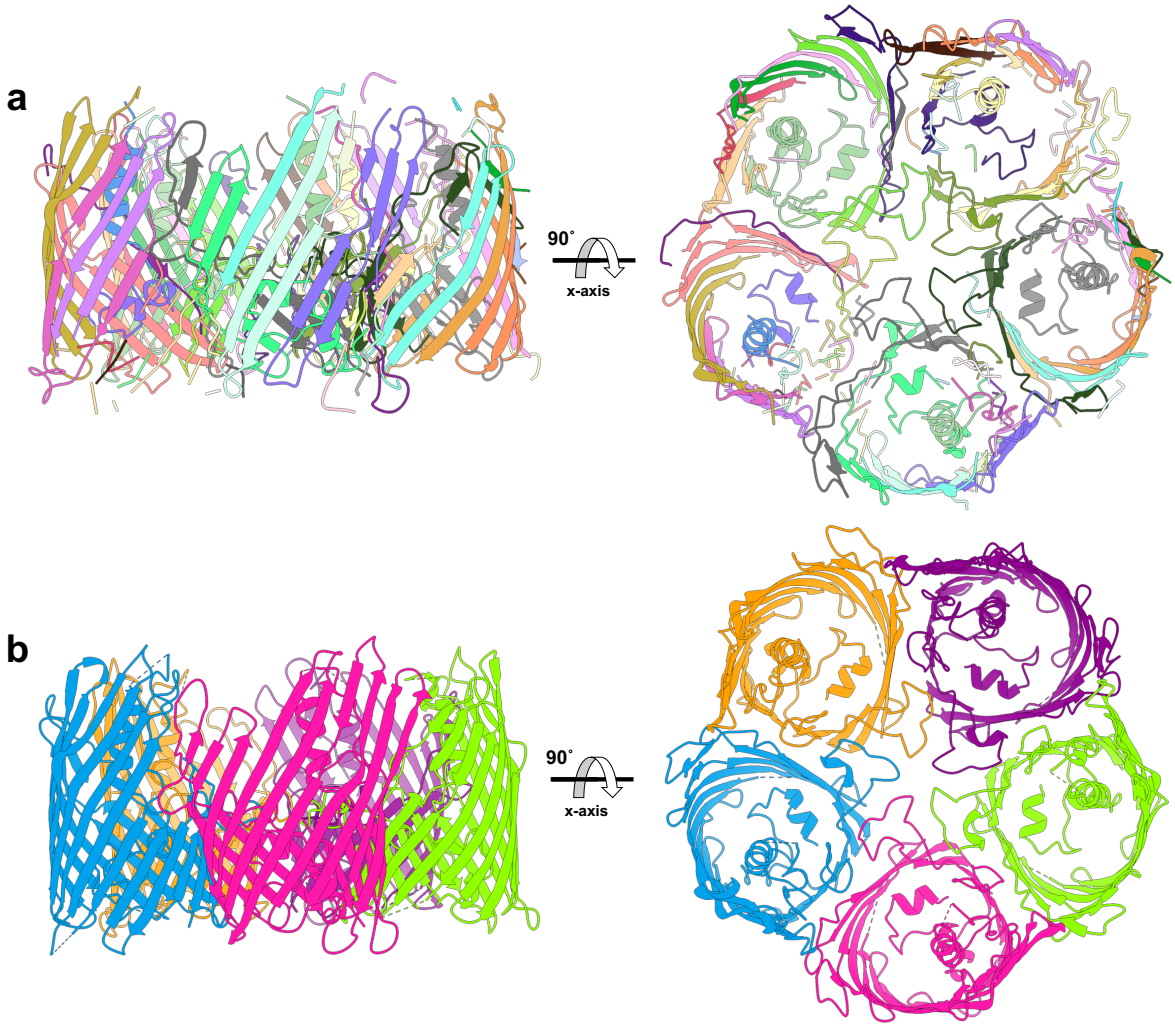


Figure 5-8: Models of Bd0427 A103His. a) Initial model generated using ModelAngelo, coloured by chain. b) Final model of Bd0427 A103His, coloured by chain.

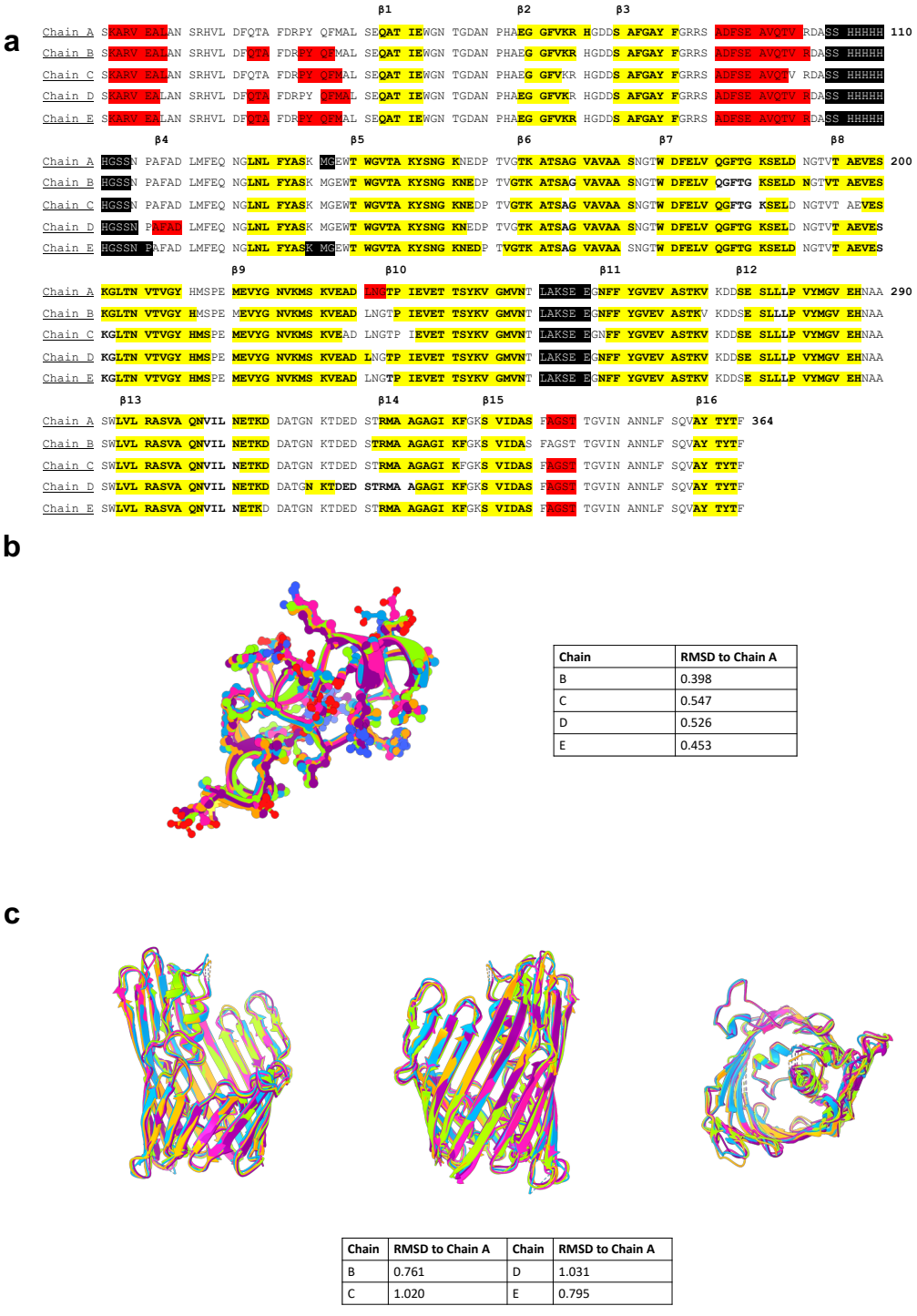


Figure 5-9: Secondary Structure Assignment and Superimposition of each Chain of Bd0427 A103His. a) Secondary structure assignment of each chain. β-strands – yellow, α-helices – red, residues that could not be resolved highlighted in black. b) Superimposition of the N-terminal plugs from Chains B-E to Chain A, including reported RMSD values. c) Superimposition of the Chains B-E to Chain A, including reported RMSD values.

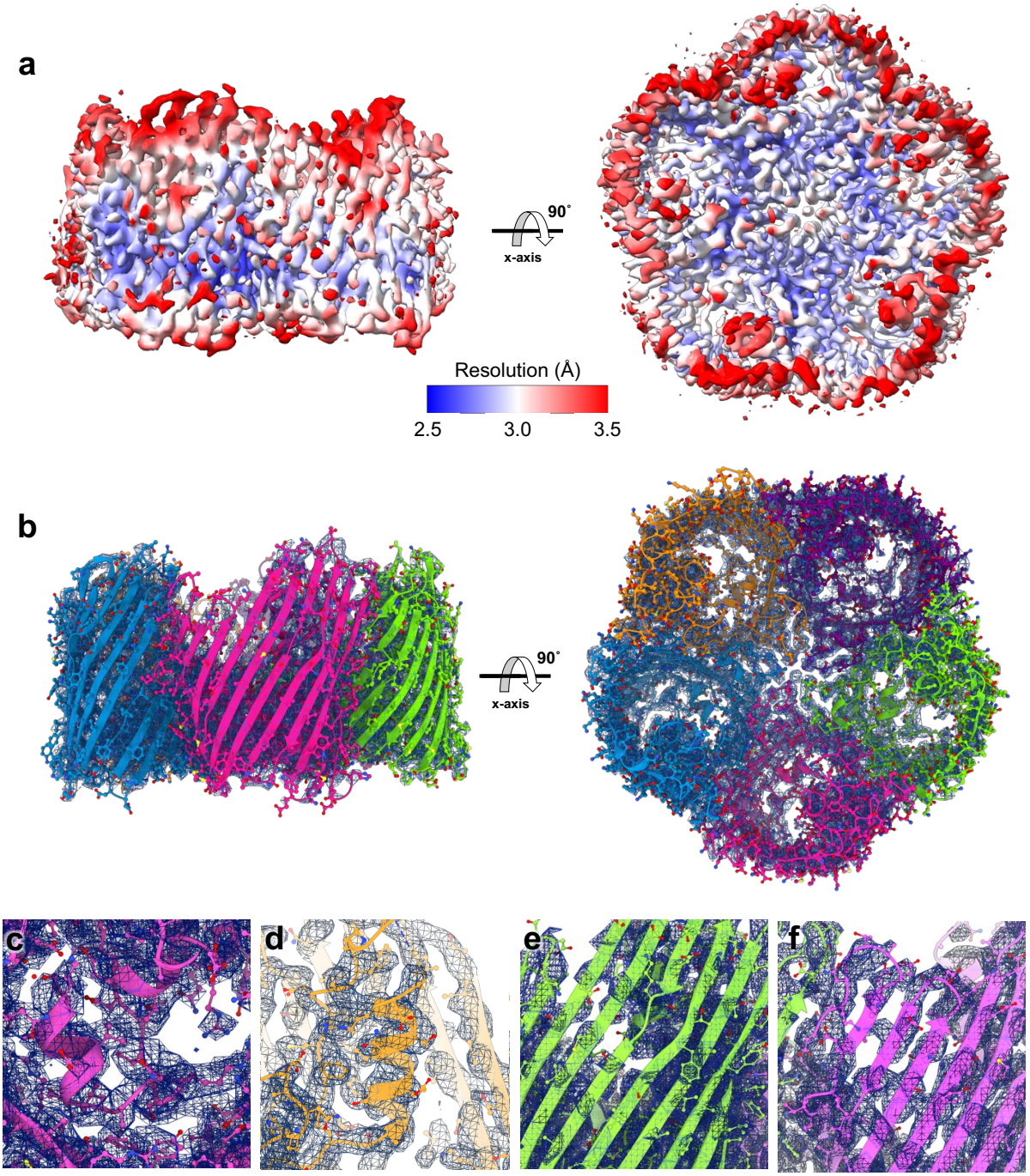


Figure 5-10: Bd0427 A103His Local Resolution and Fit to Cryo-EM Density Map. a) A local resolution map (lv. 0.15) of Bd0427 A103His, showing areas of higher and lower resolution. b) Bd0427 A103His model fit into the cryo-EM density map (lv. 0.15). c-f) Regions of Bd0427 A103His model in the cryo-EM density map. N-terminal plug of chain E (c), loop 2 of chain B (d), Outer face of chain D (e), β-strands 11-14 of chain C (f). Map lv. 0.1.

superimposed with Chain A acting as the reference (Figure 5-9b-c). This showed that despite the differences in secondary structure assignment the N-terminal plugs of each chain are very similar, with all the RMSD values below 0.55 Å (Figure 5-9b). The RMSD values from the

superimposition of all the residues in each chain, to Chain A, were larger (between 0.7 Å and 1.1 Å) representative of a greater difference across the whole chain. These differences mostly occur in the loop regions of the structure (Figure 5-9c), with a small level of shift in the position of the β-strands and the helix of loop 2. To assess the range of resolutions within the map, the local resolution was calculated using Cryosparc (Figure 5-10a). This shows that residues of the β-strands within the hydrophobic band are at the highest resolution at ~2.6 Å, with the loop regions being at the lowest resolution ~ 3.5 Å. It can also be seen that the resolution is lower for the top (above hydrophobic band) of β-strands 7-14 and the N-terminal plug at ~3.0 Å, this suggests that these parts are more mobile, resulting in lower resolution. As a result of the variable resolution there are regions where there is density for both the backbone and sidechains of the residues (Figure 5-10c), regions where the density for the backbone and bulky sidechains can be seen (Figure 5-10d-f) and regions where only density for backbone can be seen (Figure 5-10e).

5.2.8 Unmodelled Density in Cryo-EM Map

Within the cryo-EM density map there are 16 unmodelled densities (Figure 5-11). Five of these are in the lumen of each chain towards the bottom of the loop 2 α-helix (Figure 5-11a), all but the density in chain A (blue) form long chain-like density. These densities correspond to a volume of positive density seen in the crystal structure FO-FC map and in common with this the ligand would be able to form hydrogen bonds, salt-bridges and/or electrostatic interactions with the sidechains in proximity, as well as hydrogen bonding to the backbone of β-strands 7 and 8. The remaining 11 unmodelled densities, are found in the pocket under the extended loop 8 and represent long-chain like ligands. The densities seen here closely resemble the 10 densities modelled as octane (representative of the detergent tails of β-OG) in the crystal

structure, and as such it is likely that these densities also represent ordered detergent tails of DDM. As with the crystal structure the density is surrounded by hydrophobic sidechains, enabling hydrophobic interactions with the detergent tail.

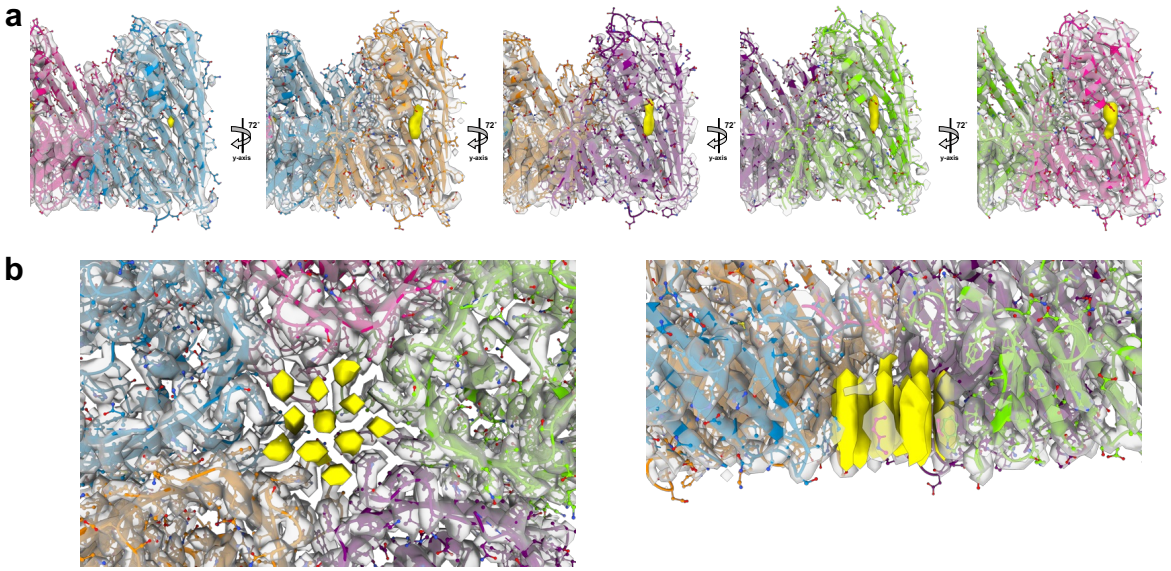


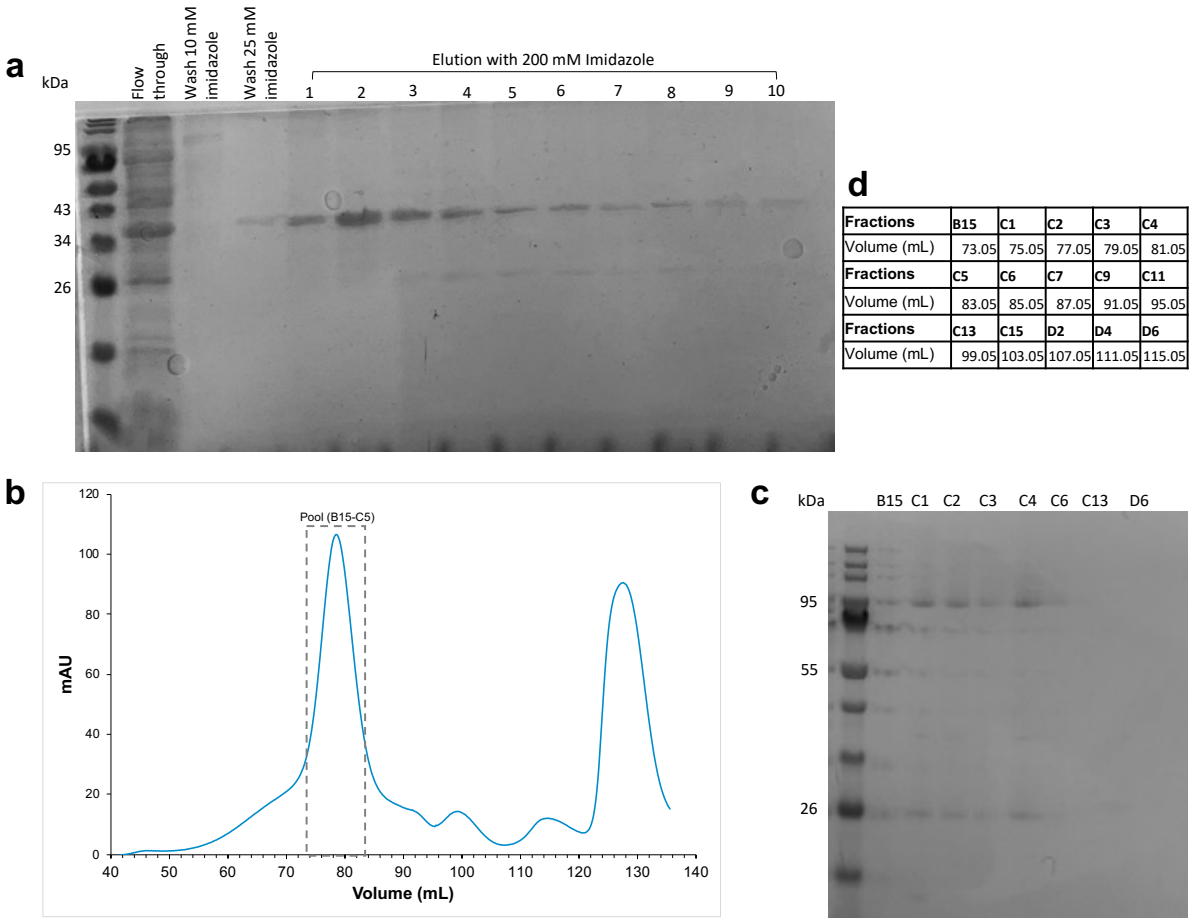
Figure 5-11: Unmodelled Density in the Cryo-EM Map. a) Unmodelled density (yellow) below the α-helix from loop 2, packing against the barrel wall. b) Unmodelled density (yellow) in the hydrophobic pocket below loop 8. Map lv. 0.15.

5.2.9 Liposome Swelling Assay

To determine if Bd0427 could transport substrates, and to see if there was any specificity for a particular class or size of substrate, a liposome swelling assay was conducted. This compared the swelling rates of proteoliposomes reconstituted with Bd0427 to proteoliposomes reconstituted with OmpF, a well characterised model porin.

For the liposome swelling assay Bd0427 *pelB* A103His was solubilised as in 5.2.3 and purified in 0.03 % (w/v) DDM using IMAC and SEC (HiLoad 16/60 Superdex 200 pg column) (Figure 5-12). The protein eluted from the SEC column in multiple polydisperse non-baseline resolved peaks (Figure 5-12b). The fractions of the peaks were run on a semi-native PAGE, which gave bands at the molecular weight of 95 and 26 KDa for the fractions B15-C4 (Figure 5-12c). These

fractions were pooled and concentrated to 5.6 mg.mL^{-1}. Bd0427 *pelB* A103His and OmpF were reconstituted into proteoliposomes, as in section 2.10, alongside the production of liposomes with the same volume of DDM buffer.



d

Fractions	B15	C1	C2	C3	C4
Volume (mL)	73.05	75.05	77.05	79.05	81.05
Fractions	**C5**	**C6**	**C7**	**C9**	**C11**
Volume (mL)	83.05	85.05	87.05	91.05	95.05
Fractions	**C13**	**C15**	**D2**	**D4**	**D6**
Volume (mL)	99.05	103.05	107.05	111.05	115.05

Figure 5-12: Purification of Bd0427 A103His in 0.03 % DDM. a) SDS-PAGE of IMAC purification, samples boiled at 100 ˚C for 10 min. b) Size exclusion chromatogram from HiLoad 16/60 Superdex 200 PG column eluted in 50 mM tris (pH 8), 150 mM NaCl and 0.03 % (w/v) DDM. c) Semi-native PAGE of SEC peaks. d) Conversion of volume to fractions ran on the semi-native PAGE.

The liposome control and proteoliposomes reconstituted with OmpF or Bd0427 *pelB* A103His were then used in LSAs, as described in Section 2.10, producing an initial mean swelling rate for each substrate (Table 5-2). The swelling rate through OmpF was set as the positive control, with the uptake of L-arabinose set to 100 %. All other data was reported relative to this value.

Initially it could be seen that Bd0427 was not as permeable as OmpF, with the relative permeability of each substrate lower for proteoliposomes containing Bd0427 than for OmpF (Figure 5-13a). From this assay it was clear that glucose and maltose are not readably transported by Bd0427, with the relative permeability of these two substrates being -3 % and 0 %, respectively (Figure 5-13a). The relative permeabilities of substrates glycine and alanine given by Bd0427 were 47 %, indicating that these substrates are able to diffuse through Bd0427 but to a lesser extent than OmpF (Figure 5-13a). If the relative permeability of substrates through Bd0427 are plotted against the molecular weight of the substrate, a trend can be seen (Figure 5-13b). This indicates that as the molecular weight of the substrate increases the

Table 5-2: Initial Rates and Relative Permeability of Substrates through OmpF and Bd0427 *pelB* A103His

Substrate	Mw (g/mol)	Initial Swelling Rate (OD400/min)		Relative Permeability (%)	
		OmpF	Bd0427	OmpF	Bd0427
12 mM L-arabinose	150.13	0.0640±0.007	0.0080±0.000	100	13
15 mM Glycine	75.05	0.0698±0.004	0.0298±0.007	109	47
15 mM Alanine	89.09	0.0820±0.010	0.0300±0.004	128	47
12 mM Leucine	131.17	0.0580±0.010	0.0100±0.000	91	16
8 mM Lysine	146.19	0.0260±0.031	0.0060±0.004	69	9
12 mM Glucose	180.16	0.0520±0.006	-0.0020±0.000	81	-3
15 mM Maltose	342.30	0.0180±0.000	0.0002±0.000	28	0

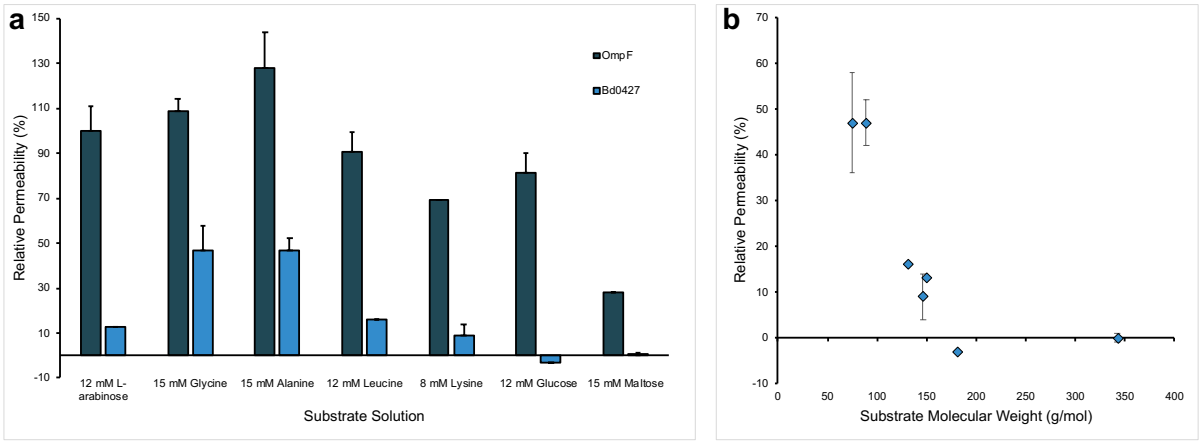


Figure 5-13: Permeability of Proteoliposomes Reconstituted with Bd0427 *pelB* A103His. The relative permeability of seven different substrates through Bd0427 *pelB* A103His compared to the permeability through OmpF (a) and molecular weight of the substrates (b). All permeability has been compared to the swelling rate of L-arabinose through OmpF. The error bars show the standard deviation across the repeats.

permeability decreases, suggesting Bd0427 selects substrates based on size with a potential cut off between 150 g.mol^{-1} and 180 g.mol^{-1}.

5.2.10 Preliminary Characterisation by Electrophysiology (Collaboration)

Preliminary characterisation of Bd0427 ΔSP and Bd0427 *pelB* A103His by electrophysiology was conducted with support from our collaborators (BvdB). Both constructs were purified using SEC in 50 mM Tris (pH 8), 150 mM NaCl and DDM at 0.3 % or 0.03 % (w/v), respectively, (section 3.2.1.2.2 and 5.2.9) and concentrated to ~ 5 mg.mL^{-1}. The protein was serially diluted in 1 % Genapole X100 before addition to the *cis* side of the cuvette (Figure 2-8).

Bd0427 ΔSP showed channel insertion and opening, identified by a change in amplitude, into the artificial membrane at a 10^{-6} dilution (Figure 5-14). From these data, it could be seen that there were short openings, observed as spikes, at ~5 and ~40 pA, at 50 mV, which corresponds

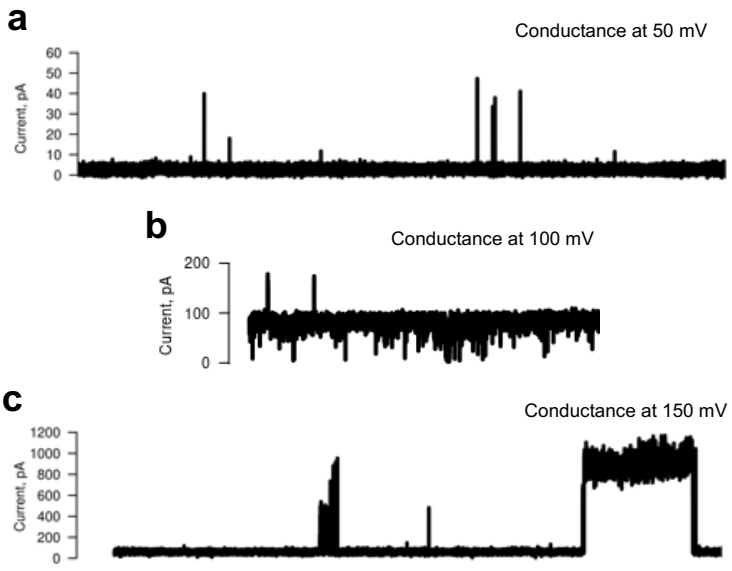


Figure 5-14: Representative Conductance of Bd0427 ΔSP. All conductance measurements were measured in 10 mM HEPES (pH 7) and 1 M KCl. a) Conductance at 50 mV showing short opening events. b) Conductance at 100 mV showing an open channel.c) Conductance at 150 mV showing short openings as well as an opening of a highly conducting state.

to channel conductance value of ~100 pS and ~1 nS, respectively (Figure 5-14a), as well as more sustained openings (Figure 5-14b-c), at 100 mV and 150 mV. The opening at 100 and 150 mV were at ~ 100 pA and ~900 pA, respectively, again giving a conductance value of ~1 nS but also revealing a larger opening with a conductance value of ~6 nS.

Bd0427 *pelB* A103His inserted into the artificial membrane at a dilution of 10^{-3} and 10^{-4} fold (Figure 5-15). In these data, channel opening with a conductance value of ~1 nS (opening at 50 pA at 50 mV and 90 pA at 75 mV) was seen at multiple potentials (Figure 5-15), along with examples of lower conducting states at ~600 pS as seen in the trace at 75 mV (Figure 5-15b). In the conductance trace at 50 mV (Figure 5-15a), a higher conductance was observed towards the end of the measurement, representing a highly conducting state of ~3 pS. This could be representative of a larger opening of the channel or the observation of an additional channel.

From the preliminary conductance measurements of both constructs it was observed that Bd0427 forms small openings with conductance between 100 pS and 1 nS, with the possibility

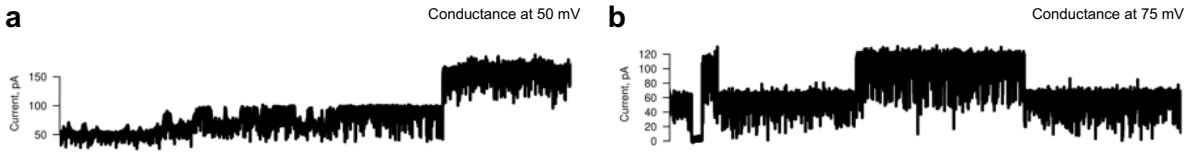


Figure 5-15: Representative Conductance of Bd0427 *pelB* A103His. All conductance measurements were measured in 10 mM HEPES (pH 7) and 1 M KCl. a) Conductance at 50 mV showing multiple opening events, with either a new insertion event towards the end or an opening to a higher conductance. b) Conductance at 75 mV, showing partial opening, closing, and opening at a higher conductance.

that larger openings are formed with conductance between 3 and 6 nS. To fully characterise the channels conductance, more measurements need to be taken.

5.2.11 Generation of *B. bacteriovorus* In-Frame Deletion Strain (Collaboration)

Construction of *B. bacteriovorus* in-frame deletion strain of *bd0427* (genome position in *B. bacteriovorus* HD100 seen in Figure 5-16a) involved the generation of a suicide plasmid

containing fragments of 1 kb upstream and 1 kb downstream of *bd0427* (Figure 5-16b) to enable homologous recombination and deletion of the gene. The two fragments were successfully amplified by PCR, as shown by the bands that ran at the same distance as the 1 kb band of the marker (Figure 5-16c). These fragments were used with the cut plasmid (pK18*mobsacB*), at a 5:1 and 7:1 (insert:vector) ratio, in Gibson Assembly. This resulted in 12 white colonies from the 5:1 reaction and 8 colonies from the 7:1 reaction, using colony PCR (Figure 5-16d) 11 of these colonies were shown to have the insert. Colony 7 (7:1) plasmid was sequenced, and this confirmed that the plasmid had been successfully constructed.

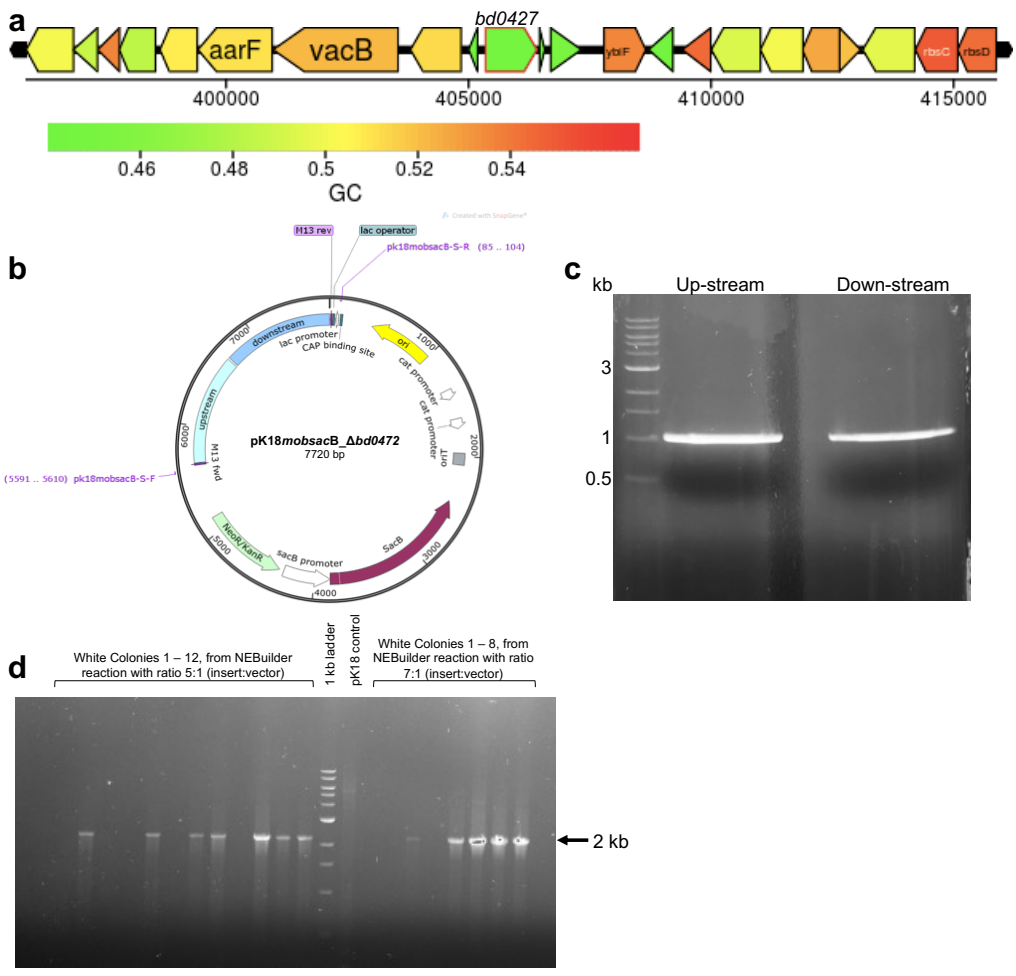


Figure 5-16: Position of *bd0427* in *B. bacteriovorus* HD100 Genome, Plasmid Map and PCR Gels. a) Regions surrounding *bd0427* in the *B. bacteriovorus* genome. b) Plasmid map of pK18*mobsac*B_Δ*bd0427* including showing where the sequence primers anneal. c) Agarose gel of amplification of upstream and downstream fragments. d) Colony PCR gel confirming positive Gibson Assembly.

The construct was then transformed into *E. coli* S17-1 for conjugation into *B. bacteriovorus* HD100 by our collaborators (RES). Initially merodiploid strains were produced but following multiple rounds of growth on sucrose, providing a selection pressure to either integrate the fragments from the plasmid or remove the plasmid from the bacterium, a second cross-over event occurs. The in-frame deletion was initially confirmed by PCR before the genome was sequenced. The sequencing results confirmed the in-frame deletion strain had successfully been generated. As an in-frame deletion was successfully generated *bd0427* cannot be essential.

5.3 Discussion

5.3.1 Bd0427 Cryo-EM Structure has the same Oligomeric State and Fold as Bd0427 Crystal Structure

The cryo-EM structure of Bd0427 is a homopentamer, identical in nature to the crystal structure. This validates the previously unseen oligomeric state for 16-stranded β-barrels in the outer membrane. The construct for the cryo-EM structure has been expressed in the membrane of an *E. coli* expression strain and purified using DDM. This is an independent method to the one used for the crystal structure of refolded Bd0427 purified in β-OG. Therefore, the cryo-EM

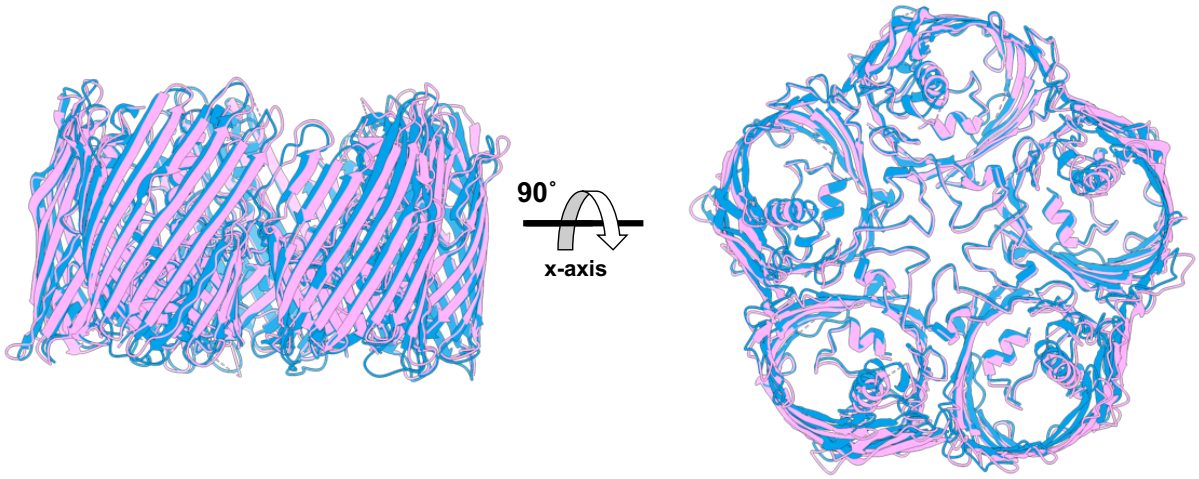


Figure 5-17: Superimposition of Bd0427 Crystal Structure (blue) and Cryo-EM Structure (pink).

structure validates the oligomeric state of Bd0427, removing the possibility that it is an artifact of refolding or crystal packing.

Comparing the two structures revealed that they have a very similar overall fold and can be superposed with an RMSD value of 0.62 Å (323 C_α pairs) (Figure 5-17). The main differences are in the secondary structure assignment and the position of the loops and turns, particularly loops 1, 2 and 7, which are all large loops of 11 - 30 residues. The differences in secondary structure assignment are found in the extracellular facing regions of β-strands 7-13, and all α-helices. These are all found within the lower resolution regions of the cryo-EM map, which could explain these variations (Figure 5-10a). The disparity in the position of loops and turns is caused by their inherently flexible nature and is commonly observed when comparing similar structures. In summary the cryo-EM structure did not provide us with any significant additional information about the fold of Bd0427, due to its similarity with the previously discussed crystal structure.

5.3.2 Potential Ligand Binding Below α-helix from Loop 2

In both the Bd0427 crystal and cryo-EM structures there is a volume of chain-like density below the α-helix from loop 2 (Figure 5-18). As this volume is found in both structures it is probable that this density represents a ligand binding site. In both structures (Figure 5-18) the ligand would be in proximity to form hydrogen bonds with Arg88 and the nitrogens and carbonyls of the backbone, with the crystal structure showing additional hydrogen bonds could be formed with the ligand to Lys135 and Asp42 (Figure 5-18a). Given the surrounding hydrophilic environment it is unlikely to represent a detergent molecule, and with no other chain-like molecules within the buffers the volume has remained unmodelled.

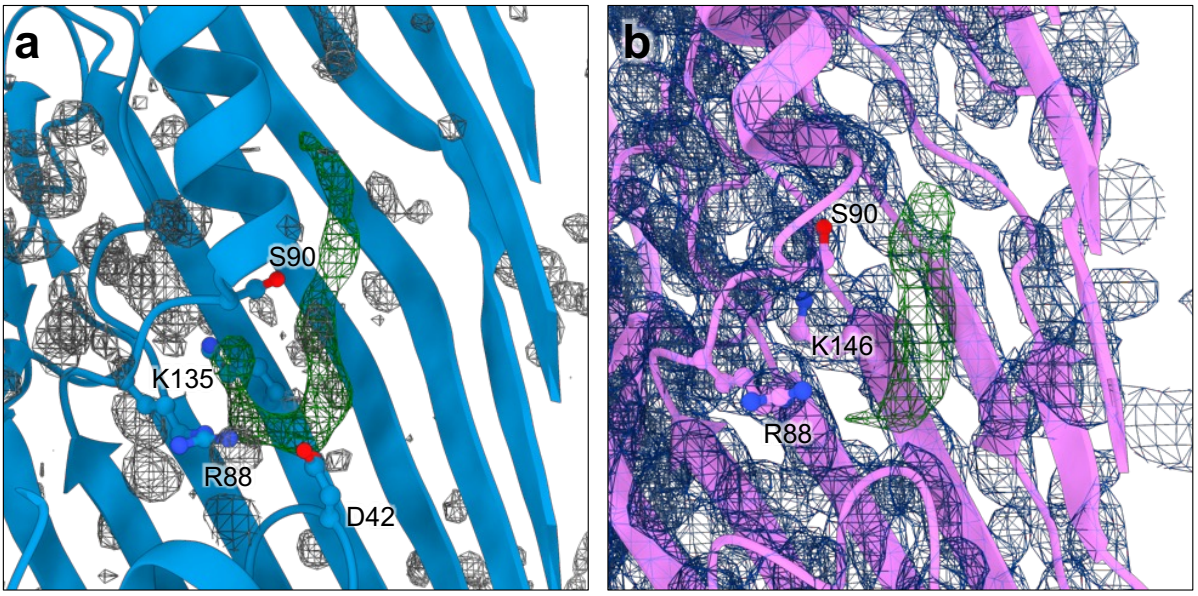


Figure 5-18: Potential Ligand Binding Sites. a) Bd0427 crystal structure with mFo-DFc map (3.0 σ), with volume of positive density coloured in green. b) Bd0427 cryo-EM structure in EM density map (map lv. 0.15) with unmodelled blob coloured in green.

5.3.3 Bd0427 Forms a Channel

The crystal and cryo-EM structure of Bd0427 both contain an N-terminal plug, that when viewed as a surface model shows no *'open'* channel (Figure 5-19), this was confirmed for the crystal structure using MOLEonline (Figure 4-12).

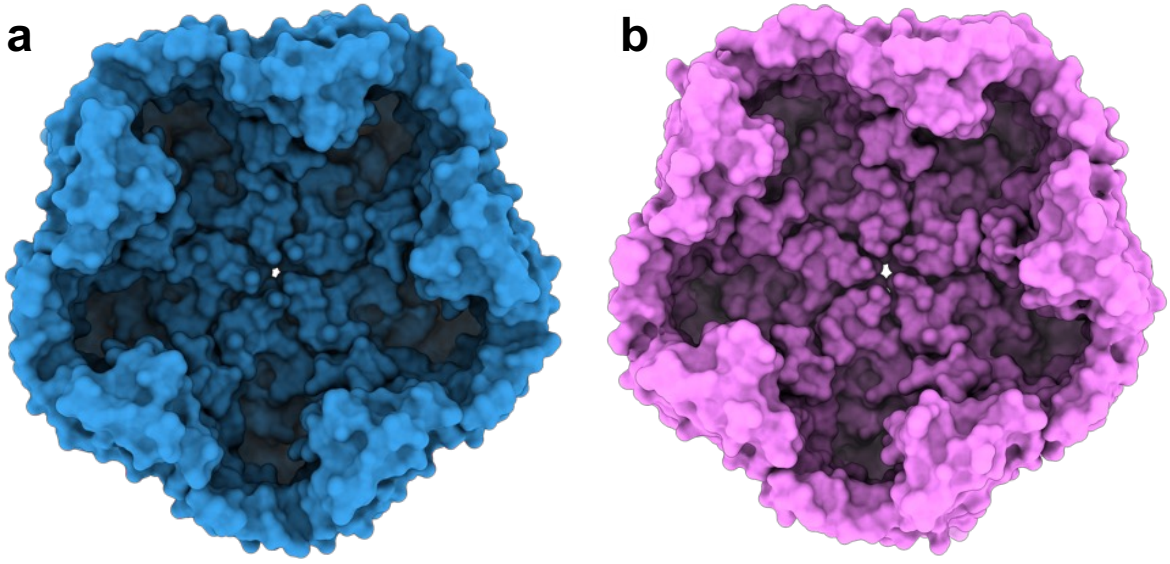


Figure 5-19: Surface Models of Bd0427 Crystal (blue) and Cryo-EM (pink) Structures.

However, the liposome swelling assays indicate that Bd0427 can enable diffusion of glycine and alanine into a proteoliposomes leading to liposome swelling (Figure 5-13). This provides the first evidence that Bd0427 could form a channel, though this channel is most probably smaller than OmpF, as OmpF is able to transport substrates with larger molecular weights. Bd0427 also enables diffusion of glycine and alanine to a lesser extent, suggesting Bd0427 is less permeable compared to OmpF. Although the liposome swelling assay provided some interesting results it is important to mention some of the limitations with the method used for this technique. During the preparation of the proteoliposomes the proteoliposomes are dried under vacuum overnight which could damage the protein, there is also no extrusion of the liposomes to select for liposomes of a specific size, meaning there will be proteoliposomes of different sizes in the assay. These could have an impact on the results, however all proteoliposomes were treated the same.

The preliminary electrophysiology experiments also support that Bd0427 forms a channel, observed by changes in amplitude across the membrane as ions pass from one side of the cuvette to the other. As this is only preliminary data based on a small number of channel insertions firm conclusions cannot be drawn. To provide conclusive data more channel insertions need to be collected with the conductance values compared to see how many channels were observed with specific conductance values. Nevertheless, these preliminary data indicate that Bd0427 could form a channel with an approximate conductance value of between 0.1 – 1 nS, with the potential of forming a larger channel with an approximate conductance value between 3 – 6 nS. These larger conductance values could also represent multiple channels opening at the same time or additional channel insertion events.

The OmpF protomer forms a channel with a conductance of ~ 1 nS (Liko et al., 2018; Phale et al., 1997; Wang et al., 2020), this is similar to the proposed smaller conductance value of

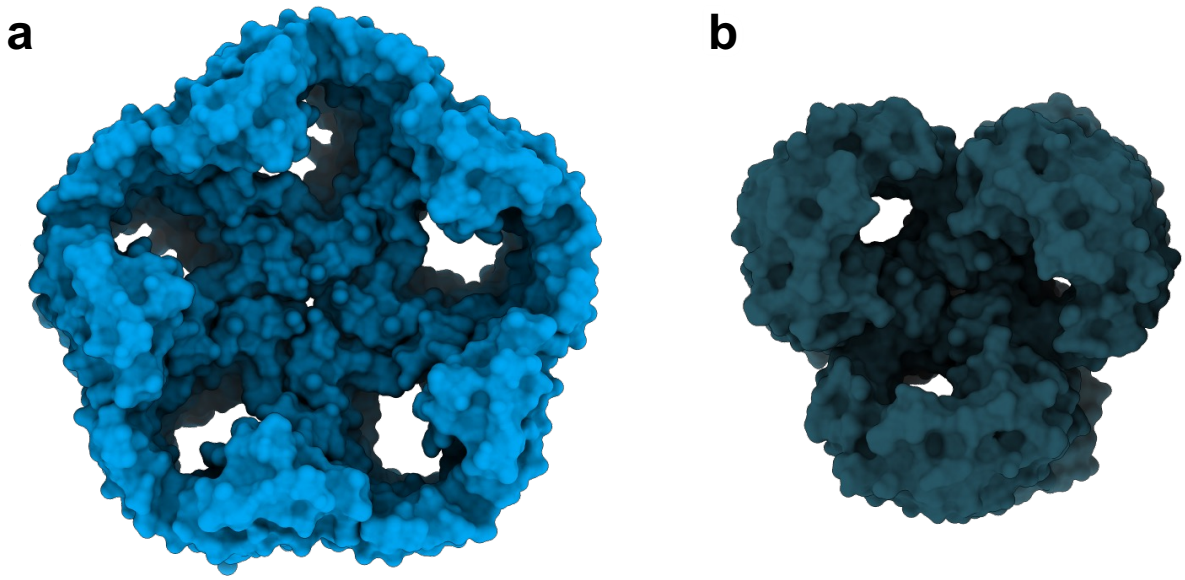


Figure 5-20: Surface of Bd0427 with N-terminal Plug Deleted (blue) and OmpF, 2OMF (teal). Proteins are in scale to each other.

Bd0427. Comparing the OmpF structure to Bd0427 with the residues for the N-terminal plug deleted (Figure 5-20), it would suggest that smaller conductance value does not represent the entire plug being removed, as this channel is observably larger than OmpF. It is a possibility that a larger conductance value could represent displacement of the whole plug, though as previously discussed the larger change in amplitude could be caused by other protomers in the homopentamer opening or an additional insertion event.

The main conclusion from these two experiments is that Bd0427 is capable of forming channels with the liposome swelling assays indicating Bd0427 allows diffusion of glycine and alanine.

5.3.4 *bd0427* is not an Essential Gene

The in-frame deletion of gene *bd0427* in *B. bacteriovorus* was successfully generated, this indicating that Bd0427 most probably does not perform an essential function in the *B. bacteriovorus* life cycle. Though it is possible that the function of the deletion of *bd0427* is compensated by another unidentified gene.

In terms of identification of a function for Bd0427, assuming the gene is not compensated for, it makes it unlikely that Bd0427 acts as the sole method for export of nutrients from the cytoplasm to the periplasm as this would be an essential step in predation. It is possible that Bd0427 could be responsible for disruption of the proton motive force across the IM, as this may not be critical to completing the life cycle. There is also a high possibility that the protein completes a yet unidentified function particularly as the Bd0427 fold has homologs outside of the world of micro-predators. Further investigation of the in-frame deletion strain is required to ascertain if a phenotype can be observed, perhaps in competition with wild-type strains. n with wild-type strains.

CHAPTER SIX: STRUCTURAL CHARACTERISATION OF BD1904 FAMILY

6.1 Foreword

Upon entry into the bdelloplast the most upregulated gene is *bd1904* (Lambert et al., 2010). This gene encodes a soluble protein with a long signal peptide, destined for the periplasm. Proteins with long signal peptides have previously been shown to delay protein folding (Jiang et al., 2011; Szabady et al., 2005), potentially preventing the protein being active in the cytoplasm. Despite generation of an in-frame deletion strain of *bd1904* in *B. bacteriovorus* (Sockett, unpublished) and the structure being characterised (Harding, unpublished), revealing



Figure 6-1: Bd1904 Structure. Structure from previous work by Harding (unpublished).

a structure with three β-hairpin arms, that come off a core containing a β-sheet (Figure 6-1), the function remains unknown. From a PSI-BLAST (section 6.2.1) three additional proteins in *B. bacteriovorus* HD100 have been identified with a sequence similarity of below 20 % (Bd1645 19.95 %, Bd2126 19.86 %, Bd2778 16 %). In this Chapter the structures of these homologs will be characterised to determine if they are truly structural homologs to Bd1904. If they are structural homologs, the information can be used to identify features of the protein, important for function and possible relevant self-association states also.

6.2 Results

6.2.1 PSI-BLAST Result

Following four iterative rounds of PSI-BLAST (Schaffer, 2001), three additional proteins from *B. bacteriovorus* HD100 were identified: Bd1645, Bd2126 and Bd2778. Following an alignment using ClustalW (Larkin et al., 2007) it could be seen that a number of residues are

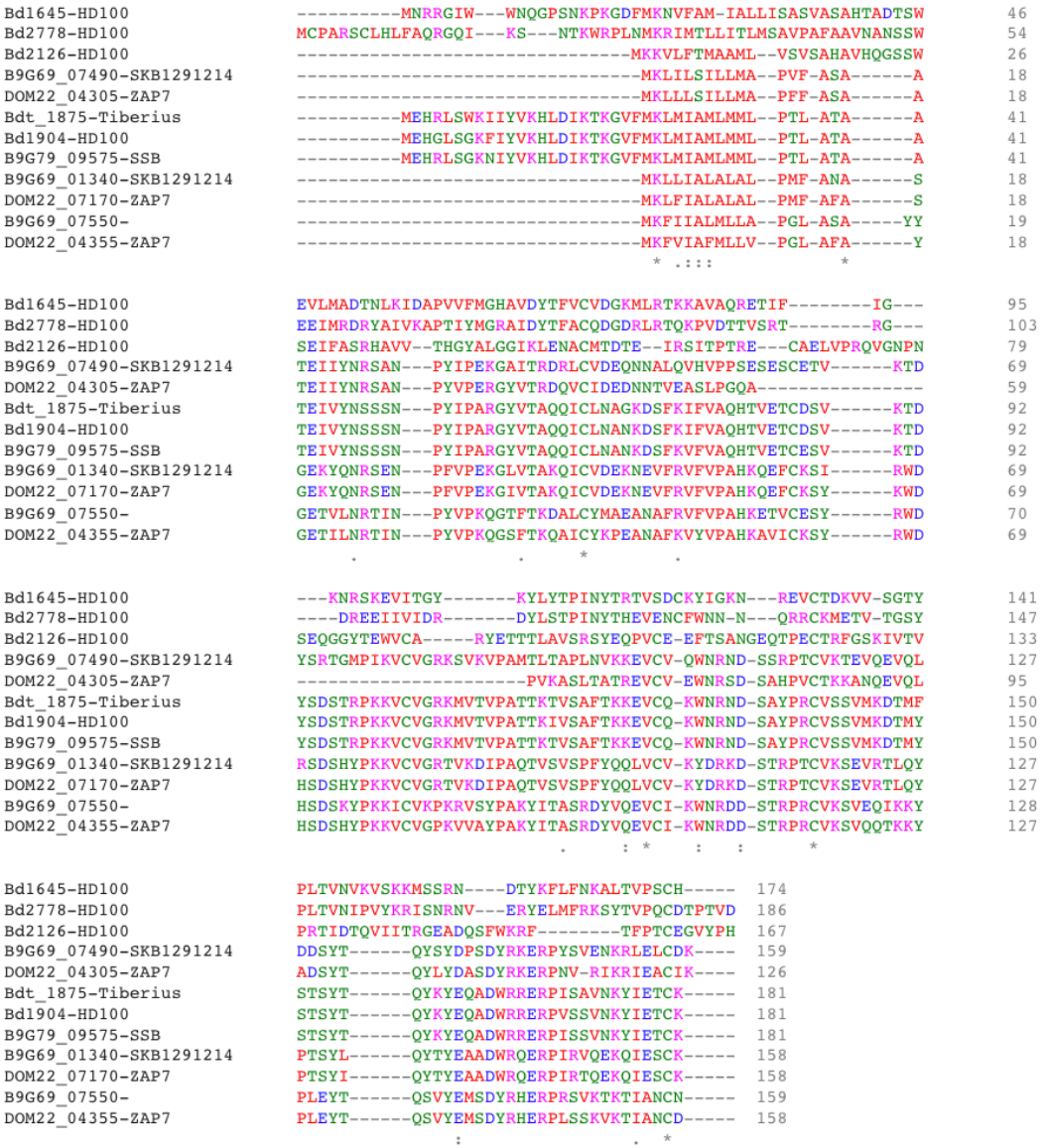
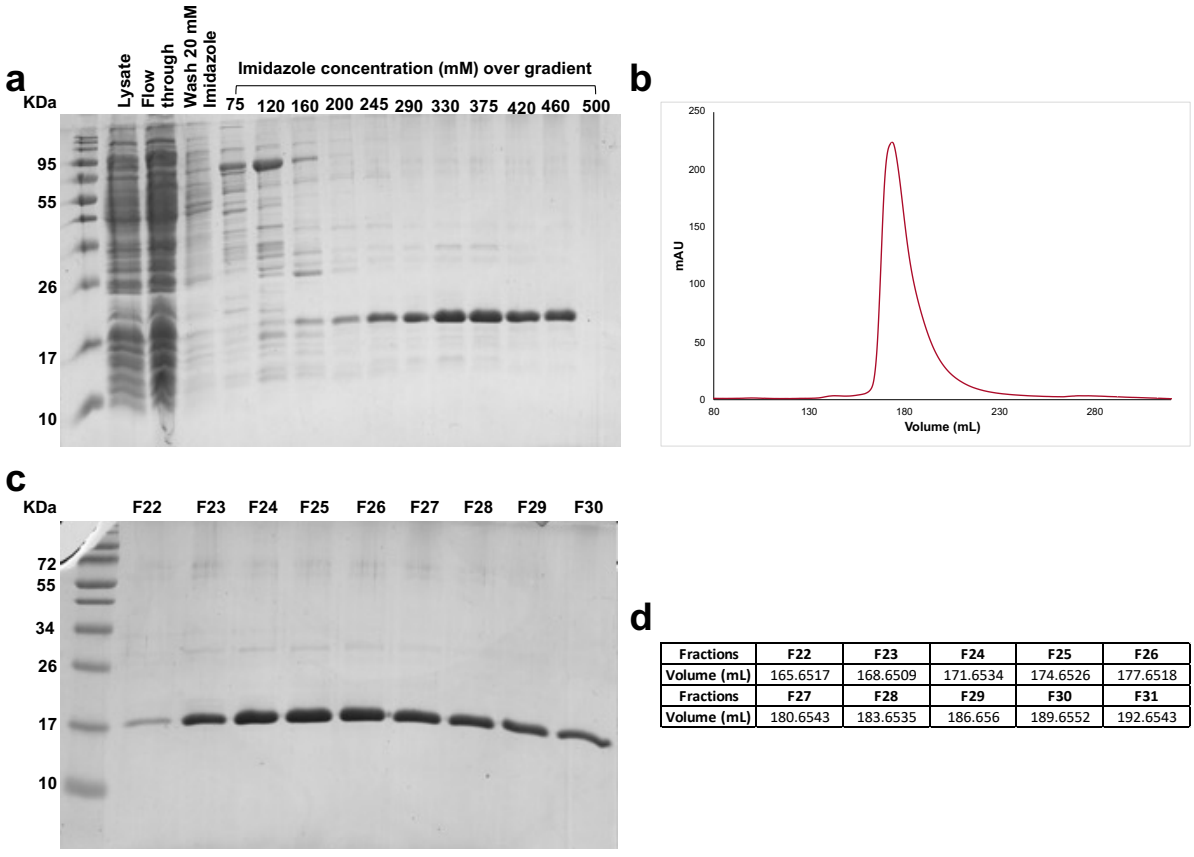


Figure 6-2: Sequence Alignment using ClustalW of Homologous Proteins to Bd1904, from *B. bacteriovorus* Strains. Asterisk (*) indicates positions which have a fully conserved residue. Colon (:) indicates conservation between amino acids of strongly related properties. Period (.) indicates conservation between amino acids of weakly related properties. Red, small and hydrophobic. Blue, acidic. Magenta, basic. Green, hydroxyl, sulfhydryl, amine, and glycine.

conserved, most notably four of the cysteine residues (Figure 6-2; fourth omitted in variants that have a shorter '*arm*'). These cysteine residues form disulfide bonds in Bd1904 that stabilise the β-hairpin arms and link the C-terminal tail to the core of the proteins.

6.2.2 Purification and Crystallisation Trials of Bd1645

Full length Bd1645 with a C-terminal His_6-tag was expressed as described in Table 2-6. The protein was initially purified using IMAC purification on 5 mL HisTrap HP column. The protein was washed and eluted using an imidazole gradient on a ÄKTA pure, with fractions with a mAU at 280 nm above 0 run on an SDS-PAGE. The SDS-PAGE (Figure 6-3a) revealed a band at ~17 KDa (Mw of Bd1645 construct 16.38 KDa) which eluted as the major band with



Fractions	F22	F23	F24	F25	F26
Volume (mL)	165.6517	168.6509	171.6534	174.6526	177.6518
Fractions	F27	F28	F29	F30	F31
Volume (mL)	180.6543	183.6535	186.656	189.6552	192.6543

Figure 6-3: Purification of Bd1645 by IMAC and SEC. a) SDS-PAGE of IMAC purification on HiTrap HP 5 mL column, using an imidazole gradient from 20 – 500 mM Imidazole. b) Size exclusion chromatogram from HiLoad 26/60 Superdex 75 PG column eluted in 20 mM Citrate (pH 6), 300 mM NaCl. c) SDS-PAGE of SEC peak. d) Conversion of volume to fractions ran on the SDS-PAGE. Protein Ladder = Color Prestained Protein Standard, Broad Range (10 250 KDa) (NEB, P7719S).

an imidazole concentration of 200 mM or greater. These fractions were pooled and dialysed into 20 mM Citrate (pH 6) and 300 mM NaCl; before concentration for further purification on a HiLoad 26/60 Superdex 75 PG. Bd1645 eluted from SEC as a monodisperse peak with a trailing shoulder (Figure 6-3b). The fractions of the peak were run on an SDS-PAGE, which resulted in a single band at ~17 KDa (Figure 6-3c); these fractions were pooled and concentrated for crystallisation experiments.

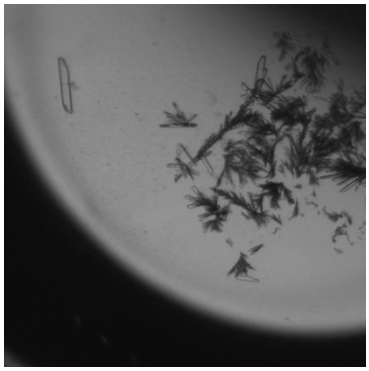
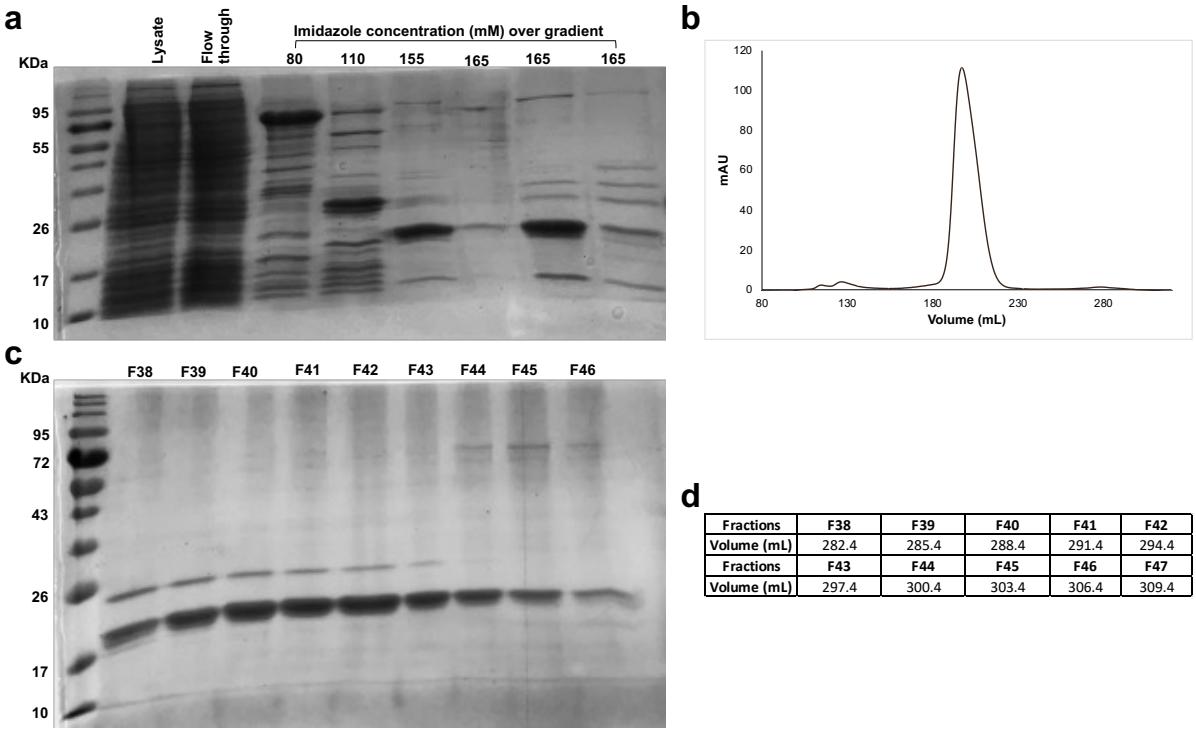


Figure 6-4: Crystal from Bd1645 Crystallisation Screen. Bd1645 in 20 mM MES (pH 6), 300 mM NaCl at 5.8 mg.mL^{-1}. Crystal condition 20 % (v/v) PPG 400, 10 % (v/v) 1-propanol.

Bd1645 was used in sitting-drop protein crystallisation experiments at 4.5, 10, 20 and 28 mg.mL^{-1} in BCS Screen, JCSGplus, MIDASplus, Morpheus and PACT Premier (Molecular Dimensions). These crystal screens resulted in no protein crystals. The protein was then dialysed into 20 mM MES (pH 6), 300 mM NaCl and concentrated to 5.8 mg.mL^{-1}. This sample was used in sitting drop crystallisation experiments using BCS Screen, JCSGplus and MIDASplus (Molecular Dimensions). This resulted in a single crystal (Figure 6-4) in 20 % (v/v) PPG 400, 10 % (v/v) 1-propanol. The crystal was looped and cryo-cooled before use in X-ray diffraction experiments. The crystal did not result in a diffraction pattern.

6.2.3 Purification and Crystallisation Trials of Bd2126

Bd2126 was expressed in TB media using IPTG to induce expression (Table 2.6). Following expression in the periplasm, the protein was extracted and initially purified using IMAC, on a 5 mL HisTrap HP column. The protein was eluted using an imidazole gradient on an ÄKTA pure, as described in Table 2.10, with fractions of the peaks run on an SDS-PAGE to determine



d

Fractions	F38	F39	F40	F41	F42
Volume (mL)	282.4	285.4	288.4	291.4	294.4
Fractions	F43	F44	F45	F46	F47
Volume (mL)	297.4	300.4	303.4	306.4	309.4

Figure 6-5: Purification of Bd2126 by IMAC and SEC. a) SDS-PAGE of IMAC purification on HiTrap HP 5 mL column, using an imidazole gradient from 20 – 500 mM Imidazole. b) Size exclusion chromatogram from HiLoad 26/60 Superdex 75 PG column eluted in 20 mM HEPES (pH 7.5), 300 mM NaCl. c) SDS-PAGE of SEC peak. d) Conversion of volume to fractions ran on the SDS-PAGE. Protein Ladder = Color Prestained Protein Standard, Broad Range (10 250 KDa) (NEB, P7719S).

purity of the samples. The SDS-PAGE (Figure 6-5a) showed a major band at ~17 KDa in the imidazole gradient, which corresponded to the molecular weight of Bd2126 (17.590 KDa), as well as impurities at both smaller and larger molecular weights. To improve the purity before crystallisation trials the fractions at 155 mM Imidazole to 165 mM Imidazole were pooled and concentrated for a final purification step. This was performed on HiLoad 26/60 Superdex 75 PG where Bd2126 eluted as a monodispersed peak (Figure 6-5b). The peak fractions were run on an SDS-PAGE, again to assess purity, revealing a major band at ~17 KDa, as well as a minor band at ~26 KDa in the first 6 fractions (38-43) of the peak and a minor band at ~75 KDa in the final three fractions (Figure 6-5c). The peak fractions were pooled and concentrated for crystallisation.

Concentrated Bd2126 was used in both sitting and hanging drop crystallisation in BCS Screen JCSGplus, PACT Premier, MIDASplus and Morpheus (Molecular Dimensions) at 5 or 9 $mg.mL^{-1}$. From these protein crystallisation trials one condition produced crystals, 100 mM Tris (pH 8) and 50 % (v/v) Pentaerythritol propoxylate (5/4 PO/OH), using 5 $mg.mL^{-1}$ Bd2126 (Figure 6-6). These crystals were harvested and cryo-protected in 20 % (v/v) ethylene glycol before X-Ray diffraction at Beamline ID30B at The European Synchrotron Radiation Facility (ESRF). When subjected to X-Rays these crystals produced no diffraction pattern.

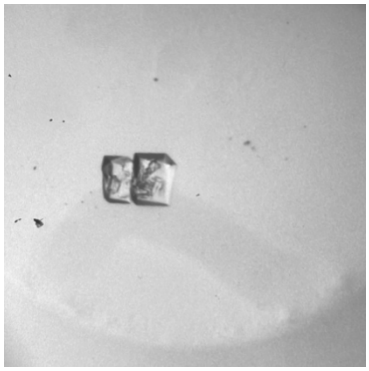


Figure 6-6: Crystals from Protein Crystallisation Trials with Bd2126. Crystals grown in 0.1 M Tris (pH 8) and 50 % (v/v) Pentaerythritol propoxylate (5/4 PO/OH).

6.2.4 Purification and Crystallisation Trials of Bd2778

Bd2778 was expressed in AI media with sugar mix C (Table 2-6) and purified using IMAC on a 5 mL His-Trap HP column. Following washing with imidazole the protein was eluted using buffer containing 50 mM EDTA. EDTA was used as there was less precipitation when the protein was eluted with EDTA compared to an imidazole gradient, during initial screening. An SDS-PAGE was run to determine purity of the elution fractions, which indicated that Bd2778 (17.523 KDa) was the major band at ~17 KDa, with two minor bands at 26 and 34 KDa (Figure

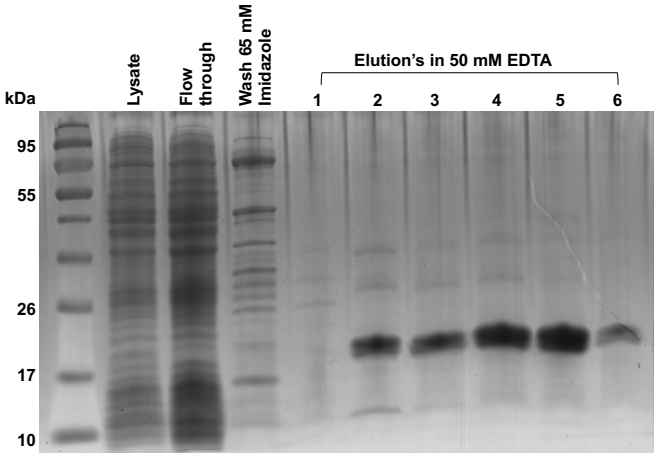


Figure 6-7: IMAC Purification of Bd2778 Method 1. SDS-PAGE of IMAC PURIFICATION, eluted using 20 mM HEPES (pH 7.5), 300 mM NaCl and 50 mM EDTA. Protein Ladder = Color Prestained Protein Standard, Broad Range (10 250 KDa) (NEB, P7719S).

6-7). The protein was pooled and dialysed into 20 mM HEPES (pH 7.5), 300 mM $MgCl_2$ (the buffer found to result in the least precipitation, before sample concentration via PEG dialysis. This method has been coined 'method 1'.

Protein crystallisation trays were set up at 4.5, 5, 5.6, 6 and 7 $mg.mL^{-1}$ in BCS Screen, JCSGplus, MIDASplus, Mini Screen, Morpheus, PACT premier and ProPlex (Molecular Dimensions). This resulted in crystals in 1 M sodium citrate, 100 mM sodium cacodylate (pH 6.5) (JCSGplus) and 100 mM NaCl, 100 mM HEPES (pH 7) and 25 % (w/v) SOKALAN PA 25 CL (MIDASplus) from a starting protein concentration of 6 $mg.mL^{-1}$ (Figure 6-8). These crystals were harvested and cryo-protected in the crystal screen buffer with 20 % (v/v) etheylene glycol and cryo-cooled ready for data collection.

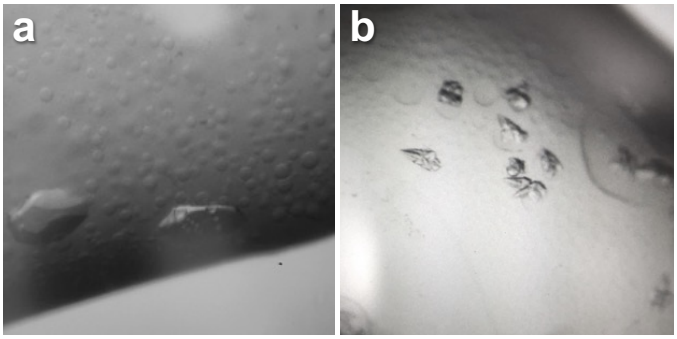
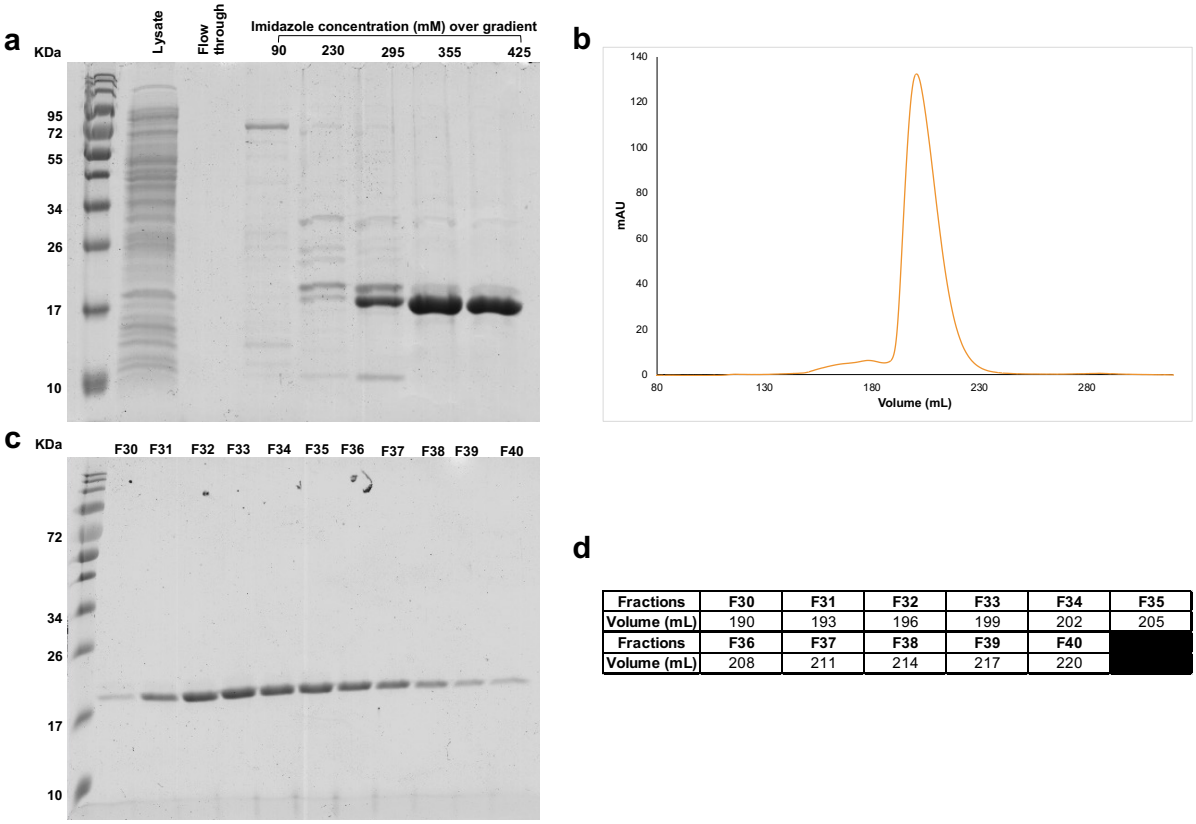


Figure 6-8: Bd2778 Crystals. Crystals in a) JCSGplus E1 (1 M sodium citrate, 100 mM sodium cacodylate (pH 6.5)) and b) MIDASplus B3 (100 mM NaCl, 100 mM HEPES (pH 7) and 25 % (w/v) SOKALAN PA 25 CL).

Following data collection (section 6.2.5), the crystal structure could not be determined as the phase problem could not be solved using Bd1904 as a search model and no heavy-atom soaked crystal was created to solve the phases using SAD. Repeats of the JCSGplus E1 crystal condition were made, but none of these attempts resulted in any crystals. Therefore, the purification method was further optimised, indicating the protein was cold sensitive and at room temperature the protein could be purified by IMAC using an imidazole gradient on the ÄKTA pure with minimal precipitation, this new method was coined 'method 2'. Fractions of the peaks were run on an SDS-PAGE to determine purity, this indicated that at 355 mM Imidazole Bd2778 was the major band at ~17 KDa, with minor bands at higher molecular weights (Figure



Fractions	F30	F31	F32	F33	F34	F35
Volume (mL)	190	193	196	199	202	205
Fractions	F36	F37	F38	F39	F40	
Volume (mL)	208	211	214	217	220	

Figure 6-9: Purification of Bd2778 by IMAC and SEC. a) SDS-PAGE of IMAC purification on HiTrap HP 5 mL column, using an imidazole gradient from 20 – 500 mM Imidazole. b) Size exclusion chromatogram from HiLoad 26/60 Superdex 75 PG column eluted in 20 mM HEPES (pH 7.5), 300 mM NaCl. c) SDS-PAGE of SEC peak. d) Conversion of volume to fractions ran on the SDS-PAGE. Protein Ladder = Color Prestained Protein Standard, Broad Range (10 250 KDa) (NEB, P7719S).

6-9a). Bd2778 was further purified by size exclusion chromatography, using HiLoad 26/60 Superdex 75 PG column. The protein was eluted using 20 mM HEPES (pH 7.5) and 300 mM NaCl as a monodisperse peak (Figure 6-9b). Fractions of the peak were run on an SDS-PAGE, showing a single band at ~17 KDa (Figure 6-9c). These fractions were pooled and concentrated for crystallisation. Protein crystallisation was trialled in three different buffers (Table 2-12) at concentrations between 3.75 – 8.8 mg.mL^{-1} in BCS Screen, PACT Premier, ProPlex, JCSGplus, MIDASplus, Morpheus (Molecular Dimensions) and Classic 1, 2, ,3, 4, 6, 7, 8, and 10 (Jena Bioscience). This yielded one condition with crystals (Figure 6-10), 20 mM Na/K phosphate

with 20 % (w/v) PEG 3350, from Bd2778 at 7 mg.mL^{-1} in 20 mM HEPES (pH 7.5), 300 mM $MgCl_2$. These were harvested and cryo-protected in 20 % (v/v) glycerol. The crystals were screened at Diamond Light Source and produced no detectable diffraction pattern.

6.2.5 Data Collection and Model Building of Bd2778

The crystals underwent data collection at Diamond Light Source in Oxford, on the I03 Beamline. The crystals from MIDASplus B3 (Figure 6-8b) diffracted to a resolution of 4.6 Å (Figure 6-11c-d) and the crystals from JCSGplus E1 (Figure 6-8a) diffracted to a resolution of 2.0 Å (Figure 6-11a-b) both in the space group P 4_1 2_1 2.

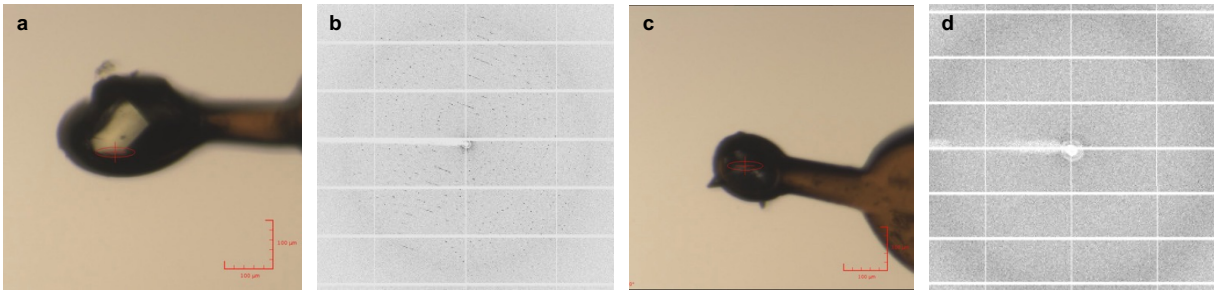


Figure 6-11: Crystal Collected on and Diffraction Pattern. a&b) Crystal and diffraction pattern from JCSGplus E1. c&d) Crystal and diffraction pattern from MIDASplus B3.

Following the release of ColabFold, the dataset from the Bd2778 JCSGplus crystal was determined using molecular replacement in Phenix Phaser (McCoy et al., 2007) with a trimmed ColabFold model of Bd2778. The ColabFold model was trimmed at the N-terminus to residue 66 and C-terminus from residue 181 as well as residues, 95-113, 124-143, and 153-173 (Figure 6-12). The phases for Bd2778 were solved with a TFZ of 8.2, searching for 8 copies in the ASU

R113

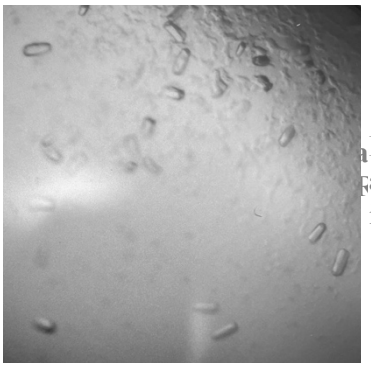


Figure 6-10: Crystals of Bd2778 using Method 2. Crystals of Bd2778 at 7 mg.mL^{-1} in 20 mM HEPES (pH 7.5), 300 mM $MgCl_2$, crystallised in 20 mM Na/K phosphate, 20 % (w/v) PEG 3350.

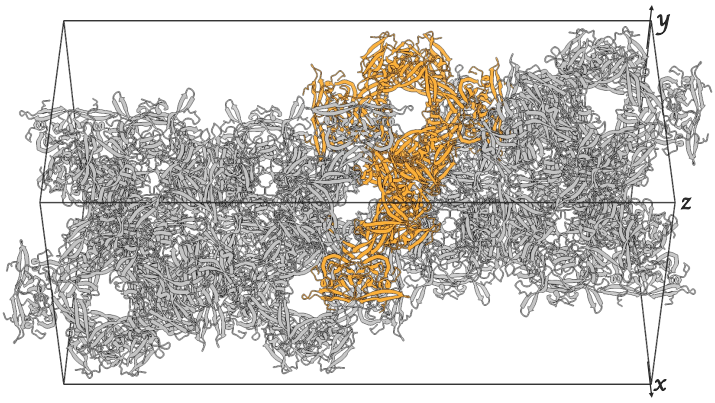


Figure 6-13: Bd2778 Unit Cell.

(Matthews Coefficient predicted 22 copies in ASU) in the enantiomeric space group P 4_3 2_1 2. This was then used in Phenix Autobuild (Terwilliger et al., 2008) with non-crystallographic symmetry (NCS) operators from Find NCS From Density (Terwilliger, 2013). This resulted in a model with 1260 residues with a correlation coefficient (CC) of 0.59, which was built manually using Coot resulting in a model with 15 copies of Bd2778. The model underwent a cycle of refinement in Phenix Refine before Density Modification. The density modified map was used in CCP4 MolRep (Vagin and Teplyakov, 2010, 1997) with the previously built model used as a fixed model, and two copies of Bd2778 being used as the search model. This resulted in 18 copies of Bd2778 in the ASU, with 2716 residues. The model was completed with iterative cycles of refinement in Phenix Refine and manual building in Coot. The final model contained 2509 residues, 1290 waters and 27 ligands, with 18 copies of Bd2778 in the ASU (Figure 6-13). Chain O was built missing the N-terminal residue (Ala47), chains D, H, I, and Q were built missing the C-terminal residue (Asp186), chain G is missing the two C-terminal residues and chain J is missing the three C-terminal residues. A summary of the data collection and processing statistics are summarised in Table 6-1.

Table 6-1: Summary of Data Collection and Processing Statistics of Bd2778.

Data collection of Bd2778					
Diffraction Source			Diamond Light Source		
Beamline/wavelength (Å)			IO3/0.89		
Total rotation range (˚)			360		
Images collected			3600		
Exposure time per image			0.020s		
Transmission (%)			100		
Space group			P 4_3 2_1 2		
Cell dimensions					
A (Å)	B (Å)	C (Å)	146.95	146.95	335.08
α (˚)	β (˚)	γ (˚)	90	90	90
Shell			Overall	Outer shell	
Resolution Range (Å)			51.34-1.99	2.06-1.99	
Total number of reflections			6781488	690919	
Total number of unique reflections			249542	24605	
R_{merge} (within I+/I-)			0.138	4.12	
R_{meas} (within I+/I-)			0.141	4.20	
$R_{p.i.m.}$ (within I+/I-)			0.0269	0.787	
I/σ(I)			14.5	0.42	
CC Half			0.999	0.411	
CC*			1	0.763	
Multiplicity			27.2	28.1	
Completeness (%)			99.93	99.72	
Wilson B-factor			48.12		
Refinement statistics					
Reflections used in refinement			249435	24583	
Reflections used for R-free			12415	1178	
R-work			0.2095	0.3675	
R-free			0.2322	0.3898	
CC(work)			0.962	0.622	
CC(free)			0.953	0.655	
Protein residues			2509		
RMS(bonds) (Å)			0.007		
RMS(angles) (˚)			0.92		
Average B-factors (Å^2)			65.30		
Ramachandran favoured (%)			98.10		
Ramachandran allowed (%)			1.74		
Ramachandran outliers (%)			0.16		
Rotamer outliers (%)			1.63		
Clashscore			7.68		

The final electron density was readily interpretable with well-defined density for the backbone and most sidechains (Figure 6-14). The only region of poor-defined density was for the β-hairpin turn of arm 1 of chains D, H, I, M, P, Q, where the density for the backbone could not be seen at a 1.5 sigma level (Figure 6-14b).

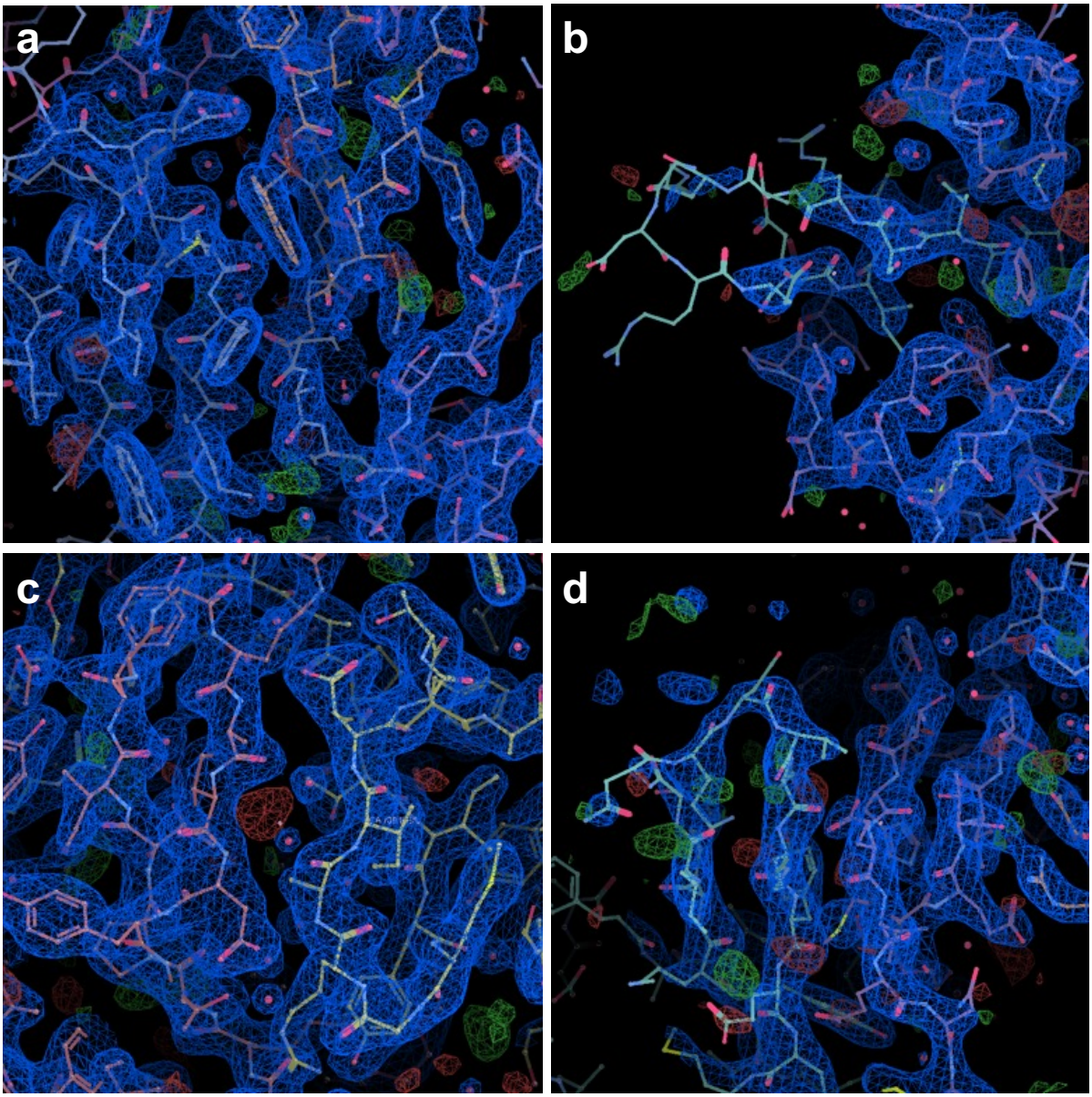


Figure 6-14: 2mFo-DFc Electron Density Map of Bd2778 1.5 σ. a) Interface 2, featuring chains F, D, and I. b) Arm 1 of chain H. c) Dimer interface of chain K and L. d) Arm 2 of chain G.

6.2.6 Structure of Bd2778

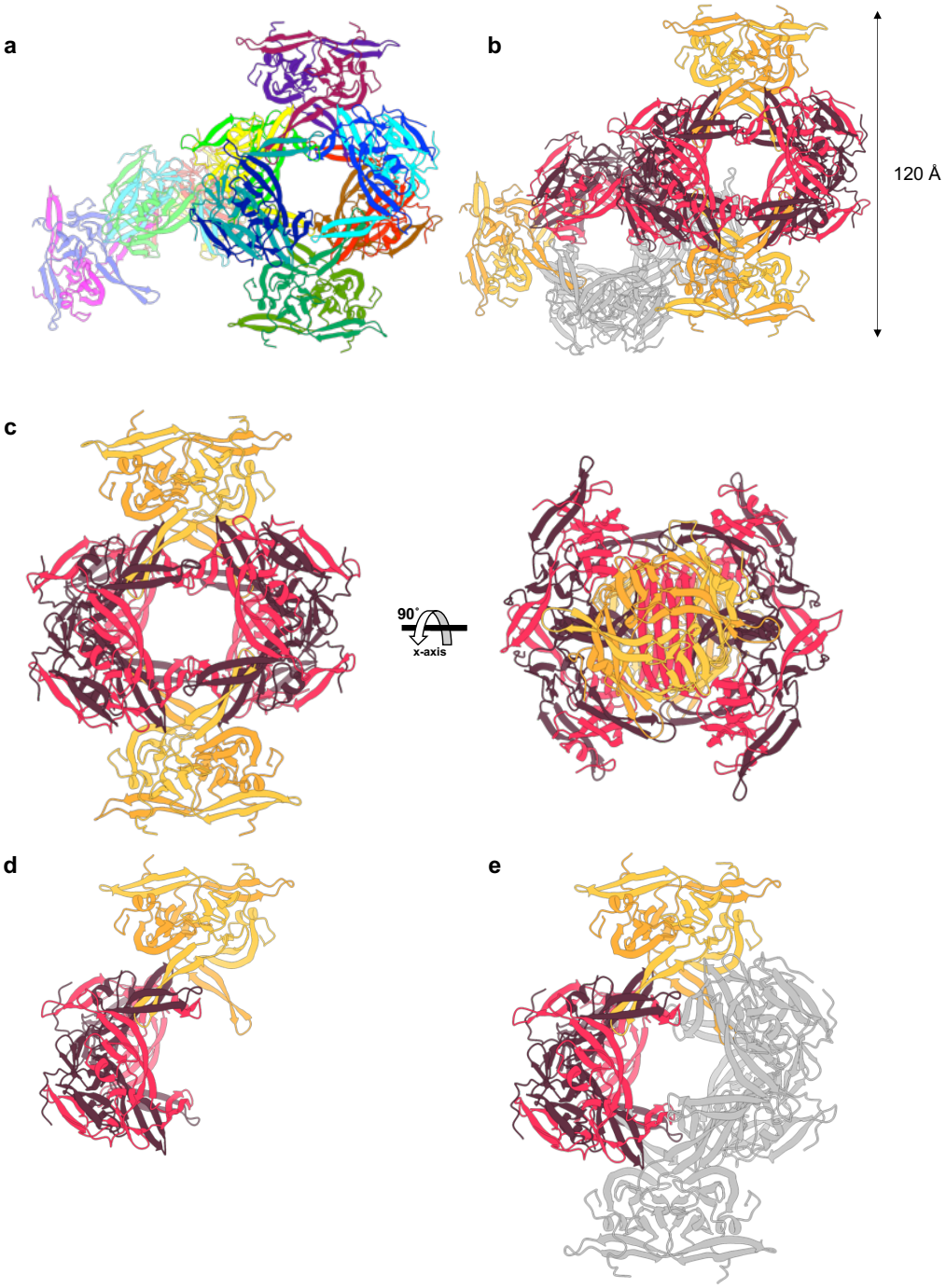


Figure 6-15: Structure of Bd2778. a) Copies of Bd2778 in the ASU coloured by chain. b) 18 copies of Bd2778 in the ASU coloured by forms, with symmetry copies in grey to complete the second dodecamer. c) Complete dodecamer rotated around the x-axis. d) The half of a dodecamer. e) Half the dodecamer with the symmetry copies in grey to complete the dodecamer.

Bd2778 packs with 18 copies in the ASU, which form a dodecamer and half a dodecamer, each made up of six dimers (Figure 6-15). Each dodecamer forms a '*lantern-like*' fold with a central solvent channel (height 27 Å, width 50 Å C_α to C_α).

The protomer forms a short (6-residue) α-helix near the N-terminus, followed by a β-strand which leads into a hydrophobic core formed of β-hairpin followed by a 3-residue α-helix and a three-stranded twisted β-sheet (β-strands 4, 5, and 8) (Figure 6-16). From the core there are three β-hairpin arms, with arm 2 stabilised by a disulfide and arm 3 stabilised by β-strand 1 (Figure 6-16). The C-terminal tail is also stabilised by a disulfide linkage to the core of the fold (Figure 6-16).

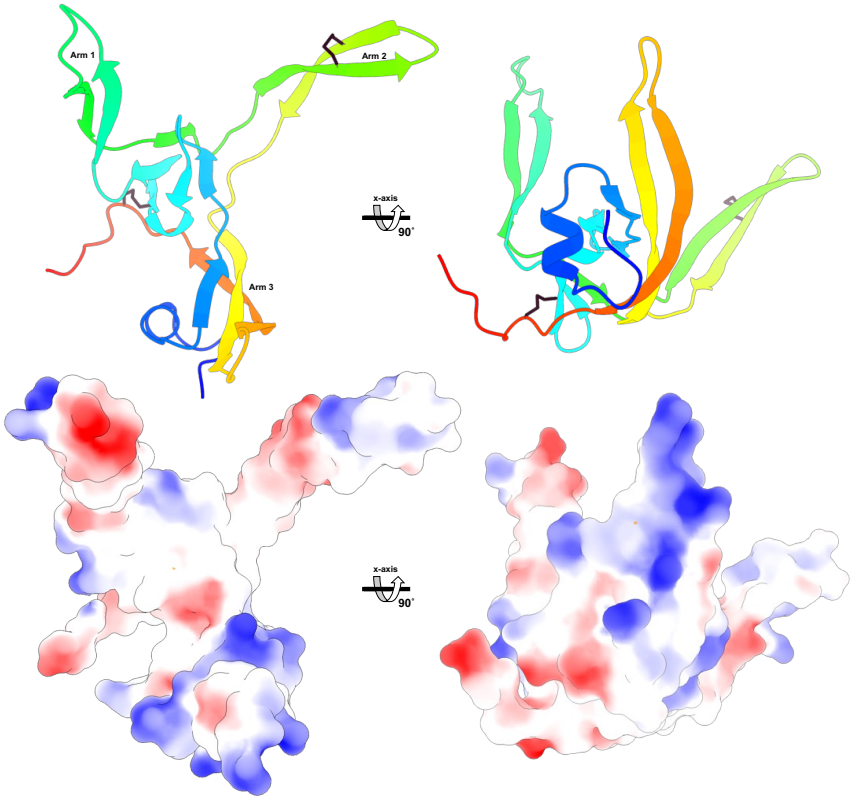


Figure 6-16: Fold of Bd2778 Protomer and Electrostatic Surface. Ribbon diagram coloured with Jones rainbow, to show N-terminus to C-terminus. Disulfides are coloured in purple.

Analysis of each chain revealed they can be split into three different subgroups each representing a specific conformation or '*form*'. '*Form 1*' includes chains A, B, E, F, K and L (Figure 6-17a), '*form 2*' includes chains C, G, J, N, O and R (Figure 6-17b) and '*form 3*'

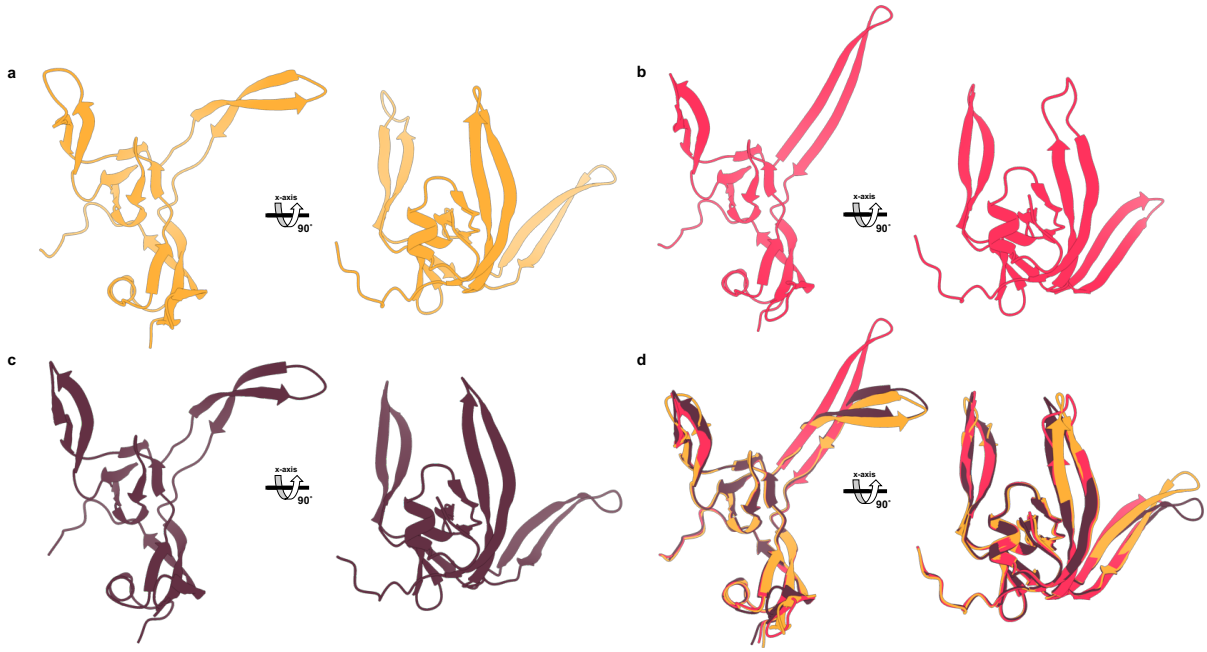


Figure 6-17: Different Forms of Bd2778. a) Representation of chains A, B, E, F, L and K, '*form 1*'. b) Representation of chains C, G, J, N, O, R, '*form 2*'. c) Representation of chains D, H, I, M, P, Q, '*form 3*'. d) Superimposition of chain E, C and D, using Needleman-Wunsch alignment algorithm.

includes chains D, H, I, M, P and Q (Figure 6-17c). The main difference between these '*forms*' is in the positions of arms 1 and 2, this can be seen clearly when the chains are superimposed (Figure 6-17d). The superimposition of chains C and D on chain E gave an RMSD across 123 and 121 pruned atom pairs of 0.56 Å and 0.59 Å respectively, and across the whole chain an RMSD of 3.9 Å and 1.4 Å respectively. This shows that 121-123 residues are in very similar positions with a low RMSD and that there is only significant movement in 17-19 residues across

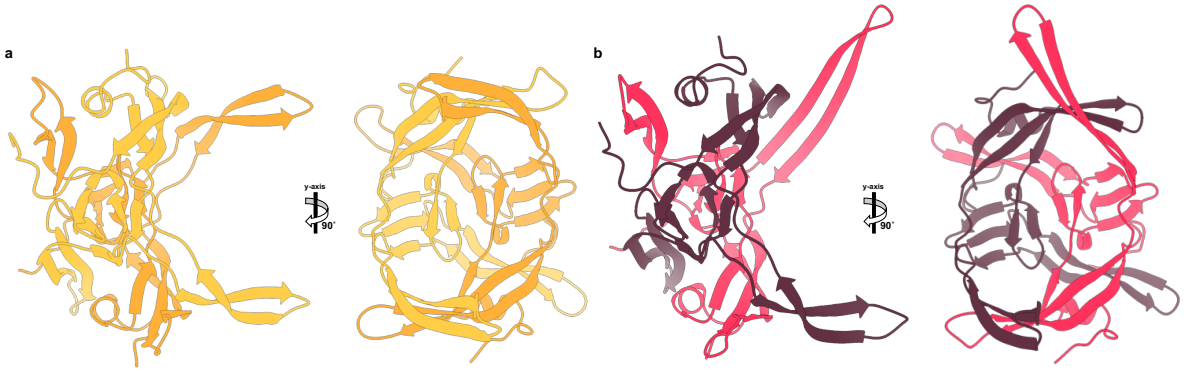


Figure 6-18: The Two Dimer Forms of Bd2778. a) Dimer made from chains A&B, E&F, and K&L. b) Dimer made from chains C&D, G&H, I&J, M&N, O&P, and Q&R.

the three forms, with a higher RMSD. As a result of these three different forms, the previously mentioned dimers that make the dodecamer are two different dimers with different conformations (Figure 6-18). The first (dimer X) is a homodimer made of two chains from form 1 (A&B, E&F, K&L) and is found at the '*top*' and '*bottom*' of the '*lantern-like*' dodecamer (Figure 6-18a, Figure 6-15). The second (dimer Y) is made up of one chain from each of the other forms (C&D, G&H, I&J, M&N, O&P, Q&R) and forms the two '*sides*' of the '*lantern*' (Figure 6-18a, Figure 6-15).

6.2.7 Dimer Interfaces of Bd2778

To characterise the two dimers, the protein:protein interfaces of each dimer were assessed using PISA (Table 6-2).

Table 6-2: Interface Statistics

Interface	**Dimer X**		**Dimer Y**	
Chain	A	B	C	D
Interface residues	62	61	63	56
Interface atoms	235	237	248	231
Interface area (Å^2)	2335.3		2356.5	
No. Hydrogen bonds	20		24	
No. salt bridges	11		9	
ΔG_{int} (kcal/mol)	-27.8		-24.8	
ΔG_{int} P-value	0.013		0.027	

Dimer X involved 61-62 residues from each chain, consisting of 20 hydrogen bonds, 11 salt bridges as well as hydrophobic interactions (Figure 6-19a-c). To construct Dimer X, hydrogen bonds were mostly formed between the carbonyls and nitrogens of the backbone of each chain particularly between the two β3-strands as well as arms 1 and 2 (Figure 6-19a). The salt bridges occur between arm 1 and arm 2 (Glu125 and Arg166) of each chain, the N-terminal α-helix and arm 3 (Glu55 and Arg100), and between both arms 3 (Lys92 and Asp95) to help stabilise the dimer (Figure 6-19b).

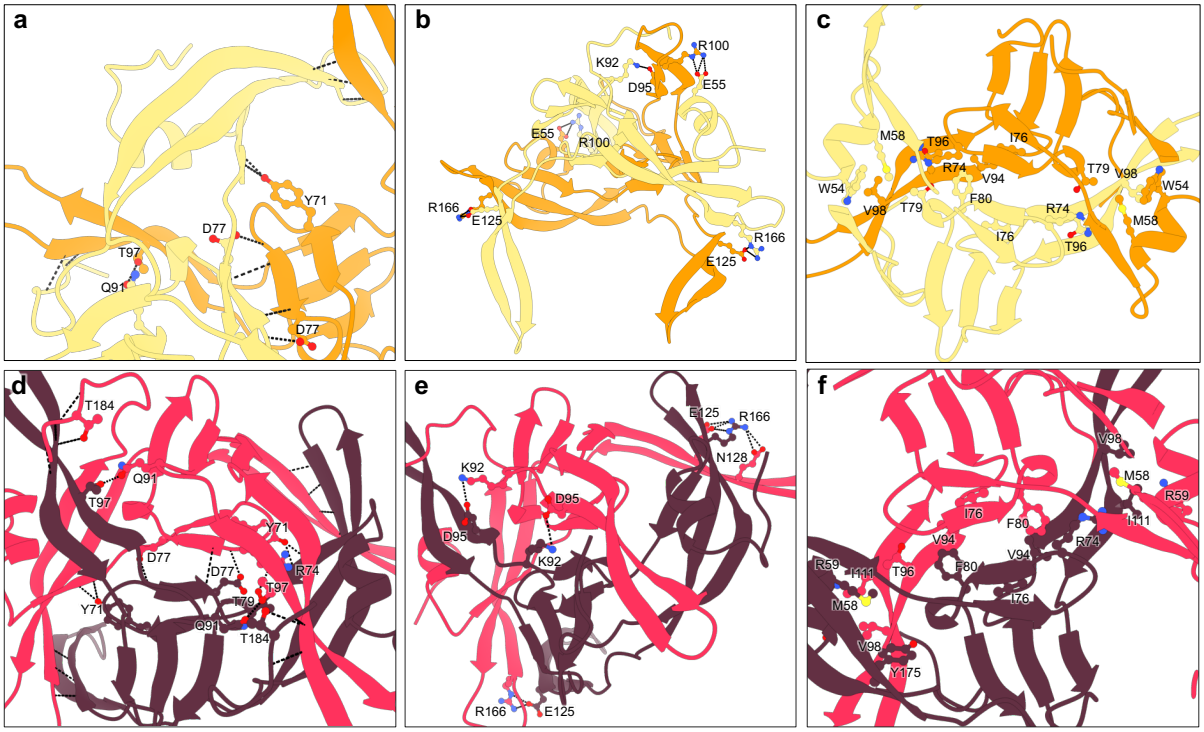


Figure 6-19: Interface of Bd2778 Two Dimer Conformations, X (orange and yellow) and Y (pink and purple). Hydrogen bonds (a), salt bridges (b) and hydrophobic interactions (c) between chains A and B. Hydrogen bonds (d), salt bridges (e) and hydrophobic interactions (f) between chains C and D.

Dimer Y involved 63 residues from chain C and 56 residues from chain D comprising of 24 hydrogen bonds, 9 salt bridges and hydrophobic interactions (Figure 6-19d-f). In this dimer most of the hydrogen bonds involved sidechain to backbone (carbonyls or nitrogens), with only 9 backbone-backbone hydrogen bonds with 6 of these occurring between arms 2 and 3 (Figure 6-19d). There are fewer salt bridges in dimer Y, than dimer X, though they mostly involve the same residues (Asp95, Lys92, Glu125, Arg166), with Asn128 (chain C) also forming a salt bridge with Arg166 (chain D) due to the conformational change in arm 2 (Figure 6-19e).

6.2.8 Formation of the Dodecamer

The full dodecamer is made up of 12 copies of Bd2778 (chains A, B, C, D, E, F, G, H, I, J, Q & R), and forms a near symmetrical oligomer (Figure 6-20). Two copies of dimer X sit at the '*top*' (chains A & B) and '*bottom*' (chains E & F) of the '*lantern-like*' dodecamer, with four copies of dimer Y forming the '*sides*' and '*central-hole'* (Figure 6-20a).

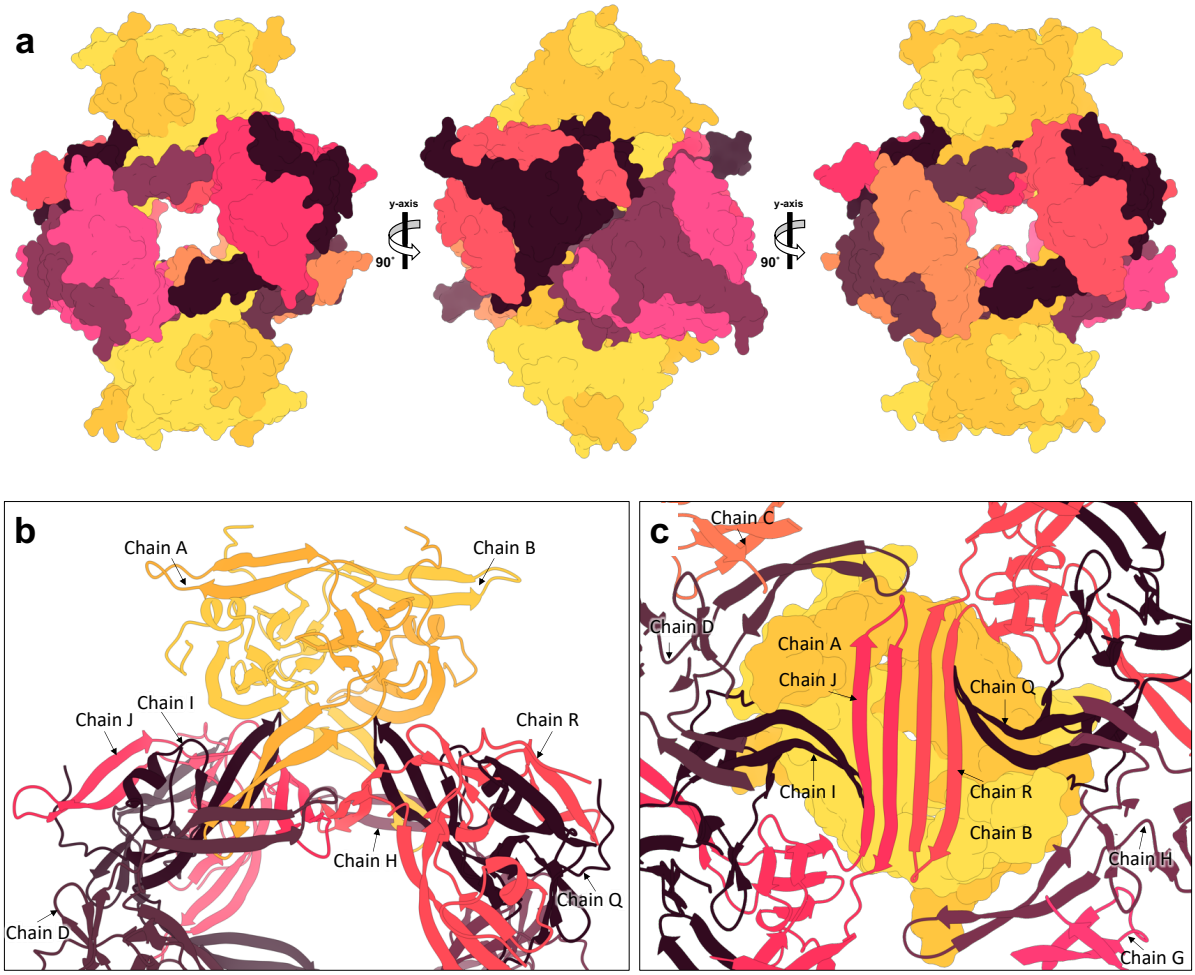


Figure 6-20: Formation of Bd2778 Dodecamer. a) Bd2778 dodecamer surface, coloured by chain. b) Bd2778 cartoon showing chains involved in forming unique interfaces. c) Chains A and B (dimer X) shown as surface models, surrounded by four copies of dimer Y (chains, C, D, G, H, I, J, Q, R).

To form this dodecamer there are four truly unique interfaces. *Interface 1* involves all of arm 2 from each '*form 2*' of dimer Y forming a β-sheet, containing 4 β-strands (Figure 6-21a-b). *Interface 2* is the β-hairpin turn of arm 2 from dimer X, forming interactions with three chains from two copies of dimer Y (Figure 6-21c). *Interface 3* involves the β-hairpin of arm 3 from '*form 3*' of dimer Y interacting with the core of one chain from dimer X, and the sidechains from loop between the α-helix and β-strand 1, of dimer Y, interacting with arm 3 from one chain from dimer X (Figure 6-21d). *Interface 4* is formed by the packing of two *'form 3'* protomers (Figure 6-20a, dark purple and purple of middle pannel).

To form the β-sheet in *Interface 1,* ten backbone to backbone (carbonyls to nitrogens) hydrogen bonds form between β-strand 10 of arm 2 (chain R) and β-strand 10 of arm 2 (chain J). The interface is reinforced by three hydrogen bonds between sidechains (JTyr122:RGln135, JGln135:RTyr122, JArg136:RThr144) and two salt bridges (JLys139:RGlu141, JGlu141:Rlys139) (Figure 6-21a-b). A salt brige is also formed between Glu141 (chain A/B) and Arg136 (chain J/R).

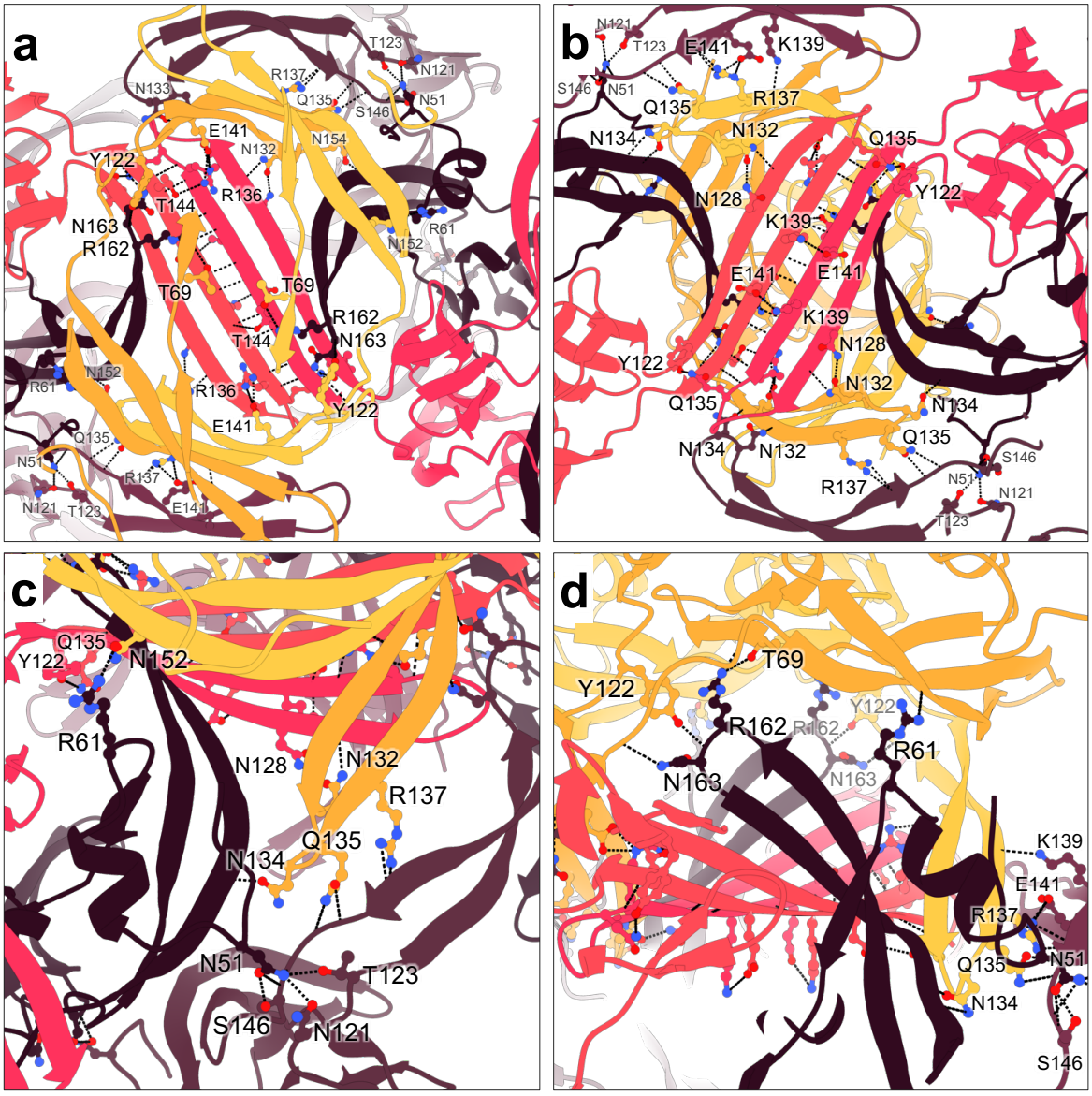


Figure 6-21: Hydrogen Bonds and Salt Bridges Involved in Forming the Dodecamer. a&b) Interface 1. c) Interface 2. d) Interface 3. This figure includes chains A (dark orange), B (orange), D (purple), H (purple), I (dark purple), J (dark pink), Q (dark purple), R (pink).

Interface 2 involves the β-hairpin turn of arm 2 from either chain A, B, E or F passing through a '*hole*' created by 3 chains from two copies of dimer Y (chains D, I and J or H, Q and R or G, H and Q or C, D and I, respectively) which forms both hydrogen bonds (Figure 6-21c) and hydrophobic interactions (Figure 6-22) to create the interface. To discuss this interface the interactions formed between chains A, D, I and J will be focused on. The '*hole*' is stablised by hydrogen bonding between chains D and I, with IAsn51 forming four hydrogen bonds; one with the backbone of chain D and three with sidechains from chain D (Asn121, Thr123, Ser146). Arm 2 of chain A forms six hydrogen bonds (Asn132, Asn134, Gln135, Arg137) with the backbone of chains D, I and J as well as forming one hydrogen bond between the sidechains of AAsn132 and JAsn128 (Figure 6-21c). As well as the hydrogen bonding there are hydrophobic interactions at the interface, including pi-teeing between APhe130 and JTrp131, and ATrp131 and IPhe171, as well as van der Waals contacts between the hydrophobic residues (Figure 6-22).

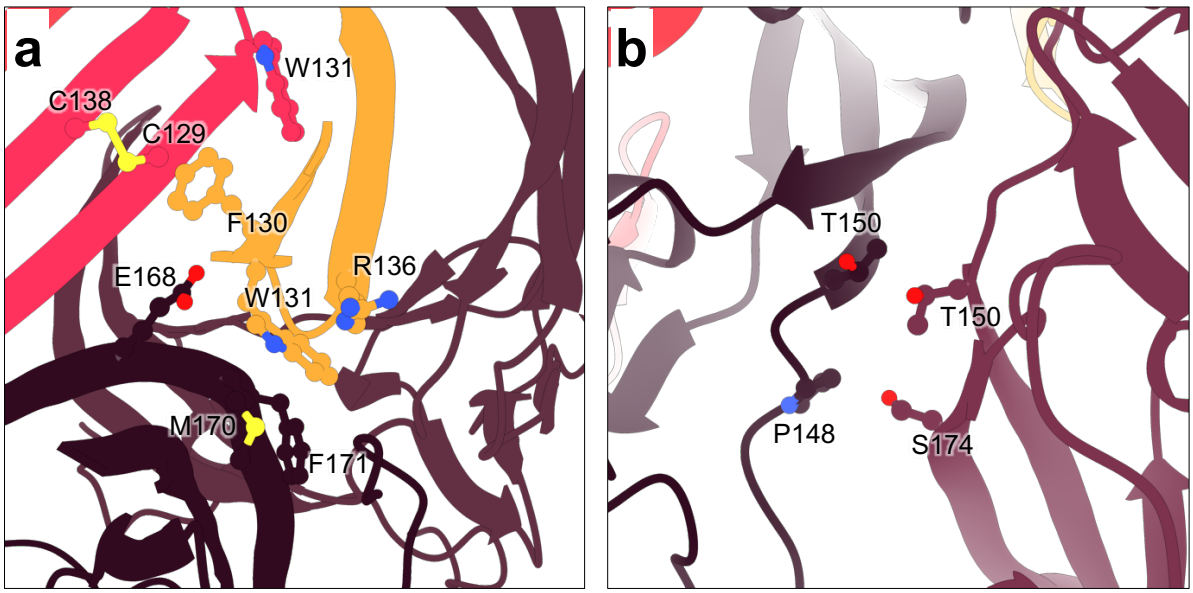


Figure 6-22: Hydrophobic Interactions of Bd2778 Dodecamer Interfaces. a) Interactions of *interface 2* showing chains, A (orange), D (purple), I (dark purple), and J (dark pink). b) Interactions of *interface 4*, showing chains Q (dark purple) and H (purple).

Chains A and Q, B and I, E and D, or F and H each form an *interface 3*. In this interface, sidechains (Arg162, Asn163) from the β-hairpin turn of arm 3 of '*form 3*' form hydrogen bonds with the backbone of '*form 1*' (β1 strand and loop following arm 2), with Arg162 also forming a hydrogen bond from the amide to carboxyl group of Thr69 (Figure 6-20d). The sidechain of Tyr122 from '*form 1*' forms a hydrogen bond to the backbone of the β-hairpin turn of arm 3 of '*form 3*'. The final interaction is a hydrogen bond between Arg61 of '*form 3*' (loop after α-helix 1) and the backbone of arm 3 of '*form1*' (Figure 6-20d).

6.2.9 Potential Ligand Binding Sites of Bd2778

In the Bd2778 structure there are two potential ligand binding sites. The first is around Lys66 (Figure 6-23). In every chain within the structure there is binding of PEG of varying lengths

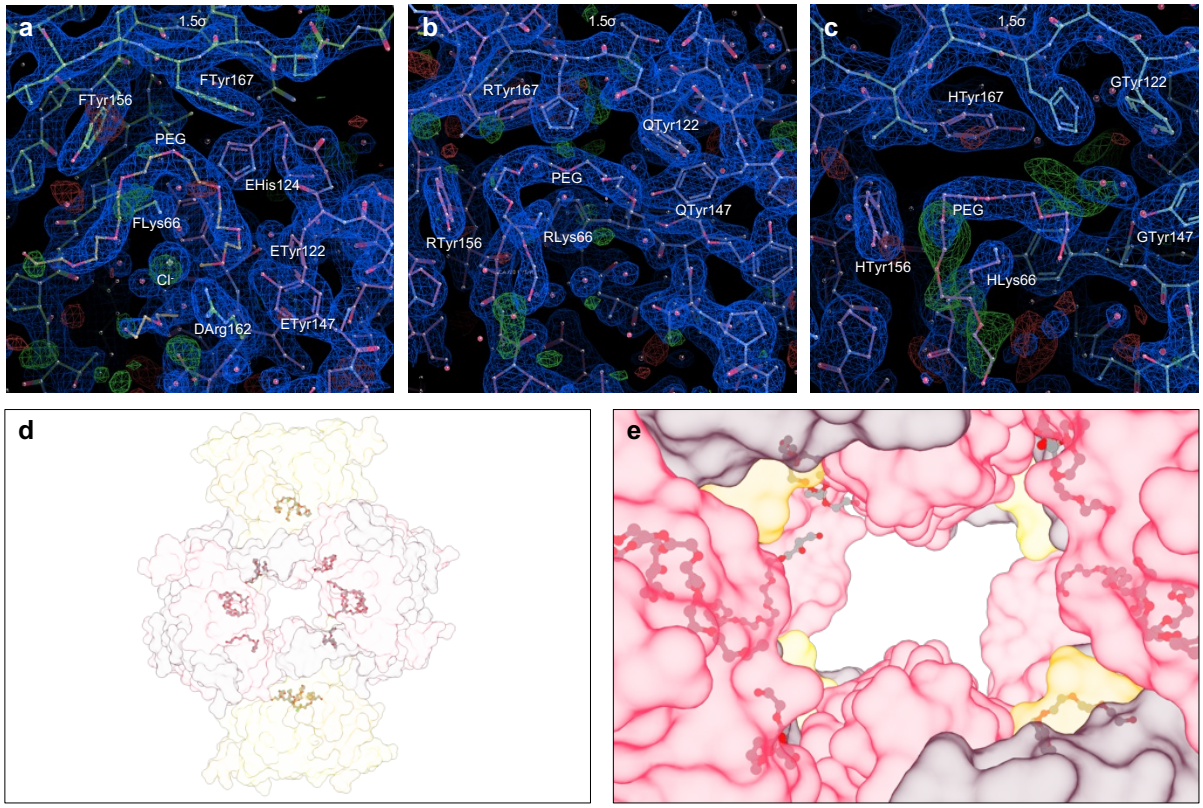


Figure 6-23: Ligand Binding Site of PEG, around Lys66. a) 2mFo-DFc and mFo-DFc density map showing PEG binding around Lys66 in chain F, '*form 1*'. b) 2mFo-DFc density map showing PEG binding around Lys66 in chain R, '*form 2*'. c) 2mFo-DFc and mFo-DFc density map showing PEG binding around Lys66 in chain H, '*form 3*'. d) Surface of Bd2778 dodecamer showing where the PEG binds in the overall structure. e) PEG within the 'central-channel' of the dodecamer.

around the sidechain of Lys66 connected by hydrogen bond from the amino group. In all '*forms*' the PEG forms hydrophobic interactions with Try156 and Try167 of the same chain, and in '*forms 1* and *2*' Try122 and Try147, from the other chain in their dimer pair, also form hydrophobic interactions with PEG (Figure 6-23a-b). In '*form 3*' the PEG molecule is shorter resulting in no interactions with Try122 and Try147 (Figure 6-23c). As well as the PEG molecules in all chains of '*form 1*' there is a Cl^- ion co-ordinated by Arg162 of '*form 3*' located within the binding site (Figure 6-23a). Investigating these sites further it can be seen that the PEG molecules are found in the centre of Dimer X and within the '*central-channel*' of the '*lantern-like*' dodecamer (Figure 6-23d-e).

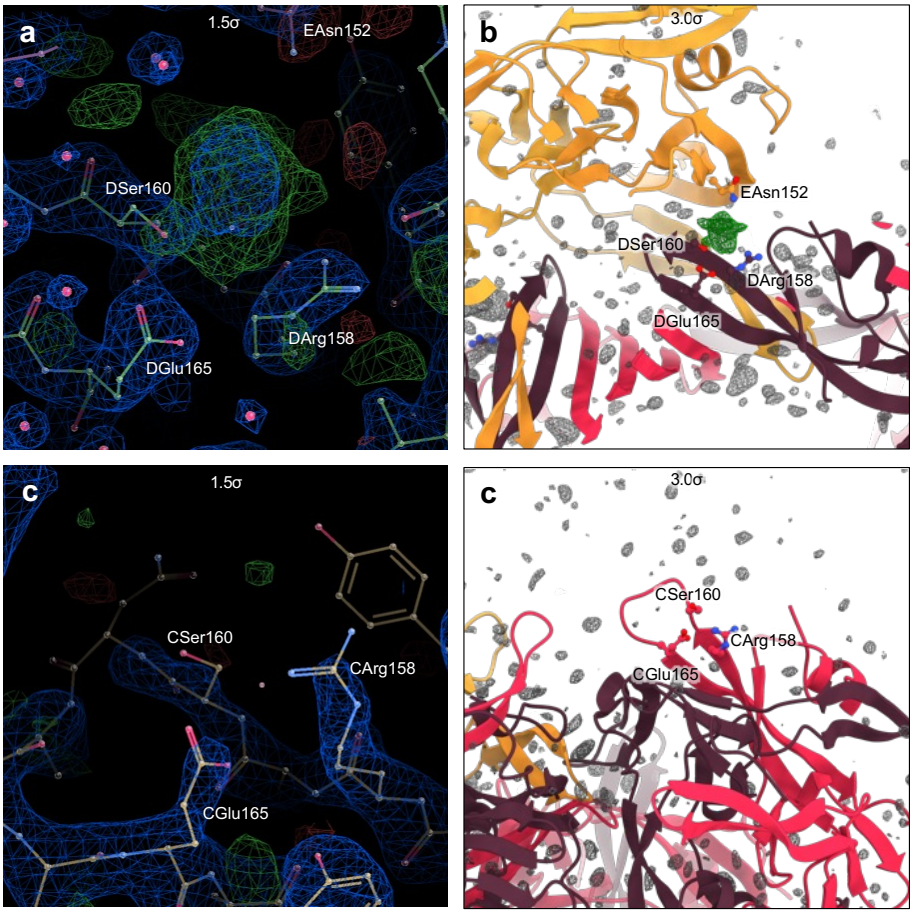


Figure 6-24: Potential Ligand Binding Site next to Ser160 of '*Form 3*'. a) 2mFo-DFc and mFo-DFc map of volume of positive density next to Ser160, in Chain D (*form 3*). b) Bd2778 dodecamer with mFo-DFc map, volume of positive density coloured in green. c) 2mFo-DFc and mFo-DFc map of volume of positive density next to Ser160, in Chain C (*form 2*). d) Bd2778 dodecamer with mFo-DFc map.

The second ligand binding site is found in '*form 3*' at Ser160, where there is a globular volume of positive density (Figure 6-24a-b). Hydrogen bonds could form between the unknown ligand and Ser160 and Glu165. This is not seen in the other chain ('*form 2*') of dimer Y, this could be to do with its location on the external face of the dodecamer (Figure 6-24c-d).

6.2.10 Conservation of Bd2778

The conservation of residues in Bd2778 were calculated using the Consurf server (Figure 6-25). This indicated that the residues in the β-hairpin arms (Thr97-Arg111, Phe130-Arg137, lys157-

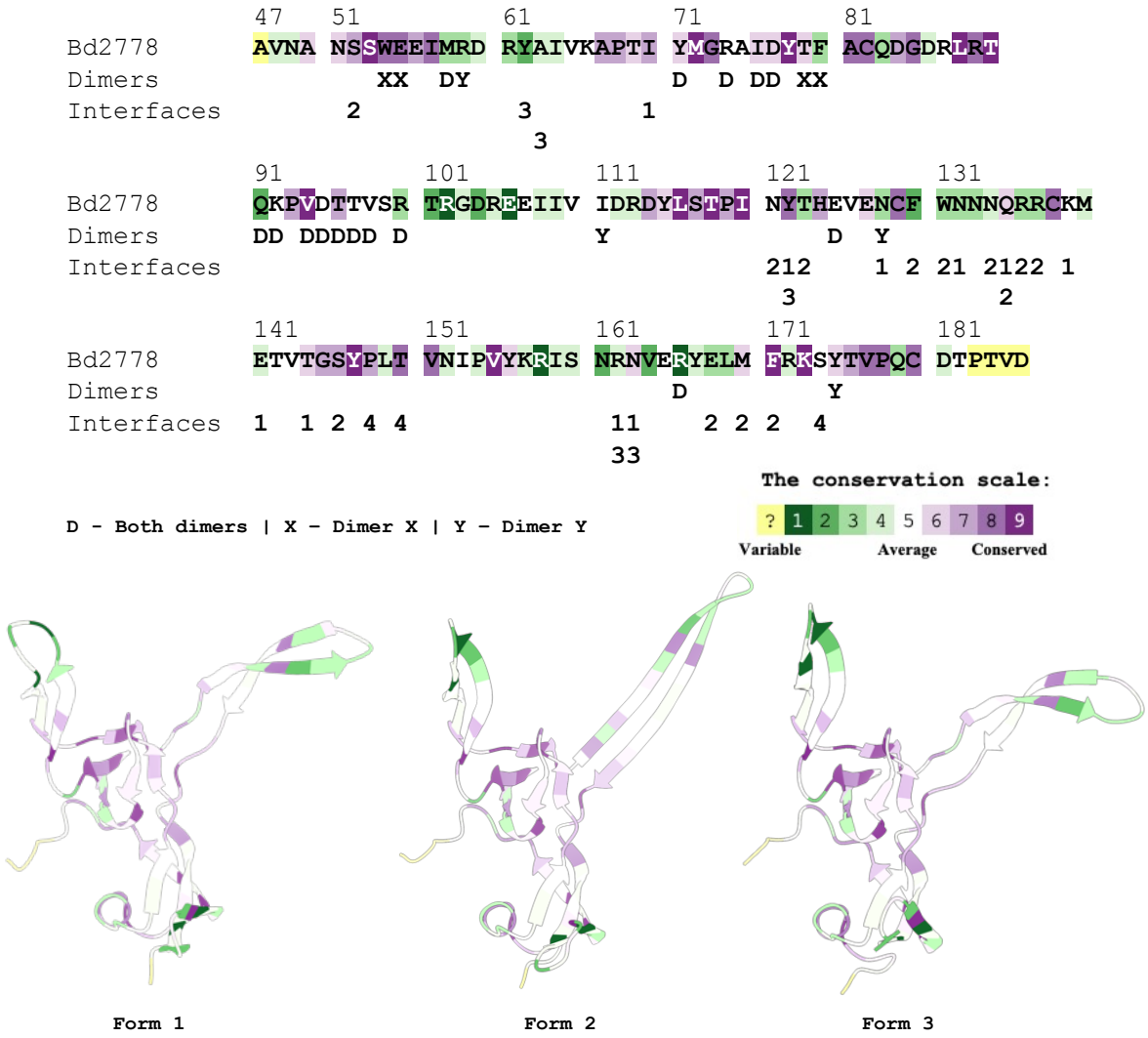


Figure 6-25: Conserved Residues of Bd2778. Sequence coloured by conservation, identifying the residues with sidechains involved in either the dimers or dodecamer interface. The sequence conservation is then mapped onto cartoons of each Bd2778 conformation.

Leu169) and the two final residues in the α-helix followed by the loop (Met58-Lys66) were the areas within the structure of the most variability (Figure 6-25). The conservation within the residues with sidechains involved in the interface, was variable with approximately 50 % of residues conserved (Figure 6-25). One interesting, conserved residue involved in *interface 4* is Thr150, a highly conserved (score of 9) residue that due to its position within the structure is unlikely to play a functional role, therefore it's conservation supports the observed oligomeric arrangement. Looking at the residues involved in the binding of PEG in each dimer; it was observed that all residues apart from Lys66 are conserved (Figure 6-26), with these tyrosine residues exposed to solvent.

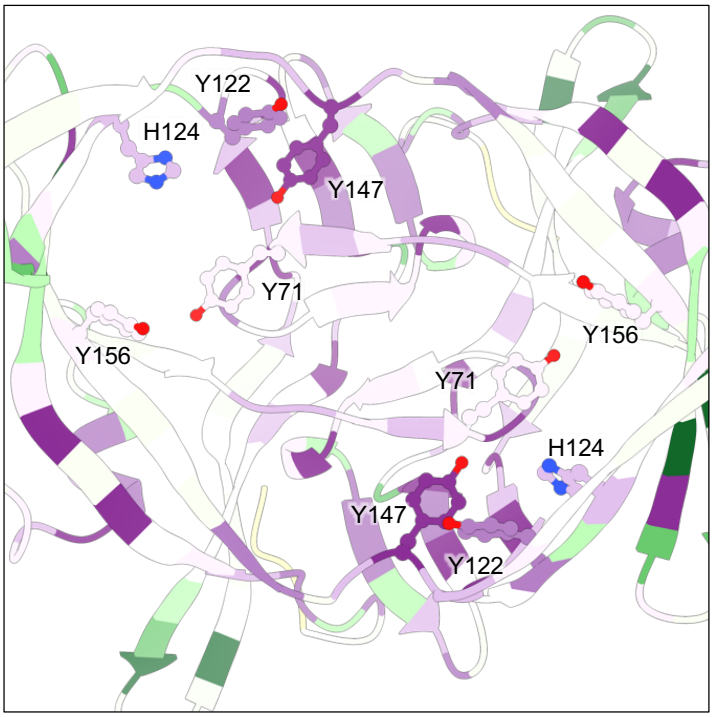


Figure 6-26: Conserved Residues Involved in the Binding of PEG. This shows the PEG binding site from dimer Y.

6.3 Discussion

6.3.1 Bd2778 a Divergent Structural Homolog of Bd1904

The three conformations of Bd2778 were compared to Bd1904, and there are clear similarities in the structures, but when superposed had an RMSD value of 22.3 Å (*form 1*), 22.7 Å (*form*

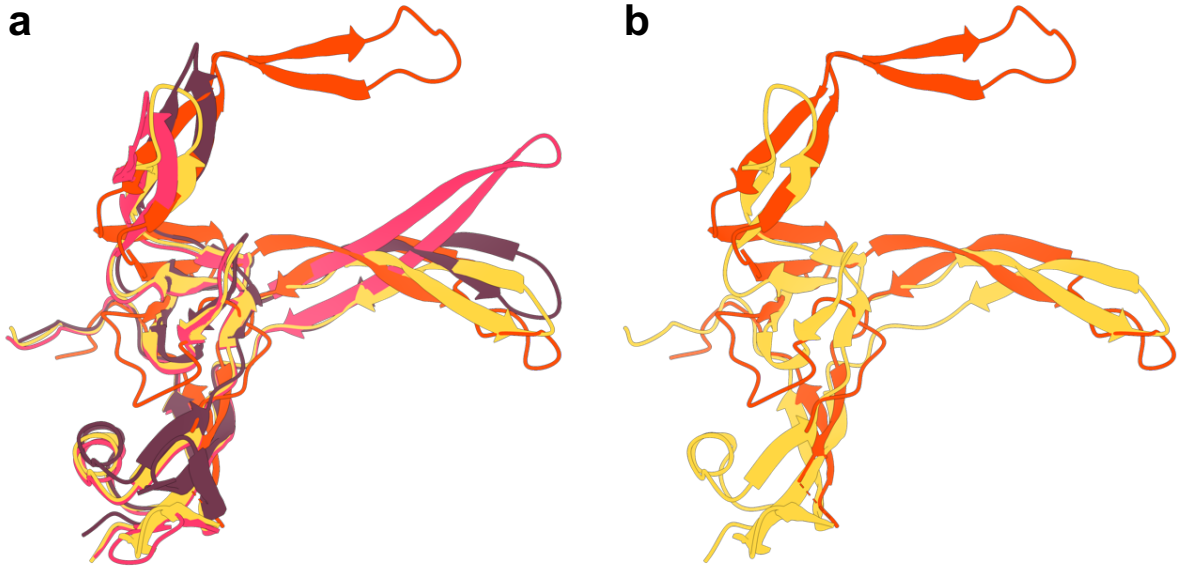


Figure 6-27: Superposition of Bd1904 to Bd2778 Forms. a) Bd1904 in dark orange superposition with Bd2778 '*form 1*' (yellow), '*form 2*' (pink), '*form 3*' (purple). b) Bd1904 superimposed with Bd2778 '*form 1*'.

2), and 21.7 Å (*form 3*) across 110 C_α pairs are indicative of the structural differences in the two proteins (Figure 6-27).

Both Bd1904 and Bd2778 have three β-hairpin arms, which whilst they do not align to give a low RMSD value most probably due to their flexible nature, are in very similar positions within

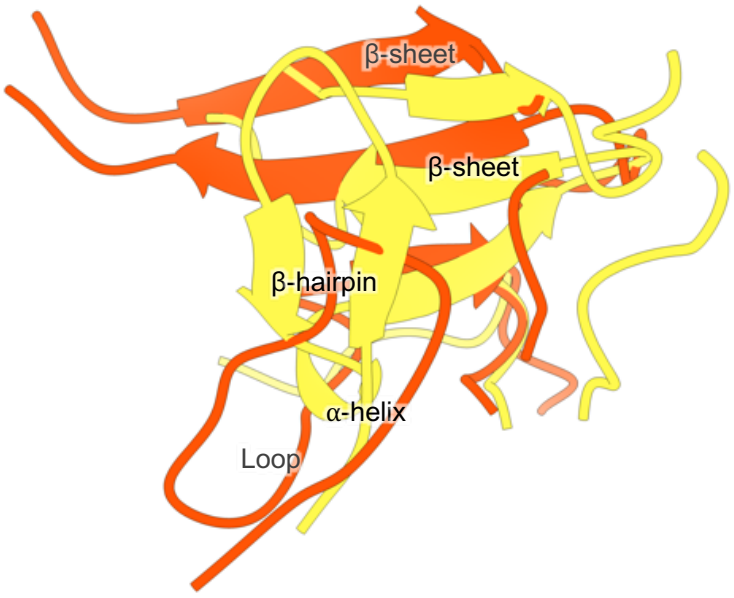


Figure 6-28: Bd1904 and Bd2778 Core Structure. Bd1904 dark orange with grey labels and Bd2778 yellow with black labels.

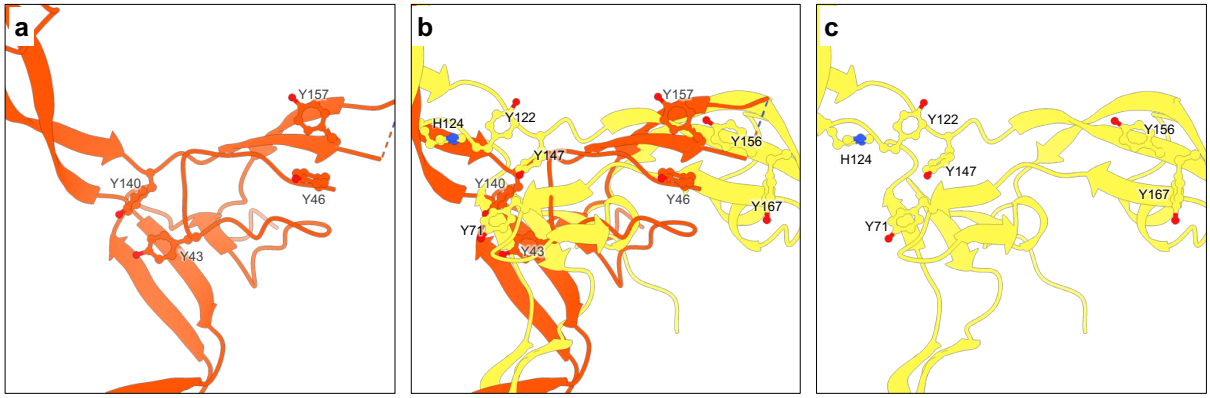


Figure 6-30: Solvent Exposed Tyrosine Residues in Bd1904 compared to Exposed Tyrosine and Histidine Residues Involved in the Binding of PEG in Bd2778. a) Bd1904 showing solvent exposed tyrosine residues. b) Bd1904 (dark orange) and Bd2778 (yellow) superimposed with solvent exposed tyrosine and histidine resides shown. c) Bd2778 (yellow) with solvent exposed tyrosine and histidine resides shown.

their respective structures. From comparison it can be observed that Bd1904 has a longer arm 1 than Bd2778, which explains the lack of conservation of the 5th and 6th cysteine residues, as they are not required to stabilise Bd2778's shorter arm. The core of the two proteins both contain a three-stranded twisted β-sheet between arms 1 and 2, with the main difference being Bd2778 which forms a β-hairpin followed by a short α-helix, whilst the Bd1904 structure forms a loop in this region (Figure 6-28).

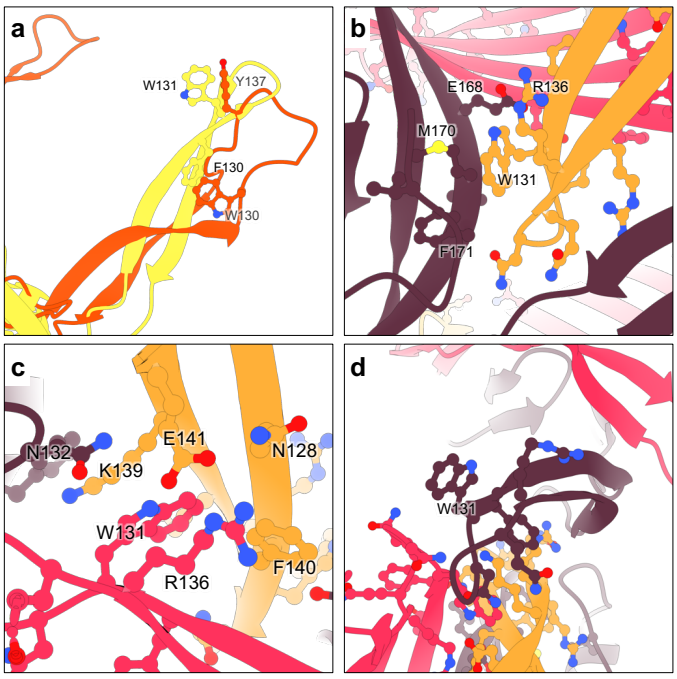


Figure 6-29: Tryptophan 131 in Bd2778. a) Tryptophan residues in arm 2 of Bd2778 (yellow) and Bd1904 (dark orange). b-d) Trp131 in Bd2778 '*form 1*' (b), '*form 2*' (c) and *'form 3'* (d).

In Bd2778 the tyrosine residues involved in the binding of PEG that are conserved and exposed to solvent, are observed in similar positions to the tyrosine residues exposed to solvent in Bd1904 (Figure 6-30). This poses the question as to whether these surface exposed tyrosine residues are important to the function of these two proteins. They could be involved in the binding of a ligand to prevent the hydrophobic aromatic ring of these residues being exposed to solvent, which would be energetically unfavourable. Both structures also have an exposed tryptophan (Trp130 (Bd1904) and Trp131 (Bd2778)) in arm 2, as well as an additional residue containing an aromatic ring (Tyr137 (Bd1904) and Phe130 (Bd2778)), these residues again would typically be found in hydrophobic regions of a protein fold (Figure 6-29a). In Bd2778 the tryptophan is found in three enviroments, one per form (Figure 6-29b-d), with the residue playing an important role in the packing of the interface for the dodecamer.

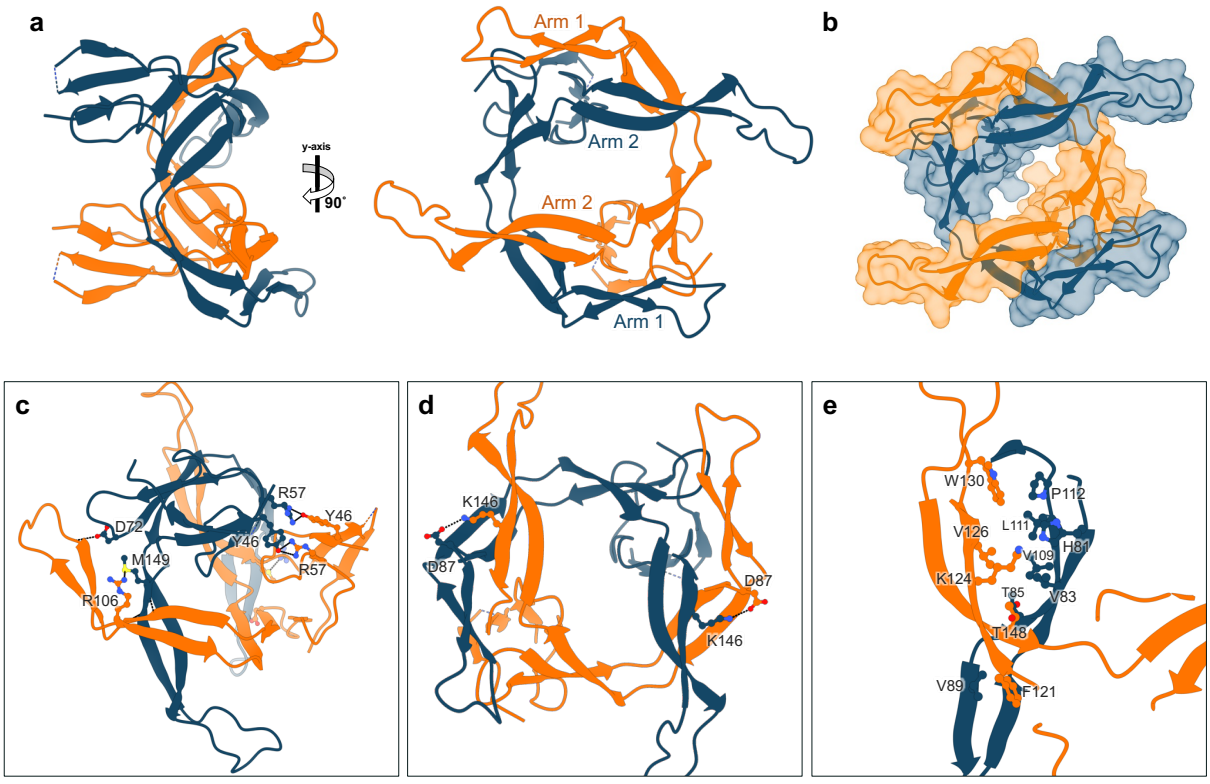


Figure 6-31: Bd1904 Dimer, including Interactions involved in Dimer Formation. a) Bd1904 dimer, crystal structure determined by Harding (unpublished). b) Bd1904 dimer surface. Hydrogen bonds (c), salt bridges (d) and hydrophobic interactions (e) between the two chains in the dimer.

Looking at the crystal contacts in the Bd1904 structure it can be seen that Bd1904 packs as a dimer (Harding, unpublished). Therefore, the structure of this dimer (Figure 6-31) was also compared to the dimer forms of Bd2778 (Figure 6-18). Immediately it can be observed that the Bd1904 dimer has a different interface to Bd2778. Bd2778 forms a dimer by the packing of each β-strand 3, from the β-hairpin in the core of each protomer, as well as interractions from arms 1 and 2. Bd1904 does not contain a β-hairpin in the core and within the dimer forms nearly all interactions (hydrogen bonding, salt bridges, and van der Waals) between just arms 1 and 2, with two hydrogen bonds (Try46 and R57) occuring between the cores of each protomer (Figure 6-31c-e). This results in the formation of a small hole (~ 10 Å) between the two protomers (Figure 6-31b), not seen in Bd2778.

The symmetry within the Bd1904 crystal revealed that a Bd1904 potentally forms a multimer, though this runs coincident to the P 6_4 axis (Figure 6-32). This multimer produces a long '*tube-like*' structure formed of two left handed helices running alongside each other, made up of multiple copies of the Bd1904 dimer (Figure 6-32). This results in the formation of a central

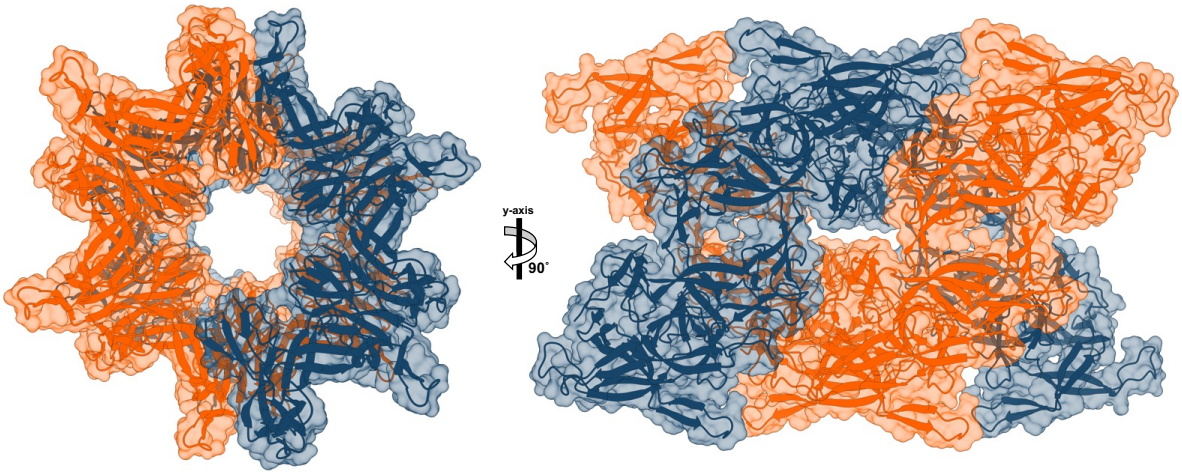


Figure 6-32: Bd1904 Multimer, observed from Crystal Symmetry Copies. Bd1904 structure determined by Harding (unpublished).

solvent channel (~35 Å, C_α-C_α) down the z-axis of the unit cell. Therefore both oligomers of Bd1904 and Bd2778 form a solvent filled channel. If the oligomers are native and not a result of crystal packing, could these channels be relevant to function?

6.3.2 PEG Binding Mimics Crown Ether Binding Motif

All protomers of Bd2778 bound PEG molecules of varying lengths around a lysine residue (Lys66), forming van der Waals contacts with the aromatic rings of tyrosine residues. This binding of PEG around a lysine residue has previously been observed in anti-PEG antibodies (Figure 6-33), where it was shown to mimic binding of a crown ether (Lee et al., 2020), well known to bind to surface exposed lysine or arginine sidechains (Gokel et al., 2004).

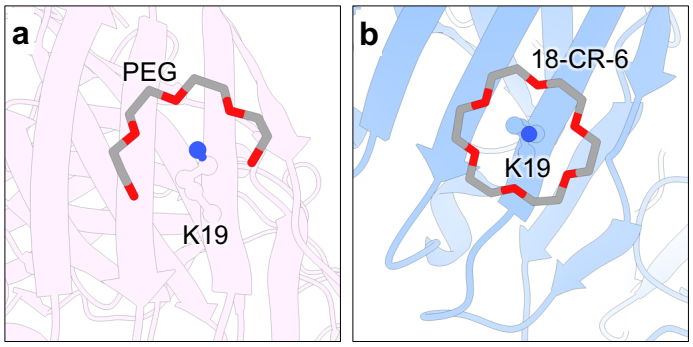


Figure 6-33: Binding of PEG and Crown Ether around Lysine in Anti-PEG Antibodies. a) PEG around lysine residue (PDB: 6JU0). b) Crown ether around lysine residue (PDB: 6JWC).

As PEG and crown ethers are synthetic compounds, these are not going to be ligands of Bd2778 and given the common binding of crown ether (a PEG-like compound) to lysine/arginine sidechains, it is probable that this is not acting as a mimic for a native ligand. However, as mentioned above it is possible that the site PEG is bound to is a binding site for a yet unknown ligand due to the conserved and exposed tyrosine residues that are seen in both Bd2778 and Bd1904.

6.3.3 Working Hypotheses of Bd1904 and Bd2778 Function

The gene *bd1904* is the most highly upregulated gene upon entry into the prey's periplasm, suggesting that the protein it encodes has a role between the invasion and growth phases of the life cycle. Structural homologs to Bd1904 identified by Foldseek (Figure 6-34), only gave

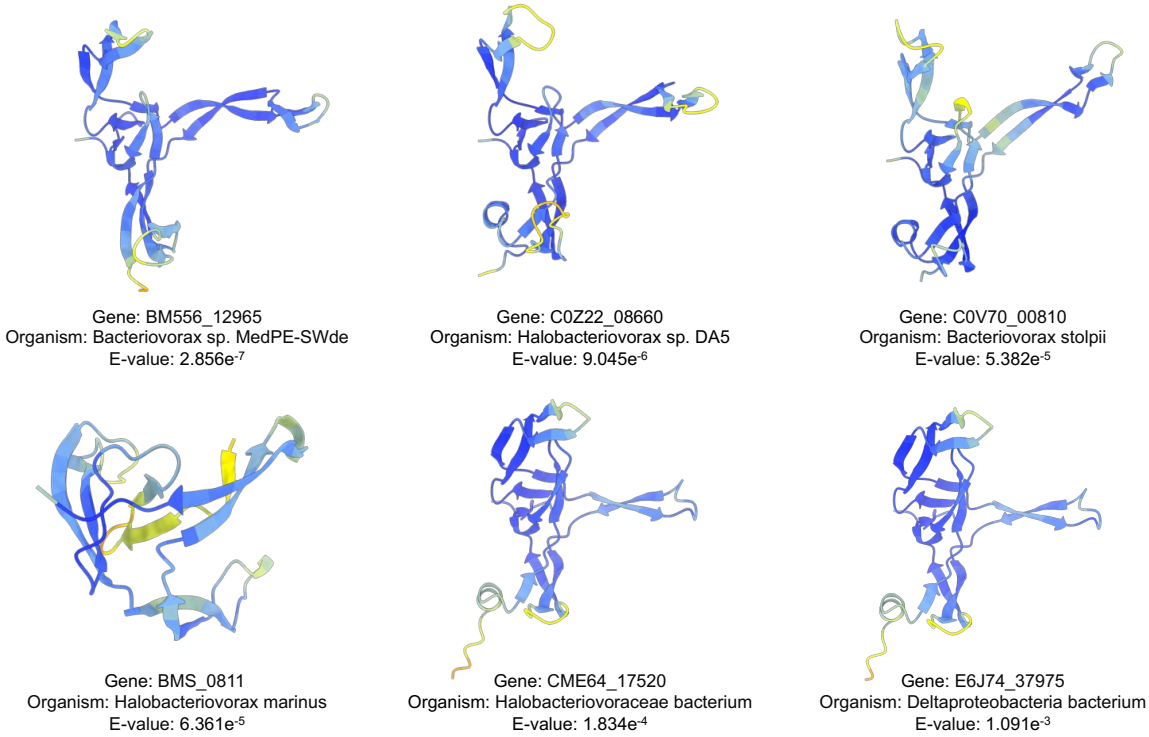


Figure 6-34: Results from Foldseek using Bd1904 as the Search Model. Models coloured using AlphaFold colour palette. Only one protein is shown from each species.

results from closely related micro-predators, also with unknown functions. This suggests that the fold may be specific for a predatory function. Also, the divergent structural homolog Bd2778 identified additional hits from *Elusimicrobia* bacterium, *Hahella sp.* and *Marinobacterium aestuarii* (Figure 6-35), again with no predicted function, which have not been observed to be predators (Bae et al., 2018; Méheust et al., 2020; Sam et al., 2017). Therefore, if Bd1904 and Bd2778 do possess the same function, it is therefore possible that the role may not be specific to predation of bacteria. The search with both Bd1904 and Bd2778 on Foldseek identified two proteins (C0V70_00810 and E6J74_37975) that were common to both searches. As Foldseek identifies structural homologs to protein models, if Bd1904 and Bd2778 produce common results this supports the finding that they are structural homologs of one another.

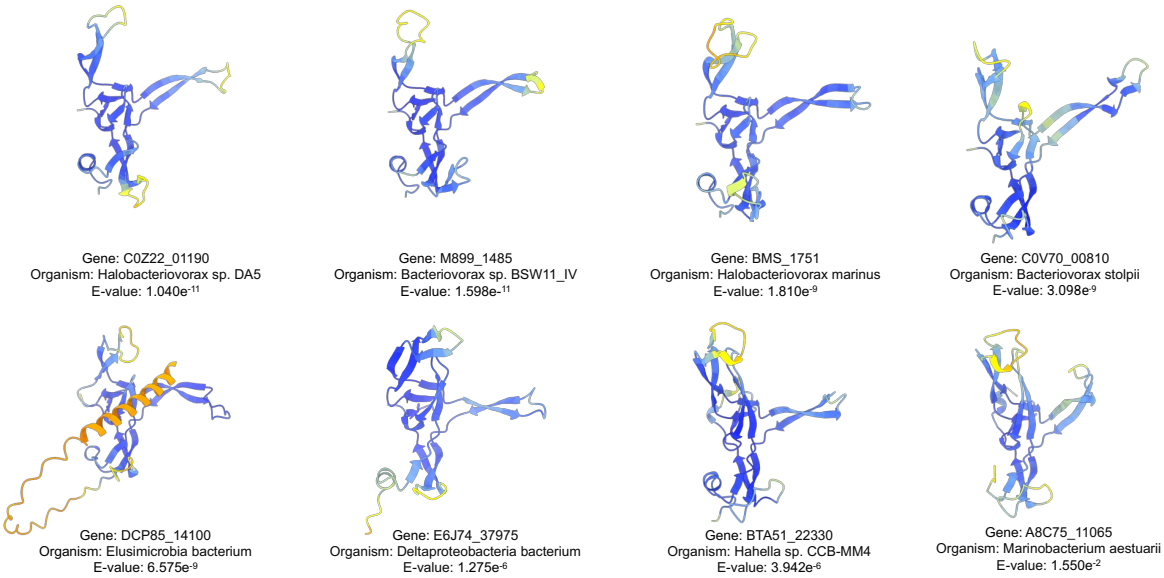


Figure 6-35: Results from Foldseek using Bd2778 as a Search Model. Models coloured using AlphaFold colour palette. Only one protein is shown from each species.

As none of these proteins aided identification of a function, we have proposed two working hypotheses for the function of Bd1904 and Bd2778 based on observations of the structure.

The first is proposed due to the formation of the Bd2778 dodecamer and Bd1904 multimer. Is it possible that all the Bd1904 family form multimers? If so, the proteins could be forming these multimers to act as structural proteins. Based on the timing of the upregulation of *bd1904*, it is possible that these proteins are expressed upon entry into the prey, to be released by the predator from the periplasm where they oligomerise to reinforce the bdelloplast structure.

The second hypothesis is due to the β-hairpin arms of the dimeric Bd2778 and monomeric Bd1904 being reminiscent of DNA-binding protein, integration host factor (IHF) (Figure 6-36).

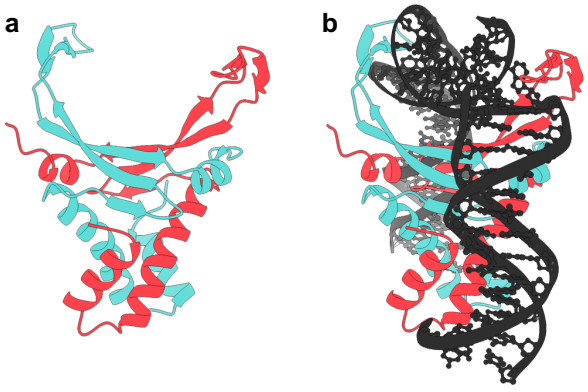


Figure 6-36: Integration Host Factor, with (b) and without (a) DNA. PDB 1IHF.

Is it possible that the Bd1904 family binds DNA within the prey? Both the Bd1904 and Bd2778 '*arms*' contain hydrophobic sidechains which could interact with DNA bases and Bd1904 contains a proline residue in the β-hairpin turn (Figure 6-37b) which is characteristic of IHF proteins, where the proline is inserted in between base pairs at the kink in the bent DNA (Figure 6-37a)(Rice et al., 1996; Yildiz et al., 2006). The binding of DNA could also explain the exposed tyrosine sidechains from the core of the protein. These sidechains, as with the hydrophobic sidechains in the '*arm*', would be able to stack with the bases in DNA preventing their exposure to the solvent. If Bd2778 and Bd1904 can bind DNA in a non-specific manner, it is possible that they are secreted into prey during predation to prevent transcription if they bind DNA. If the proteins do bind DNA, it would explain why they would have long signal peptides to delay folding, as binding of these proteins to *B. bacteriovorus* DNA would be toxic to the predator.

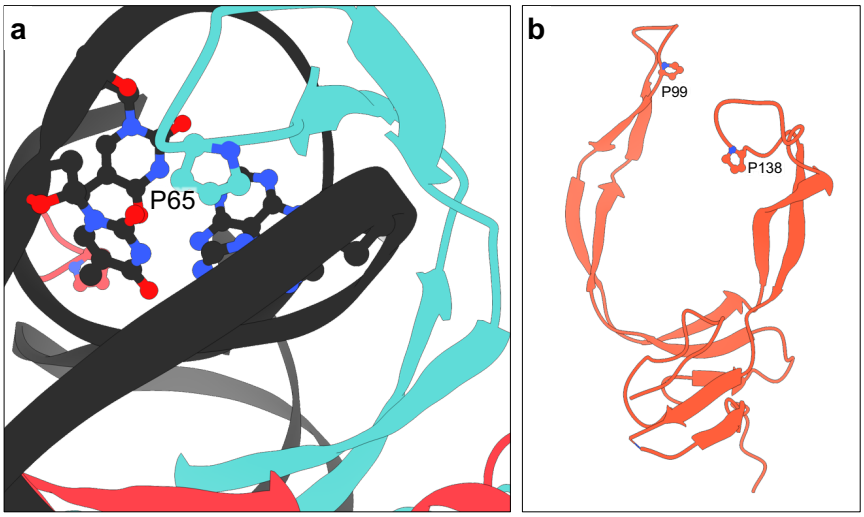


Figure 6-37: Proline Residues in β-Hairpin Turn of IHF (a) and Bd1904 (b).

6.3.4 ColabFold Models of Bd1645 and Bd2126 Confirm they are Structural Homologs to Bd1904 and Bd278

As AlphaFold has been shown to provide a good model for Bd2778, enabling solving of the phase problem of the diffraction dataset, the ColabFold models of Bd1645 and Bd2126 (Figure 6-39) have been analysed to confirm if they are structural homologs. This reveals that both have

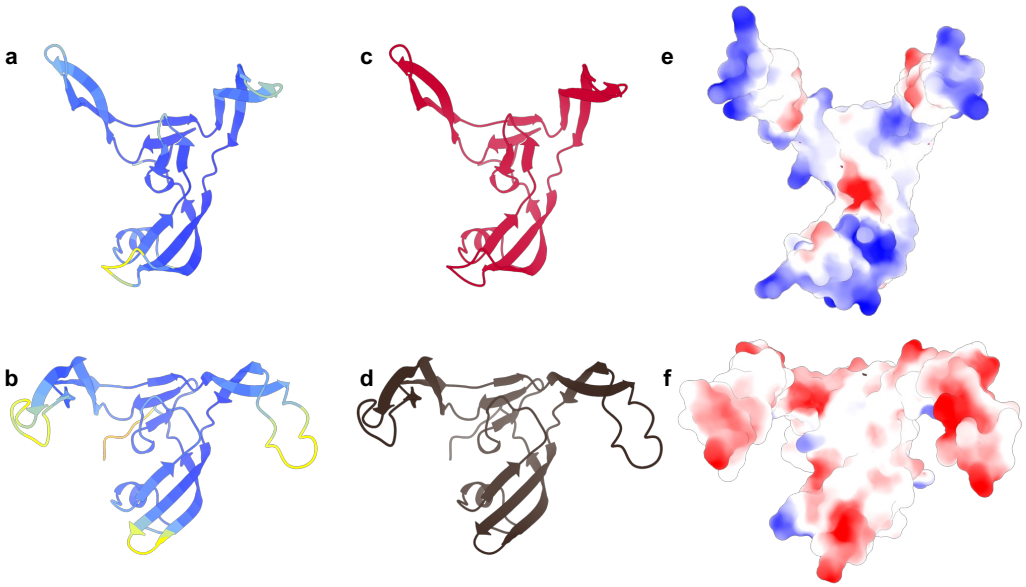


Figure 6-39: Predicted Structures of Bd1645 and Bd2126 by ColabFold. a&b) Bd1645 (a) and Bd2126 (b) coloured to show model confidence. c&d) Overall fold of Bd1645 (c) and Bd2126 (d). e&f) Surface model of Bd1645 (e) and Bd2126 (f) coloured to show electrostatics.

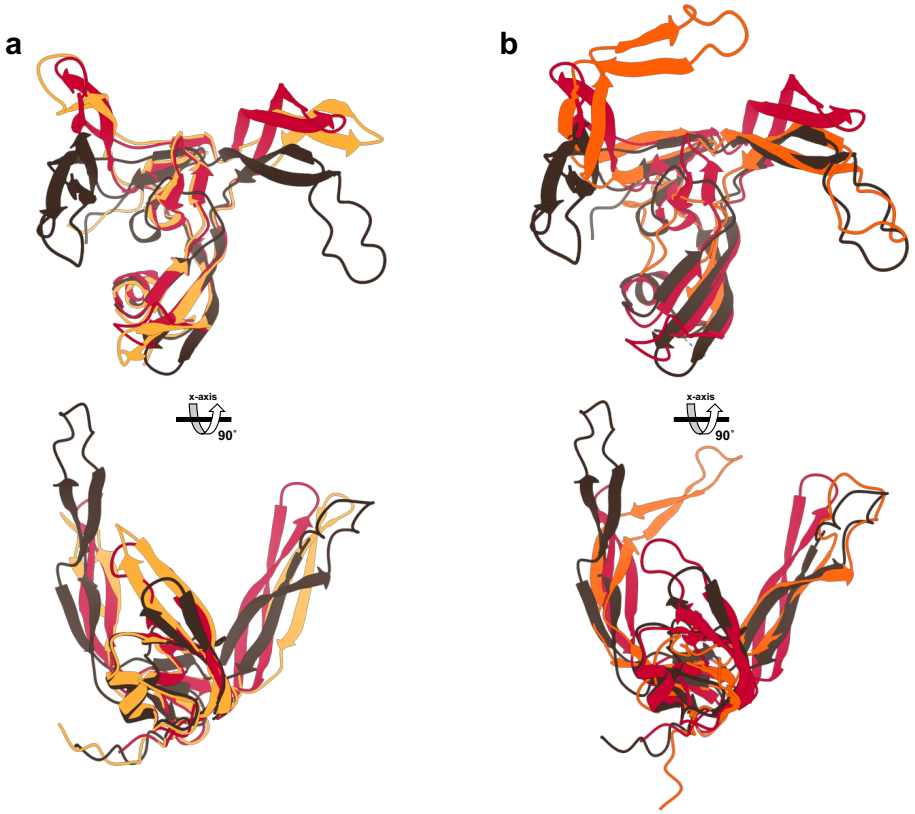


Figure 6-38: Superposition of Bd1645 and Bd2126 to Bd2778 (a) and Bd1904 (b). a) Bd1645 in red, Bd2126 in brown, and Bd2778 in orange. b) Bd1645 in red, Bd1904 in dark orange, and Bd2126 in brown.

three β-hairpin arms, with a central core containing a three-stranded β-sheet and a short α-helix (Figure 6-39c-d). Bd1645 also has a β-hairpin (Figure 6-39c) within the core similar to Bd2778; Bd2126 instead contains a loop in this region (Figure 6-39d). Similar to Bd1904, Bd2126 has a longer '*arm 1*' than Bd1645 and Bd2778, with the longer arm stabilised by a disulphide. These similarities can be readily observed when the models are superposed (Figure 6-38). Bd1645 and Bd2126 have RMSD values of 3.0 Å (131 C_α pairs) and 9.2 Å (128 C_α pairs), respectively, when aligned to Bd2778 (Figure 6-38a) and 4.9 Å (18 C_α pairs) and 12.9 Å (117 C_α pairs), respectively, when aligned to Bd1904 (Figure 6-38b). This enabled us to conclude that these two proteins are structural homologs to Bd1904 and Bd2778.

Looking at the other features observed in Bd1904 and Bd2778, Bd1645 was also found to have residues with aromatic sidechains in similar positions to the tyrosine residues involved in the binding of PEG (Figure 6-40a-b). Whereas Bd2126 only had two exposed tyrosine residues in similar positions (Figure 6-40c-d), this along with the significantly different electrostatic surface (mostly patches of negative charge) suggests that Bd2126 may have a different role to the other three proteins. It was also observed that Bd1645 and Bd2126 did not contain a tryptophan residue in '*arm 2*', though both did contain at least one residue with an aromatic sidechain towards the end of the '*arm*' (Figure 6-41a). Bd2126 did contain a tryptophan in *'arm*

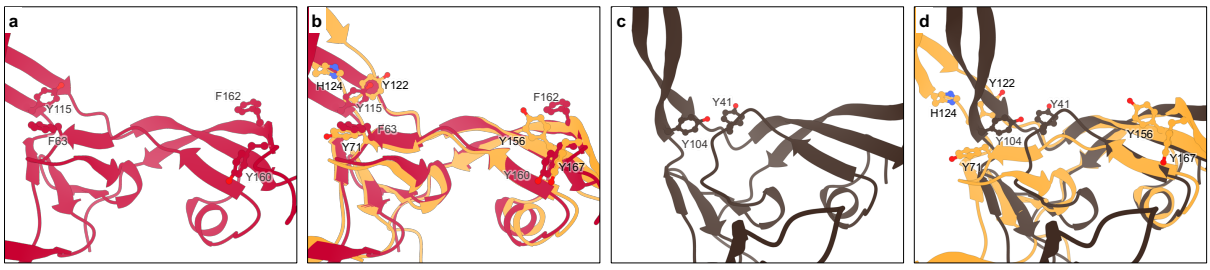


Figure 6-40: Solvent Exposed Residues with Aromatic Rings Involved in Binding PEG in Bd2778 compared to Residues in Bd1645 and Bd2126. a) Bd1645 exposed tyrosine and phenylalanine residues. b) Comparison of Bd1645 (red) exposed residues with aromatic rings to Bd2778 (orange). a) Bd2126 exposed tyrosine residues. b) Comparison of Bd2126 (brown) exposed tyrosine residues to Bd2778 (orange).

1' in a similar position to the position of the tryptophan residue in *'arm 2'* for Bd1904 and Bd2778 (Figure 6-41b). This results in hydrophobic residues being present in at least one arm in each protein, as it would be unfavourable it is likely that these residues have a structural or functional role in the protein.

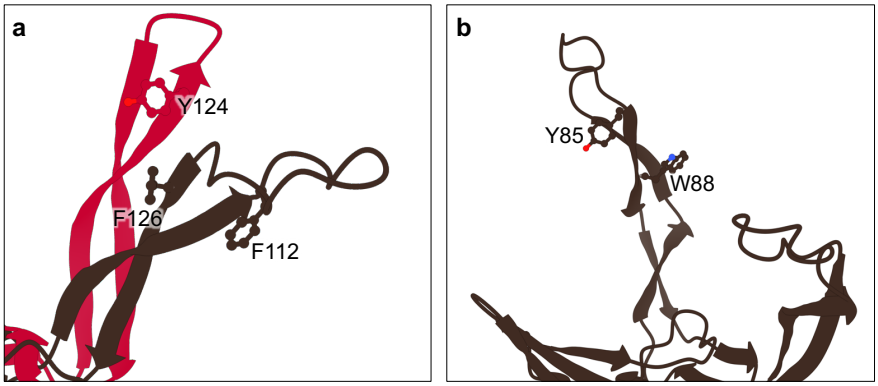


Figure 6-41: Residues with Aromatic Sidechains within '*Arm 2*' (a) and '*Arm 1*' (a) in Bd1645 (red) and Bd2126 (brown).

CHAPTER SEVEN: CONCLUDING REMARKS AND FUTURE DIRECTIONS

7.1 Identification of a New Superfamily of Oligomeric Outer Membrane Proteins

One of the most exciting discoveries during these studies was that Bd0427 was shown to form a homopentamer - a previously unseen oligomeric arrangement for this type of outer membrane protein. This was observed using two independent methods: refold of Bd0427 from inclusion body material in β-OG, which was used to grow protein crystals destined for X-ray diffraction, and solubilisation and purification of Bd0427 with a His-tag in a turn, expressed natively in the membrane for use in cryo-EM. The homopentamer was also predicted by AlphaFold, even when asked to search for a dimer and trimer the interface remained mostly unchanged (Figure 4-24). This all resulted in a high degree of confidence that this is the native oligomeric state of Bd0427 and not an artifact of any of the techniques used. The oligomer also revealed that the pentameric interface was unique to this homopentamer, involving loops not used by homotrimers (loops 7 and 8), as well as using loops used by homotrimers in different ways (see Table 4-4). This provided the initial evidence that Bd0427 is a new class of OMP.

To further investigate the protein, Bd0427 was initially compared to the current database of published protein models which did not reveal any structural homologs (Figure 4-18), again supporting that a new class had been found. A search using Foldseek however, did identify several structural homologs from the AlphaFold database of predicted structures (Figure 4-23). This revealed that the Bd0427 fold, 16-stranded β-barrel with an N-terminal plug, that is asymmetric in height, with a long loop 8 that folds away from the barrel, is seen not only in

closely related predators (*B. exovorus*, *Bacteriovorax* sp. and *Halobacteriovorax* sp.) but also in other non-predatory families/species, indicating the fold is more widespread than just being found in predatory bacteria. When these protein sequences identified by Foldseek were folded as dimers, pentamers or hexamers using a combination of ColabFold and AlphaFold, the predicted oligomeric state of these proteins were found to be either pentamers (from *Bacteriovorax* sp. and *Halobacteriovorax* sp.) or hexamers (from the non-predatory families) (Figure 4-25), confirming that the protein fold was found across several families of bacteria and that these all form either pentamers or hexamers, concluding that a new class of OMP had been identified.

Although not the original intention of this PhD, it would now be interesting to further investigate this new class of OMPs. Firstly, to confirm that the predicted structural homologs from the AlphaFold database are indeed structural homologs. Then secondly, to confirm the oligomeric states of a selection of these proteins (both pentameric and hexameric proteins). This can then be used to study the differences between the interfaces of the pentameric and hexameric proteins, as well as enable investigation into the similarities in the proteins that form pentamers and the proteins that form hexamers. This will aid analysis into why this fold can form either pentamers or hexamers. Thirdly, as two of the proteins identified by Foldseek have predicted functions it would be of particular interest to determine the structure of these two proteins, especially as the oligomeric state of TDE_2508 (*Treponema*) could not be predicted by AlphaFold. This would reveal any features that can be associated to their predicted functions.

7.2 What has been learnt about the Function of Bd0427

In this thesis I have determined that Bd0427 forms an N-terminally plugged 16-stranded β-barrel, which does not form an '*open*' channel when viewed as a surface model. This N-terminal

plug has homology to the N-terminal plug in FadL (van den Berg et al., 2004), which has been shown to completely block the channel with observations that transport occurs through an opening in the channel wall (Hearn et al., 2009). Despite this we have shown that Bd0427 does generate a pore in the membrane, as observed by electrophysiology, though more measurements are required to fully characterise the conductance of the pore. Alongside this we have shown that glycine and alanine are clearly transported by Bd0427, with larger substrates such as glucose and maltose not transported (Figure 5-13). This provides initial evidence that Bd0427 functions as a porin, but with the selection likely to be for small substrates (<90 $g.mol^{-1}$). Therefore, it is unlikely that Bd0427 is the sole transporter of nutrients from the cytoplasm to the periplasm of a bdelloplast during predation, as many nutrients will be larger than 90 $g.mol^{-1}$. This is further supported by the in-frame deletion of *bd0427* which demonstrated the gene was not essential. If this gene product was the sole nutrient transporter *bd0427* would be essential, meaning one of the original hypotheses for the function of this protein has been disputed.

The other original hypothesis for Bd0427 was that it disrupted the proton motive force across the inner membrane of a bdelloplast, as seen with PorB from *N. gonorrhoea*. This could still be a possible function as *B. bacteriovorus* could potentially still complete the predatory life cycle without disrupting the proton motive force of the prey cell. However, this is unlikely to be a function of this protein fold in the non-predatory bacteria that possess Bd0427 homologues. This casts some doubt on whether the function is specifically related to the predation cycle, and therefore involved in the disruption of the proton motive force.

From the structural homologs identified by Foldseek, two proteins (TDE_2508 and FSU_2404) were found to have predicted functions. As previously discussed, the proteins are predicted to regulate biofilm formation and adhesion to epithelia and be involved in vesicle formation,

respectively. Whilst it is possible that Bd0427 performs one of these functions there is currently no clear evidence to support this.

Together this means at this time we have no data to confirm a function of Bd0427 but three potential functions can be hypothesised:

1. Bd0427 could still disrupt the proton motive force across the IM of the bdelloplast
2. Bd0427 may be involved in the formation of vesicles utilised by *B. bacteriovorus*
3. Bd0427 could be an adhesion protein

It is also very possible that the function is yet to be identified. To further the identification of a function more experimentation is required. Firstly, it is important to fully characterise the pore by electrophysiology to provide a conductance value for Bd0427 which can be compared to other porins. Modelling could also be used to predict what movement within the protein channel would provide the measured conductance value, aiding identification of whether the plug is be removed or not. Electrophysiology experiments of a disulfide locked plug could also be conducted alongside this to look for differences in the conductance value of the mutant and wild-type protein. Secondly, the in-frame deletion strain needs to be fully characterised to determine if there is any change to the rate of predation, number of progeny produced or any other visible differences. The strain could also be tested in competition with wild-type *B. bacteriovorus* to see if there is a deficit resulting in it being out competed by the wild-type. Hypothesis 1 and 2 could be looked at specifically to see if the prey cell dies at a different time point to the wild-type and to see if there is reduced formation of vesicles during predation.

7.3 Bd1904, Bd1645, Bd2126 and Bd2778 are Structural Homologs

This thesis set out to determine if Bd1645, Bd21216 and Bd2778 were structural homologs of Bd1904. The determined structure of Bd2778 and ColabFold models of Bd1645 and Bd2126 indicate that these proteins are all structural homologs of Bd1904. This has enabled the identification of similar features between these proteins such as the exposed aromatic sidechains from the core of the protein and the tryptophan residue found in '*arm 2*' or '*arm 1*'. These features all have the potential to be important for a shared function. During these studies no function for these homologs has been identified but we have proposed two potential working hypotheses: (1) they act as structural proteins that support the bdelloplast or (2) they are DNA binders, potentially to block translation.

To continue the investigation into the function of these proteins it would be useful to validate the models of Bd1645 and Bd2126 by determining their structure. This may also give insight into whether these form oligomeric states, which can then be compared to the Bd2778 dimer and dodecamer, as well as the Bd1904 dimer and potential higher oligomer. This may aid determination as to whether hypothesis (1) is true. Alongside this, experiments can be conducted to determine if any of the proteins are capable of binding double-stranded or single-stranded DNA, to determine if hypothesis (2) could be true.

LITERATURE CITED

Abiko, Y., Nagano, K., Yoshida, Y., Yoshimura, F., 2014. Major Membrane Protein TDE2508 Regulates Adhesive Potency in Treponema denticola. PLoS ONE 9, e89051. https://doi.org/10.1371/journal.pone.0089051

Abraham, E.P., Chain, E., 1940. An Enzyme from Bacteria able to Destroy Penicillin. Nature 146, 837–837. https://doi.org/10.1038/146837a0

Abram, D., e Melo, J.C., Chou, D., 1974. Penetration of *Bdellovibrio bacteriovorus* into Host Cells. J. Bacteriol. 118, 663–680. https://doi.org/10.1128/jb.118.2.663-680.1974

Afonine, P.V., Grosse-Kunstleve, R.W., Echols, N., Headd, J.J., Moriarty, N.W., Mustyakimov, M., Terwilliger, T.C., Urzhumtsev, A., Zwart, P.H., Adams, P.D., 2012. Towards automated crystallographic structure refinement with *phenix.refine*. Acta Crystallogr. D Biol. Crystallogr. 68, 352–367. https://doi.org/10.1107/S0907444912001308

Afonine, P.V., Poon, B.K., Read, R.J., Sobolev, O.V., Terwilliger, T.C., Urzhumtsev, A., Adams, P.D., 2018. Real-space refinement in *PHENIX* for cryo-EM and crystallography. Acta Crystallogr. Sect. Struct. Biol. 74, 531–544. https://doi.org/10.1107/S2059798318006551

Allen, W.J., Corey, R.A., Oatley, P., Sessions, R.B., Baldwin, S.A., Radford, S.E., Tuma, R., Collinson, I., 2016. Two-way communication between SecY and SecA suggests a Brownian ratchet mechanism for protein translocation. eLife 5, e15598. https://doi.org/10.7554/eLife.15598

Allen, W.J., Watkins, D.W., Dillingham, M.S., Collinson, I., 2020. Refined measurement of SecA-driven protein secretion reveals that translocation is indirectly coupled to ATP turnover. Proc. Natl. Acad. Sci. 117, 31808–31816. https://doi.org/10.1073/pnas.2010906117

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. https://doi.org/10.1128/AEM.00426-11

Atterbury, R.J., Tyson, J., 2021. Predatory bacteria as living antibiotics – where are we now? Microbiology 167. https://doi.org/10.1099/mic.0.001025

Avidan, O., Petrenko, M., Becker, R., Beck, S., Linscheid, M., Pietrokovski, S., Jurkevitch, E., 2017. Identification and Characterization of Differentially-Regulated Type IVb Pilin Genes Necessary for Predation in Obligate Bacterial Predators. Sci. Rep. 7, 1013. https://doi.org/10.1038/s41598-017-00951-w

Baek, M., DiMaio, F., Anishchenko, I., Dauparas, J., Ovchinnikov, S., Lee, G.R., Wang, J., Cong, Q., Kinch, L.N., Schaeffer, R.D., Millán, C., Park, H., Adams, C., Glassman, C.R., DeGiovanni, A., Pereira, J.H., Rodrigues, A.V., van Dijk, A.A., Ebrecht, A.C., Opperman, D.J., Sagmeister, T., Buhlheller, C., Pavkov-Keller, T., Rathinaswamy, M.K., Dalwadi, U., Yip, C.K., Burke, J.E., Garcia, K.C., Grishin, N.V., Adams, P.D., Read, R.J., Baker, D., 2021. Accurate prediction of protein structures and interactions

using a three-track neural network. Science 373, 871–876. https://doi.org/10.1126/science.abj8754

Banks, E.J., Valdivia-Delgado, M., Biboy, J., Wilson, A., Cadby, I.T., Vollmer, W., Lambert, C., Lovering, A.L., Sockett, R.E., 2022. Asymmetric peptidoglycan editing generates cell curvature in Bdellovibrio predatory bacteria. Nat. Commun. 13, 1509. https://doi.org/10.1038/s41467-022-29007-y

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. https://doi.org/10.1016/j.ygeno.2007.06.002

Barel, G., Sirota, A., Volpin, H., Jurkevitch, E., 2005. Fate of Predator and Prey Proteins during Growth of *Bdellovibrio bacteriovorus* on *Escherichia coli* and *Pseudomonas syringae* Prey. J. Bacteriol. 187, 329–335. https://doi.org/10.1128/JB.187.1.329-335.2005

Bauer, B.W., Shemesh, T., Chen, Y., Rapoport, T.A., 2014. A "Push and Slide" Mechanism Allows Sequence-Insensitive Translocation of Secretory Proteins by the SecA ATPase. Cell 157, 1416–1429. https://doi.org/10.1016/j.cell.2014.03.063

Beck, S., Schwudke, D., Appel, B., Linscheid, M., Strauch, E., 2005. Characterization of outer membrane protein fractions of *Bdellovibrionales*. FEMS Microbiol. Lett. 243, 211–217. https://doi.org/10.1016/j.femsle.2004.12.006

Beck, S., Schwudke, D., Strauch, E., Appel, B., Linscheid, M., 2004. *Bdellovibrio bacteriovorus* Strains Produce a Novel Major Outer Membrane Protein during Predacious Growth in the Periplasm of Prey Bacteria. J. Bacteriol. 186, 2766–2773. https://doi.org/10.1128/JB.186.9.2766-2773.2004

Behrens, S., 2001. The SurA periplasmic PPIase lacking its parvulin domains functions in vivo and has chaperone activity. EMBO J. 20, 285–294. https://doi.org/10.1093/emboj/20.1.285

Berg, B. van den, Clemons, W.M., Collinson, I., Modis, Y., Hartmann, E., Harrison, S.C., Rapoport, T.A., 2004. X-ray structure of a protein-conducting channel. Nature 427, 36–44. https://doi.org/10.1038/nature02218

Berleman, J.E., Chumley, T., Cheung, P., Kirby, J.R., 2006. Rippling Is a Predatory Behavior in *Myxococcus xanthus*. J. Bacteriol. 188, 5888–5895. https://doi.org/10.1128/JB.00559-06

Bertani, B., Ruiz, N., 2018. Function and Biogenesis of Lipopolysaccharides. EcoSal Plus 8, ecosalplus.ESP-0001-2018. https://doi.org/10.1128/ecosalplus.ESP-0001-2018

Biou, V., 2023. Lipid-membrane protein interaction visualised by cryo-EM: A review. Biochim. Biophys. Acta BBA - Biomembr. 1865, 184068. https://doi.org/10.1016/j.bbamem.2022.184068

Bitto, E., McKay, D.B., 2002. Crystallographic Structure of SurA, a Molecular Chaperone that Facilitates Folding of Outer Membrane Porins. Structure 10, 1489–1498. https://doi.org/10.1016/S0969-2126(02)00877-8

Blot, N., Berrier, C., Hugouvieux-Cotte-Pattat, N., Ghazi, A., Condemine, G., 2002. The Oligogalacturonate-specific Porin KdgM of Erwinia chrysanthemi Belongs to a New Porin Family. J. Biol. Chem. 277, 7936–7944. https://doi.org/10.1074/jbc.M109193200

Bonfiglio, G., Neroni, B., Radocchia, G., Pompilio, A., Mura, F., Trancassini, M., Di Bonaventura, G., Pantanella, F., Schippa, S., 2019. Growth Control of Adherent-Invasive Escherichia coli (AIEC) by the Predator Bacteria Bdellovibrio bacteriovorus: A New Therapeutic Approach for Crohn's Disease Patients. Microorganisms 8, 17. https://doi.org/10.3390/microorganisms8010017

Borgnia, M.J., Subramaniam, S., Milne, J.L.S., 2008. Three-Dimensional Imaging of the Highly Bent Architecture of *Bdellovibrio bacteriovorus* by Using Cryo-Electron Tomography. J. Bacteriol. 190, 2588–2596. https://doi.org/10.1128/JB.01538-07

Breukink, E., Nouwen, N., van Raalte, A., Mizushima, S., Tommassen, J., de Kruijff, B., 1995. The C Terminus of SecA Is Involved in Both Lipid Binding and SecB Binding. J. Biol. Chem. 270, 7902–7907. https://doi.org/10.1074/jbc.270.14.7902

Bulieris, P.V., Behrens, S., Holst, O., Kleinschmidt, J.H., 2003. Folding and Insertion of the Outer Membrane Protein OmpA Is Assisted by the Chaperone Skp and by Lipopolysaccharide. J. Biol. Chem. 278, 9092–9099. https://doi.org/10.1074/jbc.M211177200

Burnham, J.C., Hashimoto, T., Conti, S.F., 1968. Electron Microscopic Observations on the Penetration of *Bdellovibrio bacteriovorus* into Gram-negative Bacterial Hosts. J. Bacteriol. 96, 1366–1381. https://doi.org/10.1128/jb.96.4.1366-1381.1968

Caroff, M., Karibian, D., 2003. Structure of bacterial lipopolysaccharides. Carbohydr. Res. 338, 2431–2447. https://doi.org/10.1016/j.carres.2003.07.010

Catipovic, M.A., Bauer, B.W., Loparo, J.J., Rapoport, T.A., 2019. Protein translocation by the SecA ATPase occurs by a power-stroke mechanism. EMBO J. 38. https://doi.org/10.15252/embj.2018101140

Catipovic, M.A., Rapoport, T.A., 2020. Protease protection assays show polypeptide movement into the SecY channel by power strokes of the SecA ATPase. EMBO Rep. 21. https://doi.org/10.15252/embr.202050905

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. https://doi.org/10.1016/j.mib.2020.07.002

Cavallo, F.M., Jordana, L., Friedrich, A.W., Glasner, C., van Dijl, J.M., 2021. *Bdellovibrio bacteriovorus* : a potential 'living antibiotic' to control bacterial pathogens. Crit. Rev. Microbiol. 47, 630–646. https://doi.org/10.1080/1040841X.2021.1908956

Chen, R., Henning, U., 1996. Aperiplasmic protein (Skp) of Escherichia coli selectively binds a class of outer membrane proteins. Mol. Microbiol. 19, 1287–1294. https://doi.org/10.1111/j.1365-2958.1996.tb02473.x

Coates, A.R., Halls, G., Hu, Y., 2011. Novel classes of antibiotics or more of the same?: New antibiotic classes are urgently needed. Br. J. Pharmacol. 163, 184–194. https://doi.org/10.1111/j.1476-5381.2011.01250.x

Collier, D.N., Bankaitis, V.A., Weiss, J.B., Bassford, P.J., 1988. The antifolding activity of SecB promotes the export of the E. coli maltose-binding protein. Cell 53, 273–283. https://doi.org/10.1016/0092-8674(88)90389-3

Cowan, S.W., Schirmer, T., Rummel, G., Steiert, M., Ghosh, R., Pauptit, R.A., Jansonius, J.N., Rosenbusch, J.P., 1992. Crystal structures explain functional properties of two E. coli porins. Nature 358, 727–733. https://doi.org/10.1038/358727a0

Cromer, D.T., Liberman, D., 1970. Relativistic Calculation of Anomalous Scattering Factors for X Rays. J. Chem. Phys. 53, 1891–1898. https://doi.org/10.1063/1.1674266

Crystallography Made Crystal Clear, 2006. . Elsevier. https://doi.org/10.1016/B978-0-12-587073-3.X5000-4

Czaplewski, L., Bax, R., Clokie, M., Dawson, M., Fairhead, H., Fischetti, V.A., Foster, S., Gilmore, B.F., Hancock, R.E.W., Harper, D., Henderson, I.R., Hilpert, K., Jones, B.V., Kadioglu, A., Knowles, D., Ólafsdóttir, S., Payne, D., Projan, S., Shaunak, S., Silverman, J., Thomas, C.M., Trust, T.J., Warn, P., Rex, J.H., 2016. Alternatives to antibiotics—a pipeline portfolio review. Lancet Infect. Dis. 16, 239–251. https://doi.org/10.1016/S1473-3099(15)00466-1

Dalbey, R.E., Wang, P., Kuhn, A., 2011. Assembly of Bacterial Inner Membrane Proteins. Annu. Rev. Biochem. 80, 161–187. https://doi.org/10.1146/annurev-biochem-060409-092524

Dashiff, A., Junka, R.A., Libera, M., Kadouri, D.E., 2011. Predation of human pathogens by the predatory bacteria Micavibrio aeruginosavorus and Bdellovibrio bacteriovorus: Predation by M. aeruginosavorus and B. bacteriovorus. J. Appl. Microbiol. 110, 431–444. https://doi.org/10.1111/j.1365-2672.2010.04900.x

Davidov, Y., Huchon, D., Koval, S.F., Jurkevitch, E., 2006. A new ?-proteobacterial clade of Bdellovibrio-like predators: implications for the mitochondrial endosymbiotic theory. Environ. Microbiol. 8, 2179–2188. https://doi.org/10.1111/j.1462-2920.2006.01101.x

Davidov, Y., Jurkevitch, E., 2004. Diversity and evolution of Bdellovibrio-and-like organisms (BALOs), reclassification of Bacteriovorax starrii as Peredibacter starrii gen. nov., comb. nov., and description of the Bacteriovorax–Peredibacter clade as Bacteriovoracaceae fam. nov. Int. J. Syst. Evol. Microbiol. 54, 1439–1452. https://doi.org/10.1099/ijs.0.02978-0

Dekker, C., Kruijff, B. de, Gros, P., 2003. Crystal structure of SecB from Escherichia coli. J. Struct. Biol. 144, 313–319. https://doi.org/10.1016/j.jsb.2003.09.012

Deo, P., Chow, S.H., Hay, I.D., Kleifeld, O., Costin, A., Elgass, K.D., Jiang, J.-H., Ramm, G., Gabriel, K., Dougan, G., Lithgow, T., Heinz, E., Naderer, T., 2018. Outer membrane vesicles from Neisseria gonorrhoeae target PorB to mitochondria and induce apoptosis. PLOS Pathog. 14, e1006945. https://doi.org/10.1371/journal.ppat.1006945

Diedrich, D.L., Duran, C.P., Conti, S.F., 1984. Acquisition of Escherichia coli outer membrane proteins by Bdellovibrio sp. strain 109D. J. Bacteriol. 159, 329–334. https://doi.org/10.1128/jb.159.1.329-334.1984

Diedrich, D.L., Portnoy, C.A., Conti, S.F., 1983. Bdellovibrio possesses a prey-derived OmpF protein in its outer membrane. Curr. Microbiol. 8, 51–56. https://doi.org/10.1007/BF01567314

Dodson, E., 2003. Is it jolly SAD? Acta Crystallogr. D Biol. Crystallogr. 59, 1958–1965. https://doi.org/10.1107/S0907444903020936

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. https://doi.org/10.1016/j.cell.2022.02.016

Duncan, M.C., Gillette, R.K., Maglasang, M.A., Corn, E.A., Tai, A.K., Lazinski, D.W., Shanks, R.M.Q., Kadouri, D.E., Camilli, A., 2019. High-Throughput Analysis of Gene Function in the Bacterial Predator Bdellovibrio bacteriovorus. mBio 10, e01040-19. https://doi.org/10.1128/mBio.01040-19

Eksztejn, M., Varon, M., 1977. Elongation and cell division in Bdellovibrio bacteriovorus. Arch. Microbiol. 114, 175–181. https://doi.org/10.1007/BF00410781

Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K., 2010. Features and development of *Coot*. Acta Crystallogr. D Biol. Crystallogr. 66, 486–501. https://doi.org/10.1107/S0907444910007493

Evans, K.J., Lambert, C., Sockett, R.E., 2007. Predation by *Bdellovibrio bacteriovorus* HD100 Requires Type IV Pili. J. Bacteriol. 189, 4850–4859. https://doi.org/10.1128/JB.01942-06

Evans, P., McCoy, A., 2008. An introduction to molecular replacement. Acta Crystallogr. D Biol. Crystallogr. 64, 1–10. https://doi.org/10.1107/S0907444907051554

Faller, M., Niederweis, M., Schulz, G.E., 2004. The Structure of a Mycobacterial Outer-Membrane Channel. Science 303, 1189–1192. https://doi.org/10.1126/science.1094114

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. https://doi.org/10.1128/JB.00914-10

Findlay, J.S., Flick-Smith, H.C., Keyser, E., Cooper, I.A., Williamson, E.D., Oyston, P.C.F., 2019. Predatory bacteria can protect SKH-1 mice from a lethal plague challenge. Sci. Rep. 9, 7225. https://doi.org/10.1038/s41598-019-43467-1

Flemming, A., 1929. On the Antibacterial Action of Cultures of a Penicillium, with Special Reference to their Use in the Isolation of B. influenzæ. Br. J. Exp. Pathol. 10, 226–236.

Flower, T.G., Hurley, J.H., 2021. Crystallographic molecular replacement using an in silico-generated search model of SARS-CoV-2 ORF8. Protein Sci. 30, 728–734. https://doi.org/10.1002/pro.4050

Friedel, G., 1913. Sur les symétries cristallines que peut révéler la diffraction des rayons Röntgen. Comptes Rendus 157, 1533–1536.

Galdiero, S., Falanga, A., Cantisani, M., Tarallo, R., Elena Della Pepa, M., D'Oriano, V., Galdiero, M., 2012. Microbe-Host Interactions: Structure and Role of Gram-Negative Bacterial Porins. Curr. Protein Pept. Sci. 13, 843–854. https://doi.org/10.2174/138920312804871120

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., Bairoch, A., 2005. Protein Identification and Analysis Tools on the ExPASy Server, in: Walker, J.M. (Ed.), The Proteomics Protocols Handbook. Humana Press, Totowa, NJ, pp. 571–607. https://doi.org/10.1385/1-59259-890-0:571

Gay, P., Le Coq, D., Steinmetz, M., Ferrari, E., Hoch, J.A., 1983. Cloning structural gene sacB, which codes for exoenzyme levansucrase of Bacillus subtilis: expression of the gene in Escherichia coli. J. Bacteriol. 153, 1424–1431. https://doi.org/10.1128/jb.153.3.1424-1431.1983

Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., Smith, H.O., 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods 6, 343–345. https://doi.org/10.1038/nmeth.1318

Gildea, R.J., Beilsten-Edmands, J., Axford, D., Horrell, S., Aller, P., Sandy, J., Sanchez-Weatherby, J., Owen, C.D., Lukacik, P., Strain-Damerell, C., Owen, R.L., Walsh, M.A., Winter, G., 2022. *xia* 2. *multiplex*: a multi-crystal data-analysis pipeline. Acta Crystallogr. Sect. Struct. Biol. 78, 752–769. https://doi.org/10.1107/S2059798322004399

Grosse-Kunstleve, R.W., Adams, P.D., 2001. Patterson correlation methods: a review of molecular replacement with *CNS*. Acta Crystallogr. D Biol. Crystallogr. 57, 1390–1396. https://doi.org/10.1107/S090744490101246X

Gu, Y., Li, H., Dong, H., Zeng, Y., Zhang, Z., Paterson, N.G., Stansfeld, P.J., Wang, Z., Zhang, Y., Wang, W., Dong, C., 2016. Structural basis of outer membrane protein insertion by the BAM complex. Nature 531, 64–69. https://doi.org/10.1038/nature17199

Guerrero, R., Pedrós-Alió, C., Esteve, I., Mas, J., Chase, D., Margulis, L., 1986. Predatory prokaryotes: Predation and primary consumption evolved in bacteria. Proc. Natl. Acad. Sci. 83, 2138–2142. https://doi.org/10.1073/pnas.83.7.2138

Guerrini, F., Romano, V., Valenzi, M., Di Giulio, M., Mupo, M.R., Sacco, M., 1982. Molecular parasitism in the Escherichia coli-Bdellovibrio bacteriovorus system: translocation of the matrix protein from the host to the parasite outer membrane. EMBO J. 1, 1439–1444. https://doi.org/10.1002/j.1460-2075.1982.tb01335.x

Hahn, M.W., Schmidt, J., Koll, U., Rohde, M., Verbarg, S., Pitt, A., Nakai, R., Naganuma, T., Lang, E., 2017. Silvanigrella aquatica gen. nov., sp. nov., isolated from a freshwater lake, description of Silvanigrellaceae fam. nov. and Silvanigrellales ord. nov., reclassification of the order Bdellovibrionales in the class Oligoflexia, reclassification of the families Bacteriovoracaceae and Halobacteriovoraceae in the new order Bacteriovoracales ord. nov., and reclassification of the family Pseudobacteriovoracaceae in the order Oligoflexales. Int. J. Syst. Evol. Microbiol. 67, 2555–2568. https://doi.org/10.1099/ijsem.0.001965

Hanahan, D., 1983. Studies on transformation of Escherichia coli with plasmids. J. Mol. Biol. 166, 557–580. https://doi.org/10.1016/S0022-2836(83)80284-8

Hartl, F.-U., Lecker, S., Schiebel, E., Hendrick, J.P., Wickner, W., 1990. The binding cascade of SecB to SecA to SecYE mediates preprotein targeting to the E. coli plasma membrane. Cell 63, 269–279. https://doi.org/10.1016/0092-8674(90)90160-G

Hearn, E.M., Patel, D.R., Lepore, B.W., Indic, M., van den Berg, B., 2009. Transmembrane passage of hydrophobic compounds through a protein channel wall. Nature 458, 367–370. https://doi.org/10.1038/nature07678

Hennecke, G., Nolte, J., Volkmer-Engert, R., Schneider-Mergener, J., Behrens, S., 2005. The Periplasmic Chaperone SurA Exploits Two Features Characteristic of Integral Outer

Membrane Proteins for Selective Substrate Recognition. J. Biol. Chem. 280, 23540–23548. https://doi.org/10.1074/jbc.M413742200

HM Government, 2019. The UK's 20-Year Vision for Antimicrobial Resistance. Open Government Licence, UK.

Hobley, L., Summers, J.K., Till, R., Milner, D.S., Atterbury, R.J., Stroud, A., Capeness, M.J., Gray, S., Leidenroth, A., Lambert, C., Connerton, I., Twycross, J., Baker, M., Tyson, J., Kreft, J.-U., Sockett, R.E., 2020. Dual Predation by Bacteriophage and *Bdellovibrio bacteriovorus* Can Eradicate *Escherichia coli* Prey in Situations where Single Predation Cannot. J. Bacteriol. 202. https://doi.org/10.1128/JB.00629-19

Holdbrook, D.A., Burmann, B.M., Huber, R.G., Petoukhov, M.V., Svergun, D.I., Hiller, S., Bond, P.J., 2017. A Spring-Loaded Mechanism Governs the Clamp-like Dynamics of the Skp Chaperone. Structure 25, 1079-1088.e3. https://doi.org/10.1016/j.str.2017.05.018

Holm, L., 2020. Using Dali for Protein Structure Comparison, in: Gáspári, Z. (Ed.), Structural Bioinformatics, Methods in Molecular Biology. Springer US, New York, NY, pp. 29–42. https://doi.org/10.1007/978-1-0716-0270-6_3

Holmes, A.H., Moore, L.S.P., Sundsfjord, A., Steinbakk, M., Regmi, S., Karkey, A., Guerin, P.J., Piddock, L.J.V., 2016. Understanding the mechanisms and drivers of antimicrobial resistance. The Lancet 387, 176–187. https://doi.org/10.1016/S0140-6736(15)00473-0

Horne, J.E., Radford, S.E., 2022. Roll out the barrel! Outer membrane tension resolves an unexpected folding intermediate. Cell 185, 1107–1109. https://doi.org/10.1016/j.cell.2022.03.001

Huber, D., Jamshad, M., Hanmer, R., Schibich, D., Döring, K., Marcomini, I., Kramer, G., Bukau, B., 2017. SecA Cotranslationally Interacts with Nascent Substrate Proteins *In Vivo*. J. Bacteriol. 199. https://doi.org/10.1128/JB.00622-16

Huber, D., Rajagopalan, N., Preissler, S., Rocco, M.A., Merz, F., Kramer, G., Bukau, B., 2011. SecA Interacts with Ribosomes in Order to Facilitate Posttranslational Translocation in Bacteria. Mol. Cell 41, 343–353. https://doi.org/10.1016/j.molcel.2010.12.028

Hutter, C.A.J., Lehner, R., Wirth, Ch., Condemine, G., Peneff, C., Schirmer, T., 2014. Structure of the oligogalacturonate-specific KdgM porin. Acta Crystallogr. D Biol. Crystallogr. 70, 1770–1778. https://doi.org/10.1107/S1399004714007147

Isturiz, R., Sings, H.L., Hilton, B., Arguedas, A., Reinert, R.-R., Jodar, L., 2017. Streptococcus pneumoniae serotype 19A: worldwide epidemiology. Expert Rev. Vaccines 16, 1007–1027. https://doi.org/10.1080/14760584.2017.1362339

Jamali, K., Kimanius, D., Scheres, S.H.W., 2022. A graph neural network approach to automated model building in cryo-EM maps. ArXiv. https://doi.org/10.48550/arXiv.2210.00006

Jault, P., Leclerc, T., Jennes, S., Pirnay, J.P., Que, Y.-A., Resch, G., Rousseau, A.F., Ravat, F., Carsin, H., Le Floch, R., Schaal, J.V., Soler, C., Fevre, C., Arnaud, I., Bretaudeau, L., Gabard, J., 2019. Efficacy and tolerability of a cocktail of bacteriophages to treat burn wounds infected by Pseudomonas aeruginosa (PhagoBurn): a randomised, controlled, double-blind phase 1/2 trial. Lancet Infect. Dis. 19, 35–45. https://doi.org/10.1016/S1473-3099(18)30482-1

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. https://doi.org/10.1038/s41586-021-03819-2

Kadouri, D., Venzon, N.C., O'Toole, G.A., 2007. Vulnerability of Pathogenic Biofilms to *Micavibrio aeruginosavorus*. Appl. Environ. Microbiol. 73, 605–614. https://doi.org/10.1128/AEM.01893-06

Kadouri, D.E., To, K., Shanks, R.M.Q., Doi, Y., 2013. Predatory Bacteria: A Potential Ally against Multidrug-Resistant Gram-Negative Pathogens. PLoS ONE 8, e63397. https://doi.org/10.1371/journal.pone.0063397

Kaplan, M., Chang, Y.-W., Oikonomou, C.M., Nicolas, W.J., Jewett, A.I., Kreida, S., Dutka, P., Rettberg, L.A., Maggi, S., Jenson, G.J., 2022. Dynamic structural adaptations enable the endobiotic predation of bdellovibrio bacteriovorus. bioRxiv. https://doi.org/10.1101/2022.06.13.496000

Killian, J.A., von Heijne, G., 2000. How proteins adapt to a membrane–water interface. Trends Biochem. Sci. 25, 429–434. https://doi.org/10.1016/S0968-0004(00)01626-1

Kim, S., Malinverni, J.C., Sliz, P., Silhavy, T.J., Harrison, S.C., Kahne, D., 2007. Structure and Function of an Essential Component of the Outer Membrane Protein Assembly Machine. Science 317, 961–964. https://doi.org/10.1126/science.1143993

Kimanius, D., Dong, L., Sharov, G., Nakane, T., Scheres, S.H.W., 2021. New tools for automated cryo-EM single-particle analysis in RELION-4.0. Biochem. J. 478, 4169–4185. https://doi.org/10.1042/BCJ20210708

Korndörfer, I.P., Dommel, M.K., Skerra, A., 2004. Structure of the periplasmic chaperone Skp suggests functional similarity with cytosolic chaperones despite differing architecture. Nat. Struct. Mol. Biol. 11, 1015–1020. https://doi.org/10.1038/nsmb828

Kotra, L.P., Golemi, D., Amro, N.A., Liu, G.-Y., Mobashery, S., 1999. Dynamics of the Lipopolysaccharide Assembly on the Surface of *Escherichia c oli*. J. Am. Chem. Soc. 121, 8707–8711. https://doi.org/10.1021/ja991374z

Koval, S.F., Hynes, S.H., Flannagan, R.S., Pasternak, Z., Davidov, Y., Jurkevitch, E., 2013. Bdellovibrio exovorus sp. nov., a novel predator of Caulobacter crescentus. Int. J. Syst. Evol. Microbiol. 63, 146–151. https://doi.org/10.1099/ijs.0.039701-0

Koval, S.F., Williams, H.N., Stine, O.C., 2015. Reclassification of Bacteriovorax marinus as Halobacteriovorax marinus gen. nov., comb. nov. and Bacteriovorax litoralis as Halobacteriovorax litoralis comb. nov.; description of Halobacteriovoraceae fam. nov. in the class Deltaproteobacteria. Int. J. Syst. Evol. Microbiol. 65, 593–597. https://doi.org/10.1099/ijs.0.070201-0

Kumar, M., Sarma, D.K., Shubham, S., Kumawat, M., Verma, V., Nina, P.B., Jp, D., Kumar, S., Singh, B., Tiwari, R.R., 2021. Futuristic Non-antibiotic Therapies to Combat Antibiotic Resistance: A Review. Front. Microbiol. 12, 609459. https://doi.org/10.3389/fmicb.2021.609459

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. https://doi.org/10.1038/s41564-017-0029-y

Kyte, J., Doolittle, R.F., 1982. A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157, 105–132. https://doi.org/10.1016/0022-2836(82)90515-0

Lambert, C., Chang, C.-Y., Capeness, M.J., Sockett, R.E., 2010a. The First Bite— Profiling the Predatosome in the Bacterial Pathogen Bdellovibrio. PLoS ONE 5, e8599. https://doi.org/10.1371/journal.pone.0008599

Lambert, C., Evans, K.J., Till, R., Hobley, L., Capeness, M., Rendulic, S., Schuster, S.C., Aizawa, S.-I., Sockett, R.E., 2006. Characterizing the flagellar filament and the role of motility in bacterial prey-penetration by Bdellovibrio bacteriovorus. Mol. Microbiol. 60, 274–286. https://doi.org/10.1111/j.1365-2958.2006.05081.x

Lambert, C., Fenton, A.K., Hobley, L., Sockett, R.E., 2011. Predatory Bdellovibrio Bacteria Use Gliding Motility To Scout for Prey on Surfaces. J. Bacteriol. 193, 3139–3141. https://doi.org/10.1128/JB.00224-11

Lambert, C., Hobley, L., Chang, C.-Y., Fenton, A., Capeness, M., Sockett, L., 2008. A Predatory Patchwork: Membrane and Surface Structures of Bdellovibrio bacteriovorus, in: Advances in Microbial Physiology. Elsevier, pp. 313–361. https://doi.org/10.1016/S0065-2911(08)00005-2

Lambert, C., Ivanov, P., Sockett, R.E., 2010b. A Transcriptional "Scream" Early Response of E. coli Prey to Predatory Invasion by Bdellovibrio. Curr. Microbiol. 60, 419–427. https://doi.org/10.1007/s00284-009-9559-8

Lambert, C., Lerner, T.R., Bui, N.K., Somers, H., Aizawa, S.-I., Liddell, S., Clark, A., Vollmer, W., Lovering, A.L., Sockett, R.E., 2016. Interrupting peptidoglycan deacetylation during Bdellovibrio predator-prey interaction prevents ultimate destruction of prey wall, liberating bacterial-ghosts. Sci. Rep. 6, 26010. https://doi.org/10.1038/srep26010

Lambert, C., Smith, M.C.M., Sockett, R.E., 2003. A novel assay to monitor predator-prey interactions for Bdellovibrio bacteriovorus 109 J reveals a role for methyl-accepting chemotaxis proteins in predation. Environ. Microbiol. 5, 127–132. https://doi.org/10.1046/j.1462-2920.2003.00385.x

Lecker, S.H., Driessen, A.J., Wickner, W., 1990. ProOmpA contains secondary and tertiary structure prior to translocation and is shielded from aggregation by association with SecB protein. EMBO J. 9, 2309–2314. https://doi.org/10.1002/j.1460-2075.1990.tb07402.x

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 Pathog. 8, e1002524. https://doi.org/10.1371/journal.ppat.1002524

Li, L., Park, E., Ling, J., Ingram, J., Ploegh, H., Rapoport, T.A., 2016. Crystal structure of a substrate-engaged SecY protein-translocation channel. Nature 531, 395–399. https://doi.org/10.1038/nature17163

Liko, I., Degiacomi, M.T., Lee, S., Newport, T.D., Gault, J., Reading, E., Hopper, J.T.S., Housden, N.G., White, P., Colledge, M., Sula, A., Wallace, B.A., Kleanthous, C., Stansfeld, P.J., Bayley, H., Benesch, J.L.P., Allison, T.M., Robinson, C.V., 2018. Lipid binding attenuates channel closure of the outer membrane protein OmpF. Proc. Natl. Acad. Sci. 115, 6691–6696. https://doi.org/10.1073/pnas.1721152115

Lomize, A.L., Pogozheva, I.D., Mosberg, H.I., 2011. Anisotropic Solvent Model of the Lipid Bilayer. 2. Energetics of Insertion of Small Molecules, Peptides, and Proteins in Membranes. J. Chem. Inf. Model. 51, 930–946. https://doi.org/10.1021/ci200020k

Mahmoud, K.K., Koval, S.F., 2010. Characterization of type IV pili in the life cycle of the predator bacterium Bdellovibrio. Microbiology 156, 1040–1051. https://doi.org/10.1099/mic.0.036137-0

Matilla, M.A., Krell, T., 2018. The effect of bacterial chemotaxis on host infection and pathogenicity. FEMS Microbiol. Rev. 42. https://doi.org/10.1093/femsre/fux052

Matthews, B.W., 1968. Solvent content of protein crystals. J. Mol. Biol. 33, 491–497. https://doi.org/10.1016/0022-2836(68)90205-2

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., Read, R.J., 2007. *Phaser* crystallographic software. J. Appl. Crystallogr. 40, 658–674. https://doi.org/10.1107/S0021889807021206

Medina, A.A., Shanks, R.M., Kadouri, D.E., 2008. Development of a novel system for isolating genes involved in predator-prey interactions using host independent derivatives of Bdellovibrio bacteriovorus 109J. BMC Microbiol. 8, 33. https://doi.org/10.1186/1471-2180-8-33

Meek, R.W., Cadby, I.T., Moynihan, P.J., Lovering, A.L., 2019. Structural basis for activation of a diguanylate cyclase required for bacterial predation in Bdellovibrio. Nat. Commun. 10, 4086. https://doi.org/10.1038/s41467-019-12051-6

Merritt, E., 2012. X-ray Anomalous Scattering.

Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., Steinegger, M., 2022. ColabFold: making protein folding accessible to all. Nat. Methods 19, 679–682. https://doi.org/10.1038/s41592-022-01488-1

Miroux, B., Walker, J.E., 1996. Over-production of Proteins inEscherichia coli: Mutant Hosts that Allow Synthesis of some Membrane Proteins and Globular Proteins at High Levels. J. Mol. Biol. 260, 289–298. https://doi.org/10.1006/jmbi.1996.0399

Modi, N., Ganguly, S., Bárcena-Uribarri, I., Benz, R., van den Berg, B., Kleinekathöfer, U., 2015. Structure, Dynamics, and Substrate Specificity of the OprO Porin from Pseudomonas aeruginosa. Biophys. J. 109, 1429–1438. https://doi.org/10.1016/j.bpj.2015.07.035

Moreira, D., Zivanovic, Y., López-Archilla, A.I., Iniesto, M., López-García, P., 2021. Reductive evolution and unique predatory mode in the CPR bacterium Vampirococcus lugosii. Nat. Commun. 12, 2454. https://doi.org/10.1038/s41467-021-22762-4

Mori, H., Ito, K., 2006. Different modes of SecY–SecA interactions revealed by site-directed *in vivo* photo-cross-linking. Proc. Natl. Acad. Sci. 103, 16159–16164. https://doi.org/10.1073/pnas.0606390103

Mu, D.-S., Wang, S., Liang, Q.-Y., Du, Z.-Z., Tian, R., Ouyang, Y., Wang, X.-P., Zhou, A., Gong, Y., Chen, G.-J., Van Nostrand, J., Yang, Y., Zhou, J., Du, Z.-J., 2020. Bradymonabacteria, a novel bacterial predator group with versatile survival strategies in saline environments. Microbiome 8, 126. https://doi.org/10.1186/s40168-020-00902-0

Munita, J.M., Arias, C.A., 2016. Mechanisms of Antibiotic Resistance. Microbiol. Spectr. 4, 4.2.15. https://doi.org/10.1128/microbiolspec.VMBF-0016-2015

Muñoz-Dorado, J., Marcos-Torres, F.J., García-Bravo, E., Moraleda-Muñoz, A., Pérez, J., 2016. Myxobacteria: Moving, Killing, Feeding, and Surviving Together. Front. Microbiol. 7. https://doi.org/10.3389/fmicb.2016.00781

Negus, D., Moore, C., Baker, M., Raghunathan, D., Tyson, J., Sockett, R.E., 2017. Predator Versus Pathogen: How Does Predatory *Bdellovibrio bacteriovorus* Interface with the Challenges of Killing Gram-Negative Pathogens in a Host Setting? Annu. Rev. Microbiol. 71, 441–457. https://doi.org/10.1146/annurev-micro-090816-093618

Ni, D., Wang, Y., Yang, X., Zhou, H., Hou, X., Cao, B., Lu, Z., Zhao, X., Yang, K., Huang, Y., 2014. Structural and functional analysis of the β-barrel domain of BamA from *Escherichia coli*. FASEB J. 28, 2677–2685. https://doi.org/10.1096/fj.13-248450

Nikaido, H., 2003. Molecular Basis of Bacterial Outer Membrane Permeability Revisited. Microbiol. Mol. Biol. Rev. 67, 593–656. https://doi.org/10.1128/MMBR.67.4.593-656.2003

Noinaj, N., Kuszak, A.J., Buchanan, S.K., 2015. Heat Modifiability of Outer Membrane Proteins from Gram-Negative Bacteria, in: Buchanan, S., Noinaj, N. (Eds.), The BAM Complex, Methods in Molecular Biology. Springer New York, New York, NY, pp. 51–56. https://doi.org/10.1007/978-1-4939-2871-2_4

O'Neill, J., 2014. Antimicrobial Resistance: Tackling a crisis for the health and wealth of nations, Review on Antimirobial Resistance. Wellcome Trust & HM Government, London, UK.

Paetzel, M., Karla, A., Strynadka, N.C.J., Dalbey, R.E., 2002. Signal Peptidases. Chem. Rev. 102, 4549–4580. https://doi.org/10.1021/cr010166y

Pan, X., Kage, H., Martin, K., Nett, M., 2017. Herpetosiphon gulosus sp. nov., a filamentous predatory bacterium isolated from sandy soil and Herpetosiphon giganteus sp. nov., nom. rev. Int. J. Syst. Evol. Microbiol. 67, 2476–2481. https://doi.org/10.1099/ijsem.0.002141

Parales, R.E., Luu, R.A., Hughes, J.G., Ditty, J.L., 2015. Bacterial chemotaxis to xenobiotic chemicals and naturally-occurring analogs. Curr. Opin. Biotechnol. 33, 318–326. https://doi.org/10.1016/j.copbio.2015.03.017

Patel, G.J., Behrens-Kneip, S., Holst, O., Kleinschmidt, J.H., 2009. The Periplasmic Chaperone Skp Facilitates Targeting, Insertion, and Folding of OmpA into Lipid Membranes with a Negative Membrane Surface Potential. Biochemistry 48, 10235–10245. https://doi.org/10.1021/bi901403c

Pathania, M., Acosta-Gutierrez, S., Bhamidimarri, S.P., Baslé, A., Winterhalter, M., Ceccarelli, M., van den Berg, B., 2018. Unusual Constriction Zones in the Major Porins OmpU and

OmpT from Vibrio cholerae. Structure 26, 708-721.e4. https://doi.org/10.1016/j.str.2018.03.010

Peltola, H., 2000. Worldwide *Haemophilus influenzae* Type b Disease at the Beginning of the 21st Century: Global Analysis of the Disease Burden 25 Years after the Use of the Polysaccharide Vaccine and a Decade after the Advent of Conjugates. Clin. Microbiol. Rev. 13, 302–317. https://doi.org/10.1128/CMR.13.2.302

Pérez, J., Contreras-Moreno, F.J., Marcos-Torres, F.J., Moraleda-Muñoz, A., Muñoz-Dorado, J., 2020. The antibiotic crisis: How bacterial predators can help. Comput. Struct. Biotechnol. J. 18, 2547–2555. https://doi.org/10.1016/j.csbj.2020.09.010

Pérez, J., Moraleda-Muñoz, A., Marcos-Torres, F.J., Muñoz-Dorado, J., 2016. Bacterial predation: 75 years and counting!: Bacterial predation. Environ. Microbiol. 18, 766–779. https://doi.org/10.1111/1462-2920.13171

Phale, P.S., Schirmer, T., Prilipov, A., Lou, K.-L., Hardmeyer, A., Rosenbusch, J.P., 1997. Voltage gating of *Escherichia coli* porin channels: Role of the constriction loop. Proc. Natl. Acad. Sci. 94, 6741–6745. https://doi.org/10.1073/pnas.94.13.6741

Punjani, A., Rubinstein, J.L., Fleet, D.J., Brubaker, M.A., 2017. cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. Nat. Methods 14, 290–296. https://doi.org/10.1038/nmeth.4169

Raghunathan, D., Radford, P.M., Gell, C., Negus, D., Moore, C., Till, R., Tighe, P.J., Wheatley, S.P., Martinez-Pomares, L., Sockett, R.E., Tyson, J., 2019. Engulfment, persistence and fate of Bdellovibrio bacteriovorus predators inside human phagocytic cells informs their future therapeutic potential. Sci. Rep. 9, 4293. https://doi.org/10.1038/s41598-019-40223-3

Read, R.J., 2001. Pushing the boundaries of molecular replacement with maximum likelihood. Acta Crystallogr. D Biol. Crystallogr. 57, 1373–1382. https://doi.org/10.1107/S0907444901012471

Rendulic, S., Jagtap, P., Rosinus, A., Eppinger, M., Baar, C., Lanz, C., Keller, H., Lambert, C., Evans, K.J., Goesmann, A., Meyer, F., Sockett, R.E., Schuster, S.C., 2004. A Predator Unmasked: Life Cycle of *Bdellovibrio bacteriovorus* from a Genomic Perspective. Science 303, 689–692. https://doi.org/10.1126/science.1093027

Rittenberg, S.C., Shilo, M., 1970. Early Host Damage in the Infection Cycle of *Bdellovibrio bacteriovorus*. J. Bacteriol. 102, 149–160. https://doi.org/10.1128/jb.102.1.149-160.1970

Rogers, M., Ekaterinaki, N., Nimmo, E., Sherratt, D., 1986. Analysis of Tn7 transposition. Mol. Gen. Genet. MGG 205, 550–556. https://doi.org/10.1007/BF00338097

Rohou, A., Grigorieff, N., 2015. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. J. Struct. Biol. 192, 216–221. https://doi.org/10.1016/j.jsb.2015.08.008

Romanowski, E.G., Stella, N.A., Brothers, K.M., Yates, K.A., Funderburgh, M.L., Funderburgh, J.L., Gupta, S., Dharani, S., Kadouri, D.E., Shanks, R.M.Q., 2016. Predatory bacteria are nontoxic to the rabbit ocular surface. Sci. Rep. 6, 30987. https://doi.org/10.1038/srep30987

Russo, R., Kolesnikova, I., Kim, T., Gupta, S., Pericleous, A., Kadouri, D., Connell, N., 2018. Susceptibility of Virulent Yersinia pestis Bacteria to Predator Bacteria in the Lungs of Mice. Microorganisms 7, 2. https://doi.org/10.3390/microorganisms7010002

Sampedro, I., Parales, R.E., Krell, T., Hill, J.E., 2014. *Pseudomonas* chemotaxis. FEMS Microbiol. Rev. n/a-n/a. https://doi.org/10.1111/1574-6976.12081

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, 874. https://doi.org/10.1038/s42003-021-02399-1

Schäfer, A., Tauch, A., Jäger, W., Kalinowski, J., Thierbach, G., Pühler, A., 1994. Small mobilizable multi-purpose cloning vectors derived from the Escherichia coli plasmids pK18 and pK19: selection of defined deletions in the chromosome of Corynebacterium glutamicum. Gene 145, 69–73. https://doi.org/10.1016/0378-1119(94)90324-7

Schäfer, U., Beck, K., Müller, M., 1999. Skp, a Molecular Chaperone of Gram-negative Bacteria, Is Required for the Formation of Soluble Periplasmic Intermediates of Outer Membrane Proteins. J. Biol. Chem. 274, 24567–24574. https://doi.org/10.1074/jbc.274.35.24567

Schirmer, T., Keller, T.A., Wang, Y.-F., Rosenbusch, J.P., 1995. Structural Basis for Sugar Translocation Through Maltoporin Channels at 3.1 Å Resolution. Science 267, 512–514. https://doi.org/10.1126/science.7824948

Schooley, R.T., Biswas, B., Gill, J.J., Hernandez-Morales, A., Lancaster, J., Lessor, L., Barr, J.J., Reed, S.L., Rohwer, F., Benler, S., Segall, A.M., Taplitz, R., Smith, D.M., Kerr, K., Kumaraswamy, M., Nizet, V., Lin, L., McCauley, M.D., Strathdee, S.A., Benson, C.A., Pope, R.K., Leroux, B.M., Picel, A.C., Mateczun, A.J., Cilwa, K.E., Regeimbal, J.M., Estrella, L.A., Wolfe, D.M., Henry, M.S., Quinones, J., Salka, S., Bishop-Lilly, K.A., Young, R., Hamilton, T., 2017. Development and Use of Personalized Bacteriophage-Based Therapeutic Cocktails To Treat a Patient with a Disseminated Resistant Acinetobacter baumannii Infection. Antimicrob. Agents Chemother. 61, e00954-17. https://doi.org/10.1128/AAC.00954-17

Schulz, G.E., 2002. The structure of bacterial outer membrane proteins. Biochim. Biophys. Acta BBA - Biomembr. 1565, 308–317. https://doi.org/10.1016/S0005-2736(02)00577-1

Schwudke, D., Linscheid, M., Strauch, E., Appel, B., Zähringer, U., Moll, H., Müller, M., Brecker, L., Gronow, S., Lindner, B., 2003. The Obligate Predatory Bdellovibrio bacteriovorus Possesses a Neutral Lipid A Containing α-D-Mannoses That Replace Phosphate Residues. J. Biol. Chem. 278, 27502–27512. https://doi.org/10.1074/jbc.M303012200

Seccareccia, I., Kost, C., Nett, M., 2015. Quantitative Analysis of Lysobacter Predation. Appl. Environ. Microbiol. 81, 7098–7105. https://doi.org/10.1128/AEM.01781-15

Segrest, J.P., De Loof, H., Dohlman, J.G., Brouillette, C.G., Anantharamaiah, G.M., 1990. Amphipathic helix motif: Classes and properties. Proteins Struct. Funct. Genet. 8, 103–117. https://doi.org/10.1002/prot.340080202

Shatzkes, K., Chae, R., Tang, C., Ramirez, G.C., Mukherjee, S., Tsenova, L., Connell, N.D., Kadouri, D.E., 2015. Examining the safety of respiratory and intravenous inoculation

of Bdellovibrio bacteriovorus and Micavibrio aeruginosavorus in a mouse model. Sci. Rep. 5, 12899. https://doi.org/10.1038/srep12899

Shatzkes, K., Singleton, E., Tang, C., Zuena, M., Shukla, S., Gupta, S., Dharani, S., Onyile, O., Rinaggio, J., Connell, N.D., Kadouri, D.E., 2016. Predatory Bacteria Attenuate Klebsiella pneumoniae Burden in Rat Lungs. mBio 7, e01847-16. https://doi.org/10.1128/mBio.01847-16

Shatzkes, K., Singleton, E., Tang, C., Zuena, M., Shukla, S., Gupta, S., Dharani, S., Rinaggio, J., Kadouri, D.E., Connell, N.D., 2017a. Examining the efficacy of intravenous administration of predatory bacteria in rats. Sci. Rep. 7, 1864. https://doi.org/10.1038/s41598-017-02041-3

Shatzkes, K., Tang, C., Singleton, E., Shukla, S., Zuena, M., Gupta, S., Dharani, S., Rinaggio, J., Connell, N.D., Kadouri, D.E., 2017b. Effect of predatory bacteria on the gut bacterial microbiota in rats. Sci. Rep. 7, 43483. https://doi.org/10.1038/srep43483

Shiratori, T., Suzuki, S., Kakizawa, Y., Ishida, K., 2019. Phagocytosis-like cell engulfment by a planctomycete bacterium. Nat. Commun. 10, 5529. https://doi.org/10.1038/s41467-019-13499-2

Sklar, J.G., Wu, T., Kahne, D., Silhavy, T.J., 2007. Defining the roles of the periplasmic chaperones SurA, Skp, and DegP in *Escherichia coli*. Genes Dev. 21, 2473–2484. https://doi.org/10.1101/gad.1581007

Skolnick, J., Gao, M., Zhou, H., Singh, S., 2021. AlphaFold 2: Why It Works and Its Implications for Understanding the Relationships of Protein Sequence, Structure, and Function. J. Chem. Inf. Model. 61, 4827–4831. https://doi.org/10.1021/acs.jcim.1c01114

Skubák, P., 2022. CCP4 Cloud 1.8 Documentation, Section 9.1 Crank-2.

Skubák, P., Pannu, N.S., 2013. Automatic protein structure solution from weak X-ray data. Nat. Commun. 4, 2777. https://doi.org/10.1038/ncomms3777

Snellen, J.E., Starr, M.P., 1974. Ultrastructural aspects of localized membrane damage in Spirillum serpens VHL early in its association with Bdellovibrio bacteriovorus 109D. Arch. Microbiol. 100, 179–195. https://doi.org/10.1007/BF00446316

Sockett, R.E., 2009. Predatory Lifestyle of *Bdellovibrio bacteriovorus*. Annu. Rev. Microbiol. 63, 523–539. https://doi.org/10.1146/annurev.micro.091208.073346

Starr, M.P., Baigent, N.L., 1966. Parasitic Interaction of *Bdellovibrio bacteriovorus* with Other Bacteria. J. Bacteriol. 91, 2006–2017. https://doi.org/10.1128/jb.91.5.2006-2017.1966

Stolp, H., Petzold, H., 1962. Untersuchungen über einen obligat parasitischen Mikroorganismus mit lytischer Aktivität für Pseudomonas-Bakterien. J. Phytopathol. 45, 364–390. https://doi.org/10.1111/j.1439-0434.1962.tb02050.x

Stolp, H., Starr, M.P., 1965. Bacteriolysis. Annu. Rev. Microbiol. 19, 79–104. https://doi.org/10.1146/annurev.mi.19.100165.000455

Stolp, H., Starr, M.P., 1963. Bdellovibrio bacteriovorus gen. et sp. n., a predatory, ectoparasitic, and bacteriolytic microorganism. Antonie Van Leeuwenhoek 29, 217–248. https://doi.org/10.1007/BF02046064

Straley, S.C., Conti, S.F., 1977. Chemotaxis by Bdellovibrio bacteriovorus toward prey. J. Bacteriol. 132, 628–640. https://doi.org/10.1128/jb.132.2.628-640.1977

Subbarao, G.V., van den Berg, B., 2006. Crystal Structure of the Monomeric Porin OmpG. J. Mol. Biol. 360, 750–759. https://doi.org/10.1016/j.jmb.2006.05.045

Suo, Y., Hardy, S.J.S., Randall, L.L., 2015. The Basis of Asymmetry in the SecA:SecB Complex. J. Mol. Biol. 427, 887–900. https://doi.org/10.1016/j.jmb.2014.12.008

Talley, B.G., McDade, R.L., Diedrich, D.L., 1987. Verification of the protein in the outer membrane of Bdellovibrio bacteriovorus as the OmpF protein of its Escherichia coli prey. J. Bacteriol. 169, 694–698. https://doi.org/10.1128/jb.169.2.694-698.1987

Tanabe, M., Nimigean, C.M., Iverson, T.M., 2010. Structural basis for solute transport, nucleotide regulation, and immunological recognition of *Neisseria meningitidis* PorB. Proc. Natl. Acad. Sci. 107, 6811–6816. https://doi.org/10.1073/pnas.0912115107

Taylor, G., 2003. The phase problem. Acta Crystallogr. D Biol. Crystallogr. 59, 1881–1890. https://doi.org/10.1107/S0907444903017815

Taylor, G.L., 2010. Introduction to phasing. Acta Crystallogr. D Biol. Crystallogr. 66, 325–338. https://doi.org/10.1107/S0907444910006694

Terwilliger, T.C., 2013. Finding non-crystallographic symmetry in density maps of macromolecular structures. J. Struct. Funct. Genomics 14, 91–95. https://doi.org/10.1007/s10969-013-9157-7

Terwilliger, T.C., Grosse-Kunstleve, R.W., Afonine, P.V., Moriarty, N.W., Zwart, P.H., Hung, L.-W., Read, R.J., Adams, P.D., 2008. Iterative model building, structure refinement and density modification with the *PHENIX AutoBuild* wizard. Acta Crystallogr. D Biol. Crystallogr. 64, 61–69. https://doi.org/10.1107/S090744490705024X

Thomashow, M.F., Rittenberg, S.C., 1978. Intraperiplasmic growth of Bdellovibrio bacteriovorus 109J: solubilization of Escherichia coli peptidoglycan. J. Bacteriol. 135, 998–1007. https://doi.org/10.1128/jb.135.3.998-1007.1978

Tormo, A., Almirón, M., Kolter, R., 1990. surA, an Escherichia coli gene essential for survival in stationary phase. J. Bacteriol. 172, 4339–4347. https://doi.org/10.1128/jb.172.8.4339-4347.1990

Tudor, J.J., Karp, M.A., 1994. Translocation of an outer membrane protein into prey cytoplasmic membranes by bdellovibrios. J. Bacteriol. 176, 948–952. https://doi.org/10.1128/jb.176.4.948-952.1994

Vagin, A., Teplyakov, A., 2010. Molecular replacement with *MOLREP*. Acta Crystallogr. D Biol. Crystallogr. 66, 22–25. https://doi.org/10.1107/S0907444909042589

Vagin, A., Teplyakov, A., 1997. *MOLREP*: an Automated Program for Molecular Replacement. J. Appl. Crystallogr. 30, 1022–1025. https://doi.org/10.1107/S0021889897006766

van den Berg, B., Black, P.N., Clemons, W.M., Rapoport, T.A., 2004. Crystal Structure of the Long-Chain Fatty Acid Transporter FadL. Science 304, 1506–1509. https://doi.org/10.1126/science.1097524

van den Berg, B., Prathyusha Bhamidimarri, S., Dahyabhai Prajapati, J., Kleinekathöfer, U., Winterhalter, M., 2015. Outer-membrane translocation of bulky small molecules by passive diffusion. Proc. Natl. Acad. Sci. 112. https://doi.org/10.1073/pnas.1424835112

van den Ent, F., Löwe, J., 2006. RF cloning: A restriction-free method for inserting target genes into plasmids. J. Biochem. Biophys. Methods 67, 67–74. https://doi.org/10.1016/j.jbbm.2005.12.008

van Kempen, M., Kim, S.S., Tumescheit, C., Mirdita, M., Gilchrist, C.L.M., Söding, J., Steinegger, M., 2022. Foldseek: fast and accurate protein structure search (preprint). Bioinformatics. https://doi.org/10.1101/2022.02.07.479398

Varadi, M., Anyango, S., Deshpande, M., Nair, S., Natassia, C., Yordanova, G., Yuan, D., Stroe, O., Wood, G., Laydon, A., Žídek, A., Green, T., Tunyasuvunakool, K., Petersen, S., Jumper, J., Clancy, E., Green, R., Vora, A., Lutfi, M., Figurnov, M., Cowie, A., Hobbs, N., Kohli, P., Kleywegt, G., Birney, E., Hassabis, D., Velankar, S., 2022. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. Nucleic Acids Res. 50, D439–D444. https://doi.org/10.1093/nar/gkab1061

Veith, P.D., Glew, M.D., Gorasia, D.G., Chen, D., O'Brien-Simpson, N.M., Reynolds, E.C., 2019. Localization of Outer Membrane Proteins in *Treponema denticola* by Quantitative Proteome Analyses of Outer Membrane Vesicles and Cellular Fractions. J. Proteome Res. 18, 1567–1581. https://doi.org/10.1021/acs.jproteome.8b00860

Vonrhein, C., Flensburg, C., Keller, P., Sharff, A., Smart, O., Paciorek, W., Womack, T., Bricogne, G., 2011. Data processing and analysis with the *autoPROC* toolbox. Acta Crystallogr. D Biol. Crystallogr. 67, 293–302. https://doi.org/10.1107/S0907444911007773

Walton, T.A., Sandoval, C.M., Fowler, C.A., Pardi, A., Sousa, M.C., 2009. The cavity-chaperone Skp protects its substrate from aggregation but allows independent folding of substrate domains. Proc. Natl. Acad. Sci. 106, 1772–1777. https://doi.org/10.1073/pnas.0809275106

Walton, T.A., Sousa, M.C., 2004. Crystal Structure of Skp, a Prefoldin-like Chaperone that Protects Soluble and Membrane Proteins from Aggregation. Mol. Cell 15, 367–374. https://doi.org/10.1016/j.molcel.2004.07.023

Wang, J., Terrasse, R., Bafna, J.A., Benier, L., Winterhalter, M., 2020. Electrophysiological Characterization of Transport Across Outer-Membrane Channels from Gram-Negative Bacteria in Presence of Lipopolysaccharides. Angew. Chem. Int. Ed. 59, 8517–8521. https://doi.org/10.1002/anie.201913618

Waso, M., Reyneke, B., Havenga, B., Khan, S., Khan, W., 2021. Insights into Bdellovibrio spp. mechanisms of action and potential applications. World J. Microbiol. Biotechnol. 37, 85. https://doi.org/10.1007/s11274-021-03054-x

Weichenberger, C.X., Rupp, B., 2014. Ten years of probabilistic estimates of biocrystal solvent content: new insights *via* nonparametric kernel density estimate. Acta Crystallogr. D Biol. Crystallogr. 70, 1579–1588. https://doi.org/10.1107/S1399004714005550

WHO, 2017. Global Priority List of Antibiotic-Resistant Bacteria to Guide Research, Discovery, and Development of New Antibiotics. WHO Press, Geneva, Switzerland.

WHO, 2015. Global Action Plan on Antimicrobial Resistance. WHO Press, Geneva, Switzerland.

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. https://doi.org/10.1016/j.cub.2016.09.067

Winter, G., Beilsten-Edmands, J., Devenish, N., Gerstel, M., Gildea, R.J., McDonagh, D., Pascal, E., Waterman, D.G., Williams, B.H., Evans, G., 2022. DIALS as a toolkit. Protein Sci. 31, 232–250. https://doi.org/10.1002/pro.4224

Wright, A., Hawkins, C.H., Änggård, E.E., Harper, D.R., 2009. A controlled clinical trial of a therapeutic bacteriophage preparation in chronic otitis due to antibiotic-resistant *Pseudomonas aeruginosa* ; a preliminary report of efficacy. Clin. Otolaryngol. 34, 349–357. https://doi.org/10.1111/j.1749-4486.2009.01973.x

Wu, E.L., Fleming, P.J., Yeom, M.S., Widmalm, G., Klauda, J.B., Fleming, K.G., Im, W., 2014. E. coli Outer Membrane and Interactions with OmpLA. Biophys. J. 106, 2493–2502. https://doi.org/10.1016/j.bpj.2014.04.024

Wu, R., Bakelar, J.W., Lundquist, K., Zhang, Z., Kuo, K.M., Ryoo, D., Pang, Y.T., Sun, C., White, T., Klose, T., Jiang, W., Gumbart, J.C., Noinaj, N., 2021. Plasticity within the barrel domain of BamA mediates a hybrid-barrel mechanism by BAM. Nat. Commun. 12, 7131. https://doi.org/10.1038/s41467-021-27449-4

Wu, T., Malinverni, J., Ruiz, N., Kim, S., Silhavy, T.J., Kahne, D., 2005. Identification of a Multicomponent Complex Required for Outer Membrane Biogenesis in Escherichia coli. Cell 121, 235–245. https://doi.org/10.1016/j.cell.2005.02.015

Xu, X., Wang, S., Hu, Y.-X., McKay, D.B., 2007. The Periplasmic Bacterial Molecular Chaperone SurA Adapts its Structure to Bind Peptides in Different Conformations to Assert a Sequence Preference for Aromatic Residues. J. Mol. Biol. 373, 367–381. https://doi.org/10.1016/j.jmb.2007.07.069

Yakimov, M.M., Merkel, A.Y., Gaisin, V.A., Pilhofer, M., Messina, E., Hallsworth, J.E., Klyukina, A.A., Tikhonova, E.N., Gorlenko, V.M., 2022. Cultivation of a vampire: ‘ *Candidatus* Absconditicoccus praedator.’ Environ. Microbiol. 24, 30–49. https://doi.org/10.1111/1462-2920.15823

Ye, J., van den Berg, B., 2004. Crystal structure of the bacterial nucleoside transporter Tsx. EMBO J. 23, 3187–3195. https://doi.org/10.1038/sj.emboj.7600330

Yeoman, C.J., Fields, C.J., Lepercq, P., Ruiz, P., Forano, E., White, B.A., Mosoni, P., 2021. *In Vivo* Competitions between Fibrobacter succinogenes, Ruminococcus flavefaciens, and Ruminoccus albus in a Gnotobiotic Sheep Model Revealed by Multi-Omic Analyses. mBio 12, e03533-20. https://doi.org/10.1128/mBio.03533-20

Yildiz, Ö., Vinothkumar, K.R., Goswami, P., Kühlbrandt, W., 2006. Structure of the monomeric outer-membrane porin OmpG in the open and closed conformation. EMBO J. 25, 3702–3713. https://doi.org/10.1038/sj.emboj.7601237

Zeng, Y., Wang, J., Yang, Chunyan, Ding, M., Hamilton, P.B., Zhang, X., Yang, Caiyun, Zhnag, L., Dai, X., 2021. A Streptomyces globisporus strain kills Microcystis

aeruginosa via cell-to-cell contact. Sci. Total Environ. 769, 144489. https://doi.org/10.1016/j.scitotenv.2020.144489

Zeth, K., Diederichs, K., Welte, W., Engelhardt, H., 2000. Crystal structure of Omp32, the anion-selective porin from Comamonas acidovorans, in complex with a periplasmic peptide at 2.1 Å resolution. Structure 8, 981–992. https://doi.org/10.1016/S0969-2126(00)00189-1

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. https://doi.org/10.1038/nmeth.4193

Zimmer, J., Nam, Y., Rapoport, T.A., 2008. Structure of a complex of the ATPase SecA and the protein-translocation channel. Nature 455, 936–943. https://doi.org/10.1038/nature07335

APPENDIX

1. Primers used in this Study

Table 0-1: Primers used in this Study

Primer Name	Description of Use	Primer Sequence (5'-3')
bd0427_p28_F	PCR amplification of bd0427 for insertion into pET28a	ccctctagaaataattttgtttaacttaagaaggagatataccatgtctaaagcacgtgtagaggctttg
bd0427_p28_R		agccggatctcagtggtggtggtggtggtgctcgagttagaaagtgtaagtgtaagcaacttg
bd0427_p28_NH_F	PCR amplification of bd0427 for insertion into pET28a with a thrombin N-terminal His-tag, used with bd0427_p28_R	catcatcacagcagcggcctggtgccgcgcggctctaaagcacgtgtagaggcttg
bd0427_p28_CH_Q5_F	Insertion of Tev cleavable C-terminal His_8-tag into pET28a-bd0427	ccaccaccaccaccaccaccactaactcgagcaccaccac
bd0427_p28_CH_Q5_R		cctccttggaagtacagattctcgccgaaagtgtaagtgtaagcaacttg
bd0427_p28_A103_His_Q5_F	Insertion of His_6-tag after residue A103 into pET28a-bd0427	ccaccacggctcatcgaatccagctttcgctgatc
bd0427_p28_A103_His_Q5_R		tgatgatgatgggaagaggcatcacgaacagtctg
bd0427_p28_K260_His_Q5_F	Insertion of His_6-tag after residue K260 into pET28a-bd0427	ccaccacggctcatcggacgactctgagtccctg
bd0427_p28_K260_His_Q5_R		tgatgatgatgggaagattttactttagtagaagctacttctac
bd0427_p28_A29C_Q5_F	Substitution, A29C into pET28a-bd0427	gaggctttgtgtaactcacgtcac
bd0427_p28_A29C_Q5_R		aactcacgtcac
bd0427_p28_S287C_Q5_F	Substitution, S287C into pET28a-bd0427	tcttcgtgcatgcgttgctcaaaa
bd0427_p28_S287C_Q5_R		accaaccaagaagccgca
bd0427_p28_Dplug_F	Deletion of N-terminal plug in pET28a-bd0427	gaacaagcgactatcgaatg
bd0427_p28_Dplug_R		catggtatatctccttcttaaag
bd1645_p26_F	PCR amplification of bd1645 for insertion into pET26	cctcgctgcccagccggcgatggcccacacagctgacacttcctgggaag
bd1645_p26_R		ctcagtggtggtggtggtggtgctcgaggtggcaggaagggacagtcaaagc

bd2126_p26_F	PCR amplification of bd2126 for insertion into pET26	cgctgcccagccggcgatggccgttcaccagggcagttcc
bd2126_p26_R		ggtggtggtggtggtgctcgaggtgaggataaacgccttcaca
bd2778_p26_F	PCR amplification of bd2778 for insertion into pET26	cctcgctgcccagccggcgatggccgctgtgaacgccaactcttcctgg
bd2778_p26_R		ctcagtggtggtggtggtggtgctcgaggtccacagtcggagtgtcacactg
bd0427_KO_up_F	PCR amplification of upstream region of *bd0427* in *B. bacteriovorus* HD100, for generation of in-frame deletion strain	taaaacgacggccagtgccaataagtggacaaagtggc
bd0427_KO_up_R		cttagaaagttttcattgttgcatccttttcgttt
bd0427_KO_dw_F	PCR amplification of downstream region of *bd0427* in *B. bacteriovorus* HD100, for generation of in-frame deletion strain	aacaatgaaaactttctaagacctagtcttaaagtttaagaa
bd0427_KO_dw_R		acagctatgacatgattacgtaatttaagttcccccaaac
pk18mobsacB_S_F	pK18mobsacB sequencing primers	gtgctgcaaggcgattaagt
pk18mobsacB_S_F		attaggcaccccaggcttta
bd0427_KO_S_F	Sequencing primers for determination of in-frame deletion in *B. bacteriovorus*	ggatctttcagtctttacgcc
bd0427_KO_S_R		acaccgcttcccagtctaaa

2. Plasmids used in this Study

Table 0-2: Plasmids used in this Study

Plasmid Name	Description	Source
pET28a	Bacterial expression vector with T7 promoter, encoding an N-terminal His_6-tag followed by a thrombin cleavage site, and contains an optional C-terminal His_6-tag. This vector contains a gene encoding Kanamycin resistance.	Novagen
pET26b(+)	Bacterial expression vector with T7 promoter that carries an N-terminal *pelB* signal sequence for potential periplasmic localisation, plus an optional C-terminal His_6-tag. The vector carries a kanamycin resistance gene.	Novagen
pK18mobsacB	Suicide vector used to clone PCR fragments into (KanR, lacZa and sacB)	(Schäfer et al., 1994)

pET28a_*bd0427*	Modified pET28a vector for the expression of *bd0427* (signal peptide removed (1 - 60 bp)). This vector contains a gene encoding Kanamycin resistance.	This study
pET28a_*bd0427*_N His	Modified pET28a vector for the expression of *bd0427* (signal peptide removed (1 - 60 bp)) with a thrombin cleavable N-terminal His_6-tag. This vector contains a gene encoding Kanamycin resistance.	This study
pET28a_*bd0427*_C His	Modified pET28a vector for the expression of *bd0427* (signal peptide removed (1 - 60 bp)) with a Tev cleavable C-terminal His_8-tag. This vector contains a gene encoding Kanamycin resistance.	This study
pET28a_*bd0427*_A 103His	Modified pET28a vector for the expression of *bd0427* (signal peptide removed (1 - 60 bp)) with a SSHHHHHHSS-tag after A103. This vector contains a gene encoding Kanamycin resistance.	This study
pET28a_*bd0427*_K 260His	Modified pET28a vector for the expression of *bd0427* (signal peptide removed (1 - 60 bp)) with a SSHHHHHHSS-tag after K260. This vector contains a gene encoding Kanamycin resistance.	This study
pET28a_*bd0427*_A 29C-S287C	Modified pET28a vector for the expression of *bd0427* (signal peptide removed (1 - 60 bp)) with two cysteine mutations (A29C and S287C). This vector contains a gene encoding Kanamycin resistance.	This study
pET28a_*bd0427*_Δ plug	Modified pET28a vector for the expression of *bd0427* (signal peptide removed (1 - 60 bp)) with N-terminal plug deleted (residues 20 - 31). This vector contains a gene encoding Kanamycin resistance.	This study
pET29b_*bd0427_pelB*_A103His	Modified pET29b (Twist Biosciences) vector for the expression of *bd0427* (signal peptide removed (1 - 60 bp) and replaced with *pelB*) in *E. coli* outer membrane with a SSHHHHHHSS-tag	Twist Biosciences as part of this study

	after A103. This vector contains a gene encoding Kanamycin resistance.	
pET26b_*bd1645*	Modified pET26b vector for the expression of *bd1645* (signal peptide removed (1-117 bp)) with an N-terminal *pelB* signal sequence and a C-terminal His_6-tag. This vector contains a gene encoding Kanamycin resistance.	This study
pET26b_*bd2126*	Modified pET26b vector for the expression of *bd2126* (signal peptide removed (1-57 bp)) with an N-terminal *pelB* signal sequence and a C-terminal His_6-tag. This vector contains a gene encoding Kanamycin resistance.	This study
pET26b_*bd2778*	Modified pET26b vector for the expression of *bd2778* (signal peptide removed (1-138 bp)) with a N-terminal *pelB* signal sequence and a C-terminal His_6-tag. This vector contains a gene encoding Kanamycin resistnance.	This study

3. Strains used in this Study

Table 0-3: Strains used in this Study

Strain	Description	Source
***E. coli* XL10-Gold**	Ultracompetent cells designed for cloning large plasmids and ligated DNA with the highest transformation efficiency possible. Genotype: Tetrdelta- (*mcr*A)183 delta-(*mcr*CB-*hsd*SMR-*mrr*)173 *end*A1 *sup*E44 *thi*-1 *rec*A1 *gyr*A96 *rel*A1 *lac* Hte [F´ *pro*AB *lac*IqZDM15 Tn10 (Tetr) Amy Camr]	Agilent
***E. coli* DH5α**	Competent cells designed for efficient transformations, for subcloning DNA. Genotype: F^- Φ80*lac*ZΔM15 Δ(*lac*ZYA-argF) U169 *rec*A1 *end*A1 *hsd*R17(rk-, mk+) *pho*A *sup*E44 *thi*-1 *gyr*A96 *rel*A1 λ-	NEB
***E. coli* BL21(DE3)**	Chemically competent *E. coli* strain for T7 expression systems, deficient in proteases Lon and OmpT. Genotype: *fhu*A2 [lon] *omp*T *gal* (λ DE3) [dcm] Δ*hsd*S λ DE3 = λ sBamHIo ΔEcoRI-B int::(*lac*I::PlacUV5::T7 *gene*1) i21 Δ*nin*5	NEB

***E. coli* Codon Plus(DE3) RIPL**	Chemically competent *E. coli* strain derived from BL21(DE3) cells, for T7 expression systems, with improved codon bias. Genotype: F– *omp*T *hsd*S(rB– mB–) dcm+ Tetr *gal* λ(DE3) *end*A Hte [argU *pro*L Camr] [*arg*U *ile*Y *leu*W Strep/Specr]	Agilent
***E. coli* C43(DE3) pLysS**	Chemically competent *E. coli* strain derived from C41(DE3) cells, for T7 expression systems, effective in expressing toxic and membrane proteins. Genotype: F – *omp*T *hsd*SB (*r*B- *m*B-) *gal dcm* (DE3) pLysS (CmR)	(Miroux and Walker, 1996)
***E. coli* S17-1**	*E. coli* strain for conjugation of plasmids into *B. bacteriovorus* HD100 and to act as prey for *B. bacteriovorus.* Genotype: *thi*, *pro*, *hsd*R– , *hsd*M+ , *rec*A; integrated plasmid RP4- Tc::Mu-Kn::Tn7)	(Hanahan, 1983)
***E. coli* S17-1:pZMR100**	S17-1 strain containing the plasmid pZMR100 (kan^R) to confer kanamycin resistance to prey	(Rogers et al., 1986)
***B. bacteriovorus* HD100**	*B. bacteriovorus* Type strain, genome sequenced, wild-type	(Rendulic et al., 2004)
B. bacteriovorus* HD100 Δ*bd0427	*B. bacteriovorus* HD100 containing in-frame deletion of *bd0427*	This study in collaboration with Sockett, R.L. Lab, University of Nottingham, UK.

4. Optimisation Trays used in this Study

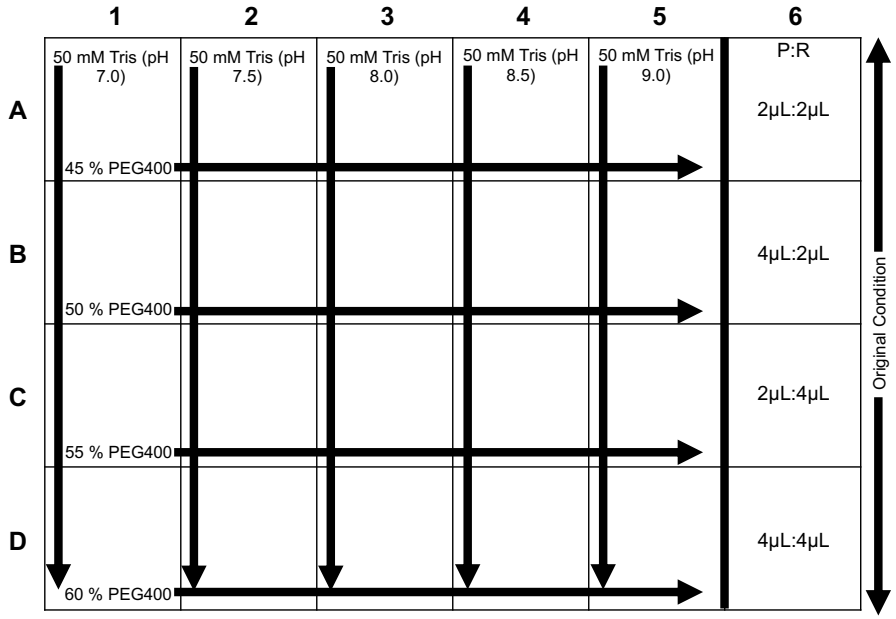
4.1 Seeding trays

Bd0427 ΔSP in 0.3 % (w/v) DDM in MIDASplus and Morpheus. Stock made from MIDASplus G5 crystal. Used at 1 in 100 and 1 in 1000 dilutions. 1 µL protein, 0.5 µL seed stock, 1.5 µL precipitant.

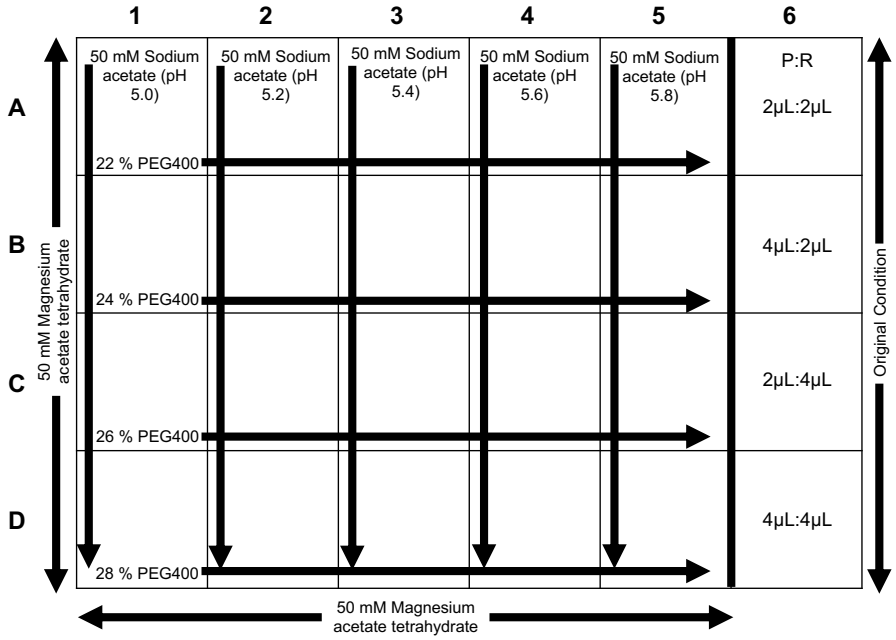
Bd2778 in JCSGplus. Stock made from JCSGplus E1 crystal. Used at 1 in 1000 dilution. 1 µL protein, 0.5 µL seed stock, 1.5 µL precipitant.

Bd2778 in JCSGplus and PACT. Stock made from PACT E10. Used at 1 in 10 and 1 in 100 dilutions in Mosquito Tray. 200 nL protein, 100 nL seed stock, 300 nL precipitant.

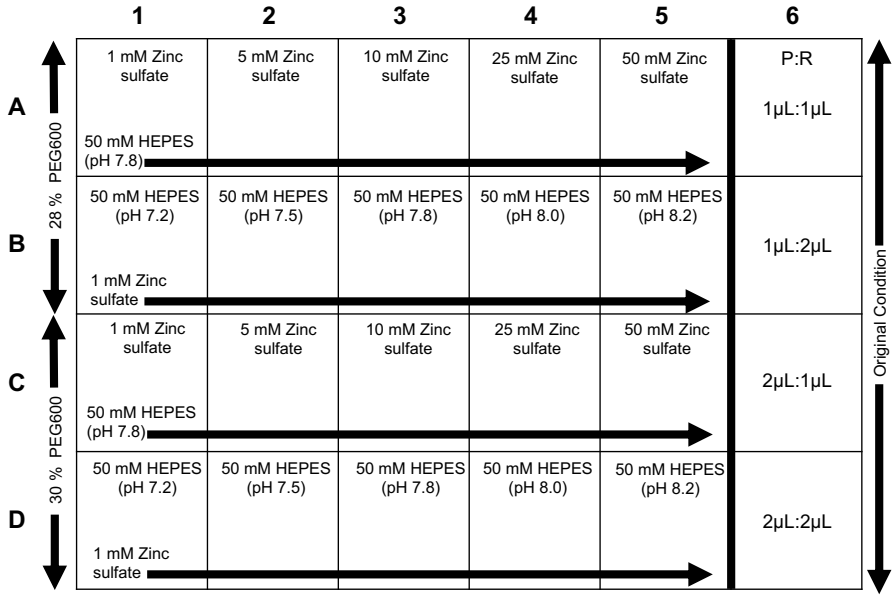
4.2 Bd0427 ΔSP in 0.3 % (w/v) DDM – Optimisation of MIDASplus G5 (55 % (v/v) PEG400)

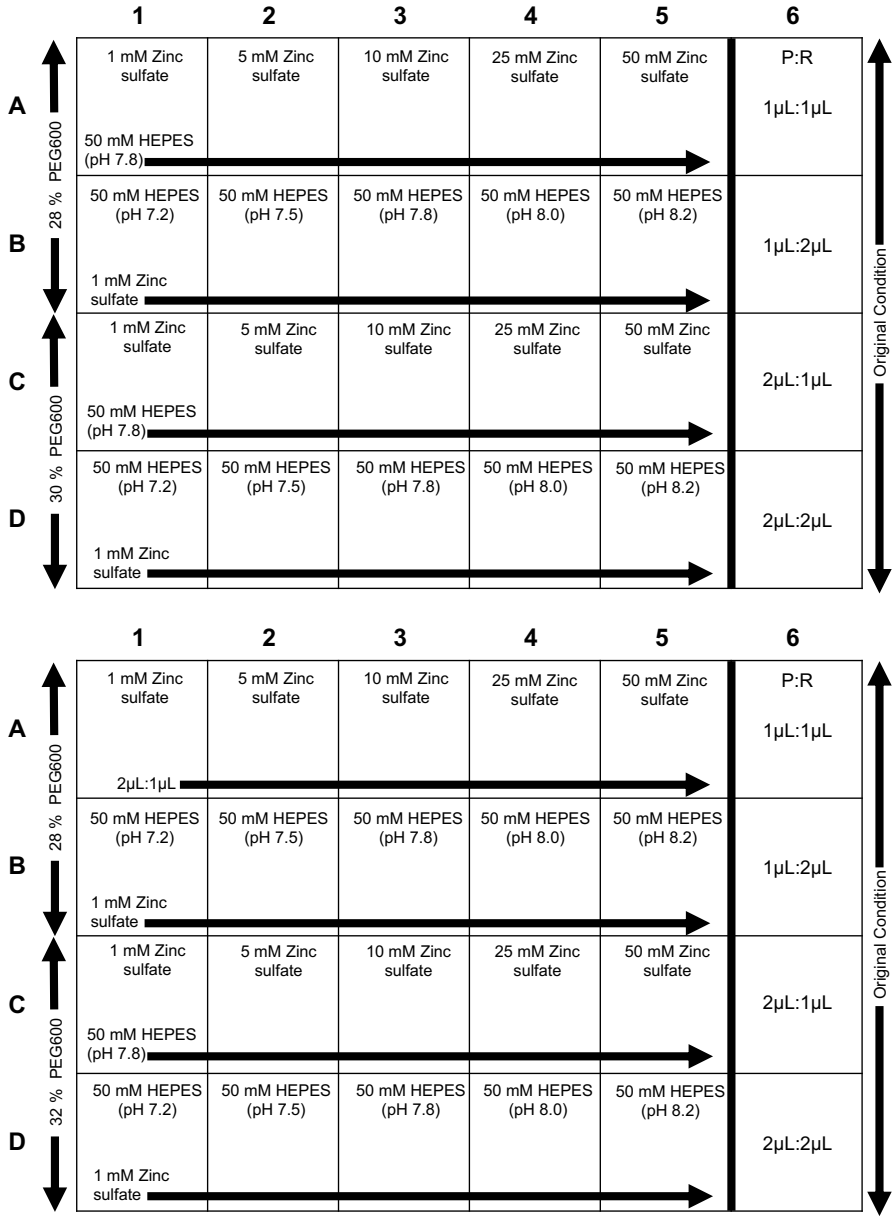


4.3 Bd0427 ΔSP in 50 mM β-OG - Optimisation of MemGold D5 (0.05 M Mg acetate dihydrate, 0.05 M Sodium acetate (5.4), 24 % PEG400)

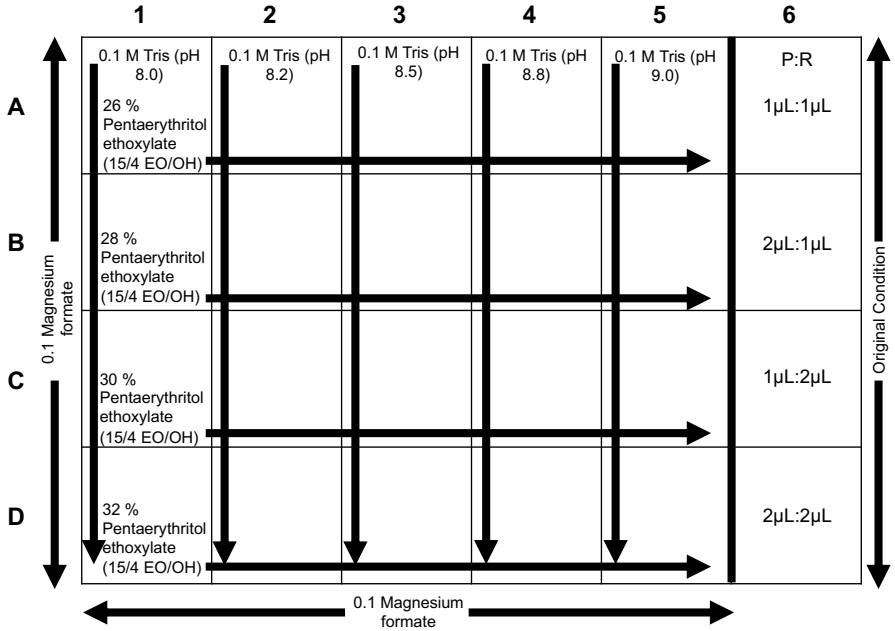


4.4 Se-Met Bd0427 ΔSP in 50 mM β-OG - Optimisations of MemGold G12 (1 mM Zinc sulfate heptahydrate, 0.05 M HEPES (7.8), 28 % (v/v) PEG600), sitting drop

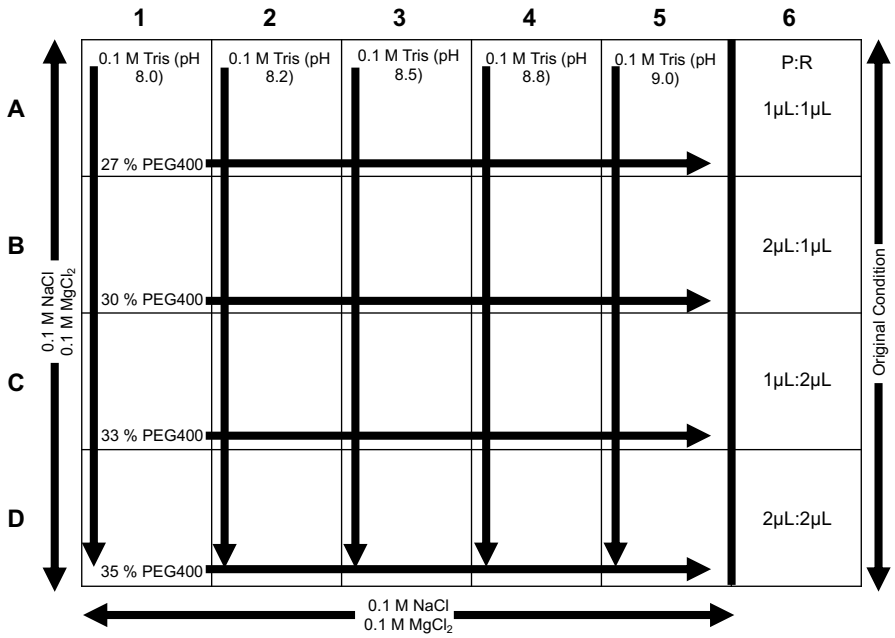




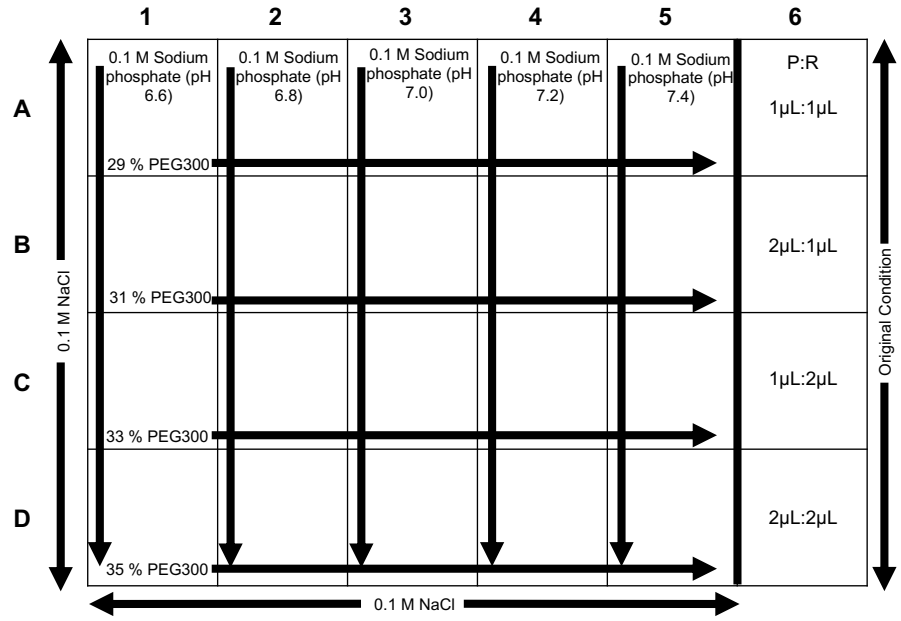
4.5 Se-Met Bd0427 ΔSP in 50 mM β-OG - Optimisation of MIDASplus C11 (0.1 M Magnesium formate dihydrate, 0.1 M Tris (8.5), 30 % (v/v) Pentaerythritol ethoxylate (15/4 EO/OH), sitting drop



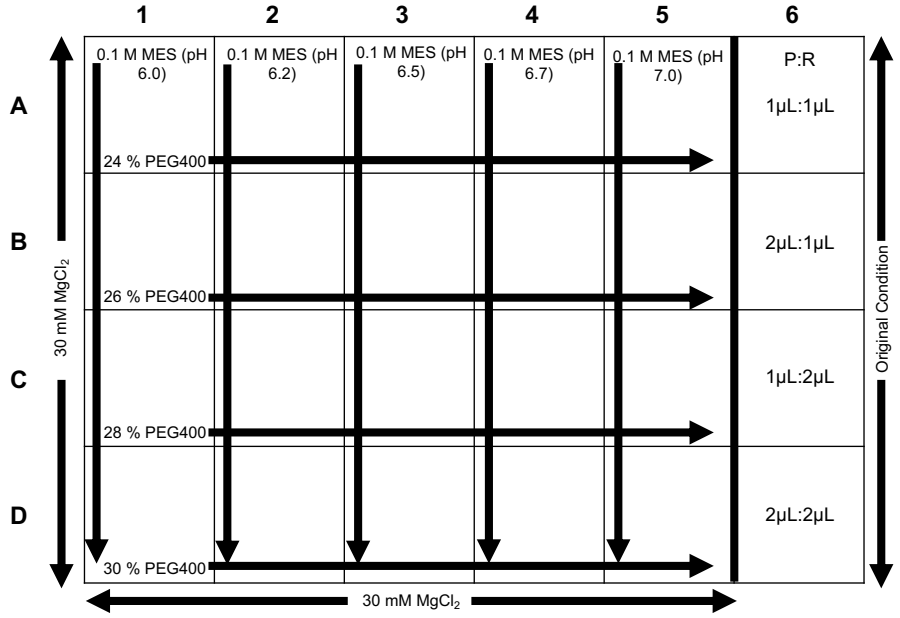
4.6 Se-Met Bd0427 ΔSP in 50 mM β-OG - Optimisation of MemGold G3 (0.05 M Magnesium chloride hexahydrate, 0.1 M NaCl, 0.1 M Tris (8.5), 33 % (v/v) PEG 400), sitting drop



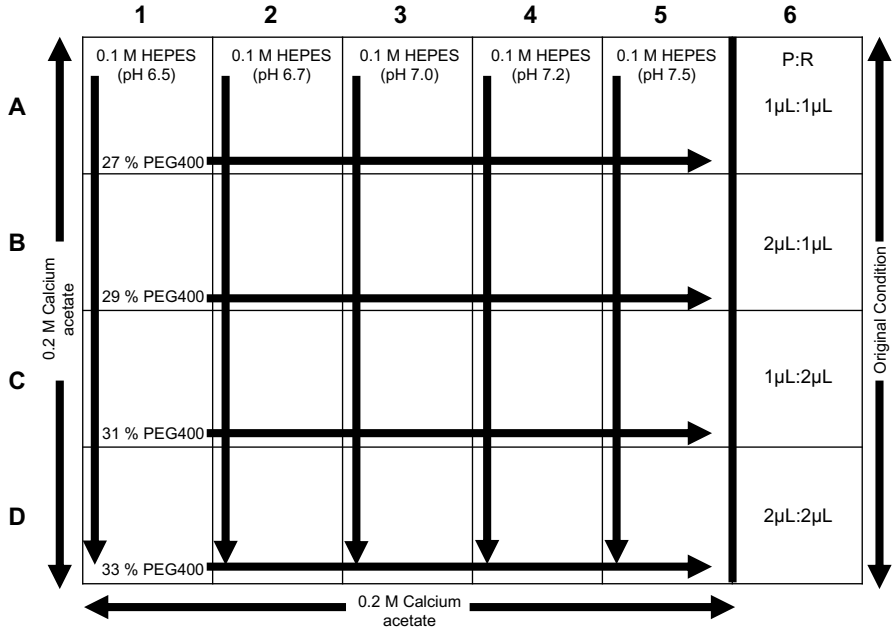
4.7 Se-Met Bd0427 ΔSP in 50 mM β-OG - Optimisation of MemGold H11 (0.1 M NaCl, 0.1 M Sodium Phosphate (pH 7), 33 % PEG300), sitting drop



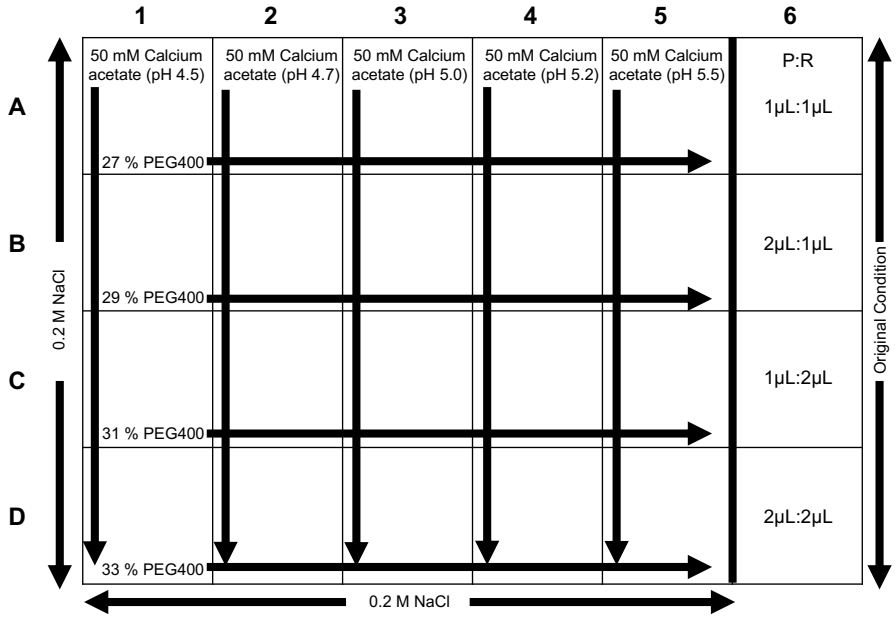
4.7 Se-Met Bd0427 ΔSP in 50 mM β-OG - Optimisation of MemGold2 B6 (0.03 M MgCl2, 0.1 M MES (6.5), 28 % (v/v) PEG400), sitting drop



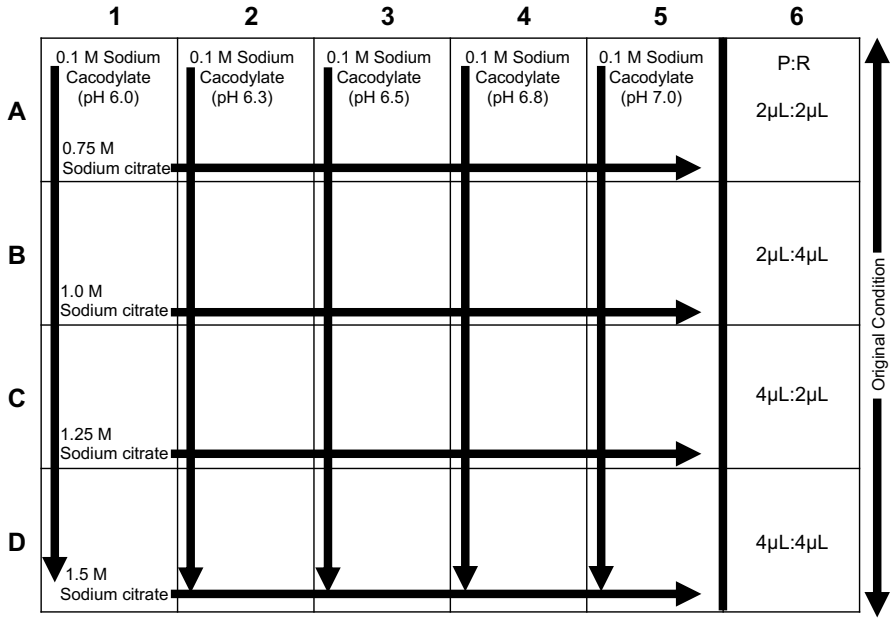
4.8 Se-Met Bd0427 ΔSP in 50 mM β-OG - Optimisation of MemGold2 D12 (0.2 M Calcium acetate, 0.1 M HEPES (7), 24 % PEG400), sitting drop



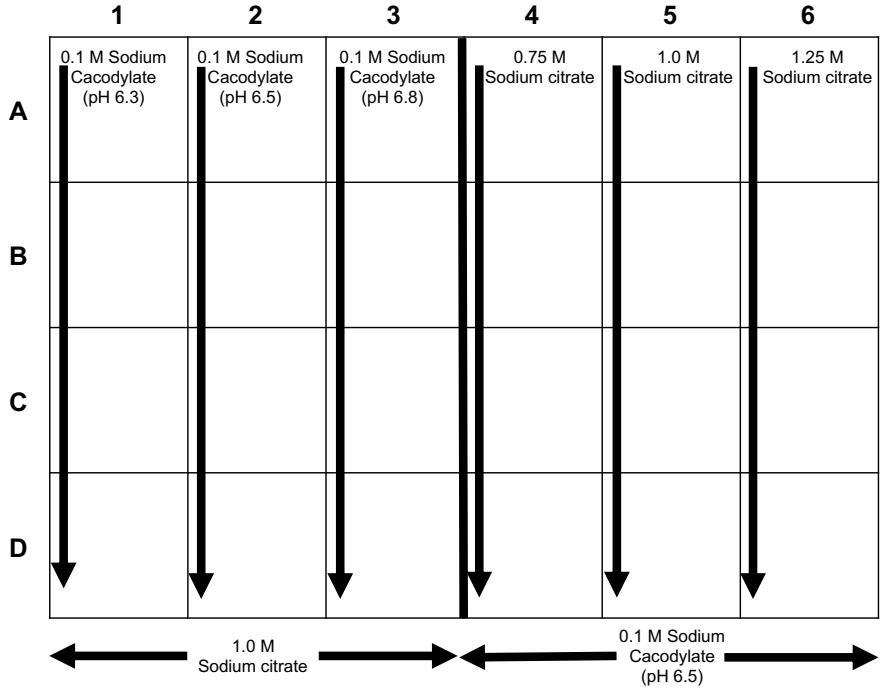
4.9 Se-Met Bd0427 ΔSP in 50 mM β-OG - Optimisation of MemGold2 E11 (0.2 M NaCl, 50 mM Calcium acetate (pH 5), 29 % PEG400), sitting drop



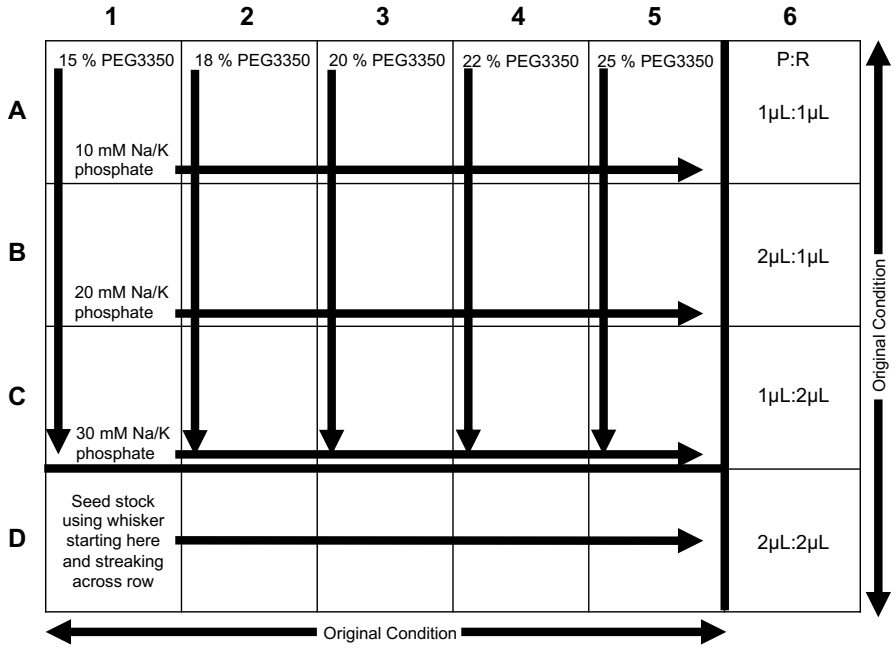
4.10 Bd2778 - Optimisation of JCSGplus E1 (0.1 M sodium cacodylate (6.5), 1 M sodium citrate), JCSGplus E1 seed stock whiskered into each drop



4.11 Bd2778 - Optimisation of JCSGplus E1 (0.1 M sodium cacodylate (6.5), 1 M sodium citrate), JCSGplus E1 seed stock whiskered into each column to dilute



4.12 Bd2778 - Optimisation of PACT E10 (20 mM Na/K phosphate, 20 % PEG3350)


1
2
3
4
5
6
15 % PEG3350
18 % PEG3350
20 % PEG3350
22 % PEG3350
25 % PEG3350
P:R
A
1µL:1µL
10 mM Na/K phosphate
B
2µL:1µL
20 mM Na/K phosphate
C
1µL:2µL
30 mM Na/K phosphate
D
Seed stock using whisker starting here and streaking across row
2µL:2µL
Original Condition
Original Condition