The Characterisation of Proteins Key to the Hunting and Invasion of Prey by *Bdellovibrio bacteriovorus*

By

Hannah Louise Walters-Morgan

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

Institute of Microbiology and Infection
School of Biosciences
University of Birmingham
March 2021

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This unpublished thesis/dissertation is copyright of the author and/or third parties.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

The work presented in this thesis centres on proteins involved in the hunting and invasion of prey by *Bdellovibrio bacteriovorus HD100* (*Bdellovibrio*), an Oligoflexia bacterium that predates Gram-negative bacteria (Hahn *et al.*, 2017). In particular this work focuses on the characterisation of the Cache domain MCP family, Bd1081, Bd2503, Bd2504 and Bd3092 (as well as *SKB1291214* homologue Bd1081^{S}), and a family of 18 putative transcription factors (TFs) with a domain of unknown function (DUF4423) specifically, Bd1172, Bd1885, Bd2625 and Bd3720.

The primary aim of this thesis was the study of how *Bdellovibrio* uses chemotactic information to locate prey. Prior to this work *Bdellovibrio* was thought to locate prey cells using hydrodynamic forces, while chemotaxis was solely for the location of small essential molecules (Jashnsaz *et al.*, 2017). However, the structural and biochemical characterisation of the Cache domain MCP family revealed them to be highly specific to only 1 or 2 ligands, where the ligands sensed by Bd2503 and Bd2504 are Gram-negative centric. This allowed us to develop two working hypotheses for the hunting of prey by *Bdellovibrio*, the 'Chum in the Water' and the 'Lions at the Watering Hole' Mechanisms, both of which are presented here.

The secondary focus of the work presented here was to characterise and determine the function of DUF4423. The 18 member TF family containing this domain within *Bdellovibrio* were originally thought to be involved in switching the gene expression from attack phase to growth phase upon prey invasion. By obtaining a partial structure of Bd2625 we were able to confirm DUF4423 contained a winged helix-turn-helix like motif within the domain, and combined with EMSAs this allowed us to attribute the function of DNA binding to the DUF4423 domain. EMSAs of Bd1172, Bd1885 and in particular Bd3720 demonstrated that this family did not bind DNA in a specific manner as initially hypothesised. This lead us to developing a new hypothesis for how these proteins function within *Bdellovibrio*; where they act

as a set of novel nucleoid associated proteins that bind DNA in a non-sequence specific manner and regulate gene expression by chromatin remodelling.

Acknowledgements

Firstly, I would like to thank my supervisor, Andy Lovering. The guidance and support you have given me during my PhD has been invaluable. You have always been there as a willing sounding board for my ideas and theories, and have supported me through every step of my PhD. Without you I would not be the scientist I am today, and I could not have asked for a better boss, mentor and friend to guide me through these last four years.

I would also like to thank the past and present members of the Lovering Lab. Ian Cadby and Simon Caulton, you have both been amazing post-docs and have taught me so much during my time here. Rebecca Parr, Amber Wilson and Mauricio Valdivia Delgado, thank you for keeping me sane when things did not work out and for the hive mind that I relied on some much during those late night decisions. To the rest of the 7th floor, I am so glad I got to work along-side my friends for so long, it has been a real privilege and I will miss you all.

I would like to thank Dave Grainger for his support as my second supervisor, as well as the Grainger Lab and Jack Bryant for their help with my transcription factor project. Additionally, I would like to thank Patrick Moynihan for his help and specialist knowledge of HLPC.

Finally, I wish to thank my family and friends outside the lab. Mum and Dad, thank you for supporting my scientific curiosity growing up, without your constant belief in me (and probably that well timed phone call) I would never had the confidence to turn my passion into a career that I whole-heartedly love. Sophie, Ife and Alice, thank you for always being people I can rely on when things go wrong and those I want to celebrate with when they do work out, you made keeping going so much easier.

Table of Contents

Chapter 1 – Introduction ***1***

1.1 *Bdellovibrio bacteriovorus* and Related Predators **1**

1.1.1 Overview 1

1.1.2 Methods of predation 3

1.1.2.1 Phagocytosis-like Predation 5

1.1.2.2 Epibiotic Predation 5

1.1.2.3 Group Attack Predation 6

1.1.2.4 Endobiotic Predation 8

1.1.3 Bdellovibrio bacteriovorus Lifecycle 10

1.1.4 Host Independent Lifecycle 13

1.1.5 *Bdellovibrio bacteriovorus* – The 'Living Antibiotic' 15

1.2 Chemotaxis **18**

1.3 Commitment to Invasion **33**

1.3.1 Cyclic di-GMP Degradation and Receptors 34

1.3.2 Attack Phase to Growth Phase Switch 38

1.4 Aims **41**

Chapter 2 – Materials and Methods ***42***

2.1 Media and Solutions **42**

2.2 Primer Design and Restriction-Free Cloning **44**

2.3 Q5 Site-Directed Mutagenesis **47**

2.4 Sequencing **48**

2.5 Recombinant Protein Expression **48**

2.5.1 Glycerol Stocks 48

2.5.2 Expression Trials 48

2.5.3 Large Batch Expression 49

2.6 Purification **49**

2.7 SDS-PAGE **51**

2.8 Differential Scanning Fluorimetry **52**

2.9 Crystallisation **52**

2.10 Building Crystal Structures **53**

2.10.1 Cache Domain Chemoreceptor Structural Determination 53

2.10.2 Structural Determination of Partial Bd2625 Structure 55

2.11 High-performance liquid chromatography **55**

2.12 Isothermal Titration Calorimetry **56**

2.13 Electrophoretic Mobility Shift Assay **57**

2.13.1 Screening EMSAs 59

2.13.2 Qualitative EMSAs 60

2.13.3 Specificity of Bd3720 61

2.13.4 Identifying DNA Sequences/Motifs Responsible for the Binding of Bd3720 62

2.13.5 Competition EMSAs 65

2.13.6 Bd3720 and Bd1885 mutant EMSAs 65

2.14 Circular Dichroism Spectroscopy 67

Chapter 3 – The Structural Determination of Bdellovibrio bacteriovorus Chemoreceptors 68

3.1 Introduction 68

3.2 Results 72

3.2.1 Structural Determination of Bd2503^{LBD} 72

3.2.1.1 Purification of Bd2503^{LBD} 72

3.2.1.2 Differential Scanning Fluorimetry of Bd2503^{LBD} 73

3.2.1.3 Structure Determination and Model Building of Bd2503^{LBD} 75

3.2.2 Structural Determination of Bd2504^{LBD} 88

3.2.2.1 Purification of Bd2504^{LBD} 88

3.2.2.2 Data Collection and Structure Building of Bd2504^{LBD} 88

3.2.3 Structural Determination of Bd1081$^{S\text{-}LBD}$ 94

3.2.3.1 Purification of Bd1081$^{S\text{-}LBD}$ 94

3.2.3.2 Data Collection and Structure Building of Bd1081$^{S\text{-}LBD}$ 98

3.2.4 Structural Determination Bd3092 105

3.2.4.1 Purification of Bd3092 105

3.2.4.2 Data Collection and Structure Building of Bd3092 106

3.2.5 Fold and Sequence Similarity 111

3.3 Discussion 113

3.3.1 The Polyamine Precedent 113

3.3.2 dCache Sensory Domain Structural Comparison 117

Chapter 4 – Testing the ligand binding of Bdellovibrio bacteriovorus Chemoreceptors... 125

4.1 Introduction 125

4.2 Results 126

4.2.1 High-Throughput Liquid Chromatography 126

4.2.2 Structural Determination of Bd1081 Ligand Binding Domain 129

4.2.2.1 Purification of Bd1081^{LBD} 129

4.2.2.2 Data Collection and Structure Building of Bd1081^{LBD} 130

4.2.3 Isothermal Titration Calorimetry 138

4.2.3.1 Bd2503 140

4.2.3.2 Bd2504 142

4.2.3.3 Bd1081^{LBD} 144

4.2.3.4 Bd1081$^{S\text{-}LBD}$ 146

4.2.3.5 Bd3092 147

4.2.4 Identifying Amino Acids Involved in the Variation of Ligand Specificity Between Bd1081^{LBD} and Bd1081$^{S\text{-}LBD}$ 149

4.2.4.1 Changes in Amino acid Sequence Between Bd1081^{LBD} and Bd1081$^{S\text{-}LBD}$ Ligand Binding Pocket 149

4.2.4.2 Purification and Isothermal Titration Calorimetry of Bd1081$^{S\text{-}LBD}$ D120G 152

4.2.4.3 Purification and Isothermal Titration Calorimetry of Bd1081$^{S\text{-}LBD}$ V126E 153

4.2.4.4 Purification and Isothermal Titration Calorimetry of Bd1081$^{S\text{-}LBD}$ Double 154

4.2.5 Bd3126^{LBD} – A dCache Histidine Kinase 155

4.3 Discussion 160

4.3.1 Quaternary Amine Compounds Precedent 160

4.3.2 Conservation of the Dimeric Interface 166
4.3.3 Modelling of Ligands Determined via ITC 169
4.3.4 Conformational Change Between Apo and Bound dCache Structures 184
4.3.5 Working Hypotheses 187
4.3.5.1 The 'Chum in the Water' Hypothesis 187
4.3.5.2 The 'Lions at the Watering Hole' Hypothesis 189
4.3.6 Bd3126 – Known Fold, Novel Use 190

Chapter 5 – Attack Phase to Growth Phase Switch Transcription Factors in Bdellovibrio bacteriovorus 193

5.1 Introduction 193
5.1.1 DUF4423 193
5.1.1.1 DUF4423 in Non-predators 194
5.1.1.2 DUF4423 in predators 195
5.1.2 Predicted Architecture 195
5.1.3 Effectors 197

5.2 Results 201
5.2.1 Gene Expression and Purification 201
5.2.1.1 Cloning of Transcription Factor constructs 201
5.2.1.2 Purification of Bd1905 202
5.2.1.3 Purification of Bd1897 203
5.2.1.4 Purification of Bd1172 204
5.2.1.5 Purification of Bd1885 205
5.2.1.6 Purification of Bd2625 206
5.2.1.7 Purification of Bd3720 207
5.2.2 Determination of Partial Bd2625 Structure 208
5.2.3 Mutagenesis of Bd2625 224
5.2.4 Mutagenesis of Bd1885 226
5.2.5 Electrophoretic Mobility Shift Assays 228
5.2.5.1 Screening EMSAs 228
5.2.5.2 Qualitative EMSAs 231
5.2.5.3 Specificity of Bd3720 234
5.2.5.4 Identifying DNA Sequences/Motifs Responsible for the Binding of Bd3720 238
5.2.5.5 Competition EMSAs 240
5.2.6 Identifying Amino Acids Involved in DNA binding 241
5.2.7 Mutation of Bd3720 245
5.2.7.1 Cloning and Purification of Bd3720 polyA recognition helix mutants 245
5.2.7.2 Circular Dichroism Spectroscopy 248
5.2.7.3 Electrophoretic Mobility Shift Assays 250
5.2.8 Cloning and Purification of Effectors 253
5.2.8.1 Bd2625/Bd2624 253
5.2.8.2 Bd1896 254
5.2.8.3 Bd0723 E122A 255

5.3 Discussion 257
5.3.1 Proteins with Multiple, Tandem DNA-binding Domains 257
5.3.2 DUF4423 – A Novel DNA Binding Domain 260
5.3.3 Updated Working Hypothesis 265
5.3.3.1 From self-regulating AP/GP Switch Transcription Factors 265

5.3.3.2 To Hypothesised Environmentally Controlled Nucleoid-Associated Proteins..... 266
5.3.4 Effectors 279

Chapter 6 – Summary 284

6.1 Chemotaxis 284
6.2 Novel Nucleoid-Associated Protein Family 286

References..... i

Appendices..... xxxv

Appendix I – Bd2503_Native Crystal Data Log File..... xxxv
Appendix II – Bd2503_Cadaverine Crystal Data Log File..... xxxvi
Appendix III – Bd2503_Arginine Crystal Data Log File..... xxxvii
Appendix IV – Bd2504 Crystal Data Log File..... xxxviii
Appendix V – Bd1081^{S} Crystal Data Log File..... xxxix
Appendix VI – Bd3092 Crystal Data Log File..... xl
Appendix VII – Bd1081 Crystal Data Log File..... xli
Appendix VIII – Bd2625 Crystal Data Log File..... xlii
Appendix IX – Crystal Data Summary Table..... xliii
Appendix X – DNA Fragment Sequences Used in EMSAs..... xlv
Appendix XI – CD Fitted Plot l
Appendix XII – Collaboration Work lii

Table of Figures

Chapter 1

Figure 1.1	Methods of Predation	4
Figure 1.2	Lifecycle of *Bdellovibrio bacteriovorus*	11
Figure 1.3	Simplified Chemosensory Pathway in *Escherichia coli*	19
Figure 1.4	Possible Mechanism of Transmembrane Signal Transduction	24
Figure 1.5	Cartoon of the 5 Ligand Binding Domains in *Bdellovibrio bacteriovorus*	30
Figure 1.6	Sequence Alignment of $Bd2622^{LBD}$ and $Bd3192^{LBD}$ to TII0287	32
Figure 1.7	Simplified Pathway of Two Component Systems	37
Figure 1.8	Phylogenic Tree of the AP/GP Switch Transcription Factor Family	39

Chapter 3

Figure 3.1	Representative Structure of dCache and sCache Domain	70
Figure 3.2	Structures of dCache Domains Bound to a Variety of Ligands	71
Figure 3.3	Purification of $Bd2503^{LBD}$	72
Figure 3.4	Melting Curve of DSF Additive Screen	74
Figure 3.5	Comparison of mmHK1S-$Z2^{LBD}$ Structure to $Bd2503^{LBD}$_Native Structure and Refined Density Map	78
Figure 3.6	Dimerisation of $Bd2503^{LBD}$	79
Figure 3.7	Structure and Ligand Binding Pocket of $Bd2503^{LBD}$ with Cadaverine	80
Figure 3.8	Anomalous Difference Density Map of $Bd2503^{LBD}$ Ligand Binding Pocket	81
Figure 3.9	Structure and Ligand Binding Pocket of $Bd2503^{LBD}$ with L-arginine	87
Figure 3.10	Ligand Binding Pocket of $Bd2503^{LBD}$ with Cadaverine Bound Compared to $Bd2503^{LBD}$ with L-arginine Bound	87
Figure 3.11	Purification of $Bd2504^{LBD}$	88
Figure 3.12	Dimerisation of $Bd2504^{LBD}$	92
Figure 3.13	Structure and Ligand Binding Pocket of $Bd2504^{LBD}$ with Putrescine	93
Figure 3.14	Anomalous Difference Density Map of $Bd2504^{LBD}$ Ligand Binding Pocket	94
Figure 3.15	Alignment of the sequence of *Bdellovibrio* dCache Domains	96
Figure 3.16	Purification of $Bd1081^{S\text{-}LBD}$	97

Figure 3.17	Asymmetric Unit of Bd1081$^{S\text{-}LBD}$	101
Figure 3.18	Structure and Ligand Binding Pocket of Bd1081$^{S\text{-}LBD}$ with Bis tris	103
Figure 3.19	Anomalous Difference Density Map of Bd1081$^{S\text{-}LBD}$ Ligand Binding Pocket	103
Figure 3.20	Ligand Binding Pocket of Bd1081$^{S\text{-}LBD}$ Compared to mmHK1S-Z2	104
Figure 3.21	Purification of Bd3092^{LBD}	105
Figure 3.22	Dimerisation of Bd3092^{LBD}	109
Figure 3.23	Structure and Ligand Binding Pocket of Bd3092^{LBD} with Acetate	110
Figure 3.24	Anomalous Difference Density Map of Bd3092^{LBD} Ligand Binding Pocket	110
Figure 3.25	Chemical Structure of Polyamines and Related Molecules	114
Figure 3.26	Structural Comparison of McpU^{LBD}, Mlp37^{LBD}, Bd2503^{LBD} and Bd2504^{LBD}	117
Figure 3.27	Ligand Binding Pocket of Bd2503^{LBD} with Cadaverine Bound Compared to Bd2504^{LBD} with Putrescine Bound	119
Figure 3.28	Ligand Binding Pockets of Bd2503^{LBD} and Bd2504^{LBD} with Van der Waals shown	120
Figure 3.29	Comparison of Bd2503^{LBD}, Bd2504^{LBD} and McpU^{LBD} Ligand Binding Pockets	121
Figure 3.30	Alignment of dCache Domain Proteins	123
Figure 3.31	Ligand Binding Pocket of Bd3092^{LBD} with Van der Waals shown	124

Chapter 4

Figure 4.1	HPLC Traces of Proteins and Standards	128
Figure 4.2	Purification of Bd1081^{LBD}	130
Figure 4.3	Dimerisation of Bd1081^{LBD}	134
Figure 4.4	Structure and Ligand Binding Pocket of Bd1081^{LBD} with TMAO	136
Figure 4.5	Anomalous Difference Density Map of Bd1081^{LBD} Ligand Binding Pocket	136
Figure 4.6	TorT and TmoX bound to TMAO	137
Figure 4.7	Chemical Structures of Molecules used in ITC	139
Figure 4.8	Raw ITC Data and Binding Curve of Bd2503^{LBD} with Cadaverine	140
Figure 4.9	Raw ITC Data and Binding Curve of Bd2503^{LBD} with DAP	141
Figure 4.10	Raw ITC Data and Binding Curve of Bd2503^{LBD} with Putrescine	141
Figure 4.11	Raw ITC Data and Binding Curve of Bd2503^{LBD} with L-Arginine	142
Figure 4.12	Raw ITC Data and Binding Curve of Bd2504^{LBD} with Putrescine	143
Figure 4.13	Raw ITC Data and Binding Curve of Bd2504^{LBD} with Cadaverine	143

Figure 4.14	Raw ITC Data and Binding Curve of Bd2504^{LBD} with DAP	144
Figure 4.15	Raw ITC Data and Binding Curve of Bd1081^{LBD} with Choline	145
Figure 4.16	Raw ITC Data and Binding Curve of Bd1081^{LBD} with DAP	145
Figure 4.17	Raw ITC Data and Binding Curve of Bd1081^{LBD} with Bis tris	146
Figure 4.18	Raw ITC Data and Binding Curve of Bd1081$^{S\text{-}LBD}$ with Choline	147
Figure 4.19	Raw ITC Data and Binding Curve of Bd3092^{LBD} with Propanoic acid	148
Figure 4.20	Raw ITC Data and Binding Curve of Bd3092^{LBD} with Acetate	148
Figure 4.21	Raw ITC Data and Binding Curve of Bd3092^{LBD} with Thioacetate	149
Figure 4.22	Sequence Alignment of Bd1081^{LBD} and Bd1081$^{S\text{-}LBD}$	150
Figure 4.23	Ligand Binding Pocket of Bd1081^{LBD} and Bd1081$^{S\text{-}LBD}$	151
Figure 4.24	Purification of Bd1081$^{S\text{-}LBD}$ D120G	152
Figure 4.25	Purification of Bd1081$^{S\text{-}LBD}$ V125E	153
Figure 4.26	Purification of Bd1081$^{S\text{-}LBD}$ Double	154
Figure 4.27	Purification of Bd3126^{LBD}	156
Figure 4.28	Bd3126^{LBD} Crystals	157
Figure 4.29	ITC Binding Curves of Bd3126^{LBD}	159
Figure 4.30	Chemical Structure of QACs and Related Molecules	160
Figure 4.31	Structural Comparisons of McpX^{LBD}, Mlp37^{LBD} and Bd1081^{LBD}	165
Figure 4.32	Sequence Alignment of Bd1081^{LBD}, Bd2503^{LBD} and Bd2504^{LBD}	167
Figure 4.33	Structural Alignment of Conserved Amino acids in Dimerisation Helices in Bd1081^{LBD}, Bd2503^{LBD} and Bd2504^{LBD}	168
Figure 4.34	Helical Wheel Diagram of Dimerisation Helices in Bd1081^{LBD}	168
Figure 4.35	Structural Alignment of Amino acids Conserved in Dimerisation in Bd1081^{LBD}, Bd2503^{LBD} and Bd2504^{LBD}	169
Figure 4.36	Model of Bd2503^{LBD} with DAP	172
Figure 4.37	Model of Bd1081^{LBD} with Choline	174
Figure 4.38	Model of Bd1081$^{S\text{-}LBD}$ with Choline	175
Figure 4.39	Structural Comparisons of Bd1081^{LBD}:TMAO, Bd1081^{LBD}:Choline, Bd1081$^{S\text{-}LBD}$:Choline and McpX^{LBD}:Proline betaine	177
Figure 4.40	Model of Bd1081^{LBD} with DAP	179
Figure 4.41	Structural Comparisons of Bd1081^{LBD}:TMAO, Bd1081^{LBD}:DAP, Bd2504^{LBD}:Putrescine, McpU^{LBD}:Putrescine and Mlp37^{LBD}:Taurine	181
Figure 4.42	Model of Bd3092^{LBD} with Propanoic acid	184

Figure 4.43 Conformational Change Between Apo and Bound States of the dCache fold 186
Figure 4.44 The Chum in the Water Predation Mechanism 188
Figure 4.45 The Lions at the Watering Hole Predation Mechanism 190

Chapter 5

Figure 5.1 Schematic of Different DUF4423 Protein Architectures 194
Figure 5.2 Predicted Secondary Structure of 18 Member TF Family 197
Figure 5.3 Gene Overlap Between TFs and Effectors 200
Figure 5.4 Phylogenic Tree of the AP/GP Switch Transcription Factor Family 201
Figure 5.5 Expression of Bd0089+GFP 202
Figure 5.6 Expression and Purification of Bd1905 203
Figure 5.7 Purification of Bd1897 204
Figure 5.8 Purification of Bd1172 205
Figure 5.9 Purification of Bd1885 206
Figure 5.10 Purification of Bd2625 207
Figure 5.11 Purification of Bd3720 208
Figure 5.12 Fit of Initial Bd2625 Structure to Initial Density Map 210
Figure 5.13 Platinum Sites in Partial Bd2625 Structure 211
Figure 5.14 Comparison of Predicted to Observed Bd2625 Structure 212
Figure 5.15 Tetrameric Partial Bd2625 Structure 213
Figure 5.16 Hydrogen Bonding Pattern that Stabilises Dimerisation of Partial Bd2625 Structure 214
Figure 5.17 Side Chain Packing Pattern that Stabilises Dimerisation of Partial Bd2625 Structure 214
Figure 5.18 Hydrophobic Interaction that Stabilises Dimerisation of Partial Bd2625 Structure 215
Figure 5.19 π-π Interaction that Stabilises Dimerisation of Partial Bd2625 Structure 216
Figure 5.20 Hydrophobic Interaction that Stabilises Tetramerisation of Partial Bd2625 Structure 216
Figure 5.21 Hydrogen Bonding Pattern that Stabilises DNA Recognition Helix of Partial Bd2625 Structure 217
Figure 5.22 Diagonal Presentation of Recognition Helices in Partial Bd2625 Structure 218

Figure 5.23	Sequence Alignment of Bd2625 DUF4423 to Homologues in Other Bdellovibrio Strains	219
Figure 5.24	Structural Comparison Between Partial Bd2625 Structure and iscR	221
Figure 5.25	Comparison of Structural Differences of Partial Bd2625 Structure and iscR DNA Binding	222
Figure 5.26	Composite Model of Partial Bd2625 Structure bound to DNA	223
Figure 5.27	Purification of Bd2625 NT-WingCT	225
Figure 5.28	Purification of Bd2625 NT-VAS and Bd2625 WingNT-VAS	226
Figure 5.29	Purification of Bd1885NT	227
Figure 5.30	Screening EMSAs of Bd1172	229
Figure 5.31	Screening EMSAs of Bd1885	230
Figure 5.32	Qualitative EMSAs of Bd1172	232
Figure 5.33	Qualitative EMSAs of Bd1885	233
Figure 5.34	Qualitative EMSAs of Bd3720	234
Figure 5.35	Upstream DNA Fragment Screening EMSAs of Bd3720	236
Figure 5.36	Specificity Screening EMSAs of Bd3720	238
Figure 5.37	Fragment Screening of Bd3720	239
Figure 5.38	Schematic of *bd3721*^{U} Fragments	240
Figure 5.39	Schematic of Possible Bd3720 Binding Sites in *bd3721*^{U}	240
Figure 5.40	Competition EMSAs of Bd3720	241
Figure 5.41	Purification of Bd3720 polyA HTH	246
Figure 5.42	Purification of Bd3720 polyA WHTH	247
Figure 5.43	Purification of Bd3720 polyA Double	247
Figure 5.44	CD Spectra of Bd3720 and Bd3720 polyA Mutants	249
Figure 5.45	Screening EMSAs of Bd3720 polyA Mutants	251
Figure 5.46	Screening EMSAs of Bd1885NT	252
Figure 5.47	Purification of Bd2625/Bd2624	253
Figure 5.48	Purification of Bd1896	255
Figure 5.49	Purification of Bd0723 E112A	256
Figure 5.50	Structures of Multiple DNA Binding Domain Protein in *Eukaryotes*	258
Figure 5.51	Structures of Multiple DNA Binding Domain Protein in *Prokaryotes*	260
Figure 5.52	Structural Comparison of Partial Bd2625 Structure, PF0864 and DN05_10610	263

Figure 5.53 Differential Gene Expression of the TF Family at Varying Point in the Bdellovibrio Lifecycle 268
Figure 5.54 Mechanisms of Chromatin Condensation by NAPs 270
Figure 5.55 Mechanism of by which the Fis Tetramer Bridges DNA 271
Figure 5.56 Structural Comparison Between Bd2625 and Predicted DUF4423 277
Figure 5.57 Predicted Full Length Structure of Novel NAP Family 278

Chapter 6

Figure 6.1 Proposed Mechanism of Prey Hunting by *Bdellovibrio* 285

Table of Tables

Chapter 1

Table 1.1	List of Methyl-accepting Chemotaxis Protein in *Bdellovibrio bacteriovorus*	26
Table 1.2	List of Che protein in *Bdellovibrio bact*eriovorus	27
Table 1.3	List of Genes in the *bd3126-bd3116* Operon	37

Chapter 2

Table 2.1	Forward and Reverse Primers used in Gene Cloning	44
Table 2.2	Standard Cycling Steps for Primary Restriction Free Cloning PCR	45
Table 2.3	Standard Cycling Steps for Secondary Restriction Free Cloning PCR	46
Table 2.4	Standard Cycling Steps for Colony PCR	46
Table 2.5	Forward and Reverse Primers used in Site Directed Mutagenesis	47
Table 2.6	Standard Cycling Steps for Site Directed Mutagenesis PCR	47
Table 2.7	Buffers used for Protein Dialysis	50
Table 2.8	Molecular Weight and Extinction Coefficient for Protein Constructs	51
Table 2.9	Crystallisation Conditions that Yielded Structures	53
Table 2.10	Buffers used in ITC Experiments	56
Table 2.11	ITC Experiments Performed	57
Table 2.12	Forward and Reverse Primers Used for DNA Fragment Production	58
Table 2.13	EMSA Gel Recipes	58
Table 2.14	Bd1172, Bd1885 and Bd2625 Screening EMSAs	59
Table 2.15	Bd1172, Bd1885 and Bd3720 Qualitative EMSAs	60
Table 2.16	Bd3720 Specificity Screening EMSAs	61
Table 2.17	Bd3720 Specificity Test EMSAs	62
Table 2.18	Primers for Truncated $bd3721^{U}$ DNA Fragment Production	63
Table 2.19	Truncated $bd3721^{U}$ DNA Fragment Screening EMSAs	64
Table 2.20	Bd3720 Competition EMSAs	65
Table 2.21	Bd1885NT and Bd3720 polyA Mutant Screening EMSAs	66
Table 2.22	Bd3720 WT and polyA Mutant Concentration used in CD	67

Chapter 3

Table 3.1	Crystallisation Conditions for $Bd2503^{LBD}$_Native	75
Table 3.2	Diffraction dataset specifications for $Bd2503^{LBD}$_Native	76
Table 3.3	Final Refinement Statistics for $Bd2503^{LBD}$_Native	77
Table 3.4	Crystallisation Conditions for $Bd2503^{LBD}$ with Cadaverine	81
Table 3.5	Diffraction dataset specifications for $Bd2503^{LBD}$ with Cadaverine	82
Table 3.6	Final Refinement Statistics for $Bd2503^{LBD}$ with Cadaverine	83
Table 3.7	Crystallisation Conditions for $Bd2503^{LBD}$ with L-arginine	84
Table 3.8	Diffraction dataset specifications for $Bd2503^{LBD}$ with L-arginine	84
Table 3.9	Final Refinement Statistics for $Bd2503^{LBD}$ with L-arginine	85
Table 3.10	Crystallisation Conditions for $Bd2504^{LBD}$ with Putrescine	89
Table 3.11	Diffraction dataset specifications for $Bd2504^{LBD}$ with Putrescine	90
Table 3.12	Final Refinement Statistics for $Bd2504^{LBD}$ with Putrescine	91
Table 3.13	Crystallisation Conditions for $Bd1081^{S\text{-}LBD}$	98
Table 3.14	Diffraction dataset specifications for $Bd1081^{S\text{-}LBD}$	99
Table 3.15	Final Refinement Statistics for $Bd1081^{SLBD}$	100
Table 3.16	Crystallisation Conditions for $Bd3092^{LBD}$	106
Table 3.17	Diffraction dataset specifications for $Bd3092^{LBD}$	107
Table 3.18	Final Refinement Statistics for $Bd3092^{LBD}$	108
Table 3.19	HHpred Results for $Bd1081^{S\text{-}LBD}$, $Bd2503^{LBD}$ and $Bd2504^{LBD}$	111
Table 3.20	DALI Results for $Bd1081^{S\text{-}LBD}$	112
Table 3.21	DALI Results for $Bd2503^{LBD}$	112
Table 3.22	DALI Results for $Bd2504^{LBD}$	112

Chapter 4

Table 4.1	Crystallisation Conditions for $Bd1081^{LBD}$ with TMAO	131
Table 4.2	Diffraction dataset specifications for $Bd1081^{LBD}$ with TMAO	132
Table 4.3	Final Refinement Statistics of $Bd1081^{LBD}$ with TMAO	133
Table 4.4	Summary of ITC Results	139
Table 4.5	HHpred results of $Bd3126^{LBD}$	155
Table 4.6	Crystallisation Conditions for $Bd3126^{LBD}$	157

Table 4.7 BlastP of Bd1081LBD, Bd2503LBD and Bd2504LBD Across Other Bdellovibrio Strains 182

Table 4.8 dCache Proteins with the Highest Sequence Identity to $Bd3126^{LBD}$ 191

Chapter 5

Table 5.1 Crystallisation Conditions for Partial Bd2625 Structure 208

Table 5.2 Diffraction Dataset Specifications for Partial Bd2625 Structure Derivatised with Pt 209

Table 5.3 Final refinement statistics for Partial Bd2625 Structure diffraction dataset 210

Table 5.4 Crystallisation Conditions of Bd1885NT 227

Table 5.5 Screening EMSAs of Bd1172 229

Table 5.6 Screening EMSAs of Bd1885 230

Table 5.7 Qualitative EMSAs of Bd1172 231

Table 5.8 Qualitative EMSAs of Bd1885 232

Table 5.9 Qualitative EMSAs of Bd3720 234

Table 5.10 Upstream DNA Fragment Screening EMSAs of Bd3720 235

Table 5.11 Specificity Screening EMSAs of Bd3720 237

Table 5.12 Fragment Screening of Bd3720 238

Table 5.13 Competition EMSAs of Bd3720 241

Table 5.14 Predicted DNA Recognition Amino acids in Each Member of the TF Family 243

Table 5.15 Predicted DNA Recognition Amino acids in Bd1885 Homologues in Other Bdellovibrio Strains 244

Table 5.16 polyA Mutations of Bd3720 Recognition Helices 245

Table 5.17 CD Spectra Analysis and Predicted Secondary Structure of Bd3720 WT and Mutants 250

Table 5.18 Screening EMSAs of Bd3720 polyA Mutants 251

Table 5.19 Screening EMSAs of Bd1885NT 252

Table 5.20 DALI Results for Partial Bd2625 Structure 262

Table 5.21 Topsearch Results for Partial Bd2625 Structure 264

Table 5.22 Differential Gene Expression of the TF Family at Varying Points in the *Bdellovibrio* Lifecycle 267

Table 5.23 AT Richness of DNA Fragments Used in EMSAs 275

Table 5.24 Differential Gene Expression of the Effector Family at Varying Point in the *Bdellovibrio* Lifecycle 282

Table 5.25 Ratio of NAPs:Effectors Across Other *Bdellovibrio* Strains 283

List of Abbreviations

2xLB	2x Lysogeny Broth
4HB	Four Helical Bundles
AI	Auto-Induction Media
AIEC	Adherent-Invasive *E. coli*
AP	Attack Phase
APS	Ammonium Persulphate
BALO	*Bdellovibrio* and Like Organisms
Bdellovibrio	*Bdellovibrio bacteriovorus*
BLAST	Basic Local Alignment Tool
Cache	Calcium Channels and Chemotaxis Receptors
CD	Circular Dichroism
c-di-GMP	Cyclic diguanylate
DAM	DNA Adenine Methylation
DAP	1,3-diaminopropane
dCache	Double Cache Domain
dNTPs	Deoxynucleotides
DSF	Differential Scanning Fluorimetry
DUF	Domain of Unknown Function
EAL	Glutamate-Alanine-Leucine
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
EMSA(s)	Electrophorectic Mobility Shift Assay(s)
EPS	Extracellular Polysaccharides
FAD	Flavin Adenine Dinucleotide
FL	Full Length
GABA	γ-Aminobutyric acid
GAS	Group A *Streptococcus*
GFP	Green Fluorescent Protein
GGDEF	Glycine-Glycine-Aspartate-Glutamate-Phenylalanine
GP	Growth Phase
HAMP	Histidine kinase, Adenylate cyclases, Methyl-accepting proteins and Phosphatases
HD	Host-dependent
HDGYP	Histidine-Aspartate-Glycine-Tyrosine-Proline
HI	Host-independent
HIT	Host-interaction
HPLC	High-Performance Liquid Chromatography
HTH	Helix-Turn-Helix
Kd	Disassociation Constant
KLDR	Kinase Ligase Digestion Reaction
KO	Knockout
LB	Lysogeny Broth
LBD	Ligand Binding Domain
LBP	Ligand Binding Pocket
IMAC	Immobilised Metal Affinity Chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITC	Isothermal Titration Calorimetry
MCP	Methyl-accepting Chemotaxis Protein

MD	Macrodomain(s)
MH	Methylated Helical
MWCO	Molecular Weight Cut Off
NAP(s)	Nucleoid-associated Protein(s)
OD	Optical Density
ORF	Open Reading Frame
PAS	Per, Arnt and Sim
PBP	Periplasmic Binding Protein
PC	Phosphatidylcholine
PCR	Polymerase Chain Reaction
PEG	Polyethylene Glycol
Poly dI-dC	poly(deoxyinosinic-deoxycytidylic)
RHH	Ribbon-Helix-Helix
RHS	Recombination Hotspot
RMSD	Root Mean Square Deviation
sCache	Single Cache Domain
SDM	Site Directed Mutagenesis
SDS-PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
SDW	Sterile Distilled Water
SEC	Size Exclusion Chromatography
T4SS	Type IV Secretion System-like Structure
TAE	Tris/Acetate/EDTA
TB	Terrific Broth
TEMED	Tetramethylethylenediamine
TEV	Tobacco Etch Virus
TF(s)	Transcription Factor(s)
Tm	Annealing Temperature
TMAO	Trimethylamine N-oxide
tNCS	Translational Non-crystallographic Symmetry
QACs	Quaternary Ammonium Compounds
WapA	Wall Associated Protein A
WHTH	Winged Helix-Turn-Helix
WT	Wild Type
ZBD	Z-DNA Binding Domain

Chapter 1 – Introduction

1.1 *Bdellovibrio bacteriovorus* and Related Predators

1.1.1 Overview

Competition for resources has long been observed within macro-biology, however for a long-time similar competition at a micro- level was left unappreciated. Nowadays, we consider the understanding of the parallels seen between the macro- and microworld, in terms of both symbiosis and antagonism, as common knowledge (Guerrero et al., 1986). Whether intra- or interspecies, competition is normally driven by nutritional requirements for replication rather than space limitations (Hibbing et al., 2010; Stannek et al., 2014). This bacteria-bacteria antagonism generally involves the use of a bacteriocin to inhibit growth or lyse sensitive competitors (Martin, 2002; Pérez et al., 2016).

In the 1940s, *Myxococcus xanthus* was the first bacterium to be identified to belong to a group of specialist bacteria that would further emphasise the similarities between the macro- and microworlds (Beebe, 1941). First deemed parasitic, we are now starting to understand these bacteria for what they are, micro-predators, actively hunting and killing their prey in order to survive rather than leeching nutrients from a host (Pérez et al., 2016). Predatory bacteria are a group of predominately Gram-negative bacteria originally thought to be unique to the class δ-proteobacteria, however, as more are discovered this is also proving false. While they do have a niche lifestyle, hunting and living almost exclusively off other microorganisms, they are found in almost every environment, from soil to fresh and salt water, as well as sewage and animal biomes (Davidov et al., 2006).

The continued study into predatory bacteria has not only revealed a plethora of different lifestyles, but also the wide-reaching implications they have on ecosystems and evolution, as well as possible

beneficial uses. Micro-predators have been accredited with having a crucial role in population control as well as effecting biodiversity (Pérez et al., 2016; Casida, 1989). In 2006, Lueders *et al.* used C^{13} to show that protists and bacteriophage are not entirely responsible for bacterial mortality, and that predatory bacteria, like *M. xanthus*, effected the overall biomass (Lueders et al., 2006; Kandel et al., 2014). A secondary study in 2021 similarly concluded that that predatory bacteria had a top-down effect on non-predatory bacterial populations (Hungate et al., 2021). Likewise, studies into biodiversity have shown that predatory bacteria play a role in affecting the biodiversity of an ecosystem and promote mean richness (Chen et al., 2011; Griffin et al., 2013). Predatory bacteria may also hold the answers to one of the most fundamental questions in biology; how did multicellular life evolve? By 1986, the idea of bacterial symbiosis as a precursor to eukaryotic evolution was well established (Guerrero et al., 1986). A major obstacle to this theory has been that prokaryotes are unable to perform phagocytosis. This has led to postulation that a neomuran, an ancestral precursor to archaebacteria and eukaryotes, containing something akin to a periplasm was invaded by a predatory bacterium that proceeded to stably replicate, forming the precursor to mitochondria. This mitochondrial precursor would most likely have come from the class *α-proteobacteria*, as this class still has the most mitochondrial reminiscent respiratory chain (Cavalier-Smith, 2009; Davidov and Jurkevitch, 2009; Yutin et al., 2008). Further support of this hypothesis was found by Von Dohlen *et al.* in 2001, as they found that the symbiotic bacteria within mealybugs, were *β-proteobacteria* with *γ-proteobacteria* stably living symbiotically inside generation after generation (Von Dohlen et al., 2001). However, in 2019 Shiratori *et al.* discovered and described *Candidatus Uab amorphum*, a bacterium belonging to the *Planctomycetes* phylum that is capable of predation by a phagocytosis-like mechanism (Shiratori et al., 2019). Investigations into the relationship between predators and prey has also suggested a link between predation and human pathogen virulence. The mechanisms that aid pathogens, may have arisen from predator evasion, for example the development of an intracellular lifestyle to prevent being engulfed by protists. However, any evidence of this seems to point towards micro-eukaryotes being the driving factor rather than predatory bacteria (Erken et al.,

2013). Studies into predatory bacteria have also elucidated a potential use of them as a 'living antibiotic'. As we are currently at the precipice of a world-wide antibiotic resistance crisis, the need for novel antibiotics has never been more critical. Predatory bacteria have been shown to kill multiple human pathogens within laboratory environments without the prey developing resistance. More recent studies are beginning to show that these predators may be effective killers of pathogens *in vivo* (Raghunathan et al., 2019).

1.1.2 Methods of predation

Despite all fulfilling a similar niche within the environment there are many different methods of predation, normally categorised as: Group attack, Epibiotic and Endobiotic (Figure 1.1) (Pérez et al., 2016). However, due to the previously mentioned work by Shiratori *et al.* a fourth type, Phagocytosis-like, must be added (Shiratori et al., 2019). As further studies into predatory bacteria have taken place and more have been identified, it has been realised that the methods of predation utilised are as diverse as the bacteria that utilise them. It is firmly believed that predation strategy influences bacterial morphology. Predators that attack as a group seem to be roughly similar in size to their prey, whilst those that attack one-on-one are notably smaller than prey, with the recent exception of the phagocytotic predator *C. Uab amorphum* (Shiratori et al., 2019). Nevertheless, one similarity between all predators seems to be that most are highly motile in order to actively hunt down their prey (Martin, 2002; Pérez et al., 2016). When Pasternak *et al.* compared the genomes of 11 sequenced predatory bacteria they found a high percentage of adhesins, proteases and metabolites that enable them to attach to and break down their prey (Pasternak et al., 2013). Each method of predation has a 'model' organism best characterised for that mechanism.

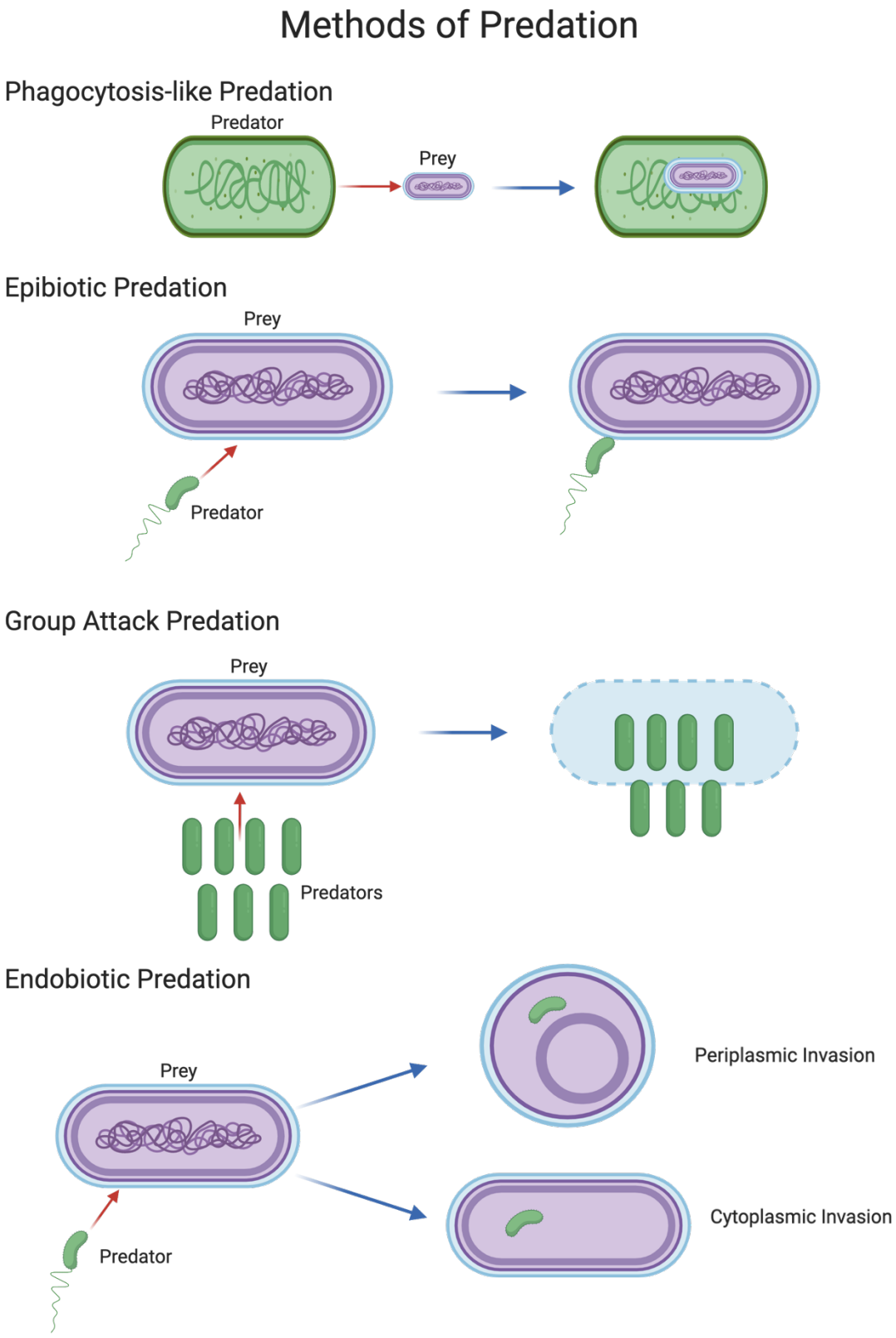


Figure 1.1 Depictions of the methods of predation used by predatory bacteria (prey shown in purple and predators in green). Red arrows show the direction of predator attack on prey. Created with BioRender.com

1.1.2.1 Phagocytosis-like Predation

This is the most recently identified type of predation, known to only be utilised by aforementioned *C. Uab amorphum*, a marine dwelling Gram-negative bacterium shown to engulf bacteria and micro-eukaryotes. Unlike previous examples of bacterial endocytosis, *C. Uab amorphum* forms a double membrane vacuole via invagination of both the outer and inner membrane. Although structurally different to a eukaryotic vacuole, *C. Uab amorphum* does seem to digest the engulfed organism in a specific compartment and utilise the substituent parts of the prey. Despite having few genes homologous to those identified in eukaryotic phagocytosis, *C. Uab amorphum* does share some similar features with phagocytotic eukaryotes, namely, a large flexible morphology and amoeboid movement (Shiratori et al., 2019).

1.1.2.2 Epibiotic Predation

Epibiotic predation is simply defined as an external attack, and is therefore the broadest group of predatory tactics.

Firstly there is 'Cell-to-Cell' predation, observed to be performed by obligate predators, *Vampirococcus, Micavibrio aeruginosavorus, Bdellovibrio exovorus* and *Vampirovibrio chlorellavorus*. This form of epibiotic predation requires the sustained attachment of the predator to the host cell throughout predation. After attachment, it is thought the predators utilise a Type IV secretion system-like structure (T4SS) to form a bridge between the predator and the host (Pérez et al., 2016). The cell contents of the host are then digested and transported into the prey cell to be used for growth and replication. This has resulted in some of the most prey-specific predators: *Vampirococcus* and *V.* chlorellavorus only predate *Chromatium* and *Chlorella* respectively. *M. aeruginosavorus* and *B. exovorus* are broader range predators, predating a few Gram-negative bacteria, but are still far more selective than most. All are water-borne, and the smallest of the predators, with an average genome size of 2.7 Mbp. As a result, the metabolic incompleteness identified upon the sequencing of *M.*

aeruginosavorus that ensures its obligate predatory lifestyle, was almost expected (Ganuza et al., 2016; Guerrero et al., 1986; Martin, 2002; Pérez et al., 2016; Wang et al., 2011).

Lysobacter carries out predation via 'Lytic Enzymes', and similar to the previous predators mentioned requires cell-to-cell contact for the commencement of predation. However, instead of maintaining contact throughout predation, *Lysobacter* lyses prey on contact, absorbing the nutrients released into the surrounding environment. It also inhabits a far more diverse environment being both water- and soil-borne, and hunts a far broader range of prey, from prokaryotes to small eukaryotes (fungi, *Oomycetes*, and nematodes) (de Bruijn et al., 2015; Christensen and Cook, 1978; Daft and Stewart, 1973; Expósito et al., 2015).

Like *Lysobacter*, facultative predator *Cupriavidus necator* uses lytic enzymes to predate its prey of prokaryotes and fungi, however they utilise 'Lysis via Diffusible Factors' that they secrete into the environment. This is a viable method for *C. necator* as it exclusively inhabits soil, and therefore the lytic enzymes are less susceptible to the environment than in the case of *Lysobacter* due to a reduced diffusion effect. *C. necator* has harnessed a method normally used to inhibit neighbouring bacterial colony growth and repurposed it to enable access to nutrients in times of starvation stress (Casida, 1989; Seccareccia et al., 2016; Pérez et al., 2016). Similarly, Gram-positive *Bacillus subtilis*, has been known to use this form of extreme antagonism to cannibalise non-*Spo0A* containing *B. subtilis* and some Gram-negative bacteria when the surrounding area has become nutritionally depleted (Pérez et al., 2016).

1.1.2.3 Group Attack Predation

The tactics employed by group attack predators are often reminiscent of epibiotic predation, as the sole distinguishing feature is that the predators must solely attack as a group and that all in the group must benefit from predation. Hence, there is some debate as to whether *Lysobacter* are group -or

epibiotic predators, as studies have shown them predating as both a group and one-on-one (Daft and Stewart, 1973).

Most similar to *Lysobacter* is *Ensifer adhaerens* which predates via the 'Picket Fence' mechanism. Similar to *Lysobacter*, *E. adhaerens* uses surface lytic enzymes to lyse prey, however *E. adhaerens* forms multi-predator rings around the prey colony, reminiscent of a picket fence. Although a ubiquitous bacterium, it was originally thought to only predate *Micrococcus luteus*, however it is now believed to predate a wider range of Gram-negative and Gram-positive bacteria (Casida, 1982; Jurkevitch, 2007; Pérez et al., 2016; Casida, 1988).

One of the most well-known predators is *Myxococcus xanthus,* a soil dwelling *δ-proteobacterium* that facultatively predates on a broad range of bacteria and some fungi via the 'co-operative predation' method. As a group-dependent predator it is consequently highly social, forming fruiting bodies and utilising two modes of gliding, one for individual hunting and another for social motility (Morgan et al., 2010; Youderian and Hartzell, 2007). This style of predation was originally denoted as wolfpack hunting, but in actuality *M. xanthus* seems to change its predatory tactics depending on its prey. When *M. xanthus* encounters a prey biofilm with an exopolysaccharide matrix high in galactoglucans it uses the wolfpack technique, surrounding the colony before entering and moving through it, lysing nearby prey cells via secreted agents. Alternatively, when this is not the case, *M. xanthus* predates via frontal attack entering at one side of the colony. Similarly, to many predators, prey contact is the required stimulus for predation, which in the case of *M. xanthus* occurs via Kil pilin complexes (a Tad-like pilin system), the nearby prey are then degraded by the release of hydrolytic enzymes into the environment. This floods the area with nutrients and as a result this form of predation is highly affected by the surrounding environment, and is therefore limited to environments with low moisture levels. Genomic analysis found about 9% of the 9.14 Mbp *M. xanthus* genome was dedicated to prey

degradation, containing about 300 genes for degradative enzymes (Beebe, 1941; Martin, 2002; Pérez et al., 2016; Thiery and Kaimer, 2020; Seef et al., 2021).

Saprospira is also a facultative predator, but predates via 'Ixotrophy' when the nutrient levels of its coastal marine habit become depleted, as like many predators it lacks the ability to biosynthesize some essential amino acids. Like *M. xanthus, Saprospira* forms multicellular complexes that move by gliding, however they catch their prey of bacteria and protists via direct contact, lysing prey in the "sticky" substance used for gliding (Lewin, 1997; Pérez et al., 2016; Saw et al., 2012).

The final type of group predation is 'Secretion of Secondary Metabolites' and has only been found to be performed by *Streptomyces.* This predation method is very similar to that performed by *C. necator*, as metabolites are secreted into the environment lysing both bacterial and fungal prey, the cellular contents that leak into the surrounding area are then used by *Streptomyces.* In contrast to *C. necator*, *Streptomyces* must form colonies in order to act in a predatory manner, this is supported by their 'social' behaviour, forming multicellular biofilms (Pérez et al., 2016).

1.1.2.4 Endobiotic Predation

Endobiotic predation is the most divergent category of predation, as the predator enters the prey cell replicating within either its cytoplasm or periplasm before lysing the prey in order to exit.

Daptobacter, a rod-shaped bacterium with a single polar flagellum, was discovered at the same time as *Vampirococcus* in sulphurous lakes. It is the only predator known to predate from within the prey cytoplasm, and only predates *Chromatiaceae*, growing in the shielded stable environment of its cytoplasm, before dividing and lysing the cell to release its progeny (Guerrero et al., 1986; Martin, 2002).

Lastly there is periplasmic predation, which is performed by *Bdellovibrio bacteriovorus, Halobacteriovorax* and *Peredibacteria*. The latter two were originally misidentified as *Bdellovibrio* strains, and therefore all three are collectively referred to as *Bdellovibrio* and like organisms (BALOs), along with *M. aeruginosavorus* and *B. exovorus* (Pasternak et al., 2014; Bonfiglio et al., 2020b). The best-characterised organism of this type is *Bdellovibrio bacteriovorus (Bdellovibrio)*. *Bdellovibrio* was discovered in 1962 by Stolp and Petzold, and originally placed in the *δ-proteobacteria* family, however it is now considered an *Oligoflexia* (Hahn et al., 2017). *Bdellovibrio* is a Gram-negative obligate predator about 0.25-0.5 by 0.5-2.5 µM in size and vibrio shaped, observed to be highly flexible and capable of both outer and inner membrane alteration (Borgnia et al., 2008). *Bdellovibrio,* like the rest of its distant relatives, is highly motile and possesses a membrane-sheathed single polar flagellum in the shape of a diminishing sine wave (Lambert et al., 2006; Seidler and Starr, 1968). The flagellum is used to move at speeds of up to 160 µm/s to swim towards prey. However, Bdellovibrio has also been observed moving in local areas using the much slower, 15–20 µm/h, process of gliding. This can be implemented to leave lysed cells, move towards prey in biofilms and inspect the prey cell before invasion (Lambert et al., 2011).

Originally thought to be purely water-borne, it was many years before the diversity of *Bdellovibrio* habitats was fully realized, and is now considered to be ubiquitous (Markelova, 2010). Sequencing of its genome in 2004 revealed a large genome for its cell volume - approximately 3.78 Mbp, hypothesised to encode 3584 proteins in a single circular chromosome (Rendulic et al., 2004). The *Bdellovibrio* genome is so large in size it expresses gyrases to add twists into the genome to enable it to be housed within the small cell volume (Butan et al., 2011; Hsieh et al., 2010). The sequencing of the genome allowed 55% of the genes to be assigned a specific function by homology, and an additional 11% are homologous to other uncharacterized proteins. This leaves 34% unassigned, presumably because they are highly specific to *Bdellovibrio* and its predatory lifestyle (Rendulic et al., 2004). Since *Bdellovibrio* is itself a Gram-negative bacterium, some of these unassigned genes will

encode protection mechanisms so that *Bdellovibrio* does not become damaged during predation, as well as to prevent one *Bdellovibrio* from trying to invade another (Lambert et al., 2015).

With only distinct relatives that predate though a variety of different mechanisms to compare it to, understanding how *Bdellovibrio* functions has been a battle slowly getting easier due to advancements in scientific techniques. As well as this, the discovery of four, far closer relatives: *Tiberius,* from the Tiber river in 2012, *Bdellovibrio SKB1291214* and *SSB218315* from Mexican soil in 2016, and *Bdellovibrio Zap7* from a drain in America in 2019 should also help provide more understanding of such a unique bacterium (Hobley et al., 2012b; Oyedara et al., 2016).

1.1.3 Bdellovibrio bacteriovorus Lifecycle

Bdellovibrio has two phases to its lifecycle, firstly an attack phase (AP) during which, as described above, *Bdellovibrio* is highly motile, and secondly a host dependent growth phase (GP). GP takes place within the host cell, and is the timepoint at which predator replication occurs. The *Bdellovibrio* lifecycle involves four distinct stages (Figure 1.2): hunting, attachment/entry, predation, and exit, and is reminiscent of a general viral lifecycle. *Bdellovibrio* is thought to naturally hunt prey which can be in both planktonic (single free-swimming cells) and biofilm (syntrophic cell masses bound to a surface) states, and has been observed doing thus *in vitro* (Kadouri and O'Toole, 2005).

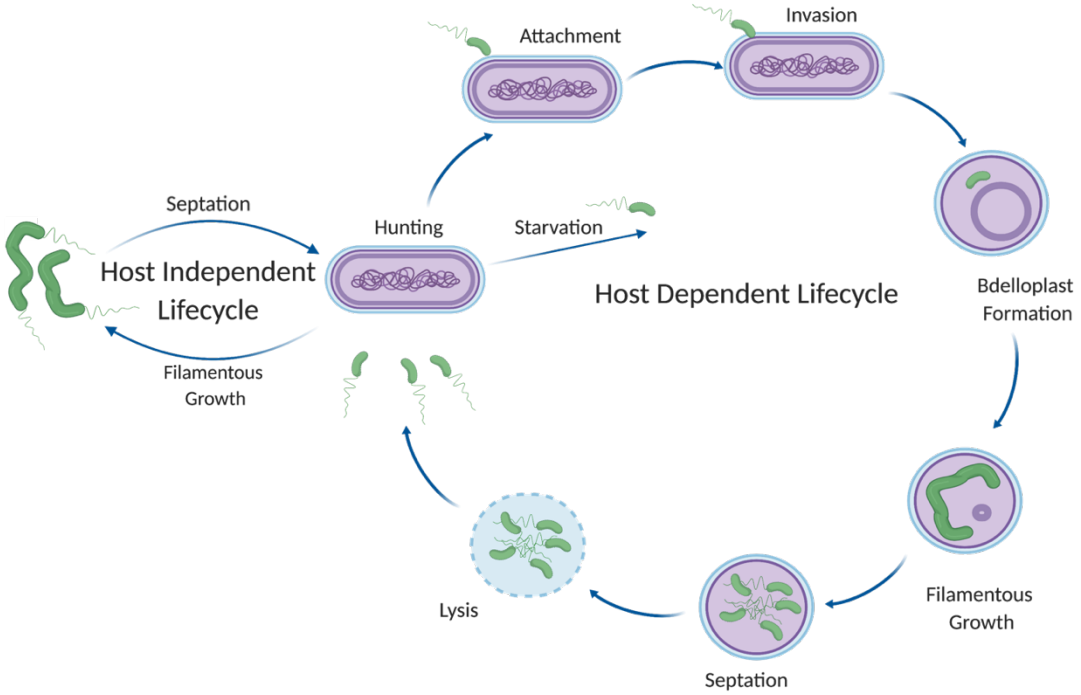


Figure 1.2 The host dependent and independent life cycle of Bdellovibrio bacteriovorus. Created with BioRender.com

After hunting down and locating a prey cell, *Bdellovibrio* attaches to cryptic prey envelope entities (Starr and Baigent, 1966). A small pore, about 200 nm in diameter, is then "bored" into the host; this has been shown to be a tight entry port for the *Bdellovibrio* (Sockett, 2009). Peptidoglycan modification by *Bdellovibrio* at the edge of the entry pore then occurs to stabilise the opening (Kuru et al., 2017). A Type IV pilus is then deployed, and it is thought *Bdellovibrio* uses this pilus to pull itself into the host cell. Electron microscopy images have also suggested that *Bdellovibrio* hydrolyses its own cell wall to assist with entry (Lambert et al., 2008). The use of a Type IV pilus for cell adhesion is well characterized in *M. xanthus* which possesses both *pilQ* and *pilT* genes, which encode a retractable Type IV pilus. *Bdellovibrio* has three genes that share sequence homology to these genes, *bd0867* to *pilQ,* and *bd1510* and *bd3852* to *pilT* (Evans et al., 2007; Lambert et al., 2006). Once inside, the pore is resealed and the host is remodelled into a bdelloplast (Kuru et al., 2017). This remodelling was originally hypothesised (~1970s era) to be due to the de-acylation of the peptidoglycan cell wall, but the remodelling is actually caused by de-crosslinking and re-crosslinking. Lambert *et al.* in 2016

identified two members of the penicillin binding protein family, Bd0816 and Bd3459, which were subsequently shown to de-crosslink the cell wall, rounding the host cell and making it an osmotically stable bdelloplast ready for *Bdellovibrio* growth (Lerner et al., 2012). The rounding has the secondary effect of also preventing a second *Bdellovibrio* from invading the host (Lambert et al., 2015). Kuru *et al.* later found that two LTDPs, Bd0886 and Bd1176, were responsible for replacing the traditional 3-4 crosslinks with 3-3 crosslinks (Kuru et al., 2017). Around the same time, Bd0468 and Bd3279 were discovered to perform the aforementioned de-acylation, however they were shown to be involved in marking the prey cell wall for progeny exit rather than bdelloplast formation (Lambert et al., 2016). Along with these penicillin binding proteins, Bd3460 was also characterised. Bd3460 was found to have an ankyrin-like fold and to be expressed in the periplasm, where it binds to Bd0816 and Bd3459. This binding inhibits the de-crosslinking of the *Bdellovibrio* cell wall, and is one of the mechanisms *Bdellovibrio* uses to protect itself during predation (Lambert et al., 2015).

The contents of the bdelloplast are then degraded by hydrolases (Lambert et al., 2006). Approximately 20 minutes post-invasion, the host cell dies due to the interruption of the electron transport chain (Kuru et al., 2017). For energy conservation, molecules are degraded and modified as little as possible before incorporation into the growing *Bdellovibrio*. This is especially true for fatty acids, the synthesis of which is especially energy-taxing because *Bdellovibrio* are lipid rich. Bdellovibrio saves about 16% of the total energy that would be required for growth by recycling fatty acids, and those that cannot be recycled get modified rather than completely degraded (Kuenen and Rittenberg, 1975). Nucleoside monophosphates (NMPs) are also not degraded, and instead are taken up intact and used for biosynthesis. Like fatty acid synthesis, the production of the phosphoester bond in NMPs is energy-taxing and by recycling these small molecules energy expenditure is further reduced (Rittenberg and Langley, 1975).

While the host is being degraded *Bdellovibrio* grows into a large helical filament. Growth of the cell wall takes place in large patches all along the edge of the filamentous cell, excluding the two poles. These areas of growth become focused as predation continues, with peptidoglycan modification occurring at the sites marked for septation (Kuru et al., 2017). Once the host is exhausted septation takes place. For a long time it was argued whether the daughter cells were produced sequentially or concurrently. However, work by Fenton *et al.* in 2010 proved the filamentous parent separated into daughter cells simultaneously. The average invasion event yields 4-6 progeny, and unlike most bacteria, *Bdellovibrio* can septate into odd numbers of cells to maximise the number of progeny produced from each predatory cycle. As the septation of the daughter cells occurs they develop flagella, and once septation has ceased the bdelloplast is lysed and the progeny are released into the environment (Burnham et al., 1970; Fenton et al., 2010).

Although this predatory lifecycle is the norm for *Bdellovibrio, some strains, like Bdellovibrio tiberius*, have been observed to be facultative predators capable of growing host-independently (HI) in nutrient rich media (Figure 1.2). This transition from host-dependent (HD) to HI is reversible in *B. tiberius*, however within the commonly studied lab strain *Bdellovibrio bacteriovorus HD100* this change is far more nuanced, requiring mutation (Capeness et al., 2013).

1.1.4 Host Independent Lifecycle

In 1992, Cotter and Thomashow identified, by recombination studies, locus mutations that cause naturally occurring HI *Bdellovibrio*. This locus was denoted the HIT, or Host-Independent locus, and mutations within led to *Bdellovibrio* that could grow in a nutrient rich environment without the need for prey (Cotter and Thomashow, 1992b). However, because all instances of host independent HD100 have been due to mutation within the HIT locus this change from HD to HI is irreversible (Capeness et al., 2013). Bd0109 is a *Bdellovibrio* recombination hotspot (RHS) protein originally identified when *Bdellovibrio* was sequenced and is encoded within the only predicted HIT locus ORF to produce the HI

phenotype. This region was found to code for *bd0108* and *bd0109* (Rendulic et al., 2004). Mutational experiments on *bd0108* and *bd0109*, showed that deletion of *bd0108* caused the attachment pili to not be deployed, whereas a naturally occurring 42 bp deletion caused greater protrusion of this pilus (Capeness et al., 2013). A *bd0109* mutant was lethal to *Bdellovibrio*. Bd0108 and Bd0109 have been shown to directly, although weakly, interact with one another, in a 3:1 stoichiometry (Capeness et al., 2013). Due to the outcome of the mutational experiments, it is thought that Bd0109 controls the formation of the pilus, whilst Bd0108 regulates Bd0109 by either sequestering Bd0109 or outcompeting pilin binding, therefore controlling pilus extrusion/retractions. This hypothesis is supported by *bd0108* and *bd0109* sharing an operon with *bd0110* and *bd0111*, which are tad pilus related genes (Capeness et al., 2013; Prehna et al., 2014). However, the link between pili loss and host independence is still unknown.

However, an accurate prediction of the role of Bd0109 within *Bdellovibrio* is hampered because the RHS family is a diverse group of proteins with a wide range of functions. A similar protein, MXAN_6679, which is almost 6x larger, is found in *M. xanthus*. It is thought to be secreted, but the function of this RHS protein, as with Bd0109, is still unknown (Youderian and Hartzell, 2007). Further research has shown that Gram-negative RHS proteins, and their distant cousins, Wall associated protein As in Gram-positive bacteria, are used to inhibit growth of neighbouring bacteria. This is done using a cleavable toxin:antitoxin region at the C-terminus, which is almost synonymous with this protein motif (Koskiniemi et al., 2013). What we do know is that although a secretory toxin could be the function within wolfpack predator *M. xanthus*, which is known to use secreted toxins for predation, Bd0109 lacks this C-terminal toxin region and it would not directly aid in the method of predation utilised by *Bdellovibrio*.

1.1.5 *Bdellovibrio bacteriovorus* – The 'Living Antibiotic'

Bdellovibrio and other predators are thought to have evolved as bacterial population regulators (Markelova, 2010). It is this predatory lifestyle, inability to invade human cells and an apparent inability of prey to form long term resistance to *Bdellovibrio* invasion, that has led to interest in *Bdellovibrio* as an antibacterial agent (Shatzkes et al., 2015). Some temporary resistance has been shown in nature, via S-layer production, but this protection is quickly lost upon the removal of the predator from the environment, and has therefore been denoted as 'plastic' resistance (Sockett, 2009). Novel antibiotics, like *Bdellovibrio*, are needed more than ever before, as we face an antimicrobial resistance crisis that could result in 10 million global deaths per year by 2050 (World Health Organisation, 2019). In February 2017, the WHO released a document laying out which antibiotic resistant bacteria they consider to be the highest threat. All of the bacteria listed as a critical concern were Gram-negative, including *Acinetobacter baumannii*, *Pseudomonas aeruginosa* and *Enterobacteriaceae* (World Health Organisation, 2017).

However, due to lack of specificity there are concerns about the long-term effects of *Bdellovibrio* on the Gram-negative bacteria naturally found in the gut. Therefore, Harini *et al.* proposed using it as a topical treatment for periodontal infection (Harini et al., 2013). In spite of these concerns, Atterbury *et al.* showed that *Bdellovibrio* had very little effect on the development of chicks, despite a shift in gut microflora to predominantly Gram-positive bacteria (Atterbury et al., 2011). Genetic studies on *Bdellovibrio* have revealed a large family of putative adhesins that are hypothesised to bind different prey, which could be used to modify *Bdellovibrio* specificity further lessening these concerns (Rendulic et al., 2004). Additional studies have shown that *Bdellovibrio* is effective at alleviating disease burden in both rats and zebrafish. Willis *et al.* showed that *Bdellovibrio* worked in tandem with the natural immune system, boosting the survival rate of zebrafish infected with a lethal dose of *Shigella* by 35% (Willis et al., 2016). Similar findings were shown by Shatzkes *et al.*, who found that both *Bdellovibrio* and *Micavibrio* could clear *Klebsiella pneumoniae* from rat lungs (Shatzkes et al., 2016). Willis *et al.*

also showed that *Bdellovibrio* were completely cleared by the zebrafish immune system within 48 hrs of administering *Bdellovibrio*. This is supported by Raghunathan *et al.* that showed *Bdellovibrio* are taken up and destroyed by macrophages, similarly to pathogenic bacteria (Raghunathan et al., 2019). Particular interest has been shown in using *Bdellovibrio* as a treatment for the chronic inflammatory disease, Crohn's. Crohn's sufferers have been shown to have an increased proportion of Gram-negative bacteria, as well as a particular persistent type of *Escherichia coli,* Adherent-Invasive *E. coli* (AIEC). The 'amphibiotic' nature of *Bdellovibrio*, being able to both balance the gut microbiome and kill AIEC make it an ideal Crohn's disease treatment (Bonfiglio et al., 2020a). However, any use of *Bdellovibrio* as an antibiotic would likely be limited to a single use as immunity within the patient towards *Bdellovibrio* is likely to arise.

In recent years there has been a shift in focus from *Bdellovibrio* being a Gram-negative only antibiotic, despite being shown to only predate Gram-negative bacteria, to a broad-spectrum antibiotic with specific focus on *Staphylococcus aureus. S. aureus*, was listed as high concern by WHO, in the above-mentioned report (World Health Organisation, 2017). In 2014, a reduction in *S. aureus* biofilms was observed when *Bdellovibrio* were applied (Iebba et al., 2014). Biofilms offer an added level of protection to bacteria from antibiotics. These findings were corroborated by Monnappa *et al.*, who reported that the addition of a HI *Bdellovibrio* caused a 75% reduction in biofilm formation and dispersal of already formed biofilms. They attributed this in part to Bd2269 and Bd2321, serine proteases secreted from *Bdellovibrio*. When they applied this treatment to a *S. aureus* infected cell line they found a 5-fold decrease in *S. aureus* invasion and no effect on the viability of MCF-10a, a human breast epithelial cell line (Monnappa et al., 2014). In 2018, this was further expanded upon when HI *Bdellovibrio* secreting Bd2269 and Bd2692 were found to be able to sustain themselves on the nutrients acquired from the degradation of *S. aureus* biofilms (Im et al., 2018). This is reminiscent of findings from Tang *et al.*, who showed that *Pseudoalteromonas* sp. strain CF6-2 could predate a variety of Gram-positive bacteria by lysing their cell wall using pseudoalterin, a metalloprotease (Tang

et al., 2020). Even though *Bdellovibrio* is not actively predating these bacteria, unlike *Pseudoalteromonas*, merely disrupting the biofilm may increase the effectiveness of traditional antibiotics on normally resistant colonies, as the extracellular matrix formed around cells in a biofilm offers increased protection against antibiotics.

As well as an antibiotic, *Bdellovibrio* is also being investigated as a food preservative to help with food security. Since *Bdellovibrio* is a Gram-negative centric predator, it would not affect the use of Gram-positive lactic acid bacteria or yeast in food production. In addition to this a study by Ottaviani *et al.* has shown that *Bdellovibrio* can significantly reduce the level of toxic and multidrug resistant *E. coli* strains on meat matrices. This study was conducted at 30 °C, with the eventual aim being to use the *Bdellovibrio* to help preserve food before and during the packing process to increase shelf life, thus reducing food waste (Ottaviani et al., 2020).

1.2 Chemotaxis

Approximately half of all *Prokaryotes* perform taxis, the sensing and movement towards or away from stimuli, and are able to sense a wide variety of these stimuli including chemicals, pH, osmolarity, temperature and redox potential, in order to locate the optimal environment (Bi and Sourjik, 2018). Although studies into bacterial taxis began in the late 1800s, it was not until the 1980s that the first structures of chemotaxis receptors (split into their periplasmic and cytoplasmic regions) and their downstream sensing proteins were elucidated (Kim et al., 1999). The most well-characterised chemotaxis pathway is that of model bacterium *Escherichia coli* where taxis are forms of biased random walk, that alternates between smooth swimming and brief periods of tumbling. *E. coli* utilises 5 chemoreceptors, 4 of which transduce the signal from the periplasm to the cytoplasm and 1 cytoplasmic receptor, along with 6 proteins that make up the cytoplasmic signalling pathway (Figure 1.3) (Bi and Sourjik, 2018). This has been expanded on with the chemosensory pathway proteins being further classified into core proteins (chemoreceptors, CheA, CheY, CheZ, CheW, CheR and CheB) and auxiliary proteins (CheC, CheD, CheV, CheX and CheZ) (Wuichet and Zhulin 2010).

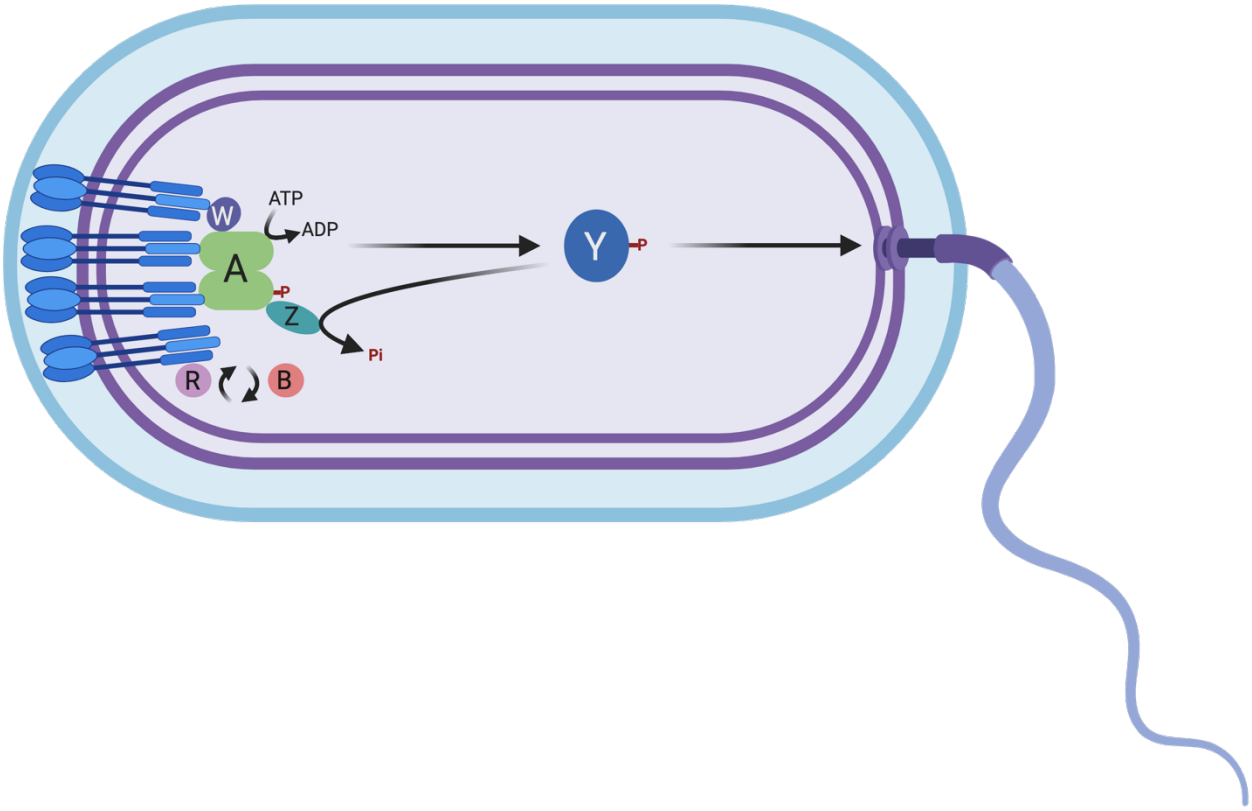


Figure 1.3 Simplified diagram of model chemotaxis components and pathway within model bacterium *Escherichia coli* (Adapted from Baker *et al.,* 2006).

Unlike in eukaryotic cells where dimers are formed upon ligand binding, the cytoplasmic region of prokaryotic chemoreceptors remain in dimeric pairs, in large chemoreceptor arrays at one or both polar ends of the bacterial cell (Kim et al., 1999). These arrays couple with shared effectors CheA and CheW to integrate the signal and increase the sensitivity of the system (Bi and Sourjik, 2018). Within model organism *E. coli* CheA, a histidine kinase protein, and CheW, an accessory protein, bind to the cytoplasmic signalling domain of the chemoreceptors at a basal rate (Wadhams and Armitage, 2004). As a result, CheA undergoes autophosphorylation, which without CheW cannot occur as it acts as a physical bridge between the chemoreceptor and CheA (Hazelbauer and Lai, 2010). CheA then phosphorylates effector protein CheY and the phospho-CheY then diffuses to the flagella C-ring protein FliM and causes the flagella to rotate in the opposite direction and the cell to tumble (Webre et al., 2003). Dissociation of CheY from the flagella returns them to normal anti-clockwise movement and the *E. coli* to swim smoothly in its new direction. Without the presence of a stimulant,

phosphorylation of CheA occurs at a basal level of 50%, which promotes frequent tumbling of the cell as well as making the chemotaxis pathway hypersensitive (Baker et al., 2006). When an attractant binds, demethylation of the receptor by CheB, a methylesterase, occurs which reduces the ability of CheA to bind to the chemoreceptor, causing a reduction in downstream phosphorylation, promoting smooth swimming in the same direction. However, CheB is also phosphorylated by CheA, and therefore due to the decrease in CheA binding and subsequent phosphorylation, there is a decrease in CheB phosphorylation and activity. This results in an increase in methylation state via the constitutively active methyltransferase CheR, which results in a return to the basal methylation state and consequently tumbling to occur at the normal rate despite no change in attractant concentration (Wadhams and Armitage, 2004). It is common for chemotaxis to adopt a sensory cascade, the pathway described here, as it allows for the integration of many different stimuli into one outcome (the changing of flagella movement) and adaption to persistent signals.

The main role of CheR and CheB is to adapt the system to a persistent chemotactic stimulus. CheR binds preferentially to the cytoplasmic helical bundles in chemoreceptors when they are in a looser conformation, which are commonly associated with a repellent, antagonist or no ligand bound states (Bi and Sourjik, 2018; Yu et al., 2015). This increases the overall level of methylation at approximately 4 specifically conserved glutamic acid residues and promotes tumbling and changes in direction. CheA is highly dependent upon this methylation for interaction with the methyl-accepting domain within chemoreceptors, and therefore increased methylation results in increased CheA activity. When an attractant binds, there is immediate inhibition of CheA activity, and phospho-CheA engages with and phosphorylates CheB instead of CheY. This phosphorylation activates CheB, enabling it to demethylate the chemoreceptor back to baseline levels, and promote smooth swimming. The kinetic rates of CheR and CheB are much lower than that of the chemoreceptors binding their chemoeffectors, and phosphorylation of CheA and CheY. This allows for a pseudo-short-term memory to be established as the current level of ligand binding can be compared to the methylation state caused by the previous

ligand binding levels. This leads to either a reduction or increase in the function of CheR and CheB (Bi and Lai, 2015; Webre et al., 2003). This temporal mechanism of gradient sensing enables bacteria to compensate for their size, which is too small to form a spatial head-to-tail gradient sensing mechanism where chemoreceptors are present at both poles of the cell. Without this pseudo-memory mechanism bacteria would be unable to establish if a chemotactic signal is getting stronger or weaker (Webre et al., 2003).

As well as the well-established flagellar motility chemosensory pathway (explained above for model organism *E.coli*), two other broad families of chemosensory pathways have been characterised based on the output controlled, Type IV pili based motility and other cellular functions. These other cellular functions include, development, biofilm formation, cell morphology, cell-cell interactions, and flagella biosynthesis (Wuichet and Zhulin, 2010). Some of these have been characterised in the predatory bacteria *M. xanthus.* Unlike the single chemosensory pathway in *E. coli*, *M. xanthus* has 8 distinct gene clusters with 8 separate CheA proteins, indicating 8 chemosensory pathways. The two best understood pathways are *Frz* and *Dif*. The *Frz* pathway is the alternative to the *E. coli* model and is responsible for reversals of *M. xanthus.* Akin to tumbling in flagellated bacteria, reversals happen at a basal rate of 7-8 times per minute, causing a random change in direction for *M. xanthus*. Similar to the *E. coli* model, methylation levels control the *Frz* pathway, however the sensing of attractants, macromolecule like peptides, causes increased methylation and reversal suppression. Conversely, binding of repellents like DMSO, causes demethylation and increased reversal. The *Dif* pathway is thought to control the production of extracellular polysaccharides (EPS) which are used in social motility to retract the type IV pilus used for s-motility. This retraction is thought to cause a positive feedback loop that increases the amount of EPS made, and in turn causes more pilus retraction. Little is known about the Che3, 4, 6 and 7 pathways, responsible for developmental gene expression, social motility regulation, type IV pilus assembly, and temperature resistance and viable spore production, respectively. Nothing is known about the final two pathways Che5 and Che8, other than that the gene

cluster for Che8 lacks an identifiable MCP. However 13 out of the 21 MCPs in *M. xanthus* are not within a gene cluster and any could control the Che8 pathway (Zusman et al., 2007).

Although the overall pathway of chemotaxis is fairly conserved across all bacteria, even in the divergent pathway of *M. xanthus*, there is variation within the chemoreceptors themselves. This variation is isolated to the periplasmic ligand binding domain (LBD), and genetic studies have revealed over 80 possible motifs/folds for this domain (Collins et al., 2014; Upadhyay et al., 2016; Ortega et al., 2017a). The majority of chemoreceptors fall into three categories; Cache (Calcium channels and Chemotaxis receptors), 4HB (4 Helical Bundle), and PAS (named for the three proteins it was found to occur in; Per, Arnt and Sim), with diminishing abundance (Ortega et al., 2017a). It is this variety in LBDs that allows for the sensing of a wide range of stimuli. The cytoplasmic region which consists of the MCP (Methyl-accepting Chemotaxis Proteins) domain is where the majority of sequence conservation occurs, as a variety of external stimuli are directed into one common output (Bi and Sourjik, 2018).

The conformational change that is thought to transduce the chemoeffector signal across the membrane is still not completely understood, with 4 possible mechanisms currently being proposed; Piston displacement, Scissoring motion, Helical rotation, and Supercoil unwinding of helices (Figure 1.4). The least favoured of these is Supercoil unwinding, which is the partial unwinding of chemoreceptor helices which in turn changes the cytoplasmic domains (Bi and Lai, 2015). The issue with this theory is that tightening of the α-helices and stabilisation of the chemoreceptor structure is often seen upon ligand binding (Yu et al., 2015). Scissoring and Helical rotation both involve rotational movements. Within the Scissoring theory, this rotation is of one monomer with respect to the other and occurs parallel to the membrane. Whereas within the Helical rotation theory, binding of the ligand causes a counter-clockwise twist in the second transmembrane helix perpendicular to the membrane. There is evidence to support both mechanisms, NpSRII from *Natronomonas pharaonis* and McpB from *Bacillus subtilis* have been proposed to use Scissoring and Helical rotation respectively (Szurmant et

al., 2004; Moukhametzianov et al., 2006). The last theory is the Piston displacement mechanism which has supporting evidence from multiple studies. The proposed principle of Piston displacement involves a small downward shift of the second transmembrane helix with respect to the first. Mutation experiments targeting the aromatic residues in the second transmembrane helix cause the shift of this helix away, towards the membrane. This suggests that these aromatics interact with the hydrophobic region of the membrane and cause the up/down shift of the helix (Bi and Lai, 2015; Ottemann et al., 1999). Molecular dynamics simulation experiments by Yu *et al.* also concluded that signal transduction occurs via a Piston displacement mechanism. However, their model also suggested that a highly linked rotation of the two monomers (Scissoring) also occurs. This is supported by similar experiments by Hall *et al.*, who looked at the cytoplasmic helices and noticed a similar piston like motion upon chemoeffector binding (Yu et al., 2015). All of the different mechanisms are proposed from studies on different LBDs, the Piston studies focus on 4 helical bundles, whereas the Helical rotation model looked chiefly at a dCache domain, and the Scissoring example originates from a phototaxis protein (Bi and Lai, 2015).

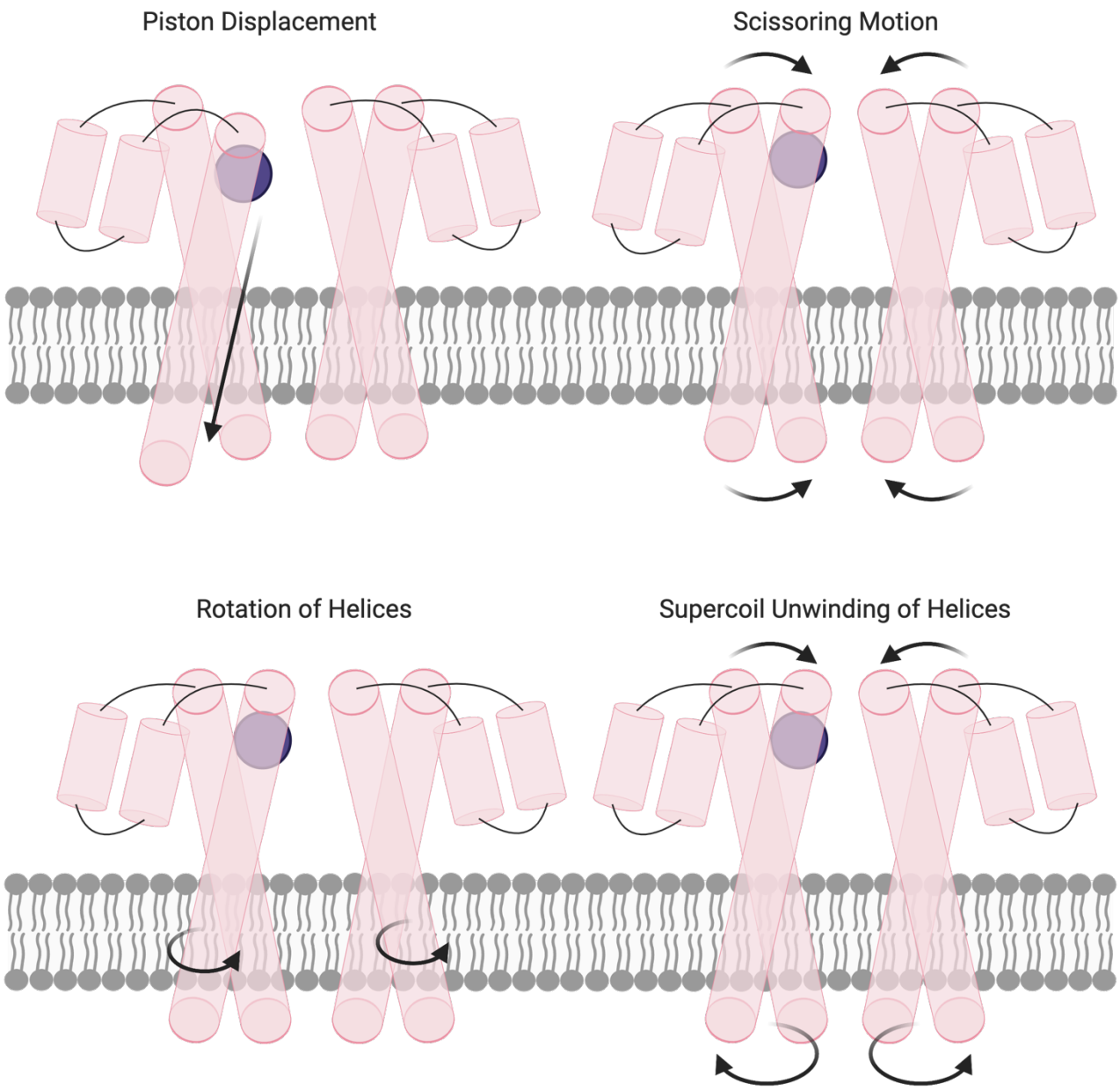


Figure 1.4 Diagram of the possible mechanism of transmembrane signal transduced performed upon ligand binding (purple) by bacterial chemoreceptors, modelled with a 4HB schematic (pink) (Adapted from Bi and Lai, 2015). Created with BioRender.com

Despite extensive research into chemotaxis, there are still gaps in our understanding of the pathway, namely how the signal is transduced across the membrane. It is therefore unsurprising that whether *Bdellovibrio* utilises chemotactic information for prey hunting is still under debate. This is because when compared with neutrophils, a bacterial hunter proven to use chemotaxis, *Bdellovibrio* are considered too fast and too small, and thus putatively unable to utilize a sensory gradient and are too

susceptible to outside forces (Jashnsaz et al., 2017a). Studies into whether or not *Bdellovibrio* hunts or randomly bumps into prey started in the late 1970s. In 1977 *Bdellovibrio* was shown to be attracted to some amino acids: Asp, Cys, Gln, Gly, His, Lys and Thr. LaMarre *et al.* also concluded that the relatively low concentrations of prey in their marine environment would make random bump an unfavourable method for survival (LaMarre et al., 1977). However, experiments by the same lab later that year could not produce a chemotactic response to a variety of whole cell bacterial strains, including known prey like *E. coli*. Therefore, they concluded that *Bdellovibrio* was in fact not chemotactic (Straley and Conti, 1977). In 1979, they expanded their small molecule chemotaxis research and found *Bdellovibrio* was attracted to acetate, propionate, thioacetate, malonate, cis-oxlate, D-glucose-6-phosphate, acetyl-coenzyme A, ammonium, barium, manganese II and potassium ions. They also found it was repelled by n-caproate, alanine, isoleucine, leucine, phenylalanine, tyrosine, cobalt II chloride and hydronium ions. Since none of the molecules found to be attractant were specific to prey waste or lysed cells, but instead were ubiquitous nutrients, it was again concluded that *Bdellovibrio* did not utilise chemotactic information to hunt prey. They instead proposed that *Bdellovibrio* predominantly relied on aerotaxis as well as chemoattraction to small ions to sustain the cell during long periods of famine, and repulsion from hydronium ions (Straley et al., 1979).

Since then the *Bdellovibrio* genome has been sequenced and 22 MCP (Garst et al., 2013; Lambert et al., 2008), transmembrane chemoreceptors capable of transducing a signal across the membrane (Ulrich and Zhulin, 2005), were identified. This has since been revised to 19 MCPs (Table 1.1) (Gumerov et al., 2020). Work by Borgnia *et al.* showed that these cluster in large arrays at the anterior pole, as is typical of most chemotaxis clusters in monoflagellated bacteria (Borgnia et al., 2008). Multiple copies of all the necessary Che proteins were also identified, with the exception of CheZ (Table 1.2). Unlike the sensory proteins these are found in gene clusters around the genome. CheZ abolishes the phospho-CheY signal but has only been identified in γ-*proteobacteria*. *Bdellovibrio* does possess one

copy of both CheZ homologues (CheC, Bd3101, and CheX, Bd1823) as well as a single copy of CheC enhancer CheD, Bd2829 (Lambert et al., 2008; Muff and Ordal, 2007). As a result of *Bdellovibrio* possessing two CheA proteins, it is predicted to have two chemosensory pathways, one of which is F7 in nature and feeds into the flagella regulatory pathway and another which is uncharacterised (Gumerov et al., 2020). It is therefore impossible to predict which MCPs are involved in chemotaxis and which are a part of the uncharacterised chemosensory pathway. *Bdellovibrio* also possesses a fully working flagellum and several gliding proteins, implying that it has the necessary machinery to sense and track prey (Garst et al., 2013).

Gene	Number of Amino Acids	Predicted Ligand Binding Domain	Genomic Context
bd0121	495	4HB	Operon with hypothetical protein, *bd0120*
bd0262	471	4HB	Operon with hypothetical protein, *bd0260*
bd0663	549	4HB	Operon with Lipase LipA, *bd0664*, and choloylglycine hydrolase, *bd0665*
bd0932	511	4HB	Operon with MerR family transcriptional, *bd0931*
bd1081	634	dCache	
bd1126	479	4HB	
bd1340	531	4HB	
bd1469	559	4HB	Overlap of final 4 nucleotides with MCP, *bd1470*
bd1470	532	4HB	Overlap of final 4 nucleotides with MCP, *bd1469*
bd2503	604	dCache	Overlap of final 4 nucleotides with MCP, *bd2504*
bd2504	601	dCache	Overlap of final 4 nucleotides with MCP, *bd2503*
bd2596	482	4HB	
bd2622	512	DUF3365	Operon With Sodium/Glutamate symport carrier, *bd2623*
bd2831	716	3xPAS	Operon with CheW2, *bd2828*, CheD, *bd2829*, CheR, *bd2830*
bd3092	484	sCache	Overlap of final 4 nucleotides with hypothetical protein, *bd3091*
bd3177	548	4HB	
bd3192	512	DUF3365	
bd3256	487	4HB	
bd3341	481	4HB	

Table 1.1 A full list of the methyl-accepting chemotaxis proteins in *Bdellovibrio bacteriovorus HD100* identified using MiST 3.0 (Gumerov et al., 2020), as well as their predicted ligand binding domain and their genomic context (Karunker et al., 2013).

Gene	Number of Amino Acids	Prediction Gene Annotation	Genomic Context
bd0102	130	*cheY*	Overlap of final 14 nucleotides with hypothetical protein, *bd0103*
bd1347	152	*cheY*	Operon with CheX, *bd1823*, hypothetical protein, *bd1822*
bd1825	154	*cheY*	Operon with hypothetical protein, *bd1822*, CheX, *bd1823*
bd2409	134	*cheY*	
bd3431	132	*cheY*	Operon with Sensory transduction protein, *bd3430*
bd0579	254	*cheW*	Operon with CheA, *bd0578*, CheB/CheD, *bd0580*, 2-hydroxymuconic semialdehyde dehydrogenase, *bd0581*
bd2828	491	*cheW2* (3x CheW domain)	Operon with CheD, *bd2829*, CheR, *bd2830*, MCP, *bd2831*
bd3471	165	*cheW*	
bd0578	578	*cheA*	Operon with CheW, *bd0579*, CheB/CheD, *bd0580*, 2-hydroxymuconic semialdehyde dehydrogenase, *bd0581*
bd3469	671	*cheA*	Operon with CheB, *bd3467*, CheR, *bd3468*
bd0580	520	*cheB/(cheD)*	Operon with CheA, *bd0578*, CheW, *bd0579*, 2-hydroxymuconic semialdehyde dehydrogenase, *bd0581*
bd3467	354	*cheB*	Operon with CheR, *bd3468*, CheA, *bd3469*
bd2830	274	*cheR*	Operon with CheW2, *bd2828*, CheD, *bd2829*, MCP, *bd2831*
bd3468	289	*cheR*	Operon with CheB, *bd3467*, CheA, *bd3469*
bd3101	313	*cheC*	Overlap of starting 4 nucleotides with hypothetical protein, *bd3102*
bd2829	156	*cheD*	Operon with CheW2, *bd2828*, CheR, *bd2830*, MCP, *bd2831*
bd1823	166	*cheX*	Operon with hypothetical protein, *bd1822*, CheY, *bd1825*

Table 1.2 A full list of the signalling proteins involved in the chemosensory pathways of *Bdellovibrio bacteriovorus HD100* (Lambert et al., 2008; Gumerov et al., 2020), as well as their predicted gene annotation and their genomic context (Karunker et al., 2013).

Knockout experiments with *mcp2*, a *Bdellovibrio 109J* chemotaxis gene homologous to *bd1469* in *HD100*, showed a reduction in predation (Lambert et al., 2006). Transposon mutagenesis studies by Medina *et al.* showed similar results when a *Bdellovibrio* 109J chemotaxis gene, locus NZ_CP007656,

(equivalent to *bd2503* in *HD100*) was disrupted, resulting in a reduction in biofilm predation from ~80% in WT to only 12.8% in the mutant (Medina et al., 2008). Yet, *Bdellovibrio* chemotaxis is still debated because it has never been demonstrated within a laboratory environment (Jashnsaz et al., 2017a). However the multiple copies of components throughout the genome are thought to protect *Bdellovibrio* from gene loss and/or mutation, therefore indicating the importance of chemotaxis to *Bdellovibrio* and its lifestyle (Lambert et al., 2008). *Bdellovibrio* has been shown to be relatively more attracted to lysed cells than whole cells (Jashnsaz et al., 2017b; Straley and Conti, 1977), aiding the theory that *Bdellovibrio* is attracted to nutrients used during famine rather than prey itself. The current working theory, proposed by Jashnsaz *et al.*, is that *Bdellovibrio* capitalises on the fact that both itself and its prey become 'geometrically captured' by surfaces, and thus it uses these hydrodynamic effects to reduce its search area, and increase the chance of randomly contacting prey (Jashnsaz et al., 2017b).

The 19 MCPs present in *Bdellovibrio* are predicted to possess 1 of 4 different LBD motifs - Cache, 4HB, PAS, and DUF3365 (a domain of unknown function) (Figure 1.5). As seen with all other chemotaxis proteins, the periplasmic domains are highly variable, with significant sequence similarity only seen between receptors with the same LBD motif. However, there is sequence conservation in the cytoplasmic region, with the majority focussed within the C-terminal signalling domain where CheA-like proteins interact. The domain is highly conserved because it is involved in the channelling of several different signalling ligands into a single outcome, control of the flagellum (Figure 1.3) (Collins et al., 2014). Three CheA proteins have been predicted in *Bdellovibrio*; Bd3469, Bd2406 and Bd0578. This indicates a higher complexity in *Bdellovibrio* compared to the model bacterium *E. coli* (Lambert et al., 2008). In actuality, *E. coli* has one of the simpler chemotaxis systems, as on average bacteria have 14 chemoreceptors, and the more diverse the habitat a species occupies the more receptors they commonly have (Ortega et al., 2017a). A better comparison model system for *Bdellovibrio* would be *Pseudomonas aeruginosa* or *Pseudomonas putida* which have 26 and 27 MCPs respectively, far

closer in number to *Bdellovibrio* than that of *E. coli*, as well as having a more diverse range of LBDs. *Pseudomonads* also have multiple CheA-like proteins that interact with their chemoreceptors and regulate a variety of different pathways. Within *P. aeruginosa* there are 4 chemosensory pathways, only one of which is involved with flagellum-mediated motility, the F6 pathway, as the F7 pathway has recently been shown to be involved in virulence (Gumerov et al., 2019, Güvener et al., 2006, Matilla et al., 2021, Ortega et al., 2017b), 1 that regulates c-di-GMP level modulation and 1 for Type IV pilus-mediated movement, which is associated with "crawling" over surfaces (Ortega et al., 2017a). This could explain why *Bdellovibrio* is predicted to have 2 CheA-like proteins as it would enable the regulation of different outcomes depending on stimuli, especially as like *Pseudomonads*, *Bdellovibrio* has been shown to glide as well as swim. Additionally, the annotation of one of *Bdellovibrio*'s chemosensory pathways as F7 strongly indicates a dependency on chemotactic information, which is supported by early chemotaxis assays (Straley and Conti, 1974; LaMarre et al., 1977; Varon and Zeigler, 1978; Straley et al., 1979), however a link between predation and chemotaxis has yet to be identified.

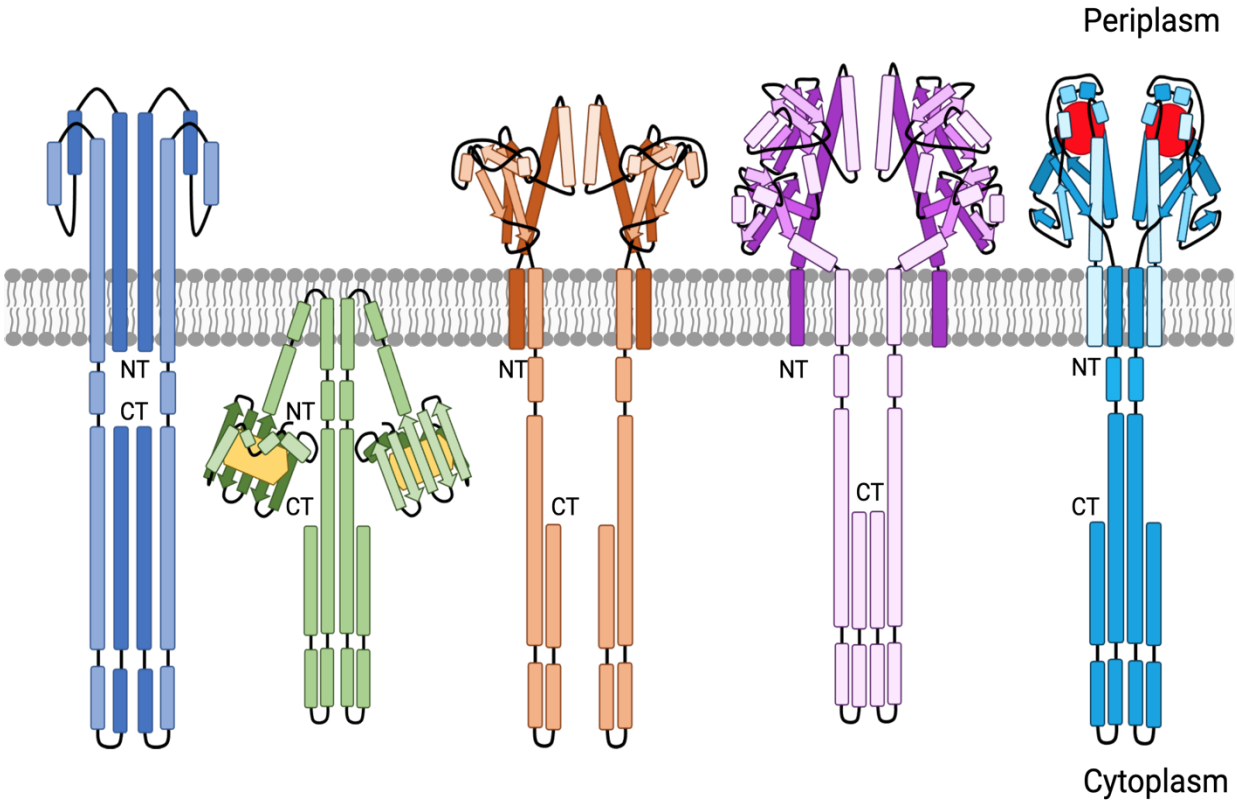


Figure 1.5 Cartoon representation of the 5 ligand binding domain in *Bdellovibrio*, left to right; 4HB, PAS with a FAD cofactor bound in the ligand binding pocket (yellow), both Cache domain variants, single (orange) and double (pink), DUF3365 with a haem cofactor in the ligand binding pocket (red). Partially created with BioRender.com

The 4HB domain is one of the best-characterised LBD due to its widespread abundance and because it is one of only two LBDs in *E. coli*. The 4HB domains have been shown to bind a wide range of chemoeffectors, either directly or indirectly via the binding of an intermediate Periplasmic Binding Protein (PBP), which complexes the chemoeffector. Binding of the ligand takes place in the interface between the two four helical bundles (Baker et al., 2006). The PAS domain is always located in the cytoplasmic region and is either held to the membrane by a membrane anchor or completely cytosolic. Using a flavin adenine dinucleotide (FAD) cofactor bound to the ligand binding pocket, the PAS domain allows for oxygen recognition via redox state changes (Ortega et al., 2017a). The Cache-superfamily is best described as either two PAS-like domains fused together, double Cache (dCache), or a single PAS-like domain, single Cache (sCache). This structural similarity has led to the reclassification of unusual periplasmic PAS receptors into the Cache-superfamily (Upadhyay et al., 2016). Although some Cache

receptors have been shown to bind PBPs they traditionally bind ligands directly (in a buried cavity), which in dCache LBDs is located in the distal lobe. Unlike 4HBs, binding to one monomer does not prevent binding to the other, because the ligand binding site is not at the dimer interface. The more enclosed nature of the binding site generally leads to a larger number of contacts with the ligand which can result in a higher ligand specificity compared to 4HBs (Ortega et al., 2017a) (although this is not always the case). The most divergent of the chemoreceptors found in *Bdellovibrio* are Bd2622 and Bd3192, which have a predicted LBD motif from the DUF3365 family. The binding domain is homologous to a haem binding protein, Tll0287, from Cyanobacteria (Motomura et al., 2017) (Figure 1.6). It is therefore thought that like the PAS domain which uses FAD to recognise oxygen levels, Bd2622 and Bd3192 use a bound haem group. However, where PAS domains indirectly sense oxygen and other ions via redox potential, the haem group in Bd2622 and Bd3192 will putatively directly bind these ions (Ortega et al., 2017a).

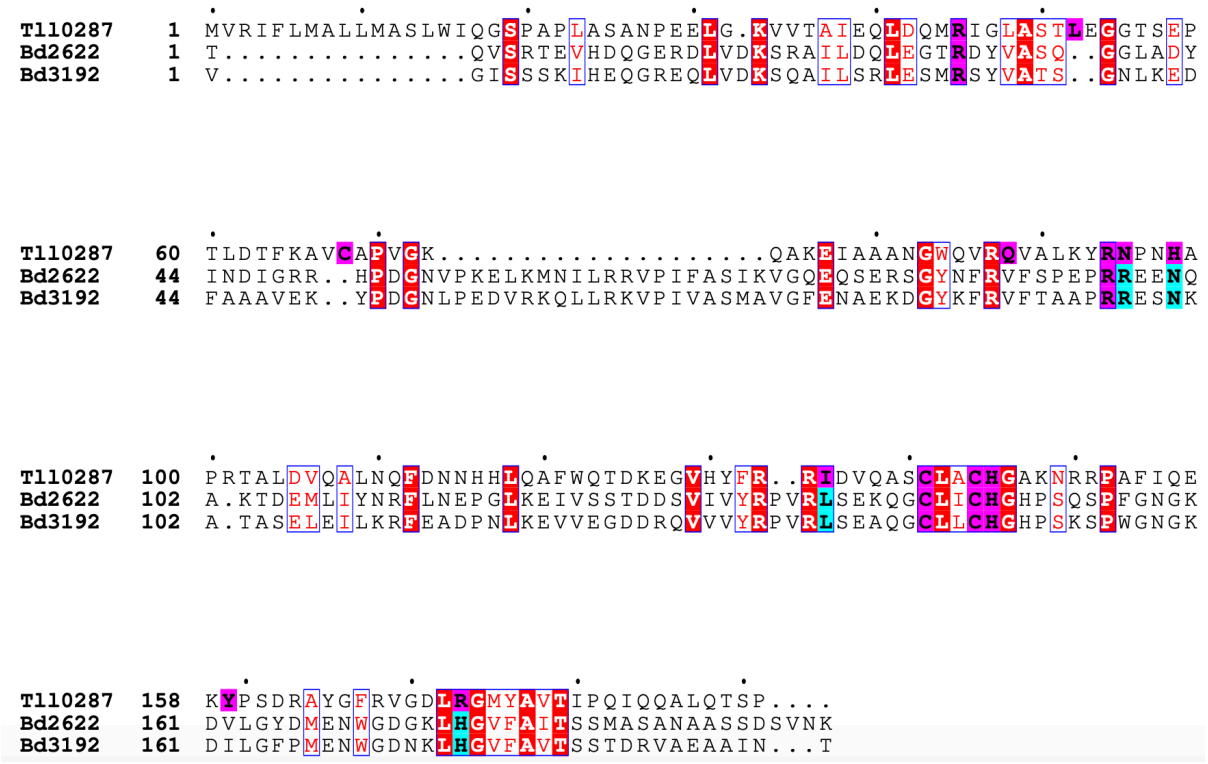


Figure 1.6 Sequence alignment of the ligand binding domains of Bd3192 and Bd2622 with haem binding protein TII0287 (5B82 (Motomura et al., 2017)) produced using ESPript 3.0 (Robert and Gouet, 2014). The binding site residues in TII0287 are highlighted in pink, with those strictly conserved in Bd2622 and Bd3129 also highlighted in pink. Amino acids in Bd2622 and Bd3129 homologous to the binding aite residues in TII0287 are highlighted in blue.

Concerning the non-MCP chemoreceptors identified via *Bdellovibrio* genomic sequencing, in 2014 *Bdellovibrio* was observed to be capable of community organisation. When in a nutrient poor environment, they will exhaust a prey supply and move on. However, if the environment is nutrient-rich the “wolves” become “sheepdogs” and hunting *Bdellovibrio* organise themselves into a predatory core, surrounded by bdelloplasts and un-predated cells. The *Bdellovibrio* sustain themselves within the nutrient rich environment and the bdelloplasts stall at the GP stage (Ferguson et al., 2014). Similar to all bacteria subject to fitness costs, *Bdellovibrio* would not have a wealth of chemotactic receptors unless these contributed to its survival. We therefore propose that at least some of these receptors may be involved in the higher-level organisation observed by Ferguson *et al.* This is supported by RNA-seq data which established the differential gene expression during AP and GP, and found some components of chemotaxis to be unregulated during GP (Karunker et al., 2013).

1.3 Commitment to Invasion

In the 1990s it was hypothesised that two signals were necessary for *Bdellovibrio* to move from AP to GP; one that switches the cell from a hunting to a growth form, and another that starts DNA synthesis to enable replication (Gray and Ruby, 1991). This theory was expanded upon by Rotem *et al.* that determined the first signal must be associated with the prey envelope, and the second with the prey cytoplasm (Rotem et al., 2016).

This physical separation of the two signals would allow time for bdelloplast formation, with the first signal indicating that invasion has started and that peptidoglycan remodelling proteins need to be expressed in order for bdelloplast formation to occur. The second signal would signify that invasion has taken place and the bdelloplast has been formed, and therefore *Bdellovibrio* can undergo growth and replication. Thus, the following mechanism for the commitment to invasion and the AP/GP switch was proposed. While in the hunting attack phase, expression of Bd0108 keeps the Type IVa pilus extended. Once stable prey association has been established, the first signal causes Bd0108 repression, causing pilus retraction. This first signal causes a shift from AP to a transitionary AP/GP phase and the repression of *mer*RNA (massively expressed riboswitch RNA) (Karunker et al., 2013)), a putative c-di-GMP sink, and the genes involved in invasion commitment and bdelloplast formation are induced. Once the *Bdellovibrio* has invaded the prey cell and bdelloplast remodelling has taken place, the secondary signal found in the prey cytoplasm diffuses into the periplasm across the now completely permeable prey inner membrane. Genes associated with the GP phase are expressed, as well as repression of flagellum-associated and bdelloplast formation genes, and the complete silencing of *mer*RNA (Rotem et al., 2016). This theory is supported by the presence of spontaneously occurring non-replicative HI cells first described in the 1970s (Cotter and Thomashow, 1992a). Further support for this came in 2014, when genetically identical host independent and dependent *B. tiberius* were naturally observed in the highly polluted Tiber river. The HI *B. tiberius* had not gone through

genetic mutation like normal HI cells, indicating that this high nutrient environment could perhaps contain the second signal (Ferguson et al., 2014).

1.3.1 Cyclic di-GMP Degradation and Receptors

Bacteria are known to use the secondary messenger cyclic-diguanylate (c-di-GMP) for a wide range of functions including the switch from planktonic to biofilm growth in *Proteobacteria* and control of pathogenic lifecycles in a range of bacteria (Tamayo et al., 2007). *Bdellovibrio* is no exception to this and has been shown to be heavily reliant on c-di-GMP, possessing 15 PilZ (c-di-GMP receptors), 5 GGDEF domain diguanylyl cyclase proteins, and 6 HDGYP proteins and 1 EAL protein, both of which are phosphodiesterases for breaking down c-di-GMP. A study by Hobley *et al.* in 2012 investigated the importance of these GGDEF domain-containing proteins and found them, and a proportion of c-di-GMP, to localize to specific cellular locations, which is thought to be the cause of the discrete GGDEF phenotypes observed in the study. Of the 5 GGDEF domain proteins identified, only 4 have an active enzymatic GGDEF domain, Bd0367 (DgcA), Bd0742 (DgcB), Bd1434 (DgcC) and Bd3766 (DgcD). Knockouts (KO) of each of these GGDEF domains produced very distinct phenotypes. Δ*bd0367* abolished gliding, resulting in *Bdellovibrio* that could enter and predate, but progeny that could not escape the post-predation detritus. Δ*bd0742* resulted in *Bdellovibrio* that could not enter and predate prey cells but had no mutations in the HIT locus and therefore were not wild type HI cells, implying Bd0742 is essential to prey entry. Δ*bd1434 Bdellovibrio* were obligate predators, and could not switch to a HI lifecycle in a protein rich/prey absent environment, and therefore Bd1434 is a likely HD to HI gate. The last GGDEF KO, Δ*bd3766*, did not result in a noticeable phenotype (Hobley et al., 2012a). These findings are supported by Karunker *et al.*, who found Bd0367, Bd0742 and Bd1434 all had AP-specific expression, whereas Bd3766 did not seem to be expressed in either the attack or growth phase (Karunker et al., 2013). The other predicted GGDEF, Bd3125, was found to be non-enzymatic. The *bd3125* KO resulted in a reduction in the speed of *Bdellovibrio* entry into the prey cell. Due to

similarities in KO phenotype between *bd0742* and *bd3125*, a link between Bd0742 and Bd3125 was established.

Along with an enzymatically active GGDEF domain, Bd0742 has an N-terminal forkhead (FHA) domain, known for binding phosphopeptides. Work by Meek *et al.*, found a 30 amino acid N-terminal tail in Bd0742, which was hypothesized to become phosphorylated when *Bdellovibrio* comes into contact with prey (Meek et al., 2019). Bd3148 is thought to be the kinase responsible for the phosphorylation of tail residue Thr14, as it is highly upregulated just prior to invasion (Karunker et al., 2013). Bd3148 also shares 26% homology to a GGDEF in distant predatory relative *Sorangium cellulosum*, whose sequence is Kinase-Tail-FHA-GGDEF, essentially forming a fused variation of the *Bdellovibrio* pathway. Phospho-Thr14 enables binding of residues L13–S19 of the N-terminal tail to the FHA cleft. This causes productive dimerisation of Bd0742 and activation of the GGDEF domain, however it is still unclear if the FHA binds its own N-terminal tail, or that of the other Bd0742 in the dimer. The subsequent c-di-GMP production is thought to cause commitment to invasion, however, it also causes auto-feedback inhibition by binding to the I-site of Bd0742 on the opposite face to the GGDEF site. Binding of c-di-GMP to the I-sites causes dimerisation of Bd0742 on the opposite face, breaking apart the GGDEF half-sites and ceasing c-di-GMP signalling (Meek et al., 2019). The phenotypically-linked Bd3125 is also thought to bind the c-di-GMP produced by Bd0742. A small subgroup of GGDEF domain proteins are known to act as c-di-GMP receptors, via this I-site, instead of using it for the classic auto-feedback inhibition. The I-site is denoted by a RxxD sequence motif, typically 5 amino acids upstream of the GGDEF domain. Bd3125 has an I-site motif, RTND, upstream from its degenerate GGDEF (GVNEF) domain. Together, Bd0742 and Bd3125 are thought to control invasion with Bd0742 controlling entry and Bd3125 controlling the speed of this process (Hobley et al., 2012a). Work by the Sockett Lab showed that Bd3125 localises at the anterior pole of *Bdellovibrio* and associates with the Type IV pili regulatory proteins RomR and TPR protein Bd2492. It is thought that this c-di-GMP pathway helps localise these proteins at the invasion pole of *Bdellovibrio* (Hobley et al., 2012a; Milner et al., 2014).

Looking at the wider genomic context of *bd0742* and *bd3125* reveals that *bd0742* is a part of a 2 gene operon with a predicted glucose-6-phosphate isomerase enzyme, *bd0741*. *bd3125*, however, is a part of a much bigger operon of 10 genes (Table 1.3), and reveals a possible secondary method for how it controls invasion speed. Within this operon are *bd3126*, *bd3123*, *bd3121* and *bd3120*. Bd3126 is predicted to be a transmembrane receptor fused to a histidine kinase, similar to CheA in the chemosensory pathway. The other three genes are homologous to components of the MlaFEDB-like complex, which is a type of ABC transporter thought to prevent accumulation of phospholipids in the outer membrane of *Proteobacteria* (Malinverni and Silhavy, 2009). Bd3123 and Bd3120 are predicted to be homologues of MlaE and MlaD respectively, and Bd3121 is predicted to have ATP-binding functionality, and therefore able to take on a similar role to MlaF. Due to this wider genomic context, a preliminary hypothesis of *Bdellovibrio* entry into prey is proposed. The N-terminal forkhead domain of Bd0742 senses prey invasion (indirectly, with a contribution from a kinase sensory domain), which activates the GGDEF domain and initiates the production of c-di-GMP, signalling to downstream effectors to commit to invasion. Meanwhile, Bd3126 binds a ligand via its periplasmic ligand binding domain, either from prey or self, which leads to the autophosphorylation of its cytoplasmic histidine kinase domain. This in turn phosphorylates the REC domain of Bd3125, which is similar to CheA phosphorylating CheY in the chemotaxis pathway. The addition of a secondary signal from Bd3126 would integrate other sensory information to the already determined invasion decision. A more tenuous extension to this hypothesis postulates that Bd3125 might then interact with the MlaFEDB-like complex of Bd3123, Bd3121 and Bd3120. Bd3125 would signal to activate this transporter complex, membrane bound Bd3120 would then shuttle phospholipids to Bd3123, which would then actively transport them into the membrane, using the ATPase function of Bd3121. This increased level of phospholipids, coupled with the already established 'softened' cell wall, would allow *Bdellovibrio* to pour itself into the prey cell, explaining how *Bdellovibrio* can enter prey through a pore smaller than its diameter and how Bd3125 might ultimately control membrane fluidity and invasion speed.

Gene	Predicted Functional Domain(s)
bd3126	Sensor Kinase containing a dCache LBD
bd3125	GGDEF and CheY-like_superfamily
bd3124	Alanine Racemase
bd3123	MlaE_permease
bd3121	ABC-type transport system ATP-binding protein
bd3120	MlaD
bd3119	TPR
bd3118	Hypothetical protein
bd3117	HD-hydrolase
bd3116	Serine acetyltransferase

Table 1.3 List of genes in the *bd3126* – *bd3116* predicted operon along with the predicted function of each gene

The proposed Bd3126/Bd3125 pathway would represent a typical two component system where the periplasmic LBD is fused to a cytoplasmic histidine kinase domain (Figure 1.7), and therefore ligand binding directly causes autophosphorylation. A two-component sensory system leading to changes in c-di-GMP levels is a common alternative to the usual helix-turn-helix domain of transcription factors as the regulatory output (Ulrich et al., 2005). However, it normally involves an enzymatically functional GGDEF protein as seen in the Wsp pathway for biofilm formation in *Pseudomonas aeruginosa*. The Wsp chemosensory system is very similar to a standard chemotaxis pathway, except the output regulator, WspR, increases the levels of c-di-GMP upon phosphorylation instead of changing the direction of flagellar rotation (Malone et al., 2007). This standard chemosensory pathway is a specialised two component system, verging on a sensory

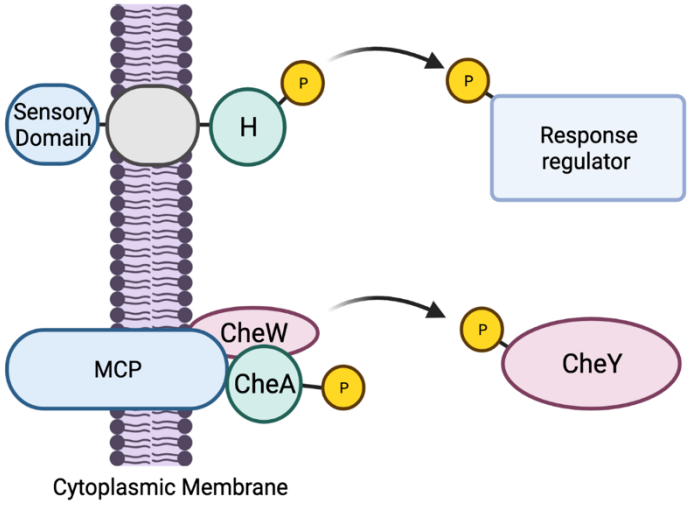


Figure 1.7 Cartoon comparing simple two component systems (top) to a simplified version of the sensory cascade used in the chemotaxis pathway (bottom) (Adapted from Kirby, 2009). Created with BioRender.com

cascade (Kirby, 2009). This increased complexity in the chemosensory pathway allows for an 'analogue' type signal that enables prokaryotes to adapt to persistent chemosensory signals. However, this adaptive sensory system would not be necessary for the 'digital decision' of invasion, and so the less complex standard two component system would be enough to channel this secondary information to the invasion pathway.

1.3.2 Attack Phase to Growth Phase Switch

Following the release of the *Bdellovibrio* genome in 2004 many studies have attempted to characterise elements of the AP/GP switch. One study using microarrays was one of the first to show the dramatic split between the AP and GP content of the genome. This study reported downregulation of typically active phase genes during growth, the top three being flagellar assembly, bacterial chemotaxis and Type III secretion. Additionally, 30 minutes into growth they saw an upregulation in ribosomal genes along with those needed for pyrimidine and purine metabolism (Lambert et al., 2010). A similar AP/GP genomic split has been found in the related BALO *M. aeruginosavorus* (Pérez et al., 2016). Building on this, Karunker *et al.* in 2013 showed an almost complete dichotomy of the genome using RNA-seq. This identified 140 AP promoters that seemed to possess a sigma-like binding site, two thirds of which were highly similar to *E. coli* sigma 28, FilA, a flagellar promoter. This suggests an evolution of FilA to become a global promoter within *Bdellovibrio*, with the rest of the genes thought to be controlled by other sigma factors (Karunker et al., 2013). Although we are starting to understand how AP genes are expressed, we know far less about the AP/GP switch.

The sequencing of the HD100 genome also revealed 18 proposed transcription factors (TFs) thought to be a part of a transcription factor superfamily predicted to contain a helix-turn-helix (HTH) motif and a DUF4423 domain (Figure 1.8). Outside of *Bdellovibrio*, DUF4423 is found in a variety of bacteria that cycle between feast and famine, the majority of which also possess a biphasic/dimorphic lifecycle, which is either predatory or a stress response to starvation (Enos et al., 2017; Jun et al., 2007; Knauber

et al., 2008; McCauley et al., 2015; Mohr et al., 2012; Mu et al., 2020; Sharma et al., 2016; Wang et al., 2020; Wiegand et al., 2018).

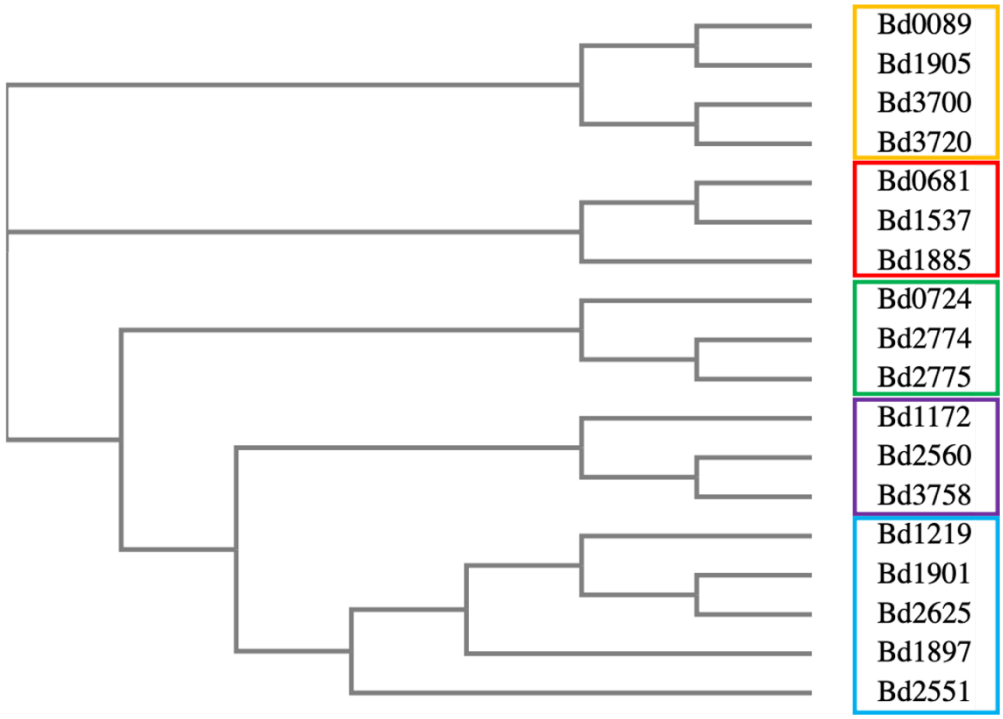


Figure 1.8 Phylogenic tree created by Tcoffee (Notredame et al., 2000) showing 18 predicted transcription factors containing the DUF4423 domain, coloured boxes denote 'subfamilies' which were determined via sequence identity.

These transcription factors were originally proposed to operate in the switch between AP and GP, as a third of them have been identified as being a part of the predatosome. A further 5 have also been shown to be downregulated in HI cell lines, even if they are not necessarily upregulated within the first 30 minutes immediately after invasion (Lambert et al., 2010). However, it is unclear whether their expression would be induced by either the first or second invasion cue. The 2013 Karunker *et al.* study identified 5 as being upregulated 3 hrs post-invasion, when compared to AP levels, again supporting the idea that this family is involved in the regulation of growth and predation (Karunker et al., 2013). Contrary to this, a new study has shown that 3 of these transcription factors, all of which had been labelled as silent by the Karunker study, are upregulated during times of predator starvation

(Sathyamoorthy et al., 2020). This may indicate that the family has a wider, environmentally dependent regulatory role.

1.4 Aims

My focus within the cycle falls on the attack phase and AP/GP switch, looking into: chemotaxis within the hunting phase, Bd3126, which we propose is involved in prey entry, and the transcription factor family that may induce the AP/GP switch.

The main aim of this project is to better understand what *Bdellovibrio* is chemotactically responsive to through structural studies of the LBD of Cache domain MCPs. The hope is to identify lifestyle specific ligands to better understand the role of chemotaxis within the *Bdellovibrio* lifecycle. The overarching purpose in using the Cache domain MCPs is to ascertain why the transposon mutagenesis study of Bd2503 gave rise to such a strongly reduced biofilm predation phenotype. Along with bound structures of the LBD of chemoreceptors and their ligand, apo structure(s) of the LBD will hopefully be obtained in order to elucidate how the signal is transduced across the membrane. This structural work will inform future *in vivo* assays performed by our collaborators using WT and knockouts, with the desire of determining whether these Cache domain MCP chemoreceptors are involved in chemotaxis, and if so, whether the ligands found are attractants or repellents.

Secondary aims for this project are to identify the ligand binding domain of the sensory protein Bd3126, and possibly characterise the ligand(s) it binds. Doing so will fill in some fundamental gaps in the proposed Bd3126/Bd3125 pathway. It will also lay down the foundations for further studies including KOs of *bd3126* and *bd3125*-CheY-like domain to see if the phenotype of slow predation persists in order to confirm the proposed additions to the invasion mechanism. Structural and biochemical studies of the proposed 18 member family of TFs in order to ascribe a function to the DUF4423 domain will also be undertaken. Further studies with this family will aim to confirm the predicted function and characterise the DNA-binding specificity of this group of putative TFs.

Chapter 2 – Materials and Methods

2.1 Media and Solutions

- All growth media was supplemented with 100 μg/mL kanamycin and 10 mL of an overnight culture for purposes of protein expression.
- LB agar – 10 g/L Tryptone, 5 g/L Yeast, 171 mM NaCl and 20 g/L Agar. This was dissolved in water and autoclaved to sterilise. It was cooled to 60 °C before kanamycin was added and the plates poured.
- Kanamycin – 100 mg/mL made with distilled water and filter sterilised.
- Auto-Induction Media (AI) – 10 g/L Tryptone, 5 g/L Yeast Extract, 25 mM Na_2HPO_4, 25mM KH_2PO_4, 50 mM NH_4Cl. Prior to use 1L of AI was supplemented with 20 mL Sugar mix, 2 mM $MgSO_4$ and 2.5 mL trace metals.
- Sugar mixes were made by mixing the sugar and glycerol with half the total volume of water, then incubated on a heated stirrer at 150 °C until the solution became clear. The remaining water was then added before autoclaving to sterilise. Sugar mixes A-E allow tuneable expression via the addition of 40% w/v glucose, with a maximum final concentration of 2 mM supplied:

Sugar Mix	Lactose (g/L)	Glucose (g/L)	Glycerol (g/L)
N	100	25	250
A	95	Variable	185
B	190	Variable	185
C	95	Variable	277.5
D	190	Variable	277.5
E	3.8	Variable	11.1

- Trace metals solution – 130 mM Na_2EDTA, 80 mM $ZnSO_4$, 180 mM H_3BO_3, 30 mM $MnCl_2$, 7 mM $CoCl_2$, 6 mM $(NH_4)_6Mo_7O_{24}$, made with distilled water and filter sterilised.
- 2x Lysogeny Broth (2xLB) – 20 g/L Tryptone, 10 g/L Yeast Extract and 171 mM NaCl

- Lysogeny Broth (LB) – 10 g/L Tryptone, 5 g/L Yeast Extract and 171 mM NaCl
- Terrific Broth (TB) – 12 g/L Tryptone, 24 g/L Yeast Extract, 54 mM K_2HPO_4, 16 mM KH_2PO_4 and 43 mM Glycerol.
- Isopropyl β-D-1-thiogalactopyranoside (IPTG) – 1 M made with distilled water and filter sterilised.
- Buffer A – 20 mM Hepes pH 7.5, 20 mM Imidazole and 300 mM NaCl
- Buffer A+B_1 – 20 mM Hepes pH 7.5, 50 mM Imidazole and 300 mM NaCl
- Buffer A+B_2 – 20 mM Hepes pH 7.5, 100 mM Imidazole and 300 mM NaCl
- Buffer B – 20 mM Hepes pH 7.5, 300 mM Imidazole and 300 mM NaCl
- 10x SDS Running buffer – 250 mM Tris Base pH 8.8, 1.92 M Glycine and 43.8 mM SDS
- Loading dye – 100 mM Tris HCl pH 6.8, 4% w/v SDS, 0.2% w/v bromophenol blue, 20% v/v glycerol and 200 mM β-mercaptoethanol.
- Gel Stain – 0.9 g Brilliant blue, 4 mL 37% v/v HCl and 996 mL dH_2O
- Dansyl-chloride solution – 28 mM dansyl-chloride dissolved in HPLC grade acetone
- 50x TAE Buffer – 2 M Tris base, 1 M Acetic acid, 22 mM Na_2EDTA
- Native Gel Loading Dye (EMSA Gel Ladder) – 62.5 mM Tris-HCl pH 6.8, 25% v/v Glycerol, 1 g/L Bromophenol blue

2.2 Primer Design and Restriction-Free Cloning

Primers were designed for: 6 transcription factors (*bd0681*, *bd1897*, *bd2625*, *bd2774*, *bd3720* and *bd3758*); *bd2625/bd2624* (a transcription factor/effector pair); and truncations of *bd2625* and *bd1885*. Primers were also designed for: the Cache domain chemoreceptor family (*bd1081*^{LBD}, *bd2504*^{LBD}, *bd3092*^{LBD}) and the *bd1081*^{LBD} homologue from *Bdellovibrio SKB1291214 B9G69_16370*^{LBD} (denoted *bd1081*$^{S\text{-}LBD}$), as well as a histidine kinase predicted to be involved in sensing invasion, *bd3126*^{LBD} (Table 2.1).

Primer	Plasmid	Forward (5' – 3')	Reverse (5' – 3')	Tm (°C)
bd0681	pET41c	GTTTAACTTTAAGAAGGAGATATACATATGAA TATATTTGATTTCAAGTCTTATAAGAC	GCTGCACTACCGCGTGGCACAAGCTTAACCCG GAAAAGATCGATATTAAAGCAATAAGC	46
bd1897	pET41c	GGTTTAACTTTAAGAAGGAGATATACATATGT CGGGACTTGAGAATCTGCCTTATTTTGC	GCTGCACTACCGCGTGGCACAAGCTTTTTTGC TTTCGGGGGAAGAAGAGGGAAAC	59
bd2625	pET41c	GTTTAACTTTAAGAAGGAGATATACATATGCA CGGCAAGCGCCCTTTGAGTGAATTTTACG	GCTGCACTACCGCGTGGCACAAGCTTTAGCGG CTTCTCACTTTCAGTAACAGTCAGAGGG	62
bd2774	pET41c	GTTTAACTTTAAGAAGGAGATATACATATGGA TCTCTCACTACTTCTGCAAGACGA	GCTGCACTACCGCGTGGCACAAGCTTGACGG GGAAGGCGGCCGTGCAGATATTGAA	58
bd3720	pET41c	GTTTAACTTTAAGAAGGAGATATACATATGGC GACGGTAGTTAAGAAACCTTTGATAC	GCTGCACTACCGCGTGGCACAAGCTTTTCTCCC ACACTTCCGTCTTTTACAGAGGAATTTTGC	57
bd3758	pET41c	GTTTAACTTTAAGAAGGAGATATACATATGAA TAAGACGTCCGAATTTCTATTAAACG	GCTGCACTACCGCGTGGCACAAGCTTTTTTTG CACCTTCGTTACCGGGAACAATTGCACG	52
bd1885-N Terminal	pET41c	GTTTAACTTTAAGAAGGAGATATACATATGAC CAAAAAACATCATTACAGAG	CTCAGTGGTGGTGGTGGTGGTGCTCGAGAGA CAAAGCCAGATCCACAAAGAATTC	49
bd1885-C Terminal	pET28a	ATGGGCAGCAGCCATCATCATCATCATCACAG CAGCGGCGGTGAAGATCAATTTGCGGTGATTT CTGAC	TCAGTGGTGGTGGTGGTGGTGCTCGAGTTAG TGTGGTTGTACGTCGGTCATTTCAAAGAA	56
bd2625/ bd2624	pET28a	GCAGCGGCCTGGTGCCGCGCGGCAGCCATAT GGGGGATTTTGTGTCTAATCAATTCG	CAGTGGTGGTGGTGGTGGTGCTCGAGCTGGG GAAATCCGCAGCGGGACTTC	54
bd2625 NT his Wing region	pET28-MBP	GAAAATCTATACTTCCAAGGATCCCTCAGTAA AACTTTGGCGGAAAAAGAGCTGGAACTGATC	GCCCCAAGGGGTTATGCTAGCTATTTGATAAA ACCCAGATCCAAAAGCAGTTCCAGACT	59
bd2625 NT his	pET28a	ATGGGCAGCAGCCATCATCATCATCATCACAG CAGCGGCGGGGATTTTGTGTCTAATCAATTCG GTGCC	TCAGTGGTGGTGGTGGTGGTGCTCGAGTTATC ATAGCGGCTTCTCACTTTCAGTAACAGTCAGA GG	67
bd2625 NT his NT to wing CT	pET28a	ATGGGCAGCAGCCATCATCATCATCATCACAG CAGCGGCGGGGATTTTGTGTCTAATCAATTCG GTGCC	TCAGTGGTGGTGGTGGTGGTGCTCGAGTTACT ATTTGATAAAACCCAGATCCAAAAGCAGTTCC AGACT	64
bd2625 NT his NT wing to CT	pET28a	ATGGGCAGCAGCCATCATCATCATCATCACAG CAGCGGCAAAACTTTGGCGGAAAAAGAGCTG GAACTGATC	TCAGTGGTGGTGGTGGTGGTGCTCGAGTTATC ATAGCGGCTTCTCACTTTCAGTAACAGTCAGA GG	65
bd1081^{LBD}	pET41c	GTTTAACTTTAAGAAGGAGATATACATATGCG GAACAGTCACGTTAACGCGGAAAAAGC	GCTGCACTACCGCGTGGCACAAGCTTGGCTTC AGCCAGAACAGTTTTGGTCG	56
bd1081$^{S\text{-}LBD}$	pET28a	AGTTTGAACACAGCAGAGCGCCAAGC	AGCATCTGCCATGACAGTCTTAAGTGGAGTAC G	62
bd2504^{LBD}	pET41c	GTTTAACTTTAAGAAGGAGATATACATATGCA GGATTCATTCAAAGACGCC	GCTGCACTACCGCGTGGCACAAGCTTGGCCCC CGCCAGAATGGTCTTTTC	55
bd3092^{LBD}	pET41c	GTTTAACTTTAAGAAGGAGATATACATATGGT GCGTGAAAGCATGTATGAAGAC	GCTGCACTACCGCGTGGCACAAGCTTATTGCG GAAGTTGGCGATGGCC	54
bd3126^{LBD}	pET28a	ATGGGCAGCAGCCATCATCATCATCATCACAG CAGCGGCACGGACAGTTTTGTCTATGGAGATC GAATGAGTTGG	TCAGTGGTGGTGGTGGTGGTGCTCGAGTTAG CGCCCCTTCATCGCCTCTTTCAG	61

Table 2.1 Table of genes cloned, the plasmid they were inserted into and the primers used. The annealing temperatures of the gene specific portion of the primers (Tm) are annotated.

Primary restriction free cloning PCR was performed using the primers from Table 2.1 to amplify specific genes using *Bdellovibrio bacteriovorus HD100/Bdellovibrio SKB1291214* genomic DNA as a template. Each reaction consisted of: 6.6 μL Phusion HiFi buffer, 3.3 μL DMSO (when GC content of target was >50%), 4mM dNTPs, 66.6 ng genomic DNA, 1.3 μL 1:10 primer mix (forward and reverse primer from Table 2.1), 0.6 units of Phusion polymerase and SDW to produce a final volume of 33 μL. The following PCR protocol was utilised:

Temperature (°C)	Time (mins)	Rounds
98	4	1
98	2	35
Tm ± 2	1.5	
72	3	
72	10	1

Table 2.2 Standard cycling steps for primary restriction free cloning PCR.

The PCR products were separated by gel electrophoresis using a 1% w/v agarose in 1x TAE gel with 1 in 10,000 Safeview stain at 100 V for 45 min. Gel samples consisted of 5 μL PCR product to 2 μL of DNA loading dye (#B7025 from NEB). If a gene product of the expected molecular weight was present on the gel, the residual PCR product was purified using Qiagen QIAquick PCR Purification kit as per the manufacturer's guidelines. The PCR product was then quantified on a microvolume spectrophotometer.

Secondary restriction free cloning PCR used 600 ng of this purified PCR product, 20 μL Phusion HiFi buffer, 4 μL dNTPs, 10 ng pET plasmid, 1 μL Phusion polymerase and SDW to produce a final volume of 100 μL. The total volume was split into 3 reactions in order to run a gradient of annealing temperatures, which were dictated by the annealing temperatures of the plasmid specific portion of the primers (Tm of pET28a – 66 °C and pET41c – 53 °C). The following PCR protocol was utilised:

Temperature (°C)	Time (mins)	Rounds
98	4	1
98	4	35
Tm ± 2	1.5	
72	10	
72	10	1

Table 2.3 Standard cycling steps for secondary restriction free cloning PCR.

The 3 reactions were then combined before adding 1 μL of DpnI and incubating at 37 °C for 2 hrs. The digested product was purified as before and concentrated to ~ 100 ng/µL using a vacuum centrifuge. 5 μL of purified and concentrated product was added to 80 μL of competent *E. coli* XL10 cells and incubated on ice for 1 hr. The cells were then heat-shocked at 42 °C for 90 secs before placing on ice for 5 min. 800 μL of LB was then added before the cells were incubated at 37 °C and 180 rpm for 1.5 hrs. The cells were then centrifuged for 3 min at 11,600 g, and the resulting pellet was resuspended with minimal LB before plating on LB agar with 100 μg/mL kanamycin and grown overnight at 37 °C.

Single colonies from this plate were then selected for colony PCR using a sterile inoculation loop and streaked onto a new gridded selection plate before being immersed in 30 μL of MyTeq red reaction mix. The MyTaq red reaction mix consisted of 4 μL generic pET F:R primers mix, 50 μL MyTaq red and 46 μL SDW. The MyTaq red reactions were then amplified using colony PCR:

Temperature (°C)	Time (secs)	Rounds
94	60	1
98	5	30
55	5	
72	40	

Table 2.4 Standard cycling steps for colony PCR.

These products were separated by gel electrophoresis in the same manner as before, but without the use of loading dye. Colonies observed to carry the desired gene construct were selected for plasmid extraction. Cultures of 10 mL LB with 100 µL/mg kanamycin were inoculated with the chosen colony and grown at 37 °C overnight. The plasmid was then extracted using a Qiagen QIAprep Spin Miniprep

Kit as per manufacturer's instructions. The concentration of the extracted plasmid was measured and an aliquot sent for sequencing (see below Section 2.4). If the desired gene was present in the extracted plasmid, transformation into the expression cell line *E. coli* BL21 took place. This was achieved as noted previously for the *E. coli* XL10 cells, the only variation being that 2 μL of plasmid was mixed with 50 μL of competent *E. coli* BL21 cells.

2.3 Q5 Site-Directed Mutagenesis

Primers for site-directed mutagenesis were designed using the online tool NEBasechanger for: *bd0723*, to stabilise the protein construct, *bd1081*$^{S\text{-}LBD}$ and *bd2625*, to stabilise oligomerisation, and prospective binding site mutations of *bd1081*$^{S\text{-}LBD}$ and *bd3720* (Table 2.5).

Primer	Plasmid	Forward (5' – 3')	Reverse (5' – 3')	Tm (°C)
***bd0723* E122A**	pET41c	GGTCCTGCACgcaCTGCTGCCAA	AGCAAAGTCCTTTGTTTTTCGTTCAACTTTTTC	69
***bd0723* E122Q**	pET41c	GGTCCTGCACcaaCTGCTGCCAA	AGCAAAGTCCTTTGTTTTTCGTTCAAC	67
***bd1081*$^{S\text{-}LBD}$ T37C**	pET41c	GATCGCAAGAtgtCTGGCGCAGCACTTGG	CCCAGTCCCTGCTCGACC	67
***bd1081*$^{S\text{-}LBD}$ D119G**	pET41c	GCCACTTATGggtTACGACAAAC	GTCAAAGTTGCTTTACCTTTTTC	61
***bd1081*$^{S\text{-}LBD}$ V125E**	pET41c	CAAACCAGGTgagGGTGATTACTAC	TCGTAATCCATAAGTGGC	59
***bd1081*$^{S\text{-}LBD}$ D119G V125E**	pET41c	accaggtgagGGTGATTACTACTTAGTTCC	ttgtcgtaaccCATAAGTGGCGTCAAAGTTG	59
***bd2625* W103C**	pET28a	TCGCCAGCTGtGAGCATTTCG	TCAGTTCCAGCTCTTTTTCC	63
***bd3720* HTH recognition helix polyA**	pET41c	ctggcggcTGTTTTTCCCGCAAGAGAAAGCTCAG	gtaggcggcGGCCATCCCGGCCTCCTT	70
***bd3720* WHTH recognition helix polyA**	pET41c	attgaggcgGCCTTGGCTTTCCTGGTCC	tgccgccgcGGAAATCCGGCCCCGCAG	70

Table 2.5 Table of point mutation site-directed mutagenesis, plasmid, primers and their annealing temperature (Tm).

Site-directed mutagenesis was performed using the following protocol with each PCR reaction consisting of 1 μL plasmid, 1.25 μL 1 in 10 dilutions of each primer, 12.5 μL 2x Q5 master mix and 10 μL SDW:

Temperature (°C)	Time (mins)	Rounds
98	1.5	1
62	0.5	25
Tm $\pm$ 2	7	
72	3	

Table 2.6 Standard cycling steps for site directed mutagenesis PCR.

A Kinase Ligase Digestion Reaction (KLDR) was then performed for 3 hrs at 20 °C, each reaction consisted of 1 μL PCR product, 1 μL T4 ligase buffer, 0.5 μL DpnI, 0.75 μL Polynucleotide kinase, 0.75 μL T4 ligase and 6 μL SDW. The product of this reaction was then transformed into *E. coli* XL10 and then *E. coli* BL21 cell lines as above.

2.4 Sequencing

Clones were sequenced by Source Bioscience to determine if cloning had been successful. 5 μL of 100 ng/L plasmid was submitted for analysis using Source Bioscience own T7 Forward and Reverse primers.

2.5 Recombinant Protein Expression

2.5.1 Glycerol Stocks

Overnight cultures of 10 mL LB with 100 μL/mg kanamycin were inoculated with a single colony and grown at 37 °C overnight. Equal volumes of overnight culture and 50% v/v glycerol were combined and flash-frozen in liquid nitrogen for storage at -80 °C.

2.5.2 Expression Trials

Overnight cultures of 10 mL LB with 100 μL/mg kanamycin were inoculated with a single colony and grown at 37 °C overnight. Trial batches of 20 mL of AI, TB, 2xLB and LB media were produced with 100 μL/mg kanamycin and inoculated with 200 μL of the overnight cultures. AI media was trialled with the full range of sugar mixes and glucose added at a final concentration of 0, 1 and 2 mM to sugar mixes A-E. The trial cultures were grown at 37 °C until an OD_{600} of 0.6-0.8 was achieved. LB, 2xLB and TB trial cultures were induced with either 0.1 or 1 mM IPTG and all trial cultures were then grown overnight at 18 °C. The OD_{600} was again measured using a 1:16 dilution. Gel samples were taken pre-induction

and after overnight growth. The samples were centrifuged for 5 min at 11,600 g and the cell pellet produced was subsequently resuspended in SDS-PAGE loading dye (amount used was calculated via Equation 1). Samples were prepared and ran on SDS-PAGE gels as described in Section 2.7. Media showing strong expression of the protein after growing overnight were selected for further large scale experiments.

$$\textbf{Equation 1}: \text{SDS} - \text{PAGE dye } (\mu\text{l}) \; = \; \frac{\text{OD}_{600} \text{ x Sample Volume (mL)}}{0.015}$$

2.5.3 Large Batch Expression

Overnight cultures of 10 mL LB with 100 μg/mL kanamycin were inoculated with a single colony and grown at 37 °C overnight. These overnight cultures were added to 1 L of media with 100 μg/mL kanamycin and shaken at 180 rpm and 37 °C until an OD_{600} of 0.6–0.8 was reached. The temperature was reduced to 18 °C, with induction via 1 mM IPTG if necessary, before growth overnight.

2.6 Purification

All proteins underwent the same IMAC purification steps. The large 1 L cultures were centrifuged at 6891 g and 4 °C for 15 mins. The pellets produced were then resuspended in 35 mL buffer A per 17 g of cell pellet. Cell lysis via sonication was performed on ice for 2 mins, 20 sec on and 20 sec off, at an intensity of 40%. The lysed cells were subsequently centrifuged at 32,826 g and 4 °C for 1hr. The supernatant was applied to Ni-NTA resin (gravity flow) equilibrated with 5 column volumes of buffer A. The column was then sequentially washed with 10 column volumes of buffer A, A+B_1 and A+B_2. The protein was then eluted in four 5 mL fractions using buffer B. SDS-PAGE gel samples were collected of each fraction with varying ratios of sample:SDS-PAGE dye: lysate and supernatant 2 μL:18 μL, flow through 5 μL:10 μL, buffer A-B_4 10 μL:10 μL, and ran on gels as stated in Section 2.7. Fractions containing protein were dialysed using 10 kDa MWCO dialysis tubing in 2 L of buffer overnight at 4 °C.

The dialysis buffer used was dependent on each proteins' stability and theoretical pI as determined by Expasy Protparam (Gasteiger et al., 2005) (Table 2.7):

Protein	pI	Buffer
Bd0089 + GFP	6.18	20 mM Hepes pH 7.5 and 500 mM NaCl
Bd1172	9.16	50 mM TS Citrate pH 6.0 and 150 mM NaCl
Bd1081^{LBD}	5.69	20 mM Hepes pH 7.5 and 300 mM NaCl
Bd1081$^{S\text{-}LBD}$	5.83	20 mM Hepes pH 7.5 and 300 mM NaCl
Bd1885	8.40	20 mM Bis Tris pH 5.5, 500 mM NaCl and 2.5 mM β- Mercaptoethanol
Bd1885 N – terminal	9.80	20 mM Bis Tris pH 6.5, 500 mM NaCl and 2.5 mM β- Mercaptoethanol
Bd1896	6.27	20 mM Bicine pH 8.0, 500 mM NaCl, 5% v/v glycerol, 2.5 mM β- Mercaptoethanol and 1 mM EDTA
Bd1897	9.61	20 mM Bis Tris pH 5.5, 500 mM NaCl and 2.5 mM β- Mercaptoethanol
Bd2503^{LBD}	4.74	20 mM Hepes pH 7.5 and 300 mM NaCl
Bd2504^{LBD}	5.19	20 mM Bis Tris pH 7.5 and 300 mM NaCl
Bd2625	9.26	20 mM Tris pH 8.5 and 500 mM NaCl
Bd2625 NT his NT to CT Wing	7.12	20 mM Bis Tris pH 6.5, 500 mM NaCl and 2.5 mM β- Mercaptoethanol
Bd3092^{LBD}	7.08	20 mM Sodium acetate pH 4.6 and 300 mM NaCl 20 mM Bicine pH 8.0 and 300 mM NaCl
Bd3126^{LBD}	9.39	20 mM Bis Tris pH 6.5 and 300 mM NaCl
Bd3720^{LBD}	7.11	20 mM Hepes pH 6.1, 500 mM NaCl, 1 mM EDTA and 5% v/v glycerol

Table 2.7 Table of buffers used for dialysis and SEC.

After dialysis, proteins were concentrated to 5–35 mg/mL using Vivaspin 20 concentrators of varying MWCO. The concentrators were spun at 3220 g and 4 °C for 20 min intervals to ensure regular resuspension to minimise protein aggregation. Final protein concentration was determined using a microvolume spectrophotometer measuring absorbance at a wavelength of 280 nm, with the extinction coefficient estimated by Expasy Protparam (Gasteiger et al., 2005) (Table 2.8).

Protein	Molecular Weight (kDa)	Extinction Coefficient M^{-1} cm^{-1}
Bd0089 + GFP	60.1	42415
Bd0723 HA	31.3	30855
Bd1172	29.2	15470
$Bd1081^{LBD}$	31.1	49390
$Bd1081^{S\text{-}LBD}$	30.8	36900
$Bd1081^{S\text{-}LBD}$ D119G	30.8	36900
$Bd1081^{S\text{-}LBD}$ V125E	30.8	36900
$Bd1081^{S\text{-}LBD}$ D119G V125E	30.8	36900
Bd1885	30.7	27055
Bd1885 NT Domain	9.19	2980
Bd1896	40.4	45965
Bd1897	30.8	28545
$Bd2503^{LBD}$	29.4	38390
$Bd2504^{LBD}$	33.1	50880
Bd2625	29.1	16960
Bd2625 NT-his NT to CT Wing	19.1	13980
$Bd3092^{LBD}$	18.3	32430
$Bd3126^{LBD}$	24.7	28420
Bd3720	33.7	38390
Bd3720 HTH recognition helix polyA	33.6	38390
Bd3720 WHTH recognition helix polyA	33.4	38390
Bd3720 Double recognition helix polyA	33.4	38390

Table 2.8 Table of Molecular Weights and Extinction Coefficients for each protein purified.

SEC was performed using either a HiLoad 26/600 Superdex 75 pg or 16/60 Superdex 200 pg column (both by GE Healthcare) depending on the molecular weight and oligomerisation state of the protein under study. The columns were ran at either 2 mL/min for the 26/600 or 1 mL/min for the 16/60 with the same buffer used for dialysis.

2.7 SDS-PAGE

SDS-PAGE gels were produced using BioRad gel-casting apparatus. The resolving layer consisted of: 2 mL 30% v/v 37.5:1 Acrylamide:Bisacrylamide, 1 mL 1.5 M Tris pH 8.8, 40 µL 10% v/v SDS, 40 µL 10% w/v APS, 4 µL TEMED, and 0.9 mL dH_2O. After allowing the resolving layer to set the stacking layer was added, this consisted of: 0.5 mL 30% v/v 37.5:1 Acrylamide:Bisacrylamide, 0.625 mL 0.5 M Tris pH 6.8, 25 µL 10% v/v SDS, 25 µL 10% w/v APS, 2.5 µL TEMED, and 1.33 mL dH_2O. All samples were prepared

with SDS-PAGE loading dye and boiled for 10 mins before loading. A volume of 10 μL was loaded for gel samples, alongside 3 μL of NEB P#7719 protein marker. All gels were ran at 180 V until 1 cm from the end (~ 1 hr). Gels were boiled for 2 min in distilled water, before boiling for 2 min with 'Gel Stain' two times. The gels were then de-stained by boiling with distilled water washes until the protein bands were clearly visible.

2.8 Differential Scanning Fluorimetry

Two types of differential scanning fluorimetry (DSF) were performed, buffer (for general stability) and additive (for iterative improvement). Both were ran in a 96-well PCR plate in a qPCR machine over a temperature course of 25–90 °C increasing by 1 °C/min. For buffer screening 30 μL of a homemade buffer screen was added to 10 μL of protein-dye mix. The protein-dye mix was made with 0.38 mg/mL protein and 1.8 μL 5000x Sypro Orange in a final volume of 1.2 mL obtained using SEC buffer. The additive screen used a 1 in 4 dilution of Molecular Dimensions Rubic screen (Boivin et al., 2013), and 40 μL of the protein-dye mix was added to 10 μL of this diluted additive screen. The protein-dye mix was made with 0.38 mg/mL protein and 1.6 μL 5000x Sypro Orange in a final volume of 4.2 mL obtained using the buffer identified by the buffer screen.

2.9 Crystallisation

Crystallography for all proteins proceeded in the same manner with all proteins at an initial screening concentration between 5-35 mg/mL. Screening used 96-well sitting drop crystallography trays, with 100 μL of commercial condition in the well and 1.5-2 μL of protein in a 1:1 ratio with the well condition in each drop. The trays were sealed with tape and left at 18 °C. Homemade screens were dependent on the initial hit from the commercial screen, but were invariably performed in a 24-well hanging drop trays. The protein concentration was kept the same as the initial hit, but the protein drop size was

increased to 4 μL and the drop ratio was varied between 2:1 to 1:2 protein:reservoir. The trays were sealed with vacuum-greased cover slips and left at 18 °C. The crystal conditions for each structure determination are provided in Table 2.9:

Protein	Screen	Drop Condition
Bd1081^{LBD} with 1 mM TMAO	MIDAS^{+} G6	0.2 M Lithium citrate tribasic tetrahydrate, 35% v/v Glycerol ethoxylate and 1 mM TMAO
Bd1081$^{S\text{-}LBD}$	JCSG^{+} H10	0.1 M Bis Tris pH 5.5, 25% w/v PEG 3350, 0.2 M Ammonium acetate, 1.1 mM DAP
Bd2503^{LBD} _Native	Homescreen around JCSG^{+} H7	0.1 M Bis Tris pH 5.7, 23% w/v PEG 3350 and 0.2 M Ammonium Sulphate
Bd2503^{LBD} with 1.2 mM Cadaverine	JCSG^{+} E10	0.1 M Bicine pH 9.0, 10% w/v PEG 6000 and 1.2 mM Cadaverine
Bd2503^{LBD} with 50 mM L-Arginine	Morpheus A4	0.1 M Imidazole/MES pH 6.5, 0.03 M $MgCl_2$, 0.03 M $CaCl_2$, 12.5% v/v MPD, 25% w/v PEG 1000 and 12.5% w/v PEG 3350, and 50 mM L-Arginine
Bd2504^{LBD} with 1.2 mM Putrescine	Homescreen around JCSG^{+} H8	0.1 M Bis Tris pH 5.5, 25% w/v PEG 3350, 0.2 M Lithium Sulphate and 1.20 mM Putrescine
Bd2625	Morpheus A7	0.1 M Hepes/MOPS pH 7.5, 0.03 M $MgCl_2$, 0.03 M $CaCl_2$, 20% v/v Glycerol, 20% w/v PEG 4000
Bd3092^{LBD}	BCS G11	0.1 M Hepes pH 7.5, 0.2 M Sodium/Potassium phosphate pH 7.5, 22,5% w/v PEG Smear Medium and 10% v/v Glycerol

Table 2.9 Table of crystallisation conditions that yielded a protein structure.

2.10 Building Crystal Structures

2.10.1 Cache Domain Chemoreceptor Structural Determination

Bd2503^{LBD} _Native crystals were soaked with a cryoprotectant containing 25 mM Cadmium chloride for 5 min before flash-cooling. Data for Bd2503^{LBD} _Native were collected at Diamond Light Source beamline i04 to 1.5 Å (Appendix I). The structure was solved using a single copy of 3LI8 (a dCache receptor from *Methanosarcina mazei* (Zhang and Hendrickson, 2010)), via molecular replacement using Phenix PHASER (McCoy et al., 2007). Due to the low TFZ scores of 7.3, the correct solution was identified using 50 rounds of density modification and auto-building in SHELXE (Spek, 2009). The structure was then built and refined using COOT (Emsley et al., 2010) and the Phenix suite (Adams et

al., 2010) (Appendix IX). Two further Bd2503^{LBD} data sets were collected at Diamond Light Source, Bd2503^{LBD} supplemented with 2 mM cadaverine on beamline i04 to 2.07 Å (Appendix II), and Bd2503^{LBD} supplemented with 50 mM L-arginine on beamline i03 to 2.26 Å (Appendix III). These structures were both solved using a single copy of the Bd2503^{LBD} _Native structure via molecular replacement using Phenix PHASER. The structures were then refined as before (Appendix IX).

Data for Bd2504^{LBD} were collected at Diamond Light Source on beamline i04 to 2.2 Å (Appendix IV). The structure was solved by obtaining the raw data from the Diamond Light Source server and truncating the dataset to only the first 2000 images because of radiation damage. This truncated range was then processed with XDS (Kabsch, 2010). Bd2503^{LBD}_Cadaverine was processed using CCP4 Sculptor (Potterton et al., 2018) to create a search model for molecular replacement using Phenix PHASER (McCoy et al., 2007). The final structure was refined in the same manner as for Bd2503^{LBD} with COOT (Emsley et al., 2010) and the Phenix suite (Adams et al., 2010) (Appendix IX).

Data for Bd1081$^{S\text{-}LBD}$ were collected to 1.55 Å at Diamond Light Source on beamline i04 (Appendix V). The structure was then solved using a single copy of the Bd2504^{LBD} distal lobe and helical backbone (range D20 to S168) as a search model to perform molecular replacement using Phenix PHASER (McCoy et al., 2007). After the distal lobe had been placed, the partial solution was fixed and the proximal lobe was placed by molecular replacement using the proximal lobe of Bd2503^{LBD} (range V15 to A28 and D182 to V274) and Phenix PHASER. The structure was then refined in an identical manner to Bd2503^{LBD} and Bd2504^{LBD} (Appendix IX). Data for Bd1081^{LBD} were collected to 2.5 Å at Diamond Light Source on beamline i03 (Appendix VII). The structure was solved using a single copy of Bd1081^{LBD} to perform molecular replacement using Phenix PHASER (McCoy et al., 2007). The structure was then refined as per the other dCache domain proteins (Appendix IX).

Data for Bd3092^{LBD} were collected at Diamond Light Source on beamline i04 to 2.3 Å (Appendix VI). Because Bd3092^{LBD} is a single Cache domain, the dCache domain proteins could not be used as a molecular replacement model. 4K08 (a chemotaxis receptor from *Anaeromyxobacter dehalogenans* (Pokkuluri et al., 2013)) was processed with CCP4 Sculptor (Potterton et al., 2018) to produce a search model. Building and refinement of Bd3092^{LBD} was carried out in the same manner as the other Cache domain proteins (Appendix IX).

2.10.2 Structural Determination of Partial Bd2625 Structure

A Bd2625 crystal was soaked with cryoprotectant containing 5 mM Platinum(IV) chloride for 1 hr before flash-cooling in liquid nitrogen. Data for Bd2625 were collected to 2.07 Å at Diamond Light Source on beamline i03 (Appendix VIII). Due to a lack of structural homology between Bd2625 and existing PDB structures, molecular replacement was not possible and so classical heavy atom phasing was used, via single wavelength anomalous dispersion (SAD). Phenix Autosol (Terwilliger et al., 2009) was used to determine the positions of the platinum atoms soaked into Bd2625, and obtain an initial map (FOM=0.204) and structure. This initial structure was improved using Phenix Autobuild (Terwilliger et al., 2008), and refined in the same manner as the previous structures (Appendix IX).

2.11 High-performance liquid chromatography

High-performance liquid chromatography (HPLC) was used to confirm the presence of polyamines bound to Bd2503^{LBD} and Bd2504^{LBD}. The polyamines were derivatised with dansyl-chloride in order to detect them via HPLC. 150 μL samples of protein and pure polyamine at 250 μM were boiled at 100 °C for 10 mins before centrifugation at 11,600 g for 5 mins. 50 μL of these samples were then added to 100 μL of 100 mM sodium bicarbonate and 200 μL dansyl-chloride solution. After the addition of dansyl-chloride solution the samples were kept in the dark to avoid photobleaching. The samples were agitated for 30 secs and then heated at 60 °C for 5 mins. In order to quench the excess dansyl-chloride

50 μL of 100 mg/ml L-proline was added and left to equilibrate for 30 min. The derivatised polyamines were extracted using 250 μL of toluene, agitated for 1 min before centrifugation at 11,600 g and 4 °C for 10 mins. This process was repeated three times with the organic phase being extracted each time. The samples were dried before resuspension with 50 μL HPLC grade acetonitrile. The HPLC was ran at 254 nm with a C18 column at 40 °C. A gradient of 0.1% v/v formic acid in dH_2O to 0.1% v/v formic acid in acetonitrile was run over 20 mins with a flow rate of 0.175 mL/min.

2.12 Isothermal Titration Calorimetry

Isothermal Titration Calorimetry (ITC) was used to determine the binding specificity of the Cache chemoreceptors: $Bd1081^{LBD}$ HD100, SKB1291214 homologues and mutants, $Bd2503^{LBD}$, $Bd2504^{LBD}$ and $Bd3092^{LBD}$ (Table 2.11). All proteins were putatively converted to the apo-form via three rounds of dialysis, established via HPLC, before dialysing into the appropriate ITC buffer (Table 2.10):

Protein	Buffer
$Bd1081^{LBD}$	20 mM Hepes pH 7.5 and 150 mM NaCl
$Bd1081^{S\text{-}LBD}$	20 mM Hepes pH 7.5 and 150 mM NaCl
$Bd1081^{S\text{-}LBD}$ D119G	20 mM Hepes pH 7.5 and 150 mM NaCl
$Bd1081^{S\text{-}LBD}$ V125E	20 mM Hepes pH 7.5 and 150 mM NaCl
$Bd1081^{S\text{-}LBD}$ D119G V125E	20 mM Hepes pH 7.5 and 150 mM NaCl
$Bd2503^{LBD}$	20 mM Hepes pH 7.5 and 150 mM NaCl
$Bd2504^{LBD}$	20 mM Bis Tris pH 7.5 and 150 mM NaCl
$Bd3092^{LBD}$	20 mM Bicine pH 8.0 and 150 mM NaCl
$Bd3126^{LBD}$	20 mM Hepes pH 7.5 and 150 mM NaCl

Table 2.10 Table of buffers used for ITC for each protein construct.

Ligand	Protein								
	Bd1081 LBD	Bd1081^{S-} LBD	Bd1081^{S-} LBD D119G	Bd1081^{S-} LBD V125E	Bd1081^{S-} LBD Double	Bd2503 LBD	Bd2504 LBD	Bd3092 LBD	Bd3126 LBD
1,3-Diamino-propane									
Acetate									
Arginine									
Argininio-succinate									
Betaine									
Bis tris									
Cadaverine									
Choline									
Diaminopimelic acid									
Ethanolamine									
Histidine									
Lysine									
Norspermidine									
Pentanediol									
Propionic acid									
Putrescine									
Spermidine									
Thioacetate									
TMAO									
Urea									

Table 2.11 Table of ITC experiments ran. Binding experiments highlighted in blue were performed, those in grey were not.

The experiments were ran on a VP-ITC Micro-Calorimeter with proteins at a concentration of 80-150 μM and ligands at a concentration of 600-1000 μM. The injection volume was 8 μL with a 4 min recovery between each injection, for a total of 17 injections. An initial delay of 1 min was set, with a Differential Power (DP) of 15 set for the reference cell, and the machine temperature set to either 15 or 25 °C. All ligand samples were prepared using the same buffer used for protein dialysis to reduce noise. Protein and ligand samples were degassed for 45 mins before loading. All datasets were analysed using Origin 7.0 using 'One Set of Sites' curve fitting model (Edwards, 2002).

2.13 Electrophoretic Mobility Shift Assay

Electrophoretic Mobility Shift Assays (EMSAs) were ran to determine the DNA binding capability of the family of 18 transcription factors thought to be involved in the switch from AP to GP (DNA fragment sequences in Appendix X). All DNA fragments were produced using the primary restriction free cloning outlined in section 2.2 (Primers in Table 2.12). All DNA fragments were purified using the

Qiagen PCR purification kit as per the manufacturer's instructions and concentrations were measured using a microvolume spectrophotometer before use in the assays.

Upstream Region of Protein	Forward Primer 5' – 3'	Reverse Primer 5' – 3'	Tm (°C)	Length (bp)
bd0089^{U}	gaaacagtttctggcagtcta	tccttcagaaagaaggtgaca	61	254
bd1172^{U1}	gaagttgtcggcatctcgcgtttg	agacccaacctattttagggcagt	66	340
bd1172^{U2}	aaatgaacctccgcaacaatgtgc	agacccaacctattttagggcagt	66	164
bd1885^{U1}	cacgaagctcatcgcgatagtggg	gagagctcctctttagtgaggctc	66	340
bd1885^{U2}	ttaaagcgtcctcaacacgcaacc	gagagctcctctttagtgaggctc	66	186
*bd1885*NT	gtttaactttaagaaggagatatacatat gaccaaaaaacatcattacagag	ctcagtggtggtggtggtggtgctcgagag acaaagccagatccacaaagaattc	61	291
bd2625^{U}	attgcagcgttccctttgtttag	cgacctgagggatttttttacattt	62	91
bd2626^{U}	ccgttatcttccctttgatgggattg	gttacgatctccgtcctctcattc	64	224
bd3721^{U}	aagtcggtgtcctttgtgattgga	agtccctccctgttaaagaatccg	66	361
gltF	tatggtaccatgaaatacgtattgatgatgctt gtg	tattctagatcatggatttttcacctcaccaag	67	264

Table 2.12 Table of the primers used to obtain DNA fragments for EMSAs, along with the Tm used to produce them via PCR and the length of the resulting DNA fragment. All fragments were named after the gene upstream.

Gels of varying percentage, depending on DNA fragment length, were produced as per the table below (Table 2.13):

Gel percentage	30% v/v Acrylamide (mL)	5x TAE (mL)	10% w/v APS (µL)	TEMED (µL)	dH_2O (mL)
5	4.15	2.5	250	50	18.05
8	6.66	2.5	250	50	15.54
10	8.33	2.5	250	50	13.87
20	16.6	2.5	250	50	5.6

Table 2.13 Table of EMSA gel recipes.

DNA was used at a constant final amount of 25 ng loaded onto the gel, and the range of protein concentration used was dependent on the EMSA being performed. Gels were pre-run for 45 mins at 80 V to equilibrate the gel and remove any unreacted APS. DNA:protein reactions had a total volume of 20 µL and were incubated at 20 °C for 30 mins before loading 10 µL onto the gel. Lanes containing DNA only, as a control, and native loading dye, as a substitute for a gel ladder, were included. The gels were ran at 80 V for 30 mins and then at 120 V until the dye front was 1 cm from the bottom of the

gel. The gels were then stained for 30 mins with 60 mL 1x TAE and 1/10,000 ethidium bromide. The gels were then de-stained by washing with distilled water, once for 20 min and a further 3 times for 15 mins. Gels were imaged using a Bio-Rad Molecular Imager Gel Doc XR set for ethidium bromide with filter 1 and faint band optimisation.

2.13.1 Screening EMSAs

Screening EMSAs of Bd1172, Bd1885 and Bd2625, with the DNA fragments of upstream regions were performed to establish if DNA binding occurred (Table 2.14). This was not performed for Bd3720 as previous studies (performed by Charlotte Manning while in the Lovering Lab) had already demonstrated binding.

Ratio DNA:Protein	Protein (μM)					
	Bd1172		Bd1885		Bd2625	
	DNA Fragment					
	bd1171^{U1}	*bd1171*^{U2}	*bd1885*^{U1}	*bd1885*^{U2}	*bd2625*^{U}	*bd2626*^{U}
1:0	0	0	0	0	0	0
1:0.001	0.00001	0.00002	0.00001	0.00002	0.00008	0.00003
1:0.01	0.00011	0.00023	0.00011	0.00020	0.00083	0.00034
1:0.1	0.0011	0.0023	0.0011	0.0020	0.0083	0.0034
1:1	0.011	0.023	0.011	0.020	0.083	0.034
1:10	0.11	0.23	0.11	0.20	0.83	0.34
1:100	1.11	2.31	1.11	2.03	8.33	3.38
1:1000	11.1	23.1	11.1	20.3	83.3	33.8
	Gel Percentage					
	8	10	8	10	20	10

Table 2.14 Table of screening EMSAs ran for Bd1172, Bd1885 and Bd2625 with the DNA fragments upstream, and the percentage of the gel used. Listed in the table are the concentration of protein used, the concentration of protein at a ratio of 1:1 represents the concentration of DNA in each reaction.

2.13.2 Qualitative EMSAs

Once binding had been established, refined qualitative EMSAs were ran in triplicate, across a narrower protein concentration range, to establish a preliminary *Kd* (Table 2.15). The bound fraction was calculated by determining band intensity of the Free DNA in each well using Fiji (Schindelin et al., 2012), where the DNA only control was considered representative of the Total DNA. The bound fraction was then determined by subtracting the Free DNA from the Total DNA. The bound fraction was then subsequently plotted against protein concentration, and the *Kd* determined using GraphPad Prism 9.0.0 'One site – total' or 'Specific binding with Hill slope' (www.graphpad.com).

Protein (μM)		
Bd1172	**Bd1885**	**Bd3720**
0	0	0
0.05	0.05	1.00
0.10	0.10	1.10
0.15	0.15	1.20
0.20	0.20	1.30
0.25	0.25	1.40
0.40	0.40	1.50
0.55	0.55	1.60
0.75	0.75	1.70
1.00	1.00	1.80
1.75	1.75	2.00
2.50	2.50	2.50
5.00	5.00	5.00
DNA Fragment (nM)		
***bd1171*^{U2}**	***bd1885*^{U2}**	***bd3720*^{U}**
23.10	20.36	10.49
Gel Percentage		
10	10	6

Table 2.15 Table of qualitative EMSAs ran for Bd1172, Bd1885 and 3720 with the DNA fragment upstream, and the percentage of the gel used.

2.13.3 Specificity of Bd3720

To understand the specificity of this transcription factor family the ability of Bd3720 to bind to distal DNA fragments (not directly upstream) was also tested (Table 2.16).

Ratio DNA:Protein	DNA Fragment Tested (μM of Bd3720)				
	bd0089^{U}	*bd1171*^{U1}	*bd1885*^{U1}	*bd2625*^{U}	*bd2626*^{U}
1:0	0	0	0	0	0
1:0.001	0.00001	0.00001	0.00001	0.00004	0.00002
1:0.01	0.00015	0.00011	0.00011	0.00042	0.00017
1:0.1	0.0015	0.0011	0.0011	0.0042	0.0017
1:1	0.015	0.011	0.011	0.042	0.017
1:10	0.15	0.11	0.11	0.42	0.17
1:100	1.49	1.11	1.11	4.16	1.69
1:1000	14.9	11.1	11.1	41.6	16.9
Gel Percentage					
	10	8	8	20	10

Table 2.16 Table of screening EMSAs ran for Bd3720 with a variety of DNA fragments and the percentage of the gel used. Listed in the table are the concentration of Bd3720 used, the concentration of protein at a ratio of 1:1 represents the concentration of DNA in each reaction.

To further access specificity, two additional gels were ran (Table 2.14). Gel 1 compared the binding capacity of Bd3720 to Bd3092^{LBD}, a protein that had been characterised as a non-DNA binding chemoreceptor. Bd3092^{LBD} was used as a control as it carries an overall similar charge to Bd3720, and had been shown to form a dimer of similar molecular weight to monomeric Bd3720. Gel 1 also assessed the ability of Bd3720 to bind to a coding DNA fragment, *gltF,* to further assess the specificity of Bd3720. Gel 2 was similar to the Bd3720:*gltF* experiment, but tested the ability of Bd3720 to bind to a section of DNA encoding a different member of the transcription factor family, *bd1885NT*. The concentration of Bd3720 was maintained at 5 μM across all experiments. This concentration of Bd3720 was in slight excess of what had been observed as necessary to achieve full binding of 25 ng of DNA in previous experiments, therefore allowing for a clear result to be obtained. All DNA was loaded with a final amount of 25 ng in the well, as this amount had been shown to be easily discernible on a gel post staining.

Gel 1			
Well	**Protein**	**Protein (μM)**	**DNA**
1	Ladder		
2	Blank		
3	-	0	*bd3721*^{U}
4	Bd3720	5	*bd3721*^{U}
5	Bd3092^{LBD}	5	*bd3721*^{U}
6	-	0	*GltF*
7	Bd3720	5	*GltF*
Gel 2			
Well	**Protein**	**Protein (μM)**	**DNA**
1	Ladder		
2	Blank		
3	-	0	*bd3721*^{U}
4	Bd3720	5	*bd3721*^{U}
5	-	0	*bd1885N*^{U}
6	Bd3720	5	*bd1885N*^{U}

Table 2.17 Table of screening EMSAs ran to further access the specificity of Bd3720. Gel 1 tested if overall charge was responsible for DNA binding, rather than a specific motif, by comparing Bd3720 to Bd3092^{LBD}. Gel 1 and Gel 2 further accessed the specificity of Bd3720 to a specific DNA sequence, by testing the binding capacity of Bd3720 with coding DNA fragments.

2.13.4 Identifying DNA Sequences/Motifs Responsible for the Binding of Bd3720

In order to identify a possible specific sequence recognised by Bd3720, fragments that represented 50 bp sequential truncations of *bd3721*^{U} were produced. 5' truncations of *bd3721*^{U} were produced using new forward primers and the original *bd3721*^{U} reverse primer, and 3' truncations of *bd3721*^{U} were produced using new reverse primers and the original *bd3721*^{U} forward primer (Table 2.18). To ensure that the smaller fragments would still be easily visualised on the gel, the concentration of DNA was increased to a final amount of 50 ng per well, this was kept constant for all DNA lengths. In order to maintain the established ratio of DNA:Protein, the concentration of Bd3720 was also increased to a final concentration of 10 μM (Table 2.19).

Forward Primers 5’ – 3’		Tm (°C)	Length (bp)
F1	ccctgccgcgatgacaaa	66	309
F2	ctggttgagggcctcggg	67	259
F3	ttgatcccgcgatgagacgc	67	213
F4	actagataaatactgaacacgaaaatta	61	164
F5	ttttgtggagaatggcgataa	61	115
F6	tcgataccgattcaacccgaa	64	66
Reverse Primers 5’ – 3’		**Tm (°C)**	**Length (bp)**
R1	gggttgaatcggtatcgaact	63	313
R2	tcgccattctccacaaaaagt	63	264
R3	cgtgttcagtatttatctagttaaattgt	61	218
R4	tctcatcgcgggatcaattga	64	165
R5	aggccctcaaccagccc	67	116
R6	atcgcggcaggggcatt	67	64

Table 2.18 Table of the primers used to obtain truncated DNA fragments of *bd3721*^{U} for EMSA experiments, along with the Tm used to produce them via PCR, and the length of the fragment yielded. All fragments produced were named after the specific truncation primer used.

Large Fragments – 8% gel		
Well	**Protein (μM)**	**DNA**
1	Ladder	
2	Blank	
3	0	F1
4	10	F1
5	0	F2
6	10	F2
7	0	R1
8	10	R1
9	0	R2
10	10	R2
Medium Fragments – 10% gel		
Well	**Protein (μM)**	**DNA**
1	Ladder	
2	Blank	
3	0	F3
4	10	F3
5	0	F4
6	10	F4
7	0	R3
8	10	R3
9	0	R4
10	10	R4
Small Fragments – 20% gel		
Well	**Protein (μM)**	**DNA**
1	Ladder	
2	Blank	
3	0	F5
4	10	F5
5	0	F6
6	10	F6
7	0	R5
8	10	R5
9	0	R6
10	10	R6

Table 2.19 Table of gels ran to identify specific sequences that Bd3720 binds to in *bd3721*^{U}. Each DNA truncation is named after the primer used.

2.13.5 Competition EMSAs

In order to determine the preliminary *Kd* for specific binding of Bd3720 to DNA EMSAs with a non-specific competitor, poly(deoxyinosinic-deoxycytidylic) (poly dI-dC), were performed (Table 2.20). This assay was performed in the same manner as the Bd3720 qualitative EMSAs but with the addition of poly dI-dC at a final amount of 50 ng loaded onto the gel.

Concentration of Bd3720 in Well (μM)														
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**
Dye	-	0	0.05	0.10	0.15	0.20	0.25	0.40	0.55	0.75	1.00	1.75	2.50	5.00
***bd3721^{U}* (ng)**														
25														
poly(dI-dC) (ng)														
50														
Gel Percentage														
8														

Table 2.20 Table of competition EMSAs to remove non-specific binding and establish a preliminary *Kd* for the specific binding interaction between Bd3720 and *bd3721^{U}*.

2.13.6 Bd3720 and Bd1885 mutant EMSAs

With the intention of determining whether one or both DNA binding domains were involved in DNA binding, screening EMSAs were ran with mutated transcription factors (Table 2.21). Initially this assay was performed with full length Bd3720 with polyA mutations designed to remove the amino acids thought to be required for the specific binding of DNA to the recognition helices in both the HTH and DUF4423 domains. A double mutant carrying both sets of polyA mutations was also produced and tested. Subsequently this assay was performed using the N-terminal HTH domain of Bd1885.

Ratio DNA:Protein	Protein			
	Bd1885NT Domain	Bd3720 HTH recognition helix polyA	Bd3720 WHTH recognition helix polyA	bd3720 Double recognition helix polyA
	DNA Fragment			
	bd1885^{U}	*Bd3721^{U}*		
1:0	0	0	0	0
1:0.001	0.00001	0.00001	0.00001	0.00001
1:0.01	0.00011	0.0001	0.0001	0.0001
1:0.1	0.0011	0.001	0.001	0.001
1:1	0.011	0.01	0.01	0.01
1:10	0.11	0.1	0.1	0.1
1:100	1.11	1.05	1.05	1.05
1:1000	11.1	10.5	10.5	10.5
Gel Percentage				
	8	8	8	8

Table 2.21 Table of screening EMSAs ran for Bd1885NT and Bd3720 with polyA mutations to the HTH, WHTH and both recognition helices. All EMSAs were all ran with the DNA fragment upstream. Listed in the table are the concentration of protein used, the concentration of protein at a ratio of 1:1 represents the concentration of DNA in each reaction.

2.14 Circular Dichroism Spectroscopy

Bd3720 WT and polyA recognition helix mutants were dialysed into 10 mM Sodium phosphate pH 7.5 and 100 mM ammonium sulphate to reduce interference to the circular dichroism (CD) spectra caused by sodium chloride in the SEC buffer. CD experiments were ran on a JAS.CO J-1500 CD spectrometer by Jake Carter. The scans were ran between 400–185 nm with a scanning speed of 200 nm/min. All CD scans were ran using a 150 μL cuvette and were averaged over 3 repeats. The concentrations used are listed below (Table 2.22). Buffer blanks were subtracted from the data to mitigate any remaining interference caused by the buffer. Analysis of the traces was preformed using Dichroweb (Miles et al., 2021) CDSSTR method (Compton and Johnson, 1986) with the SMP180 reference set (Abdul-Gader et al., 2011), fitted plots are presented in Appendix XI.

Protein	Concentration (mg/mL)
Bd3720 WT	1.00
Bd3720 polyA HTH	1.35
Bd3720 polyA WHTH	1.12
Bd3720 polyA Double	2.35

Table 2.22 Concentrations of Bd3720 WT and polyA mutants used in CD analysis.

Chapter 3 – The Structural Determination of Bdellovibrio bacteriovorus Chemoreceptors

3.1 Introduction

As previously stated, the ability of *Bdellovibrio* to use chemotaxis to assist with prey hunting has never been investigated (on a mechanistic basis) and is under much debate. Unlike traditional chemotaxis towards high concentrations of small molecules, the hunting of mobile prey would require *Bdellovibrio* to not only determine the general direction of a nutrient source but also predict where its prey is moving towards (Jashnsaz et al., 2017). Nevertheless, it is unlikely that an obligate predatory bacterium locates its prey merely by random chance, especially as genomic sequencing has shown that *Bdellovibrio* contains the genes for two separate chemosensory pathways (Gumerov et al., 2020). An alternative use for chemotaxis could be for the location of nutrients by a naturally occurring HI equivalent form of *Bdellovibrio*. However, the apparent complexity of its sensory system seems to indicate that at least one pathway would allow for prey location in its primary state of prey dependence.

Two studies in *Bdellovibrio* 109J showed that the interruption of chemosensory genes reduced biofilm predation (Lambert et al., 2006; Medina et al., 2008). The equivalent genes in *Bdellovibrio bacteriovorus* HD100, *bd1469* and *bd2503*, possess a 4HB and dCache LBD respectively. As the enclosed nature of the Cache ligand binding pocket (LBP) generally leads to a larger number of contacts with the ligand which can result in a higher ligand specificity compared to 4HBs (Ortega et al., 2017a) (although this is not always the case), Bd2503 and its homologues, Bd1081 and Bd2504, as well as the sCache Bd3092, were chosen for structural studies to determine their respective ligands.

The Cache domain was first reported by Anantharaman and Aravind in 2000, with a predicted structure analogous to the PAS domain (Figure 3.1b). These are small molecule binding extracellular/periplasmic domains, with very few exceptions (unlike the PAS domain, which mostly rely on prosthetic groups like cytochromes and flavins to sense redox changes, something that is indicative of its cytoplasmic nature). Ligands are bound within a β-sheet pocket, which is located in the distal lobe of dCache domain. The Cache domain at its core is a small molecule sensor that is capable of transducing a signal across bacterial membranes. The effect this produces is reliant on a cytoplasmic domain fused to the Cache domain (Anantharaman and Aravind, 2000). All four of the *Bdellovibrio* homologues contain a MCP cytoplasmic region (Figure 3.1a); which consists of a HAMP helix (a motif in Histidine kinase, Adenylate cyclases, **Methyl-accepting proteins** and Phosphatases), a Methylated Helical (MH) bundle and a signalling tip (Bi and Sourjik, 2018). The structure of the cytoplasmic region of MCPs is highly conserved, and as this project aims to identify the ligands of these proteins rather than to determine the mechanism of signal transduction, the LBD for each was isolated and expressed independently of the cytoplasmic region.

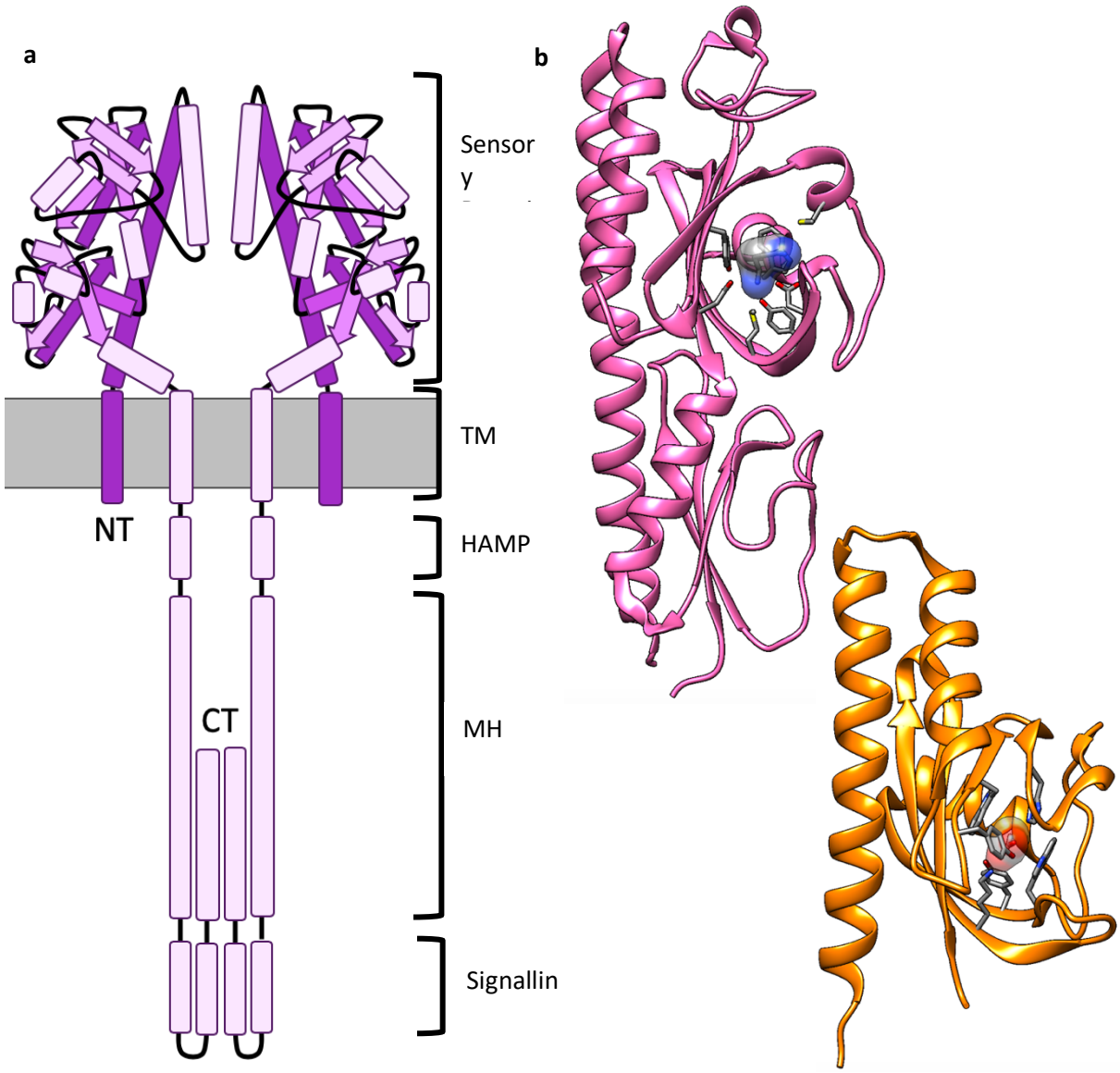


Figure 3.1 (a) Diagram of a dCache chemoreceptor dimer with domains indicated. **(b)** Structural examples of dCache (*P. putida* McpU^{LBD} (6F9G) (Gavira et al., 2018) (Pink)) and sCache (*A. dehalogenans* Adeh_3718 (4K08) (Pokkuluri et al., 2013) (Orange)) ligand binding domains.

Cache domains have been confirmed as small molecule binders, with dCache domains being confirmed to bind (Figure 3.2): Amino acids, Quaternary Ammonium Compounds (QACs), Di-/Tri-carboxylic acids, γ-Aminobutyric acid (GABA), Polyamines, Taurine, Purines, Galactose, and Thiamine (Ortega et al., 2017a). Conservely, sCache have been observed to bind Urea and C2/C3 organic acids, and more recently C4 di-carboxylic acids (Martín-Mora et al., 2018b; Ortega et al., 2017a). Numerous studies into Cache domains have shown their specificity to generally one particular group of related

small molecules (Ortega et al., 2017a), with some binding just a single ligand, for example McpG in *P. putida* binds only GABA (Reyes-Darias et al., 2015). It is this tight range of specificity to non-ion small molecules that should enable us to shed light on the predatory use of chemotaxis by *Bdellovibrio.*

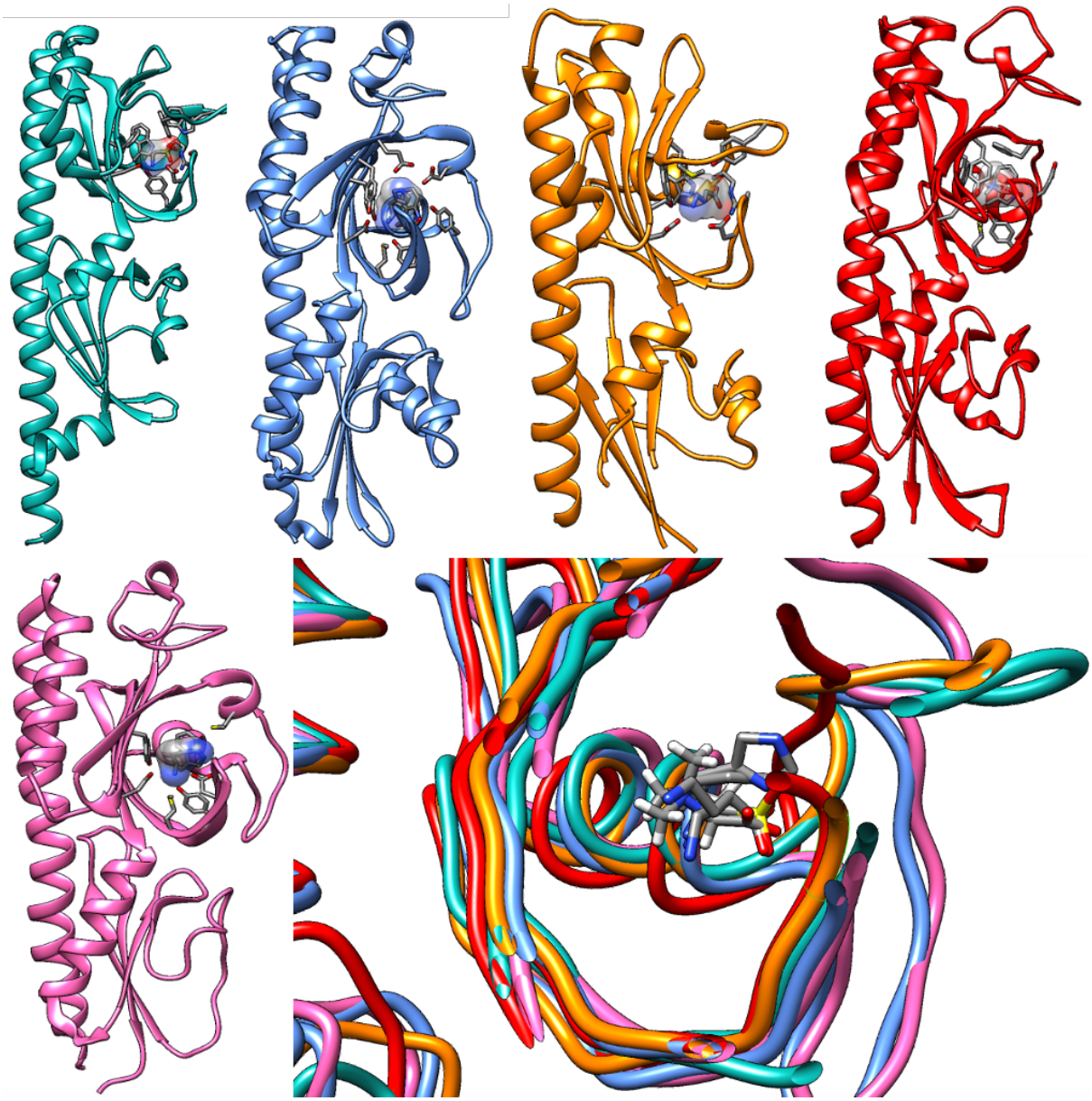


Figure 3.2 dCache domains bound to a variety of different ligands within the membrane distal pocket, with the overlay showing they all adopt the same horizontal binding conformation across the pocket. (Teal – V. cholerae Mlp37 with taurine (5AVF) (Nishiyama et al., 2016), Blue – P. aeruginosa TlpQ with histamine (6FU4) (Corral-Lugo et al., 2018), Orange – P. aeruginosa PctC with GABA (5LTV) (Gavira et al., 2020), Red – S. meliloti with proline betaine (6D8V) (Shrestha et al., 2018), Pink – *P. putida* McpU^{LBD} with putrescine (6F9G) (Gavira et al., 2018))

3.2 Results

3.2.1 Structural Determination of Bd2503^{LBD}

3.2.1.1 Purification of Bd2503^{LBD}

The ligand binding domain of Bd2503^{LBD} was expressed in AI media with sugar mix N and purified using nickel affinity chromatography eluted with 300 mM Imidazole (Figure 3.3a). The purified protein was dialysed into 20 mM Hepes pH 7.5 and 300 mM NaCl, prior to a size exclusion step with a HiLoad 26/60 Superdex 75 pg (Figure 3.3c). SEC yielded two peaks, a small peak eluting at 120 ml and a much larger peak eluting at 165 ml, these were assumed to be a dimeric and monomeric peak respectively as a result of the post-SEC gel and the elution volume.

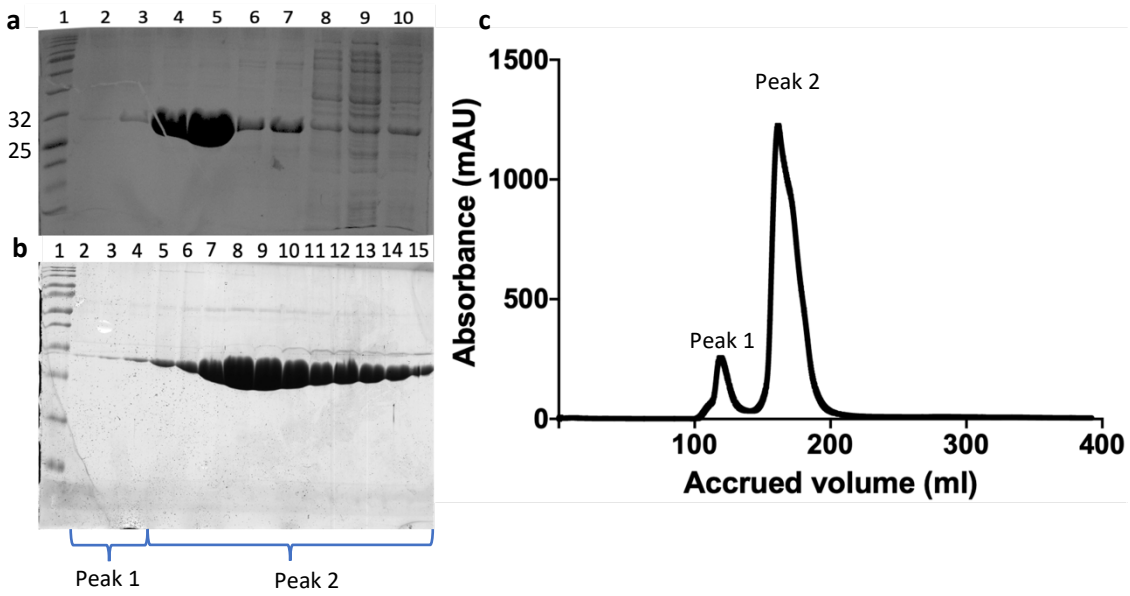


Figure 3.3 (a) SDS-PAGE gel of Bd2503^{LBD} IMAC purification. **(b)** SDS-PAGE gel of peak 1 and peak 2 of Bd2503^{LBD} post size exclusion purification. **(c)** UV trace Bd2503^{LBD} size exclusion purification showing both the dimer and monomer peaks.

3.2.1.2 Differential Scanning Fluorimetry of Bd2503^{LBD}

Since crystals were not being obtained in the initial buffer of 20 mM Hepes, 300 mM NaCl and 5% v/v glycerol, DSF was used to screen for a more favourable buffer condition. The screen did not indicate a more favourable buffer, so additive screening was pursued to see if a small molecule would help stabilise the protein and allow crystallisation. The DSF additive screening was performed in 20 mM Imidazole pH 7.0 and 300 mM NaCl, as this produced the best buffer screen result, in order to reduce thermal shift interference caused by the buffer when determining a viable small molecule. The additive screen proved far more successful, with large shifts produced for 46 mM L-Arginine, 46 mM L-Glutamic acid, 4.6 mM L-Glutamic acid/L-Arginine, and 46 mM L-Glutamic acid/L-Arginine (Figure 3.4). Subsequent crystal screens were set up with 50 mM L-Arginine, 50 mM L-Glutamic acid and 50 mM L-Glutamic acid/L-Arginine, to help protein stabilisation and crystal formation.

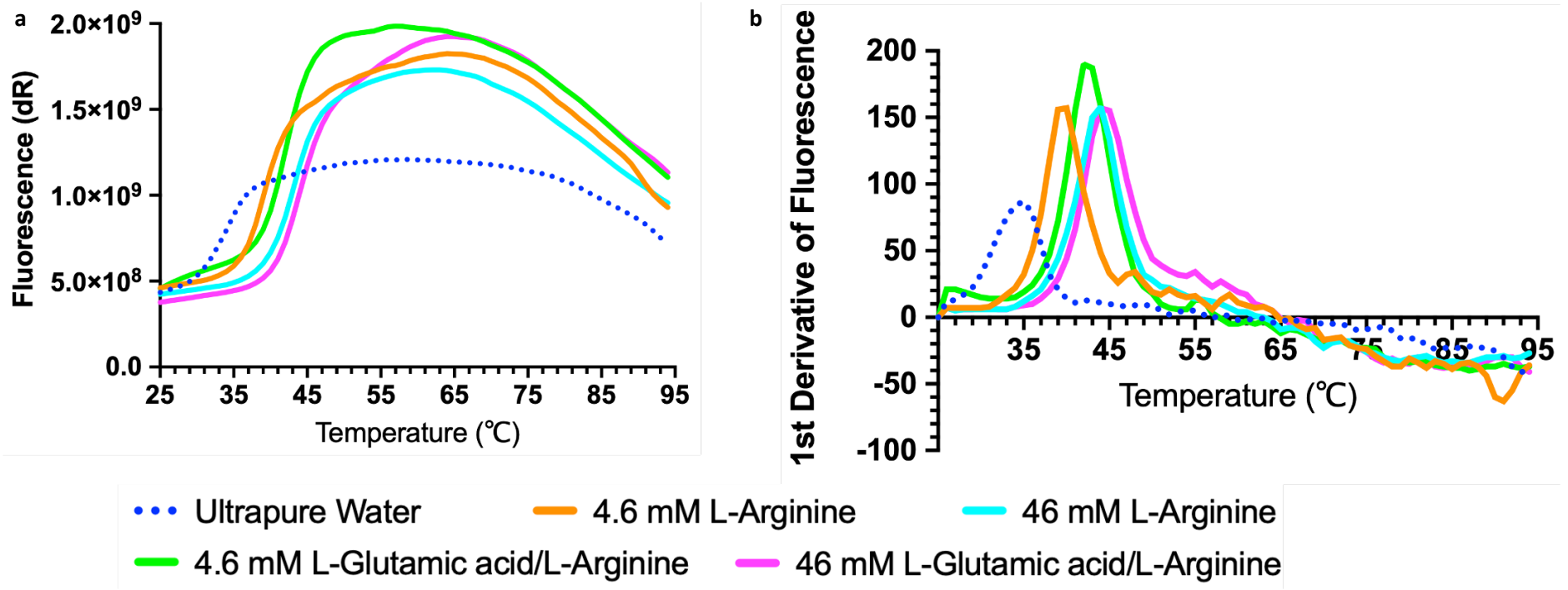


Figure 3.4 DSF melting curve of the additives that shifted the melting point 6 °C or more, with water negative control in dark blue, **(a)** Raw DSF melting curves, **(b)** First derivative plot of DSF melting curves.

3.2.1.3 Structure Determination and Model Building of Bd2503^{LBD}

The DSF screen buffer was not used for crystallisation, but the purified protein was supplemented with L-Glutamine and/or L-Arginine as determined by the additive screen. An initial hit in JCSG^{+} H7 with Bd2503^{LBD} at 17 mg/ml supplemented with 50 mM L-Arginine and a 1:1 drop ratio was obtained. This condition was then optimised using a homescreen varying Bis tris pH and PEG 3350 final w/v concentration, while keeping protein, L-Arginine and Ammonium sulphate concentration the same. Diffraction data were obtained from the following conditions:

Method	Hanging drop vapour diffusion
Temperature (°C)	18
Plate Type	24-Well Greiner Combo plate
Protein Concentration (mg/ml)	17
Reservoir Solution	0.1 M Bis tris pH 5.7, 23% w/v PEG 3350 and 0.2 M Ammonium sulphate
Drop Volume (µl)	2
Drop Ratio	1:1
Volume of Reservoir (µl)	500
Crystallisation Morphology	3D Pear shape

Table 3.1 Crystallisation conditions for Bd2503^{LBD} _Native.

Crystals were soaked with a cryoprotectant containing 20% v/v ethylene glycol and 25 mM Cadmium chloride. The crystals were supplemented with heavy atoms, as all existing dCache structural models had low sequence identity. Diffraction data were obtained at Diamond Light Source to the following specifications:

Diffraction Source	Diamond Light Source – i04		
Cryoprotectant	20% v/v Ethylene Glycol		
Heavy Atom	25 mM Cadmium Chloride		
Wavelength (Å)	1.400		
Rotation of Crystal (°)	360		
Rotation per Image (°)	0.1		
Images Collected	3600		
Exposure per Image (sec)	0.04		
Transmission (%)	26.36		
Detector Resolution (Å)	2.5		
Space Group	C 1 2 1		
Cell Dimensions			
a, b, c (Å)	111.32	35.43	136.08
α, β, γ (°)	90.0	106.47	90.0
Resolution Range (Å)	1.90 – 65.32		
Wilson B-Factor	20.30		
	Overall	High	
Total Reflections	53377	168	
Unique Reflections	29473	167	
R_{merge}	0.050	0.000	
R_{meas}	0.071	0.000	
$R_{p.i.m}$	0.050	0.000	
$I/\sigma I$	7.4	0.7	
CC(1/2)	0.990	0.800	
Multiplicity	4.4	1.1	
Completeness	73.0	6.1	

Table 3.2 Diffraction dataset specifications for $Bd2503^{LBD}$ _Native.

Once the diffraction data had been obtained (log file in Appendix I) an initial search model was created using a single copy of the mmHK1S-$Z2^{LBD}$ distal lobe (a chemotaxis receptor from *Methanosarcina mazei* with 25% homology over a range of 278 amino acids (3LI8 (Zhang and Hendrickson, 2010)) with CCP4 Chainsaw. Since the initial molecular replacement model had weak sequence homology, experimental phasing using SIR/MIR Phenix AutoSol (Terwilliger et al., 2009) to locate the supplemented cadmium ions was attempted. A search criteria for approximately 6 Cd ions (number estimated based on number of monomers in asymmetric unit), with the wavelength set to 1.4 Å, and f' = 0.067 e and f'' = 3.929 e. However, the AutoSol map produced was not readily interpretable with an FOM of 0.1. Molecular replacement using Phenix PHASER (McCoy et al., 2007) and a 3LI8 model produced a putative solution with a TFZ-score of 7.3 and the two monomer copies placed in the classical Cache dimer conformation. As the initial map produced by the molecular replacement and

anomalous signal from the Cadmium was poor, 50 rounds of density modification in SHELXE (Spek, 2009) was employed, this was only possible because of the high resolution of the data sets. The combination of molecular replacement, density modification and weak anomalous signal yielded an interpretable map with a FOM of 0.4. The proximal lobe of $Bd2503^{LBD}$ was then built using a combination of Phenix AutoBuild (Terwilliger et al., 2008) and manual building in COOT (Emsley et al., 2010). The model was then refined manually using COOT and Phenix Refine (Afonine et al., 2012) (Table 3.3).

Resolution (Å)	1.4
Number of Copies in Asymmetric Unit	2
R_{Work}	0.2588
R_{Free}	0.2842
R.M.S Deviations	
Bond Length (Å)	0.01
Bond Angles (°)	1.01
Ramachandran Plot	
Favoured Regions(%)	96.01
Allowed Regions(%)	3.99
Outliers (%)	0.00

Table 3.3 Final refinement statistics for $Bd2503^{LBD}$ _Native.

The asymmetric subunit of the native crystals contained two copies of $Bd2503^{LBD}$, forming a symmetrical dimer around the N-terminal transmembrane helix. Overlaying the initial model against the built structure of $Bd2503^{LBD}$ showed that the initial model deviated from the refined map reinforcing that the true structure of $Bd2503^{LBD}$ had been determined with limited bias from the molecular replacement model (Figure 3.5). The majority of both chains were built, for Chain A residues 12-243 and 247-276 were placed. In the gap there was some partial density that allowed for an approximation of the loop to be placed. This inferior electron density for loops can be expected as they are often highly flexible. Chain B was fully built from residue 12-277, which helped with the fitting of the missing amino acids in Chain A. Missing amino acids at both termini is a common occurrence in protein structures.

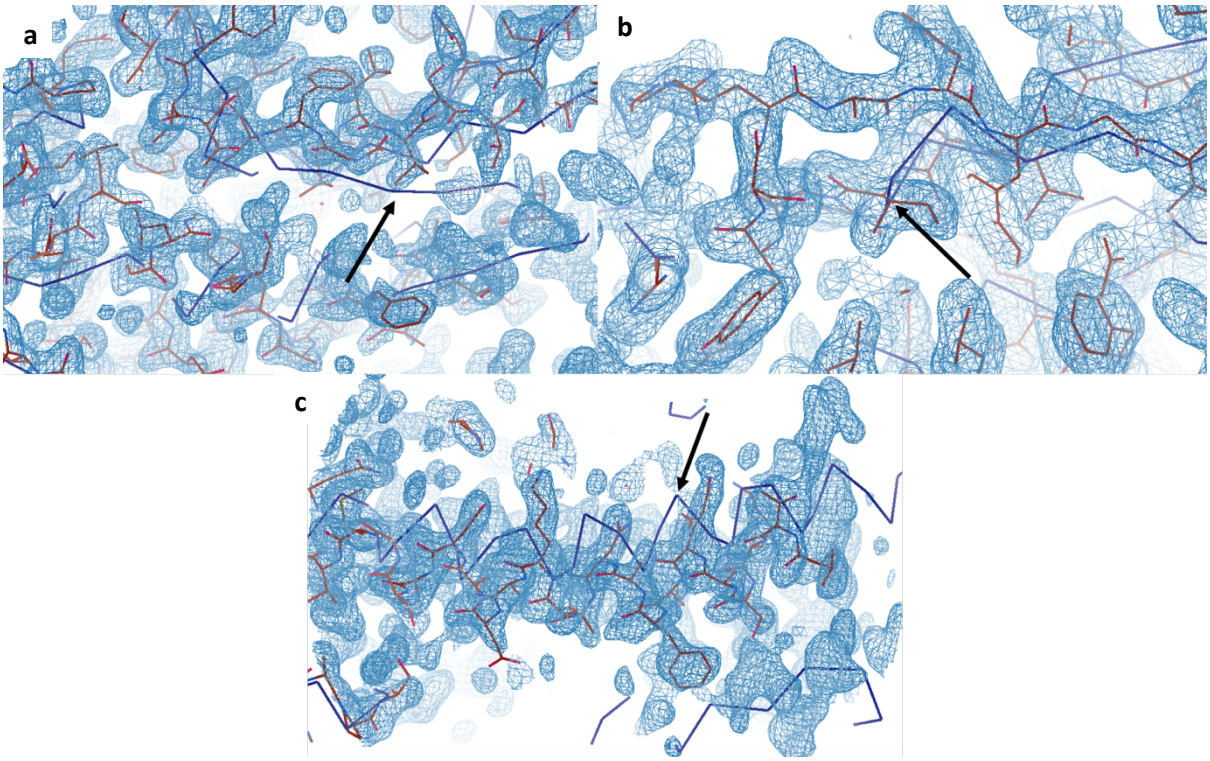


Figure 3.5 Deviations between the structure of mmHK1S-Z2^{LBD} (3LI8 in dark blue (Zhang and Hendrickson, 2010)) and the refined 2FOFCWT PH2FOFCWT map (blue) shows that the true Bd2503^{LBD} (orange) structure has been built. **(a)** Shows a β-strand (identified by a black arrow) in the bottom lobe of mmHK1S-Z2^{LBD} that does not match the refined map, whereas Bd2503^{LBD} fits well within the density. **(b)** Shows a loop (identified by a black arrow) in the bottom lobe of mmHK1S-Z2^{LBD} that does not extend as far as that in Bd2503^{LBD}. **(c)** Shows that the N-terminal transmembrane helix (identified by a black arrow) of mmHK1S-Z2^{LBD} does not bend at the same angle as Bd2503^{LBD}, and therefore does not fit the refined map as well as Bd2503^{LBD} does.

The two monomers are held in a dimeric conformation by 11 hydrogen bonds as well as hydrophobic packing (Figure 3.6), K14 forming a hydrogen bond with the backbone carboxyl of E12 is the only bond not formed by both monomers. Sidechain-sidechain hydrogen bonds form between: the hydroxyl group of T46 and the carboxyl of N64, and the terminal amine group of R138 and the carboxyl of Q68. The hydrogen bond formed between R138 and Q68 is the only intermolecular hydrogen bond where both residues involved are not found in the transmembrane helix, as R138 forms part of the membrane distal lobe of the chemoreceptor. Sidechain-backbone hydrogen bonds also formed between: the amide group of N24 and the carboxyl of A21, Q38 sidechain amide and Q68 backbone carboxyl, and Q39 sidechain amine and the backbone carboxyl of M65.

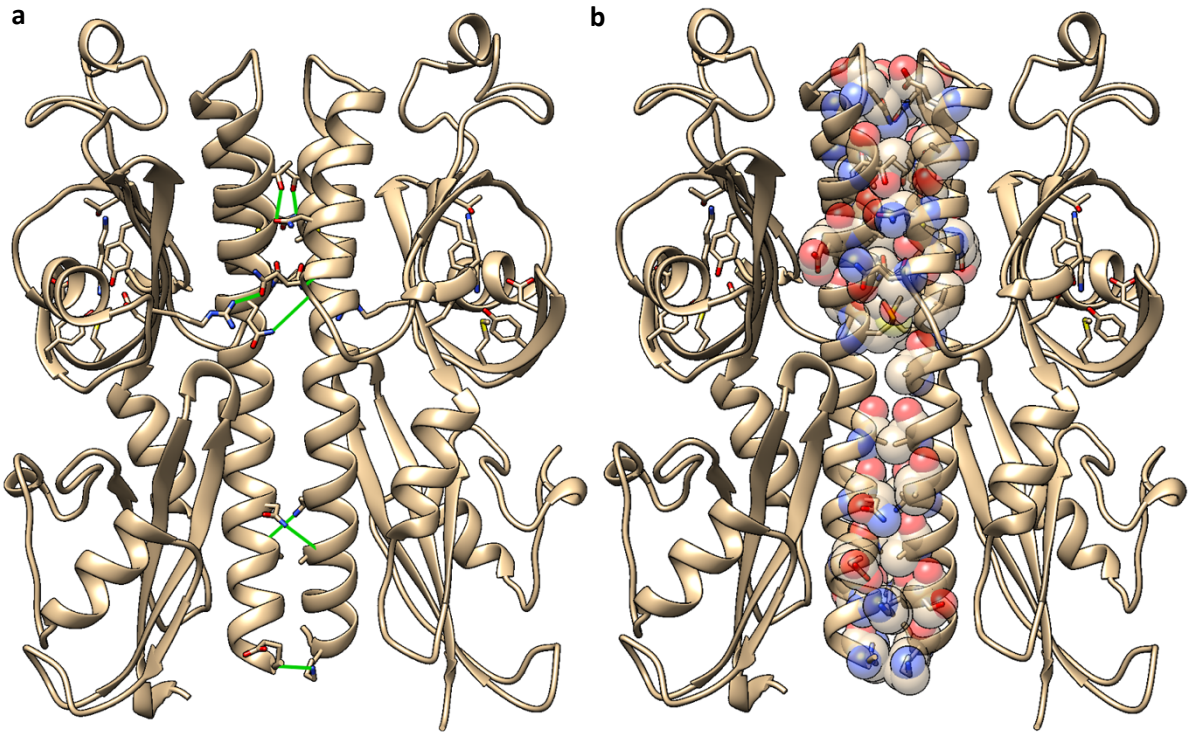


Figure 3.6 (a) Bd2503^{LBD} dimer with hydrogen bonds formed to stabilise dimer interaction shown in green. These hydrogen bonds are mostly limited to the transmembrane helix, with the exception of distal lobe R138 bonding with Q68 from the other monomer transmembrane helix. **(b)** Bd2503^{LBD} dimer with hydrophobic packing between the two monomers displayed.

Having built Bd2503^{LBD} (Figure 3.7) the refined map showed a region of strong density (Figure 3.8) in the classic dCache chemotaxis receptor binding pocket. The ligand was determined to be cadaverine due to a discernible 5-carbon chain and a ligand binding domain rich in hydrogen bond accepting side chains indicating terminal amine groups. Binding of cadaverine takes place in the top lobe of Bd2503^{LBD}, with the ligand axis placed vertically with respect to the expected orientation of dCache domain in the membrane. The binding pocket contains three aromatic amino acids (W76, Y106 and F128) forming a cage around the 5-carbon chain of cadaverine. Four amino acids (Y106, Y144, D146, and D175) at the bottom of the binding pocket form hydrogen bonds with one of the amine groups in cadaverine. Four amino acids (A77, T78, G104 and D121) binds the distal amine group of cadaverine, again through hydrogen bonding. The binding of polyamines by dCache domain chemoreceptors is supported by the literature, however the only other incident of a chemotaxis sensory domain bound to a polyamine shows a far more bent conformation than the rigid elongated pose observed for

cadaverine in Bd2503^{LBD} (Gavira et al., 2018; Ortega et al., 2017a). However, a small molecule structure with cadaverine, using the Cambridge Structural Database, shows that 2 of the 5 related structures deposited, YAMLON (bis(1,5-Diaminopentane dichloride dihydrate)) and FISJOG (bis(Pentane-1,5-diammonium) hexachloro-platinum (iv) dichloride), adopt a similar rigid straight conformation.

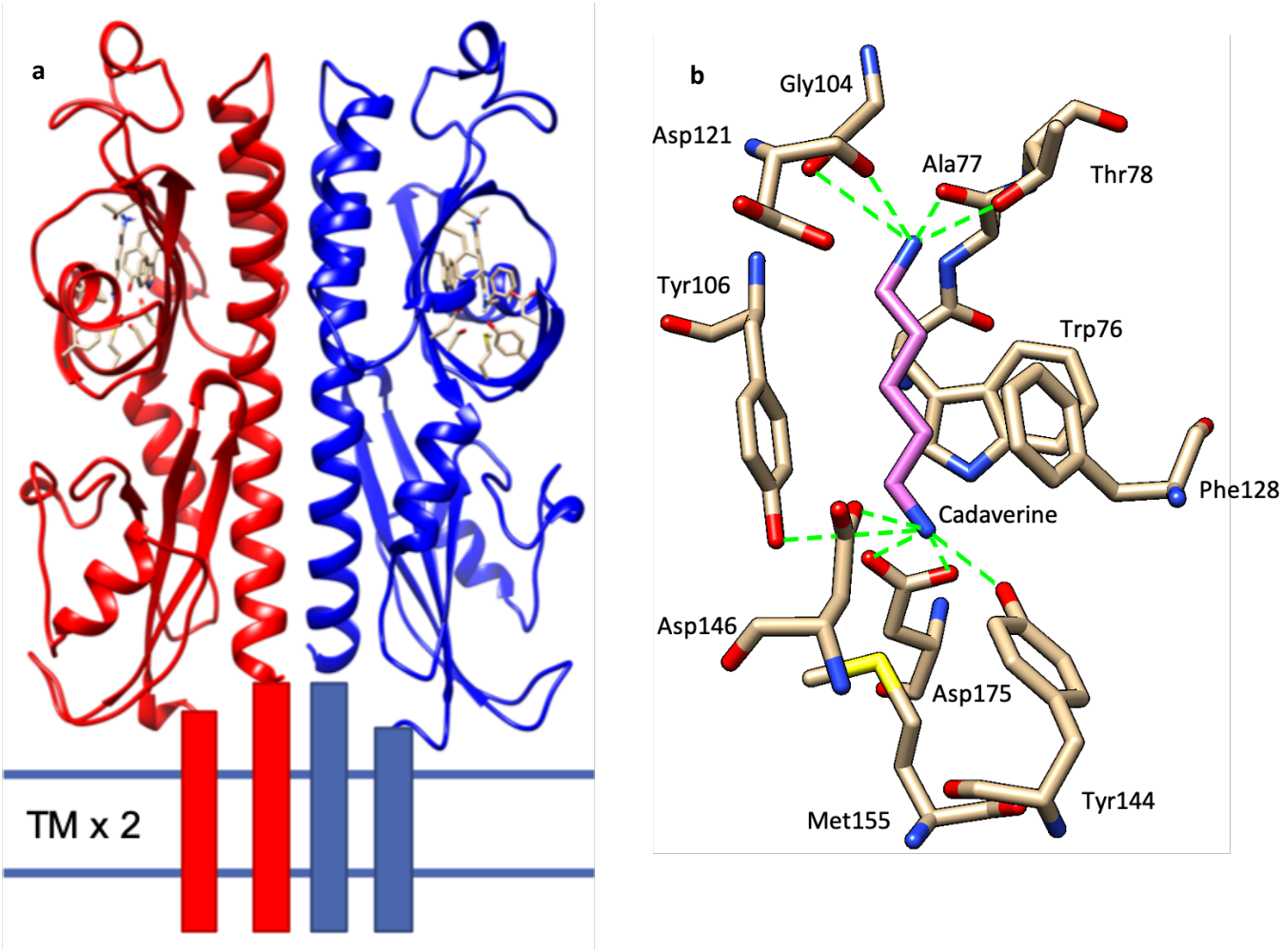


Figure 3.7 (a) Bd2503^{LBD} dimerisation. **(b)** Close up of Bd2503^{LBD} binding pocket with cadaverine (pink) bound and hydrogen bonds shown in green.

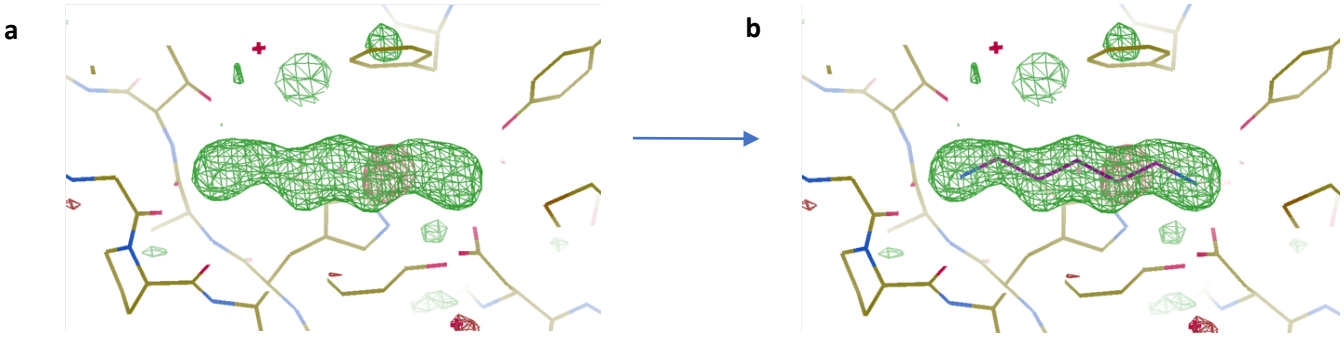


Figure 3.8 Coot FOFCWT PHFOFCWT density map of the binding pocket with sigma set to 3.0. **(a)** High resolution, unaccounted for, density in the ligand binding pocket. **(b)** Cadaverine fitted into this density.

As the original crystal structure of $Bd2503^{LBD}$ was obtained with only the co-purified cadaverine available for complexation, subsequent crystallisation trays were set up supplemented with 2 mM cadaverine. This was done in order to saturate the binding sites available and improve crystal homogeneity in order to improve the resolution of the density map of the proximal lobe, which was significantly decreased compared to the top lobe in the native crystal data set. This lead to crystallisation in $JCSG^{+}$ E10 with the following conditions:

Method	Sitting drop vapour diffusion
Temperature (°C)	18
Plate Type	96 Well CombiClover 4 chamber crystallization plates
Protein Concentration (mg/ml)	17
Reservoir Solution	0.1 M Bicine pH 9.0, 10% w/v PEG 6000 and 2 mM cadaverine
Drop Volume (µl)	1.5
Drop Ratio	1:1
Volume of Reservoir (µl)	100
Crystallisation Morphology	3D Pear shape

Table 3.4 Crystallisation conditions for $Bd2503^{LBD}$ supplemented with cadaverine.

Crystals were soaked with a cryoprotectant containing 20% v/v ethylene glycol and 2 mM cadaverine, and the crystal diffraction data were collected at Diamond Light Source:

Diffraction Source	Diamond Light Source – i04		
Cryoprotectant	20% v/v Ethylene Glycol		
Wavelength (Å)	0.9795		
Rotation of Crystal (°)	360		
Rotation per Image (°)	0.15		
Images Collected	2400		
Exposure per Image (sec)	0.05		
Transmission (%)	79.91		
Detector Resolution (Å)	2		
Space Group	P 1		
Cell Dimensions			
a, b, c (Å)	62.30	69.27	74.21
α, β, γ (°)	91.64	97.75	111.85
Resolution Range (Å)	2.07 – 50.41		
Wilson B-Factor	33.75		
	Overall	High	
Total Reflections	189838	3603	
Unique Reflections	58590	1280	
R_{merge}	0.028	0.607	
R_{meas}	0.040	0.859	
$R_{p.i.m}$	0.028	0.607	
I/σI	17.6	1.2	
CC(1/2)	0.999	0.577	
Multiplicity	3.2	2.8	
Completeness	84.6	37.0	

Table 3.5 Diffraction dataset specifications for $Bd2503^{LBD}$ supplemented with cadaverine.

Once the crystal data had been obtained (log file in Appendix II) the structure was obtained using Phenix PHASER (McCoy et al., 2007) and the original co-purified cadaverine complex data set structure was used as a molecular replacement model, producing an initial structure with a TFZ of 47.0. COOT (Emsley et al., 2010) and the suite of Phenix programs (Adams et al., 2010) were then used to refine the structure (Table 3.6).

Resolution	2.07
Number of Copies in Asymmetric Unit	4
R_{Work}	0.2442
R_{Free}	0.2851
R.M.S Deviations	
Bond Length (Å)	0.02
Bond Angles (°)	1.52
Ramachandran Plot	
Favoured Regions(%)	93.97
Allowed Regions(%)	5.28
Outliers (%)	0.75

Table 3.6 Final refinement statistics for Bd2503^{LBD} supplemented with cadaverine.

Although supplementing with cadaverine did lower the overall resolution of the data set, it did improve the overall density of the bottom domain in the 4 copies of the asymmetric unit which formed two dimers, 2 of which were able to be built as a continuous chain, Chain B from 13-274 and Chain D from 14-274 (with the full length construct being 290 amino acids). However, despite the overall improvement in the bottom domain, Chains A and C were missing more residues from this flexible loop than the native structure. Chain A residues 17-242 and 248-274, and Chain C residues 18-241 and 251-274 were placed.

Having obtained a structure of Bd2503^{LBD} with its ligand bound, we then aimed to produce an apo structure to establish how the distal domain changed conformation upon ligand binding. Apo Bd2503^{LBD} was produced using multiple dialysis steps (See Chapter 2.12 and Chapter 4.2.1). Attempts to crystallise apo Bd2503^{LBD} using the same crystallisation conditions as the complex form were unsuccessful, as were commercial crystal screens of Bd2503^{LBD} Apo set up using seed crystal stocks obtained from the Bd2503^{LBD} Native homescreen. Therefore, Bd2503^{LBD} Apo was screened with 50 mM L-Arginine as the original Native crystals had been, to determine if this would stabilise the protein and enable crystallisation. This led to crystallisation in Morpheus A4 with the following conditions:

Method	Sitting drop vapour diffusion
Temperature (°C)	18
Plate Type	96 Well CombiClover 4 chamber crystallization plates
Protein Concentration (mg/ml)	17
Reservoir Solution	0.1 M Imidazole/MES pH 6.5, 0.03 M $MgCl_2$, 0.03 M $CaCl_2$, 12.5% v/v MPD, 25% w/v PEG 1000 and 12.5% w/v PEG 3350
Drop Volume (µl)	1.5
Drop Ratio	1:1
Volume of Reservoir (µl)	100
Crystallisation Morphology	Octahedron

Table 3.7 Crystallisation conditions for $Bd2503^{LBD}$ supplemented with L-arginine.

Crystals did not require cryoprotecting before flash-cooling due to the makeup of the crystal drop providing innate cryoprotection, data were collected at Diamond Light Source:

Diffraction Source	Diamond Light Source – i03		
Cryoprotectant	25% w/v PEG 1000		
Wavelength (Å)	0.9763		
Rotation of Crystal (°)	360		
Rotation per Image (°)	0.10		
Images Collected	3600		
Exposure per Image (sec)	0.05		
Transmission (%)	79.91		
Detector Resolution (Å)	2		
Space Group	C 1 2 1		
Cell Dimensions			
a, b, c (Å)	149.10	68.87	129.23
α, β, γ (°)	90.00	117.41	90.00
Resolution Range (Å)	2.32 – 66.18		
	Overall	High	
Total Reflections	368240	26378	
Unique Reflections	54768	4048	
R_{merge}	0.099	2.293	
R_{meas}	0.118	2.757	
$R_{p.i.m}$	0.063	1.512	
$I/\sigma I$	7.4	1.5	
CC(1/2)	0.996	0.469	
Multiplicity	6.7	6.5	
Completeness	100.0	100.0	

Table 3.8 Diffraction dataset specifications for $Bd2503^{LBD}$ supplemented with L-arginine.

Once the diffraction data had been obtained (log file in Appendix III) the structure was obtained using Chain C from the Bd2503^{LBD} Cadaverine complex dataset as a molecular replacement model using Phenix PHASER (McCoy et al., 2007). COOT (Emsley et al., 2010) and the suite of Phenix programs (Adams et al., 2010) were then used to refine the structure (Table 3.9).

Resolution	2.07
Number of Copies in Asymmetric Unit	4
R_{Work}	0.2866
R_{Free}	0.3336
R.M.S Deviations	
Bond Length (Å)	0.01
Bond Angles (°)	1.62
Ramachandran Plot	
Favoured Regions(%)	89.35
Allowed Regions(%)	9.1
Outliers (%)	1.55

Table 3.9 Final refinement statistics for Bd2503^{LBD} supplemented with L-arginine.

Bd2503^{LBD} with L-Arginine yielded 4 copies in the asymmetric unit, and like the Bd2503^{LBD} Cadaverine complex formed two dimers (Figure 3.9a). Of the three Bd2503^{LBD} structures, Bd2503^{LBD} with L-Arginine had the worst density in the bottom lobe and was the most disordered. Of the four chains, only Chain C was fully built, residues 10-280, with the exception of the terminal regions. However unlike the previous structures, the top of the second transmembrane helix could be visualised. Chain C dimerised with partially built Chain A, in which residues could only be placed confidently in 4 sections, 11-179, 193-218, 224-236 and 247-271. Chain B was only missing a small section in the flexible loop, with residues placed at positions 14-239 and 244-280. Chain B dimerised with the worst copy, Chain D , that had residues built in positions 10-91, 97-106, 115-160, 153-160, 163-232 and 247-271. This pattern of complete and incomplete chains dimerising together holds true for the other two structures, indicating that dimerisation seems to affect protein flexibility and therefore the data obtained from the resulting crystals.

Upon refinement, the LBP was found to be occupied, not empty as had been desired (Figure 3.9b). The density was determined to be the L-Arginine added to try and stabilise the protein for crystallisation. Superimposing the $Bd2503^{LBD}$ Cadaverine complex with $Bd2503^{LBD}$:L-Arginine revealed that there was no variation in conformation of the top lobe between the two structures (Figure 3.10). The L-Arginine was bound in an identical location and similar conformation as cadaverine, with the ligand axis running parallel to the transmembrane helix, however there is a slight bend in the side chain between C_β and N_ε of the L-Arginine to allow it to fit into the pocket. Most of the same amino acids identified to interact with cadaverine were involved in binding L-Arginine, W76, Y106 and F128 form the aromatic cage, whilestD146 and M155, and D146 and D175, bind the amine group of the arginine sidechain via hydrogen bonds and salt bridges respectively. A77, T78, G104, P105 and D121 bind the amine group of the backbone and T78 and N123 bind the backbone carboxylic group, which faces away from the transmembrane helix, and towards the outside of the pocket. $Bd2503^{LBD}$ is thought to have bound L-Arginine only because the extreme excess it was added in, about 100 fold (this is supported by preliminary ITC experiments in Chapter 4.2.3). Firstly, the close proximity of D121 to the carboxylic group of L-Arginine upon binding would be undesirable, and is indicative of a LBP tailored to amine binding and ligand specificity, instead of a more promiscuous ligand range. Secondly, although insistences of higher ligand versatility in dCache proteins have been described (Webb et al., 2017; Martín-Mora et al., 2018a), a chemoreceptor that recognises both polyamines and, their molecular precursor, amino acids, has yet to be reported. However further experimentation into ligand specificity of the *Bdellovibrios* dCaches was performed, which included amino acids and L-Arginine in particular. Although not the apo structure desired, $Bd2503^{LBD}$ bound with L-Arginine did demonstrate how flexible the LBP was and how much it could expand to accommodate a ligand unlikely to be relevant *in vivo* (demonstrated by ITC showing weak affinity in Chapter 4.2.3).

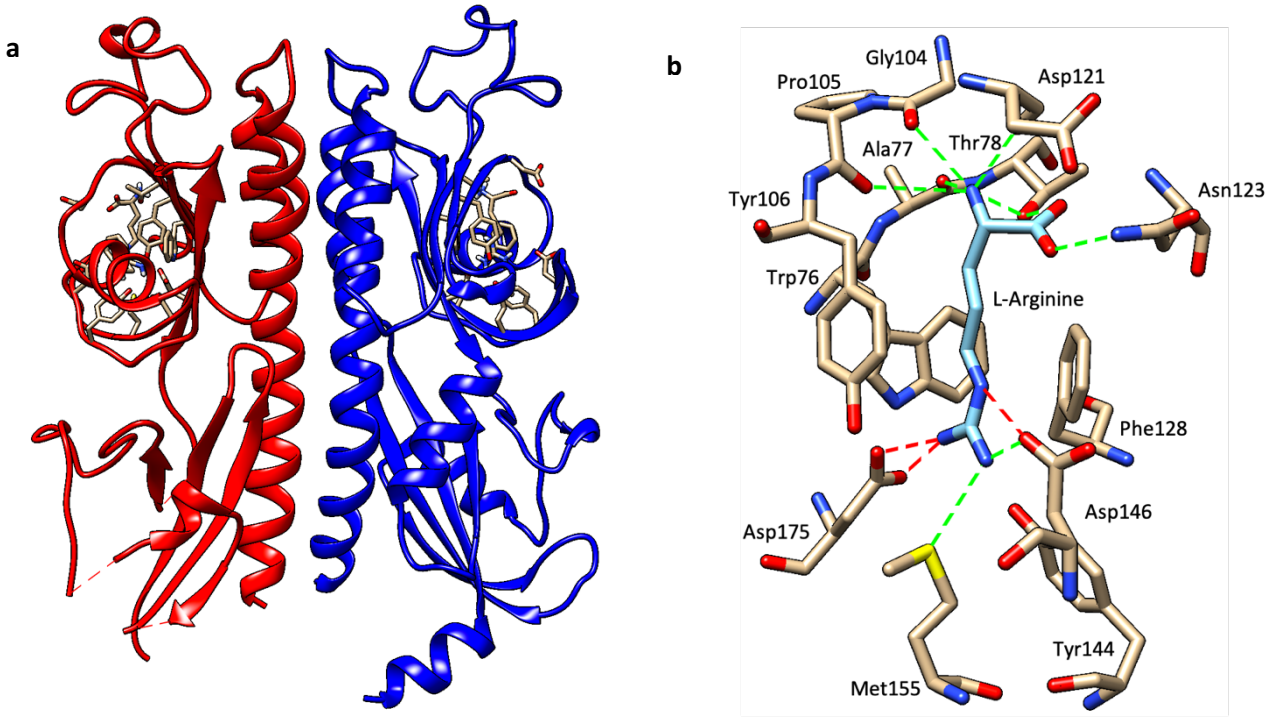


Figure 3.9 (a) Bd2503^{LBD} dimerisation. **(b)** Close up of Bd2503^{LBD} binding pocket with L-Arginine (light blue) bound, with stabilising hydrogen bonds shown in green and salt bridges in red.

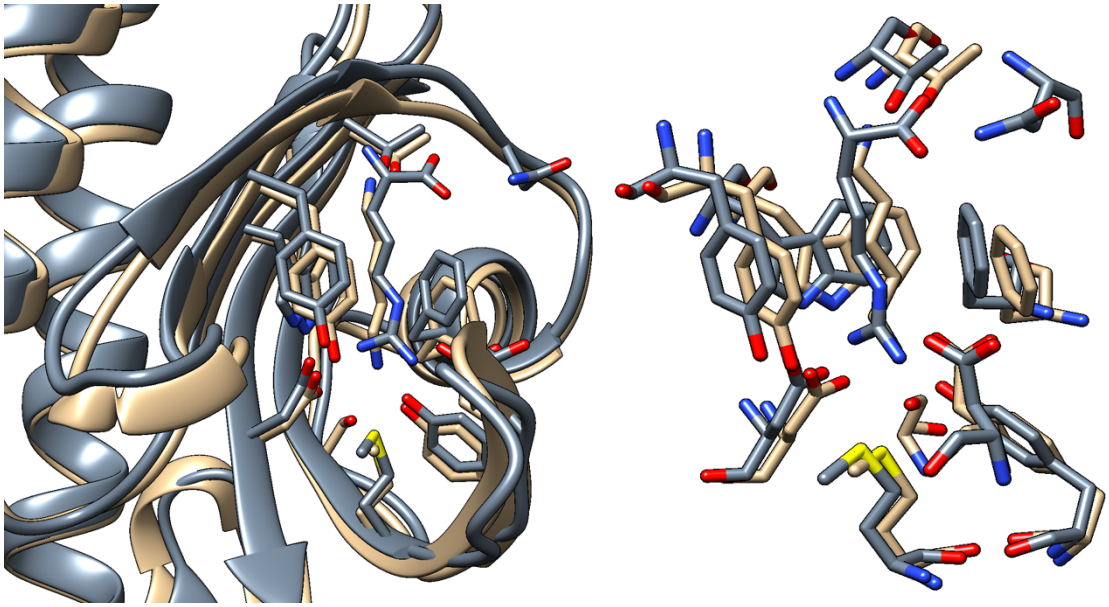


Figure 3.10 Overlay of Bd2503^{LBD} with cadaverine (tan) and Bd2503^{LBD} with L-arginine (silver) bound, showing very little conformational change as well as the difference between the amino acids used for binding the two molecules.

3.2.2 Structural Determination of Bd2504^{LBD}

3.2.2.1 Purification of Bd2504^{LBD}

The ligand binding domain of Bd2504^{LBD} was expressed in AI media with sugar mix N and purified using nickel affinity chromatography eluted with 100 mM Imidazole (Figure 3.11a). The purified protein was dialysed into 20 mM Bis tris pH 7.5 and 300 mM NaCl, before a SEC step with a HiLoad 26/60 Superdex 75 pg (Figure 3.11b). SEC yielded two peaks of similar size, one eluting at 120 ml and one peak eluted at 165 ml, these were assumed to be a dimeric and monomeric peak respectively as a result of the post-SEC gel and the elution volume.

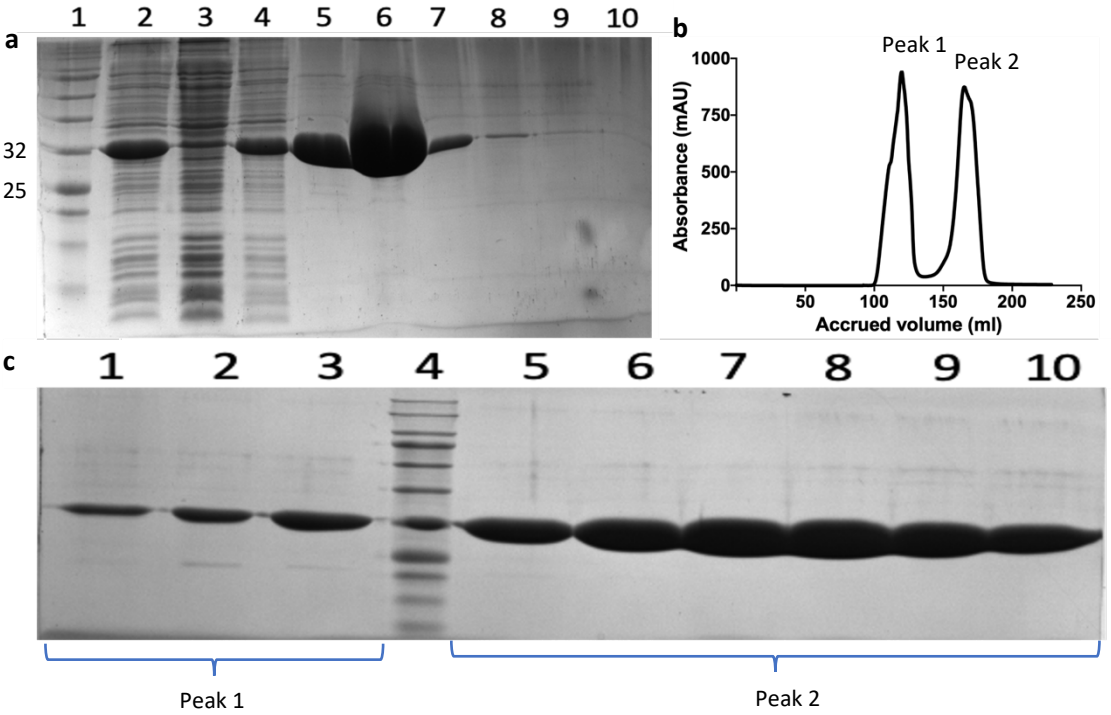


Figure 3.11 (a) SDS-PAGE gel of Bd2504^{LBD} IMAC purification. **(b)** UV trace Bd2504^{LBD} size exclusion purification showing both the monomer and dimer elutions. **(c)** SDS-PAGE gel of peak 1 and peak 2 of Bd2504^{LBD} post size exclusion purification.

3.2.2.2 Data Collection and Structure Building of Bd2504^{LBD}

Initial attempts to crystallise Bd2504^{LBD} were unsuccessful, as Bd2504^{LBD} proved very unstable compared to Bd2503^{LBD} and precipitated heavily when in crystal trays. Unlike Bd2503^{LBD}, addition of 50 mM L-Arginine did not improve stability to allow crystallisation. Using Tcoffee (Notredame et al.,

2000) to align Bd2503^{LBD} and Bd2504^{LBD} sequences showed Bd2504^{LBD} was predicted to have tryptophan residues at positions 106 and 130, instead of the less bulky aromatics Y106 and F128 found forming the aromatic cage in Bd2503^{LBD}. This was predicted to reduce the volume available for ligand-binding within the LBP of Bd2504^{LBD} in comparison to Bd2503^{LBD}. Therefore the protein was supplemented with 1.2 mM putrescine, a 4-carbon chain version of cadaverine. A concentration of 1.2 mM was chosen as it was twice that of the protein concentration being used in the crystal trays, therefore saturating the ligand binding domains available, increasing the chances of crystal homogeneity. Initial hits were in JCSG^{+} H9; 0.1 M Bis-Tris pH 5.5, 25% w/v PEG 3350 and 0.2 M NaCl, with Bd2504^{LBD} at 20 mg/ml supplemented with 1.2 mM putrescine and in a 1:1 drop ratio. This condition was then optimised using a homescreen varying Bis tris pH and PEG 3350 w/v concentration, while keeping protein, putrescine and Lithium sulphate concentration constant. Diffraction data were obtained from the following condition:

Method	Hanging drop vapour diffusion
Temperature (°C)	18
Plate Type	24-Well Greiner Combo plate
Protein Concentration (mg/ml)	20
Reservoir Solution	0.1 M Bis tris pH 5.5, 25% w/v PEG 3350 and 0.2 M Lithium Sulphate
Drop Volume (µl)	2
Drop Ratio	1:1
Volume of Reservoir (µl)	500
Crystallisation Morphology	3D Pear shape

Table 3.10 Crystallisation conditions for Bd2504^{LBD} with putrescine.

Crystals were soaked with a cryoprotectant containing 20% v/v ethylene glycol and 1.2 mM putrescine to maintain ligand binding during flash-cooling. Crystal data were obtained at Diamond Light Source with the following specifications:

Diffraction Source	Diamond Light Source – i04		
Cryoprotectant	20% v/v Ethylene Glycol		
Wavelength (Å)	0.9795		
Rotation of Crystal (°)	360		
Rotation per Image (°)	0.1		
Images Collected	3600		
Exposure per Image (sec)	0.04		
Transmission (%)	60.04		
Detector Resolution (Å)	2.0		
Space Group	P 6 1		
Cell Dimensions			
a, b, c (Å)	107.72	107.72	106.47
α, β, γ (°)	90.0	90.0	120.0
Resolution Range (Å)	2.05 – 70.17		
Wilson B-Factor	46.949		
	Overall	High	
Total Reflections	913139	45489	
Unique Reflections	45978	2258	
R_{merge}	0.075	2.067	
R_{meas}	0.079	2.177	
$R_{p.i.m}$	0.025	0.682	
$I/\sigma I$	18.3	1.5	
CC(1/2)	0.996	0.816	
Multiplicity	19.9	20.1	
Completeness	100.00	99.1	

Table 3.11 Diffraction dataset specifications for Bd2504^{LBD} supplemented with putrescine.

Once the diffraction data had been obtained (log file in Appendix IV) the raw images were acquired, with only the first 2000 images being used and processed as a result of radiation damage which caused deterioration in the data. This raw data was processed using XDS (Kabsch, 2010). The previously obtained Bd2503^{LBD} structure which shares 30% sequence identity with Bd2504^{LBD}, was then processed using CCP4 Sculptor (Potterton et al., 2018), before use as an initial model for molecular replacement using Phenix PHASER (McCoy et al., 2007), yielding a TFZ of 6.2. Although the TFZ was poor, the placement of the two copies was in the classical dCache dimer pose, and thus the model was built using Phenix Autobuild (Terwilliger et al., 2008) and COOT (Emsley et al., 2010), before the final structure was refined with Coot and Phenix Refine (Afonine et al., 2012) similarly to the protocol used for Bd2503^{LBD} (Table 3.12).

Resolution	2.02
Number of Copies in Asymmetric Unit	2
R_{Work}	0.2721
R_{Free}	0.3134
R.M.S Deviations	
Bond Length (Å)	0.01
Bond Angles (°)	1.10
Ramachandran Plot	
Favoured Regions(%)	94.92
Allowed Regions(%)	4.7
Outliers (%)	0.38

Table 3.12 Final refinement statistics for $Bd2504^{LBD}$ supplemented with putrescine.

The asymmetric subunit contained two copies of $Bd2504^{LBD}$, which like $Bd2503^{LBD}$ formed a symmetrical dimer around the N-terminal transmembrane helix. Both Chains A and B were continuously built from residue 2 to 264. However the density between residues 211 and 242 was poorer than the rest of the structure, however there was enough intermittent density to build this section of the structure. The two monomers are held in a dimeric confirmation by 13 hydrogen bonds along with hydrophobic packing (Figure 3.12), all of which are formed between the transmembrane helices unlike $Bd2503^{LBD}$. Sidechain-sidechain hydrogen bonds form between: the carboxyl group of E6 and the amine group of K10 and the amine group of K46 and E36 carboxyl group as well as the backbone carboxyl group of Q41. Sidechain-sidechain bonds were also observed between the amide of Q13 and the hydroxyl group of S17, whist the other monomer formed a intermolecular hydrogen bond from the Q13 amide group to the carbonyl of Q14, instead of S17. Like $Bd2503^{LBD}$, $Bd2504^{LBD}$ had one hydrogen bond that was not mirrored and only formed once, between the amide group of Q41 and the carbonyl group of Q49. Side chain-backbone hydrogen bounds also formed between: the amide group of Q25 and the carboxyl of A24, and the non-terminal sidechain amine group of R31 and the backbone carboxyl of E53.

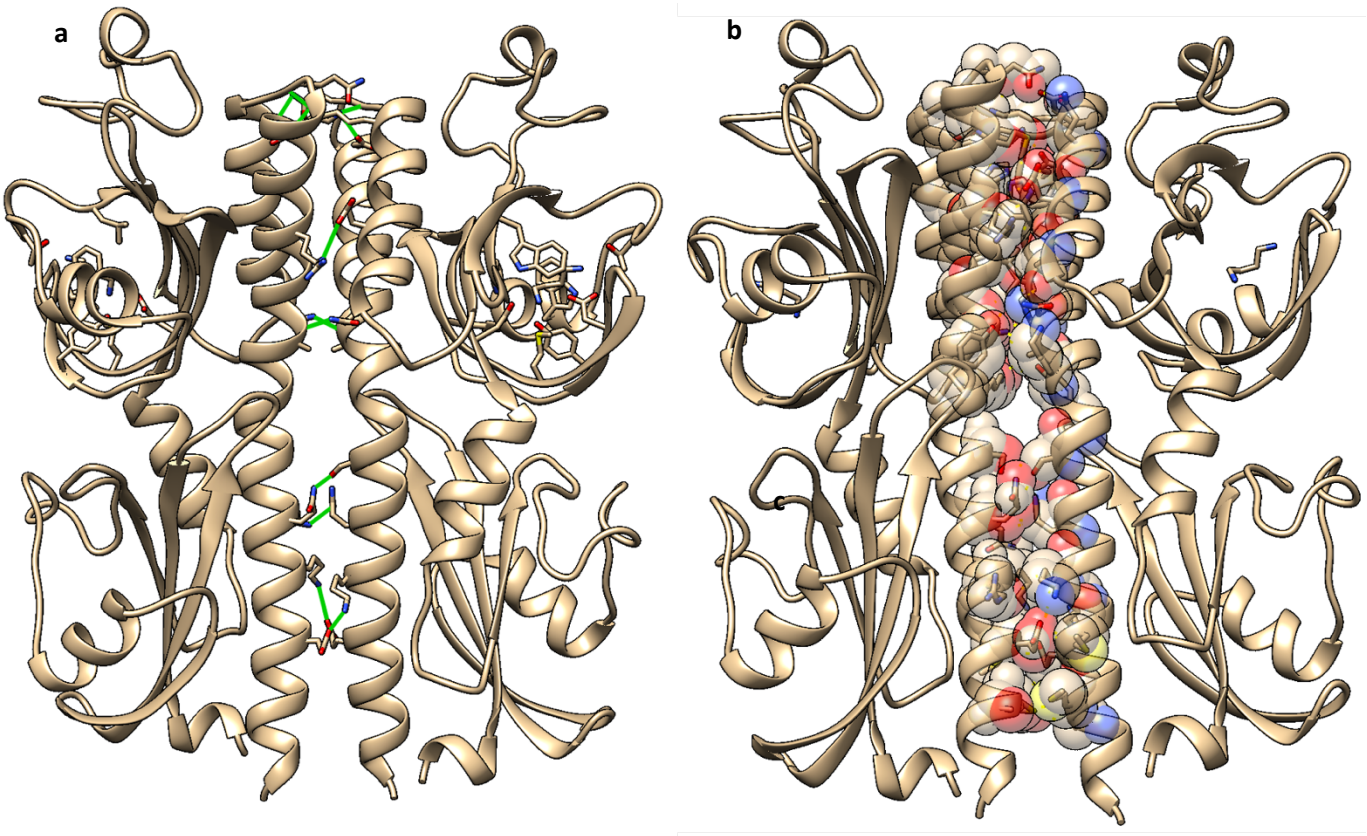


Figure 3.12 (a) Bd2504^{LBD} dimer with hydrogen bonds formed to stabilise dimer interaction shown in green. These hydrogen bonds are limited to the transmembrane helix **(b)** Bd2503^{LBD} dimer with hydrophobic packing between the two monomers displayed.

Having built Bd2504^{LBD} (Figure 3.13a) the refined map showed a region of strong density in the same region as cadaverine (Figure 3.13b). Although the area of density was less well-defined than cadaverine in Bd2503^{LBD}, the ligand was modelled as putrescine, as Bd2504^{LBD} had been supplemented with putrescine prior to crystallisation (Figure 3.14). Binding of putrescine takes place again in the distal lobe, but in a bent conformation. While modelling of Bd2504^{LBD} was underway, another polyamine binding dCache protein was identified by Gavira *et al* in *Pseudomonas putida*, who found McpU^{LBD} bound Putrescine (*Kd* = 2 μM), Cadaverine (*Kd* = 22 μM), and Spermidine (*Kd* = 4.5 μM) (Martín-Mora et al., 2018a). The structure of McpU^{LBD} bound to Putrescine showed the same bent putrescine binding proposition as seen between Bd2504^{LBD} and Putrescine, therefore supporting this observed binding conformation (Gavira et al., 2018). Like Bd2503^{LBD}, Bd2504^{LBD} had a LBP rich in hydrogen bond donors. The three aromatic amino acids (W76, W106, and W13), predicted to form

the cage around putrescine were confirmed. Fewer amino acids hold putrescine within the LBP, with a total of 5 hydrogen bonds supporting the binding. Four amino acids (Y146, D148, M157 and D176) towards the bottom of the pocket bind one putrescine amine group, whereas L119 and D148 bind the other. Unlike Bd2503^{LBD} which binds cadaverine vertically within the pocket parallel to the transmembrane helix, Bd2504^{LBD} binds the second amine group with amino acids on the opposite face of the pocket to the long transmembrane helix. It is this positioning of the binding amino acids that causes putrescine to assume this bent conformation.

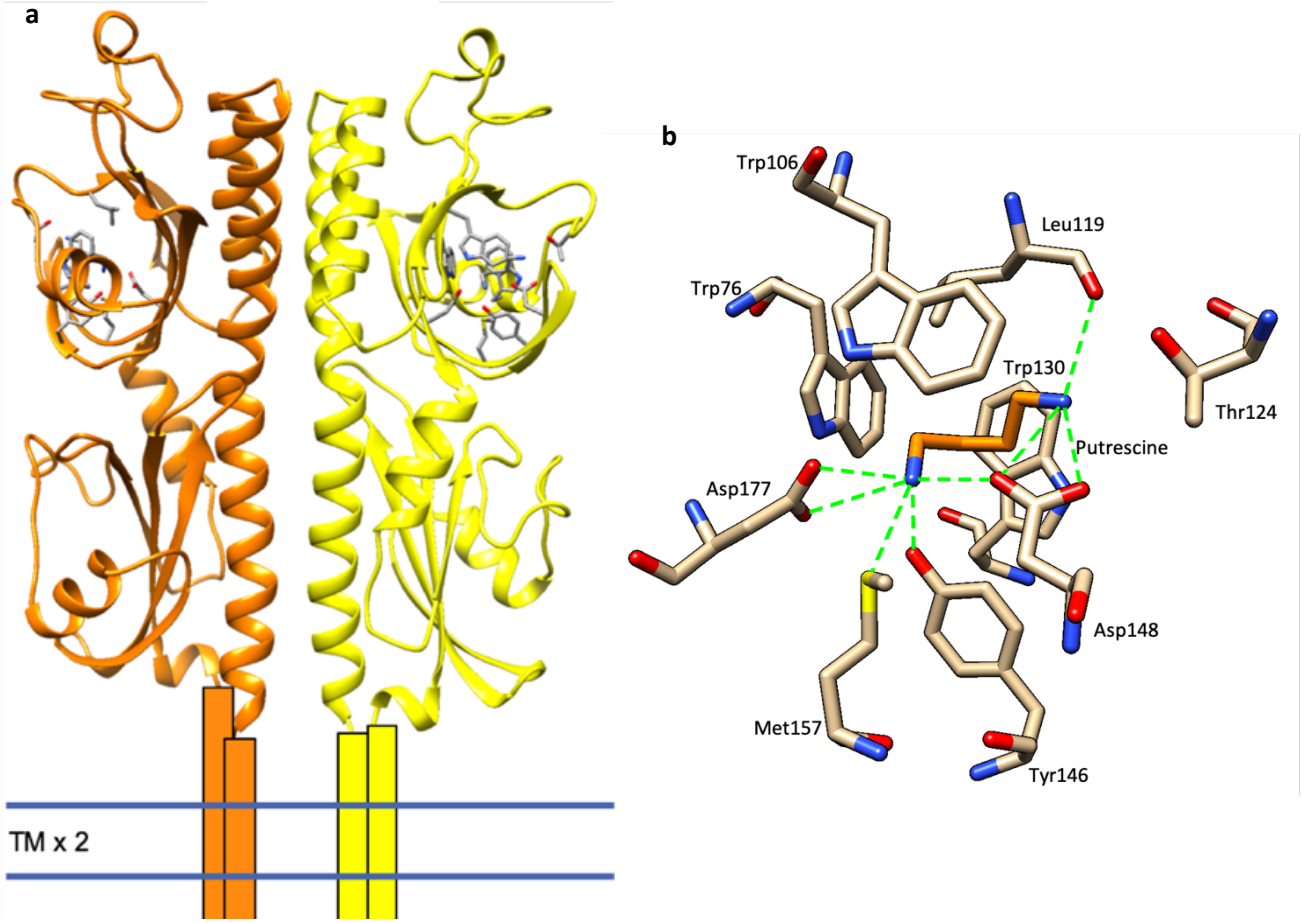


Figure 3.13 (a) Bd2504^{LBD} dimerisation. **(b)** Close up of Bd2504^{LBD} binding pocket with putrescine bound (orange), with the binding pocket amino acids shown and hydrogen bonds shown in green.

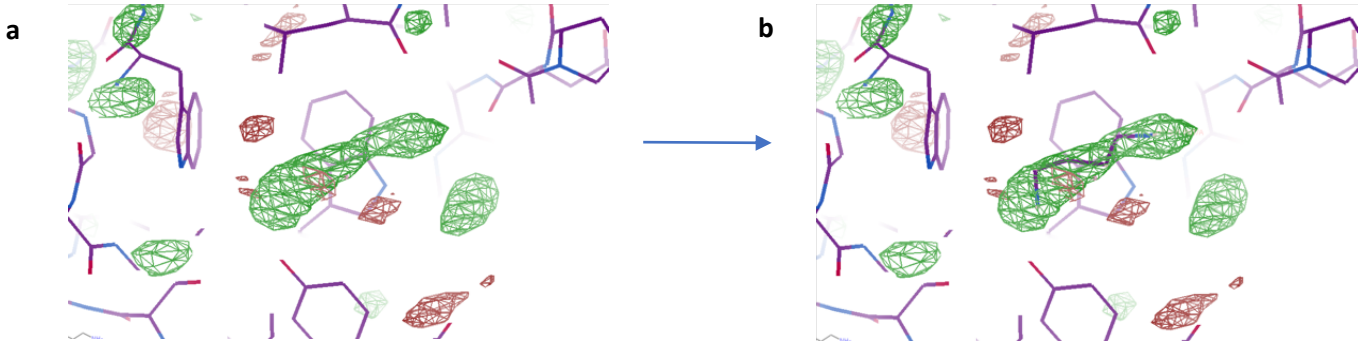


Figure 3.14 Coot FOFCWT PHFOFCWT density map of the binding pocket with sigma set to 3.0. **(a)** High resolution, unaccounted for, density in the ligand binding pocket. **(b)** Putrescine fitted into this density.

3.2.3 Structural Determination of Bd1081$^{S\text{-}LBD}$

3.2.3.1 Purification of Bd1081$^{S\text{-}LBD}$

Since the other two chemoreceptor homologues, Bd2503^{LBD} and Bd2504^{LBD}, had required a bound ligand for crystallisation, an attempt similar to before with Bd2504^{LBD}, to predict the specificity of Bd1081^{LBD} was preformed, by analysing the amino acid variation within the binding pocket (Figure 3.15). This identified a ligand binding site with an aromatic make up more similar to Bd2503^{LBD} than Bd2504^{LBD}, with conservation of the tryptophan and tyrosine but substitution of the phenylalanine for a second tyrosine. Although this would make the LBP less tightly packed than Bd2504^{LBD}, the sequence alignment also shows that Bd1081^{LBD} has an extended loop, two amino acids larger than Bd2503^{LBD}, the same as Bd2504^{LBD}. Therefore Bd1081^{LBD} is likely to have a more closed-off LBP, especially as this loop in Bd1081^{LBD} contains another tyrosine, unlike the two glutamic acids found in Bd2504^{LBD}. After this insertion, Bd1081^{LBD} has a lysine instead of the threonine found in Bd2504^{LBD}, and the conserved aspartate in Bd2503^{LBD} and Bd2504^{LBD} is a tyrosine in Bd1081^{LBD}. Both lysine and tyrosine are far bulkier side chains than the amino acids normally found in these positions. This would give rise to limited space for ligand binding, and would suggest Bd1081^{LBD} binds a smaller molecule/polyamine than Bd2504^{LBD}. Studies into *Bdellovibrio in-vivo* have shown that it is attracted to *Acinetobacter baumannii*. 1,3-diaminopropane (DAP, a 3 carbon polyamine) is the most abundant polyamine in *A. baumanii* and is used for gliding motility (Skiebe et al., 2012), which suggests one of the chemoreceptors in

Bdellovibrio senses DAP. As a results of the theory that Bd1081 possess a space-restricted LBD it was predicted that Bd1081 binds DAP.

Bd1081^{LBD} would not crystallise with either it's predicted ligand, DAP, or without. Since, attempts to elucidate the ligand from sequence alignment with homologues Bd2503^{LBD} and Bd2504^{LBD} had failed, the Bd1081^{LBD} homologue from *Bdellovibrio* SKB1291214 was cloned into *E. coli*. This protein was denoted Bd1081$^{S\text{-}LBD}$, as it originates from the SKB1291214 strain (Oyedara et al., 2016), and shares 61% identity with Bd1081^{LBD} dCache domain, and 73% within the ligand binding pocket. It was hoped that these minor variations would allow for crystallisation whilst still being representative of the HD100 Bd1081^{LBD} structure and ligand specificity. As the variations identified via the sequence alignment of Bd2503^{LBD}, Bd2504^{LBD} and Bd1081^{LBD} held true for Bd1081$^{S\text{-}LBD}$, Bd1081$^{S\text{-}LBD}$ was also thought to bind DAP.

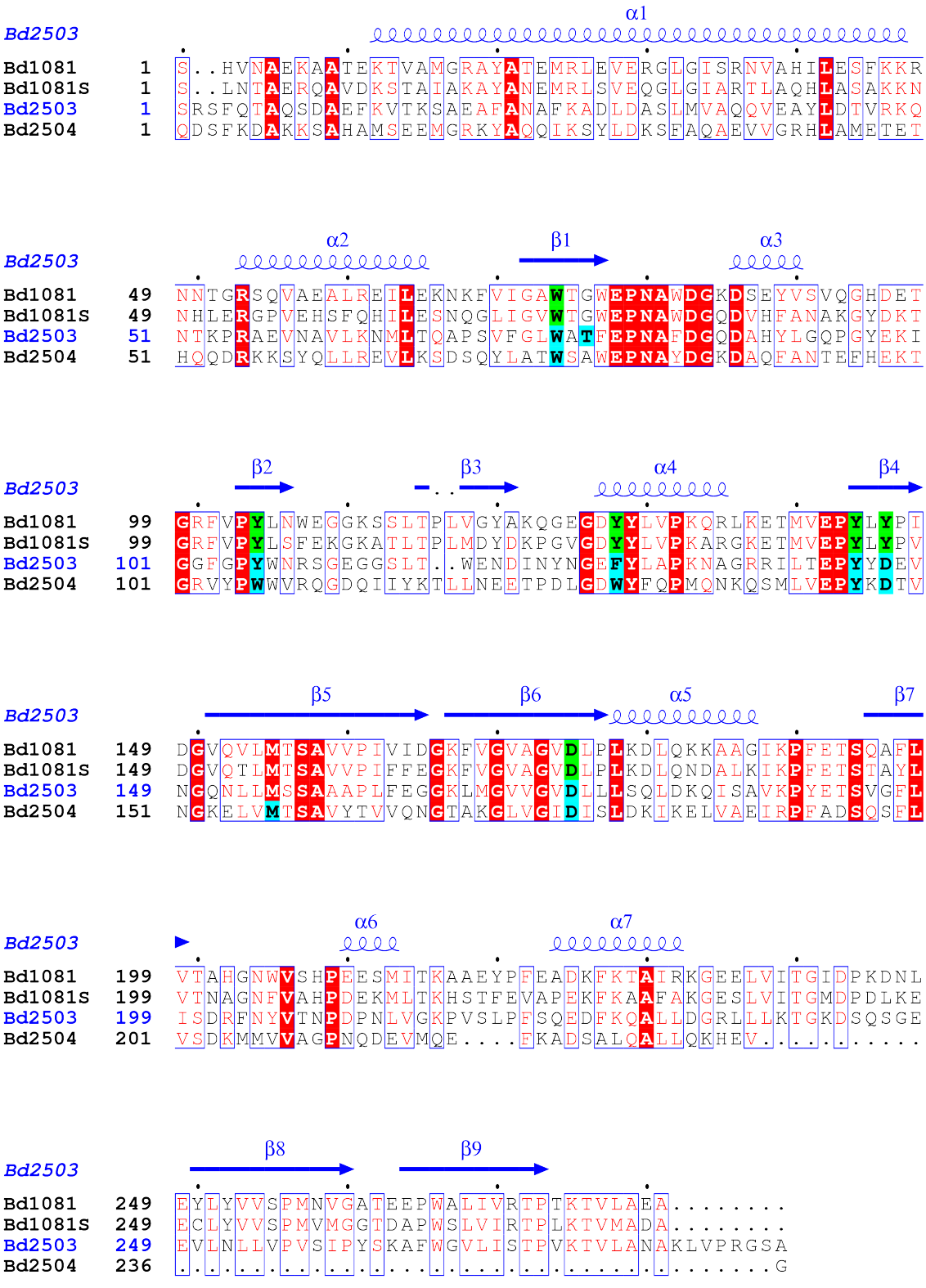


Figure 3.15 Alignment of Bd1081^{LBD}, Bd1081$^{S\text{-}LBD}$, Bd2503^{LBD} and Bd2504^{LBD}, with the secondary structure of Bd2503^{LBD} depicted above (generated using ESPript 3.0 (Robert and Gouet, 2014)). Amino acids in red boxes are strictly conserved, while those in red are chemically similar. Highlighted in blue are amino acid sidechains confirmed via structural analysis to be involved in ligand binding. Highlighted in green are amino acids predicted to be involved in ligand binding by sequence alignment.

The ligand binding domain of Bd1081$^{S\text{-}LBD}$ was expressed in AI media with sugar mix N and purified using nickel affinity chromatography eluted with as little as 20 mM imidazole; 100 and 300 mM imidazole eluted fractions only were used in order to optimise purity (Figure 3.16a). The purified protein was dialysed into 20 mM Hepes pH 7.5 and 300 mM NaCl, prior to SEC with a HiLoad 26/60 Superdex 75 pg (Figure 3.16c). SEC yielded one peak at 170 ml, which was assumed to be a monomeric peak as a result of the SEC elution volume.

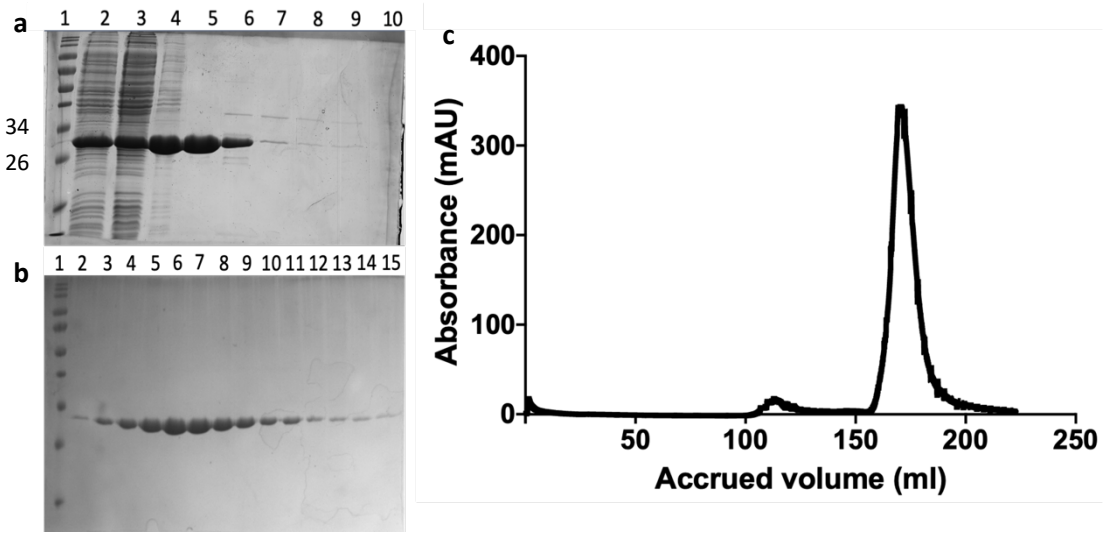


Figure 3.16 (a) SDS-PAGE gel of Bd1081$^{S\text{-}LBD}$ IMAC purification. **(b)** SDS-PAGE gel of peak 1 and peak 2 of Bd1081$^{S\text{-}LBD}$ post SEC purification. **(c)** UV trace Bd1081$^{S\text{-}LBD}$ SEC purification.

3.2.3.2 Data Collection and Structure Building of Bd1081$^{S\text{-}LBD}$

Crystal trays were therefore set up in native conditions, as well as with 1.1 mM DAP or 1.1 mM cadaverine, as the structure of Bd2503^{LBD} with L-Arginine had shown that the binding pocket could expand significantly. Multiple hits in JCSG^{+} were obtained for Bd1081$^{S\text{-}LBD}$, with the best being in JCSG^{+} H10 with Bd1081$^{S\text{-}LBD}$ at 17 mg/ml supplemented with 1.1 mM DAP and in a 1:1 drop ratio. This condition did not require any optimisation:

Method	Sitting drop vapour diffusion
Temperature (°C)	18
Plate Type	96 Well CombiClover 4 chamber crystallization plates
Protein Concentration (mg/ml)	20
Reservoir Solution	0.1 M Bis tris pH 5.5, 25% w/v PEG 3350, 0.2 M Ammonium acetate
Drop Volume (µl)	1.5
Drop Ratio	1:1
Volume of Reservoir (µl)	100
Crystallisation Morphology	3D Pear shape

Table 3.13 Crystallisation conditions for Bd1081$^{S\text{-}LBD}$.

Crystals were soaked with a cryoprotectant containing 20% v/v ethylene glycol and 1.1 mM DAP to maintain ligand binding during fishing. Crystal data were obtained at Diamond Light Source to the following specifications:

Diffraction Source	Diamond Light Source – i04		
Cryoprotectant	20% v/v Ethylene Glycol		
Wavelength (Å)	0.9795		
Rotation of Crystal (°)	360		
Rotation per Image (°)	0.1		
Images Collected	3600		
Exposure per Image (sec)	0.05		
Transmission (%)	100		
Detector Resolution (Å)	2.0		
Space Group	P 4 1		
Cell Dimensions			
a, b, c (Å)	72.88	72.88	101.58
α, β, γ (°)	90.0	90.0	90.0
Resolution Range (Å)	1.55 – 99.9		
Wilson B-Factor	23.95		
	Overall	High	
Total Reflections	831323	19987	
Unique Reflections	76532	3677	
R_{merge}	0.068	0.931	
R_{meas}	0.074	1.143	
$R_{p.i.m}$	0.029	0.647	
$I/\sigma I$	12.3	0.7	
CC(1/2)	0.999	0.797	
Multiplicity	10.9	5.4	
Completeness	99.9	97.3	

Table 3.14 Diffraction dataset specifications for Bd1081$^{S\text{-}LBD}$.

Once the diffraction data had been obtained (log file in Appendix V) initial search models were built using the distal lobe of Bd2503^{LBD}, which shared 35% sequence identity, and Bd2504^{LBD}, which shared 28%. As both were fairly good models, molecular replacement was performed with both models independently using Phenix PHASER (McCoy et al., 2007). Bd2504^{LBD} was found to be the better initial model, producing a preliminary solution with a higher TFZ, 8.9, and LLG score, 72.22. Once the two copies of the distal lobe had been placed, they were fixed in position and the proximal lobe was searched for using the proximal lobe of Bd2503^{LBD} and Phenix PHASER, producing a solution with a TFZ of 10.0. Phenix Autobuild (Terwilliger et al., 2008) was then run, alongside manual building in COOT (Emsley et al., 2010) to model missing residues. Final refinement used COOT and Phenix Refine (Afonine et al., 2012) (Table 3.15).

Resolution	1.55
Number of Copies in Asymmetric Unit	2
R_{Work}	0.2361
R_{Free}	0.2542
R.M.S Deviations	
Bond Length (Å)	0.01
Bond Angles (°)	1.37
Ramachandran Plot	
Favoured Regions(%)	96.52
Allowed Regions(%)	2.75
Outliers (%)	0.75

Table 3.15 Final refinement statistics for $Bd1081^{S\text{-}LBD}$.

$Bd1081^{S\text{-}LBD}$ had two copies in the asymmetric unit, however unlike the other dCache proteins, it did not form a dimer. Both chains were continuously built, residues for Chain A were placed from 3 to 279 and for Chain B from 6 to 278. There were no areas of patchy or missing density, again supporting the hypothesis that dimerisation interferes with the ordering of the proximal lobe in one or both of the chains. Only one hydrogen bond formed between the two $Bd1081^{S\text{-}LBD}$ monomers, likely due to them crystallising at a 90° angle to one another (Figure 3.17). This formed between the backbone carboxyl of K189 and the sidechain amide group of N22.

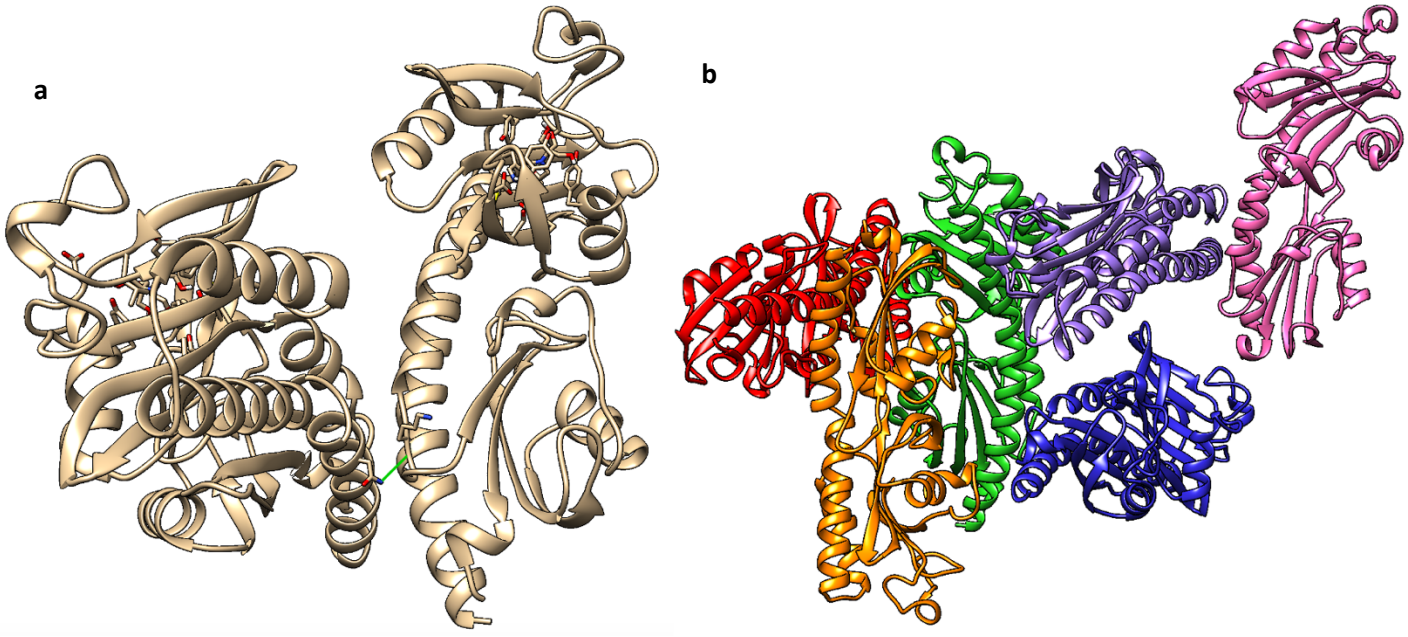


Figure 3.17 (a) Bd1081$^{S\text{-}LBD}$ contacts with the hydrogen bond formed between the two monomers shown in green. These hydrogen bonds are limited to the transmembrane helix. **(b)** Three copies of the asymmetric unit, shows that Bd1081$^{S\text{-}LBD}$ does not form the classic chemoreceptor dimer even when symmetry copies are viewed.

Having built Bd1081$^{S\text{-}LBD}$ (Figure 3.18) the refined map showed a region of strong density in the ligand binding pocket (Figure 3.19). However, the density was determined not to belong to DAP as predicted but Bis tris from the crystal screen. This has been deemed an artefact of crystallisation, rather than a true substrate, much akin to Zhang and Henderickson's structure for mmHK1S-Z2^{LBD} bound with Bis tris (Zhang and Hendrickson, 2010). Firstly, instead of forming a dimer, as is normally seen upon crystallisation, the two copies of Bd1081$^{S\text{-}LBD}$ crystallised at right angles to one another. This coupled with only a monomeric peak on SEC indicates that Bd1081$^{S\text{-}LBD}$ did not co-purify with its ligand. This indicates that Bis tris is unlikely to be a true ligand. Secondly, the LBP is once again hydrogen bond donator rich, meaning that Bis tris, which is also a hydrogen bond donator is unlikely to be a tight-binding ligand. This is supported by the last point, which is that unlike the other chemoreceptors where the sidechains are drawn into the LBP, the sidechains are flexed out of the pocket. This is due to repulsion between the Bis tris molecule and the sidechains, rather than the normal attraction forming hydrogen bonds pulling the sidechains in. When Bd1081$^{S\text{-}LBD}$ was compared to mmHK1S-Z2^{LBD},

very little difference between the two chemoreceptors was seen, with the majority isolated to the membrane proximal lobe (Figure 3.20). Within the distal lobe, a minor shift can be seen in the top most β-strands, and a slight change to the binding orientation of the Bis tris. Despite this, the proximal lobe seems to adopt the same conformation in both proteins, the angle of the domain with respect to the distal lobe, is however shifted causing the helix that attaches the two domains together to bend more within $Bd1081^{S\text{-}LBD}$. However the structure does seem to confirm predictions that $Bd1081^{S\text{-}LBD}$, and therefore likely $Bd1081^{LBD}$, has a cage formed from five aromatics, one tryptophan, W74, and four tyrosine residues (Y104, Y128, Y144 and Y146). Hydrogen bonds are formed between M118, Y144, Y146 and D175, and Bis tris to hold it in the binding pocket.

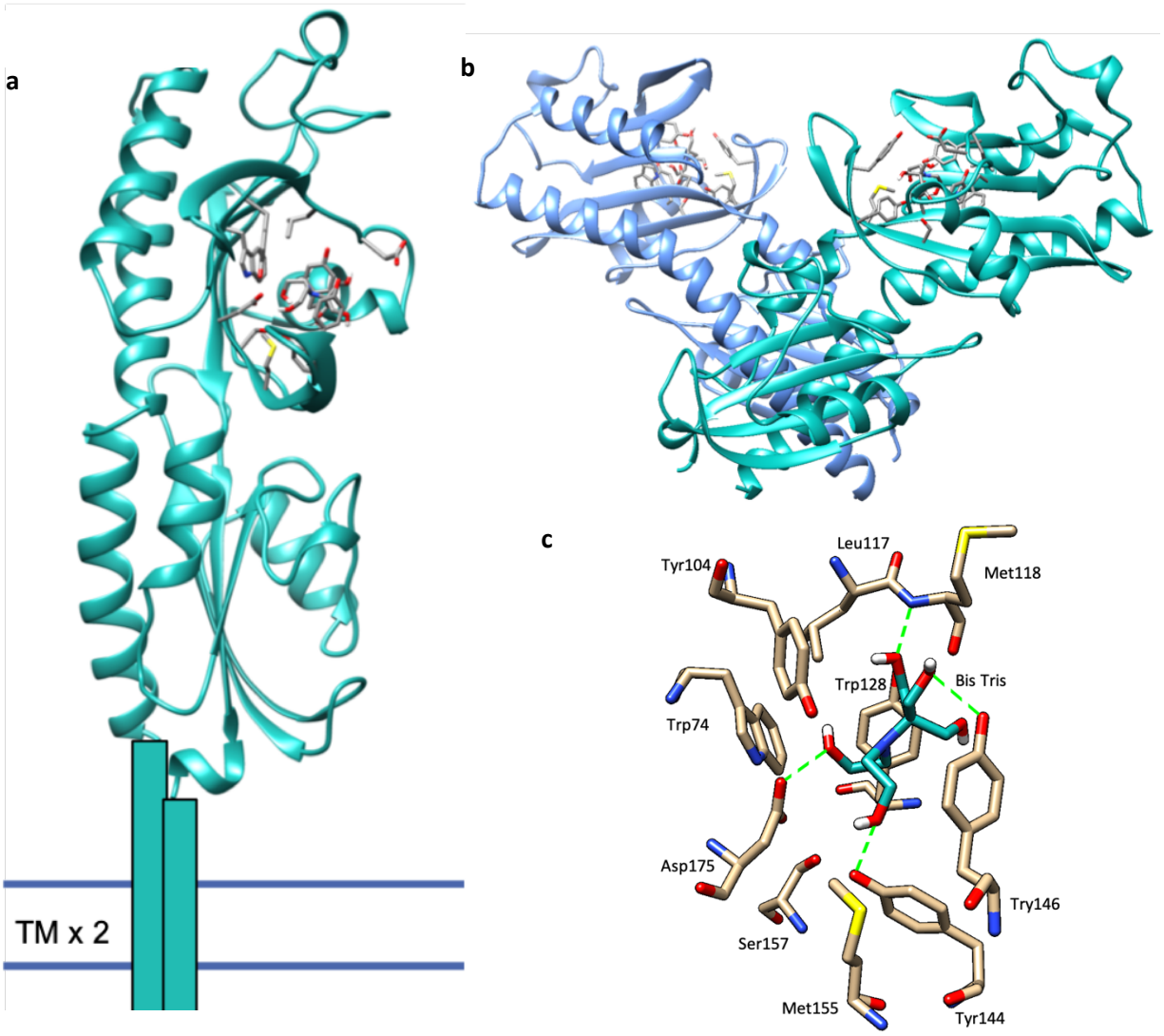


Figure 3.18 (a) $1081^{S\text{-}LBD}$ monomer with Bis tris in the binding pocket. **(b)** $1081^{S\text{-}LBD}$ asymmetric unit with contact due to crystallisation. **(c)** Close up of $Bd1081^{S\text{-}LBD}$ binding pocket with Bis tris (cyan) with hydrogen bonds shown in green

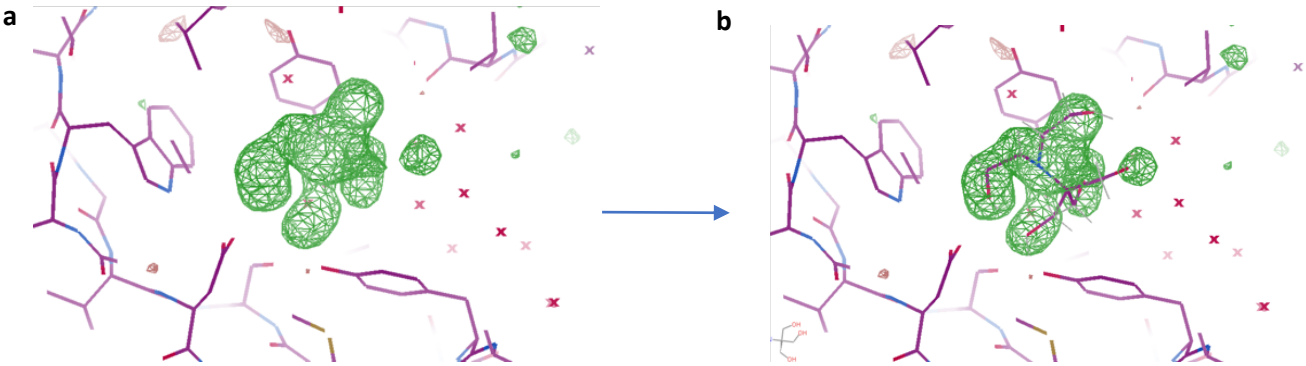


Figure 3.19 Coot FOFCWT PHFOFCWT density map of the binding pocket with sigma set to 3.0. **(a)** High resolution, unaccounted for, density in the ligand binding site. **(b)** Bis tris fitted into this density.

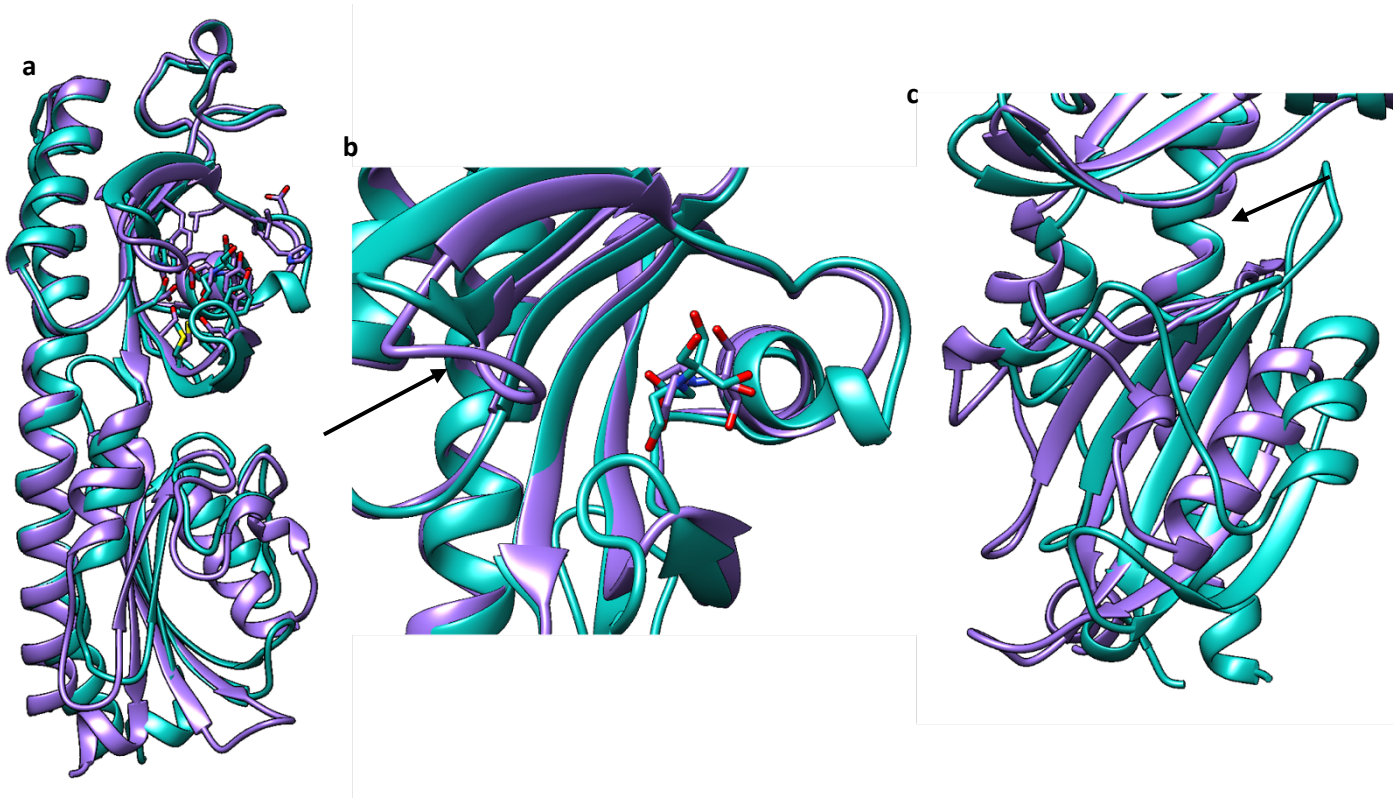


Figure 3.20 Overlay of Bd1081$^{S\text{-}LBD}$ (teal) and mmHK1S-Z2^{LBD} (purple). **(a)** 1081$^{S\text{-}LBD}$ and mmHK1S-Z2^{LBD} monomer with Bis tris in the binding pocket. **(b)** Close up of the distal lobe of 1081$^{S\text{-}LBD}$ and mmHK1S-Z2^{LBD}, an arrow (black) denotates the shifted β-strands. **(c)** Close up of the proximal lobe Bd1081$^{S\text{-}LBD}$ and mmHK1S-Z2^{LBD} showing the increased bend in the linking helix (black arrow) in Bd1081$^{S\text{-}LBD}$ compared to mmHK1S-Z2^{LBD}.

3.2.4 Structural Determination Bd3092

3.2.4.1 Purification of Bd3092

The ligand binding domain of Bd3092 was expressed in AI media with sugar mix N and purified using nickel affinity chromatography and eluted with as little as 20 mM imidazole; 50, 100 and 300 mM imidazole eluted fractions only were used to increase protein purity (Figure 3.21a). The purified protein was dialysed into 20 mM Bicine pH 8.0 and 300 mM NaCl, before SEC with a HiLoad 26/60 Superdex 75 pg (Figure 3.21c). SEC yielded one peak eluting at 165 ml, which was assumed to be a monomeric peak as a result of the SEC elution volume.

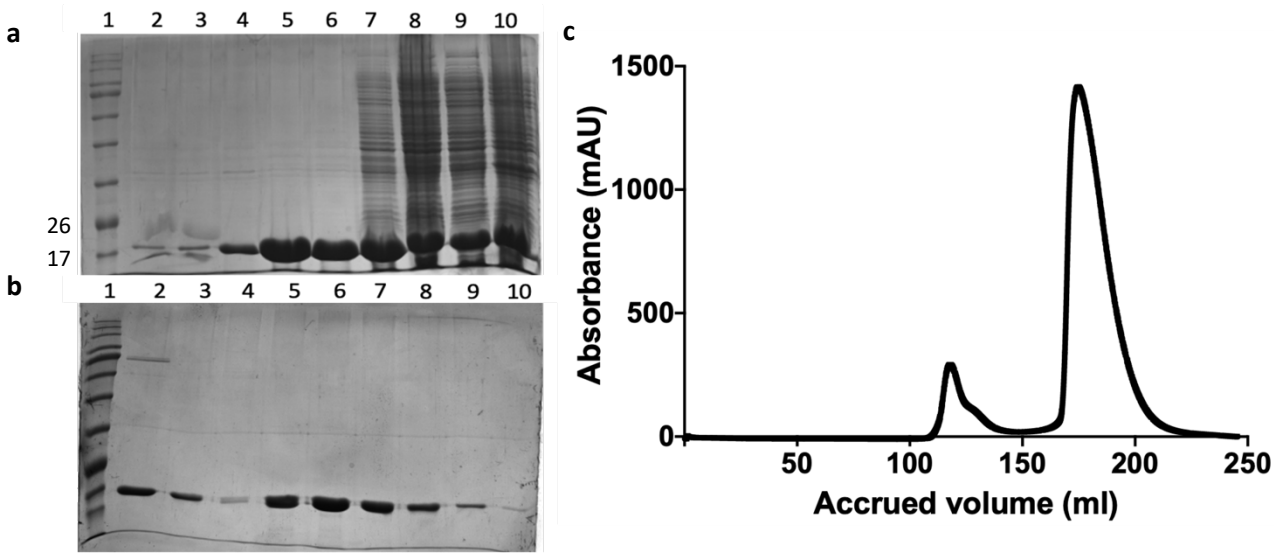


Figure 3.21 (a) SDS-PAGE gel of Bd3092 IMAC purification. **(b)** SDS-PAGE gel of peak 1 and peak 2 of Bd3092 post SEC purification. **(c)** UV trace Bd3092 SEC, showing both the monomer and dimer peak.

3.2.4.2 Data Collection and Structure Building of Bd3092

Bd3092 was originally crystallised in 20 mM Bicine pH 8.0 and 300 mM NaCl at 15 mg/ml in JSCG^{+} in condition D6. However due to closer analysis of the sequence of the sCache Bd3092, compared to the other three chemoreceptor proteins, Bd3092 appeared more likely to be a carboxylate acceptor, rather than a polyamine or related nitrogenous compound binder. This was supported by the literature that stated sCache domain chemotaxis proteins are known small carboxylic acid binders (Ortega et al., 2017a). Therefore the buffer was changed from Bicine to 20 mM Sodium acetate pH 4.6 and 300 mM NaCl, this change was tolerated well by the protein because the theorectical pI of the protein was 7.08. This led to crystals in BCS G11, with Bd3092 at 18 mg/ml and in a 1:1 drop ratio. These crystals were used for data collection without further optimisation:

Method	Sitting drop vapour diffusion
Temperature (°C)	18
Plate Type	96 Well CombiClover 4 chamber crystallization plates
Protein Concentration (mg/ml)	18
Reservoir Solution	0.1 M Hepes pH 7.5, 0.2 M Sodium/Potassium phosphate pH 7.5, 22,5% w/v PEG Smear Medium and 10% v/v Glycerol
Drop Volume (µl)	1.5
Drop Ratio	1:1
Volume of Reservoir (µl)	100
Crystallisation Morphology	Lozenge

Table 3.16 Crystallisation conditions for Bd3092.

Crystals were soaked with a cryoprotectant containing 20% v/v ethylene glycol. Crystal data were obtained at Diamond Light Source to the following specifications:

Diffraction Source	Diamond Light Source – i04		
Cryoprotectant	20% v/v Ethylene Glycol		
Wavelength (Å)	0.9795		
Rotation of Crystal (°)	360		
Rotation per Image (°)	0.1		
Images Collected	3600		
Exposure per Image (sec)	0.04		
Transmission (%)	79.91		
Detector Resolution (Å)	2.2		
Space Group	P 1 2_1 1		
Cell Dimensions			
a, b, c (Å)	55.53	157.29	60.99
α, β, γ (°)	90.0	102.39	90.0
Resolution Range (Å)	2.15 – 157.29		
Wilson B-Factor	25.92		
	Overall	High	
Total Reflections	363768	15101	
Unique Reflections	54863	2554	
R_{merge}	0.164	1.476	
R_{meas}	0.195	1.814	
$R_{p.i.m}$	0.106	1.039	
I/σI	7.2	1.7	
CC(1/2)	0.993	0.547	
Multiplicity	6.6	5.9	
Completeness	99.7	93.6	

Table 3.17 Diffraction dataset specifications for Bd3092.

Data for Bd3092 were collected to 2.3 Å (log file in Appendix VI). As Bd3092 is a single Cache domain, the other chemotaxis proteins could not be used as a molecular replacement model. Adeh_3718 (a chemotaxis receptor from *Anaeromyxobacter dehalogenans* (strain 2CP-C) with 49% sequence homology (4K08 (Pokkuluri et al., 2013)) was processed with CCP4 Sculptor (Potterton et al., 2018) to produce a model, which was then used for molecular replacement using Phenix PHASER (McCoy et al., 2007), before refining manually with COOT (Emsley et al., 2010) and automatically with Phenix Refine (Afonine et al., 2012) (Table 3.18).

Resolution	2.15
Number of Copies in Asymmetric Unit	4
R_{Work}	0.2438
R_{Free}	0.2743
R.M.S Deviations	
Bond Length (Å)	0.01
Bond Angles (°)	1.11
Ramachandran Plot	
Favoured Regions(%)	97.05
Allowed Regions(%)	2.79
Outliers (%)	0.16

Table 3.18 Final refinement statistics for Bd3092.

Bd3092 has four copies in its asymmetric subunit forming two separate dimers around the N-terminal transmembrane helices in the same manner as $Bd2503^{LBD}$ and $Bd2504^{LBD}$. However, like $Bd1081^{S\text{-}LBD}$ there were no sections of interrupted density, as there was no bottom lobe and the binding domain lobe was stabilised via ligand binding during crystallisation. All four copies were built from residue 1-163, meaning that none of the domain is missing. The two monomers are held in a dimeric conformation by 8 hydrogen bonds and hydrophobic packing (Figure 3.22), and like $Bd2504^{LBD}$, all were formed between the transmembrane helices. Sidechain-sidechain hydrogen bonds formed between: the carboxyl group of E26 and the amine group of K51, the hydroxyl groups of T27 and T27, as well as the carbonyl group of Q44 and the amide group of Q30. A sidechain-backbone hydrogen bond was also formed between the sidechain carboxyl group of D18 and the backbone amide group of S55.

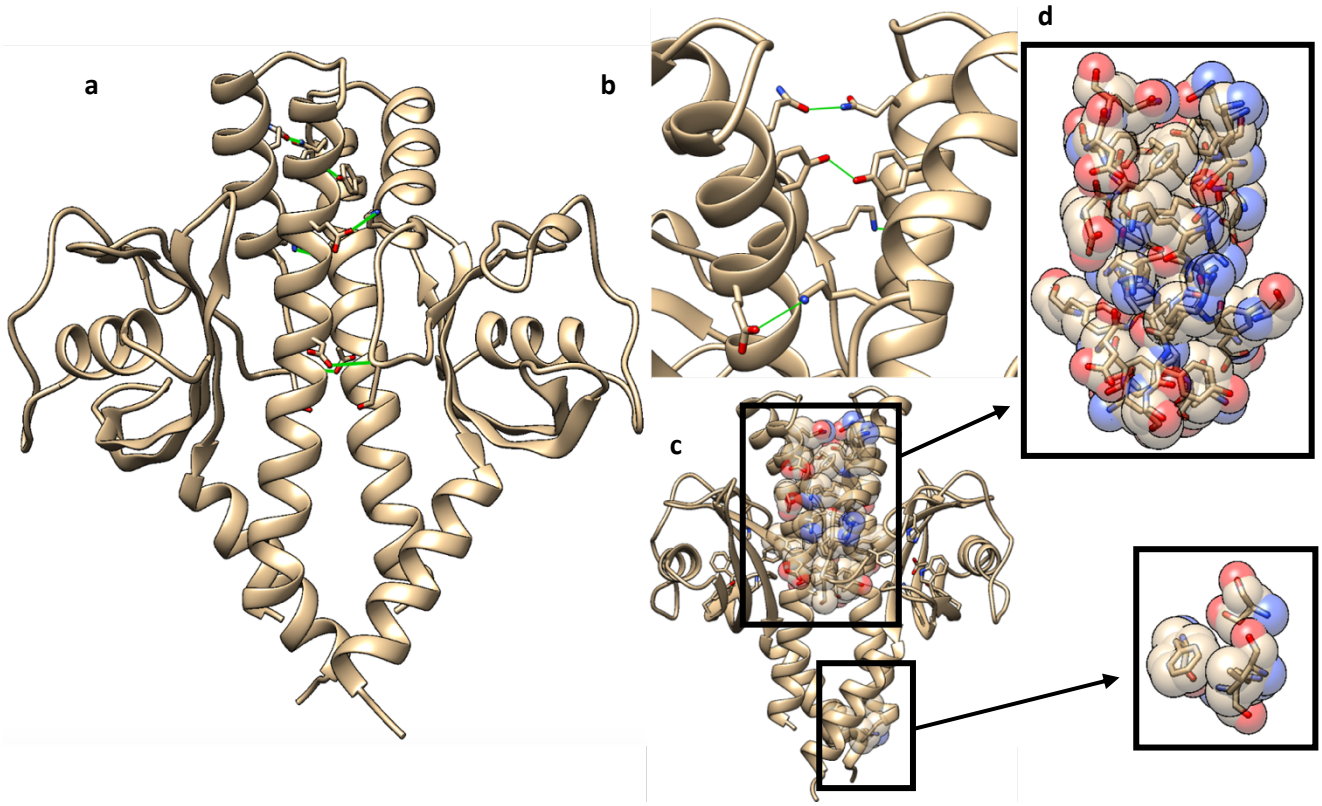


Figure 3.22 (a) Bd3092 dimer with hydrogen bonds formed to stabilise dimer interaction shown in green. These hydrogen bonds are limited to the transmembrane helix **(b)** Close up of the most membrane-distal hydrogen bonds. Bd3092 dimer with hydrophobic interactions displayed with **(c)** and without **(d)** the ribbon displayed.

Having built Bd3092 (Figure 3.23) the refined map did show an area of strong density equivalent to an acetate molecule in the binding site as predicted, which was subsequently modelled in (Figure 3.24). Bd3092, like the other Cache domain chemoreceptors utilises aromatic residues in this pocket (Y59, Y113 and W115) in order to bind acetate, however, they do not form a true cage. The two tyrosine residues point inwards forming hydrogen bonds with the ligand carboxyl group, instead of flanking the ligand with their aromatic rings as W115 does within this pocket and is observed in the dCache domain aromatic cages. Also, as predicted, the LBP contains hydrogen bond acceptors, H72 and K126, instead of the hydrogen bond donor amino acids seen in the dCache proteins. Each of these amino acids forms a stabilising hydrogen bond with one of the two oxygen groups of acetate.

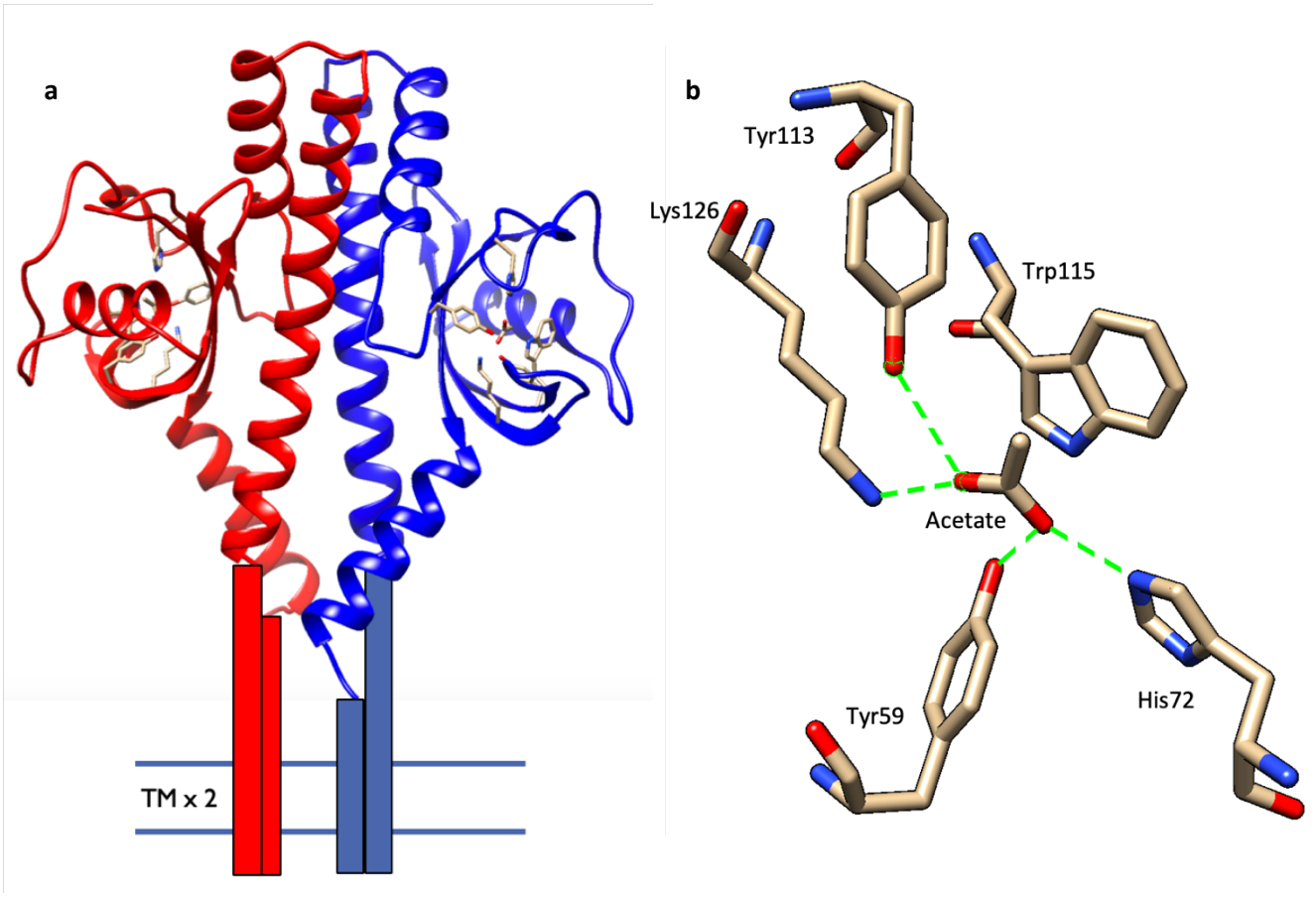


Figure 3.23 (a) Bd3092 dimerisation. **(b)** Close up of Bd3092 binding pocket with acetate bound (blue), with the binding pocket amino acids shown and hydrogen bonds in green.

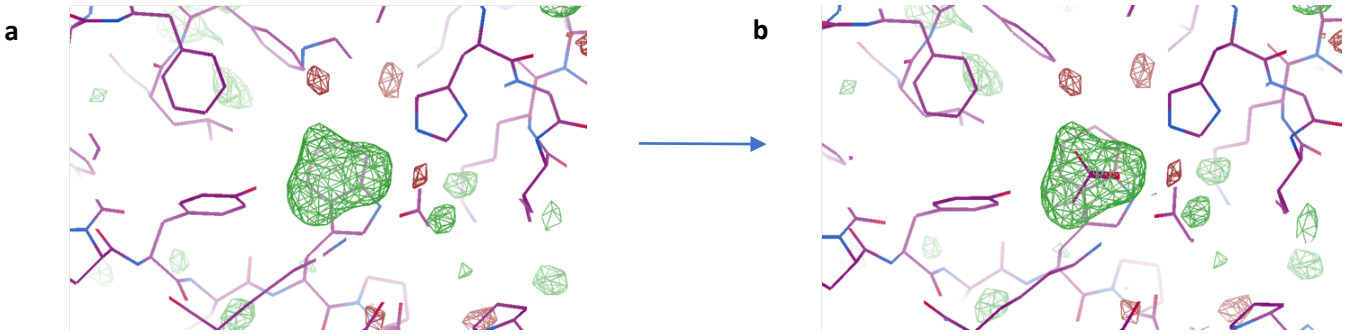


Figure 3.24 Coot FOFCWT PHFOFCWT density map of the binding pocket with sigma set to 3.0. **(a)** High resolution, unaccounted for, density in the ligand binding site. **(b)** Acetate fitted into this density.

3.2.5 Fold and Sequence Similarity

A DALI protein server search (Holm, 2020) and HHpred (Zimmermann et al., 2018) sequence alignment of existing structures was performed on all of the dCache chemoreceptors to find proteins with similar structure and sequence to identify possible ligands for Bd1081^{LBD} and Bd1081$^{S\text{-}LBD}$. The HHpred sequence alignment found many of the same top 5 hits for all three proteins (Table 3.19). Of the 7 proteins, 6 of them had been crystallised with a ligand molecule in the LBP. However there was a low likelihood they were possible ligands for Bd1081$^{S\text{-}LBD}$ based on the amino acids in the binding pocket. A Bis tris like molecule, as seen in mmHK1S-Z2^{LBD} and the Bd1081$^{S\text{-}LBD}$, had already been ruled out due to the repulsion between LBP side chains and Bis tris. For similar reasons, a diol like MPD seen in pHK1S-Z16, was also ruled out. The other four ligands bound were all amino acids, or related, molecules. As seen with Bd2503^{LBD}, it is possible for these dCache domain proteins to accommodate an amino acid in their binding site, however again the carboxylic acid group in amino acids would be somewhat undesirable within the hydrogen bond donating rich LBP. However if bound in a similar orientation as seen in Bd2503^{LBD}, an amino acid change of Asp120 to Gly120 as observed between Bd1081$^{S\text{-}LBD}$ and Bd1081^{LBD} might allow for this to occur.

Protein	Ligand Bound	Bd1081$^{S\text{-}LBD}$		Bd2503^{LBD}		Bd2504^{LBD}	
		Length of Alignment	Sequence ID %	Length of Alignment	Sequence ID %	Length of Alignment	Sequence ID %
CtaA^{LBD}	L-Isoleucine	235	19	247	17	238	18
Mcp^{LBD}_N	L-Alanine			237	17		
McpX^{LBD}	L-Proline Betaine					261	20
mmHK1S-Z2^{LBD}	Bis tris	267	34	278	25	266	24
pHK1S-Z16^{LBD}	MPD	240	11	251	12	241	10
PscC^{LBD}	L-Proline	244	22	250	14	245	18
TlpC^{LBD}		245	13				

Table 3.19 Top 5 HHpred (Zimmermann et al., 2018) unique sequences alignments for each of the three solved dCache domains found in *Bdellovibrio*.

Similar results were obtained with the DALI (Holm, 2020) structural analysis (Table 3.20-3.22). DALI ranks structure similarity hits based on a z-score, with a good score generally being considered above (n/10)-4, where n is the length of the query (Holm et al., 2008). Again the top five unique results were analysed, all of which are well above this threshold indicating that the dCache domains within *Bdellovibrio* are structurally very similar to other dCache domains already solved. Similar ligands were again identified by this approach. However, instead of the hydrophobic amino acids identified by sequence homology, TlpQ which binds histamine, a positively charged amine similar to amino acids, was identified. This would be a far more plausible ligand to bind Bd1081$^{S\text{-}LBD}$ or Bd1081^{LBD}.

Protein Name	Ligand Bound	Chain ID	PDB	Z-Score	RMSD	Length of Alignment	Sequence ID %
mmHK1S-Z2^{LBD}		A	3LI8	30.6	2.2	255	35
mmHK1S-Z3^{LBD}		E	3LIB	30.1	2.6	259	32
McpA^{LBD}		A	6PZJ	27.6	2.9	257	26
TlpQ^{LBD}	L-Histamine	A	6FU4	27.6	2.9	267	21
McpX^{LBD}	L-Proline Betaine	A	6D8V	27.5	3.4	263	25

Table 3.20 Top 5 DALI protein server (Holm, 2020) unique hits for Bd1081$^{S\text{-}LBD}$, threshold z-score = 23.7

Protein Name	Ligand Bound	Chain ID	PDB	Z-Score	RMSD	Length of Alignment	Sequence ID %
TlpQ^{LBD}	L-Histamine	B	6FU4	27.8	2.8	256	27
mmHK1S-Z3^{LBD}		E	3LIB	27.2	2.7	253	27
McpX^{LBD}	L-Proline Betaine	A	6D8V	27.1	3.0	256	26
McpA^{LBD}		A	6PZJ	26.6	2.9	252	26
McpU^{LBD}	Putrescine	A	6F9G	26.6	2.9	252	26

Table 3.21 Top 5 DALI protein server (Holm, 2020) unique hits for Bd2503^{LBD}, threshold z-score = 22.6

Protein Name	Ligand Bound	Chain ID	PDB	Z-Score	RMSD	Length of Alignment	Sequence ID %
mmHK1S-Z3^{LBD}		H	3LIB	29.4	2.3	254	26
McpA^{LBD}		A	6PZJ	29.0	2.2	251	24
TlpQ^{LBD}	L-Histamine	B	6FU4	29.0	2.1	258	21
McpU^{LBD}	Putrescine	A	6F9G	27.5	2.0	241	19
mmHK1S-Z2^{LBD}		A	3LI8	27.4	2.4	253	23

Table 3.22 Top 5 DALI protein server (Holm, 2020) unique hits for Bd2504^{LBD}, threshold z-score = 22.3

3.3 Discussion

3.3.1 The Polyamine Precedent

Prior to this work the relationship between *Bdellovibrio* and polyamines had been limited to speculations made in the 1980s that the polyamines putrescine and spermidine might be involved in stabilisation of the bdelloplast (Varon et al., 1983). While, the original *in-vivo* studies into chemotactic molecules were limited to common small molecules, non-specific to *Bdellovibrio*'s predatory lifecycle (LaMarre et al., 1977; Straley and Conti, 1974; Straley et al., 1979). This study is the first to identify specific ligands for a family of chemoreceptors within *Bdellovibrio*.

Polyamines are comprised of a hydrocarbon backbone with at least two primary amine groups, with longer carbon chains (6+) having additional secondary amine groups (Figure 3.25). At physiological pH these amine groups form co-ordinate bonds with free hydrogens, becoming protonated and giving the molecule an overall positive charge, making them polycationic (Park and Igarashi, 2013; Shah and Swiatlo, 2008). The polyamine family has been linked to a variety of functions within bacteria: DNA and RNA stabilisation, cell membrane integrity, stress response and porin regulation (Kamio and Nakamura, 1987; Shah and Swiatlo, 2008).

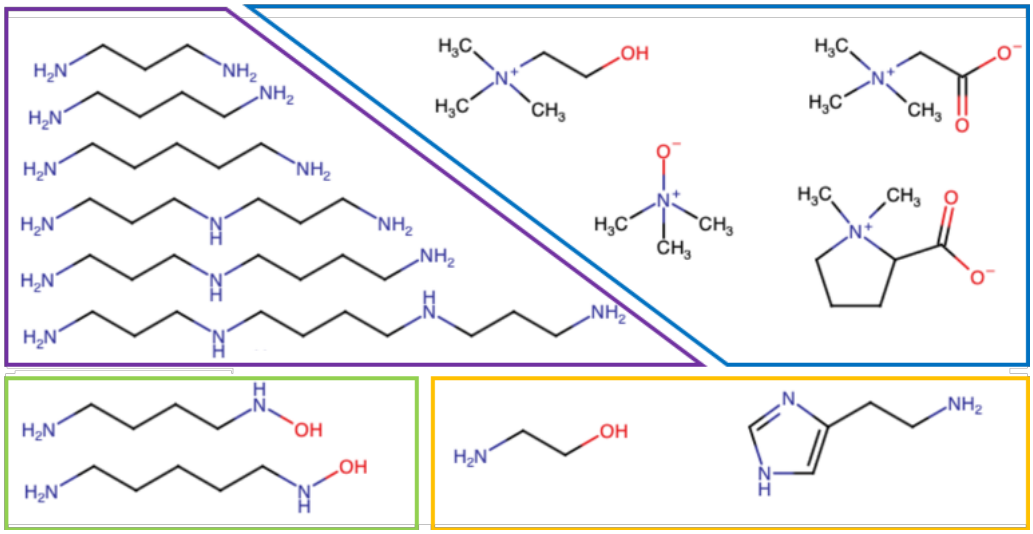


Figure 3.25 The polyamine family and chemically similar molecules. Purple box – Polyamines (top to bottom) bound by McpU^{LBD} in *P. putida*; 1,3 – Diaminopropane, Putrescine, Cadaverine, Norspermidine, Spermidine, and Spermine. Green box – Hydroxypolyamines (top to bottom), hydroxyputrescine and hydroxycadaverine. Blue box – Quaternary Ammonium Compounds (clockwise) bound by McpX^{LBD} in *S. meliloti*; Choline, Betaine, Proline betaine, and TMAO. Yellow box – Other amines (left to right), ethanolamine and histamine (bound by TlpQ in *P. aeruginosa)*.

Traditionally polyamines have been associated with functionally supporting small cations, Mg^{2+} and Ca^{2+}, with the stabilisation of DNA and RNA to improve replication, transcription, and translation. The involvement of polyamines improves the maintenance of DNA conformation compared to small ions alone. The hydrocarbon chain provides a defined distance between the positive charges, therefore forming electrostatic bridges between the negatively charged phosphates of the DNA backbone, stabilising the required conformation. Similarly the translation of RNA is improved by the involvement of polyamines, as they not only stabilise the quaternary structure of RNA but also interact with ribosomes to increase codon accuracy (Shah and Swiatlo, 2008).

Multiple studies have also shown that the binding of polyamines to porins causes inhibition of solute exchange. Putrescine and spermidine have been shown to bind to OmpF and OmpC aspartate residues, changing the porin conformation and causing it to close. Spermidine has been shown to have a greater effect than putrescine on porin activity. Dela Vega and Delcour in 1996 found that

polyamines also had an effect on the chemotaxis of *E. coli*, with diminishing effect as the chain length grew shorter (Spermine > Spermidine > Putrescine ~ Cadaverine). However, they concluded this effect was likely due to polyamine closure of the porins, rather than direct binding to *E. coli's* chemoreceptors (Dela Vega and Delcour, 1996; Lopes and Sourjik, 2018; Shah and Swiatlo, 2008).

Long chain polyamines, spermidine and spermine, have also been shown to play a role as free radical scavengers to aid Superoxidative dismutase in the protection of DNA from oxidative stress. Further uses of polyamines as an iron control agent are seen in *Bacillius anthracis,* which uses spermidine in the control of free iron levels, and *Vibrio cholerae/vulnificus,* both of which use norspermidine as an iron scavenger (Shah and Swiatlo, 2008). *Yersina pestis*, *Vibrio* and *Bacillus subtilis* require polyamines for biofilm formation. *Y. pestis* needs intrinsic putrescine production to enable attachment to solid surfaces in order to form biofilms, whereas *Vibrio* rely on their predominant polyamine, norspermidine, for biofilm formation. Arginine decarboxylases in *Bacillus subtilis*, and their downstream product spermidine, have been shown to be essential for biofilm formation, whereas the slightly shorter norspermidine prevents biofilm formation (Burrell et al., 2010; Hobley et al., 2014; Shah and Swiatlo, 2008).

The Gram-negative bacterium *Veillonella alcalescens* requires cadaverine or putrescene to be covalently linked to 40% of its peptidoglycan layer, via D-glutamic acid α-carbons, for normal growth (Kamio and Nakamura, 1987). Another Gram-negative bacterium that relies on cadaverine is *Selenomonas ruminantium* which uses covalent linking of cadaverine to the peptidoglycan to anchor the outer membrane to the cell wall (Kojima et al., 2011). Moreover, in gut bacteria cadaverine and putrescene are almost exclusively produced by Gram-negative bacteria, conversely this specificity is not seen with the longer polyamines. Also, Gram-negative bacteria can exclusively derive cadaverine and putrescene from amino acids (Pugin et al., 2017). Due to the Gram-negative centric production of

short polyamines, and the dependency bacteria seem to have on them, *Bdellovibrio* using these molecules as a functional chemotactic ligand is a viable hypothesis.

Polyamines are not unheard of as chemotactic attractants as both TAAR13c in zebrafish (Li et al., 2015) and McpU^{LBD} in *Pseudomonas putida* (Gavira et al., 2018) have already been shown to bind polyamine ligands. However, the only polyamine chemoreceptor that has been modelled with its ligand, McpU^{LBD}, binds putrescine at a distorted angle, similar to *Vibrio cholerae* Mlp37^{LBD} binding taurine. This contrasts to the straight vertical binding of cadaverine within the Bd2503^{LBD} binding pocket, but is reminiscent of how Bd2504^{LBD} binds putrescine (Figure 3.26) (Gavira et al., 2018; Nishiyama et al., 2016). Although, taurine is not a polyamine, it is often tested alongside them, as it is similarly derived from amino acid modification. *P. putida* actually uses the McpU^{LBD} receptor to navigate towards favourable environments (Corral-Lugo et al., 2016). Therefore, *Bdellovibrio* could use this same ligand to follow a similar trajectory as its prey to locate them, akin to lions hunting prey at a watering hole in the macro-world. Similar to this *Bdellovibrio* has been seen to predate planktonic *Acinetobacter baumannii* within a laboratory environment (Dashiff et al., 2011). The predominant polyamine produced by *Acinetobacter* is 1,3-diaminopropane (DAP), which has a 3 carbon backbone. DAP has been found to be involved in *A. baumannii* motility (Skiebe et al., 2012), and if DAP is found to be one of the polyamines *Bdellovibrio* responds to, this may explain why *Bdellovibrio* can so readily hunt *A. baumannii.*

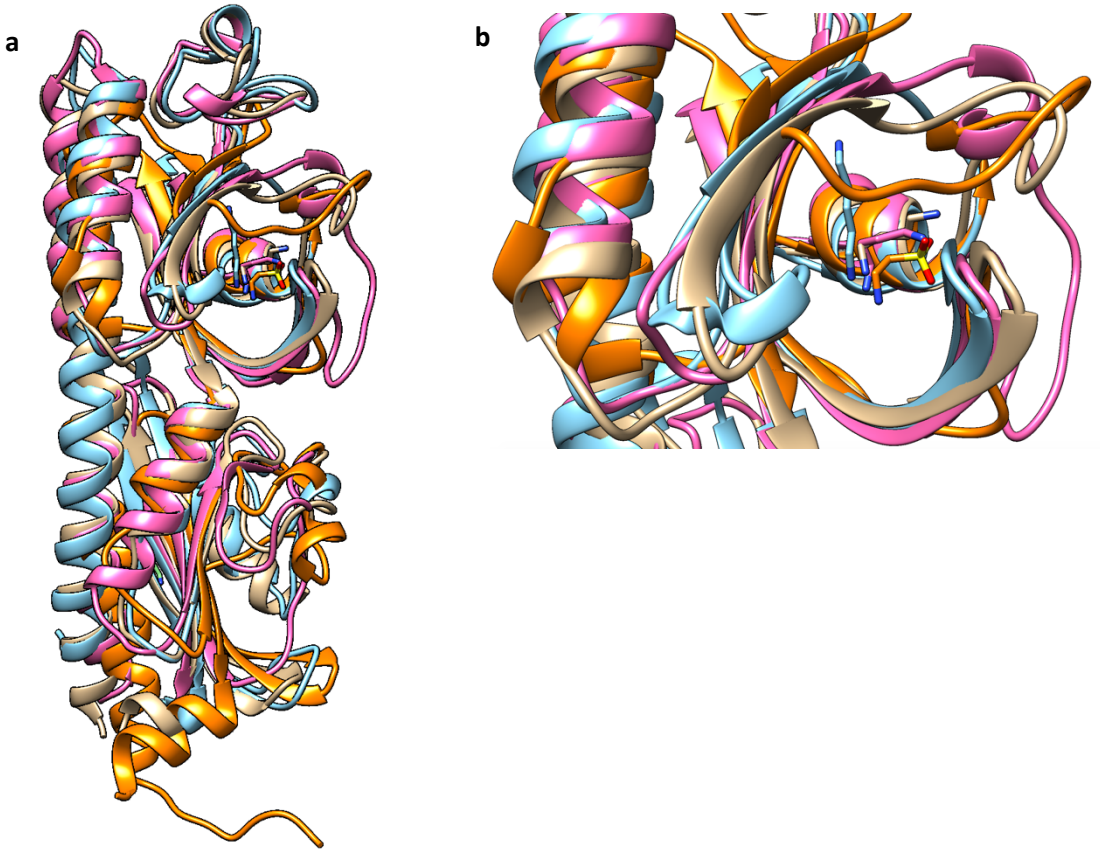


Figure 3.26 (a) Structural comparison of McpU^{LBD} (pink (Gavira et al., 2018)) Mlp37 (orange (Nishiyama et al., 2016)), Bd2503^{LBD} (blue), and Bd2504^{LBD} (tan). **(b)** Close up of the ligand binding pocket, showing that Bd2504^{LBD} binds putrescine in a conformation similar to McpU^{LBD} and Mlp37 binding putrescene and taurine respectively, but at a 90° angle to the binding of cadaverine within the Bd2503^{LBD} pocket.

3.3.2 dCache Sensory Domain Structural Comparison

There is little difference in the number of amino acids used to bind cadaverine and putrescine by Bd2503^{LBD} and Bd2504^{LBD} respectively, and sequence alignment reveals that the binding amino acids are equivalent to one another. However the positioning of them within the ligand binding pocket ultimately determines the binding conformation of the ligand. An overlay of Bd2504^{LBD} against Bd2503^{LBD} reveals an extended loop in Bd2504^{LBD} formed by amino acids L109 to T113 (Figure 3.27). The conformation of this loop creates a ligand binding pocket that is shorter and wider in Bd2504^{LBD} compared to Bd2503^{LBD}. This causes putrescine to assume a bent conformation when binding side-on within the Bd2504^{LBD} pocket, as opposed to the straight rigid binding of cadaverine parallel to the transmembrane helix within Bd2503^{LBD}. This variation within the ligand binding pocket likely also

confers the binding specificity of the chemoreceptors. Putrescine would be too short to bind in the Bd2503^{LBD} LBP, whereas cadaverine is too long to fit into the LBP of Bd2504^{LBD}. This in turn would explain why Bd2504^{LBD} would not crystallise when supplemented with cadaverine, but would with putrescine, despite the overall similarity between the two molecules. This is supported by analysis of the binding pockets of both when Van der Waals surfaces are calculated, which shows a limited capacity for the ligand to bind within the pocket, along with the filling of the pocket upon binding (Figure 3.28). Conversely McpU^{LBD} in *Pseudomonas putida* has been shown to bind both putrescine (*Kd* = 2 μM) and cadaverine (*Kd* = 22 μM) with fairly high affinity (Martín-Mora et al., 2018a). As already stated Bd2504^{LBD} binds putrescine in a conformation similar to McpU^{LBD} binding putrescine (Gavira et al., 2018), and structural alignment of the two LBPs shows that this is because McpU^{LBD} utilises a similar binding pocket to Bd2504^{LBD} (Figure 3.29a). Further analysis of the LBP in McpU^{LBD} reveals that the amino acid sidechains positioned in the distal region of the pocket are hydrogen bond donor-rich. However, none of these amino acids are positioned in a manner that would allow McpU^{LBD} to bind cadaverine in a conformation similar to that observed for Bd2503^{LBD}, therefore suggesting that cadaverine likely binds within the McpU^{LBD} LBP in a conformation similar to putrescine in both the McpU^{LBD} and Bd2504^{LBD} structures (Figure 3.29b).

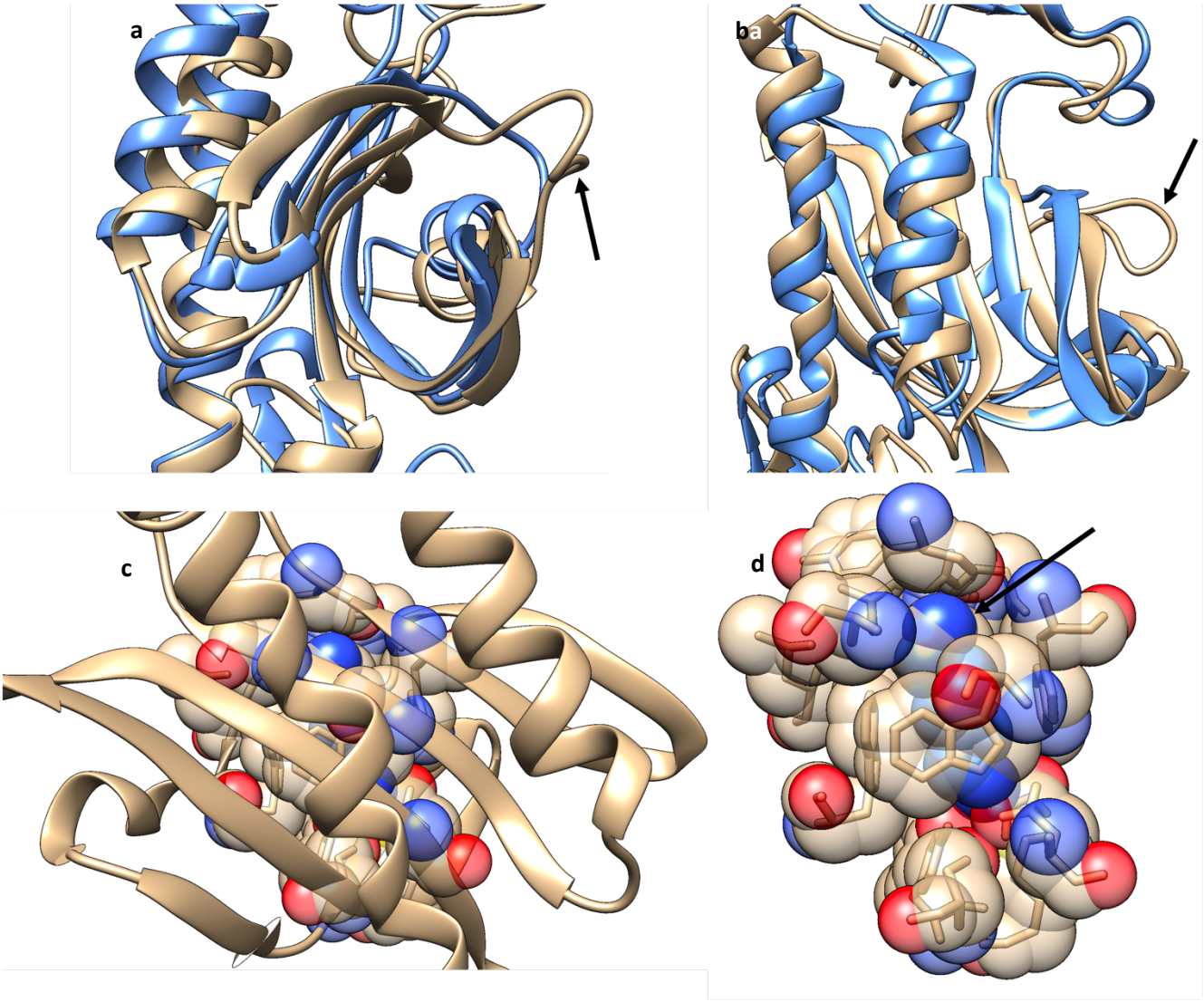


Figure 3.27 (a) Close up overlay of the binding pocket of Bd2503^{LBD} (blue) and Bd2504^{LBD} (tan) superimposed against one another. An arrow points to the extended loop that closes off the binding pocket of Bd2504^{LBD} when compared to Bd2503^{LBD}. **(b)** Represents a 45° turn compared to **(a)**, with the black arrow indicating the extended loop in Bd2504^{LBD} compared to Bd2503^{LBD}. **(c)** Close up of Bd2504^{LBD} (tan) with cadaverine (blue) from the Bd2503^{LBD} structure superimposed within the LBP, at 90° with respect to **(a)**. Cadaverine and the amino acids involved in binding putrescine to Bd2504 along with those in the loop that closes off the binding pocket (L120 and N121) are shown with Van der Waals surfaces. **(d)** Depicts the same as **(c)** but without the secondary structure backbone. An arrow indicates where L119 in Bd2504^{LBD} (tan) clashes with Cadaverine (blue) as a result of the extended loop in Bd2504^{LBD} forcing L119 further into the pocket in Bd2504^{LBD} compared to Bd2503^{LBD}, this in turn would prevent Bd2504 from binding cadaverine leading to ligand specificity.

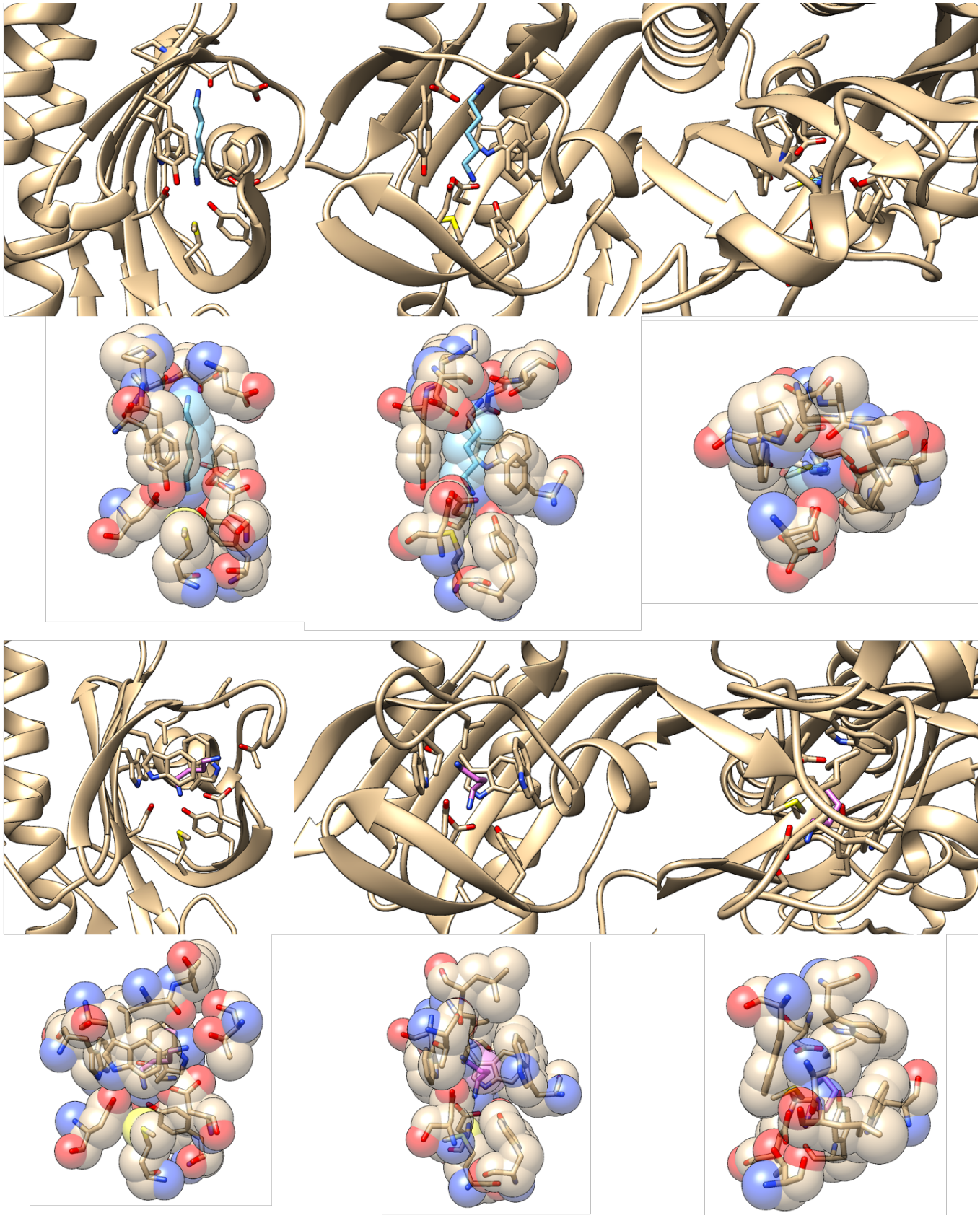


Figure 3.28 Displaying the binding pockets of Bd2503^{LBD} with cadaverine (top) and Bd2504^{LBD} with putrescine (bottom) with Van der Waals surfaces shows that the ligand fills the pocket upon binding. It is this limited space within the pocket that adds to ligand specificity.

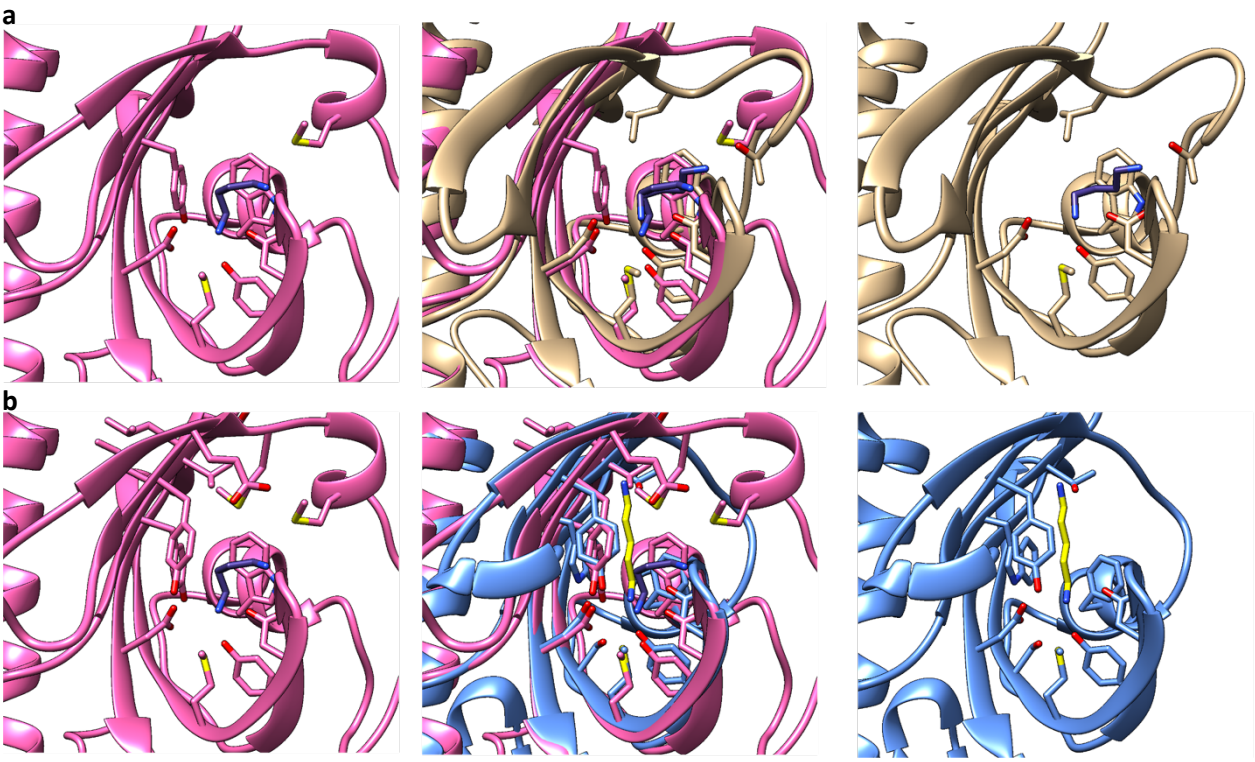


Figure 3.29 (a) Close up of the binding pockets of Bd2504^{LBD} (tan) superimposed against McpU^{LBD} (pink (Gavira et al., 2018)) both with putrescine bound. Both structures show a similar pattern of amino acids involved in stabilising the binding of putrescine within the pocket. **(b)** Close up of the binding pockets of Bd2503^{LBD} (blue) with cadaverine bound superimposed against McpU^{LBD} (pink (Gavira et al., 2018)) with putrescine bound. In addition to the amino acids known to be involved in putrescine binding within McpU^{LBD}, those amino acids in the distal region of the LBP are shown. This demonstrates none are positioned in a manner that would confer binding of cadaverine within the pocket in a confirmation similar to Bd2503^{LBD}, suggesting cadaverine is bound in a manner similar to putrescine in McpU ^{LBD}.

Alignment of all three dCache HD100 chemoreceptors along with Bd1081^{SLBD}, McpU^{LBD} and McpX^{LBD} allowed for prediction of the amino acids involved in ligand binding in Bd1081^{LBD} and Bd1081$^{S\text{-}LBD}$. The structure of the Bd1081$^{S\text{-}LBD}$ Bis tris complex was used to verify if the amino acids predicted through sequence homology would be physically able to bind a ligand within the LBP. McpU^{LBD} was chosen for this alignment as it is the other structurally determined polyamine binder. McpX^{LBD} was selected as it was the best HHpred (Zimmermann et al., 2018) sequence match for Bd1081^{LBD} found in bacteria with a confirmed ligand. This revealed 17 highly conserved amino acids across all 6 sequences (Figure 3.30), 3 of which were within the N-terminal transmembrane helix, and the other 14 within the membrane distal lobe. A previous alignment by Shrestha *et al*, comparing McpX^{LBD} to 4 different dCache

chemoreceptors identified 8 strictly conserved amino acids, all but the last of which are located within the distal lobe of the domain (Shrestha et al., 2018). However, three of these were not conserved throughout all the chemoreceptors compared here. The amino acid equivalent to T174 in $MlpX^{LBD}$ was only found in $Bd2504^{LBD}$ (T144), while the equivalent amino acid to G227 in $MlpX^{LBD}$ was in none of the *Bdellovibrio* chemoreceptors but was in $McpU^{LBD}$ (G253). The final amino acid, Y162 in $MlpX^{LBD}$ was seen in all but $McpU^{LBD}$ (L187), where the tyrosine was substituted for leucine. This divergency between polyamine binders and other dCache domains with known structure may help dictate which amino acids are truly conserved. However, it does not seem to help with the identification of possible ligands for $Bd1081^{LBD}$ and $Bd1081^{S\text{-}LBD}$, as both sequences share as much sequence homology with the polyamine binders as $McpX^{LBD}$ in terms of the key ligand binding amino acids. Conversely, at the same time shows the same key amino acids can be used for the binding of multiple ligand families. As such, $Bd1081^{LBD}$ and $Bd1081^{S\text{-}LBD}$ could be quaternary ammonium compound binders, supported by identification of a tyrosine in $Bd1081^{LBD}$ and $Bd1081^{S\text{-}LBD}$ positioned equivalently to the key binding amino acid F153 in $McpX^{LBD}$, within the area of highest binding variance between the aligned dCache domain proteins.

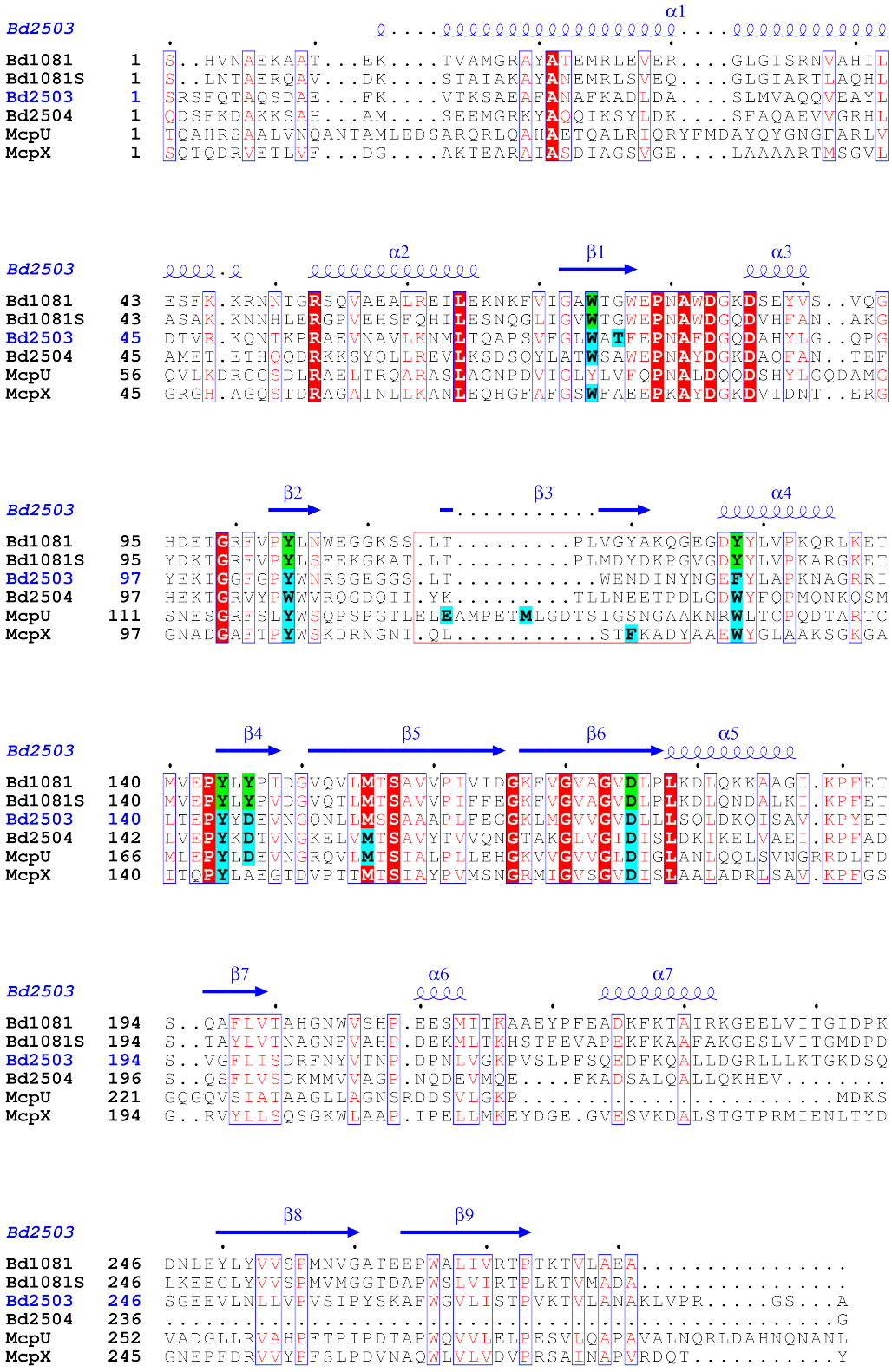


Figure 3.30 Alignment of Bd1081^{LBD}, Bd1081$^{S\text{-}LBD}$, Bd2503^{LBD} and Bd2504^{LBD}, McpU^{LBD} and McpX^{LBD}, with the secondary structure of Bd2503^{LBD} depicted above (generated using ESPript 3.0 (Robert and Gouet, 2014)). Amino acids in red boxes are strictly conserved, while those in red are chemically similar amino acids. Highlighted in blue are amino acid sidechains confirmed via structural analysis to be involved in ligand binding. Highlighted in green are amino acids predicted to be involved in ligand binding by sequence alignment. The red box indicates the area of high involvement in binding capability and alignment variance.

Bd3092 binding acetate does not match up with the other *Bdellovibrio* bound structures, however, as it is a sCache, it was unlikely to bind polyamines. Furthermore, binding acetate aligns with our knowledge from the literature of sCache domain ligand preferences. Studies by Straley *et al* in 1979 did show *Bdellovibrio* to be attracted to acetate and similar small molecules, propionate and thioacetate (Straley et al., 1979). This is supported by analysis of the binding pocket with Van der Waals2. surfaces modelled, which shows that although the pocket is filled around the carboxylic acid group and methyl group, F95 which makes contact with the methyl group has room to flex away from the methyl group and expand the binding pocket, meaning there is physically space for the binding of a longer chain organic acids within the pocket (Figure 3.31). Acetate is a very useful molecule for bacteria as it can be utilised for a number of purposes as a carbon source or a precursor to acetyl-CoA, which can be used in signalling or the production of fatty acids (Matsuoka and Shimizu, 2013; Zhang et al., 2019; Cronan and Thomas, 2009). Although not as specific to Gram-negative bacteria as the polyamines cadaverine and putrescine, acetate would make a good attractant for *Bdellovibrio* either to hunt down prey, or as a small molecule *Bdellovibrio* can use to partly sustain themselves between kills.

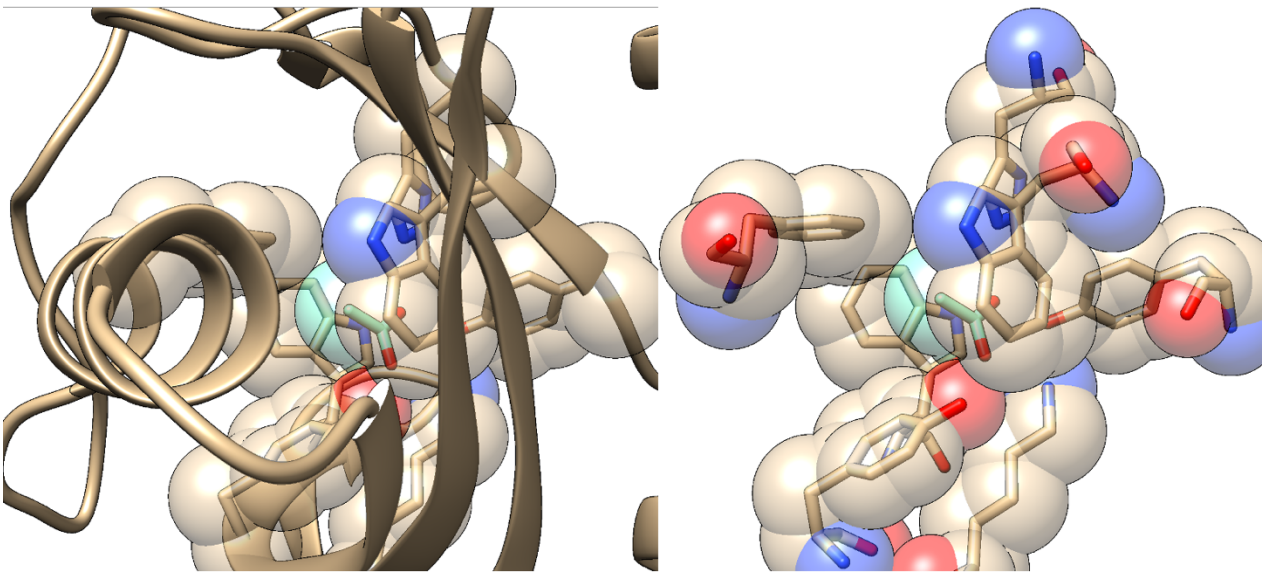


Figure 3.31 Displaying the binding pockets of Bd3092 with acetate with Van der Waals surfaces that the ligand does fill the pocket upon binding, however F95 which makes contact with the methyl group can move to allow a longer organic acid ligand to bind.

Chapter 4 – Testing the ligand binding of *Bdellovibrio bacteriovorus* Chemoreceptors

4.1 Introduction

After the determination of ligands for 3 of the 4 Cache domain chemoreceptor proteins, via x-ray crystallography, biochemical techniques were applied to verify ligand assignment and probe specificity. Firstly, high-throughput liquid chromatography (HPLC) was used to confirm the assignment of cadaverine as the ligand that had co-purified and crystallised with Bd2503. A secondary purpose for using HPLC was to establish a method by which apo forms of the chemoreceptors could be obtained/confirmed. These apo chemoreceptors were subsequently used in isothermal titration calorimetry experiments to determine ligand binding specificity and affinity for each Cache chemoreceptor. The possible ligands predicted for $Bd1081^{LBD}$ and $Bd1081^{S\text{-}LBD}$ via HHpred (Zimmermann et al., 2018) and DALI (Holm, 2020) searches were also tested via ITC and crystallography. Upon the identification of ligands for $Bd1081^{LBD}$ and homologue $Bd0181^{S\text{-}LBD}$, a disparity in specification and affinity was identified between the two (discussed below). This led to the production of three mutants of $Bd1081^{S\text{-}LBD}$: $Bd1081^{S\text{-}LBD}$ D120G, $Bd1081^{S\text{-}LBD}$ V126E, and $Bd1081^{S\text{-}LBD}$ D120G/V126E, in order to ascertain the cause of this difference between the two homologues.

4.2 Results

4.2.1 High-Throughput Liquid Chromatography

To confirm the assignment of cadaverine and putrescine as the bound ligands in the crystal structures of Bd2503 and Bd2504, respectively, HPLC was performed. This was especially important for Bd2503 which had been co-purified natively with cadaverine. A secondary aim was to verify that the method by which apo chemoreceptor proteins were made was effective.

Firstly, polyamine standards of cadaverine, putrescine and spermidine were ran to determine if the method of derivatisation using dansyl chloride allowed for the ligands to be detected by HPLC. This also meant that the retention times for each polyamine could be determined, and therefore conclude whether the HPLC method being employed would result in adequate separation between peaks to be able to differentiate between the polyamines. Specific peaks for each polyamine were seen at the following retention times: 15.1 min for cadaverine, 14.9 min for putrescine and 16.4 min for spermidine (Figure 4.1a); these were similar to times previously reported in the literature (Moret et al., 2005). The differences between each retention time were considered sufficient to differentiate between each polyamine. A final standard of imidazole was also ran, this is because imidazole contains an amine group, which could be derivatised by the dansyl chloride (Moret et al., 2005), and therefore theoretically produce a specific peak which overlaps with those of the polyamines being tested, which would prevent ligand identification. Imidazole did not produce a specific peak, with the only major peak being produced coinciding with the common retention time peak at 12.8 min. This peak is thought to be caused by the addition of proline during derivatisation to quench the action of the dansyl chloride (Moret et al., 2005).

Although imidazole was shown to not produce a peak that interfered with the assignment of the polyamine ligands, steps were taken to limit the presence of it within the HPLC samples in order to

prevent it acting as a competitor of the native ligands during derivatisation. This led to an adapted purification method whereby following elution of contaminants from the IMAC column using 100 mM imidazole, the column was washed with 50 column volumes of 20 mM Hepes pH 7.5 and 300 mM NaCl to remove any imidazole present. Bd2503/Bd2504 samples were then eluted using 20 mM Hepes pH 7.5, 300 mM NaCl and 50 mM EDTA, before dialysis and SEC. Native samples of Bd2503 and Bd2504 prepared in this way were analysed using HPLC, demonstrating no peak at the specific retention times for the polyamines (Figure 4.1b and Figure 4.1c), with the exception of a very small peak in the Bd2503 native sample at the retention time for cadaverine. Samples of the proteins prepared in the same manner were supplemented with cadaverine and putrescine, and left to equilibrate before derivatisation with dansyl chloride. These samples were prepared to determine if binding of the ligands to the protein interfered with the derivatisation process. These samples produced a peak at the specific retention time of their expected ligand, indicating that the proteins were not interfering with derivatisation. Therefore it was concluded that washing the protein with 50 column volumes of imidazole-free buffer removed the ligand that was co-purified with the protein. Identification of a similar small peak coinciding with the retention time of putrescine in the sample of Bd2504 supplemented with cadaverine seems to support this theory (Figure 4.1c). Furthermore, Bd2504 was found to crystallise only upon supplementation with putrescine.

These observations led to a reproducible method for producing apo Cache proteins, wherein three rounds of dialysis with SEC buffer for 12 hours was applied post-size exclusion. This method was tested with Bd2503 using HPLC, which was shown to lack a peak for cadaverine at the specific retention time (Figure 4.1b). Having successfully obtained apo protein, re-crystallisation of Bd2503 supplemented with cadaverine was undertaken to further confirm that the original ligand identification was correct. This led to the crystallisation of Bd2503 with 2 mM cadaverine in commercial screen condition JCSG$^+$ E10 which was reported in the previous chapter.

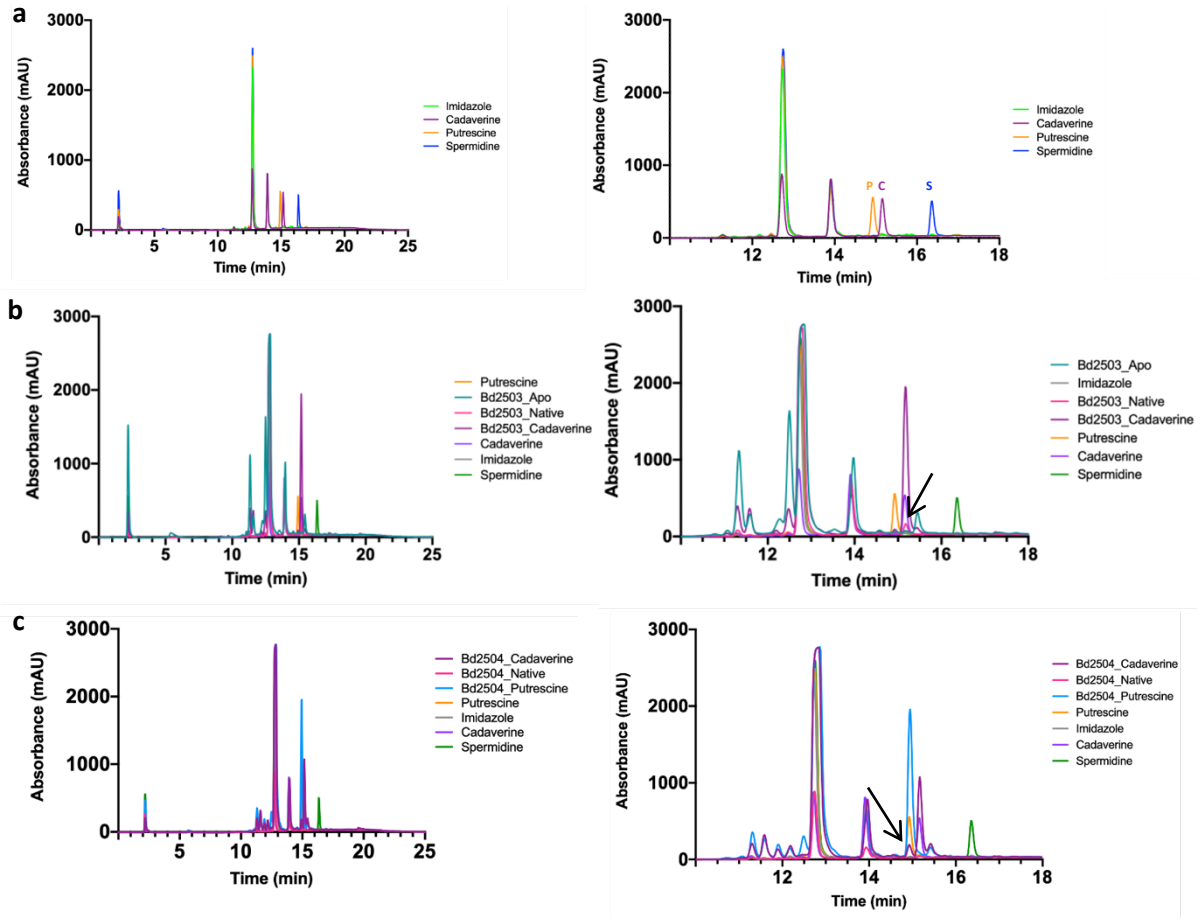


Figure 4.1 HPLC traces, in all figures the righthand panel displays the whole traces, and the lefthand panel the area of the trace with the specific retention times for polyamines (Moret et al., 2005). The specific peaks for each polyamine standard are indicated with an initial, P = putrescine, C = cadaverine, and S = Spermidine. **(a)** HPLC traces of polyamine standards cadaverine, putrescine and spermidine, as well as imidazole. **(b)** HPLC traces of Bd2503 native, apo and supplemented with cadaverine, as well as the standards. The black arrow indicates the small cadaverine peak in the Bd2503 native sample. **(c)** HPLC traces of Bd2503 native, and supplemented with cadaverine or putrescine, as well as the standards. The black arrow indicates the small putrescine peak in the Bd2504 sample supplemented with cadaverine.

4.2.2 Structural Determination of Bd1081 Ligand Binding Domain

Initial attempts to crystallise the ligand binding domain (LBD) of Bd1081 ($Bd1081^{LBD}$) were unsuccessful, and therefore the initial structure for this dCache was obtained for homologue $Bd1081^{S\text{-}LBD}$ from *Bdellovibrio SKB1291214* (61% sequence identity between dCache domains). Analysis of the $Bd1081^{S\text{-}LBD}$ structure (Chapter 3.2.5), along with sequence comparisons of $Bd1081^{LBD}$ against $Bd2503^{LBD}$, $Bd2504^{LBD}$, $Bd1081^{S\text{-}LBD}$, $McpX^{LBD}$ and $McpU^{LBD}$ (Chapter 3.3.2) were performed in order to identify the possible ligand(s) for $Bd1081^{LBD}$. Crystallisation of the $Bd1081^{LBD}$ was then reattempted using this information to obtain a structure of the Bd1081 receptor ligand binding domain from *Bdellovibrio bacteriovorus HD100*, as future whole cell chemotaxis studies would be performed using this strain, and therefore ligand binding had to ultimately be studied using the $Bd1081^{LBD}$. A secondary aim was to further support the apo confirmation of the $Bd1081^{S\text{-}LBD}$ structure via comparison to the bound structure of close homologue $Bd1081^{LBD}$.

4.2.2.1 Purification of $Bd1081^{LBD}$

$Bd1081^{LBD}$ was expressed in AI media with sugar mix N and purified using nickel affinity chromatography, eluted with 100 mM Imidazole (Figure 4.2a). The purified protein was dialysed into 20 mM Hepes pH 7.5 and 300 mM NaCl, before a SEC step with a HiLoad 26/60 Superdex 75 pg (Figure 4.2b). SEC yielded two peaks, a dimer peak eluting at 120 ml and a monomer peak eluting at 160 ml.

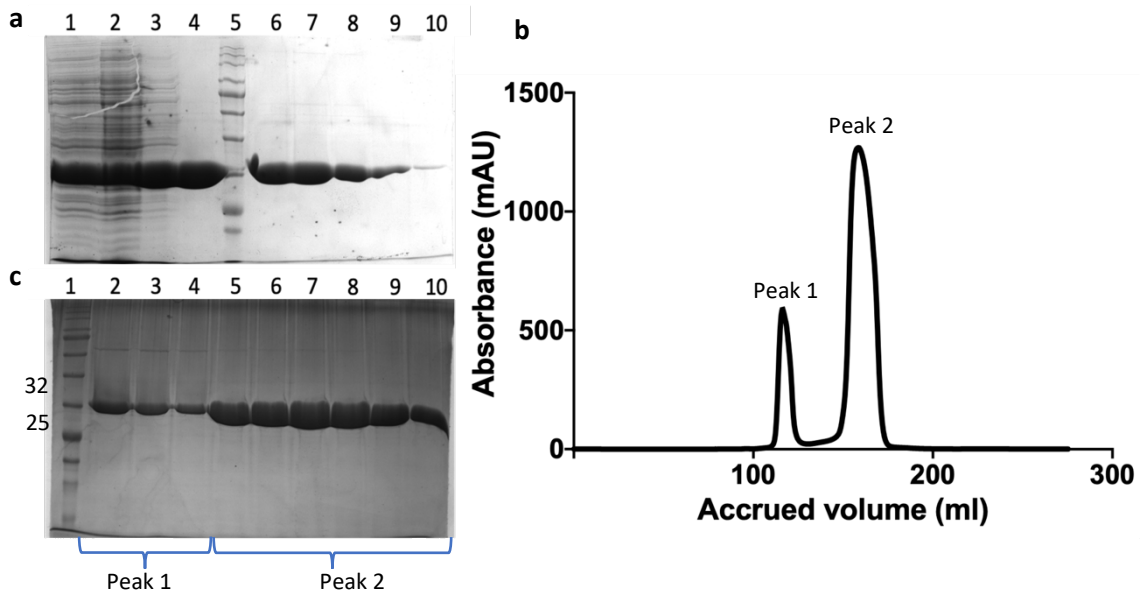


Figure 4.2 (a) SDS-PAGE gel of Bd1081 IMAC purification. **(b)** SDS-PAGE gel of peak 1 and peak 2 of Bd1081 post SEC purification. **(c)** UV trace Bd1081 SEC purification showing both the monomer and dimer elution fractions.

4.2.2.2 Data Collection and Structure Building of Bd1081^{LBD}

As previously discussed in Chapter 3.2.5, a DALI search (Holm, 2020) of Bd1081$^{S\text{-}LBD}$ identified McpX^{LBD} as having a similar structure (z-score = 27.5, good score threshold for Bd1081^{SLBD} > 23.7 (Holm et al., 2008)), indicating that quaternary ammonium compounds (QACs) could be a possible ligand for Bd1081^{LBD} and Bd1081^{SLBD}. McpX^{LBD} was shown to bind tightest to choline and glycine betaine (Webb et al., 2017). Both compounds contain a straight chain polycarbon backbone and a terminal nitrogen that is positively charged at physiological pH, similar to polyamines (Shrestha et al., 2018). However, choline was ultimately chosen over glycine betaine as a possible Bd1081^{LBD} /Bd1081$^{S\text{-}LBD}$ ligand as it carries an overall positive charge, and has an alcohol group instead of the carboxyl group zwitterion of glycine betaine. Therefore, choline along with readily available related molecule trimethylamine N-oxide (TMAO), were used in co-crystallisation trials. Both ligands were used at a concentration in 2-fold excess, resulting in trays being supplemented with 1.1 mM choline or TMAO. Initial hits were in BCS F4; 0.1 M Sodium chloride 0.1 M Sodium formate 0.1 M Bis-Tris propane 8.5, 25 % v/v PEG Smear

Medium, with Bd1081^{LBD} at 17 mg/ml supplemented with 1.1 mM TMAO and a 1:1 drop ratio. This condition did not require optimisation:

Method	Sitting drop vapour diffusion
Temperature (°C)	18
Plate Type	96 Well CombiClover 4 chamber crystallization plates
Protein Concentration (mg/ml)	17
Reservoir Solution	0.1 M Sodium chloride, 0.1 M Sodium formate, 0.1 M Bis-Tris propane 8.5, 25 % v/v PEG Smear Medium and 1.1 mM TMAO
Drop Volume (µl)	2
Drop Ratio	1:1
Volume of Reservoir (µl)	100
Crystallisation Morphology	3D Pear shape

Table 4.1 Crystallisation conditions for Bd1081^{LBD} supplemented with TMAO.

Crystals were soaked with a cryoprotectant containing 20% ethylene glycol and 1.1 mM TMAO to maintain ligand binding during flash-cooling. Diffraction data were obtained at Diamond Light Source with the following specifications:

Diffraction Source	Diamond Light Source		
Cryoprotectant	20% Ethylene Glycol		
Wavelength (Å)	0.8900		
Rotation of Crystal (°)	360		
Rotation per Image (°)	0.1		
Images Collected	3600		
Exposure per Image (sec)	0.002		
Transmission (%)	100		
Detector Resolution (Å)	2.0		
Space Group	P 1		
Cell Dimensions			
a, b, c (Å)	83.59	84.33	95.68
α, β, γ (°)	101.2	113.4	93.3
Resolution Range (Å)	2.63 – 85.35		
Wilson B-Factor	32.26		
	Overall	High	
Total Reflections	247631	12937	
Unique Reflections	67976	3434	
R_{merge}	0.137	2.260	
R_{meas}	0.193	3.197	
$R_{p.i.m}$	0.137	2.260	
$I/\sigma I$	3.8	0.3	
CC(1/2)	0.995	0.352	
Multiplicity	3.6	3.8	
Completeness	98.7	98.6	

Table 4.2 Diffraction dataset specifications for $Bd1081^{LBD}$ supplemented with TMAO

Once the diffraction data had been obtained (log file in Appendix VII) an initial search model was built from $Bd1081^{S\text{-}LBD}$ (previous chapter) using CCP4i2 Sculptor (Potterton et al., 2018). Molecular replacement was performed with the model using Phenix PHASER (McCoy et al., 2007), with search parameters set to find 8 copies based on Matthews coefficient results from CCP4i2 Estimate cell content (Potterton et al., 2018). Eight copies were placed with a final TFZ of 21.9 and a LLG score of 1555. Regions that had been removed by CCP4i2 Sculptor were then rebuilt by Phenix Autobuild

(Terwilliger et al., 2008) and manual building in COOT (Emsley et al., 2010). Final refinement used COOT and Phenix Refine (Afonine et al., 2012) (Table 4.3).

Resolution (Å)	2.4
Number of Copies in Asymmetric Unit	8
R_{Work}	0.1898
R_{Free}	0.2540
R.M.S Deviations	
Bond Length (Å)	0.01
Bond Angles (°)	1.03
Ramachandran Plot	
Favoured Regions(%)	92.35
Allowed Regions(%)	6.20
Outliers (%)	1.45

Table 4.3 Final refinement statistics for $Bd1081^{LBD}$ supplemented with TMAO.

The asymmetric subunit contained 8 copies of $Bd1081^{LBD}$, which like Bd2503 and Bd2504 (previous chapter) formed symmetrical dimers around the N-terminal transmembrane helix. All chains were built as a continuous, single polypeptide; chain A was built from residues 12-277, chain B 7-278, chain C 10-280, chain D 12-276, chain E 10-277, chain F 11-276, chain G 8-279, and chain H 9-275. However, two of the chains had relatively poor electron density for the proximal lobe, residues 243-249 of chain F and 241-248 of chain H, compared to the rest of the structure, similar to that observed for Bd2503 and Bd2504 (previous chapter). Unlike the structures of the previous dCache domain proteins, an area of poor density was also observed in the top lobe of chain D between residues 93 and 100. However, in all three cases there was enough continuous density to build these sections of the structure when a complete copy/chain was superimposed as a guide. Two monomers are held in a dimer by 9 hydrogen bonds, all of which form in the top half of the transmembrane helices, along with hydrophobic packing which occurs along the entire length of these helices (Figure 4.3). All of the hydrogen bonds formed between sidechains, with the exception of the bond between the sidechain of S44 and backbone of E58. Hydrogen-bonding between the monomers demonstrates divergence

from absolute symmetry, bonds between residues S44-E58, K47-E62, and E48-K66 are implicated in both chains, whilst residues Q55-Q55/N50 and N67-H40 form hydrogen bonds from one chain only.

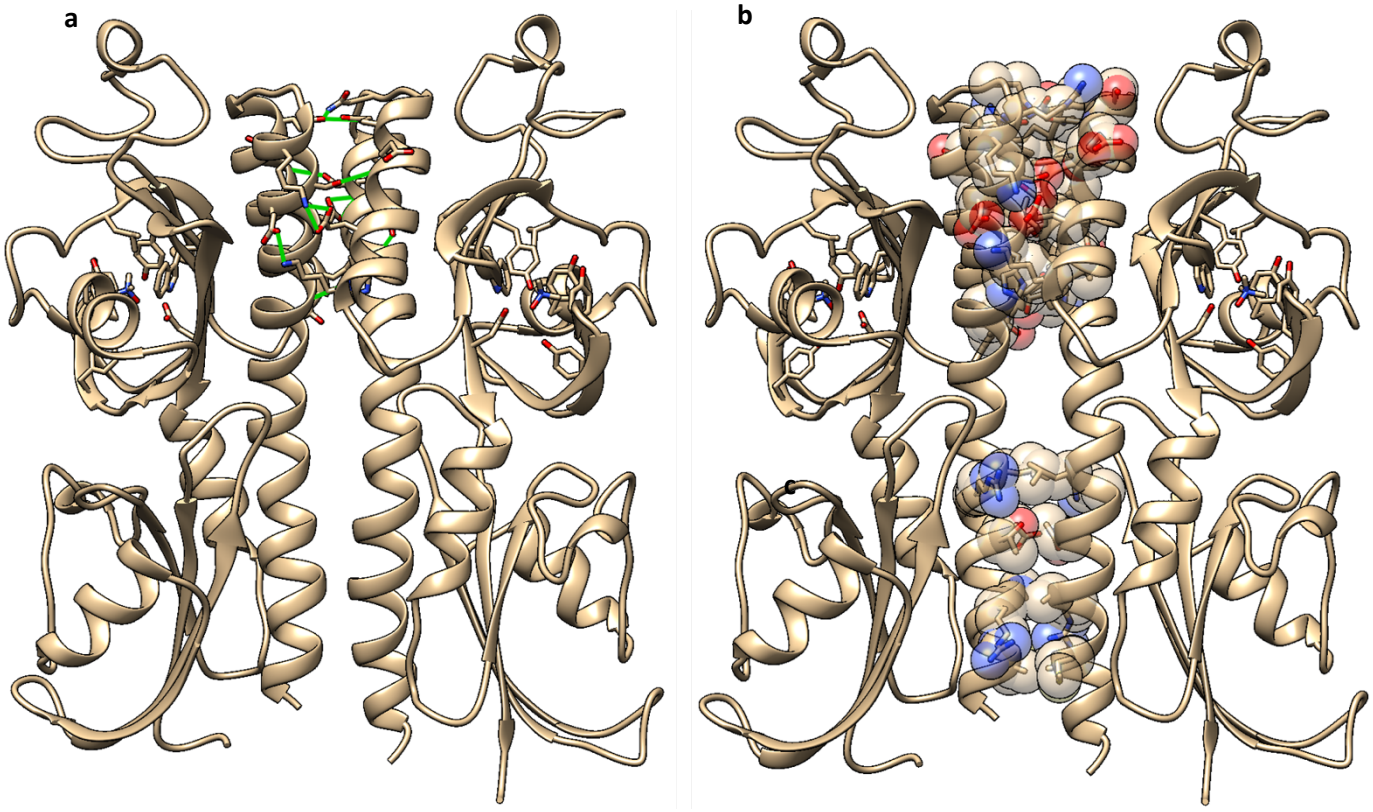


Figure 4.3 (a) Bd2504 dimer with hydrogen bonds formed to stabilise dimer interaction shown in green. These hydrogen bonds are limited to the transmembrane helix **(b)** Bd2503 dimer with hydrophobic packing between the two monomers displayed.

Having built Bd1081^{LBD} (Figure 4.4a) the refined map showed a region of strong density in the classical dCache binding pocket (Figure 4.4b). Although the density was not particularly well defined, it was modelled as TMAO as Bd1081^{LBD} had been supplemented with TMAO prior to crystallisation (Figure 4.5). On account of TMAO being a smaller molecule than the previously modelled cadaverine and putrescine, and the resolution of the Bd1081^{LBD} map being lower than Bd2503 and Bd2504, the orientation of TMAO could not be assigned as accurately as previous ligands. Instead of being able to orientate the ligand using both the density and pocket environment, the reduced resolution meant that TMAO had to be placed based solely on the properties of the pocket environment. Therefore, the exact orientation of TMAO within the pocket many differ slightly compared to the Bd1081^{LBD}:TMAO

model presented here. Binding of TMAO occurs in the top lobe, as is typical of dCache chemoreceptors. McpX of *Sinorhizobium meliloti* has been shown to bind quaternary ammonium compounds (QAC) reminiscent of TMAO. The structure of McpX bound to proline betaine showed the same binding pose within the LBP as that observed for $Bd1081^{LBD}$ and TMAO (Shrestha et al., 2018). Unlike Bd2503 and Bd2504, no hydrogen bonds were formed between the hydrogen bond donating rich LBP and TMAO. Instead, a combination of cation-π interactions and hydrophobic packing is relied upon for complex formation (Figure 4.6). Cation-π interactions form between positively charged groups or ions and aromatic groups within 6 Å of each other (Gallivan and Dougherty, 1999). Within the $Bd1081^{LBD}$ ligand pocket, these interactions form between the positively charged central ammonium group of TMAO and the aromatic cage amino acids (W74, Y104, Y128 and Y146) all of which are between 4.48–4.81 Å away from the ammonium group of TMAO. This mechanism of binding appears to be fairly universal for the binding of TMAO and occurs in both TorT, a periplasmic binding protein which interacts with MCP TorS in *Vibrio parahaemolyticus* (3O1H (Moore and Hendrickson, 2012)), and TmoX, a TMAO transporter in *Ruegeria pomeroyi* (4XZ6 (Li *et al.*, 2015)).

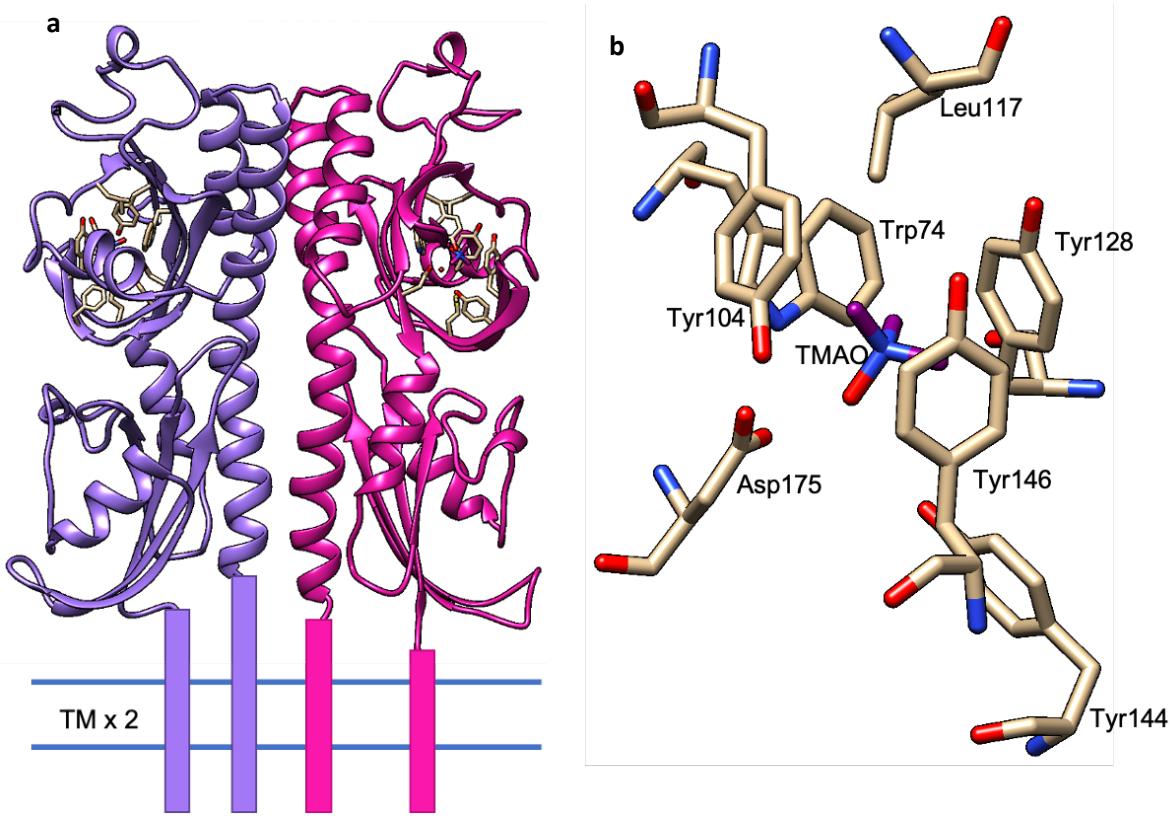


Figure 4.4 (a) Bd1081 dimerisation. **(b)** Close up of Bd1081 binding pocket with TMAO placed in hypothetical binding orientation (purple), with the binding pocket amino acids shown.

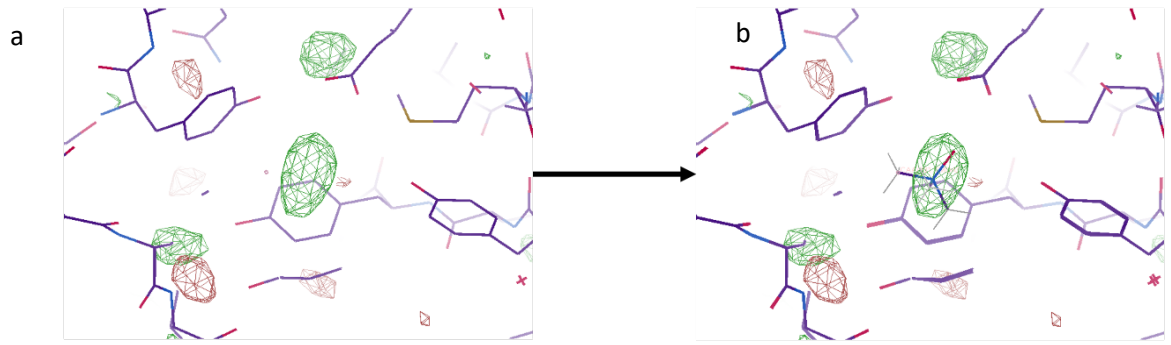


Figure 4.5 FOFCWT PHFOFCWT density map of the binding pocket with sigma set to 3.0. **(a)** High resolution, unaccounted for, density in the LBP. **(b)** TMAO fitted into this density.

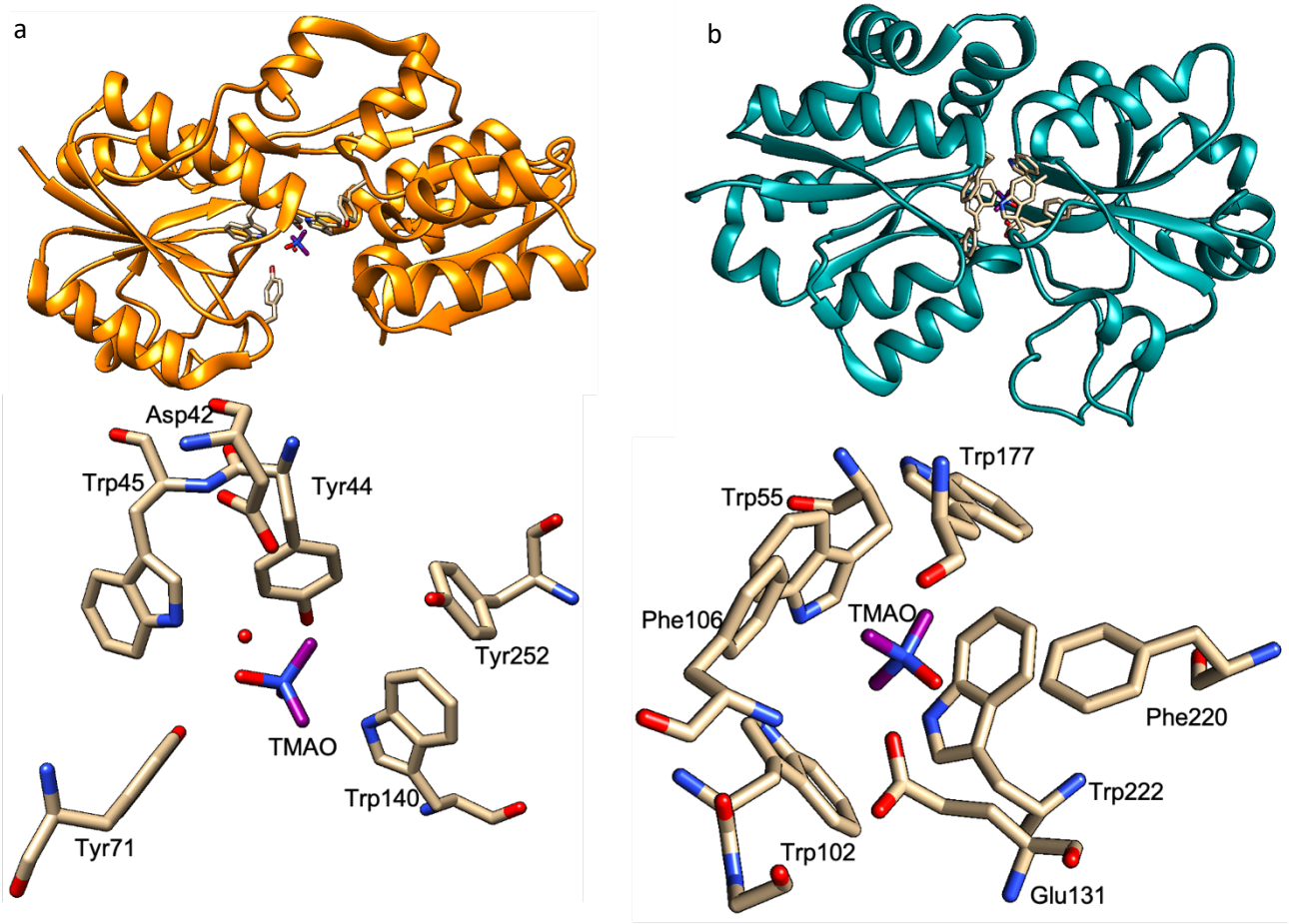


Figure 4.6 (a) Top – monomer of TorT, TMAO periplasmic binding protein. Bottom – Close up of binding pocket with TMAO showing that TorT relies on cation-π interactions to co-ordinate the binding of TMAO (3O1H (Moore and Hendrickson, 2012)). **(b)** Top – monomer of TmoX, TMAO periplasmic binding protein. Bottom – Close up of binding pocket with TMAO showing that TmoX relies on cation-π interactions to co-ordinate the binding of TMAO (4XZ6 (Li *et al.*, 2015)).

4.2.3 Isothermal Titration Calorimetry

To determine ligand specificity, isothermal titration calorimetry (ITC) using the apo form of each chemoreceptor LBD was performed. Standard experimental conditions were: temperature set to 25 °C, with a cell concentration of 80 μM and an injectant concentration of 600 μM, however optimisation of these condition was sometimes necessary to improve curve sigmoidicity in order to calculate the *Kd*. All data was baseline corrected using a reference dataset obtained by injecting ligand at the same concentration as in the experiment into a cell containing buffer only.

Given that dCache domain proteins have been demonstrated to bind a wider variety of ligands than sCache domain proteins (Ortega et al., 2017a), $Bd1081^{LBD}$, Bd2503 and Bd2504 were assayed with a larger range of ligands (Figure 4.7) (Table 4.4). On the basis of the crystal structures of Bd2503 and Bd2504, all of the dCache proteins were tested for specificity to polyamines, 1,3-diaminopropane (DAP) (3C), cadaverine (5C), norspermidine (6C), spermidine (7C) and putrescine (4C). Specificity for the presence of multiple amine/hydroxyl groups was also tested via diaminopimelic acid, ethanolamine, and pentanediol. As a result of the $Bd1081^{LBD}$ structure all dCache proteins were tested for specificity to QACs: glycine betaine, choline and TMAO. In order to establish if the occupancy of the $Bd1081^{S\text{-}LBD}$ LBP by Bis tris was indeed an artifact of crystallisation the binding between the *HD100* dCache receptors and Bis tris was also tested. Likewise, the structure of Bd2503 with arginine led to specificity tests of all dCache proteins with amino acids possessing amine containing sidechains, arginine, histidine and lysine, as well as an arginine derivative, argininosuccinate, that *Bdellovibrio* had been previously shown to be chemotactic towards (Straley et al., 1979).

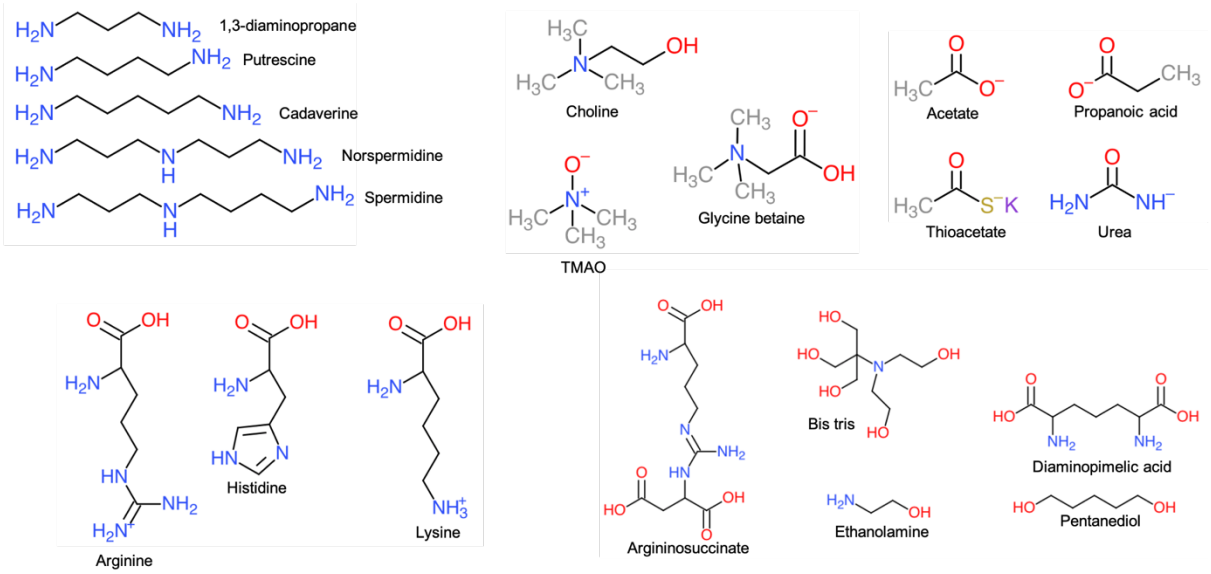


Figure 4.7 The chemical structure of all ligands used for ITC.

Ligand	Protein							
	Bd1081^{LBD}	Bd1081$^{S\text{-}LBD}$	Bd1081$^{S\text{-}LBD}$ D119G	Bd1081$^{S\text{-}LBD}$ V125E	Bd1081$^{S\text{-}LBD}$ Double	Bd2503^{LBD}	Bd2504^{LBD}	Bd3092^{LBD}
1,3-Diaminopropane						$3.11x10^{5} \pm 2.87x10^{4}$		
Acetate								$2.02x10^{4} \pm 9.73x10^{3}$
Arginine								
Argininiosuccinate								
Betaine								
Bis tris								
Cadaverine						$1.09x10^{6} \pm 1.78x10^{5}$		
Choline								
Diaminopimelic acid								
Ethanolamine								
Histidine								
Lysine								
Norspermidine								
Pentanediol								
Propanoic acid								$2.04x10^{5} \pm 8.69x10^{4}$
Putrescine								
Spermidine								
Thioacetate								
TMAO								
Urea								

Table 4.4 Summary of ITC experiments, green boxes reflect experiments that showed binding and include the Ka (M) obtained, red boxes are experiments that showed no significant binding, and grey boxes are experiments that were not ran.

4.2.3.1 Bd2503

Bd2503 was found to only bind the polyamines cadaverine and DAP. Despite similar chemical properties, no binding was observed between Bd2503 and putrescine (Figure 4.10), nor was significant binding detected between Bd2503 and arginine (Figure 4.11), supporting the conclusion that in spite of the crystal structure arginine is not a true ligand of Bd2503. ITC experiments with Bd2503 and cadaverine were performed under standard conditions and a curve was fitted to the ITC data using 'One Set of Sites' in Origin 7.0 (Edwards, 2002), with the *Kd* calculated to be ~1 μM (Figure 4.8). Testing of Bd2503 binding to DAP was performed under standard conditions with the exception of the temperature which was set to 15 °C, in order to produce exothermic peaks only, a curve was fit in the same manner as Bd2503 with DAP and a preliminary *Kd* estimated to be ~5 μM (Figure 4.9).

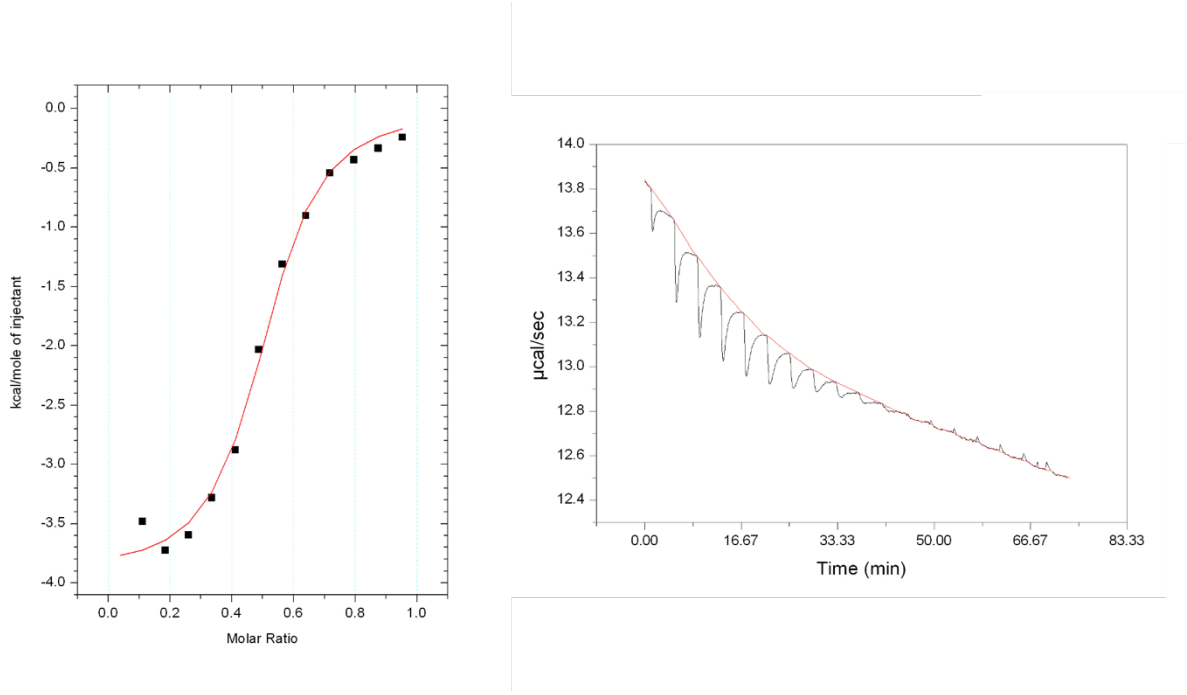


Figure 4.8 Binding of Bd2503 to cadaverine was performed at concentrations of 80 μM and 600 μM respectively, with 8 μL injections. Left is the normalised integrated data with curve fitted, right is the raw ITC titration data. The *Kd* was established to be ~0.92 μM.

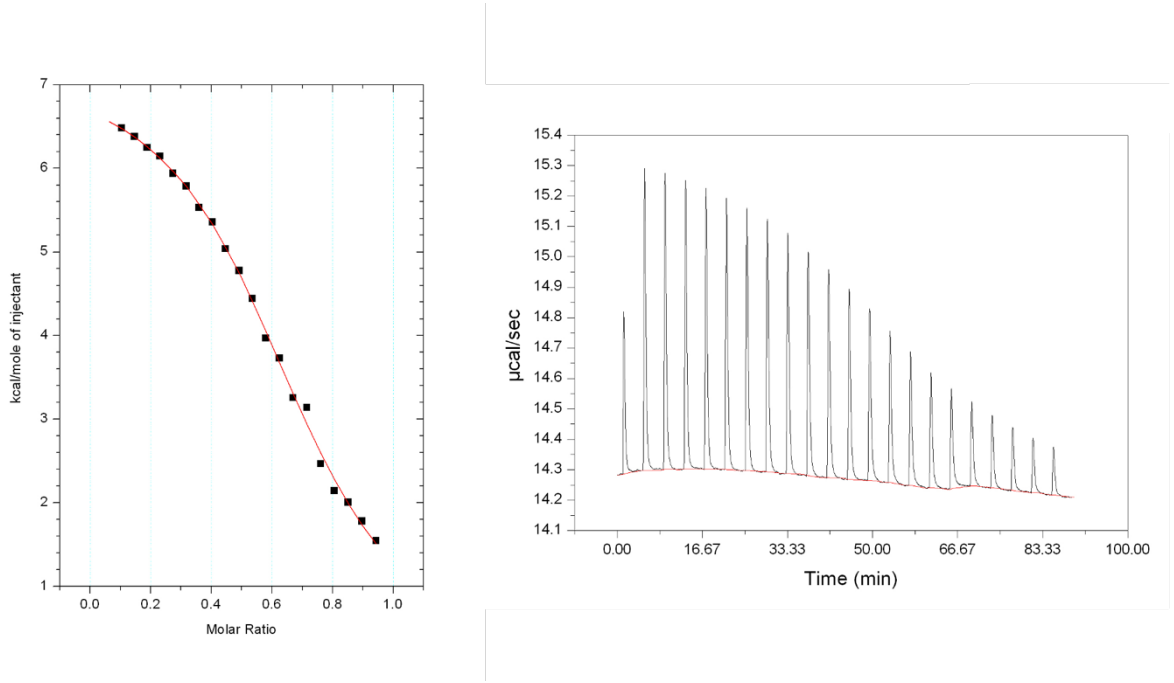


Figure 4.9 Binding of Bd2503 to DAP was performed at concentrations of 80 µM and 600 µM respectively, with 8 µL injections. Left is the normalised integrated data with curve fitted, right is the raw ITC titration data. A preliminary *Kd* was estimated to be ~5 µM.

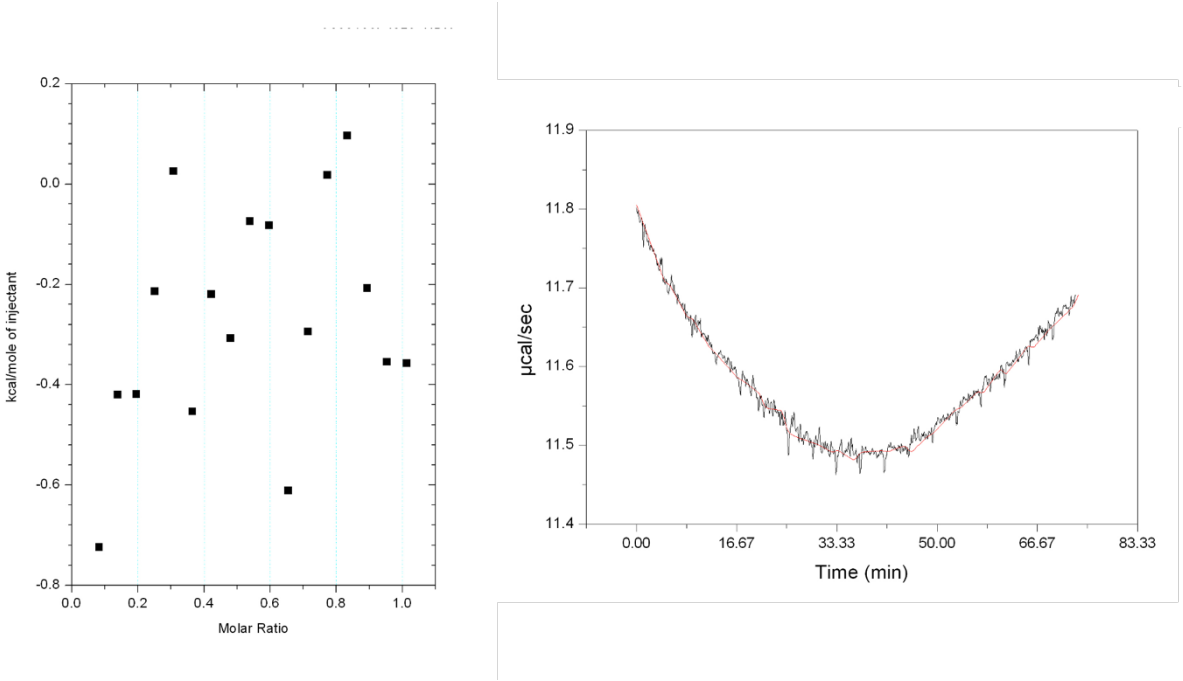


Figure 4.10 Preliminary ITC data of Bd2503 with putrescine performed at concentrations of 80 µM and 600 µM respectively with 8 µL injections, showed no binding at this temperature. Left is the normalised integrated data, right is the raw ITC titration data. Data is representative of all non-binding data for Bd2503.

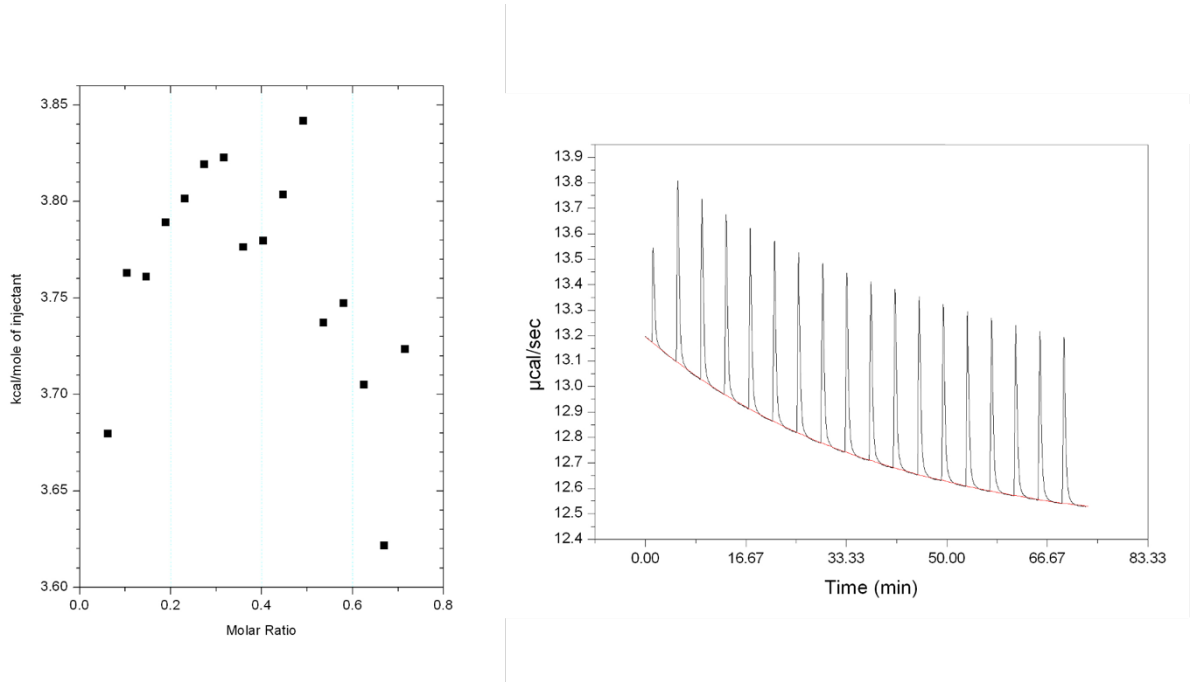


Figure 4.11 Preliminary ITC data of Bd2503 with arginine were performed at concentrations of 80 μM and 600 μM respectively, with 8 μL injections, left is the normalised integrated data, right is the raw ITC titration data.

4.2.3.2 Bd2504

Bd2504 was only shown to bind putrescine, and despite similar chemical properties no binding was observed with other polyamines (Figure 4.13 and Figure 4.14) or ethanolamine, thus lacking a similar pattern of C_n and C_{n+2} binding, as seen with the Bd2503 ability to bind both DAP and cadaverine. ITC experiments between Bd2504 and putrescine were performed with Bd2504 at 100 μM in the cell and putrescine at 1000 μM in the injection port, temperature kept at 25 °C as standard. A curve was fitted to the ITC data using 'One Set of Sites' in Origin 7.0 (Edwards, 2002), however the data did not meet the requirements necessary to determine a *Kd* value (Figure 4.12).

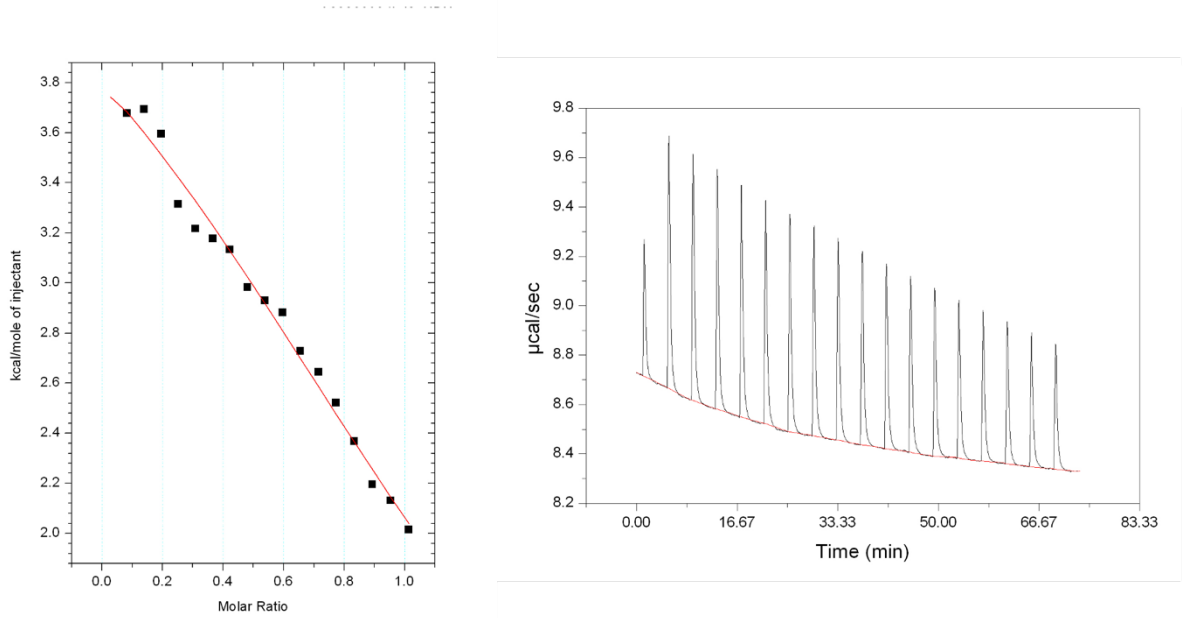


Figure 4.12 ITC of Bd2504 with putrescine were performed at concentrations of 100 µM and 1000 µM respectively, with 8 µL injections, left is the normalised integrated data, right is the raw ITC titration data. This preliminary data suggests binding between Bd2504 and putrescine.

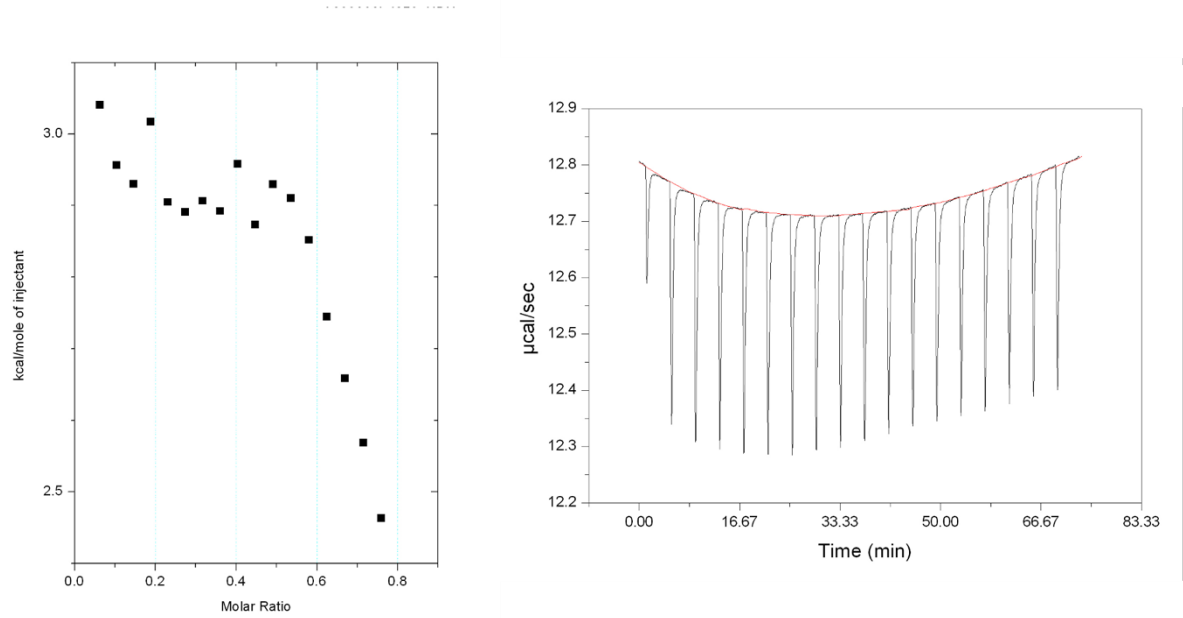


Figure 4.13 Preliminary ITC data of Bd2504 with cadaverine performed at concentrations of 80 µM and 600 µM respectively with 8 µL injections, left is the normalised integrated data with curve fitted, right is the raw ITC titration data.

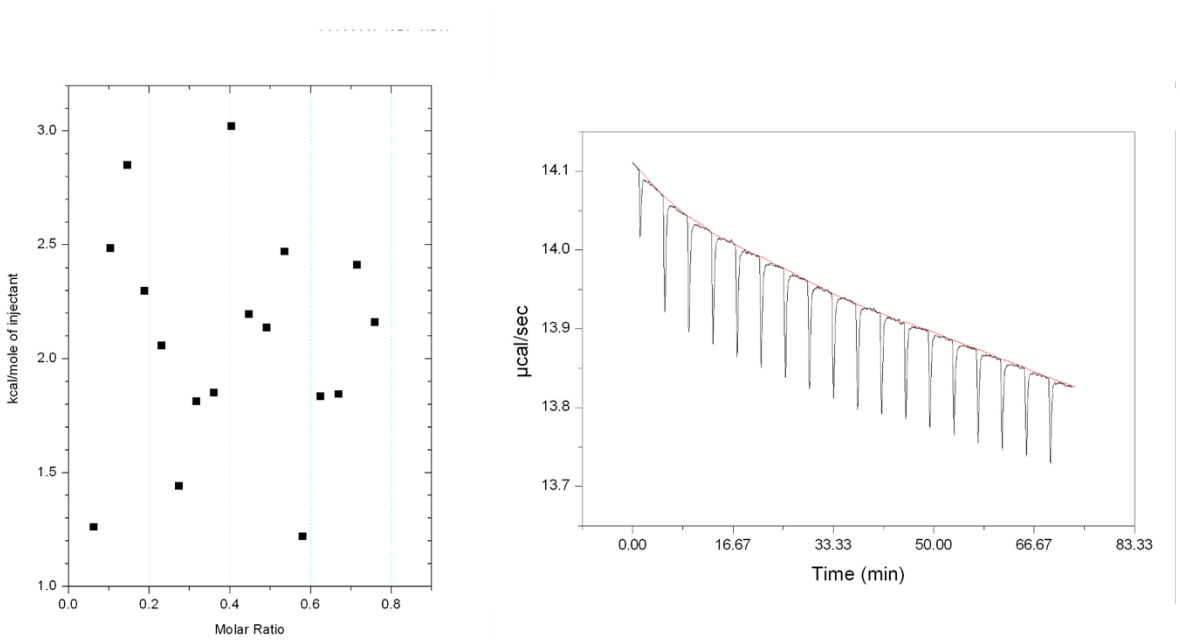


Figure 4.14 Preliminary ITC data of Bd2504 with DAP performed at concentrations of 80 µM and 600 µM respectively with 8 µL injections, left is the normalised integrated data with curve fitted, right is the raw ITC titration data. Data is representative of all non-binding data for Bd2504.

4.2.3.3 Bd1081^{LBD}

Preliminary ITC data only showed binding between Bd1081^{LBD} and one QAC, choline, as well as suggesting an affinity to DAP. No binding was detected between Bd1081^{LBD} and Bis tris (Figure 4.17), supporting the conclusion that a naturally occurring Bis tris-like compound is not the true ligand of Bd1081^{LBD} homologues, further suggesting the binding of Bis tris is an artefact of crystallography. All ITC experiments with Bd1081^{LBD} were performed at 15 °C. Experiments between Bd1081^{LBD} (at 80 µM) and choline (Figure 4.15) and DAP (Figure 4.16) were performed at 700 µM and 600 µM respectively. Curves were fitted to both data sets using 'One Set of Sites' in Origin 7.0 (Edwards, 2002), however the data did not met the requirements necessary to determine a *Kd* value.

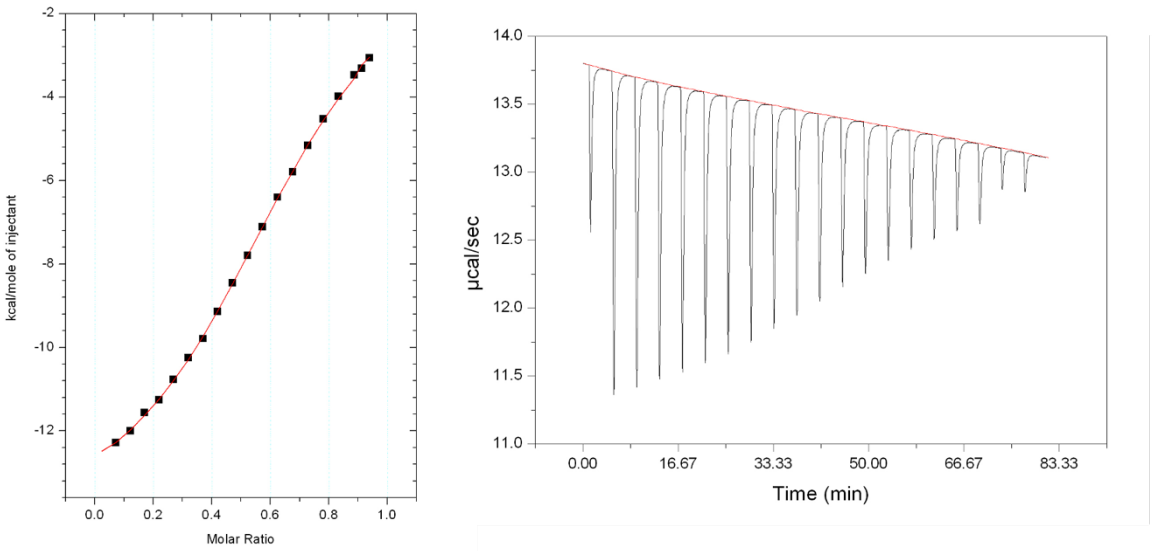


Figure 4.15 ITC of Bd1081 with choline were performed at concentrations of 80 µM and 700 µM respectively, with 8 µL injections, left is the normalised integrated data, right is the raw ITC titration data. This preliminary data suggests binding between Bd2504 and putrescine.

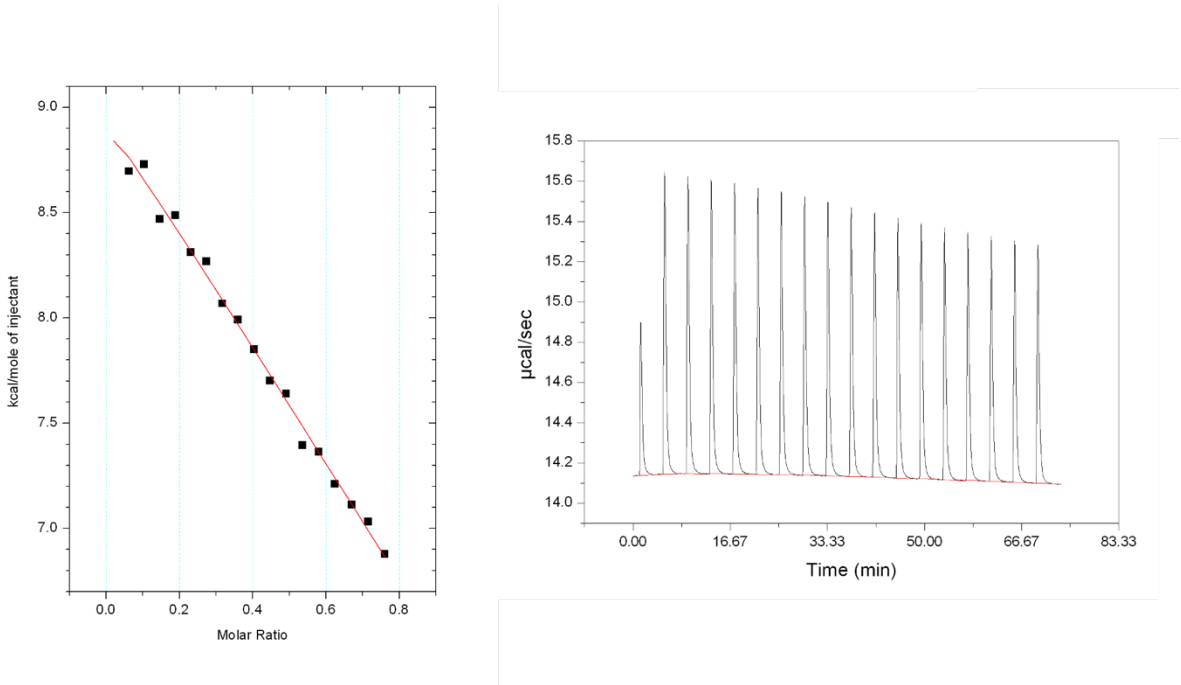


Figure 4.16 ITC of Bd1081 with DAP were performed at concentrations of 80 µM and 600 µM respectively, with 8 µL injections, left is the normalised integrated data, right is the raw ITC titration data. This preliminary data suggests binding between Bd1081 and DAP.

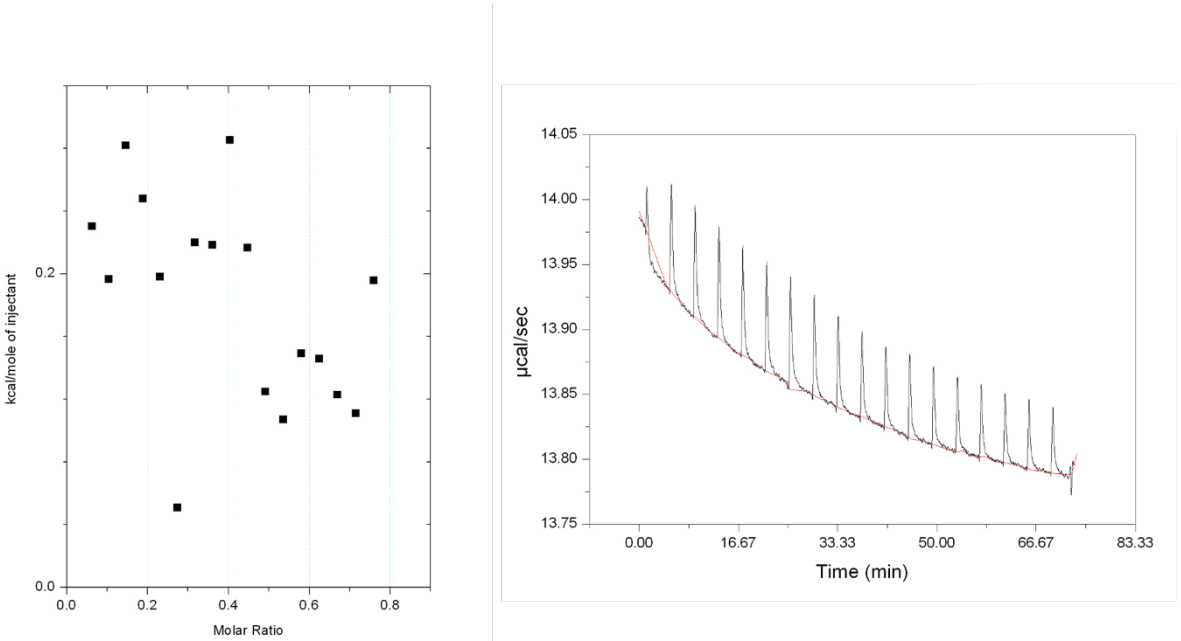


Figure 4.17 Preliminary ITC data of Bd1081^{LBD} with Bis tris performed at concentrations of 80 µM and 600 µM respectively, with 8 µL injections, left is the normalised integrated data with curve fitted, right is the raw ITC titration data. Data is representative of all non-binding data for Bd1081^{LBD}.

4.2.3.4 Bd1081$^{S\text{-}LBD}$

Limited testing of Bd1081$^{S\text{-}LBD}$ was undertaken, chiefly because of the high homology between Bd1081$^{S\text{-}LBD}$ and Bd1081^{LBD}. Having shown that Bd1081^{LBD} bound to both choline and DAP, the binding of Bd1081$^{S\text{-}LBD}$ to both compounds was tested. All experiments with Bd1081$^{S\text{-}LBD}$ were undertaken using standard experimental conditions. Preliminary data only showed binding between Bd1081$^{S\text{-}LBD}$ and choline, a curve was fit using 'One Set of Sites' in Origin 7.0 (Edwards, 2002) however a *Kd* could not be established (Figure 4.18). Since a discrepancy in binding affinity between Bd1081^{LBD} and Bd1081$^{S\text{-}LBD}$ was observed, Bd1081$^{S\text{-}LBD}$ binding was also tested against other QAC ligands, TMAO and glycine betaine, despite Bd1081^{LBD} demonstrating little affinity for these. Again, no significant binding affinity between Bd1081$^{S\text{-}LBD}$ and these QACs was detected.

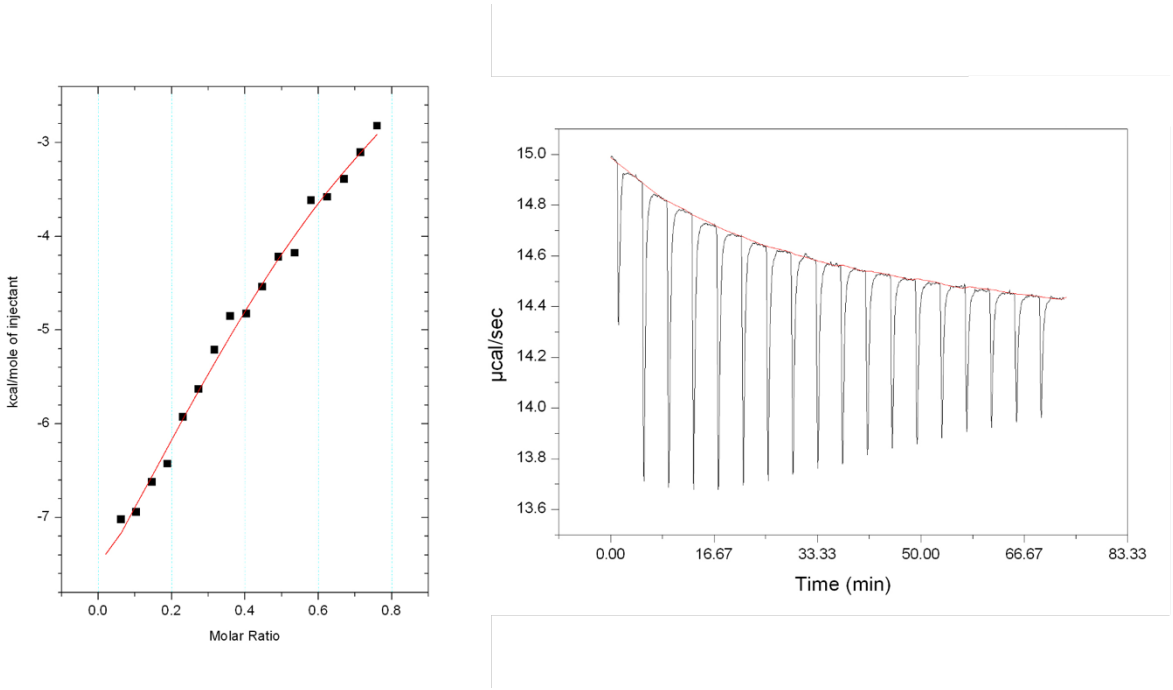


Figure 4.18 ITC of Bd1081^{S} with choline were performed at concentrations of 80 µM and 600 µM respectively, with 8 µL injections, left is the normalised integrated data, right is the raw ITC titration data. This preliminary data suggests binding between Bd1081^{S} and choline.

4.2.3.5 Bd3092

sCache domain proteins are known to bind C2/C3 organic acids and urea (Ortega et al., 2017a), therefore acetate, which Bd3092 had been crystallised with, along with propanoic acid and urea were used to test the specificity of Bd3092. Since Bd3092 had been crystallised with acetate, the sulphate derivative thioacetate, previously identified as a chemoattractant (Straley et al., 1979), was also used to further test ligand specificity. Bd3092 was found to bind propanoic acid and acetate, but not urea or thioacetate (Figure 4.21). ITC experiments between Bd3092 and propanoic acid were performed with a cell concentration of 150 µM and ligand concentration of 800 µM, at 25 °C (Figure 4.19). A curve was fitted to the ITC data using 'One Set of Sites' in Origin 7.0 (Edwards, 2002) and the *Kd* was estimated to be ~5 µM. Testing of Bd3092 binding to acetate was performed at 15 °C with Bd3092 at 80 µM and acetate at 800 µM, a curve was fit identically to Bd3092 with propanoic acid and a preliminary *Kd* was estimated to be ~50 µM (Figure 4.20). This revealed that although binding of acetate to the LBP was enough to stabilise the structure of Bd3092 for crystallisation, Bd3092 has a

stronger affinity for propanoic acid than acetate. Both propanoic acid and acetate have been previously identified as chemo-attractants (Straley et al., 1979).

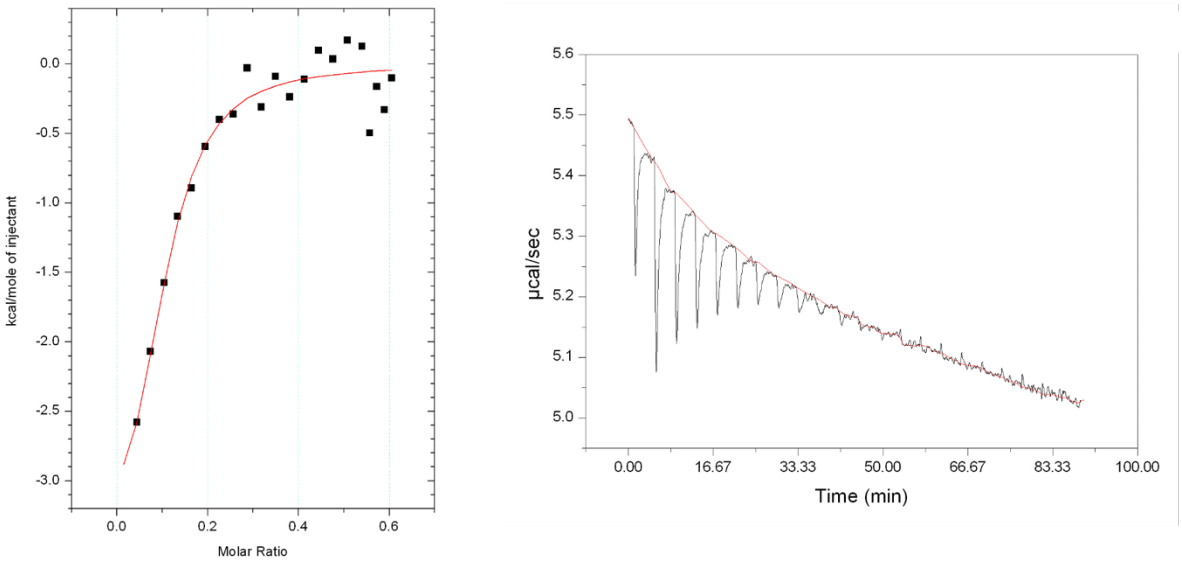


Figure 4.19 Binding of Bd3092 to propanoic acid was performed at concentrations of 150 µM and 800 µM respectively, with 8 µL injections. Left is the normalised integrated data with curve fitted, right is the raw ITC titration data. The *Kd* was established to be ~5 µM.

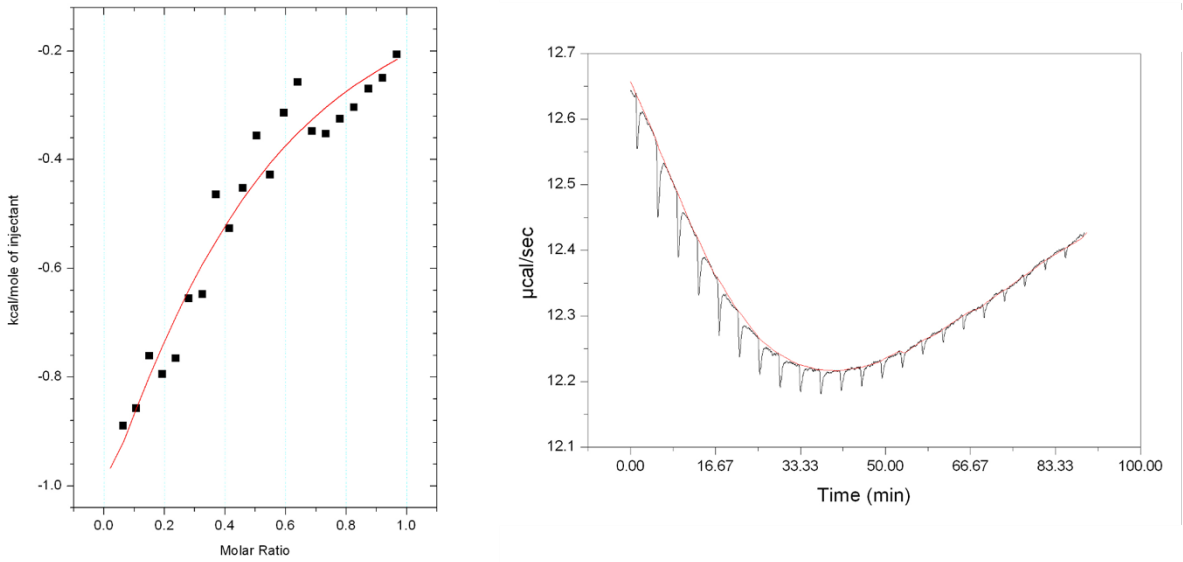


Figure 4.20 Binding of Bd3092 to acetate was performed at concentrations of 80 µM and 800 µM respectively, with 8 µL injections, left is the normalised integrated data with curve fitted, right is the raw ITC titration data. A preliminary *Kd* of ~50 µM was estimated.

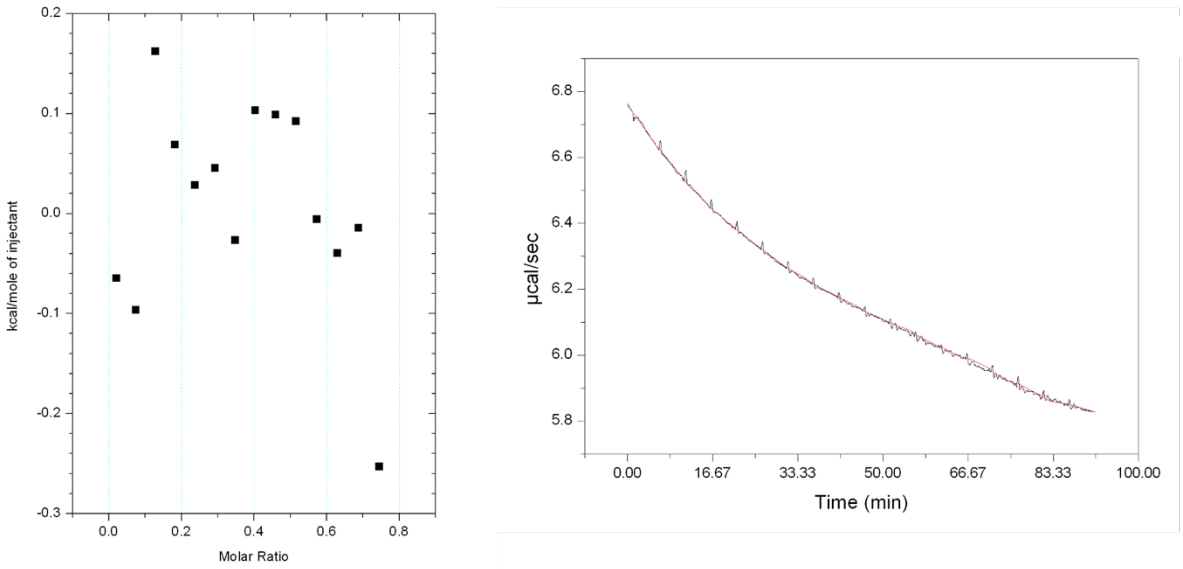


Figure 4.21 Preliminary ITC data of Bd3092 with thioacetate performed at concentrations of 80 µM and 600 µM respectively with 8 µL injections, left is the normalised integrated data with curve fitted, right is the raw ITC titration data. Data is representative of all non-binding data for Bd3092.

4.2.4 Identifying Amino Acids Involved in the Variation of Ligand Specificity Between Bd1081^{LBD} and Bd1081$^{S\text{-}LBD}$

4.2.4.1 Changes in Amino acid Sequence Between Bd1081^{LBD} and Bd1081$^{S\text{-}LBD}$ Ligand Binding Pocket

The ITC revealed a binding discrepancy between Bd1081^{LBD} and homologue Bd1081$^{S\text{-}LBD}$. This resulted in sequence (Figure 4.22) and structural alignments (Figure 4.23) of Bd1081^{LBD} and Bd1081$^{S\text{-}LBD}$ being performed in order to ascertain which amino acids may be responsible. Sequence alignment revealed 13 amino acids in the LBP region that were different between Bd1081^{LBD} and Bd1081$^{S\text{-}LBD}$. However, when these differences were examined within the aligned structure, only two, D120/G120 and V126/E126, were considered to be in positions that could affect ligand binding. Bd1081^{LBD} had been shown to bind choline with a higher affinity than Bd1081$^{S\text{-}LBD}$, therefore these positions in Bd1081$^{S\text{-}LBD}$ were mutated to resemble Bd1081^{LBD} in order to phenocopy Bd1081^{LBD}. Bd1081$^{S\text{-}LBD}$ D120G, Bd1081$^{S\text{-}}$

^{LBD} V126E, and a double mutant, Bd1081$^{S\text{-}LBD}$ D120G V126E (denoted Bd1081$^{S\text{-}LBD}$ Double) were all successfully cloned and verified via sequencing.

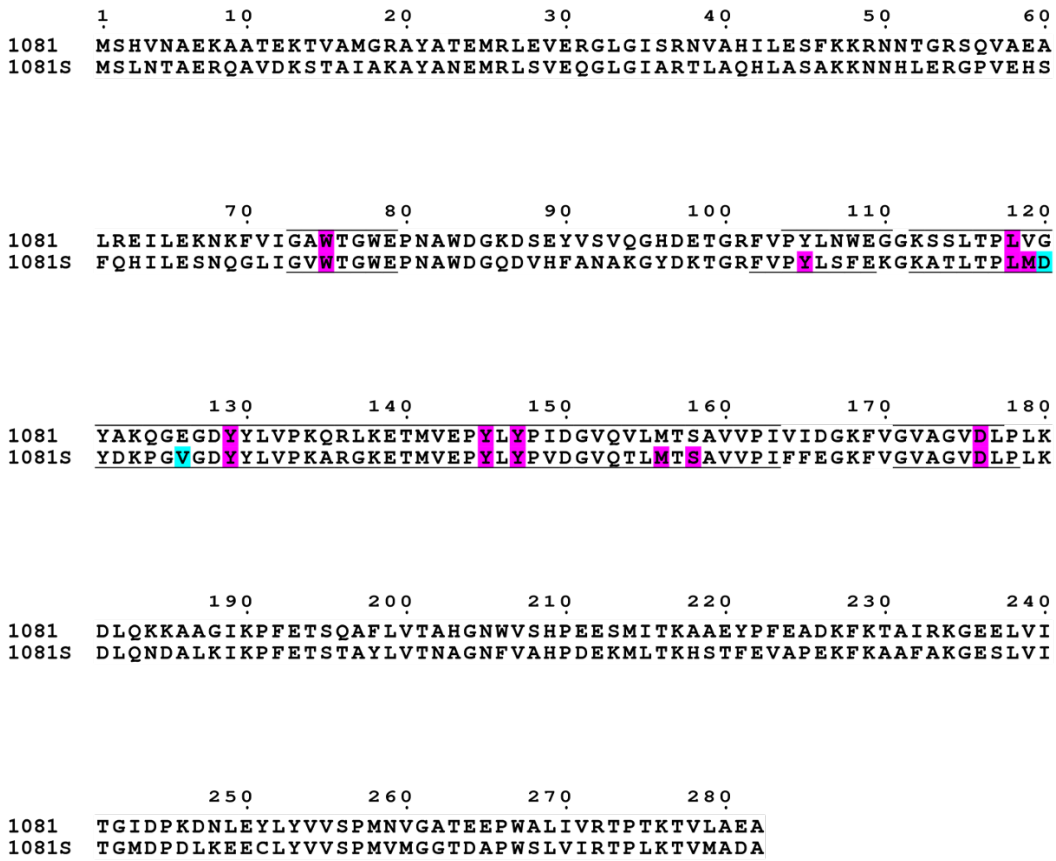


Figure 4.22 Alignment of Bd1081^{LBD} and Bd1081$^{S\text{-}LBD}$ (generated using ESPript 3.0 (Robert, X. and Gouet, P. (2014)). Amino acids in a black box make up the LBPs in Bd1081^{LBD} and Bd1081$^{S\text{-}LBD}$. Highlighted in pink are amino acid sidechains thought to be involved in ligand binding based on the structures of Bd1081^{LBD} and Bd1081$^{S\text{-}LBD}$, in complex with TMAO and Bis tris respectively. Highlighted in blue are amino acids chosen for mutation based on their position in the ligand binding domain, which although are not directly implicated in binding by the structures, may affect ligand binding specificity.

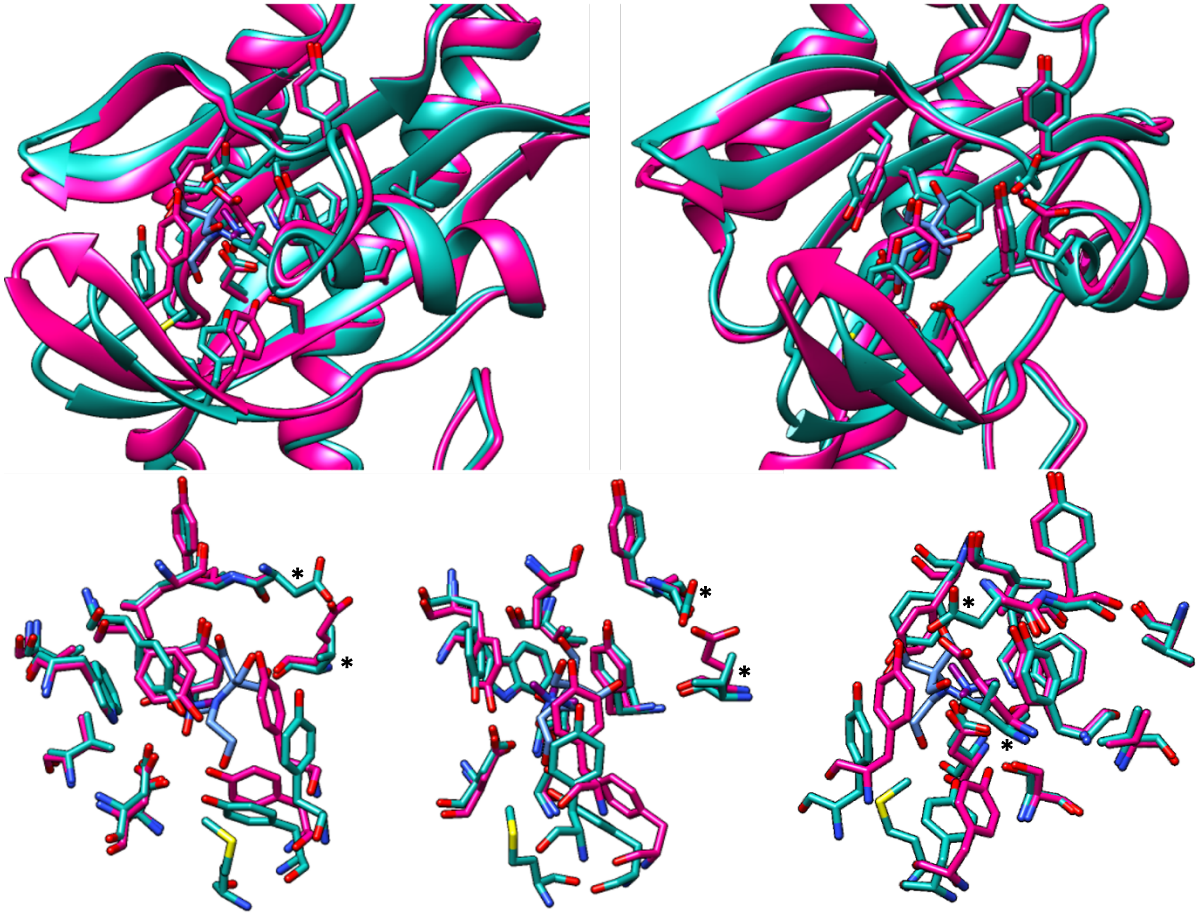


Figure 4.23 Close up of the LBPs of Bd1081^{LBD} and Bd1081$^{S\text{-}LBD}$ from structural alignment. Bd1081^{LBD} (pink) bound to TMAO (purple). Bd1081$^{S\text{-}LBD}$ (teal) bound to Bis tris (blue). The amino acids displayed are those thought to be involved in ligand binding and stabilisation within the LBP. The amino acids mutated in Bd1081$^{S\text{-}LBD}$ are denoted by a *.

4.2.4.2 Purification and Isothermal Titration Calorimetry of Bd1081$^{S\text{-}LBD}$ D120G

The LBD of Bd1081$^{S\text{-}LBD}$ D120G was expressed in AI media with sugar mix N and purified using nickel affinity chromatography, eluted with 100 mM Imidazole (Figure 4.24a). The purified protein was dialysed into 20 mM Hepes pH 7.5 and 300 mM NaCl, before SEC with a HiLoad 26/60 Superdex 75 pg (Figure 4.24c). SEC yielded one peak at 160 ml. ITC experiments with both choline and glycine betaine showed no binding.

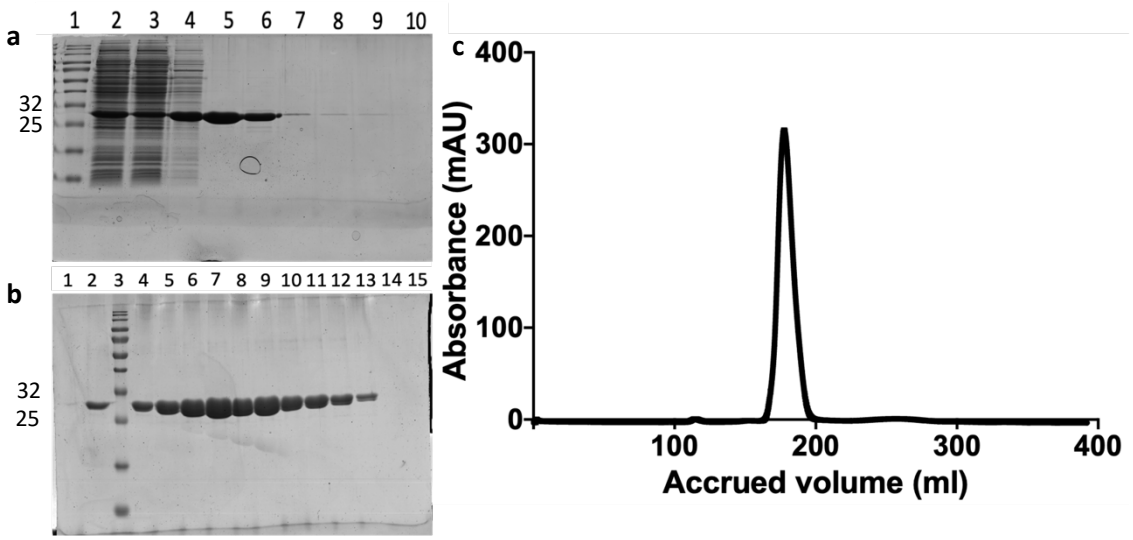


Figure 4.24 (a) SDS-PAGE gel of Bd1081$^{S\text{-}LBD}$ D120G IMAC purification. **(b)** SDS-PAGE gel of Bd1081$^{S\text{-}LBD}$ D120G post SEC purification. **(c)** UV trace of SEC purification showing Bd1081$^{S\text{-}LBD}$ D120G elutes as a single monomeric peak.

4.2.4.3 Purification and Isothermal Titration Calorimetry of Bd1081$^{S\text{-}LBD}$ V126E

The LBD of Bd1081$^{S\text{-}LBD}$ V126E was expressed in AI media with sugar mix N and purified using nickel affinity chromatography, eluted with 100 mM Imidazole (Figure 4.25a). The purified protein was dialysed into 20 mM Hepes pH 7.5 and 300 mM NaCl, before SEC with a HiLoad 26/60 Superdex 75 pg (Figure 4.25c). SEC yielded one peak at 160 ml. ITC experiments with both choline and glycine betaine showed no binding.

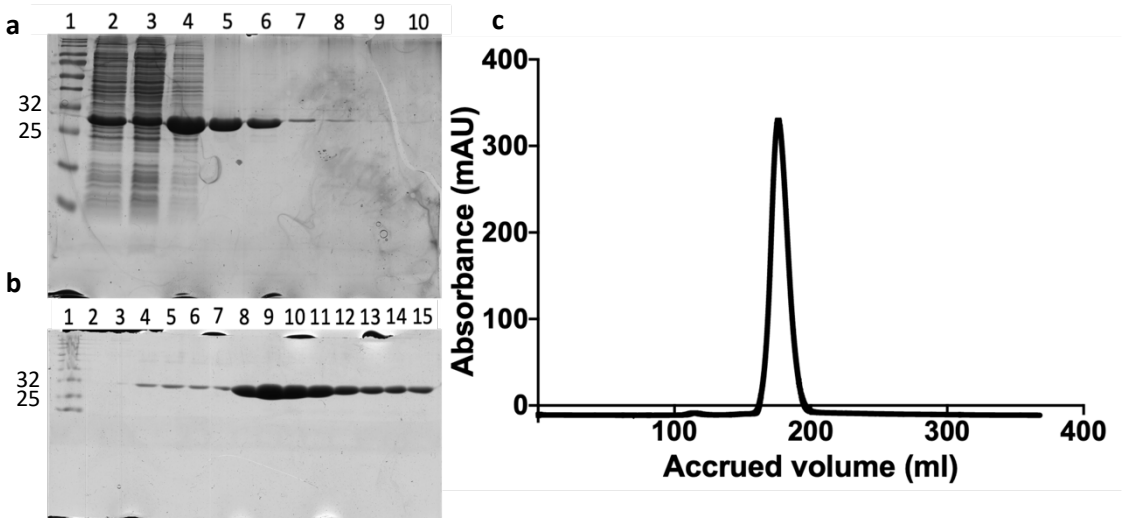


Figure 4.25 (a) SDS-PAGE gel of Bd1081$^{S\text{-}LBD}$ V126E IMAC purification. **(b)** SDS-PAGE gel of Bd1081$^{S\text{-}LBD}$ V126E post SEC purification. **(c)** UV trace of SEC purification showing Bd1081$^{S\text{-}LBD}$ V125E elutes as a single monomeric peak.

4.2.4.4 Purification and Isothermal Titration Calorimetry of Bd1081$^{S\text{-}LBD}$ Double

The LBD of Bd1081$^{S\text{-}LBD}$ Double was expressed in AI media with sugar mix N and purified using nickel affinity chromatography, eluted with 100 mM Imidazole (Figure 4.26a). The purified protein was dialysed into 20 mM Hepes pH 7.5 and 300 mM NaCl, before SEC with a HiLoad 26/60 Superdex 75 pg (Figure 4.26c). SEC yielded one peak at 160 ml. ITC experiments with both choline and glycine betaine showed no binding.

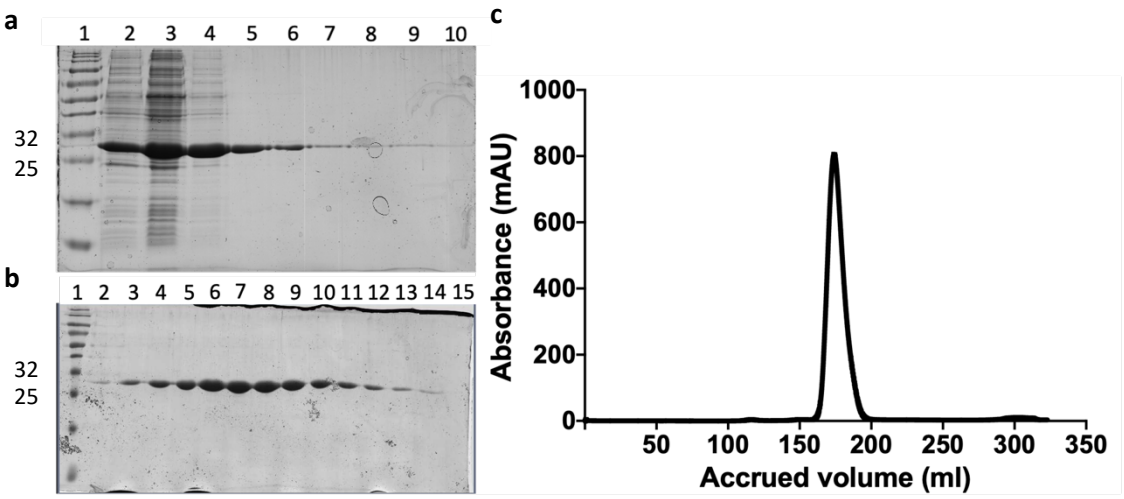


Figure 4.26 (a) SDS-PAGE gel of Bd1081$^{S\text{-}LBD}$ Double IMAC purification. **(b)** SDS-PAGE gel of Bd1081$^{S\text{-}LBD}$ Double post SEC purification. **(c)** UV trace of SEC purification showing Bd1081^{S} Double elutes as a single monomeric peak.

4.2.5 Bd3126^{LBD} – A dCache Histidine Kinase

In contrast to the other dCache domain proteins in *Bdellovibrio*, Bd3126 is a histidine kinase rather than an MCP, and prior to this work the fold of the periplasmic LBD had not been identified. A BlastP search (Altschul et al., 1990) did not reveal any homologues outside of *Bdellovibrio* and those identified did not have an annotated LBD. However, a HHpred search (Zimmermann et al., 2018), identified exclusively dCache-containing proteins as the top hits against the LBD of Bd3126 (Table 4.5). As a result of this, Bd3126 was discerned as containing a putative dCache fold as its LBD.

PDB Code	Organism	Protein	Ligand	HHpred Identity (%)
3LIF	*Rhodopseudomonas palustris*	HK1S-Z16	MPD	11
6IOU	*Vibrio cholerae*	Mlp24	Serine	12
6MNI	*Pseudomonas syringae*	PscC	Proline	8
4WY9	*Campylobacter jejuni*	Tlp1	-	13
6E09	*Helicobacter pylori*	TlpA	-	10
5LTV	*Pseudomonas aeruginosa*	PctC	GABA	13
3LI9	*Methanosarcina mazei*	HK1S-Z2	Bis tris	10
3LIC	*Shewanella oneidensis*	HK1S-Z6	1,2-ethanediol	12
4XMR	*Campylobacter jejuni*	Tlp3	Isoleucine	11
5LTX	*Pseudomonas aeruginosa*	PctA	Methionine	12

Table 4.5 Table of HHpred results for Bd3126 ligand binding domain, all of the identified proteins have been shown to contain a dCache domain via x-ray crystallography.

The LBD of Bd3126 (Bd3126^{LBD}) was expressed in AI media with sugar mix N and purified using nickel affinity chromatography, eluted with 100 mM Imidazole (Figure 4.27a). The purified protein was dialysed into 20 mM Hepes pH 7.5, 300 mM NaCl, 20 mM Imidazole and 50 µl TEV at 10 mg/ml for cleavage of the His-tag overnight. Bd3126^{LBD} was then repurified using IMAC, this time eluting in the flow through and the wash step with 20 mM imidazole (Figure 4.27b). The eluted fractions were dialysed into 20 mM Hepes pH 7.5 and 300 mM NaCl before SEC with a HiLoad 26/60 Superdex 75 pg (Figure 4.27c). SEC yielded one peak at 190 ml.

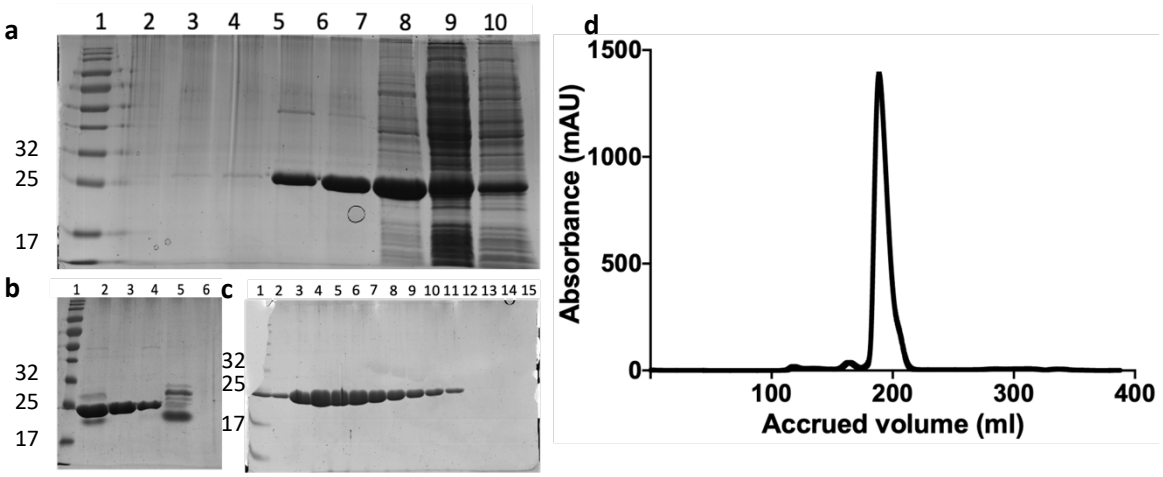


Figure 4.27 (a) SDS-PAGE gel of Bd3126^{LBD} IMAC purification. **(b)** SDS-PAGE gel of Bd3126^{LBD} post overnight cleavage of the His-tag using TEV and repurifying using IMAC. **(c)** SDS-PAGE gel of Bd3126^{LBD} post SEC purification. **(d)** UV trace of SEC purification showing Bd3126^{LBD} elutes as a single monomeric peak.

Initial attempts to crystalise Bd3126^{LBD} were unsuccessful. However, because Bd3126^{LBD} is predicted to have a dCache domain, and the SEC trace indicates a monomeric species, it was assumed crystallisation had not occurred because it was in the apo form, as the same had been true for Bd1081^{LBD}. Comparision between Bd3126^{LBD} and the other dCache domain proteins in *Bdellovibrio*, showed Bd3126^{LBD} had a higher sequence homology to Bd1081$^{S\text{-}LBD}$ (19.6%) and Bd2503^{LBD} (19.2%), than Bd1081^{LBD} (16.7%) and Bd2504^{LBD} (13%). When compared to dCache proteins outside *Bdellovibrio* the two proteins with the highest sequence homology are McpU^{LBD} from *Pseudomonas putida* at 20.7% which binds polyamines (Gavira *et al* 2018), and Mcp_N^{LBD} from *Vibrio cholerae* at 17.2% which binds amino acids (Patskovsky *et al* 2008). This resulted in crystal trays of Bd3126^{LBD} being set up with cadaverine, amino acids (L-glutamic acid and D-/L-alanine), and QACs (TMAO and choline). Crystals only formed in trays supplemented with QACs. Crystals supplemented with TMAO formed in Proplex E10; 0.2 M ammonium sulphate, 0.1 M Tris pH 8.5, 12% w/v PEG 8000, and JCSG^{+} A8; 0.2 M Ammonium formate and 20% w/v PEG 3350. Crystals supplemented with choline formed only in Proplex E10. In all conditions Bd3126^{LBD} was at 20 mg/ml and TMAO or choline was added at a 2x concentration of 1.6 mM, with a drop ratio of 1:1. These initial crystals possessed good physical characteristics (Figure 4.28) and were not optimised further:

Crystal	Proplex E10 with TMAO	JCSG$^+$ A8 with TMAO	Proplex E10 with Choline
Method	Sitting drop vapour diffusion	Sitting drop vapour diffusion	Sitting drop vapour diffusion
Temperature (°C)	18	18	18
Plate Type	96 Well CombiClover 4 chamber crystallization plates	96 Well CombiClover 4 chamber crystallization plates	96 Well CombiClover 4 chamber crystallization plates
Protein Concentration (mg/ml)	20	20	20
Reservoir Solution	0.2 M ammonium sulphate, 0.1 M Tris pH 8.5, 12% w/v PEG 8000 and 1.6 mM TMAO	0.2 M Ammonium formate, 20% w/v PEG 3350 and 1.6 mM TMAO	0.2 M ammonium sulphate, 0.1 M Tris pH 8.5, 12% w/v PEG 8000 and 1.6 mM choline
Drop Volume (µl)	2	2	2
Drop Ratio	1:1	1:1	1:1
Volume of Reservoir (µl)	100	100	100
Crystallisation Morphology	3D Pear shape	3D Pear shape	3D Pear shape

Table 4.6 Crystallisation conditions for Bd3126^{LBD}

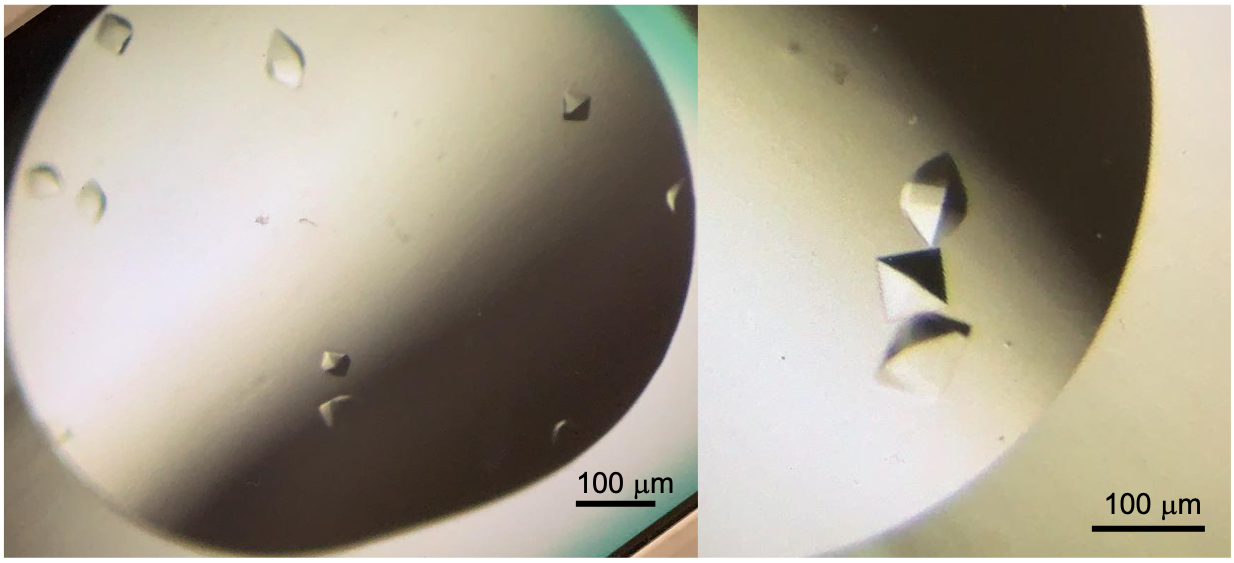


Figure 4.28 (a) Examples of Bd3126^{LBD} crystals obtained from Proplex E10 with choline (left) and TMAO (right)

Crystals were soaked with a cryoprotectant containing 20% ethylene glycol, either 1.6 mM TMAO or choline (to maintain ligand binding during flash-cooling), and 10 mM Potassium tetrachloroplatinate (II) for heavy atom derivatisation. Diffraction data were obtained at Diamond Light Source for all three crystal conditions, but because of inherent crystallographic issues ($P4_12_12$ with tNCS that approximates $P4_22_12$) the data is currently unsolved. Initial preliminary ITC was performed at 25 °C with Bd3126 at a concentration of 80 μM and ligands at 600 μM with 8 μL injections. As with the other ITC data discussed in this thesis, the data was corrected using a reference dataset where the cell contained buffer only. This data suggested that Bd3126 may bind choline, but not TMAO or glycine betaine (Figure 4.29).

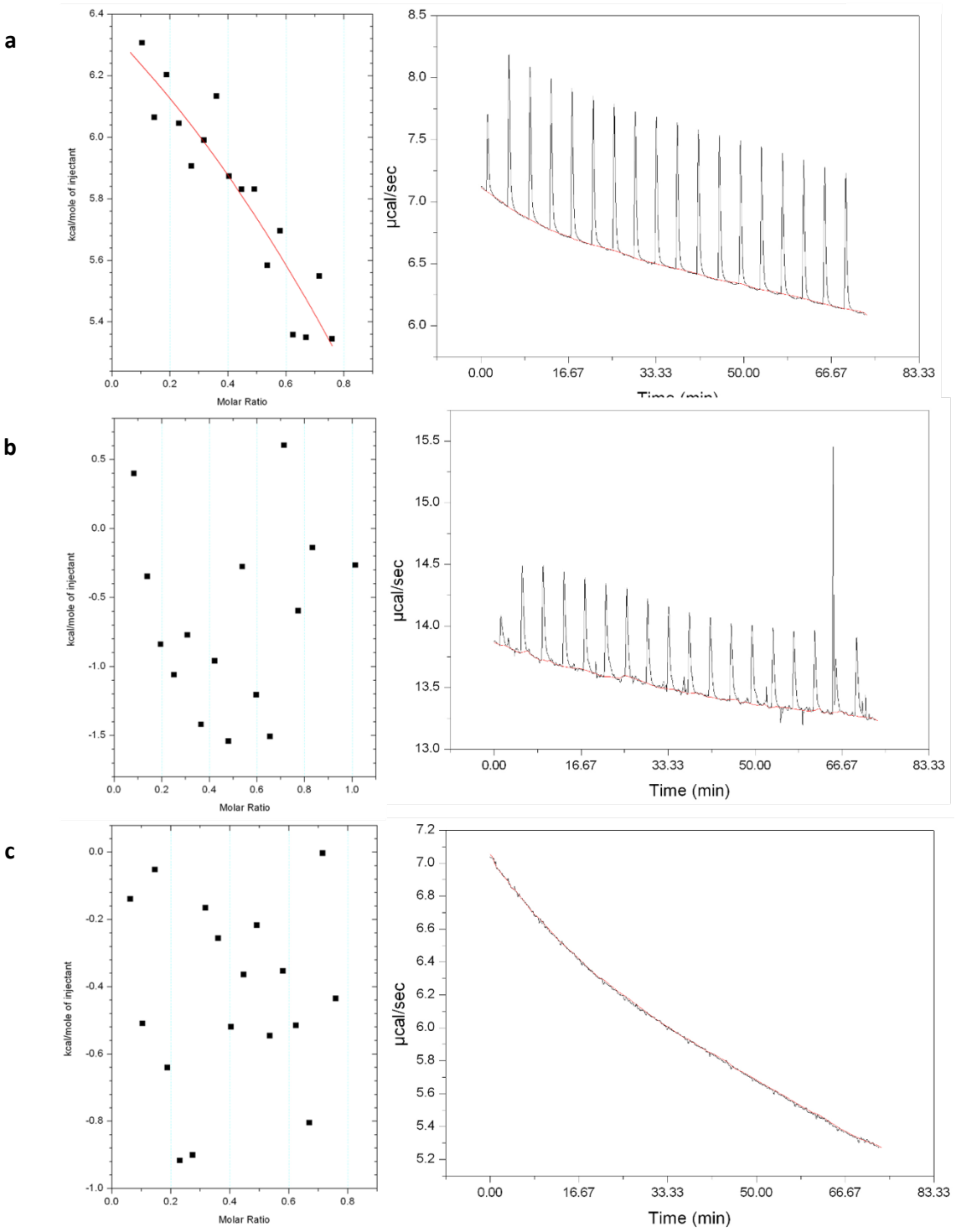


Figure 4.29 Preliminary ITC data of testing the binding of Bd3126^{LBD} to 3 quaternary ammonium compounds **(a)** choline **(b)** TMAO **(c)** glycine betaine, left is the normalised integrated data and right is the raw ITC titration data. This preliminary data suggests that Bd3126^{LBD} may bind choline, but does not bind TMAO or glycine betaine.

4.3 Discussion

4.3.1 Quaternary Amine Compounds Precedent

Unlike polyamines, there is no previously identified link between *Bdellovibrio* and quaternary ammonium compounds (QACs), as either a chemotactic ligand or as a small molecule used by *Bdellovibrio*. In contrast to polyamines, QACs are branched compounds and the common feature of the chemical family is a central nitrogen atom. This nitrogen forms an 'ammonium ion' where all of the hydrogens have been replaced by organic groups, e.g. methyl- and ethyl- groups (Bureš, 2019). This results in a permanent cationic charge similar to an ammonium ion, which gave rise to the family name of quaternary ammonium compounds (Webb et al., 2017). Many large complex QACs have been developed as antimicrobial surfactants, where the naturally occurring family is more defined and are generally smaller and significantly simpler (Figure 4.30) (Bureš, 2019). Despite no previous link to *Bdellovibrio* many bacteria naturally utilise QACs (Kiene, 1998).

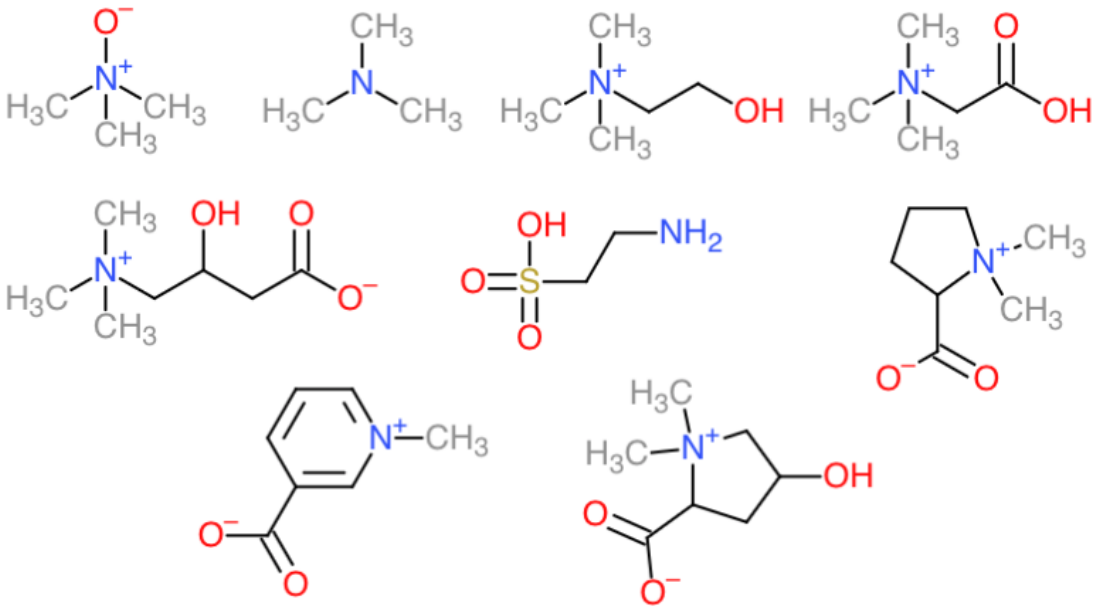


Figure 4.30 The chemical structure of some naturally occurring quaternary ammonium compounds and related molecules (top to bottom, left to right), TMAO, TMA, Choline, Glycine betaine, Carnitine, Taurine, Proline betaine, Trigonelline, Betonicine.

QACs are also associated with protein stabilisation in response to osmotic and temperature stress (Shrestha et al., 2018). Most bacteria use the following as osmoprotectants (those in bold are QACs): **Carnitine**, **Choline-O-sulphate**, Dimethylsulfoniopropionate, Ectoine, **Glycine betaine**, Mannitol, N-acetyl glutaminyl glutamine amide, Proline, **Stachydrine**, Taurine, Trihalose. Although many simple QACs have an essential role within bacteria, most do not have a method of producing them *de novo*, and the QAC glycine betaine is in low concentrations environmentally (Wargo, 2013). However, the glycine betaine precursor choline (Kiene, 1998) is in much higher concentrations environmentally, and most bacteria contain at least one pathway for metabolising choline into glycine betaine. It is thought that one of the major roles of choline in bacteria is as a source of glycine betaine (Wargo, 2013). Glycine betaine has been shown to structurally and functionally stabilise proteins, and it is via this action that glycine betaine is thought to act as a protectant against both osmotic and temperature pressures (Sakamoto and Murata, 2001). *Pseudomonas* and *Burkholderia* have both been shown to store small amounts of choline and glycine betaine, possibly for use in future stress responses (Wargo, 2013).

Despite having a major role as a glycine betaine precursor in many bacteria, choline is also used in the production of phosphatidylcholine (PC) (Mora-Ortiz and Claus, 2017). Although more commonly associated with eukaryotic membranes, 15% of organisms belonging to the domain *Bacteria* synthesize and use PC within their membranes. The percentage of total membrane made up of PC varies from 0-4% in *Pseudomonas aeruginosa* to 73% in *Acetobacter aceti* (Geiger et al., 2013), both of which are Gram-negative and capable of being targeted by *Bdellovibrio*. Many bacteria are also capable of using choline as a sole carbon source, which has led to the secondary theory that bacteria store choline and glycine betaine as innocuous nitrogen and carbon sources to be used during starvation stress (Wargo, 2013).

Gram-positive bacteria use teichoic acid and lipoteichoic acid in the production of their cell walls, (Brown et al., 2013). *Streptococcus pneumoniae*, uses choline in a niche way, attaching it to both teichoic and lipoteichoic acid to modify its cell wall (Baur *et al* 2009). This allows choline anchored virulence proteins to attach to the cell wall. In times of choline stress the production of polyamine transporter PotD is increased. This is thought to be because putrescine can be incorporated into the cell walls instead, however this prevents these virulence proteins from associating with the surface and is thought to attenuate virulence (Shah and Swiatlo, 2008).

A final use for choline is as a precursor for TMAO, via CutA producing TMA, which is then oxidised by TMA monooxygenase to TMAO (Fennema et al., 2016; Mora-Ortiz and Claus, 2017). Alternatively, TMA can be produced via another QAC, carnitine being metabolised by CntA (Jameson et al., 2018). TMAO can also act as an osmoprotectant, functioning to stabilise proteins by counteracting the denaturing effects of small molecules like urea (Li et al., 2015). TMAO is also imported, and is used as a terminal electron acceptor via TMAO reductase in the electron transport chain during anaerobic growth (Kracke et al., 2015).

Despite the ubiquitous and essential nature of QACs, only one QAC chemoreceptor has so far been identified (Webb et al., 2017). McpX is one of eight chemoreceptors identified in *Sinorhizobium meliloti* (Shrestha et al., 2018). It has been shown to contain a dCache LBD which is used to bind a large range of naturally occurring QACs; choline (138 nm), glycine betaine (1.3 μM), proline betaine (3.8 μM), trigonelline (88.5 μM), betonicine (2.3 mM), as well as the amino acid proline (45.2 μM) (Webb et al., 2017; Shrestha et al., 2018). *S. meliloti* forms bacterioids in the nodules of *Leguminosae* plants. The seeds of these plants exude a variety of chemicals during germination, including QACs, and it has been shown that *S. meliloti* use these signals to locate viable host plants. As well as enabling the location of viable hosts, QACs play at least two other important roles for *S. meliloti*, which can use

proline betaine as a nutrient source and also use choline in the production of PC for use in their cell membrane (Geiger et al., 2013; Shrestha et al., 2018).

Although McpX is the only other QAC chemoreceptor to be identified, the unrelated protein TmoX from *Ruegeria pomeroyi DSS-3* has been shown to bind TMAO in a similar manner to McpX and $Bd1081^{LBD}$. TmoX is a part of an ATP-binding cassette transporter, TmoXWV, which is specific for TMAO. The binding of TMAO is coordinated via cation-π interactions between 4 tryptophan residues in TmoX and the central nitrogen of TMAO (Li et al., 2015). This is the same binding pattern that occurs in McpX (Shrestha et al., 2018) and $Bd1081^{LBD}$, as well as being reminiscent of the aromatic cage seen to help co-ordinate binding in other dCache domain proteins (Ortega et al., 2017a). An aspartate (D131) has also been shown to make a hydrogen bond with the polar oxygen of TMAO (Li et al., 2015), similar to the aspartate (D208 in McpX) that is highly conserved across dCache proteins (Shrestha et al., 2018). This agreement shows that despite the differences in function, the mechanism of binding QACs is conserved and highly dependent on the presence of an aromatic cage. This similarity is common across all dCache proteins and indicates how this particular subfamily of dCache chemoreceptors could have evolved.

Alignment of $Bd1081^{LBD}$ with $McpX^{LBD}$ (6D8V (Shrestha *et al.*, 2018)) shows that the central nitrogen of TMAO in $Bd1081^{LBD}$ aligns with that of proline betaine in the $McpX^{LBD}$ structure (Figure 4.31). The positioning of TMAO within $Bd1081^{LBD}$ also means that the negatively charged oxygen is facing in the same direction as in the $McpX^{LBD}$:proline betaine structure, even if they do not align as a result of ligand structural variation. $Mlp37^{LBD}$ binds taurine (5AVF (Nishiyama *et al.*, 2016)), which is a tertiary ammonium compound, with a ligand conformation similar to that of polyamine binder $McpU^{LBD}$ (6F9G (Gavira *et al.*, 2018)) and $Bd2504^{LBD}$. When $Mlp37^{LBD}$ is compared to both QAC binding dCache proteins, variations in ligand binding positions can be seen, with the QAC binders positioning the nitrogen group towards the centre of the pocket, instead of the proximal end of the pocket as

observed with polyamine binding dCaches. Both TMAO and proline betaine are shown to bind horizontally across the LBP, instead of the vertical binding in Bd2503^{LBD}, or the bent confirmation traditionally seen in other dCache proteins, *e.g.* McpU^{LBD}, Mlp37^{LBD} and Bd2504^{LBD}. This divergence in ligand binding may explain how QAC binding dCache domain proteins bind QACs with a higher affinity than polyamines (choline *Kd* = 8.85 μM, DAP *Kd* = 83.33 μM in Bd1081) or amino acids (choline *Kd* = 138 nm, proline 45.2 μM in McpX^{LBD} (Webb *et al* 2017)); despite QAC-, polyamine- and amino acid-binding dCache proteins all using a similar binding mechanism.

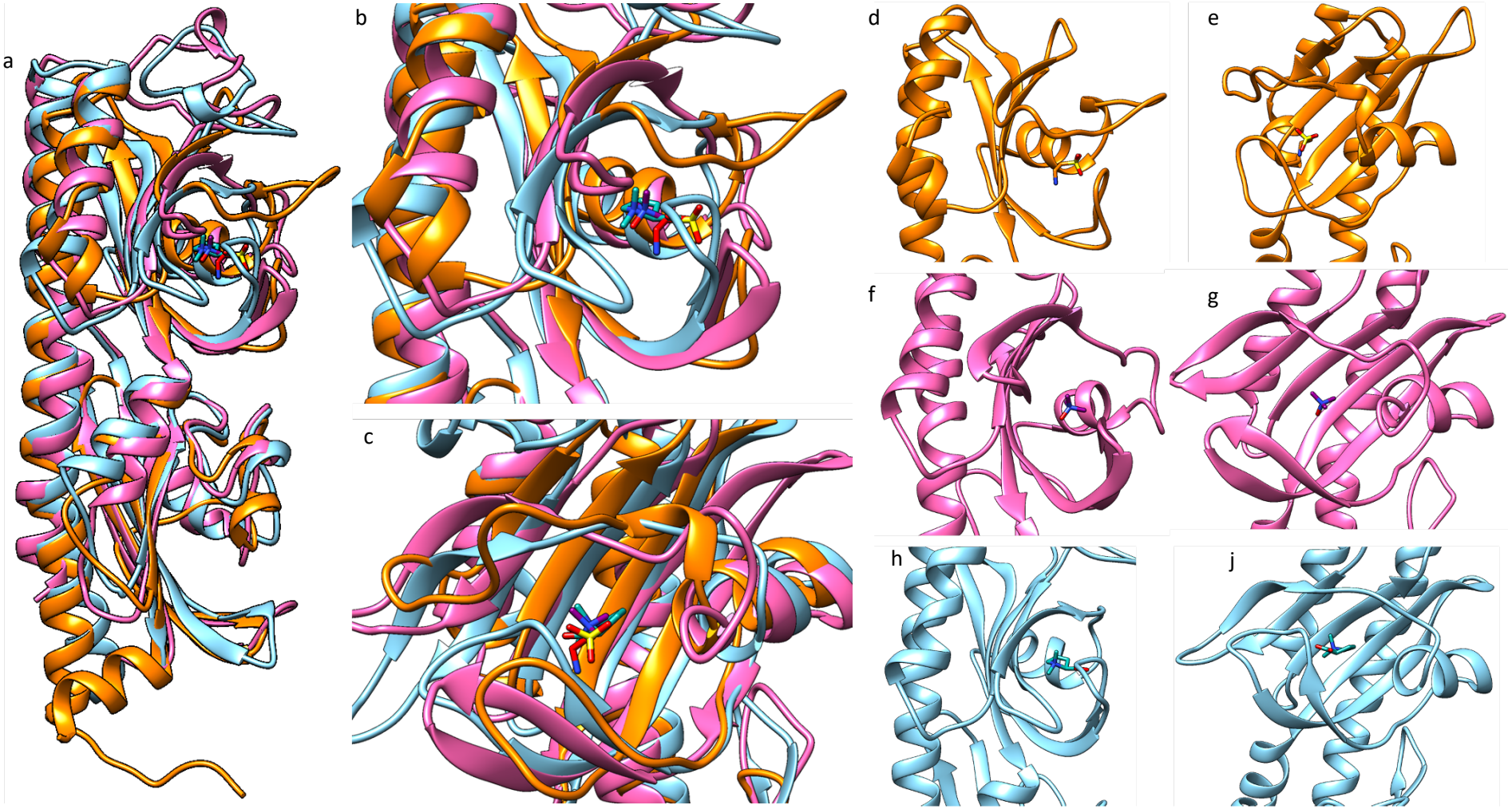


Figure 4.31 (a) Structural comparison of McpX^{LBD} (blue), Mlp37^{LBD} (orange) and Bd1081^{LBD} (pink). **(b)** Close up of the LBP, showing that Bd1081^{LBD} binds TMAO in a conformation where the nitrogen aligns with the proline betaine nitrogen in McpX^{LBD}. Alignment of the QAC binders with Mlp37^{LBD}, which binds taurine, shows a slight divergence between the ligand position in the pocket of QAC binders and other dCache domain chemoreceptors, like Mlp37^{LBD} which binds its ligand more akin to polyamine binding dCache proteins. **(c)** 90° rotation of (b). **(d)** Close up of the LBP pocket of Mlp37^{LBD} with taurine bound. **(e)** 90° rotation of (d). **(f)** Close up of the LBP pocket of Bd1081^{LBD} with TMAO bound. **(g)** 90° rotation of (f). **(h)** Close up of the LBP pocket of McpX^{LBD} with proline betaine bound. **(j)** 90° rotation of (h).

4.3.2 Conservation of the Dimeric Interface

The dimeric interface between dCache domains is made up of two N-terminal α-helices. The first of which is the periplasmic region of the receptor's N-terminal transmembrane helix, and is truncated to this portion in the constructs used for these experiments. The second of these helices is much shorter, comprising 13 amino acids on average and approximately ½ the length of the distal lobe, connected to the transmembrane helix by a turn region. Analysis of the dimeric interface using qtPISA (Krissinel and Henrick, 2007) determined a buried interface of 1448.1 A^2, 1309.5 A^2 and 1427.8 A^2 for $Bd1081^{LBD}$, $Bd2503^{LBD}$ and $Bd2504^{LBD}$ respectively. This software also calculated a low hydrophobic P-value for each dCache ($Bd1081^{LBD}$ = 0.185, $Bd2503^{LBD}$ = 0.183, and $Bd2504^{LBD}$ = 0.013), indicating that the interaction surface is likely to be hydrophobic and also suggests the interaction is pertinent to protein function (Krissinel, 2010). Within the constructs produced herein there are 20 conserved amino acids within the dimeric interface, 17 of which were identifiable in all 3 dCache dimers, $Bd1081^{LBD}$, $Bd2503^{LBD}$ and $Bd2504^{LBD}$ (Figure 4.32). The majority of these conserved amino acids are involved not in dimerisation, but the stabilisation of the dCache domain fold by interacting with the proximal and ligand binding distal lobes (Figure 4.33 and Figure 4.34). Of these conserved amino acids, $Bd2503^{LBD}$ utilises the fewest, 5, in the dimerization interface. All but one of these, N24, is located in the distal section of the protein, 3 of which are in the second helix, V61, N64, M65, and 1 in the turn region, E57 (Figure 4.35). $Bd2504^{LBD}$ uses 6 of the conserved amino acids; 1 in the transmembrane helix, Q24, 2 in the turn region, Q52 and K57, and 3 in the second helix, L61, E64 and V65. $Bd1081^{LBD}$ utilises 7 of the conserved amino acids in dimerisation; 2 in the transmembrane region, T22 and R25, 2 in the turn region, N50 and Q55, and 3 in the second helix, A59, E62 and I63.

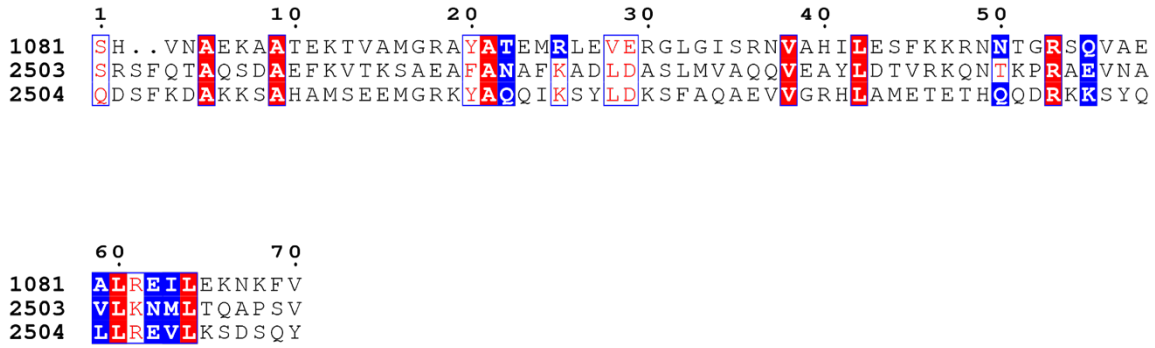


Figure 4.32 Alignment of the dimeric region of each dCache construct, highlighted in red boxes are amino acids that are strictly conserved, whilst those in red text are chemically similar (generated using ESPript 3.0 (Robert and Gouet, 2014)). The amino acids highlighted in blue are involved in dimerisation. The solved structures for Bd1081^{LBD}, Bd2503^{LBD} and Bd2504^{LBD} start at K12, E12 and K8 respectively.

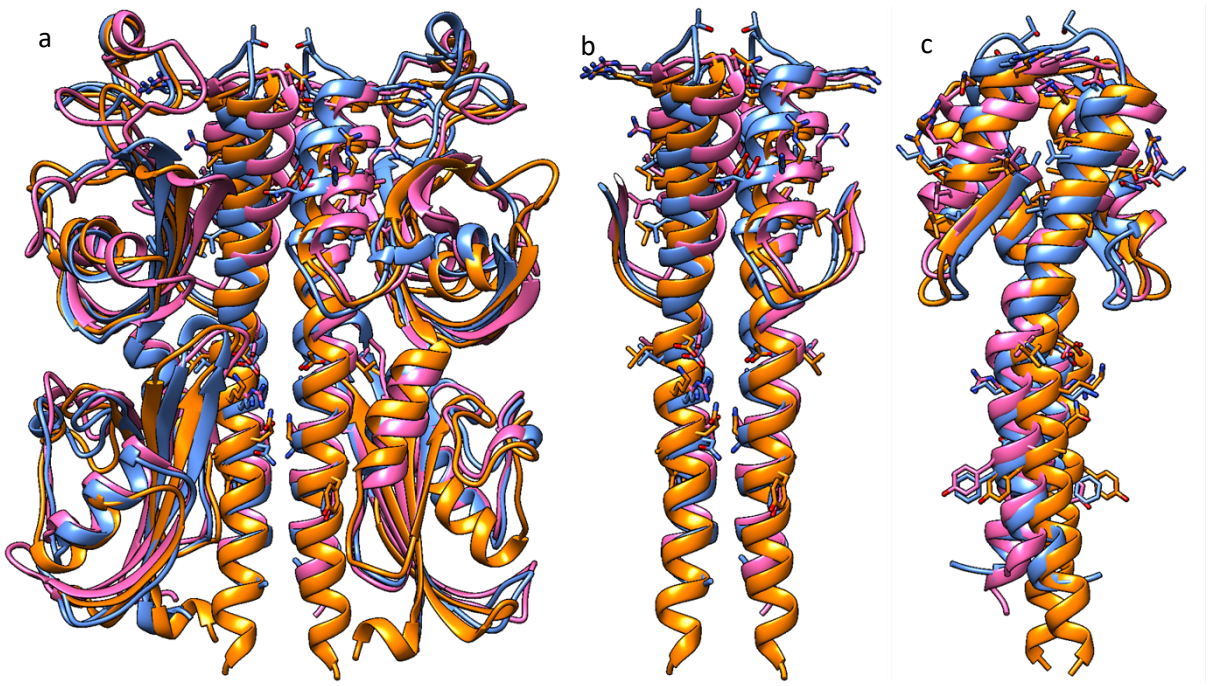


Figure 4.33 Alignment of all three dimeric dCache structures, Bd1081^{LBD} (pink), Bd2503^{LBD} (blue) and Bd2504^{LBD} (orange). **(a)** All conserved amino acids in the dimerisation region showing that the majority face towards the dCache fold. **(b)** Clipped structure of aligned dCache structures showing only the two helices involved in dimerisation. Removal of the dCache fold highlights that the majority of the conserved amino acids face away from the dimerisation interface. **(c)** 90 ° rotation of (b).

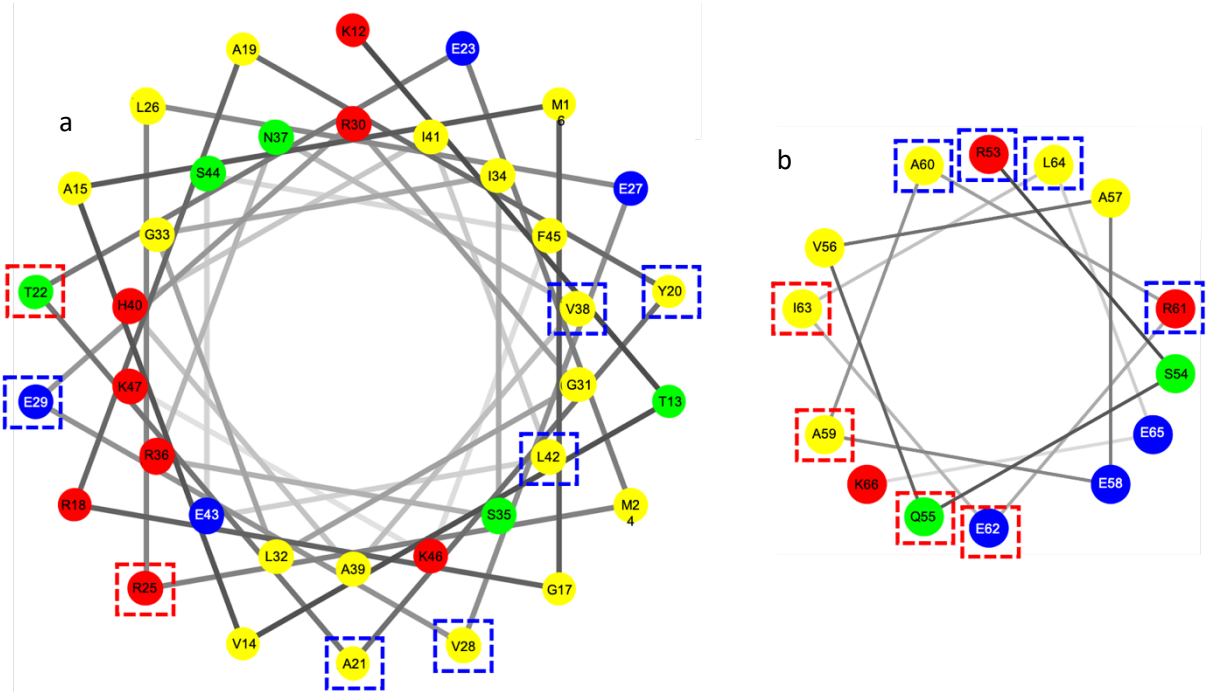


Figure 4.34 Helical wheel diagram of the two helices Bd1081^{LBD} involved in dimerisation showing that the majority of the conserved amino acids are not involved in dimerisation. The highly conserved amino acids are highlighted by boxes, those in red boxes are involved in dimerisation, whilst the ones in blue boxes are only highly conserved. **(a)** Transmembrane helix. **(b)** Helix involved in connecting the LBP to the transmembrane region.

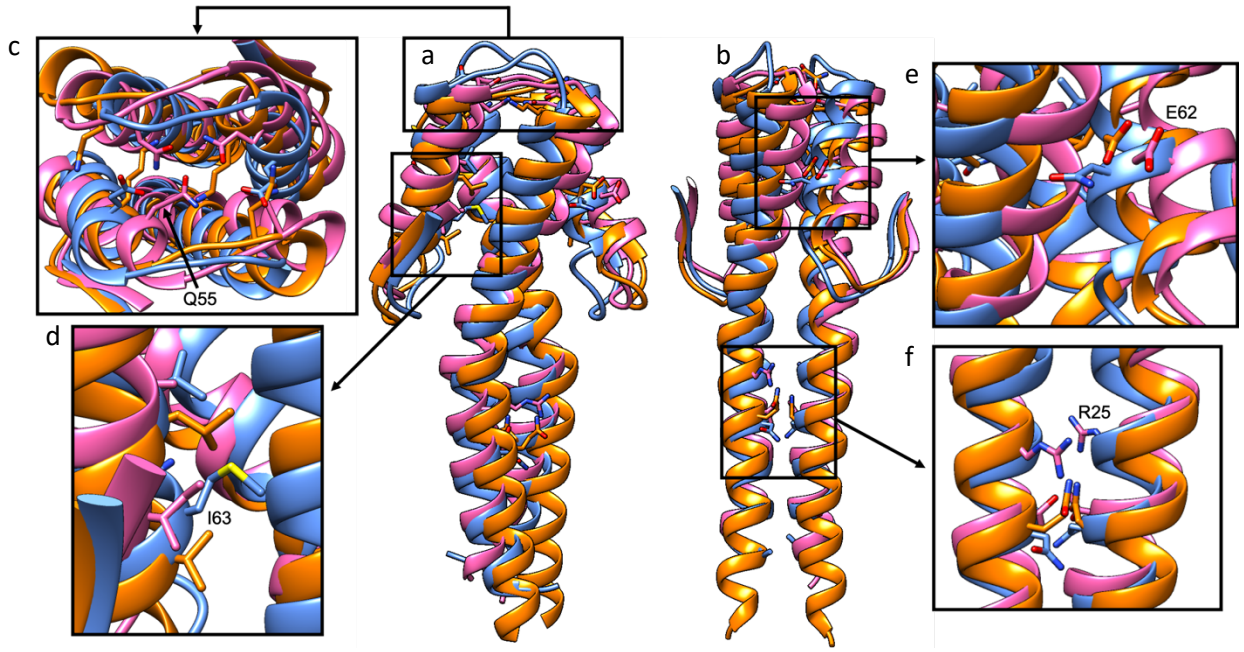


Figure 4.35 (a) Alignment of all three dimeric dCache structures, Bd1081^{LBD} (pink), Bd2503^{LBD} (blue) and Bd2504^{LBD} (orange), showing only the conserved amino acids involved in dimerisation within each dCache structure. **(b)** 90 ° rotation of (a). **(c)** Close up of the amino acids involved in dimerisation located in the turn region, Q55 in Bd1081^{LBD} labelled. **(d)** and **(e)** Close up of amino acids involved in dimerisation located in the second helix, I63 and E62 in Bd1081^{LBD} labelled respectively. **(f)** Close up of amino acids involved in dimerisation located in the proximal half of the transmembrane region, R25 in Bd1081^{LBD} labelled.

4.3.3 Modelling of Ligands Determined via ITC

The ITC experiments showed binding between the Cache proteins and only 1-2 ligands that were in most cases highly related. However, without preforming high-throughput ligand screening the true specificity of these Cache domain proteins cannot be determined. Bd2503^{LBD} was found to bind both 5C cadaverine and 3C DAP polyamines, whereas Bd2504^{LBD} was shown to bind only 4C putrescine with high enough affinity to obtain a confident *Kd*. As a result of the similarity in chemical properties between cadaverine and DAP, binding between Bd2503^{LBD} and DAP is thought to take place in the same LBD via a similar amino acid usage. The ITC experiments also showed that Bd2503^{LBD} did not bind putrescine, which has led to the postulation that binding between Bd2503^{LBD} and DAP is not the result of contraction of the LBP, but bridging of the difference in distance by a water molecule. The formation of a hydrogen bond between one of the terminal amines and a water molecule, and then the

subsequent formation of hydrogen bonds between the water molecule and amino acids within the LBD would allow for DAP to bind vertically within the LBP in a conformation similar to the binding of cadaverine. The binding of DAP to Bd2503^{LBD} was modelled by fitting DAP within the electron density of cadaverine in the Bd2503^{LBD}:cadaverine complex (Figure 4.36). Although a water molecule could theoretically replace the binding of either terminal amine group, one was modelled at the distal end of the pocket, rather than binding at the proximal end of the pocket. This was because previous comparisons of Bd2503^{LBD} to other polyamine binding proteins (Chapter 3.3.2) revealed the distal binding position of a terminal amine group was conserved more readily than that of the proximal end. This is supported by the Bd2503^{LBD}:arginine structure, where the sidechain amine group forms hydrogen bonds with the same distal amino acids that are involved in the binding of cadaverine. However the proximal amino acids that bind the amino acid backbone of arginine show more variation and flexibility in binding. The model of Bd2503^{LBD} with DAP showed complete conservation of the amino acids used in the binding of cadaverine, with amino acids A77, T78, G104, D121 all forming hydrogen bonds with the water molecule, and Y144, D146, M155 and D175 all forming hydrogen bonds with DAP. The same amino acids, W76, Y106 and F128, were all shown to form the same aromatic cage around the 3C backbone of DAP, as was seen formed around the 5C backbone of cadaverine. The only difference in amino acid binding pattern between the Bd2503^{LBD} cadaverine structure and the DAP model is an additional hydrogen bonded form between the introduced water molecule and P105. However, P105 was shown to be involved in the binding to arginine in the Bd2503^{LBD}:arginine structure.

Although no *Kd* could be established, ITC experiments did show some binding response between Bd2504^{LBD} and cadaverine. This might suggest that unlike Bd2503^{LBD} with putrescine, where the polyamine is too small to bind even with flexing of the pocket and too large for a water molecule to bridge the gap, cadaverine maybe able to bend enough for a small expansion in the pocket size in Bd2504^{LBD} to allow binding to occur. Bd2504^{LBD} is thought to not be able to bind DAP as the single

carbon truncation is enough to prevent the formation of the hydrogen bonds required to stabilise binding at the distal amine group. The high specificity provided by the variation in ligand binding conformation may indicate why the excludon gene arrangement of Bd2503 and Bd2504 has arisen. A search of Bd2503 and Bd2504 in the excludon arrangement using MultigeneBlast *bacteriovorus* (Medema et al., 2013), did not identify any other instances of this excludon architecture outside of *Bdellovibrio HD100*, *tiberius* and *SSB218315.*

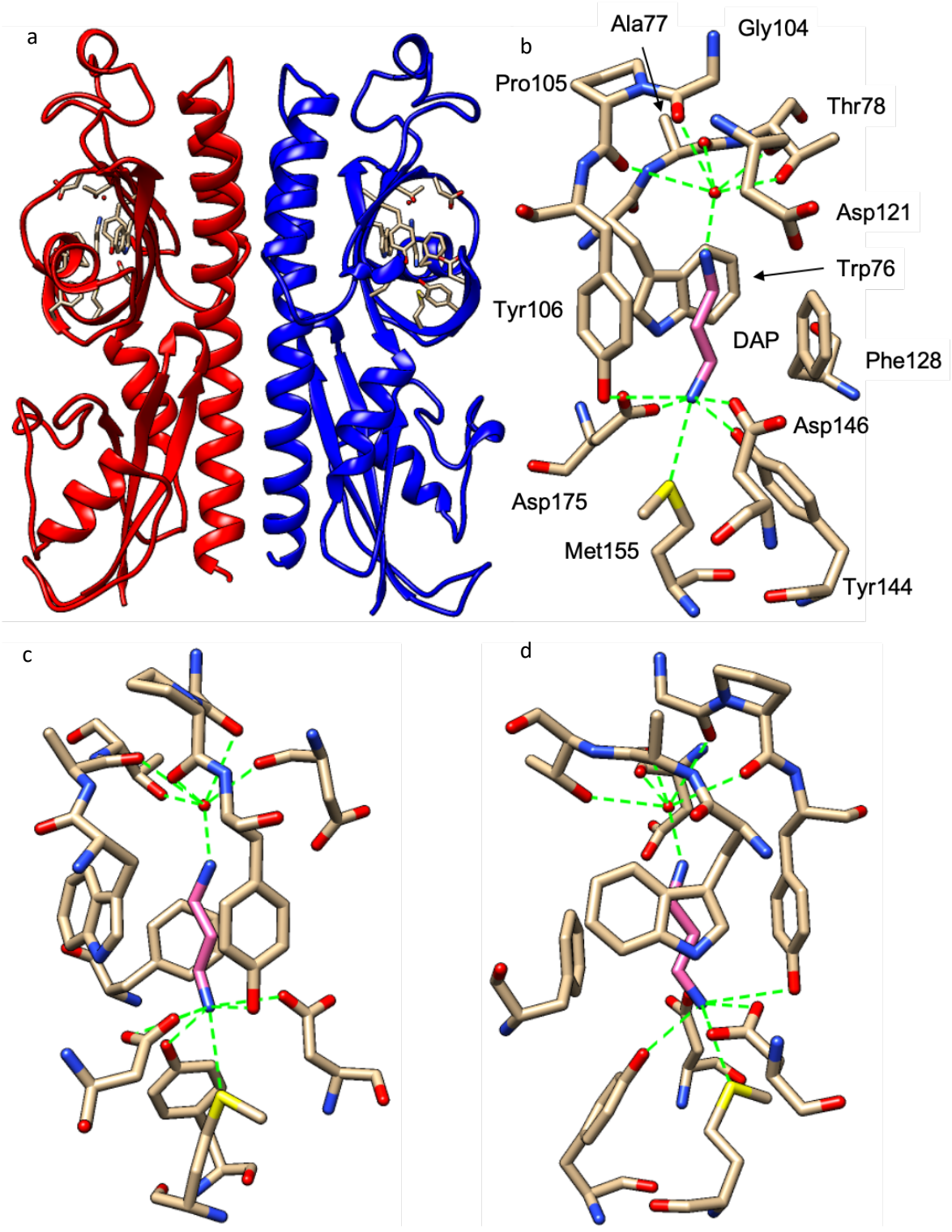


Figure 4.36 (a) Model of the binding between Bd2503^{LBD} and DAP. **(b-d)** Close up of Bd2503^{LBD} binding pocket with DAP (pink) binding modelled, showing a water molecule bridging the gap between the proximal terminal amine group and the proximal LBP amino acids. Hydrogen bonds are shown in green.

ITC showed that both *HD100* Bd1081^{LBD} and SKB1291214 homologue Bd1081$^{S\text{-}LBD}$ bound choline, but neither were shown to bind to TMAO, as suggested by the Bd1081^{LBD}:TMAO structure. However, because of the chemical similarity between TMAO and choline, Bd1081^{LBD} binding to choline was assumed to occur in a manner similar to that presented in the Bd1081^{LBD}:TMAO structure. Choline was fit into the TMAO density within the Bd1081^{LBD}:TMAO complex (Figure 4.37), using the position of the ammonium group in the Bd1081^{LBD}:TMAO structure as a guide, as this seems to be conserved in other complexes. Similar to the 'binding' of TMAO in the Bd1081^{LBD}:TMAO structure, no hydrogen bonds are predicted to form between Bd1081^{LBD} and choline, with the model again appearing to rely on cation-π interaction to coordinate binding. The same amino acids involved in these cation-π interactions in the Bd1081^{LBD}:TMAO structure, W74, Y104, Y128 and Y146, are involved in binding choline in the Bd1081^{LBD}:choline model. However, the length of these interactions vary more in the Bd1081^{LBD}:choline model, 3.97 – 5.10 Å, than the Bd1081^{LBD}:TMAO structure, 4.48 – 4.81 Å. This is likely a result of modelling the ligand but may reflect how choline binds within the pocket as choline is less symmetrical than TMAO, which may cause bond lengths to become skewed within the pocket. Hydrophobic packing from L117 and D175 is also conserved in the Bd1081^{LBD}:choline model compared to Bd1081^{LBD}:TMAO, along with additional packing from S157 present in the Bd1081^{LBD}:choline complex. The apparent conservation of protein:ligand contacts between Bd1081^{LBD}:choline compared to Bd1081^{LBD}:TMAO seems to suggest that small changes in ligand structure along with small differences in position are enough to effect specificity of ligand binding, therefore resulting in no ligand binding between Bd1081^{LBD} and TMAO in ITC experiments.

Choline was modelled into Bd1081$^{S\text{-}LBD}$ by superimposing Bd1081^{LBD} and using the position of choline within the prior Bd1081^{LBD}:choline model as a guide (Figure 4.38). Similarly to both Bd1081^{LBD}:TMAO and Bd1081^{LBD}:choline, binding of choline within Bd1081$^{S\text{-}LBD}$ is co-ordinated by cation-π interaction and hydrophobic packing instead of hydrogen bonding. Conversely, the Bd1081$^{S\text{-}LBD}$:choline model does not utilise 4 cation-π interactions, merely 3 between W74, Y104 and Y128 with a distances

between 4.16 – 5.82 Å. However, as the structure of Bd1081$^{S\text{-}LBD}$ used is considered the apo form, based on both the binding pocket of Bd1081$^{S\text{-}LBD}$ and non-physiological Bis tris being hydrogen bond donor rich (Chapter 3.2.3.2), this variation within the binding cage is likely a result of conformational changes within the LBP between apo and bound states. Hydrophobic packing using L117, S157 and D175 is conserved between the Bd1081$^{S\text{-}LBD}$:choline model and the Bd1081^{LBD}:choline model.

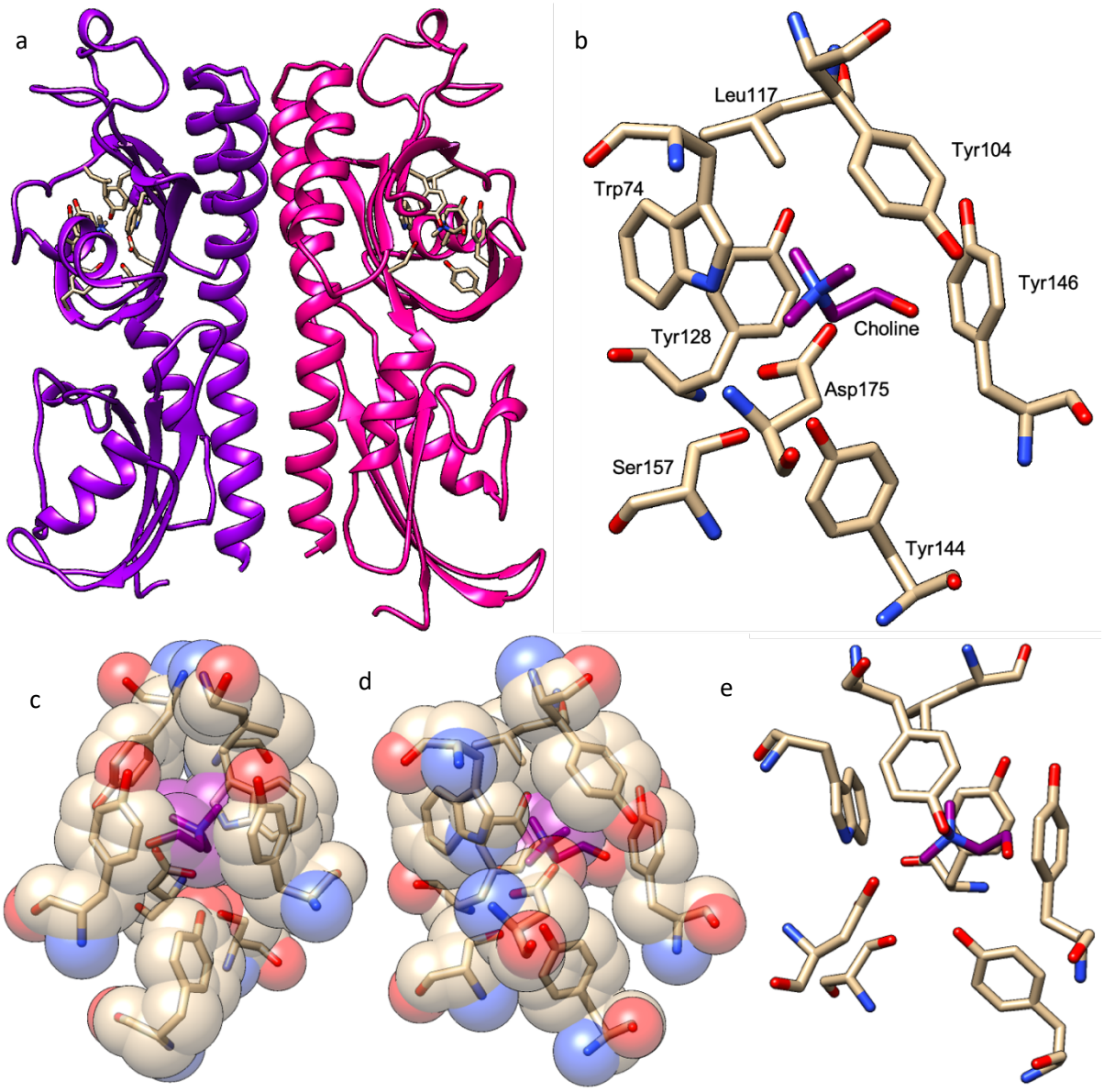


Figure 4.37 (a) Model of the binding between Bd1081^{LBD} and choline. **(b)** Close up of Bd1081^{LBD} binding pocket with choline (purple) binding modelled. **(c)** Display of the binding pockets of Bd1081^{LBD} with choline with Van der Waals surface showing that the ligand fills the pocket upon binding even if no hydrogen bonds form. **(d)** 90° rotation of (c). **(e)** 90° rotation of (b).

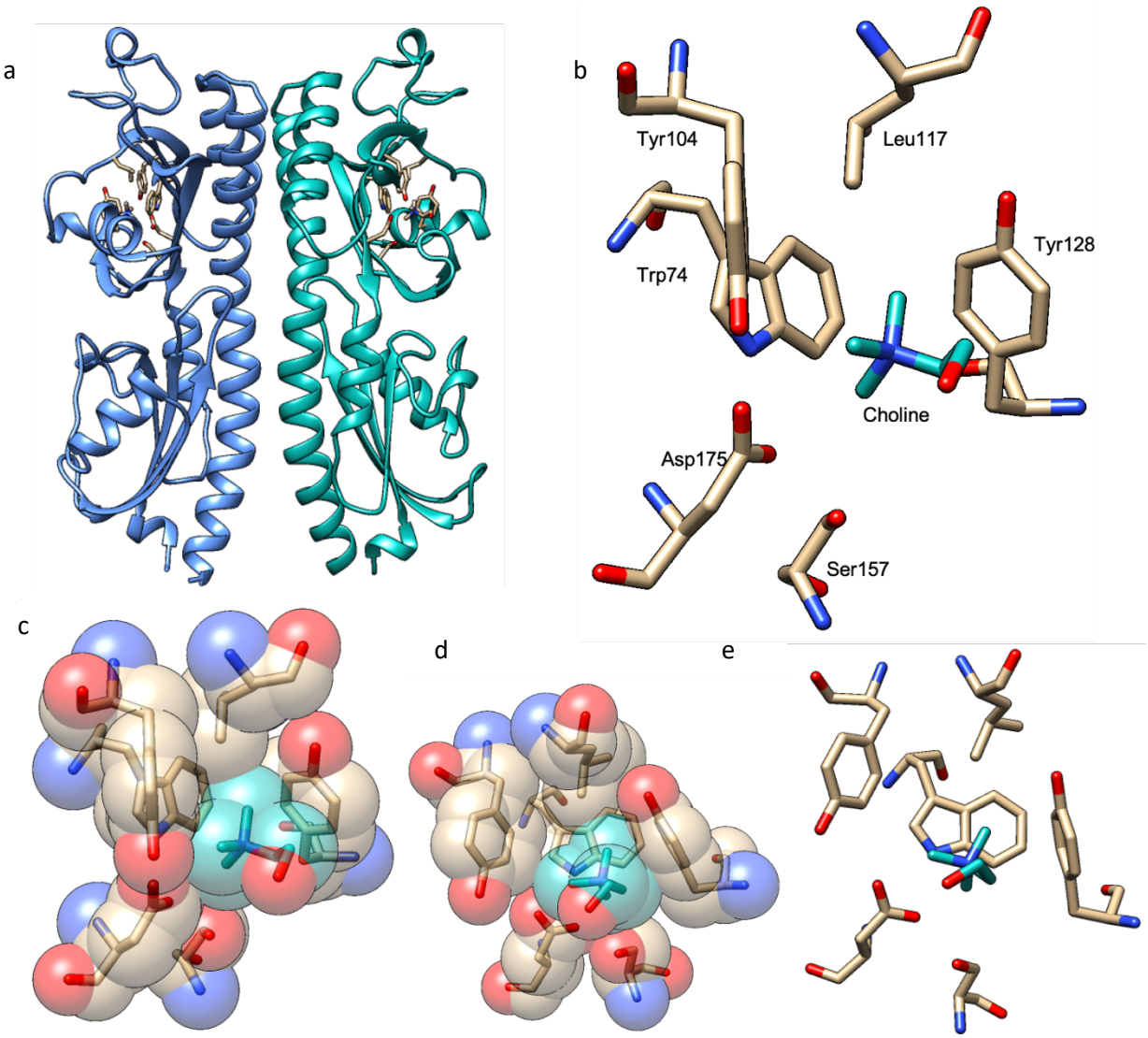


Figure 4.38 (a) Model of the binding between Bd1081$^{S\text{-}LBD}$ and choline. **(b)** Close up of Bd1081$^{S\text{-}LBD}$ binding pocket with choline (blue) binding modelled. **(c)** Display of the binding pockets of Bd1081$^{S\text{-}LBD}$ with choline with Van der Waals surface showing that the ligand fills the pocket upon binding even if no hydrogen bonds form. **(d)** 90° rotation of (c). **(e)** 90° rotation of (b).

Both Bd1081^{LBD}:choline and Bd1081$^{S\text{-}LBD}$:choline models were superimposed against the structures of Bd1081^{LBD}:TMAO and McpX^{LBD}:proline betaine (6D8V (Shrestha *et al.*, 2018)) (Figure 4.49). The ammonium group of all 4 ligands aligned within a similar central position within the LBP. The alcohol group in choline in both models aligns with the carboxylic acid group of proline betaine in the McpX^{LBD}:proline betaine structure, whilst the polar charge faces in the same direction in these models as the negative charge on the TMAO in the Bd1081^{LBD}:TMAO structure. This alignment of the structures shows a consistent binding of QACs horizontally across the LBP within the model, furthering the hypothesis this is the conformation in which QAC-binding dCache domain proteins orient QACs. This alignment also shows that the LBP within Bd1081^{LBD} and Bd1081$^{S\text{-}LBD}$ is narrower and longer than that of McpX^{LBD}, and has more similar dimensions to polyamine binding dCache LBPs. This may be the reason that McpX^{LBD} has been shown to be able to bind a wide range of QACs, whereas Bd1081^{LBD} only binds choline with even the small chemical variance between choline and glycine betaine being enough to abolish binding.

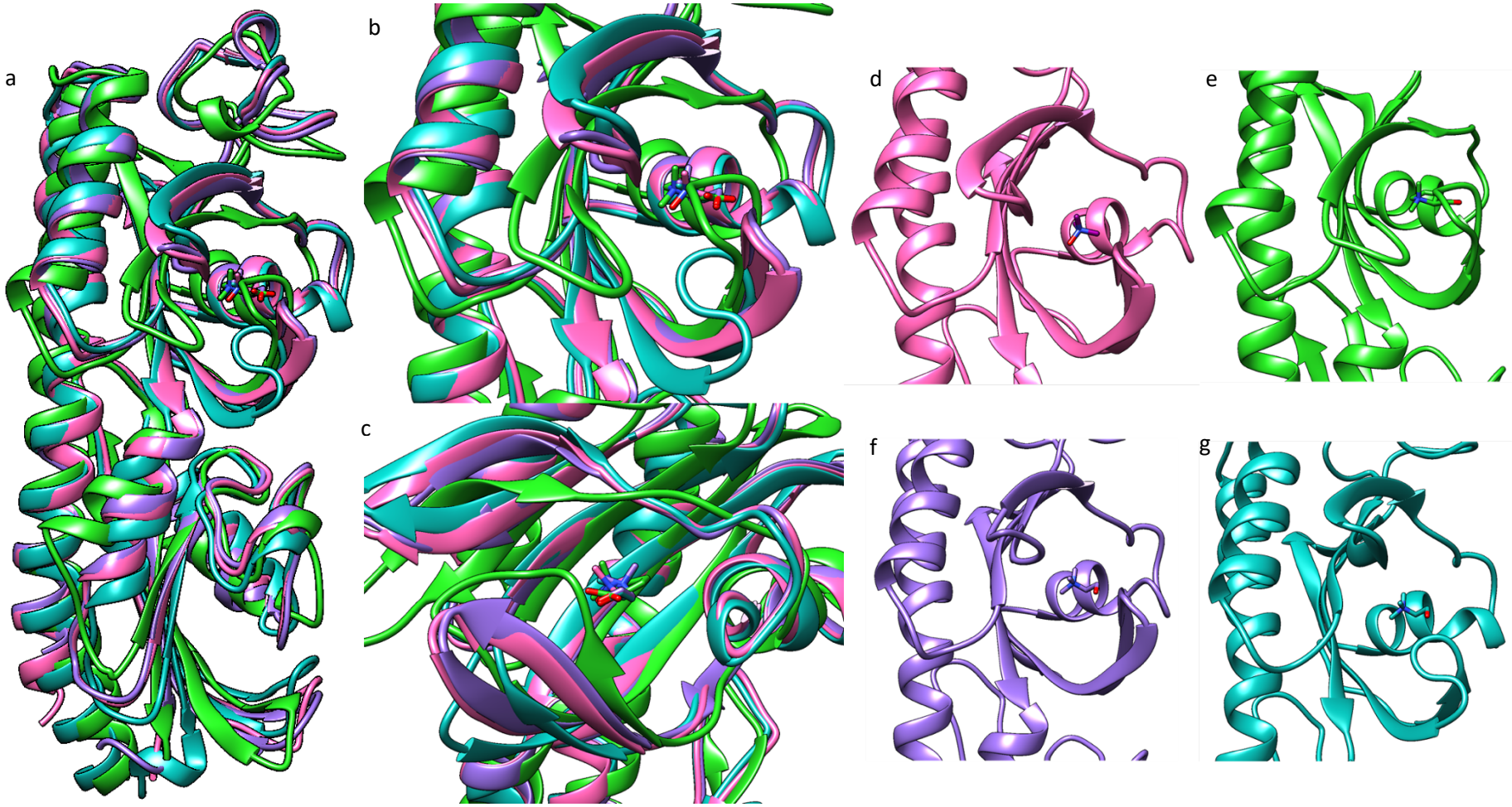


Figure 4.39 (a) Structural comparison of Bd1081^{LBD}:TMAO (pink), Bd1081^{LBD}:choline (purple), Bd1081$^{S\text{-}LBD}$:choline (teal), and McpX^{LBD}:proline betaine (green). **(b)** Close up of the LBP, showing that the ammonium groups in the model of Bd1081^{LBD}:choline and Bd1081$^{S\text{-}LBD}$:choline aligned with Bd1081^{LBD}:TMAO and McpX^{LBD}:proline betaine. Models also conserve the flat horizontal binding of the ligand across the binding pocket and the directionality of the oxygen group. **(c)** 90° rotation of (b). **(d)** Close up of the LBP pocket of Bd1081^{LBD} with TMAO bound. **(e)** Close up of the LBP pocket of McpX^{LBD} with proline betaine bound. **(f)** Close up of the LBP pocket of Bd1081^{LBD} with choline bound. **(g)** Close up of the LBP pocket of Bd1081$^{S\text{-}LBD}$ with choline bound.

ITC also demonstrated affinity between Bd1081^{LBD} and DAP, although the binding was significantly worse than with QACs. DAP was fit over the TMAO density within the Bd1081^{LBD}:TMAO structure (Figure 4.40). Unlike with the Bd1081^{LBD}:TMAO structure and Bd1081^{LBD}:choline model, binding of DAP to Bd1081^{LBD} involves the formation of a hydrogen bond between Tyr144 and the proximal terminal amine group. Amino acids W78, Y128 and Y146, which form part of the aromatic cage, form cation-π interactions with the distal terminal amine group with distances of 5.06 Å, 3.18 Å and 5.52 Å respectively, similar to both the Bd1081^{LBD}:choline model and Bd1081^{LBD}:TMAO structure. The final amino acid in the aromatic cage, Y104, stabilises the poly-carbon backbone via hydrophobic packing. Similarly to the Bd1081^{LBD}:TMAO structure, L117 and D175 stabilise binding via hydrophobic packing, along with M155, which is conserved in polyamine binding. The resulting binding pattern is a composite of both QAC and polyamine LBPs.

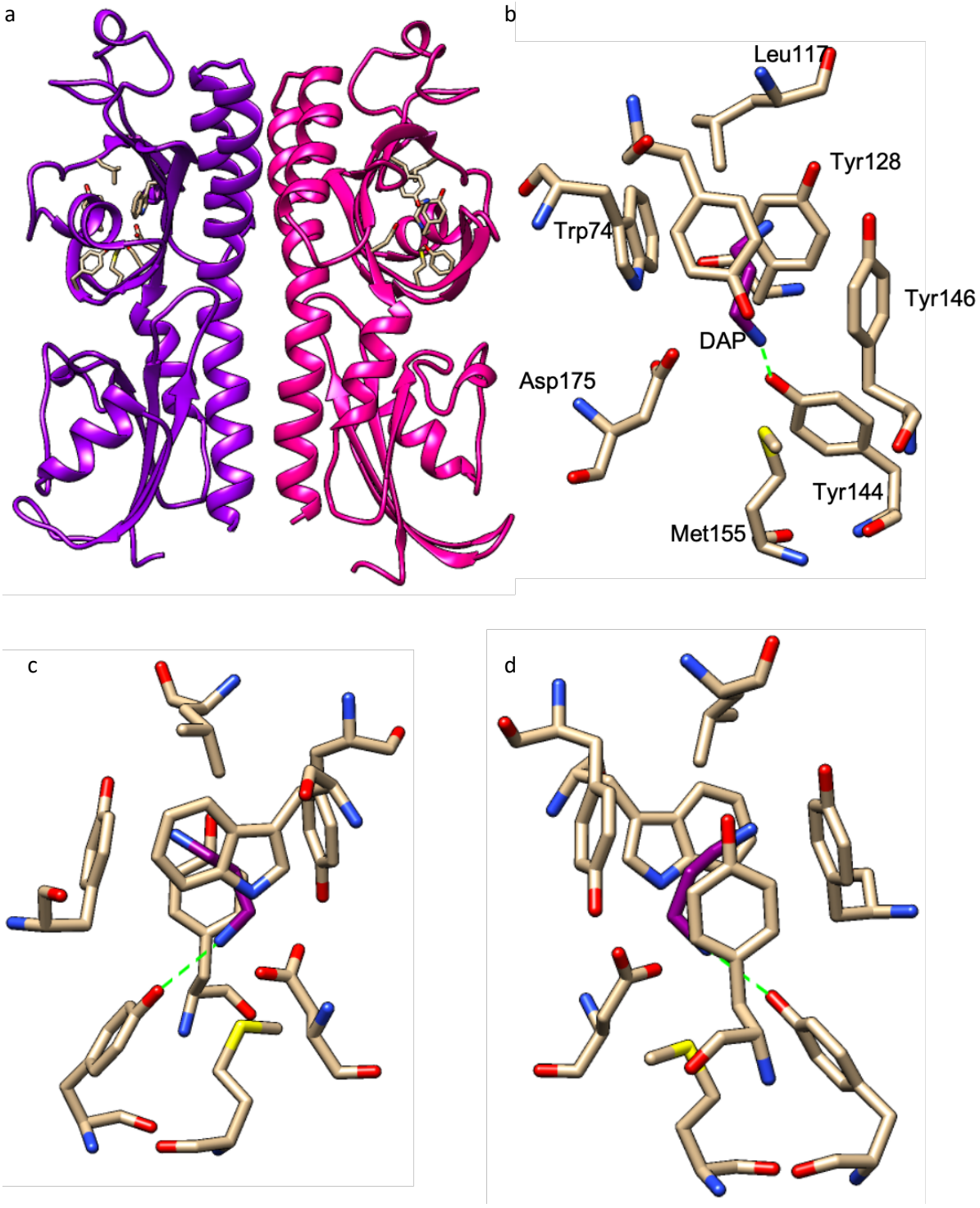


Figure 4.40 (a) Model of the binding between Bd1081^{LBD} and DAP. **(b-d)** Close up of Bd1081^{LBD} binding pocket with DAP (purple) binding modelled, showing the single hydrogen bond that is

The model of Bd1081^{LBD}:DAP was superimposed against the structures of Bd1081^{LBD}:TMAO, Bd2504^{LBD}:putrescine, McpU^{LBD}:putrescine (6F9G (Gavira *et al.*, 2018)) and Mlp37^{LBD}:taurine (5AVF (Nishiyama *et al.*, 2016)) (Figure 4.41). This alignment shows the proximal terminal amine group of DAP in the Bd1081^{LBD}:DAP model aligns with the same amine group in both the polyamine and the taurine binding dCache receptors, rather than the ammonium group of TMAO in the Bd1081^{LBD}:TMAO complex. The DAP molecule also adopts the 90° bend seen for the other ligands of this domain, which would enable binding of the distal amine via cation-π interaction to stabilise the ligand within the pocket. However, it also conserves the general directionality of the distal amine group shown in other polyamine binding proteins.

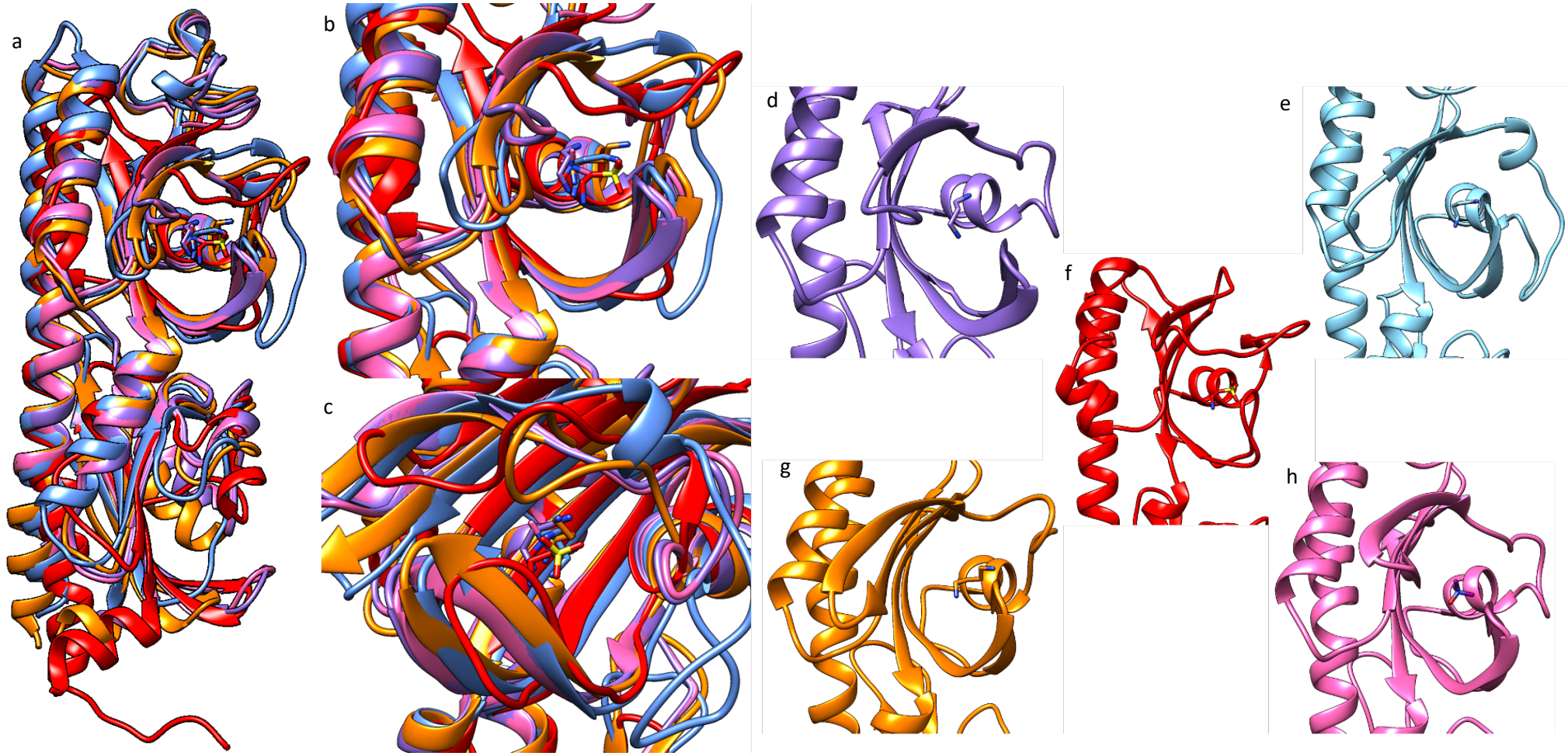


Figure 4.41 (a) Structural comparison of Bd1081^{LBD}:TMAO (pink), Bd1081^{LBD}:DAP (purple), Bd2504^{LBD}:putrescine (orange), McpU^{LBD}:putrescine (blue) and Mlp37^{LBD}:taurine (red). **(b)** Close up of the LBP, showing that the proximal terminal amine group of the Bd1081^{LBD}:DAP model aligns with other polyamine and related ligand binding dCaches, rather than the ammonium group of TMAO in the Bd1081^{LBD}:TMAO complex, despite fitting of DAP to the same map density. Bd1081^{LBD}:DAP model also adopts a 90° ligand bend traditionally seen in polyamine binding of dCache domain proteins and conserves the general directionality of the distal amine group. **(c)** 90° rotation of (b). **(d)** Close up of the LBP pocket of Bd1081^{LBD} with DAP bound. **(e)** Close up of the LBP pocket of McpU^{LBD} with putrescine bound. **(f)** Close up of the LBP pocket of Mlp37^{LBD} with taurine bound. **(g)** Close up of the LBP pocket of Bd2504^{LBD} with putrescine bound. **(h)** Close up of the LBP pocket of Bd1081^{LBD} with TMAO bound.

As previously stated, the overall dimensions of the LBP of Bd1081^{LBD} are more similar to other polyamine binding dCache pockets, than the QAC binding McpX^{LBD}. This may be the reason why Bd1081^{LBD} can bind very selectively to a single polyamine and a single QAC, and would be similar to the cross ligand family binding demonstrated for McpX^{LBD}, binding not only a wide range of QACs but also proline (Webb *et al* 2017). This cross-family binding has led to postulation within our lab that QAC binding dCache domain proteins may have arisen from amino acid binding dCache proteins. In *Bdellovibrio* the dCache domains of Bd2503 and Bd2504 are as similar to Bd1081, 33.7% and 30.4% respectively, as they are to each other, 30.2%. This coupled with the cross ligand binding ability shown by Bd1081, has led to us to generate the hypothesis that either/both Bd2503 and Bd2504 arose from gene duplication of Bd1081. This is supported by Bd1081 being overwhelmingly the most conserved dCache MCP in *Bdellovibrio* strains, with only epibiotic variants, *Bdellovibrio exovorus, qaytius* and *ArHS* (Deeg et al., 2018; Pérez et al., 2016; Sangwan et al., 2015) being shown to possess no significant homologues of Bd1081, Bd2503 or Bd2504 (Table 4.7). Of the two polyamine binding dCache proteins it is more likely Bd2504 arose first in *Bdellovibrio*, this is because it utilizes the more classical polyamine binding conformation which appears to be shared with Bd1081 when binding DAP. Bd1081 and Bd2504 are also present in one *Bdellovibrio* strain, *SKB1291214*, without Bd2503, but the opposite (Bd1081 and Bd2503 without Bd2504) could not be identified in any strain via BlastP (Altschul et al., 1990).

dCaches present	*Bdellovibrio* strains
No homology to Bd1081, Bd2503 or Bd2504	*Bdellovibrio ArHS*, *Bdellovibrio exovorus* and *Bdellovibrio qaytius*
Only Bd1081	*Bdellovibrio NC01*, *Bdellovibrio W* and *Bdellovibrio zap7*
Bd1081 and Bd2504	*Bdellovibrio SKB1291214*
All 3: Bd1081, Bd2503 and Bd2504	*Bdellovibrio bacteriovorus HD100*, *Bdellovibrio tiberius* and *Bdellovibrio* SSB218315

Table 4.7 BlastP search (Altschul et al 1990) for the presence of Bd1081, Bd2503 and Bd2504 in *Bdellovibrio* strains.

ITC experiments showed Bd3092^{LBD} bound 2C and 3C organic acids, acetate and propanoic acid. Both of these chemicals have been previously identified as eliciting a strong chemoattractant effect from *Bdellovibrio* in early chemotaxis assays undertaken by Straley *et al*. This study also identified thioacetate as a strong chemoattractant (Straley et al., 1979), however the small chemical difference between acetate and thioacetate was shown to be enough to prevent binding of thioacetate to Bd3092^{LBD}, highlighting the specificity of this sCache receptor. Propanoic acid was fit into the acetate density within the Bd3092^{LBD}:acetate structure (Figure 4.42), using the position of the carboxylic group in acetate in the Bd3092^{LBD}:acetate structure as a guide, as this is utilised in the formation of hydrogen bonds to bind the ligand. All the amino acids used in binding acetate, are conserved in the binding of propanoic acid with Y79, H72, Y113 and K126 continuing to form hydrogen bonds with the carboxylic group. However, an aromatic cage like conformation is present in the Bd3092^{LBD}:propanoic acid model but not in the Bd3092^{LBD}:acetate structure, as a result of additional hydrophobic packing from W61 on the opposite face of the binding pocket to W115, which is present in the Bd3092^{LBD}:acetate structure. Further hydrophobic packing is provided by L94, F95 and F98, which results in propanoic acid being more tightly enclosed within the pocket than acetate. Unlike both polyamine and QAC binding dCache receptors, there are multiple examples of organic acid binding sCache domains proteins including McpP^{LBD} in *Pseudomonase putida*, McpV^{LBD} in *Sinorhizobium meliloti* and PscD^{LBD} in *Pseudomonas syringae* which all bind both acetate and propanoic acid (Martín-Mora et al., 2018; Ortega et al., 2017a). All of these examples have been shown to bind a wide range of small organic acids, with little difference in *Kd* magnitude reported for both acetate and propanoic acid (García et al., 2015; Martín-Mora et al., 2018). Conversely, there is an order of magnitude difference in the tightness of binding to Bd3092^{LBD} between acetate and propanoic acid, with Bd3092^{LBD} showing a higher affinity for propanoic acid. This large variation in *Kd* may indicate that Bd3092^{LBD} binds a smaller range of ligands than previously identified organic acid sCache receptors, however further experiments are required to confirm this.

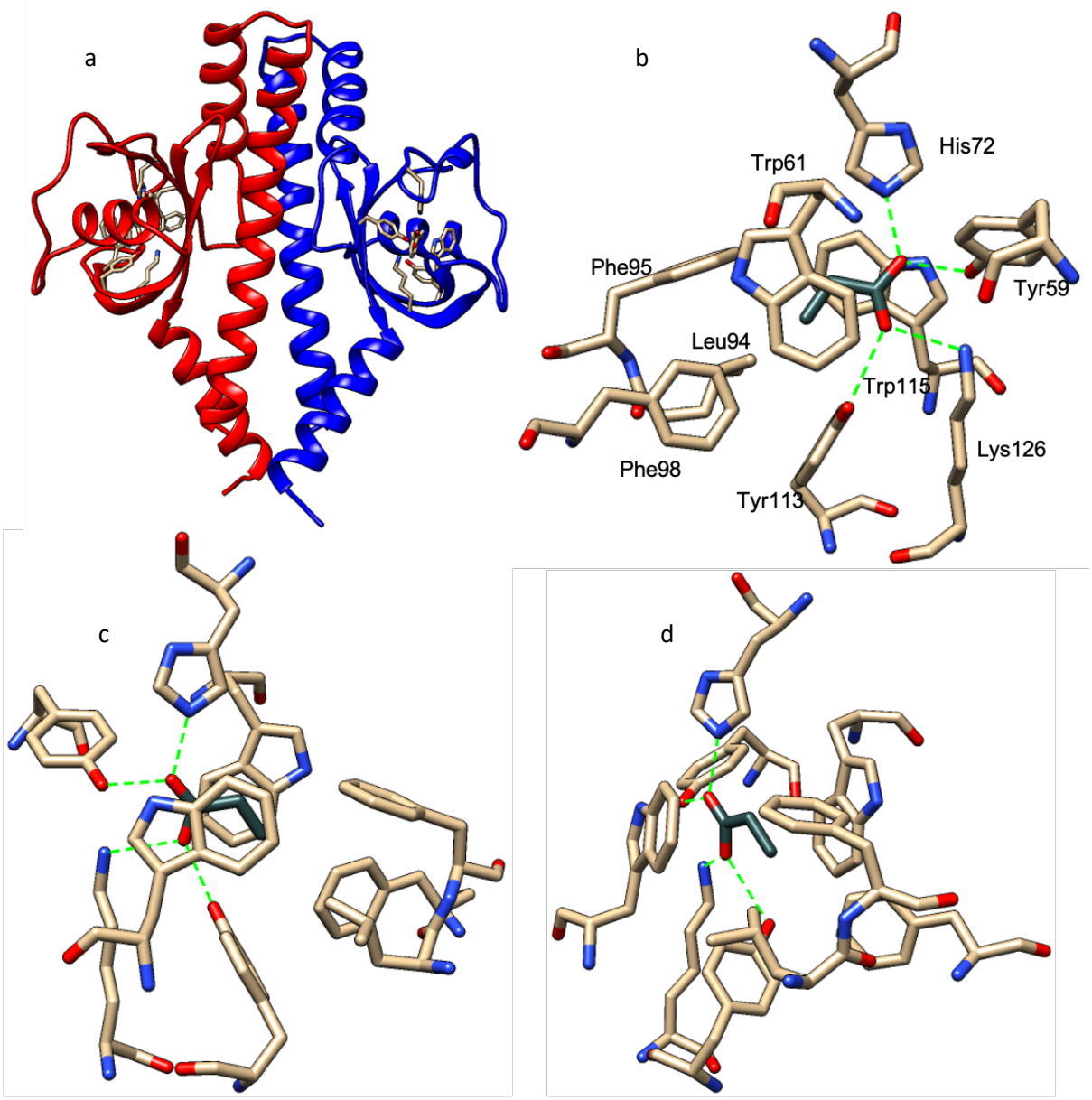


Figure 4.42 (a) Model of the binding between Bd3092^{LBD} and propanoic acid. **(b-d)** Close up of Bd3092^{LBD} binding pocket with propanoic acid (grey) binding modelled, showing hydrogen bonds in green.

4.3.4 Conformational Change Between Apo and Bound dCache Structures

Although not an original aim for this thesis, obtaining both 'apo' and bound conformations of the dCache receptor homologues has allowed for the development of a hypothesis on how signal transduction occurs across the membrane upon ligand binding. Alignment of the bound dCache domains onto the apo dCache reveals movement of amino acids Y144-M155 within a β-sheet located at the distal end of LBP away from the membrane and across the centre of the pocket, causing the

pocket to enclose more around the ligand (Figure 4.43). A conformational change of the α-helix (K179–K187 in Bd1081^{S}) that connects the two lobes of the dCache fold is also apparent. The α-helix in bound state is displaced towards the proximal lobe and away from the membrane, this seems to be caused by tightening of the helix at the proximal end which causes amino acid L178 to extend the helix. This is thought to be a direct result of the distal β-sheet shifting and causing a conformational change of the proximal lobe upon ligand binding. The transmembrane α-helix (R7–R25 and V28–K47 in Bd1081^{S}) also shifts sideways slightly, this slight conformational changes causes the clockwise tightening of the transmembrane α-helix. This results in the formation of a continuous α-helical backbone, removing the two amino acid kink observed midway in the α-helix within the apo structure. This would allow full interaction between the two monomers to occur, causing the formation of the dimer-stabilising hydrogen bonds previously discussed. All current theories of transmembrane signal transduction focus on changes in the second transmembrane α-helix with respect to the first transmembrane α-helix (Chapter 1.2) (Bi and Lai, 2015). Although the structures presented here include very little of this second transmembrane α-helix, they do show slight conformational changes within the first transmembrane α-helix upon ligand binding. Within these structures the opposite movement to what is proposed in the both the Helical rotation and Piston displacement theories is seen within the first transmembrane α-helix (Bi and Lai, 2015), where tightening of the α-helix via a clockwise rotation would cause displacement of this α-helix upwards with respect to the second transmembrane helix. To our knowledge this is the first time a role in transmembrane signal transduction has been shown for the first transmembrane helix, however it may be an artefact caused by domain truncation in our crystallization constructs.

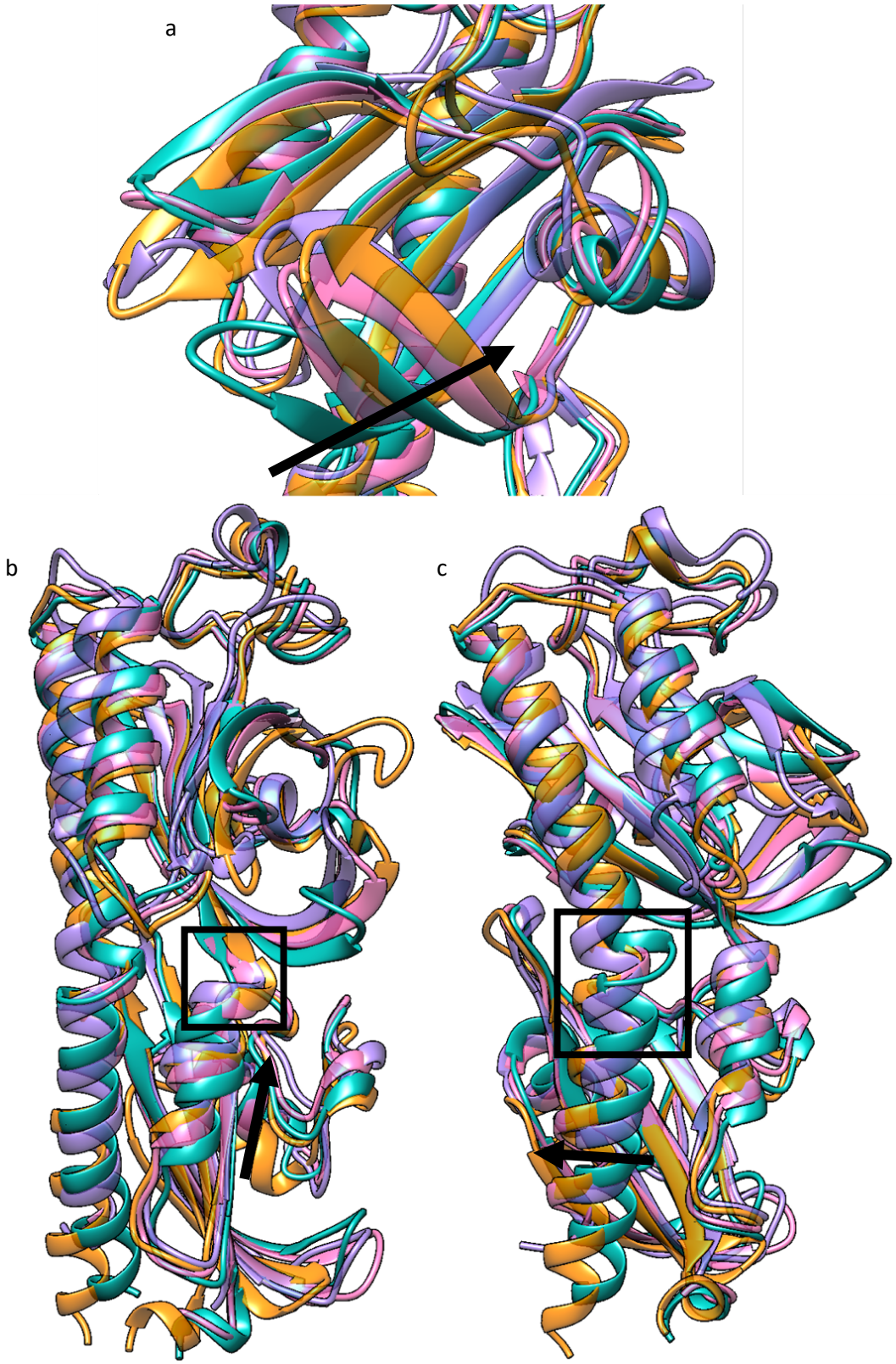


Figure 4.43 Structural comparison of 'apo' (Bd1081$^{S\text{-}LBD}$) and bound (Bd1081^{LBD}, Bd2503^{LBD} and Bd2504^{LBD}) forms of dCache domain (In all models, Bd1081$^{S\text{-}LBD}$ – teal, Bd1081^{LBD} – pink, Bd2503^{LBD} – purple, Bd2504^{LBD} – orange) **(a)** Close up of LBP with arrow showing movement of β-sheet upon ligand binding. **(b)** All dCache monomers aligned with arrow showing that the α-helix connecting the two lobes of the domain moves towards the distal lobe, with the change in helix wrapping highlighted by a black box. **(c)** 90° rotation from (b) with arrow showing movement of the backbone transmembrane region sideways, which allows for the formation of a continuous α-helix removing the kink seen midway in the apo structure (highlighted by the black box).

4.3.5 Working Hypotheses

As previously stated, the current theory of *Bdellovibrio* hunting is that it occurs by taking advantage of hydrodynamic forces to keep it in an area of high prey density to increase the chances of prey collision. This likely does aid in the hunting of planktonic prey, where the challenge of tracking prey is increased by prey motility. Thus *Bdellovibrio* does not have to move to where the prey is, but where it is going to be, making the likelihood of *Bdellovibrio* chemotactically hunting planktonic prey unlikely (Jashnsaz et al., 2017). However, our study into dCache domain specificity has revealed that their ligands are not only a part of the universally essential polyamine and QAC family, but cadaverine and putrescine are also Gram-negative centric (Pugin et al., 2017). Based on the prior study of a Bd2503 homologue knockout leading to decreased biofilm predation (Lambert et al., 2006) coupled with the prey centric nature of some of the ligands, two new working hypotheses were formed for biofilm predation. It is conjectured that future chemotaxis assays will support the role of Bd2503 in prey chemotaxis, and by extension the Cache family, where biofilm predation is not only influenced by hydrodynamic forces but also nuanced by chemotactic information.

4.3.5.1 The 'Chum in the Water' Hypothesis

The first hypothesis presented in this thesis is based on the polyamine family found to be ligands for Bd2503 and Bd2504, with a possibility that Bd1081 may also utilise QAC sensory information in this manner. This hypothesis is that either hydrodynamics leads to the location of an initial prey cell, which results in lysis following active predation, or high cell density causes lysis of some cells within the colony. Both of these would result in the release of polyamines into the environment and this in turn would trigger recruitment of more *Bdellovibrio* to the local area via chemotaxis. The recruited *Bdellovibrio* would then proceed to progress through multiple predation cycles in the localized area of the biofilm (Figure 4.44). Initial prey cell lysis resulting in biofilm predation would only occur if the

initial prey was a part of a biofilm. The similarities between this and the general perception of how sharks hunt prey has led this predation strategy being denoted as the 'Chum in the Water' Hypothesis.

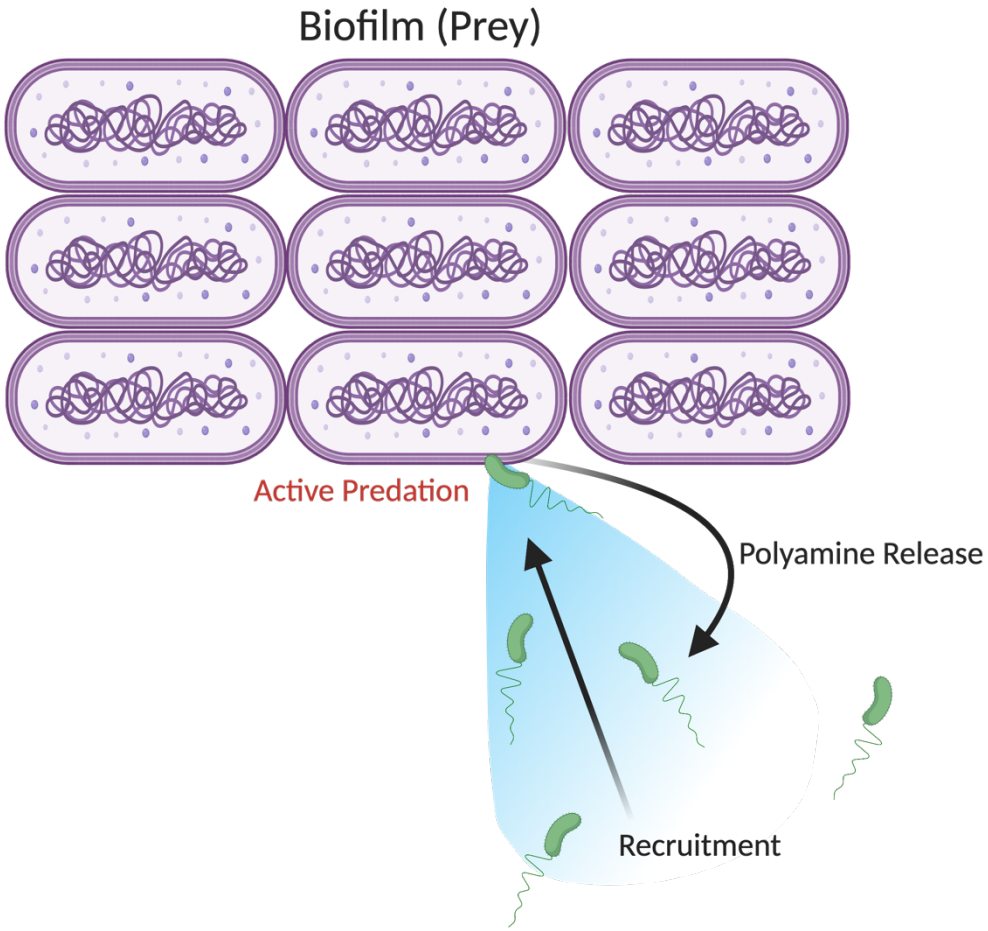


Figure 4.44 Diagram showing the first working hypothesis for chemotaxis influencing biofilm predation, 'Chum in the Water', in which the active predation of prey causes the release of polyamine, and possibly QACs, into the environment and consequently the recruitment of more *Bdellovibrios* to the site of predation. Created with BioRender.com

In the initial chemotactic experiments it was believed that *Bdellovibrio* was attracted to lysed prey cells because of the very high prey density, rather than to the whole prey cells because no chemotactic response could be elicited below 10^8 cells/ml (Jashnsaz et al., 2017; LaMarre et al., 1977; Varon and Zeigler, 1978). This would support this hypothesis, that chemotaxis is utilised for biofilm predation triggered by the lysis of an initial prey cell, as it would indicate that the chemotactic signal is located within the prey cell rather than being excreted/secreted from the prey. Secondary support for this comes from *Bdellovibrio* being shown to be chemotactically attracted towards an isolated extract of

non-prey species *Saccharomyces cerevisiae* (Straley and Conti, 1974). As with prey, yeast also utilises and synthesises cadaverine and putrescine (Valdés-Santiago and Ruiz-Herrera, 2014; Walters and Cowley, 1996), again suggesting the chemo-attractant that *Bdellovibrio* is responding to are these polyamines released upon lysis of the cells. Both of these previous studies would suggest a positive/attractant chemotactic effect from cadaverine and putrescine as noted in the presented theory.

4.3.5.2 The 'Lions at the Watering Hole' Hypothesis

The second hypothesis presented is based on the ligands identified as binding to Bd3092, acetate and propanoic acid, and Bd1081, choline. Both small organic acids and choline are essential carbon sources for bacteria and known ligands of chemoreceptors (Leone et al., 2015; Webb et al., 2017; Wolfe, 2005). Prey are likely to actively seek out and form biofilms in areas with high concentrations of these small molecules (García et al., 2015; Shrestha et al., 2018). This hypothesis proposes that *Bdellovibrio* follow these same small molecules to locate areas likely to be high in prey concentration (Figure 4.45). Once within these areas, as with the 'Chum in the Water' mechanism, *Bdellovibrio* would be able to use hydrodynamic forces to find individual prey (Jashnsaz et al., 2017). The use of an environmental resource to find prey is reminiscent of lions seeking out prey at a resource they need, like a watering hole, this has led to this predation strategy being named the 'Lions at the Watering Hole' Hypothesis.

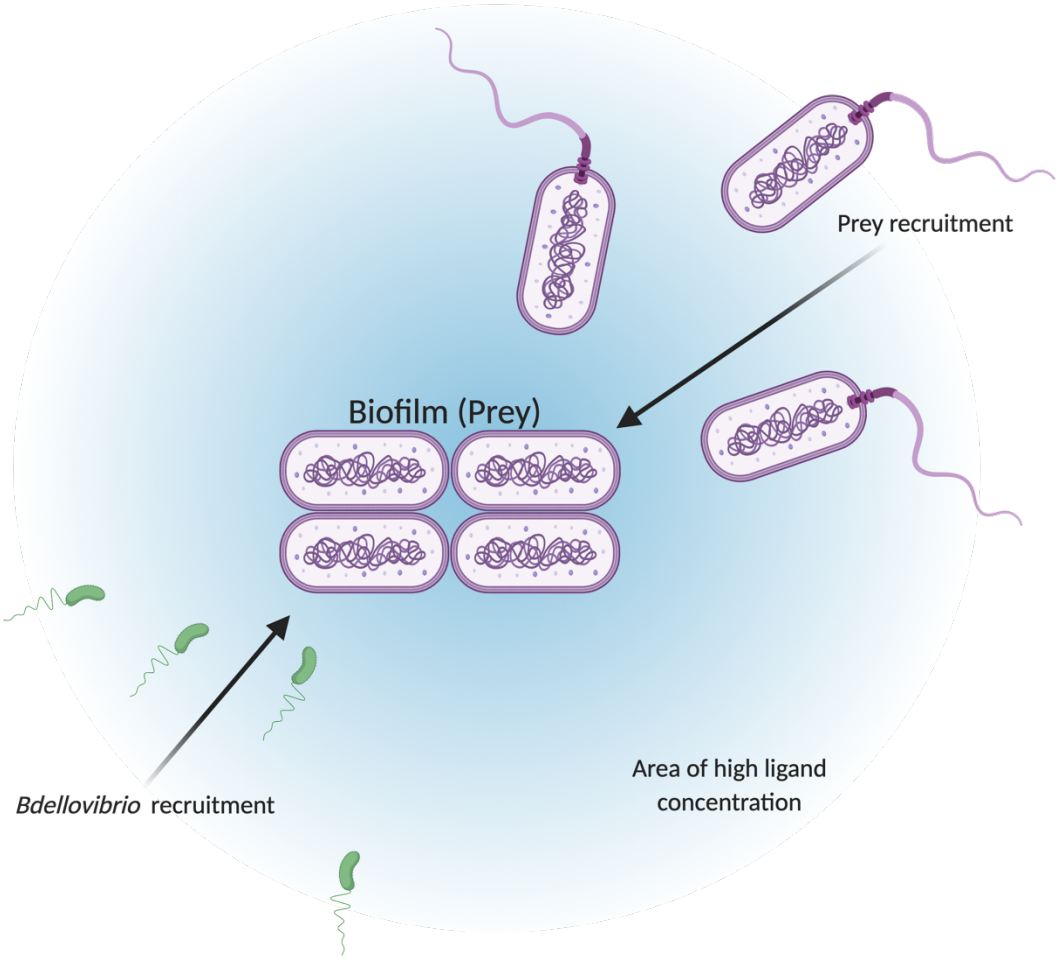


Figure 4.45 Diagram showing second working hypothesis for chemotaxis influencing biofilm predation, 'Lions at the Watering Hole', in which *Bdellovibrio* move towards an area high in ubiquitous nutrients in order to increase the likelihood of finding prey. Created with BioRender.com

4.3.6 Bd3126 – Known Fold, Novel Use

As previously mentioned (Chapter 1.3.1), Bd3126 is thought to work with Bd3125, a degenerative GGDEF domain containing REC protein, to create a standard two component system, similar to the GGDEF-REC protein system used by the Wsp pathway of *Pseudomonas aeruginosa* (Malone et al., 2007). However unlike with this pathway, the response regulator protein, Bd3125, is thought to be activated by two signals; firstly by binding c-di-GMP produced by Bd0742 (Meek et al., 2019), and secondly by phosphorylation of the REC domain by Bd3126. The activated CheY-like Bd3125 is then thought to activate a pathway that regulates invasion speed (Hobley et al., 2012). Although the Bd3126/Bd3125 pathway is predicted to be less complex than the Wsp pathway, producing a 'digital'

rather than 'analogue' signal, the identified LBD within Bd3126 is more complex than the 4HB normally associated with c-di-GMP signalling. The inclusion of the dCache domain within this pathway suggests an increased ligand specificity compared to the broad range of ligands accepted by 4HBs, which in turn should allow *Bdellovibrio* increased invasion control.

As with the other dCache domain proteins within *Bdellovibrio*, the crystallisation of Bd3126^{LBD} without its ligand was unsuccessful. This lead to both bioinformatic and experimental studies in order to predict the ligand family Bd3126^{LBD} binds. Sequence comparison of Bd3126^{LBD} to other dCache showed receptors that bound polyamines, amino acids and QACs had the highest sequence identity (Table 4.8).

PDB Code	Organism	Protein	Ligand Family	Identity (%)
6F9G	*Pseudomonas putida*	McpU	Polyamines	20.7
-	*Bdellovibrio bacteriovorus HD100*	Bd2503	Polyamines	19.2
3C8C	*Vibrio cholerae*	Mcp_N	Amino acids	17.2
5AVF	*Vibrio cholerae*	Mlp37	Taurine	16.8
-	*Bdellovibrio bacteriovorus HD100*	Bd1081	QACs	16.7

Table 4.8 Table of dCache proteins with the highest sequence identity to Bd3126 with a Known Structures.

Considering that the predicted LBP of Bd3126 is likely to carry an overall positive charge, binding to polyamines was thought to be unlikely, with preliminary ITC with cadaverine supporting this by demonstrating a negative result. Although amino acids were shown via DSF to stabilise Bd3126^{LBD}, they did not improve crystallisation. Although alignment between Bd1081^{LBD} and Bd3126^{LBD} did not reveal strong sequence identity, preliminary ITC data did show binding affinity between choline and Bd3126^{LBD}, and supplementing Bd3126^{LBD} with both TMAO and choline led to successful crystallisation. Bd3126 possesses significantly fewer aromatic residues than Bd1081, which are the amino acids traditionally involved in coordinating QAC binding. However, the overall positive charge of the LBD may compensate for this lack of aromatic residues, whilst also increasing selectivity to QACs that carry a constitutive negatively charged group. Nevertheless, if Bd3126 does sense QACs there is presently

not an obvious universal role for QACs in Gram-negative bacteria that suggests why this family of molecules would be sensed by *Bdellovibrio* upon initiation of invasion.

Chapter 5 – Attack Phase to Growth Phase Switch Transcription Factors in *Bdellovibrio bacteriovorus*

5.1 Introduction

Bdellovibrio bacteriovorus HD100 possesses a family of 18 proposed transcription factors (TFs), comprised of a helix-turn-helix (HTH) DNA binding domain and DUF4423 (domain of unknown function). These 18 putative TFs are located throughout the genome, and similar to the diversification of chemosensory proteins, it is thought that the high copy number could help protect against loss/mutation of the family, indicating a dependency on this family (Lambert et al., 2008). This multiplicity, paired with the upregulation of a third of the family during the AP/GP lifestyle switch (Lambert et al., 2010), has led to the hypothesis that these proteins are key to the predatory lifestyle.

5.1.1 DUF4423

Across all genetic insistences of this domain it almost always occurs as an isolated domain, with the second most common architecture, where it is preceded by a HTH domain, only occurring about 6% of the time (Figure 5.1). The DUF4423 domain is almost exclusively found in bacteria that shift between periods of feast and famine. This supports the current working hypothesis that these TFs are involved in lifestyle switches within *Bdellovibrio*. Seventy percent of the genomes predicted to encode proteins with DUF4423 have a biphasic lifestyle, 70% of which are predatory, and 8% dimorphic (Blum et al., 2021).

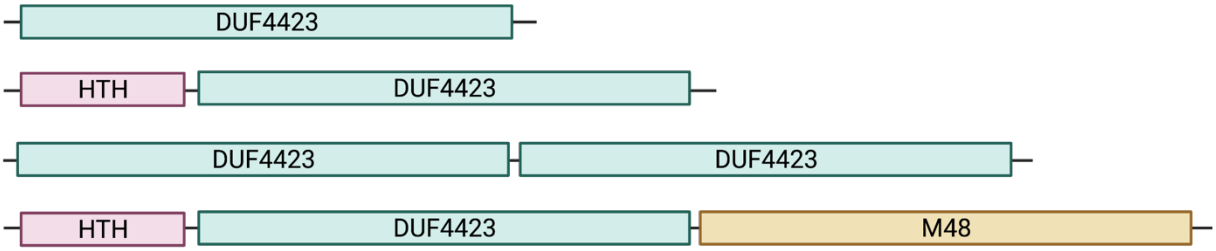


Figure 5.1 Schematic of different domain architectures predicted (Blum et al., 2021). Created with BioRender.com

5.1.1.1 DUF4423 in Non-predators

The most prevalent occurrence of DUF4423 is within *Fibrobacter succinogenes*, an anaerobic Gram-negative bacterium, found in cattle rumen. *F. succinogenes* has 38 DUF4423 proteins all of which are predicted to contain the DUF4423 domain alone, and is 1 of 3 bacteria that are both predicted to have the domain and not utilise a biphasic or dimorphic lifecycle. *F. succinogenes* requires direct binding, via its outer membrane, to cellulose in order for digestion to take place, enabling it to grow and replicate. Although *F. succinogenes* is not a predator, the necessity to directly attach to its substrate, via type IV pili (Jun et al., 2007), to take up nutrients is reminiscent of epibiotic predators like *Vampirococcus* (Pérez et al., 2016), and the release of enzymes for degradation is similar to wolfpack *Myxococcus xanthus* (Martin, 2002). This attachment and release, to move on to the next substrate, means *F. succinogenes* often cycles between feast and famine. Non-predatory *Polyangium fumosum and Sorangium cellulosum* are both *Myxococcales* that form fruiting bodies (Knauber et al., 2008; Mohr et al., 2012), and are predicted to have the HTH-DUF4423 variant, with *S. cellulosum* being the only bacteria predicted to have a double HTH domain alongside DUF4423 in both of its homologous TF proteins. Another non-predatory *Myxobacterium*, *Sandaracinus amylolyticus,* also forms fruiting bodies under starvation stress (Sharma et al., 2016) and has both the DUF4423 only and HTH-DUF4423 variants of these proteins, whereas dimorphic *Planctomyes bekefii* (Wiegand et al., 2018) has only the DUF4423 type.

5.1.1.2 DUF4423 in predators

Outside of *F. succinogenes*, BALOs are the only bacteria to contain a significant number of DUF4423 proteins, with the majority possessing it as both a singular domain and a HTH-DUF4423 variant. *Bradymonadales bacterium V1718* and *Persicimonas caeni* (Mu et al., 2020; Wang et al., 2020) are the only two examples to have just DUF4423 only type proteins. BALO *Pseudobacteriovorax antillogorgiicola* (Bonfiglio et al., 2020) is the most prevalent predatory carrier of DUF4423 proteins with 28 copies, 2/3 of which are DUF4423 only and 1/3 HTH-DUF4423. *P. antillogorgiicola* also has the only detectable occurrence of a double DUF4423 domain protein. As already discussed, *Bdellovibrio bacteriovorus HD100* possesses 18 proteins containing a DUF4423 motif, all of which are HTH-DUF4423, making it the only predatory bacterium predicted to exclusively use this family architecture; the reasons behind this are currently obscure.

5.1.2 Predicted Architecture

HHpred (Zimmermann et al., 2018) predicts structural similarities between DUF4423 and winged helix (WHTH) DNA binding domains, and has thus led to the theory that DUF4423 could be a DNA binding domain. Thus HTH:DUF4423 containing proteins putatively possess 2 separate DNA binding motifs. This, coupled with its prevalence in bacteria that undergo starvation stress and lifestyle changes, may indicate the involvement of DUF4423 in the control of genes that enable the switch between different lifecycle stages of *Bdellovibrio.* These proteins had already been labelled as putative TFs due to the presence of the N-terminal HTH domain, and the domain fusion with DUF4423 supports this predicted role in DNA binding. However, the (broader) rarity with which multiple DNA binding domains are found within a single protein (whichever domain they use) raises questions about why *Bdellovibrio* would need such a protein architecture in high prevalence.

Secondary structure and fold prediction software Xtalpred (Slabinski et al., 2007) (Figure 5.2a), predicted these TFs to have a fairly typical 3 helix HTH domain, consisting of three helices at the N-terminus of the protein. The latter two of these helices (H2 and H3) make up the actual HTH motif and are connected by a short turn. Typically the whole motif consists of about 20 amino acids, but is instead usually only 19 residues long in these DUF4423-associated TFs. The second helix (H3) in the motif usually interacts with the major groove of the DNA, via the amino acid sidechains forming hydrogen bonds and hydrophobic interactions, whilst the first helix (H2) stabilises this interaction (Matthews et al., 1982) (Figure 5.2b). As previously stated, the predicted secondary structure of DUF4423 shares homology to the WHTH motif, and is likely to contain a WHTH-like motif within the domain. A WHTH is an extended form of the basic 3 helix HTH domain where the 2 wings are formed by 3-4 β-strands (Figure 5.2c). Similarly, like the HTH mode of recognition, the major groove of DNA is recognized by the third helix (H3), with the second helix (H2) and wings (W1 and W2) making stabilising contact with the minor groove or the backbone of the DNA (Gajiwala and Burley, 2000). Unlike the prototypical WHTH motif which is H1-S1-S2-H2-H3-S3-S4 (Donaldson et al., 1996), the DUF4423 domain within Bdellovibrio seems to lack the 2 β-strands that are typically between H1 and H2 of the motif, and as such it has been denoted as a WHTH-like motif.

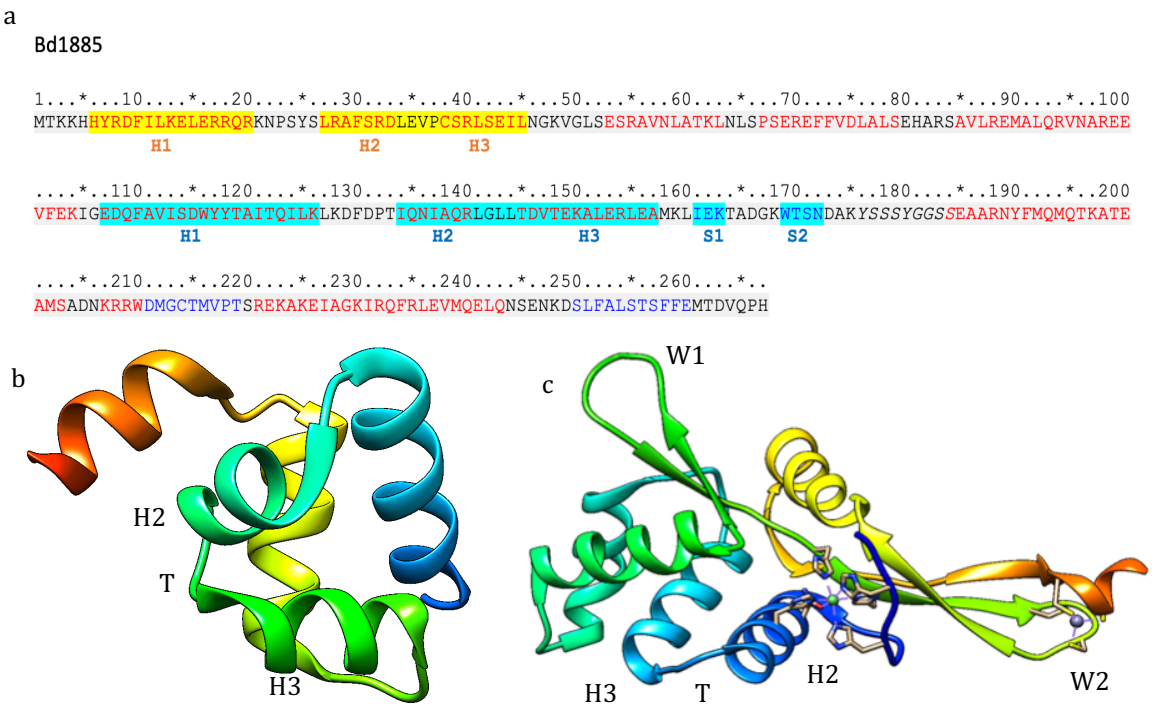


Figure 5.2 In all figures; H = helix, S = Strand, T = turn, W = Wing.**(a)** Sequence of the archetypical TF, Bd1885, showing Xtalpred (Slabinski et al., 2007) prediction of the secondary structure elements, annotated to show the likely positions of the α/β regions of the two motifs. The HTH motif is highlighted in yellow, the WHTH-like motif in DUF4423 highlighted in blue. **(b)** An example of a helix-turn-helix structure, adapted from 5J2Y (Zhao et al., 2017). **(c)** An example of winged-helix structure, adapted from 4I7H (Makthal et al., 2013).

5.1.3 Effectors

Eleven of the eighteen TFs in HD100 have a neighbouring gene that is conserved, which possesses a secretion signal and thus could represent an effector-like function. Unlike DUF4423, these effectors are found exclusively within BALOs, and have no predicted homology to any characterised domains, which was ascertained by no significant matches from HHPred analysis (Zimmermann et al., 2018). However, there is one variant of DUF4423 proteins (which has been predicted in an uncharacterised bacterium), that has a M48 metalloprotease fused to the end of a HTH-DUF4423 protein, (Uniprot ID: A3B70_06150). The poor effector HHpred matches identified were often characterised as hydrolases, and possessed a M48 domain, for example; Q74D82 in *Geobacter sulfurreducens* (3C37 (Kuzin et al., 2008)) with a score of 28.8, HtpX in *Vibrio parahaemolyticus* (3CQB (Wu et al., 2008)) with a score of

22.36, and BepA in *E. coli* (6SAR (Bryant et al., 2019)) with a score of 27.97. This has led to the idea that the effectors may function as a novel metalloprotease.

Metalloproteases are a diverse group of proteins that all cleave peptide bonds through the use of a metal ion, which is commonly zinc. There are over 70 families of metalloproteases, with the majority using the 5 amino acid motif HExxH to bind the metal ion (Rawlings and Barrett, 1995). Cleavage of peptide bonds occurs in a single step via a metal ion co-factor, bound by two histidines, a glutamate (functioning as a general base/acid), and a solvent molecule (Cerdà-Costa and Gomis-Rüth, 2014).

Although the HExxH motif is the most common motif for metal ion binding in metalloproteases, there is some variation to this, the most well-studied families being M14, M15, M16, M17, and M24. M14 proteins, from the Carboxypeptidase A family, utilise a distant histidine as the secondary histidine involved in metal binding. This forms a divergent HEx_nH motif which is used in the binding of a metal ion, identified in human carboxypeptidase A as H69, E72 and distant H196. This family also uses a distant glutamine, E270, as the catalytic residue instead of the one within the HExxH motif (Rawlings and Barrett, 1995). Structure determination of a member of the M15 subgroup, Zinc D-Ala-D-Ala carboxypeptidase from *Streptomyces* (1LBU (Charlier et al., 2006)), showed that it utilised 3 histidines to bind zinc, H153, H194 and H196. M16 members, *e.g.* Pitrilysins, contain an inverse HExxH domain, HxxEH (Rawlings and Barrett, 1995). M17, the Leucyl aminopeptidase family, bind two co-catalytic zinc ions using conserved residues identified in the crystal structure of bovine lens leucine aminopeptidase (1BPN (Kim and Lipscomb, 1993)) as D255, D332 and E334 which bind one ion, whilst K250, D255, D273 and E334 bind the second. Lastly, the M24 subgroup, Methionyl aminopeptidases, bind two cobalt ions, using residues D97, D108, H177, E204 and E235, as identified via the structure of Methionine aminopeptidase from *E. coli* (1MAT (Roderick and Matthews, 1993)). Although the effectors in *Bdellovibrio* do not fit well with any of these families motifs, this does demonstrate a vast

number of proteins that bind metal ions and function as metalloproteases, without using an absolutely conserved HExxH motif.

Seven of these effectors have an interesting 5 nucleotide overlap between the stop codon of the TF and the start of the effector, XX-ATG-AAX (Figure 5.3). This could allow for co-translation of the two proteins, the outcome similar to the M48 fusion protein outlined above (with the added complication that the TF is cytoplasmic and the effector presumably not). Within the *Feline Calicivirus* genome there is the same 5 nt overlap between the VP1 and VP2 capsid proteins. It was found that single and double basepair mutations of the TGA region still allowed VP2 expression, as long as there was termination of VP1 synthesis. In total this demonstrated that VP2 synthesis was dependent on VP1 synthesis (Luttermann and Meyers, 2007). This could suggest a similar expression dependency between the *Bdellovibrio* TFs and effectors, and produce a similar output to the HTH-DUF4423-M48 fused protein. Metallopeptidases generally function as non-specific general peptidases, and therefore are tightly regulated (Cerdà-Costa and Gomis-Rüth, 2014). If these effectors are metallopeptidases, this tight regulation would continue to hold true in *Bdellovibrio* as all are either in an operon or overlapping with one of the proposed DUF4423 TFs.

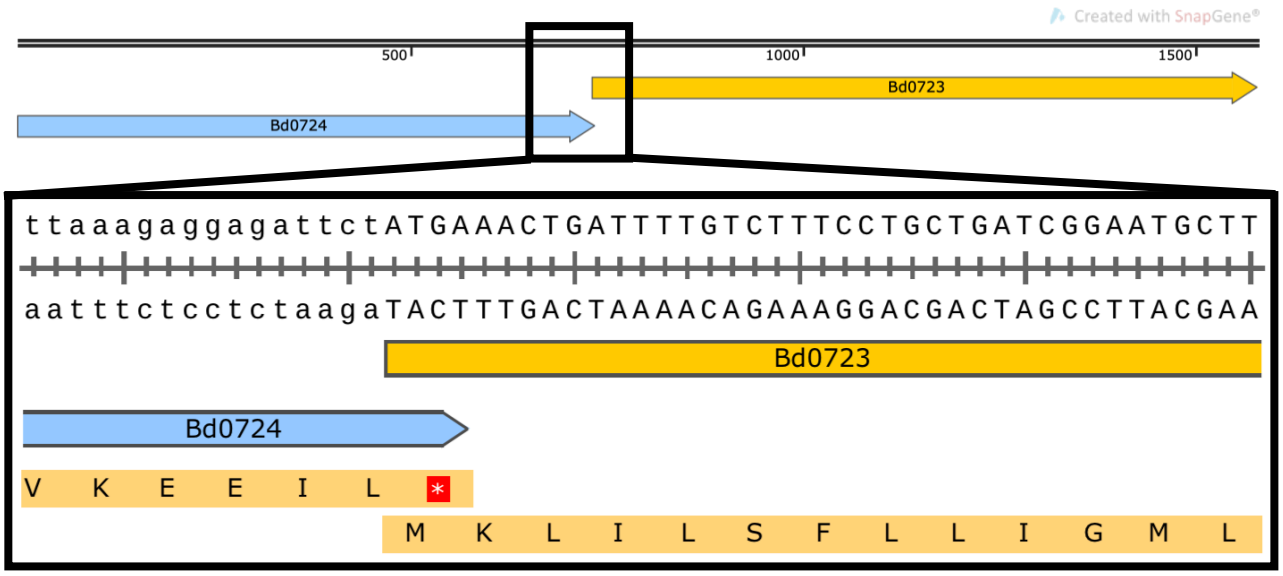


Figure 5.3 Diagram depicting gene overlap using Bd0724/Bd0723 as an example.

5.2 Results

5.2.1 Gene Expression and Purification

5.2.1.1 Cloning of Transcription Factor constructs

The 18 member TF family, were sub-divided into 5 subfamilies by sequence homology using a phylogenic tree (Figure 5.4), of which only 3 subfamilies had genes previously cloned by our group (*bd0089*, *bd1905*, *bd1172*, and *bd1885)*.

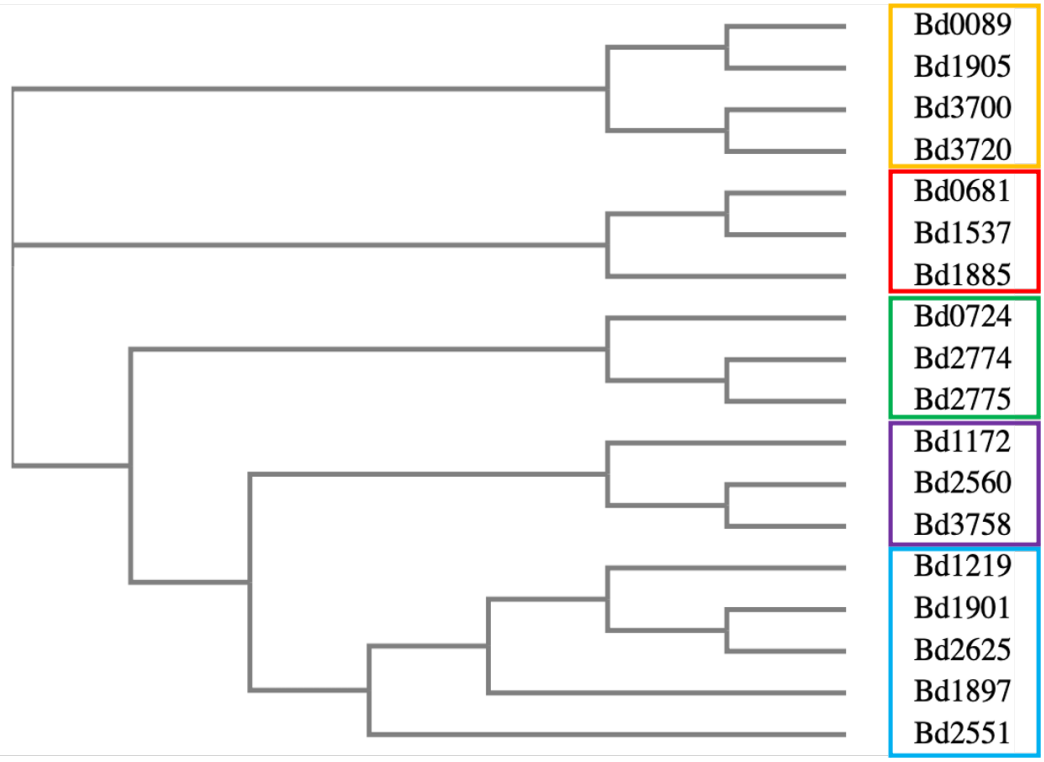


Figure 5.4 Phylogenic tree created by Tcoffee (Notredame et al., 2000) showing how 18 predicted transcription factors split into 'subfamilies', coloured boxes denote 'subfamilies'.

To try and ensure that a pure stable protein was obtained for at least one member of each subfamily, multiple genes from each were targeted for cloning. This also ensured both TFs with and without a downstream effector were obtained. Therefore a further 6 genes; *bd0681*, *bd1897*, *bd2625*, *bd2774, bd3720*, and *bd3758*, were targeted for cloning. Of the 6 new genes, 3 were successfully cloned, *bd1897*, *bd2625*, *bd3720*, meaning all but one subfamily was successfully cloned into BL21 expression strains.

Due to the low expression levels of Bd0089, even when fused with GFP to enable stabilisation, it was not investigated further (Figure 5.5).

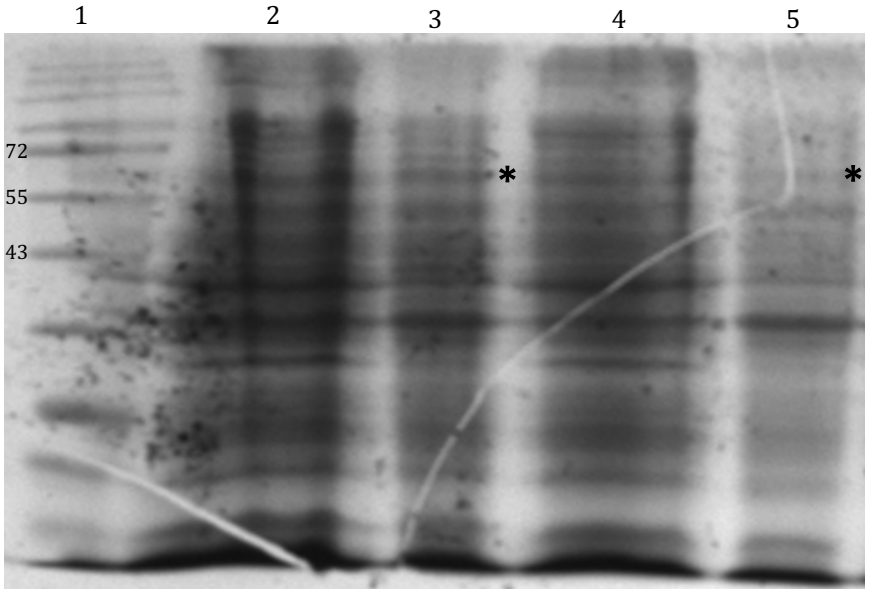


Figure 5.5 SDS-PAGE expression gel Bd0089+GFP showing no expression post induction in either TB (lane 2 pre and lane 3 post induction) or AI with sugar mix N (lane 4 pre and lane 5 post induction). * indicates the expected position of the target protein.

5.2.1.2 Purification of Bd1905

An expression trial of full length Bd1905 in TB and AI media, induced with 1 mM IPTG or sugar mix N respectively, showed high expression of Bd1905 in both media (Figure 5.6a). However, IMAC purification resulted in a very low yield and impure product (Figure 5.6b). Therefore, Bd1905 was rejected for further study.

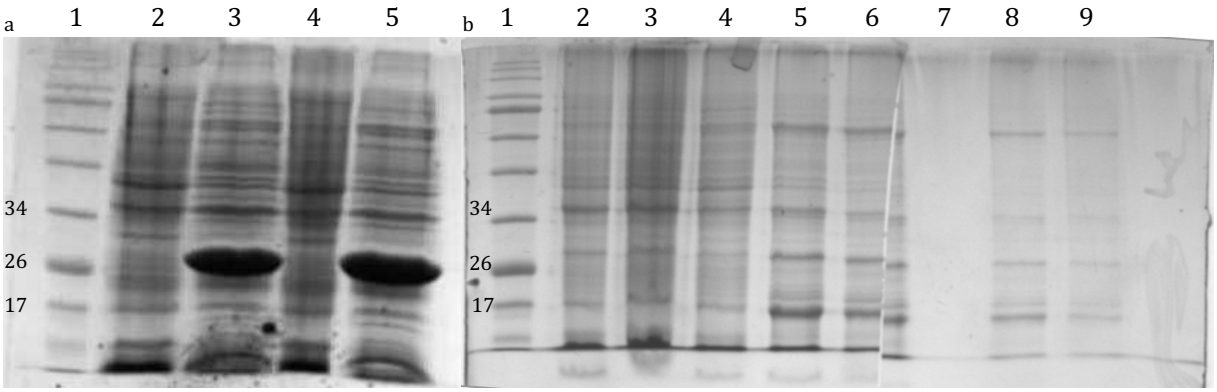


Figure 5.6 SDS-PAGE gel of Bd1905 expression and IMAC purification. **(a)** Expression gel showing expression post induction in both TB (lane 2 pre and lane 3 post induction) and AI with sugar mix N (lane 4 pre and lane 5 post induction). **(b)** Gel of IMAC purification, showing no/little soluble Bd1905.

5.2.1.3 Purification of Bd1897

Bd1897 was expressed in AI media with sugar mix N and purified with a high yield using nickel affinity chromatography, eluted with 100–500 mM Imidazole (Figure 5.7). Due to protein instability thought to be caused by 2 unpaired cysteines in the DUF4423 domain, 2.5 mM β-mercaptoethanol was added to Bd1897 elution fractions, before dialysis of the purified protein into 20 mM Hepes pH 6.0, 500 mM NaCl, 2.5 mM β-mercaptoethanol. Although this went some way to achieving adequate protein stabilisation, Bd1897 still readily aggregated upon concentration and as such was never successfully ran on SEC. Although SEC could not be performed, and in spite of some contamination, crystallisation of Bd1897 was attempted but was ultimately unsuccessful.

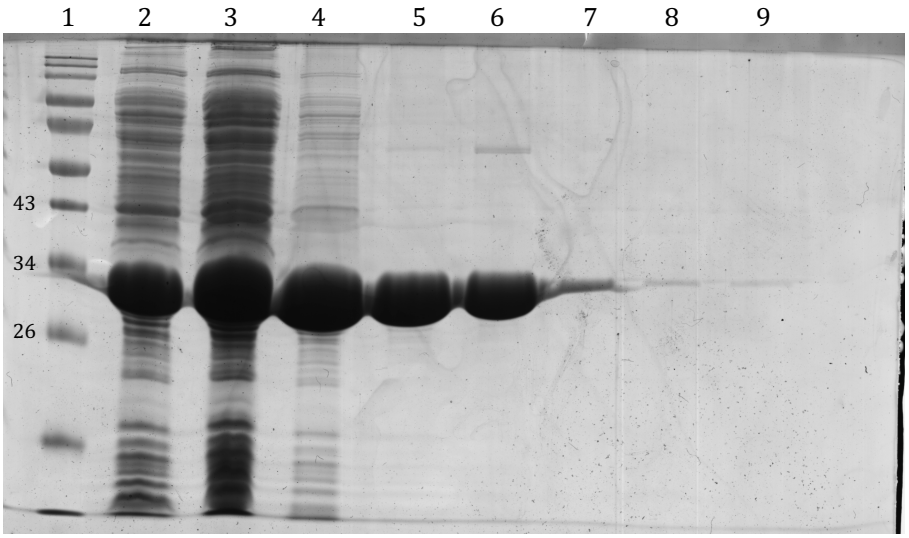


Figure 5.7 SDS-PAGE gel of IMAC purification of Bd1897

5.2.1.4 Purification of Bd1172

Bd1172 was expressed in AI media with sugar mix N and purified with a low pure yield using nickel affinity chromatography, eluted with 500 mM Imidazole (Figure 5.8). Due to protein instability thought to be partially caused by an unpaired cysteine in the DUF4423 domain, 2.5 mM β-mercaptoethanol was added to Bd1172 elution fractions, before dialysis of the purified protein into 20 mM Bicine pH 8.0, 500 mM NaCl and 2.5 mM β-mercaptoethanol. Although this went some way to achieving adequate protein stabilisation, Bd1897 still readily aggerated upon concentration and as such was never successfully run on SEC. Although SEC could not be performed, and in spite of some contamination, crystallisation of Bd1172 was attempted but was ultimately unsuccessful.

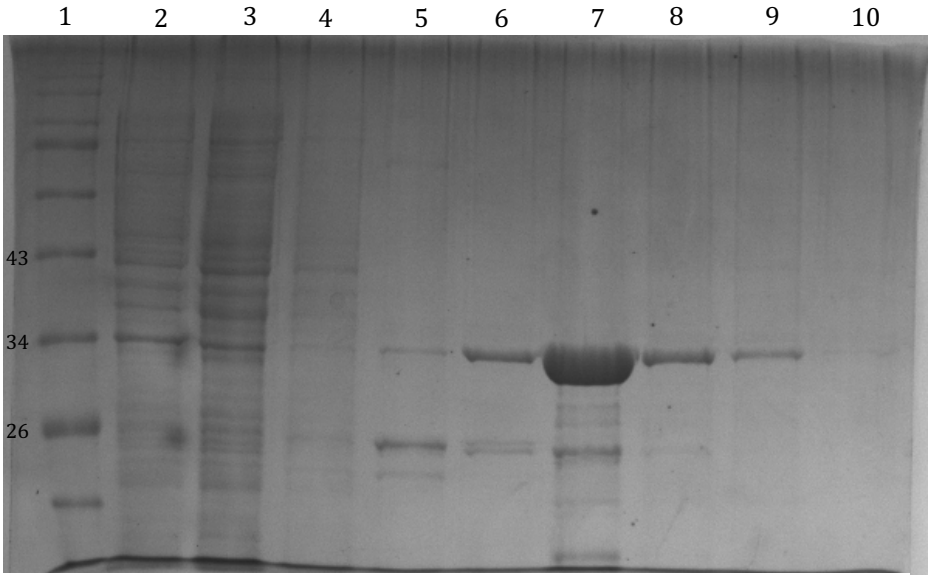


Figure 5.8 SDS-PAGE gel of IMAC purification of Bd1172

5.2.1.5 Purification of Bd1885

Bd1885 full length was expressed in AI media with sugar mix E and purified with high pure yields using nickel affinity chromatography and eluted with 100 mM Imidazole (Figure 5.9). However, Bd1885 had the same stability issues as the other transcription factors, precipitating about 15 mins after elution. This was predicted to be due to an unpaired cysteine in both the HTH and DUF4423 domains, and buffer trials with β-mercaptoethanol seemed to reduce this precipitation and help stabilise the protein. Subsequently, 2.5 mM β-mercaptoethanol was added to eluted Bd1885, before dialysis into 20 mM Bicine pH 8.0, 500 mM NaCl, and 2.5 mM β-mercaptoethanol. However, like most other members of the transcription factor family, Bd1885 would still precipitate upon concentrating and could not be ran on SEC. Although size exclusion could not be performed, crystallisation of Bd1885 was attempted but was unsuccessful.

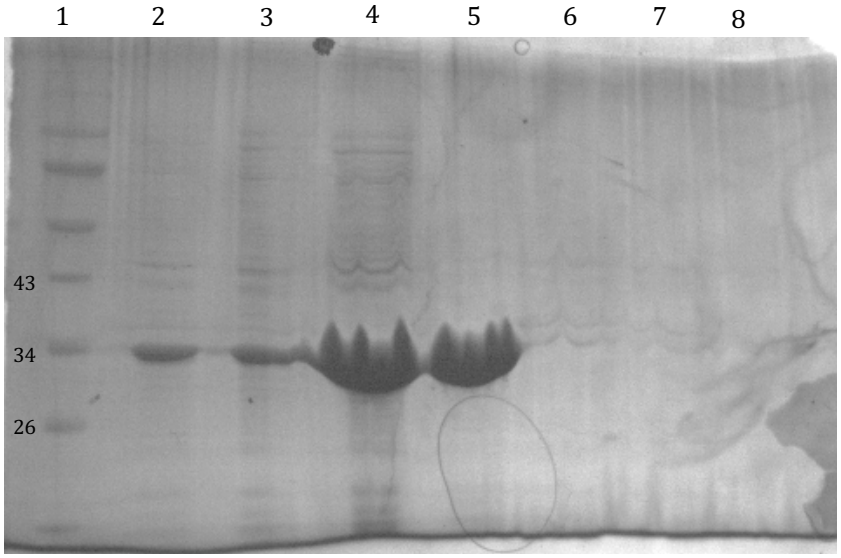


Figure 5.9 SDS-PAGE gel of IMAC purification of Bd1885

5.2.1.6 Purification of Bd2625

Full length Bd2625 was expressed in AI media with sugar mix N and purified using nickel affinity chromatography and eluted with 500 mM Imidazole (Figure 5.10). Similarly to the rest of the transcription factor family, an unpaired cysteine in the HTH domain resulted in stability issues. Therefore, 2.5 mM β-mercaptoethanol was added to eluted fractions of Bd2625, before dialysis into 20 mM Tris pH 8.5, 500 mM NaCl and 2.5 mM β-mercaptoethanol. This stabilised Bd2625 significantly better than previous members of the family and enabled successful concentration and SEC of Bd2625. SEC with a HiLoad 26/60 Superdex 75 pg yielded two peaks of similar height, higher molecular weight contaminants coming off as one peak at 120 ml and a monomer peak at 145 ml. SEC also removed lower molecular weight contaminants, visible as multiple small peaks on the UV trace.

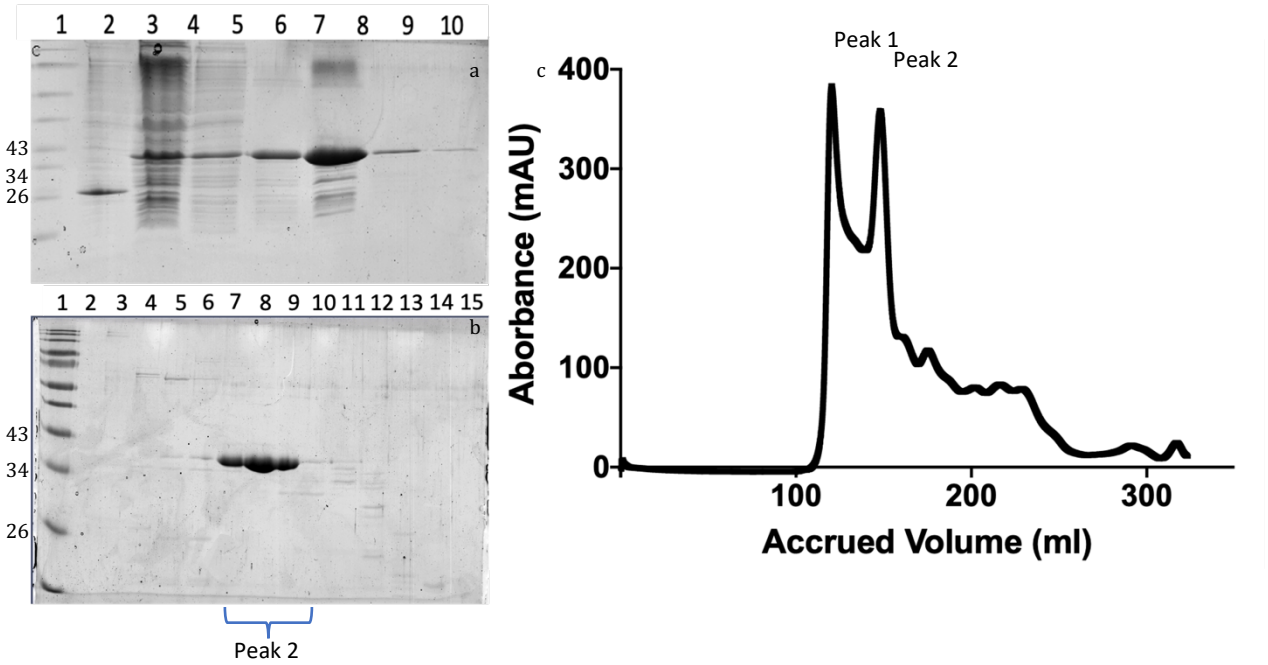


Figure 5.10 (a) SDS-PAGE gel of Bd2625 IMAC purification. **(b)** SDS-PAGE gel of Bd2625 post SEC purification. **(c)** UV trace of Bd2625 SEC, showing monomeric Bd2625 (peak 2) being separated from contaminant proteins.

5.2.1.7 Purification of Bd3720

Bd3720 was expressed in AI media with sugar mix N and purified using nickel affinity chromatography and eluted with 500 mM Imidazole (Figure 5.11). An unpaired cysteine in both the HTH and DUF4423 domains caused instability of Bd3720, which was ameliorated with the addition of 2.5 mM β-mercaptoethanol to eluted fractions of Bd3720, before dialysis into 20 mM Bicine pH 8.0, 500 mM NaCl and 2.5 mM β-mercaptoethanol. Similarly to Bd2625, this stabilised Bd3720 enough for concentration and SEC to be performed. SEC of Bd3720 was performed using a HiLoad 16/60 Superdex 200 pg as use of a HiLoad 26/60 Superdex 75 pg suggested Bd3720 formed a tetramer, which could not be fully separated from the void volume of the HiLoad 26/60 Superdex 75 pg column. Three peaks eluted from the HiLoad 16/60 Superdex 200 pg column, a tetramer peak at 60 ml, and two slightly smaller contaminant peaks at 115 and 120 ml. Post SEC crystallisation attempts of Bd3720 were ultimately unsuccessful.

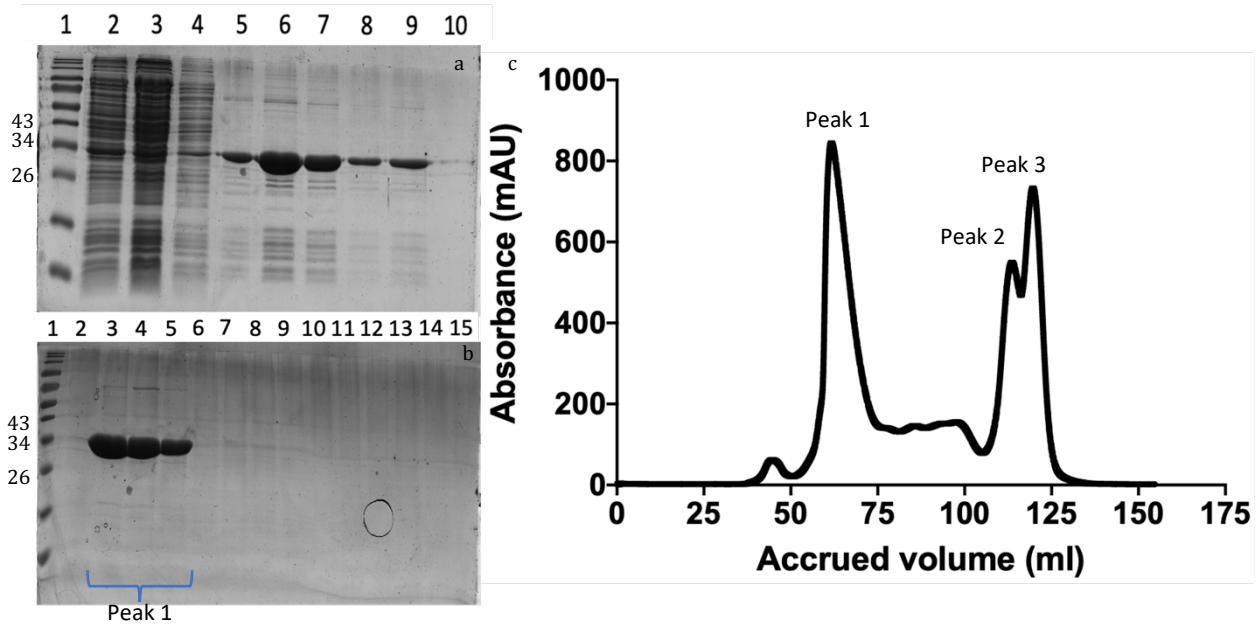


Figure 5.11 (a) SDS-PAGE gel of Bd3720 IMAC purification. **(b)** SDS-PAGE gel of Bd3720 post SEC purification. **(c)** UV trace of Bd3720 SEC, showing tetrameric Bd3720 (peak 1) being separated from contaminant proteins.

5.2.2 Determination of Partial Bd2625 Structure

Crystals of Bd2625 were obtained in commercial crystal condition Morpheus D7; 0.1 M Sodium Hepes/Mops (acid) pH 7.5, 0.02M 1,6-Hexanediol, 0.02M 1-Butanol, 0.02M 1,2-Propanediol, 0.02M 2-Propanol, 0.02M 1,4-Butanediol, 0.02M 1,3-Propanediol, 20% Glycerol and 10% PEG 4000, at a protein concentration of 10 mg/ml and drop ratio of 1:1. This crystal screen was carried out by graduate student Charlotte Manning under my instruction. These initial crystals possessed good physical characteristics and did not require any optimisation:

Method	Sitting drop vapour diffusion
Temperature (°C)	18
Plate Type	96 Well CombiClover 4 chamber crystallization plates
Protein Concentration (mg/ml)	10
Reservoir Solution	0.1 M Sodium Hepes/Mops (acid) pH 7.5, 0.02M 1,6-Hexanediol, 0.02M 1-Butanol, 0.02M 1,2-Propanediol, 0.02M 2-Propanol, 0.02M 1,4-Butanediol, 0.02M 1,3-Propanediol, 20% Glycerol and 10% PEG 4000
Drop Volume (µl)	2
Drop Ratio	1:1
Volume of Reservoir (µl)	100
Crystallisation Morphology	Cuboid

Table 5.1 Crystallisation Conditions for Bd2625.

Crystals were innately cryoprotected, and thus crystals were only supplemented with 5 mM Potassium tetrachloroplatinate (II), and soaked for 1 hr, before being fished. The crystals were supplemented with heavy atoms as existing protein structures with either a HTH or WHTH had low sequence homology, and poor representation across the entire length of Bd2625. Diffraction data were obtained at Diamond Light Source beamline i03 to the following specifications (Table 5.2):

Diffraction Source	Diamond Light Source i03		
Cryoprotectant	None		
Derivative	5 mM Potassium tetrachloroplatinate(II)		
Wavelength (Å)	0.89		
Rotation of Crystal (°)	360		
Rotation per Image (°)	0.1		
Images Collected	3600		
Exposure per Image (sec)	0.05		
Transmission (%)	39.95		
Detector Resolution (Å)	2.7		
Space Group	P 3_1 2 1		
Cell Dimensions			
a, b, c (Å)	52.84	52.84	208.31
α, β, γ (°)	90	90	120
Resolution Range (Å)	2.30 – 44.70		
Wilson B-Factor	40.98		
	Overall	High	
Total Reflections	264787	8748	
Unique Reflections	15463	925	
R_{merge}	0.061	1.554	
R_{meas}	0.064	1.729	
$R_{p.i.m}$	0.020	0.738	
I/σI	21.6	1.1	
CC(1/2)	1.00	0.554	
Multiplicity	17.1	9.5	
Completeness	97.5	83.4	

Table 5.2 Diffraction dataset specifications for Bd2503 derivatised with Pt.

Once the diffraction data had been obtained (log file in Appendix VIII) experimental phasing using SAD/MAD Phenix AutoSol (Terwilliger et al., 2009) was performed to locate the supplemented platinum atoms. A search criteria based on 1 copy of Bd2625 with approximately 13 Pt ions (number estimated based on full occupancy of all histidine, methionine and cysteine residues in asymmetric unit), with the wavelength set to 0.89 Å, and f' = -7.322 e and f'' = 12.08 e. Autosol produced a map

with a FOM of 0.204, however in spite of this, a limited range of residues 92–149 were built by Autosol. Four copies of this region of Bd2625 were placed, and fit well within the density of the initial map (Figure 5.12). This revealed it was not one copy of full length Bd2625 as initially predicted, but 4 copies of a truncated domain of Bd2625. This truncation contained the predicted DNA recognition helices within the DUF4423 WHTH-like motif. Eight Pt ions were identified, 4 of these were in 2 supersites between 2 monomers in the asymmetric unit and 2 symmetry copies (Figure 5.13a). The other 4 Pt were in single occupancy sites, again coordinated between 2 copies of the monomer (Figure 5.13b). The model was then adjusted manually in COOT (Emsley et al., 2010) and refined using Phenix (Afonine et al., 2012) (Table 5.3).

Resolution (Å)	1.9
Number of Copies in Asymmetric Unit	4
R_{Work}	0.2429
R_{Free}	0.2793
R.M.S Deviations	
Bond Length (Å)	0.01
Bond Angles (°)	1.28
Ramachandran Plot	
Favoured Regions(%)	98.21
Allowed Regions(%)	1.79
Outliers (%)	0.00

Table 5.3 Final refinement statistics for Bd2625 diffraction dataset.

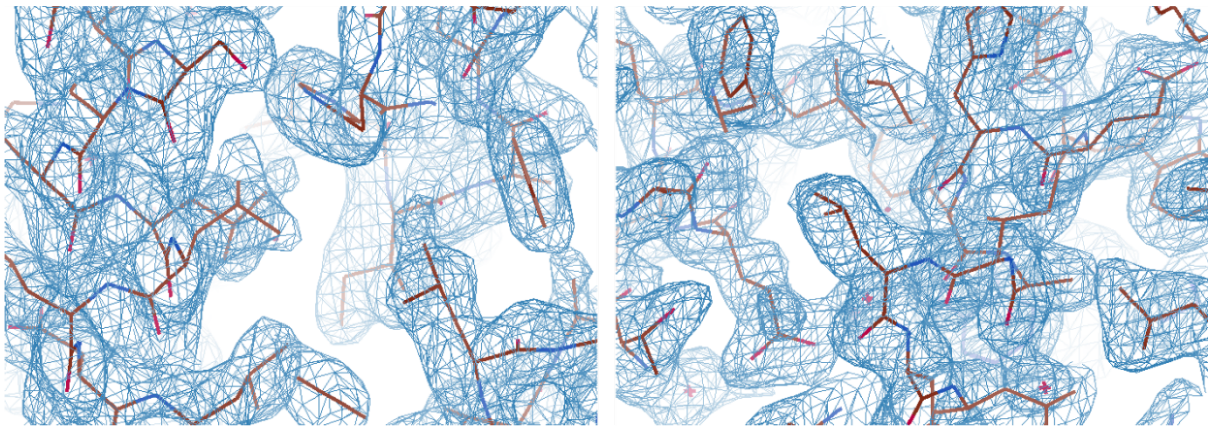


Figure 5.12 Initial 2FOFCWT PH2FOFCWT map (blue) and structure of Bd2625 structure (orange) that was built by Phenix AutoSol using Pt sites. Phenix AutoSol fit most of the truncated Bd2625 structure with good fit into the density.

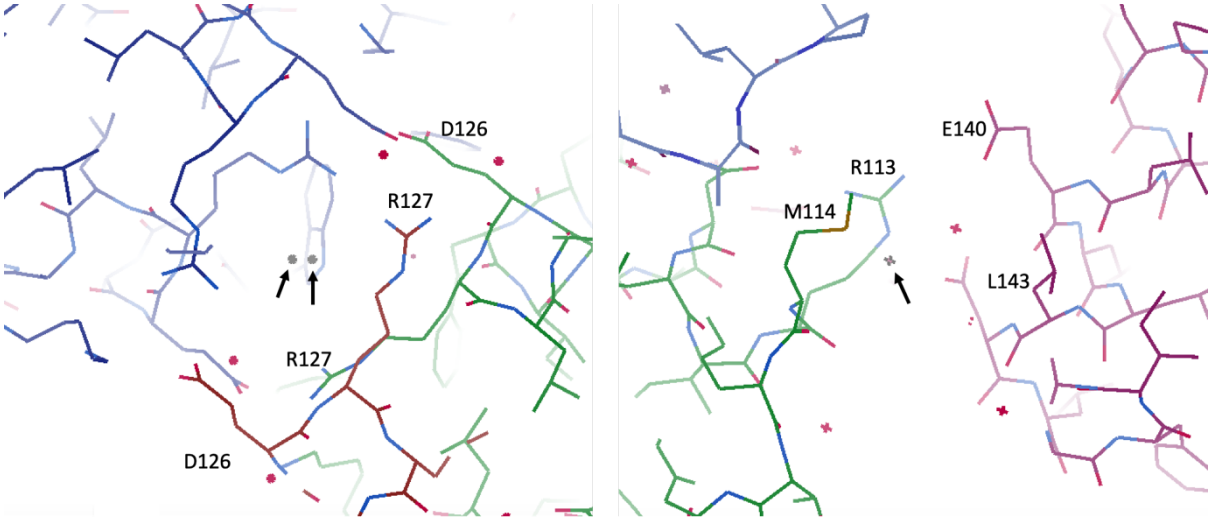


Figure 5.13 Initial structure of Bd2625 structure that was built by Phenix Autosol using Pt sites, indicated by arrows. **(a)** Supersite containing two Pt ions, with some of the surrounding amino acids labelled (Chain 3- Red, Chain 4 - green, Symmetry copy – blue). **(b)** Single Pt ion containing site, with some of the surrounding amino acids labelled (Chain 1 – Red, Chain 4 – green, Symmetry copy – blue).

The asymmetric subunit of the crystals contained 4 partial copies of Bd2625, forming a tetramer from two dimers. For all chains residues 92–149 were placed, with a final R_{work} and R_{free} of 0.24 and 0.28 respectively, and Ramachandran statistics of 98.21% in the preferred region. The DUF4423 region consists of amino acids 85–246, with amino acids 103–189 making up the predicted WHTH-like motif and 121–144 the core HTH DNA binding motif. Therefore, although the entire DUF4423 was not crystallised, the partial structure did contain the proposed central DNA recognition site (Figure 5.14). Very little difference was seen between the secondary structure predicted by Xtalpred (Slabinski et al., 2007) and that observed in the crystal structure (Figure 5.15). The most notable variation is that the last two amino acids are not folded into a β-strand. However, these amino acids are still likely to form a β-strand, and the discrepancy in secondary structure is likely to be due to the protein being cleaved mid-way through this secondary structure element.

```
           1...*...10....*...20....*...30....*...40....*...50....*
Predicted  MGDFVSNQFGALLKNKLEEIQKKNPRFSFRSLAKKVGISPGCLNELMHGKRPLSE
Observed
           ...60....*...70....*...80....*...90....*..100....*..110
Predicted  FYANKIVLGLELGAEERNEVYSLISTRSRKFAAQKTLAEKELELIASWEHFAILN
Observed                                      TLAEKELELIASWEHFAILN

           ....*..120....*..130....*..140....*..150....*..160....*
Predicted  LIRMKTFKPEPEWIAERLALPLEKVQQSLELLLDLGFIKRKGNSIARSVASLATT
Observed   LIRMKTFKPEPEWIAERLALPLEKVQQSLELLLDLGFIK

           ..170....*..180....*..190....*..200....*..210....*..220
Predicted  TDIPSEALVQAHVSDMHKAIEVLKRTPIDRRDYSAITMAINPHKMEEAKNLIKKF
Observed

           ....*..230....*..240....*..250....*
Predicted  RRKFSMLVEEGDMTEVYNLNIQFFPLTVTESEKPL
Observed
```

Figure 5.14 Bd2625 sequence showing the secondary structure predicted by Xtalpred (Slabinski et al., 2007) against the secondary structure observed, helices depicted in red and sheets in blue.

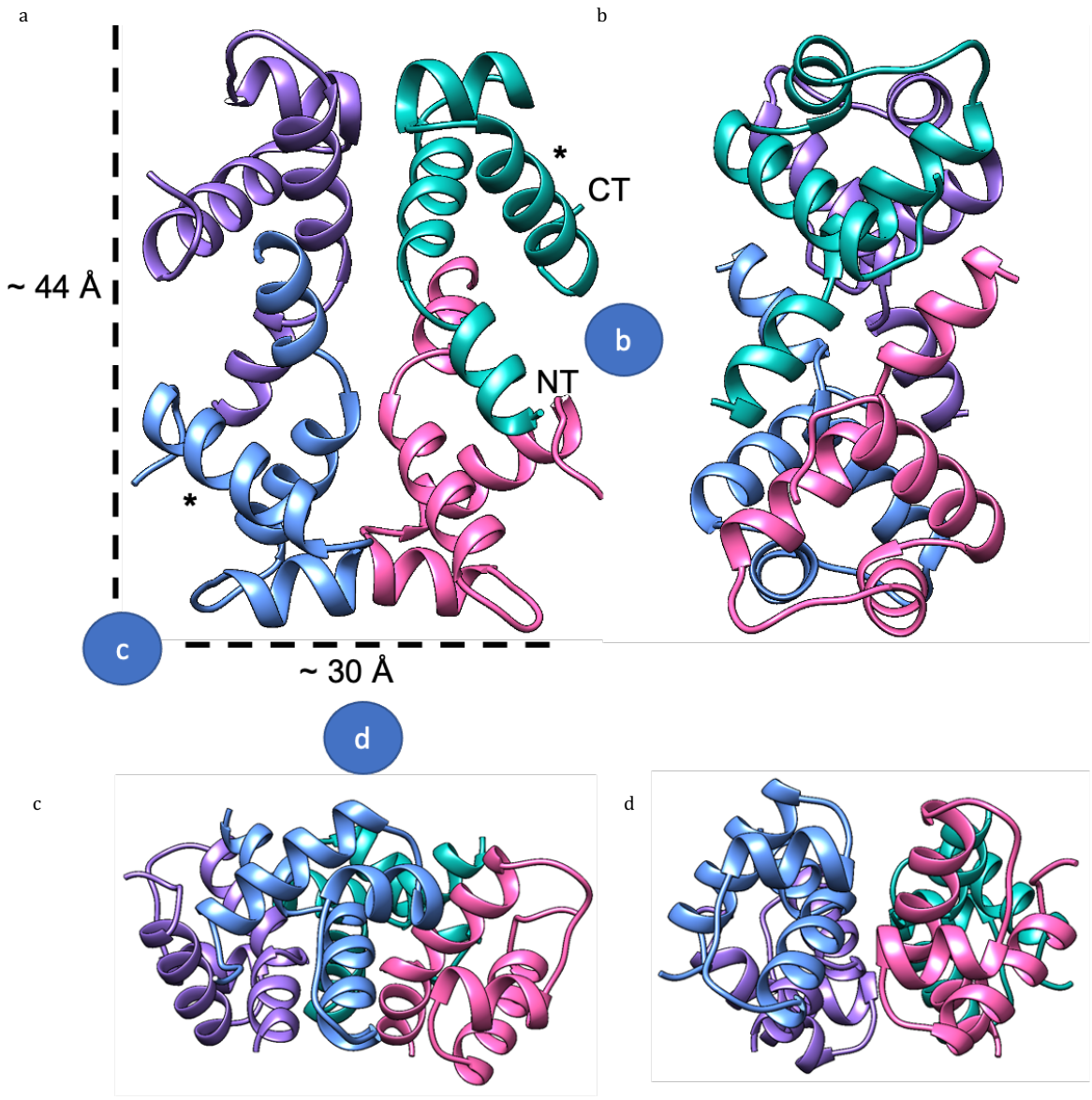


Figure 5.15 (a) Bd2625 tetramer with the N-terminal (NT), C-terminal (CT) labelled and recognition helices labelled (*) on the teal monomer. Circles denoting the angle at which the other images were taken are also shown. **(b)** Bd2625 tetramer at 90° rotation from (a) right side. **(c)** Bd2625 tetramer at 90° rotation from (a) left corner. **(d)** Bd2625 tetramer at 90° rotation from (a) bottom side.

The two monomers dimerise via an interface comprised of 8 hydrogen bonds (Figure 5.16, Figure 5.17), 4 from each monomer, as well as hydrophobic packing. All of the hydrogen bonds are formed between the sidechains of one monomer and the backbone of another, and the majority are located in H2 of the HTH motif within the domain. Three hydrogen bonds are formed between the R127 sidechain and the backbone carbonyl groups of E136 and R127, all located within H2 of the HTH motif.

The final hydrogen bond is formed between the N110 sidechain within H1 of the HTH motif and the backbone carbonyl of A129 from the turn region of the HTH motif.

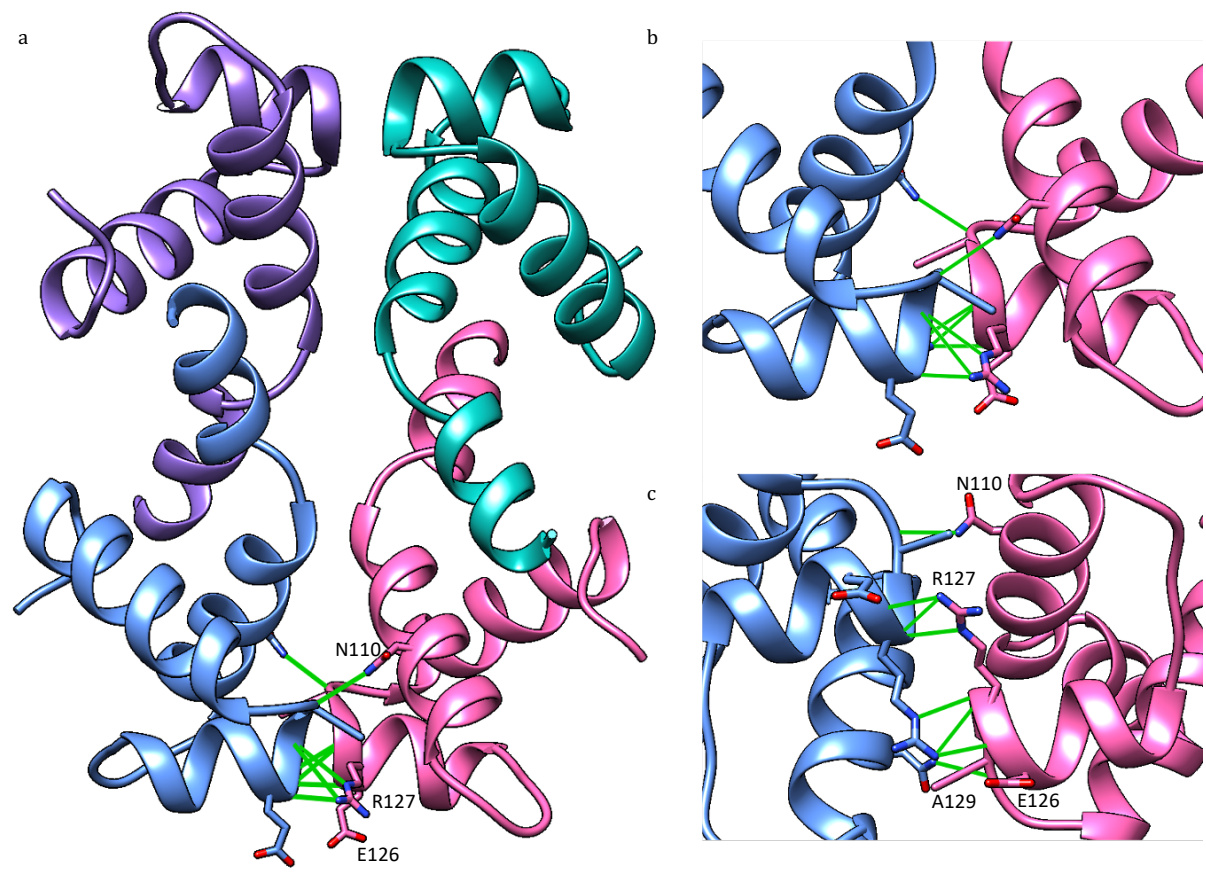


Figure 5.16 (a) Bd2625 tetramer showing hydrogen bonds that stabilise the dimerisation of two monomers. Hydrogen bonds shown in green between the blue and pink monomers only, although these hydrogen bonds are present in all monomers. **(b)** Close-up of the hydrogen bonds between the two monomers. **(c)** Close-up of the hydrogen bonds between the two monomers, 90° rotation from (b).

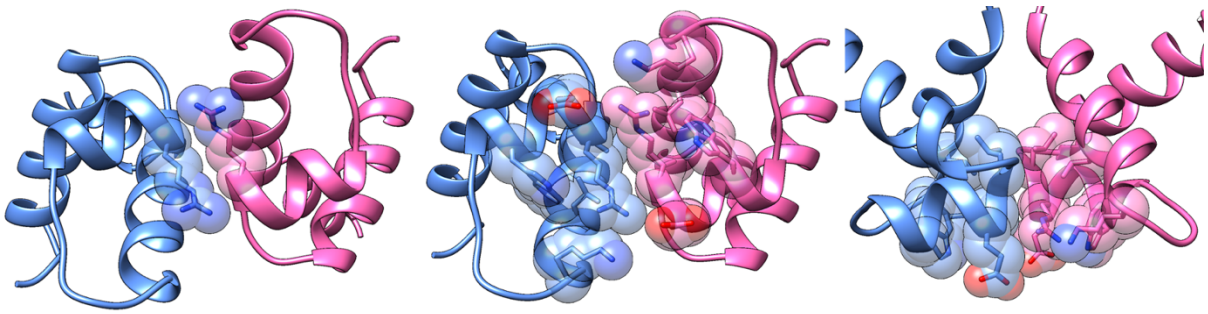


Figure 5.17 (a) The hydrophobic packing between R127 equivalents, which form hydrogen bonds between each monomer, stabilising dimerisation. **(b)** Amino acids in the environment surrounding R127 which aid in the hydrophobic packing via dimerisation. **(c)** 90° rotation from (b).

The majority of the hydrophobic packing that helps to stabilise the dimeric interaction occurs in H2 and the turn of the HTH motif, and involves the amino acids involved in hydrogen bonding and those in the surrounding environment (Figure 5.18). Outside of this region, π-π interactions between W103 and F106 from each monomer further stabilises dimerisation (Figure 5.19). Tetramerisation of the two dimers is stabilised largely by hydrophobic interactions as far as our partial domain structure indicates (Figure 5.20).

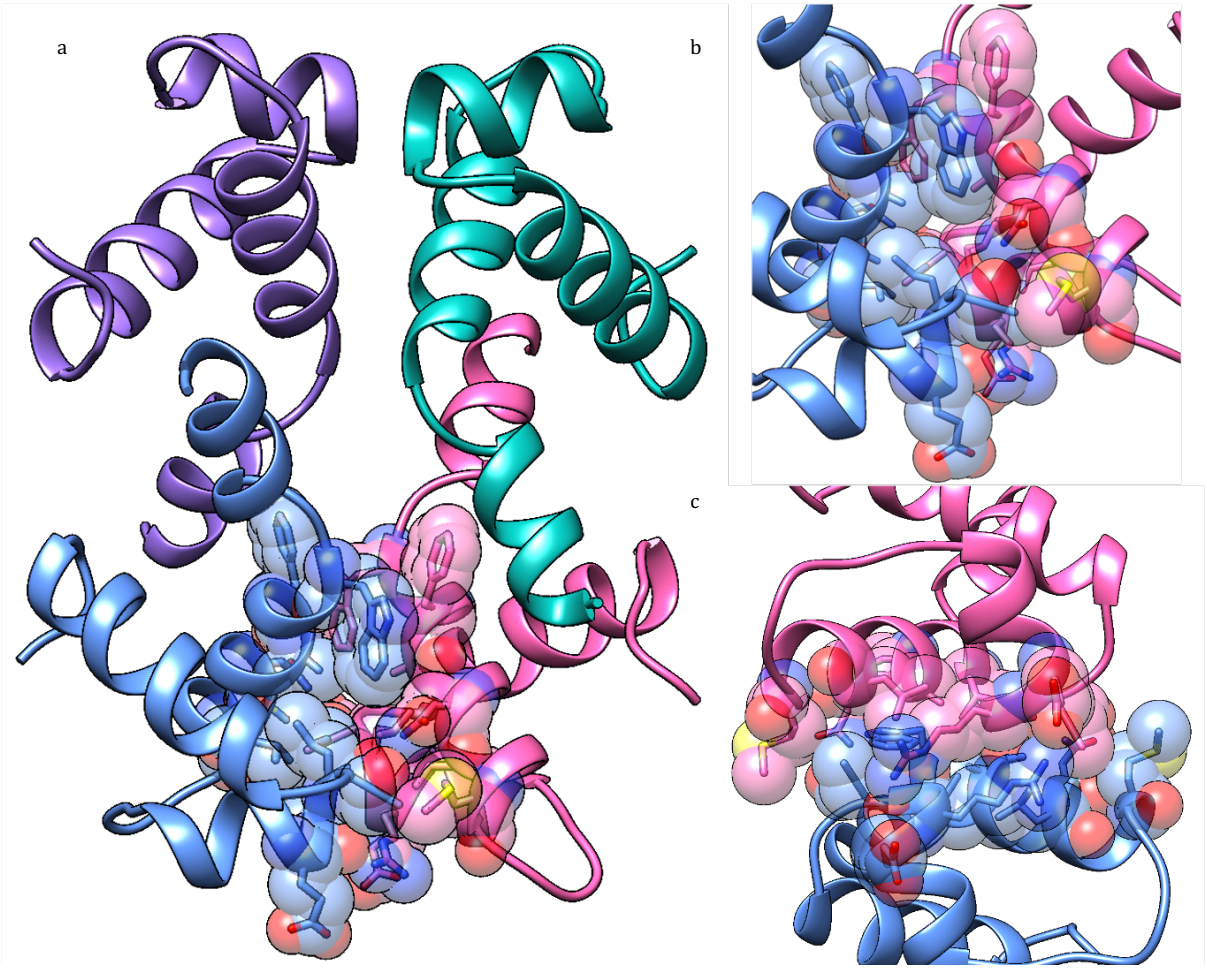


Figure 5.18 (a) Bd2625 tetramer showing hydrophobic interactions that stabilise the dimerisation of two monomers. Packing is only shown between the blue and pink monomers, but would also be present between the purple and teal monomers. **(b)** Close-up of the packing. **(c)** Close-up of the packing shown from the inner core.

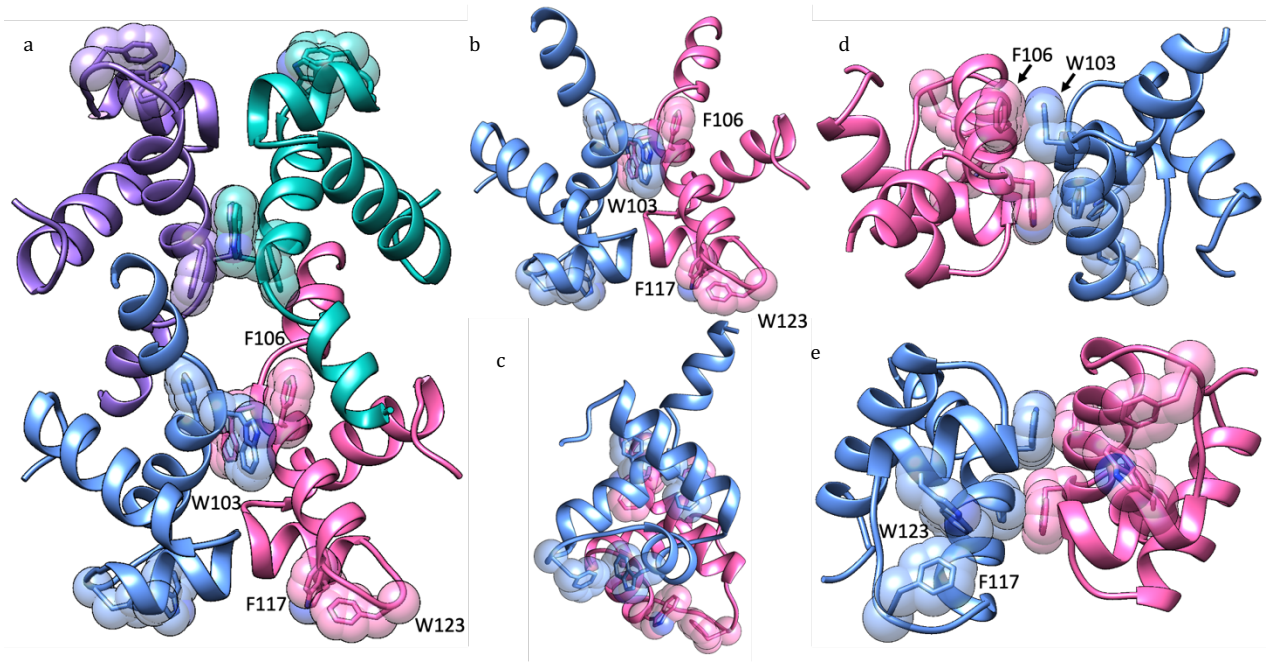


Figure 5.19 Bd2625 tetramer highlighting the importance of the packing between W103 and F106 in dimer packing, as well as F117 and W123 in stabilisation of the HTH motif. **(a)** Position in Bd2526 tetramer **(b)** Position within Bd2625 dimer. **(c)** 90° rotation of (b) to the right. **(d)** 90° rotation of (b) downward. **(e)** 90° rotation of (b) upward.

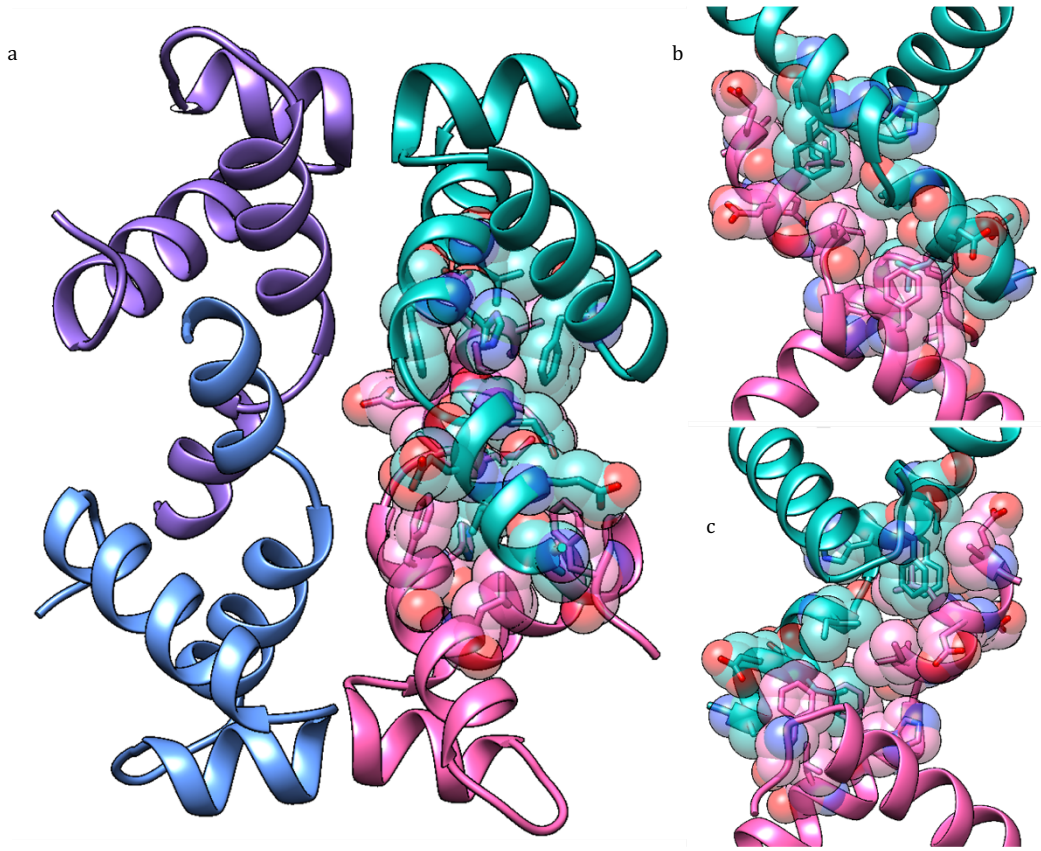


Figure 5.20 (a) Bd2625 tetramer showing hydrophobic interactions that stabilise the dimerisation of two dimers. Packing only shown between the pink and teal monomers, but would also be present between the blue and purple monomers. **(b)**Close-up of the packing shown from the inner core **(c)**Close-up of the packing shown from the outer edge

The refined partial Bd2625 structure showed that each of the four predicted recognition helices in each monomer are readily available for DNA binding. The position of these recognition helices are stabilised by an intra-molecular hydrogen bond forming between S123 and H105 (Figure 5.21). Further hydrogen bonding between L99, S102 and H105 distorts the polypeptide backbone of amino acids L92–M130. As a result of this distortion, two separate helices, L92–I110 and W103–M114, form instead of one continuous one, which positions H105 so it can bind S123. These recognition helices are presented with two on each face of the tetramer, each set diagonally across from one another, chain A across from chain C, and chain B across from chain D (Figure 5.22). It was therefore hypothesised that each diagonal monomer work as a pair to bind DNA, chains AC and BD acting in concert, despite no direct contact between the two individual monomers present in the partial structure. If this held true for the full length structure, it would mean that productive DNA binding was dependant on tetramerisation allowing two monomers to be brought together with the correct juxtaposition to bind DNA.

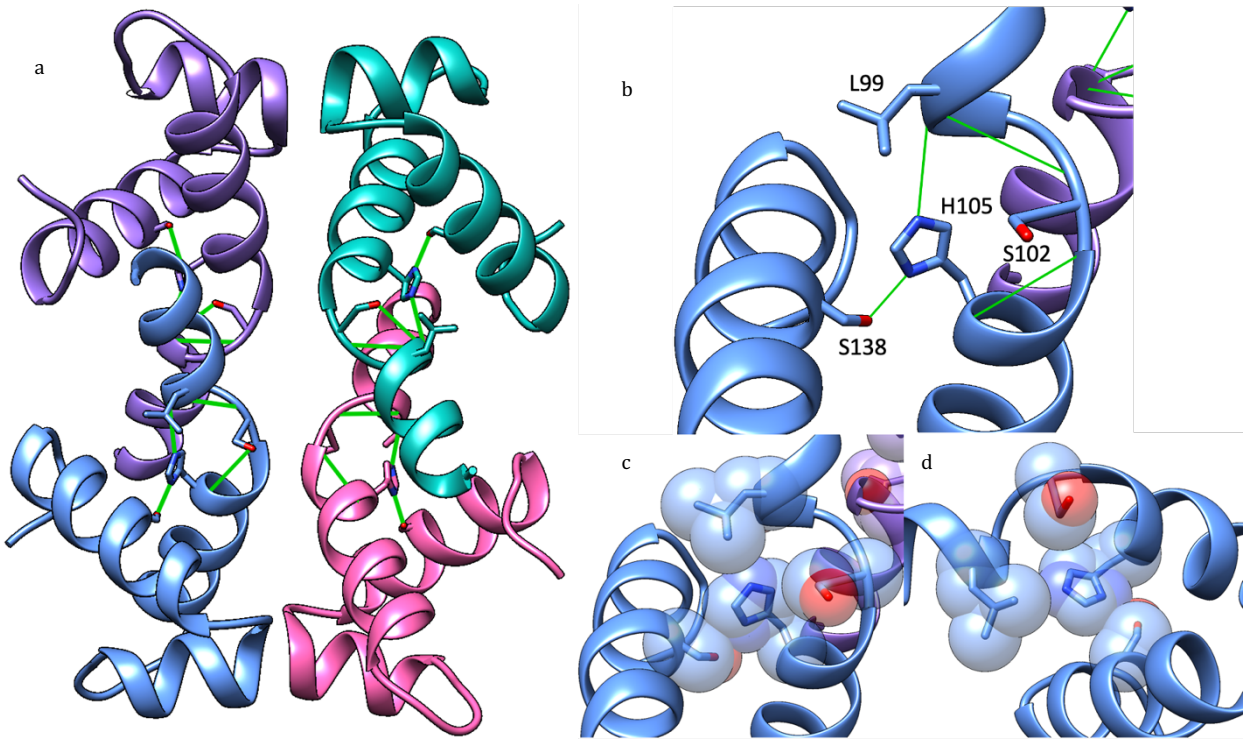


Figure 5.21 (a) Bd2625 tetramer showing the amino acids involved in the stabilisation of the monomeric structure via hydrogen bonding. **(b)**Close-up of hydrogen bonding pattern. **(c)**Close-up of amino acids involved monomeric stabilisation shown as hydrophobic packing. **(d)** Rotation of (c).

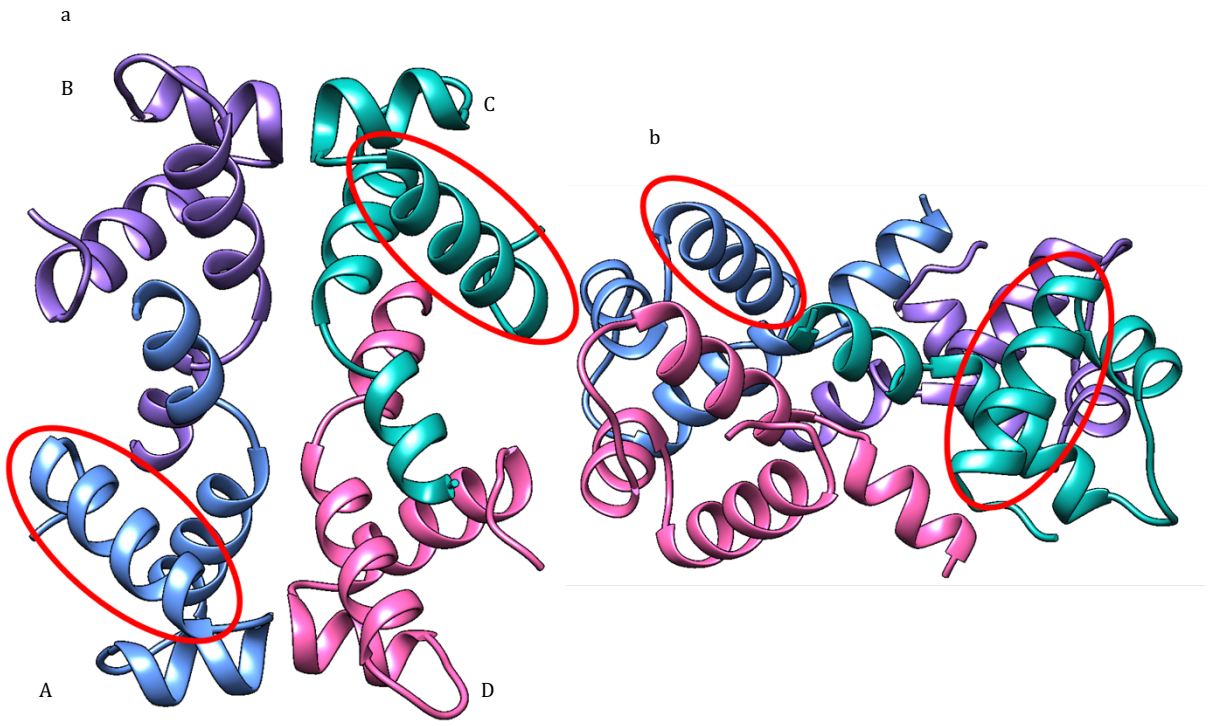


Figure 5.22 Bd2625 tetramer showing freely available recognition helices (indicated by red ovals). Each pair of recognition helices are situated diagonally opposite one another on each face of the tetramer. Therefore diagonal monomers are likely work together to bind DNA.

The DUF4423 domain of Bd2625 was aligned with the closest DUF4423 sequence homologues in other *Bdellovibrio* strains which were identified via BlastP (Altschul et al., 1990) (Figure 5.23). Amino acids W103, F106, N110 and R127, all of which are involved in dimerisation, were highly conserved across all 9 *Bdellovibrio* strains compared, with W103 and R127 being strictly conserved (Figure 5.23). Whilst F106 was not strictly conserved, *Bdellovibrio exovorus* was the only *Bdellovibrio* strain compared to not use an aromatic residue at this position, this highlights the importance of the π-π interaction between W103 and F106 in the formation of the dimer. Similarly, of the amino acids involved in HTH stabilisation, only the two aromatic residues, F117 and W123, are highly conserved. However, as two of the amino acids involved in kinking the L92-M114 region into two separate helices do so via backbone hydrogen bonding this does not necessarily indicate that the HTH motif conformation varies greatly between DUF4423 homologues.

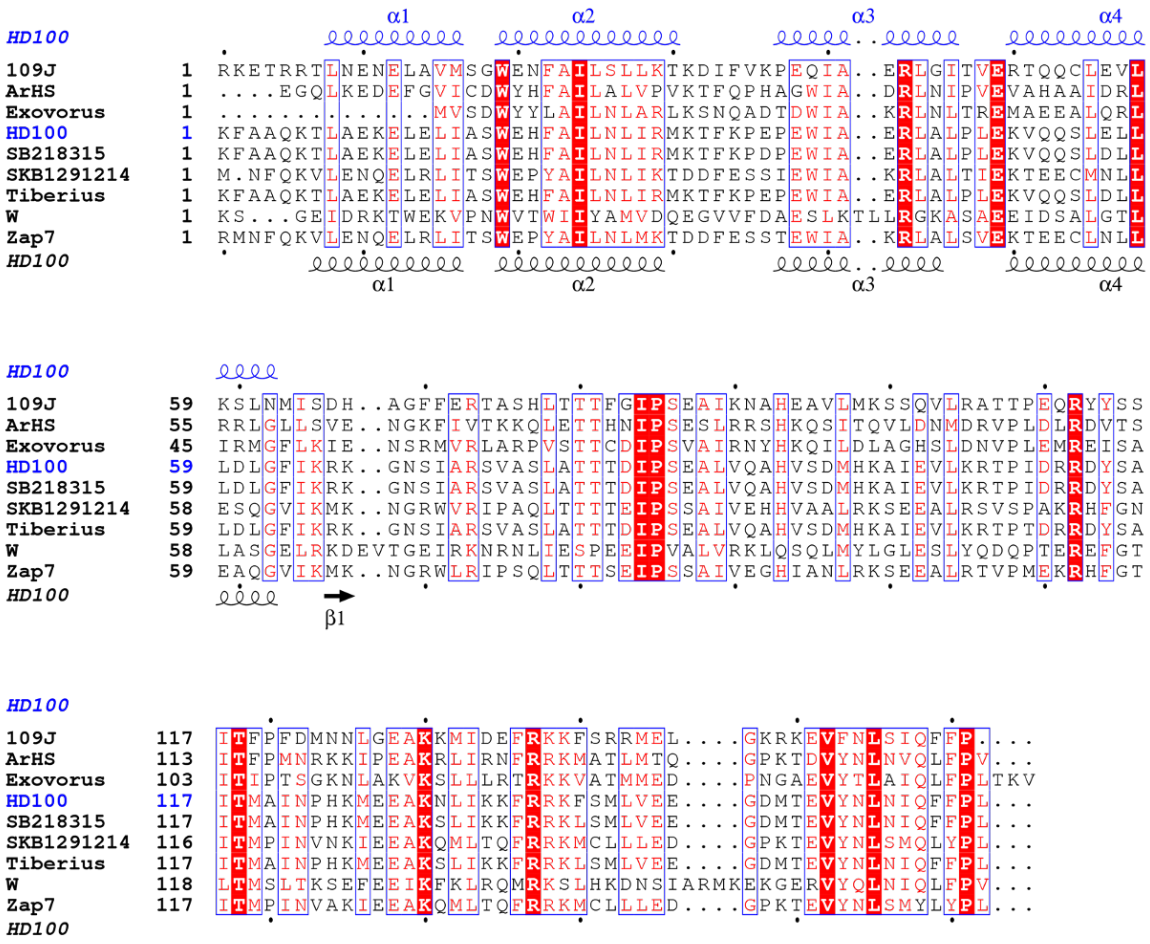


Figure 5.23 Alignment of Bd2625 DUF4423 to the closest DUF4423 homologues in other *Bdellovibrio* strains (generated using ESPript 3.0 (Robert and Gouet, 2014), where all homologues are HTH-DUF4423 variants with the exception of *Bdellovibrio exovorus* and *Bdellovibrio SKB1291214* which are DUF4423 only variants. Overlapped in blue above the alignment is the secondary observed structure, below in black is the predicted structure of Bd2625 taken from Xtalpred (Slabinski et al., 2007)

The tetramer was then submitted for structural alignment searches using the 'Topsearch' server (Wiederstein et al., 2014). The advantage of 'Topsearch' over 'DALI' (Holm, 2020), is that it not only takes into account fold similarity, but also the oligomerisation state of the protein. 'Topsearch' identified WHTH DNA binding protein IscR from *E. coli* (4HF1, Rajagopalan *et al.*, 2013) as the characterised protein with the highest structural homology to Bd2625, with a corrected S_t score of 37.4 (this score is purely used to rank matches against one another where the maximum value of St is equal to target length (Sippl and Wiederstein, 2008)). The crystal structure of IscR is a dimer, but when

superimposed against a diagonal pair of Bd2625, both recognition helices align fairly well with those of IscR that insert into the major groove of the DNA fragment co-crystallised (Figure 5.24). There is an almost prefect alignment between the HTH motifs of one monomer from each structure, however the second HTH motif in the second monomer of the DNA binding pair does not align as well. The variation between the second recognition helices has a relative distance difference between 2.90 Å (at L132 in Bd2625 and L39 in IscR), and 3.95 Å (D144 in Bd2625 and K51 in IscR). This difference in the second recognition helices means that the second recognition helix in Bd2625 cannot make contact with the DNA in the current confirmation. This model suggests that either Bd2625 would bend DNA more than IscR upon binding of DNA to both DNA recognition helices (Figure 5.25) or that Bd2635 undergoes a small induced fit/conformational change upon binding. This could be achieved as Bd2625 does not have a prototypical WHTH. As already stated, the WHTH-like motif in the DUF4423 domain is predicted to have a shorter β-wing than is typical of a WHTH, with the wing region in Bd2625 being predicted to be 2 amino acids shorter. Due to the change in angle between the second recognition helix of Bd2625 and that of IscR, an RMSD difference of 3.6 Å is observed between the start of the wing in Bd2625, I58, and that of IscR, L54. This would change the position of the β-wing, and paired with the shortening, these changes would be enough to allow the DNA to bend further in order to make contact with the second recognition helix before the β-wing inserted into the adjacent minor groove to make further stabilising contacts.

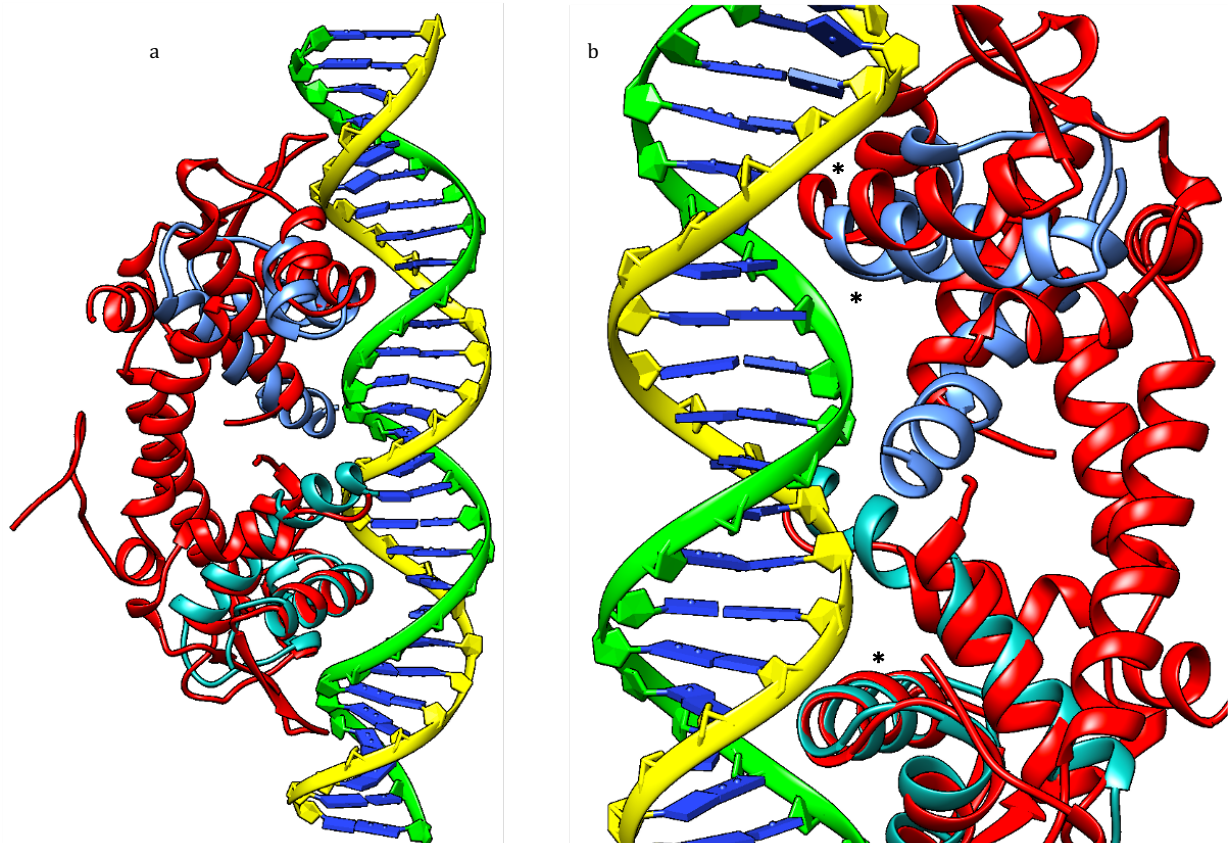


Figure 5.24 Bd2625 diagonal pair (blue and teal) superimposed with IscR:DNA complex (IscR (4HF1) in red, DNA in yellow and green (Rajagopalan et al., 2013)). Alignment using the recognition helices of both HTH-motifs within each dimer shows good alignment of the recognition helices of one monomer, and insertion of both helices into the major groove of the DNA. Recognition helices denoted by a * in (b), the well-aligned helices in the bottom HTH motif (teal and red) share a denotation, whilst the divergent helices (blue and red) are denoted by separate *.

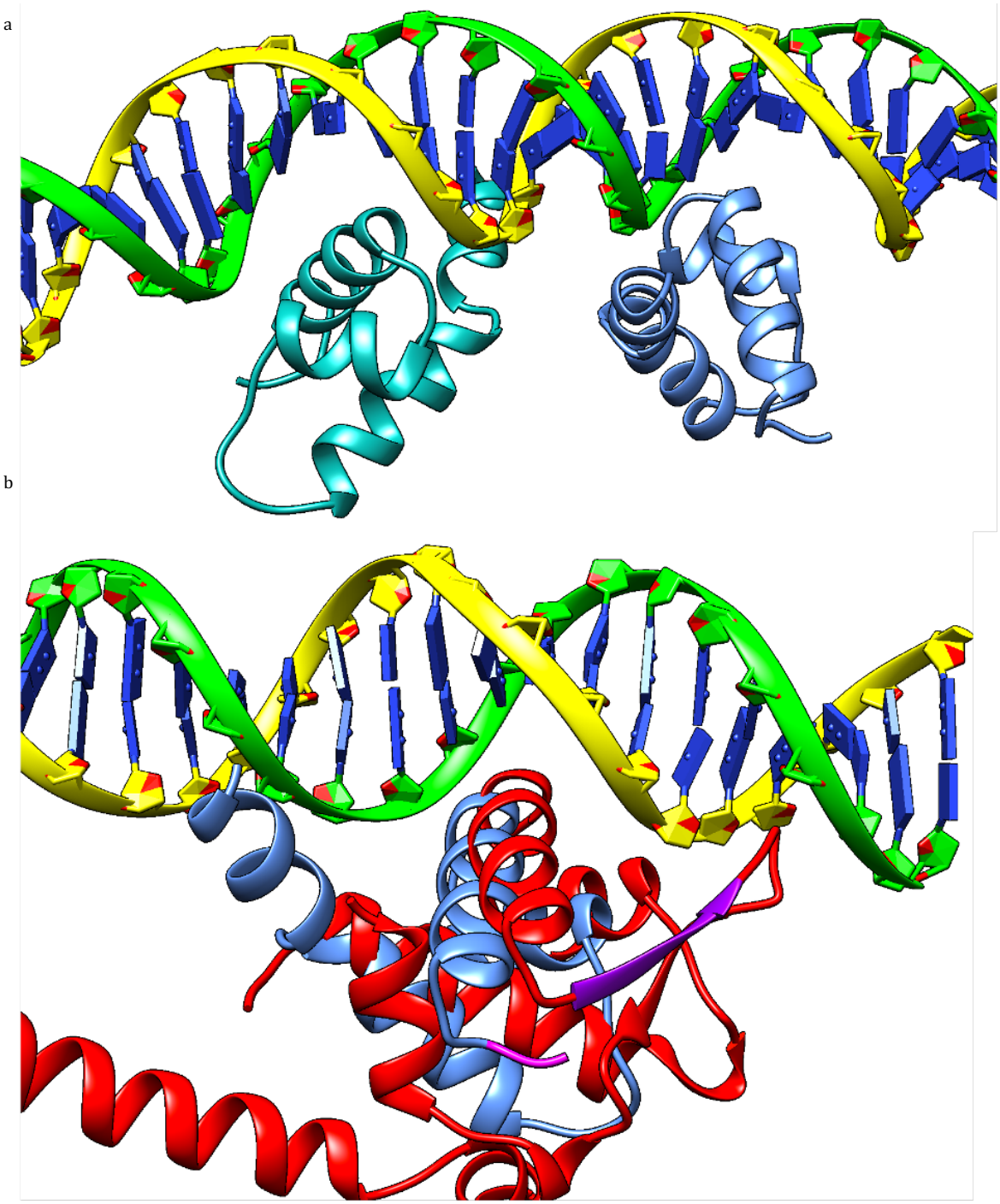


Figure 5.25 (a) Composite model created by superimposing Bd2625 DNA binding pair (blue and teal) onto IscR (not depicted) bound to a DNA fragment (4HF1, yellow and green (Rajagopalan et al., 2013)), demonstrates a fairly good match to the expected protein:DNA interface. **(b)** Close-up of difference between second HTH motifs in IscR (red) dimer and Bd2625 (blue), shows that Bd2625 would bend DNA more than IscR in order for both recognition helices to bind DNA and that the change in the position of the β-wing between IscR (purple) and Bd2625 (first 2 amino acids in pink) would allow this to occur.

Modelling of two copies of IscR against the Bd2625 tetramer showed that the tetramer could bind DNA at both faces, without the DNA binding at one face interfering with binding at the other (Figure 5.26). This holds true even when the difference in binding angle and bending of the DNA when in complex with IscR is compared to the predicted binding with Bd2625. This arrangement results from the two DNA fragments being modelled at almost right angles to one another, with the tetramer sandwiched in the middle. This conformation, if it holds true for the full length protein, might suggest that these TFs are involved in DNA looping or bridging.

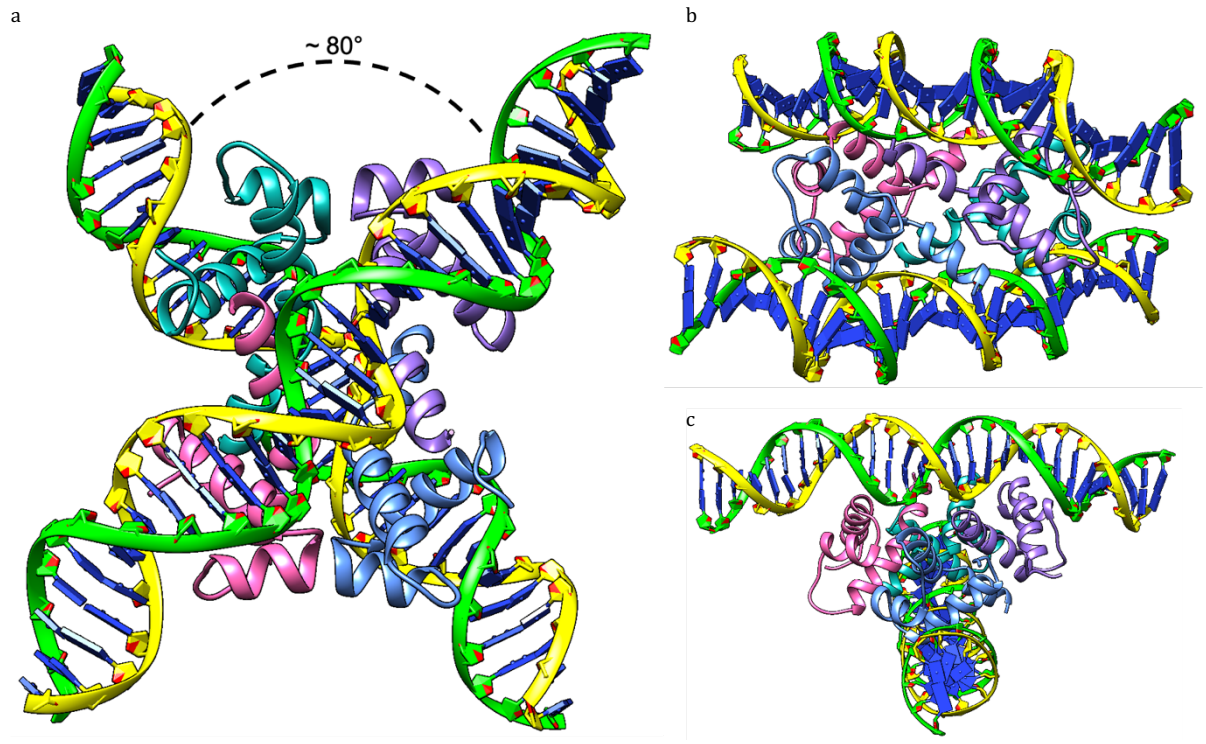


Figure 5.26 Composite model of Bd2625 tetramer (binding pairs blue and teal, pink and purple) with DNA fragment from IscR (4HF1, yellow and green (Rajagopalan et al., 2013)), indicated that the tetramer could be able to bind DNA at both faces without the DNA fragments interfering with one another, suggesting these TFs may be involved in DNA looping or bridging.

5.2.3 Mutagenesis of Bd2625

Efforts to obtain a crystal of full length Bd2625 (Bd2625 FL) had proved unsuccessful, therefore domain truncations of Bd2625 were engineered with the aim of obtaining crystals of the remainder of the protein.

From information of our partial structure, three new constructs were designed; N-terminus to the C-terminus of the partial structure (denoted NT-WingCT, M1-K149), N-terminus of the partial structure to the C-terminus (denoted WingNT-CT, K90-L255), and the partial structure region with a TEV cleavable MBP (denoted Wing+MBP, K90-K149). These new constructs were made with an N-terminal His-tag in pET28a rather than pET41c, as oligomerisation is thought to involve the C-terminus and there was concern that a C-terminal His-tag may be interfering with this interaction and thus potentially interfering with crystallisation. As previously stated, NT-WingCT and WingNT-CT were designed to obtain the "missing domains", whereas Wing+MBP was designed to recreate the original crystals to seed either the full length or other constructs. Cloning was unsuccessful for both WingNT-CT and Wing+MBP, however the NT-WingCT was obtained. NT-WingCT was expressed in AI with sugar mix N, and purified with nickel affinity chromatography, and eluted with 500 mM Imidazole, identically to full length Bd2625 (Figure 5.27a). Eluted fractions of NT-WingCT were supplemented with 2.5 mM β-mercaptoethanol, as this construct contained the sole cysteine of Bd2625, before dialysis into 20 mM Tris pH 8.5, 500 mM NaCl and 2.5 mM β-mercaptoethanol. Similarly to Bd2625, the addition of β-mercaptoethanol stabilised NT-WingCT enough for SEC to be trialled. SEC with a HiLoad 26/60 Superdex 75 pg yielded two larger peaks of concurrent size, monomeric NT-WingCT was eluted in the first of these peaks at 170 ml (Figure 5.27b) . Higher molecular weight contaminants were eluted in the small initial peaks before the monomeric peak, whereas lower molecular weight contaminants were eluted in multiple small peaks after the monomeric peak and in the second large peak at 290 ml. Attempts to crystallise NT-WingCT were unsuccessful.

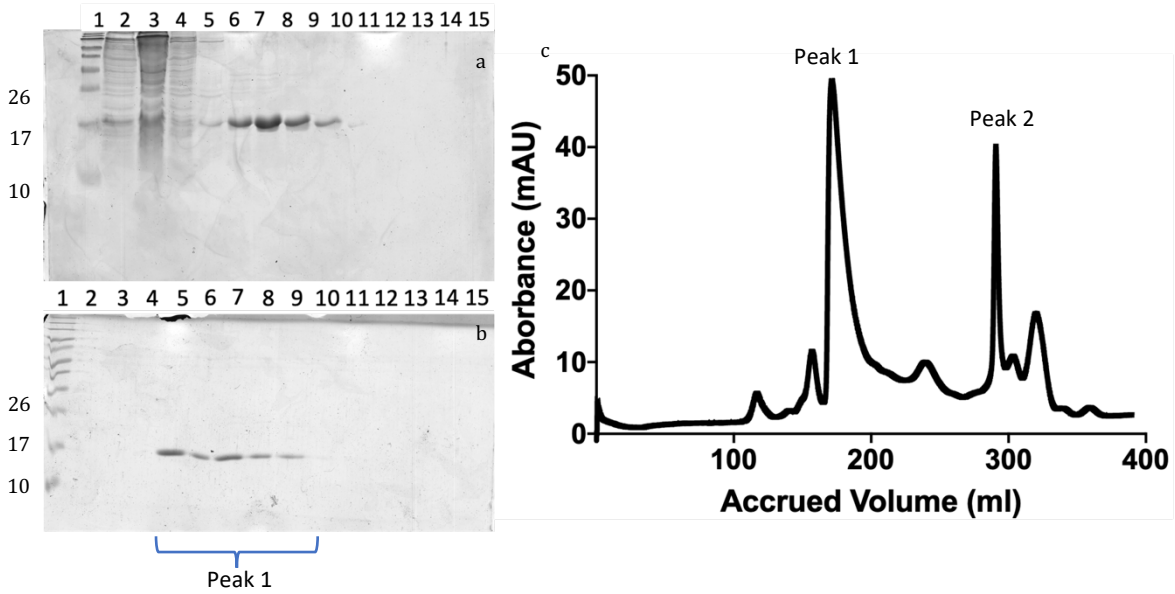


Figure 5.27 (a) SDS-PAGE gel of NT-WingCT IMAC purification. **(b)** SDS-PAGE gel of NT-WingCT post SEC purification. **(c)** UV trace of SEC purification of NT-WingCT, showing monomeric NT-WingCT (peak 1) being separated from contaminant proteins.

Due to the lack of success of the N-terminal His-tag constructs, a second set of Bd2625 truncated constructs were designed. These were also cloned into pET28a, but had a TEV cleavable C-terminal His-tag, to alleviate any instability that may be caused by the His-tag. The two new constructs developed were: N-terminus to the predicted W2 of the WHTH-like motif (denoted NT-VAS, M1-S161) and N-terminus of the partial structure to the same C-terminus (denoted WingNT-VAS, K90-S161). Both were successfully cloned, expressed and purified in the same way as Bd2625 FL and NT-WingCT. Both NT-VAS and WingNT-VAS eluted with 500 mM Imidazole, however whereas NT-VAS produced a similar yield to Bd2625 FL, WingNT-VAS yielded relatively little protein (Figure 5.28) and as such, the WingNT-VAS construct was abandoned. The eluted fractions of NT-VAS were again supplemented with 2.5 mM β-mercaptoethanol before dialysis into 20 mM Tris pH 8.5, 500 mM NaCl and 2.5 mM β-mercaptoethanol and TEV cleavage of the C-terminal His-tag overnight. Unfortunately the removal of the tag destabilised the construct completely and led to precipitation of the protein.

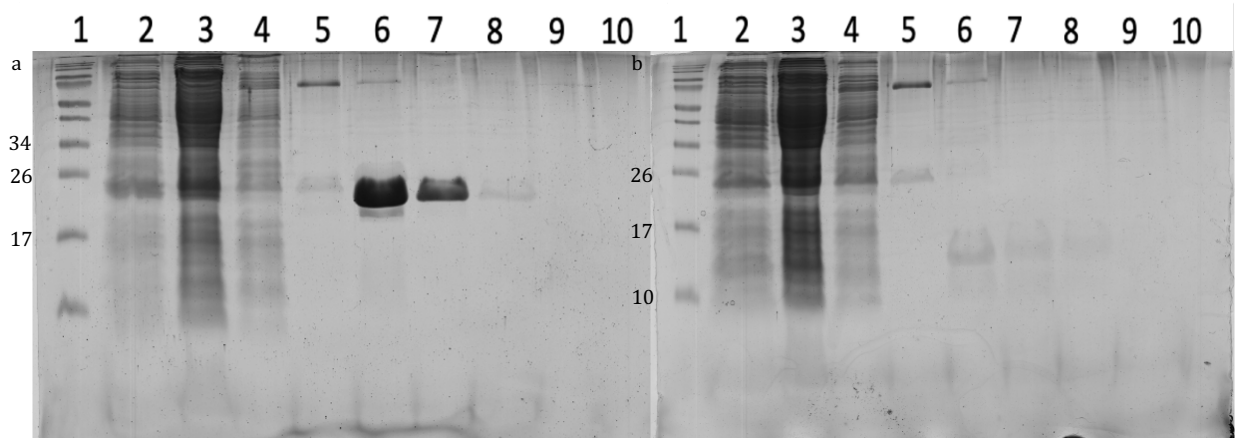


Figure 5.28 (a) 15% SDS-PAGE gel of NT-VAS IMAC purification. **(b)** 20% SDS-PAGE gel of WingNT-VAS IMAC purification

5.2.4 Mutagenesis of Bd1885

In tandem with the Bd2625 truncations, truncations of Bd1885 were also attempted in order to obtain a two-domain structure. These truncations were performed on Bd1885, in spite of its stability issues, as it has the most representative sequence of the whole family. Splitting Bd1885 into its two constituent domains also sought to help stabilise the protein and improve the chances of crystallisation. The two constructs aimed to isolate the HTH domain (denoted Bd1885NT, M1-S79) and the DUF4423 domain (denoted Bd1885CT, G106-H267). Both constructs were originally designed for pET41c with a C-terminal His-tag, however cloning was only successful for Bd1885NT. This led to redesigning of the Bd1885CT construct to enable cloning into pET28a with an N-terminal His-tag. However, cloning was still unsuccessful, and the construct was abandoned. Bd1885NT was expressed identically to the full length construct, in AI media with sugar mix E. Bd1885NT was expressed and purified with a very high yield, being eluted between 100–500 mM Imidazole (Figure 5.29). Eluted fractions were supplemented with 2.5 mM β-mercaptoethanol, as the construct contained a single unpaired cysteine. However, Bd1885NT still retained some of the same stability issues of Bd1885 FL, as such SEC was not performed. Unlike all of the other transcription factor constructs, Bd1885NT was most stable in an acidic buffer, rather than one close to its pI, and was dialysed into 20 mM Bis-Tris pH 5.5, 500 mM NaCl and 2.5 mM β-mercaptoethanol. Despite not being stable enough for SEC,

crystallisation was still attempted. Crystals were obtained in a MIDAS^{+} G4 at a concentration of 35 mg/ml in a 1:1 drop ratio (Table 5.4). These crystals were shot at Diamond without added cryoprotectant due to the 30% v/v Glycerol ethoxylate in the screen condition which should act as an intrinsic cryoprotectant. However, this did not cryoprotect the crystal well enough, resulting in ice formation affecting the diffraction pattern, as such no data set was collected. These crystals have so far proved to be non-reproducible.

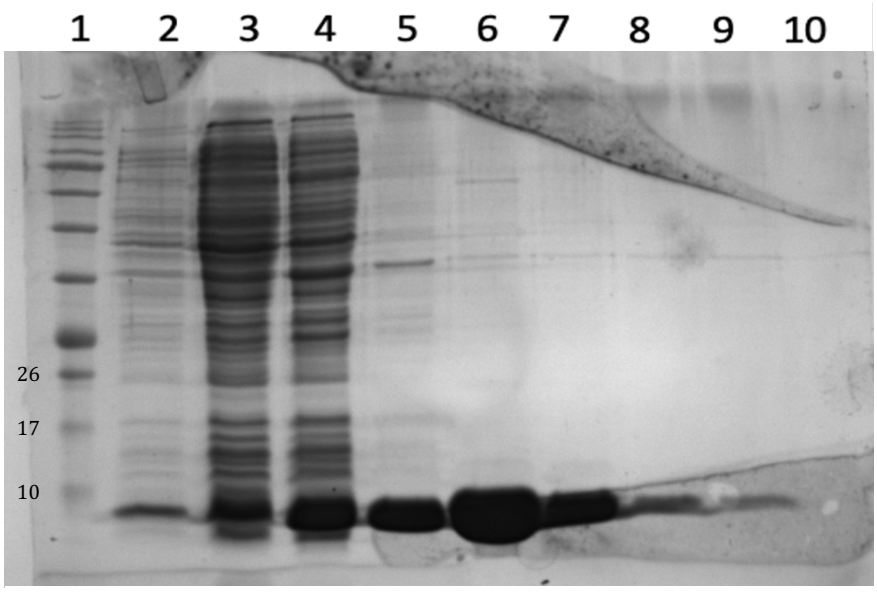


Figure 5.29 SDS-PAGE gel of IMAC purification of Bd1885NT

Method	Sitting drop vapour diffusion
Temperature (°C)	18
Plate Type	96 Well CombiClover 4 chamber crystallization plates
Protein Concentration (mg/ml)	35
Reservoir Solution	0.1 M Tris pH 8.5, and 30% Glycerol exthoxylate
Drop Volume (µl)	2
Drop Ratio	1:1
Volume of Reservoir (µl)	100
Crystallisation Morphology	Cuboid

Table 5.4 Table of conditions used to crystallise Bd1885NT

5.2.5 Electrophoretic Mobility Shift Assays

5.2.5.1 Screening EMSAs

In order to test the DNA binding capabilities of this family of proposed transcription factors Electrophoretic Mobility Shift Assays (EMSAs) were performed on one member from each subfamily that had been successfully cloned. Therefore, initial screening EMSAs using a broad range of 1000:1–0.0001:1 protein to DNA concentration were conducted on Bd1172, Bd1885, and Bd2625, this was possible due to the low concentrations of protein needed, therefore negating the protein stability issues discussed previously for Bd1172 and Bd1885. These screening EMSAs were not reproduced using Bd3720 as previous work had shown that Bd3720 bound DNA. Bd2625 was tested with the two intergenic regions upstream of the *bd2625* coding region, $bd2625^{U}$ (Appendix X), ranging from -91–0 nucleotide bases upstream of the *bd2625* protein coding region, and also $bd2626^{U}$ (Appendix X), ranging from -224–0 nucleotide bases upstream of the *bd2626* protein coding region, as the intergenic region between *bd2625* and *bd2626* is very small and as such Bd2625 may bind upstream of *bd2626*. Neither trial EMSA indicated strong DNA binding.

The previous studies with Bd3720, performed by Charlotte Manning under my supervision, showed binding to a DNA fragment of 340 bp located from -340–0 nucleotide bases upstream of the *bd3721* protein coding region, denoted $bd3721^{U}$. The upstream region of *bd3721* (Appendix X) was chosen over *bd3720* as the intergenic region is only 17 bp, and it was therefore considered unlikely that Bd3720 would bind to the region upstream of itself. Due to the previous work on Bd3720, binding of Bd1172 to DNA was tested against two DNA fragments, $bd1172^{U1}$ (Appendix X), 340 bp directly upstream from *bd1772*, and $bd1172^{U2}$ (Appendix X), the 164 bp intergenic region between *bd1772* and *bd1171*. Bd1772 was shown to shift both $bd1172^{U1}$ and $bd1172^{U2}$ with similar efficacy (Table 5.5 and Figure 5.30), requiring a similar 100:1 protein to DNA concentration ratio for a shift to start to be

observable. However, an initial binding *Kd* could be more reliably obtained for *bd1172*^{U2} and therefore future binding experiments of Bd1172 were performed exclusively with *bd1172*^{U2}.

DNA fragment	Concentration of Bd1172 in Well (μM)									
	1	2	3	4	5	6	7	8	9	10
bd1172^{U1}	Dye	-	0	0.00001	0.00011	0.0011	0.011	0.11	1.11	11.1
bd1172^{U2}	Dye	-	0	0.00002	0.00023	0.0023	0.023	0.23	2.31	23.1
	Ratio of Protein:DNA									
	Dye	-	0:1	0.001:1	0.01:1	0.1:1	1:1	10:1	100:1	1000:1

Table 5.5 Concentration of Bd1172 in each well of screening EMSAs performed with *bd1172*^{U1} and *bd1172*^{U2}. The concentration of *bd1172*^{U1} and *bd1172*^{U2} were kept constant at a final concentration of 11.1 nM and 23.1 nM respectively.

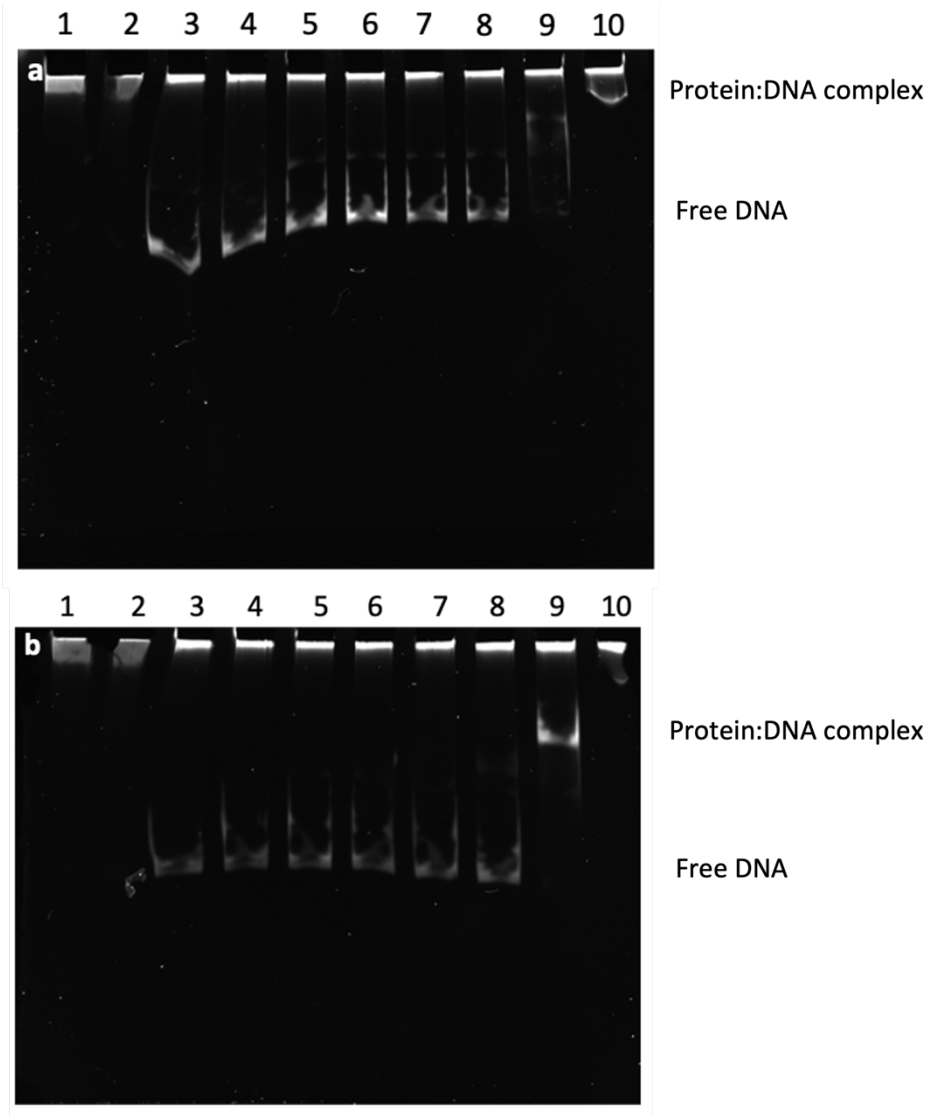


Figure 5.30 (a) EMSA showing Bd1172 shifting *bd1172*^{U1} at a protein:DNA ratio between 10–100:1. **(b)** EMSA showing Bd1172 shifting *bd1172*^{U2} at a protein:DNA ratio between 10–100:1.

Similarly to Bd1172, screening with Bd1885 was performed with two DNA fragments: *bd1885*^{U1} (Appendix X), 340 bp directly upstream from *bd1885*, and *bd1885*^{U2} (Appendix X), the 186 bp intergenic region between *bd1885* and *bd1886*. Bd1885 was shown to shift *bd1885*^{U2} with the same efficiency as Bd1172 with both *bd1172*^{U} fragments (Table 5.6, Figure 5.31), requiring a protein:DNA ratio of 100:1. The binding experiment with *bd1885*^{U1} was inconclusive due to damaged wells causing misloading of the DNA only control. Consequently, all future EMSAs with Bd1885 were performed with *bd1885*^{U2}.

DNA fragment	Concentration of Bd1172 in Well (μM)									
	1	2	3	4	5	6	7	8	9	10
bd1885^{U1}	Dye	-	0	0.00001	0.00011	0.0011	0.011	0.11	1.11	11.1
bd1885^{U2}	Dye	-	0	0.00002	0.00020	0.0020	0.020	0.20	2.04	20.4
	Ratio of Protein:DNA									
	Dye	-	0:1	0.001:1	0.01:1	0.1:1	1:1	10:1	100:1	1000:1

Table 5.6 Concentration of Bd1885 in each well of screening EMSAs performed with *bd1885*^{U1} and *bd1885*^{U2}. The concentration of *bd1885*^{U1} and *bd1885*^{U2} were kept constant at a final concentration of 11.1 nM and 20.4 nM respectively.

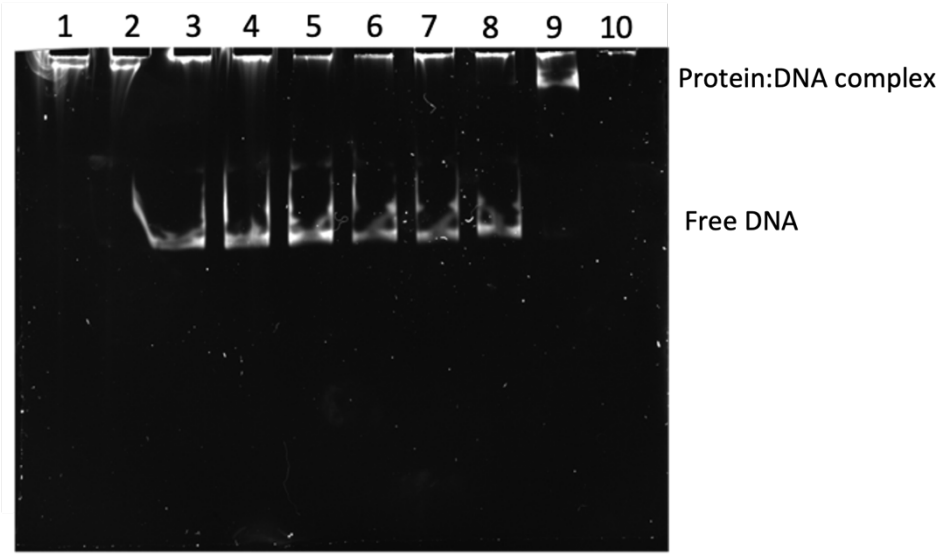


Figure 5.31 (a) EMSA showing Bd1885 shifting *bd1885*^{U1} at a protein:DNA ratio between 10–100:1.

5.2.5.2 Qualitative EMSAs

Qualitative EMSAs were then carried out in triplicate to establish a *Kd* for Bd1172, Bd1885 and Bd3720. Since the screening EMSA had shown Bd1172 shifted DNA at a concentration between 0.23–2.31 μM, the qualitative EMSAs were conducted between 0.05–5 μM to ensure a well-populated binding curve. The EMSA gels showed a steady shift of the unbound DNA band as the protein concentration increased (Table 5.7, Figure 5.32). It was also noted that there were some intermediate shifts of the DNA, which may indicate that the stoichiometry is not 1:1, however it could also indicate that DNA bridging or looping occurs (Cournac and Plumbridge, 2013; Watanabe et al., 2013). The intensity of the unbound DNA band was measured using Fiji (Schindelin et al., 2012), and normalised using the DNA only band as the fully unbound standard. This allowed for the fraction of DNA bound by each concentration of Bd1172 to be calculated. An affinity curve of Bd1172 against the bound fraction of *bd1172*^{U2} was plotted using the mean of triplicate experimental repeats, and a curve was fitted using 'One site – total' using GraphPad Prism 9.0.0 for Mac OS X (www.graphpad.com), with the Bmax set to 1. This analysis was applied as the data fit the expected pattern of a one site binding model, and although the multiple shift bands might indicate a non 1:1 stoichiometry it can also indicate DNA looping (Cournac and Plumbridge), and therefore 'One-site' binding was applied due to the data fitting the expected pattern. This method of analysis also allowed for a *Kd* to be estimated without the removal of non-specific background binding. Bmax was set to 1, as the total bound fraction should not exceed 1, and therefore consequently Bmax cannot. The *Kd* of Bd1172 with *bd1172*^{U2} was estimated to be 0.14 μM, however this does not fit with the 50% shift point on the gel which approximates the *Kd* to be between 1 – 1.75 μM.

Concentration of Bd1172 in Well (μM)														
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Dye	-	0	0.05	0.10	0.15	0.20	0.25	0.40	0.55	0.75	1.00	1.75	2.50	5.00

Table 5.7 Concentration of Bd1172 in each well of qualitative EMSAs performed with *bd1172*^{U2}. The concentration of *bd1172*^{U2} was kept constant at a final concentration of 23.1 nM.

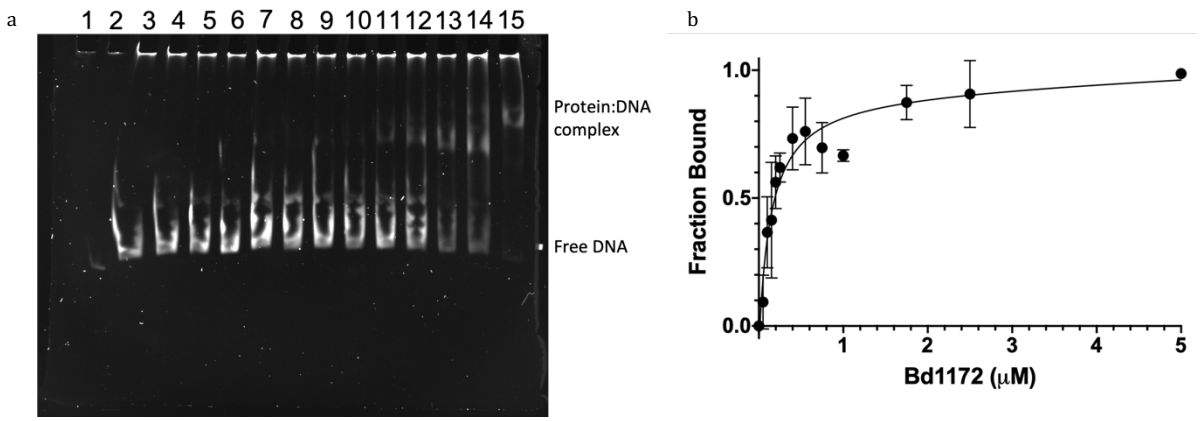


Figure 5.32 (a) Representative qualitative EMSA showing Bd1172 binding *bd1172*^{U2}. **(b)** The affinity of Bd1172 to *bd1172*^{U2}. The mean of three experimental replicate assays was calculated and plotted, along with SD. In order to calculate the *Kd* a curve was fitted using GraphPad Prism – 'One site – Total' with Bmax set to 1. This produced an estimated *Kd* of 0.14 μM, however this does not fit with the 50% shift point on the gel which approximates the Kd to be between 1 – 1.75 μM.

As the screening EMSA of Bd1885 had shown a shift in DNA at a concentration similar to Bd1172 (0.20–2.04 μM), the same protein concentration range of 0.05–5 μM was used for the qualitative EMSAs of Bd1885. Similarly to Bd1172, the gels showed the same steady shift from unbound to fully bound, however far less intermediate banding was present (Table 5.8, Figure 5.33). The intensity of the unbound DNA was measured using Fiji (Schindelin et al., 2012) and normalised using the DNA only sample as the fully unbound standard. An affinity curve of Bd1885 against *bd1885*^{U2} fraction bound was plotted using the mean of in triplicate repeats, and a curve fitted using 'One site – total' using GraphPad Prism 9.0.0 for Mac OS X (www.graphpad.com), with the Bmax set to 1. The constraints were applied to the Bd1885 data in the same manner as the Bd1172 data. The *Kd* of Bd1885 with *bd1885*^{U2} was estimated to be 0.12 μM, however this does not fit with the 50% shift point on the gel which approximates the *Kd* to be between 0.55 – 1.00 μM.

Concentration of Bd1885 in Well (μM)														
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Dye	-	0	0.05	0.10	0.15	0.20	0.25	0.40	0.55	0.75	1.00	1.75	2.50	5.00

Table 5.8 Concentration of Bd1885 in each well of qualitative EMSAs performed with *bd1885*^{U2}. The concentration of *bd1885*^{U2} was kept constant at a final concentration of 20.4 nM.

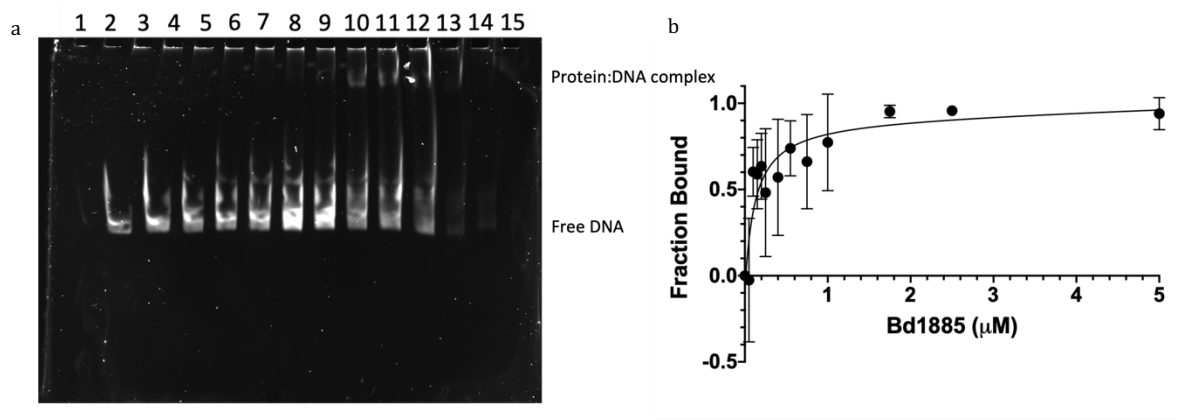


Figure 5.33 (a) Representative qualitative EMSA showing Bd1885 binding *bd1885*^{U2}. **(b)** The affinity of Bd1885 to *bd1885*^{U2}. The mean of three experimental replicate assays was calculated and plotted, along with SD. In order to calculate the *Kd* a curve was fitted using GraphPad PRISM – 'One site – Total' with Bmax set to 1. This produced an estimated *Kd* of 0.12 μM, however this does not fit with the 50% shift point on the gel which approximates the *Kd* to be between 0.55 – 1.00 μM.

The DNA fragment for Bd3720 EMSAs was expanded from the 340 bp upstream of *bd3721* used in the original trials to 361 bp upstream of *bd3721*, *bd3721*^{U}, to allow for more consistent production of the fragment. Unlike with Bd1172 and Bd1885, a qualitative EMSAs of Bd3720 at concentrations between 0.25–10 μM had been previously undertaken, and an initial *Kd* estimate of 1.74 μM determined. However, in contrast to Bd1172 and Bd1885 the shift between unbound and bound was not readily reproducible, and further refinement of Bd3720 concentration as well as technical repeats were necessary to determine a more accurate *Kd*. Thus, qualitative EMSAs at a concentration range of 1–5 μM were trialled to see if a more reliable shift could be produced. A steady shift was still not achieved, with full DNA binding happening suddenly at either 1.6 or 1.7 μM, varying between individual experiments (Table 5.9, Figure 5.34). The gels were processed in the same way as the other qualitative gels, and an affinity curve of Bd3720 against *bd3721*^{U} fraction bound was plotted using the mean of quadruplicate repeats. Since a steady shift was still not achieved, the data displayed a sigmoidal trend rather than a traditional one site binding curve. For this reason, the 'One site – Total' curve could not be fit, and instead a curve was fitted using 'Specific binding with Hill slope' using GraphPad Prism 9.0.0 for Mac OS X (www.graphpad.com). Although this allowed for the *Kd* to be more accurately estimated,

as it takes into account the sigmoidicity of the curve, it does not allow for background non-specific binding. In spite of this a refined *Kd* estimate of 1.34 μM was calculated, which was to be used in further experimentation of Bd3720 against *bd3721*^{U} with the addition of a competitor. However, as with the other qualitative EMSAs performed, this did not tally with the 50% shift point seen on the gel which estimates the *Kd* to be between 1.6 – 1.7 μM.

Concentration of Bd3720 in Well (μM)														
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Dye	-	0	1.00	1.10	1.20	1.30	1.40	1.50	1.60	1.70	1.80	2.00	2.50	5.00

Table 5.9 Concentration of Bd3720 in each well of qualitative EMSAs performed with *bd3721*^{U}. The concentration of *bd3721*^{U} was kept constant at a final concentration of 10.5 nM.

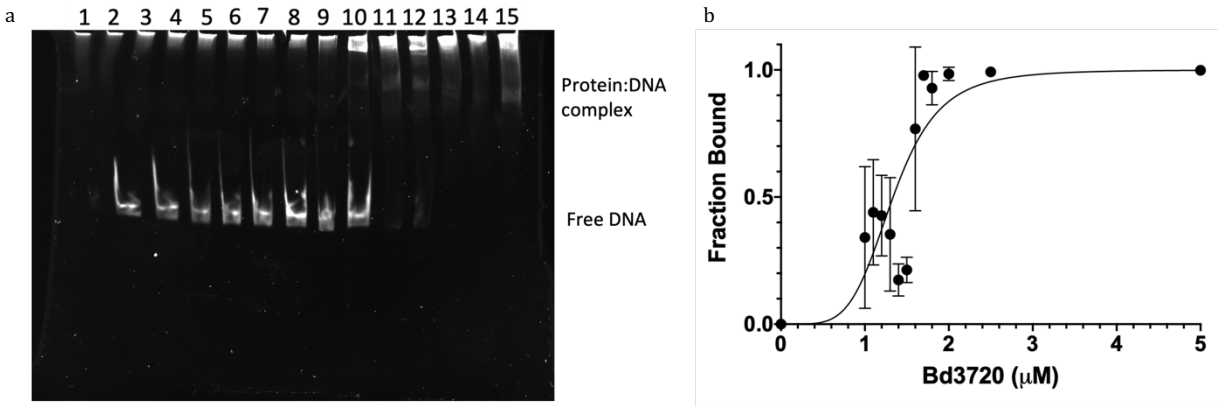


Figure 5.34 (a) Representative qualitative EMSA showing Bd3730 binding *bd3721*^{U}. **(b)** The affinity of Bd3720 to *bd3721*^{U}. The mean of three experimental replicate assays was calculated and plotted, along with SD. In order to calculate the *Kd* a curve was fitted using GRAPHPAD PRISM – 'Specific binding with Hill slope' with Bmax set to 1. The *Kd* was estimated to be 1.34 μM, however this does not fit with the 50% shift point on the gel which approximates the Kd to be between 1.6 – 1.7 μM.

5.2.5.3 Specificity of Bd3720

In order to start to determine the specificity of this transcription factor family to particular DNA sequences, the ability of Bd3720 to bind a variety of different DNA fragments was assayed. Firstly, screening EMSAs of Bd3720 with other upstream regions; *bd0089*^{U} (Appendix X), *bd1172*^{U1}, *bd1885*^{U1}, *bd2625*^{U}, and *bd2627*^{U}, were trialled (Table 5.10). *bd1172*^{U1} and *bd1885*^{U1} were chosen over the shorter *bd1172*^{U2} and *bd1885*^{U2} to limit any negative effect fragment length may have on Bd3720

binding to DNA, as they are more similar in length to the known binder *bd3720*^{U}. These EMSAs demonstrated binding between Bd3720 and all of the variant DNA fragments (Figure 5.35). *bd0089*^{U} and *bd2627*^{U} were both shifted with a similar efficacy to *bd3720*^{U} with full binding happening between a 10–100:1 protein:DNA ratio. Although *Bd1172*^{U1}, *bd1885*^{U1} and *bd2626*^{U} all bound to Bd3720, they bound 10x less efficiently, achieving full shift between 100–1000:1 protein to DNA concentration. This promiscuity of binding to the upstream regions, led to postulation that Bd3720 might be less specific than first thought, potentially due to the widespread nature of the TF family throughout the genome.

DNA fragment	Concentration of Bd3720 in Well (μM)									
	1	2	3	4	5	6	7	8	9	10
***Bd0089*^{U}**	Dye	-	0	0.00001	0.00011	0.0011	0.011	0.11	1.11	14.9
***bd1172*^{U1}**	Dye	-	0	0.00001	0.00011	0.0011	0.011	0.11	1.11	11.1
***bd1885*^{U1}**	Dye	-	0	0.00001	0.00011	0.0011	0.011	0.11	1.11	11.1
***Bd2625*^{U}**	Dye	-	0	0.00004	0.00042	0.0042	0.042	0.42	4.16	41.6
***Bd2626*^{U}**	Dye	-	0	0.00002	0.00017	0.0017	0.017	0.17	1.69	16.9
Ratio of Protein:DNA										
	Dye	-	0:1	0.001:1	0.01:1	0.1:1	1:1	10:1	100:1	1000:1

Table 5.10 Concentration of Bd3720 in each well of screening EMSAs performed with each upstream region. The concentration of DNA was kept constant at the same concentration of protein in the 1:1 protein to DNA ratio well.

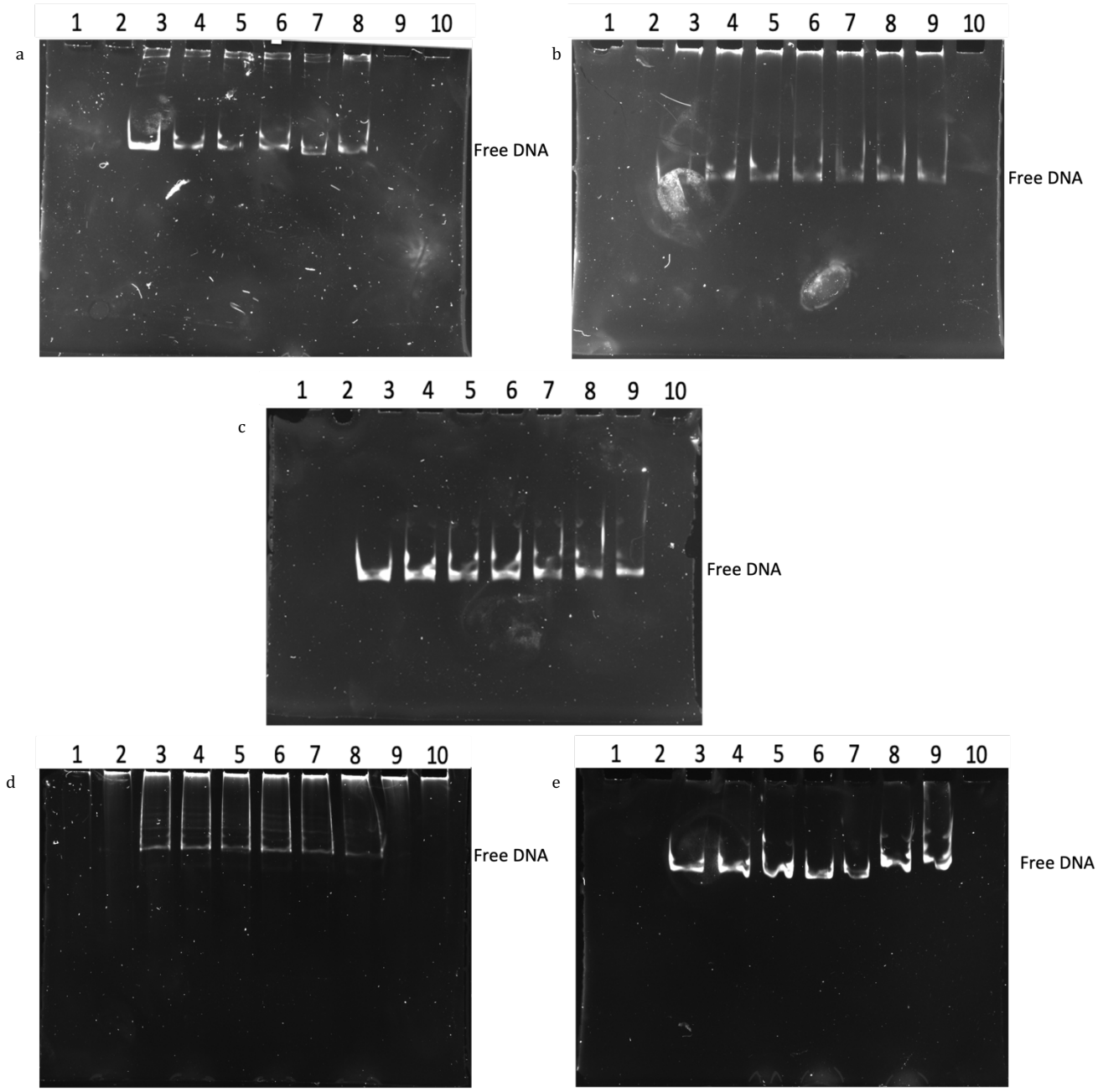


Figure 5.35 Screening EMSAs of Bd3720 against a variety of upstream DNA fragments were used to assess specificity of Bd3720 binding. **(a)** Bd3720 binding *bd0089*^{U}, with full binding occurring at a protein to DNA ratio between 10 - 100:1 **(b)** Bd3720 binding *bd1172*^{U1}, with full binding occurring at a protein to DNA ratio between 100 - 1000:1 **(c)** Bd3720 binding *bd1885*^{U1}, with full binding occurring at a protein to DNA ratio between 100 - 1000:1 **(d)** Bd3720 binding *bd2625*^{U}, with full binding occurring at a protein to DNA ratio between 10 - 100:1 **(e)** Bd3720 binding *bd2627*^{U}, with full binding occurring at a protein to DNA ratio between 100 - 1000:1

To test the specificity of Bd3720 further, an additional two EMSA experiments were ran (Table 5.11). The first EMSA compared the binding capacity of Bd3720 to negative control sCache chemoreceptor domain of Bd3092. Bd3092 has been characterised to bind propanoic acid and acetate (Chapters 3 and 4 of this thesis), and is therefore unlikely to specifically bind DNA. This paired with a similar overall charge and molecular weight between the Bd3092 dimer and monomeric Bd3720, makes it an ideal control to show that Bd3720 is interacting with the DNA via DNA binding motifs, not non-specifically. This EMSA showed, as predicted, that at a concentration of protein approximately 4x the *Kd* of Bd3720 binding to *bd3721*^{U}, Bd3720 was able to fully bind the DNA, and Bd3092 was not (Figure 5.36a). This implies that these transcription factors interact with DNA in a specific manner (likely via at least one of their two predicted DNA binding domains). The specificity of Bd3720 to particular DNA fragments was explored further by testing the binding capacity of Bd3720 for other *Bdellovibrio* genes. Firstly, this was trialled against *gltF* (*bd2373*) (Appendix X), a gene thought to be involved in *Bdellovibrio* gliding motility, and therefore a control gene that would be differentially expressed throughout the lifecycle. Similarly to the *bd3720* upstream regions Bd3720 fully shifted the *gltF* gene, again at a concentration 4x the *Kd* of Bd3720 binding to *bd3721*^{U} (Figure 5.36a). Secondly, Bd3720 binding was tested against the HTH domain of *bd1885*, *bd1885NT* (Appendix X), to see if these transcription factors could 'self' bind. Similarly to *gltF*, Bd3720 was able to bind and shift *bd1885NT* (Figure 5.36b). Although a full screen was not done with either gene fragment, both genes where shifted with a similar efficacy as the upstream regions of *bd0089*^{U} and *bd2625*^{U}. There is very little homology across the 8 DNA fragments Bd3720 has been shown to bind, adding additional support to the conclusion that Bd3720 exhibits fairly non-specific DNA binding properties.

Gel		Concentration of Protein in Well (μM)						
		1	2	3	4	5	6	7
a	Protein	Dye	-	-	Bd3720	Bd3092	-	Bd3720
	DNA	Dye	-	*bd3721*^{U}	*bd3721*^{U}	*bd3721*^{U}	*gltF*	*gltF*
b	Protein	Dye	-	-	Bd3720	-	Bd3720	
	DNA	Dye	-	*bd3721*^{U}	*bd3721*^{U}	*Bd1885NT*	*Bd1885NT*	

Table 5.11 Contents of each well in specificity test EMSA. Both Bd3720 and Bd3092 were used and a final concentration of 5 μM, while all DNA kept at a constant 25 ng per well.

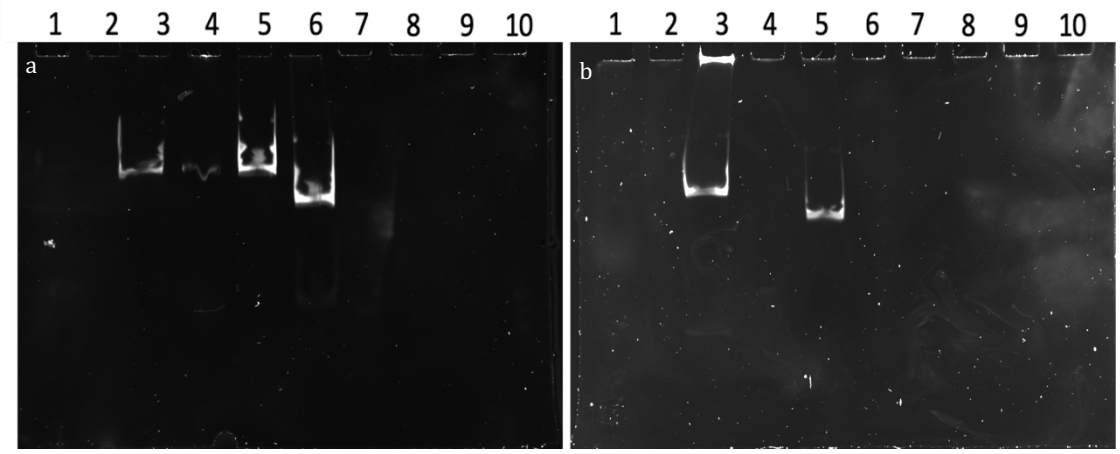


Figure 5.36 (a) EMSA testing binding of Bd3720 and chemotaxis protein Bd3092 to *bd3721*^{U}, showing that despite overall similar charge and size between Bd3720 and dimeric Bd3092, only Bd3720 is capable of DNA binding. EMSA also shows that Bd3720 is capable of binding to *gltF* gene fragments. **(b)** EMSA testing Bd3720 binding specificity, comparing the binding of Bd3720 to *bd1885NT* gene fragment against its binding to *bd3721*^{U}, with both binding to Bd3720.

5.2.5.4 Identifying DNA Sequences/Motifs Responsible for the Binding of Bd3720

Although the specificity tests seemed to show a lack of specificity to a particular DNA fragment, truncations of *bd3721*^{U} were produced in order to investigate if any sequence specificity could be elucidated. The truncations were designed by sequentially truncating *bd3721*^{U} by 50 bp from either the 5' or 3' end (Appendix X). All truncations were made using the original forward or reverse primer, and a new truncating primer, which the fragments were subsequently named after. Truncations F1-3 and R1-3, all fully gel-shifted at the tested ratio, whereas F4 and R4 only showed partial binding. F6, R5 and R6 all showed no binding to bd3720, while no F5 band was present, even in the DNA control (Table 5.12, Figure 5.37).

DNA fragment size		Concentration of Bd1172 in Well (μM)									
		1	2	3	4	5	6	7	8	9	10
Large	Bd3720 (μM)	Dye	-	0	10	0	10	0	10	0	10
	DNA			F1	F1	F2	F2	R1	R1	R2	R2
Medium	Bd3720 (μM)	Dye	-	0	10	0	10	0	10	0	10
	DNA			F3	F3	F4	F4	R3	R3	R4	R4
Small	Bd3720 (μM)	Dye	-	0	10	0	10	0	10	0	10
	DNA			F5	F5	F6	F6	R5	R5	R6	R6

Table 5.12 Concentration of Bd3720 in each well of fragment screening EMSAs along with the fragment screened. The final amount of DNA loaded in each well was kept constant at 50 ng.

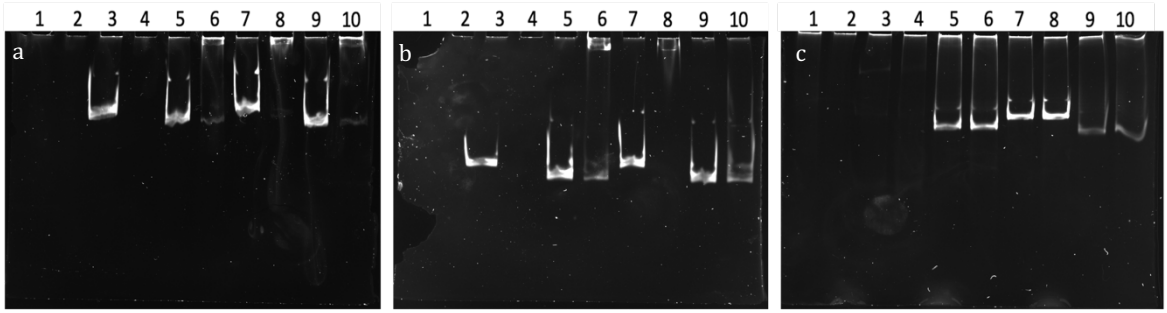


Figure 5.37 (a) EMSA testing Bd3720 binding to large *bd3721*^{U} truncation fragments; F1, F2, R1, and R2, all of which have been shown to bind. **(b)** EMSA testing Bd3720 binding to medium *bd3721*^{U} truncation fragments; F3, F4, R3, and R4. Both R3 and R4 have been shown to fully bind at this ratio of protein to DNA, whereas F4 and R4, although still shown to bind, have done so to a lesser extent. **(c)** EMSA testing Bd3720 binding to small *bd3721*^{U} truncation fragments; F5, F6, R5, and R6, all of which ran well, except F5 which is not present even in the control DNA only well. For the samples that did run, none of the bands shifted, indicating that Bd3720 seems unable to bind these fragments.

These experiments identified a 179 bp region that could potentially contain the bd3720 binding site(s) (Figure 5.38). Both the SEC profile of Bd3720 and the partial crystal structure of Bd2625 suggest that this transcription factor family forms tetramers, and the presence of two DNA binding sites in each monomer suggest that a total of 8 binding sites per protomer are possible. Eight sites, consisting of 6 nucleotides each, were identified by Dr. Jack Bryant (J Bryant 2020, personal communication, 5 November), positioned into two groups of four, with 28 bp between sites, the whole DNA motif totalling 131 bp, which would allow for the DNA to wrap around the Bd3720 tetramer as proposed by the Bd2625 partial structure (Figure 5.39). The two groups follow the same general pattern of: Site–11 bp–Site–6 bp–Site–11 bp–Site. Four of the sites contain a DNA adenine methylation (DAM) site (GATC); S1, S4 and S6 (tgatca) and S2 (tgatcc), whereas the other 4 possess high sequence similarity to these sites; S3 (tgagac), S5 (tgaaca), S7 (tgaact), and S8 (tggcga). Of these sites only S1 – 3 remain present in fragment R4, whilst S5 – 8 are present in fragment F4; this reduction of binding sites could explain the reduced binding ability seen when assayed with Bd3720. However, these sites do not explain Bd3720 binding to the other fragments (*gltF*, *bd0089*^{U}, *bd1172*^{U2}, *bd1885*^{U1}, *bd1885NT*, *bd2625*^{U}, *bd2626*^{U}). *gltF* does contain a possible 8 sites that follow the above pattern but *bd1885NT*

and *bd0089*^{U} contain only 2, and the other 4 fragments (*bd1172*^{U2}, *bd1885*^{U1}, *bd2625*^{U}, *bd2626*^{U}) contain no significant matches to this consensus.

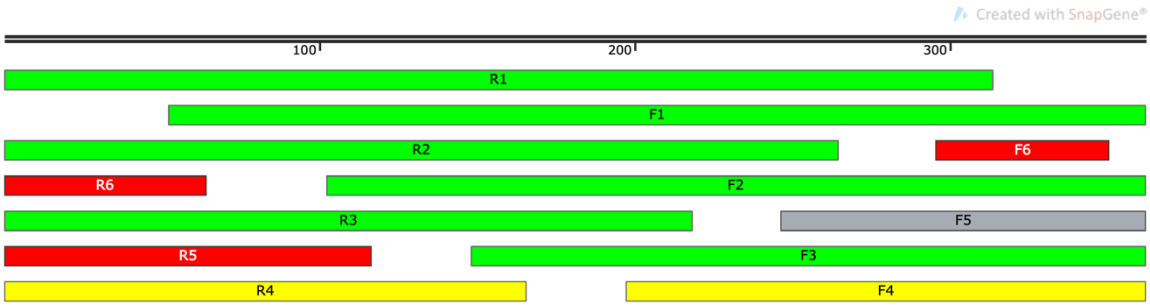


Figure 5.38 Section of *bd3721*^{U} of each truncation experiment. The result of the binding EMSA is represented by the fragment colour, green for full shift, yellow for partial shift, red for no shift, and grey for inconclusive.

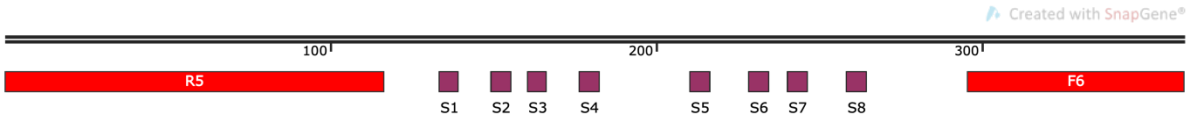


Figure 5.39 Position of the possible binding sites identified (S1-8) in purple. S1 – 3 overlap with R4 and S5 – 8 overlap with F4.

5.2.5.5 Competition EMSAs

In conjunction with the Bd3720 specificity screening, competition EMSAs with poly(deoxyinosinic-deoxycytidylic) (poly(dI-dC)) were carried out, in order to ascertain the specific binding *Kd* of Bd3720 with *bd3721*^{U}. These experiments were conducted with poly(dI-dC) in 2x excess to *bd3721*^{U} in order for the poly(dI-dC) to outcompete *bd3721*^{U} for the non-specific binding sites, meaning that any shift seen in the *bd3721*^{U} would be due to specific binding alone. Since the total binding qualitative EMSAs analysis estimated the *Kd* to be approximately 1.34 μM while visual gel inspection estimated it to be between 1.6 – 1.7 μM, the competition EMSAs were carried out with a Bd3720 concentration between 0.05–5 μM to ensure good datapoint distribution across the binding curve (Table 5.13). Unlike with the total binding gels, a steady shift of unbound to bound was seen, however the bands were less clean (Figure 5.40). This is due to background noise caused by the staining of the poly(dI-dC) by the ethidium bromide, as poly(dI-dC) is a mixed weight polymer, and although the average size is 1.2–3

kb, and would not be ran onto the gel, some of the lower weight polymer, <550 bp, will be, and upon staining will cause higher background. The gel images were analysed using Fiji (Schindelin et al., 2012), and normalised using the DNA only band as a fully unbound standard. However, a *Kd* of Bd3720 against the bound fraction of *bd3721*^{U} was unable to be accurately calculated due to poor staining of the bands due to the background noise caused by the poly(dI-dC).

Concentration of Bd3720 in Well (μM)														
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Dye	-	0	0.05	0.10	0.15	0.20	0.25	0.40	0.55	0.75	1.00	1.75	2.50	5.00

Table 5.13 Concentration of Bd3720 in each well of competition EMSAs performed with *bd3721*^{U}. The concentration of *bd3721*^{U} and poly(dI-dC) were kept constant at a final concentration of 10.5 nM and 2.5 ng/μl respectively.

a

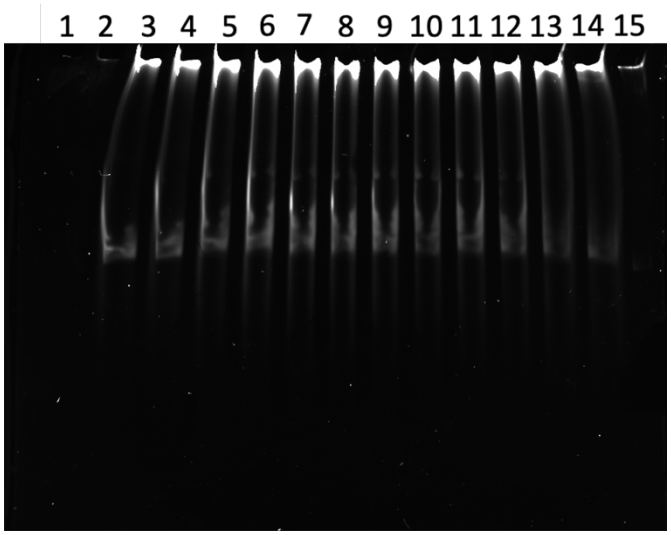


Figure 5.40 (a) Representative competition EMSA showing Bd3720 binding *bd3721*^{U}.

5.2.6 Identifying Amino Acids Involved in DNA binding

Sequence comparisons of the two predicted DNA binding sites across all 18 homologues within the *HD100* strain were aligned against HTH and WHTH reference proteins. This was primarily done in order to determine the amino acids putatively involved in DNA recognition, with a secondary interest in using this information to mutate out the corresponding amino acids to determine which of the two domains was more important for DNA recognition.

The amino acids putatively involved in DNA recognition via the HTH domain (Table 5.14), were estimated purely from the sequence alignment of the whole family against 3 HTH domain containing proteins; RsaL from *Pseudomonas aeruginosa* (5J2Y (Zhao *et al.*, 2017)), plcR from *Bacillus thuringiensis* (3U3W (Grenha *et al.*, 2013)), and cII from *Escherichia phage lambda* (1ZS4 (Jain *et al.*, 2005)). These HTH were chosen because a crystal structure of them bound to DNA has been determined, as such the predicted presence of the HTH domain had been confirmed, along with the specific amino acids involved in DNA binding. All are also predicted to have structural homology to the family based upon statistically-significant HHpred results, with a sequence ID% of 12% for RsaL & cII, and 11% for plcR (to Bd2625). The recognition amino acids in the WHTH-like motif of the DUF4423 (Table 5.14) were identified based on a combination of whole family alignment with RovA, a WHTH domain from *Yersinia pseudotuberculosis* (4AIK (Quade *et al.*, 2012)), and the partial model of Bd2625. RovA was used due to the same reasoning given for RsaL usage for the HTH domain. A composite model of the partial structure of Bd2625 superimposed upon IscR from *E. coli* (4HF1 (Rajagopalan *et al.*, 2013)) was produced. As previously stated, IscR had been identified by Topsearch (Wiederstein et al., 2014) as having a similar fold and oligomerisation state to a DUF4423 dimer. This agreement meant that superimposition could be performed by aligning the recognition helices of each dimer, L132–D144 in Bd2625 and S38–N52 in IscR. This produced a very good fit between one HTH motif in the dimer, however there was some deviation in the other. The RMSD between the two recognition helices, is smallest measured at 3.0 Å at L132 in the major groove and increases to 4.7 Å at D144 at the opposite end of the helix, as seen in Figure 5.24. This superimposition allowed for the DNA fragment crystallised with IscR to be used to predict which amino acids within the Bd2625 recognition helix would be in closest contact to the nucleotide bases. The recognition amino acids predicted from alignment (in bold): ^{132}**LEK**VQ**Q**SLELLLD^{144} in Bd2625, agreed well with those predicted from the composite model, ^{132}L**EK**VQ**Q**SL**E**LLLD^{144}. Of the two methods used to predict the amino acids key to DNA binding via the DUF4423 domain, the composite model predictions were favoured. This is because the composite model allows for the position of amino acids relative to a DNA strand to be

taken into consideration. However, the general agreement between the different methods of prediction, with regards to which amino acids are involved in DNA binding via the DUF4423 recognition helix, gives credence to both the structure- and sequence-based approaches. This cross-family comparison revealed higher sequence homology throughout the HTH recognition helix than the WHTH recognition helix.

Gene	Helix-turn-Helix alignment	Winged Helix-turn-Helix alignment
Bd0089	WAFELGFEHRSFLRQVVIGRRA-LTDTTAAQISQRM	-ADVLEVEKGIVILTRLNLAEQVG--A-GYRATVSSFKVPS
Bd1905	WTRHLGLSYPSALSMILNRKRH-IGDKLLETMCQYF	-PIL-G----------MKTSEVLALEP
Bd3700	FSAAANIKSPNYLKMIIEGKRN-LSDDMIGKFGKAL	-ASEDEIDNALTTLITSGDLRRDEVTG-ELKKARSLTESPE
Bd3720	VAKEAGMA-SGYLPMCFSRKRK-LSKRSFVKLLPHL	RISQKEIEEALAFLVQFKFIEKNA--DGKYLVVQKQLDCHE
*Bd0681**	LSEAANTQ-KSYLSACLNGKGQ-LSLDQAFGVAEYL	-LNQASVATIIGYLEKMELIKKSG--D-KYRWNSSNIHLE-
Bd1537	LVEAMRIS-TSLFSQILKGEKN-LSSEQGLEAAAFF	-VPLEKVDTAIQFLIEVGLCERKN--G-QLYYKQGF--THL
Bd1885	FSRDLEVP-CSRLSEILNGKVG-LSESRAVNLATKL	-LLTDVTEKALERLEAMKLIEKTA--DGKWTSNDAKYSSSY
*Bd0724**	FARLLKTD-PSSLSKVLNGLRI-PSDETVEKWVDCM	-MTEEQIQKTIEYLESHKILHRST--D-NGNYLPSNLTT-V
*Bd2774**	FANSLNID-AGSLARVMSGQRV-VKDATARQLMTGL	-ISIEELQKRLLELERLGLLSATD--T-HMV----------
*Bd2775**	FAKSLDQD-AGTLTRVMSGQRV-VKQPTALAVMQAL	-LAPEELESRLISLEKIGAIAIKN--S-TVEVISKNLST-L
*Bd1172**	FAQSLSIS-PAQLSQLLSGKRN-FTTEVLGTISRGL	-ISVSDARAALERLRRMDIIED-G--K-ALKQKSAPLNVVS
*Bd2560**	LALKAQIP-SGRVSEILNGTRP-LSPYYKGKITQAL	-ITTTQARSALKRLQKLKLIEEHQ--G-RFVRTNTFLTTSK
*Bd3758**	FAKILGIP-SSRLSDYINGRRI-MTMSVGKQVIKGL	-IPVEVAQAAIERLCRLGLLQIED--G-KFIVTHKQLETSH
*Bd1219**	LAQRVEIS-PGHLSEIFNNKRP-VSSANFNKLVAGL	-VPSERTQTVMNKLLQLGLVEMNS--PTTYVKTVRSLASES
*Bd1897**	YAQFLEMN-PGTLSSIIKGKRL-VPLQEAAKVMQKL	-IPAERVNACLQTLQDLQVITRKG--P-VWERTHKNLYSPH
Bd1901	LAKKVAIS-PGCLNELMHGKRP-LSEFYASKIVLGL	-IPLEKVQHSLDLLMNLGFIKMQG--T-TFVRTVASLGTTS
*Bd2625**	LAKKVGIS-PGCLNELMHGKRP-LSEFYANKIVLGL	-LPLEKVQQSLELLLDLGFIKRKG--N-SIARSVASLATTT
*Bd2551**	IAARMGIS-SGRLSEILSEKRP-LTAYYADKFCAAL	-LTTRELTVLLDAMTQTGLIEWDN--S-AWRPTHNEATTGY
RsaL	FWSRFGIS-QSCGSRFENGENLPFPIYL----LLHF	
RovA		-IEQPSLVRTLDQLEEKGLITRHT--S-ANDRRAKRI----

Table 5.14 Comparison of the 18 transcription factors' DNA binding sites along with HTH reference RsaL (5J2Y (Zhao et al., 2017)), from *Pseudomonas aeruginosa*, and winged helix reference RovA (4AIK (Quade et al., 2012)), from *Yersinia pseudotuberculosis*, which were used to identify the possible binding sites. Each 'subfamily' is grouped by colour, and the DNA binding domain amino acids are coloured the same, bold black for the reference. The region in bold is the entire sequence of the recognition helix, highlighted yellow for the reference structure.

A comparison of a single TF against homologous proteins from other *Bdellovibrio* strains was performed to see if this relationship between the two domains continued to hold true. Bd1885 was selected for this analysis, as it was considered the most representative of the family in terms of matching the family sequence consensus. Using BlastP (Altschul et al., 1990), likely Bd1885 homologues were identified for 8 different *Bdellovibrio* strains (Table 5.15). The WHTH recognition amino acids were again shown to be more variable than the HTH recognition amino acids, as overall sequence identity decreased. Since these homologues would be presumed to have similar functions

and specificity across the different strains, this increased variation seen in the WHTH compared to the HTH may be suggestive of an alternative specificity mechanism. A study by McIver and Myles on a double DNA binding domain protein, Mga, found that the HTH_4 domain was conserved across all Group A *Streptococcus* (GAS) homologues whilst the HTH_3 domain was not. They concluded that as Mga fulfilled the same role in all strains the HTH_4 domain was more critical for DNA binding than HTH_3 (McIver and Myles, 2002). Since a similar conservation usage/difference is observed here between the HTH and DUF4423 domains, this could be suggestive of a similar importance being placed on HTH over the DUF4423 in terms of DNA binding. However, where GAS has only one Mga, *Bdellovibrio* has a family of these TFs, therefore the HTH domain may be more critical for the sustained binding of DNA, as per the McIver and Myles study, whereas the variation in the DUF4423 region may be responsible for the initial recognition of a DNA binding site.

Organism	Protein	Helix-turn-Helix alignments	Winged Helix-turn-Helix alignment	Sequence Identity (%)
Bdellovibrio bacteriovorus HD100	Bd1885	RDLEVP**C**-**S**RL**SE**ILNGKVGL	AQRLGLLT**DV**TE**K**AL**E**RLEAMKLIEKT	–
Bdellovibrio tiberius	Bdt_1852	RDLEVP**C**-**S**RL**SE**ILNGKVGL	SRRLGLLT**DV**TE**K**AL**E**RLEAMKLVEKT	93.63
Bdellovibrio sp. SSB218315	B9G79_09665	RDLEVP**C**-**S**RL**SE**ILNGKVGL	AQRLGLTL**DV**AE**K**AI**E**RLEAMKLVEKN	91.01
Bdellovibrio bacteriovorus sp. ArHS	OM95_11865	RDLEVP**C**-**S**RL**SE**ILNGKVGL	SQRLGLTL**DT**TE**K**AL**E**RLENVGLVTKS	73.58
Bdellovibrio bacteriovorus 109J	AZI87_12470	RDLEVP**C**-**S**RL**SE**ILNGKVGL	SQRLGLPV**DT**VE**K**AL**E**RLQNVGLVTKS	73.31
Bdellovibrio W	BDW_01115	QALGVS**SK**S**F**L**RF**AILGKRTI	ADLLSVSA**EE**VM**G**HL**E**ELKNQGFV-KC	35.36
Bdellovibrio sp. SKB1291214	B9G69_16615	RDLELN**A**-**S**RL**SE**IMNGKVGL	ADRLGLPH**YE**VD**D**IV**A**RLVENKMLDT-	33.96
Bdellovibrio exovorus	A11Q_1193	RDLNFS**N**-**S**RL**SY**ILNKSAGI	SAILGVKE**VV**VK**K**AL**R**QLVDNGYLLQK	28.57
Bdellovibrio sp. Zap7	DOM_06770	RDIGVS**A**-**S**RL**SE**LLNSKVGL	AKKLGISV**AE**AE**A**TV**A**RLMKVGLLKE-	22.05

Table 5.15 Comparison of Bd1885 from *Bdellovibrio bacteriovorus* HD100 to its homologs in different *Bdellovibrio* strains. The predicted DNA binding domain amino acids are in bold black and the recognition helix are highlighted in yellow.

5.2.7 Mutation of Bd3720

5.2.7.1 Cloning and Purification of Bd3720 polyA recognition helix mutants

Having identified the amino acids in the two DNA recognition helices that could be potentially involved in DNA binding, poly-alanine (polyA) recognition helix mutants of Bd3720 were created in order to elucidate how the domains were involved in binding. Three mutants were produced; a polyA variant of the HTH recognition helix (polyA HTH) (^{47}AAYLAA^{52}), a polyA variant of the WHTH recognition helix (polyA WHTH) (^{158}AAAIEA^{163}), and a polyA variant of both recognition helices (polyA Double) (Table 5.16). As stated in the previous section the amino acids predicted to be involved in DNA recognition within the HTH domain were based solely from sequence alignments - for Bd3720 these were S47, G48, P51 and M52. However, within DUF4423 there is some discrepancy between the alignment predictions and the composite model, as a result the amino acids predicted to bind DNA by both methods were marked for mutation. In the specific case of Bd3720, this resulted in only the amino acids predicted by alignment being mutated; Q158, K159, E160, and E163, as the final amino acid identified by the composite model, A166, was already an alanine. All mutants were confirmed by sequencing.

Mutant	polyA HTH	polyA WHTH
Original Sequence	KGFS**LRQVAKE**AGMA**SGYLPMCF**SRKR	FKED**PAWIQDRL**RGRIS**QKEIEEALAFLVQ**FKFI
Mutated Sequence	KGFS**LRQVAKE**AGMA**AAYLAACF**SRKR	FKED**PAWIQDRL**RGRIS**AAAIEAALAFLVQ**FKFI

Table 5.16 A table showing the polyA mutation of the HTH recognition helix and the WHTH recognition helix. The polyA Double mutant contained both recognition helix mutations. The amino acids shown in bold make up the helices within the HTH motif of both domains, whilst those underlined are in the turn region. Highlighted in red are the amino acids that were mutated.

Similarly to wild-type Bd3720, Bd3720 polyA HTH was expressed in AI media with sugar mix N and purified using nickel affinity chromatography and eluted with 500 mM Imidazole (Figure 5.41a). The eluted fractions of Bd3720 polyA HTH were supplemented with 2.5 mM β-mercaptoethanol, prior to dialysis into 20 mM Bicine pH 8.0, 500 mM NaCl and 2.5 mM β-mercaptoethanol, and was stable enough for SEC. SEC was performed using a HiLoad 16/60 Superdex 200 pg column, with the aim of

purifying the sample further and also confirming correct folding and tetramerisation. A large peak was eluted at 65 ml, similar to the 60 ml at which the Bd3720 WT tetramer peak was eluted, indicating that Bd3720 polyA HTH still formed a tetramer (Figure 5.41c). A few other small peaks came off before and after which consisted of contaminating proteins.

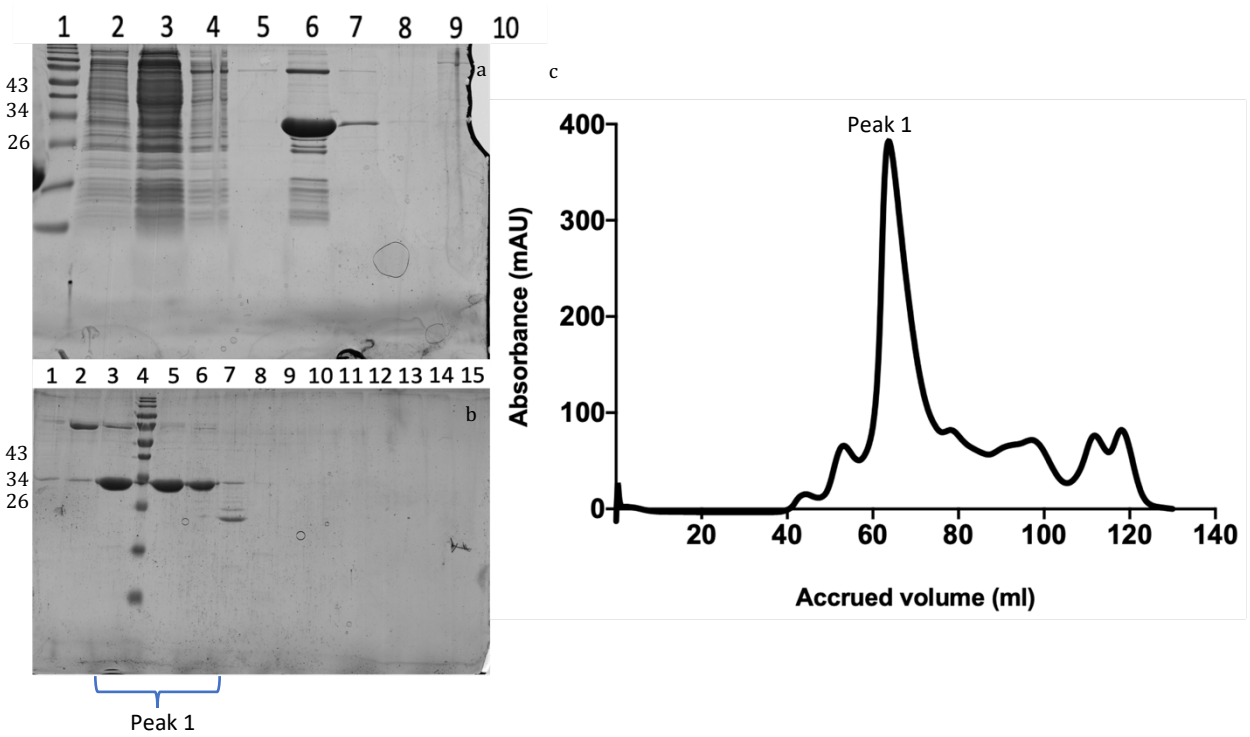


Figure 5.41 (a) SDS-PAGE gel of Bd3720 polyA HTH IMAC purification. **(b)** SDS-PAGE gel of Bd3720 polyA HTH post SEC purification. **(c)** UV trace of Bd3720 polyA HTH SEC, showing tetrameric Bd3720 polyA HTH (peak 1) being separated from contaminant proteins.

Bd3720 polyA WHTH was expressed and purified in the same manner as Bd3720 polyA HTH and WT (Figure 5.42a). Identically to Bd3720 polyA HTH, Bd3720 polyA WHTH eluted as one major peak alongside some much smaller peaks (Figure 5.42c). The large peak at ~60 ml is again consistent with correct folding and tetramerisation.

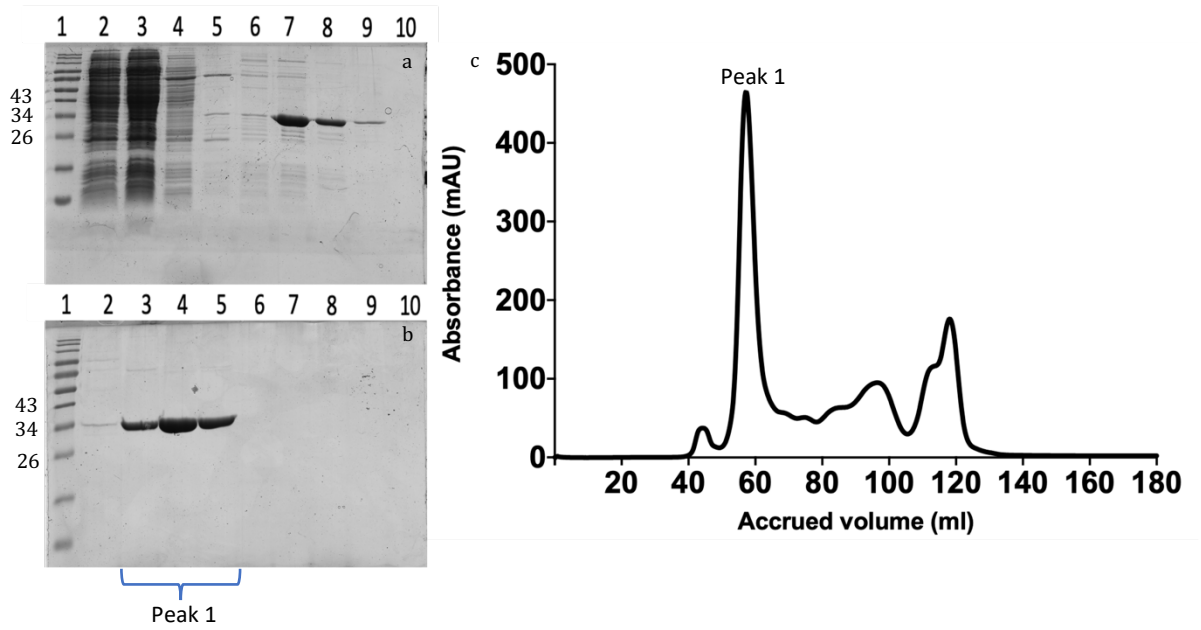


Figure 5.42 (a) SDS-PAGE gel of Bd3720 polyA WHTH IMAC purification. **(b)** SDS-PAGE gel of Bd3720 polyA WHTH post SEC purification. **(c)** UV trace of Bd3720 polyA WHTH SEC, showing tetrameric Bd3720 polyA WHTH (peak 1) being separated from contaminant proteins.

Bd3720 polyA double was expressed and purified in the same manner as Bd3720 WT and the other two mutants (Figure 5.43a). Identically to the other single-site mutants, Bd3720 polyA double eluted as one major peak, alongside some smaller peaks (Figure 5.43c). The large peak at 60 ml confirmed that double-site mutation had not affected the ability to tetramerise.

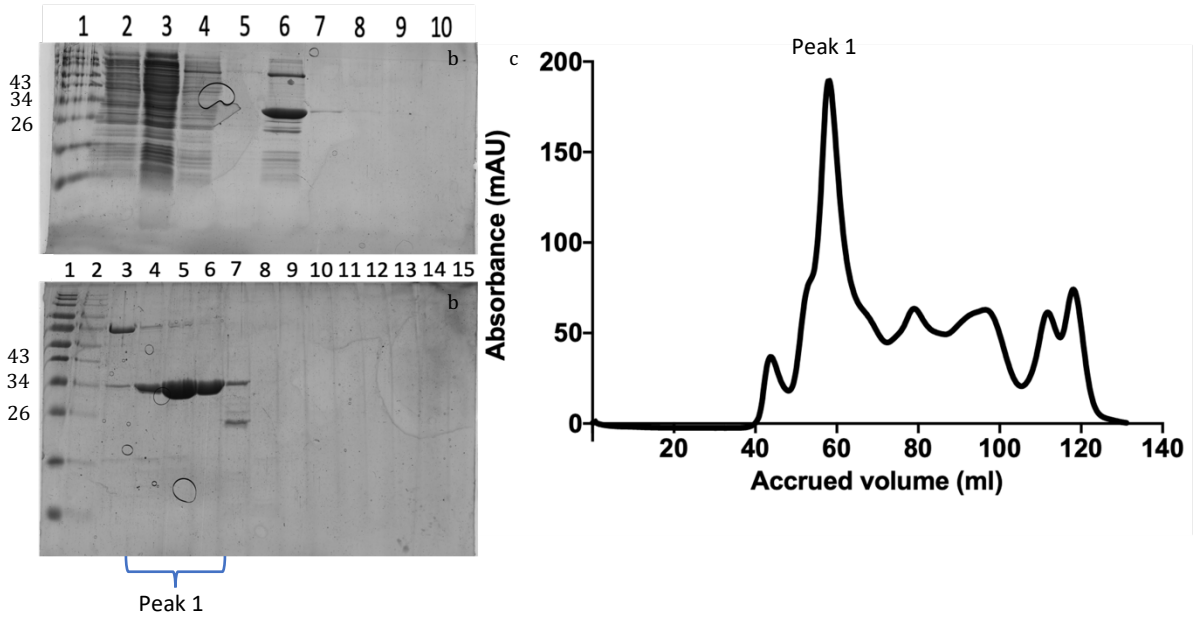


Figure 5.43 (a) SDS-PAGE gel of Bd3720 polyA Double IMAC purification. **(b)** SDS-PAGE gel of Bd3720 polyA Double post SEC purification. **(c)** UV trace of Bd3720 polyA Double SEC, showing tetrameric Bd3720 polyA Double (peak 1) being separated from contaminant proteins.

5.2.7.2 Circular Dichroism Spectroscopy

To complement the SEC results that suggested the mutants were still capable of tetramerisation, circular dichroism was used to show that all mutants still had a similar secondary structure profile to Bd3720 WT, and therefore demonstrate that the polyA mutations had not significantly affected folding. CD is a good, non-invasive method to determine the nature of folded protein, distinguishing between α-helix and β-sheet content. All proteins were dialysed into 10 mM Sodium phosphate pH 7.5 and 100 mM $(NH_4)_2SO_4$, to reduce interference to the CD run caused by the SEC buffer. Firstly, a Bd3720 WT spectra was obtained (Figure 5.44a). The CD spectra trace produced had a peak with a double minima, one at 210 nm and another at 223 nm, similar to minima at ~208 nm and ~222 nm that are characteristic of a predominantly α-helical protein (Greenfield, 2007). However, analysis of this trace using Dichroweb (Miles et al., 2021) CDSSTR method (Compton and Johnson, 1986) using the SMP180 reference set (Abdul-Gader et al., 2011) (Fitted plots included in Appendix XI) reported only 8% helical content (Table 5.17a). This contrasts with both the predicted helical content from Xtalpred (Slabinski et al., 2007), which is 54% (Table 5.17b), and the partial Bd2625 crystal structure which depicts 16.5% of the helical content of the whole protein (Figure 5.14). Both values exceed the helical content reported by CD, therefore indicating that the CD data may not be reliable, despite good fitting of the data by the reference set (Appendix XI). Despite this, traces were still obtained for all three mutants and analysed in the same manner as the WT. Analysis was performed in the same way as for Bd3720 WT, because the data had been well fitted (Fitted plots included in Appendix XI). This analysis reported a conservation in the percentage of strands and unordered regions for all three mutants in comparison to WT (Table 5.17a). Further structural conservation was seen for Bd3720 polyA HTH, which reported the same percentage of helical content and similar percentage of turns. Bd3720 polyA WHTH and polyA Double both reported a significant decrease in helical content in comparison to WT, with polyA WHTH showing an increase in Turn and Unordered content and polyA Double showing an increase in just Turn content (Table 5.17a), suggesting that both polyA WHTH and polyA Double mutations altered the structure of Bd3720. However, the structural content predicted

by Xtalpred (Slabinski et al., 2007) (Table 5.17b) for each mutant was similar to that of WT, and therefore were similarly significantly different to the structural content reported from the CD data (Table 5.17a). As a result of this screening EMSAs with the mutants were performed, as it could not be reliably determined whether the CD data reported a true reflection of the structural content of the mutants, because it was unlikely to be representative of the WT.

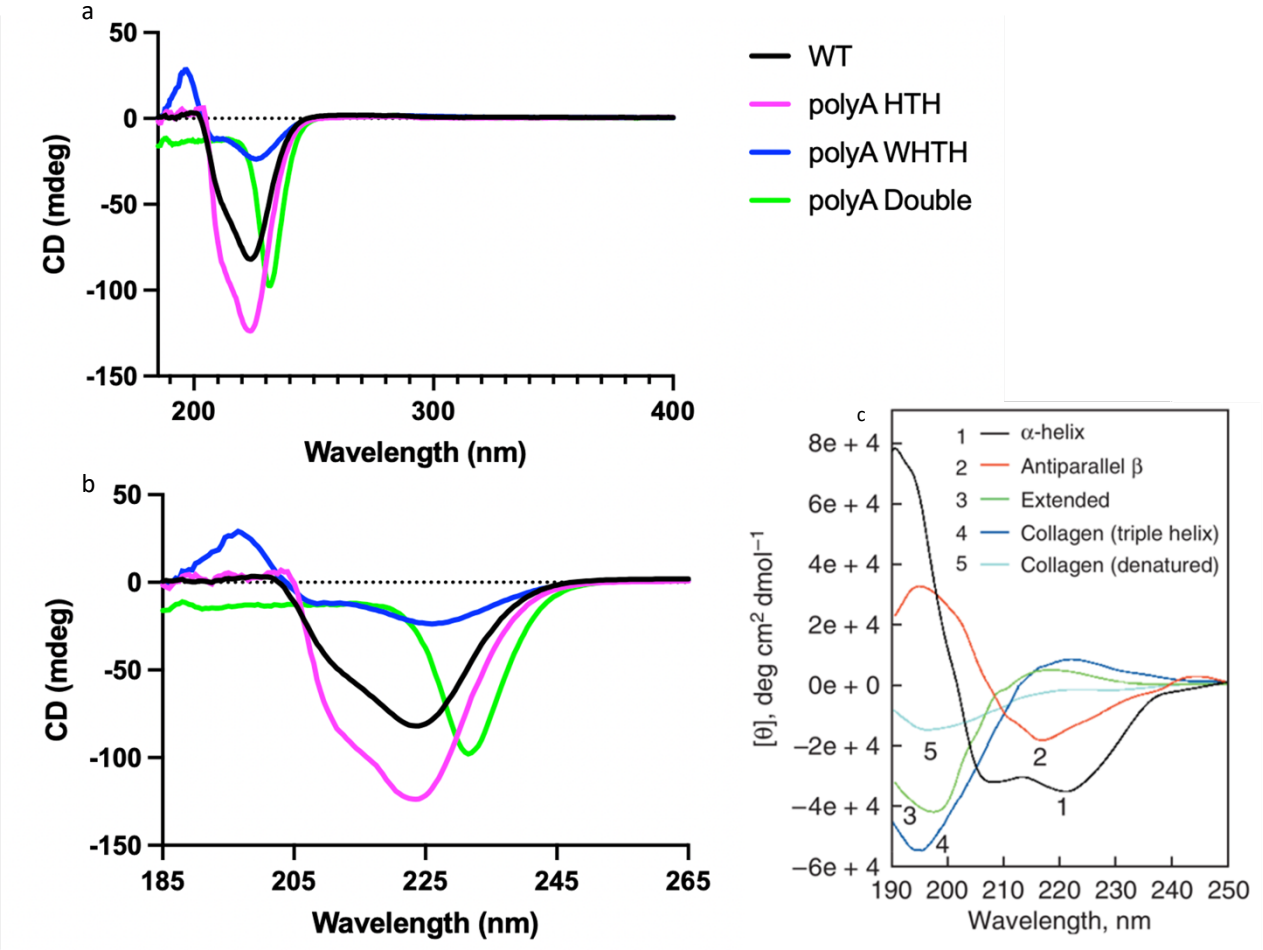


Figure 5.44 CD spectra for Bd3720 WT and polyA mutants. All traces, with the exception of Bd3720 polyA Double, have a similar peak with double minima at ~208 nm and ~222 nm, standard of α-helical proteins. **(a)** Full CD spectra. **(b)**Close-up of wavelengths where α-helical minima were found. **(c)** Standard CD curves created using poly-l-Lys for secondary structure comparison, taken from Greenfield, 2007.

a

Protein	Helix (%)	Strand (%)	Turns (%)	Unordered (%)	Total (%)
Bd3270 WT	8	41	11	41	101
Bd3720 polyA HTH	8	41	10	42	101
Bd3720 polyA WHTH	4	41	12	43	100
Bd3720 polyA Double	3	41	14	41	99

b

Protein	Helix (%)	Strand (%)	Loops (%)	Unordered (%)
Bd3270 WT	54	11	24	11
Bd3720 polyA HTH	54	11	23	12
Bd3720 polyA WHTH	53	12	23	12
Bd3720 polyA Double	53	12	24	11

Table 5.17 (a) CD spectra analysis for Bd3720 WT and polyA mutants. Analysis was preformed using Dichroweb server (Miles et al., 2021) and SMP180 (Abdul-Gader et al., 2011) as a reference dataset. The percentage of both regular and distorted helices (Sreerema et al., 1999) were combined and listed as Total Helix, and the same was done for the regular and distorted sheets (Sreerema et al., 1999) which were combined to produce Total Strand. **(b)** The secondary structure was predicted by Xtalpred (Slabinski et al., 2007) for Bd3720 WT and Mutants. This shows a large disagreement between the CD results and the predicted secondary structure, the latter of which is in line with partial crystal structure.

5.2.7.3 Electrophoretic Mobility Shift Assays

In order to test if the polyA mutations of the domains influenced DNA binding, screening EMSAs were performed on the mutants, using a broad range of 1000:1–0.0001:1 protein to DNA concentrations (Table 5.18). All mutants were ran with *bd3721*^{U}, so the effects of the polyA mutations could be directly compared to the competition EMSA of Bd3720 WT. Bd3720 polyA WHTH and Bd3720 Double bound *bd3721*^{U} with a similar efficacy as observed between Bd1172 and *bd1172*^{U2} and Bd1885 and *bd1885*^{U2} when similar screening experiments were performed, fully shifting the DNA at some point between a 10–100:1 Protein:DNA concentration ratio (Figure 5.45). Although Bd3720 polyA HTH did not fully shift DNA at a 100:1 protein to DNA concentration, the band was very faint, and therefore could be due to experimental variation, and as such a similar binding capacity for *bd3721*^{U} by all mutants was assumed (Figure 5.45a). As these EMSAs only screened the effect the binding site mutation had on DNA binding no *Kd* calculation was made.

DNA fragment	Concentration of Bd3720 polyA mutants in Well (μM)									
	1	2	3	4	5	6	7	8	9	10
bd3721^{U}	Dye	-	0	0.00001	0.00011	0.0011	0.011	0.11	1.05	10.5
	Ratio of Protein:DNA									
	Dye	-	0:1	0.001:1	0.01:1	0.1:1	1:1	10:1	100:1	1000:1

Table 5.18 Concentration of protein used was the same for all Bd3720 polyA mutants. The concentration of *bd3721*^{U} was kept constant at a final concentration of 10.5 nM.

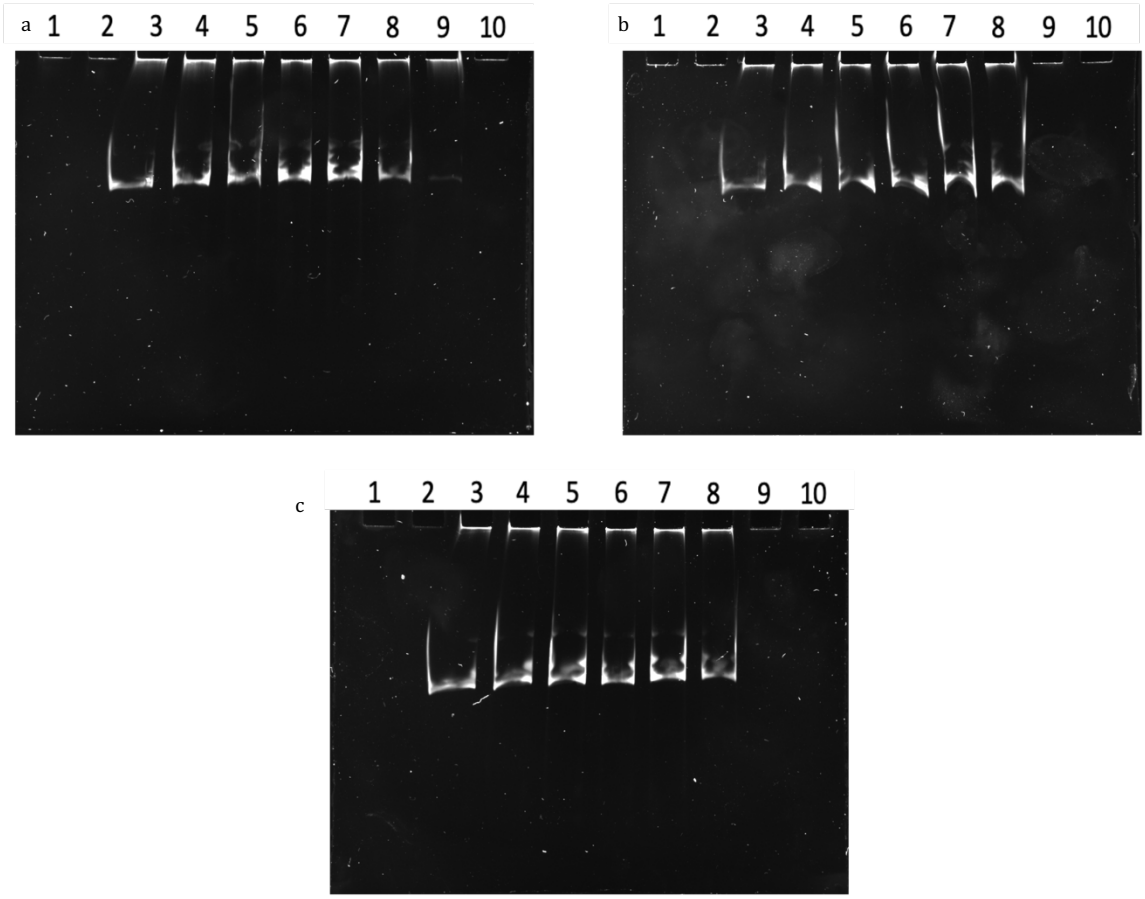


Figure 5.45 Screening EMSAs of Bd3720 polA mutants against *bd3721*^{U}. **(a)** Bd3720 polyA HTH binding *bd3721*^{U}, with almost full binding occurring at a protein to DNA ratio between 10 - 100:1. **(b)** Bd3720 polyA WHTH binding *bd3721*^{U}, with full binding occurring at a protein to DNA ratio between 10 - 100:1. **(c)** Bd3720 polyA Double binding *bd3721*^{U}, with full binding occurring at a protein to DNA ratio between 10 - 100:1.

A similar experiment was ran with the HTH domain only truncation of Bd1885, Bd1885NT, against *bd1885*^{U1}. The gel was ran in the same way as the Bd3720 polyA mutants, however unlike them Bd1885NT did not bind the DNA with any real avidity (Table 5.19, Figure 5.46a). In fact, it was not until the gel was analysed and the plotting of the affinity curve of Bd1885NT against *bd1885*^{U1} that any shift was apparent (Figure 5.46b). The affinity curve showed a weak trend similar to that of a binding event,

however no *Kd* could be estimated from it. Although the polyA mutations did not disrupt the binding between the recognition helices and the DNA, the removal of one of the binding domains did have a detrimental effect on binding. This observation could be for a number of reasons including: the loss of cooperative binding that is normally present between the two DNA binding domains, or the loss of ability to form higher oligomers due to the removal of the region that allows for the tetramerisation, which the DUF4423 has been shown to be at least involved in. The latter reason is unlikely given our coherent SEC profiles for the WT and mutant proteins.

DNA fragment	Concentration of Bd1885NT in Well (μM)									
	1	2	3	4	5	6	7	8	9	10
bd1885^{U1}	Dye	-	0	0.00001	0.00011	0.0011	0.011	0.11	11.1	11.1
	Ratio of Protein:DNA									
	Dye	-	0:1	0.001:1	0.01:1	0.1:1	1:1	10:1	100:1	1000:1

Table 5.19 Concentration of Bd1885NT polyA mutants used the was the same for all as the same DNA fragment, *bd1885*^{U1}, was used for all protein concentration is dependent on DNA concentration. The concentration of *bd1885*^{U1} was kept constant at a final concentration of 10.5 nM.

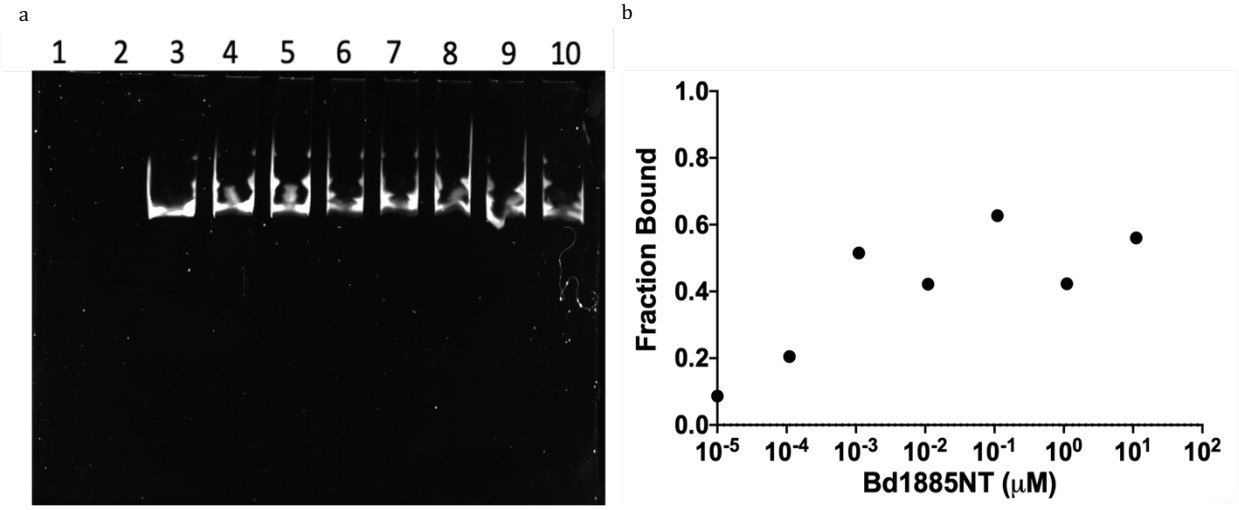


Figure 5.46 (a) EMSA showing the binding of Bd1885NT with *bd1886*^{U1} **(b)** The affinity of Bd1885NT to *bd1885*^{U1} plotted using GraphPad Prism, a weak binding trend can be seen, but a *Kd* could not be estimated.

5.2.8 Cloning and Purification of Effectors

5.2.8.1 Bd2625/Bd2624

Additionally to the transcription factor family three of the effectors, which are encoded downstream of the TFs, were also chosen for investigation. This was initiated via an attempt at co-expression of Bd2625 and Bd2624. Bd2625, is one of the 7 proteins that have a 5 nucleotide overlap with a downstream effector, CGCTATGAtc. Cloning of the genes using the native overlap was successful with both genes being cloned together into pET28a. pET28a was chosen as it allowed for a thrombin cleavable N-terminal his-tag to be added to Bd2625, whilst a non-cleavable C-terminal his-tag was added to Bd2624. This would have allowed for both to be co-purified using affinity binding to the his-tags, and then separated by the removal of the tag from Bd2625. However, although cloning was successful, the expression in LB and AI was not successful. Even so, purification was attempted in case expression was merely poor, nevertheless, neither Bd2625 or Bd2624 were observed on the gel (Figure 5.47). It was theorised that the expression of them as a pair would increase expression, similar to the expression of VP1 and VP2 in *Feline Calicivirus* (Luttermann and Meyers, 2007), however, this appears to not have occurred.

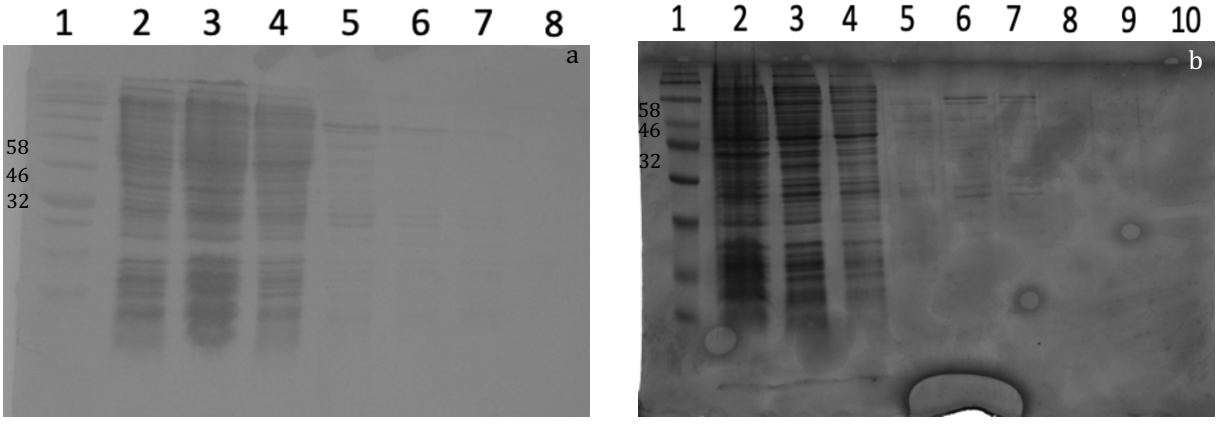


Figure 5.47 (a) Expression of Bd2625/Bd2624. The first three lanes use AI media, there is no change between pre-induction, lane 1, and post-induction, lane 2 and 3. The second three use LB media, again there is no change between pre-induction, lane 4, and post-induction, lane 5 and 6. All gel samples were normalised with respect to their OD using Equation 1 outlined in Chapter 2.5.2 **(b)** Purification was still attempted in case the genes were just poorly expressed indistinguishable from the background. This was not the case, and neither Bd2625 or Bd2624 were produced.

5.2.8.2 Bd1896

After the lack of success with Bd2625/Bd2624 co-expression, focus was moved to Bd1896. Previous studies of Bd1896 by Jack Wright (prior Lovering lab project student), demonstrated expression and purification in large yields, although this resulted in very unstable protein (Figure 5.48). This was compounded by a significant amount of contamination still being present post-IMAC purification, meaning Bd1896 required SEC before any further experiments could be done. Similarly to the TFs, Bd1896 has 3 cysteines in the protein sequence, meaning at least one was likely to be unpaired. Therefore a similar approach to the TFs was applied during IMAC purification, where fractions eluted with 500 mM Imidazole were supplemented with 2.5 mM β-mercaptoethanol. Although this stopped immediate precipitation, Bd1896 remained too unstable for SEC. Due to these effectors being predicted to have a similar role to the M48 protease in the HTH-DUF4423-M48 predicted protein, and the theory that they contain a divergent HExxH domain (^{157}HELIY^{161} in Bd1896), Bd1896 elution fractions were subsequently supplemented with 1 mM EDTA as well as 2.5 mM β-mercaptoethanol. It was hypothesized that some of the Ni^{2+} ions in the IMAC purifications were being stripped from the resin by Bd1896 and were then causing the protein to aggerate. The addition of both 1 mM EDTA and 2.5 mM β-mercaptoethanol caused Bd1896 to be stabilised enough for SEC using a HiLoad 26/60 Superdex 75 pg, with SEC buffer; 20 mM Bicine pH 9.0, 500 mM NaCl, 1 mM EDTA, and 2.5 mM β-mercaptoethanol. This removed most of the contamination, resulting in one elution peak at 125 ml, with two shoulders caused by the small amount of contamination. Although the protein was stabilised enough for SEC, it still could not be frozen directly, and so 5% glycerol was added to the SEC buffer to enable this. Crystallisation was attempted, but no crystalline material was produced.

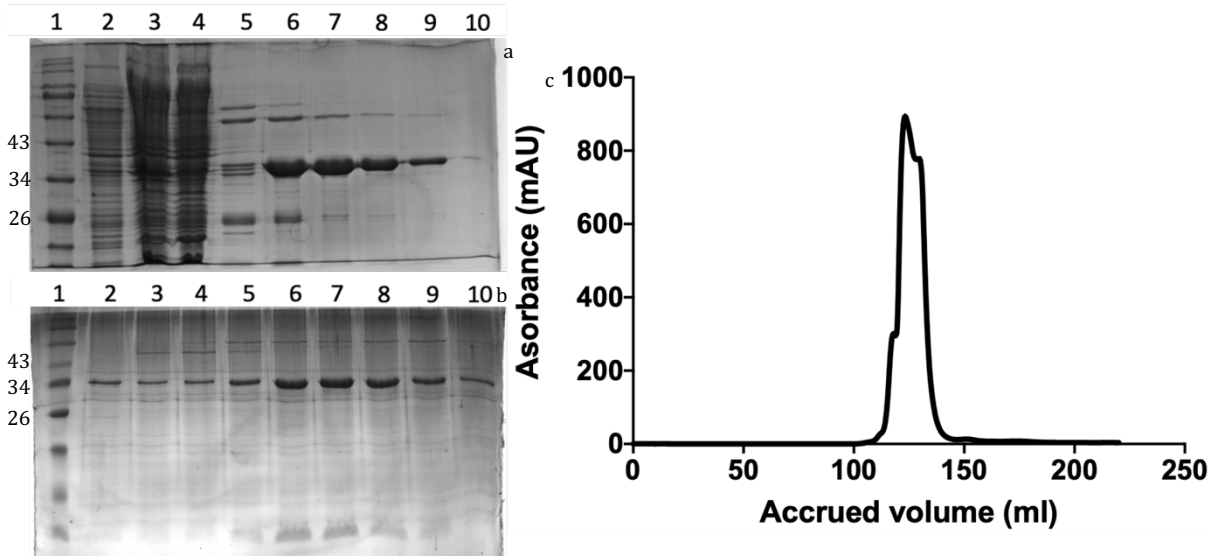


Figure 5.48 (a) SDS-PAGE gel of Bd1896 IMAC purification. **(b)** SDS-PAGE gel of Bd1896 post SEC purification. **(c)** UV trace of Bd1896 SEC, showing tetrameric Bd1896 being separated from contaminant proteins.

5.2.8.3 Bd0723 E122A

Lastly, two mutants of Bd0723 were designed. These aimed to mutate out the divergent proposed HExxH metal binding motif (^{121}HELLP^{125} in Bd0723), by mutating E112 to either A or Q, which is isosteric. It was thought that the removal of the E122 would prevent Ni^{2+} ions from purification destabilising the protein, and also make the protein itself less toxic to the *E. coli* expressing it, increasing the likelihood of correct folding. Bd0723 was chosen for this over Bd1896, as it had a cluster of 6 cysteines, likely to pair, which would hopefully make the protein more stable without the need for additives. Cloning of Bd0725 E122A was successful, however Bd0725 E122Q was not. Bd0723 E122A was expressed in AI media with sugar mix N and purified with a high yield using nickel affinity chromatography, eluted with 500 mM Imidazole (Figure 5.49). No additional additives were required and Bd0725 E122A was dialysed into 20 mM Bicine pH 9.0 and 500 mM NaCl overnight based on pI similarity to Bd1896. However, Bd0723 E122A precipitated overnight, albeit with enough protein remaining for a SEC run, after centrifugation to remove aggregate. SEC was performed in 20 mM Hepes pH 7.5 and 500 mM NaCl. A small peak at 120 ml consistent with previous effector Bd1896 SEC,

was thought to be Bd0723, whilst the later peaks were attributed to contaminant. However, due to the low yield post-dialysis, this could not be confirmed using SDS-PAGE. Bd0723 E122A purification was repeated using 20 mM Hepes pH 7.5 and 500 mM NaCl SEC buffer for dialysis. However due to time constrains SEC was not repeated, and all crystallisation attempts using pure protein from IMAC have been unsuccessful.

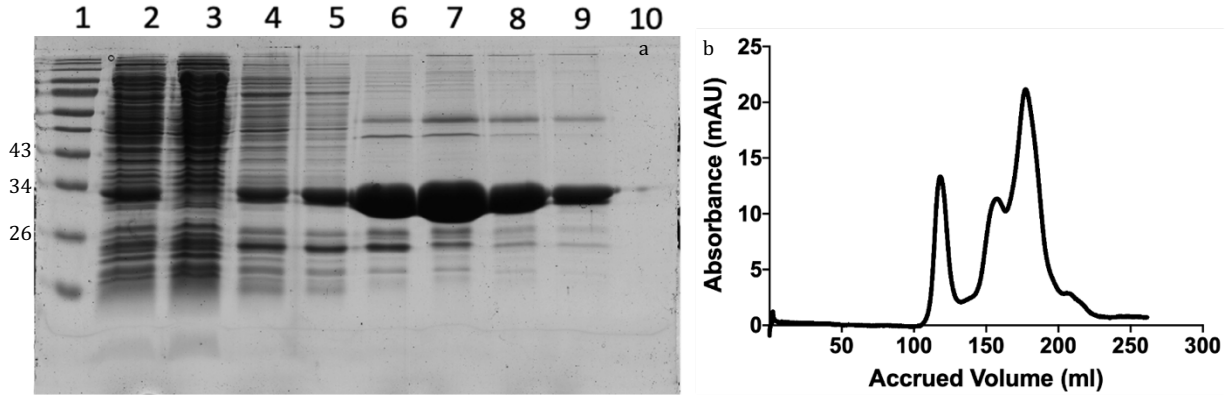


Figure 5.49 (a) SDS-PAGE gel of Bd0723 E112A IMAC purification. **(b)** UV trace of Bd0723 E112A SEC, showing tetrameric Bd0723 E112A (peak 1) being separated from contaminant proteins.

5.3 Discussion

5.3.1 Proteins with Multiple, Tandem DNA-binding Domains

This family of transcription factors is not the first set of proteins to be characterised with multiple DNA domains within a single protein. Searches of both Uniprot (The UniProt Consortium, 2019) and the literature show that the specific DNA motif has very little bearing on whether multiple DNA binding domains are predicted, however, these examples reveal that both domains commonly originate from the same family.

Multiple DNA binding domains are not restricted to prokaryotes, and are found in some of the most essential pathways across all life. DAI (DNA-dependent activator of IFN-regulatory factors) found in mammalian cells regulates the immune response via the binding of foreign DNA to both of the Z-DNA binding domains (ZBDs) (Figure 5.50a). Binding to both of the ZBDs is required for full conversion of Z-DNA to B-DNA, which is thought to trigger IFN production. As well as the two ZBDs, there is a third DNA binding domain. This domain does not follow the general rule of all the DNA binding domains originating from the same family, and very little is known about its role in the immune response (Ha 2008). OCT4 contains two tandem POU DNA binding domains, POU_S and POU_{HD}, which are related to the HTH superfamily (Figure 5.50c) (Yesudhas et al., 2017). When OCT4 comes into contact with free DNA, its two POUs bind the DNA nucleotides at two tandem sites. However, when working in conjunction with SOX2, OCT4 has been shown to be involved in nucleosome unwrapping. OCT4 achieves this by binding with POU_S to the entry-exit of the nucleosome, while SOX2 binds upstream. This action causes the DNA to peel away from H2B. The continued interaction of the DNA with H2A prevents POU_{HD} from binding. However, as these observations come from crystal structures, the role of POU_{HD} is still unclear, but in theory it could interact with the DNA if it was further unwrapped from the nucleosome (Michael et al., 2020). Wor1 has been shown to be upregulated in *Candida albicans* when it undergoes a white to opaque phenotypic switch. Wor1 contains a double WOPR box, $WOPR_a$

and $WOPR_b$, both of which have been shown to be essential for DNA binding (Figure 5.50b) (Lohse et al., 2010). A later study by Zhang *et al* in 2014, revealed that the two domains interconnect in order to bind DNA, explaining why both were necessary for DNA binding to occur (Zhang et al., 2014).

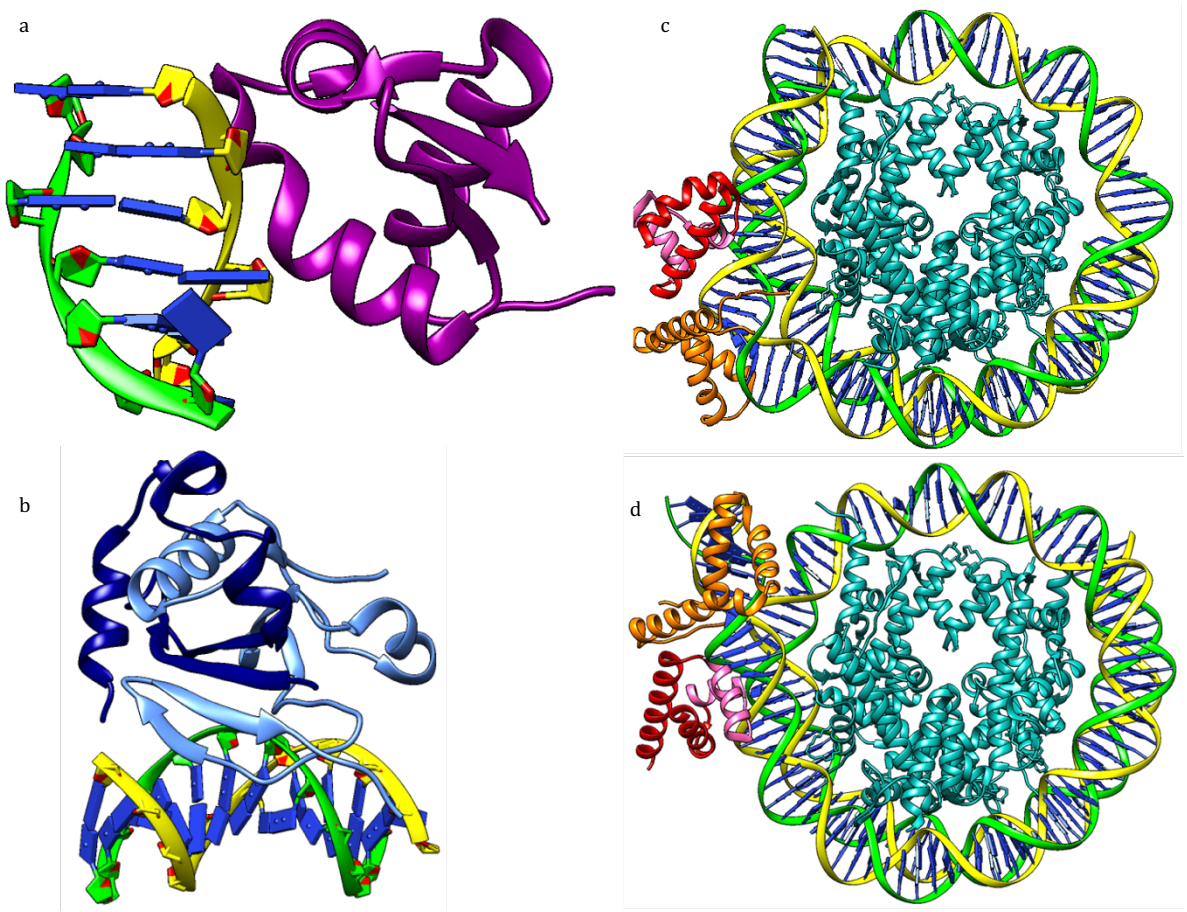


Figure 5.50 Structures of examples of multiple DNA binding domain proteins in *Eukaryotes* (in all figures DNA is shown in green and yellow) **(a)** Structure of a single ZBD domain of DAI (purple) bound to DNA (3EYI (Sung *et al.*, 2008)). **(b)** Structure of Wor1, showing how $WORP_a$ (light blue) and $WORP_b$ (dark blue) interconnect in order to bind DNA (4QTJ (Zhang *et al.*, 2014)). **(c)** Structure of SOX2 (orange) and partial OCT4 (Red with POU_s domain in pink) bound to DNA wrapped around histone complex (teal) (6YOV (Michael *et al.*, 2020)). **(d)** Shows how the SOX2 and OCT4 work together to unwrap DNA from the nucleosome (colourings the same as (c)) (6T90 (Michael et al., 2020)).

Within prokaryotes the best-characterised family of multiple DNA binding domain proteins is the AraC family. AraC from *E. coli* was the first to be characterised, and is involved in the regulation of genes involved in the transportation and regulation of L-arabinose (Lee et al., 1987). Since AraC contains and utilises two distinct HTH motifs, it led to the coining of the term 'bipartite' as a description for these

types of proteins (McIver and Myles, 2002). Each monomer of AraC contains two HTH motifs at the C-terminal end, and a N-terminal arabinose binding domain that allows for dimerisation (Figure 5.51a). Upon dimerisation and in the absence of L-arabinose, the two pairs of HTH lay against the dimerisation domain, and are presented on opposite faces. This causes binding to two half sites, $araO_2$ and $araI_1$, which are positioned 210 bp away from one another. This results in looping of the DNA and repression of the L-arabinose utilisation genes as it prevents RNA polymerase from binding. Binding of L-arabinose to AraC causes a conformational change in the dimer, and rebinding of AraC to $araI_1$ and $araI_2$ instead. This ultimately leads to the activation of the L-arabinose operon (Lee et al., 1987; Rodgers and Schleif, 2009). Also within the AraC family are MarA and Rob, both of which are involved in the regulation of multiple antibiotic resistance genes. Both proteins have two HTH DNA binding domains, however, where MarA is capable of bipartite binding of two adjacent major DNA grooves (Figure 5.51c) (Rhee et al., 1998), Rob is not similar, with the second HTH recognition helix binding to the phosphate backbone (Figure 5.51b) (Kwon et al., 2000). This is thought to be due to a β-loop in Rob that prevents the DNA from bending to the same extent as is necessary in MarA to engage both HTH domain, and would also be required for both domains in Rob to bind to the DNA major grooves (Kwon et al., 2000; Rhee et al., 1998). One additional double DNA binding domain protein is Mga (Figure 5.51d), found in GAS, which is involved in the regulation of multiple virulence factors (McIver and Myles, 2002). GAS can occupy a range of different environments, and therefore a mechanism for tuneable virulence gene control is necessary. Mga was predicted to have 4 possible HTH motifs, but via poly-alanine mutation of the recognition helices, it was shown that just the HTH_3 and HTH_4 motifs were necessary for DNA binding. HTH_4 was essential for DNA binding, whereas removal of HTH_3 had a smaller effect. This finding, paired with the complete conservation of HTH_4 across GAS homologues, has led to the theory that HTH_4 is the major domain responsible for DNA binding. It was hypothesized by McIver and Myles that a double DNA binding domain is present to increase specificity to the broad range of promoters that Mga binds to control virulence, as they have been shown to have a low sequence similarity (McIver and Myles, 2002). Although none of these uses of

double DNA binding domains are completely applicable for the use of these TFs by *Bdellovibrio*, there are parts that are reminiscent of *Bdellovibrio's* requirements; namely the need of global-acting transcription factors and the constant cycle of environmental and morphological change.

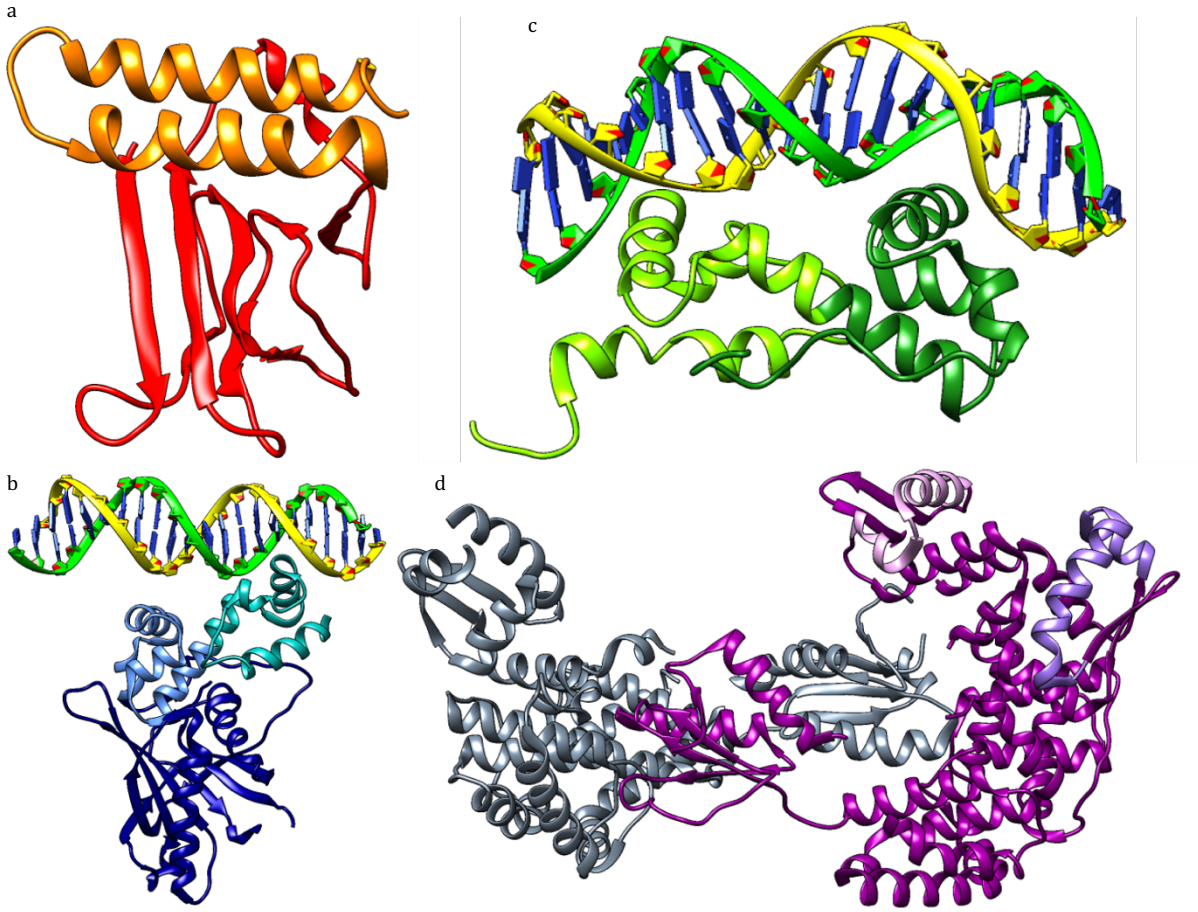


Figure 5.51 Structures of examples of multiple DNA binding domain protein in *Prokaryotes* (in all figures DNA is shown in green and yellow) **(a)** Partial structure of AraC (Red) showing only one of the HTH motifs (orange) (2ARA (Soisson, 1997)). **(b)** Structure of Rob (blue) bound to DNA, showing both HTH motifs (teal and light blue) (1D5Y (Kwon *et al.*, 2000)). **(c)** Structure of MarA with both HTH motifs (light and dark green) bound to DNA (1XS9 (Dangi *et al.*, 2004)). **(d)** Structure of Mga dimer (one monomer dark purple, one grey). The HTH motifs involved in DNA binding, HTH3 and HTH4, are coloured pink and light purple respectively (3SQN (Osipiuk *et al.*, 2011)).

5.3.2 DUF4423 – A Novel DNA Binding Domain

The partial crystal structure of Bd2625 confirmed that the DUF4423 fold does contain a HTH motif, and therefore is likely to partially consist of a WHTH-like motif. Although the polyA mutation EMSAs

did not confirm the presence of the predicted N-terminal HTH domain, or show which of the two DNA binding domains is responsible for DNA binding, the removal of the DUF4423 region completely from Bd1885NT did have a detrimental effect on DNA binding. This assay showed that the N-terminal domain is capable of binding DNA, and is thus likely to consist of the HTH domain that was predicted, and suggests that both the HTH and the DUF4423 domains are required for full DNA binding function.

The crystal structure of Bd2625 proposed a tetrameric DNA binding core, where the recognition helices on the same face are positioned diagonally to one another. This suggests that tetramerisation is essential for DNA binding, as the two monomers predicted to interact with a single DNA duplex are only brought into the correct orientation upon tetramerisation and do not make any direct contacts with one another. Although oligomerisation is often necessary for gene regulation, it is generally dimers that interact with a single DNA site, whereas higher oligomerisation is needed for more complex DNA regulation, e.g. the AraC family (Lee et al., 1987; Rodgers and Schleif, 2009), or histone binding (Michael et al., 2020). Although unusual, the observed tetramerisation is not likely to be an artefact of crystallography. This is supported by Bd2625 FL which was shown to be monomeric in solution by SEC (Figure 5.10) and did not bind DNA, whereas Bd3720 and Bd3720 mutants all formed tetramers and bound DNA. If tetramerisation is key to DNA binding, it may suggest why Bd2625 was the only TF to not show any DNA binding when assayed (Bd2625 tetramerization being reversible/in equilibrium or dependent on buffer conditions).

The partial Bd2625 structure was analysed using DALI (Holm, 2020) and Topsearch (Wiederstein et al., 2014) servers. DALI ranks matches based on fold similarity, via the calculation of a z-score which is based on RMSD and corrected against query length (Holm, 2020). Since such a 'Strong match' limit is dependent upon query length, and is generally defined as either a sequence ID above 20% or a z-score above (n/10)-4, where n is the length of the query (Holm et al., 2008). In the case of Bd2625, this would mean a z-score ≥1.8, however the minimum z-score for a fold to be considered similar is

normally set to >2 (Holm et al., 2008), the top 5 hits for DALI were all above this minimum limit of 2. Three of the hits, PF0864, ADAR and DN052_10610, all contained a canonical WHTH motif, whilst ORF112 recognises DNA via a ZBD domain like DAI (Sung et al., 2008), but unlike DAI works as a dimer rather than having two consecutive DNA binding domains (Table 5.20). Only 2 of the hits had a sequence ID above 20%, PF0864 (Yang et al., 2006) and DN052_10610 (Prabaharan et al., 2019). Both of these proteins possess a WHTH domain, however when superimposed with the tetramer (using the recognition helix pairs from each structure), a good alignment could only be achieved between one monomer of each structure (Figure 5.52). Within the aligned monomers, both structures aligned well with amino acids W103–K149 of Bd2625, further confirming the HTH motif within DUF4423.

Protein Name	DNA Binding Motif	Chain ID	PDB	Z-Score	RMSD	Length of Alignment	Sequence ID %
PF0864	WHTH	B	2IA0	8.1	1.7	50	20
ADAR	WHTH	C	3F21	7.7	1.3	50	14
MMAR_2991	-	A	5ZUL	7.7	1.3	50	16
DN052_10610	WHTH	A	6J05	7.6	2.3	54	26
ORF112	ZBD	C	4HOB	7.5	2.9	56	18

Table 5.20 Top 5 DALI server unique hits for Bd2625

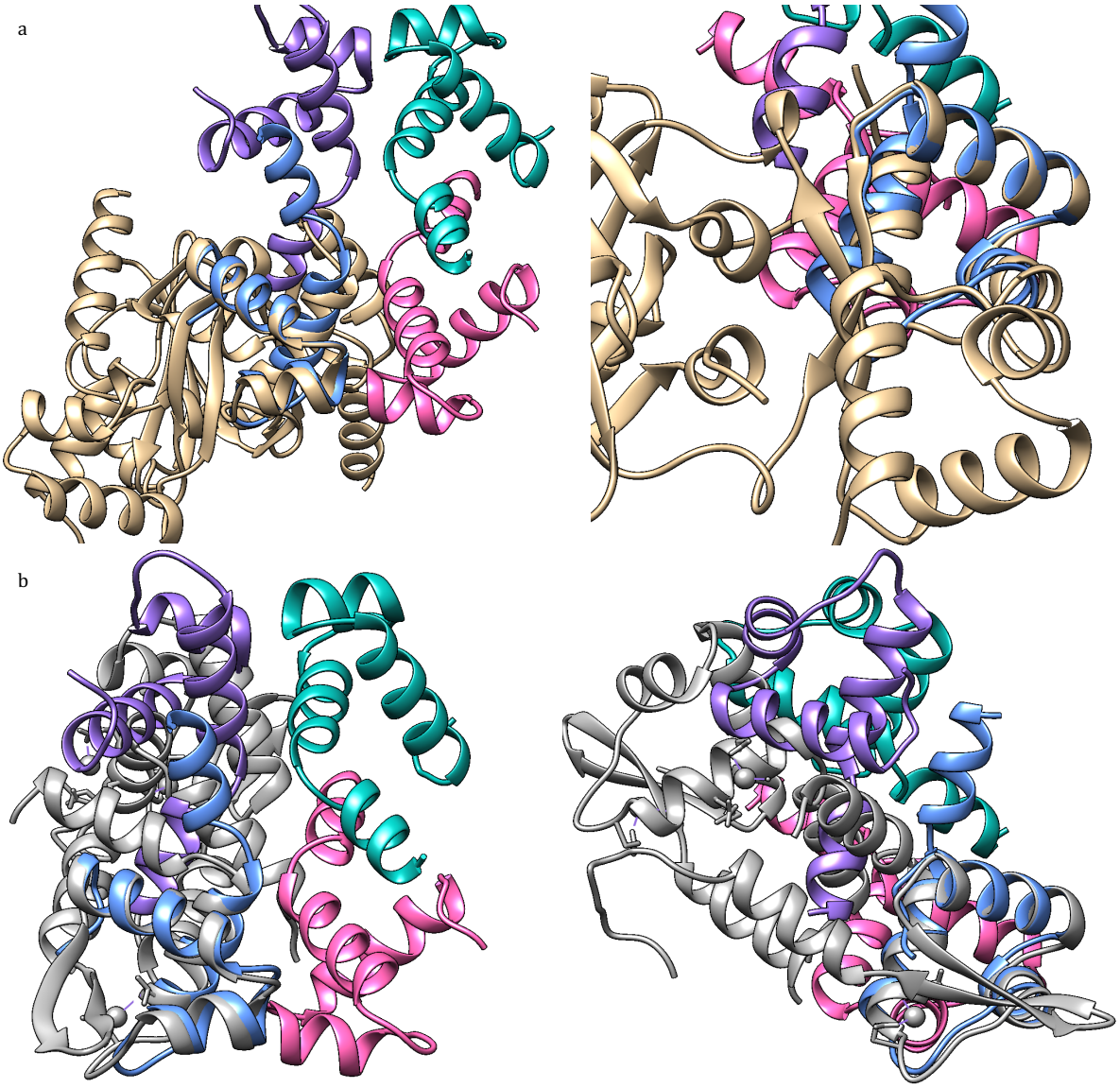


Figure 5.52 (a) PF0864 (2IA0, tan (Yang et al., 2006)) superimposed with Bd2625 based upon the recognition helices from each dimer (Bd2625 – blue and teal), E36–E48 and L132–D144 respectively. Only one HTH motif in each monomer aligns. **(b)** DN052_10610 (6J05, grey (Prabaharan et al., 2019)) superimposed with Bd2625 based upon the recognition helices from each dimer (Bd2625 – blue and teal), H50–H62 and L132–D144 respectively. Only one HTH motif

Due to the poor alignment across the whole tetramer, a Topsearch server (Wiederstein et al., 2014) search was used to find oligomeric matches. Unlike DALI, Topsearch does not compare fold similarity alone, but instead how similar folds are conserved across multimeric interfaces. Topsearch ranks hits based on a S-score, which is calculated by scoring similarity between structures using Gaussian functions (Sippl and Wiederstein, 2012). Therefore, like DALI, the similarity score is length dependent,

but unlike DALI, Topsearch uses the total length of the multimer, which in the case of Bd2625 results in a maximum S-score of 232. The S-score is reported corrected by either the query (S_q), Bd2625, or the target (S_t) multimer length. Of the top 5 hits, only IscR contains a recognisable DNA binding domain, which is again a WHTH domain (Table 5.21). As discussed above, superimposition using the two recognition helices in each DNA binding pair, L132–D144 for Bd2625 and S38–N52 for IscR, led to good alignment across the two recognition helices of Bd2625 (Figure 5.24). One helix anchors the alignment, with the other shifted by 2.90 Å at L132 to 3.95 Å at D144. The monomer with the better fit to the recognition helix of IscR aligned well across the whole HTH motif, W103–K149, with the respective IscR WHTH. However, the poor overall fit from all the proteins identified across the two searches shows that the orientation of the two recognition helices, diagonally opposite each other and having no contacts between the two monomers that bind a single DNA site, is novel and currently restricted to our DUF4423 structure.

Protein Name	DNA Binding Motif	Chain ID	PDB	S	S_t	RMS	Sequence ID %
frdB	-	1	2BS3	99.8	4.3	4.0	9.0
gag-pol	-	2	5KZB	99.4	23.9	3.9	9.2
IscR	WHTH	1	4HF1	96.4	37.4	3.3	13.6
RHO	-	2	2ZIY	94.6	12.8	3.5	6.7
pcnB	-	1	3AQM	88.4	11.4	3.4	10.9

Table 5.21 Top 5 Topsearch server (Wiederstein et al., 2014) unique hits for Bd2625

From the composite model with IscR (Figure 5.24) we can make predictions about the specificity of binding between the DUF4423 DNA binding motif and DNA, if this confirmation holds true for the full length structure in the presence of DNA. The predicted hydrogen bond pattern of Bd2625 to DNA, from the IscR model, shows only one hydrogen bond between the recognition helix, N137, and the nucleotide bases. This may suggest that DUF4423 binds DNA in a largely base neutral manner, instead favouring shape readout as a method of DNA recognition (Lin and Guo, 2019), followed by hydrogen bonding to subsequently stabilise the interaction with DNA, similar to WOPR (Zhang et al., 2014). Further comparison of the composite model to existing literature supports the lack of specificity

shown in the EMSAs. Recognition of DNA via a helix, instead of a strand or coil, is generally associated with a reduced sequence specificity (Lin and Guo, 2019). Both recognition sites within these TFs are predicted to interact with DNA via helices, with the predicted recognition helix in the DUF4423 domain being supported by the composite model. A bias of hydrogen bonding to one DNA strand over the other is also associated with specificity to a range of DNA motifs (Lin and Guo, 2019). This again supports the results from EMSAs conducted with Bd3720, where binding to a broad range of DNA fragments was observed, and suggests that these TFs are not as specific as first hypothesized.

5.3.3 Updated Working Hypothesis

5.3.3.1 From self-regulating AP/GP Switch Transcription Factors

The initial working hypothesis of this project postulated that this family of TFs were involved in the attack phase to growth phase switch (AP/GP switch). This hypothesis was guided by upregulation of 6 of the TF genes in a study by Lambert *et al* (Lambert et al., 2010). This study compared gene expression 30 minutes post-invasion of a host cell to HI *Bdellovibrio*. On account of the strong upregulation, *bd1219*, *bd1885*, *bd1897*, *bd1905*, *bd2775*, and *bd3758*, were all deemed a part of the "predatosome" (Lambert et al., 2010) and annotated as TFs involved in the AP/GP switch. A later study, by Karunker *et al*, showed that a further 5 of these TFs, *bd0089*, *bd1885*, *bd1901*, *bd2625*, *bd3700*, and *bd3720*, were upregulated 3 hrs into growth (Karunker et al., 2013). Since all TF were thought to be expressed during similar times of the lifecycle this led to the assumption that the diversification into a large family meant that each protein was highly specific, acting at a limited number of binding sites, most likely one locus. The HTH domain sequence is more conserved across all members of this family than the DUF4423 region. This conservation pattern may, similar to other studies (McIver and Myles, 2002), suggest that the HTH domain is more important for DNA binding than the WHTH-like motif. However, because of the multiple copies of this family within *Bdellovibrio*, the working assumption at the start of this project was that the sequence variation within the WHTH-like motif would allow for specificity

to a (variable) particular DNA sequence. The HTH domain would then be postulated to provide secondary contacts to increase the tightness of binding between these TFs and DNA. Considering most TFs are self-regulating and the RNA-seq data from Karunker *et al* revealed tight regulation of the predicted operons that these TFs were a part of, this led to our use of their upstream regions in the testing of their DNA binding capabilities. Although the TFs did bind these regions, we are now aware that these TFs show a far greater binding diversity than originally thought, and that our initial assumption of them being highly specific growth phase regulators is likely not true.

5.3.3.2 To Hypothesised Environmentally Controlled Nucleoid-Associated Proteins

In 2020, a further paper on the differential gene expression of *Bdellovibrio* was published by Jurkevitch *et al*, which compared the regulation of genes during starvation with that of newly released progeny. This study identified a further three TF genes, *bd0724*, *bd1172*, and *bd2560* that were upregulated during starvation arrest (Sathyamoorthy et al., 2020), thus identifying a role for this subgrouping of TFs in the regulation for genes involved in starvation management. This new role, allied with the findings that Bd3720 was not as site-specific as first expected, led to our reinterpretation of the differential expression of the TFs genes reported in previous studies (Table 5.22, Figure 5.53). Instead of hypothesising that the differential expression of these genes in HI *Bdellovibrio* is due solely to their regulation of specific growth genes, we now propose they are broad-acting transcription regulating proteins, likely controlled by environmental factors and acting via a low-specificity nucleoid-associated protein (NAP) like mechanism.

Protein	(Lambert et al., 2010)	(Karunker et al., 2013)	(Sathyamoorthy et al., 2020)
Bd0089	Downregulated in HI	Upregulated 3 hrs after invasion	
Bd0681	Downregulated in HI	Silent	
Bd0724	Upregulated in HI	Silent	Upregulated during predator starvation
Bd1172	No difference	Silent	Upregulate during predator starvation
Bd1219	Upregulated 30 mins into invasion	Silent	
Bd1537	No difference	Silent	
Bd1885	Upregulated 30 mins into invasion	Upregulated 3 hrs after invasion	
Bd1897	Upregulated 30 mins into invasion	Silent	
Bd1901	Downregulated in HI	Upregulated 3 hrs after invasion	
Bd1905	Upregulated 30 mins into invasion	Silent	
Bd2551	Upregulated in HI	Silent	
Bd2560	Downregulated in HI	Silent	Upregulated during predator starvation
Bd2625	No difference	Upregulated 3 hrs after invasion	
Bd2774	Downregulated in HI	Silent	
Bd2775	Upregulated 30 mins into invasion	Silent	
Bd3700	Upregulated in HI	Upregulated 3 hrs after invasion	
Bd3720	Upregulated in HI	Upregulated during attack phase and 3 hrs after invasion	
Bd3758	Upregulated 30 mins into invasion	Silent	

Table 5.22 The differential genes expression of the transcription factors as reported in Lambert *et al.* 2010, HI expression compared to 30 minutes into invasion, Karunker *et al.* 2013, attack phase expression compared to growth phase expression, and Sathyamoorthy *et al.* 2020, the gene expression of freshly release progeny compared to 10 hours post-release starvation *Bdellovibrio*.

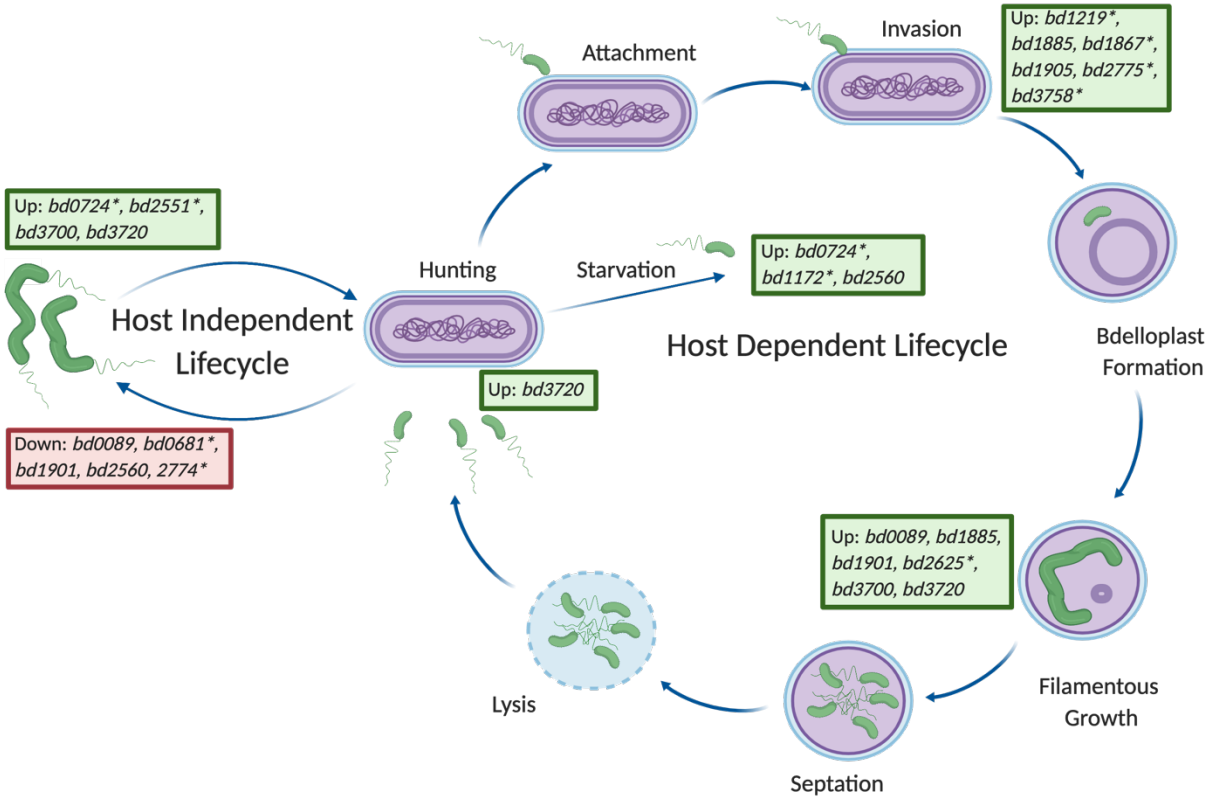


Figure 5.53 Lifecycle of *Bdellovibrio* showing the differential expression of the HTH-DUF4423 family. Green boxes are genes shown to be upregulated at that particular point in the lifecycle, whereas red boxes are genes that are downregulated (Data obtained from Lambert et al., 2010; Karunker et al., 2013; Sathyamoorthy et al., 2020). Created with BioRender.com

NAPs act in a histone-like manner, and are one group of proteins which organise and compact chromatin in prokaryotes and some archaea (Luijsterburg et al., 2006). There are five levels of chromatin organisation that have been described in the model organism *E. coli*. Firstly, large scale organisation of chromatin results in four distinct structured macrodomains (MD); Ter, Ori, Left and Right, along with 2 less-structured domains which flank the Ori MD. MDs are maintained in a defined cellular position throughout the cell cycle (Cournac and Plumbridge, 2013). The organisation of the Ter MD in *E. coli* is due the highly specific binding of MatP to *matS*. Interaction with a single *matS* site is mediated by a dimer of MatP. Originally it was thought tetramerisation of MatP using the C-terminal coiled coil domain caused the formation of bridges across distant *matS* sites, condensing the chromosome to form the Ter MD (Dupaigne et al., 2012). However, it now appears that condensation occurs because the binding of MatP to *matS* causes the exclusion of condensin MukBEF from the *ter*

region which causes the formation of the Ter MD (Lioy et al., 2018). Like the DUF4423 TFs, MatP interacts with DNA via 2 N-terminal DNA binding motifs, a four helical bundle (4HB) motif and a Ribbon-Helix-Helix (RHH) motif, although only the 4HB interacts specifically with the *matS* sequence (Dupaigne et al., 2012). A similar protein, MaoP, interacting with *maoS* sites in the Ori region is responsible for organising this into a MD (Duigou and Boccard, 2017). There is no homology between any *Bdellovibrio* protein to either MatP or MaoP, however, there is agreement between Bd3095 and ParB, an alternative mechanism for organisation used in *Myxococcus xanthus* (Harms et al., 2013). Although *Bdellovibrio* appears to utilise this high level chromatin organisation, the non-specific manner in which the DUF4423 TFs bind DNA make it unlikely for them to function at this high level. This is because the non-specific binding would allow widespread interaction across the entire chromosome causing aberrant condensation.

The next two levels of DNA organisation do not require NAP proteins for stabilisation. The first involves large scale loops that are stabilised by the membrane and/or RNA. Additionally, supercoiling, which functions over ~10 kb distances, splits the chromatin into topological domains. Both intermediate and small loops are stabilised by NAPs and TFs respectively. Transcription factors are mostly responsible for these small loops, and are the protein type most commonly associated with transcription regulation, but there is an increasing awareness of NAPs acting as global transcription regulators (Cournac and Plumbridge, 2013).

NAPs are generally basic with a low molecular weight and are present in high abundance within the cell, and therefore used to be described as histone-like despite the lack of structural similarities (Luijsterburg et al., 2006). NAPs act to compact DNA by; bridging (H-NS, Lrp, SMC, and Fis), bending (HU, IHF, and Fis), and wrapping (HU and Lrp) (Figure 5.54) (Luijsterburg et al., 2006, 2008).

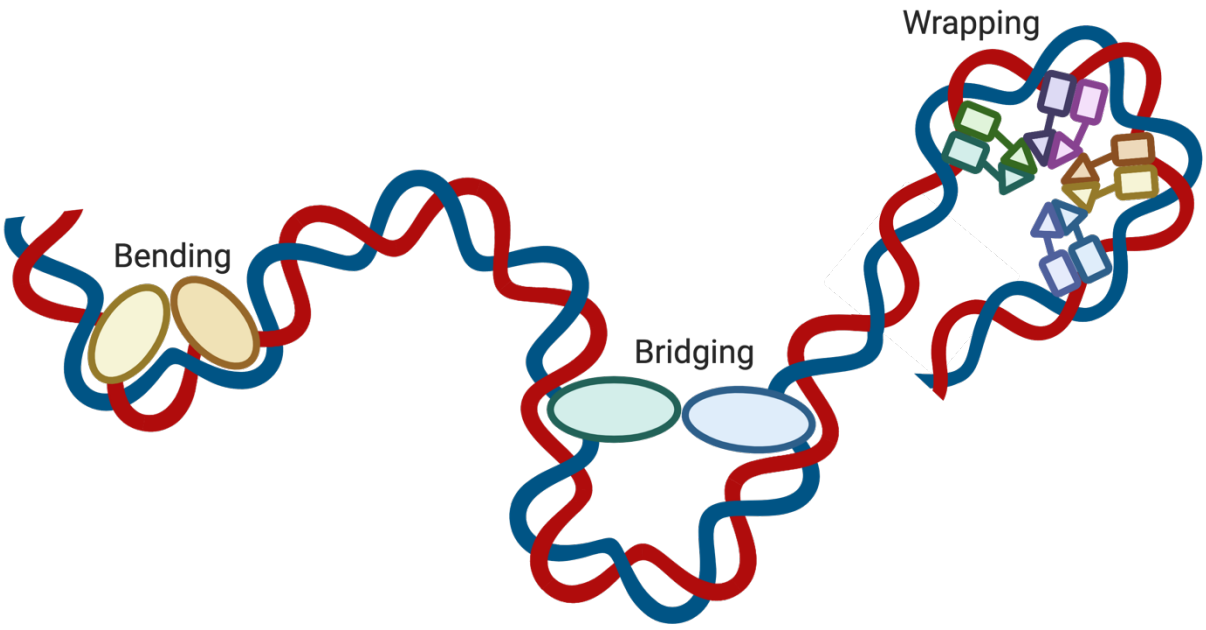


Figure 5.54 Illustration of the mechanisms of DNA bending, bridging and wrapping controlled by NAPs (Figure adapted from Luijsterburg *et al.*, 2006, 2008). Created with BioRender.com

Bending proteins HU and IHF are highly-related proteins that interact with DNA by forming homo- and hetero- dimers respectively (Luijsterburg et al., 2008). Both proteins bind DNA by inserting a β-arm from each monomer into the major groove of DNA. IHF recognises a fairly well conserved AT-rich binding site, 5'-(A/T)ATCAANNNNTT(A/G)-3', where the less conserved 3' end is generally also AT rich. HU binds non-specifically but has a higher affinity for supercoiled or structurally distorted DNA (Luijsterburg et al., 2006), *e.g.* AT rich domains that have an inherent curvature (Carrera and Azorin, 1994). The other commonly conserved DNA bender in Gram-negative bacteria is Fis, which again binds DNA as a dimer via a HTH motif in each monomer. Fis regulates a large range of cellular functions including transcription, recombination, and replication, as well as directly or indirectly affecting the RNA levels of approximately 20% of chromosomal genes (Hancock et al., 2016). Like IHF, Fis is considered to be specific, despite recognising a very degenerate 15 bp sequence with high affinity (Luijsterburg et al., 2006; Skoko et al., 2006). Continued studies into Fis using Chip-seq, have revealed that only a few bases are conserved within the Fis binding site (Stella et al., 2010). This site includes a 5 bp AT rich core at positions -2 – +2, where 0 is the centre of the 15 bp site, with a guanidine at -7 and a cytosine at +7. The guanidine and cytosine have been shown to be directly contacted by the

recognition helix. The only other conserved factor within the binding site is that an adenine at -4 and a thiamine at +4 which prevent the binding of Fis (Hancock et al., 2016; Stella et al., 2010). AT richness is known to cause inherent curvature to DNA which narrows the minor groove (Carrera and Azorin, 1994; Rettig et al., 2013). This has led to a proposed mechanism for Fis binding that involves Fis first scanning the genome for a narrowed minor groove caused by the AT rich core of the binding site. The two recognition helices in the dimer then insert into the adjacent major grooves, and form stabilising H-bonds between the α-helices and the DNA. This causes the DNA to bend further as the recognition helices are positioned 10 Å closer together than the major grooves in straight DNA (Luijsterburg et al., 2006). Crystal structures have shown that the recognition helices of Fis do not change conformation upon DNA binding, and the crystallised DNA fragment is more bent than inherent AT richness would naturally induce, leading to the conclusion that Fis induces additional bending upon binding (Stella et al., 2010). A proposed protein-protein interaction between Fis dimers is thought to cause further DNA condensation by a bridging effect (Figure 5.55) (Luijsterburg et al., 2006, 2008).

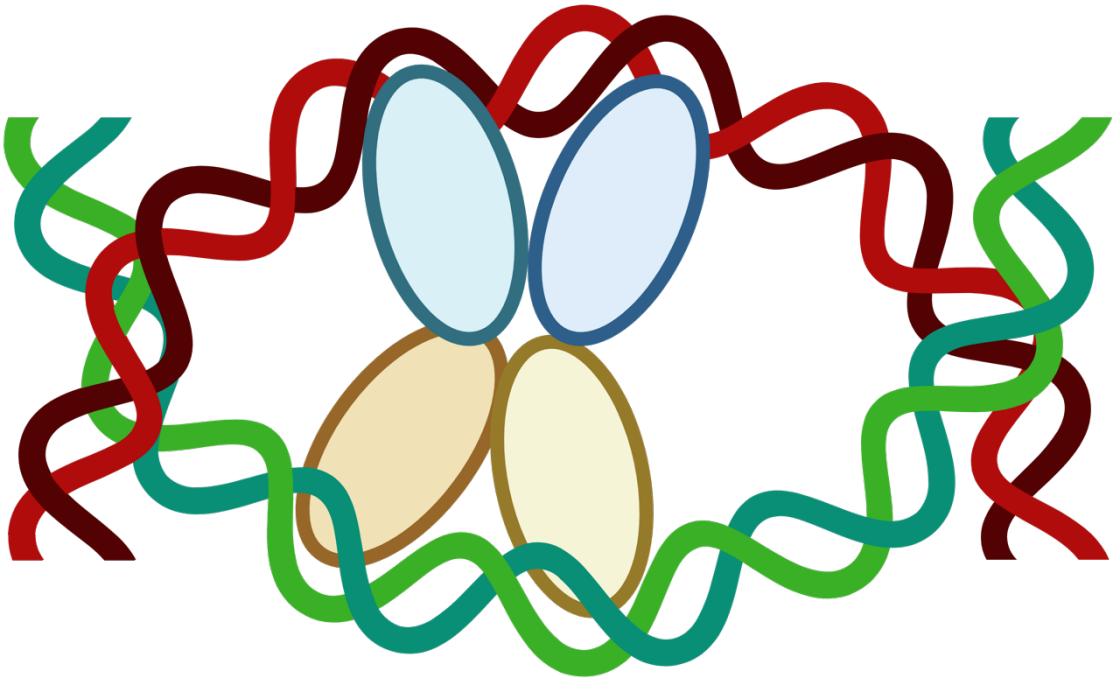


Figure 5.55 Illustration of the mechanism by which the Fis tetramer bridges DNA, each dimer (orange and yellow/teal and blue) binds one strand of DNA, causing the DNA to bend. The DNA is then bridged by tetramerisation of the two dimers via the C-terminal ends. (Adapted from Luijsterburg *et al.*, 2006). Created with BioRender.com

DNA bridging protein SMC is the exception to the general rule of NAPs being of low molecular weight with an average molecular weight of 150-200 kDa. The N-terminal and C-terminal regions of SMC combine to form an ATP cassette, whilst the centre region forms a highly flexible hinge involved in dimerisation. SMC proteins are thought to trap two strands of DNA within the dimer, forming a ring-like structure around the DNA, instead of directly interacting with the DNA bases (Luijsterburg et al., 2006). Like Fis, Lrp has been shown to be a global regulator, controlling the expression of at least 10% of genes in *E. coli,* which are involved in the transportation, degradation and biosynthesis of amino acids, as well as the production of pili and porins, and sugar transport (Thaw et al., 2006). Lrp is comprised of a N-terminal 3 helical HTH motif and C-terminal leucine response domain that is involved in the formation of multiple oligomers: dimers, tetramers, octamers and hexadecamers. The octamers, made of 4 dimers, forms a disc shape that can either bridge DNA by binding 2 dimers on opposite faces to the DNA, or wrap DNA around all 4 dimers (Figure 5.54). Binding of DNA to the hexadecamer causes the DNA to wrap around the protein complexes approximately two times, and is the closest prokaryote DNA organisation method to the histone:DNA complexes found in eukaryotes (Luijsterburg et al., 2006). Like most NAPs, Lrp appears to bind DNA in a non-sequence dependent manner, instead preferentially binding to inherently curved DNA. Despite being named after its leucine binding domain, the majority of gene regulation by Lrp does not require the binding of leucine (Thaw et al., 2006). One gene regulated by Lrp that is dependent on leucine is *fimA*, which is activated upon the binding of Lrp to the *fimA* promoter *fimX*, this is thought to exclude global-repressor H-NS from binding thus allowing expression of *fimA* (McFarland et al., 2008). The final DNA bridging protein normally present in Gram-negative bacteria is H-NS. Unlike other NAPs, H-NS has an atypical DNA binding motif, binding DNA via a turn formed between a strand and helix (Luijsterburg et al., 2006), and yet H-NS still seems to bind DNA via a shape readout mechanism rather than sequence specificity, again preferring AT rich tracts (Luijsterburg et al., 2008). H-NS has also been shown to have a very broad binding range, not only binding double-stranded DNA, but also single-stranded DNA and RNA, albeit with decreasing affinity. Despite this, H-NS is considerably more specific than other NAPs, and

is reported to control approximately 30 genes, all involved in managing environmental stress (Williams and Rimsky, 1997). Although originally thought to only regulate genes via repression, further work has shown that as many genes are positively regulated by H-NS as are repressed (Atlung and Ingmer, 1997).

Although H-NS is reported as being conserved across Gram-negative bacteria (Luijsterburg et al., 2006), analysis of the *Bdellovibrio bacteriovorus HD100* genome using Uniprot (The UniProt Consortium, 2019), Psi-Blast (Altschul, 1997) and BackPhyre (Kelley et al., 2015), failed to identify any homologues to H-NS within *Bdellovibrio*. Additionally, similar searches for IHF and HU proteins in *Bdellovibrio* revealed significantly fewer copies, 4, compared to model organism *E. coli* which possesses 12. Knockout studies of NAPs have shown that the removal of one family has a limited effect on phenotype, and that other NAPs are instead upregulated to maintain gene regulation and chromatin condensation (Luijsterburg et al., 2006). Therefore this natural reduction in *Bdellovibrio* of IHF, HU and H-NS NAPs may imply that *Bdellovibrio* has a novel family of NAPs to compensate. This conclusion of *Bdellovibrio* having a set of potential novel NAPs is partly supported by the identification of CdbA and CdbB in the related deltaproteobacterium *Myxococcus xanthus* (Skotnicka et al., 2020). CdbA/B both possess a RHH DNA binding motif, and are not present in *Bdellovibrio*. CdbA seems to have a more important role in the chromosomal organisation of *M. xanthus* than CdbB, as deletion of *cdbA* has a detrimental effect on chromosomal organisation, resulting in a phenotype in which cells are unable to complete division. *cdbB* deletion does not appear to cause a distinct phenotypic difference from *M. xanthus* WT. Unlike most NAPs, CdbA binds to a consensus sequence that is GC rich. This was partly expected due to *M. xanthus* having a GC rich genome with an overall GC content of 69%. However, CdbA binding to DNA has shown to be regulated by the binding of c-di-GMP via the RHH motif, which has led to speculation that this is a secondary reason as to why CdbA binds to a GC rich consensus as opposed to the more prototypical AT rich of most NAPs. The theory of c-di-GMP outcompeting the binding of DNA to the RHH domain, in order to regulate function, is supported by a

higher affinity between CdbA and c-di-GMP (~83 nM) compared to DNA (0.4–4 μM) (Skotnicka et al., 2020). This, paired with gene expression data that shows both *cdbA* and *cdbB* are constitutively expressed (Muñoz-Dorado et al., 2019; Sharma et al., 2021), has led to the hypothesis that fluctuations in c-di-GMP concentrations throughout the *M. xanthus* lifecycle allows for an analogue regulation of genes controlled by CdbA, as an on/off transcription pattern has been shown to be detrimental to *M.* xanthus (Skotnicka et al., 2020).

Despite possessing a prototypical fold, and being approximately twice the molecular weight for most common NAPs, we hypothesize that the DUF4423 family act as global gene regulators by chromosomal organisation. There is a wealth of evidence that the global regulation of gene transcription caused by NAPs remodelling DNA is in response to, and controlled by, environmental signals (Hołówka and Zakrzewska-Czerwińska, 2020; Luijsterburg et al., 2006, 2008). This is best understood and described by the mechanisms of regulation utilised by CdbA and Lrp (Skotnicka et al., 2020; Thaw et al., 2006). Due to lifecycle-dependent differential expression of the DUF4423 family, it is highly likely that an environmental signal governs transcription of the family, rather than post-translationally/allosterically controlling binding to DNA as seen with CdbA. Although there is no structural similarity between CdbA and the DUF4423 family, they are still roughly comparable as both are found in predators, use prototypical DNA binding domains and have similar DNA binding affinities. The similar DNA affinity of both proteins may indicate that the binding of DNA to the DUF4423 family is actually out-competed by a small molecule binding to one or both DNA binding domains, similar to c-di-GMP and CdbA, as they do not have a recognisable ligand binding domain, as discernible in Lrp and SMC. However, unlike CdbA, we hypothesize that the DUF4423 family binds the AT rich regions that are stereotypically bound by NAPs in a non-sequence dependent manner (Hołówka and Zakrzewska-Czerwińska, 2020). Although the fragment assay did seem to imply some level of sequence specificity as DNA binding was eventually abolished, Bd3720 was able to bind to all ‘full-length’ DNA fragments it was assayed with. AT enrichment in the DNA fragments, as determined by

the difference to overall GC-content of *Bdellovibrio* which is 50.6% (Pasternak et al., 2014), was compared to the approximate ratio of TF to DNA needed for full binding to occur. This showed a general trend where the higher the GC content within the DNA fragment, the more protein was required for binding saturation (Table 5.23). This would explain why Bd3720 was able to bind all DNA fragments assayed, but why some bound with an apparent lower affinity than others.

Protein	DNA	GC %	AT %	ΔGC %	Protein (μM)	DNA (μM)	Ratio
Bd3720	*bd2626*^{U}	49.1	50.9	1.5	16.9	0.0169	1000
Bd3720	*gltF*	48.9	51.1	1.7	5	0.0143	350
Bd3720	*bd1172*^{U1}	47.1	52.9	3.5	11.1	0.0111	1000
Bd3720	*bd1885NT*	46	54	4.6	5	0.0160	313
Bd3720	*bd1885*^{U1}	45	55	5.6	11.1	0.0111	1000
Bd1172	*bd1172*^{U2}	43.9	56.1	6.7	5	0.0231	216
Bd1885	*bd1885*^{U2}	43.5	56.5	7.1	5	0.0204	245
Bd3720	*bd3721*^{U}	41.8	58.2	8.8	1.6	0.0105	152
Bd3720	*bd0089*^{U}	41.7	58.3	8.9	1.49	0.0149	100
Bd3720	*bd2625*^{U}	39.6	60.4	11	4.16	0.0416	100

Table 5.23 Table comparing the ratio of protein to DNA to obtain full shift in EMSA, compared to the AT richness of the fragment.

The presentation of the recognition helices in pairs on either face of Bd2625 tetramer crystal structure is reminiscent of the proposed Fis tetramer bridging mechanism (Figure 5.55). This, coupled with the apparent preference to AT rich DNA fragments has led to the conclusion that the DUF4423 family may locate DNA binding sites, bind to them and bridge DNA in a manner similar to Fis. Fis is one of the few ubiquitous NAPs *Bdellovibrio* is predicted to possess by Uniprot (The UniProt Consortium, 2019), Psi-Blast (Altschul, 1997) and BackPhyre (Kelley et al., 2015). Like Fis, the DUF4423 family would locate target DNA binding sites via an intrinsically narrowed minor groove, due to AT richness that causes natural bending to occur within the DNA. Binding would then be stabilised by the few hydrogen bonds predicted to form between the DNA and the DUF4423 DNA binding domain. Unlike Fis, DUF4423 is unlikely to bend the DNA further, because the two recognition helices are approximately 34 Å apart from one another, unlike Fis where they are about 25 Å apart (Luijsterburg et al., 2006). The major

grooves in DNA are 34 Å apart (Luijsterburg et al., 2008), implying that both recognition helices could bind two adjacent DNA major grooves without needing the DNA to bend significantly. Recently Yang *et al.* released a predicted structure of the DUF4423 domain (calculated by co-evolution analysis) (Figure 5.56), which predicts a WHTH motif (Yang et al., 2020) and supports our structural and biochemical conclusions that the domain is involved in DNA recognition. However, upon alignment of the HTH motif in both structures the body of the DUF4423 fold (predicted form) is shown to clash with the neighbouring copies in the tetramer (experimental form). If the tetrameric conformation we observe is conserved in the full-length protein, we predict that the body of the DUF4423 fold modelled by Yang *et al.* twists in the opposite direction post wing in order to maintain this oligomerisation (Figure 5.57). A small section, L92–S102, of the linker between the two DNA binding domains, H48–S102, is present in the partial structure of Bd2625, the angle of which suggests that the HTH domains from each monomer would not come into contact with another monomer. Thus, the HTH motif is more likely to make contact with the DNA than form protein-protein interactions (Figure 5.57). If the HTH motif did bind DNA, which the Bd1885N-*bd1885*^{U1} EMSA is supportive of, it would also likely be non-specific in nature, as suggested by the relatively poor affinity determined, however at present this is unconfirmed, but future ChIPseq experiments could be used to address the specificity of both the HTH and DUF4423 domains and may identify a sequence-specific recognition site.

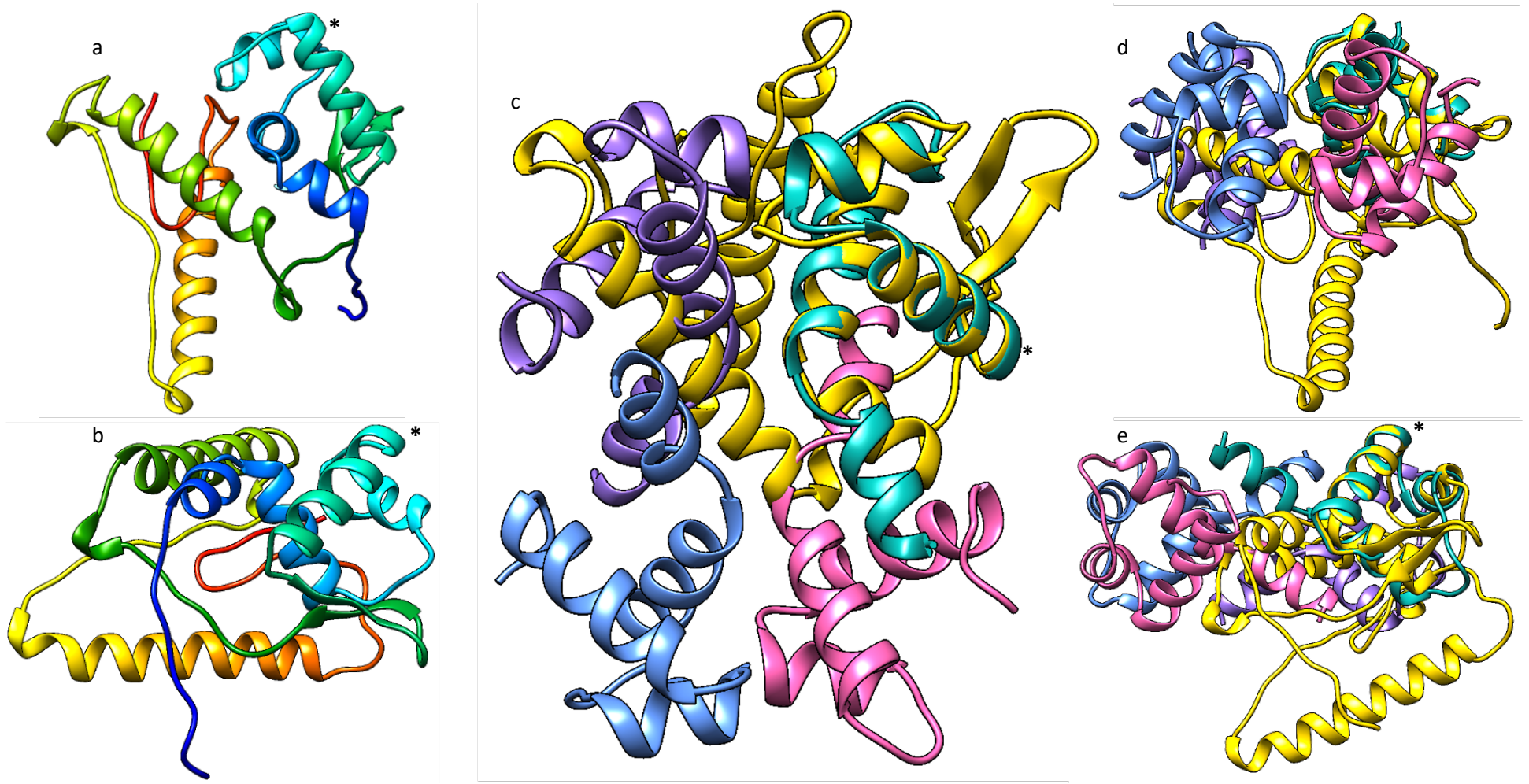


Figure 5.56 (a) DUF4423 predicted structure (Coloured Jones Rainbow N-terminal to C-terminal) (Yang et al., 2020) **(b)** DUF4423 predicted structure at 90° rotation from (a) **(c-e)** Bd2625 tetramer (blue, purple, teal and pink) superimposed with the DUF4423 predicted structure (yellow) (Yang et al., 2020). Alignment of the predicted DUF4423 structure with the partial Bd2625 structure shows that the body of the predicted fold after the HTH would clash with the neighbouring copies of in the tetramer, indicating the domain is likely to be twisted in the opposite direction post wing. Recognition helices denoted by a * in (a-c) and (e).

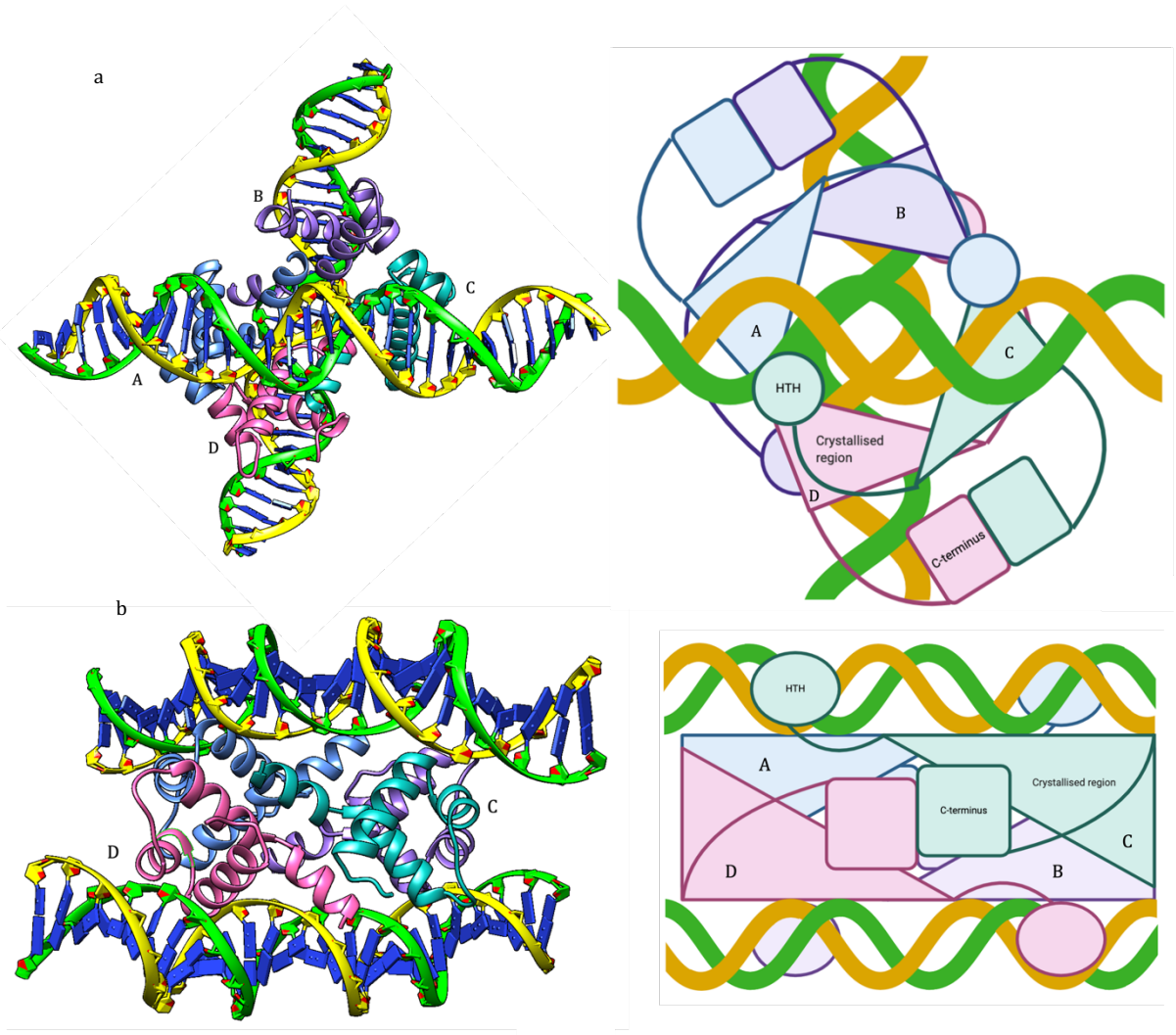


Figure 5.57 Predicted full length structure of the DUF4423 family, based on the partial structure of Bd2625. The predicted interaction of the HTH domain with the same DNA strand contacted by the DUF4423. Most NAPs and TFs have an oligomerisation region, and the C-terminus of the DUF4423 region is shown as fulfilling this role as there are only a few hydrogen bonds stabilising tetramerisation in the crystallised region. **(a)** Top view of full length prediction next to partial Bd2625 structure and modelled DNA (4HF1 (Rajagopalan et al., 2013)), where Chain A (Blue) and Chain C (Green) bind to the top strand of DNA. **(b)** Side view of full length prediction next to partial Bd2625 structure and modelled DNA (4HF1 (Rajagopalan et al., 2013)), where Chain D (Pink) and Chain C (Green) are front most in the figure but interacting with the bottom and top DNA strands respectively. Created with BioRender.com

This mechanism of binding DNA via a shape-readout mechanism (Zhang et al., 2014) would explain why the poly-ala mutations of the recognition helices in Bd3720 made little difference to the binding affinity. Since shape-readout does not rely on a specific DNA sequence, it stands to reason the amino acid sequence within the recognition helix is less important to the binding of the DNA, as long as the

recognition domain is maintained so initial interaction can occur and the subsequent stabilising contacts can be formed. This also supports the conclusion that if the HTH motif binds DNA, it would be in a non-specific manner as neither the polyA HTH or Double mutant drastically changed the binding affinity compared to either WT or polyA WHTH.

Bridging of DNA by NAPs is generally considered to repress the expression of genes, by preventing the access of RNA polymerase to DNA (Hołówka and Zakrzewska-Czerwińska, 2020). In some cases, as seen with the xenogeneic silencing performed by H-NS, a counter-silencing mechanism is undertaken, where removal of a NAP (or other DNA-silencing complex) from DNA leads to transcription without RNA polymerase directly interacting with the NAP:DNA complex (Wiechert et al., 2019). A number of factors have been observed to interfere with NAP:DNA complexes including temperature, osmolarity and TFs (Flores-Ríos et al., 2019). However this mechanism of gene regulation by NAPs has currently only been observed in the regulation of genes obtained by horizontal gene transfer (Will et al., 2015). Therefore, the proposed mechanism of the DUF4423 family forming cross-bridges in DNA implies that the proteins repress the transcription of genes involved in the attack phase, as the majority of the family are upregulated during invasion and growth. Those members upregulated during starvation most likely have a role in swimming arrest. As stated above, Gram-negative conserved DNA bridging protein H-NS is not present in *Bdellovibrio*. Due to the broad activity range of H-NS, and its repressor capabilities, expansion of this repression mechanism into a large family would provide *Bdellovibrio* greater control of gene regulation, allowing it to more readily adapt to the highly variable environments it may encounter.

5.3.4 Effectors

One of the proteins identified by HHpred (Zimmermann et al., 2018) to have homology to several of the effectors is HtpX from *Vibrio parahaemolyticus* (3CQB), a M48 heat shock metallopeptidase (Wu et al., 2008). HtpX in *E. coli* is thought to work in conjunction with FtsH, an ATP-dependent

metallopeptidase, in the removal of aberrant membrane proteins. Removal of either HtpX or FtsH has been shown to trigger the two component Cpx stress response system. HtpX undergoes self-degradation, and the removal of the bound zinc is enough to eliminate this activity (Sakoh et al., 2005). If the effectors are metallopeptidases they may also be capable of self-degradation. This may explain why EDTA could stabilise Bd1896, as the EDTA would remove the nickel ion bound to the active site, abolishing its catalytic ability, and preventing self-degradation. Although metallopeptidases commonly use zinc, nickel is a known substitute (Cerdà-Costa and Gomis-Rüth, 2014), further supporting this as an explanation that 'natively' prepared Bd1896 was unstable.

A protein related to HtpX, PG2197, in *Porphyromonas gingivalis* has been shown to be upregulated in disease states (Walters et al., 2009). PG2197 is also a M48 metalloprotease and has been linked to virulence as one of 18 genes shown to be upregulated 5 and 30 mins into adherence to human coronary artery endothelial cells. However, PG2197 is not necessary for invasion of *P. gingivalis* into endothelial cells. The link between HtpX and Cpx in *E. coli* and the homology between HtpX and PG2197, has led to it being theorised that the loss of PG2197 causes the activation of the Cpx pathway, and the upregulation of virulence factors (Walters et al., 2009). Similar to this, the most common expression timepoint of these effectors in *Bdellovibrio* is 30 mins into invasion, observed for ~1/3 of family members (Table 5.24). The predicted structural similarity between these effectors and metallopeptidases involved in the stress response may suggest a similar link in *Bdellovibrio.* This hypothesis is supported by the identification of these effectors in other *Bdellovibrio* strains (Table 5.25). Most strains of *Bdellovibrio* possess about 2/3 as many effectors as NAPs, with *Bdellovibrio exovorus* having slightly fewer. However, no effectors could be identified in *Bdellovibrio str. W* via BlastP (Altschul et al., 1990) or PSI-Blast (Altschul, 1997). *Bdellovibrio str. W*, has been shown to be the only related strain to undergo encapsulation as a stress response to starvation. Encapsulation involves the entry of a host prey cell by *Bdellovibrio str. W*, but without the predator initiating lysis at 3-4 hours (Tudor and Conti, 1977a). Instead, *Bdellovibrio str. W* becomes dormant and remodels the

host into a protective layer termed a 'bdellocyst' (Tudor and Conti, 1977b). The bdellocyst provides increased resistance to temperature fluctuations and cellular disruption caused by dehydration. Encapsulation allows *Bdellovibrio str. W* to remain viable for longer under starvation stress (Tudor and Conti, 1977a). Upon the increase of nutrients in the environment, the bdellocyst germinates and a single viable *Bdellovibrio str. W* is released (Tudor and Conti, 1977b). It could therefore be theorised that if these effectors are involved in a stress response mechanism as is shown in other metallopeptidases, the reason they are not found in *Bdellovibrio str. W* is because this strain utilises different stress-related mechanisms.

All of the effectors, apart from Bd1536, are predicted to encode a signal peptide. Signal peptides are associated with either the translocation of proteins to the periplasm or secretion from the cell. Therefore, although there is a possible role for these effectors to behave like similar membrane-associated protein HtxP, there is a possible role for these proteins to be exported into the prey and degrade prey protein during predation. However, the differential expression of some of these effectors in the HI lifecycle (Bd1170 and Bd1171 are upregulated, Bd1896 and Bd2561 downregulated) (Table 5.24), may further suggest a role in a stress response-like pathway.

Protein	(Lambert et al., 2010)	(Karunker et al., 2013)	(Sathyamoorthy et al., 2020)
Bd0682	Up regulated 30 mins into invasion	Silent	-
Bd0723*	-	Up regulated during attack phase and 3 hrs after invasion	-
Bd1170	Up regulated in HI	Up regulated in attack phase	-
Bd1171*	Up regulated in HI	Silent	-
Bd1218*	Up regulated 30 mins into invasion	Silent	-
Bd1896	Down regulated in HI	Silent	-
Bd2552*	-	Silent	-
Bd2561	Down regulated in HI	Silent	-
Bd2624*	-	Up regulated 3 hrs after invasion	-
Bd2776*	Up regulated 30 mins into invasion	Silent	-
Bd3759*	Up regulated 30 mins into invasion and in HI	Silent	-
Bd1536^	Up regulated 30 mins into invasion	Silent	-

Table 5.24 The differential gene expression of the putative effectors as reported in Lambert et al., 2010; Karunker et al., 2013; Sathyamoorthy et al., 2020. * signifies proteins with a stop/start overlap with NAPs, ^ signifies proteins not associated with a neighbouring NAP.

A search for homologues to these effectors in *Bdellovibrio* strains other than *HD100* via BlastP (Altschul et al., 1990) and PSI-Blast (Altschul, 1997), led to the discovery of a few proteins with sequence homology that did not have a neighbouring NAP as expected (Table 5.25). These proteins were: Bd1536 in *HD100*, A11Q_1652 in *B. exovorus*, B9G69_14775 in strain *SKB1291214*, and DOM22_16375 and DOM_16540 in strain *Zap7*. They were subsequently named 'orphans', and although *bd1536* in *HD100* has a neighbouring TF, *bd1537*, we have labelled it as an orphan because RNA-seq data has shown these proteins to be encoded in opposite direction within the genome (Karunker et al., 2013). These orphans have led to the new hypothesis that these proteins are not controlled by the upstream NAPs as originally hypothesised. This proposition is supported by divergent differential expression being observed between both the NAPs (Table 5.22) and the effectors (Table 5.24). Only 4 effectors are shown to have the same expression pattern as their NAP, *bd1218*, *bd2561*, *bd2624*, *bd2776*, 3 of which possess the 5 nt gene overlap, *bd1218*, *bd2624*, *bd2776*. Metallopeptidases generally exhibit a

broad activity range, and therefore are tightly regulated in order to prevent adverse effects within the cell (Cerdà-Costa and Gomis-Rüth, 2014). The updated hypothesis that the DUF4423 proteins are broad acting NAPs rather than specific TFs does not agree with the tight regulation required for metallopeptidases, further suggesting that expression of the two are not always linked. Seven of the 59 proteins identified across the 7 strains were found to have an identifiable HExxH motif. This further supports the theory that these are a group of metallopeptidases, in which the divergent effectors most likely use a distant histidine to stabilise the metal binding, as observed for the M14 or M15 groupings (Rawlings and Barrett, 1995).

	***Bdellovibrio* strain**						
Protein	***HD100***	***exovorus***	***SKB1291214***	***SSB218315***	***tiberius***	***W***	***ZAP7***
	Bd0682	A11Q_35	B9G69_04940!	B9G79_00890	Bdt0651		DOM22_02770
	Bd0723*	A11Q_209	B9G69_05340	B9G79_06190	Bdt0689		DOM22_06775
	Bd1171*	A11Q_1652^	B9G69_13580	B9G79_06470	Bdt1208		DOM22_08745!
	Bd1170		B9G69_16620	B9G79_09615	Bdt1393		DOM22_10840
	Bd1218*		B9G69_17140	B9G79_11705	Bdt1530		DOM22_11780!
	Bd1896		B9G69_17750!	B9G79_12645	Bdt1863		DOM22_12020
	Bd2552*		B9G69_18260	B9G79_14840	Bdt2538		DOM22_12120
	Bd2561		B9G69_14775^		Bdt2759		DOM22_12140
	Bd2624*				Bdt3642		DOM22_12470
	Bd2776*						DOM22_12485
	Bd3759*						DOM22_12960
	Bd1536^						DOM22_14110
							DOM22_14600
							DOM22_16745
							DOM22_17580!
							DOM22_18095
							DOM22_18270!
							DOM22_18555!
							DOM22_16375^
							DOM22_16540^
Ratio	18:12	6:3	12:8	12:7	15:9	4:0	31:20

Table 5.25 Table of effectors in 7 *Bdellovibrio* strains, with the ratio of HTH-DUF4423 proteins to effector proteins noted for each strain.* signifies proteins with a stop/start overlap with NAPs, ^ signifies proteins not associated with a neighbouring NAPs, ! signifies proteins with an identifiable HExxH motif.

Chapter 6 – Summary

6.1 Chemotaxis

The work presented here in Chapters 3 and 4 of this thesis adds to previous work on *Bdellovibrio* chemotaxis from the 1970s (Straley and Conti, 1974; LaMarre et al., 1977; Varon and Zeigler, 1978; Straley et al., 1979), and starts to build a more conclusive picture of the role of chemosensing in *Bdellovibrio* prey hunting by presenting the first ligands with a tangible prey specificity. By using both structural and biophysical techniques we have determined the ligands and preliminary binding *Kd* for 5 Cache domain proteins ($Bd1081^{LBD}$ showed binding to both choline and DAP; $Bd1081^{S\text{-}LBD}$ showed binding affinity for choline; $Bd2503^{LBD}$ cadaverine *Kd* ~ 1 μM and DAP *Kd* ~ 3 μM; $Bd2504^{LBD}$ binding to putrescine; and $Bd3092^{LBD}$ acetate *Kd* ~ 50 μM and propanoic acid *Kd* ~ 5 μM).

If the MCPs studied within this thesis are found to be involved in chemotaxis, we propose the following two mechanisms for prey hunting, 'Chum in the Water' and 'Lions at the Watering Hole' (Figure 6.1), which could be used in tandem and enable chemosensory information to influence *Bdellovibrio* in locating viable prey. The 'Chum in the Water' hypothesis is centred on the sensing of polyamines, and possible quaternary ammonium compounds (QACs), and relies on the release of these sensory molecules into the local area upon active predation. This would result in the recruitment of further *Bdellovibrio* to the area, resulting in large scale biofilm predation. It would be interesting to monitor predation of any prey species known to secrete polyamines rather than sequester them tightly at wall/outer membrane. Whereas the 'Lions at the Watering Hole' hypothesis is based on the sensing of QACs and organic acids which are essential molecules for most bacteria. Therefore prey are likely to move towards and accumulate in areas that contain high concentrations of these molecules (García et al., 2015; Shrestha et al., 2018). Herein *Bdellovibrio* would move towards areas with high

concentrations of these molecules to increase the likelihood of prey collision by hydrodynamic forces (Jashnsaz et al., 2017).

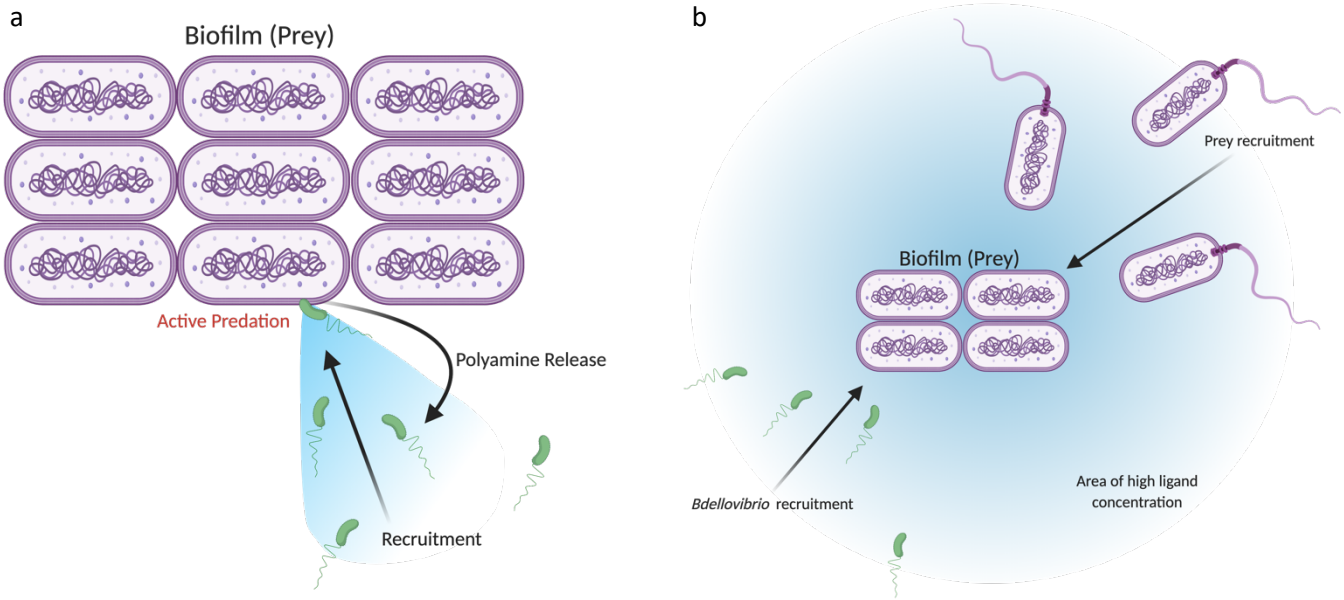


Figure 6.1 (a) Diagram of the 'Chum in the Water' *Bdellovibrio* hunting mechanism, in which the active predation of prey causes the release of polyamines, and possibly QACs, into the environment and consequently the recruitment of more *Bdellovibrios* to the site of predation. **(b)** Diagram of the 'Lions at the Watering Hole' *Bdellovibrio* hunting mechanism, in which *Bdellovibrio* move towards an area high in ubiquitous nutrients in order to increase the likelihood of finding prey. Created with BioRender.com

Although the work in this thesis has allowed for the development of theories on how *Bdellovibrio* utilise chemotaxis to hunt prey, there is still work to be done on validating this. Most notably chemosensory assays with whole cell *Bdellovibrio bacteriovorus HD100 WT* and pure solutions of the ligands presented here need to be performed to assess if the assumption laid out in these hypotheses that these ligands are chemo-attractants is correct. Knockout strains of *Bdellovibrio* for each dCache, HD100 *Δbd1081*, HD100 *Δbd2503-bd2504,* have already been developed by Rebecca Parr in collaboration with the Sockett and Hobley laboratories, however full testing of these is yet to take place. These knockouts would mean any chemotactic response to these ligands seen in WT could be attributed to a particular chemoreceptor as a result of a lack of overlapping specificity observed in our ITC experiments. *In vivo* Cache mutants wherein additional cysteines are introduced into the

transmembrane helices via point mutations to lock the chemoreceptor transmembrane helices into bound/apo forms could also be developed. If no chemotactic response by these mutants to their specific ligand(s) was observed it would support the conclusion made here that a piston movement is necessary for chemotactic response. Phenotypes could also be verified by use of point mutants in the cache binding pockets. Additional crystallographic work to obtain the structures of the Cache proteins with the additional ligands identified by ITC (Bd2503^{LBD}:DAP, Bd1081^{LBD}:choline, Bd1081^{LBD}:DAP, Bd1081$^{S\text{-}LBD}$:choline and Bd3092^{LBD}:propanoic acid) should also be undertaken. This is especially important in the case of Bd2503^{LBD}:DAP, as it would either confirm the predicted use of a water molecule as a bridge within the LBP or identify how the ligand is coordinated, e.g. showing DAP binds in the more classical polyamine binding confirmation as observed for Bd2504^{LBD}:putrescine and McpU^{LBD}:putrescine.

6.2 Novel Nucleoid-Associated Protein Family

The work presented here in Chapter 5 of this thesis enabled the partial structure of the DUF4423 domain to be established, as well as the function of DNA binding to be being attributed to both the DUF4423 domain and the HTH-DUF4423 family of proteins within *Bdellovibrio*. EMSAs with Bd3720 established that this family of proteins was far less specific than originally theorised, which led to the recategorization of these proteins from highly specific AP/GP switch transcription factors to non-specific proteins that function in a nucleoid associated protein-like (NAP) manner to regulate transcription at the genome level.

However, this work is only the beginning of truly understanding the role of this novel group of NAPs. Additional EMSAs using polyAT and polyGC DNA constructs would enable confirmation of the current theory that this family preferentially binds to AT rich regions and therefore confirm shape-readout as the mechanism by which this family recognises binding sites within DNA. Broadening the testing of

DNA fragment specificity via EMSAs to other members of the family would also validate that this family acts via non-specific interactions with DNA. Further crystallographic studies would be beneficial, both with and without DNA bound. This should not only allow the full-length structure of the DUF4423 domain to be obtained but also the full HTH-DUF4423 architecture, which would enable us to understand how the two domains interact with one another. A full-length DNA-bound structure would enable us to determine if both domains interact with DNA as currently hypothesised. This would also reveal if any conformational changes occur upon the binding of DNA and whether the predicted tetrameric structure is maintained, whilst also starting to allow us to validate the hypothesis that these NAPs regulate transcription via DNA bridging. Further work using scanning atomic force microscopy of this family bound to DNA should enable validation of this hypothesis.

Phenotyping work using *Bdellovibrio bacteriovorus HD100* knockouts of these NAPs would also provide a wealth of information on their role within *Bdellovibrio*, although the duplication of genes may render this a difficult task. Knockouts of *cdbA*, which is thought to fulfil a similar role within *Myxococcus xanthus,* resulted in a strong phenotype of aberrant DNA organisation (Skotnicka et al., 2020), and if the same proved true for this family in *Bdellovibrio* it would support the theory that they function in a NAP-like manner. Differential gene expression studies between knockouts and WT (potentially via RNAseq) would enable a better understanding of which genes are regulated by these NAP-like proteins, and therefore whether they are responsible for the repression or activation of the genes they regulate. NAPs have been shown to be involved in activation, counter-silencing and repression (Hołówka and Zakrzewska-Czerwińska, 2020; McFarland et al., 2008; Flores-Ríos et al., 2019; Wiechert et al., 2019). However because of the expression pattern of the family (Lambert et al., 2010; Sathyamoorthy et al., 2020; Karunker et al., 2013) and DNA bridging being most-commonly associated with gene repression (Hołówka and Zakrzewska-Czerwińska, 2020), we propose that this family of proteins is most likely involved in gene repression. Finally, experiments into the environmental factor(s) that regulates the expression and/or DNA binding capacity of these NAPs

needs to be performed to understand how the action of these NAPs is controlled and directed, as the non-specific binding described in this thesis rules out DNA sequence specificity as a manner of controlling NAP:DNA interaction. Although counter-silencing has currently only been described as a mechanism of regulating genes obtained via horizontal gene transfer (Will et al., 2015; Wiechert et al., 2019) it does highlight a possible mechanism by which the interaction between DNA and this family of NAPs could be controlled by another protein(s).

References

1. Abdul-Gader A, Miles AJ, Wallace BA. A reference dataset for the analyses of membrane protein secondary structures and transmembrane residues using circular dichroism spectroscopy. Bioinformatics. 2011 Jun 15;27(12):1630-6. doi: 10.1093/bioinformatics/btr234.

2. Adams, P.D., Afonine, P. V., Bunkóczi, G., et al. (2010) PHENIX : a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D Biological Crystallography*, 66 (2): 213–221. doi:10.1107/S0907444909052925.

3. Afonine, P. V., Grosse-Kunstleve, R.W., Echols, N., et al. (2012) Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallographica Section D Biological Crystallography*, 68 (4): 352–367. doi:10.1107/S0907444912001308.

4. Altschul, S. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 25 (17): 3389–3402. doi:10.1093/nar/25.17.3389.

5. Altschul, S.F., Gish, W., Miller, W., et al. (1990) Basic local alignment search tool. *Journal of Molecular Biology*, 215 (3): 403–410. doi:10.1016/S0022-2836(05)80360-2.

6. Anantharaman, V. and Aravind, L. (2000) Cache - A signaling domain common to animal Ca2+-channel subunits and a class of prokaryotic chemotaxis receptors. *Trends in Biochemical Sciences*, 25 (11): 535–537. doi:10.1016/S0968-0004(00)01672-8.

7. Atlung, T. and Ingmer, H. (1997) H-NS: a modulator of environmentally regulated gene expression. *Molecular Microbiology*, 24 (1): 7–17. doi:10.1046/j.1365-2958.1997.3151679.x.

8. Atterbury, R.J., Hobley, L., Till, R., et al. (2011) Effects of orally administered Bdellovibrio bacteriovorus on the well-being and Salmonella colonization of young chicks. *Applied and Environmental Microbiology*, 77 (16): 5794–5803. doi:10.1128/AEM.00426-11.

9. Baker, M.D., Wolanin, P.M. and Stock, J.B. (2006) Signal transduction in bacterial chemotaxis. *BioEssays*, 28 (1): 9–22. doi:10.1002/bies.20343.

10. Beebe, J.M. (1941) *Studies on the myxobacteria: I, Distribution in Iowa soils and description of a new species; II, Myxobacteria as bacterial parasites; III, The morphology and cytology of Myxococcus xanthus sp.n.* Iowa State University, Digital Repository. doi:10.31274/rtd-180813-14570.

11. Bi, S. and Lai, L. (2015) Bacterial chemoreceptors and chemoeffectors. *Cellular and Molecular Life Sciences*, 72 (4): 691–708. doi:10.1007/s00018-014-1770-5.

12. Bi, S. and Sourjik, V. (2018) Stimulus sensing and signal processing in bacterial chemotaxis. *Current Opinion in Microbiology*, 45: 22–29. doi:10.1016/j.mib.2018.02.002.

13. Boivin, S., Kozak, S. and Meijers, R. (2013) Optimization of protein purification and characterization using Thermofluor screens. *Protein Expression and Purification*, 91 (2): 192–206. doi:10.1016/j.pep.2013.08.002.

14. Bonfiglio, G., Neroni, B., Radocchia, G., et al. (2020a) Growth control of adherent-invasive escherichia coli (Aiec) by the predator bacteria bdellovibrio bacteriovorus: A new therapeutic approach for crohn's disease patients. *Microorganisms*, 8 (1): 1–14. doi:10.3390/microorganisms8010017.

15. Bonfiglio, G., Neroni, B., Radocchia, G., et al. (2020b) Insight into the possible use of the predator bdellovibrio bacteriovorus as a probiotic. *Nutrients*, 12 (8): 1–18. doi:10.3390/nu12082252.

16. Borgnia, M.J., Subramaniam, S. and Milne, J.L.S. (2008) Three-dimensional imaging of the highly bent architecture of Bdellovibrio bacteriovorus by using cryo-electron tomography. *Journal of Bacteriology*, 190 (7): 2588–2596. doi:10.1128/JB.01538-07.

17. Brown, S., Santa Maria, J.P. and Walker, S. (2013) Wall teichoic acids of gram-positive bacteria. *Annual Review of Microbiology*, 67 (6): 313–336. doi:10.1146/annurev-micro-092412-155620.

18. de Bruijn, I., Cheng, X., de Jager, V., et al. (2015) Comparative genomics and metabolic profiling of the genus Lysobacter. *BMC Genomics*, 16 (1): 1–16. doi:10.1186/s12864-015-2191-z.

19. Bryant, J.A., Cadby, I.T., Chong, Z.-S., et al. (2019) Structure-function analysis of the Escherichia coli β–barrel assembly enhancing protease BepA suggests a role for a self-inhibitory state. *bioRxiv*. doi:10.1101/689117.

20. Bureš, F. (2019) Quaternary Ammonium Compounds: Simple in Structure, Complex in Application. *Topics in Current Chemistry*, 377 (3): 1–21. doi:10.1007/s41061-019-0239-2.

21. Burnham, J.C., Hashimoto, T. and Conti, S.F. (1970) Ultrastructure and cell division of a facultatively parasitic strain of Bdellovibrio bacteriovorus. *Journal of Bacteriology*, 101 (3): 997–1004. doi:10.1128/jb.101.3.997-1004.1970.

22. Burrell, M., Hanfrey, C.C., Murray, E.J., et al. (2010) Evolution and multiplicity of arginine decarboxylases in polyamine biosynthesis and essential role in Bacillus subtilis biofilm formation. *Journal of Biological Chemistry*, 285 (50): 39224–39238. doi:10.1074/jbc.M110.163154.

23. Butan, C., Hartnell, L.M., Fenton, A.K., et al. (2011) Spiral architecture of the nucleoid in Bdellovibrio bacteriovorus. *Journal of Bacteriology*, 193 (6): 1341–1350. doi:10.1128/JB.01061-10.

24. Capeness, M.J., Lambert, C., Lovering, A.L., et al. (2013) Activity of Bdellovibrio hit locus proteins, bd0108 and bd0109, links type IVa pilus extrusion/retraction status to prey-independent growth signalling. *PLoS ONE*, 8 (11). doi:10.1371/journal.pone.0079759.

25. Carrera, P. and Azorin, F. (1994) Structural characterization of intrinsically curved AT-rich DNA sequences. *Nucleic Acids Research*, 22 (18): 3671–3680. doi:10.1093/nar/22.18.3671.

26. Casida, L.E. (1982) Ensifer adhaerens gen. nov., sp. nov.: A bacterial predator of bacteria in soil. *International Journal of Systematic Bacteriology*, 32 (3): 339–345. doi:10.1099/00207713-32-3-339.

27. Casida, L.E. (1988) Minireview: Nonobligate bacterial predation of bacteria in soil. *Microbial Ecology*, 15 (1): 1–8. doi:10.1007/BF02012948.

28. Casida, L.E. (1989) Protozoan Response to the Addition of Bacterial Predators and Other Bacteria to Soil †. *Applied and Environmental Microbiology*, 55 (8): 1857–1859. doi:10.1128/aem.55.8.1857-1859.1989.

29. Cavalier-Smith, T. (2009) Predation and eukaryote cell origins: A coevolutionary perspective. *International Journal of Biochemistry and Cell Biology*, 41 (2): 307–322. doi:10.1016/j.biocel.2008.10.002.

30. Cerdà-Costa, N. and Gomis-Rüth, F.X. (2014) Architecture and function of metallopeptidase catalytic domains. *Protein Science*, 23 (2): 123–144. doi:10.1002/pro.2400.

31. Charlier, P., Wery, J.-P., Dideberg, O., et al. (2006) "Streptomyces Albus G D -Ala- D -Ala Carboxypeptidase." In *Handbook of Metalloproteins*. Chichester, UK: John Wiley & Sons, Ltd. doi:10.1002/0470028637.met032.

32. Chen, H., Athar, R., Zheng, G., et al. (2011) Prey bacteria shape the community structure of their predators. *ISME Journal*, 5 (8): 1314–1322. doi:10.1038/ismej.2011.4.

33. Christensen, P. and Cook, F.D. (1978) Lysobacter, a new genus of nonfruiting, gliding bacteria with a high base ratio. *International Journal of Systematic Bacteriology*, 28 (3): 367–393. doi:10.1099/00207713-28-3-367.

34. Collins, K.D., Lacal, J. and Ottemann, K.M. (2014) Internal Sense of Direction: Sensing and Signaling from Cytoplasmic Chemoreceptors. *Microbiology and Molecular Biology Reviews*, 78 (4): 672–684. doi:10.1128/mmbr.00033-14.

35. Compton, L.A. and Johnson, W.C., Jr. (1986) Analysis of protein circular dichroism spectra for secondary structure using a simple matrix multiplication. Anal. Biochem. 155, 155-167. doi: 10.1016/0003-2697(86)90241-1

36. Corral-Lugo, A., De la Torre, J., Matilla, M.A., et al. (2016) Assessment of the contribution of chemoreceptor-based signalling to biofilm formation. *Environmental Microbiology*, 18 (10): 3355–3372. doi:10.1111/1462-2920.13170.

37. Corral-Lugo, A., Matilla, M.A., Martín-Mora, D., et al. (2018) High-Affinity Chemotaxis to Histamine Mediated by the TlpQ Chemoreceptor of the Human Pathogen Pseudomonas aeruginosa Hazelbauer, G.L. and Msadek, T. (eds.). *mBio*, 9 (6). doi:10.1128/mBio.01894-18.

38. Cotter, T.W. and Thomashow, M.F. (1992a) A conjugation procedure for Bdellovibrio bacteriovorus and its use to identify DNA sequences that enhance the plaque-forming ability of a spontaneous host-independent mutant. *Journal of Bacteriology*, 174 (19): 6011–6017. doi:10.1128/jb.174.19.6011-6017.1992.

39. Cotter, T.W. and Thomashow, M.F. (1992b) Identification of a Bdellovibrio bacteriovorus genetic locus, hit, associated with the host-independent phenotype. *Journal of Bacteriology*, 174 (19): 6018–6024. doi:10.1128/jb.174.19.6018-6024.1992.

40. Cournac, A. and Plumbridge, J. (2013) DNA Looping in Prokaryotes: Experimental and theoretical approaches. *Journal of Bacteriology*, 195 (6): 1109–1119. doi:10.1128/JB.02038-12.

41. Cronan, J.E. and Thomas, J. (2009) *Chapter 17 Bacterial Fatty Acid Synthesis and its Relationships with Polyketide Synthetic Pathways*. In pp. 395–433. doi:10.1016/S0076-6879(09)04617-5.

42. Daft, M.J. and Stewart, W.D.P. (1973) Light and Electron Microscpoe Observations on Algal Lysis by Bacterium CP-I. *New Phytologist*, 72 (4): 799–808. doi:10.1111/j.1469-8137.1973.tb02055.x.

43. Dangi, B., Gronenborn, A.M., Rosner, J.L., et al. (2004) Versatility of the carboxy-terminal domain of the α subunit of RNA polymerase in transcriptional activation: use of the DNA contact site as a protein contact site for MarA. *Molecular Microbiology*, 54 (1): 45–59. doi:10.1111/j.1365-2958.2004.04250.x.

44. Dashiff, A., Junka, R.A., Libera, M., et al. (2011) Predation of human pathogens by the predatory bacteria Micavibrio aeruginosavorus and Bdellovibrio bacteriovorus. *Journal of Applied Microbiology*, 110 (2): 431–444. doi:10.1111/j.1365-2672.2010.04900.x.

45. Davidov, Y., Huchon, D., Koval, S.F., et al. (2006) A new α-proteobacterial clade of Bdellovibrio-like predators: Implications for the mitochondrial endosymbiotic theory. *Environmental Microbiology*, 8 (12): 2179–2188. doi:10.1111/j.1462-2920.2006.01101.x.

46. Davidov, Y. and Jurkevitch, E. (2009) Predation between prokaryotes and the origin of eukaryotes. *BioEssays*, 31 (7): 748–757. doi:10.1002/bies.200900018.

47. Deeg, C.M., Le, T.T., Zimmer, M.M., et al. (2018) From the inside out: An epibiotic Bdellovibrio predator with an expanded genomic complement. *bioRxiv*, 202 (8). doi:10.1101/506055.

48. Von Dohlen, C.D., Kohler, S., Alsop, S.T., et al. (2001) Mealybug β-proteobacterial endosymbionts contain γ-proteobacterial symbionts. *Nature*, 412 (6845): 433–436. doi:10.1038/35086563.

49. Donaldson, L.W., Petersen, J.M., Graves, B.J., et al. (1996) Solution structure of the ETS domain from murine Ets-1: A winged helix-turn-helix DNA binding motif. *EMBO Journal*, 15 (1): 125–134. doi:10.1002/j.1460-2075.1996.tb00340.x.

50. Duigou, S. and Boccard, F. (2017) Long range chromosome organization in Escherichia coli: The position of the replication origin defines the non-structured regions and the Right and Left macrodomains. *PLoS Genetics*, 13 (5). doi:10.1371/journal.pgen.1006758.

51. Dupaigne, P., Tonthat, N.K., Espéli, O., et al. (2012) Molecular Basis for a Protein-Mediated DNA-Bridging Mechanism that Functions in Condensation of the E. coli Chromosome. *Molecular Cell*, 48 (4): 560–571. doi:10.1016/j.molcel.2012.09.009.

52. Edwards, P.M. (2002) Origin 7.0: Scientific graphing and data analysis software. *Journal of Chemical Information and Computer Sciences*, 42 (5): 1270. doi:10.1021/ci0255432.

53. Emsley, P., Lohkamp, B., Scott, W.G., et al. (2010) Features and development of Coot. *Acta Crystallographica Section D Biological Crystallography*, 66 (4): 486–501. doi:10.1107/S0907444910007493.

54. Enos, B.G., Anthony, M.K., DeGiorgis, J.A., et al. (2017) Prey range and genome evolution of Halobacteriovorax marinus predatory bacteria from an estuary. *bioRxiv*, 3 (1): 1–14. doi:10.1101/180265.

55. Erken, M., Lutz, C. and McDougald, D. (2013) The Rise of Pathogens: Predation as a Factor Driving the Evolution of Human Pathogens in the Environment. *Microbial Ecology*, 65 (4): 860–868. doi:10.1007/s00248-013-0189-0.

56. Evans, K.J., Lambert, C. and Sockett, R.E. (2007) Predation by Bdellovibrio bacteriovorus HD100 requires type IV pili. *Journal of Bacteriology*, 189 (13): 4850–4859. doi:10.1128/JB.01942-06.

57. Expósito, R.G., Postma, J., Raaijmakers, J.M., et al. (2015) Diversity and activity of Lysobacter species from disease suppressive soils. *Frontiers in Microbiology*, 6 (NOV): 1–12. doi:10.3389/fmicb.2015.01243.

58. Fennema, D., Phillips, I.R. and Shephard, E.A. (2016) Trimethylamine and trimethylamine N-oxide, a Flavin-Containing Monooxygenase 3 (FMO3)-mediated host-microbiome metabolic axis implicated in health and disease. *Drug Metabolism and Disposition*, 44 (11): 1839–1850. doi:10.1124/dmd.116.070615.

59. Fenton, A.K., Kanna, M., Woods, R.D., et al. (2010) Shadowing the actions of a predator: Backlit fluorescent microscopy reveals synchronous nonbinary septation of predatory Bdellovibrio inside prey and exit through discrete bdelloplast pores. *Journal of Bacteriology*, 192 (24): 6329–6335. doi:10.1128/JB.00914-10.

60. Ferguson, M.A., Núñez, M.E., Kim, H.J., et al. (2014) Spatially organized films from Bdellovibrio bacteriovorus prey lysates. *Applied and Environmental Microbiology*, 80 (23): 7405–7414. doi:10.1128/AEM.02423-14.

61. Flores-Ríos, R., Quatrini, R. and Loyola, A. (2019) Endogenous and Foreign Nucleoid-Associated Proteins of Bacteria: Occurrence, Interactions and Effects on Mobile Genetic Elements and Host's Biology. *Computational and Structural Biotechnology Journal*, 17: 746–756. doi:10.1016/j.csbj.2019.06.010.

62. Gajiwala, K.S. and Burley, S.K. (2000) Winged helix proteins. *Current Opinion in Structural Biology*, 10 (1): 110–116. doi:10.1016/S0959-440X(99)00057-3.

63. Gallivan, J.P. and Dougherty, D.A. (1999) Cation-π interactions in structural biology. *Proceedings of the National Academy of Sciences of the United States of America*, 96 (17): 9459–9464. doi:10.1073/pnas.96.17.9459.

64. Ganuza, E., Sellers, C.E., Bennett, B.W., et al. (2016) A novel treatment protects Chlorella at commercial scale from the predatory bacterium Vampirovibrio chlorellavorus. *Frontiers in Microbiology*, 7 (JUN): 1–13. doi:10.3389/fmicb.2016.00848.

65. García, V., Reyes-Darias, J.A., Martín-Mora, D., et al. (2015) Identification of a chemoreceptor for C2 and C3 carboxylic acids. *Applied and Environmental Microbiology*, 81 (16): 5449–5457. doi:10.1128/AEM.01529-15.

66. Garst, E., Kivensen, V., Hughes, E., et al. (2013) Exploration of Bdellovibrio Chemotaxis and Predation using Microfluidics. *Biophysical Journal*, 104 (2): 639a. doi:10.1016/j.bpj.2012.11.3529.

67. Gasteiger, E., Hoogland, C., Gattiker, A., et al. (2005) "Protein Identification and Analysis Tools on the ExPASy Server." In *The Proteomics Protocols Handbook*. Totowa, NJ: Humana Press. pp. 571–607. doi:10.1385/1-59259-890-0:571.

68. Gavira, J.A., Gumerov, V.M., Rico-Jiménez, M., et al. (2020) How Bacterial Chemoreceptors Evolve Novel Ligand Specificities Groisman, E.A. (ed.). *mBio*, 11 (1). doi:10.1128/mBio.03066-19.

69. Gavira, J.A., Ortega, Á., Martín-Mora, D., et al. (2018) Structural Basis for Polyamine Binding at the dCACHE Domain of the McpU Chemoreceptor from Pseudomonas putida. *Journal of Molecular Biology*, 430 (13): 1950–1963. doi:10.1016/j.jmb.2018.05.008.

70. Geiger, O., López-Lara, I.M. and Sohlenkamp, C. (2013) Phosphatidylcholine biosynthesis and function in bacteria. *Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids*, 1831 (3): 503–513. doi:10.1016/j.bbalip.2012.08.009.

71. Gray, K.M. and Ruby, E.G. (1991) "Intercellular Signalling in the Bdellovibrio Developmental Growth Cycle." In *Microbial cell-cell interactions*. pp. 333–366. doi:10.1016/0962-8924(92)90253-J.

72. Greenfield, N.J. (2007) Using circular dichroism spectra to estimate protein secondary structure. *Nature Protocols*, 1 (6): 2876–2890. doi:10.1038/nprot.2006.202.

73. Grenha, R., Slamti, L., Nicaise, M., et al. (2013) Structural basis for the activation mechanism of the PlcR virulence regulator by the quorum-sensing signal peptide PapR. *Proceedings of the National Academy of Sciences*, 110 (3): 1047–1052. doi:10.1073/pnas.1213770110.

74. Griffin, J.N., Byrnes, J.E.K. and Cardinale, B.J. (2013) *Effects of predator richness on prey suppression : a meta-analysis Author (s): John N . Griffin , Jarrett E . K . Byrnes and Bradley J . Cardinale Published by : Wiley on behalf of the Ecological Society of America Stable URL : http://www.jstor.org/stab*., 94 (10): 2180–2187.

75. Guerrero, R., Pedros-Alio, C., Esteve, I., et al. (1986) Predatory prokaryotes: Predation and primary consumption evolved in bacteria. *Proceedings of the National Academy of Sciences of the United States of America*, 83 (7): 2138–2142. doi:10.1073/pnas.83.7.2138.

76. Hahn, M.W., Schmidt, J., Koll, U., et al. (2017) Silvanigrella aquatica gen. nov., sp. nov., isolated from a freshwater lake, description of silvanigrellaceae fam. nov. and silvanigrellales ord. nov., reclassification of the order bdellovibrionales in the class oligoflexia, reclassification of the famil. *International Journal of Systematic and Evolutionary Microbiology*, 67 (8): 2555–2568. doi:10.1099/ijsem.0.001965.

77. Hancock, S.P., Stella, S., Cascio, D., et al. (2016) DNA sequence determinants controlling affinity, stability and shape of DNA complexes bound by the nucleoid protein Fis. *PLoS ONE*, 11 (3): 1–24. doi:10.1371/journal.pone.0150189.

78. Harini, K., Ajila, V. and Hegde, S. (2013) Bdellovibrio bacteriovorus: A future antimicrobial agent? *Journal of Indian Society of Periodontology*, 17 (6): 823–825. doi:10.4103/0972-124X.124534.

79. Harms, A., Treuner-Lange, A., Schumacher, D., et al. (2013) Tracking of Chromosome and Replisome Dynamics in Myxococcus xanthus Reveals a Novel Chromosome Arrangement. *PLoS Genetics*, 9 (9): 19–23. doi:10.1371/journal.pgen.1003802.

80. Hazelbauer, G.L. and Lai W. (2010) Bacterial chemoreceptors: providing enhanced features to two-component signaling. Current Opinion in Microbiology, 13 (2): 124-132. doi.org/10.1016/j.mib.2009.12.014.

81. Hibbing, M.E., Fuqua, C., Parsek, M.R., et al. (2010) Bacterial competition: surviving and thriving in the microbial jungle. *Nature Reviews Microbiology*, 8 (1): 15–25. doi:10.1038/nrmicro2259.

82. Hobley, L., Fung, R.K.Y., Lambert, C., et al. (2012a) Discrete cyclic di-GMP-dependent control of bacterial predation versus axenic growth in Bdellovibrio bacteriovorus. *PLoS Pathogens*, 8 (2). doi:10.1371/journal.ppat.1002493.

83. Hobley, L., Kim, S.H., Maezato, Y., et al. (2014) Norspermidine is not a self-produced trigger for biofilm disassembly. *Cell*, 156 (4): 844–854. doi:10.1016/j.cell.2014.01.012.

84. Hobley, L., Lerner, T.R., Williams, L.E., et al. (2012b) Genome analysis of a simultaneously predatory and prey-independent, novel Bdellovibrio bacteriovorus from the River Tiber, supports in silico predictions of both ancient and recent lateral gene transfer from diverse bacteria. *BMC Genomics*, 13 (1): 1–13. doi:10.1186/1471-2164-13-670.

85. Holm, L. (2020) DALI and the persistence of protein shape. *Protein Science*, 29 (1): 128–140. doi:10.1002/pro.3749.

86. Holm, L., Kääriäinen, S., Rosenström, P., et al. (2008) Searching protein structure databases with DaliLite v.3. *Bioinformatics*, 24 (23): 2780–2781. doi:10.1093/bioinformatics/btn507.

87. Hołówka, J. and Zakrzewska-Czerwińska, J. (2020) Nucleoid Associated Proteins: The Small Organizers That Help to Cope With Stress. *Frontiers in Microbiology*, 11 (April): 1–7. doi:10.3389/fmicb.2020.00590.

88. Hsieh, T.J., Yen, T.J., Lin, T.S., et al. (2010) Twisting of the DNA-binding surface by a β-strand-bearing proline modulates DNA gyrase activity. *Nucleic Acids Research*, 38 (12): 4173–4181. doi:10.1093/nar/gkq153.

89. Hungate, B.A., Marks, J.C., Power, M.E., et al. (2021) The Functional Significance of Bacterial Predators. *bioRxiv*. doi:10.1101/2021.02.22.432408.

90. Iebba, V., Totino, V., Santangelo, F., et al. (2014) Bdellovibrio bacteriovorus directly attacks Pseudomonas aeruginosa and Staphylococcus aureus cystic fibrosis isolates. *Frontiers in Microbiology*, 5 (JUN): 1–9. doi:10.3389/fmicb.2014.00280.

91. Im, H., Dwidar, M. and Mitchell, R.J. (2018) Bdellovibrio bacteriovorus HD100, a predator of Gram-negative bacteria, benefits energetically from Staphylococcus aureus biofilms without predation. *ISME Journal*, 12 (8): 2090–2095. doi:10.1038/s41396-018-0154-5.

92. Jain, D., Kim, Y., Maxwell, K.L., et al. (2005) Crystal Structure of Bacteriophage λcII and Its DNA Complex. *Molecular Cell*, 19 (2): 259–269. doi:10.1016/j.molcel.2005.06.006.

93. Jameson, E., Quareshy, M. and Chen, Y. (2018) Methodological considerations for the identification of choline and carnitine-degrading bacteria in the gut. *Methods*, 149: 42–48. doi:10.1016/j.ymeth.2018.03.012.

94. Jashnsaz, H., Anderson, G.G. and Pressé, S. (2017a) Statistical signatures of a targeted search by bacteria. *Physical Biology*, 14 (6): 065002. doi:10.1088/1478-3975/aa84ea.

95. Jashnsaz, H., Al Juboori, M., Weistuch, C., et al. (2017b) Hydrodynamic Hunters. *Biophysical Journal*, 112 (6): 1282–1289. doi:10.1016/j.bpj.2017.02.011.

96. Jun, H.S., Qi, M., Gong, J., et al. (2007) Outer membrane proteins of Fibrobacter succinogenes with potential roles in adhesion to cellulose and in cellulose digestion. *Journal of Bacteriology*, 189 (19): 6806–6815. doi:10.1128/JB.00560-07.

97. Jurkevitch, E. (2007) *Predatory Prokaryotes*. In Jurkevitch, E. (ed.) . Microbiology Monographs. Springer Berlin Heidelberg. doi:10.1007/978-3-540-38582-0.

98. Kabsch, W. (2010) XDS. *Acta Crystallographica Section D Biological Crystallography*, 66 (2): 125–132. doi:10.1107/S0907444909047337.

99. Kadouri, D. and O'Toole, G.A. (2005) Susceptibility of biofilms to Bdellovibrio bacteriovorus attack. *Applied and Environmental Microbiology*, 71 (7): 4044–4051. doi:10.1128/AEM.71.7.4044-4051.2005.

100. Kamio, Y. and Nakamura, K. (1987) Putrescine and cadaverine are constituents of peptidoglycan in Veillonella alcalescens and Veillonella parvula. *Journal of Bacteriology*, 169 (6): 2881–2884. doi:10.1128/JB.169.6.2881-2884.1987.

101. Kandel, P.P., Pasternak, Z., van Rijn, J., et al. (2014) Abundance, diversity and seasonal dynamics of predatory bacteria in aquaculture zero discharge systems. *FEMS Microbiology Ecology*, 89 (1): 149–161. doi:10.1111/1574-6941.12342.

102. Karunker, I., Rotem, O., Dori-Bachash, M., et al. (2013) A Global Transcriptional Switch between the Attack and Growth Forms of Bdellovibrio bacteriovorus. *PLoS ONE*, 8 (4). doi:10.1371/journal.pone.0061850.

103. Kelley, L.A., Mezulis, S., Yates, C.M., et al. (2015) The Phyre2 web portal for protein modeling, prediction and analysis. *Nature Protocols*, 10 (6): 845–858. doi:10.1038/nprot.2015.053.

104. Kiene, R.P. (1998) Uptake of choline and its conversion to glycine betaine by bacteria in estuarine waters? *Applied and Environmental Microbiology*, 64 (3): 1045–1051. doi:10.1128/aem.64.3.1045-1051.1998.

105. Kim, H. and Lipscomb, W.N. (1993) Differentiation and identification of the two catalytic metal binding sites in bovine lens leucine aminopeptidase by x-ray crystallography. *Proceedings of the National Academy of Sciences*, 90 (11): 5006–5010. doi:10.1073/pnas.90.11.5006.

106. Kim, K.K., Yokota, H. and Kim, S.H. (1999) Four-helical-bundle structure of the cytoplasmic domain of a serine chemotaxis receptor. *Nature*, 400 (6746): 787–792. doi:10.1038/23512.

107. Kirby, J.R. (2009) Chemotaxis-like regulatory systems: Unique roles in diverse bacteria. *Annual Review of Microbiology*, 63: 45–59. doi:10.1146/annurev.micro.091208.073221.

108. Knauber, T., Doss, S.D., Gerth, K., et al. (2008) Mutation in the rel gene of Sorangium cellulosum affects morphological and physiological differentiation. *Molecular Microbiology*, 69 (1): 254–266. doi:10.1111/j.1365-2958.2008.06285.x.

109. Kojima, S., Kaneko, J., Abe, N., et al. (2011) Cadaverine covalently linked to the peptidoglycan serves as the correct constituent for the anchoring mechanism between the outer membrane and peptidoglycan in Selenomonas ruminantium. *Journal of Bacteriology*, 193 (9): 2347–2350. doi:10.1128/JB.00106-11.

110. Koskiniemi, S., Lamoureux, J.G., Nikolakakis, K.C., et al. (2013) Rhs proteins from diverse bacteria mediate intercellular competition. *Proceedings of the National Academy of Sciences of the United States of America*, 110 (17): 7032–7037. doi:10.1073/pnas.1300627110.

111. Kracke, F., Vassilev, I. and Krömer, J.O. (2015) Microbial electron transport and energy conservation - The foundation for optimizing bioelectrochemical systems. *Frontiers in Microbiology*, 6 (JUN): 1–18. doi:10.3389/fmicb.2015.00575.

112. Krissinel, E. (2010) Crystal contacts as nature's docking solutions. *Journal of Computational Chemistry*, 31 (1): 133–143. doi:10.1002/jcc.21303.

113. Krissinel, E. and Henrick, K. (2007) Inference of Macromolecular Assemblies from Crystalline State. *Journal of Molecular Biology*, 372 (3): 774–797. doi:10.1016/j.jmb.2007.05.022.

114. Kuenen, J.G. and Rittenberg, S.C. (1975) Incorporation of long chain fatty acids of the substrate organism by Bdellovibrio bacteriovorus during intraperiplasmic growth. *Journal of Bacteriology*, 121 (3): 1145–1157. doi:10.1128/jb.121.3.1145-1157.1975.

115. Kuru, E., Lambert, C., Rittichier, J., et al. (2017) Fluorescent D-amino-acids reveal bi-cellular cell wall modifications important for Bdellovibrio bacteriovorus predation. *Nature Microbiology*, 2 (12): 1648–1657. doi:10.1038/s41564-017-0029-y.

116. Kuzin, A.P., Chen, Y., Seetharaman, J., et al. (2008) *X-ray structure of the putative Zn-dependent peptidase Q74D82 at the resolution 1.7 A*.

117. Kwon, H.J., Bennik, M.H.J., Demple, B., et al. (2000) Crystal structure of the Escherichia coli Rob transcription factor in complex with DNA. *Nature Structural Biology*, 7 (5): 424–430. doi:10.1038/75213.

118. LaMarre, A.G., Straley, S.C. and Conti, S.F. (1977) Chemotaxis toward amino acids by Bdellovibrio bacteriovorus. *Journal of Bacteriology*, 131 (1): 201–207. doi:10.1128/JB.131.1.201-207.1977.

119. Lambert, C., Cadby, I.T., Till, R., et al. (2015) Ankyrin-mediated self-protection during cell invasion by the bacterial predator Bdellovibrio bacteriovorus. *Nature Communications*, 6: 1–9. doi:10.1038/ncomms9884.

120. Lambert, C., Chang, C.Y., Capeness, M.J., et al. (2010) The first bite - Profiling the predatosome in the bacterial pathogen Bdellovibrio. *PLoS ONE*, 5 (1): 1–16. doi:10.1371/journal.pone.0008599.

121. Lambert, C., Evans, K.J., Till, R., et al. (2006) Characterizing the flagellar filament and the role of motility in bacterial prey-penetration by Bdellovibrio bacteriovorus. *Molecular Microbiology*, 60 (2): 274–286. doi:10.1111/j.1365-2958.2006.05081.x.

122. Lambert, C., Fenton, A.K., Hobley, L., et al. (2011) Predatory Bdellovibrio bacteria use gliding motility to scout for prey on surfaces. *Journal of Bacteriology*, 193 (12): 3139–3141. doi:10.1128/JB.00224-11.

123. Lambert, C., Hobley, L., Chang, C.-Y., et al. (2008) "A Predatory Patchwork: Membrane and Surface Structures of Bdellovibrio bacteriovorus." In *Advances in Microbial Physiology*. pp. 313–361. doi:10.1016/S0065-2911(08)00005-2.

124. Lambert, C., Lerner, T.R., Bui, N.K., et al. (2016) Interrupting peptidoglycan deacetylation during Bdellovibrio predator-prey interaction prevents ultimate destruction of prey wall, liberating bacterial-ghosts. *Scientific Reports*, 6 (April): 1–19. doi:10.1038/srep26010.

125. Lee, N., Francklyn, C. and Hamilton, E.P. (1987) Arabinose-induced binding of AraC protein to araI2 activates the araBAD operon promoter. *Proceedings of the National Academy of Sciences of the United States of America*, 84 (24): 8814–8818. doi:10.1073/pnas.84.24.8814.

126. Leone, S., Sannino, F., Tutino, M.L., et al. (2015) Acetate: Friend or foe? Efficient production of a sweet protein in Escherichia coli BL21 using acetate as a carbon source. *Microbial Cell Factories*, 14 (1): 1–10. doi:10.1186/s12934-015-0299-0.

127. Lerner, T.R., Lovering, A.L., Bui, N.K., et al. (2012) Specialized peptidoglycan hydrolases sculpt the intra-bacterial niche of predatory Bdellovibrio and increase population fitness. *PLoS Pathogens*, 8 (2). doi:10.1371/journal.ppat.1002524.

128. Lewin, R.A. (1997) Saprospira grandis: A flexibacterium that can catch bacterial prey by "Ixotrophy." *Microbial Ecology*, 34 (3): 232–236. doi:10.1007/s002489900052.

129. Li, C.Y., Chen, X.L., Shao, X., et al. (2015a) Mechanistic insight into trimethylamine N-oxide recognition by the marine bacterium Ruegeria pomeroyi DSS-3. *Journal of Bacteriology*, 197 (21): 3378–3387. doi:10.1128/JB.00542-15.

130. Li, Q., Tachie-Baffour, Y., Liu, Z., et al. (2015b) Non-classical amine recognition evolved in a large clade of olfactory receptors. *eLife*, 4: 1–17. doi:10.7554/eLife.10441.

131. Lin, M. and Guo, J.T. (2019) New insights into protein-DNA binding specificity from hydrogen bond based comparative study. *Nucleic acids research*, 47 (21): 11103–11113. doi:10.1093/nar/gkz963.

132. Lioy, V.S., Cournac, A., Marbouty, M., et al. (2018) Multiscale Structuring of the E. coli Chromosome by Nucleoid-Associated and Condensin Proteins. *Cell*, 172 (4): 771-783.e18. doi:10.1016/j.cell.2017.12.027.

133. Lohse, M.B., Zordan, R.E., Cain, C.W., et al. (2010) Distinct class of DNA-binding domains is exemplified by a master regulator of phenotypic switching in Candida albicans. *Proceedings of the National Academy of Sciences of the United States of America*, 107 (32): 14105–14110. doi:10.1073/pnas.1005911107.

134. Lopes, J.G. and Sourjik, V. (2018) Chemotaxis of Escherichia coli to major hormones and polyamines present in human gut. *The ISME Journal*, 12 (11): 2736–2747. doi:10.1038/s41396-018-0227-5.

135. Lueders, T., Kindler, R., Miltner, A., et al. (2006) Identification of Bacterial Micropredators Distinctively Active in a Soil Microbial Food Web. *Applied and Environmental Microbiology*, 72 (8): 5342–5348. doi:10.1128/AEM.00400-06.

136. Luijsterburg, M.S., Noom, M.C., Wuite, G.J.L., et al. (2006) The architectural role of nucleoid-associated proteins in the organization of bacterial chromatin: A molecular perspective. *Journal of Structural Biology*, 156 (2): 262–272. doi:10.1016/j.jsb.2006.05.006.

137. Luijsterburg, M.S., White, M.F., Van Driel, R., et al. (2008) The major architects of chromatin: Architectural proteins in bacteria, archaea and eukaryotes. *Critical Reviews in Biochemistry and Molecular Biology*, 43 (6): 393–418. doi:10.1080/10409230802528488.

138. Luttermann, C. and Meyers, G. (2007) A bipartite sequence motif induces translation reinitiation in feline calicivirus RNA. *Journal of Biological Chemistry*, 282 (10): 7056–7065. doi:10.1074/jbc.M608948200.

139. Makthal, N., Rastegari, S., Sanson, M., et al. (2013) Crystal Structure of Peroxide Stress Regulator from Streptococcus pyogenes Provides Functional Insights into the Mechanism of Oxidative Stress Sensing*. *Journal of Biological Chemistry*, 288 (25): 18311–18324. doi:10.1074/jbc.M113.456590.

140. Malinverni, J.C. and Silhavy, T.J. (2009) An ABC transport system that maintains lipid asymmetry in the Gram-negative outer membrane. *Proceedings of the National Academy of Sciences of the United States of America*, 106 (19): 8009–8014. doi:10.1073/pnas.0903229106.

141. Malone, J.G., Williams, R., Christen, M., et al. (2007) The structure–function relationship of WspR, a Pseudomonas fluorescens response regulator with a GGDEF output domain. *Microbiology*, 153 (4): 980–994. doi:10.1099/mic.0.2006/002824-0.

142. Markelova, N.Y. (2010) Predacious bacteria, Bdellovibrio with potential for biocontrol. *International Journal of Hygiene and Environmental Health*, 213 (6): 428–431. doi:10.1016/j.ijheh.2010.08.004.

143. Martín-Mora, D., Fernández, M., Velando, F., et al. (2018a) Functional annotation of bacterial signal transduction systems: Progress and challenges. *International Journal of Molecular Sciences*, 19 (12): 1–22. doi:10.3390/ijms19123755.

144. Martín-Mora, D., Ortega, Á., Pérez-Maldonado, F.J., et al. (2018b) The activity of the C4-dicarboxylic acid chemoreceptor of Pseudomonas aeruginosa is controlled by chemoattractants and antagonists. *Scientific Reports*, 8 (1): 1–12. doi:10.1038/s41598-018-20283-7.

145. Martin, M.O. (2002) Predatory prokaryotes: An emerging research opportunity. *Journal of Molecular Microbiology and Biotechnology*, 4 (5): 467–477.

146. Matilla, M.A., Martín-Mora, D., Gavira, J.A., et al. (2021) Pseudomonas aeruginosa as a Model To Study Chemosensory Pathway Signaling. Microbiology and Molecular Biology Review, 85 (1). doi:10.1128/MMBR.00151-20

147. Matsuoka, Y. and Shimizu, K. (2013) Catabolite regulation analysis of Escherichia coli for acetate overflow mechanism and co-consumption of multiple sugars based on systems biology approach using computer simulation. *Journal of Biotechnology*, 168 (2): 155–173. doi:10.1016/j.jbiotec.2013.06.023.

148. Matthews, B.W., Ohlendorf, D.H., Anderson, W.F., et al. (1982) Structure of the DNA-binding region of lac repressor inferred from its homology with cro repressor. *Proceedings of the National Academy of Sciences of the United States of America*, 79 (5): 1428–1432. doi:10.1073/pnas.79.5.1428.

149. McCauley, E.P., Haltli, B. and Kerr, R.G. (2015) Description of pseudobacteriovorax antillogorgiicola gen. Nov., sp. nov., a bacterium isolated from the gorgonian octocoral Antillogorgia elisabethae, belonging to the family pseudobacteriovoracaceae fam. nov., within the order Bdellovibrionales. *International Journal of Systematic and Evolutionary Microbiology*, 65 (2): 522–530. doi:10.1099/ijs.0.066266-0.

150. McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., et al. (2007) Phaser crystallographic software. *Journal of Applied Crystallography*, 40 (4): 658–674. doi:10.1107/S0021889807021206.

151. McFarland, K.A., Lucchini, S., Hinton, J.C.D., et al. (2008) The leucine-responsive regulatory protein, Lrp, activates transcription of the fim operon in Salmonella enterica

serovar typhimurium via the fimZ regulatory gene. *Journal of Bacteriology*, 190 (2): 602–612. doi:10.1128/JB.01388-07.

152. McIver, K.S. and Myles, R.L. (2002) Two DNA-binding domains of Mga are required for virulence gene activation in the group A streptococcus. *Molecular Microbiology*, 43 (6): 1591–1601. doi:10.1046/j.1365-2958.2002.02849.x.

153. Medema, M.H., Takano, E. and Breitling, R. (2013) Detecting sequence homology at the gene cluster level with multigeneblast. *Molecular Biology and Evolution*, 30 (5): 1218–1223. doi:10.1093/molbev/mst025.

154. Medina, A.A., Shanks, R.M. and Kadouri, D.E. (2008) Development of a novel system for isolating genes involved in predator-prey interactions using host independent derivatives of Bdellovibrio bacteriovorus 109J. *BMC Microbiology*, 8: 1–11. doi:10.1186/1471-2180-8-33.

155. Meek, R.W., Cadby, I.T., Moynihan, P.J., et al. (2019) Structural basis for activation of a diguanylate cyclase required for bacterial predation in Bdellovibrio. *Nature Communications*, 10 (1). doi:10.1038/s41467-019-12051-6.

156. Michael, A.K., Grand, R.S., Isbel, L., et al. (2020) Mechanisms of OCT4-SOX2 motif readout on nucleosomes. *Science*, 368 (6498): 1460–1465. doi:10.1126/science.abb0074.

157. Miles, A.J, Ramalli, S.G, and Wallace, B.A. (2021) DichroWeb, a website for calculating protein secondary structure from circular dichroism spectroscopic data. Protein Science. doi.org/10.1002/pro.4153

158. Milner, D.S., Till, R., Cadby, I., et al. (2014) Ras GTPase-Like Protein MglA, a Controller of Bacterial Social-Motility in Myxobacteria, Has Evolved to Control Bacterial Predation by Bdellovibrio. *PLoS Genetics*, 10 (4). doi:10.1371/journal.pgen.1004253.

159. Mohr, K.I., Garcia, R.O., Gerth, K., et al. (2012) Sandaracinus amylolyticus gen. nov., sp. nov., a starch-degrading soil myxobacterium, and description of Sandaracinaceae fam. nov. *International Journal of Systematic and Evolutionary Microbiology*, 62 (5): 1191–1198. doi:10.1099/ijs.0.033696-0.

160. Monnappa, A.K., Dwidar, M., Seo, J.K., et al. (2014) Bdellovibrio bacteriovorus inhibits staphylococcus aureus biofilm formation and invasion into human epithelial cells. *Scientific Reports*, 4: 1–8. doi:10.1038/srep03811.

161. Moore, J.O. and Hendrickson, W.A. (2012) An asymmetry-to-symmetry switch in signal transmission by the histidine kinase receptor for TMAO. *Structure*, 20 (4): 729–741. doi:10.1016/j.str.2012.02.021.

162. Mora-Ortiz, M. and Claus, S.P. (2017) Choline Theft—An Inside Job. *Cell Host and Microbe*, 22 (3): 253–255. doi:10.1016/j.chom.2017.08.017.

163. Moret, S., Smela, D., Populin, T., et al. (2005) A survey on free biogenic amine content of fresh and preserved vegetables. *Food Chemistry*, 89 (3): 355–361. doi:10.1016/j.foodchem.2004.02.050.

164. Morgan, A.D., MacLean, R.C., Hillesland, K.L., et al. (2010) Comparative analysis of myxococcus predation on soil bacteria. *Applied and Environmental Microbiology*, 76 (20): 6920–6927. doi:10.1128/AEM.00414-10.

165. Motomura, T., Suga, M., Hienerwadel, R., et al. (2017) Crystal structure and redox properties of a novel cyanobacterial heme protein with a His/Cys heme axial ligation and a Per-Arnt-Sim (PAS)-like domain. *Journal of Biological Chemistry*, 292 (23): 9599–9612. doi:10.1074/jbc.M116.746263.

166. Mu, D.-S., Wang, S., Liang, Q.-Y., et al. (2020) *Bradymonabacteria, a novel bacterial predator group with versatile survival strategies in saline environments*., pp. 1–15. doi:10.21203/rs.2.20535/v1.

167. Muff, T.J. and Ordal, G.W. (2007) The CheC phosphatase regulates chemotactic adaptation through CheD. *Journal of Biological Chemistry*, 282 (47): 34120–34128. doi:10.1074/jbc.M706432200.

168. Muñoz-Dorado, J., Moraleda-Muñoz, A., Marcos-Torres, F.J., et al. (2019) Global transcriptome analysis of the Myxococcus xanthus multicellular developmental program. *bioRxiv*, pp. 1–27. doi:10.1101/564641.

169. Nishiyama, S.I., Takahashi, Y., Yamamoto, K., et al. (2016) Identification of a Vibrio cholerae chemoreceptor that senses taurine and amino acids as attractants. *Scientific Reports*, 6: 1–11. doi:10.1038/srep20866.

170. Notredame, C., Higgins, D.G. and Heringa, J. (2000) T-coffee: a novel method for fast and accurate multiple sequence alignment 1 1Edited by J. Thornton. *Journal of Molecular Biology*, 302 (1): 205–217. doi:10.1006/jmbi.2000.4042.

171. Ortega, Á., Zhulin, I.B. and Krell, T. (2017a) Sensory Repertoire of Bacterial Chemoreceptors. *Microbiology and Molecular Biology Reviews*, 81 (4): 1–28. doi:10.1128/mmbr.00033-17.

172. Ortega, D. R., Fleetwood, A. D., Krell, T., Harwood, C. S., Jensen, G. J., & Zhulin, I. B. (2017b). Assigning chemoreceptors to chemosensory pathways in Pseudomonas aeruginosa. Proceedings of the National Academy of Sciences of the United States of America, 114(48): 12809–12814. doi.org/10.1073/pnas.1708842114

173. Osipiuk, J., Wu, R., Jedrzejczak, R., et al. (2011) *Putative Mga Family Transcriptional Regulator from Enterococcus faecalis.*

174. Ottaviani, D., Pieralisi, S., Angelico, G., et al. (2020) Bdellovibrio bacteriovorus to control Escherichia coli on meat matrices. *International Journal of Food Science and Technology*, 55 (3): 988–994. doi:10.1111/ijfs.14355.

175. Oyedara, O.O., De Luna-Santillana, E. de J., Olguin-Rodriguez, O., et al. (2016) Isolation of Bdellovibrio sp. from soil samples in Mexico and their potential applications in control of pathogens. *MicrobiologyOpen*, 5 (6): 992–1002. doi:10.1002/mbo3.382.

176. Park, M.H. and Igarashi, K. (2013) Polyamines and their metabolites as diagnostic markers of human diseases. *Biomolecules and Therapeutics*, 21 (1): 1–9. doi:10.4062/biomolther.2012.097.

177. Pasternak, Z., Njagi, M., Shani, Y., et al. (2014) In and out: An analysis of epibiotic vs periplasmic bacterial predators. *ISME Journal*, 8 (3): 625–635. doi:10.1038/ismej.2013.164.

178. Pasternak, Z., Pietrokovski, S., Rotem, O., et al. (2013) By their genes ye shall know them: Genomic signatures of predatory bacteria. *ISME Journal*, 7 (4): 756–769. doi:10.1038/ismej.2012.149.

179. Pérez, J., Moraleda-Muñoz, A., Marcos-Torres, F.J., et al. (2016) Bacterial predation: 75 years and counting! *Environmental Microbiology*, 18 (3): 766–779. doi:10.1111/1462-2920.13171.

180. Pokkuluri, P.R., Dwulit-Smith, J., Duke, N.E., et al. (2013) Analysis of periplasmic sensor domains from Anaeromyxobacter dehalogenans 2CP-C: Structure of one sensor domain from a histidine kinase and another from a chemotaxis protein. *MicrobiologyOpen*, 2 (5): 766–777. doi:10.1002/mbo3.112.

181. Potterton, L., Agirre, J., Ballard, C., et al. (2018) CCP 4 i 2: the new graphical user interface to the CCP 4 program suite. *Acta Crystallographica Section D Structural Biology*, 74 (2): 68–84. doi:10.1107/S2059798317016035.

182. Prabaharan, C., Kandavelu, P., Packianathan, C., et al. (2019) Structures of two ArsR As(III)-responsive transcriptional repressors: Implications for the mechanism of derepression. *Journal of Structural Biology*, 207 (2): 209–217. doi:10.1016/j.jsb.2019.05.009.

183. Prehna, G., Ramirez, B.E. and Lovering, A.L. (2014) The lifestyle switch protein Bd0108 of Bdellovibrio bacteriovorus is an intrinsically disordered protein. *PLoS ONE*, 9 (12): 1–24. doi:10.1371/journal.pone.0115390.

184. Pugin, B., Barcik, W., Westermann, P., et al. (2017) A wide diversity of bacteria from the human gut produces and degrades biogenic amines. *Microbial Ecology in Health and Disease*, 28 (1): 1353881. doi:10.1080/16512235.2017.1353881.

185. Quade, N., Mendonca, C., Herbst, K., et al. (2012) Structural Basis for Intrinsic Thermosensing by the Master Virulence Regulator RovA of Yersinia*. *Journal of Biological Chemistry*, 287 (43): 35796–35803. doi:10.1074/jbc.M112.379156.

186. Raghunathan, D., Radford, P.M., Gell, C., et al. (2019) Engulfment, persistence and fate of Bdellovibrio bacteriovorus predators inside human phagocytic cells informs their future therapeutic potential. *Scientific Reports*, 9 (1): 1–16. doi:10.1038/s41598-019-40223-3.

187. Rajagopalan, S., Teter, S.J., Zwart, P.H., et al. (2013) Studies of IscR reveal a unique mechanism for metal-dependent regulation of DNA binding specificity. *Nature Structural and Molecular Biology*, 20 (6): 740–747. doi:10.1038/nsmb.2568.

188. Rawlings, N.D. and Barrett, A.J. (1995) *Evolutionary families of metallopeptidases*. In pp. 183–228. doi:10.1016/0076-6879(95)48015-3.

189. Rendulic, S., Jagtap, P., Rosinus, A., et al. (2004) Supplementary Materials Rendulic 2004. *Science (New York, N.Y.)*, 303 (5658): 689–92. Available at: http://science.sciencemag.org/content/sci/suppl/2004/01/29/303.5658.689.DC1/Rendulic.SOM.pdf.

190. Rettig, M., Germann, M.W., Wang, S., et al. (2013) Molecular Basis for Sequence-Dependent Induced DNA Bending. *ChemBioChem*, 14 (3): 323–331. doi:10.1002/cbic.201200706.

191. Reyes-Darias, J.A., García, V., Rico-Jiménez, M., et al. (2015) Specific gamma-aminobutyrate chemotaxis in pseudomonads with different lifestyle. *Molecular Microbiology*, 97 (3): 488–501. doi:10.1111/mmi.13045.

192. Rhee, S., Martin, R.G., Rosner, J.L., et al. (1998) A novel DNA-binding motif in MarA: The first structure for an AraC family transcriptional activator. *Proceedings of the National Academy of Sciences of the United States of America*, 95 (18): 10413–10418. doi:10.1073/pnas.95.18.10413.

193. Rittenberg, S.C. and Langley, D. (1975) Utilization of nucleoside monophosphates per se for intraperiplasmic growth of Bdellovibrio bacteriovorus. *Journal of Bacteriology*, 121 (3): 1137–1144. doi:10.1128/jb.121.3.1137-1144.1975.

194. Robert, X. and Gouet, P. (2014) Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Research*, 42 (W1): W320–W324. doi:10.1093/nar/gku316.

195. Roderick, S.L. and Matthews, B.W. (1993) Structure of the cobalt-dependent methionine aminopeptidase from Escherichia coli: a new type of proteolytic enzyme. *Biochemistry*, 32 (15): 3907–3912. doi:10.1021/bi00066a009.

196. Rodgers, M.E. and Schleif, R. (2009) Solution structure of the DNA binding domain of AraC protein. *Proteins: Structure, Function and Bioinformatics*, 77 (1): 202–208. doi:10.1002/prot.22431.

197. Rotem, O., Nesper, J., Borovok, I., et al. (2016) An extended cyclic di-GMP network in the predatory bacterium Bdellovibrio bacteriovorus. *Journal of Bacteriology*, 198 (1): 127–137. doi:10.1128/JB.00422-15.

198. Sakamoto, A. and Murata, N. (2001) Update on Genetic Engineering The Use of Bacterial Choline Oxidase , a Glycinebetaine- Synthesizing Enzyme , to Create Stress-Resistant Transgenic Plants 1 STRESS TOLERANCE FOR GENETIC ENGINEERING OF THE. *Society*, 125 (January): 180–188.

199. Sakoh, M., Ito, K. and Akiyama, Y. (2005) Proteolytic activity of HtpX, a membrane-bound and stress-controlled protease from Escherichia coli. *Journal of Biological Chemistry*, 280 (39): 33305–33310. doi:10.1074/jbc.M506180200.

200. Sangwan, N., Lambert, C., Sharma, A., et al. (2015) Arsenic rich Himalayan hot spring metagenomics reveal genetically novel predator-prey genotypes. *Environmental Microbiology Reports*, 7 (6): 812–823. doi:10.1111/1758-2229.12297.

201. Sathyamoorthy, R., Kushmaro, Y., Rotem, O., et al. (2020) To hunt or to rest: prey depletion induces a novel starvation survival strategy in bacterial predators. *ISME Journal*, 15 (1): 109–123. doi:10.1038/s41396-020-00764-2.

202. Saw, J.H.W., Yuryev, A., Kanbe, M., et al. (2012) Complete genome sequencing and analysis of Saprospira grandis str. Lewin, a predatory marine bacterium. *Standards in Genomic Sciences*, 6 (1): 84–93. doi:10.4056/sigs.2445005.

203. Schindelin, J., Arganda-Carreras, I., Frise, E., et al. (2012) Fiji: an open-source platform for biological-image analysis. *Nature Methods*, 9 (7): 676–682. doi:10.1038/nmeth.2019.

204. Seccareccia, I., Kovács, Á.T., Gallegos-Monterrosa, R., et al. (2016) Unraveling the predator-prey relationship of Cupriavidus necator and Bacillus subtilis. *Microbiological Research*, 192: 231–238. doi:10.1016/j.micres.2016.07.007.

205. Seef, S., Herrou, J., Boissier, P. de, et al. (2021) A Tad-like apparatus is required for contact-dependent prey killing in predatory social bacteria. *bioRxiv*. doi:10.1101/2021.02.25.432843.

206. Seidler, R.J. and Starr, M.P. (1968) Structure of the flagellum of Bdellovibrio bacteriovorus. *Journal of bacteriology*, 95 (5): 1952–1955. doi:10.1128/jb.95.5.1952-1955.1968.

207. Shah, P. and Swiatlo, E. (2008) A multifaceted role for polyamines in bacterial pathogens. *Molecular Microbiology*, 68 (1): 4–16. doi:10.1111/j.1365-2958.2008.06126.x.

208. Sharma, G., Khatri, I. and Subramanian, S. (2016) Complete Genome of the Starch-Degrading Myxobacteria Sandaracinus amylolyticus DSM 53668T. *Genome biology and evolution*, 8 (8): 2520–2529. doi:10.1093/gbe/evw151.

209. Sharma, G., Yao, A.I., Smaldone, G.T., et al. (2021) Global gene expression analysis of the Myxococcus xanthus developmental time course. *Genomics*, 113 (1P1): 120–134. doi:10.1016/j.ygeno.2020.11.030.

210. Shatzkes, K., Chae, R., Tang, C., et al. (2015) Examining the safety of respiratory and intravenous inoculation of Bdellovibrio bacteriovorus and Micavibrio aeruginosavorus in a mouse model. *Scientific Reports*, 5 (August): 1–12. doi:10.1038/srep12899.

211. Shatzkes, K., Singleton, E., Tang, C., et al. (2016) Predatory bacteria attenuate Klebsiella pneumoniae burden in rat lungs. *mBio*, 7 (6): 1–9. doi:10.1128/mBio.01847-16.

212. Shiratori, T., Suzuki, S., Kakizawa, Y., et al. (2019) Phagocytosis-like cell engulfment by a planctomycete bacterium. *Nature Communications*, 10 (1): 1–11. doi:10.1038/s41467-019-13499-2.

213. Shrestha, M., Compton, K.K., Mancl, J.M., et al. (2018) Structure of the sensory domain of McpX from Sinorhizobium meliloti, the first known bacterial chemotactic sensor for quaternary ammonium compounds. *Biochemical Journal*, 475 (24): 3949–3962. doi:10.1042/BCJ20180769.

214. Sippl, M.J. and Wiederstein, M. (2008) A note on difficult structure alignment problems, *Bioinformatics*. 24 (3): 426-427, https://doi.org/10.1093/bioinformatics/btm622

215. Sippl, M.J. and Wiederstein, M. (2012) Detection of Spatial Correlations in Protein Structures and Molecular Complexes. *Structure*, 20 (4): 718–728. doi:10.1016/j.str.2012.01.024.

216. Skiebe, E., de Berardinis, V., Morczinek, P., et al. (2012) Surface-associated motility, a common trait of clinical isolates of Acinetobacter baumannii, depends on 1,3-diaminopropane. *International Journal of Medical Microbiology*, 302 (3): 117–128. doi:10.1016/j.ijmm.2012.03.003.

217. Skoko, D., Yoo, D., Bai, H., et al. (2006) Mechanism of Chromosome Compaction and Looping by the Escherichia coli Nucleoid Protein Fis. *Journal of Molecular Biology*, 364 (4): 777–798. doi:10.1016/j.jmb.2006.09.043.

218. Skotnicka, D., Steinchen, W., Szadkowski, D., et al. (2020) CdbA is a DNA-binding protein and c-di-GMP receptor important for nucleoid organization and segregation in Myxococcus xanthus. *Nature Communications*, 11 (1): 1–19. doi:10.1038/s41467-020-15628-8.

219. Slabinski, L., Jaroszewski, L., Rychlewski, L., et al. (2007) XtalPred: a web server for prediction of protein crystallizability. *Bioinformatics*, 23 (24): 3403–3405. doi:10.1093/bioinformatics/btm477.

220. Sockett, R.E. (2009) Predatory lifestyle of Bdellovibrio bacteriovorus. *Annual Review of Microbiology*, 63: 523–539. doi:10.1146/annurev.micro.091208.073346.

221. Soisson, S.M. (1997) Structural Basis for Ligand-Regulated Oligomerization of AraC. *Science*, 276 (5311): 421–425. doi:10.1126/science.276.5311.421.

222. Spek, A.L. (2009) Structure validation in chemical crystallography. *Acta Crystallographica Section D Biological Crystallography*, 65 (2): 148–155. doi:10.1107/S090744490804362X.

223. Sreerema, N., Venyaminov, S.Y., and Woody, R.W. (1999) Estimation of the number of helical and strand segments in proteins using CD spectroscopy. Protein Sci. 8, 370-380. doi: 10.1110/ps.8.2.370

224. Stannek, L., Egelkamp, R., Gunka, K., et al. (2014) Monitoring intraspecies competition in a bacterial cell population by cocultivation of fluorescently labelled strains. *Journal of Visualized Experiments*, (83): 2–9. doi:10.3791/51196.

225. Starr, M.P. and Baigent, N.L. (1966) Parasitic interaction of Bdellovibrio bacteriovorus with other bacteria. *Journal of bacteriology*, 91 (5): 2006–2017. doi:10.1128/jb.91.5.2006-2017.1966.

226. Stella, S., Cascio, D. and Johnson, R.C. (2010) The shape of the DNA minor groove directs binding by the DNA-bending protein Fis. *Genes and Development*, 24 (8): 814–826. doi:10.1101/gad.1900610.

227. Straley, S.C. and Conti, S.F. (1974) Chemotaxis in Bdellovibrio bacteriovorus. *Journal of Bacteriology*, 120 (1): 549–551. doi:10.1128/jb.120.1.549-551.1974.

228. Straley, S.C. and Conti, S.F. (1977) Chemotaxis by Bdellovibrio bacteriovorus toward prey. *Journal of Bacteriology*, 132 (2): 628–640. doi:10.1128/jb.132.2.628-640.1977.

229. Straley, S.C., LaMarre, A.G., Lawrence, L.J., et al. (1979) Chemotaxis of Bdellovibrio bacteriovorus toward pure compounds. *Journal of Bacteriology*, 140 (2): 634–642. doi:10.1128/jb.140.2.634-642.1979.

230. Sung, C.H., Kim, D., Hwang, H.Y., et al. (2008) The crystal structure of the second Z-DNA binding domain of human DAI (ZBP1) in complex with Z-DNA reveals an unusual binding mode to Z-DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 105 (52): 20671–20676. doi:10.1073/pnas.0810463106.

231. Tamayo, R., Pratt, J.T. and Camilli, A. (2007) Roles of Cyclic Diguanylate in the Regulation of Bacterial Pathogenesis. *Annual Review of Microbiology*, 61 (1): 131–148. doi:10.1146/annurev.micro.61.080706.093426.

232. Tang, B.L., Yang, J., Chen, X.L., et al. (2020) A predator-prey interaction between a marine Pseudoalteromonas sp. and Gram-positive bacteria. *Nature Communications*, 11 (1). doi:10.1038/s41467-019-14133-x.

233. Terwilliger, T.C., Adams, P.D., Read, R.J., et al. (2009) Decision-making in structure solution using Bayesian estimates of map quality: the PHENIX AutoSol wizard. *Acta Crystallographica Section D Biological Crystallography*, 65 (6): 582–601. doi:10.1107/S0907444909012098.

234. Terwilliger, T.C., Grosse-Kunstleve, R.W., Afonine, P. V., et al. (2008) Iterative model building, structure refinement and density modification with the PHENIX AutoBuild wizard. *Acta Crystallographica Section D Biological Crystallography*, 64 (1): 61–69. doi:10.1107/S090744490705024X.

235. Thaw, P., Sedelnikova, S.E., Muranova, T., et al. (2006) Structural insight into gene transcriptional regulation and effector binding by the Lrp/AsnC family. *Nucleic Acids Research*, 34 (5): 1439–1449. doi:10.1093/nar/gkl009.

236. The UniProt Consortium (2019) UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research*, 47 (D1): D506–D515. doi:10.1093/nar/gky1049.

237. Thiery, S. and Kaimer, C. (2020) The Predation Strategy of Myxococcus xanthus. *Frontiers in Microbiology*, 11 (January). doi:10.3389/fmicb.2020.00002.

238. Tudor, J.J. and Conti, S.F. (1977a) Characterization of bdellocysts of Bdellovibrio sp. *Journal of Bacteriology*, 131 (1): 314–322. doi:10.1128/jb.131.1.314-322.1977.

239. Tudor, J.J. and Conti, S.F. (1977b) Ultrastructural changes during encystment and germination of Bdellovibrio sp. *Journal of Bacteriology*, 131 (1): 323–330. doi:10.1128/jb.131.1.323-330.1977.

240. Ulrich, L.E., Koonin, E. V. and Zhulin, I.B. (2005) One-component systems dominate signal transduction in prokaryotes. *Trends in Microbiology*, 13 (2): 52–56. doi:10.1016/j.tim.2004.12.006.

241. Ulrich, L.E. and Zhulin, I.B. (2005) Four-helix bundle: a ubiquitous sensory module in prokaryotic signal transduction. *Bioinformatics*, 21 (Suppl 3): iii45–iii48. doi:10.1093/bioinformatics/bti1204.

242. Upadhyay, A.A., Fleetwood, A.D., Adebali, O., et al. (2016) Cache Domains That are Homologous to, but Different from PAS Domains Comprise the Largest Superfamily of Extracellular Sensors in Prokaryotes. *PLoS Computational Biology*, 12 (4): 1–21. doi:10.1371/journal.pcbi.1004862.

243. Valdés-Santiago, L. and Ruiz-Herrera, J. (2014) Stress and polyamine metabolism in fungi. *Frontiers in Chemistry*, 1 (JAN): 1–10. doi:10.3389/fchem.2013.00042.

244. Varon, M., Fine, M. and Stein, A. (1983) Effect of polyamines on the intraperiplasmic growth of Bdellovibrio at low cell densities. *Archives of Microbiology*, 136 (2): 158–159. doi:10.1007/BF00404792.

245. Varon, M. and Zeigler, B.P. (1978) Bacterial predator-prey interaction at low prey density. *Applied and Environmental Microbiology*, 36 (1): 11–17. doi:10.1128/aem.36.1.11-17.1978.

246. Dela Vega, A.L. and Delcour, A.H. (1996) Polyamines decrease Escherichia coli outer membrane permeability. *Journal of Bacteriology*, 178 (13): 3715–3721. doi:10.1128/jb.178.13.3715-3721.1996.

247. Wadhams, G., Armitage, J. (2004) Making sense of it all: bacterial chemotaxis. Nat Rev Mol Cell Biol 5, 1024–1037. https://doi.org/10.1038/nrm1524.

248. Walters, D.R. and Cowley, T. (1996) Formation of cadaverine derivatives in Saccharomyces cerevisiae. *FEMS Microbiology Letters*, 145 (2): 255–259. doi:10.1016/S0378-1097(96)00420-X.

249. Walters, S., Bélanger, M., Rodrigues, P.H., et al. (2009) A member of the peptidase M48 superfamily of Porphyromonas gingivalis is associated with virulence in vitro and in vivo. *Journal of Oral Microbiology*, 1 (2009): 1–8. doi:10.3402/jom.v1i0.2021.

250. Wang, S., Mu, D. and Du, Z.J. (2020) Persicimonas caeni gen. nov., sp. nov., the Representative of a Novel Wide-Ranging Predatory Taxon in Bradymonadales. *Frontiers in Microbiology*, 11 (April): 1–13. doi:10.3389/fmicb.2020.00698.

251. Wang, Z., Kadouri, D.E. and Wu, M. (2011) Genomic insights into an obligate epibiotic bacterial predator: Micavibrio aeruginosavorus ARL-13. *BMC Genomics*, 12 (1): 453. doi:10.1186/1471-2164-12-453.

252. Wargo, M.J. (2013) Homeostasis and catabolism of choline and glycine betaine: Lessons from Pseudomonas aeruginosa. *Applied and Environmental Microbiology*, 79 (7): 2112–2120. doi:10.1128/AEM.03565-12.

253. Watanabe, M., Ni, S., Lindenberger, A.L., et al. (2013) Characterization of the Stoichiometry of HMGA1/DNA Complexes. *The Open Biochemistry Journal*, 7: 73–81. doi:10.2174/1874091x01307010073.

254. Webb, B.A., Karl Compton, K., Castañeda Saldaña, R., et al. (2017) Sinorhizobium meliloti chemotaxis to quaternary ammonium compounds is mediated by the chemoreceptor McpX. *Molecular Microbiology*, 103 (2): 333–346. doi:10.1111/mmi.13561.

255. Webre, D.J., Wolanin, P.M. and Stock, J.B. (2003) Bacterial chemotaxis. *Current Biology*, 13 (2): R47–R49. doi:10.1016/S0960-9822(02)01424-0.

256. Wiechert, J., Filipchyk, A., Hünnefeld, M., et al. (2019) Deciphering the rules underlying xenogeneic silencing and counter-silencing of Lsr2-like proteins. *bioRxiv*, 11 (1): 1–17. doi:10.1101/751156.

257. Wiederstein, M., Gruber, M., Frank, K., et al. (2014) Structure-Based Characterization of Multiprotein Complexes. *Structure*, 22 (7): 1063–1070. doi:10.1016/j.str.2014.05.005.

258. Wiegand, S., Jogler, M. and Jogler, C. (2018) On the maverick Planctomycetes. *FEMS Microbiology Reviews*, 42 (6): 739–760. doi:10.1093/femsre/fuy029.

259. Will, W.R., Navarre, W.W. and Fang, F.C. (2015) Integrated circuits: How transcriptional silencing and counter-silencing facilitate bacterial evolution. *Current Opinion in Microbiology*, 23: 8–13. doi:10.1016/j.mib.2014.10.005.

260. Williams, R.M. and Rimsky, S. (1997) Molecular aspects of the E. coli nucleoid protein, H-NS: A central controller of gene regulatory networks. *FEMS Microbiology Letters*, 156 (2): 175–185. doi:10.1016/S0378-1097(97)00397-2.

261. Willis, A.R., Moore, C., Mazon-Moya, M., et al. (2016) Injections of Predatory Bacteria Work Alongside Host Immune Cells to Treat Shigella Infection in Zebrafish Larvae. *Current Biology*, 26 (24): 3343–3351. doi:10.1016/j.cub.2016.09.067.

262. Wolfe, A.J. (2005) The Acetate Switch. *Microbiology and Molecular Biology Reviews*, 69 (1): 12–50. doi:10.1128/mmbr.69.1.12-50.2005.

263. World Health Organisation (2017) *Global Priority List of Antibiotic-Resistant Bacteria to Guide Research, Discovery , and Development of New Antibiotics*. Available at: https://www.who.int/medicines/publications/global-priority-list-antibiotic-resistant-bacteria/en/.

264. World Health Organisation (2019) *No Time to Wait: Securing the Future from Drug-Resistanct Infections*. Available at: https://www.who.int/antimicrobial-resistance/interagency-coordination-group/final-report/en/.

265. Wu, R., Clancy, S. and Joachimiak, A. (2008) *The Crystal Structure of Heat Shock Protein HtpX Domain from Vibrio parahaemolyticus RIMD 2210633.*

266. Wuichet, K. and Zhulin, I. (2010) Origins and Diversification of a Complex Signal Transduction System in Prokaryotes. Science Signaling, 3 (128). DOI: 10.1126/scisignal.2000724

267. Yang, H., Chang, J., Liu, Z.J., et al. (2006) *Crystal Structure of the Transcriptional Regulatory Protein PF0864: and Asnc Family Member from Pyrococcus Furiosus*.

268. Yang, J., Anishchenko, I., Park, H., et al. (2020) Improved protein structure prediction using predicted interresidue orientations. *Proceedings of the National Academy of Sciences*, 117 (3): 1496–1503. doi:10.1073/pnas.1914677117.

269. Yesudhas, D., Batoo, M., Anwar, M.A., et al. (2017) Proteins recognizing DNA: Structural uniqueness and versatility of DNA-binding domains in stem cell transcription factors. *Genes*, 8 (8). doi:10.3390/genes8080192.

270. Youderian, P. and Hartzell, P.L. (2007) Triple mutants uncover three new genes required for social motility in Myxococcus xanthus. *Genetics*, 177 (1): 557–566. doi:10.1534/genetics.107.076182.

271. Yu, D., Ma, X., Tu, Y., et al. (2015) Both piston-like and rotational motions are present in bacterial chemoreceptor signaling. *Scientific Reports*, 5: 1–8. doi:10.1038/srep08640.

272. Yutin, N., Makarova, K.S., Mekhedov, S.L., et al. (2008) The deep archaeal roots of eukaryotes. *Molecular Biology and Evolution*, 25 (8): 1619–1630. doi:10.1093/molbev/msn108.

273. Zhang, S., Yang, W., Chen, H., et al. (2019) Metabolic engineering for efficient supply of acetyl-CoA from different carbon sources in Escherichia coli. *Microbial Cell Factories*, 18 (1): 130. doi:10.1186/s12934-019-1177-y.

274. Zhang, S., Zhang, T., Yan, M., et al. (2014) Crystal structure of the WOPR-DNA complex and implications for Wor1 function in white-opaque switching of Candida albicans. *Cell research*, 24 (9): 1108–1120. doi:10.1038/cr.2014.102.

275. Zhang, Z. and Hendrickson, W.A. (2010) Structural Characterization of the Predominant Family of Histidine Kinase Sensor Domains. *Journal of Molecular Biology*, 400 (3): 335–353. doi:10.1016/j.jmb.2010.04.049.

276. Zhao, J., Gan, J., Zhang, J., et al. (2017) *Molecular insight into the regulatory mechanism of the quorum-sensing repressor RsaL in Pseudomonas aeruginosa*. doi:10.2210/pdb5J2Y/pdb.

277. Zimmermann, L., Stephens, A., Nam, S.-Z., et al. (2018) A Completely Reimplemented MPI Bioinformatics Toolkit with a New HHpred Server at its Core. *Journal of Molecular Biology*, 430 (15): 2237–2243. doi:10.1016/j.jmb.2017.12.007.

278. Zusman, D.R., Scott, A.E., Yang, Z., et al. (2007) Chemosensory pathways, motility and development in Myxococcus xanthus. *Nature Reviews Microbiology*, 5 (11): 862–872. doi:10.1038/nrmicro1770.

Appendices

Appendix I – Bd2503_Native Crystal Data Log File

For mx14692v36/xCtagJCSGH9HSB511/SAD	Overall	Low	High
High resolution limit	1.41	3.83	1.41
Low resolution limit	44.99	45.01	1.43
Completeness	90.4	99.5	46.1
Multiplicity	5.6	6.1	3.0
I/sigma	12.7	31.7	1.0
Rmerge(I)	0.060	0.047	0.844
Rmerge(I+/-)	0.057	0.045	0.769
Rmeas(I)	0.066	0.051	1.029
Rmeas(I+/-)	0.068	0.054	1.055
Rpim(I)	0.026	0.020	0.572
Rpim(I+/-)	0.037	0.029	0.717
CC half	0.998	0.998	0.494
Wilson B factor	17.752		
Anomalous completeness	87.0	97.9	35.6
Anomalous multiplicity	3.0	3.3	1.7
Anomalous correlation	-0.193	-0.175	-0.142
Anomalous slope	0.665		
dF/F	0.041		
dI/s(dI)	0.547		
Total observations	503353	31370	6713
Total unique	89276	5152	2259

Assuming spacegroup: C 1 2 1

Unit cell: 111.39 35.44 135.73 90.0 106.76 90.0

Appendix II – Bd2503_Cadaverine Crystal Data Log File

	Overall	Low	High
High resolution limit	2.07	5.62	2.07
Low resolution limit	50.41	50.43	2.11
Completeness	84.6	96.2	37.0
Multiplicity	3.2	3.3	2.8
I/sigma	17.6	39.3	1.2
Rmerge(I)	0.036	0.022	0.763
Rmerge(I+/-)	0.028	0.018	0.607
Rmeas(I)	0.044	0.026	0.945
Rmeas(I+/-)	0.040	0.025	0.859
Rpim(I)	0.024	0.014	0.546
Rpim(I+/-)	0.028	0.018	0.607
CC half	0.999	0.999	0.577
Wilson B factor	33.750		
Anomalous completeness	74.7	83.3	27.6
Anomalous multiplicity	1.7	1.8	1.6
Anomalous correlation	0.172	-0.011	0.027
Anomalous slope	1.047		
dF/F	0.053		
dI/s(dI)	0.860		
Total observations	189838	10978	3603
Total unique	58590	3324	1280

Assuming spacegroup: P 1

Unit cell: 62.300 69.270 74.210 91.640 97.750 111.850

Appendix III – Bd2503_Arginine Crystal Data Log File

	Overall	InnerShell	OuterShell
Low resolution limit	66.18	66.18	2.32
High resolution limit	2.26	10.11	2.26
Rmerge (within I+/I-)	0.099	0.056	2.293
Rmerge (all I+ and I-)	0.104	0.057	2.457
Rmeas (within I+/I-)	0.118	0.066	2.757
Rmeas (all I+ & I-)	0.113	0.063	2.680
Rpim (within I+/I-)	0.063	0.035	1.512
Rpim (all I+ & I-)	0.044	0.025	1.056
Rmerge in top intensity bin	0.063	-	-
Total number of observations	368240	4167	26378
Total number unique	54768	654	4048
Mean((I)/sd(I))	7.4	15.0	1.5
Mn(I) half-set correlation CC(1/2)	0.996	0.990	0.469
Completeness	100.0	98.9	100.0
Multiplicity	6.7	6.4	6.5
Mean(Chi^2)	0.81	0.72	0.61
Anomalous completeness	99.8	98.6	99.2
Anomalous multiplicity	3.4	3.5	3.3
DelAnom correlation between half-sets	-0.258	-0.283	0.064
Mid-Slope of Anom Normal Probability	0.652	-	-

Average unit cell: 149.10 68.87 129.23 90.00 117.41 90.00

Space group: C 1 2 1

Appendix IV – Bd2504 Crystal Data Log File

For mx14692v52/xPut43mincryo1/SAD	Overall	Low	High
High resolution limit	2.02	5.48	2.02
Low resolution limit	70.17	70.21	2.05
Completeness	100.0	100.0	99.1
Multiplicity	19.9	19.2	20.1
I/sigma	18.3	57.2	1.5
Rmerge(I)	0.076	0.040	2.117
Rmerge(I+/-)	0.075	0.039	2.067
Rmeas(I)	0.078	0.041	2.171
Rmeas(I+/-)	0.079	0.041	2.177
Rpim(I)	0.018	0.009	0.483
Rpim(I+/-)	0.025	0.013	0.682
CC half	0.996	0.991	0.816
Wilson B factor	46.949		
Anomalous completeness	100.0	100.0	99.1
Anomalous multiplicity	10.1	10.0	10.2
Anomalous correlation	-0.247	-0.268	0.030
Anomalous slope	0.881		
dF/F	0.038		
dI/s(dI)	0.662		
Total observations	913139	45581	45489
Total unique	45978	2369	2258

Assuming spacegroup: P 61

Other likely alternatives are:

P 65

Unit cell: 107.7238 107.7238 106.4671 90.0 90.0 120.0

Appendix V – Bd1081^{S} Crystal Data Log File

For AUTOMATIC/DEFAULT/SAD	Overall	Low	High
High resolution limit	1.55	4.21	1.55
Low resolution limit	101.56	101.77	1.58
Completeness	99.9	100.0	97.3
Multiplicity	10.9	12.9	5.4
I/sigma	12.3	35.3	0.7
Rmerge(I)	0.069	0.039	1.021
Rmerge(I+/-)	0.068	0.038	0.931
Rmeas(I)	0.072	0.040	1.131
Rmeas(I+/-)	0.074	0.042	1.143
Rpim(I)	0.020	0.011	0.469
Rpim(I+/-)	0.029	0.016	0.647
CC half	0.999	0.999	0.797
Wilson B factor	23.948		
Anomalous completeness	99.8	100.0	96.3
Anomalous multiplicity	5.5	6.7	2.7
Anomalous correlation	-0.412	-0.442	-0.003
Anomalous slope	0.450		
dF/F	0.033		
dI/s(dI)	0.362		
Total observations	831323	50660	19987
Total unique	76532	3923	3677

Assuming spacegroup: P 41

Unit cell: 72.88036 72.88036 101.5829 90.0 90.0 90.0

Appendix VI – Bd3092 Crystal Data Log File

For mx14692v52/x5mincryo1/SAD	Overall	Low	High
High resolution limit	2.15	5.85	2.15
Low resolution limit	157.29	157.71	2.19
Completeness	99.7	100.0	93.6
Multiplicity	6.6	6.6	5.9
I/sigma	7.2	13.9	1.7
Rmerge(I)	0.174	0.066	1.590
Rmerge(I+/-)	0.164	0.064	1.476
Rmeas(I)	0.189	0.072	1.749
Rmeas(I+/-)	0.195	0.076	1.814
Rpim(I)	0.073	0.028	0.720
Rpim(I+/-)	0.106	0.041	1.039
CC half	0.993	0.997	0.547
Wilson B factor	25.919		
Anomalous completeness	99.6	100.0	93.4
Anomalous multiplicity	3.4	3.4	3.0
Anomalous correlation	-0.146	-0.287	0.006
Anomalous slope	0.795		
dF/F	0.103		
dI/s(dI)	0.619		
Total observations	363768	18535	15101
Total unique	54863	2822	2554

Assuming spacegroup: P 1 21 1

Unit cell: 55.5279 157.2917 60.989 90.0 102.392 90.0

Appendix VII – Bd1081 Crystal Data Log File

	Overall	InnerShell	OuterShell
Low resolution limit	85.350	85.350	2.679
High resolution limit	2.634	7.147	2.634
Rmerge (all I+ & I-)	0.163	0.033	2.852
Rmerge (within I+/I-)	0.137	0.027	2.260
Rmeas (all I+ & I-)	0.192	0.039	3.326
Rmeas (within I+/I-)	0.193	0.039	3.197
Rpim (all I+ & I-)	0.100	0.020	1.704
Rpim (within I+/I-)	0.137	0.027	2.260
Total number of observations	247631	12432	12937
Total number unique	67976	3405	3434
Mean(I)/sd(I)	3.8	15.9	0.3
Completeness	98.7	99.2	98.6
Multiplicity	3.6	3.7	3.8
CC(1/2)	0.995	0.999	0.352
Anomalous completeness	97.5	95.7	97.8
Anomalous multiplicity	1.8	1.9	1.9
CC(ano)	-0.024	-0.043	0.016
|DANO|/sd(DANO)	0.530	0.480	0.526

Assuming spacegroup: P 1

Unit cell: 83.59 84.33 95.68 101.2 113.4 93.3

Appendix VIII – Bd2625 Crystal Data Log File

	Overall	InnerShell	OuterShell
Low resolution limit	44.70	44.70	2.36
High resolution limit	2.30	10.29	2.30
Rmerge (within I+/I-)	0.061	0.021	1.554
Rmerge (all I+ and I-)	0.066	0.028	1.660
Rmeas (within I+/I-)	0.064	0.022	1.729
Rmeas (all I+ & I-)	0.068	0.029	1.753
Rpim (within I+/I-)	0.020	0.007	0.738
Rpim (all I+ & I-)	0.016	0.007	0.548
Rmerge in top intensity bin	0.029	-	-
Total number of observations	264787	3440	8748
Total number unique	15463	233	925
Mean((I)/sd(I))	21.6	68.1	1.1
Mn(I) half-set correlation CC(1/2)	1.000	1.000	0.554
Completeness	97.5	99.2	83.4
Multiplicity	17.1	14.8	9.5
Mean(Chi^2)	0.94	0.65	0.68
Anomalous completeness	97.1	100.0	81.3
Anomalous multiplicity	9.3	11.0	5.0
DelAnom correlation between half-sets	0.648	0.815	0.015
Mid-Slope of Anom Normal Probability	1.087	-	-

Average unit cell: 52.84 52.84 208.31 90.00 90.00 120.00

Space group: P 31 2 1

Appendix IX – Crystal Data Summary Table

Summary table of data collection and refinement statistics

	Bd2503_ Native	Bd2503_ Cadaverine	Bd2503_ Arginine	Bd2504	Bd1081[S]	Bd3092	Bd1081	Bd2625
Data collection								
Space group	C 1 2 1	P1	C 1 2 1	P 6 1	P 4 1	P 1 2_1 1	P 1	P 3_1 2 1
Cell dimensions								
a, b, c (Å)	111.3, 35.4, 136.1	62.3, 69.3, 74.2	149.1, 68.9, 129.2	107.7, 107.7, 106.5	72.9, 72.9, 101.6	55.5, 157.3, 61.0	83.6, 84.3, 95.7	52.84, 52.84, 208.31
α, β, γ (°)	90, 106.5, 90	91.6, 97.8, 111.9	90, 117.4, 90	90, 90, 120	90, 90, 90	90, 102.4, 90	101.2, 113.4, 93.3	90, 90, 120
Resolution (Å)	2.5	2.1	2.32	2.05	1.55	2.15	2.63	2.30
R_{merge}	0.05 (0.00)	0.028 (0.607)	0.099 (2.293)	0.075 (2.067)	0.068 (0.931)	0.164 (1.476)	0.137 (2.260)	0.061 (1.554)
R_{pim}	0.05 (0.00)	0.028 (0.607)	0.063 (1.512)	0.025 (0.682)	0.029 (0.647)	0.106 (1.039)	0.137 (2.260)	0.020 (0.738)
$I/\sigma I$	7.4 (0.7	17.6 (1.2)	7.4 (1.5)	18.3 (1.5)	12.3 (0.7)	7.2 (1.7)	3.8 (0.3)	21.6 (1.1)

CC(1/2)	0.990 (0.80)	0.999 (0.577)	0.996 (0.469)	0.996 (0.816)	0.999 (0.797)	0.993 (0.547)	0.995 (0.352)	97.5 (83.4)
Completeness (%)	73.0 (6.1)	84.6 (37.0)	100 (100)	100 (99.1)	99.9 (97.3)	99.7 (93.6)	98.7 (98.6)	17.1 (9.5)
Multiplicity	4.4 (1.1)	3.2 (2.8)	6.7 (6.5)	19.9 (20.1)	10.9 (5.4)	6.6 (5.9)	3.6 (3.8)	17.1 (9.5)
Refinement								
Resolution (Å)	1.4	2.07	2.07	2.02	1.55	2.15	2.4	1.90
R_{work}/R_{free}	25.9/28.4	24.4/28.5	28.7/33.4	27.2/31.3	23.6/25.4	24.4/27.4	19.0/25.4	24.3/27.9
Bond lengths (Å)	0.01	0.02	0.01	0.01	0.01	0.01	0.01	0.01
Bond angles (°)	1.01	1.52	1.62	1.10	1.37	1.11	1.03	1.28
Ramachandran Favoured (%)	96.01	93.97	89.35	94.92	96.52	97.05	92.35	98.21
Ramachandran Outliers (%)	0.00	0.75	1.55	0.38	0.75	0.16	1.45	0.00

*Values in parentheses are for highest-resolution shell.

Appendix X – DNA Fragment Sequences Used in EMSAs

bd0089^{U}

GAAACAGTTTCTGGCAGTCTAAGAATTTCGGATTCTTCTTTTGATGAAGTTCTGGAAGTCAAAGTTGC
GATTACGAAACACTAGATAAATTCCAATACTGCGAATCAGAAAAGATGAACTTGAGAGCAAAATGCTC
TCAAGTTTTCGCTTTACTCCCGCCTCAAGGTGCGCGCTAATTGTGCCATGACATCAAGTGCGGCGACG
ACCAAACCTCAGTTAAAAGACTATCTGAATGTCACCTTCTTTCTGAAGGA

bd1172^{U1}

GAAGTTGTCGGCATCTCGCGTTTGGCTTCGTGAGAAAGCTCCTGTGCGCGCAGCTGTTCACGTTTTTT
CGCATTTTTCATCATGCCCAAGATCAGAGCACCGACAATCACCGCGAAAATAAGAACGATAGCCCCCA
CCAAAATTTCCATCTGCTGAGCATGCCCAGGTGACATCATAAATGAACCTCCGCAACAATGTGCGTCA
TTAAGATAAGAAAACACAACCTATGTACCGCATTTTAGAAGCCGAGACCAGAGCATTTTCCCAGCGTT
TCAAAGTCTGAAACACTGTTTACAACATGTCGTTGACCGCCGGAACTGCCCTAAAATAGGTTGGGTCT

bd1172^{U2}

AAATGAACCTCCGCAACAATGTGCGTCATTAAGATAAGAAAACACAACCTATGTACCGCATTTTAGAA
GCCGAGACCAGAGCATTTTCCCAGCGTTTCAAAGTCTGAAACACTGTTTACAACATGTCGTTGACCGC
CGGAACTGCCCTAAAATAGGTTGGGTCT

bd1885^{U1}

CACGAAGCTCATCGCGATAGTGGGACGACACCAAAACTGTGCAATCTTTACCAAAAGCCACCGAAGAC
AAAAGCAAAACCACGGCAACAGCAACAGTCGATTTCAATGAATTCAAAAGGCACCCTTGCGATGTGAA
AGCACCTTCGTATCGCATTTAAAGCGTCCTCAACACGCAACCGAGCTCTCCATGCAATTCTTACTTAA
GTGTCTAGATTCACGACAGGTTCCAAAACCGGAATCTCGTGGAAACTTTTGGAACAAAATGCAGAACT
AGAGTCGCTCCCCTAGTGGACACAGCAAAATTCATATATATAAAGAGCCTCACTAAAGAGGAGCTCTC

bd1885^{U2}

TTAAAGCGTCCTCAACACGCAACCGAGCTCTCCATGCAATTCTTACTTAAGTGTCTAGATTCACGACA
GGTTCCAAAACCGGAATCTCGTGGAAACTTTTGGAACAAAATGCAGAACTAGAGTCGCTCCCCTAGTG
GACACAGCAAAATTCATATATATAAAGAGCCTCACTAAAGAGGAGCTCTC

bd1885NT

ATGACCAAAAAACATCATTACAGAGACTTTATCCTGAAAGAACTGGAACGACGCCAGCGCAAAAATCC
CTCCTACTCCCTTCGCGCTTTCTCTCGTGATCTGGAAGTTCCTTGTTCTCGCCTGTCTGAGATTTTAA
ACGGAAAAGTCGGCTTGTCTGAAAGCCGCGCTGTCAATCTTGCAACAAAACTGAACTTATCCCCTTCC
GAGCGTGAATTCTTTGTGGATCTGGCTTTGTCT

bd2625^{U}

ATTGCAGCGTTCCCTTTGTTTAGGGTACGGAACTCTTCGTCTATTTTAACCTTGCAACTTTGGTATTA
AATGTAAAAAATCCCTCAGGTCG

bd2626^{U}

CCGTTATCTTCCTTTGATGGGATTGGGGGACGACCTTGTCCCGGTATTGAATGACGCTGCAGAAACAA
AAAAAGGTTAACCCGATTGCCACTTTTTTAGGTCTTACCAGAGGGAAACTTGAGCATTCCTATATTCT
TGGTGGAACACTTTCCGGCCTAGGGTTGATGCGACAGGGTTGCTCTTGGGCCGGGGGATCAACAGAAT
GAGAGGACGGAGATCGTAAC

bd3721^{U}

AAGTCGGTGTCCTTTGTGATTGGATTATAGAAGCCCTGAAAATATAGAATGCCCCTGCCGCGATGACA
AAGAAAAACAGGGCATCAAAAGGCAAAGTCGGGGCTGGTTGAGGGCCTCGGGTCTCGGGGTGAGATGA
TCACAAAATCAATTGATCCCGCGATGAGACGCTATTTTATTGATCATAAGTATACAATTTAACTAGAT
AAATACTGAACACGAAAATTAAAATGATCACACTTGTGAACTTTTTGTGGAGAATGGCGATAAGTAAA
GTATCCTGATTTCAGATTGCAGTTCGATACCGATTCAACCCGAAGATTCATCACAGACAGGATGTCGG
ATTCTTTAACAGGGAGGGACT

bd3721^{U}-F1

CCCTGCCGCGATGACAAAGAAAAACAGGGCATCAAAAGGCAAAGTCGGGGCTGGTTGAGGGCCTCGGG
TCTCGGGGTGAGATGATCACAAAATCAATTGATCCCGCGATGAGACGCTATTTTATTGATCATAAGTA
TACAATTTAACTAGATAAATACTGAACACGAAAATTAAAATGATCACACTTGTGAACTTTTTGTGGAG
AATGGCGATAAGTAAAGTATCCTGATTTCAGATTGCAGTTCGATACCGATTCAACCCGAAGATTCATC
ACAGACAGGATGTCGGATTCTTTAACAGGGAGGGACT

bd3721^{U}-F2

CTGGTTGAGGGCCTCGGGTCTCGGGGTGAGATGATCACAAAATCAATTGATCCCGCGATGAGACGCTA
TTTTATTGATCATAAGTATACAATTTAACTAGATAAATACTGAACACGAAAATTAAAATGATCACACT
TGTGAACTTTTTGTGGAGAATGGCGATAAGTAAAGTATCCTGATTTCA

bd3721^{U}-F3

TTGATCCCGCGATGAGACGCTATTTTATTGATCATAAGTATACAATTTAACTAGATAAATACTGAACA
CGAAAATTAAAATGATCACACTTGTGAACTTTTTGTGGAGAATGGCGATAAGTAAAGTATCCTGATTT
CAGATTGCAGTTCGATACCGATTCAACCCGAAGATTCATCACAGACAGGATGTCGGATTCTTTAACAG
GGAGGGACT

bd3721^{U}-F4

ACTAGATAAATACTGAACACGAAAATTAAAATGATCACACTTGTGAACTTTTTGTGGAGAATGGCGAT
AAGTAAAGTATCCTGATTTCAGATTGCAGTTCGATACCGATTCAACCCGAAGATTCATCACAGACAGG
ATGTCGGATTCTTTAACAGGGAGGGACT

bd3721^{U}-F5

TTTTGTGGAGAATGGCGATAAGTAAAGTATCCTGATTTCAGATTGCAGTTCGATACCGATTCAACCCG
AAGATTCATCACAGACAGGATGTCGGATTCTTTAACAGGGAGGGACT

bd3721^{U}-F6

TCGATACCGATTCAACCCGAAGATTCATCACAGACAGGATGTCGGATTCTTTAACAGGGAGGGACT

bd3721^{U}-R1

AAGTCGGTGTCCTTTGTGATTGGATTATAGAAGCCCTGAAAATATAGAATGCCCCTGCCGCGATGACA
AAGAAAAACAGGGCATCAAAAGGCAAAGTCGGGGCTGGTTGAGGGCCTCGGGTCTCGGGGTGAGATGA
TCACAAAATCAATTGATCCCGCGATGAGACGCTATTTTATTGATCATAAGTATACAATTTAACTAGAT
AAATACTGAACACGAAAATTAAAATGATCACACTTGTGAACTTTTTGTGGAGAATGGCGATAAGTAAA
GTATCCTGATTTCAGATTGCAGTTCGATACCGATTCAACCC

bd3721^{U}-R2

AAGTCGGTGTCCTTTGTGATTGGATTATAGAAGCCCTGAAAATATAGAATGCCCCTGCCGCGATGACA
AAGAAAAACAGGGCATCAAAAGGCAAAGTCGGGGCTGGTTGAGGGCCTCGGGTCTCGGGGTGAGATGA
TCACAAAATCAATTGATCCCGCGATGAGACGCTATTTTATTGATCATAAGTATACAATTTAACTAGAT
AAATACTGAACACGAAAATTAAAATGATCACACTTGTGAACTTTTTGTGGAGAATGGCGA

bd3721^{U}-R3

AAGTCGGTGTCCTTTGTGATTGGATTATAGAAGCCCTGAAAATATAGAATGCCCCTGCCGCGATGACA
AAGAAAAACAGGGCATCAAAAGGCAAAGTCGGGGCTGGTTGAGGGCCTCGGGTCTCGGGGTGAGATGA
TCACAAAATCAATTGATCCCGCGATGAGACGCTATTTTATTGATCATAAGTATACAATTTAACTAGAT
AAATACTGAACACG

bd3721^{U}-R4

AAGTCGGTGTCCTTTGTGATTGGATTATAGAAGCCCTGAAAATATAGAATGCCCCTGCCGCGATGACA
AAGAAAAACAGGGCATCAAAAGGCAAAGTCGGGGCTGGTTGAGGGCCTCGGGTCTCGGGGTGAGATGA
TCACAAAATCAATTGATCCCGCGATGAGA

bd3721^{U}-R5

AAGTCGGTGTCCTTTGTGATTGGATTATAGAAGCCCTGAAAATATAGAATGCCCCTGCCGCGATGACA
AAGAAAAACAGGGCATCAAAAGGCAAAGTCGGGGCTGGTTGAGGGCCT

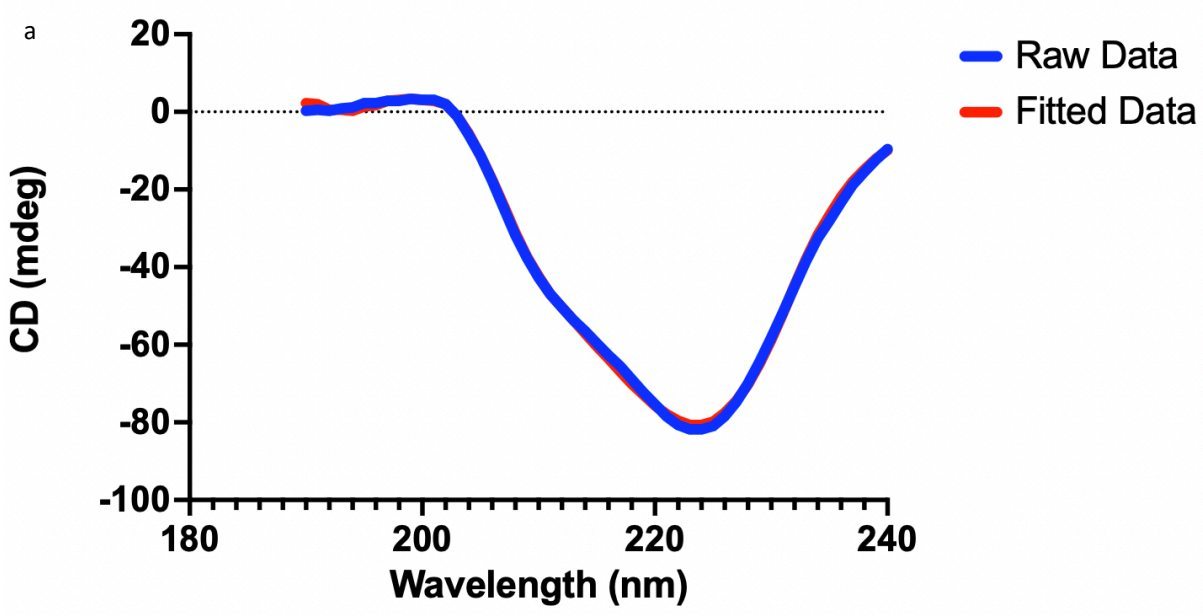
bd3721^{U}-R6

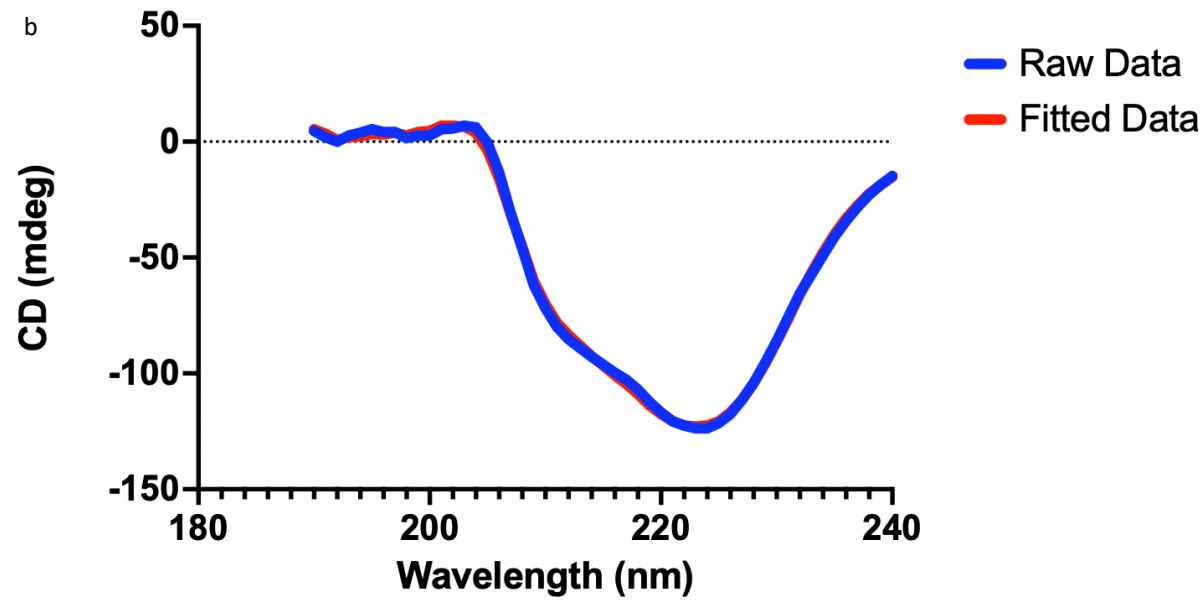
AAGTCGGTGTCCTTTGTGATTGGATTATAGAAGCCCTGAAAATATAGAATGCCCCTGCCGCGAT

gltF

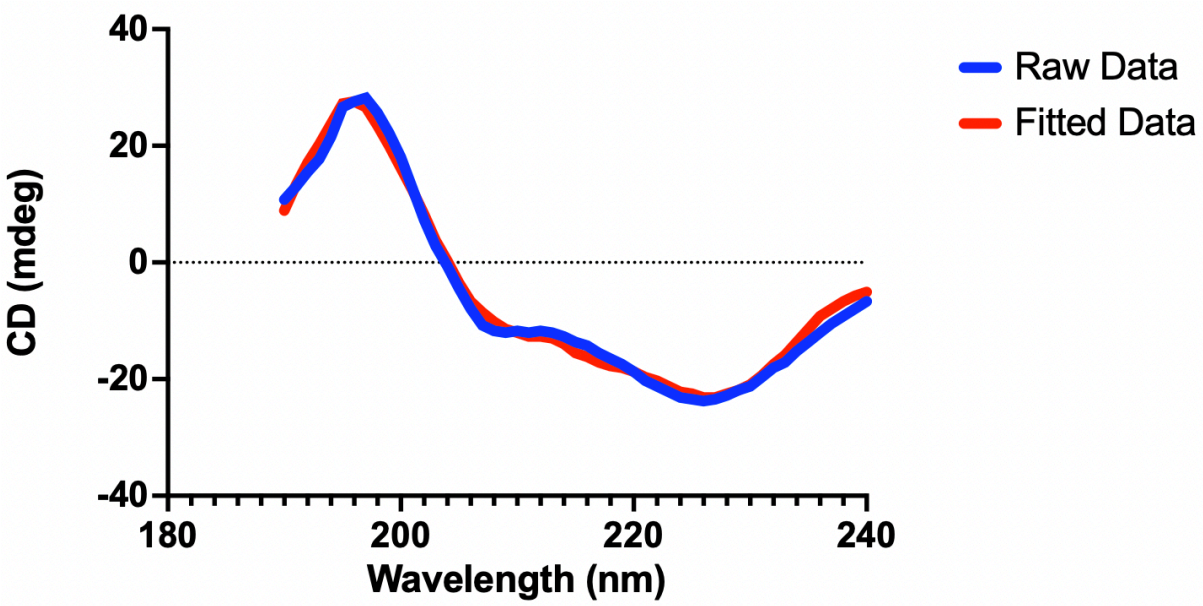
ATGAAATACGTATTGATGATGCTTGTGATGGCAGGCTCCGTCCAGGCGGCTCCTGAAAAAGCGGCATC
TGATAAAATGGCACCAAAAAAGAATCAGGACGTGAACTTCGAGGACCTTCTGGTCAGCGGTCAGTATC
ACTTTTCTGATGAGGCCGTGACCACGGTGGAAGAAGACAAGGTTCTGAACAGCCTTATCGGCGTCAGA
GCAGACTTCAAGGATCGCATGAAACAATCTGCGACCCTTGGTGAGGTGAAAAATCCATGA

Appendix XI – CD Fitted Plot





c



d

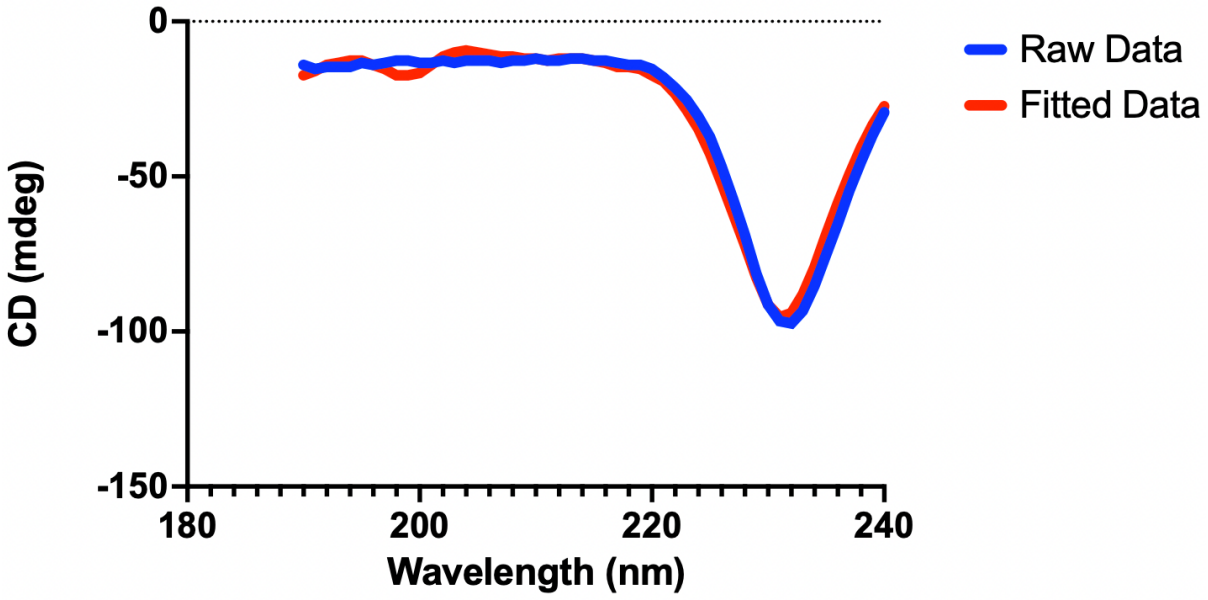


Figure XI CD data fit plots for **(a)** Bd3720 WT **(b)** Bd3720 HTH **(c)** Bd3720 WHTH **(d)** Bd3720 Double

Appendix XII – Collaboration Work