How do bacteria glide?

A structural perspective into the gliding motility of

Bdellovibrio bacteriovorus

by

LIAM RICHARD MEAD

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

DOCTOR OF PHILOSOPHY



Institute of Microbiology and Infection
School of Biosciences
College of Life and Environmental Sciences
University of Birmingham
December 2023

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties.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

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

Abstract

Bdellovibrio bacteriovorus is a bacterial predator that invades and kills Gram-negative bacteria. *B. bacteriovorus's* ability to kill pathogenic strains highlights the possibility of its use as a therapeutic and biocontrol agent. Motility is key for efficient predation; *B. bacteriovorus* uses gliding motility to traverse biofilms in search of prey and to escape the exhausted prey cell. Gliding is controlled by cGAMP signalling which regulates the formation of an envelope-spanning gliding machinery called the Agl-Glt complex. Whilst genetic, microscopy and interaction studies have provided initial insights into the assembly of the gliding machinery, the mechanisms by which the proteins of the Agl-Glt complex interact to facilitate gliding remain to be uncovered.

This work primarily focusses on the structural characterisation of the components of the Agl-Glt complex. The primary targets for this study are GltD1 and GltE1, periplasmic components of the Agl-Glt complex proposed to transduce force across the cell envelope. We demonstrate that GltD1 and GltE1 interact with nanomolar affinity and present the crystal structures of GltE1 and the complex of GltD1:GltE1. The GltD1:GltE1 complex provides the first structural insights into the connectivity within the Agl-Glt complex and provides information on orientation of the Agl-Glt complex in the cell envelope. A secondary focus of this study was on Bd1996 and MglA, proteins associated with the regulation of gliding motility. We experimentally confirmed that the PilZ domain of Bd1996 binds to both c-di-GMP and cGAMP, which represents the first case of a PilZ domain with dual affinity for both second messengers. We also present a combined experimental and computational characterisation of the ras-like GTPase MglA and reveal that MglA undergoes GTP dependent conformational changes to interact with binding partners. Finally, we present the crystal structure of Bd1431, a nuclease belonging to the *staphylococcal* nuclease family, that is upregulated during invasion. Bd1431 contains a novel dimeric fold previously unseen in the staphylococcal nuclease family.

Acknowledgements

Firstly, I would like to thank Professor Andy Lovering. You have taught me so much over the years and your support and guidance have been invaluable in helping me develop as a scientist. I truly appreciate all the opportunities you have provided me throughout my scientific journey. I sincerely believe I am unlikely to ever have a better mentor and I am profoundly grateful.

Secondly, I would like to thank Dr Simon Caulton. Since day one, you have always pushed me to improve and I truly value your guidance and friendship. Though I still don't believe you could beat an elephant in a fight, I do thank you for teaching me the most important lesson… Don't mess with Poitín.

To Rebecca, Amber, Mauricio, Hannah, Matt and Stella, thank you for being the best lab mates. You have been there through the turbulent ups and downs and I couldn't have asked for a better group to go through it all with. To the rest of the 7th floor, thank you for helping to create the best environment someone could hope to complete a PhD in and most importantly, thank you for humouring my countless, pointless hypothetical scenarios. To Matt and Cat, thank you for being there since day one of our PhDs and for being (a still scientific) escape from structural biology.

Finally, I thank my family. Mom and Dad, thank you for your constant, unwavering support and encouragement that have got me to the point of submitting this thesis. Grace, I can't put into words how much you have helped me over the last four years. Without you, I certainly would never have completed this PhD. Thank you.

Table of Contents

Chapter 1 1

Introduction

1.1 Predatory Bacteria 1
1.1.1 Group Attack 1
1.1.2 Engulfment 2
1.1.3 Epibiotic Predation 2
1.1.4 Endobiotic Predation 3

1.2 The lifecycle of *Bdellovibrio bacteriovorus* 5

1.3 The use of *B. bacteriovorus* as therapeutic agents 7

1.4 Bacterial Motility 10
1.4.1 Flagellar swimming 12
1.4.2 TFP Twitching 13
1.4.3 Gliding Motility 15

1.5 Gliding in *M. xanthus* 16
1.5.1 The regulation of *M. Xanthus* motility 17
1.5.2 The identification of the gliding machinery 19
1.5.3 The Agl-Glt complex 21
1.5.3.1 AglRQS 22
1.5.3.2 Cytoplasmic Glt Components 23
1.5.3.3 IM Glt components 24
1.5.3.4 Periplasmic Glt components 26
1.5.3.5 OM Glt components 29
1.5.4 Proposed models of gliding motility 30
1.5.4.1 The slime secretion model 30
1.5.4.2 Focal Adhesion Model 32
1.5.4.3 The Helical Rotor Model 34
1.5.4.4 FA vs HR models 35
1.5.5 The current model of *M. xanthus* gliding 37
1.5.6 Gliding in *B. bacteriovorus* 40

1.6 An introduction to c-di-GMP signalling 44
1.6.1 C-di-GMP signalling in *B. bacteriovorus* 47

1.7 Thesis Overview 52

Chapter 2 53

Materials and Methods

2.1 General methods 53
2.1.1 Molecular Cloning 53
2.1.1.1 Restriction cloning 53
2.1.1.2 Restriction-free cloning 55
2.1.1.3 Site-directed mutagenesis 56
2.1.1.4 Transformation 57
2.1.1.5 Colony PCR 57

2.1.1.6 Sequencing 58
2.1.1.7 Gene synthesis 58
2.1.2 Cell culture 58
2.1.2.1 Culture Media 58
2.1.2.2 Antibiotic Stocks 59
2.1.2.3 Expression Trials 59
2.1.2.4 SDS-PAGE 60
2.1.2.5 Largescale expression protocol 61
2.1.3 Protein Purification 62
2.1.3.1 Immobilised metal affinity chromatography 62
2.1.3.2 Size exclusion chromatography 63
2.1.4 Crystallisation and diffraction experiments 63
2.1.5 AlphaFold modelling 64
2.2 Methods for Chapter 3 (Structural characterisation of the *B. bacteriovorus* gliding machinery) 64
2.2.1 Cloning, expression and purification 64
2.2.2 GltD1:GltE1 ITC 67
2.2.3 GltD1:GltE1 Circular Dichroism 67
2.2.4 GltD1:GltE1 Crystallisation and data collection 68
2.2.5 GltD1:GltE1 Cryo-electron microscopy 69
2.2.6 Purification of MBP-GltG1$_{491–556}$ 69
2.2.7 GltC1 inclusion body and refold 70
2.3 Methods for Chapter 4 (Structural characterisation of proteins associated with the regulation of gliding) 70
2.3.1 Cloning, expression and purification 70
2.3.2 Bd1996 and MglA$_{Bd:}$Bd2492 Crystallisation and data collection 72
2.3.3 Bd1996 Pull-down assays 73
2.3.4 Bd1996:c-di-GMP/cGAMP ITC 75
2.3.5 Phage display 76
2.4 Methods for Chapter 5 (Structural characterisation of the nuclease Bd1431) 77
2.4.1 Cloning, expression and purification 77
2.4.2 Bd1431 Crystallisation and data collection 77
2.4.4 Bd1431 activity assays 80
Chapter 3 81
Structural characterisation of the *B. bacteriovorus* gliding machinery
3.1 Introduction 81
3.1.1 Structural characterisation of GltD 81
3.1.2 Chapter Aims 89
3.2 Results 92
3.2.1 AF modelling of GtlD1-GltE1 92
3.2.2 Expression and purification of GltD1 and GltE1 93
3.2.3 ITC analysis of GltD1-GltE1 interaction 99
3.2.4 Crystallisation of GltE1 106
3.2.4.1 Crystal Form 1 Structure Solution 108
3.2.4.2 The overall fold of GltE1 (Crystal Form 1) 108

3.2.4.3 Putative ligand binding pockets 110
3.2.4.4 Crystal Form 2 Structure Solution 114
3.2.4.5 The overall fold of GltE1 (Crystal Form 2) 115
3.2.4.6 Ligand binding 115
3.2.4.7 GltE1 Crystal Form 2 exhibits a pseudodomain swap 119
3.2.5 Crystallisation of GltD1:GltE1 complex 120
3.2.5.1 The overall fold of $GltD1_{Cterm}$:GltE1 128
3.2.5.2 $GltD1_{Cterm}$:GltE1 Interface 129
3.2.6 Further insights from AF modelling 136
3.2.6.1 The OM Glt complex 137
3.2.6.2 Periplasmic Glt complex 140
3.2.7 Experimental validation of GltD1:GltE1:GltG1 interaction 145
3.2.8 Expressing GltC1 149

3.3 Discussion 153
3.3.1 GltD1:GltE1 interaction – connectivity and orientation of the Agl-Glt complex 153
3.3.2 Similarity between the *B. bacteriovorus* GltD paralogs 155
3.3.3 Insights from AF modelling 159
3.3.4 Proposed mechanism of gliding 162
3.3.5 Conclusions and Future Work 167

Chapter 4 169

Structural characterisation of proteins associated with the regulation of gliding

4.1 Introduction 169
4.1.1 Hybrid GYF:PilZ proteins may regulate gliding in *B. bacteriovorus* 169
4.1.2 The role of MglA in *M. xanthus* gliding motility 175
4.1.3 The role of MglA in *B. bacteriovorus* 182
4.1.4 Chapter Aims 184

4.2 Results 185
4.2.1 Expression and purification of Bd1996 constructs 185
4.2.2 Crystallisation and structure solution of Bd1996-PilZ 190
4.2.3 Bd1996-PilZ Overall Fold 192
4.2.4 Further attempts to crystallise full-length Bd1996 193
4.2.5 AF insights into Bd1996 dimerisation 197
4.2.6 Validation of Bd1996 self-inhibition 202
4.2.7 Bd1996 pull-down assays 206
4.2.8 Experimental Validation of Bd1996:DgcA interaction 212
4.2.9 Bd1996 Phage Display 215
4.2.10 AF model of $MglA_{Bd}$:Bd2492 complex 217
4.2.11 Purification of $MglA_{Bd}$:$Bd2492_{205–353}$ 218
4.2.12 Crystallisation of $MglA_{Bd}$:$Bd2492_{205–353}$ 220
4.2.13 $MglA_{Bd}$:$Bd2492_{205–353}$ structure solution 221
4.2.14 Structural analysis of $MglA_{Bd}$ 223

4.3 Discussion 228
4.3.1 ApoBd1996 is monomeric but may exhibit self-inhibition 228
4.3.2 Pull-down assays identify known gliding proteins as Bd1996 binding partners 229

4.2.3 $MglA_{Bd}$ appears to undergo GTP-dependent conformational changes to interact with Bd2492 232
4.2.4 Conclusions and Future studies 236

Chapter 5 239

Structural characterisation of the nuclease Bd1431

5.1 Introduction 239
5.1.1 The importance of secreted enzymes for predation. 239
5.1.2 Importance of nucleases 239
5.1.3 An introduction to Bd1431 243
5.1.4 Introduction to SNase 243
5.1.5 Structure and mechanism of SNase 244
5.1.6 Chapter Aims 246

5.2 Results 247
5.2.1 Expression and purification of Bd1431 247
5.2.2 Crystallisation 249
5.2.3 Bd1431 Structure solution 250
5.2.3.1 Bd1431 Overall Fold and comparison to conventional SNase fold 253
5.2.3.2 Active Site 256
5.2.3.3 Dimer interface 258
5.2.4 Validation of the biological relevance of dimerisation 262
5.2.5 Crystal structure of $Bd1431_{F120D}$ 264
5.2.6 Crystal structure of Bd1431-TDP complex 270
5.2.7 Bd1431 Activity assay 275

5.3 Discussion 277
5.3.1 Bd1431 exhibits a novel dimeric SNase fold 277
5.3.2 The insertion of hydrophobic amino acids to Loop 2–3 promotes dimerisation 278
5.3.3 Bd1431 is a catalytically active nuclease 283
5.3.4 Potential role for Bd1431 in *B. bacteriovorus* 284
5.3.5 Future Studies 284

List of Figures

Figure 1.1: Mechanisms of bacterial predation 4
Figure 1.2: Overview of *B. bacteriovorus's* predatory lifecycle 7
Figure 1.3: Overview of *B. bacteriovorus* motility 11
Figure 1.4: Schematic of diagram of a flagellum 13
Figure 1.5: Cartoon diagram of a Gram-negative TFP 15
Figure 1.6: The roles of *B. bacteriovorus's* gliding motility 16
Figure 1.7: Assembly of the Agl-Glt complex based on cell fractionation studies and bioinformatic analysis 22
Figure 1.8: TonB-dependent transport 25
Figure 1.9: Lipoprotein synthesis 28
Figure 1.10: The focal adhesion model of gliding 34
Figure 1.11: The helical model of gliding 35
Figure 1.12: Assembly of complete gliding machineries from pre-formed subcomplexes 39
Figure 1.13: Assembly of the *B. bacteriovorus* Agl-Glt complex and overview of the four *B. bacteriovorus* gliding operons 41
Figure 1.14: Chemical structure of c-di-GMP 45
Figure 1.15: PilZ:c-di-GMP interaction 47
Figure 1.16: Chemical structure of cGAMP 49
Figure 1.17: Sequence alignment of the GGDEF domains of *B. bacteriovorus's* active DGCs 50
Figure 1.18: DgcA can produce both c-di-GMP and cGAMP 51

Figure 3.1: Crystal structure of full-length GltD4 83
Figure 3.2: Crystal packing in GltD structures 84
Figure 3.3: GltD3 cryo-EM structures 86
Figure 3.4: TPR ligand binding 87
Figure 3.5: Schematic diagram of *B. bacteriovorus's* four GltD genes 88
Figure 3.6: Multiple sequence alignment of the four *B. bacteriovorus* GltD genes 91
Figure 3.7: AF models of GltD1-GltE1 complex 93
Figure 3.8: Purification of GltD1 95
Figure 3.9: Purification of GltE1 97
Figure 3.10: GltD1:GltE1 SEC interaction assay 98
Figure 3.11: ITC binding for GltD1-GltE1 interaction 102
Figure 3.12: Location of $GltD1_{A762D}$ and $GltE1_{N200R}$ mutants on GltD1:GltE1 AF model 103
Figure 3.13: CD spectra 105
Figure 3.14: GltE1 crystals 106
Figure 3.15: The crystal structure of GltE1 Crystal Form 1 109
Figure 3.16: Sequence comparison of GltE1 TPR motifs and consensus TPR sequence 109
Figure 3.17: PDBSum analysis of GltE1 Crystal Form 1 110
Figure 3.18: GltE1 contains a hydrophobic groove that lines the concave face 111
Figure 3.19: The convex face of GltE1 contains a stretch of positive charge that could house ligand interactions 112

Figure 3.20: Crystal structure of GltE1 Crystal Form 2 116
Figure 3.21: Packing of ASU copies of GltE1 Crystal Form 2 117
Figure 3.22: Interaction between ASU copies of GltE1 Crystal Form 2 118
Figure 3.23: Crystal Form 2 pseudodomain swap 119
Figure 3.24: Micrograph of GltD1:GltE1 complex 120
Figure 3.25: Purification of $GltD1_{Cterm}$ 122
Figure 3.26: ITC binding for $GltD1_{Cterm}$:GltE1 interaction 123
Figure 3.27: GltD1:GltE1 co-purification 125
Figure 3.28: GltD1:GltE1 crystals 126
Figure 3.29: $GltD1_{Cterm}$:GltE1 crystal structure 128
Figure 3 30: Sequence comparison of $GltD1_{Cterm}$ TPR motifs and consensus TPR sequence 129
Figure 3.31: Hydrophobic interactions at the $GltD1_{Cterm}$-GltE1 interface 133
Figure 3.32: Further interactions at the $GltD1_{Cterm}$-GltE1 interface 135
Figure 3.33: AF model of OM and periplasmic components of the Glt1 operon 136
Figure 3.34: AF model of GltA1-GltB1-Bd0875 interaction 138
Figure 3.35: AF model of the GltA1-GltB1-GltC1 interaction 139
Figure 3.36: AF model of the GltC1-$GltD_{125–469}$-GltF1-$GltG1_{491–707}$ interaction 142
Figure 3.37: Predicted conformational change of the TonB domain of GltG1 upon GltF1 binding 143
Figure 3.38: AF model of the GltD1-GltE1-$GltG1_{491–707}$ interaction 143
Figure 3.39: Composite AF model of the OM-periplasmic components of the Glt1 operon 144
Figure 3.40: AF model of the region of the GltD1:GltE1:GltG1 complex targeted for experimental validation 145
Figure 3.41: Purification of MBP-$GltG1_{491–556}$ 147
Figure 3.42: ITC interactions for MBP-$GltG1_{491–556}$ 148
Figure 3.43: GltC1 Expression trial 150
Figure 3.44: GltC1 Buffer screen 150
Figure 3.45: IMAC purification of GltC1 lysed in buffer A containing 20 mM sodium cholate and 20% glycerol 151
Figure 3.46: $GltC1_{\Delta 480–529}$ expression trial 151
Figure 3.47: Refolding trial of GltC1 152
Figure 3.48: AF model of GltD_MX:GltE_MX 154
Figure 3.49: Superimposition of $GltD1_{Cterm}$-GltE1 crystal structure and GltD3 cryo-EM structures 156
Figure 3.50: Structural comparison of *B. bacteriovorus* GltD paralogs 158
Figure 3.51: Crystal structures of TonB:ton-box interactions 161
Figure 3.52: Possible arrangement of Agl-Glt complex arrays 165
Figure 3.53: Proposed mechanism of gliding 166

Figure 4.1: Ligand binding mechanism of GYF domains 170
Figure 4.2: GYF domain sequence alignment and domain architecture 172
Figure 4.3: Partial crystal structure of Bd1996 174
Figure 4.4: Proposed Bd1996 dimerisation prevents c-di-GMP/cGAMP binding 175
Figure 4.5: The conformational states of $MglA_{MX}$ 181
Figure 4.6: $MglA_{Bd}$ and $MglA_{MX}$ sequence alignment 183
Figure 4.7: Purification of Bd1996 186

Figure 4.8: IMAC Purification of Bd1996-PilZ 188
Figure 4.9: SEC Purification of Bd1996-PilZ 189
Figure 4.10: Bd1996-PilZ crystals 190
Figure 4.11: Overall fold of Bd1996-PilZ 192
Figure 4.12: Crystal packing of Bd1996-PilZ 193
Figure 4.13: Crystals of Bd1996 194
Figure 4.14: Crystal packing of Bd1996-PilZ Crystal Form 2 197
Figure 4.15: AF model of Bd1996 198
Figure 4.16: AF model of Bd1996 GYF domains 199
Figure 4.17: AF model of the Bd1996-Dimer 201
Figure 4.18: Bd1996 PilZ:GYF-I domain interaction 202
Figure 4.19: Purification of Bd1996-GYF 204
Figure 4.20: ITC binding for Bd1996-PilZ:Bd1996-GYF interaction 205
Figure 4.21: Bd1996 Pull-down assays 206
Figure 4.22: Bd1996:DgcA AF model 211
Figure 4.23: ITC binding for Bd1996 c-di-GMP interaction 213
Figure 4.24: Bd1996 cGAMP interaction 214
Figure 4.25: AF models of $MglA_{Bd}$:Bd2492 complex 218
Figure 4.26: Purification of $MglA_{Bd}$:$Bd2492_{205–353}$ 220
Figure 4.27: $MglA_{Bd}$:$Bd2492_{205–353}$ crystals 221
Figure 4.28: Crystal structure of $MglA_{Bd}$ 225
Figure 4.29: Electron density for bound nucleotide 226
Figure 4.30: AF model of Bd1482:AglR3 interaction 232
Figure 4.31: Superimposition of $MglA_{MX}$-GTPγS to $MglA_{MX}$:MglB and $MglA_{Bd}$:Bd2492 AF model 235

Figure 5.1: Structure of SNase fold 245
Figure 5.2: SNase active site 246
Figure 5.3: IMAC purification of Bd1431 248
Figure 5.4: SEC purification of Bd1431 249
Figure 5.5: Bd1431 crystals 250
Figure 5.6: Overall fold of Bd1431 solved to 2.1 Å 254
Figure 5.7: Each Bd1431 monomer contains two disulphide bonds 255
Figure 5.8: Bd1431 active site 256
Figure 5.9: Crystal packing disrupts calcium binding in Bd1431 chain B 257
Figure 5.10: Bd1431 dimer interface 259
Figure 5.11: Schematic of hydrophobic packing at the dimer interface 260
Figure 5.12: Schematic of hydrogen bonding at the dimer interface 261
Figure 5.13: Second dimer interface identified by PISA 262
Figure 5.14: Comparison of SEC chromatograms between Bd1431 and $Bd1431_{F120D}$ 263
Figure 5.15: $Bd1431_{F120D}$ dimer interface 268
Figure 5.16: Comparison of SEC chromatograms between $Bd1431_{F120D}$ in high and low salt buffers 270
Figure 5.17: TDP binding to Bd1431 dimer 274
Figure 5.18: Electron density for TDP 274
Figure 5.19: Hydrogen bonding at the ligand binding pocket 275
Figure 5.20: Bd1431 activity assay 276

Figure 5.21: Helix swapped dimer of $SNase_{\Delta 114-119}$ 278
Figure 5.22: $Bd1431_{\Delta 59-66}$ AF model 279
Figure 5.23: SNase dimer 282
Figure 5.24: Double stranded DNA aligned to TDP 283

List of Tables

Table 1.1:	Summary of studies demonstrating the potential of *B. bacteriovorus* as a therapeutic and biocontrol agent	9
Table 1.2:	Alternative gene names used components of the Agl-Glt	17
Table 1.3:	Agl-Glt complex encoding genes in *B. bacteriovorus*	42
Table 2.1:	Thermocycling conditions for PCR	53
Table 2.2:	Thermocycling conditions for second-round PCR	54
Table 2.3:	Thermocycling conditions for Q5 SDM	56
Table 2.4:	Thermocycling conditions for colony PCR	57
Table 2.5:	Recipes for making 8 SDS PAGE gels	60
Table 2.6:	Information for constructs used in Chapter 3	66
Table 2.7:	Parameters used for GltD1:GltE1 ITC experiments	67
Table 2.8:	Information for constructs used in Chapter 4	71
Table 2.9:	Parameters used for Bd1996:c-di-GMP ITC experiments	76
Table 2.10:	Parameters used for Bd1996:cGAMP ITC experiments	76
Table 2.11:	Information for constructs used in Chapter 5	77
Table 2.12:	Crystal screen used for Bd1431 streak seeding	78
Table 3.1:	Sequence similarity between *B. bacteriovorus* GltD paralogs	88
Table 3.2:	Data collection and refinement statistics for GltE1 Crystal Form 1	107
Table 3.3:	Data collection and refinement statistics for GltE1 Crystal Form 2	113
Table 3.4:	Data collection and refinement statistics for the GltD1:GltE1 complex	127
Table 3.5:	Hydrogen bonding across $GltD1_{Cterm}$:GltE1 interface	130
Table 3.6:	Salt bridges at the $GltD1_{Cterm}$:GltE1 interface	131
Table 4.1:	Data collection and refinement statistics for Bd1996-PilZ	191
Table 4 2:	Data collection and refinement statistics for Bd1996 PilZ Crystal Form 2	195
Table 4.3:	Top 50 interactors identified from pull-down assays using Bd1996 as bait	208
Table 4.4:	Top 50 interactors identified from pull-down assays using Bd1996-GYF as bait	209
Table 4.5:	Top 10 enriched peptide sequences following three rounds of panning	216
Table 4.6:	Data collection and refinement statistics for $MglA_{Bd}$	222
Table 5.1:	Summary of *B. bacteriovorus's* degradative enzymes	239
Table 5.2:	Data collection and refinement statistics for Bd1431	252
Table 5.3:	Data collection and refinement statistics for $Bd1431_{F120D}$	265
Table 5.4.	Data collection and refinement statistics for the Bd1431-TDP complex	272

List of Abbreviations

A. baumannii	*Acinetobacter baumannii*
AF	AlphaFold
Agl	adventurous gliding
AI	auto-induction
A-motility	adventurous motility
APS	ammonium persulfate
ASU	asymmetric unit
BALOs	*Bdellovibrio* and like organisms
BCS	basic chemical space
bp	base pair
CCCP	carbonyl cyanide-m-chlorophenylhydrazone
c-di-GMP	cyclic-di-GMP
Cgl	contact-dependent gliding
Che	chemotaxis
CIM	cytoplasmic/IM
CM	cytoplasmic membrane
DNA	deoxyribonucleic acid
E. coli	*Escherichia coli*
eDNA	extracellular DNA
EM	electron microscopy
ESKAPE	*Enterobacter, Klebsiella, Acinetobacter, Pseudomonas, Enteric*
FA	focal adhesion
FHA	forkhead-associated
GAP	guanine nucleotide hydrolysis activating protein
GEF	guanine exchange nucleotide factor
Glt	gliding transducer
GMP	Guanosine 3',5'-cyclic monophosphate
GTP	Guanosine-5'-triphosphate
HR	helical rotor
IM	inner membrane
IMAC	immobilised metal affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
K. pneumoniae	*Klebsiella pneumoniae*
LB	Luria Bertani
LGT	lipoprotein diacylglyceryl transferase
LLG	log-likelihood gain
LNT	lipoprotein NN-acyl transferase
LSP	lipoprotein signal peptidase
M. xanthus	*Myxococcus xanthus*
MASC	major spore coat
MCP	methyl-accepting chemotaxis protein
Mgl	mutual gliding
MIDAS	metal-ion-dependant adhesion site
MWCO	molecular weight cut-off
NETs	neutrophil extracellular traps
Nfs	needed for sporulation

OB-fold	oligonucleotide-binding fold
OD	optical density
OM	outer membrane
P. aeruginosa	*Pseudomonas aeruginosa*
P. tolaasi	*Pseudomonas tolaasi*
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDB	Protein Data Bank
PEG	polyethylene glycol
PG	peptidoglycan
PISA	Protein Interfaces, Surfaces and Assemblies
PMF	proton-motive force
POM	periplasmic/OM
qRT-PCR	real-time quantitative RT-PCR
RMSD	root mean square deviation
RNA	ribonucleic acid
RPM	revolutions per minute
RT-PCR	reverse transcription PCR
S. aureus	*Staphylococcus aureus*
SDS	sodium dodecylsulphate
SEC	size exclusion chromatography
S-motility	social motility
SNase	Staphylococcal nuclease
SPR	surface plasmon resonance
TB	terrific broth
TBDT	TonB-dependent transporter
TDP	Thymidine 3’5’ diphosphate
TEMED	tetramethylethylenediamine
TFP	type IV pili
TFZ	translation function Z-score
TIRF	total internal reflection fluorescence
TPR	tetratricopeptide repeat
UV	ultraviolet
VWA	von Willebrand adhesin

Chapter 1

Introduction

1.1 Predatory Bacteria

Predatory bacteria are naturally antibacterial strains that survive in complex ecosystems[1]. These micropredators can either be obligate or facultative predators; obligate predators need to consume prey cells to survive, whilst facultative predators can adopt both a predatory and a saprophytic lifestyle[2]. Following successful attack, bacterial predators utilise the prey cell's macromolecules for their own growth and division[1]. Predatory bacteria are characterised based on the mechanism of predation they utilise, the most common mechanisms include group attack, engulfment, epibiotic predation and endobiotic predation[3] (Figure 1.1).

1.1.1 Group Attack

Group attack is a predation mechanism utilised by *Myxococcus xanthus (M. xanthus)* and *Lysobacter* and refers to a scenario in which communities of predatory bacteria work together to kill prey cells[4,5]. This can either be achieved through direct predatory contact between members of the group and the prey cell or through the release of degradative enzymes by the group to lyse and decompose the prey which releases nutrients for the predatory cells[3,6]. The most well-characterised example of group attack predation can be seen with *M. xanthus*. *M. xanthus* cells form social predatory swarms that attack prey cells in a "wolfpack" fashion[5]. When *M. xanthus* swarms encounter prey cells, members of the swarm release an array of antibiotics and degradative enzymes, resulting in the death and lysis of the prey cell[7]. This releases

macromolecules from the prey cell and thus, the entire swarm can benefit from the spoils of predation[8].

1.1.2 Engulfment

A recently identified predatory mechanism involves the phagocytosis-like engulfment of prey cells by the predator. This process has been exclusively observed in the planctomycete bacterium, *Candidatus Uab amorphum*[9]*.* This mechanism of predation remains largely uncharacterised, although light microscopy and electron microscopy (EM) has uncovered the basic predatory process. Prey cells are first engulfed by surface invagination of the predator cell. The prey cell then becomes enclosed within a phagosome-like vacuole, where digestion of the prey cell is postulated to occur[9]. Despite the limited studies on *Candidatus Uab amorphum*, this newly uncovered predatory bacterium appears to have a broad prey range including both Gram-negative and Gram-positive bacteria, as well as small Eukaryotic cells[9]*.* The mechanism of action of both group attack predation and predatory engulfment results in relatively non-specific predation.

1.1.3 Epibiotic Predation

Epibiotic predation refers to an extracellular form of predation in which predatory cells attach to the exterior of a prey cell, secrete hydrolytic enzymes into the prey and extract the cell content, leading to the division of the predatory cell into two daughter cells through binary fission[10]. Members of the epibiotic predator group include *Bdellovibrio exovorus and Vampirococcus*[1]. Additionally, some members of the Candidate Phyla Radiation superphylum have been identified as ultra-small epibiotic predators[11,12].

A hybrid predatory approach also exists which combines features of epibiotic and group predation; this so-called 'picket fence' predation strategy is employed by *Ensifer adhaerens* and the Betaproteobacterium predatory strain 679–2 and involves small groups of predators attaching in unison to the prey outer membrane (OM)[3,13].

1.1.4 Endobiotic Predation

Endobiotic predation refers to the direct invasion of a predator into a prey cell[3]. Following attachment, endobiotic predators invade the prey cell and occupy the prey's cytoplasm (*e.g., Daptobacter*[14]) or periplasm (*e.g., Bdellovibrio*[15]), where the predatory lifecycle is carried out[8]. Following the utilisation of the prey's macromolecules for growth and division, progeny predator cells escape from the exhausted prey shell in search of new prey[15]. *Bdellovibrio* and like organisms (BALOs) are the most well-studied of the endobiotic predators. *Bdellovibrio bacteriovorus*, in particular, acts as a model organism for bacterial predation.

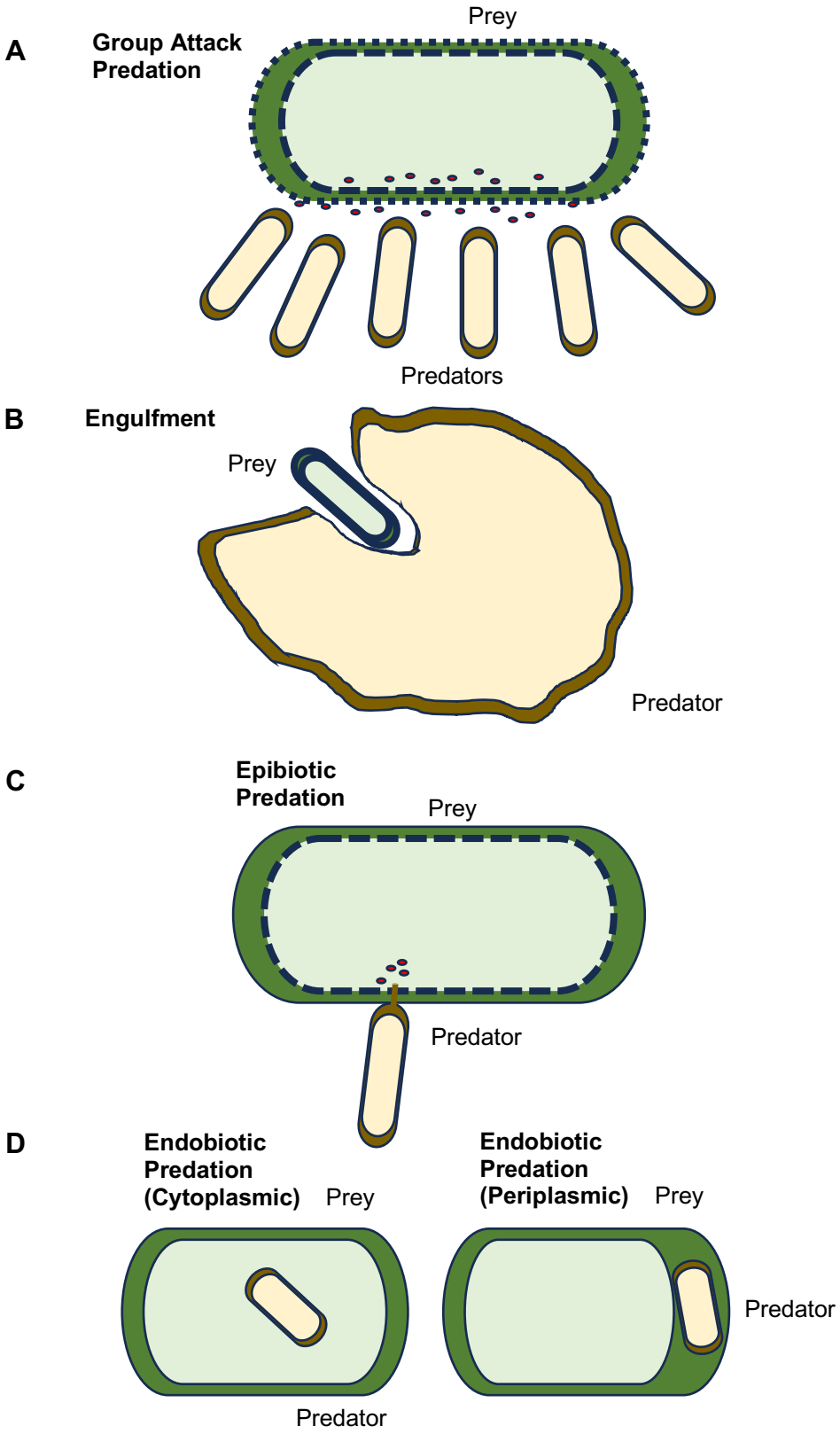


Figure 1.1: Mechanisms of bacterial predation.
Predatory bacteria can be classified into distinct groups based on their predation strategy. Common predation strategies include **(A)** group attack, **(B)** engulfment, **(C)** epibiotic predation and **(D)** endobiotic predation. Predators are coloured brown/beige and prey are coloured green. Red circles represent degradative enzymes released by predators to facilitate prey death.

1.2 The lifecycle of *Bdellovibrio bacteriovorus*

Bdellovibrio bacteriovorus is a small Gram-negative bacterium which inhabits a wide range of aquatic and terrestrial ecosystems[16]. In these environments, *B. bacteriovorus* is an obligate predator of other Gram-negative bacteria and carries out a multi-faceted endobiotic predatory lifecycle which sees *B. bacteriovorus* invade into the prey cells periplasm (Figure 1.2)[15]. Unless otherwise stated, *B. bacteriovorus* will refer to the strain *B. bacteriovorus HD100* throughout this thesis.

The predatory process of *B. bacteriovorus* consists of four main stages: the attack phase, invasion phase, growth phase and division and escape phase[17]. During the attack phase, highly motile *B. bacteriovorus* are (presumed to be) chemotactically driven towards areas of high prey density using either flagellar swimming or surface gliding motility[18]. Upon encountering a prey cell, a transient recognition event occurs, likely mediated through surface adhesins on *B. bacteriovorus's* OM[19]. If the attached cell is identified as suitable prey, then attachment becomes irreversible[15]. This irreversible attachment of *B. bacteriovorus* to a prey cell is mediated by type IV pili (TFP) and marks the start of predation[17].

In order to initiate invasion, *B. bacteriovorus* creates a small pore in the prey envelope, through which the invading cell passes through the prey's OM and peptidoglycan (PG) cell wall to reside within the periplasm[20]. The exact mechanism of entry remains elusive. Initial hypotheses suggested that TFP were responsible for entry, as the deletion of PilA (the major protein of the pilus filament) abolished predation[21,22].

However, *B. bacteriovorus* that express non-retractable pilus mutants are still able to invade prey[23].

Once inside the prey, peptidoglycan is deposited around the invasion pore[19,24]. The pore is subsequently resealed to prevent leakage of usable prey resources from the cell[15]. The prey cell's PG is then remodelled to form a spherical structure called the *Bdelloplast*[15]. The formation of the *Bdelloplast* creates an environment which is more resilient to osmotic pressure and also acts as a signal to prevent competition between multiple *B. bacteriovorus* for one prey cell[25]. Remarkably, a recent study has shown that *B. bacteriovorus* utilises lytic transglycosylases to cleave the septum of dividing prey cells, increasing the efficiency of predation by allowing access to the whole prey cell[26].

Following the formation of the *Bdelloplast*, the invading cell releases a cocktail of hydrolytic enzymes; including proteases, nucleases, glycanases and lipases to break the prey cell down[16,27]. The released macromolecules are used for bidirectional growth of the invading cell. Upon the exhaustion of prey resources, the growing filament septates and divides into multiple progeny cells, with the number of progeny being dependent upon the size of the prey cell[28]. The progeny cells then lyse the exhausted prey cell envelope and exit the cell in search of new prey. In total, the entire predatory process takes around 2–3 hours[15]. Although *B. bacteriovorus* are obligate predators, spontaneous host-independent (H-I) mutants can grow and divide outside of prey. These H-I strains can be grown in rich media and have been proven an invaluable tool during genetic studies to uncover genes that are essential for predation[15,29].

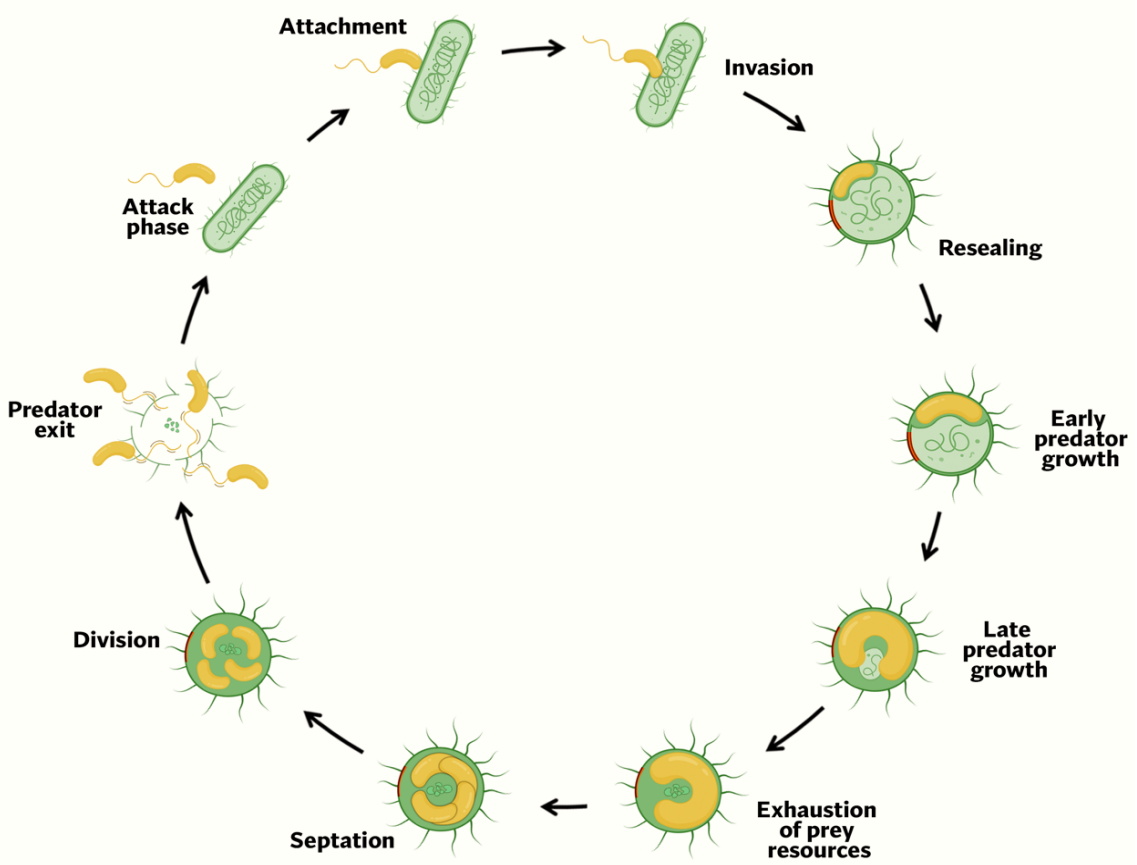


Figure 1.2: Overview of *B. bacteriovorus*'s predatory lifecycle
Schematic diagram of *B. bacteriovorus* (coloured yellow) predating upon a Gram-negative prey cell (coloured green). Predation begins when a highly motile *B. bacteriovorus* cell encounters a prey cell. Following a transient recognition event, *B. bacteriovorus* irreversibly attaches to the prey's OM. Following attachment, the *B. bacteriovorus* cell creates a small pore in the prey's OM, through which *B. bacteriovorus* invades into the prey's periplasm. During invasion, the prey cell wall is modified to form an osmotically stable spherical structure called the Bdelloplast. In the Bdelloplast, *B. bacteriovorus* releases a cocktail of hydrolytic enzymes to break the prey cell down from within, subsequently utilising the prey cell's nutrients for growth. Upon exhaustion of prey nutrients, the growing *B. bacteriovorus* cell septates and divides into progeny cells. These progeny cells lyse and exit the exhausted prey cell to begin the predatory cycle again.

1.3 The use of *B. bacteriovorus* as therapeutic agents

The ever-increasing prevalence of antibiotic-resistant bacteria has threatened a global health and economic crisis. The acquisition of antibiotic resistance is a natural and inevitable phenomenon[30]. However, the misuse of antibiotics within the healthcare, veterinary and agriculture sectors has led to a significant increase in the selective

pressure applied to bacteria. Overall, this has led to an increased occurrence of antibiotic resistance and the emergence of multi-drug resistant bacteria[31,32].

At present, multi-drug resistant bacteria account for at least 700,000 of annual deaths across the globe[1]. Studies suggest that the number of annual deaths caused by multi-drug resistant bacteria could rise to 10 million by 2050[33,34], highlighting the true scope of the antibiotic-resistant bacteria crisis. The problem is further compounded by the lack of new antibiotics being produced by the pharmaceutical industry[35]. The inevitable increase in the occurrence of multi-drug resistant bacteria highlights the need for alternative therapies to treat infections, otherwise, we face the risk of returning to a pre-antibiotic era in which hospitalisations and deaths from previously treatable infections are commonplace. As such, considerable effort is currently being applied into the development of alternatives to antibiotics, including but not limited to vaccines, phage therapy, CRISPR-Cas editors and the use of bacterial predators[36,37].

B. bacteriovorus's innate ability to kill a broad range of Gram-negative bacteria has earmarked its potential as a therapeutic agent for use in healthcare, veterinary and agriculture applications[38]. *B. bacteriovorus's* therapeutic capability is highlighted by several key points. Mainly, *B. bacteriovorus* can kill a large number of pathogenic Gram-negative bacteria, including members of the multi-drug resistant ESKAPE pathogens which have been highlighted as a critical priority by the World Health Organisation[17,39]. *B. bacteriovorus* has also been observed to disrupt and prevent the formation of biofilms[40], emphasising its utility within both a clinical setting and as a biocontrol agent. Moreover, there have been no reported cases of prey developing

resistance to predation by *B. bacteriovorus*[1]. *B. bacteriovorus's* multi-faceted predatory process utilises a vast array of highly specialised degradative enzymes, making it difficult for prey to evolve conventional resistance mechanisms[41]. A list of studies demonstrating *B. bacteriovorus's* therapeutic potential is provided in Table 1.1.

Table 1.1: Summary of studies demonstrating the potential of *B. bacteriovorus* as a therapeutic and biocontrol agent.

Findings	Ref
B. bacteriovorus can kill *Pseudomonas aeruginosa* and *Serratia marcescens* (responsible for causing ocular infections).	42
B. bacteriovorus can predate and kill *Fusobacterium nucleatum* and *Aggregatibacter actinomycetemcomitans* (responsible for causing periodontal infections).	43
B. bacteriovorus and Bacteriophage can work synergistically to kill *E. coli*.	44
B. bacteriovorus and *Violacein* (as a dual treatment) can kill polymicrobial pathogenic communities.	45
B. bacteriovorus can prevent the formation of *E. coli, K. pneumoniae, P. aeruginosa* and *A. baumannii* biofilms and have some effect on dispersing already established biofilms.	40
B. bacteriovorus can kill *Yersinia pestis* (responsible for lung infection in mice) and protect mice from a lethal plague challenge.	46,47
B. bacteriovorus pre-exposure can have prophylactic activity against adherent-invasive *E. coli* in a *Galleria mellonella* moth model of infection.	48
B. bacteriovorus can prey on *K. pneumoniae* (an antibiotic-resistant human-pathogen) and reduce bacterial burden in rat lungs.	49,50
B. bacteriovorus can kill *shigella* in Zebrafish Larvae infection model.	51
B. bacteriovorus can reduce *E. coli* O157:H7 and Salmonella in lettuce and carrots and can be used to prevent post-harvest spoilage.	52
B. bacteriovorus can prey upon and kill *P. tolaasi* on mushroom surfaces and can be used to prevent post-harvest spoilage.	53
B. bacteriovorus (intranasal or intravenous) inoculation led to no reduction in mouse viability.	54

1.4 Bacterial Motility

Motility is key to bacterial life and plays essential roles in resource acquisition, colonisation, pathogenesis, biofilm formation and progeny dispersal[55]. Bacteria have evolved several motility mechanisms to facilitate their lifecycles. These mechanisms can be classified into those which enable cells to move freely through aqueous solutions or those which enable movement across solid surfaces, with each type of motility being driven by independent mechanisms, although the regulation of these systems can overlap[55,56].

B. bacteriovorus's lifecycle requires the predator to be highly motile, with motility playing a key role in prey encounter, prey entry and progeny escape[15]. *B. bacteriovorus* utilises combinations of flagellar swimming[57], TFP retraction[21] and gliding motility[58], with the motility systems being used at any time being dependent on the stage of the lifecycle and/or the local environment of the *B. bacteriovorus* cell[59] (Figure 1.3).

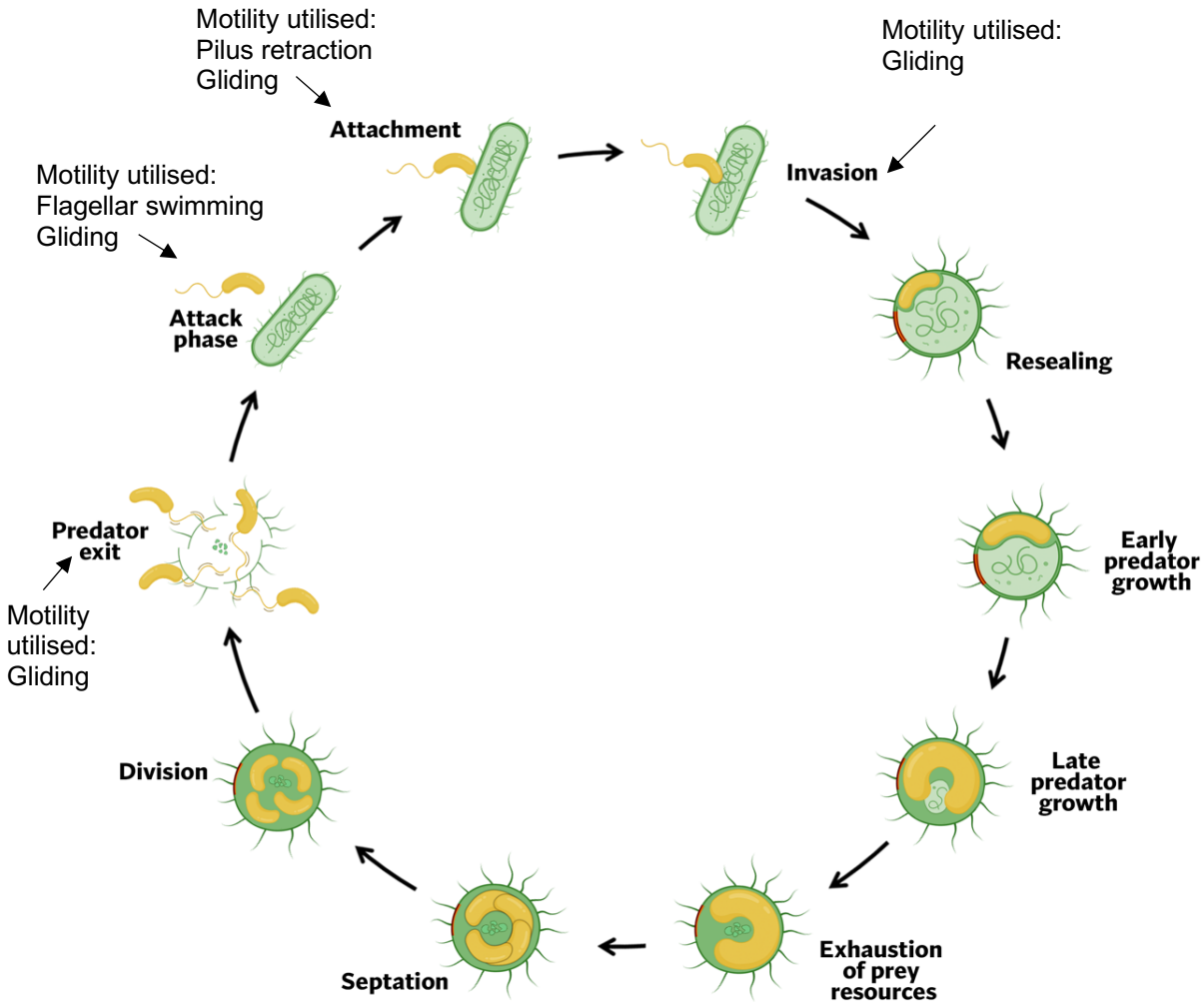


Figure 1.3: Overview of *B. bacteriovorus* motility
Schematic diagram depicting the type of motility utilised at different stages of *B. bacteriovorus's* lifecycle. *B. bacteriovorus* utilises combinations of flagellar swimming, TFP retraction and gliding motility to facilitate predation. Flagellar swimming plays a key role during the attack phase, allowing *B. bacteriovorus* cells to encounter prey in liquid environments[57]. TFP are responsible for the irreversible attachment of *B. bacteriovorus* to the prey's OM[21-23]. Gliding motility is essential for predation and allows *B. bacteriovorus* cells to navigate solid surfaces[58]. *B. bacteriovorus* cells use gliding motility during the attack phase to encounter prey on biofilms[58]. Gliding motility is also used to scout the surface of the prey cell for a suitable entry/attachment site[58]. Finally, gliding motility is utilised by progeny *B. bacteriovorus* cells to escape the exhausted prey cell[58].

1.4.1 Flagellar swimming

Flagellar swimming is governed by multiprotein complexes called flagellum. The flagellum comprises around 30 different proteins which form four main structural units: the basal body, the rod, the hook and the filament[55,60,61] (Figure 1.4). The basal body spans the cell envelope and is composed of a rotor and stators[61]. The stator units surround the rotor and consist of MotA and MotB which are embedded in the inner membrane (IM)[61]. Together, MotA and MotB form H^+ channels, in which a MotA pentamer forms around a MotB dimer[62]. The passage of H^+ through the channel transduces a proton-motive force (PMF) that causes rotation of the MotA pentamer around MotB resulting in the rotation of the motor[60]. This rotation is then transmitted through the rod to the hook, which ultimately results in the rotation of the flagellar filament, which drives the cell forward[63].

B. bacteriovorus contains a single sheathed polar flagellum that permits rapid swimming motility, allowing the cell to reach swimming speeds up to 160 μms^{-1}[57,64]. This rapid swimming is essential for prey encounter, as non-flagellate mutants are observed to be non-predatory in liquid environments[57]. However, predation is still observed when non-flagellate mutants are pipetted directly onto prey cells[57], revealing that whilst the flagellum is essential for prey encounter, it is not essential for invasion. Until recently, *B. bacteriovorus* was assumed to shed its flagellum prior to invasion, however, electron tomography has revealed that *B. bacteriovorus* retracts and presumably recycles its flagella upon prey entry[65].

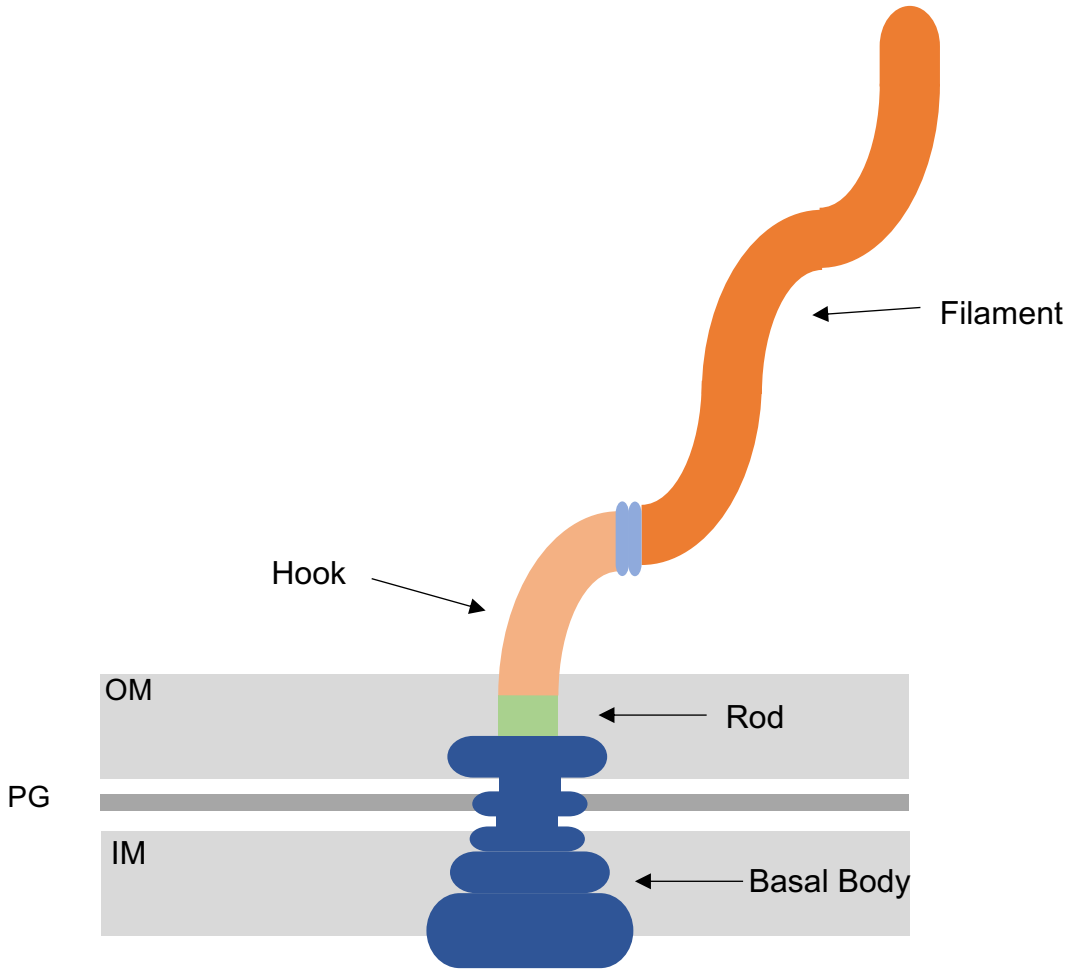


Figure 1.4: Schematic of diagram of a flagellum
The flagellum consists of four parts: the basal body, rod, hook and filament. The basal body comprises a MotAB PMF driven motor. Passage of protons results in rotation of the motor which is transferred through the rod and hook, ultimately leading to rotation of the filament which drives propulsion of the cell.

1.4.2 TFP Twitching

Whilst flagellar rotation enables cells to swim through liquid environments, many bacteria have evolved mechanisms to move across solid surfaces. The most common form of surface motility is TFP-mediated twitching[55]. TFP play diverse roles in adhesion, virulence, biofilm formation, predation, chemotaxis and DNA uptake[55,66]. The TFP machinery is embedded throughout the cell envelope and forms a series of

structural rings through which the pilus can pass (Figure 1.5)[67]. The pilus itself is a filament composed of PilA monomers[67]. Pilus formation is controlled by PilB (the assembly ATPase) which regulates the assembly of thousands of PilA monomers into a helical filament[66]. The elongating pilus passes through a pore in the OM and adheres to a target site. Once attached, the pilus is then retracted by PilT (the retraction ATPase). This retraction pulls the cell towards the point of adhesion in a fashion similar to that of a grappling hook[66]. The TFP of *B. bacteriovorus* are located on the invasive pole of the cell[21]. Whilst TFP are essential for predation[21], they appear to be involved in prey recognition and/or attachment rather than prey invasion as originally hypothesised[23].

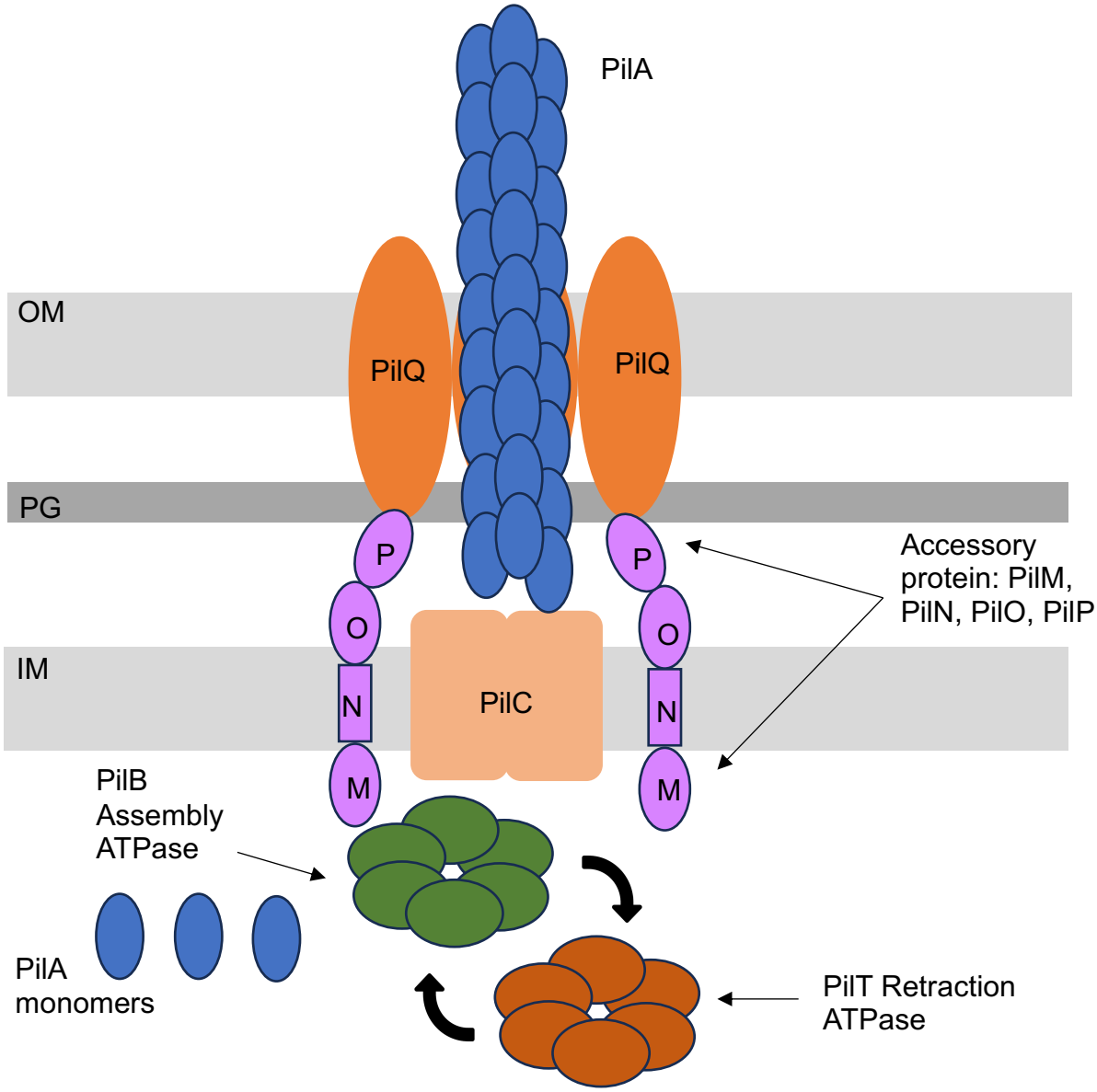


Figure 1.5: Cartoon diagram of a Gram-negative TFP
Cytoplasmic PilA monomers are assembled into a functional filament through the action of PilB[67] The extending filament passes through an OM ring formed by PilQ to attach to a target site[67]. Retraction of the pilus is mediated by the PilT which withdraws the pilus back into the IM and pulls the bacteria towards the pilus binding site[67]. Accessory proteins PilM-P help guide the filament through the periplasm[67].

1.4.3 Gliding Motility

Gliding is the process by which cells navigate solid surfaces without the use of external appendages[58]. Gliding was identified in *B. bacteriovorus* when mutants defective in flagellar swimming maintained slow movement (~16μm/h) across biofilms[58]. Surface

gliding enables *B. bacteriovorus* to traverse biofilms, scout prey surfaces for an entry point and to escape exhausted prey shells following predation (Figure 1.6)[58]. The mechanisms that govern gliding in *B. bacteriovorus* remain to be fully characterised and much of our current understanding is extrapolated from genetic similarities to the gliding motility of *M. xanthus*.

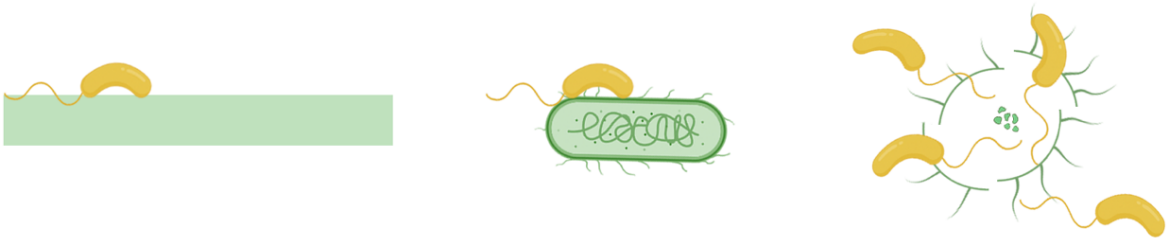


Figure 1.6: The roles of *B. bacteriovorus* gliding motility
B. bacteriovorus cells utilise gliding motility to traverse solid surfaces. Gliding motility is important for: **(1)** prey encounter in biofilms, **(2)** scouting the prey surface to identify a suitable entry site, **(3)** allowing progeny cells to escape the exhausted prey cell (as flagellar may tangle in the close confines of the prey cell).

1.5 Gliding in *M. xanthus*

M. xanthus lack flagella but movement across surfaces is essential to facilitate the previously described group attack predation[68]. Initial research into *M. xanthus* surface motility began following the identification of a semi-motile strain of *M. xanthus*[69]. This strain was observed to be non-motile when alone, but motile when the cells were apposed within a swarm[69]. In subsequent years, extensive mutagenesis work on motility mutants revealed the existence of two genetically distinct motility systems[68,70]. These systems were classified as social- (S) and adventurous- (A) motility. S-motility involves swarms of cells moving together, in a process driven by TFP[71]. In contrast, A-motility refers to the movement of individual cells away from the main swarm to scout out nutrient-rich environments[71]. A-motility is a form of 'true' gliding motility, wherein

cells move across surfaces without the aid of surface appendages. From here onwards, A-motility will be referred to exclusively as gliding motility.

1.5.1 The regulation of *M. Xanthus* motility

To date, 51 genes have been implicated in the regulation of gliding motility through transposon screening[72]. Over the decades of research into *M. xanthus* gliding, multiple names have been used to describe the same gliding genes (Table 1.2). For clarity, all the gene names used in this thesis will be the reclassified names first denoted in the study by Luciano *et al.*, 2011[72], irrespective of the name of the gene used at the time of the studies cited.

Table 1.2: Alternative gene names used components of the Agl-Glt complex

Gene name	Alternative name(s)	Protein Length (AA)	Domains
GltA	-	256	OM β-barrel
GltB	-	275	OM β-barrel
GltC	AgnA	673	TPR
GltD	AgmU	1218	TPR
GltE	AglT	478	TPR
GltF	CglF	89	Disordered
GltG	PglI	640	TonB
GltH	CglE/AgmV	209	OM β-barrel
GltI	AgmK	3822	TPR
GltJ	AgmX	674	TonB, ZnR, GYF
GltK	AgmO/CglC	184	Unknown
AglR	-	245	MotA/ExbB
AglQ	-	162	MotB/ExbD
AglS	-	194	MotB/ExbD
AglZ	-	1395	Coiled-coil
Prefaces – Glt = Gliding Transducer, Agl/Agm/Agn = A motility, Cgl = contact dependent gliding, Pgl = Partial gliding			

Both S-motility and gliding are under the regulation of the Frz chemosensory system[73]. The Frz system controls cell reversals, which ultimately governs the direction of movement. A cell reversal causes a 180° rotation in the direction of movement which results from an inversion of cell polarity[74]. Following cell reversal, the previous lagging pole becomes the new leading pole[74]. The Frz system comprises homologs of conventional bacterial chemotaxis (Che) pathway proteins: FrzA/B (CheW homolog), FrzE (CheA/Y Homolog), FrzF (CheR Homolg), FrzG (CheB Homolog), FrzX/Z (CheY Homolg) and FrzCD[74,75]. In brief, signalling through the Frz system begins when a ligand binds FrzCD, a cytoplasmic Methyl-accepting Chemotaxis Protein (MCP) receptor[74,76]. The activation of FrzCD results in the autophosphorylation of the histidine kinase FrzE which, in turn, transfers a phosphoryl group to FrzX[74–79]. FrzX then regulates downstream polarity switching through its interactions with MglA & MglB[80]. The modulation of cell reversals through Frz signalling creates directed movements, where the binding of an 'attractant' to FrzCD decreases the frequency of reversals, allowing movement to continue in a given direction for longer[74,81]. Conversely, the binding of a 'repellent' to FrzCD increases the frequency of cell reversals, providing greater opportunity for cells to reorientate themselves towards more favourable conditions[74,82].

MglA is a ras-like GTPase that is essential for regulating polarity in both S-motility and gliding[83]. Both S- and gliding motility require the assembly of polarly localised motility machineries[83,84]. The localisation of such complexes is governed by MglA, which works in concert with MglB and the RomRX complex[85]. Together, MglB and RomRX regulate nucleotide-binding to MglA[83]. MglB acts as a cognate guanine nucleotide hydrolysis

activating protein (GAP), activating MglA's intrinsically low GTPase activity[83,86]. RomRX acts as a guanine exchange nucleotide factor (GEF) that promotes the binding of GTP to MglA[87]. When bound to GTP, MglA is recruited to the leading pole by RomRX, whilst MglB localises to the lagging pole[83,85]. The spatial separation of MglA and MglB ensures the maintenance of a polarity axis by preventing GTP-activated MglA from binding to the lagging pole[83,87]. Upon cell reversal, MglA and MglB inversely oscillate to define a new leading and lagging pole[88].

At the leading pole, MglA interacts directly with AglZ, a cytoplasmic protein essential for gliding[89,90]. Fluorescently-labelled AglZ was observed to form periodic clusters along the length of the cell[91]. These clusters appeared to assemble at the leading pole and disperse at the lagging pole[91]. Notably, the AglZ clusters remained stationary relative to the underlying surface, with gliding cells appearing to move through the clusters[91]. It was proposed that these clusters contained gliding machinery that assemble at the leading pole and then disassembles as it reaches the lagging cell pole[91]. AglZ clusters have been utilised as a marker in almost all subsequent studies trying to underpin the mechanism of gliding, with the localisation of proteins to AglZ clusters being used as an indicator for whether proteins are incorporated into the gliding machinery.

1.5.2 The identification of the gliding machinery

Over the past 15 years, the mechanisms that govern gliding motility have become better understood, owing largely to the identification of the genes that encode the structural components of the gliding machinery[72]. Re-evaluation of the 51 gliding-

implicated genes searching for those encoding likely components of an envelope spanning machinery identified 14 genes that were clustered in three regions of the *M. xanthus* genome[72]. These clusters were denoted Motor 1 (M1), Gliding transducer 1 (G1) and Gliding transducer 2 (G2). The genes within the G1 and G2 clusters were assigned the prefix Glt (Gliding transducer) with GltD-J belonging to G1 and GltA-C & GltK belonging to G2. Finally, the M1 cluster comprises three genes: AglR, AglQ and AglS[72].

Prior to the identification of the G1 and G2 clusters, the Glt genes had already been implicated in gliding through transposon screening[92,93]. However, whether these genes played regulatory or structural roles remained unclear. The deletion of each Glt gene resulted in *M. xanthus* cells that retained intact S-motility but were unable to glide[72]. However, these cells were completely non-motile when the deletions were introduced in Δ*pilA* strains (S-motility deficient)[72]. The deletion of the Glt genes in Δ*frzE* strains also resulted in completely non-motile *M. xanthus* cells[72]. Previous observations demonstrated that gliding defects caused by mutations in regulatory genes could be suppressed by a second mutation which disrupts the Frz system[75]. For example, the deletion of AglZ abolishes gliding motility but a second mutation to FrzE restores gliding[72,89]. The complete lack of motility in Glt/FrzE double mutants provided compelling evidence that the Glt genes encode structural components of the gliding machinery[72].

1.5.3 The Agl-Glt complex

Cell fractionation experiments revealed that the Glt proteins localise to cytoplasm (GltI), IM (GltG/J), periplasm (GltC/D/E/F) and OM (GltA/B/H/K), consistent with the proteins interacting to form an envelope-spanning complex[72,94–96]. Moreover, fluorescently-labelled GltA/B/C/D/F and I have all been observed to localise to AglZ clusters[72,90,97–99]. The products of the M1 cluster (AglRQS) also localise to the IM and AglR and GltG have been shown to interact via a bacterial two-hybrid assay, demonstrating a direct link between Agl and Glt proteins[72]. Taken together, the genetic, localisation and interaction studies have led to the proposal that the 14 genes of the G1, G2 and M1 clusters encode the Agl-Glt complex, an envelope-spanning machinery that localises to AglZ clusters to facilitate gliding[72] (Figure 1.7).

Interestingly, an additional operon containing homologues of the Glt genes was also identified[72]. This operon has since been identified as encoding for the 'needed for sporulation' (Nfs) system (genes NfsA-NfsG are homologues of GltA-GltG)[72], an envelope-spanning machinery responsible for the distribution of spore coat proteins as *M. xanthus* undergoes sporulation in the face of starvation[100]. It is notable that two extremely similar systems perform such distinct roles and demonstrates a strong evolutionary link between gliding and sporulation[100].

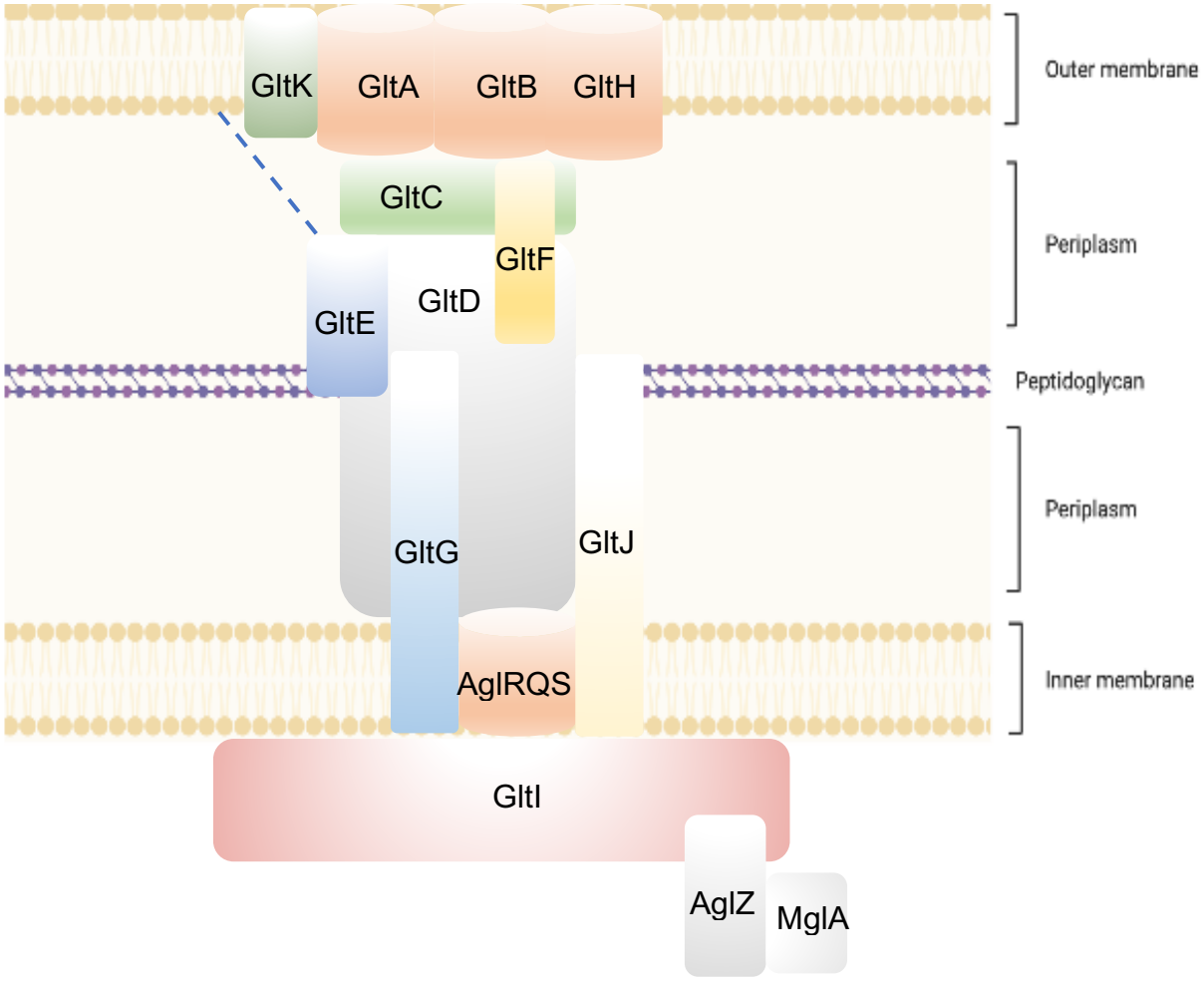


Figure 1.7: Assembly of the Agl-Glt complex based on cell fractionation studies and bioinformatic analysis
The Agl-Glt complex spans the cell envelope to transduce the force generated by the AglRQS motor to the cell exterior to power gliding. The complex likely assembles from distinct subcomplexes: GltA/B/C/H/K form the OM complex; GltD/E/F form the periplasmic hub; and GltH/J and AglRQS form the IM motor complex.

1.5.3.1 AglRQS

Prior to their identification as part of the Agl-Glt complex, the AglRQS genes were identified as essential gliding genes during studies attempting to identify the energy source of gliding[92,98,101]. Gliding in *M. xanthus* is powered by a PMF-powered motor; this was uncovered when treatment of *M. xanthus* cells with CCCP (which disrupts the PMF) immediately halted gliding[68,94,98,101]. As previously described, flagellar rotation is

driven by the PMF-powered MotAB motor, suggesting that both bacterial gliding and swimming are governed by a common form of molecular motor. Scouring the *M. xanthus* genome in search of MotAB homologs identified an operon encoding AglRQS, where AlgR is a MotA homolog and AglQ and S are MotB homologs[98,102]. Notably, the deletion of each of the AglRQS genes abolished gliding but not S-motility[72,98,102]. Subsequent work revelated that the components of AglRQS interact to form a functional proton channel[98] and a point mutation in AglQ (D28N), which was predicted to prevent proton conduction through the channel, immediately halted gliding[98]. Tracking of fluorescently-labelled AglQ has revealed that AglQ assembles at the leading pole and localises to the AglZ clusters distributed along the length of the cell[98]. AglQ/AglZ clusters were observed to disperse upon disruption of the PMF[98]. Meanwhile, fluorescently-labelled AglR was observed to form a continuous, twisted helix spanning the length of the cell when imaged using high-resolution deconvolution microscopy[101], suggesting that the AglRQS motor traverses a helical path to localise to AglZ clusters.

1.5.3.2 Cytoplasmic Glt Components

GltI, a tetratricopeptide repeat (TPR)-rich protein, is the only member of the Glt complex that localises solely to the cytoplasm[94]. The TPR motif is a 34-residue, helix-turn-helix fold that is prevalent in proteins specialised for mediating protein-protein interactions[103]. The basic TPR unit is built around the consensus sequence W4-L7-G8-Y11-A20-F24-A27-P32[94,104]. Whilst this sequence defines many TPR motifs, none of these residues are absolutely conserved, with variations allowed at each position[105]. The structural arrangement of three or more TPR motifs produces a TPR superhelix[94].

The repeating 'consensus' sequence defines the consistent periodicity of the superhelix and confers the amphipathicity of the superhelix, where the concave face contains a hydrophobic groove, whilst the convex face is hydrophilic[103].

GltI is the largest component of the Glt complex. Sequence analysis suggests that GltI comprises at least 23 TPR repeats that likely form an elongated superhelical scaffold[94]. The deletion of GltI results in defective localisation of AglZ and MglA[106,90,94]. The large, TPR-rich nature of GltI suggests that the protein may act as a cytoplasmic scaffold that interacts with cytoplasmic proteins involved in the regulation of gliding such as AglZ and MglA.

1.5.3.3 IM Glt components

Together, GltG and GltJ constitute the (non-motor) IM components of the Glt complex[72]. GltG and GltJ both share homology to TonB[94]. TonB belongs to the TonB/ExbB/ExbD (TBD) system, an IM motor complex responsible for the PMF-powered transport of large nutrients across the OM into the periplasm[107,108]. TonB is tethered to the IM via a transmembrane helix, an intrinsically disordered region then emanates from the IM helix and spans the periplasm to connect to a C-terminal domain[107,108].

TonB interacts with TonB-dependent transporters (TBDTs), 22-stranded β-barrels that are N-terminally plugged[108]. When bound to their target ligand, TBDTs undergo a conformational change which unmasks a ton-box motif[108]. The periplasmic domain of TonB binds to the exposed ton-box, providing a direct tether between the TBDT and

the IM[108] (Figure 1.8). ExbB and ExbD share homology to MotA and MotB respectively, where the conduction of protons results in the rotation of the ExbB/D motor[108]. This rotation results in the mechanical pulling on TonB which unfolds the TBDT plug, opening the pore and allowing the passage of the bound nutrient into periplasm[108].

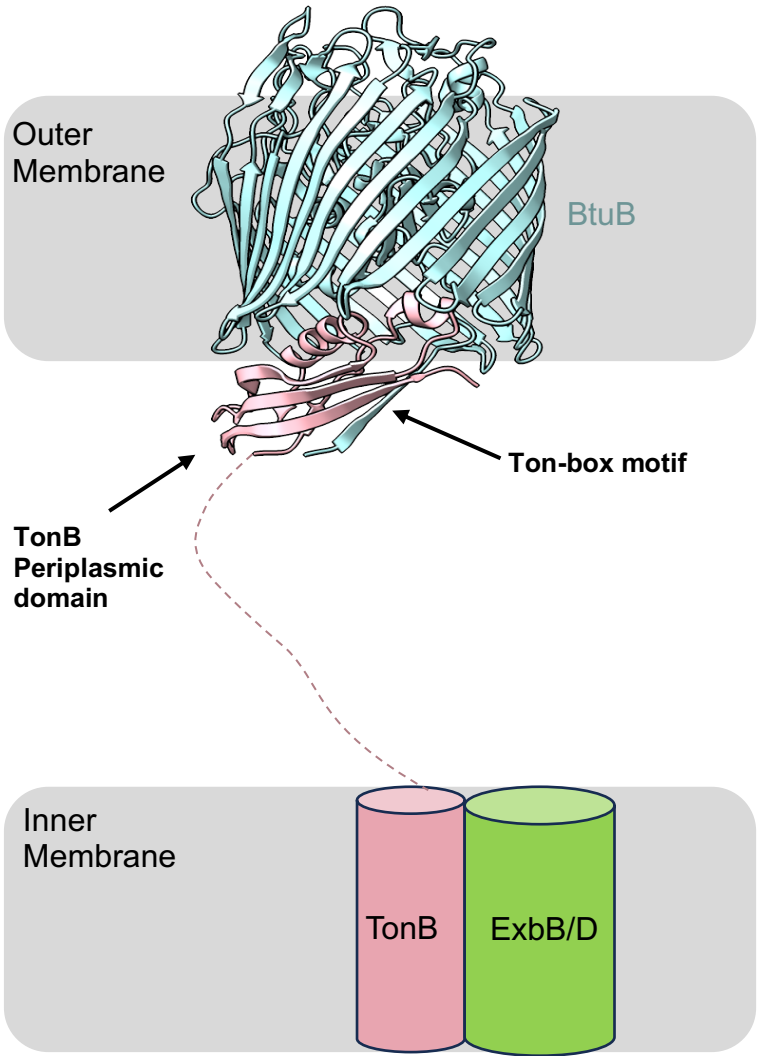


Figure 1.8: TonB-dependent transport
Upon binding its target ligand, the TBDT undergoes a conformational change to present the ton-box motif at periplasmic face of the OM. The periplasmic domain of TonB binds the ton-box motif via β-strand complementation. Rotation of the ExbB/ExbD motor subsequently pulls on TonB which results in the opening of the pore to permit nutrient transport. Example presented above is the crystal structure of the Vitamin B12 transporter BtuB (coloured blue) in complex with the periplasmic domain of TonB (coloured pink) (PDB: 2GSK).

The presence of two TonB homologs within the IM is indicative of a mechanism for how the force generated by the AglRQS motor is translated from the IM across the cell

envelope via the Glt apparatus to the OM. Notably, GltF is predicted to contain a ton-box motif, marking it as a potential binding partner for the periplasmic domains of GltG/GltJ[97]. Whilst GltG and GltJ both share a strong homology to TonB, they exhibit variation in their N-terminal domains. GltG contains a cytoplasmic forkhead-associated (FHA) domain, a domain known to bind phosphopeptides[94]. In contrast, GltJ contains a cytoplasmic GYF domain, a domain known to bind to polyproline sequences on binding partners[94] (GYF domains will be discussed further in Chapter 4). Both FHA and GYF domains are well reported to interact with a broad range of binding partners, making identification of their downstream binding partners difficult[94].

1.5.3.4 Periplasmic Glt components

The periplasmic component of the Glt complex comprises GltC, GltD, GltE and GltF.

GltC contains a Type I signal peptide for secretion to the periplasm via the SEC pathway[109]. In cell fractionation studies, GltC can be detected in both the soluble and OM fractions, suggesting that GltC directly interacts with the OM components of the Glt complex[99,7,96,94]. GltC is predicted to contain up to 9 TPR motifs, supporting a role for GltC in connecting the periplasmic and OM components of the Glt complex.

GltD is the largest periplasmic component of the Glt complex (>1000 aa) and is predicted to contain up to 18 TPR motifs. Like GltC, GltD also contains a Type I signal peptide for secretion to the periplasm[94]. Although, GltD has also been identified in both IM and OM fractions[7,72,94,96]. The large size and TPR-rich fold of GltD, along with the discrepancies in localisation, suggest that GltD may act as an interaction hub to bridge

the IM and OM components of the Glt complex. Consistent with this idea, pull-down experiments against GltD have identified AglZ, GltE, GltI and GltJ as GltD binding partners[94,95]. Fluorescently-labelled GltD also localises to AglZ clusters as well as being observed to follow the same helical path as AglR when tracked using high-resolution deconvolution microscopy[101], suggesting that the periplasmic components of the Glt machinery also traverse a helical path to localise to AglZ.

GltE, also a TPR protein, is predicted to contain up to 8 TPR motifs. In contrast to GltD, GltE is predicted to contain a lipoprotein signal peptide (Figure 1.9). Lipoproteins are translated in the cytoplasm as prolipoproteins which contain a signal peptide that is capped at the C-terminal end by a lipobox[110,111]. The lipobox contains a conserved cysteine residue which is the site of post-translational modification and finally acts as the N-terminus of the mature lipoprotein[110–112]. The signal peptide traffics the prolipoprotein to the IM where Lipoprotein diacylglyceryl transferase (Lgt) catalyses the attachment of diacylglycerol to the conserved cysteine residue[110–112]. The signal peptide is then cleaved from the proliporotein by lipoprotein signal peptidase II (Lsp)[112]. Finally, lipoprotein N-acyl transferase (Lnt) adds an acyl group to the conserved cysteine residue, yielding a mature triacylated lipoprotein that is tethered to a membrane[112]. Most lipoproteins are translocated to the OM by the Lol pathway[111], however, the presence of a Lol avoidance signal leads to the retention of the lipoprotein in the IM[113]. The nature of Lol avoidance signals vary between bacteria. For example, an aspartate at the +2 position of the mature lipoprotein in *E.coli* results in IM retention, whilst a lysine and serine at positions +3 and +4 acts as a Lol avoidance signal in *P. aeruginosa*[113].

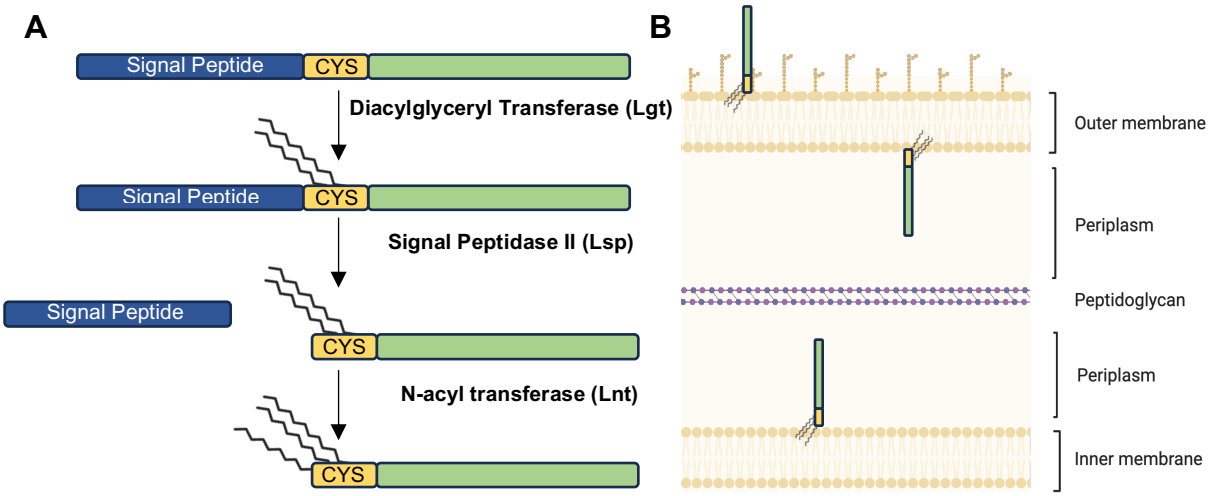


Figure 1.9: Lipoprotein synthesis
A – Overview of the lipoprotein synthesis pathway. Following translation in the cytoplasm, the prolipoprotein is trafficked across the IM via the SEC pathway. At the IM, the prolipoprotein is diacylated by the activity of Lgt. The signal peptide is subsequently cleaved by Lsp. Finally, a third acyl chain is added to the conserved cysteine residue by Lnt, yielding a mature, triacylated lipoprotein.
B – Possible localisation of lipoproteins. Mature lipoproteins can remain at the inner leaflet of the IM or can be trafficked to the inner or outer leaflet of the OM.

GltE possesses none of the reported Lol avoidance signals, suggesting that it would be translocated to the OM. However, it is worth noting that Lol avoidance has not been specifically characterised in *M. xanthus* and therefore, an unorthodox Lol avoidance sequence could result in the IM retention of GltE[114]. Pull-down assays have identified GltE as an interacting partner of GltD[95], suggesting that GltE would likely be tethered to the periplasmic face of the membrane it is trafficked to. The deletion of GltE results in defective GltD localisation, leading to suggestions that GltE either regulates the folding of GltD or plays a role in protecting GltD from the action of periplasmic proteases[106].

GltF is the smallest component of the Glt complex (89 aa) and contains a Type I signal peptide for secretion to the periplasm. The deletion of GltF results in defective GltD

localisation and completely abolishes gliding motility[72]. As previously mentioned, GltF contains a ton-box motif which is the likely binding partner of GltG/J[94].

1.5.3.5 OM Glt components

GltA, GltB, GltH and GltK form the OM components of the Glt complex. GltA, GltB and GltH are all homologs of OmpA and as such, likely contain a β-barrel fold.[94]. GltA and GltB have been shown to interact[99]. Notably, the deletion of either GltA or GltB prevents the remaining protein from localising to the OM[94,99]. GltA and GltB also interact directly with GltC, supporting the notion that GltC functions to link the OM and periplasmic Glt components[99]. Fluorescently-labelled GltA, GltB and GltC were observed to localise to AglZ clusters at sites distributed across the length of the cell[99]. However, in the absence of AglZ, GltA/B are dispersed throughout the OM[99], suggesting that AglZ is required for the recruitment of the GltA/B/C subcomplex into the gliding machinery.

GltK is predicted to contain a lipoprotein signal peptide and was detected in OM vesicles[99]. Proteinase K shaving experiments demonstrated that GltK resides on the inner leaflet of the OM[94,99]. GltK shares homology to Tgl, an OM lipoprotein responsible for integrating the OM components of the TFP system into the OM[115]. Consistent with this homology, the deletion of GltK results in the IM retention of GltA/B, demonstrating that GltK functions to integrate GltA/B into the OM[99], thus playing a key role in the assembly of the Glt complex. The deletion of GltK also resulted in defective AlgZ localisation, with no stationary AglZ clusters being present in the absence of GltK[99].

Prior to being renamed as part of the Agl-Glt complex, GltK was named CglC, with the prefix Cgl signifying contact-dependent gliding[70]. The Cgl prefix was assigned to genes in which the gliding defect caused by their deletion could be restored upon contact with WT *M. xanthus* cells[70]. This restoration was mediated by the transfer of OM lipoproteins from a donor to a recipient cell[115]. The process of OM lipoprotein exchange is utilised by *M. xanthus* to aid in the regulation of social behaviours[116]. In the context of GltK, the transfer from a $GltK^{+}$ donor to a Δ*gltK* strain resulted in GltA/B/C and AglZ all localising to periodic clusters along the length of the cell[99]. This suggests that the OM transfer of GltK to donor cells can permit the OM incorporation of GltA/B, thus stimulating the recruitment of the gliding machinery into a functional unit[99].

1.5.4 Proposed models of gliding motility

Over the decades of research into *M. xanthus* gliding motility, several models have been proposed as to how the gliding machinery enables gliding motility.

1.5.4.1 The slime secretion model

The slime secretion model marked the first proposed mechanism for gliding motility[117]. This model proposed that gliding was driven by the secretion of exopolysaccharide (EPS) slime from the lagging pole of the cell[118]. The hydration of this slime would then act to propel the cell forward[68,118]. The model was based on the observation that gliding *M. xanthus* cells leave a trail of slime behind them, with the rate of the slime release being consistent with the speed of gliding motility[119,120]. Notably, *M. xanthus* cells have been seen to be inclined to follow predetermined slime trails[117,120]. The slime secretion model was supported by EM studies that identified nozzle-like structures at the poles

of the cell[121]. The model was also supported by the fact that some non-gliding mutants were observed to deposit slime trails at both poles of the cell[68,93]. It could be assumed that, if the slime secretion drove gliding, the release and hydration of slime at both ends of the cell would result in no net movement.

Despite reasonable evidence and logical reasoning behind the slime secretion model, recent evidence rules out slime secretion as the mechanism of gliding. In the mass of mutagenesis studies to identify gliding genes, all of the genes that encode a structural component of the gliding machinery belong to the Agl-Glt complex and to date, no genes have been identified that function to form a nozzle. Furthermore, gliding speed is independent of cell length[68,122]. Treatment of *M. xanthus* cells with cephalexin, an antibiotic that inhibits cell wall septation, results in the formation of filamentous *M. xanthus* cells. These filamentous cells glide at the same rate as WT *M. xanthus* cells but do not produce more slime[122]. If gliding was driven by slime secretion, it would be expected that the gliding speed would reduce as cell size increases. The observation that defective gliding mutants leave slime trails at either end of the cell is likely to be a result of hyperreversals, as all the genes resulting in this phenotype have since been implicated in the regulation of cell polarity[93]. Finally, the slime secretion model proposes that slime secretion at the lagging pole drives gliding, however, known gliding genes (MglA, AglZ, Agl-Glt) form periodic clusters along the length of cell and not just at the lagging pole. There remains a clear association between slime and gliding, however, it appears that slime secretion is not the key driver of gliding.

1.5.4.2 Focal Adhesion Model

The focal adhesion (FA) model of gliding originated from the observation that stationary AglZ clusters remained fixed relative to the underlying surface during gliding[91]. From these observations, Mignot and colleagues[91] drew comparisons to the FA-based gliding mechanism utilised by Apicomplexan parasites such as *Plasmodium* and *Toxoplasma*[91]. In brief, apicomplexan gliding is driven by an actin/myosin-based motor system in which myosin, tethered to a membrane complex, transports actin-associated surface adhesins from the leading to the lagging pole[123]. The net effect of the rearward trafficking of adhesins fixed to an extracellular surface results in the forward locomotion of the parasite[123]. The observed similarities between AglZ cluster dynamics and apicomplexan gliding resulted in the proposal of the FA model of *M. xanthus* gliding[91]. This model stipulated that AglZ clusters contained gliding machinery that spans the cell envelope to form FA sites to the underlying substratum. Motor proteins would then drive the rearward movement of surface-attached gliding complexes, generating forward motion[91].

The FA model was strengthened when the bacterial actin homolog MreB was identified as essential for gliding[124–126]. The treatment of *M. xanthus* cells with A22, an antibiotic that inhibits MreB polymerisation, rapidly immobilised gliding cells[126,127]. Initial research suggested that MreB polymerises into a helix that forms along the length of the cell[124,128]. This MreB helix was proposed to form a track that is responsible for the helical localisation patterns observed for AglR and GltD[101]. However, whilst A22 treatment prevented gliding, it did not prevent the helical localisation of AglR and GltD[97], demonstrating that MreB does not form the gliding track. Moreover, the helical

arrangement of MreB has since been invalidated[129]. The role of MreB in gliding appears to be regulatory as evidenced by the fact that MreB, MglA and AglZ have all been shown to associate at the leading pole to promote the formation of the Agl-Glt complex[90].

The FA model draws from the slime secretion model and the close association between gliding and slime, assuming that a component of the gliding machinery binds to extracellular slime to create a direct link between the gliding machinery and the cell exterior[94,130]. Since the proposal of the FA model, the Agl-Glt complex has been identified as the gliding machinery, with fluorescently-labelled AglQ and GltA/B/C/D/F & I all being observed to localise to AglZ clusters[91,98,99]. The FA model proposes that the Agl-Glt complex is assembled at the leading pole and travels along helical tracks towards the lagging pole. Upon reaching the ventral surface of the cell, the Agl-Glt complex would make direct contact with extracellular slime forming a FA site[94,97]. Surface-coupled gliding complexes are then trafficked backwards along a helical track, generating the traction necessary for gliding (Figure 1.10)[90,94,127]. GltA/GltB/GltH were originally proposed as possible mediators of adhesion[94,131], however, no direct role in adhesion can be associated with GltA/B/H as the deletion of any of these genes prevents the assembly of the Glt complex and thus the FA[99].

The rearward trafficking of surface attached Agl-Glt complexes highlights a flaw in the FA model. The transport of an envelope-spanning complex along the cell length would require continuous movement through the cell membranes and peptidoglycan. Such movement would presumably require remodelling of the peptidoglycan. However, to

date, no known peptidoglycan modifying proteins have been implicated in gliding[99,117]. It is therefore difficult to comprehend how a cell would maintain its integrity during the transport of the Agl-Glt complex.

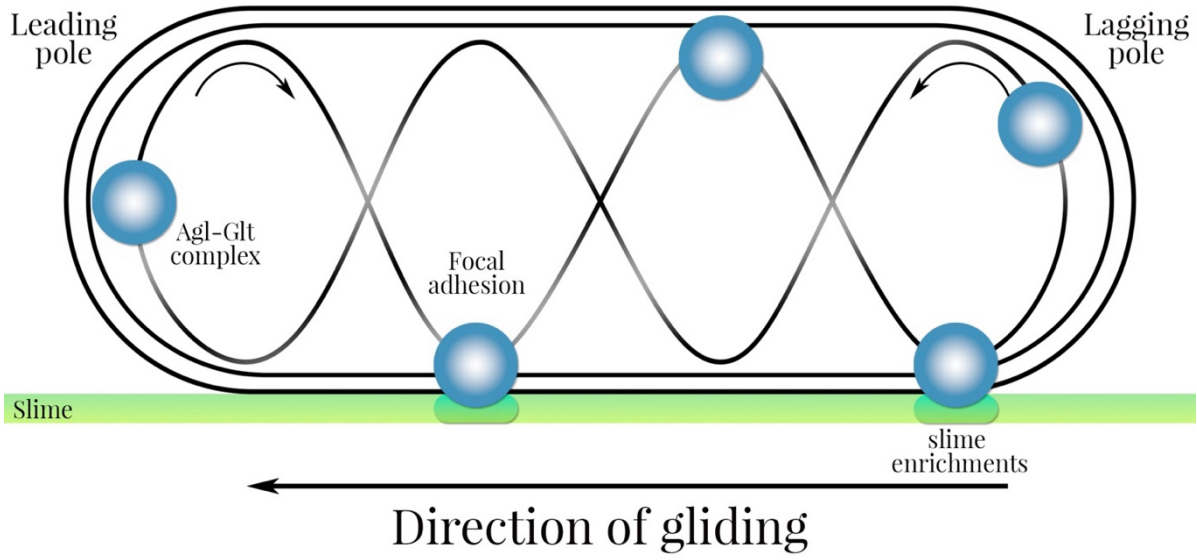


Figure 1.10: The focal adhesion model of gliding
The Agl-Glt complex (blue circles) assembles at the leading pole and is then trafficked along helical tracks by the activity of the AglRQS motor. As the complex reaches the ventral side of the cell, it interacts directly with EPS slime to form a FA site. The rearward trafficking of surface immobilised gliding complexes results in forward propulsion of the cell.

1.5.4.3 The Helical Rotor Model

Several years after the publication of the FA model of gliding, the helical rotor (HR) model was presented (Figure 1.11)[102]. The HR model was proposed to provide an alternative explanation for the helical localisation observed for both AglR and GltD[102]. The HR model proposes that the AglRQS motor pushes the Agl-Glt complex along helical tracks[102]. This helical movement distorts the cell envelope at the ventral side of the cell, creating 'ripples' at the cell surface[102]. These surface 'ripples' would push against slime, generating the propulsion that drives gliding. Consistent with the HR model, periodic surface deformations in gliding cells have been observed by total

internal reflection fluorescence (TIRF) microscopy and EM[101,132,133]. The HR model argues that the clusters of gliding proteins observed are not due to the formation of FAs, but instead, are a result of increased drag forces experienced by the gliding machinery as it reaches the ventral side of the cell[102]. The increased drag forces slows down the movement of the gliding machinery, resulting in 'traffic jams' that result in the appearance of stationary clusters of gliding proteins[102].

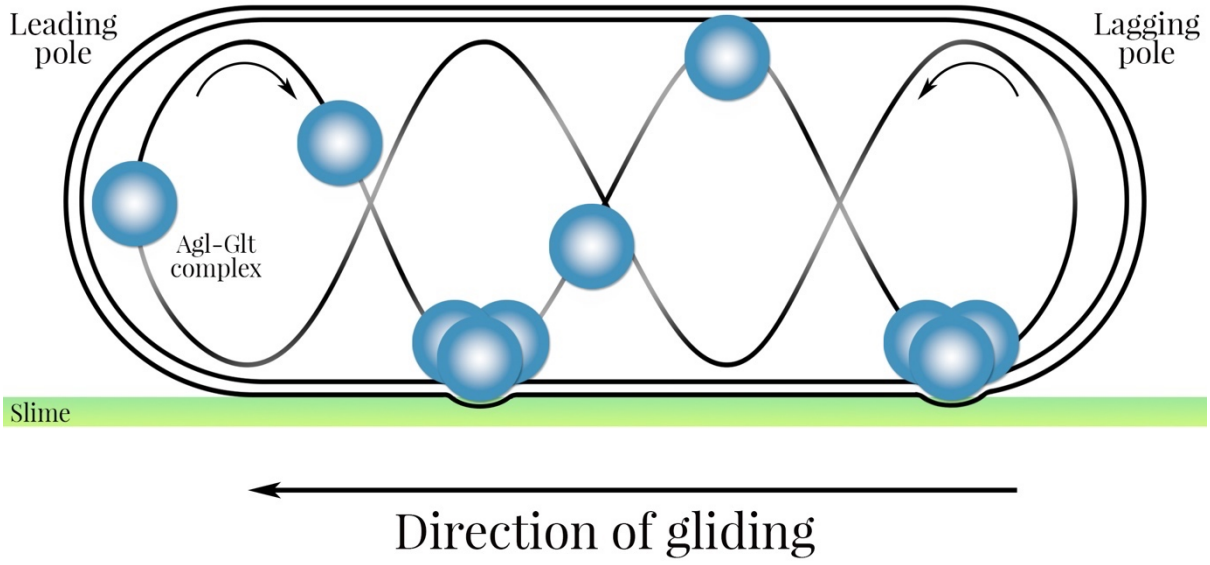


Figure 1.11: The helical model of gliding
The Agl-Glt complex (blue circles) assembles at the leading pole and is then trafficked along helical tracks by the activity of the AglRQS motor. Upon reaching the ventral surface, the gliding machinery pushes against the PG to distort the OM to create surface ripples. These ripples create drag forces that propel the cell forward. The increased drag slows the movement of the gliding machinery creating traffic jams that appear as stationary clusters.

1.5.4.4 FA vs HR models

Mathematical modelling was used to further scrutinise the FA and HR models of gliding[134]. These mathematical models primarily focussed on the fact that the key difference between the FA and HR models lies in the nature of the interaction between the Agl-Glt complex and the extracellular substratum. The FA model would exhibit

strong coupling mediated through an adhesin, whilst the HR model would depend upon viscous drag coupling[134]. A model was developed to simulate how cells that were either elastically- (FA) or viscously- (HR) coupled to the substratum would respond to cell-cell collisions[134]. In the elastic coupling model, cells were observed to be difficult to displace following cell collision, presumably a result of the tight association between the cell and the substratum[134]. In contrast, in the viscous coupling model, cells were observed to move over a much greater distance[134] due to the weaker interactions with the substratum. Comparison of each model with experimental observations of isolated cell-cell collisions revealed that cells closely follow the behaviour simulated in the elastic-coupling model[134], suggesting that the interactions between *M. xanthus* cells and the substratum are elastic in nature. These findings were consistent with the FA model of gliding.

Further validation of the FA model came following the identification of CglB as a candidate adhesin responsible for linking the Agl-Glt complex to the substratum[135]. CglB is an OM lipoprotein that is predicted to contain a von Willebrand A (VWA) domain, a fold that is also found in the adhesins that mediate gliding in apicomplexan parasites[135]. The VWA domain contains a metal-ion-dependent adhesion site (MIDAS) motif that co-ordinates a divalent cation, often Mg^{2+} or Mn^{2+}. The binding of a cation to the MIDAS site results in conformational changes that facilitate high affinity binding to a ligand[135].

Fluorescent CglB was observed to localise to stationary AglZ clusters, indicative of incorporation into the gliding machinery[135]. The deletion of CglB abolished gliding but

did not prevent the formation of the gliding machinery, with AglZ still observed to form clusters along the cell. However, these clusters were not stationary, instead moving continuously from the leading to lagging pole[135]. This indicates that whilst it is not essential for the formation or trafficking of the Agl-Glt complex, CglB likely interacts with the underlying substratum and thus immobilises the gliding machinery to form FA sites[135].

Immunolabelling confirmed that CglB localised to the cell surface, which is consistent with CglB functioning as a surface exposed adhesin[135]. CglB has been shown to interact with the OM Glt components via pull-down assays and the OM Glt components appear to regulate the retention of CglB at the cell surface, shielding CglB from protease activity[135]. In Δ*gltA/B/K* strains, CglB is rapidly proteolysed; this degradation is prevented if *M. xanthus* cells are grown in the presence of EDTA, suggesting that a metalloprotease is responsible for proteolysis[135]. Whilst the OM Glt components interact with CglB, they are not responsible for its surface exposure, as CglB localises to the cell surface in Δ*gltA/B/H* cells when grown in media containing EDTA[135]. The mechanism by which CglB is trafficked to the cell surface to interact with the Glt complex remains to be determined.

1.5.5 The current model of *M. xanthus* gliding

Both the FA and HR hypothesise that the Agl-Glt complex assembles at the leading pole and the subsequent activity of the AglRQS motor would power the rearward movement of the complex[91]. This hypothesis is largely based on the observation that AlgZ clusters originate at the leading pole. However, current evidence supports a view

where the Agl-Glt complex is divided into distinct sub-complexes that dock together to form a complete complex to facilitate gliding[97].

Evidence of functional sub-complexes emerged when, in addition to forming stationary clusters, AglZ was observed to form dynamic clusters that migrated from the leading to the lagging pole, following the same helical trajectory previously observed for AglR and GltD[97]. Co-localisation studies of AglZ, AglQ, GltC, and GltD demonstrated that GltC and GltD only associated with AglZ at stationary clusters, whilst AglQ localised to AglZ in both stationary and dynamic clusters[97]. Gliding was only observed when at least one stationary AglZ cluster was present and gliding speed increased in accordance with an increase in the number of stationary AglZ clusters. In contrast, the number of dynamic clusters had no impact on gliding speed[97], indicating that stationary AglZ clusters likely correspond to fully assembled gliding machineries that generate the traction forces necessary for gliding. Further mutagenesis studies that probed the impact of each component of the Agl-Glt complex on AglZ localisation dynamics revealed that the deletion of the cytoplasmic (GltI) and IM components (GltG/J) had the largest impact on AglZ localisation, whilst the deletion of periplasmic and OM components only caused minor disruptions to AglZ localisation[97]. These findings are indicative of a degree of structural separation between the cytoplasmic/IM (CIM) and periplasmic/OM (POM) components of the Agl-Glt complex. Dynamic AglZ clusters likely represent motor subcomplexes comprising the AglRQS motor, the CIM Glt components and regulatory proteins (including AlgZ and MglA). These motor complexes would assemble at the leading pole and travel along a helical track towards the lagging pole. At points along this track, the CIM motor complex interacts with the

POM components of the Glt complex to form a complete and functional gliding machinery. This is presumed to result in conformational changes in the OM Glt components, allowing for surface exposure of CglB which subsequently binds to the substratum and immobilises the Agl-Glt complex and thus forming a FA site (Figure 1.12).

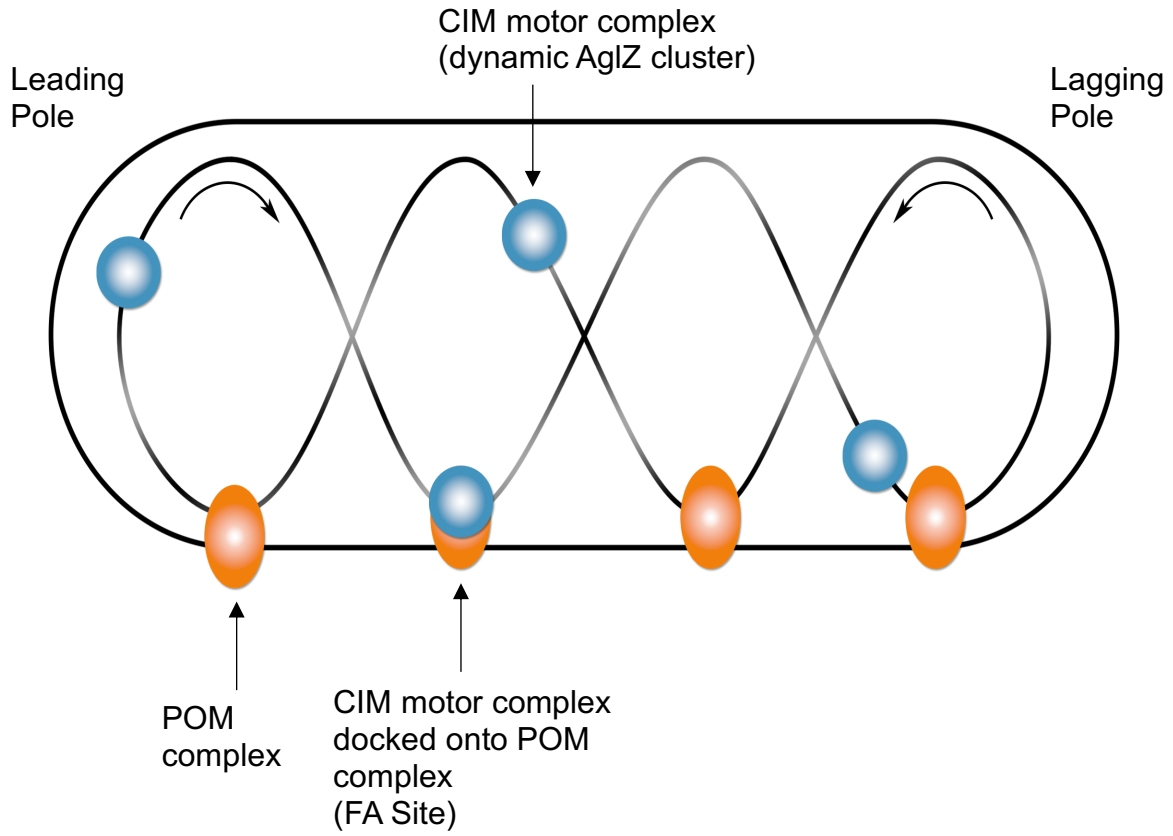


Figure 1.12: Assembly of complete gliding machineries from pre-formed subcomplexes
Dynamic AglZ clusters comprising AglZ, cytoplasmic/IM Glt proteins and AglRQS form an IM motor complex (blue circles). The IM motor complex traverses a helical track until reaching a periplasmic/OM Glt complex (orange circles), onto which the IM complex will dock to form a complete gliding machinery.

The mechanisms by which the elements of Agl-Glt complex interact with each other remain to be uncovered. However, the observed interaction between AglR and GltG as well as GltG/J's homology to TonB suggest that the force generated by the AglRQS motor will be transferred from the motor through GltG to GltF, with which GltG would bind the ton-box motif. GltD and GltF have also been shown to interact[72],

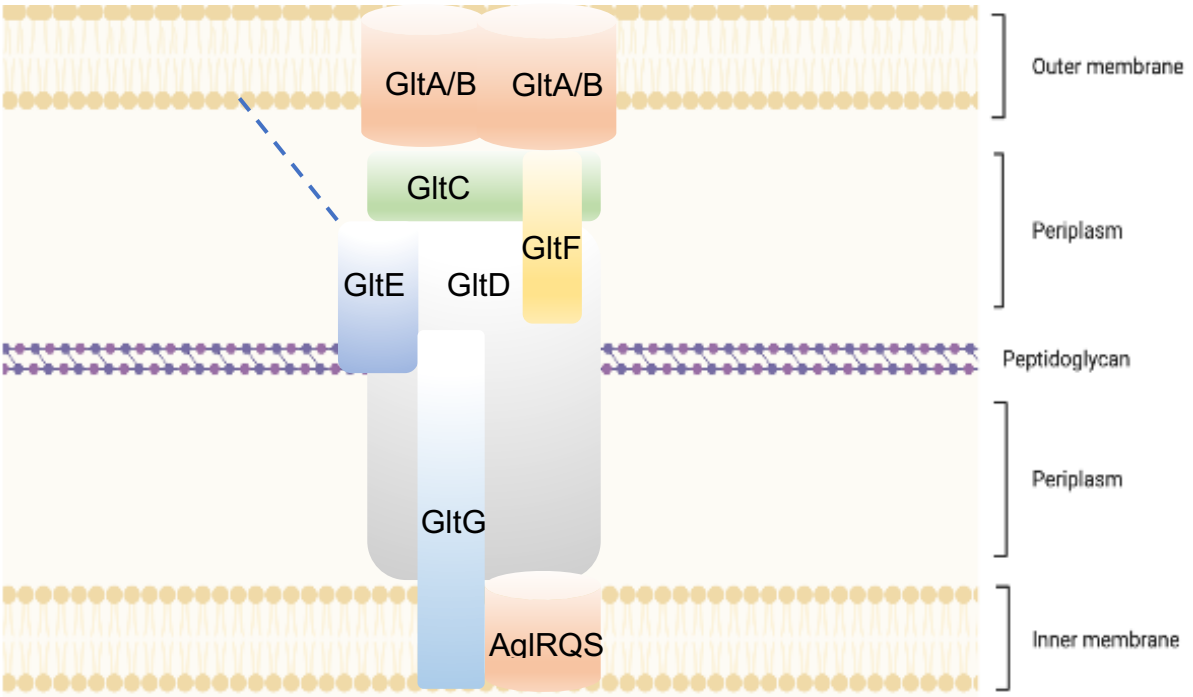
demonstrating how the force from the motor would be transferred across the periplasm to the OM. Whilst this is a compelling hypothesis, it remains to be functionally and structurally validated.

It remains to be seen how the Agl-Glt complex moves laterally through the peptidoglycan without harming the integrity of the cell, especially given the fact that no currently identified gliding proteins contain peptidoglycan remodelling domains[99]. Moreover, the inhibition of peptidoglycan synthesis has no impact on a cell's ability to glide[129], suggesting that the modification of peptidoglycan is not a part of the gliding process. One possibility is that the periplasmic/OM Glt components are less dynamic than previously thought and instead, form 'arrays' throughout the periplasm and OM. The flexible periplasmic domain of GltG/J could then bypass the peptidoglycan to interact with the periplasmic-OM complex to form a FA site. The motor could then progress along the cell until it reaches the next periplasmic-OM complex, creating a new FA. This model would see the motor progress backwards along the cell, passing from one periplasmic-OM complex to the next, creating a new FA site with each 'step'. This would result in the forward locomotion of the cell without the need for remodelling of the peptidoglycan.

1.5.6 Gliding in *B. bacteriovorus*

Gliding in *B. bacteriovorus* shares key similarities to *M. xanthus*, suggesting that the two organisms share a common mechanism. Genes encoding for components of the Agl-Glt complex are conserved in *B. bacteriovorus*[136]. However, *B. bacteriovorus's* Agl-Glt encoding genes are contained within operons, as opposed to the distinct gene

clusters seen in *M. xanthus*. Notably, *B. bacteriovorus* contains four gliding operons[137] (named Glt1–4), although, none of these contain a full complement of Agl-Glt genes as there are no genes encoding for GltH–K present within *B. bacteriovorus's* genome[58]. These differences could suggest that the Agl-Glt genes common to both *M. xanthus* and *B. bacteriovorus* represent the core gliding machinery whilst GltH-K could represent accessory elements that are specific for *M. xanthus* gliding. The four gliding operons are summarised in Table 1.3 and Figure 1.13.



Gliding Operon	Genes	Function
	GltA GltB GltC GltD GltE GltF GltG AglRQS Adhesin	
bd0828-bd0838 (Glt1)		Essential gliding operon
bd0412-bd0420 (Glt2)	x x x	Prey entry
bd1473-bd1483 (Glt3)	x x x	Biofilm predation
bd2368-bd2377 (Glt4)	x x x	HI Biofilm navigation

Figure 1.13: Assembly of the *B. bacteriovorus* Agl-Glt complex and overview of the four *B. bacteriovorus* gliding operons
Top: Schematic diagram demonstrating the possible arrangement of the Agl-Glt complex in *B. bacteriovorus*.
Bottom: Table summarising the *B. bacteriovorus* gliding operons. *B. bacteriovorus* contains four gliding operons named Glt1-Glt4. However, none of these operons contain a full complement of the Agl-Glt genes seen in *M. xanthus* (GltH-K are absent in *B. bacteriovorus*). The Glt1 operon shares the closest similarity to *M. xanthus*, containing genes encoding for GltA-GltG and AglRQS. The Glt2-4 operons lack genes encoding for GltA and GltE.

Table 1.1: Agl-Glt complex encoding genes in *B. bacteriovorus*

Gene name	Glt1 Operon	Glt2 Operon	Glt3 Operon	Glt4 Operon	Sequence identity between Glt1 and *M. xanthus* orthologs
GltA	Bd0828	X	X	X	26.9%
GltB	Bd0829	Bd0414	Bd1475	Bd2370	26.7%
GltC	Bd0831	Bd0413	Bd1474	Bd2369	24.7%
GltD	Bd0832	Bd0417	Bd1478	Bd2374	25.0%
GltE	Bd0833	X	X	X	21.5%
GltF	Unannotated	Unannotated	Bd1477	Bd2373	34.2%
GltG	Bd0834	Bd0416	Bd1476	Bd2372	38.9%
GltH	Absent in *B. bacteriovorus*				X
GltI	Absent in *B. bacteriovorus*				X
GltJ	Absent in *B. bacteriovorus*				X
GltK	Absent in *B. bacteriovorus*				X
AglQ	Bd0837	Bd0419	Bd1480	Bd2376	28.7%
AglR	Bd0836	Bd0420	Bd1481	Bd2377	43.1%
AglS	Bd0838	Bd0418	Bd1479	Bd2375	31.3%
CglB	X	Bd0412	Bd1483	Bd2368	X

Of *B. bacteriovorus's* gliding operons, Glt1 shares the strongest similarity to the *M. xanthus* gliding machinery (Table 3), containing genes encoding for GltA-G and the AglRQS motor, although no adhesin-encoding gene is present on the operon. The Glt1 operon is essential for predation and transposon screening revealed that insertions into the genes of this operon resulted in defective prey attachment/killing[138].

The Glt2–4 operons appear to represent more degenerate/specialised versions of the gliding machinery, with each operon lacking genes encoding for GltA and GltE. The absence of these genes is notable given that transposon insertions into any of the Glt1 genes resulted in defective gliding[138]. The absence of GltA in these operons is also notable, given that the deletion of GltA resulted in the IM retention of GltB in *M. xanthus*[99]. This suggests that the assembly of the OM Glt components differs between the Glt2–4 operons and the Glt complex of *M. xanthus*. In contrast to the Glt1 operon, Glt2–4 each encode a MIDAS adhesin that is homologous to CglB.

The occurrence of four gliding operons suggests the presence of distinct gliding machineries, each with different functions. Consistent with this, transcriptomics studies demonstrate that each operon has a distinct expression profile[41]. The Glt2 operon is upregulated during predation and invasion, suggesting that this operon plays a role in prey contact or entry[41,139]. In contrast, the Glt3 operon is downregulated during the transition from predation to invasion, implying that this operon plays a role in biofilm navigation and prey encounter[41]. Consistent with this, transposon insertions in the AglR and MIDAS genes of the Glt3 operon resulted in impaired predation within biofilms[136]. Finally, the Glt4 operon is upregulated during H-I growth and may be involved in the formation or navigation of H-I self-biofilms[41].

The absence of GltH–K in any of *B. bacteriovorus's* gliding operons implies that some differences exist between *M. xanthus* and *B. bacteriovorus* gliding. GltH/K form part of the OM complex in *M. xanthus*, where GltK plays a role in the localisation of GltA/B to the OM[99]. The absence of GltK in *B. bacteriovorus*, along with the fact that three of the

four gliding operons lack a GltA gene, suggests that *B. bacteriovorus's* OM Glt complex assembles in a different manner to that of *M. xanthus's*. In *M. xanthus*, the OM Glt complex has been shown to play a role in protecting CglB, ensuring its retention at the cell surface[135]. The differences in the assembly of the OM Glt complex in *B. bacteriovorus* would also suggest that there are differences in the manner in which the Glt complex interacts with the MIDAS adhesins. Prior to their renaming, both GltH/K were denoted with the Cgl prefix. This suggests that in *M. xanthus,* these genes may also play a role in swarm motility[140] which are not applicable to *B. bacteriovorus*.

The absence of a GltI gene in *B. bacteriovorus* is presumably a result of the differences in the regulation of gliding between *B. bacteriovorus* and *M. xanthus*. As previously discussed, in *M. xanthus*, the Frz system is responsible for setting the cell polarity which subsequently results in the recruitment of MglA and AglZ at the leading pole[83]. These then interact with MreB and GltI to initiate the formation of the Glt complex[89,90,127]. Despite *B. bacteriovorus* encoding MreB and MglA homologs, it lacks AglZ and all genes responsible for Frz signalling. Whilst MreB remains essential for gliding in *B. bacteriovorus*[58], MglA is not and has evolved to regulate prey invasion[137] (MglA's role in *B. bacteriovorus* will be further discussed in Chapter 4). Gliding in *B. bacteriovorus* is instead under the regulation of the cyclic dinucleotides cyclic di-GMP (c-di-GMP) and 3',3'-cyclic-GMP-AMP (cGAMP)[139,141].

1.6 An introduction to c-di-GMP signalling

C-di-GMP is a cyclic dinucleotide that acts a second messenger in the majority of bacteria[142] (Figure 1.14)[143]. Many cellular processes fall under the regulation of c-di-

GMP signalling, including but not limited to motility, virulence, biofilm formation and protein secretion[144,145]. C-di-GMP is synthesised from two GTP molecules by diguanylate cyclases (DGCs)[142,144]. DGCs contain a GGDEF domain, where catalytic activity occurs[142]. In order for synthesis to begin, DGCs respond to a stimulus which promotes the dimerisation of two DGCs through their GGDEF domains[142,146]. This dimerisation creates an active site at the GGDEF interface, to which, each domain recruits one molecule of GTP[142]. Once synthesised, c-di-GMP can occupy monomeric, dimeric or tetrameric states[142].

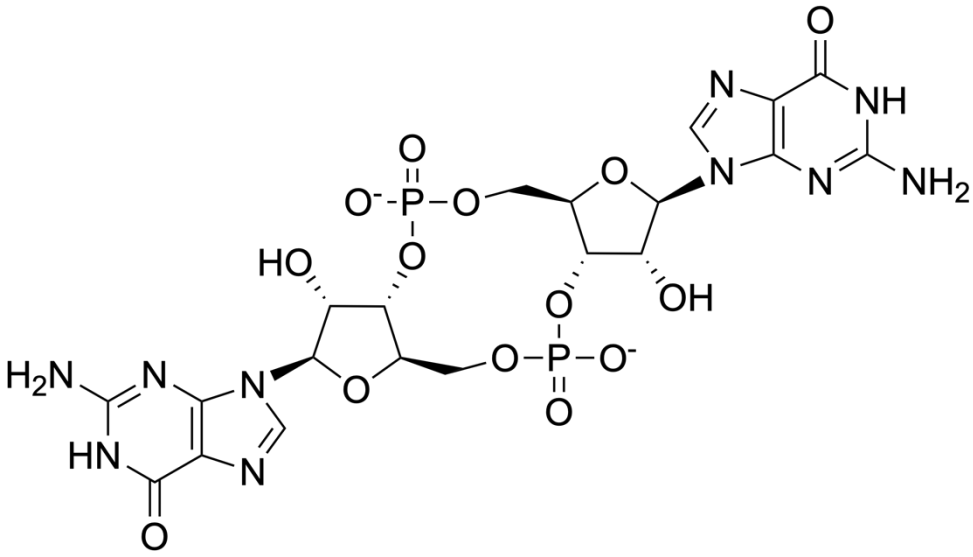


Figure 1.14: Chemical structure of c-di-GMP.
C-di-GMP is synthesised from two molecules of GTP at the GGDEF domain of DGCs. Image made using ChemDraw.

DGCs rarely consist of only GGDEF domains. The dimerisation and thus catalytic activity of DGCs is often controlled by regulatory domains[145]. Receiver (REC) domains are commonly associated with DGC regulatory domains, with the activity of some DGCs being completely dependent on phosphorylation of their REC domain, which facilitates DGC dimerisation[145,147]. The majority (>60%) of DGCs also contain an inhibitory (I) site, which consists of an RXXD motif found ~five residues upstream

of the GGDEF active site[142,145]. The binding of c-di-GMP to the I-site prevents GGDEF dimerisation and thus regulates c-di-GMP levels[142].

The breakdown of c-di-GMP is controlled by c-di-GMP specific phosphodiesterases (PDEs). PDEs contain EAL or HD-GYP domains which catalyse the breakdown of c-di-GMP to 5′-phosphonoguanylyl-(3′-5′)-guanosine (pGpG) and GMP respectively[142,144,148]. Together, DGCs and PDEs work in concert to regulate c-di-GMP signalling pathways by orchestrating variations in cellular c-di-GMP levels[144]. Owing to the prevalence of c-di-GMP signalling in bacteria, a large number of c-di-GMP receptors have been identified. Known c-di-GMP receptors include YajQ, riboswitches, degenerate PDEs and DGCs (that have lost catalytic activity), MshE domains and PilZ domains[144]. This range of receptors is, in part, responsible for the regulatory diversity of c-di-GMP signalling pathways.

PilZ domains represent both the first identified and the most abundant class of c-di-GMP receptor[142,145,149] and are heavily associated with cell motility, as evidenced by their involvement in the assembly of TFP and the regulation of flagellar swimming[150–153]. Notably, the PilZ domain of the flagellar brake protein YcgR has been shown to interact directly with the flagellar motor protein MotA to control flagellar swimming[153]. The overall fold of the PilZ domains consists of a six-stranded β-barrel capped C-terminally by an α-helix[154] (Figure 1.15). C-di-GMP binding occurs through the conserved RxxxR and (D/N)x(S/A)xxG motifs, each of which interacts with one guanine base of c-di-GMP[149,155] (Figure 1.15). Notably, these motifs permit binding to both monomeric and dimeric c-di-GMP[142].

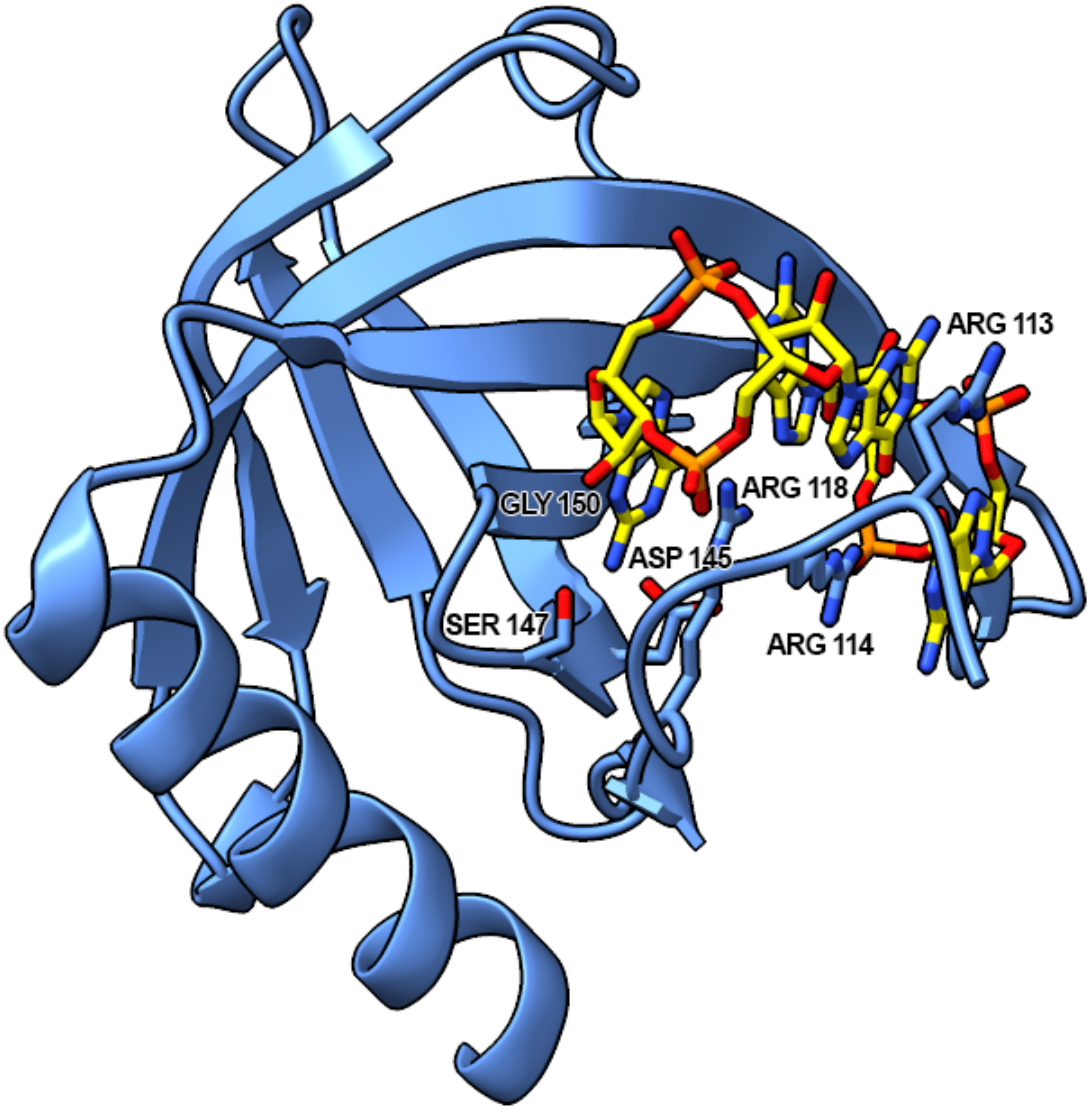


Figure 1.15: PilZ:c-di-GMP interaction
Crystal structure of the PilZ domain of YcgR in complex with c-di-GMP. The PilZ domain (coloured blue) is formed by a six-stranded β-barrel capped C-terminally by an α-helix. The PilZ domain binds to c-di-GMP through the RxxxR and (D/N)x(S/A)xxG consensus motifs. Consensus residues are shown as sticks. PDB: 5Y6G.

1.6.1 C-di-GMP signalling in *B. bacteriovorus*

B. bacteriovorus contains an extensive c-di-GMP signalling network; to date four DGCs, six HD-GYPS, one EAL, 18 PilZ domain-containing proteins, one (predicted) c-di-GMP binding riboswitch and one degenerate DGC functioning as a receptor have been identified[139,156]. The active DGCs (denoted DgcA-DgcD) are conserved in all *B.*

bacteriovorus strains, except for DgcD which is absent in the genome of *B. bacteriovorus Tiberius*[139]. In *B. bacteriovorus* HD100, the deletion of each active DGC produces a distinct phenotype which cannot be compensated for by the activity of the remaining DGCs, suggesting that each DGC regulates a specific c-di-GMP signalling pathway as opposed to being 'general purpose' enzymes observed in other organisms[139]. DgcB and DgcC have been implicated in the regulation of prey invasion and the transition to a HI lifestyle respectively, whilst no function has currently been assigned to DgcD[139].

DgcA plays a role in the regulation of motility[139,141]. The deletion of DgcA resulted in non-flagellate *B. bacteriovorus* that were unable to predate within liquid media[139]. However, when applied directly to prey on solid surfaces, Δ*dgcA* cells were observed to invade prey cells[139]. Despite successful replication, progeny cells were observed to be non-motile and could not escape prey cells[139]. As previously discussed, progeny escape is driven by gliding motility (as flagellar may tangle in the close confines of the prey shell). Complementation of Δ*dgcA* cells with the WT gene was observed to restore gliding and thus allowed for progeny escape and the continuation of the predatory cycle[139]. Interestingly, complemented cells remained non-flagellate and the role of DgcA in flagellar synthesis remains to be characterised[139].

The initial work on DgcA strongly suggested that gliding motility was under the regulation of c-di-GMP. However, DgcA has since been identified as a hybrid promiscuous (Hypr) GGDEF containing protein[157]. Hypr GGDEF domains are extremely rare, accounting for less than 1% of all GGDEF domains. These domains

possess both c-di-GMP and 3',3'-cyclic-GMP-AMP (cGAMP) synthesis activity[157] (Figure 1.16)[158].

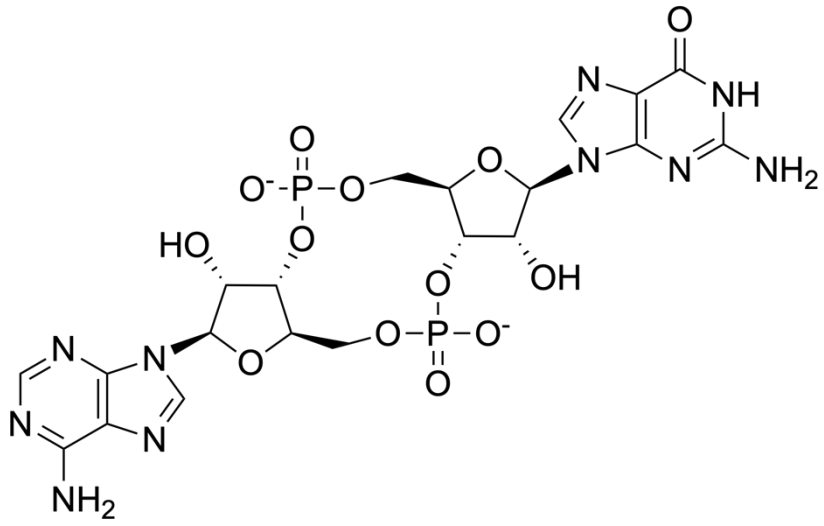


Figure 1.16: Chemical structure of cGAMP.
Hypr GGDEF domains can utilise both ATP and GTP as substrates. cGAMP is synthesised from one molecule of GTP and one molecule of ATP in Hypr GGDEF containing DGCs. Image made using ChemDraw.

Hypr GGDEF domains are characterised by a conserved serine residue located upstream of the GGDEF motif (S214 in DgcA) (Figure 1.17), which enables utilisation of both ATP and GTP[141]. Comparison of the synthase activity of DgcA and $DgcA_{S214D}$ revealed that the WT protein produced predominately cGAMP, whilst $DgcA_{S214D}$ exclusively made c-di-GMP[141]. The synthesis of cGAMP is abolished in both Δ*dgcA* and $DgcA_{S214D}$ cells, demonstrating that DgcA is the sole cGAMP synthase in *B. bacteriovorus*[141]. $DgcA_{S214D}$ exhibited the same phenotype as Δ*dgcA* cells, where progeny cells were unable to glide out of prey debris[141]. Like Δ*dgcA* cells, gliding can be restored to $DgcA_{S214D}$ cells upon complementation with the WT gene[141]. Despite being unable to glide, $DgcA_{S214D}$ cells were observed to have functional flagellar; this observation suggests that c-di-GMP and cGAMP produced by DgcA regulate distinct motility systems, with c-di-GMP regulating flagellar swimming and cGAMP regulating

gliding motility[141] (Figure 1.18). Interestingly, the expression levels of two gliding genes (Bd1480 (AglQ) and Bd3099 (MIDAS) adhesin) were reduced in Δ*dgcA* and $DgcA_{S214D}$ cells, further implicating cGAMP in the regulation of gliding[141]. Notably, the Glt3 operon contains a PilZ domain-containing protein (Bd1482). It is therefore possible that Bd1482 is the target for cGAMP signalling from DgcA.

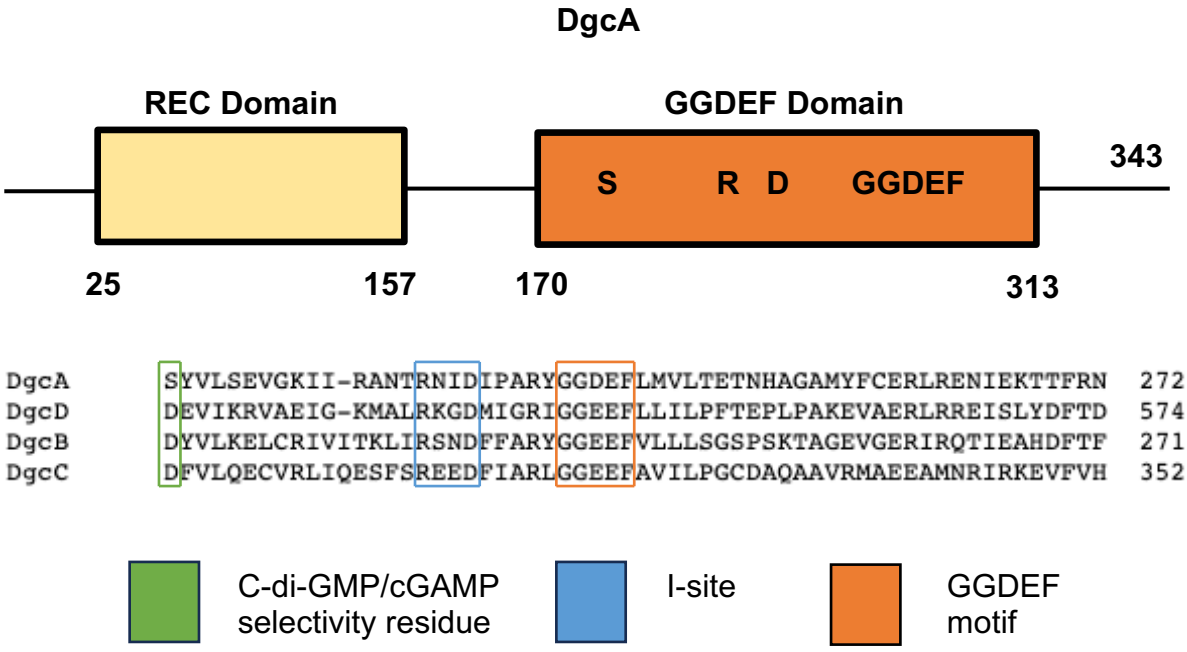


Figure 1.17: Sequence alignment of the GGDEF domains of *B. bacteriovorus's* active DGCs
The presence of S214 (green box) in DgcA in place of a conserved aspartic acid residue permits binding to both ATP and GTP and thus synthesis of both c-di-GMP & cGAMP. The position of the GGDEF active (orange box) and inhibitory site (blue box) are also shown. Figure adapted from Lowry et al. 2023[141].

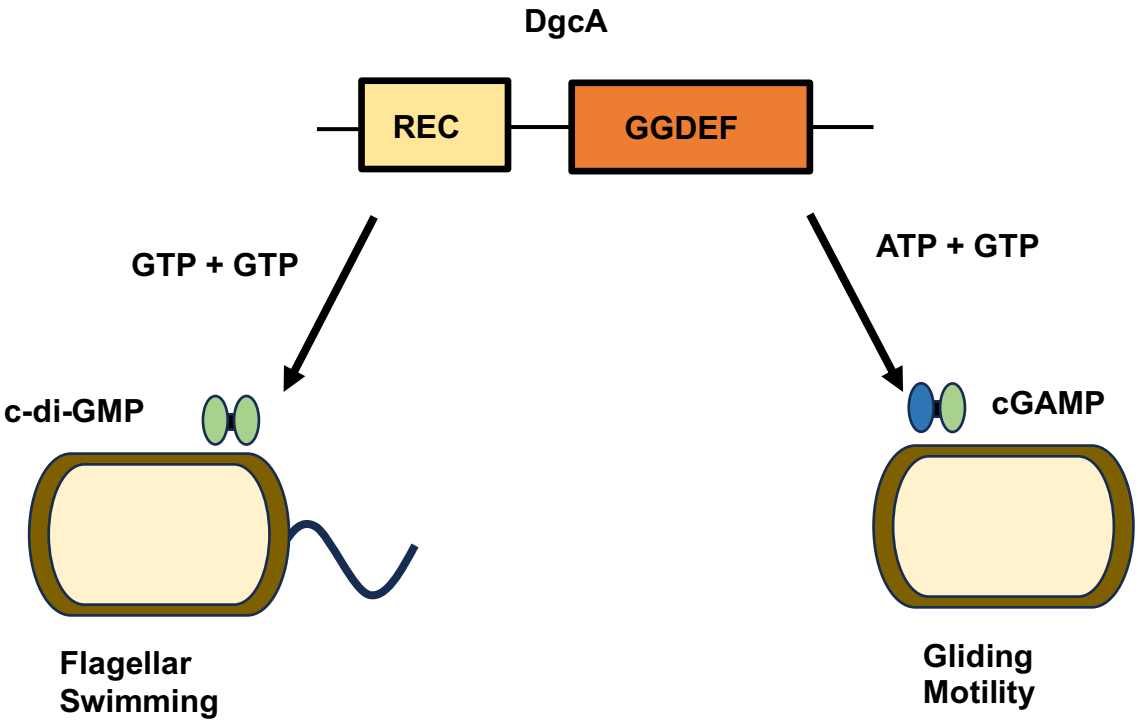


Figure 1.18: DgcA can produce both c-di-GMP and cGAMP
The Hypr GGDEF domain of DgcA has a dual c-di-GMP and cGAMP synthase activity. The deletion of DgcA abolishes both flagellar swimming and gliding motility. However, a S214D mutation renders DgcA unable to synthesise cGAMP. The disruption of cGAMP synthesis abolishes gliding motility but flagellar swimming remains intact. Taken together, these observations suggest that c-di-GMP synthesised by DgcA regulates flagellar swimming whilst cGAMP synthesised by DgcA regulates gliding motility[141].

1.7 Thesis Overview

This thesis primarily focuses on the structural and biochemical characterisation of components of the Agl-Glt complex and related regulatory proteins. The specific project aims are presented in the relevant chapters.

Chapter 3 focusses on the interaction between GltD1 and GltE1. Using isothermal titration calorimetry, we demonstrate that GltD1 and GltE1 interact with nanomolar affinity. The crystal structures of GltE1 and a truncated GltD1:GltE1 complex are presented. The GltD1:GltE1 structure represents the first structural insight into the connectivity within the Agl-Glt complex.

The focus of Chapter 4 switches to the regulation of gliding, where we investigate Bd1996, a hybrid GYF:PilZ protein, and MglA. We demonstrate that the PilZ domain of Bd1996 has dual binding affinity for c-di-GMP and cGAMP and identify DgcA as a Bd1996 binding partner using pull-down assays. The crystal structure of MglA is also presented. We demonstrate, through combined experimental and computational structural characterisation, that MglA likely undergoes GTP-dependent conformational changes to interact with its binding partner, Bd2492.

Finally, Chapter 5 presents the crystal structure of Bd1431, a nuclease belonging to the *staphylococcal* nuclease family that is used by *B. bacteriovorus* during prey invasion. Bd1431 contains a novel dimeric *staphylococcal* nuclease fold that maintains activity against nucleic acids.

Chapter 2

Materials and Methods

2.1 General methods

2.1.1 Molecular Cloning

Throughout this work, target genes were cloned from the *B. bacteriovorus* HD100 genome (DNA provided by Sockett lab, University of Nottingham) by either traditional restriction cloning or restriction-free cloning.

2.1.1.1 Restriction cloning

For restriction cloning, PCR was performed using the recommended protocol for Phusion DNA Polymerase. Each reaction contained: 10 μL 5X Phusion HF buffer, 1 μL 10 mM dNTPs, 2.5 μL 10 μM forward primer, 2.5 μL 10 μM reverse primer, 0.5 μL template (~50–250 ng), 1.5 μL DMSO and 0.5 μL Phusion DNA Polymerase (NEB). The final solution was made up to 50 μL with distilled water. The thermocycling conditions used are shown in Table 2.1. The annealing temperature used was calculated using NEB Tm calculator. In general, for each PCR, reactions were performed at the predicted Tm and 5 °C above and below the Tm.

Table 2.1: Thermocycling conditions for PCR

Step		Temperature (°C)	Time
Initial Denaturation		98	30 s
	Denaturation	98	10 s
30 Cycles	Annealing	melting temperature (Tm) ± 5	30 s
	Elongation	72	30 s per kb
Final extension		72	10 minutes
Hold		4	

Agarose gel electrophoresis was used to assess the success of the PCR. Following PCR, 5 μL PCR product was mixed with 1 μL 6X purple loading dye (NEB). Gel samples were run on 1% agarose gels (supplemented with SafeView nucleic acid stain) at 120 V for ~45 minutes. If the gel confirmed the presence of a single band, then the PCR product was purified using a Qiagen PCR purification kit. Otherwise, a Qiagen gel extraction kit was used for purification.

In order to facilitate cloning, purified PCR products and the target plasmid were subjected to restriction digest. All restriction enzymes used in this work were purchased from NEB and were active in Cutsmart buffer. In general, each restriction digest comprised ~1 μg PCR product/plasmid, 5 μL 10X cutsmart buffer and 1 μL each of the required restriction enzyme (for plasmid DNA, 1 μL Quick-CIP (NEB) was also added). The reaction was made up to 50 μL with distilled water. Restriction digests were incubated for 1–2 hours at 37 °C. Following digestion, the digested DNA was purified using a Qiagen PCR purification kit.

Ligation reactions for digested plasmids/inserts were performed using T4 DNA ligase (NEB). Where possible, ligation reactions were performed at 3:1, 5:1 and 7:1 ratios (insert:plasmid). The NEB ligation calculator was used to calculate the required amount of DNA for each ratio. Each ligation reaction contained: 70 ng plasmid DNA, the calculated amount of insert DNA, 2 μL 10X T4 DNA ligase buffer and 1 μL T4 DNA ligase. The final solution was made up to 20 μL with distilled water. Ligation reactions were incubated for 1–2 hours at room temperature or overnight at 16 °C before being transformed into chemically competent XL10 gold cells (Agilent).

2.1.1.2 Restriction-free cloning

Restriction-free (RF) cloning allows a target gene to be cloned into a plasmid in a two-step process that does not require the use of restriction enzymes[159]. The first PCR reaction was performed as described for restriction cloning. The product of the first PCR reaction is a mega-primer that contains overlaps that are complementary to the target plasmid.

The second PCR reaction was performed with a molar mega-primer:plasmid ratio of 20:1. The second-round PCR mixture contained: 4 μL 5X Phusion HF buffer, 0.4 μL 10 mM dNTPs, 0.2 μL Phusion DNA Polymerase and the required amounts of mega-primer and plasmid. The final solution was made up to 20 μL with distilled water.

The thermocycling conditions used for the second-round PCR are shown in Table 2.2. Again, for each PCR, reactions were performed at the predicted Tm and 5 °C above and below the Tm. To digest template (methylated) DNA, the products of second-round PCR were incubated with 1 μL DpnI (NEB) for one hour at 37 °C. The post-DpnI digest sample was purified with a Qiagen PCR purification kit and transformed into chemically competent XL10-Gold cells.

Table 2.2: Thermocycling conditions for second-round PCR

Step		Temperature (°C)	Time
Initial Denaturation		98	30 s
30 Cycles	Denaturation	98	8 s
	Annealing	melting temperature (Tm) ± 5	20 s
	Elongation	72	30 s per kb
Final extension		72	5 minutes
Hold		4	

2.1.1.3 Site-directed mutagenesis

All site-directed mutagenesis (SDM) performed in this work followed the recommended protocol for Q5 high-fidelity DNA Polymerase (NEB). The NEBaseChanger tool was used to design primers suitable for the required SDM. Each Q5 reaction contained: 12.5 μL Q5 High-Fidelity 2X Master Mix, 1.25 μL 10 μM Forward Primer, 1.25 μL 10 μM Reverse Primer and 0.5 μL template DNA (~5 ng). The final solution was made up to 25 μL with distilled water.

The thermocycling conditions for Q5 can be seen in Table 2.3. For each PCR, reactions were performed at the predicted Tm and 5 °C above and below the Tm. The PCR product of the Q5 reaction was ran on a 1% agarose gel as previously described. If an appropriately sized band was observed, the PCR product was used in a Kinase, Ligase, DpnI (KLD) reaction. Each KLD reaction contained: 1 μL Q5 PCR product, 1 μL 10 X T4 Ligase buffer, 0.75 μL Polynucleotide Kinase (NEB), 0.75 μL T4 DNA Ligase, 0.5 μL DpnI and 6 μL distilled water. KLD reactions were left to incubate at room temperature for 1–2 hours before being transformed into chemically competent XL10-Gold cells.

Table 2.3: Thermocycling conditions for Q5 SDM

Step		Temperature (°C)	Time
Initial Denaturation		98	30 s
35 Cycles	Denaturation	98	10 s
	Annealing	melting temperature (Tm) ± 5	30 s
	Elongation	72	30 s per kb
Final extension		72	2 minutes
Hold		4	

2.1.1.4 Transformation

During transformation, 50–100 ng plasmid DNA was added to 50 μL competent XI10 gold cells which were left to incubate on ice for 10 minutes. Cells were heat shocked for 45 seconds at 42 °C and subsequently allowed to recover for 10 minutes on ice. Following this, 950 μL sterile LB was added to the cells which were subsequently incubated at 37 °C for one hour in an incubator shaking at 180 RPM. Cells were pelleted by centrifugation at 16,200 x g. Pellets were resuspended in 50 μL LB before being spread onto LB/agar plates supplemented with the required antibiotic. Plates were incubated overnight at 37 °C.

2.1.1.5 Colony PCR

Colony PCR was used to assess the success of restriction/RF cloning and was performed following manufactures instructions for MyTaq red mix. Colonies were picked with a sterile wire loop/pipette tip and dipped into 25 μL colony PCR reaction mixture containing 12.5 μL MyTaq Red mastermix, 1 μL 10 μM forward primer and 1 μL10 μM reverse primer. The thermocycling conditions for colony PCR can be seen in Table 2.4. The products of the colony PCR were run on a 1% agarose gel as previously described.

Table 2.4: Thermocycling conditions for colony PCR

Step		Temperature (°C)	Time
Initial Denaturation		98	10 mins
35 Cycles	Denaturation Annealing Elongation	98 55 72	15 s 20 s 30 s per kb
Hold		4	∞

2.1.1.6 Sequencing

Following successful colony PCR (or SDM), the desired colonies were used to inoculate 10 mL LB supplemented with the appropriate antibiotic. Cultures were grown overnight at 37 °C in an incubator shaking at 180 RPM. The cells were harvested by centrifugation at 3,220 x g. Recombinant plasmids were purified using the Qiagen spin miniprep kit, following manufacturer's instructions. Purified plasmid DNA was concentrated to ~100 ng/μL and sent for Sanger sequencing at SourceBiosciences using appropriate primers. Constructs that were confirmed to contain the correct insert sequence were transformed into chemically competent BL21-CodonPus (DE3)-RIPL cells (Agilent) for protein expression.

2.1.1.7 Gene synthesis

Several constructs used in this study were synthesised by Twist Bioscience. Constructs were designed using the online tool available on the Twist Bioscience website (https://ecommerce.twistdna.com/app/gene.com).

2.1.2 Cell culture

2.1.2.1 Culture Media

LB – 1 L (10 g Tryptone, 5 g Yeast extract, 5 g NaCl)

TB – 1 L (12 g Tryptone, 24 g Yeast extract, 9.4 g K_2HPO_4, 2.2 g KH_2PO_4, 4 mL glycerol)

AI – 1L (10 g Tryptone, 5 g Yeast extract, 3.55 g Na_2HPO_4, 3.4 g KH_2PO_4, 2.68 g NH_4Cl) + 20 mL sugar mix

AI Sugar mixes

Sugar mix A – 1L (95 g lactose, 250 mL glycerol, 25 g glucose)

Sugar mix B – 1L (5 g lactose, 100 mL glycerol, 2.5 glucose)

Sugar mix C – 1L (180 g lactose, 150 mL glycerol, 2.5 g glucose)

2.1.2.2 Antibiotic Stocks

Ampicillin – 100 mg/mL (in 50% ethanol)

Kanamycin – 100 mg/mL (in H2O)

Chloramphenicol – 25 mg/mL (in 70% ethanol)

Antibiotic stock solution was diluted 1 in 1000 for the working concentration of antibiotics used in supplemented media.

2.1.2.3 Expression Trials

A 10 mL LB overnight culture, supplemented with the appropriate antibiotic, was inoculated with a single colony. Overnight cultures were grown overnight at 37 °C in an incubator shaking at 180 RPM. To make glycerol stocks, 500 μL overnight culture was mixed with 500 μL sterile 50% glycerol. Glycerol stocks were stored at -80 °C.

Expression trials were carried out in LB, TB, and AI. For each trial, 10 mL of each media, supplemented with the appropriate antibiotic, was inoculated with 100 μL of overnight culture. During AI expression trials, 200 μL of sugar mix was added. Expression trial cultures were grown at 37 °C in an incubator shaking at 180 RPM until reaching an OD600 of ~0.6. As a pre-induction sample for SDS-PAGE analysis, 1 mL of the pre-induction culture was kept. Upon reaching OD, cultures were induced with

0.5 mM IPTG and incubated overnight at 18 °C in an incubator shaking at 180 RPM. Following overnight expression, 1 mL cell culture was pelleted by centrifugation at 16,200 x g for 10 minutes. Cells were resuspended in 500 μL resuspension buffer (50 mM Hepes pH 7.5, 250 mM NaCl), lysed by sonication (Fisher sonic dismembranator) (1 x 15 second pulse) and 100 μL sonicated cell solution was kept for SDS-PAGE analysis. Insoluble cell material was removed by centrifugation for 20 minutes at 16,200 x g and 100 μL cell lysate was taken for SDS-PAGE analysis.

2.1.2.4 SDS-PAGE

Protein expression levels and purity were assessed by SDS-PAGE. Throughout this thesis, 15% and 20% hand-cast SDS-PAGE gels were made using the recipes outlined in Table 2.5. Gels were cast using Bio-Rad Mini-Protean casting frames. In general, 15% gels were used for proteins over 25 KDa whilst 20% gels were used for proteins below 25 KDa.

Table 2.5: Recipes for making 8 SDS PAGE gels

Reagent	15% Gel	20% Gel	Stacking Gel
30% Acrylamide	16 mL	21.3 mL	4 mL
1.5 M Tris pH 8.8	8mL	8 mL	–
1M Tris pH 6.8	–	–	5 mL
10% SDS	320 μL	320 μL	200 μL
10% APS	320 μL	320 μL	200 μL
TEMED	32 μL	32 μL	20 μL
H_2O	7.3 mL	2 mL	10.6 mL

Samples for SDS-PAGE were diluted 5:1 in 5X SDS-PAGE loading dye (250 mM Tris pH 6.8, 10% SDS, 0.5% Bromophenol blue, 30% glycerol, 5% β-mercaptoethanol) and boiled for 10 minutes at 98 °C. Around 5–10 μL of sample was loaded to the gel. Gels were run at 200V for ~1 hour in SDS-PAGE running buffer (25 mM Tris pH 8.8, 200 mM glycine, 0.1% SDS). For reference, 3 μL Colour Prestained Protein Standard, Broad Range ladder (NEB) was loaded as a molecular weight marker. Gels were stained by boiling in Coomassie stain (0.4% acectic acid, 0.1 % Coomassie Brilliant Blue G-250) for one minute and subsequently destained by boiling in water for three minutes.

2.1.2.5 Largescale expression protocol

The expression protocol for all constructs used in this thesis followed the same general protocol. A 10 mL LB overnight culture, supplemented with the appropriate antibiotic, was inoculated with a single colony or a small quantity of glycerol stock. Overnight cultures were incubated overnight at 37 °C in an incubator shaking at 180 RPM. To inoculate 1L of growth media, 10 mL overnight culture was used, supplemented with the appropriate antibiotic. The media used for each construct was determined based on the results of the expression trial. Cell cultures were grown until they reached an OD600 of 0.6, at which point they were induced with 0.5 mM IPTG. Induced cultures were incubated overnight at 18 °C in an incubator shaking at 180 RPM. Cells were harvested by centrifugation at 8,980 x g for 20 minutes in a JLA-8.100 rotor. Cell pellets were stored in 50 mL Greiner tubes at -20 °C.

2.1.3 Protein Purification

2.1.3.1 Immobilised metal affinity chromatography

Cell pellets (from 1L of cell culture) were thawed and resuspended in 30–40 mL lysis buffer (50 mM Hepes pH 7.5, 250 mM NaCl, 1mM PMSF, 1 mM Benzamidine). Cells were lysed via sonication with the amplitude set to 40%. Cells were sonicated in six 20-second pulses, with each pulse separated by 40 seconds of recovery on ice. Lysed cells were centrifuged at 48298 xg at 4 °C for one hour in a JLA 25.5 rotor. Immobilised metal affinity chromatography (IMAC) purification was performed on a 5mL HisTrap HP column. Prior to loading the cell lysate, the HisTrap was stripped with 5 column volumes (CV) of 100 mM EDTA, washed with 5 CV of 0.5 M NaOH and recharged with 5 CV of 100 mM $NiSO_4$. The recharged column was equilibrated in 5 CV of buffer A (50 mM Hepes pH 7.5, 250 mM NaCl, 20 mM Imidazole). The cell lysate was loaded onto the HisTrap using a peristaltic pump. The column was subsequently connected to an ÄKTA pure system and was washed in buffer A until the UV Absorbance at 280 nm dropped below 100 mAU. A linear gradient from 0–100% of buffer B (50 mM Hepes pH 7.5, 250 mM NaCl, 500 mM Imidazole) across 30 CV was used to elute the protein of interest. Eluate was collected in 3 mL fractions and the peak fractions was assessed by SDS-PAGE. Fractions containing the protein of interest were pooled and concentrated using Amicon Ultra-15 spin concentrators with an appropriate molecular weight cut-off (MWCO). Protein concentration was determined using a DeNovix DS-11 nanodrop.

2.1.3.2 Size exclusion chromatography

Size exclusion chromatography (SEC) was performed using Superdex gel filtration columns. In general, for proteins above 40 KDa, a HiLoad 26/600 Superdex 200 prep grade column was used. For proteins below 40 KDa, a HiLoad 26/600 Superdex 75 prep grade column was used. For analytical SEC experiments a HiLoad 10/300 superdex 75 analytical column was used. The appropriate column was connected to an ÄKTA pure system and was equilibrated in 1.2 CV SEC buffer (20 mM Hepes pH 7.5, 150 mM NaCl). Samples were loaded onto the equilibrated column through a 5 mL loop and were eluted at a flow rate of 2 mL/min. Eluate was collected in 3 mL fractions and the peak fractions were assessed by SDS-PAGE.

2.1.4 Crystallisation and diffraction experiments

Crystallisation was performed by sitting/hanging-drop vapour diffusion. In general, initial sparse matrix screening was performed using a Mosquito (SPT Labtech) and SWISSCI 96-well 3-drop LP UVP sitting drop plates. Throughout this thesis, screening was performed using the following commercial screens: BCS, JCSG+, Memgold, Memgold2, MIDAS+, Morpheus, PACT Premier, Proplex and Wizard. When setting up the 96-well plate, 35 μL of each condition was added to the reservoir. Following this, 400 nl of protein was mixed with 400 nl of reservoir solution and the plates were sealed with crystal clear tape and stored at 18 °C.

Seeding was used to optimise initial crystal hits. To make seed stocks, crystals were resuspended in 50 μL of reservoir solution and added to a micro-seed bead tube (Hampton research). The tube was vortexed for five minutes in cycles of 30 seconds

of mixing and 30 seconds of recovery on ice. Generally, 1/10, 1/100 and 1/1000 dilutions of the seed stock were made. These seed stocks were either added directly to the drop during sparse matrix screening or used for streak seeding.

Crystals to be used in x-ray diffraction experiments were soaked in cryoprotectant before being harvested from the drop with a litholoop and cryocooled in liquid nitrogen. Information about protein-specific cryoprotectants and data collection parameters is provided later in this chapter.

2.1.5 AlphaFold modelling

All AlphaFold (AF)[160] *ab initio* structure prediction presented in this work was done using the ColabFold server. Unless otherwise specified, default parameters were used. Throughout this thesis, the confidence of AF models is presented as the average pLDDT score for a model; this score ranges from 0–100 (low-high) confidence.

2.2 Methods for Chapter 3 (Structural characterisation of the *B. bacteriovorus* gliding machinery)

2.2.1 Cloning, expression and purification

A summary of the constructs used in Chapter 3 can be seen in Table 2.6. The GltD1 construct used in this thesis was cloned by Dr Amber Wilson prior to this study. The GltE1 construct was cloned by restriction cloning into pET26b using the NcoI and XhoI restriction sites. GltE1 was located directly downstream of a PelB leader sequence for secretion of GltE1 to the periplasm. The $GltG1_{491–556}$ and GltC1 constructs were

synthesised by Twist Bioscience. All constructs used in this chapter were expressed and purified using the general protocols previously described.

To obtain a pure, stoichiometric GltD1:GltE1 complex for structural studies, a five-fold molar excess of GltE1 was added to GltD1 was allowed to incubate on ice for one hour. The GltD1:GltE1 complex was then co-eluted from a HiLoad 26/600 Superdex 200 prep grade column.

Table 2.6: Information for constructs used in Chapter 3

Construct	Cloning Method	Residues	Plasmid	Purification Tag	Cleavage site	PelB Leader sequence	Antibiotic resistance	Forward Primer	Reverse Primer
GltD1	RF cloning	25–1066	pET28a	N-terminal 6x His	No	No	Kanamycin	-	-
$GltD1_{Cterm}$	SDM	706–1066	pET28a	N-terminal 6x His	No	No	Kanamycin	AATCTGGCAGT CAATGCG	GCCGCTGCT GTGATGATG
$GltD1_{A762D}$	SDM	25–1066	pET28a	N-terminal 6x His	No	No	Kanamycin	CAGGATTC TGATCAGT TCGAGGAA	GTAAAGCTTC GCCAGCAA
$GltD1_{\Delta 1043-1066}$	SDM	25–1042	pET28a	N-terminal 6x His	No	No	Kanamycin	TAAGAATTC GAGCTCCGTC	ATCCTTGCTG TTCATGTAC
GltE1	Restriction cloning	17–281	pET26b	C-terminal 6x His	TEV	Yes	Kanamycin	CGTCATATGTC TTCCAGTCCAA CCAAAG	CGTGCTCGA GGCCCTGAA AATACAGGTT TTCTTTTATC CCAGCTTTCG C
$GltE1_{N200R}$	SDM	17–281	pET26b	C-terminal 6x His	TEV	Yes	Kanamycin	CGTTTTGAACC GTTATGGTATT GCACTGACGG C	CGAGGATCA CGCACACCC
$GltG1_{491-556}$	Gene synthesis	491–556	pET29b	N-terminal MBP	TEV	No	Kanamycin	-	-
GltC1	Gene synthesis	22–529	pET29b	N-terminal 6x His	No	Yes	Kanamycin	-	-
$GltC1_{\Delta 480-529}$	SDM	22–479	pET29b	N-terminal 6x His	No	Yes	Kanamycin	TAAGGAAAAGA TATATCGGGAG	AGCGATTTTC TTTTTAATTGA TTC

2.2.2 GltD1:GltE1 ITC

ITC was performed using a MicroCal PEAQ-ITC. To avoid buffer mismatch, GltD1 and GltE1 were simultaneously dialysed into assay buffer (20 mM Hepes pH 7.5, 150 mM NaCl). In all GltD1:GltE1 ITC experiments, the sample cell was loaded with 25 μM GltD1 and the syringe was loaded with 250 μM GltE1. GltE1 was titrated into the cell across 19 injections; the initial 0.4 μL injection was followed by 18 injections of 2 μL. The full ITC parameters can be seen in Table 2.7. A control experiment of GltE1 titrated into assay buffer was also performed. Data analysis and control subtraction was performed using the MicroCal PEAQ-ITC analysis software.

Table 2.7: Parameters used for GltD1:GltE1 ITC experiments

ITC Parameters – GltD1:GltE1	
Temperature (°C)	25
Reference power (μcal/s)	10
Stir speed (RPM)	750
Initial delay	60
Number of injections	19
Injection volume (μL) (initial injection)	2 (0.4)
Injection duration (s)	4
Injection spacing (s)	150

2.2.3 GltD1:GltE1 Circular Dichroism

Circular Dichroism (CD) spectra were generated to assess the folding state of $GltD1_{\Delta 1043-1066}$. In preparation for CD experiments, GltD1 and $GltD1_{\Delta 1043-1066}$ were dialysed overnight into 20 mM Sodium Phosphate pH 7.5. Following dialysis, GltD1 and $GltD1_{\Delta 1043-1066}$ were diluted to 0.3 mg/mL. CD was kindly performed by Lauren Garbett using a JASCO J-1500 CD Spectrophotometer.

2.2.4 GltD1:GltE1 Crystallisation and data collection

GltE1 was concentrated to 24 mg/mL and used in sparse matrix crystallisation trials. Two crystal forms of GltE1 were grown by sitting-drop vapour diffusion. Crystal Form 1 was grown in Memgold2 H10 (2.8 M Ammonium chloride, 0.075 M HEPES pH 7.5) and Crystal Form 2 was grown in Memgold2 H7 (0.1 M Tris pH 8.5 44 % v/v PEG 200). Crystal Form 1 was cryoprotected in the reservoir solution + 25% ethylene glycol, mounted on a litholoop and cryocooled in liquid nitrogen. As Crystal Form 2 was grown in a cryoprotective condition, it was harvested and cryocooled directly from the reservoir solution. Data collection for GltE1 crystals was performed at Diamond Light Source on beamline IO4. Datasets were collected with an x-ray wavelength of 0.9795 Å and transmission of 40%. The crystals were rotated 360° in oscillations of 0.1°, producing a dataset with 3600 images. Data for both crystal forms were processed using the Xia2-dials automated pipeline[161,162]. Initial phases for both crystal form 1 and 2 were estimated by molecular replacement (MR) (using PhaserMR[163]) using the AF model of GltE1 as a search model (further information provided in sections 3.2.4.1 and 3.2.4.4).

The $GltD1_{Cterm}$:GltE1 complex was concentrated to 50 mg/mL and used to set up crystallisation trials. $GltD1_{Cterm}$-GltE1 crystals were grown in Memgold2 E11 (0.25 M Magnesium formate dihydrate 0.1 M Sodium cacodylate pH 6.5 22 % w/v PEG 3000). Crystals were cryoprotected in the reservoir solution + 25% glycerol, mounted on a litholoop and cryocooled in liquid nitrogen. Data collection for $GltD1_{Cterm}$:GltE1 crystals was performed at Diamond Light Source on beamline IO4. Datasets were collected with an X-ray wavelength of 0.6888 Å and transmission of 25%. The crystals were rotated 360° in oscillations of 0.1°, producing a dataset with 3600 images. Diffraction

data were processed using the autoproc-staraniso automated pipeline[164,165]. Initial phases for GltD1$_{Cterm}$:GltE1 were estimated by MR (using PhaserMR[163]) using the crystal structure of GltE1 (Crystal Form 1) and the AF model of GltD1 as search models (further information provided in section 3.2.5).

2.2.5 GltD1:GltE1 Cryo-electron microscopy

Cryo-electron microscopy (cryo-EM) grids were prepared at the EM facility at the Institute for Structural and Chemical Biology (University of Leicester). Quantifoil R1.2/1.3 gold holey carbon grids were glow discharged for 15 seconds at 30 mA using a Quorum GloQube. In order to freeze grids, 3 μL GltD1:GltE1 (0.1–1 mg/mL) was applied to glow discharged grids. Grids were blotted for three seconds with the blot force set to 10 and were plunge-frozen in liquid ethane using an FEI Vitrobot (4 °C, 100% humidity). Grids were screened using a Titan Krios G3.

2.2.6 Purification of MBP-GltG1$_{491–556}$

Cell lysate was prepared as described in section 2.1.3.1. MBP-GltG1$_{491–556}$ was purified using a 5 mL MBPTrap. The MBPtrap was equilibrated in 5 CV of binding buffer (50 mM Hepes pH 7.5, 250 mM NaCl, 1 mM EDTA). Cell lysate was loaded onto the MBPTrap using a peristaltic pump flowing at ~2 mL/min. The MBPTrap was then washed with 5 CV of binding buffer. MBP-GltG1$_{491–556}$ was eluted with 5 CV of elution buffer (50 mM Hepes pH 7.5, 250 mM NaCl, 1 mM EDTA, 10 mM Maltose). Eluate was collected in 2 mL fractions and MBP-GltG1$_{491–556}$-containing fractions were identified using a DeNovix DS-11 nanodrop and subsequent SDS-PAGE analysis.

2.2.7 GltC1 inclusion body and refold

To promote inclusion body formation, cells were grown to an OD600 of 0.6 and GltC1 expression was induced with 2 mM IPTG. GltC1 was expressed for four hours at 37 °C in an incubator shaking at 180 RPM. Cells were harvested by centrifugation at 3220 x g. Cell pellets were resuspended in buffer C (50 mM Hepes pH 7.5, 250 mM NaCl, 1% Triton X-100, 1 mM EDTA, 20 mM DTT). Cells were lysed by sonication (as described in section 2.1.3.1) and subsequently centrifuged at 16,200 x g for 30 minutes. The supernatant was decanted, and the inclusion pellet was washed in buffer C and sonicated on ice for 15 seconds to aid resuspension. The resuspended inclusion pellet was centrifuged at 16,200 x g for 30 minutes. In general, this wash step was repeated 3–5 times until a white-coloured inclusion pellet was obtained. The inclusion pellet was then resuspended in buffer D (50 mM Hepes pH 7.5, 1 M NaCl, 1% Triton X-100, 1 mM EDTA) and centrifuged at 16,200 x g or 30 minutes. Following this final wash step, the inclusion pellet was resuspended in denaturing buffer (50 mM Hepes pH 7.5, 250 mM NaCl, 8M Urea). GltC1 was refolded by shock dilution, where the solubilised inclusion pellet was diluted 15-fold in assay buffer; this refolding solution was left to cycle overnight at 4 °C.

2.3 Methods for Chapter 4 (Structural characterisation of proteins associated with the regulation of gliding)

2.3.1 Cloning, expression and purification

A summary of the constructs used in Chapter 4 can be seen in Table 2.8 All constructs used in this chapter were expressed and purified using the general protocols previously described.

Table 2.8: Information for constructs used in Chapter 4

Construct	Cloning Method	Residues	Plasmid	Purification Tag	Cleavage site	PelB Leader sequence	Antibiotic resistance	Forward Primer	Reverse Primer
Bd1996	Restriction cloning	1–227	pCOLD1	N-terminal 6x His	TEV	No	Ampicillin	CGTCATATGGAAAACCTGTATTTTCAGGGCATGAATCCAAAAAACTGGTTCG	CGTCTCGAGTTAAACCATGGCAAACTTGC
Bd1996-PilZ	Restriction cloning	126–227	pET26b	C-terminal 6x His	No	No	Kanamycin	CGTCATATGAGTCGCCGGTCCCACCCC	CGTCTCGAGAACCATGGCAAACTTGCGCACG
Bd1996-GYF	Restriction cloning	1–117	pET26b	C-terminal 6x His	No	No	Kanamycin	CGTCATATGAATCCAAAAAACTGGTTC	CGTCTCGAGGGGAATGGAATAGATTTCC
$MglA_{Bd:}$Bd2492	Gene synthesis	$MglA_{Bd:}$: 1–197 Bd2492: 205-353	pET29b	N-terminal 6x His (Bd2492)	No	No	Kanamycin	-	-
$MglA_{BdT54A:}$Bd2492	SDM	$MglA_{Bd:}$: 1–197 Bd2492: 205-353	pET29b	N-terminal 6x His (Bd2492)	No	No	Kanamycin	TATAGAACGTGCGCTTTTCTTTGATTTC	TCTGTATTCAGCGCTAC

2.3.2 Bd1996 and $MglA_{Bd}$:Bd2492 Crystallisation and data collection

For crystallisation trials, Bd1996-PilZ was concentrated to 11mg/mL. Bd1996-PilZ crystals were grown in Morpheus G2 (0.2M Ammonium formate, 25% PPGBA 400). Bd1996-PilZ crystals were grown in a cryoprotective condition and were harvested and cryocooled directly from the reservoir solution. Data collection for Bd1996-PilZ crystals was performed at Diamond Light Source on beamline IO4. Datasets were collected with an x-ray wavelength of 0.8989 Å and transmission of 100%. The crystals were rotated 360° in oscillations of 0.1°, producing a dataset with 3,600 images. Diffraction data were processed using the autoproc-staraniso automated pipeline[164,165]. Initial phases for Bd1996-PilZ were estimated by MR (using PhaserMR[163]) using the PilZ domain of Bd1996-PM as the search model (further information provided in section 4.2.2).

Bd1996-PilZ crystals were used to create a seed stock to facilitate sparse matrix seed screening against full-length Bd1996 (15 mg/mL). The seed stock was made as described in section 2.1.4. Initial hits were grown in Proplex B9 (0.1 M Sodium cacodylate pH 6.0, 15 % PEG 4000). These hits were used to make a second seed stock for a second-round of sparse matrix screening. Bd1996 crystals were grown from the second-round of seeding in 0.1 M Sodium cacodylate pH 6.0, 15 % PEG 4000. Bd1996 crystals were cryoprotected in the reservoir solution + 25% ethylene glycol. Data collection for Bd1996 crystals was performed at Diamond Light Source on beamline IO4. Datasets were collected with an x-ray wavelength of 0.9795 Å and transmission of 60%. The crystals were rotated 360° in oscillations of 0.1°, producing

a dataset with 3,600 images. Diffraction data were processed using the FastDP automated pipeline[166].

For crystallisation trials, $MglA_{Bd:}$Bd2492 was concentrated to 30 mg/mL. $MglA_{Bd:}$Bd2492 crystals were grown in Memgold2 H1 (0.05 M Tris 8.5 28 % v/v PEG 400). $MglA_{Bd:}$Bd2492 crystals grew in a cryoprotective condition and were harvested and cryocooled directly from the reservoir solution. Data collection for $MglA_{Bd:}$Bd2492 crystals was performed at The European Synchrotron Radiation Facility (ESRF) on beamline ID30A-3. Datasets were collected with an x-ray wavelength of 0.9677 Å and transmission of 2.5%. The crystals were rotated 360° in oscillations of 0.1°, producing a dataset with 3600 images. Diffraction data were processed using the Grenades_parallelproc automated pipeline[167]. Initial phases for $MglA_{Bd}$ were estimated by MR (using PhaserMR[163]) using the AF model of $MglA_{Bd}$ as the search model (further information provided in section 4.2.13).

2.3.3 Bd1996 Pull-down assays

In preparation for pull-down assays, 5 mg Bd1996, Bd1996-GYF and BSA were dialysed overnight into coupling buffer (0.1 M $NaHCO_3$ pH 8.3, 0.5 M NaCl). A total of 0.3 g of cyanogen-bromide activated Sepharose resin was added to a 15 mL Greiner tube and washed with 14 mL ice cold 1 mM HCl for 10 minutes. The resin was pelleted by centrifugation (1 minute, 500 x g) and subsequently washed in 5 mL water followed by three 5 mL washes of coupling buffer. The resin was centrifuged (1 minute, 500 x g) between each wash step. Following washing, the total resin volume was split three ways into separate 15 mL Greiner tubes. To permit binding, 2 mg of Bd1996, Bd1996-GYF and BSA was added to the resin. Protein-resin mixtures were left gently rotating

for two hours at room temperature. Each resin was then washed with three 5 mL washes of coupling buffer for five minutes. The resin was blocked overnight in 1 M ethanolamine at 4°C. The resin was then washed for five minutes in 5 mL coupling buffer, pelleted, and then washed in 5 mL wash buffer (0.1 M acetate buffer pH 4, 0.5 M NaCl) for 10 minutes. This wash cycle was repeated three times. Finally, each resin was equilibrated in 5 mL RIPA buffer (20 mM Tris-HCl pH 8.0, 1mM EDTA, 0.5 mM EGTA, 1% Triton X-100, 0.1% sodium deoxycholate, 0.1% SDS, 140 mM NaCl, 1 mM PMSF.

Pellets from 3L of *B. bacteriovorus* HD100 cells were resuspended in 15 mL ice-cold RIPA buffer. Cells were sonicated on ice for nine 20-second cycles followed by three x 30-second cycles, with two-minute rests between each cycle. Lysed cells were centrifuged at 4 °C for one hour at 48,298 x g The lysate was collected and filtered through 0.45 μm filters. Subsequently, 5 mL *B. bacteriovorus* lysate was added to the resin and incubated at 4 °C for two hours. The resin was pelleted, resuspended in 1 mL 100 mM ammonium bicarbonate pH 8.0, and transferred to 1.5 mL Eppendorf tubes. The resin was washed three times in 1 mL 100 mM ammonium bicarbonate pH 8.0. Following the final wash step, 150 μL of resin was taken for SDS-PAGE analysis. The SDS PAGE sample was centrifuged for one minute at 500 x g and the pellet was resuspended in 50 μL 5X loading dye. Gel samples were boiled at 95 °C for five minutes and centrifuged for one minute at 500 x g. For SDS-PAGE analysis, a 10 μL sample was run on a 4–20% Mini-PROTEAN precast protein gel. Following the identification of unique bands in the Bd1996 and Bd1996-GYF samples, the Bd1996, Bd1996-GYF and BSA resins were sent to the Advanced Mass Spectrometry Facility

at the University of Birmingham for on-bead digestion followed by subsequent MS analysis.

2.3.4 Bd1996:c-di-GMP/cGAMP ITC

ITC was performed using a MicroCal PEAQ-ITC. Bd1996 was dialysed overnight into assay buffer (20 mM Hepes pH 7.5, 150 mM NaCl). C-di-GMP and cGAMP, purchased from Bio-Techne, were resuspended in dialysis buffer to concentrations of 400 μM and 1 mM respectively. For Bd1996:c-di-GMP experiments, the sample cell was loaded with 10 μM Bd1996 and the syringe was loaded with 400 μM c-di-GMP. C-di-GMP was titrated into the cell across 24 injections; the initial 0.4 μL injection was followed by 23 injections of 1.5 μL. For Bd1996:cGAMP experiments, the sample cell was loaded with 50 μM Bd1996 and the syringe was loaded with 1 mM cGAMP. cGAMP was titrated into the cell across 19 injections; the initial 0.4 μL injection was followed by 18 injections of 2 μL. The full ITC parameters can be seen in Tables 2.9 and 2.10. Control experiments of c-di-GMP/cGAMP titrated into assay buffer were also performed. Data analysis and control subtraction was performed using the MicroCal PEAQ-ITC analysis software.

Table 2.9: Parameters used for Bd1996:c-di-GMP ITC experiments

ITC Parameters – Bd1996:c-di-GMP	
Temperature (°C)	25
Reference power (μcal/s)	10
Stir speed (RPM)	750
Initial delay	60
Number of injections	14
Injection volume (μL) (initial injection)	1.5 (0.4)
Injection duration (s)	4
Injection spacing (s)	150

Table 2.10: Parameters used for Bd1996:cGAMP ITC experiments

ITC Parameters - Bd1996:cGAMP	
Temperature (°C)	25
Reference power (μcal/s)	10
Stir speed (RPM)	750
Initial delay	60
Number of injections	19
Injection volume (μL) (initial injection)	2 (0.4)
Injection duration (s)	4
Injection spacing (s)	150

2.3.5 Phage display

Phage display for Bd1996 was performed by an industry collaborator using a linear 15mer library.

2.4 Methods for Chapter 5 (Structural characterisation of the nuclease Bd1431)

2.4.1 Cloning, expression and purification

A summary of the constructs used in Chapter 5 can be seen in Table 2.11. All constructs used in this chapter were expressed and purified using the general protocols previously described.

Table 2.11: Information for constructs used in Chapter 5

Construct	Cloning Method	Residues	Plasmid	Purification Tag	Cleavage site	PelB Leader sequence	Antibiotic resistance	Forward Primer	Reverse Primer
Bd1431	RF cloning	18–166	pET26b	N-terminal 6x His	No	Yes	Kanamycin	GCCCAG CCGGCG ATGGCC ATGTCTG AACGTCT TCAAGC CC	GGTGGT GGTGGT GGTGCT CGAGGG TGCCGC CATCATA GGC
$Bd1431_{F120D}$	SDM	18–166	pET26b	N-terminal 6x His	No	Yes	Kanamycin	CGACAA ATACGAT CGCATTC TGGCG	CGCTGA ACATTGT GCAGT

2.4.2 Bd1431 Crystallisation and data collection

For crystallisation trials, Bd1431 was concentrated to 17 mg/mL. Initial hits were grown in JCSG E1 (1 M Sodium Citrate, 0.1 M Sodium Cacodylate pH 6.5) by hangingdrop vapour diffusion. Streak seeding was used to optimise initial hits. The seed stock was made as described in section 2.1.4. Streak seeding was performed in a 24-well optimisation tray. The screen conditions for Bd1431 optimisation are outlined in Table 2.12. To perform streak seeding, 100 μL Bd1431 (12 mg/mL) was mixed with 100 μL reservoir solution on plastic cover slip. Manual streaks were performed sequentially through each well of a column of the screen. A cat whisker was dipped in the seed stock and passed through the four drops of a column in a single movement.

Table 2.12: Crystal screen used for Bd1431 streak seeding

	1	2	3	4	5	6
A	1M Na citrate 0.1 M Na cacodylate pH 6.3	1M Na citrate 0.1 M Na cacodylate pH 6.5	1M Na citrate 0.1 M Na cacodylate pH 6.8	0.75 M Na citrate 0.1 M Na cacodylate pH 6.5	1 M Na citrate 0.1 M Na cacodylate pH 6.5	1.25 M Na citrate 0.1 M Na cacodylate pH 6.5
B	1M Na citrate 0.1 M Na cacodylate pH 6.3	1M Na citrate 0.1 M Na cacodylate pH 6.5	1M Na citrate 0.1 M Na cacodylate pH 6.8	0.75 M Na citrate 0.1 M Na cacodylate pH 6.5	1 M Na citrate 0.1 M Na cacodylate pH 6.5	1.25 M Na citrate 0.1 M Na cacodylate pH 6.5
C	1M Na citrate 0.1 M Na cacodylate pH 6.3	1M Na citrate 0.1 M Na cacodylate pH 6.5	1M Na citrate 0.1 M Na cacodylate pH 6.8	0.75 M Na citrate 0.1 M Na cacodylate pH 6.5	1 M Na citrate 0.1 M Na cacodylate pH 6.5	1.25 M Na citrate 0.1 M Na cacodylate pH 6.5
D	1M Na citrate 0.1 M Na cacodylate pH 6.3	1M Na citrate 0.1 M Na cacodylate pH 6.5	1M Na citrate 0.1 M Na cacodylate pH 6.8	0.75 M Na citrate 0.1 M Na cacodylate pH 6.5	1 M Na citrate 0.1 M Na cacodylate pH 6.5	1.25 M Na citrate 0.1 M Na cacodylate pH 6.5

Streak seeding was successful and optimised crystals were grown in 1 M Na citrate 0.1 M Na cacodylate pH 6.3. Optimised Bd1431 crystals were cryoprotected in the reservoir solution + 25% glycerol, mounted on a litholoop and cryocooled in liquid nitrogen. Data collection for Bd1431 crystals was performed at Diamond Light Source on beamline IO3. Datasets were collected with an x-ray wavelength of 0.8989 Å and transmission of 100%. The crystals were rotated 360° in oscillations of 0.1°, producing a dataset with 3600 images. Diffraction data were processed using the Xia2-dials automated pipeline[161,162]. Initial phases for Bd1431 were estimated by MR (using PhaserMR[163]) using the crystal structure of Staphylococcal nuclease (PDB: 1ey0) as the search model (further information provided in section 5.2.3).

For crystallisation trials, $Bd1431_{F120D}$ was concentrated to 22mg/mL. $Bd1431_{F120D}$ crystals were grown in JCSG G1 (0.3 M Magnesium formate dihydrate, 0.1 M BIS-Tris pH 5.5). Crystals were cryoprotected in the reservoir solution + 25% ethylene glycol,

mounted on a litholoop and cryocooled in liquid nitrogen. Data collection for $Bd1431_{F120D}$ crystals was performed at Diamond Light Source on beamline IO4. Datasets were collected with an x-ray wavelength of 0.9793 Å and transmission of 100%. The crystals were rotated 360° in oscillations of 0.1°, producing a dataset with 3600 images. Diffraction data were processed using the Xia2-dials automated pipeline[161,162]. Initial phases for $Bd1431_{F120D}$ were estimated by MR (using PhaserMR[163]) using the crystal structure of Bd1431 as the search model (further information provided in section 5.2.5).

For co-crystallisation of the Bd1431:Thymidine 3',5'-diphosphate (TDP) complex, Bd1431 was concentrated to 12 mg/mL. A five-fold molar excess (3.5 mM) of TDP was incubated with Bd1431 for one hour on ice. Bd1431-TDP crystals were grown in Midas E4 (0.2M Sodium thiocyanate, 0.1M HEPES pH 7.0, 40% v/v Pentaerythritol propoxylate). Bd1431-TDP crystals grew in a cryoprotective condition and were harvested and cryocooled directly from the reservoir solution. Data collection for Bd1431-TDP crystals was performed at Diamond Light Source on beamline IO4. Datasets were collected with an x-ray wavelength of 0.9795 Å and transmission of 100%. The crystals were rotated 360° in oscillations of 0.1°, producing a dataset with 3600 images. Diffraction data were processed using the Xia2-dials automated pipeline[161,162]. Initial phases for Bd1431-TDP were estimated by MR (using PhaserMR[163]) using the crystal structure of Bd1431 as the search model (further information provided in section 5.2.6)

2.4.4 Bd1431 activity assays

The nuclease activity of dimeric Bd1431 was investigated using a DNA digestion assay. Bd1431 was dialysed overnight into a buffer containing 20 mM Hepes pH 7.5, 150 mL NaCl and 5 mM $CaCl_2$. PCold1 plasmid DNA was singly digested with NdeI to produce linearised DNA. Each digestion reaction was 20 μL and contained 200 ng linearised PCold1 (3.6 nM). A serial dilution of Bd1431 was performed so that molar ratios of 100:1 10:1 1:1 1:10 1:100 1:1000 1:10000 (Protein:DNA) could be tested. Each reaction was incubated for one hour at 37 °C before being inactivated by the addition of 5mM EGTA. Agarose gel electrophoresis on a 2% gel was performed for each reaction to assess the nuclease activity. An identical assay was performed using commercial Staphylococcal Nuclease (ThermoFisher) as a positive control.

Chapter 3

Structural characterisation of the *B. bacteriovorus* gliding machinery

3.1 Introduction

3.1.1 Structural characterisation of GltD

The current understanding of gliding motility is largely dominated by microscopy and genetic studies. Although these studies have provided invaluable insights into how cells glide, progress has been slowed by the lack of structural data that can help underpin the specific interactions that mediate the assembly of the Agl-Glt complex. However, the work of Dr Amber Wilson (Lovering lab, unpublished) has made initial attempts towards a structural understanding of gliding, culminating in crystal/cryo-EM structures of GltD from the Glt3 and Glt4 operons; these structures represent the first structures of any component of the Agl-Glt complex.

As previously discussed, GltD (a TPR-rich protein) has been shown to interact with IM, periplasmic and OM Glt proteins. As such, it is predicted to function as the central interaction hub of the Agl-Glt complex, acting to permit force transduction across the cell envelope. GltD's central role as the foundational scaffold of the Agl-Glt complex made it a priority for structural investigation. The crystal structure of full-length Bd2374 (GltD4), solved to 2.8Å (Figure 3.1), revealed that the protein is composed of 18 TPR repeats that form an extended superhelix. At the N-terminus, the longer A-helix of TPR1 (the two helices of a TPR repeat are denoted A-helix and B-helix) forms a hook-like architecture that could possibly mediate protein-protein interactions. The GltD4

superhelix is capped C-terminally by a four-helix bundle that finally extends into two three-helix bundles that pack against the concave face of the superhelix. Variations in the length of TPR helices along the length of the GltD4 superhelix create an increased surface area and several clefts that could house interactions between GltD4 and the other components of the Glt complex.

When analysing the crystal packing of the full-length GltD4 structure, two main crystal contacts exist between adjacent lattice mates, with both the N-terminal hook and C-terminal four-helix bundle contacting distinct regions of the TPR superhelix on a neighbouring protomer (Figure 3.2). Interestingly, when the N- and C-terminal domains of GltD4 are crystallised separately, these major crystal contacts are conserved. The conservation of crystal contacts in the truncated structures (Figure 3.2), which all belong to a different spacegroup and presumably have more steric freedom, suggests that the contacts may be biologically relevant[168,169] and supports a notion that GltD forms arrays in gliding cells[97]. The biological relevance of the crystal packing of GltD was strengthened when a partial structure was solved of the N-terminal domain of Bd1478 (GltD3) which also maintained the same relative crystal contacts (Figure 3.2).

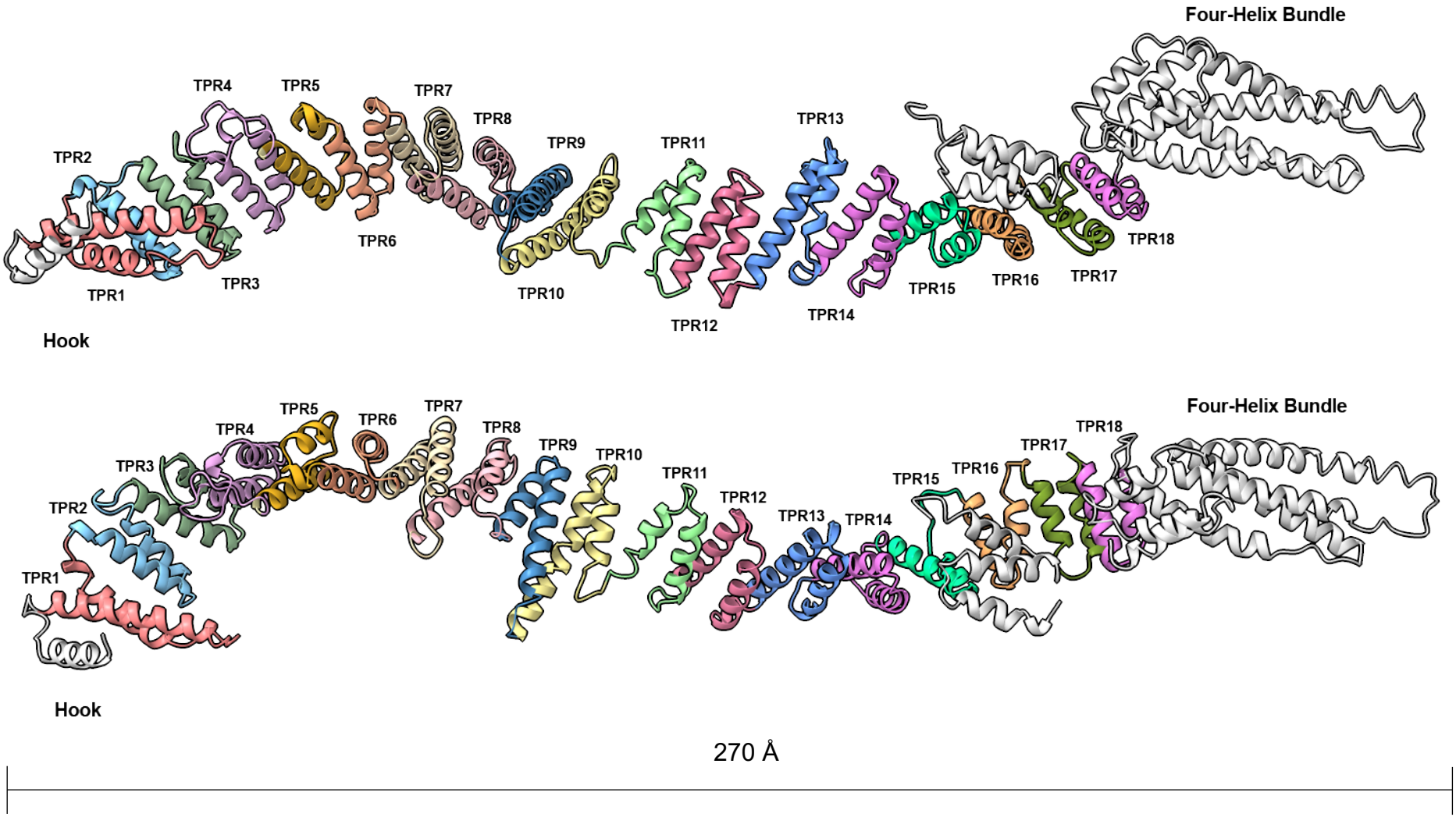


Figure 3.1: Crystal structure of full-length GltD4
GltD4 comprises 18 TPR motifs that form a superhelical architecture. The extension of TPR1 results in the formation of an N-terminal hook domain that could mediate interactions with other periplasmic Glt proteins. At the C-terminus, GltD4 is capped by a four-helix bundle. The presented structures are coloured by TPR for clarity.

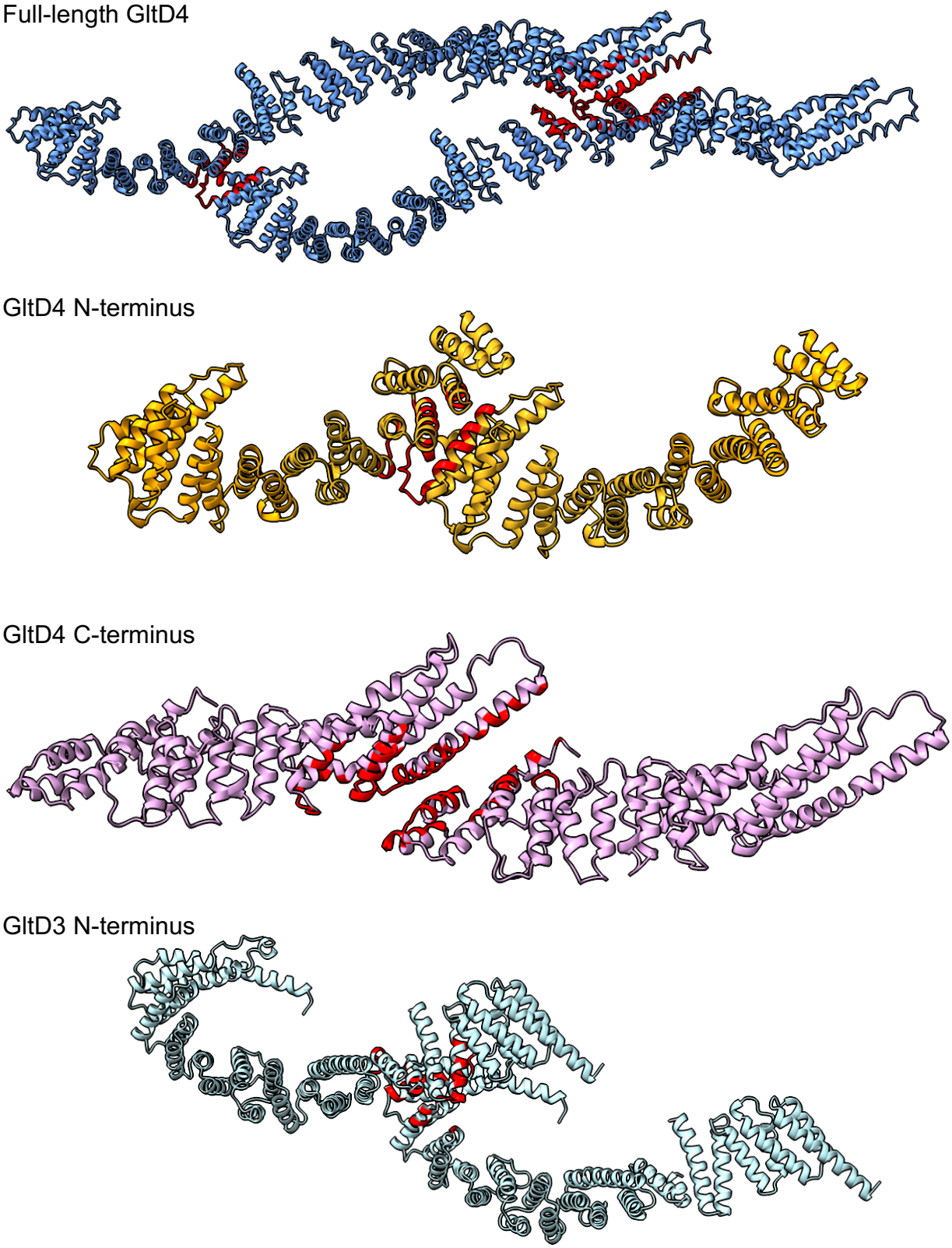


Figure 3.2: Crystal packing in GltD structures
The major crystal contacts (coloured red) are conserved in the crystal lattices of all GltD structures solved by crystallography, despite each crystal belonging to a different spacegroup. The conservation of the crystal contacts is indicative of these interactions being biologically relevant.

In addition to the crystal structures of GltD4, two structures of the C-terminal domain of GltD3 were obtained by cryo-EM (Figure 3.3). The C-terminus of GltD3 shares similarities to that of GltD4, with four TPR repeats leading into a four-helix bundle. The TPR repeats observed in the truncated structure are presumably identical to those formed by the full-length TPR superhelix. In contrast to GltD4, a flexible linker extends from the four-helix bundle of GltD3 leading into an auxiliary lobe comprising six TPR repeats. This auxiliary lobe (residues 1008–1237) packs against the TPR superhelix and the four-helix bundle. The flexible linker of GltD3 threads through the TPR motifs of the auxiliary lobe, occupying the majority of its concave face. This packing of an extended flexible peptide against the concave face of an array of TPR motifs is a common interface observed in TPR-ligand interactions. This type of interaction is seen between the regulatory cell division protein LGN and a flexible client peptide on its binding partner NuMA[170] (Figure 3.4). Comparison of the two EM structures of GltD3 reveals a conformational change between an open and closed conformation. This change corresponds to the movement of the four-helix bundle and the auxiliary lobe. This movement creates an extended cleft which could act as a binding site. In the closed conformation, this cleft is largely occupied by the flexible ‘pseudo client peptide’ (residues 936–1007), suggesting that this could act in an inhibitory manner to regulate ligand binding.

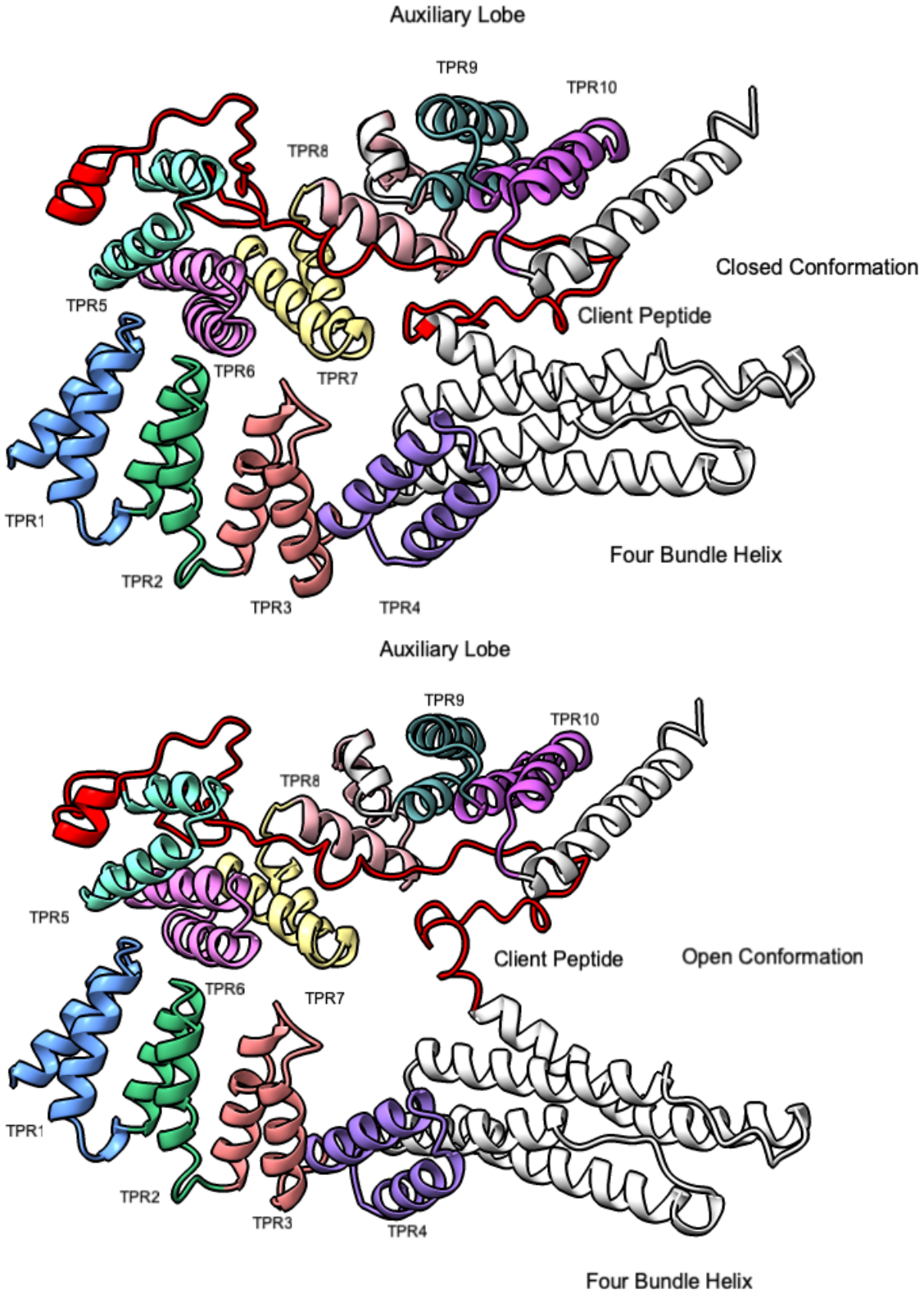


Figure 3.3: GltD3 cryo-EM structures
The structure of GltD3 was solved in two conformations by cryo-EM. The TPR body and four-helix bundle share similarity to the structure of GltD4. GltD3 contains an additional auxiliary lobe that is composed of 6 TPR motifs. A flexible peptide (coloured red) connects the four-helix bundle and auxiliary lobe and packs against the concave face of the auxiliary lobe in a manner akin to many TPR ligands. The presented structures are coloured by TPR for clarity.

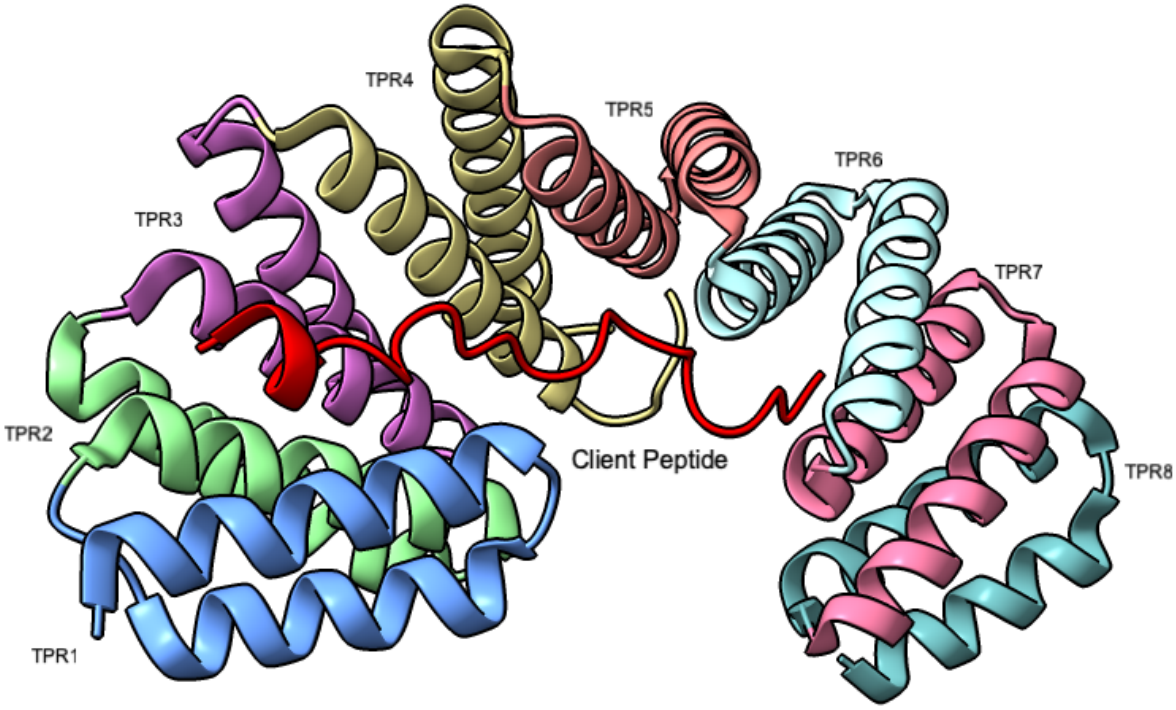


Figure 3.4: TPR ligand binding
Example of TPR ligand interactions. The flexible peptide (shown in red) of NumA threads along the concave face of the regulatory cell division protein LGN. The presented structure is coloured by TPR for clarity. (PDB: 3RO2)

GltD3 is ~200 residues larger than any of *B. bacteriovorus's* other GltD paralogs and sequence alignment demonstrates that the conservation between each GltD paralog deteriorates at the auxiliary lobe of GltD3 (Figures 3.5 & 3.6 and Table 1). The auxiliary lobe does, however, share some similarity to GltE (which is only present on the Glt1 operon). This suggests that the increased size of GltD3 could be the result of the fusion of GltD and GltE within an ancestral Glt3 operon.

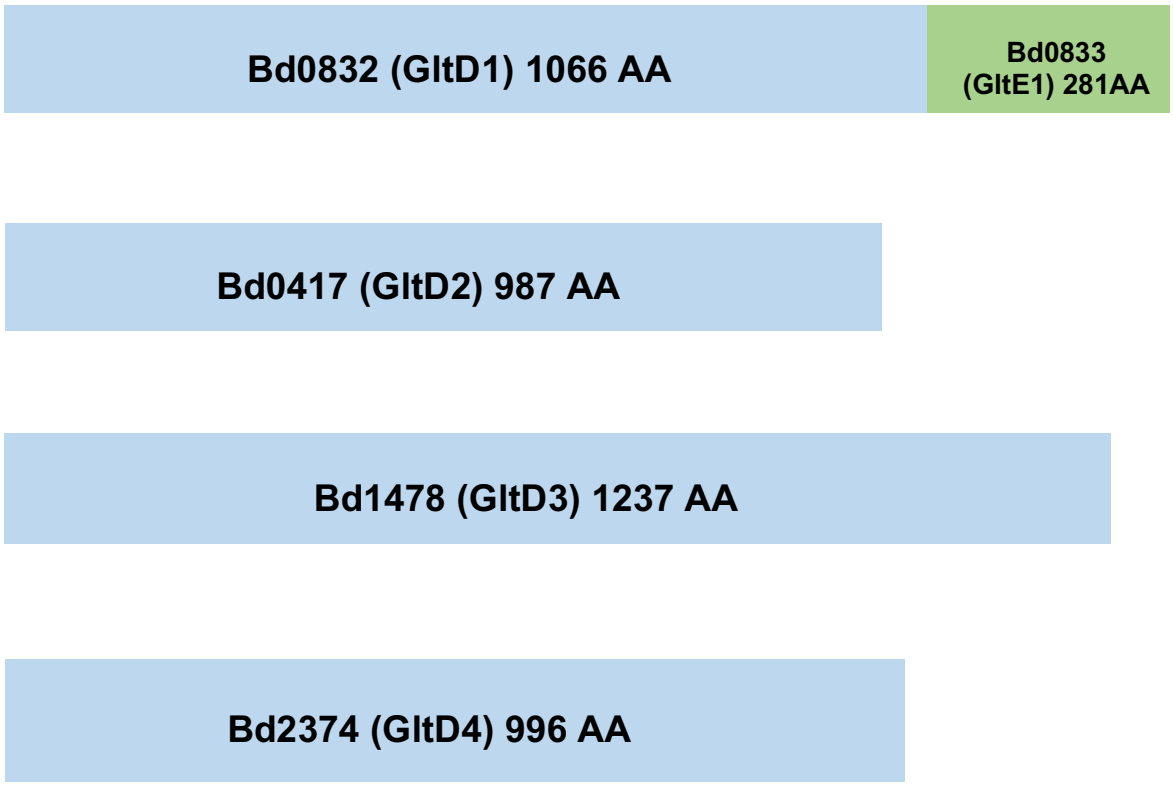
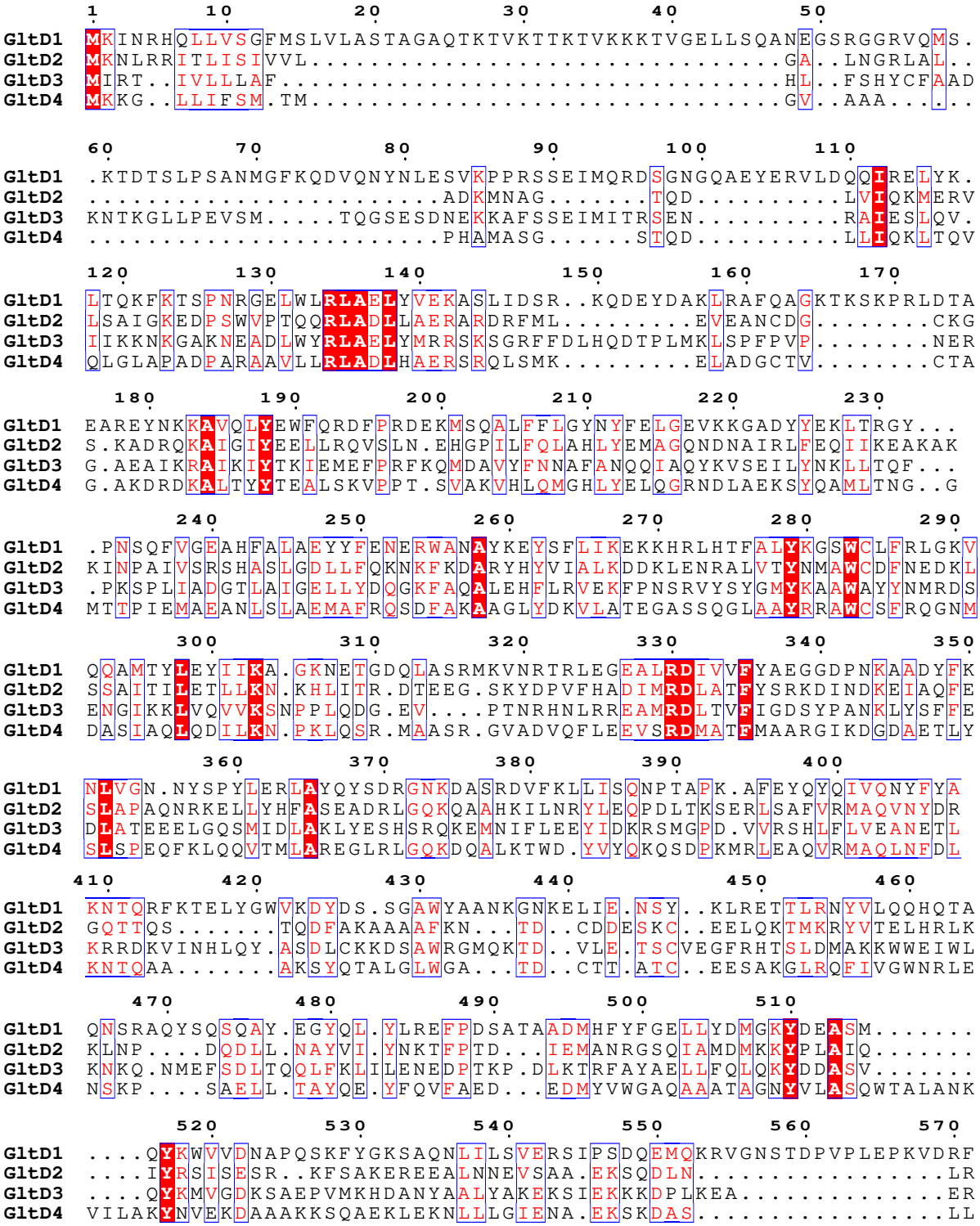
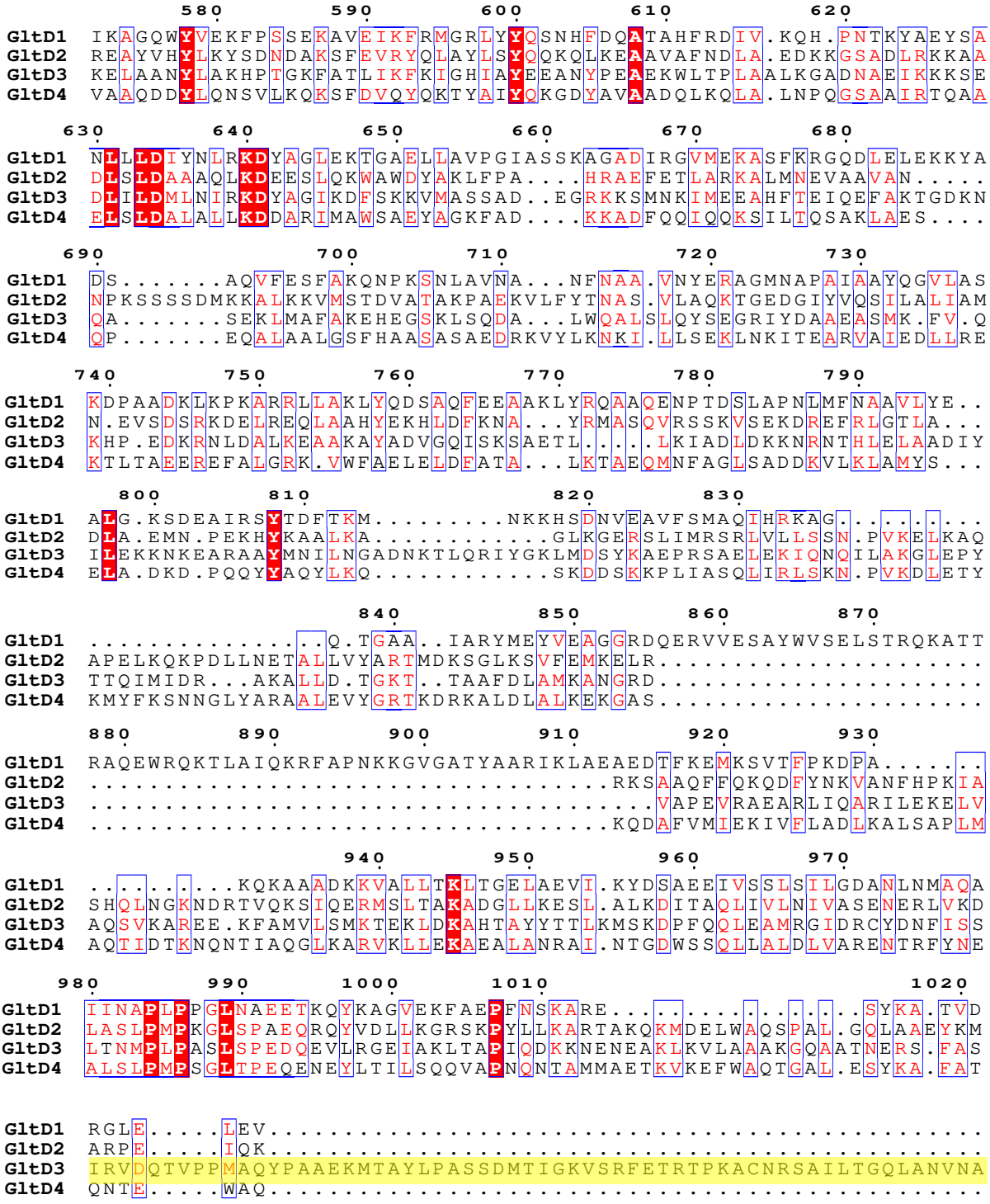


Figure 3.5: Schematic diagram of *B. bacteriovorus's* four GltD paralogs.
GltD3 is ~200 residues larger than any of *B. bacteriovorus's* other GltD paralogs. These additional residues form the auxiliary lobe of GltD3. The auxiliary lobe of GltD3 shares sequence similarity with GltE1, suggesting that the increased size of GltD3 is a result of a fusion between GltD and GltE in the Glt3 operon.

Table 3.1: Sequence identity between *B. bacteriovorus* GltD paralogs

	GltD1	GltD2	GltD3	GltD4	GltE
GltD1		20.00%	21.93%	23.44%	
GltD2	20.00%		21.78%	28.61%	
GltD3	21.93%	21.78%		25.00%	
GltD4	23.44%	28.61%	25.00%		
GltD3 Auxiliary Lobe					21.58%


GltD1
GltD2
GltD3
GltD4


580 590 600 610 620
GltD1 IKAGQWYVEKFPSSEKAVEIKFRMGRLYYQSNHFDQATAHFRDIV.KQH.PNTKYAEYSA
GltD2 REAYVHYLKYSDNDAKSFEVRYQLAYLSYQQKQLKEAAVAFNDLA.EDKKGSADLRKKAA
GltD3 KELAANYLAKHPTGKFATLIKFKIGHIAYEEANYPEAEKWLTPLAALKGADNAEIKKKSE
GltD4 VAAQDDYLQNSVLKQKSFDVQYQKTYAIYQKGDYAVAADQLKQLA.LNPQGSAAIRTQAA
630 640 650 660 670 680
GltD1 NLLLDIYNLRKDYAGLEKTGAELLAVPGIASSKAGADIRGVMEKASFKRGQDLELEKKYA
GltD2 DLSLDAAAQLKDEESLQKWAWDYAKLFPA....HRAEFETLARKALMNEVAAVAN.....
GltD3 DLILDMLNIRKDYAGIKDFSKKVMASSAD..EGRKKSMNKIMEEAHFTEIQEFAKTGDKN
GltD4 ELSLDALALLKDDARIMAWSAEYAGKFAD....KKADFQQIQQKSILTQSAKLAES....
690 700 710 720 730
GltD1 DS.......AQVFESFAKQNPKSNLAVNA...NFNAA.VNYERAGMNAPAIAAYQGVLAS
GltD2 NPKSSSSDMKKALKKVMSTDVATAKPAEKVLFYTNAS.VLAQKTGEDGIYVQSILALIAM
GltD3 QA.......SEKLMAFAKEHEGSKLSQDA...LWQALSLQYSEGRIYDAAEASMK.FV.Q
GltD4 QP.......EQALAALGSFHAASASAEDRKVYLKNKI.LLSEKLNKITEARVAIEDLLRE
740 750 760 770 780 790
GltD1 KDPAADKLKPKARRLLAKLYQDSAQFEEAAKLYRQAAQENPTDSLAPNLMFNAAVLYE..
GltD2 N.EVSDSRKDELREQLAAHYEKHLDFKNA...YRMASQVRSSKVSEKDREFRLGTLA...
GltD3 KHP.EDKRNLDALKEAAKAYADVGQISKSAETL.....LKIADLDKKNRNTHLELAADIY
GltD4 KTLTAEERFALGRK.VWFAELELDFATA...LKTAEQMNFAGLSADDKVLKLAMYS...
800 810 820 830
GltD1 ALG.KSDEAIRSYTDFTKM.........NKKHSDNVEAVFSMAQIHRKAG.....
GltD2 DLA.EMN.PEKHYKAALKA.............GLKGERSLIMRSRLVLLSSN.PVKELKAQ
GltD3 ILEKKNKEARAAYMNILNGADNKTLQRIYGKLMDSYKAEPRSAELEKIQNQILAKGLEPY
GltD4 ELA.DKD.PQQYYAQYLKQ.............SKDDSKKPLIASQLIRLSKN.PVKDLETY
840 850 860 870
GltD1 ..............Q.TGAA..IARYMEYVEAGGRDQERVVESAYWVSELSTRQKATT
GltD2 APELKQKPDLLNETALLVYARTMDKSGLKSVFEMKELR.....................
GltD3 TTQIMIDR...AKALLD.TGKT..TAAFDLAMKANGRD.....................
GltD4 KMYFKSNNGLYARAALEVYGRTKDRKALDLALKEKGAS.....................
880 890 900 910 920 930
GltD1 RAQEWRQKTLAIQKRFAPNKKGVGATYAARIKLAEAEDTFKEMKSVTFPKDPA.....
GltD2 ....RKSAAQFFQKQDFYNKVANFHPKIA
GltD3 ..................................VAPEVRAEARLIQARILEKELV
GltD4 ..................................KQDAFVMIEKIVFLADLKALSAPLM
940 950 960 970
GltD1 ..........KQKAAADKKVALLTKLTGELAEVI.KYDSAEEIVSSLSILGDANLNMAQA
GltD2 SHQLNGKNDRTVQKSIQERMSLTAKADGLLKESL.ALKDITAQLIVLNIVASENERLVKD
GltD3 AQSVKAREE.KFAMVLSMKTEKLDKAHTAYYTTLKMSKDPFQQLEAMRGIDRCYDNFISS
GltD4 AQTIDTKNQNTIAQGLKARVKLLEKAEALANRAI.NTGDWSSQLLALDLVARENTRFYNE
980 990 1000 1010 1020
GltD1 IINAPLPPGLNAEETKQYKAGVEKFAEPFNSKARE.....SYKA.TVD
GltD2 LASLPMPKGLSPAEQRQYVDLLKGRSKPYLLKARTAKQKMDELWAQSPAL.GQLAAEYKM
GltD3 LTNMPLPASLSPEDQEVLRGEIAKLTAPIQDKKNENEAKLKVLAAAKGQAATNERS.FAS
GltD4 ALSLPMPSGLTPEQENEYLTILSQQVAPNQNTAMMAETKVKEFWAQTGAL.ESYKA.FAT
GltD1 RGLE.....LEV.....
GltD2 ARPE.....IQK.....
GltD3 IRVDQTVPPMAQYPAAEKMTAYLPASSDMTIGKVSRFETRTPKACNRSAILTGQLANVNA
GltD4 QNTE.....WAQ.....

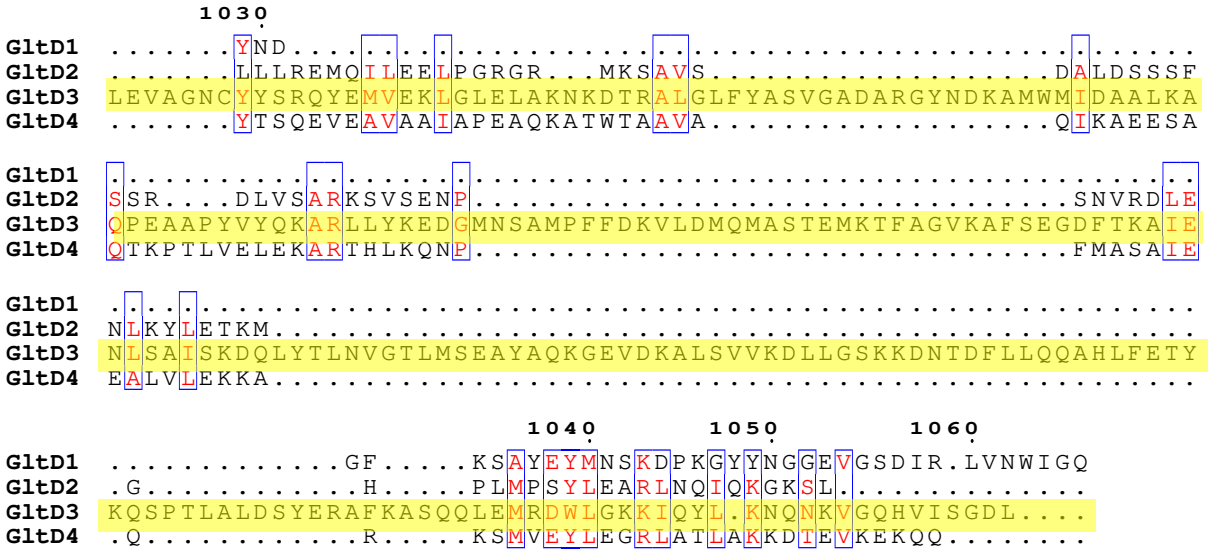


Figure 3.6: Multiple sequence alignment of the four *B. bacteriovorus* GltD paralogs
Sequence alignment of Glt1-Glt4 reveals that the conservation between each GltD paralog deteriorates at the auxiliary lobe of GltD3. The sequence for the auxiliary lobe of GltD3 is highlighted yellow. Alignment made with Espript3.0[171]. Colour code: Red background, white text = strict identity; white background, red text = similarity in a group; blue frame = similarity across groups.

3.1.2 Chapter Aims

The observed differences between the C-termini of GltD3 and GltD4 are indicative of GltD3 representing a gene fusion between both GltD and GltE. The Glt1 operon bears the closest resemblance to the *M. xanthus* gliding machinery and is the only operon to contain a gene encoding a potential GltE. In order to investigate the possibility that GltD3 represents a gene fusion event, we will probe an interaction between GltD1 and GltE1 via biophysical and structural studies. As previously mentioned in Chapter 1, GltE is a lipoprotein predicted to localise to the OM. The confirmation of an interaction between GtlD1 and GltE1 would provide the first insights into the orientation of the Agl-Glt complex with respect to the cell envelope. Structural characterisation of GltD1 via x-ray crystallography will also provide the opportunity to investigate whether the crystal contacts are also conserved amongst another GltD homolog and may provide data to strengthen the GltD array hypothesis.

3.2 Results

3.2.1 AF modelling of GtlD1-GltE1

To gain an initial insight into the structures of GltD1 and GltE1, *Ab-initio* models were generated using AF[160] (Figure 3.7). These models revealed that GltD1 is predicted to follow the same overall fold observed in the full-length crystal structure of GltD4, with an N-terminal hook leading into superhelical body comprising 18 TPR repeats which is capped at the C-terminus by a four-helix bundle. The four-helix bundle extends into a flexible tail-peptide as seen in GltD3. However, GltD1 lacks the auxiliary lobe observed at the C-terminus GltD3. The AF model of GltE1 shares strong similarity to the auxiliary lobe of GltD3, forming a superhelical architecture comprising 6 TPR repeats.

AF was further used to create a co-folded model of GltD1 and GltE1 in complex. In this model, GltE1 is seen to pack against the TPR-body and four-helix bundle of GltD1 where the flexible tail-peptide of GltD1 packs against the concave face of GltE1 in a manner similar to that seen in GltD3. A comparison of the GltD1:GltE1 AF model against the ‘open’ and ‘closed’ conformations of GltD3 show that the AF model resembles an average position of both conformations. When taken together, the insights gained from AF modelling support an interaction between GltD1 and GltE1 whilst also supporting the hypothesis that the increased size of GltD3 corresponds to a gene fusion of GltD and GltE in the Glt3 operon.

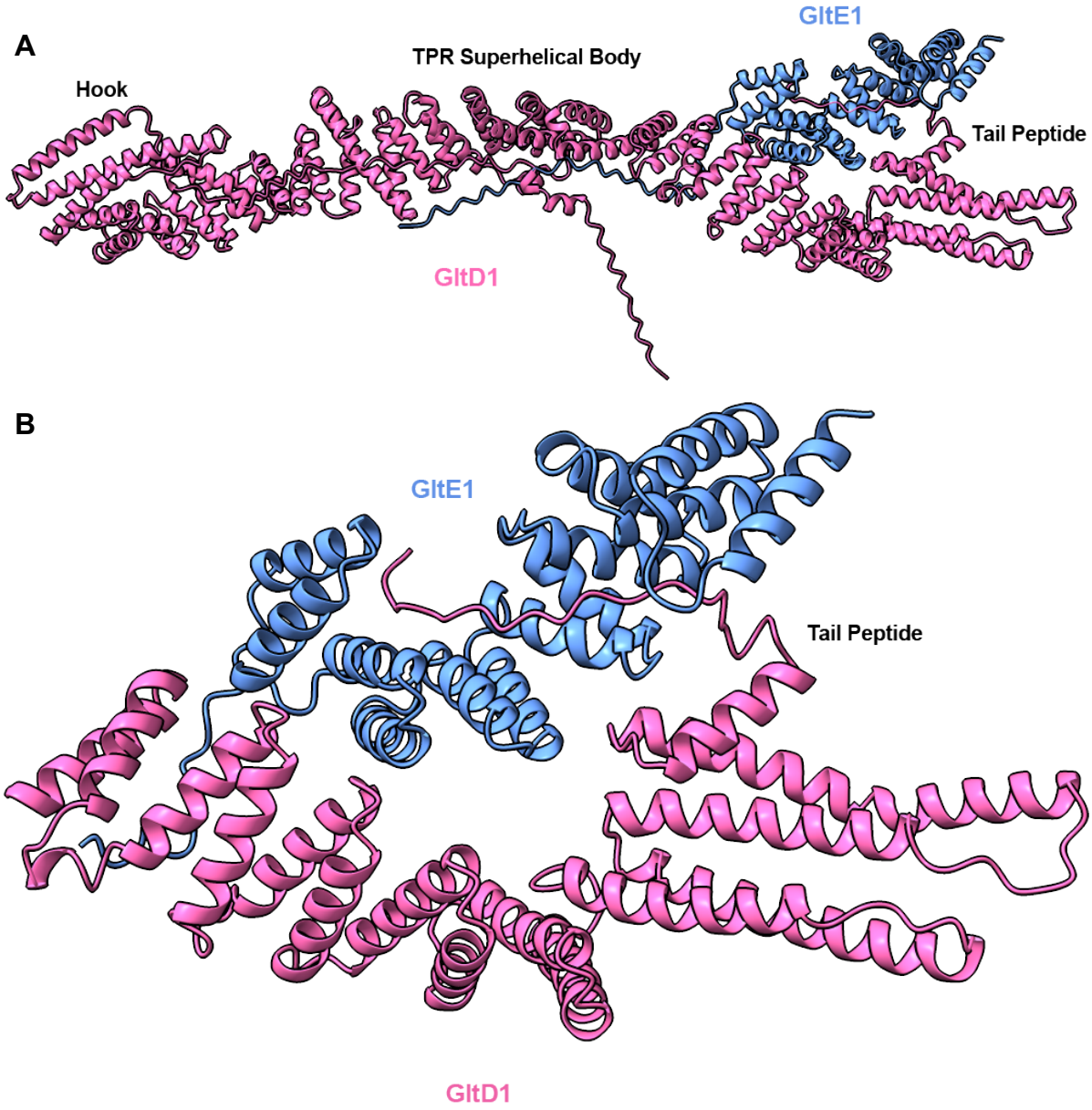


Figure 3.7: AF models of GltD1-GltE1 complex
A – Full-length model of GltD1-GltE1 interaction. GltD1 (coloured pink) and GltE1 (coloured blue). Model generated using the ColabFold server. Average pLDDT (confidence) scores for AF models are as follows: GltD1 = 84.1; and GltE1 = 84.8.
B – Close-up of GltD1-GltE1 interface. The flexible tail peptide of GltD1 packs along the concave face of the GltE1 superhelix.

3.2.2 Expression and purification of GltD1 and GltE1

The AF modelling of the GltD1:GltE1 complex provides compelling preliminary data of an interaction between the two proteins, although the nature of this interaction requires experimental validation.

Full-length GltD1 and GltE1 were expressed from the constructs described in Table 2.6. GltD1 expressed well and was initially purified using IMAC where the protein was eluted in a single population at ~17% concentration buffer B (85 mM imidazole) (Figure 3.8). SDS-PAGE confirmed the presence of GltD1 whilst also revealing some initial signs of degradation (Figure 3.8). GltD1 containing fractions were pooled and further purified by SEC. GltD1 was eluted as a monodisperse peak from HiLoad 26/600 Superdex 200 prep grade column at 168 mL, an elution volume roughly consistent with the size of monomeric GltD1 (~120 KDa) (Figure 3.8). SDS-PAGE demonstrated the high purity of the post-SEC GltD1 sample and revealed that signs of degradation persisted throughout the subsequent purification stages (Figure 3.8).

GltE1 also expressed well and was initially purified using IMAC where the protein was eluted in a single population at ~21% concentration buffer B (105 mM imidazole) (Figure 3.9). SDS-PAGE confirmed the presence of GltE1 and, similar to GltD1, also revealed some initial signs of degradation (Figure 3.9). GltE1 containing fractions were pooled and further purified by SEC. GltE1 was eluted as a monodisperse peak from HiLoad 26/600 Superdex 75 prep grade column at 172 mL, an elution volume consistent with the size of monomeric GltE1 (~31 KDa) (Figure 3.9). SDS-PAGE demonstrated the high purity of the post-SEC GltE1 sample and revealed that signs of degradation persisted throughout subsequent purification stages (Figure 3.9). The signs of degradation observed for both GltD1 and GltE1 suggests that the two proteins may stabilise each other upon interaction.

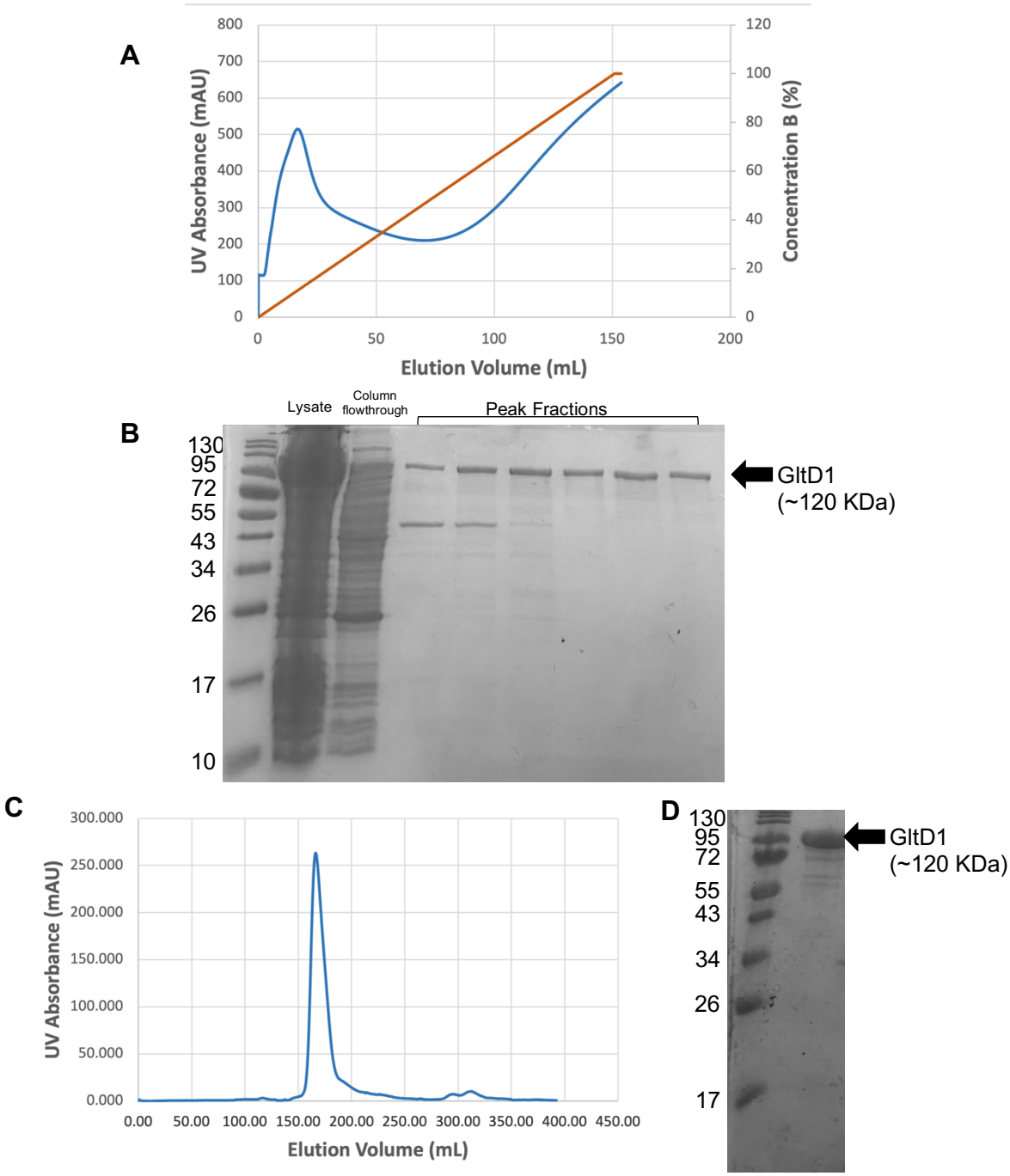


Figure 3.8: Purification of GltD1

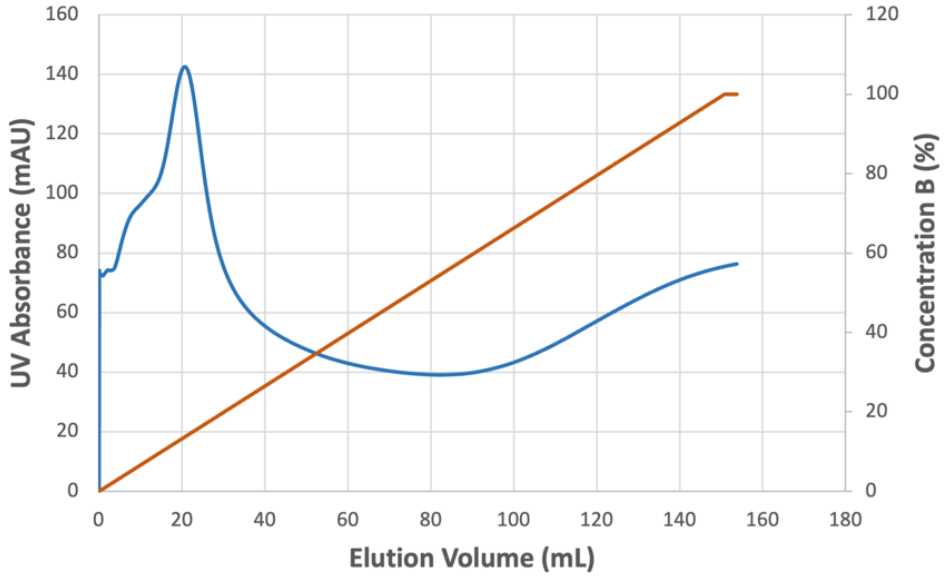
A – Chromatogram of GltD1 IMAC purification. IMAC performed on a 5 mL HisTrap HP using a gradient elution of 0-100% buffer B across 30 CV. GltD1 elutes in a single population at ~17% concentration buffer B (85 mM imidazole). Curves for UV absorbance at 280 nM (blue trace) and buffer B concentration (orange trace) are shown.

B – SDS-PAGE analysis of IMAC purification. GltD1 can be seen as a predominant band between the 95 and 130 kDa markers. The gel also reveals the presence of some minor contaminants and shows signs of GltD1 degradation. The black arrow indicates the position of GltD1.

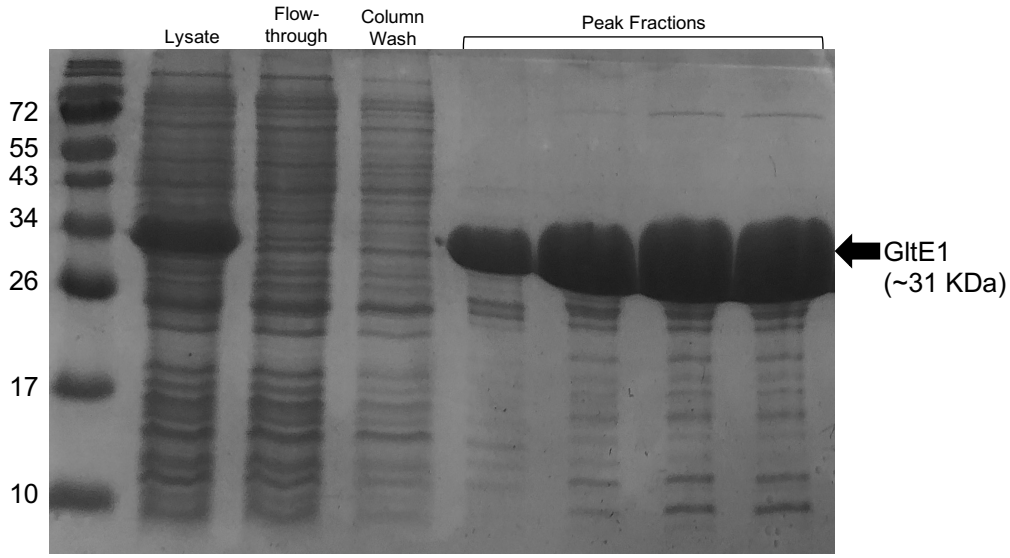
C – SEC Chromatogram for GltD1. GltD1 was eluted as a monodisperse peak from a HiLoad 26/600 Superdex 200 prep grade column at 168 mL, an elution volume consistent with the size of monomeric GltD1 (~120 KDa). Curve for UV absorbance at 280 nM shown in blue.

D – SDS-PAGE analysis of SEC peak fractions. GltD1 can be seen as a predominant band between the 95 and 130 kDa markers. The gel reveals that SEC has removed the majority of contaminants, but degradation is still present. The black arrow indicates the position of GltD1.

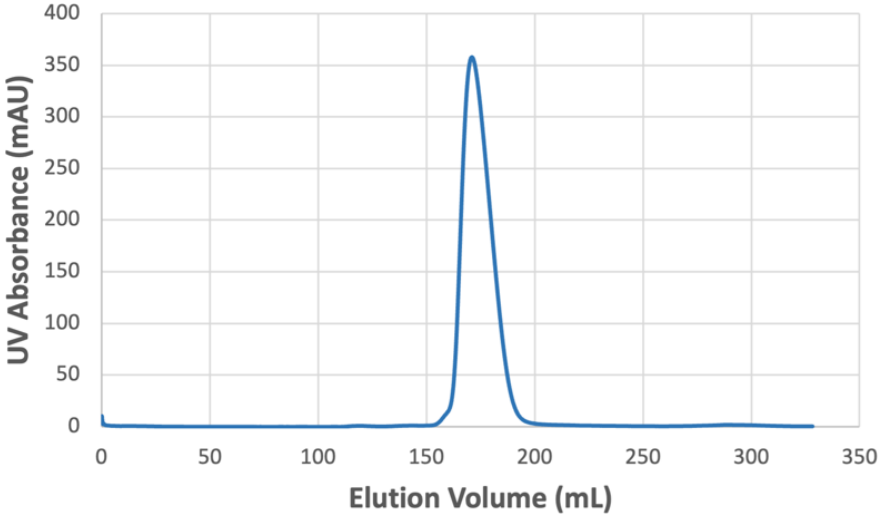
A


UV Absorbance (mAU)
Concentration B (%)
Elution Volume (mL)
0
20
40
60
80
100
120
140
160
180

B


Lysate
Flow-through
Column Wash
Peak Fractions
72
55
43
34
26
17
10
GltE1
(~31 KDa)

C


UV Absorbance (mAU)
Elution Volume (mL)
0
50
100
150
200
250
300
350
400

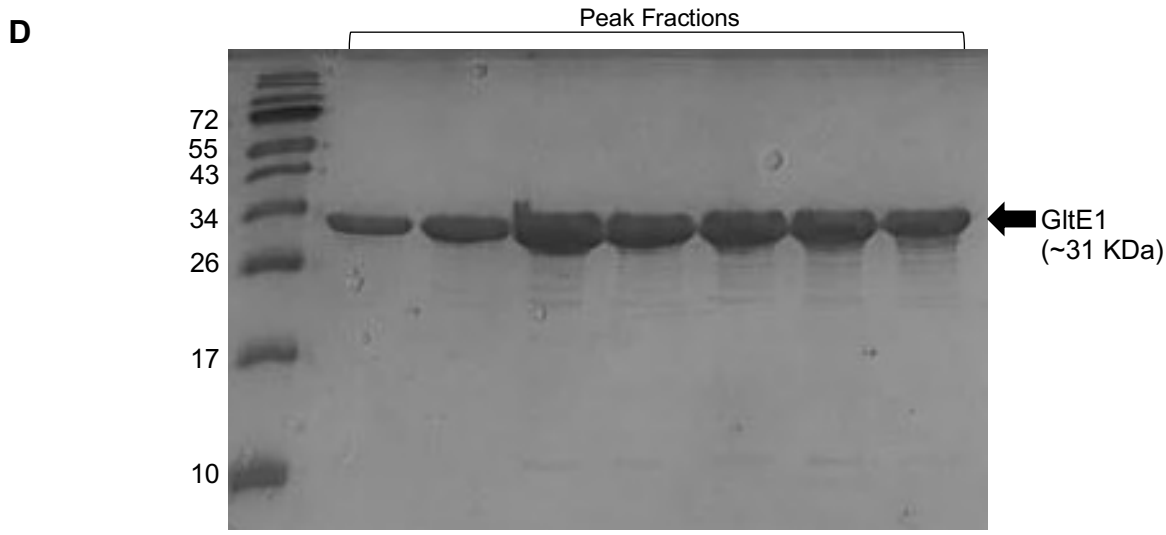


Figure 3.9: Purification of GltE1
A – Chromatogram of GltE1 IMAC purification. IMAC performed on a 5 mL HisTrap HP using a gradient elution of 0-100% buffer B across 30 CV. GltE1 elutes in a single population at ~21% concentration buffer B (105 mM imidazole). Curves for UV absorbance at 280 nM (blue trace) and buffer B concentration (orange trace) are shown.
B – SDS-PAGE analysis of IMAC purification. GltE1 can be seen as a predominant band between the 26 and 34 kDa MW markers. The gel also reveals the presence of some lower MW contaminants and shows signs of GltE1 degradation. The black arrow indicates the position of GltE1.
C – SEC Chromatogram for GltE1. GltE1 was eluted as a monodisperse peak from a HiLoad 26/600 Superdex 75 prep grade column at 172 mL, an elution volume consistent with the size of monomeric GltE1 (~31 KDa). Curve for UV absorbance at 280 nM shown in blue.
D – SDS-PAGE analysis of SEC peak fractions. GltE1 can be seen as a predominant band between the 26 and 34 kDa markers. The gel reveals that SEC has removed the majority of contaminants, but degradation is still present. The black arrow indicates the position of GltE1.

In order to assess the interaction between GltD1 and GltE1, a five-fold molar excess of GltE1 was added to GltD1. The mixed sample was left to equilibrate for one hour to allow GltE1 to saturate GltD1 before being applied to a HiLoad 26/600 Superdex 200 prep grade column. The GltD1:GltE1 sample was eluted as a monodisperse peak at 158 mL, corresponding to an approximately 10 mL shift in elution volume when compared to the chromatogram for GltD1 alone, consistent with an increase in molecular weight (Figure 3.10). The complete shift of this peak is demonstrative of a complete saturation of GltD1 with GltE1. A second peak was observed at an elution volume ~208 mL which corresponds to the excess monomeric GltE1. SDS-PAGE confirmed that the fractions underneath the first peak contained both GltD1 and GltE1

whilst the second peak contained only GltE1 (Figure 3.10). Breakdown was still observed for both proteins on SDS-PAGE, suggesting that the interaction does not fully protect the proteins from degradation.

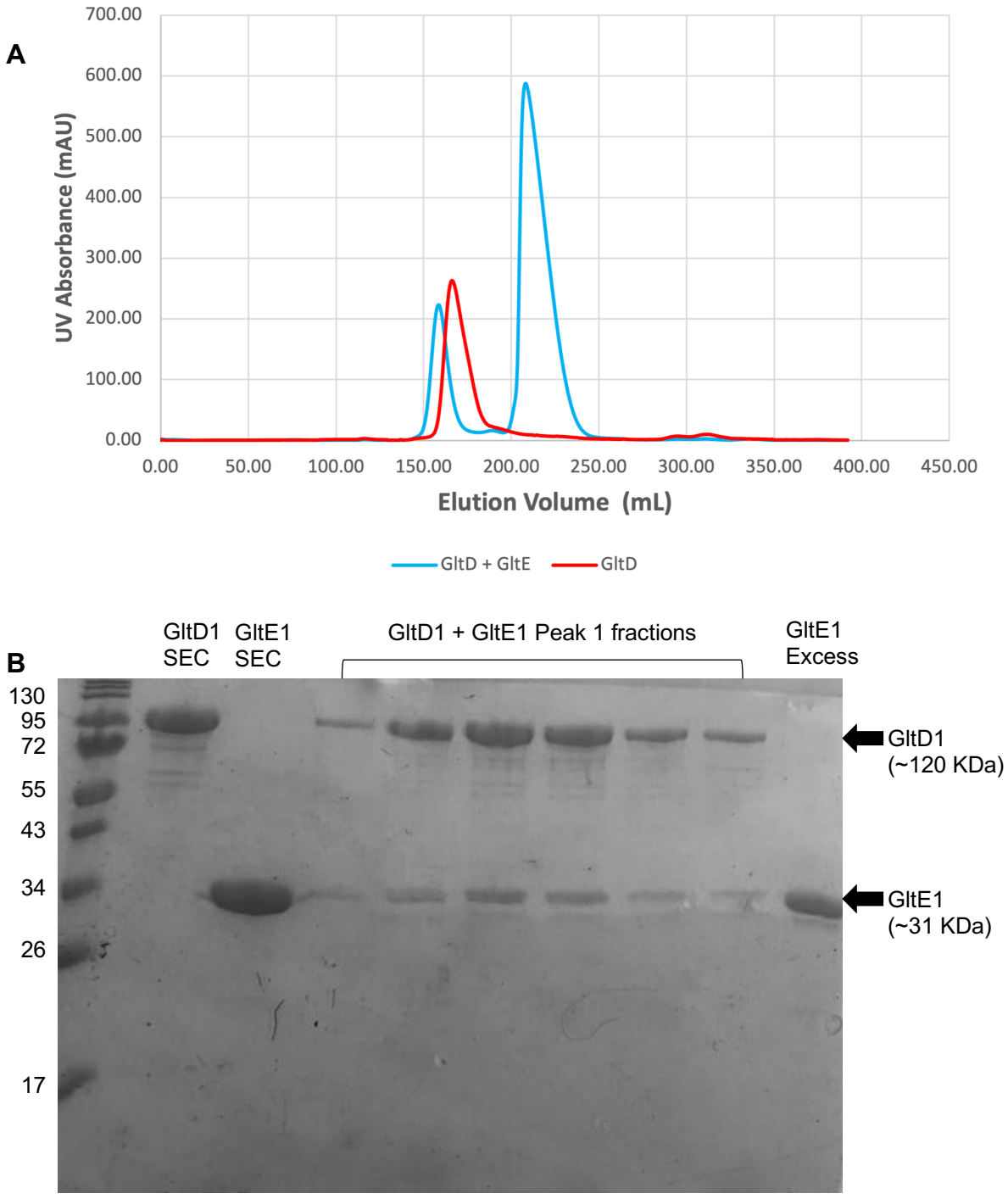


Figure 3.10: GltD1:GltE1 SEC interaction assay
A – SEC chromatogram for GltD1 (red) and GltD1:GltE1 complex (blue). SEC performed using a HiLoad 26/600 Superdex 200 prep grade column. The addition of GltE1 to GltD1 results in an ~10 mL shift in elution volume, consistent with an increase in molecular weight and an interaction between GltD1 and GltE1.
B – SDS-PAGE analysis of SEC peak fraction. The shifted peak contains both GltD1 and GltE1, confirming an interaction between GltD1 and GltE1. Black arrows indicate the position of GltD1 and GltE1.

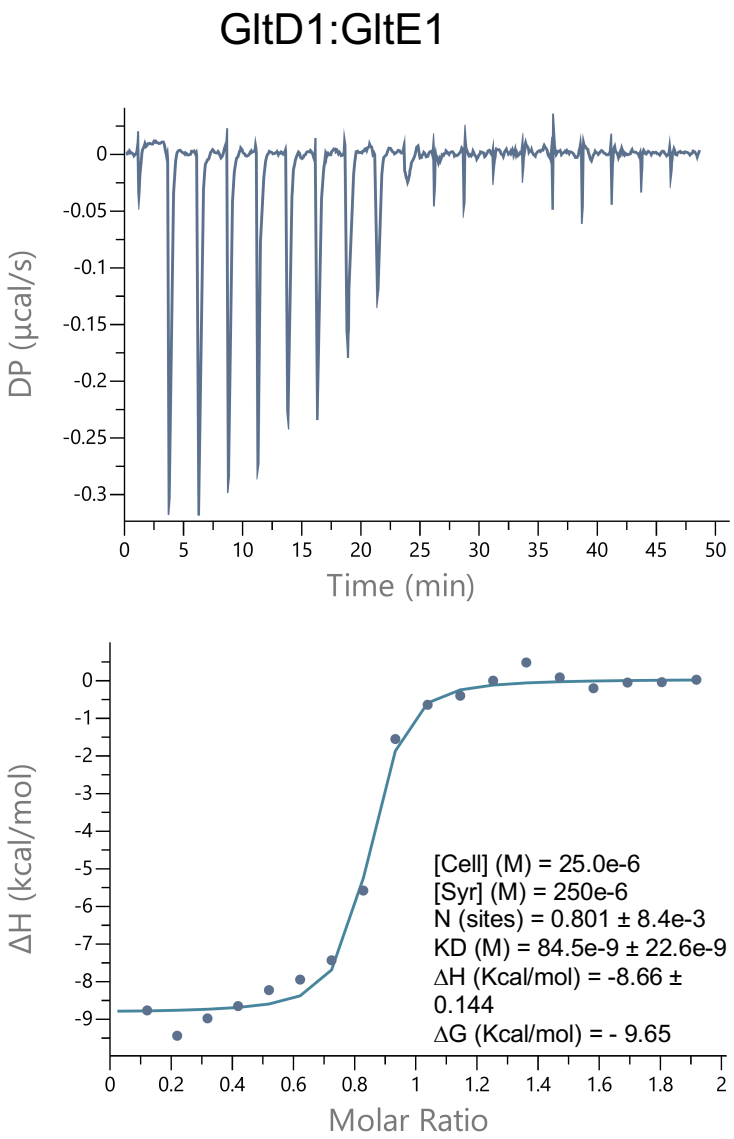
3.2.3 ITC analysis of GltD1-GltE1 interaction

The interaction between GltD1 and GltE1 was further characterised by ITC which confirmed the binding of GltD1 and GltE1 in a 1:1 stoichiometry with nanomolar affinity (Figure 3.11). Although the ITC studies confirmed the high affinity binding of GltD1 and GltE1, the nature of the interaction makes an accurate calculation of a dissociation constant (KD) difficult. Across four experiments identical in composition, the following values were recorded for the KD: 26.9 nM ± 9.7 nM, 58.2 nM ± 20.9 nM, 84.5 nM ± 22.6 nM and 125 nM ± 45.3 nM. The inaccuracy in the KD values is a result of the high affinity of the interaction which results in a steep binding isotherm. This causes high errors in the KD as the KD is calculated from the gradient of the isotherm. The normal optimisation route in this scenario would be to lower the concentration of protein in the cell. However, in the interaction between GltD1 and GltE1, the issue of the steep isotherm is compounded by the relatively small heats of binding between the two proteins, with only a small exothermic signal being detected. All optimisation attempts made with reduced cellular protein concentrations reduced the exothermic signal to undetectable levels.

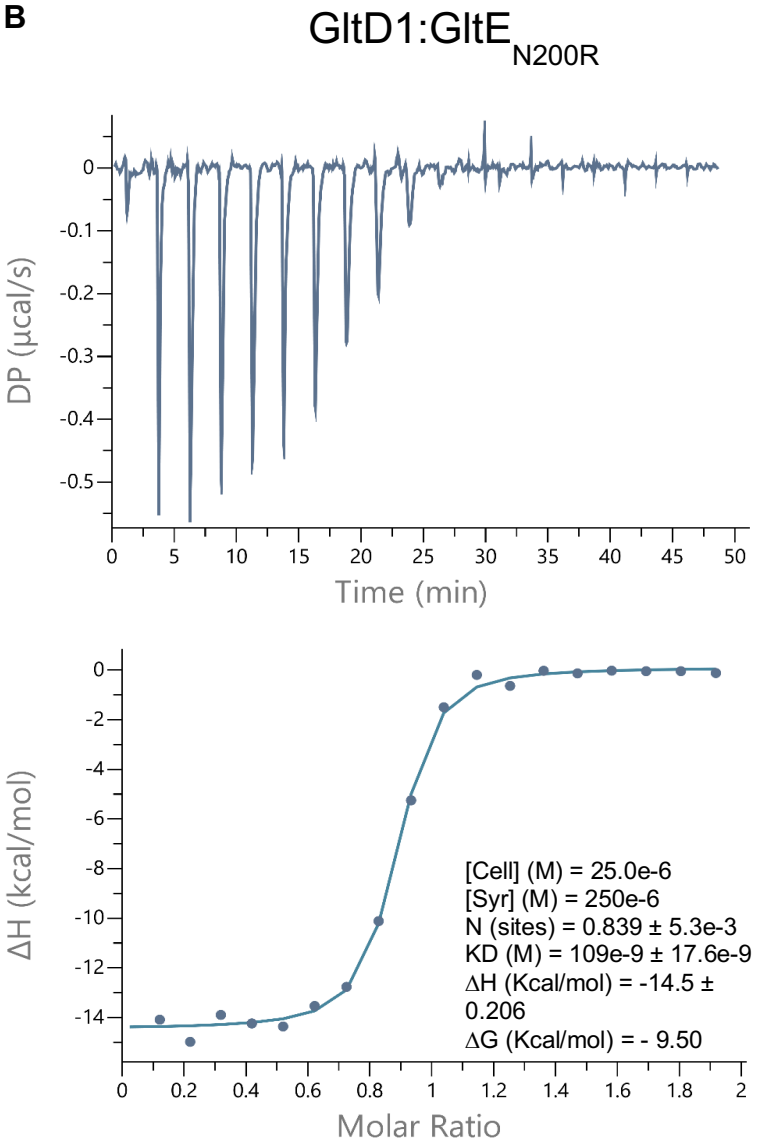
Guided by the AF model of the GltD1:GltE1 complex, two-point mutants ($GltD1_{A762D}$ and $GltE1_{N200R}$) were designed in an attempt to disrupt the interaction between the two proteins and thus provide some validation of the protein-protein interface predicted by AF. Both mutants were designed to disrupt a distinct aspect of the interface. $GltD1_{A762D}$ was designed to interfere with the packing of TPR15 of GltD1 into a hydrophobic pocket formed between TPR1 and TPR2 of GltE1 by inserting a larger negative reside into the pocket (Figure 3.12). In contrast, $GltE1_{N200R}$ was designed to disrupt the packing of the

tail peptide of GltD1 against the concave face of GltE1 (Figure 3.12). Both mutants expressed well and were purified by the same method used for the wild-type proteins. When used in ITC studies, neither GltD1$_{A762D}$ nor GltE1$_{N200R}$ ablated the interaction when tested against their respective wild type binding partner (Figure 3.11). These findings raise the possibility that either the interface between the two proteins was poorly modelled by AF, or that the insertion of a single point mutation into the interface was insufficient to cause detectable changes to the interaction. Taking a more drastic approach, a third construct was cloned in which the complete C-terminal tail peptide of GltD1 was deleted (GltD1$_{\Delta 1043\text{-}1066}$). The deletion of the C-terminal tail-peptide of GltD1 completely abolished the interaction between GltD1 and GltE1 (Figure 3.11). This is notable as the GltD1-GltE1 AF model predicts packing of GltE1 against the TPR body and four-helix bundle of GltD1 which would be expected to produce some residual signal even in the absence of the tail peptide.

A


GltD1:GltE1
DP (µcal/s)
Time (min)
ΔH (kcal/mol)
Molar Ratio
[Cell] (M) = 25.0e-6
[Syr] (M) = 250e-6
N (sites) = 0.801 ± 8.4e-3
KD (M) = 84.5e-9 ± 22.6e-9
ΔH (Kcal/mol) = -8.66 ± 0.144
ΔG (Kcal/mol) = - 9.65

B


GltD1:GltE N200R
DP (µcal/s)
Time (min)
ΔH (kcal/mol)
Molar Ratio
[Cell] (M) = 25.0e-6
[Syr] (M) = 250e-6
N (sites) = 0.839 ± 5.3e-3
KD (M) = 109e-9 ± 17.6e-9
ΔH (Kcal/mol) = -14.5 ± 0.206
ΔG (Kcal/mol) = - 9.50

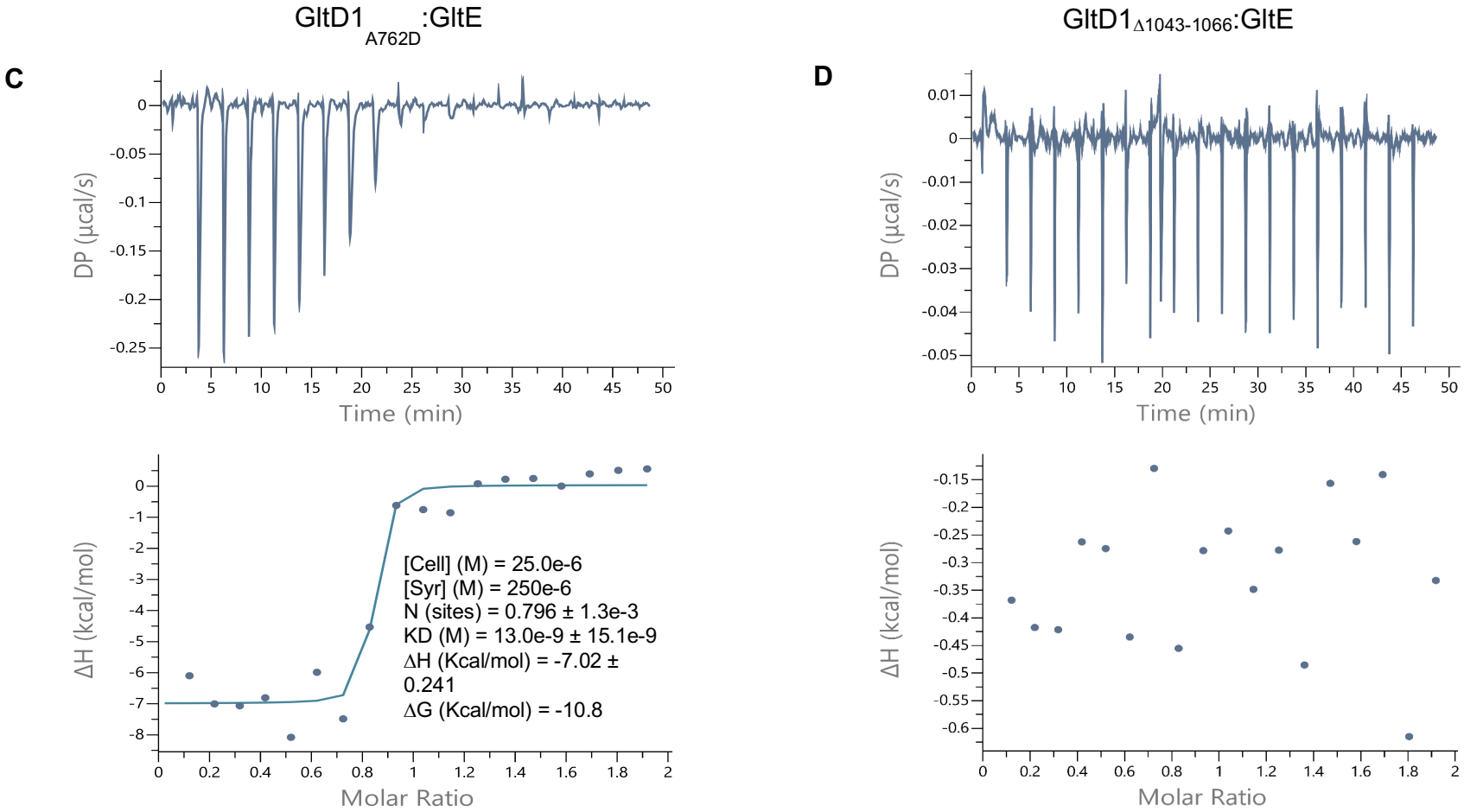


Figure 3.11: ITC binding for GltD1-GltE1 interaction
A – Binding thermogram for wildtype GltD1 and GltE1. The two proteins bind in a 1:1 ratio with nanomolar affinity.
B – Binding thermogram for GltD1:GltE$_{N200R}$ interaction. The insertion of the point mutation at the GltD1-GltE1 interface had no measurable impact on binding.
C – Binding thermogram for GltD1$_{A762D}$:GltE interaction. The insertion of the point mutation at the GltD1-GltE1 interface had no measurable impact on binding.
D – Binding thermogram for GltD1$_{\Delta 1043-1066}$:GltE interaction. The deletion of the C-terminal tail peptide of GltD1 completely abolished the interaction between GltD1 and GltE1.

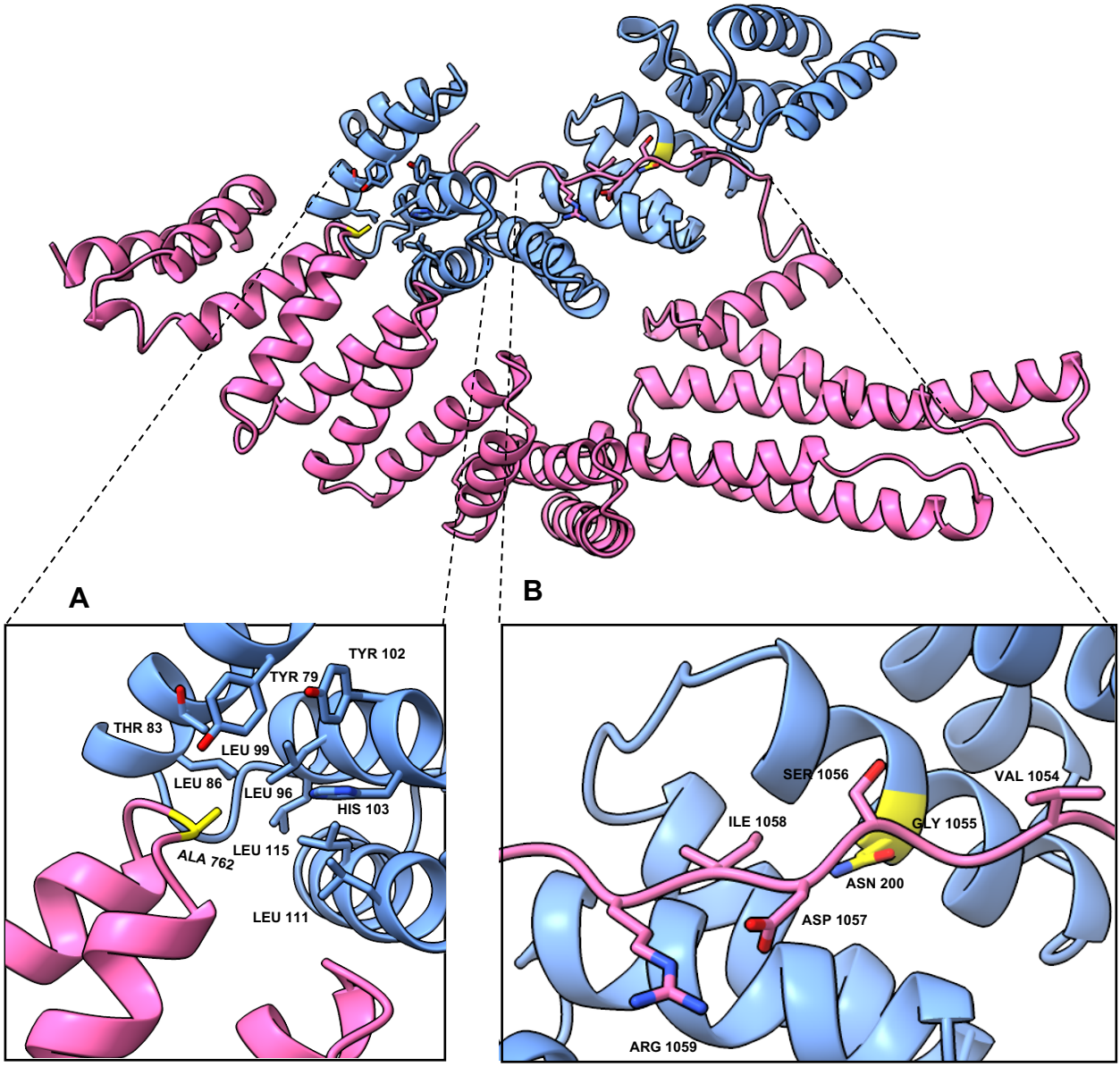


Figure 3.12: Location of GltD1$_{A762D}$ and GltE1$_{N200R}$ mutants on GltD1:GltE1 AF model
A – Ala762 (coloured yellow) is located on TPR15 of GltD1 and projects into a hydrophobic pocket formed between TPR1 and TPR2 of GltE1
B – Asn200 (coloured yellow) is located on TPR5 of GltE1 and interacts with the tail peptide of GltD1.

CD was used to assess the folding of GltD1$_{\Delta 1043\text{-}1066}$ to ensure that the observed loss of interaction with GltE1 was not a result of the deletion disrupting the overall folding of GltD1$_{\Delta 1043\text{-}1066}$. Comparison of the CD spectra of GltD1 and GltD1$_{\Delta 1043\text{-}1066}$ revealed that both proteins produce the characteristic CD spectra expected for α-helical rich proteins, with a large positive peak at ~195 nm followed by two negative peaks of a similar magnitude at ~208 nm and ~222 nm being observed (Figure 3.13). The similarities between the CD spectra for both samples demonstrate that GltD1$_{\Delta 1043\text{-}1066}$ maintains correct folding in the absence of the tail peptide.

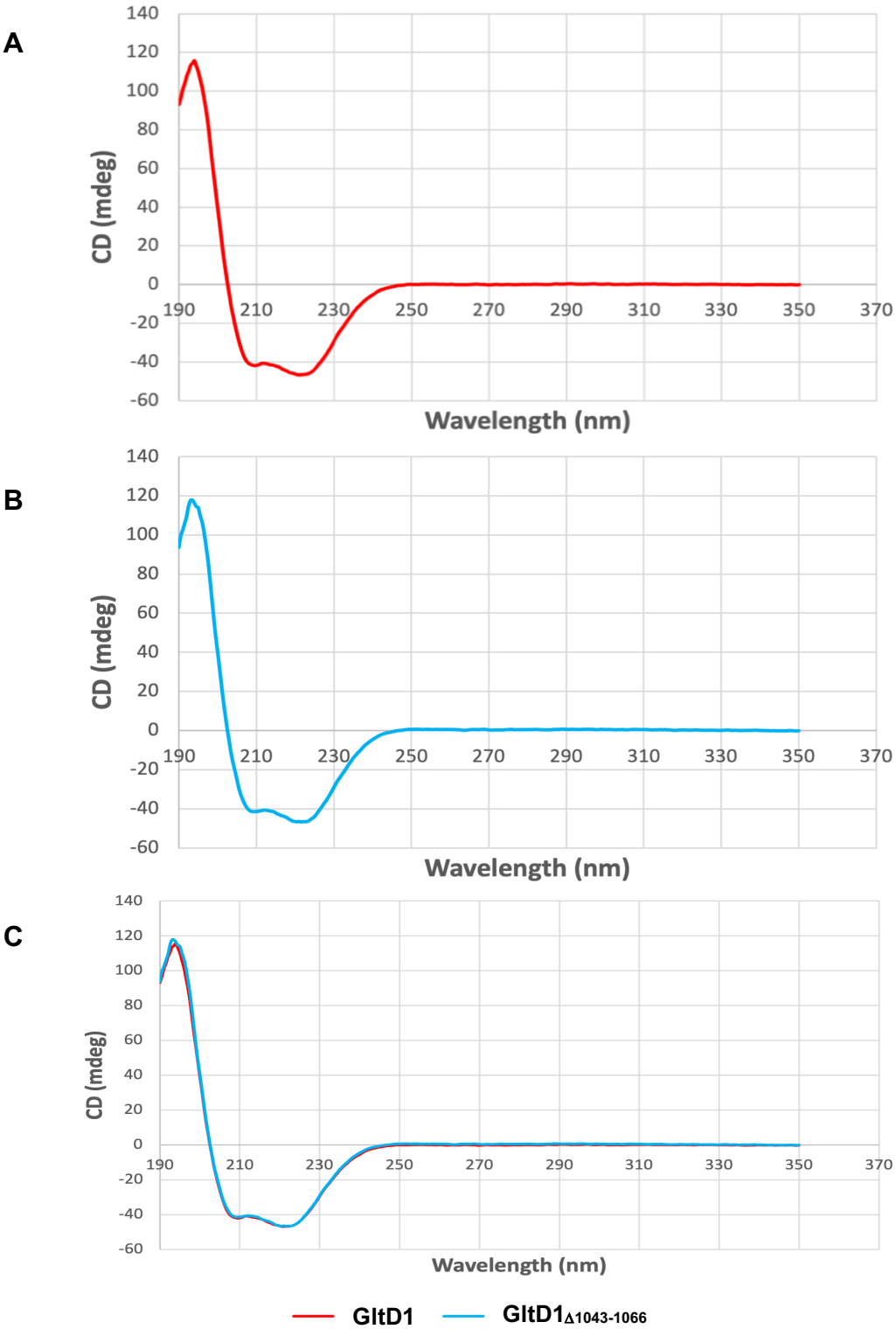


Figure 3.13: GltD1 CD spectra
A – CD spectra for GltD1: GltD1 produces a characteristic CD spectrum expected for α-helical rich proteins, with a large positive peak at ~195 nm followed by two negative peaks of a similar magnitude at ~208 nm and ~222 nm.
B – CD spectra for GltD1$_{\Delta 1043\text{-}1066}$: GltD1$_{\Delta 1043\text{-}1066}$ produces the same CD spectrum observed for GltD1, suggesting that GltD1$_{\Delta 1043\text{-}1066}$ folds correctly in the absence of the tail peptide.
C – Superimposition of GltD1 & GltD1$_{\Delta 1043\text{-}1066}$ CD spectra.

3.2.4 Crystallisation of GltE1

Attempts were made to crystallise GltD1 and GltE1 both individually and in complex. Initial sparse matrix screening was performed using a Mosquito system. Unfortunately, no crystals were grown for GltD1 or the full-length GltD1:GltE1 complex, possibly due to the fact that GltD1 is prone to degradation. Crystals for GltE1 were grown by sitting-drop vapour diffusion in the Memgold2 crystallisation screen. Two distinct crystal morphologies were observed (Figure 3.14): long thin rods (Crystal Form 1) and "layered teardrops" (Crystal Form 2). Data collection for GltE1 crystals was performed at Diamond Light Source beamline IO4.

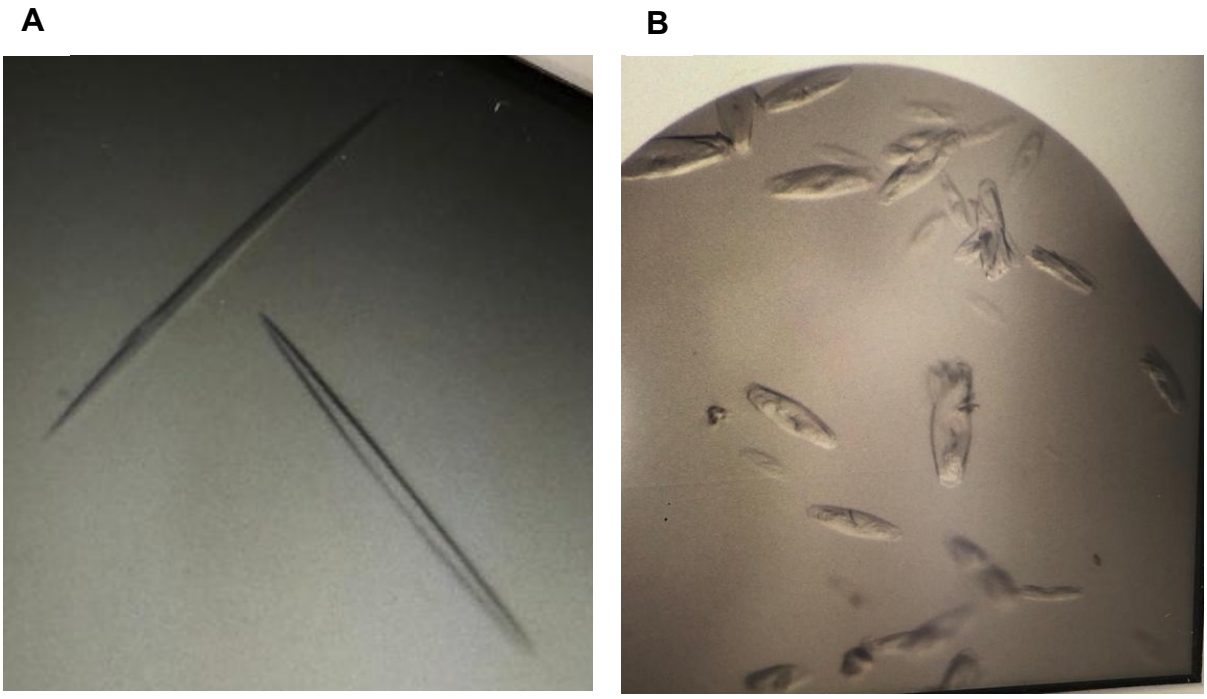


Figure 3.14: GltE1 crystals
A – GltE1 Crystal Form 1, crystallised in Memgold2 H7 (2.8 M Ammonium chloride, 0.075 M HEPES 7.5)
B – GltE1 Crystal Form 2, crystallised in Memgold2 H10 (0.1 M Tris pH 8.5, 44 % v/v PEG 200)

Table 3.2: Data collection and refinement statistics for GltE1 Crystal Form 1

Collection Statistics	
Wavelength	0.9795 Å
Resolution range	43.76 – 2.75 Å (2.848 – 2.75 Å)
Space group	R 3 2 :H
Unit cell	230.49 Å 230.49 Å 89.70 Å 90° 90° 120°
Total reflections	495244 (44910)
Unique reflections	23707 (2327)
Multiplicity	20.9 (19.3)
Completeness (%)	94.82 (65.08)
Mean I/sigma(I)	5.35 (0.36)
Wilson B-factor	70.26 $Å^2$
R-merge[a]	0.3524 (3.748)
R-meas[b]	0.3613 (3.85)
R-pim[c]	0.07896 (0.8753)
$CC^1/_2$	0.972 (0.29)
Refinement Statistics	
Reflections used in refinement	23707
Reflections used for R-free	1111
R-work[d]	0.2165
R-free[e]	0.2632
Number of non-hydrogen atoms	3688
macromolecules	3620
Ligands	34
Solvent	54
Protein residues	459
RMS(bonds)	0.007
RMS(angles)	1.01
Ramachandran favored (%)	95.82
Ramachandran allowed (%)	3.96
Ramachandran outliers (%)	0.22
Rotamer outliers (%)	1.89
Clashscore	21.78
Average B-factor	102.61 $Å^2$
Macromolecules	103.37 $Å^2$
Ligands	108.44 $Å^2$
Solvent	49.97 $Å^2$

Statistics for the highest-resolution shell are shown in parentheses.

a Rmerge = $\sum hkl \sum i |I_i(hkl) - \langle I(hkl)\rangle| / \sum hkl \sum i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the ith measurement of a reflection with indexes hkl and $\langle I(hkl)\rangle$ is the statistically weighted average reflection intensity.

b Rmeas is the Rmerge adjusted by a factor of $n(n-1)^{1/2}$ to account for multiplicity.

c R-pim = $\sum hkl [1/(n-1)]^{1/2} \sum i |I_i(hkl) - \langle I(hkl)\rangle| / \sum hkl \sum_i I_i(hkl)$ where $I_i(hkl)$ is the intensity of the ith measurement of a reflection with indexes hkl and $\langle I(hkl)\rangle$ is the statistically weighted average reflection intensity.

d R-work = $\sum ||Fo| - |Fc|| / \sum |Fo|$ where Fo and Fc are the observed and calculated structure factor amplitudes, respectively.

e R-free is the R-work calculated with 5% of the reflections chosen at random and omitted from the refinement.

3.2.4.1 Crystal Form 1 Structure Solution

Crystal Form 1 belongs to the R32 spacegroup and diffracted to a resolution of 2.75 Å. Matthews coefficient analysis indicated that two copies of GltE1 comprise the asymmetric unit (ASU). Initial phases were estimated by MR (PhaserMR)[163] using the AF model of $GltE1_{63-274}$ as an initial search model. Two copies of GltE1 (denoted chain A and B) were placed in the ASU with a final log-likelihood gain (LLG) of 447.25 and translation function Z-score (TFZ) of 22.5, indicating confident placement of the search model. The initial electron density map was poor for some regions, particularly the intra/inter TPR loops, and therefore a feature-enhanced map (FEM)[172] was generated to aid model building. Following successive rounds of model building and refinement, residues 61–290 of chain A and 63–291 of chain B were successfully modelled. All subsequent discussions will focus on chain A as a representative of both copies in the ASU.

3.2.4.2 The overall fold of GltE1 (Crystal Form 1)

The overall fold of GltE1 in Crystal Form 1 comprises 6 TPR motifs (designated TPR1-6), arranged into a superhelical conformation which is capped at the C-terminus by an additional α-helix (helix-13) that packs against hydrophobic residues lining the concave face of TPR-6 (Figure 3.15). Helix-13 is extended beyond the protein chain by the inclusion of an uncleaved TEV site (sequence: ENLYFQGLE). Each TPR is connected to the next by a four-residue loop. In general, these loops poorly fit the electron density and were modelled predominately through stereochemical restraints.

Each TPR-forming helix is of identical length, forming a uniform superhelical arrangement. However, the 6 TPR motifs of GltE1 differ from the consensus TPR sequence by varying degrees (Figure 3.16)[173]. Of the consensus residues, only the presence of a small hydrophobic residue at position 8 is conversed amongst all 6 TPR motifs of GltE1. In general, the most conserved consensus residues are Trp4 and the helix breaking Pro32 which interact to stabilise the fold[173]. This tryptophan is absent in all 6 TPR motifs, with leucine most frequently occupying this position. Similarly, Pro32 is only found in TPR1 and 2.

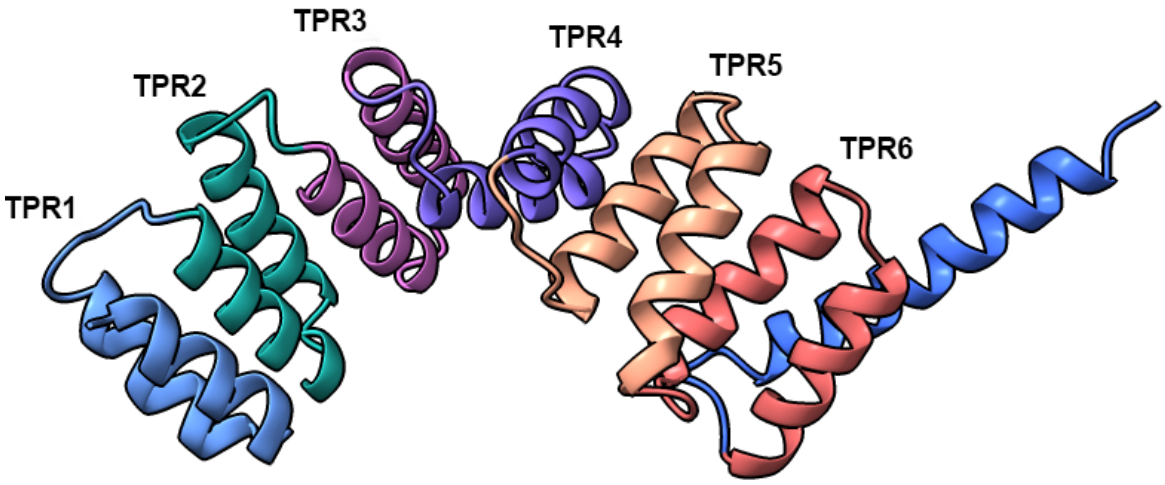


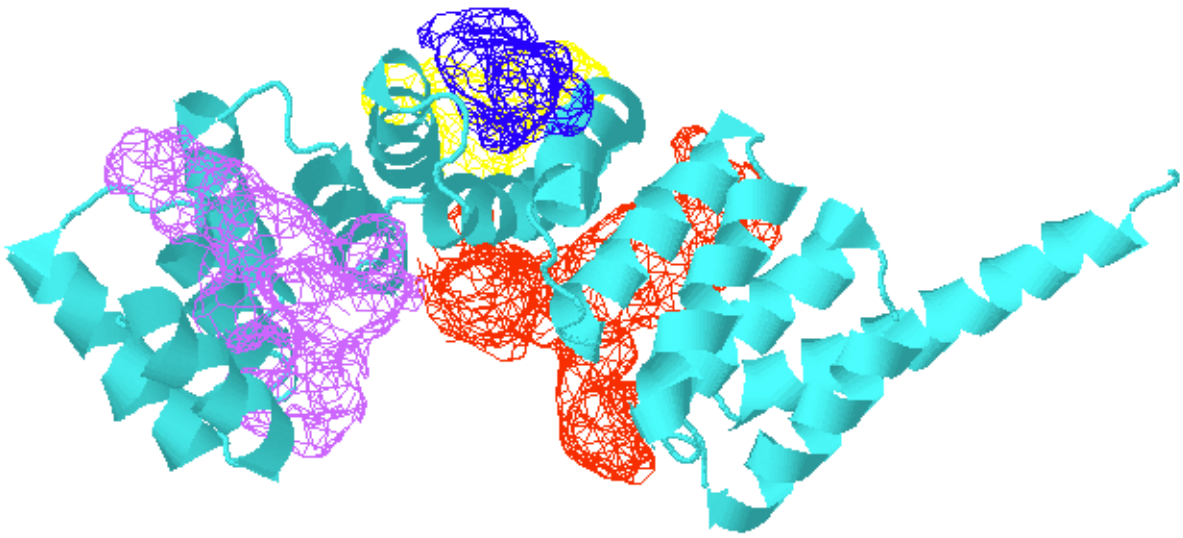
Figure 3.15: The crystal structure of GltE1 Crystal Form 1
Ribbon representation of GltE1. GltE1 comprises 6 TPR motifs (labelled TPR1–6) that form a superhelical architecture. For clarity, the structure has been coloured by each TPR.

```
Consensus_TPR    AEAWYNLGNCYYKM-GDYDEAIECYEK-ALELDPNN
TPR1             QYAALNEAIKSQSD-ER---IYQASTQ-ILTQSPND 92
TPR2             ARALNALAMYHYKR-GRFDLCRYLLGK-AISSSPKT 126
TPR3             AELYSNLGIVQLAQ-NERRDAVKSFRK-ALDINNDE 160
TPR4             AVAAANLGAIYVQE-RDFAKAGVVLET-AYRKGVRD 194
TPR5             PRVLNNYGIALTAQ-GKLDRAEDMYKA-ALKDSGNN 228
TPR6             KEVLFNYAILLVDHMQKYQDGLEVINRLKFVGG--P 262
```

Figure 3.16: Sequence comparison of GltE1 TPR motifs and consensus TPR sequence
Sequence alignment of the 6 TPR motifs of GltE1 against the consensus TPR sequence identified in Main *et al.,*[173]. Residues shown in green are TPR consensus residues. Residues shown in yellow match to any of the three most commonly occurring residues at each consensus position.

3.2.4.3 Putative ligand binding pockets

Crystal Form 1 of GltE1 was submitted to the PDBsum[174] server to calculate the volume of surface clefts (Figure 3.17). The server detected four main clefts distributed across the length of GltE1. The two largest clefts are located on the concave face of the superhelix (red and purple cleft in Figure 3.17), and likely form one extended cleft as opposed to the independent clefts identified by PDBsum. The concave face of GltE1 is lined with patches of hydrophobic residues. Residues from all 6 TPRs contribute to this interface, creating a continuous hydrophobic groove along the concave face (Figure 3.18). Notably, this hydrophobic groove is predominately occupied by the tail-peptide of GltD1 in the AF model of the GltD1:GltE1 complex.



Cleft 1: Volume = 2393.30 Å^3 (red)
Cleft 2: Volume = 1908.98 Å^3 (purple)
Cleft 3: Volume = 654.33 Å^3 (yellow)
Cleft 4: Volume = 687.66 Å^3 (blue)

Figure 3.17: PDBSum analysis of GltE1 Crystal Form 1
The structure of GltE1 Crystal Form 1 is shown in cyan. Clefts are represented as wireframe surfaces, coloured by individual cleft. PDBSum identifies four clefts on the surface of GltE1. Clefts 1 and 2 (coloured red and purple, respectively) are located on the concave face of GltE1 and likely form an extended cleft. Clefts 3 and 4 are located on the convex face of GltE1 (coloured yellow and blue, respectively).

The two remaining clefts identified by PDBsum are located in the space between TPR3 and TPR4 on the convex face of GltE1 (yellow and blue cleft in Figure 3.17). The convex surface is decorated with a stretch of positively charged residues that stems from TPR2 and follows the curve of the superhelix to TPR6, creating a continuous stretch of positive charge which could act as a binding site for one or more ligands (Figure 3.19). The region between TPR3 and TPR4 identified by PDBsum constitutes the region of the convex face with the densest collection of positive charge and closely associates with GltD1 in the AF model of the GltD1:GltE1 complex.

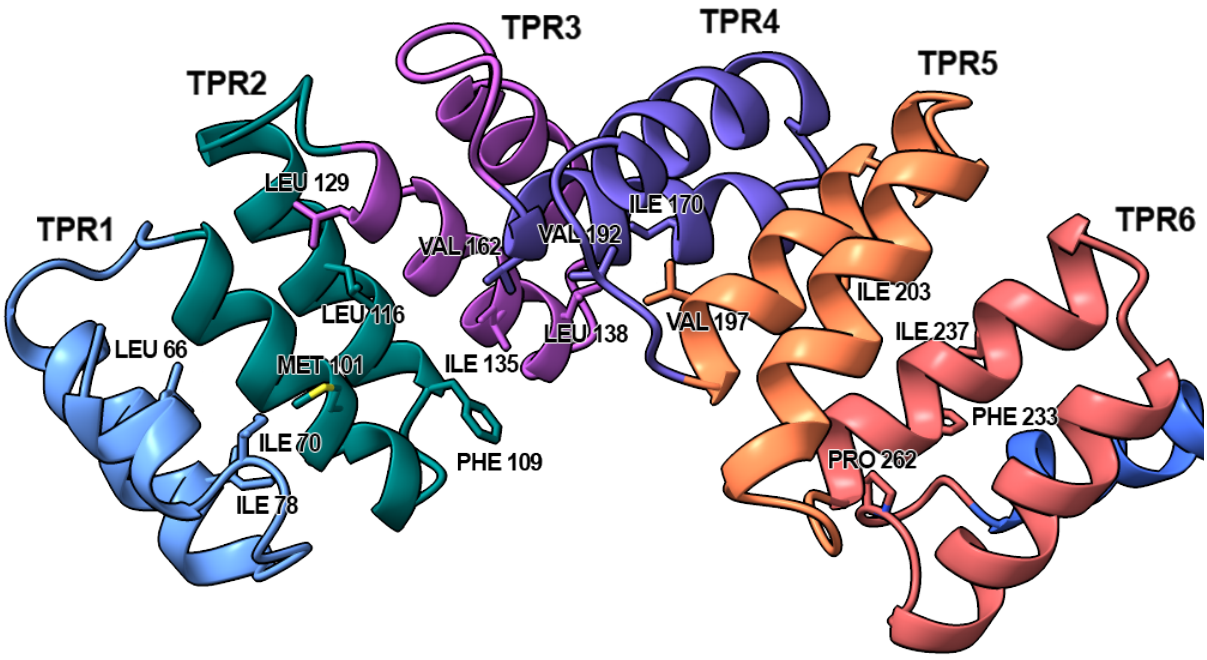


Figure 3.18: GltE1 contains a hydrophobic groove that lines the concave face
Hydrophobic residues from each of GltE1's TPR motifs create a continuous hydrophobic groove that lines the concave face of the GltE1 superhelix. This hydrophobic groove is predicted by AF to make interactions with the flexible tail peptide of GltD1. Hydrophobic residues lining the hydrophobic groove are shown as sticks. For clarity, the structure has been coloured by each TPR.

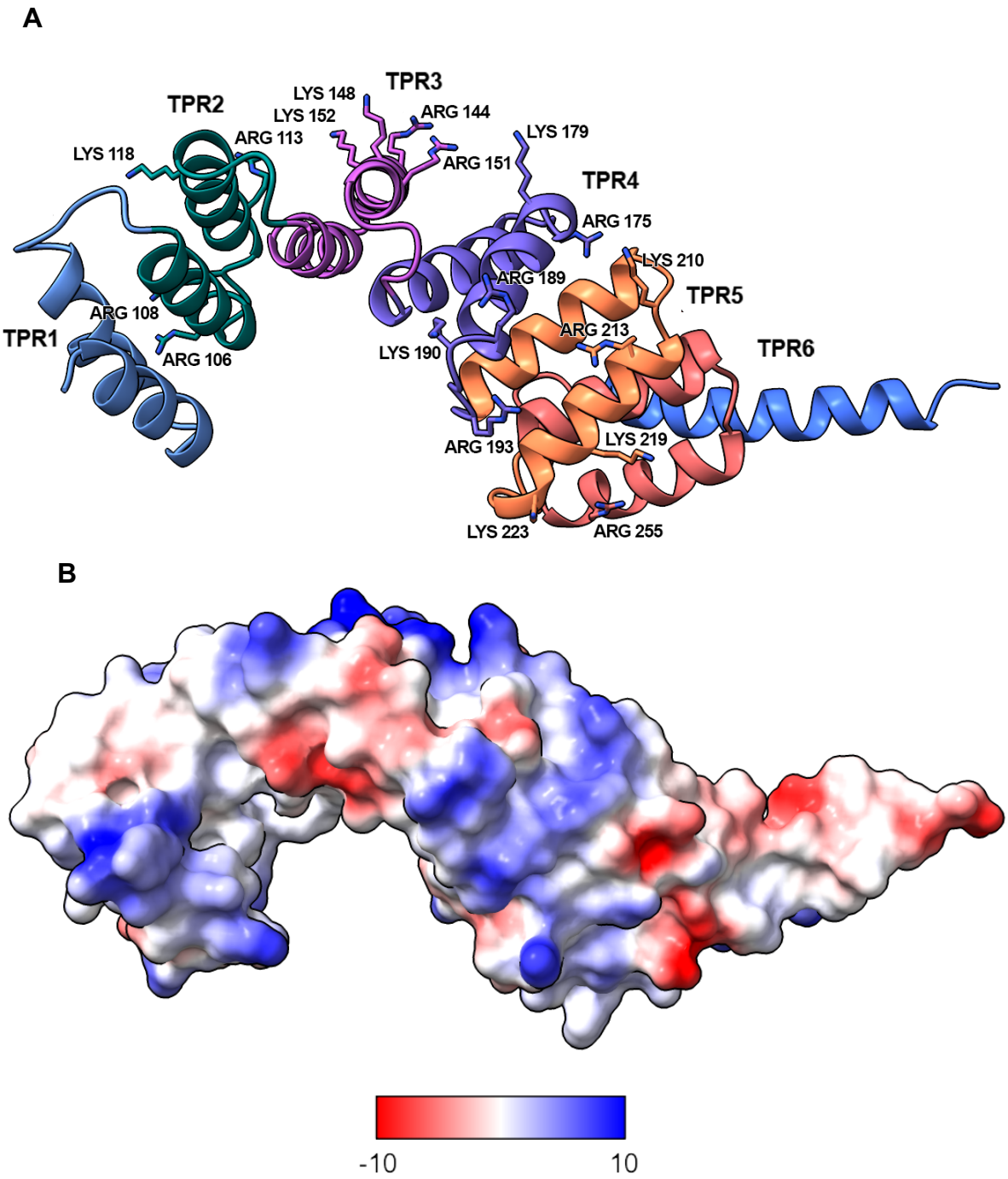


Figure 3.19: The convex face of GltE1 contains a stretch of positive charge that could house ligand interactions.
A – Ribbon representation of GltE1 with positively charged residues located on the convex face shown as sticks. For clarity, the structure has been coloured by each TPR.
B – Surface representation of GltE1 coloured according to coulombic electrostatic potential with negative and positive charge coloured in red and blue respectively. Surface representation created using ChimeraX.

Table 3.3: Data collection and refinement statistics for GltE1 Crystal Form 2

Collection Statistics	
Wavelength	0.9795 Å
Resolution range	53.37 – 2.928 Å (3.032 – 2.928 Å)
Space group	C 1 2 1
Unit cell	99.54 Å 68.55 Å 107.73 Å 90° 97.76° 90°
Total reflections	108628 (10610)
Unique reflections	15618 (1474)
Multiplicity	7.0 (7.2)
Completeness (%)	98.04 (84.97)
Mean I/sigma(I)	5.07 (0.42)
Wilson B-factor	80.46 Å^2
R-merge[a]	0.3168 (3.908)
R-meas[b]	0.3425 (4.211)
R-pim[c]	0.129 (1.562)
$CC^1/_2$	0.984 (0.273)
Refinement Statistics	
Reflections used in refinement	15618
Reflections used for R-free	720
R-work[d]	0.2764
R-free[e]	0.2850
Number of non-hydrogen atoms	3704
macromolecules	3697
Ligands	0
Solvent	7
Protein residues	472
RMS(bonds)	0.009
RMS(angles)	1.67
Ramachandran favored (%)	95.09
Ramachandran allowed (%)	4.27
Ramachandran outliers (%)	0.64
Rotamer outliers (%)	2.11
Clashscore	20.47
Average B-factor	99.50 Å^2
Macromolecules	99.57
Solvent	63.13

Statistics for the highest-resolution shell are shown in parentheses.

a Rmerge = $\sum hkl \sum i |I_i(hkl) - \langle I(hkl)\rangle| / \sum hkl \sum i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the ith measurement of a reflection with indexes hkl and $\langle I(hkl)\rangle$ is the statistically weighted average reflection intensity.
b Rmeas is the Rmerge adjusted by a factor of $n(n-1)^{1/2}$ to account for multiplicity.
c R-pim = $\sum hkl [1/(n-1)]^{1/2} \sum i |I_i(hkl) - \langle I(hkl)\rangle| / \sum hkl \sum_i I_i(hkl)$ where $I_i(hkl)$ is the intensity of the ith measurement of a reflection with indexes hkl and $\langle I(hkl)\rangle$ is the statistically weighted average reflection intensity.
d R-work = $\sum ||Fo| - |Fc|| / \sum |Fo|$ where Fo and Fc are the observed and calculated structure factor amplitudes, respectively.
e R-free is the R-work calculated with 5% of the reflections chosen at random and omitted from the refinement.

3.2.4.4 Crystal Form 2 Structure Solution

Crystal Form 2 of GltE1 belongs to the C2 spacegroup and diffracted to a resolution of 2.93Å. Matthews coefficient analysis indicated that two copies of GltE1 would be the most likely composition of the ASU. Once again, initial phases were estimated by MR (PhaserMR) using the AF model of $GltE1_{63–274}$ as an initial search model. During the initial MR, two copies of GltE1 were placed with a final LLG of 108.5 and TFZ of 8.7, corresponding to a moderately successful solution when compared to the solution for Crystal Form 1. Visual inspection of the MR solution revealed that N-terminal regions of the search model occupied areas that were absent of any electron density and clashed with symmetry copies, indicative of GltE1 occupying an alternative conformation in Crystal Form 2. Splitting the initial search model into halves ($GltE1_{64–143}$ and $GltE1_{144–274}$) improved the placement, with two copies of each half (denoted chains A-D, subsequently merged into complete chains A and B) being placed with a final LLG of 507.8 and a TFZ of 15.5. The search models fit reasonably well to the electron density and no longer clashed with symmetry copies. As with Crystal Form 1, a FEM map was generated to aid model building and after successive rounds of model building and refinement, residues 54–291 of chain A and 56–289 of chain B were successfully modelled.

3.2.4.5 The overall fold of GltE1 (Crystal Form 2)

In Crystal Form 2, GltE1 adopts an unconventional conformation deviating from the expected TPR superhelical fold (Figure 3.20). The conformational change is largely driven by the reorientation of residues 157–160, which, in Crystal Form 1 form a flexible loop that connects TPR3 and TPR4. In Crystal Form 2, however, these residues form an additional helical turn that causes helix B of TPR3 and helix A of TPR4 to form a single extended helix. The presence of this extended helix causes TPR1–3 of Crystal Form 2 to flip approximately 180° relative to their position in Crystal Form 1, disrupting the superhelical nature of the fold and elongating the concave face.

3.2.4.6 Ligand binding

The reorientation of TPR1-3 creates a large cleft into which the C-terminus of the second ASU copy of GltE1 packs (Figure 3.21). This interaction is mediated by the packing of the uncleaved TEV site into hydrophobic pockets formed along the length of TPR1, TPR2 and TPR3 (Figure 3.22). The first pocket is formed by Ile135, Leu138 and Phe150 of TPR3, into which Leu284 of the TEV sequence packs. A second pocket formed between Leu129, Ile135, Val136 of TPR3 and Ala100, Met101, Tyr104 of TPR2 interacts with Phe286 of the TEV sequence. The interaction is likely further stabilised by the packing of Gly288 against Ile70 and Met101 as well as hydrogen bonding between Asn97/Asn132 and Gln287. The interaction of the TEV site with GltE1 shares striking similarities to the AF model of the GltD1:GltE1 interaction.

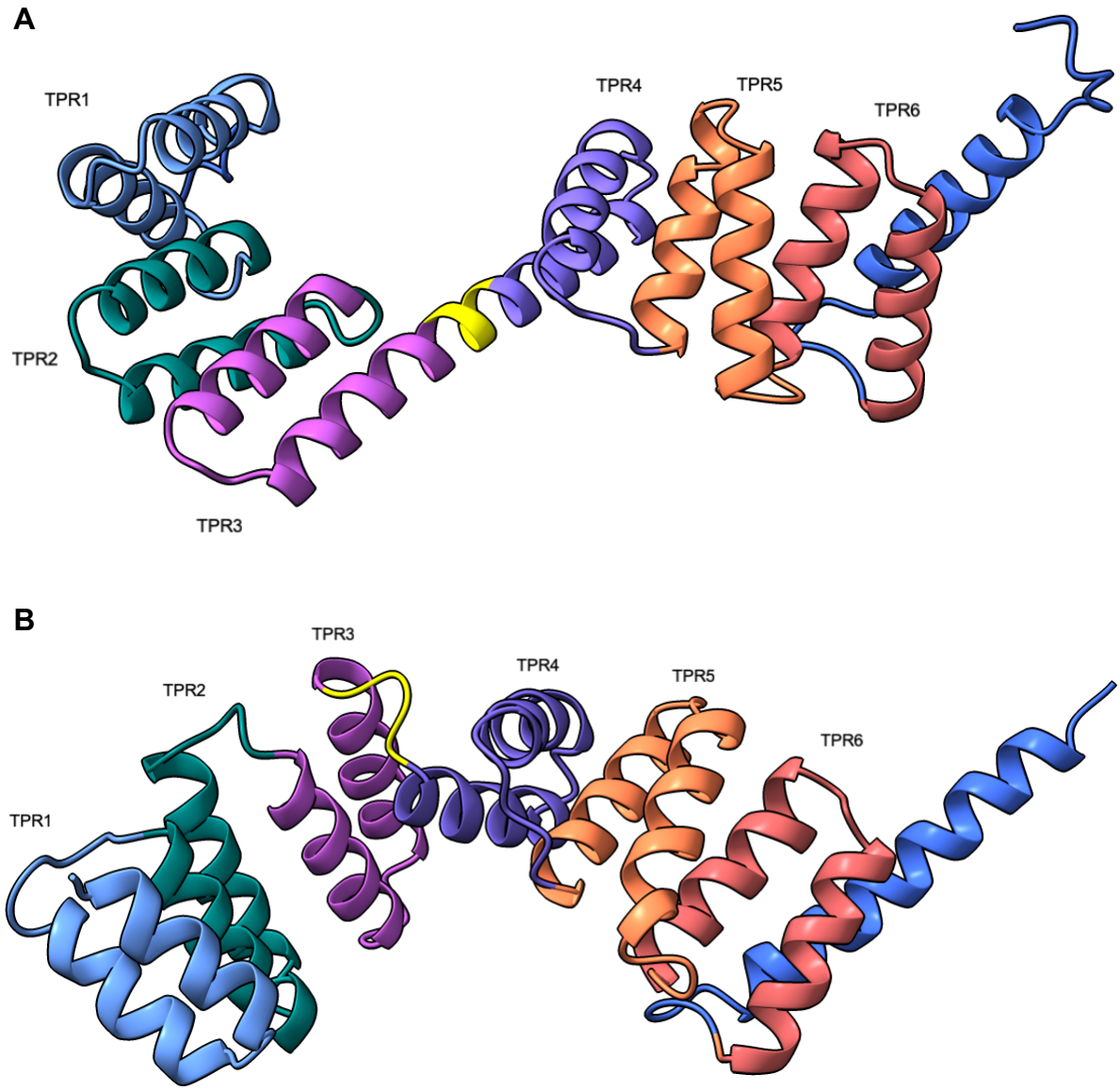


Figure 3.20: Crystal structure of GltE1 Crystal Form 2
A – Crystal structure of Crystal Form 2 coloured by each TPR. In Crystal Form 2, residues 157–160 (coloured yellow) form a helical turn as opposed to an inter-TPR loop. This results in the formation of a single, elongated α-helix formed from helix B of TPR3 and helix A of TPR4. The presence of this helix results in the reorientation of TPR1–3.
B – GltE1 Crystal Form 1 with residues 157–160 coloured yellow.

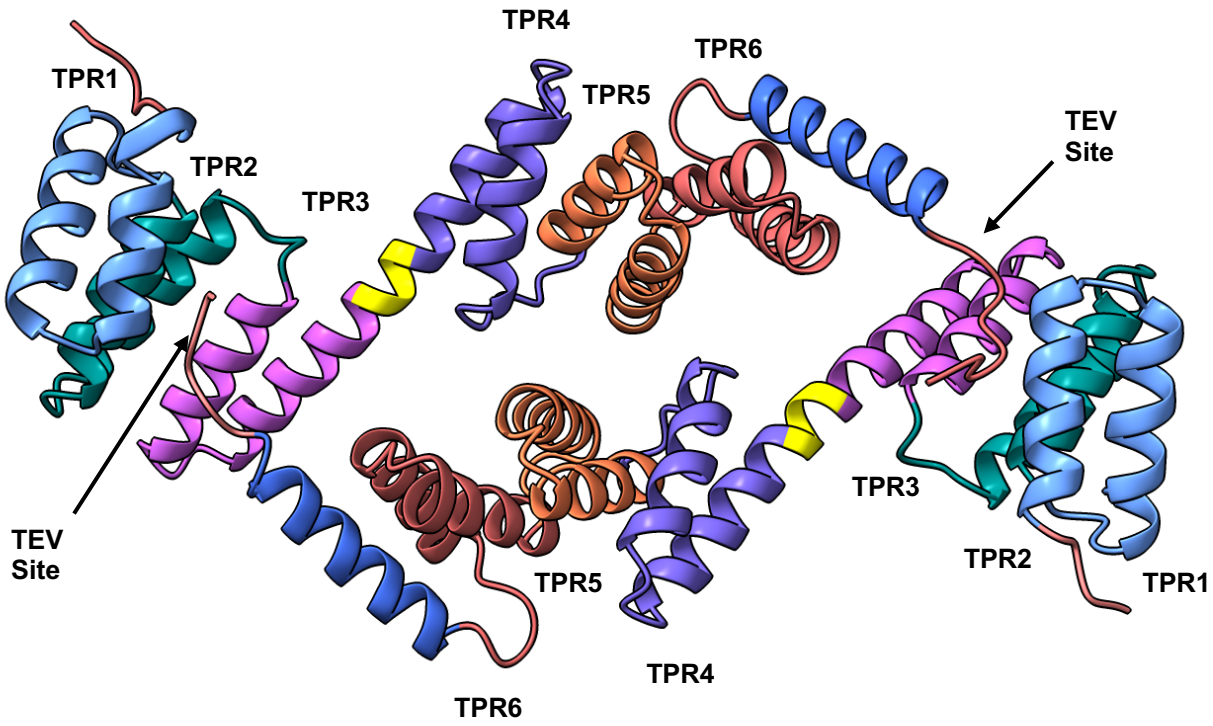


Figure 3.21: Packing of ASU copies of GltE1 Crystal Form 2
In GltE1 Crystal Form 2, the reorientation of TPR1–3 allows an uncleaved TEV site (indicated by the black arrow) of one ASU copy to pack into hydrophobic pockets formed along the concave face of TPR1, TPR2 and TPR3 of a second ASU copy. For clarity, the structure has been coloured by each TPR.

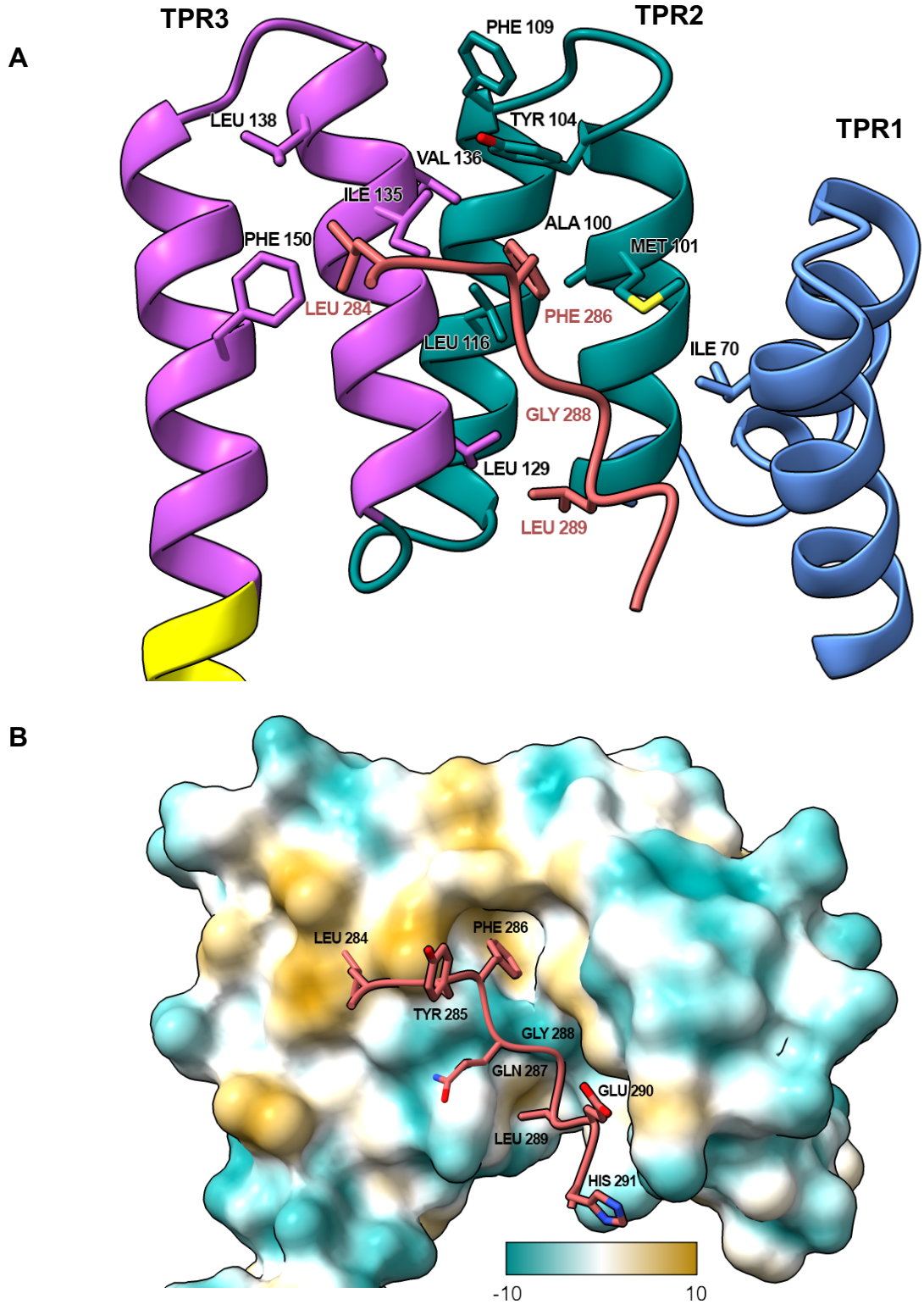


Figure 3.22: Interaction between ASU copies of GltE1 Crystal Form 2
A – Ribbon representation of the TEV site (coloured red) of one ASU copy of GltE1 Crystal Form 2 packing into a hydrophobic pocket formed along the concave face of TPR1–3 (coloured blue, green and pink respectively) of a second ASU copy. Hydrophobic residues involved in the interaction are shown as sticks. TEV site residue labels are coloured red for clarity.
B – Surface representation of the TEV site (coloured red) of one ASU copy packing into a hydrophobic pocket formed along the concave face of TPR1–3 of a second ASU copy. GltE1 surface coloured according to hydrophobicity with hydrophobic and hydrophilic residues coloured gold and cyan respectively. Surface representation created using ChimeraX.

3.2.4.7 GltE1 Crystal Form 2 exhibits a pseudodomain swap

The conformational change observed in Crystal Form 2 is likely an artefact of crystallisation and reflects the inherent flexibility of TPR domains. Notably, a pseudo-domain swap occurs between each ASU copy and a symmetry related molecule (Figure 3.23). This pseudo-domain swap is made possible due to the presence of the extended helix linking TPR3 and TPR4 which results in TPR1 from one chain packing against TPR4 of a second chain. The interaction results in the arrangement of the 6 TPR motifs in a more typical superhelical arrangement that more closely resembles Crystal Form 1, demonstrating the propensity of an array of TPR motifs to form a superhelical fold; albeit the chain runs in the opposite direction with each pseudo-domain swap adopting the arrangement TPR3-TPR2-TPR1-TPR4-TPR5-TPR6.

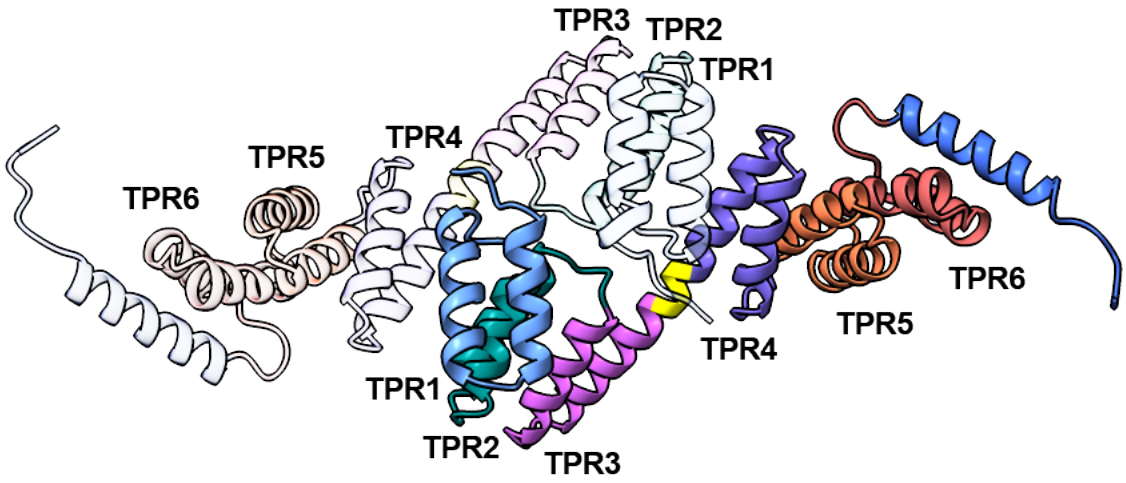


Figure 3.23: GltE 1 Crystal Form 2 pseudodomain swap
GltE1 Crystal Form 2 exhibits a pseudodomain swap which sees TPR1 of one ASU copy pack against TPR4 on a symmetry related molecule forming a conformation that more closely resembles the superhelical architecture of Crystal Form 1. The structure has been coloured by TPR and the symmetry copy is shown transparent.

3.2.5 Crystallisation of GltD1:GltE1 complex

Following unsuccessful attempts to crystallise the GltD1:GltE1 complex, we turned our attention to single-particle cryo-EM to obtain structural information for the full-length complex. Following several rounds of grid preparation and optimisation, it became apparent that issues with the heterogeneity of the sample, likely caused by GltD1's proneness to degradation, would hinder attempts to obtain a structure by cryo-EM (Figure 3.24).

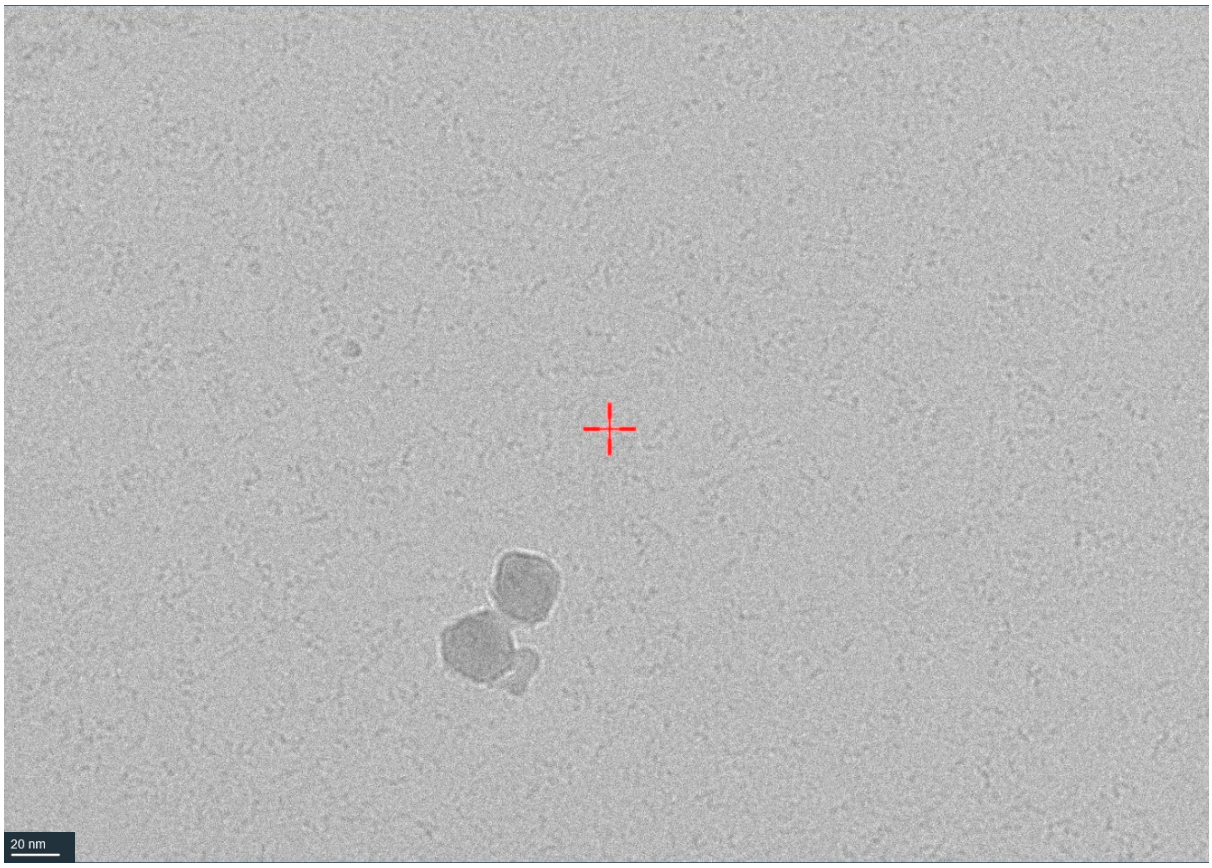


Figure 3.24: Micrograph of GltD1:GltE1 complex
Micrographs of GltD1-GltE1 complex at 0.3 mg/mL. Sample was plunge-frozen on Quantifoil R1.2/1.3 Gold Holey Carbon Grids. The grids were imaged on a Titan Krios and reveal a heterogenous sample likely caused by the degradation of GltD1.

In order to improve the chances of crystallisation, a truncated GltD1 construct was designed. The boundaries of this construct were guided by the AF model of the GltD1:GltE1 complex and were chosen to give the smallest possible span of GltD1 whilst still maintaining the entire interaction interface. $GltD1_{706–1066}$ ($GltD1_{Cterm}$) expressed well, was readily purified and maintained its interaction with GltE1 when tested by ITC (Figures 3.25 and 3.26).

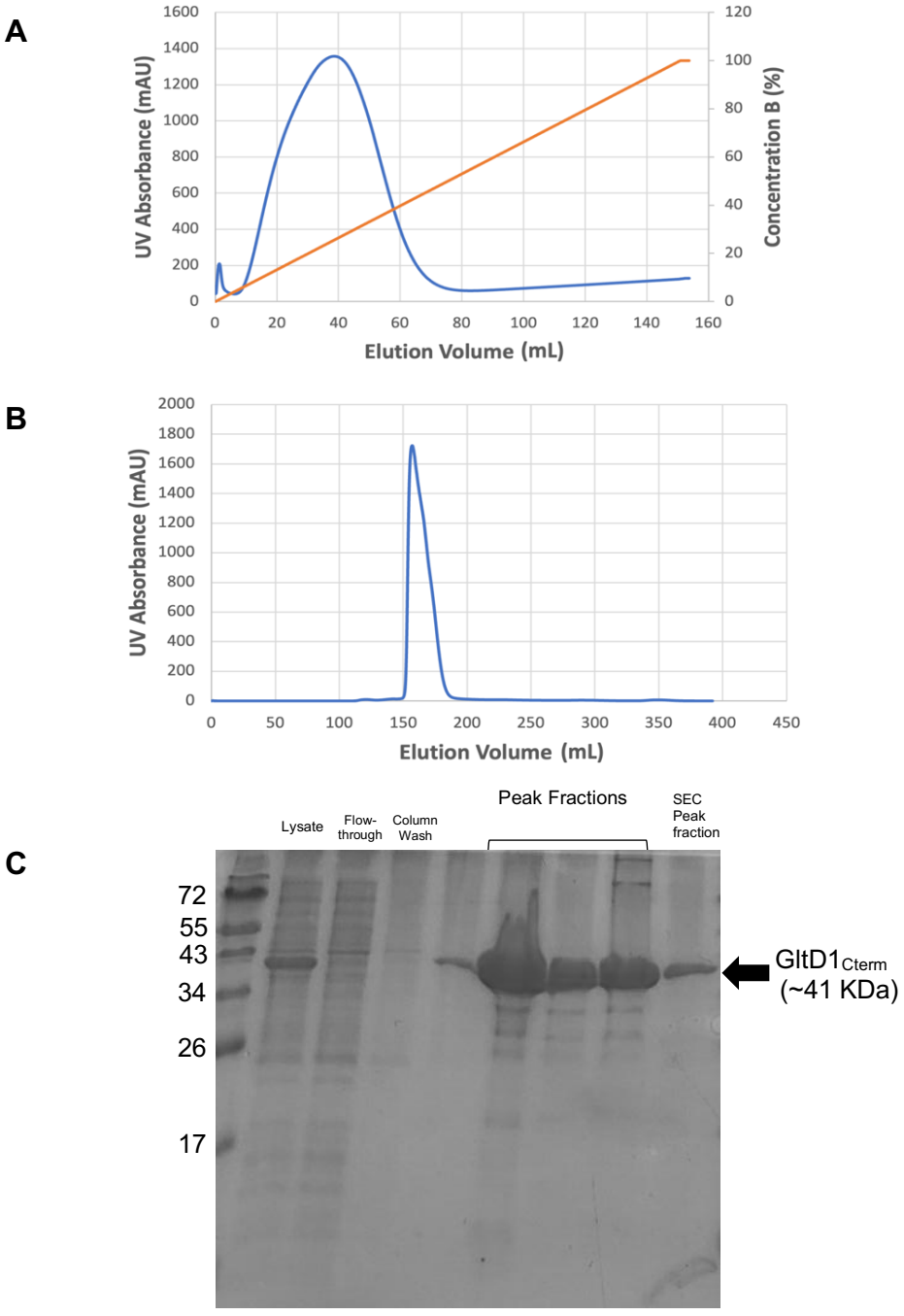


Figure 3.25: Purification of GltD1$_{Cterm}$
A – Chromatogram of GltD1$_{Cterm}$ IMAC purification. IMAC performed on a 5 mL HisTrap HP using a gradient elution of 0 –100% buffer B across 30 CV. Curves for UV absorbance at 280 nM (blue trace) and buffer B concentration (orange trace) are shown.
B – SEC Chromatogram for GltD1$_{Cterm}$. GltD1$_{Cterm}$ was eluted as a monodisperse peak from a HiLoad 26/600 Superdex 75 prep grade column at 158 mL, an elution volume consistent with the size of monomeric GltD1$_{Cterm}$ (~41 KDa). Curve for UV absorbance at 280 nM shown in blue.
C – SDS-PAGE analysis of IMAC and SEC peak fractions. GltD1$_{Cterm}$ can be seen as a predominant band between the 34 and 43 kDa markers. The gel reveals the presence of several contaminants following IMAC. The gel also reveals that the majority of contaminants are removed following SEC. Black arrow indicates the position of GltD1$_{Cterm}$.

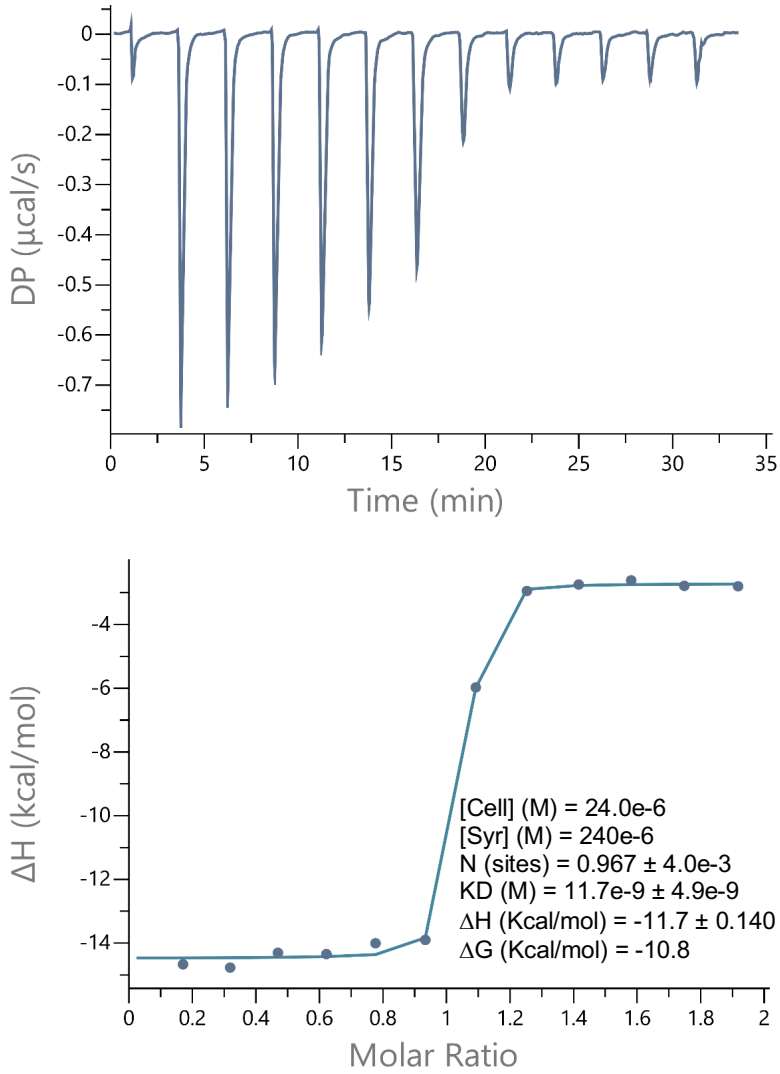
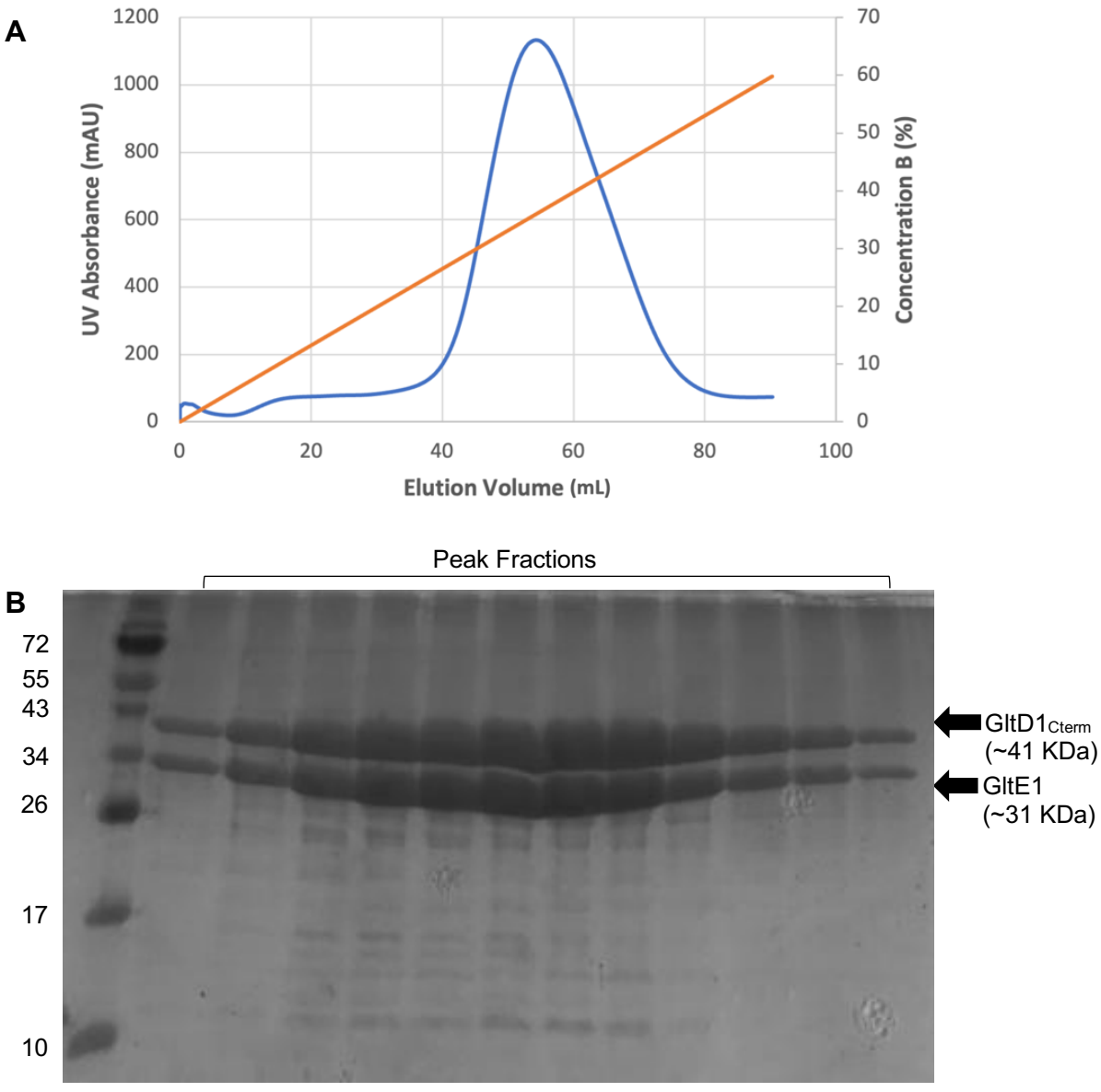


Figure 3.26: ITC binding for GltD1$_{Cterm}$:GltE1 interaction
Binding thermogram for GltD1$_{Cterm}$-GltE1 interaction. The two proteins bind in a 1:1 ratio with nanomolar affinity.

GltD1$_{Cterm}$ and GltE1 were co-purified for crystallisation trials. Separate cell pellets expressing each protein were simultaneously lysed and the mixed lysate was used during subsequent purification steps. GltD1$_{Cterm}$ and GltE1 were observed to co-elute as a single peak during IMAC (Figure 3.27). The GltD1$_{Cterm}$ and GltE1 containing fractions were pooled and further purified by SEC. GltD1$_{Cterm}$ and GltE1 eluted from a HiLoad 26/600 Superdex 200 prep grade column as a large peak at ~194 mL. SDS-PAGE demonstrated that both proteins were present in these fractions (Figure 3.27).



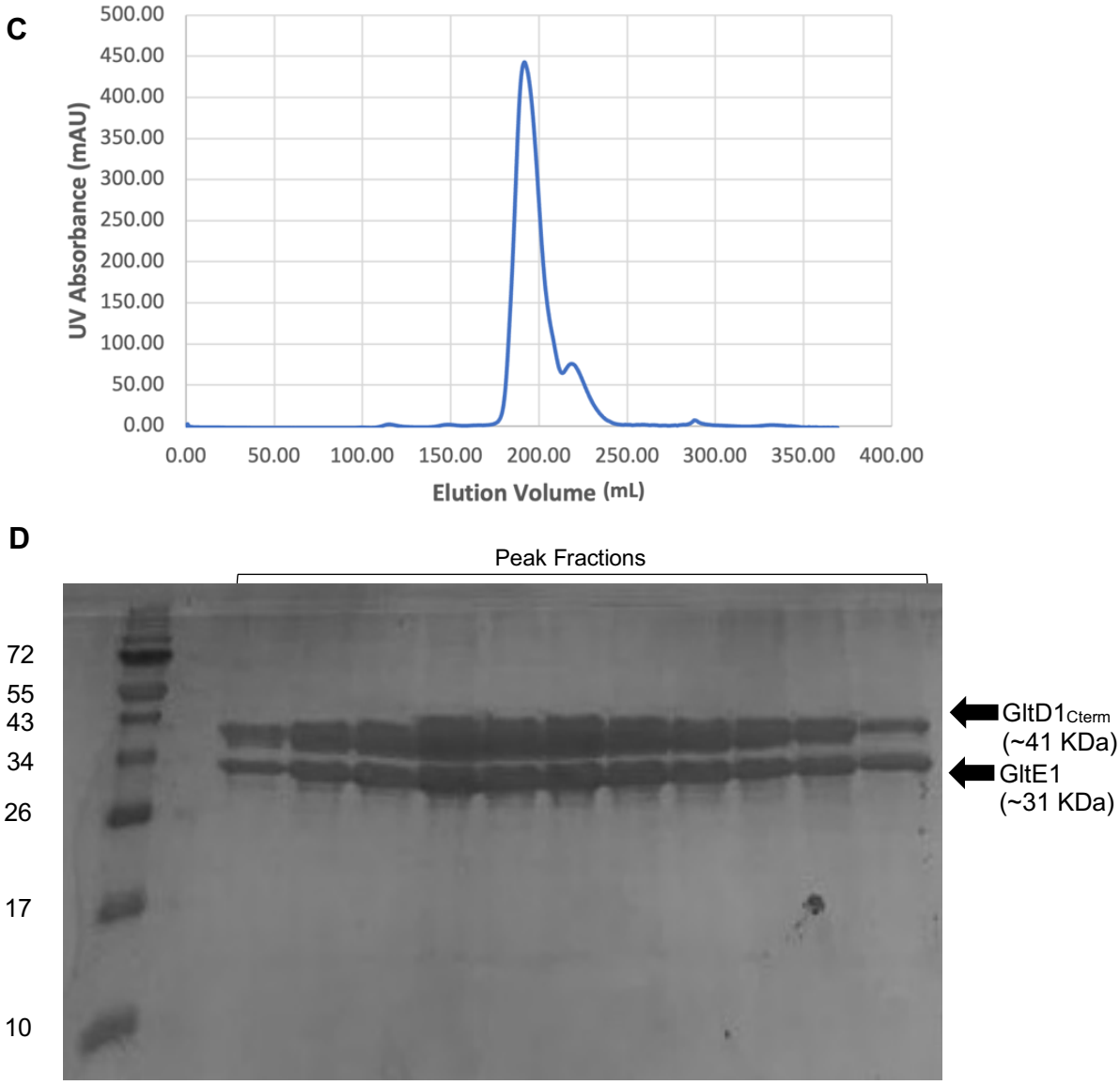


Figure 3.27: GltD1$_{Cterm}$:GltE1 co-purification
A – Chromatogram of GltD1$_{Cterm}$:GltE1 IMAC co-purification. IMAC performed on a 5 mL HisTrap HP using a gradient elution of 0-100% buffer B across 30 CV. Curves for UV absorbance at 280 nM (blue trace) and buffer B concentration (orange trace) are shown.
B – SDS-PAGE analysis of IMAC purification. GltD1$_{Cterm}$ can be seen as a predominant band between the 34 and 43 KDa MW markers. GltE1 can be seen as a predominant band between the 26 and 34 kDa MW markers. The gel reveals the presence of some lower MW contaminants. The black arrows indicate the positions of GltD1$_{Cterm}$ and GltE1.
C – SEC Chromatogram for GltD1$_{Cterm}$:GltE1. GltD1$_{Cterm}$ and GltE1 were eluted as a peak from a HiLoad 26/600 Superdex 200 prep grade column at 194mL, an elution volume consistent with GltD1$_{Cterm}$ and GltE1 interacting in a 1:1 stoichiometry. Curve for UV absorbance at 280 nM shown in blue.
D – SDS-PAGE analysis of SEC peak fractions. GltD1$_{Cterm}$ can be seen as a predominant band between the 34 and 43 KDa MW markers. GltE1 can be seen as a predominant band between the 26 and 34 kDa MW markers. The gel reveals that GltD1$_{Cterm}$ and GltE1 co-elute during SEC and shows that the majority of contaminants have been removed following SEC. The black arrows indicate the positions of GltD1$_{Cterm}$ and GltE1.

Crystals for the GltD1$_{Cterm}$:GltE1 complex were grown by sitting-drop vapour diffusion in Memgold2 E11 (Figure 3.28). These crystals belong to the P 6 2 2 spacegroup and diffracted to a resolution of 3.11Å. Matthews coefficient analysis indicated that a single copy of the GltD1$_{Cterm}$:GltE1 complex would be the most likely composition of the ASU. Initial phases were estimated by MR using Crystal Form 1 of GltE1 and the AF model of GltD1$_{Cterm}$ as search models. One copy of each search model was placed with a final LLG of 561.4 and TFZ of 22.0, demonstrating good placement of the search models. As seen for both structures of GltE1, the initial electron density map was poor in some regions therefore a FEM map was generated to aid model building. Following successive rounds of model building and refinement, residues 725–1066 of GltD1 and 61–285 of GltE1 were successfully modelled.

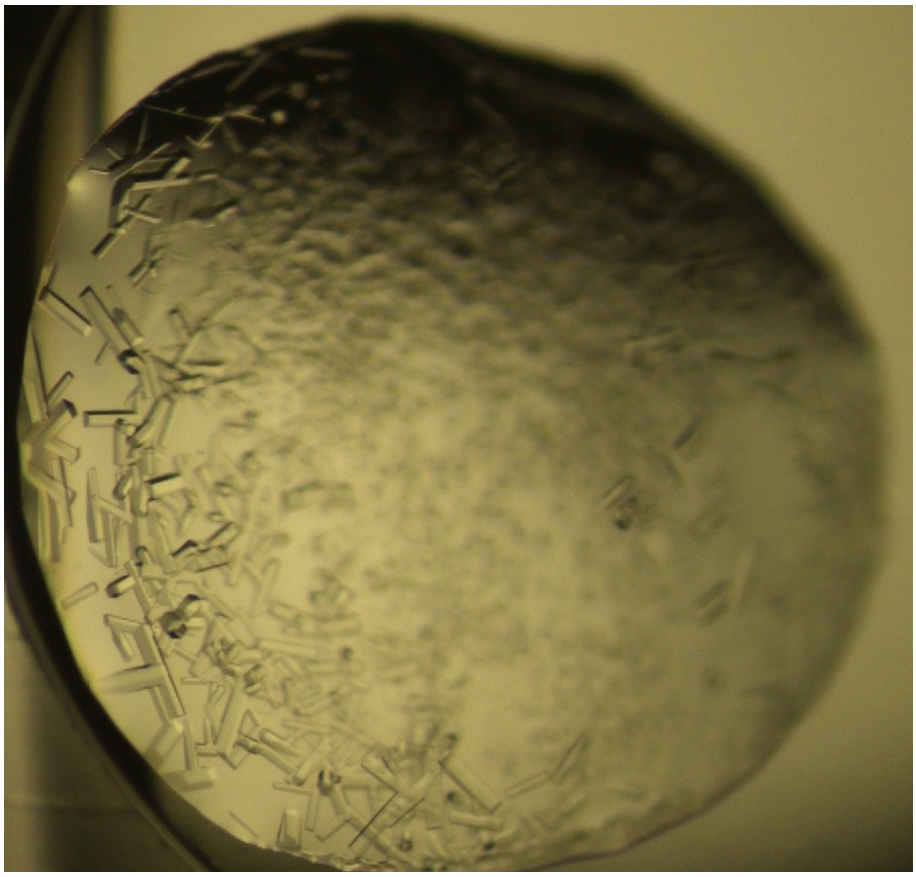


Figure 3.28: GltD1$_{Cterm}$:GltE1 crystals
GltD1$_{Cterm}$:GltE1 crystals, crystallised in 0.25 M Magnesium formate dihydrate 0.1 M Sodium cacodylate pH 6.5 22 % w/v PEG 3000.

Table 3.4: Data collection and refinement statistics for GltD1:GltE1 complex

Collection Statistics	
Wavelength	0.6888 Å
Resolution range	50.95 – 3.11 Å (3.221 – 3.11 Å)
Space group	P 6 2 2
Unit cell	155.64 Å 155.64 Å 130.91 Å 90° 90° 120°
Total reflections	430157 (23265)
Unique reflections	17271 (1646)
Multiplicity	24.9 (14.1)
Completeness (%)	98.13 (88.32)
Mean I/sigma(I)	7.42 (0.41)
Wilson B-factor	94.37 Å^2
R-merge[a]	0.3141 (2.808)
R-meas[b]	0.3205 (2.914)
R-pim[c]	0.06243 (0.7562)
$CC^1/_2$	0.998 (0.405)
Refinement Statistics	
Reflections used in refinement	17271
Reflections used for R-free	845
R-work[d]	0.2581
R-free[e]	0.3047
Number of non-hydrogen atoms	4426
macromolecules	4426
Ligands	0
Solvent	0
Protein residues	565
RMS(bonds)	0.004
RMS(angles)	0.64
Ramachandran favoured (%)	93.74
Ramachandran allowed (%)	6.08
Ramachandran outliers (%)	0.18
Rotamer outliers (%)	0.89
Clashscore	10.04
Average B-factor	117.27 Å^2
Macromolecules	117.27 Å^2

Statistics for the highest-resolution shell are shown in parentheses.

a Rmerge = $\sum hkl \sum i |I_i(hkl) - \langle I(hkl)\rangle| / \sum hkl \sum i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the ith measurement of a reflection with indexes hkl and $\langle I(hkl)\rangle$ is the statistically weighted average reflection intensity.

b Rmeas is the Rmerge adjusted by a factor of $n(n-1)^{1/2}$ to account for multiplicity.

c R-pim = $\sum hkl [1/(n-1)]^{1/2} \sum i |I_i(hkl) - \langle I(hkl)\rangle| / \sum hkl \sum_i I_i(hkl)$ where $I_i(hkl)$ is the intensity of the ith measurement of a reflection with indexes hkl and $\langle I(hkl)\rangle$ is the statistically weighted average reflection intensity.

d R-work = $\sum ||Fo| - |Fc|| / \sum |Fo|$ where Fo and Fc are the observed and calculated structure factor amplitudes, respectively.

e R-free is the R-work calculated with 5% of the reflections chosen at random and omitted from the refinement.

3.2.5.1 The overall fold of GltD1$_{Cterm}$:GltE1

In the GltD1$_{Cterm}$:GltE1 structure, GltE1 adopts the same conformation as in Crystal Form 1 (RMSD = 0.69 Å), where the 6 TPR motifs form a superhelical architecture (Figure 3.28). GltD1$_{Cterm}$ comprises 5 TPR motifs, although only helix B of TPR14 could be built (TPRs are numbered as TPR14–18 based on the numbering of TPRs as denoted in the full-length AF model of GltD1). The TPR repeats of GltD1$_{Cterm}$ extend into a four-helix bundle which leads into the C-terminal tail peptide (Figure 3.29). As with GltE1, the TPRs of GltD1$_{Cterm}$ deviate from the consensus TPR sequence by varying degrees (Figure 3.30).

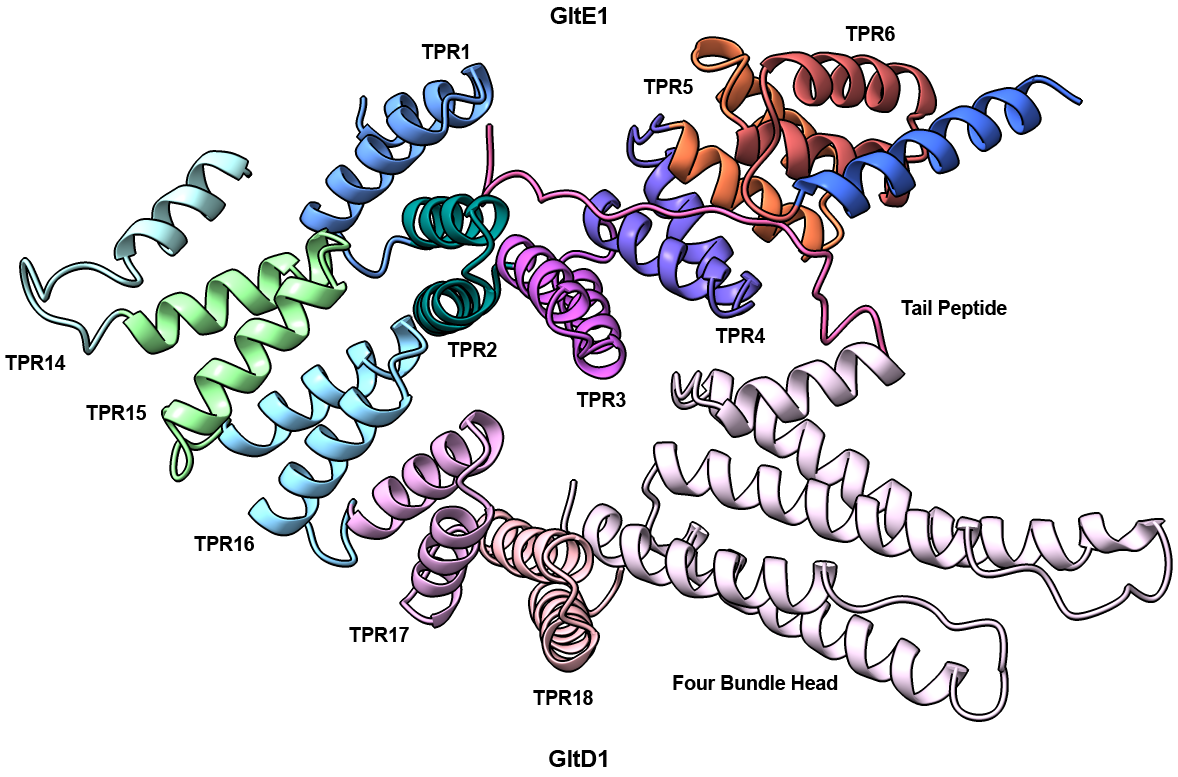


Figure 3.29: GltD1$_{Cterm}$:GltE1 crystal structure
The overall fold of the GltD1$_{Cterm}$:GltE1 complex. GltE1 interacts with the TPR superhelix, four bundle head and tail peptide of GltD1. The tail peptide (shown in pink) of GltD1 packs against the concave face of GltE1. GltD1 and GltE1 (labelled TPR 1–6) structures are coloured by TPR and the TPRs of GltD1$_{Cterm}$ are numbered based on the TPR numbering used in the full-length GltD1 AF model (TPR 14–18).

```
Consensus_TPR    AEAWYNLGNCYYKMGDYDEAIECYEKALELD-PN-----N
TPR14            VNANFNAAVNYERAGMNAPAIAAYQGVLASKDPAADK-LK 747
TPR15            PKARRLLAKLYQDSAQFEEAAKLYRQAAQEN-PT-DS--L 783
TPR16            PNLMFNAAVLYEALGKSDEAIRSYTDFTKMN-KK-HS--D 820
TPR17            VEAVFSMAQIHRKAGQTGAAIARYMEYVEAG-GR-----D 855
TPR18            VESAYWVSELSTRQKATTRAQEWRQKTLAIQ-KR-FAPNK 897
```

Figure 3.30: Sequence comparison of GltD1$_{Cterm}$ TPR motifs and consensus TPR sequence
Sequence alignment of the 5 TPR motifs of GltD1$_{Cterm}$ against the consensus TPR sequence identified in Main *et al.,*[173]. Residues shown in green are TPR consensus residues. Residues shown in yellow match to any of the three most commonly occurring residues at each consensus position.

3.2.5.2 GltD1$_{Cterm}$:GltE1 Interface

The GltD1$_{Cterm}$:GltE1 interface is extensive with a buried surface area of 2163 Å^2, as calculated by PISA[168]. The interaction is maintained by 12 hydrogen bonds, five salt bridges and extensive hydrophobic packing. Full details of the hydrogen bonding and salt bridging can be seen in Tables 3.5 and 3.6. Consistent with the AF model, our ITC data and mutational results, the interaction predominantly occurs between the tail peptide of GltD1 and the concave face of GltE1. The tail peptide threads along the hydrophobic groove that occupies the entirety of the concave face, tracing along the curve of the GltE1 superhelix (Figure 3.31). The hydrophobic nature of the interaction provides an explanation as to why only a small exothermic signal is measured during ITC experiments. To stabilise the interaction, hydrophobic residues from TPR15 and TPR16 of GltD1 pack into a hydrophobic pocket formed between TPR1 and TPR2 of GltE1. This interaction is further stabilised by hydrogen bonding between Tyr79 located on TPR1 of GltE1 and Ala762 located on TPR15 of GltD1. Numerous hydrogen bonds and salt bridges between the four-helix bundle of GltD1 and TPR3 and TPR4 of GltE1 provide additional stability at the interface (Figure 3.32). Given that the deletion of the tail peptide completely abolished the interaction between GltD1 and GltE1, it is likely

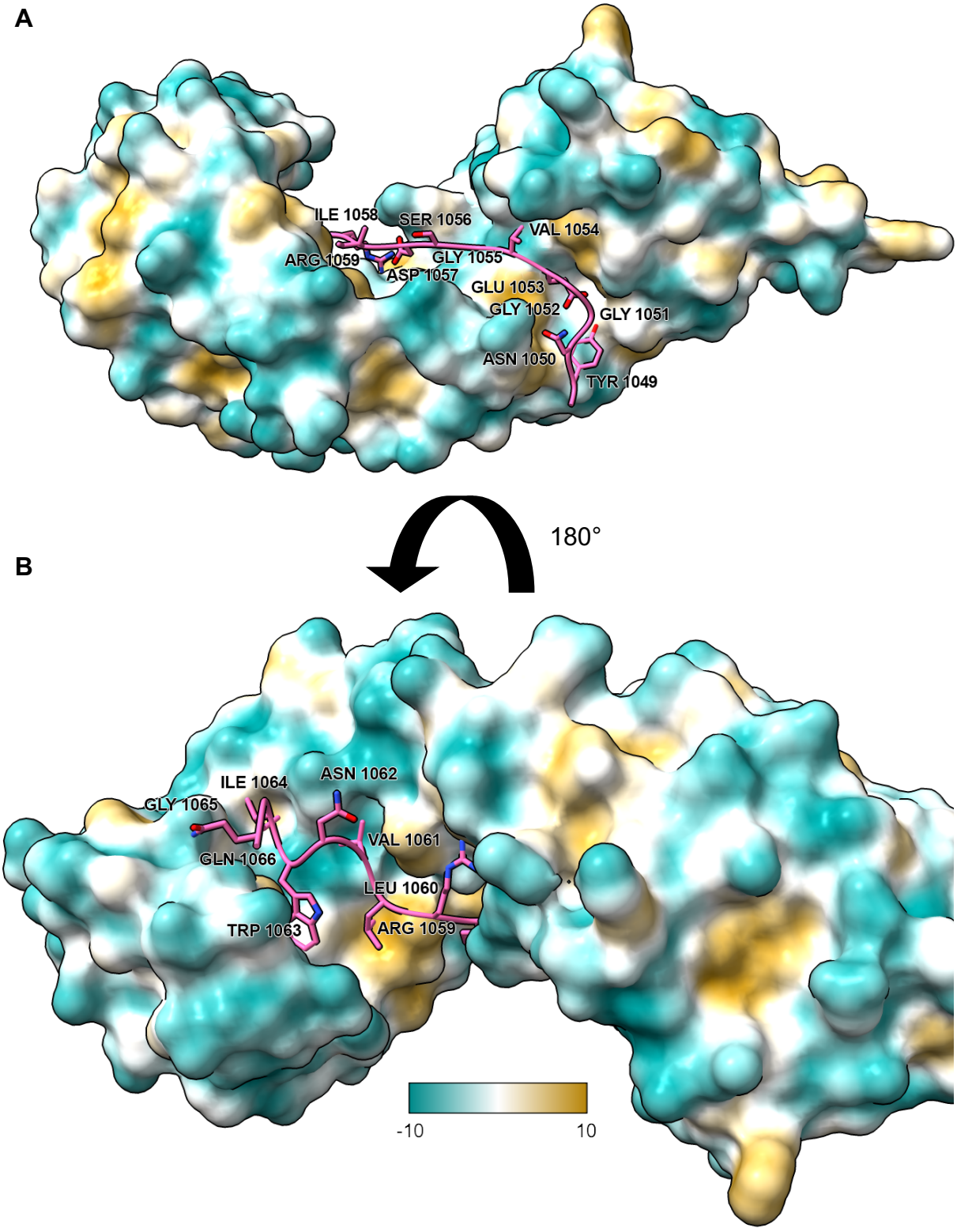
that any additional interactions between the two proteins are dependent upon (and secondary to) the tail peptide binding to GltE1.

Table 3.5: Hydrogen bonding across GltD1$_{Cterm}$:GltE1 interface

GltD1	Length (Å)	GltE1
TYR1029 [OH]	2.3	ASP176 [OD2]
VAL1054 [N]	3.3	ASN234 [OD1]
ARG1059 [N]	3.2	ASN166 [OD1]
ALA762 [O]	3.4	TYR79 [OH]
ALA797 [O]	3.2	LEU111 [N]
LEU1024 [O]	2.6	LYS179 [NZ]
GLU1027 [OE2]	2.3	TYR171 [OH]
ASN1050 [OD1]	2.8	ARG175 [NE]
GLY1051 [O]	2.4	ARG268 [NH2]
VAL1054 [O]	3.0	ASN199 [ND2]
VAL1054 [O]	2.8	ASN234 [ND2]
GLN1066 [OE1]	3.1	ARG94 [NH2]

Table 3.6: Salt bridges at the GltD1$_{Cterm}$:GltE1 interface

GltD1	Length (Å)	GltE1
ARG1059 [NH1]	2.3	ASP194 [OD2]
ARG1059 [NH2]	3.3	ASP194 [OD2]
GLU960 [OE1]	3.2	ARG144 [NH2]
GLU960 [OE2]	3.4	ARG144 [NH2]
GLU960 [OE2]	3.2	ARG144 [NE]
GLU1027 [OE1]	2.6	ARG151 [NE]
GLU1027 [OE2]	2.3	ARG151 [NH1]
GLU1027 [OE2]	2.8	LYS179 [NZ]
GLU1027 [OE2]	2.4	ARG151 [NE]
GLU1027 [OE2]	3.0	ARG151 [NH2]
ASP1057 [OD1]	2.8	ARG196 [NH2]
ASP1057 [OD2]	3.8	ARG196 [NE]


A
ILE 1058
SER 1056
VAL 1054
ARG 1059
GLY 1055
ASP 1057
GLU 1053
GLY 1052
GLY 1051
ASN 1050
TYR 1049
180°
B
ILE 1064
ASN 1062
GLY 1065
VAL 1061
GLN 1066
LEU 1060
TRP 1063
ARG 1059
-10
10

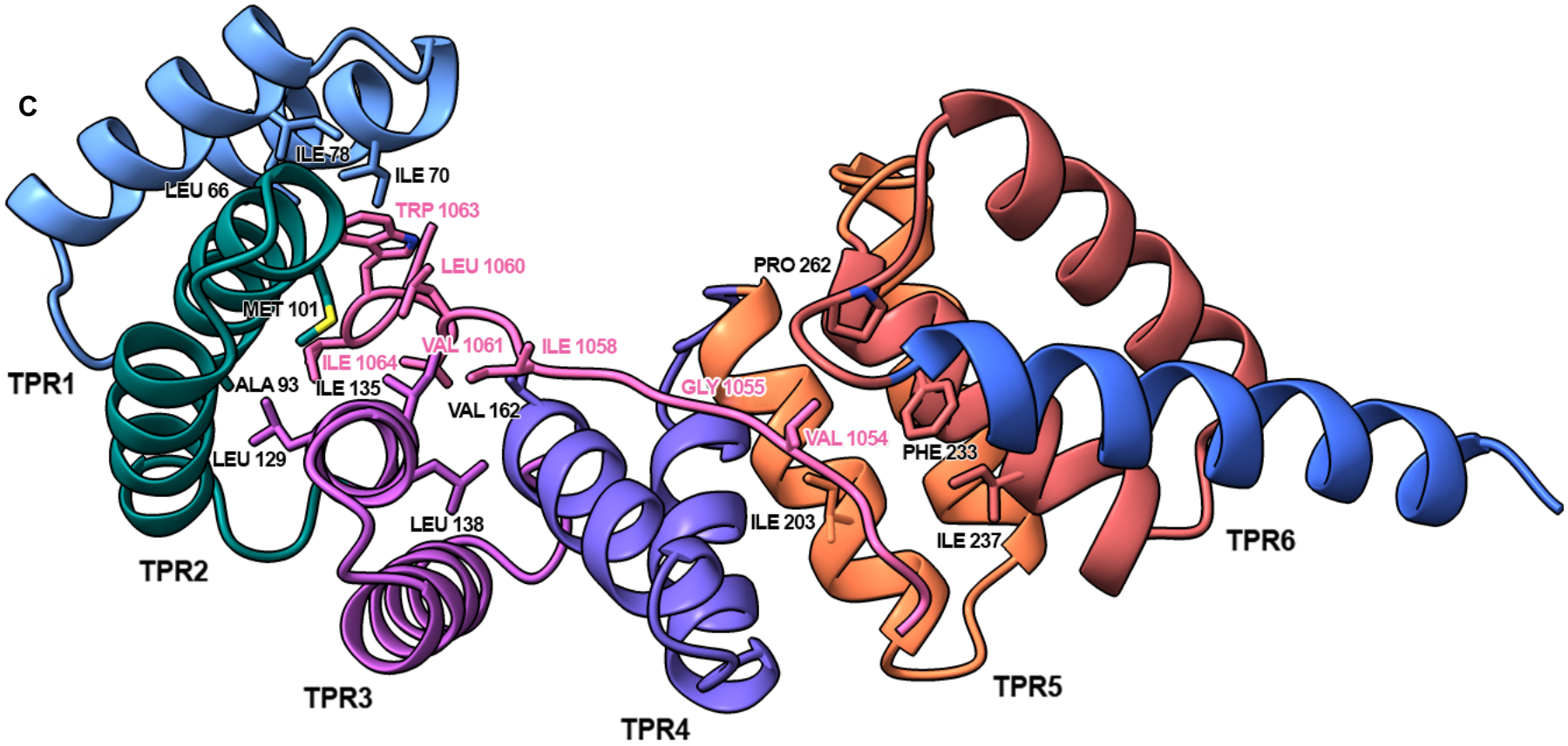
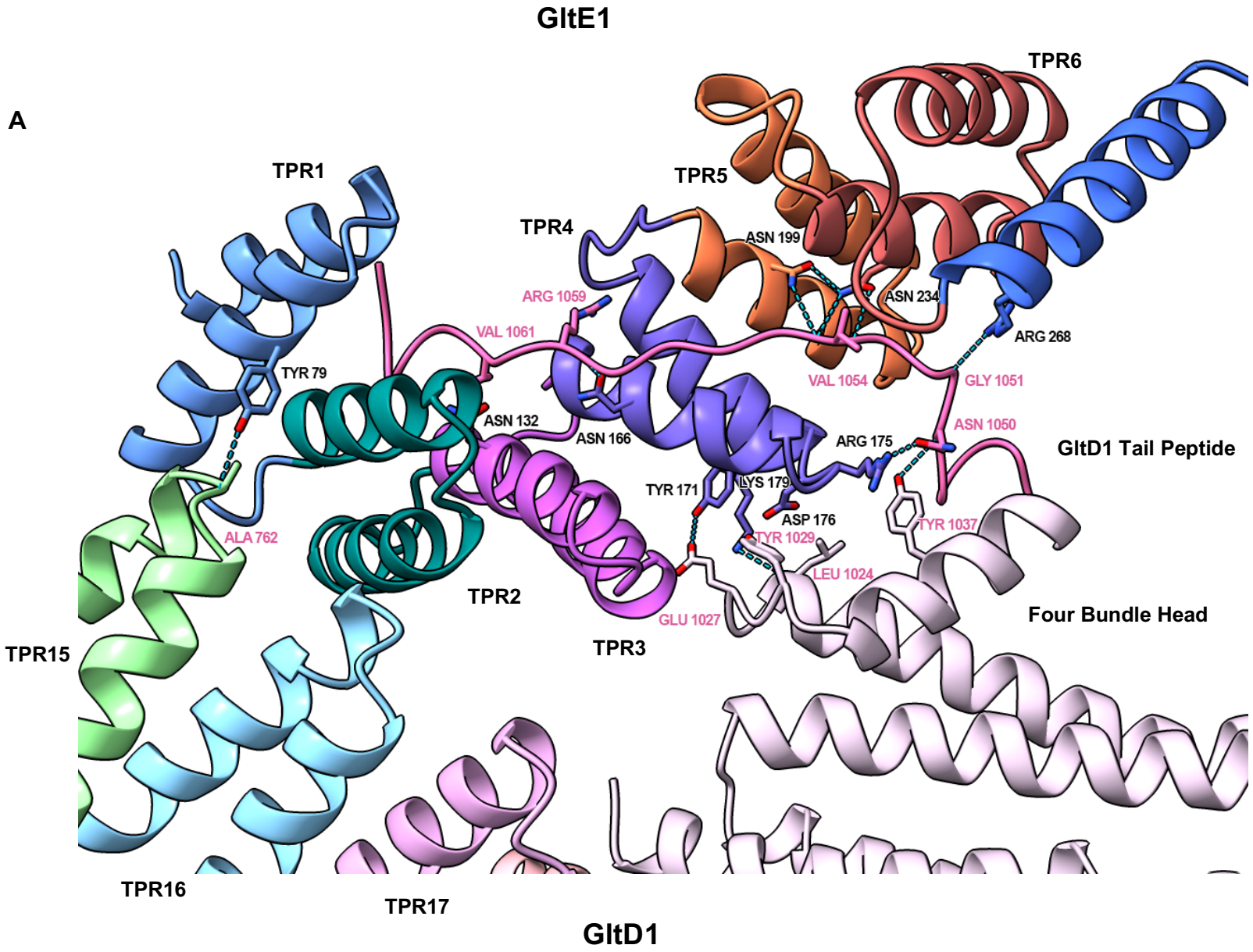


Figure 3.31: Hydrophobic interactions at the GltD1$_{Cterm}$-GltE1 interface
A & B – Cartoon/surface representation of the GltD1$_{Cterm}$-GltE1 interface. The tail peptide of GltD1 (coloured pink) packs against a continuous hydrophobic groove that forms along the concave face of GltE1. Hydrophobic residues from all 6 TPR motifs of GltE1 contribute to the interaction creating an interface with a buried surface area of 2,163 Å^2. GltE1 surface coloured according to hydrophobicity with hydrophobic and hydrophilic residues coloured gold and cyan respectively. Surface representation created using ChimeraX.
C – Cartoon representation of hydrophobic packing at the GltD1$_{Cterm}$-GltE1 interface. The tail peptide of GltD1 is coloured pink and GltD1 residue labels are also coloured pink for clarity. GltE1 structure is coloured by TPR (labelled TPR 1–6) for clarity. Residues participating in hydrophobic interactions are shown as sticks.


GltE1
A
TPR6
TPR1
TPR5
TPR4
ASN 199
ARG 1059
ASN 234
VAL 1061
ARG 268
TYR 79
VAL 1054
GLY 1051
ASN 132
ASN 166
ASN 1050
ARG 175
GltD1 Tail Peptide
TYR 171
LYS 179
ASP 176
ALA 762
TYR 1037
TYR 1029
LEU 1024
TPR2
GLU 1027
Four Bundle Head
TPR15
TPR3
TPR16
TPR17
GltD1

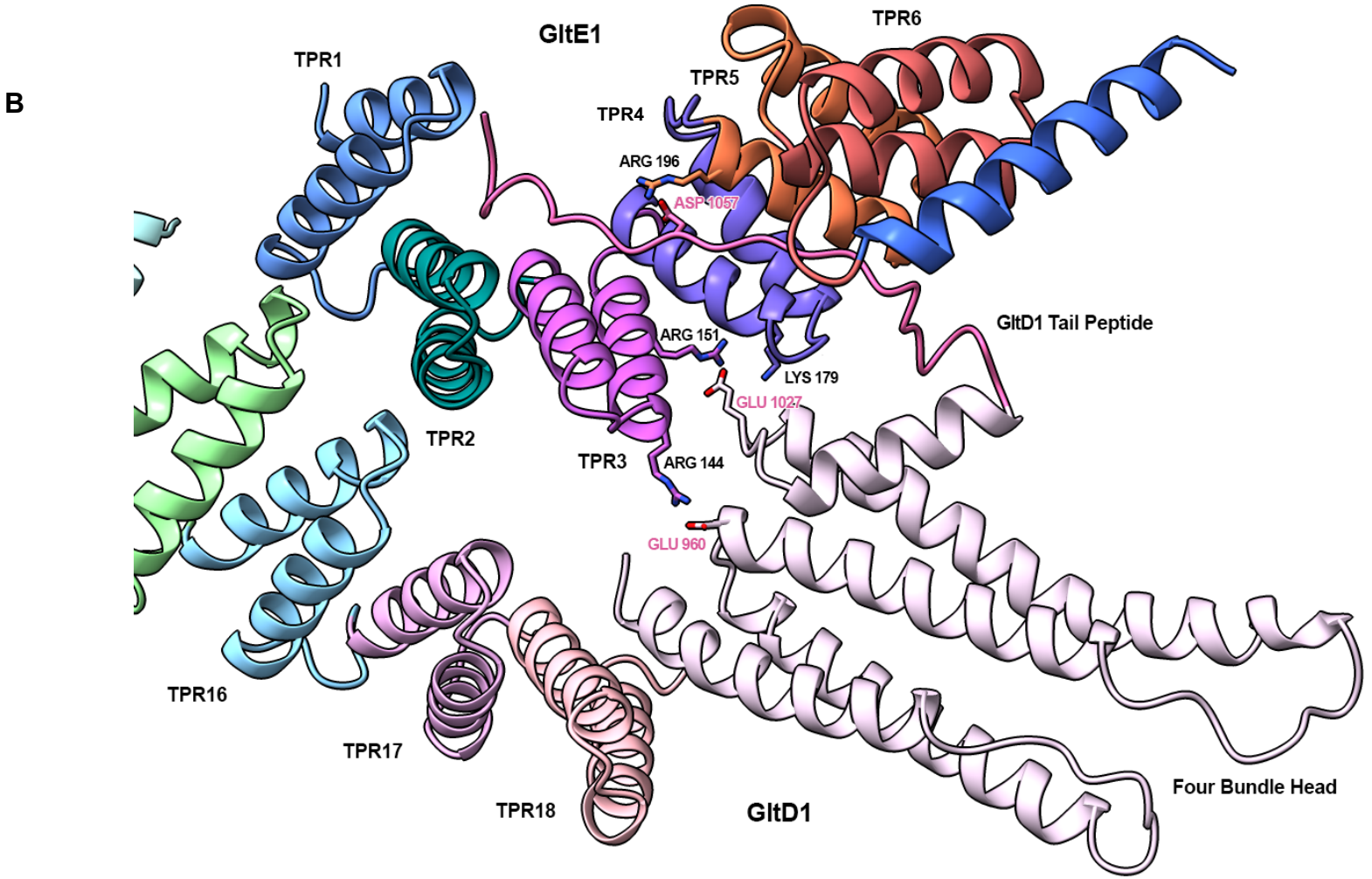


Figure 3.32: Further interactions at the GltD1$_{Cterm}$-GltE1 interface
A – Hydrogen bonding further stabilises the GltD1$_{Cterm}$-GltE1 interaction, with residues on TPR15, the four bundle head and the tail peptide of GltD1, forming hydrogen bonds with residues on the concave face of GltE1. Residues forming hydrogen bonds are shown as sticks and hydrogen bonds are shown as blue dashed lines. GltD1 residue labels are coloured pink for clarity.
B – Salt bridges at the GltD1$_{Cterm}$-GltE1 interface provide additional stability to the GltD1$_{Cterm}$-GltE1 interaction. Residues on the four bundle head and tail peptide of GltD1 form salt bridges with residues on the concave face of GltE1. Residues forming salt bridges are shown as sticks. GltD1 residue labels are coloured pink for clarity.

3.2.6 Further insights from AF modelling

Following the experimental validation of the GltD1:GltE1 interaction, AF was used to generate a model of the entire OM and periplasmic components of the Glt1 operon. An initial model was generated by simultaneously submitting the sequences of GltA1-G1 to AF2 (Figure 3.33). In the produced model, GltD1 folds back on itself, sandwiching GltE1 between the N and C terminal domains of GltD1. As previously shown, the deletion of C-terminal tail peptide of GltD1 completely abolished the interaction between GltD1 and GltE1 during ITC experiments. Therefore, the spatial arrangement of GltD1 and GltE1 in this co-folded model raised concerns about the accuracy of the model.

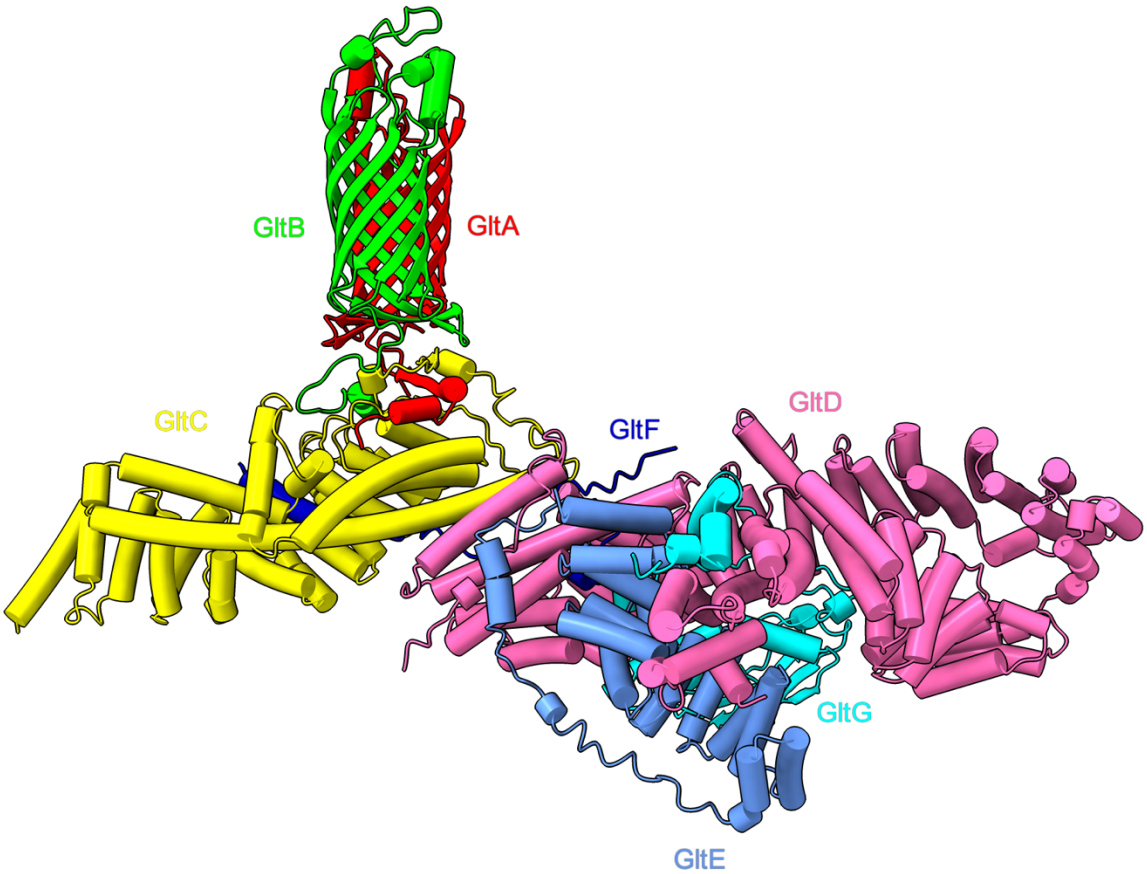


Figure 3.33: AF model of the OM and periplasmic components of the Glt1 operon
AF model generated by simultaneously submitting the sequences of GltA-G from the Glt1 operon to a local installation of AF2. In the model, GltD1 can be seen to occupy a conformation which sandwiches GltE between both termini of GltD1. This conformation contradicts experimental binding data, questioning the accuracy of the model. Average pLDDT (confidence) scores for AF models are as follows: GltA1 = 86.6; GltB1 = 85.1; GltC1 = 85.0; GltD1 = 81.2; GltE1 = 75.2; GltF1 = 60.0; and GltG1 = 65.2.

In an attempt to improve the accuracy of the model, an alternative approach was adopted where gliding proteins were iteratively co-folded (usually pairwise) and then aligned based on common overlaps to generate a composite model.

3.2.6.1 The OM Glt complex

The OM Glt components comprise GltA and GltB, which are predicted to interact to form a hub that houses further interactions with GltC and the MIDAS adhesin[99,135]. Whilst the Glt1 operon contains no gene encoding an adhesin, a genome search identifies Bd0875 as a MIDAS containing protein with sequence similarity (e.g. 24% to 31%) to gliding MIDAS adhesins. Therefore, Bd0875 was trialled as a candidate adhesin for the Glt1 operon. To model the OM interactions, models were generated for two subcomplexes (GltA1:GltB1:Bd0875 and GltA1:GltB1:GltC1) (Figures 3.34 and 3.35). These subcomplexes were then superimposed based on the positions of GltA1 and GltB1. GltA1 and GltB1 are modelled as homologous eight-stranded β-barrels that interact to form a heterodimer, with the main interaction predicted to occur at β-strands at the base of each barrel. Bd0875 is predicted to interact with GltB1 and thus appears to be a compelling candidate for the adhesin of the Glt1 operon. The interaction between Bd0875 and GltB1 appears to be mediated through β-strand complementation with a β-strand at the N-terminus of Bd0875 modelled to interact with a β-strand on GltB1 that projects out of the OM to be exposed at the outer leaflet. Notably, the validity of this interaction is strengthened when considering that Bd0875 is expressed as a lipoprotein and the interaction as modelled would place the acylated cysteine residue directly at the OM leaflet.

GltC1 is modelled as a TPR superhelix comprising 8 TPR motifs, which leads into a single large α-helix which is capped at the C-terminus by a flexible 'lasso' (residues 480–529). GltC1 is positioned at the base of the GltA1/B1 dimer and is predicted to interact directly with GltA1. A flexible N-terminal peptide from GltA1 passes through the periplasm and threads through a 'lasso'-like loop formed by the C-terminus of GltC1. A single β-strand on the N-terminal GltA1 peptide is sandwiched between two β-strands on the 'lasso' of GltC1, forming a three-stranded β-sheet. Aside from the interaction between the β-strands, hydrophobic residues on the flexible peptide of GltA1 pack into a hydrophobic groove on GltC1. Moreover, the predicted interface places oppositely charged residues on GltA1 and GltC1 within proximity of each other, suggesting salt bridging stabilises the interaction.

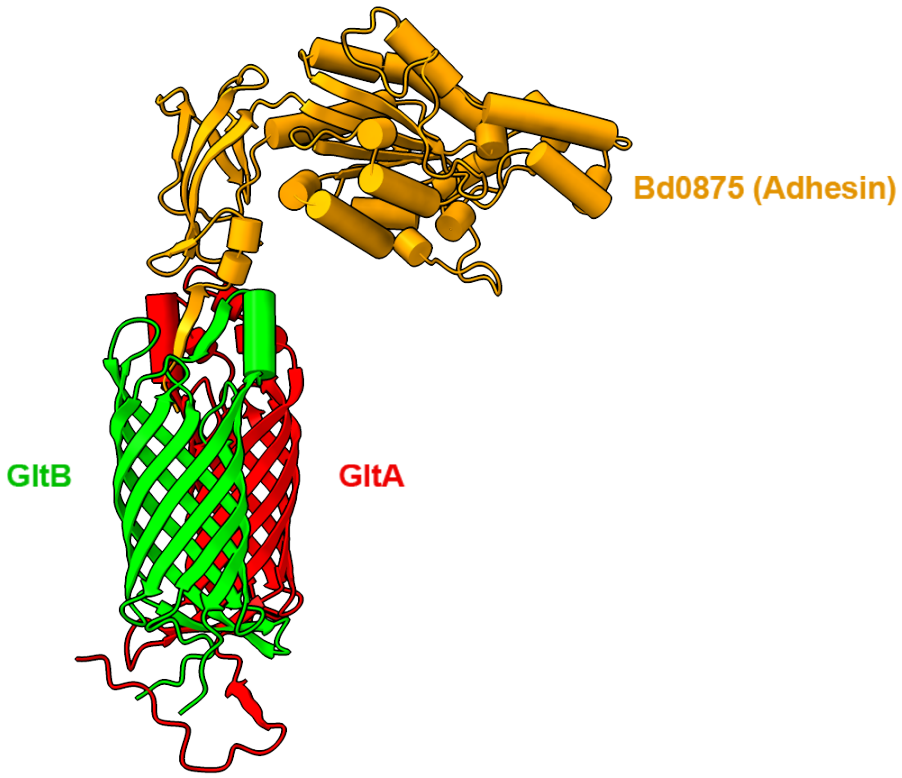


Figure 3.34: AF model of GltA1-GltB1-Bd0875 interaction
The co-folded model shows an interaction between GltA1 (coloured red) and GltB1 (coloured green) which form a heterodimer in the OM. The model also predicts an interaction between GltB1 and Bd0875 (coloured orange) which interact through β-stand complementation at the outer leaflet of the OM. Average pLDDT (confidence) scores for AF models are as follows: GltA1 = 89.8; GltB1 = 82.6; and Bd0875 = 86.5.

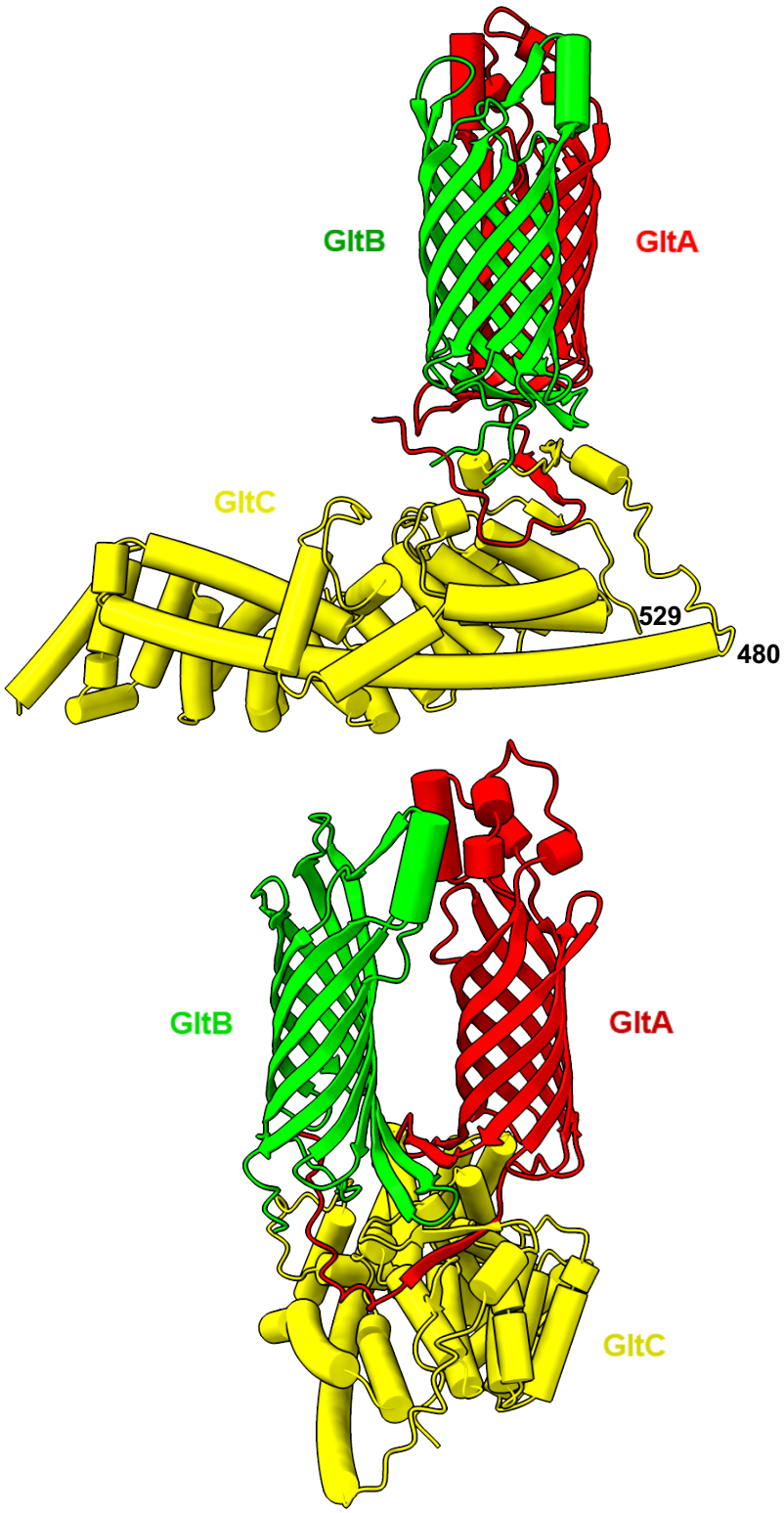


Figure 3.35: AF model of the GltA1-GltB1-GltC1 interaction
A flexible N-terminal peptide of GltA (coloured red) interacts with the C-terminal lasso (residues 480–529) of GltC1 (coloured yellow) to form a three-stranded β-sheet that creates a direct link between the OM and periplasmic Glt components. Average pLDDT (confidence) scores for AF models are as follows: GltA1 = 78.9; GltB1 = 78.9; and GltC1 = 84.4.

3.2.6.2 Periplasmic Glt complex

The periplasmic Glt components were initially co-folded in pairs to identify potential interaction sites. From these, pairwise co-folding interactions were identified between GltD1:GltC1, GltD1:GltF1, GltD1:GltG1, GltC1:GltF1 and GltE1:GltG1. Guided by these interactions, co-folded models of GltC1:GltD1$_{25–469}$:GltF1:GltG1$_{491–707}$ and GltD1:GltE1:GltG1$_{491–707}$ were generated to provide insights into these interactions in a wider biological context (unless otherwise stated, full-length sequences were utilised).

In these models, GltC1 and GltD1 pack against one another, with the interaction occurring between the large α-helix of GltC1 and the N terminal hook of GltD1 (Figure 3.36). GltF1 threads along the concave face of the GltC1 superhelix and forms a short α-helix that predominately occupies a binding pocket formed between TPR5 and TPR6 of GltC1 (Figure 3.36). GltF1 emanates from GltC1 and continues to thread along the concave face of the N-terminal hook domain of GltD1, potentially acting as a tether to link GltC1 and GltD1 together. The N-terminus of GltF1, which contains the (predicted) ton-box motif[97], protrudes out of GltD1 and binds to the TonB domain of GltG1, which packs against the concave face of GltD1 (Figure 3.36). Notably, the presence of GltF1 appears to cause a conformational change in the TonB domain. When folded alone, the TonB domain of GltG1 contains a four stranded (β-strand 1-4) β-sheet (Figure 3.37). In the presence of GltF1, β-strand 1 of the GltG1 TonB domain reorientates, allowing the ton-box motif of GltF1 to complete the fold (Figure 3.37). The GltG1 TonB domain connects to an IM α-helix via a large, disordered region (~200 residues). Surprisingly, approximately 100 residues of this disordered region can be modelled

with confidence, traversing along the concave face of GltD1, guided along the superhelical architecture (Figure 3.38). This disordered region packs against the four-helix bundle of GltD1 and occupies the cavity formed between GltD1 and GltE1 (Figure 3.38). Together, the generated models could be aligned based on common overlaps to generate a composite model of the OM-periplasmic Glt complex (Figure 3.39).

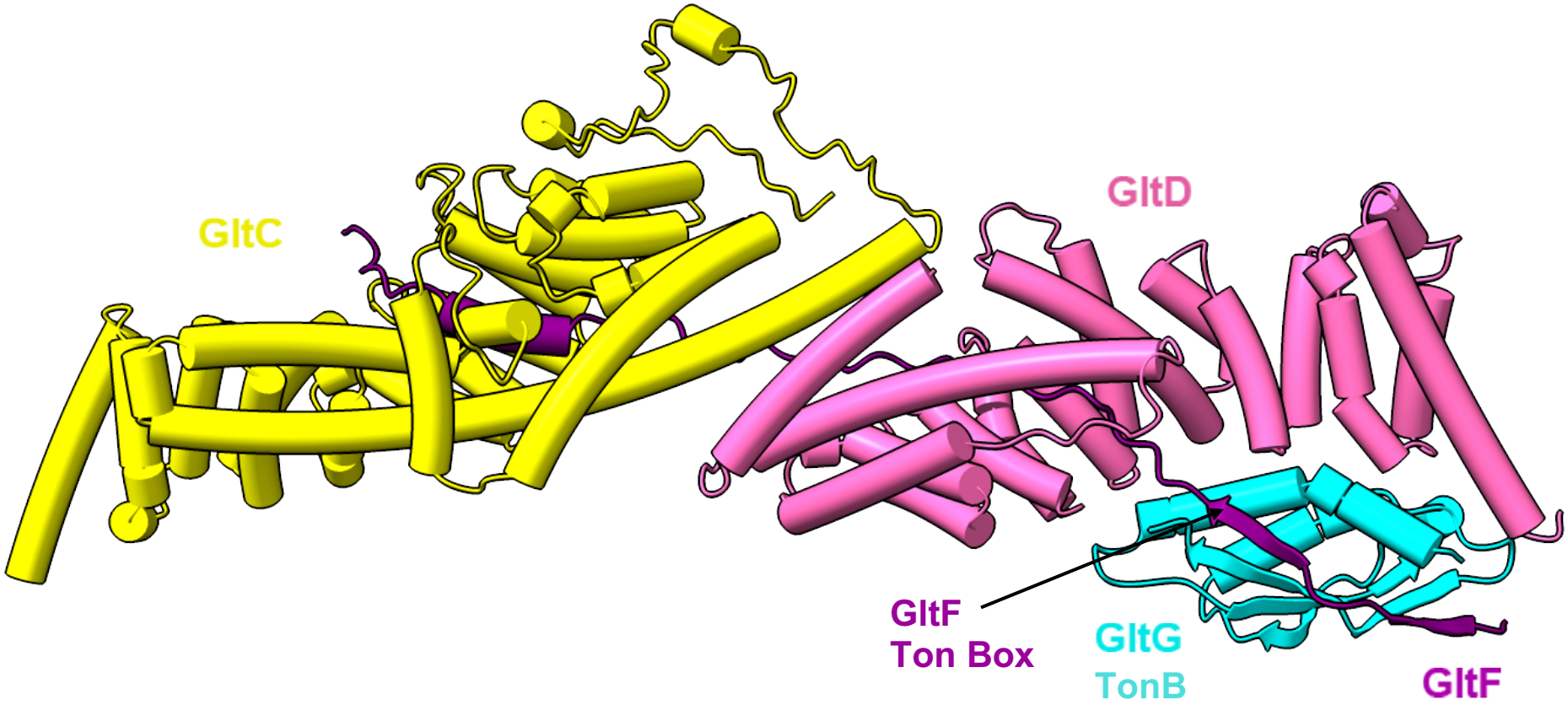


Figure 3.36: AF model of the GltC1-GltD1$_{25–469}$-GltF1-GltG1$_{491–707}$ interaction
GltF1 (coloured purple) is predicted to pack against the concave face of both GltC1 (coloured yellow) and GltD1 (coloured pink), providing a link between both proteins. The ton-box motif of GltF1 protrudes from GltD1 and binds to the TonB domain of GltG (coloured cyan). The TonB domain of GltG packs against the concave face of GltD1. Average pLDDT (confidence) scores for AF models are as follows: GltC1 = 82.6; GltD1 = 79.3; GltF1 = 67.5; and GltG1 = 77.0.

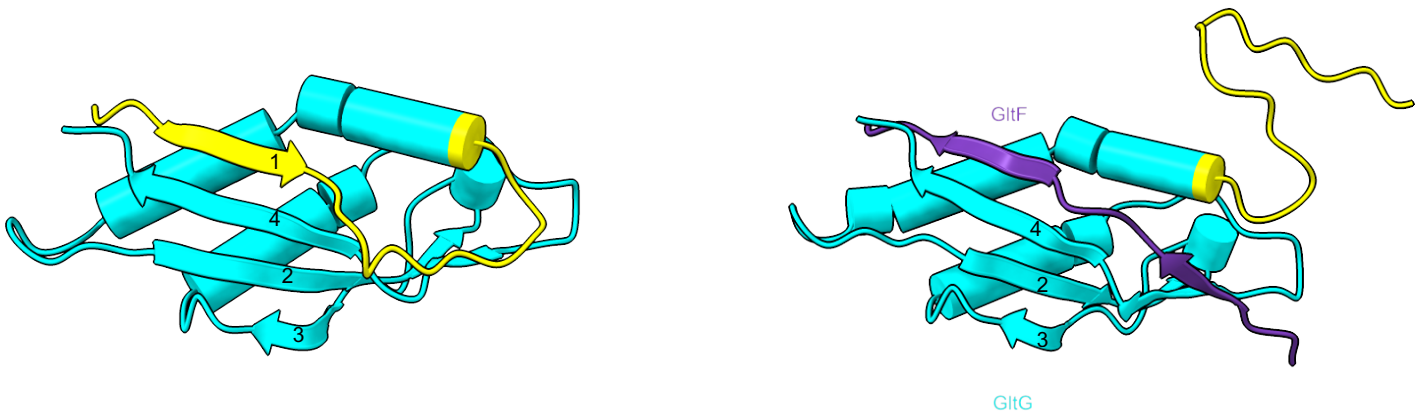


Figure 3.37: Predicted conformational change of the TonB domain of GltG1 upon GltF1 binding
Left: AF model of the TonB domain of GltG1 (coloured cyan). Average pLDDT (confidence) scores for GltG1: 76.01 **Right:** AF model of the TonB domain of GltG1 (coloured cyan) co-folded with GltF1 (coloured purple). The presence of GltF1 is predicted to cause a conformational change of the TonB domain of GltG1, causing β-strand 1 (coloured yellow) to reorientate to allow GltF1 to complete the fold. β-strands are labelled 1–4. Average pLDDT (confidence) scores for AF models are as follows: GltF1: 67.52; and GltG1: 76.99.

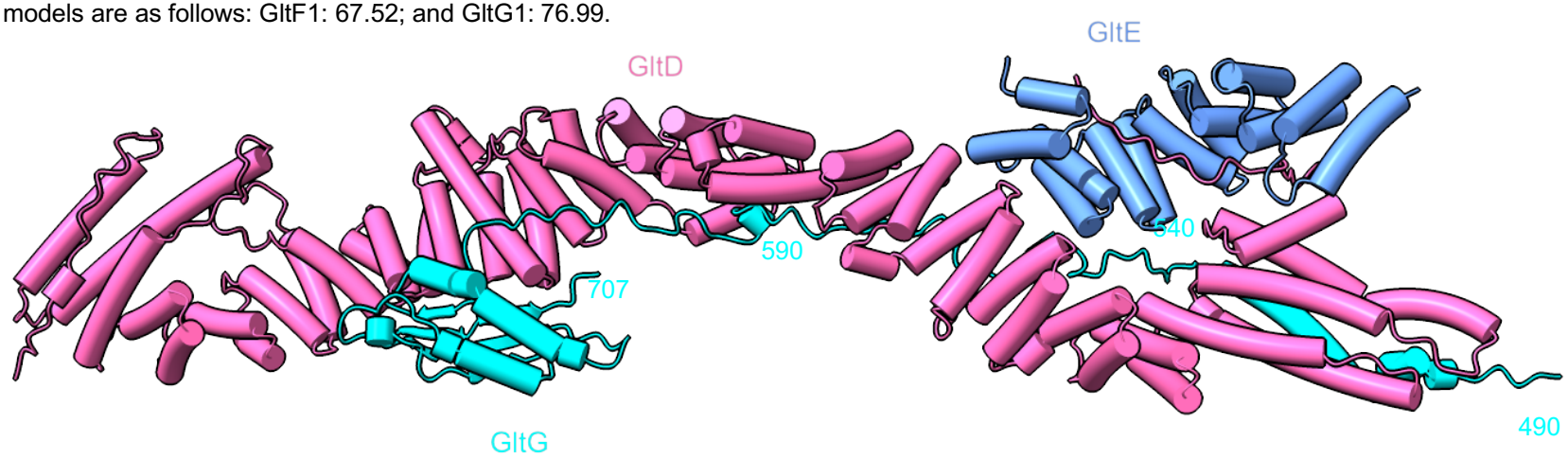


Figure 3.38: AF model of the GltD1-GltE1-GltG1$_{491–707}$ interaction
GltD1 (coloured pink) and GltE1 (coloured blue) interact as previously described. The disordered region of GltG1 (coloured cyan) traces along the concave face of GltD1 and packs into the cleft between GltD1 and GltE1. Average pLDDT (confidence) scores for AF models are as follows: GltD1 = 84.0; GltE1 = 82.0; and GltG1 = 75.4.

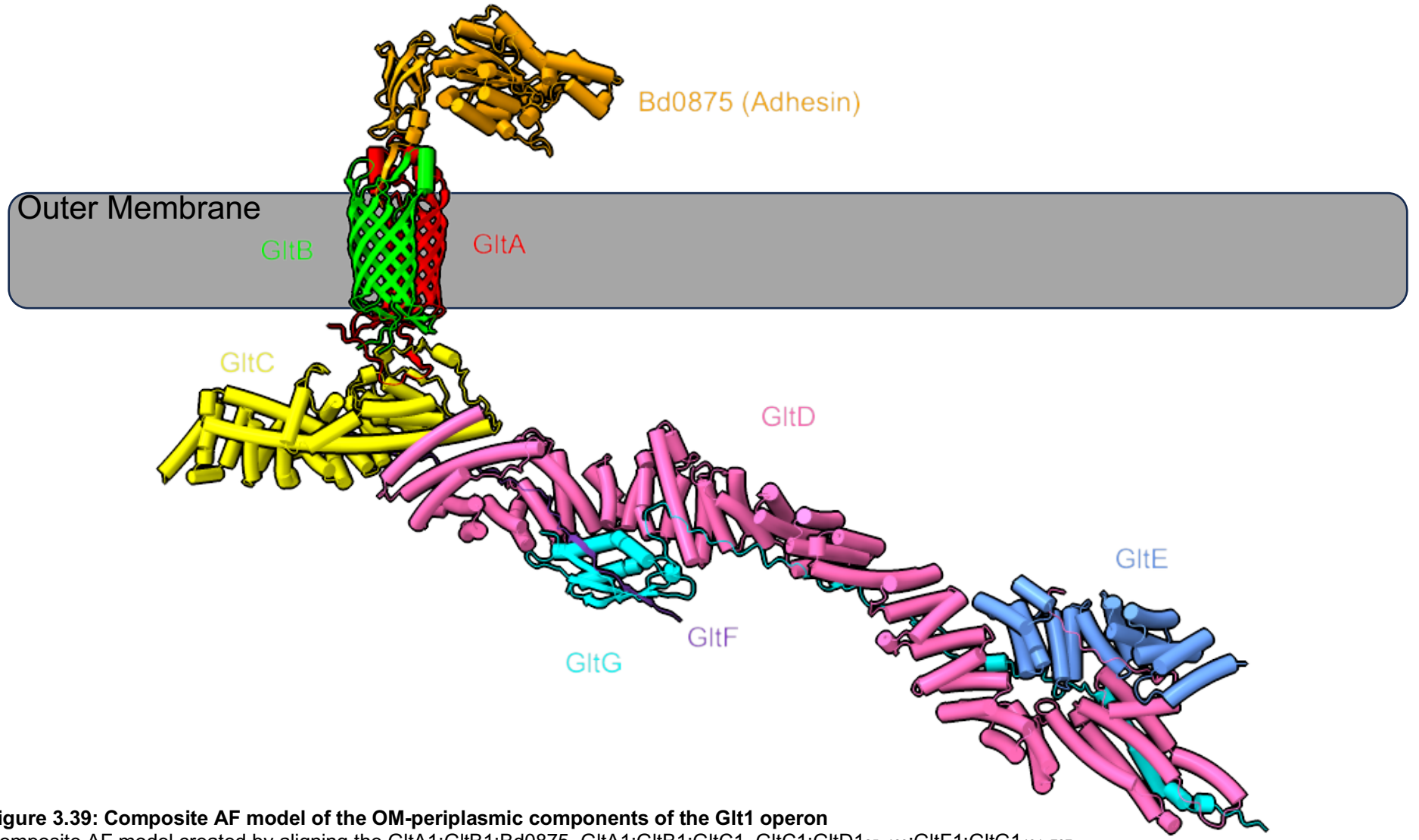


Figure 3.39: Composite AF model of the OM-periplasmic components of the Glt1 operon
Composite AF model created by aligning the GltA1:GltB1:Bd0875, GltA1:GltB1:GltC1, GltC1:GltD1$_{25–469}$:GltF1:GltG1$_{491–707}$ and GltD1:GltE1:GltG1$_{491–707}$ models based on common overlaps. The model provides an insight into how the periplasmic and OM components of the Agl-Glt complex could interact.

3.2.7 Experimental validation of GltD1:GltE1:GltG1 interaction

The extensive AF modelling provides compelling insights into the connectivity within the Glt complex, although these interactions require experimental validation. The GltD1:GltE1:GltG1 interaction was targeted for validation, as this would complement the already obtained structural data and provide a complete characterisation of this region of the Glt complex (Figure 3.40). To investigate this interaction, the region of GltG1 that was modelled to interact with GltD1 and GltE1 was cloned as an MBP fusion (MBP-GltG1$_{491–556}$).

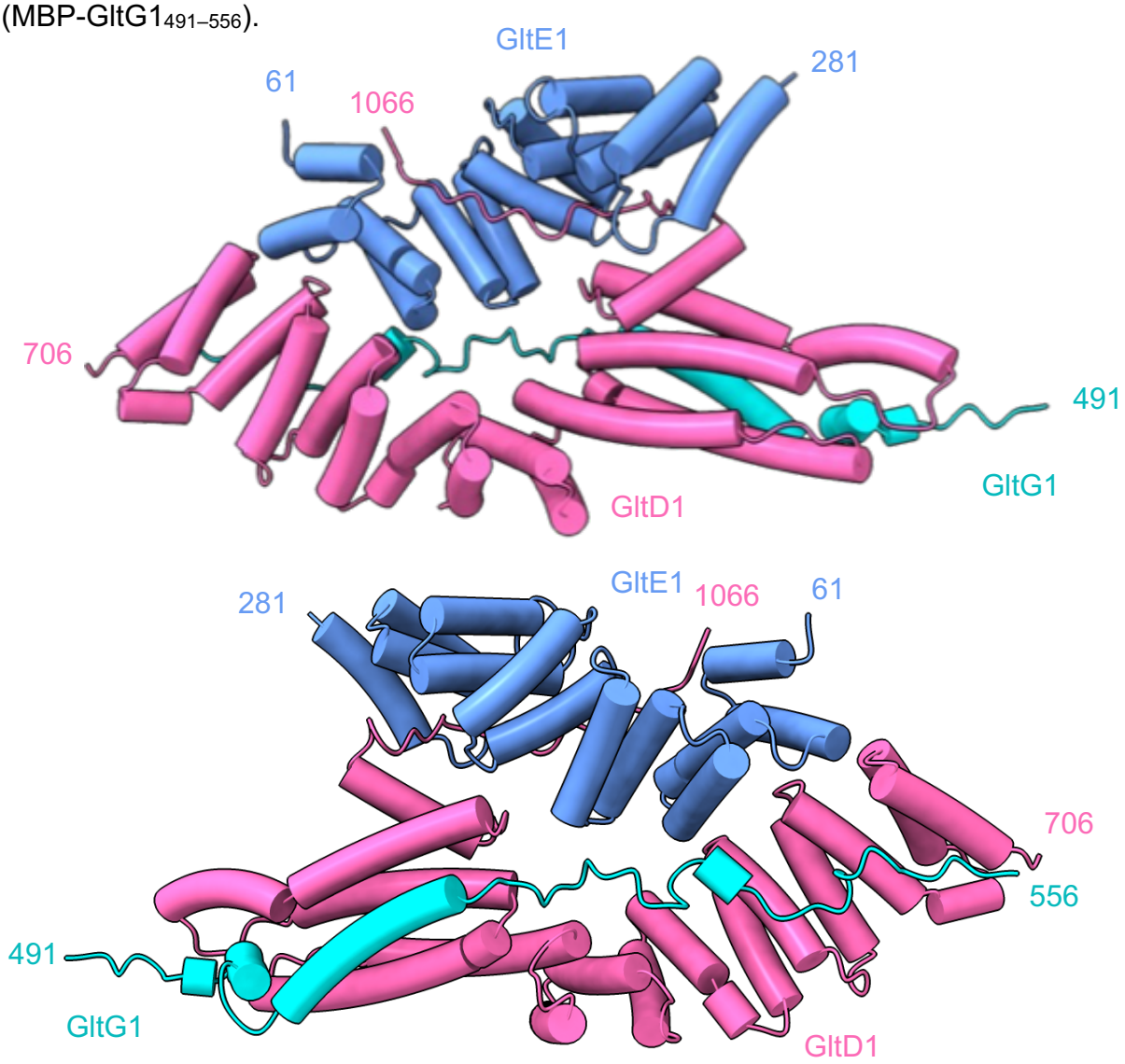
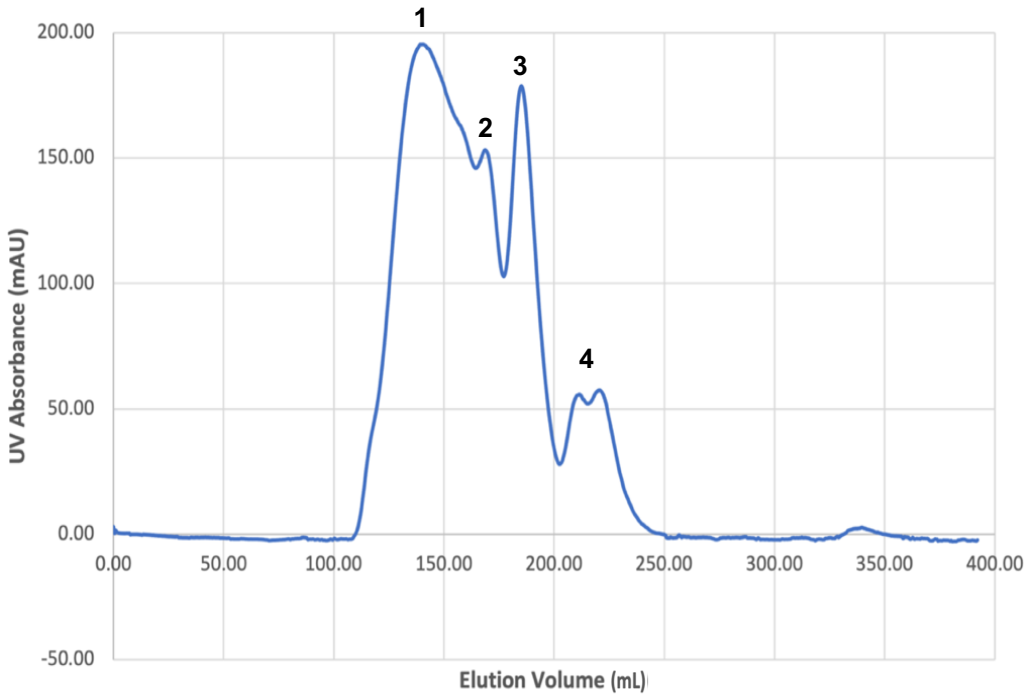


Figure 3.40: AF model of the region of the GltD1:GltE1:GltG1 complex targeted for experimental validation.
Average pLDDT (confidence) for GltD1 = 84.0, Average pLDDT (confidence) for GltE1 = 82.0, Average pLDDT (confidence) for GltG1 = 75.4.

MBP-GltG1$_{491–556}$ was initially purified using an MBPtrap. MBP-GltG1$_{491–556}$ containing fractions were pooled and further purified by SEC. MBP-GltG1$_{491–556}$ was eluted from a HiLoad 26/600 Superdex 200 prep grade column in multiple peaks (Figure 3.41). The peak at ~186 mL (peak 3) corresponds approximately to the molecular weight of monomeric MBP-GltG1$_{491–556}$ (~48 KDa) and SDS-PAGE analysis confirmed that these peak fractions contained MBP-GltG1$_{491–556}$ (Figure 3.41). The gels show clear signs of degradation, suggesting that GltG1 was being cleaved from MBP in a sub-population of the sample.

A



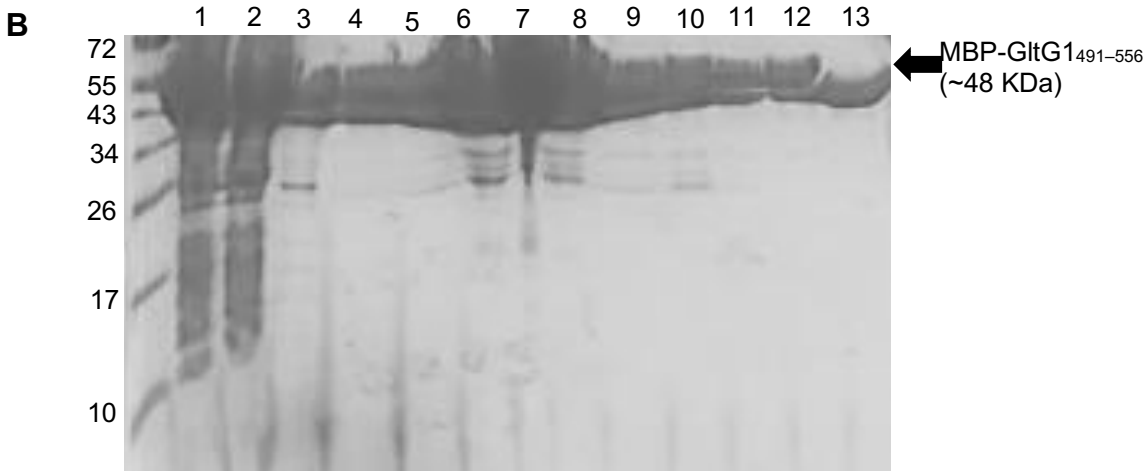


Figure 3.41: Purification of MBP-GltG1$_{491–556}$
A – Chromatogram for MBP-GltG1$_{491–556}$ SEC purification. MBP-GltG1$_{491–556}$ was eluted from a HiLoad 26/600 Superdex 200 prep grade column in four peaks (labelled 1–4). The elution volume of peak 3 (~186 mL) is consistent with the size of monomeric MBP-GltG1$_{491–556}$ (~48 KDa).
B – SDS-PAGE analysis of MBP-GltG1$_{491–556}$ purification.
Lanes: 1 = lysate, 2 = flowthrough, 3-9 = MBPtrap elution fractions, 10 = SEC peak 1, 11 = SEC peak 2, 12 = SEC peak 3, 13 = SEC Peak 4. Black arrow indicates the position of MBP-GltG1$_{491–556.}$
MBP-GltG1$_{491–556}$ can be seen as a large band between the 43 and 55 kDa markers. The gel reveals signs of degradation, suggesting that GltG1 is being cleaved from MBP in a sub-population of the sample.

ITC was used to measure the interaction between both GltD1:MBP-GltG1$_{491–556}$ and GltD1:GltE1:MBP-GltG1$_{491–556.}$ In both scenarios, no interaction was observed between GltG1$_{491–556}$ and GltD1/GltE1 (Figure 3.42). When GltD1 is modelled alone by AF, a flexible N-terminal peptide (GltD1 residues 1–100) packs against the concave face of the TPR superhelix, occupying the same position that GltG1 was modelled to occupy. To rule out any possible inhibitory effects of this peptide, the ITC experiments were repeated using GltD1$_{Cterm}$ which lacks this possible inhibitory peptide. In these experiments, there was still no observable interaction between GltG1$_{491–556}$ and GltD1:GltE1.

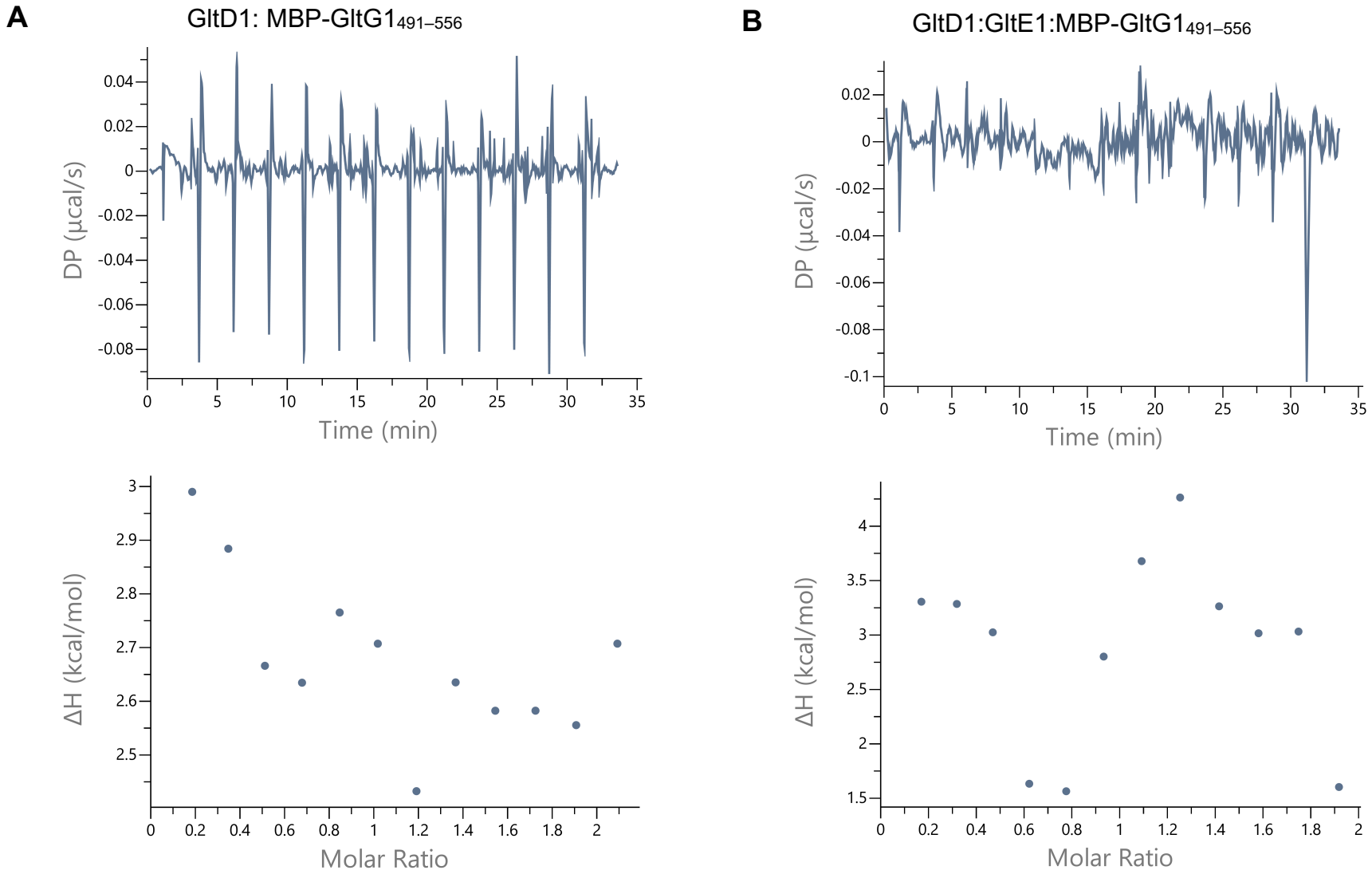


Figure 3.42: ITC interactions for MBP-GltG1$_{491–556}$
A – Binding thermogram for GltD1: MBP-GltG1$_{491–556}$ interaction. No binding was observed for the two proteins.
B – Binding thermogram for GltD1:GltE1: MBP-GltG1$_{491–556}$ interaction. No binding was observed for the two proteins.

3.2.8 Expressing GltC1

In order to probe the GltC1:GltD1 interaction, attempts were made to express and purify GltC1. GltC1 was initially cloned as a full-length construct (~60 KDa) with a PelB signal peptide for secretion to the periplasm. Expression trials revealed poorly soluble expression of full-length GltC1 in all types of media trialled (Figure 3.43). As GltC1 is a peripheral membrane protein, a buffer screen was performed to determine whether it was possible to increase the yield of soluble protein by preventing GltC1 from interacting with any cell membranes (Figure 3.44). A small amount of soluble GltC1 appeared to be present when the cells were lysed into a buffer containing 50 mM Hepes, 250 mM NaCl, 20 mM sodium cholate and 20% glycerol (Figure 3.44). An initial IMAC purification performed in the optimised buffer revealed that GltC1 was extremely sensitive and prone to degradation, with the small quantity of full-length GltC1 obtained being observed to precipitate shortly after IMAC (Figure 3.45). Based on AF modelling, the majority of the interaction between GltC1 and the OMP GltA1 occurs through the C-terminal lasso. In an attempt to improve solubility, the lasso was deleted ($GltC1_{\Delta 480-529}$). However, additional expression trials and buffer screening revealed that deletion of the lasso led to no improvement in the solubility of GltC1 (Figure 3.46).

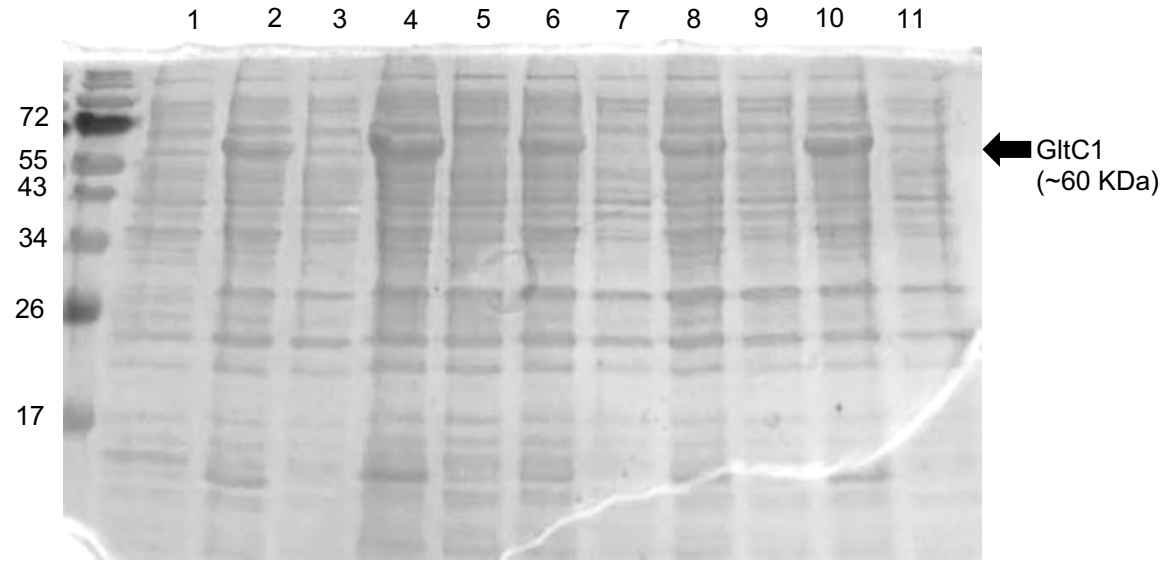


Figure 3.43: GltC1 Expression trial
Expression trial for GltC1. Lanes: 1 = Pre-induction, 2 = LB whole cell, 3 = LB lysate, 4 = TB whole cell, 5 = TB lysate, 6 = AI-A whole cell, 7 = AI-A lysate, 8 = AI-B whole cell, 9 = AI-B lysate, 10 = AI-C whole cell, 11 = AI-C lysate. The black arrow indicates the position and MW of full-length GltC1.

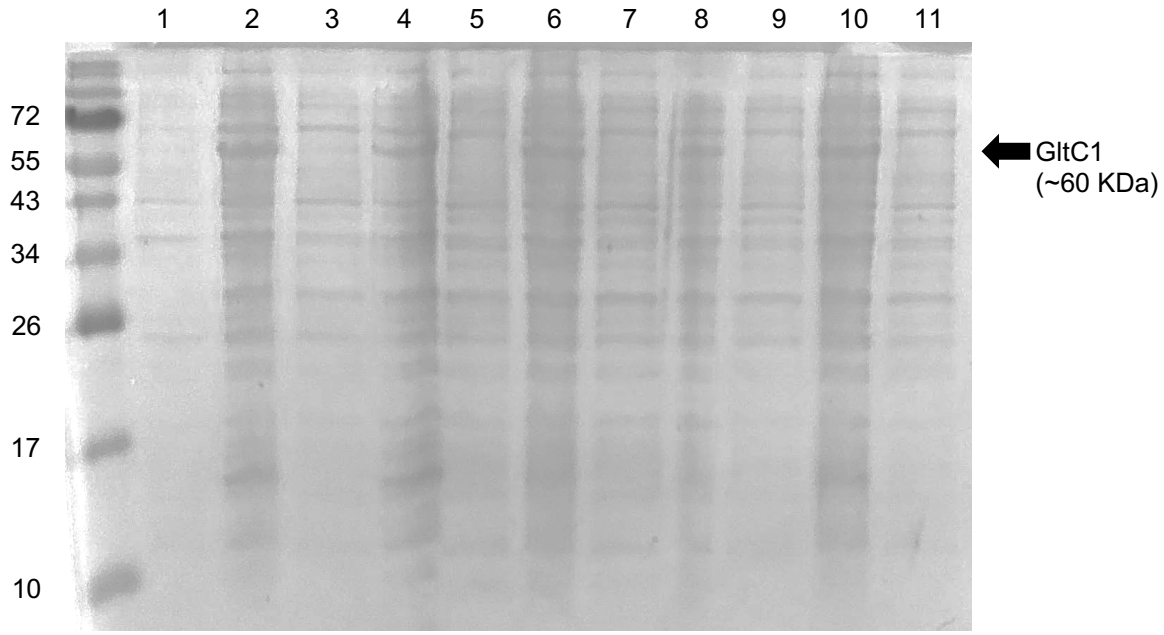


Figure 3.44: GltC1 Buffer screen
Buffer screen for GltC1. All buffers were based on buffer A (50 mM Hepes, 250 mM NaCl, 20 mM imdazole pH 7.5) + an additive to disrupt the interaction between GltC1 and the cell membrane. Lanes: 1 = buffer A lysate, 2 = 20 mM sodium cholate, 20% glycerol whole cell, 3 = 20 mM sodium cholate, 20% glycerol lysate, 4 = 1 M NaCl whole cell, 5 = 1 M NaCl lysate 6 = 2 M NaCl whole cell, 7 = 2M NaCl lysate, 8 = 2% triton x-100 whole cell, 9 = 2% triton x-100 lysate, 10 = 2% DDM whole cell, 11 = 2% DDM lysate. The black arrow indicates the position and MW of full-length GltC1.

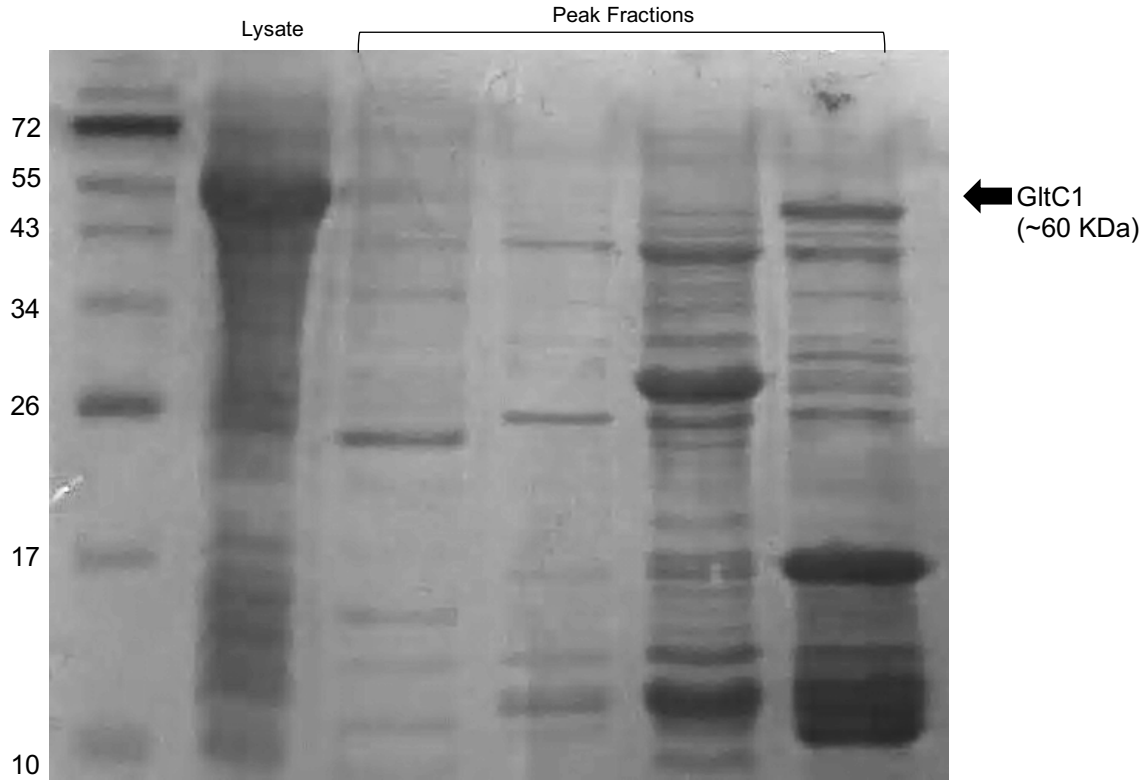


Figure 3.45: IMAC purification of GltC1 lysed in buffer A containing 20 mM sodium cholate and 20% glycerol
The optimised buffer appears to increase the solubility of GltC1. Despite the increased solubility, GltC1 was prone to degradation and the small amount of full-length GltC1 that was obtained from IMAC precipitated shortly after elution. The black arrow indicates the position and MW of full-length GltC1.

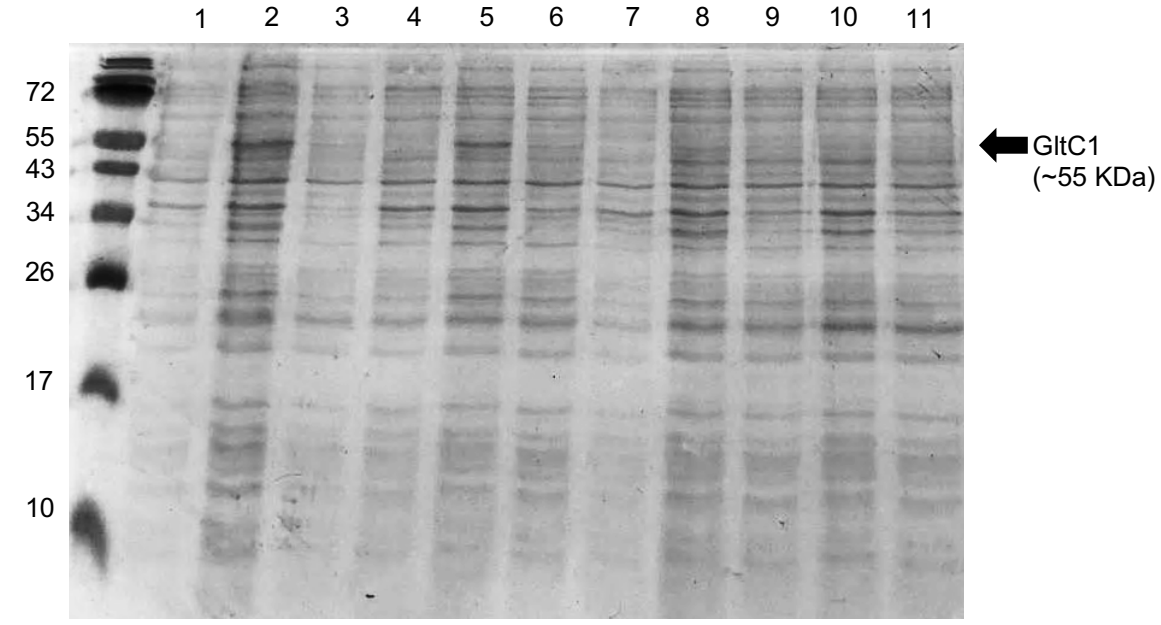


Figure 3.46: GltC1$_{\Delta 480\text{-}529}$ expression trial
Expression trial for GltC1. Lanes: 1 = Pre-induction, 2 = LB whole cell, 3 = LB lysate, 4 = TB whole cell, 5 = TB lysate, 6 = AI-A whole cell, 7 = AI-A lysate, 8 = AI-B whole cell, 9 = AI-B lysate, 10 = AI-C whole cell, 11 = AI-C lysate. The black arrow indicates the position and MW of full-length GltC1$_{\Delta 480\text{-}529}$.

As the attempts to improve the solubility of GltC1 were largely unsuccessful, the focus was switched towards refolding GltC1. Refolding trials were performed for both GltC1 alone and in the presence of GltD1. Purified inclusion pellets were solubilised in 8 M urea and GltC1 was refolded by shock dilution into 50 mM Hepes, 250 mM NaCl followed by overnight cycling at 4°C. SDS-PAGE analysis of refolded samples revealed that GltC1 only appears to refold in the presence of GltD1, suggesting that GltD1 acts as a scaffold/chaperone to aid refolding (Figure 3.47). Time restraints prevented further optimisation of GltC1 refolding, although the performed refolding trials provide a good foundation from which optimisation can be performed to facilitate downstream structural and interaction studies.

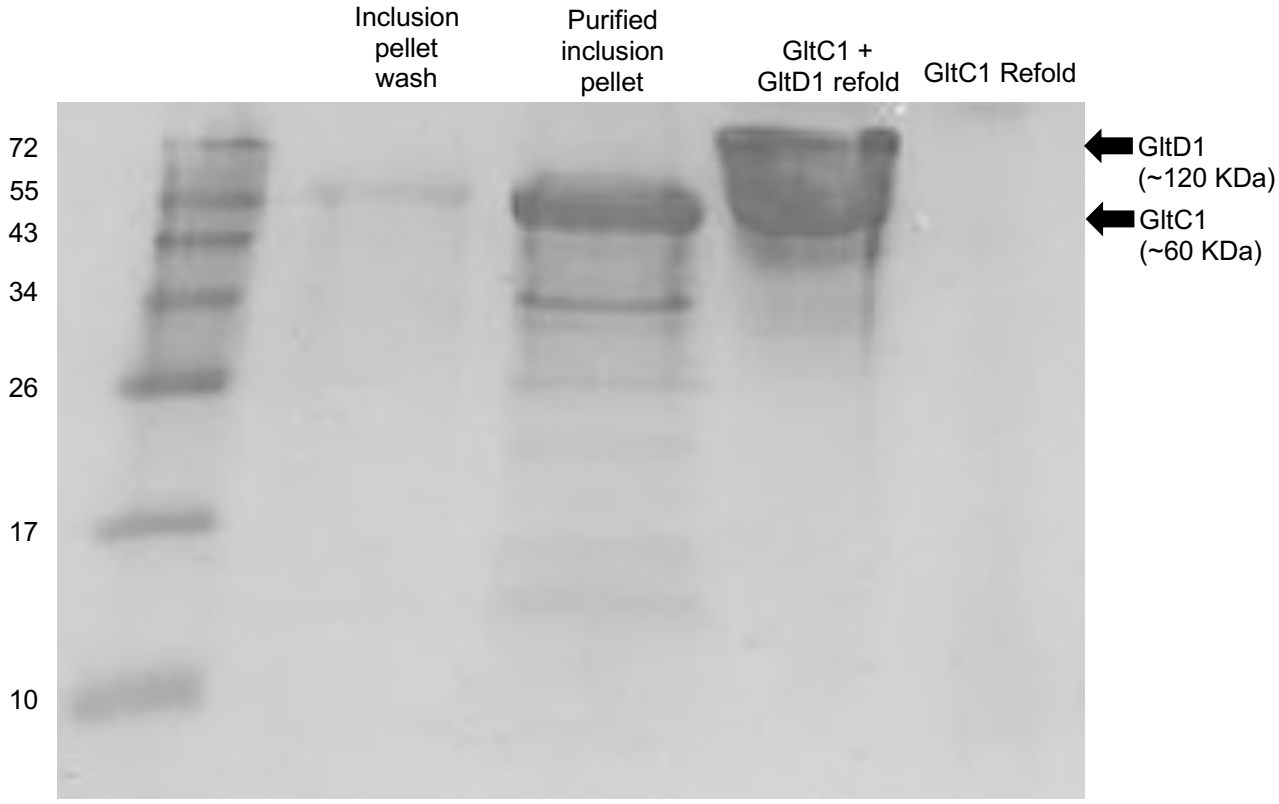


Figure 3.47: Refolding trial of GltC1
Attempts were made to refold GltC1 alone and in the presence of GltD1. Refolding was performed by shock dilution of solubilised inclusion bodies. GltC1 appears to refold in the presence of GltD1 but not when refolded alone, suggesting that GtlD1 may act as a scaffolding to aid refolding.

3.3 Discussion

3.3.1 GltD1:GltE1 interaction – connectivity and orientation of the Agl-Glt complex

The GltD1$_{Cterm}$:GltE1 structure represents the first experimental structural insight into the connectivity within the Agl-Glt complex and provides information into the potential orientation of the Agl-Glt complex within the cell envelope. As previously discussed, GltE1 is expressed as a lipoprotein and sequence analysis demonstrates that none of the known Lol avoidance signals are present in the GltE1 sequence, suggesting that GltE1 localises to the OM[72,110,112–114]. The confirmation of GltE1's interaction with GltD1 strongly suggests that GltE1 would be located on the inner leaflet of the OM. The fact that GltE1 interacts with GltD1 whilst being tethered to the OM considerably constrains the number of possible conformations that the Agl-Glt complex encoded by the Glt1 operon could occupy. The composite AF model of the periplasmic and OM components of the Glt1 operon supports the localisation of GltE1 to the OM, with GltE1 interacting at the face of GltD1 that also interacts with GltC1 and therefore by extension, the OMPs GltA1 and GltB1.

AF was used to model *M. xanthus* GltD (GltD_MX) and GltE (GltE_MX) in complex. This co-folded model revealed that GltD_MX and GltE_MX are predicted to interact in a similar manner to that seen for *B. bacteriovorus*, with a flexible tail peptide of GltD_MX threading through a hydrophobic groove along the concave face of GltE_MX (Figure 3.48). Despite the interaction being maintained, there are several structural differences between the *M. xanthus* and *B. bacteriovorus* proteins, with the largest deviation occurring in GltE_MX. GltE_MX is considerably larger than its *B.*

bacteriovorus ortholog, comprising 9 TPR repeats. This increased size leads to extension of the concave face and thus increases the size of the interface. Interface analysis with PISA[168] calculates the buried surface area of the predicted interface at 3241 Å^2, a considerable increase to buried surface area of 2163 Å^2 seen in the GltD1$_{Cterm}$:GltE1 structure. The C-terminus of GltE_MX also contains an additional elongated α-helix that could function to mediate additional interactions. The conservation of GltD:GltE interaction strongly suggests that insights into the *B. bacteriovorus* Agl-Glt complex can be extrapolated to *M. xanthus*.

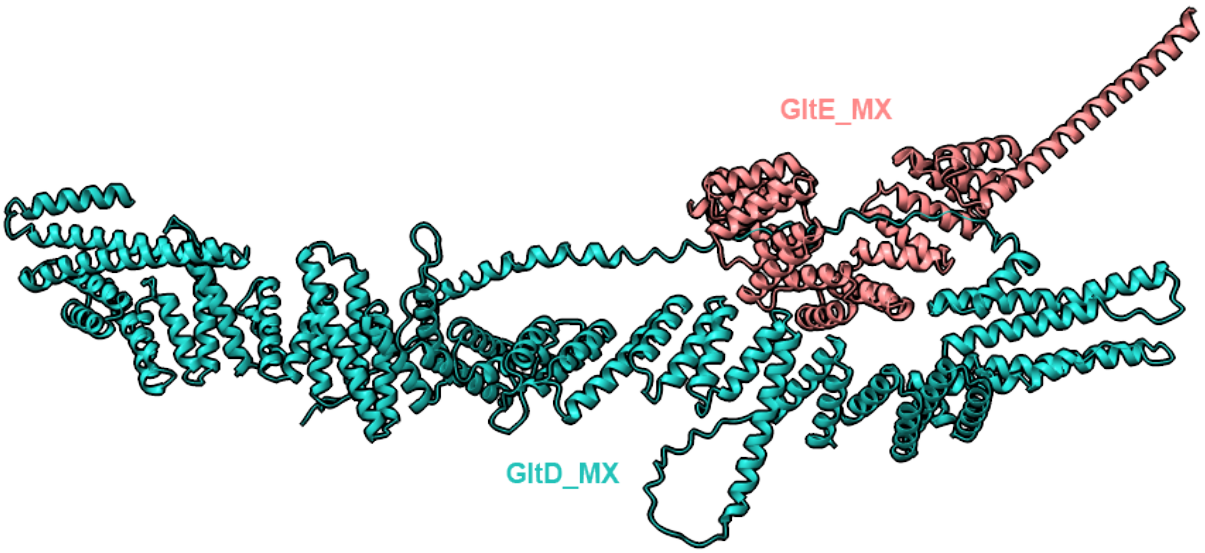


Figure 3.48: AF model of GltD_MX:GltE_MX
AF model of the GltD_MX (coloured blue) and GltE_MX (coloured pink) complex. AF predicts that the *M. xanthus* orthologs interact in a similar manner seen in the GltD1$_{Cterm}$-GltE1 crystal structure with the tail peptide of GltD_MX packing against the concave face of GltE_MX. Average pLDDT (confidence) scores for AF models are as follows: GltD_MX = 80.57; and GltE_MX = 76.98.

3.3.2 Similarity between the *B. bacteriovorus* GltD paralogs

Despite relatively low sequence identity (~22%), the structures of GltD1$_{Cterm}$:GltE1 and GltD3 show remarkable structural similarity with RMSDs of 1.21 Å (GltD3 open conformation, RMSD calculated across 104 atom pairs using ChimeraX) and 1.18 Å (GltD3 closed conformation, RMSD calculated across 104 atom pairs using ChimeraX) (Figure 3.49). In both GltD1$_{Cterm}$:GltE1 and GltD3, the binding of the flexible tail peptide to the auxiliary lobe/GltE1 occurs via the same mechanism, with the peptide occupying the entirety of the concave face of the TPR superhelix. The structural similarity between GltD1$_{Cterm}$:GltE1 and GltD3 provides compelling evidence to support the hypothesis that GltD3 represents a gene fusion of GltD and GltE, where the auxiliary lobe of GltD3 would function akin to GltE. As a result of this gene fusion, the auxiliary lobe GltD3 would not be tethered to the OM in the same manner that GltE would be. The loss of the OM tether in the Glt3 operon may confer more conformational freedom, as the auxiliary lobe is not required to stay in the same plane as the OM, although it is plausible that the auxiliary lobe of GltD3 would still occupy the same relative position of GltE in the cell envelope. The reason for the GltD:GltE fusion in the Glt3 operon remains unclear. Our ITC data shows that GltD1 and GltE1 interact with nanomolar affinity. The high affinity of this interaction suggests that once GltE1 binds to GltD1, it may not dissociate. The fusion of the two genes would likely retain the same function whilst also enforcing correct assembly.

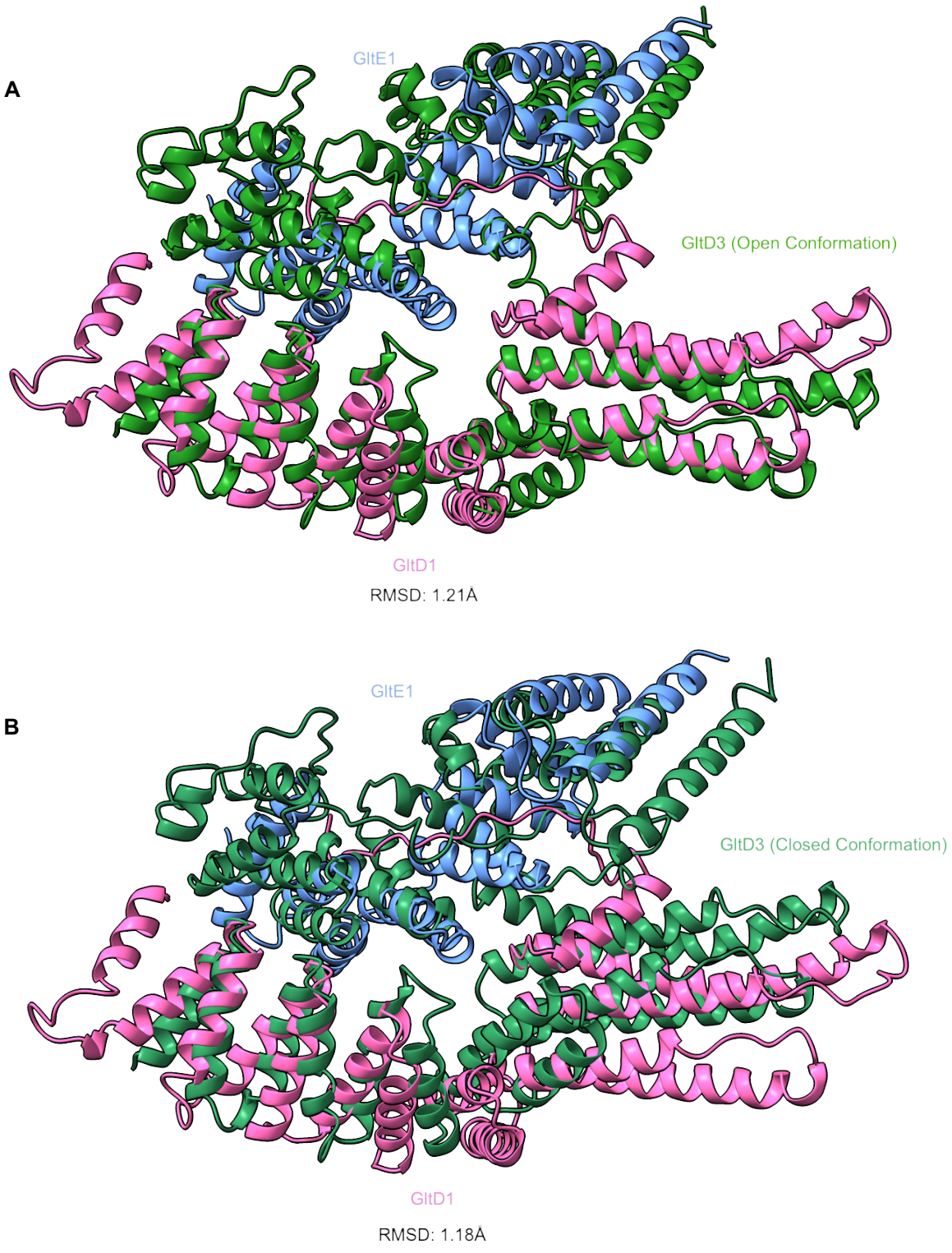


Figure 3.49: Superimposition of GltD1$_{Cterm}$-GltE1 crystal structure and GltD3 cryo-EM structures
A – Superimposition of GltD1$_{Cterm}$-GltE1 (coloured pink and blue) and GltD3 (coloured green) in the open conformation. RMSD =1.21 Å.
B – Superimposition of GltD1$_{Cterm}$-GltE1 (coloured pink and blue) and GltD3 (coloured green) in the closed conformation. RMSD =1.18 Å.

The overall function of GltE in gliding remains to be uncovered. Previous studies have suggested that GltE functions to aid the folding of GltD and/or to protect GltD from the activity of periplasmic proteases[95]. However, in this study, GltD showed signs of degradation even in the presence of GltE; albeit this is in the context of overexpressed, recombinant proteins and may not truly reflect a protective interaction that could occur within the *B. bacteriovorus* cell. Based on the composite AF model, GltE1 appears to work in conjunction with GltD1 to help guide GltG1 through the periplasm. It is possible that GltE/GltD3 auxiliary lobe may function to regulate the conformational flexibility of GltD whilst also contributing to the 'scaffolding' to aid GltG guidance through the periplasm.

Whilst $GltD1_{Cterm}$:GltE1 and GltD3 share strong structural similarities and likely represent the same functional unit, *B. bacteriovorus's* remaining GltD paralogs appear to exhibit structural deviation. The full-length crystal structure of GltD4 contains no auxiliary lobe (Figure 3.50) and the operon contains no gene encoding a potential GltE protein. As no structural information exists for GltD2, an AF model was generated which revealed that GltD2 also lacks an auxiliary domain and like GltD4, contains a helical bundle at the C-terminus (Figure 3.50). Similarly, the GltD2 operon also contains no gene encoding for GltE. These findings raise further questions as to the function of GltE. The deletion of GltE1 abolished gliding motility, demonstrating that GltE1 is essential for gliding[138]. Likewise, the fusion of GltE as the auxiliary lobe of GltD3 implies that the auxiliary lobe plays an essential role in the Glt3 operon. It is therefore interesting that a protein that is essential for gliding in both the Glt1 and Glt3 operons is completely absent in both the Glt2 and Glt4 operons.

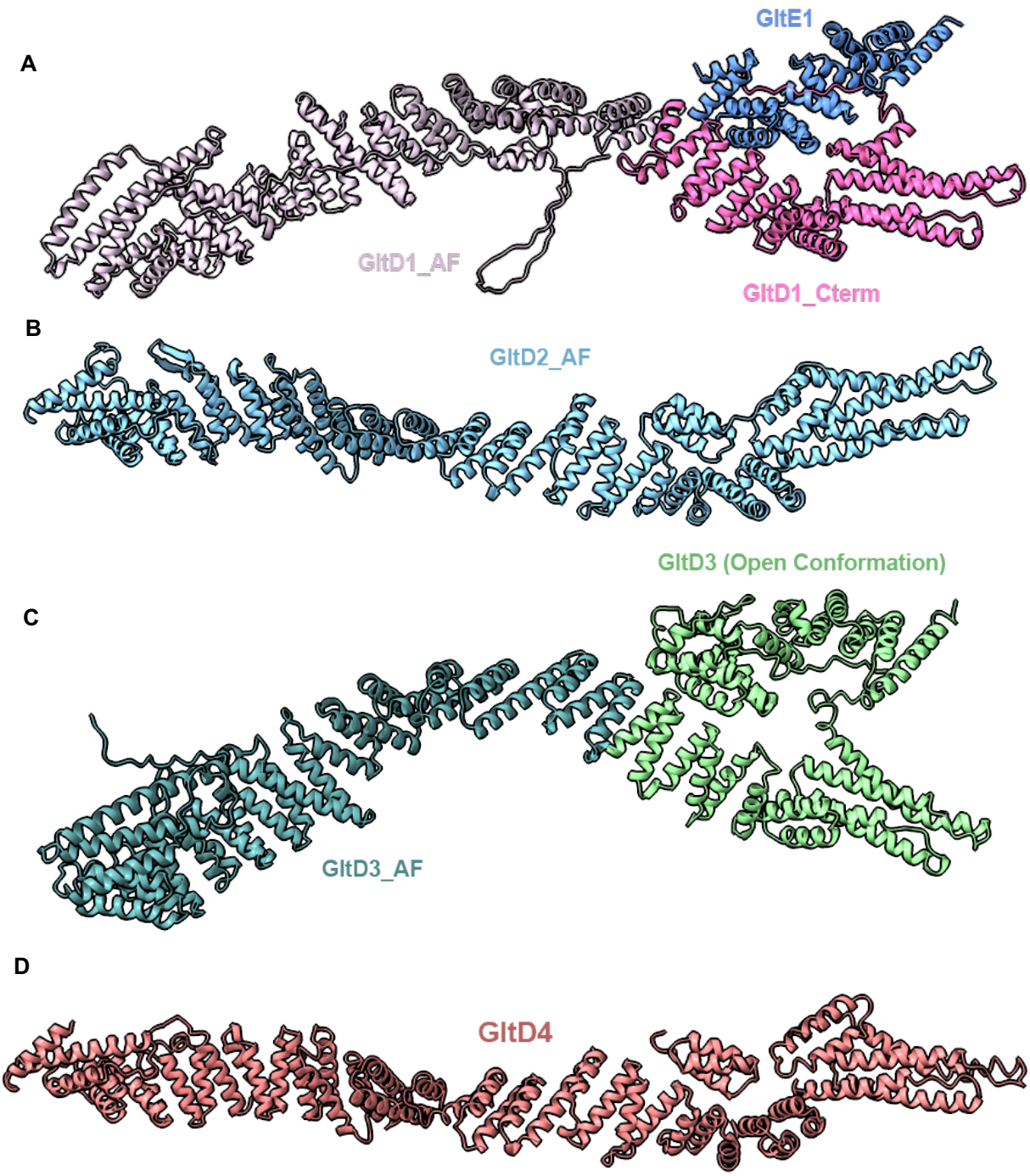


Figure 3.50: Structural comparison of *B. bacteriovorus* GltD paralogs
A – Composite model of GltD1$_{Cterm}$-GltE1 crystal structure (coloured pink and blue, respectively) and GltD1 AF model (coloured lilac). Average pLDDT (confidence) score for GltD1 AF = 84.44.
B – GltD2 AF model. Average pLDDT (confidence) score for GltD2_AF = 87.23.
C – Composite model of GltD3 EM structure (coloured light green) and GltD3 AF model (colured dark green). Average pLDDT (confidence) score for GltD3 AF = 83.56
D – Crystal structure of full-length GltD4.

It is possible that not all Glt operons encode functional gliding complexes; to date, only the Glt1 operon has been shown to be essential for gliding[138]. As discussed in Chapter 1, the Glt and nfs complexes of *M. xanthus* share a strong evolutionary history but perform distinct biological functions[100,175,176]. It is therefore possible that Glt operons that contain a GltE/GltE-like auxiliary lobe and those that lack GltE perform distinct biological roles, although this requires further study.

3.3.3 Insights from AF modelling

AF modelling has provided an initial insight into how the periplasmic and OM Glt components could assemble into a functional unit. The composite AF model generated shows how GltG1 could be guided through the periplasm by GltD1 and GltE1 to bind its putative target ton-box motif on GltF1. GltF1 also appears to link GltD1 and GltC1, the latter of which interacts with GltA1 and GltB1 in the OM. Together, the composite model shows how the Glt complex could form an interconnected complex that spans the cell envelope.

Although many of the modelled interactions still require experimental validation, the similarity to known homologous systems adds some biological strength to the composite model. The GltF1:GltG1 interaction in particular is modelled as a canonical TonB:ton-box interaction. Currently, there are two structures of TonB:TBDT complexes deposited in the PDB. These structures reveal the interaction between both TonB and the Vitamin B12 transporter BtuB and TonB and the Ferrichrome transporter FhuA[177,178]. In both structures, the TonB:ton-box interaction occurs via β-strand exchange[177,178]. The C-terminal domain of TonB contains a four stranded β-sheet. In

the apo structure, β-strand 4 forms an anti-parallel β interaction with β-strand 3[179]. However, in the presence of the ton-box motif, β-strand 4 undergoes a conformational change, allowing the ton-box motif to form a parallel β interaction with β-strand 3 to complete the fold[177,178] (Figure 3.51). This phenomenon of β-strand exchange is also observed when GltG1 is modelled in the presence of GltF1.

In this study, attempts were made to validate the GltD1:GltE1:GltG1$_{491–556}$ interaction that was observed in the AF model. Despite being confidentially modelled by AF, no interaction was seen during experimental attempts to validate the interaction using ITC. Whilst this initially questions the accuracy of the GltD1:GltE1:GltG1 AF model, the lack of interaction can potentially be explained by considering GltG's homology to the ExbB-ExbD-TonB system. To permit nutrient transport, TonB binds to both ExbD and the ton-box motif of the TBDT[14]. Molecular dynamic simulations reveal that hairpin structures form along the length of the disordered region of TonB[179–181]. The binding of TonB to either ExbD or the ton-box motif is hypothesised to put tension on the disordered region, resulting in destabilisation of the hairpin structures to expose further binding sites[181]. It is therefore possible that, in solution, the GltG1 construct used is in an inhibited conformation, whereas AF could be modelling the disordered peptide pulled taut under tension. If this were the case, it is possible that surface plasmon resonance (SPR) may be a more appropriate technique to measure the interaction between GltG1$_{491–556}$ and GltD1 as having the option to bind GltG1$_{491–556}$ to a solid surface may allow the peptide to be tested under tension.

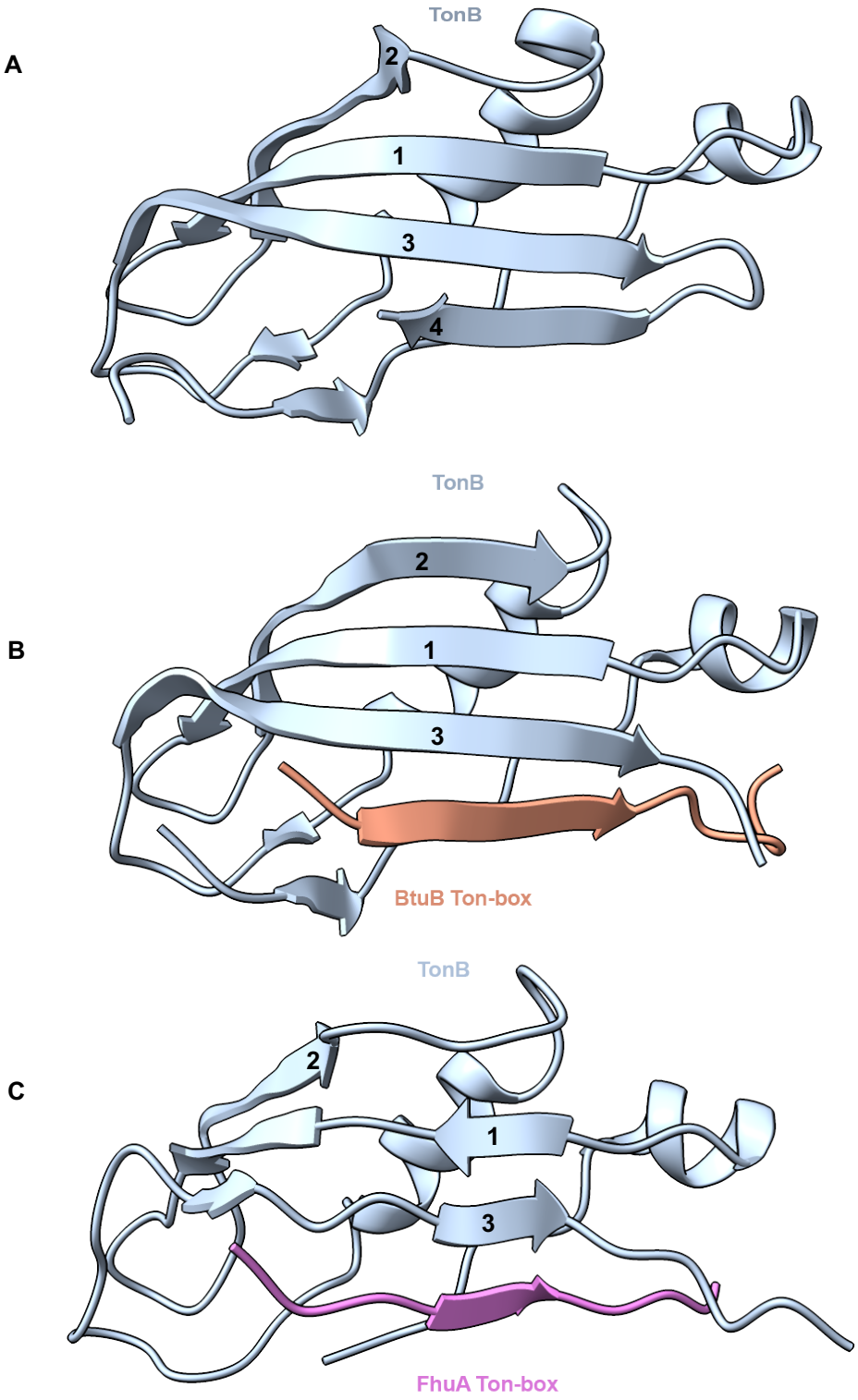


Figure 3.51: Crystal structures of TonB:ton-box interactions
In the presence of the ton-box motif, the TonB domain undergoes conformational change which sees the reorientation of β-strand 4 allowing the ton-box to complete the fold. β-strands are labelled 1–4.
A – Crystal structure of apo TonB (coloured blue) (PDB: 1XX3)
B – Crystal structure of TonB (coloured blue) binding the ton-box motif of BtuB (coloured orange) (PDB: 2GSK)
C – Crystal structure of TonB binding the ton-box motif of FhuA (coloured pink) (PDB: 2GRX)

3.3.4 Proposed mechanism of gliding

Of the proposed mechanisms of gliding, the FA model is currently the most widely accepted in the field[97,134]. Based on a range of microscopy, interaction and genetic studies, the FA model proposes that, once assembled, the Agl-Glt complex interacts with an extracellular substrate via a surface adhesin[91,94,97,134,135]. The rearward trafficking of surface attached Agl-Glt complexes is proposed to create traction to propel the cell forward[91,94,97,134,135]. Together, the structural data obtained during this work, both experimentally and computationally, provide insights into how interactions within the Agl-Glt complex encoded by the Glt1 operon could facilitate gliding.

GltG1, embedded in the IM through a single α-helix, would interact with the AglRQS motor via direct interaction with AglR. In the ExbB-ExbD-TonB system, a section of the disordered region of TonB undergoes a disorder-to-order transition to selectively bind between an ExbD dimer[181]. Based on AglR's homology to ExbD, it is reasonable to propose that the GltG1:AglR interaction occurs in a similar manner to TonB and ExbD. The disordered region of GltG1 would then pass through the periplasm following the superhelical architecture of GltD1/E1 to bind the ton-box motif of GltF1. In addition to being the target for GltG1, our AF model suggests that GltF1 functions to link GtlD1 and GltC1 together. Experimental data show that GltC1 forms an OM subcomplex with GltA1/B1[97,99]. Our AF model suggests that this subcomplex forms through direct interaction between GltC1 and GltA1. GltA1 and GltB1 then interact in the OM, forming a heterodimer. Interestingly, the Glt2-4 operons only contain a GltB gene and initial AF modelling suggests that homodimers of GltB likely form in these operons. During this work, we have identified Bd0875 as a candidate for the adhesin of the Glt1 operon. AF

predicts that Bd0875 interacts with GltB1 through β-strand complementation to permit surface exposure of the adhesin.

The binding of GltG1 to GltF1 could represent the transition to an energised gliding complex capable of transducing the energy generated by the AglRQS motor across the cell envelope. Although GltB1 appears to interact with Bd0875, it does not appear to be required for the surface exposure of Bd0875, as in *M. xanthus*, CglB localises to the cell surface in Δ*gltA/B/H* triple knockouts[135]. Therefore, energisation of the gliding machinery likely causes a conformational change in Bd0875 to permit binding to the extracellular substratum and the formation of a FA site. A similar mechanism is seen in the MIDAS adhesins used in apicomplexan gliding[182–184]. The FA model proposes that the surface-attached Agl-Glt complex is then trafficked backwards along the cell[94]. For this to be the case, the AglRQS motor would have to begin moving along the currently unidentified gliding 'tracks' towards the back of the cell. Given its interactions with GltG1, the movement of the AglRQS motor would presumably cause the same relative movement of GltG1. GltG1 would, in turn, pull on GltF1 and by extension, the remainder of the Glt complex, causing the Glt complex to follow the same trajectory as the AglRQS motor. Whilst this mechanism would fit the observations that led to the proposal of the FA model, it remains difficult to see how the gliding cell would maintain its integrity during continuous lateral movement of the Agl-Glt complex through both membranes and the peptidoglycan.

In order to address the peptidoglycan problem, we tentatively propose an alternative mechanism for gliding. This mechanism draws on the structural insights into the

assembly of the Agl-Glt complex obtained in this work and previous studies performed by our group. Given that both termini of GltD1 appear to be within proximity of the OM (the N-terminus through interaction with GltC1 and the C-terminus through interaction with GltE1), it seems unlikely that GltD1 spans the periplasm to connect the IM and OM. Instead, we propose that arrays of GltD sit above the peptidoglycan layer (at the OM side). The idea of arrays of GltD stems from our observations of conserved crystal packing in the obtained crystal structures of GltD. The only exception to this conserved crystal packing is observed in the $GltD1_{Cterm}$:GltE1 structure presented in this work. This anomaly is likely a result of the construct design (as the contacting residues are not included in the $GltD1_{Cterm}$:GltE1 construct) and does not detract from the conserved crystal contacts being indicative of biologically relevant interactions[169]. The notion of functional arrays is supported by early EM studies that identified chain-like strands in the cell envelope of *M. xanthus* cells[185,186]. The strands closely associated with the OM and contained periodic areas of additional density which could correspond to components of the Agl-Glt complex[185,186]. These strands were initially proposed to wrap helically around the cell to act as the gliding track[185,186], although it is worth considering that these strands could themselves be arrays of GltD, providing an explanation to the observed continuous helical arrangement of fluorescent GltD seen in *M. xanthus*[101,187].

We propose that each GltD1 monomer in the array would form a functional periplasmic/OM subcomplex that may resemble our composite AF model. Notably, alignment of our composite model to the crystal lattice of GltD4 reveals that there are minimal clashes between adjacent complexes (Figure 3.52). During gliding, GltG1

would bind GltF1 and the AglRQS motor would then power the translational movement of GltG1 along the IM causing mechanical pulling on GltF1 (Figure 3.53). This pulling would continue as GltG1 moves further through the membrane until the force required to pull GltF1 any further would be greater that the strength of the interaction between GltF1 and GltG1. A similar scenario is seen in the 'wrap and pull mechanism' seen between TonB and TBDTs which causes partial unplugging of the transport channel[108]. At this point, GltG1 would dissociate from GltF1 and would progress through the membrane until it reaches the next GltD1 in the array to continue gliding (Figure 3.53). Such mechanism would explain the available cytological and structural data without the requirement for continuous remodelling of the peptidoglycan, although considerable additional work is required to validate this mechanism.

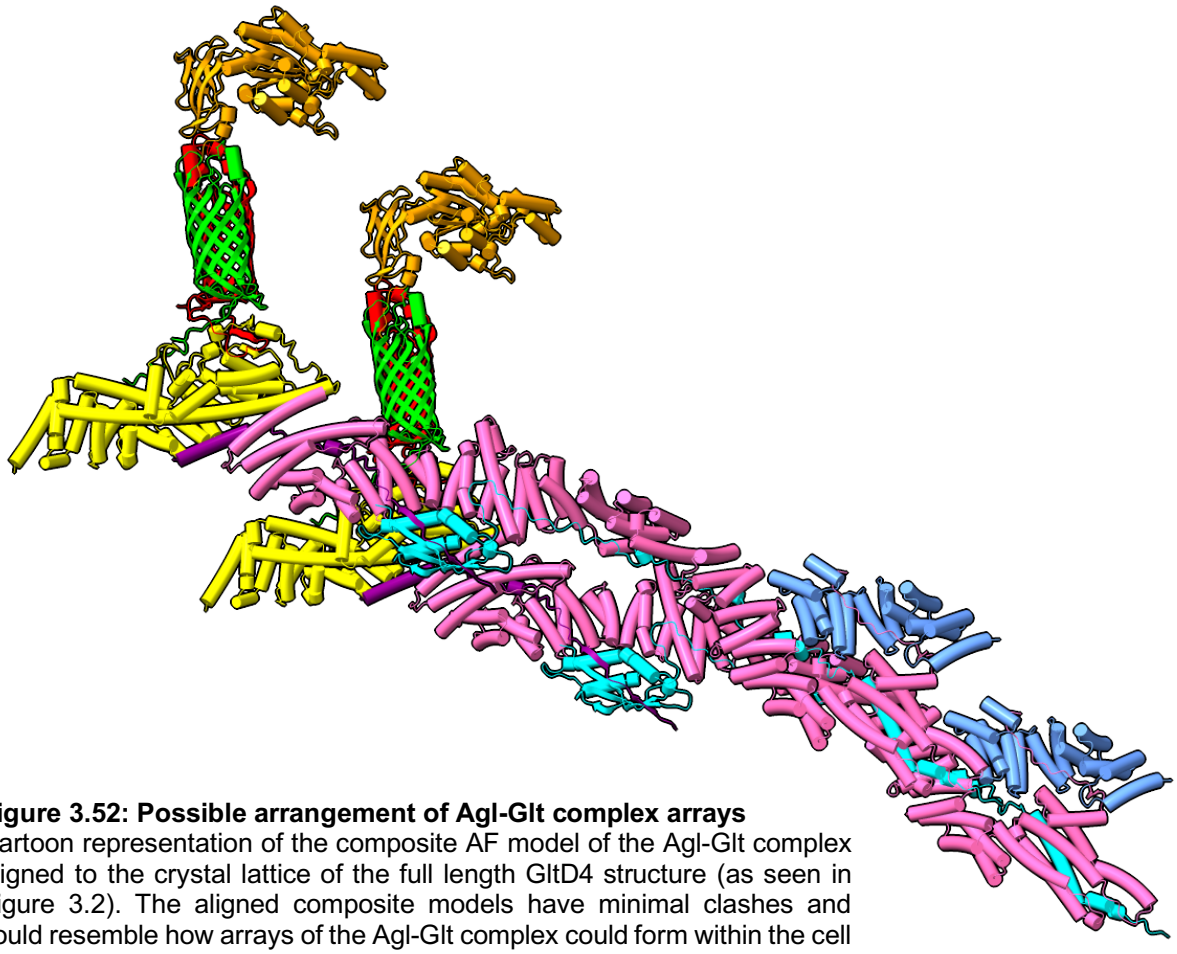


Figure 3.52: Possible arrangement of Agl-Glt complex arrays
Cartoon representation of the composite AF model of the Agl-Glt complex aligned to the crystal lattice of the full length GltD4 structure (as seen in Figure 3.2). The aligned composite models have minimal clashes and could resemble how arrays of the Agl-Glt complex could form within the cell envelope.

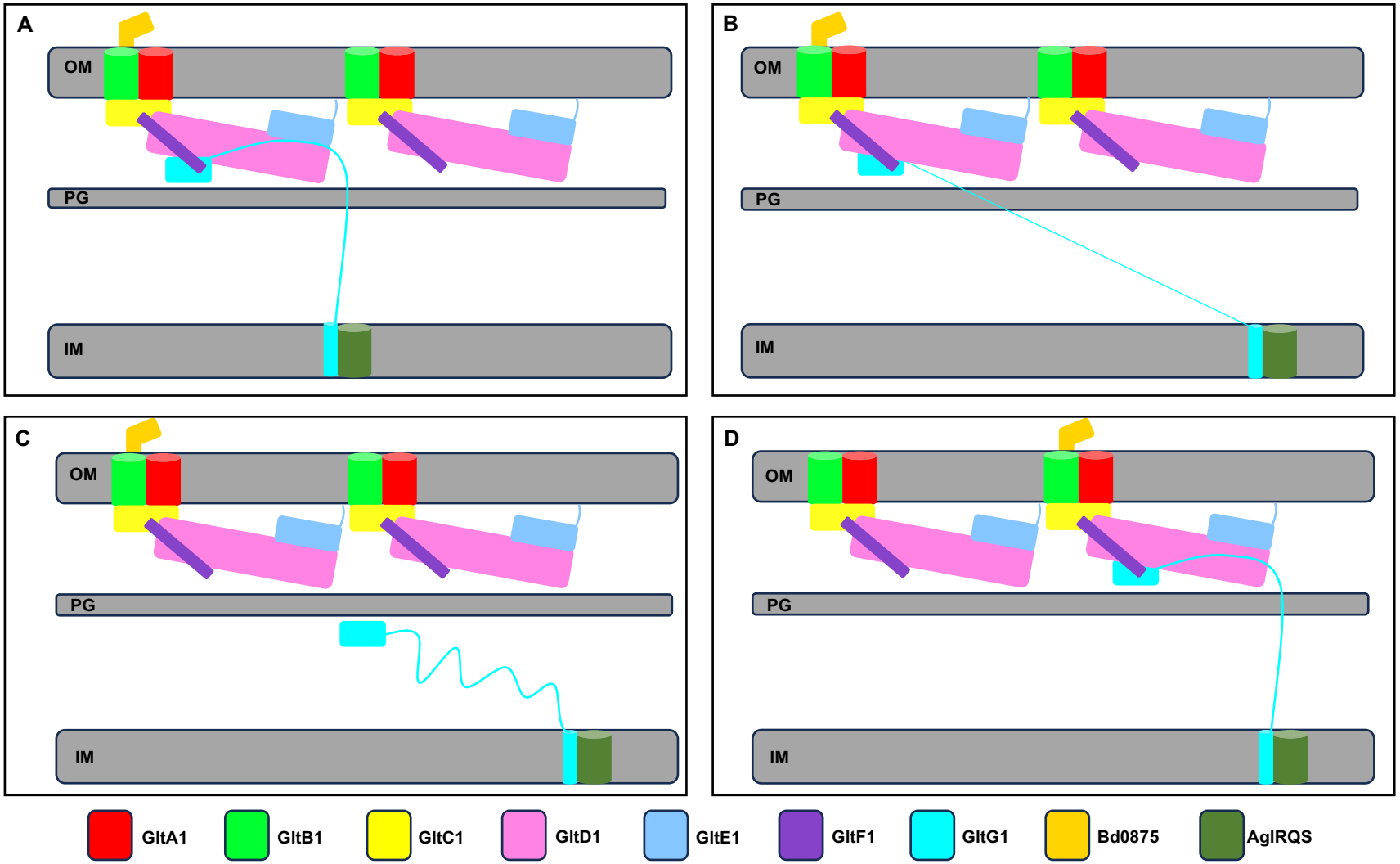


Figure 3.53: Proposed mechanism of gliding
GltG1 binds to GltF1 to facilitate gliding (**A**). The AglRQS motor would then traffic GltG1 through the IM causing mechanical pulling on GltF1 (**B**). This pulling would continue as GltG1 moves further through the membrane until the force required to pull GltF1 any further would be greater that the strength of the interaction between GltF1 and GltG1. At this point, GltG1 would dissociate from GltF1 (**C**) and would progress through the membrane until it reaches the next GltD1 in the array to continue gliding (**D**).

3.3.5 Conclusions and Future Work

The work presented in this chapter provides structural insights into the assembly of the Agl-Glt complex. The validation of the GltD1:GltE1 interactions provides information into the assembly of the gliding complex whilst also strengthening the hypothesised gene fusion event in the Glt3 operon. The GltD1:GltE1 interaction also helps orientate the Agl-Glt complex within the cell envelope and challenges the previous hypothesis that GltD spans the periplasm to interact with both the IM and OM. The AF models generated during this project provided further preliminary information into the assembly of the Agl-Glt complex and allow possible mechanisms of gliding to be proposed.

Moving forward, this work will focus on validating the interactions that have been modelled in this project. In order to facilitate this goal, the composite AF model will be a powerful tool to design constructs to validate the predicted interactions. Our preliminary GltC1 refolding trials provide a promising foundation to test the GltA1:GltC1 and GltC1:GltD1 interactions. The fact that GltC1 only appears to refold in the presence of GltD1 could be considered as initial evidence of GltC1:GltD1 interaction, although the possibility of this being an artefact of refolding cannot yet be ruled out. If GltC1 can be reliably refolded, it will open the possibility of investigating the GltC1:GltD1:GltF1 interaction which would mark a crucial step of understanding the connectivity within the Agl-Glt complex. If pursuits to optimise GltC1 refolding are unsuccessful, the focus should then turn towards GltA1:GltB1:GltC1 co-expression as a means to characterise the OM subcomplex as a whole.

Aside from further structural and functional characterisation of the Glt1 operon, it would be prudent to investigate whether the differences observed between *B. bacteriovorus's*

gliding operons represent functional diversification of the operons. This would have to be largely driven through collaboration with other laboratories specialised in *in vivo* mutagenesis to determine whether we can begin to underpin the distinct function of each Glt operon and their roles in predation.

Chapter 4

Structural characterisation of proteins associated with the regulation of gliding

4.1 Introduction

4.1.1 Hybrid GYF:PilZ proteins may regulate gliding in *B. bacteriovorus*

As previously discussed, gliding in *B. bacteriovorus* is under the regulation of cGAMP signalling[141]. To date, the nature of the cGAMP signalling pathway that regulates gliding remains elusive. However, the presence of a PilZ domain containing protein (Bd1482) on the Glt3 operon provides a compelling candidate for a possible cGAMP receptor responsible for the regulation of gliding. In addition to the C-terminal PilZ domain, Bd1482 also contains an N-terminal GYF domain, a fold that is specialised to recognise and bind proline-rich sequences (PRS)[188], and is part of the domain structure of the myxococcal gliding protein GltJ[94,189].

GYF domains are prevalent in eukaryotes and are characterised by a conserved W-X-Y-X_{6-11}-GP[YF]-X4-[MV]-X2-W-X3-[GN]YF motif[188]. Structurally, GYF domains are composed of a single α-helix and four β-strands that form an anti-parallel β-sheet[190]. The mechanism of GYF domain binding to PRS was uncovered by work on the human T-cell adhesion molecule CD2 and its binding partner CD2BP2, where the binding of CD2BP2 to CD2 regulates the production of interleukin-2 (IL–2)[188,191]. The determination of the NMR structure of the GYF domain of CD2BP2 in complex with a CD2 peptide revealed that the GYF:PRS interaction occurs via the packing of a PPG

recognition sequence of the CD2 peptide into a hydrophobic groove formed by a region of aromatic residues housed on the GYF domain (Figure 4.1)[188,191].

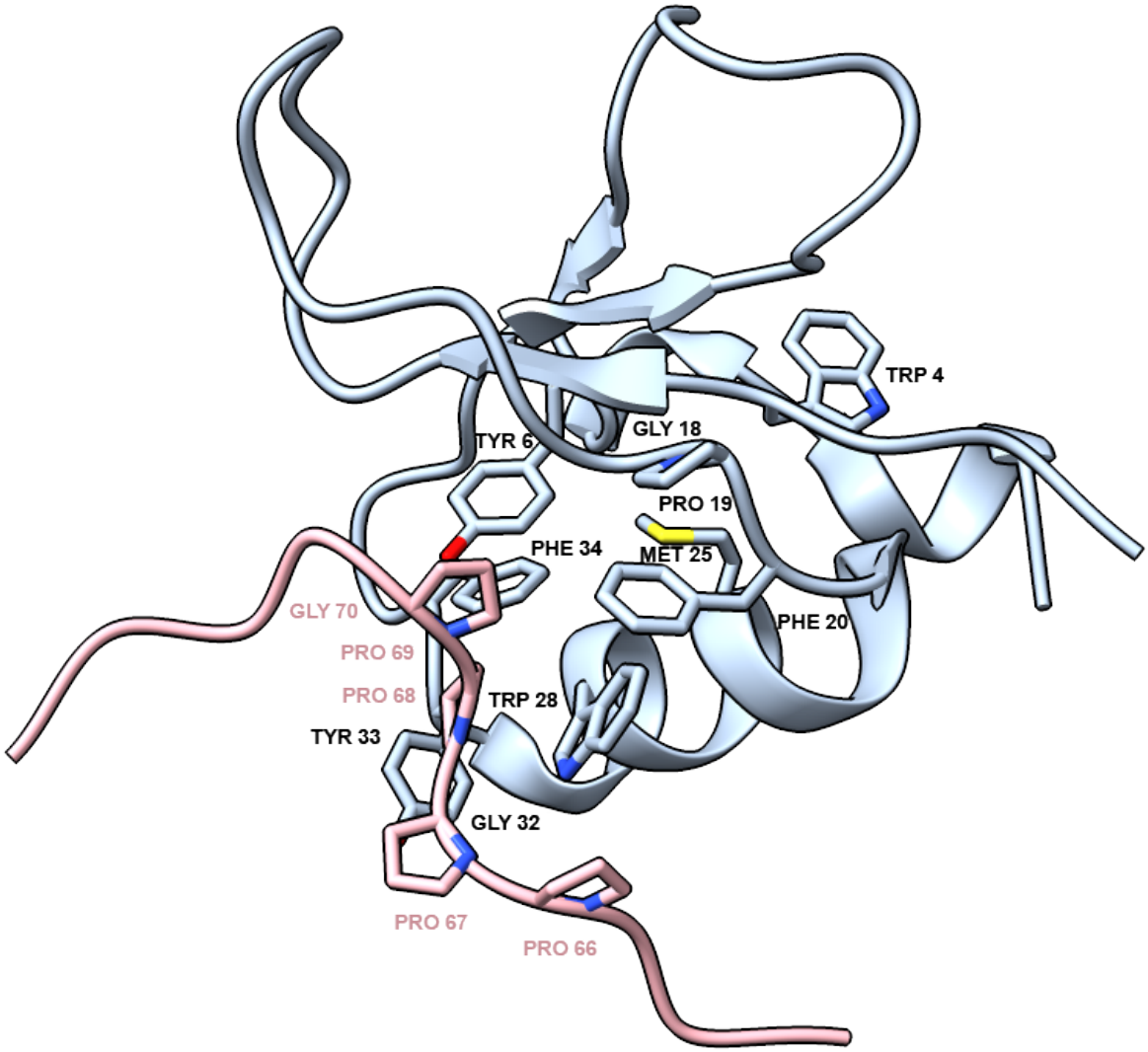


Figure 4.1: Ligand binding mechanism of GYF domains
The GYF domain of CDBP2 (coloured blue) interacts with a PRS (coloured pink) on a CD2 peptide. The PPG recognition sequence packs into a hydrophobic groove formed by the consensus GYF residues. GYF consensus residues and PRS are shown as sticks. (PDB: 1L2Z)

Although present in many prokaryotic genomes, the roles of bacterial GYF domains are poorly characterised. Notably, the only current publication on bacterial GYF domains focuses on the cytoplasmic GYF domain of *M. xanthus* GltJ[189]. Interestingly, the GYF domain of GltJ deviates from the consensus GYF sequence observed in eukaryotes[189]. (Figure 4.2). Nevertheless, NMR assignment of the GltJ GYF domain

indicates that the arrangement of the secondary structure is consistent with that of eukaryotic GYF domains[189]. The overall preservation of the GYF fold, despite the degenerate consensus sequence, could be indicative of differences between ligand binding mechanisms between prokaryotic and eukaryotic GYF domains[189].

In *M. xanthus*, the deletion of GltJ results in defective AglQ, AglZ and GltI localisation[97,189], highlighting that GltJ likely plays a role in the assembly of the Agl-Glt complex[189]. Whilst *B. bacteriovorus* contains no homologue of GltJ, it is plausible to propose that the GYF domains of Bd1482 and GltJ perform similar roles in the regulation of gliding. In contrast to Bd1482, GltJ contains no PilZ domain; though, this is presumably a result of *M. xanthus* gliding being under the regulation of Frz signalling cascades as opposed to cGAMP signalling[75,117,192].

Searching the *B. bacteriovorus* genome for Bd1482 homologs identifies Bd1996 and Bd3100 as hybrid GYF:PilZ proteins. Notably, the genes encoding for Bd1482 and Bd3100 are located adjacent to genes encoding for MIDAS adhesins (Bd1483 and Bd3099 respectively); though it is worth noting that no other genes encoding for additional components of the Agl-Glt complex are located near the Bd3099–Bd3100 gene cluster. In contrast, Bd1996 is a standalone gene; though it has been implicated in the regulation of motility through observations that Bd1996 is the most upregulated PilZ domain-containing protein as cells transition from a mobile to immobile state when under starvation conditions[193]. Sequence comparison of *B. bacteriovorus's* three hybrid GYF:PilZ proteins reveals that both Bd1996 and Bd3100 contain a tandem arrangement of two GYF domains followed by a PilZ domain, whereas Bd1482

contains only one GYF and one PilZ domain (Figure 4.2). Similar to GltJ, the *B. bacteriovorus* GYF domains deviate considerably from the consensus sequence observed in Eukaryotic GYF domains (Figure 4.2). Interestingly, Bd1996 and Bd3100 both contain one GYF domain that shares strong similarity to the consensus sequence of the GltJ GYF domain (Bd1996-GYF-I & Bd3100-GYF-II) and one degenerate GYF domain that shares little similarity to either the Eukaryotic or GltJ GYF sequence (Bd1996-GYF-II & Bd3100-GYF-I) (Figure 4.2).

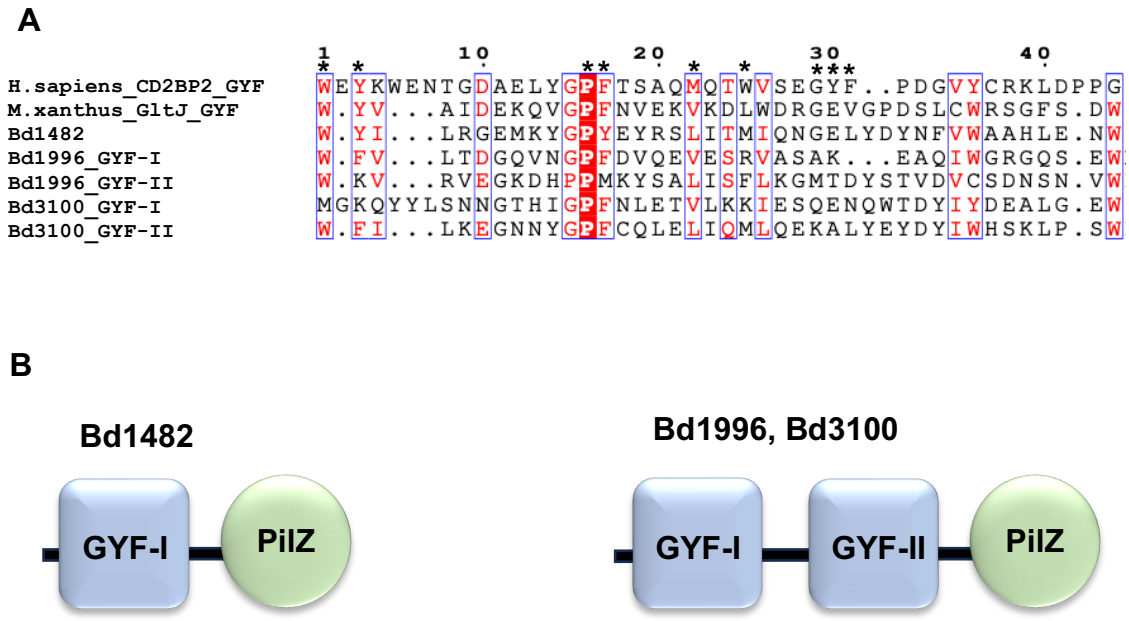


Figure 4.2: GYF domain sequence alignment and domain architecture
A – Sequence alignment between the GYF domains of CD2BP2, GltJ and *B. bacteriovorus's* hybrid GYF:PilZ proteins. GYF consensus residues are marked with an asterisk. Alignment made with Espript3.0[171]. Colour code: Red background, white text = strict identity; white background, red text = similarity in a group; blue frame = similarity across groups.
B – Domain architecture of *B. bacteriovorus's* hybrid GYF:PilZ proteins. Bd1482 contains one GYF and one PilZ domain. In contrast, Bd1996 and Bd3100 contain a tandem arrangement of two GYF domains followed by a PilZ domain.

During previous projects in our lab, we have been unable to produce soluble Bd1482 and Bd3100, with both proteins expressing poorly. However, attempts to structurally characterise Bd1996 have been more successful, with a 3.2 Å x-ray diffraction dataset previously being collected. However, this dataset was observed to be twinned and the combination of low-resolution data and the considerable twin fraction (0.245) meant that it was only possible to build a partial model of Bd1996 (Bd1996-PM). Bd1996-PM comprises two copies of Bd1996 in the ASU, of which only the PilZ domains and a region of the GYF-II domain of one copy can be modelled (Figure 4.3). The two ASU copies of the PilZ domain appear to dimerise via interactions between the aromatic side chains of Tyr195 and Tyr201 of each copy (Figure 4.3). Superimposition of a PilZ:c-di-GMP complex (PDB:5y6g) reveals that the observed dimerisation would occlude the binding of c-di-GMP to the PilZ domain of Bd1996, suggesting that the presence of c-di-GMP/cGAMP may cause monomerisation of Bd1996, potentially freeing the GYF domains to bind a PRS on its binding partner (Figure 4.4). Whilst this partial model provides an initial insight into c-di-GMP signalling via Bd1996, this mechanism requires further validation and the binding partner(s) for Bd1996 remain to be identified.

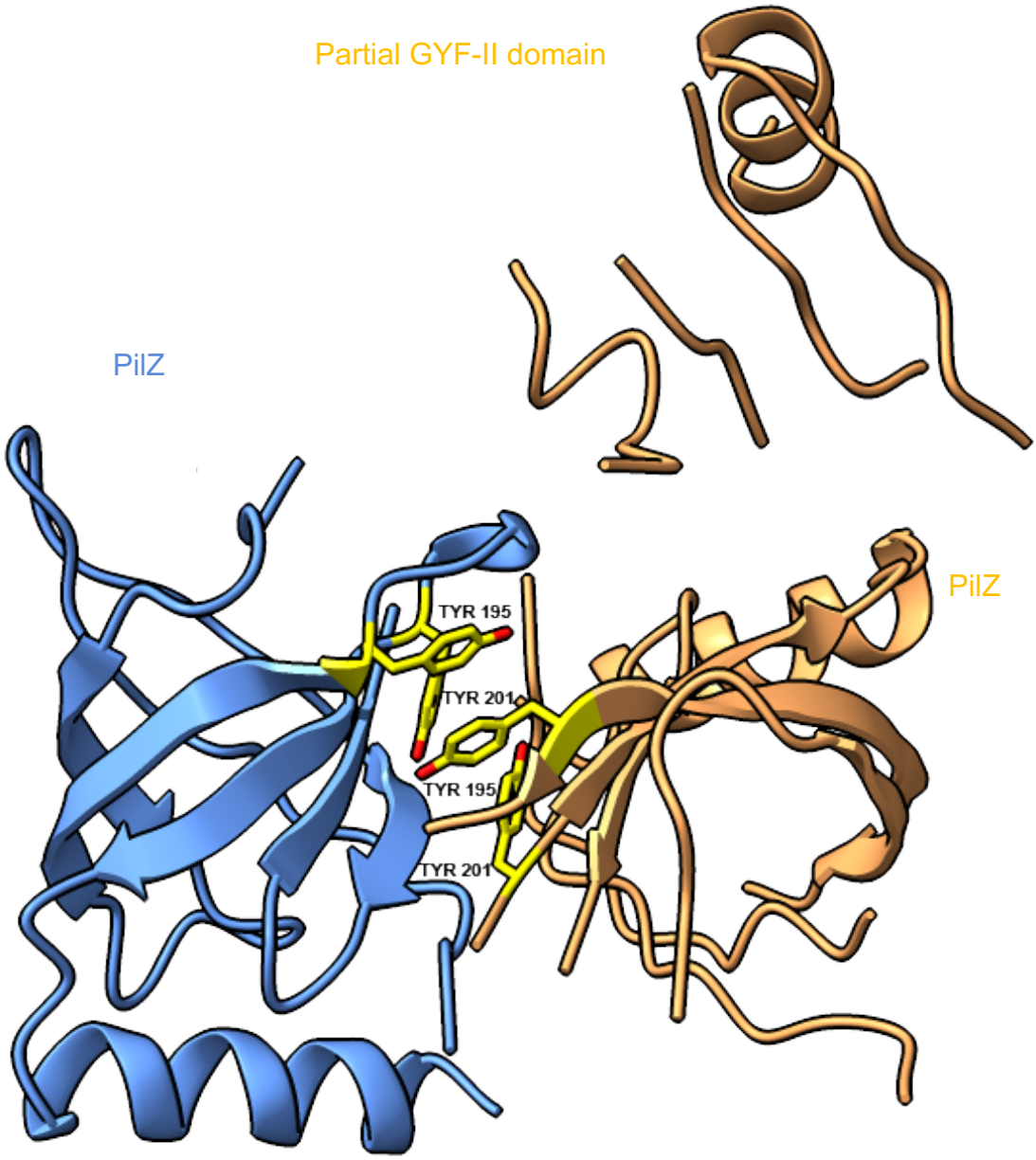


Figure 4.3: Partial crystal structure of Bd1996
In the current crystal structure of Bd1996, only the PilZ domains and a region of one of the GYF-II domains can be built. The partial model suggests that Bd1996 dimerises through an interaction between the side chains of TYR195 and TYR201 of each copy of the PilZ domain. TYR195 and TYR201 are coloured yellow.

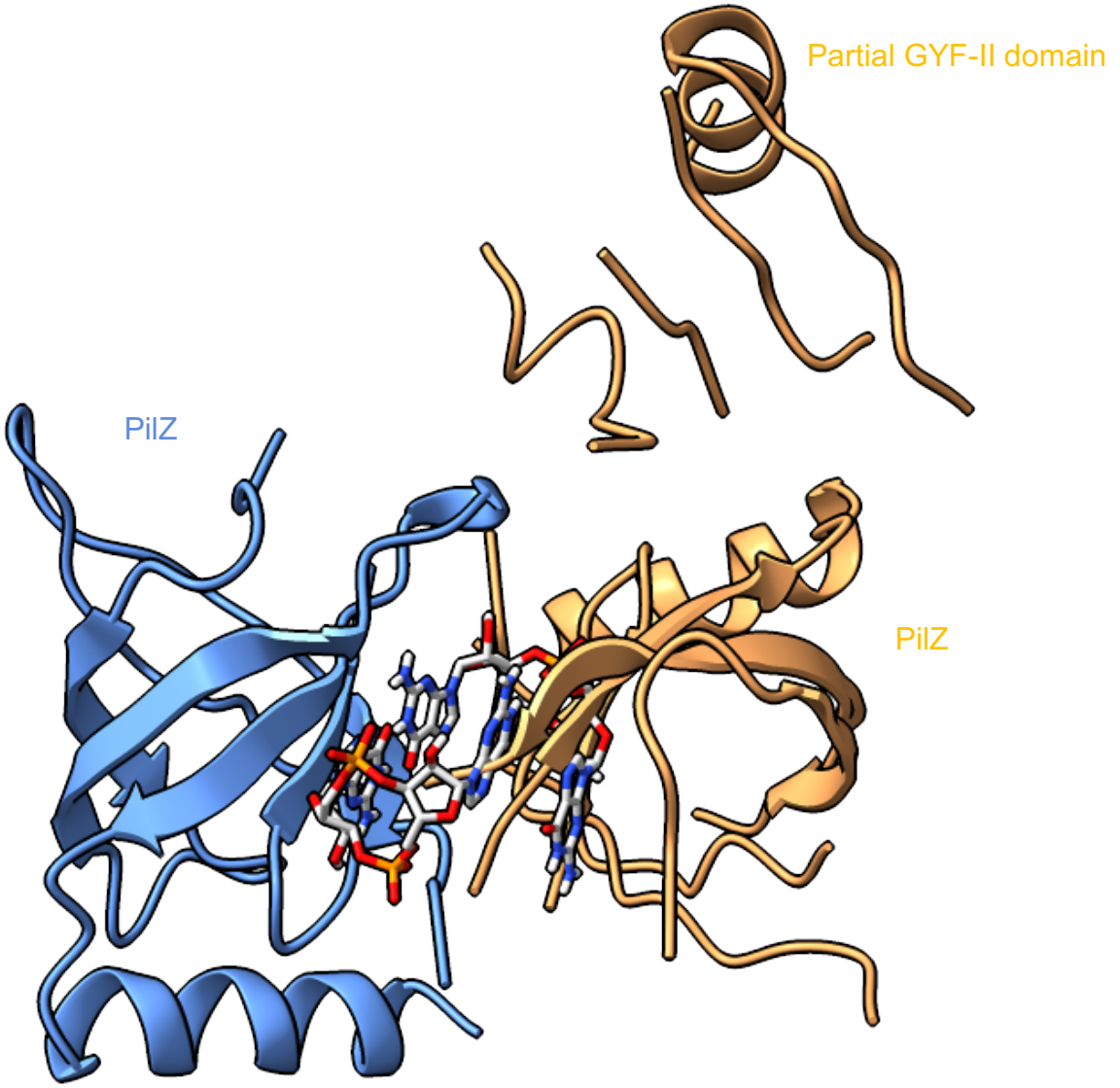


Figure 4.4: Proposed Bd1996 dimerisation prevents c-di-GMP/cGAMP binding
Superimposition of c-di-GMP (from PDB:5y6g) to the partial structure of Bd1996 reveals that the proposed dimerisation of the PilZ domains would prevent the binding of c-di-GMP to the PilZ domains. This suggests that the binding of c-di-GMP/cGAMP would cause Bd1996 to monomerise.

4.1.2 The role of MglA in *M. xanthus* gliding motility

As previously discussed in Chapter 1, the direction of *M. xanthus* gliding is regulated by the activity of the small ras-like GTPase MglA and its cognate GAP MglB, which localise to opposite cell poles to create a polarity axis[83]. GTP activated MglA regulates the recruitment of the Agl-Glt complex at the leading pole[80,83]. During gliding, MglA

travels with the Agl-Glt complex towards the lagging pole[83]. Upon reaching the lagging pole, MglB binds to MglA, triggering the hydrolysis of GTP to GDP and ultimately, the disassembly of the Agl-Glt complex[83,124].

M. xanthus MglA ($MglA_{MX}$) has been subject to considerable structural characterisation, with structures of $MglA_{MX}$-GDP and $MglA_{MX}$-GTPγS (non-hydrolysable GTP analogue) being reported[194]. Structurally, $MglA_{MX}$ comprises a seven-stranded β-sheet and five α-helices that pack around the β-sheet (Figure 4.5)[194]. Like most small GTPases, $MglA_{MX}$ contains a P-loop, where GTP/Mg^{2+} binding occurs, and two switch regions (switch 1 and 2) that interact with regulatory proteins and undergo conformational changes as $MglA_{MX}$ transitions between inactive (GDP-bound) and active (GTP-bound) states (Figure 4.5)[83]. The structure of the $MglA_{MX}$-GTPγS-MglB complex has also been reported[86,194] and the comparison between $MglA_{MX}$-GDP, $MglA_{MX}$-GTPγS and $MglA_{MX}$-GTPγS-MglB structures provides detailed insights into how the interaction between $MglA_{MX}$ and MglB regulates GTP hydrolysis and thus the assembly and disassembly of the Agl-Glt complex.

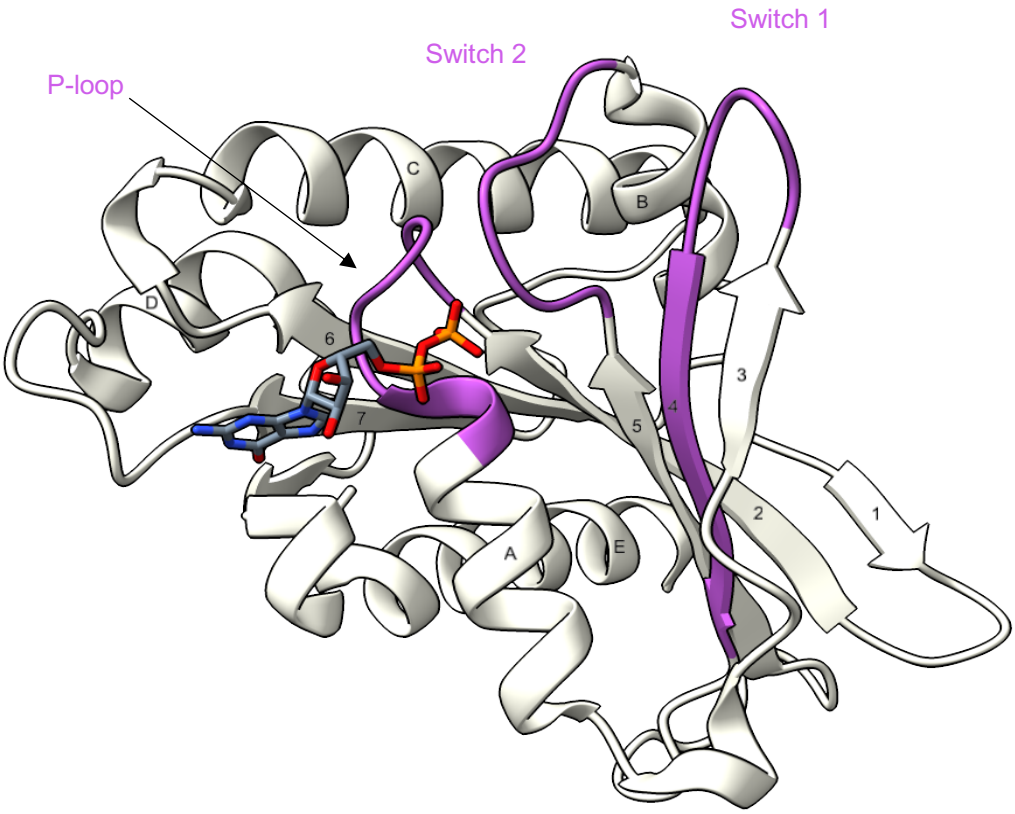
In the inactive (GDP-bound) form of $MglA_{MX}$, the switch 1 region occupies a catalytically null conformation, in which Arg53, a residue essential for catalysis, projects away from the nucleotide-binding pocket (Figure 4.5)[194]. Likewise, switch 2 occupies a self-inhibitory conformation that prevents Mg^{2+} co-ordination and blocks the nucleotide-binding pocket to an extent where there is insufficient space for the γ-phosphate of GTP to bind[83,194]. In the active (GTP-bound) conformation of $MglA_{MX}$, both switch 1 and switch 2 undergo a conformational change (figure 4.5)[194]. When

bound to GTP, β-strand 4 (located on switch 1) twists, shifting the register of the strand by three residues, resulting in the surface presentation of Phe56, Phe57 and Phe59[86,194]. Switch 2 also reorientates to permit the binding of GTP and Mg^{2+} (Figure 4.5)[194]. Finally, Gln82 located on switch 2 reorientates to face into the nucleotide-binding pocket and co-ordinates the water molecule that will ultimately act as the attacking nucleophile[86].

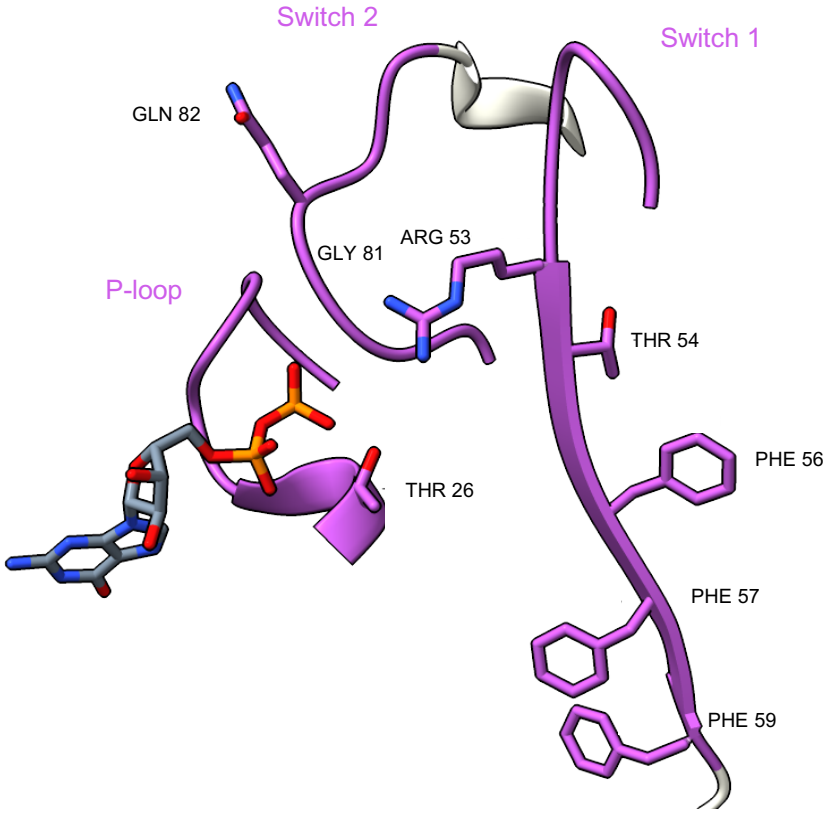
The structure of the $MglA_{MX}$-GTPγS-MglB complex reveals that $MglA_{MX}$ and MglB interact with a stoichiometry of 1:2 where $MglA_{MX}$ binds to a MglB dimer via an asymmetric interface (Figure 4.5)[86,194]. The GTP-bound conformations of switch 1 and switch 2 both interact with MglB[86,194], indicating that the binding of GTP is essential to permit interaction between the two proteins. Despite MglB acting as a GAP, there is no contact between MglB and GTP. Instead, the GAP activity of MglB is mediated by triggering conformational changes in $MglA_{MX}$ to create a full catalytic site for the hydrolysis of GTP. The binding of MglB to $MglA_{MX}$ is characterised by further conformational changes in the switch regions. The predominant changes occur on switch 1, which sees the reorientation of Arg53 into the nucleotide-binding pocket (Figure 4.5)[86,194]. Thr54 also reorientates to complete the co-ordination of the Mg^{2+} ion (Figure 4.5)[86,194].

MglA$_{MX}$-GDP

A

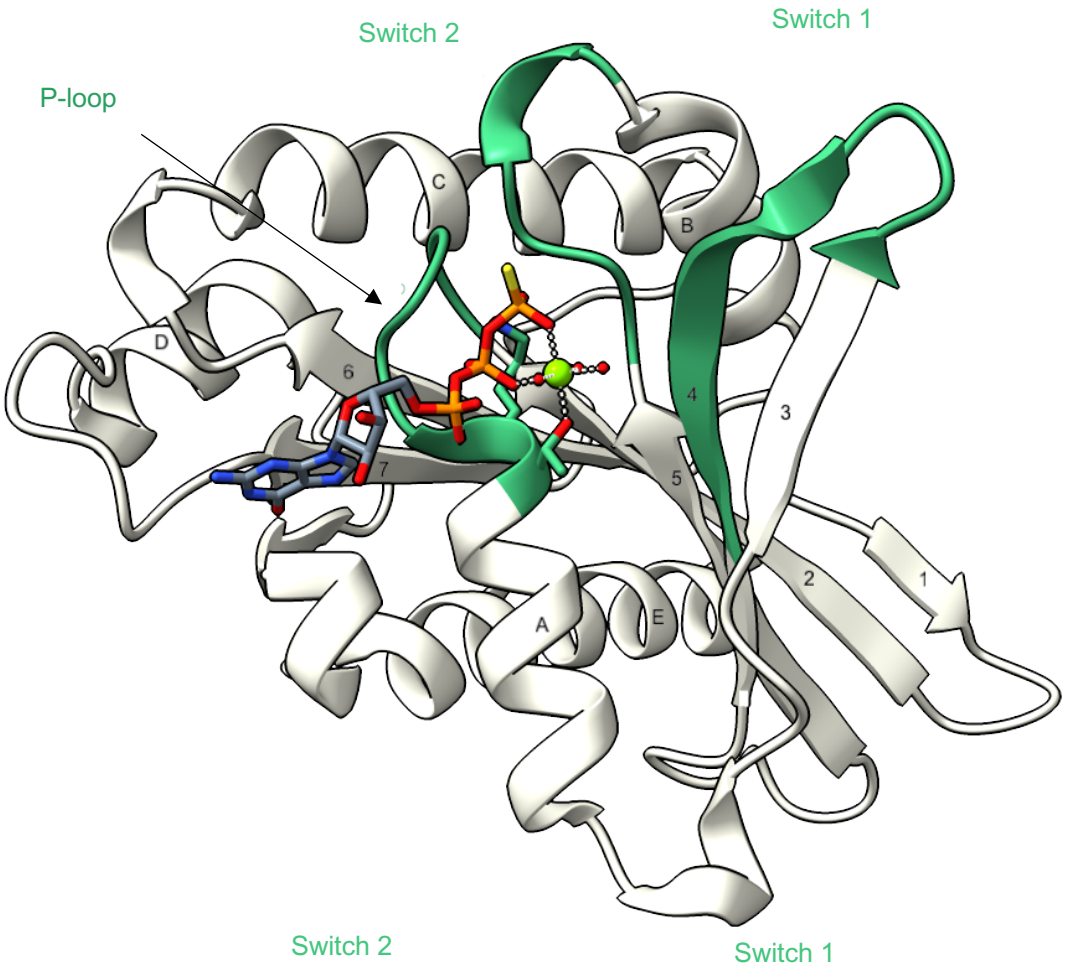


B

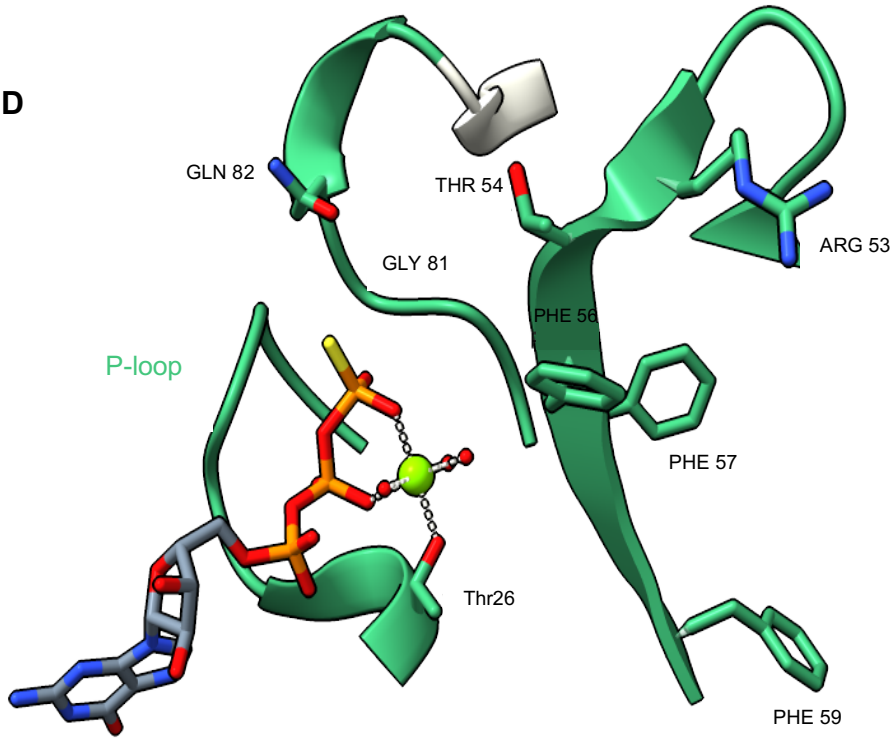


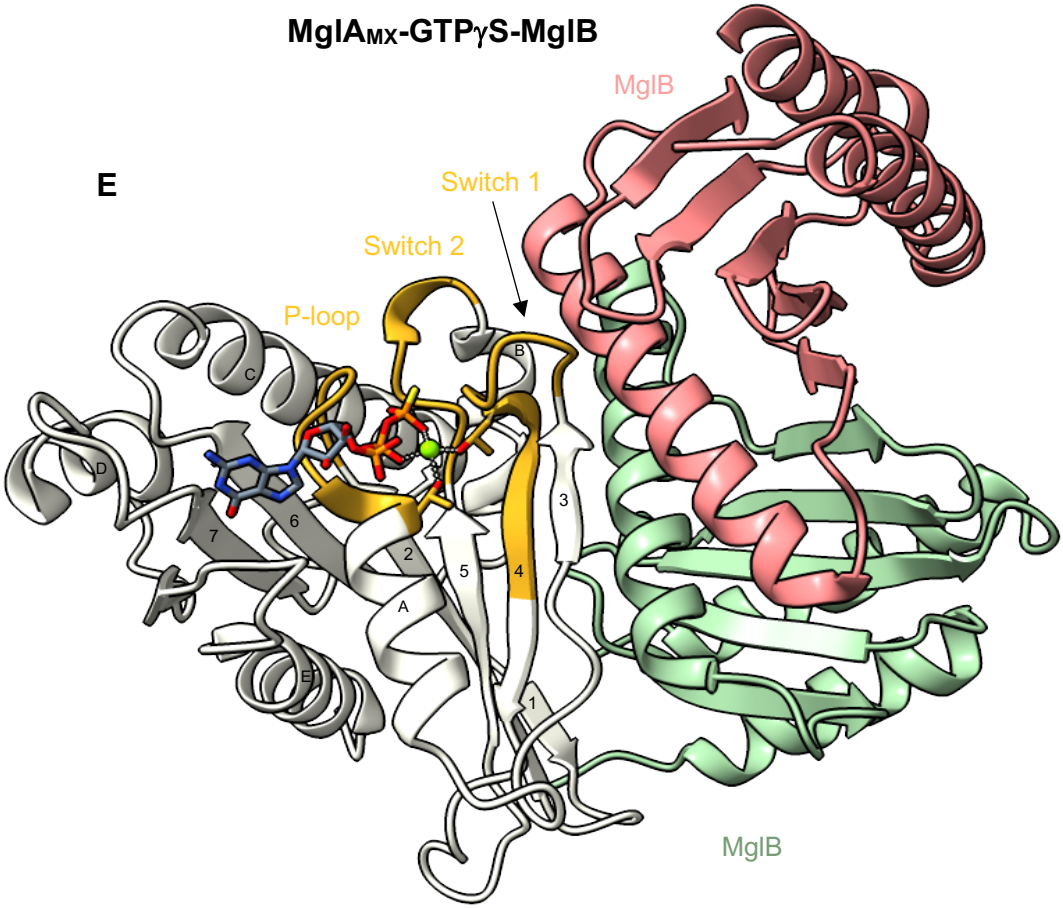
MglA$_{MX}$-GTPγS

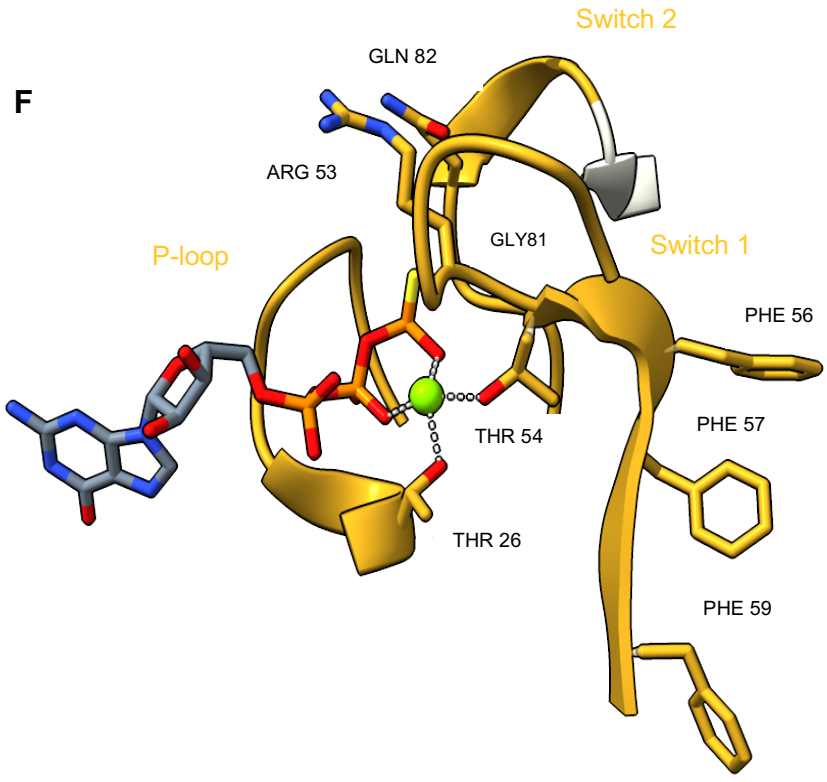
C



D




MglA$_{MX}$-GTPγS-MglB
MglB
E
Switch 1
Switch 2
P-loop
C
B
D
3
6
7
2
5
4
A
1
MglB


Switch 2
GLN 82
F
ARG 53
GLY81
P-loop
Switch 1
PHE 56
PHE 57
THR 54
THR 26
PHE 59

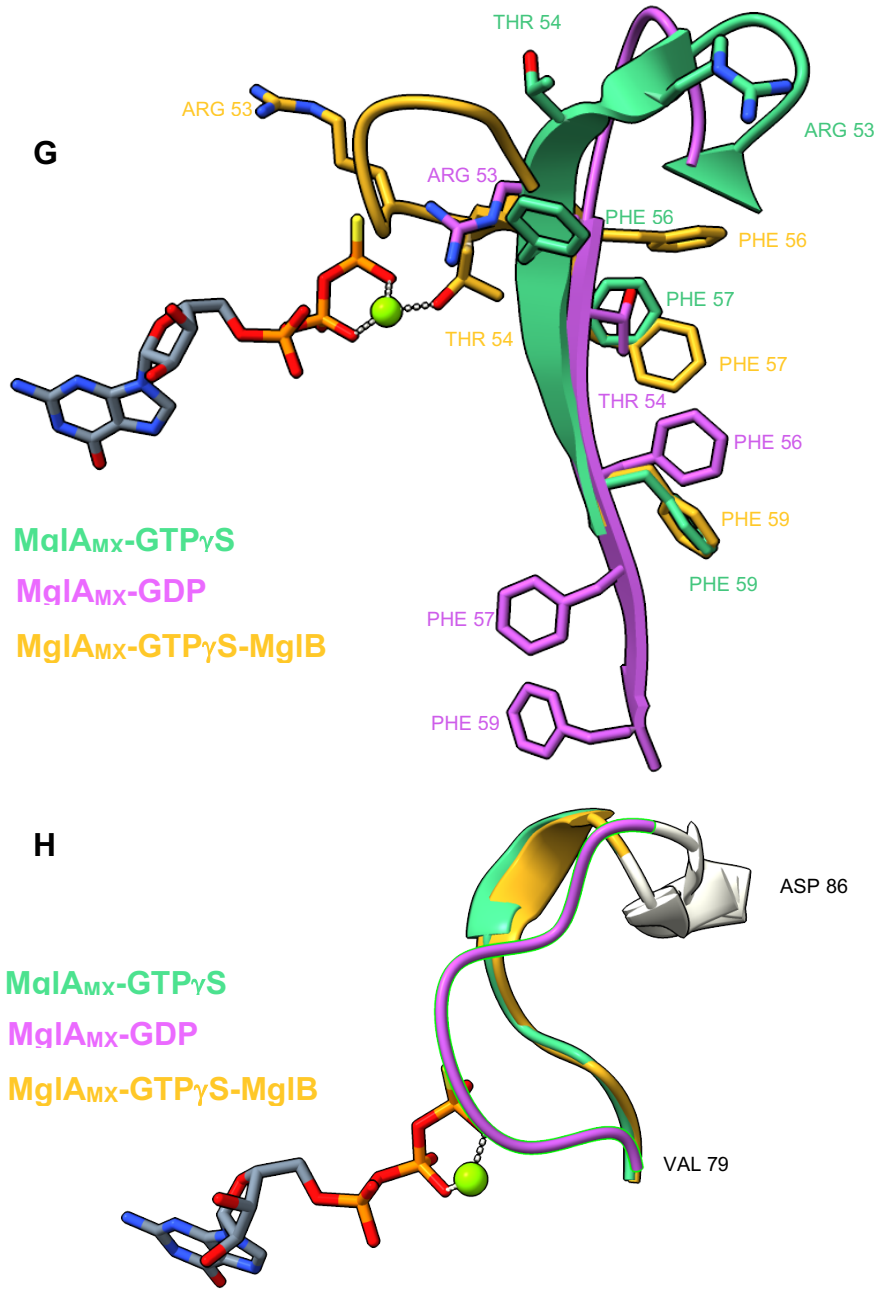


Figure 4.5: The conformational states of MglA$_{MX}$
A–B – Crystal structure of MglA$_{MX}$-GDP (PDB: 6HJO). In this conformation, both switches 1 and 2 are in an inactive state, incompatible with GTP and Mg^{2+} binding.
C–D – Crystal structure of MglA$_{MX}$-GTPγS (PDB: 6H17). In the presence of GTPγS, both switches undergo conformational changes to permit GTP and Mg^{2+} binding; though the co-ordination shell of Mg^{2+} is incomplete.
E–F – Crystal structure of MglA$_{MX}$-GTPγS-MglB (PDB: 6H5B). Upon MglB binding, the switch regions of MglA$_{MX}$ are further re-modelled. Arg53 and Thr54 both reorientate to complete the nucleotide/Mg^{2+} binding sites respectively.
G – Superimposition of the switch 1 region from the MglA$_{MX}$-GDP, MglA$_{MX}$-GTPγS and MglA$_{MX}$-GTPγS-MglB structures.
H – Superimposition of the switch 2 region from the MglA$_{MX}$-GDP, MglA$_{MX}$-GTPγS and MglA$_{MX}$-GTPγS-MglB structures.
Figure adapted from Galicia *et al.,* (2019)[194].

4.1.3 The role of MglA in *B. bacteriovorus*

Like *M. xanthus*, gliding in *B. bacteriovorus* is bidirectional[58], with the direction of gliding presumably controlled by changes in cell polarity. Although *B. bacteriovorus* contains a gene encoding MglA ($MglA_{Bd}$), it lacks an MglB homologue[137]. The absence of MglB and all genes encoding for the Frz signalling system suggests that an alternative mechanism for regulating cell reversals exists in *B. bacteriovorus*[137]. Interestingly, whilst $MglA_{Bd}$ has been implicated in the regulation of cell reversals, it is not essential for gliding[137]. This is in stark contrast to *M. xanthus* where the deletion of MglA abolishes both gliding and S-motility[124]. Instead, the deletion of $MglA_{Bd}$ results in *B. bacteriovorus* cells that can attach to, but not invade into prey[137]. Notably, $\Delta mglA_{Bd}$ cells were observed to have drastically reduced TFP populations at the invading pole[137]. The non-predatory phenotype of $\Delta mglA_{Bd}$ cells, together with the reduction, but not complete lack of TFP, suggests that $MglA_{Bd}$ plays a regulatory role in TFP mediated prey invasion.

$MglA_{Bd}$ and $MglA_{MX}$ (Figure 4.6) share considerable sequence similarity, with the switch regions and catalytic site being conserved between both proteins (Figure 4.6). The P-loop is also well-conserved with the exception of Ser21 in $MglA_{Bd}$ (Gly21 *M. xanthus* MglA). The presence of Ser21 on the P-loop of $MglA_{Bd}$ has led to the proposal that $MglA_{Bd}$ is permanently active, as the equivalent Gly to Ser mutation in the oncogenic Ras GTPase results in Ras being trapped in a permanently active conformation[137,195]. The notion of permanently active $MglA_{Bd}$ has been proposed as an explanation for the lack of MglB in *B. bacteriovorus*[137].

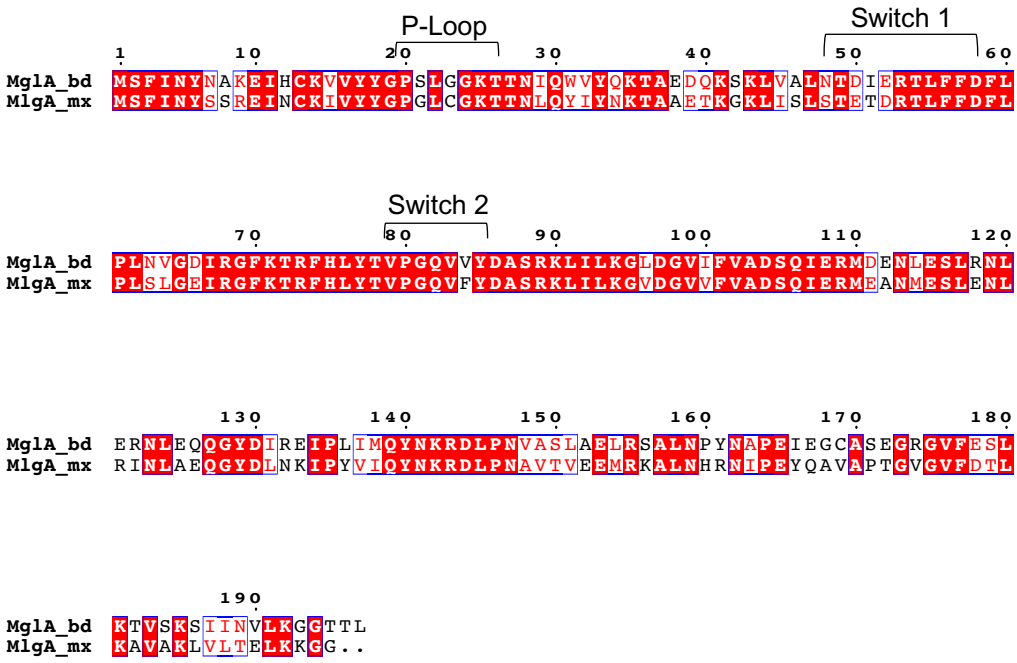


Figure 4.6: $MglA_{Bd}$ and $MglA_{MX}$ sequence alignment
$MglA_{Bd}$ shares considerable similarity to $Mgla_{MX}$ (~65% sequence identity), with the switch regions, P-loop and catalytic site being conserved between both proteins. Positions of the P-Loop and switch 1 and 2 are shown. Alignment made with Espript3.0[171]. Colour code: Red background, white text = strict identity; white background, red text = similarity in a group; blue frame = similarity across groups.

$MglA_{Bd}$ has been shown to interact directly with Bd2492, a TPR-containing protein that is essential for predation[137]. Notably, Bd2492 shares homology to SgmX, a myxococcal protein that interacts directly with $MglA_{MX}$ to regulate the formation of TFP[196,197]. Our lab has made previous attempts to structurally characterise the $MglA_{Bd}$:Bd2492 interaction in order to determine how $MglA_{Bd}$ and Bd2492 function to regulate predation in *B. bacteriovorus*. In previous works, it was found that $MglA_{Bd}$ could not be expressed alone. However, co-expression of $MglA_{Bd}$:Bd2492 from a pETDuet-1 plasmid resulted in the soluble expression of both proteins. Extensive crystallisation trials have previously been performed for $MglA_{Bd}$:Bd2492, but to date, no crystals have been produced.

4.1.4 Chapter Aims

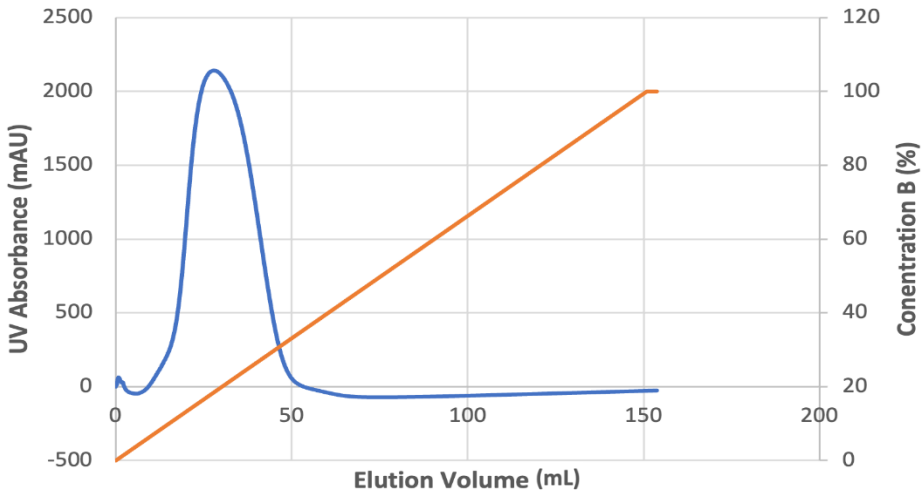
B. bacteriovorus's hybrid GYF:PilZ proteins provide compelling candidates for the cGAMP receptors involved in the regulation of gliding motility. Previous structural characterisation of Bd1996 has resulted in a partial structure which provides initial insights into Bd1996-c-di-GMP/cGAMP signalling, suggesting that cyclic dinucleotide-binding triggers conformational change to permit downstream signalling. In this chapter, we will continue to pursue complete structural characterisation of Bd1996 to gain clearer insights into the potential signalling mechanism. We will also use pull-down assays and phage display to attempt to identify Bd1996 binding partners. Following the successful identification of binding partners, attempts will be made to validate these interactions. Finally, we will continue our attempts to determine the structure of the $MglA_{Bd}$:Bd2492 complex in order to gain a structural perspective for how $MglA_{Bd}$ has evolved to regulate a distinct biological process. In order to do this, we will utilise newly available bioinformatic tools to aid construct design to produce new constructs more amenable to crystallisation.

4.2 Results

4.2.1 Expression and purification of Bd1996 constructs

Due to the low resolution and twinned nature of the previously collected diffraction data for Bd1996, attempts were made to grow different crystal forms. A new construct for Bd1996 with a cleavable N-terminal His-Tag was cloned (Table 2.8). This construct expressed well and was initially purified using IMAC, where the protein eluted as a single peak at ~20% concentration buffer B (100 mM imidazole) (Figure 4.7). SDS-PAGE analysis confirmed the presence of Bd1996 and the peak fractions were pooled and further purified by SEC. Bd1996 was eluted as a monodisperse peak from HiLoad 26/600 Superdex 75 prep grade column at ~188 mL, an elution volume roughly consistent with the size of monomeric Bd1996 (~26 KDa) (Figure 4.7). SDS-PAGE demonstrated the high purity of the post-SEC Bd1996 sample (Figure 4.7). Purified Bd1996 was used in sparse matrix crystallisation trials but unfortunately, no further crystals for the full-length protein were grown.

A



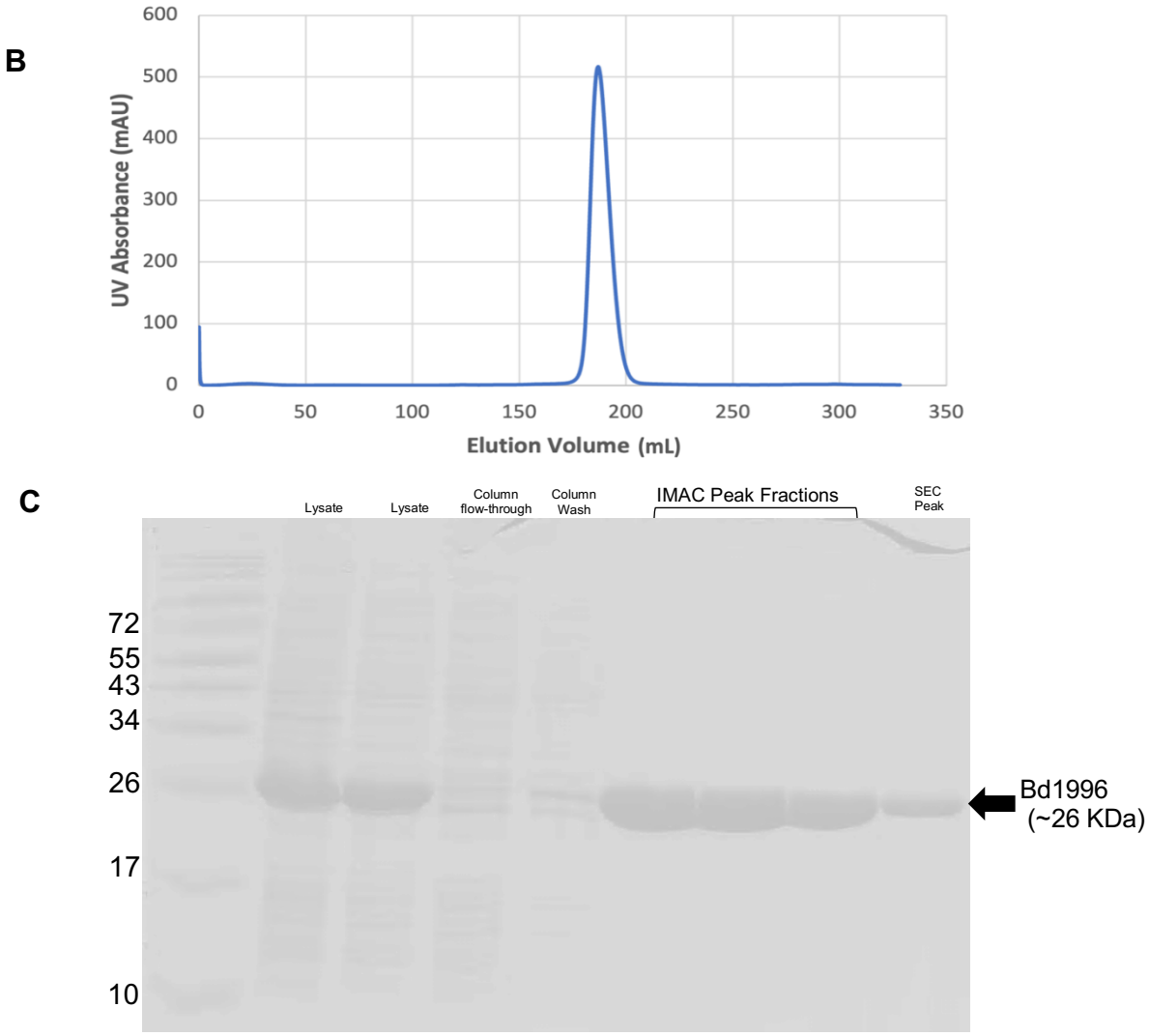


Figure 4.7: Purification of Bd1996
A – Chromatogram of Bd1996 IMAC purification. IMAC performed on a 5 mL HisTrap HP using a gradient elution of 0-100% buffer B across 30 CV. Bd1996 elutes in a single population at ~20% concentration buffer B (100 mM imidazole). Curves for UV absorbance at 280 nM (blue trace) and buffer B concentration (orange trace) are shown.
B – SEC Chromatogram for Bd1996. Bd1996 was eluted as a monodisperse peak from a HiLoad 26/600 Superdex 75 prep grade column at 188 mL, an elution volume consistent with the size of monomeric GltE1 (~26 KDa). Curve for UV absorbance at 280 nM shown in blue.
C – SDS-PAGE analysis of IMAC and SEC peak fractions. Bd1996 can be seen as a predominant band in line with the 26 kDa marker. The gel reveals that Bd1996 is extremely pure following both IMAC and SEC. Black arrow indicates the position of Bd1996.

In the Bd1996-PM model, the possible dimerisation of Bd1996 appears to occur through the PilZ domain. As no crystals were grown for the full-length Bd1996 construct, the PilZ domain was cloned alone (Bd1996-PilZ) in an attempt to improve the likelihood of crystallisation and provide a tool to study Bd1996-PilZ dimerisation. Bd1996-PilZ expressed well and was initially purified using IMAC, where the protein eluted as a single peak at ~50% concentration buffer B (250 mM imidazole) (Figure 4.8). SDS-PAGE analysis confirmed the presence of Bd1996-PilZ and the peak fractions were pooled and further purified by SEC (Figure 4.9). Bd1996-PilZ was eluted as a monodisperse peak from HiLoad 26/600 Superdex 75 prep grade column at ~220 mL, an elution volume roughly consistent with the size of monomeric Bd1996-PilZ (~12 KDa). The SEC elution profiles of Bd1996 and Bd1996-PilZ suggest that dimerisation of the PilZ domain does not occur in solution. SDS-PAGE demonstrated the high purity of the post-SEC Bd1996-PilZ sample (Figure 4.9).

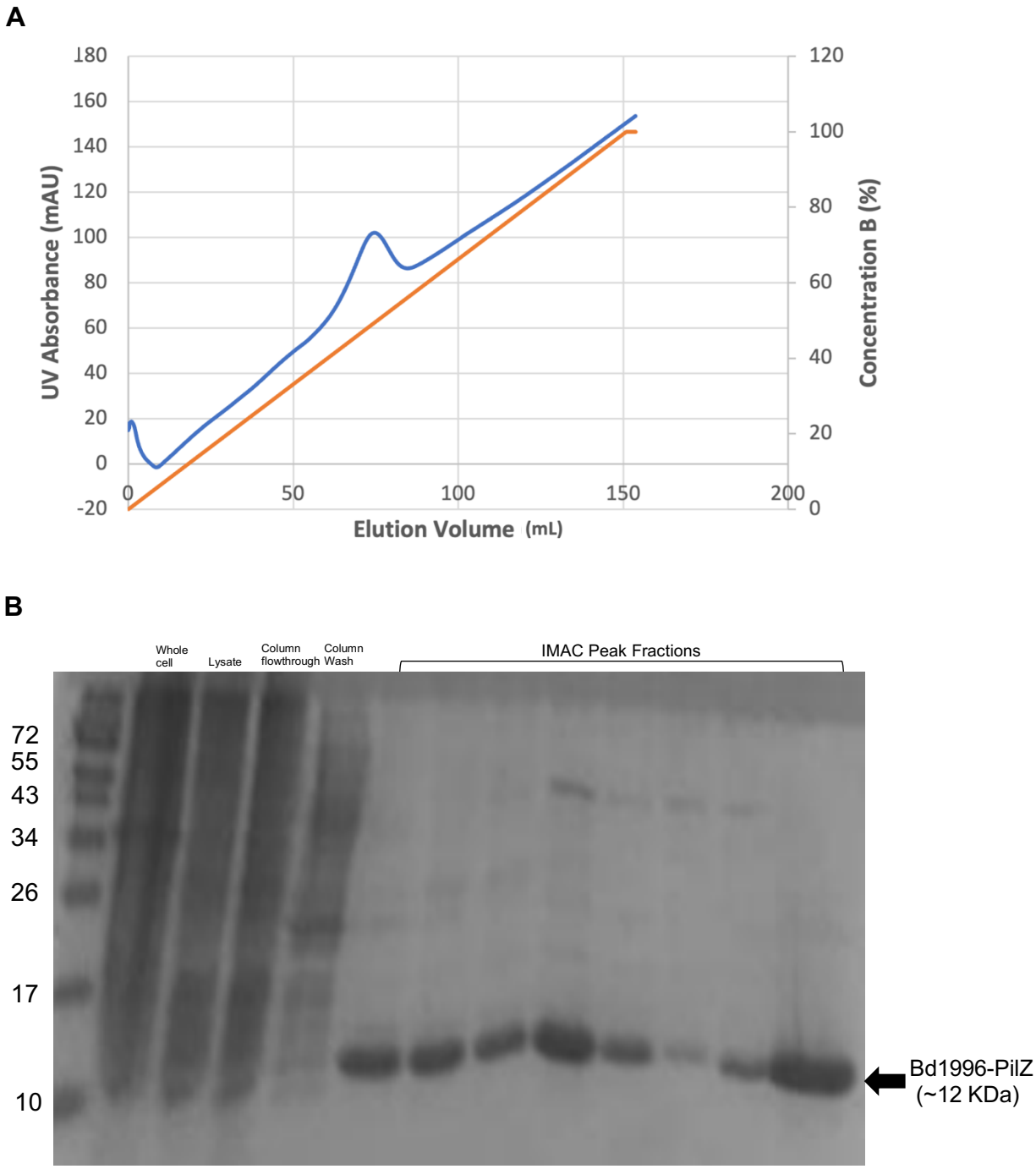


Figure 4.8: IMAC Purification of Bd1996-PilZ
A – Chromatogram of Bd1996-PilZ IMAC purification. IMAC performed on a 5 mL HisTrap HP using a gradient elution of 0-100% buffer B across 30 CV. Bd1996-PilZ elutes in a single population at ~50% concentration buffer B (250 mM imidazole). Curves for UV absorbance at 280 nM (blue trace) and buffer B concentration (orange trace) are shown.
B – SDS-PAGE analysis of IMAC purification. Bd1996-PilZ can be seen as a predominant band between the 10 and 17 KDa MW markers. The gel also reveals the presence of some higher MW contaminants. The black arrow indicates the position of Bd1996-PilZ.

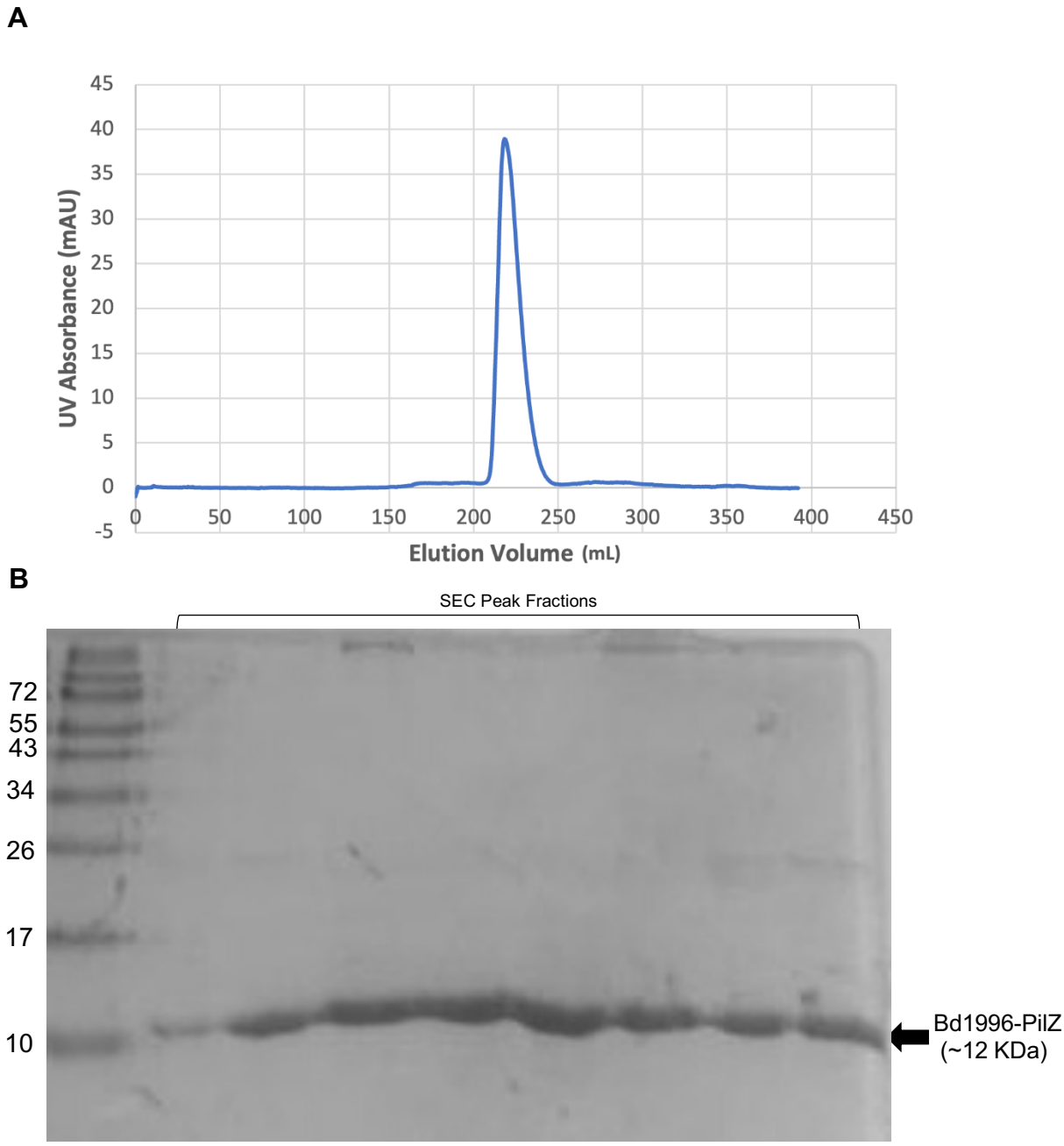


Figure 4.9: SEC Purification of Bd1996-PilZ
A – SEC Chromatogram for Bd1996-PilZ. Bd1996-PilZ was eluted as a monodisperse peak from a HiLoad 26/600 Superdex 75 prep grade column at 220 mL, an elution volume consistent with the size of monomeric Bd1996-PilZ (~12 KDa). Curve for UV absorbance at 280 nM shown in blue.
B – SDS-PAGE analysis of SEC peak fractions. Bd1996-PilZ can be seen as a predominant band between the 10 and 17 KDa markers. The gel reveals that SEC has removed the majority of contaminants. The black arrow indicates the position of Bd1996-PilZ.

4.2.2 Crystallisation and structure solution of Bd1996-PilZ

Bd1996-PilZ readily formed rod-shaped crystals during sparse-matrix crystallisation trials (Figure 4.10). Data collection for Bd1996-PilZ crystals was performed at Diamond Light Source beamline IO3. Bd1996-PilZ crystals belong to the $P3_221$ spacegroup and diffracted to a resolution of 2.8Å. Matthews coefficient analysis indicated that one copy of Bd1996-PilZ comprises the ASU. Initial phases were estimated by MR (PhaserMR)[163] using the PilZ domain from the Bd1996-PM structure as an initial search model. One copy of Bd1996-PilZ was placed in the ASU with a final LLG of 115.1 and TFZ of 10.7, indicating confident placement of the search model. The initial electron density map was of good quality and following successive rounds of model building and refinement, residues 131–221 of Bd1996 were built.



Figure 4.10: Bd1996-PilZ crystals
Representative rod-shaped crystal of Bd1996-PilZ grown in Morpheus G2 (0.2 M Ammoniium formate, 25% PPGBA 400).

Table 4.1: Data collection and refinement statistics for Bd1996-PilZ

Collection Statistics	
Wavelength	0.8989 Å
Resolution range	41.82 – 2.801 Å (2.901 – 2.801 Å)
Space group	P 32 2 1
Unit cell	83.64 Å 83.64 Å 46.97 Å 90° 90° 120°
Total reflections	93705 (9714)
Unique reflections	4858 (328)
Multiplicity	19.3 (20.2)
Completeness (%)	95.66 (68.33)
Mean I/sigma(I)	23.46 (1.47)
Wilson B-factor	58.26 Å^2
R-merge[a]	0.07661 (0.9093)
R-meas[b]	0.07873 (0.9329)
R-pim[c]	0.01801 (0.2073)
$CC^1/_2$	1 (0.964)
Refinement Statistics	
Reflections used in refinement	4649
Reflections used for R-free	188
R-work[d]	0.2093
R-free[e]	0.2551
Number of non-hydrogen atoms	689
macromolecules	686
Ligands	0
Solvent	11
Protein residues	91
RMS(bonds)	0.010
RMS(angles)	1.26
Ramachandran favoured (%)	94.38
Ramachandran allowed (%)	5.62
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	1.30
Clashscore	8.77
Average B-factor	58.26 Å^2
Macromolecules	58.33 Å^2
Solvent	40.84 Å^2

Statistics for the highest-resolution shell are shown in parentheses.

a Rmerge = $\sum hkl \sum i \, |I_i(hkl) - \langle I(hkl)\rangle| / \sum hkl \sum i \, I_i(hkl)$, where $I_i(hkl)$ is the intensity of the ith measurement of a reflection with indexes hkl and $\langle I(hkl)\rangle$ is the statistically weighted average reflection intensity.

b Rmeas is the Rmerge adjusted by a factor of $n(n-1)^{1/2}$ to account for multiplicity.

c R-pim = $\sum hkl [1/(n-1)]^{1/2} \sum i |I_i(hkl) - \langle I(hkl)\rangle| / \sum hkl \sum_i I_i(hkl)$ where $I_i(hkl)$ is the intensity of the ith measurement of a reflection with indexes hkl and $\langle I(hkl)\rangle$ is the statistically weighted average reflection intensity.

d R-work = $\sum ||Fo| - |Fc|| / \sum |Fo|$ where Fo and Fc are the observed and calculated structure factor amplitudes, respectively.

e R-free is the R-work calculated with 5% of the reflections chosen at random and omitted from the refinement.

4.2.3 Bd1996-PilZ Overall Fold

As expected, Bd1996-PilZ is a canonical PilZ domain, consisting of a six-stranded β-barrel capped at the C-terminus by a single α-helix (Figure 4.11). As only a single copy of Bd1996-PilZ occupies the ASU, any dimerisation would have to occur through interaction with a symmetry-related molecule. Analysis of the crystal packing reveals that none of the Bd1996-PilZ copies in the crystal lattice interact via the same interface observed for Bd1996-PM (Figure 4.12).

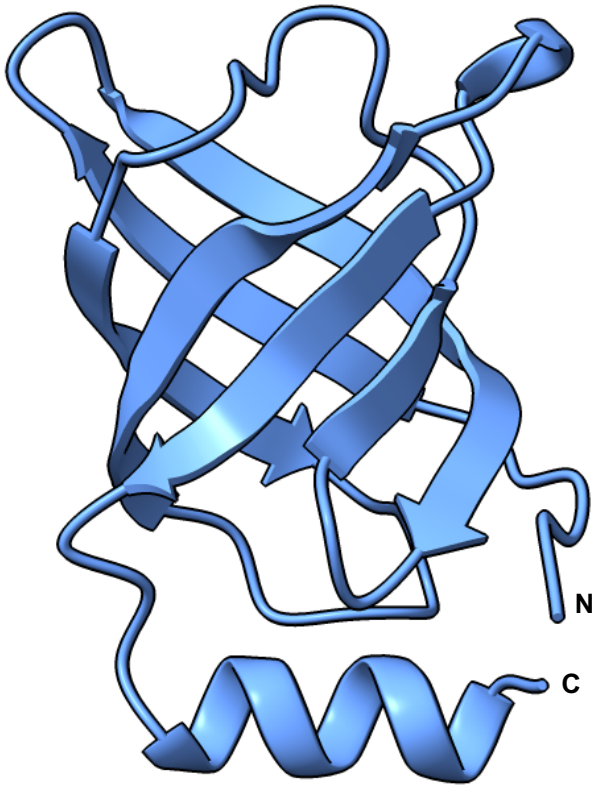


Figure 4.11: Overall fold of Bd1996-PilZ
Crystal structure of Bd1996-PilZ solved to 2.8 Å. Bd1996-PilZ is a canonical PilZ domain consisting of a six-stranded β-barrel and a C-terminal α-helix.

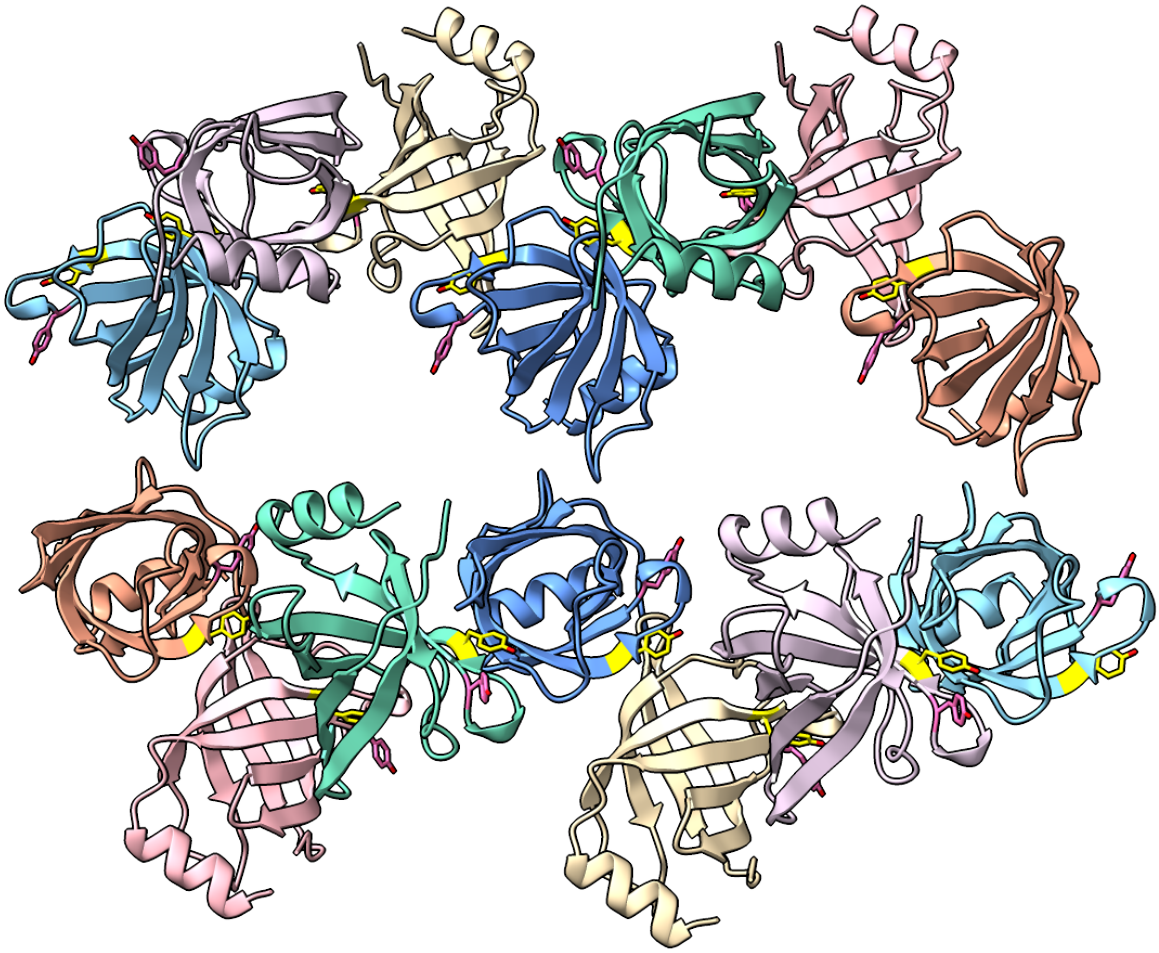


Figure 4.12: Crystal packing of Bd1996-PilZ
Crystal packing between Bd1996-PilZ symmetry neighbours. Tyr195 is coloured yellow and Tyr201 is coloured pink. The crystal packing indicates that none of the Bd1996-PilZ copies interact in the manner observed in the Bd1996-PM.

4.2.4 Further attempts to crystallise full-length Bd1996

Following the successful structural elucidation of Bd1996-PilZ, the remaining crystals were used to facilitate crystal seeding trials for full-length Bd1996. The seed stocks were added to spare matrix screens to maximise the likelihood of crystalising Bd1996. From these initial sparse matrix seeding trials, initial hits were grown in Proplex B9 (0.1 M Sodium cacodylate, 15% PEG 4000) (Figure 4.13). These crystals were clusters of needles that emanated from a larger rod-like crystal. The rod-like sections of these

crystal were isolated but unfortunately did not diffract. The remaining crystals were used as seeds for another round of seeding. This second-round of seeding yielded larger rod-shaped crystals, again in a condition containing 0.1 M Sodium cacodylate, 15 % PEG 4000 (Figure 4.13).



Figure 4.13: Crystals of Bd1996
A – Bd1996 crystals grown from the first round of seeding with Bd1996-PilZ crystals, grown in Proplex B9 (0.1 M Sodium cacodylate, 15 % PEG 4000).
B – Optimised Bd1996 crystals grown from the second-round of seeding grown in Proplex B9 (0.1 M Sodium cacodylate, 15 % PEG 4000).

Table 4.2: Data collection and refinement statistics Bd1996 PilZ for Crystal Form 2

Collection Statistics	
Wavelength	0.9795 Å
Resolution range	29.5 – 2.48 Å (2.569 – 2.48 Å)
Space group	P 65
Unit cell	68.13 Å 68.13 Å 46.55 Å 90° 90° 120°
Total reflections	91602 (8560)
Unique reflections	4452 (439)
Multiplicity	20.6 (19.5)
Completeness (%)	99.69 (99.77)
Mean I/sigma(I)	25.59 (3.92)
Wilson B-factor	52.19 $Å^2$
R-merge[a]	0.0896 (0.6849)
R-meas[b]	0.09185 (0.703)
R-pim[c]	0.02008 (0.1565)
$CC^1/_2$	0.999 (0.952)
Refinement Statistics	
Reflections used in refinement	4466
Reflections used for R-free	234
R-work[d] (Unrefined)	0.2420
R-free[e] (Unrefined)	0.3130
Number of non-hydrogen atoms	792
macromolecules	775
Ligands	0
Solvent	17
Protein residues	102
RMS(bonds) (Unrefined)	0.009
RMS(angles) (Unrefined)	1.46
Ramachandran favoured (%) (Unrefined)	95.00
Ramachandran allowed (%) (Unrefined)	4.00
Ramachandran outliers (%) (Unrefined)	1.00
Rotamer outliers (%) (Unrefined)	2.33
Clashscore (Unrefined)	8.41
Average B-factor (Unrefined)	67.06 $Å^2$
Macromolecules	66.93 $Å^2$
Solvent	72.64 $Å^2$

Statistics for the highest-resolution shell are shown in parentheses.

a Rmerge = $\sum hkl \sum i |I_i(hkl) - \langle I(hkl)\rangle| / \sum hkl \sum i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the ith measurement of a reflection with indexes hkl and $\langle I(hkl)\rangle$ is the statistically weighted average reflection intensity.
b Rmeas is the Rmerge adjusted by a factor of $n(n-1)^{1/2}$ to account for multiplicity.
c R-pim = $\sum hkl [1/(n-1)]^{1/2} \sum i |I_i(hkl) - \langle I(hkl)\rangle| / \sum hkl \sum_i I_i(hkl)$ where $I_i(hkl)$ is the intensity of the ith measurement of a reflection with indexes hkl and $\langle I(hkl)\rangle$ is the statistically weighted average reflection intensity.
d R-work = $\sum ||Fo| - |Fc|| / \sum |Fo|$ where Fo and Fc are the observed and calculated structure factor amplitudes, respectively.
e R-free is the R-work calculated with 5% of the reflections chosen at random and omitted from the refinement.

Data collection for the Bd1996 crystals obtained from the second round of seeding was performed at Diamond Light Source beamline IO4. These Bd1996 crystals belong to $P6_5$ spacegroup and diffracted to 2.48 Å. Whilst attempting to characterise the ASU contents using Matthews coefficient analysis, it became apparent that the unit cell was too small to accommodate full-length Bd1996, suggesting that only a fragment of Bd1996 had been crystallised. Fortuitously, as the fragment crystallised in a condition containing Sodium cacodylate, there was an anomalous signal from the arsenic atoms and initial phases were estimated using SAD (Crank2)[198]. As the Bd1996 fragment was of unknown composition, the number of substructure atoms in the ASU was set to 4 (as Bd1996 contains four cysteine residues). Of the four requested As sites, only two were found with a mean figure of merit (FOM) of 0.59 and the correct hand was selected automatically by Crank2. Initial phase estimation was followed by a round of automated model building using Buccaneer[199] (the full-length sequence of Bd1996 was provided as an input). Buccaneer successfully built 102 residues (R-work = 0.24, R-free = 0.31, FOM = 0.92) and visual inspection of the model revealed that the crystallised fragment was the PilZ domain (Bd1996-PilZ Crystal Form 2). Again, analysis of the crystal packing revealed that none of the PilZ copies in the crystal lattice interact with one and other via the same interaction observed in Bd1996-PM (Figure 4.14). As one structure of Bd1996-PilZ had already been refined, this second structure was not fully refined.

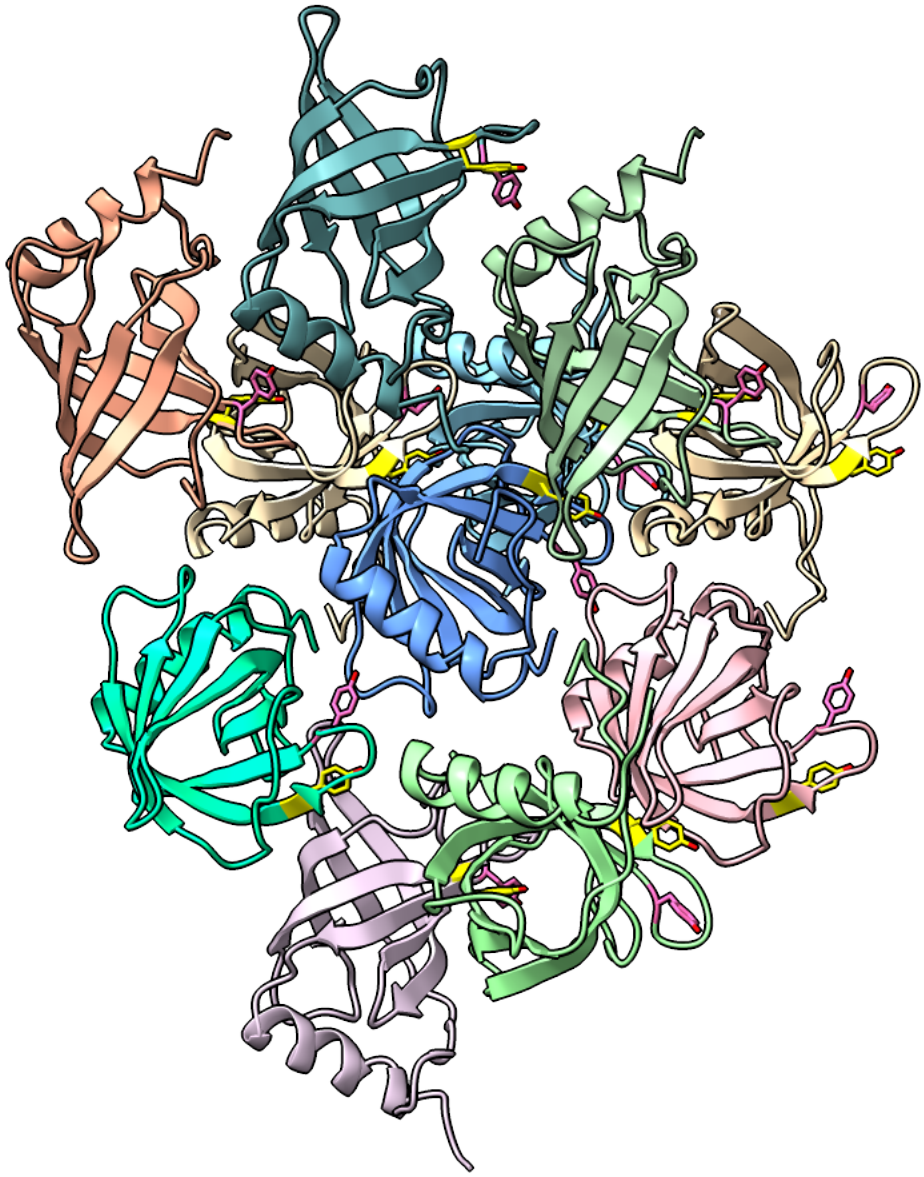


Figure 4.14: Crystal packing of Bd1996-PilZ Crystal Form 2
Crystal packing between Bd1996-PilZ Crystal Form 2 symmetry neighbours. Tyr195 is coloured yellow and Tyr201 is coloured pink. The crystal packing indicates that none of the Bd1996-PilZ copies interact in the manner observed in the Bd1996-PM

4.2.5 AF insights into Bd1996 dimerisation

As attempts to crystalise full-length Bd1996 were unsuccessful, AF[160] was used to generate a model of Bd1996 (Bd1996-AF). In the Bd1996-AF model, both GYF domains and the PilZ domain are confidently (Average pLDDT >70) modelled (Figure 4.15). The GYF-I domain is modelled to contain three β-strands that form an anti-

parallel β-sheet which is enclosed on each side by a short α-helix. Consistent with the sequence analysis, the GYF-I domain contains a region of aromatic residues that form a hydrophobic pocket similar to that seen in the NMR structure of the CD2BP2 GYF domain. In addition to the conserved consensus residues, Trp36 and Trp43 are both predicted to contribute to the hydrophobic pocket, highlighting a degree of structural diversity between the Bd1996-GYF-I domain and the CD2BP2 GYF domain (Figure 4.16). The GYF-II domain is also predicted to possess the same overall fold as the GYF-I domain but consistent with the sequence analysis, appears to represent a more degenerate GYF domain that lacks the majority of the consensus residues (GYF-II domain contains 2 out of 10 consensus residues) and therefore does not possess the characteristic hydrophobic pocket to facilitate binding to PRS (Figure 4.16).

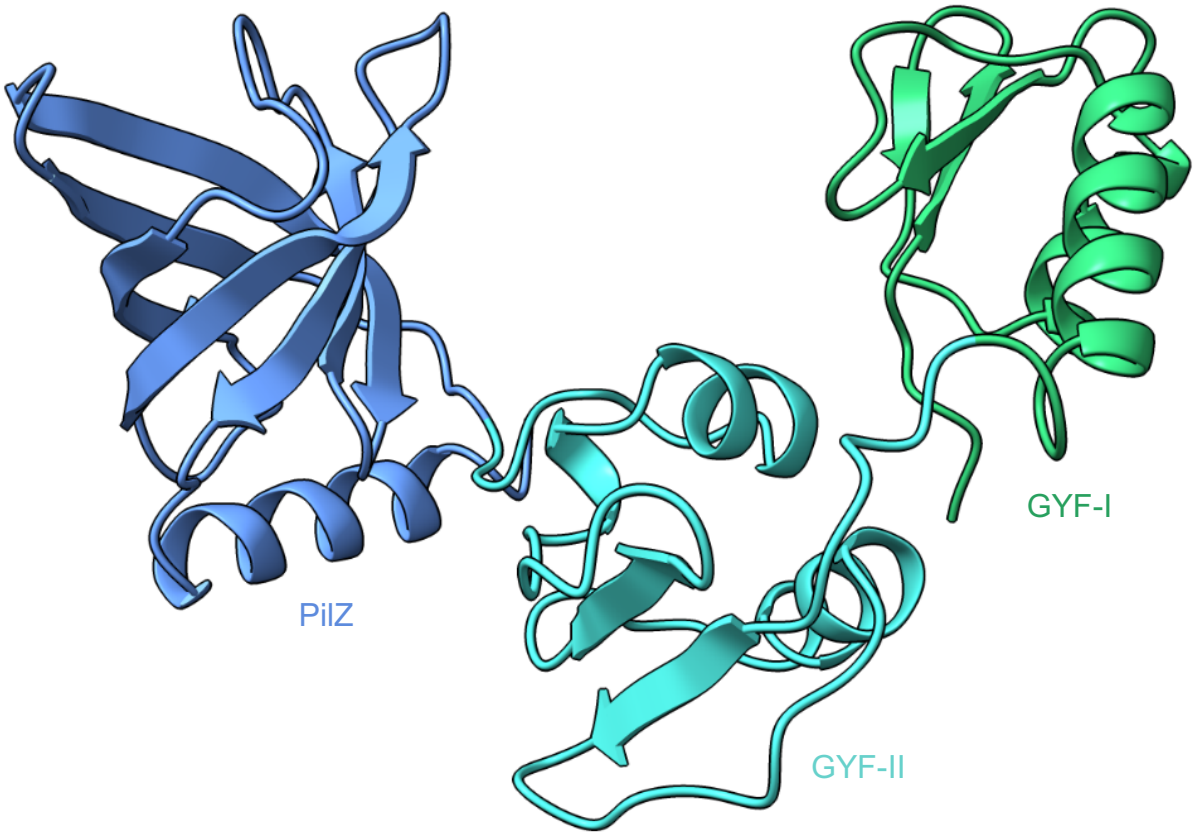


Figure 4.15: AF model of Bd1996
Cartoon representation of the Bd1996 AF model. AF confidently models the GYF-I domain (coloured green), GYF-II domain (coloured cyan) and PilZ domain (coloured blue) of Bd1996. Average pLDDT (confidence) score for Bd1996 = 86.5.

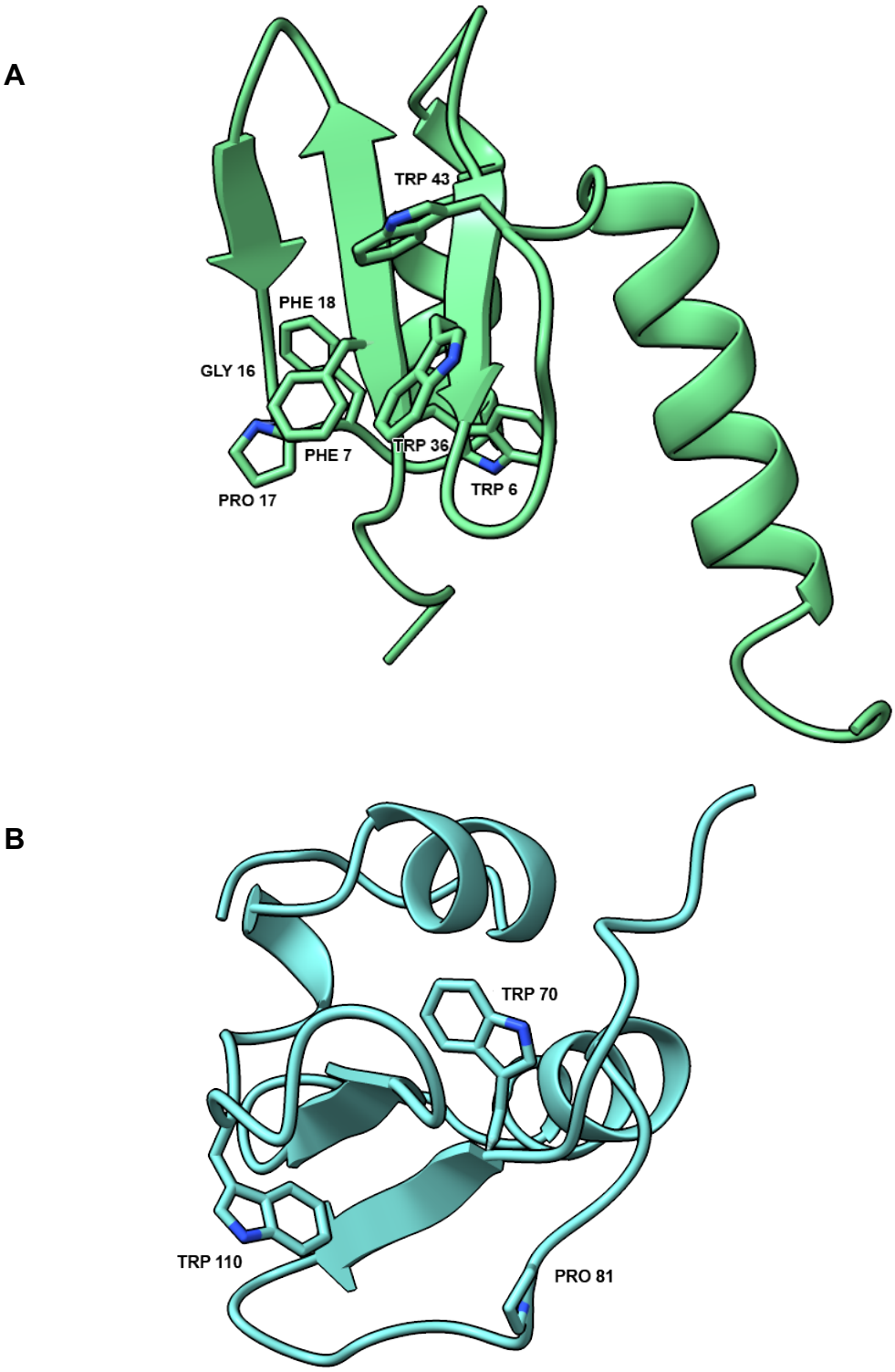


Figure 4.16: AF model of Bd1996 GYF domains
A – Cartoon representation of the AF model of the Bd1996 GYF-I domain. The GYF-I domain contains a characteristic hydrophobic pocket that will likely interact with a PRS on a binding partner. Conserved GYF consensus residues shown as sticks. Average pLDDT (confidence) score for Bd1996 GYF-I domain = 83.92
B – Cartoon representation of the AF model of the Bd1996 GYF-II domain. Consistent with sequence analysis, the GYF-II domain is predicted to be a degenerate GYF domain that lacks a hydrophobic pocket to recognise a PRS. Conserved GYF consensus residues shown as sticks. Average pLDDT (confidence) score for Bd1996 GYF-I domain = 88.62

The lack of any observed dimerisation in two crystal structures of Bd1996-PilZ and during SEC suggests that the interaction observed in Bd1996-PM is an artefact of crystal packing. However, as the GYF domains could only be poorly modelled in Bd1996-PM, we cannot discount a scenario where the interaction between the GYF domains is responsible for bringing the PilZ domains within close proximity to interact. To further probe the likelihood of Bd1996 dimerisation, AF^{160} was used to co-fold two chains of Bd1996 (Bd1996-Dimer) (Figure 4.17). The model of the Bd1996-Dimer predicts an interaction between both copies of Bd1996, however, the interaction occurs via a different mechanism than seen in Bd1996-PM, with Tyr195 and Tyr201 not participating in the interaction at the dimer interface (Figure 4.17). Instead, the predicted interaction occurs between the PilZ domain of one chain and the GYF-I domain of the second chain. In the Bd1996-Dimer model, Pro134 located on β-strand 1 of the PilZ domain packs into a hydrophobic pocket formed by Phe7, Trp36 & Trp43 of the GYF-I domain, in a manner reminiscent to the binding of a GYF domain to its PRS (Figure 4.18). This interaction would block c-di-GMP binding to the PilZ domain and could possibly represent an inhibitory mechanism occurring between two copies of Bd1996.

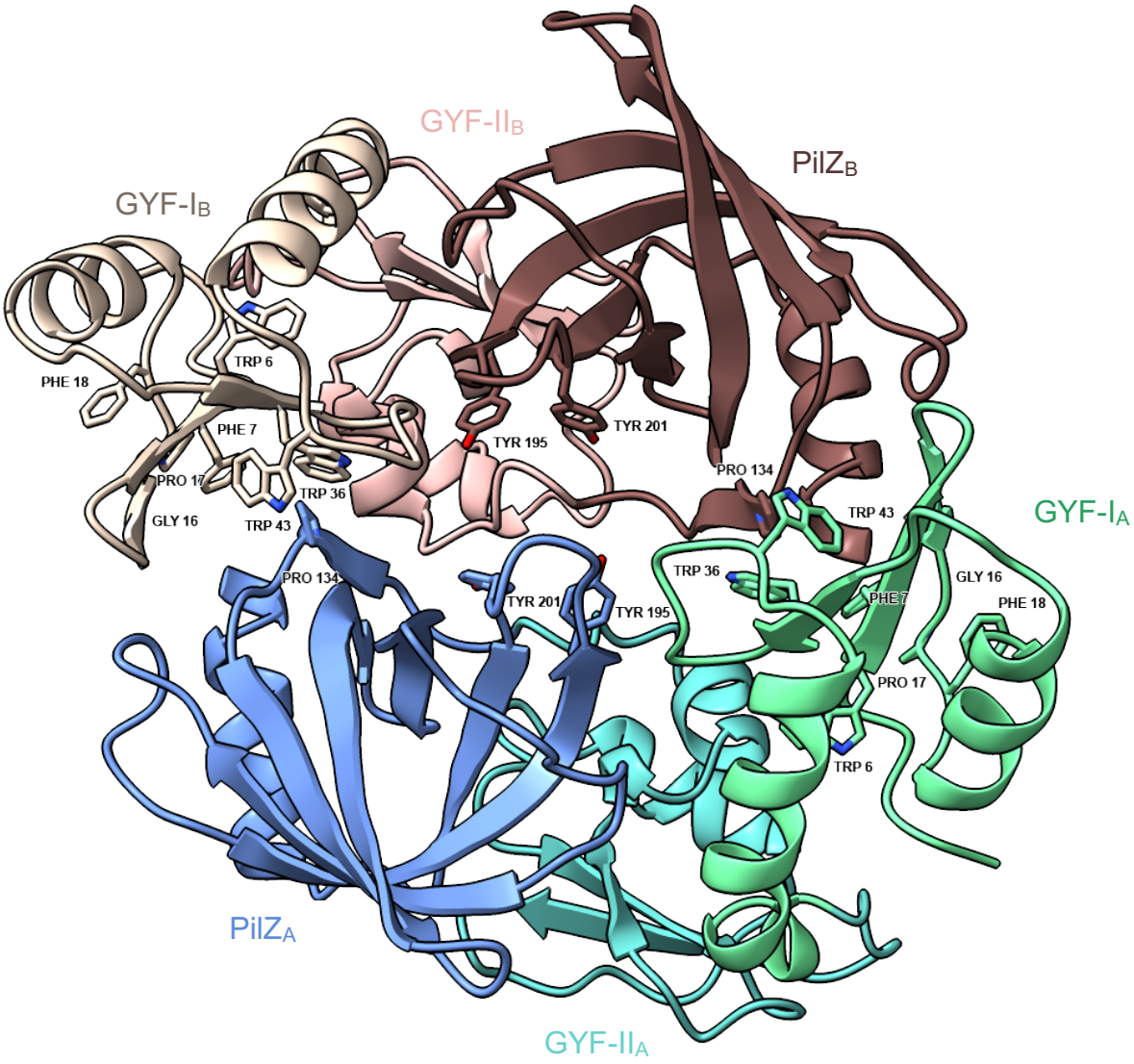


Figure 4.17: AF model of the Bd1996-Dimer
The AF model of the Bd1996-Dimer predicts that the two chains of Bd1996 interact through the PilZ domain of one chain and the GYF-I domain of the second chain. Pro134 of the PilZ domain packs into a hydrophobic pocket formed by Phe7, Trp36 and Trp43 of the GYF-I domain. Chains labelled A and B for clarity. Average pLDDT (confidence) for Bd1996 = 78.23.

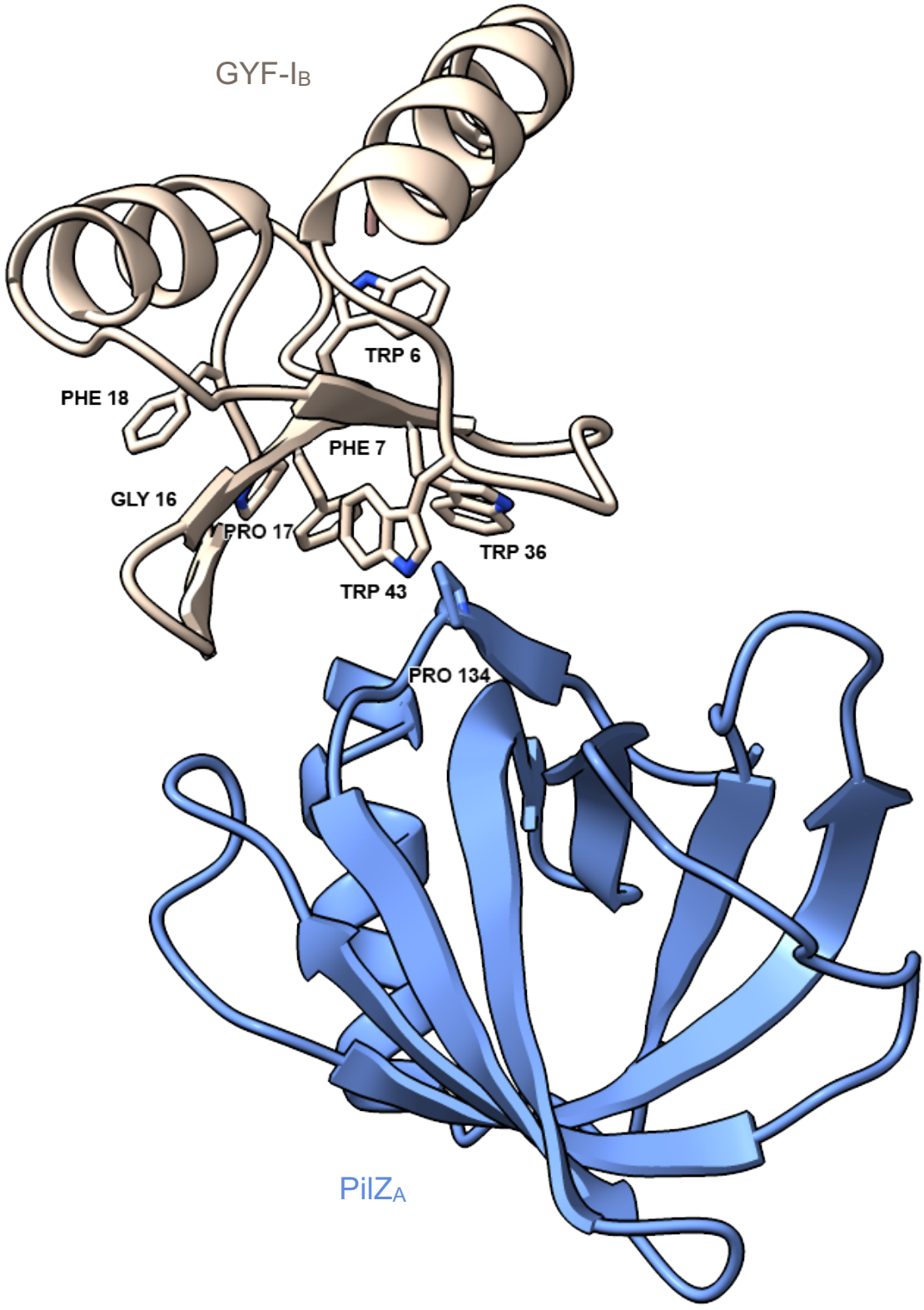
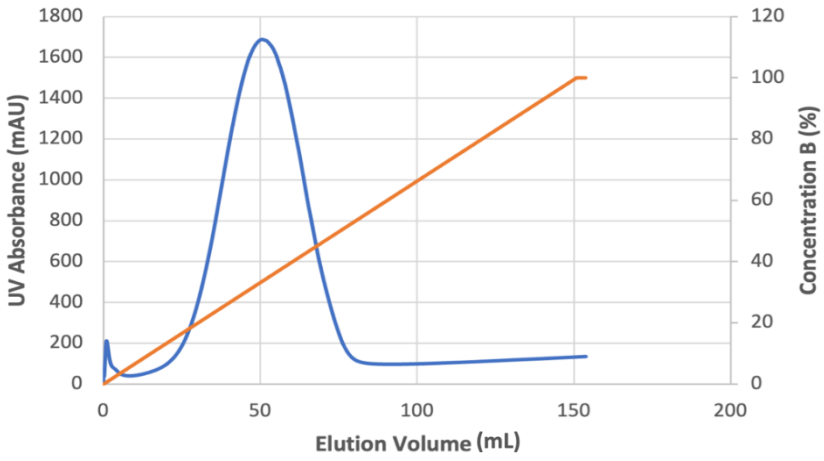


Figure 4.18: Bd1996 PilZ:GYF-I domain interaction
The AF model of the Bd1996-Dimer predicts that the two chains of Bd1996 interact through the PilZ domain of one chain and the GYF-I domain of the second chain. PRO134 of the PilZ domain packs into a hydrophobic pocket formed by Phe7, Trp36 & Trp43 of the GYF-I domain. Chains labelled A and B for clarity. Average pLDDT (confidence) for Bd1996 = 78.23.

In order to further investigate whether the PilZ:GYF interaction observed in the Bd1996-Dimer model was biologically relevant, a construct containing only the GYF domains of Bd1996 was cloned (Bd1996-GYF). Bd1996-GYF expressed well and was initially purified using IMAC, where the protein eluted as a single peak at ~35% concentration buffer B (175 mM imidazole) (Figure 4.19). SDS-PAGE analysis confirmed the presence of Bd1996-GYF and the peak fractions were pooled and further purified using a HiLoad 26/600 Superdex 75 prep grade SEC column. During SEC, a small void peak was observed; though the majority of the sample was eluted as a large peak at ~178 mL, an elution volume roughly consistent with the size of monomeric Bd1996-GYF (~15 KDa) (Figure 4.19). SDS-PAGE demonstrated the high purity of the post-SEC Bd1996-GYF sample (Figure 4.19).

A



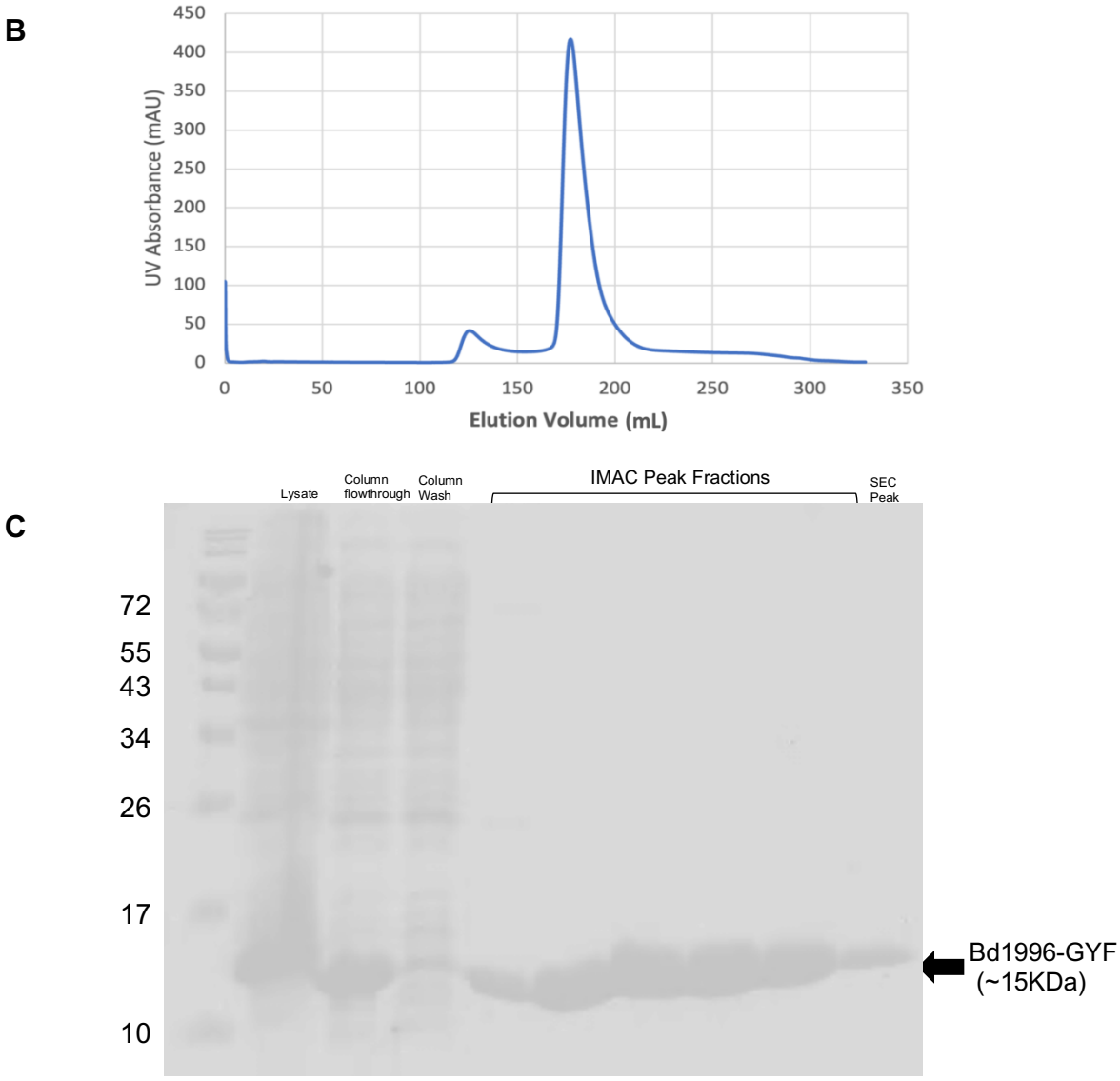


Figure 4.19: Purification of Bd1996-GYF
A – Chromatogram of Bd1996-GYF IMAC purification. IMAC performed on a 5 mL HisTrap HP using a gradient elution of 0-100% buffer B across 30 CV. Bd1996-GYF elutes in a single population at ~35% concentration buffer B (175 mM imidazole). Curves for UV absorbance at 280 nM (blue trace) and buffer B concentration (orange trace) are shown.
B – SEC Chromatogram for Bd1996-GYF. Bd1996-GYF was eluted as a monodisperse peak from a HiLoad 26/600 Superdex 75 prep grade column at 178 mL, an elution volume consistent with the size of monomeric Bd1996-GYF (~15 KDa). Curve for UV absorbance at 280 nM shown in blue.
C – SDS-PAGE analysis of IMAC and SEC peak fractions. Bd1996-GYF can be seen as a predominant band between the 10 and 17 KDa markers. The gel reveals that Bd1996-GYF is extremely pure following both IMAC and SEC. The black arrow indicates the position of Bd1996-GYF.

ITC was used to measure the potential interaction between Bd1996-PilZ and Bd1996-GYF. During the ITC experiments, a small endothermic signal was measured upon the addition of Bd1996-GYF to Bd1996-PilZ and low affinity binding was detected (Figure 4.20). The low affinity of the interaction produces a non-sigmoidal binding thermogram, making the calculations of the KD and stoichiometry inaccurate. Nevertheless, there appears to be a low affinity interaction between Bd1996-GYF to Bd1996-PilZ in solution.

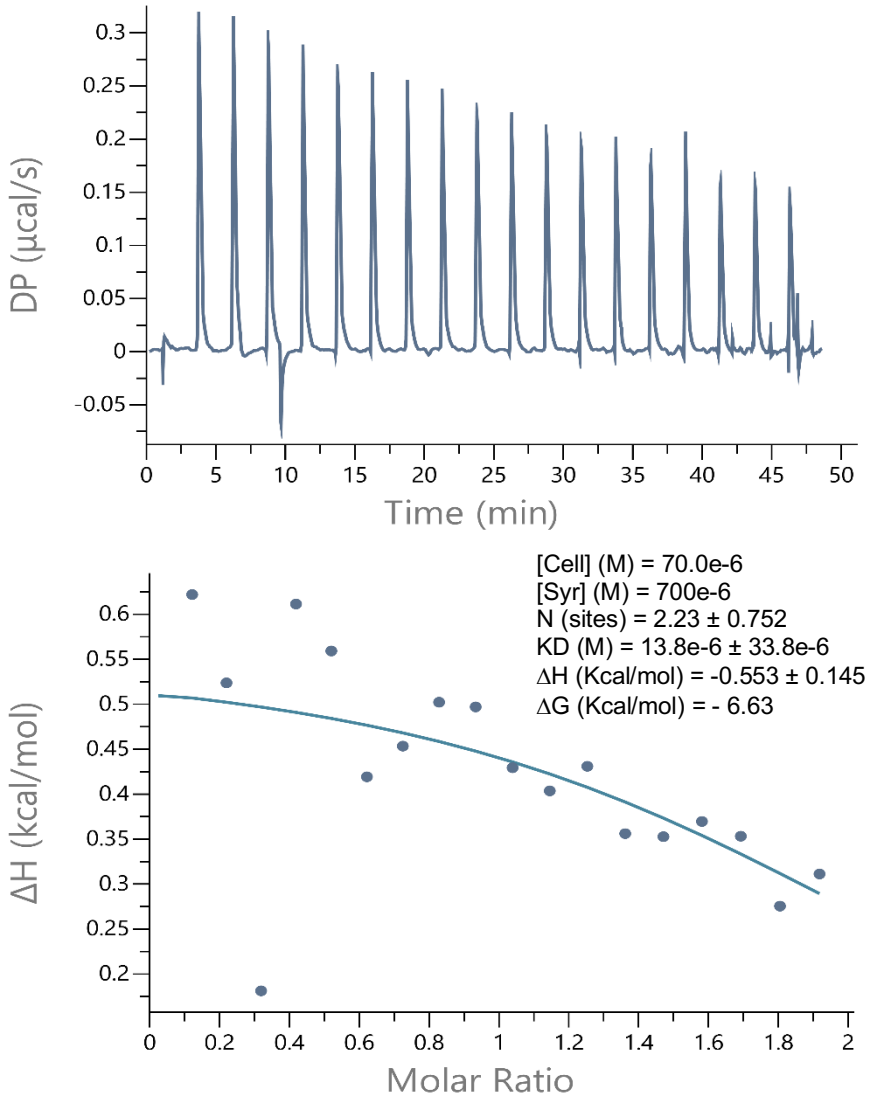


Figure 4.20: ITC binding for Bd1996-PilZ:Bd1996-GYF interaction
Binding thermogram for Bd1996-PilZ:Bd1996-GYF. The two proteins interact with low affinity. As the interaction is not sigmoidal, accurate calculation of KD and stoichiometry is not possible.

4.2.7 Bd1996 pull-down assays

A pull-down assay against *B. bacteriovorus* lysate was performed to identify potential Bd1996 binding partners. To account for the possibility that the binding of c-di-GMP/cGAMP to the PilZ domain could result in conformational change to expose the GYF binding site, the pull-down assay was performed against both full-length Bd1996 and Bd1996-GYF. A BSA control sample was included for comparison, although upon completion of the pull-down assay, SDS-PAGE revealed that the BSA had not bound to the resin and this sample was instead considered as a resin control. The pull-down assays were performed in conjunction with another lab member using the same lysate to identify binding partners against Bd0109, a cryptic *B. bacteriovorus* protein[200]. As these pull-down assays were performed at the same time, Bd0109 was included as another control experiment. SDS-PAGE revealed distinct bands for both full-length Bd1996 and Bd1996-GYF that were not present in either control sample (Figure 4.21).

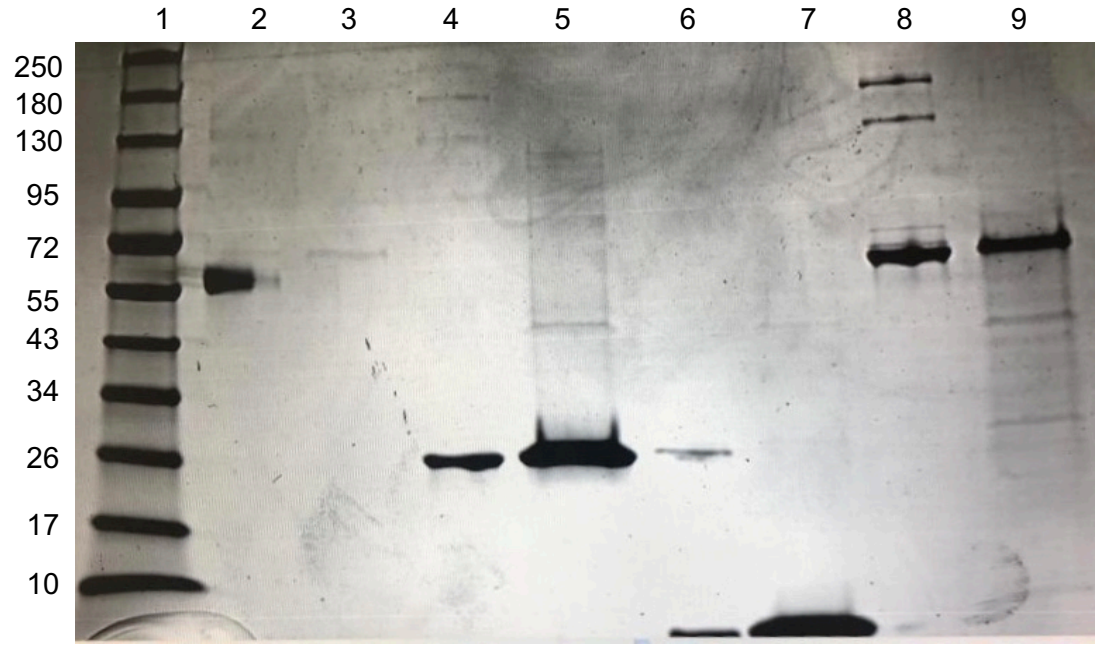


Figure 4.21: Bd1996 Pull-down assays
SDS-PAGE gel of the pull-down assay using Bd1996 and Bd1996-GYF as bait.
Lanes: 1 = ladder, 2 = purified BSA sample, 3 = BSA resin, 4 = purified Bd1996, 5 = Bd1996 resin, 6 = purified Bd1996-GYF, 7 = Bd1996-GYF resin, 8 = purified Bd0109, 9 = Bd0109 resin. Distinct bands can be seen for both full-length Bd1996 and Bd1996-GYF that are not present in either the resin or Bd0109 control samples.

Mass spectrometry (MS) was used to identify the binding partners that were detected through the pull-down assays. The resins used to perform the pull-down assays were sent to the Advanced Mass Spectrometry Facility at the University of Birmingham for on bead digestion followed by subsequent MS analysis. In total, four samples were sent for MS: BSA/resin control, Bd0109 control, Bd1996 and Bd1996-GYF. The majority of the proteins identified through MS were identified as housekeeping genes, ribosomal proteins or chaperones. Other hits were discounted due to the (predicted) subcellular localisation being inconsistent with Bd1996 binding. Finally, many intrinsically disordered proteins were identified which likely pertain to false positives due to the abundance of proline residues in disordered regions.

However, a number of known gliding genes were identified (Tables 4.3 and 4.4). Both GltC and AglR from the Glt1 operon were present in the Bd1996-GYF sample. GltC being located at the periplasmic face of the OM makes it unlikely to be a direct binding partner of Bd1996. However, AglR localises to the IM and is predicted to contain a cytoplasmic region[101]. Direct interaction between a component of the AglRQS motor and Bd1996-GYF would be consistent with Bd1996 playing a role in the regulation of gliding. Another hit, present in both the Bd1996 and Bd1996-GYF samples was DgcA, the Hypr GGDEF enzyme that synthesises c-di-GMP/cGAMP to regulate flagellar motility and gliding motility respectively[141].

Table 4.3: Top 50 interactors identified from pull-down assays using Bd1996 as bait

Proteins that were identified in control samples have been removed. Known gliding proteins are highlighted in yellow. PSM = peptide spectrum match.

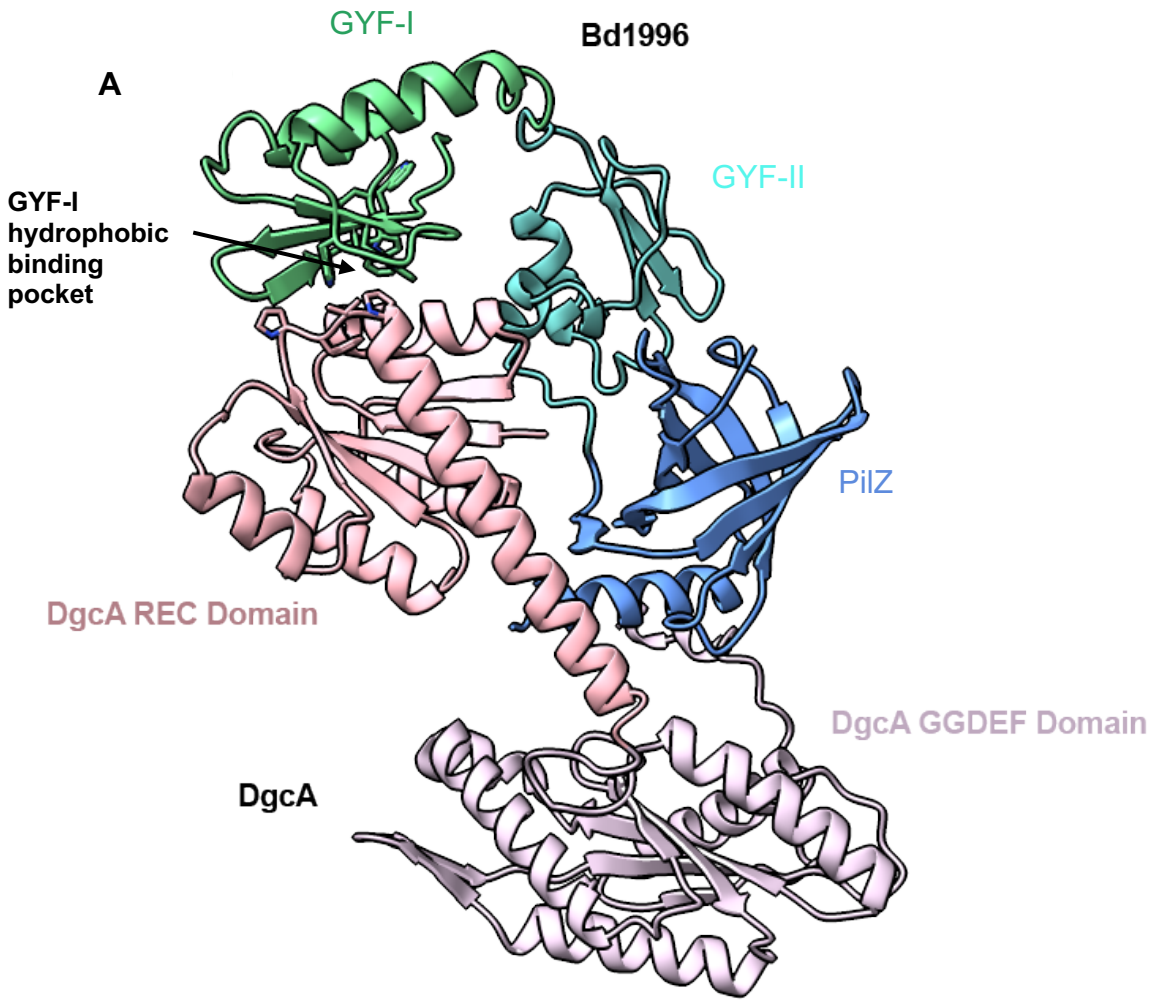
Gene	# PSMs	# Peptides	Uniprot Accession	Gene	# PSMs	# Peptides	Uniprot Accession
Bd3031	23	18	Q6MIW4	malK	4	4	Q6MNM2
Bd0998	13	8	Q6MP66	pfkA	4	4	Q6MI00
Bd1159	10	7	Q6MNS8	Bd1776	4	4	Q6MM67
Bd3101	9	8	Q6MIQ0	rfaD	4	3	Q6MHK5
Bd0233	8	6	Q6MR66	Bd2486	4	3	Q6MKB8
recG	7	5	Q6MKR1	Bd2767	4	3	Q6MJK8
fixP	6	4	Q6MK03	Bd0760	4	2	Q6MPT1
sucC	6	6	Q6MGW1	Bd1238	4	2	Q6MNL0
dppD	6	5	Q6MNM5	appD	4	2	Q6ML33
thrS	6	3	Q6MML0	pepA	3	3	Q6MK58
rpsL	6	2	Q6MJ11	carB	3	3	Q6MPJ5
fabG	5	5	Q6MLJ7	nuoC	3	3	Q6MIR5
mcpA	5	4	Q6MJE7	Bd1933	3	3	Q6MLS3
pilN	5	4	Q6MPI9	Bd2796	3	3	Q6MJI1
pmm	5	4	Q6MRQ8	Bd0737	3	3	Q6MPV3
Bd2402	5	5	Q6MKI8	flil	3	3	Q6MHY3
rhlE	5	5	Q6MR64	Bd2483	3	3	Q6MKC1
Bd2077	5	5	Q6MLD5	Bd3483	3	3	Q6MHQ7
pfkB	5	4	Q6MPV6	choA	3	3	Q6MPV4
putA	5	4	Q6MNK1	pilM	3	3	Q6MPJ0
Bd2913	5	2	Q6MJ70	ddl	3	3	Q6MQ94
hutI	5	2	Q6MJP9	Bd0367 (DgcA)	3	2	Q6MQU2
Bd2723	4	4	Q6MJP7	nrtD	3	2	Q6MPL0
rmlB	4	4	Q6MLY4	Bd1320	3	2	Q6MND8
Bd3385	4	3	Q6MHZ9	folE2	3	2	Q6MKE9

Table 4.4: Top 50 interactors identified from pull-down assays using Bd1996-GYF as bait

Proteins that were identified in control samples have been removed. Known gliding proteins are highlighted in yellow. PSM = peptide spectrum match.

Gene	# PSMs	# Peptides	Uniprot Accession	Gene	# PSMs	# Peptides	Uniprot Accession
Bd3031	18	17	Q6MIW4	Bd2532	3	3	Q6MK78
Bd0998	9	6	Q6MP66	lon2	3	3	Q6MGP8
Rhs	9	9	Q6MQX4	fabG	3	2	Q6MLJ7
fixo	7	4	Q6MK03	mcpA	3	2	Q6MJE7
Bd1159	6	6	Q6MNS8	coaE	3	3	Q6MIK8
tolB	6	6	Q6MRB2	rplQ	3	2	Q6MJ37
Bd2005	6	3	Q6MLK3	Bd1075	3	2	Q6MNZ7
recG	5	5	Q6MKR1	Pkn	2	2	Q6MIK7
Bd0233	5	4	Q6MR66	Bd0220	2	2	Q6MR79
Bd2723	5	3	Q6MJP7	rmlB	2	2	Q6MLY4
Rph	5	3	Q6MJR9	Bd3070	2	2	Q6MIS8
Bd0082	5	4	Q6MRJ7	pepA	2	2	Q6MK58
Bd1591	5	2	Q6MMN5	Bd1697	2	2	Q6MMD8
Pmm	4	4	Q6MRQ8	carB	2	2	Q6MPJ5
Bd3101	4	4	Q6MIQ0	murA	2	2	Q6MRK7
nuoc	4	3	Q6MIR5	rpsQ	2	2	Q6MJ23
PilN	4	3	Q6MPI9	Bd1079	2	2	Q6MNZ3
Bd3385	4	3	Q6MHZ9	feoA	2	2	Q6MLY7
tyrS	4	4	Q6MNQ2	cbaB	2	2	Q6MI75
dnaK	4	4	Q6MNF8	Bd1971	2	2	Q6MLN6
Kup	4	4	Q6MLL0	pal	2	2	Q6MQ51
cit	4	4	Q6MK98	Bd2485	2	2	Q6MKB9
Bd2053	4	4	Q6MLF8	aglR (Bd0836)	2	2	Q6MPL4
purE	4	3	Q6MIY6	Bd0831	2	2	Q6MPL8
Bd1443	3	3	Q6MN20	Bd0367 (DgcA)	2	2	Q6MQU2

Attempts were made to model the potential Bd1996:AglR1 and Bd1996:DgcA interactions with AF[160]. The AF model of Bd1996:AglR1 places both proteins within proximity of each other, though there is no observed interaction between the hydrophobic pocket of the Bd1996-GYF-I domain and a PRS on AglR1. Remarkably, in agreement with the pull-down data, the Bd1996:DgcA model provides insights into a potential interaction between Bd1996 and DgcA with residues 125–128 of the REC domain of DgcA (sequence PFVP) packing into the hydrophobic pocket formed by Phe7, Trp36 and Trp43 of the GYF-I domain (Figure 4.22). The interaction as modelled would block the binding of c-di-GMP/cGAMP to the PilZ domain of Bd1996 and could represent an early interaction in a possible Bd1996:DgcA signalling pathway.



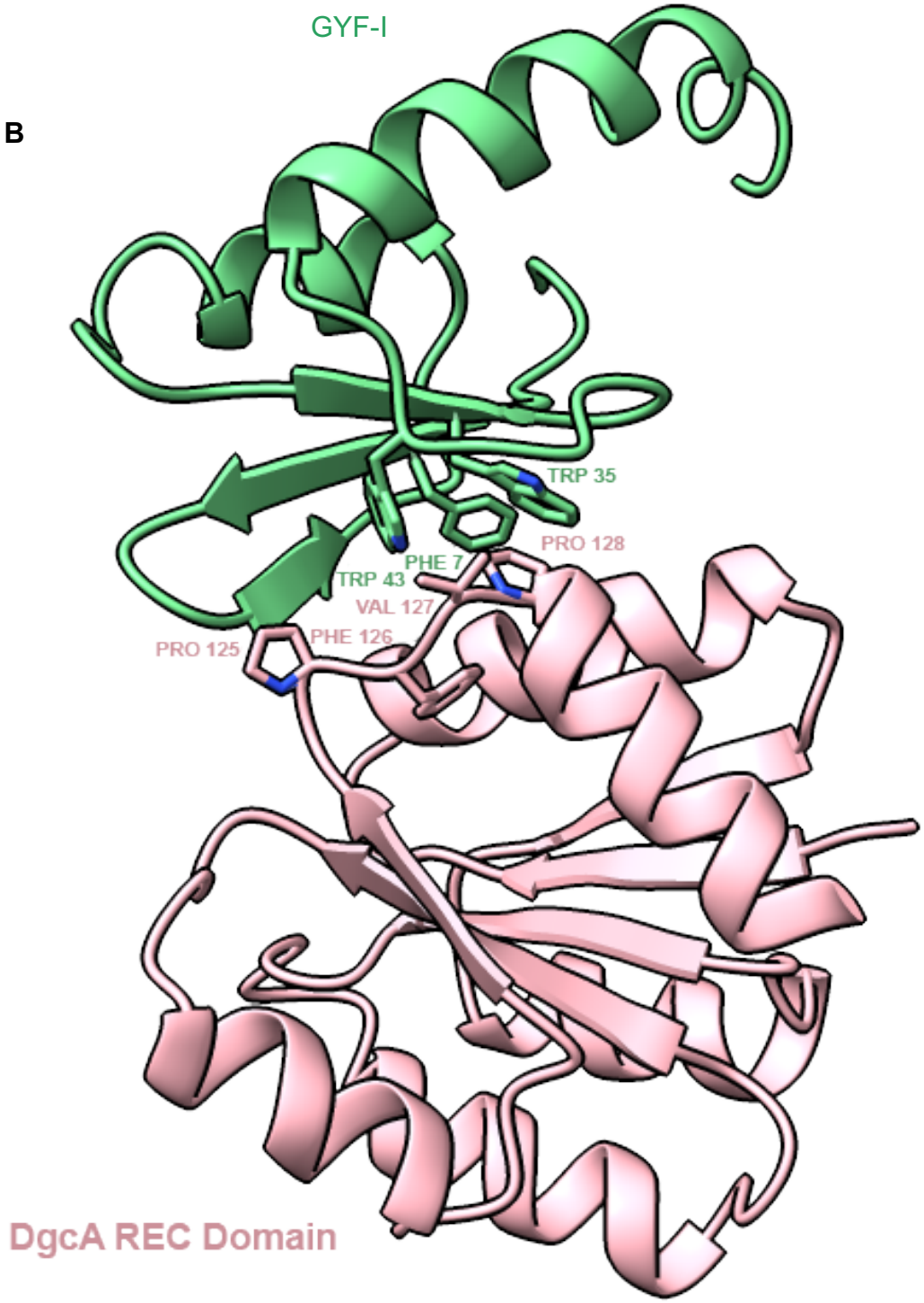


Figure 4.22: Bd1996:DgcA AF model
A – AF model of the interaction between full-length Bd1996 and DgcA. Average pLDDT (confidence) for Bd1996 = 67.3. Average pLDDT (confidence) for DgcA = 75.0.
B – Close-up of the interaction between residues 125–128 (sequence: PFVP) of the REC domain of DgcA with the hydrophobic pocket formed by Phe7, Trp36 and Trp43 of the GYF-I domain of Bd1996. Residues involved in the interaction are shown as sticks.

4.2.8 Experimental Validation of Bd1996:DgcA interaction

In order to experimentally validate the Bd1996:DgcA interaction, a full-length DgcA construct was synthesised; but to date, no DgcA expression has been observed in any of the expression trials performed (data not shown). Although we were unable to produce DgcA, we decided to examine the cyclic dinucleotide substrate range of Bd1996 under the premise that, as DgcA is a Hypr GGDEF enzyme[141], it is plausible that if Bd1996 interacts with DgcA then Bd1996 could bind to both c-di-GMP and cGAMP. ITC was used to examine the interaction between Bd1996 and c-di-GMP/cGAMP. As both cyclic dinucleotides can form oligomers, the cell and syringe concentrations were chosen so that molar ratios of up to 4:1 (c-di-GMP/cGAMP:Bd1996) could be tested.

ITC confirmed that Bd1996 can bind to both c-di-GMP and cGAMP with KD values of 22.6 nM ± 3.58 nM and 1.94 μM ± 404 nM respectively (Figures 4.23 and 4.24). The stoichiometry of the interactions is such that Bd1996 appears to bind dimeric forms of both C-di-GMP and cGAMP. The confirmation that both cyclic dinucleotides bind to Bd1996 suggests that both secondary messengers can potentially regulate distinct processes through Bd1996. To the best of our knowledge, this finding represents the first reported case of a PilZ domain binding to both c-di-GMP and cGAMP.

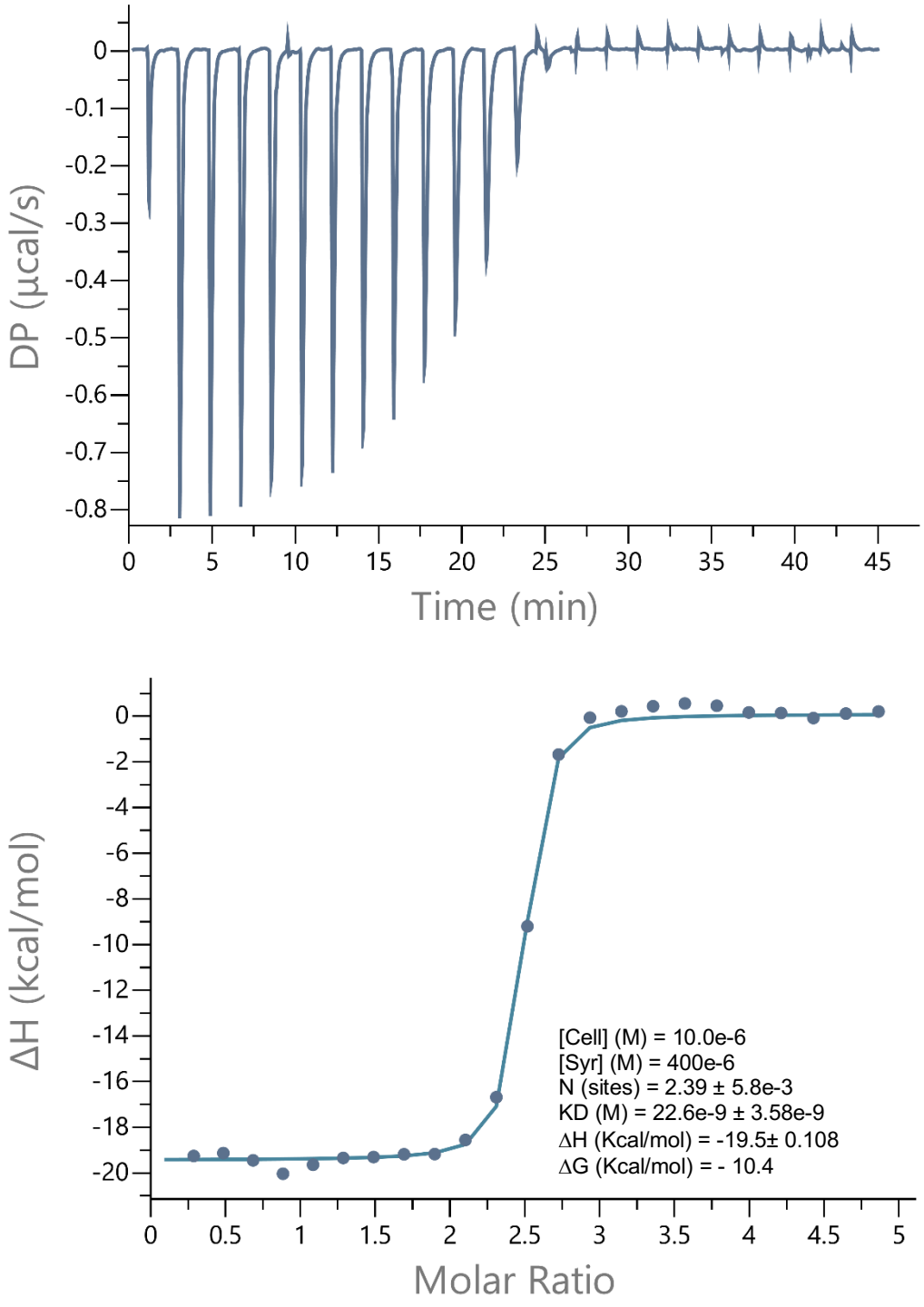


Figure 4.23: ITC binding for Bd1996 c-di-GMP interaction
Binding thermogram for Bd1996 and c-di-GMP. Bd1996 and c-di-GMP interact in a 1:2 ratio with nanomolar affinity.

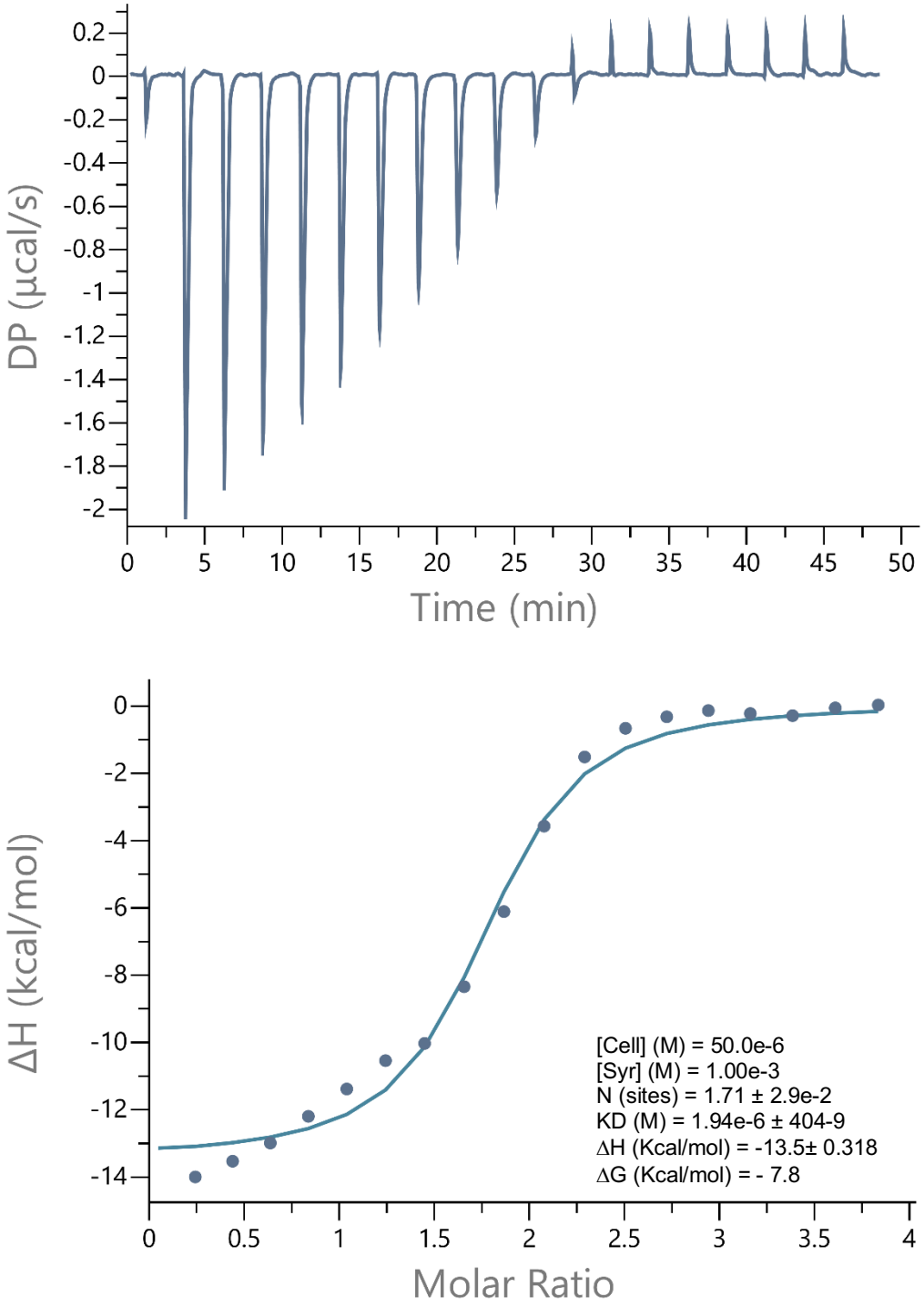


Figure 4.24: Bd1996 cGAMP interaction
Binding thermogram for Bd1996 and cGAMP. Bd1996 and cGAMP interact in a 1:2 ratio with micromolar affinity.

4.2.9 Bd1996 Phage Display

In conjunction with the pull-down assays, phage display was also utilised as an alternative approach to identify Bd1996 binding partners. Whilst the pull-down assays act to identify specific protein interactors, phage display provides the opportunity to independently determine what motifs Bd1996 can recognise[201]. Identified motifs can then be searched in the *B. bacteriovorus* genome to identify potential binding partners. Although phage display represents a less direct method to determine binding partners, the technique is not impacted by transient gene expression levels associated with the distinct stages of the *B. bacteriovorus* lifecycle or protein solubility and may identify motifs in binding partners that were not present in sufficient quantities in the lysate used for pull-down assays.

Phage display was performed with an industrial collaborator who performed the assay on a supplied sample of purified full-length Bd1996. Following three rounds of panning against a linear 15mer library, a prevalent WxRxWWE motif was identified for the Bd1996 sample but not in control experiments (Table 4.5). The identified WxRxWWE differs considerably from the expected proline-rich motif, raising concerns as to the biological relevance of the identified motif. Submission of the WxRxWWE sequence to the MOTIF Search server revealed that the WxRxWWE motif was not present in any protein expressed by *B. bacteriovorus* HD100. Moreover, a literature search identified that the motif DxRxWWQ forms part of a polystyrene binding peptide that is a commonly observed false positive in phage display studies[202]. Taken together, from the lack of biological relevance and the similarity to a known false positive, it was

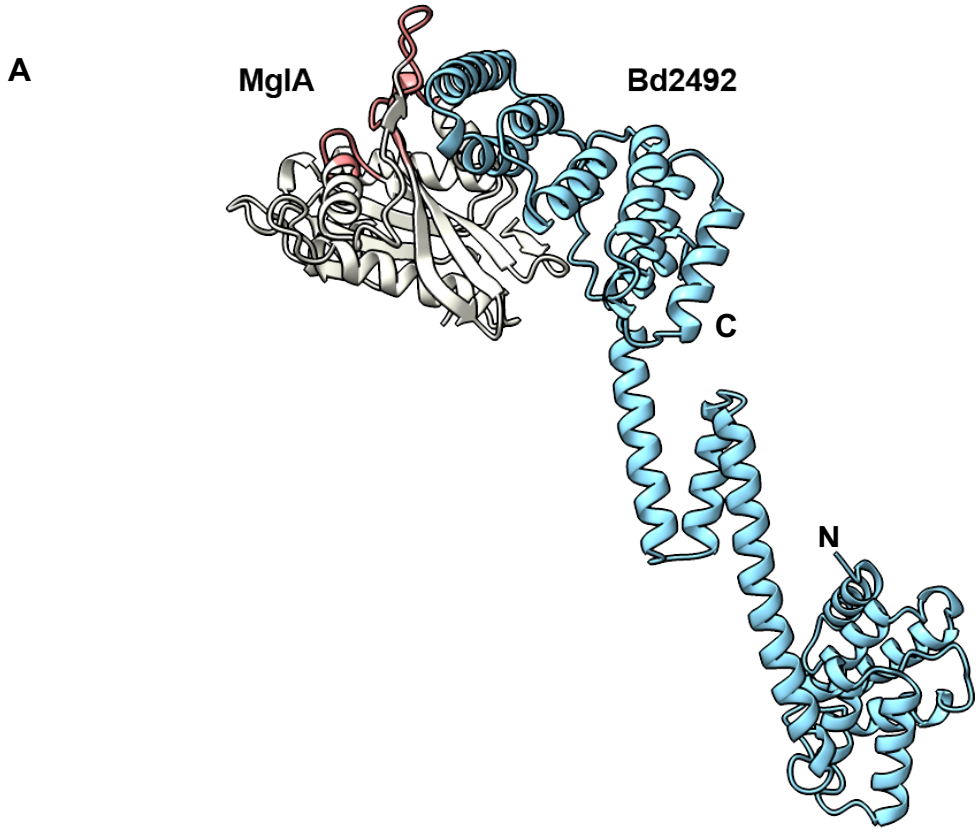
concluded that the identified WxRxWWE motif was also a false positive. There were no other enriched motifs identified in the phage display study.

Table 4.5: Top 10 enriched peptide sequences following three rounds of panning.
All 10 sequences contain the WxRxWWE motif.

sequence
MWERYWWEHPFQQDD
KVEQWVRHWWEYPHW
SGYYWERYWWENTEV
HTFMWQRYWWETAVD
SGYYWERYWWERSV
KVDQWVRHWWEYPHW
KVQQWVRHWWEYPHW
KVEQWVRHWWERSV
KVEQWGRHWWEYPHW
GYQDSWTRYWWEKVD

4.2.10 AF model of $MglA_{Bd}$:Bd2492 complex

As previous attempts to crystallise the $MglA_{Bd}$:Bd2492 complex have been unsuccessful, an AF model was generated to gain insights into the interaction that may inform the design of a new construct which is more amenable for crystallisation. Consistent with the high sequence similarity, the predicted structure of $MglA_{Bd}$ bares strong resemblance to the overall fold of $MglA_{MX.}$ The Bd2492 model comprises 7 TPR repeats (Figure 4.25). The overall arrangement of these TPR repeats is such that three N-terminal TPR repeats are spatially separated from four C-terminal TPR repeats by a helical linker region consisting of three α-helices. The majority of the $MglA_{Bd}$:Bd2492 interaction is predicted to occur between $MglA_{Bd}$ and the four C-terminal TPRs of Bd2492, where $MglA_{Bd}$ packs against the concave face of the C-terminal TPRs. As such, a construct was designed and synthesised that permitted the co-expression of full-length $MglA_{Bd}$ and the four C-terminal TPRs of Bd2492 ($Bd2492_{205–353}$).



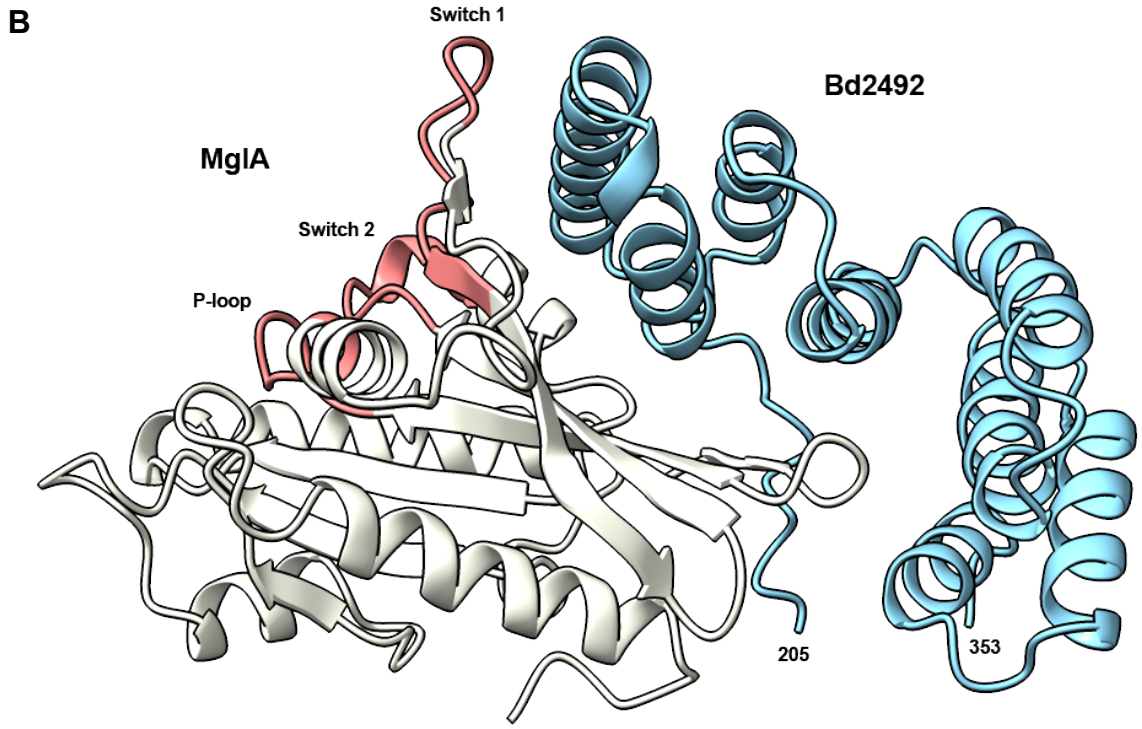


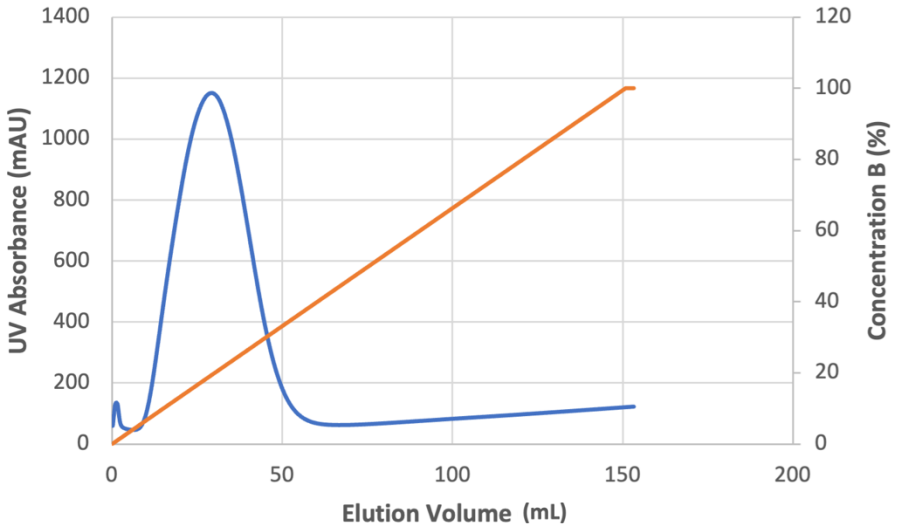
Figure 4.25: AF models of $MglA_{Bd}$:Bd2492 complex
A – Full-length AF model of the $MglA_{Bd}$:Bd2492 complex. The interaction predominately occurs between the four C-terminal TPR motifs of Bd2492 and $MglA_{Bd}$. The switch regions of $MglA_{Bd}$ are coloured pink. Average pLDDT (confidence) score for $MglA_{Bd}$ = 83.3. Average pLDDT (confidence) score for Bd2492 = 82.5.
B – Close-up of the $MglA_{Bd}$:Bd2492 interaction showing the $Bd2492_{205–353}$ region that was cloned.

4.2.11 Purification of $MglA_{Bd}$:$Bd2492_{205–353}$

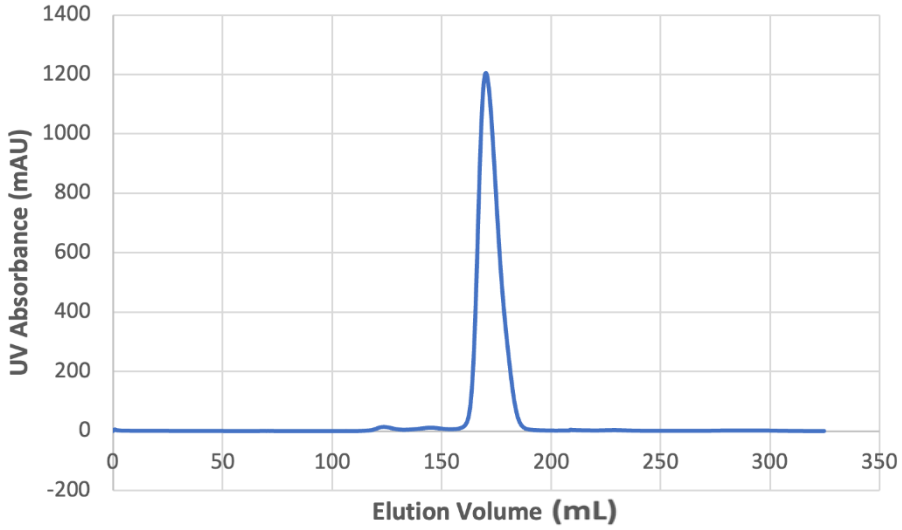
$MglA_{Bd}$ and $Bd2492_{205–353}$ both expressed well from the co-expression construct described in Table 2.8. $MglA_{Bd}$ & $Bd2492_{205–353}$ were initially purified by IMAC where a single peak was observed at ~22% concentration buffer B (110 mM imidazole) (Figure 4.26). SDS-PAGE analysis of the peak fractions confirmed that both $MglA_{Bd}$ & $Bd2492_{205–353}$ were present (Figure 4.26). Only $Bd2492_{205–353}$ contained a His-tag, therefore the presence of both proteins in the IMAC elution fractions confirms that $Bd2492_{205–353}$ maintains its interaction with $MglA_{Bd}$. $MglA_{Bd}$:$Bd2492_{205–353}$-containing fractions were pooled and further purified by SEC. $MglA_{Bd}$:$Bd2492_{205–353}$ was eluted as a monodisperse peak from HiLoad 26/600 Superdex 75 prep grade column at 170 mL (Figure 4.26). Consistent with the AF model, the elution volume of

$MglA_{Bd}$:$Bd2492_{205–353}$ during SEC suggests that $MglA_{Bd}$ and $Bd2492_{205–353}$ interact with a 1:1 stoichiometry ($MglA_{Bd}$: ~23 KDa, $Bd2492_{205–353:}$ ~17 KDa). SDS-PAGE demonstrated the high purity of the post-SEC $MglA_{Bd}$:$Bd2492_{205–353}$ sample (Figure 4.26).

A



B



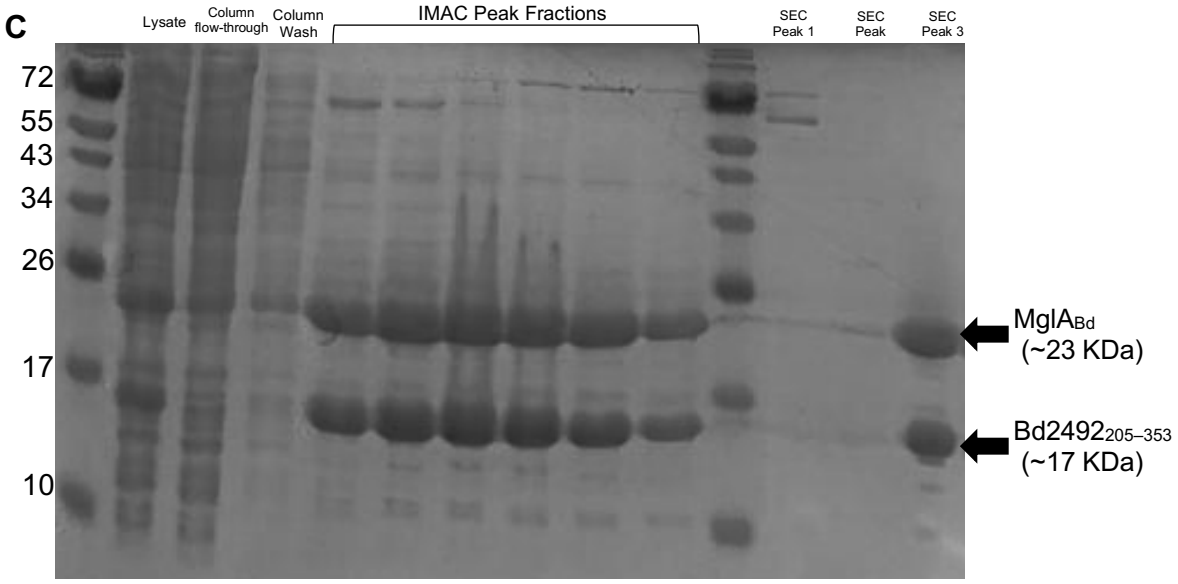


Figure 4.26: Purification of MglA$_{Bd}$:Bd2492$_{205–353}$
A – Chromatogram of MglA$_{Bd}$:Bd2492$_{205–353}$ IMAC purification. IMAC performed on a 5 mL HisTrap HP using a gradient elution of 0-100% buffer B across 30 CV. MglA$_{Bd}$:Bd2492$_{205–353}$ elutes in a single population at ~22% concentration buffer B (110 mM imidazole). Curves for UV absorbance at 280 nM (blue trace) and buffer B concentration (orange trace) are shown.
B – SEC Chromatogram for MglA$_{Bd}$:Bd2492$_{205–353}$. MglA$_{Bd}$:Bd2492$_{205–353}$ was eluted as a monodisperse peak from a HiLoad 26/600 Superdex 75 prep grade column at 170 mL, an elution volume consistent with MglA$_{Bd}$ and Bd2492$_{205-353}$ interacting with a 1:1 stoichiometry. Curve for UV absorbance at 280 nM shown in blue.
C – SDS-PAGE analysis of IMAC and SEC peak fractions. SDS-PAGE confirms that both MglA$_{Bd}$ and Bd2492$_{205–353}$ are present in IMAC peak fractions. MglA$_{Bd}$ can be seen as a predominant band between the 17 and 26 KDa markers while Bd2492$_{205–353}$ can be seen as a predominant band between the 10 and 17 KDa markers. The gel also confirms the presence of a number of contaminates following IMAC. SDS-PAGE also reveals that both MglA$_{Bd}$ and Bd2492$_{205–353}$ are present in SEC peak fractions and reveals that SEC removes the majority of contaminants. The black arrows indicate the positions of MglA$_{Bd}$ and Bd2492$_{205–353}$.

4.2.12 Crystallisation of MglA$_{Bd}$:Bd2492$_{205–353}$

Purified MglA$_{Bd}$:Bd2492$_{205–353}$ was used in sparse matrix crystallisation trials. Following extensive screening, crystals with a diamond morphology were observed in condition H1 of the Memgold 2 screen (50 mM Tris 8.5, 28% v/v PEG 400) (Figure 4.27). Data collection for MglA$_{Bd}$:Bd2492$_{205–353}$ crystals was performed at ESRF beamline ID30A.

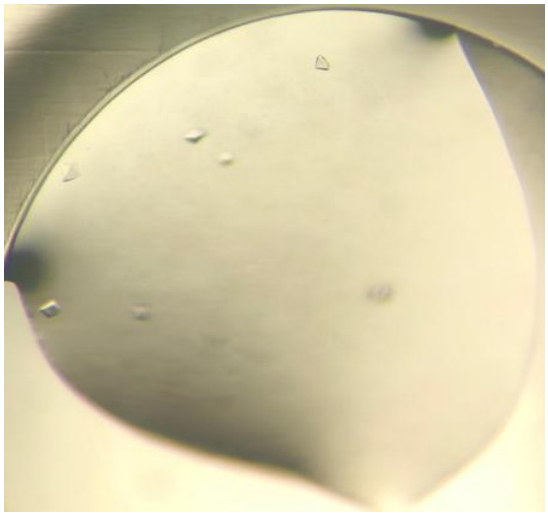


Figure 4.27: $MglA_{Bd}$:$Bd2492_{205–353}$ crystals
Crystals of $MglA_{Bd}$:$Bd2492_{205–353}$ grown in Memgold 2 H1 (50 mM Tris 8.5, 28 % v/v PEG 400).

$MglA_{Bd}$:$Bd2492_{205–353}$ crystals belong to the P 4_3 2_1 2 spacegroup and diffracted to 2.03Å. Matthews coefficient analysis indicated that one copy of the $MglA_{Bd}$:$Bd2492_{205–353}$ complex comprises the ASU. Initial phases were estimated by MR (PhaserMR)[163] using the AF models of $MglA_{Bd}$ and $Bd2492_{205–353}$ as initial search models. During MR, one copy of $MglA_{Bd}$ was placed in the ASU with a final LLG of 610.81 and TFZ of 28.7 indicating confident placement of the $MglA_{Bd}$ search model. However, a copy of $Bd2492_{205–353}$ could not be placed by PhaserMR. Inspection of the MR solution revealed that $MglA_{Bd}$ fits the initial electron density map well. Analysis of $MglA_{Bd}$ packing in the crystal lattice demonstrated that there would be insufficient space in the ASU to accommodate a copy of $Bd2492_{205–353}$, revealing that only $MglA_{Bd}$ had crystallised. Following successive rounds of model building and refinement, residues 2–197 (excluding 49–52, presumed disordered) of $MglA_{Bd}$ were successfully modelled.

Table 4.6: Data collection and refinement statistics for $MglA_{Bd}$

Collection Statistics	
Wavelength	0.8989 Å
Resolution range	61.76 – 2.031 Å (2.104 – 2.031 Å)
Space group	P 43 21 2
Unit cell	87.34 Å 87.34 Å 73.37 Å 90° 90° 90°
Total reflections	254496 (25246)
Unique reflections	18778 (1779)
Multiplicity	13.6 (14.2)
Completeness (%)	99.54 (97.47)
Mean I/sigma(I)	18.59 (0.65)
Wilson B-factor	51.85 $Å^2$
R-merge[a]	0.08889 (3.536)
R-meas[b]	0.09244 (3.668)
R-pim[c]	0.02508 (0.9693)
$CC^1/_2$	0.998 (0.354)
Refinement Statistics	
Reflections used in refinement	18776
Reflections used for R-free	926
R-work[d]	0.2161
R-free[e]	0.2476
Number of non-hydrogen atoms	1638
macromolecules	1557
Ligands	95
Solvent	42
Protein residues	192
RMS(bonds)	0.016
RMS(angles)	1.17
Ramachandran favoured (%)	97.34
Ramachandran allowed (%)	2.66
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	1.19
Clashscore	14.33
Average B-factor	61.45 $Å^2$
Macromolecules	62.41 $Å^2$
Ligands	60.80 $Å^2$
Solvent	54.69 $Å^2$

Statistics for the highest-resolution shell are shown in parentheses.

a Rmerge = $\sum hkl \sum i |I_i(hkl) - \langle I(hkl)\rangle| / \sum hkl \sum i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the ith measurement of a reflection with indexes hkl and $\langle I(hkl)\rangle$ is the statistically weighted average reflection intensity.

b Rmeas is the Rmerge adjusted by a factor of $n(n-1)^{1/2}$ to account for multiplicity.

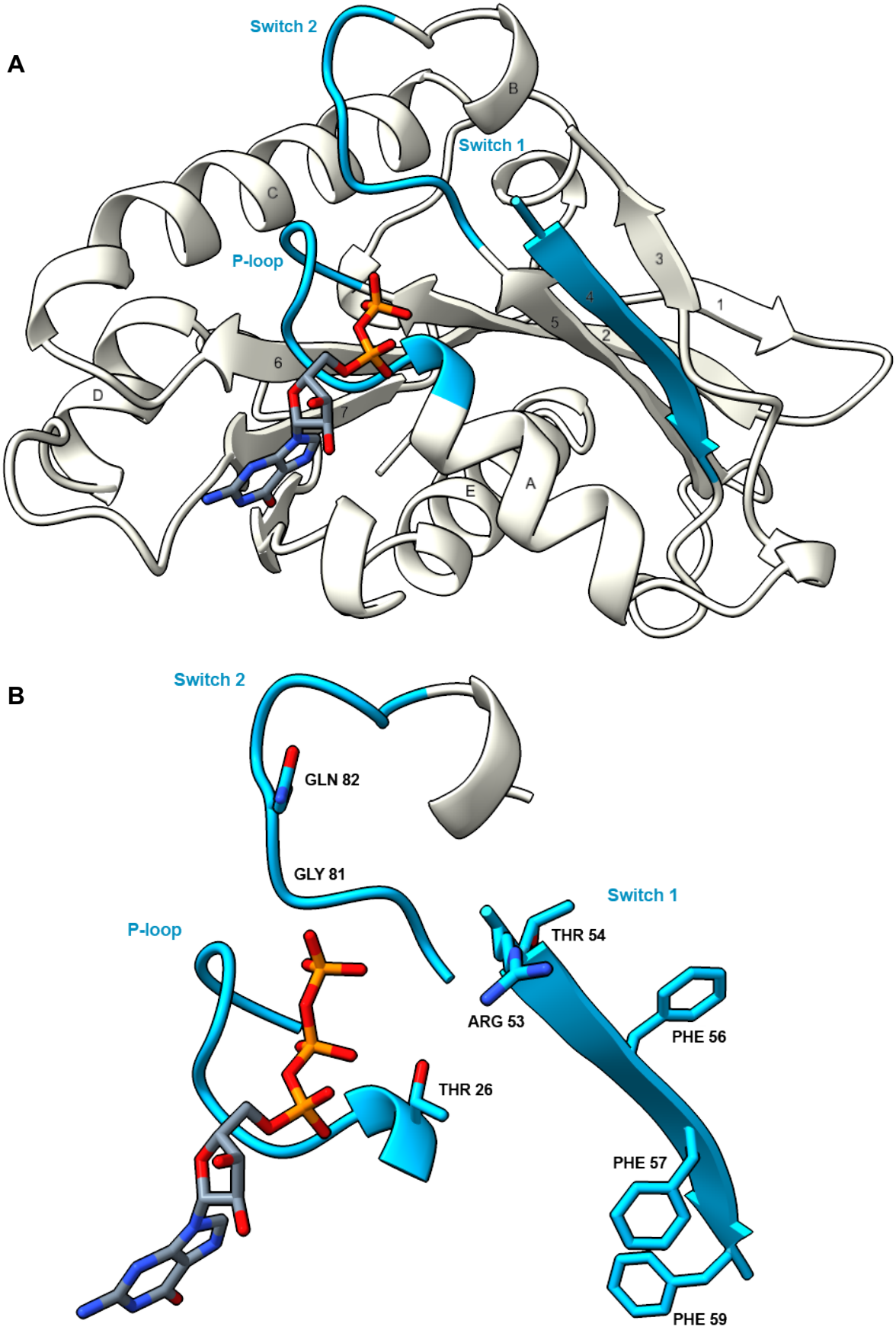
c R-pim = $\sum hkl [1/(n-1)]^{1/2} \sum i |I_i(hkl) - \langle I(hkl)\rangle| / \sum hkl \sum_i I_i(hkl)$ where $I_i(hkl)$ is the intensity of the ith measurement of a reflection with indexes hkl and $\langle I(hkl)\rangle$ is the statistically weighted average reflection intensity.

d R-work = $\sum ||Fo| - |Fc|| / \sum |Fo|$ where Fo and Fc are the observed and calculated structure factor amplitudes, respectively.

e R-free is the R-work calculated with 5% of the reflections chosen at random and omitted from the refinement.

4.2.14 Structural analysis of $MglA_{Bd}$

$MglA_{Bd}$ contains a canonical small GTPase fold that shares strong similarity to $MglA_{MX}$ (RMSD: 1.236 across all atom pairs). Like $MglA_{MX}$, the fold comprises a seven-stranded β-sheet (denoted 1–7) and five α-helices (denoted A–E) that pack around the central β-sheet (Figure 4.28). Surprisingly, $MglA_{Bd}$ appears to be in a mixed active/inactive conformation where the switch 1 region adopts the state observed in the $MglA_{MX}$-GDP structure whilst the switch 2 region occupies the state observed in the $MglA_{MX}$-GTPγS structure (Figure 4.28). The inactive conformation of switch 1 sees the γ-phosphate co-ordinating Arg53 and the Mg^{2+} co-ordinating Thr54 face away from nucleotide-binding site and metal binding site respectively (Figure 4.28). The active conformation of switch 2 is positioned to accommodate binding of the γ-phosphate of GTP. Likewise, Gln82 is positioned to co-ordinate the nucleophilic water molecule, although electron density for the attacking water molecule is not identifiable. Consistent with this mixed GDP/GTP-bound conformation, there is clear electron density for GDP in the nucleotide-binding pocket and also weaker (but significant) density for the γ-phosphate of GTP, suggesting that a sub-population of $MglA_{Bd}$ is bound to GTP (Figure 4.29). As such, both GDP and GTP were modelled in the nucleotide-binding pocket with occupancies of 0.8 and 0.2 respectively. Despite the switch 2 region being in an active conformation, no Mg^{2+} is present in the $MglA_{Bd}$ structure presumably due to occlusion of the Mg^{2+} binding pocket by the inactive conformation of switch 1.


A
Switch 2
B
Switch 1
C
P-loop
3
4
1
5
2
6
D
7
E
A
B
Switch 2
GLN 82
GLY 81
Switch 1
P-loop
THR 54
ARG 53
PHE 56
THR 26
PHE 57
PHE 59

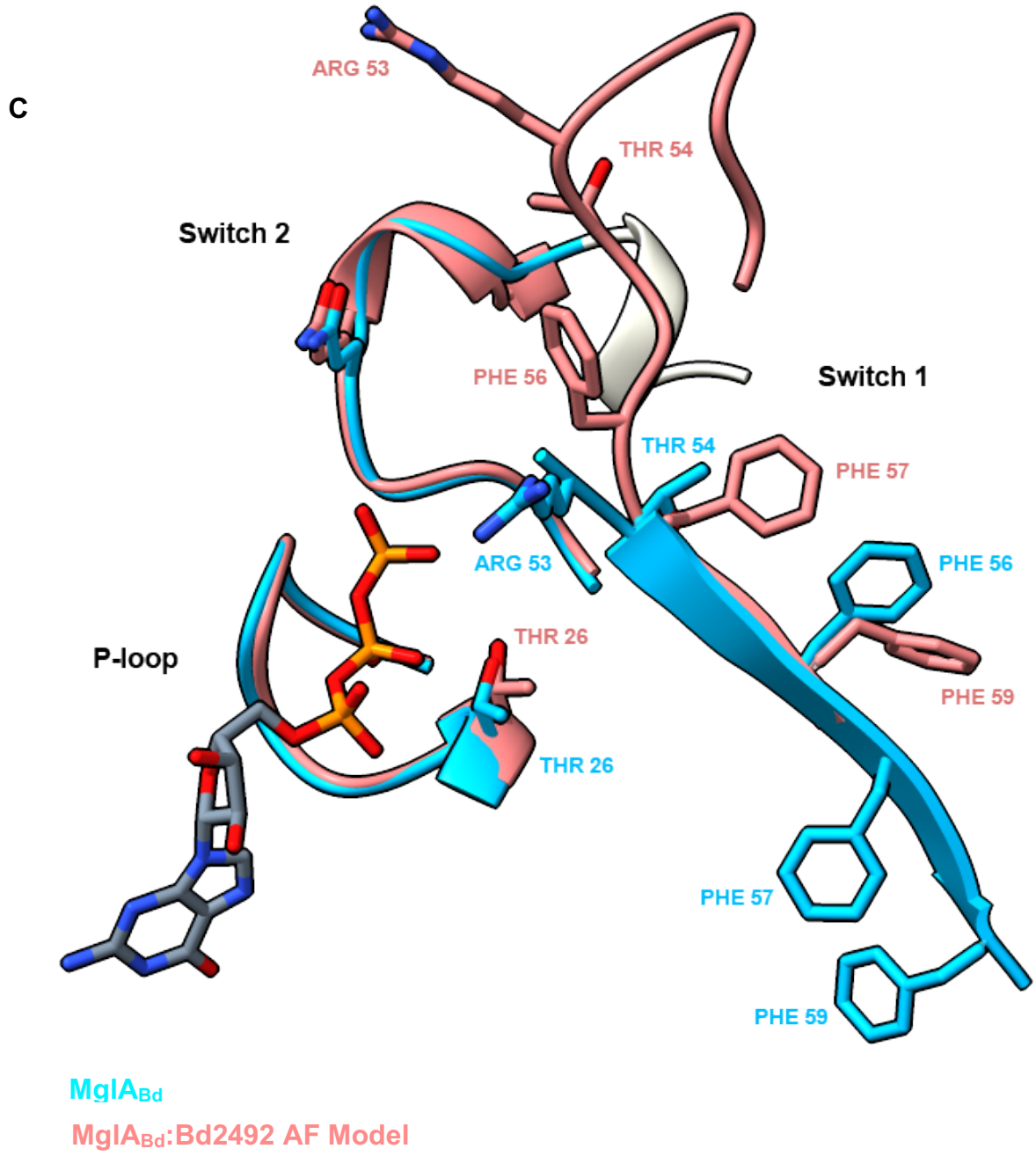


Figure 4.28: Crystal structure of MglA$_{Bd}$
A – Overall fold of MglA$_{Bd}$ bound to GDP. Switch regions and P-loop are coloured blue. β-strands are labelled 1–7, α-helices are labelled A–E.
B – Close-up of the switch regions of MglA$_{Bd}$. Switch 1 occupies the inactive/GDP-bound conformation. Switch 2 occupies the active/GTP-bound conformation. GTP has been modelled to help visualise active conformation of switch 2.
C – Superimposition of the switch 1 and switch 2 regions from the crystal structure of MglA$_{Bd}$ (coloured blue) and the AF model of MglA$_{Bd}$:Bd2492 (coloured red). The switch 1 region of the MglA$_{Bd}$:Bd2492 AF model occupies the active/GTP-bound conformation.

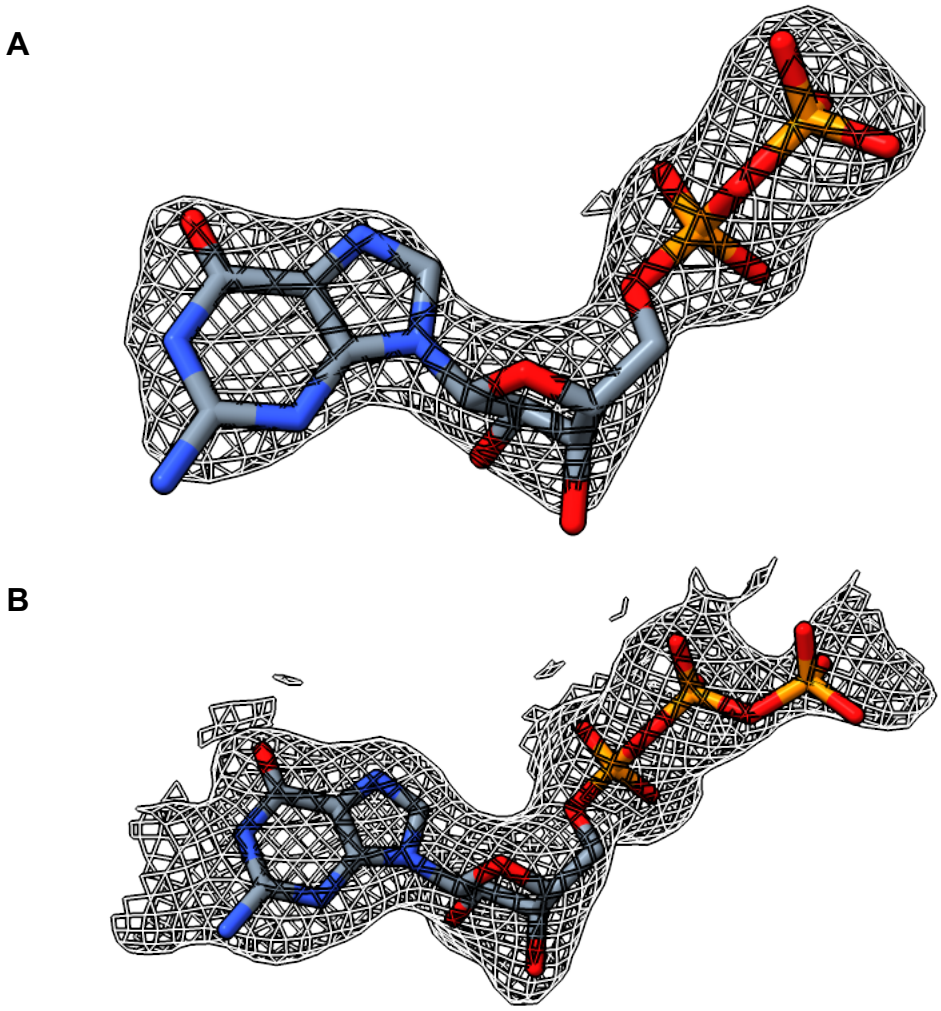


Figure 4.29: Electron density for bound nucleotide
A – Refined 2fo-fc electron density of GDP shown in ChimeraX contoured to 1.76 σ.
B – Refined 2fo-fc electron density of GTP shown in ChimeraX contoured to 1.03 σ.

Remarkably, comparison of the structure of $MglA_{Bd}$ with the AF model of the $MglA_{Bd}$:Bd2492 complex reveals that the switch 1 region of $MglA_{Bd}$ is predicted to be in the active/GTP-bound conformation when in complex with Bd2492, although when modelled alone, $MglA_{Bd}$ occupies the inactive conformation. In the $MglA_{Bd}$:Bd2492 AF model, the switch 1 region of $MglA_{Bd}$ exhibits the three-residue register shift seen in the $MglA_{MX}$-GTPγS-MglB structure (Figure 4.28). This register shift results in the surface presentation of Phe57 and Phe59 which are modelled as contributing to the

hydrophobic packing at the $MglA_{Bd}$:Bd2492 interface (Figure 4.28). Whilst switch 1 is predicted to undergo the same register shift observed in $MglA_{MX}$, the interaction between $MglA_{Bd}$ and Bd2492 as modelled by AF does not see the movement of Arg53 to complete the active site nor does it see the movement of Thr54 to complete Mg^{2+} co-ordination, although it is unclear whether this a result of lower-confidence modelling of this region by AF.

The comparisons between the crystal structure of $MglA_{Bd}$ and the $MglA_{Bd}$:Bd2492 AF model suggests that the interaction between the two proteins is dependent upon $MglA_{Bd}$ being in the GTP-bound conformation. To increase the likelihood of crystallising $MglA_{Bd}$:$Bd2492_{205–353}$, a number of point mutants were designed with the aim of trapping $MglA_{Bd}$ in a GTP-bound conformation. In the $MglA_{MX}$ literature, the Q82L mutation is commonly used to create an inactive MglA construct[80,203]. Two mutants (T26A, T54A) were also designed to prevent Mg^{2+} binding and thus disrupt GTP hydrolysis. In this project, of the three designed constructs, we were only able to clone the $MglA_{BdT54A}$ construct. $MglA_{BdT54A}$:$Bd2492_{205–353}$ was expressed and purified from the co-expression vector and behaved identically to the wild-type complex. $MglA_{BdT54A}$:$Bd2492_{205–353}$ crystallised readily, but upon x-ray diffraction, it was observed that the crystals belonged to the same spacegroup and had the same unit cell dimensions as the wild-type $MglA_{Bd}$:$Bd2492_{205–353}$ complex, suggesting that again only $MglA_{BdT54A}$ had crystallised. MR using the crystal structure of $MglA_{Bd}$ as a search model confirmed that only $MglA_{BdT54A}$ had crystallised and the structure was not refined or pursued further.

4.3 Discussion

4.3.1 ApoBd1996 is monomeric but may exhibit self-inhibition

Prior to this work, previous structural data obtained by our lab suggested that the PilZ domain of Bd1996 dimerises in its apo-form. The nature of this dimerisation would see the occlusion of the c-di-GMP/cGAMP binding site leading to the proposal that, in the presence of c-di-GMP/cGAMP, Bd1996 will monomerise to permit binding which will likely regulate downstream signalling through interactions between the Bd1996 GYF domains and its binding partner(s). However, the work presented in this chapter rules out the possibility of the apoprotein dimerising, with no PilZ dimerisation being observed in two crystal structures or during SEC experiments. AF modelling of a Bd1996 dimer also fails to predict an interaction between the two PilZ domains. It does, however, predict a possible self-inhibitory mechanism in which the PilZ domain of one Bd1996 copy occupies the putative ligand binding pocket of the GYF-I domain of the second copy. Attempts to validate this interaction using ITC suggest that a low affinity interaction exists between the Bd1996-PilZ and Bd1996-GYF constructs used in this work. Further work is required to determine whether the interaction measured by ITC is the same as predicted by AF. This can be done potentially via mutagenesis of the PilZ proline (Pro134), which is 95% conserved (based on sequence alignment of 58 BLAST[204] hits that share >30% sequence identity and 100% coverage to Bd1996).

4.3.2 Pull-down assays identify known gliding proteins as Bd1996 binding partners

Pull-down assays identified AglR1, GltC1 and DgcA as potential Bd1996 binding partners. The subcellular localisation of GltC1 is inconsistent with its potential as a direct binding partner of Bd1996 and may have been identified as a false positive.

AF modelling of Bd1996:DgcA provides a compelling interaction, with a PFVP sequence of DgcA occupying the putative ligand binding pocket of the Bd1996 GYF-I domain. Direct interaction between DgcA and Bd1996 could provide a partial explanation for the specific c-di-GMP signalling pathways observed in *B. bacteriovorus* [139] as direct interaction between GGDEF enzymes and c-di-GMP targets would prevent c-di-GMP spillover. Direct interactions between PilZ domain-containing proteins and GGDEF domain-containing proteins have previously been reported[150,205,206], although none of these interactions occur through GYF domains. Notably, co-folded AF models of DgcA:Bd1482 and DgcA:Bd3100 do not appear to interact.

The Bd1996:DgcA interaction (as modelled by AF) would inhibit the binding of c-di-GMP/cGAMP to Bd1996, with the c-di-GMP/cGAMP binding site of Bd1996 being completely obscured by the REC domain of DgcA. One possible scenario could be that when the GYF domain of Bd1996 binds to DgcA, the synthesis of c-di-GMP/cGAMP is triggered and the presence of the newly synthesised cyclic dinucleotide triggers the release of Bd1996 from DgcA. C-di-GMP/cGAMP-Bd1996 is then free to bind a downstream target to regulate flagellar or gliding motility. Conversely, Bd1996 may act to occlude the REC phosphorylation site of DgcA, allowing for c-di-GMP/cGAMP hydrolases to terminate signalling effectively.

Consistent with a Bd1996:DgcA interaction, we have demonstrated that Bd1996 binds to both c-di-GMP and cGAMP. Whilst Bd1996 appears to have higher affinity for c-di-GMP, the KD values for the binding of both c-di-GMP and cGAMP are within the range of intracellular concentrations (estimated to be in the low micromolar range)[142,207]. Bd1996 has been observed to be the most upregulated PilZ domain-containing protein as *B. bacteriovorus* cells transition to an immobile state when facing starvation[193]. This suggests that Bd1996 could respond to c-di-GMP and cGAMP produced by DgcA to halt both flagellar and gliding motility to completely immobilise cells; although this would be dependent on DgcA being able to simultaneously regulate c-di-GMP and cGAMP signalling.

The mechanisms that stimulate c-di-GMP vs cGAMP production by DgcA remain to be uncovered. In the case of GacA, a Hypr GGDEF enzyme from *Geobacter sulfurreducens* that shares similarity to DgcA (~50% sequence identity), c-di-GMP/cGAMP synthesis is dependent on available levels of ATP/GTP[157,208]. It is therefore possible that a similar mechanism occurs in *B. bacteriovorus* where variations in the ATP/GTP concentrations has some bearing on DgcA-mediated signalling. As previously discussed in Chapter 1, phosphorylation of REC domains triggers the dimerisation of GGDEF domains stimulating diguanylate cyclase activity[209]. However, substitution of the phosphorylation residue of GacA has no impact on c-di-GMP/cGAMP production[157,208] and it has been hypothesised that phosphorylation of this residue regulates which product will be synthesised[208], though this remains to be validated. The REC domain of DgcA contains a consensus phosphorylation residue. Importantly, the mutation of the DgcA phosphorylation

residue (D74E) causes a switch in product production, with $DgcA_{D74E}$ producing predominantly c-di-GMP, supporting a role for REC phosphorylation in the regulation of product identity[141]. DgcA has also been shown to interact directly with the PDE Bd1971[210]. Moreover, Bd2325, a specific cGAMP PDE,[211] has been identified as essential for predation[212]. Together, these findings suggest that PDE activity could dictate whether c-di-GMP or cGAMP signalling occurs at any one time[141].

AglR1 was also identified as a hit in the Bd1996 pull-down assays. However, we were unable to experimentally validate the Bd1996:AglR1 interaction in the time available to complete this project. Although no obvious interaction was observed in the Bd1996:AglR1 AF model, a component of the AglRQS motor would be a logical downstream target for a protein proposed to regulate gliding motility. As such, it was decided that the interaction between Bd1482 and Bd1481 (AglR3) would also be modelled. Remarkably, the Bd1482:AglR3 AF model predicts that a PRS (Residues 228–232, sequence: PPSPG) located on a C-terminal tail peptide of Bd1482 packs into the binding pocket of the Bd1482 GYF domain (Figure 4.30). Sequence comparison of AglR1 and AglR3 reveals that AglR1 lacks a proline-rich tail peptide, providing a possible explanation for the lack of interaction in the Bd1996:AglR1 AF model. Nevertheless, the identification of a Bd1996:AglR1 interaction by pull-down assay and a possible Bd1482:AglR3 interaction by AF modelling provides an exciting foundation for future study.

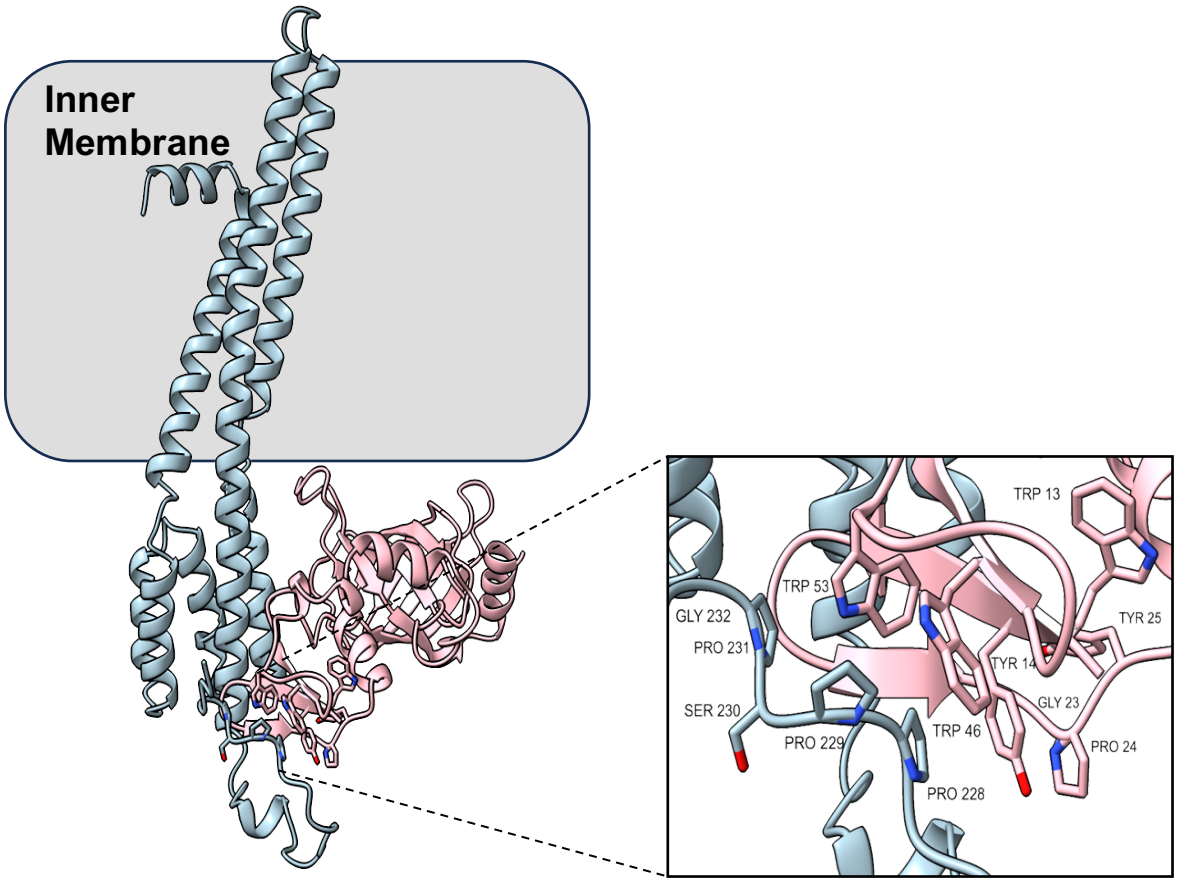


Figure 4.30: AF model of Bd1482:AglR3 interaction
AF predicts an interaction between Bd1482 and AglR3 with a PRS (coloured grey) on the C-terminal tail of AglR3 binding to the hydrophobic pocket on the GYF domain of Bd1482 (coloured pink). Average pLDDT (confidence) scores for AF models are as follows: Bd1482 = 72.2; and AglR3 = 84.2.

4.2.3 $MglA_{Bd}$ appears to undergo GTP-dependent conformational changes to interact with Bd2492

Previous studies have proposed that the Gly21Ser substitution in the P-loop of $MglA_{Bd}$ results in $MglA_{Bd}$ being trapped in a GTP-bound state[137]. The equivalent substitutions in $MglA_{MX}$ (G21V) and Ras (G12S) result in GTPases which are insensitive to both intrinsic and GAP-mediated GTPase activity[195,213]. Therefore, the presence of permanently active $MglA_{Bd}$, which is insensitive to GTPase activity, has been hypothesised to be the reason as to why *B. bacteriovorus* lacks a MglB homologue[137]. However, the structural data presented in this chapter challenges the hypothesis of

permanently GTP-bound $MglA_{Bd}$. Firstly, despite being in a mixed active/inactive conformation, our $MglA_{Bd}$ crystal structure appears to be predominately GDP-bound, indicative of intrinsic GTPase activity occurring during crystallisation. Moreover, the structural insights gained from the comparisons between the crystal structure of $MglA_{Bd}$ and the $MglA_{Bd}$:Bd2492 AF model suggest that despite $MglA_{Bd}$ and $MglA_{Mx}$ regulating distinct processes, both proteins undergo the same GTP-dependent conformational changes to regulate interactions with their binding partners. This is a remarkable observation; particularly given the fact that the same conformational change occurs to license interactions with both monomeric and dimeric binding partners, highlighting the versatility of MglA.

These conformational changes provide an explanation as to why only $MglA_{Bd}$ crystallised despite the $MglA_{Bd}$:$Bd2492_{205–353}$ interaction being maintained throughout all purification steps. All of the $MglA_{Bd}$ crystals produced in this work grew extremely slowly, taking ~6–8 months on average. During this time, the low intrinsic GTPase activity of $MglA_{Bd}$ would cause the hydrolysis of GTP. The transition of $MglA_{Bd}$ from GTP- to GDP-bound state would result in a conformational change in switch 1 region, breaking the interaction with Bd2492. Presumably, as the concentration of $MglA_{Bd}$-GDP increases, the protein nucleates. This results in the formation of $MglA_{Bd}$-GDP crystals which may also leave some residual GTP as observed in the experimental maps. It may be of interest to probe newer and older samples of $MglA_{Bd}$:Bd2492 via HPLC to correlate oligomerisation state and age with nucleotide content.

Conventional GTPase:GAP interactions are usually characterised by the insertion of an arginine/asparagine residue from the GAP into the active site of the GTPase to complete the catalytic site that subsequently permits hydrolysis of GTP[83,214,215]. However, the $MglA_{MX}$:MglB interaction is unconventional in the fact that the GAP activity of MglB is mediated by triggering a conformational change in MglA to form a full catalytic site within MglA[83]. It is therefore possible that although *B. bacteriovorus* lacks a MglB homologue, the interaction between $MglA_{Bd}$ and a binding partner could result in the conformational changes required for GAP activity. The current AF model of the $MglA_{Bd}$:Bd2492 interaction suggests that the binding of Bd2492 to $MglA_{Bd}$ does not trigger GAP activity as Arg53 and Thr54 are modelled projecting away from the nucleotide/Mg^{2+} binding site. Comparison of the $MglA_{MX}$-GTPγS and $MglA_{MX}$-GTPγS-MglB crystal structures to the $MglA_{Bd}$:Bd2492 AF model reveals why the $MglA_{MX}$:MglB interaction promotes GAP activity, but the $MglA_{Bd}$:Bd2492 interaction may not (Figure 4.31). When superimposing the $MglA_{MX}$-GTPγS structure to the $MglA_{MX}$-GTPγS-MglB structure, it can be seen that one MglB monomer clashes with the GTP-bound conformation of switch 1 seen in the $MglA_{MX}$-GTPγS structure. To prevent this clash, in the presence of MglB, switch 1 reorientates and in doing so, completes the active site. In contrast, in the $MglA_{Bd}$:Bd2492 AF model, whilst Bd2492 interacts with surface-exposed Phe57 and Phe59 of the active conformation of switch 1, it does not clash with the remainder of the switch 1 region, meaning that complete reorientation of the switch 1 region is not strictly necessary. It is unclear, however, whether this is due to poor *ab initio* modelling of the interaction. Experimental determination of the $MglA_{Bd}$:Bd2492 structure will therefore provide a more complete insight into the interaction.

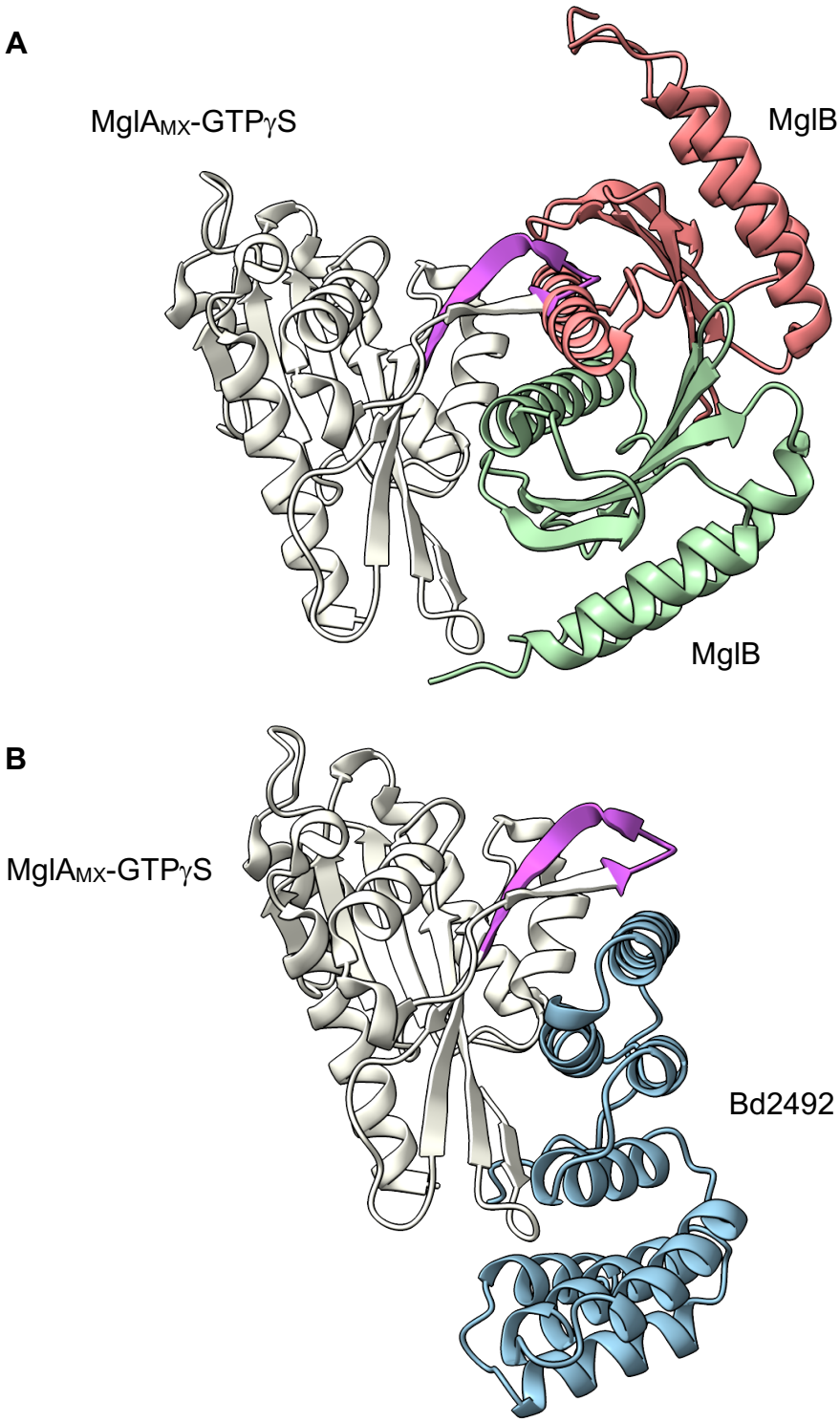


Figure 4.31: Superimposition of MglA$_{MX}$-GTPγS to MglA$_{MX}$:MglB and MglA$_{Bd}$:Bd2492 AF model

A – Superimposition of the MglA$_{MX}$-GTPγS and MglA$_{MX}$:MglB crystal structures. The GTP-bound conformation of switch 1 (coloured purple) clashes with the upper MglB monomer, explaining the observed remodelling of switch 1 in the presence of MglB.

B – Superimposition of the MglA$_{MX}$-GTPγS and MglA$_{Bd}$:Bd2492 AF model. As Bd2492 is monomeric, the GTP-bound conformation of switch 1 does not clash at the interface, therefore switch 1 is not remodelled and no GAP activity is observed

In our crystal structure, $MglA_{Bd}$ occupies a mixed active/inactive conformation. A similar mixed state has also been observed for $MglA_{MX}$[194]. In *M. xanthus*, this mixed active/inactive conformation of MglA has been associated with a sub-population of MglA-GTP that is insensitive to the GAP activity of MglB and cannot interact productively with the Agl-Glt complex[194]. Whilst interesting, the function of this apparently inactive population of $MglA_{MX}$-GTP remains to be fully characterized.

4.2.4 Conclusions and Future studies

The work presented in this chapter provides further structural characterisation of Bd1996, although a full-length experimental structure remains elusive. We have identified DgcA and AglR as possible Bd1996 binding partners and future work will focus on validating these interactions. To date, all attempts to produce recombinant DgcA have been unsuccessful. However, these attempts have currently focussed on expressing the full-length DgcA protein. As the interaction appears to occur mainly between Bd1996 and the REC domain of DgcA, there may be more success in working on a REC-only DgcA construct. Once an expression and purification protocol has been established for DgcA, the Bd1996:DgcA interaction can be validated by ITC and/or crystallography. To the best of our knowledge, Bd1996 represents the first reported case of a PilZ domain with dual affinity for c-di-GMP and cGAMP. Therefore, structural characterisation of the Bd1996:DgcA complex would provide the first structural insights into this novel signalling pathway. It may also be possible to probe conformational differences in the presence of c-di-GMP/cGAMP to identify how each cyclic dinucleotide facilitates signalling.

In terms of validating an interaction with AglR, it appears that Bd1482 may be a more compelling target based on our AF modelling of the Bd1482:AglR3 interaction. However, historically, we have had difficulty producing recombinant Bd1482. If Bd1482 expression cannot be optimised, the focus should therefore shift towards working with collaborators to validate the Bd1482:AglR3 interaction *in vivo*. Obtaining structural data for the potential Bd1996:AglR1 and/or Bd1482:AglR3 interaction will help describe the chronology behind cGAMP activation/inactivation of AglRQS complex, providing key insights into the regulation of gliding motility.

It is likely that Bd1996 interacts with more binding partners than those identified in this chapter. The pull-down assays performed in this study used *B. bacteriovorus* lysate and as such, would only predominately identify soluble proteins (with AglR being an apparent exception). Moreover, the lysate used would only reflect the gene expression profile that was present at the time of harvesting. Moving forward, it would be beneficial to perform more Bd1996 pull-down assays against *Bdellovibrio* cells that have been harvested at distinct timepoints; e.g., as cells face starvation. Despite being unsuccessful in this work, phage display still represents a good opportunity to identify Bd1996 binding motifs. In this work, we were restricted by our collaboration to only sending one sample and as such, full-length Bd1996 was chosen due to concerns around the stability of the Bd1996-GYF construct. Phage display should be repeated for both full-length Bd1996 and Bd1996-GYF with additional steps included to prevent possible polystyrene/plasticware false positives.

The combined experimental and computational structural characterisation of the $MglA_{Bd}$:Bd2492 complex has revealed that despite *B. bacteriovorus* lacking an MglB gene, $MglA_{Bd}$ still appears to be involved in a GTPase cycle and requires GTP-dependent conformational changes to interact with binding partners. Moving forward, the primary focus should be on the experimental structural determination of the $MglA_{Bd}$:Bd2492 complex. In *M. xanthus*, $MglA_{Q82L}$ is unresponsive to both intrinsic and GAP-activated GTPase activity and therefore the cloning of an $MglA_{BdQ82L}$ construct likely represents the best opportunity of crystallising the $MglA_{Bd}$:Bd2492 complex. Alternatively, a designed disulphide to physically prevent dissociation of the heterodimer could be used. Complete structural characterisation of the $MglA_{Bd}$:Bd2492 complex will provide key insights into how $MglA_{MX}$ and $MglA_{Bd}$ utilise a common interface to regulate diverse biological processes.

Chapter 5

Structural characterisation of the nuclease Bd1431

5.1 Introduction

5.1.1 The importance of secreted enzymes for predation.

B. bacteriovorus's genome is highly specialised for predation, encoding for a large number of (putative) secreted degradative enzymes. Bioinformatic comparison of 115 Gram-negative strains revealed that *B. bacteriovorus* had the largest secretome of the compared genomes, with ~42% of *B. bacteriovorus's* genome encoding for secreted proteins[2,216]. Within this secretome is a large arsenal of around 293 hydrolytic enzymes which forms part of *B. bacteriovorus's* predatosome[2,16] (Table 5.1).

Table 5.1: Summary of *B. bacteriovorus's* degradative enzymes

Enzyme Class	*B. bacteriovorus* HD100
DNAse	20
RNAse	9
Lipase	15
Glycanases	10
Peptide and protease	150
Other hydrolytic enzymes	89

5.1.2 Importance of nucleases

Amongst the array of degradative enzymes used by *B. bacteriovorus*, nucleases play a key role in replication within the *Bdelloplast*. A replicating *B. bacteriovorus* is required to synthesise multiple genomes from predating upon a single prey cell[217]. It is therefore

clear that alongside DNA synthesis, the breakdown and recycling of prey nucleic acids is of great importance[217]. Consistent with this, *B. bacteriovorus* contains predicted nucleoside porins for the uptake of prey nucleosides[218].

The early work of Rittenburg and colleagues demonstrated that the degradation of prey nucleic acids within the *Bdelloplast* is a highly regulated process that involves the sequential synthesis and release of nucleases from *B. bacteriovorus*[219,220]. Within the first hour of invasion, prey cell DNA is degraded into ~800 bp fragments which are retained within the *Bdelloplast*[27,220,221]. The formation of these fragments is a result of the accumulation of single- and double-strand cuts driven by the regulated release of endonucleases with different specificities[221]. The rapid formation of these intermediately sized fragments occurs before the initiation of *B. bacteriovorus* nucleic acid synthesis[220,221]. Exonucleases further degrade these fragments into mononucleotides at a rate comparable to that of *B. bacteriovorus* DNA synthesis, suggesting that these mononucleotides are incorporated rapidly into progeny DNA[27,221]. At the time of prey cell lysis, around 70% of prey DNA has been incorporated into progeny DNA[220,221].

B. bacteriovorus's genome contains 20 genes encoding for predicted nucleases[27], although, which amongst these nucleases are responsible for the degradation of prey DNA remains to be confirmed. However, it is reasonable to assume that such nucleases would be secreted from *B. bacteriovorus* into the prey cell, although the transit of any *B. bacteriovorus* enzyme across the intact prey IM remains an open question. Five of *B. bacteriovorus's* nucleases are predicted to contain N-terminal

signal peptides for secretion to the periplasm and beyond. The remaining nucleases share homology to known housekeeping genes[217]. Of the secreted nucleases, three share homology to members of the Endonuclease_I family: Bd1244, Bd0934 and Bd3507[27,217]. In contrast, Bd1431 and Bd1934 are predicted to be members of the *Staphylococcal nuclease* (SNase) family[217].

The activities of four of *B. bacteriovorus's* secreted nucleases (Bd0934, Bd1244, Bd1934 and Bd3507) have been characterised in two independent studies[217,27]. Transcriptomic analysis revealed that Bd1244 and Bd1934 were upregulated during the early stages of invasion[41,217]. Further RT-PCR analysis of secreted nuclease expression throughout *B. bacteriovorus's* lifecycle demonstrated that Bd1244 and Bd1934 expression is induced upon contact with prey[27,217] highlighting these nucleases as candidates for carrying out the initial degradation of prey DNA[217]. However, both single and double deletion mutants of Bd1244 and Bd1934 showed that neither nuclease was essential for predation and that the deletions had no impact on *B. bacteriovorus's* ability to degrade prey DNA[217]. Indicating that these enzymes are not responsible for the breakdown of prey DNA within the *Bdelloplast* unless their activity is compensated for by the upregulation of *B. bacteriovorus's* other secreted nucleases.

Although Bd1244 and Bd1934 do not appear to play a role in the degradation of prey DNA, they are involved in the formation and degradation of biofilms[217]. *B. bacteriovorus* is able to form self-biofilms as well as successfully predate within prey biofilms[222–224]. *B. bacteriovorus* biofilms are degraded upon treatment with proteinase K and DNase

I, indicating that proteins and nucleic acids form structural components of the biofilm[223]. Extracellular DNA (eDNA) is widely reported as a key structural component of many biofilms and the roles of endonucleases in biofilm formation have been documented in other bacteria[217,225]. The deletion of Bd1244 led to less efficient self-biofilm formation, suggesting that Bd1244 plays a possible role in processing eDNA to help shape biofilm architecture[217]. It is possible that that the loss of a biofilm shaping nuclease may result in an accumulation of eDNA which prevents additional *B. bacteriovorus* cells from integrating into the biofilm, reducing the overall biomass of biofilms[217].

Intriguingly, the deletion of Bd1244 and Bd1934 led to the increased clearance of established prey biofilms[217]. The authors explained this surprising result by proposing that the lack of these nucleases hinders *B. bacteriovorus's* ability to traverse biofilms, resulting in a more localised killing within a biofilm[217].

The role of Bd0934 and Bd3507 in prey DNA degradation has also been characterised[27]. RT-PCR analysis revealed that Bd0934 and Bd3507 expression is observed during the mid-stages of intraperiplasmic growth[27]. Notably, the expression of Bd0934 and Bd3507 correlates with the initiation of *B. bacteriovorus* DNA synthesis. This led to the proposal that these nucleases could act to break down prey DNA fragments into single nucleotides to be incorporated into progeny DNA[27]. However, both single and double deletions of Bd0934 and Bd3507 revealed that neither of these nucleases were essential for predation[27]. qRT-PCR analysis revealed that neither Bd1244 or Bd1934 were upregulated in the Δ*bd0934* Δ*bd3507* strain, suggesting that compensatory action of these nucleases was unlikely[27]. However, Bd1431 was

observed to be upregulated in the Δ*bd0934* Δ*bd3507* strain[27], suggesting that this nuclease may be able to compensate for the nuclease activity in the double deletion strain.

5.1.3 An introduction to Bd1431

Bd1431 featured in both previous studies which characterised *B. bacteriovorus's* secreted nucleases. Despite being upregulated following invasion[41], Bd1431 was deemed unlikely to be involved in prey DNA degradation as it does not contain a signal peptide for secretion. Instead, Bd1431 contains a lipoprotein signal peptide for membrane localisation. Bd1431 is a member of the SNase family. SNases form a group of non-specific endo/exonucleases which are active against RNA and both single and double stranded DNA[226,227].

5.1.4 Introduction to SNase

SNase was first identified in 1956[228] and was characterised as a secreted, calcium dependant, 5'-phosphodiesterase with activity against RNA and single- and double-stranded DNA[228,229]. Since its discovery, SNase has been extensively studied as a model protein for studying protein folding kinetics, thermodynamic stability and structure-function relationships[230–234]. It has since been found that *S. aureus* expresses two secreted nucleases: Nuc and Nuc2[226]. Nuc is secreted by all clinical isolates of *S. aureus*[235], a feature which has been utilised to develop PCR-based analysis methods to detect for *S. aureus* contamination and to design Nuc-activated probes to track *S. aureus* infections[226,235–243]. Once secreted from *S. aureus,* Nuc acts as a defence mechanism to prevent the *S. aureus* cell from becoming entrapped in neutrophil

extracellular traps (NETs) [235,244–247] and also plays a role in processing eDNA in *S. aureus* biofilms [235,243,248] Nuc2 is a membrane-associated, extracellular nuclease which shares the same biochemical characteristics as Nuc, although its biological function remains to be characterised[226,242].

5.1.5 Structure and mechanism of SNase

Structural studies of SNase span back to the early days of protein crystallography. The first crystals of SNase were reported in 1966[249], followed by 4 Å electron density maps of both native SNase and SNase in complex with the inhibitor Thymidine 3'5' diphosphate (TDP) in 1969[250]. The first structure of SNase, solved to 2 Å, was published in 1970[251]. At the time, SNase was the tenth unique protein structure to be determined[252]. The structure revealed that the overall fold of SNase consists of three α-helices which form around a five-stranded, antiparallel β-barrel (Figure 5.1). This beta barrel arrangement would later become characterised as an oligonucleotide-binding fold (OB-fold). This structure of SNase was assigned the now obsoleted PDB code (1SNS) and was part of a group of 20 structures which became the first models deposited within the PDB[252–254]. Owing to its relatively small size (149 amino acids), lack of disulphide bonds and its ability to be easily purified in large quantities, SNase has since been extensively studied in structural biology[231,234]. To date, there are over 300 entries for SNase in the PDB, many of which are a result of extensive mutagenesis work designed to gain insights into protein folding.

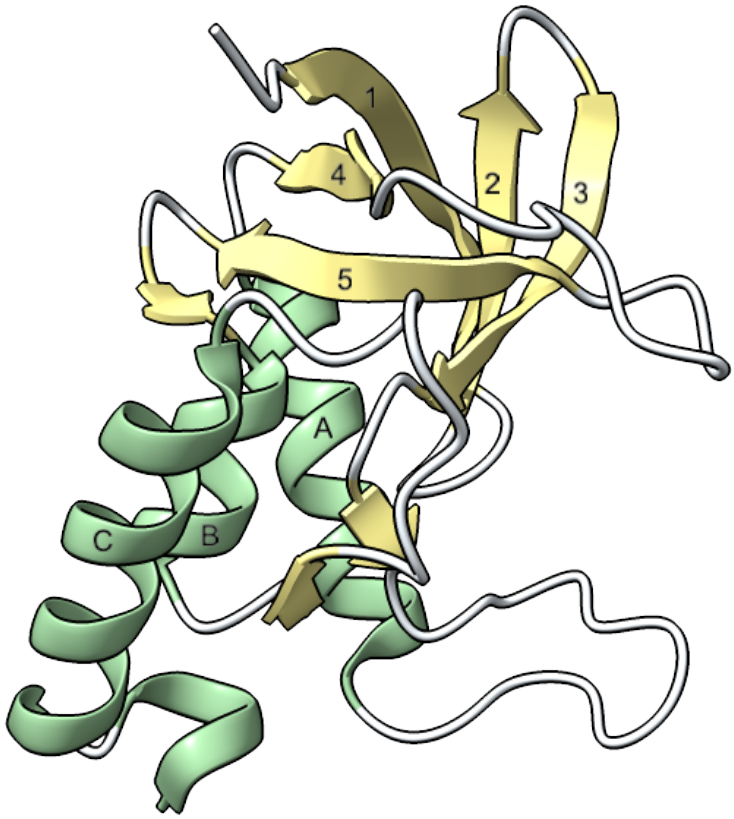


Figure 5.1: Structure of SNase fold
The overall architecture of the conventional SNase fold. β-strands (coloured yellow) have been labelled 1–5 and α-helices (coloured green) have been labelled A–C. PDB:1ey0

The publication of the 1.5 Å crystal structure of SNase in complex with Ca^{2+} and TDP provided high-resolution information into the interaction between SNase and its substrate and led to the proposal of a mechanism of action[255]. TDP binds into a pocket located beneath β-strands 2 and 3[255]. Upon ligand binding, Tyr113 and Tyr115 undergo a conformational change to enclose the binding pocket, providing a cavity into which the thymine ring packs[234]. The interaction is stabilised by hydrogen bonding between Arg35, Arg87 and the oxygen atoms of the 5' phosphate, whilst the 3' phosphate oxygens are held in place through interactions with Lys84 and Tyr85[255]. The calcium ion is hexacoordinated through the carbonyl oxygen of Thr41, the side chain carboxyl oxygens of Asp21 and Asp40, two water molecules and an oxygen from

the 5' phosphate[255]. Nuclease activity is conferred via nucleophilic attack, where Glu43 deprotonates the water molecule co-ordinated between Glu43 and the 5'-phosphate. This results in the formation of a hydroxide ion that ultimately carries out the nucleophilic attack, resulting in the hydrolysis of the 5' phosphodiester bond[255].

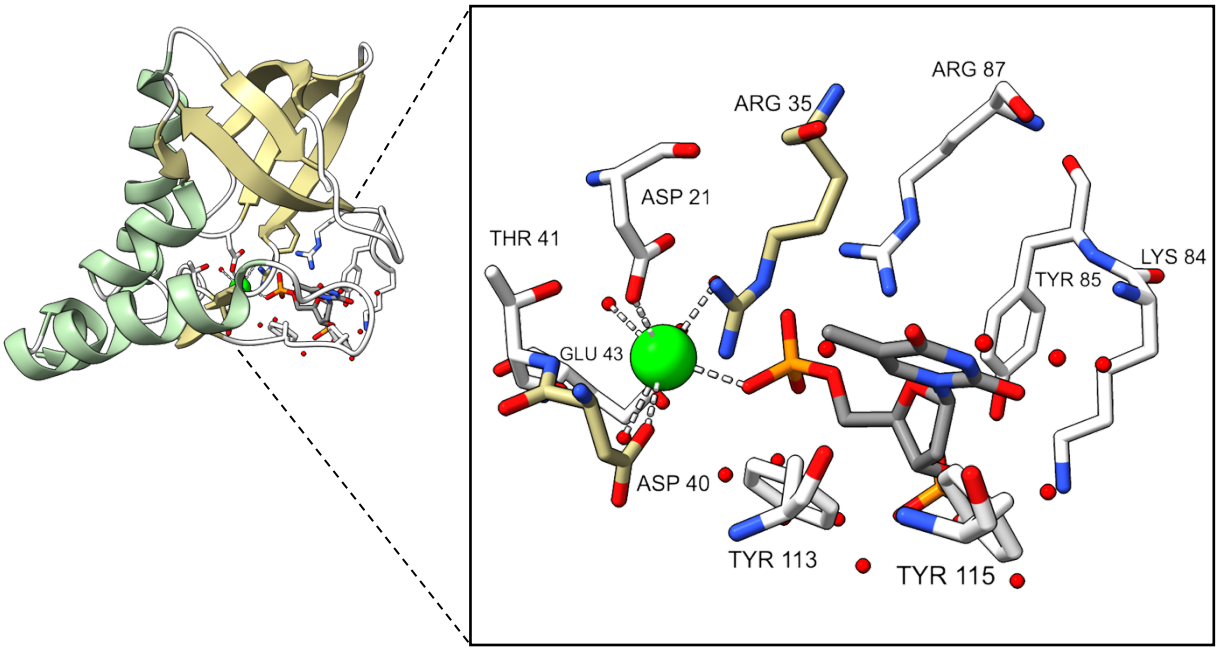


Figure 5.2: SNase active site
Co-ordination of Ca^{2+} and TDP in the SNase active site. The Ca^{2+} is co-ordinated by the sidechain carboxylates of Asp21 & Asp40 and the backbone carbonyl oxygen of Thr41. Arg35, Lys84, Tyr85 and Arg87 are responsible for nucleotide-binding whilst Glu43 acts as a general base to deprotonate the water molecule that subsequently acts as a nucleophile.

5.1.6 Chapter Aims

Bd1431 represents the only uncharacterised secreted nuclease within *B. bacteriovorus's* genome. As SNases are generally secreted from the cell, it is of interest to uncover why Bd1431 remains tethered to a membrane, expressed as a lipoprotein. The purpose of this study is to structurally characterise Bd1431 and attempt to underpin its possible function within *B. bacteriovorus*. If possible, attempts will be made to characterise nuclease-nucleotide interactions to determine whether Bd1431 differs from SNase in terms of engagement with nucleic acids.

5.2 Results

5.2.1 Expression and purification of Bd1431

Bd1431 was cloned as a soluble construct, with the native lipoprotein signal peptide being replaced with a PelB leader sequence for secretion to the periplasm. Bd1431 expressed well and was initially purified using IMAC. From the IMAC purification, two peaks were observed corresponding to two populations of protein (Figure 5.3). SDS-PAGE identified that the majority of Bd1431 was present within the second elution peak (Figure 5.3). Bd1431-containing fractions were pooled and further purified by SEC. Bd1431 was eluted as a monodisperse peak from HiLoad 26/600 Superdex 75 prep grade column at an elution volume of ~171 mL (Figure 5.4). When compared to standard curves, this elution volume corresponds to a molecular weight that is larger than expected for a Bd1431 monomer (~17 KDa). SDS-PAGE demonstrated the high purity of the post-SEC Bd1431 sample with the bulk of contaminant proteins being removed (Figure 5.4).

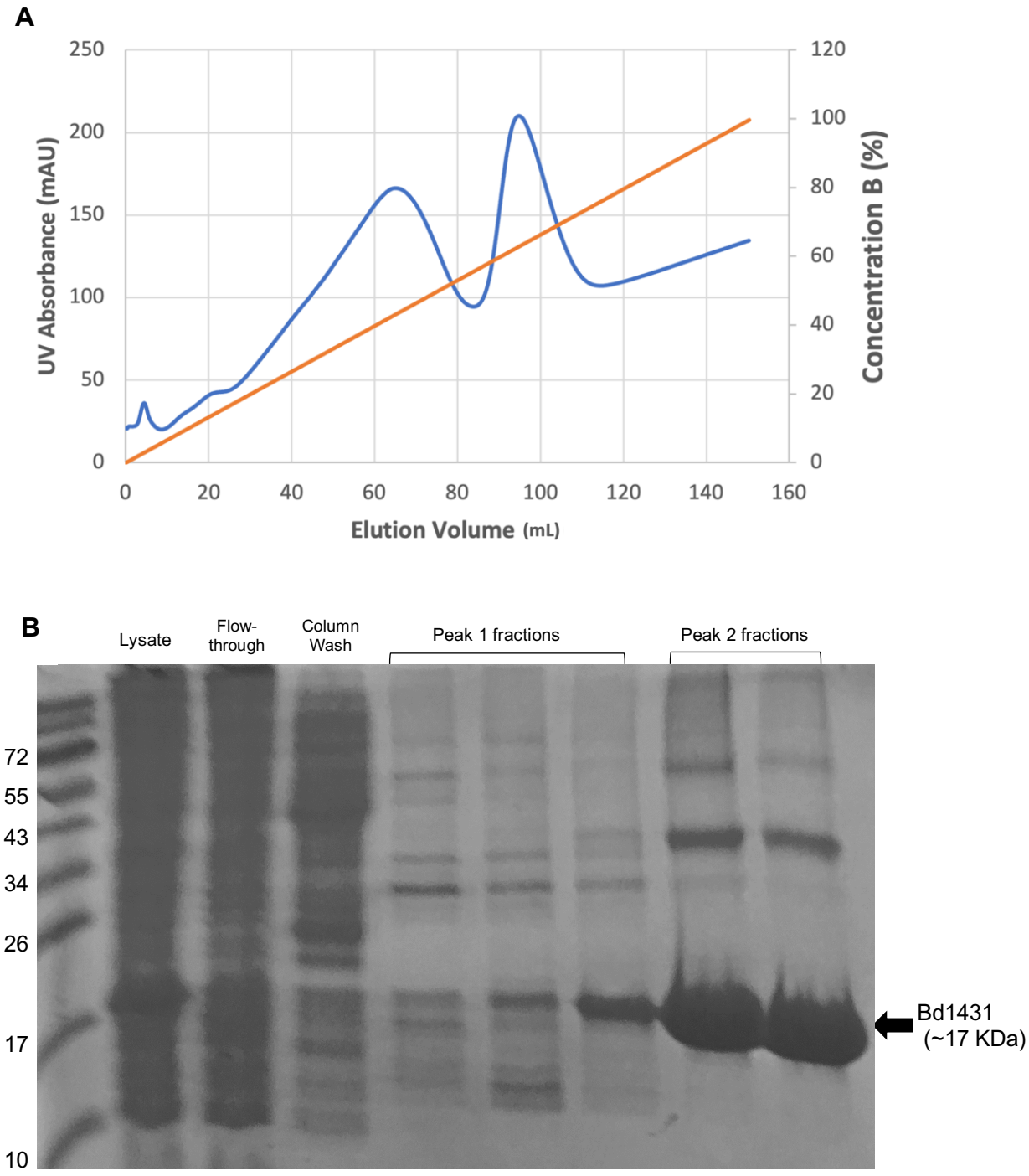


Figure 5.3: IMAC purification of Bd1431
A – Chromatogram of Bd1431 IMAC purification. IMAC performed on a 5 mL HisTrap HP using a gradient elution of 0-100% buffer B across 30 CV. Two peaks are observed on the chromatogram, corresponding to two populations of protein. Curves for UV absorbance at 280 nM (blue trace) and buffer B concentration (orange trace) are shown.
B – SDS-PAGE analysis of IMAC purification. SDS-PAGE reveals that Bd1431 is present in the second IMAC elution peak. Bd1431 can be seen as a predominant band in line with the 17 KDa MW marker. The gel also reveals the presence of some higher MW contaminants. The black arrow indicates the position of Bd1431.

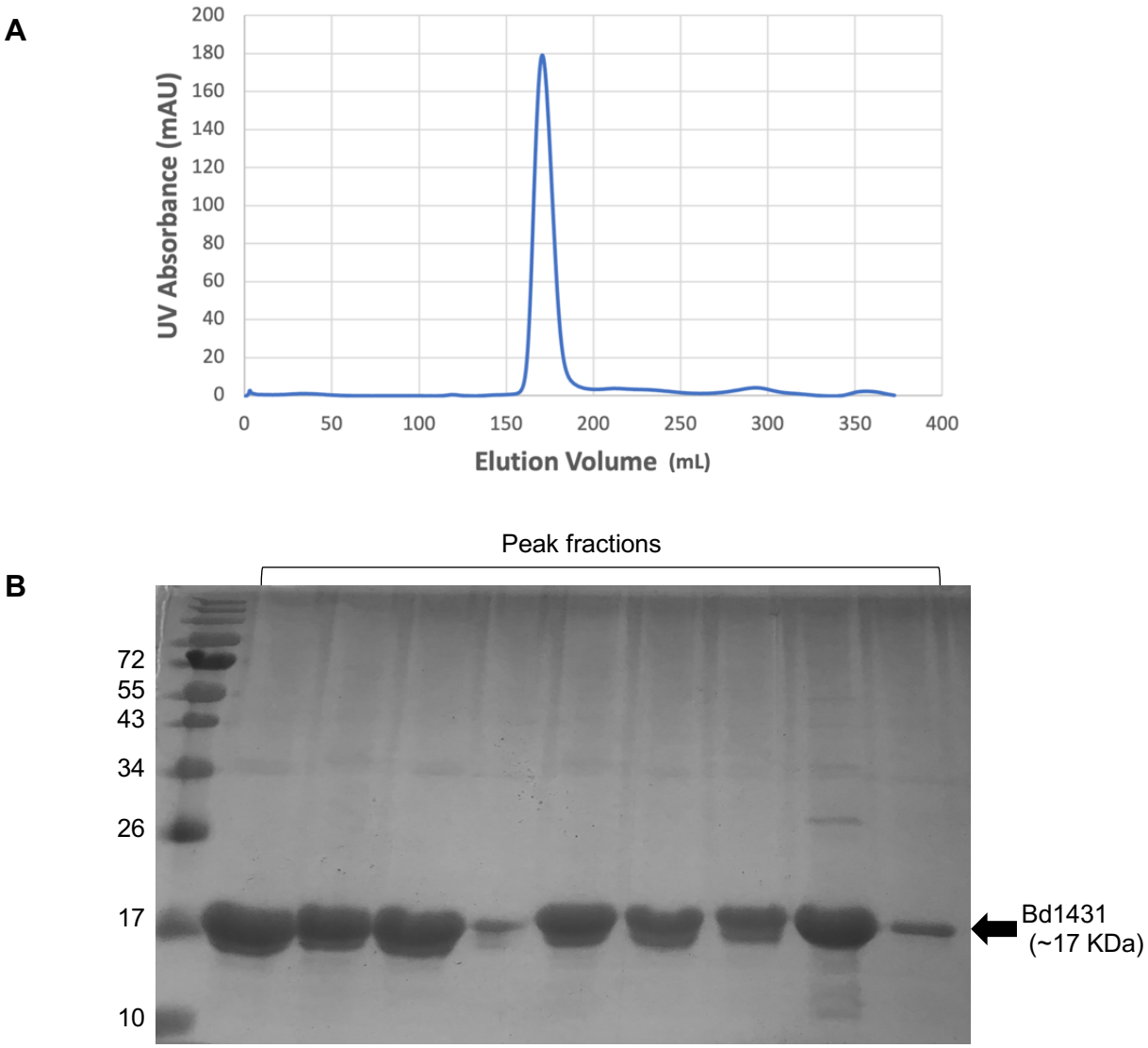


Figure 5.4: SEC purification of Bd1431
A – SEC Chromatogram for Bd1431. Bd1431 was eluted as a monodisperse peak from a HiLoad 26/600 Superdex 75 prep grade column at 171 mL, this elution volume corresponds to a molecular weight that is larger than expected for a Bd1431 monomer (~17 KDa). Curve for UV absorbance at 280 nM shown in blue.
B – SDS-PAGE analysis of SEC peak fractions. Bd1431 can be seen as a predominant band in line with the 17 KDa MW marker. The gel reveals that SEC has removed the majority of contaminants. The black arrow indicates the position of Bd1431.

5.2.2 Crystallisation

Initial sparse matrix screening was performed for Bd1431 using a Mosquito system. Clusters of needles were grown in JCSG E1 (1M Sodium Citrate, 0.1 M Sodium

Cacodylate pH 6.5); amongst these clusters, several individual crystals were present (Figure 5.5). Individual crystals were harvested and any remaining needle clusters were used as seeds for streak seeding. Optimisation by streak seeding was successful, reducing the abundance of clusters and promoting the formation of large rod-shaped crystals (Figure 5.5). Data collection for Bd1431 crystals was performed at Diamond Light Source beamline IO4.

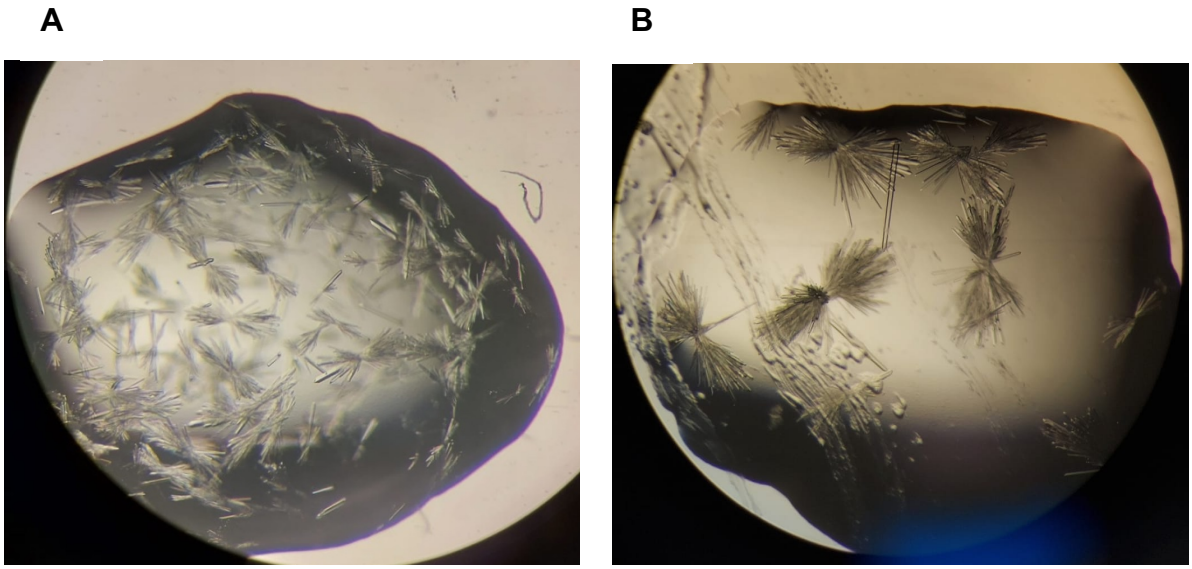


Figure 5.5: Bd1431 crystals
A – Initial crystal hits used as seeds for streak seeding. Crystallised in 1M Sodium Citrate, 0.1 M Sodium Cacodylate pH 6.5.
B – Optimised crystals used in x-ray diffraction experiments. Crystallised in 1M Sodium Citrate, 0.1 M Sodium Cacodylate pH 6.3.

5.2.3 Bd1431 Structure solution

Bd1431 crystals belonged to the P 2_1 2_1 2_1 spacegroup and diffracted to a relatively high resolution of 2.1Å. Matthews coefficient analysis indicated that two copies of Bd1431 is the most likely composition of the ASU. Initial phases were estimated by MR (PhaserMR)[163] using the crystal structure of SNase (PDB: 1ey0) as an initial search model. This model was first processed in Chainsaw[256] (CCP4)[257] to prune non-

conserved residues. Two copies of Bd1431 (denoted chain A & B) were placed in the ASU with a final LLG of 139.3 and TFZ of 12.1, indicating confident placement of the search model. The initial electron density map was of high quality and permitted model building. Following successive rounds of model building and refinement, residues 20–162 of chain A and 27–160 of chain B were successfully modelled.

Table 5.2: Data collection and refinement statistics for Bd1431

Collection Statistics	
Wavelength	0.8989 Å
Resolution range	49.28 – 2.1 Å (2.175 – 2.1 Å)
Space group	P 21 21 21
Unit cell	58.18 Å 70.49 Å 92.68 Å 90° 90° 90°
Total reflections	300821 (28003)
Unique reflections	22890 (2244)
Multiplicity	13.1 (12.5)
Completeness (%)	99.88 (99.38)
Mean I/sigma(I)	12.08 (0.69)
Wilson B-factor	49.98 Å^2
R-merge[a]	0.1377 (2.069)
R-meas[b]	0.1433 (2.157)
R-pim[c]	0.03921 (0.6036)
$CC^1/_2$	0.999 (0.606)
Refinement Statistics	
Reflections used in refinement	22864
Reflections used for R-free	1136
R-work[d]	0.2091
R-free[e]	0.2445
Number of non-hydrogen atoms	2240
macromolecules	2110
Ligands	113
Solvent	81
Protein residues	269
RMS(bonds)	0.004
RMS(angles)	0.68
Ramachandran favored (%)	97.72
Ramachandran allowed (%)	2.28
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	1.28
Clashscore	8.41
Average B-factor	63.74 Å^2
Macromolecules	63.74 Å^2
Ligands	83.83 Å^2
Solvent	51.53 Å^2

Statistics for the highest-resolution shell are shown in parentheses.

a Rmerge = $\sum hkl \sum i |I_i(hkl) - \langle I(hkl)\rangle| / \sum hkl \sum i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the ith measurement of a reflection with indexes hkl and $\langle I(hkl)\rangle$ is the statistically weighted average reflection intensity.

b Rmeas is the Rmerge adjusted by a factor of $n(n-1)^{1/2}$ to account for multiplicity.

c R-pim = $\sum hkl [1/(n-1)]^{1/2} \sum i |I_i(hkl) - \langle I(hkl)\rangle| / \sum hkl \sum_i I_i(hkl)$ where $I_i(hkl)$ is the intensity of the ith measurement of a reflection with indexes hkl and $\langle I(hkl)\rangle$ is the statistically weighted average reflection intensity.

d R-work = $\sum ||Fo| - |Fc|| / \sum |Fo|$ where Fo and Fc are the observed and calculated structure factor amplitudes, respectively.

e R-free is the R-work calculated with 5% of the reflections chosen at random and omitted from the refinement.

5.2.3.1 Bd1431 Overall Fold and comparison to conventional SNase fold

Both ASU copies of Bd1431 pack across a non-crystallographic two-fold axis to form a homodimer (Figure 5.6). Each monomer consists of a characteristic SNase fold comprising three α-helices (helix A–C) which pack around a five-stranded (β-strand 1–5) β-barrel. As five strands represents the minimum number required to form a β-barrel, the five β-strands of Bd1431 are highly twisted to permit closure[234,258]. At the N-terminus, residues 20–43 form a flexible peptide that would extend to the cysteine that tethers Bd1431 to its target membrane when expressed as a lipoprotein. The secondary structure elements are connected through a series of flexible loops that are key to activity and dimerisation. An 11-residue loop (residues 58–69) connects β-strands 2 and 3 (loop 2–3) and forms a major component of the dimer interface. A second 11-residue loop (residues 74–85) connects β-strand 3 to a α-helix A (loop 3–A). Loop 3–A lines the of edge the nucleotide-binding pocket and houses residues essential for Ca^{2+} co-ordination and catalytic activity. A shorter nine-residue loop connects β-strands 4 and 5 (loop 4–5) which forms the lower boundary of the dimer interface. A final loop formed between α-helix B and C (loop B–C) forms the lower boundary of nucleotide-binding pocket.

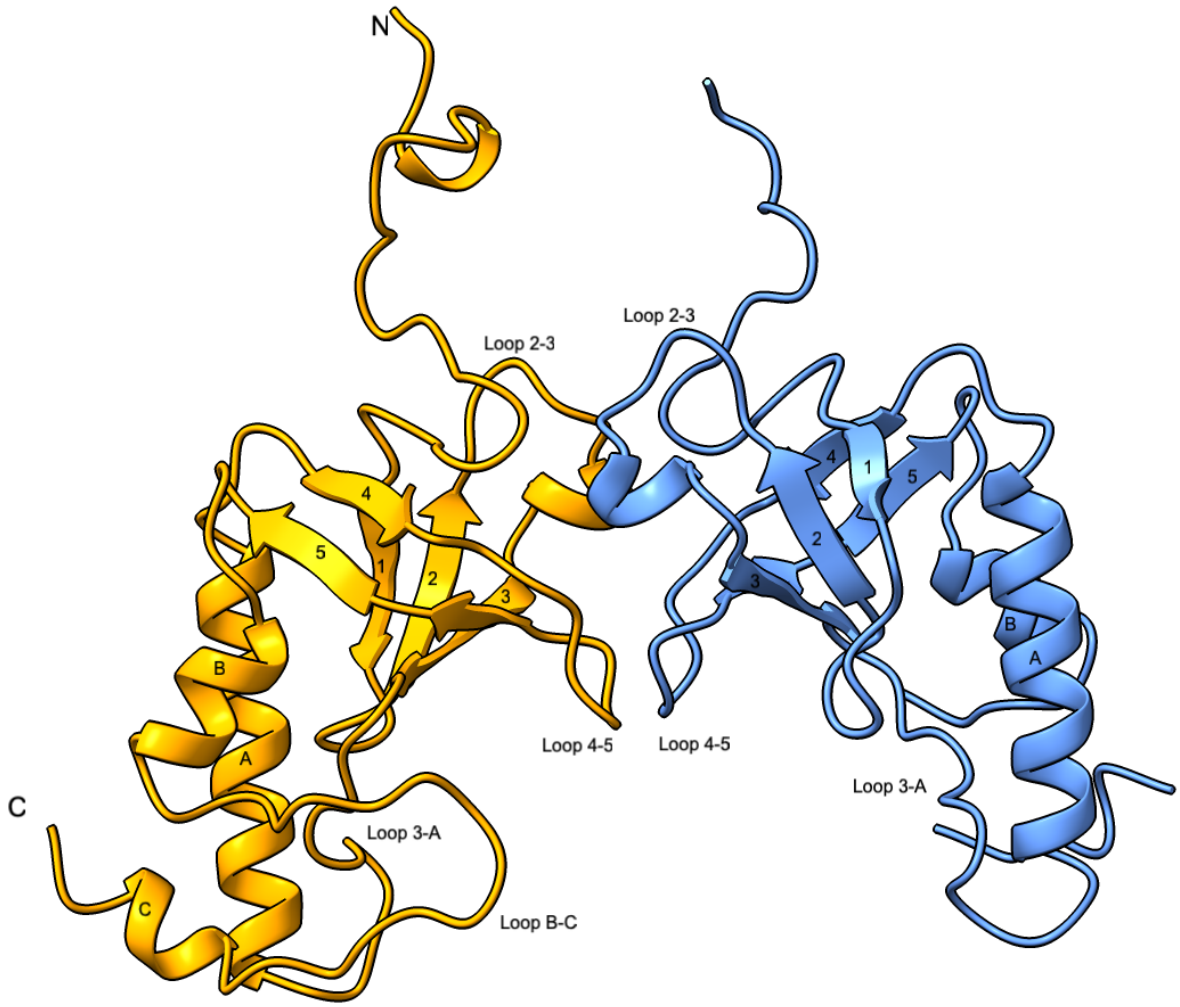


Figure 5.6: Overall fold of Bd1431 solved to 2.1 Å
The overall fold of the Bd1431 dimer with each monomer interacting across a non-crystallographic two-fold axis. β-strands are labelled 1–5 and α-helices are labelled A-C. Loops are labelled based on the secondary structure elements they connect, e.g., the loop between β-strands 3 and α-helix is denoted loop 3–A.

The structure of Bd1431 exhibits several deviations from the characteristic SNase fold. SNase's lack of disulphide bonds was a key contributing factor to its extensive use in protein folding studies. However, Bd1431 contains two disulphide bonds per monomer (Figure 5.7). The first disulphide bond forms between Cys33-Cys42 and tethers two sections of the flexible N-terminal peptide, creating a 'kinked' loop in the peptide that runs parallel to β-strand 4. The presence of this disulphide likely provides a degree of

rigidity to the N-terminal peptide and could help ensure the correct orientation of Bd1431 relative to the membrane. The second disulphide forms between Cys87-Cys157 connecting helices A and C, which likely restricts the movement of loop B–C.

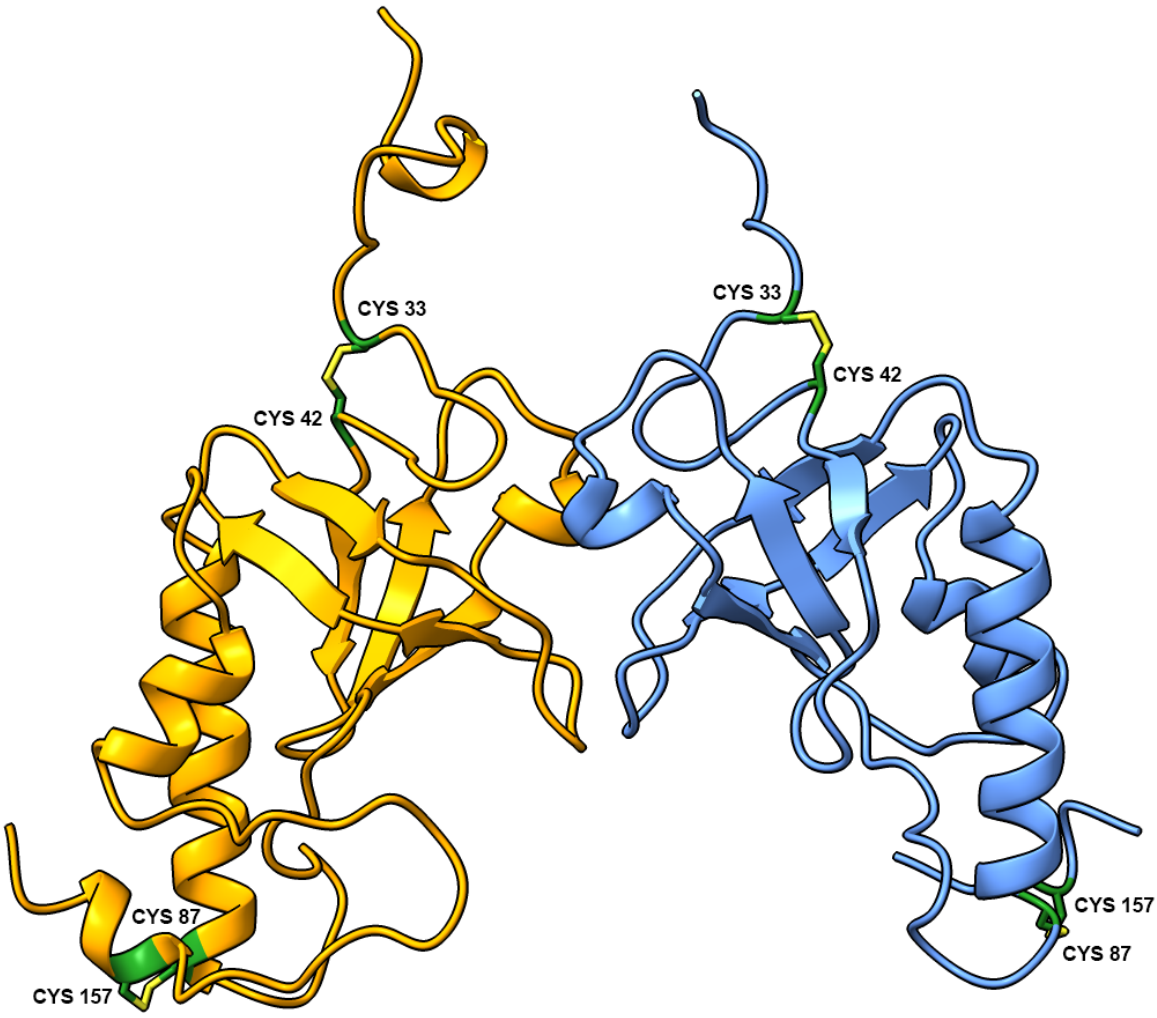


Figure 5.7: Each Bd1431 monomer contains two disulphide bonds
Bd1431 contains disulphide bonds between Cys33-Cys42 and Cys87-157 coloured green and yellow. The disulphide bond between Cys33-Cys42 stabilises the flexible N-terminal peptide and creates a 'kinked' loop that runs parrel to β-strand and forms part of the dimer interface. The disulphide bond between Cys87-157 connects α-helices A and C and could function stabilise loop B–C.

5.2.3.2 Active Site

All residues that are essential for calcium binding and nucleotide-binding/cleavage are conserved in Bd1431. The Ca^{2+} binding site is comprised of Asp52, Asp77 and Thr78 (Figure 5.8). Asp52 is located on the short turn between β-strand 1–2 whilst Asp77 and Thr78 are located on loop 3A. In chain A, a single Ca^{2+} occupies the binding site, co-ordinated by the sidechain carboxylates of Asp52 & Asp77 and the backbone carbonyl oxygen of Thr78. In chain B, crystal packing disrupts the Ca^{2+} binding site, preventing Ca^{2+} binding. The nature of the crystal contacts brings Arg-161 from a symmetry related molecule into close proximity of Asp52 which reorientates away from the Ca^{2+} binding site in order to form a salt bridge (Figure 5.9). The catalytic residues comprise Arg72, located on β-strand 3, Glu80, located on loop 3A and Arg121 located on loop 4–5.

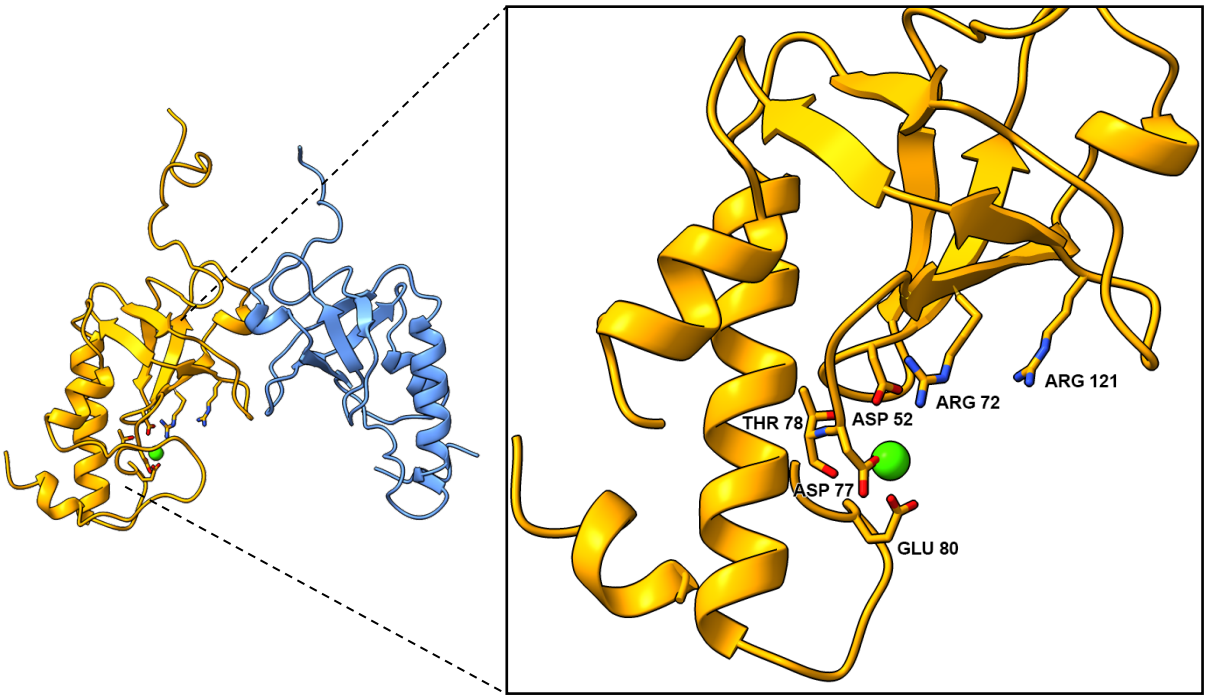


Figure 5.8: Bd1431 active site
Bd1431 contains all residues essential for Ca2+ and nucleotide-binding. The Ca2+ is co-ordinated by the sidechain carboxylates of Asp52 & Asp77 and the backbone carbonyl oxygen of Thr78 with bond lengths of 2.32 Å, 2.5 Å and 2.55 Å respectively. Arg72 and Arg121 are responsible for nucleotide-binding whilst Glu80 acts as a general base to deprotonate the water molecule that subsequently acts as a nucleophile.

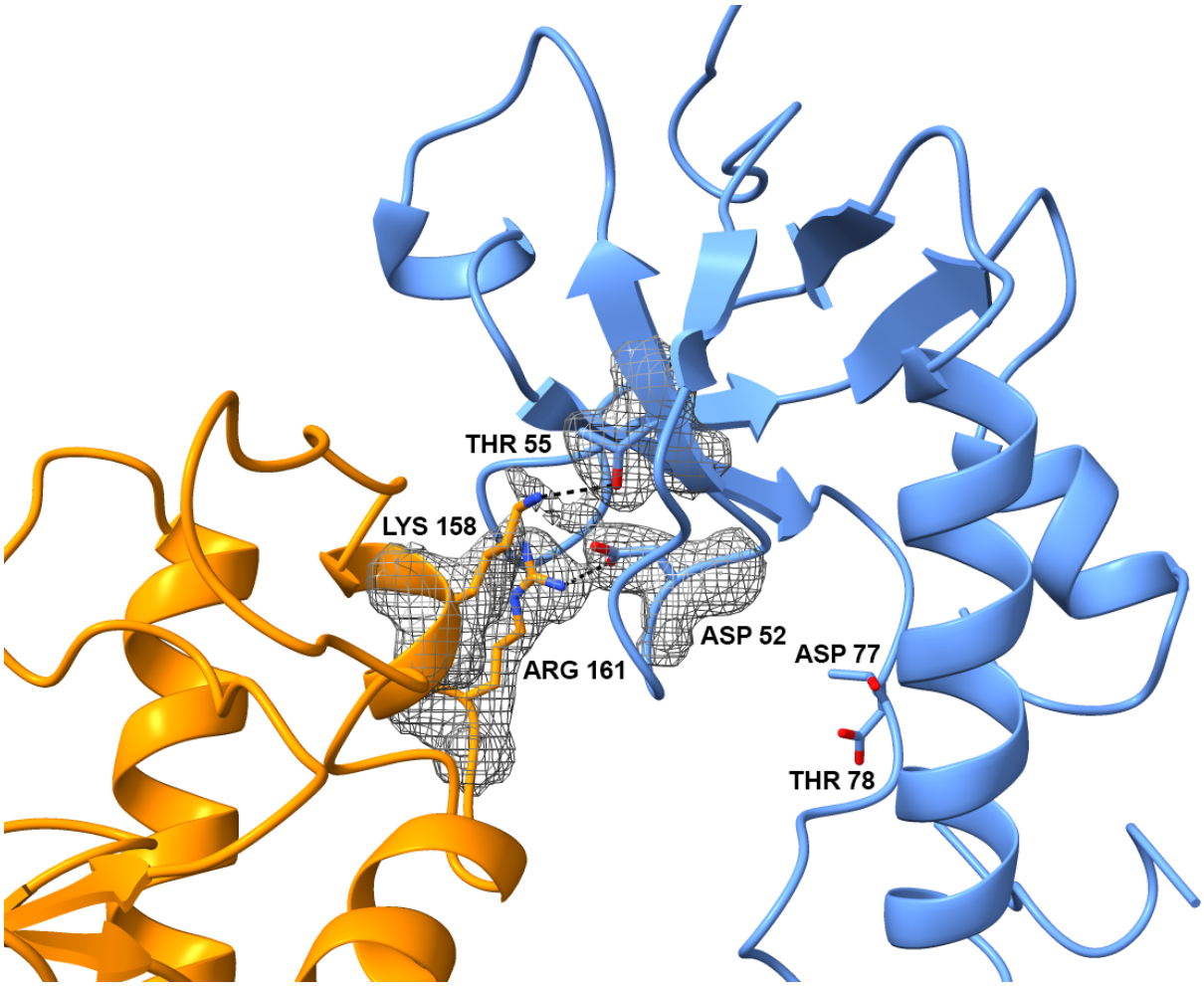
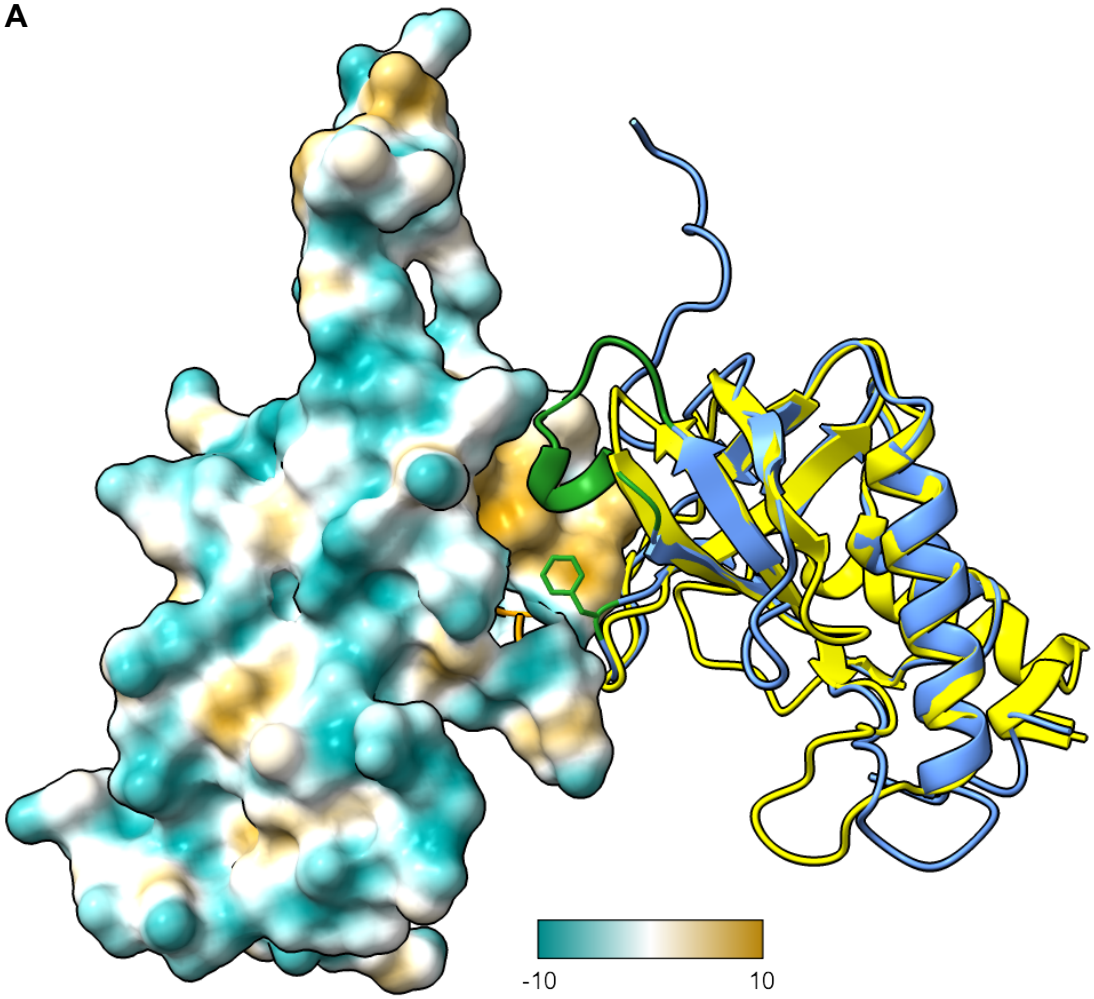


Figure 5.9: Crystal packing disrupts calcium binding in Bd1431 chain B
Asp52 on Chain B salt bridges with Arg161 on a symmetry copy of chain A. This interaction results in the reorientation of Asp52 away from the calcium binding site, preventing calcium binding. Electron density shown in ChimeraX (contoured to ~1.3 σ) to demonstrate correct modelling of Asp52.

5.2.3.3 Dimer interface

The dimer interface is extensive with a buried surface area of 835 Å^2 per monomer, as calculated by PISA[168]. The interaction is maintained by 12 hydrogen bonds and extensive hydrophobic packing between monomers. The interface is formed predominately by loop 2–3 and loop 4–5. Sequence alignment of Bd1431 and SNase demonstrates that both loops deviate from the conventional SNase sequence to form the dimer interface (Figure 5.10). In SNase, β-sheets 2–3 are connected by a short turn formed by two residues. In contrast, loop 2–3 of Bd1431 is elongated by the insertion of a stretch of hydrophobic residues (residues 59–66, sequence PNVPALIG) that form key interactions at the dimer interface. On loop 2–3, a single helical turn is formed by Ala63, Leu64, Ile65 and Gly66. The helical turns of each monomer pack against each other, whilst Leu64, emanating from the helical turn, packs against Val114 and Ile122 on the opposite monomer (Figure 5.11). Loop 4–5 is generally well conserved between SNase and Bd1431. However, in Bd1431, the substitution of a glycine residue to a phenylalanine (Phe120) stabilises the interaction between monomers through additional hydrophobic packing and face-to-face pi stacking between Phe120 of each monomer. Hydrogen bonding between Lys118 and Phe120 of opposite monomers also stabilise the lower boundary of the dimer interface (Figure 5.12). Additionally, Arg116 located on loop 4–5 hydrogen bonds with Lys67 and Lys68 on loop 2–3. Finally, the ‘kinked’ loop formed by the disulphide between Cys33-Cys42 faces into the dimer interface and forms hydrogen bonds with residues on the helical turn of loop 2–3 (Figure 5.12).

A



B

```
Bd1431 TVNIPNVPALIGKKI
SNase  KLMY------KGQPM
Bd1431 RDKYFRILA
SNase  TDKYGRGLA
```

Figure 5.10: Bd1431 dimer interface
A – Bd1431 dimer with chain A shown as a hydrophobic surface. The structure of SNase (PDB: 1ey0 coloured yellow) is superimposed onto Bd1431 chain B (coloured blue). The inserted stretch of hydrophobic residues on loop 2–3 and Phe120 on loop 4–5 are coloured green. Bd1431 surface coloured according to hydrophobicity with hydrophobic and hydrophilic residues coloured gold and cyan respectively. Surface representation created using ChimeraX.
B – Sequence alignment of loop 2–3 and loop 4–5 between SNase and Bd1431. The inserted stretch of hydrophobic residues on loop 2–3 and Phe120 on loop 4–5 are highlighted green.

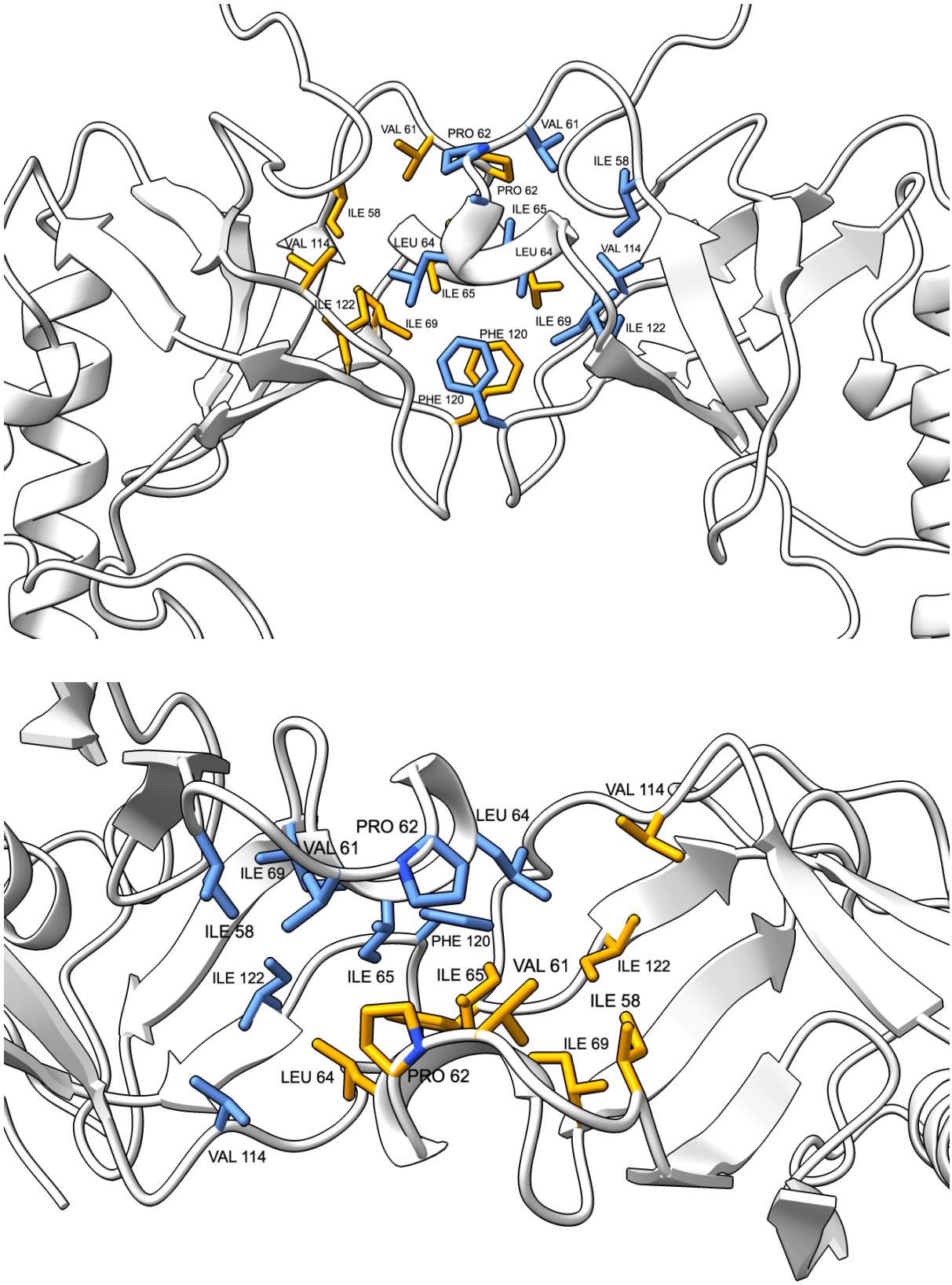


Figure 5.11: Schematic of hydrophobic packing at the dimer interface
Hydrophobic interactions at the dimer interface. Cartoons are coloured white for clarity, residues from chain A are coloured orange and residues from chain B are coloured blue. Hydrophobic residues are housed on loops 2–3 and 4–5.

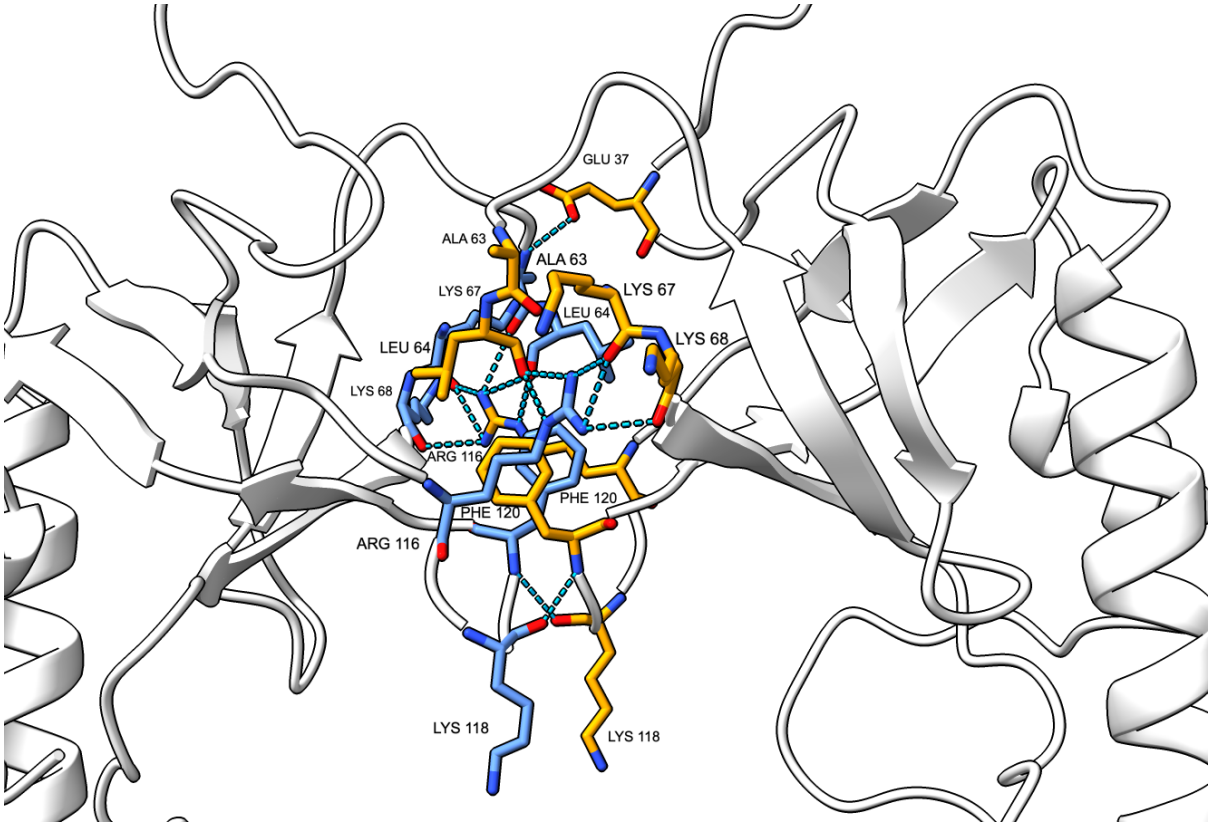


Figure 5.12: Schematic of hydrogen bonding at the dimer interface
Hydrogen bonding at the dimer interface. Cartoons are coloured white for clarity, residues from chain A are coloured orange and residues from chain B are coloured blue. Hydrogen bonds are depicted as dashed lines.

Interestingly, whilst carrying out interface analysis with PISA[168], the software identified a second interface across a crystallographic two-fold axis (Figure 5.13). This interface is considerably larger with a buried surface area of 1224 Å^2 per monomer. In total, there are 12 hydrogen bonds, one salt bridge and additional hydrophobic interactions at this interface. The interaction across this second interface appears more likely to be a consequence of crystal packing than a biologically relevant dimer interface. At this interface, the flexible N-terminal peptides face towards each other, packing alongside the exterior surface of β-sheets 4–5 and α-helix B. As these N-terminal peptides function to tether Bd1431 to a membrane, it seems unlikely that a dimer interface which

sees this peptide run in opposite directions between monomers would exist within the cell.

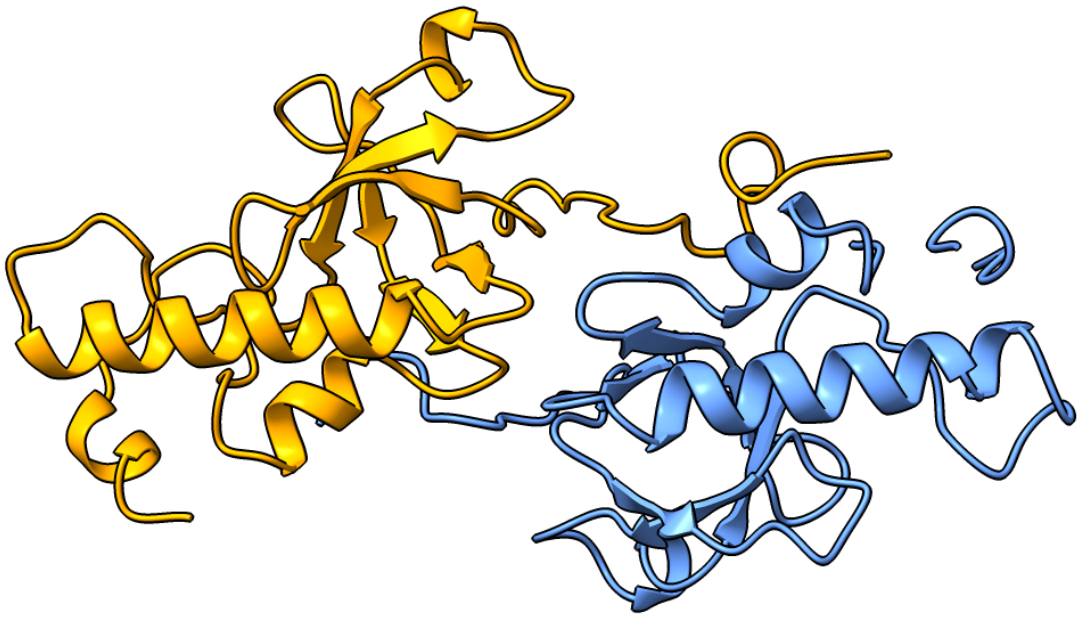


Figure 5.13: Second dimer interface identified by PISA
Possible dimer interface where Bd1431 monomers pack across a two-fold crystallographic axis. The nature of the interface suggests that this interface is an artefact of crystal packing.

5.2.4 Validation of the biological relevance of dimerisation

The occurrence of amino acid insertions and substitutions at the dimer interface is highly suggestive that the interface is biologically relevant and has evolved over time to facilitate dimerisation. The dimerisation of Bd1431 also provides an explanation to the observation that during SEC, Bd1431 elutes at an elution volume corresponding to a molecular weight larger than a Bd1431 monomer. To further probe the biological relevance of the dimer interface, $Bd1431_{F120D}$ was cloned with the assumption that the insertion of a negatively charged residue at the interface would prevent dimerisation.

Bd1431$_{F120D}$ expressed well and was purified as readily as the wild-type protein. The impact of the point mutation on dimerisation was initially assessed using SEC (Figure 5.14). Comparison of the SEC chromatograms for Bd1431 and Bd1431$_{F120D}$ revealed a clear shift in the elution volume between both samples. Wildtype Bd1431 eluted at 172 mL whilst Bd1431$_{F120D}$ eluted at 195 mL. When comparing to a standard curve, this ~23 mL shift in elution volume is roughly consistent with a shift from dimeric to monomeric Bd1431 and suggests that the F120D point mutation has successfully disrupted the dimer interface.

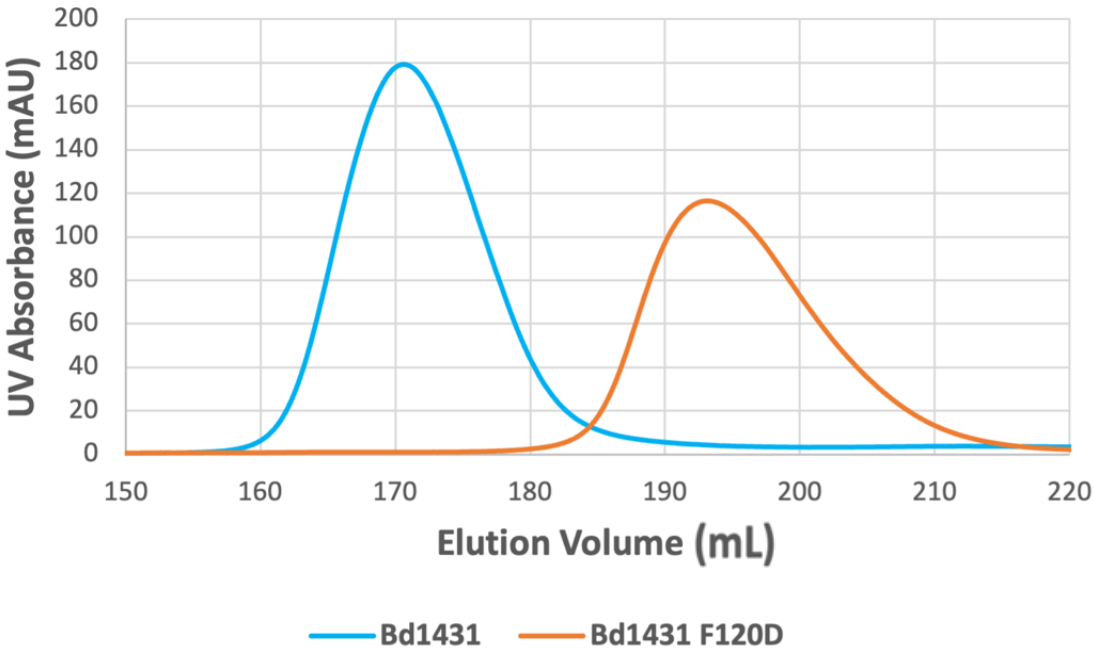


Figure 5.14: Comparison of SEC chromatograms between Bd1431 and Bd1431$_{F120D}$
The SEC elution profile of Bd1431$_{F120D}$ (coloured orange) and Bd1431 (coloured blue). The substitution of the hydrophobic phenylalanine at position 120 for aspartic acid disrupts the hydrophobic packing at the dimer interface and results in Bd1431$_{F120D}$ eluting as a monomer during SEC.

5.2.5 Crystal structure of Bd1431$_{F120D}$

To confirm that Bd1431$_{F120D}$ was unable to dimerise, the mutant was structurally characterised. Bd1431$_{F120D}$ crystals were grown by sitting-drop vapour diffusion. These crystals belong to the C 2 2 2_1 space group and diffracted to a resolution of 2.09Å. Matthews coefficient analysis indicated that three copies of Bd1431 is the most likely composition of the ASU. Initial phases were estimated by MR using one copy of the Bd1431 crystal structure as a search model. Three copies of Bd1431 (denoted chain A, B & C) were placed in the ASU with a final LLG of 2170 and TFZ of 45.3, indicating confident placement of the search model. The initial electron density map was of high-quality and permitted model building. Following successive rounds of model building and refinement, residues 28–161 of chain A, 28–161 of chain B and 27–165 of chain C were successfully modelled.

Table 5.3: Data collection and refinement statistics for Bd1431$_{F120D}$

Collection Statistics	
Wavelength	0.9793 Å
Resolution range	50.16 – 2.09 Å (2.165 – 2.09 Å)
Space group	C 2 2 2_1
Unit cell	117.89 Å 166.29 Å 69.10 Å 90° 90° 90°
Total reflections	507920 (29053)
Unique reflections	40575 (3987)
Multiplicity	12.5 (7.3)
Completeness (%)	97.22 (72.75)
Mean I/sigma(I)	10.26 (0.27)
Wilson B-factor	53.55 Å^2
R-merge[a]	0.1042 (2.6)
R-meas[b]	0.1085 (2.806)
R-pim[c]	0.03001 (1.021)
$CC^1/_2$	0.999 (0.369)
Refinement Statistics	
Reflections used in refinement	40575
Reflections used for R-free	1924
R-work[d]	0.2132
R-free[e]	0.2471
Number of non-hydrogen atoms	3193
macromolecules	3132
Ligands	0
Solvent	61
Protein residues	401
RMS(bonds)	0.009
RMS(angles)	1.42
Ramachandran favored (%)	97.46
Ramachandran allowed (%)	2.54
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	1.42
Clashscore	7.10
Average B-factor	73.37 Å^2
Macromolecules	73.70 Å^2
Solvent	56.39 Å^2

Statistics for the highest-resolution shell are shown in parentheses.

a Rmerge = $\sum hkl \sum i\ |I_i(hkl) - \langle I(hkl)\rangle| / \sum hkl \sum i\ I_i(hkl)$, where $I_i(hkl)$ is the intensity of the ith measurement of a reflection with indexes hkl and $\langle I(hkl)\rangle$ is the statistically weighted average reflection intensity.

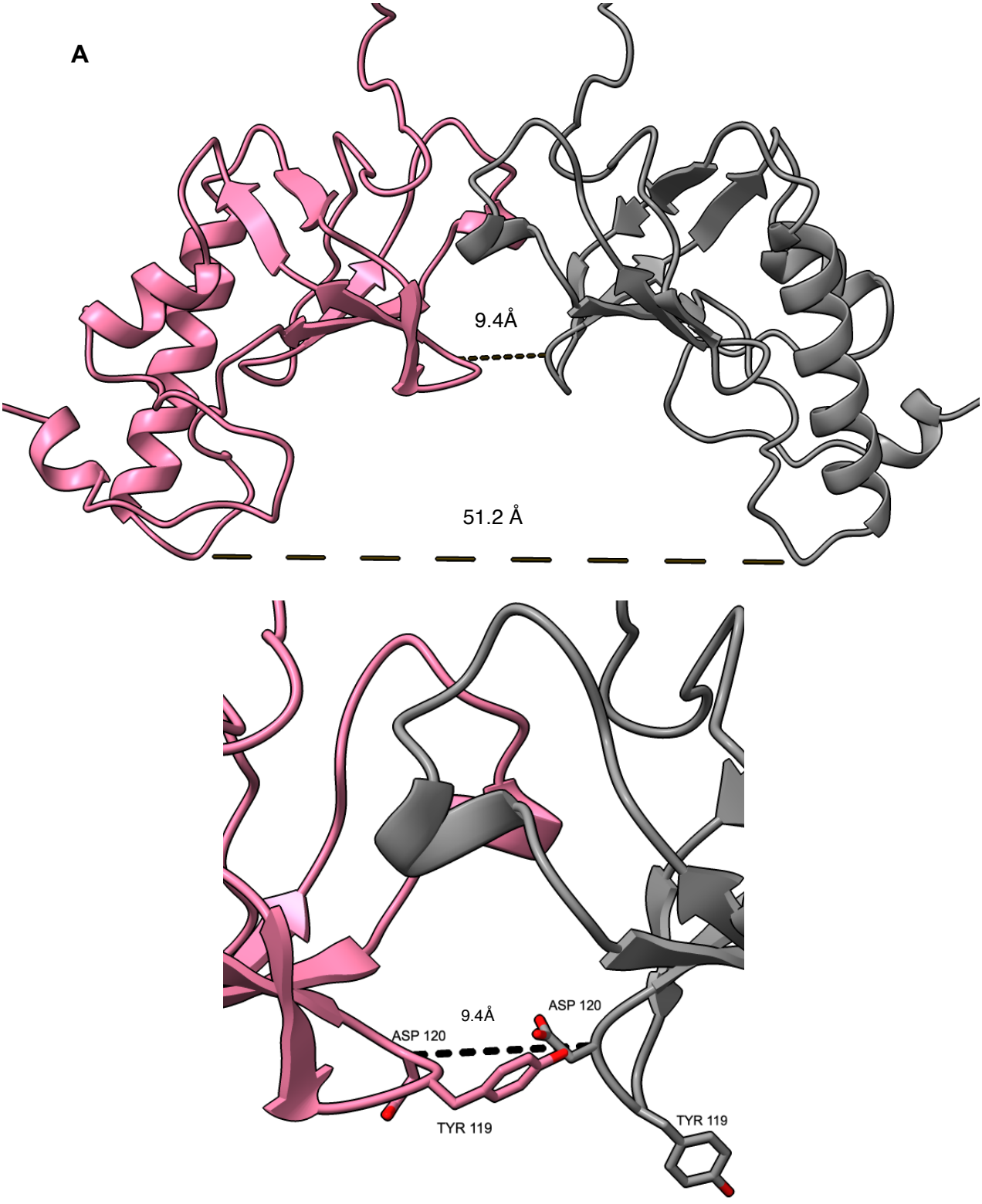
b Rmeas is the Rmerge adjusted by a factor of $n(n-1)^{1/2}$ to account for multiplicity.

c R-pim = $\sum hkl [1/(n-1)]^{1/2} \sum i |I_i(hkl) - \langle I(hkl)\rangle| / \sum hkl \sum_i I_i(hkl)$ where $I_i(hkl)$ is the intensity of the ith measurement of a reflection with indexes hkl and $\langle I(hkl)\rangle$ is the statistically weighted average reflection intensity.

d R-work = $\sum ||Fo| - |Fc|| / \sum |Fo|$ where Fo and Fc are the observed and calculated structure factor amplitudes, respectively.

e R-free is the R-work calculated with 5% of the reflections chosen at random and omitted from the refinement.

Despite Bd1431$_{F120D}$ appearing monomeric in solution, the dimer interface was maintained in the crystal, with each ASU copy of Bd1431$_{F120D}$ packing across a two-fold crystallographic axis (Figure 5.15). The point mutation has little impact on the overall fold of Bd1431, with the aligned wild-type and mutant structures having an RMSD of 0.58 Å. The insertion of the aspartic acid residue does indeed perturb the lower boundary of the dimer interface. In one monomer, Asp120 adopts the same orientation as the wild-type Phe120, pointing into the dimer interface. This results in the reorientation of Asp120 on the opposite monomer, which turns to face towards the nucleotide-binding pocket and pushes Tyr119 towards the interface. The repulsive interaction between D120 on opposing chains pushes the lower boundary of the dimer interface apart by ~5.5 Å (Figure 5.15). The point mutation also has an impact on the overall spatial arrangement of the Bd1431 dimer, pushing the two chains further apart and causing an ~10 Å expansion of the central cavity between each monomer (distance measured from Lys84) (Figure 5.15). Despite this, the F120D mutation has no impact on the upper boundary of the interface with the helical turns of both monomers being ~6.5 Å apart in both Bd1431 and Bd1431$_{F120D}$. This suggests that the undisrupted dimer interface acts as a pivot point to permit the lower half of dimer to swing apart. Remarkably, the second interface that was observed in the crystal structure of Bd1431 is also maintained in the Bd1431$_{F120D}$ structure.


A
9.4Å
51.2 Å
9.4Å
ASP 120
ASP 120
TYR 119
TYR 119

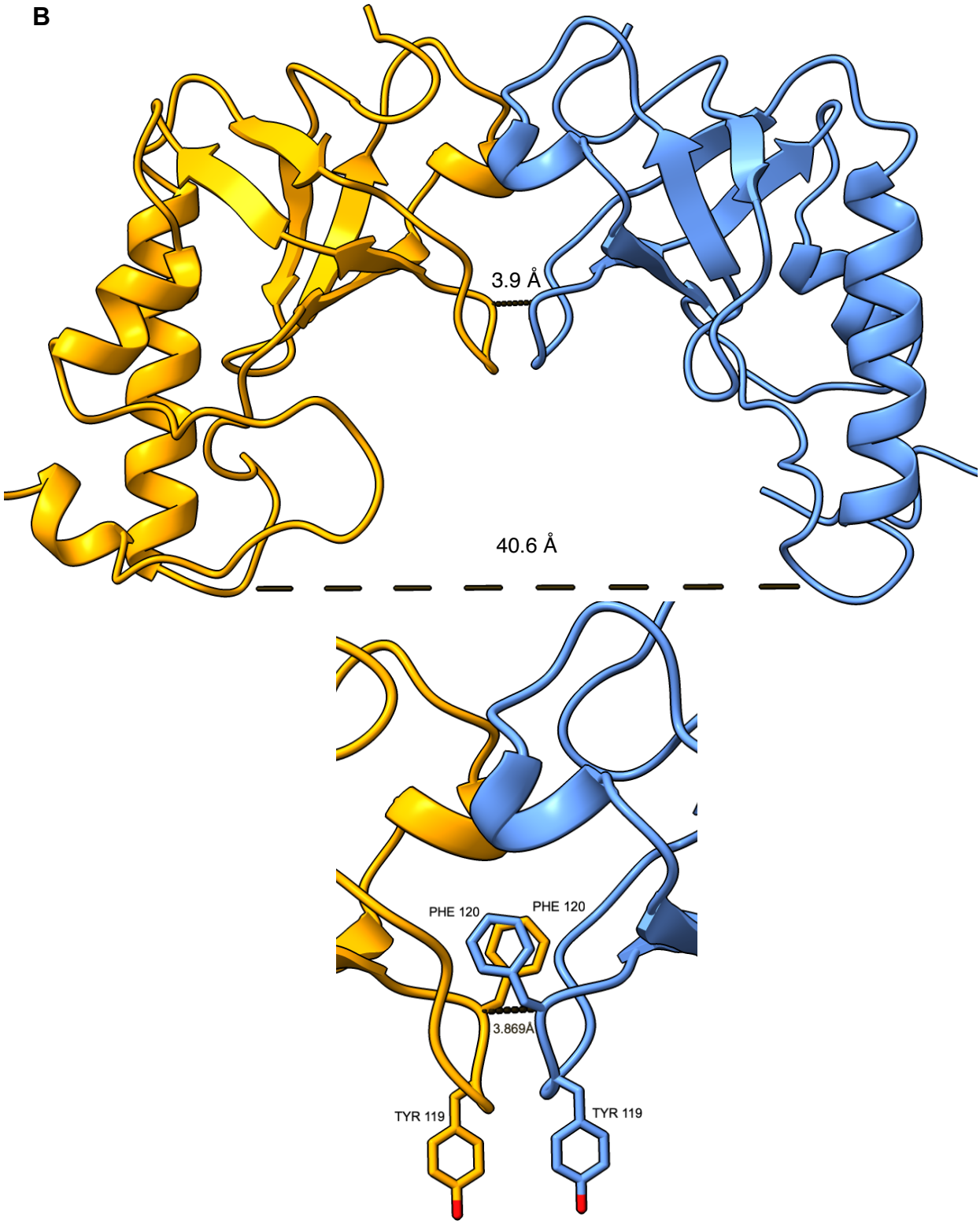


Figure 5.15: Bd1431$_{F120D}$ dimer interface
A – The repulsive forces of an aspartic acid residue at the dimer interface pushes the lower boundary of the dimer interface apart by ~5.5 Å causing an overall 10 Å increases in the distance between the bottom of the Bd1431 dimer.
B – Distances between wild-type Bd1431 monomers

Whilst initially surprising, the observation that Bd1431$_{F120D}$ appears monomeric in solution but dimeric in a crystal lattice could perhaps be explained by the nature of the condition in which Bd1431$_{F120D}$ was crystallised. Bd1431$_{F120D}$ crystallised in a condition containing 0.1 M BIS-Tris 5.5 and 3.0 M NaCl. The high salt concentration, which would increase further over the course of a vapour diffusion experiment, could potentially increase the strength of the hydrophobic packing at the interface to the point where the effect of the F120D mutation is trivial.

To assess the impact of salt concentration on oligomeric state, Bd1431$_{F120D}$ was loaded onto a HiLoad 10/300 superdex 75 analytical column in both low (150 mM NaCl) and high salt (1M NaCl) conditions. In low salt, Bd1431$_{F120D}$ behaved as previously observed, eluting at a volume consistent with monomeric Bd1431 (Figure 5.16). However, following overnight dialysis into a high salt buffer, a population of Bd1431$_{F120D}$ shifted to an elution volume consistent with wild-type dimeric Bd1431, the remainder of the sample was eluted as a monomer (Figure 5.16).

Taken together, these results suggest that the F120D point mutation successfully disrupts the dimer interface in solution and that the occurrence of dimeric Bd1431$_{F120D}$ in the crystal lattice is likely a result of the high salt concentration in which the protein crystallised. The successful disruption of the dimer interface demonstrates that the interface is biologically relevant and not an artefact of crystal packing. Notably, AF also folds Bd1431 confidently as a dimer and maintains the same dimer interface observed in the crystal structure of Bd1431 further strengthening the biological relevance of the

dimer interface. Importantly, the second interface that was observed in the crystal structure of Bd1431 is not modelled by AF.

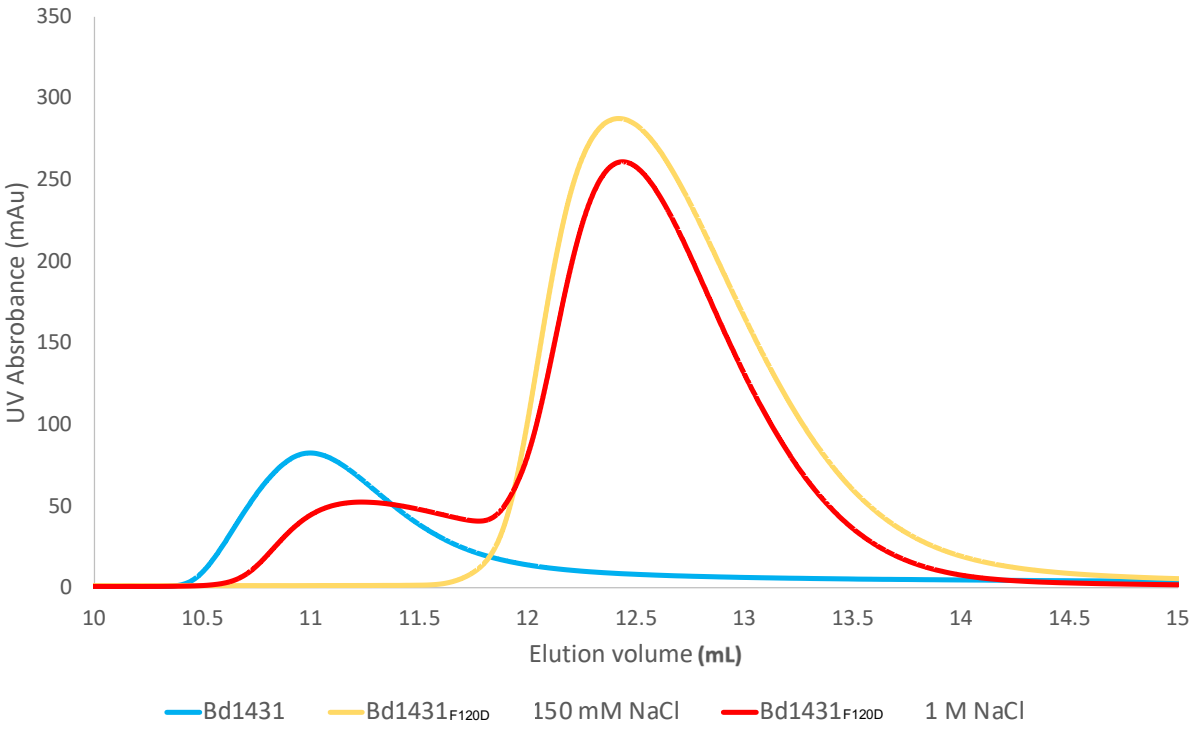


Figure 5.16: Comparison of SEC chromatograms between Bd1431$_{F120D}$ in high and low salt buffers
The SEC elution profiles of Bd1431$_{F120D}$ in high salt buffer (coloured red), low salt buffer (coloured yellow) and wild-type Bd1431 (coloured blue). In the presence of high salt concentrations, a population of Bd1431$_{F120D}$ shift towards the elution volume observed for wild-type Bd1431 indicating dimer formation.

5.2.6 Crystal structure of Bd1431-TDP complex

As previously stated, SNases are active against RNA and both single- and double-stranded DNA. However, the arrangement of chains in the Bd1431 dimer results in the active sites of both monomers being within close proximity. This raises questions as to whether both monomers are active against substrate at the same time.

To investigate ligand binding, Bd1431 was co-crystallised in the presence of the nucleotide TDP. Bd1431-TDP crystals were grown by sitting-drop vapour diffusion. These crystals belong to the C 2 spacegroup and diffracted to a resolution of 2.04 Å. Matthews coefficient analysis indicated that eight copies of Bd1431 is the most likely composition of the ASU. Initial phases were estimated by MR using one copy of the Bd1431 crystal structure as a search model. Only seven copies of Bd1431 (denoted chain A–G) were placed in the ASU with a final LLG of 18280 and TFZ of 103.7, indicating confident placement of the search model. The initial electron density map was of high quality and clear density was visible for TDP (Figure 5.17). Six of the ASU copies formed dimers across a non-crystallographic two-fold axis. The remaining copy was placed facing an area of weak electron density. Following successive rounds of model building and refinement, the final (eighth) copy of Bd1431 (chain H) could be placed by superimposition of a Bd1431 dimer over the unpaired copy of Bd1431. Refinement with this superimposed copy confirmed the correct placement of the eighth copy, though the electron density remained relatively poor. The presence of the poorly defined chain resulted in refinement statistics that were slightly higher than expected for the resolution of the data. All copies of Bd1431 contained positive electron density for TDP (Figure 5.17) and further discussion of the Bd1431-TDP complex will focus on the dimer formed by chains A & B as a representative of all dimers in the ASU.

Table 5.4. Data collection and refinement statistics for the Bd1431-TDP complex

Collection Statistics	
Wavelength	0.9795 Å
Resolution range	46.27 – 2.04 Å (2.113 – 2.04 Å)
Space group	C 1 2 1
Unit cell	167.85 Å 119.90 Å 72.90 Å 90° 93.46° 90°
Total reflections	644545 (65490)
Unique reflections	91505 (9121)
Multiplicity	7.0 (7.2)
Completeness (%)	98.96 (91.06)
Mean I/sigma(I)	14.91 (0.33)
Wilson B-factor	53.40 Å^2
R-merge[a]	0.07431 (3.131)
R-meas[b]	0.08025 (3.375)
R-pim[c]	0.0301 (1.251)
CC$^1/_2$	0.999 (0.366)
Refinement Statistics	
Reflections used in refinement	91504
Reflections used for R-free	4512
R-work[d]	0.2161
R-free[e]	0.2602
Number of non-hydrogen atoms	8456
macromolecules	8327
Ligands	6
Solvent	123
Protein residues	1038
RMS(bonds)	0.020
RMS(angles)	1.69
Ramachandran favored (%)	95.79
Ramachandran allowed (%)	3.33
Ramachandran outliers (%)	0.88
Rotamer outliers (%)	5.29
Clashscore	25.20
Average B-factor	84.54 Å^2
Macromolecules	85.10 Å^2
Ligands	132.81 Å^2
Solvent	58.35 Å^2

Statistics for the highest-resolution shell are shown in parentheses.

a Rmerge = $\sum hkl \sum i |I_i(hkl) - \langle I(hkl)\rangle| / \sum hkl \sum i I_i(hkl)$, where $I_i(hkl)$ is the intensity of the ith measurement of a reflection with indexes hkl and $\langle I(hkl)\rangle$ is the statistically weighted average reflection intensity.

b Rmeas is the Rmerge adjusted by a factor of $n(n-1)^{1/2}$ to account for multiplicity.

c R-pim = $\sum hkl [1/(n-1)]^{1/2} \sum i |I_i(hkl) - \langle I(hkl)\rangle| / \sum hkl \sum_i I_i(hkl)$ where $I_i(hkl)$ is the intensity of the ith measurement of a reflection with indexes hkl and $\langle I(hkl)\rangle$ is the statistically weighted average reflection intensity.

d R-work = $\sum ||Fo| - |Fc|| / \sum |Fo|$ where Fo and Fc are the observed and calculated structure factor amplitudes, respectively.

e R-free is the R-work calculated with 5% of the reflections chosen at random and omitted from the refinement.

The structures of Bd1431 and the Bd1431-TDP complex are remarkably similar, sharing an RMSD of 0.431 Å (RMSD calculated across 131 atom pairs using ChimeraX), demonstrating that the major elements of secondary structure on the whole are unchanged. One molecule of TDP occupies the binding pocket of each monomer and the binding mechanism is the same as typically seen for SNases (Figure 5.18). The thymine ring packs into a cleft formed between loop 3–A and loop B–C. In the Bd1431-TDP complex Tyr118 and Lys119 form hydrogen bonds with the 3'-phosphate of TDP to stabilise ligand binding (Figure 5.19). The 5'-phosphate is buried deeper into the binding pocket and is held in place by hydrogen bonds from the side chains of Arg72 and Arg121 (Figure 5.19). Hydrogen bonds from Asp117 and Lys118 further stabilise Arg121. Upon nucleotide-binding in SNase, two tyrosine residues (Tyr113 and Tyr115) rotate to enclose the left side of the ligand binding pocket (Figure 5.19). In Bd1431, these two conserved tyrosine residues (Tyr144 and Tyr146) adopt the nucleotide bound conformation in both the apo- and ligand-bound structures. The calcium binding pocket is largely unchanged in the Bd1431-TDP complex, with the Ca^{2+} being hexacoordinated by the sidechain carboxylates of Asp52 & Asp77, the backbone carbonyl oxygen of Thr78, two water molecules and finally, through an ionic interaction with an oxygen from the 5'-phosphate. A water molecule bridges the side chain of Glu80 to the Ca^{2+} (Figure 5.19). Notably, no clear density is visible for a water molecule that could bridge Glu80 to the 5'-phosphate of TDP to facilitate nucleophilic attack.

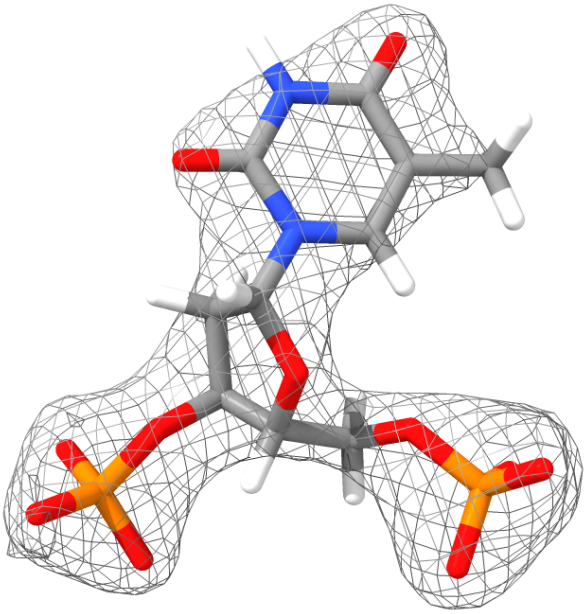


Figure 5.17: Electron density for TDP
Refined 2fo-fc electron density map of TDP shown in ChimeraX contoured to 1.5 σ.

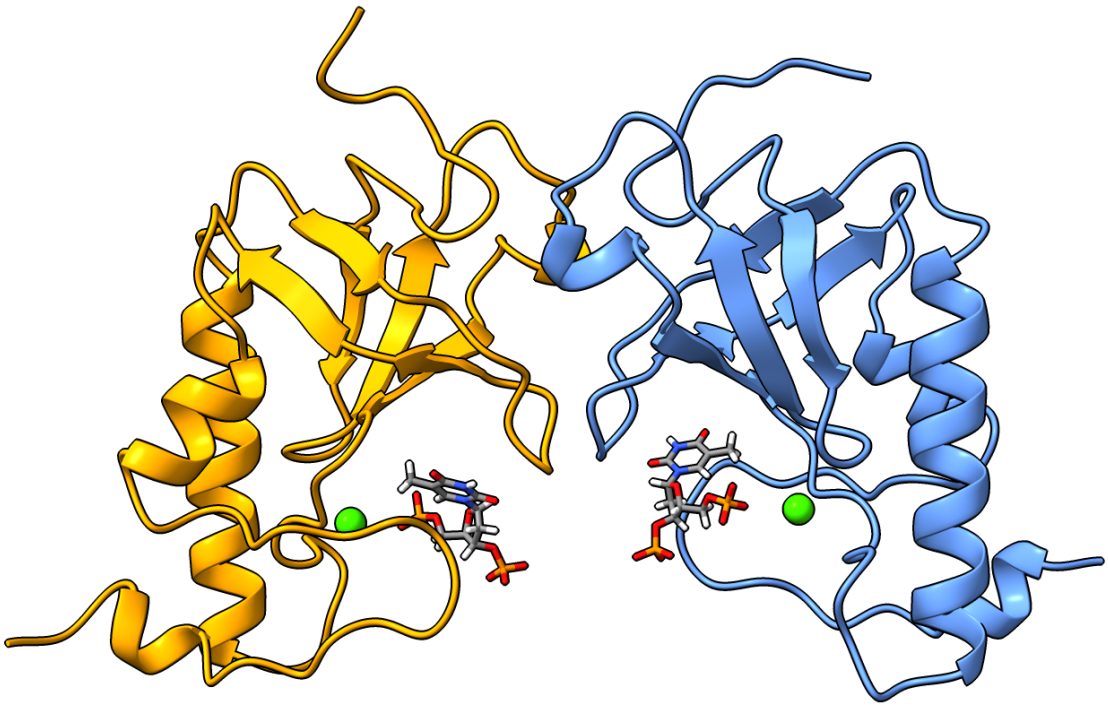


Figure 5.18: TDP binding to Bd1431 dimer
The nucleotide-binding pocket of each monomer of Bd1431 is occupied by one Ca^{2+} and one molecule of TDP.

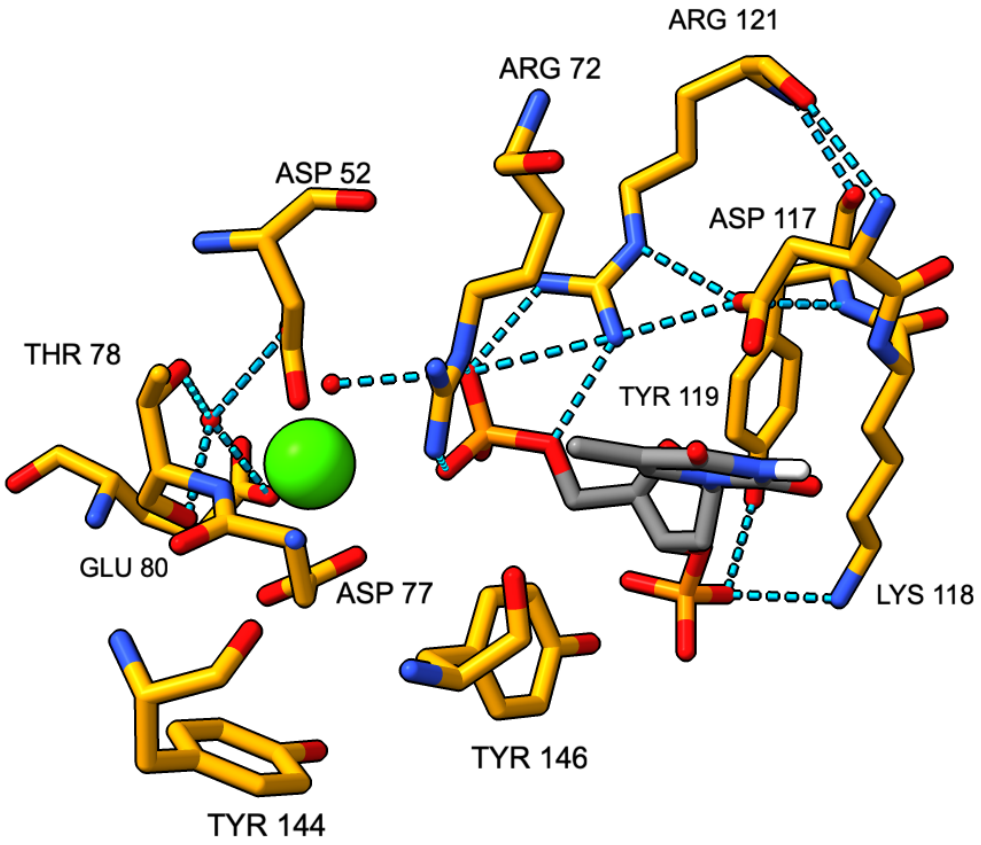


Figure 5.19: Hydrogen bonding at the ligand binding pocket
Bd1431-TDP interactions. Arg72 and Arg121 each form a pair of hydrogen bonds with oxygen atoms on the 5' phosphate. Lys118 and Tyr119 hydrogen bond with oxygen atoms on the 3' phosphate. Glu80 acts as a general base to deprotonate the water molecule that subsequently acts as a nucleophile. Residues of chain A are coloured orange; water molecules are shown as red circles; and hydrogen bonds are shown as dashed lines.

5.2.7 Bd1431 Activity assay

To assess nuclease activity, serial dilutions of Bd1431 were incubated with fixed concentrations of linearised double-stranded plasmid DNA for one hour before the reaction was stopped by the addition of EGTA. Over the duration of the assay, Bd1431 completely degraded the plasmid DNA when applied in excess (100:1 & 10:1) or at stoichiometric levels (Figure 5.20). Activity was maintained up until a hundred-fold excess of DNA; after which point, little activity was observed in the timeframe of the assay (Figure 5.20). An EGTA control experiment confirmed that Ca^{2+} is essential for Bd1431 activity (Figure 5.20).

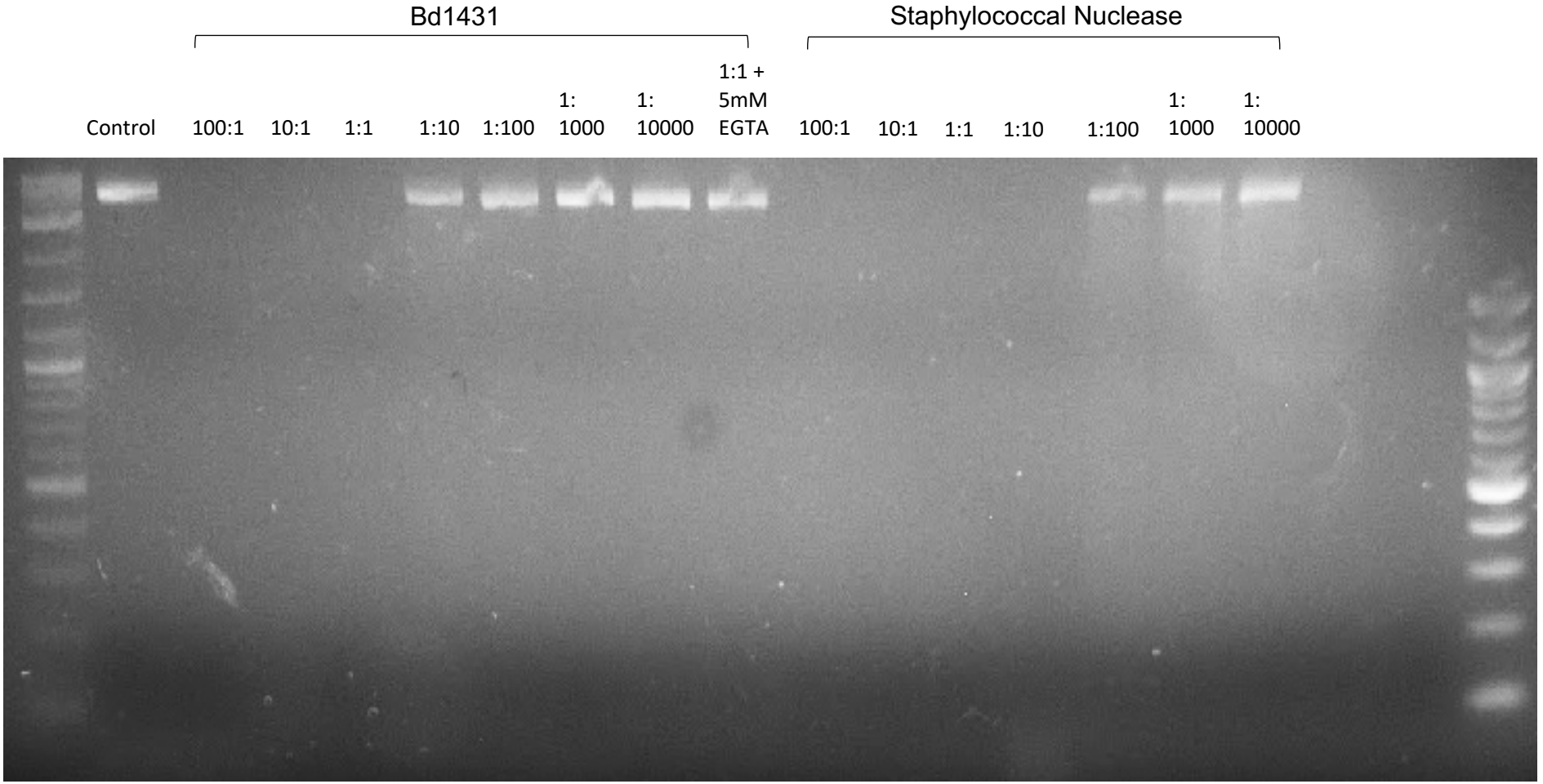


Figure 5.20: Bd1431 activity assay
Agarose gel of Bd1431 activity assay. Serial dilutions of Bd1431 were incubated with a fixed concentration (3.6 nm) of linearised plasmid DNA for one hour at 37 °C. The reaction was ended by the addition of 5mM EGTA. Activity was observed up until the 1:100 reaction. Molar ratio of Nuclease:DNA (Linear plasmid). Control experiments using commercial SNase were also included.

5.3 Discussion

5.3.1 Bd1431 exhibits a novel dimeric SNase fold

There are currently over 300 structures of SNases deposited in the PDB. Despite this extensive structural characterisation, none of the determined structures represent a biologically relevant dimer. The novelty of the Bd1431 dimer was examined using Topsearch[259] to search for similar interfaces amongst the known SNase structures. A search against both the ASU and biological assembly libraries confirmed that no SNase structure currently deposited in the PDB shares the same interface observed in Bd1431.

One structure of a dimeric SNase exists in the PDB that is a result of the extensive mutagenesis work that SNase was subjected to[260]. Whilst attempting to study the effects of mutations on the folding of SNase, it was observed that the deletion of residues 114-119 (located on loop B–C) resulted in the formation of a tightly associated SNase dimer[260]. The deletion of residues 114–119 resulted in the reorientation of helix C away from its native position. To stabilise the nuclease fold, helix C of another copy of SNase occupies the typical helix position and two monomers of SNase interact to form a helix swapped dimer (Figure 5.21)[260]. The helix swapping brings loops 4–5 of each monomer within proximity to form a second interface to stabilise the interaction. Whilst this structure provides interesting insight into the impact of mutation on oligomeric state, it bears no relevance to the evolution of the dimer interface observed in Bd1431.

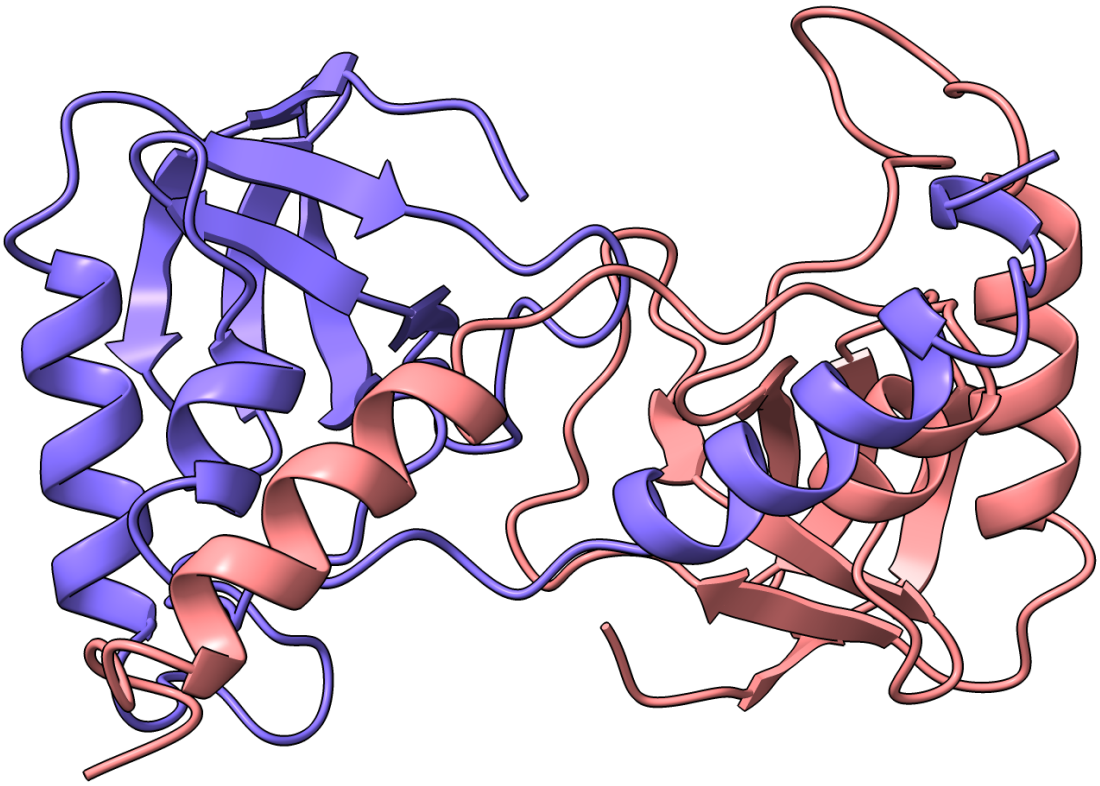


Figure 5.21: Helix swapped dimer of SNase$_{\Delta 114–119}$
The deletion of residues 114–119 of SNase results in the formation of a helix swapped dimer. Helix C of each monomer occupies its normal position on the opposite monomer creating a tightly associated dimer interface. PDB:1SND[260].

5.3.2 The insertion of hydrophobic amino acids to Loop 2–3 promotes dimerisation

The evolution of the oligomeric state of proteins likely stems from protein monomers acquiring chance mutations to surface residues, resulting in the formation of interfaces to mediate monomer-monomer interactions. Subsequent selection pressure towards oligomers that confer a fitness advantage likely drives the evolution of optimised interfaces[261].

In the case of Bd1431, the insertion of a stretch of hydrophobic amino acids (residues 59–66) into loop 2–3 creates a hydrophobic interface previously unseen in the SNase superfamily. A second substitution of a glycine residue on loop 4–5 to a phenylalanine further stabilises the interface. Attempts were made to produce a Bd1431 construct which contained a shortened loop 2–3, with the inserted hydrophobic residues being replaced with the sequence GSG to create a short turn; however, this construct did not produce a soluble protein. The impact of the deletion of loop 2–3 was instead assessed using AF[160] to model two copies of $Bd1431_{\Delta 59\text{-}66.}$ Under these circumstances, $Bd1431_{\Delta 59\text{-}66}$ was folded as two monomers with no interaction retained between each copy, strengthening the notion that these residues are important for dimerisation.

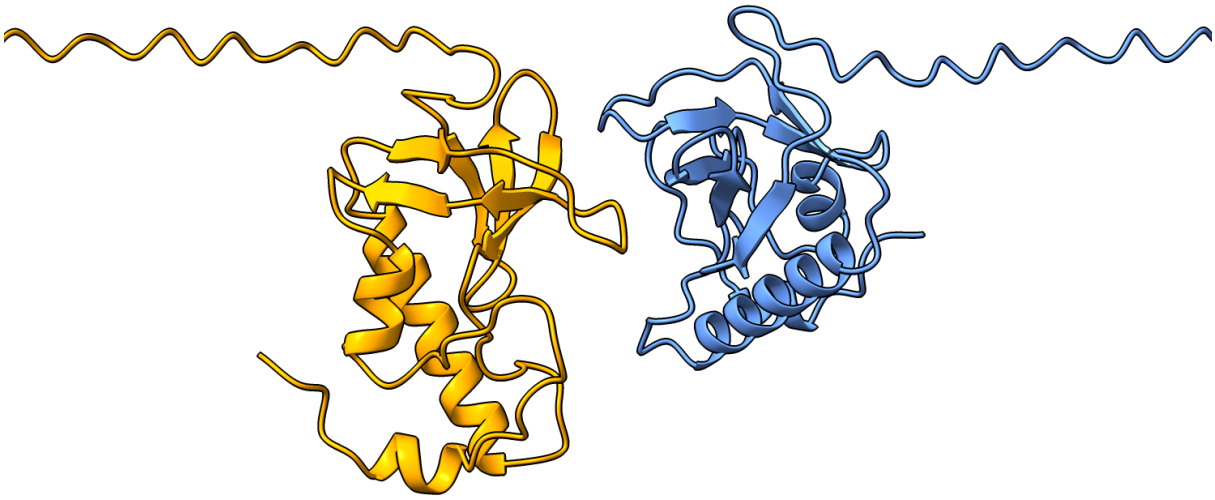
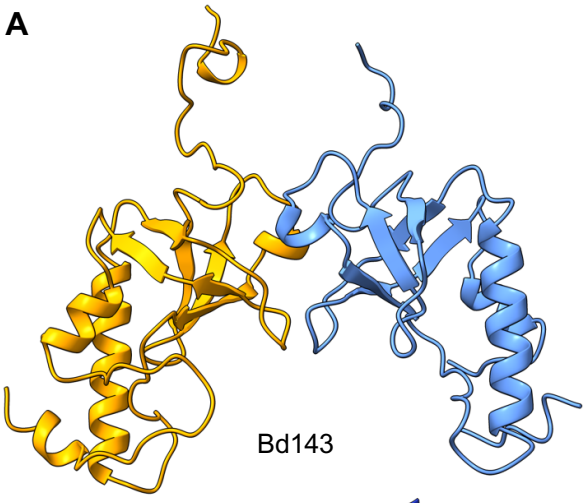


Figure 5.22: $Bd1431_{\Delta 59–66}$ AF model
The deletion of residues 59–66 prevents AF folding Bd1431 as a dimer, abolishing all interaction between Bd1431.

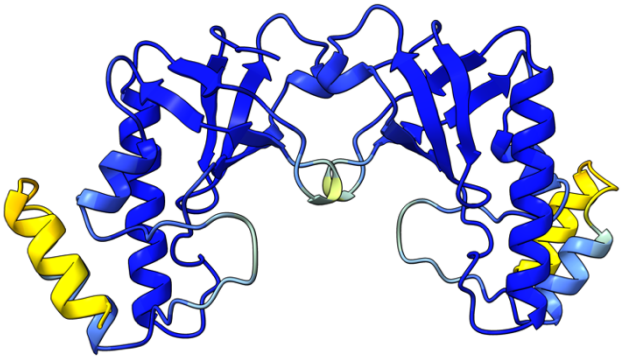
BLAST[204] analysis of the Bd1431 sequence against a non-redundant database identified SNases from a diverse range of bacteria that also possess a hydrophobic insertion in loop 2–3. Five of such proteins were arbitrarily chosen and were folded as

dimers using AF[160]. All five proteins were modelled as dimers and maintained the same dimer interface observed in Bd1431. Notably, in four of the five selected nucleases, the insertion of a hydrophobic region in loop 2–3 is also complemented by presence of a phenylalanine residue on loop 4–5. In the fifth sequence, this position is occupied by a tyrosine residue, suggesting that the presence of an aromatic residue on loop 4–5 is co-evolutionarily linked to the hydrophobic insertion. These findings demonstrate that dimeric SNases are likely widespread throughout nature but are yet to be fully characterised.

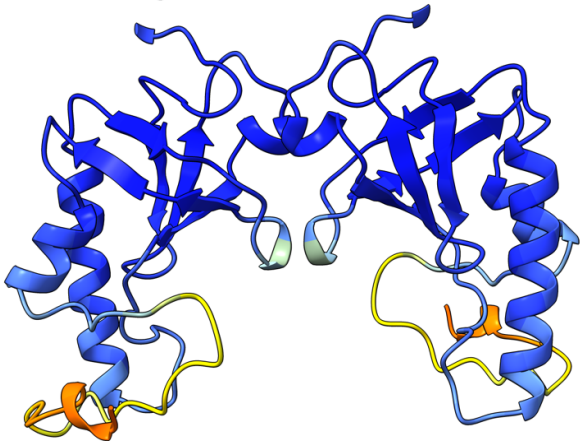
A



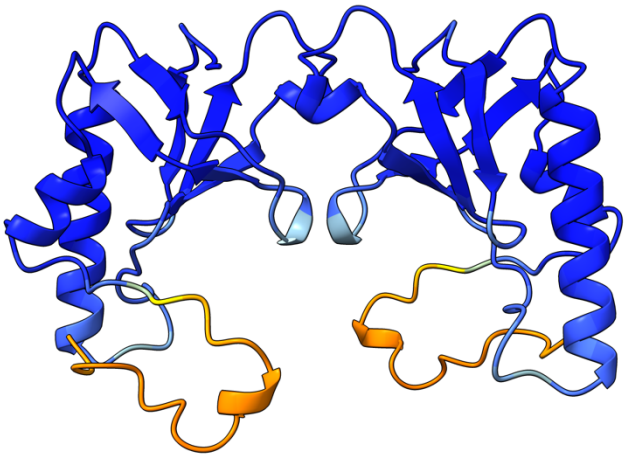
Bd143



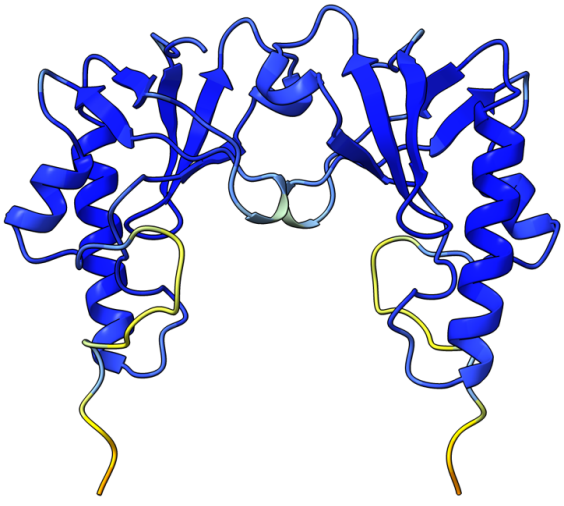
MCI5132304.1 MAG: thermonuclease family protein [Candidatus Electrothrix sp. EH2]
Sequence Identity to Bd1431: 44%



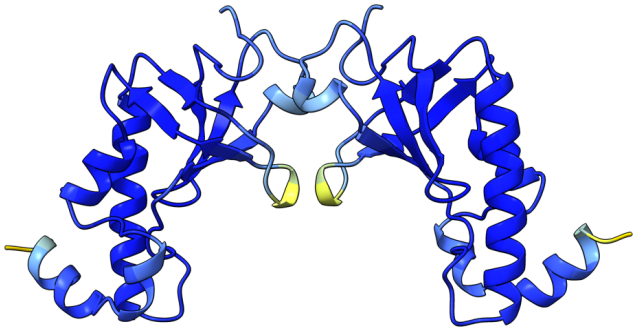
MBC76035.1 MAG: nuclease [Halobacteriovoraceae bacterium]
Sequence Identity to Bd1431: 54%



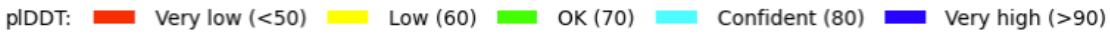
MBF0620955.1 MAG: thermonuclease family protein [Magnetococcales bacterium]
Sequence Identity to Bd1431: 36%



MCK5017954.1 MAG: thermonuclease family protein [Candidatus Peribacteraceae bacterium] Sequence Identity to Bd1431: 38%



MBL6621331.1 MAG: thermonuclease family protein [Rickettsiales bacterium]
Sequence Identity to Bd1431: 34%



B

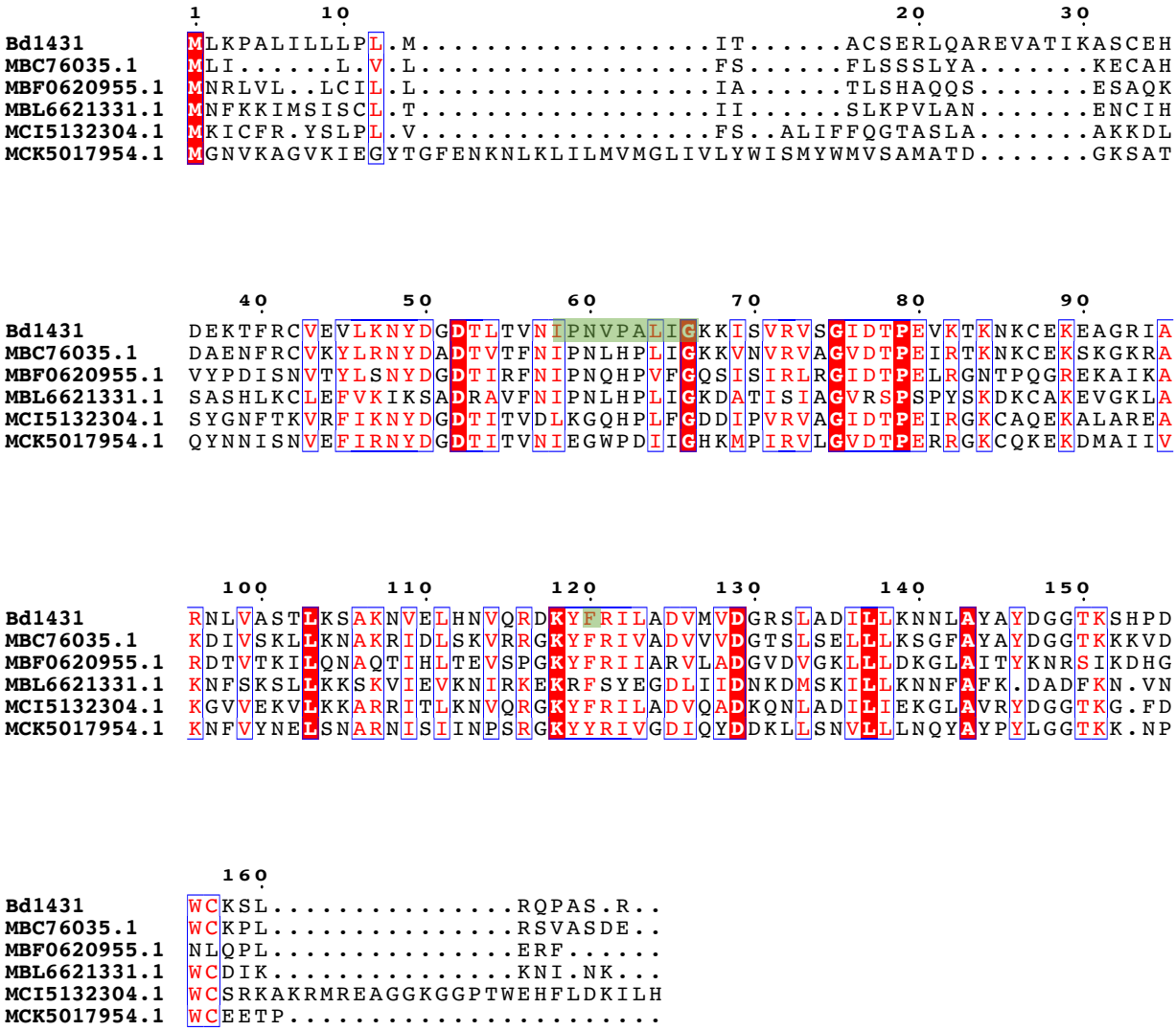


Figure 5.23: SNase dimer

A – AF models of SNases identified by BLAST search. The models are coloured by confidence (plDDT). In each case, AF folds the SNase as a dimer, with the interaction mediated by the same interface observed in Bd1431.

B – Sequence alignment of Bd1431 and the five selected SNases. For Bd1431, the inserted stretch of hydrophobic residues on loop 2–3 and Phe120 on loop 4–5 are highlighted green. Alignment made with Espript3.0[171]. Colour code: Red background, white text = strict identity, white background; red text = similarity in a group; blue frame = similarity across groups.

5.3.3 Bd1431 is a catalytically active nuclease

Activity assays demonstrate that Bd1431 is active against double stranded plasmid DNA. In addition to this, the crystal structure of the Bd1431-TDP complex revealed that both Bd1431 monomers competently bound a single copy of TDP. This suggests that both monomers would be simultaneously capable of catalytic activity. However, crude superimposition of double-stranded DNA to the TDP molecule bound in the Bd1431 active site reveals considerable clashes between the double helix and the dimer interface, questioning whether both monomers are capable of binding substrates at the same time. This also raises the question as to whether the dimer is retained in the presence of a nucleic acid that is larger than a single nucleotide. Although, the lack of major conformational change seen in the Bd1431-TDP complex would suggest that the splitting of the Bd1431 dimer, even in the presence of larger substrates, would be unlikely.

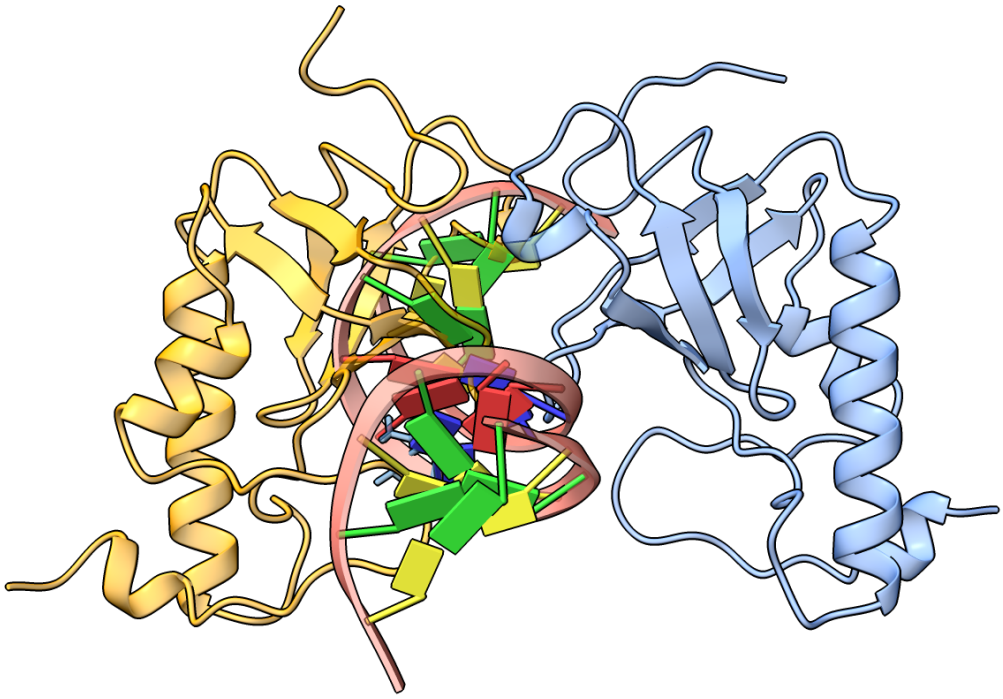


Figure 5.24: Double stranded DNA aligned to TDP
A thymine base from double stranded DNA (PDB: 1BNA) was aligned to the position of TDP in the Bd1431-TDP structure. The positioning of the DNA double helix demonstrates considerable clashes between both monomers of Bd1431.

5.3.4 Potential role for Bd1431 in *B. bacteriovorus*

Within *B. bacteriovorus* cells, Bd1431 is expressed as a lipoprotein and contains a signal peptide that possesses none of the conventional Lol avoidance signals, suggesting that Bd1431 localises to the OM. Although, it is unclear as to whether Bd1431 would be localised on the inner or outer leaflet. However, it is worth noting that the full complement of genes encoding for the Lol pathway have not been identified in *B. bacteriovorus*[262], indicating that differences in the regulation of lipoprotein localisation may occur. Despite this, localisation to the OM would appear to be most consistent with protein sequence features. Additionally, in *Bacteroides* it has been observed that the presence of acidic residues in the six amino acids following the conserved cysteine residue results in the surface localisation of lipoproteins[263]. Bd1431 contains a glutamic residue at the +2 position, although whether this system exists in *B. bacteriovorus* remains to be seen. The upregulation of Bd1431 following the deletion of Bd0934 and Bd3507[27] is indicative of compensatory activity. In order for Bd1431 to compensate for the activity of secreted nucleases, it would likely have to be located at the OM, presumably on the outer leaflet in order to gain access to prey nucleic acids. If this were the case, the dimerisation of Bd1431 could occur for increased stability of a surface-exposed enzyme. Whilst this may seem a compelling scenario, it remains to be experimentally validated.

5.3.5 Future Studies

Moving forward, a key step towards underpinning the function of Bd1431 will be to determine its cellular localisation. Tracking fluorescently-labelled Bd1431 during predation would likely provide key insights into its role in the predatory life cycle.

Secondly, Bd1431 has yet to be subjected to the same genetic characterisation as *B. bacteriovorus*'s other secreted nucleases and therefore, the generation of Δ*bd1431* strains will help uncover the function of Bd1431. The interaction between Bd1431 and its substrates can also be further characterised by crystallising Bd1431 in the presence of a biologically relevant substrate as opposed to TDP; such a study would be a step towards understanding how the SNase fold interacts with its substrate.

(1) Pérez, J.; Contreras-Moreno, F. J.; Marcos-Torres, F. J.; Moraleda-Muñoz, A.; Muñoz-Dorado, J. The Antibiotic Crisis: How Bacterial Predators Can Help. *Computational and Structural Biotechnology Journal* **2020**, *18*, 2547–2555. https://doi.org/10.1016/j.csbj.2020.09.010.
(2) Bratanis, E.; Andersson, T.; Lood, R.; Bukowska-Faniband, E. Biotechnological Potential of Bdellovibrio and Like Organisms and Their Secreted Enzymes. *Frontiers in Microbiology* **2020**, *11*.
(3) Pérez, J.; Moraleda-Muñoz, A.; Marcos-Torres, F. J.; Muñoz-Dorado, J. Bacterial Predation: 75 Years and Counting! *Environmental Microbiology* **2016**, *18* (3), 766–779. https://doi.org/10.1111/1462-2920.13171.
(4) Seccareccia, I.; Kost, C.; Nett, M. Quantitative Analysis of Lysobacter Predation. *Appl Environ Microbiol* **2015**, *81* (20), 7098–7105. https://doi.org/10.1128/AEM.01781-15.
(5) Berleman, J. E.; Kirby, J. R. Deciphering the Hunting Strategy of a Bacterial Wolfpack. *FEMS Microbiol Rev* **2009**, *33* (5), 942–957. https://doi.org/10.1111/j.1574-6976.2009.00185.x.
(6) Thiery, S.; Kaimer, C. The Predation Strategy of Myxococcus Xanthus. *Front Microbiol* **2020**, *11*, 2. https://doi.org/10.3389/fmicb.2020.00002.
(7) Berleman, J. E.; Allen, S.; Danielewicz, M. A.; Remis, J. P.; Gorur, A.; Cunha, J.; Hadi, M. Z.; Zusman, D. R.; Northen, T. R.; Witkowska, H. E.; Auer, M. The Lethal Cargo of Myxococcus Xanthus Outer Membrane Vesicles. *Frontiers in Microbiology* **2014**, *5*.
(8) Bauer, A.; Forchhammer, K. Bacterial Predation on Cyanobacteria. *Microb Physiol* **2021**, 1–10. https://doi.org/10.1159/000516427.
(9) Shiratori, T.; Suzuki, S.; Kakizawa, Y.; Ishida, K. Phagocytosis-like Cell Engulfment by a Planctomycete Bacterium. *Nat Commun* **2019**, *10* (1), 5529. https://doi.org/10.1038/s41467-019-13499-2.
(10) Pasternak, Z.; Njagi, M.; Shani, Y.; Chanyi, R.; Rotem, O.; Lurie-Weinberger, M. N.; Koval, S.; Pietrokovski, S.; Gophna, U.; Jurkevitch, E. In and out: An Analysis of Epibiotic vs Periplasmic Bacterial Predators. *ISME J* **2014**, *8* (3), 625–635. https://doi.org/10.1038/ismej.2013.164.
(11) Nielsen, P. H.; Singleton, C. M. Parasitic Bacteria Control Foam Formation. *Nat Microbiol* **2021**, *6* (6), 701–702. https://doi.org/10.1038/s41564-021-00914-y.
(12) Kamada, S.; Wakabayashi, R.; Naganuma, T. Phylogenetic Revisit to a Review on Predatory Bacteria. *Microorganisms* **2023**, *11* (7), 1673. https://doi.org/10.3390/microorganisms11071673.
(13) CASIDA, L. E. Ensifer Adhaerens Gen. Nov., Sp. Nov.: A Bacterial Predator of Bacteria in Soil†. *International Journal of Systematic and Evolutionary Microbiology* **1982**, *32* (3), 339–345. https://doi.org/10.1099/00207713-32-3-339.
(14) Guerrero, R.; Pedrós-Alió, C.; Esteve, I.; Mas, J.; Chase, D.; Margulis, L. Predatory Prokaryotes: Predation and Primary Consumption Evolved in Bacteria. *Proc Natl Acad Sci U S A* **1986**, *83* (7), 2138–2142.
(15) Sockett, R. E. Predatory Lifestyle of Bdellovibrio Bacteriovorus. *Annu Rev Microbiol* **2009**, *63*, 523–539. https://doi.org/10.1146/annurev.micro.091208.073346.
(16) Rendulic, S.; Jagtap, P.; Rosinus, A.; Eppinger, M.; Baar, C.; Lanz, C.; Keller, H.; Lambert, C.; Evans, K. J.; Goesmann, A.; Meyer, F.; Sockett, R. E.; Schuster, S. C. A Predator Unmasked: Life Cycle of Bdellovibrio Bacteriovorus from a Genomic Perspective. *Science* **2004**, *303* (5658), 689–692. https://doi.org/10.1126/science.1093027.

(17) Cavallo, F. M.; Jordana, L.; Friedrich, A. W.; Glasner, C.; van Dijl, J. M. Bdellovibrio Bacteriovorus: A Potential 'Living Antibiotic' to Control Bacterial Pathogens. *Critical Reviews in Microbiology* **2021**, *47* (5), 630–646. https://doi.org/10.1080/1040841X.2021.1908956.
(18) Lambert, C.; Smith, M. C. M.; Sockett, R. E. A Novel Assay to Monitor Predator-Prey Interactions for Bdellovibrio Bacteriovorus 109 J Reveals a Role for Methyl-Accepting Chemotaxis Proteins in Predation. *Environ Microbiol* **2003**, *5* (2), 127–132. https://doi.org/10.1046/j.1462-2920.2003.00385.x.
(19) Caulton, S. G.; Lovering, A. L. Bacterial Invasion and Killing by Predatory Bdellovibrio Primed by Predator Prey Cell Recognition and Self Protection. *Curr Opin Microbiol* **2020**, *56*, 74–80. https://doi.org/10.1016/j.mib.2020.07.002.
(20) Borgnia, M. J.; Subramaniam, S.; Milne, J. L. S. Three-Dimensional Imaging of the Highly Bent Architecture of Bdellovibrio Bacteriovorus by Using Cryo-Electron Tomography. *J Bacteriol* **2008**, *190* (7), 2588–2596. https://doi.org/10.1128/JB.01538-07.
(21) Evans, K. J.; Lambert, C.; Sockett, R. E. Predation by Bdellovibrio Bacteriovorus HD100 Requires Type IV Pili. *J Bacteriol* **2007**, *189* (13), 4850–4859. https://doi.org/10.1128/JB.01942-06.
(22) Mahmoud, K. K.; Koval, S. F. Characterization of Type IV Pili in the Life Cycle of the Predator Bacterium Bdellovibrio. *Microbiology* **2010**, *156* (4), 1040–1051. https://doi.org/10.1099/mic.0.036137-0.
(23) Chanyi, R. M.; Koval, S. F. Role of Type IV Pili in Predation by Bdellovibrio Bacteriovorus. *PLoS One* **2014**, *9* (11), e113404. https://doi.org/10.1371/journal.pone.0113404.
(24) Kuru, E.; Lambert, C.; Rittichier, J.; Till, R.; Ducret, A.; Derouaux, A.; Gray, J.; Biboy, J.; Vollmer, W.; Van Nieuwenhze, M.; Brun, Y. V.; Sockett, R. E. Bi-Cellular Wall Modifications during Bdellovibrio Bacteriovorus Predation Include Pore-Formation and L,D-Transpeptidase Mediated Prey Strengthening. *Nat Microbiol* **2017**, *2* (12), 1648–1657. https://doi.org/10.1038/s41564-017-0029-y.
(25) Lerner, T. R.; Lovering, A. L.; Bui, N. K.; Uchida, K.; Aizawa, S.-I.; Vollmer, W.; Sockett, R. E. Specialized Peptidoglycan Hydrolases Sculpt the Intra-Bacterial Niche of Predatory Bdellovibrio and Increase Population Fitness. *PLOS Pathogens* **2012**, *8* (2), e1002524. https://doi.org/10.1371/journal.ppat.1002524.
(26) Banks, E. J.; Lambert, C.; Mason, S. S.; Tyson, J.; Radford, P. M.; McLaughlin, C.; Lovering, A. L.; Sockett, R. E. An MltA-Like Lytic Transglycosylase Secreted by Bdellovibrio Bacteriovorus Cleaves the Prey Septum during Predatory Invasion. *J Bacteriol 205* (4), e00475-22. https://doi.org/10.1128/jb.00475-22.
(27) Bukowska-Faniband, E.; Andersson, T.; Lood, R. Studies on Bd0934 and Bd3507, Two Secreted Nucleases from Bdellovibrio Bacteriovorus, Reveal Sequential Release of Nucleases during the Predatory Cycle. *J Bacteriol* **2020**, *202* (18), e00150-20. https://doi.org/10.1128/JB.00150-20.
(28) Makowski, Ł.; Trojanowski, D.; Till, R.; Lambert, C.; Lowry, R.; Sockett, R. E.; Zakrzewska-Czerwińska, J. Dynamics of Chromosome Replication and Its Relationship to Predatory Attack Lifestyles in Bdellovibrio Bacteriovorus. *Appl Environ Microbiol* **2019**, *85* (14), e00730-19. https://doi.org/10.1128/AEM.00730-19.
(29) Cotter, T. W.; Thomashow, M. F. Identification of a Bdellovibrio Bacteriovorus Genetic Locus, Hit, Associated with the Host-Independent Phenotype. *J Bacteriol* **1992**, *174* (19), 6018–6024.

(30) Munita, J. M.; Arias, C. A. Mechanisms of Antibiotic Resistance. *Microbiol Spectr* **2016**, *4* (2), 10.1128/microbiolspec.VMBF-0016–2015. https://doi.org/10.1128/microbiolspec.VMBF-0016-2015.
(31) Rawson, T. M.; Wilson, R. C.; O'Hare, D.; Herrero, P.; Kambugu, A.; Lamorde, M.; Ellington, M.; Georgiou, P.; Cass, A.; Hope, W. W.; Holmes, A. H. Optimizing Antimicrobial Use: Challenges, Advances and Opportunities. *Nat Rev Microbiol* **2021**, *19* (12), 747–758. https://doi.org/10.1038/s41579-021-00578-9.
(32) Prestinaci, F.; Pezzotti, P.; Pantosti, A. Antimicrobial Resistance: A Global Multifaceted Phenomenon. *Pathog Glob Health* **2015**, *109* (7), 309–318. https://doi.org/10.1179/2047773215Y.0000000030.
(33) Price, R. O'Neill Report on Antimicrobial Resistance: Funding for Antimicrobial Specialists Should Be Improved. *Eur J Hosp Pharm* **2016**, *23* (4), 245–247. https://doi.org/10.1136/ejhpharm-2016-001013.
(34) de Kraker, M. E. A.; Stewardson, A. J.; Harbarth, S. Will 10 Million People Die a Year Due to Antimicrobial Resistance by 2050? *PLoS Med* **2016**, *13* (11), e1002184. https://doi.org/10.1371/journal.pmed.1002184.
(35) Årdal, C.; Balasegaram, M.; Laxminarayan, R.; McAdams, D.; Outterson, K.; Rex, J. H.; Sumpradit, N. Antibiotic Development - Economic, Regulatory and Societal Challenges. *Nat Rev Microbiol* **2020**, *18* (5), 267–274. https://doi.org/10.1038/s41579-019-0293-3.
(36) Gupta, R.; Sharma, S. Role of Alternatives to Antibiotics in Mitigating the Antimicrobial Resistance Crisis. *Indian J Med Res* **2022**, *156* (3), 464–477. https://doi.org/10.4103/ijmr.IJMR_3514_20.
(37) Kumar, M.; Sarma, D. K.; Shubham, S.; Kumawat, M.; Verma, V.; Nina, P. B.; JP, D.; Kumar, S.; Singh, B.; Tiwari, R. R. Futuristic Non-Antibiotic Therapies to Combat Antibiotic Resistance: A Review. *Front Microbiol* **2021**, *12*, 609459. https://doi.org/10.3389/fmicb.2021.609459.
(38) Atterbury, R. J.; Tyson, J. Predatory Bacteria as Living Antibiotics – Where Are We Now? *Microbiology* **2021**, *167* (1). https://doi.org/10.1099/mic.0.001025.
(39) Santajit, S.; Indrawattana, N. Mechanisms of Antimicrobial Resistance in ESKAPE Pathogens. *Biomed Res Int* **2016**, *2016*, 2475067. https://doi.org/10.1155/2016/2475067.
(40) Sun, Y.; Ye, J.; Hou, Y.; Chen, H.; Cao, J.; Zhou, T. Predation Efficacy of Bdellovibrio Bacteriovorus on Multidrug-Resistant Clinical Pathogens and Their Corresponding Biofilms. *Jpn. J. Infect. Dis.* **2017**, *70* (5), 485–489. https://doi.org/10.7883/yoken.JJID.2016.405.
(41) Lambert, C.; Chang, C.-Y.; Capeness, M. J.; Sockett, R. E. The First Bite— Profiling the Predatosome in the Bacterial Pathogen Bdellovibrio. *PLoS One* **2010**, *5* (1), e8599. https://doi.org/10.1371/journal.pone.0008599.
(42) Shanks, R. M. Q.; Davra, V. R.; Romanowski, E. G.; Brothers, K. M.; Stella, N. A.; Godboley, D.; Kadouri, D. E. An Eye to a Kill: Using Predatory Bacteria to Control Gram-Negative Pathogens Associated with Ocular Infections. *PLOS ONE* **2013**, *8* (6), e66723. https://doi.org/10.1371/journal.pone.0066723.
(43) Dashiff, A.; Kadouri, D. E. Predation of Oral Pathogens by Bdellovibrio Bacteriovorus 109J. *Mol Oral Microbiol* **2011**, *26* (1), 19–34. https://doi.org/10.1111/j.2041-1014.2010.00592.x.
(44) Hobley, L.; Summers, J. K.; Till, R.; Milner, D. S.; Atterbury, R. J.; Stroud, A.; Capeness, M. J.; Gray, S.; Leidenroth, A.; Lambert, C.; Connerton, I.; Twycross, J.;

Baker, M.; Tyson, J.; Kreft, J.-U.; Sockett, R. E. Dual Predation by Bacteriophage and Bdellovibrio Bacteriovorus Can Eradicate Escherichia Coli Prey in Situations Where Single Predation Cannot. *J Bacteriol* **2020**, *202* (6). https://doi.org/10.1128/JB.00629-19.

(45) Im, H.; Choi, S. Y.; Son, S.; Mitchell, R. J. Combined Application of Bacterial Predation and Violacein to Kill Polymicrobial Pathogenic Communities. *Sci Rep* **2017**, *7*. https://doi.org/10.1038/s41598-017-14567-7.

(46) Findlay, J. S.; Flick-Smith, H. C.; Keyser, E.; Cooper, I. A.; Williamson, E. D.; Oyston, P. C. F. Predatory Bacteria Can Protect SKH-1 Mice from a Lethal Plague Challenge. *Sci Rep* **2019**, *9*. https://doi.org/10.1038/s41598-019-43467-1.

(47) Russo, R.; Kolesnikova, I.; Kim, T.; Gupta, S.; Pericleous, A.; Kadouri, D. E.; Connell, N. D. Susceptibility of Virulent Yersinia Pestis Bacteria to Predator Bacteria in the Lungs of Mice. *Microorganisms* **2018**, *7* (1). https://doi.org/10.3390/microorganisms7010002.

(48) Bonfiglio, G.; Neroni, B.; Radocchia, G.; Pompilio, A.; Mura, F.; Trancassini, M.; Di Bonaventura, G.; Pantanella, F.; Schippa, S. Growth Control of Adherent-Invasive Escherichia Coli (AIEC) by the Predator Bacteria Bdellovibrio Bacteriovorus: A New Therapeutic Approach for Crohn's Disease Patients. *Microorganisms* **2019**, *8* (1). https://doi.org/10.3390/microorganisms8010017.

(49) Baker, M.; Negus, D.; Raghunathan, D.; Radford, P.; Moore, C.; Clark, G.; Diggle, M.; Tyson, J.; Twycross, J.; Sockett, R. E. Measuring and Modelling the Response of Klebsiella Pneumoniae KPC Prey to Bdellovibrio Bacteriovorus Predation, in Human Serum and Defined Buffer. *Sci Rep* **2017**, *7*. https://doi.org/10.1038/s41598-017-08060-4.

(50) Shatzkes, K.; Singleton, E.; Tang, C.; Zuena, M.; Shukla, S.; Gupta, S.; Dharani, S.; Onyile, O.; Rinaggio, J.; Connell, N. D.; Kadouri, D. E. Predatory Bacteria Attenuate Klebsiella Pneumoniae Burden in Rat Lungs. *mBio* **2016**, *7* (6). https://doi.org/10.1128/mBio.01847-16.

(51) Willis, A. R.; Moore, C.; Mazon-Moya, M.; Krokowski, S.; Lambert, C.; Till, R.; Mostowy, S.; Sockett, R. E. Injections of Predatory Bacteria Work Alongside Host Immune Cells to Treat Shigella Infection in Zebrafish Larvae. *Curr Biol* **2016**, *26* (24), 3343–3351. https://doi.org/10.1016/j.cub.2016.09.067.

(52) Olanya, O. M.; Niemira, B. A.; Cassidy, J. M.; Boyd, G.; Uknalis, J. Pathogen Reduction by Predatory Bacteria and Survival of Bdellovibrio Bacteriovorus and Escherichia Coli on Produce and Buffer Treated with Low-Dose Gamma Radiation. *LWT* **2020**, 109630. https://doi.org/10.1016/j.lwt.2020.109630.

(53) Saxon, E. B.; Jackson, R. W.; Bhumbra, S.; Smith, T.; Sockett, R. E. Bdellovibrio Bacteriovorus HD100 Guards against Pseudomonas Tolaasii Brown-Blotch Lesions on the Surface of Post-Harvest Agaricus Bisporus Supermarket Mushrooms. *BMC Microbiol.* **2014**, *14*, 163. https://doi.org/10.1186/1471-2180-14-163.

(54) Shatzkes, K.; Chae, R.; Tang, C.; Ramirez, G. C.; Mukherjee, S.; Tsenova, L.; Connell, N. D.; Kadouri, D. E. Examining the Safety of Respiratory and Intravenous Inoculation of Bdellovibrio Bacteriovorus and Micavibrio Aeruginosavorus in a Mouse Model. *Scientific Reports* **2015**, *5* (1), 12899. https://doi.org/10.1038/srep12899.

(55) Wadhwa, N.; Berg, H. C. Bacterial Motility: Machinery and Mechanisms. *Nat Rev Microbiol* **2022**, *20* (3), 161–173. https://doi.org/10.1038/s41579-021-00626-4.

(56) Kearns, D. B. A Field Guide to Bacterial Swarming Motility. *Nat Rev Microbiol* **2010**, *8* (9), 634–644. https://doi.org/10.1038/nrmicro2405.

(57) Lambert, C.; Evans, K. J.; Till, R.; Hobley, L.; Capeness, M.; Rendulic, S.; Schuster, S. C.; Aizawa, S.-I.; Sockett, R. E. Characterizing the Flagellar Filament and the Role of Motility in Bacterial Prey-Penetration by Bdellovibrio Bacteriovorus. *Mol Microbiol* **2006**, *60* (2), 274–286. https://doi.org/10.1111/j.1365-2958.2006.05081.x.
(58) Lambert, C.; Fenton, A. K.; Hobley, L.; Sockett, R. E. Predatory Bdellovibrio Bacteria Use Gliding Motility To Scout for Prey on Surfaces ▿ . *J Bacteriol* **2011**, *193* (12), 3139–3141. https://doi.org/10.1128/JB.00224-11.
(59) Negus, D.; Moore, C.; Baker, M.; Raghunathan, D.; Tyson, J.; Sockett, R. E. Predator Versus Pathogen: How Does Predatory Bdellovibrio Bacteriovorus Interface with the Challenges of Killing Gram-Negative Pathogens in a Host Setting? *Annu Rev Microbiol* **2017**, *71*, 441–457. https://doi.org/10.1146/annurev-micro-090816-093618.
(60) Chaban, B.; Hughes, H. V.; Beeby, M. The Flagellum in Bacterial Pathogens: For Motility and a Whole Lot More. *Seminars in Cell & Developmental Biology* **2015**, *46*, 91–103. https://doi.org/10.1016/j.semcdb.2015.10.032.
(61) Morimoto, Y. V.; Minamino, T. Structure and Function of the Bi-Directional Bacterial Flagellar Motor. *Biomolecules* **2014**, *4* (1), 217–234. https://doi.org/10.3390/biom4010217.
(62) Santiveri, M.; Roa-Eguiara, A.; Kühne, C.; Wadhwa, N.; Hu, H.; Berg, H. C.; Erhardt, M.; Taylor, N. M. I. Structure and Function of Stator Units of the Bacterial Flagellar Motor. *Cell* **2020**, *183* (1), 244-257.e16. https://doi.org/10.1016/j.cell.2020.08.016.
(63) Minamino, T.; Imada, K. The Bacterial Flagellar Motor and Its Structural Diversity. *Trends Microbiol* **2015**, *23* (5), 267–274. https://doi.org/10.1016/j.tim.2014.12.011.
(64) Iida, Y.; Hobley, L.; Lambert, C.; Fenton, A. K.; Sockett, R. E.; Aizawa, S.-I. Roles of Multiple Flagellins in Flagellar Formation and Flagellar Growth Post Bdelloplast Lysis in Bdellovibrio Bacteriovorus. *Journal of Molecular Biology* **2009**, *394* (5), 1011–1021. https://doi.org/10.1016/j.jmb.2009.10.003.
(65) Kaplan, M.; Chang, Y.-W.; Oikonomou, C. M.; Nicolas, W. J.; Jewett, A. I.; Kreida, S.; Dutka, P.; Rettberg, L. A.; Maggi, S.; Jensen, G. J. Bdellovibrio Predation Cycle Characterized at Nanometre-Scale Resolution with Cryo-Electron Tomography. *Nat Microbiol* **2023**, *8* (7), 1267–1279. https://doi.org/10.1038/s41564-023-01401-2.
(66) Melville, S.; Craig, L. Type IV Pili in Gram-Positive Bacteria. *Microbiol Mol Biol Rev* **2013**, *77* (3), 323–341. https://doi.org/10.1128/MMBR.00063-12.
(67) Craig, L.; Forest, K. T.; Maier, B. Type IV Pili: Dynamics, Biophysics and Functional Consequences. *Nat Rev Microbiol* **2019**, *17* (7), 429–440. https://doi.org/10.1038/s41579-019-0195-4.
(68) Nan, B.; Zusman, D. R. Uncovering the Mystery of Gliding Motility in the Myxobacteria. *Annu Rev Genet* **2011**, *45*, 21–39. https://doi.org/10.1146/annurev-genet-110410-132547.
(69) Burchard, R. P. Gliding Motility Mutants of Myxococcus Xanthus. *J Bacteriol* **1970**, *104* (2), 940–947.
(70) Hodgkin, J.; Kaiser, D. Cell-to-Cell Stimulation of Movement in Nonmotile Mutants of Myxococcus. *Proc Natl Acad Sci U S A* **1977**, *74* (7), 2938–2942.
(71) Nan, B.; McBride, M. J.; Chen, J.; Zusman, D. R.; Oster, G. Bacteria That Glide with Helical Tracks. *Curr Biol* **2014**, *24* (4), R169–R173. https://doi.org/10.1016/j.cub.2013.12.034.
(72) Luciano, J.; Agrebi, R.; Le Gall, A. V.; Wartel, M.; Fiegna, F.; Ducret, A.; Brochier-Armanet, C.; Mignot, T. Emergence and Modular Evolution of a Novel Motility

Machinery in Bacteria. *PLoS Genet* **2011**, *7* (9), e1002268. https://doi.org/10.1371/journal.pgen.1002268.
(73) Mercier, R.; Mignot, T. Regulations Governing the Multicellular Lifestyle of Myxococcus Xanthus. *Current Opinion in Microbiology* **2016**, *34*, 104–110. https://doi.org/10.1016/j.mib.2016.08.009.
(74) Mauriello, E. M. F. How an Unusual Chemosensory System Forms Arrays on the Bacterial Nucleoid. *Biochemical Society Transactions* **2020**, *48* (2), 347–356. https://doi.org/10.1042/BST20180450.
(75) Bustamante, V. H.; Martínez-Flores, I.; Vlamakis, H. C.; Zusman, D. R. Analysis of the Frz Signal Transduction System of Myxococcus Xanthus Shows the Importance of the Conserved C-Terminal Region of the Cytoplasmic Chemoreceptor FrzCD in Sensing Signals. *Molecular Microbiology* **2004**, *53* (5), 1501–1513. https://doi.org/10.1111/j.1365-2958.2004.04221.x.
(76) Guiseppi, A.; Vicente, J. J.; Herrou, J.; Byrne, D.; Barneoud, A.; Moine, A.; Espinosa, L.; Basse, M.-J.; Molle, V.; Mignot, T.; Roche, P.; Mauriello, E. M. F. A Divergent CheW Confers Plasticity to Nucleoid-Associated Chemosensory Arrays. *PLOS Genetics* **2019**, *15* (12), e1008533. https://doi.org/10.1371/journal.pgen.1008533.
(77) Blackhart, B. D.; Zusman, D. R. "Frizzy" Genes of Myxococcus Xanthus Are Involved in Control of Frequency of Reversal of Gliding Motility. *Proceedings of the National Academy of Sciences* **1985**, *82* (24), 8767–8770. https://doi.org/10.1073/pnas.82.24.8767.
(78) Inclán, Y. F.; Vlamakis, H. C.; Zusman, D. R. FrzZ, a Dual CheY-like Response Regulator, Functions as an Output for the Frz Chemosensory Pathway of Myxococcus Xanthus. *Molecular Microbiology* **2007**, *65* (1), 90–102. https://doi.org/10.1111/j.1365-2958.2007.05774.x.
(79) Scott, A. E.; Simon, E.; Park, S. K.; Andrews, P.; Zusman, D. R. Site-Specific Receptor Methylation of FrzCD in Myxococcus Xanthus Is Controlled by a Tetra-Trico Peptide Repeat (TPR) Containing Regulatory Domain of the FrzF Methyltransferase. *Molecular Microbiology* **2008**, *69* (3), 724–735. https://doi.org/10.1111/j.1365-2958.2008.06323.x.
(80) Nan, B.; Bandaria, J. N.; Guo, K. Y.; Fan, X.; Moghtaderi, A.; Yildiz, A.; Zusman, D. R. The Polarity of Myxobacterial Gliding Is Regulated by Direct Interactions between the Gliding Motors and the Ras Homolog MglA. *Proceedings of the National Academy of Sciences* **2015**, *112* (2), E186–E193. https://doi.org/10.1073/pnas.1421073112.
(81) Bonner, P. J.; Shimkets, L. J. Phospholipid Directed Motility of Surface-Motile Bacteria. *Molecular Microbiology* **2006**, *61* (5), 1101–1109. https://doi.org/10.1111/j.1365-2958.2006.05314.x.
(82) Guzzo, M.; Agrebi, R.; Espinosa, L.; Baronian, G.; Molle, V.; Mauriello, E. M. F.; Brochier-Armanet, C.; Mignot, T. Evolution and Design Governing Signal Precision and Amplification in a Bacterial Chemosensory Pathway. *PLOS Genetics* **2015**, *11* (8), e1005460. https://doi.org/10.1371/journal.pgen.1005460.
(83) Dinet, C.; Mignot, T. Unorthodox Regulation of the MglA Ras-like GTPase Controlling Polarity in Myxococcus Xanthus. *FEBS Letters* **2023**, *597* (6), 850–864. https://doi.org/10.1002/1873-3468.14565.
(84) Mauriello, E. M. F.; Zusman, D. R. Polarity of Motility Systems in Myxococcus Xanthus. *Curr Opin Microbiol* **2007**, *10* (6), 624–629. https://doi.org/10.1016/j.mib.2007.09.012.

(85) Carreira, L. A. M.; Szadkowski, D.; Müller, F.; Søgaard-Andersen, L. Spatiotemporal Regulation of Switching Front–Rear Cell Polarity. *Current Opinion in Cell Biology* **2022**, *76*, 102076. https://doi.org/10.1016/j.ceb.2022.102076.
(86) Miertzschke, M.; Koerner, C.; Vetter, I. R.; Keilberg, D.; Hot, E.; Leonardy, S.; Søgaard-Andersen, L.; Wittinghofer, A. Structural Analysis of the Ras-like G Protein MglA and Its Cognate GAP MglB and Implications for Bacterial Polarity. *EMBO J* **2011**, *30* (20), 4185–4197. https://doi.org/10.1038/emboj.2011.291.
(87) Szadkowski, D.; Harms, A.; Carreira, L. A. M.; Wigbers, M.; Potapova, A.; Wuichet, K.; Keilberg, D.; Gerland, U.; Søgaard-Andersen, L. Spatial Control of the GTPase MglA by Localized RomR–RomX GEF and MglB GAP Activities Enables Myxococcus Xanthus Motility. *Nat Microbiol* **2019**, *4* (8), 1344–1355. https://doi.org/10.1038/s41564-019-0451-4.
(88) Baranwal, J.; Lhospice, S.; Kanade, M.; Chakraborty, S.; Gade, P. R.; Harne, S.; Herrou, J.; Mignot, T.; Gayathri, P. Allosteric Regulation of a Prokaryotic Small Ras-like GTPase Contributes to Cell Polarity Oscillations in Bacterial Motility. *PLOS Biology* **2019**, *17* (9), e3000459. https://doi.org/10.1371/journal.pbio.3000459.
(89) Mauriello, E. M. F.; Nan, B.; Zusman, D. R. AglZ Regulates Adventurous (A-) Motility in Myxococcus Xanthus through Its Interaction with the Cytoplasmic Receptor, FrzCD. *Mol Microbiol* **2009**, *72* (4), 964–977. https://doi.org/10.1111/j.1365-2958.2009.06697.x.
(90) Treuner-Lange, A.; Macia, E.; Guzzo, M.; Hot, E.; Faure, L. M.; Jakobczak, B.; Espinosa, L.; Alcor, D.; Ducret, A.; Keilberg, D.; Castaing, J. P.; Lacas Gervais, S.; Franco, M.; Søgaard-Andersen, L.; Mignot, T. The Small G-Protein MglA Connects to the MreB Actin Cytoskeleton at Bacterial Focal Adhesions. *J Cell Biol* **2015**, *210* (2), 243–256. https://doi.org/10.1083/jcb.201412047.
(91) Mignot, T.; Shaevitz, J. W.; Hartzell, P. L.; Zusman, D. R. Evidence That Focal Adhesion Complexes Power Bacterial Gliding Motility. *Science* **2007**, *315* (5813), 853–856. https://doi.org/10.1126/science.1137223.
(92) Youderian, P.; Burke, N.; White, D. J.; Hartzell, P. L. Identification of Genes Required for Adventurous Gliding Motility in Myxococcus Xanthus with the Transposable Element Mariner. *Molecular Microbiology* **2003**, *49* (2), 555–570. https://doi.org/10.1046/j.1365-2958.2003.03582.x.
(93) Yu, R.; Kaiser, D. Gliding Motility and Polarized Slime Secretion. *Mol Microbiol* **2007**, *63* (2), 454–467. https://doi.org/10.1111/j.1365-2958.2006.05536.x.
(94) Islam, S. T.; Mignot, T. The Mysterious Nature of Bacterial Surface (Gliding) Motility: A Focal Adhesion-Based Mechanism in Myxococcus Xanthus. *Seminars in Cell & Developmental Biology* **2015**, *46*, 143–154. https://doi.org/10.1016/j.semcdb.2015.10.033.
(95) Nan, B.; Mauriello, E. M. F.; Sun, I.-H.; Wong, A.; Zusman, D. R. A Multi-Protein Complex from Myxococcus Xanthus Required for Bacterial Gliding Motility. *Mol Microbiol* **2010**, *76* (6), 1539–1554. https://doi.org/10.1111/j.1365-2958.2010.07184.x.
(96) Kahnt, J.; Aguiluz, K.; Koch, J.; Treuner-Lange, A.; Konovalova, A.; Huntley, S.; Hoppert, M.; Søgaard-Andersen, L.; Hedderich, R. Profiling the Outer Membrane Proteome during Growth and Development of the Social Bacterium Myxococcus Xanthus by Selective Biotinylation and Analyses of Outer Membrane Vesicles. *J Proteome Res* **2010**, *9* (10), 5197–5208. https://doi.org/10.1021/pr1004983.
(97) Faure, L. M.; Fiche, J.-B.; Espinosa, L.; Ducret, A.; Anantharaman, V.; Luciano, J.; Lhospice, S.; Islam, S. T.; Tréguier, J.; Sotes, M.; Kuru, E.; Van Nieuwenhze, M. S.;

Brun, Y. V.; Théodoly, O.; Aravind, L.; Nollmann, M.; Mignot, T. The Mechanism of Force Transmission at Bacterial Focal Adhesion Complexes. *Nature* **2016**, *539* (7630), 530–535. https://doi.org/10.1038/nature20121.
(98) Sun, M.; Wartel, M.; Cascales, E.; Shaevitz, J. W.; Mignot, T. Motor-Driven Intracellular Transport Powers Bacterial Gliding Motility. *Proc Natl Acad Sci U S A* **2011**, *108* (18), 7559–7564. https://doi.org/10.1073/pnas.1101101108.
(99) Jakobczak, B.; Keilberg, D.; Wuichet, K.; Søgaard-Andersen, L. Contact- and Protein Transfer-Dependent Stimulation of Assembly of the Gliding Motility Machinery in Myxococcus Xanthus. *PLoS Genet* **2015**, *11* (7), e1005341. https://doi.org/10.1371/journal.pgen.1005341.
(100) Agrebi, R.; Wartel, M.; Brochier-Armanet, C.; Mignot, T. An Evolutionary Link between Capsular Biogenesis and Surface Motility in Bacteria. *Nat Rev Microbiol* **2015**, *13* (5), 318–326. https://doi.org/10.1038/nrmicro3431.
(101) Nan, B.; Chen, J.; Neu, J. C.; Berry, R. M.; Oster, G.; Zusman, D. R. Myxobacteria Gliding Motility Requires Cytoskeleton Rotation Powered by Proton Motive Force. *Proc Natl Acad Sci U S A* **2011**, *108* (6), 2498–2503. https://doi.org/10.1073/pnas.1018556108.
(102) Nan, B.; Chen, J.; Neu, J. C.; Berry, R. M.; Oster, G.; Zusman, D. R. Myxobacteria Gliding Motility Requires Cytoskeleton Rotation Powered by Proton Motive Force. *Proc Natl Acad Sci U S A* **2011**, *108* (6), 2498–2503. https://doi.org/10.1073/pnas.1018556108.
(103) Cerveny, L.; Straskova, A.; Dankova, V.; Hartlova, A.; Ceckova, M.; Staud, F.; Stulik, J. Tetratricopeptide Repeat Motifs in the World of Bacterial Pathogens: Role in Virulence Mechanisms. *Infect Immun* **2013**, *81* (3), 629–635. https://doi.org/10.1128/IAI.01035-12.
(104) Lamb, J. R.; Tugendreich, S.; Hieter, P. Tetratrico Peptide Repeat Interactions: To TPR or Not to TPR? *Trends Biochem Sci* **1995**, *20* (7), 257–259. https://doi.org/10.1016/s0968-0004(00)89037-4.
(105) Zeytuni, N.; Zarivach, R. Structural and Functional Discussion of the Tetra-Trico-Peptide Repeat, a Protein Interaction Module. *Structure* **2012**, *20* (3), 397–405. https://doi.org/10.1016/j.str.2012.01.006.
(106) Nan, B.; Mauriello, E. M. F.; Sun, I.-H.; Wong, A.; Zusman, D. R. A Multi-Protein Complex from Myxococcus Xanthus Required for Bacterial Gliding Motility. *Mol Microbiol* **2010**, *76* (6), 1539–1554. https://doi.org/10.1111/j.1365-2958.2010.07184.x.
(107) Celia, H.; Botos, I.; Ni, X.; Fox, T.; De Val, N.; Lloubes, R.; Jiang, J.; Buchanan, S. K. Cryo-EM Structure of the Bacterial Ton Motor Subcomplex ExbB–ExbD Provides Information on Structure and Stoichiometry. *Commun Biol* **2019**, *2* (1), 1–6. https://doi.org/10.1038/s42003-019-0604-2.
(108) Ratliff, A. C.; Buchanan, S. K.; Celia, H. The Ton Motor. *Front Microbiol* **2022**, *13*, 852955. https://doi.org/10.3389/fmicb.2022.852955.
(109) Kaushik, S.; He, H.; Dalbey, R. E. Bacterial Signal Peptides- Navigating the Journey of Proteins. *Frontiers in Physiology* **2022**, *13*.
(110) Juncker, A. S.; Willenbrock, H.; von Heijne, G.; Brunak, S.; Nielsen, H.; Krogh, A. Prediction of Lipoprotein Signal Peptides in Gram-Negative Bacteria. *Protein Sci* **2003**, *12* (8), 1652–1662.
(111) Zückert, W. R. Secretion of Bacterial Lipoproteins: Through the Cytoplasmic Membrane, the Periplasm and Beyond. *Biochimica et Biophysica Acta (BBA) -*

Molecular Cell Research **2014**, *1843* (8), 1509–1516. https://doi.org/10.1016/j.bbamcr.2014.04.022.
(112) Narita, S.-I.; Tokuda, H. Bacterial Lipoproteins; Biogenesis, Sorting and Quality Control. *Biochim Biophys Acta Mol Cell Biol Lipids* **2017**, *1862* (11), 1414–1423. https://doi.org/10.1016/j.bbalip.2016.11.009.
(113) Konovalova, A.; Silhavy, T. J. Outer Membrane Lipoprotein Biogenesis: Lol Is Not the End. *Philos Trans R Soc Lond B Biol Sci* **2015**, *370* (1679), 20150030. https://doi.org/10.1098/rstb.2015.0030.
(114) Bhat, S.; Zhu, X.; Patel, R. P.; Orlando, R.; Shimkets, L. J. Identification and Localization of Myxococcus Xanthus Porins and Lipoproteins. *PLoS ONE* **2011**, *6* (11), e27475. https://doi.org/10.1371/journal.pone.0027475.
(115) Nudleman, E.; Wall, D.; Kaiser, D. Cell-to-Cell Transfer of Bacterial Outer Membrane Lipoproteins. *Science* **2005**, *309* (5731), 125–127. https://doi.org/10.1126/science.1112440.
(116) Pathak, D. T.; Wei, X.; Bucuvalas, A.; Haft, D. H.; Gerloff, D. L.; Wall, D. Cell Contact–Dependent Outer Membrane Exchange in Myxobacteria: Genetic Determinants and Mechanism. *PLOS Genetics* **2012**, *8* (4), e1002626. https://doi.org/10.1371/journal.pgen.1002626.
(117) Nan, B.; Zusman, D. R. Uncovering the Mystery of Gliding Motility in the Myxobacteria. *Annu Rev Genet* **2011**, *45*, 21–39. https://doi.org/10.1146/annurev-genet-110410-132547.
(118) Burchard, R. P. Gliding Motility of Prokaryotes: Ultrastructure, Physiology, and Genetics. *Annu Rev Microbiol* **1981**, *35*, 497–529. https://doi.org/10.1146/annurev.mi.35.100181.002433.
(119) Hoiczyk, E.; Baumeister, W. The Junctional Pore Complex, a Prokaryotic Secretion Organelle, Is the Molecular Motor Underlying Gliding Motility in Cyanobacteria. *Curr Biol* **1998**, *8* (21), 1161–1168. https://doi.org/10.1016/s0960-9822(07)00487-3.
(120) Burchard, R. P. Trail Following by Gliding Bacteria. *J Bacteriol* **1982**, *152* (1), 495–501.
(121) Wolgemuth, C.; Hoiczyk, E.; Kaiser, D.; Oster, G. How Myxobacteria Glide. *Curr Biol* **2002**, *12* (5), 369–377. https://doi.org/10.1016/s0960-9822(02)00716-9.
(122) Sun, H.; Yang, Z.; Shi, W. Effect of Cellular Filamentation on Adventurous and Social Gliding Motility of Myxococcus Xanthus. *Proc Natl Acad Sci U S A* **1999**, *96* (26), 15178–15183.
(123) Heintzelman, M. B. Gliding Motility in Apicomplexan Parasites. *Seminars in Cell & Developmental Biology* **2015**, *46*, 135–142. https://doi.org/10.1016/j.semcdb.2015.09.020.
(124) Mauriello, E. M. F.; Mouhamar, F.; Nan, B.; Ducret, A.; Dai, D.; Zusman, D. R.; Mignot, T. Bacterial Motility Complexes Require the Actin-like Protein, MreB and the Ras Homologue, MglA. *EMBO J* **2010**, *29* (2), 315–326. https://doi.org/10.1038/emboj.2009.356.
(125) Jones, L. J.; Carballido-López, R.; Errington, J. Control of Cell Shape in Bacteria: Helical, Actin-like Filaments in Bacillus Subtilis. *Cell* **2001**, *104* (6), 913–922. https://doi.org/10.1016/s0092-8674(01)00287-2.
(126) Mauriello, E. M. F.; Mignot, T.; Yang, Z.; Zusman, D. R. Gliding Motility Revisited: How Do the Myxobacteria Move without Flagella? *Microbiol Mol Biol Rev* **2010**, *74* (2), 229–249. https://doi.org/10.1128/MMBR.00043-09.

(127) Mauriello, E. M. F.; Mouhamar, F.; Nan, B.; Ducret, A.; Dai, D.; Zusman, D. R.; Mignot, T. Bacterial Motility Complexes Require the Actin-like Protein, MreB and the Ras Homologue, MglA. *EMBO J* **2010**, *29* (2), 315–326. https://doi.org/10.1038/emboj.2009.356.
(128) Garner, E. C.; Bernard, R.; Wang, W.; Zhuang, X.; Rudner, D. Z.; Mitchison, T. Coupled, Circumferential Motions of the Cell Wall Synthesis Machinery and MreB Filaments in B. Subtilis. *Science* **2011**, *333* (6039), 222–225. https://doi.org/10.1126/science.1203285.
(129) Fu, G.; Bandaria, J. N.; Le Gall, A. V.; Fan, X.; Yildiz, A.; Mignot, T.; Zusman, D. R.; Nan, B. MotAB-like Machinery Drives the Movement of MreB Filaments during Bacterial Gliding Motility. *Proc Natl Acad Sci U S A* **2018**, *115* (10), 2484–2489. https://doi.org/10.1073/pnas.1716441115.
(130) Ducret, A.; Valignat, M.-P.; Mouhamar, F.; Mignot, T.; Theodoly, O. Wet-Surface–Enhanced Ellipsometric Contrast Microscopy Identifies Slime as a Major Adhesion Factor during Bacterial Surface Motility. *Proceedings of the National Academy of Sciences* **2012**, *109* (25), 10036–10041. https://doi.org/10.1073/pnas.1120979109.
(131) Smith, S. G. J.; Mahon, V.; Lambert, M. A.; Fagan, R. P. A Molecular Swiss Army Knife: OmpA Structure, Function and Expression. *FEMS Microbiology Letters* **2007**, *273* (1), 1–11. https://doi.org/10.1111/j.1574-6968.2007.00778.x.
(132) Lünsdorf, H.; Schairer, H. U. Frozen Motion of Gliding Bacteria Outlines Inherent Features of the Motility Apparatus. *Microbiology* **2001**, *147* (4), 939–947. https://doi.org/10.1099/00221287-147-4-939.
(133) Pelling, A. E.; Li, Y.; Shi, W.; Gimzewski, J. K. Nanoscale Visualization and Characterization of Myxococcus Xanthus Cells with Atomic Force Microscopy. *Proc Natl Acad Sci U S A* **2005**, *102* (18), 6484–6489. https://doi.org/10.1073/pnas.0501207102.
(134) Balagam, R.; Litwin, D. B.; Czerwinski, F.; Sun, M.; Kaplan, H. B.; Shaevitz, J. W.; Igoshin, O. A. Myxococcus Xanthus Gliding Motors Are Elastically Coupled to the Substrate as Predicted by the Focal Adhesion Model of Gliding Motility. *PLoS Comput Biol* **2014**, *10* (5), e1003619. https://doi.org/10.1371/journal.pcbi.1003619.
(135) Islam, S. T.; Jolivet, N. Y.; Cuzin, C.; Belgrave, A. M.; My, L.; Fleuchot, B.; Faure, L. M.; Mahanta, U.; Kezzo, A. A.; Saïdi, F.; Sharma, G.; Fiche, J.-B.; Bratton, B. P.; Herrou, J.; Nollmann, M.; Shaevitz, J. W.; Durand, E.; Mignot, T. Unmasking of the von Willebrand A-Domain Surface Adhesin CglB at Bacterial Focal Adhesions Mediates Myxobacterial Gliding Motility. *Sci Adv* **2023**, *9* (8), eabq0619. https://doi.org/10.1126/sciadv.abq0619.
(136) Medina, A. A.; Shanks, R. M.; Kadouri, D. E. Development of a Novel System for Isolating Genes Involved in Predator-Prey Interactions Using Host Independent Derivatives of Bdellovibrio Bacteriovorus 109J. *BMC Microbiology* **2008**, *8* (1), 33. https://doi.org/10.1186/1471-2180-8-33.
(137) Milner, D. S.; Till, R.; Cadby, I.; Lovering, A. L.; Basford, S. M.; Saxon, E. B.; Liddell, S.; Williams, L. E.; Sockett, R. E. Ras GTPase-Like Protein MglA, a Controller of Bacterial Social-Motility in Myxobacteria, Has Evolved to Control Bacterial Predation by Bdellovibrio. *PLoS Genet* **2014**, *10* (4), e1004253. https://doi.org/10.1371/journal.pgen.1004253.
(138) Duncan, M. C.; Gillette, R. K.; Maglasang, M. A.; Corn, E. A.; Tai, A. K.; Lazinski, D. W.; Shanks, R. M. Q.; Kadouri, D. E.; Camilli, A. High-Throughput Analysis of Gene

Function in the Bacterial Predator Bdellovibrio Bacteriovorus. *mBio* **2019**, *10* (3). https://doi.org/10.1128/mBio.01040-19.
(139) Hobley, L.; Fung, R. K. Y.; Lambert, C.; Harris, M. A. T. S.; Dabhi, J. M.; King, S. S.; Basford, S. M.; Uchida, K.; Till, R.; Ahmad, R.; Aizawa, S.-I.; Gomelsky, M.; Sockett, R. E. Discrete Cyclic Di-GMP-Dependent Control of Bacterial Predation versus Axenic Growth in Bdellovibrio Bacteriovorus. *PLOS Pathogens* **2012**, *8* (2), e1002493. https://doi.org/10.1371/journal.ppat.1002493.
(140) Pathak, D. T.; Wall, D. Identification of the cglC, cglD, cglE, and cglF Genes and Their Role in Cell Contact-Dependent Gliding Motility in Myxococcus Xanthus. *J Bacteriol* **2012**, *194* (8), 1940–1949. https://doi.org/10.1128/JB.00055-12.
(141) Lowry, R. C.; Hallberg, Z. F.; Till, R.; Simons, T. J.; Nottingham, R.; Want, F.; Sockett, R. E.; Hammond, M. C.; Lambert, C. Production of 3′,3′-cGAMP by a Bdellovibrio Bacteriovorus Promiscuous GGDEF Enzyme, Bd0367, Regulates Exit from Prey by Gliding Motility. *PLOS Genetics* **2022**, *18* (5), e1010164. https://doi.org/10.1371/journal.pgen.1010164.
(142) Römling, U.; Galperin, M. Y.; Gomelsky, M. Cyclic Di-GMP: The First 25 Years of a Universal Bacterial Second Messenger. *Microbiol Mol Biol Rev* **2013**, *77* (1), 1–52. https://doi.org/10.1128/MMBR.00043-12.
(143) *KEGG COMPOUND: C16463*. https://www.genome.jp/dbget-bin/www_bget?cpd:C16463 (accessed 2023-12-29).
(144) Jenal, U.; Reinders, A.; Lori, C. Cyclic Di-GMP: Second Messenger Extraordinaire. *Nat Rev Microbiol* **2017**, *15* (5), 271–284. https://doi.org/10.1038/nrmicro.2016.190.
(145) Schirmer, T.; Jenal, U. Structural and Mechanistic Determinants of C-Di-GMP Signalling. *Nat Rev Microbiol* **2009**, *7* (10), 724–735. https://doi.org/10.1038/nrmicro2203.
(146) Römling, U.; Amikam, D. Cyclic Di-GMP as a Second Messenger. *Current Opinion in Microbiology* **2006**, *9* (2), 218–228. https://doi.org/10.1016/j.mib.2006.02.010.
(147) Paul, R.; Abel, S.; Wassmann, P.; Beck, A.; Heerklotz, H.; Jenal, U. Activation of the Diguanylate Cyclase PleD by Phosphorylation-Mediated Dimerization. *J Biol Chem* **2007**, *282* (40), 29170–29177. https://doi.org/10.1074/jbc.M704702200.
(148) Christen, M.; Christen, B.; Folcher, M.; Schauerte, A.; Jenal, U. Identification and Characterization of a Cyclic Di-GMP-Specific Phosphodiesterase and Its Allosteric Control by GTP. *J Biol Chem* **2005**, *280* (35), 30829–30837. https://doi.org/10.1074/jbc.M504429200.
(149) Galperin, M. Y.; Chou, S.-H. Structural Conservation and Diversity of PilZ-Related Domains. *Journal of Bacteriology* **2020**, *202* (4), 10.1128/jb.00664-19. https://doi.org/10.1128/jb.00664-19.
(150) Llontop, E. E.; Cenens, W.; Favaro, D. C.; Sgro, G. G.; Salinas, R. K.; Guzzo, C. R.; Farah, C. S. The PilB-PilZ-FimX Regulatory Complex of the Type IV Pilus from Xanthomonas Citri. *PLOS Pathogens* **2021**, *17* (8), e1009808. https://doi.org/10.1371/journal.ppat.1009808.
(151) Alm, R. A.; Bodero, A. J.; Free, P. D.; Mattick, J. S. Identification of a Novel Gene, pilZ, Essential for Type 4 Fimbrial Biogenesis in Pseudomonas Aeruginosa. *J Bacteriol* **1996**, *178* (1), 46–53.
(152) Paul, K.; Nieto, V.; Carlquist, W. C.; Blair, D. F.; Harshey, R. M. The C-Di-GMP Binding Protein YcgR Controls Flagellar Motor Direction and Speed to Affect Chemotaxis by a "Backstop Brake" Mechanism. *Mol Cell* **2010**, *38* (1), 128–139. https://doi.org/10.1016/j.molcel.2010.03.001.

(153) Han, Q.; Wang, S.-F.; Qian, X.-X.; Guo, L.; Shi, Y.-F.; He, R.; Yuan, J.-H.; Hou, Y.-J.; Li, D.-F. Flagellar Brake Protein YcgR Interacts with Motor Proteins MotA and FliG to Regulate the Flagellar Rotation Speed and Direction. *Frontiers in Microbiology* **2023**, *14*.
(154) Benach, J.; Swaminathan, S. S.; Tamayo, R.; Handelman, S. K.; Folta-Stogniew, E.; Ramos, J. E.; Forouhar, F.; Neely, H.; Seetharaman, J.; Camilli, A.; Hunt, J. F. The Structural Basis of Cyclic Diguanylate Signal Transduction by PilZ Domains. *EMBO J* **2007**, *26* (24), 5153–5166. https://doi.org/10.1038/sj.emboj.7601918.
(155) Chou, S.-H.; Galperin, M. Y. Diversity of Cyclic Di-GMP-Binding Proteins and Mechanisms. *J Bacteriol* **2015**, *198* (1), 32–46. https://doi.org/10.1128/JB.00333-15.
(156) Rotem, O.; Nesper, J.; Borovok, I.; Gorovits, R.; Kolot, M.; Pasternak, Z.; Shin, I.; Glatter, T.; Pietrokovski, S.; Jenal, U.; Jurkevitch, E. An Extended Cyclic Di-GMP Network in the Predatory Bacterium Bdellovibrio Bacteriovorus. *J Bacteriol* **2015**, *198* (1), 127–137. https://doi.org/10.1128/JB.00422-15.
(157) Hallberg, Z. F.; Wang, X. C.; Wright, T. A.; Nan, B.; Ad, O.; Yeo, J.; Hammond, M. C. Hybrid Promiscuous (Hypr) GGDEF Enzymes Produce Cyclic AMP-GMP (3', 3'-cGAMP). *Proc Natl Acad Sci U S A* **2016**, *113* (7), 1790–1795. https://doi.org/10.1073/pnas.1515287113.
(158) *KEGG COMPOUND: C20640*. https://www.genome.jp/dbget-bin/www_bget?C20640 (accessed 2023-12-29).
(159) van den Ent, F.; Löwe, J. RF Cloning: A Restriction-Free Method for Inserting Target Genes into Plasmids. *Journal of Biochemical and Biophysical Methods* **2006**, *67* (1), 67–74. https://doi.org/10.1016/j.jbbm.2005.12.008.
(160) Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; Bridgland, A.; Meyer, C.; Kohl, S. A. A.; Ballard, A. J.; Cowie, A.; Romera-Paredes, B.; Nikolov, S.; Jain, R.; Adler, J.; Back, T.; Petersen, S.; Reiman, D.; Clancy, E.; Zielinski, M.; Steinegger, M.; Pacholska, M.; Berghammer, T.; Bodenstein, S.; Silver, D.; Vinyals, O.; Senior, A. W.; Kavukcuoglu, K.; Kohli, P.; Hassabis, D. Highly Accurate Protein Structure Prediction with AlphaFold. *Nature* **2021**, *596* (7873), 583–589. https://doi.org/10.1038/s41586-021-03819-2.
(161) Winter, G. Xia2: An Expert System for Macromolecular Crystallography Data Reduction. *Journal of Applied Crystallography* **2010**, *43* (1), 186–190. https://doi.org/10.1107/S0021889809045701.
(162) Winter, G.; Waterman, D. G.; Parkhurst, J. M.; Brewster, A. S.; Gildea, R. J.; Gerstel, M.; Fuentes-Montero, L.; Vollmar, M.; Michels-Clark, T.; Young, I. D.; Sauter, N. K.; Evans, G. DIALS: Implementation and Evaluation of a New Integration Package. *Acta Cryst D* **2018**, *74* (2), 85–97. https://doi.org/10.1107/S2059798317017235.
(163) McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J. Phaser Crystallographic Software. *J Appl Cryst* **2007**, *40* (4), 658–674. https://doi.org/10.1107/S0021889807021206.
(164) Vonrhein, C.; Flensburg, C.; Keller, P.; Sharff, A.; Smart, O.; Paciorek, W.; Womack, T.; Bricogne, G. Data Processing and Analysis with the *autoPROC* Toolbox. *Acta Crystallogr D Biol Crystallogr* **2011**, *67* (4), 293–302. https://doi.org/10.1107/S0907444911007773.
(165) *STARANISO anisotropy & Bayesian estimation server*. https://staraniso.globalphasing.org/cgi-bin/staraniso.cgi (accessed 2023-12-20).

(166) Winter, G.; McAuley, K. E. Automated Data Collection for Macromolecular Crystallography. *Methods* **2011**, *55* (1), 81–93. https://doi.org/10.1016/j.ymeth.2011.06.010.
(167) Monaco, S.; Gordon, E.; Bowler, M. W.; Delagenière, S.; Guijarro, M.; Spruce, D.; Svensson, O.; McSweeney, S. M.; McCarthy, A. A.; Leonard, G.; Nanao, M. H. Automatic Processing of Macromolecular Crystallography X-Ray Diffraction Data at the ESRF. *J Appl Crystallogr* **2013**, *46* (Pt 3), 804–810. https://doi.org/10.1107/S0021889813006195.
(168) Krissinel, E.; Henrick, K. Inference of Macromolecular Assemblies from Crystalline State. *J Mol Biol* **2007**, *372* (3), 774–797. https://doi.org/10.1016/j.jmb.2007.05.022.
(169) Krissinel, E. Crystal Contacts as Nature's Docking Solutions. *J Comput Chem* **2010**, *31* (1), 133–143. https://doi.org/10.1002/jcc.21303.
(170) Zhu, J.; Wen, W.; Zheng, Z.; Shang, Y.; Wei, Z.; Xiao, Z.; Pan, Z.; Du, Q.; Wang, W.; Zhang, M. LGN/mInsc and LGN/NuMA Complex Structures Suggest Distinct Functions in Asymmetric Cell Division for the Par3/mInsc/LGN and Gαi/LGN/NuMA Pathways. *Mol Cell* **2011**, *43* (3), 418–431. https://doi.org/10.1016/j.molcel.2011.07.011.
(171) Robert, X.; Gouet, P. Deciphering Key Features in Protein Structures with the New ENDscript Server. *Nucleic Acids Research* **2014**, *42* (W1), W320–W324. https://doi.org/10.1093/nar/gku316.
(172) Afonine, P. V.; Moriarty, N. W.; Mustyakimov, M.; Sobolev, O. V.; Terwilliger, T. C.; Turk, D.; Urzhumtsev, A.; Adams, P. D. FEM: Feature-Enhanced Map. *Acta Crystallogr D Biol Crystallogr* **2015**, *71* (Pt 3), 646–666. https://doi.org/10.1107/S1399004714028132.
(173) Main, E. R. G.; Xiong, Y.; Cocco, M. J.; D'Andrea, L.; Regan, L. Design of Stable Alpha-Helical Arrays from an Idealized TPR Motif. *Structure* **2003**, *11* (5), 497–508. https://doi.org/10.1016/s0969-2126(03)00076-5.
(174) Laskowski, R. A.; Jabłońska, J.; Pravda, L.; Vařeková, R. S.; Thornton, J. M. PDBsum: Structural Summaries of PDB Entries. *Protein Sci* **2018**, *27* (1), 129–134. https://doi.org/10.1002/pro.3289.
(175) Wartel, M.; Ducret, A.; Thutupalli, S.; Czerwinski, F.; Gall, A.-V. L.; Mauriello, E. M. F.; Bergam, P.; Brun, Y. V.; Shaevitz, J.; Mignot, T. A Versatile Class of Cell Surface Directional Motors Gives Rise to Gliding Motility and Sporulation in Myxococcus Xanthus. *PLOS Biology* **2013**, *11* (12), e1001728. https://doi.org/10.1371/journal.pbio.1001728.
(176) Müller, F. D.; Schink, C. W.; Hoiczyk, E.; Cserti, E.; Higgs, P. I. Spore Formation in Myxococcus Xanthus Is Tied to Cytoskeleton Functions and Polysaccharide Spore Coat Deposition. *Mol Microbiol* **2012**, *83* (3), 10.1111/j.1365-2958.2011.07944.x. https://doi.org/10.1111/j.1365-2958.2011.07944.x.
(177) Shultis, D. D.; Purdy, M. D.; Banchs, C. N.; Wiener, M. C. Outer Membrane Active Transport: Structure of the BtuB:TonB Complex. *Science* **2006**, *312* (5778), 1396–1399. https://doi.org/10.1126/science.1127694.
(178) Pawelek, P. D.; Croteau, N.; Ng-Thow-Hing, C.; Khursigara, C. M.; Moiseeva, N.; Allaire, M.; Coulton, J. W. Structure of TonB in Complex with FhuA, E. Coli Outer Membrane Receptor. *Science* **2006**, *312* (5778), 1399–1402. https://doi.org/10.1126/science.1128057.
(179) Sean Peacock, R.; Weljie, A. M.; Peter Howard, S.; Price, F. D.; Vogel, H. J. The Solution Structure of the C-Terminal Domain of TonB and Interaction Studies with

TonB Box Peptides. *J Mol Biol* **2005**, *345* (5), 1185–1197. https://doi.org/10.1016/j.jmb.2004.11.026.
(180) Virtanen, S. I.; Kiirikki, A. M.; Mikula, K. M.; Iwaï, H.; Ollila, O. H. S. Heterogeneous Dynamics in Partially Disordered Proteins. *Phys. Chem. Chem. Phys.* **2020**, *22* (37), 21185–21196. https://doi.org/10.1039/D0CP03473H.
(181) Zinke, M.; Lejeune, M.; Mechaly, A.; Bardiaux, B.; Boneca, I. G.; Delepelaire, P.; Izadi-Pruneyre, N. *Ton Motor Conformational Switch and Peptidoglycan Role in Bacterial Nutrient Uptake*; preprint; Microbiology, 2023. https://doi.org/10.1101/2023.08.11.552980.
(182) Boucher, L. E.; Bosch, J. The Apicomplexan Glideosome and Adhesins -- Structures and Function. *J Struct Biol* **2015**, *190* (2), 93–114. https://doi.org/10.1016/j.jsb.2015.02.008.
(183) Song, G.; Springer, T. A. Structures of the Toxoplasma Gliding Motility Adhesin. *Proc Natl Acad Sci U S A* **2014**, *111* (13), 4862–4867. https://doi.org/10.1073/pnas.1403059111.
(184) Song, G.; Koksal, A. C.; Lu, C.; Springer, T. A. Shape Change in the Receptor for Gliding Motility in Plasmodium Sporozoites. *Proc Natl Acad Sci U S A* **2012**, *109* (52), 21420–21425. https://doi.org/10.1073/pnas.1218581109.
(185) Lünsdorf, H.; Reichenbach, H. Ultrastructural Details of the Apparatus of Gliding Motility of Myxococcus Fulvus (Myxobacterales). *Microbiology* **1989**, *135* (6), 1633–1641. https://doi.org/10.1099/00221287-135-6-1633.
(186) Freese, A.; Reichenbach, H.; Lünsdorf, H. Further Characterization and in Situ Localization of Chain-like Aggregates of the Gliding Bacteria Myxococcus Fulvus and Myxococcus Xanthus. *J Bacteriol* **1997**, *179* (4), 1246–1252. https://doi.org/10.1128/jb.179.4.1246-1252.1997.
(187) Nan, B.; McBride, M. J.; Chen, J.; Zusman, D. R.; Oster, G. Bacteria That Glide with Helical Tracks. *Curr Biol* **2014**, *24* (4), R169–R173. https://doi.org/10.1016/j.cub.2013.12.034.
(188) Kofler, M. M.; Freund, C. The GYF Domain. *The FEBS Journal* **2006**, *273* (2), 245–256. https://doi.org/10.1111/j.1742-4658.2005.05078.x.
(189) Attia, B.; Serrano, B.; Bornet, O.; Guerlesquin, F.; My, L.; Castaing, J.-P.; Mignot, T.; Elantak, L. 1H, 13C and 15N Chemical Shift Assignments of the ZnR and GYF Cytoplasmic Domains of the GltJ Protein from Myxococcus Xanthus. *Biomol NMR Assign* **2022**, *16* (2), 219–223. https://doi.org/10.1007/s12104-022-10083-6.
(190) Freund, C.; Dötsch, V.; Nishizawa, K.; Reinherz, E. L.; Wagner, G. The GYF Domain Is a Novel Structural Fold That Is Involved in Lymphoid Signaling through Proline-Rich Sequences. *Nat Struct Biol* **1999**, *6* (7), 656–660. https://doi.org/10.1038/10712.
(191) Freund, C.; Kühne, R.; Yang, H.; Park, S.; Reinherz, E. L.; Wagner, G. Dynamic Interaction of CD2 with the GYF and the SH3 Domain of Compartmentalized Effector Molecules. *EMBO J* **2002**, *21* (22), 5985–5995. https://doi.org/10.1093/emboj/cdf602.
(192) Mauriello, E. M. F.; Nan, B.; Zusman, D. R. AglZ Regulates Adventurous (A-) Motility in Myxococcus Xanthus through Its Interaction with the Cytoplasmic Receptor, FrzCD. *Mol Microbiol* **2009**, *72* (4), 964–977. https://doi.org/10.1111/j.1365-2958.2009.06697.x.
(193) Sathyamoorthy, R.; Kushmaro, Y.; Rotem, O.; Matan, O.; Kadouri, D. E.; Huppert, A.; Jurkevitch, E. To Hunt or to Rest: Prey Depletion Induces a Novel Starvation Survival Strategy in Bacterial Predators. *ISME J* **2021**, *15* (1), 109–123. https://doi.org/10.1038/s41396-020-00764-2.

(194) Galicia, C.; Lhospice, S.; Varela, P. F.; Trapani, S.; Zhang, W.; Navaza, J.; Herrou, J.; Mignot, T.; Cherfils, J. MglA Functions as a Three-State GTPase to Control Movement Reversals of Myxococcus Xanthus. *Nat Commun* **2019**, *10* (1), 5300. https://doi.org/10.1038/s41467-019-13274-3.
(195) Sei, S.; Mussio, J. K.; Yang, Q.; Nagashima, K.; Parchment, R. E.; Coffey, M. C.; Shoemaker, R. H.; Tomaszewski, J. E. Synergistic Antitumor Activity of Oncolytic Reovirus and Chemotherapeutic Agents in Non-Small Cell Lung Cancer Cells. *Molecular Cancer* **2009**, *8* (1), 47. https://doi.org/10.1186/1476-4598-8-47.
(196) Bautista, S.; Schmidt, V.; Guiseppi, A.; Mauriello, E. M. F.; Attia, B.; Elantak, L.; Mignot, T.; Mercier, R. FrzS Acts as a Polar Beacon to Recruit SgmX, a Central Activator of Type IV Pili during Myxococcus Xanthus Motility. *EMBO J* **2023**, *42* (1), e111661. https://doi.org/10.15252/embj.2022111661.
(197) Potapova, A.; Carreira, L. A. M.; Søgaard-Andersen, L. The Small GTPase MglA Together with the TPR Domain Protein SgmX Stimulates Type IV Pili Formation in M. Xanthus. *Proceedings of the National Academy of Sciences* **2020**, *117* (38), 23859–23868. https://doi.org/10.1073/pnas.2004722117.
(198) Skubák, P.; Araç, D.; Bowler, M. W.; Correia, A. R.; Hoelz, A.; Larsen, S.; Leonard, G. A.; McCarthy, A. A.; McSweeney, S.; Mueller-Dieckmann, C.; Otten, H.; Salzman, G.; Pannu, N. S. A New MR-SAD Algorithm for the Automatic Building of Protein Models from Low-Resolution X-Ray Data and a Poor Starting Model. *IUCrJ* **2018**, *5* (2), 166–171. https://doi.org/10.1107/S2052252517017961.
(199) Cowtan, K. The Buccaneer Software for Automated Model Building. 1. Tracing Protein Chains. *Acta Cryst D* **2006**, *62* (9), 1002–1011. https://doi.org/10.1107/S0907444906022116.
(200) Capeness, M. J.; Lambert, C.; Lovering, A. L.; Till, R.; Uchida, K.; Chaudhuri, R.; Alderwick, L. J.; Lee, D. J.; Swarbreck, D.; Liddell, S.; Aizawa, S.-I.; Sockett, R. E. Activity of Bdellovibrio Hit Locus Proteins, Bd0108 and Bd0109, Links Type IVa Pilus Extrusion/Retraction Status to Prey-Independent Growth Signalling. *PLoS One* **2013**, *8* (11), e79759. https://doi.org/10.1371/journal.pone.0079759.
(201) Azzazy, H. M. E.; Highsmith, W. E. Phage Display Technology: Clinical Applications and Recent Innovations. *Clinical Biochemistry* **2002**, *35* (6), 425–445. https://doi.org/10.1016/S0009-9120(02)00343-0.
(202) Qiang, X.; Sun, K.; Xing, L.; Xu, Y.; Wang, H.; Zhou, Z.; Zhang, J.; Zhang, F.; Caliskan, B.; Wang, M.; Qiu, Z. Discovery of a Polystyrene Binding Peptide Isolated from Phage Display Library and Its Application in Peptide Immobilization. *Sci Rep* **2017**, *7*, 2673. https://doi.org/10.1038/s41598-017-02891-x.
(203) Zhang, Y.; Franco, M.; Ducret, A.; Mignot, T. A Bacterial Ras-Like Small GTP-Binding Protein and Its Cognate GAP Establish a Dynamic Spatial Polarity Axis to Control Directed Motility. *PLOS Biology* **2010**, *8* (7), e1000430. https://doi.org/10.1371/journal.pbio.1000430.
(204) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic Local Alignment Search Tool. *J Mol Biol* **1990**, *215* (3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.
(205) Guzzo, C. R.; Dunger, G.; Salinas, R. K.; Farah, C. S. Structure of the PilZ–FimXEAL–c-Di-GMP Complex Responsible for the Regulation of Bacterial Type IV Pilus Biogenesis. *Journal of Molecular Biology* **2013**, *425* (12), 2174–2197. https://doi.org/10.1016/j.jmb.2013.03.021.

(206) Yang, F.; Tian, F.; Li, X.; Fan, S.; Chen, H.; Wu, M.; Yang, C.-H.; He, C. The Degenerate EAL-GGDEF Domain Protein Filp Functions as a Cyclic Di-GMP Receptor and Specifically Interacts with the PilZ-Domain Protein PXO_02715 to Regulate Virulence in Xanthomonas Oryzae Pv. Oryzae. *MPMI* **2014**, *27* (6), 578–589. https://doi.org/10.1094/MPMI-12-13-0371-R.
(207) Hobley, L.; Fung, R. K. Y.; Lambert, C.; Harris, M. A. T. S.; Dabhi, J. M.; King, S. S.; Basford, S. M.; Uchida, K.; Till, R.; Ahmad, R.; Aizawa, S.-I.; Gomelsky, M.; Sockett, R. E. Discrete Cyclic Di-GMP-Dependent Control of Bacterial Predation versus Axenic Growth in Bdellovibrio Bacteriovorus. *PLOS Pathogens* **2012**, *8* (2), e1002493. https://doi.org/10.1371/journal.ppat.1002493.
(208) Hallberg, Z. F.; Chan, C. H.; Wright, T. A.; Kranzusch, P. J.; Doxzen, K. W.; Park, J. J.; Bond, D. R.; Hammond, M. C. Structure and Mechanism of a Hypr GGDEF Enzyme That Activates cGAMP Signaling to Control Extracellular Metal Respiration. *Elife* **2019**, *8*, e43959. https://doi.org/10.7554/eLife.43959.
(209) Jenal, U.; Malone, J. Mechanisms of Cyclic-Di-GMP Signaling in Bacteria. *Annual Review of Genetics* **2006**, *40* (1), 385–407. https://doi.org/10.1146/annurev.genet.40.110405.090423.
(210) Cadby, I. T.; Basford, S. M.; Nottingham, R.; Meek, R.; Lowry, R.; Lambert, C.; Tridgett, M.; Till, R.; Ahmad, R.; Fung, R.; Hobley, L.; Hughes, W. S.; Moynihan, P. J.; Sockett, R. E.; Lovering, A. L. Nucleotide Signaling Pathway Convergence in a cAMP-Sensing Bacterial c-Di-GMP Phosphodiesterase. *EMBO J* **2019**, *38* (17), e100772. https://doi.org/10.15252/embj.2018100772.
(211) Wright, T. A.; Jiang, L.; Park, J. J.; Anderson, W. A.; Chen, G.; Hallberg, Z. F.; Nan, B.; Hammond, M. C. Second Messengers and Divergent HD-GYP Phosphodiesterases Regulate 3′,3′-cGAMP Signaling. *Molecular Microbiology* **2020**, *113* (1), 222–236. https://doi.org/10.1111/mmi.14412.
(212) Tudor, J. J.; Davis, J. J.; Panichella, M.; Zwolak, A. Isolation of Predation-Deficient Mutants of Bdellovibrio Bacteriovorus by Using Transposon Mutagenesis. *Appl Environ Microbiol* **2008**, *74* (17), 5436–5443. https://doi.org/10.1128/AEM.00256-08.
(213) Leonardy, S.; Miertzschke, M.; Bulyha, I.; Sperling, E.; Wittinghofer, A.; Søgaard-Andersen, L. Regulation of Dynamic Polarity Switching in Bacteria by a Ras-like G-Protein and Its Cognate GAP. *EMBO J* **2010**, *29* (14), 2276–2289. https://doi.org/10.1038/emboj.2010.114.
(214) Bos, J. L.; Rehmann, H.; Wittinghofer, A. GEFs and GAPs: Critical Elements in the Control of Small G Proteins. *Cell* **2007**, *129* (5), 865–877. https://doi.org/10.1016/j.cell.2007.05.018.
(215) Scheffzek, K.; Ahmadian, M. R. GTPase Activating Proteins: Structural and Functional Insights 18 Years after Discovery. *Cell Mol Life Sci* **2005**, *62* (24), 3014–3038. https://doi.org/10.1007/s00018-005-5136-x.
(216) Song, C.; Kumar, A.; Saleh, M. Bioinformatic Comparison of Bacterial Secretomes. *Genomics Proteomics Bioinformatics* **2009**, *7* (1–2), 37–46. https://doi.org/10.1016/S1672-0229(08)60031-5.
(217) Lambert, C.; Sockett, R. E. Nucleases in Bdellovibrio Bacteriovorus Contribute towards Efficient Self-Biofilm Formation and Eradication of Preformed Prey Biofilms. *FEMS Microbiol Lett* **2013**, *340* (2), 109–116. https://doi.org/10.1111/1574-6968.12075.

(218) Barabote, R. D.; Rendulic, S.; Schuster, S. C.; Saier, M. H. Comprehensive Analysis of Transport Proteins Encoded Within the Genome of Bdellovibrio Bacteriovorus. *Genomics* **2007**, *90* (4), 424–446. https://doi.org/10.1016/j.ygeno.2007.06.002.
(219) Hespell, R. B.; Miozzari, G. F.; Rittenberg, S. C. Ribonucleic Acid Destruction and Synthesis during Intraperiplasmic Growth of Bdellovibrio Bacteriovorus. *J Bacteriol* **1975**, *123* (2), 481–491. https://doi.org/10.1128/jb.123.2.481-491.1975.
(220) Matin, A.; Rittenberg, S. C. Kinetics of Deoxyribonucleic Acid Destruction and Synthesis During Growth of Bdellovibrio Bacteriovorus Strain 109D on Pseudomonas Putida and Escherichia Coli. *J Bacteriol* **1972**, *111* (3), 664–673.
(221) Rosson, R. A.; Rittenberg, S. C. Regulated Breakdown of Escherichia Coli Deoxyribonucleic Acid during Intraperiplasmic Growth of Bdellovibrio Bacteriovorus 109J. *J Bacteriol* **1979**, *140* (2), 620–633.
(222) Kadouri, D.; O'Toole, G. A. Susceptibility of Biofilms to Bdellovibrio Bacteriovorus Attack. *Appl Environ Microbiol* **2005**, *71* (7), 4044–4051. https://doi.org/10.1128/AEM.71.7.4044-4051.2005.
(223) Medina, A. A.; Kadouri, D. E. Biofilm Formation of Bdellovibrio Bacteriovorus Host-Independent Derivatives. *Res Microbiol* **2009**, *160* (3), 224–231. https://doi.org/10.1016/j.resmic.2009.02.001.
(224) Iebba, V.; Totino, V.; Santangelo, F.; Gagliardi, A.; Ciotoli, L.; Virga, A.; Ambrosi, C.; Pompili, M.; De Biase, R. V.; Selan, L.; Artini, M.; Pantanella, F.; Mura, F.; Passariello, C.; Nicoletti, M.; Nencioni, L.; Trancassini, M.; Quattrucci, S.; Schippa, S. Bdellovibrio Bacteriovorus Directly Attacks Pseudomonas Aeruginosa and Staphylococcus Aureus Cystic Fibrosis Isolates. *Frontiers in Microbiology* **2014**, *5*.
(225) Whitchurch, C. B.; Tolker-Nielsen, T.; Ragas, P. C.; Mattick, J. S. Extracellular DNA Required for Bacterial Biofilm Formation. *Science* **2002**, *295* (5559), 1487. https://doi.org/10.1126/science.295.5559.1487.
(226) Tang, J.; Zhou, R.; Shi, X.; Kang, M.; Wang, H.; Chen, H. Two Thermostable Nucleases Coexisted in Staphylococcus Aureus: Evidence from Mutagenesis and in Vitro Expression. *FEMS Microbiol Lett* **2008**, *284* (2), 176–183. https://doi.org/10.1111/j.1574-6968.2008.01194.x.
(227) Yang, W. Nucleases: Diversity of Structure, Function and Mechanism. *Q Rev Biophys* **2011**, *44* (1), 1–93. https://doi.org/10.1017/S0033583510000181.
(228) Cunningham, L.; Catlin, B. W.; de Garilhe, M. P. A Deoxyribonuclease of Micrococcus Pyogenes1. *J. Am. Chem. Soc.* **1956**, *78* (18), 4642–4645. https://doi.org/10.1021/ja01599a031.
(229) Cotton, F. A.; Bier, C. J.; Day, V. W.; Hazen, E. E.; Larsen, S. Some Aspects of the Structure of Staphylococcal Nuclease. I. Crystallographic Studies. *Cold Spring Harb Symp Quant Biol* **1972**, *36*, 243–249. https://doi.org/10.1101/sqb.1972.036.01.032.
(230) Calderon, R. O.; Stolowich, N. J.; Gerlt, J. A.; Sturtevant, J. M. Thermal Denaturation of Staphylococcal Nuclease. *Biochemistry* **1985**, *24* (22), 6044–6049. https://doi.org/10.1021/bi00343a004.
(231) Hynes, T. R.; Hodel, A.; Fox, R. O. Engineering Alternative Beta-Turn Types in Staphylococcal Nuclease. *Biochemistry* **1994**, *33* (17), 5021–5030. https://doi.org/10.1021/bi00183a004.
(232) Shortle, D.; Lin, B. Genetic Analysis of Staphylococcal Nuclease: Identification of Three Intragenic "Global" Suppressors of Nuclease-Minus Mutations. *Genetics* **1985**, *110* (4), 539–555.

(233) Truckses, D. M.; Somoza, J. R.; Prehoda, K. E.; Miller, S. C.; Markley, J. L. Coupling between Trans/Cis Proline Isomerization and Protein Stability in Staphylococcal Nuclease. *Protein Sci* **1996**, *5* (9), 1907–1916.
(234) Hynes, T. R.; Fox, R. O. The Crystal Structure of Staphylococcal Nuclease Refined at 1.7 Å Resolution. *Proteins: Structure, Function, and Bioinformatics* **1991**, *10* (2), 92–105. https://doi.org/10.1002/prot.340100203.
(235) Lister, J. L.; Horswill, A. R. Staphylococcus Aureus Biofilms: Recent Developments in Biofilm Dispersal. *Front Cell Infect Microbiol* **2014**, *4*, 178. https://doi.org/10.3389/fcimb.2014.00178.
(236) Alarcón, B.; Vicedo, B.; Aznar, R. PCR-based Procedures for Detection and Quantification of Staphylococcus Aureus and Their Application in Food. *Journal of Applied Microbiology* **2006**, *100* (2), 352–364. https://doi.org/10.1111/j.1365-2672.2005.02768.x.
(237) Brakstad, O. G.; Aasbakk, K.; Maeland, J. A. Detection of Staphylococcus Aureus by Polymerase Chain Reaction Amplification of the Nuc Gene. *Journal of Clinical Microbiology* **1992**, *30* (7), 1654–1660. https://doi.org/10.1128/jcm.30.7.1654-1660.1992.
(238) El-Sayed, A.; Alber, J.; Lämmler, C.; Bönner, B.; Huhn, A.; Kaleta, E. F.; Zschöck, M. PCR-Based Detection of Genes Encoding Virulence Determinants in Staphylococcus Aureus from Birds. *Journal of Veterinary Medicine, Series B* **2005**, *52* (1), 38–44. https://doi.org/10.1111/j.1439-0450.2004.00814.x.
(239) Hernandez, F. J.; Huang, L.; Olson, M. E.; Powers, K. M.; Hernandez, L. I.; Meyerholz, D. K.; Thedens, D. R.; Behlke, M. A.; Horswill, A. R.; McNamara, J. O. Non-Invasive Imaging of Staphylococcus Aureus Infections with a Nuclease-Activated Probe. *Nat Med* **2014**, *20* (3), 301–306. https://doi.org/10.1038/nm.3460.
(240) Palomares, C.; Torres, M. J.; Torres, A.; Aznar, J.; Palomares, J. C. Rapid Detection and Identification of Staphylococcus Aureus from Blood Culture Specimens Using Real-Time Fluorescence PCR. *Diagnostic Microbiology and Infectious Disease* **2003**, *45* (3), 183–189. https://doi.org/10.1016/S0732-8893(02)00542-4.
(241) Ramesh, A.; Padmapriya, B. P.; Chrashekar, A.; Varadaraj, M. C. Application of a Convenient DNA Extraction Method and Multiplex PCR for the Direct Detection of Staphylococcus Aureus and Yersinia Enterocolitica in Milk Samples. *Molecular and Cellular Probes* **2002**, *16* (4), 307–314. https://doi.org/10.1006/mcpr.2002.0428.
(242) Kiedrowski, M. R.; Crosby, H. A.; Hernandez, F. J.; Malone, C. L.; McNamara, J. O.; Horswill, A. R. Staphylococcus Aureus Nuc2 Is a Functional, Surface-Attached Extracellular Nuclease. *PLoS ONE* **2014**, *9* (4), e95574. https://doi.org/10.1371/journal.pone.0095574.
(243) Kiedrowski, M. R.; Kavanaugh, J. S.; Malone, C. L.; Mootz, J. M.; Voyich, J. M.; Smeltzer, M. S.; Bayles, K. W.; Horswill, A. R. Nuclease Modulates Biofilm Formation in Community-Associated Methicillin-Resistant Staphylococcus Aureus. *PLoS One* **2011**, *6* (11), e26714. https://doi.org/10.1371/journal.pone.0026714.
(244) Berends, E. T. M.; Horswill, A. R.; Haste, N. M.; Monestier, M.; Nizet, V.; von Köckritz-Blickwede, M. Nuclease Expression by Staphylococcus Aureus Facilitates Escape from Neutrophil Extracellular Traps. *J Innate Immun* **2010**, *2* (6), 576–586. https://doi.org/10.1159/000319909.
(245) Thammavongsa, V.; Missiakas, D. M.; Schneewind, O. Staphylococcus Aureus Degrades Neutrophil Extracellular Traps to Promote Immune Cell Death. *Science* **2013**, *342* (6160), 863–866. https://doi.org/10.1126/science.1242255.

(246) Papayannopoulos, V. Neutrophil Extracellular Traps in Immunity and Disease. *Nat Rev Immunol* **2018**, *18* (2), 134–147. https://doi.org/10.1038/nri.2017.105.
(247) Mutua, V.; Gershwin, L. J. A Review of Neutrophil Extracellular Traps (NETs) in Disease: Potential Anti-NETs Therapeutics. *Clin Rev Allergy Immunol* **2021**, *61* (2), 194–211. https://doi.org/10.1007/s12016-020-08804-7.
(248) Mann, E. E.; Rice, K. C.; Boles, B. R.; Endres, J. L.; Ranjit, D.; Chandramohan, L.; Tsang, L. H.; Smeltzer, M. S.; Horswill, A. R.; Bayles, K. W. Modulation of eDNA Release and Degradation Affects Staphylococcus Aureus Biofilm Maturation. *PLoS One* **2009**, *4* (6), e5822. https://doi.org/10.1371/journal.pone.0005822.
(249) Cotton, F. A.; Hazen, E. E.; Richardson, D. C. Crystalline Extracellular Nuclease of Staphylococcus Aureus. *Journal of Biological Chemistry* **1966**, *241* (19), 4389–4390. https://doi.org/10.1016/S0021-9258(18)99732-2.
(250) Arnone, A.; Bier, C. J.; Cotton, F. A.; Hazen, E. E.; Richardson, D. C.; Richardson, J. S. THE EXTRACELLULAR NUCLEASE OF Staphylococcus Aureus: STRUCTURES OF THE NATIVE ENZYME AND AN ENZYME-INHIBITOR COMPLEX AT 4 Å RESOLUTION*. *Proc Natl Acad Sci U S A* **1969**, *64* (2), 420–427.
(251) Arnone, A.; Bier, C. J.; Cotton, F. A.; Day, V. W.; Hazen, E. E.; Richardson, D. C.; Richardson, J. S.; Yonath, A. A High Resolution Structure of an Inhibitor Complex of the Extracellular Nuclease of Staphylococcus Aureus. *Journal of Biological Chemistry* **1971**, *246* (7), 2302–2316. https://doi.org/10.1016/S0021-9258(19)77221-4.
(252) Richardson, J. S.; Richardson, D. C. Studying and Polishing the PDB's Macromolecules. *Biopolymers* **2013**, *99* (3), 170–182. https://doi.org/10.1002/bip.22108.
(253) Bernstein, F. C.; Koetzle, T. F.; Williams, G. J.; Meyer, E. F.; Brice, M. D.; Rodgers, J. R.; Kennard, O.; Shimanouchi, T.; Tasumi, M. The Protein Data Bank: A Computer-Based Archival File for Macromolecular Structures. *J Mol Biol* **1977**, *112* (3), 535–542. https://doi.org/10.1016/s0022-2836(77)80200-3.
(254) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res* **2000**, *28* (1), 235–242. https://doi.org/10.1093/nar/28.1.235.
(255) Cotton, F. A.; Hazen, E. E.; Legg, M. J. Staphylococcal Nuclease: Proposed Mechanism of Action Based on Structure of Enzyme—Thymidine 3′,5′-Bisphosphate—Calcium Ion Complex at 1.5-Å Resolution. *Proc Natl Acad Sci U S A* **1979**, *76* (6), 2551–2555.
(256) Schwarzenbacher, R.; Godzik, A.; Grzechnik, S. K.; Jaroszewski, L. The Importance of Alignment Accuracy for Molecular Replacement. *Acta Crystallogr D Biol Crystallogr* **2004**, *60* (Pt 7), 1229–1236. https://doi.org/10.1107/S0907444904010145.
(257) Agirre, J.; Atanasova, M.; Bagdonas, H.; Ballard, C. B.; Baslé, A.; Beilsten-Edmands, J.; Borges, R. J.; Brown, D. G.; Burgos-Mármol, J. J.; Berrisford, J. M.; Bond, P. S.; Caballero, I.; Catapano, L.; Chojnowski, G.; Cook, A. G.; Cowtan, K. D.; Croll, T. I.; Debreczeni, J. É.; Devenish, N. E.; Dodson, E. J.; Drevon, T. R.; Emsley, P.; Evans, G.; Evans, P. R.; Fando, M.; Foadi, J.; Fuentes-Montero, L.; Garman, E. F.; Gerstel, M.; Gildea, R. J.; Hatti, K.; Hekkelman, M. L.; Heuser, P.; Hoh, S. W.; Hough, M. A.; Jenkins, H. T.; Jiménez, E.; Joosten, R. P.; Keegan, R. M.; Keep, N.; Krissinel, E. B.; Kolenko, P.; Kovalevskiy, O.; Lamzin, V. S.; Lawson, D. M.; Lebedev, A. A.; Leslie, A. G. W.; Lohkamp, B.; Long, F.; Malý, M.; McCoy, A. J.; McNicholas, S. J.; Medina, A.; Millán, C.; Murray, J. W.; Murshudov, G. N.; Nicholls, R. A.; Noble, M. E. M.;

Oeffner, R.; Pannu, N. S.; Parkhurst, J. M.; Pearce, N.; Pereira, J.; Perrakis, A.; Powell, H. R.; Read, R. J.; Rigden, D. J.; Rochira, W.; Sammito, M.; Sánchez Rodríguez, F.; Sheldrick, G. M.; Shelley, K. L.; Simkovic, F.; Simpkin, A. J.; Skubak, P.; Sobolev, E.; Steiner, R. A.; Stevenson, K.; Tews, I.; Thomas, J. M. H.; Thorn, A.; Valls, J. T.; Uski, V.; Usón, I.; Vagin, A.; Velankar, S.; Vollmar, M.; Walden, H.; Waterman, D.; Wilson, K. S.; Winn, M. D.; Winter, G.; Wojdyr, M.; Yamashita, K. The CCP4 Suite: Integrative Software for Macromolecular Crystallography. *Acta Cryst D* **2023**, *79* (6), 449–461. https://doi.org/10.1107/S2059798323003595.

(258) Richardson, J. S. The Anatomy and Taxonomy of Protein Structure. *Adv Protein Chem* **1981**, *34*, 167–339. https://doi.org/10.1016/s0065-3233(08)60520-3.

(259) Wiederstein, M.; Gruber, M.; Frank, K.; Melo, F.; Sippl, M. J. Structure-Based Characterization of Multiprotein Complexes. *Structure* **2014**, *22* (7), 1063–1070. https://doi.org/10.1016/j.str.2014.05.005.

(260) Green, S. M.; Gittis, A. G.; Meeker, A. K.; Lattman, E. E. One-Step Evolution of a Dimer from a Monomeric Protein. *Nat Struct Mol Biol* **1995**, *2* (9), 746–751. https://doi.org/10.1038/nsb0995-746.

(261) D'Alessio, G. The Evolutionary Transition from Monomeric to Oligomeric Proteins: Tools, the Environment, Hypotheses. *Prog Biophys Mol Biol* **1999**, *72* (3), 271–298. https://doi.org/10.1016/s0079-6107(99)00009-7.

(262) Heidari Tajabadi, F.; Medrano-Soto, A.; Ahmadzadeh, M.; Salehi Jouzani, G.; Saier, M. H. Comparative Analyses of Transport Proteins Encoded within the Genomes of Bdellovibrio Bacteriovorus HD100 and Bdellovibrio Exovorus JSS. *J Mol Microbiol Biotechnol* **2017**, *27* (6), 332–349. https://doi.org/10.1159/000484563.

(263) Lauber, F.; Cornelis, G. R.; Renzi, F. Identification of a New Lipoprotein Export Signal in Gram-Negative Bacteria. *mBio* **2016**, *7* (5), 10.1128/mbio.01232-16. https://doi.org/10.1128/mbio.01232-16.