Biophysical characterisation of the *Campylobacter jejuni* β-barrel assembly machinery (BAM) complex mediated outer membrane transport pathway

Gurdip Singh Matharu, MBiolSci, AFHEA

Thesis submitted to the University of Nottingham
for the degree of Doctor of Philosophy

School of Life Sciences
University of Nottingham
Initially submitted July 2021

Declaration of Originality

I, Gurdip Matharu, declare that this thesis is an original piece of work written entirely by myself. All of the results arise from my own research unless otherwise stated. Any information and conclusions derived from other works have been referenced appropriately.

Abstract

Outer membrane proteins in Gram-negative bacteria perform many functions that are key to bacterial survival, extracellular transport, and virulence. In order for these proteins to get to the outer membrane where they perform their function, they engage in a transport pathway starting from the ribosome, traversing the inner membrane to the periplasm where they are folded and assembled into the outer membrane. Several transport pathways contribute to this final folding and assembly step, however, it was recently established that the majority of outer membrane folding and assembly in Gram-negative bacteria is attributed to the β-barrel assembly machinery (BAM) complex. This PhD aims to extend the present knowledge on the BAM complex through biophysical and computational characterisation of the key proteins involved in the BAM-mediated transport pathway of *Campylobacter jejuni*.

This thesis describes the preparation of *C. jejuni* BamA, BamD, Cj1289, DegP and LptD from cloning through to expression and purification, and subsequent experimental characterisation using circular dichroism and small angle X-ray scattering where folding, melting temperatures, complexation and protein-protein interactions are analysed. In tandem with the experimental wet work, homology models were generated and *in silico* docking runs were used to analyse the docking interface between BamA and BamD, and to prepare models for molecular dynamics (MD) simulations. Further experiments in the pipeline for elucidating structural insights through cryogenic transmission electron microscopy and MD simulations are briefly described.

The results presented in this work suggest three main findings. Firstly, Landau free energy expansion models may provide more accurate melting temperature estimations than the current Boltzmann analysis methodology. Secondly, that there is an interaction between Cj1289 (SurA-like chaperone) and BamD not previously cited in the literature. Thirdly, given knowledge of an approximate binding position, MD simulations may be a potential alternative to rigid docking that provides additional information on docking site and conformational sampling occurring prior to the final expected binding position.

Acknowledgments

First, I would like to thank my supervisor Dr Boyan Bonev for the opportunity to work on this challenging and diverse project, and for his continuing support, discussions about science, unwavering enthusiasm and understanding.

Much of this project was helped along by researchers both inside and outside of the lab. I would like to thank the lab members for any advice given during this project, and to Dr Vivien Yeh for being the other person who stays late in the lab. From the Wheatley lab, I would like to thank Amelia Townley and Sophie Rockett for advice about cloning, and running expression tests on proteins. From the Chalmers and Bolt labs, Jack Braithwaite, Andy Cubbon, and both PIs for the different cells types that were used in this project. Many thanks to Phil Bardalang for the FPLC demo, Stephen Hall (Engineering) for use of a fermenter, James Doutch for helping me with the SAXS analysis, and to the other beamline scientists involved in this project.

A big thank you to Isabel Moraes (NPL) for the three-month industry placement, and Andrew Quigley for allowing me to use the (MPL) Harwell facilities. Both opportunities came quite late in the project, but were extremely useful experiences.

A thank you to the University of Nottingham for the opportunity to teach and plan out projects for undergraduates; and to Jack and Micha, for their hard work, amusing mess-ups, and for teaching me things about myself and where I need to develop.

On a personal note, a massive thank you to Dr Will Gilder (chiropractor), Joe Pheasant (sports physio), Harry Andrews (specialist physio), and to all the other medical personal involved for helping me to recover from severe spinal injuries that occurred at the start of the doctoral programme. Also, to the many people who have kept me going throughout this process. You are greatly appreciated.

A thank you to the examiners Dr Steve Atkinson and Dr Tim Knowles for their advice and feedback and to Steve for going through the many corrections.

Finally, I would like to thank the BBSRC and Nottingham DTP for funding and facilitating this project, and the Nottingham HPC and Midlands HPC Plus clusters for use of their high-performance computational infrastructures.

Table of contents

Declaration of Originality **2**

Abstract **3**

Acknowledgments **4**

Table of contents **5**

List of figures **17**

List of tables **22**

List of abbreviations **24**

Amino acids **27**

1 Introduction **28**

1.1 Campylobacteriosis **28**

1.2 OMP transport story **29**

1.3 The BAM Complex **31**

1.3.1 Essential conserved components 32

1.3.1.1 BamA 32

1.3.2 POTRA domains 33

1.3.3 Extracellular Loop 6 35

1.3.4 β-barrel 35

1.3.5 BamD 37

1.3.6 Accessory components 38

1.3.6.1 BamB 39
1.3.6.2 BamC 39
1.3.6.3 BamE 40
1.3.7 Other potentially associated and alternative subunits 40
1.4 Chaperones 41
1.4.1 Skp 42
1.4.2 SurA 42
1.4.3 DegP 43
1.4.4 BepA 43
1.5 Mechanisms 44
1.5.1 Overview of current models 44
1.5.1.1 Disruptase 44
1.5.1.2 β-augmentation 45
1.5.2 Substrate recognition 47
1.6 Current progress in the *C. jejuni* BAM complex 49
1.7 Thesis aims 50
1.7.1 Specific chapter aims 50
2 Materials and methods 52
2.1 Materials 52
2.1.1 Reagents 52
2.1.2 Sterilisation 52
2.1.3 Media 52
2.1.4 Antibiotics 53
2.1.5 Buffers and stains 54

2.1.6 Kits 55

2.1.7 Detergents and synthetic lipids 55

2.1.8 Bacterial strains and plasmids 57

2.1.9 Molecular weight ladders 60

2.1.10 Equipment 60

2.1.11 Primers 61

2.1.12 Computing resources 64

2.1.13 HPC Midlands Plus acknowledgement 66

2.2 Methods 67

2.2.1 Cloning 67

2.2.1.1 Gene sequence 67

2.2.1.2 Operon search 67

2.2.1.3 Protein sequence 67

2.2.1.4 Determination of signal sequence 68

2.2.1.5 *C. jejuni* plate culture 68

2.2.1.6 *C. jejuni* glycerol stocks 69

2.2.1.7 Preparation of DNA template 69

2.2.1.8 PCR 69

2.2.1.9 Agarose gel electrophoresis 71

2.2.1.10 PCR clean-up 71

2.2.1.11 Restriction digest 71

2.2.1.12 Digest check 72

2.2.1.13 Gel Extraction 72

2.2.1.14 Ligation 72

2.2.1.15 $CaCl_2$ (and $MgCl_2$) chemically competent cell preparation 73

2.2.1.16 Heat shock transformation 74
2.2.1.17 Plasmid mini-preps 74
2.2.1.18 DNA quantification 74
2.2.1.19 Plasmid screening 74
2.2.1.20 Sanger sequencing 75
2.2.1.21 ChromasLite 75
2.2.1.22 Snapgene viewer 75
2.2.1.23 A Plasmid Editor 75
2.2.2 Bacterial overexpression, protein purification and preparation for further experiments 75
2.2.2.1 *E. coli* overnight cultures 75
2.2.2.2 Glycerol stocks 75
2.2.2.3 Protein overexpression 76
2.2.2.4 Expression testing of new constructs 76
2.2.2.5 Pellet lysis 76
2.2.2.6 Cell disruption 77
2.2.2.7 Sonication 77
2.2.2.8 Protein purification 77
2.2.2.9 Membrane protein purification 77
2.2.2.10 Soluble protein purification 78
2.2.2.11 IMAC purification 78
2.2.2.12 SDS-PAGE gels 79
2.2.2.13 Coomassie blue staining 80
2.2.2.14 Western blotting 80
2.2.2.15 Gel scan 81
2.2.2.16 Dialysis 81
2.2.2.17 FPLC 82

2.2.2.18 Column preparation 82
2.2.2.19 System degassing 82
2.2.2.20 Column and FPLC storage 83
2.2.2.21 Size exclusion chromatography 83
2.2.2.22 Concentration of proteins 83
2.2.3 Screening and optimisation of recovery of proteins 84
2.2.3.1 Detergent screening 84
2.2.3.2 Liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis 86
2.2.3.3 BamA IMAC resin choice and imidazole concentration optimisation 86
2.2.3.4 BamA TALON resin binding time optimisation IMAC screen 86
2.2.3.5 BamA Buffer and NaCl concentration IMAC screen 87
2.2.3.6 BamA 1 M salt and cofactor IMAC screen 87
2.2.3.7 Generation of *ΔbamA* and *ΔbamD* competent cells 88
2.2.3.8 Generation of constructs for complementation assay 88
2.2.3.9 BamD sucrose gradient assay 88
2.2.3.10 Cross-species complementation assay 89
2.2.4 Circular dichroism 89
2.2.4.1 Preparation of proteins 89
2.2.4.2 Annealing of *in silico* homology models 90
2.2.4.3 *In-silico* predictions of secondary structure assignments 91
2.2.4.4 Generation of *in silico* predictions of secondary structure assignments 91
2.2.4.5 Chirascan spectra data collection 92
2.2.4.6 Thermal unfolding 92
2.2.4.7 Predictions of mixture secondary structure estimations 93

2.2.4.8 Protein concentrations, molecular weights, and extinction coefficients 93
2.2.5 Small angle X-ray scattering 93
2.2.5.1 Sample preparation and instrumentation 94
2.2.5.2 Complexation analysis 96
2.2.5.3 Data collection 96
2.2.6 *In silico* docking and molecular dynamics simulations 97
2.2.6.1 Generation of homology models 97
2.2.6.2 Discovery studio/ UCSF Chimera 97
2.2.6.3 ZDock 97
2.2.6.4 pyDock 98
2.2.6.5 RosettaDock (Docking 2) 98
2.2.6.6 Editing of PDBs 99
2.2.6.7 Protonation state 99
2.2.6.8 CHARMM-GUI 99
2.2.6.9 Debye length calculation 101
2.2.6.10 MD Force fields 102
2.2.6.11 Molecular dynamics simulations 102
2.2.6.12 Analysis of molecular dynamics simulations 103
3 Cloning, expression, purification, screening, and optimisation 105
3.1 Chapter aims 105
3.2 Results 105
3.2.1 Cloning 105
3.2.1.1 Homology searches and target selection 105
3.2.1.2 Determination of the signal peptide leader sequence 106

3.2.1.3 Constructs 106
3.2.2 Confirmation of constructs 109
3.2.2.1 Western blots 109
3.2.2.2 LC-MS/MS 110
3.2.3 Protein expression 110
3.2.4 Protein purification 111
3.2.4.1 *C. jejuni* BamA 111
3.2.4.2 *C. jejuni* BamD 114
3.2.4.3 Cj1289 116
3.2.4.4 *C. jejuni* DegP 117
3.2.4.5 *C. jejuni* LptD 117
3.2.5 Functional assays 119
3.2.5.1 *C. jejuni* BamD interaction with lipid 119
3.2.5.2 Cross-species complementation assay 120
3.3 Discussion 122
3.4 Summary 123
4 Secondary structure and melting temperature characterisation of the BAM complex in *C. jejuni* using circular dichroism 124
4.1 Chapter aims 124
4.2 Introduction 125
4.2.1 Introduction to Circular Dichroism 125
4.2.2 Biological background to experiment 128
4.2.3 Thermal Degradation experiments 129
4.2.4 Thermal stability expectations 130

4.2.5 Derivation of critical points of folded and unfolded states during thermal degradation....132
4.2.6 Landau free energy expansion models....134
4.2.7 CD analysis....136
4.3 Results....137
4.3.1 In-silico predictions of protein SSAs....137
4.3.2 Chirascan spectrophotometer experiments....138
4.3.2.1 BamA....140
4.3.2.2 BamD....140
4.3.2.3 Cj1289....146
4.3.2.4 BamA interacts with BamD....150
4.3.2.5 BamA does not show a co-stabilising interaction with Cj1289....152
4.3.2.6 BamD interacts with Cj1289 in 1:1 stoichiometry....159
4.3.2.7 Secondary structure stability of BamA-BamD-Cj2189 1:1:1 stoichiometric mixture exceeds experimental temperature range....167
4.3.2.8 Landau models and tau....170
4.4 Discussion....176
4.5 Summary....184
4.6 Limitations....185
5 Small angle X-ray scattering analysis of protein-protein interaction in the BAM mediated outer membrane transport pathway....186
5.1 Chapter aims....186
5.2 Introduction....187
5.2.1 Biological background....188

5.2.1.1 Interactions expected from the *E. coli* BAM mediated transport pathway 188
5.2.1.2 Interactions expected from the CD data 189
5.2.2 The SAXS curve 189
5.2.2.1 Guinier analysis 189
5.2.2.2 Kratky analysis 190
5.2.2.3 Flexibility analysis 191
5.2.2.4 Pair density distribution function 191
5.2.2.5 Complexation analysis 192
5.3 Results 193
5.3.1 BamD 193
5.3.2 Cj1289 193
5.3.3 BamA, DegP and LptD 197
5.3.3.1 BamD-Cj1289 199
5.3.3.2 Other mixtures 203
5.3.4 Homology model fitting to experimental data 203
5.3.5 FoxS and GAJOE cross-validation of SAXS profiles 208
5.3.6 Complexation analysis 212
5.4 Discussion 218
5.5 Summary 224
6 Chapter 6: *In silico* investigation of the *C. jejuni* BamA-BamD docking interface using rigid body docking and Debye-length modulated molecular dynamics simulations 226
6.1 Chapter aims 226

6.2 Introduction 228

6.3 Results 229

6.3.1 Determining the docking contact interface 229

6.3.2 Rigid body docking 231

6.3.2.1 ZDock rigid body docking runs 231

6.3.2.2 PyDock rigid body docking runs 233

6.3.2.3 RosettaDock rigid body and local refinement docking runs 234

6.3.3 Molecular dynamics-based docking runs 235

6.3.3.1 MD Docking run A0D0 238

6.3.3.2 MD Docking run A0D0 normal 238

6.3.3.3 MD Docking run A0D1 241

6.3.3.4 MD Docking run A0D2 243

6.3.3.5 MD Docking run A90D0 245

6.3.3.6 MD Docking run A90D1 245

6.3.3.7 MD Docking run A90D2 248

6.3.3.8 MD Docking run A180D0 248

6.3.3.9 MD docking run A180D1 251

6.3.3.10 MD docking run A180D2 253

6.3.3.11 MD docking run A270D0 253

6.3.3.12 MD docking run A270D1 256

6.3.3.13 MD docking run A270D2 258

6.4 Discussion 261

6.5 Summary 265

7 Concluding remarks and future work 267

7.1 Concluding remarks 267

7.2 Future work 268

7.2.1 Extensions to current work....268
7.2.2 Cryo-TEM....269
7.2.3 Unanalysed molecular dynamics simulations....271
8 Appendices....273
8.1 Chapter 1 appendices....273
8.1.1 Operon search....273
8.2 Chapter 2 appendices....274
8.2.1 Expression testing variables....274
8.2.2 Optimised expression conditions....275
8.2.3 Experiment run sheet....276
8.3 Chapter 3 appendices....278
8.3.1 pET26b+ vector map....278
8.3.2 Protein sequences of proteins in this study....279
8.3.3 LC-MS/MS Mascot analysis of protein in this study....281
8.3.4 Biotools analysis of LC-MS/MS....282
8.3.5 Further optimisation of the BamA purification process....287
8.3.6 BamA detergent exchange screen for future structural determination by crystallography....291
8.3.7 BamA reverse IMAC purification following detergent exchange....292
8.3.8 LptD initial detergent screening....293
8.4 Chapter 4 appendices....294
8.4.1 *In silico* SSA calculations....294
8.4.2 Goodness of fit - SSEs....296

8.4.3 BamD concentration inaccurate 297

8.4.4 Individual protein *vs.* protein mixture thermal degradation and SSE comparison panels 298

8.4.5 Boltzmann melting temperature estimation is wavelength dependent 302

8.5 Chapter 5 appendices 303

8.6 Chapter 6 appendices 306

8.6.1 Docking interface analysis for residues in contact 306

8.6.2 Rigid body docking outputs 308

8.6.3 MD scaling runs 313

8.6.4 BamA and BamD structural elements by residue number 315

8.6.5 Starting positions – MD docking assay 316

8.6.6 Trajectory RMSD for docking runs (pre-PBC processing) 317

8.6.7 Trajectory RMSF for docking runs (pre-PBC processing) 319

8.7 Chapter 7 appendices 321

8.7.1 Cryo-TEM 2D classes 321

8.7.2 MD simulations to be analysed 322

8.8 Doctoral training programme appendices 324

8.8.1 PIP 324

8.8.2 DTP Lab Rotation Poster 326

8.8.3 PGR symposium poster 327

8.9 References 328

List of figures

Figure 1.1: *E. coli* BAM-mediated outer membrane transport pathway. 30
Figure 1.2: The structural features of BamA. 32
Figure 1.3: 'Open' and 'Closed' states of the BAM complex. 36
Figure 1.4: Structural features of *C. jejuni* BamD. 37
Figure 1.5: Evolutionary conservation of the BAM complex subunits. 41
Figure 1.6: Overview of the hybrid barrel formation and budding mechanism. 46
Figure 1.7: *C. jejuni* BAM-mediated outer membrane transport pathway. 49
Figure 3.1: Signal sequence determination of *C. jejuni* BamA, Cj1289 and BamD. 107
Figure 3.2: Signal sequence determination of *C. jejuni* DegP and LptD. 108
Figure 3.3: SDS-PAGE and Western blot analysis of IMAC purified proteins. 109
Figure 3.4: BamD repeat Western blot. 110
Figure 3.5: BamA purification. 112
Figure 3.6: CjBamA detergent screening. 113
Figure 3.7: CjBamD PBS *v*s. TBS Western blot. 115
Figure 3.8: CjBamD purification. 115
Figure 3.9: Cj1289 purification. 116
Figure 3.10: CjDegP purification. 117
Figure 3.11: CjLptD purification. 118
Figure 3.12: CjLptD detergent screen. 118
Figure 3.13: CjBamD interacts with DOPC LUVs. 119
Figure 4.1: Circular dichroism spectra of myoglobin, hen egg lysozyme, triosephosphate isomerase and chymotrypsin. 128
Figure 4.2: BAM-mediated outer membrane transport pathway in *C. jejuni*. 129
Figure 4.3: Thermal melt profile expectations of co-stabilisation and co-melting, thermal proximity coaggregation, and non-interaction between proteins. 133
Figure 4.4: The spinodal of the transition of the folded to unfolded state of the α-helical (or β-strand) content of a protein. 135
Figure 4.5: Buffer 1-n spectral scans. 138
Figure 4.6: 1-n Spectral scan of all proteins pre-melt. 139
Figure 4.7: Thermal degradation spectra, Boltzmann fitting and SSE of BamA. 141

Figure 4.8: Thermal degradation spectra of BamD in both non-detergent [NaF-Tris] and detergent [SB12] buffer types shows increased CD signal stability in the presence of SB12.....142
Figure 4.9: Boltmann fitting analysis of BamD in both non-detergent [NaF-Tris] and detergent [SB12] buffer types. 144
Figure 4.10: Secondary structure estimation of BamD in non-detergent [NaF-Tris] and detergent [SB12] buffer types. 145
Figure 4.11: Thermal degradation profiles of Cj1289 in both buffer types. 147
Figure 4.12: Boltmann fitting analysis of Cj1289 in both non-detergent [NaF-Tris] and detergent [SB12] buffer types. 148
Figure 4.13: Secondary structure estimation of Cj1289 in non-detergent [NaF-Tris] and detergent [SB12] buffer types. 149
Figure 4.14: Thermal degradation spectra, Boltzmann fitting and SSE of a 1:1 BamA-BamD mixture. 151
Figure 4.15: BamA-BamD mixture additive spectra versus experimental data.....153
Figure 4.16: Thermal degradation profiles of BamA-Cj1289 mixtures in 1:1 and 1:3 stiochiometries. 154
Figure 4.17: Boltmann fitting analysis of BamA-Cj1289 mixture in 1:1 and 1:3 stoichiometry in detergent [SB12] buffer. 156
Figure 4.18: Secondary structure estimation of BamA-Cj1289 mixture 1:1 and 1:3 stoichiometry.....157
Figure 4.19: BamA-Cj1289 mixture additive spectra versus experimental data. 158
Figure 4.20: Thermal degradation spectra, Boltzmann fitting and SSE of a 1:1 BamD-Cj1289 mixture.....160
Figure 4.21: Thermal degradation spectra of BamD-Cj1289 mixtures in 3:1 and 1:3 stoichiometries in non-detergent and detergent buffer backgrounds. 161
Figure 4.22: Boltmann fitting analysis of BamD-Cj1289 mixture in 1:3 and 3:1 stoichiometry in NaF-Tris and SB12 buffer.....162
Figure 4.23: Secondary structure estimation of BamD-Cj1289 mixtures in 3:1 and 1:3 stoichiometries in non-detergent and detergent buffer backgrounds. 164
Figure 4.24: BamD and Cj1289 (1:1) mixture additive spectra versus experimental data. 165
Figure 4.25: BamD and Cj1289 1:3 and 3:1 stoichiometry mixture additive spectra versus experimental data. 166

Figure 4.26: BamA: BamD: Cj1289 [1:1:1] mixture thermal degradation spectra, SSE and Boltzmann fitting. 168
Figure 4.27: BamA:BamD:Cj1289 (1:1:1) mixture additive spectra versus experimental data. 169
Figure 4.28: Landau models of BamA, BamD, and BamA:BamD mixture. 171
Figure 4.29: Landau models of Cj1289 and Cj1289: [protein] 1:1 mixtures. 172
Figure 4.30: Landau models of the BamA:BamD:Cj1289 1:1:1 mixture and summary table of derived Landau parameters. 173
Figure 4.31: Calculation and plot of tau for all proteins and their mixtures. 175
Figure 5.1: BamD SAXS analysis. 194
Figure 5.2: Cj1289 SAXS analysis. 195
Figure 5.3: Cj1289 P(r) calculation and refinement. 196
Figure 5.4: Cj1289 DAMMIF model comparison to rigid body structure. 198
Figure 5.5: BamD-Cj1289 SAXS analysis. 200
Figure 5.6: BamD-Cj1289 P(r) calculation and refinement. 201
Figure 5.7: BamD-Cj1289 bead modelling. 202
Figure 5.8: Experimental data validation of Cj1289, BamD and BamD-Cj1289 mixture using model structures. 204
Figure 5.9: Experimental data validation of DegP using monomer, hexamer, dodecamer and 24-mer model structures. 206
Figure 5.10: Experimental data validation of BamA and LptD using homology model structures. 207
Figure 5.11: Cj1289 FoxS fitting cross-validation. 209
Figure 5.12: Cj1289 GAJOE fitting and validation. 210
Figure 5.13: BamD-Cj1289 FoxS fitting cross-validation. 211
Figure 5.14: Complexation analyses of 1:1 protein mixtures that indicate interaction. 213
Figure 5.15: Complexation analyses of 1:1 protein mixtures that indicate no interaction. 214
Figure 5.16: Complexation analyses of DegP and BamA mixtures in 6:1 and 24:1 stoichiometries at 16 °C and 30°C. 216
Figure 5.17: ConSurf analysis of Cj1289 conserved residues. 219
Figure 5.18: Cj1289 DAMMIN model fitting. 220
Figure 6.1: *E. coli* models and *C. jejuni* I-TASSER homology model output. 230

Figure 6.2: Rigid body docking input and output models. 232
Figure 6.3: MD docking orientation translations. 237
Figure 6.4: A0D0 MD docking run. 239
Figure 6.5: A0D0 normal MD docking run. 240
Figure 6.6: A0D1 MD docking run. 242
Figure 6.7: A0D2 MD docking run. 244
Figure 6.8: A90D0 docking run. 246
Figure 6.9: A90D1 MD docking runs. 247
Figure 6.10: A90D2 MD docking run. 249
Figure 6.11: A180D0 MD docking run. 250
Figure 6.12: A180D1 MD docking run. 252
Figure 6.13: A180D2 MD docking run. 254
Figure 6.14: A270D0 MD docking run. 255
Figure 6.15: A270D1 MD Docking run. 257
Figure 6.16: A270D2 MD docking run. 259
Figure 8.1: Operon search around *bamA* and *bamD* gene loci. 273
Figure 8.2: pET26b+ vector map. 278
Figure 8.3: SDS-PAGE excised gel bands for LC-MS/MS analysis. 281
Figure 8.4: Biotools confirmation of BamA (LC-MS/MS) 282
Figure 8.5: Biotools confirmation of Cj1289 (LC-MS/MS) 283
Figure 8.6: Biotools confirmation of LptD (LC-MS/MS) 284
Figure 8.7: Biotools confirmation of DegP (multimer, LC-MS/MS) 285
Figure 8.8: Biotools confirmation of DegP (monomer, LC-MS/MS) 286
Figure 8.9: BamA IMAC pulldown optimisation – Resin, wash buffer imidazole concentration, and elution volume. 287
Figure 8.10: TALON pulldown binding optimisation of DDM extracted BamA. 288
Figure 8.11: BamA buffer and salt concentration IMAC screening. 289
Figure 8.12: 1 M NaCl and buffer cofactor TALON IMAC binding screen. 290
Figure 8.13: CjBamA follow-up detergents screen in Cymal-5, DDM, LDAO and LMNG. 291
Figure 8.14: Reverse IMAC purification of CjBamA. 292
Figure 8.15: LptD initial detergent screen. 293
Figure 8.16: BamD [NaF-Tris] SSE with concentration reduced by one third 297

Figure 8.17: Thermal degradation profiles of BamA and BamA: [protein] mixtures. 298
Figure 8.18: SSE profiles of BamA and BamA: [protein] mixtures. 299
Figure 8.19: Thermal degradation profiles of BamD in both buffers and BamD: [protein] mixtures. 300
Figure 8.20: SSE profiles of BamD (both buffers) and BamD: [protein] mixtures. 301
Figure 8.21: Boltzmann fitting of BamA and Cj1289 (1:1) mixture CD signal at 192 nm and 209 nm illustrates how wavelength alters Boltzmann derived T_m. 302
Figure 8.22: BamA SAXS analysis. 303
Figure 8.23: DegP SAXS analyis. 304
Figure 8.24: LptD SAXS analysis. 305
Figure 8.25: Residues involved in *E. coli* BamA-BamD interaction interface. 306
Figure 8.26: Approximation of the residues involved in the CjBamAD docking interface by superimposition to homologous structures. 307
Figure 8.27: ZDock rigid body docking output. 308
Figure 8.28: pyDock rigid body modelling outputs. 309
Figure 8.29: RosettaDock rigid body modelling ouput. 310
Figure 8.30: MD scaling runs A0D0-A90D2. 313
Figure 8.31: MD scaling runs A180D0-A270D2. 314
Figure 8.32: Trajectory RMSD A0D0-A90D1 MD runs (PBC wrap). 317
Figure 8.33: Trajectory RMSD A90D2-A270D2 MD runs (PBC wrap). 318
Figure 8.34: Trajectory RMSF A0D0-A90D1 MD runs (PBC wrap). 319
Figure 8.35: Trajectory RMSF A90D2-A270D2 MD runs (PBC wrap). 320
Figure 8.36: BamAD Cryo-TEM screening. 321

List of tables

Table 2.1: Bacteriological media used in this study 53
Table 2.2: Antibiotics used in this study.....53
Table 2.3: Composition of buffers used in this study 55
Table 2.4: Kits used in this study.....55
Table 2.5: Detergents and synthetic lipids used in this study 56
Table 2.6: Bacterial strains and plasmids used in this study.....59
Table 2.7: Molecular weight ladders used in this study.....60
Table 2.8: Equipment purchased from this grant to enable this study 60
Table 2.9: List of primers used in this study.....63
Table 2.10: Webservices and software used in this study.....66
Table 2.11: High performance computing resources used in this study 66
Table 2.12: Standard taq PCR reaction set up.....69
Table 2.13: Standard taq PCR cycle conditions.....70
Table 2.14: Standard Q5 PCR reaction set up.....70
Table 2.15: Standard Q5 PCR cycle conditions.....70
Table 2.16: Standard restriction digest set up 72
Table 2.17: Standard Ligation reaction set up 73
Table 2.18: Stacking gel composition and set up.....79
Table 2.19: Separating gel composition and set up.....80
Table 2.20: The critical micelle concentrations (CMC), extraction and purification concentrations of detergents used in this study.....84
Table 2.21: Concentrations, molecular weights and extinction coefficients of proteins used in this study.....93
Table 2.22: SAXS data collection beamline parameters.....97
Table 2.23: Vector alignments used in PDB set up to add membrane bilayer 99
Table 3.1: Cross-species complementation assay - *C. jejuni* pET20::*bamA* vector supplementation of native BamA depleted (JWD3) *E. coli* K12.....121
Table 3.2: Cross-species complementation assay - *C. jejuni* pET26b::*bamD* vector supplementation of native BamD depeleted (JCM90) *E. coli* K12.....121
Table 4.1: Averaged *in silico* estimated SSAcomponent values of BamA, BamD, Cj1289 and 1:1 protein mixtures.....137

Table 4.2: Experimental secondary structure estimation component values of all proteins in this study at 20°C before thermal degradation. 139
Table 5.1: Summary table of interactions between proteins in this study. 217
Table 6.1: Residues important for docking of CjBamA-CjBamD. 260
Table 8.1: Expression screening conditions tested in this study. 274
Table 8.2: Optimised expression conditions for proteins in this study. 275
Table 8.3: Circular dichroism experiment list. 277
Table 8.4: Sample molecular weights and extinction coefficients. 277
Table 8.5: Protein sequences of BamA, Cj1289, BamD, DegP and LptD. 280
Table 8.6: Calculated secondary structures for Cj1289, BamD-Cj1289 (1:1), BamA-Cj1289 (1:1), and BamA-BamD-Cj1289 (1:1:1) mixtures from homology model PDB files. 294
Table 8.7: Calculated secondary structures for BamA, BamD and BamAD from homology model and equilibrated MD simulations last step PDB files. 295
Table 8.8: Goodness of fit of secondary structure estimations. 296
Table 8.9: Residues within contact distance (≤ 4Å) in the selected rigid body docking outputs. 312
Table 8.10: BamA and BamD structural elements by residue number. 315
Table 8.11: MD simulations to be analysed. 323

List of abbreviations

APS	Ammonium Persulfate
Ar	Aromatic residue
BAM	β-barrel assembly machinery complex
BamA	β-barrel assembly machinery complex factor A, BamA
BamD	β-barrel assembly machinery complex factor D, BamD
BepA	β-Barrel Assembly-Enhancing Protease, BepA
BSA	Bovine Serum Albumin
C12E8	Polyoxyethylene 8
C12E9	Polyoxyethylene 9
CD	Circular Dichroism
CHARMM-GUI	Chemistry at Harvard Macromolecular Mechanics - Graphics User Interface
CHS	Cholesterol hemisuccinate
CIP	Calf Intestinal Phosphatase
Cj1289	*C. jejuni* SurA-like chaperone, Cj1289
CjBamA	*C. jejuni* β-barrel assembly machinery complex factor A
CjBamD	*C. jejuni* β-barrel assembly machinery complex factor D
CjDegP	*C. jejuni* Periplasmic serine endoprotease P, DegP
CjLptD	*C. jejuni* LPS-assembly protein D, LptD
CMC	Critical Micelle Concentration
Cryo-TEM	Cryogenic Transmission Electron Microscopy
C-score	Confidence score
CVs	Column Volumes
Cymal-5	5-Cyclohexyl-1-Pentyl-β-D-Maltoside
DCD	DesignCAD (file)
DDM	n-Dodecyl-β-D-Maltoside
DegP	Periplasmic serine endoprotease P, DegP
DM	N-Decyl-β-D-Maltopyranoside
D_{max}	Maximum particle diameter
DMPC	Phosphatidylcholine (14:0/ 14:0)
DMSO	Dimethylsulfoxide

DNA	Deoxyribonucleic acid
DOPC	Phosphatidylcholine (18:1/ 18:1)
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
EtOH	Ethanol
FC12	Fos-choline 12
Fha	Filamentous haemagglutinin adhesin
FPLC	Fast Protein Liquid Chromatography
GLB	Gel Loading Buffer
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	High Performance Liquid Chromatography
IM	Inner Membrane
IMS	Industrialised methylated spirits
IMAC	Immobilised Metal Anion Chromatography
IMP	Inner Membrane Protein
IPTG	Isopropyl β-$_D$-1-thiogallactopyranoside
LB	Luria Bertani
LC-MS/MS	Liquid Chromatography Tandem Mass Spectrometry
LDAO	Lauryldimethylamine-N-oxide
LMNG	Lauryl Maltose Neopentyl Glycol
LptD	Lipopolysaccharide-assembly protein factor D, LptD
M/W	Molecular Weight
MD	Molecular Dynamics
MES	2-(N-morpholino)ethanesulfonic acid
MWCO	Molecular Weight Cut Off
NAMD	Nanoscale Molecular Dynamics
NEB	New England Biolabs
Ni-NTA	Nickel-Nitrilotriacetic acid
OD	Optical Density
OGNG	n-Octyl-β-D-Glucopyranoside/ n-Nonyl-β-D-Glucopyranoside
OM	Outer Membrane
OMP	Outer Membrane Protein

PAL	Peptidoglycan-Associated Lipoprotein
PBC	Periodic Boundary Conditions
PBS	Phosphate Buffered Saline
PC	Phosphatidylcholine
PCR	Polymerase Chain Reaction
PE	Phosphoethanolamine
PES	Polyethersulfone
PG	Phosphoglycerol
pI	Isoelectric point
PIPES	1,4-Piperazinediethanesulfonic acid
POPE	1-Palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine
POPG	2-Oleoyl-1-palmitoyl-sn-glycero-3-phospho-rac-(1-glycerol)
POTRA	Polypeptide Transport Associated (domain)
PPM	Positioning of protein in membranes (webserver)
rIMAC	Reverse Immobilised Metal Anion Chromatography
R_g	Radius of gyration
RMSD	Root mean square deviation
RMSF	Root mean square fluctuation
rpm	revolutions per minute
SAXS	Small Angle X-ray Scattering
SB12	Sulfobetaine-12
SDS	Sodium Dodecyl Sulfate
SDS-PAGE	Sodium Dodecyl Sulfate - Polyacrylamide Gel Electrophoresis
Sec	Secretory (translocase)
SEC	Size Exclusion Chromatography
SSA	Secondary structure assignment
SSE	Secondary structure estimation
TAE	Tris Acetate EDTA
TALON	Carboxymethylated aspartic acid (trade name TALON)
TBS	Tris Buffered Saline
TBST	Tris buffered saline plus tween20
TE	Tris-EDTA

TEMED	N,N,N′,N′-Tetramethylethylenediamine
TEV	Tobacco Etch Virus
T_m	Melting temperature
TPR	TetratricoPeptide Repeat(s)
UV	Ultraviolet (light)
ver.	Version
VMD	Visual Molecular Dynamics
WB	Western Blot

Amino acids

Amino acid	**Three letter code**	**Single-letter code**
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartate	Asp	D
Cysteine	Cys	C
Glutamate	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

1 Introduction

Gram-negative bacteria are enclosed by two membranes, an inner cytoplasmic membrane (IM), also common to Gram-positive bacteria, and an outer membrane (OM) that faces the environment, separated by the periplasm. The IM is composed of a phospholipid bilayer, forms a permeability barrier between the inside and outside of the cell, and carries out the majority of membrane-associated metabolic functions via membrane associated and embedded integral membrane proteins . The OM is asymmetric, consisting of phospholipids on the inner leaflet and lipopolysaccharides on the outer leaflet of the membrane, respectively, and acts as a protective barrier to the extracellular environment [1–3]. Additional to forming a protective barrier around the cell, the OM also contains a number of embedded proteins, making up about 50% of the OM by mass, that perform a number of functions such as nutrient uptake, signal transduction, and protein translocation [1–5]. Depending on their interaction with the OM, these proteins can be split into three groups: lipoproteins that are attached to the OM via a lipidated N-terminal cysteine, porins and export channels, and outer membrane proteins that contain a characteristic transmembrane β-barrel domain [1, 3, 6].

How these outer membrane proteins (OMPs) were inserted into the outer membrane was unknown until recently, with the assembly and insertion of the majority of OMPs into the OM of Gram-negative bacteria now attributed to the β-Barrel Assembly Machinery (BAM) complex. The work conducted thus far in Bacteria on this enigmatic system has been carried out in *Escherichia coli*, *Haemophilus ducreyi*, *Neisseria gonorrhoea* and *Rhodothermus marinus*, which are important organisms for food poisoning, sexual health, and industrial biotechnology, respectively [6–10]. However, there are no published structures available on the most commonly reported food-borne gastrointestinal pathogen in the US and European Union, *C. jejuni* [11, 12], which is the main focus of this body of work.

1.1 Campylobacteriosis

Campylobacter are Gram-negative, spiral shaped, microaerophilic commensal organisms present in the gut flora of wild birds and mammals, and are responsible for campylobacteriosis in humans. Campylobacteriosis is the most common food

borne infection that affects developed countries, and costs an estimated €2.4 billion in the EU and $1.8 billion in the US in lost productivity each year [11–14]. Poultry are the natural reservoir for Campylobacter species, and are estimated to be responsible for 80% of human campylobacteriosis cases, with two species *C. jejuni* and *Campylobacter coli* being responsible in around 95% of those cases [11, 15]. The majority of campylobacteriosis cases result in mild gastroenteritis, fever, headaches, watery or bloody diarrhoea, and vomiting, which usually resolve within 10 days. However, around 1-2 % of cases are more severe and require hospitalisation and usage of macrolide, tetracycline or fluoroquinolone antibiotics, for which there has been increasing reports of resistance in *C. jejuni* and *C. coli* [16, 17]. In rare instances, Campylobacter infection can lead to more serious autoimmune conditions, such as Guillain-Barré syndrome and Miller-Fisher syndrome, that result in nerve damage and may ultimately lead to paralysis [12, 14]. With the increase in antibiotic resistance to first line treatments becoming more prevalent, finding new therapeutic targets for the development of novel antimicrobials is becoming increasingly important.

1.2 OMP transport story

While integral IM proteins span the membrane via several hydrophobic transmembrane α-helical secondary structures , the β-barrel, composed of amphipathic anti-parallel β-sheets, predominates integral OMPs [4, 5]. Despite these differences, both classes of protein are synthesised on a ribosome, exhibiting some degree of co-translational folding in the cytoplasm, as precursors with N-terminal signal sequences that target them to the IM-embedded secretory (Sec) translocon for transport into or across the inner membrane, respectively [18, 19]. The precursors are released by the ribosome and bound by the cytoplasmic chaperones SecA and/or SecB in *E. coli*, which prevent further folding and block potential sites of aggregation, and deliver the precursor to the SecYEG complex for translocation across or into the inner membrane [19–21].

During translocation the distinction is made between inner membrane proteins (IMPs), which are released laterally via the lumen of the translocation machinery directly into the lipid bilayer of the inner membrane, and OMPs which are translocated across the membrane into the periplasm, where their signal sequence is

cleaved by a signal peptidase [19, 21–23]. It is proposed that the high abundance of hydrophobic alpha-helical content present in inner membrane proteins triggers a lateral release into the inner membrane during translocation across SecYEG; a process that presumably is not triggered by the mostly beta-sheet content present in OMPs and probably why α-helical structures are generally absent in OMPs [19, 22, 24]. Once in the periplasm the partially folded OMPs are picked up by the periplasmic chaperone network, which in *E. coli*, consists of the proline isomerase SurA, the prefoldin-like chaperone Skp, and the serine endoprotease DegP [23, 25–27]. The OMP is then delivered to the BAM complex for BAM-assisted insertion into the outer membrane (figure 1.1). How the BAM complex assembles these OMP substrates into the OM, a membrane devoid of proton motive force and ATP energy sources, has remained a mystery [28, 29].

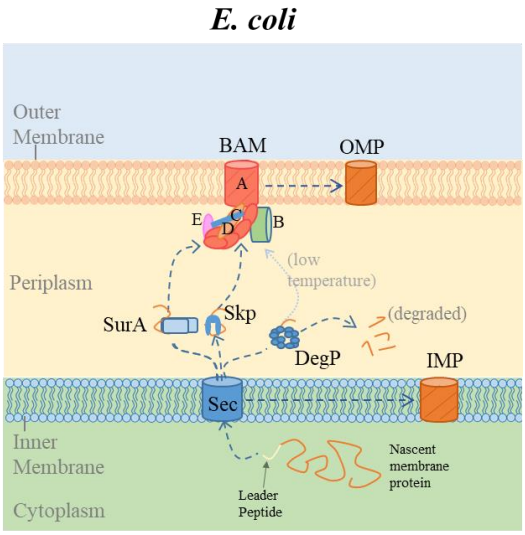


Figure 1.1: *E. coli* BAM-mediated outer membrane transport pathway. An unfolded nascent membrane protein is recognised by the Sec translocase *via* the leader peptide. If the protein is an IMP then the protein is folded and assembled into the IM. If the protein is an OMP, then the protein is picked up by either SurA or Skp and transported to the BAM complex for folding and assembly into the OM, or, if it is unfolded, it is picked up by DegP and degraded.

1.3 **The BAM Complex**

The first member of the BAM complex to be identified was Omp85 in *Neisseria meningitis*, where *omp85* was shown to be essential through gene deletion experiments in which no viable deletion mutant was obtainable, and through gene knockdowns experiments where subsequent protein depletion resulted in the build-up of unassembled OMPs in the periplasm [28]. Following this finding, an Omp85 homolog, named YaeT after the corresponding gene in *E. coli*, was found to be essential for OMP biogenesis [30, 31]. YaeT formed the core subunit of a five-subunit complex, which in *E. coli* is composed of YaeT, YfgL, NlpB, YfiO and SmpA proteins [29, 30]. These subunits were later renamed after the BAM complex, as BamA (YaeT), BamB (YfgL), BamC (NlpB), BamD (YfiO) and BamE (SmpA). Homologues of Omp85/ BamA were subsequently found in a number of Gram-negative bacteria, mitochondria (Tob55/Sam50) and chloroplasts (Toc75), suggesting widespread evolutionary conservation through the bacterial and eukaryotic domains of life, and its importance in the assembly of β-barrel proteins [30, 32–36].

There are now several crystal structures of the complete BAM complex available in *E. coli* and homologues in *N. meningitis*, *H. ducreyi* and *R. Marinus* [6–10, 37], and structural details on how subunits interact are now broadly known. The complete complex is approximately 200 kDa in size in these organisms, but evolutionary studies show that the number of subunits, the subunit size, and even variation in the subunits themselves that comprise the BAM complex can exist [36, 38]. While there can be wide variation in the primary structure of subunits, several stretches of sequence and the secondary structures are broadly conserved between organisms, suggesting key features to function. Mechanisms for assembly have been proposed around these features, and the requirements for different subunits have been investigated. Two subunits, BamA and BamD, are essential in most Gram-negative bacteria, with BamB, BamC and BamE acting as accessory components [26, 30–32].

1.3.1 Essential conserved components

1.3.1.1 BamA

The β-barrel membrane protein BamA, formerly known as YaeT in *E. coli*, is an 810 residue 90.6 kDa integral membrane protein that forms the central component of the BAM complex. BamA is highly conserved, with homologues found in mitochondria and chloroplasts, in Gram-negative bacteria, and in the thermophile *Thermus Thermophilus*. Notable exceptions include Gram-negative Fermicutes and Actinobacteria, and Archaea, where no homologues have been found [29, 33, 34, 39–43]. In *E. coli*, *N. meningitis*, and *H. ducreyi* BamA (or Omp85/ D15 respectively) is made up of five periplasmic N-terminal polypeptide transport associated (POTRA) domains and a membrane embedded C-terminal β-barrel domain composed of sixteen anti-parallel amphipathic β-sheets joined by periplasmic and extracellular loops [37, 44–47].

The key structural features of BamA (figure 1.2) are the POTRA domains, which are proposed to be involved in substrate recognition, and the junction between the β1-β16 strands, which is thought to open to allow insertion of incoming (substrate) proteins into the membrane. In addition, the extracellular loops and the conserved VRGF/Y motif have been shown to be essential, and are hypothesised to form part of an intricate insertion and assembly mechanism [46, 48, 49].

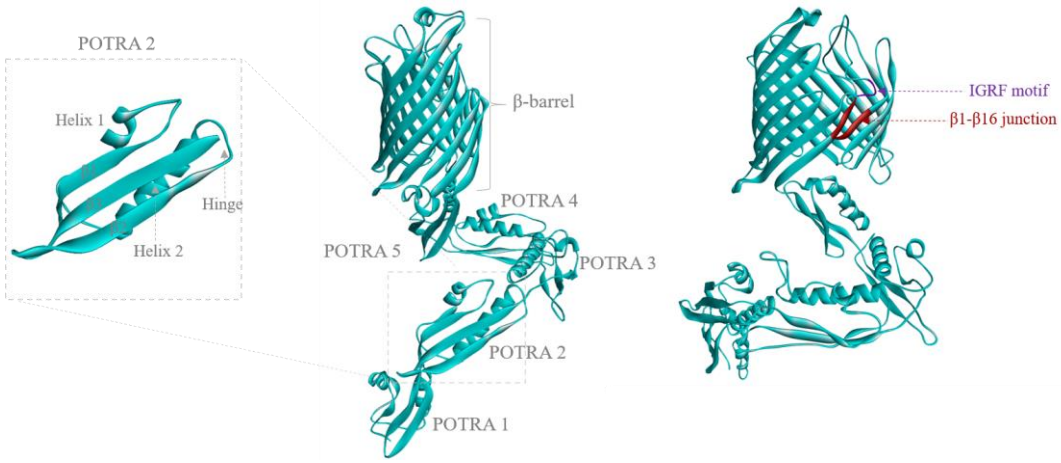


Figure 1.2: The structural features of BamA. Homology model of *C. jejuni* BamA. Protein domain structure, secondary structure and key features (the β1- β16 junction [red] and IGRF motif [purple]) central to BamA's function in folding and assembly are annotated.

1.3.2 **POTRA domains**

The POTRA domain is a conserved element among the β-barrel transport machinery. Though the sequence of each POTRA is in fact variable, there is a characteristic β-α-α-β-β fold structure, with conserved residues found in the hydrophobic core and loop structural elements of the motif [50–52]. Initial studies in the yeast homologue Sam50 indicated a role for POTRA domain in unfolded OMP substrate recognition [35], and nuclear magnetic resonance (NMR) spectroscopy data have demonstrated that several POTRA domains can form a surface using their β-sheets to accommodate an incoming peptide [50]. Truncations of the chloroplast homologue, Toc75, have suggested its POTRA domains have a role as docking sites for proteins targeted to, or across, membranes [53].

Deletion mutants in *E. coli* revealed loss of POTRA 2 is the best tolerated with little effect on growth rate and membrane OMP abundance; whereas deletion of POTRA 1 results in poor growth and a reduction in OMPs in the membrane [54]. Deletions of POTRA domain 3 and 4 are lethal unless a wild-type copy of BamA is introduced on an external gene element; and deletion of POTRA 5 is lethal even when the cell is supplemented with an encoded wild-type copy [55]. In contrast, deletion of the only POTRA of mitochondrial Sam50, an Omp85 homologue, in *S. cerevisiae* had no effect on cell growth or OMP assembly [43], and deletions of all but POTRA 5, the POTRA nearest the β-barrel, in *N. meningitidis* are tolerated [28]; with the other subunits postulated to help mediate folding of larger substrate OMPs. It is possible that this discrepancy may be related to differences in the distance between the outer and inner membranes in each of these systems, as the POTRA domain have been proposed to scan the IM for OMPs [51].

Superimposition of POTRA domains one and two in *E. coli* reveal the same basic fold, but with slight differences. POTRA domain 1 is slightly shorter, with a smaller α1 helix which is about half the length of the same helix in POTRA domain 2, with a slight (13°) change in orientation [50] relative to the second α-helix domain. The interdomain angles also differ between structures depending on the technique used, with crystal structures showing the POTRA domains as a ‘J’ or ‘fishhook’-like structure [55] and solution structures showing a more ‘Z’ like conformation [50]. The

main points of flexibility are the hinge region between POTRA domains 2 and 3, where there is a stretch of peptides acting as a linker sequence, and between POTRA domains 4 and 5, with other links between POTRA domains being comparatively rigid [52]. However, the extent of interdomain flexibility is debated, with different techniques giving different results. Several independent crystal structures show similar positioning and angles of POTRA domains between structures, and a POTRA 2-3 interdomain flexibility of angles between 47° and 106°. In contrast, molecular dynamics (MD) simulations and NMR show a greater degree of flexibility and freedom of movement in all but the POTRA4-5 joint, which is roughly in agreement with crystal structures [50–52, 55, 56]. Knowledge of the exact flexibility and conformational arrangement of these POTRA domains is important as one of the proposed functions of the BAM complex is to work as a 'disruptase', whereby localised disorder is generated by the POTRA domain in the membrane to aid OMP insertion [57].

MD simulations, on a roughly 0.5 μs scale, of BamA and the associated POTRA domains have raised the possibility of potential interactions of POTRA domains 1-3 with the inner leaflet of the OM [56]. This potential interaction was supported by docking of BamD and BamB crystal structures after membrane interaction had occurred and stabilised; though, whether this interaction will still occur after other subunits are added to the simulation before membrane interaction is established, was not explored *via* MD. The same paper [56] also mentions a lack of lateral opening over the time course of their simulations (540 ns). This is not surprising based on current knowledge of the lateral opening mechanism being dependent on incoming substrate β-strands displacing the hydrogen bonding at the β1-β6 junction. However, it does suggest that lateral opening of the β1-β16 junction is not simply triggered by gross movement of the POTRA domains towards the membrane.

Further information on the potential roles of the POTRA domains are inferred from crystal structures that show the POTRA domains blocking the periplasmic opening of the BamA β-barrel and the extracellular loops forming a closed dome, thereby preventing random solutes from traversing the OM. These suggestions arise from the crystal structures of *N. gonorrhoea* BamA where POTRA5 is directly below the β-

barrel, as opposed to the structure of *H. ducreyi* BamA where POTRA 4 and 5 have swung away from the β-barrel exposing the periplasmic facing pore opening [7].

1.3.3 Extracellular Loop 6

The loops connecting the anti-parallel β-sheets of the BamA β-barrel are exposed to the periplasmic space or extracellular milieu, and are another feature of interest in the function of BamA. Evidence suggests the loops exists in three confirmations, forming a dome at the extracellular surface, interacting with a conserved arginine residue within the pore, and potentially extending through the pore to protrude in the periplasm [7, 44, 45].

Studies into filamentous haemagglutinin adhesin (Fha) secretion by FhaC, an omp85 protein family member, in *Bordetella pertussis* identified a functionally relevant external loop, loop 6, that contained a conserved VRGF/Y motif at its tip [58]. Deletion of the loop or of the conserved sequence in FhaC resulted in reduced Fha secretion [45], indicating a role in assembly. In a similar experiments in *E. coli*, deletion of extracellular loop 6, or the conserved VRGF/Y motif contained within, resulted in a lethal phenotype [46]. Further mutational analysis of extracellular loop 6 in BamA revealed that the conserved RGF motif is crucial for OMP assembly, with arginine being most important, followed by phenylalanine and then glycine [46]. Deletion of RGF or replacement with AAA resulted in hypersensitivity to vancomycin and rifampin, implying increased membrane permeability and/or membrane defects, and reduced OMP levels, which were subsequently confirmed to be due to folding defects. The sensitivity, reduction in OMP levels, and folding defect severity differed by substitution, with AAA> RGF> AGA> AAF> RAA> EGF> RGA> AGF> RGF in order of severity. Substitutions with RGA and AGF show the least defects when compared to the wild-type RGF, suggesting that the presence of arginine or phenylalanine in this motif, but not necessarily both, is required for proper function [46].

1.3.4 β-barrel

Several crystal structures are now available for the BAM complex. These crystal structures suggest two key conformations that are thought to represent the externally ‘closed’ (resting/ acceptor) and externally 'open' (active/ insertion) states for

substrate OMP binding during BAM-assisted OMP biogenesis. In the externally 'closed' or acceptor state, the BamA β-barrel is closed to the extracellular environment but open to incoming OMPs on the periplasmic side of the OM. In the alternative conformation, the external loops, which guard the external face of the β-barrel from solutes and help to prevent random solute transport, are displaced from their resting conformation opening the β-barrel to the external environment [44, 48]. The 'open' state has only been observed in the absence of BamB, a 42 kDa non-essential β-propeller motif lipoprotein, which has led to the suggestion that BamB binding and dissociation acts as a switch whose presence regulates the confirmation and therefore the insertion process mediated by the central component BamA [59, 60]. However, this is disputed as a recent cryogenic transmission electron microscopy (Cryo-TEM) structure has been solved which shows the 'open' state with BamB present; with the author suggesting that BamB presence is unlinked to the external pore state (‘open’ or ‘closed’), and that the confirmations in the presence/ absence of BamB are potential artefacts of the crystallisation process[44]. The β-barrel forms a ‘closed’ structure via hydrogen bonding between the β1-β16 strands, which is absent in the ‘open’ structure [37, 51] (figure 1.3).

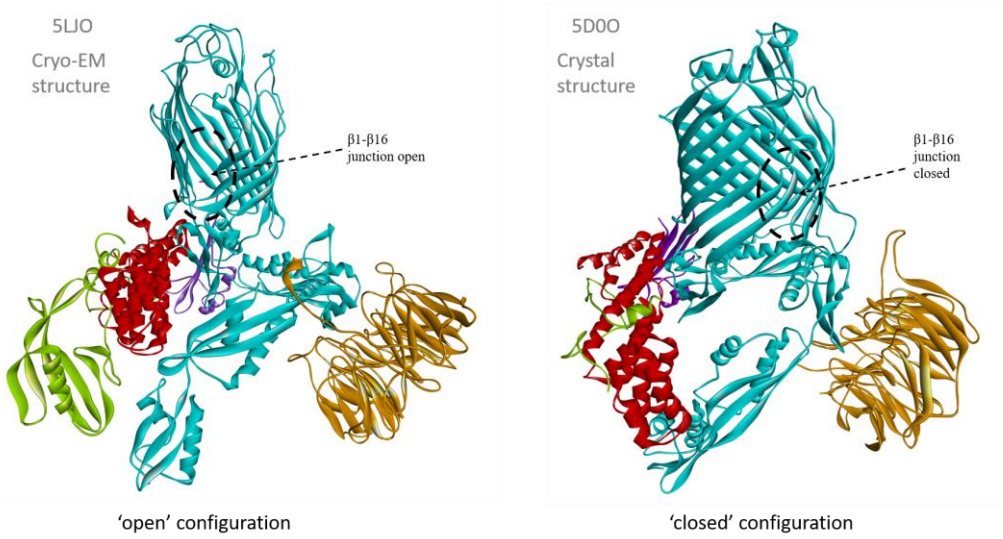


Figure 1.3: ‘Open’ and ‘Closed’ states of the BAM complex. The 5LJO Cryo-TEM structure of *E. coli* BamABCDE showing the ‘open’ state, where the interaction between the β1-β16 junction is disrupted (left), and the 5D0O crystal structure of *E. coli* BamABCDE showing the ‘closed’ state where the β1-β16 strands of the β-barrel are interacting. BamA (light blue), BamB (orange), BamC (green), BamD (red) and BamE (purple) are shown.

Interestingly, the number of hydrogen bonds that close the β-barrel varies by organism, with the β1-β16 junction being closed by eight hydrogen bonds in *H. ducreyi*, similar to the hydrogen bonding displayed between the other strands of the β-barrel, but only two hydrogen bonds in *N. gonorrhoea*. This data, and knowledge of lateral opening occurring in the Sec translocon, led several authors to suggest that this junction acts as an access point to the hydrophobic interior of the OM [19, 21, 37, 43, 44]. Strong evidence for lateral opening in the assembly mechanism comes from a loss of BAM function on the insertion of a synthetic disulphide crosslink between the β1-β16 strand junction, with several mutants showing no growth unless a strong reductant, tris(2-carboxyethyl)phosphine (TCEP), was added [61].

1.3.5 **BamD**

BamD, formerly known as YfiO in *E. coli*, is a 245 residue 27.8 kDa protein and is the only other essential subunit in the BAM complex in all Gram-negative organisms studied thus far, with the notable exception of *T. thermophilus* which only has an integral BamA subunit, which is essential and sufficient alone [30, 32, 36, 42]. The structure of BamD consists of 9-10 α-helices joined by short loops that are arranged in pairs into five repeating helix-turn-helix, or tetratricopeptide (TPR), motifs, which pack against each other (figure 1.4).

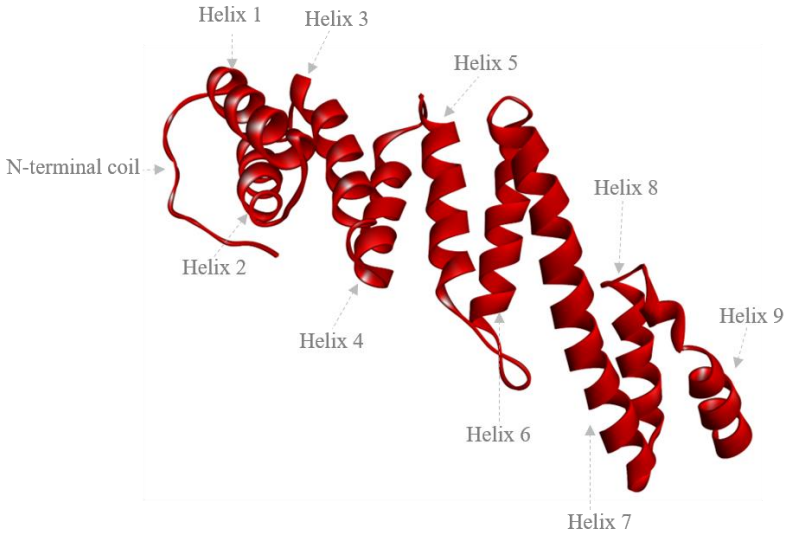


Figure 1.4: Structural features of *C. jejuni* BamD. Homology model of *C. jejuni* BamD. BamD is composed of nine α-helices, of which helix 7 is the largest and features a mid-helix kink, and a unordered N-terminal coil.

In *E. coli* each of these helices are between 13-16 residues in length, with the exception of α7 in TPR domain 4 which is 29 residues in length. BamD has the same overall structure elements in *R. marinus*, except the α-helices are between 10-20 residues in length, with α7 being 28 residues. This effects the width of the deep groove formed between TPR3 and TPR4 (34.2Å in *E. coli* and 24Å in *R. marinus*), which contains a stretch of conserved residues and is therefore thought to act as a protein-protein interaction surface [6, 37, 62]. The α7 helix serves to split the N- and C-terminal regions, with the C-terminal TPR4, TPR5 and groove between TPR3 and TPR4 predicted to form associations with BamA, BamC, and BamE, as shown by TPR deletion experiments. The N-terminal region (TPR1-3) is proposed to interact with incoming OMP substrates [37, 62]. Coinciding with this there are several conserved aromatic residues contained within the first three TPRs, which are conserved in the signature sequence motif discussed later in this review [63–66].

The number of BAM subunits was estimated in a large mRNA and mass spectrometry peptide fingerprinting proteomics assay in *E. coli*, with approximate cell copy numbers of 200 units of BamA, 140 units of BamB, 180 units of BamC, 165 units of BamD and 187 units of BamE predicted to be present [67]. This coincides with suggestions that the stoichiometry of the BAM complex is roughly a 1:1:1:1:1 ratio of all BamA-E subunits [27, 44, 67]. Though the subunits may be predicted to be roughly equal in number, they are not all essential. Only deletion of BamA and BamD are lethal *in vivo*, and knockdown experiments reveal that decreasing the concentration of either of these essential BAM subunits in *E. coli* have the same negative, but not additive, effect on biogenesis [27, 31, 32, 62]. However, the reduction of the accessory components BamB, BamC and BamE produces a less severe and to varying degrees, with deletion of BamB producing a synthetically lethal phenotype and BamC and BamE deletions producing less severe outcomes [26, 30, 65, 68, 69].

1.3.6 **Accessory components**

On the periplasmic side of the OM, BamA is associated with four proteins, the essential lipoprotein BamD, and the ‘accessory’ proteins BamB, BamC, and BamE. They are termed ‘accessory’ proteins as deletion of BamB, BamC, or BamE does not

cause lethality *in vivo* [26, 30, 41, 65, 68, 69]. However, BamB, BamC and BamE still have important roles as deletion reduces OMP abundance in the membrane, which has been shown to cause increased sensitivity to small toxic compounds such as vancomycin, and thereby implies the formation of structural defects in the membrane [70, 71].

1.3.6.1 **BamB**

BamB is a 392 residue 41.9 kDa peripheral accessory lipoprotein that binds the BAM complex via direct interaction with the central BamA subunit [29, 32, 72]. This interaction was shown to be dependent on the presence of POTRA domains 3-5 at the N-terminus of BamA [59], with more recent crystal structures illustrating contact to occur between the β-propeller and β-sheet of POTRA 3 [8, 9]. Through these contacts the POTRA domain act as scaffolding for associations with accessory proteins, and their deletion prevents BamB association [55], and significantly reduces OMP assembly success.

Crystal structures of BamB reveal a WD40 repeat domain composed of an orthodox eight-bladed propeller structure, with blade eight composed of β-strands from the N and C-terminal residues [29, 40, 44]. Each blade consists of four anti-parallel amphipathic β-strands separated by short loop junctions, with the innermost β-strand connecting to the outermost of the next blade via a long loop, and the overall structure similar to a short cylinder. The loops facing the upper, membrane-facing, side of the short cylindrical structure are longer (5-19 residues) when compared to the loop protruding on the periplasmic side. Attachment to the periplasmic side of the OM is mediated via a N-terminal palmitate lipid attachment site located at the top of blade eight of the β-propeller structure [29, 40, 44].

1.3.6.2 **BamC**

The accessory subunit BamC is a 344 residue 35 kDa peripheral outer membrane protein lipidated with N-acyl-diacylglycerol at cysteine 25 to facilitate membrane attachment. It is composed of 3 domains, a disordered N-terminal domain (residues 25-100), a N-terminal helix grip domain (residues 101-210), and an equivalent C-terminal helix grip domain (residues 229-346), with the latter two domains separated by an 18 residue flexible linker [66, 73, 74]. The unstructured N-terminal domain contains

the majority of the conserved residues in BamC, and interacts extensively with BamD on its longitudinal axis on the opposite side to the groove made by TPR3 and TPR4 of BamD. Site-directed mutagenesis has revealed that residues 227-245 of BamD are necessary for this interaction [32, 65, 66]. The two helix grip domains are similar, with the N-terminal domain consisting of two α-helices and five anti-parallel β-sheets, and the C-terminus having an equivalent structure plus an additional β-sheet and a seven residue 3_{10} helix between strands β3 and β4 [65, 66, 73]. The flexible linker is susceptible to cleavage by subtilisin, breaking BamC into two fragments of 12.2 and 14.5 kDa respectively [74]. The proposed function of BamC is to stabilise the BamA-BamD interaction along with BamE, with the whole BamCDE sub-complex interacting with BamA through POTRA five [49, 55, 65, 66].

1.3.6.3 **BamE**

The crystal structure of BamE revealed a monomeric 113 residue 12.3 kDa peripheral membrane protein with an anti-parallel β-sheet packed against a pair of α-helices, or α-α-β-β-β-α fold, with secondary structure elements made up of residues 40-45, 52-58, 72-78, 88-95 and 100-106, respectively. NMR was used to study the effects of site-specific mutations on BamE localisation to the outer membrane, with a phenylalanine to alanine substitution at residue 74 in the predicted phosphatidylglycerol binding site, F74A, nullifying membrane binding function [41, 69]. However, BamE was determined not to be an essential subunit of the BAM complex, as deletion of BamE has limited or no effect on growth at 30°C, the proposed function of BamE is to stabilise the BamA-BamD interaction. This function has been determined through data that show decreased BamA-BamD interaction in Δ*bamE* backgrounds, and synthetic lethal phenotypes in Δ*bamE* Δ*bamA101* (knockdown) double mutants which are rescued by overexpression of BamD [26, 49, 69, 75].

1.3.7 **Other potentially associated and alternative subunits**

Bioinformatic analyses of the genomes of proteobacteria have revealed that BamB is well conserved in α-γ proteobacteria, but not δ- or ε-proteobacteria. BamC is conserved in β- and γ- but absent from α-, δ-, and ε-proteobacteria. BamE is well conserved in α-γ, but not δ- or ε-proteobacteria [36, 38]. This finding and the pulldown of an unexpected 17.5 kDa BAM complex-associated lipoprotein in the α-proteobacterium *Caulobacter crescentus*, later named BamF, have led Lithgow *et al.*

(2012) [38] to suggest that alternative subunits of the BAM complex may have evolved among the proteobacteria (figure 1.5). The authors suggest that BamF is loosely related to BamC, and that co-purification with BamD occurs in membrane preparations under solubilising conditions of 0.75% dodecylmaltoside.
Previous work by the same group also identified a 20 kDa peptidoglycan-associated protein, Peptidoglycan-associated lipoprotein (PAL), which has been suggested to assist in the transfer of OMPs from the Sec translocon to the BAM complex. PAL deletion mutants result in membrane defects, similar to loss of BamB, BamC, or BamE in *E. coli* [76], thus raising the possibility of a much larger coordinated complex between the IM and OM. PAL is related to other proteins YiaD (*E. coli*) and RmpM (*N. meningitidis*), which may serve a similar role *in vivo* in these organisms [76, 77].

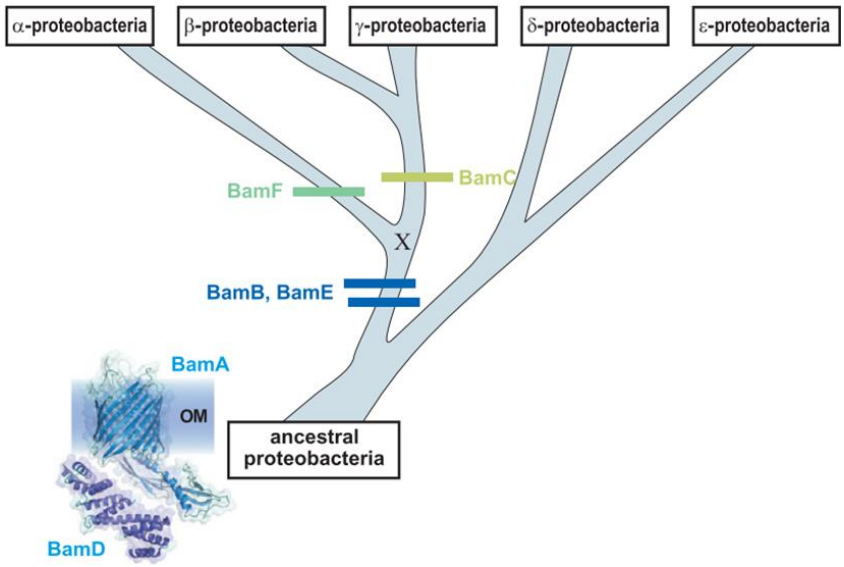


Figure 1.5: Evolutionary conservation of the BAM complex subunits. ε-proteobacteria, such as *C. jejuni*, have only the two essential components of the BAM complex. The same subunits are available for δ-proteobacteria. β- and γ-proteobacteria feature subunits BamA-E, and α-proteobacteria have BamF instead of BamC. (Taken from Anwari *et al.*, 2012[38])

1.4 Chaperones

Several chaperones have been identified in the periplasm that work together to guide nascent OMPs from the Sec translocon to the BAM complex for insertion. The main chaperones shown to be involved are the proline isomerases SurA and FkpA, the

prefoldin-like chaperone Skp and the protease DegP, with SurA and Skp being put forward as the most important [70, 78, 79]. An emerging chaperone, the β-Barrel Assembly-Enhancing Protease, BepA, has been shown to interact with BamA, though its exact role with SurA still present is still under investigation[80, 81].

1.4.1 **Skp**

Skp has been shown to bind unfolded OMPs in the periplasm while they are still engaged with the Sec translocon[82], and may assist in OMP release from the Sec machinery, resulting in a soluble periplasmic intermediate. Deletion of the *skp* gene resulted in reduced concentrations of OMPs in the outer membrane [70, 83–85].

1.4.2 **SurA**

SurA was first identified as an essential protein for the survival of *E. coli* in the stationary phase [86], but is not necessary for growth under normal conditions; though deletion of the *surA* gene under normal growth conditions results in reduced concentrations of OMPs in the OM [87, 88]. Peptide binding studies have indicated that periplasmic SurA preferentially binds to peptides containing aromatic (Ar) residues and an aromatic-x-aromatic sequence is typically found in the C-terminal sequences of Bacterial OMPs [83, 89]. It is not fully understood whether this activity is attributed to the N-terminal or peptidyl-prolyl isomerase domain, with recent evidence, in the form of 1.3 Å /3.4Å crystal complexes of SurA/ SurAΔP1 (parvulin domain 1) with highly aromatic peptides containing the signal sequence Ar-X-Ar-X-P, suggesting structural plasticity and adaptability of SurA depending on the motif recognised [88, 89].

There is debate on the role that Skp and SurA play in the process of transporting OMPs from Sec to the OM. Some authors hypothesise that the two proteins act in a sequential manner during transport, i.e. Skp initially binds the translocated OMP and then is displaced by SurA which shuttles the protein to the BAM complex in the OM [3, 28]. Alternatively, they may form part of parallel periplasmic chaperone pathways, and thus have a redundant interchangeable function. This view is supported by deletion studies which show viable cells with reduction, but not absence, of activity, as assessed by levels of OMPs in the OM, when either *skp* or *surA* are deleted [25, 70, 84, 87, 88]. The latter view is further supported by the knowledge that a *surA*/ *skp* or

surA/ *degP* double mutant is synthetically lethal [25]. Proponents of the sequential viewpoint suggest that this synthetic lethality can be explained by an increased requirement for the holding chaperone (Skp) when folding is comprised by deletion of a folding chaperone (SurA), and a greater efficiency of folding being required in the opposite scenario [3, 70]. This view is supported by information that suggests SurA likely interacts with the BAM complex and Skp likely interacts with the Sec translocon. [28]

1.4.3 **DegP**

DegP is a 474 residue 49.4 kDa periplasmic serine protease that acts as a chaperone at lower temperatures on the same pathway as Skp, and serves as a quality control factor at elevated temperatures degrading partially degraded, aggregated and misfolded OMPs in the periplasm [90–92]. In BAM complex subunit deletion backgrounds DegP knockouts are synthetically lethal [25, 27]. This synthetic lethality is thought to be due to the build-up of aggregated OMPs in the periplasm and the ensuing periplasmic stress response, in which DegP, and other proteins are up-regulated to degrade aggregated and misfolded OMPs, and small regulatory RNAs are produced to inhibit OMP translation [71, 90, 92–94].

1.4.4 **BepA**

Identified as one of the 114 genes in the *E. coli* σ^E regulon[95], BepA, is a 51 kDa periplasmic protein, composed of an N-terminal protease domain and a TPR domain, that shows dual chaperone and protease activity [96]. BepA has been observed to pulldown strongly with BamA and other BAM complex members[96]. The active site is under elucidation and mutational studies have proposed the active site to be formed by a zinc atom and a HEXXH motif in the protease domain [96]. Deletion studies of the TPR domain show diminished BepA function, suggesting the TPR domain is essential to function, and indicate that the TPR domain modulates interaction between BepA and BamA[97]. In addition, a full length double cysteine mutant of BepA has been shown to degrade BamA under SurA depletion conditions [81].

1.5 Mechanisms

1.5.1 Overview of current models

Four potential model mechanisms for BAM-mediated insertion of OMPs have been proposed. The ‘superpore’ model, hypothesises a concerted insertion mechanism where multimeric complexes of BamA work together to insert substrate OMPs. The ‘BamA-assisted’ or ‘disruptase’ model, proposes that BamA primes the outer membrane for spontaneous insertion of OMPs. The ‘export’, model entails usage of the BamA β-barrel pore to export substrate OMPs to the extracellular surface, where it then is inserted. The ‘budding’, model formulates the use of the BamA β-barrel as a frame on which substrate OMPs fold and form a hybrid β-barrel, from which the substrate β-barrel then buds off directly into the membrane [29, 41, 55].

The 'disruptase' and 'budding' (or the 'β-augmentation' model), are the two predominate models in light of the steric hindrances observed in the 'multimeric' model and the small pore size of the BamA β-barrel, with a consensus view that multiple mechanisms of action for OMP assembly by the BAM complex exist [5, 7, 48, 61, 98–100].

1.5.1.1 Disruptase

Outer membrane proteins are able to fold spontaneously into the membrane, albeit at a slower rate than when assisted. In the disruptase mechanism, the BAM complex is thought to lower the kinetic barrier to folding, assembly and insertion of OMPs into the OM [7, 56, 100].

The native lipid membrane in *E. coli* is composed of phosphoglycerol (PG) and phosphoethanolamine (PE) in an approximate 80:20 ratio, with small amounts of cardiolipin[101, 102]. The periplasmic lipid head groups of these lipids, in particular PE, are thought to act as a kinetic barrier hindering spontaneous insertion of OMPs into the membrane [7, 56, 100]. This concept is supported by experiments *in vitro* where introduction of host PE and/or PG lipids into a synthetic Phosphatidylcholine (PC) membrane, a lipid not normally found in the OM of *E. coli* that allows for efficient spontaneous folding of OMPs, reduced the rate of spontaneous insertion in a dose

dependent manner. Addition of a pre-folded BamA subunit into the synthetic membrane improved the efficiency in all combinations of PC and PG/ PE/ PE+PG tested suggesting the BAM complex accelerates the folding and assembly of OMPs into the outer membrane [100, 103].

Electron microscopy data adds further credence to this notion, showing that the membrane surrounding BamA β-barrel has an asymmetric hydrophobic thickness, with the membrane adjacent to the β1-β16 junction being between 12Å-16Å, and the opposite side between 24Å-25Å in thickness [7, 30, 56]. Further support comes from MD simulations of *N. meningitidis* BamA, where the lipids within 12Å of the C-terminal strand (where the membrane has a decreased thickness) showed a three-fold decrease in order when compared to the opposite side [7]; which would presumably reduce the kinetic barrier to insertion by allowing partially folded OMPs easier access to the membrane. There is also the possibility that the POTRA domain of BamA can interact with the membrane causing further destabilisation, as discussed earlier [56, 100, 103]. The BAM complex, in this mechanism, is therefore proposed to accelerate OMP insertion and assembly into the OM by reducing the kinetic barrier generated by the presence of native *E. coli* lipids by disturbing the packing of lipid head groups in an area proximal to the β1-β16 junction [56, 60, 100, 103].

1.5.1.2 **β-augmentation**

The substrate OMPs are delivered to the BAM complex for assembly and insertion into the OM. The substrate OMP is recognised by the BAM complex, possibly via the deep groove formed by TPR3 and TPR4 of BamD, the hydrophobic pocket or aromatic residues of BamB, or the POTRA domain of BamA, using the β-signal sequence at the C-terminal β-strand of the incoming OMP [29, 40, 51, 52, 62–65, 104]. The partially folded OMP is then guided to the BamA β-barrel opening on the periplasmic side, where it is then threaded section-by-section into the lumen of the β-barrel. A lateral opening event is triggered either by the presence of the substrate OMP in the lumen of the β-barrel or via concerted movement of the POTRA domain and extracellular loops [7, 45, 50, 55, 105]. The most C-terminal strand of the incoming OMP pairs with the exposed β1 strand of the BamA β-barrel, and forms a hybrid barrel one β-strand or β-hairpin at a time. Once enough of the strands have been

inserted into the hybrid barrel, and therefore incorporated into the membrane, the substrate OMP's β-barrel buds away from the BamA β-barrel and spontaneously closes using its final β-strand [29, 43, 44, 61]. The budding event has been proposed to be triggered by formation of the last β-strand of the substrate OMP β-barrel, which has a greater affinity for the first β-strand of the substrate β-barrel than the β-16 strand of the BamA β-barrel as a greater number of hydrogen bonds can be formed. The folding process is then complete and the membrane protein diffuses away from the site of insertion and assembly [7, 29, 48, 61] (figure 1.6).

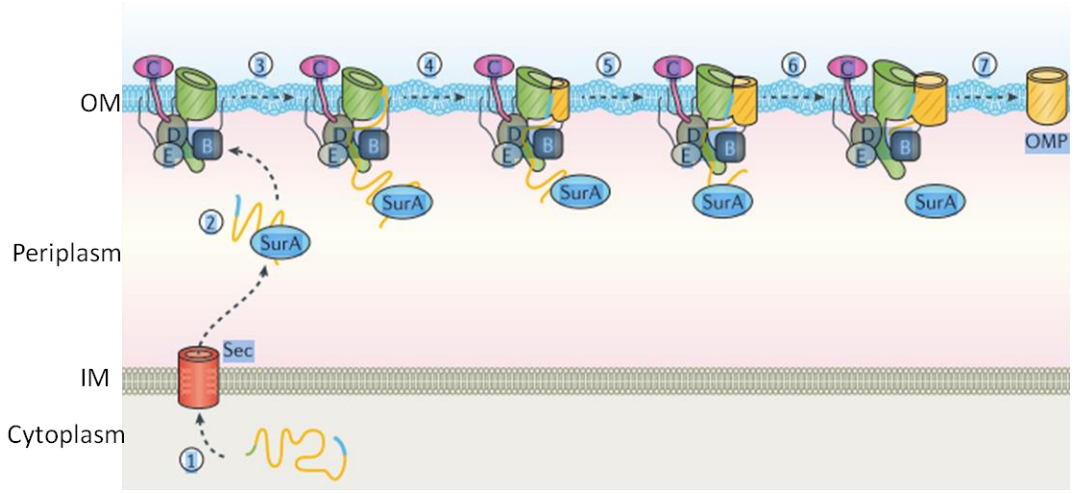


Figure 1.6: Overview of the hybrid barrel formation and budding mechanism. Overview of the hybrid barrel and budding mechanism. (1) The partially folded OMP is transported to the Sec translocon for transport across the IM. The leader sequence (green) is cleaved by a signal peptidase. (2) The OMP is bound by SurA (or Skp) in the periplasm. (3) The OMP is guided to the BAM complex for transport and the β-signal sequence (light blue) docks onto either the POTRA domains, BamB or BamD, and the process of assembly begins. (4, 5) The leading C-terminal β-sheet displaces the β-16 strand of BamA and forms hydrogen bonds with the β1 strand. One β-strand is assembled at a time and the hybrid barrel is extended. (6) The extended hybrid barrel nears completion, and the final β-strand of the substrate OMP β-barrel is close to the first strand of its native β-barrel. The budding process begins. (7) The assembly process is now complete. The substrate OMP 'buds off' and diffuses away from the BAM complex. (Taken from Noinaj *et al.*, 2017)

1.5.2 **Substrate recognition**

OMP substrate recognition is mediated by the presence of a sorting signal, more recently termed the 'β-signal', that exists on the C-terminal of incoming substrates to the BAM complex. The β-signal consists of an approximately 10 residue peptide that, to our knowledge, invariably has an aromatic residue, typically phenylalanine but occasionally tryptophan, at position 1 (most C-terminal), and hydrophobic residues at positions 3 (tyrosine favoured), 5,7 and 9 [28, 33, 63, 64]. In the mutational experiments conducted by Struyvé *et al.* (1991)[63], the phenylalanine (hydrophobicity score (h-score) of 1.19) was substituted for valine (h-score 1.08), tyrosine (h-score 0.26), and asparagine, with the tyrosine substitution being the best tolerated, and the valine substitution resulting in a much lower OMP uptake into the membrane. This work illustrated that it is far more important to have a residue of a similar type (i.e. aromatic) than of a similar hydrophobicity at position 1 of the β-signal sequence [63]. The asparagine substitution had a drastically negative effect on OMP uptake and insertion into the membrane by Omp85; suggesting position 1 is unlikely to be a hydrophilic amino acid. Further investigation by de Cock *et al.* (1997) showed that the deletion of phenylalanine at position 1 outright results in blocked or incomplete assembly and accumulation of the substrate OMP in the periplasm *in vivo* [104]. The evidence supporting the importance of the C-terminal signal sequence was extended by measuring channel conductance across the membrane with and without a substrate OMP that had the C-terminal signature motif, and by use of a peptide fragment corresponding to this motif. The experiments showed increased channel conductance in a dose dependent manner with increased substrate OMP, and that an increase in channel conductance was triggered by the peptide fragment alone, implying that the C-terminal sequence alone is sufficient to activate the insertion machinery mechanism[64].

Investigation into C-terminal signal (or 'β-signal') sequences highlighted the presence of characteristic residues that either enable or block recognition by the BAM complex in a species dependent manner. Species specificity was demonstrated in cross-species BAM mediated membrane insertion experiments where PorA from *N. meningitidis* was unable to be inserted and assembled by BamA of *E. coli* until the signal sequence was altered to match that of *E. coli* OM proteins. The majority of

N. meningitidis proteins were found to contain a positive residue, arginine or lysine, at position two and a histidine at position three of the β-signal contrasting with glutamine at position two and tyrosine at position three in *E. coli* [63, 64, 104]. While the histidine was determined not to be a widespread feature of Neisserial proteins, the substitution of the positive residue at position two for glutamine enabled cross-species insertion of *N. meningitidis* PorA by BamA of *E. coli*[64]. In addition, substitution of the glutamine at position two of PhoE (an *E. coli* OMP) to lysine prevented insertion by BamA of *E. coli* [64]. Although anecdotal exceptions exist where substitution was not required [106], overall it is accepted that the 'β-signal' is species specific and that manipulation of the β-signal may be required for the insertion and assembly success of non-native proteins by the BAM complex.

There has been limited progress on this work to date, with additional coming from co-crystallisation binding experiments with SurA and BamA-D binding inhibition via peptides derived from BamA. The co-crystallisation experiments with signal peptides and SurA [88] illustrate that the signal sequence can be varied on the current theme as outlined by Struyvé *et al.* (1991) [63], Ar-X-Ar-X-Z-X-Z-X-Z-X (where Z represents a hydrophobic residue), to include additional aromatic residues in the patterns. With the preferred patterns being Ar-X-Ar and Ar-Ar, extensions on these short motifs in the form Ar-X-Ar-X-Ar and Ar-Ar-Ar-Ar, and longer patterns where a pair of aromatic residues is separated from a third by five random residues, Ar-Ar-X5-Ar and Ar-X5-Ar-X-Ar [88]. In the BamA-D binding inhibition experiments, Hagan *et al.* (2015) [107], identified a peptide sequence derived from the C-terminal residues 765-779 of BamA, SAGIALQW, that inhibited folding of BamA into proteoliposomes in cases where the complete BAM complex or BamD had already been incorporated. This peptide also inhibited transport of OmpA by BamA in a similar experiment. The SAGIALQW peptide sequence is proposed to interfere with BamD binding, thereby acting as a competitive inhibitor of BamA-D association and of OMP assembly [107]. The sequence was found to have the same inhibitory effect *in vivo*, increasing sensitivity to vancomycin and rifampin and increasing expression of the DegP, which is indicative of membrane defects (due to reduced OMP assembly) and the periplasmic stress response, respectively [70, 71, 91, 92, 107]. This work indicates that mimicry of the signal sequence is a viable strategy in designing anti-microbials.

1.6 Current progress in the *C. jejuni* BAM complex

In comparison to *E. coli*, *H. ducreyi*, *N. meningitidis* and *R. marinus*, very little work on the BAM complex has been carried out in any species of Campylobacter. Evolutionary studies of the BAM complex show that species of Campylobacter are predicted to have only the BamA and BamD subunits [36, 76, 108, 109]. Potential gene homologs were previously identified in *C. jejuni* for BamA (*cj0129*), BamD (*cj0174c*), and four candidates that could potentially act as chaperones *cj0694* (putative PpiD homologue), *cj1289* (putative SurA-like chaperone/ SalC homologue), *cj0596* (putative PEB4 homologue), and *cj1228* (putative DegP homologue) [108, 109]. The pathway of the *C. jejuni* BAM-mediated OM pathway is similar to that in *E. coli*, however it represents a simplified system as Skp and accessory BAM subunits are absent and SurA is replaced with Cj1289 (figure 1.7).

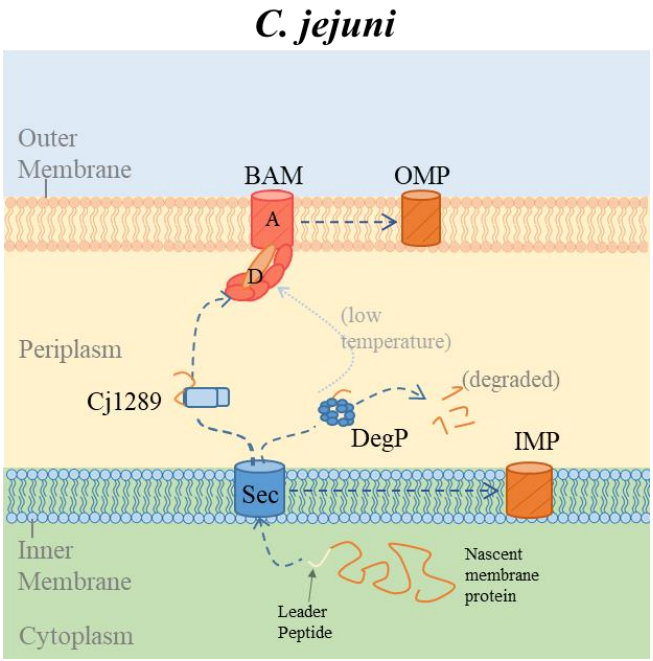


Figure 1.7: *C. jejuni* BAM-mediated outer membrane transport pathway. Similar to the *E. coli* BAM mediated OM transport pathway, SecYEG is responsible for the folding and assembly of IMPs. The situation changes when the protein arrives in the periplasm as there is only one main chaperone to shuttle it to the BAM complex, Cj1289.

Previous work by Kale *et al.* (2011) determined the crystal structure for Cj1289[108] prior to the start of this project, and Zarzecka *et al.* (2020)[110] determined the structure of *C. jejuni* DegP.

Operons are groups of clustered genes that are co-transcribed and are a central feature of bacterial gene regulation[111]. A later investigation into the *bamA* and *bamD* operons (this work) identified several genes for histidine synthesis (hisN), fatty acid metabolism (accD), transcription (*rlmH*), transcription regulation (*dksA*, *dus*), protease activity (*lon*), and an unidentified putative membrane protein (*NCTC11168_00123*). The unidentified putative membrane protein was homology modelled using AlphaFold 2[112, 113], and a transmembrane protein receptor composed of an N-terminal alpha-helical hairpin and C-terminal TPR domain was identified (appendix 8.1.1, figure 8.1). The potential role of this protein in the BAM-mediated OMP biogenesis pathway in *C. jejuni* is unknown and warrants further investigation in future projects.

1.7 Thesis aims

The objective of this project is to characterise the β-barrel assembly machinery complex in *C. jejuni* using biophysical and computational techniques, informing on the structure, dynamics, and interactions in the BAM-mediated OMP biogenesis pathway to gain insight into functional aspects of this key complex. As the BAM complex is an essential and conserved machinery necessary for cell functionality and survival, it presents as a key target for antimicrobial therapies, and the outcomes of this project may have future relevance to treatments of the autoimmune syndrome Guillain-Barré that presents in rare cases of infection with the pathogen *C. jejuni*.

1.7.1 Specific chapter aims

In chapter 3, we aim to describe the isolation of the genes of interest from *C. jejuni*, construct formation, and the progressive optimisation of the protein expression, protein purification, detergent screening, and recovery of these proteins in order to enable further biophysical characterisation.

In chapter 4, we aim to analyse the isolated *C. jejuni* proteins of interest using circular dichroism (CD) to characterise the secondary structures and thermal degradation melting temperatures of the proteins individually and in mixture, and compare these to the known values present in the literature to assess the folding, stability and any interactions of these recombinantly expressed proteins. Further analysis will be applied using the recently developed Landau free energy models to determine the transition cooperativities of these proteins and their mixtures as a validation of any suspected interaction, and this work represents one of the first known data sets to be analysed with this technique.

In chapter 5, we aim to utilise small angle X-ray scattering (SAXS) to generate dummy atom model structures of the proteins in this study, and further probe for protein-protein interactions and aim to validate any interactions suspected from the previous chapter.

In Chapter 6 we aim to use an *in silico* approach to determine the docking interface between *C. jejuni* BamA and BamD. Homology models for *C. jejuni* BamA and BamD will be generated and a rigid docking-based approach will be applied to determine an appropriate docking configuration. If multiple configurations are found, we aim to resolve these using MD simulations.

2 Materials and methods

2.1 Materials

2.1.1 Reagents

All chemicals and culture media used in this study were purchased from Sigma-Aldrich or Fisher/ Fisher Scientific unless stated otherwise. All primers utilised in cloning in this study were purchased from Sigma-Aldrich. All plastic disposables used in this study were purchased from Starstedt unless stated otherwise. Deionised water was used to prepare all general solutions, MilliQ (18 mΩ) water was used in all molecular cloning work, and all solutions used in biophysical work were filter sterilised through a 0.22 µm polyethersulfone (PES) syringe filter (Sartorius).

2.1.2 Sterilisation

All pipette tips, glassware, growth media, buffers, and bacteriological equipment were purchased either pre-sterilised or autoclaved at 121°C for 15 minutes on a standard media cycle. All heat-labile reagents were filter sterilised through 0.22 µm syringe filters in aseptic conditions. All bacteriological waste and chemical waste were disposed as per internal School of Life Sciences safety regulations. All post-culture solutions were disposed of after a minimum one hour in suspension in a 1 % Distel™ solution. Where hazardous culture media was used (e.g. blood), the solutions were left in suspension in a 2 % Distel™ solution for a minimum of 3 hours before disposal. All surfaces that may have been in contact with bacteriological solutions were cleaned down with either 70 % ethanol, 70 % industrialised methylated spirits (IMS) or 1 % Distel™. All bacteriological work was carried out in aseptic conditions unless stated otherwise.

2.1.3 Media

Media used in this study for bacterial growth were prepared as in table 2.1. The pre-mixed powders or individual components (where pre-mixed powder was not available) were weighed dry before addition of dH_2O, and the pH was checked and adjusted (as necessary) prior to autoclaving.

Medium	Composition
LB Broth	10 g/L peptone, 5 g/L NaCl, 5 g/L yeast extract
LB Agar	10 g/L peptone, 5 g/L yeast extract, 15 g/L agarose
Columbia agar	23 g/L special nutrient substrate, 1 g/L starch, 5 g/L NaCl, 15 g/L agar
Thioglycolate broth	L-cysteine 0.5 g/L, NaCl 2.5 g/L, 5.5 g/L glucose, 5.0 g/L yeast extract, 15 g/L pancreatic digest of casein, 0.5 g/L sodium thioglycolate
Terrific broth	12 g/L Tryptone, 24 g/L yeast extract, 9.4 g/L K_2HPO_4, 2.2 g/L KH_2PO_4
SOC medium	3.6 g/L Dextrose, 0.186 g/L KCl, 4.8 g/L $MgSO_4$, 20 g/L tryptone, 5 g/L yeast extract

Table 2.1: Bacteriological media used in this study

2.1.4 Antibiotics

Antibiotic stock solutions were prepared by dissolving antibiotic salt in solvent followed by filter sterilisation (table 2.2). Stocks were stored in aliquots at - 20°C. Stocks were thawed and added to media up to the working concentrations.

Antibiotics	Stock solution (1000x)	Mixture composition (if applicable)	Final working concentration
Ampicillin	100 mg/ml in dH_2O	-	100 µg/ml
Kanamycin	50 mg/ml in dH_2O	-	50 µg/ml
Chloramphenicol	34 mg/ml in ethanol (95 %)	-	34 µg/ml
Campylobacter Preston selective supplement	-	Values per vial: Polymyxin B 2500 U, Rifampicin 5 mg, Trimethoprim 5 mg, Cycloheximide 50 mg	1 vial/500 ml

Table 2.2: Antibiotics used in this study

2.1.5 Buffers and stains

The compositions of the buffers and solutions used in this study are shown in table 2.3. All buffers made up to desired volume with dH_2O.

Buffer or stain	Composition	Used for
Buffer		
TBS	150 mM NaCl, 50 mM Tris pH 7.6	Protein purification
PBS	137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4 pH 7.2	Protein purification
HEPES	115 mM NaCl, 1.2 mM $CaCl_2$, 1.2 mM $MgCl_2$, 2.4 mM K_2HPO_4, 20 mM HEPES pH 7.0	Protein purification
MES	30 mM MES (free acid) pH 6.7	Protein purification
TE	10 mM Tris, 1 mM EDTA	Primer storage
Detergent	150 mM NaF, 30 mM NaF pH 7.6	CD experiments
Non-detergent	0.5 % SB12, 150 mM NaF, 30 mM Tris pH 7.6	CD experiments
LC-MS/MS	50 mM TCEP in 10 % methanol	Band storage for LC-MS/MS
Gel loading	150 mM Tris, 8 M urea, 2.5 % SDS (w/v), 20 % (v/v) glycerol, 10 mM DTT, 0.1 % bromophenol blue (w/v)	Sodium Dodecyl Sulfate - Polyacrylamide Gel Electrophoresis (SDS-PAGE) sample loading
SDS-PAGE	250 mM Tris, 1.85 M glycine, 1 % SDS (w/v)	SDS-PAGE
Western blot transfer	25 mM Tris, 200 mM glycine, 20 % (v/v) methanol	Western blot transfer
Blocking	3 % Bovine serum albumin (BSA)	Blot blocking, primary antibody incubation
TBST	150 mM NaCl, 50 mM Tris, 0.05 % (v/v) Tween-20	Secondary antibody incubation, wash steps

Detergent lysis	20 mM HEPES pH 7.5, 500 mM NaCl, 5 % glycerol, 10 mM imidazole pH 7.5	Detergent screen
Detergent wash	20 mM HEPES, 500 mM NaCl, 5 % glycerol, 25 mM imidazole, 3 x CMC detergent pH 7.5	Detergent screen
Detergent elution	20 mM HEPES, 500 mM NaCl, 5 % glycerol, 300 mM imidazole, 3 x CMC detergent pH 7.5	Detergent screen
Stain		
SYBR™ SAFE	SYBR™ Safe 98, DMSO (50 ppm)	Stain DNA gels
Coomassie stain	50 % (v/v) Methanol, 20% (v/v) glacial acetic acid, 0.12 % (w/v) Coomassie brilliant blue G250 dye	Stain SDS-PAGE gels
Destain	10 % methanol (v/v), 10 % acetic acid	Destain SDS-PAGE gels

Table 2.3: Composition of buffers used in this study

2.1.6 Kits

The kits used in this study are listed in table 2.4

Kit	**Manufacturer/ distributor**
QIAprep® spin miniprep	QIAGEN
QIAquick PCR purification	QIAGEN
QIAquick gel extraction	QIAGEN

Table 2.4: Kits used in this study

2.1.7 Detergents and synthetic lipids

All detergents and synthetic lipids (table 2.5) were purchased in powder form and stored as per manufacturer's recommendations. Lipid vials were sealed and wrapped with parafilm after use to prevent water ingress and stored at -20°C.

Detergent/ Lipid	Formal name	Manufacturer/ distributor
Lipid		
DMPC	1,2-dimyristoyl-sn-glycero-3-phosphocholine (14:0/ 14:0 phosphatidylcholine)	Avanti lipids
DOPC	1,2-dioleoyl-sn-glycero-3-phosphocholine (18:1/ 18:1 phosphatidylcholine)	Avanti lipids
Detergent		
DDM	n-Dodecyl-β-D-Maltoside	Anatrace
DDM + CHS	n-Dodecyl-β-D-Maltoside + Cholesterol hemisuccinate	Anatrace
DM	N-Decyl-β-D-Maltopyranoside	Anatrace
DM + CHS	N-Decyl-β-D-Maltopyranoside + Cholesterol hemisuccinate	Anatrace
OG	N-Octyl-β-D-Glucopyranoside	Anatrace
LMNG	Lauryl Maltose Neopentyl Glycol	Anatrace
OGNG + CHS	n-Octyl-β-D-Glucopyranoside/ n-Nonyl-β-D-Glucopyranoside + Cholesterol hemisuccinate	Anatrace
LDAO	Lauryldimethylamine-N-oxide or N,N-Dimethyldodecylamine N-oxide	Anatrace
C12E8	Polyoxyethylene 8	Anatrace
C12E9	Polyoxyethylene 9	Anatrace
Cymal-5	5-Cyclohexyl-1-Pentyl-β-D-Maltoside	Anatrace
FC12	Dodecylphosphocholine or Fos-choline 12	Avanti lipids
SB12	N-Dodecyl-*N*,*N*-dimethyl-3-ammonio-1-propanesulfonate or Sulfobetaine-12	Sigma
Brij-35	Tricosaethylene glycol mono-n-dodecyl ether	Sigma

Table 2.5: Detergents and synthetic lipids used in this study

2.1.8 **Bacterial strains and plasmids**

The Bacterial strains or plasmids used in this study are listed in table 2.6.

Bacterial strain or plasmid	**Relevant characteristics**	**Restriction sites**	**Antibiotic resistance**	**Source**
Bacterial strains				
Neb5α	*fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*			
BL21	*fhuA2 [lon] ompT gal [dcm] ΔhsdS*			Laboratory collection
BL21(DE3)	*fhuA2 [lon] ompT gal (λ DE3)[a] [dcm] ΔhsdS*			Gift from Chalmers lab
C41(DE3)	*F - ompT hsdSB (rB- mB-) gal dcm (λ DE3)[a]*			Gift from Williams lab
pLysS (BL21(DE3))	*MiniF lysY (CamR) / fhuA2 lacZ::T7 gene1 [lon] ompT gal sulA11 R(mcr-73::miniTn10--TetS)2 [dcm] R(zgb-210::Tn10--TetS) endA1 Δ(mcrC-mrr)114::IS10*		Chloramphenicol	Gift from Chalmers lab
Rosetta(DE3)	*F- ompT hsdSB(rB- mB-) gal dcm (λ DE3)[a] pRARE (Cam^R)*			Gift from Bolt lab

Lemo21(DE3)	*fhuA2 [lon] ompT gal (λ DE3)*[a] *[dcm]* Δ*hsdS/ pLemo*(Cam^R)[b]		Chloramphenicol	Gift from Chalmers lab
JCM290	*ΔbamD* arabinose inducible *paraBAD* regulated gene deletion mutant		Ampicillin	Gift from Ian Henderson
JWD3	*ΔbamA* arabinose inducible *paraBAD* regulated gene deletion mutant		Kanamycin	Gift from Ian Henderson
Plasmids				
pMH1	pET20b+ with BamA construct	NcoI and BamHI	Ampicillin	Matt Hardcastle
pET26b+	pET26b+ base vector		Kanamycin	Merck Millipore
p26b::*bamA* (pGM1)	pET26b+ with re-cloned *bamA* gene, N-terminal 10His-TEV, and C-terminal $(TAA)_3$	BamHI and EcoRI	Kanamycin	This Work
p26b::*bamD* (pGM2)	pET26b+ with *bamD* gene and C-terminal TEV-6His-$(TAA)_3$	NcoI and BamHI	Kanamycin	This Work
p26b::*cj1289* (pGM3)	pET26b+ with *cj1289* gene and N-terminal TEV-10His	BamHI and EcoRI	Kanamycin	This Work
p26b::*degP* (pGM4)	pET26b+ with *degP* gene and C-terminal TEV-6His	NcoI and EcoRI	Kanamycin	This Work
p26b::*lptD* (pGM5)	pET26b+ with *lptD* gene and C-terminal 10His-TEV-TAA	NcoI and BamHI	Kanamycin	This Work

p20b::*bamA* (pGM1B)	pET20b+ with re-cloned *bamA* gene, N-terminal 10His-TEV, and C-terminal $(TAA)_3$	NcoI and EcoR1	Ampicillin	This work

[a]λ DE3 = λ sBamHIo ΔEcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5

[b]pLemo = pACYC184-*PrhaBAD-lysY*

Table 2.6: Bacterial strains and plasmids used in this study

2.1.9 Molecular weight ladders

The molecular weight ladders used in this project are listed in table 2.7.

Ladder	Manufacturer	Ladder range	Use
PAGERuler™ Plus pre-stained protein ladder	Thermo Scientific	10 kDa to 250 kDa	SDS-PAGE gels
SeeBlue™ Pre-stained Protein Standard	Invitrogen	3 kDa to 198 kDa	SDS-PAGE gels
SeeBlue™ Plus2 Pre-stained Protein Standard	Invitrogen	3 kDa to 198 kDa	SDS-PAGE gels
Quick-Load® Purple 1 kb DNA Ladder	NEB	0.5 kb to 10 kb	Agarose gels/ cloning
Quick-Load® 100 bp DNA Ladder	NEB	100 bp to 1517 bp	Agarose gels/ cloning

Table 2.7: Molecular weight ladders used in this study

2.1.10 Equipment

Several pieces of equipment were purchased (table 2.8) to enable this PhD from the project grant, including:

Equipment	Manufacturer
LifeEco PCR machine	BioER
Benchtop microfuge, GT 1R benchtop centrifuge	Fisher
AnaeroJar™, protein gel tank and accessories	ThermoScientific
PowerPac 200 and PowerPac 300	BioRad
DNA gel tank	BioRad
20 L Static incubator (x2)	VWR
Orbital incubator	Stuart
All kits and enzymes	QIAgen, NEB

Table 2.8: Equipment purchased from this grant to enable this study

2.1.11 **Primers**

Several primers were used in this project, and are listed in table 2.9.

Primer name	**Sequence**	**Annealing temperature**	**Use**
bamA forward	GCGGGATCCGGCAACCATTAAAGATATT AAATTTATCGG	62°C	BamA amplification
bamA reverse	GCGGAATTCTTATTATTAAAGCGTGTT CCAAGATTAAATTC	61°C	BamA amplification
bamD forward	GGGCCATGGGGTGTAGTACGAAAAATGATGAAGG	60°C	BamD amplification
bamD reverse	CCCGGATCCAAATCTTTCAAATATACTTCTATACCAAG G	60°C	BamD amplification
cj1289 forward	GATCGGATCCGGCAAATACCATTAATGCTATAGC	59°C	Cj1289 amplification
cj1289 reverse	GATCGAATTCGCTCTTAAATATTCTATATTGATTTTAG AGCG	61°C	Cj1289 amplification
degP forward	GATCGGATCCAGCTAATCGTGTGAATCCTG	60°C	DegP amplification
degP reverse	GATCGAATTCTTTAAGCACAAGCAAAGTCG	61°C	DegP amplification
lptD forward	GATCCCATGGGGGCACAAGTTGATATTTACGC	59°C	LptD amplification
lptD reverse	GATCGGATCCGCAATTTTGTTTTCACTTTCTGC	61°C	LptD amplification

N-terminal tag forward	GAGAATCTGTACTTTCAGAGCGCCATGGATATCGG	55°C	TEV tag insertion N-terminal
N-terminal tag reverse	GTGATGGTGATGGTGATGATGATGGTGGTGGGCCATGGCCATCGC	59°C	10 His tag insertion N-terminal
C-terminal tag forward *lptD*	CATCATCATCACCATCACCACCACCACCACTGAGATCCGGCTGC	61°C	10 His tag insertion C-terminal for *lptD*
C-terminal tag reverse *lptD*	GCTCTGAAAGTACAGATTCTCAATTTTGTTTTCACTTTCTGCTAAAG	60°C	TEV tag insertion C-terminal for *lptD*
C-terminal tag forward *bamD*	CATCATCACCATCACCACTAATAATAAGGGGATCCGAATTCG	59°C	6 His tag insertion C-terminal for *bamD*
C-terminal tag reverse *bamD*	GCTCTGAAAGTACAGATTCTCAAATCTTTCAAATATACTTCTATACCAAG	58°C	TEV tag insertion C-terminal for *bamD*
C-terminal tag forward *degP*	GCTCTGAAAGTACAGATTCTCGCTCGAATTCGCTTTAAGC	62°C	TEV tag insertion C-terminal for *degP*
C-terminal tag reverse *degP*	CATCATCACCATCACCACGCCTCCGTCGACAAG	63°C	TEV tag insertion C-terminal for *degP*
bamA optimise forward	GCAACCATTAAAGATATTAAATTTATCGG	60°C	Remove redundant N-terminal vector sequence

cj1289 optimise forward	GCAAATACCATTAATGCTATAGC	59°C	Remove redundant N-terminal vector sequence
bamA/ cj1289 optimise (TEV) rev	GCTCTGAAAGTACAGATTCTC	60°C	Remove redundant N-terminal vector sequence
bamA SS 817F	TTAGATGTTCAAGTATCCTCCCC	59°C,	Sequencing primer to bridge T7P and T7T
bamA SS 364R	TATACCTGTGATAGTAACCTTTGC	58°C,	Sequencing primer to check start of *bamA* gene
lptD SS 740F	ACAAACCGTGGTTTTG	55°C,	Sequencing primer to bridge T7P and T7T
degP SS 764F	CAATATGAAAATTTTATACAAACAG	55°C,	Sequencing primer to bridge T7P and T7T good sequence
T7P	TAATACGACTCACTATAGGGG	55°C,	Standard sequencing primer
T7T	CTAGTTATTGCTCAGCGG	56°C,	Standard sequencing primer
TEV forward	GAGAATCTGTACTTTCAGAGC	60°C,	Colony PCR check

Table 2.9: List of primers used in this study. Table of the primers used in this project to generate the constructs with a description of their use. Restriction sites are in red, tags are in green, spacer sequences to keep the construct in frame are in blue, and stop codons are in purple.

2.1.12 **Computing resources**

There were three types of computer resource used in this project, web services (table 2.10), software (table 2.10), and high performance computing (table 2.11) resources.

Web service or software	**Web address of web service or software source**	**Use**
Web service		
BLAST/ BLASTP	https://blast.ncbi.nlm.nih.gov/Blast.cgi	Homology searches
EMBOSS transeq	https://www.ebi.ac.uk/Tools/st/emboss_transeq/	DNA to protein sequence translation
SignalP	http://www.cbs.dtu.dk/services/SignalP/	Determine signal leader peptide
LipoP	http://www.cbs.dtu.dk/services/LipoP/	Determine lipoprotein signal leader peptide
NEB T_m calculator	https://tmcalculator.neb.com/#!/main	Primer T_m calculator
NEB BioCalc	https://nebiocalculator.neb.com/#!/ligation	Ligation calculator
Expasy ProtParam	https://web.expasy.org/protparam/	Protein M/W, pI, extinction coefficient calculator
Reverse complement	https://www.bioinformatics.org/sms/rev_comp.ht ml	Create reverse complement of DNA
Debye-calculator	http://www.surfchem.info/calculate/Debye/	Debye length calculator
Uniprot	https://www.uniprot.org/	Find protein/ gene information
RCSB	https://www.rcsb.org/	Find protein crystal structures
I-TASSER	https://zhanglab.dcmb.med.umich.edu/I-TASSER/	Generate homology model

Phyre2	http://www.sbg.bio.ic.ac.uk/phyre2/ht ml/ page.cgi?id=index	Generate homology model/ cross-check for I-TASSER
CHARMM-GUI	https://charmm-gui.org/?doc=input	Generate input file for MD simulations
ZDock	https://zlab.umassmed.edu/zdock/	Dock protein
PyDock	https://life.bsc.es/pid/pydockweb	Evaluate/ re-rank docking
ROSIE	https://rosie.rosettacommons.org/docking 2/submit	Dock protein (alternative)
STRIDE	http://webclu.bio.wzw.tum.de/cgi-bin/stride/stridecgi.py	Generate protein SSA
DSSP	https://www3.cmbi.umcn.nl/xssp/	Generate protein SSA
PROSS	http://folding.chemistry.msstate.edu/utils/ runpross.ht ml	Generate protein SSA
2Struc	https://2struccompare.cryst.bbk.ac.uk/inde x.php	Generate protein SSA
Software		
Snapgene viewer	https://www.snapgene.com/snapgene-viewer/	View DNA sequence files, design primers
A Plasmid Editor	https://jorgensen.biology.utah.edu/wayne d/ape/	Prepare gene sequence for Snapgene viewer
Chromas Lite	https://www.softpedia.com/get/Science-CAD/Chromas-Lite.sht ml	Visualise sequencing chromatograms
Discovery Studio	https://discover.3ds.com/discovery-studio-visualizer-download	Visualise PDB files, generate images
UCSF Chimera	https://www.cgl.ucsf.edu/chimera/	Visualise PDB files, generate images
VMD	https://www.ks.uiuc.edu/Research/vmd/	Visualise/ analyse MD simulation trajectories
Protein Data Bank (PDB) File Editor	https://sourceforge.net/projects/pdbeditorj l/	Edit PDB files

NAMD	https://www.ks.uiuc.edu/Research/namd/ (pre-installed on HPC)	Run MD simulation
CDapps	https://www.diamond.ac.uk/Instruments/Soft-Condensed-Matter/B23/manual/Beamline-software.ht ml	Analyse CD data
SCATTER	https://www.bioisis.net/scatters/download	Analyse SAXS data
ATSAS	https://www.embl-hamburg.de/biosaxs/atsas-online/download.php	Analyse SAXS data

Table 2.10: Webservices and software used in this study

System name	Owner	Type of resource	Resource(s)
Verlet	University of Nottingham (Bonev Lab)	GPU	NVIDIA Tesla K10, Tesla K40, Tesla K80
Athena (HPC-Midlands-Plus)	Midlands Consortium	CPU	CPU/ core: Intel Xeon E5-2680v4 at 2.40 GHz (base) with 128 GB of memory.

Table 2.11: High performance computing resources used in this study

2.1.13 HPC Midlands Plus acknowledgement

We acknowledge the use of Athena at high performance computing (HPC) Midlands plus, which was funded by the EPSRC on grant EP/P020232/1, in this research in the generation of MD simulations using NAMD, as part of the HPC Midlands Plus consortium. The research consumed 1,355,366.833 CPU core hours, using SAFE project A14-0036.

2.2 Methods

2.2.1 Cloning

2.2.1.1 Gene sequence

The sequence for the *C. jejuni bamA*, *bamD*, *degP*, *cj1289*, and *lptD* genes were found through homology search using a nucleotide Basic Local Alignment Search Tool (BLAST) search against the *E. coli* homologous gene sequences, or the *C. crescentus* sequence for *bamF*, with an organism filter criterion of *C. jejuni* (taxonomic identifier: 197). The top result was used to design primers once the signal sequence length was determined through the SignalP and/ or LipoP servers (section 2.2.1.3). Where no matching gene sequence was found, the threshold of the blast search was lowered (more dissimilar sequences) or a similar search was run on a protein BLAST search using the *C. jejuni* (taxonomic identifier: 197) database criterion and *E. coli* protein sequences. Unfortunately, the three publications on the involvement of BepA in OMP biogenesis and degradation were not identified prior to thesis submission, and *C. jejuni* BepA (sequence ID: VTQ53926.1) does not feature in this project.

2.2.1.2 Operon search

The sequences of the identified *C. jejuni bamA* and *bamD* genes were searched against the genome assembly for *C. jejuni* NCTC11168 chromosome 1. The gene loci were identified for *bamA* (129,796…132,015) and *bamD* (1,008,795…1,009,442) and genes in the same reading direction were queried using BLAST.

2.2.1.3 Protein sequence

Where DNA sequences were found, the protein sequence was generated by converting the gene sequence using the EMBL-EBI EMBOSS Transeq tool[114] and codon table 11 (Bacteria). The outputs were screened using a protein BLAST search (taxonomic identifier: 197) to determine the appropriate sequence, and confirmed using a translational BLAST search. These protein sequences were used to generate homology models.

2.2.1.4 **Determination of signal sequence**

The SignalP and LipoP servers identified the signal peptide and cleavage sites based on a neural network trained against sets of prokaryotic Gram-negative sequences, and the algorithm outputs a graph with probabilities of the expected cleavage site and signal peptide type. The current versions at time of use were SignalP *ver.* 4.0 and LipoP *ver.* 1.0. Where SignalP gives a high probability of a lipoprotein signal peptide, or where it is inconclusive and the protein is known to interact with lipid *in vivo* (BamA, BamD, LptD), the SignalP analysis has been followed up with analysis using the LipoP server.

2.2.1.5 ***C. jejuni* plate culture**

C. jejuni culture was in accordance with Public Health England guidance for *C. jejuni* NCTC 11322 (www.phe-culturecollections.org.uk/collections/nctc.aspx) and guidance from the literature[115]. *C. jejuni* NCTC 11322 was sourced as a LYFO DISK™ (TCS Biosciences) and was selected as it is satisfied the criterion of being clinically relevant (i.e. faecal isolate from a patient) while being a commonly used industrial strain. Columbia agar was made up to 42 g/L for a 200 ml volume with 190 ml dH_2O (10 ml blood to be added later, see below) and autoclaved. To prepare Oxoid™ Preston Campylobacter selective antibiotic supplement (SR0117E, ThermoFisher) to isolate *C. jejuni*, add 5 ml of sterilised MQ to one vial of pre-prepared powder in aseptic conditions, shake gently until dissolved and keep on ice until use. Retrieve media and transfer to waterbath set to 50°C. Once cooled to 50°C, transfer to laminar flow cabinet and aseptically add 10 ml lysed horse blood (5% total volume) and 2 ml of Preston Campylobacter selective supplement. Swirl and pour ten plates to 20 ml/ plate, and leave plates open to set and dry for up to 10 minutes. Place one LYFO DISK™ of *C. jejuni* NCTC 11322 (TCS Biosciences) onto one plate for 5 minutes to allow it to absorb moisture from the plate. Once lyophilised pellet has hydrated, use streak plating technique to spread hydrated pellet across plate. Place plate lid down in Anaerojar™ plate holder insert. Open the Campygen™ gas generating sachet, quickly transfer into the sachet slot and close the Anaerojar™ as per manufacturer's instructions. The sachet will generate the atmosphere (approximately 10 % CO_2, 2-5 % O_2, 85-88 % N_2 by composition) required to grow microaerophilic organisms[116]. Once the Anaerojar™ is airtight transfer to stationary incubator set at 42°C and culture for 48-72 hours at 42°C. If

further plates are required, they are subcultured from this plate or alternatively the LYFO DISK™ can be dissolved in 1 ml sterile thioglycolate broth and 100 µl spread onto each plate required.

2.2.1.6 *C. jejuni* glycerol stocks

Transfer 5 ml sterile thioglycolate broth into a sterile universal tube and add 50 µl Preston antibiotic supplement. Transfer part of one colony using a sterile tip and incubate universal 48-72 hours at 42°C 180 rpm checking for growth at regular intervals. After the culture becomes turbid, transfer 2 x 750 µl into sterile Eppendorf tubes and centrifuge at 4000 rpm for 10 minutes at 4°C in benchtop microcentrifuge. Prepare 5 ml of sterile thioglycolate broth mixed with 40 % glycerol in a 50:50 mixture, resuspend pellet in 750 µl and snap freeze.

2.2.1.7 Preparation of DNA template

Take part of one well-grown healthy colony with a toothpick and place in a PCR tube containing 100 µl Tris-EDTA (TE) buffer. Place tube in PCR machine and heat for 5 minutes at 95°C to lyse all cells. Pipette into 10 x 10 µl aliquots and snap-freeze all aliquots that will not be used in this session.

2.2.1.8 PCR

Assemble reaction components on ice as below. Gently flick tubes to mix the reaction and centrifuge as required. Quickly transfer the reactions to a thermocycler preheated to the denaturation temperature. Standard Taq reaction (tables 2.12 and 2.13) and Q5 reaction (tables 2.14 and 2.15) set ups are shown.

Component	**25µl reaction**	**Final concentration**
10x std. Taq buffer	2.5 µl	1x
10mM dNTPs	0.5 µl	200 µM
10µM Forward primer	0.5 µl	0.2 µM
10µM Reverse primer	0.5 µl	0.2 µM
Template DNA	Variable	<10 ng/ part of 1 colony
***Taq* DNA polymerase**	0.125 µl	0.025 U/µl
Nuclease free water	To 25 µl	

Table 2.12: Standard taq PCR reaction set up

Step	Temperature	Time
Initial Denaturation	95°C	30 seconds*
30 cycles:		
Denaturation	95°C	5-30 seconds
Annealing	45-68°C (as appropriate)	15-60 seconds
Extension	68°C	1 minute/kb
Final Extension	68°C	5 minutes
Hold	4-10°C	∞

* For colony PCR, the initial denaturation time is changed to 5 minutes.

Table 2.13: Standard taq PCR cycle conditions

Component	25µl reaction	Final concentration
5x Q5 reaction buffer	5.0 µl	1x
10mM dNTPs	0.5 µl	200 µM
10µM Forward primer	1.25 µl	0.5 µM
10µM Reverse primer	1.25 µl	0.5 µM
Template DNA	Variable	<10 ng/ 1 colony
Q5 DNA polymerase	0.25 µl	0.02 U/µl
Nuclease free water	To 25 µl	

Table 2.14: Standard Q5 PCR reaction set up

Step	Temperature	Time
Initial Denaturation	98°C	30 seconds*
25-35 cycles:		
Denaturation	98°C	5-10 seconds
Annealing	50-72°C (as appropriate)	10-30 seconds
Extension	72°C	30 seconds/kb
Final Extension	72°C	2 minutes
Hold	4-10°C	∞

* For colony PCR, the initial denaturation time can be increased up to 3 minutes.

Table 2.15: Standard Q5 PCR cycle conditions

2.2.1.9 **Agarose gel electrophoresis**

All DNA molecular cloning products were run on agarose gels to check amplification, size and digestion during the cloning process. To prepare a 1 % gel, mix 0.25 g agarose with 25 ml of Tris-Acetate-EDTA (TAE) in a small conical flask, note weight, and microwave for 2 minutes on high. Weigh conical flask after heating and replenish lost dH_2O. Prepare the agarose gel cassette for casting. Once the solution has cooled to around 50-60°C, add 2.5 µl SYBR™ Safe DNA gel stain (ThermoFisher), swirl conical to mix, and pour gel in prepared cassette. Comb the gel to remove any bubbles, add in comb, and leave to set. Cover tank with foil to prevent premature degradation of SYBR safe. Once the gel has set, fill the tank with (1x) TAE buffer, remove comb and then add the ladder and samples. For DNA <1.5 kb in length a 100 bp ladder (NEB, #N0551s) was run alongside samples on a 1.2 % agarose gel. For DNA >1.5 kb in length a 1 kb ladder (NEB, #N0552s) was run alongside samples on a 1 % agarose gel. All restriction digest products were run on a 0.7 % agarose gel unless otherwise stated. Standard run conditions for all runs were (1x) TAE buffer, 80 v for 45 minutes.

2.2.1.10 **PCR clean-up**

As per manufacturer's instructions, (QIAGEN, 'QIAquick PCR Purification using a Microcentrifuge' protocol), with the exceptions that PE is replaced with ice-cold 80 % ethanol and the elution buffer is heated to 55°C prior to application to column.

2.2.1.11 **Restriction digest**

A typical restriction digest reaction set up is shown (table 2.16).

Component	25 µl Reaction	50 µl Reaction	Final concentration
10 x buffer (CutSmart®)	2.5	5	1x
Restriction enzyme 1	0.5	1	≥5 U/25 µl reaction ≥10 U/50 µl reaction
Restriction enzyme 2 (optional)	0.5	1	≥5 U/25 µl reaction ≥10 U/50 µl reaction

Calf Intestinal Phosphatase (optional)	0.5	1	≥5 U/25 µl reaction ≥10 U/50 µl reaction
DNA template	Variable (0.5 µg)	Variable (1 µg)	Variable
dH2O	Up to 25 µl	Up to 50 µl	

Table 2.16: Standard restriction digest set up

For coherent ends, digest up to 1 µg of plasmid with the appropriate enzymes for cloning (where both enzymes use the same buffer) for 30 minutes at 37°C in the heat block or water bath. Add 1 µl calf intestinal phosphatase (CIP) to the reaction and incubate for another 30 minutes at 37°C. Simultaneously digest up to 1 µg of insert with the corresponding enzymes for 1 hour at 37°C in the heat block or water bath. For blunt ends the CIP does not need to be added, otherwise the procedure is the same as for coherent ends. If Time-Saver™ enzymes from NEB (or a similar product from a different manufacturer) are used, reduce the digestion time as per manufacturer's recommendations. Following the reaction, the enzymes can be heat inactivated as per manufacturer's instructions, or removed by running on an agarose gel using a loading dye that contains sodium dodecyl sulphate (SDS).

2.2.1.12 **Digest check**

To check the digest, add 3 µl of digested product to 7 µl MilliQ water and 2 µl purple loading dye (6x) and run on a 0.8 % agarose gel for 45 minutes. Alternatively, proceed to the gel extraction stage.

2.2.1.13 **Gel Extraction**

As per manufacturer's instructions (QIAGEN, 'QIAquick Gel Extraction using a Microcentrifuge' protocol), with the exceptions that PE is replaced with ice-cold 80 % ethanol and the elution buffer is heated to 55°C prior to application to column.

2.2.1.14 **Ligation**

Set up ligation reaction(s) on ice using standard set up shown in table 2.17. Gently mix the reaction by pipetting up and down and microfuge briefly or by flicking the tube. For cohesive ends, incubate for 10 minutes at room temperature or 16°C

overnight. For blunt ends, incubate for 2 hours at room temperature or 16°C overnight. Heat inactivate the DNA ligase at 65°C for 10 minutes. Chill on ice and transform 0.5-2 µl of the reaction into 20 µl competent cells. Ligation reactions were set up in 1:1, 1:3, and 1:5 vector to insert ratios using the NEBioCalculator to calculate molar ratios.

Component	**Typical 20 µl Reaction**
T4 DNA Ligase Buffer (10X)	2 µl
Vector DNA	50 ng
Insert DNA	37.5 ng
Nuclease-free water	to 20 µl
T4 DNA Ligase	1 µl

Table 2.17: Standard Ligation reaction set up

2.2.1.15 $CaCl_2$ (and $MgCl_2$) chemically competent cell preparation

From a working stock plate inoculate a single colony into 5 ml LB broth medium, and incubate overnight at 37°C and shaking at 180 rpm. The following morning inoculate 100 ml of Luria-Bertani (LB) broth to OD_{600} to 0.02 using the overnight culture, and incubate suspension at 37°C shaking (180 rpm). Once culture reaches $OD_{600} \approx 0.40$, cool culture on ice for 10 minutes. Split the culture between two 50 ml Falcon tubes and pellet cells on a pre-chilled rotor (4°C) at 4,000 x *g* for 10 minutes. Gently re-suspend cell pellets in 25 ml ice-cold 0.1 M $CaCl_2$ medium, and incubate on ice for 30 minutes. Pellet cells as before on a pre-chilled rotor at 4,000 x *g* for 10 minutes. Gently re-suspend pellets in 2.5 ml of ice-cold 0.1 M $CaCl_2$, 15 % glycerol medium each and leave on ice for 10 minutes. Aliquot cells on ice into pre-labelled Eppendorf tubes and either use immediately or flash-freeze in liquid nitrogen for long term storage at -80°C.

For $CaCl_2$ and $MgCl_2$ chemically competent cell protocol substitute the first 0.1 M $CaCl_2$ wash for 0.1 M $MgCl_2$, add in a second wash with 25 ml ice-cold 0.1 M $CaCl_2$, and replace final resuspension in 0.1 M $CaCl_2$ and 15 % glycerol with 0.85 M $CaCl_2$ + 15 % glycerol.

These preparations were used to generate competent cells during this study for the following strains: Neb5α, DH5α, BL21, BL21(DE3), C41(DE3), pLysS, Rosetta(DE3), BL21(DE3)ai, and Δ*ABCF*.

2.2.1.16 **Heat shock transformation**

Thaw *E. coli* aliquot on ice for 10 minutes. To 50 µl of competent cells, add 0.5-1 µl of a standard DNA miniprep or up to a maximum of 5 µl of a ligation mixture. Incubate cell preparations on ice for 30 minutes. Heat shock cells at 42°C for 1 minute, and place on ice for 2 minutes. Add 900 µl LB broth to each cell preparation and incubate at 37°C shaking at 180 rpm for 1 hour. Pellet cells on tabletop microcentrifuge at 4,000 x *g* for 10 minutes at room temperature. Discard 900 µl from each cell preparation and re-suspend cells into their respective remaining media (approximately 50 µl) onto the appropriate resistance/ control plate using a sterile glass spreader or disposable inoculating loop. Incubate plates in a static incubator at 37°C.

2.2.1.17 **Plasmid mini-preps**

As per manufacturer's instructions (QIAGEN, 'Plasmid DNA Purification using the QIAprep Spin Miniprep Kit and a Microcentrifuge' protocol), with the exceptions that PE is replaced with ice-cold 80 % ethanol and the elution buffer is heated to 55°C prior to application to column.

2.2.1.18 **DNA quantification**

The plasmid and insert preparations were quantified using a NanoDrop 2000/2000c spectrophotometer (ThermoFisher) at 260 nm using the eluting medium as a blank for the measurement.

2.2.1.19 **Plasmid screening**

The resultant plasmids were screened to check the genes, TEV- and His-tags using combinations of restriction enzymes with restriction sites either flanking the gene or DNA element, or using one flanking restriction site and one internal restriction site (where available) prior to Sanger sequencing.

2.2.1.20 Sanger sequencing

Sanger sequencing was carried out by the DeepSeq lab in Nottingham at cost, and analysed sequences were received as a chromatogram with the fluorescence signals (.ab1), and a sequence file (.dna). For more details on the sequencing, reactions and procedure contact the DeepSeq lab (D106, Queen's medical school).

2.2.1.21 ChromasLite

Sanger sequencing chromatogram files (.ab1) were viewed using a freeware program, ChromasLite, which was used to check noisy sequence regions for incorrect annotations that may cause sequence frame shifts.

2.2.1.22 Snapgene viewer

Sanger sequencing sequence output files (.dna) were viewed using the free version of Snapgene viewer. This software was chosen for ease of use and its graphic user interface, however the free version lacks DNA sequence editing tools and alternative software was used for this purpose (see 2.2.1.22).

2.2.1.23 A Plasmid Editor

A Plasmid Editor was used to add/ remove DNA to model the changes after adding in sequence from inverse PCR reactions, to check for potential frame shifts, and to adjust primer designs.

2.2.2 Bacterial overexpression, protein purification and preparation for further experiments

2.2.2.1 *E. coli* overnight cultures

Several molecular biology techniques require overnight cultures. For starter cultures, 10 ml of LB broth with an appropriate antibiotic was inoculated with either a 5 µl loopful of glycerol stock or part of a single colony from an agar plate, and incubated for 14-16 hours at 37°C at 180 rpm. For plasmid preparations and glycerol stocks 5 ml LB was used.

2.2.2.2 Glycerol stocks

Prepare a 5 ml LB overnight culture and centrifuge culture the following morning at 2500 x *g*, discarding supernatant. Re-suspend pelleted cells in 3 ml filter sterilised

40 % glycerol: LB (1:1) mixture by gently swirling and leave to incubate at room temperature for 10 minutes to allow cells to adapt to the glycerol. 0.5-1.0 ml aliquots were pipetted into sterile Eppendorf tubes or 'Cryo-Vials', immediately snap frozen in liquid nitrogen and transferred to the -80°C freezer for long-term storage.

2.2.2.3 **Protein overexpression**

Prepare 5 x 750 ml LB to 50 µg/ml kanamycin using stock and a starting OD_{600} of 0.02 using overnight BamA starter cultures in 2.5 L conical flasks, and incubate cultures at 30°C 180 rpm in orbital shaking incubator. After the first 2 hours, the OD_{600} was followed every 40 minutes until an OD_{600} of 0.6 ± 0.05 was reached. The cultures were cooled to the induction temperature (16°C), and left to incubated at this temperature for 30 minutes before overnight induction with 0.4 mM isopropyl β-$_D$-1-thiogallactopyranoside (IPTG) for 14-16 hours. Cultures were incubated after induction for 14-16 hours overnight. The following morning the cultures were collected, transferred to 1 L centrifuge buckets and centrifuged at 4000 x *g* for 30 minutes at 4°C in a pre-chilled centrifuge. Pellets were collected and were either processed as in pellet lysis or transferred to 50 ml Falcon tubes and frozen for later processing. Frozen culture pellets were kept at -80°C until required.

For BamD, Cj1289, DegP and LptD protein overexpressions substitute conditions as required for optimum expression (appendix 8.2.2).

2.2.2.4 **Expression testing of new constructs**

Due to the low yield of the recombinantly expressed *C. jejuni* proteins, all protein in this study were systematically screened to find the best conditions for expression. Expression testing variables (appendix 8.2.1) and optimised expression conditions (appendix 8.2.2) are shown.

2.2.2.5 **Pellet lysis**

Thaw pellets if required. Add 4 ml ice-cold buffer for each gram of pellet being processed (as appropriate), and resuspend pellets to generate suspensions using a Dounce homogeniser. Collate all suspensions into 50 ml Falcon tubes/ a 250 ml duran (as appropriate) add 1/1000 volume of protease inhibitor cocktail (P8849, ThermoFisher), lysozyme from a 50 mg/ml stock to 0.1 mg/ml, and $MgCl_2$ from a

1 M stock to 10 mM. Add a small amount of DNAse I and proceed to cell disruption (volumes > 50 ml) or sonication (volumes < 50 ml) to lyse the cells.

2.2.2.6 **Cell disruption**

Turn on cell disruptor and cooling system 30 minutes before use to allow the system to cool down. Prepare cell disruptor (Constant Cell Disruption Systems) by flushing the system with 200 ml ice-cold buffer at 10-15 kPSI to remove any debris and contaminants left from previous users runs. Place sample in sample chamber, change pressure to 36 kPSI and start cell disruptor. Collect cell lysate as it passes out of the system in a Falcon tube on ice, changing tube as necessary. Run collected material through for a second pass. Clean down system by running through 200 ml buffer, 100 ml dH_2O and 100 ml 70 % ethanol at 10-15 kPSI.

2.2.2.7 **Sonication**

Prepare sonicator for use, cleaning sonicator tip with 70 % IMS and then dH_2O. Place sample Falcon tube on ice in beaker and immerse end of sonicator tip into culture. Sonicate using an on/off cycle of 1 second/ 2 seconds at 20 amplitude (approximately 30 W) for 6 minutes total sonication time. Assess sample for opacity and colour changes, and sonicate for a further three minutes using the same settings (as required) or proceed to protein purification.

2.2.2.8 **Protein purification**

Balance collected cell lysate between the Falcon tubes and centrifuge at 10,000 *g* for 10 minutes at 4°C in a pre-cooled benchtop centrifuge to remove cell debris. Repeat as necessary until no cell debris is pelleted. Depending on protein type and localisation follow either the membrane protein or soluble protein purification procedure and then with IMAC purification.

2.2.2.9 **Membrane protein purification**

Transfer the soluble fraction of the lysed cell preparation to 12 ml centrifuge tubes (Beckman Coulter) for ultracentrifugation. Balance centrifuge tubes within 0.01 g, load into ultracentrifuge rotor and centrifuge at 30000 rpm (SW41Ti, Beckman Coulter) for 2 hours at 4°C in a pre-chilled ultracentrifuge to isolate membrane component of lysate. Discard supernatant and retain membrane pellet. Solubilise

membrane pellet in 10 ml 0.5 % sulfobetaine-12 (SB12) for BamA or 0.5 % Fos-Choline 12 (FC12) for LptD (as appropriate) for 2 hours in the cold room on a rotating wheel. Centrifuge the solubilised lysate at 38,000 rpm (SW41Ti, Beckman Coulter) for 1 hour at 4°C to remove insoluble contaminants and collect supernatant in a 15 ml Falcon tube.

2.2.2.10 **Soluble protein purification**

Transfer lysed cell supernatant to 30 ml centrifuge tubes (Beckman Coulter), balance tubes to within 0.01 g, and centrifuge at 30,000 x *g* (JA-25:50, Avanti) for 20 minutes at 4°C to remove any residual cell debris. Repeat as necessary. Transfer supernatant into 50 ml Falcon tube(s), as appropriate.

2.2.2.11 **IMAC purification**

Prepare 100 µl of nickel-nitrilotriacetic acid (Ni-NTA) resin (Bio-Rad) per mg of expected protein in a 10 ml Pierce™ centrifuge column (ThermoFisher) by applying 3 x 5 column volumes (CVs) dH_2O, and 10 CVs of binding buffer (30 mM imidazole in TBS/ PBS ± 0.25 % SB12/ 0.14 % FC12, as appropriate) to prepare resin for lysate addition, discarding flowthrough. Cap column, add 1 ml of binding buffer, pipette to re-suspend resin and transfer to 15 ml/ 50 ml Falcon tube containing centrifuged cell lysate supernatant. Add imidazole from stock to bring solution up to 30 mM total imidazole, transfer tube to rotating wheel and bind on rotating wheel at 15 rpm for 2 hours in the cold room. Collect Falcon tube post-bind and transfer the lysate-resin mixture to the centrifuge column, and allow resin to settle. Once resin is settled, remove the cap from the column, and collect the flowthrough in a clean 15 ml Falcon tube. After lysate flowthrough is collected, apply 3 x 5 CVs of binding buffer to the column to wash off any non-specifically bound proteins, collect flowthrough in a clean 15 ml Falcon tube. Cap the end of the column and add 500 µl of elution buffer (250 mM imidazole in TBS/ PBS ± 0.25 % SB12/ 0.14 % FC12, as appropriate), and leave to incubate for 3 minutes. De-cap the column, collect flowthrough in clean Eppendorf, and label as ‘elution 1’. Apply a further 500 µl elution buffer to the resin, collect in a clean Eppendorf as ‘elution 2’ and repeat to collect ‘elution 3’. Cap end of column and add 500 µl elution buffer, cap the top of the column and store in the fridge until elution has been analysed on SDS-PAGE. Store elution fractions in fridge at 4°C until required.

2.2.2.12 **SDS-PAGE gels**

All protein extractions were run on a polyacrylamide gel to check for expression of the protein of interest with resolving gel percentage chosen based on molecular weight (M/W). All gel components used to make the stacking (table 2.18) and resolving gels (table 2.19) were added into two separate 50 ml Falcon tubes except tetramethylethylenediamine (TEMED), which was added prior to pouring into gel casting cassettes (Bio-Rad). Resolving gel was added into gel cassette, overlaid with propan-2-ol to remove bubbles and level the solution, and left to set for 25 minutes at room temperature. Once polymerised, the propan-2-ol was removed, and the gel rinsed with dH_2O and blotted dry to remove any residual propan-2-ol. The stacking gel was poured on top of the separating gel, the comb was fitted and the stacking gel was left to polymerise for 25 minutes.

While the stacking gels are setting, prepare samples in 4x gel loading buffer (GLB) and heat the samples on a holding PCR cycle at 95°C for 10 minutes. Set up gel plates for running in gel tank and load samples into their corresponding wells as planned. Run gel at 100 V for 25 minutes for stacking gel and then 140 V for 45/ 55/ 65 minutes for 10/ 12/ 15 % gels (respectively) for the separating gel. Following run completion, proceed with Coomassie blue staining or Western blotting procedures, as required.

	Stacking gel (%)			
Component	**1 Gel**	**2 Gels**	**3 Gels**	**4 Gels**
dH_2O	1.4875 ml	2.975 ml	4.4625 ml	5.95 ml
0.5 M Tris-HCl, pH 6.8	0.675 ml	1.25 ml	1.875 ml	2.5 ml
10 % (w/v) SDS	25 µl	50 µl	75 µl	100 µl
Acrylamide/ Bis-acrylamide (30 %/ 0.8 %, w/v)	335 µl	670 µl	1005 µl	1340 µl
10 % (w/v) APS	25 µl	50 µl	75 µl	100 µl
TEMED	2.5 µl	5 µl	7.5 µl	10 µl

Table 2.18: Stacking gel composition and set up

(For 2 Gels)	**Separating gel (%)**			
Component	**8%**	**10%**	**12%**	**15%**
dH_2O	4.6 ml	3.8 ml	3.2 ml	2.2 ml
Acrylamide/ Bis-acrylamide (30 %/ 0.8 %, w/v)	2.6 ml	3.4 ml	4 ml	5 ml
1.5 M Tris (pH 8.8)	2.6 ml	2.6 ml	2.6 ml	2.6 ml
10 % (w/v) SDS	100 µl	100 µl	100 µl	100 µl
10 % (w/v) APS	100 µl	100 µl	100 µl	100 µl
TEMED	10 µl	10 µl	10 µl	10 µl

Table 2.19: Separating gel composition and set up

2.2.2.13 Coomassie blue staining

After SDS-PAGE the gel was transferred to a clean plastic container, Coomassie stain was added to cover the gel and the gel was left on a rocker (Stuart) at 20 rpm for 15-20 minutes. The stain was removed and the gel rinsed twice with dH_2O to remove residual Coomassie stain to prevent further staining. To quickly check gel bands, dH_2O was added to cover gel, and the container was microwaved on high (700 W microwave) for 2 minutes. After checking the gel bands, the dH_2O was replaced with destain buffer, and the gel left overnight on a rocker to destain at 20 rpm.

2.2.2.14 Western blotting

Nitrocellulose (Amersham) and Whatmann filter paper were cut to size and left to soak in Western transfer buffer (200 mM glycine, 25 mM tris, 20 % (v/v) methanol) along with the Western cassette sponges for 5 minutes. The cassette was assembled with the nitrocellulose membrane facing towards the anode and the SDS-PAGE gel facing towards the cathode sandwiched between the Whatmann filter paper and then cassette sponges. A roller was used to flatten the gel onto the membrane and remove any bubbles and the assembled cassette was placed in the gel tank. The protein bands were transferred at 80 V for 2 hours at room temperature (or overnight at 16°C at 25 V) in ice-cold Western blot transfer buffer kept chilled using an ice pack. The gel

band transfer from gel to blot was checked post-transfer. The membrane was then rinsed three times in TBS (pH 7.5) and blocked in blocking buffer (3 % bovine serum albumin in TBS) for 1 hour at room temperature on the rocker at 20 rpm. The blocking buffer was removed and the blot was transferred to a plastic sheet using clean tweezers. Three sides of the plastic were sealed using a bag sealer and 2 ml of fresh blocking buffer with a 1/5000 dilution of mouse Penta-His tag (primary) antibody (Li-Cor) was added. Bubbles were removed, the plastic sheet was sealed off and the blot was left on a rocker to incubate with the primary antibody for 2 hours at room temperature at 20 rpm. The blot was then removed, rinsed three times in TBS, and washed three times in TBS + 0.05% Tween-20 (TBST) for 10 minutes on a rocker, replacing the TBST between each wash cycle. The blot was transferred into a plastic sheet, and sealed as before with 2 ml of TBST with a 1/10000 dilution of IRDye® 680LT goat anti-mouse fluorescent (secondary) antibody (Li-Cor). The sealed plastic sheet was wrapped in foil, and left on the rocker for a 1 hour secondary antibody incubation at room temperature. The blot was removed, washed three times with TBST for 10 minutes, replacing the TBST between each wash cycle, and the blot was scanned.

2.2.2.15 **Gel scan**

All SDS-PAGE gels and nitrocellulose membranes were scanned on an Odyssey Li-Cor unless otherwise stated.

2.2.2.16 **Dialysis**

Dialysis was used to either remove excess salt or detergent, carry out buffer exchange or to dialyse out detergent when shifting membrane proteins into lipid. As the smallest protein in this study was 27 kDa, a 10-14 kDa MWCO dialysis tubing was selected for the desalting process and a 20 kDa MWCO for the detergent exchange or removal processes.

For the desalting process, dialysis was carried out in the sample buffer minus the salt to be removed or at the new salt concentration. For dialysis, a minimum ratio of 100:1 (buffer: sample) for 2 hours on a stirrer in the cold room was utilised in all cases with a buffer change after the first hour, unless stated otherwise.

For the detergent exchange process, dialysis was carried out in the new detergent at the appropriate concentration; however, a minimum dialysis time of 72 hours was utilised with a buffer change at each 24 hour interval to allow for slow diffusion and exchange of detergent from the sample.

To shift the membrane protein into lipid, the detergent was removed by dialysing against the detergent-free buffer for 72 hours at 4°C on a plate stirrer, changing out the buffer for detergent-free buffer at regular intervals.

2.2.2.17 **FPLC**

To separate the protein expression target from the IMAC contaminants a fast protein liquid chromatography (FPLC) system using either ion-exchange (IEX) chromatography columns (5 ml Q-HP or SP, Cytiva) or size exclusion chromatography (SEC) columns (Hi-load 16/600 120 ml S200 or S75, Cytiva) was utilised. All FPLC work was performed on an ÄKTA prime or ÄKTA start (Cytiva) in accordance with manufacturer's recommendations for operating parameters. Fractions were assessed for protein concentration using the inbuilt UV spectrometer calibrated at 280 nm.

2.2.2.18 **Column preparation**

The columns were stored in 20 % ethanol (± 0.2 M sodium acetate) as per manufacturer's instructions. To prepare the column for use with the samples, the column is first flushed through with 1.5 CVs (180 ml, SEC columns) or 5 CVs (IEX columns) dH_2O, and then equilibrated with the same volume of the buffer conditions corresponding to the protein being run.

2.2.2.19 **System degassing**

As the lab FPLC is kept at room temperature, lab temperature variations can cause bubbles to form in the tubing. The FPLC was degassed thoroughly as per manufacturer's instructions before every run to remove any air bubbles that may elevate column or system pressure levels.

2.2.2.20 **Column and FPLC storage**

After use, the FPLC unit and columns were exchanged into a 20 % ethanol (± 0.2 M sodium acetate) solution (as appropriate) using the same volumes and flow rate as in the column preparation.

2.2.2.21 **Size exclusion chromatography**

Size exclusion chromatography is a type of liquid chromatography that uses a stationary solid phase consisting of either a gel matrix or porous silica/ polymer beads (with random infrastructure but discreet sizing of pores) to separate proteins based on their size by selectively altering their progression through the mobile (liquid) phase of the column. The sizing of the porous stationary phase determines which proteins can enter the porous silica, which in turn alters how much time the protein of interest can spend in association with the stationary phase before elution. As larger proteins spend less time in association with the solid phase of a given size, this causes proteins to elute in a roughly sequential manner, largest to smallest, assuming appropriate stationary phase selection. The exception to this general rule is that all proteins that are larger than the pores of the stationary phase will elute at a very similar, if not the same, time (or excluded volume), and that the different shapes of the proteins may impact the exact timings of elution from the column for any given size of protein.

The SEC column was prepared using the appropriate buffer for the protein of interest and purified protein sample was injected via the sample loop. The FPLC was run at 1 ml/minute for 1.5 CVs, with 2 ml elution fractions collected after the void volume (approximately 40 ml) had elapsed. Where the elution point was previously determined, collection of fractions was delayed until the appropriate elution point. Elution fractions were assessed using SDS-PAGE and sample containing fraction were spin concentrated and either prepared for experiment or snap frozen in liquid nitrogen (soluble monomeric periplasmic proteins only).

2.2.2.22 **Concentration of proteins**

Proteins were concentrated with Amicon Ultra centrifugation filters (EMD Millipore) or the equivalent Vivaspin centrifugation filters (Cytiva). Several sizes and volumes were used from the ranges as appropriate with consideration given for detergent micelle sizes to minimise concentration of detergent (where appropriate).

2.2.3 Screening and optimisation of recovery of proteins

2.2.3.1 Detergent screening

The use of detergents as lipid analogues for membrane proteins in solution is dictated by their critical micelle concentration (CMC). CMCs of detergents used in detergent screening (table 2.20) are shown.

Detergent/ additive	Stock concentration	Critical micelle concentration	Extraction concentration	Purification concentration
DDM	10 %	0.17 mM (0.0087 %)	1 %	0.03 %
DDM + CHS	10 % + 2 %	0.17 mM (0.0087 %)	1 % + 0.2 %	0.03 % + 0.01 %
DM	10 %	1.8 mM (0.087 %)	1 %	0.25 %
DM + CHS	10 % + 2 %	1.8 mM (0.087%)	1 % + 0.2 %	0.25 % + 0.01 %
OG	10 %	18-20 mM (0.53 %)	1 %	1.50 %
LMNG	10 %	0.01 mM (0.001 %)	1 %	0.01 %
OGNG + CHS	10 % + 2 %	1.02 mM (0.058 %)	1 % + 0.2 %	0.20 % + 0.04 %
LDAO	10 %	1-2 mM (0.023 %)	1 %	0.07 %
C12E8	10 %	0.09 mM (0.0048 %)	1 %	0.015 %
C12E9	10 %	0.05 mM (0.03 %)	1 %	0.09 %
Cymal-5	10 %	2.4 mM (0.12 %)	1 %	0.40 %
Fos-Choline 12	10 %	1.5 mM (0.047 %)	1 %	0.14 %

Table 2.20: The critical micelle concentrations (CMC), extraction and purification concentrations of detergents used in this study

110 µl of carboxymethylated aspartic acid (TALON™) metal affinity resin was prepared for each detergent condition for every protein to be detergent screened.

Briefly, the required volume of 50 % TALON™ resin was added into a Falcon tube, centrifuged at 300 x *g* for 2 minutes at 4°C, and the storage buffer supernatant was discarded. The resin was washed with 3 x 5 CVs of dH_2O, inverting tube to resuspend the resin, and centrifuged at 300 *g* for 2 minutes at 4°C. This washing step was repeated replacing dH_2O with lysis buffer, then after discarding the supernatant, lysis buffer was added in a 1:1 ratio (buffer: resin), and the resin stored at 4°C until use.

Membranes (isolated from 500 ml overexpression culture) were resuspended in 25 ml detergent lysis buffer with one protease inhibitor cocktail tablet added, and membranes were homogenised using a Dounce homogeniser. 900 µl of resuspended membrane was added into each well of a row of a 96 deep well block according to number of samples (one row/ protein). 100 µl of the appropriate (10x) extraction concentration detergent was added, one per column, and the block was sealed with a foil seal, and left at 450 rpm on a platform shaker for 1 hour at 4°C. The block was centrifuged at 3500 x *g* for 30 minutes at 4°C to pellet any remaining insoluble material, and supernatant was transferred into a new deep well block. 100 µl of equilibrated TALON™ resin was added into each well using a cut tip, and the block was covered with a foil seal and left to incubate for 1 hour at 4°C at 450 rpm to allow for protein binding. The solubilised membrane - resin mixture was transferred into the filter plate (2 ml long drip 25 µM pore size, Crawford Scientific) wells keeping the same layout and placed on top of a 'waste' deep well block. The block assembly was centrifuged at 300 x *g* at 4°C for 2 minutes, and flowthrough collected in the waste block. 1 ml of detergent screen wash buffer was prepared in the corresponding detergent (3x CMC, purification concentration) in two clean 96 deep well blocks mirroring the layout of the original block. 500 µl of detergent wash buffer (with detergent added) was transferred into the filter plate, and the block assembly was centrifuged at 300 x *g* for 1 minute at 4°C. Flowthrough was discarded, and the wash step was repeated twice discarding the flowthrough each time. The block assembly was centrifuged at 500 x *g* for 3 minutes at 4°C to remove any remaining wash buffer. A microtitre plate with 100 µl detergent elution buffer and add 10 µl of (10x) purification concentration detergent was prepared, mirroring the layout of the original block. The filter plate was placed on top of a clean microtitre plate and 50 µl of elution buffer was added directly onto the resin. The plate was sealed with a foil

seal, and the plate assembly was sealed together with parafilm. The assembly was incubated for 20 minutes at 4°C on a platform shaker at 450 rpm. The seal was removed and the plate assembly was centrifuged at 500 x *g* for 3 minutes at 4°C to elute the protein(s). 10 µl of sample was mixed with 5 µl of (4x) GLB and samples were run on an SDS-PAGE gel.

2.2.3.2 Liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis

Samples were prepared by running spin concentrated FPLC purified samples for BamA, BamD, Cj1289, DegP and LptD on a 12 % SDS-PAGE gel for 20 minutes at 100 V (to clear stacking gel) and 55 minutes at 140 V. Bands were carefully excised with 4 mm x 1 mm gel cutting tips and bands were sequentially transferred to a PCR plate with each well containing 100 µl LC-MS/MS buffer. The band excision order was noted; PCR plate foil sealed and sent off for analysis to Oxford Mass Spectrometry Research Facility on dry ice.

2.2.3.3 BamA IMAC resin choice and imidazole concentration optimisation

400 mg of BamA overexpression membrane isolate was solubilised in 10 ml 1 % DDM overnight. 1 ml TALON™ resin was added to 2 gravity flow columns and the columns were washed using 15 CVs dH_2O, 5 CVs TBS, and 5 CVs 1 % DDM in TBS in turn. 5 ml solubilised membrane isolate (200 mg membrane in 1 % DDM in TBS) was added to each column, and the columns were left on a rotating wheel at 15 rpm for 4 hours at 4°C for protein binding. The columns were placed upright in 15 ml Falcon tubes to allow the resin to settle for 5 minutes. The solubilised membrane DDM solution was allowed to flowthrough and retained. The samples were then sequentially washed with 10 CVs of wash buffer with increasing concentrations of imidazole (10 mM to 70 mM) with each wash, and eluted using 0.5 CVs of elution buffer 10 times to estimate the correct recovery volume.

2.2.3.4 BamA TALON resin binding time optimisation IMAC screen

500 mg of BamA overexpression membrane isolates were solubilised in 10 ml 1% DDM for 2 hours at 4°C on a rotating wheel. The solubilised membrane was then split equally between five gravity flow columns each containing 0.5 ml pre-equilibrated TALON™ resin. The columns were all made up to 10 ml by adding 8 ml ice-cold TBS, capped and left to bind in the cold room at 4°C on a rotating

wheel. One sample was recovered and processed through to elution at 0.5 hours, 1 hour, 2 hours, 4 hours and after overnight binding time points.

2.2.3.5 **BamA Buffer and NaCl concentration IMAC screen**

A total of 1200 mg of BamA membrane was solubilised overnight at 4°C on a rotating wheel at 15 rpm in a total of 12.5 ml 1 % DDM. 1 ml of overnight lysate and a 1 ml TALON™ resin bed volume was set up in either TBS, phosphate buffered saline (PBS), 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), or 2-(N-morpholino)ethanesulfonic acid (MES) with varying salt conditions (150 mM, 300 mM, 500 mM) in the pulldown screen. All samples were made up to 5 ml with their respective buffer and salt conditions using stock solutions and were batch bound onto their respective TALON™ resin aliquots in 12 x 10 ml Pierce gravity flow columns. Samples were then washed with 15 CVs (15 ml) of their respective buffer/ salt condition. All samples were eluted in 3 CVs of their respective base buffer (TBS, PBS, MES, HEPES) supplemented to 150 mM NaCl and 250 mM imidazole. Samples during the elution process were analysed via SDS-PAGE.

2.2.3.6 **BamA 1 M salt and cofactor IMAC screen**

A previously eluted sample (5 ml at ~0.2 $mgml^{-1}$) was dialysed overnight in 8,000 Da MWCO dialysis tubing suspended in 0.03 % DDM buffer in a cold room at 4°C on a stirring plate. The resulting solution was split evenly into two and applied to 1 ml of pre-equilibrated TALON™ resin, made up to either 1 M NaCl or supplemented with cofactors (1 mM ATP, 10 mM MgCl2, and 1 mM MnCl2) and allowed to batch bind in the cold room at 4°C on a rotating wheel at 10 rpm for 2 hours. The flowthrough was collected by gravity flow, and the columns were washed with 20 resin volumes of wash buffer (TBS + 0.03 % DDM + 10 mM imidazole pH 7.0) adjusted to 1 M NaCl or supplemented with cofactors to mimic the binding buffer conditions. Both columns were then eluted with three resin volumes of elution buffer (TBS + 0.03 % DDM + 250 mM imidazole pH 7.0). 15 µl of the flowthrough, wash, and elutions were loaded onto a gel for analysis.

2.2.3.7 **Generation of *ΔbamA* and *ΔbamD* competent cells**

The *ΔbamA* and *ΔbamD* constructs were grown overnight in LB broth supplemented with 50 µg/ml kanamycin (*ΔbamA*)/ 100 µg/ml ampicillin (*ΔbamD*) and 0.05 % arabinose. This overnight culture was used to inoculate 100 ml LB with the appropriate antibiotic, but lacking arabinose to deplete protein production. The cultures were grown until an optical density (OD) at 600 nm of 0.4 and competent cells were made as per $CaCl_2$ and $MgCl_2$ chemically competent cell protocol, aliquoted, frozen in liquid nitrogen and stored at -80°C for later use.

2.2.3.8 **Generation of constructs for complementation assay**

1 µg of the pET26b::*bamA* construct and 1 µg of pMH1 was digested with NcoI and EcoRI, the cut pET20b vector and pET26b::*bamA* insert were recovered by gel extraction, ligated and transformed into Neb5α using the standard cloning procedures. The putative transformants were screen by colony PCR with the TEV forward and *bamA* reverse primers. Five colonies were selected and used to inoculate 5 ml overnight cultures. Plasmid preparations were made, screened by Sanger sequencing and the plasmid selected for the *ΔbamA* strain supplementation. As the *ΔbamD* strain has ampicillin resistance, the pET26b::*bamD* construct (kanamycin resistance) made previously was suitable for this assay, and no further construct needed to be made.

2.2.3.9 **BamD sucrose gradient assay**

2.2.3.9.1 Preparing large unilamellar vesicles

Assemble mini-extruder (Avanti Lipids) components according to manufacturer's instructions and plug in extrusion heating block (Avanti Lipids). Set extrusion heating-block to just above the gel phase-transition temperature for the lipid being used. Check syringes for leaks and wash syringes with buffer. Fill one syringe with buffer and leave the other empty. Insert both syringes into the Teflon O-rings, pass buffer through once to check extruder components for leaks, and remove sample buffer. Add lipid suspension mixture to the same gas tight syringe as before and reinsert into the O-ring. Place syringe/ O-ring assembly onto the extruder heating block for 5-10 minutes to allow lipid suspension mixture to equilibrate to above the phase transition temperature (for DOPC this step is not required as gel to liquid crystalline phase transition temperature is -17°C). Gently push the plunger of the

filled syringe until the entire sample has passed through to the second (empty) syringe. Gently push the plunger of the newly filled syringe until the entire sample has passed back into the original syringe. Repeat this back and forth process a minimum of four times (for ten membrane passes total), and then carry out one final pass to the (originally) empty syringe to prevent any of the starting mixture from being present in the final large unilamellar vesicle (LUV) sample.

2.2.3.9.2 Sucrose gradient protocol

Prepare 60 % stock solution of sucrose, and dilute to generate 5-50 % sucrose stocks in 5 % increments with experimental buffer solution. Place sucrose stocks and centrifuge tubes on ice, and incubate for 30 minutes. Chill BamD sample and the LUV preparation for 10 minutes, mix to the desired concentration, and leave to incubate for 20 minutes. Chill centrifuge rotor. Using a cut tip, layer sucrose gradient from highest density to lowest density in two centrifuge tubes. Check tube weight and balance to within 0.01 g. Incubate tubes on ice for 10 minutes. Add 500 µl sample to one tube and label sample, and the equivalent weight in buffer to the other tube. Insert tubes into chilled rotor, and centrifuge at 150,000 x *g* for 17 hours at 4°C. Retrieve centrifuge tubes, and slowly remove 0.5 ml aliquots from the top down with the cut pipette tip from the meniscus of the sucrose gradient. Once all sucrose is collected, rinse tube once with 1 ml sample buffer, and run samples on SDS-PAGE for analysis.

2.2.3.10 Cross-species complementation assay

A standard transformation procedure was used in triplicate using the *ΔbamA* and *ΔbamD* competent cells stocks, pET20b::*bamA* and pET26b::*bamD* constructs, and pET20b+ (ampicillin^{R}) and pET26b+ (kanamycin^{R}) empty vectors to generate complementation data.

2.2.4 Circular dichroism

2.2.4.1 Preparation of proteins

Proteins were recombinantly expressed, extracted, purified and prepared as outlined earlier. The buffer component NaCl was incompatible with CD due to its absorbance in the experimental wavelength range and was exchanged by dialysis with NaF, which has lower absorption and scattering in CD experiments. Where detergent

buffers were required, the detergent concentration was reduced to approximately 1.8x CMC to lower the influence of the detergent absorbance on the spectra collected. However, this was not reduced to 1.0x CMC to allow for changes in detergent CMC during the thermal degradation (20°C-37°C), which is parabolic for a number of detergents with typical minima around 300-310 K (27-37°C), and require increasing detergent concentrations to maintain micelle formation the further temperature deviates is from this pivot point[117]. For these reasons, the concentration was set to 0.25 % SB12 (0.134 % is 1x CMC at 25°C) for the thermal degradation experiments. To this end, BamA samples were spin concentrated for the experiment using a 100 kDa molecular weight cut-off (MWCO) spin concentrator tube to ensure the excess SB12 micelles (micelle average size of 55 molecules or approximately 18.5 kDa) did not concentrate in the sample. BamA was then made up to volume with the detergent buffer (0.25 % SB12, 150 mM NaF, 30 mM Tris), and dialysed to 0.25 % SB12 over 72 hours in the cold room. Where mixtures of proteins were tested, the buffers were matched and diluted with the appropriate buffer for the mixture. For the BamD and Cj1289 CD experiments, this resulted in two separate preparations in the detergent (0.25 % SB12, 150 mM NaF, 30 mM Tris, pH 7.6) and non-detergent (150 mM NaF, 30 mM Tris pH7.6) buffers. BamD and Cj1289 preparations were concentrated in 10 kDa MWCO spin concentrators, the samples were split into two and dialysed against either the non-detergent buffer (for 3 hours, with 2 changes) or detergent buffer (for 48 hours, with 1 buffer change, total 96 hours) at 4°C in the cold room on a plate stirrer. Concentrations were checked post-dialysis using a UV/Vis spectrophotometer (NanoDrop™ 2000c, ThermoFisher) against the last dialysis exchange buffer.

2.2.4.2 **Annealing of *in silico* homology models**

Homology models were generated for BamA and BamD using I-TASSER[118], and the Cj1289 crystal structure model (3RGC) was obtained from the RCSB repository. These PDB files were prepared for MD simulations using the ‘Chemistry at Harvard Macromolecular Mechanics - Graphics User Interface' (CHARMM-GUI) web interface[119–121] with proteins placed into either aqueous (water and ions; default for soluble proteins), aqueous and detergent (water, ions and FC12 or SB12, as appropriate), or aqueous and lipid (water, ions, POPE:POPG 80:20 membrane[102]) environments, and files were parameterised using the CHARMM36 forcefields[122].

The simulations were carried out using NAMD[123] on the Midlands-HPC-Plus server for a minimum of 100 ns for all proteins in this study in each environment, and the last step was recovered and converted into a PDB file for SSA server submission.

2.2.4.3 *In-silico* predictions of secondary structure assignments

Prior to the experimental runs, *in silico* characterisations of protein secondary structure were made using PDBs of I-TASSER homology models of BamA and BamD, and the crystal structure model of Cj1289 (3RGC) or MD-annealed last-step PDB equivalents in multiple background environments, with analysis using secondary structure assignment (SSA) algorithms DSSP[124], STRIDE[125], PSEA[126], STICKS[127], and PROSS[128].

For full summary of SSA predictions for all proteins in all environments, see appendices (appendix 8.4.1). To relate these outputs to the output that the analysis software (CDapps) provides, the results have been collated into four categories: α-helical (3_{10}-, α-, π-), β-sheet (β-sheet, β-bridge), turn, other (unordered/ coil).

2.2.4.4 Generation of *in silico* predictions of secondary structure assignments

The homology models for each protein in the CD experiments were assessed prior to the experiments using SSA algorithms to gain an initial estimate of secondary structure. The secondary assignment algorithms used are DSSP, STRIDE, PSEA, STICKS and PROSS. DSSP, dictionary of secondary structure of proteins, gives eight assignments of secondary structure (3_{10}-helix [G], α-helix [H], π-helix [I], turns [T], extended β-sheet [E], β-bridge [B], and bend [S]) by analysis of hydrogen bonds (which are defined by an energy lower than -0.5 kcal/mol). STRIDE, "STRuctural IDEntification" method, uses the same categories but defines structure by analysing hydrogen bonds and classification of backbone geometry torsion angles using a Ramachandran map. P-SEA assigns secondary structures by analysing the Cα dihedral angles in the protein and creating a two angle criteria for each secondary structure[126]. STICKS finds a set of line segments without externally defined secondary structure, and uses the segments as a novel basis for secondary structure definition, taking the average rise/residue from a set of line segments along each axis to characterise the segment. This has the advantage that secondary structure elements are described by a continuous value and thereby not restricted to standard structure

classes[127]. PROSS makes secondary structure assignments based on backbone dihedral angles and their plot on a Phi/Psi grid in torsion angle space[128].

2.2.4.5 **Chirascan spectra data collection**

To investigate the folded state of the recombinantly expressed proteins, samples were delivered to Diamond Light Source to undergo far-UV circular dichroism experiments. The experiments could not be attended in person due to external influences (Covid), and an experimental run sheet was provided to the beamline scientist to perform the experimental runs (appendix 8.2.3).

Initial CD spectra were obtained for all buffers and protein samples at 25°C ± 0.25°C to preview attainable data quality. CD spectra were obtained for each thermal degradation experiment for all buffers and protein samples from 20°C to 89°C in increments of 3°C, with a two-minute equilibration time at each temperature increment. An additional measurement was taken at 20°C post-experiment after samples gradually cooled over a 5-10 minute period (unmonitored). In all experiments, data collection was from 260 nm to 180 nm in stepped ramp mode (data pitch, 1 nm; bandwidth, 1 nm; slit, 3 mM; response time, 25 µs; integration time, 1 second [individual proteins]/ 2 seconds [protein mixtures]) using a 0.01 cm path length quartz cuvette cell and a Chirascan Plus CD spectrometer (Applied Photophysics, U.K.). The instruments were routinely calibrated with d-camphor sulphonic acid for optical rotation and wavelength by the staff at Diamond Light Source B23 Beamline. Reported spectra have been averaged at each point, and corrected by subtraction of the buffer CD spectrum.

2.2.4.6 **Thermal unfolding**

The following of thermal unfolding, given at an appropriate wavelength, has been initially estimated using a Boltzmann two-state exchange model[129]:

$$A + \frac{(B-A)}{1+\frac{exp(x-x_0)}{\delta x}} \qquad \text{(Equation 2.1)}$$

Where x is the temperature, x_0 is the melting temperature, and δx is the width of the thermal transition using an in-built non-linear square fitting algorithm in CDapps[130]. In all cases, Boltzmann fittings were carried out at the wavelength where the largest change in CD signal occurred throughout the temperature range of the thermal denaturation experiment for each sample. All secondary structure estimations were

performed using the CONTIN/LL[131, 132] algorithm and the SMP56 (SP43 + 13) CD spectra reference set, unless stated otherwise.

2.2.4.7 Predictions of mixture secondary structure estimations

To determine whether the changes in the secondary structure estimations observed in the mixtures relative to the individual proteins signify interactions or are the result of a simple non-interacting mixture of the two particles, the individual secondary structure estimations were added together in a 1:1 stoichiometry with equal weighting. These plots were then compared to the mixture secondary structure estimation profiles to see if the secondary structure estimations were due to an interaction between the proteins or the result of mixing two non-interacting proteins together.

2.2.4.8 Protein concentrations, molecular weights, and extinction coefficients

The protein concentrations were measured prior to the experiment on a NanoDrop 2000c UV-Vis spectrophotometer (Thermo Scientific) at 280 nm using 2 µl sample, measured in triplicate (average value shown), washing the measurement pedestal with buffer between measurements. The protein molecular weights and extinction coefficients were calculated from the primary sequences using the Expasy ProtParam service. The concentration measurements and calculations are shown (table 2.21).

	Measurement average ($mgml^{-1}$)	Molecular weight (kDa)	Extinction coefficient ($\varepsilon/1000$)
BamA	0.811	84.75	98.56
BamD	0.903	25.36	36.33
Cj1289	0.919	32.52	17.88

Table 2.21: Concentrations, molecular weights and extinction coefficients of proteins used in this study

2.2.5 Small angle X-ray scattering

Proteins were recombinantly expressed, extracted, purified and prepared as outlined earlier. The experiment was planned using the information available on the Diamond

Light Source Beamline 21 website (https://www.diamond.ac.uk/Instruments/Soft-Condensed-Matter/small-angle/B21/Planning.ht ml).

All samples were made up in 96 well plates to 2 % glycerol and to 1 mM dithiothreitol (DTT) on ice, foil sealed and stored at 4°C in the fridge overnight before sample drop off on ice. Buffers for individual protein runs were prepared from spin concentrator flow through or dialysis buffer of each protein. Buffers for mixtures were prepared using either the dialysis buffer (where available, preferred) or by mixing the two buffers at the same ratios as the protein sample mixture.

2.2.5.1 **Sample preparation and instrumentation**

Samples were prepared in 96 well plates for high throughput SAXS in concentration gradients ranging from 2 mg/ml (where possible) to 0.125 mg/ml for each protein and 1 mg/ml to 0.25 mg/ml for each protein mixture to be tested. 30°C was chosen as the standard pre-test equilibration temperature as any potential models would be compared to homology models annealed by MD simulations at the standard practice *in silico* temperature of 303.15 K (or 30°C).

2.2.5.1.1 SAXS analysis

To analyse the scattering data, the raw SAXS data was imported into ScÅtter, a java-based application developed by Robert Rambo at Diamond Light Source (https://www.bioisis.net/scatters/download), and the standard processes built into the program were used to generate plots on the analysis and P(r) tabs (see https://www.bioisis.net/tutorials/1 for details).

2.2.5.1.2 Assessment of sample quality

Samples were initially assessed by plotting scattering data on $\log_{10}$ and Power-law plots. These plots highlight the low-q and high-q range, respectively, and the scattering data should show similar profiles for all concentrations tested for a single protein on each plot. The Kratky plots were used to determine whether the samples were folded, and whether there is a degree of flexibility to their structure. The flexibility plots were used in a similar vein to gain semi-quantitative information on flexibility and to gain the parameter q_{max}. Guinier fittings were then carried out to assess the degree of aggregation and determine whether the samples were amenable to further analysis. Where the Porod region of the SAXS curve was of sufficient

quality, the Porod exponent was derived and a rough calculation of molecular weight was made.

The profiles were then compared to the homology models generated computationally for each protein, using the base homology model generated by I-TASSER[118] or Phyre2[133], and a MD simulations annealed model to assess whether the SAXS curve obtained is reflective of expectations for the structure and environment of the sample. The highest concentration sample was used for all analyses of the proteins in this study in each instance unless stated otherwise.

2.2.5.1.3 Comparison to homology models

To determine whether the SAXS data curve obtained is justified by the sample, a homology model for the protein was made (with and without annealing in a suitable background environment via MD simulations, where available) and compared using ScÅtter to the experimental data curve obtained for each protein. The experimental data was also compared to the homology models using the FoxS server[134, 135] where any dummy atom models were generated from the experimental P(r) obtained.

To prepare the homology models (where the models had been annealed using MD simulations), the appropriate Debye length had to be determined for the hydration shell around the sample. For further information on the Debye length calculation, see methods (2.2.6.9).

2.2.5.1.4 Bead Modelling

A bead model is a representation of the particle in a defined search volume through a collection of beads, with each bead belonging to either the particle or solvent, which produces a similar scattering pattern to fit the experimental data. Where a suitable P(r) function was determined for a SAXS data set, a dummy atom or bead models were produced using DAMMIF[136] (ATSAS *ver.* 3.0) via the ScÅtter model tab. The DAMMIF models were then sorted and superimposed to identify outliers using DAMSEL (via SUPCOMB[137]) through minimization of atom vector distances using sequential model pairings. The selected models are then averaged using DAMAVER[138] to generate a probability map, and either filtered using DAMFILT[138] or refined using a combination of DAMSTART[138] (to generate the input file) and DAMMIN[139] (to refine the model). The final models were then manually superimposed onto the available homology models to gauge the goodness of fit.

2.2.5.2 **Complexation analysis**

The complex assessment function in ScÅtter is used to test whether a complex is formed, and works by carrying out a linear addition of the individual spectra at different concentrations to make an assessment on whether a complex has been formed in the mixed sample. This is reflected in the both the projected overlay on the scattering curve and in the residuals shown. A near perfect fit to the scattering curve and a linear residual (or one with an even distribution of data points) throughout indicates that no interaction has taken place, and an imperfect SAXS curve fitting and/or skewed residual indicates formation of a complex. This can be observed throughout the q-range sampled, however, at low-q, as our samples in this study show aggregation, it is difficult to resolve whether any skewed residuals in this region are due to aggregation or complex formation. As such, the low-q region (below 0.02 in this instance) has been omitted from our analysis and whether a mixture of proteins forms a complex will be gauged based from the residuals from the remaining low-q, mid and high-q data. In addition, the data has been trimmed back in each instance to remove excessively noisy data in the high-q range, which may make interpretation difficult, with all samples shown within q-limits of $0.2 \leq q \leq 0.28$.

2.2.5.3 **Data collection**

The beamline parameters for SAXS experiments in this study are shown (table 2.22).

Instrument/ data processing	Instrument: B21 Data processing: Dawn (https://dawnsci.org/)
Wavelength (Å)	13.1 KeV (~0.946 Å)
Beamsize (µm)	Spherical 50 µm full width at half maximum at the detector, 1 x 0.25 mm at the sample
Camera length (m)	3.9 m
q measurement range ($Å^{-1}$)	0.0045-0.35
Absolute scaling method	Scaled to water scatter
Normalization	Normalised to an integrating photo diode on the beamstop

Monitoring for radiation damage	Monitoring changes on multiple short exposures
Exposure time	28 x 1 second exposures for BioSAXS
Sample configuration	Sample is measured flowing through an in-vacuum 1.5 mm capillary
Injection volume (µl)	40 µl
Capillary type/dimension/ manufacturer	Hilga 1.5 mm diameter, 10 µm wall thickness, silica glass

Table 2.22: SAXS data collection beamline parameters

2.2.6 *In silico* docking and molecular dynamics simulations

2.2.6.1 Generation of homology models

As no crystal structures are available for *C. jejuni* BamA, BamD, DegP or LptD, all structures were modelled *in silico* by comparison to existing homologous templates (template modelling) or by recognition of peptide motifs (ab-initio modelling). The initial models were generated from the primary sequence using I-TASSER[140] (https://zhanglab.ccmb.med.umich.edu/I-TASSER/) without restraints for both BamA and BamD. Models were selected using the internal model rating system (confidence score, C-score), with the highest scoring model taken forward for docking. A second modelling server, Phyre2[133], was used to cross-check I-TASSER models. Later in the project, the protein models were further verified with the 14th critical assessment of structural prediction competition winner, AlphaFold 2[112, 113].

2.2.6.2 Discovery studio/ UCSF Chimera

Discovery studio and UCSF Chimera[141] were used to generate the homology model figures for this study.

2.2.6.3 ZDock

ZDock[142] is a webserver used to generate docked protein complexes from two input models. Models were submitted to the ZDock server without restraints, or with all residues except the restraints identified by superimposition of *C. jejuni* BamAD with homologous *E. coli* models. The first hundred models were generated from the

transformation matrix and submitted models, and were screened for the appropriateness of the model. Where BamD was directly bound to the β-barrel, the docking model was non-permissible as there is no evidence of BamD permeating or invading the membrane. Other models that were inappropriate were those that had singular binding points as there was no prior information to suggest these were important binding states for the BAM complex mechanism. Residues in contact were analysed with a distance restraint of 5 Å using UCSF Chimera or Discovery Studio.

2.2.6.4 **pyDock**

pyDock or pyDockWeb[143] is a protein-protein docking webserver requiring a ligand and receptor input, with the option for the use of restraints. The first docking run uses no restraints, and the second run applies the identified residues from superimposition of the *C. jejuni* and *E. coli* BamAD structures as distance restraints (within 6 Å). The first hundred models were generated, manually screened to find appropriate docking configurations, and the top two results were chosen. Residues in contact were analysed with a distance restraint of 5 Å using UCSF Chimera or Discovery Studio.

2.2.6.5 **RosettaDock (Docking 2)**

The ROSIE Docking 2[144] (RosettaDock) webserver (https://rosie.graylab.jhu.edu/docking2) performs a local docking search of conformations near the given starting conformation of the Protein Data Bank (PDB) input files and was used to perform local docking on BamA and BamD. A combined PDB was generated using the I-TASSER generated homology models for BamA and BamD in Discovery Studio, chains were edited using a PDB file editor and assigned different chain identifiers. The homology models were separated by 13 Å and placed in a configuration similar to the expected docking confirmation, saved as a PDB file, and used as the input to the Docking 2 server for local protein-protein docking. The top ten results alongside the energies were analysed, and two representative models were chosen. Residues in contact were analysed with a distance restraint of 5 Å using UCSF Chimera or Discovery Studio.

2.2.6.6 **Editing of PDBs**

Where docked homology models list both proteins under the same 'ChainID' it is necessary to assign separate chain identifiers to allow programs to recognise each protein as a separate object (i.e. CHARMM-GUI, UCSF Chimera, VMD). The program 'Protein Data Bank (PDB) File Editor', available from source forge (https://sourceforge.net/projects/pdbeditorjl/), was used for this purpose and chain identifiers were renamed to A (BamA) and D (BamD) prior to simulation set up.

2.2.6.7 **Protonation state**

The protonation state of homology models were assessed prior to local environment assembly using CHARMM-GUI using the PDB2PQR[145] web server. The PDB2PQR webserver uses the PROPKA[146] algorithm to calculate the pKa for each residue in the protein and then assigns a protonation state based on the user-defined pH, which for CHARMM-GUI assemblies is pH 7.0, so this was selected for all homology models. In addition, the webserver checks the PDB residue by residue for errors and inconsistencies (e.g. missing hydrogens) and attempts to repair these issues, all of which are output on a log of the process alongside any changes to protonation state.

2.2.6.8 **CHARMM-GUI**

Simulation set up was performed using the CHARMM-GUI webserver (https://charmm-gui.org/) solvator or membrane builder input menus. Where membranes were added to the simulation to encapsulate the protein, the protein PDBs were aligned by vector to take a set orientation for membrane placement (table 2.23). Where a vector has not been listed for a particular simulation, then the positioning of proteins in membranes (PPM)[147] webserver was used to predict of the protein in the PDB was used.

	Vector alignment	
Simulation	Residue 1	Residue 2
A0_D0/ D1/ D2	Gly480	Lys544
A90_D0/ D1/ D2	Val 468	Ile543
A180_D0/ D1/ D2	Gly422	Lys456
A270_D0/ D1/ D2	Leu665	Met697

Table 2.23: Vector alignments used in PDB set up to add membrane bilayer

2.2.6.8.1 Membrane builder

Membrane builder is one of the solvation options available on CHARMM-GUI and was used to assemble an integral membrane protein into a membrane bilayer and place this assembly in a water box. The process consists of six steps, wherein the protein coordinates are first (1) read from the PDB file, (2) the protein is orientated, (3) the lipid or detergent composition is assigned in a defined unit cell, (4) the components are built onto the original PDB, (5) the system is assembled and (6) the equilibration and trajectory production scripts are generated. The process is semi-automated and requires input from the user at the beginning of each step. In the first step, the PDB file for the protein is loaded onto the CHARMM-GUI interface and it was checked whether all residues are being read correctly. In this study there were few modifications that need to be applied to the proteins, so the default options were chosen for modifications in step one. The membrane proteins were orientated in step two using the "align by vector" orientation method, and two residues were picked to assign the orientation vector. In addition, pore water was generated, as this would likely exist in the system *in vivo*. In step 3, the box type, the water thickness, ratio of lipid components and system size were defined. In all cases where lipid is used in the system a 80:20 ratio of 1-Palmitoyl-2-Oleoyl-sn-Glycero-Phosphoethanolamine (POPE): 1-Palmitoyl-2-Oleoyl-*sn*-Glycero-3-Phosphogylcerol (POPG) is used to represent a simplified membrane (values present in *E. coli*[101, 102]). The system was then packed and checked for correct protein insertion into the membrane (correct orientation, membrane placement, and system size), and the lipid placement method and ion concentration, type, and placement method were chosen. In the remaining steps, the components were then built into the system, and the file structure options (i.e. NAMD) and equilibration options (automatic grid generation, NPT ensemble, temperature) were chosen. The input files generated were downloaded and then uploaded onto high performance computation systems to carry out equilibration, energy minimization and trajectory runs.

2.2.6.8.2 MD docking runs

For the MD docking runs the membrane builder was used to assemble the combined BamA/BamD PDBs into solvated systems. A membrane patch size was estimated to be around 140 Å through iterative assembly attempts until the proteins lay within the patch size constraint. The TIP3 water molecules were added to this membrane

assembly by the distance placement method to assemble a water box, and the KCl ions were then added by the replacement method before full system assembly. The scripts for the assembled systems were then generated using the CHARMM36m force fields, automatic grid generation for particle mesh Ewald (PME) fast Fourier transformation, and isothermal-isobaric (NPT) ensemble options. Scripts were then generated for execution of simulation using NAMD at 303.15 K.

2.2.6.8.3 Solution builder

CHARMM-GUI solution builder is used to assemble soluble proteins in a water box for MD simulations. It is a simplified set up compared to membrane builder consisting of 3 steps, (1) read PDB coordinates, (2) solvate in water box, (3) build components into system, assemble and output input files.

2.2.6.8.4 Micelle builder

Where detergents are needed to be added to the system the micelle builder solvation option is chosen instead of membrane builder, both set ups share the same overall steps to assemble the system and generate the input files. In this mode of solvation, a set number of detergent molecules was chosen in step 3 rather than a membrane patch size. The number of detergent molecules required were calculated by: (1) measuring the β-barrel width in Discovery Studio using the monitor distance function and modelling the β-barrel as an elliptical cylinder, (2) calculating the surface area minus the area for the two ellipses, and (3) measuring the diameter of a single SB12 molecule (CPK representation) within an empty SB12 micelle. The surface area from (2) was divided by the diameter of 4 Å obtained in (3) to give an estimate of 148 SB12 molecules to fit around a BamA β-barrel and 250 SB12 molecules to fit around the LptD β-barrel. The final detergent molecule placement was checked to ensure there were no obvious gaps.

2.2.6.9 **Debye length calculation**

MD simulations come at a cost in CPU hours, and it would be beneficial to keep this at a minimum. The MD docking experiments attempt to expedite the timeframe required for docking interactions through modulation of the Debye length, the distance within which electrostatic interactions are permissible/ the extent of shielding of the molecules, by reducing salinity of the system, which in turn increases the Debye length and allows interactions to occur over a larger distance.

Increasing the Debye length allows BamD to be placed further away from BamA in the combined PDB/ start point of the simulation, potentially allowing for more extensive sampling or rearrangements prior to protein-protein interaction.

The permittivity of the solvent of the system (water) was referenced from previous work on the dielectric constant of water for 303.15 K (or 30°C). Alteration of the salt concentration of the system, from 0.15 M to 0.05 M, was used to adjust the Debye length so that interactions could occur from a longer distance (13.5 Å instead of 7.8 Å at 303.15 K), increasing the distance for sampling and rearrangements to approximately 8 Å for sampling and rearrangements to occur. The proteins residues are considered to be interacting when they are within a contact distance of 5 Å.

The equation used to calculate the Debye length is as follows:

$$\delta = \frac{1}{\kappa} = \left(\frac{\varepsilon\kappa T}{e^2 \sum_i c_i z_i^2}\right)^{\frac{1}{2}} = \left(\frac{\varepsilon\kappa T}{e^2 N_\alpha 2C}\right)^{\frac{1}{2}} \qquad \text{(Equation 2.2)}$$

Where C is concentration in moles per litre, e is the electron charge, κ is the Boltzmann constant (1.380649×10^{-23} $J\cdot K^{-1}$), Nα is Avogadro's constant (6.02214076×10^{23} $mol^{-1)}$, and ε is the permittivity of solute at a given temperature T in Kelvin.

2.2.6.10 **MD Force fields**

All simulation were performed with CHARMM(34+) force fields, with earlier simulations using CHARMM34 and later simulations using CHARMM36.

2.2.6.11 **Molecular dynamics simulations**

All systems were equilibrated for 750 ps at 1 fs/step and a further 750 ps at 2 fs/step as an isothermal-isovolumetric canonical ensemble (NVT) on a Supermicro server with NVIDIA K80 and K40 general purpose graphical processing units. The equilibrated systems were moved to the HPC-Midlands-Plus server, where the systems were tested for scaling to determine the appropriate core allocation prior to production runs. All atom MD simulations productions runs were carried out at 2 fs/step at 303.15 K under NPT ensemble using Langevin piston control. All runs were carried out on Intel Xeon E5-2680v4 chipsets under account A14-0036 using NAMD 2.12. MD docking production runs were carried out for 100 ns per system.

2.2.6.11.1 MD Scaling runs

Scaling runs were performed to determine the most cost-efficient resource allocation using either 256, 504 or 756 core allocations for 1 ns production runs for all MD runs simulated. The CPU hours (or wall clock times) were plotted graphically as seconds per atom per nanosecond production ($s^{-1}atom^{-1}ns^{-1}$) for each scaling run (appendix 8.6.3). Theoretically, there should be a doubling in speed when generating a set length trajectory when cores are doubled for the run (i.e. clock wall time reduced by half). The 1 ns wall clock times were extrapolated for 50 or 100 ns and the most efficient core allotment in terms of CPU hours was chosen. Where the wall clock for a run exceeded 100 hours, the trajectory length was reduced or the core allotment increased to fit the user limits assigned by HPC-Midlands-Plus. In practice, where there was a ≥80 % speed up upon doubling the cores the higher core allotment would be accepted.

2.2.6.12 **Analysis of molecular dynamics simulations**

Analysis of MD simulations was performed in the Visual Molecular Dynamics (VMD) software package *v*1.9.3 (available from https://www.ks.uiuc.edu/Research/vmd/).

2.2.6.12.1 Trajectory concatenation

Trajectories were concatenated using VMD by matching DesignCAD (DCD) trajectory output file frequencies of the 1 ns runs to the 50 ns production runs for all systems, and written to new DCD files. The concatenation was carried out by trimming the 1 ns DCD file (10 frames out of 500 kept to match DCD frequency of 50 ns production run) and adding all frames from the 50 ns production run(s) by sequentially loading the trajectory DCD file types and the protein structure file (PSF) for that system. The combined trajectory was then saved to a new DCD file.

2.2.6.12.2 Calculation of trajectory root mean square deviation (RMSD)

The root mean square deviation (RMSD) is a measurement of the differences between a starting or reference structure and a second structure. In the case of this study, it is a comparison of the starting MD annealed structure prior to generation of a trajectory and the resulting structure after a 50 ns production run. The RMSD from the starting point in the trajectories were calculated using VMD utilising the following formulae for the calculation of the RMSD over a set trajectory length:

$$RMSD = \left[\frac{\sum_{i=1}^{N} w_i \|x_i - y_i\|^2}{N \sum_{i=1}^{N} w_i}\right]^{\frac{1}{2}} \quad \text{(Equation 2.3)}$$

2.2.6.12.3 Calculation of residue root mean square fluctuation (RMSF)

The root mean square fluctuation (RMSF) calculation measures how much a particular residue changes or fluctuates over time or the trajectory length. The RMSF was plotted as a function of residue number, and indicated which residues contribute to a given molecular motion. The formulae for the calculation are:

$$RMSF = \sqrt{\frac{1}{T}\sum_{t_j=1}^{T}(x_i(t_j) - x_i^{ref})^2} \quad \text{(Equation 2.4)}$$

2.2.6.12.4 Visualisation of trajectories and protein-protein contact interfaces

All trajectory outputs were visualised using VMD (*ver.* 1.9.3). RMSD and RMSF were calculated using VMD, and the final frame of each trajectory output were saved as PDB files for analysis of the contact interface. Where necessary, the trajectories were unwrapped from their periodic boundary conditions (PBC) using the in-built PBC tools plugin before RMSD and RMSF calculation. This removed trajectory drift across the periodic boundaries. The unprocessed data for RMSD (appendix 8.6.6) and RMSF (appendix 8.6.7) are shown for reference. The final frame PDB files were analysed for residues in a set contact distance on UCSF Chimera (*ver.* 1.15).

3 Cloning, expression, purification, screening, and optimisation

3.1 Chapter aims

At the start of the project, the pMH1 construct and knowledge regarding of the stability of *C. jejuni* BamA (CjBamA) in SB12 was inherited from previous work in the lab on this project[148]. However, a poor initial result in preliminary CD thermal degradation experiments featuring BamA, and unstable expression, yields and degradation of *C. jejuni* BamA (pMH1 construct; data not shown), led to doubts regarding the construct, which was later identified to contain 16 in-frame stop codons by Sanger sequencing. The project was subsequently restarted nine months into the PhD, and ambitiously restructured to structurally characterise several proteins within *C. jejuni* BAM-mediated OMP biogenesis pathway and test protein-protein interactions using biophysical and computational approaches. The following chapter describes the details of the molecular cloning, expression testing, purification optimisation, detergent screening, and other optimisation work to prepare the target proteins for biophysical characterisation.

3.2 Results

3.2.1 Cloning

3.2.1.1 Homology searches and target selection

Homology searches were performed using BLAST and BLASTp on *E. coli* (*bamA, bamD, surA, skp, degP, lptD* and other genes) or *C. crescentus (bamF*) DNA sequences and protein translations against the *C. jejuni* taxonomic identifier. Homology sequences were found for *C. jejuni bamA* and *bamD*, but no homology sequences were found for *E. coli bamB, bamC, bamE or bamF* in *C. jejuni*, which coincided with findings in the literature regarding BAM subunits in ε-proteobacteria[36]. In addition, homology sequences for *degP*, *lptD*, and two *surA*-like genes (*cj1289* and *peb4*[108]) were identified, of which *cj1289* features in this project as Kale *et al.* (2011)[108] suggest Cj1289 is the more structurally similar protein to SurA, but no homologue for *skp* or for a *skp*-like gene has been identified, and it appears this chaperone is not a part of the *C. jejuni* OM transport system.

3.2.1.2 **Determination of the signal peptide leader sequence**

The primary sequences for the proteins in this study were submitted to the SignalP[149, 150] and/ or LipoP[151] webservers to determine leader peptide sequences, which were removed in the primer design as an *E. coli pelB* leader peptide sequence was vector-encoded. The cleavage sites were determined to be between positions 17-18 for *C. jejuni* BamA (CjBamA), BamD (CjBamD), and Cj1289 (figure 3.1). For *C. jejuni* DegP (CjDegP) and LptD (CjLptD) multiple potential cleavage sites were found (figure 3.2). The highest probability cleavage site for CjDegP (position 16-17; figure 3.2(b)) and the prior of the equally probable cleavage sites (position 17-18; figure 3.2(d)) for CjLptD was chosen based on LipoP server outputs and with consideration of the CjBamA, CjBamD and Cj1289 cleavage site predictions.

3.2.1.3 **Constructs**

Primers were used to isolate the gene sequences (methods, table 2.9) from *C. jejuni* NCTC11322 and several of the identified genes were cloned into the multiple cloning site of a pET26b+ kanamycin resistance vector (appendix 8.3.1). The corresponding protein sequences were homology modelled using I-TASSER and equilibrated using MD simulations to approximate structure and determine whether C- and N-terminal tag placement of Tobacco Etch Virus (TEV) and poly-Histidine (His-) tags was suitable for each protein (data not shown).

N-terminal (10His-TEV) tag placement was suitable for CjBamA, avoiding the possibility of the tag becoming unavailable by dipping into the C-terminal β-barrel. Other considerations were whether a C-terminal tag would interfere with the β1-β16 junction, unfold or destabilise the β-barrel. Similar considerations were made for the membrane protein CjLptD, and homology models indicated the tag should be C-terminally placed. For CjBamD, the tag was added at the C-terminal end, as the extreme C-terminus does not appear to be directly involved in the BamA-BamD binding interaction in *E. coli* (5AYW, 5D0O, 5D0Q, 5LJO). In the cases of the selected chaperones, Cj1289 and CjDegP, both C-terminal and N-terminal tags would likely be accessible and unlikely to affect function, so the first construct cloned was taken forward. All constructs were confirmed by Sanger sequencing, and translated protein sequences for the genes of interest (appendix 8.3.2) are available.

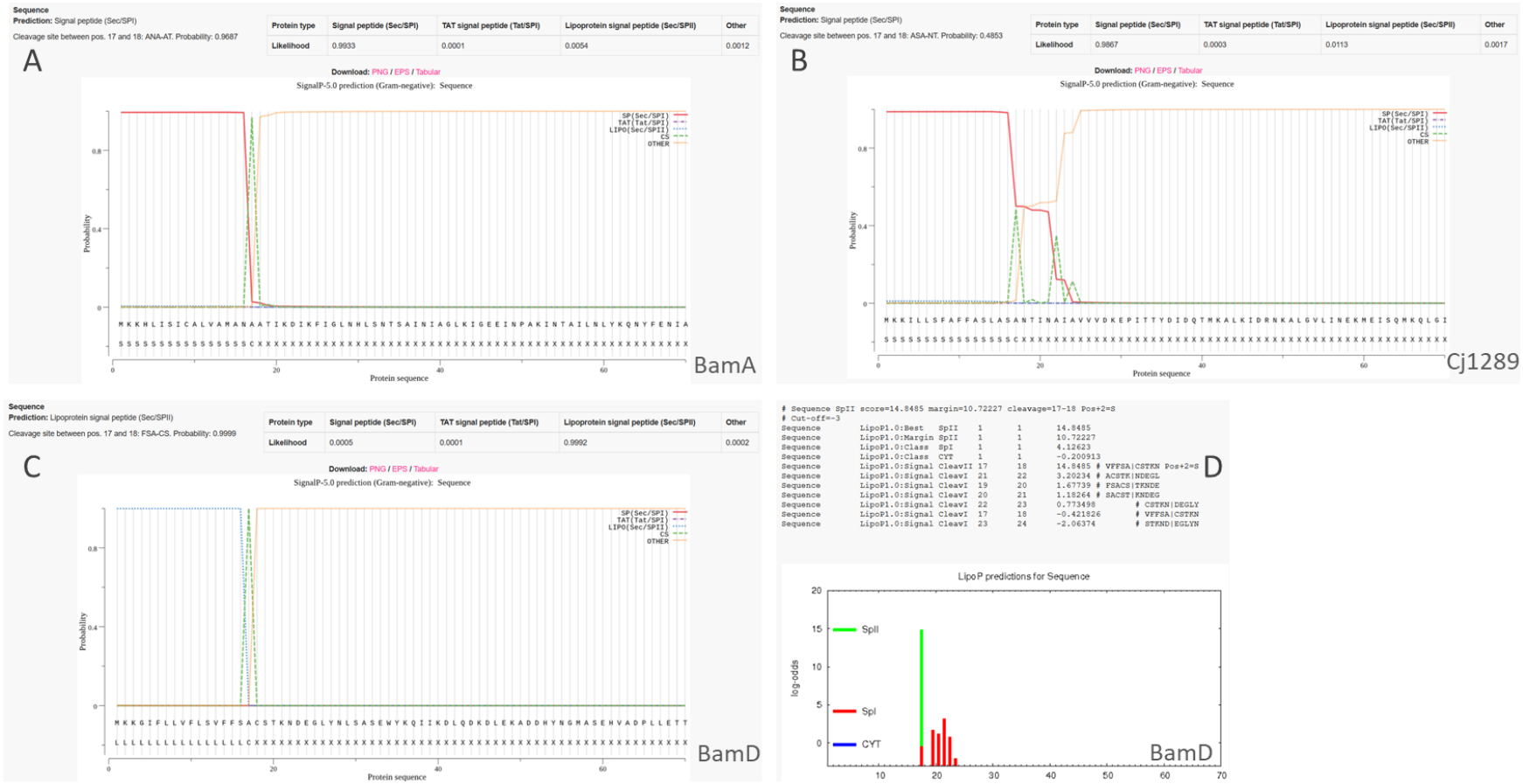


Figure 3.1: Signal sequence determination of *C. jejuni* BamA, Cj1289 and BamD. The protein sequences of BamA, Cj1289, and BamD were submitted to the SignalP server and predicted signal peptide cleavage sites were indicated between residue positions 17-18 for each of BamA (A), Cj1289 (B), and BamD (C), respectively. The LipoP confirmation of cleavage between position 17-18 is shown for BamD (D).

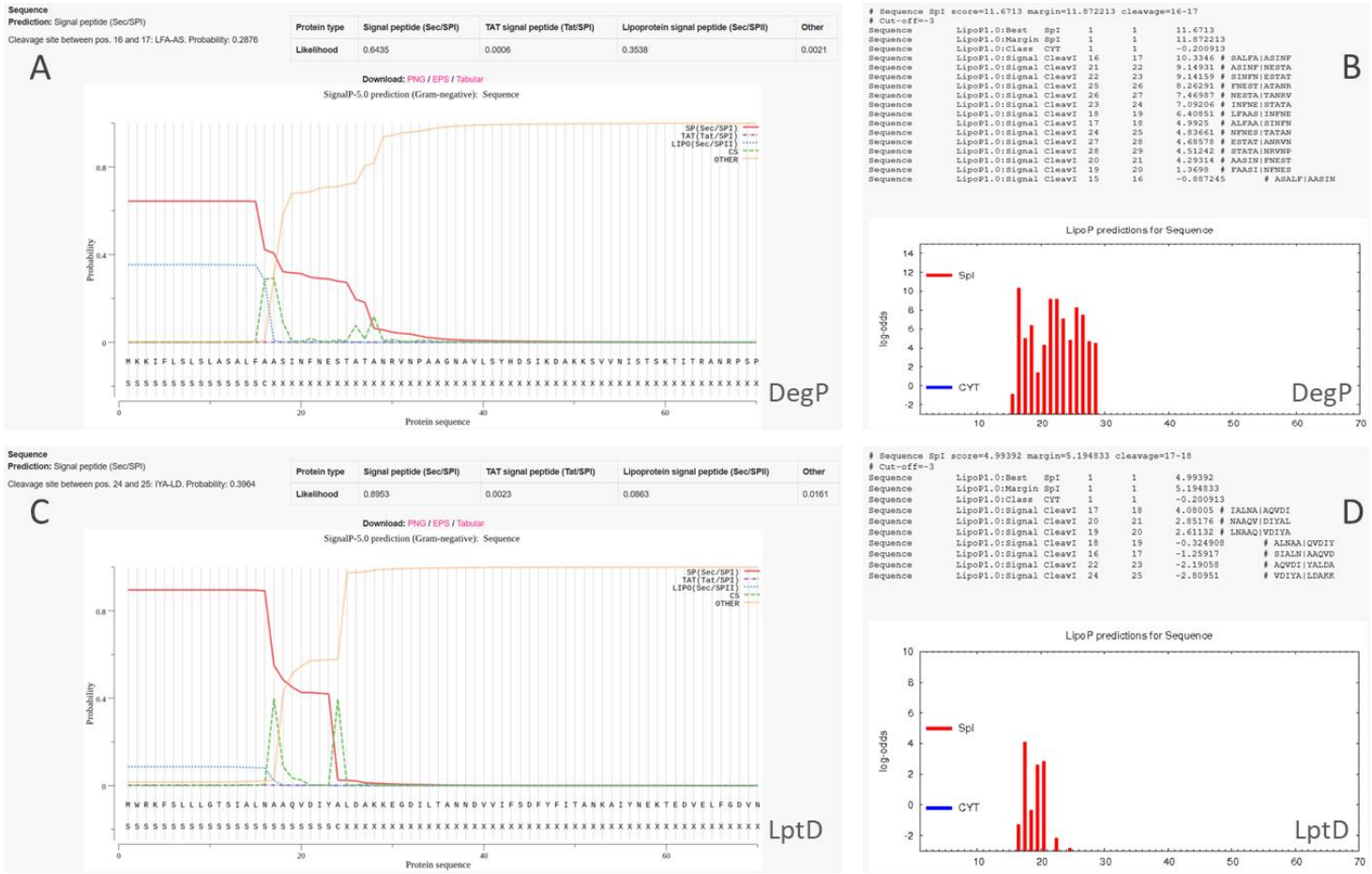


Figure 3.2: Signal sequence determination of *C. jejuni* DegP and LptD. The protein sequences of DegP and LptD were submitted to the SignalP and LipoP servers. Predicted signal peptide cleavage sites are indicated by SignalP between residue positions 16-17 for DegP (A) and 17-18/ 24-25 for LptD (C). LipoP predictions indicate optimal cleavage sites are between residues 16-17 for DegP (B), and 17-18 for LptD (D).

3.2.2 Confirmation of constructs

The expression of the target proteins was confirmed using a combination of Western blots (figure 3.3) and liquid chromatography tandem mass spectrometry (LC-MS/MS) of excised SDS-PAGE bands (appendix 8.3.3) which was outsourced to Oxford Mass Spectrometry Research Facility.

3.2.2.1 Western blots

A Western blot using a 'Penta-His' tag antibody was performed with post-expression immobilised metal ion affinity chromatography (IMAC) purified samples of the target proteins with bovine serum albumin (BSA) negative control (figure 3.3). On the initial Western blot signals were observed for CjBamA (84 kDa), Cj1289 (35 kDa), CjLptD (80 kDa) and CjDegP (55 kDa), but no signal was observed for BamD (25 kDa) or for the negative control BSA (as expected).

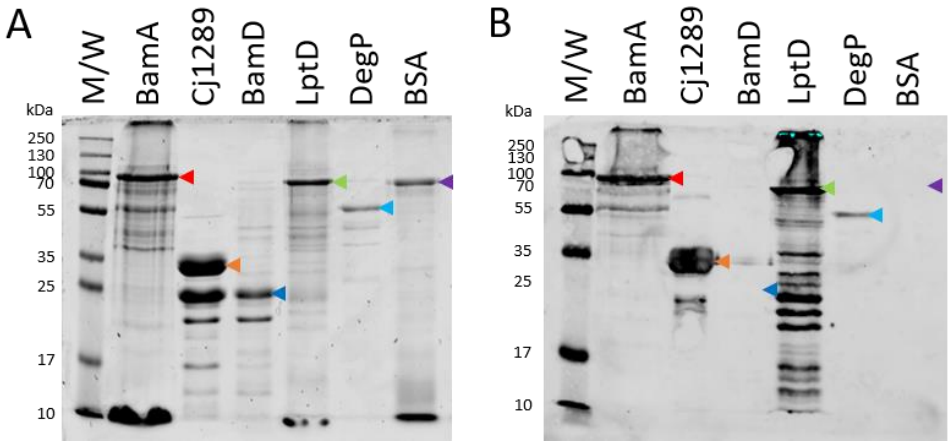


Figure 3.3: SDS-PAGE and Western blot analysis of IMAC purified proteins. IMAC purifications were run on two SDS-PAGE gels in duplicate, and stained with Coomassie (A) or transferred for Western blotting (B) against 'Penta-His' tag antibody. Target bands (or their expected positions) are indicated by arrowheads for *C. jejuni* BamA (red), Cj1289 (orange), BamD (dark blue), LptD (green), DegP (light blue) and for the negative control BSA (purple). Western blot signal is observed for all proteins at the correct target band size except BamD and BSA.

A weak signal was present for CjBamD on the Western blot and the blot was repeated using a tris-buffered saline (TBS) IMAC purified CjBamD sample, and a signal was obtained (figure 3.4).

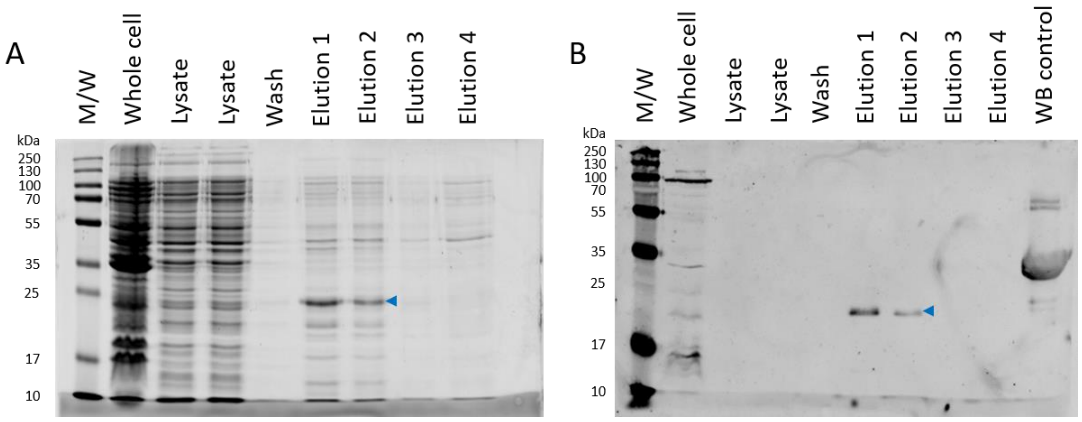


Figure 3.4: BamD repeat Western blot. A SDS-PAGE gel (A) and Western blot (B) were directly run after a CjBamD IMAC purification in TBS. The target band for CjBamD is indicated (blue arrowhead), and a signal is observed for the first two IMAC elutions on the Western blot (B).

3.2.2.2 **LC-MS/MS**

Proteins were submitted for LC-MS/MS analysis (appendix 8.3.3) against the MASCOT database as a further confirmation to the DNA sequencing and Western blots. None of the proteins in this study were positively identified when the peptide fingerprints were initially searched against the MASCOT (*E. coli*) database (data not shown), however positive identifications were found for all proteins when the peptide fingerprints were cross-referenced against the provided protein sequences on a Biotools bioinformatics search (appendix 8.3.4), with the exception of CjBamD, which was identified as cAMP activated global transcriptional regulator.

3.2.3 **Protein expression**

After assembly of the target gene into the pET26b+ vector backbone, the constructs were screened for expression under a large variety of conditions (appendix 8.2.1) in a systematic process to optimise expression. Once optimal conditions for protein expression were determined from the small-scale expressions, these were scaled up for validation at typical volumes for overexpression, and further optimised to determine optimal conditions (appendix 8.2.2).

3.2.4 Protein purification

Proteins in this study were purified using a combination of IMAC, anion exchange chromatography (AEX), and SEC FPLC purifications before concentrating in spin concentrator tubes with defined MWCO prior to experiments.

3.2.4.1 *C. jejuni* BamA

Expression and purification of the integral membrane protein CjBamA represented a major challenge in this project. After several iterations of expression optimisation yield was still poor (100-300 µg/L), but expression was stable and bulk overexpression and purification (in SB12) generated acceptable yields to accomplish planned CD and SAXS experiments. IMAC purification of CjBamA shows isolation of the 84 kDa target band and a major contaminant band around 55 kDa (figure 3.5(a)). The IMAC elutions were further purified by SEC on a preparative S200 column (on the low pressure FPLC available in lab) with 1.5 ml elutions of the target band broadly in fractions 1-18, and the majority of the contaminant bands eluting after fraction 15 (figure 3.5(b)). Fractions 10-14 were loaded onto an AEX column and eluted in 5 ml fractions, with isolation of the target band in fractions 19-20 around 26-30% buffer B (figure 3.5(c)). The sample prior to final spin concentration using a 100 kDa MWCO spin concentrator tube is shown (figure 3.5(d)).

Due to the particularly low yield of CjBamA, several further optimisations were made to the purification process, including IMAC resin choice (appendix 8.3.5, figure 8.9), resin binding (appendix 8.3.5, figure 8.10), purification buffer and salt concentration IMAC screening (appendix 8.3.5, figure 8.11), and further salt and buffer cofactor screening (appendix 8.3.5, figure 8.12) optimisations. The results from these were used broadly across the other protein preparations, and were repeated where necessary (data not shown).

After optimisation in SB12, CjBamA was further screened against twelve different detergents to assess how yield is affected by the extraction detergent and to determine appropriate detergents for structural characterisation by X-ray crystallography at the membrane protein laboratory (MPL) at Harwell.

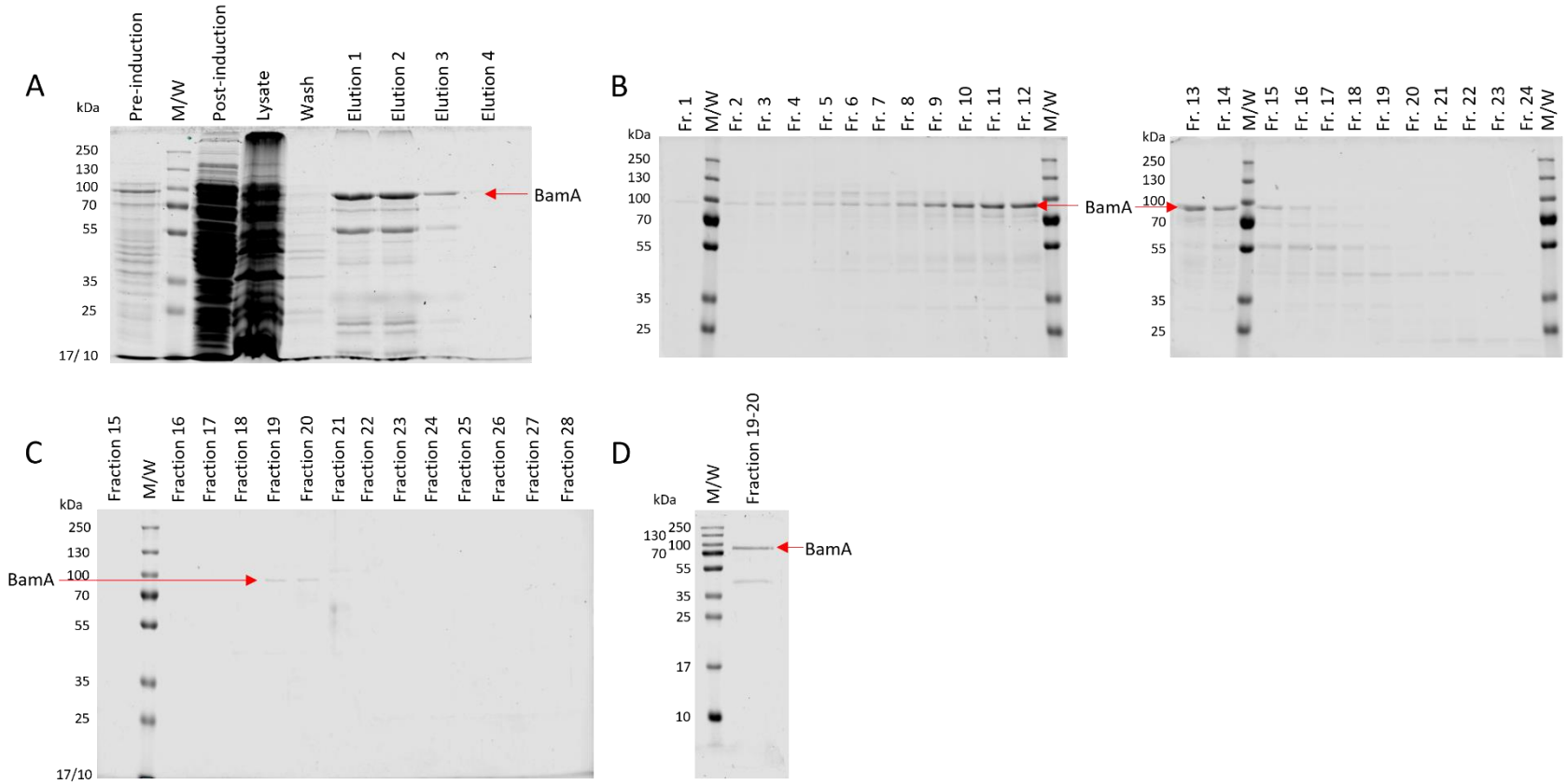


Figure 3.5: BamA purification. BamA was purified using a strategy of IMAC, SEC and AEX. BamA (A) IMAC Ni-NTA purification, (B) S200 column SEC purification, (C) Q-HP column AEX purifcation, and (D) collected fractions prior to final pre-experiment spin concentration (with a 100 kDa MWCO filter spin concentrator tube) are shown.

SB12 is absent from these screens as there are few successful examples of crystallisation using SB12 (MPL, personal communication). IMAC purifications of CjBamA in each detergent were run in parallel and the post-purification elution samples were analysed by SDS-PAGE and tryptophan fluorescence (figure 3.6).

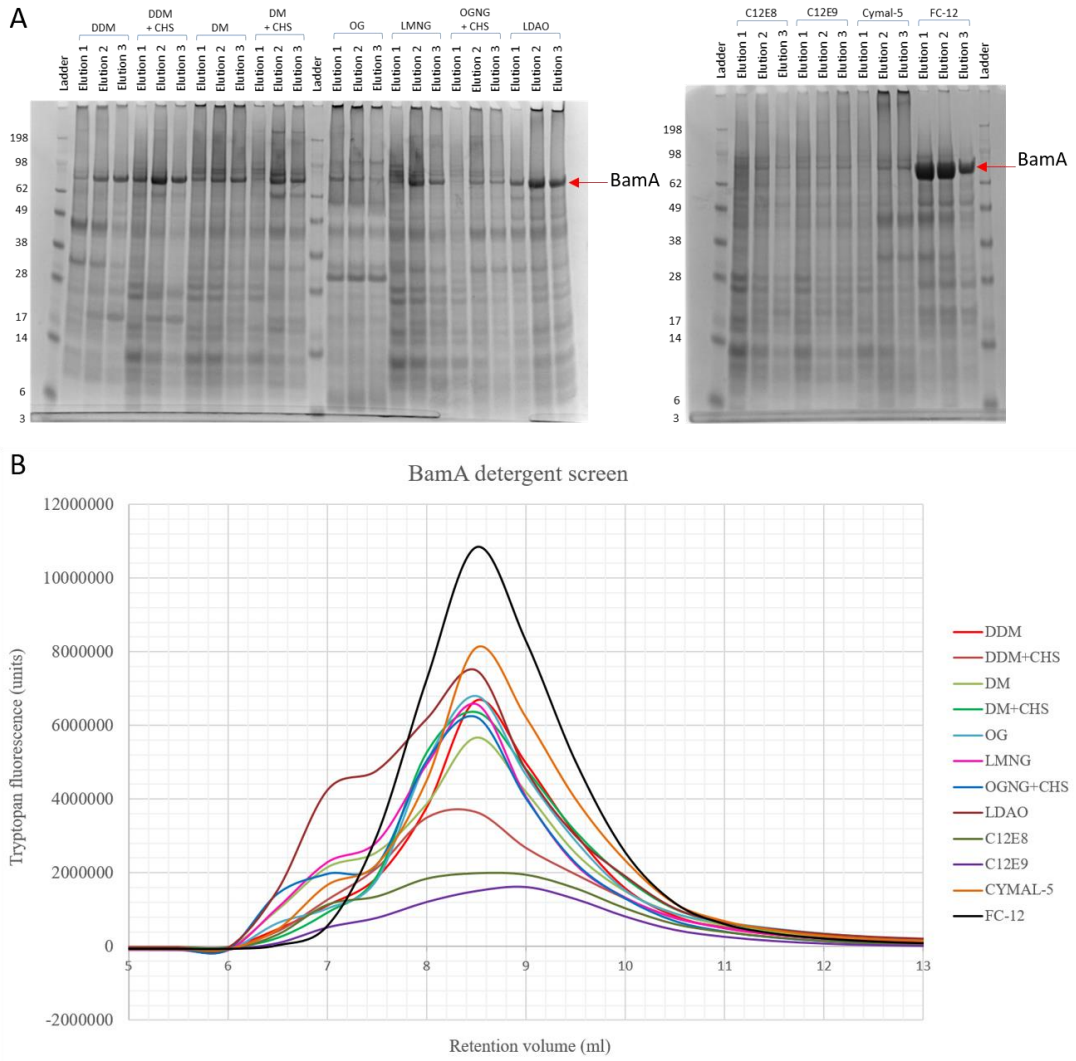


Figure 3.6: CjBamA detergent screening. CjBamA was screened against twelve detergents in preparation for CD, SAXS and future structural experiments. (A) SDS-PAGE of IMAC purifications of CjBamA in each detergent. (B) Tryptophan fluorescence profile of IMAC purified BamA sample elutions from a Seepax SRT 300 SEC column. FC12, LDAO, and Cymal-5 show good extraction on SDS-PAGE and the highest tryptophan fluorescence signal. The weakest fluorescence profiles are obtained from samples extracted in DDM (+ CHS), C12E8, and C12E9.The detergent screens revealed CjBamA is extracted and purified well in FC12, Cymal-5 and LDAO, and that C12E9, C12E8 and DDM+CHS should be avoided for

extraction and purification. Unfortunately, while FC12 extracts CjBamA very well, it is similar to SB12 in that it has poor success in crystallisation attempts of membrane proteins (MPL, personal communication). A follow-up detergent screen was performed with CjBamA extracted initially in FC12 and detergent exchanged into the four next best performing detergents (Cymal-5, LDAO, DDM and LMNG) to determine appropriate conditions for future crystallography screens (appendix 8.3.6).

CjBamA membrane extraction in FC12 followed by detergent exchange was found to be a viable strategy for preparation of CjBamA into suitable detergents for crystallography. However, the spin concentrated sample requires further processing for purity. To improve sample purity post-SEC in the selected detergents, a reverse IMAC (rIMAC) strategy was utilised as a final polishing step (appendix 8.3.7). The flowthrough containing de-tagged CjBamA after TEV protease treatment was retained and CjBamA samples were observed to be significantly purified (>90%) in all detergents tested with few contaminants remaining.

The majority of the detergent screen work was carried out during a short visit to the MPL in Harwell and, due to an endemic phage issue, CjBamA required re-screening of expression conditions in the phage-resistant (Lemo21^{R}) cell line available. Expression of CjBamA was relatively unaffected by the change in expression cell strain type, and conditions mimicking known optimal expression conditions were shown to be suitable (data not shown).

3.2.4.2 *C. jejuni* BamD

Expression and purification of the lipoprotein CjBamD is relatively more straightforward and with higher yields (300-500 µg/L) than CjBamA, however there are instability issues during the purification and the presence of a low concentration (0.5-1.0x CMC) of SB12 is recommended if purified in TBS (see Chapter 4).

The inconsistent Western blot result was investigated, and a consistent signal was obtained through buffer screening when using PBS (pH 7.6) instead of TBS (the purification of BamD (figure 3.7). Purification of CjBamD is similar to CjBamA and requires IMAC, AEX (Q-HP), and SEC (S75) purification steps (figure 3.8). The IMAC purification is relatively impure and contains four major contaminant bands at approximately 42 kDa, 60 kDa, 100 kDa and >250 kDa, respectively (figure 3.8(a)).

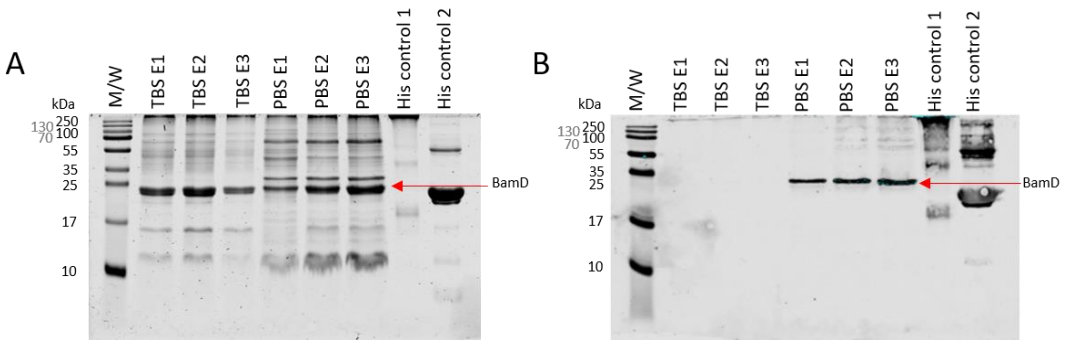


Figure 3.7: CjBamD PBS *vs.* TBS Western blot. An overexpression of CjBamD was split into two and purified independently in TBS and PBS. The (A) SDS-PAGE gel analysis and accompanying (B) Western blot of the IMAC elution fractions are shown. No signal is observed when CjBamD is purified in the TBS (pH 8.0), but is observed when purified in PBS (pH 7.6).

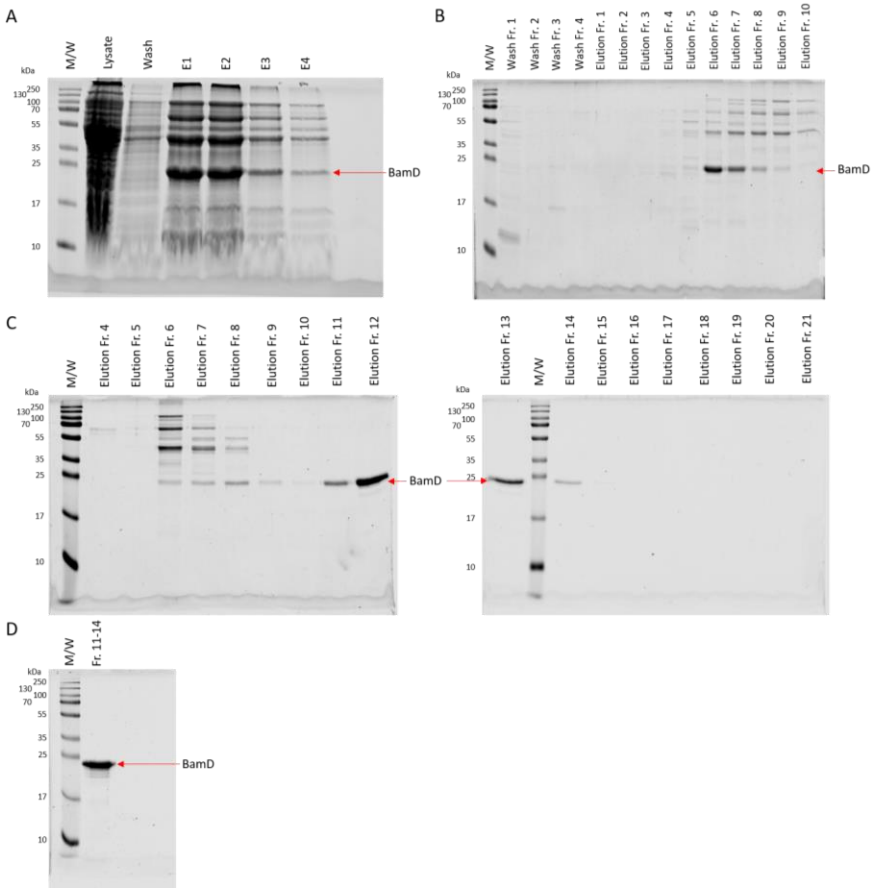


Figure 3.8: CjBamD purification. CjBamD (25.4 kDa) was purified using a strategy of IMAC, AEX and SEC. CjBamD (A) IMAC Ni-NTA purification, (B) Q-HP column AEX purification, (C) S75 column SEC purification, and (D) collected fractions prior to pre-experiment final spin concentration (with a 5 kDa MWCO filter spin concentrator tube) are shown.

The band at 42 kDa was removed through AEX (figure 3.8(b)), and the remaining bands were removed by SEC (figure 3.8(c)). Fractions 11-14 were retained and the sample before further spin concentration (figure 3.8(d)) is shown.

3.2.4.3 **Cj1289**

Cj1289, a peptidyl prolyl transferase and suspected chaperone in the OMP biogenesis pathway, was the highest yielding (0.5-1.0 mg/L) protein in this study, and the purification was similarly a three-step purification, requiring IMAC (figure 3.9(a)), SEC (figure 3.9(c)) and AEX (figure 3.9(d)) FPLC purification runs. A representative purification (figure 3.9) is shown below.

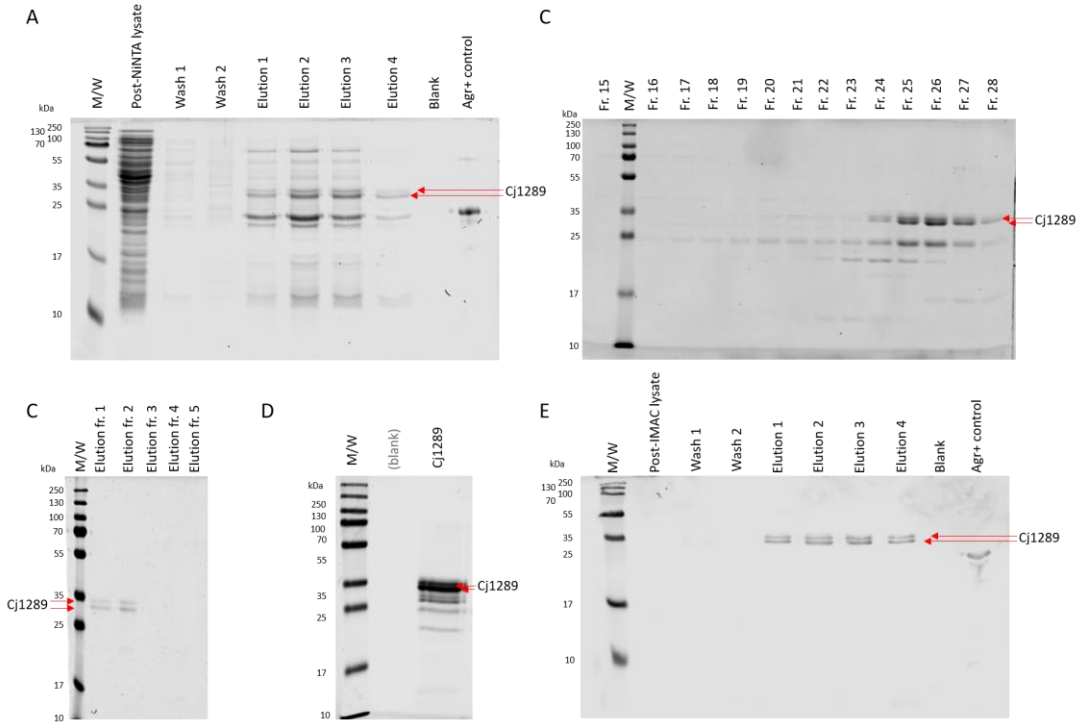


Figure 3.9: Cj1289 purification. Cj1289 (32.5 kDa) was purified using a strategy of IMAC, SEC, and AEX. Cj1289 (A) IMAC Ni-NTA purification, (B) S75 column SEC purification, (C) Q-HP column AEX purification, (D) end sample after purification (post-thaw), and a repeat Western blot (E) to confirm both bands are the His-tagged Cj1289 is included.

A repeat Western blot has been included as two bands are present for Cj1289 around 35 kDa when concentration is low. Both bands show a positive Western blot signal (figure 3.9(e); Coomassie in figure 3.9(a)) and are His-tagged Cj1289, which was

confirmed by the mid-band excision sent for LC-MS/MS (appendix 8.3.3). This was due to plasmid reversion and recyclisation during a later cloning step where the extreme N-terminal 'PMDIGINSDP' (approximately 1.1 kDa) vector encoded sequence was removed in preparation for future structural work (data not shown), and could not be addressed at the time.

3.2.4.4 *C. jejuni* DegP

CjDegP, a dual function temperature dependent chaperone/ protease, was purified using IMAC, SEC and AEX purification methodologies, and a representative purification is shown below (figure 3.10). CjDegP elutes between fractions 18-28 during SEC and fraction 19 during AEX column preparations. The final sample was spin concentrated prior to experiment using a 50 kDa MWCO spin concentrator tube (data not shown).

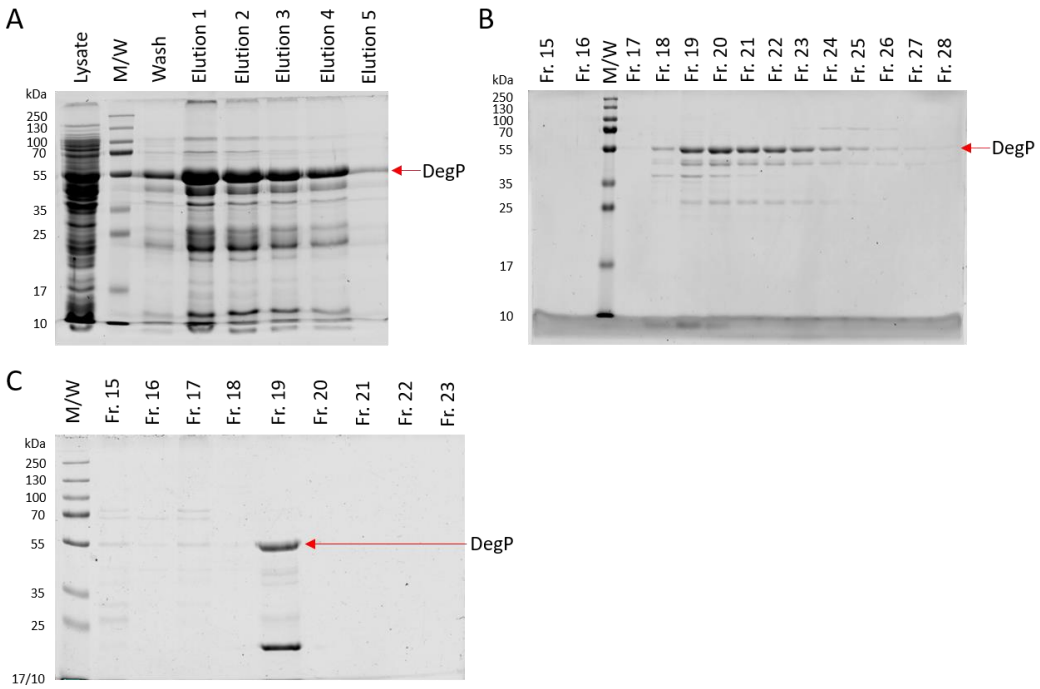


Figure 3.10: CjDegP purification. CjDegP (55 kDa) was purified using a strategy of IMAC, SEC, and AEX. CjDegP (A) IMAC Ni-NTA purification, (B) S200 column SEC purification, and (C) Q-HP column AEX purification.

3.2.4.5 *C. jejuni* LptD

Expression of LptD, an integral membrane protein, gave low yielding preparations post-purification (100-200 µg/L), but was the simplest to purify only requiring two

steps, an IMAC and SEC S200 FPLC run. A representative purification is shown below (figure 3.11).

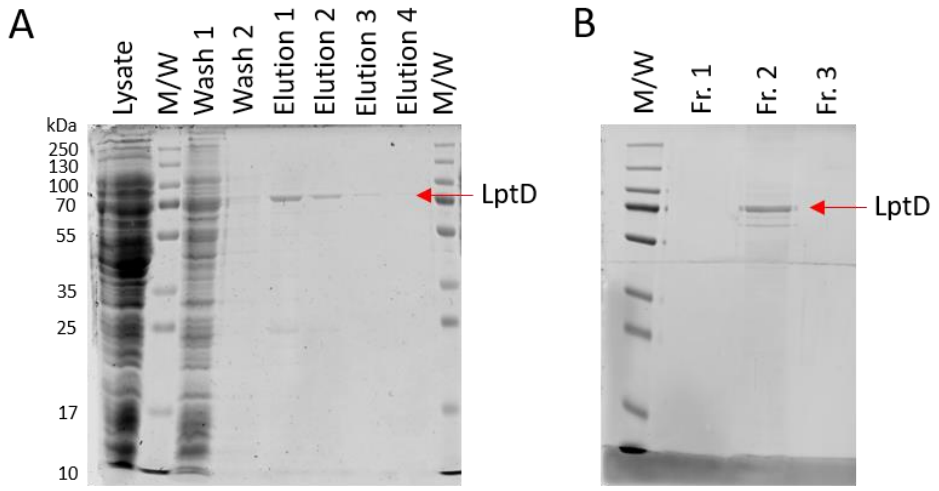


Figure 3.11: CjLptD purification. CjLptD (80 kDa) was purified using a two-step strategy of IMAC and SEC (S200). CjLptD (A) IMAC Ni-NTA purification and (B) S200 column SEC purification, and (C) Q-HP column AEX purification.

In preparing samples for the SAXS mitigation experiments, CjLptD expressed reasonably well for a membrane protein, but the recovery post-IMAC was poor when extracted and purified in SB12 buffers, which were required for buffer matching to CjBamA and CjBamD. An initial detergent screen was performed while at the MPL (appendix 8.3.8), and this was investigated further in overexpression purification screen using DDM, FC12 and SB12, which were available in the lab and were suitable for CjBamA and CjBamD. The resulting IMAC elutions were run on a SDS-PAGE gel (figure 3.12), and show that CjLptD degrades when purified in SB12, and FC12 should be used for all purifications of CjLptD.

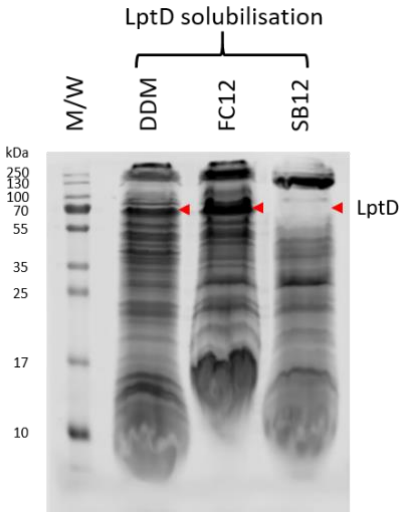


Figure 3.12: CjLptD detergent screen. Membrane isolates from a single CjLptD batch overexpression were extracted in either DDM, FC12, or SB12, and lysates were examined for the presence of CjLptD by SDS-PAGE. CjLptD is extracted well by both DDM and FC12, but was not present in the SB12 extraction. SB12 should not be used for isolation of CjLptD during extraction.

3.2.5 Functional assays

3.2.5.1 *C. jejuni* BamD interaction with lipid

To test the lipophilic properties of CjBamD (and attempt to remove the degradation products) a preliminary experiment was performed on a sample containing a preparation of LUVs incubated with CjBamD and then run on a continuous sucrose gradient for 17 hours (figure 3.13). If CjBamD interacts with the DOPC LUVs (low density) then migration on the sucrose gradient will be retarded and the CjBamD will be found in the first few fractions upon recovery, otherwise CjBamD protein (high density) should be found in the 60% gradient (bottom) fraction. The majority of the CjBamD band is found in the first 2.5 ml post-sucrose gradient, with trace amounts found in other fractions (likely due to incomplete recovery of LUVs by pipetting), showing CjBamD interacts with DOPC LUVs (as do the degradation products).

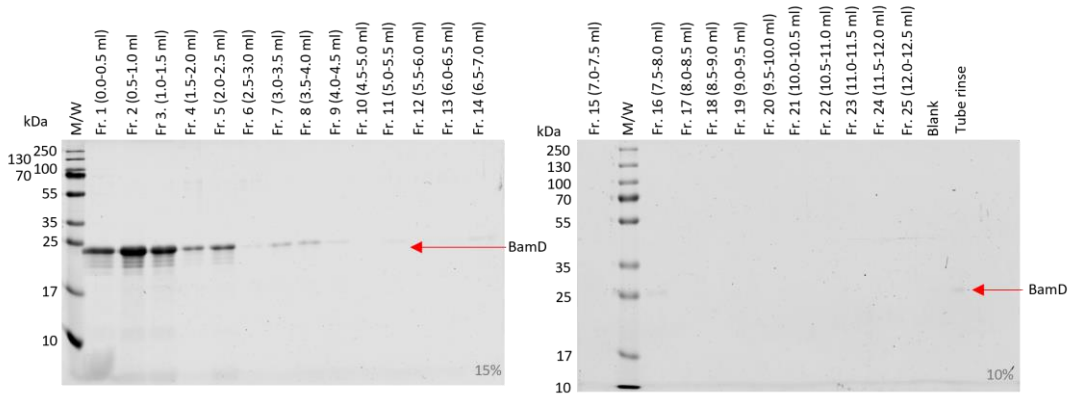


Figure 3.13: CjBamD interacts with DOPC LUVs. A preparatoin of LUVs was mixed with CjBamD, incubated for 30 minutes on ice and run on a continuous sucrose gradient for 17 hours. CjBamD progress through the sucrose gradient is retarded significantly and the majority of the CjBamD sample is found in the first few fractions retrieved in a top-down manner indicating interaction with the DOPC phospholipids. The trace amounts of CjBamD can be found in the end of experiment centrifuge tube rinse are likely due to incomplete recovery of LUVs by pipetting.

3.2.5.2 **Cross-species complementation assay**

The function of the BAM complex is to fold and assemble other proteins. This functionality is difficult to test directly without knowing the exact interaction partners or substrate OMPs. One method of testing the functionality was to supplement *E. coli ΔbamA* or *ΔbamD* conditional deletion mutants with vector-encoded *C. jejuni* homologues. Given the structural similarity of the *C. jejuni* homology models to the *E. coli* structures, it may be feasible to have a functional chimeric BAM complex[106] composed of proteins from multiple organisms. The experiment will be a quantitative assessment as BamA and BamD are essential proteins and deletions of either gene result in a lethal phenotype (as discussed in the introduction). The *ΔbamA* (JWD3) and *ΔbamD* (JCM290) mutants, generously gifted by Ian Henderson, have a *paraBAD* regulated arabinose inducible *E. coli bamA* and *bamD* gene, respectively.

Both mutants were grown without arabinose, and it was confirmed that two overnight subculture rounds are required to deplete the cells of *E. coli* BamA and BamD and generate the lethal phenotype (data not shown). The *ΔbamA* and *ΔbamD* competent were then generated after the first round of overnight subculture from the culture stock. Aliquots of *E. coli ΔbamA* and *ΔbamD* competent cells were transformed with vector encoded *C. jejuni bamA and bamD* constructs, pET20b::*bamA* (pGM1B) and pET26b::*bamD* (pGM2), respectively, and plated on plates containing both ampicillin (A100) and kanamycin (K50), but lacking arabinose. Several negative controls and a positive control were run alongside these transformations and all cell counts, and repeat experiment cell counts are listed for the CjBamA (table 3.1) and CjBamD (table 3.2) cross-species complementation assays are shown.

Unfortunately, no transformants were obtained in the absence of arabinose (Ara) in the growth media, which strongly indicates that the expression of *C. jejuni* BamA and BamD proteins are not able to complement the loss of *E. coli* BamA or BamD, respectively.

pET20b::*bamA* (Ampicillin^R) into Δ*bamA* (JWD3, Kanamycin^R)				
Plate	Condition	1st run	2nd run	3rd run
Transformation 1	LB Ara -ve, K50 A100 + IPTG	-	-	- (1)*
Transformation 2	LB Ara -ve, K50 A100 + IPTG	-	-	-
Control 1 (+)	LB Ara +ve, K50	+++	+++	+++
Control 2 (+/-, depletion dependent)	LB Ara -ve, K50	+++	+	+
Control 3 (-)	LB Ara +ve, A100	-	-	-
Control 4 (-)	LB Ara -ve, A100	-	-	-
Control 5 (-)	LB Ara +ve, A100 K50	-	-	-
Control 6 (-)	LB Ara -ve, A100 K50	-	-	-

* Small colony, likely contaminant

Table 3.1: Cross-species complementation assay - *C. jejuni* pET20::*bamA* vector supplementation of native BamA depleted (JWD3) *E. coli* K12.

pET26b::*bamD* (Kanamycin^R) into Δ*bamD* (JCM290, Ampicillin^R)				
Plate	Condition	1st run	2nd run	3rd run
Transformation 1	LB Ara -ve, K50 A100 + IPTG	-	-	-
Transformation 2	LB Ara -ve, K50 A100 + IPTG	-	-	-
Control 1 (+)	LB Ara +ve, A100	+++	+++	+++
Control 2 (+/-, depletion dependent)	LB Ara - A100	+++	+	-
Control 3 (-)	LB Ara+ K50	-	-	-
Control 4 (-)	LB Ara- K50	-	-	-
Control 5 (-)	LB Ara+ A100 K50	-	-	-
Control 6 (-)	LB Ara- A100 K50	-	-	-

Table 3.2: Cross-species complementation assay - *C. jejuni* pET26b::*bamD* vector supplementation of native BamD depeleted (JCM90) *E. coli* K12.

3.3 **Discussion**

The initial cloning phase of this project to isolate the genes of interest had taken several months, which hindered the progress of the project past the initial stages. The cloning issues were likely due to Campylobacter species having a high genome plasticity[152], and cloning was impacted by the *C. jejuni* NCTC 11322 source strain choice as there was a lack of sequencing data on this strain in the NCBI databank. A more appropriate strain for initial gene isolation would be NCTC 11168, as it is the academic strain of choice for genomic sequencing, however this was not known at the early phases of the project when the strain was purchased, and NCTC 11322 was deemed appropriate being a human clinic isolate and a strain readily used in industry. Following isolation of genes and cloning, the expression of constructs was confirmed by Western blots and LC-MS/MS analysis.

The LC-MS/MS successfully identified *C. jejuni* BamA, Cj1289, LptD and DegP after the peptide fingerprints were cross-referenced against their primary sequences using Biotools. The Western blots, with the exception of CjBamD, showed consistently positive signals for all proteins. The Western blot signal variability on BamD samples was perplexing, with positive signals if performed following purification and loss/ lack of signal on older samples. This may be due to gradual tag degradation, a lower stability of CjBamD in TBS relative to PBS (due to pH differences), or a low half-life of the protein and degradation as the sample ages (EXPASY ProtParam prediction, data not shown). Considering the variability in the Western blot signal for CjBamD, the LC-MS/MS negative result was likely due to degradation of CjBamD as it was stored prior to use at 4°C over the Christmas period while ordering was unavailable. CjBamD will need to be resubmitted for LC-MS/MS identification when the project receives further funding.

An alternative explanation may be inherent instability of CjBamD due to the construct design, which factored removal of the signal peptide at the cleavage site identified by SignalP/ LipoP, and removed all amino acids prior to the N-terminal cysteine resulting in an immature or improperly targeted lipoprotein. After further investigation, the amino acids prior to the N-terminal cysteine which were removed (MKKGIFLLVFLSVFFSA[C]) contained the lipobox motif, a sequence of four amino acids that target lipoproteins to the lipoprotein biosynthetic pathway[153]. In *E.*

coli the consensus lipobox motif sequence is $(LVI)_{-3}(ASTVI)_{-2}(GAS)_{-1}C_{+1}$[154–156]. While the sequence prior to the N-terminal cysteine of *C. jejuni* BamD does not match this consensus *E. coli* sequence completely, featuring only three (SAC) of the four consensus residues, previous work by Jin *et al.* (2004)[157] in *C. jejuni* on JlpA identified a "L-F-S-A-C" lipobox sequence at amino acids 14-18. This coincides well with the 'VFFSAC' sequence observed in *C. jejuni* BamD given a leucine to valine substitution and an additional phenylalanine, and indicates the loss of the targeting sequence and this may not allow for the correct targeting and localisation of BamD when expressed with our current construct. The pET26b+ encoded *pelB* leader sequence (MKYLLPTAAAGLLLLAAQPAMA) prior to the *bamD* gene shares insufficient similarity to compensate for this absent sequence. However, the absence of the lipobox sequence did not hinder BamD binding to DOPC in the sucrose gradient analysis, and it is unknown what affect it will have on BamD in the subsequent experiments.

The lack of transformants in the cross-species complementation assay may be due to the *C. jejuni* proteins not being able to engage in cross-species interactions with the *E. coli* BAM complex members, or being unable to recognise *E. coli* folding and assembly target OMPs due to differences in the β-signal sequence between organisms (as discussed in introduction).

3.4 Summary

The project was restarted nine months into the PhD and restructured to include several proteins involved in the *C. jejuni* BAM OMP biogenesis pathway. This chapter covered all work from cloning through to optimised purifications, detergent screening and basic functional assays. Suitable conditions were found for purification of all proteins, FC12 was observed as the best extraction detergent for both BamA and LptD, and identified several others that BamA can be detergent exchanged into prior to further experiments. In the functional assays, the cross-species complementation assay showed a negative result for a functioning *E. coli* and *C. jejuni* chimeric BAM complex, likely due to lack of interaction or differences in β-signal sequence, and BamD was observed to migrate with DOPC lipid vesicles on a continuous sucrose gradient confirming BamD lipid interaction despite being expressed without a lipobox motif.

4 Secondary structure and melting temperature characterisation of the BAM complex in *C. jejuni* using circular dichroism

4.1 Chapter aims

In this chapter, we sought to determine the secondary structure profiles, assess the state of folding through thermal degradation, and probe and characterise any interactions between the proteins of the BAM mediated outer membrane transport pathway in order to progress the project towards the structural determination of these proteins.

In these experiments, we aim to use CD spectroscopy to do the following:

1) Characterise the secondary structure of the proteins that have been recombinantly overexpressed
2) Follow the thermal deformation of protein secondary structures to determine the melting temperature or transition point that describes the folded to unfolded state transition
3) Analyse the CD spectrum and secondary structure estimations of mixtures of these recombinant proteins to test for interaction through interpretation of the changes relative to the individual spectra and melting temperatures obtained
4) Compare these experimentally obtained melting temperatures to those in the literature to determine the appropriateness of these values for the type of protein being studied
5) Determine the cooperativity of melting transition(s) to infer whether these proteins interact and to what extent they stabilise the secondary structures
6) Compare the recently published Landau free energy expansion model analysis of CD to the current established Boltzmann two state equation model to determine its relative suitability for studying thermal melting parameters

4.2 Introduction

4.2.1 Introduction to Circular Dichroism

Electromagnetic radiation is termed ‘light’ if the wavelength range falls with the visible or easily accessible UV wavelength range (~170nm-800 nm). Light can be referred to as a wave form consisting of a direction of propagation and both an electrical and magnetic component (at 90° to each other) which vibrate or oscillate perpendicular to the direction of travel[158]. In natural or unpolarised light, the electric field oscillations occur in all planes that are perpendicular to the direction of travel. By passing light through planar filters, the light is only able to oscillate in one plane and the resultant light is referred to as plane polarised light. If the horizontal and vertical components of this light are retarded out of phase with one another then circularly polarised light with constant magnitude is generated. Circularly polarised light is chiral, that is the left- and right-circularly polarised light are mirror images and cannot be superimposed onto one another. Whether this circularly polarised light interacts with a molecule depends on its optical activity (ability to rotate the plane of polarisation of polarised light) and chirality (“handedness”). The vast majority of biological molecules are chiral and exhibit optical activity, including the amino acids that make up proteins. The differential absorption of left- and right-circularly polarised by a chiral molecule is referred to as the molecule’s circular dichroism. CD spectroscopy is a differential absorption technique that analyses the differences in absorption of left- and right-circularly polarised light by a given molecule[158, 159].

The study of secondary structure of protein by CD utilises dichroic polarimeters that are composed of several modules including a light source, filter or monochromator, polarizer, quarter-phase plate, sample loader, and detector. The light source (helium lamp, xenon lamp, or synchrotron) generates light that is passed through a filter or monochromator to filter the light to a set spectral wavelength. The monochromatic light passes through a polarizer to generate linearly polarised light, which is incident on a birefringent material (such as a fused silica quarter-phase wave plate) at 45° to the plane of the incident light. The fused silica is a piezoelectric material that responds to electrical current in the form of an induced stress when oscillating at its resonant frequency (typically 50 kHz). This alternating stress causes the fused silica quarter wave plate to retard the vertical component with respect to the horizontal

component of the incident light to generate left-circularly polarised light, and *vice versa* to generate right-circularly polarised light of the same intensity. This light is incident on the sample, which depending on its properties, differentially absorbs left- and right-handed circularly polarised light (or photons). The unabsorbed photons are incident on a photomultiplier tube, which produces a current of magnitude proportional to the number of incident photons, thereby giving the intensity of unabsorbed photons. This current is detected by a lock-in amplifier, which determines the sign (+ or -) of the signal generated. These measurements are used to generate a typical CD spectrum[158, 160].

The absorption of unpolarised light by a molecule can be described by the absorbance, A, and is defined as:

$$\boldsymbol{A = log_{10}\left(\frac{I_0}{I}\right)} \qquad \text{(Equation 4.1)}$$

Where I_0 is the intensity of the incident light prior to reaching the sample, and I is the intensity of the light after it has travelled a distance, l, through the sample. As the incident left- and right-circularly polarised light has the same intensity and we are looking at the differential absorption, the equation becomes

$$\boldsymbol{\Delta A(\lambda) = A_L(\lambda) - A_R(\lambda)} \qquad \text{(Equation 4.2)}$$

Where ΔA is a function of the incident wavelength. Beer-Lamberts law states where an absorbing molecule has a molar concentration, c, and the sample has thickness or pathlength l, the absorbance is given by:

$$\boldsymbol{A = \varepsilon \cdot c \cdot l} \qquad \text{(Equation 4.3)}$$

Where ε is the molar extinction coefficient. The molar extinction coefficient can be described in terms of circularly polarised light to give the molar circular dichroism, Δε:

$$\boldsymbol{\Delta\varepsilon = \varepsilon_L - \varepsilon_R} \qquad \text{(Equation 4.4)}$$

By substituting the molar circular dichroism into Beer-Lamberts law (Equation 4.3):

$$\boldsymbol{A(\lambda) = [\varepsilon_L(\lambda) - \varepsilon_R(\lambda)] \cdot c \cdot l} \qquad \textbf{(}\text{Equation 4.5)}$$

Where ε(λ) is the molar extinction coefficient with units $cm^{-1}M^{-1}$.

Commercially available CD instruments measure the differential absorbance, ΔA, which is readily converted to Δε by:

$$\Delta\boldsymbol{\varepsilon} = \boldsymbol{\varepsilon_L} - \boldsymbol{\varepsilon_R} = \frac{\boldsymbol{A_L - A_R}}{\boldsymbol{c \cdot l}} \qquad \text{(Equation 4.6)}$$

However, most instruments also give the ellipticity, an angular measure of CD, which has a relationship to ΔA given by:

$$\boldsymbol{\theta = 32.98 \cdot \Delta A} \qquad \text{(Equation 4.7)}$$

Where θ is in degrees. In order to normalise against the effects of pathlength and concentration, the molar ellipticity can be further defined as:

$$[\boldsymbol{\theta}] = \frac{\boldsymbol{100 \cdot \theta}}{\boldsymbol{c \cdot l}} = \boldsymbol{3298 \cdot \Delta\varepsilon} \qquad \text{(Equation 4.8)}$$

The units for molar ellipticity are degrees cm^2/ $dmol$[159].

For the study of protein CD, the wavelengths of radiation that are useful can be divided into three regions: the far UV (below 250 nm, including the vacuum UV below 200 nm), the near UV (250-300 nm) and the near UV-visible region (300-700 nm). Proteins possess several chromophores of interest for the analysis of CD, the amide bond, the disulphide bond, aromatic sidechains, and prosthetic groups, which correspond to these regions. The signal obtained from far-UV CD is dominated by the amide bond transitions, and analysis of this signal gives information about the conformation of the (poly)peptide backbone and therefore the secondary structure of proteins. Whereas the signal in the near-UV region, where contributions from aromatic side chains and disulphide bonds are prevalent, gives information about the tertiary structure of proteins. The analysis of CD bands above 300 nm give information on the conformation and environment of prosthetic groups (e.g. heme- or flavo-proteins)[161].

Different types of secondary structure of proteins give rise to different characteristic signals in the analysis of far-UV CD (figure 4.1). The experimental CD spectrum for the α-helix has two characteristic negative bands at 222 nm and 208 nm, and one characteristic positive band near 193 nm. The CD spectrum for β-sheets has two characteristic negative bands near 215 nm and 175 nm, and a positive band near 195 nm. Due to the flexible geometry of the backbone of β-sheets in polypeptides,

these values are given approximate values and may be red-shifted in practice. A CD spectrum analysis of proteins also gives values of composition in turns and unordered secondary structures, which can have spectrum that vary depending on the type of turn or unordered structure analysed[162, 163].

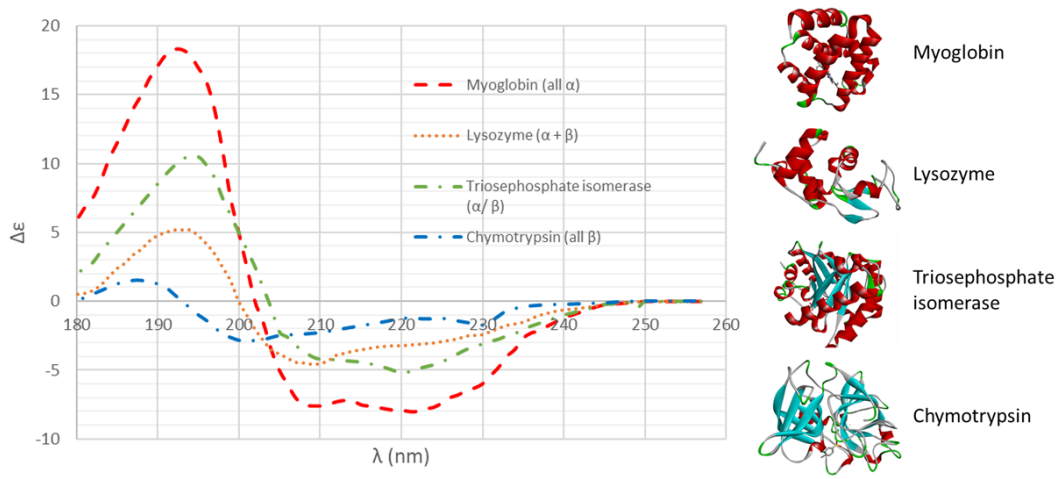


Figure 4.1: Circular dichroism spectra of myoglobin, hen egg lysozyme, triosephosphate isomerase and chymotrypsin. Circular dichroism spectra of myoglobin (H:0.77, B:0.0, T:0.12, U:0.11), hen egg lysozyme (H:0.40, B:0.11, T:0.20, U:0.29), triosephosphate isomerase (H:0.45, B:0.16, T:0.12, O:0.27), and chymotrypsin (H:0.10, B:0.39, T:0.18, U:0.33) illustrating how different abundances of secondary structure types in proteins affect the output CD spectra. (H = helix, B = beta-sheet, T = turn, U = unordered; redrawn and adapted from Johnson 1988[164])

4.2.2 Biological background to experiment

In *E. coli*, BamA and BamD have been shown to interact[5, 8, 10, 37, 44], and the available chaperones SurA, Skp, and DegP are thought to interact with the BAM system. SurA is thought to be the direct mediator in the shuttling process of protein substrate OMPs to the BAM complex via interaction with BamA[70, 165, 166]. As we are testing homologues of these proteins in *C. jejuni*, they should display the same or similar interactions. Therefore, we expect *C. jejuni* BamA and BamD to interact to form the BAM complex, and Cj1289, which is described as a SurA-like protein[108], to interact with the BAM complex via interaction with BamA. As DegP has a dual functionality as a chaperone and a protease[90, 92], we expect that there will be an interaction with LptD and any other substrate OMPs that are destined for the BAM mediation

transport pathway for the chaperone activity, and potentially any of the proteins in this study for the protease activity (figure 4.2).

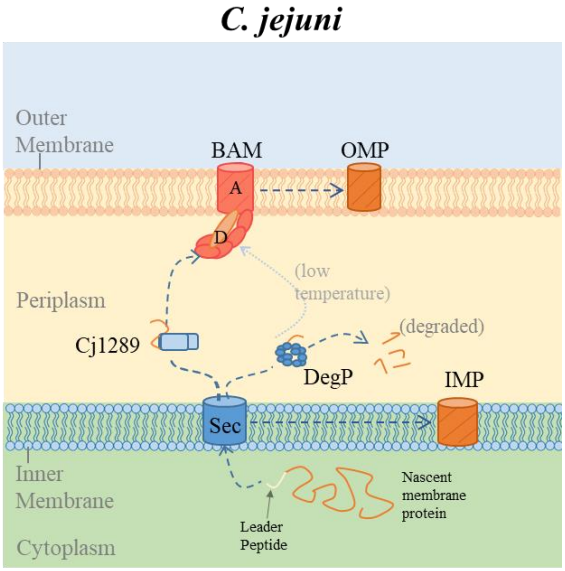


Figure 4.2: BAM-mediated outer membrane transport pathway in *C. jejuni*. The nascent membrane protein is transported to the periplasm by the Sec translocase, where it is bound by either DegP or Cj1289, and transported to BAM for folding and targeted insertion into the membrane.

To test the interactions, we employ the use of CD thermal degradation experiments where we individually, or as mixtures, thermally degrade the proteins and compare the data for shifts in the melting temperature, CD spectra and secondary structure estimation (SSE). If the melting temperature for the mixture exceeds that for the individual components, we can make the reasonable assumption that the proteins interact and stabilise each other.

4.2.3 Thermal Degradation experiments

The thermal stability of proteins can be assessed by CD by following the secondary structure changes as a function of temperature. The structural changes can be used to describe the thermal unfolding of proteins, the folded to unfolded transition, by comparison of the 'folded' spectrum and 'unfolded' spectrum, typically by following the change in α-helical content at a specific wavelength (e.g. α-helices have characteristic CD signal near 193 nm or 209 nm). The thermal degradation experiments were planned broadly in line with the experimental guidelines set out in Greenfield (2006)[167], with the CD spectra being recorded at 2°C intervals from 20°C

to 89°C and on return to 20°C post-experiment. The beamline specialist application, CDapps[130], is utilised to analyse the raw CD spectra to generate and display secondary structure as a function of temperature throughout the thermal degradation experiments.

4.2.4 Thermal stability expectations

Recently, Mateus *et al.* (2018) performed thermal proteome profiling of *E. coli* and published the "meltome" (or thermal degradation profiles) of *E. coli* proteins, thereby establishing some relevant guidelines by which to assess the melting temperatures obtained through thermal degradation experiments. The average melting temperature (T_m) followed a trend of OMPs > periplasmic proteins > IMPs > cytosolic proteins. The authors have related this trend to the protein's structure, function and/ or locality, highlighting a trend wherein the T_m of proteins increases the further the protein localises from the nucleus. Within this overarching trend, the low temperature outliers (<52°C) are represented by enzymes involved in DNA replication, and high temperature outliers are represented by periplasmic chaperones (FkpA, Skp, etc.) and β-barrels, which make up some of the high temperature extreme outliers. They posited that in proteins that engage in substrate OMP or ligand binding (e.g. periplasmic chaperones and ligand-binding subunits of ATP-binding cassette transporters), there may be a co-operative co-stabilising effect upon binding, reflected in their relatively higher melting points, and that this trend exists in all sub-localisations within the cell (i.e. whenever proteins are part of complexes). Through comparisons of different complexes, they attempted to deconvolute this behaviour, stating that spatial localisation has a greater effect on co-melting temperature than simply complex formation. That is to say, a complex formed of two proteins from the same localisation may co-stabilise well whereas a complex formed of two proteins from different localisations may display bi-phasic (or n-phasic, where n is the number of proteins in the complex) rather than co-melting behaviour[168]. In addition, they found that β-barrel proteins demonstrate a bi-phasic melting behaviour, with one fraction melting at a lower temperature and the second fraction being stable to temperatures up to 87°C.

These findings have relevance to our current experiment, as mixtures of BamA (intrinsic OMP with β-barrel domain and periplasmic β-α-α-β-β fold POTRA

domains), BamD (α-helical lipoprotein), and Cj1289 (periplasmic chaperone composed of α + β components) which have closely related spatial localisations are tested. Potentially, the periplasmic domain of BamA will be stabilised by and co-melt with the lipoprotein BamD or the periplasmic chaperone Cj1289, or a combination of both. Considering the above in the context of the findings from Mateus *et al.* (2018)[168], there would be an expectation of a high T_m for the β-barrel domain of BamA, and likely a bi-phasic profile between the periplasmic POTRA domains (predominantly α/β folds and turns) and the β-barrel domain (predominantly β-barrel structure linked by turns and flexible loops). In addition, there is the consideration that these different domains may not be distinguishable if stabilisation of BamA POTRA domains and BamD/ Cj1289 result in co-melting. Overall, if the localisation is the major determinant of T_m, we would expect that BamA > BamD > Cj1289. However, if structure were the major determinant of T_m, then it may be more reasonable to expect BamA > Cj1289 > BamD, as BamD is purely α-helical and Cj1289 is a mixture of α/β structures.

An alternative point to consider is that a complex formed of proteins from different localisations may undergo a phenomenon termed thermal proximity co-aggregation (TPCA), where the melting of one component (that has lower thermostability) can destabilise and negatively affect the thermostability of the remaining complex members[169].

Taking this into consideration for the proposed experiment, interaction of BamA POTRA domains with BamD/ Cj1289 may cause the periplasmic domain of BamA to destabilise at lower temperatures if the co-stabilising effect is lower than that required to bridge the individual (POTRA to BamD/ Cj1289) T_m difference (assuming the POTRA domains have a higher T_m than BamD/ Cj1289).

Previous work from our lab[170] suggests the use of Landau modelling to resolve data displaying this biphasic behaviour, and theorise that the biphasic phenomena can be explained by independent melting of the α-helical (low temperature melt, approximately 45°C) and β-sheet (high temperature melt, approximately 65-70°C) secondary structures.

The third option is that the proteins will not interact, in which case we will see a thermal degradation profile that is a combination of the profiles of both proteins. These different possibilities are illustrated in the following figure (figure 4.3).

A final point to consider is the finding that *E. coli* BamC was found to be completely thermostable over the temperatures tested (40°C-87°C), which may be down to its high flexibility (shown in crystal structure 5do0)[168]. Therefore, flexible or unordered structures may not show corresponding changes or these changes may not be easily distinguished during a thermal degradation experiment in this temperature range.

4.2.5 Derivation of critical points of folded and unfolded states during thermal degradation

The current theory used in thermal degradation experiments to derive the T_m is the Boltzmann two state equation. The Boltzmann two state equation models the structural changes during thermal degradation as a single change in structure, going from the initial starting state, folded, to the end state, unfolded, considers both states and picks a midpoint value between these states, which it defines as the T_m for the transition[167]. While this appears a logical approach, it does not explain the biphasic melt behaviour exhibited by β-barrel proteins upon thermal degradation,[168] and potentially similar behaviours for proteins composed of mixtures of α-helical and β-strand content. This raised the question of whether proteins that consisted of both α-helical and β-strand structures should be treated in this manner, leading to the proposition of expanding the current theory on thermal degradation in CD by treating the α-helical and β-strand melts independently[170].

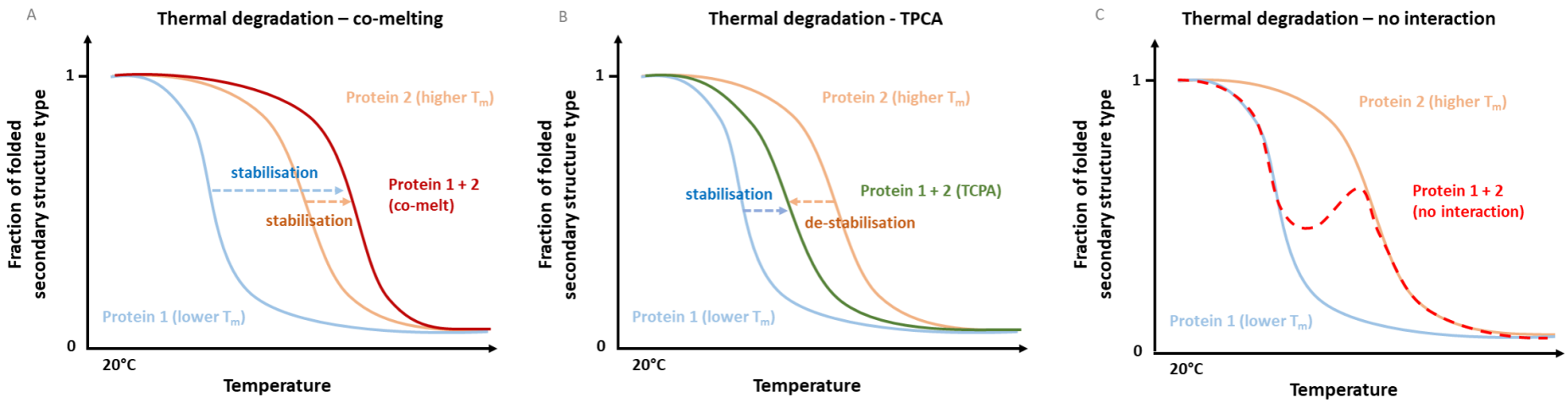


Figure 4.3: Thermal melt profile expectations of co-stabilisation and co-melting, thermal proximity coaggregation, and non-interaction between proteins. Different experimental outcomes are observed depending on the interaction phenomena exhibited between protein 1 and 2 depending on whether they interact, and the specific type of interaction. Outcome one (A), the proteins interact strongly and show co-stabilisation and cooperative melting at a right-shifted T_m relative to the more thermally stable protein 2. Outcome two (B), TPCA, the proteins interact and stabilise up to a point, resulting in increased thermal stability of protein 1 and a right-shift of the profile (T_m) of protein 1. However, the stabilisation of protein 1 is not as pronounced [as in (A)] and so it begins melting at only a slightly higher temperature relative to its base profile, and the denaturation of protein 1 has a destabilising effect on protein 2, causing the profile (T_m) of protein 2 to become left-shifted. Outcome 3 (C), there is no interaction is observed between protein 1 and 2, and so a stepped profile is shown for the thermal degradation of the protein mixture reflecting the individual melting temperatures of each protein.

4.2.6 **Landau free energy expansion models**

A novel alternative to the Boltzmann two state equation analysis of melting point and secondary structure changes during thermal degradation is the use of Landau free energy expansion models pioneered by the Bonev lab group in Hussain *et al.* (2018)[170]. The difference between the methods lies in the method of fitting the CD data to the melting point, Boltzmann uses the transition midpoint and Landau modelling derives several parameters and ignores the post-thermal melt state. These factors in turn, affect the T_m derivation. Instead of following the transition of the protein as a whole to determine the melting point, the Landau free energy expansion model monitors a single type of secondary structure, and calculates the melting point for that structural type by examining the total change for that structure over the entire temperature range sampled in isolation. The structure (S+, S-) and temperature (T+, T-) parameters are calculated to define the start and end points of the folded to unfolded transition for a secondary structure type, alongside a further parameter, tau (τ, equation 4.9) to generate an equation of state defining the spinodal fit to the data. The equation of state requires substitution of tau into the Gibbs-Helmholtz equation (equation 4.10), usually used to model thermal denaturation of proteins[167], to derive two independent equations representing the folded to unfolded phase transitions for α-helical and β-strand structures (equation 4.11). The fit is improved through iterative changes in the estimated T_m value, which alters fitting of the spinodal to the changes in secondary structure in the data, giving the final estimation of T_m for the structural transition[170].

$$\boldsymbol{\tau = 1 - \frac{T}{T_m}} \qquad \text{(Equation 4.9)}$$

$$\boldsymbol{\Delta G = \Delta H\left(1 - \frac{T}{T_m}\right) - \Delta C_p\left[(T_m - T) + T \cdot ln\left(\frac{T}{T_m}\right)\right]} \qquad \text{(Equation 4.10)}$$

$$\boldsymbol{\alpha = \frac{3{s_+}^2}{T_c - T_+}} \quad \text{and} \quad \boldsymbol{\beta = \frac{2{s_+}^3}{T_+ - T_m}} \qquad \text{(Equation 4.11)}$$

By ignoring the latter (progressively unfolded) states and focussing on the transition of the initial folded state, the Landau free energy expansion model hopes to provide more accurate and informative T_m estimations for transition of the pre-experiment secondary structures.

An illustration of this theory is given in figure 4.4, where the spinodal is fitted and used to derive terms T_+, S_+, T_- and S_-, which define the positions of the folded and unfolded states in the transition, and that of the metastable state between these states for a given secondary structure type. The addition or removal of further energy to this derived metastable state, by increasing or decreasing the temperature to cause thermal fluctuations of the secondary structures to exceed or regress from the threshold set out by either boundary, would result in the secondary structure leaving this equilibrium point and becoming either unfolded or remaining folded.

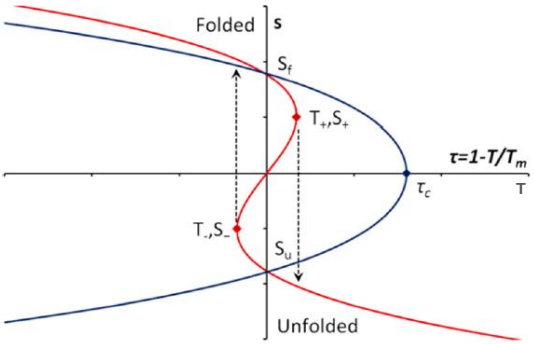


Figure 4.4: The spinodal of the transition of the folded to unfolded state of the α-helical (or β-strand) content of a protein. The points T+ and T- denote the temperatures of transition of state, and S+ and S- represent the point at which the structure transitions. These points represent the last points of the metastable state of equilibrium between the folded and unfolded states, with the midpoint around the critical point, tc (where $\Delta G_{\rightarrow 0}$). Passing these points causes the metastable state between folded and unfolded to collapse, and one state dominates. (Taken from Hussain *et al.*, 2018[170]; where T = temperature, S = structure, T_m = melting temperature, and τ = transition cooperativity of melting; Note: Axis amended.)

A further advantage of this method is the derivation of τ, which describes the cooperativity of folded to unfolded state transition, and is a normalised value that can be plotted on an absolute scale, between zero and 0.5, enabling comparison of any one material to another. The upper limit of this scale can be represented by the ice to water transition, which would be considered a highly cooperative or 'hard' transition. Where proteins fall on this scale will be mapped during this study, and it would be intuitive to expect that two interacting proteins would have higher cooperativity than individual proteins. The impact of multiple domains on cooperativity and the influence of detergent will also be investigated.

Both Boltzmann and Landau expansion models will be used in the analysis of samples in this study.

4.2.7 **CD analysis**

Observing the change in the spectrum at approximately 193/208 nm or near 215 nm can give an estimate to the changes in secondary structure in α-helical and β-strand content during the thermal degradation experiments. To obtain more precise estimates of the change in secondary structures, several computer algorithms were developed that compare the spectra obtained to a reference database, and these are utilised in current analyses[163]. The common algorithms in use are SELCON3[171, 172], CDSSTR/ VARSLC[173, 174], and CONTIN/LL[131, 132]. For the purposes of this study we utilise CONTIN[131]/LL[132], which fits the CD of unknown proteins using a linear combination of reference spectra from the reference set, and is considered better than alternatives at providing estimates for β-sheets[129, 163]. The difference between the original CONTIN[131] and CONTIN/LL[132], is the latter organises the reference database by similarity to the experimental dataset obtained and applies a weighting, eliminating the distantly rated proteins to construct smaller and better tailored reference data sets with which to do the linear combination[175]. The reference set used in this analysis was the SMP 56 reference data set, comprising of 43 soluble proteins and 13 membrane proteins[176, 177]. The protein structures can be defined by the categories set out in the DSSP algorithm by Kabsch and Sanders[124], which are split into helix type (α, 3_{10}, pi), β-strand (parallel, anti-parallel), turn (i+3, i+4, i+5 hydrogen bonded, non-hydrogen bonded), and unordered (polyproline I, polyproline II, unordered/ extended/ flexible). However, in practice more simplified outputs are recorded such as ordered and disordered α-helical, ordered and disordered β-strand, turns, and unordered[172]. The program used in this analysis, CDapps[130], simplifies this further in to a standard four category secondary structure output, α-helical, β-strand, turn, and unordered.

4.3 Results

4.3.1 In-silico predictions of protein SSAs

Secondary structure assignments were made prior to the experiments using available structures. As the MD annealed models showed lower standard deviation than the unequilibrated homology models across the SSA algorithms, and may better represent the proteins in the aqueous (and where appropriate, a combined aqueous and detergent) environment of the experiment, where possible, these averaged outputs were selected for each protein and protein mixture (table 4.1). As expected, BamA was predominantly estimated to be β-sheet, BamD and Cj1289 were estimated to be mainly α-helical, and mixtures have values averaged between their individual components. For all individual SSA server outputs, averages, and standard deviation calculations see appendices (appendix 8.4.1).

Sample	α-helical	β-sheet	Turns	Unordered
BamA [MD, SB12]	0.152	0.435	0.074	0.339
BamD [MD]	0.485	0.000	0.345	0.170
BamD [MD, SB12]	0.673	0.010	0.092	0.226
Cj1289 [MD]	0.520	0.097	0.367	0.016
Cj1289 [MD, SB12]	-	-	-	-
BamA + BamD [MD, SB12]	0.264	0.342	0.078	0.314
BamA + Cj1289 [PDB]	0.321	0.178	0.204	0.297
BamD + Cj1289 [MD]	0.625	0.094	0.062	0.216
BamA + BamD + Cj1289 [PDB]	0.338	0.289	0.075	0.297

Table 4.1: Averaged *in silico* estimated SSAcomponent values of BamA, BamD, Cj1289 and 1:1 protein mixtures. Annealed MD models of BamA, BamD, Cj1289 and mixtures thereof, were analysed by SSA programs to estimate secondary structure. Predicted α-helical, β-sheet, turn and unordered compositions are shown.

SSA estimation of unannealed assembled PDB models are shown where MD models were not available. Cj1289 [MD, SB12] MD simulation PDB unavailable.

4.3.2 Chirascan spectrophotometer experiments

As a preliminary step, 1-n spectral scans of the relevant background absorbance signal associated with the two buffers used in the experiments, a non-detergent buffer (NDB; 150 mM NaF, 30 mM Tris pH 7.6) and detergent buffer (DB; 0.25 % SB12, 150 mM NaF, 30 mM Tris pH 7.6) were obtained for buffer subtraction of sample spectra (figure 4.5). The individual proteins and protein mixtures (in 1:1 stoichiometry) were then assessed by 1-n spectra scans to check the signal quality (figure 4.6) and determine the pre-melt SSE at 20°C (table 4.2). BamD and Cj1289 were assessed in both non-detergent and detergent buffers to determine the effect, if any, on the thermal stability of either protein before a mixture was made with BamA. The goodness of fit was calculated as a standard deviation for SSEs in each thermal degradation experiment at each temperature, and tabulated for reference (appendix 8.4.2).

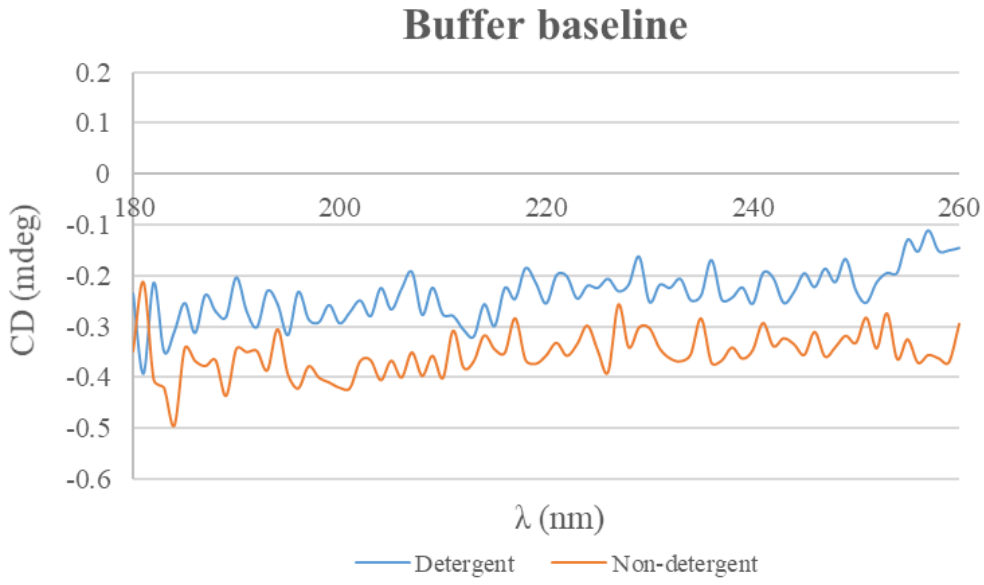


Figure 4.5: Buffer 1-n spectral scans. The detergent and non-detergent sample buffers were scanned prior to the thermal degradation protein runs to establish a baseline value for buffer subtraction of the signal in the following experiments.

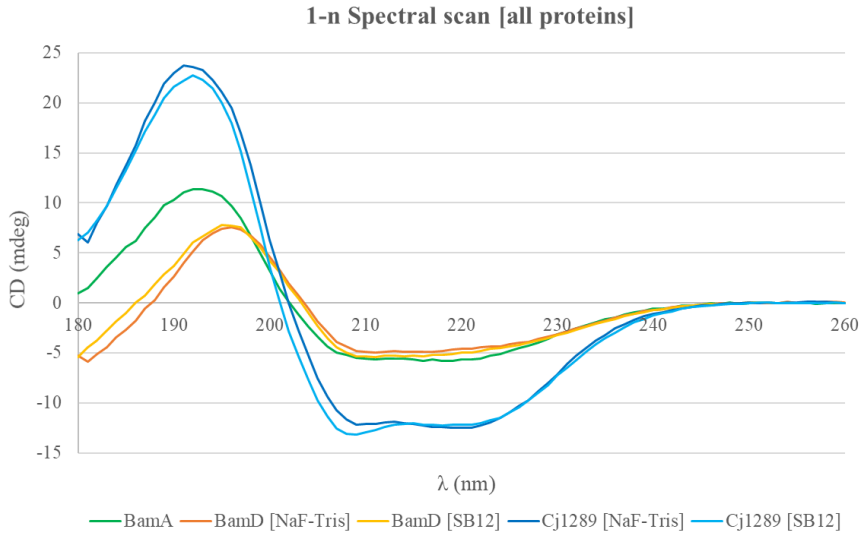


Figure 4.6: 1-n Spectral scan of all proteins pre-melt. All proteins were scanned at 20°C between 180-260 nm prior to the thermal degradation experiment to check the signal quality of the sample. BamD and Cj1289 are shown in both buffer backgrounds as both will be used in following thermal melts. All proteins are buffer subtracted, and show signal at all wavelengths without any obvious signal gaps, indicating suitable sample quality for CD experiments.

Sample	α-helical	β-sheet	Turns	Unordered
BamA	0.259	0.252	0.207	0.283
BamD	0.337	0.219	0.193	0.251
BamD [SB12]	0.449	0.133	0.177	0.241
Cj1289	0.427	0.135	0.156	0.282
Cj1289 [SB12]	0.465	0.096	0.162	0.277
BamA + BamD	0.311	0.195	0.214	0.280
BamA + Cj1289	0.321	0.178	0.204	0.297
BamD + Cj1289	0.396	0.140	0.190	0.273
BamA + BamD + Cj1289	0.235	0.268	0.205	0.292

Table 4.2: Experimental secondary structure estimation component values of all proteins in this study at 20°C before thermal degradation. The 1-n spectra scans for all proteins and their combinations were analysed using Cdapps with the CONTIN/LL algorithm and SMP56 reference set to generate SSEs. Component values for α-helices, β-sheet, turn and unordered structures are shown.

4.3.2.1 **BamA**

BamA secondary structure is relatively stable throughout the experimental temperature range. The thermal degradation experiment spectra show a gradual collapse of structure around 193 nm and a decrease in the negative signal between 205-225 nm as temperature increases throughout the experimental temperature range (figure 4.7(a)). The post-melt thermal degradation spectra reading is similar to that of the protein at 41°C, showing a lower positive CD signal in the 187-200 nm range and increased negative signal in the 208-220 nm range. The unfolding of BamA throughout the temperature range of the experiment was followed using a Boltzmann estimation at 192 nm (figure 4.7(b)). The estimated T_m was determined to be 44.19°C for BamA. The BamA SSE (figure 4.7(c)) shows α-helical content gradually decreasing throughout the experiment from 26 % to 018 %. Similarly, the β-stand, turn and unordered content increases from 25 % to 30 % throughout the experiment compensating for the lost α-helical content. The β-sheet and α-helical content reverts partially on sample return to 20°C towards starting secondary structure component values. The SSE goodness of fit diminished at discrete intervals through the experimental temperature range (appendix 8.4.2).

4.3.2.2 **BamD**

The addition of SB12 appears to increase stability of BamD throughout the temperature range of the experiment. In the thermal degradation experiment under non-detergent buffer conditions the CD signal starts blue-shifting around 53-56°C. The largest blue-shift occurs at 89°C and a partial red-shift when sample is returned to 20°C post-experiment (figure 4.8(a)). There is a loss of signal at 195 nm, suggesting a decrease in β-sheet content as temperature increases, and increased signal at (+)190 nm/ (-)208 nm/ (-)222 nm suggesting increased α-helical content as temperature increases. The post-experiment signal is intermediate between the pre-melt and 89°C spectra CD signal, resembling the CD signal intensity of either 29°C or 68°C. In contrast, there is less signal change at positive 195 nm and at negative 208 nm, 215 nm and 220 nm in the detergent buffer sample (figure 4.8(b)). The blue-shift in CD spectra occurs at a similar point (approximately 47°C), so the structural change causing this phenomenon may not be affected significantly by buffer background. On return to 20°C both spectra show a positive peak centred on 190 nm with gain in signal around 185 nm and loss of content at 195-200 nm.

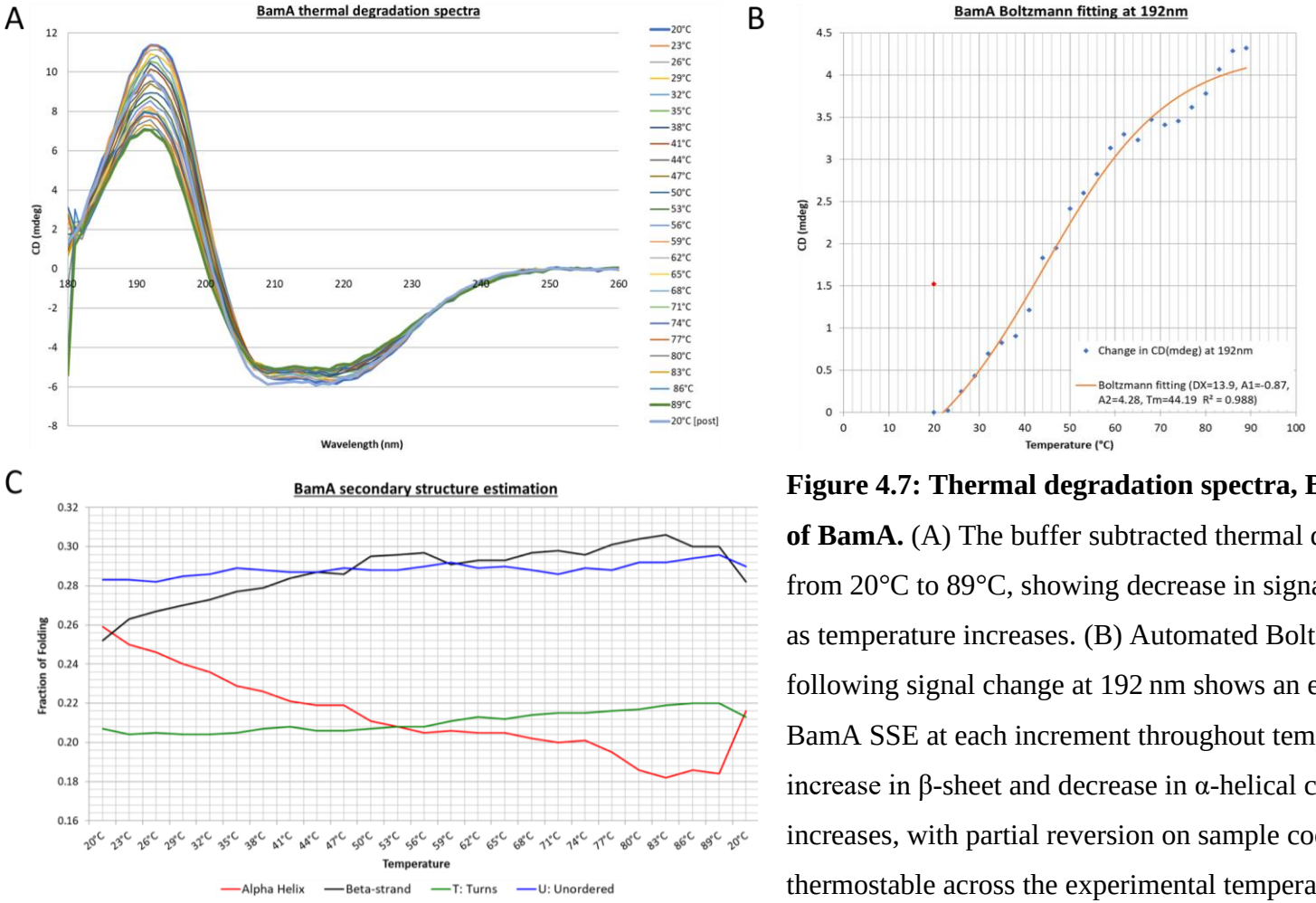


Figure 4.7: Thermal degradation spectra, Boltzmann fitting and SSE of BamA. (A) The buffer subtracted thermal degradation spectra of BamA from 20°C to 89°C, showing decrease in signal at 193 nm and 205-225 nm as temperature increases. (B) Automated Boltzmann fitting of BamA following signal change at 192 nm shows an estimated T_m of 44.19°C. (C) BamA SSE at each increment throughout temperature range. Gradual increase in β-sheet and decrease in α-helical content as temperature increases, with partial reversion on sample cooling. BamA is relatively thermostable across the experimental temperature range.

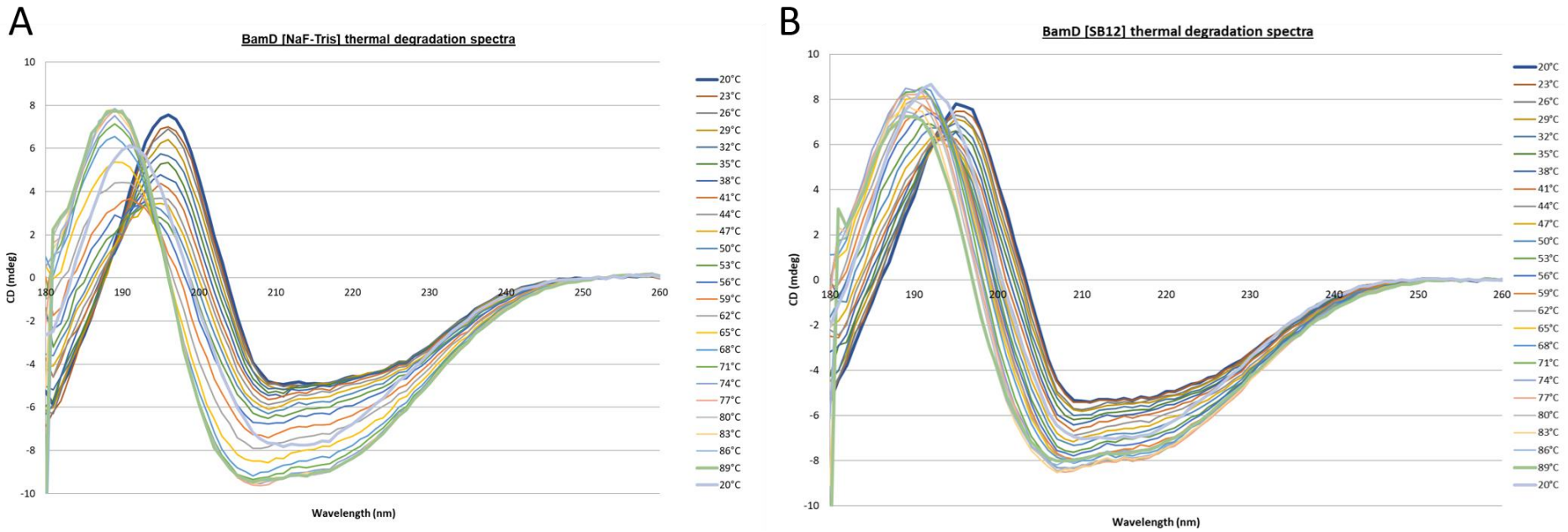


Figure 4.8: Thermal degradation spectra of BamD in both non-detergent [NaF-Tris] and detergent [SB12] buffer types shows increased CD signal stability in the presence of SB12. Buffer subtracted thermal degradation spectra of BamD in non-detergent (A) and detergent (B) buffers from 20℃ to 89°C. In both buffers the spectra exhibit a blue-shift after 53°C [NaF-Tris] and 47°C [SB12], respectively. A smaller, but earlier, blue-shift is observed in the DB sample, with less signal loss around 195 nm before blue-shift of the spectra. A larger CD signal change is observed in the NDB sample throughout, indicating increased stability of BamD secondary structures in the presence of detergent.

The CD spectra are more tightly packed together in the detergent sample, blue shift slightly less and have much less signal loss at 195 nm before blue shift of spectra, possibly indicating an improvement in thermal stability in the detergent buffer background before the structural transition to the unfolded state.

The unfolding of BamD as temperature was increased was followed using a Boltzmann estimation at 196 nm in NDB (figure 4.9(a)) and DB (figure 4.9(b)) backgrounds. The estimated T_m was determined to be 41.98°C (NDB) and 42.8°Cfor BamD. The post-experiment spectra point upon return to 20°C shows a change in CD signal by 4.5 mdeg relative to the pre-experiment measurement independent of buffer background (figure 4.9).

In the BamD [NaF-Tris] SSE (figure 4.10(a)) the α-helical content oscillates around 34 % until 62°C and then begins to increase to a 41 % total content at 80°C, and oscillates around 40 % until the upper temperature limit at 89°C. The β-sheet content gradually decreases to a low of 6 % total content at 74°C, levelling off and oscillating around 8 % total content after 80°C. Turn content increases steadily by 6 % total content to 25 % total content from 29°C to 62°C, and then returns to approximately starting value by 77°C and oscillates around 20 % ±1 % total content until the upper temperature limit. A decrease in both turn and β-sheet content appear to compensate for the increased α-helical content. Unordered content increases steadily in several stages by roughly 8 % total content throughout the experiment, with a pronounced increase after 65°C. The structure compositions are restored to within 2-4 % of starting values on sample return to 20°C post-experiment, indicating a subpopulation of the BamD sample refolds to the original structure, but a large subpopulation does not as the thermal degradation spectrum is blue shifted significantly (refer figure 4.8(a)). The SSE goodness of fit was good throughout, with minor depreciation at 23°C, and 47°C (appendix 8.4.2).

In the BamD [SB12] SSE (figure 4.10(b)) the α-helical content decreases gradually throughout the experiment, with sudden increases of 5-6 % total content between 56-59°C and 77-83°C. This signal change is mirrored in the change in β-sheet content, with a small reversible increase between 65°C and 77°C. Turn content increases throughout the experiment with a pronounced change after 77°C. Unordered content increases gradually throughout.

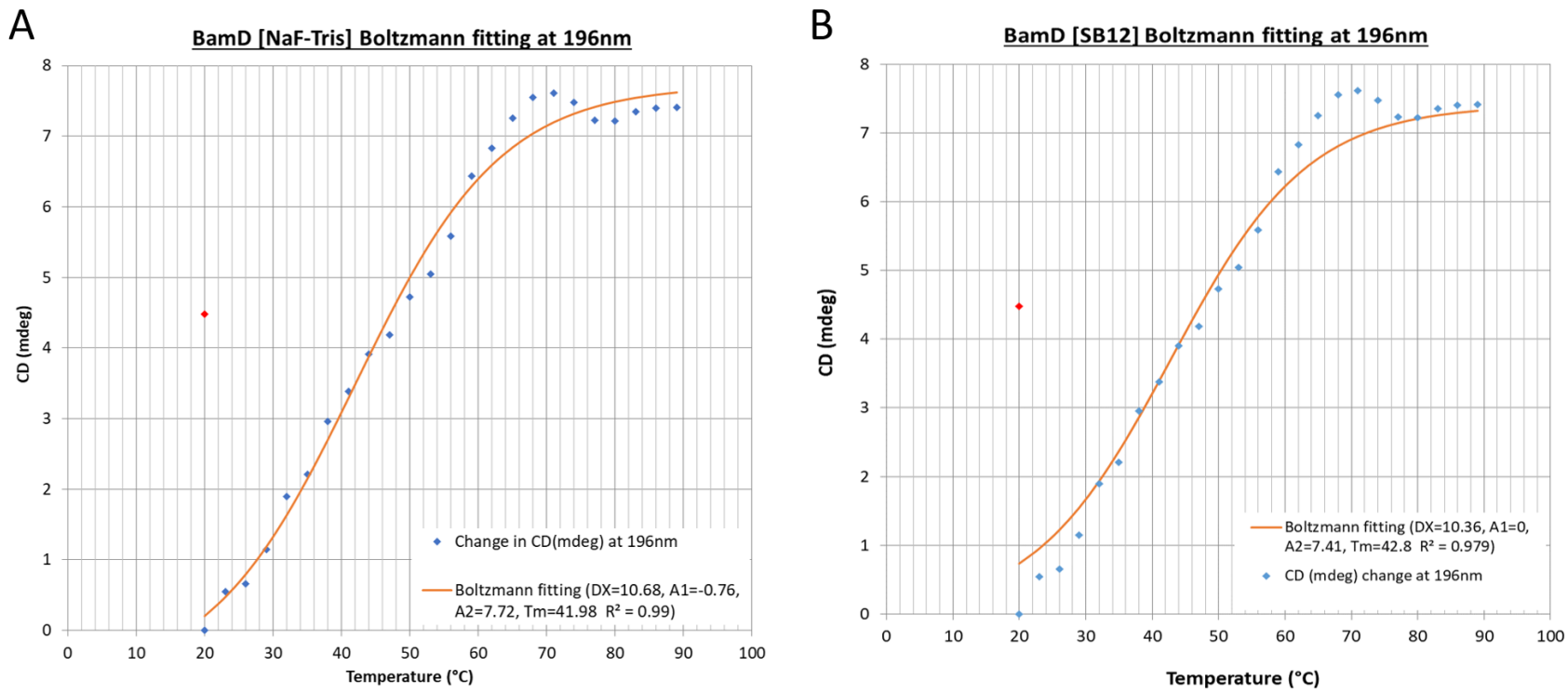


Figure 4.9: Boltmann fitting analysis of BamD in both non-detergent [NaF-Tris] and detergent [SB12] buffer types. An automated Boltzmann fitting of the BamD thermal degradation following changes at 196 nm with sample in NDB (A) and DB (B) background. The Boltzmann-derived T_m was determined to be 41.98°C (NDB, [NaF-Tris]) and 42.8°C (DB, [SB12]). The post-experiment spectra point upon return to 20°C is shown in red, and both samples differ to the pre-experiment measurement by +4.5 mdeg independent of the buffer background. The addition of detergent has no significant effect on the Boltzmann two state equation estimation of T_m for BamD.

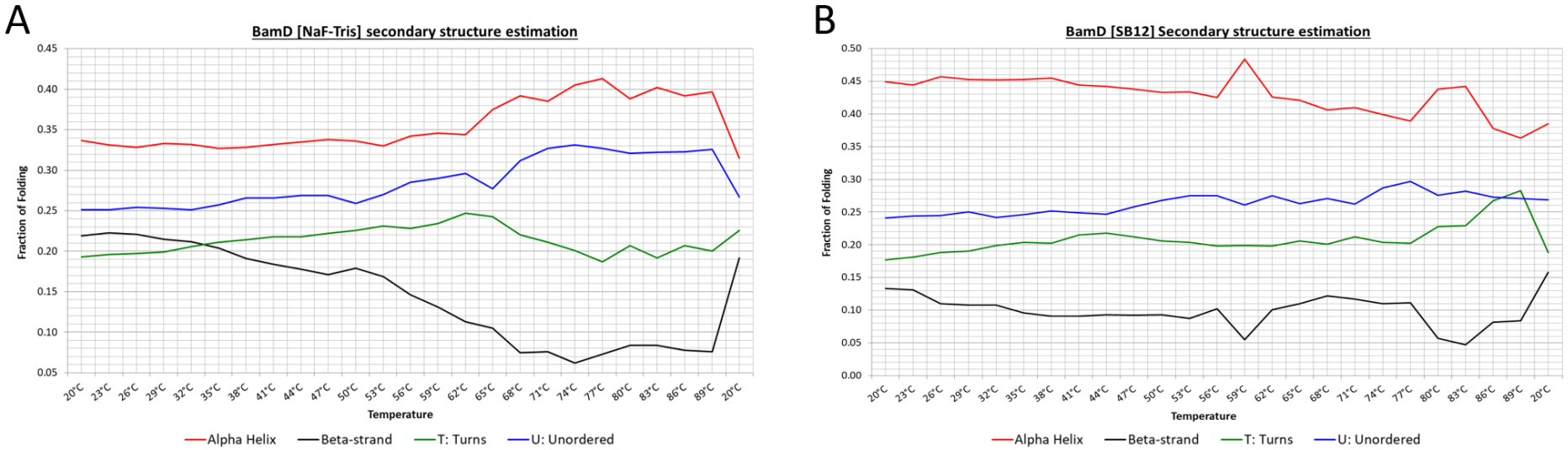


Figure 4.10: Secondary structure estimation of BamD in non-detergent [NaF-Tris] and detergent [SB12] buffer types. BamD SSE at each increment throughout the 20°C-89°C temperature range of the experiment in NDB [NaF-Tris] and DB [SB12] buffer backgrounds. (A) β-sheet content gradually decreases from 26°C to 77°C in the NDB sample, which coincides with increases in turn and unordered secondary structure initially and increases in α-helical and unordered secondary structures after 65°C. (B) In constrast to the non-detergent sample, the DB sample remains relatively stable in terms of secondary structure content until 59°C, where there appears to be a transition of β-sheet to α-helical structure temporarily, before larger changes around 77°C, with decreases in β-sheet content coinciding with increases in α-helical and turn secondary structures. Both samples show partial reversion of secondary structure composition towards sample starting values upon sample cooling, though these changes are more pronounced in the NDB sample. BamD secondary structures are stabilised in the presence of detergent.

The structure compositions are restored towards to within 2-4 % of starting values on sample return to 20°C post-experiment, but similarly to the non-detergent background sample, this likely indicates only a proportion of the sample has refolded. The SSE goodness of fit was good throughout, with some depreciation in the fitting at 59°C, and 77°C (appendix 8.4.2).

4.3.2.3 **Cj1289**

Addition of detergent improves the refolding of Cj1289, but reduces the Boltzmann T_m estimation. There is a gradual collapse of CD signal around 190 nm throughout the temperature range of the thermal degradation experiment in both DB and NDB samples (figure 4.11). The post-melt thermal degradation spectra is similar to the spectra at 68°C in the NDB sample and to 29/ 32°C in the DB sample, suggesting far better refolding of Cj1289 or a larger proportion of refolded Cj1289 proteins after degradation in the presence of detergent. The detergent sample CD spectra appear more compact throughout the experiment across temperatures at all wavelengths (e.g. 12.65 mdeg [DB] *vs.* 16.31 mdeg [NDB] change at 190 nm from 20°C to 89°C), possibly suggesting improved thermal stability in the presence of detergent.

The Boltzmann two state equation estimation (figure 4.12) gives a different perspective in terms of thermal stability, with the estimation suggesting the addition of detergent reduces the T_m from 60.04°C (NDB sample, figure 4.12(a)) to 34.04°C (DB sample, figure 4.12(b)). However, it does show a similar view on improved refolding in the presence of detergent, with the post-experiment CD measurement matching the pre-experiment CD signal far better (approximately +2.1 mdeg *vs.* approximately +12.8 mdeg difference to starting values).

In Cj1289 [NaF-Tris] SSE (figure 4.13(a)), the α-helical content decreases gradually from 32°C onwards by 22.5 % total content, with concurrent increase in turn content and from 50°C β-sheet content (by 14.5 % total content) and a gradual increase in turn content after 29°C to offset the change. The exchange of α-helical and β-sheet content becomes sharper after 68°C and both α-helical content and β-sheet content level off after 83°C. The α-helical and β-sheet content shows only roughly 5 % return to pre-melt value post sample cooling, suggesting poor refolding and irreversible thermal degradation in the NDB background. The SSE goodness of fit was good throughout, with minor depreciation at 59°C, and 83°C (appendix 8.4.2).

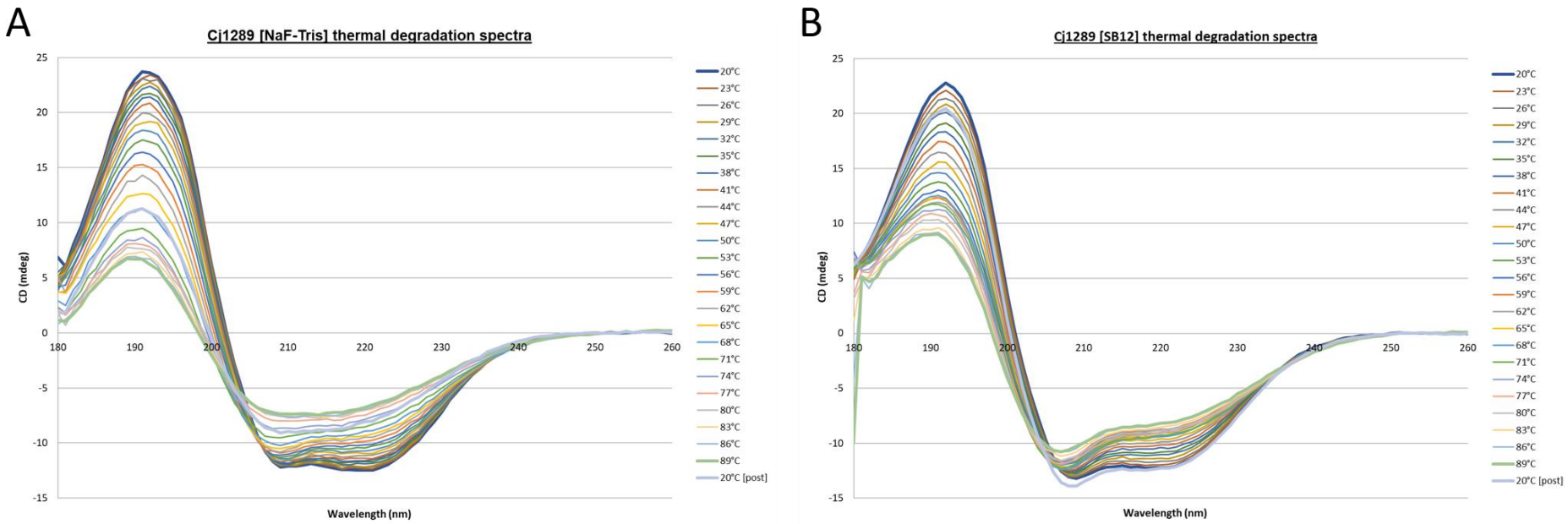


Figure 4.11: Thermal degradation profiles of Cj1289 in both buffer types. Buffer subtracted thermal degradation spectra of Cj1289 in non-detergent (A) and detergent (B) buffers from 20°C to 89°C. In both buffers there is a gradual collapse of Cj1289 CD signal at 190 nm and between 205 nm-225 nm as temperature increases through the experimental temperature range. The change in signal is less pronounced in the DB sample throughout the wavelength range measured, potentially indicating increased stability of Cj1289 in the presence of SB12.

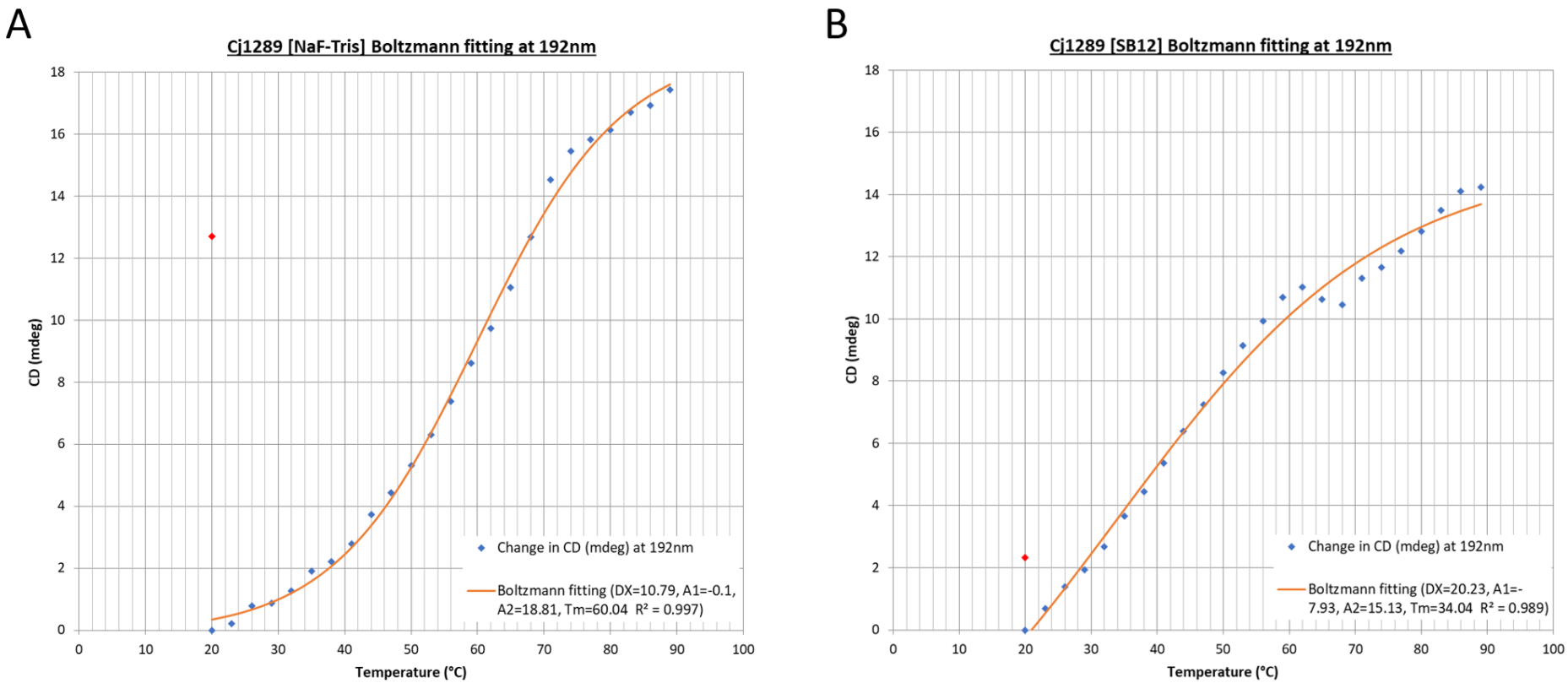


Figure 4.12: Boltmann fitting analysis of Cj1289 in both non-detergent [NaF-Tris] and detergent [SB12] buffer types. An automated Boltzmann fitting of the Cj1289 thermal degradation following changes at 192 nm with sample in NDB (A) and DB (B) background. The Boltzmann-derived T_m was determined to be 60.04°C (NDB, [NaF-Tris]) and 34.04°C (DB, [SB12]). The post-experiment spectra measurement point at 20°C is shown in red, and a change of 12.69 mdeg (NDB) and 2.32 mdeg (DB) is observed from pre-experiment values at 192 nm, respectively.

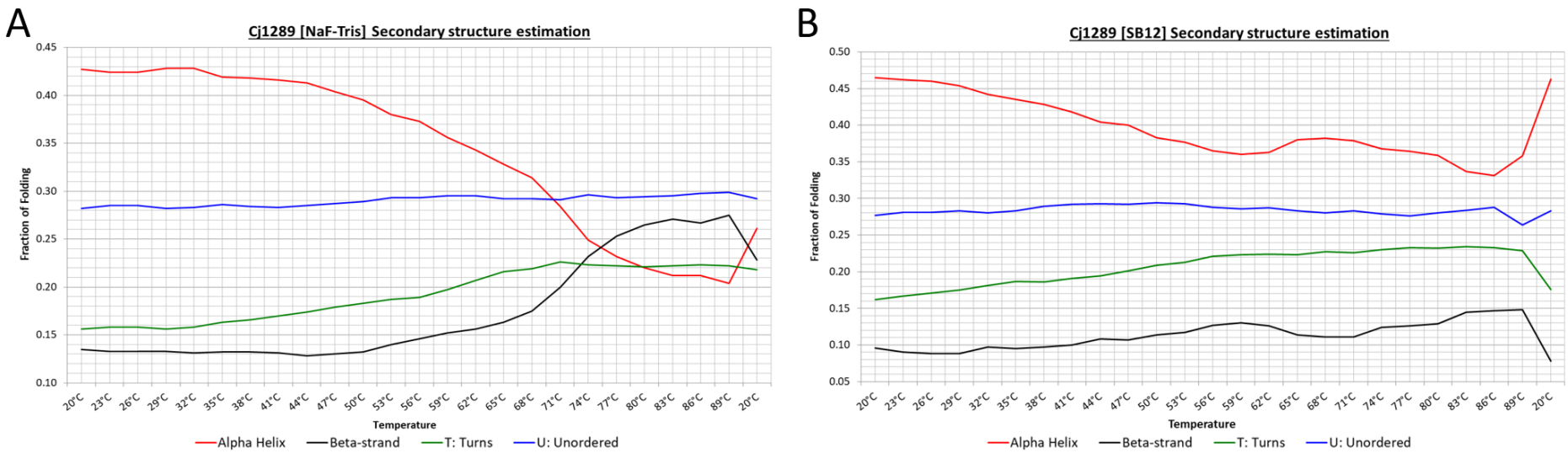


Figure 4.13: Secondary structure estimation of Cj1289 in non-detergent [NaF-Tris] and detergent [SB12] buffer types. Cj1289 SSE at each increment throughout the 20°C-89°C temperature range of the experiment in NDB [NaF-Tris] and DB [SB12] buffer backgrounds. (A) α-helical secondary structure gradually decreases from 32°C onwards by 22.5% total content in the NDB sample, which coincides with increases an initial increase in turn, and after 50°C, β-sheet secondary structures by 6.5% and 14.5%, respectively. Unordered content minorly increases approximately 1.5% over the experiment temperature range. (B) In constrast to the NDB sample, the DB sample shows a smaller decrease in α-helical secondary structure over two stages, from 26°C to 62°C and from 65°C to 86°C, which are mirrored by increases in β-sheet and turn secondary structure. Cj1289 is stabilised in SB12. The NDB sample shows little reversion (approximately 5% for β-sheet/ α-helices) of secondary structures towards starting values upon sample cooling, in sharp contrast to the DB sample which shows values close to starting values post-sample cooling, suggesting a significantly larger proportion of Cj1289 proteins refold post-experiment in the presence of SB12.

Unordered structural content remains relatively steady throughout, varying by ± 1-2 %, and turn content increased by around 7 % overall.

In the detergent sample SSE (figure 4.13(b)), the α-helical content decreases gradually in two phases, the first starting at around 26°C until 62°C, and the second from 65°C until 86°C. This change is mirrored in the β-sheet component, which increases gradually in two phases (29°C to 59°C, 71°C to 89°C), and a gradual increase in turn content. The unordered content stays relatively stable throughout around 28 % ± 1-2 %. The structural content changes less throughout relative to the starting sample when compared to the non-detergent sample. The post-experiment spectrum shows that all structural content returns to approximately pre-melt values (within 1 % for helical, within 2 % for turn, within 2 % for β-sheet, and exactly for unordered)). This suggests good refolding and a large proportion of refolded Cj1289 proteins after thermal degradation. The SSE goodness of fit was good throughout except at 77°C, where it was poor (appendix 8.4.2).

4.3.2.4 **BamA interacts with BamD**

The thermal degradation spectra show the largest change in signal is observed around 190-194 nm (figure 4.14(a)), with +11.1 mdeg pre-melt CD signal decreasing to +8.2 mdeg at 89°C and increasing to +10.6 mdeg post-experiment at 20°C. The CD signal is also slightly blue-shifted after 62°C. As blue shifting of the spectra was not observed in the BamA thermal degradation, but was prominent in the BamD thermal degradation, it is likely due to the BamD component. The CD signal stays relatively stable throughout the temperature range compared to either BamA or BamD individually, and the limited blue shifting of the spectra indicates greatly increased thermal and/ or structural stability of BamD.

The Boltzmann two state equation estimation at 194 nm (figure 4.14(b)) shows an increase in T_m when compared to the individual components of 2.6-6.0°C. The pre- and post-experiment spectra point differs by +1.6 mdeg, suggesting reasonably good refolding post-melt. There is a clear break in continuity after 59°C, suggesting that there are two transitions in the thermal degradation, from (1) 20°C to 71°C, (2) 71°C onwards, though the experiment temperature range does not allow the second curve to be fully plotted.

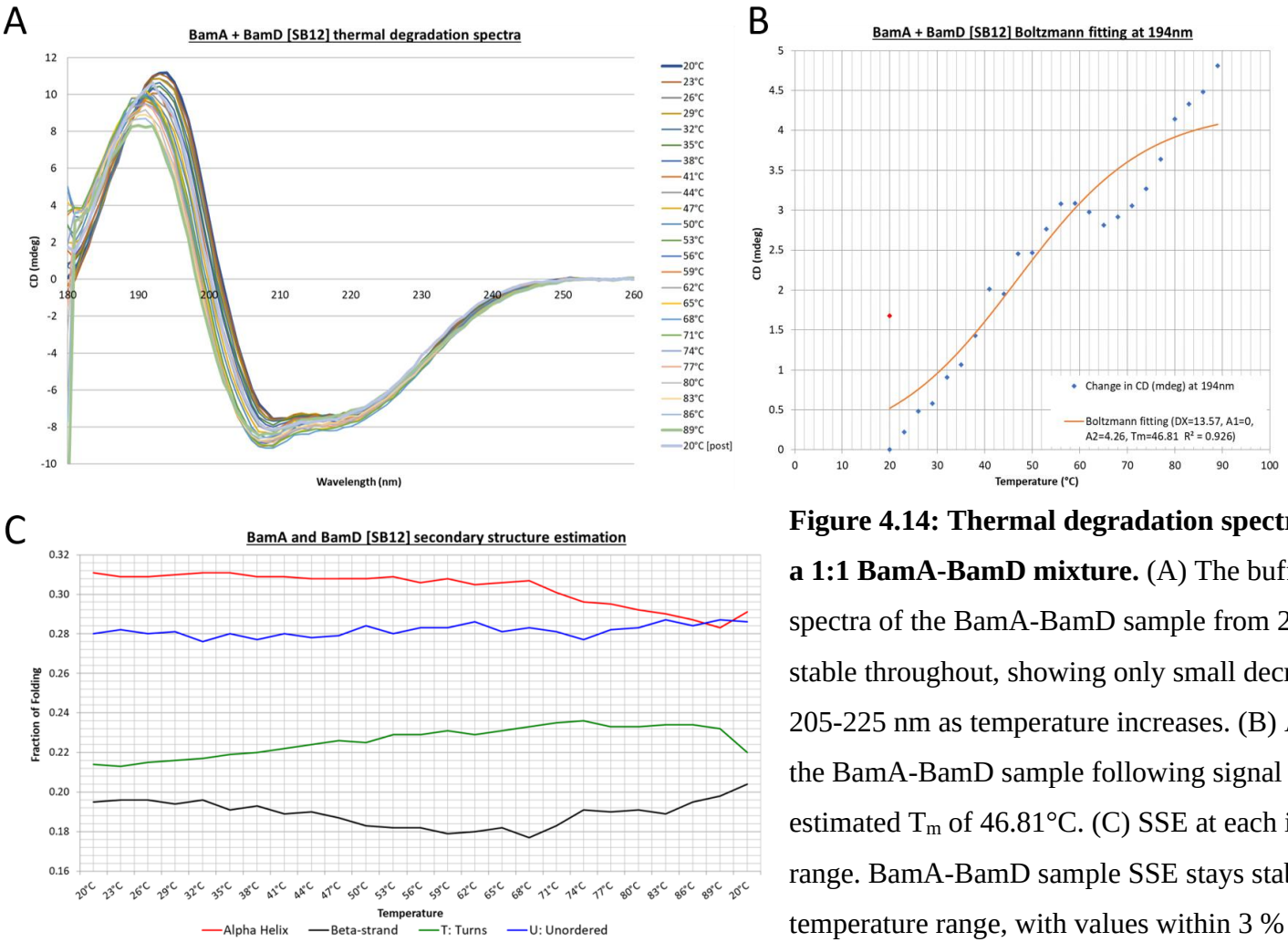


Figure 4.14: Thermal degradation spectra, Boltzmann fitting and SSE of a 1:1 BamA-BamD mixture. (A) The buffer subtracted thermal degradation spectra of the BamA-BamD sample from 20°C to 89°C. The spectra stays stable throughout, showing only small decreases in signal at 190-194 nm and 205-225 nm as temperature increases. (B) Automated Boltzmann fitting of the BamA-BamD sample following signal change at 194 nm, showing an estimated T_m of 46.81°C. (C) SSE at each increment throughout temperature range. BamA-BamD sample SSE stays stable throughout the experimental temperature range, with values within 3 % of starting values.

In contrast to the individual protein degradations the BamA and BamD (1:1) mixture SSE (figure 4.14(c)) shows α-helical and β-sheet content decrease very little (~1 %), remaining comparatively stable up until 68°C, at which point α-helical content decreases and β-sheet increases to offset this change. The post-experiment measurement shows β-sheet returning to pre-melt value and alpha helices decreasing by 3 % overall. Turn content increases gradually by 2 % and unordered content oscillates around 28 % throughout. The secondary structure content remains more stable overall in the BamA-BamD mixture compared to either of the BamA or BamD individually. The SSE goodness of fit depreciated after 77°C (appendix 8.4.2).

To determine whether the SSE was simply the additive result of mixing the proteins, the spectra were either added together with equal weighting (figure 4.15(a)) and compared to the experimental SSE (figure 4.15(b)). The additive SSE suggests there would be a gradual decrease of α-helical content and increase of β-sheet content with more pronounced changes between 56-59°C and 77-86°C, turn content would rapidly increase after 83°C, and unordered content would increase slowly throughout. As the BamA and BamD mixture does not match the individual or additive (non-interaction) SSE patterns, remaining more stable throughout for all secondary structure content, it is highly likely that the proteins interact to stabilise each other and melt in cooperatively as observed in protein co-melting phenomena.

4.3.2.5 **BamA does not show a co-stabilising interaction with Cj1289**

The BamA-Cj1289 1:1 (figure 4.16(a)) and 1:3 (figure 4.16(b)) mixtures show a large change in CD signal at 190 nm and between 208-225 nm, with +14.0/+12.9 mdeg pre-melt CD signal decreasing to +6.4/+5.3 mdeg at 89°C and increasing to +11.1/+10.1 mdeg post-experiment at 20°C in the 1:1/1:3 stoichiometry samples, respectively. The thermal degradation spectra look to be more similar to those for Cj1289 alone, but the post-experiment 20°C signal matches the 41°C spectra as in the BamA individual melt. As characteristics of both individual protein degradation profiles are prevalent in the mixture to the same extent as the individual proteins, this may suggest that there is no interaction or stabilisation between BamA and Cj1289 and that both proteins undergo degradation independently. As both CD profiles are similar, there appears to be no additional impact (stabilisation or de-stabilisation) in changing the stoichiometric ratio.

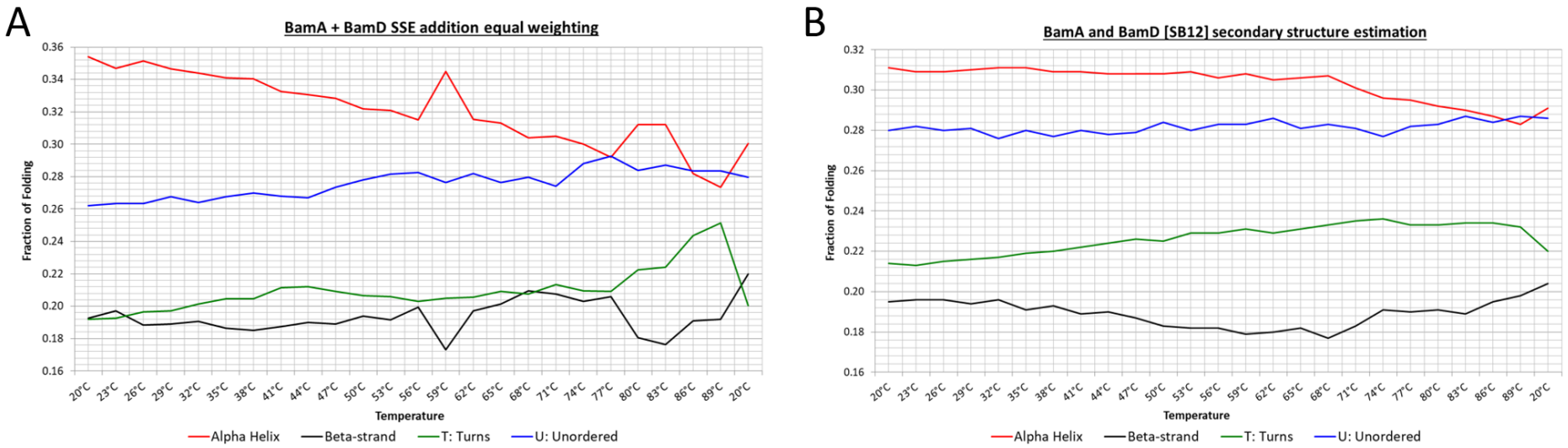


Figure 4.15: BamA-BamD mixture additive spectra versus experimental data. To predict the SSE of a non-interacting mixture of BamA and BamD, the individual spectra were added together in equal weighting (A). The actual experimental data for the mixing of BamA and BamD (B) is shown to aid comparison. The additive SSE (A) predicts a gradual decrease of α-helical content and increase of β-sheet content, with pronounced changes between secondary structure types from 56-59°C and 77-86°C, and a sharp increase in turn content to compensate for the sharp decrease in α-helical content after 77°C. Additionally, unordered content was predicted to increase slowly throughout the experiment, but remains stable in the experimental SSE. As the BamA and BamD mixture experimental SSE (B) does not match the individual protein or additive (non-interaction) SSE pattern (A), with secondary structures remaining more stable throughout, it is highly likely that BamA and BamD interact to co-stabilise their structures and undergo co-melting.

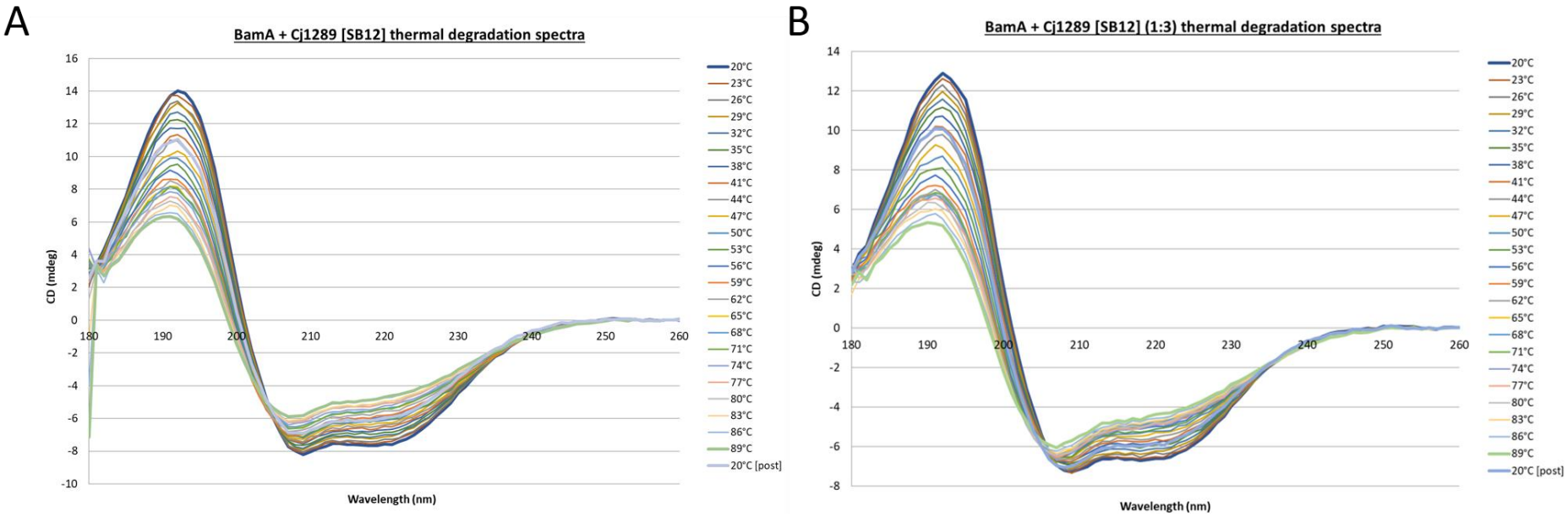


Figure 4.16: Thermal degradation profiles of BamA-Cj1289 mixtures in 1:1 and 1:3 stiochiometries. Buffer subtracted thermal degradation spectra of BamA-Cj1289 mixtures in 1:1 (A) and 1:3 (B) stoichiometric ratios from 20°C to 89°C. In both stoichiometries there was a gradual collapse of CD signal at 190 nm and between 205 nm-225 nm as temperature increases through the experimental temperature range, similar to the Cj1289 [SB12] individual protein spectra. The changes in signal were similar regardless of stoichiometry, and increasing Cj1289 relative to BamA only increases the likeness of the spectra to the Cj1289 [SB12] individual melt.

The Boltzmann two state equation estimation at 192 nm for the BamA-Cj1289 mixture in 1:1 (figure 4.17(a)) and 1:3 (figure 4.17(b)) stoichiometric ratios gives an estimated T_m of 46.84°C ± 0.55°C. The small difference in estimated T_m between the stoichiometries suggests that the altering the stoichiometry does not affect the interaction properties between BamA and Cj1289, corroborating lack of interaction. The discontinuity in the data at 71°C in both experiments suggests there are two structural transitions. The first was likely the melt for Cj1289 (and potentially the BamA POTRA domains), and the second for BamA β-barrel.

In the BamA-Cj1289 mixture SSE (figure 4.18), there was a large decrease in α-helical and concurrent increase in β-sheet structural content throughout the experiment temperature range, with values intersecting at 62°C (1:1) and 47°C (1:3). The unordered content stays stable through the majority of the experiment around 29 %±1 % with the exception of an initial dip (after 32°C for 1:1, 26°C for 1:3), and turn content gradually increases by 2.5 % overall. There was a large change between the post-experiment 20°C and 89°C, with a large increase in alpha helical content and decrease in all other structural types. The initial dip in unordered content and post-experiment structural composition change are more pronounced in the 1:3 stoichiometric ratio sample. This mimics the change observed for Cj1289 alone, but the intersection of α-helical and β-sheet is earlier (62°C [1:1]/ 47°C [1:3] *vs.* 72-74°C [Cj1289 individually]), likely due to contribution of the BamA degradation which follows a similar pattern of decreasing α-helices/ increasing β-sheet. While the change in structure appears larger in the 1:3 stoichiometric ratio sample, the overall content transition is 0.119 [1:1]/ 0.115 [1:3] (helices) and 0.095 [1:1]/ 0.076 [1:3] (sheet) was quite similar between the samples. The SSE goodness of fit was poor at discrete intervals for both stoichiometries (appendix 8.4.2).

The additive spectra show a different pattern of change in secondary structure (figure 4.19) to the experimental data for β-sheet and turn structure, but similar changes in α-helical structure with increasing temperature regardless of stoichiometry. The similarity in α-helical structure change (roughly 12% during the experiment) suggests that the spectra are simply the result of the additive signal from the individual proteins (i.e. no interaction), however the differences in β-sheet and turn structure may suggest there is some form of (likely transient) interaction (i.e. TPCA) between the proteins that cannot be explain by simple mixing.

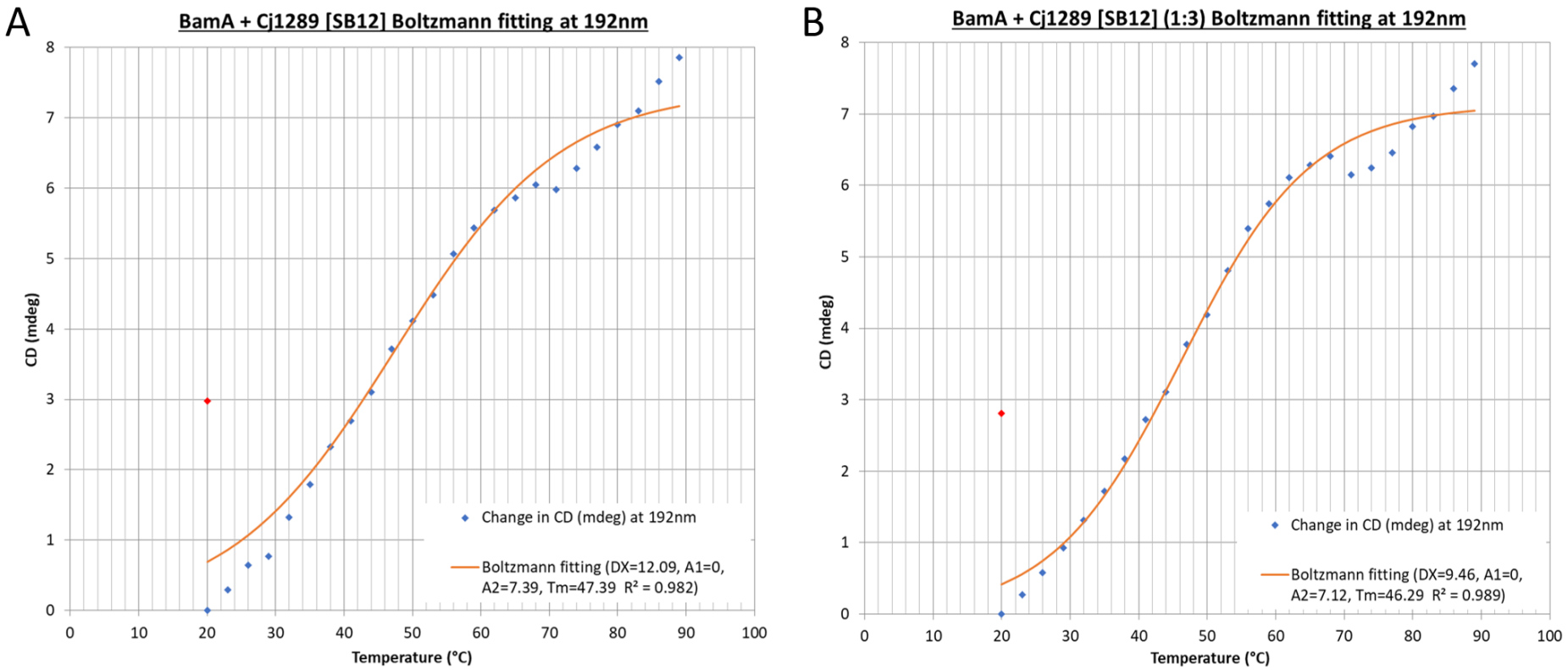


Figure 4.17: Boltmann fitting analysis of BamA-Cj1289 mixture in 1:1 and 1:3 stoichiometry in detergent [SB12] buffer. An automated Boltzmann fitting of the BamA-Cj1289 thermal degradation following changes at 192 nm with sample mixture in 1:1 (A) and 1:3 (B) stiochiometry in DB background. The Boltzmann-derived T_m was determined to be 47.39°C (1:1) and 46.29°C (1:3). The post-experiment spectra measurement point at 20°C is shown in red, and a change of 2.97 mdeg (1:1) and 2.81 mdeg (1:3) is observed from pre-experiment values at 192 nm, respectively. The similarity between the fittings suggest alteration of stiochiometry has little effect on structural transition.



Figure 4.18: Secondary structure estimation of BamA-Cj1289 mixture 1:1 and 1:3 stoichiometry. BamA-Cj1289 mixture SSEs at each increment throughout the 20°C-89°C temperature range of the experiment in 1:1 (A) and 1:3 stiochiometry (B) in DB background. (A) α-helical secondary structure gradually decreases from starting temperature by 12 % total content in the 1:1 stiochiometry sample, which coincides with a 10 % increase in β-sheet secondary structures after 35°C and gradual 2 % increase of turn content as temperature increases. Unordered content oscillates between 28-30 % throughout the experiment temperature range. (B) The 1:3 stoichiometry sample shows a similar change in structure to the 1:1 stiochiometry sample as temperature increases, except the transition point between α-helical and β-sheet structure composition occurs earlier (47°C *vs.* 62°C), which may reflect the increased proportion of Cj1289 in the sample, as Cj1289 showed reduced thermal stability in the presence of SB12 when melted individually. Both samples show restoration of approximately 7% α-helical and loss of approximately 3% β-sheet structures after sample cooling, with turn strucutral content returning to starting values and unordered conent approximately 2% less than at the experiment start point. Given this data, a BamA and Cj1289 co-stabilising interaction is unlikely.

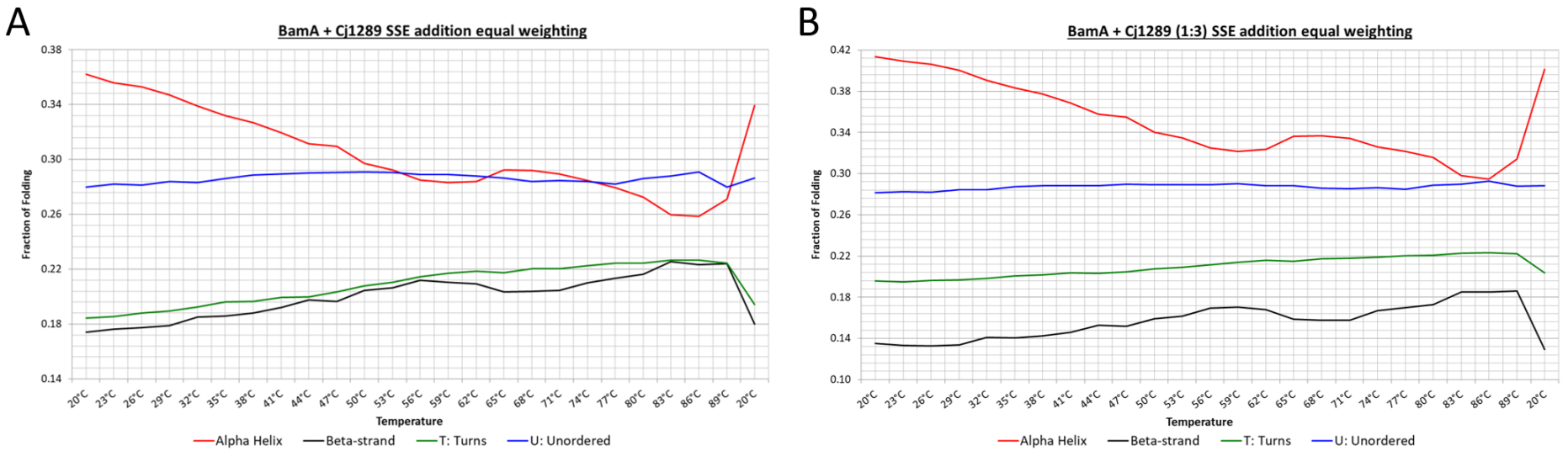


Figure 4.19: BamA-Cj1289 mixture additive spectra versus experimental data. To predict the SSE of a non-interacting mixture of BamA and Cj1289, the individual spectra were added together in equal weighting (A) or 1:3 stoichiometry (B). The additive SSEs for the 1:1 (A) and 1:3 (B) stoichiometry mixtures of BamA-Cj1289 predict an approximately 12% decrease in α-helical structure similar to the experimental data, however, a lesser increase in β-sheet (5% *vs.* 8-9%) and greater increase in turn (3-4.5% *vs.* 2%) secondary structures is predicted relative to the experimental data (refer figure 4.19). The difference being the lack of increase in β-sheet structure after 50°C/ 56°C and lack of decrease of turn structure around 53°C/ 47°C in the 1:1/1:3 stoichiometry samples, respectively. The similarity between the α-helical structure decay suggests the experimental SSEs can be explained by the summative non-interacting spectra of BamA and Cj1289, however, the difference in β-sheet and turn structure content suggests some form of interaction occurred to facilitate the increase in formation of β-sheet structures.

4.3.2.6 **BamD interacts with Cj1289 in 1:1 stoichiometry**

In the 1:1 BamD-Cj1289 mixture (figure 4.20(a)) in NDB background, there is a gradual decrease in signal around 190-195 nm, with a progressive blue shift of the peak from 195 nm to 190 nm as the temperature increases. Upon sample return to 20°C post-degradation, there is a minor red shift and the CD signal increases to similar levels to the 62°C spectra. This suggests some attributes, such as the poor refolding post thermal degradation in Cj1289 (NDB) and (blue- and red- shift) observed in BamD (NDB) have bled through into the spectra of the mixture. Changing the stoichiometric ratio of BamD-Cj1289 to either 1:3 or 3:1 (figure 4.21) results in the CD spectra taking up similar patterns to the more abundant protein in the mixture. In stoichiometries with an excess of BamD in the mixture (figure 4.21(a) and (c)) there is a bleed through of the BamD thermal degradation profile, as shown by the blue shift of the spectra around 190 nm, suggesting that the stoichiometry is unlikely to be one where multiple BamD monomers interact with a single Cj1289 monomer. However, when the stoichiometry ratio is reversed with three Cj1289 molecules to a single BamD (figure 4.21(b) and (d)), the thermal denaturation profile appears similar to the 1:1 stoichiometric ratio profile (figure 4.21(a)). This could suggest that more than one Cj1289 protein can interact with a single BamD. The post-experiment CD signal at 20°C shows a better CD signal recovery in the presence of detergent.

The Boltzmann two state equation estimation at 193 nm for the 1:1 mixture (figure 4.20(b)) shows a T_m estimation (50.13°C) intermediate to those obtained for the individual proteins (41.98°C [BamD] and 60.04°C [Cj1289]). The post-experiment CD signal at 20°C is similarly intermediate (+6.8 mdeg compared to +4.5 mdeg [BamD]/ +12.8 mdeg [Cj1289]), with a slight skew in CD signal to that of the BamD sample. The data looks continuous with no obvious break in continuity in the data until 74°C, after which the data becomes noisy. Changing the stoichiometric ratio of BamD-Cj1289 to either 1:3 or 3:1 (figure 4.22), increases the estimated T_m by 6-8°C in the NDB background. The addition of detergent reduced the estimated T_m regardless of the stoichiometric ratio sampled, and reduces the change in CD signal at 20°C post-experiment compared to pre-experiment values.

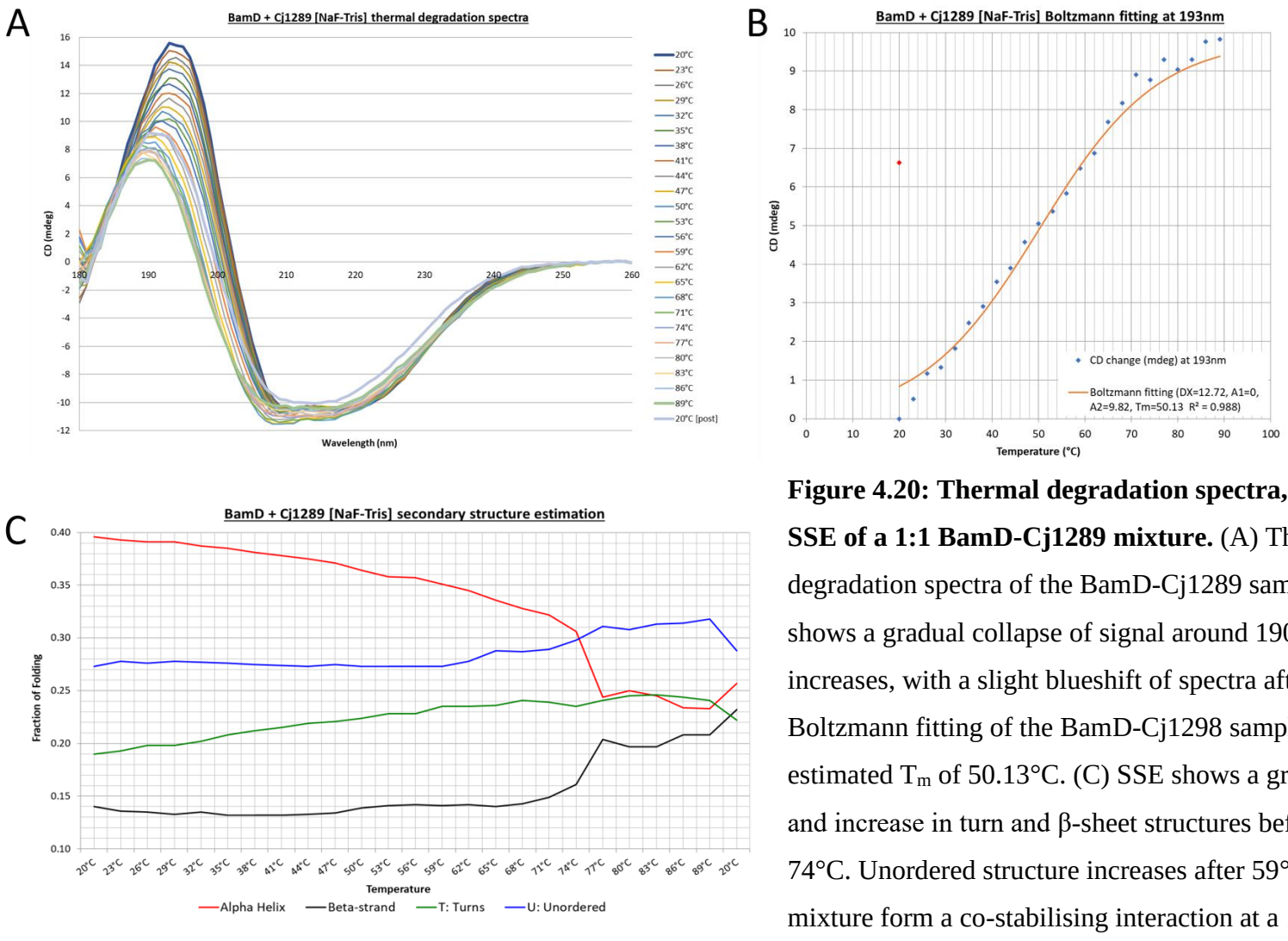


Figure 4.20: Thermal degradation spectra, Boltzmann fitting and SSE of a 1:1 BamD-Cj1289 mixture. (A) The buffer subtracted thermal degradation spectra of the BamD-Cj1289 sample from 20°C to 89°C shows a gradual collapse of signal around 190 nm as temperature increases, with a slight blueshift of spectra after 59°C. (B) Automated Boltzmann fitting of the BamD-Cj1298 sample at 193 nm, showing an estimated T_m of 50.13°C. (C) SSE shows a gradual decrease in α-helical, and increase in turn and β-sheet structures before a sharp change after 74°C. Unordered structure increases after 59°C. BamD and Cj1289 in mixture form a co-stabilising interaction at a 1:1 stoichiometry.

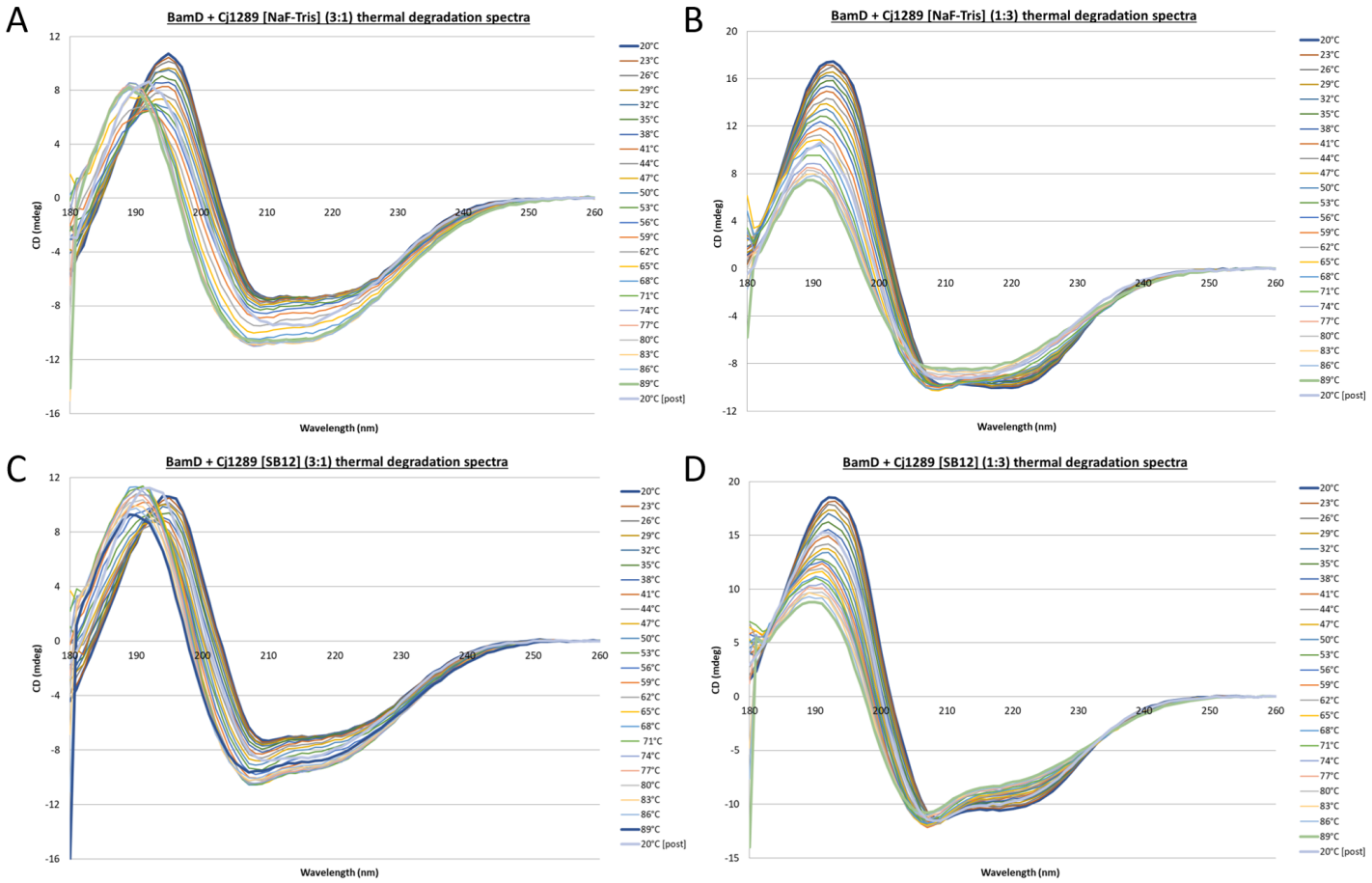


Figure 4.21: Thermal degradation spectra of BamD-Cj1289 mixtures in 3:1 and 1:3 stoichiometries in non-detergent and detergent buffer backgrounds. The spectra more closely resemble the protein in higher concentration in the BamD-Cj1289 mixtures, with 3:1 NDB (A) and 3:1 DB (C) more closely resembling the profile of BamD, and 1:3 NDB (B) and 1:3 DB (D) more closely resembling the profile of Cj1289.

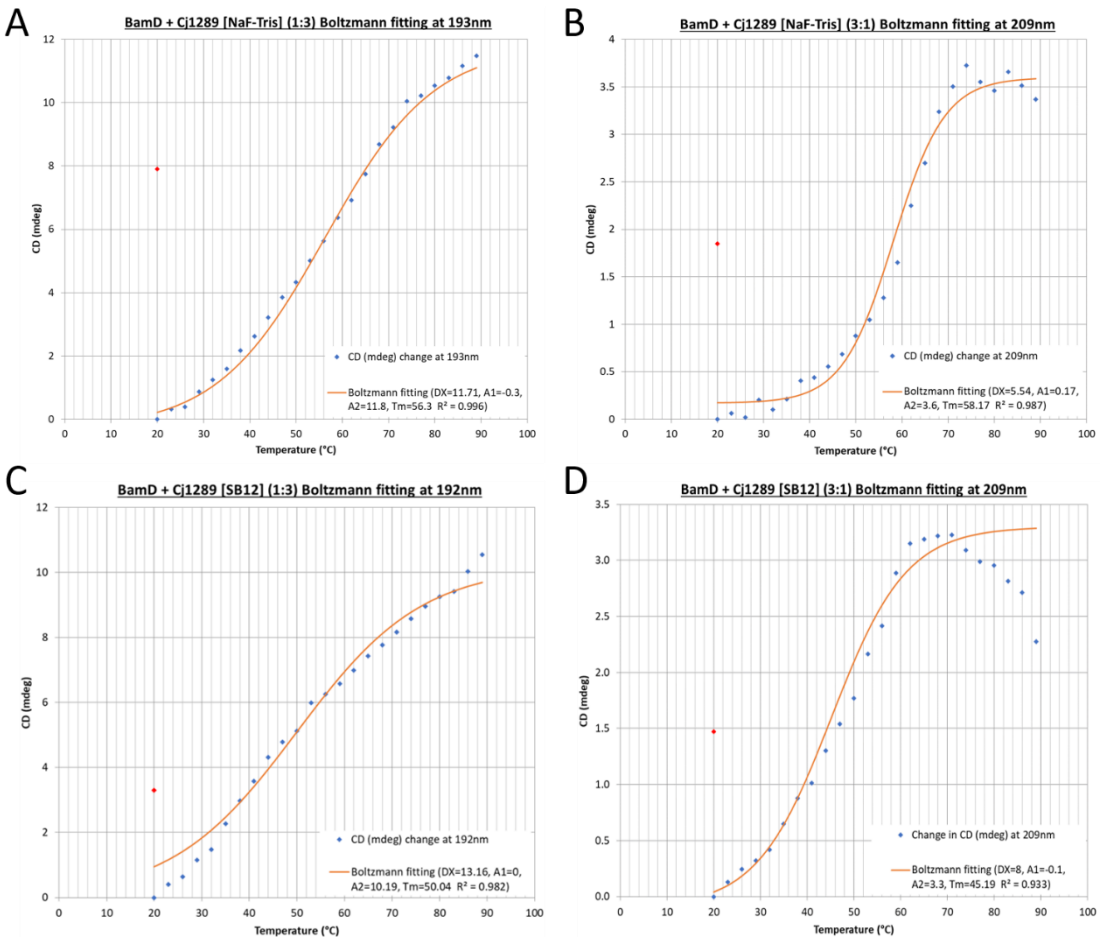


Figure 4.22: Boltmann fitting analysis of BamD-Cj1289 mixture in 1:3 and 3:1 stoichiometry in NaF-Tris and SB12 buffer. An automated Boltzmann fitting of the BamD-Cj1289 thermal degradation following changes at 193 nm (A) and192 nm (C) for the 1:3 stiochiometry mixtures and 209 nm for the 3:1 (B, D) stoichiometry mixtures. The Boltzmann-derived T_m was determined to be (A) 56.30°C (1:3, NDB), (B) 58.17°C (3:1, NDB), (C) 50.04 (1:3, DB), and (D) 58.17°C (3:1, DB), respectively. In both stoichiometries, the addition of detergent reduces the automated Boltzmann T_m calculated relative to the NDB samples by 6.26°C (1:3) and 12.98°C (3:1), respectively. The post-experiment spectra measurement at 20°C (red diamond) suggests the addition of detergent likely improves refolding outcomes for both sample stoichiometries, with smaller differences in CD signal (mdeg) post-experiment relative to the starting samples shown in the DB samples.

The SSE for the 1:1 BamD: Cj1289 mixture in the NDB background (figure 4.20(c)) behaves differently to the individual components, showing a Cj1289-like α-β transition, but with SSE staying stable until 74°C as opposed to a gradual transition for Cj1289, and a β-sheet to α-helical transition for BamD in the NDB background. This increased stabilisation over Cj1289 and the difference in transition compared to BamD, indicate that an interaction is occurring at a 1:1 stoichiometry between BamD and Cj1289. Contrary to this, the altered stoichiometry BamD and Cj1289 mixture SSEs in NDB background show structural changes similar to the higher abundance component, with the 1:3 BamD: Cj1289 SSE (figure 4.23(a)) resembling the Cj1289 SSE, and the 3:1 BamD: Cj1289 SSE (figure 4.23(b)) resembling BamD SSE. The addition of detergent reduces the change in structure with temperature in both stoichiometries (figure 4.23(c) and (d)).

The additive equal weighting SSE prediction (figure 4.24(a)) for the 1:1 BamD: Cj1289 mixture suggests there would be a gradual decrease of α-helical content throughout and β-sheet content until 68°C, where there is an inflection point and subsequent increase in β-sheet until the end of the experiment. Turn content would increase after 29°C with an inflection point at 65°C, followed by a decrease, later oscillating around 21 % from 77°C until experiment end. Unordered content gradually increases after 32°C throughout up until the upper temperature limit, and decreases by 5 % total content on sample cooling. In the actual experiment (figure 4.24(b)), the SSE is quite different, as described above, which suggests that BamD and Cj1289 interact, and thermally degrade with a co-melting phenomenon.

The additive equal weighting predictions for SSE were carried out for the 1:3 and 3:1 BamD: Cj1289 stoichiometry samples in both buffer backgrounds and the predictions show gross similarities to the actual experimental SSEs in all cases (figure 4.25). This indicates that at least part of the experimental SSEs are the result of non-interacting particles undergoing melting individually. The minor differences between the additive and experimental SSEs at all BamD-Cj1289 stoichiometries are likely due to the 1:1 interaction between BamD and Cj1289 stabilising a sub-population of the sample, and the differences, which resemble the individual protein melts in all cases, are likely due to the non-interacting excess of the higher abundance protein. This indicates that the interaction between BamD and Cj1289 is 1:1 and that additional units of either protein does not lead to increased stability.

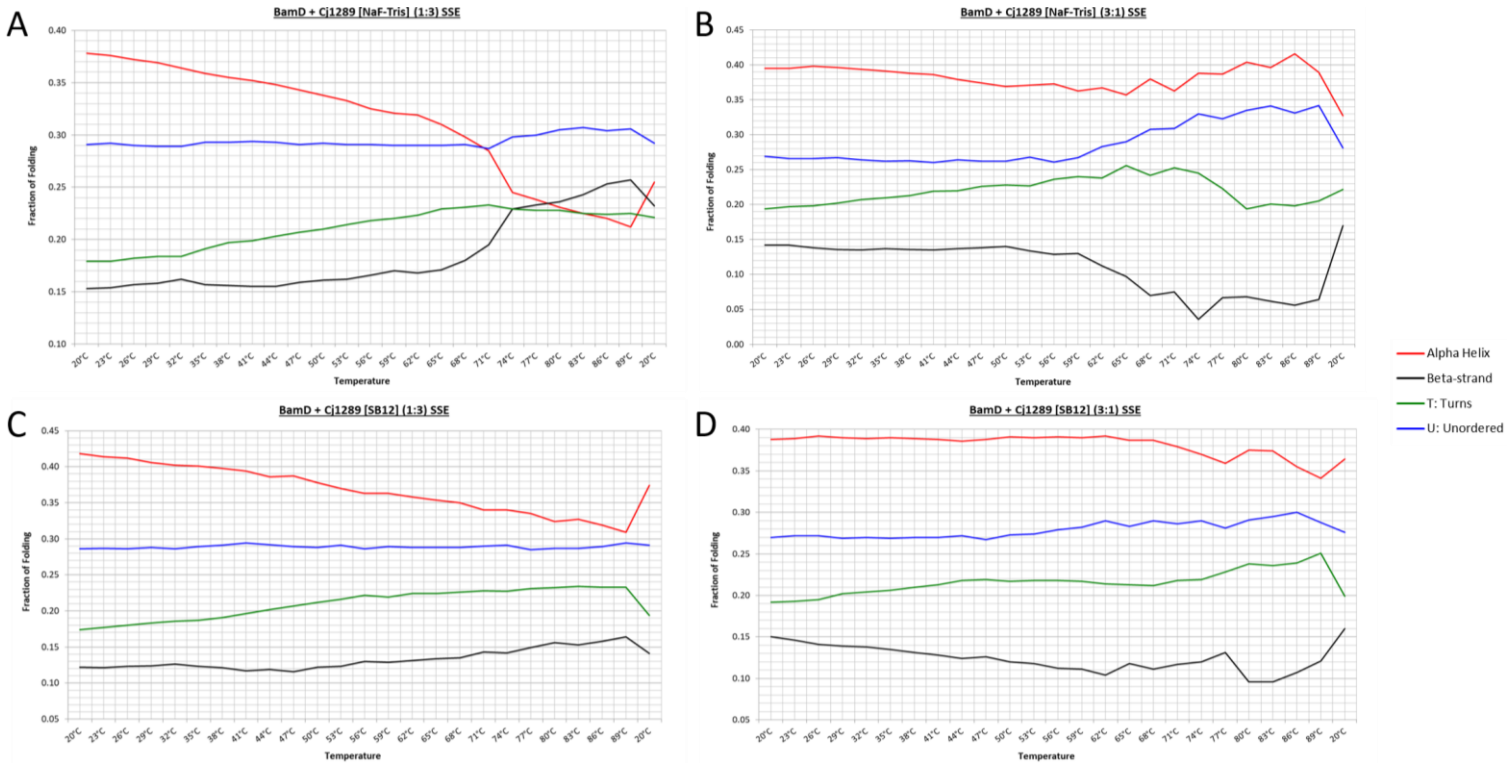


Figure 4.23: Secondary structure estimation of BamD-Cj1289 mixtures in 3:1 and 1:3 stoichiometries in non-detergent and detergent buffer backgrounds. BamD-Cj1289 mixture SSEs of thermal degradation experiments in 1:3 (A, C) and 3:1 stiochiometries (B, D) in non-detergent (A, B) and detergent (C, D) buffer backgrounds. There is a gradual decrease in α-helical and increase in β-sheet and turn structure in the 1:3 NDB sample throughout the temperature range (A), whereas the 3:1 NDB (B) sample remains stable until 65°C before a decrease in α-helical structure and until 80°C before an increase in β-sheet structure is observed. Adding detergent to either stoichiometric mixture (C, D) of BamD and Cj1289 results in stabilisation of the secondary structure composition throughout the temperature range relative to the NDB samples.

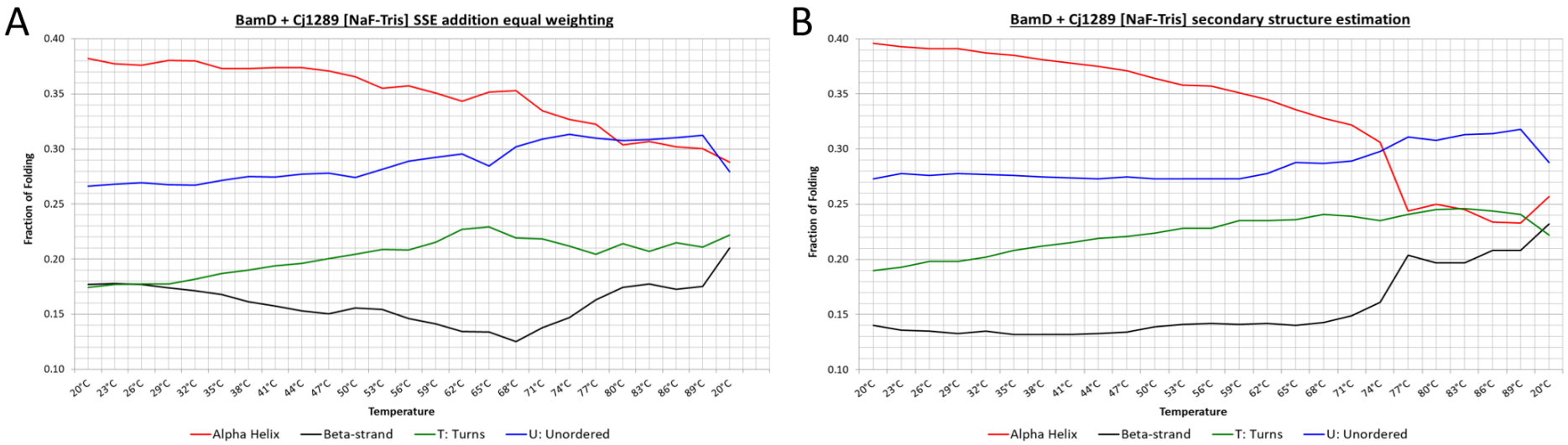


Figure 4.24: BamD and Cj1289 (1:1) mixture additive spectra versus experimental data. To predict the SSE of a non-interacting spectral addition of BamD and Cj1289, the individual spectra were added together in equal weighting (A). The experimental SSE for the 1:1 mixture of BamD and Cj1289 (B) is shown to aid comparison. The additive SSE (A) predicts a gradual decrease of α-helical content by 8% throughout and increase in unordered content throughout and β-sheet content after 68°C to compensate. The experimental SSE (B) shows α-helical structure decreasing gradually between 32°C-74°C before sharply undergoing an α-helical to β-sheet transition at 77°C, and reaching final composition values. The experimental SSE (B) does not match the BamD (NDB) SSE, appears more stable than the Cj1298 (NDB) SSE, and cannot be explained by the additive (non-interaction) SSE (A), and therefore, it is highly likely that BamD and Cj1289 interact to co-stabilise their structures.

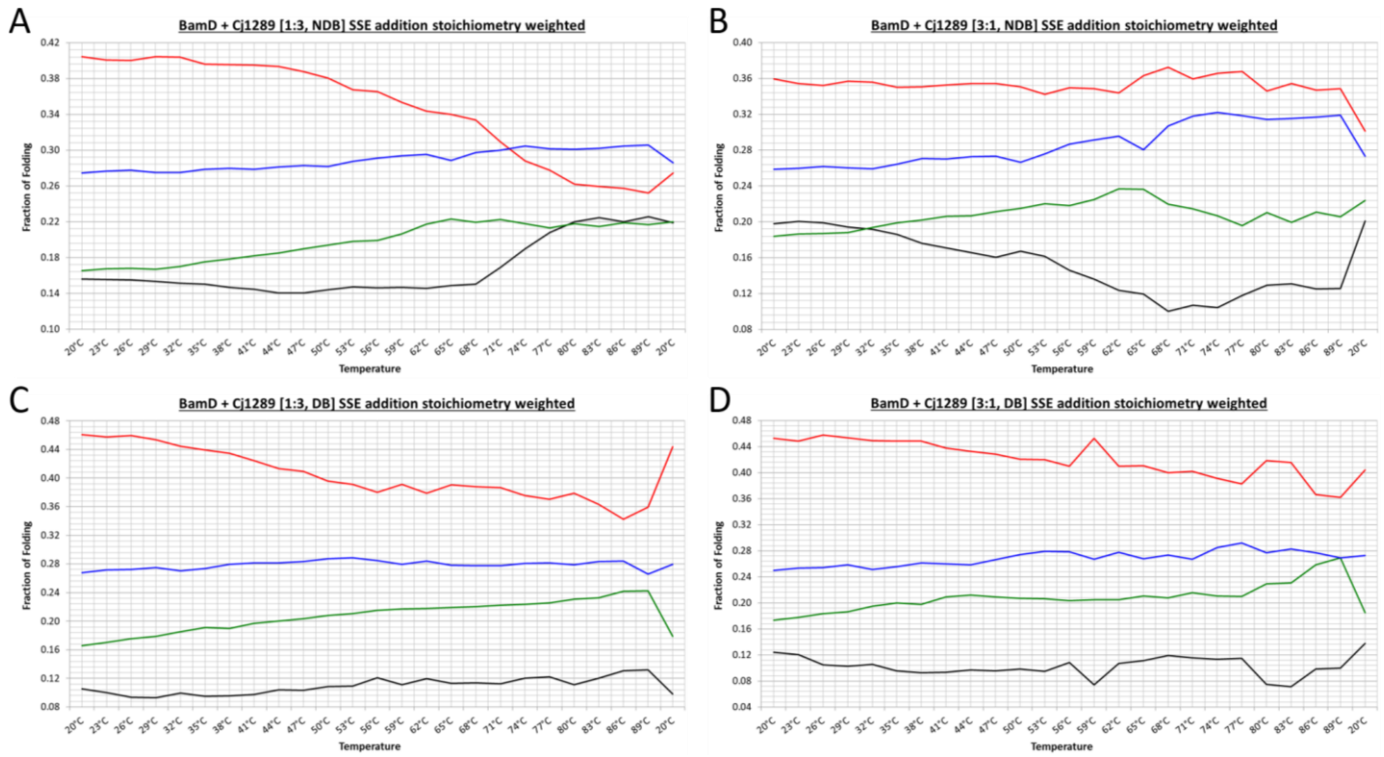


Figure 4.25: BamD and Cj1289 1:3 and 3:1 stoichiometry mixture additive spectra versus experimental data. To predict the SSE of a non-interacting spectral addition of BamD and Cj1289, the individual spectra were added together in either 1:3 (A, C) or 3:1 (B, D) stoichiometric ratios in non-detergent [NDB] (A, B) and SB12 [DB] (C, D) buffer backgrounds. Comparison of the predicted to the experimental SSEs reveals gross similarities in all stoichiometries, suggesting that all stoichiometries can be largely explained by the additive spectra of BamD and Cj1289 in the relevant background at that stoichiometry. The minor differences are likely due to the impact of the 1:1 BamD: Cj1289, and this suggests that the interaction is 1:1 and not affected by stoichiometry changes.

4.3.2.7 Secondary structure stability of BamA-BamD-Cj2189 1:1:1 stoichiometric mixture exceeds experimental temperature range

The thermal degradation profile of BamA: BamD: Cj1289 [1:1:1] mixture in detergent buffer (figure 4.26(a)) shows a small decrease in CD signal around 195 nm with progressive blue-shifting of the spectra over the temperature range of the experiment. Overall, the thermal degradation spectra present a very stable CD signal throughout the experimental temperature range, similar to the BamA-BamD [1:1] mixture. The post-experiment 20°C spectrum is similar to the 35°C spectrum with a minor blue shift. The relatively small change in CD signal suggests a strong co-stabilising interaction between the proteins, with the blue-shift of the spectra likely oweing to the presence of BamD.

The stability of the mixture to thermal degradation is highlighted by the BamA-BamD-Cj1289 mixture SSE (figure 4.26(b)), which shows less than a 4% change in α-helical and β-sheet secondary structure over the experimental temperature range. Unordered and turn structural content are similarly stable, with an approximately 2% increase as temperature increases after 35°C and 44°C, respectively. The largest change in structural content was an approximately 2 % α-helical to β-sheet transition observed from 62°C-89°C.

Contrary to the stability shown in the spectra and SSE, the Boltzmann two state equation estimation at 193 nm for the BamA:BamD:Cj1289 [1:1:1] mixture (figure 4.26(c)) shows an estimated T_m of 40.8°C. The post-experiment 20°C spectra point (red diamond) shows an approximate 1.1 mdeg change in signal from starting value, indicating a reasonably similar structure post-melt relative to the starting structure. This was likely as the overall structure did not change significantly over the temperature range of the thermal degardation experiment. The low T_m estimation was likely due to Cj1289, which has a low T_m in the DB background, rather than a significant change in the overall sample. The goodness of fit depreciated at discrete intervals (appendix 8.4.2). The predictive (figure 4.27(a)) and experimental SSE (figure 4.27(b)) for the 1:1:1 BamA:BamD:Cj1289 mixture show similar initial values and changes in structural composition for turn and unordered structures. However, the experimental α-helical and β-sheet initial compositions differ greatly and the α/β transition after 59°C is absent in the prediction.

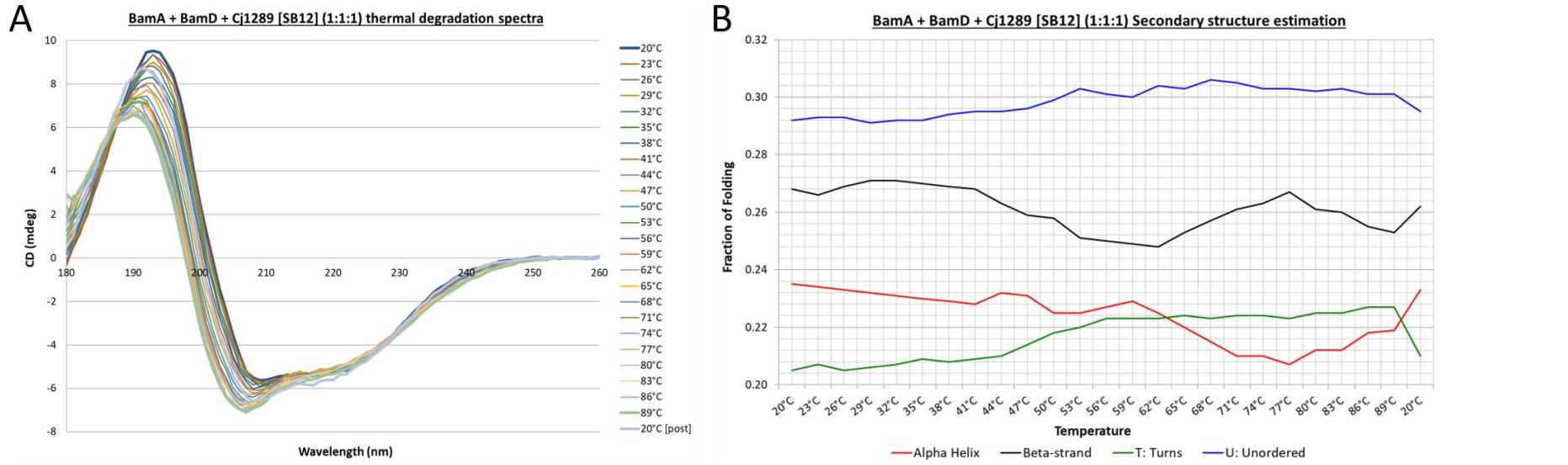


Figure 4.26: BamA: BamD: Cj1289 [1:1:1] mixture thermal degradation spectra, SSE and Boltzmann fitting. (A) Buffer subtracted thermal degradation profile of BamA: BamD: Cj1289 [1:1:1] mixture in detergent buffer, showing a very stable CD signal (< 4 mdeg change) throughout the experimental temperature range. (B) BamA-BamD-Cj1289 mixture SSE, showing less than 4% change in α-helical and β-sheet, and approximately 2% change in turn and unordered structure overall, with a α/β transition observed after 62°C. (C) Automated Boltzmann fitting of the BamA: BamD: Cj1298 mixture sample at 193 nm, showing an estimated T_m of 40.80°C. The post-experiment 20°C spectra point (red diamond) shows a roughly 1.1 mdeg change in signal from starting value. The sample is thermostable.

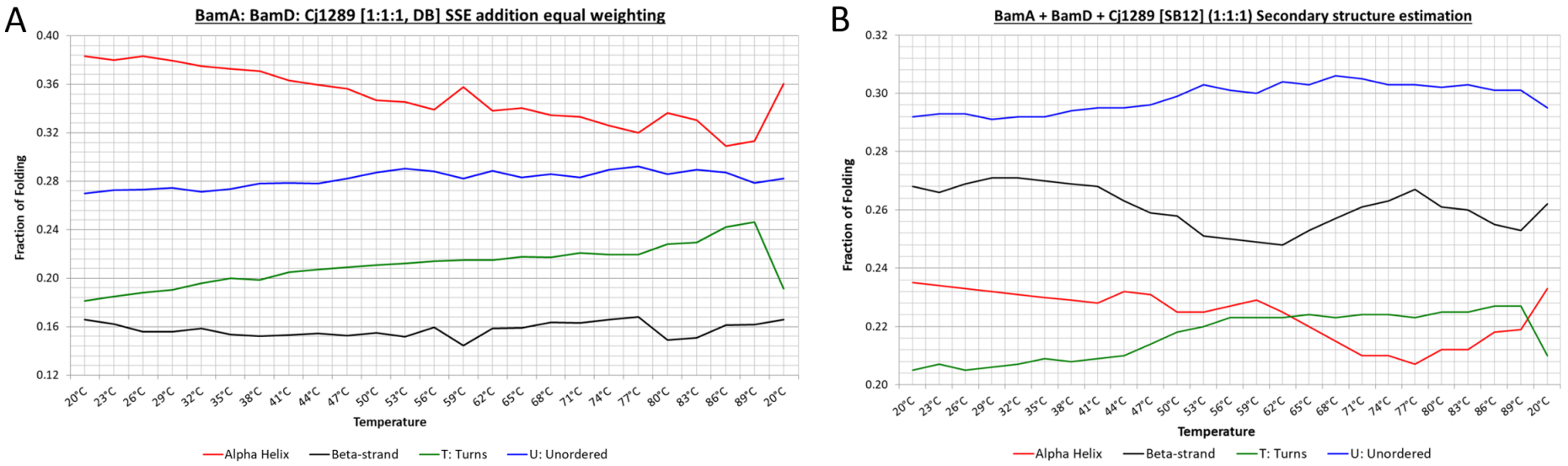


Figure 4.27: BamA:BamD:Cj1289 (1:1:1) mixture additive spectra versus experimental data. To predict the SSE of a non-interacting spectral addition of BamA, BamD and Cj1289, the individual spectra were added together in equal weighting (A). The actual experimental data for the mixing of BamA, BamD and Cj1289 [1:1:1] (B) is shown to aid comparison. The additive SSE (A) predicts a gradual decrease of α-helical content by 8% and turn structural content to increase by 6% over the experimental temperature range. Unordered structure increases by 2% overall and β-sheet structural content remains constant throughout the SSE. The experimental SSE (B) shows an initial gradual decrease in α-helical structure similar to the prediction, before a sharp α/β transition from 59°C to 77°C, which then gradually reverts until the end of the experimental temperature range. The turn and unordered structural content change is similar to the prediction, with the exception of the lack of a sharp increase in turn content between 86-89°C. The experimental SSE α/β transition and strikingly different α-helical and β-sheet content values at experiment start do not match predictions; therefore, it is likely that BamA, BamD and Cj1289 interact and enforce a structural change.

These factors suggest an interaction between BamA, BamD and Cj1289 that alters the structural content and melting behaviour to temperature relative to the individual protein thermal degradation experiments.

4.3.2.8 **Landau models and tau**

Landau models were produced from the SSE data from each of the experimental models as an alternative to the Boltzmann two state equation analysis to estimate T_m. The change in α-helical secondary structure content was examined for each sample, except BamD [NDB] where the change in β-sheet content was followed, as this secondary structure type showed a decrease over the experimental temperature range, which is (currently) required for Landau modelling calculations. The equation of state (equation 4.11) was determined for each sample and the spinodal representing the equation of state fitted by iterative adjustment of the T_m value until a good fit was obtained. The secondary structure estimations for α-helical and β-sheet secondary structures were plotted graphically with the calculated equation of state spinodal fitting the data for the first structural transition of each of the individual and mixed thermal degradation experiments. Landau models of BamA, BamD, and BamA: BamD mixture (figure 4.28), Cj1289, and Cj1289: BamA and Cj1289: BamD mixtures (figure 4.29), and BamA: BamD: Cj1289 mixture (figure 4.30) are shown. The points at which the spinodals meet the x-axis determine the T_+ temperature parameters. While it is possible to fit multiple spinodals to each plot (e.g. BamA, figure 4.29(a)), after the first structural transition the overall structure is not representative of the initial structure pre-melt, and so this has not been carried out (except BamA α-helical SSE as an example).

The T_m calculation shows that the BamA and BamD mixture have the highest T_m of all of the samples in the study, and the triple mixture of BamA, BamD, and Cj1289 the lowest calculated T_m (summarised in figure 4.30(c)). Overall, from highest to lowest T_m the analysis shows BamA-BamD > BamD [SB12] > BamD + Cj1289 > BamD (β) > Cj1289 > BamA > BamA + Cj1289 / Cj1289 [SB12] > BamA + BamD + Cj1289. In contrast, the T_+ calculations (summarised in figure 4.30(c)) shows a slightly different summary, with the analysis showing (highest to lowest T_+) BamA + BamD + Cj1289 > BamA + BamD > BamD [SB12] > BamD + Cj1289 > Cj1289 > Cj1289 [SB12] > BamA > BamD (beta) > BamA + Cj1289.

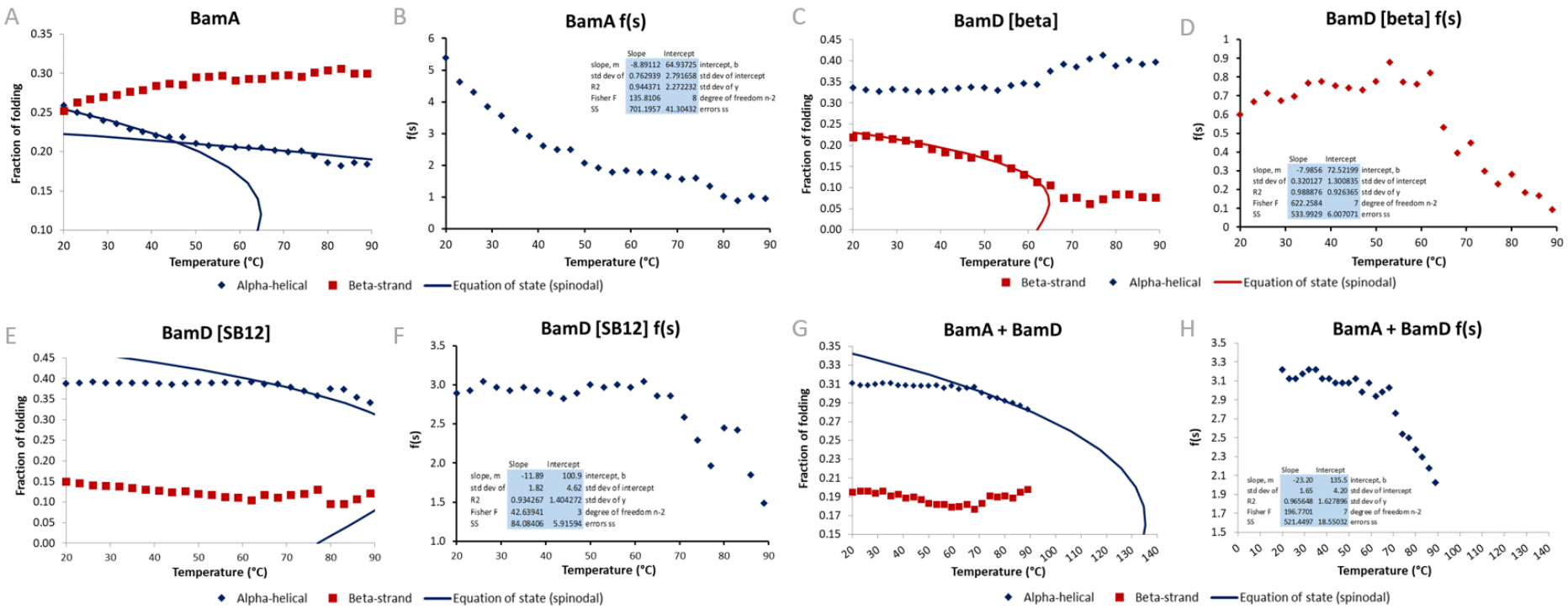


Figure 4.28: Landau models of BamA, BamD, and BamA:BamD mixture. Landau free energy expansion models for BamA, BamD and the BamA:BamD [1:1] mixture. Equations of state (spinodals) have been fitted by iterative adjustment to a decreasing secondary structure type to calculate the T_+ for (A) BamA, (C) BamD [NDB], (E) BamD [DB], and (G) BamA:BamD mixture [1:1], with the point at which the spinodal (blue line of fit; red line in BamD [beta]) meets the x-axis determining T_+. The accompanying f(s) *vs.* temperature first derivative plots for (B) BamA, (D) BamD [NDB], (F) BamD [DB], and (H) BamA:BamD show the relevant parameters and statistics for the Landau models are shown for each sample.

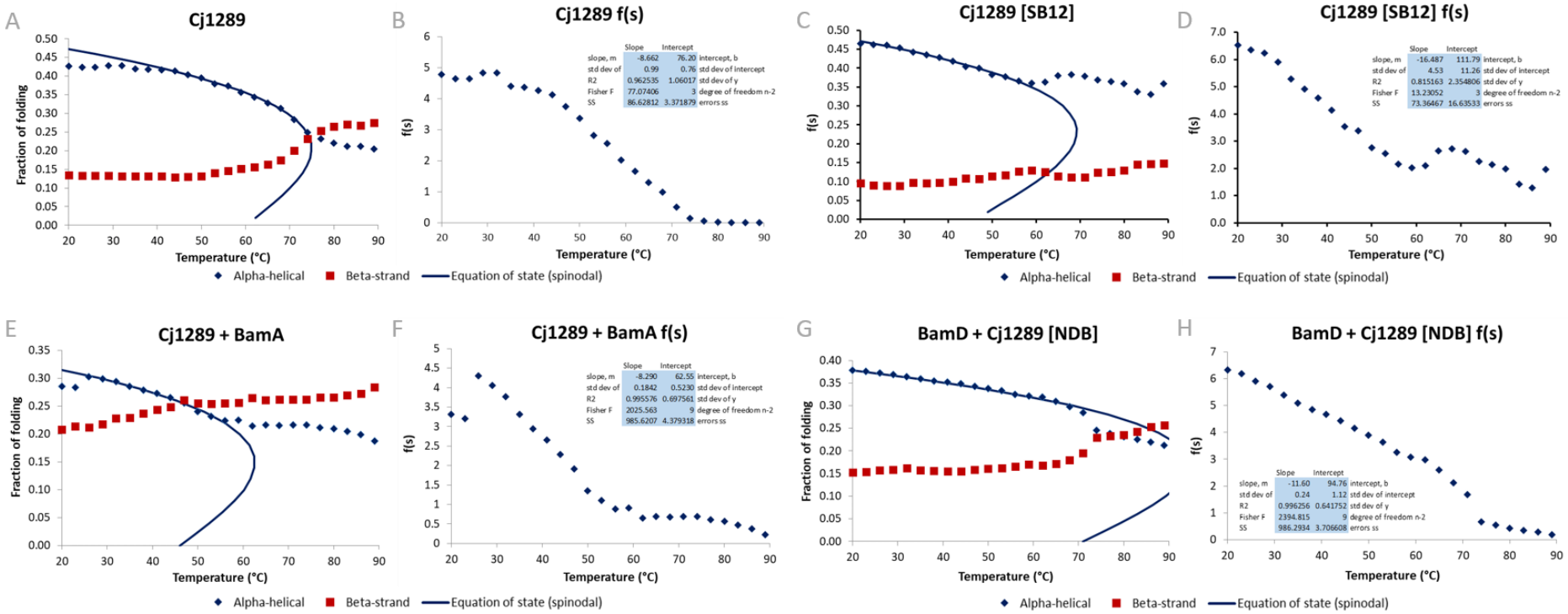
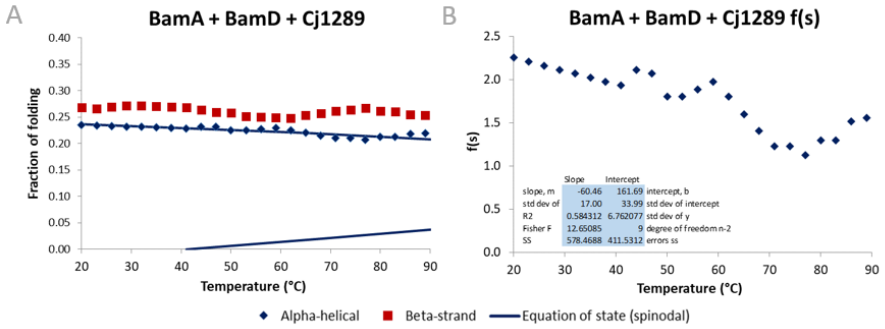


Figure 4.29: Landau models of Cj1289 and Cj1289: [protein] 1:1 mixtures. Landau free energy expansion models for Cj1289 [NDB, DB], and Cj1289-BamA [1:1], and Cj1289-BamD [1:1] mixtures. Equations of state (spinodals) have been fitted by iterative adjustment to α-helical secondary structure to calculate the T_+ for (A) Cj1289 [NDB], (C) Cj1289 [DB], and the (E) Cj1289-BamA [1:1, DB], and (G) Cj1289-BamD [1:1] mixtures, with the point at which the spinodal (blue line of fit) meets the x-axis determining T_+. The accompanying function of structure, f(s), *vs.* temperature first derivative plots for (B) Cj1289 [NDB], (D) Cj1289 [DB], (F) Cj1289-BamA [1:1, DB], and (H) Cj1289-BamD [1:1] mixtures show the relevant parameters and statistics for the Landau models are shown for each sample.

A



B

C

Sample	T_m	T_+	S_+
BamA [SB12]	47	64.86	0.121
BamD (β-sheet)	62	64.77	0.067
BamD [SB12]	77	100.82	0.208
Cj1289	60	74.92	0.205
Cj1298 [SB12]	46	72.09	0.208
BamA + BamD [SB12]	89	135.43	0.163
BamA + Cj1289 [SB12]	46	62.56	0.149
BamD + Cj1289	71	94.48	0.170
BamA + BamD + Cj1289 [SB12]	41	161.81	0.133

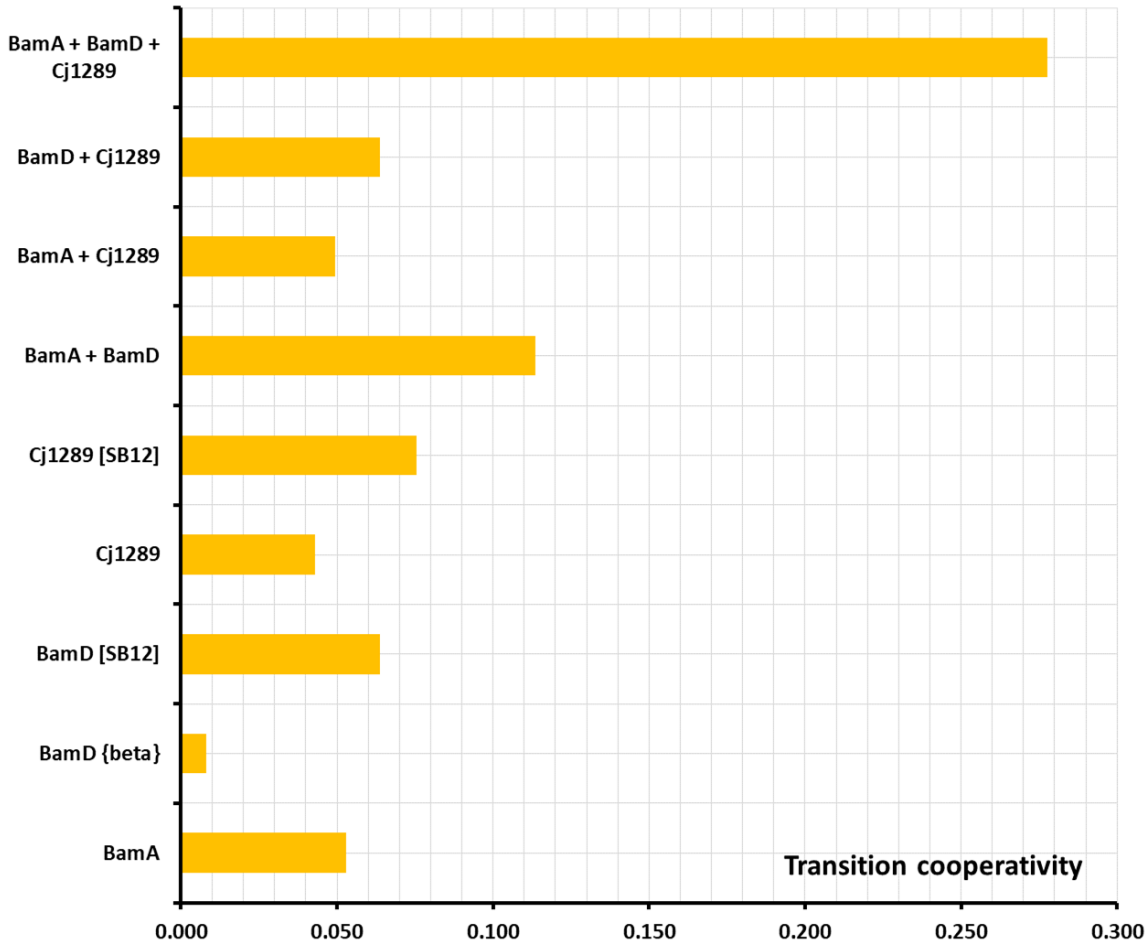
Figure 4.30: Landau models of the BamA:BamD:Cj1289 1:1:1 mixture and summary table of derived Landau parameters. Landau free energy expansion model for the BamA: BamD: Cj1289 1:1:1 mixture (A), the accompanying f(s) *vs.* temperature first derivative plot (B), and a table summarising derived paramaters for each analysis (C) are shown. Equations of state (spinodals) have been fitted by iterative adjustment of T_m to α-helical secondary structure to calculate the T_+ for (A) BamA-BamD-Cj1289 1:1:1 mixture [DB]. Usually the point at which the spinodal meets the x-axis determines T_+, however, as the mixture fails to show significant structural change over the experimental temperature range tested the T_+ parameter has been calculated from the spinodal peak (Dr B. Bonev, personal communication, 2020). (B) The accompanying f(s) *vs.* temperature first derivative plot for the BamA-BamD-Cj1289 mixture. Taking the slope of the first eight points and projecting forwards will generate the 161.81°C value returned for T_+, relevant statistics for the fitting are shown. (C) Reference table for all T_m, T_+, and S_+ values derived during the Landau modelling of each individual protein and protein:protein [1:1] mixture in this study.

A distinct advantage of Landau modelling was the derivation of tau (figure 4.31(a)), a measure of the transition cooperativities (as discussed in the introduction). The cooperativities of transition (figure 4.31(b), x-axis scaled back to fit data) show that the 'softest' or lowest energy transitions can be found in the individual protein thermal degradations. Where mixing of these proteins results in a co-stabilising interaction there is an increase in the negative cooperativity of transition (i.e. proteins co-stabilise to resist structural change or require higher energy input to unfold), and the 'hardest' or highest energy transition was observed in the BamA: BamD: Cj1289 [1:1:1] mixture sample which far exceeds the individual components. Our experimental results suggest that mixing two proteins hypothesised to interact together generally results in a higher (negative) transition cooperativity, as was the case with the BamA: BamD [1:1], BamD: Cj1289 [1:1], and BamA: BamD; Cj1289 [1:1:1] samples, which appear to be co-stabilising interactions. An exception to this observation was the BamA-Cj1289 sample, which shows a lower intermediate absolute (or positive) transition cooperativity when compared to the BamA and Cj1289 τ parameters, suggesting non-interaction, or a destabilising TPCA-type interaction.

A

Sample	T_m (K)	T_+ (K)	Tau ($1-T_m/T_+$)
BamA	320.150	338.010	0.053
BamD-beta	335.150	337.917	0.008
BamD-SB12	350.150	373.965	0.064
Cj1289	333.150	348.071	0.043
Cj1289 [SB12]	319.150	345.240	0.076
BamA+BamD	362.150	408.584	0.114
BamA + Cj1289	319.150	335.714	0.049
BamD + Cj1289	344.150	367.626	0.064
BamA + BamD + Cj1289	314.150	434.961	0.278

B



[Dr B Bonev, personal communications, 2020]

Figure 4.31: Calculation and plot of tau for all proteins and their mixtures. Landau model derived parameters T_m and T_+ were converted to Kelvin and used to derive τ, a dimensionless scaled value of measurement for the cooperativity of transition (A), which can take values of 0 (system undergoing critical unfolding) to 0.5 (system showing an ideally cooperative transition). The derived τ values have been plotted on this scale (x-axis cut back to fit data) for each protein individually and their combinations to show their relative cooperativities of transition (B).

4.4 **Discussion**

The predictions carried out *in silico* pre-experiment based off the PDB models did not match up well with the experimental data. This could be because the proteins in solution did not necessarily take up the same structures in solution as that observed in the crystal structures, which the homology modelling programs reference as a template to fold the primary structure information submitted, or unannealed homology models used. Alternatively, the crystal structure models might have slightly different torsion angles than those allowed by the SSA program algorithm (due to packing and process of crystallisation), causing the structures to be incorrectly classified. The MD annealed model SSEs were slightly better predictions, but overall, they also did not compare well against the experimental SSE, suggesting that the discrepancy may be caused by other factors (e.g. concentration, impurities in the sample, simplicity of MD model), or that the MD trajectories sampled were too small in duration to sample the configuration present in solution. Unfortunately, to our knowledge, no program exists in CD analysis currently that allow CD spectra to be matched against individual frames of MD trajectories to allow for the determination of the appropriate frame/conformational state to fit the data, so an arbitrary point/ frame after RMSD values stabilised, or after 100 ns, was chosen.

After further analysis, the differences in the initial BamD sample SSA and SSE could be explained by an incorrect concentration value, as reducing the concentration by one third allows the SSA to coincide with reasonably with the experimental SSE (large gain in α-helical and decrease in β-sheet SSA to near zero; appendix 8.4.3). However, after comparison of multiple methods (such as Bradford, BCA and micro-biuret methods) little difference in concentration measurement was found (data not shown), so the NanoDrop pre-experiment was likely correct. An alternative explanation, that takes this concentration value and the observed instability of BamD into account, was that a portion of the sample degraded either prior to the experiment or upon initial heating to 20°C, or was aggregated and sedimented in the quartz cuvette and was not sampled during the experiment, effectively changing the concentration. However, due to external influences (Covid restrictions), the protein state prior to the experiment could not be checked for partial aggregation, and

limited access was available on-site for the experiment to recover samples to confirm this.

While the pre-experiment predictions were not suitable, the CD experiments proceeded well, and all individual protein samples and mixtures showed good pre-experiment spectra with little noise and no breaks in the data (troughs and peaks) between 260 nm and 180 nm.

The blue shift in the CD spectra after 44°C in the BamD thermal degradation experiment (figure 4.8) was a unique phenomenon during these experiments that was not observed in either BamA or Cj1289. The blue shift can be explained by the loss of β-sheet structures, which have a negative absorption around 180 nm, increases in unfolded structures that have a positive absorption between 180 nm and 190 nm, and the formation of polyproline II type structures, which have negative absorption below 190 nm[178]. As the data suggests the formation of polyproline II type structures, which was found to be the dominant conformation in several short model peptides[179], this indicates a sub-population of the BamD was irreversibly degraded into short peptides. On post-experiment cooling, a sub-population of the BamD sample that was able to refold reforms some of the original β-sheet content, and so the spectra red shifts in response. In addition, the blue shift was a useful property in these experiments, as monitoring of the blue shift became an indicator of co-stabilising protein-protein interactions.

The addition of SB12, an amphiphilic detergent, had two major effects on the thermal degradations of BamD and Cj1289, firstly, to alter the Boltzmann and Landau estimations for T_m, and secondly, to improve the CD signal post-melt significantly for Cj1289.

For BamD, SB12 appears to improve stability of the CD spectra, with the CD signal at 190 nm decaying more slowly in the DB sample (figure 4.8(b)) as the temperature increases compared to the NDB sample (figure 4.8(a)), and the Boltzmann (figure 4.9) and Landau modelling analyses (figure 4.28) both showing a higher T_m in the presence of SB12. However, the Boltzmann analysis did not show any change in CD signal compared to the non-detergent sample on the post-experiment measurement, suggesting there was no effect of SB12 on the refolding on BamD, and blue shifting of spectra still occurs, albeit to a lesser degree. The influence of SB12 on BamD was

understandable as detergents can be considered good substitutes for lipid[180], and as a lipoprotein, BamD would be expected to interact with lipid. These interactions, whether Van der Waals, hydrogen bonding or charge-charge interactions would result in the stabilisation of BamD, improving thermal stability and resistance to thermally induced irreversible degradation.

For Cj1289, SB12 similarly reduces the CD signal loss at each temperature point (figure 4.11), but unlike BamD, refolding of Cj1289 was vastly improved and T_m (both Boltzmann and Landau) estimations are greatly reduced in the presence of SB12. As SB12 is zwitterionic, it may play a role in refolding through interaction with charged and non-polar regions of proteins, thereby limiting cross interactions, inappropriate folding and aggregation. These features may enable better refolding of Cj1289, but conversely, if Cj1289 is a dimer in solution, as the 3RGC crystal structure suggests, then molecules of SB12 could interfere with the binding of two Cj1289 monomers or cause dissociation of a Cj1289 dimer. Interference with binding of between Cj1289 monomers and/ or dissociation of a Cj1289 dimer would explain the large reduction in the T_m estimation as a monomer would likely have a lower thermal stability than a (likely co-stabilising) dimer of Cj1289.

A final observation on the influence of the addition of SB12 was the change of the secondary structure assignment at experiment start (20°C), with an increase in alpha helical content by 10 % (BamD)/ 4 % (Cj1289) offset by a decrease in estimated β-sheet content, improving the coalescence between predicted SSAs and observed SSEs (refer tables 4.1 and 4.2).

Summarising the individual protein thermal degradation experiments, BamA was reasonably thermostable over the experimental temperature range, BamD had increased thermal lability without detergent, and Cj1289 showed a gradual melting phenomena which was stabilised above 65°C in the presence of SB12. The individual protein melts show the following T_m trend according to Boltzmann two state equation analysis, with Cj1289 [NDB] > BamA > BamD [DB] > BamD [NDB] > Cj1289 [DB].

Once the individual protein thermal degradation experiments had concluded, several mixtures of the individual proteins were thermally degraded to determine if and what type of interaction occurs between the proteins in the mixture.

The results indicate that BamA and BamD show a co-stabilising interaction, BamA and Cj1289 do not interact regardless of stoichiometry, and BamD and Cj1289 show a co-stabilising interaction, in likely a 1:1 stoichiometry. These results are best illustrated by the side-by-side comparison of the CD spectra and SSE for BamA (appendix 8.4.4, figure 8.17 and 8.18, respectively), and BamD (appendix 8.4.4, figure 8.19 and 8.20, respectively), but for the Cj1289 spectra (figure 4.11) and SSE (figure 4.13), the original figures will be referred to.

BamA and BamD show a co-stabilising interaction. The BamA mixture CD spectra shows that the BamA individual thermal melt exhibits absorption flattening around 193 nm with increasing temperature (figure 8.17(a)), which was stabilised in the BamA: BamD mixture (figure 8.17(b)) but features blue shifting of spectra (similar to BamD individually). This was further illustrated through comparison of the SSEs, where BamA (figure 8.18(a)) shows a gradual α-helical to β-sheet transition over the temperature range of the experiment, which was largely stabilised when in mixture with BamD (figure 8.18(b)). The Boltzmann two state equation shows a 2.6°C increase in T_m (46.81°C *vs.* 44.2°C) for this stabilisation, which does not match the greatly increased SSE stability, however Landau modelling determines an increase in T_m of 44.8°C (89°C *vs.* 44.2°C) which better encapsulates the increased stabilisation. This interaction was expected from the literature from the structures available in *E. coli* on the BAM complex[7–9, 44]

BamA does not exhibit a co-stabilising interaction with Cj1289 in mixture. Comparison of the CD spectra of BamA (figure 8.17(a)) to the BamA: Cj1289 mixture (figure 8.17(c)) shows an increase in the absorption flattening around 193 nm in the thermal degradation of the sample mixture relative to BamA individually, indicating no co-stabilisation of structures. The increase in absorption flattening resembles the individual Cj1289 thermal degradation experiment spectra in SB12 (figure 4.11(b)), and comparison of the SSEs of BamA (figure 8.17(a)), Cj1289 in SB12 (figure 4.13(b)), and the BamA: Cj1289 mixture (figure 8.18(c)), show an intermediate profile between the individual SSE changes. This corroborated by the Boltzmann estimation (figure 4.17) and Landau modelling summation (figure 4.30(c)) which both show a T_m of 46.5°C ± 0.8°C, which was nearly identical to the individual melting temperatures of BamA (47°C, Boltzmann and Landau) and Cj1289 in SB12 (47°C, Landau). Altogether, this strongly indicates no interaction

between BamA and Cj1289. This does not coincide with the literature as a recent structure of *E. coli* BAM and SurA demonstrate an interaction[181], but in the context of the experiment, where BamA was reasonably thermostable across the experimental temperature range, it makes sense, as without a degree of unfolding in the BamA secondary structure Cj1289 will likely not exhibit its chaperone function.

BamD and Cj1289 show a co-stabilising interaction in mixture. The limited blue shift of BamD in the BamD-Cj1289 CD spectra relative to the individual BamD CD spectra (figure 8.19(a) and (c), respectively), the increased stability in the associated SSEs (figure 8.20(a) and (c), respectively), strongly indicate an interaction which stabilises the structural change of BamD during thermal degradation. The Landau models corroborate this interaction, showing a 10°C in the T_m of the mixture relative to the individual components; however, the Boltzmann analysis gives an intermediate value relative to the individual components. Given the increased stability of the mixture SSE (figure 8.20(c)) relative to the individual components (BamD, figure 8.19(a); Cj1289, figure 4.11(a)) the Boltzmann estimation was likely incorrectly underestimating the T_m for the BamD-Cj1289 mixture. Taken together this indicates a co-stabilising interaction between BamD and Cj1289. This was a reasonable result given the instability observed in the BamD sample, and Cj1289 likely acts as a holdase stabilising BamD as it begins the β-sheet to α-helix transition observed around 50°C. Kale *et al.* (2011)[108] stated the Cj1289 was structurally the most similar protein to SurA in *C. jejuni* and SurA is known for its holdase activity[182], so this interaction is in agreement with the literature and a holdase functionality of Cj1289 is likely cause of the structural stabilisation observed in the BamD-Cj1289 mixture.

Whether this interaction should occur is a different topic, as discussed in the previous chapter the lack of a lipobox motif in the construct design may be resulting in the instability of BamD due to improper targeting and maturation of the lipoprotein. The data indicates this interaction occurs, however, as this was unexpected pre-experiment, it is difficult to quantify without an experiment re-run with an improved BamD construct featuring a lipobox motif if this would occur with a mature and correctly targeted BamD.

The BamA: BamD: Cj1289 [1:1:1] mixture shows a highly co-stabilising interaction. This interaction, is likely composed of two interactions, that of BamA and BamD, and BamD and Cj1289, given the prior findings in this study. The thermal degradation profile appears similar to the BamA-BamD degradation profile (figure 4.14(a)) with an increased absorption flattening of the CD signal across the experimental temperature range. This increased absorption flattening was likely caused by addition of Cj1289, as among the individual thermal denaturation spectra Cj1289 had the largest change in CD signal as the temperature increased (figure 4.11(b)). The Boltzmann analysis predicts a low T_m for the mixture (figure 4.26(c)), which coincided well with the T_m from the Landau analysis of the α-helical secondary structure component (figure 4.30). However, the SSE shows a relatively stable secondary structure composition throughout the temperature range of the experiment (fluctuating by 3% at most), suggesting this was an underestimation, which was corroborated by the Landau T_+ and transition cooperativity values, or that the α-helical component melts separately to the rest of the structure. The experimental temperature range was likely insufficient to accurately generate the correct T_m value. Biologically, this could make sense if Cj1289, in its holdase functionality, was delivering a partially unfolded BamD to BamA for binding, which may be a reflection of the α-helical to β-sheet transition after 62°C that was more pronounced in the triple mixture than the BamA-BamD mixture.

The goodness of fit data shows an overall trend where the fit begins to diminish after approximately 50°C. In the majority of thermal degradation experiments this coincides with when the major transitions in secondary structure occur, and the increased variability is likely due to sub-populations of the sample having undergone the change and another sub-population within the sample still undergoing the change. This is the most likely scenario as there are several values after 50°C that show very small standard deviations, and the more stable samples tend to have better goodness of fits throughout the experimental temperature range (appendix 8.4.2).

The Boltzmann two state equation estimated melting temperatures showed a very different picture to the Landau model derived melting temperatures. Pre-experiment, based on knowledge from the literature[168] and the concept that interaction between proteins would result in stabilisation, it was thought that the T_m estimations would follow the pattern β-barrel + [co-stabilising protein] > β-barrel > [protein] + [co-

stabilising protein] > lipoprotein > TPCA> periplasmic protein. However, in the experiment, the highest individual protein estimated T_m was for Cj1289 (60.04°C), a periplasmic protein, and BamA (β-barrel) and proteins that were thought to interact in a stabilising manner were estimated to have much lower T_m (<50°C). In contrast, the Landau model predicts the BamA-BamD mixture to be have the highest T_m, which was likely higher than the temperature range of the experiment. The triple mixture has a lower predicted T_m than this mixture; however, this was very likely underestimated looking at the very limited change in secondary structure. If you consider the T_+ estimation, the triple mixture would likely have the highest T_m, and the Landau models would predict A + C + D (T_+) > A + D > D [SB12] > D + C [NaF-Tris] > C [NaF-Tris] > A > C [SB12] / A + C [SB12]. The prediction from the Landau model was in good agreement with pre-experiment expectations, with the highest T_m belonging to the mixtures that interact. It should be noted that α-helical content was used to generate the T_m estimations and so BamA has a lower T_m than would be expected if the β-sheet content were sampled. The BamA and Cj1289 mixture being the lowest in the T+ estimation points toward TPCA type phenomena, but at approximately 2°C lower than BamA it had a minimal effect.

Unlike the disagreement in overall T_m estimations, both the Boltzmann equations and Landau models agree that the T_m for BamD increases and the T_m for Cj1289 decreases in the presence of SB12. This was logical, as the detergent likely serves as a lipid substitute and interaction surface for BamD, but also conversely as an interfering particle to Cj1289 dimerisation.

The biggest concern in using the T_m estimations from the Boltzmann two-state equation was that the T_m can change quite dramatically depending on the wavelength selected. A good illustration of changing T_m when using Boltzmann was a comparison of measurements for α-helical structure at 192 nm and 209 nm (appendix 8.4.5), where an 8°C change in T_m was found.

The standard methodology used in picking the wavelength to analyse was to look for the wavelengths with the biggest change in CD signal over the experimental temperature range, and this was consistently applied across all samples. This can then be manually fine-tuned to the wavelength with the best correlation of data. This method of picking the wavelength to monitor may or may not retrieve the highest T_m

value available, and may result in different wavelengths being selected across samples. The issues with this are that different wavelengths sample different secondary structures or combinations of secondary structures, so variation in the wavelength used between samples means data should not be directly compared. Secondly, the Boltzmann estimation was based on the overall structural change and considers the final product, which causes a problem when multiple transitions occur as troughs or peaks in the correlation can skew the auto-fitting. This may mean that the Boltzmann equation T_m estimations likely work better for single domain proteins and proteins that exhibit high cooperativity of transition, but might be inaccurate in T_m prediction for more complex multidomain proteins or those with several secondary structure types, each of which might individually go through distinct transitions during thermal degradation. While some of this can be mitigated using multiple manual fittings to resolve each transition, these can be subjective and may vary from user to user as to how much variation in signal indicates a transition and where the fitting should start and end. The last problem is that even if all the transitions are correctly fitted, unless you have ideal proteins consisting completely of one secondary structure type, the contribution of individual secondary structures or what this transition represents in not known or immediately obvious.

The Landau models have numerous advantages over Boltzmann fittings. Firstly, the different secondary structures can all be followed individually, and T_m can be calculated for the transition of each secondary structure type. Secondly, the method is wavelength independent, removing all influence of wavelength choice on the result. Thirdly, values for T_+ and τ can be derived to give information on the cooperativity of melting. Lastly, these values can be directly compared, particularly the transition cooperativity, which is placed on a dimensionless scale.

However, Landau models are not without their issues either. To obtain a T_m estimation there needs to be a significant change in a secondary structure type, and (currently) it must be a decrease in that structure type. In addition, the methodology is not currently widely available or accessible. However, these limitations are in the process of being addressed [Dr B. Bonev, personal communication, 2021].

Overall, the results suggest that the Landau models give a more consistent prediction on T_m (as wavelength choice is not a factor), and in this data set, which contains

periplasmic proteins, lipoproteins and OMPs, the output better fits the theoretical expectations. The data gathered here strongly advocates for the use of Landau models as either a secondary verification of the current established Boltzmann two state equation methodology, and as the case may be in the future, a replacement for it entirely.

4.5 **Summary**

In this chapter, the CD of the proteins in this study and of the proteins in several mixtures were assessed during thermal degradation experiments between 20°C and 89°C at 3°C intervals.

The samples in the Chirascan experiment all showed good 1-n spectral scans, indicating good sample quality. However, the protein SSEs did not match up accurately against the pre-experiment *in silico* estimations. In the case of BamD, the concentration, partial aggregation or a sub-population of degraded subunits may have led to the gross inaccuracy in the estimation, and this may be a reflection of the instability of BamD in the current construct design, which lacks a lipobox motif.

In the thermal degradation experiments, structural changes were mapped against temperature for individual proteins and mixtures of proteins, and melting temperatures and cooperativity parameters were derived using either Boltzmann two state equation or Landau modelling analyses. The BamA-BamD, BamD-Cj1289, and BamA-BamD-Cj1289 mixtures all resulted in co-stabilising interactions displaying increased thermal stability of secondary structures and reduced CD signal decay over the experimental temperature range relative to the individual components. The co-stabilising interaction was corroborated by their increased T_m and τ-values, which indicate an increased degree of cooperativity in the thermal denaturation experiments relative to the individual components. Conversely, the data for the BamA and Cj1289 mixture indicates a lack of interaction, likely as BamA was relatively thermostable over the temperature range of the experiments.

The addition of SB12 improves refolding in Cj1289, but has a negative effect on T_m, which may make sense if Cj1289 was a dimer in solution as the detergent molecule may prevent or interfere with cross-interactions required for binding. The opposite effect was observed from the addition of SB12 on BamD, with no impact observed

on refolding, but a significant increase in T_m and secondary structure stability as would be expected for a lipoprotein interacting with a lipid substitute.

The Boltzmann two state equation prediction of T_m was heavily influenced by the wavelength sampled and the T_m estimations are erratic, with periplasmic Cj1289 having higher T_m than two interacting proteins and β-barrel proteins individually and in mixture, and so, do not match expectations gleaned from the knowledge in the literature. The Landau free energy expansion models estimate T_m based on the rate of change in a particular secondary structure and several derived parameters. The Landau models show a trend in T_m more consistent with expectations made from knowledge in the literature, and as there is no wavelength dependency, they are more reliable predictors of T_m than the current established Boltzmann two state equation.

Overall, this aspect of Covid mitigation went well and good insight was gained into the *C. jejuni* BAM-mediated outer membrane transport pathway.

4.6 **Limitations**

For the thermal degradation experiments involving mixtures of proteins, particularly in the case of the Bam-BamD-Cj1298 and BamA-BamD samples, the temperature range used in the experiment was insufficient to cause the proteins to undergo full thermal degradation. This limits the Landau model analysis, as their parameter derivations are dependent on samples showing a decrease in a specific secondary structure type in the experiment temperature range, resulting in potential inaccurate derivation of the equation of state and transition cooperativities.

The refolding of the proteins was not followed when the temperature of the sample was reduced from 89°C back to 20°C, so it cannot be determined if the unfolded to folded transition was similar to the folded to unfolded transition.

5 Small angle X-ray scattering analysis of protein-protein interaction in the BAM mediated outer membrane transport pathway

5.1 Chapter aims

In this chapter, we sought to further characterise the proteins in this study using SAXS, and validate the interaction profiles determined by CD.

In this set of experiments, we aim to use SAXS to:

1) Obtain an experimental SAXS curve for each protein in this study, and use this to:
    a. Infer the flexibility, state of folding, and overall shape of each protein
    b. Cross-validate the experimental data to simulated SAXS curves generated from available structural models, homology models and annealed MD simulation models for each protein
2) Obtain experimental SAXS curves for mixtures of proteins and use these profiles to infer and cross-validate interactions between proteins in this study through comparison to individual protein SAXS experimental data
3) Build bead models of proteins or protein mixtures (where applicable) and relate these models to the available structural models

5.2 Introduction

The proteins in this study are either β-barrel integral membrane proteins (BamA, LptD), lipoproteins (BamD), or oligomeric flexible chaperones (Cj1289 [monomer or dimer], DegP [multiple oligomeric states]), and all proteins are expected to be involved in protein-protein interactions, molecule or protein transport and/ or folding and assembly. All of these types of proteins present challenges that make structural characterisation via crystallisation challenging, whether due to presence of detergents, lipids, their flexibility, ability to self-assemble into multiple oligomeric forms, their low yields, or a combination of these factors. An alternative method of characterising protein structures is SAXS, which requires less sample than X-ray crystallography, and unlike crystallography is a solution technique, allowing for measurements in near physiological conditions, albeit at lower resolution. Samples are exposed to a highly collimated X-ray beam, and the coherent scattering is detected at low angle (0.5°-10°) to the incident beam. The SAXS profile obtained can provide information on the overall shape, size, state of folding, and flexibility of proteins, information on organisation, stoichiometries and interaction in mixtures of proteins, and how these properties change as a function of time or protein concentration [183–185].

However, being a solution technique, the molecules that constitute the sample are tumbling in solution, resulting in the elastic scattering of the molecule of interest and bulk solvent being sampled from random sample orientations, leading to an isotropic (or averaged) profile of the molecule of interest. As a result of this averaging, SAXS is inherently a contrast method where the scattering signal is generated through the difference in the average electron density between the molecule of interest and the background bulk solvent, with greater differences in electron density leading to better signal-to-noise at collection. This can be given by the equation:

$$\Delta\rho(r) = \rho(r) - \rho(s) \qquad \text{(Equation 5.1)}$$

Where Δρ(r) is the averaged electron density, ρ(r) is the electron density of the molecule of interest and ρ(s) is the electron density of the bulk solvent.

As proteins have an average electron density of ~0.44 $e^-/Å^3$ and water has an average electron density of ~0.33 $e/Å^3$, the resulting contrast is small in ideal

conditions, so a high protein concentration (1-10 mg/ml) is typically required to gain a good signal to noise ratio[185–188]. As the proteins in our study are involved in binding or protein-protein interactions, given the potential of aggregates (which scatter more strongly than smaller particles) at higher concentrations, and the difficulty in obtaining these high concentrations with these proteins, we used protein concentrations at the lower end of this range.

In this study, SAXS is used as both a standalone and a complementary structural method to CD for the characterisation of the proteins in this study and interactions in the *C. jejuni* BAM mediated transport pathway. SAXS allows these interactions to be probed in a solution environment in near physiological conditions and can typically provide low-resolution information on particles sized between 10 Å-1000 Å[184]. As the SAXS curves can be recreated from PDB files, the existing structural models in the literature, and homology and MD simulations annealed models generated in this study can be used to validate the SAXS experimental data obtained.

5.2.1 Biological background

5.2.1.1 Interactions expected from the *E. coli* BAM mediated transport pathway

BamA and BamD are known to interact in *E. coli* [5, 8, 10, 37, 44], and LptD is thought to interact with BamA[189] for its assembly. The chaperone SurA acts as an intermediary by receiving proteins in the periplasm and transferring the proteins to the BAM complex members[26, 165, 166]. DegP is known to interact with misfolded and unfolded proteins as part of its protease function[90, 92], however, under the correct conditions (i.e. low temperature) can serve as an auxiliary chaperone[25, 190]. BamD is known to have some interactions with OMPs through cross-linking studies, bioinformatics and this function is inferred due to its tetratricopeptide repeat domains[30, 166, 191], but it is unknown whether there is interaction with chaperones SurA and DegP, or with LptD specifically.

5.2.1.2 Interactions expected from the CD data

Our CD data suggests that co-stabilising interactions occur between BamA and BamD, and BamD and Cj1289, but a destabilising TPCA type interaction occurs between Cj1289 and BamA. The BamA-BamD interaction matches that of the *E. coli* homologues, but the destabilising interaction between BamA and Cj1289 does not mimic the expected *E. coli* BamA-SurA interaction profile. The interaction of BamD with SurA is not well documented in *E. coli*, but our CD data suggests interaction between BamD and Cj1289.

Unfortunately, no samples could be run for DegP or LptD individually or in mixture to test potential interactions with CD due to external factors, so it is unknown what interactions, if any, these have with the other proteins in this study. Based on the available information from experiments with *E. coli* homologues, LptD is expected to interact with BamA, and DegP may interact with LptD, BamA or BamD as part of its chaperone function in the BAM mediated transport pathway. No interaction is documented between LptD and BamD or SurA in *E. coli*, but as *C. jejuni* lacks Skp and Cj1289 acts as the main chaperone, it would be intuitive to expect LptD and Cj1289 to interact. Another possibility is that LptD directly interacts with BamA, not requiring the chaperones, or that LptD folds independently of the BAM mediated OM transport system in *C. jejuni*.

5.2.2 The SAXS curve

The raw scattering data can be represented as log10 and Power-law plots, which emphasise different q-ranges of the signal, and transformed into a number of plots for analysis. The standard plots used in this analysis are the Guinier, Kratky, flexibility, and pair-distance distribution function plots. In addition, these raw scattering curves can be utilised to assess complex formation.

5.2.2.1 Guinier analysis

The radius of gyration, R_g, is the average electron density weighted by the squared distance of the scatters from the centre of mass of the object. Analysis of the Guinier region, the low-q region, gives information on the overall size of the molecule of interest through derivation of the intensity at zero scattering vector, I(0), and the volume of the particle. R_g is smaller and more accurate for compact globular

particles compared to elongated, extended or flexible particles[185, 188]. The Guinier plot, ln(I(q)) *vs.* q^2, of the low-q region should show a linear fitting with minimal trimming from low-q, with slope given by $-\frac{{R_g}^2}{3}$ and $q \cdot R_g \leq 1.3$ for the fitting[192]. The overall equation for Guinier analysis is given by:

$$\lim_{q \to 0} I(q) = I(0) \cdot exp\left(-\frac{1}{3} \cdot q^2 \cdot {R_g}^2\right) \qquad \text{(Equation 5.2)}$$

Where the scattering vector, q, is given by:

$$q = \frac{4\pi sin\theta}{\lambda} \qquad \text{(Equation 5.3)}$$

Where 2θ is the scattering angle, λ is the wavelength of radiation, and q is distance measured in either $Å^{-1}$ or nm^{-1}.

The Guinier approximation is sensitive to aggregation and inter-particle interactions (which manifest as upturns at low-q), concentration dependence (downturn at low-q), radiation damage (noise) and poor buffer subtraction, but is independent of the shape of the molecule of interest[185, 188].

5.2.2.2 **Kratky analysis**

The scattering curve can be transformed and plotted as $q^2 \cdot I(q)$ *vs.* q, a Kratky plot, to give information on the flexibility and foldedness of proteins.

The Kratky plot can be normalised by multiplying q by the particle's R_g and multiplying I(q) by $(q \cdot R_g)^2$ and dividing I(q) by I(0) to generate a dimensionless Kratky plot ($(q \cdot R_g)^2 \cdot I(q)/I(0)$ *vs.* $q \cdot R_g$). This plot normalises scattering profiles by mass and concentration, providing semi-quantitative analysis of flexibility and state of folding[193, 194]. For compact globular proteins there will be a peak at 1.732 with a magnitude of $3e^{-1}$ (1.104) with a Gaussian distribution.

In both plots, a high-q region that tends towards zero on the y-axis after the hyperbolic rise indicates the sample is folded, if the high-q region levels off before this the sample is folded but has flexible elements, and if the scattering curve rises and levels off in the high-q the sample in unfolded or flexible[193]. Where samples show an uptick in the low-q region on the Guinier approximation, indicative of aggregation or inter-particle effects, normalised Kratky, which is dependent on a Guinier analysis of the low-q region, will not be carried out.

5.2.2.3 **Flexibility analysis**

A method of understanding the flexibility of a sample is to transform the scattering data into a series of plots, the Kratky-Debye plot ($q^2 \cdot I(q)$ *vs.* q^2), the SIBYLS plot ($q^3 \cdot I(q)$ *vs.* q^3) and the Porod-Debye plot ($q^4 \cdot I(q)$ *vs.* q^4)[195, 196]. The q-range is sampled towards low-q until one of the plots level off. A plateau in the Kratky-Debye plot indicates a disordered or flexible particle; a plateau in the Porod-Debye plot indicates a compact or ordered particle, and a plateau in the SIBYLS plot indicates a particle with a mixture of ordered and flexible regions or domains. A further parameter that can be derived from this analysis is q_{max}, which defines the upper q end-point of the Guinier plot.

$$q_{max} = \frac{M}{R_g} \qquad \text{(Equation 5.4)}$$

Where M is typically 1.3 for globular proteins.

5.2.2.4 **Pair density distribution function**

The scattering intensity, I(q), can be indirect Fourier transformed into a pair-distance distribution function, P(r), which is a histogram of all electron pair distances, r, between points within the molecule of interest. The P(r) is the real space representation of the distances in the molecule being investigated, and uses all of the experimental data range. By integrating the function over all values of r, the R_g and I(0) can be determined. Good agreement between Guinier approximated and P(r) derived R_g and I(0) is indicative of good quality data. The P(r) function is given by:

$$P(r) = \frac{r}{2\pi^2} \int_0^\infty I(q) q sin(qr) dq \qquad \text{(Equation 5.5)}$$

The parameters are determined by ScÅtter via a fitting routine regularised through maximisation of the smoothness of the P(r) distribution in this study[187, 197–199]. In practice, using ScÅtter, some of the noisy high-q data is trimmed and the maximal particle diameter (D_{max}) representing the largest electron pair-distance in the molecule, is reassigned iteratively until the smoothness of the distribution is maximised.

5.2.2.5 **Complexation analysis**

The complex assessment function in ScÅtter is used to test whether a complex is formed, and works by carrying out a linear addition of the individual spectra at different concentrations to make an assessment on whether a complex has been formed in the mixed sample. This is reflected in both the overlay of the scattering curves and in the residuals throughout the q-range sampled, with near perfect fits to the scattering curve and linear residuals (or one with an even distribution of data points) throughout indicating that no interaction occurs, and an imperfect SAXS curve fitting and/or skewed residual indicating formation of a complex.

5.3 Results

5.3.1 BamD

SAXS analysis of the BamD sample shows the signal only varies by intensity with increasing concentration indicating the overall shape was independent of concentration, and good signal quality extends to high and low-q in the 1 mg/ml sample (figures 5.1(a) and (b), respectively). The flexibility assessment (figure 5.1(c)) indicates the overall structure is mainly rigid and compact, but that there may be an element of flexibility to consider (Porod-Debye followed closely by SIBYLS). The derived q_{max} from the flexibility plots was 0.1360. The Kratky plot (figure 5.1(d)) has a single hyperbolic rise, indicating a single domain protein, but the data plot does not tend to the x-axis, corroborating that the structure has some degree of flexibility. There was some aggregation apparent in the Guinier region, but the trimmed profile fits the curve well in the range of $\sim 0.2 < (q \cdot R_g)^2 < 1.6$ (figure 5.1(e)). Further analysis was restricted due to the trimming required and the relatively small M/W of BamD.

5.3.2 Cj1289

SAXS analysis of the Cj1289 sample the overall shape was independent of concentration, and good signal quality extends to high-q and low-q at concentrations above 0.5 mg/ml (figures 5.2(a) and (b), respectively). The flexibility assessment (figure 5.2(c)) indicates the overall structure is mainly rigid and compact, but has flexible elements to consider. The derived q_{max} from the flexibility plots was 0.1294. Kratky plot analysis (figure 5.2(d)) shows a single hyperbolic rise that does not tend to the x-axis, indicating a single domain protein with flexible elements. There was minor aggregation in the Guinier region and the trimmed profile was fitted in the range of $0.1 < (q \cdot R_g)^2 < 1.7$ with even distribution in the residuals between ± 0.02 (figure 5.2(e)). As the Guinier fitting range was appropriate, calculation of the P(r) and building of a representative model was attempted using the inbuilt ScÅtter P(r) functionality and DAMMIF (ATSAS *ver.* 3.0), respectively. Iterative testing of maximum particle diameter values gave a D_{max} of 63 Å, generating a smooth P(r) model with a left skewed Gaussian-like distribution with a real *vs.* reciprocal space R_g comparison of 24.13 *vs.* 25.81 Å (figure 5.3(a)).

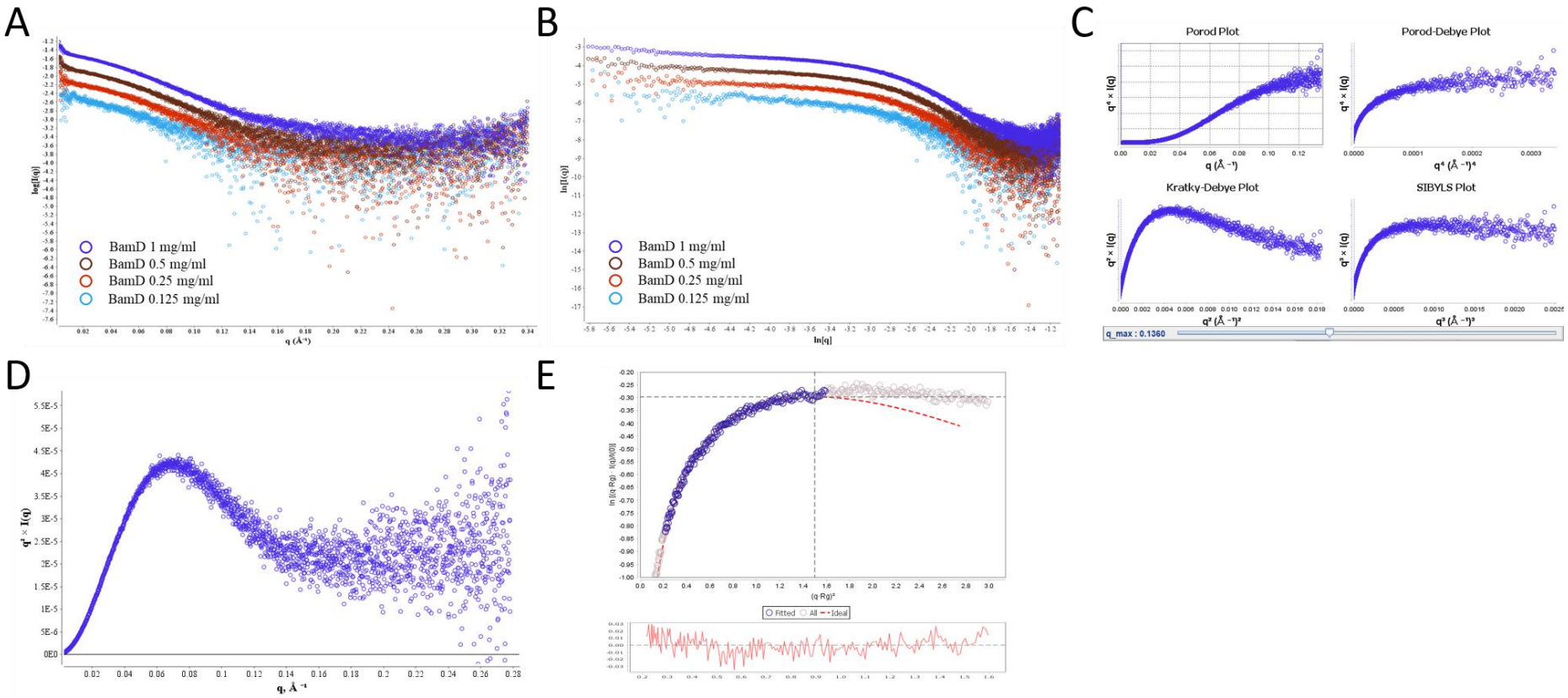


Figure 5.1: BamD SAXS analysis. BamD SAXS $\log_{10}$ (A), Power-law (B), flexibility (C), Kratky (D), and Guinier peak (E) analyses. Data signal in the mid-q (A) and low-q (B) are noisy below 0.5 mg/ml. The flexibility plot (C) suggests a rigid compact shape with potentially some flexibility. The Kratky plot (D) has a single peak, suggesting a rigid compact shape and single domain protein. As the peak does not tend to the x-axis this suggests flexibility. The Guinier fitting (E) was good ($0.2 < q < 1.6$) with a few points excluded after trimming of uptick at low-q.

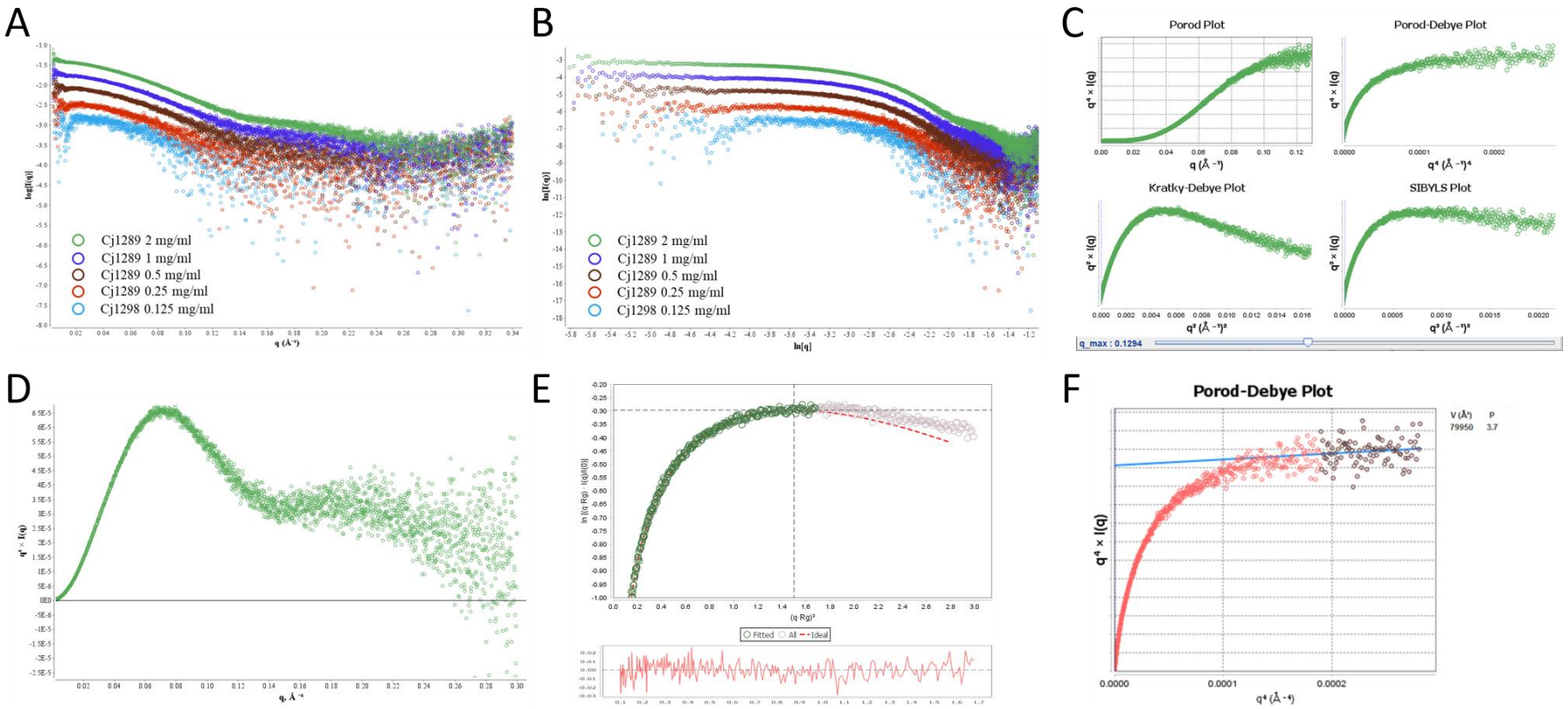
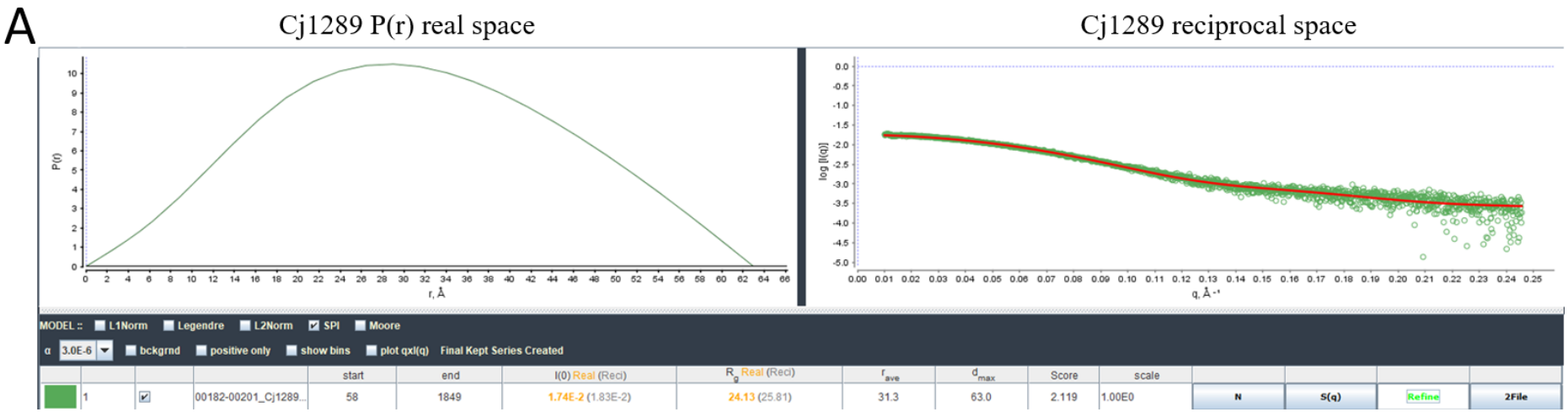


Figure 5.2: Cj1289 SAXS analysis. Cj1289 SAXS $\log_{10}$ (A), Power-law (B), flexibility (C) Kratky (D), Guinier peak (E) and Porod-Debye (F) analyses. Data signal in the mid-q (A) and low-q (B) are noisy below 2 mg/ml. The flexibility plot (C) suggests a rigid compact shape with potentially some flexibility. The Kratky plot (D) has a single peak, suggesting a rigid compact shape and single domain protein. As the peak does not tend to the x-axis this suggests flexibility. The Guinier fitting (E) was good with a very few points excluded after trimming of uptick at low-q. The Porod-Debye plot (F) suggests a volume of 79950Å^3 (~40 kDa) and Porod-exponent of 3.7.



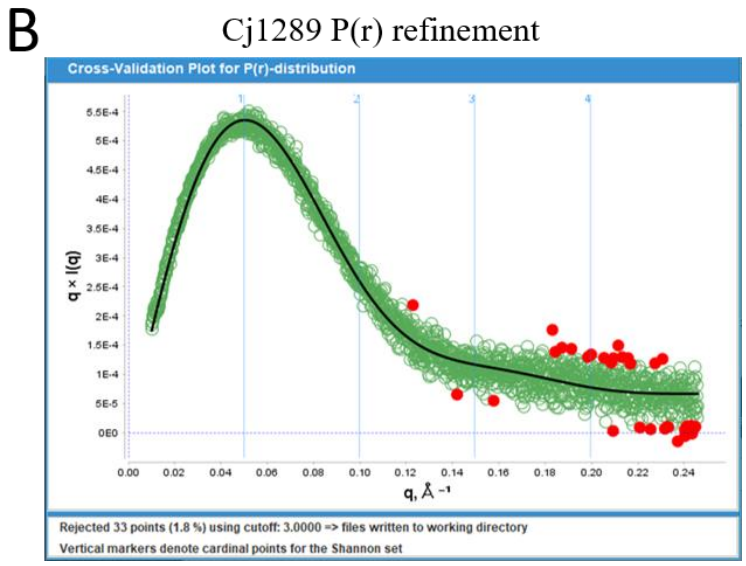


Figure 5.3: Cj1289 P(r) calculation and refinement. The Cj1289 (A) Porod real space and reciprocal space fittings to the SAXS curve and (B) Shannon refinement of the fitting are shown. (a) The real space and reciprocal space fittings show good agreement for R_g and I(0), with the real space fit suggesting an overall globular or flat disk type structure possibly with some minor elongation (left pane), and a close fit of the model projected SAXS data to the actual data (right pane). (B) The fitting was further refined and 33 points were rejected from the 1792 points used for the fitting mainly in the high-q region. The majority of these rejected points likely reflect the noise in the signal in the data at high-q.

The Shannon channel refinement of the model results in the rejection of 33 points (1.84 %) of the 1791 data points used to generate the initial distribution (figure 5.3(b)), reducing the Chi squared score for the model to 2.119. The distribution was used to generate a model using DAMMIF, which was further processed (via DAMSEL, DAMSUP, DAMAVER and DAMMIN) using the ATSAS (*ver.* 3.0) package to produce a refined density model that looks like a ratchet or spring clamp (figure 5.4). Comparison of the homology model to the density by manual alignment resulting in a good overall fit in terms of shape and size, but the density suggests a larger space and has a more extended structure than the available structural models.

5.3.3 BamA, DegP and LptD

The Guinier fittings were poor for BamA, DegP and LptD, and limited analysis to the standard $\log_{10}$, Power-law, flexibility, Kratky, Guinier fitting and (where appropriate) Porod-Debye analyses. The plots for these analyses for BamA, DegP and LptD (appendix 8.5, figures 8.22, 8.23 and 8.24, respectively) are shown. These results will be briefly described. The $\log_{10}$ and Power-law plots confirmed the overall shape was independent of concentration and good signal quality extended to low-q at 1 mg/ml for BamA, DegP and LptD (figures 8.22(b), 8.23(b), and 8.24(b), respectively), and to high-q for BamA and LptD (figures 8.22(a) and 8.24(a), respectively). A noisy high-q region was observed for DegP (figure 8.23(a)). The flexibility assessment indicated the overall structure is mainly rigid and compact, but has flexible elements to consider for each of BamA, DegP and LptD. The derived q_{max} from the flexibility plots was 0.2251 for BamA, 0.0427 for DegP (due to noise at high-q), and 0.2406 for LptD (figures 8.22(c), 8.23(c) and 8.24(c), respectively). Kratky plot analyses showed two peaks (the first elongated, and the second hyperbolic) for BamA and LptD, which indicated multidomain proteins, and a shoulder at low-q in the single hyperbolic peak for DegP, which indicates the presence of smaller and larger multimeric complexes (figures 8.22(d), 8.24(d) and 8.23(d), respectively). The Guinier peak fittings were poor for BamA, DegP and LptD (figures 8.22(e), 8.23(e) and 8.24(e), respectively), with very few points available for fitting.

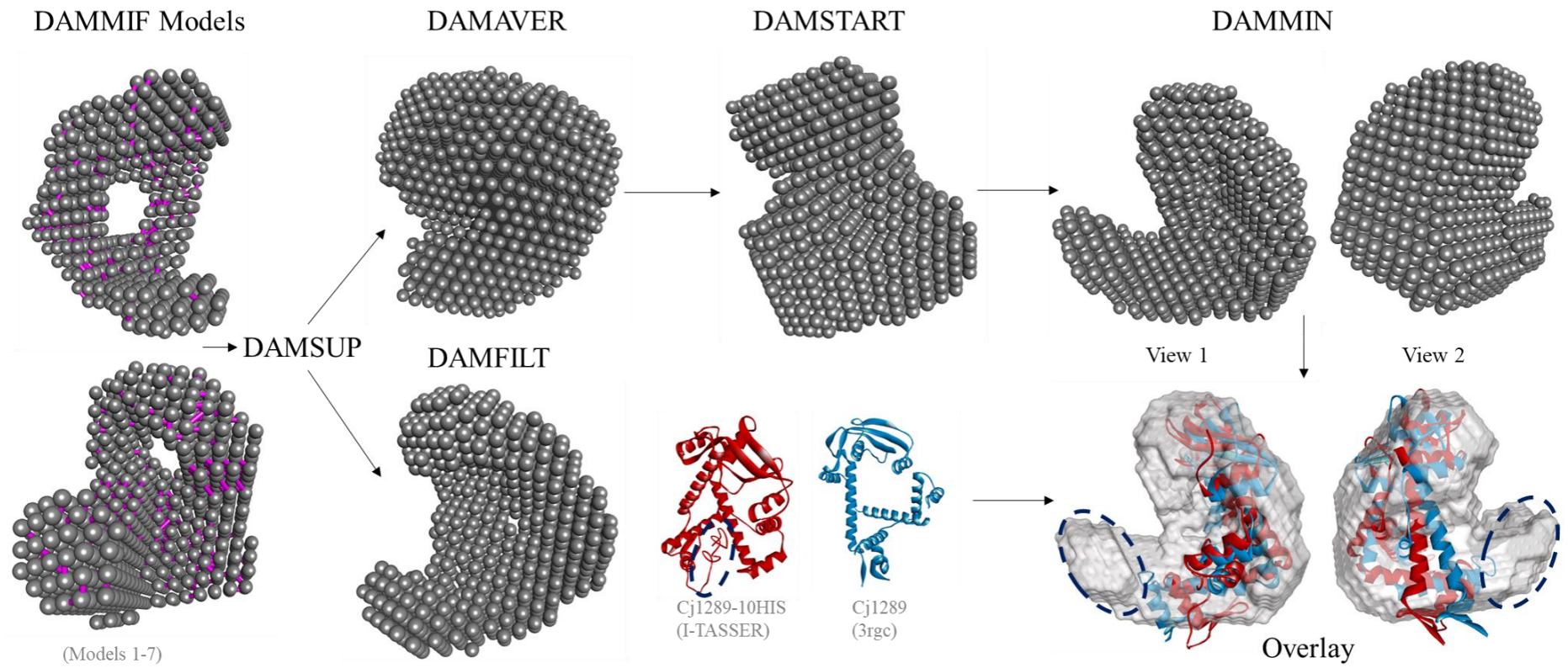


Figure 5.4: Cj1289 DAMMIF model comparison to rigid body structure. Dummy atom modelling process of Cj1289. Models are shown for each stage in the modelling process and homology models are overlaid onto the final model for comparison. The fit is good for the Cj1289 crystal structure model (PDB: 3RGC), and the unnaccounted density may be explained by the 10-His tag or conformational differences in solution relative to the crystal structure model.

The Porod-Debye plot for BamA (figure 8.22(f)) indicated a volume of 288092 $Å^3$, which corresponds to 144 kDa and fits expectations if the 150-180 SB12 molecules (M/W 335.5 Da) required around the BamA β-barrel are considered. (135,077-145,142 Da); the Porod-exponent was 3.8.

5.3.3.1 **BamD-Cj1289**

The $\log_{10}$ and Power-law plots (figures 5.5(a) and 5.5(b), respectively) shows the overall shape that the SAXS curve obtained was independent of concentration for the BamD-Cj1289 mixture. The signal at mid-q to high-q is poor below 1 mg/ml and is noisy at high-q in the 1 mg/ml sample. The flexibility assessment (figure 5.5(c)) shows that the Porod-Debye plot levels off first, indicating the overall structure in mainly rigid and compact. The derived q_{max} from the flexibility plots was 0.1461. The Kratky plot (figure 5.5(d)) shows a single hyperbolic rise, indicating a single domain protein, and the data plot appears to tend to the x-axis (at $q^2 \cdot I(q) = 0$) indicating the structure is folded and either rigid or with limited flexibility (difficult to discern due to noisy mid-q to high-q region). There was some aggregation apparent in the Guinier region in the untrimmed profile, but the profile was fitted manually with some trimming at low-q, and fits the curve well in the Guinier peak analysis (figure 5.5(e)) in the range of $0.1 < (q \cdot R_g)^2 < 1.7$ with residuals between +0.06 to -0.02.

As the Guinier fitting range ($0.1 \leq q \cdot R_g^2 \leq 1.7$) was appropriate for modelling, calculation of the P(r) and building of a representative model was attempted using ScÅtter and DAMMIF (ATSAS *ver.* 3.0), respectively. Iterative testing of maximum particle diameter values gave a D_{max} of 102 Å (figure 5.6(a)), generating a smooth P(r) model with a left skewed Gaussian-like distribution with extended arm (which may indicate a rigid core with an extended structure). The R_g calculation returned a real *vs.* reciprocal space comparison of 30.99 *vs.* 30.71 Å, and refinement of the model (figure 5.6(b)) results in the rejection of 45 points (2.10 %) of the 2104 data points used to generate the initial distribution, bringing the Chi squared score to 2.762. Following a similar modelling scheme as previously, a refined model of the BamD-Cj1289 sample was generated using DAMMIN (figure 5.7).

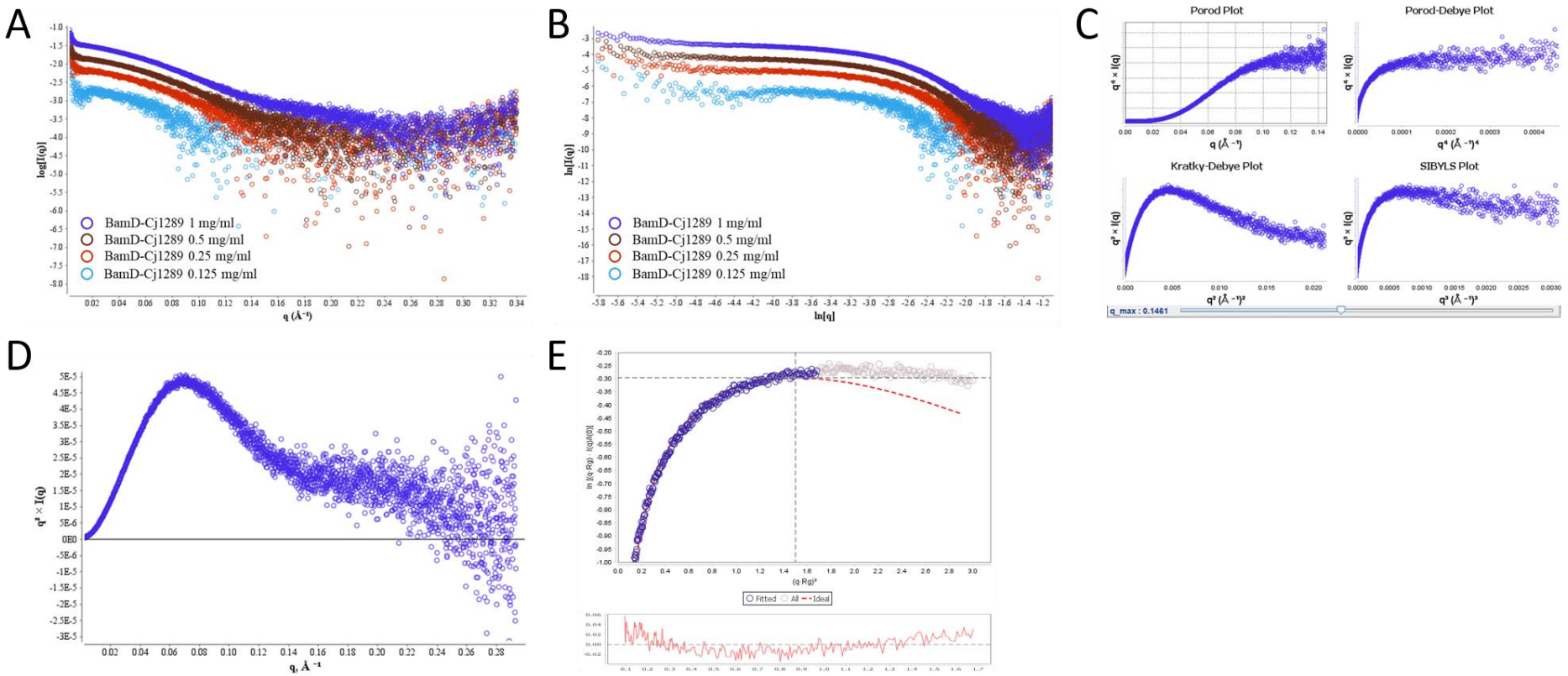
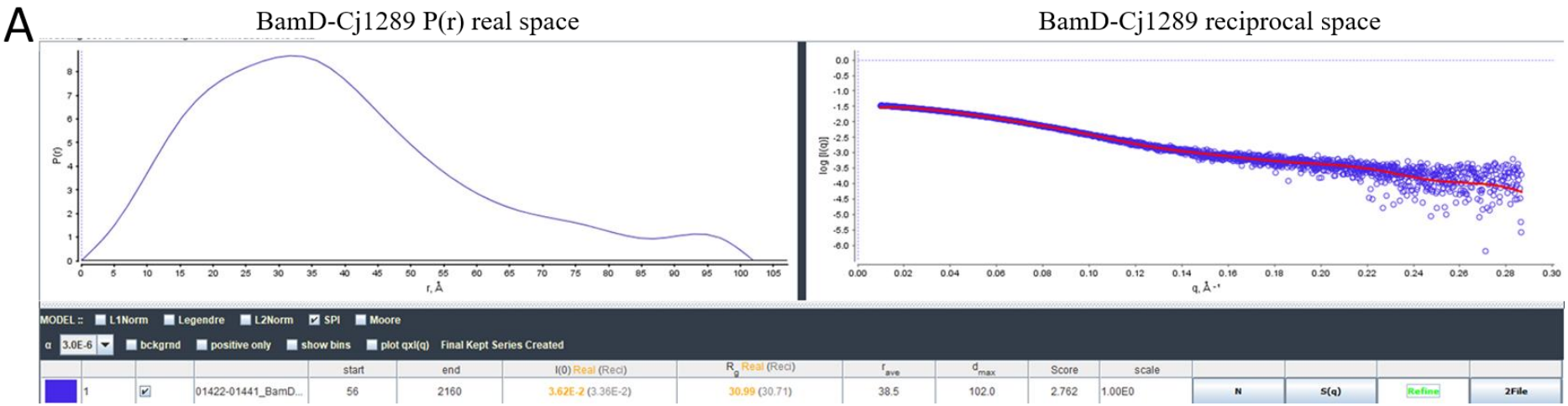


Figure 5.5: BamD-Cj1289 SAXS analysis. BamD-Cj1289 mixture $\log_{10}$ (A), Power-law (B), flexibility (C), Kratky (D), and Guinier peak (E) analyses. Data signal in the mid-q (A) and low-q (B) are continuous at 1 mg/ml. The flexibility plot (C) suggests a rigid compact shape with potentially some flexibility. The Kratky plot (D) has a single peak in the low-mid-q suggesting a rigid compact shape. Flexibility is uncertain due to the noise at high-q, but the plot appears to tend to the x-axis suggesting limited flexibility. The Guinier fitting (E) was good ($0.1 \leq q \cdot R_g^2 \leq 1.7$) with very few points excluded after trimming and residuals are between -0.02 and +0.04 (except the first data point at +0.06).



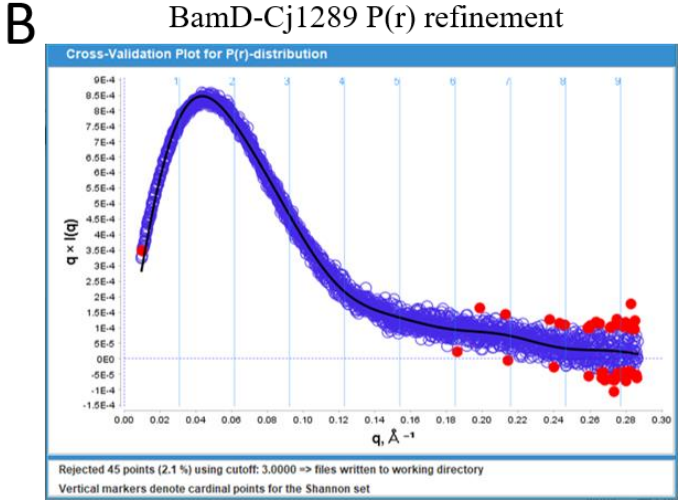


Figure 5.6: BamD-Cj1289 P(r) calculation and refinement. The BamD-Cj1289 (A) Porod real space and reciprocal space fittings to the SAXS curve and (B) Shannon refinement of the fitting are shown. (A) The real space and reciprocal space fittings show good agreement for R_g and I(0), with the real space fit suggesting an overall globular structure with an elongated aspect or long rod-like skew (left pane), and a close fit of the model projected SAXS data to the actual data (right pane). (B) The fitting was further refined and 45 points were rejected from the 2105 points used for the fitting, mainly in the high-q region. The majority of these rejected points likely reflect the noise in the signal in the data at high-q.

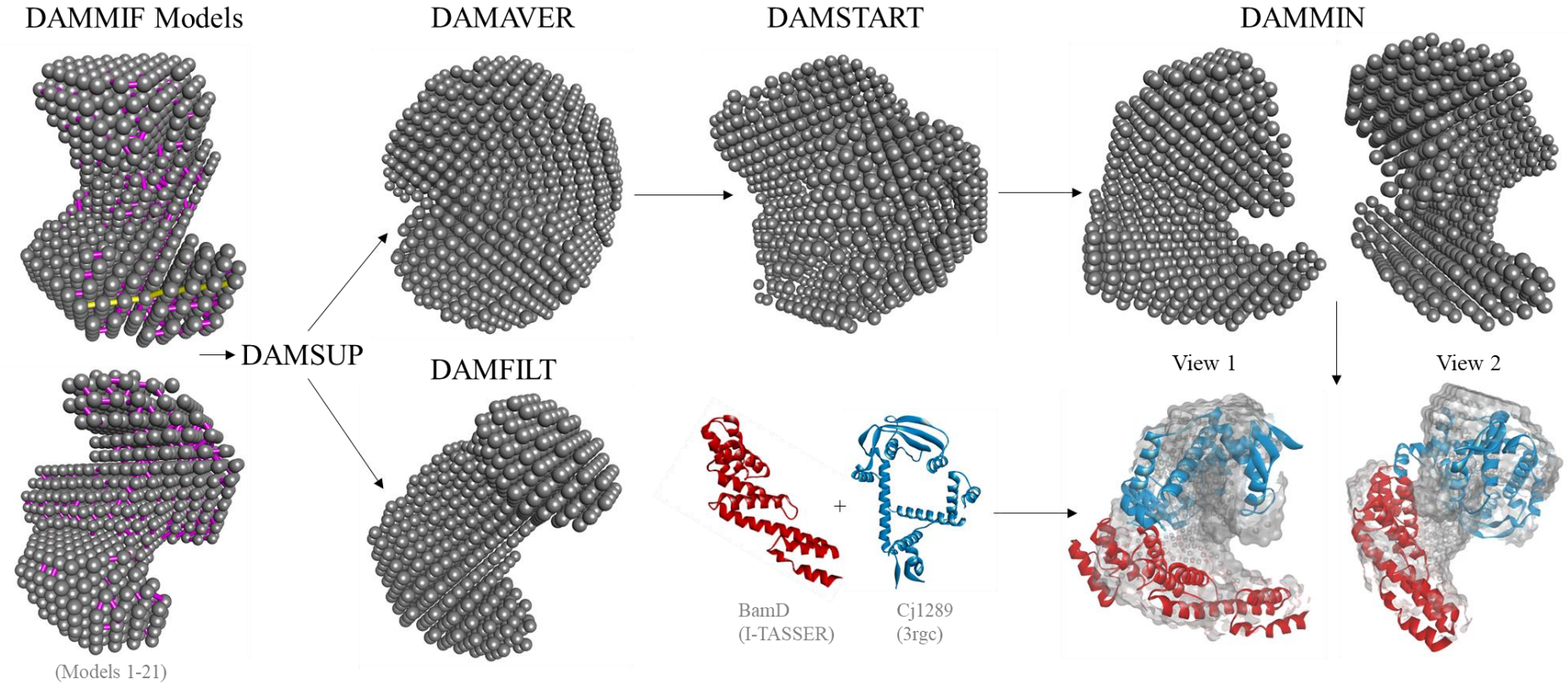


Figure 5.7: BamD-Cj1289 bead modelling. Dummy atom modelling process of the BamD-Cj1289 mixture. The models and associated programs used in their generation are shown for each stage in the modelling process and homology models are overlaid onto the final model for comparison. The fit was good for the unnanealed models of Cj1289 (PDB: 3rgc) and the I-TASSER generated homology model of BamD, though the slight mismatch of both structures to the density may suggest conformation rearrangements of the proteins in solution or as a result of interaction relative to the available structural models.

Comparison of the density model generated to the homology model by manual alignment (figure 5.7) shows a very good fit of the homology model to the density in terms of size, shape and space filling, and confirms the potential binding interaction between Cj1289 and BamD observed in the last chapter. The density model was slightly more compact than the BamD rigid homology and Cj1289 3RGC crystal structure models, and there are small gaps in the fitting of the homology models to the density between the proteins.

5.3.3.2 **Other mixtures**

The other mixtures in this study (totalling 13 mixture samples at four concentrations each with associated buffer runs) had increasing severity of aggregation to varying degrees due to aggregation present in the individual samples used. As the aggregation limited (and possibly skewed) any further analysis past the basic plots and complexation analysis (below, does not rely solely on low-q Guinier region analysis), these results have not been shown.

5.3.4 **Homology model fitting to experimental data**

To examine whether the experimental SAXS data curves were accurate descriptions of the proteins in this study (and *vice versa*), crystal structures or homology models were used to generate experimental SAXS profiles for comparison. The experimental SAXS curves were left untrimmed for the fitting and the maximal range of the simulated *in silico* SAXS profiles was used for comparison. The profiles were visually compared after using the scale function to align the profiles on ScÅtter.

The alignments on ScÅtter show good profile fits of models to experimental data for the 3RGC chain A monomer to Cj1289 (figure 5.8(a)) with a very good fit throughout the profile, and the 3RGC dimer to Cj1289 (figure 5.8(b)) which shows a disagreement in fitting at mid-q but fits excellently at low-q and high-q. Overall, the 3RGC crystal structure chain A monomer provides a better fit to the experimental data than the full 3RGC dimer model, indicating Cj1289 is a monomer in solution contrary to the published crystal structure model.

A rigid homology model generated through I-TASSER (based on available *E. coli* structures) and a hydrated annealed MD trajectory projection of this homology model were fitted to the BamD experimental SAXS data.

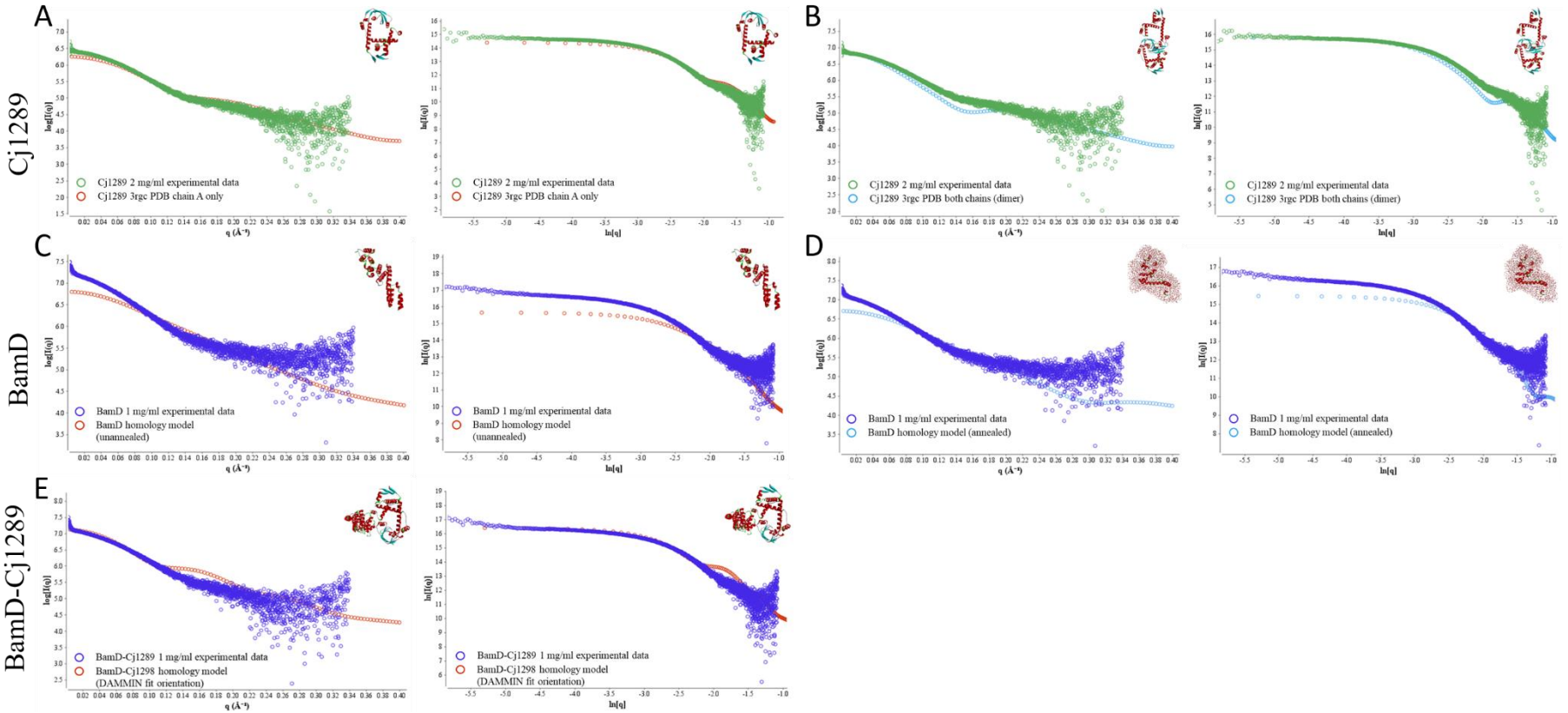


Figure 5.8: Experimental data validation of Cj1289, BamD and BamD-Cj1289 mixture using model structures. Comparison of the experimental data to available models via simulated SAXS curves on $\log_{10}$ (left panel) and Power-law (right panel) plots for (A) Cj1289 (3RGC chain A/ monomer), (B) Cj1289 (3RGC dimer), (C) BamD (homology model), (D) BamD (MD annealed homology model with hydration layer), (E) BamD-Cj1289 (orientated as in DAMMIN density). Good SAXS signal cross-validation was observed for Cj1289 monomer (A) and minor disagreements for the dimer (B). BamD models (C, D) disagree at low and high-q, and BamD-Cj1289 (E) models disagree slightly at mid-q.

The rigid homology model for BamD fits the experimental data mid-q fairly well but shows deviation at low- and high-q (figure 5.8(c)). The MD annealed model shows an excellent overlapping profile at mid-q, lessens the disagreement to the experimental data at low-q and high-q, and models the upward turn at high-q observed in the experimental data (figure 5.8(d)). Overall, neither available model fits very closely to the experimental data at low- and high-q, though the hydrated annealed model appears the better fit.

The BamD homology model and Cj1289 3RGC monomer were manually assembled to approximate a BamD-Cj1289 homology model in the DAMMIN density model orientation, and was fitted to the BamD-Cj1289 mixture experimental profile (figure 5.8(e)). The simulated profile shows an overall good fit to the experimental data, with good fitting at low-q and high-q and disagreement at mid q. The relatively good fitting, particularly at low-q, indicates that the experimental profile was the result of the mixture of these two proteins in close proximity or in complex.

Several models, either homology model, *E. coli* crystal or *C. jejuni* Cryo-TEM structures were fitted to the experimental SAXS curve obtained for *C. jejuni* DegP. The DegP monomer homology model (figure 5.9(a)) fits the profile well in the region $0.04 \leq q \leq 0.24$, but deviates outside this range at very low and high-q. Hydration and MD annealing of the homology model monomer improves the fit slightly at $q \geq 0.12$ but causes slightly larger disagreement at $q \leq 0.10$ (figure 5.9(b)). The assembled hexamer (built on *E. coli* hexamer model 1KY9 from *C. jejuni* DegP homology model monomers) shows a similar fit to the homology model monomer with slight improvements at $0.04 \leq q \leq 0.08$ (figure 5.9(c)). The *E. coli* 12-mer fits discrete small sections of q, but presents an overall poor fit to the *C. jejuni* DegP profile (figure 5.9(d)). The *C. jejuni* 24-mer cryo-TEM structure (6z05; figure 5.9(e)) fits a small discrete section of low-q, but was otherwise a poor fit. Overall, the *C. jejuni* DegP model comparison suggests *C. jejuni* DegP was present as a monomer or hexamer in solution under the conditions screened (30°C). The BamA and LptD homology models fit poorly to the experimental data (figure 5.10), though for LptD this was mitigated somewhat (limited fit to mid-q) by the hydrated and FC12 assembled LptD model (figure 5.10(b)). Due to size of the BamA MD simulation model in SB12 (~820,000 atoms) and hardware restrictions, it was not available for comparison to the BamA model.

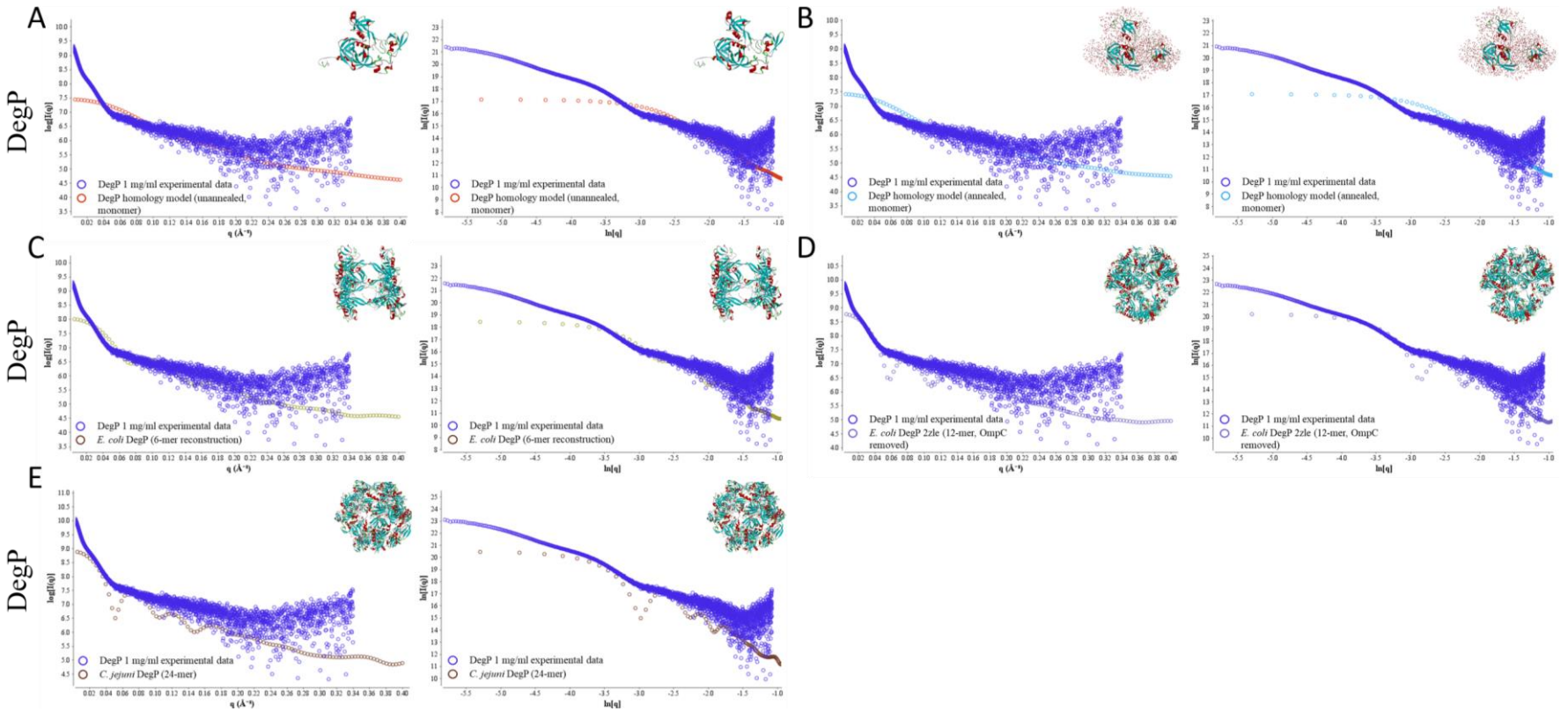


Figure 5.9: Experimental data validation of DegP using monomer, hexamer, dodecamer and 24-mer model structures. Comparison of experimental data to available models via simulated SAXS curves displayed on $\log_{10}$ (left panel) and Power-law (right panel) plots for DegP (A) *C. jejuni* homology model monomer, (B) hydrated MD annealed monomer, (C) approximated hexamer (built on 1KY9 *E. coli* model), (D) *E. coli* 12-mer (2ZLE, OmpC removed), and (E) *C. jejuni* 24-mer [6z05]). The monomer (A) and hexamer (E) fit the experimental data reasonably well in the mid-q and high-q, but aggregation prevents fitting at low-q and uptick at high-q was unaccounted for by the model structures.

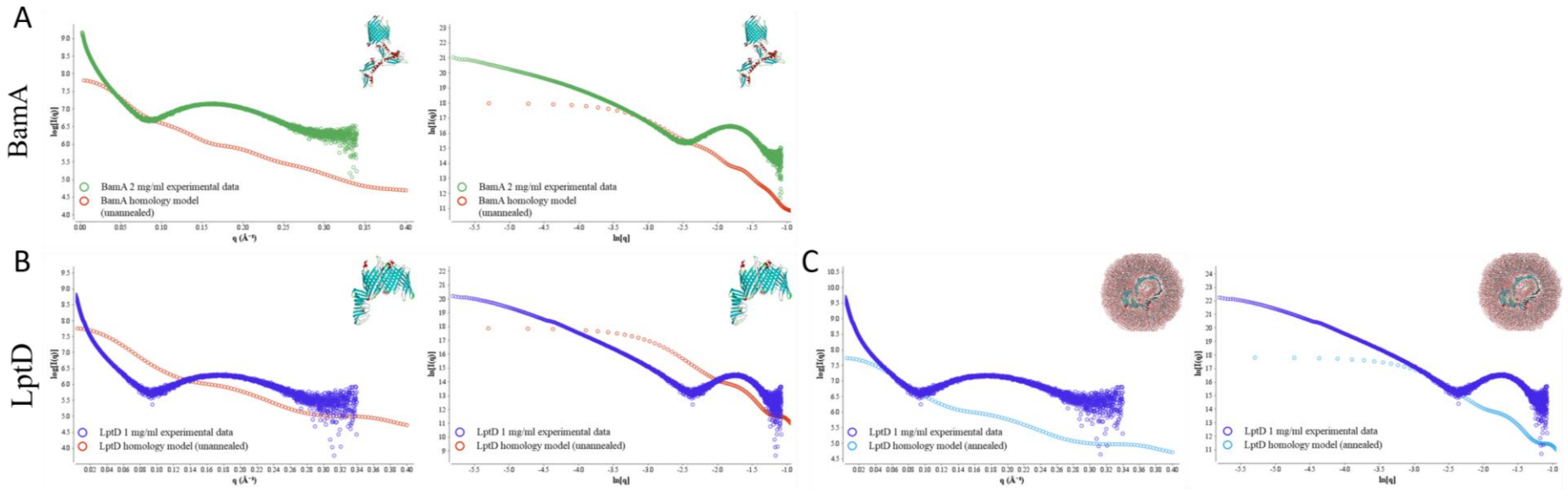
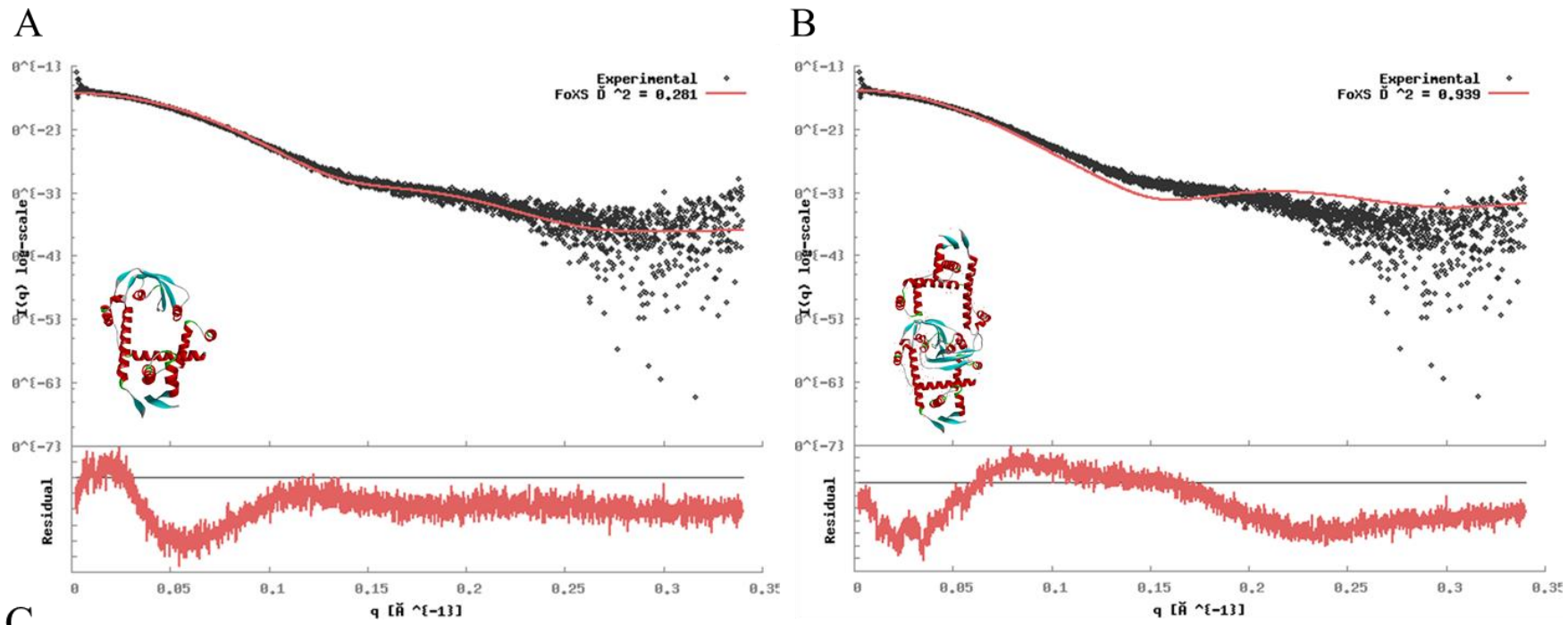


Figure 5.10: Experimental data validation of BamA and LptD using homology model structures. Comparison of the experimental data to available homology models via simulated SAXS curves displayed on $\log_{10}$ (left panel) and Power-law (right panel) plots for (A) BamA homology model (unannealed), (B) LptD homology model (unannealed) and (C) LptD homology (hydrated MD annealed in FC12).

5.3.5 FoxS and GAJOE cross-validation of SAXS profiles

The alignments for Cj1289 and BamD-Cj1289 were validated using the FoxS server[135] (figures 5.11 and 5.13, respectively), and for Cj1289 using GAJOE[200, 201] (figure 5.12). Fitting of the 3RGC chain A monomer to the Cj1289 experimental profile using FoxS (figure 5.11) returns a Chi squared value of 0.28, which indicates either potential overfitting or lower error than expected on average, and 0.94 for the 3RGC dimer fitting. The overall profile fit of the monomer was excellent throughout the profile, and good with small disagreements at mid to high-q for the dimer. The atom radii scaling parameter (C_1) is 1.05, the maximal allowed value ($0.98 \leq C_1 \leq 1.05$), for both fits to Cj1289, suggesting that the model is too compact to give the experimental data or potential overfitting of the data. The hydration shell density parameter (C_2) is +4.00 for the 3rgc chain A monomer fitting indicating a higher hydration shell density than the bulk solvent of 0.388 e/Å^3, and -2.00 for the 3RGC dimer indicating a lower hydration shell density than the bulk solvent of 0.307 e/Å^3. As these are the maximal and minimal values allowed for the hydration shell density parameter in FoxS, this may indicate overfitting of the data, missing electron density (+4.00), model inaccuracies, or a reflection of surface charge (-2.00). A secondary validation using GAJOE (figure 5.12), shows a good profile fitting over the range of q sampled, and from the MD simulation of Cj1289 (3520 frames, 352 ns) used as an input, identifies two models for Cj1289 than contribute to the SAXS profile observed (figure 5.12(b)). The first model shows a state similar to the 3RGC crystal structure monomer that presents a ‘closed’ conformation, and the second model identified shows an ‘open’ conformation of Cj1289 that requires further investigation. A calculation of which structure predominates in solution indicates the second ‘open’ model predominating in a roughly 2:1 ratio (figure 5.12(c)). For the BamD-Cj1289 model cross-validation using FoxS (figure 5.13), a Chi squared value of 0.93 was given, indicating a good fitting using FoxS and that experimental errors fall within predictions for the data set. The atom radii scaling parameter deviates from its nominal value by 2 % (1.02) indicating the scaling of the atomic radii needs to be increased slightly to fit the model to the density and that the model is slightly too compact for the experimental data, however this is well within the given limits. The hydration shell density parameter is -2.00, indicating a lower hydration shell density than the bulk solvent, and may be reflective of the surface charge properties.



C

Model	χ^2	C_1	C_2	R_g	atoms
3rgc chain A	0.28	1.05	4.00	22.66	1898
3rgc dimer	0.94	1.05	-2.00	30.23	3803

Figure 5.11: Cj1289 FoxS fitting cross-validation. FoxS fitting of Cj1289 (A) 3rgc chain A monomer, and (B) 3rgc dimer to the experimental data profile. Associated fitting output values for each fitting (C) are shown. The FoxS fitting suggests that the simulated profile for the Cj1289 (3RGC conformation) monomer fits the experimental data closely (χ^2=0.28), and the dimer fits reasonably well (χ^2=0.94) within normal experimental error.

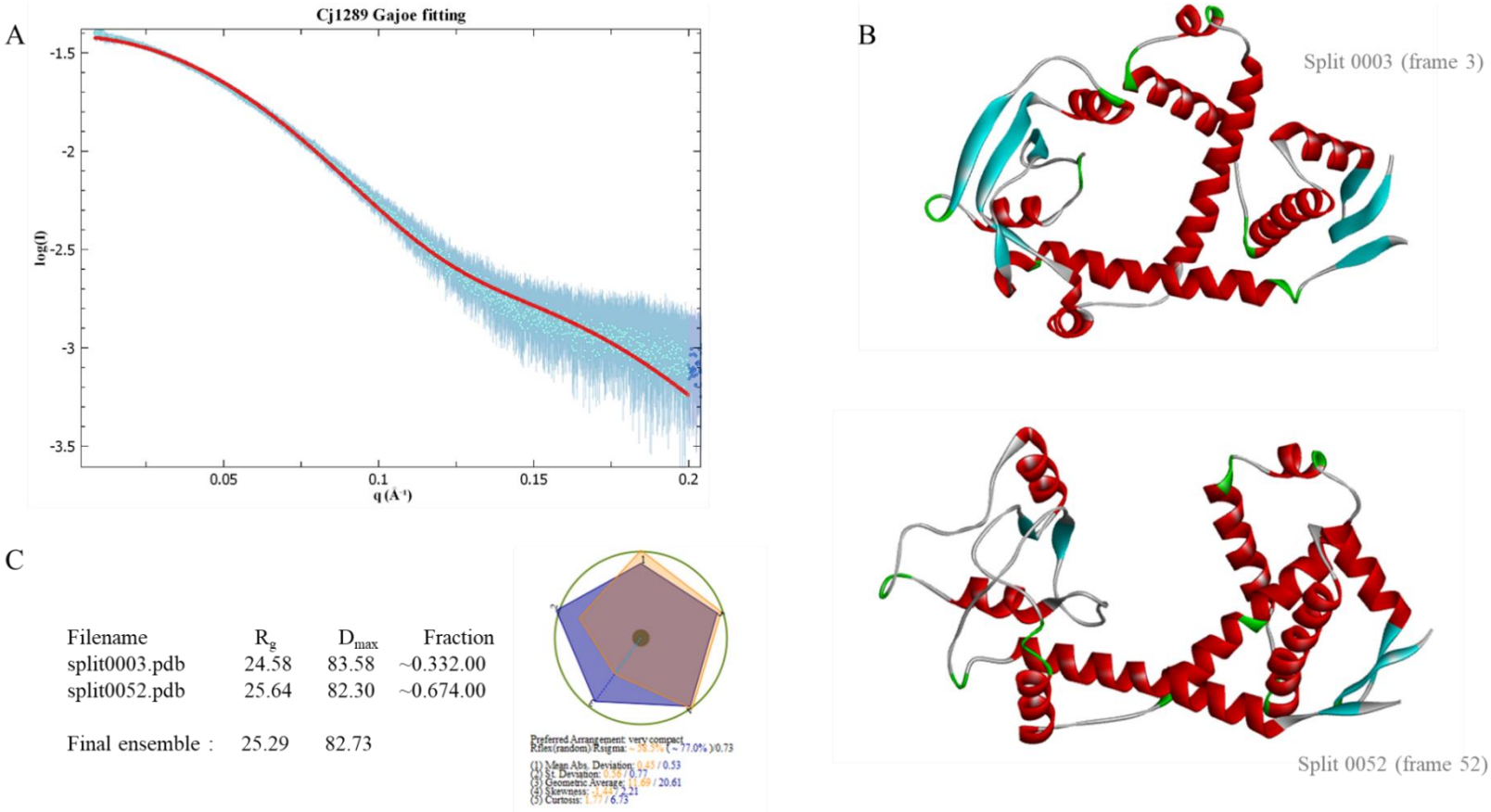
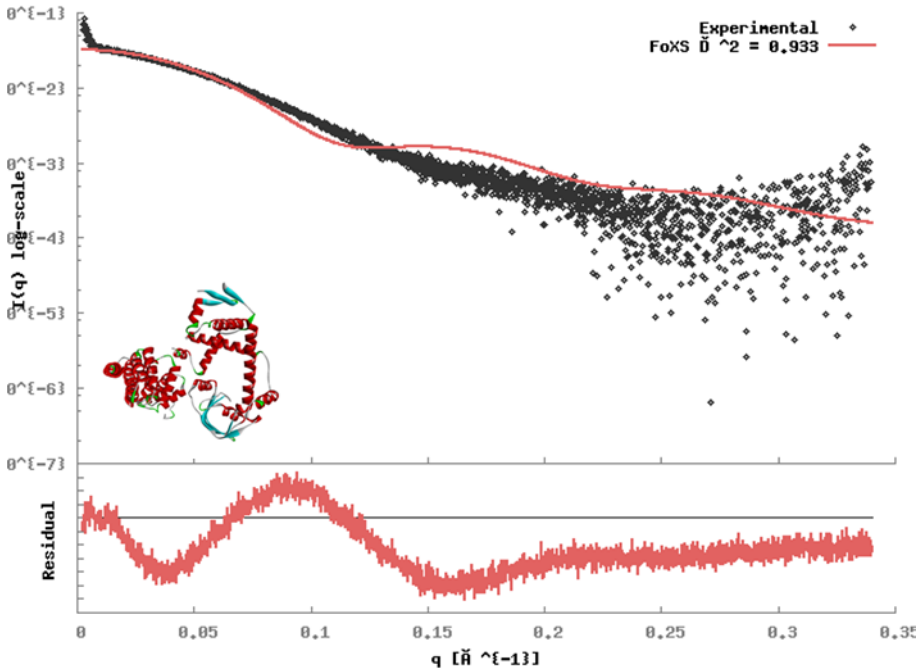


Figure 5.12: Cj1289 GAJOE fitting and validation. A 352 ns Cj1289 MD simulation was used as an input to GAJOE to fit and validate the SAXS data. The fitting to the SAXS profile (A), two models that contribute to the fitting (B), and statistics for the fitting (C) are shown. The two models that make up the fitting appear to illustrate an 'open' and 'closed' confirmation, of which the 'open' confirmation (B, split 0052) was calculated to contribute in a 2:1 ratio (C, fraction) relative to the 'closed' conformation to the profile observed experimentally.

A



B

Model	χ^2	C_1	C_2	R_g	atoms
BamD-Cj1289 homology model	0.93	1.02	-2.00	27.94	3565

Figure 5.13: BamD-Cj1289 FoxS fitting cross-validation. FoxS fitting of an approximately assembled BamD-Cj1289 homology model (manually orientated as in DAMMIN model) to BamD-Cj1289 experimental data (A) and associated fitting output values (B) are shown. The χ^2 and C_1 values suggest this is a reasonable approximation of the data.

5.3.6 Complexation analysis

The proteins in this study were mixed in a 1:1 stoichiometry in their associated buffers at 30°C to examine complex formation and infer protein-protein interaction. In addition, 6:1 and 24:1 stoichiometry samples with equilibration before measurement at either 16°C or 30°C were generated for the DegP-BamA mixture to observe whether different structures predominate at different stoichiometric ratios and temperatures. The DegP-BamA samples prepared at 16°C were equilibrated at this temperature as DegP functionality is thought to be determined by a temperature "switch" around 22°C, below which it exhibits a chaperone function, and above which a protease function[91]. Within this characterisation, there is an intermediate temperature range between 22°C and 28°C where DegP is thought to prevent aggregation of misfolded proteins, and greater than 28°C DegP exhibits an efficient protease activity[202].

The complex assessment function in ScÅtter was used to test whether a complex was formed in each mixture, and the data has been trimmed back in each instance to remove excessively noisy data at the low-q and high-q range, which may skew interpretation, with all samples shown within q-limits of $0.2 \leq q \leq 0.28$.

The BamA complexation analysis residuals suggest that there are interactions with all of the remaining proteins in this study (figure 5.14(a-d), respectively).

In contrast to this, the BamD complexation analysis evenly distributed residuals and well fitted SAXS curves suggest that, except BamA (figure 5.14(a)), there are no interactions between BamD and the other proteins in this study (figure 5.15, panels a, b, d, respectively). The result of mixing BamD with Cj1289 (figure 5.15(a)) or LptD (figure 5.15(d)) is inconclusive, but as the shift from linear residuals was subtle and the projected SAXS curve and experimental data fit well in both cases, they have both been marked as a non-interactions.

LptD complexation analysis suggests interaction with BamA (figure 5.14(d)) and both chaperones (figures 5.14(e) and (f)), suggesting that there is interaction with all proteins in this study, with the possible exception being BamD (figure 5.15(d)).

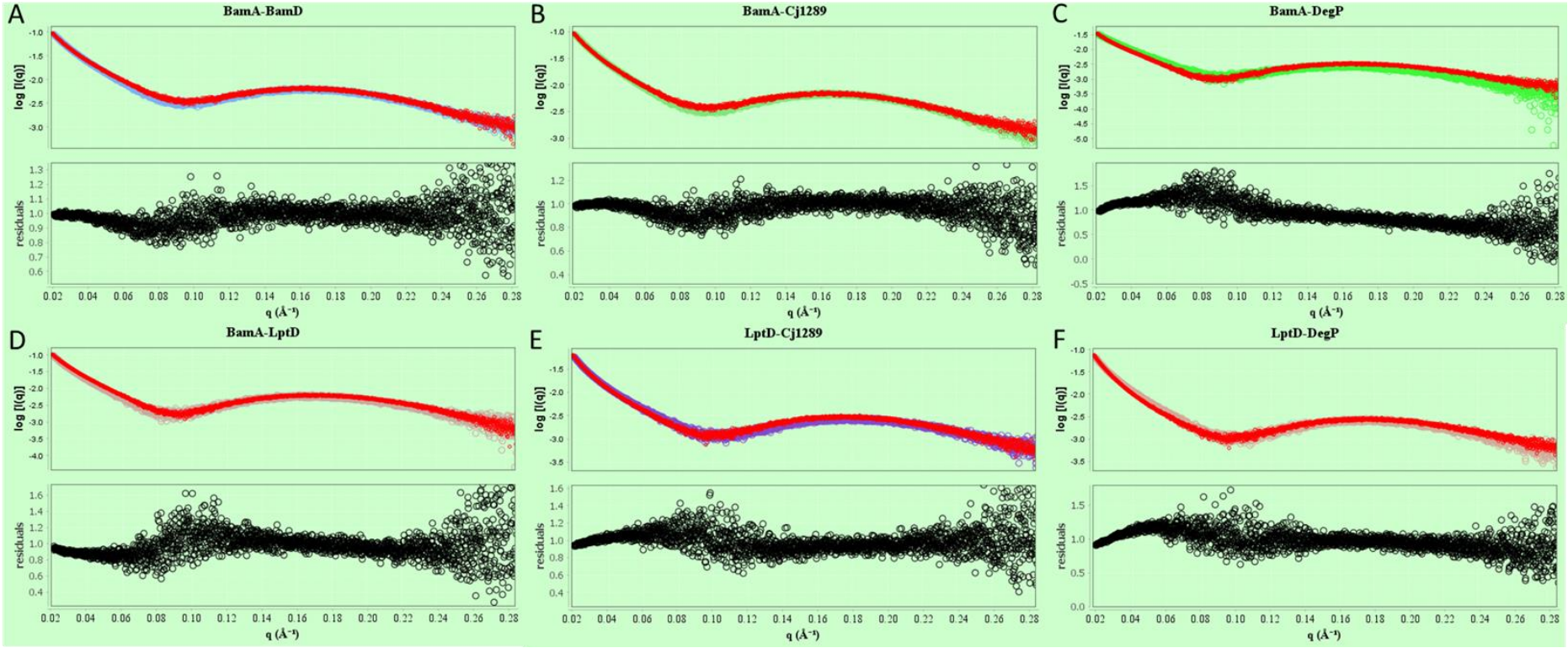


Figure 5.14: Complexation analyses of 1:1 protein mixtures that indicate interaction. Complexation analysis showing protein:protein mixture SAXS profiles superimposed onto simulated profiles and residuals of (A) BamA-BamD, (B) BamA-Cj1289, (C) BamA-DegP, (D) BamA-LptD, (E) LptD-Cj1289, and (F) LptD-DegP. All profiles show deviation from linearity and homogeneity in the residuals indicating the protein:protein 1:1 mixtures tested all form complexes or interact.

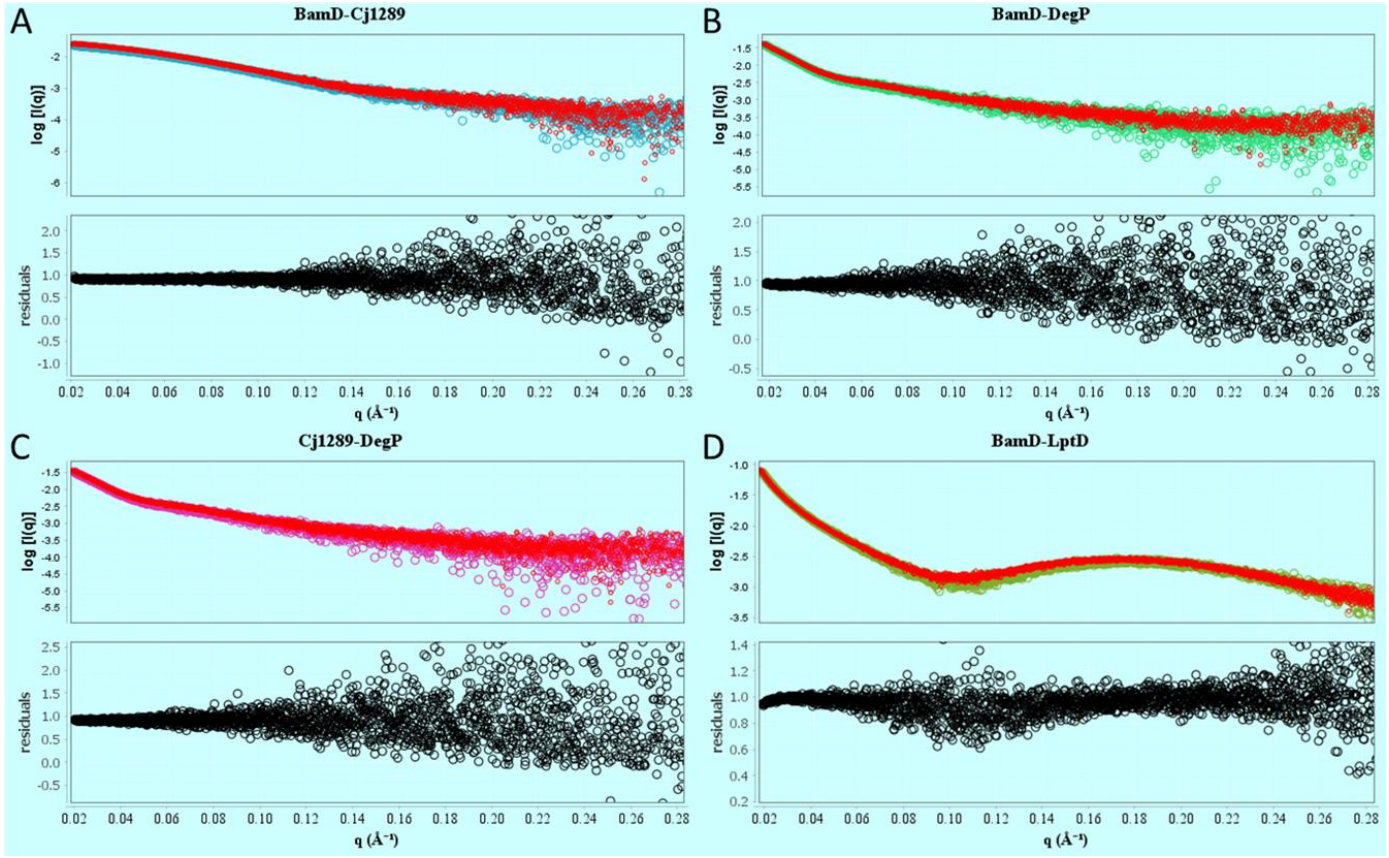


Figure 5.15: Complexation analyses of 1:1 protein mixtures that indicate no interaction. Complexation analysis showing protein:protein mixture SAXS profiles superimposed onto simulated profiles and residuals of (A) BamD-Cj1289, (B) BamD-DegP, (C) Cj1289-DegP, and (D) BamD-LptD. All profiles show experimental profiles closely matched with simlated non-interaction SAXS profiles and linear and (roughly) homogeneous residuals indicating the protein:protein 1:1 mixtures tested do not form complexes or interact.

For the chaperone Cj1289, the complexation analysis (figure 5.14(b) and (e); figure 5.15(a) and (c)) suggests that there is interaction with BamA and LptD (both β-barrel proteins), but not with BamD (a BAM complex member) or DegP (chaperone), mixtures of which with Cj1289 show linear residuals.

For the conditional chaperone DegP, the complexation analysis (figure 5.14(c) and (f); figure 5.15(b) and (c)) suggests that there is interaction with BamA and LptD (both β-barrel proteins), but not with Cj1289 (a chaperone) or BamD (the other BAM complex member). Increasing the stoichiometric ratio of DegP relative to BamA in mixture (figure 5.16(a) and (c) *vs.* (b) and (d)) results in an increased deviation from a linear residual and a deviation of the SAXS curve from the experimental data at q-values between 0.04-0.06 and greater than 0.20, with a similar change overall irrespective of temperature. This trend is observed when going from a 1:1 stoichiometric ratio of DegP: BamA (single kink at approximately q = 0.08, figure 5.14(c)) to a 6:1 stoichiometry (kinks in residuals at approximately q = 0.03 and 0.08, figure 5.16(a) and (c)), and again in a more pronounced manner in the 24:1 stoichiometric sample (figure 5.16(b) and (d)). Changing the sample temperature (figure 5.16, panels (a) and (b) *vs.* (c) and (d)), from 30°C to 16°C during the sample measurement seems to increase the rate of change in the residuals in the 6:1 at approximately q =0.08 sample comparison and between q = 0.04 and 0.10 in the 24:1 sample comparison.

The data is summarised in table 5.1.

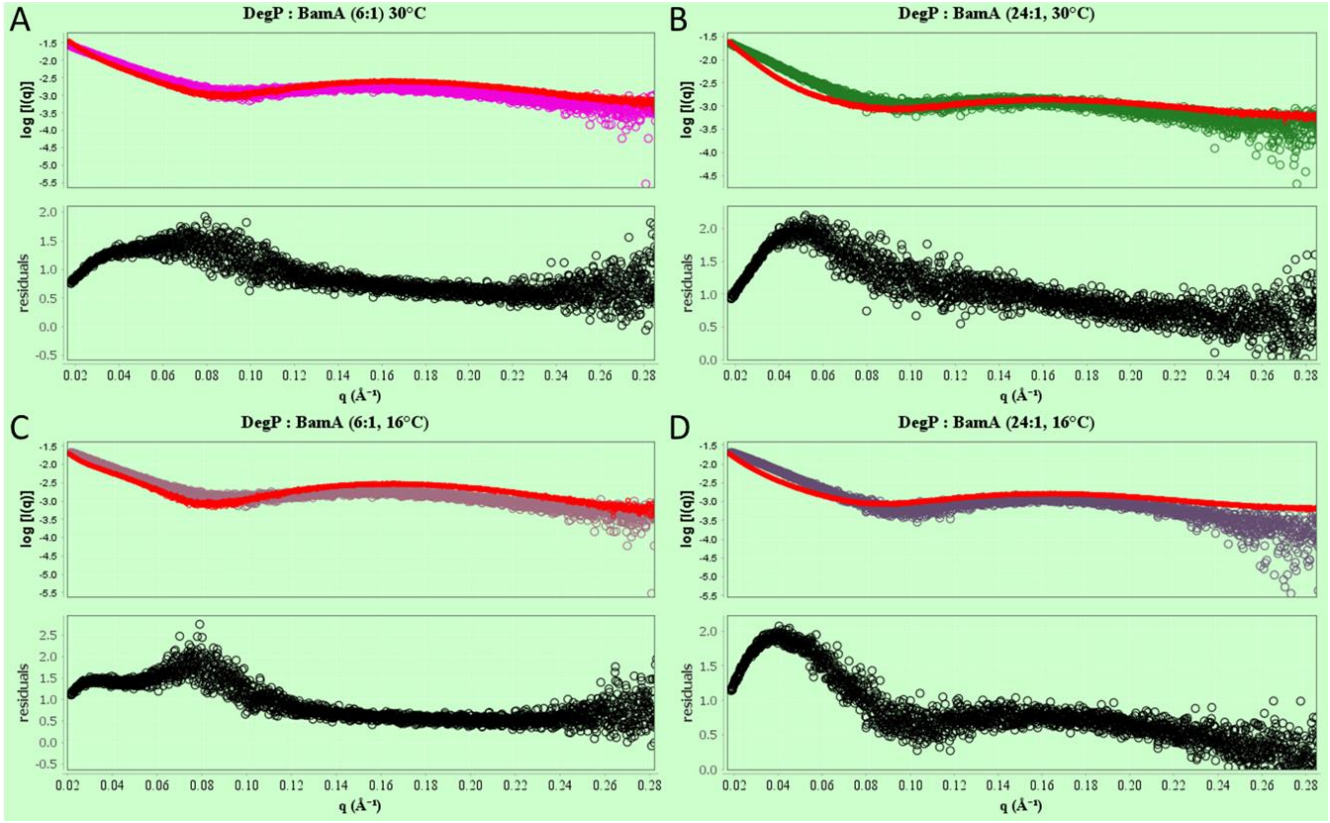


Figure 5.16: Complexation analyses of DegP and BamA mixtures in 6:1 and 24:1 stoichiometries at 16 °C and 30°C. Complexation analysis showing protein:protein mixture SAXS profiles superimposed onto simulated profiles and residuals of DegP-BamA mixtures at 30°C in (A) 6:1 and (B) 24:1 stoichiometries and at 16°C in (C) 6:1 and (D) 24:1stoichiometries. All profiles show deviation in the experimental and simulated (non-interaction) SAX profile overlays, and large deviation from linear homogenous distribution in the residuals indicating that the protein-protein mixtures tested result in complex formation or interaction.

	BamA	BamD	cj1289	DegP	LptD	
BamA		+	-			CD
BamD	+		+			
cj1289	+	-/+[#]				
DegP	+	-	-			
LptD	+	-	+	+		
	SAXS					

[#] Complexation analysis shows subtle changes in residuals which may suggest lack of interaction, but the P(r) based bead model generated for BamD and Cj1289 superimposes well onto the homology model of BamD and crystal structure of Cj1289.

Table 5.1: Summary table of interactions between proteins in this study.

5.4 **Discussion**

The SAXS analysis of BamA suggests a mostly rigid multidomain folded protein with a degree of flexibility. This fits the expectations coming into the experiment as the β-barrel and POTRA 5 should be fairly stable, the POTRA domain and β-barrel can be considered two domains, and the POTRA domain are known to exhibit flexibility due to their elongated spacers between adjoining POTRA domain[50] and the β-barrel has several flexible loops, but otherwise limited flexibility overall. The volume estimation divided by two gives an estimated molecule weight of 144.06 kDa, which is reasonable for a BamA molecule with an estimated 150-180 SB12 molecules forming the detergent micelle around the β-barrel (135 kDa - 145 kDa combined). However, the fit between the experimental SAXS profile and homology model is partial at best ($0.04 \leq q \leq 0.10$), likely due to using an unannealed rigid homology model lacking the SB12 environment of the sample (due to hardware limitation in rendering ~800,000 atoms) and to aggregation observed in the Guinier region, which limited further analysis.

BamD SAXS analysis indicates a single domain folded protein with a mixture of rigid and flexible components, providing a reasonable fit within expectations for a single domain protein composed of a series of helices connected by flexible linker sequences, but with an overall rod-like and compact shape in crystal structures and homology models. The experimental data shows partial overlap limited to the mid-q range with both the rigid homology model and annealed homology model in water, with the annealed model appearing the better fit by eye. The aggregation in the Guinier region is substantial and limits the Guinier fit to the range $0.2 \leq q \cdot R_g^2 \leq 1.6$, which might remove too much information at low-q to allow modelling considering its low molecular weight ($q \cdot R_g^2$ needs to extend to at least 0.1).

The Cj1289 experimental data profile shows a mostly rigid single domain folded protein with some flexible components, in line with expectation for a protein with a toroidal-like compact shape featuring a protruding arm like extension composed of helices and sheets linked by 5-10 amino acid disordered linker or flexible loop like structures. The 3RGC[108] crystal structure chain A monomer is a good fit to the experimental data across the entire q-range of the experimental data set, whereas the dimer observed in the 3RGC crystal structure appears to diverge from the

experimental data in the mid-q range. This is further confirmed by analysis using the FoxS server, and it is likely that Cj1289 is a monomer in solution when in isolation from other proteins. Further validation with GAJOE indicates two conformations are found that jointly can describe the Cj1289 SAXS profile in solution, a ‘closed’ state resembling the 3RGC crystal structure, and an ‘open’ conformation, with the latter predominating in solution at a 2:1 ratio. Potentially these two states represent the two main states for the chaperone function of Cj1289, with the predominant ‘open’ state resembling an open spring clamp/ C-clamp and allowing access to the three helices containing multiple charged residues and aromatic residues, and the ‘closed’ state preventing access. Relating these states to residue conservation, a ConSurf analysis (figure 5.17) revealed several conserved residues along three helices that may not be accessible in the ‘closed’ state and become available in the ‘open’ state, the two N-terminal helices, and a highly conserved (DPRLLGLLS) helix positioned below the three β-strands that collapse to generate the ‘open’ conformation. The conserved residues highlight a potential mechanism for the opening and closing of Cj1289 involving deformation of the β-strands above the ‘DPRLLGLLS’ conserved helix, movement of the N-terminal helices towards the proximal helix, and the movement and formation of a kink in the C-terminal helix.

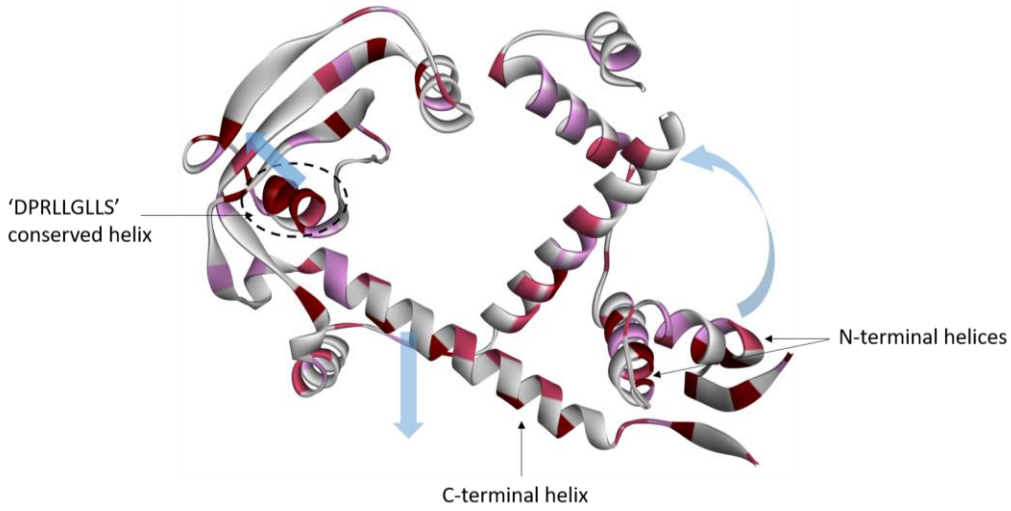


Figure 5.17: ConSurf analysis of Cj1289 conserved residues. The ConSurf webserver was used to identify evolutionary conservation of residues. Residues with a conservation score of ‘7’ (pink), ‘8’ (Indian red), and ‘9’ (dark red) have been recoloured on the 3RGC model for visualisation. Dotted circle denotes the ‘DPRLLGLLS’ residue sequence and blue arrows indicate movement of helices from the ‘closed’ to ‘open’ conformation of Cj1289.

The obtained Guinier fit after trimming for Cj1289 ($0.1 \leq (q \cdot Rg)^2 \leq 1.7$) allowed for further modelling and lead to a density model that looks like a ratchet or spring clamp. When the homology model and 3RGC crystal structure model are overlaid initially onto the bead model, there is a reasonably good fit except for coverage of the density in the arm-like protrusion. This missing coverage of the density model was initially thought to be due to the TEV-10His N-terminal tag that remained on the protein (not included in the MD annealed models), and due to the compact crystal structure model used, as the annealed model which has been relaxed (18 ns) appears to fill the density envelope space better (figure 5.18). An alternative explanation is the envelope density output is due to a combination of the 3RGC ('closed') and 52 ns MD ('open') conformations, as the 'open' conformation is slightly narrow for the main section of the density but fills the arm-like projection density well, and the 'closed' fits the main section well but does not account for the arm-like projection. Overall, both the 3RGC and annealed models fit the experimentally derived density well, but it is better explained by combination of 'closed' and 'open' conformations.

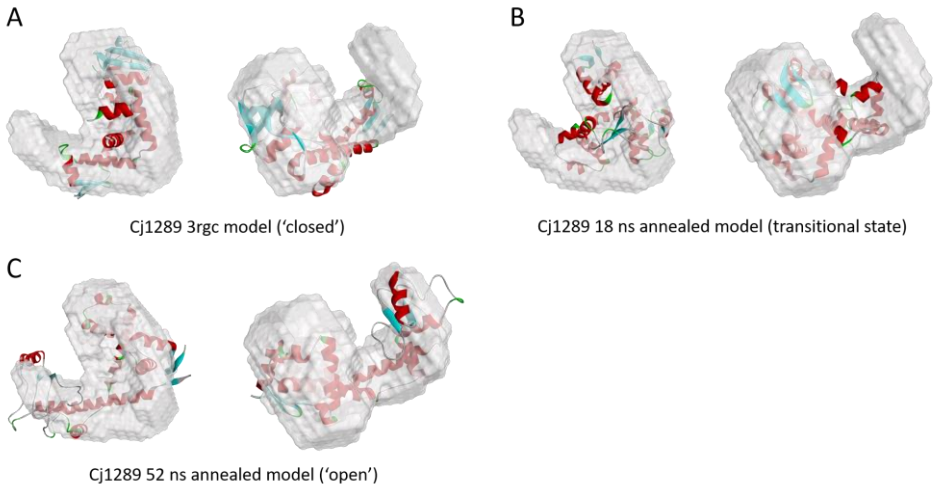


Figure 5.18: Cj1289 DAMMIN model fitting. Side-on and top view manual overlays of Cj1289 (A) 3RGC crystal structure (rigid, 'closed'), (B) 3RGC 18 ns MD annealed (potential transitional state), (C) 3RGC 52 ns MD annealed ('open') models onto the GAJOE envelope output. The 'closed' model (A) and 18 ns annealed models (B) fill the main section of the envelope density well, but do not account well for the arm-like projection density. The 'open' (C) model accounts for the arm-like projection density and provides a good fit to the density overall, but leaves some gaps in the main section of the envelope. Likely the experimental density is due to a combination of 'open' and 'closed' Cj1289 monomer states.

Analysis of the LptD SAXS profile indicates a multi-domain mostly rigid folded protein with some flexible components. This is a reasonable fit to the expectations for LptD as it is composed of a β-barrel domain and a periplasmic domain, and should have relatively stable structure in a detergent/ artificial membrane environment overall with some flexible loops. The available homology model is a poor fit to the experimental data, as shown by the SAXS profile comparison where only the mid-q region is fit well when the hydrated FC12 assembled model is used. The severe aggregation in the Guinier region makes any further analysis difficult.

The DegP SAXS profile obtained suggests a single domain folded protein with a mixture of rigid and flexible components. It was expected that DegP would be a mixture of flexible and rigid components as the homology model (confirmed by the recent Cryo-TEM structure 6z05[110]) shows a series of small globular α/β domains connected by long flexible/disordered linker sequences, so this fits well with expectations. The shoulder observed on the hyperbolic rise in the Kratky plot may indicate the presence of another oligomeric species. If the latter, this indicates potential heterogeneity in the sample or that DegP can transition between states in the experimental conditions used or generally in solution. Multiple models of different oligomeric states were fit to the experimental data, and an approximation of the *E. coli* hexamer structure[203] (rebuilt with *C. jejuni* DegP monomers in Discovery studio from deposited single chain PDB structures, assembly not made available on RCSB) fits the experimental data curve over a longer q-range overall than the monomer and other models. This mirrors the evidence of a hexamer resting state conformation in *E. coli*, and is in line with pre-experiment expectations. The severe aggregation in the Guinier region makes further analysis difficult.

The BamD and Cj1289 mixture showed a single domain structure composed of rigid and flexible components on the standard plots similar to the individual components. At first glance, the BamD-Cj1289 dummy atom model looked similar to the Cj1289 model with a clamp like appearance, but with an extra rod-shaped extension tacked onto the end of the arm-like protrusion observed previously. The BamD homology model fit well into this density and with some straightforward rotational optimisation of the Cj1289 model, these two individual models placed together (in Discovery studio) fit the envelope density well, with only a small area between the proteins not covered. This missing coverage is likely due to conformational differences between

the rigid and compact structural models used (3RGC, and BamD homology model) and their relaxed or alternative conformations in solution, local structural rearrangements post-interaction, and because of the TEV-His tags that were left on the proteins (due to time restrictions). The dummy atom model looks slightly too small for BamD as the helices protrude either side of the model space, but the overall fit is good, and this is likely due to conformational or local rearrangements of BamD (note conformational differences in models in 5.8(c) and (d)). This assembled orientation of BamD-Cj1289 fits the SAXS experimental data reasonably well, with a deviation at mid-q on both the ScÅtter overlay of the data (figure 5.8(e)) and from the FoxS server interpretation (figure 5.13(a)). The R_g of the manual Guinier fitting (26.8 Å), FoxS R_g (27.94 Å), and ScÅtter auto-R_g of the homology model used for the fitting (27.7Å) are in relatively good agreement. The FoxS server returned a very good χ^2 score (0.93) and atomic radii scaling parameter score (C_1, 1.02) well within allowable limits, but the hydration shell density parameter score (-2.00) suggests that the hydration layer around the molecule was of lower density than the bulk solvent. This may or may not occur dependent on surface charge, could be due to inaccuracies in the available models, or may just be reflective of the external surface properties of the proteins. Overall, the fit of the roughly assembled model to the experimental data is good, and suggests that Cj1289 interacts with BamD, which *in vivo* would likely be for the transport of BamD from SecYEG to the inner leaflet of the OM.

The majority of the protein: protein mixture samples in this study were unable to be analysed due to varying degrees of aggregation present in the individual protein profiles (BamA, DegP, LptD) which carried over to the mixture, or a lack of suitable models (BamA, DegP). While some of these issues were caused by (Covid-related) equipment unavailability and abrupt re-scheduling of the experiments, the issues with aggregation indicate changes to final steps in the purification strategy and preparation of the samples are required prior to repetition of experiments.

The complexation analysis corroborates *C. jejuni* BamA and BamD interaction, which agrees with the expectation from the literature and with the increased thermal stabilisation of the sample mixture relative to the individual proteins observed in the CD experiments.

However, the complexation analysis disagrees with the CD data on whether BamA and Cj1289 interact, with no interaction indicated in the CD data and interaction indicated in the SAXS data. Unfortunately, due to the severe aggregation observed in the BamA-Cj1289 mixture (data not shown), a dummy atom model could not be produced to help support or refute the interaction, and therefore requires further investigation with non-aggregated monodisperse samples (pending dynamic light scattering or equivalent instrument access post-Covid restrictions).

In addition, the complexation analysis predicts an interaction between BamA and DegP, and BamA and LptD. The BamA and DegP has not been elaborated on in the literature, but is possible due to its dual function as a chaperone and as a protease. As this was the first experiment with limited sample quantities available, and the samples were not able to be recovered post-experiment, it is unknown whether the chaperone or protease function resulted in this interaction. The BamA and LptD interaction has been posited in the literature in *E. coli*, so the interaction indicated by the complexation analysis was expected, although whether this occurs directly or via a chaperone interaction is unknown. Interactions between both chaperones and LptD are indicated by the analysis, which is expected as chaperones may be involved in the transport of LptD and therefore would be likely to interact with LptD prior to insertion and assembly by BAM at the membrane.

A potential contradictory result of the complexation analysis is that it does not indicate and interaction between BamD and Cj1289. There is a potential change in the residuals around q = 0.14 (figure 5.15(a)), and the residuals appear to skew downwards at high-q, but these changes are subtle and can be easily over-interpreted, and the rest of the q-range shows a linear and homogeneous distribution of points, so the result has been marked as a non-interaction. An alternative explanation is that the conformation of the proteins when in complex are only subtly different, as evidenced by the roughly assembled BamD-Cj1289 model fitting the experimental profile well during cross-validation (figures 5.8(e) and 5.13(a)), and so only subtle change would be apparent in the complexation analysis. However, the density model fitting provides a strong indication that these proteins do interact, and supports the interaction indicated by increases in thermal stability in the CD data. Further investigation with higher concentration samples or using a different technique may be required to deconvolve this discrepancy.

In addition, the complexation analysis indicates a lack of interaction between BamD and LptD, BamD and both chaperones, and between both chaperones. The lack of interaction between chaperones is expected as it would likely not serve any biological purpose, unless they are initially required to fold one another, but this has not been shown in the literature. As we have no data of a triple mixture between BamD, Cj1289 and a target OMP, this will require further investigation as it is in contradiction to the literature on BamD interaction with incoming OMPs[107, 204]. The lack of interaction with LptD, however, is not concerning as the complexation analysis suggests BamA interacts with LptD, so BamD may not be required for the transfer of LptD between chaperones and the BAM complex.

5.5 Summary

The series of experiments detailed in this chapter was the first data collection of these samples. Basic characterisation was carried out in the majority of samples, with P(r) being determined for Cj1289 and the BamD-Cj1289 mixture. Dummy atom models were subsequently built for Cj1289 and the BamD-Cj1289 mixture, with both models showing good or excellent cross-validation when simulated SAXS profiles generated from homology models were compared to the experimental data. Further cross-validation using an independent software, FoxS, via its webserver corroborated this cross-validation and confirmed the fit and suitability of the available structural models to explain the experimental data for both Cj1289 and the BamD-Cj1289 mixture. Cj1289 was determined to be present as a monomer in solution, contrary to the dimer presentation observed in the 3RGC Cj1289 crystal structure model. Additional analysis with GAJOE determined that the structural state of the monomer exists in two major conformations in solution, a ‘closed’ conformation that potentially prevents access to the inner conserved surfaces, and a previously unknown ‘open’ conformation derived from MD data generated in this study that allows access to these surfaces. Further analysis of the experimentally derived Cj1289 density suggests that the density is explained well by a combination of ‘open’ and ‘closed’ Cj1289 monomers. A potential model mechanism for the transition between these conformational states was proposed based on conserved residues identified by a ConSurf analysis. A roughly assembled BamD-Cj1289 interaction model cross-validated well and the good density fitting supports an

interaction between BamD and Cj1289. This interaction would fit in the literature as a transport mechanism for BamD from SecYEG to the inner leaflet of the OM.

Unfortunately, it was not possible to generate dummy atom models for BamA, LptD and DegP, and mixtures containing these proteins due to varying degrees of aggregation, and for BamD due to insufficient signal intensity. In addition, it was not possible with the available BamA and LptD models to fit the experimental scattering data well, and better models in the correct environment with annealing need to be generated for BamA and LptD to make a proper assessment on the experimental data. On a similar theme, although the roughly assembled model fit the experimental data well, it would be beneficial to have a MD generated model for the BamD-Cj1289 interaction for fitting purposes.

However, The BamD-Cj1289 model requires additional validation as the complexation analysis indicates a lack of interaction between the proteins, but comparisons of homology models SAXS profiles to the experimental data show a good fit to the data curve. The complexation analysis indicates interaction between BamA and all other proteins in this study, and both chaperones with LptD. The analysis indicates a lack of interaction between BamD and LptD, BamD and both chaperones, and between both chaperones.

Unfortunately, limited analysis of other samples could be carried out due to aggregation in the sample sets; however, this data serves as good preliminary data set for future experiments.

6 Chapter 6: *In silico* investigation of the *C. jejuni* BamA-BamD docking interface using rigid body docking and Debye-length modulated molecular dynamics simulations

6.1 Chapter aims

In the previous chapter, we used small angle scattering to assess interactions of the proteins in this study and built bead models for Cj1289 and the BamD-Cj1289 mixture. The *C. jejuni* BamA-BamD mixture did not result in the generation of a bead model, however the interaction between BamA and BamD was indicated by the complexation analysis at mid-q (size, structure) corroborating the CD inference and the interaction can be modelled using *in silico* approaches.

In this chapter we build homology models of BamA and BamD, analyse the models available from *E. coli* to approximate the residues involved in the *C. jejuni* BamA-BamD docking site, dock the proteins with and without these restraints using existing docking programs, and perform an unbiased MD simulations-based docking assay as a novel alternative approach to traditional rigid body docking.

In this set of experiments, we aim to:

1) Analyse existing *E. coli* experimental structural models of the BAM complex to assess which residues contribute to the binding of BamA and BamD
2) Generate homology models for *C. jejuni* BamA (CjBamA) and BamD (CjBamD)
3) Compare the homology models to existing structures and determine which residues are relevant for the docking interaction in the homology models.
    a. Use these identified residues as restraint inputs for rigid body docking
        i. Assess the unbiased and restrained input generated models for feasibility and pick appropriate model outputs
        ii. Revise docking site restraints (as required)
    b. Estimate the relative positions of BamA and BamD prior to docking

        i. Use this approximation to build a combined PDB model
4) Prepare the combined PDB model for MD simulations using the CHARMM-GUI web interface
5) Perform a series of orientationally unbiased MD simulations to assess to whether the rigid body docked outputs can be replicated by these simulations
    a. Assess the feasibility and conditions required for this to occur, and whether it will occur in a reasonable timeframe
        i. Can this process be expedited by modulation of Debye layer properties
    b. Whether the orientation of the models needs to be known for the docking to occur

6.2 Introduction

From the previous chapters we have acquired data that suggest that *C. jejuni* BamA and BamD interact. This was the expected result as the literature on homologous systems show a BamA and BamD interaction. To assess how these proteins interact we attempted to use SAXS to build a bead model in the previous chapter, however the data showed aggregation and this inhibited the bead modelling building process for the BamA-BamD interaction.

An alternative to building a model from experimental data is to use an *in silico* approach, where models representing the proteins are generated through homology modelling using known homologous structures as templates, and docking these protein models using current and established rigid-body docking approaches.

The homology models will be generated using I-TASSER and cross-validated with Phyre2. To determine which residues contribute to the *C. jejuni* BamAD docking interaction, docked experimentally determined *E. coli* structural models of the BAM complex available on RCSB (5AYW[9], 5D0O[8], 5D0Q[8], 5LJO[44]) will be screened for residues in contact with a distance restraint of 5 Å[205]. Using the derived docking interface residue output from *E. coli*, we superimpose the *C. jejuni* models onto the *E. coli* experimental structures (5LJO) to make an approximation of the residues involved in the docking interaction. The rigid body docking can be carried out by a number of pre-existing programs, and in this study we utilise ZDock[206], pyDock[207], and Rosetta (Docking 2)[144] as these will perform *ab initio* docking with no prior knowledge on the docking site, but have the option to bias the docking procedure with derived residue restraints.

Following on from the rigid docking we explore the use of MD simulations generated via NAMD to simulate the docking event from multiple orientations in an unbiased manner to deconvolve the multiple rigid docking states. In this chapter we aim to probe which residues are involved in docking in the *C. jejuni* BamAD interaction, and run a feasibility study of using MD simulations for this process utilising a modulated Debye length during system set up to minimise computation cost.

6.3 Results

6.3.1 Determining the docking contact interface

Prior to commencing the docking runs, the crystal and Cryo-TEM structures of the *E. coli* BAM complex were examined to better understand the docking interface and to generate an approximation of residues involved in the *C. jejuni* BamA and BamD interact. The *E. coli* 5AYW, 5D0O, 5D0Q crystal structures, and 5LJO Cryo-TEM structures were obtained in PDB format from RCSB (figure 6.1(a)), and BamB, BamC, BamE and any repeat chains were removed prior to analysis of the docking interface.

The binding interface for *E. coli* BamAD was examined in UCSF Chimera[141] using a 5 Å distance restraint for all four *E. coli* experimental structural models and identified residues and their occurrence were tabulated and represented graphically (appendix 8.6.1, figure 8.25). The identified binding interface of BamA is composed of POTRA 1 (β-strand 1 - helix 1 loop and helix 1, helix 2 - β-strand 1 loop), POTRA 2 (β-strand 1, helix 1-2 loop, β-turn and β-strand 2-3 loop), POTRA 5 (β-strand 1, helix 1, helix 1-2 loop), and the β-strand 2-3 and 4-5 loop of the β-barrel. The binding interface of BamD is composed of the start of helix 1, end of helix 2, the helix 2-3 loop, start of helix 3, helix 4, the turn and loop between helix 5 and 6, second half of helix 7 and helix 8.

Homology models of *C. jejuni* BamA and BamD were generated using I-TASSER[118], selecting the model with the best confidence (C-score) parameter for BamA (figure 6.1(b)) and BamD (figure 6.1(c)). The BamA and BamD homology models appear to have a similar overall structure to the *E. coli* experimental structures, differing predominantly by sequence length (BamA, 803 *vs.* 739 residues; BamD, 243 *vs.* 215 residues). To determine the equivalent residues in contact between *C. jejuni* BamA and BamD a molecular overlay function was utilised to superimpose the models (*C. jejuni* BamA to 5LJO chain A and *C. jejuni* BamD to 5LJO chain D) and approximate residues involved in the *C. jejuni* BamAD interface based on structural similarity, local proximity and residue chemistry. Derived values were tabulated and represented graphically (appendix 8.6.1, figure 8.26)). The residues identified were then used as restraints for the rigid body docking process.

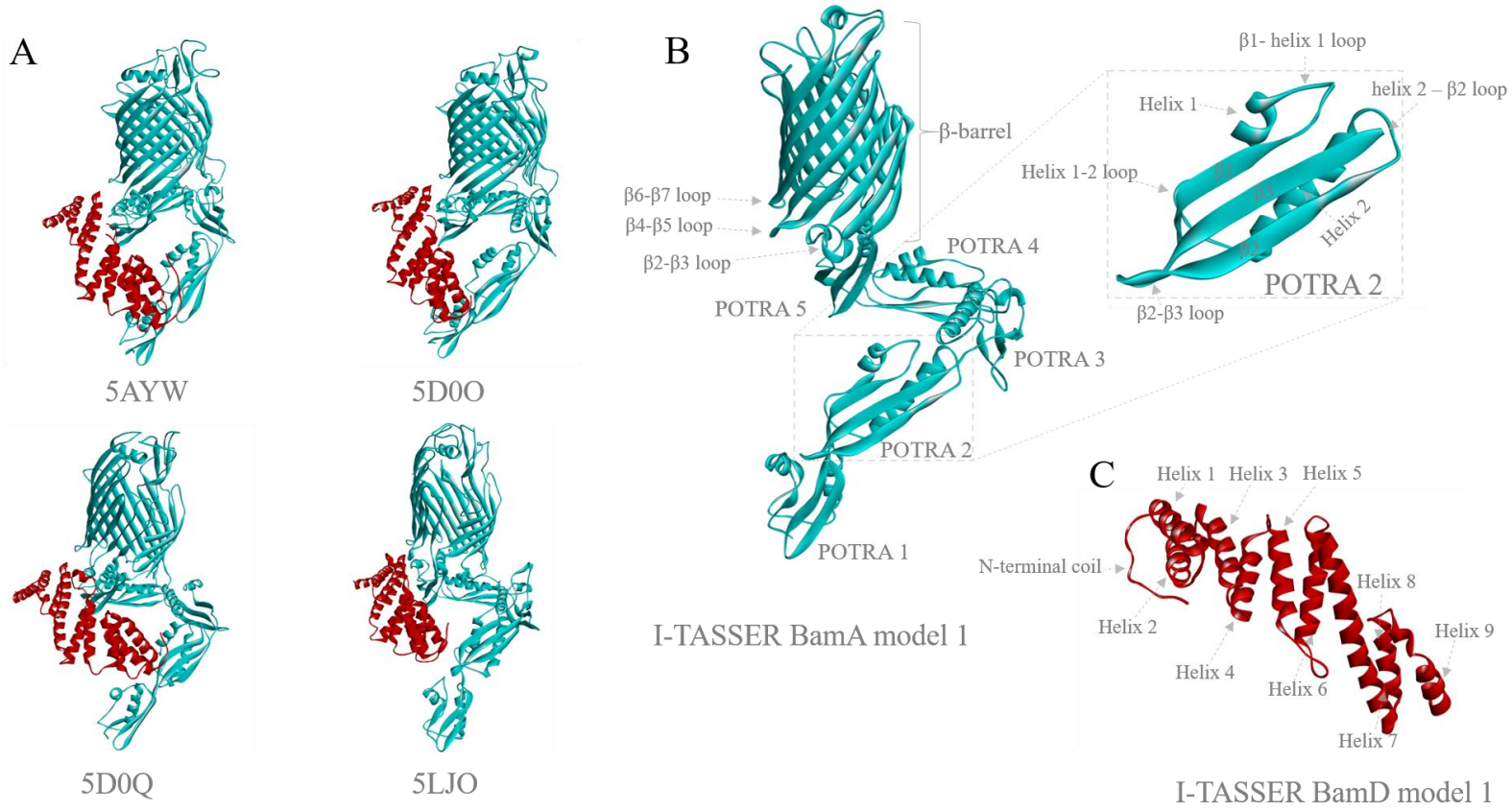


Figure 6.1: *E. coli* models and *C. jejuni* I-TASSER homology model output.The *E. coli* structural models used to determine residue contacts at the BamA-BamD docking interface (A), and the structurally annotated *C. jejuni* I-TASSER homology models for BamA (B), and BamD (C) are shown. Note that all structures of BamA are shown in light blue and all structures of BamD are shown in red.

6.3.2 Rigid body docking

Three web services were used for rigid body docking, ZDock (ver. 3.0.2), pyDock(Web), and RosettaDock (Docking 2). These services were chosen as they require no prior knowledge of the docking interface, but accept restraints or approximate orientations if known. In both ZDock and pyDock, these residues were used as docking restraints, and for RosettaDock (Docking 2) a combined PDB model with separated protein chains and chain identifiers (hereafter named 'A0D0') was generated in Discovery Studio for a local protein-protein docking run. Input PDBs are shown for ZDock and pyDock (figure 6.2(a)) and RosettaDock (figure 6.2(d)).

6.3.2.1 ZDock rigid body docking runs

The previously identified residue restraints for binding were used to set up two rigid body docking runs. In the first docking run they were identified as 'binding site residue' restraints, and in the second run, all residues other than the identified residue restraints were blocked. The first docking run resulted in the top 100 outputs mostly consisting of non-permissible or inappropriate binding states, where some part of BamD was against the β-barrel of BamA where the membrane would normally be found. Of the few permissible docking states, the majority showed single binding sites and less than five results where the binding site occupied was similar to the *E. coli* crystal and Cryo-TEM structure presentation. The second run had far fewer of these inappropriate configurations, however, the top 10 results (appendix 8.6.2, figure 8.27) showed only a singular binding point, inconsistent with the expectations from the *E. coli* models. After screening the top 100 results, two models were chosen that approximately matched the pre-experiment expectations (figure 6.2(b)). Complex 31 showed binding at the approximate regions shown in the other *E. coli* template models, however, helix 7 was pointed down and away from POTRA5, suggesting an inward rotation of approximately 180° of BamD on its axial plane that changes which part of helix 7 contacts POTRA5. POTRA 1 to helix 4 (BamD) and POTRA 2 to helix 2 (BamD) composed the second binding site interaction. As a similar region of BamA was bound, this binding position may be a variation to the binding configuration observed in the *E. coli* models.

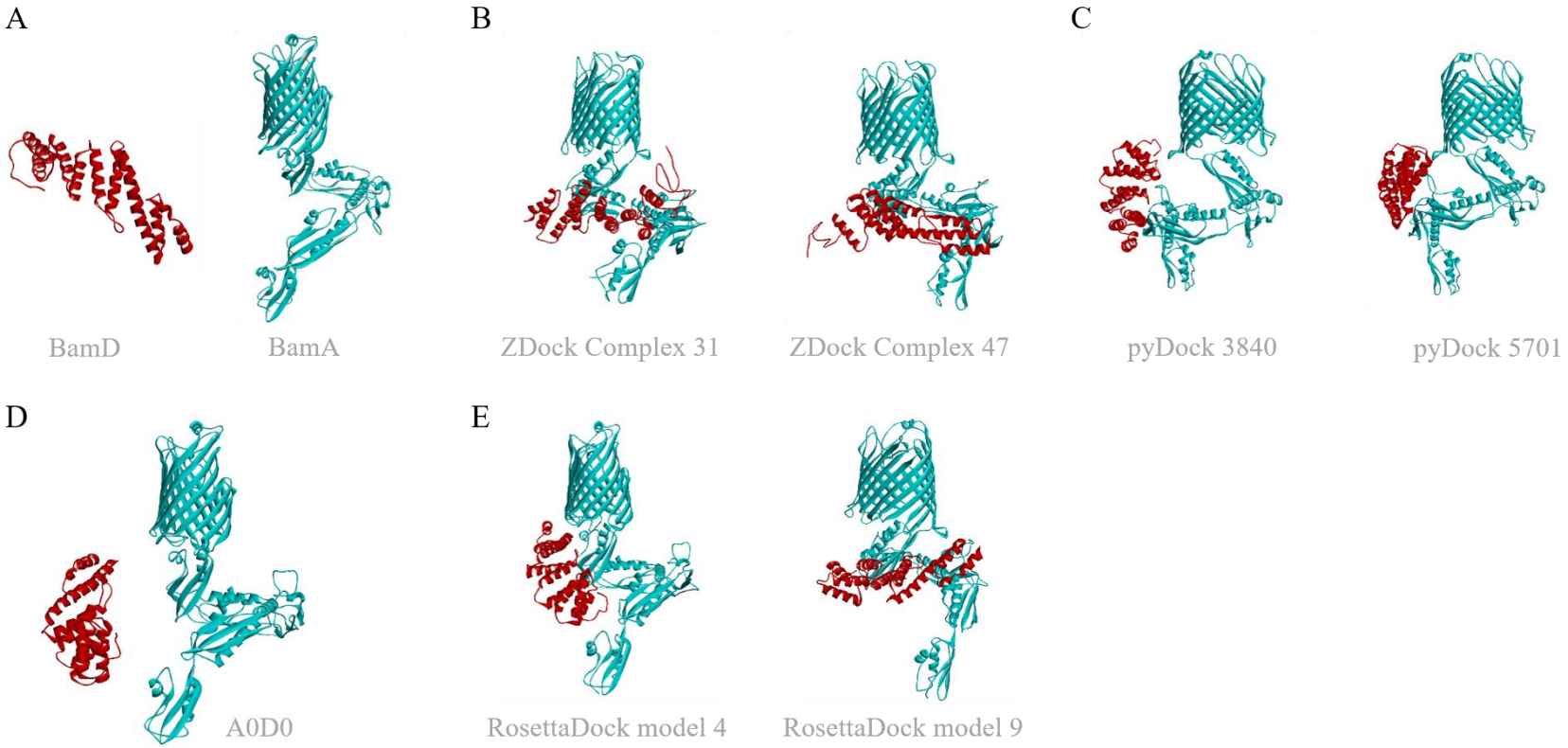


Figure 6.2: Rigid body docking input and output models. The homology model structures for *C. jejuni* BamA (light blue) and BamD (red) used as input structures for rigid body docking for ZDock and pyDock (A), the two selected docked structures generated using restraint inputs in ZDock (B) and pyDock (C), the combined PDB input model (D), and the two selected docked structures generated by RosettaDock (E) are shown. All rigid body docking programs are able to generate viable docked models that resemble the available *E. coli* experimental structures given restraints derived from superimposition (ZDock, pyDock) or the approximated docking orientation (RosettaDock).

Complex 47 was another which bound to the approximately correct position on BamA, however, in this case BamD was rotated 180° clockwise from the frontal viewpoint, as if BamD was flipped over. The contact interface is complex of helix 2 and 4 (BamD) with POTRA 5 and helix 7 (BamD) with POTRA 1. While this does not match the expected output from the *E. coli* models, as BamD binds BamA at the approximately correct areas and BamB, BamC and BamE are not present in *C. jejuni* as restraints, this model should be taken into consideration. The residues in contact (within 4 Å) are tabulated (appendix 8.6.2, table 8.9) alongside the other rigid body docking outputs.

6.3.2.2 **PyDock rigid body docking runs**

The first pyDock(Web) rigid body docking was carried out with no restraints similar to the first ZDock run, and resulted in a similar output, with the majority of models showing BamD being bound in an impermissible configuration approximately midway up the BamA β-barrel (appendix 8.6.2, figure 8.28(b)). The two most suitable models available among the top 100 results were chosen (appendix 8.6.2, figure 8.28 (c)) to be reviewed. The first model, 8196 (complex 81), showed a similar (but slightly high) binding position to the expected outcome, and the second model, 8238 (complex 82), shows a similar contact interface to the flipped BamD orientation, but both models protrude on where the membrane would normally be placed and were inviable. The second pyDockWeb run utilised the previously identified residues as distance restraints (which the service defined as within 6 Å), and resulted in more viable docking configurations. Two models, output 3840 (complex 38) and 5701 (complex 57), were chosen to be reviewed after manually screening the 100 outputs (figure 6.2(c)). Output 3840 shows a 'flipped' orientation of CjBamD relative to the expected configuration and a contact interface between the β2-β3 loop (BamA) and helix 1 (BamD), POTRA 3 helix 1-2 loop (BamA) to the helix 5-6 loop (BamD), and between POTRA 3 β-sheet and helix 7 (BamD). Output 5701 shows a contact interface between the β2-β3 loop (BamA) and the helix 2-3 loop/ helix 4 (BamD), and several helices of BamD with POTRA 2 and 3 of BamA. While this model also shows a flipped over orientation of BamD it is 180° rotated from the former model with the opposite surface of helix 7 forming part of the contact interface. The residues in contact for outputs 3840 and 5701 are listed (appendix 8.6.2, table 8.9).

6.3.2.3 **RosettaDock rigid body and local refinement docking runs**

The final rigid body docking run was carried out with the Rosetta Docking 2 web service, and differs from the previous rigid body docking runs as a combined PDB must be used as an input compared to individual PDBs of each protein for the former two. A combined PDB was set up with the relative orientations of CjBamA and CjBamD resembling the *E. coli* experimental structure configuration, but with proteins separated by 13.5 Å, and was used as the input file (figure 6.2(d)) to RosettaDock. This input model (termed 'A0D0') was later used in the MD docking runs. The top 10 outputs (appendix 8.6.2, figure 8.29(c)) showed eight models with permissible configurations, of which model 4 and 9 replicated the two major contact regions observed across the *E. coli* structures and approximated the expected binding position well. Model 4 shows the expected contact interface, as observed in the *E. coli* experimental structures, with interaction between POTRA 5 (BamA) and helix 7 (BamD) on one site, and POTRA 1-2 (BamA) with helix 1-2 (BamD). Model 9 shows another example of the 'flipped' BamD structure, with a contact interface between POTRA 5 and helix 3-4 (BamD), and POTRA 2 with helices 7 and 9 (BamD). Both structures appear viable and their total scores (appendix 8.6.2. figure8.29(b)) are similar. The residues in contact for models 4 and 9 within 4 Å are tabulated (appendix 8.6.2, table 8.9).

The docking results show multiple orientations of CjBamD relative to CjBamA docking, with two viable configurations docking to similar overall areas of CjBamA. As the *E. coli* experimental structures are of five component models, the other components may restrict the potential binding configurations. This restriction does not apply to our two-component model where a greater surface area is available for binding, so these alternative orientations of BamD (axially rotated and 'flipped') relative to BamA should be considered. To investigate whether a docking event leading to the alternative configurations being stabilised will occur when other options are available, a systematic unbiased MD based docking methodology was developed in which the relative orientation of BamD to BamA is adjusted so different surfaces are sampled for docking.

6.3.3 Molecular dynamics-based docking runs

Current established rigid body docking methodologies utilise energy values and (optionally) residue restraints to determine correct docking configurations, but the models undergoing docking stay static throughout the process[208]. The premise behind this methodology is that MD simulations takes energies, electrostatics, bond parameters, solvent, lipid surfaces, conformational changes and other factors into account[209], and so should be able to resolve whether these interactions are permissible and may have several advantages and be more reliable over rigid body docking. The clear disadvantages of this approach are the computation cost[210], that the Debye length or molecular shielding may prevent the molecules from interacting if they are placed too far apart[211] or may prevent rearrangements or conformational sampling before a docking position is reached if the molecules are placed too closely together. In addition, the process is more time-consuming, requiring several further steps and requires knowledge of MD simulations. So, to allow for the interactions of BamA and BamD to occur over a longer distance the Debye length was modulated by changing the salt concentration modelled in the system set up from physiological (0.15 M) to 0.05 M KCl, changing the effective Debye length from 7.82 Å to 13.55 Å. The molecules were placed (at least partially) within this distance relative to one another to enable interaction. The salt concentration was not decreased further as the distance calculation length parameter for long-range electrostatics in the MD simulation had to be considered as well as the increased cost for simulation of an expanded unit cell. BamD was placed between 12.5 Å - 13 Å to BamA in all the sampled starting configurations, with the exception of the control run (A0D0 normal) where placement had to be within 7.8 Å due to using physiological salt conditions.

The rigid docking results suggest that there may be alternative configurations for the docking of BamA and BamD. An unbiased systematic approach was adopted wherein the initial orientation (A0D0), which represents the expected binding configuration from available structures, was iteratively adjusted by either rotating BamD relative to BamA or vice versa. BamA was sequentially rotated 90° clockwise around the y-axis, giving four potential orientations at 0° (A0), 90° (A90), 180° (A180), and 270° (A270). In tandem with this, the orientation of BamD was either in the initial expected orientation (D0) or adjusted by 180° rotation through a central

point around the x-axis (D1, flipped orientation), or by 180° rotation on its axial plane (D2, axially rotated orientation). This gave rise to twelve potential docking orientations which were tested and a control condition using an adjusted A0D0 starting position under physiological salt conditions. These translations are shown graphically in figure 6.3.

The thirteen combined PDB models of BamA and BamD were set up (Discovery Studio) and assigned chain identifiers (PDB file editor), the models were assembled into the solvated (80:20 POPE:POPG[101, 102]) membrane bilayer systems using the CHARMM-GUI[119, 121] membrane builder (bilayer builder) web interface using NPT ensemble at 303.15 K with configuration for execution using NAMD[212].

Prior to the production runs, the systems were equilibrated and MD scaling runs were carried out to determine a cost efficient core allocation for the system size. The MD scaling runs were normalised by the number of atoms in the systems and graphed as cores per s/ atom/ ns (appendix 8.6.3). In all runs, a 504 core allocation showed the best compromise between cost efficiency while remaining within the maximum wall clock allowed for all production runs on the Midland-Plus-HPC server. After the completion of the production runs, all systems were analysed using VMD[213] (*ver*. . 1.9.3) for calculation of the root mean square deviation (RMSD) and root mean square fluctuation (RMSF) and UCSF Chimera (ver. 1.15) to assess which residues were in contact after 51 ns (including 1 ns scaling run).

For reference, a list of the structural elements of BamA and BamD *vs.* residue number has been tabulated (appendix 8.6.4) and the starting positions for all MD docking runs are shown (appendix 8.6.5) with the membrane removed. Please note as MD trajectories are calculated and can be outputted at much smaller timescales than nanoseconds, in the following experiments ‘frames’ and nanoseconds are used interchangeably as units, where 10 frames is equivalent to 1 ns.

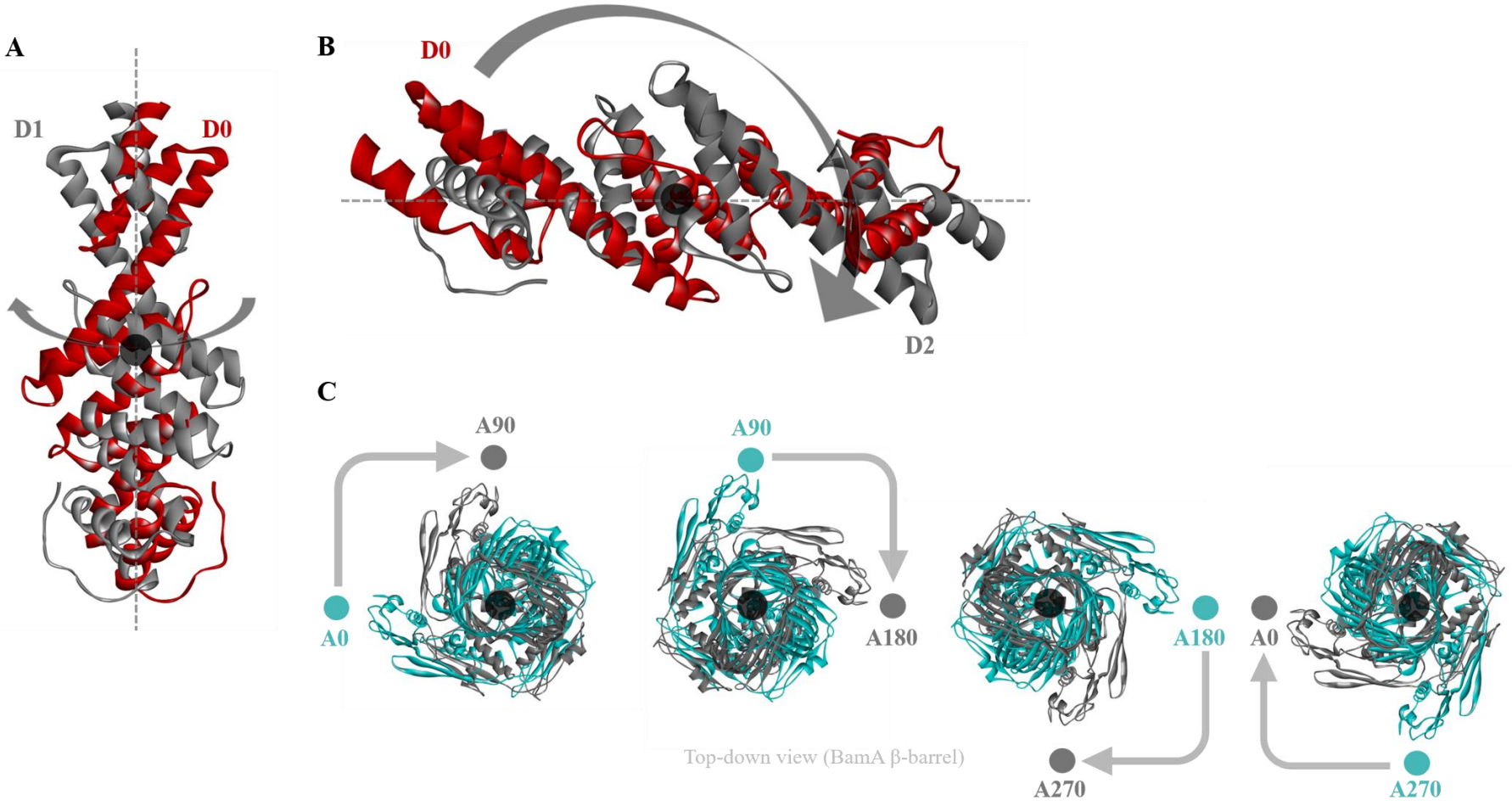


Figure 6.3: MD docking orientation translations. The translations applied to the CjBamD (red, grey [rotated]) and CjBamA (light blue, grey [rotated]) proteins are shown graphically for the D0-D1 axial rotation (A), D0-D2 flipped orientation (B), and all BamA translations (C).

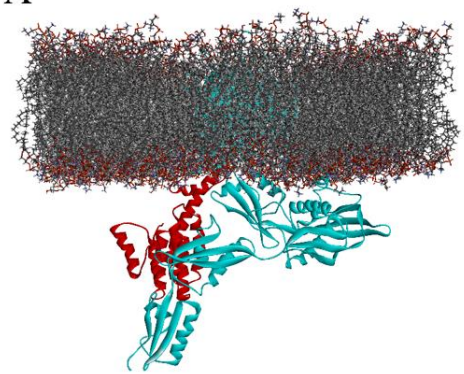
6.3.3.1 **MD Docking run A0D0**

This run started in the expected orientation that would lead to docking (figure 6.4(a)). After the 50 ns production run (figure 6.4(b)), run A0D0 shows a contact interface between BamD helix 1 and 2, helix 1-2 loop, and the N-terminal coil with BamA POTRA 1 helices, and POTRA 2 helix 1 at binding site 1. At binding site 2 there is an interaction between BamD helix 7 and the helix 5-6 loop with the POTRA 4-5 loop/hinge region, and between helix 7 and the β-barrel β6-7 loop. This ending position appears quite similar to the *E. coli* crystal and Cryo-TEM structures, with minor differences such as BamD helix 7 interacting with the β-barrel β6- β7 loop instead of the β4- β5 loop, and the binding position moved slightly closer to the end of POTRA 1 instead of on POTRA 2. This is reflected in the 17 residues of BamA and 19 residues of BamD that match the initial prediction for residues involved in the binding interface (figure 6.4(c) and (d)). The docking run appears to find this docking configuration immediately after the system equilibration, and the 250 frames it takes for the RMSD to stabilise (figure 6.4(e)) is the transition from this initial binding configuration to flattening onto the membrane surface whilst maintaining this configuration. The RMSF (figure 6.4(f)) indicates that the elements that contribute the most to these changes are BamA POTRA 1, β-barrel β7- β8 and β11-12 extracellular loops, and BamD helix 4, helix 5, and helix 7.

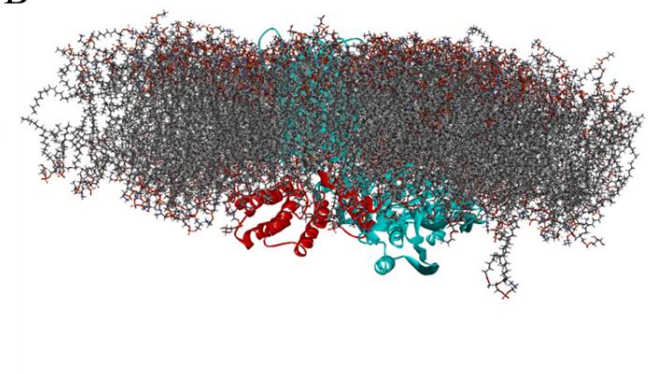
6.3.3.2 **MD Docking run A0D0 normal**

The salt concentration was increased in this run to physiological salt concentration (0.15 M KCl) which required a reduction in separation distance at the start of the run, but otherwise is a replicate of the orientation observed in the A0D0 run (figure 6.5(a)) and serves as a control run. Comparing the end point of this run to the A0D0 MD docking run, a variation on the expected docking configuration is found (figure 6.5(b)), however, it does not occur immediately. The BamD molecule goes through a sampling phase early in the trajectory, during which the BamD molecule rotates approximately 120-180° outward on its axial axis and BamD helices are sampled by POTRA 1 and ‘flicked’ into position (data available on request). This results in a similar binding position, but BamD helix 7 is facing away from the BamA β-barrel causing a pronounced kink to form to mediate the helix 7 to β-barrel β6-β7 loop.

A



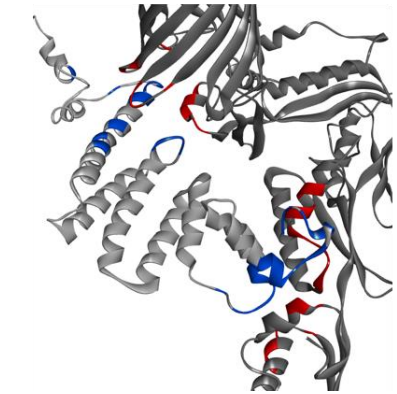
B



C

	Residues in contact (≤ 4 Å)	
Model	CjBamA	CjBamD
A0D0	**32S**, **34T**, **35S**, 58L, 59N, 62K, **63Q**, **65Y**, 102K, 103Q, 105E, 106S, 109G, 110I, **111K**, **112R**, **113G**, **114T**, 115L, **354S**, **355R**, **357V**, **359R**, 360V, **363R**, 471R, 474D, 475S, 476Q, 507Q, **509A**, **510R**, **511Y**, 554N, 555D, 556T	18C, 19S, 20T 21K, 23D, 24E, 25G, 26L, **28N**, **29L**, **30S**, **31A**, 32S, 33E, **34W**, 35Y, 57M, **58A**, **59S**, **60E**, **61H**, **62V**, **63A**, 66L, **120A**, **121V**, **122P**, **123N**, 158K, 161L, 165Y, 172N, **173L**, 175Q, **176R**, **177I**, **181E**, 202K

D



E



F

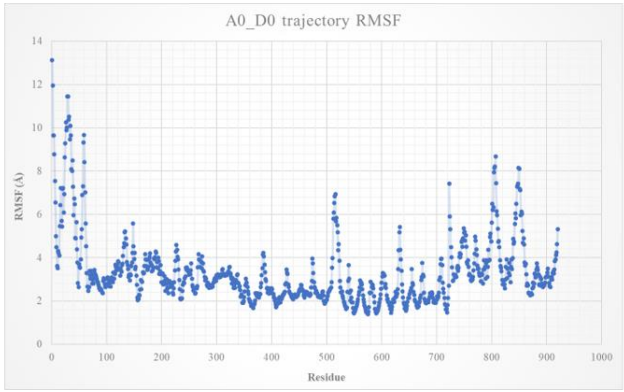
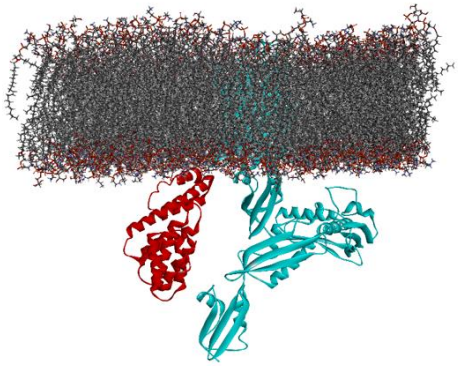
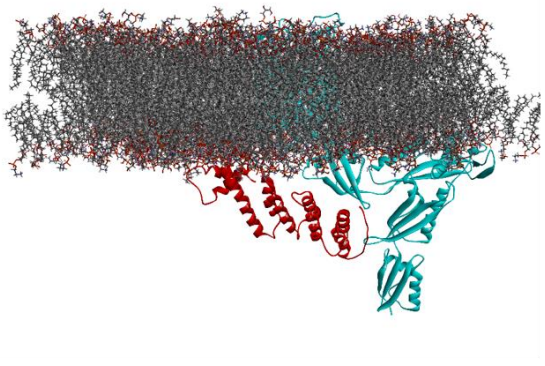


Figure 6.4: A0D0 MD docking run. The A0D0 docking run starting position post-equilibration (A), end position post-50 ns trajectory (B), residues in contact post-run (C), residues in contact post-run [graphic] (D), trajectory RMSD change by frame (E) and RMSF per residue (F) are shown. The A0D0 CjBamA-CjBamD docking run results in a very similar end configuration to the *E. coli* experimental structures.

A



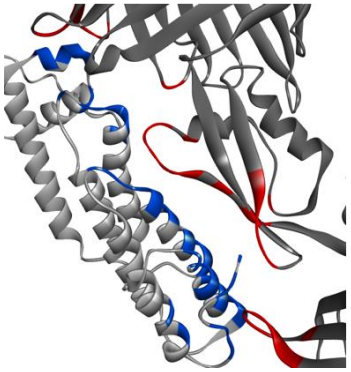
B



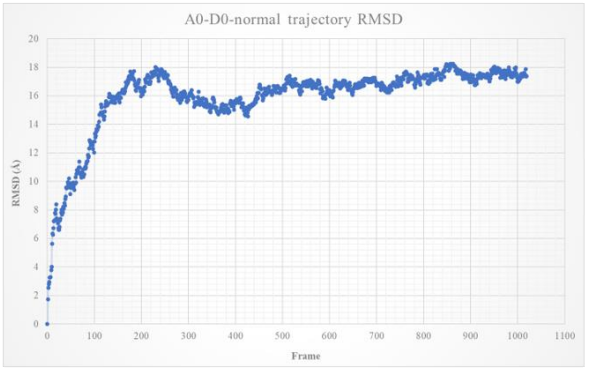
C

Model	Residues in contact (≤ 4 Å)	
	CjBamA	CjBamD
A0D0 normal	147A, 148T, 149L, **150E**, **151N**, **152T**, **153D**, 156E, **346R**, **347N**, **349I**, **351S**, **352G**, **353N**, **354S**, **355R**, **356T**, **358D**, **369E**, 397K, 398E, 399E, 400K, 401V, 407D, 475S, 476Q, 506R, 507Q, 508F, **509A**, **510R**, **511Y**, 559Y, **560Y**	18C, 20T, 21K, 24E, 25G, 26L, **27Y**, **34W**, 37Q, 48E, 49K, 51D, 52D, 53H, 55N, 56G, **59S**, **60E**, 82E, 83E, 84Y, **85K**, **86L**, **89F**, **93E**, **96K**, **97K**, **122P**, **124R**, **125N**, **126Q**, **127A**, **130L**, 172N, **173L**, 175Q, **176R**, **177I**, **178G**, **180D**, **181E**

D



E



F

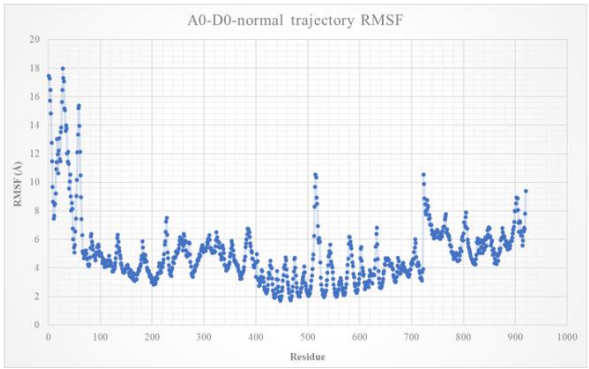


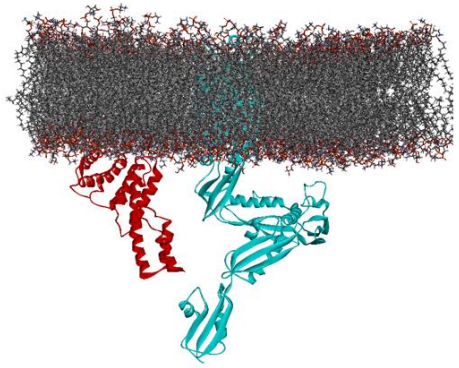
Figure 6.5: A0D0 normal MD docking run. The A0D0 normal (salt concentration) docking run starting position (A), end position post-50 ns trajectory (B), residues in contact post-run [graphic] (C), highlighted residues in contact post-run (D), trajectory RMSD change by frame (E) and RMSF per residue (F) are shown. The A0D0 normal CjBamA-CjBamD docking run results in a similar end configuration to the *E. coli* experimental structures just with CjBamD (red) slightly more underneath CjBamA (light blue).

The BamD helix 6 and helix 5-6 loop interact with β-barrel β4- β5 is also present at binding site 1. The second binding site is composed of an interaction between the N-terminal pre-helix coil, helix 2 and helix 4 with BamA POTRA 2 β-strands, and helix 2 and helix 4 with POTRA 5 β-strands. The POTRA domain are arranged in a Y-shape compared to the L-shape observed in the A0D0 docking run. Overall, the A0D0 normal docking run is more underneath the β-barrel and has less contact with the lipid bilayer inner leaflet. In addition, unlike the former run, the POTRA domain and BamD do not flatten against the membrane. This configuration does not look like the crystal or Cryo-TEM *E. coli* structures, but considering the sampling phase and the approximately correct binding position, it may be a genuine docking configuration. A slightly higher number of the pre-experiment predicted binding residues are also present (BamA, 19 residues; BamD 22 residues; figure 6.5(c) and (d)). There are a twice as many frames available in this run that the other runs, however the position captured at frame 520 (same point as others) is approximately maintained from frame 200 onwards (until frame 1020; figure 6.5(e)). The RMSF (figure 6.5(f)) shows increased movement of the structure overall, with the majority of movement understandably shown by POTRA 1, as well as the β-barrel β6- β7 loop and the BamD N-terminal pre-helix coil as in A0D0.

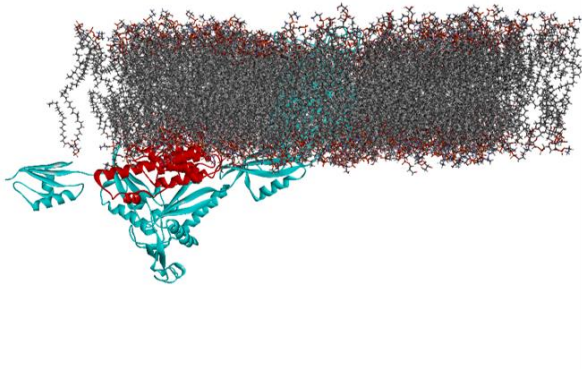
6.3.3.3 **MD Docking run A0D1**

The A0D1 run, where BamD is outwardly rotated on its axial axis (figure 6.6(a)), shows the first example of the extended POTRA domains flattened against the membrane bilayer inner leaflet (figure 6.6(b)). There is a contact interface between BamD and POTRA domain 2-4, with the BamD N-terminal coil, helix 1, helix 2, and the helix 3-4 loop bridging POTRA domain 2-4 to hold this configuration. BamD helix 8 interacting with POTRA 2 and the helix 4-5 loop interacting with POTRA 5 also contribute to stabilising this configuration (figure 6.6(c) and (d)). The surfaces not involved in these interactions appear to be interacting with the membrane inner leaflet, which include BamA POTRA domain 1, 2, and 5 and the unused surfaces of BamD. POTRA domain 3 and 4 are not interacting with the membrane as they are blocked by BamD, which appears to facilitate the interaction of the POTRA domain with the inner leaflet of the membrane acting like a suction cup to the lipid surface (data available on request).

A



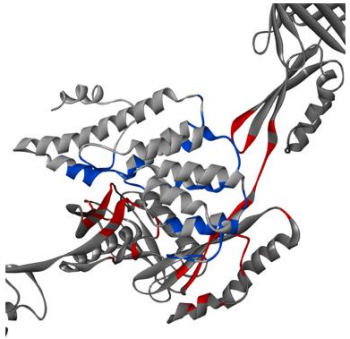
B



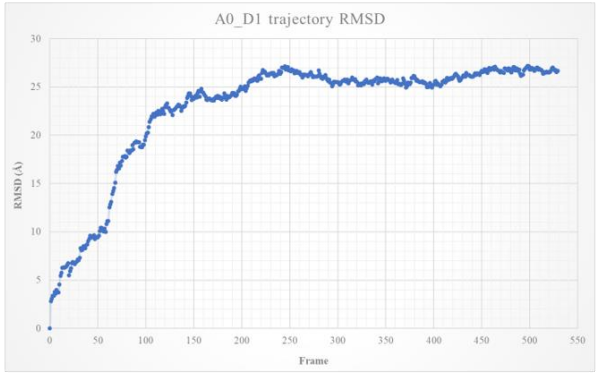
C

	Residues in contact (≤ 4 Å)	
Model	CjBamA	CjBamD
A0D1	90K, 92T, 93I, 94T, 95G, 96I, 99N, 100D, 101R, 102K, 105E, **111K**, **112R**, 139D, 140T, 141I, 143E, 145K, 156E, 158E, 160I, 161V, 162N, 163R, 164G, 165E, 166N, 241T, 242Y, 243T, 264S, 265I, 266S, 267I, 268E, 269N, 270P, 271L, 272F, 273D, 274D, 275K, 280T, 283D, 286S, 313Y, 332T, 336K, 337V, 338I, 339P, 341D, 342K, 343V, 344Y, 346R, 399E, 400K, 401V, 402D, 404T, **405H**	18C, 19S, 20T, 21K, 22N, 23N, 24E, **27Y**, **29L**, **30S**, **31A**, 50A, 51D, 54Y, 55N, 56N, **58A**, **59S**, **60E**, **62V**, **63A**, 67L, 81E, 83E, 84Y, **85K**, **86L**, **89F**, 90Y, **93E**, **96K**, **97K**, **98F**, **99G**, **100N**, 101S, 102R, 103N, 104A, 105D, 108R, **123N**, **124R**, **125N**, **126Q**, **127A**, **128L**, **130L**, 131E, 141K, 142D, 143Y, 145Y, **185I**, **186Y**, 187K, **188Q**, **189R**, 190L, 191Q, **192E**, 193S

D



E



F

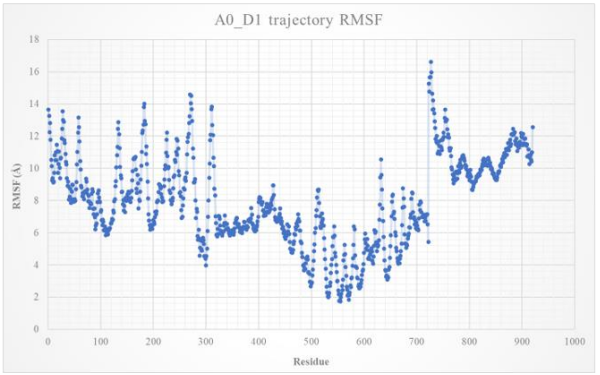


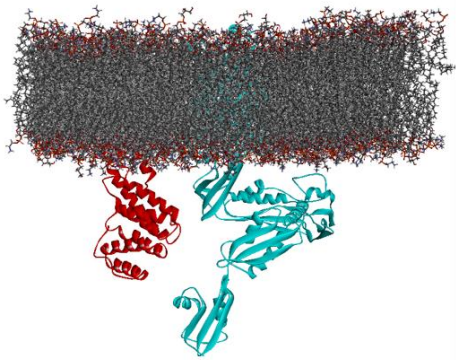
Figure 6.6: A0D1 MD docking run. The A0D1 docking run starting position (A), end position post-50 ns trajectory (B), residues in contact post-run (C), residues in contact post-run [graphic] (D), trajectory RMSD change by frame (E) and RMSF per residue (F) are shown. The A0D1 CjBamA-CjBamD docking run results in a very different end configuration to the *E. coli* experimental structures with CjBamD (red) in contact with the membrane and the CjBamA (light blue) POTRA domain extending and flattening.

The RMSD stabilises around 250 frames (figure 6.6(e)). During this lead up to the stabilisation, BamD flattens against the membrane (50 frames), and POTRA 2 interacts with BamD (100 frames).Following on, BamD moves towards BamA (125 frames), and BamD and BamA move to their final position with POTRA 2 moving around BamD to interact with the membrane causing POTRA 1 to move towards the membrane and flatten (125-300 frames). This position, where BamD stabilises POTRA 3 and 4 allowing extension of the BamA POTRA domain, is maintained until the end of the simulation. This large degree of movement is reflected in the RMSF (figure 6.6(f)), which shows large fluctuations in BamA POTRA domain 1-4 and BamD from the N-terminal coil to helix 2. As this position is vastly different to the *E. coli* structures, different residues form the docking interface than expected for BamA, however, though thirty of the expected BamD residues contribute to this interaction, which might be a reflection of the extent of the BamD surface in contact with BamA.

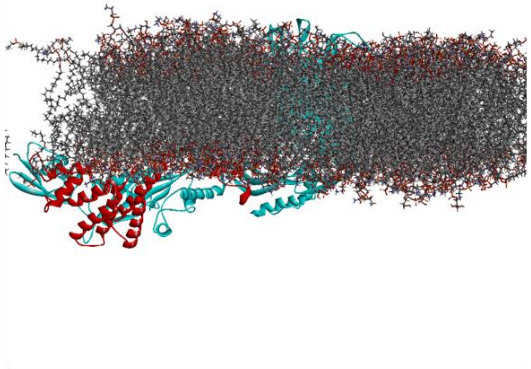
6.3.3.4 **MD Docking run A0D2**

The A0D2 docking run, featuring the 'flipped' orientation of BamD (figure 6.7(a)), shows a second example of the elongated POTRA domain against the membrane inner leaflet presentation, albeit in a slightly different configuration (figure 6.7(b)). The docking interface for this run is between BamD helix 7 and the helix 6-7 loop with BamA POTRA 2, with no other structural elements in contact (figure 6.7(c) and (d)). The RMSD (figure 6.7(e)) takes a comparatively longer time to stabilise compared with the previous runs at 400 frames or 40 ns. The longer transition to a stable RMSD can be explained by viewing the trajectory (data available upon request). BamD is interacting with the membrane at the start of the trajectory by helix 7. BamD then flattens against the membrane inner leaflet (frame 50), BamD and BamA move closer together (frames 50-200) while the BamA POTRA domain are sampling the environment. The BamA POTRA 2 β-strand 2-3 loop interacts with the BamD helix 6-7 loop and helix 7 (frames 200-250), this interaction is then slowly adjusted bringing the POTRA domain into contact with the membrane while they extend past BamD, stabilising the BamA POTRA 2 - BamD helix 7 interaction (400 frames). This configuration is then maintained until the end of the production run.

A



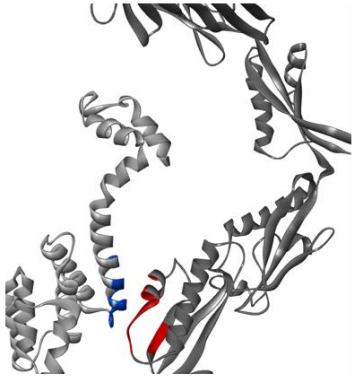
B



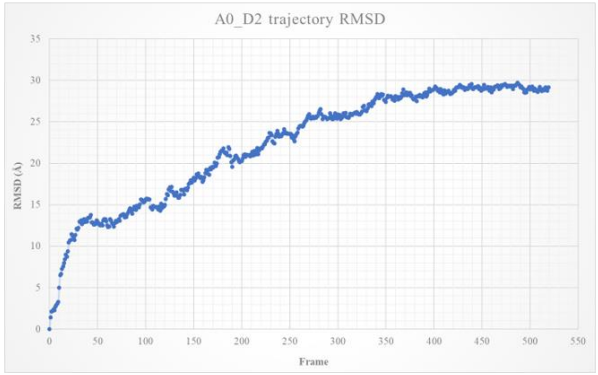
C

	Residues in contact (≤ 4 Å)	
Model	CjBamA	CjBamD
A0D2	**89A**, 90K, 102K, 105E, 106S, 108L, 109G, 110I, **111K**, **112R**, **113G**	145Y, 146T, 147E, 148Y, 149E, 150P, 151L, 153Q, 154T, 157T

D



E



F

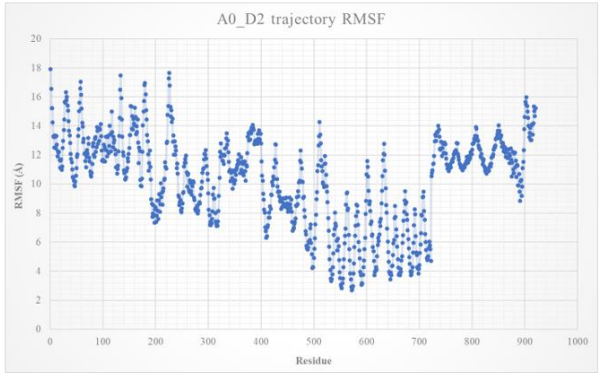


Figure 6.7: A0D2 MD docking run. The A0D2 docking run starting position (A), end position post-50 ns trajectory (B), residues in contact post-run (C), residues in contact post-run [graphic] (D), trajectory RMSD change by frame (E) and RMSF per residue (F) are shown. The A0D2 CjBamA-CjBamD docking run results in a very different end configuration to the E. coli experimental structures with CjBamD in contact with the membrane and bringing the CjBamA POTRA domain in contact with the membrane.

The RMSF (figure 6.7(f)) shows an average residue fluctuation of around 12 Å from the initial position, and so all structural elements of both proteins contribute to this RMSD change. As this position varies greatly to the *E. coli* model structures and there is a small contact interface between the two proteins, there are very few of the expected residues involved in the interaction.

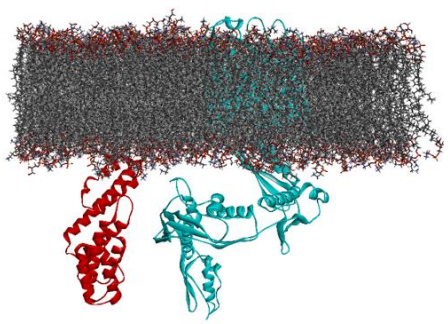
6.3.3.5 **MD Docking run A90D0**

The next run, A90D0, returns BamD to the expected docking orientation but rotates BamA around the y-axis by 90° (figure 6.8(a)), presenting a different surface for BamD to sample. At the end of the production run (figure 6.8(b)) there are no residues within 4 Å (figure 6.8(c) and (d)), however, the proteins did interact during the trajectory (data available on request). The interaction starts with BamD helix 7 interacting with the outside of POTRA 2 (50 frames), BamD helix 7 then move down to sample the outside of POTRA 1 (50-350 frames), and then slides off POTRA 1 to find the final configuration observed at the end of the trajectory. The RMSD stabilises after 200 frames (figure 6.8(e)), and the RMSF (figure 6.8(f)) shows BamD contributes to this change in RMSD while BamA stays relatively stable throughout the trajectory.

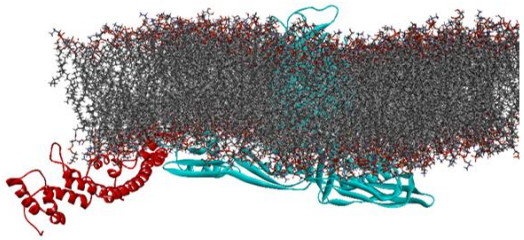
6.3.3.6 **MD Docking run A90D1**

The A90D1 docking run starts with BamD axially rotated and BamA rotated through 90° (figure 6.9(a)), and is another run that ends up with an extended POTRA presentation using BamD to adhere to the membrane inner leaflet (figure 6.9(b)). Ten of the expected BamA residues and 18 of the expected BamD residues contribute to this protein-protein interaction (figure 6.9(c)). The docking interface between the proteins consists of BamA POTRA 1 with BamD helix 6 and helix 8, BamA POTRA 2 β2-β3 loop with BamD helix 6, BamD helix 4 bridging POTRA domain 1 and 2, and BamD helix 2 interacting with POTRA 2 (figure 6.9(d)). The surfaces of POTRA 2 not used for the BamA-BamD interaction along with POTRA domain 3 and 4 are flattened against and interact with the membrane inner leaflet. The same applies to the unused surfaces of BamD. The BamD interaction prevents POTRA 1 from interacting with the membrane. The RMSD (figure 6.9(e)) is stabilised after 350 frames or 35 ns.

A



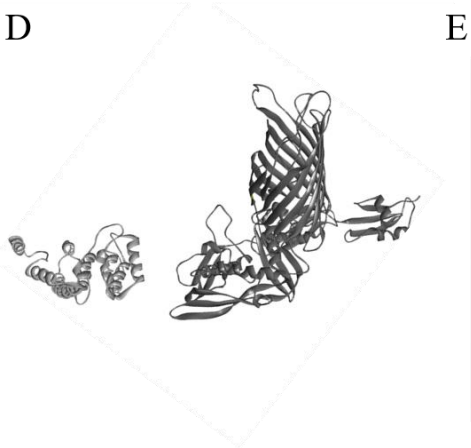
B



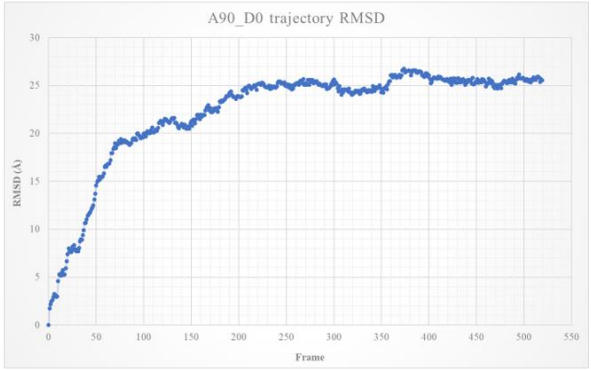
C

	Residues in contact (≤ 4 Å)	
Model	CjBamA	CjBamD
A90D0	None	None

D



E



F

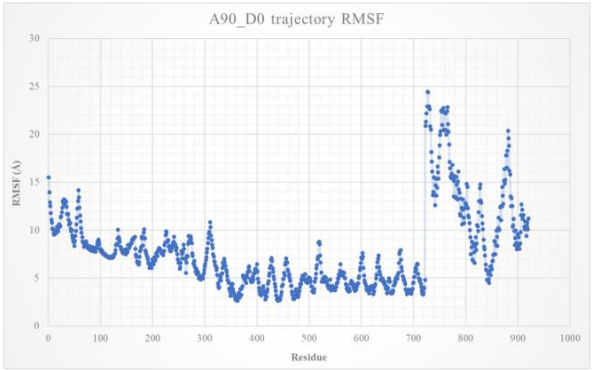
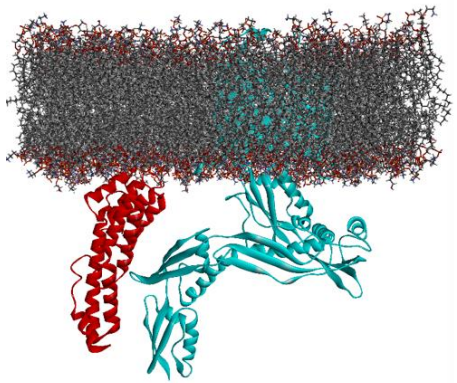
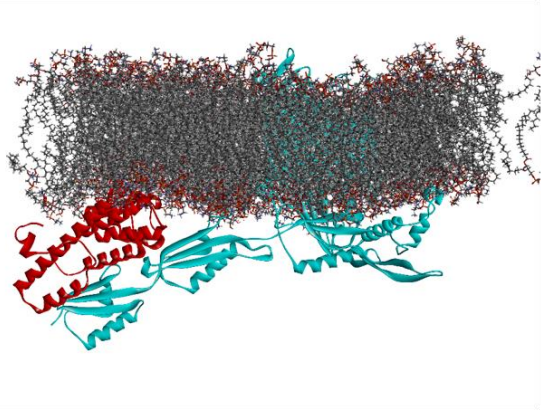


Figure 6.8: A90D0 docking run. The A90D0 docking run starting position (A), end position post-50 ns trajectory (B), residues in contact post-run (C), residues in contact post-run [graphic] (D), trajectory RMSD change by frame (E) and RMSF per residue (F) are shown. CjBamA-CjBamD interact transiently during the run as CjBamD slides along POTRA domain 2 and 1 before reaching the end configuration where no residues are found within contact distance. A stable docking configuration is not achieved within 50 ns from the A90D0 starting orientation.

A



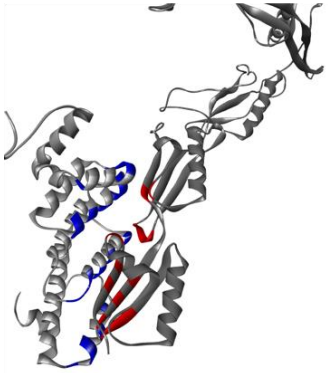
B



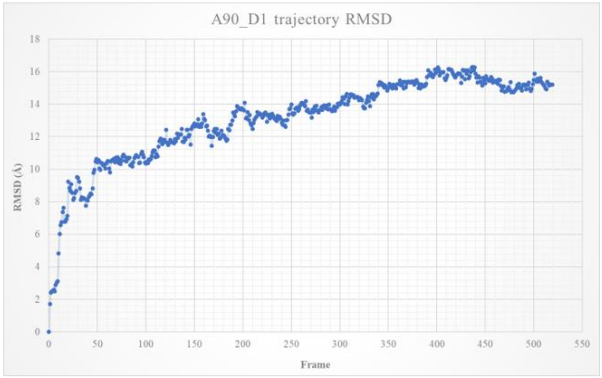
C

	Residues in contact (≤ 4 Å)	
Model	CjBamA	CjBamD
A90D1	24K, 26I, **29N**, **30H**, **31L**, 72E, 73N, 74N, 75N, 77I, 79E, 81I, **89A**, 90K, **112R**, **113G**, 149L, **150E**, **151N**, **152T**, **153D**	54Y, **63A**, **64D**, **65P**, **85K**, **86L**, **89F**, 92D, 95N, **96K**, **97K**, **99G**, **124R**, **125N**, **126Q**, **127A**, **130L**, 134K, 138T, 141K, 142D, 179H, **181E**, **182S**, **185I**, **189R**

D



E



F

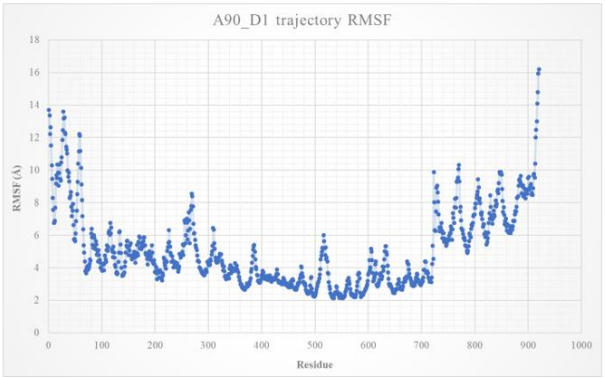


Figure 6.9: A90D1 MD docking runs. The A90D1 docking run starting position (A), end position post-50 ns trajectory (B), residues in contact post-run (C), residues in contact post-run [graphic] (D), trajectory RMSD change by frame (E) and RMSF per residue (F) are shown. The A90D1 docking run has a very different end configuration to the *E. coli* structures with CjBamD appearing to be pressed against the membrane by CjBamA POTRA domain 1 and 2, and has the majority of its residues in contact with either CjBamA or the membrane inner leaflet.

During the trajectory, BamD helices 8 and 9 interact with POTRA 1 (0-50 frames), BamD helix 8 then samples POTRA 1 (50-100 frames) after which BamD helix 8 and 9 interact with POTRA 1 and 2 and pull both POTRA domain closer (100-300 frames), leading to the final presentation. The RMSF (figure 6.9(f)) shows that BamD helix 9 and BamA POTRA 1 contribute the most to the RMSD values over the trajectory

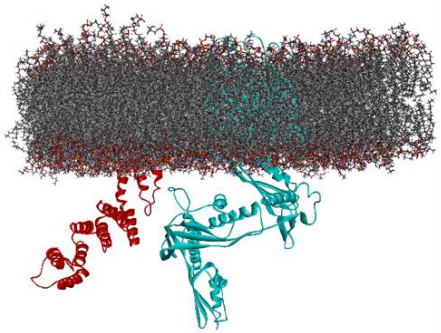
6.3.3.7 **MD Docking run A90D2**

The A90D2 starts with BamA rotated through 90° and BamD in flipped orientation relative to A0D0 (figure 6.10(a)) and the end point for the docking run shows an extended S-like shape for the periplasmic POTRA domain against the membrane inner leaflet, with stabilisation of POTRA 1 and 2 by BamD (figure 6.10(b)). The contact interface between the proteins at the end of the production run are BamD helix 8-9 loop with POTRA 2, BamD helix 3 and helix 5 and the helix 4-5 loop with POTRA 1, and helices 6 and 7 and the helix 6-7 loop with POTRA 1 (figure 6.10(c) and (d)). POTRA domain 3-5 are extended and flattened against the membrane inner leaflet, as are the unused surfaces of BamD helices 5-9. The RMSD (figure 6.10(e)) shows stabilisation after 350 frames (35 ns). At the start of the trajectory, BamD helices 6 and 7 and the helix 6-7 loop interact with POTRA 1 (frames 0-50). This configuration is then sampled (frames 50-300), after which BamA POTRA 1 moves towards BamD (frames 300-350) forming an interaction with helices 5-7 and the helix 3-4 loop, facilitated by BamD helices 1 and 2, stabilising the final configuration until the run ends. The RMSF (figure 6.10(f)) shows POTRA 1 and BamD helices 1-3 contribute the most to the change in RMSD over the trajectory.

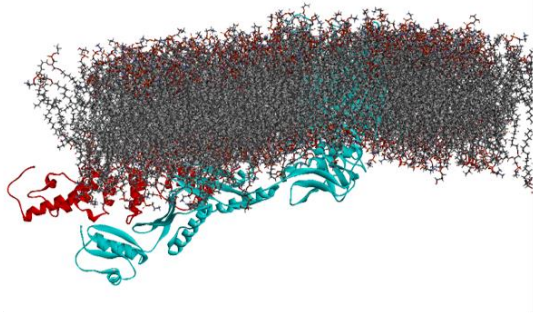
6.3.3.8 **MD Docking run A180D0**

In the next docking run, A180D0, BamA was rotated another 90° and BamD returned to the original orientation (figure 6.11(a)). At the end of the production run (figure 6.11(b)), the BamA POTRA domain are extended with all POTRA domain in contact with the membrane inner leaflet, supported at the side by the interaction of BamD with POTRA domain 2 and 3, giving a similar end configuration to A0D2. The interaction interface is between the BamD helix 5-6 loop, helix 7 and helix 8 with POTRA 3, and the helix 7-8 loop and helix 9 with POTRA 2.

A



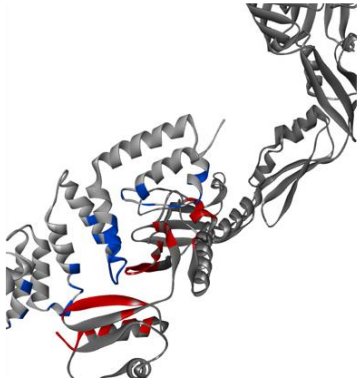
B



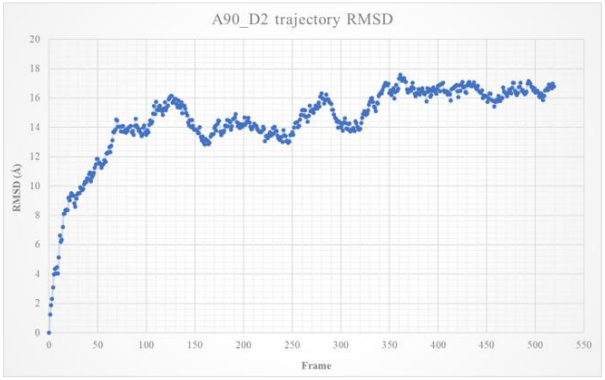
C

	Residues in contact (≤ 4 Å)	
Model	CjBamA	CjBamD
A90D2	18A, 19T, 20I, 21K, 49N, 50P, 51A, 54N, 58L, 61Y, 69I, 70A, 71V, 72E, 73N, 74N, 75N, 81I, **89A**, 90K, 92T, 94T, 102K, 103Q, 105E, 106S, **112R**, **113G**, 145K, 146K, 147K, 148T, 149L, **150E**, **151N**, **152T**, **154G**, 156E, 157L, 158E, 160I	**65P**, **68E**, 72I, **100N**, 102R, 103N, 105D, 106Y, 109Y, 133Q, 136I, 137D, 138T, 140L, 141K, 142D, 143Y, 144P, 145Y, 146T, 147E, 148Y, 149E, 150P, 152V, 153Q, 191Q, **192E**, 195F, 198Q, 200I, 203P, 204E, 207W, 208Y

D



E



F

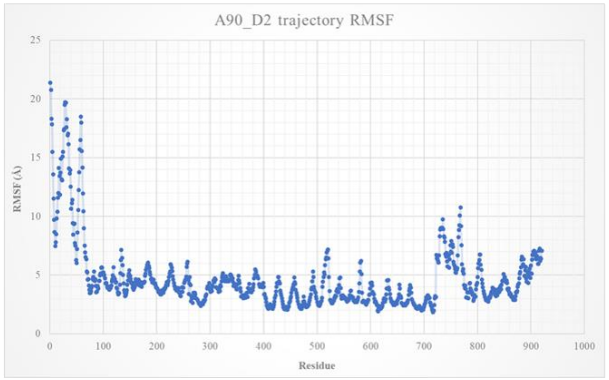
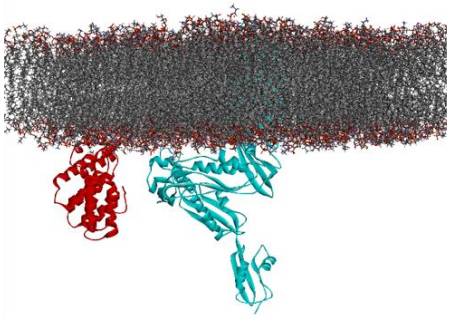
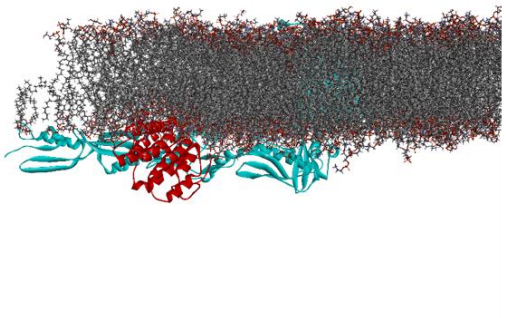


Figure 6.10: A90D2 MD docking run. The A90D2 docking run starting position (A), end position post-50 ns trajectory (B), residues in contact post-run (C), residues in contact post-run [graphic] (D), trajectory RMSD change by frame (E) and RMSF per residue (F) are shown. The A90D2 CjBamA-CjBamD docking run is another example of CjBamD interacting with the membrane inner leaflet with CjBamA POTRA domain extended and situated below CjBamD.

A



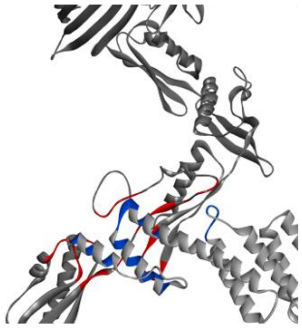
B



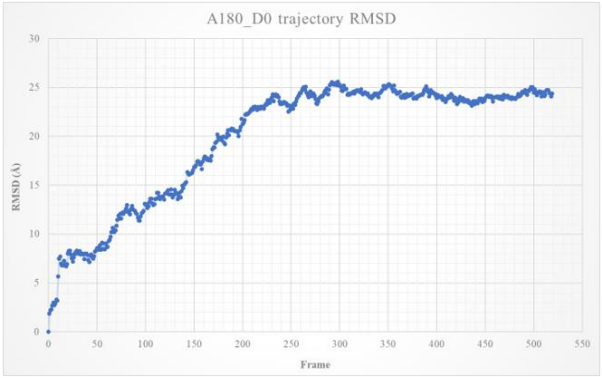
C

	Residues in contact (≤ 4 Å)	
Model	CjBamA	CjBamD
A180D0	95G, 96I, 97A, 98S, 99N, 100D, 101R, 137Y, 138F, 161V, 162N, 163R, 165E, 166N, 169I, 170D, 171N, 172V, 173H, 175S, 177A, 178K, 179K, 192K, 193E, 194K, 195E, 196F, 197M, 242Y, 243T, 244D, 245T, 247Q, 248A, 249N	**122P**, **123N**, **124R**, **125N**, **126Q**, **170I**, **173L**, **174Y**, **176R**, **177I**, **178G**, **179H**, **181E**, **182S**, **185I**, **186Y**, **189R**, 190L, **192E**, 193S, 207W, 211I, 214R, 215F

D



E



F

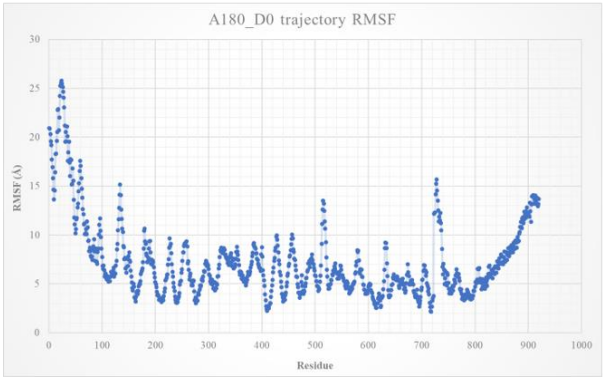


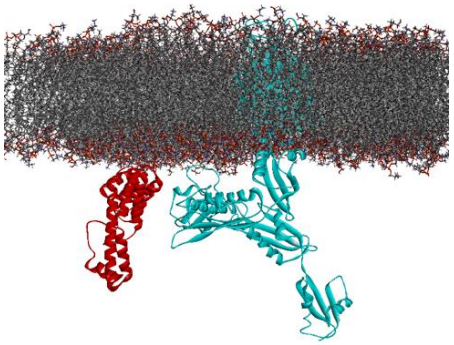
Figure 6.11: A180D0 MD docking run. The A180D0 docking run starting position (A), end position post-50 ns trajectory (B), residues in contact post-run (C), residues in contact post-run [graphic] (D), trajectory RMSD change by frame (E) and RMSF per residue (F) are shown. The A180D0 CjBamA-CjBamD docking shows interaction of both proteins with the membrane inner leaflet, with the POTRA domain extended and underneath/ side-on with CjBamD.

None of the expected BamA residues are involved in this interaction, however, 18 of the expected BamD residues are involved (figure 6.11(c) and (d)).The RMSD (figure 6.11(e)) stabilises after 270 frames (27 ns). In the lead up to this final configuration, BamD is initially interacting with the lipid membrane through helix 7 and 9, BamD flattens against the membrane (50 frames), the POTRA domain move towards BamD and the N-terminal coil mediates this transition by intermittent interaction with POTRA 4 (100 frames). This position is stabilised by BamD helix 7 and helix 8 and the POTRA domain begin to rise towards the membrane (150 frames) with POTRA 2 interacting after 200 frames and POTRA 1 after 270 frames. During these frames, the N-terminus of BamD moves away from the POTRA domain returning to its earlier flattened position against the membrane. This configuration is maintained until the end of the run. The RMSF (figure 6.11(f)) shows POTRA domain 1 and 2, β-barrel β-strands 6 and 7, and the N-terminal coil of BamD contribute most to the change in RMSD.

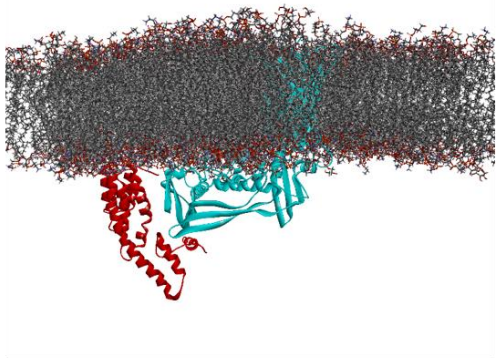
6.3.3.9 **MD docking run A180D1**

The A180D1 has BamA rotated 180° and BamD axially rotated relative to A0D0 (figure 6.12(a)) and the docking run shows extended L-shape POTRA domain against the membrane supported by BamD interacting with the membrane via part of helices 1-7 (figure 6.12(b)). The initial position shows part of BamD helices 1-7 and the POTRA 3 helix 1-2 loop interacting with the membrane. The interface between the proteins at the end of the run consists of BamD helices 4, 6, 8 and 9 interacting with the BamA POTRA β-strands. Very few of the previously identified residues contribute to this interaction (figure 6.12(c) and (d)). The RMSD (figure 6.12(e)) shows stabilisation after 100 frames. From the initial position the POTRA domain rise up towards BamD and the membrane and the BamD helix 8-9 loop interacts with POTRA 3 (0-100 frames). After this the helix 8-9 loop disengages from POTRA 3 and is replaced by an interaction between BamD helix 8 and 9 and POTRA 3, and this position is stabilised until run end. The RMSF (figure 6.12(f)) shows POTRA domain 1 and 3, BamD helices 8 and 9, and the BamD N-terminal coil contribute the most to changes observed in the RMSD profile.

A



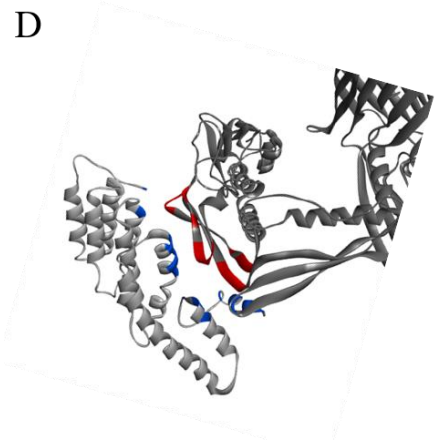
B



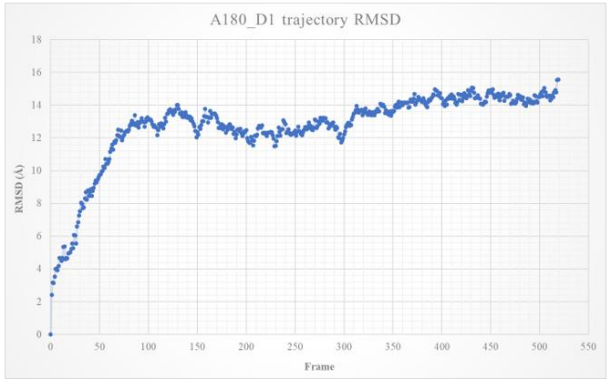
C

	Residues in contact (≤ 4 Å)	
Model	CjBamA	CjBamD
A180D1	170D, 171N, 173H, 175S, 176G, 177A, 178K, 179K, 231D, 232V, 233Q, 235S, 242Y, 243T, 244D, 245T, 255K, 256E, 257G, 258K	18C, **96K**, 134K, 137D, 138T, 141K, **189R**, **192E**, 194E, 204E, 207W, 208Y, 210S, 211I, 212F, 213E

D



E



F

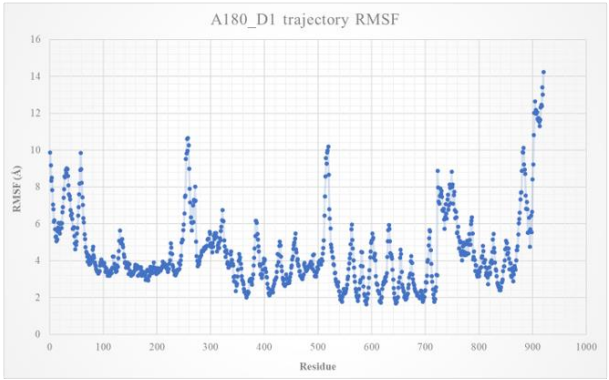


Figure 6.12: A180D1 MD docking run. The A180D1 docking run starting position (A), end position post-50 ns trajectory (B), residues in contact post-run (C), residues in contact post-run [graphic] (D), trajectory RMSD change by frame (E) and RMSF per residue (F) are shown. The A180D1 CjBamA-CjBamD docking run shows a stabilised position where both proteins interact with the membrane, with CjBamA POTRA domain forming an L-shape and interacting via POTRA 3 and CjBamD interacting via helices 1-7 and inter-helix loop structures.

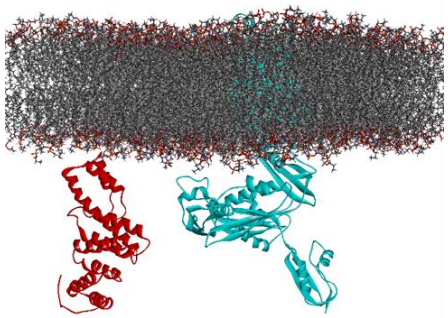
6.3.3.10 **MD docking run A180D2**

The A180D2 docking run starts with BamA rotated 180° and BamD in flipped orientation relative to A0D0 (figure 6.13(a)) and behaves similarly to the other 'D2'docking runs, with the POTRA domain up against the membrane following an interaction of BamD and BamA (figure 6.13(b)), mediated by the progressive interaction of BamD helix 7 with the membrane. The interaction interface at the end of the run is between BamD helix 1, the helix 2-3, 4-5 and 6-7 loops and helix 4 with POTRA 3. None of the expected residues for BamA and ten expected residues for BamD are involved in this interaction (figure 6.13(c) and (d)). POTRA domain 1-3 and 5, and the unused BamD surface are in contact with the lipid membrane post-run. The RMSD (figure 6.13(e)) shows stabilisation after 200 frames. There is no initial contact between BamD or the POTRA domain and the lipid membrane. A sampling phase occurs after run start (50 frames); the POTRA 3 helix 1-2 loop interacts with the membrane triggering the POTRA domain to become compact (50-125 frames). After which BamA and BamD move closer together and interact through helix 7 and then helices 2, 3, 4 and 5 (BamD) and POTRA 3 (frames 125-250). The proteins are held in this position and then the POTRA domain move progressively closer to the membrane as BamD helix 7 increasingly interacts with the membrane (BamD flattens) and the POTRA domain interact with the lipid surface. This configuration is maintained until the end of the run. The RMSF (figure 6.13(f)) suggest that BamD helices 1-4 contribute the most to the RMSD profile.

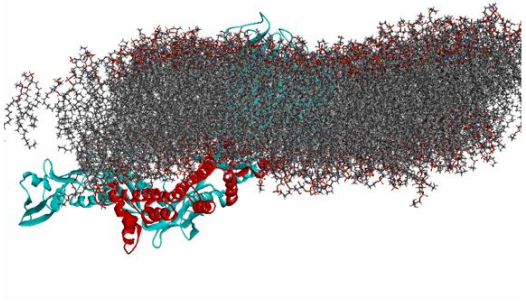
6.3.3.11 **MD docking run A270D0**

BamA is once again rotated 90° around the y-axis for this final series of docking runs and BamD is returned to the original orientation (figure 6.14(a)). The A270D0 docking run is a legitimate non-interacting docking orientation (figures 6.14(b)) with no residues in contact post-run (figure 6.14(c) and (d)). The RMSD (figure 6.14(e)) stabilises after 150 frames. At the starting position, BamD helices 8 and 9 are close to the lipid membrane and this likely results in BamD flattening against the membrane. Around frame 120 POTRA 1 and BamD helix 7 try to interact but are later repelled, this similarly occurs for the β-barrel β12- β13 loop and BamD helices 7, 8 and 9 around 250 frames only for the proteins to slowly separate until the end of the trajectory. The RMSF (figure 6.14(f)) shows the BamD N-terminal coil and POTRA 1 contribute the most to the change from the initial position.

A



B



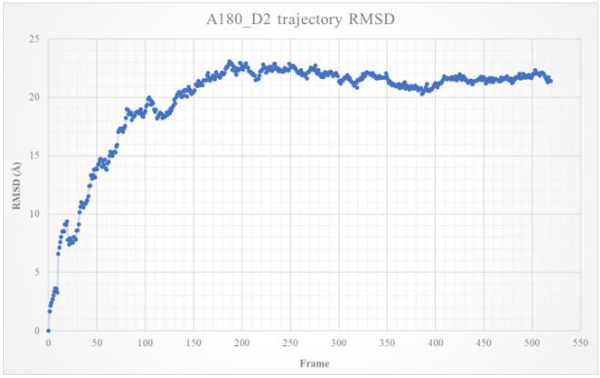
C

	Residues in contact (≤ 4 Å)	
Model	CjBamA	CjBamD
A180D2	170D, 171N, 172V, 173H, 174L, 175S, 178K, 181S, 182Y, 183S, 184D, 186E, 187P, 192K, 193E, 223E, 226K, 227K, 238Y, 240K, 242Y, 249N, 251T, 295E	35Y, **59S**, **60E**, **61H**, **62V**, **63A**, **64D**, **65P**, **98F**, **99G**, **100N**, 101S, 102R, 103N, 104A, 105D, 106Y, 145Y, 146T, 147E, 148Y, 151L

D



E



F

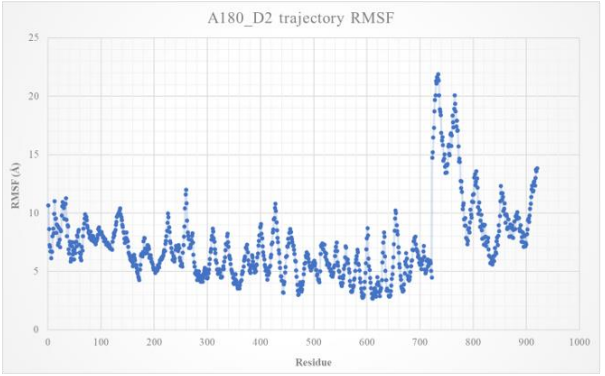
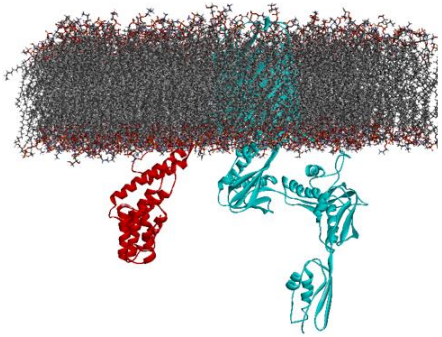
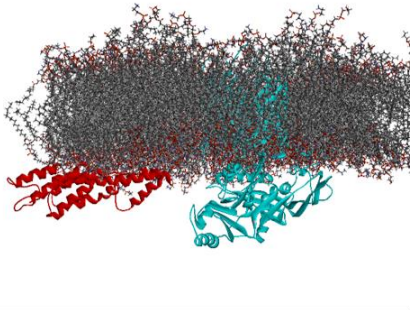


Figure 6.13: A180D2 MD docking run. The A180D2 docking run starting position (A), end position post-50 ns trajectory (B), residues in contact post-run (C), residues in contact post-run [graphic] (D), trajectory RMSD change by frame (E) and RMSF per residue (F) are shown. The A180D2 CjBamA-CjBamD docking run end configuration shows CjBamD interacting with the membrane via helix 7 and with CjBamA POTRA 3 via helix 1 and several loop structures.

A



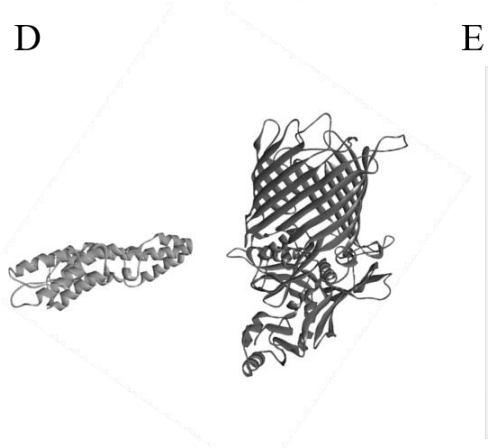
B



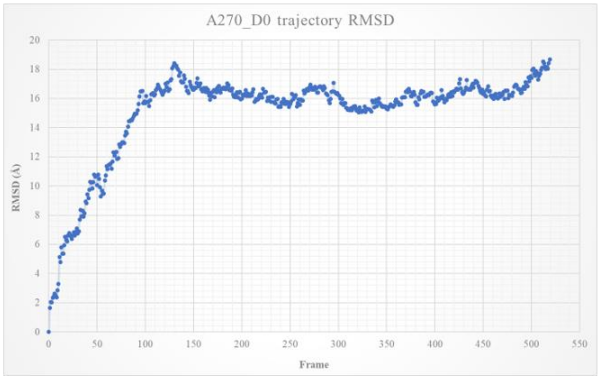
C

	Residues in contact (≤ 4 Å)	
Model	CjBamA	CjBamD
A270D0	None	None

D



E



F

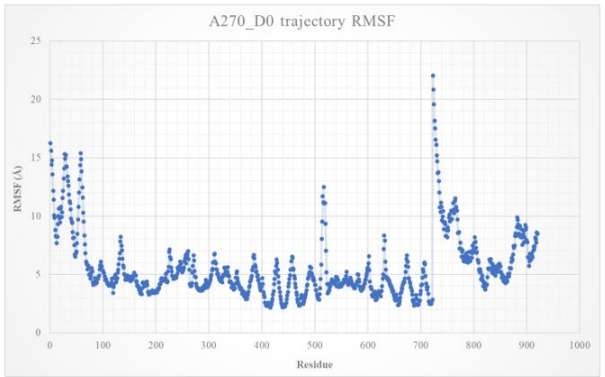


Figure 6.14: A270D0 MD docking run. The A270D0 docking run starting position (A), end position post-50 ns trajectory (B), residues in contact post-run (C), residues in contact post-run [graphic] (D), trajectory RMSD change by frame (E) and RMSF per residue (F) are shown. The A270D0 CjBamA-CjBamD docking run shows no interaction between CjBamA and CjBamD throughout the 50 ns docking run from the initial starting orientation.

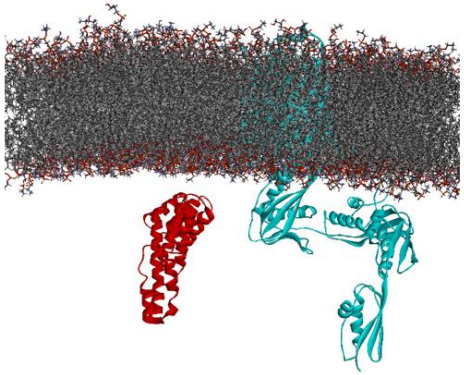
6.3.3.12 **MD docking run A270D1**

The penultimate run, A270D1, where BamA is rotated 270° and BamD axially rotated relative to A0D0 (figure 6.15(a)), behaves differently from previous runs, with BamD starting initially side-on and gradually working its way under the POTRA domain to eventually bridging the POTRA domain and causing its elongation, with BamD supporting from underneath (figure 6.15(b)). Some of the expected BamD docking residues contribute to the protein-protein interaction but no expected residues are found for BamA in this configuration (figure 6.15(c)).

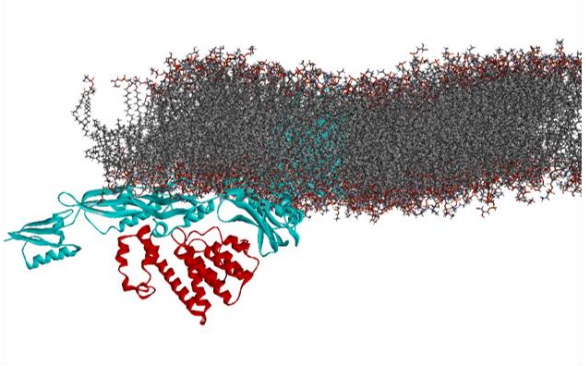
The contact interface between the proteins at the end of the trajectory is between POTRA 5 and BamD helix 4, POTRA 4 to BamD helix 2 and N-terminal coil, POTRA 3 to the BamD helix 1-2 and 3-4 loops and helix 9, and POTRA 2 to the helix 7-8 loop (figure 6.15(c) and (d)). The RMSD (figure 6.15(e)) shows stabilisation after 200 frames. However, this run shows very large changes in configuration up until frame 400. At the start point, BamD is not in contact with the membrane but the POTRA 3 helix 1-2 loop starts close to the membrane. From this starting point, BamD moves away from the membrane to interact with POTRA 4 via helix 4, and the POTRA domain begin to spread out (0-100 frames). The POTRA 3 loop interacts with the membrane, which causes the POTRA domain to start to elongate, and gradually POTRA 2 and then POTRA 1 interact with the membrane. Simultaneously BamD moves from a side-on to POTRA 3 to underneath POTRA 3 (frames 100-150). BamD continues to move underneath the POTRA domain and uses helix 7 to snag onto POTRA 2, bridging POTRA domain 2 and 4 (frames 150-200). A sampling phase occurs, and eventually BamD pushes POTRA 4 flat against the membrane or vice versa (frame 200-300).

During the next hundred frames the final configuration is formed, with interactions between BamD helix 1 and POTRA 4, helices 2-6 with POTRA 3, and helix 7 with POTRA 2. This pushes the BamA POTRA domain flush against the membrane, and this configuration is maintained until the end of the run. The RMSF (figure 6.15(f)) shows that fluctuations in configuration of BamD are mainly responsible for the RMSD change over the trajectory, and that the RMSF is the largest for this run out of all of the docking runs.

A



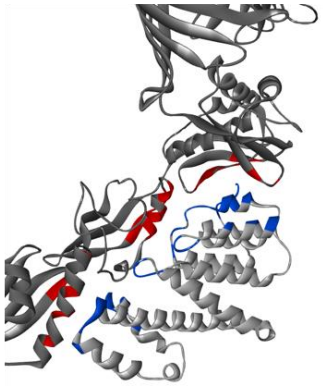
B



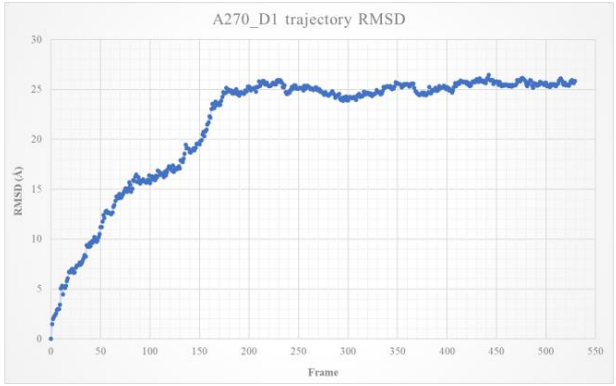
C

	Residues in contact (≤ 4 Å)	
Model	CjBamA	CjBamD
A270D1	122K, 125I, 126E, 129K, 139D, 141I, 142V, 143E, 166N, 167I, 218S, 219R, 221A, 222D, 225M, 226K, 230L, 231D, 232V, 233Q, 234V, 235S, 317E, 319Y, 321D, 322I, 323Q, 324K, 325N, 336K, 338I, 400K, 403D	18C, 19S, 20T, 21K, 22N, 23D, 24E, 45K, 47L, 48E, 50A, 51D, 52D, 53H, 54Y, 55N, 56G, 57M, **59S**, **60E**, 79M, 80D, 81E, 82E, **89F**, **93E**, **96K**, **97K**, **120A**, **121V**, **122P**, **123**N, **176R**, **177I**, **178G**, **179H**, **181E**, 207W, 210S, 211I, 213E, 214R, 215F

D



E



F

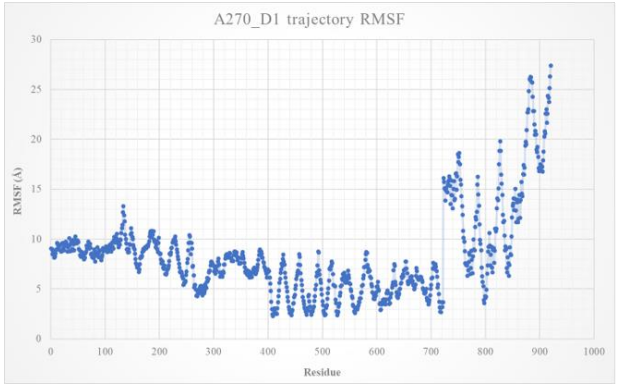


Figure 6.15: A270D1 MD Docking run. The A270D1 docking run starting position (A), end position post-50 ns trajectory (B), residues in contact post-run (C), residues in contact post-run [graphic] (D), trajectory RMSD change by frame (E) and RMSF per residue (F) are shown. The A270D1 CjBamA-CjBamD docking run results in CjBamA (light blue) POTRA domain interacting with the membrane in an extended configuration, but CjBamD (red) does not interact with the membrane and instead supports POTRA domain 1-3 from underneath.

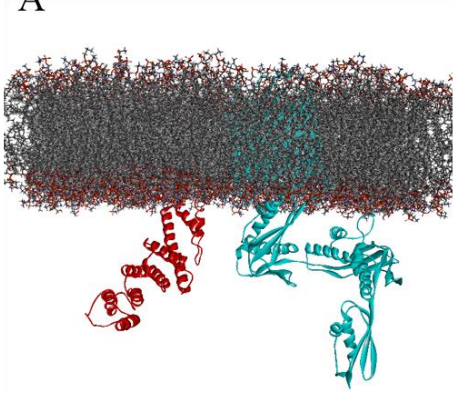
6.3.3.13 **MD docking run A270D2**

The final docking run, A270D2, where BamA is rotated through 270° and BamD is in 'flipped' orientation (figure 6.16(a)), behaves similarly to the other 'D2' runs. After 50 ns (figure 6.16(b)) the contact interface between the proteins is between the BamD helix 3-4, 5-6 and 6-7 loop and POTRA 5, BamD helix 9 and the N-terminal coil to POTRA 4, and BamD helix 2 to β-barrel β6-β7 and β8-9 loops. Some of the expected BamD docking residues contribute to the protein-protein interaction but no expected residues are found for BamA in this configuration (figure 6.16(c) and (d)). The RMSD (figure 6.16(e)) stabilises after 250 frames. The initial start point places the POTRA 3 helix 1-2 loop and BamD helices 7 and 9 close to the membrane.

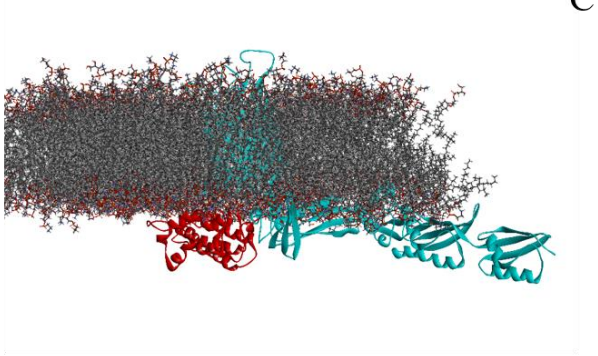
After the start of the run BamD moves closer to POTRA 4 (0-100 frames), then BamD helix 9 interacts with POTRA 4, and helix 6 and the helix 4-5 loop with POTRA 5 (frames 100-150). Simultaneously the POTRA domain begin to rise and then flatten against the membrane. This configuration goes through a sampling phase until the end of the run from this point, and BamD helix 9 eventually interacts with POTRA 4, giving rise to the final configuration. The RMSF (figure 6.16(f)) show POTRA 1 and BamD from the N-terminus to helix 4 contribute the most to the RMSD throughout the run.

The A0D0 and A0D0 normal (salt concentration) docking runs show very similar docking final docking positions overall to the expectation identified from the homologous *E. coli* BAM experimental structures. Comparison of the residues in contact in these final docked positions to RosettaDock model 4, the rigid body docking model that best resembled the pre-experiment expectations, show several common residues at the docking interface. Back comparison to the residue restraints determined from superimposition show roughly 50% of the residues from rigid body docking and the MD docking models could be identified from superimposition, and that several residues are present at the docking interface between CjBamA and CjBamD across all methods. The key residues important for docking of CjBamA and CjBamD derived from *in silico* methods are outlined below (table 6.1).

A



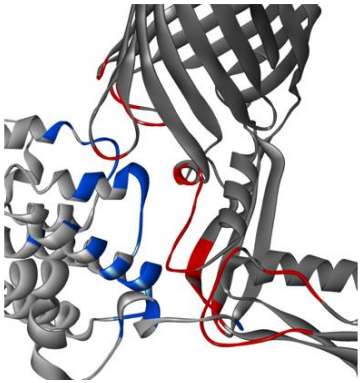
B



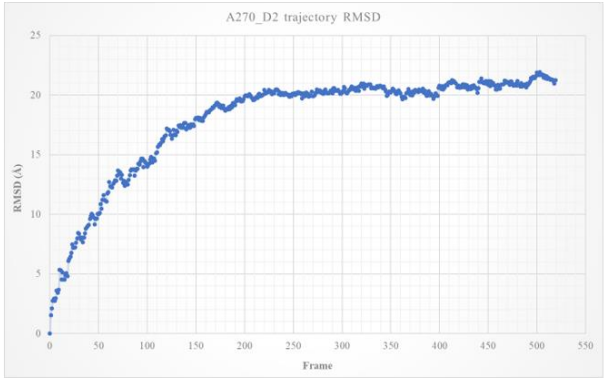
C

	Residues in contact (≤ 4 Å)	
Model	CjBamA	CjBamD
A270D2	269N, 270P, 271L, 272F, 273D, 274D, 275K, 276Q, 277N, 278A, 281V, 282K, 283D, 284L, 342K, 343V, 344Y, 346R, 347N, 348V, 359R, 362R, 363R, 365L, 366Y, 367I, 368T, 369E, 370G, 371N, 372L, 404T, 405H, 510R, 511Y, 556T, 557D, 558D, 559Y, 563R, 564S, 598D, 599Y, 672R	**59S**, **60E**, **62V**, **63A**, **64D**, **65P**, 92D, **93E**, **96K**, **97K**, **98F**, **99G**, **100N**, 101S, 102R, 104A, 105D, 108R, 111K, **115K**, 131E, 134K, 137D, 140L, 141K, 142D, 143Y, 144P, 145Y, 146T, 204E, 207W, 214R, 215F

D



E



F

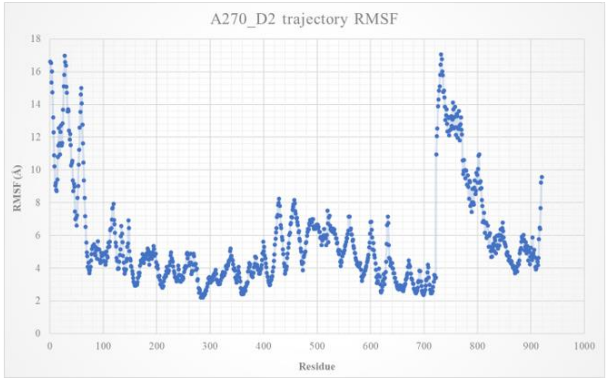


Figure 6.16: A270D2 MD docking run. The A270D0 docking run starting position (A), end position post-50 ns trajectory (B), residues in contact post-run (C), residues in contact post-run [graphic] (D), trajectory RMSD change by frame (E) and RMSF per residue (F) are shown. The A270D0 docking run end position does not resemble the *E. coli* structures, with CjBamD (red) situated underneath the CjBamA (light blue) β-barrel pore interacting via a periplasmic loop and POTRA domain 4 and 5, and the POTRA domains are extended against the membrane.

	Residues in contact (≤ 4 Å)	
Model	**CjBamA**	**CjBamD**
A0D0	**32S**, **34T**, **35S**, 58L, 59N, 62K, **63Q**, **65Y**, 102K, 103Q, 105E, 106S, 109G, 110I, **111K**, **112R**, **113G**, **114T**, 115L, **354S**, **355R**, **357V**, **359R**, 360V, **363R**, 471R, 474D, 475S, 476Q, 507Q, **509A**, **510R**, **511Y**, 554N, 555D, 556T	18C, 19S, 20T, 21K, 23D, 24E, 25G, 26L, **28N**, **29L**, **30S**, **31A**, 32S, 33E, **34W**, 35Y, 57M, **58A**, **59S**, **60E**, **61H**, **62V**, **63A**, 66L, **120A**, **121V**, **122P**, **123N**, 158K, 161L, 165Y, 172N, **173L**, 175Q, **176R**, **177I**, **181E**, 202K
A0D0 normal	147A, 148T, 149L, **150E**, **151N**, **152T**, **153D**, 156E, **346R**, **347N**, **349I**, **351S**, **352G**, **353N**, **354S**, **355R**, **356T**, **358D**, **369E**, 397K, 398E, 399E, 400K, 401V, 407D, 475S, 476Q, 506R, 507Q, 508F, **509A**, **510R**, **511Y**, 559Y, **560Y**	18C, 20T, 21K, 24E, 25G, 26L, **27Y**, **34W**, 37Q, 48E, 49K, 51D, 52D, 53H, 55N, 56G, **59S**, **60E**, 82E, 83E, 84Y, **85K**, **86L**, **89F**, **93E**, **96K**, **97K**, **122P**, **124R**, **125N**, **126Q**, **127A**, **130L**, 172N, **173L**, 175Q, **176R**, **177I**, **178G**, **180D**, **181E**
Rosetta Dock model 4	**30H**, **87T**, **112R**, **113G**, **114T**, **151N**, **152T**, **346R**, **347N**, **348V**, **349I**, **350I**, **358D**, **359R**, **361I**, **362R**, **363R**, **364E**, **365L**, **366Y**, **368T**, **369E**, **371N**, **509A**, **510R**, **511Y**, 512Y, 555D, 557D, 558D, **561L**, 564S, 565G, 566I, 594Q, 595G, 596L	20T, 21K, 22N, 23D, 24E, 26L, **121V**, **122P**, **123N**, **124R**, **125N**, 165Y, 168S, **169T**, **170I**, 172N, **173L**, **174Y**, 175Q, **176R**, **177I**, 202K, 206P, 209R, 213E
Key residues for docking	112R, 113G, 114T, 151N, 152T, 346R, 347N, 349I, 354S, 355R, 358D, 359R, 363R, 369E, 476Q, 507Q, 509A, 510R, 511Y	18C, 20T, 21K, 23D, 24E, 25G, 26L, 34W, 59S, 60E, 121V, 122P, 123N, 124R, 125N 163Y, 172N, 173L, 174Y, 175Q, 176R, 177I, 202K

Table 6.1: Residues important for docking of CjBamA-CjBamD. Table of residues identified at the docking interface between CjBamA and CjBamD in key models. Residues in bold were identified by superimposition of *C. jejuni* homology models and *E. coli* experimental structures. Residues are highlighted if there is agreement between rigid body docking model and one of the MD runs examined (yellow), agreement between both MD runs (cyan), and agreement between the rigid body docking model and both MD runs (red).

6.4 Discussion

The purpose of this chapter was to establish the correct docking positions for the *C. jejuni* BamA-BamD interaction, and determine the residues involved in this interaction using an *in silico* approach. Rosetta model four gives the closest visual representation to the *E. coli* structure docking confirmation of all of the rigid docking models in this study, and has the highest number of predicted residues involved in the docking interaction. However, the input file was heavily biased as a starting configuration is required, so it is unknown whether this docking configuration would be found organically if the input requirement was similar to that of ZDock and pyDock where no relative reference positions (orientation, spatial positioning) of the proteins are given. There would likely be more non-permissible models without the strong bias applied, but the result is what we were initially looking for, a similar docking position to the *E. coli* models. However, this is not the complete story as in the process of identifying this docking configuration, a few alternatives or variations on the docking configuration were found that could not be dismissed easily as they used similar residues or bound to similar positions to the expected outcome. These variations, particularly the axially rotated and 'flipped' orientations, appeared frequently among the results, so this was factored into the unbiased MD docking simulations starting configurations. This decision was supported by consideration of the differences between the *E. coli* five-component structures that place several restraints on BamD binding relative to the *C. jejuni* two-component BAM complex, where far more BamA surface area is accessible.

Reducing salt concentration (and thereby molecular shielding) expedites protein-protein interactions significantly without drastically changing the docked end configuration, reducing associated costs, but may result in loss of pre-docking interaction information. The major comparative in this work is that of the A0D0 (reduced salt concentration, Debye length altered) and A0D0 normal (physiological salt) MD trajectory runs, which both represent similar outcomes to the expected orientations of the BamA and BamD proteins observed in the *E. coli* experimental structures, but with the models separated at the start of the trajectory. In the A0D0 model where the Debye length has been altered to allow interactions over a longer distance, the protein-protein interaction occurs after equilibration (1.5 ns, occurs

before scaling and production runs) and conforms to the expected output of docking based on the crystal structure models. As both models had the same starting position and were assembled using the same vector alignment, the Debye length altered MD docking run (A0D0) approaches this initial 'endpoint' faster and/ or more directly than the physiological salt trajectory run. However, a caveat of this expedited time course for finding the docking configuration is that some extra information is lost on the interaction between CjBamD and CjBamA, as can be observed in the A0D0 normal trajectory (data available on request). The A0D0 normal model also finds this protein-protein interaction and conformation, however, the model exhibits multiple sampling phases, where POTRA 1 and 2 are sampled before POTRA 1 'flicks' BamD into place, before the final docking position is achieved. This points to an advantage of docking via MD simulations as opposed to docking via rigid docking programs (regardless of whether they have semi-flexible sampling component/ "local refinement") as new information on how the interaction occurs and steps towards the final conformation can be visualised. However, this run takes 200 ns (POTRA 1 engaged with BamD) or 400 ns (POTRA 1 disengaged from BamD). As a system of this size (298, 025 atoms) requires large computational resources (or a long run time) and there is no guarantee this extra information will reveal itself, it may be more cost effective to modulate the Debye length if the end result is all that is required.

Both CjBamA and CjBamD interact with the membrane during the MD docking runs except when they are in physiological salt (time-restricted) or are physically restricted. Multiple orientations and configurations of BamD relative to BamA, and vice versa, were sampled during the MD docking simulations, with the majority of runs showing protein interaction with the membrane inner leaflet. From the thirteen runs, only two runs have minimal interaction of BamD with the membrane, A0D0 normal and A270D1. In the former it appears to be due to the final position directly underneath the β-barrel, which acts as a physical restriction preventing BamD from interacting with the lipid surface. Though this may be partly due to the physiological salt conditions (increased molecular shielding/ reduced Debye length) and experimental limitations (timescale), as a longer 600 ns trajectory of BamAD from docking onwards in physiological salt conditions shows a flattening phenomenon and membrane interaction (data not shown). In the latter, BamD starts near the lipid interface but appears to interact more strongly with POTRA 4, eventually dragging

CjBamD away from the membrane, bridging the POTRA domain and forcing the POTRA domain to extend and flatten against the membrane. The former is an interesting configuration as BamD helix 7 is directly under the β-barrel, which could be a potential delivery mechanism of substrate OMPs to BamA for folding and assembly. The latter is also interesting, but for a different reason, as the β-barrel is exposed to the periplasm with the POTRA domain moved out of the way of the β-barrel opening. As BamA is a membrane protein and one of the mechanisms in the literature relies on membrane interaction, and BamD is a lipoprotein, some interaction with the membrane is expected to occur, but whether either of these configurations are biologically relevant will need to be probed in follow up MD simulations.

CjBamD helix 7 and CjBamA POTRA 3 helix 1-2 loop mediate interaction with the membrane. Several mechanistic details of the protein-lipid interaction were found during the course of these production runs. The first is that the common triggers that cause the proteins to associate with the membrane are BamD helix 7 (helix 6-7 loop end) and/ or helix 9, and the BamA POTRA 3 helix 1-2 loop. When either of these BamD helices come close to the membrane they form an association, and something similar to a lever or pivot/ fulcrum mechanism is found where BamD progressively flattens against the membrane with increasing interaction of helix 7 with the membrane. In this way helix 7 acts as both the anchor to the membrane (via helix 7 and the helix 6-7 loop) and levers BamD flat against the membrane. This may be because it is the largest helix in BamD and has the largest surface area, and BamD maintains an overall rod-like shape. On the other hand, in run A270D1 helix 7 initially mediates the bridging of POTRA domain before another helix makes an interaction with POTRA 2 freeing up helix 7, and then movement in helix 7 causes the POTRA domain to be pushed flush against the membrane. Whether this apparent leverage function is biologically relevant is unknown, however there should be a reason behind the kink present in helix 7, and with the exception of the BamA derived peptide (*E. coli* BamA residues 765–779)[107] BamD has not yet been shown to be used for binding of other proteins in the literature. The BamA POTRA 3 helix 1-2 loop appears to be the trigger for the non-BamD mediated lipid association, causing the POTRA domain to rise towards the membrane inner leaflet. This interaction is found in A180D1, A180D2, A270D1, and A270D2 in this study and

triggers the POTRA domain to rise in the absence of BamD. Whether either of these triggers are influenced by the increased Debye length or are biologically is unknown. However, in the case of BamA there is a mechanism related to membrane destabilisation and the runs in which the elongated lipid associated POTRA domain occurs the lipid membrane appears to display some curvature (data not shown, analysis method not known). Whether this extension of the POTRA domain facilitates this mechanism by applying a stress to the membrane will need to be investigated in future work.

There are some commonalities between all of the docking runs, such as the structures frequently involved in protein-protein interactions. These common interaction points are BamD helix 7 and the majority of the helix loops (on the helix 6-7 loop side), and also, helix 1 and helix 9 where bridging of the POTRA domain occurs. The most active structural elements (in terms of RMSF) appear to be POTRA 1, N-terminal BamD to helix 3, helix 7 and helix 9. POTRA 1 can be explained by the hinges between POTRA domain and the overall mobility of the POTRA domain, but as BamD appears quite compact and rigid the amount of conformational change observed in the MD docking runs is surprising, and notably so in A270D1.

BamD appears to be repelled when interacting with the outside (β-strands) of POTRA 1 and 2 (A90D0), but there is a very favourable interaction with POTRA 3 (A180D0, A180D2, A270D1, A270D2) and POTRA 4 (A180D0, A270D1, A270D2) from a similar configuration. However, when BamD interacts with POTRA 1 from the inside (mainly the A0D0/ D1/ D2) an interaction is formed. The reasons for this may be because in some of these runs POTRA 4 was not associated with the membrane and so BamD could sample above or below POTRA 3 and 4 or there may be some underlying reason why the interaction does not work with POTRA 1 or 2 on the outside. In the A90D0 run, it appeared that BamD slowly made its way clear of POTRA 2 and 1 from the outside and would loop around to bind to the inside of the POTRA 1, but the trajectory did not extend past 50 ns, so this it is unknown whether BamD would have re-orientated and bound.

Despite screening the original (D0), axially rotated (D1), and ‘flipped’ (D2) states of BamD identified in rigid body docking against several rotational orientations of BamA (original [A0], 90° [A90], 180° [A180], and 270° [A270]), the expected

docking outcome was only observed in the A0D0 and A0D0 normal MD docking runs. This indicates either that the timeframe was insufficient or that an orientational restriction needs to be applied for the interaction between BamA and BamD to replicate the configuration observed in the *E. coli* experimental structures, and more widely, to any future MD docking runs.

Overall, the advantages of this methodology are the improved environment (hydration, membrane), observation of sampling prior to the end docking configuration, no requirement for knowledge about residue restraints, which instead is replaced by a rougher approximate spatial orientation, and that docking components have several properties modelled during docking (energies, bond vibration, conformational changes, etc.). The disadvantages are that prior knowledge is required regarding homology modelling, established protocols, MD simulation, software, and HPC script submission, and the cost of MD simulation (approximately £80 for the roughly 350,000 atom 50 ns A0D0 504 core CPU simulation at £0.02/ CPU hour). Though recent GPGPU implementations significantly reduce this cost by drastically reducing CPU requirements[214], it is advisable to orientationally bias any MD docking runs with knowledge from the literature.

6.5 Summary

In this work we utilise superimposition of *C. jejuni* homology models of BamA and BamD onto *E. coli* experimental structures to derive a list of residues that facilitate interaction at the docking interface, which were then used as residue restraint inputs in established rigid body docking programs to probe the docking interface between CjBamA and CjBamD. During rigid body docking, multiple configurations for the docked model were identified, and as *C. jejuni* BAM is a two-component system, these variation were potentially viable.

To investigate the viability of these variations an unbiased MD-based docking strategy was employed whereby multiple orientations and relative configurations of BamD (original, axially rotated and ‘flipped’) and BamA (rotations through 270° in 90° increments) were sampled to determine which docking states occur. The A0D0 docking orientation in physiological (A0D0 normal) or reduced salt conditions (A0D0) resulted in CjBamAD docked structures that closely resembled the docked structure observed in *E. coli* experimental structures. However, despite the increased

flexibility and surface availability of the two-component *C. jejuni* BAM system, the 'flipped' and axially rotated configurations of BamD in combination with the progressively rotated BamA configurations were not found to form the docked configuration observed in the rigid body docking outputs in the timeframe of the experiment. Sampling of other orientations either resulted in previously unknown configurations, an interaction that led to lipid association or, in one case each, an unfavourable or undocked configuration. The deviations from the expected outcome suggest that this methodology has somewhat similar caveats to existing rigid body docking, where residue restraints are typically required to derive meaningful results. However, instead of precise residue information, MD docking requires rough spatial and orientational biasing to achieve the desired outcome.

Some mechanistic details were unravelled during the course of the MD docking runs, such as the hinge 'flick' action of POTRA 1 to move BamD into the expected docking configuration (A0D0 normal), and the main structures involved in protein-protein interactions and protein-lipid interaction were identified. The important residues, which are prevalent in both the rigid docking top model (RosettaDock model 4) and the A0D0 MD docking runs that replicate the expected docking orientation observed in the *E. coli* experimental structures well, were elucidated for the *C. jejuni* BAM system.

Modulation of the Debye layer through changing the salt concentration expedited docking to the end configuration (A0D0 *vs.* A0D0 normal) and allows for a reduced cost simulation, but may prevent observation of additional pre-docking interaction information.

7 Concluding remarks and future work

7.1 Concluding remarks

Prior to the commencement of this project there was already significant progress in the literature characterising the *E. coli* BAM complex, ranging from experimental structural models to well theorised mechanisms published by several large groups. Since the onset of the project, several other BAM structures have been determined and some of the mechanistic models have structural validation. The unique angles of this project were a lack of literature on the *C. jejuni* BAM complex, and that it was a simplified two-component model rather than the five-component models characterised. The main aims of this body of work at the onset was to characterise the expression, purification of BamA, follow up on previous circular dichroism work in the lab, and get to the stage where preliminary attempts at structural characterisation via X-ray crystallography or Cryo-TEM were in progress. During the course of this pursuit, several other related proteins were cloned, expressed, purified and characterised, and the project expanded beyond its original intentions to encompass the major proteins involved in the *C. jejuni* BAM-mediated outer membrane transport system. The availabilities of these additional constructs pushed the project towards characterisation of the interaction between proteins in this pathway to determine whether the interaction network mimics that found in *E. coli*.

The original aim was achieved in doing this work, and prior to project closure, structural characterisation was attempted by Cryo-TEM for the *C. jejuni* BamAD complex. In addition, several other proteins were isolated and characterised using CD and SAXS, to varying degrees of success. The main findings of the project are the interaction between the *C. jejuni* SurA-like chaperone Cj1289 and BamD, dummy atom model outputs and the conformations of Cj1289, the utility of Landau modelling analysis and its benefits over Boltzmann analysis for prediction of T_m, and that docking through MD simulations may capture mechanistic details that are not available through rigid body docking. The relevance to the literature of a BamD-Cj1289 interaction, as shown by the dummy atom model and increased thermal parameters during thermal degradation, is it provides a mechanism by which BamD is delivered to the membrane. The Cj1289 dummy atom model analysis shows a

closed state for Cj1289 is representative of the crystal structure (3rgc[108]), and an 'open' state, which predominates in solution over the 'closed' state and is the likely base state of Cj1289 in solution before binding of suitable protein substrates. The Landau free energy expansion models used in this work predict the melting temperature more accurately, or with less variation, than the current Boltzmann wavelength based analysis employed for the dataset analysed when compared to values typically found for similar proteins in the literature[168]. This was achieved through isolation of individual secondary structures for analysis, allowing for discrimination between α-helical and β-sheet melting temperatures, and suggests that the Landau free energy expansion models have advantages over the current Boltzmann analysis for determining melting temperature. This discrimination between structure melting temperatures and the ability to directly compare the 'hardness' of melting, strongly suggest that Landau free energy expansion models should be used alongside, as an expansion to, or (eventually) instead of the current established Boltzmann analysis of CD. The MD docking runs exhibited an array of different ending configurations, some of which led to potential mechanistic details being observed. In particular, the flicking motion of POTRA 1 to help engage BamD in the correct docking position could represent a mechanism for loading BamD onto BamA, and illustrates that docking in this way through MD simulations can lead to extra mechanistic details being observed that are lost when using rigid body docking.

Overall, significant progress in the characterisation of the *C. jejuni* BAM complex has been achieved during the course of this project and the project is now well poised to deliver on its more ambitious objectives in the future.

7.2 Future work

7.2.1 Extensions to current work

Several CD thermal degradation experiments were performed in chapter 4, but several others could not be run due to unforeseeable circumstances of beam time cancellation and the ensuing abrupt re-scheduling of experiments. The follow up work for CD should include a repeat of the BamD CD runs, and the missing spectral scans and thermal degradation experiments for DegP and LptD, and the DegP-LptD, DegP-BamA, DegP-BamD, LptD-BamA, LptD-BamD, and LptD-Cj1289 mixtures.

This will lead to the further characterisation of these proteins and progression in the characterisation of the interaction network of *C. jejuni*.

Several proteins in chapter 5 were unable to fully characterised using SAXS due to aggregation uptick observed in the Guinier region. These experiments require repeating once facility access to dynamic light scattering and funding resumes.

If suitable yield is obtainable, the results in both of these chapters should be supplemented with surface plasmon resonance, analytical ultracentrifugation, or isothermal titration calorimetry data on proteins that are shown to interact through complexation analysis or those that lead to dummy atom models to further validate models.

The follow-up work on chapter 6 involves further analysis of the MD trajectories using a combination of principal component analysis to characterise protein-lipid contacts and by determining a suitable methodology to examine any stresses applied to the membrane by the lipid-associated MD docking runs. In addition, the MD simulations that did not lead to a docked structure may need to be extended to see if the result remains as a non-docked structure. This *in silico* work could be supplemented through the use of either carbene footprinting[215] or biotin labelling[216] protein-protein interaction wet lab experiments, which would provide sufficient corroboration to the MD simulation results.

This thesis covered data obtained by several methodologies, but while it appears a thorough and extensive account of work undertaken in this project, it does not cover data that was obtained for cryogenic transmission electron microscopy (Cryo-TEM) of a *C. jejuni* BamAD mixture and a large number of MD simulations, which are yet to be analysed. The following brief sections describe the background, current progress, challenges and future work required to complete analysis on these data sets.

7.2.2 **Cryo-TEM**

An ambitious goal of this project was to produce a structure of the *C. jejuni* BamAD complex. Crystallisation was initially considered, however, with considerations to the yield of the proteins obtained and the quantity and concentration of proteins required for crystallisation, this experimental theme was temporarily shelved and

other, more achievable, goals were established. Another factor in this decision was the inherent flexibility of the BamA POTRA domains and loops, and the lack of the other complex members (BamB, BamC, and BamE) observed in *E. coli* that would likely have a role in stabilising the complex. Thus, whilst the two-component BAM complex has fewer complex members to purify in order to achieve this theme of investigation, the increased flexibility will likely decrease the likelihood of a successful outcome.

Alternative methods of obtaining an experimental structure for this system include NMR spectroscopy, wide-angle X-ray scattering (WAXS), and cryogenic transmission electron microscopy (Cryo-TEM). The size of the BamAD complex would make structure determination via NMR into its own non-trivial project that would likely require the entire length of the PhD (Dr B. Bonev, personal communication, 2017). Due to this and the initial difficulties with the proteins and the subsequent time remaining, this avenue of structure determination was not pursued despite the speciality of the Bonev lab in NMR. WAXS was not attempted for similar reasons to crystallography, as it requires a reasonably large volume and concentration that were unfeasible with the typical protein yields obtained. The last option, Cryo-TEM, has been gaining momentum in the literature due to the recently improved detector technology and software allowing for higher resolution structures to be obtainable[217], and had relatively relaxed requirements in terms of sample concentration (3 µl of sample at 1 $mgml^{-1}$/ grid; this work), this presented as the achievable option in the time remaining in the PhD.

As the last experiment to this PhD project, a BamA-BamD sample mixture was prepared and submitted to a screening day at the Leicester Cryo-TEM facility for screening on a Titan Krios microscope. The sample screening was deemed to be of good quality, and the sample data was extended to an approximately 3300-micrograph data collection.

The analysis of these micrographs was attempted using the regularised likelihood optimisation **(**RELION) 3.0 software suite[218] on the Leicester HPC cluster. However, due to several factors, this aspect of the project had to be stalled. The data obtained, in and of itself, presented a major challenge in the identification of the protein for the production of 2D classes due to the small size of the complex (110 kDa), the

observed lack of contrast in the micrographs due to the detergent micelles, and the sample conformational heterogeneity. A potential set of subclasses were identified during the 2D class assignment, however, it is difficult to progress to 3D models due to the sample heterogeneity observed in the 2D classes (appendix 8.7.1).

The future work on this section of the project will be to carry on this analysis in RELION to determine if a 3D model can be isolated, and use the analysis of this dataset to optimise the sample further (as required). An alternative methodology that may be better able to resolve heterogenous sample is CryoSPARC, which features a heterogenous refinement algorithm[219]. One potential sample optimisation is the use of the "GraDeR" protocol[220], which uses the membrane mimetic LMNG as the choice detergent, coupled with a sucrose gradient and washes to produce a better contrast sample for analysis via Cryo-TEM. This detergent was chosen due its propensity to act similarly to a lipid molecule and stay associated with proteins when subjected to the protocol preparation requirements. In chapter 3, LMNG did not work particularly well for the extraction and purification of the *C. jejuni* BAM complex, but the potential for a FC12 extracted and purified sample that is detergent exchanged into LMNG has not been tested. Further approaches to consider are the recent trends in use of FABs in Cryo-TEM[221, 222], which serve to stabilise and increase the size of the complexes, preventing the particles from adopting several heterogeneous conformations and making particles more visible on micrographs to aid particle picking.

7.2.3 Unanalysed molecular dynamics simulations

Several MD simulations were carried out during the course of this study that are yet to be analysed, and so, have not been presented in this thesis. These MD simulations are summarised by the run, environment and production length (appendix 8.7.2).

The purpose of the majority of these runs were to investigate the structural dynamics of the proteins involved in this study, protein-protein interactions (BamA-BamD), how this interaction changes with a change in the simulated environment, and for SAXS model validation and analysis.

The majority of the simulations ran into issues caused by the periodic boundary conditions, which has hindered the analysis as it effected the RMSD, RMSF and

other potential analyses. However, this issue was addressed for the analysis of the docking runs in chapter six, and the methodology for removing drift preventing simulation analysis is now known.

The future work on this segment of the project can be divided into the two parts. The first part involves working out how to remove the drift without causing artefacts to occur in the membrane because of the unwrapping, and how to stop the drift from occurring in future simulations. The second part is to carry out the RMSD, RMSF, and further analyses to investigate whether the cyclic behaviour cited in the literature for the BAM complex is shown in the simulations, and investigate the structural dynamics, with particular emphasis on the lipid interactions and the effect of lipopolysaccharide.

8 Appendices

8.1 Chapter 1 appendices

8.1.1 Operon search

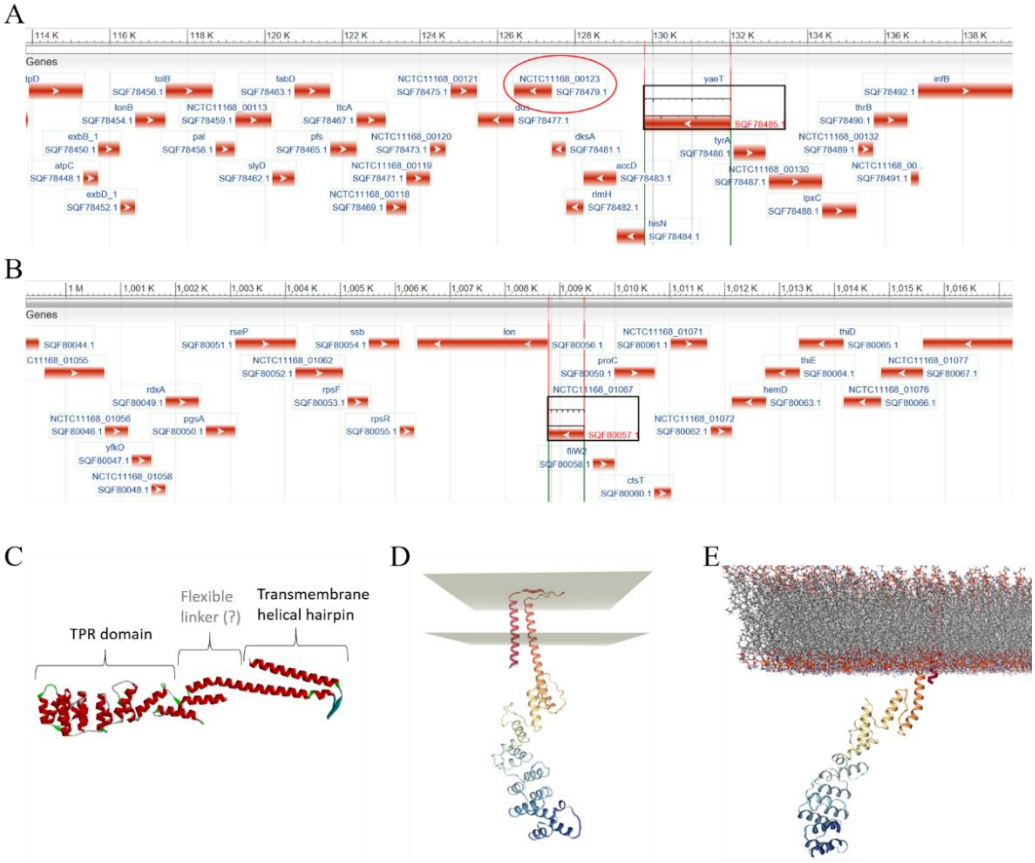


Figure 8.1: Operon search around *bamA* and *bamD* gene loci. The operons containing the *bamA* and *bamD* gene loci were searched for genes that were likely to be co-transcribed. Several genes were identified around the *bamA* (*yaeT*) gene locus (A), including genes involved in histidine synthesis (*hisN*), fatty acid synthesis (*accD*), transcription (*rlmH*), transcription regulation (*dksA*, *dus*) and one unidentified gene for a putative membrane protein (*NCTC11168_00123*). The gene for *lon*, an ATP-dependent serine protease, was identified around the *bamD* (*yfiO*) gene locus (B). The putative membrane protein was submitted to AlphaFold 2 for *ab initio* homology modelling, and the relaxed top ranked output, composed of an N- terminal transmembrane helical hairpin and a C-terminal TPR domain is shown (C). The output was assembled in CHARMM-GUI, and the PPN server membrane placement (D) and placement in a POPE: POPG membrane bilayer (E) are shown.

8.2 Chapter 2 appendices

8.2.1 Expression testing variables

Condition	Values tested
Media	LB, terrific broth
Cell type	BL21, BL21(DE3), C41(DE3), pLysS [BL21(DE3)], Lemo21, Lemo21^{R}, BL21(DE3)ai, ΔABCF*
OD induction point	0.4, 0.6, 0.8, 1.0, 1.5
Aeration volume (relative to media volume)	1:1, 2:1, 3:1, 5:1
IPTG (mM)	0, 0.1, 0.2, 0.4, 0.6, 0.8, 1.0
Glucose (%)	0, 0.1, 0.2, 0.4, 0.6, 0.8, 1.0
Arabinose (%)	0, 0.1, 0.2, 0.4, 0.6, 0.8, 1.0 (Bl21(DE3)ai only)
Rhamnose (mM)	0, 100, 250, 500, 750, 1000 (Lemo21 strains only)
Temperature (growth, °C)	37, 30, 27
Temperature (induction, °C)	37, 25, 20, 18, 16
Induction time	3 hours, 4 hours, 5 hours, 8 hours, O/N
Detergent (extraction)	DDM, DDM + CHS, DM, DM + CHS, OG, LMNG, OGNG + CHS, LDAO, C12E8, C12E9, Cymal-5, FC12, SB12
Detergent (purification)	DDM, DDM + CHS, DM, DM + CHS, OG, LMNG, OGNG + CHS, LDAO, C12E8, C12E9, Cymal-5, FC12, SB12

*BamA only

Table 8.1: Expression screening conditions tested in this study.

8.2.2 **Optimised expression conditions**

Protein	**Media**	**Expression cell type**	**OD_{600} at induction**	**Aeration volume**	**IPTG (mM)**	**Glucose**	**Temperature (growth, °C)**	**Temperature (induction, °C)**	**Induction time**	**Detergents used**
BamA	LB[a]	BL21(DE3)	0.60-0.65	3-5x	0.4[b]	0%	30	16	O/N	FC12 or SB12
BamD	Terrific broth	C41(DE3)	0.8-1.0	2-3x	0.4[b]	0.1%	30	16	O/N	-/ SB12
Cj1289	Terrific broth	BL21(DE3)	0.4-0.6	2-3x	0.4[b]	0.1%	37	20	O/N	-
DegP	Terrific broth	pLysS	0.4-0.5	2-3x	1.0	0.1%	37	37	4 hours (max)	-
LptD	LB[a]	BL21(DE3)	0.6	3-5x	0.4	0%	30	18	O/N	FC12

[a] Terrific broth gives slightly better yields, but there is roughly double the pellet weight to process making LB expression the streamlined option with the equipment available in the Medical School (rotor volume Optima SW41Ti).

Table 8.2: Optimised expression conditions for proteins in this study.

8.2.3 **Experiment run sheet**

No.	Run	Protein 1	Protein 2	Stoichiometry	Concentration ($mgml^{-1}$)	Buffer
1	detergent buffer	-	-	-		0.25 % SB12, 150 mM NaF 30 mM Tris
2	Non-detergent buffer	-	-	-		150 mM NaF, 30 mM Tris
3	BamA	BamA	-	-	0.821	0.25 % SB12, 150 mM NaF 30 mM Tris
4	BamD	BamD	-	-	0.903	150 mM NaF, 30 mM Tris
5	Cj1289	Cj1289	-	-	0.919	150 mM NaF, 30 mM Tris
6	BamD SB12	BamD	-	-	0.903	0.25 % SB12, 150 mM NaF 30 mM Tris
7	Cj1289 SB12	Cj1289	-	-	0.919	0.25 % SB12, 150 mM NaF 30 mM Tris
8	BamD + Cj1289	BamD	Cj1289	1:1		150 mM NaF, 30 mM Tris
9	BamA + BamD	BamA	BamD	1:1		0.25 % SB12, 150 mM NaF 30 mM Tris
10	BamA + Cj1289	BamA	Cj1289	1:1		0.25 % SB12, 150 mM NaF 30 mM Tris

11	BamA + Cj1289	BamA	Cj1289	1:3		0.25 % SB12, 150 mM NaF 30 mM Tris
12	BamD + Cj1289	BamD	Cj1289	1:3		0.25 % SB12, 150 mM NaF 30 mM Tris

Table 8.3: Circular dichroism experiment list.

Protein	M/W (kDa)	Extinction Coefficient (ε/1000)
BamA	84.75	98.56
BamD	25.36	36.33
Cj1289	32.52	17.88

Table 8.4: Sample molecular weights and extinction coefficients.

8.3 Chapter 3 appendices

8.3.1 pET26b+ vector map

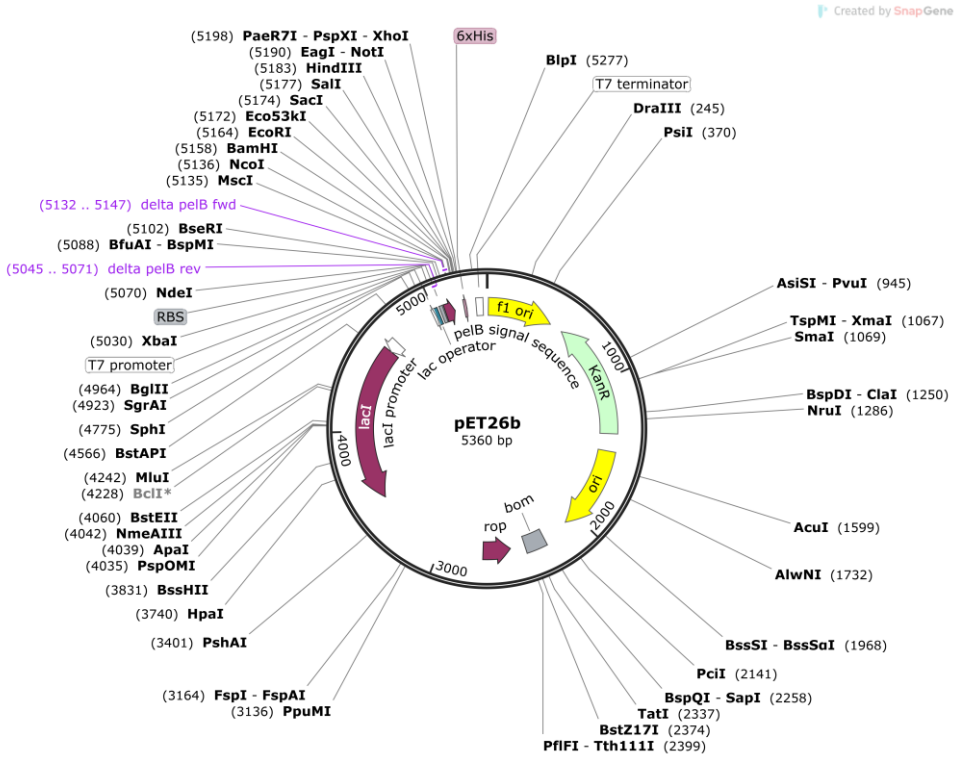


Figure 8.2: pET26b+ vector map. pET26b+ (Novagen) vector map. Map display constructed in SnapGene viewer, restriction sites and gene features shown.

8.3.2 **Protein sequences of proteins in this study**

Protein	Sequence
BamA	MKKHLISICALVAMANAATIKDIKFIGLNHLSNTSAINIAGLKIGEEINPAKINTAILNLYKQNYFENIAVENNNGILEIIVTEK PTIAKVTITGIASNDRKQVESILGIKRGTLLDEGNIKEAIERIKAYYEAKSYFDTIVEYKKKTLENTDGLELEFIVNRGENIIIDN VHLSGAKKFSYSDIEPAVVNKEKEFMGWMWGRNDGKLKVFELSNDSSRIADEYMKKGYLDVQVSSPYLKTYTDTYQAN LTYFIKEGKPYKIKSISIENPLFDDKQNAQTVKDLRSSAGKTINIEDIRKDVKTIETQSADLGYAFVEVYPDIQKNDQTQEAT VVFKVIPHDKVYIRNVIISGNSRTVDRVIRRELYITEGNLYNRTDLSESKNALKRTSYFDDVNIKEEKVDDTHIDLIVDVKEA STGAISGGIGYGSSDGILLNASLSDTNIFGSGIKSSVSVDKSDDTLSGRISLVNPRVLDSQYSLGGTLYSNDYEWDNYSEKNY GFDITIGRQFARYYNVSLTYNLEQSDIYHLSPTLLRTGYELGKSIKSSITPAITFNDTDDYYLPRSGIIASTSLEYAGLGGDQEF ISSSSKFNFYQGLQDYIGYDLIYRYKASFYKVWDEGYLPINQRIYLGGIRSIRGFESRTVSPKNQWGDEIGGTIAFANSVELSF PLIDRIKLRGSVFFDYGMIGRKNLDEIKRMSTGIGIEWITPIGPLQLVFAKPLNDKKGDDTNSFEFNLGTRF
Cj1289	MKKILLSFAFFASLASANTINAIAVVVDKEPITTYDIDQTMKALKIDRNKALGVLINEKMEISQMKQLGIVVNDLELDDAIN KMLAQNKTTLNAFKANLKSKNQSYEQFRTNFKKDLEKRKLYEKIASMAKTDFSDDGAKKFFEQNKDKFTFYTQINANIYL SNNPQTLENIKNTKKTILKPQNASLNTSNADPRLLGLLSQIPVGSFSPVLNGKNGYELYEVKSKDGTQTPEYEQVKNEVLNA YVSEQRQNFIQDYFDKLRSKINIEYLR
BamD	MKKGIFLLVFLSVFFSACSTKNDEGLYNLSASEWYKQIIKDLQDKDLEKADDHYNGMASEHVADPLLETTLIILAQAHMDE EEYKLAEFYLDEYNKKFGNSRNADYIRYLKIKAKFDAFAVPNRNQALMLESQKEIDTFLKDYPYTEYEPLVQTMLTKFNL AVFYLNSTIENLYQRIGHDESAQIYKQRLQESEFYQQSIIKPELPWYRSIFERF

DegP	MKKIFLSLSLASALFAASINFNESTATANRVNPAAGNAVLSYHDSIKDAKKSVVNISTSKTITRANRPSPLDDFFNDPYFKQF FDFDFPQRKGKNDKEVVSSLGSGVIISKDGYIVTNNHVVDDADTITVNLPGSDIEYKAKLIGKDPKTDLAVIKIEANNLSAIT FTNSDDLMEGDVVFALGNPFGVGFSVTSGIISALNKDNIGLNQYENFIQTDASINPGNSGGALVDSRGYLVGINSAILSRGG GNNGIGFAIPSNMVKDIAKKLIEKGKIDRGFLGASILALQGDTKKAYKNQEGALITDVQKGSSADEAGLKRGDLVTKVNDK VIKSPIDLKNYIGTLEIGQKISLSYERDGENKQASFILKGEKENPKGVQSDLIDGLSLRNLDPRLKDRLQIPKDVNGVLVDSV KEKSKGKNSGFQEGDIIIGVGQSEIKNLKDLEQALKQVNKKEFTKVWVYRNGLATLLVLK
LptD	MWRKFSLLLGTSIALNAAQVDIYALDAKKEGDILTANNDVVIFSDFYFITANKAIYNEKTEDVELFGDVNILRGQNERSHSD YAKINLNSNQADFSNFFFSNNNLEVWFQSKTSHLNDKVFESKISAVSSCNVEDPDWEIRFSKGWLNRETNFVHLYNARLYV KNTPVFYLPYFGFSADTHRQSGLLIPKIVLKSSEGLYYEQPIYIATQENWDLELDPQIRTNRGFGLYSILRFLDSPYSTGELNF GAFRENSSYFHDENLKNQTHYGIELKYSRDDLIKSLLSDNFQEGLWIDATYLNDVDYLNLGSRDYRDLNSLVTSKINYFLA DENNFYGAYARYYIDTSKLSNNTTLQEYPSFQYHRFLNNLFDERLRYSFDASFHNFYRPAGSYANELNLDLPISYHNAFFG DFLHFTFTERFYASFVNYSNDPERNHEHYFRNTHDFNLYTDLSKAYENFFHTLNLGVNYILPGAKSGKITQDYLEEYDKEN EHTSLYAVQYFYNNEGQKKLKHRISLDYLNKQNEFYELENLLTYYFNENININLNSEVLYSYEQSRFTNVISQIEVNTNSKFN WMFSHAYQNDEYGKYSFIGTRANYIATPNYNLFGGIWFDTQRAHTNMWELGYTYQRKCWNYSLMYRERIDPQLTSGGIT AKNQSGIYFIFNFYPLGGVKYDFSLAESENKI

Table 8.5: Protein sequences of BamA, Cj1289, BamD, DegP and LptD. Signal leader peptide sequences are highlighted in yellow.

8.3.3 LC-MS/MS Mascot analysis of protein in this study

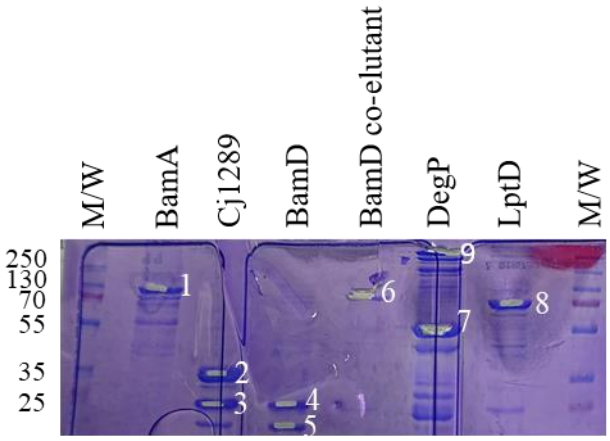


Figure 8.3: SDS-PAGE excised gel bands for LC-MS/MS analysis. Gel bands for proteins in this study were excised and analysed via LC-MS/MS. The corresponding excised gel bands are shown for BamA (84 kDa), Cj1289 (32 kDa, 25 kDa), BamD (25 kDa, 17 kDa), BamD co-elutant (approximately 85 kDa), DegP (monomer, 55 kDa; multimer, approximately 300-400 kDa) and LptD (80 kDa).

8.3.4 **Biotools analysis of LC-MS/MS**

GM1 BamA

Biotools results

Positive hit for BamA and coverage of provided sequence.

Tree hierarchy	Meas. M/z	Calc. MH+	Meas. Mr	Calc. Mr	Int.	z	Dev.(Da)	Dev.(ppm)	Score	MascotS...	Rt(min)	Range	P	Sequence
MSMS 129	632.759	1264.610	1263.502	1263.602	257358.0...	2+	-0.100	-79.115	380	-	5.300	137 - 146	0	SYFDTIVEYK
MSMS 124	734.789	1468.732	1467.562	1467.725	124935.0...	2+	-0.162	-110.518	264	-	5.000	181 - 193	0	FSYSDIEPAVVNK
MSMS 60	573.766	1146.611	1145.516	1145.604	381082.0...	2+	-0.088	-76.554	263	-	2.400	92 - 102	0	VTITGIASNDR
MSMS 134	578.229	1155.579	1154.442	1154.572	126233.0...	2+	-0.130	-112.313	133	-	5.500	498 - 507	0	NYGFDITIGR
MSMS 58	577.203	1153.548	1152.390	1152.541	1563635....	2+	-0.151	-130.812	65	-	2.300	211 - 220	0	VFELSNDSSR
MSMS 67	487.205	973.531	972.394	972.524	282129.0...	2+	-0.130	-133.323	44	-	2.600	292 - 299	0	TINIEDIR
MSMS 59	399.705	798.483	797.394	797.476	307785.0...	2+	-0.082	-102.307	36	-	2.300	466 - 472	0	ISLVNPR
MSMS 51	485.703	970.520	969.390	969.513	282129.0...	2+	-0.123	-126.621	32	-	1.900	45 - 53	0	IGEEINPAK

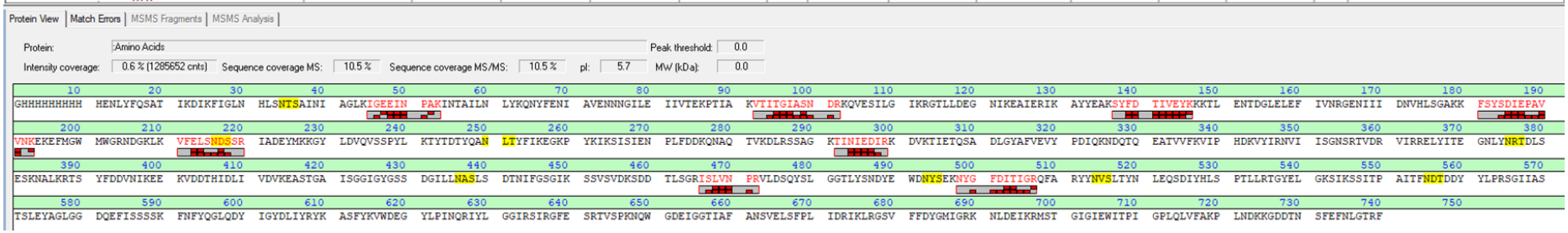


Figure 8.4: Biotools confirmation of BamA (LC-MS/MS)

GM2 Cj1289

Biotools results

Positive hit for cj1289 and coverage of provided sequence.

Tree hierarchy	Meas. M/z	Calc. MH+	Meas. Mr	Calc. Mr	Int.	z	Dev.(Da)	Dev.(ppm)	Score	MascotS...	Rt(min)	Range	P	Sequence
MSMS 147	934.938	1868.996	1867.860	1867.989	7617637....	2+	-0.129	-68.929	1034	-	6.600	79 - 95	0	QLGIVVNDLELDDAINK
MSMS 173	675.345	2023.195	2023.012	2022.188	4112884....	3+	0.824	407.369	553	-	7.500	208 - 227	0	LLGLLSQIPVGGFSPILNGK
MSMS 148	623.980	1868.996	1868.917	1867.989	3924958....	2+	0.927	496.479	194	-	6.600	79 - 95	0	QLGIVVNDLELDDAINK
MSMS 75	647.335	1940.020	1938.982	1939.012	1677906...	3+	-0.031	-15.873	192	-	3.200	190 - 207	0	TILKPQNASLNTSNADPR
MSMS 49	536.142	1071.485	1070.268	1070.478	617405.0...	2+	-0.210	-195.961	77	-	2.000	115 - 122	0	NQSYEQFR
MSMS 120	659.307	1317.611	1316.598	1316.604	2470484....	2+	-0.005	-4.076	64	-	5.500	263 - 272	0	QNFIQDYFDK
MSMS 82	460.703	920.520	919.390	919.513	7039762....	2+	-0.122	-133.078	44	-	3.500	277 - 283	0	INIEYLR
MSMS 72	478.736	956.578	955.456	955.570	1253565...	2+	-0.114	-119.177	44	-	3.100	63 - 71	0	ALGVLINEK

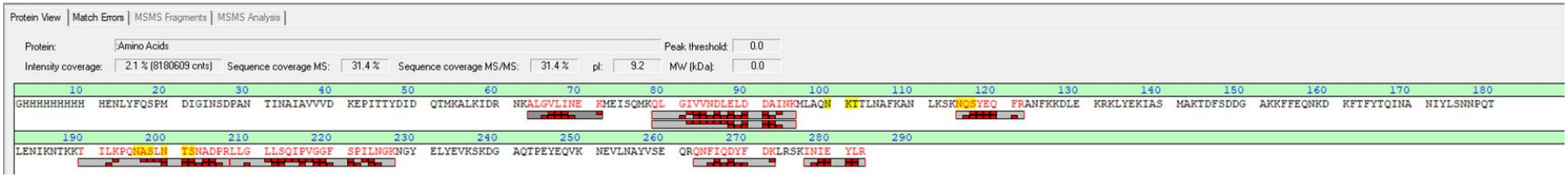


Figure 8.5: Biotools confirmation of Cj1289 (LC-MS/MS)

GM6 LptD

Biotools results

Positive hit for LptD and coverage of provided sequence.

Tree hierarchy	Meas. M/z	Calc. MH+	Meas. Mr	Calc. Mr	Int.	z	Dev.(Da)	Dev.(ppm)	Score	MascotS...	Rt(min)	Range	P	Sequence
MSMS 133	847.346	1693.876	1692.676	1692.868	750036.0...	2+	-0.192	-113.366	547	-	5.400	539 - 553	0	FTNVISQIEVNTNSK
MSMS 132	584.265	1167.579	1166.514	1166.572	1170522....	2+	-0.058	-49.438	252	-	5.400	348 - 356	0	FLNNLFDER
MSMS 61	488.765	976.531	975.514	975.524	1095508....	2+	-0.009	-9.559	61	-	2.300	298 - 306	0	DLNSLVTSK
MSMS 83	483.198	965.530	964.380	964.523	1018216....	2+	-0.143	-147.842	43	-	3.200	498 - 505	0	ISLDYLNK

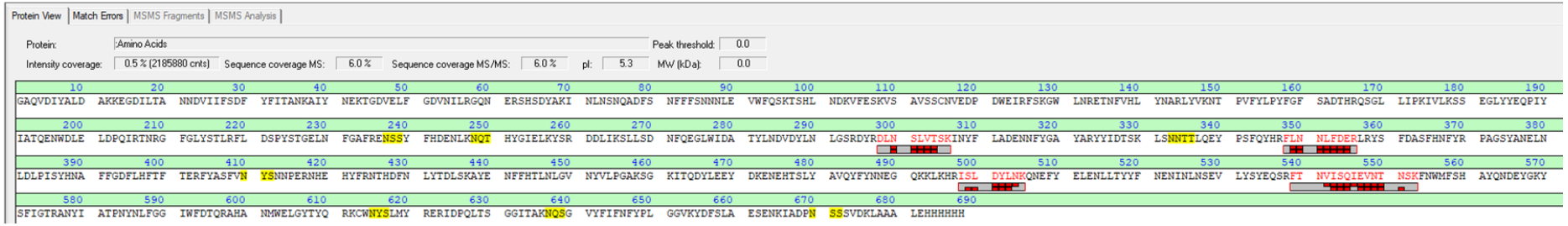


Figure 8.6: Biotools confirmation of LptD (LC-MS/MS)

GM8 DegP multimer

Biotools results

Positive hit for DegP and coverage of provided sequence.

Tree hierarchy	Meas. M/z	Calc. MH+	Meas. Mr	Calc. Mr	Int.	z	Dev.(Da)	Dev.(ppm)	Score	MascotS...	Rt(min)	Range	P	Sequence
MSMS 143	664.615	1991.008	1990.822	1990.001	1137624...	2+	0.821	412.457	3072	-	6.200	404 - 422	0	NSGFQEGDIIIGVGQSEIK
MSMS 158	766.876	1532.868	1531.736	1531.861	3454943....	2+	-0.125	-81.368	2174	-	6.800	262 - 276	0	GFLGVTILALQGDTK
MSMS 144	995.913	1991.008	1989.810	1990.001	3601874....	2+	-0.190	-95.671	2114	-	6.300	404 - 422	0	NSGFQEGDIIIGVGQSEIK
MSMS 97	878.352	1755.902	1754.688	1754.895	1258396....	2+	-0.207	-117.841	2063	-	4.100	15 - 31	0	VNPAAGNAVLSYHDSIK
MSMS 105	687.821	1374.784	1373.626	1373.777	7650637....	2+	-0.150	-109.367	1532	-	4.400	82 - 95	0	EVVSSLGSGVIISK
MSMS 43	658.250	1315.685	1314.484	1314.678	1178153...	2+	-0.194	-147.256	1532	-	1.700	281 - 292	0	NQEGALITDVQK
MSMS 60	658.269	1315.685	1314.522	1314.678	1178153...	2+	-0.156	-118.351	1019	-	2.500	281 - 292	0	NQEGALITDVQK
MSMS 131	686.874	1372.743	1371.732	1371.736	1310413...	2+	-0.003	-2.510	520	-	5.700	360 - 372	0	GVQSDLIDGLSLR
MSMS 149	538.287	1075.651	1074.558	1074.644	1579648....	2+	-0.085	-79.451	267	-	6.400	447 - 456	0	NGFATLLVLK
MSMS 110	618.706	1235.663	1235.396	1234.656	2639996...	2+	0.741	599.842	253	-	4.700	323 - 333	0	NYIGTLEIGQK
MSMS 92	618.257	1235.663	1234.498	1234.656	2639996...	2+	-0.157	-127.487	252	-	3.800	323 - 333	0	NYIGTLEIGQK
MSMS 67	572.733	1144.621	1143.450	1143.614	1042385...	2+	-0.163	-142.708	189	-	2.800	387 - 397	0	DVNGVLVDSVK
MSMS 142	816.816	1632.816	1631.616	1631.809	5438852....	2+	-0.193	-118.044	127	-	6.200	231 - 247	0	GGGNNGIGFAIPSNMVK
MSMS 132	458.194	1372.743	1371.559	1371.736	1904568....	3+	-0.177	-129.229	96	-	5.800	360 - 372	0	GVQSDLIDGLSLR
MSMS 147	681.872	2042.961	2042.593	2041.953	2973009...	2+	0.639	313.001	68	-	6.400	49 - 65	0	ANRPSPLDDFFNDPYFK
MSMS 160	668.744	1336.596	1335.472	1335.588	2236684...	2+	-0.116	-86.891	62	-	6.800	66 - 75	0	QFFDFDFSQR
MSMS 187	668.695	1336.596	1335.374	1335.588	2236684...	2+	-0.214	-160.267	61	-	7.700	66 - 75	0	QFFDFDFSQR
MSMS 155	1021.905	2042.961	2041.794	2041.953	3189572....	2+	-0.159	-77.887	52	-	6.800	49 - 65	0	ANRPSPLDDFFNDPYFK
MSMS 46	467.694	934.520	933.372	933.513	2414057....	2+	-0.141	-150.786	43	-	1.900	36 - 44	0	SVVNISTSK
MSMS 172	681.552	2042.961	2041.633	2041.953	2973009...	3+	-0.321	-157.137	33	-	7.200	49 - 65	0	ANRPSPLDDFFNDPYFK
MSMS 56	434.190	867.457	866.364	866.450	8995059....	2+	-0.085	-98.600	31	-	2.200	334 - 340	0	ISLSYER
MSMS 53	380.206	759.461	758.396	758.454	1921540....	2+	-0.057	-75.753	27	-	2.100	134 - 140	0	TDLAVIK

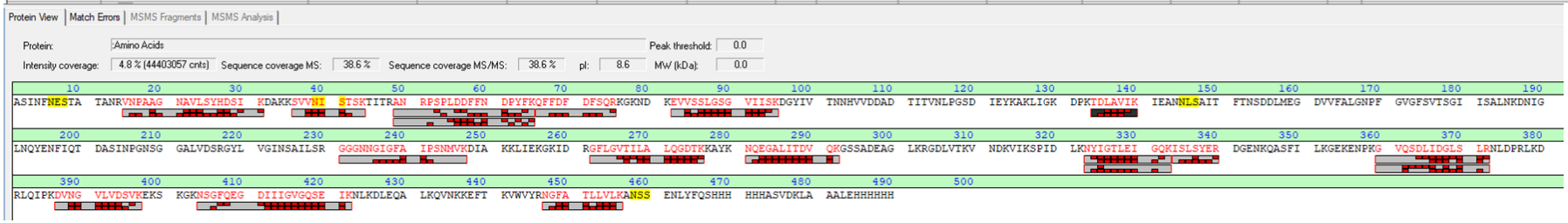


Figure 8.7: Biotools confirmation of DegP (multimer, LC-MS/MS)

GM9 DegP monomer

Biotools results

Positive hit for DegP and coverage of provided sequence.

Tree hierarchy	Meas. M/z	Calc. MH+	Meas. Mr	Calc. Mr	Int.	z	Dev.(Da)	Dev.(ppm)	Score	MascotS...	Rt(min)	Range	P	Sequence
MSMS 154	995.908	1991.008	1989.800	1990.001	4794087....	2+	-0.200	-100.696	2086	-	6.200	404 - 422	0	NSGFQEGDIIIGVGQSEIK
MSMS 123	687.823	1374.784	1373.630	1373.777	1147011...	2+	-0.146	-106.455	2046	-	4.900	82 - 95	0	EVVSSLGSGVIISK
MSMS 143	686.807	1372.743	1371.598	1371.736	1147011...	2+	-0.137	-100.196	636	-	5.800	360 - 372	0	GVQSDLIDGLSLR
MSMS 97	618.474	1235.663	1234.932	1234.656	2455518...	2+	0.277	224.028	635	-	3.800	323 - 333	0	NYIGTLEIGQK
MSMS 64	658.268	1315.685	1314.520	1314.678	2319354...	2+	-0.158	-119.873	574	-	2.500	281 - 292	0	NQEGALITDVQK
MSMS 77	572.716	1144.621	1143.416	1143.614	2468039...	2+	-0.197	-172.439	380	-	3.000	387 - 397	0	DVNGVLVDSVK
MSMS 117	618.751	1235.663	1235.486	1234.656	2455518...	2+	0.831	672.736	126	-	4.700	323 - 333	0	NYIGTLEIGQK
MSMS 43	467.681	934.520	933.346	933.513	8000190....	2+	-0.167	-178.638	43	-	1.800	36 - 44	0	SVVNISTSK
MSMS 181	681.925	2042.961	2042.752	2041.953	5926148...	2+	0.798	390.868	32	-	7.100	49 - 65	0	ANRPSPLDDFFNDPYFK
MSMS 56	434.161	867.457	866.306	866.450	3998035....	2+	-0.143	-165.540	29	-	2.200	334 - 340	0	ISLSYER
MSMS 55	380.180	759.461	758.344	758.454	1272035...	2+	-0.109	-144.313	27	-	2.200	134 - 140	0	TDLAVIK
MSMS 166	1021.870	2042.961	2041.724	2041.953	4081702....	2+	-0.229	-112.168	23	-	6.700	49 - 65	0	ANRPSPLDDFFNDPYFK

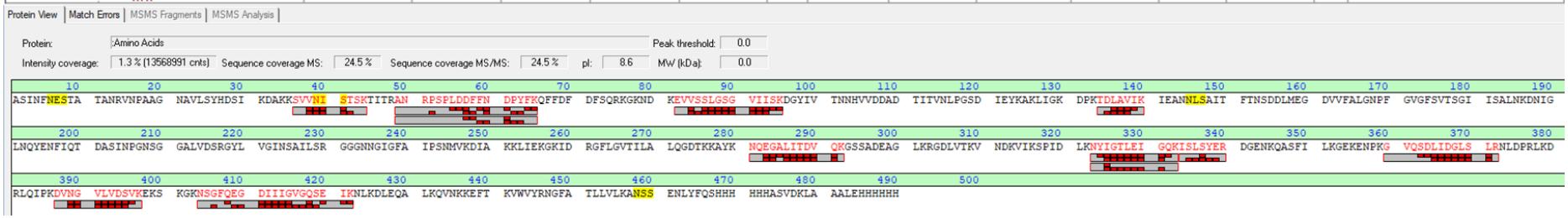


Figure 8.8: Biotools confirmation of DegP (monomer, LC-MS/MS)

8.3.5 Further optimisation of the BamA purification process

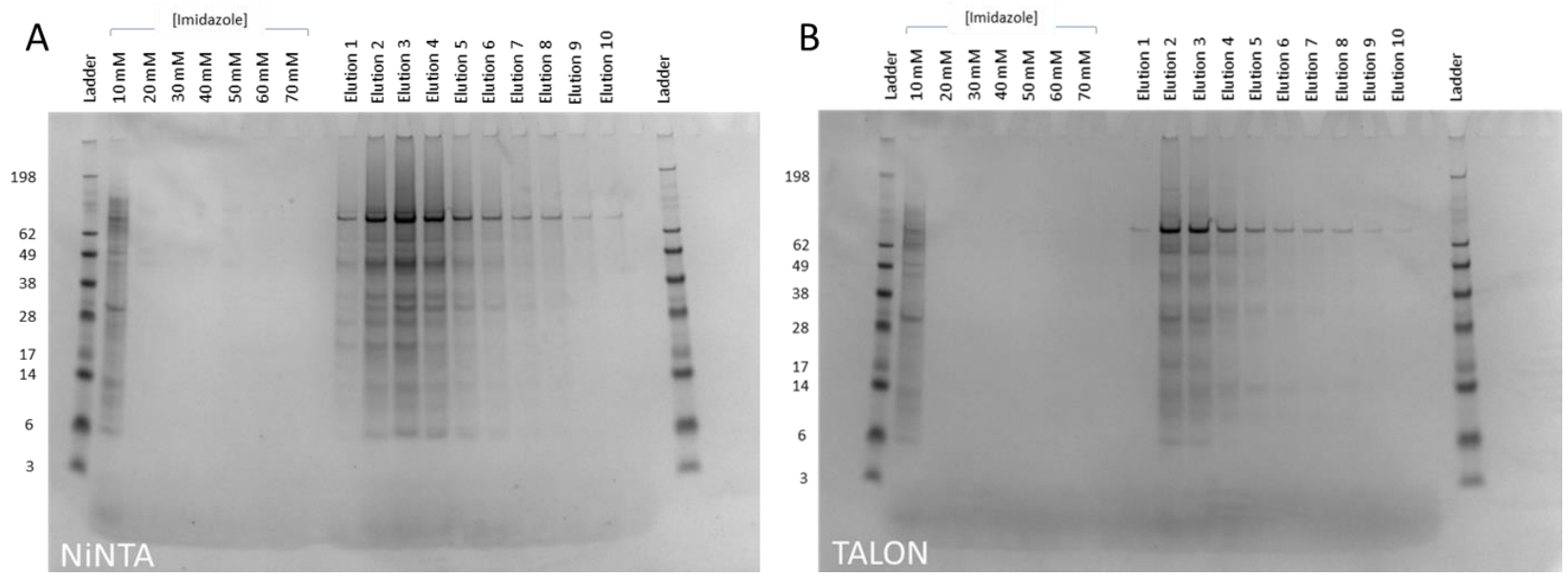


Figure 8.9: BamA IMAC pulldown optimisation – Resin, wash buffer imidazole concentration, and elution volume. Experiment to optimise resin, wash buffer imidazole concentration and elution volume for BamA IMAC purification. DDM solubilised BamA membrane was batch bound to Ni-NTA and TALON resin for 4 hours at 4°C, sequentially washed with 10 CVs of wash buffer with increasing imidazole concentrations (10 mM to 70 mM), and eluted in steps at 0.5 CVs/ step. The SDS-PAGE analysis for the Ni-NTA (A) and TALON (B) IMAC purifications are shown. TALON resin give cleaner purification, but potentially less pulldown. Concentrations of imidazole up to 50 mM remove contaminants and concentrations up to 70 mM can be used in the wash buffer without eluting BamA. As the majority of BamA protein elutes by six elution volumes (3 ml/ 3 CVs), three to four CVs should be used as the minimum elution volume on future preparations of BamA.

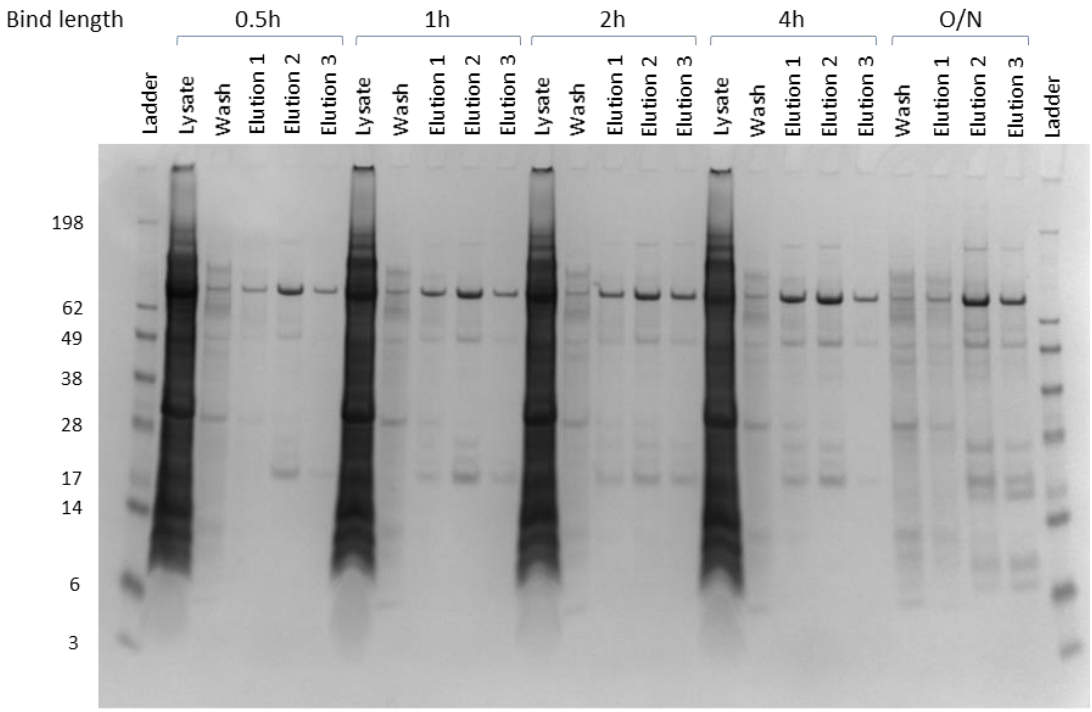


Figure 8.10: TALON pulldown binding optimisation of DDM extracted BamA. Experiment to optimise BamA TALON resin binding time for recovery and minimisation of contaminants. BamA was recovered at all binding lengths from 0.5 h to overnight binding. A 0.5 h BamA TALON binding time gives the lowest level of contaminants post-elution, and a binding time of 4 h or overnight maximises BamA recovery. Extending TALON resin binding time to overnight greatly increases contaminants in the subsequent elution. A binding time of 4 h should be used for maximal recovery of BamA whilst minimising contaminants, and a binding up to 2 h should be used to keep concentration of the 50 kDa contaminant (or degradation) band to a minimum.

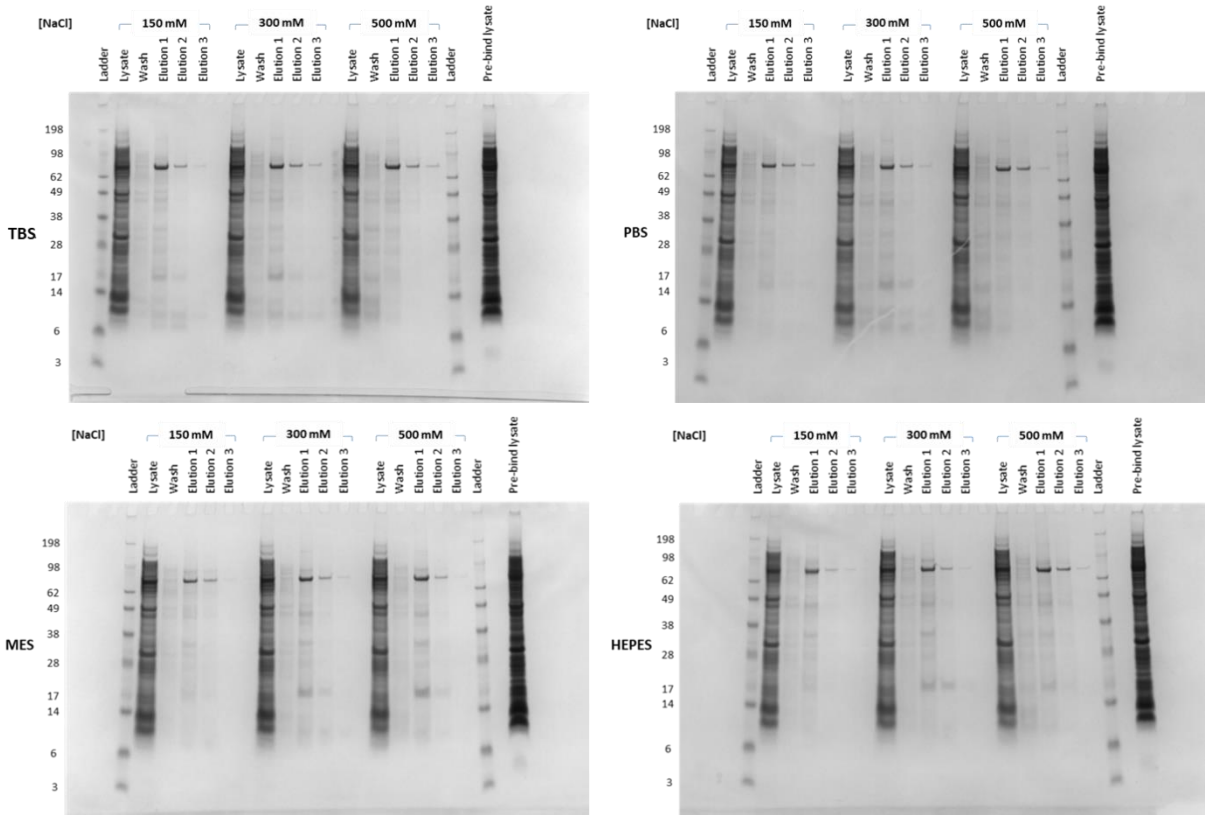


Figure 8.11: BamA buffer and salt concentration IMAC screening. Experiment to determine suitable buffer conditions (TBS, PBS, MES, HEPES) and optimal salt concentration (150 mM, 300 mM, 500 mM) for the IMAC purification of BamA. BamA purifies well in all buffers tested, however there appears to be lower pulldown in MES buffer. The 0.9 pH unit difference across the buffers tested (MES pH 6.7, HEPES pH 7.0, PBS pH 7.2, and TBS pH 7.6) appears to have no major impact on stability during purification. Contrary to expectations, increasing salt concentration does not reduce pulldown contaminants within the concentration range tested.

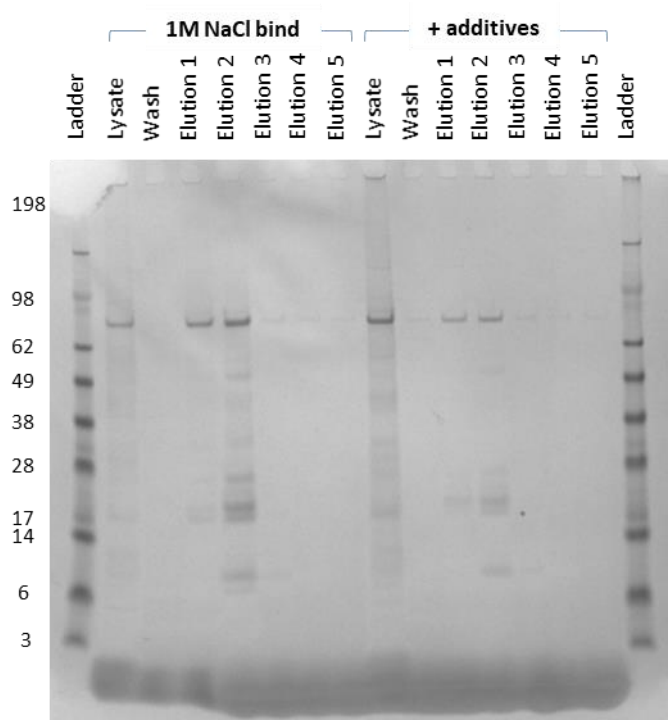


Figure 8.12: 1 M NaCl and buffer cofactor TALON IMAC binding screen. As an extension to the previous screens, an increased concentration of 1 M salt and the use of several cofactors (10 mM MgCl2, 1 mM MnCl2 and 1 mM ATP) that may be useful in removal of chaperones and other IMAC contaminants (identified in the LC-MS/MS data for BamA, data not shown) were tested. Increasing the salt concentration to 1 M does not appear to improve the contaminant pulldown during BamA purification. However, the addition of the chosen cofactors appears to reduce the banding between 28 kDa and 62 kDa.

It should be noted that bands below 28 kDa would be removed during SEC or spin concentration. The experiment is promising but needs repeating with higher concentration samples.

8.3.6 BamA detergent exchange screen for future structural determination by crystallography

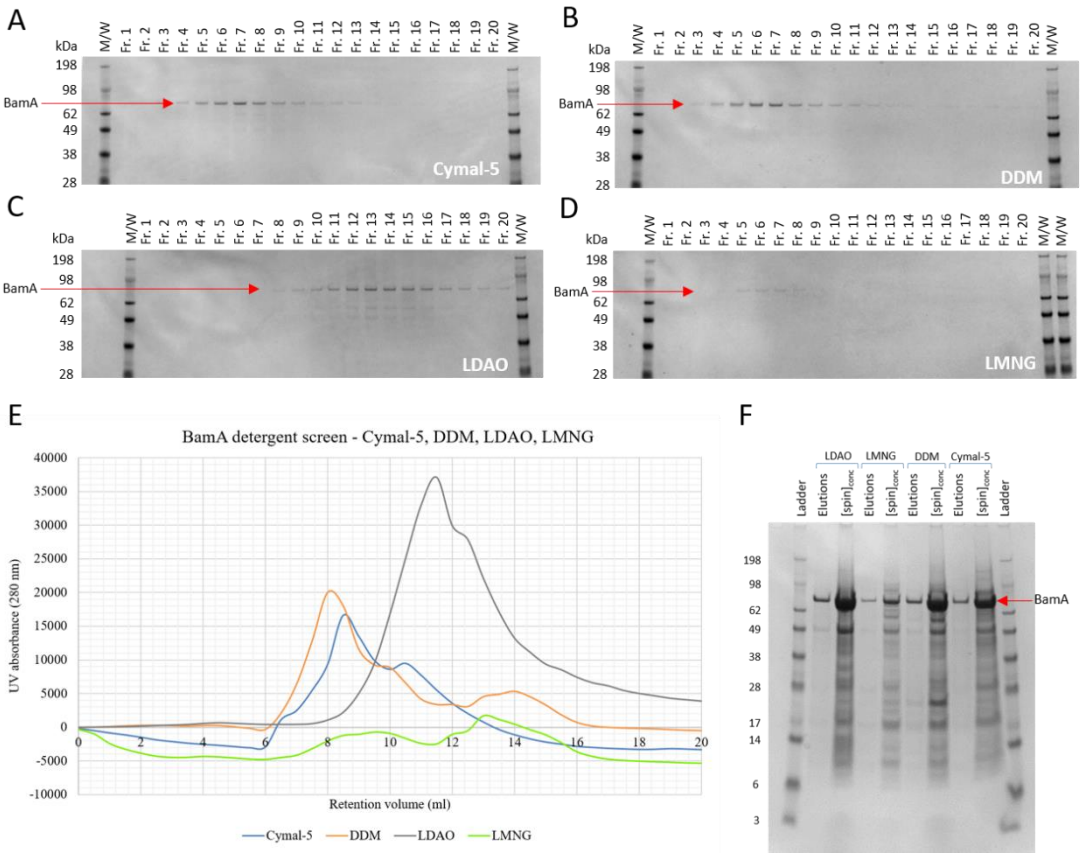


Figure 8.13: CjBamA follow-up detergents screen in Cymal-5, DDM, LDAO and LMNG. A screen to test the viability detergent exchange into suitable crystallisation detergents after membrane extraction of CjBamA in FC12. Post-detergent exchanged IMAC samples were run on a Seepax SRT 300 SEC column. SEC purified SDS-PAGE analysis of samples detergent exchanged in Cymal-5 (A), DDM (B), LDAO (C) and LMNG (D), associated tryptophan fluorescence profiles during SEC (E), and spin concentrated sample purity (F) post-SEC are shown. Membrane extraction in FC12 followed by detergent exchange into detergents suitable for crystallisation and subsequent purification in these detergents is a viable strategy for producing highly concentrated samples for crystallisation. However, while CjBamA is the major band, the spin concentrated end sample requires further processing for purity.

8.3.7 BamA reverse IMAC purification following detergent exchange

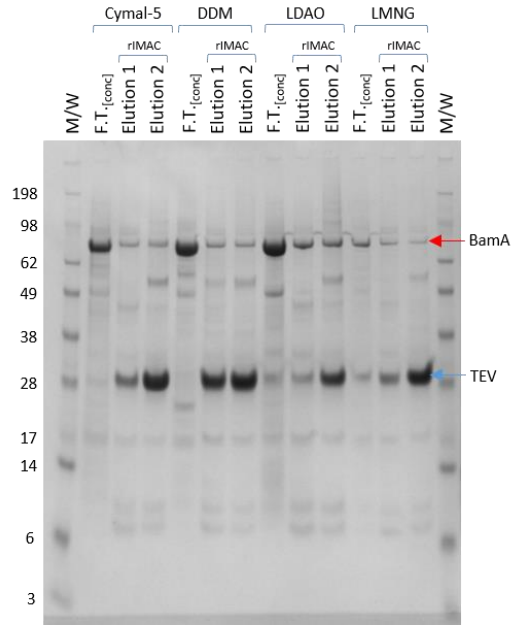


Figure 8.14: Reverse IMAC purification of CjBamA. The final polishing step for post-SEC FC12 extracted CjBamA detergent exchanged samples was a reverse IMAC (rIMAC) purification. Samples were added to Ni-NTA IMAC resins and incubated overnight with TEV protease (‘Super’-TEV, MPL). Flowthrough was collected the following morning along with IMAC elutions. The target protein band was mostly retained in the flowthrough (84 kDa) and the His-tagged TEV protease remained bound to the column before elution (28 kDa).The spin concentrated flowthroughs show significantly purified (>90%) CjBamA samples for all detergents tested.

8.3.8 **LptD initial detergent screening**

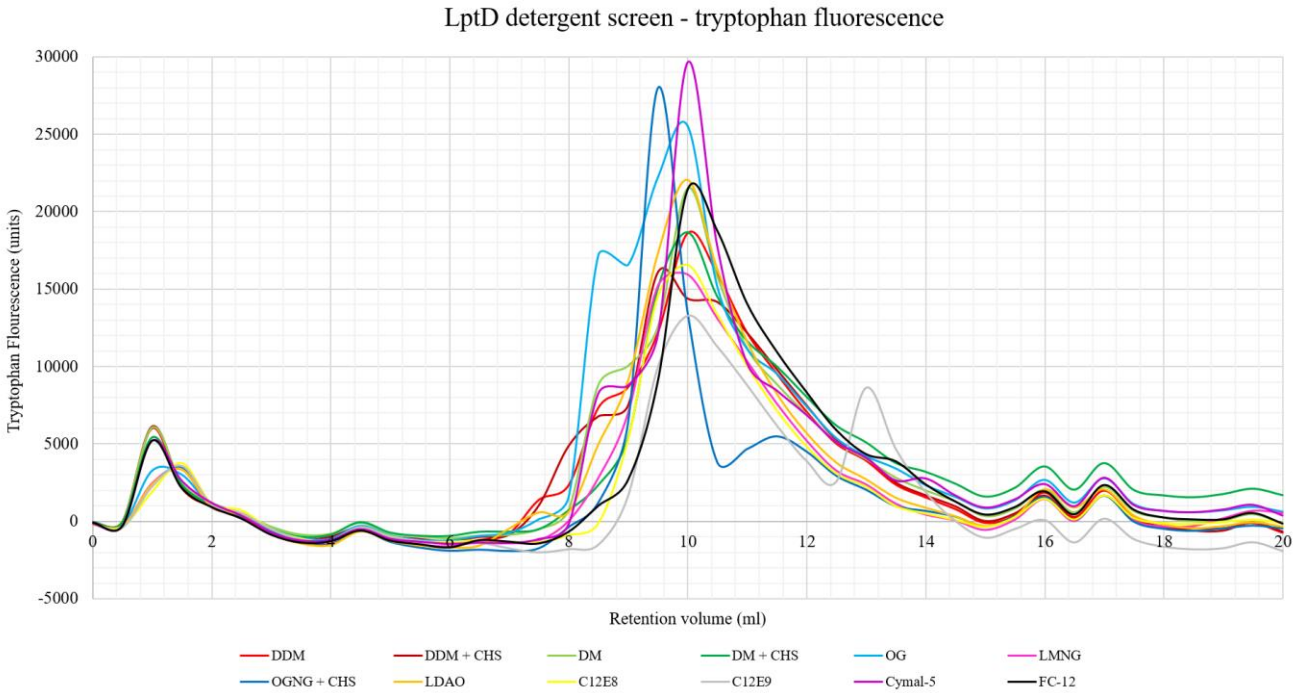


Figure 8.15: LptD initial detergent screen. A LptD overexpression membrane fraction was purified in twelve detergents in parallel. Tryptophan fluorescence profiles of LptD purified in DDM (red), DDM + CHS (dark red), DM (light green), DM + CHS (dark green), OG (light blue), LMNG (pink), OGNG + CHS (dark blue), LDAO (orange), C12E8 (yellow), C12E9 (grey), Cymal-5 (purple) and FC12 (black) are shown. The highest tryptophan fluroescence and, by inference, pulldown of LptD is observed in Cymal-5, OGNG + CHS and OG.

8.4 Chapter 4 appendices

8.4.1 *In silico* SSA calculations

	Cj1289 [PDB = 3RGC crystal structure			
	Helix	Sheet	Turn	Unordered
DSSP	0.53	0.16	0.09	0.22
STRIDE	0.56	0.20	0.09	0.15
P-SEA	0.54	0.15	0.00	0.30
STICKS	0.54	0.15	0.00	0.31
PROSS	0.52	0.17	0.07	0.24
Average	0.538	0.167	0.050	0.245
SD	0.012	0.019	0.041	0.057

	Cj1289 [PDB = 3RGC, MD]			
	Helix	Sheet	Turn	Unordered
DSSP	0.56	0.16	0.09	0.19
STRIDE	0.52	0.11	0.09	0.28
P-SEA	0.53	0.15	0.00	0.32
STICKS	0.53	0.16	0.00	0.31
PROSS	0.49	0.08	0.07	0.36
Average	0.525	0.131	0.050	0.293
SD	0.023	0.031	0.041	0.058

	BamD [MD, waterbox]			
	Helix	Sheet	Turn	Unordered
DSSP	0.53	0.03	0.14	0.30
STRIDE	0.51	0.02	0.18	0.29
P-SEA	0.52	0.03	0.13	0.33
STICKS	0.56	0.03	0.14	0.27
PROSS	0.50	0.02	0.14	0.34
Average	0.523	0.026	0.146	0.306
SD	0.020	0.005	0.017	0.026

	BamD: Cj1289 [1:1, PDB = I-TASSER/3RGC]			
	Helix	Sheet	Turn	Unordered
DSSP	0.61	0.09	0.10	0.20
STRIDE	0.65	0.11	0.13	0.11
P-SEA	0.64	0.08	0.00	0.27
STICKS	0.64	0.09	0.00	0.27
PROSS	0.59	0.09	0.08	0.23
Average	0.625	0.094	0.062	0.216
SD	0.021	0.010	0.053	0.058

	BamA: Cj1289 [1:1, PDB = I-TASSER/ 3RGC]			
	Helix	Sheet	Turn	Unordered
DSSP	0.26	0.35	0.09	0.31
STRIDE	0.27	0.38	0.19	0.16
P-SEA	0.26	0.39	0.00	0.34
STICKS	0.28	0.36	0.00	0.36
PROSS	0.23	0.26	0.09	0.42
Average	0.259	0.346	0.074	0.319
SD	0.017	0.045	0.070	0.086

	BamA: BamD: Cj1289 [1:1:1, PDB = I-TASSER/ 3RGC]			
	Helix	Sheet	Turn	Unordered
DSSP	0.33	0.29	0.10	0.28
STRIDE	0.35	0.31	0.19	0.15
P-SEA	0.34	0.32	0.00	0.33
STICKS	0.36	0.30	0.00	0.34
PROSS	0.30	0.22	0.09	0.39
Average	0.338	0.289	0.075	0.297
SD	0.022	0.036	0.070	0.083

Table 8.6: Calculated secondary structures for Cj1289, BamD-Cj1289 (1:1), BamA-Cj1289 (1:1), and BamA-BamD-Cj1289 (1:1:1) mixtures from homology model PDB files. Secondary structure assignment were derived from PDB files generated by I-TASSER using different algorithms for Cj1289 [PDB], Cj1289 [MD], BamD [MD, waterbox], BamD-Cj1289 (1:1, PDB), BamA-Cj1289 (1:1, PDB), and BamA-BamD-Cj1289 mixtures (1:1:1, PDB). Average secondary structure assignments and standard deviations of the predictions are shown.

	BamA [PDB = I-TASSER]					BamA [PDB = MD, Lipid membrane]					BamA [PDB = MD, SB12]			
	Helix	Sheet	Turn	Unordered		Helix	Sheet	Turn	Unordered		Helix	Sheet	Turn	Unordered
DSSP	0.16	0.41	0.09	0.34	DSSP	0.16	0.42	0.09	0.33	DSSP	0.14	0.42	0.09	0.35
STRIDE	0.18	0.44	0.22	0.17	STRIDE	0.16	0.43	0.22	0.20	STRIDE	0.16	0.44	0.21	0.19
P-SEA	0.16	0.47	0.00	0.36	P-SEA	0.15	0.48	0.00	0.37	P-SEA	0.15	0.48	0.00	0.37
STICKS	0.19	0.43	0.00	0.38	STICKS	0.17	0.44	0.00	0.39	STICKS	0.17	0.44	0.00	0.39
PROSS	0.13	0.29	0.09	0.49	PROSS	0.14	0.39	0.10	0.37	PROSS	0.14	0.39	0.08	0.39
Average	0.163	0.407	0.081	0.347	Average	0.155	0.432	0.081	0.332	Average	0.152	0.435	0.074	0.339
SD	0.020	0.062	0.082	0.104	SD	0.010	0.029	0.079	0.070	SD	0.011	0.030	0.076	0.074
	BamD [PDB = I-TASSER]					BamD [PDB = MD, Lipid membrane]					BamD [PDB = MD, SB12]			
	Helix	Sheet	Turn	Unordered		Helix	Sheet	Turn	Unordered		Helix	Sheet	Turn	Unordered
DSSP	0.71	0.00	0.12	0.17	DSSP	0.69	0.00	0.10	0.22	DSSP	0.68	0.00	0.14	0.19
STRIDE	0.76	0.00	0.18	0.06	STRIDE	0.70	0.00	0.10	0.20	STRIDE	0.69	0.00	0.18	0.13
P-SEA	0.76	0.00	0.00	0.24	P-SEA	0.65	0.05	0.00	0.30	P-SEA	0.65	0.05	0.00	0.30
STICKS	0.77	0.02	0.00	0.21	STICKS	0.69	0.00	0.00	0.32	STICKS	0.69	0.00	0.00	0.31
PROSS	0.68	0.00	0.10	0.23	PROSS	0.68	0.02	0.07	0.23	PROSS	0.66	0.00	0.14	0.20
Average	0.736	0.004	0.080	0.183	Average	0.682	0.014	0.052	0.254	Average	0.673	0.010	0.092	0.226
SD	0.034	0.008	0.071	0.065	SD	0.017	0.020	0.044	0.047	SD	0.015	0.020	0.077	0.069
	BamAD [PDB = I-TASSER]					BamAD [PDB = MD, Lipid membrane]					BamAD [PDB = MD, SB12]			
	Helix	Sheet	Turn	Unordered		Helix	Sheet	Turn	Unordered		Helix	Sheet	Turn	Unordered
DSSP	0.28	0.32	0.10	0.30	DSSP	0.27	0.33	0.09	0.31	DSSP	0.26	0.33	0.10	0.31
STRIDE	0.30	0.34	0.21	0.14	STRIDE	0.27	0.34	0.19	0.20	STRIDE	0.27	0.35	0.20	0.18
P-SEA	0.29	0.37	0.00	0.33	P-SEA	0.26	0.39	0.00	0.35	P-SEA	0.26	0.39	0.00	0.35
STICKS	0.31	0.34	0.00	0.34	STICKS	0.28	0.35	0.00	0.37	STICKS	0.28	0.35	0.00	0.37
PROSS	0.25	0.23	0.09	0.43	PROSS	0.26	0.31	0.09	0.34	PROSS	0.25	0.30	0.09	0.35
Average	0.287	0.321	0.080	0.311	Average	0.269	0.342	0.074	0.315	Average	0.264	0.342	0.078	0.314
SD	0.022	0.048	0.079	0.094	SD	0.009	0.026	0.070	0.062	SD	0.011	0.028	0.074	0.070

Table 8.7: Calculated secondary structures for BamA, BamD and BamAD from homology model and equilibrated MD simulations last step PDB files. Secondary structure assignment were derived from PDB files generated by I-TASSER and subsequent MD simulations (≥50 ns) in lipid- and SB12-aqueous environments using different algorithms for BamA ([PDB], [MD, lipid], [MD, SB12]), BamD ([PDB], [MD, lipid], [MD, SB12]), and BamA-BamD ([PDB], [MD, lipid], [MD, SB12]). Average secondary structure assignments and standard deviations of the predictions are shown.

8.4.2 **Goodness of fit - SSEs**

	Sample													
	BamA (DB)	BamD (NDB)	BamD (DB)	Cj1289 (NDB)	Cj1289 (DB)	BamA: BamD (DB)	BamA: Cj1289 (1:1, DB)	BamA: Cj1289 (1:3, DB)	BamD: Cj1289 (1:1, NDB)	BamD: Cj1289 (1:3, NDB)	BamD: Cj1289 (3:1, NDB)	BamD: Cj1289 (1:3, DB)	BamD: Cj1289 (3:1, DB)	BamA: BamD: Cj1289 (1:1:1, DB)
Temperature (°C)														
20	0.014	0.025	0.027	0.036	0.012	0.041	0.010	0.084	0.030	0.013	0.012	0.012	0.020	0.092
23	0.101	0.199	0.043	0.019	0.012	0.024	0.012	0.107	0.015	0.013	0.011	0.013	0.021	0.190
26	0.092	0.026	0.028	0.097	0.013	0.062	0.013	0.199	0.015	0.011	0.011	0.013	0.021	0.136
29	0.016	0.092	0.070	0.018	0.012	0.014	0.063	0.100	0.015	0.011	0.014	0.012	0.019	0.046
32	0.017	0.028	0.029	0.016	0.012	0.014	0.013	0.222	0.027	0.013	0.012	0.063	0.019	0.019
35	0.017	0.033	0.153	0.017	0.011	0.014	0.415	0.023	0.014	0.012	0.010	0.013	0.018	0.024
38	0.101	0.036	0.031	0.015	0.012	0.100	0.362	0.022	0.016	0.043	0.012	0.014	0.019	0.108
41	0.346	0.033	0.027	0.014	0.011	0.018	0.015	0.024	0.017	0.011	0.018	0.056	0.160	0.041
44	0.017	0.035	0.028	0.012	0.012	0.081	0.016	0.021	0.030	0.024	0.200	0.014	0.181	0.026
47	0.160	0.104	0.027	0.013	0.015	0.019	0.043	0.085	0.017	0.011	0.011	0.017	0.018	0.042
50	0.017	0.033	0.024	0.012	0.014	0.086	0.017	0.290	0.019	0.011	0.020	0.032	0.033	0.161
53	0.181	0.029	0.037	0.012	0.017	0.017	0.300	0.346	0.084	0.025	0.026	0.015	0.016	0.069
56	0.241	0.031	0.028	0.009	0.019	0.028	0.061	0.026	0.018	0.083	0.023	0.017	0.204	0.025
59	0.111	0.031	0.104	0.177	0.016	0.044	0.176	0.077	0.015	0.026	0.048	0.132	0.288	0.137
62	0.084	0.028	0.037	0.027	0.017	0.016	0.040	0.096	0.019	0.069	0.062	0.033	0.336	0.029
65	0.161	0.036	0.024	0.086	0.018	0.202	0.244	0.131	0.065	0.045	0.068	0.033	0.015	0.073
68	0.022	0.052	0.025	0.070	0.031	0.021	0.092	0.041	0.149	0.094	0.014	0.018	0.231	0.028
71	0.074	0.047	0.021	0.015	0.019	0.028	0.038	0.219	0.136	0.023	0.043	0.021	0.018	0.036
74	0.349	0.036	0.028	0.081	0.020	0.021	0.229	0.039	0.114	0.051	0.056	0.064	0.022	0.050
77	0.018	0.027	0.189	0.031	0.371	0.023	0.026	0.162	0.407	0.050	0.022	0.019	0.546	0.029
80	0.315	0.030	0.028	0.047	0.020	0.141	0.093	0.040	0.033	0.129	0.014	0.019	0.085	0.034
83	0.029	0.046	0.028	0.178	0.036	0.049	0.038	0.061	0.039	0.029	0.022	0.160	0.060	0.033
86	0.249	0.033	0.092	0.032	0.096	0.144	0.363	0.027	0.167	0.234	0.011	0.154	0.053	0.022
89	0.157	0.036	0.025	0.067	0.016	0.148	0.148	0.164	0.044	0.029	0.030	0.024	0.101	0.022
20	0.014	0.023	0.032	0.122	0.187	0.043	0.015	0.054	0.025	0.040	0.024	0.011	0.019	0.085
Standard deviation of measurement by temperature														

Table 8.8: Goodness of fit of secondary structure estimations. The standard deviation of secondary structure estimation was calculated for each thermal degradation experiment performed at every increment in temperature where data was collected. The calculations where the standard deviation was elevated have been marked in increasing intensities of grayscale. BamA, BamA: Cj1289 (1:1, 1:3) and BamD: Cj1289 (3:1, DB) should be noted.

8.4.3 **BamD concentration inaccurate**

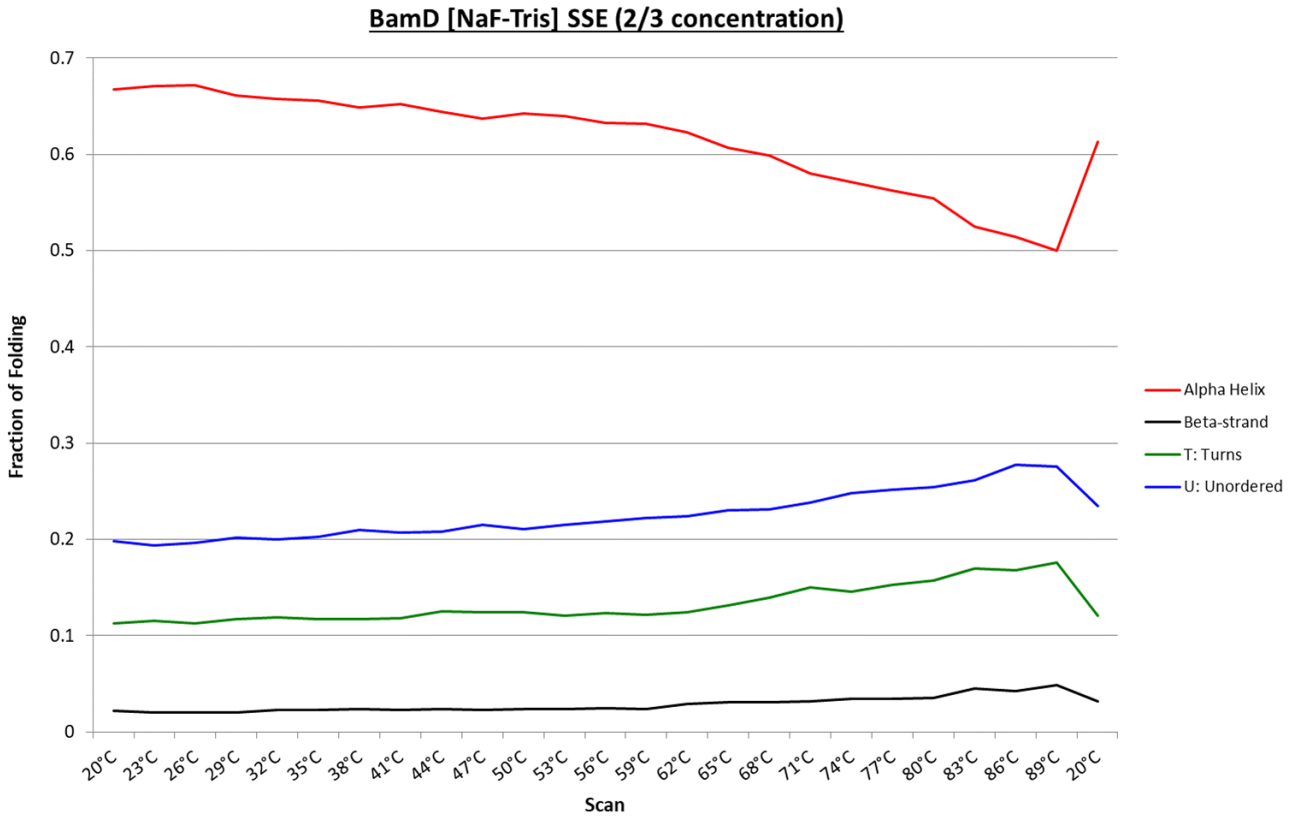


Figure 8.16: BamD [NaF-Tris] SSE with concentration reduced by one third . BamD SSE at 20°C if the concentration was reduced by a third from the original NanoDrop estimation.

8.4.4 **Individual protein *vs.* protein mixture thermal degradation and SSE comparison panels**

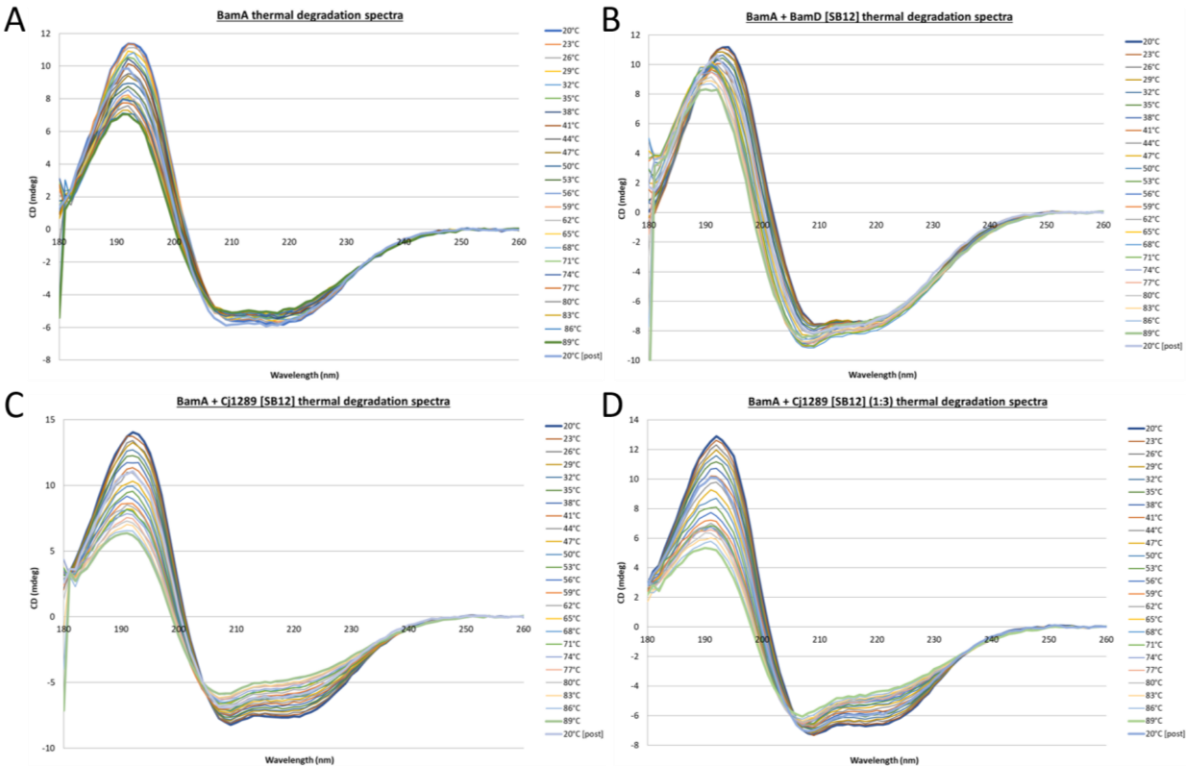


Figure 8.17: Thermal degradation profiles of BamA and BamA: [protein] mixtures. Buffer subtracted thermal degradation profiles of (A) BamA, and the BamA + (B) BamD [1:1], (C) Cj1289 [1:1], and (D) Cj1289 [1:3] mixtures in DB background. (A) BamA CD signal at 193 nm gradually decays as temperature increases, addition of (B) BamD stabilises and progressively blue-shifts the signal decay with temperature, (C) Cj1289 increases the signal decay around 193 nm, and (D) Cj1289 [1:3] decreases signal decay relative to the 1:1 BamA-Cj1289 mixture (C).

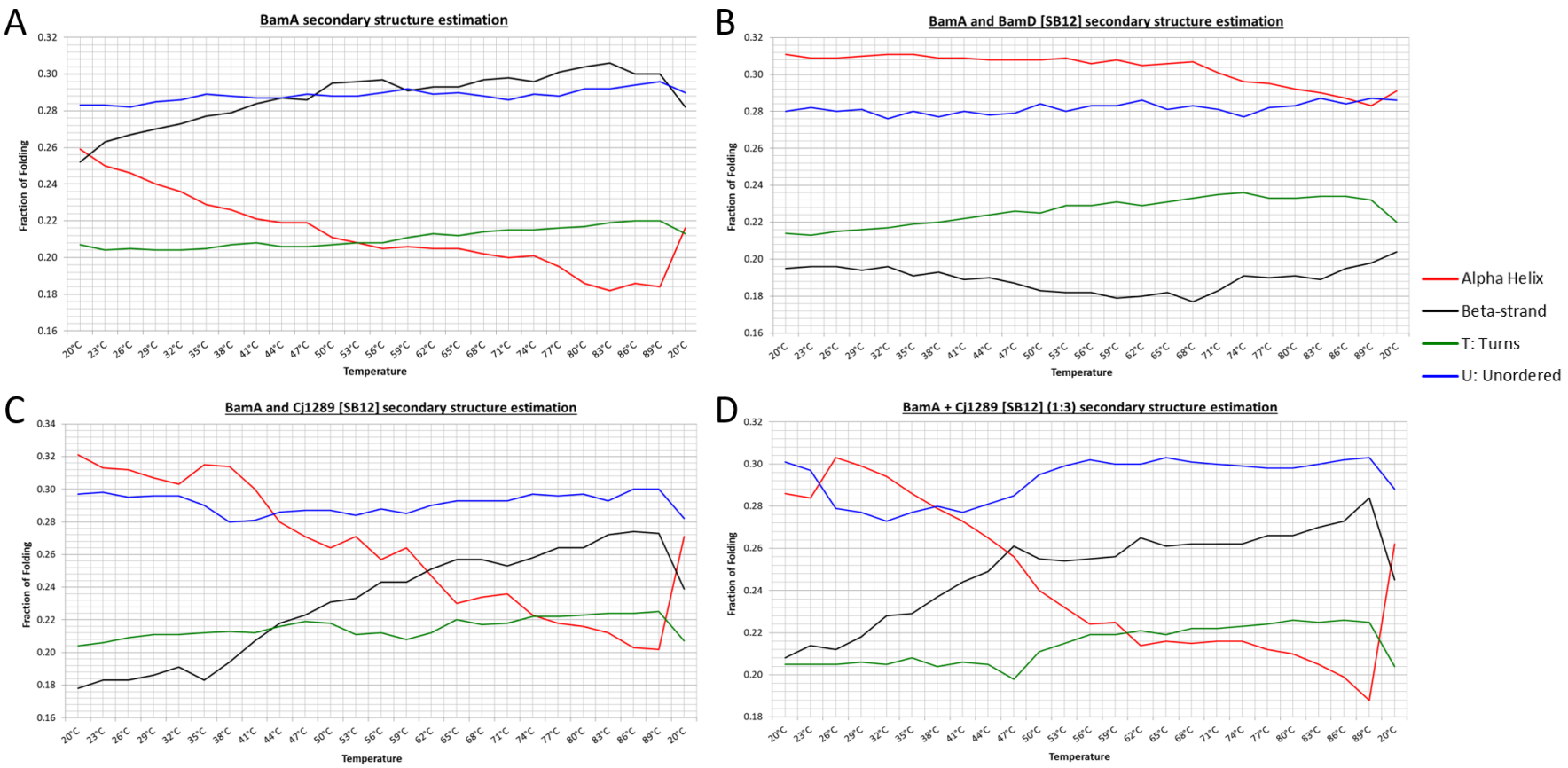


Figure 8.18: SSE profiles of BamA and BamA: [protein] mixtures. SSE profiles of (A) BamA, and the BamA + (B) BamD [1:1], (C) Cj1289 [1:1], and (D) Cj1289 [1:3] mixtures in DB background. (A) BamA shows a gradual decrease in α-helical structure and increase in β-sheet and turn structures with increasing temperature, addition of (B) BamD stabilises the SSE significantly, (C) Cj1289 increases the α-helical/ β-sheet transition significantly, and (D) Cj1289 [1:3] causes the α/ β transition to occur at lower temperatures relative to the 1:1 mixture (C).

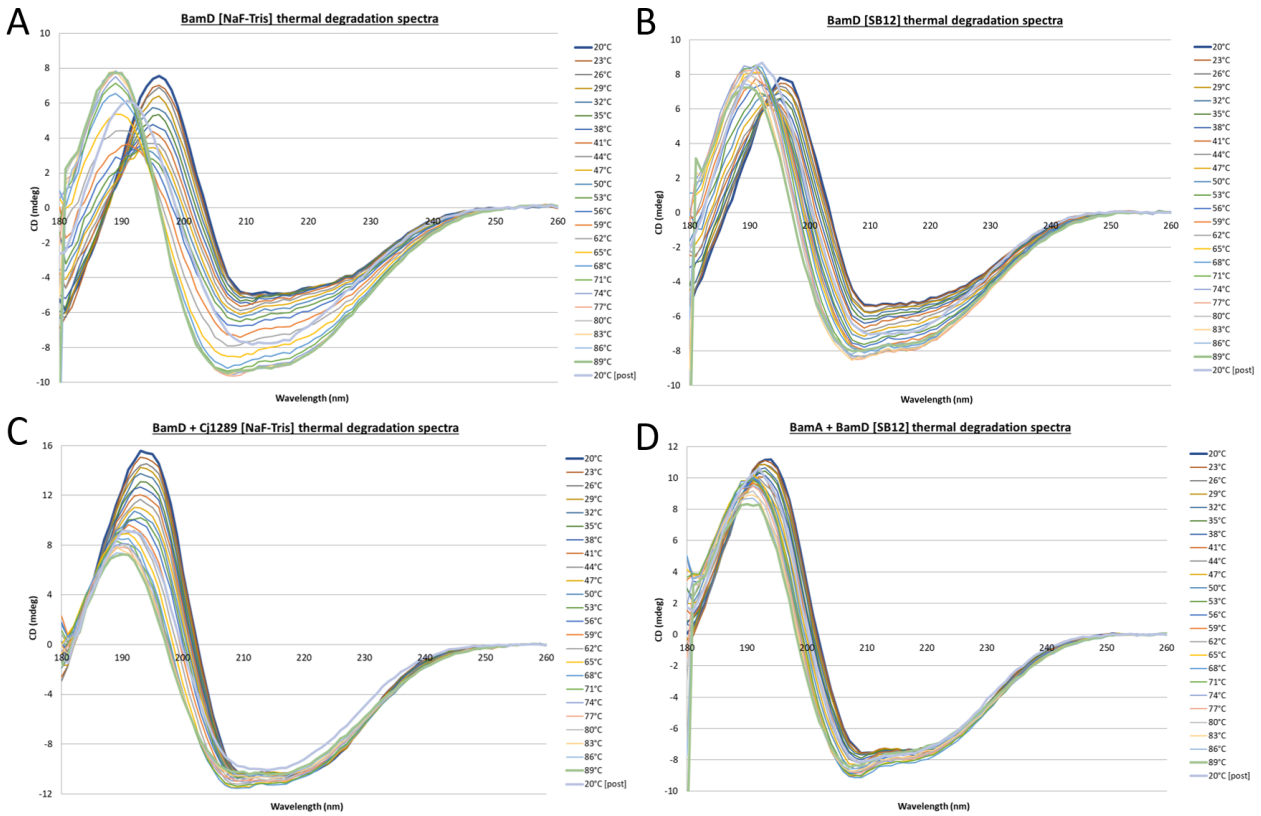


Figure 8.19: Thermal degradation profiles of BamD in both buffers and BamD: [protein] mixtures. Buffer subtracted thermal degradation profiles of (A) BamD [NDB], (B) BamD [DB], and the BamD + (C) Cj1289 [1:1, NDB], and (D) BamA [1:1, DB] mixtures. (A) BamD CD signal at 196 nm gradually decays and blue-shifts as temperature increases, (B) use of detergent buffer stabilises the BamD CD signal decay and lessens the blue-shift, (C) addition of Cj1289 increases the signal decay around 196 nm but reduces blue-shift of spectra, and (D) addition of BamA stabilises the signal decay and decreases the blue-shift.

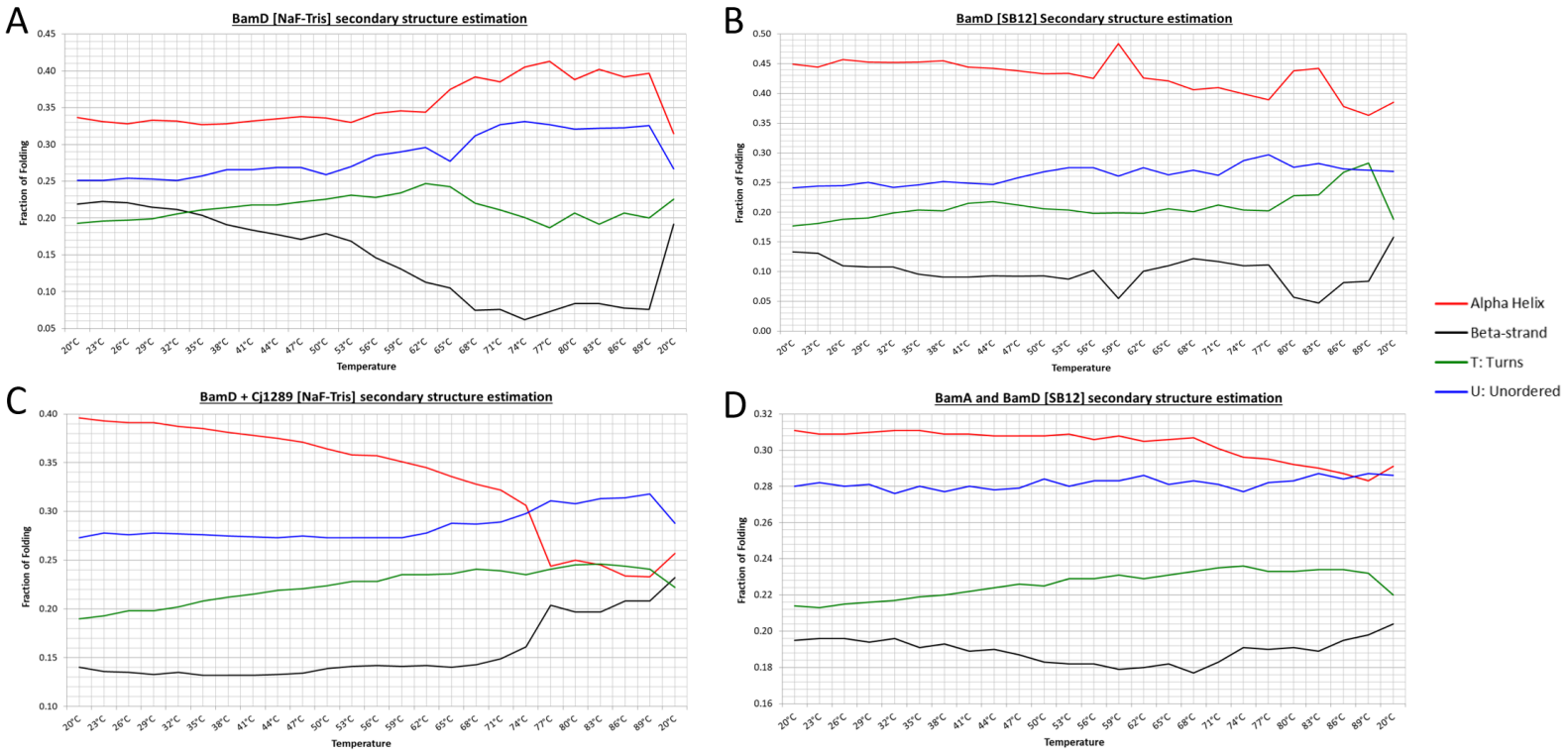


Figure 8.20: SSE profiles of BamD (both buffers) and BamD: [protein] mixtures. SSE profiles of (A) BamD [NDB], (B) BamD [DB], and the BamD + (C) Cj1289 [1:1, NDB], and (D) BamA [1:1, DB] mixtures. (A) BamD β-sheet structure gradual decreases with concurrent increases in turn and unordered structure with temperature, before an increase in α-helical after 65°C coinciding with a decrease in turn structure. (B) Addition of detergent stabilises the BamD SSE significantly, particularly β-sheet and unorderd structure, with two α/β transitions after 56°C and 77°C, respectively and a gradual decrease in α-helical overall. Addition of (C) Cj1289 stabilises the structures until a sharp α/β transition after 74°C, and (D) BamA stabilises the SSE component values significantly throughout the experimental temperature range.

8.4.5 **Boltzmann melting temperature estimation is wavelength dependent**

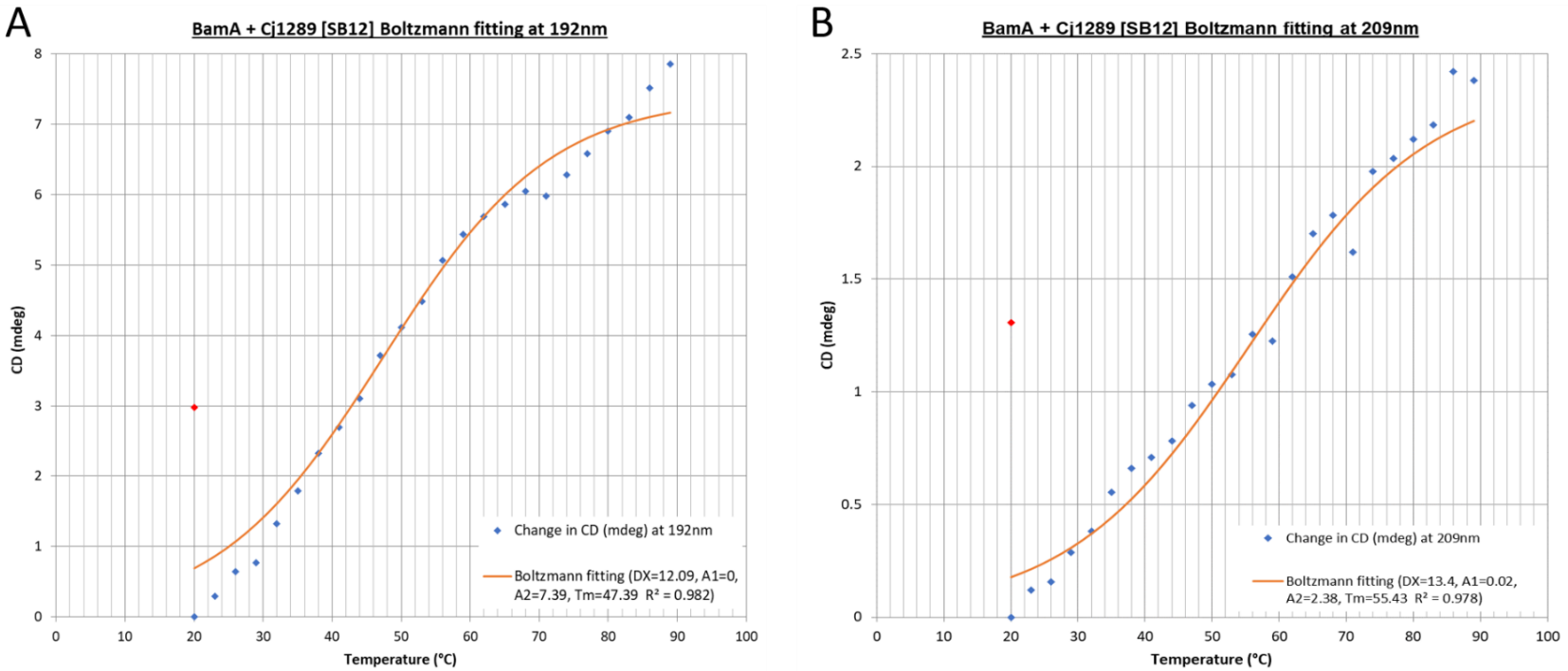


Figure 8.21: Boltzmann fitting of BamA and Cj1289 (1:1) mixture CD signal at 192 nm and 209 nm illustrates how wavelength alters Boltzmann derived T_m. An automated Boltzmann fitting was applied to the BamA-Cj1289 mixture spectral profile following the experiment at 192 nm (A) and 209 nm (B) to illustrate the impact of the wavelength choice on T_m and on the interpretation of the thermal degradation. The estimated T_m was determined to be 47.39°C (192 nm) and 55.43°C (209 nm) for the mixture in SB12 detergent buffer. The post-experiment spectra point upon return to 20°C is shown in red, and is different to the pre-experiment measurement by +3.0/ +1.3 mdeg at 192/ 209 nm.

8.5 Chapter 5 appendices

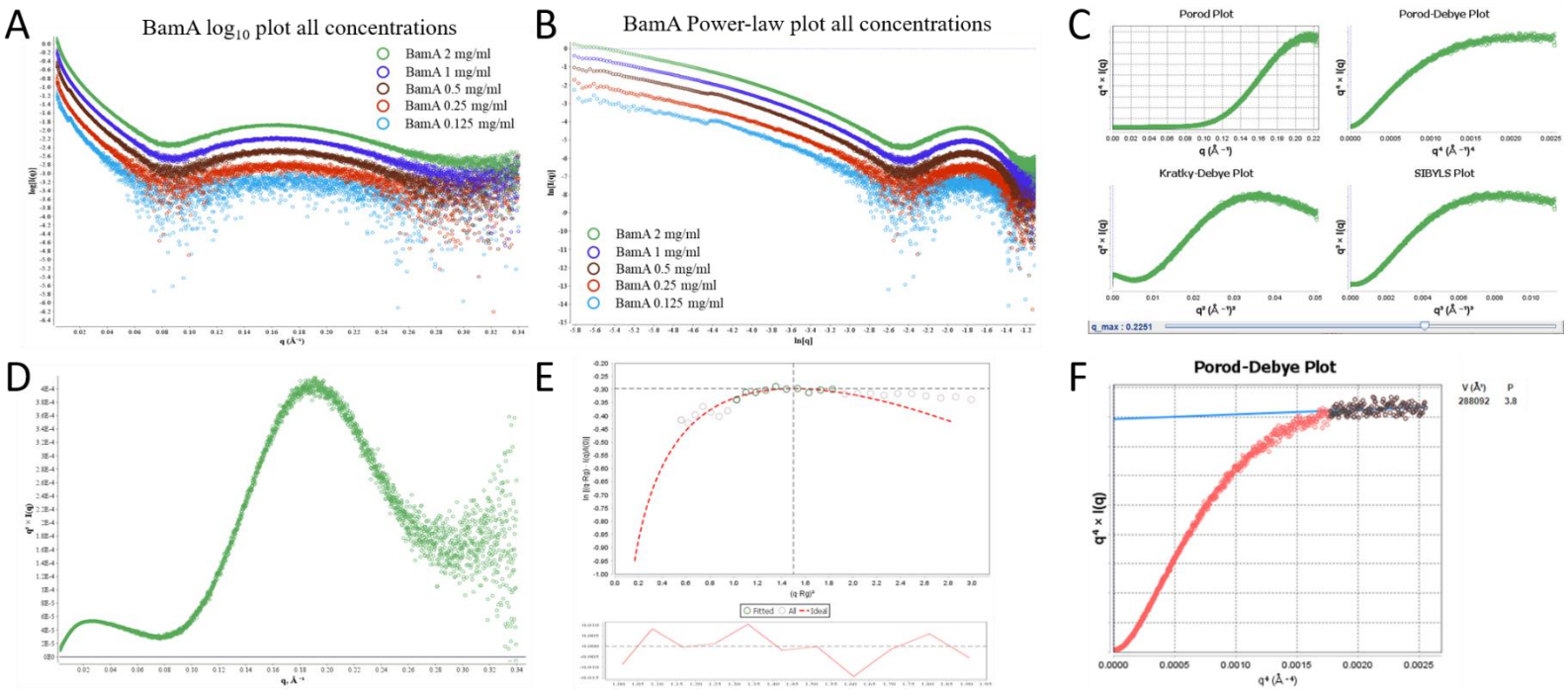


Figure 8.22: BamA SAXS analysis. BamA SAXS log_{10} (A), Power-law (B), flexibility (C) Kratky (D), Guinier peak (E), and Porod-Debye (F) analyses. Data signal in the mid-q (A) and low-q (B) are noisy below 1 mg/ml. Flexibility analysis (C) suggests a rigid compact shape with potentially some flexible elements (Porod-Debye followed closely by SIBYLS plot). The Kratky plot (D) has two peaks, suggesting a multi-domain protein. As the signal does not tend to the x-axis this suggests flexibility. The Guinier fitting (E) is poor with few points available after trimming uptick at low-q. The Porod-Debye plot (F) suggests a volume of 288092Å^3 (~144 kDa) and Porod-exponent of 3.8.

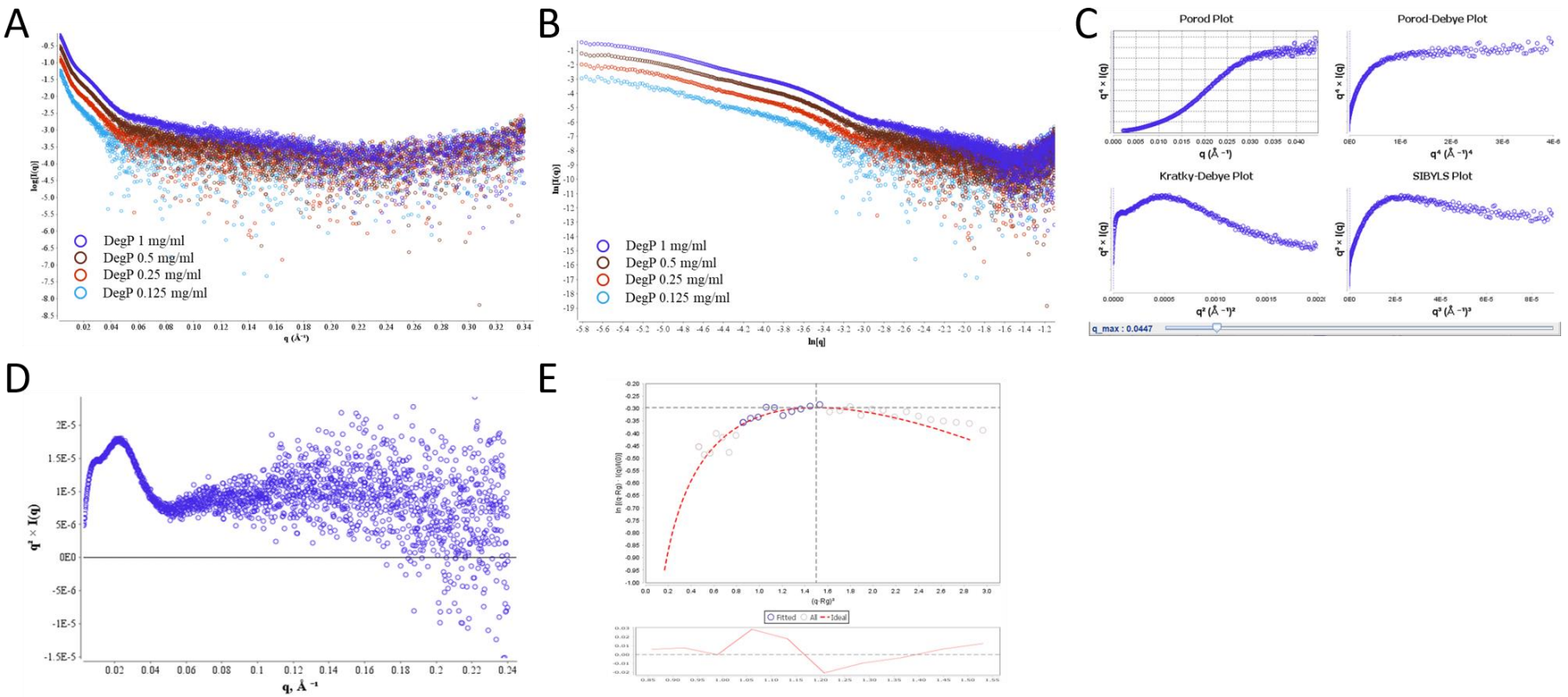


Figure 8.23: DegP SAXS analyis. DegP SAXS $\log_{10}$ (A), Power-law (B), flexibility (C), Kratky (D), and Guinier peak (E) analyses. Data signal in the mid-q (A) is noisy below 1 mg/ml and good at all concentrations tested at low-q (B). The flexibility plot (C) suggests a rigid compact shape with potentially some flexibility. The Kratky plot (D) has two peaks at low-q, suggesting a mix of larger and smaller multimeric complexes in the sample. As the peak does not tend to the x-axis this suggests flexibility. The Guinier fitting (E) is poor, with very few points available for fitting, limiting further analysis.

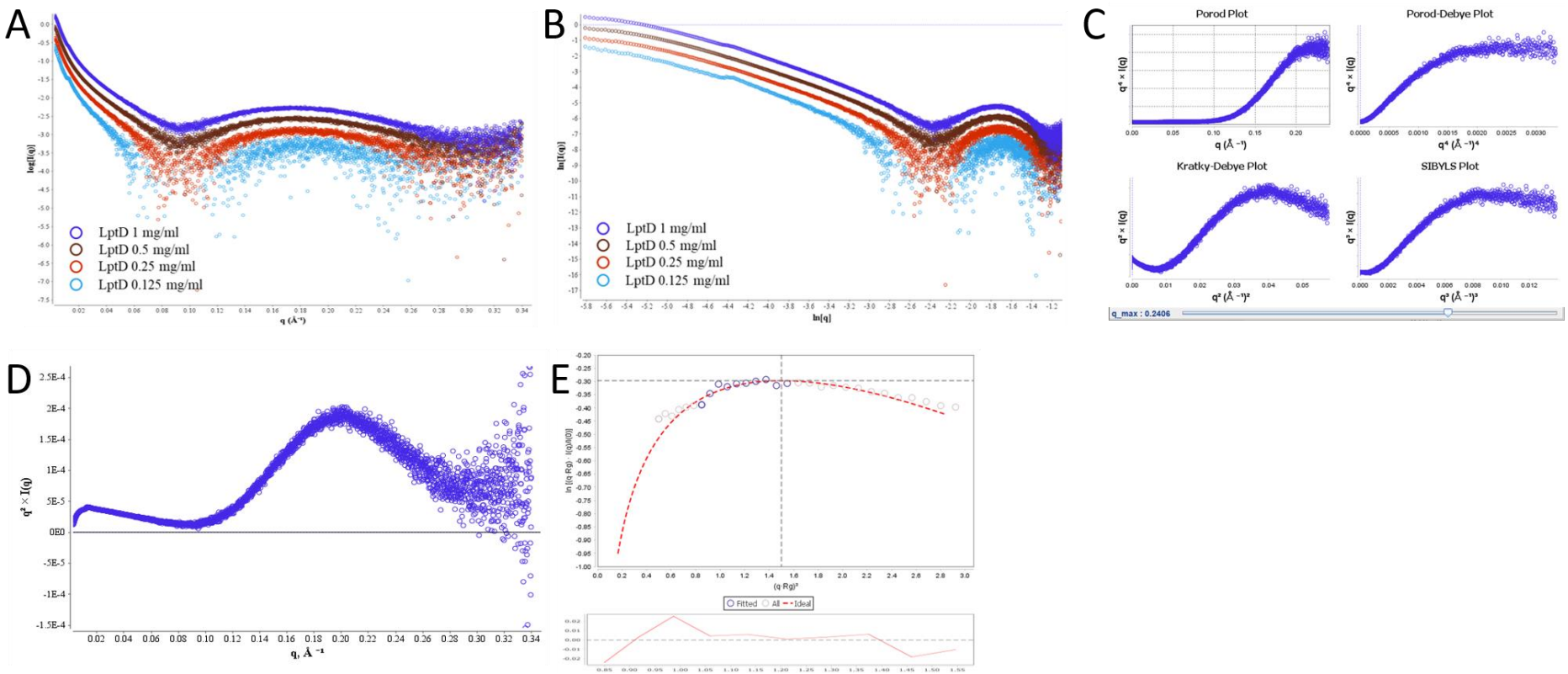


Figure 8.24: LptD SAXS analysis. LptD SAXS $\log_{10}$ (A), Power-law (B), flexibility (C), Kratky (D), and Guinier peak (E) analyses. Data in the mid-q is continuous above 0.5 mg/ml and good at all concentrations tested in the low-q (B). The flexibility plot (C) suggests a rigid compact shape with potentially some flexibility. The Kratky plot (D) has two peaks, one at low-q and another around mid- to high-q, suggesting a multidomain protein. Flexibility in the sample is uncertain due to noise at high-q. The Guinier fitting (E) is poor, with very few points available for fitting, limiting further analysis.

8.6 Chapter 6 appendices

8.6.1 Docking interface analysis for residues in contact

A

Protein/ model	Residues in contact (within 5 Å)
***E. coli* BamA**	35Q, 36R, 37V, 38A, 40G, 41A, 44L, 69T, 71N, 93T, 94I, 95A, 120R, 121V, 122G, 123E, 158L, 159P, 160R, 161N, 162R, 350R, 351K, 352I, 353R, 354F, 355E, 356G, 357N, 358D, 359T, 360S, 361K, 362D, 363A, 366R, 367R, 368E, 369M, 370R, 372M, 373E, 374G, 375A, 411Q, 417K, 449W, 450L, 451G, 452T, 480V, 481D
***E. coli* BamD**	29D, 30S, 31P, 32P, 35I, 59D, 60N, 61R, 62Y, 63P, 64F, 65G, 66P, 69Q, 72Q, 86P, 87L, 90A, 94R, 97R, 98L, 99N, 100P, 101T, 119L, 120D, 122S, 123A, 124L, 125Q, 128F, 130V, 131D, 132R, 133S, 134D, 135R, 136D, 137P, 138Q, 139H, 141R, 177Y, 178E, 180S, 181V, 184Y, 185Y, 187E, 188R, 189G, 190A, 191W, 192V, 193A, 196N, 197R, 199E, 200G, 203R

B

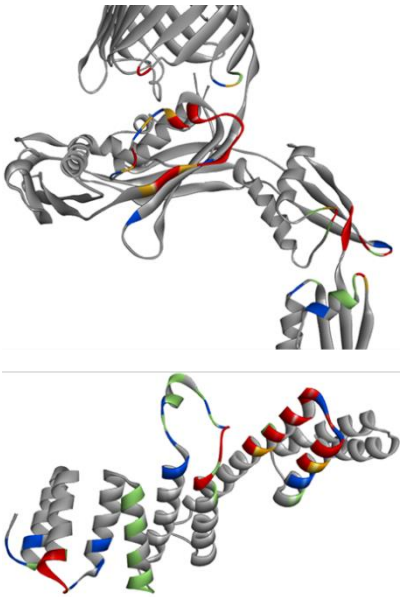


Figure 8.25: Residues involved in *E. coli* BamA-BamD interaction interface. Four available structures on RCSB of the BAM complex (5AYW, 5D0O, 5D0Q, 5LJO) were analysed for residues in the contact interface between BamA-BamD by selecting all residues with 5Å. Table of residues composing the docking interface (A) colour coded by occurrence in either blue (1 of 4 models), green (2 of 4 models), orange (3 of 4 models) or red (all four models) and a graphical representation of the residue location (B) are shown on BamA and BamD individually.

A

Protein/ model	Potential residues for binding
C. jejuni **BamA**	29N, 30H, 31L, 32S, 34T, 35S, 38N, 63Q, 65Y, 87T, 88I, 89A, 111K, 112R, 113G, 114T, 150E, 151N, 152T, 153D, 154G, 345I, 346R, 347N, 348V, 349I, 350I, 351S, 352G, 353N, 354S, 355R, 356T, 357V, 358D, 359R, 361I, 362R, 363R, 364E, 365L 366Y, 368T, 369E, 370G, 371N, 405H, 410V, 411D, 509A, 510R, 511Y, 560Y, 561L, 562P
C. jejuni **BamD**	27Y, 28N, 29L, 30S, 31A, 34W, 58A, 59S, 60E, 61H, 62V, 63A, 64D, 65P, 68E, 85K, 86L, 89F, 93E, 96K, 97K, 98F, 99G, 100N, 115K, 116F, 119F, 120A, 121V, 122P, 123N, 124R, 125N, 126Q, 127A, 128L, 130L, 166L, 167N, 169T, 170I, 173L, 174Y, 176R, 177I, 178G, 179H, 180D, 181E, 182S, 185I, 186Y, 188Q, 189R, 192E

B

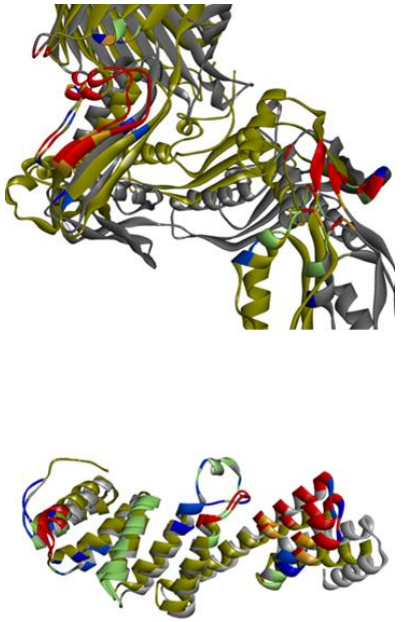


Figure 8.26: Approximation of the residues involved in the CjBamAD docking interface by superimposition to homologous structures. As the CjBamA and CjBamD homology models show significant structural similarity to the known *E. coli* structural models, the homology models were superimposed onto the 5LJO Cryo-TEM structure using a molecular overlay function to approximate the residues involved in the docking interface and derive residues restraints for rigid body docking. The equivalent residues were identified in CjBamA and CjBamD based on structural similarity, local proximity and residue chemistry that represent the docking interface observed in *E. coli* BamAD are tabulated (A) and represented graphically with annotation (B).

8.6.2 **Rigid body docking outputs**

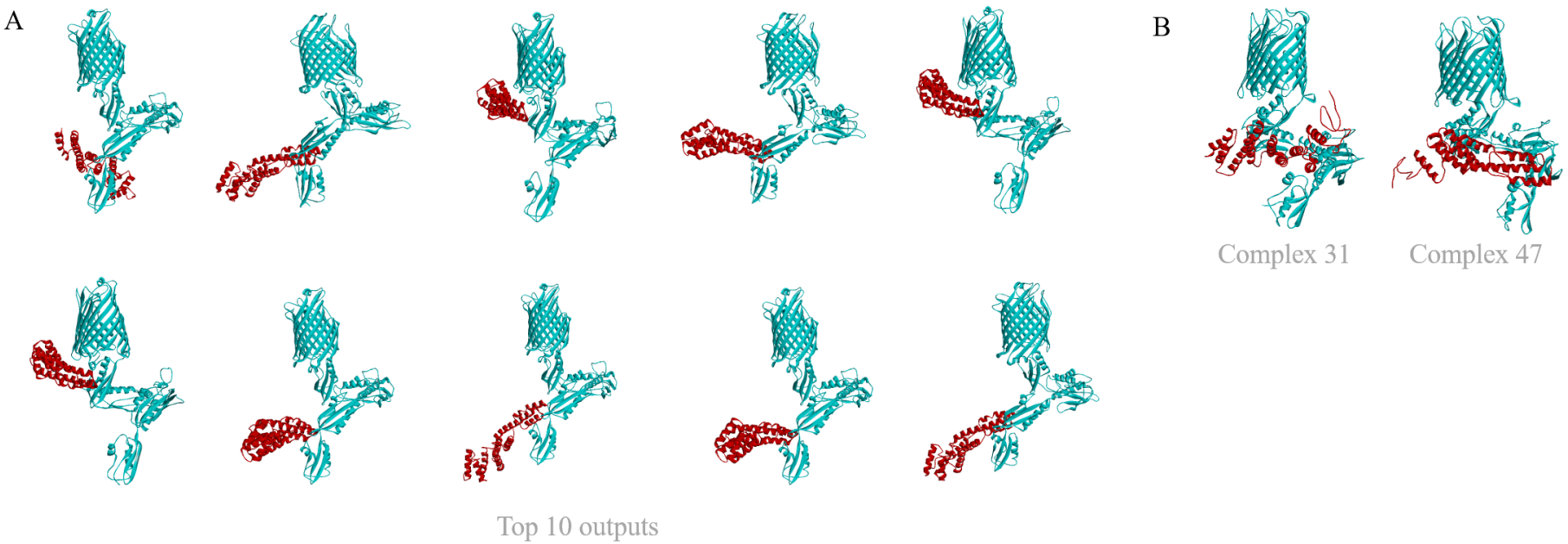
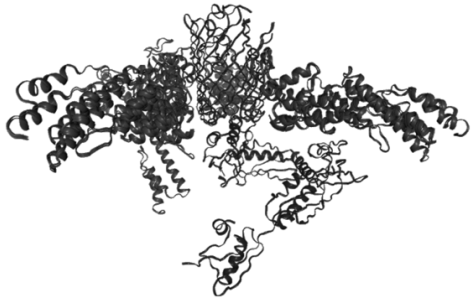


Figure 8.27: ZDock rigid body docking output. A rigid body docking run was carried out using ZDock (3.0.2) blocking all residues except residues identified as restraints previously. The top 10 outputs (A) and 2 selected models that approximately match the expected binding configuration (B) are shown.

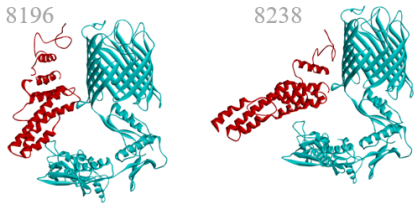
A

Conf	Electrostatics	Desolvation	VdW	Total	Rank
2190	-4.206	-59.881	73.133	-56.774	1
137	-3.352	-60.926	91.087	-55.170	2
2858	-5.664	-53.460	63.993	-52.725	3
3074	-15.273	-43.392	76.153	-51.049	4
1851	-7.021	-48.588	46.941	-50.915	5
8238	-3.999	-62.618	167.835	-49.833	6
3514	-0.786	-56.780	77.441	-49.822	7
4791	-0.460	-56.067	71.736	-49.353	8
1218	-8.185	-53.638	129.327	-48.890	9
2224	0.785	-56.201	69.315	-48.484	10
Conf	Electrostatics	Desolvation	VdW	Total	Rank

B



C

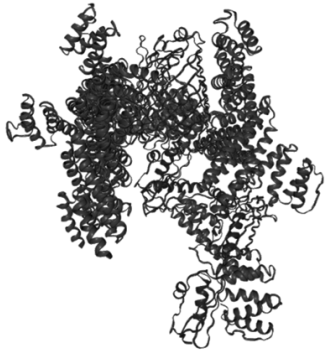


Conf	Electrostatics	Desolvation	VdW	Total	Rank
8196	-9.015	-31.880	62.046	-34.690	98
8238	-3.999	-62.618	167.835	-49.833	6

D

Conf	Electrostatics	Desolvation	VdW	relRST	Total	Rank
1682	-5.407	-43.805	2.128	32.727	-81.726	1
7872	-2.022	-74.200	81.352	12.727	-80.814	2
6470	-15.076	-50.592	-49.345	10.000	-80.603	3
457	-4.986	-67.811	87.269	16.364	-80.434	4
4869	-7.348	-71.311	29.444	4.545	-80.260	5
3994	-11.275	-60.328	52.610	11.818	-78.160	6
3998	-14.638	-41.362	-19.570	20.000	-77.957	7
4502	-4.525	-62.656	78.957	17.273	-76.558	8
9790	-6.473	-63.757	73.920	13.636	-76.474	9
6045	-1.675	-87.416	147.865	1.818	-76.122	10
Conf	Electrostatics	Desolvation	VdW	relRST	Total	Rank

E



F



Conf	Electrostatics	Desolvation	VdW	relRST	Total	Rank
3840	-17.980	-49.917	30.141	5.455	-70.338	28
5701	-14.157	-35.601	6.912	18.182	-67.249	47

Figure 8.28: pyDock rigid body modelling outputs. Two rigid body docking runs were carried out using the pyDockWeb webserver and the CjBamA and CjBamD homology models as inputs. The first run was unbiased and without restraints (A, B, C) and the second (D, E, F) used the previously identified distance restraints (within 5 Å). The top 10 models by energies (A, D, respectively), their docked models (B, E, respectively), and two chosen models (C, F, respectively) are shown for each run. The inclusion of the restraints identified from superimposition increased the validity of the models outputted.

A



B

Score data [Download original score file]

decoy	total_score	rms	Fnat	I_sc	Irms	cen_rms	dslf_ca_dih	dslf_cs_ang	dslf_ss_dih	dslf_ss_dst
proteins_0069	-716.835	12.502	NaN	-6.078	NaN	11.892	0	0	0	0
proteins_0015	-715.896	40.38	NaN	-6.852	NaN	39.007	0	0	0	0
proteins_0048	-715.54	40.098	NaN	-6.297	NaN	42.486	0	0	0	0
proteins_0678	-715.325	16.575	NaN	-5.057	NaN	14.579	0	0	0	0
proteins_0067	-715.245	10.384	NaN	-4.679	NaN	12.781	0	0	0	0
proteins_0310	-715.033	14.62	NaN	-5.269	NaN	15.418	0	0	0	0
proteins_0006	-715.007	25.51	NaN	-5.552	NaN	27.423	0	0	0	0
proteins_0056	-714.921	42.259	NaN	-5.352	NaN	37.163	0	0	0	0
proteins_0716	-714.878	14.551	NaN	-4.631	NaN	11.61	0	0	0	0
proteins_0164	-714.849	8.986	NaN	-5.064	NaN	8.906	0	0	0	0

C

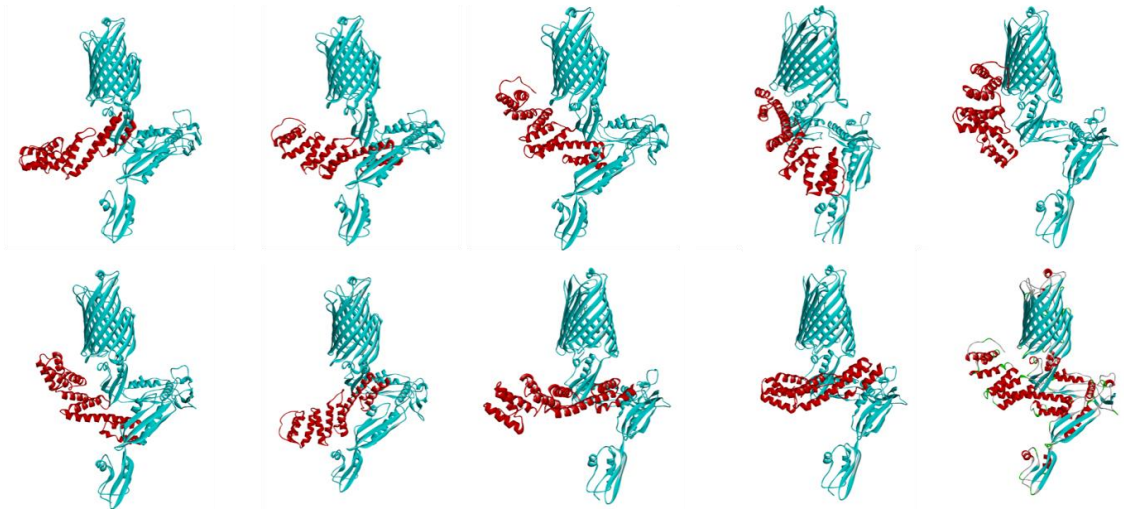


Figure 8.29: RosettaDock rigid body modelling ouput. A rigid body docking run was carried out using the ROSIE (Docking 2) webservice. The initial combined model input (A), assigned scoring output for the top 10 models (B), and top 10 docked models (C) are shown.

	Residues in contact (rigid docking)	
Rigid body docking model	***C. jejuni* BamA**	***C. jejuni* BamD**
ZDock complex 31	**30H**, **31L**, **32S**, 33N, **34T**, **35S**, **38N**, 39I, **111K**, **112R**, **113G**, **151N**, **346R**, **347N**, **349I**, 400K, 401V, 402D, **405H**, 407D	49K, 53H, 55N, 56G, **59S**, **60E**, **63A**, 72I, 90Y, 94Y, **97K**, **98F**, **122P**, **123N**, **124R**, 162A, **166L**, **169T**, **173L**
ZDock complex 47	**30H**, **31L**, **32S**, **34T**, **35S**, **38N**, 39I, 62K, **63Q**, 64N, **65Y**, **344Y**, **345I**, **346R**, **347N**, **348V**, **359R**, **362R**, **368T**, **369E**, **370G**, **371N**, **405H**, 407N	52D, 53D, 55N, 56G, 57M, **59S**, **60E**, 75A, 79M, 87A, 90Y, 91L, **93E**, 94Y, **97K**, **124R**, **125N**, **166L**, **169T**, **170I**, **173L**, **177I**
pyDock complex 3840	137Y, 163R, 165E, 168I, 194K, 196F, 200M, 201W, 202G, 204N, 242Y, 244D, 246Y, 247Q, 444F	22N, 23D, 26L, 82E, 84Y, 117D, 118A, **119F**, **120A**, **121V**, **122P**, **123N**, **124R**, 151L, 155M, 157T, 158K, 161L, 165Y, 198Q
pyDock complex 5701	96I, 97A, 99N, 137Y, 139D, 163R, 164G, 165E, 166N, 196F, 201W, 242Y, 247Q, 444F, 445G, 446S	**62V**, **63A**, 84T, **85K**, **97K**, **98F**, **124R**, **125N**, **126Q**, **127A**, **128L**, 134K, 137E, 138T, 141K, **170I**, **173L**, **174Y**, **189R**, **192E**

Rosetta Dock model 4	**30H**, **87T**, **112R**, **113G**, **114T**, **151N**, **152T**, **346R**, **347N**, **348V**, **349I**, **350I**, **358D**, **359R**, **361I**, **362R**, **363R**, **364E**, **365L**, **366Y**, **368T**, **369E**, **371N**, **509A**, **510R**, **511Y**, 512Y, 555D, 557D, 558D, **561L**, 564S, 565G, 566I, 594Q, 595G, 596L	20T, 21K, 22N, 23D, 24E, 26L, **121V**, **122P**, **123N**, **124R**, **125N**, 165Y, 168S, **169T**, **170I**, 172N, **173L**, **174Y**, 175Q, **176R**, **177I**, 202K, 206P, 209R, 213E
Rosetta Dock model 9	100D, 101R, 102K, 103Q, 105E, 106S, **111K**, **112R**, **346R**, **347N**, **349I**, **350I**, **351S**, **352G**, **353N**, **354S**, **362R**, **369E**, **370G**, 401V, 402D, **405H**, 407D, **411D**, 413K, 442N, 444F, 445G, 446S, 471R	18C, 20T, 48E, 81E, 82E, 83E, 84Y, **85K**, **124R**, **125N**, **127A**, **128L**, 161L, 165Y, 168S, 171E, 172N, **174Y**, **176R**, **179H**, **182S**, **189R**, 198Q, 202K, 205L, 206P, 209R

Table 8.9: Residues within contact distance (≤ 4Å) in the selected rigid body docking outputs. The two selected CjBamAD docked models from the ZDock, pyDock and RosettaDock rigid body docking outputs with the highest similarity to the *E. coli* experimental structure docking configuration were analysed for residues in contact (≤ 4 Å). All identified residues within this distance restraint are tabulated, and residues that match the initial predictions from the superimposition of the *C. jejuni* and *E. coli* models are shown in bold.

8.6.3 **MD scaling runs**

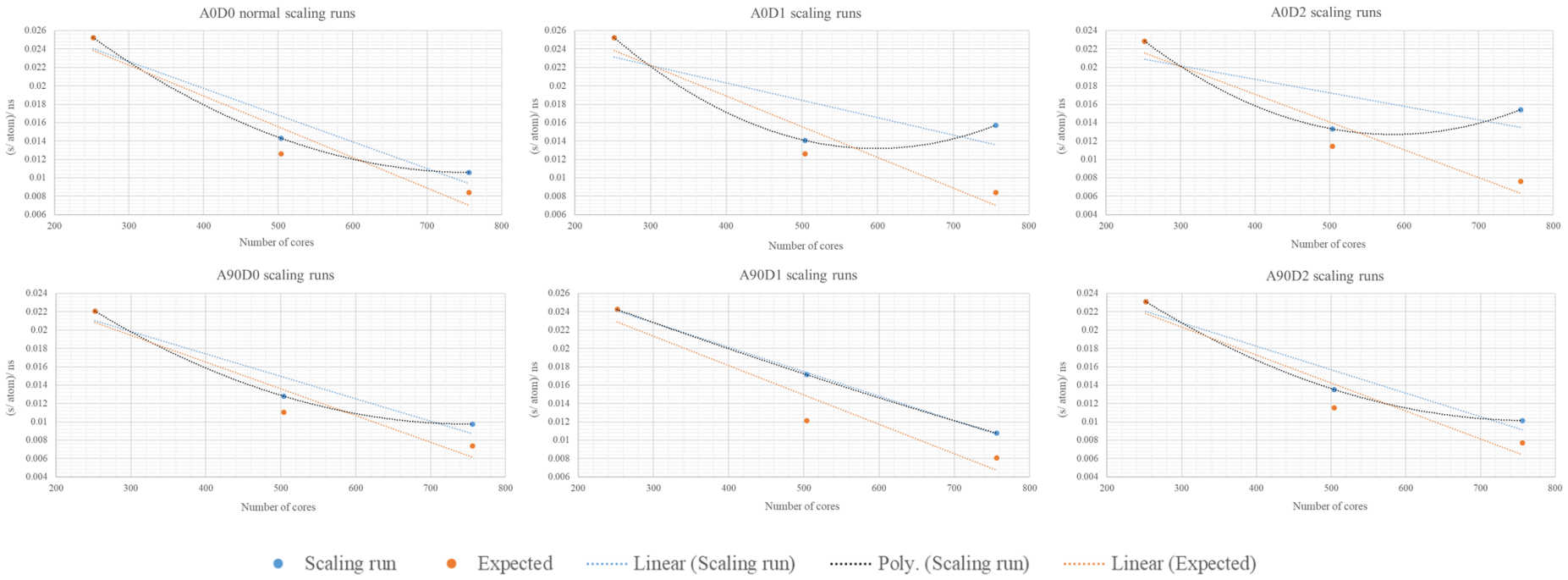


Figure 8.30: MD scaling runs A0D0-A90D2. Prior to the docking assay MD simulation trajectory production runs a series of scaling runs were carried out to determine the appropriate core allocation that delivers trajectories cost effectively while within the wall clock time constraints. 1 ns MD scaling runs where three core allocations (256, 504, 756) were assessed for simulation time are shown for MD docking runs A0D0-A90D2, and have been graphed against expectations for simulation time for the corresponding resource allocation. A 504-core allocation shows appropriate decreases (>80 %) in simulation time for near doubling of core allocation in all cases, and should be used for all simulations.

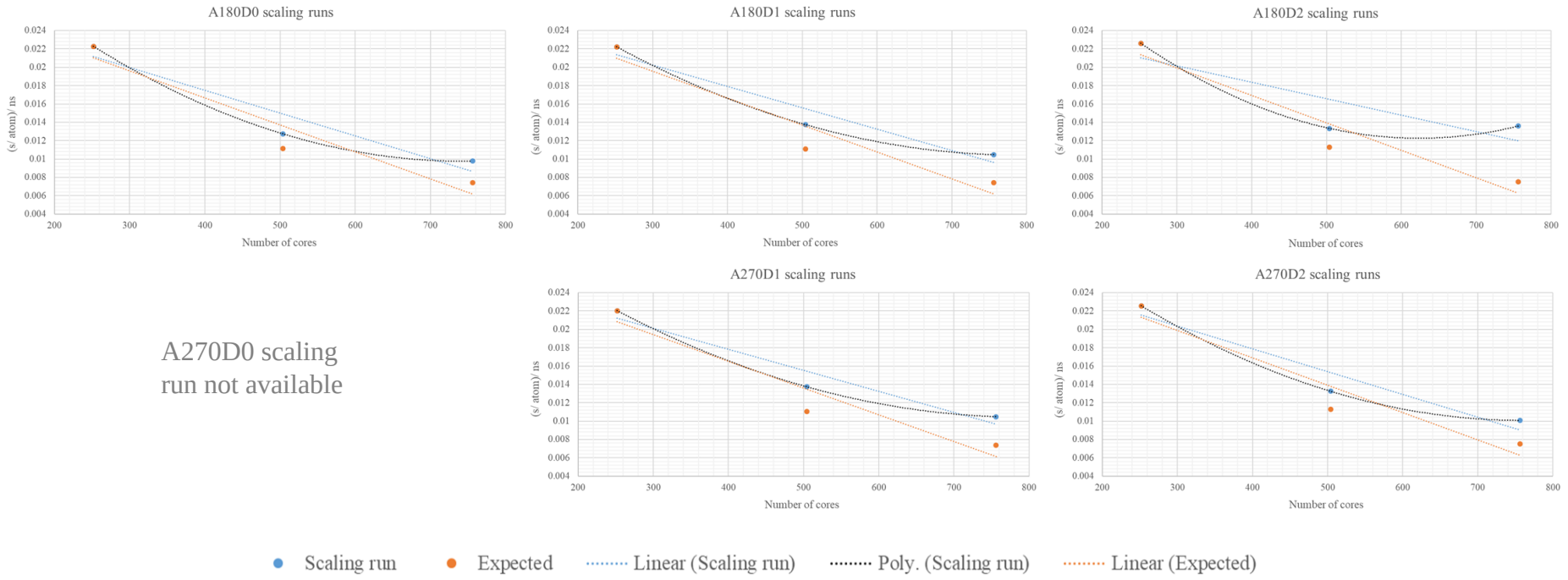


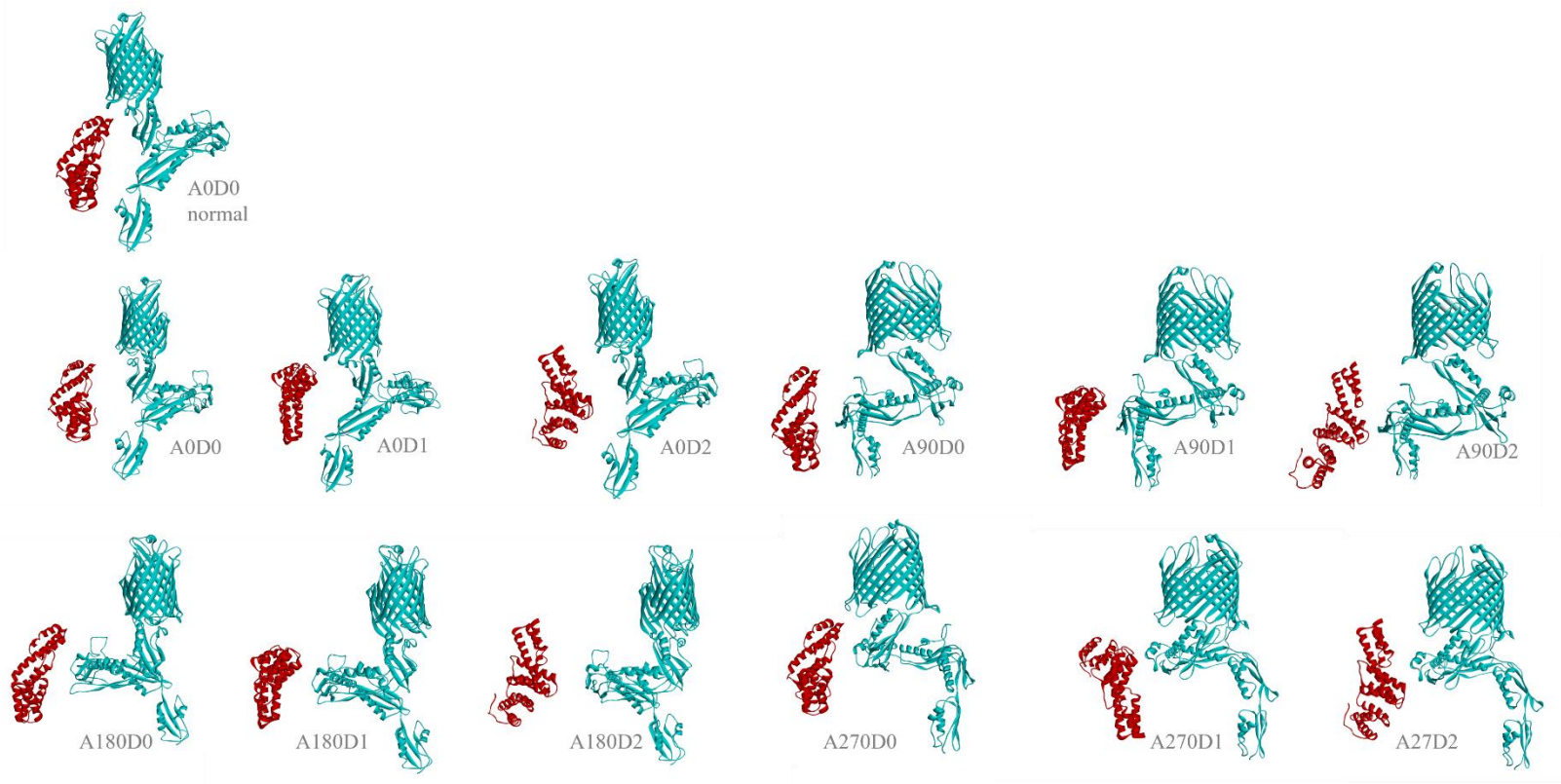
Figure 8.31: MD scaling runs A180D0-A270D2. Prior to the docking assay MD simulation trajectory production runs a series of scaling runs were carried out to determine the appropriate core allocation that delivers trajectories cost effectively while within the wall clock time constraints. 1 ns MD scaling runs where three core allocations (256, 504, 756) were assessed for simulation time are shown for MD docking runs A180D0-A270D2, and have been graphed against expectations for simulation time for the corresponding resource allocation. A 504-core allocation shows appropriate decreases (>80 %) in simulation time for near doubling of core allocation in all cases, and should be used for all simulations.

8.6.4 **BamA and BamD structural elements by residue number**

Structure	Sub-structure	Residue (start)	Residue (end)
BamA		**18**	**739**
POTRA domain		18	414
	POTRA 1	18	85
	POTRA 2	87	163
	POTRA 3	170	256
	POTRA 4	257	339
	POTRA 5	344	414
β-Barrel		418	739
	β1	418	428
	β2	431	442
	β3	449	457
	β4	460	474
	β5	477	490
	β6	493	508
	β7	512	528
	β8	541	557
	β9	565	576
	β10	583	594
	β11	603	616
	β12	658	668
	β13	676	687
	β14	696	707
	β15	711	721
	β16	731	736
BamD		**723**	**920**
	helix 1	736 (31)	750 (45)
	helix 2	756 (51)	766 (61)
	helix 3	770 (65)	783 (78)
	helix 4	789 (84)	803 (98)
	helix 5	809 (104)	823 (118)
	helix 6	832 (127)	848 (143)
	helix 7	853 (148)	883 (178)
	helix 8	885 (180)	897 (192)
	helix 9	910 (205)	920 (215)

Table 8.10: BamA and BamD structural elements by residue number. Residue lookup table for all *C. jejuni* BamA-BamD MD docking runs input files to provide reference for RMSF outputs residue structural location.

8.6.5 **Starting positions – MD docking assay**



MD docking assay run starting positions. Relative starting positions of BamA and BamD in the MD docking for each orientational sampling are shown. Membranes have been removed to enable readers to gauge distance and orientation more easily.

8.6.6 Trajectory RMSD for docking runs (pre-PBC processing)

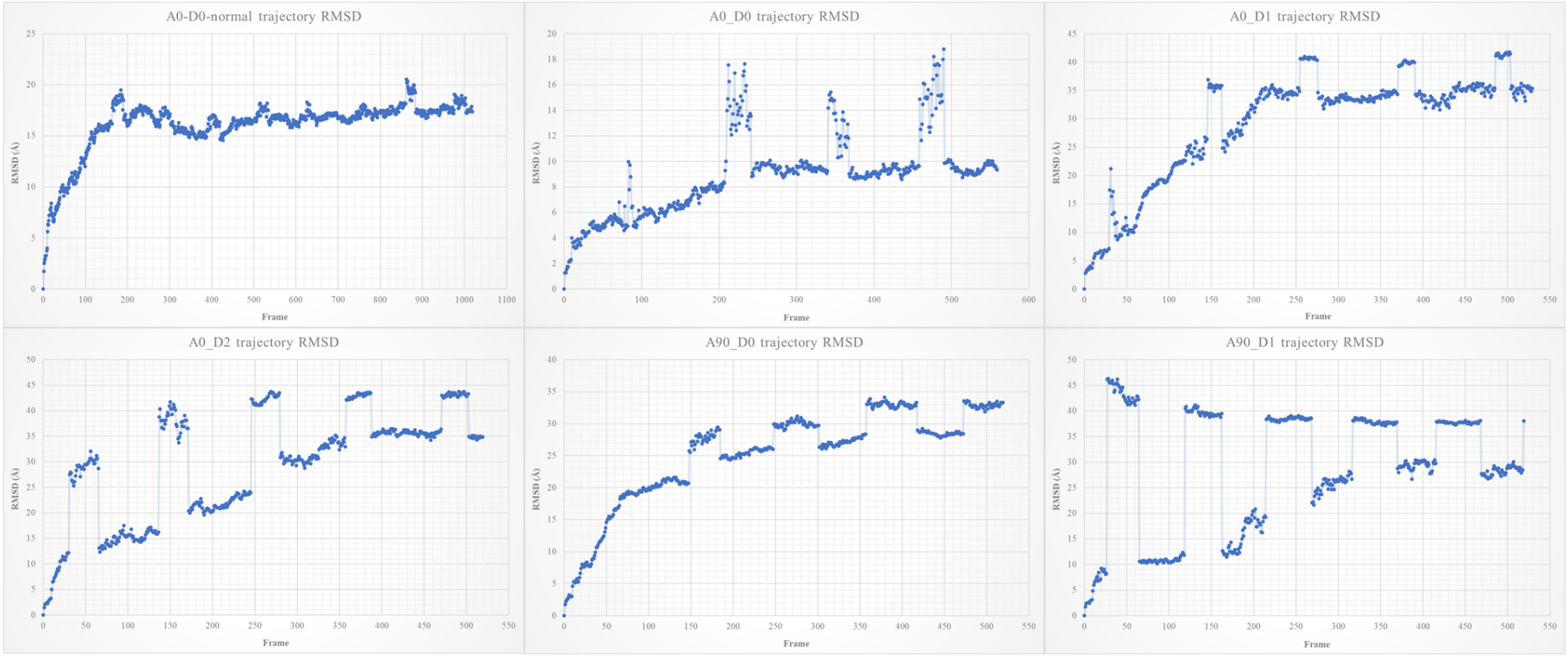


Figure 8.32: Trajectory RMSD A0D0-A90D1 MD runs (PBC wrap). The original (untreated) RMSD data for MD docking runs A0D0-A90D1 before PBCs were unwrapped to remove simulation drift and allow for further analysis of data.

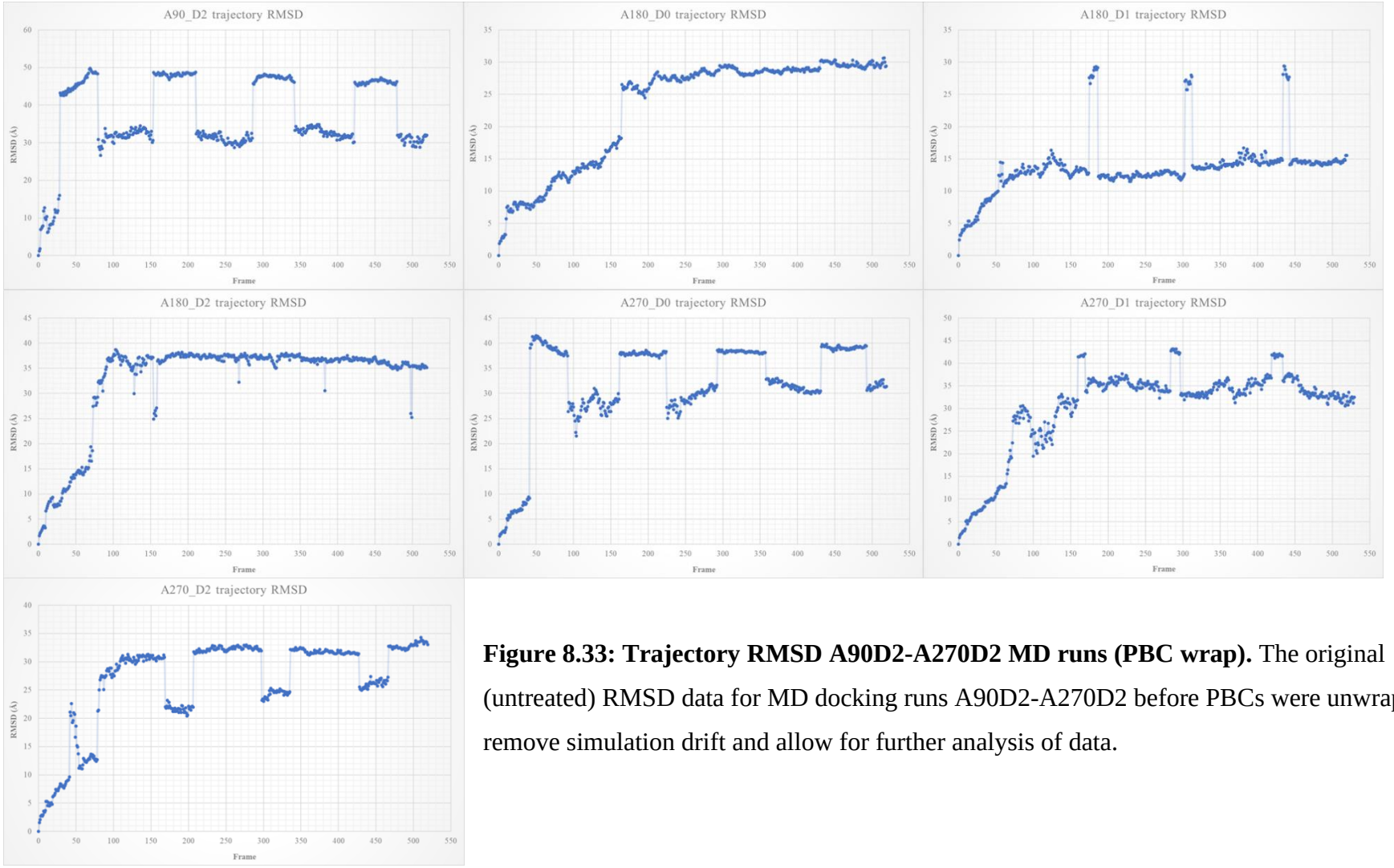


Figure 8.33: Trajectory RMSD A90D2-A270D2 MD runs (PBC wrap). The original (untreated) RMSD data for MD docking runs A90D2-A270D2 before PBCs were unwrapped to remove simulation drift and allow for further analysis of data.

8.6.7 Trajectory RMSF for docking runs (pre-PBC processing)

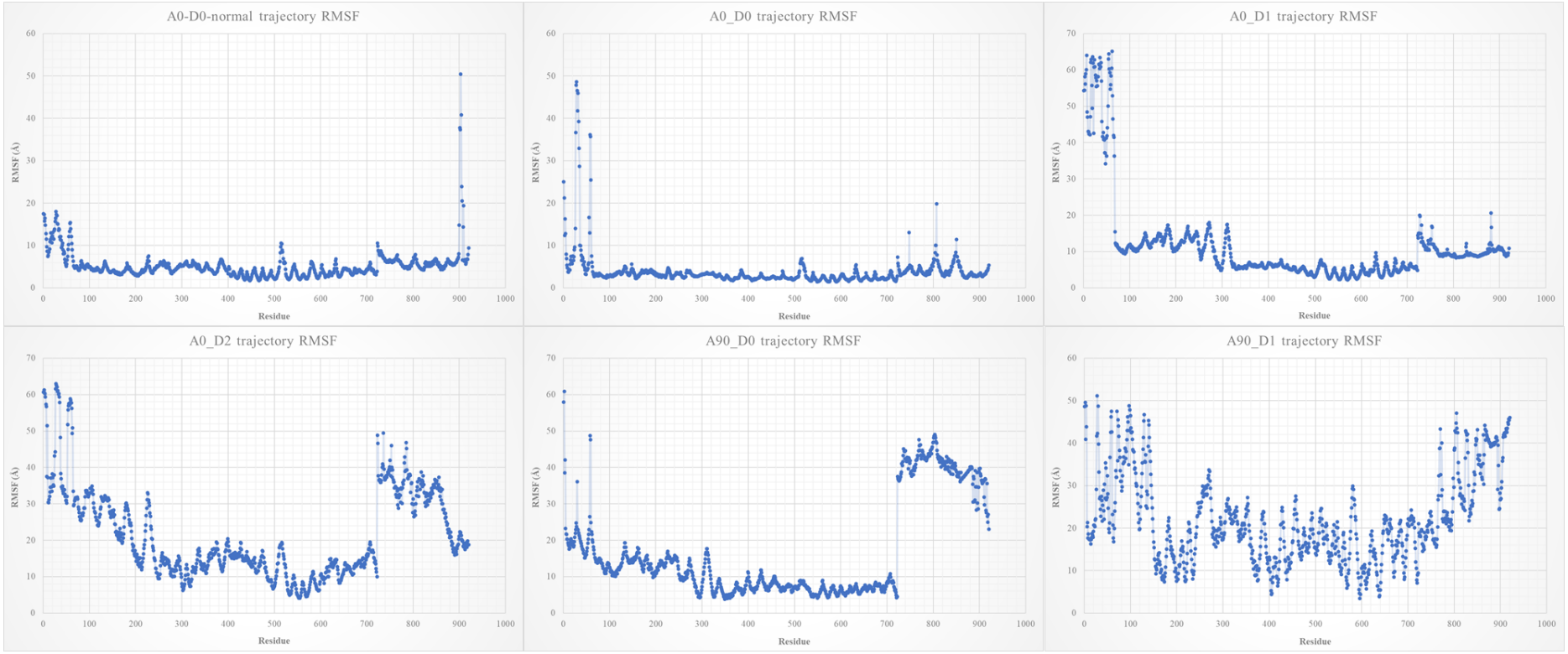


Figure 8.34: Trajectory RMSF A0D0-A90D1 MD runs (PBC wrap). The original (untreated) RMSF data for MD docking runs A0D0-A90D1 before PBCs were unwrapped to remove simulation drift and allow for further analysis of data.

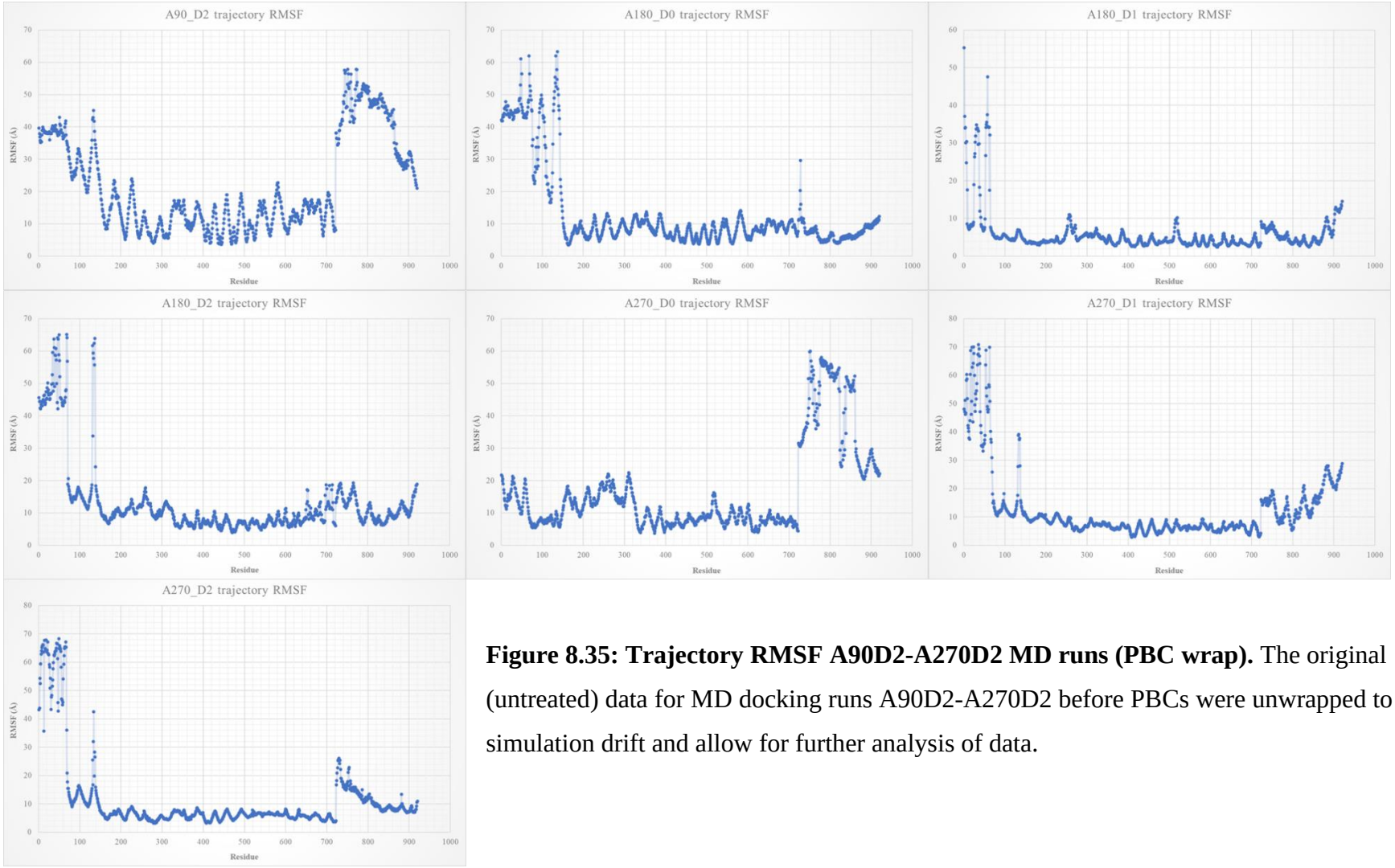


Figure 8.35: Trajectory RMSF A90D2-A270D2 MD runs (PBC wrap). The original (untreated) data for MD docking runs A90D2-A270D2 before PBCs were unwrapped to remove simulation drift and allow for further analysis of data.

8.7 Chapter 7 appendices

8.7.1 Cryo-TEM 2D classes

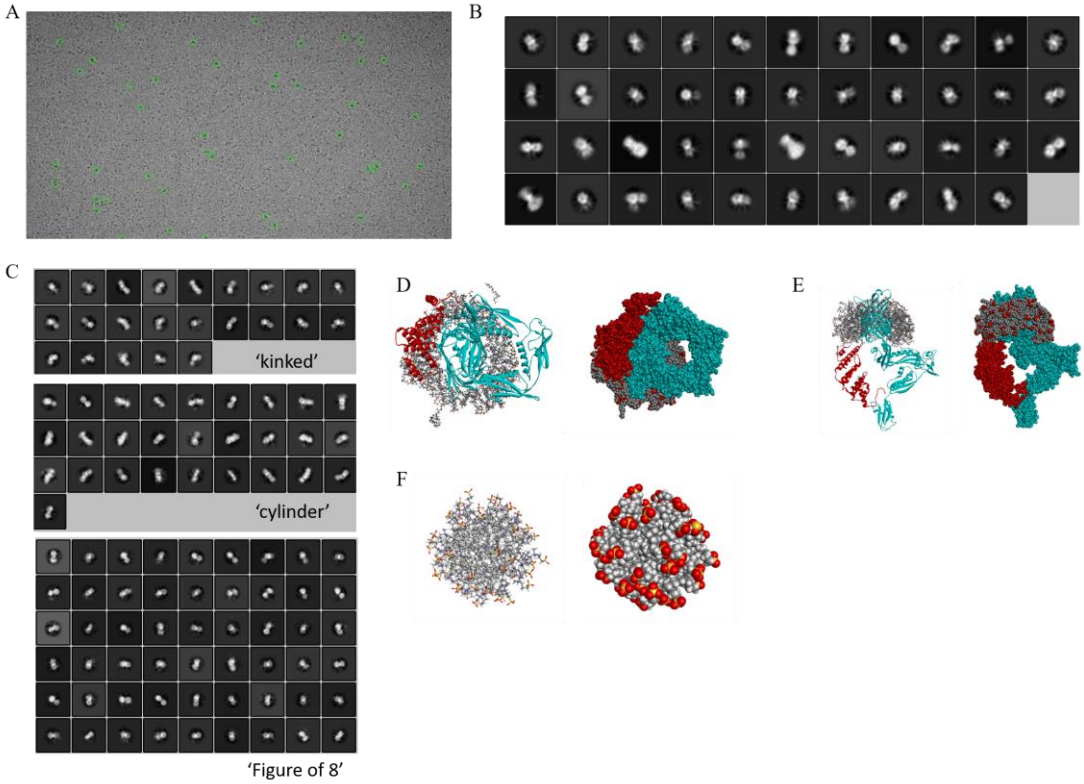


Figure 8.36: BamAD Cryo-TEM screening. A sample mixture of BamA and BamD was screened using a Titan Krios Microscope. A sample micrograph during particle picking (A), round six iteration of 2D classes showing viable classes (B), 2D class subsets showing heterogeneity in the sample (C), wire/ ribbon and CPK representations of BamAD in SB12 from an axial view (D) and side-on view (E), and a wire and CPK representation of a SB12 micelle (F) are shown. Junk particles (micelles) and heterogeneity ('kinked', 'cylinder', and 'figure of 8' poses) in the sample, which gives averaged molecules by stochastic gradient descent methodoly used by Relion in initial model building, are preventing further processing to the intial 3D model.

8.7.2 **MD simulations to be analysed**

Protein simulated	Trajectory run length	Environment	Purpose(s)
BamA	610 ns	Barrel inserted into POPE:POPG lipid membrane, system solvated in water	Examine flexibility and structural dynamics of BamA in lipid membrane Establish a baseline to compare to different environment and BamAD complex runs
BamA	150 ns	Barrel inserted into SB12 micelle; system solvated in water	Examine flexibility and structural dynamics of BamA in lipid membrane Establish a baseline to compare to different environment and BamAD complex runs SAXS analysis (validation, bead models)
BamD	500 ns	System solvated in water	Examine flexibility and structural dynamics of BamD in water Establish baseline to compare to different environment and BamAD complex runs
BamD	500 ns	BamD excluded from POPE:POPG lipid membrane, system solvated in water	Examine the tendency for BamD to associated with the lipid membrane Comparison of structural dynamics to the BamD in water simulation
BamAD	600 ns 500 ns (reversed orientation - BamD)	BamA β-barrel inserted into POPE:POPG lipid membrane, BamD and system solvated in water	Examine the flexibility and structural dynamics of the BamA-BamD complex in a solvated lipid membrane environment

	200 ns (docking run - physiological salt) 50 ns (Debye length altered)		Compare the flexibility and dynamics to the individual proteins and to the complex in different environments
BamAD	100 ns	BamA β-barrel inserted into SB12 micelle, system solvated in water	Compare flexibility and structural dynamics to the lipid environment to analyse the impact of the detergent SAXS analysis (validation, bead models)
BamAD	200 ns	Barrel inserted into POPE:POPG lipid membrane with LPS added to outer leaflet, system solvated in water	Compare flexibility and structural dynamics to the lipid environment to analyse the impact of adding LPS
DegP	100 ns	Protein solvated in water box	SAXS analysis (validation, bead models)
Cj1289	350 ns	Protein solvated in water box	SAXS analysis (validation, bead models)
LptD	50 ns	LptD in Fos-Choline 12 (FC12) micelle, system solvated in water	SAXS analysis (validation, bead models)
LptD	50 ns	LptD in SB12 micelle, system solvated in water	SAXS analysis (validation, bead models)

Table 8.11: MD simulations to be analysed. Table providing summary of unanalysed MD simulation by protein, trajectory run length, environment and the intended purpose or goal of the MD simulation.

8.8 Doctoral training programme appendices

8.8.1 PIP

Note to examiners:

This reflective statement, which describes the mandatory 3-month placement undertaken by BBSRC DTP students, is included as a compulsory appendix to the thesis. It was designed to capture the skills development which has taken place during the student's placement and the impact on their career plans it has had.

Where: National physical laboratory, Research Complex at Harwell

Project: Worked on recombinant expression of a confidential OM protein from *Mycobacterium tuberculosis*.

Outcomes: Progressed the project (membrane protein X) from gene isolation to 1 L recombinant expressions, identifying the appropriate conditions for culture, expressions, purification and pulldown (including detergent screens). The outcomes of this PIP work have revitalised a project that was shelved in 2013 and the project is now being actively pursued by NPL staff with the hope of eventual publication of a key target protein complex from a *M. tuberculosis* for exploitation as an antimicrobial target.

Skills developed:

Personal: The majority of my PhD project was done in isolation, and it was an enjoyable change of pace to work as part of a team. I developed my interpersonal and team working skills through interactions with the lab members. I delivered four presentations as part of my PhD, and delivering presentations fortnightly at the placement was useful for developing my presentation delivery, presentation slides layouts, and figure presentation, all of which are transferable skills that will serve me well as my career progresses.

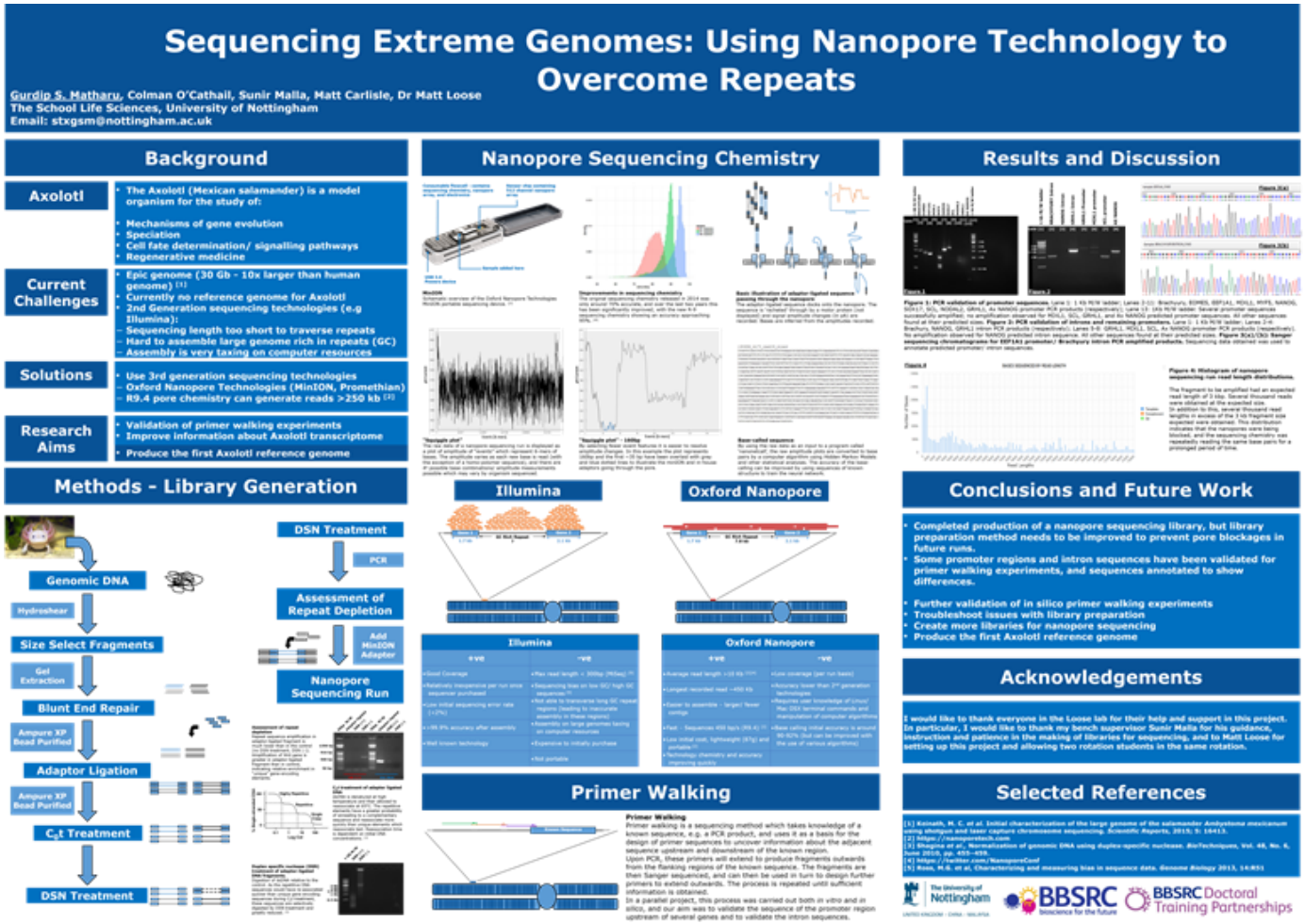
Technical: The experience has improved my working practices in terms of safety considerations, how I approach construct design and structural biology problems. I was exposed to a research institute with an industrial culture, and understand the advantages that a made for purpose structural biology institute offers to researchers and their research problems. The environment exposed me to many experience

researchers whom have to deliver projects on set time scales, which has fundamentally changed how I plan and approach my work, and my expectations. Finally, I was exposed to several high throughput approaches for generation of constructs, expression screening, and detergent screening, which I will try to implement and utilise in my future work.

Impact: I have experienced new technologies, alternative purification techniques, the effectiveness of high-throughput methods, and how to set up labs to be efficient as possible for preparation of proteins for structural analysis, broadening my knowledge and expanding my awareness in several areas that will improve my future career in science.

I feel that this placement was a good all-round experience and worked favourably in motivating myself to continue pursuing a career in science

8.8.2 **DTP Lab Rotation Poster**



8.8.3 PGR symposium poster

The University of Nottingham
UNITED KINGDOM · CHINA · MALAYSIA

Structural studies into the β-barrel assembly machinery (BAM) complex in Campylobacter

Gurdip Matharu, Rohanah Hussain, Guiliano Siligardi, Phil Williams, Weng Chan, and Boyan Bonev
School of Life Sciences, University of Nottingham

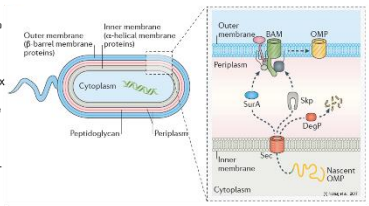
Introduction

The outer membrane (OM) in Gram negative bacteria contains several embedded proteins that perform numerous essential cellular functions such as nutrient uptake, signal transduction, and protein translocation. How these outer membrane proteins (OMPs) were assembled and inserted into the outer membrane was unknown until recently; with the assembly and insertion of an estimated 70% of OMPs in Gram negative bacteria now attributed to the β-Barrel Assembly Machinery (BAM) complex. How this complex achieves this feat in the absence of proton-motive force and ATP still remains unknown [1, 2].

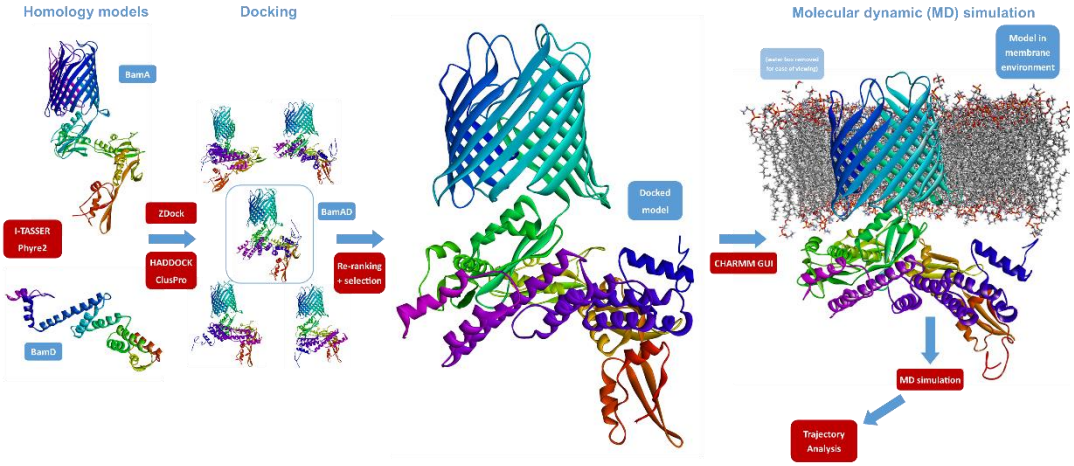
To investigate this function we focus on *Campylobacter jejuni*, a prominent pathogen in food-borne infections in developed countries [3], which, unlike E. coli, contains only the essential members of the BAM complex, BamA and BamD. Our study aims to characterise the complex through the use of structural techniques such as circular dichroism (CD), high resolution cryogenic electron microscopy (Cryo-EM), and *in silico* techniques such as molecular homology modelling, docking studies, and molecular dynamics (MD) simulations.

OMP transport story

Outer membrane proteins (OMPs) begin their journey to the outer membrane (OM) in the cytoplasm. They are picked up by SecA/B and delivered to SecYEG complex to be transported across the inner membrane. Once in the periplasm OMPs are picked up by chaperone proteins (SurA/ Skp) and delivered to the BAM complex in a partially-folded state for assembly and insertion into the OM.



In silico modelling process



Synchrotron radiation circular dichroism (SRCD)

Synchrotron radiation circular dichroism (SRCD) was used to determine the secondary structure content of the 16 strand β-barrel domain and polypeptide transport associated (POTRA) domains. Measurements were taken using a 1 to n spectra scan between 185 and 260 nm with a wavelength increment of 1nm at ambient room temperature.

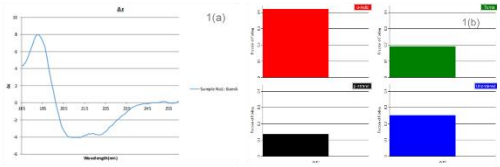


Figure 1: (a), 1 to n Spectra scan of the POTRA domains of BamA (0.8mg/ml BamA in 150mM NaF, 50mM Tris, 0.5% Sulfobetaine-12 buffer).
(b), Spectral analysis showing the POTRA domains of BamA to be 42.1% alpha helical, 13.6% β-sheet, 19.1% turns, and 25.2% unordered in secondary structure content by composition.

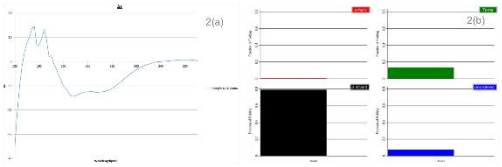


Figure 2: (a), 1 to n Spectra scan of the 16 strand β-barrel domain of BamA (0.3mg/ml BamA in 150mM NaF, 50mM Tris, 0.5% Sulfobetaine-12 buffer).
(b), Spectral analysis showing the 16 stand β-barrel of BamA to be 79.2% β-sheet, 13.3% turns and 7.5% unordered in secondary structure content by composition.

Trajectory analysis

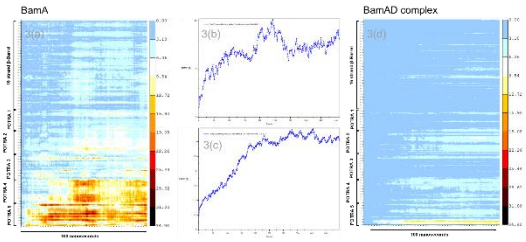


Figure 3: Trajectory analysis of C. jejuni BamA and BamAD complex over 100ns molecular dynamics simulation.
3(a) Heat map and 3(b) RMSD trace (scale 0-10Å) of BamA over 100ns MD simulation highlighting mobility of BamA POTRA domains in the absence of BamD.
3(c) RMSD trace (scale 0-3Å) and 3(d) Heat map of BamAD complex over 100ns MD simulation highlighting stabilisation of BamA POTRA domains in the presence of BamD.

Selected References

[1] Noinaj N, Gumbart JC, Buchanan SK. The β-barrel assembly machinery in motion. Nat Rev Microbiol. 2017;15(4):197–204.
[2] Walther DM, Rapaport D, Tommassen J. Biogenesis of β-barrel membrane proteins in bacteria and eukaryotes: Evolutionary conservation and divergence. Cell Mol Life Sci. 2009;66(17):2789–804.
[3] EFSA, European Food Safety Authority, and ECDC EC for DP and C, Boelaert F, Stoicescu A, Boloeil P, Georgiadis M, Mazzolini E, *et al.* Trends and Sources of Zoonoses , Zoonotic Agents and Food-borne Outbreaks in 2010. EFSA J. 2012;10(3):1–442.

Acknowledgements

I would like to thank everyone in the Bonev lab for their help and support on this project, the NMRC for access to and training in the Jeol 2100+ TEM, and Diamond Light Source for access to instrumentation and their staff members for their expertise in collecting the circular dichroism data. I would also like to thank the BBSRC Doctoral Training Partnership programme for providing the funding for this PhD project.

BBSRC bioscience for the future · BBSRC Doctoral Training Partnerships · diamond

8.9 References

1. Tamm, L. K., Hong, H., and Liang, B. (2004) Folding and assembly of β-barrel membrane proteins. *Biochim. Biophys. Acta - Biomembr. 1666*, 250–263.

2. Voulhoux, R., Bos, M. P., Geurtsen, J., *et al.* (2003) Role of a Highly Conserved Bacterial Protein in Outer Membrane Protein Assembly. *Science (80-.). 299*, 262–265.

3. Walther, D. M., Rapaport, D., and Tommassen, J. (2009) Biogenesis of β-barrel membrane proteins in bacteria and eukaryotes: evolutionary conservation and divergence. *Cell. Mol. Life Sci. 66*, 2789–2804.

4. Koebnik, R., Locher, K. P., and Van Gelder, P. (2000) Structure and function of bacterial outer membrane proteins: barrels in a nutshell. *Mol. Microbiol. 37*, 239–253.

5. Bakelar, J., Buchanan, S. K., and Noinaj, N. (2016) The structure of the β-barrel assembly machinery complex. *Science (80-.). 351*, 180–186.

6. Sandoval, C. M., Baker, S. L., Jansen, K., *et al.* (2011) Crystal Structure of BamD: An Essential Component of the β-Barrel Assembly Machinery of Gram-Negative Bacteria. *J. Mol. Biol. 409*, 348–357.

7. Noinaj, N., Kuszak, A. J., Gumbart, J. C., *et al.* (2013) Structural insight into the biogenesis of β-barrel membrane proteins. *Nature 501*, 385–390.

8. Gu, Y., Li, H., Dong, H., *et al.* (2016) Structural basis of outer membrane protein insertion by the BAM complex. *Nature 531*, 64–9.

9. Han, L., Zheng, J., Wang, Y., *et al.* (2016) Structure of the BAM complex and its implications for biogenesis of outer-membrane proteins. *Nat. Struct. Mol. Biol. 23*, 192–196.

10. Bergal, H. T., Hopkins, A. H., Metzner, S. I., *et al.* (2016) The Structure of a BamA-BamD Fusion Illuminates the Architecture of the β-Barrel Assembly Machine Core. *Structure 24*, 243–251.

11. European Food Safety Authority, E. C. for D. P. and C. (2012) The European Union Summary Report on Trends and Sources of Zoonoses, Zoonotic Agents and

Food-borne Outbreaks in 2010. *EFSA J. 10*, 1–442.

12. EFSA Panel on Biological Hazards (BIOHAZ). (2011) Scientific Opinion on Campylobacter in broiler meat production: control options and performance objectives and/or targets at different stages of the food chain. *EFSA J. 9*, 1–141.

13. Batz, M. B., Hoffman, S., and Morris, G. J. (2011) Ranking the Risks: The 10 Pathogen-Food Combinations with the Greatest Burden on Public Health. *Emerg. Pathog. Institute, Univ. Florida, Florida*, pp 1–70.

14. Bolton, D. J. (2015) Campylobacter virulence and survival factors. *Food Microbiol. 48*, 99–108.

15. Dasti, J. I., Tareen, A. M., Lugert, R., *et al.* (2010) *Campylobacter jejuni*: A brief overview on pathogenicity-associated factors and disease-mediating mechanisms. *Int. J. Med. Microbiol. 300*, 205–211.

16. Gibreel, A., and Taylor, D. E. (2006) Macrolide resistance in *Campylobacter jejuni* and *Campylobacter coli*. *J. Antimicrob. Chemother. 58*, 243–55.

17. Alfredson, D. A., and Korolik, V. (2007) Antibiotic resistance and resistance mechanisms in *Campylobacter jejuni* and *Campylobacter coli*. *FEMS Microbiol. Lett. 277*, 123–132.

18. Gloge, F., Becker, A. H., Kramer, G., *et al.* (2014) Co-translational mechanisms of protein maturation. *Curr. Opin. Struct. Biol. 24*, 24–33.

19. Berg, B. van den, Clemons, W. M., Collinson, I., *et al.* (2004) X-ray structure of a protein-conducting channel. *Nature 427*, 36–44.

20. de Keyzer, J., van der Does, C., and Driessen, A. J. M. (2003) The bacterial translocase: a dynamic protein channel complex. *Cell. Mol. Life Sci. 60*, 2034–2052.

21. Osborne, A. R., Rapoport, T. A., and van den Berg, B. (2005) Protein translocation by the Sec61/SecY channel. *Annu. Rev. Cell Dev. Biol. 21*, 529–50.

22. Papanikou, E., Karamanou, S., and Economou, A. (2007) Bacterial protein secretion through the translocase nanomachine. *Nat. Rev. Microbiol. 5*, 839–851.

23. Eppens, E. F., Nouwen, N., and Tommassen, J. (1997) Folding of a bacterial outer membrane protein during passage through the periplasm. *EMBO J. 16*, 4295–

301.

24. du Plessis, D. J. F., Nouwen, N., and Driessen, A. J. M. (2011) The Sec translocase. *Biochim. Biophys. Acta 1808*, 851–65.

25. Rizzitello, A. E., Harper, J. R., and Silhavy, T. J. (2001) Genetic evidence for parallel pathways of chaperone activity in the periplasm of *Escherichia coli*. *J. Bacteriol. 183*, 6794–800.

26. Sklar, J. G., Wu, T., Gronenberg, L. S., *et al.* (2007) Lipoprotein SmpA is a component of the YaeT complex that assembles outer membrane proteins in *Escherichia coli*. *Proc. Natl. Acad. Sci. 104*, 6400–6405.

27. Mahoney, T. F., Ricci, D. P., and Silhavy, T. J. (2016) Classifying β-Barrel Assembly Substrates by Manipulating Essential Bam Complex Members. *J. Bacteriol. 198*, 1984–1992.

28. Bos, M. P., Robert, V., and Tommassen, J. (2007) Biogenesis of the gram-negative bacterial outer membrane. *Annu. Rev. Microbiol. 61*, 191–214.

29. Heuck, A., Schleiffer, A., and Clausen, T. (2011) Augmenting β-augmentation: structural basis of how BamB binds BamA and may support folding of outer membrane proteins. *J. Mol. Biol. 406*, 659–66.

30. Wu, T., Malinverni, J., Ruiz, N., *et al.* (2005) Identification of a Multicomponent Complex Required for Outer Membrane Biogenesis in *Escherichia coli*. *Cell 121*, 235–245.

31. Doerrler, W. T., and Raetz, C. R. H. (2005) Loss of Outer Membrane Proteins without Inhibition of Lipid Export in an *Escherichia coli* YaeT Mutant. *J. Biol. Chem. 280*, 27679–27687.

32. Malinverni, J. C., Werner, J., Kim, S., *et al.* (2006) YfiO stabilizes the YaeT complex and is essential for outer membrane protein assembly in *Escherichia coli*. *Mol. Microbiol. 61*, 151–164.

33. Voulhoux, R., and Tommassen, J. (2004) Omp85, an evolutionarily conserved bacterial protein involved in outer-membrane-protein assembly. *Res. Microbiol. 155*, 129–135.

34. Estrada Mallarino, L., Fan, E., Odermatt, M., *et al.* (2015) TtOmp85, a β-Barrel Assembly Protein, Functions by Barrel Augmentation. *Biochemistry 54*, 844–852.

35. Habib, S. J., Waizenegger, T., Niewienda, A., *et al.* (2007) The N-terminal domain of Tob55 has a receptor-like function in the biogenesis of mitochondrial β-barrel proteins. *J. Cell Biol. 176*, 77–88.

36. Webb, C. T., Heinz, E., and Lithgow, T. (2012) Evolution of the β-barrel assembly machinery. *Trends Microbiol. 20*, 612–620.

37. Albrecht, R., and Zeth, K. (2011) Structural Basis of Outer Membrane Protein Biogenesis in Bacteria. *J. Biol. Chem. 286*, 27792–27803.

38. Anwari, K., Webb, C. T., Poggio, S., *et al.* (2012) The evolution of new lipoprotein subunits of the bacterial outer membrane BAM complex. *Mol. Microbiol. 84*, 832–844.

39. Schleiff, E., Jelic, M., and Soll, J. (2003) A GTP-driven motor moves proteins across the outer envelope of chloroplasts. *Proc. Natl. Acad. Sci. U. S. A. 100*, 4604–4609.

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

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

42. Nesper, J., Brosig, A., Ringler, P., *et al.* (2008) Omp85 Tt from *Thermus thermophilus* HB27: an Ancestral Type of the Omp85 Protein Family. *J. Bacteriol. 190*, 4568–4575.

43. Kutik, S., Stojanovski, D., Becker, L., *et al.* (2008) Dissecting Membrane Insertion of Mitochondrial β-Barrel Proteins. *Cell 132*, 1011–1024.

44. Iadanza, M. G., Higgins, A. J., Schiffrin, B., *et al.* (2016) Lateral opening in the intact β-barrel assembly machinery captured by cryo-EM. *Nat. Commun. 7*, 12865.

45. Clantin, B., Delattre, A.-S., Rucktooa, P., *et al.* (2007) Structure of the membrane protein FhaC: a member of the Omp85-TpsB transporter superfamily.

Science 317, 957–61.

46. Leonard-Rivera, M., and Misra, R. (2012) Conserved Residues of the Putative L6 Loop of *Escherichia coli* BamA Play a Critical Role in the Assembly of -Barrel Outer Membrane Proteins, Including That of BamA Itself. *J. Bacteriol. 194*, 4662–4668.

47. Schulz, G. E. (2000) beta-Barrel membrane proteins. *Curr. Opin. Struct. Biol. 10*, 443–7.

48. Noinaj, N., Gumbart, J. C., and Buchanan, S. K. (2017) The β-barrel assembly machinery in motion. *Nat. Rev. Microbiol. 15*, 197–204.

49. Rigel, N. W., Schwalm, J., Ricci, D. P., *et al.* (2012) BamE modulates the *Escherichia coli* beta-barrel assembly machine component BamA. *J. Bacteriol. 194*, 1002–8.

50. Knowles, T. J., Jeeves, M., Bobat, S., *et al.* (2008) Fold and function of polypeptide transport-associated domains responsible for delivering unfolded proteins to membranes. *Mol. Microbiol. 68*, 1216–1227.

51. Gatzeva-Topalova, P. Z., Walton, T. A., and Sousa, M. C. (2008) Crystal Structure of YaeT: Conformational Flexibility and Substrate Recognition. *Structure 16*, 1873–1881.

52. Gatzeva-Topalova, P. Z., Warner, L. R., Pardi, A., *et al.* (2010) Structure and Flexibility of the Complete Periplasmic Domain of BamA: The Protein Insertion Machine of the Outer Membrane. *Structure 18*, 1492–1501.

53. Ertel, F., Mirus, O., Bredemeier, R., *et al.* (2005) The Evolutionarily Related β-Barrel Polypeptide Transporters from *Pisum Sativum* and Nostoc PCC7120 Contain Two Distinct Functional Domains. *J. Biol. Chem. 280*, 28281–28289.

54. Bennion, D., Charlson, E. S., Coon, E., *et al.* (2010) Dissection of β-barrel outer membrane protein assembly pathways through characterizing BamA POTRA 1 mutants of *Escherichia coli*. *Mol. Microbiol. 77*, 1153–1171.

55. Kim, S., Malinverni, J. C., Sliz, P., *et al.* (2007) Structure and function of an essential component of the outer membrane protein assembly machine. *Science 317*, 961–4.

56. Fleming, P. J., Patel, D. S., Wu, E. L., *et al.* (2016) BamA POTRA Domain Interacts with a Native Lipid Membrane Surface. *Biophys. J. 110*, 2698–2709.

57. Sinnige, T., Weingarth, M., Renault, M., *et al.* (2014) Solid-State NMR Studies of Full-Length BamA in Lipid Bilayers Suggest Limited Overall POTRA Mobility. *J. Mol. Biol. 426*, 2009–2021.

58. Moslavac, S., Mirus, O., Bredemeier, R., *et al.* (2005) Conserved pore-forming regions in polypeptide- transporting proteins. *FEBS J. 272*, 1367–1378.

59. Jansen, K. B., Baker, S. L., and Sousa, M. C. (2015) Crystal Structure of BamB Bound to a Periplasmic Domain Fragment of BamA, the Central Component of the β-Barrel Assembly Machine. *J. Biol. Chem. 290*, 2126–2136.

60. Noinaj, N., and Buchanan, S. K. (2014) Structural insights into the transport of small molecules across membranes. *Curr. Opin. Struct. Biol. 27*, 8–15.

61. Noinaj, N., Kuszak, A. J., Balusek, C., *et al.* (2014) Lateral Opening and Exit Pore Formation Are Required for BamA Function. *Structure 22*, 1055–1062.

62. Dong, C., Hou, H. F., Yang, X., *et al.* (2012) Structure of *Escherichia coli* BamD and its functional implications in outer membrane protein assembly. *Acta Crystallogr. D. Biol. Crystallogr. 68*, 95–101.

63. Struyvé, M., Moons, M., and Tommassen, J. (1991) Carboxy-terminal phenylalanine is essential for the correct assembly of a bacterial outer membrane protein. *J. Mol. Biol. 218*, 141–8.

64. Robert, V., Volokhina, E. B., Senf, F., *et al.* (2006) Assembly factor Omp85 recognizes its outer membrane protein substrates by a species-specific C-terminal motif. *PLoS Biol. 4*, e377.

65. Kim, K. H., Aulakh, S., and Paetzel, M. (2011) Crystal Structure of β-Barrel Assembly Machinery BamCD Protein Complex. *J. Biol. Chem. 286*, 39116–39121.

66. Kim, K. H., Aulakh, S., Tan, W., *et al.* (2011) Crystallographic analysis of the C-terminal domain of the *Escherichia coli* lipoprotein BamC. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun. 67*, 1350–1358.

67. Masuda, T., Saito, N., Tomita, M., *et al.* (2009) Unbiased Quantitation of

Escherichia coli Membrane Proteome Using Phase Transfer Surfactants. *Mol. Cell. Proteomics 8*, 2770–2777.

68. Hagan, C. L., Silhavy, T. J., and Kahne, D. (2011) β-Barrel Membrane Protein Assembly by the Bam Complex. *Annu. Rev. Biochem. 80*, 189–210.

69. Knowles, T. J., Browning, D. F., Jeeves, M., *et al.* (2011) Structure and function of BamE within the outer membrane and the β-barrel assembly machine. *EMBO Rep. 12*, 123–128.

70. Sklar, J. G., Wu, T., Kahne, D., *et al.* (2007) Defining the roles of the periplasmic chaperones SurA, Skp, and DegP in *Escherichia coli*. *Genes Dev. 21*, 2473–2484.

71. Ruiz, N., and Silhavy, T. J. (2005) Sensing external stress: watchdogs of the *Escherichia coli* cell envelope. *Curr. Opin. Microbiol. 8*, 122–126.

72. Noinaj, N., Fairman, J. W., and Buchanan, S. K. (2011) The Crystal Structure of BamB Suggests Interactions with BamA and Its Role within the BAM Complex. *J. Mol. Biol. 407*, 248–260.

73. Warner, L. R., Varga, K., Lange, O. F., *et al.* (2011) Structure of the BamC Two-Domain Protein Obtained by Rosetta with a Limited NMR Data Set. *J. Mol. Biol. 411*, 83–95.

74. Albrecht, R., and Zeth, K. (2010) Crystallization and preliminary X-ray data collection of the *Escherichia coli* lipoproteins BamC, BamD and BamE. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun. 66*, 1586–1590.

75. Endo, T., Kawano, S., and Yamano, K. (2011) BamE structure: the assembly of β-barrel proteins in the outer membranes of bacteria and mitochondria. *EMBO Rep. 12*, 94–95.

76. Anwari, K., Poggio, S., Perry, A., *et al.* (2010) A Modular BAM Complex in the Outer Membrane of the α-Proteobacterium *Caulobacter Crescentus*. *PLoS One 5*, e8619.

77. Grizot, S., and Buchanan, S. K. (2004) Structure of the OmpA-like domain of RmpM from *Neisseria meningitidis*. *Mol. Microbiol. 51*, 1027–1037.

78. Walton, T. A., and 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.

79. Wülfing, C., and Plückthun, A. (1994) Protein folding in the periplasm of *Escherichia coli*. *Mol. Microbiol. 12*, 685–92.

80. Narita, S. ichiro, and Tokuda, H. (2009) Biochemical characterization of an ABC transporter LptBFGC complex required for the outer membrane sorting of lipopolysaccharides. *FEBS Lett. 583*, 2160–2164.

81. Shahrizal, M., Daimon, Y., Tanaka, Y., *et al.* (2019) Structural Basis for the Function of the β-Barrel Assembly-Enhancing Protease BepA. *J. Mol. Biol. 431*, 625–635.

82. Harms, N., Koningstein, G., Dontje, W., *et al.* (2001) The Early Interaction of the Outer Membrane Protein PhoE with the Periplasmic Chaperone Skp Occurs at the Cytoplasmic Membrane. *J. Biol. Chem. 276*, 18804–18811.

83. Qu, J., Mayer, C., Behrens, S., *et al.* (2007) The Trimeric Periplasmic Chaperone Skp of *Escherichia coli* Forms 1:1 Complexes with Outer Membrane Proteins via Hydrophobic and Electrostatic Interactions. *J. Mol. Biol. 374*, 91–105.

84. Chen, R., and Henning, U. (1996) Aperiplasmic protein (Skp) of *Escherichia coli* selectively binds a class of outer membrane proteins. *Mol. Microbiol. 19*, 1287–1294.

85. Walton, T. A., Sandoval, C. M., Fowler, C. A., *et al.* (2009) The cavity-chaperone Skp protects its substrate from aggregation but allows independent folding of substrate domains. *Proc. Natl. Acad. Sci. 106*, 1772–1777.

86. Tormo, A., Almirón, M., and Kolter, R. (1990) surA, an *Escherichia coli* gene essential for survival in stationary phase. *J. Bacteriol. 172*, 4339–4347.

87. Justice, S. S., Hunstad, D. A., Harper, J. R., *et al.* (2005) Periplasmic peptidyl prolyl cis-trans isomerases are not essential for viability, but SurA is required for pilus biogenesis in *Escherichia coli*. *J. Bacteriol. 187*, 7680–6.

88. Xu, X., Wang, S., Hu, Y.-X., *et al.* (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.

89. Bitto, E., and McKay, D. B. (2002) Crystallographic Structure of SurA, a Molecular Chaperone that Facilitates Folding of Outer Membrane Porins. *Structure 10*, 1489–1498.

90. Krojer, T., Sawa, J., Schäfer, E., *et al.* (2008) Structural basis for the regulated protease and chaperone function of DegP. *Nature 453*, 885–890.

91. Spiess, C., Beil, A., and Ehrmann, M. (1999) A temperature-dependent switch from chaperone to protease in a widely conserved heat shock protein. *Cell 97*, 339–47.

92. Kolmar, H., Waller, P. R. H., and Sauer, R. T. (1996) The DegP and DegQ periplasmic endoproteases of *Escherichia coli*: specificity for cleavage sites and substrate conformation. *J. Bacteriol. 178*, 5925–5929.

93. Cho, S.-H., Szewczyk, J., Pesavento, C., *et al.* (2014) Detecting Envelope Stress by Monitoring β-Barrel Assembly. *Cell 159*, 1652–1664.

94. Duguay, A. R., and Silhavy, T. J. (2004) Quality control in the bacterial periplasm. *Biochim. Biophys. Acta 1694*, 121–34.

95. Bury-Moné, S., Nomane, Y., Reymond, N., *et al.* (2009) Global Analysis of Extracytoplasmic Stress Signaling in *Escherichia coli*. *PLoS Genet. 5*, e1000651.

96. Narita, S., Masui, C., Suzuki, T., *et al.* (2013) Protease homolog BepA (YfgC) promotes assembly and degradation of β-barrel membrane proteins in *Escherichia coli*. *Proc. Natl. Acad. Sci. 110*.

97. Daimon, Y., Iwama-Masui, C., Tanaka, Y., *et al.* (2017) The TPR domain of BepA is required for productive interaction with substrate proteins and the β-barrel assembly machinery complex. *Mol. Microbiol. 106*, 760–776.

98. McMorran, L. M., Brockwell, D. J., and Radford, S. E. (2014) Mechanistic studies of the biogenesis and folding of outer membrane proteins *in vitro* and *in vivo*: What have we learned to date? *Arch. Biochem. Biophys. 564*, 265–280.

99. Bakelar, J., Buchanan, S. K., and Noinaj, N. (2017) Structural snapshots of the β-barrel assembly machinery. *FEBS J. 284*, 1778–1786.

100. Gessmann, D., Chung, Y. H., Danoff, E. J., *et al.* (2014) Outer membrane β-barrel protein folding is physically controlled by periplasmic lipid head groups and BamA. *Proc. Natl. Acad. Sci. 111*, 5878–5883.

101. Raetz, C. R., and Dowhan, W. (1990) Biosynthesis and function of phospholipids in *Escherichia coli*. *J. Biol. Chem. 265*, 1235–8.

102. Dowhan, W. (1997) Molecular basis for membrane phospholipid diversity: why are there so many lipids? *Annu. Rev. Biochem. 66*, 199–232.

103. Huysmans, G. H. M., Guilvout, I., Chami, M., *et al.* (2015) Lipids assist the membrane insertion of a BAM-independent outer membrane protein. *Sci. Rep. 5*, 15068.

104. de Cock, H., Struyvé, M., Kleerebezem, M., *et al.* (1997) Role of the carboxy-terminal phenylalanine in the biogenesis of outer membrane protein PhoE of *Escherichia coli*K-12. *J. Mol. Biol. 269*, 473–478.

105. Hagan, C. L., and Kahne, D. (2011) The Reconstituted *Escherichia coli* Bam Complex Catalyzes Multiple Rounds of β-Barrel Assembly. *Biochemistry 50*, 7444–7446.

106. White, D. A., Barlow, A. K., Clarke, I. N., *et al.* (1990) Stable expression of meningococcal class 1 protein in an antigenically reactive form in outer membranes of *Escherichia coli*. *Mol. Microbiol. 4*, 769–776.

107. Hagan, C. L., Wzorek, J. S., and Kahne, D. (2015) Inhibition of the β-barrel assembly machine by a peptide that binds BamD. *Proc. Natl. Acad. Sci. 112*, 2011–2016.

108. Kale, A., Phansopa, C., Suwannachart, C., *et al.* (2011) The Virulence Factor PEB4 (Cj0596) and the Periplasmic Protein Cj1289 Are Two Structurally Related SurA-like Chaperones in the Human Pathogen *Campylobacter jejuni*. *J. Biol. Chem. 286*, 21254–21265.

109. Taylor, A. J., Zakai, S. A. I., and Kelly, D. J. (2017) The Periplasmic Chaperone Network of *Campylobacter jejuni*: Evidence that SalC (Cj1289) and PpiD (Cj0694) Are Involved in Maintaining Outer Membrane Integrity. *Front. Microbiol. 8*, 531.

110. Zarzecka, U., Grinzato, A., Kandiah, E., *et al.* (2020) Functional analysis and cryo-electron microscopy of *Campylobacter jejuni* serine protease HtrA. *Gut Microbes 12*, 1–16.

111. Lim, H. N., Lee, Y., and Hussein, R. (2011) Fundamental relationship between operon organization and gene expression. *Proc. Natl. Acad. Sci. 108*, 10626–10631.

112. Jumper, J., Evans, R., Pritzel, A., *et al.* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature 596*, 583–589.

113. Varadi, M., Anyango, S., Deshpande, M., *et al.* (2022) AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res. 50*, D439–D444.

114. Madeira, F., Park, Y. mi, Lee, J., *et al.* (2019) The EMBL-EBI search and sequence analysis tools APIs in 2019. *Nucleic Acids Res. 47*, W636–W641.

115. Davis, L., and DiRita, V. (2008) Growth and Laboratory Maintenance of *Campylobacter jejuni*. *Curr. Protoc. Microbiol. 10*, 1–7.

116. Luechtefeld, N. W., Reller, L. B., Blaser, M. J., *et al.* (1982) Comparison of atmospheres of incubation for primary isolation of *Campylobacter fetus subsp. jejuni* from animal specimens: 5% oxygen versus candle jar. *J. Clin. Microbiol. 15*, 53–57.

117. Khoshnood, A., Lukanov, B., and Firoozabadi, A. (2016) Temperature Effect on Micelle Formation: Molecular Thermodynamic Model Revisited. *Langmuir 32*, 2175–2183.

118. Roy, A., Kucukural, A., and Zhang, Y. (2010) I-TASSER: a unified platform for automated protein structure and function prediction. *Nat. Protoc. 5*, 725–738.

119. Wu, E. L., Cheng, X., Jo, S., *et al.* (2014) CHARMM-GUI Membrane Builder toward realistic biological membrane simulations. *J. Comput. Chem. 35*, 1997–2004.

120. Cheng, X., Jo, S., Lee, H. S., *et al.* (2013) CHARMM-GUI Micelle Builder for Pure/Mixed Micelle and Protein/Micelle Complex Systems. *J. Chem. Inf. Model. 53*, 2171–2180.

121. Jo, S., Kim, T., Iyer, V. G., *et al.* (2008) CHARMM-GUI: A web-based graphical user interface for CHARMM. *J. Comput. Chem. 29*, 1859–1865.

122. Lee, J., Cheng, X., Swails, J. M., *et al.* (2016) CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J. Chem. Theory Comput. 12*, 405–413.

123. Phillips, J. C., Hardy, D. J., Maia, J. D. C., *et al.* (2020) Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J. Chem. Phys. 153*, 044130.

124. Kabsch, W., and Sander, C. (1983) Dictionary of protein secondary structure: Pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers 22*, 2577–2637.

125. Frishman, D., and Argos, P. (1995) Knowledge-based protein secondary structure assignment. *Proteins Struct. Funct. Genet. 23*, 566–579.

126. Labesse, G., Colloc'h, N., Pothier, J., *et al.* (1997) P-SEA: a new efficient assignment of secondary structure from Cα trace of proteins. *Bioinformatics 13*, 291–295.

127. Taylor, W. R. (2001) Defining linear segments in protein structure. *J. Mol. Biol. 310*, 1135–1150.

128. Srinivasan, R., and Rose, G. D. (1999) A physical basis for protein secondary structure. *Proc. Natl. Acad. Sci. 96*, 14258–14263.

129. Greenfield, N. J. (2006) Using circular dichroism spectra to estimate protein secondary structure. *Nat. Protoc. 1*, 2876–2890.

130. Hussain, R., Benning, K., Javorfi, T., *et al.* (2015) CDApps : integrated software for experimental planning and data processing at beamline B23, Diamond Light Source. *J. Synchrotron Radiat. 22*, 465–468.

131. Provencher, S. W., and Gloeckner, J. (1981) Estimation of globular protein secondary structure from circular dichroism. *Biochemistry 20*, 33–37.

132. van Stokkum, I. H. M., Spoelder, H. J. W., Bloemendal, M., *et al.* (1990) Estimation of protein secondary structure and error analysis from circular dichroism spectra. *Anal. Biochem. 191*, 110–118.

133. Kelley, L. A., Mezulis, S., Yates, C. M., *et al.* (2015) The Phyre2 web portal for

protein modeling, prediction and analysis. *Nat. Protoc. 10*, 845–858.

134. Schneidman-Duhovny, D., Hammel, M., Tainer, J. A., *et al.* (2013) Accurate SAXS Profile Computation and its Assessment by Contrast Variation Experiments. *Biophys. J. 105*, 962–974.

135. Schneidman-Duhovny, D., Hammel, M., Tainer, J. A., *et al.* (2016) FoXS, FoXSDock and MultiFoXS: Single-state and multi-state structural modeling of proteins and their complexes based on SAXS profiles. *Nucleic Acids Res. 44*, W424–W429.

136. Franke, D., and Svergun, D. I. (2009) DAMMIF , a program for rapid ab-initio shape determination in small-angle scattering. *J. Appl. Crystallogr. 42*, 342–346.

137. Kozin, M. B., and Svergun, D. I. (2001) Automated matching of high- and low-resolution structural models. *J. Appl. Crystallogr. 34*, 33–41.

138. Volkov, V. V, and Svergun, D. I. (2003) Uniqueness of ab initio shape determination in small-angle scattering. *J. Appl. Crystallogr. 36*, 860–864.

139. Svergun, D. I. (1999) Restoring Low Resolution Structure of Biological Macromolecules from Solution Scattering Using Simulated Annealing. *Biophys. J. 76*, 2879–2886.

140. Yang, J., Yan, R., Roy, A., *et al.* (2015) The I-TASSER Suite: protein structure and function prediction. *Nat. Methods 12*, 7–8.

141. Pettersen, E. F., Goddard, T. D., Huang, C. C., *et al.* (2004) UCSF Chimera?A visualization system for exploratory research and analysis. *J. Comput. Chem. 25*, 1605–1612.

142. Pierce, B. G., Wiehe, K., Hwang, H., *et al.* (2014) ZDOCK server: interactive docking prediction of protein-protein complexes and symmetric multimers. *Bioinformatics 30*, 1771–1773.

143. Jiménez-García, B., Pons, C., and Fernández-Recio, J. (2013) pyDockWEB: a web server for rigid-body protein–protein docking using electrostatics and desolvation scoring. *Bioinformatics 29*, 1698–1699.

144. Lyskov, S., and Gray, J. J. (2008) The RosettaDock server for local protein-

protein docking. *Nucleic Acids Res. 36*, W233–W238.

145. Dolinsky, T. J., Czodrowski, P., Li, H., *et al.* (2007) PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res. 35*, W522–W525.

146. Olsson, M. H. M., Søndergaard, C. R., Rostkowski, M., *et al.* (2011) PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput. 7*, 525–37.

147. Lomize, M. A., Pogozheva, I. D., Joo, H., *et al.* (2012) OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res. 40*, D370–D376.

148. Hardcastle, M. (2018) Structure, function and mechanism of beta-barrel protein assembly in *Campylobacter jejuni* (University of Nottingham. Theses. Life Sciences, Ed.). Thesis (PhD)--University of Nottingham.

149. Nielsen, H., Engelbrecht, J., Brunak, S., *et al.* (1997) Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng. Des. Sel. 10*, 1–6.

150. Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., *et al.* (2019) SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nat. Biotechnol. 37*, 420–423.

151. Juncker, A. S., Willenbrock, H., von Heijne, G., *et al.* (2003) Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci. 12*, 1652–1662.

152. Hull, D. M., Harrell, E., van Vliet, A. H. M., *et al.* (2021) Antimicrobial resistance and interspecies gene transfer in *Campylobacter coli* and *Campylobacter jejuni* isolated from food animals, poultry processing, and retail meat in North Carolina, 2018–2019. *PLoS One 16*, e0246571.

153. Noland, C. L., Kattke, M. D., Diao, J., *et al.* (2017) Structural insights into lipoprotein N-acylation by *Escherichia coli* apolipoprotein N-acyltransferase. *Proc. Natl. Acad. Sci. 114*.

154. Hayashi, S., and Wu, H. C. (1990) Lipoproteins in bacteria. *J. Bioenerg. Biomembr. 22*, 451–471.

155. von Heijne, G. (1989) The structure of signal peptides from bacterial lipoproteins. *"Protein Eng. Des. Sel. 2*, 531–534.

156. Babu, M. M., Priya, M. L., Selvan, A. T., *et al.* (2006) A Database of Bacterial Lipoproteins (DOLOP) with Functional Assignments to Predicted Lipoproteins. *J. Bacteriol. 188*, 2761–2773.

157. Jin, S., Joe, A., Lynett, J., *et al.* (2004) JlpA, a novel surface-exposed lipoprotein specific to *Campylobacter jejuni*, mediates adherence to host epithelial cells. *Mol. Microbiol. 39*, 1225–1236.

158. Berova, N., Nakanishi, K., and Woody, R. (2000) Circular dichroism: Principles and applications (Berova, N., Nakanishi, K., and Woody, R., Eds.) 2nd ed. book, Wiley-VCH, New York ; Chichester.

159. Woody, R. W. (1995) [4] Circular dichroism, in *Methods in Enzymology*, pp 34–71. Academic Press.

160. Nordén, B. (2010) Linear dichroism and circular dichroism : a textbook on polarized-light spectroscopy (Rodger, A., Dafforn, T., and Royal Society of Chemistry (Great Britain), Eds.). Royal Society of Chemistry, Cambridge.

161. Sreerama, N., Manning, M. C., Powers, M. E., *et al.* (1999) Tyrosine, Phenylalanine, and Disulfide Contributions to the Circular Dichroism of Proteins: Circular Dichroism Spectra of Wild-Type and Mutant Bovine Pancreatic Trypsin Inhibitor. *Biochemistry 38*, 10814–10822.

162. Wallace, B. A. (2003) Analyses of circular dichroism spectra of membrane proteins. *Protein Sci. 12*, 875–884.

163. Greenfield, N. J. Circular Dichroism Analysis for Protein–Protein Interactions, in *Protein-Protein Interactions*, pp 55–78. Humana Press, New Jersey.

164. Johnson, W. C. (1988) Secondary Structure of Proteins Through Circular Dichroism Spectroscopy. *Annu. Rev. Biophys. Biophys. Chem. 17*, 145–166.

165. Ieva, R., and Bernstein, H. D. (2009) Interaction of an autotransporter passenger domain with BamA during its translocation across the bacterial outer membrane. *Proc. Natl. Acad. Sci. 106*, 19120–19125.

166. Wang, X., Peterson, J. H., and Bernstein, H. D. (2021) Bacterial Outer Membrane Proteins Are Targeted to the Bam Complex by Two Parallel Mechanisms. *MBio 12*, 1–19.

167. Greenfield, N. J. (2006) Using circular dichroism collected as a function of temperature to determine the thermodynamics of protein unfolding and binding interactions. *Nat. Protoc. 1*, 2527–2535.

168. Mateus, A., Bobonis, J., Kurzawa, N., *et al.* (2018) Thermal proteome profiling in bacteria: probing protein state *in vivo*. *Mol. Syst. Biol. 14*, e8242.

169. Tan, C. S. H., Go, K. D., Bisteau, X., *et al.* (2018) Thermal proximity coaggregation for system-wide profiling of protein complex dynamics in cells. *Science 359*, 1170–1177.

170. Hussain, R., Hughes, C. S., Jávorfi, T., *et al.* (2018) To Boil an Egg: Substrate Binding Affects Critical Stability in Thermal Unfolding of Proteins. *J. Phys. Chem. B 122*, 2213–2218.

171. Sreerama, N., and Woody, R. W. (1993) A Self-Consistent Method for the Analysis of Protein Secondary Structure from Circular Dichroism. *Anal. Biochem. 209*, 32–44.

172. Sreerama, N., Venyaminov, S. Y. U., and Woody, R. W. (2008) Estimation of the number of α-helical and β-strand segments in proteins using circular dichroism spectroscopy. *Protein Sci. 8*, 370–380.

173. Manavalan, P., and Johnson, W. C. (1987) Variable selection method improves the prediction of protein secondary structure from circular dichroism spectra. *Anal. Biochem. 167*, 76–85.

174. Compton, L. A., and Johnson, W. C. (1986) Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication. *Anal. Biochem. 155*, 155–167.

175. Sreerama, N., and Woody, R. W. (2000) Estimation of Protein Secondary Structure from Circular Dichroism Spectra: Comparison of CONTIN, SELCON, and CDSSTR Methods with an Expanded Reference Set. *Anal. Biochem. 287*, 252–260.

176. Sreerama, N. (2004) On the analysis of membrane protein circular dichroism

spectra. *Protein Sci. 13*, 100–112.

177. Park, K., Perczel, A., and Fasman, G. D. (1992) Differentiation between transmembrane helices and peripheral helices by the deconvolution of circular dichroism spectra of membrane proteins. *Protein Sci. 1*, 1032–1049.

178. Woody, R. W. (1996) Theory of circular dichroism of proteins., in *Circular Dichroism and the Conformational Analysis of Biomolecules* (Fasman, G. D., Ed.), pp 25–67. Plenum Press, New York.

179. Shi, Z., Woody, R. W., and Kallenbach, N. R. (2002) Is polyproline II a major backbone conformation in unfolded proteins? *Adv. Protein Chem. 62*, 163–240.

180. Majeed, S., Ahmad, A. B., Sehar, U., *et al.* (2021) Lipid Membrane Mimetics in Functional and Structural Studies of Integral Membrane Proteins. *Membranes (Basel). 11*, 685.

181. Schiffrin, B., Machin, J. M., Karamanos, T. K., *et al.* (2022) Dynamic interplay between the periplasmic chaperone SurA and the BAM complex in outer membrane protein folding. *Commun. Biol. 5*, 560.

182. Behrens, S. (2001) The SurA periplasmic PPIase lacking its parvulin domains functions *in vivo* and has chaperone activity. *EMBO J. 20*, 285–294.

183. Thompson, M. K., Ehlinger, A. C., and Chazin, W. J. (2017) Analysis of Functional Dynamics of Modular Multidomain Proteins by SAXS and NMR, in *Methods in Enzymology* 1st ed., pp 49–76. Elsevier Inc.

184. Schnableggar, H., and Singh, Y. (2017) The SAXS guide: Getting acquainted with the principles (4th edition). *Ant. Paar* 1–163.

185. Putnam, C. D., Hammel, M., Hura, G. L., *et al.* (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q. Rev. Biophys. 40*, 191–285.

186. de Jeu, W. H. (2016) Basic X-Ray Scattering for Soft Matter. *Basic X-Ray Scatt. Soft Matter*. Oxford University Press, Oxford.

187. Feigin, L. A., and Svergun, D. I. (1987) Structure Analysis by Small-Angle X-

Ray and Neutron Scattering. *Struct. Anal. by Small-Angle X-Ray Neutron Scatt.* New York: Plenum Press.

188. Koch, M. H. J., Vachette, P., and Svergun, D. I. (2003) Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. *Q. Rev. Biophys. 36*, 147–227.

189. Lee, J., Xue, M., Wzorek, J. S., *et al.* (2016) Characterization of a stalled complex on the β-barrel assembly machine. *Proc. Natl. Acad. Sci. 113*, 8717–8722.

190. Costello, S. M., Plummer, A. M., Fleming, P. J., *et al.* (2016) Dynamic periplasmic chaperone reservoir facilitates biogenesis of outer membrane proteins. *Proc. Natl. Acad. Sci. U. S. A. 113*, E4794-800.

191. Blatch, G. L., and Lässle, M. (1999) The tetratricopeptide repeat: a structural motif mediating protein-protein interactions. *BioEssays 21*, 932–939.

192. Guinier, A. (1939) La diffraction des rayons X aux très petits angles : application à l'étude de phénomènes ultramicroscopiques. *Ann. Phys. (Paris). 11*, 161–237.

193. Durand, D., Vivès, C., Cannella, D., *et al.* (2010) NADPH oxidase activator p67(phox) behaves in solution as a multidomain protein with semi-flexible linkers. *J. Struct. Biol. 169*, 45–53.

194. Bernadó, P., and Svergun, D. I. (2012) Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering. *Mol. BioSyst. 8*, 151–167.

195. Glatter, O., and Kratky, O. (1982) Small-Angle X-ray Scattering. London: Academic press.

196. Rambo, R. P., and Tainer, J. A. (2011) Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers 95*, 559–571.

197. Glatter, O. (1977) A new method for the evaluation of small-angle scattering data. *J. Appl. Crystallogr. 10*, 415–421.

198. Moore, P. B. (1980) Small-angle scattering. Information content and error analysis. *J. Appl. Crystallogr. 13*, 168–175.

199. Svergun, D. I. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr. 25*, 495–503.

200. Bernadó, P., Mylonas, E., Petoukhov, M. V., *et al.* (2007) Structural Characterization of Flexible Proteins Using Small-Angle X-ray Scattering. *J. Am. Chem. Soc. 129*, 5656–5664.

201. Tria, G., Mertens, H. D. T., Kachala, M., *et al.* (2015) Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCrJ 2*, 207–217.

202. Hunke, S., and Betton, J.-M. (2003) Temperature effect on inclusion body formation and stress response in the periplasm of *Escherichia coli*. *Mol. Microbiol. 50*, 1579–1589.

203. Krojer, T., Garrido-Franco, M., Huber, R., *et al.* (2002) Crystal structure of DegP (HtrA) reveals a new protease-chaperone machine. *Nature 416*, 455–459.

204. Lee, J., Sutterlin, H. A., Wzorek, J. S., *et al.* (2018) Substrate binding to BamD triggers a conformational change in BamA to control membrane insertion. *Proc. Natl. Acad. Sci. 115*, 2359–2364.

205. Blaszczyk, M., Ciemny, M. P., Kolinski, A., *et al.* (2019) Protein–peptide docking using CABS-dock and contact information. *Brief. Bioinform. 20*, 2299–2305.

206. Pierce, B. G., Hourai, Y., and Weng, Z. (2011) Accelerating protein docking in ZDOCK using an advanced 3D convolution library. *PLoS One* (Keskin, O., Ed.) *6*, e24657.

207. Cheng, T. M.-K., Blundell, T. L., and Fernandez-Recio, J. (2007) pyDock: Electrostatics and desolvation for effective scoring of rigid-body protein-protein docking. *Proteins Struct. Funct. Bioinforma. 68*, 503–515.

208. Desta, I. T., Porter, K. A., Xia, B., *et al.* (2020) Performance and Its Limits in Rigid Body Protein-Protein Docking. *Structure 28*, 1071-1081.e3.

209. Gelpi, J., Hospital, A., Goñi, R., *et al.* (2015) Molecular dynamics simulations: advances and applications. *Adv. Appl. Bioinforma. Chem.* 37.

210. Hollingsworth, S. A., and Dror, R. O. (2018) Molecular Dynamics Simulation for All. *Neuron 99*, 1129–1143.

211. Kontogeorgis, G. M., Maribo-Mogensen, B., and Thomsen, K. (2018) The Debye-Hückel theory and its importance in modeling electrolyte solutions. *Fluid Phase Equilib. 462*, 130–152.

212. Phillips, J. C., Braun, R., Wang, W., *et al.* (2005) Scalable molecular dynamics with NAMD. *J. Comput. Chem. 26*, 1781–1802.

213. Humphrey, W., Dalke, A., and Schulten, K. (1996) VMD: Visual molecular dynamics. *J. Mol. Graph. 14*, 33–38.

214. Salomon-Ferrer, R., Götz, A. W., Poole, D., *et al.* (2013) Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput. 9*, 3878–3888.

215. Manzi, L., Barrow, A. S., Scott, D., *et al.* (2016) Carbene footprinting accurately maps binding sites in protein–ligand and protein–protein interactions. *Nat. Commun. 7*, 13288.

216. Elliott, T. S., Townsley, F. M., Bianco, A., *et al.* (2014) Proteome labeling and protein identification in specific tissues and at specific developmental stages in an animal. *Nat. Biotechnol. 32*, 465–472.

217. Kühlbrandt, W. (2014) The Resolution Revolution. *Science (80-.). 343*, 1443–1444.

218. Scheres, S. H. W. (2012) A Bayesian View on Cryo-EM Structure Determination. *J. Mol. Biol. 415*, 406–418.

219. Punjani, A., Rubinstein, J. L., Fleet, D. J., *et al.* (2017) cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat. Methods 14*, 290–296.

220. Hauer, F., Gerle, C., Fischer, N., *et al.* (2015) GraDeR: Membrane Protein Complex Preparation for Single-Particle Cryo-EM. *Structure 23*, 1769–1775.

221. Wu, S., Avila-Sakar, A., Kim, J., *et al.* (2012) Fabs Enable Single Particle cryoEM Studies of Small Proteins. *Structure 20*, 582–592.

222. Bloch, J. S., Mukherjee, S., Kowal, J., *et al.* (2021) Development of a universal

nanobody-binding Fab module for fiducial-assisted cryo-EM studies of membrane proteins. *Proc. Natl. Acad. Sci. 118*.