Structural characterisation of *Bdellovibrio bacteriovorus* predatory peptidoglycan modifying proteins



UNIVERSITY OF BIRMINGHAM

Mauricio Sebastián Valdivia Delgado

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

DOCTOR OF PHILOSOPHY

Institute of Microbiology and Infection

School of Biosciences

College of Life and Environmental Sciences

University of Birmingham

May 2022

UNIVERSITY OF
BIRMINGHAM

University of Birmingham Research Archive

e-theses repository

This unpublished thesis/dissertation is copyright of the author and/or third parties.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

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

ABSTRACT

Bdellovibrio bacteriovorus is an endobiotic bacterial predator that encodes multiple enzymes related to known peptidoglycan (PG) modifiers, which are necessary for predation to occur. Additionally, a plethora of cryptic proteins are expressed and upregulated during predation.

It has been shown that L,D-transpeptidases (LDTs) are required for bdelloplast formation and stabilisation. These LDTs cluster in six groups in *B. bacteriovorus* HD100, nevertheless, no further LDT information has been previously described.

Here we have determined and analysed the structures of LDTs Bd0566, Bd1075, Bd3334 and Bd3741, showing that even though they share a structurally conserved catalytic domain, they have acquired accessory domains and motifs characteristic to each LDT group. Moreover, we have integrated these novel structures with the previously solved structures of Bd0553, Bd0886 and Bd1402, generating a panel of seven LDTs representing the six groups of *B. bacteriovorus* HD100, allowing us to propose possible roles and the structural determinants defining the transferase/crosslinking or hydrolytic activities of each group.

Further, we report the structure of two predatory NlpC/P60 amidohydrolases/peptidases, Bd0601 and Bd1177. We have determined that Bd0601 is a DUF1460 amidohydrolase, wherein the PG stem peptide binding groove is blocked by a tyrosine residue, suggesting the need of a conformational rearrangement for substrate binding. On the other hand, we have determined that Bd1177 is a bilobular NlpC/P60, wherein the active site is located in the interface between the catalytic N-terminal domain and a novel C-terminal regulatory domain, suggesting that Bd1177 fluctuates to an open conformation upon substrate binding.

Finally, we have determined the crystallographic structure of the cryptic secreted protein Bd0675, depicting a novel fold. The solved structure depicts a prominent central acidic and aromatic groove delimited by a conserved disulphide-stabilised loop, suggesting that this protein has ligand binding functions. Based on the solved structure we have tested different candidate ligands; however further experiments are required to propose a function for Bd0675. In addition, due to its novelty, Bd0675 was used as target for multiple structure predicting algorithms in CASP14.

ACKNOWLEDGEMENTS

Such a long ride is full of challenges and difficult times. It's also full of beautifully unique moments and people that are present in-between lines in this work and would like to thank.

First of all, I want to thank my supervisor Andy, who has been supporting and mentoring me when things were going smoothly and specially when getting to this stage seemed impossible. Every meeting and chat remind me of why we commit to this career path in the beginning, just for the love of learning something new at every step.

To the scientific framework supporting this thesis. To the funding grant Becas Chile 72180329 (ANID). To Diamond Light Source, proposals MX19880 and MX26803, allowing us to collect all crystallographic data presented herein. To Ian for the structural data on three of the LDTs. I'm also thankful for the collaboration with Socket lab in Nottingham, particularly Emma, and the Bd1075 related publication, and to Dr Stéphane Mesnage and the possibility to collaborate with his group.

I'd also like to thank the 7th floor, my second home away from home, I can't imagine a better place to have spent these four years. First to the Lovering lab, Amber, Hannah, Liam, Matt, Rebecca, and Simon for being there during endless screening hours and beamtime slumber parties, and the occasional after-hours pint and transcendental talks. To the other members of the 7th floor, Pooja, Steph, Khalid, Pete, Ben, Giedre, Rachel, Dave, Lauren, Bethan, and Gaby, who made days better no matter what.

To my family, my parents Norma and Ricardo, and my brother Diego for being my never-ending source of support and listening to endless rants about my work. I know that being this far away makes things much harder, but at the same time I know I'm not alone wherever I go because you're always there supporting me.

To my dearest Gabi, Coral, Perso, Maquita and Aran. From beginning to end always there, have seen the good, the bad and the ugly about what's presented herein. Half of us have dispersed far away, but no distance is enough to dilute our friendship. The same goes for Cris, who somehow also decided that doing a PhD in another continent was the most natural thing to do. To all of you, and everyone that is not explicitly named here just due to a slip of my memory, saying that I'm thankful is definitely too little.

Finally, to Claudia, the one that's been there since the beginning of this rollercoaster ride, the surprise factor that makes even the most mundane moments unique and worth treasuring, and for believing in our decisions specially when I couldn't do it myself. For all of that, and for what's to come.

TABLE OF CONTENTS

CHAPTER 1

1. INTRODUCTION	**1**
1.1. Predatory relationships in bacteria: obligation and attack strategies	**1**
1.2. *Bdellovibrio bacteriovorus* as a model of endobiotic predation	**2**
1.3. Peptidoglycan: synthesis and modification in Gram negative bacteria	**6**
1.3.1. Synthesis of the PG cell wall precursor Lipid II	6
1.3.2. Lipid II insertion into the PG cell wall: PG synthases	10
1.3.3. PG cell wall modifications: L,D-TPases	13
1.3.4. PG cell wall hydrolytic modifications: transglycosylases, EPases, CPases and amidases	20
1.3.4.1. Transglycosylases	20
1.3.4.2. Cell wall peptidases: EPases, CPases and amidases	20
1.4. PG modifying enzymes in *B. bacteriovor*us and their connection with predation	**22**
1.4.1. DDTs Bd0816 and Bd3459 in the predatory life cycle and self-protection mechanisms	22
1.4.2. Deacetylation and prey exit mechanisms	24
1.4.3. LDTs and their role during predation	25
1.4.4. Predatory NlpC/P60 and CHAP amidohydrolases and peptidases	30
1.5. Cryptic proteins as part of the *B. bacteriovorus* predation arsenal	**32**
1.6. Research aims	**32**

CHAPTER 2

2. MATERIALS AND METHODS	**34**
2.1. General methods	**34**
2.1.1. Molecular cloning	34
2.1.2. Site-directed mutagenesis	37
2.1.3. Recombinant protein expression in native conditions	38
2.1.4. Recombinant protein expression in inclusion bodies	39
2.1.5. Protein purification in native conditions by IMAC Nickel chromatography	39
2.1.6. Protein purification from inclusion bodies by IMAC Nickel chromatography and refolding on-column	41
2.1.7. Protein purification in native conditions by MBP-Trap® chromatography	42
2.1.8. Additional purification of proteins by Size Exclusion Chromatography (SEC)	42
2.1.9. Protein concentration, quantification, and storage	43
2.1.10. Protein crystallisation screening and optimisation	43
2.1.11. X-ray diffraction data collection and analysis	44
2.1.12. Structure similarity search	45
2.1.13. Peptidoglycan docking modelling	45
2.2. LDTs	**46**
2.2.1. Bd0566	46

2.2.1.1. Crystallisation and cryoprotection 46
2.2.1.2. Data collection and structure solving 46
2.2.2. Bd1075 47
2.2.2.1. Crystallisation and cryoprotection 47
2.2.2.2. Data collection and structure solving 47
2.2.3. Bd3334 48
2.2.3.1. Crystallisation and cryoprotection 48
2.2.3.2. Data collection and structure solving 48
2.2.4. Bd3741 48
2.2.4.1. Crystallisation and cryoprotection 48
2.2.4.2. Data collection and structure solving 49
2.2.4.3. Phylogenetic analysis of LDTs 49
2.3. PG amidohydrolases 50
2.3.1. Bd0601 C94A 50
2.3.1.1. Crystallisation and cryoprotection 50
2.3.1.2. Data collection and structure solving 50
2.3.2. Bd1177 and Bd1177 C86A/Y124C/Y508C 50
2.3.2.1. Crystallisation and cryoprotection 50
2.3.2.2. Data collection and structure solving 51
2.4. Bd0675 cryptic protein 52
2.4.1. Crystallisation and cryoprotection 52
2.4.2. Data collection and structure solving 52
2.4.3. Bd0675 electrophoretic mobility shift assay (EMSA) for non-specific ssDNA binding determination 53

CHAPTER 3

3. STRUCTURAL CHARACTERISATION OF *B. bacteriovorus* L,D-TRANSPEPTIDASES 54
3.1. Introduction 54
3.1.1. LDTs with possible transpeptidase or hydrolytic activity are present in *B. bacteriovorus* strain HD100 54
3.1.2. *B. bacteriovorus* LDT group 1: LdtF (DpaA/YafK)-like putative LDTs 55
3.1.3. *B. bacteriovorus* LDT group 3: *B. bacteriovorus*' unique SH3b-LDTs 57
3.1.4. *B. bacteriovorus* LDT group 5: B. *bacteriovorus*' unique LDTs 59
3.2. Results 62
3.2.1. Group 2: Bd3741 62
3.2.1.1. Bd3741 expression, purification, and crystallisation 62
3.2.1.2. Bd3741 data collection and structure solving 63
3.2.1.3. Bd3741 overall structure and domains 65
3.2.1.4. Bd3741 catalytic domain and active site 68
3.2.1.5. Bd3741 cryptic triangular hook insertion 71
3.2.2. Group 4: Bd0566 72
3.2.2.1. Bd0566 purification and crystallisation 72
3.2.2.2. Bd0566 data collection and structure solving 74
3.2.2.3. Bd0566 overall structure and domains 75
3.2.2.4. Bd0566 catalytic domain and active site 77

3.2.3. Unique group 5: Bd3334 85
3.2.3.1. Bd3334 expression and purification 85
3.2.3.2. Bd3334 data collection and structure solving 86
3.2.3.3. Bd3334 overall structure and domains 87
3.2.3.4. Bd3334 active site and substrate binding cleft 89
3.2.4. Group 6: Bd1075 92
3.2.4.1. Bd1075 expression, purification, and crystallisation 92
3.2.4.2. Bd1075 data collection and structure solving 93
3.2.4.3. Bd1075 overall structure 96
3.2.4.4. Bd1075 LDT domain and active site 99
3.2.4.5. Bd1075 NTF2-like domain 100
3.3. Discussion **101**
3.3.1. Bd3741 is a possible LDT with crosslinking/NCDAA-exchanging activity 102
3.3.2. Bd0566 is a YcbB-related LDT with a predation-specialised structure 104
3.3.3. Bd3334 is a possible predatory YkuD2-like LDT with proposed peptidase activity 108
3.3.4. Bd1075 is a L,D-carboxypeptidase that regulates the cell shape of *B. bacteriovorus* 108
3.4. Conclusion **109**

CHAPTER 4

4. COMPARATIVE ANALYSIS OF *B. bacteriovorus* L,D-transpeptidases **110**
4.1. Introduction **110**
4.2. Results **110**
4.2.1. Clustering of *B. bacteriovorus* HD100 LDTs 110
4.2.2. Individual LDT group analysis 114
4.2.2.1. LDT group 1 114
4.2.2.2. LDT group 2 117
4.2.2.3. LDT group 3 124
4.2.2.4. LDT group 4 130
4.2.2.5. LDT group 5 133
4.2.3. LDT intergroup analysis 140
4.2.3.1. YkuD-like domains of L,D-transferases 140
4.2.3.2. YkuD- and YkuD2-like domain of L,D-hydrolases 143
4.3. Discussion **145**
4.3.1. *B. bacteriovorus* HD100 LDTs cluster in six specialised groups 145
4.3.2. *B. bacteriovorus* HD100 LDTs accessory domains as targeting and substrate specificity determinants 147
4.3.3. Active site lids and residue conservation are the features defining either the L,D-transferases or L,D-hydrolases preference of each enzyme 150
4.4. Conclusion **151**

CHAPTER 5

5. STRUCTURAL CHARACTERISATION OF *B. bacteriovorus* PUTATIVE HYDROLASES Bd0601 AND Bd1177	**153**
5.1. Introduction	**153**
5.2. Results	**155**
5.2.1. Bd0601	155
5.2.1.1. Bd0601 purification and crystallisation	155
5.2.1.2. Bd0601 C94A structure solving and refinement	157
5.2.1.3. Bd0601 C94A structure overview	159
5.2.1.4. Bd0601 C94A sugar binding pocket	162
5.2.1.5. Bd0601 C94A stem peptide binding pocket	163
5.2.1.6. Bd0601 C94A docking of PG model substrates	166
5.2.2. Bd1177	168
5.2.2.1. Bd1177 purification and crystallisation	168
5.2.2.2. Bd1177 data collection and refinement	169
5.2.2.3. Bd1177 partially proteolyzed C-terminal structure	171
5.2.2.4. Bd1177 C86A and Bd1177 C86A/Y124C/Y508C design	172
5.2.2.5. Bd1177 C86A and Bd1177 C86A/Y124C/Y508C purification and crystallisation	174
5.2.2.6. Bd1177 C86A and Bd1177 C86A/Y124C/Y508C data collection and refinement	175
5.2.2.7. Bd1177 C86A/Y124C/Y508C overall structure and N-terminal domain	176
5.2.2.8. Structural interaction between N- and C-terminal domains of Bd1177 C86A/Y124C/Y508C	178
5.2.2.9. Active site analysis and possible substrate binding residues of Bd1177 C86A/Y124C/Y508C	184
5.3. Discussion	**185**
5.3.1. Bd0601 C94A	185
5.3.1.1. Bd0601 C94A express only as a catalytically null mutant	185
5.3.1.2. Bd0601 C94A is a DUF1460 amidase with a conserved structure but divergent sequence	186
5.3.1.3. Bd0602 and Bd0603 are membrane-linked proteins that may associate with Bd0601	187
5.3.2. Bd1177 and Bd1177 C86A/Y124C/Y508C	188
5.3.2.1. Bd1177 is expressed but crystallises only as a C-terminal fragment	188
5.3.2.2. Bd1177 C86A/Y124C/Y508C is a bilobular protein with a divergent NlpC/P60 fold and a novel capping domain	189
5.3.2.3. Bd1177 C86A/Y124C/Y508C might undergo a substrate induced conformational change	190
5.4. Conclusion	**192**

CHAPTER 6

6. STRUCTURAL CHARACTERISATION OF *B. bacteriovorus* CRYPTIC SECRETED PROTEIN Bd0675	**194**
6.1. Introduction	**194**
6.2. Results	**195**
6.2.1. Bd0675 purification and crystallisation	195
6.2.2. Bd0675 structure solving and overall fold	197
6.2.3. Bd0675 hydrophobic core, disulphide bonding and coulombic surface potential	200
6.2.4. Bd0675 proposed ligand binding cleft definition and function elucidation	202
6.2.5. Bd0675 structural conservation	209
6.2.6. Bd0675 ligand identification analyses	213
6.2.7. Structural predictions of Bd0675	214
6.3. Discussion	**217**
6.3.1. Bd0675 expression, purification, and crystallisation	217
6.3.2. Bd0675 structure comprises a conserved predicted ligand binding cleft	217
6.3.3. Bd0675 structure diverges significantly from ssDNA and lipid binding proteins	218
6.3.4. Bd0675 structure prediction by CASP14 algorithms	219
6.4. Conclusion	**220**

CHAPTER 7

7. SUMMARY AND FUTURE WORK	**221**
7.1. Summary	**221**
7.1.1. L,D-transpeptidases	221
7.1.2. NlpC/P60 amidohydrolases and peptidases	223
7.1.3. Hypothetical cryptic secreted protein Bd0675	224
7.2. Future work	**225**
7.2.1. L,D-transpeptidases	225
7.2.2. NlpC/P60 amidohydrolases and peptidases	227
7.2.3. Hypothetical cryptic secreted protein Bd0675	228
7.2.4. *B. bacteriovorus* and possible applications of PG modifying enzymes	228

REFERENCES	**230**

APPENDIX I. ELECTRON DENSITY MAPS	**246**
A1. Bd3741 (group 2)	246
A2. Bd0566 (group 4)	246
A3. Bd3334 (group 5, soluble)	247
A4. Bd1075 (group 6)	247
A5. Bd1177 C-terminal domain	248

A6. Bd1177 C86A/Y124C/Y508C	248
A7. Bd0675	249
APPENDIX II. pLDDT PLOTS FOR AlphaFold2 MODELS	**250**
A8. ErfK	250
A9. Bd1763 (group 1)	250
A10. Bd2061 (group 1)	251
A11. Bd3012 (group 2)	251
A12. Bd3176 (group 2)	252
A13. Bd3639 (group 2)	252
A14. Bd1176 (group 3)	253
A15. Bd1181 (group 4)	253
A16. Bd1358 (group 4)	254
A17. Bd2908 (group 4)	254
A18. Bd0782 (group 5, lipobox SP)	255
A19. Bd2940 (group 5, soluble)	255
A20. Bd3093 (group 5, soluble)	256
APPENDIX III. RELATED PUBLICATIONS	**257**

TABLE OF FIGURES

Figure 1: *Bdellovibrio bacteriovorus* life cycle	3
Figure 2: Distribution of upregulated genes in *B. bacteriovorus* during predation	6
Figure 3: Peptidoglycan structure	7
Figure 4: UDP-MurNAc-Stem peptide biosynthesis	8
Figure 5: Synthesis of Lipid II PG precursor	9
Figure 6: PBP classes and TPase active site	11
Figure 7: L,D-TPase crosslinking reaction and domain structure	14
Figure 8: L,D-Transpeptidases (LDTs) catalytic mechanism	16
Figure 9: Structure of *M. tuberculosis* LDTs	18
Figure 10: Structure of cell wall peptidases (CWPs) CHAP and NlpC/P60 domains	21
Figure 11: Structure of Bd3459 and immunity complex between Bd3459 and Bd3460	23
Figure 12: Structure of Bd3279 and the deacetylation-specific lysozyme Bd0314	24
Figure 13: Classification of *B. bacteriovorus* LDTs based on domain diversification	26
Figure 14: *E. coli* YcbB structure and domain organisation	28
Figure 15: *H. pylori* Csd6 structure	29
Figure 16: *bd0601-0603* operon organisation and protein products	30
Figure 17: DUF1460 structure	31
Figure 18: Bd0553 crystallographic structure and electrostatic surface potential	55
Figure 19: Bd0553 active site and substrate binding residues	57
Figure 20: Bd0886 crystallographic structure, dimerization, active site, and surface representation	58
Figure 21: Bd1402 structure, dimerization, active site, and surface depiction	60
Figure 22: Bd3741-MBP purification	62
Figure 23: Bd3741 crystals	63
Figure 24: Heavy atom sites of the Bd3741 PtSAD dataset	65
Figure 25: Bd3741 crystallographic structure, topological organisation, and surface coulombic potential	66
Figure 26: B-factor colouring of Bd3741	67
Figure 27: Bd3741 active site organisation	68
Figure 28: Bd3741 substrate binding residue identification	70
Figure 29: Structural conservation of Bd3741 active site	70
Figure 30: Triangular loop of Bd3741 cryptic domain	72
Figure 31: Bd0566 purification	73
Figure 32: Crystal and diffraction pattern for the two Bd0566 datasets	74
Figure 33: Bd0566 structure and coulombic surface potential	76
Figure 34: Bd0566 catalytic domain and active site	78
Figure 35: Bd0566 substrate binding sites and catalytic domain interactions	79
Figure 36: B-factor colouring of Bd0566	80
Figure 37: Sequence alignment and structural context of Bd0566 and YcbB LDT domain	81
Figure 38: Superimposition of the Meropenem binding site of YcbB with Bd0566 structure	82

Figure 39: Citrate binding in Bd0566 active site	84
Figure 40: Bd3334 purification and gel filtration profile	85
Figure 41: Bd3334 protein crystals	86
Figure 42: Bd3334 crystallographic structure, topology, and surface representation	88
Figure 43: B-factor colouring of Bd3334 cartoon and surface representation	89
Figure 44: Bd3334 active site	90
Figure 45: Active site and possible substrate binding residues of Bd3334	91
Figure 46: Bd1075 purification	92
Figure 47: Bd1075 protein crystals	93
Figure 48: Bd1075 bromide phasing sites	94
Figure 49: Bd1075 structure, topology, electrostatic surface potential and possible ligand binding pockets	97
Figure 50: Structural variation of Bd1075 compared to Csd6 and Pgp2	99
Figure 51: Bd1075 active site and substrate binding residues	100
Figure 52: Bd1075 NTF2-like domain ligand binding pocket	101
Figure 53: Structural comparison between Bd3741, ErfK and DpaA	103
Figure 54: Structural comparison of Bd0566 and YcbB scaffolding domains	105
Figure 55: Structural comparison of Bd0566 and YcbB PGBDs	107
Figure 56: Dendrogram analysis of *B. bacteriovorus* HD100 LDTs	111
Figure 57: LDT distribution in *Bdellovibrio*-related predators	113
Figure 58: LDT group 1 sequence alignment	115
Figure 59: LDT group 1 structural divergence	116
Figure 60: Bd0553 and Bd1763 surface comparison	117
Figure 61: LDT group 2 sequence alignment	118
Figure 62: Structural comparison of group 2 of LDTs	120
Figure 63: Surface analysis of group 2	123
Figure 64: Predicted hydrophobic pocket of Bd3741	124
Figure 65: LDT group 3 sequence alignment	126
Figure 66: Structural comparison of group 3 of LDTs	127
Figure 67: SH3b domain conservation and possible ligand binding pockets	129
Figure 68: LDT group 4 sequence alignment	131
Figure 69: Structural comparison of group 4 of LDTs	133
Figure 70: LDT group 5 sequence alignment	134
Figure 71: Structural comparison between lipobox Bd1402 and soluble Bd3334	136
Figure 72: Bd0782 and Bd1402 structural comparison	137
Figure 73: Bd2940, Bd3093 and Bd3334 structural comparison	139
Figure 74: YkuD-like L,D-transferases active site comparison	142
Figure 75: L,D-hydrolase catalytic domain sequence comparison	144
Figure 76: Structural comparison between L,D-hydrolases	144
Figure 77: Structural comparison between L,D-transferases and L,D- hydrolases	150
Figure 78: Bd0601 C94A purification and gel filtration profile	156
Figure 79: Bd0601 C94A crystals	156
Figure 80: Electron density contour map of apo- and GlcNAc-Bd0601 C94A	159
Figure 81: Bd0601 C94A overall fold and topology diagram	159

Figure 82: Electrostatic surface potential and surface representation of substrate binding pockets of Bd0601 C94A	161
Figure 83: Bd0601 C94A glycan binding pocket	162
Figure 84: Sequence alignment between Bd0601 and available DUF1460 structures	163
Figure 85: Stem peptide binding groove of Bd0601 C94A	164
Figure 86: Bd0601 stem peptide binding groove sequence alignment	166
Figure 87: PG model substrate docking on Bd0601 substrate binding site	167
Figure 88: Bd1177 purification and gel filtration profile	168
Figure 89: Bd1177 crystals	169
Figure 90: Bd1177 platinum phasing sites	170
Figure 91: Bd1177 proteolyzed C-terminal structure	172
Figure 92: Bd1177 predicted structural model	173
Figure 93: Purification and gel filtration profile of Bd1177 C86A and Bd1177 C86A/Y124C/Y508C	174
Figure 94: Bd1177 C86A/Y124C/Y508C crystals	175
Figure 95: Bd1177 C86A/Y124C/Y508C crystal structure, active site, and putative catalytic core	177
Figure 96: Analysis of the interacting interface between the N- and C-terminal domains of Bd1177 C86A/Y124C/Y508C	179
Figure 97: B-factor colouring of Bd1177 C86A/Y124C/Y508C	180
Figure 98: Simulated conformational change of Bd1177	181
Figure 99: Conservation of N- and C-terminal interface interacting residues in Bd1177 homologues	183
Figure 100: Potential Bd1177 substrate binding residues	184
Figure 101: Purification and refolding of Bd0675	195
Figure 102: Crystals of Bd0675	196
Figure 103: Bd0675 structural model and topological organisation	199
Figure 104: Hydrophobic core of Bd0675	200
Figure 105: Electrostatic surface potential of Bd0675	201
Figure 106: Amino acid conservation of Bd0675 across predators	203
Figure 107: Conserved residues in Bd0675 proposed ligand binding cleft	204
Figure 108: Conserved binding pockets of Bd0675	206
Figure 109: Crystallisation ligands interactions with Bd0675	207
Figure 110: Bd0675 platinum interacting site	208
Figure 111: Bd0675 structural and sequence conservation against the top 25 Dali matches of the PDB	210
Figure 112: Structural comparison of Bd0675 with the two main folds found by Dali	211
Figure 113: PBF2 structure and ssDNA binding	212
Figure 114: cRABP2 structure and lipid binding	213
Figure 115: Electrophoretic mobility shift assay (EMSA) of Bd0675	214
Figure 116: Structural prediction of Bd0675	215
Figure 117: Comparison of the proposed ligand binding site residues of Bd0675 structure and predicted models	216
Figure 118: Summary of *B. bacteriovorus* HD100 LDT structures	222

TABLE OF TABLES

Table 1: Upregulated PG related genes in *B. bacteriovorus* predation	26
Table 2: Attribution of protein constructs used in this thesis	34
Table 3: Primer list for RF cloning of the different protein constructs into each expression vector	35
Table 4: Primer list for site-directed mutagenesis	38
Table 5: Expression of the proteins crystallised in this thesis	39
Table 6: Purification conditions of crystallised proteins	41
Table 7: Summary of LDTs expressed and crystallised	61
Table 8: Data collection, phasing, and refinement statistics for Bd3741	64
Table 9: Data collection and refinement statistics for Bd0566	75
Table 10: Data collection and refinement statistics for Bd3334	87
Table 11: Dali results for Bd3334 structural search	90
Table 12: Data collection, phasing, and refinement statistics for Bd1075	95
Table 13: Dali results for Bd1075 structural search	98
Table 14: Data collection and refinement statistics for apo- and GlcNAc-Bd0601 C94A	158
Table 15: Data collection and refinement statistics for PtSAD and native Bd1177 C-terminus	171
Table 16: Data collection and refinement statistics for Bd1177 C86A/Y124C/Y508C	176
Table 17: Data collection, phasing, and refinement statistics for Bd0675	198
Table 18: Dali results for Bd0675 structural comparison	209

LIST OF ABBREVIATIONS

°C	Degrees Celsius
Å	Angstrom
AI	Autoinduction
ASU	Asymmetric unit
BALO	*Bdellovibrio*-like organism
BLAST	Basic local alignment search tool
CASP	Critical assessment of protein structure prediction
CHAP	Cysteine-histidine amidohydrolases/peptidases
CHES	N-Cyclohexyl-2-aminoethanesulfonic acid
CPase	Carboxypeptidase
CV	Column volumes
CWA	Cell wall amidohydrolase
CWG	Cell wall glycosidase
CWH	Cell wall hydrolase
CWP	Cell wall peptidase
DAP	Diaminopimelic acid
DDT	D,D-transpeptidase
dNTP	Deoxyribonucleotide triphosphate
DUF	Domain of unknown function
EDTA	Ethylenediaminetetraacetic acid
EMSA	Electrophoretic mobility shift assay
EPase	Endopeptidase
Fw	Forward primer
Gdn-HCl	Guanidinium hydrochloride
GlcNAc	N-acetylglucosamine
GTase	Glycosyl transferase
HD	Host dependent
HEPES	4-(2-hydroxyethyl)-1-piperazineethanolsulfonic acid
HI	Host independent
IMAC	Immobilised metal affinity chromatography
IPTG	Isopropyl β-D-1-thiogalactopyranoside
kb	Kilobase
kD	Kilo Daltons
LB	Lysogeny broth
LDT	L,D-transpeptidase
LLG	Log likelihood gain
M	Molar concentration
MBP	Maltose-binding protein
MLD	GlcNAc(β1-4)-MurNAc(1,6-anhydro)-L-Ala-γ-D-Glu-mDAP-D-Ala
MPD	2-methyl-2,4-pentanediol
MR	Molecular replacement
MurNAc	N-acetylmuramic acid
Mw	Molecular weight
MWCO	Molecular wight cut off
NCBI	National Center for Biotechnology Information

NCDAA	Noncanonical D-amino acid
NlpC/P60	New lipoprotein C/Protein of 60-kD
NTF2	Nuclear-transport factor 2
OD_{600}	Optical density at 600 nm wavelength
PBP	Penicillin-binding protein
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PG	Peptidoglycan
PGBD	Peptidoglycan binding domain
PIPES	Piperazine-N,N'-bis(2-ethanesulfonic acid)
pLDDT	Percentage of Local Distance Difference Test
RMSD	Root-mean-square deviation of atomic positions
rpm	Revolutions per minute
Rv	Reverse primer
SAD	Single-wavelength anomalous dispersion
SDS-PAGE	Sodium dodecyl sulphate- polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SES	Solvent excluded surface
SH3b	Bacterial Src-homology domain 3
SP	Signal peptide
Ta	Annealing temperature
TB	Terrific broth
TEV	Tobacco-etch virus
TFZ	Translation function Z-score
Tm	Melting temperature
TPase	Transpeptidase
U	Enzymatic units
x *g*	Times gravity acceleration
ΔG°	Standard Gibbs free energy
ε	Extinction coefficient

CHAPTER 1

1. INTRODUCTION

1.1. Predatory relationships in bacteria: obligation and attack strategies

Prey-predator relationship dynamics are crucial components of the thermodynamic flow of ecosystem networks [1]. Predation as an ecological interaction is intuitively related with eukaryotic multicellular organisms, while prokaryotes as bacteria are usually considered as only primary nutrient producers and recyclers by means of their degradative and saprophytic metabolism [2], [3]. However, since prokaryotes have colonised the planet for almost four billion years, emergence of prokaryotic and bacterial predatory relationships during their evolution is not surprising. Predatory *Archaea* and bacteria are proposed to be a selective factor for bacterial cell shape and size for both prey and predator. Furthermore, bacterial predators can be found practically in every environment, including open ocean, soils, plant roots, depositions and even harsher environments as highly contaminated soils and water or extreme weather [3].

Predatory bacteria can be taxonomically divided into nonobligate (or facultative) and obligate predators [3], [4]. Moreover, predators can be further classified according to their predation strategies: i) group or "wolfpack" attack, in which a cluster of predators secrete hydrolytic factors and metabolites to degrade prey, ii) epibiotic predation, wherein predators attach themselves to the cell envelope whilst consuming the prey content from the outside, and iii) endobiotic predation, in which cells penetrate the periplasm or cytoplasm of the prey to metabolise the prey cell content.

The best characterised facultative predators are the saprophytes belonging to the *Myxococcales*, particularly, *Myxococcus xanthus*, which exhibits a "wolfpack" strategy of predation [5]. It has been suggested that the secretion of hydrolytic enzymes or lysing factors by a quorum of predatory bacteria confers *M. xanthus* a wide range of prey options, and even by direct contact with prey [3], [5], [6]. However, in this wolfpack strategy the whole quorum of predators is benefited instead of only the cells producing lytic factors, which is proposed to not be sustainable in prey-limiting conditions and is hypothesised to be maintained only by nonobligate predators [7]. Other less characterised facultative predators include unclassified bacteria from the *Lysobacter* and *Cytophaga* genera, also exhibiting a "wolfpack" predation strategy [8].

On the other hand, obligate predators' cell replication is metabolically linked with predation and thus, is mandatory for cell cycle progression. At present, the *Bdellovibrio* and like organisms (here onwards, BALOs) are the best characterised obligate predators, comprising mainly δ-proteobacteria and some α-proteobacteria such as *Micavibrio* spp [3]. In general, BALOs have a narrower prey scope than facultative predators, predating on other Gram-negative bacteria by means of either an epibiotic or endobiotic attack strategy [9], [10]. BALOs as *Micavibrio* spp., *Vampirovibrio chlorellavorus* and *Bdellovibrio exovorus* deploy an epibiotic predatory strategy [11], [12]. However, most BALOs, including most *Bdellovibrio*, *Bacteriovorax*, *Hallobacteriovorax* and *Peridibacter* species, use an endobiotic attack mechanism [4], [9].

To date, the best characterised endobiotic predator corresponds to *Bdellovibrio bacteriovorus*, which has received attention due to its potential use as a pathogen controlling agent [13], [14], [15].

1.2. Bdellovibrio bacteriovorus as a model of endobiotic predation

Bdellovibrio bacteriovorus, specifically strain HD100, was first isolated in 1962 and sequenced in 2004 [16], [17]. This organism is a ubiquitously found predatory Gram-negative *Oligoflexia* which predates on other Gram-negative bacteria [18].

To succeed as a micro-predator in aquatic and terrestrial niches, *B. bacteriovorus* has developed a complex lifecycle consisting of seven key stages: (A) location of the prey and initiation of the attack phase (AP), (B) attachment to- and entry of prey, (C) invasion of periplasm, (D) bdelloplast formation, (E) filamentous growth inside the bdelloplast, (F) exhaustion of bdelloplast resources and septation, and (G) bdelloplast lysis and release of progeny (**Figure 1**) [18]–[20]. It is worth noting that certain *B. bacteriovorus* strains can switch between a host dependent and a host independent (HI) lifestyle (H).

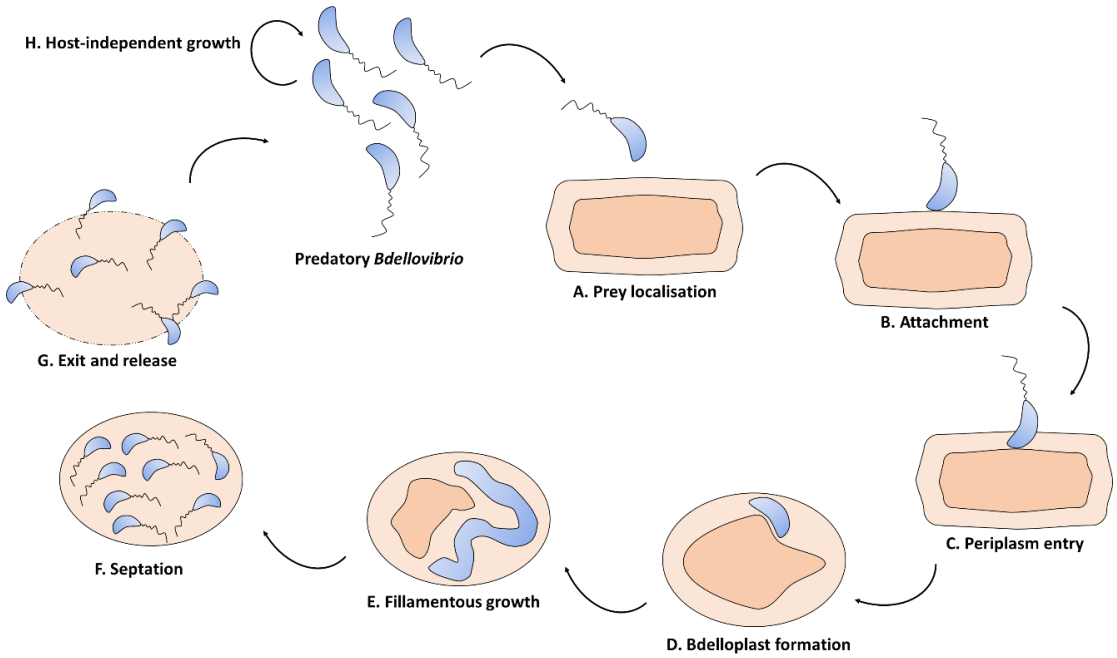


Figure 1: *Bdellovibrio bacteriovorus* life cycle. Summary of the life stages of *B. bacteriovorus*, consisting of (**A**) location of the prey, (**B**) attachment, (**C**) penetration of the outer membrane and periplasm colonisation, (**D**) formation of the bdelloplast, (**E**) filamentous growth inside the bdelloplast by degrading and metabolising prey components, (**F**) septation, (**G**) lysis and exit of progeny from bdelloplast. (**H**) Certain strains can grow separately or simultaneously as a biofilm, independently from a host (HI growth).

B. bacteriovorus differentially expresses genes at each stage with functions including, but not restricted to, chemosensory, motility, hydrolytic and metabolic activities [21]. Nevertheless, the molecular mechanisms underlying each phase of the cycle and switch between host-dependent (HD), and HI lifestyles require further characterisation. For instance, it is known that in strain HD100, loss of function or deletion mutations in the gene *bd0108*, which encodes an intrinsically disordered protein, are related with producing the HI lifestyle phenotype due to its possible regulatory effect on the protrusion/retraction of the invasion type IV pilus [22], [23].

Prey encounter has been proposed to be achieved, at least partially, by chemotactic responses towards amino acids and high concentrations of bacteria, independent to their identity (prey or not prey), as demonstrated by Straley *et al.* (1977 and 1979) [24], [25]. Nevertheless, swimming-compromised mutants have been proven to retain their predation competence, therefore, mechanical interaction between prey and predator does not completely account for prey invasion mechanisms [26].

The commitment of *B. bacteriovorus* to initiate the predatory cycle has been linked to the second messenger cyclic-di-GMP (cdg), which has also been observed to regulate change in lifestyle in non-predatory bacteria, promoting motility, biofilm formation, virulence, among other events [27]–[29]. Specifically, the GGDEF motif-domain containing enzyme DgcB (Diguanylate Cyclase B) has been proposed as a master regulator of predation in *B. bacteriovorus*. Meek *et al.* (2019) have solved the crystallographic structure of DgcB, revealing an asymmetric dimer composed of a GGDEF-motif domain fused to a sensory fork head-associated (FHA) domain, wherein the activity of the enzyme is triggered by phosphorylation of a conserved threonine residue in the disordered N-terminal tail of the FHA domain [30]. This regulatory phosphorylation is proposed to be carried-out by the kinase Bd3148, is fused to DgcB in other species [30].

Regarding the molecular mechanisms by which the predator attaches to- and penetrates the outer membrane of the prey, these are still to be determined. Nonetheless, there is evidence indicating that protein synthesis is required for prey invasion, even in highly viscous environments or by motility-compromised mutant predators, suggesting that adhesion dedicated proteins, among other factors, participate in this process [31], [32].

Moreover, by tracking the incorporation of fluorescently labelled D-amino acid derivatives, it has been observed that *B. bacteriovorus* opens a pore-like structure at the contact point between the prey and predator's outer membranes. The predator is proposed then to squeeze through the pore by using type IV pili, the entry pore being sealed afterwards by the nested predator, initiating the bdelloplast formation stage [33], [34].This stage requires glycanase and peptidase activities, with subsequent modification of the peptidoglycan (PG) cell wall of the prey, which are described in detail below [19], [33].

The bdelloplast is a rounded-up structure formed 30-45 minutes after the beginning of the AP, comprising the infected prey and the *B. bacteriovorus* settled in the periplasmic niche, which now has proceeded to metabolise the contents of said structure to support its own filamentous growth and concomitant division [18], [19]. This structure allows the predator to grow by utilising the prey's metabolites and macromolecules in an efficient manner, avoiding their leakage and consequent competition with other predators, having a role in supporting and protecting the predator during this stage of *B. bacteriovorus*' life cycle.

In fact, for periplasm invasion and colonisation, the predator must not only adapt to physical and chemical changes with respect to the planktonic environment but also the bdelloplast structure

formed upon invasion must be osmotically stable to sustain *Bdellovibrio*'s growth inside the invaded cell. In this regard, it has been determined that formation and stability of the bdelloplast is mainly achieved by modification of the PG wall structure of the invaded prey by several gene products produced during the predator's attack [33], [35], [36].

The first three described modifications of the prey cell wall key for niche formation and maintenance were identified by Thomashow and Rittenberg (1978): i) endopeptidase activities hydrolysing stem peptide crosslinks [37], ii) deacetylase activities acting specifically on prey cell wall, differentiating prey and predator's cell wall [38], and (iii) replacement of the cell wall- OM crosslink mediated by the Braun lipoprotein by the acylation of PG with long chain fatty acids, mainly palmitic acid and oleic acid [39].

Transcriptomic analysis of *B. bacteriovorus* strain HD100 revealed the upregulation of at least 240 genes during predation (30 minutes after initiating the attack phase) respect to the HI strain [21]. Approximately 10% of upregulated protein coding genes have putative PG modifying activities, including predicted 3-4 and 3-3 stem peptide crosslinking enzymes (referred as D,D- and L,D-Transpeptidases, from here onwards DDTs and LDTs, respectively), endo- and carboxypeptidases and amidases, among other hydrolytic activities (see **Figure 2**) [21]. Moreover, it is worth noting that only 35% of the upregulated transcriptome (or predatosome) has a demonstrated or putative function, whereas the role of the remaining 65% of genes are cryptic and require further examination.

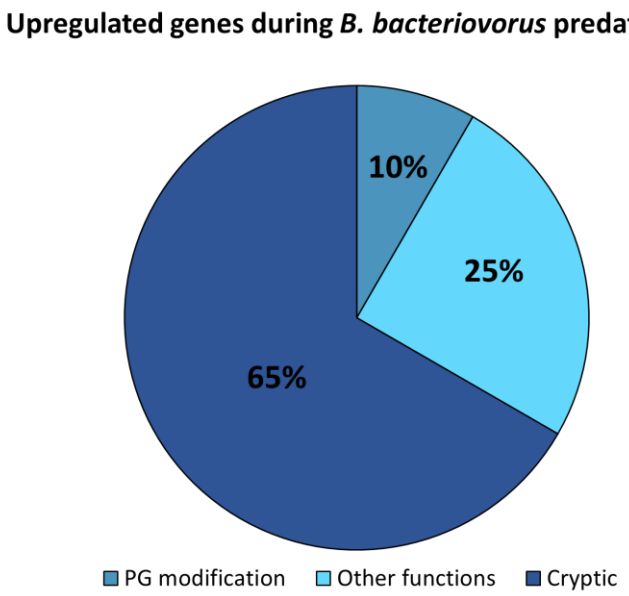


Figure 2: Distribution of upregulated genes in *B. bacteriovorus* during predation. Percentage of upregulated genes during *Bdellovibrio* predation. Around 10% of the total overexpressed genes are related with PG modification, whereas about 65% of the genes encode hypothetical proteins without predicted functions.

Evidently, one of the main targets of *B. bacteriovorus*, in particular during the early stages of predation, corresponds to prey PG cell wall. Thus, obtaining molecular insight about the PG cell wall modification processes is key for understanding *B. bacteriovorus* life cycle at the microscopic level.

1.3. Peptidoglycan: synthesis and modification in Gram negative bacteria

1.3.1. Synthesis of the PG cell wall precursor Lipid II

Peptidoglycan is a complex polymer present in both Gram-positive and Gram-negative bacteria, consisting of repeating disaccharide subunits of N-acetylglucosamine (GlcNAc) and N-acetylmuramic acid (MurNAc), wherein the disaccharide is further modified at the D-lactate group in MurNAc C_3 with a three to five amino acid long oligopeptides constituted by D- and L-amino acids, which crosslink the different PG units forming a dynamic mesh structure in the periplasm [40]–[43] (**Figure 3**).

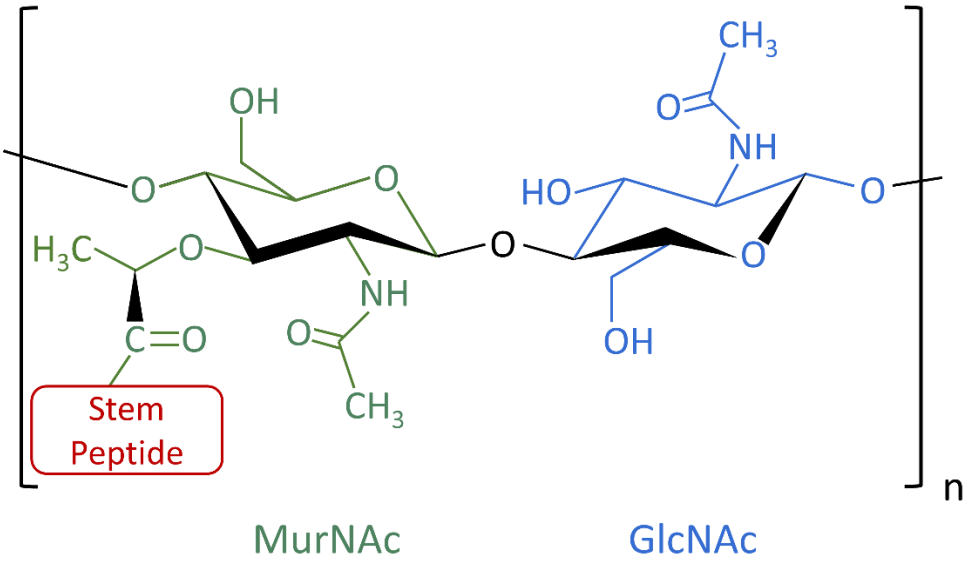


Figure 3: Peptidoglycan structure. PG is a β1-4 glycan polymer formed by "n" alternating units of MurNAc (green) and GlcNAc (blue) sugars, wherein the D-lactate radical at position C_3 of the MurNAc ring is substituted with a three to five amino acid long stem peptide (red).

In Gram-negative bacteria, the PG precursors are synthesised in the cytoplasm by the enzymes of the Mur pathway. Specifically, a UDP activated precursor of N-acetylglucosamine (UDP-GlcNAc) is converted to UDP- N-acetylmuramic acid (UDP-MurNAc) in the cytoplasm by addition and subsequent reduction of a phosphoenolpyruvate (PEP) group at C_3, catalysed by the transferase and dehydrogenase MurA and MurB, respectively [44], [45] (**Figure 4**).The UDP-MurNAc C_3 oxypropanoic (D-lactate) radical is then modified by the addition of an L-Alanine (L-Ala) residue catalysed by the ATP-dependent MurC [46], [47].

The L-Ala modified UDP-MurNAc is then extended by the sequential peptide bond condensation of γ-D-Glutamate (D-Glu), *meso*-diaminopimelate (*meso*-2,6-diaminoheptanedioate, here onwards mDAP), and a D-Ala/D-Ala dimer, catalysed by the MurD-F ATP-dependent amino acid ligases, forming a pentapeptide commonly referred as stem peptide [48]–[50] (**Figure 4**). In Gram positive bacteria, the mDAP at position 3 is usually replaced by a L-lysine residue [51].

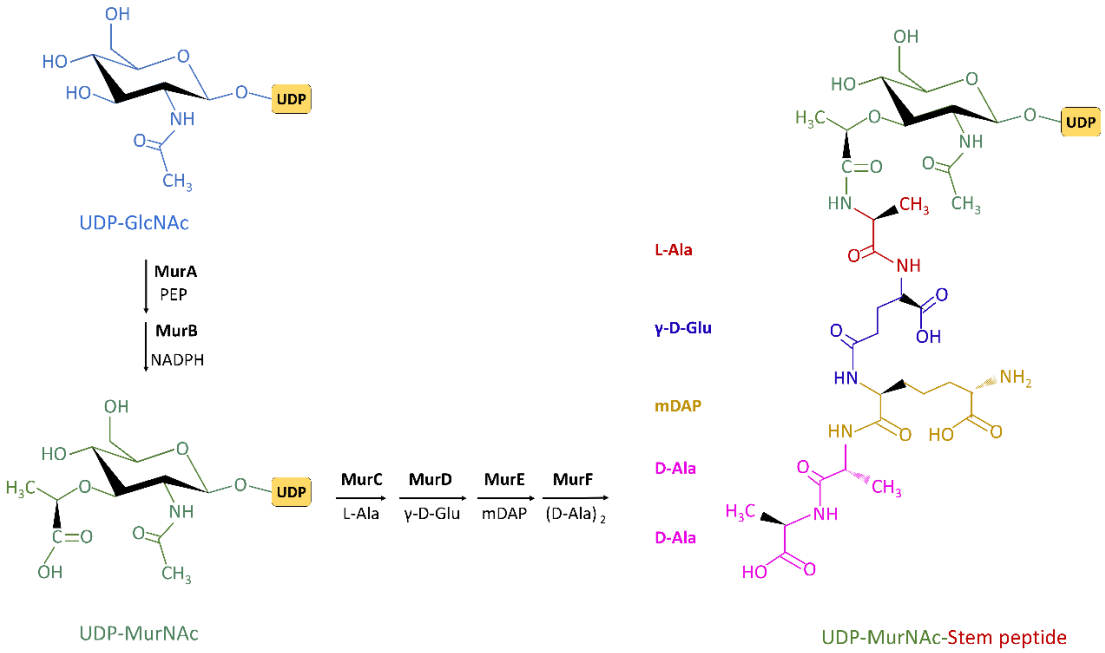


Figure 4: UDP-MurNAc-Stem peptide biosynthesis. The UDP-MurNAc-Stem peptide precursor is synthesised in the cytoplasm by the Mur pathway enzymes. MurA and MurB catalyse the conversion of UDP-GlcNAc into UDP-MurNAc, adding a carboxy radical substituting position C_3 of the MurNAc ring. Finally, the amino acids L-Ala, γ-D-Glu, mDAP γ $(D\text{-}Ala)_2$ are added to the active carboxy group by MurC-F ligases, forming the PG stem peptide precursor UDP-MurNAc-Stem peptide.

In the next step of the pathway, the UDP-MurNAc-pentapeptide precursor is attached to a C_{55} isoprenyl pyrophosphate lipid carrier on the cytoplasmic side of the inner membrane leaflet via a pyrophosphate linkage formed by releasing UMP, the reaction catalysed by the glycosyltransferase MraY, forming a molecule named Lipid I [52]. Later, the glycosyltransferase MurG forms a β-1,4 glycosidic bond between Lipid I and GlcNAc to form a C_{55}-PP activated-GlcNAc-β-1,4-MurNAc-pentapeptide molecule named Lipid II [53] (**Figure 5**). Lipid II is transported across the inner membrane, from the cytoplasm to the periplasm, by the MurJ flippase and added to a nascent PG chain in the periplasm [54], [55].

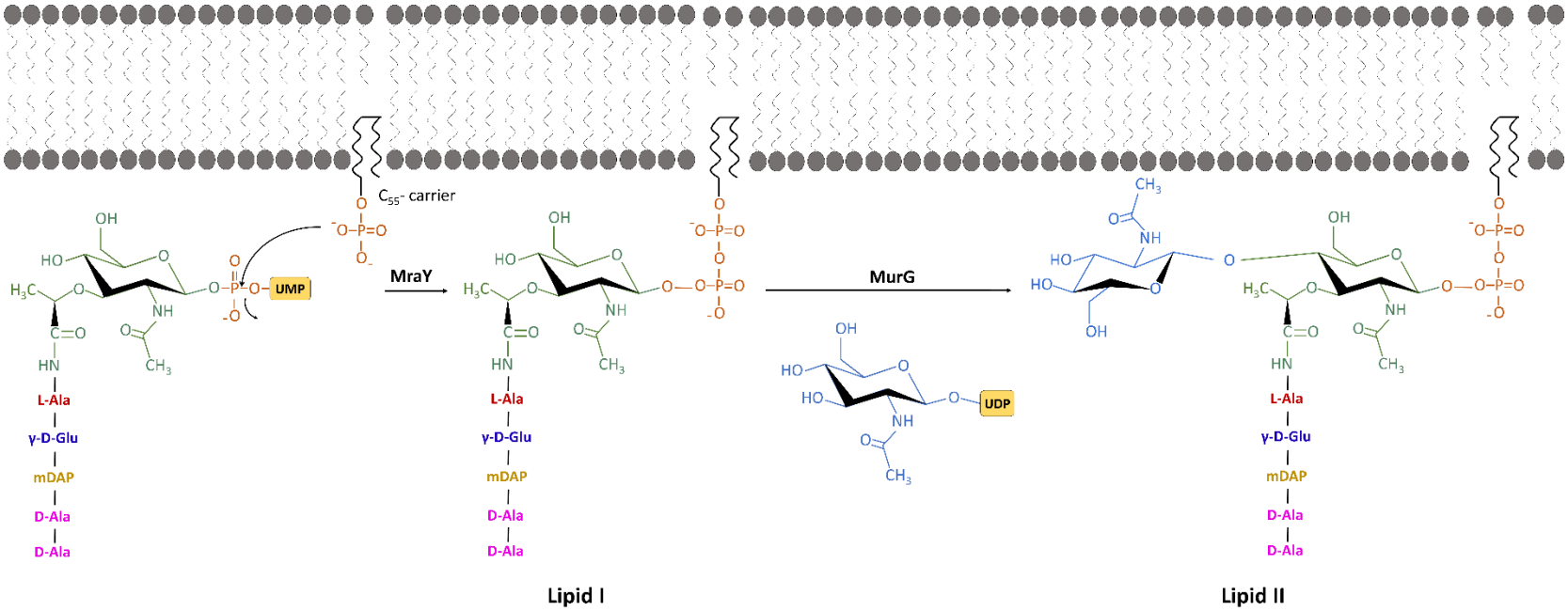


Figure 5: Synthesis of Lipid II PG precursor. The UDP-MurNAc-stem peptide molecule is linked to the cytoplasmic side of the inner membrane by the condensation of a pyrophosphate bond with a C_{55}- undecaprenyl carrier, catalysed by MraY, forming the Lipid I molecule. Finally, MurG links UDP-GlcNAc to Lipid I, producing the Lipid II, which will be flipped across the inner membrane to the periplasm for its incorporation into the PG sacculus.

1.3.2. Lipid II insertion into the PG cell wall: PG synthases

Once in the periplasm, the peptide-disaccharide portion of Lipid II is assimilated into the pre-existent PG sacculus by transglycosylation, forming a new β-1,4-GlcNAc-MurNAc linkage. This step is achieved by the elimination of the pyrophosphate lipid carrier, which is subsequently hydrolysed and recycled to the cytoplasmic side of the inner membrane [40], [56]. The nascent PG chain is incorporated into the cell wall mesh by crosslinking of stem peptides between PG chains using the terminal D-Ala of the accepting stem peptide as leaving group.The *de novo* transglycosylation and crosslinking of Lipid II is catalysed by the PG synthases, commonly referred as Penicillin Binding Proteins (PBPs) for being firstly identified due to their inactivation by β-lactams [41], [56], [57]. PBPs are monotopic membrane glycosyltransferases (GTases, EC:2.4.1.129) and/or TPases (EC:3.4.16.4) and are divided into three categories: i) class A, corresponding to bifunctional GTases-TPases, ii) class B, comprising monofunctional TPases, and iii) monofunctional GTases (**Figure 6A**). *E. coli* has three class A PBPs (PBP1a, PBP1b and PBP1c), two class B PBPs (PBP2 and PBP3, which participate in cell elongation or division, respectively) and one monofunctional GTase (MtgA) [56], [57].

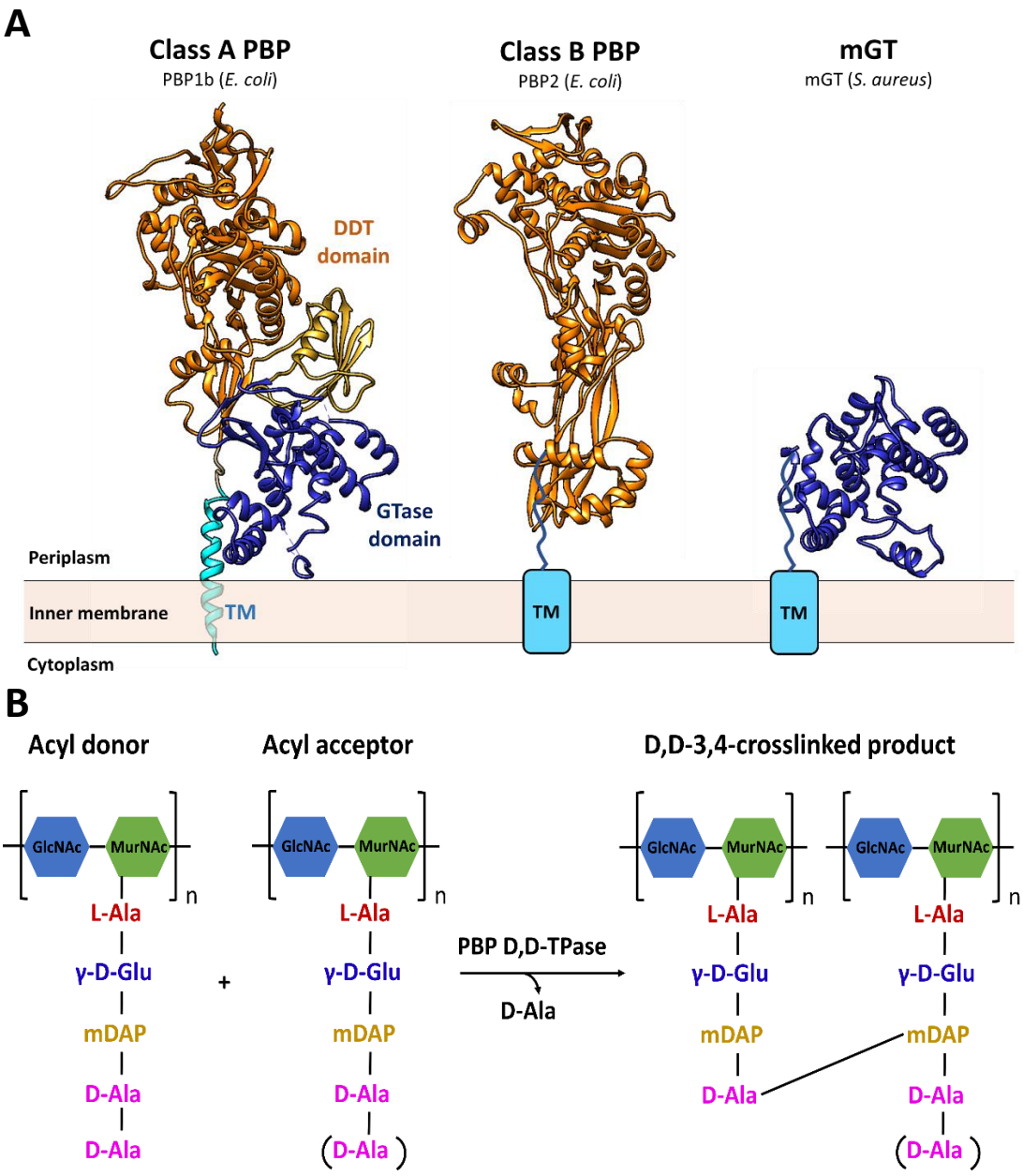


Figure 6: PBP classes and TPase active site. A) Membrane associated PG synthases or PBPs are divided in i) class A, comprising bifunctional GTase-TPase enzymes, ii) class B, consisting of monofunctional TPases and iii) monofunctional GTases (mGT). One representative protein structure per class is depicted (PBP1b of *E. coli*, PDB ID: 3VMA; PBP2 of *E. coli*, PDB ID: 6G9F; and mGT of *S. aureus*, PDB ID: 3HZS). The GTase domain is presented in blue, the TPase domain in orange and transmembrane portion (TM) in cyan. Additionally, PBP1b possess an accessory UB2H domain (in yellow), with activity regulation function. **B)** D,D-TPase reaction. The PBP TPase crosslinks the fourth amino acid (D-Ala) of the acyl donor molecule with the mDAP at third position in the acyl acceptor by the elimination of the terminal D-Ala in the acyl donor stem peptide.

The transglycosylation is the first incorporation reaction to occur, in which Lipid II utilises the C_4 hydroxyl group of the GlcNAc sugar ring as a nucleophile and the lipid pyrophosphate carrier at position C_1 of the MurNAc of another Lipid II as leaving group, elongating a pre-existent PG chain by two GlcNAc-MurNAc units.

Then, for the major biosynthetic PG synthases, PBPs of class A, the elongated PG chain is then displaced to the TPase potion of the PBP for it to be crosslinked to an adjacent PG chain in the periplasm, as proposed based on the PBP1b crystal structure and evidence supporting that only polymerised glycan chains undergo crosslinking [58].

The TPase domain of the PBPs corresponds to a serine transpeptidase that catalyses the crosslinking reaction between the mDAP in the third position of the donor stem peptide and the D-Ala in the fourth position of the acceptor stem peptide using the D-Ala in the fifth position of the acceptor stem peptide as a leaving group [42], [58], [59] (**Figure 6B**). Due the resulting bond possessing a D,D-linkage stereochemistry, PBPs are alternatively denominated DDTs.

The catalytic domain of the DDT is formed by two subdomains, wherein the active site sits in the interface channel between them, defining an acyl donor and acyl acceptor binding site [41], [58]. The first subdomain comprises a five stranded β-sheet bounded by three α-helices arranged in an orthogonal bundle, whereas the second subdomain is mainly composed by α-helices. The active site possesses the signature motif SXXK, wherein the Ser residue acts as the nucleophile group while K is proposed to act as the general base [41], [57], [58].

DDTs follow a ping-pong mechanism in which the catalytic serine is deprotonated by the general base and attacks the carbonyl amide of the D-Ala at position four, displacing the D-Ala at position five by an addition-elimination mechanism. The bound acyl-enzyme intermediate is finally attacked by the ε-amino group of the mDAP of the acceptor peptide, releasing the D,D-crosslinked product [58]. Alternatively, a carboxypeptidation reaction, catalysed by D,D-carboxypeptidases (D,D-CPases) occurs via a water molecule acting as a Lewis base instead of a second PG molecule, hydrolysing the acyl-enzyme intermediate and generating a PG tetrapeptide product [59]. PG tetrapeptides may also be used as substrates for D,D-transpeptidation as depicted in **Figure 6B**.

In *E. coli*, the PG cell wall remains connected to the inner membrane via the multiprotein complexes in charge of PG synthesis during division (the divisome, associated with PBP1b and

PBP3) and elongation (elongasome, associated with PBP1a and PBP2). During cell grow and division, PBPs interact with specific activating proteins such as LpoA (elongasome) and LpoB (divisome), and the Tol-Pal outer membrane constricting complex [40], [56]. It has also been proven that class B PBPs (bPBPs) act in complex with shape, elongation, division, and sporulation (SEDS)-family glycosyltransferase proteins, as shown for the *Thermus thermophilus* RodA-PBP2 SEDS-bPBP complex, which participates in PG synthesis sustaining cell growth [60].

1.3.3. PG cell wall modifications: L,D-TPases

While 3-4 crosslinking and hydrolysis by PBPs is the most studied type of PG modification, 3-3 crosslinks have been reported in different model organisms, formed between the mDAP of a tetrapeptide PG chain and the mDAP of a tripeptide PG chain, using the D-Ala at position four of the donor as leaving group (**Figure 7A**) [61], [62], [63].

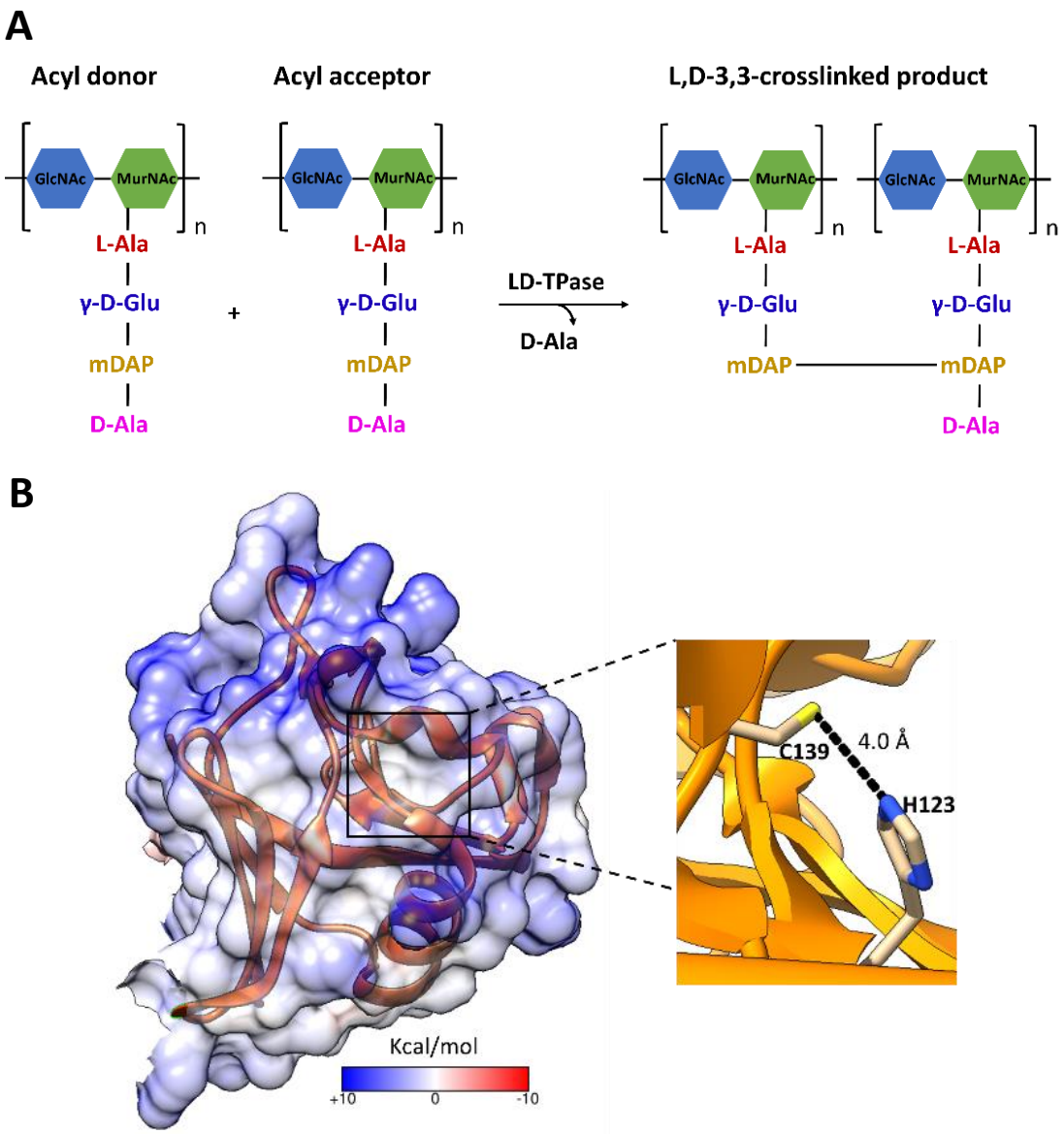


Figure 7: L,D-TPase crosslinking reaction and domain structure. **A)** 3-3 crosslinks catalysed by LDTs. The mDAP of the acyl acceptor is transferred to the mDAP of the acyl donor using the terminal D-Ala of the donor as leaving group. **B)** The YkuD LDT from *B. subtilis* (PDB ID: 1Y7M) is depicted in ribbon and surface representations, wherein the catalytic dyad C139-H123 has a 4.0 Å distance between the thiol of C139 and the $N^{\varepsilon 2}$ of H123. Coulombic surface potential scale from basic (+10 Kcal/mol) to acidic (-10 Kcal/mol) is indicated.

The enzymes catalysing the condensation of 3-3 crosslinks were firstly characterised in *Enterococcus faecium* as part of a β-lactam resistance machinery and defined as LDTs [64]. Since this group of enzymes are insensitive to non-carbapenem β-lactams, they have been more thoroughly studied in organisms of clinical relevance, such as multidrug resistant bacteria strains, i.e. *Mycobacterium tuberculosis*, *Klebsiella pneumoniae*, *Bacillus subtilis* and *Helicobacter pylori*, among others [62], [65]–[68].

At present, the solved structures for the LDTs share a common catalytic core denominated YkuD-like LDT domain. The LDT domain is composed by a basic eight strand β-barrel capped by an α-helix (**Figure 7B**) [69]. The active site of the LDT comprises a conserved cysteine acting as a nucleophile and a histidine as a general base, that may be further polarised by a third residue, usually an aspartate group. The ping-pong mechanism followed by LDTs is similar to the one of the DDTs as shown in **Figure 8**.

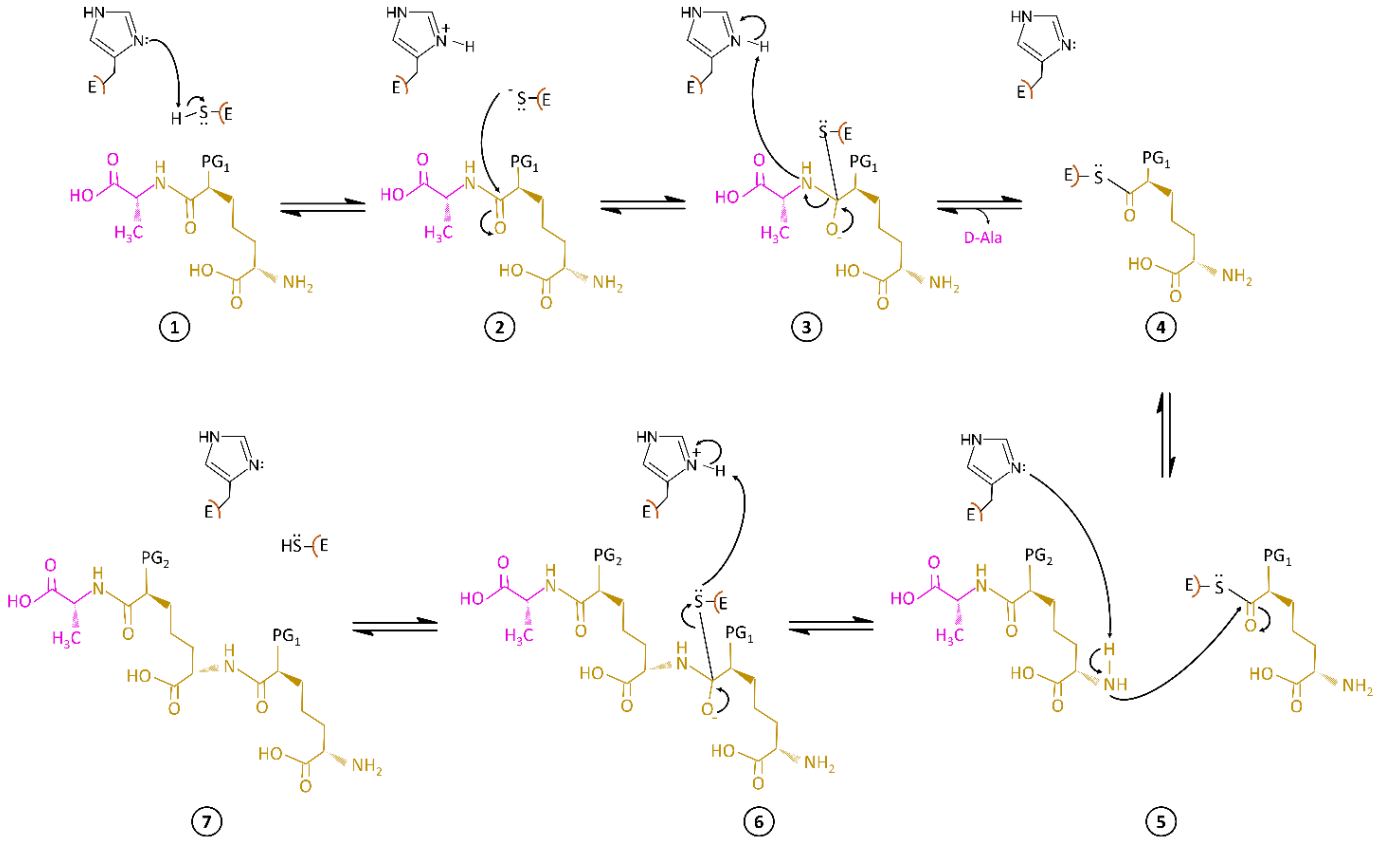


Figure 8: L,D-Transpeptidases (LDTs) catalytic mechanism. LDTs catalyse the crosslinking of two PG stem tetrapeptides (D-Ala in magenta and mDAP in yellow) by a ping-pong mechanism. The catalytic cysteine is deprotonated by a conserved histidine to be a competent nucleophile (**step 1**). The sulfhydryl anion attacks the carbonyl of the amide bond between D-Ala and mDAP (**step 2**), generating tetrahedral intermediate (**step 3**) which is converted into an acyl-enzyme intermediate by the elimination of the terminal D-Ala (**step 4**). A second PG tetrapeptide molecule enters the active site at **step 5**, wherein the ε-amine of the mDAP residue is activated by the histidine and attacks the acyl-enzyme bond, forming a second tetrahedral intermediate (**step 6**). Finally, the tetrahedral intermediate uses the enzyme as leaving group and the acid-base state of the cysteine and histidine are recovered and the 3-3 crosslinked PG molecule released (**step 7**). It is worth noting that for L,D-CPases at **step 5**, a water molecule could be used as a Lewis base instead of a second PG molecule, generating a PG tripeptide as product of the reaction.

As with the PBPs, the first PG tetrapeptide enters the acyl donor site, wherein the catalytic cysteine is deprotonated by a conserved histidine acting as a general base (**step 1**). The nucleophilic cysteine anion then attacks the peptidyl carbonyl group of the mDAP residue of the tetrapeptide (**step 2**), forming a tetrahedral oxyanion intermediate (**step 3**). This first intermediate is resolved by eliminating the D-Ala of the tetrapeptide, recovering the deprotonated state of the active site histidine (**step 3**), forming an acyl-enzyme intermediate (**step 4**). Next, the second PG tetrapeptide enters the active site, wherein the catalytic histidine deprotonates the ε-amino of the mDAP of the tetrapeptide which attacks the carbonyl sulphide group (**step 5**) and generates a new tetrahedral intermediate (**step 6**). This final intermediate is resolved by releasing the enzyme (**step 6**), generating a final PG tetrapeptide-tripeptide 3-4 crosslinked product (**step 7**). Analogously to D,D-CPases, at **step 5**, a water molecule may act as a Lewis base following a similar mechanism, realising a PG-tripeptide as product of the reaction. This alternative trimming reaction is catalysed by enzymes denominated L,D-CPases [70].

Functionally, enzymes with 3-3 PG crosslinking activity are related with specialised cell wall growing mechanisms. In *Agrobacterium tumefaciens*, the LDT Atu0845 localises at the growing tip of the cell, co-localising with the pole organising protein PopZ, whereas PodJ accumulates at the old pole of the cell [71]. Additionally, LDTs are responsible of thickening the new cell compartments during unipolar polar growth [72].

LDTs are also part of *Mycobacteria* cell growth machinery, wherein the mature cell wall exhibits mostly 3-3 crosslinks [73]. *M. tuberculosis* encodes five LDT paralogues ($Ldt_{Mt1\text{-}Mt5}$), where Ldt_{Mt1} and Ldt_{Mt2} have been proposed to be part of the elongasome along with the DDTs PonA1 and PonA2 [74].

Loss of Ldt_{Mt1} and Ldt_{Mt2} in *M. tuberculosis* has been shown to result in shorter and less rigid cells [74], [75]. Further, these two LDTs are required for maintenance of infection, immune system evasion and antibiotic resistance [76], [77]. Additionally, loss of Ldt_{Mt5}, which has a nonredundant function with Ldt_{Mt2}, has been proven to compromise *M. tuberculosis* cell wall integrity [75].

The structures of Ldt_{Mt1}, Ldt_{Mt2}, Ldt_{Mt3} and Ldt_{Mt5} have been experimentally determined (PDB ID: 4JMN, 6RLG, 6D4K and 4Z7A, respectively) [75], [78]–[80]. These enzymes comprise a catalytic YkuD-like domain and one or two bacterial immunoglobulin-like beta sandwich domains (BIg-like domain) (**Figure 9**).

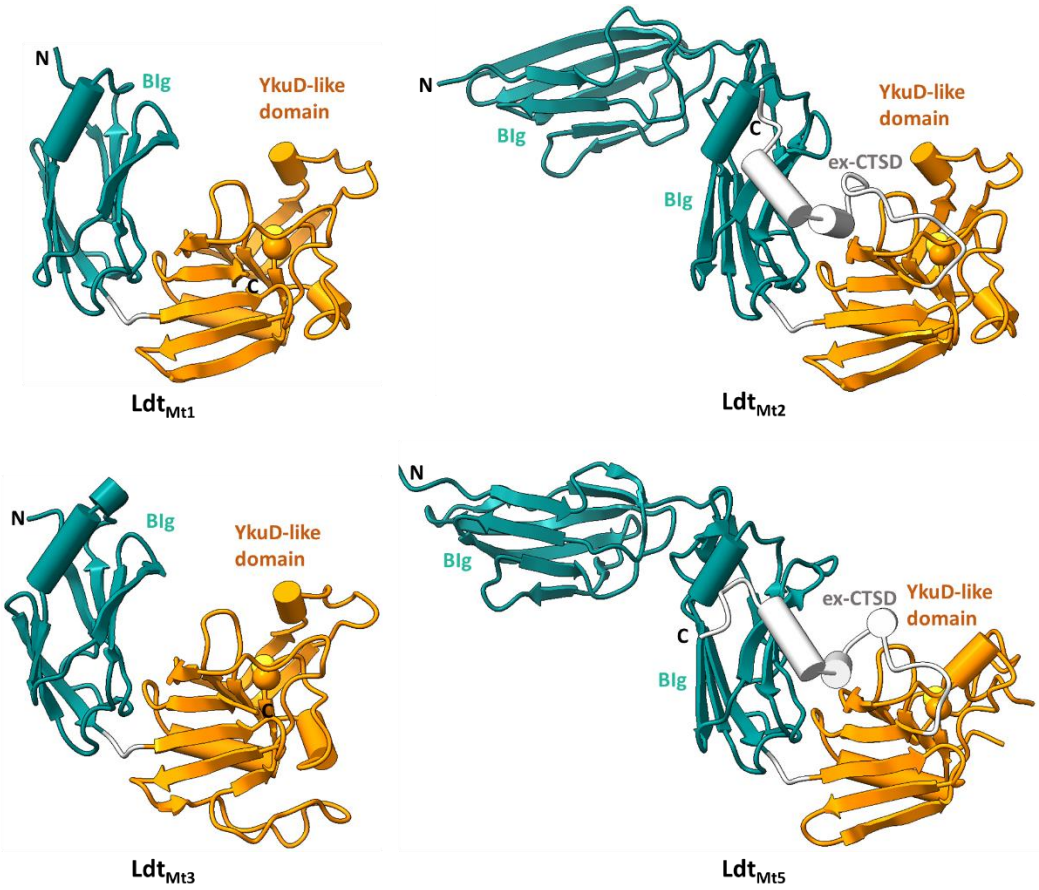


Figure 9: Structure of *M. tuberculosis* LDTs. *M. tuberculosis* encodes five LDT paralogues, comprising one or two BIg N-terminal domains (in teal) and the C-terminal YkuD-like domain (in orange). Ldt_{Mt2} and Ldt_{Mt5} comprise an additional N-terminal BIg domain in a different orientation and an extended C-terminal subdomain (ex-CTSD, in white). Ldt_{Mt1}, Ldt_{Mt2}, Ldt_{Mt3} and Ldt_{Mt5} PDB IDs are 4JMN, 6RLG, 6D4K and 4Z7A, respectively.

The four structures share a structurally similar YkuD-like catalytic core and active site. Ldt_{Mt2} and Ldt_{Mt5} contain an additional N-terminal BIg domain and an extended C-terminal subdomain (ex-CTSD). Nonetheless, Ldt_{Mt2} and Ldt_{Mt5} ex-CTSDs modify the relative orientation of the BIg domain closest to the YkuD-like domain when compared to Ldt_{Mt1} and Ldt_{Mt3} [75], [79]. Moreover, orientation differences of the active site residues between LdtMt2 and LdtMt5 and amino acid composition of the ex-CTSD are hypothesised to affect substrate specificity and/or interactions with other proteins [75].

From a virulence perspective, in *Salmonella enterica* serovar *Typhi*, the secretion of the typhoid toxin is limited to 3-3 crosslinked cell wall sites [81]. The translocation across the bacterial envelope of the typhoid toxin requires the 3-3- crosslinking activity of YcbB and the localised

activity of the N-acetyl muramidase TtsA, which cleaves the β1,4-glycosidic bonds between MurNAc and GlcNAc only on 3-3 crosslinked PG molecules.

In *E. faecium* and *E. coli*, 3-3 crosslinks confer β-lactam resistance via a mechanism involving the production of the alarmone (p)ppGpp in response to the inactivation of PBPs [82]. Under these conditions, it is proposed that YcbB (LdtF) forms a transient complex with the inactivated PBP1b DDT and the CPase PBP5. PBP1b elongates a new PG pentapeptide chain by its still active GT domain while PBP5 generates a PG tetrapeptide that acts as substrate for YcbB [82], [83]. As a result, the PBP inactivated route of cell wall synthesis is bypassed by producing a PG sacculus enriched in 3-3 crosslinks.

Additionally in *E.* coli, three LDTs, YcfS, YbiS and ErfK, crosslink the PG cell wall to the outer membrane (OM) via the Braun lipoprotein or Lpp1 [61], [84]. The Lpp1 attachment of the PG cell wall has been implicated in maintaining envelope integrity, where Δ*lpp* or non-cross linkable Lpp1 mutants display a membrane blebbing phenotype, increased susceptibility to toxic or membrane destabilising agents and release of periplasmic content to the external environment [85]–[89]. Furthermore, Lpp1 variants with increased lengths have been shown to result in the impairment of the OM-IM stress sensor system RcsF-IgaA, blocking the activation of the Regulation of Capsule Synthesis (Rcs) stress response system [90].

Lpp1 is membrane-linked protein synthesised as a preprolipoprotein and is 57 residues long in its processed state [85], [91]. The prepro-Lpp1 comprises an N-terminal type II signal peptide containing a LXXC lipobox, wherein the cysteine residue is modified for anchoring LPP to the inner leaflet of the OM [85]. The cysteine of the lipobox is first diacylated by a lipoprotein glycerol transferase (Lgt), and then the signal peptide is cleaved by a lipoprotein signal peptidase (Lsp) up to the modified cysteine, residue that is finally N-acylated by an apolipoprotein N-acyltransferase (Lpn), generating a triacylated membrane-anchored Lpp1.

The OM-bound Lpp1 forms an acidic α-helical trimeric bundle, wherein the ε-amine of the C-terminal lysine of part of the Lpp1 population is crosslinked to the mDAP of a PG tripeptide [92], [93]. The terminal lysine is part of a highly conserved KYRK motif, corresponding to the only significantly basic region of the protein, and is essential for efficient attachment of the PG sacculus to Lpp1. It has been shown that the KYRK motif of Lpp1 forms an unstructured loop wherein polarity altering mutations in any of the residues produce a diminished PG-Lpp1 crosslinking phenotype [89]. Interestingly, *B. bacteriovorus* HD100 does not encode any

recognizable Lpp1 homologue, thus membrane integrity might be maintained by sacculus attachment via other factors.

1.3.4. PG cell wall hydrolytic modifications: transglycosylases, EPases, CPases and amidases

PG cell wall shape is further sculpted by hydrolytic activities required for polarised cell growth and division, among other processes. These enzymes were originally denominated autolysins, currently Cell Wall Hydrolases (CWHs), and are responsible for hydrolysing glycosidic bonds in the glycan chain of the cell wall (transglycosylases), crosslinks between PG molecules and detaching stem peptides from the glycan chain (peptidases and amidases, respectively) [94].

1.3.4.1. Transglycosylases

Transglycosylases are often referred as cell wall or PG glycosidases (CWGs or PPGs, EC 3.2.1) and catalyse the hydrolysis of the β1,4-glycosidic bonds between two sugar units, releasing the corresponding hemiacetals or hemiketals [95], [96]. The CWGs can be divided respect to their catalytic domains into two broad categories: i) N-acetylglucosaminidases and ii) lysozymes.

N-acetylglucosaminidases hydrolyse the glycosidic bonds between N-acetyl-β-D-glucosamine and the adjacent saccharide by means of an acid-base mechanism [97], catalysed by either an aspartate-histidine or glutamate-aspartate catalytic dyad for GHF-3 or GHF-73 CWG, respectively [98], [99]. On the other hand, lysozymes or N-acetylmuramidases cleave the β1,4-glycosidic bond generating either a reducing open MurNAc or a cyclic 1,6-anhydromuramic acid (anhMurNAc) and are characterised by one of three domains in bacteria: GHF-22, GHF-24, and GHF-25 [94], [100].

1.3.4.2. Cell wall peptidases: EPases, CPases and amidases

Cell wall peptidases (CWPs, EC 3.4) hydrolyse bonds between amino acids of the PG stem peptide [96], and are divided in three groups: i) EPases, cleaving peptide bonds between amino acids within the stem peptide, including crosslinks between two PG molecules, ii) CPases, cleaving and consequently releasing the C-terminal amino acid of the stem peptide and iii) amidohydrolases or amidases, which hydrolyse the MurNAc-stem peptide bond, separating the glucan and peptide portion of the PG substrate [94].

EPases and CPases can be further classified in LD or DD CWPs depending on the chirality of the target bond, mirroring LDTs and DDTs [56]. In fact, several TPases have also CPase or EPase

activity, wherein the reaction mechanism proceeds analogously to TPases, except that the second substrate is a water molecule instead of a second PG molecule.

Currently, ten CWPs signature domains have been reported, with four of these being the most specifically implicated in *Bdellovibrio* endobiotic predators. These domains correspond to the Cysteine Histidine-dependent Amidohydrolases/Peptidases (CHAP), New lipoprotein C/Protein of 60-kD (NlpC/P60) and the D,D- and L,D-TPase domains (**Figure 10**) [94].

Firstly, CHAP domains consist of approximately 120 amino acids folded into a papain-like fold of alpha/beta/alpha topology, having cysteine amidase or peptidase activity, as illustrated by the structure of the CHAP domain of LysGH15 from *Staphylococcus* phage G15 (PDB ID: 4OLK, **Figure 10A**) [101]–[103]. Of the three above individualised domains, CHAPs are by far the less studied group, however, its presence has been usually detected in multi-modular CWH, alongside multiple PG binding motifs (as SH3b and LysM domains) and other amidase or peptidase domains [104]–[107]. The crystal structure of a handful of CHAP domains have been solved and deposited in the PDB, depicting a structural calcium (II) ion coordinated in an hexahedral geometry, suggested to be responsible for stability and activity of the CHAP enzyme, modulating the conformational change of the protein from a closed to an open state [103], [108], [109].

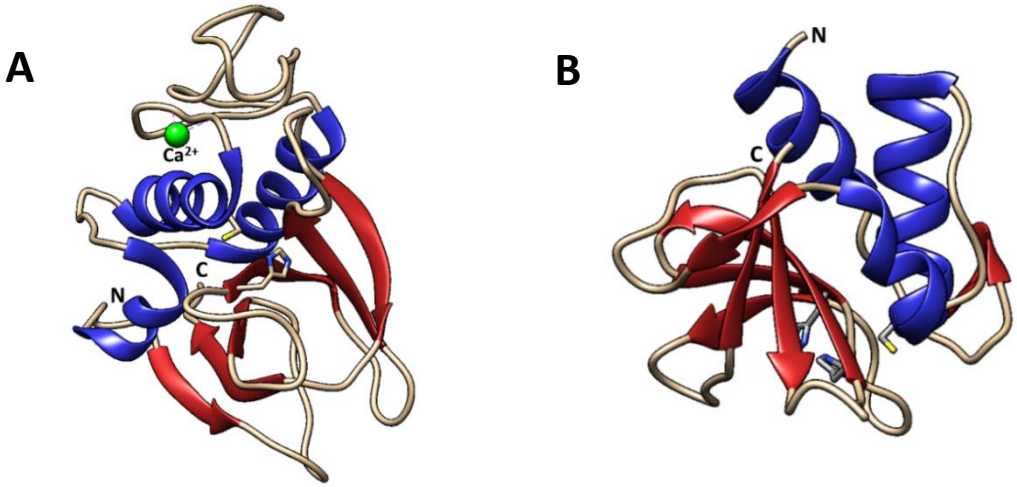


Figure 10: Structure of cell wall peptidases (CWPs) CHAP and NlpC/P60 domains. A) CHAP domain structure of LysGH15 from *Staphylococcus* phage G15 (PDB ID: 4OLK). α-helices are coloured blue while β-strands are coloured red. The CHAP domain adopts a papain-like fold with a catalytic dyad formed by a cysteine and a histidine in the active site. A structural calcium (II) ion coordinated with a hexahedral geometry is coloured in green. **B)** NlpC/P60 domain structure of the PG hydrolase SagA from *Enterococcus faecium* (PDB ID: 6B8C). α-helices are coloured blue while β-strands are coloured red. The domain adopts a papain-like fold as CHAP domains; however, no calcium ion is required for structural stability of the protein. The active site of the domain with a catalytic triad formed by a cysteine and two histidines in the active site.

CHAP domains have been implicated in cell lysis processes and have been more thoroughly studied for their antimicrobial potential. For instance, the LysK endolysin produced by the staphylococcal phage K has been biochemically characterised, determining that even though the enzyme contains both a CHAP and an amidase domain, a CHAP only truncation harbours most of the enzymatic activity [108] [110]. On the other hand, the CHAP domain of Cse of *Streptococcus thermophilus* has been implicated in cell separation by hydrolysing the mature septa due to its endopeptidase activity [111].

Secondly, NlpC/P60 domains structural features are illustrated in **Figure 10B**, depicting the solved structure of the catalytic domain of the PG hydrolase SagA from *Enterococcus faecium* (PDB ID: 6B8C) [112]. This domain also adopts a papain-like fold and is, in consequence, sometimes grouped with the CHAP domain despite both types of domain to differ largely in aminoacidic sequence and absence of structural metal ions stabilising the domain fold [94]. For this group of enzymes, it is postulated that the active site is formed by a catalytic triad formed by a cysteine and two histidine residues, wherein the catalytic domain shifts from a closed to an open conformation is regulated by a loop capping the active site of the enzyme [107], [112], [113].

NlpC/P60 domains are more widespread than CHAP domains, however in bacteria and phage they have been connected with PG hydrolysis. In mycobacteria, the NlpC/P60 endopeptidase RipA is involved in cytokinesis, being responsible for septa hydrolysis [114]. Additionally, endopeptidases LytA and LytB are responsible for the positioning of the septum and consequent cell division in *Lactobacillus plantarum* [115].

In summary, the PG cell wall dynamic is highly regulated by a plethora of enzymatic activities, which modify its structure depending on the multiple factors as the environment and cell cycle stage

1.4. PG modifying enzymes in *B. bacteriovorus* and their connection with predation

1.4.1. DDTs Bd0816 and Bd3459 in the predatory life cycle and self-protection mechanisms

Two homologous genes, *bd0816* and *bd3459*, have been identified to be upregulated 15 minutes after the start of the invasion process, corresponding to the attachment and invasion stage of the predatory cycle [35].

A double in-frame deletion mutant predator strain Δ*bd0816*Δ*bd3459* shows almost no bdelloplast rounding and slowed-down invasion [35]. The double in-frame deletion strain displays a greater number of multiple invasions into one cell, suggesting that invasion rate and bdelloplast rounding are linked to single invasions of *B. bacteriovorus* into *E. coli*.

These genes encode PBP4 homologues with D,D-CPase and D,D-EPase activities, against PG tetrapeptide derivatives, wherein the nucleophilic attack is performed by the catalytic S70 [35].

As shown in **Figure 11A**, Bd3459 (PDB ID: 3V39) is organised in three domains: a DDT catalytic domain (in orange), domain II (in grey) and domain III (in teal). The re-entrant domain III is unique to Bd3459 and has been implicated in the adaptation of this enzyme to predation, allowing for the active site to be more accessible compared to PBP4 and in consequence, highly reactive towards the prey cell wall.

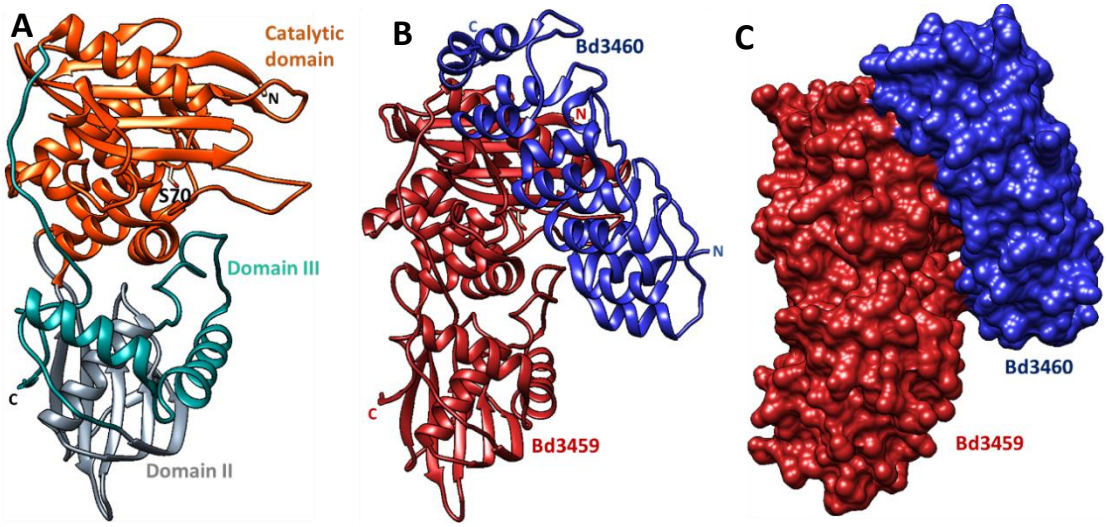


Figure 11: Structure of the DDT Bd3459 and immunity complex between Bd3459 and Bd3460. A) Ribbon diagram of Bd3459 structure (PDB ID: 3V39). The Bd3459 crystallographic structure depicts a DDT catalytic domain (in orange), domain II (in grey) and domain III (in teal). The catalytic S70 is highlighted in the structure. **B)** Bd3459-Bd3460 immunity complex (PDB ID: 5CEC). Bd3460 is an ankyrin domain containing protein (in blue) that blocks the active site of Bd3459 (in red), protecting *B. bacteriovorus* from cell lysis. This complex is secreted to the prey's periplasm, where it dissociates, releasing the Bd3459 EPase and CPase activity. A similar immunity complex is formed between Bd0816 and Bd3460 (PDB ID: 5CER). **C)** Surface representation of the immunity complex between Bd3459 (in red) and Bd3460 (in blue).

Considering its high reactivity, it has been proven that Bd3459 is expressed alongside an ankyrin domain containing protein, Bd3460 (PDB ID: 5CEA), which acts as an immunity protein, blocking access to the Bd3459 active site and protecting the predator from self-rounding (PDB ID: 5CEC **,**

Figure 11B and **C**) [36]. This protein complex is also formed between Bd0816 and Bd3460 (PDB ID: 5CER).

The ankyrin-mediated protection is proposed to involve the retention of Bd3460 in *B. bacteriovorus*' periplasm and the selective secretion of Bd0816/Bd3459 to prey mediated by an unknown mechanism, which is hypothesised to be related to the induced fit of Bd3460 upon complex formation [36].

1.4.2. Deacetylation and prey exit mechanisms

Further modifications are carried out during predation in order to differentiate the otherwise similar PG cell walls from prey and predator. As firstly observed by Thomashow and Rittenberg (1979), and precisely identified by Lambert and Lerner *et al.* (2016), *B. bacteriovorus* produces two prey-targeted GlcNAc deacetylases, Bd0468 and Bd3279 of the CE-4 group [38], [116]. Particularly, Bd3279 presents an accessory N-terminal domain (NTD) capping the active site of the enzyme located in the C-terminal domain (CTD), a feature only observed in *B. bacteriovorus* and closely related predators (**Figure 12A**).

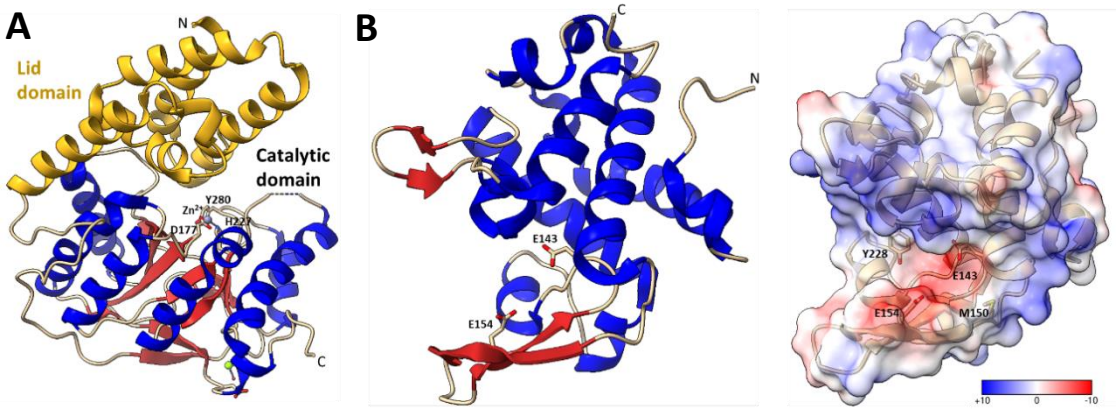


Figure 12: Structure of PG deacetylase Bd3279 and the deacetylation-specific lysozyme Bd0314. A) Ribbon representation of Bd3279 structure (PDB ID: 5JP6). The fold depicts an NTD lid domain (in yellow) partially covering the access to the catalytic CTD (TIM barrel, α-helices in blue and β-strands in red). Active site residues D177, H227 and Y280 are shown coordinating a Zn (II) on (in grey). **B)** Ribbon (left, coloured by secondary structure, α-helices in blue and β-strands in red) and electrostatic surface charge model (right) of the DslA Bd0314 (PDB ID: 6TAB). The catalytic dyad E143/E154 defines a substrate binding cleft, wherein residues M150 and Y228 limit the access to deacetylated PG by steric blockage.

It has been observed that double deletion mutants Δ*bd0428*Δ*bd3279* do not have a predation-compromising phenotype, however, leave bacterial-shaped prey cell ghosts after exit, indicating

that late-acting enzymes involved in predation recognise the deacetylation of prey cell wall specifically [116].

One of the enzymes that has been demonstrated to require such modification of its substrate is the deacetylation-specific lysozyme A (DslA) Bd0314, which is deployed by *B. bacteriovorus* as an "exit" enzyme at the end of the predatory cycle [117]. This hypothesis is supported by time-lapse fluorescent experiments showing that Δ*bd0314* predators have an overall delay in exit time at the end of the cycle and that exogenous addition of Bd0314 produces premature bdelloplast lysis.

The crystallographic structure of Bd0314 depicts a conservation of the general architecture of the enzyme, enabling its grouping with lysozymes, with a proposed inverting mechanism of hydrolysis, utilising residues E143 and E154 as the catalytic dyad of Bd0314, which differs from the Glu-Asp dyad of the currently characterised lysozymes (**Figure 12B**). The specificity of the enzyme is proposed to arise from the steric blockage of acetylated GlcNAc rings *c* and *e* by residues Y228 and M150, limiting the substrate cleft only to deacetylated GlcNAc-MurNAc polymers [117].

1.4.3. LDTs and their role during predation

To this date, the Bd3459, Bd0816-Bd3460 mechanism of prey cell wall modification is the most thoroughly characterised at the molecular level. Nonetheless, as demonstrated by Lambert *et al.* (2011), transcriptomic studies reveal that a large number of other genes related with PG cell wall modification are upregulated in *B. bacteriovorus* predation, wherein a considerable proportion corresponds to evolutionary divergent LDTs (**Table 1**) [21].

Table 1: Upregulated PG related genes in *B. bacteriovorus* predation. Relative upregulation position in the predatosome of each protein and their predicted function or domains are listed.

Relative upregulation position	Protein name	Predicted function or domains
5	Bd3176	LDT
11	Bd1176	LDT
13	Bd3459	DDT
14	Bd0601	DUF1460 NlpC/P60 amidase
15	Bd0553	LDT
16	Bd0816	DDT
29	Bd2480	Predicted CHAP protein
30	Bd0602	Ankyrin domain protein
35	Bd0603	Undetermined
40	Bd1358	LDT
49	Bd2940	LDT
67	Bd3460	Ankyrin immunity protein

B. bacteriovorus HD100 encodes 19 putative LDT genes related to predation, each comprising the YkuD-like LDT core domain and diversified with additional putative or cryptic domains. Based on these additional domains and sequence conservation, *B. bacteriovorus*' LDTs are classified in six groups (**Figure 13**).

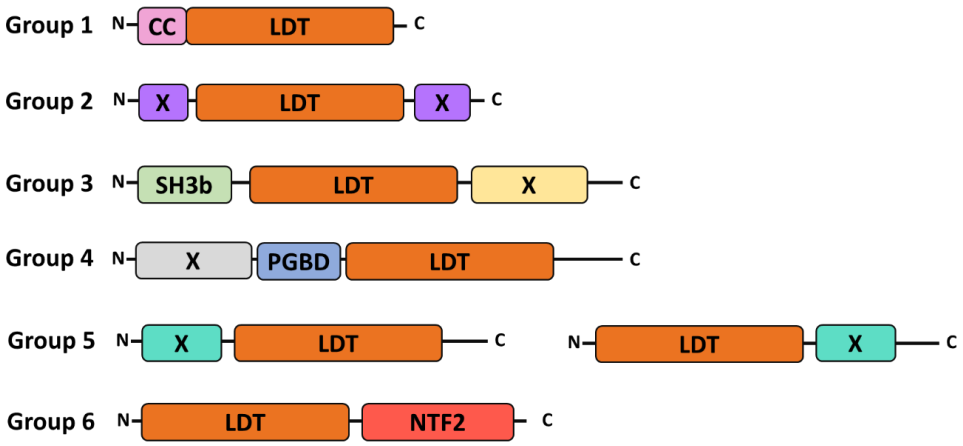


Figure 13: Classification of *B. bacteriovorus* LDTs based on domain diversification. *B. bacteriovorus*' encodes 19 LDTs that can be divided in six groups based on sequence diversification and domain acquisition, all sharing a similar LDT YkuD catalytic core. Cryptic domains are indicated with "X". Group 1 is related to *E. coli* LdtF (DpaA or YafK) and is formed by Bd0553, Bd1763 and Bd2061 comprising a predicted N-terminal coiled-coil (CC) domain. Group 2 is related to the PG-Lpp1 *E. coli* crosslinkers ErfK/Ybis possess two cryptic domains and is formed by Bd3012, Bd3176, Bd3639 and Bd3741. Group 3 is conserved only in *B. bacteriovorus* and comprises an N-terminal SH3b domain and a cryptic C-terminal domain, including Bd0886 and Bd1176. Group 4 is related to *E. coli* YcbB and comprises two accessory domains: a cryptic N-terminal scaffolding domain and a PGBD; this group includes Bd0566, Bd1181, Bd1358 and Bd2908. Group 5 is unique for *B. bacteriovorus* with a cryptic domain that may be located at the N-terminus for Bd0782, Bd1402 and Bd3334 or at the C-terminus for Bd2940 and Bd3093. Group 6 comprised only by Bd1075 whose sequence is related to *H. pylori* Csd6 and is characterised by a C-terminal NTF2-like domain.

Group 1 comprises LDTs Bd0553, Bd1763 and Bd2061 which sequences are distantly related to LdtF (DpaA or YafK) *E. coli* carboxypeptidase, which has been shown to hydrolyse the Lpp1-PG crosslink [118]. These LDTs are predicted to possess a N-terminal coiled-coil (CC) domain followed by the YkuD LDT catalytic domain. The CC domain might contribute to its localisation by association with other CC domain containing proteins in the prey's periplasm [119]–[121]

Group 2 formed by Bd3012, Bd3176, Bd3639 and Bd3741 are related to the Lpp1-PG crosslinkers ErfK/Ybis of *E. coli* and are predicted to incorporate two cryptic domains specific for *B. bacteriovorus* [89].

Group 3 is exclusive for *B. bacteriovorus* and comprises the LDTs Bd0886 and Bd1176. These two enzymes incorporate an N-terminal SH3b domain and a cryptic C-terminal domain. In intracellular pathogens, SH3b domains have been associated with protein oligomerisation via interaction with other SH3b domains, either promoting invasion by interacting with SH3 cell receptors or supporting intracellular survival by modulating specific signalling pathways [122]. However, in the context of PG modifying enzymes, it has been determined that for the lysostaphin from *Staphylococcus simulans* and the *Listeria* phage endolysin, among other CWHs, SH3b domains function as PG binding motifs, participating in substrate recognition and channelling, determining the specificity of each enzyme [123], [124].

Bd0886 and Bd1176 have been proven to be relevant for bdelloplast integrity [33]. It has been noted that even though a double knock-out *B. bacteriovorus* strain Δ*bd0886*Δ*bd1176* still forms a rounded bdelloplast, shows less incorporation of the fluorescent alanine analogue FDAA and is more susceptible to osmotic shock compared to a wild-type predator, suggesting that Bd0886 and Bd1176 participate in stabilising and maintaining the bdelloplast structure [33].

Group 4 containing Bd0566, Bd1181, Bd1358 and Bd2908 is related to the *E. coli* 3-3 PG crosslinker YcbB. These enzymes comprise three domains: 1) an N-terminal putative scaffolding domain, 2) a putative PG binding domain (PGBD) and 3) the LDT catalytic domain [64], [83]. The structure of the *E. coli* YcbB is presented in **Figure 14** (PDB ID: 6NTW).

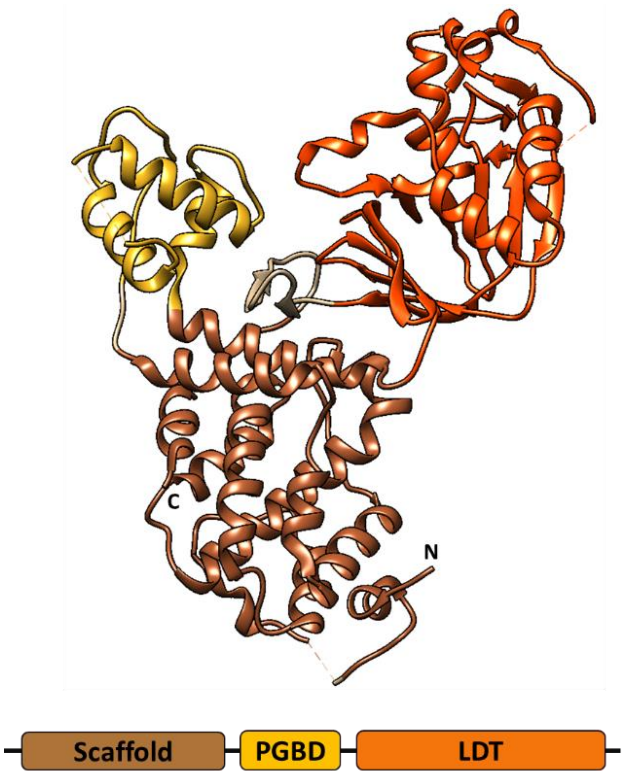


Figure 14: *E. coli* YcbB structure and domain organisation. Ribbon representation of YcbB crystallographic structure (PDB ID: 6NTW). The YcbB structure is re-entrant and is formed by a scaffolding domain (in brown), a PG binding domain (PGBD, in yellow) and the catalytic LDT domain (in orange).

The YcbB structure was solved in a Meropenem acylated form (catalytic residue C528) and possesses a discontinuous arrange of domains or re-entrant topology. The scaffolding domain (in brown) has been demonstrated to bind to the PBP PG synthesis machinery under increased alarmone (p)ppGpp concentrations caused by β-lactam stress, as described in section 1.3.3 [64], [83]. The PGBD (in yellow) is hypothesised to direct the second PG substrate towards the acyl acceptor site of the LDT domain (in orange), which is supported by results showing that a Δ*PGBD* YcbB variant does not rescue a β-lactam sensitive *E. coli* strain, in contrast to the wild type enzyme [83]. Additionally, the LDT catalytic domain of YcbB possesses a flexible capping subdomain that defines the acyl donor and acceptor binding pockets in the active site.

Group 5 of exclusive *B. bacteriovorus*' LDTs comprises a YkuD2-like catalytic domain with an additional N-terminal or C-terminal cryptic domain for Bd0782, Bd1402 and Bd3334, or Bd2940 and Bd3093, respectively.

Finally, group 6 is composed only by Bd1075, which has been shown to participate in defining the cell shape of *B. bacteriovorus*, and in consequence may impact predation and bdelloplast morphology [125]. Bd1075 is evolutionarily related to *H. pylori* Csd6, which has been implicated as a cell-shape determination, wherein Csd6 structure is shown in **Figure 15** (PDB ID: 4XZZ) [68].

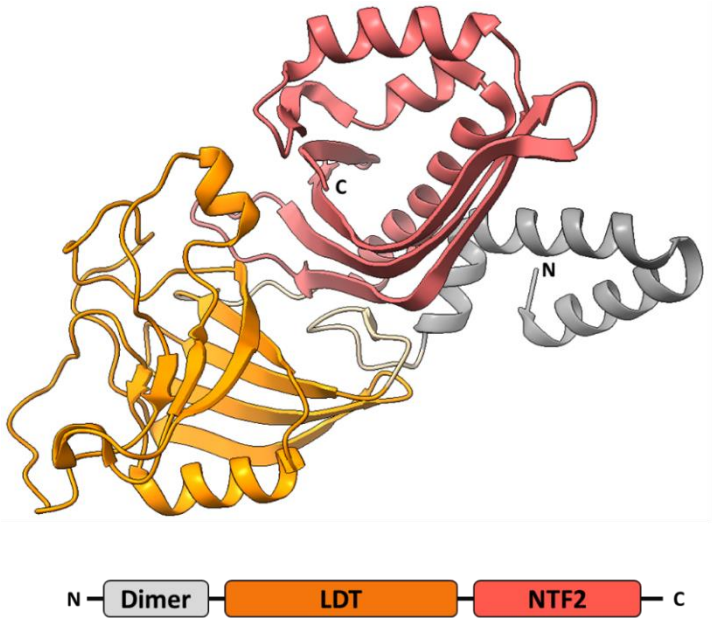


Figure 15: *H. pylori* Csd6 structure. Ribbon representation of Csd6 crystallographic structure (PDB ID: 4XZZ). The overall Csd6 structure is formed by three domains: a dimerization domain (in grey), the YkuD LDT catalytic domain (in orange) and the NTF2-like domain (in red).

Csd6 has been shown to regulate *H. pylori* cell shape by acting as a carboxypeptidase. The C-terminal nuclear transport factor 2-like domain (NTF2) is proposed to bind pseudaminic acid, participating in the deglycosylation of flagellin FlaA.

The implications of the LDT superfamily specialisation and their specific role in the predatory life cycle of *B. bacteriovorus* is, therefore, one of the main focuses of this work. The exact task of each group of enzymes and their unique domains in the modification of the PG cell wall, or if certain groups of LDTs are of more importance depending on the prey to be invaded is also a subject of discussion. Current information regarding the predation scope of *B. bacteriovorus* shows that most prey's share a similar PG stem peptide chemistry, wherein position 3 is occupied by a mDAP residue [126], [127]. Interestingly, spirochetes are resistant to *B. bacteriovorus* predation, and coincidentally substitute the mDAP at position 3 in the stem peptide for an L-

ornithine residue, however the available evidence does not allow to correlate this substitution with predation resistance.

1.4.4. Predatory NlpC/P60 and CHAP amidohydrolases and peptidases

Upon analysing the *B. bacteriovorus* predatosome, the genes contained in the *bd0601-bd0602-bd0603* operon (from hereon, *bd0601-0603* operon) are among the most upregulated genes during predation, with respect to the host independent *B. bacteriovorus* mutant strain (**Table 1**) [21].

The *bd0601-0603* operon encodes three proteins containing N-terminal signal peptides (SP): Bd0601 which is a NlpC/P60 amidase (SecI SP), Bd0602 which is an ankyrin domain containing protein, and the cryptic protein Bd0603 (both Bd0602 and Bd0603 containing lipobox SPs) (**Figure 16**).

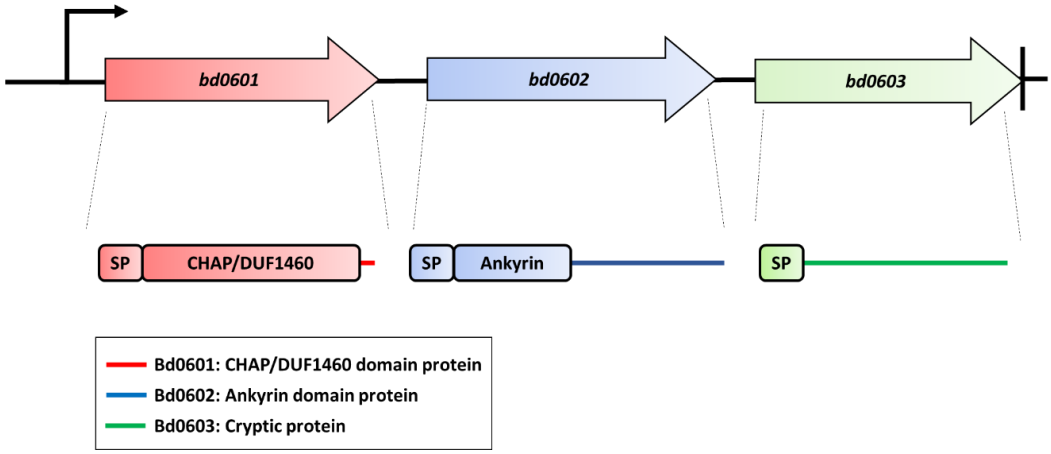


Figure 16: ***bd0601-0603*** **operon organisation and protein products.** The *bd0601-0603* operon (in arrows) encodes three periplasmic proteins (in blocks): Bd0601 (red), a NlpC/P60 protein of the DUF1460 sub-family with possible amidase/peptidase activity; Bd0602 (blue), an ankyrin domain containing protein; and Bd0603 (green), a cryptic protein with no hypothesised function. SP indicates a signal peptide and DUF stands for Domain of Unknown function

Bd0601 is a 33 kD NlpC/P60 domain protein from the sub-family of the DUF1460 domains, which are proposed to have amidohydrolase and peptidase activity, conferred by a catalytic triad of a cysteine and two histidine residues [113]. The related crystal structure of the DUF1460 domain of the *Bacteroides uniformis* AmiA is presented in **Figure 17** (PDB ID: 4H4J).

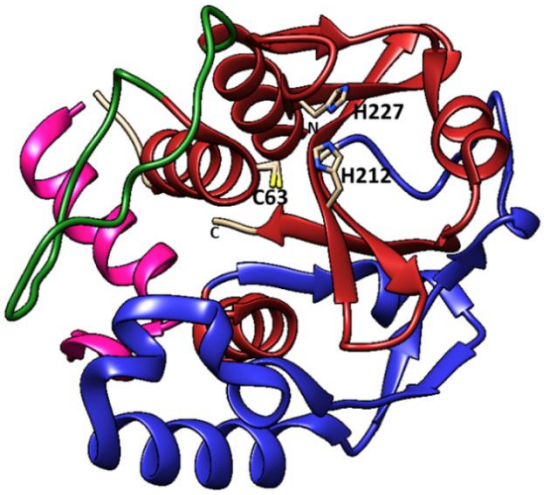


Figure 17: DUF1460 structure. The conserved core structure of a NlpC/P60 domain of the DUF1460 sub-family is presented (PDB ID: 4H4J). The NlpC/P60 domain depicts a papain-like alpha/beta/alpha topology (in red) with the conserved cysteine (C63) and two histidine residues (H212 and H227) indicated in the centre of the protein. Three DUF1460 specific insertions are highlighted in the structure (in green, blue and magenta), forming a pocket to accommodate the disaccharide portion of the PG substrate.

As reported by Xu *et al.* (2014) [113], compared to the reported structure of the NlpC/P60 cell wall hydrolase from *Clostridium difficile* (PDB ID: 4HPE), the DUF1460 structure possess three accessory insertions (in green, magenta and blue) surrounding the central papain fold (in red), forming a substrate binding pocket adjacent to the active site. Specifically, the *C. difficile* NlpC/P60 is a γ-D-Glu-mDAP endopeptidase that hydrolyses a bond within the PG stem peptide to avoid compromising the integrity of the PG cell wall [128], whereas the crystallised DUF1460 proved to hydrolyse the amide bond linking the stem peptide to the polyglycan portion of the PG substrate. Nonetheless, Bd0601 shares 29% sequence similarity to AmiA, showing multiple insertions and deletions in the DUF1460 putative domain, implying that this enzyme may have evolved specifically for predation [113].

Additionally, the *bd0601-0603* operon also encodes a putative ankyrin domain protein, Bd0602. Ankyrin motifs, consisting of a Helix-Turn-Helix (HTH) motif wherein the two helices are arranged in an antiparallel manner, projecting a loop perpendicularly to form a hairpin-like structure, which along with the surface of the "concave" portion of the helices are responsible for the ankyrin binding to its partner(s) [129]. It is possible to propose that Bd0602 could act as an immunity protein for Bd0601, protecting the predator from the amidase activity proposed for DUF1460 proteins, similarly to the Bd3460 immunity protein and Bd3459/Bd0816 PBP endopeptidase [36].

Besides the NlpC/P60 amidases, structurally related CHAP amidases are also upregulated in the predatosome with respect to a HI *B. bacteriovorus* mutant strain. The putative CHAP amidase Bd2480, and its homologue Bd1177, has been identified as part of the predatosome in spite of retaining a poor sequence similarity with the canonical CHAP domains due to going through several permutations and evolutionary specialisations, and possibly the acquisition of novel uncharacterised domains [94], [130].

Therefore, one of the aims of the present work is to contribute to the understanding of the NlpC/P60 and CHAP domain amidohydrolases, including their regulation by possible ankyrin mediated self-protection mechanisms.

1.5. Cryptic proteins as part of the *B. bacteriovorus* predation arsenal

As shown in **Figure 2**, over 50% of the *B. bacteriovorus* predatosome gene products are annotated as hypothetical proteins with no putative domains, obscuring the thorough understanding of the predatory life cycle [21].

Alongside Bd0603, which is part of the bd0601-0603 operon, and the poorly conserved CHAP domain proteins Bd2480 and Bd1177, other cryptic proteins are hypothetically secreted inside the prey, including the 36th most upregulated gene *bd0675*.

The gene *bd0675* encodes a 14 kD secreted protein with four conserved cysteines and no (identifiable) predicted domain. Proteins similar to Bd0675 are found in different *B. bacteriovorus* strains and in other predators such as *Halobacteriovorax* spp [131], but no information regarding their function is available. Therefore, structural approaches are expected to aid in the clarification of Bd0675 function in *B. bacteriovorus* predation.

1.6. Research aims

The general aim of this work is to structurally characterise PG modifying enzymes deployed by *B. bacteriovorus* during predation.

Firstly, this will be focused on the structural determination of novel LDTs from groups 2 and 4-6, which in combination with previously solved LDTs of group 1, 3 and 5 will allow discussing about their role during *B. bacteriovorus* lifecycle. Furthermore, structural data covering the full repertoire of *B. bacteriovorus* LDTs will be used for stablishing relationships regarding their evolution and enzymatic specificity.

Secondly, structural studies will be directed on two putative NlpC/P60 proteins: the putative DUF1460 amidohydrolase Bd0601, and the evolutionary divergent putative NlpC/P60 Bd1177. Bd0601 structural studies will be centred on the possible predatory adaptations suffered by this enzyme encoded by the upregulated operon *bd0601-bd0603*. On the other hand, Bd1177 structural data will be used to determine the novel aspects defining this possible novel family of CWPs.

Finally, this thesis will attempt to assign a biological function to the cryptic protein Bd0675 using the information provided by its three-dimensional structure, which is expected to adopt a previously uncharacterised fold.

CHAPTER 2

2. MATERIALS AND METHODS

2.1. General methods

2.1.1. Molecular cloning

The list of the protein constructs reported in this research is shown in **Table 2**, indicating protein constructs cloned during this thesis and constructs previously cloned and kindly provided by Dr. Ian Cadby.

Table 2: Attribution of protein constructs used in this thesis. Expression vectors used are indicated. Previously cloned LDTs were kindly provided by Dr. Ian Cadby.

Protein construct	Expression vector	Attribution
Bd0553	pET28a(+)	Provided by Dr. Ian Cadby
Bd0566	pET41c(+)	This thesis
Bd0886	pET28a(+)	Provided by Dr. Ian Cadby
Bd1075	pET41c(+)	Provided by Dr. Ian Cadby
Bd1402	pET41c(+)	Provided by Dr. Ian Cadby
Bd3334	pET28a(+)	Provided by Dr. Ian Cadby
Bd3741-MBP	pET28a(+)	This thesis
Bd0601-0603, codon optimised	pET21d(+)	This thesis. Gene synthesis (Genescript©)
Bd0601, codon optimised	pET28a(+)	This thesis
Bd1177	pET28a(+)	This thesis
Bd1177 C86A	pET28a(+)	This thesis
Bd1177 C86A/Y124C/Y508C	pET28a(+)	This thesis
Bd0675	pET28a(+)	This thesis

The genes of interest were cloned in *E. coli* expression vectors by a restriction free (RF) cloning approach [132], using the http://www.rf-cloning.org/ online tool for primer design. Briefly, primers to amplify each target gene from *B. bacteriovorus* strain HD100 genomic DNA were designed comprising a flanking region complimentary to the desired destination plasmid.

Thus, in a first PCR reaction, a PCR product consisting of the target gene flanked by the destination plasmid complimentary regions is generated. Later, the PCR product is purified and used in a second PCR reaction as a unique megaprimer inserting the target gene in the designated position in the destination plasmid. Finally, the original methylated plasmid is eliminated by DpnI digestion, transformed into calcium-competent XL10 gold cells, and plated in LB-agar plates supplemented with 60 µg/mL of the corresponding selection antibiotic.

The primers designed and used in this thesis for RF cloning and the destination vectors are listed in **Table 3**. The nucleotide sequence for each gene was retrieved from

http://xbase.warwick.ac.uk/ database and primers were desired considering the amino acid sequence of the mature exported proteins, thus removing the signal peptide (SP) coding sequence of the gene.

Table 3: Primer list for RF cloning of the different protein constructs into each expression vector. Fw: Forward primer, Rv: Reverse primer, Tm: melting temperature and Ta: annealing temperature.

Protein construct	Destination Vector	Fw (5'-3')	Fw Tm/°C	Rv (5'-3')	Rv Tm/°C	Ta/ °C
Bd0566 C-terminal His_6 tag	pET28a(+)	AATTTTGTTTA ACTTTAAGAAG GAGATATACCA TGCAGAATTCA AAGGACTACTT GGTC	56	GCCGCTGCTG TGATGATGAT GATGATGCCA GGCACCCAGT CTTAAC	63	55
Bd0601 C94A N-terminal His_6 tag	pET28a(+)	GGCAGCAGCC ATCATCATCAT CATCACACGCA ATCGCACAAAG ATC	63	CGCGCGGCAC CAGGCCGCTG CTTTAAAGGC CGCCCACAAA	73	60
Bd0675 N-terminal His_6 tag	pET28a(+)	AAGGAGATAT ACCATGCATCA TCATCATCATC ACGGTGGCAA TGACTTTGTGA G	61	CGCGCGGCAC CAGGCCGCTG CTTTAGAACT GGCAAGTCGT GG	73	55
Bd0675 C-terminal His_6 tag	pET28a(+)	AATTTTGTTTA ACTTTAAGAAG GAGATATACCA TGGGTGGCAA TGACTTTGTGA G	56	GCCGCTGCTG TGATGATGAT GATGATGGAA CTGGCAAGTC GTGGA	63	55
Bd1177 N-terminal His_6 tag	pET28a(+)	GGCAGCAGCC ATCATCATCAT CATCACCAGAC GGCTTCCAAAG CA	63	CGCGCGGCAC CAGGCCGCTG CTTTATCTCAG GGAATTCAGC AAACG	73	57
Bd3741 N-terminal His_6-MBP-TEV cleavable tag	pET28a(+)-MBP	AGGATCCCTCA GTACAGGCTG CCAATCCCGCT CTGAACTTGA	59	GTCTTCTAGA AGGTTCTTTAA GTAATCTGTC AGTATTAATAT TCCTCTACGAT AACGATT	61	55

Each target gene was amplified in the first round PCR using a Phusion high-fidelity DNA polymerase (APExBIO©, cat# K1031) in a 50 µl reaction mixture consisting of 1X Phusion HF Buffer, 0.2 mM dNTPs, 0.5 µM of each primer, 0.4 ng of HD100 genomic DNA stock and 0.02 U/µL of enzyme. The amplification conditions were: i) an initial denaturation of 98 °C for 30 seconds;

ii) 30 cycles of denaturation, annealing and extension of 98 °C for 15 second, 55-60 °C (target dependent, Ta as listed in **Table 3**) for 15 seconds and 72 °C for 30 seconds per Kb of target gene, respectively; and iii) a final extension of 72 °C for 5 minutes.

The quality of the amplification was assessed by running 5 µL of PCR product mixed with 1 µL of 6X DNA loading buffer (30% glycerol and 0.25% bromophenol blue in 1X TAE buffer) in 1% agarose gels prepared and run in 1X TAE buffer (1 mM EDTA, 40 mM Tris and 20 mM acetic acid) and visualised with SafeView Nucleic Acid Stain® (NBS Biologicals™, cat# NBS-SV1) according to the manufacturer guidelines. Single band PCR products were purified using the QIAquick PCR Purification Kit (QIAGEN©, cat# 28104) according to the manufacturer instructions.

Subsequently, the second round PCR was carried out in a 20 µL reaction mixture with the same components and concentrations as the first round but replacing primers and HD100 DNA template with the first-round product and destination vector in an insert: plasmid molar ratio of 20. In general, for inserts of approximately 1 kb, 220 ng of insert and 60 ng of a plasmid of 5-6 kb were used. The amplification programme consisted in i) an initial denaturation of 98 °C for 30 seconds; ii) 15 cycles of denaturation and extension of 98 °C for 15 seconds and 72 °C for 30 seconds per kb of final construct (insert plus plasmid); and iii) a final extension of 72 °C for 5 minutes. This second PCR does not comprise an annealing step.

To digest the parental unmethylated plasmid DNA, 2 U of DpnI (1 µL of a 2 U/µL stock, NEB©, cat# R0176S) were added to the PCR amplification product and incubated at 37°C for three hours. Afterward, 10 µl of the digestion product were incubated on ice with 100 µL of XL10 gold chemo competent cells for 30 minutes, heat-shocked at 42 °C for 50 seconds, incubated on ice again for 5 minutes and recovered in 1 mL of LB broth for 2 hours at 37 °C with constant agitation. The recovered cells were pelleted at 11000 x *g* for 1 minute, resuspended in 100 µL of supernatant, plated in LB plates with the corresponding selection antibiotic and incubated overnight at 37°C.

Positive colonies were screened by PCR using T7 sequencing primers (T7 Forward: 5' TAATACGACTCACTATAGGG 3' and T7 reverse: 5' GCTAGTTATTGCTCAGCGG 3') and resuspending each colony, after replicating each clone in a LB-antibiotic plate, in a 10 µl reaction mixture consisting of 1X MyTaq™ Red Mix (2X stock mix, Bioline©, cat# BIO-25043) and 0.4 µM of each T7 primer. The PCR programme consisted of i) an initial denaturation of 95 °C for 5 minutes; ii) denaturation, annealing and extension of 95 °C for 15 seconds, 55 °C for 15 seconds and 72 °C

for 30 seconds per Kb, respectively; and iii) final extension of 72 °C for 5 minutes. The 10 µL of each PCR products were analysed on 1% agarose gels.

Finally, positive clones were inoculated in 5 mL of LB-antibiotic and grown overnight at 37 °C and constant agitation for plasmid purification using the GenElute™ Plasmid Miniprep Kit (Sigma®, cat# PLN350) according to the manufacturer's indications and confirmed by sequencing in SourceBioscience©.

Successfully cloned gene constructs were finally transformed into calcium competent BL21-CodonPlus (DE3)-RIPL expression strain (from here onwards BL21-RIPL strain, Agilent Technologies©, cat# 230280) as indicated for the XL10 gold strain.

2.1.2. Site-directed mutagenesis

Point mutation for the Bd1177 C86A construct were introduced by adapting the Q5 site-directed mutagenesis kit guidelines (Q5-SDM, NEB©, cat# E0554S).

Briefly, template DNA was amplified using the mutagenesis primers designed using the NEB© online tool (available at https://nebasechanger.neb.com/) in a 25 µL PCR reaction mix consisting of 1X Q5 PCR mix, 0.5 µM of each primer and 10 ng of template DNA. The PCR amplification conditions were as follow: initial denaturation at 98 °C for one minute; followed by 25 cycles of denaturation at 98 °C for 30 seconds, 30 seconds of primer annealing at a target dependent temperature and a 72 °C extension step for five minutes (30 seconds per kb of the DNA template); and a three-minute final extension step at 72 °C. DNA amplification was corroborated by visualising 5 µL of PCR product in 1% agarose gels.

A T4 kinase/ T4 ligase/ DpnI (KLD) reaction was performed for assembling the mutagenized PCR product and digestion of the parental wild-type DNA. The 10 µL KDL reaction mixture consisted of 1 µL of site directed mutagenesis product, 1X of T4 ligase buffer (NEB©, cat#M0202S), 300 U of T4 ligase (1 µL of a 400 U/µL stock, NEB©, cat#M0202S), 7.5 U of T4 polynucleotide kinase (0.75 µL of a 10 U/µL stock, NEB©, cat# M0201S) and 10 U of DpnI (0.5 µL of a 20 U/µL stock, NEB©, cat#R0176S). The reaction mixture was incubated at room temperature for one hour and transformed in XL10 gold chemo competent cells and plated on the corresponding antibiotic. Procedures for transformation, selection, PCR screening and sequencing of positive clones are indicated in section 2.1.1.

The engineered Y124C/Y508C disulphide bond mutation was introduced in the Bd1177 C86A construct by RF-cloning as indicated in section 2.1.1, designing the Y124C and Y508C mutations in the forward and reverse primer, respectively.

The mutagenesis primers for the Bd1177 C86A and Bd1177 C86A/Y124C/Y508C are listed in **Table 4**.

Table 4: Primer list for site-directed mutagenesis. Fw: Forward primer, Rv: Reverse primer, Tm: melting temperature and Ta: annealing temperature.

Protein mutant	Fw (5'-3')	Fw Tm/°C	Rv (5'-3')	Rv Tm/°C	Ta/ °C
Bd1177 C86A	GTTTGCCGATGCTG CGGATCTGCCTTAC TATCTGCGC	72	AGCCGCAGATCCGC CGGA	75	72
Bd1177 C86A/Y124C/Y508C	GGGATCTGAAAACA AAGACCCTCGCTGT TCCAAATTCGGCAA TGC	56	CAGTGAATAATCCA TGCGGGCGTCACAG CGACGCTCCCACTG	61	55

2.1.3. Recombinant protein expression in native conditions

Each BL21-RIPL *E. coli* strain containing the corresponding gene in an expression plasmid was inoculated in 10 mL of LB media supplemented with 60 µg/mL of antibiotic and grown overnight at 37 °C and 180 rpm constant agitation. The next day, the 10 mL of overnight cultures were inoculated in 1 L of expression media, in 2.5 L baffled flasks, supplemented with 60 µg/mL of antibiotic and grown at 37 °C and 180 rpm of constant agitation until OD_{600}= 0.6-0.8.

Depending on the protein, two types of media and inductions were used: i) Terrific Broth (TB) media composed by 2.4% yeast extract, 1.2% tryptone, 54 mM K_2HPO_4, 16.2 mM KH_2PO_4 and 0.4% glycerol; in this case after reaching an OD_{600} of 0.6-0.8, the cultures were induced with 0.5 mM of IPTG at 18°C overnight, and ii) Autoinduction (AI) media comprising 0.5% yeast extract, 1% tryptone, 25.4 mM Na_2HPO_4, 25 mM KH_2PO_4, 50.5 mM NH_4Cl and 2.5 mL of trace metal mixture per litre of media, wherein the trace metal mixture consists in 130 mM $Na_2EDTA•2H_2O$, 80 mM $ZnSO_4•7H_2O$, 180 mM H_3BO_3, 30 mM $MnCl_2•4H_2O$, 7 mM $CoCl_2•6H_2O$, 6 mM $CuSO_4•5H_2O$ and 0.8 mM $(NH_4)_6(Mo_7O_2)_4•4H_2O$). For autoinduction, media was supplemented with either Sugar Mix A (0.19% lactose and 0.37% glycerol, final concentration in media) or Sugar Mix N (0.05% glucose, 0.2% lactose and 0.5% glycerol, final concentration in media) from the start of the cell growth. The expression conditions for each protein are summarised in **Table 5**.

Table 5: Expression of the proteins crystallised in this thesis. Each protein construct, with the indicated tags and position in the protein sequence, was tested for expression in AI and TB media. *Bd0675 proved to be toxic when expressed in native conditions and was expressed in inclusion bodies for subsequent purification and refolding.

Protein	Purification tag	Expression media	Inducer	Induction condition
Bd0566	C-terminal His_6-tag	AI	Sugar Mix N	18 °C overnight
Bd0601 C94A	C-terminal His_6-tag	AI	Sugar Mix A	18 °C overnight
Bd0675*	N-terminal His_6-tag	TB	IPTG	37 °C, four hours
Bd1075	C-terminal His_6-tag	AI	Sugar Mix N	18 °C overnight
Bd1177	N-terminal His_6-tag	TB	IPTG	18 °C overnight
Bd1177 C86A	N-terminal His_6-tag	TB	IPTG	18 °C overnight
Bd1177 C86A/Y124C/Y508C	N-terminal His_6-tag	TB AI	IPTG	18 °C overnight
Bd3334	N-terminal His_6-tag	AI	Sugar Mix N	18 °C overnight
Bd3741	N-terminal His_6-MBP TEV cleavable tag		Sugar Mix N	18 °C overnight

After starting the induction, the cultures were incubated overnight at 18 °C and 180 rpm of constant agitation. The induced cultures were harvested the next day by centrifugation in a JLA-8.1000 rotor at 4,000 rpm (3,060 x *g*) for 20 minutes. The supernatants were then discarded and each pellet coming from 1L of culture were scraped into 50 mL conical centrifuge tubes and stored at -80 °C until required.

2.1.4. Recombinant protein expression in inclusion bodies

Poorly folded or highly toxic proteins (see **Table 5**) were expressed as inclusion bodies. Briefly, 10 mL of LB-antibiotic overnight grown cultures were inoculated in 1L of TB media and grown at 37 °C and 180 rpm of constant agitation until OD_{600} reached 0.6-0.8. The cultures were then induced with 1.0 mM IPTG for four hours at 37 °C and 180 rpm of constant agitation to cause the formation of inclusion bodies. The induced cultures were then harvested and stored as indicated for expression of native recombinant proteins.

2.1.5. Protein purification in native conditions by IMAC Nickel chromatography

Each stored pellet was thawed at room temperature and resuspended up to a final volume of 35-40 mL with pH 8.0 lysis buffer (50 mM Tris pH 8.0, 300 mM NaCl and 20 mM imidazole). For proteins with an isoelectric point (pI) between 7 and 8, and therefore predicted to be less soluble in pH 8.0 lysis buffer, were lysed in a modified lysis buffer consisting of 50 mM sodium phosphate buffer pH 6.8, 300 mM NaCl and 20 mM imidazole. Cells were then disrupted by sonication using

a Fisher Scientific™ Model 505 Sonic dismembrator, using six pulses of 45 seconds at 40% amplitude, with 1.5 minutes of rest on ice between pulses. Cell debris was then removed by centrifugation at 20,000 rpm (48,400 x *g*) for 1 hour in a JA25.50 rotor. The supernatant was loaded into a 5 ml HisTrap® FF column (GE Healthcare life science™), pre equilibrated with five column volumes (CV) of lysis buffer, at a flowrate of 1 mL per minute. The column was then washed with 20 CV of lysis buffer to remove contaminants and later eluted with a 20 CV imidazole gradient between lysis buffer and elution buffer (50 mM Tris pH 8.0, or 500 mM sodium phosphate buffer pH 6.8, 300 mM NaCl and 500 mM imidazole) on an AKTA™ Pure system operated by the Unicorn® software, monitoring elution by UV absorption at 280 nm, collecting 3 mL fractions.

Fractions were analysed by 15% SDS-PAGE to assess purity, pooled, and dialysed overnight at 4°C against 2 L of buffer (protein dependent) in a 12-14 kD or 3.5 kD MWCO dialysis tubing (Spectrum™ Spectra/Por™, cat#10142634) and concentrated for Size Exclusion Chromatography (SEC) purification. The dialysis buffers for each protein are the same as the buffers employed for SEC and are protein dependent, as detailed in **Table 6**.

Table 6: Protein theoretical molecular weight (Mw), isoelectric point (pI) and molar extinction coefficient (ε) considering all the cysteine pairs as oxidised, calculated for each protein sequence using the https://web.expasy.org/protparam/ online tool. The dialysis and final stabilisation buffer for each protein is also specified.

Protein	Mw/Da	pI	ε/$M^{-1}cm^{-1}$	Purification strategy	Dialysis/ SEC buffer
Bd0566	56,079.53	9.25	101,425	1) Nickel affinity, 2) SEC	25 mM sodium citrate buffer pH 6.0, 200 mM NaCl and 2.5% glycerol
Bd0601 C94A	32,332.91	8.97	37,360	1) Nickel affinity, 2) SEC	15 mM HEPES-NaOH pH 7.5, 200 mM NaCl
Bd0675	14,673.40	5.02	9,190	1) Nickel affinity and refolding on column, 2) SEC	25 mM HEPES-NaOH pH 7.5, 200 mM NaCl, 2.5% glycerol
Bd1075	36,154.90	8.93	52,830	1) Nickel affinity, 2) SEC	25 mM sodium citrate buffer pH 6.0, 200 mM NaCl and 3.5 mM β-mercaptoethanol
Bd1177	59,218.67	8.07	88,615	1) Nickel affinity, 2) SEC	15 mM MES-NaOH pH 6.0 and 200 mM NaCl
Bd1177 C86A	59,186.61	8.10	88,615	1) Nickel affinity, 2) SEC	15 mM MES-NaOH pH 6.0 and 200 mM NaCl
Bd1177 C86A/Y124C/Y508C	59,066.54	8.04	85,760	1) Nickel affinity, 2) SEC	15 mM MES-NaOH pH 6.0 and 200 mM NaCl
Bd3334	26,874.21	9.45	39,435	1) Nickel affinity, 2) SEC	20 mM Tricine pH 8.0 and 200 mM NaCl
Bd3741 (MBP tag removed)	25,351.66	9.75	26,360	1) MBP affinity, 2) TEV cleavage in 1.5 M NaCl, 3) SEC in 1.5 M NaCl, 4) Desalting by dialysis	25 mM HEPES-NaOH pH 7.5, 1.5 M NaCl, 2.5% glycerol (steps 2 and 3) 25 mM HEPES-NaOH pH 7.5, 200 mM NaCl, 2.5% glycerol (step 4)

2.1.6. Protein purification from inclusion bodies by IMAC Nickel chromatography and refolding on-column

Thawed pellets were resuspended to a final volume of 35-40 mL of denaturant lysis buffer (6M Gdn-HCl, 100 mM Tris pH 8.0, 300 mM NaCl, 10 mM imidazole and 3.5 mM β-mercaptoethanol) and left tumbling overnight at room temperature to disaggregate the inclusion bodies. The next day, resuspended pellets were sonicated using 8 pulses of 45 seconds at 40% amplitude, with 1.5 minutes of rest on ice between pulses.

Cell debris was removed by centrifugation at 20,000 rpm (48,400 x *g*) for 1 hour in a JA25.50 rotor, and the supernatant was loaded into a 5 mL HisTrap® FF column (GE Healthcare life

science™), pre equilibrated with 5 CV of denaturant lysis buffer, at a flowrate of 1 mL per minute. The column was washed with 10 CV of lysis buffer to remove contaminants.

To refold the protein, the column was washed with 10 CV of lysis buffer with decreasing concentrations of urea and β-mercaptoethanol as follows: 4 M urea and 3 mM of β-mercaptoethanol, 2 M urea and 1 mM β-mercaptoethanol, 1 M urea and no β-mercaptoethanol, and lysis buffer without urea or β-mercaptoethanol. The refolded protein was eluted with 5 CV of native lysis buffer supplemented with 500 mM imidazole. The eluted protein was finally dialysed and concentrated for loading in a SEC column for purification of the protein in the correct conformation, as indicated for natively purified proteins.

2.1.7. Protein purification in native conditions by MBP-Trap® chromatography

MBP fusion proteins were purified using an MBP-Trap® FF 5 mL column (GE Healthcare life science™). In summary, the bacterial pellets were lysed as indicated for native proteins, in a lysis buffer chosen depending on the isoelectric point (pI) of the target protein (see **Table 6**). The lysates were then loaded in the column and washed with 20 CV of the specific lysis buffer. The target protein was then eluted with 5 CV of lysis buffer supplemented with 10 mM maltose in 1 CV fractions. The protein containing fractions (determined by absorbance at 280 nm, A_{280}) were analysed by SDS-PAGE, pooled, supplemented with 3 mM of β-mercaptoethanol and mixed with TEV protease in a mass ration of 1:100 (TEV: fusion protein). The TEV cleavage reactions were incubated overnight at 4 °C and constant agitation. To separate the MBP and the un-cleaved MBP tagged protein from the successfully cleaved one by SEC.

For Bd3741, the TEV cleavage reaction and subsequent SEC was carried out with in high salt buffer (25 mM HEPES-NaOH pH 7.5, 1.5 M NaCl, 2.5% glycerol and 3 mM β-mercaptoethanol) to separate the cleaved protein from the MBP.

2.1.8. Additional purification of proteins by Size Exclusion Chromatography (SEC)

All proteins were further purified by SEC at room temperature to eliminate contaminant proteins, separate solubility fusion tags and to get rid of soluble aggregates of the protein. Depending on molecular weight differences between the target proteins and contaminants, either a Superdex® 75 26/600 or Superdex® 200 16/600 column (GE healthcare life science™) was used for SEC purification, both connected to an AKTA™ Pure system and operated by the Unicorn® software.

The columns were equilibrated with 1.2 CV of filtered and degassed buffer (see **Table 6**), after which the samples were loaded and eluted at a 2 mL/min flow rate and collected into 3 mL fractions after the void volume has passed. The collected fractions corresponding to the correct peak were then analysed by SDS-PAGE, pooled and concentrated.

2.1.9. Protein concentration, quantification, and storage

Protein concentration was calculated by measuring A_{280} absorbance in a DeNovix™ DS-11 FX nanodrop, using the theoretical molecular weight and molar extinction coefficient (ε) calculated from the aminoacidic sequence of each protein using the online tool https://web.expasy.org/protparam/ (see **Table 6**).

Proteins were concentrated after dialysis in Vivaspin 15R, 10,000 (cat# VS15RH02) or 5,000 (cat# VS15RH12) MWCO ultrafiltration centrifuge (Sartorius©) units at 3,500 x *g* at 4 °C up to a volume between 2-3 mL for SEC. The final purified proteins were then concentrated by the same method up to a concentration of 15-35 mg/mL (protein dependent) and aliquoted in 200 µL aliquots. Unless otherwise indicated, protein aliquots were flash frozen with liquid nitrogen and stored at -80°C until needed.

For Bd0566, protein aliquots required the supplementation of 5% ethylene glycol as cryoprotectant to be stored frozen at -80°C and avoid precipitation after thawing.

In the case of Bd0675, protein aliquots were concentrated to 35 mg/mL. It has been noted that this protein reversibly precipitates at 4 °C or lower temperatures and thus is best stored at room temperature.

2.1.10. Protein crystallisation screening and optimisation

Proteins were screened for crystallisation conditions at different protein concentrations at 18 °C by the sitting drop method using a Mosquito® LV crystal automatic nanolitre handler, each drop consisting of 0.5 µL of protein solution and 0.5 µL of crystal screen solution, or manually by the sitting drop method using 2 µL of protein sample and 2 µL of crystallisation solution in 96 well sitting drop trays (cat# MD15-XPT-40, Molecular Dimensions™). The screenings with positive crystallisation conditions are detailed for each protein individually.

Optimisation for the crystallisation conditions were carried out, if required, either by hanging drop customised screen optimisation in 24 well cell culture plates using a 2 µL: 2 µL protein

solution: crystallisation solution ratio and/or microseed matrix screening (MMS) in 96 well sitting drop trays (cat# MD15-XPT-40, Molecular Dimensions™) using a 1.5 µL:0.5 µL: 1 µL of protein solution: protein seed: crystallisation solution ratio.

For the MMS crystallisation protocol, the seed was prepared by selecting suboptimal crystals from a positive condition (usually needles or spherolites) and transferred along with their mother liquor to a 1.5 mL microcentrifuge tube and crushed with a 2.381 mm microseeding glass beads (Molecular Dimensions®, cat# MD2-14) in a total volume of 60 µL solution of mother liquor, flash frozen with liquid nitrogen and stored at -80 °C until needed. For crystal seeding, the seed solution was tested for dilutions 1:10 or 1:100 volume times in mother liquor.

For native crystal collection, each positive crystallisation condition mother liquor was replaced with crystallisation solution supplemented with a final concentration of 20-30% of cryoprotectant if required. For proteins lacking a reported evolutionarily related protein structure, a heavy metal phasing strategy was adopted, wherein the crystallisation solution was supplemented with the cryoprotectant plus 10 mM of K_2PtCl_4 and incubated for one hour at 18 °C, prior to crystal collection. For compliance of platinum-derived chemicals handling, these concentrated stocks were prepared by Dr. Andrew Lovering.

2.1.11. X-ray diffraction data collection and analysis

Crystallographic data was collected on IO3 and IO4 Beamlines at Diamond Light Source (Oxford, United Kingdom). All data for native crystals were collected in routine data collection procedures: IO3 (3600 images, 0.1° Ω Oscillations, 100% transmission, 0.05 s exposure, optimal beamline wavelength of 0.9762 Å) and IO4 (3600 images, 0.1° Ω Oscillations, 100% transmission, 0.05 s exposure, 0.9763 Å wavelength). The resolution was initially limited between 2.0-2.5 Å and posteriorly adjusted depending on the quality of the diffraction observed during the initial crystal screening. The diffraction data was auto processed by the ISPyB automated pipelines Xia2-3dii, Xia2-dials and Autoproc, and the datasets with best resolution and statistical parameters (completeness and minimal radiation damage) were selected for structure solving.

For heavy metal phasing experiments, data was collected on the IO3 and IO4 Beamlines at Diamond Light Source in the following conditions: i) for IO3 3600 images, 0.1° Ω Oscillations, 25 or 40% transmission, 0.05 s exposure and wavelength at the platinum absorption edge of 0.8900 Å with an initial resolution limited to 2.5 Å and adjusted to 2.0 Å, and ii) for IO4 3600 images, 0.1°

Ω Oscillations, 17.34% transmission, 0.05 s exposure and wavelength at the platinum absorption edge of 0.91188 Å with an initial resolution limited to 2.0 Å and adjusted to 1.5 Å. The datasets autoprocessed by the Xia2-dials, Xia2-3dii and Autoproc ISPyB automated pipelines exhibiting anomalous signal correlation coefficients (CC_{anom}) values of 0.3 or more for the inner shell (over 5 Å of resolution) were selected for experimental phasing.

Data quality was assessed by analysing parameters as completeness (number of collected reflections against number of theoretically possible reflections for the given crystal symmetry), multiplicity (average number of measurements per symmetrically unique reflection), R_{meas} and $\langle I \rangle / \sigma I$ (which estimate dispersion of unique or merged intensities, respectively) and $CC_{1/2}$ (correlation between intensity estimates from half datasets and reflects the signal/noise ratio of the collected data) [133]. Representative electron density map views ($2F_0$-F_0, σ= 1.50) for each solved structure are shown in the **Appendix I** section.

2.1.12. Structure similarity search

Solved structures were submitted to the Dali server (http://ekhidna2.biocenter.helsinki.fi/dali/) for comparison with structures deposited in the PDB. Protein hits with a Z score of over 6 were considered and further analysed using UCSF Chimera and ChimeraX by comparing secondary structure, residue conservation and electrostatic surface analysis [134], [135].

2.1.13. Peptidoglycan docking modelling

Based on the evolutionary conservation of sequence and protein structures, PG binding was modelled by molecular docking using the AutoDock Vina software and HADDOCK server, independently [136], [137].

For AutoDock Vina (version 1.1.2), target protein and ligand coordinates are converted to PDBQT files with AutoDock Tools (version 1.5.4). AutoDock tools was additionally used to define polar hydrogens for protein structure and ligand files, setting the rotatable bonds in the PG ligands and defining the search box size for the protein receptor, which is centred in enzyme putative active sites. For initial docking protocols, all the amide bonds in the PG ligands were considered rotatable. Docking models were then run in AutoDock Vina considering each protein as a fixed structure while the ligand is rotated and translated to overlap the geometric centre of the ligand with the predicted pocket centres of the enzyme. Finally, refined models are selected based on

the most favourable binding affinity energy and RMSD calculated relative to the lowest energy model.

The HADDOCK2.2 easy interface approach was used by defining catalytic residues and evolutionary conserved residues of the enzyme as the 'active' residues for docking. Standard parameters were used by calculating 1,000 structures, of which 200 were further refined in water using 2,500 steps at 300 °K followed by 600 cooling steps.

2.2. LDTs

2.2.1. Bd0566

2.2.1.1. Crystallisation and cryoprotection

Crystallisation conditions for Bd0566 were screened at 15 mg/mL protein concentration using the JCSG-plus™, BCS™, ProPlex™ and MIDAS-plus™ (Molecular Dimensions©, cat# MD1-37, MD1-104, MD1-38 and MD1-106, respectively). The first condition found with 2.45 Å diffracting crystals corresponded to MIDAS-plus™ B1 (0.1 M sodium formate and 20% Sokalan® CP 45) which were cryoprotected with 20% ethylene glycol. The second 2.22 Å crystals were found after the 2.45 Å dataset in ProPlex™ F12 (0.1M Na•HEPES pH 7.0 and 1.5 M $(NH_4)_2SO_4$), which was cryoprotected with 20% ethylene glycol prior to collection.

2.2.1.2. Data collection and structure solving

For the first Bd0566 dataset, diffraction data was collected on the IO4 beamline under the previously described conditions for native collection.

The 2.45 Å dataset autoprocessed by the Xia2-3diii automated ISPyB pipeline was solved by Molecular Replacement (MR) in the $P6_5$ spacegroup using the reported structure for *E. coli* YcbB (PDB ID: 6NTW) as search model, which has an overall sequence identity of 28% with Bd0566 [138].

For solving the structure, the YcbB structure model was manually split in two fragments corresponding to two of the three reported domains of the protein: i) the LDT domain (6NTW Fragment 1) comprising residues 155-371, while omitting regions 172-175, 209-277 and 350-354, and ii) the PGBD domain (6NTW Fragment 2) from residues 214-266 and 315-366. A remaining alpha only region of 6NTW (denominated Fragment 3, spanning between residues 376-575) was not employed for MR. Fragments 1 and 2 were then used as search models in a tandem MR

search strategy with Phenix Phaser [139], searching first for a single copy of Fragment 1 (TFZ= 12.3 and LLG= 247.559), followed by searching a single copy of Fragment 2 (TFZ= 7.8 and LLG= 161.234), finding a single copy of the assembled molecule in the asymmetric unit. The obtained model was then refined by Phenix Refine employing Fragments 1, 2 and 3 as rigid bodies for the tracing of the scaffolding domain to be defined. Finally, the refined MR model was subjected to 15 cycles of auto building by Phenix Autobuild. A capping subdomain, as defined for YcbB, spanning from residue 295 to 355 was not traced by Phenix and had to be manually built in WinCoot and subjected to multiple cycles of building and refinement [140].

In the case of the 2.22 Å dataset, diffraction data was collected at the IO3 beamline under native conditions as previously indicated. For data solving, the resulting 2.45 Å structure of Bd0566 was used for MR and automated building in Phenix for the 2.22 Å dataset in the spacegroup I222, finding two copies of Bd0566 in the asymmetric unit (TFZ= 117.1, LLG= 20,104). The capping subdomain was not automatically traced by Phenix Autobuild and was also manually built and refined for this dataset using WinCoot and Phenix Refine, respectively.

2.2.2. Bd1075

2.2.2.1. Crystallisation and cryoprotection

Crystallisation conditions for Bd1075 were screened with the JCSG-plus™, BCS™ and MIDAS-plus™ (Molecular Dimensions©, cat# MD1-37, MD1-104 and MD1-106, respectively) at 25 mg/mL of protein concentration. Crystals were obtained in the BCS™ H8 condition (0.1M Tris pH 7.8, 0.1 M NaBr, 0.1M KSCN, 25% PEG Smear Broad) and cryoprotected with 25% ethylene glycol prior to collection.

2.2.2.2. Data collection and structure solving

Native datasets were collected at the IO3 beamline under conditions stated in section 2.1.11, autoprocessed by the Xia2-dials ISPyB automated pipeline at a maximum resolution of 1.55 Å. Nevertheless, attempts to solve the crystallographic structure of Bd1075 by MR were unsuccessful. Therefore, anomalous datasets at the bromide edge were collected for the same crystal, taking advantage of the co-crystallised halide.

Briefly, data was collected on the IO3 beamline, collecting 3600 images at 0.1° Ω Oscillations, 100% transmission, 0.04 s exposure and wavelength of 0.9100 Å with an initial resolution limited to 2.0 Å and adjusted to 1.7 Å. The dataset autoprocessed by the Xia2-3dii ISPyB automated

pipeline exhibiting an overall and outer shell anomalous CC values of 0.3 was selected for experimental phasing. The Bromide-phasing dataset was solved using CRANK2 on the CCP4 online software suite, with parameters set to default [141]. The Bd1075 bromide phased structure was solved in the $P2_1$ spacegroup, finding two copies of the enzyme in the asymmetric unit (ASU) (TFZ= 152.6, LLG= 31,852). Loops were manually built on WinCoot, and the resulting model was used for MR solution of the native dataset, which was subjected to multiple cycles of building and refinement via WinCoot and Phenix refine, respectively.

2.2.3. Bd3334

2.2.3.1. Crystallisation and cryoprotection

Bd3334 was screened for crystallisation at 30 mg/mL using the JCSG-plus™ and BCS™ screens (Molecular Dimensions©, cat# MD1-37 and MD1-104, respectively). Crystals were obtained in JCSG-plus™ condition B2 (20% PEG 3350 and 0.2 M NaSCN) and cryoprotected with mother liquor supplemented with 25% ethylene glycol prior to crystal collection.

2.2.3.2. Data collection and structure solving

Native datasets were collected at the IO3 beamline as described in 2.1.11. The collected dataset was autoprocessed by the Xia2-3dii ISPyB pipeline up to a maximum resolution of 1.70 Å.

The Bd3334 dataset was solved by MR in the $P2_12_12_1$ spacegroup by using the crystallographic structure of Bd1402 as search model (single Bd1402 chain, residues 125-163 and 170-264), finding a single copy of Bd3334 in the ASU (TFZ= 22.0, LLG= 354.896). The MR solution was auto built using Phenix Autobuild and subjected to multiple cycles of building and refinement in WinCoot and Phenix refine, respectively.

2.2.4. Bd3741

2.2.4.1. Crystallisation and cryoprotection

Crystallisation conditions for Bd3741 were screened with the JCSG-plus™, BCS™, ProPlex™ and MIDAS-plus™ (Molecular Dimensions©, cat# MD1-37, MD1-104, MD1-38 and MD1-106, respectively) at 10, 15, 20 and 35 mg/mL of protein concentration, obtaining isomorphous crystals in 45 of the 96 conditions of the BCS™ screen. Crystals were then cryoprotected with mother liquor supplemented with 30% of glycerol (for native data collection) or alternatively

phased with 10 mM K_2PtCl_4 and 30% glycerol and incubated for one hour at 18 °C prior to crystal collection.

2.2.4.2. Data collection and structure solving

Platinum-derived experimental phasing of the Bd3741 structure (PtSAD, platinum Single-wavelength Anomalous Diffraction dataset) was undertaken at the IO3 beamline under conditions indicated in section 2.1.11.

A platinum derivative dataset was collected for crystals obtained in condition D8 of BCS™ screen (0.1 M MES pH 6.2, 15% PEG Smear Broad, 0.15 M $CaCl_2 \bullet 2H_2O$ and 5% glycerol) and autoprocessed by the Xia2-dials ISPyB automated pipeline at a maximum resolution of 1.91 Å. This dataset was solved using the Autosol programme of the Phenix suite, defining dispersion coefficient (f') and absorption coefficient (f'') values of -7.012 and 12.06, respectively (estimated from http://www.bmsc.washington.edu/scatter/AS_form.html, applicable for all other phasing atoms), searching initially for seven platinum sites. The Bd3741 platinum-derivatized structure was then solved in the $P2_3$ spacegroup and subjected to five cycles of automated building in Phenix Autobuild. Finally, loops were manually built in WinCoot before the structure was used as a MR model for solving the native dataset.

The Autoproc autoprocessed native dataset for Bd3741 was collected for crystals grown in BCS™ screen condition C12 (0.1 M PIPES pH 7.0, 22.5% PEG Smear Medium, 0.1 M $CaCl_2 \bullet 2H_2O$ and 0.1 M $MgCl_2 \bullet 6H_2O$). This dataset was solved by MR using Phenix Phaser, employing the PtSAD determined structure as search model. The Bd3741 native structure was solved at a maximum resolution of 1.80 Å in the $P2_3$ spacegroup, finding one copy of the protein in the asymmetric unit (TFZ= 85.5, LLG= 11,211). Finally, loops were manually built in WinCoot, and the model was refined via successive cycles of Phenix Refine.

2.2.4.3. Phylogenetic analysis of LDTs

For the dendrogram analysis of LDTs, protein sequences were aligned using the MEGA11® software with the Clustal Omega algorithm. The aligned sequences were grouped using the Neighbor-Joining algorithm with 500 iterations for the generation of each dendrogram.

2.3. PG amidohydrolases

2.3.1. Bd0601 C94A

2.3.1.1. Crystallisation and cryoprotection

Bd0601 C94A was concentrated to 35 mg/mL and screened for crystallisation conditions, obtaining crystals in MIDAS-Plus™ and Morpheus™ commercial screens (Molecular Dimensions©, cat# MD1-106 and MD1-46, respectively). Two crystal forms were obtained for Bd0601 C94A, MIDAS-Plus™ condition G3 (40 % glycerol ethoxylate) and Morpheus™ condition F10 (Buffer system 3 pH 8.5, 0.12 M monosaccharides and 30% ethylene glycol- PEG 8000 precipitant mix). No cryoprotection was required prior to crystal collection.

2.3.1.2. Data collection and structure solving

Native datasets for Bd0601 C94A were collected at the IO3 beamline as described in 2.1.11. The MIDAS-Plus™ G3 dataset was autoprocessed by the Xia2-Multiplex pipeline to a maximum resolution of 3.95 Å, by merging two independent data collection experiments. This first collected dataset was solved by MR in the $P2_1$ spacegroup, using the *Legionella pneumophila* LPG0564 hypothetical protein (PDB ID: 2IM9, residues 75-89, 118-210 and 217-341) as a search model, finding two copies of the enzyme in the ASU (TFZ= 12.0, LLG= 141.107). The MR solution was subjected to automated building via Phenix Autobuild, and multiple cycles of manual building via WinCoot and refinement with Phenix refine.

The Morpheus™ F10 dataset was autoprocessed by the Autoproc ISPyB pipeline to a maximum resolution of 2.05 Å. This dataset was solved in the $P2_1$ spacegroup by MR using the abovementioned Bd0601 C94A model structure (TFZ= 55.3, LLG= 3,977). The obtained solution was finally subjected to multiple cycles of manual building and refinement using WinCoot and Phenix refine, respectively.

2.3.2. Bd1177 and Bd1177 C86A/Y124C/Y508C

2.3.2.1. Crystallisation and cryoprotection

Bd1177 (referred as Bd1177 proteolyzed C-terminus, after crystallographic data solving) was screened for crystallisation conditions at concentrations of 10, 15 and 20 mg/mL at 18 °C. Crystals were obtained in the JCSG-plus™, BCS™ and ProPlex™ commercial screens (Molecular Dimensions©, cat# MD1-37, MD1-104 and MD1-38, respectively). Crystals obtained at 20 mg/mL in condition F1 of the ProPlex™ commercial screen (0.1M Na•HEPES pH 7.0 and 20% PEG 8000)

were cryoprotected with mother liquor supplemented with 30% glycerol and harvested for native data collection. For structure determination, crystals were soaked in cryoprotective mother liquor supplemented with 10 mM K_2PtCl_4 for one hour at 18 °C before harvesting the crystals.

Bd1177 C86A/Y124C/Y508C catalytically null and disulphide-locked mutant were screened for crystallisation conditions at 15 and 20 mg/mL concentrations at 18 °C. Crystals were obtained only in condition A7 of the JCSG-plus™ commercial screen (Molecular Dimensions©, cat# MD1-37) containing 0.1M CHES pH 9.5 and 20% PEG 8000. Crystals were cryoprotected with mother liquor supplemented with 30% glycerol and harvested for data collection.

2.3.2.2. Data collection and structure solving

Native and platinum phased crystals were diffracted at the IO4 beamline under the respective conditions indicated in section 2.1.11 and processed by the ISPyB Xia2-3dii automated pipeline.

For Bd1177 platinum derivatised crystals, PtSAD data was solved using the programme Autosol of the Phenix suite, defining f' and f'' values of -7.030 and 12.08, respectively, at a beam wavelength of 0.8900 Å, and initially searching for seven anomalous atom sites in the unit cell. The final platinum phased model was solved in the $P4_32_12$ spacegroup finding one copy of Bd1177 C-terminus fragment the ASU. Furthermore, the obtained model was subjected to three automated building cycles in Phenix Autobuild. Loops and disulphide bonds were then built manually using WinCoot, and the model was further refined using Phenix Refine. The PtSAD model was used for MR solving of the native Bd1177 dataset at a maximum resolution 1.70 Å in the same crystallisation condition (TFZ= 49.8, LLG= 16,870), and further refined using Phenix Refine.

Bd1177 C86A/Y124C/Y508C triple mutant native datasets were collected at a beam wavelength of 0.9795 Å with parameters as indicated in section 2.1.11. Diffraction data was autoprocessed by the Xia2-3dii automated pipeline and solved by molecular replacement at 2.2 Å resolution in the P1 spacegroup with four copies of the protein in the ASU, using the Bd1177 C-terminus fragment as search model (TFZ= 31.8, LLG= 11,353). The MR model and initial electron density map were subjected to ten cycles of automated building and refinement in Phenix Autobuild. This model was then further manually built in WinCoot and refined using Phenix Refine.

2.4. Bd0675 cryptic protein

2.4.1. Crystallisation and cryoprotection

Crystallisation conditions for Bd0675 were screened in JCSG-plus™, BCS™, ProPlex™ and MIDAS-plus™ (Molecular Dimensions©, cat# MD1-37, MD1-104, MD1-38 and MD1-106, respectively) at 10, 15 and 20 mg/mL of protein concentration with no positive results. One condition with crystalline needles was found at 35 mg/mL in ProPlex™ H9 (0.1 M Imidazole pH 7.0 and 50% MPD). Based on this observation, Bd0675 was re-screened at 35 mg/mL in the Jena Biosciences© MPD and MPD/alcohol-based screens JBScreen Classic 7 and JBScreen Classic 8 (cat# CS-107L and CS-108L, respectively), obtaining needles and microcrystals in conditions D6 (50% MPD) for JBScreen Classic 7, and A1 (50% MPD, 15% ethanol and 10 mM sodium acetate) and A4 (55% MPD) for JBScreen Classic 8.

For optimisation, Bd0675 needles of condition JBScreen Classic 7 D6 were crushed as indicated for the MMS screening protocol (section 2.1.10) and used as seeds in 1:10 and 1:100 dilutions in the rescreening of Bd0675 in JBScreen Classic 7 and 8 screens. Crystals were obtained in most screening conditions for both 1:10 and 1:100 seed dilutions, which consisted mainly of conditions with > 30% MPD and thus did not require cryoprotection.

Additionally, since no model with significant sequence similarity to Bd0675 has been deposited in the Protein Data Bank (PDB), crystals were soaked in mother liquor supplemented with 10 mM K_2PtCl_4 for one hour at 18°C prior to crystal collection.

2.4.2. Data collection and structure solving

Native and platinum phased crystals were diffracted at the IO4 beamline under the respective conditions indicated in section 2.1.11.

The platinum phasing dataset was collected for crystals grown in JBScreen 7 condition D5 (47% MPD and 2% tert-butanol) using a 1:10 seed dilution and diffracted at a resolution of 1.64 Å and autoprocessed by the ISPyB Xia2-dials automated pipeline.

For data solving, PtSAD data were solved using the programme Autosol of the Phenix suite, defining f' and f'' values of -7.251 and 10.79, respectively, at a beam wavelength of 0.91188 Å, and initially searching for seven anomalous atom sites in the unit cell. The final platinum phased model was solved in the $C222_1$ spacegroup finding one copy of the protein in the asymmetric

unit. Furthermore, the obtained model was subjected to three automated building cycles in Phenix Autobuild. Loops and disulphide bonds were then built manually using WinCoot, and the model was further refined using Phenix.

Finally, the PtSAD obtained model was used for MR solution of the native dataset of Bd0675, at a maximum resolution 1.50 Å in the same crystallisation condition. The collected data was processed by the ISPyB Xia2-3dii automated pipeline and solved in the $C222_1$ spacegroup using Phenix Phaser by searching for one copy of Bd0675 in the asymmetric unit (TFZ= 51.7 and LLG= 10,692). Loops, disulphide bonds and ligands were then manually built in WinCoot and finally refined via successive Phenix Refine cycles.

2.4.3. Bd0675 electrophoretic mobility shift assay (EMSA) for non-specific ssDNA binding determination

For testing the possible single-stranded DNA (ssDNA) binding function of Bd0675, oligo dT nucleotides of 15, 25, 35, 45 and 55 nucleotides were synthesised by Sigma-Aldrich© and assayed by EMSA.

Briefly, 0 to 100 µM of Bd0675 were incubated with 100 ng of each ssDNA oligo probe in a buffer comprising 50 mM Tris pH 8.0, 100 mM NaCl and 5 mM of $MgCl_2$ and incubated at room temperature for one hour. Samples were then supplemented to a final glycerol concentration of 5% and loaded on 15% acrylamide-TAE gels, run at 60 V constant for 15 minutes and subsequently at 120 V until complete resolution of the DNA ladder. Finally, gels were stained with SafeView Nucleic Acid Stain® (NBS Biologicals™, cat# NBS-SV1) and visualised for electrophoretic mobility shift of the oligo probes.

CHAPTER 3

3. STRUCTURAL CHARACTERISATION OF *B. bacteriovorus* L,D-TRANSPEPTIDASES

3.1. Introduction

3.1.1. LDTs with possible transpeptidase or hydrolytic activity are present in *B. bacteriovorus* strain HD100

LDTs (EC 2.3.2.B14) catalyse stem peptide crosslinking, D-Ala exchange with other non-canonical D-amino acids (NCDAAs) or cell wall attachment to the outer membrane (OM), whereas L,D-endopeptidases (EC 3.4.19.11) and L,D-carboxypeptidases (EC 3.4.17.13) hydrolyse PG stem peptides [61], [69], [84], [82], [111], [131], [132]. These three types of enzymes are predicted to share a similar catalytic YkuD-like domain (PF03734) consisting of an eight-stranded β-barrel fold with a conserved Cys-His catalytic dyad [144].

Recently, a subgroup of YkuD-like domains has been re-classified as YkuD2 (PF13645). YkuD domains bear the catalytic cysteine-centred signature motif (S/T)XGCϕRϕ, whereas YkuD2 domains vary to SXGCXAϕ. In this regard, it has been proposed that the change of an arginine residue two positions after the catalytic cysteine for an alanine residue defines if the YkuD-like domain of the enzyme will act as a crosslinker/NCDAA exchanger or as a peptidase, respectively [118], [142]. This has been proven for LdtF (DpaA, formerly known as YafK), which is the only *E. coli* LDT with peptidase activity, hydrolysing the Lpp1-PG crosslink [84], [118], [142].

Of the 19 putative LDTs encoded by *B. bacteriovorus* strain HD100, only LDTs from group 5 are officially classified as YkuD2-like enzymes and thus are hypothesised to be hydrolases. Further, LDTs from group 1 also possess the activity-defining alanine residue while still classified as YkuD and could be suggested to act as hydrolases. Interestingly, the group 6 LDT, Bd1075, showcases a valine residue in the respective position. Bd1075 has been shown to exhibit carboxypeptidase activity is further discussed in sections 3.2.4.4 and 3.2.4.5 [125]. In contrast, LDTs of groups 2-4 possess a highly conserved arginine residue in the YkuD-like motif, suggesting that these enzymes act as LDTs and/or NCDAA exchangers.

Previous unpublished work of our group includes the determination of the crystallographic structures of Bd0553 (group 1), Bd0886 (group 3) and Bd1402 (group 5)

by Dr. Ian Cadby, which will be presented in the next introductory sections. Preliminary data also suggests that these three putative LDTs localise at the bdelloplast (unpublished).

This previous work along with the structural results obtained in this thesis described in this chapter will be contrasted in **Chapter 4** to analyse the diversification of the LDT family in *B. bacteriovorus* and their possible roles in predation.

3.1.2. *B. bacteriovorus* LDT group 1: LdtF (DpaA/YafK)-like putative LDTs

Group 1 of *B. bacteriovorus* HD100 comprises the LdtF-like LDTs Bd0553, Bd1763 and Bd2061. The structure of Bd0553 has been determined at a maximum resolution of 2.7 Å and is depicted in **Figure 18**.

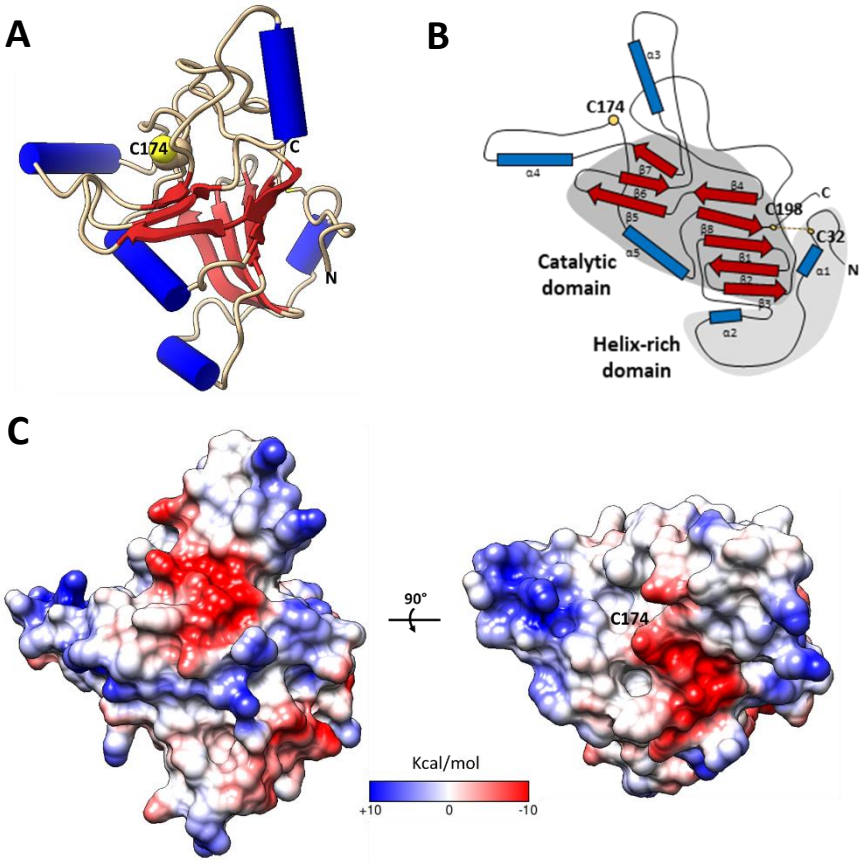


Figure 18: Bd0553 crystallographic structure and electrostatic surface potential. A) Cartoon representation of the 2.7 Å structure of Bd0553 coloured by secondary structure (α-helices in blue, β-strands in red). Bd0553 adopts a YkuD capped β-barrel fold. Catalytic cysteine C174 is shown in yellow sphere representation. **B)** Topology diagram of Bd0553. Two distinct domains are identifiable in the structure: the helix-rich domain (corresponding to the initially predicted coiled-coil domain) comprising alpha helices α1 and α2, and the YkuD-like LDT domain formed by alpha helix α5 and beta strands β1-8. Catalytic cysteine C174 is highlighted in yellow. The helix-rich domain is connected to the YkuD-like domain by the C32-C198 disulphide bond. **C)** Electrostatic surface potential of Bd0553 in a scale of +10 to -10 Kcal/mol in 90° angles apart dispositions. The catalytic C174 location is indicated. The putative substrate binding pocket of is limited by the basic alpha helix α4 and by an acidic patch formed by beta strands β4-6.

The Bd0553 solved structure traces 174 of the 183 residues of the mature protein (residues 21-204, respectively). Bd0553 displays the archetypical YkuD-like LDT capped β-barrel fold formed by alpha helix α5 and beta strands β1-8, wherein the catalytic cysteine C174 is located in the putative substrate binding cleft delimited by alpha helices α3 and α4, and beta strand β7 (**Figure 18B**).

In addition to the LDT catalytic core, Bd0553 possess a domain comprising residues 35-55 (including helices α1 and α2) which was predicted to form a coiled-coil fold based on its amino acid sequence. Interestingly, this portion of the structure does not adopt a coiled-coil domain fold, wherein the two four residues-long helices α1 and α2 are connected by an eight residues-long loop. Thus, this domain has been renamed as the helix-rich domain. The overall Bd0553 structure appears to be further stabilised by the disulphide bond formed between C32 and C198.

The electrostatic surface potential representation of Bd0553 is depicted in **Figure 18C** in two 90° apart dispositions. The Bd0553 putative substrate binding cleft (centred on residue C174) is delimited by the basic helix α4, which might participate in substrate binding, particularly, the stem peptide portion of the PG substrate, and by the acidic patch formed by beta strands β4-6.

Analysis of the predicted substrate binding site of Bd0553 shows a solvent-exposed pocket with the conserved catalytic dyad C174-H150, wherein the thiol group of C174 is located at 3.9 Å distance to $N^{\varepsilon 2}$ of H150 (**Figure 19A**). The imidazole ring of H150 is positioned by the 2.8 Å hydrogen bonding interaction between its $N^{\delta 1}$ and the backbone carboxylate group of G151.

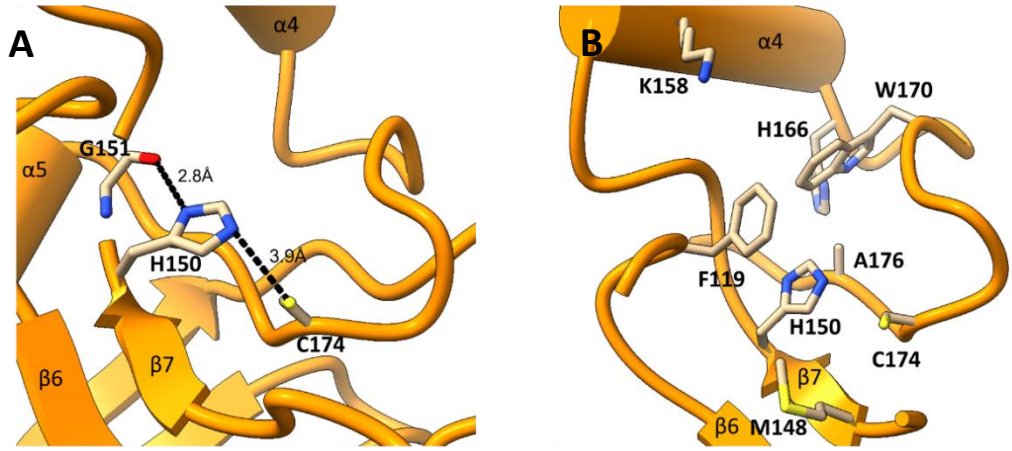


Figure 19: Bd0553 active site and substrate binding residues. A) Bd0553 active site. The catalytic dyad C174-H150 is located in a pocket formed between alpha helix α4 and beta strand β7. The imidazole ring of H150 is positioned by its interaction with the backbone carboxylate group of G151. **B)** Bd0553 proposed substrate binding residues. Residues in proximity to the C174-H150 catalytic dyad are shown. Residue A176 is part of the SXGCXAϕ motif and has been proposed as the defining feature of LDTs with peptidase activity. Residues H166, K158 and W170 are part of a structural insertion additional to the YkuD-like conserved core and are exclusive to group 1 of *B. bacteriovorus*' LDTs.

As illustrated in **Figure 19B**, aside from the catalytic dyad, Bd0553 substrate pocket potentially comprises residue K158 from alpha helix α4, M148 from beta strand β7, H166 and a hydrophobic pocket formed by residues F119, W170 and A176. Furthermore, A176 corresponds to the conserved alanine residue of the SXGCXAϕ motif proposed to define YkuD-like folds with peptidase activity [118], [142]. In fact, unpublished enzyme activity data of our group suggests that Bd0553 has carboxypeptidase and endopeptidase activity on PG stem peptides.

3.1.3. *B. bacteriovorus* LDT group 3: *B. bacteriovorus'* unique SH3b-LDTs

Group 3 of putative LDTs comprises Bd0886 and Bd1176, which are exclusive to *B. bacteriovorus*, and are predicted to possess a N-terminal SH3b domain. The crystallographic structure of Bd0886 has been determined at a maximum resolution of 2.39 Å and is depicted in **Figure 20A**.

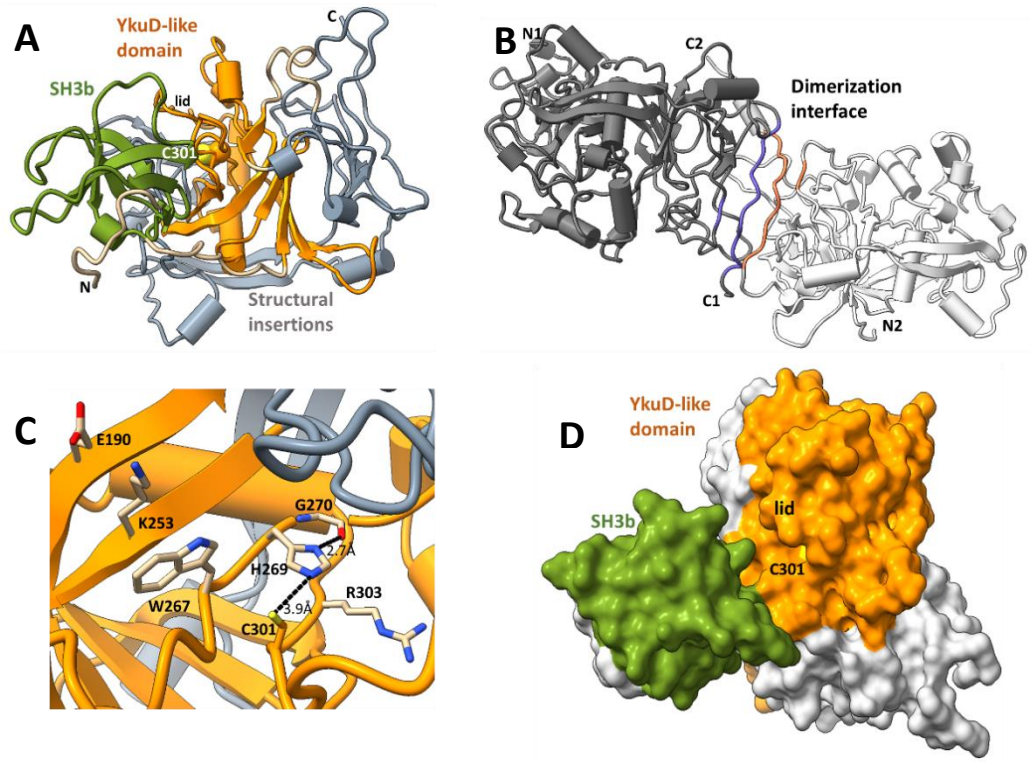


Figure 20: Bd0886 crystallographic structure, dimerization, active site, and surface representation. A) Cartoon representation of Bd0886 overall crystallographic structure. Bd0886 structure comprises a N-terminal SH3b domain (residues 39-121, in green) and the YkuD-like LDT domain (residues 131-193 and 250-323, in orange). Structural features unique for Bd0886 are shown in grey. Catalytic residue C301 is shown in yellow sphere representation. Active site access is proposed to be regulated by the region named "lid". **B)** Bd0886 possible homodimer structure (chain 1 in grey, chain 2 in white). Bd0886 forms a homodimer in the ASU wherein residues 478-486, which are predicted to be disordered in the three-dimensional structure, form an antiparallel β-sheet-like structure between both Bd0886 monomers (in purple and yellow for chains A and B, respectively). **C)** Bd0886 active site. The active site of Bd0886 comprises the conserved catalytic dyad C301-H269 (3.9Å distance), with the carboxylate of the G270 backbone orienting and polarising the imidazole ring of H269 (2.7Å distance). Sidechains of residues oriented towards the catalytic dyad are shown in stick representation (E190, K253, W267 and R303). Residue R303 corresponds to the arginine residue of the (S/T)XGCϕRϕ motif, proposed to be the residue defining LDTs with crosslinking or NCDAA-exchanging activity. **D)** Surface overview of Bd0886. Catalytic C301 and proposed "lid" are indicated. Surfaces of SH3b (in green) and LDT domains (in orange) are shown.

Bd0886 crystallographic structure depicts a homodimer in the ASU wherein chain A traces 449 of the 456 residues of the mature protein (residues 30-486), whereas chain B traces 361 residues (residues 125-486). Bd0886 chain B lacks traceable electron density for the SH3b domain, suggesting it is highly mobile in the crystal lattice or has been proteolyzed during the crystallisation process.

Bd0886 structure comprises a N-terminal SH3b domain (residues 39-121, in green), the YkuD-like LDT domain (residues 131-193 and 250-323, in orange) and unique structural

insertions (residues 194-249 and 324-486, in grey). The SH3b domain of Bd0886 is connected to the YkuD-like domain by a ten residues-long linker, forming a cleft between the SH3b domain and the Bd0886 putative substrate binding site, suggesting that the SH3b domain participates in substrate recognition and binding, as observed in *S. simulans* lysostaphin, *S. aureus* endolysins and *Listeria* phage endolysin, among others [122]–[124].

In the ASU, Bd0886 forms a tail-to-tail dimer mediated by residues 478-486 (**Figure 20B**, chain A in grey, chain B in white). These residues are part of the unique structural features of Bd0886 and are predicted to be disordered in the three-dimensional structure. Nevertheless, these residues form an antiparallel β-sheet like structure between each Bd0886 monomer (in purple for chain A and yellow for chain B, respectively), which could suggest that such crystallographic dimer is thermodynamically favourable.

Bd0886 active site is shown in **Figure 20C**, including the archetypical catalytic dyad C301-H269. The imidazole ring of H269 is oriented and polarised by its hydrogen bonding interaction with the backbone carbonyl group of G270. Residues oriented towards Bd0886 catalytic residues are highlighted, including E190, K253, W267 and R303, the latter being the conserved arginine of the proposed LDT crosslinkers and/or NCDAA-exchangers motif (S/T)XGCϕRϕ [118], [142]. Interestingly, as shown in **Figure 20D**, catalytic C301 is not solvent accessible due to the presence of a loop region defined as the active site "lid" (residues 280-299).

3.1.4. *B. bacteriovorus* LDT group 5: *B. bacteriovorus'* unique LDTs

Group 5 of LDTs is likewise unique to *B. bacteriovorus*, comprising Bd0782, Bd1402, Bd2940, Bd3093 and Bd3334. These LDTs are predicted YkuD2-like LDTs with the proposed hydrolytic motif SXGCXAϕ. The crystallographic structure of Bd1402 has been determined at a maximum resolution of 2.35Å and is depicted in **Figure 21A**.

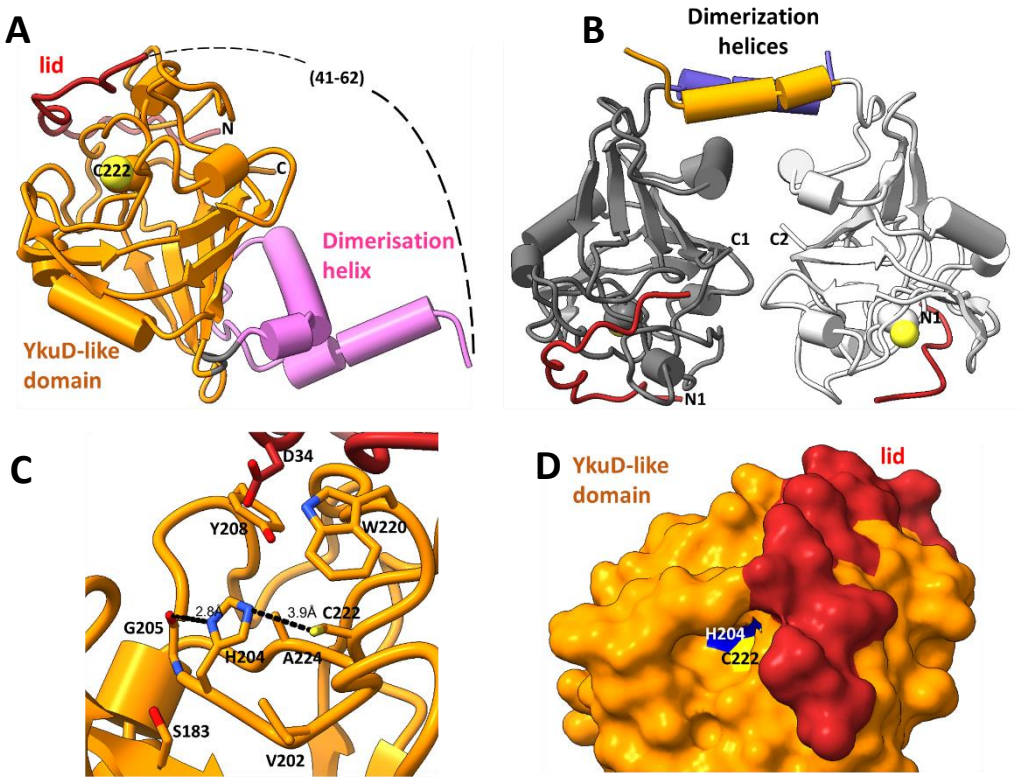


Figure 21: Bd1402 structure, dimerization, active site, and surface depiction. A) Cartoon representation of Bd1402 three-dimensional structure. Bd1402 structure comprises an N-terminal active site lid (residues 23-40, in red), a dimerization helical domain (residues 63-104, in pink) and the C-terminal YkuD2-like LDT domain (residues 108-248, in orange). The active site lid region is connected to the dimerization helix by a 21 residues-long linker not traceable in the experimental electron density map. Catalytic C222 is highlighted in yellow sphere representation. **B)** Bd1402 dimer representation. Chain A (in grey) and B (in white) associate via the amphipathic helix comprising residues 65 to 75 (in purple and yellow for chain A and B, respectively). Catalytic C222 and active site lid are highlighted in yellow and red, respectively. **C)** Bd1402 active site visualisation. The catalytic dyad of Bd1402 comprises residues C222 and H204, within a 3.9Å distance between the thiol group and the imidazole ring. H204 is oriented by its 2.8Å hydrogen bonding interaction with the G205 backbone carbonyl group. Residues S183, V202, Y208, W220 and A224 of the YkuD2-like LDT domain and D34 from the active site lid are proposed to be relevant for activity. Residue A224 corresponds to the hydrolytic LDT defining position in the SXGCXAϕ motif. **D)** Surface representation of Bd1402. YkuD2-like LDT domain of Bd1402 is shown in orange and the active site lid in red. Access to the active site appears to be blocked by the N-terminal active site lid of Bd1402.

Bd1402 three-dimensional structure traces residues 23-40 and 63-248 of the mature protein (residues 20-248). The Bd1402 fold comprises a N-terminal proposed active site lid (residues 23-40, in red) which is connected by a non-traceable 21 residues-long predicted loop to a helical dimerization domain (residues 63-104, in pink), and the YkuD2-like domain (residues 108-248, in orange).

As shown in **Figure 21B**, Bd1402 forms a dimer in the ASU. The dimeric state of Bd1402 has been observed in solution as per shown by preliminary gel filtration chromatography

experiments (unpublished data). Based on the crystallographic data, the dimer forms via the amphipathic helix spanning from residue 65 to 75 (in purple and yellow for chains A and B, respectively).

Bd1402 active site comprises the catalytic dyad C222-H204 at 3.9Å distance, wherein the imidazole ring is oriented and polarised by its hydrogen bonding interaction with the backbone carbonyl group of G205 (**Figure 21C**). Residues potentially participating in catalysis are highlighted, comprising S183, V202, Y208, W220 and A224 wherein the latter alanine residue is part of the proposed YkuD2-like hydrolytic motif SXGCXAϕ. Additionally, the active site lid residue D34 is shown to be oriented towards the active site and might participate in substrate selection. In fact, as shown in **Figure 21D**, the active site access is blocked by the proposed active site lid, and therefore its displacement is probably required for enzymatic activity.

Thus, to complete a panel of representative *B. bacteriovorus* LDTs, additional representative LDTs for each group have been crystallised and their structures determined. A summary of the LDTs tested for soluble expression, crystallisation and the respective structure attribution is shown in **Table 7**.

Table 7: Summary of LDTs tested for soluble expression, crystallisation and the corresponding structure attribution. S= soluble, I= insoluble, - =not applicable.

LDT	Solubility	Crystallised	Structure attribution
Bd0553	S	Yes	Dr. Ian Cadby
Bd3012	I	-	-
Bd3176	I	-	-
Bd3741	S	Yes	This thesis
Bd0886	S	Yes	Dr. Ian Cadby
Bd0566	S	Yes	This thesis
Bd1181	S	No	This thesis
Bd1358	I	-	-
Bd2908	I	-	-
Bd0782	S	No	This thesis
Bd1402	S	Yes	Dr. Ian Cadby
Bd2940	I	-	-
Bd3093	I	-	-
Bd3334	S	Yes	This thesis
Bd1075	S	Yes	This thesis

The structures of Bd3741 (group 2), Bd0566 (group 4), Bd3334 (group 5) and Bd1075 (group 6) are presented and analysed in this chapter.

3.2. Results

3.2.1. Group 2: Bd3741

3.2.1.1. Bd3741 expression, purification, and crystallisation

The ErfK/Ybis-like group of LDTs comprises proteins Bd3176 and Bd3741, that are insoluble when recombinantly expressed in *E. coli*. As an alternative, both proteins have been cloned as N-tagged MBP fusions, obtaining soluble protein for Bd3176 and Bd3741 expressed using AI media. However, after TEV overnight cleavage at 4°C of the MBP tag, only Bd3741 remained in solution and properly folded according to its SEC profiles and was further pursued for crystallisation and structure determination (**Figure 22A** and **22B** for purification and gel filtration profiles, respectively).

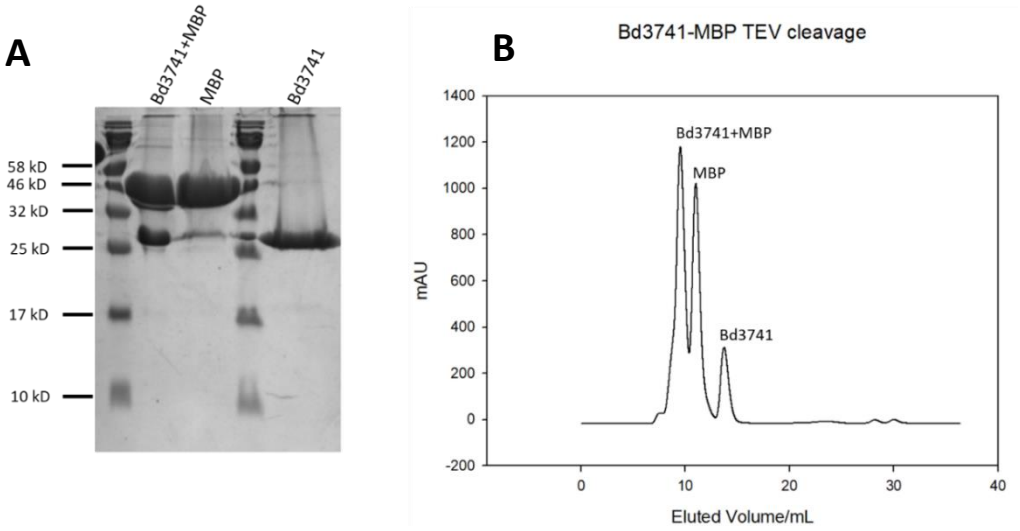


Figure 22: Bd3741-MBP purification. A) Bd3741-MBP was purified by maltose affinity chromatography and cleaved overnight by TEV digestion at 4°C. The cleaved samples were loaded onto a Superdex® 75 10/300 GL column to separate the MBP and Bd3741. **B)** Three peaks were observed in SEC that were analysed by SDS-PAGE, corresponding each to i) cleaved Bd3741 attached to the MBP, ii) MBP only and iii) Bd3741 only.

The Bd3741 SEC profile depicts three peaks corresponding to MBP only, Bd3741 only and Bd3741 cleaved but still associated to MBP (**Figure 22B**). The cleaved MBP and Bd3741 have been better separated by increasing the concentration of NaCl from 0.2 to 1.5 M during the SEC run and later dialysed against a final buffer comprised of 25 mM HEPES pH 7.5, 200 mM NaCl and 2.5 % glycerol.

Bd3741 was screened for crystallisation conditions at 10, 15, 20 and 35 mg/mL. Crystals were obtained in 46 of the 96 conditions of the BCS™ screen only at 35 mg/mL of

concentration, as shown in **Figure 23**, corresponding to condition A2 of the BCS™ screen (0.1 M phosphate/citrate pH 5.5 and 30% PEG smear low). All crystals had the same morphology and similar sizes across the screen.

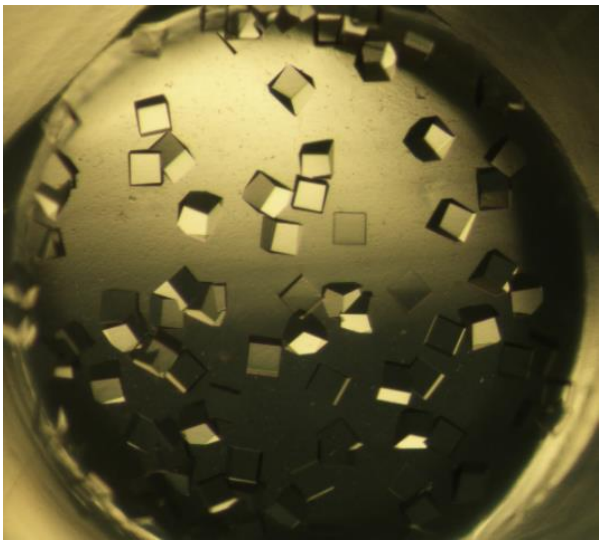


Figure 23: Bd3741 crystals. Bd3741 was concentrated to 35 mg/mL in a final buffer comprising 25 mM citrate pH 6.0, 200 mM NaCl and 2.5% glycerol and crystallised in the BCS™ screen (Molecular dimensions), obtaining isomorphous crystal in 46 of the 96 conditions. Representative crystals obtained in condition A2 of the BCS™ screen (0.1 M phosphate/citrate pH 5.5 and 30% PEG smear low).

3.2.1.2. Bd3741 data collection and structure solving

Diffraction experiments were carried out at the IO4 beamline as indicated in section 2.1.9, collecting the platinum phased and native Bd3741 datasets for crystallisation conditions D8 (0.1 M MES pH 6.2, 15% PEG Smear Broad, 0.15 M $CaCl_2 \bullet 2H_2O$ and 5% glycerol) and C12 (0.1 M PIPES pH 7.0, 22.5% PEG Smear Medium, 0.1 M $CaCl_2 \bullet 2H_2O$ and 0.1 M $MgCl_2 \bullet 6H_2O$) of the BCS™ screen, respectively.

Diffraction data was autoprocessed by the automated pipelines Xias2-dials for the PtSAD experiment and Autoproc for Bd3741 native crystals and used for structure solving. Specifically, the PtSAD dataset was solved using the programme Autosol of the Phenix suite and after one round of Autobuild, this model was used for MR solving of the native dataset using Phasing-MR. Statistical results for data collection, phasing and refinement are shown in **Table 8**.

Table 8: Data collection, phasing, and refinement statistics for Bd3741. Statistics for PtSAD and native Bd3741 crystallographic data are shown.

	Bd3741 PtSAD	Bd3741 Native
Data Collection		
Resolution range (Å)	54.59 - 1.91	66.81 - 1.80
Space group	P 2_3	P 2_3
Cell dimensions		
a, b, c (Å)	94.56, 94.56, 94.56	94.55, 94.55, 94.55
α, β, γ (°)	90, 90, 90	90, 90, 90
Total reflections	606045 (4996)	1022670 (97989)
Unique reflections	21854 (945)	26422 (2635)
Completeness (%)	98.60 (84.50)	99.97 (100.00)
Redundancy	27.7 (5.3)	38.7 (37.2)
<I>/σI	14.5 (0.3)	45.27 (3.23)
R_{meas}	0.129 (2.793)	0.06118 (0.3024)
$CC_{1/2}$	1.00 (0.60)	1 (0.971)
CC_{anom}	0.8 (-0.1)	
Refinement		
Maximum resolution (Å)		1.80
Number of reflections used		26420 (2635)
Number of reflections for R_{free}		1244 (131)
R_{work}		0.1893
R_{free}		0.2088
RMSD		
Bonds (Å)		0.012
Angles (°)		1.25
Number of non-hydrogen atoms		
Protein		1720
Ligand/ion		26
Water		198
B-factors		
Protein		34.38
Ligand/ion		42.34
Water		38.89
Ramachandran favoured (%)		97.13
Ramachandran allowed (%)		2.87
Ramachandran outliers (%)		0.00
Rotamer outliers (%)		6.08

*Values in parenthesis correspond to the highest-resolution shell

Eleven platinum sites were found for the PtSAD dataset, with a figure of merit (FOM) of 0.35. The density modified contour map for the heavy atom sites ($2F_0$-F_0, σ= 1.50) is shown in **Figure 24**.

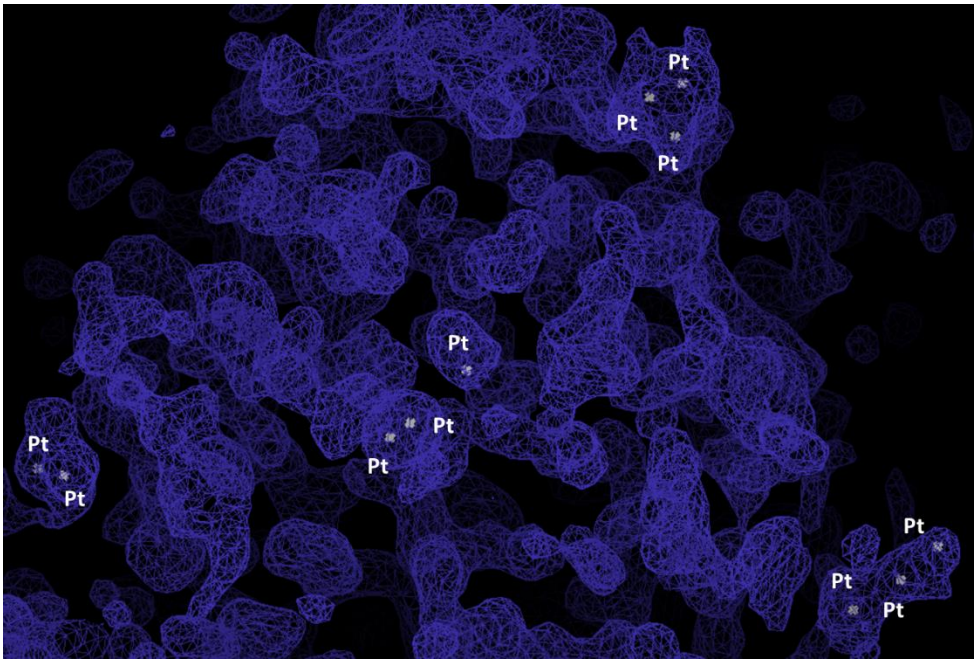


Figure 24: Heavy atom sites of the Bd3741 PtSAD dataset. Eleven heavy atom sites were found for the PtSAD Bd3741 dataset (white). The density modified map ($2F_0$-F_0, σ=1.50) is shown (blue mesh).

3.2.1.3. Bd3741 overall structure and domains

The crystallographic structure of Bd3741 is depicted in **Figure 25A**, covering 211 of the 223 residues of the mature protein sequence (excluding the 22 residues of the signal peptide, mature proteins residues numbered 23 to 246). Besides A23, residues 105 to 116 are missing from the model, which are presumably part of a flexible loop of the Bd3741 LDT catalytic domain. As depicted in **Figure 25B**, Bd3741 possesses the characteristic YkuD-like catalytic core described for most LDTs, which is formed by eight β-strands organised in a funnel-like β-barrel that is capped by an α-helix. The catalytic dyad formed by H165 and C190 is located on the outer side of the barrel forming a putative substrate binding groove. β-strand 6 works as a hinge connecting both sides of the LDT β-barrel.

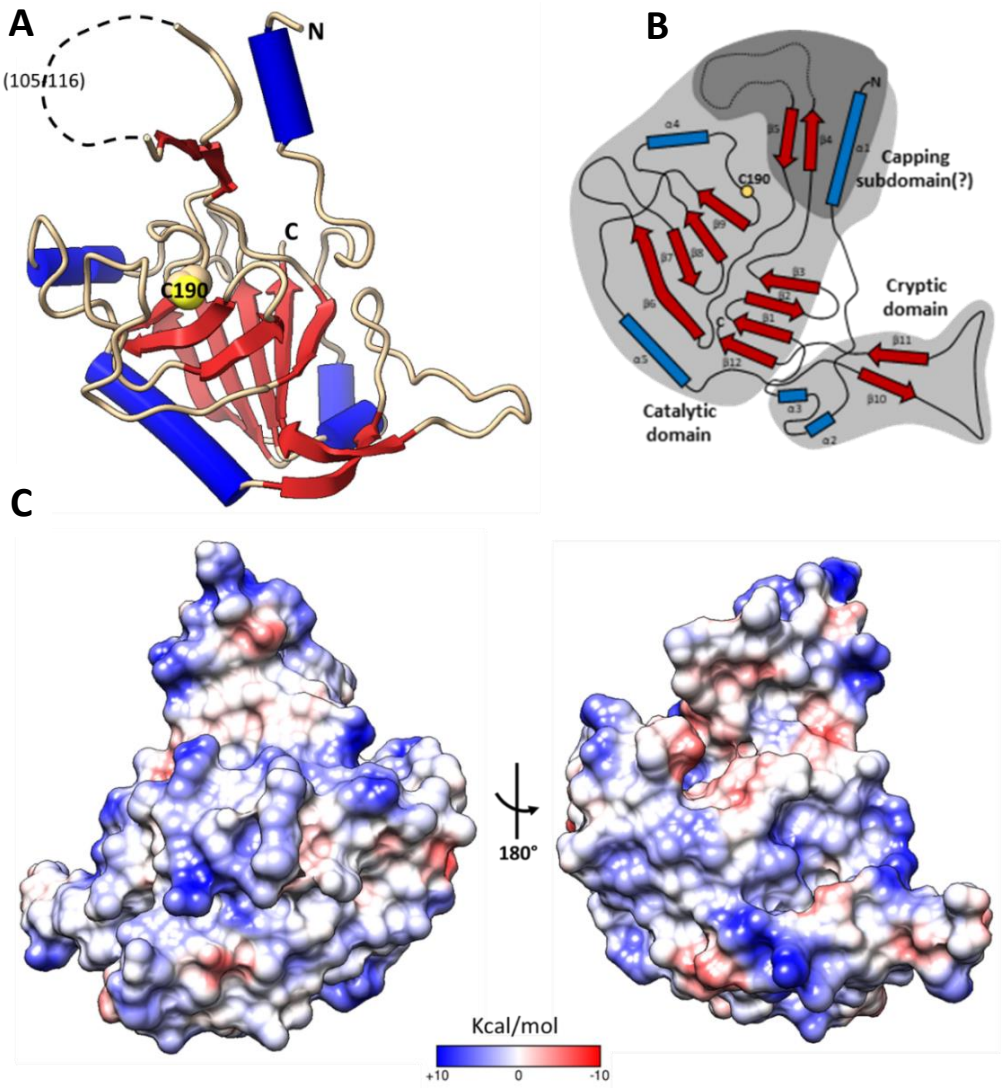


Figure 25: **Bd3741 crystallographic structure, topological organisation, and surface coulombic potential. A)** Bd3741 cartoon representation structure, wherein β-strands are coloured in red and α-helices in blue. N and C-terminal. The catalytic C190 is highlighted in yellow. A loop spanning from residue 105 to 116 is not traceable in the electron density map (represented with dashed lines). **B)** Topological and domain organisation of Bd3741, listing α-helices in blue from 1 to 5 and β-strands in red from 1 to 12. The catalytic C190 is highlighted in yellow in the diagram. Bd3741 is defined by two structural domains: the catalytic LDT domain, comprising the central β-funnel and the active site which is capped by a subdomain formed by α-helix 1 and a loop absent in the electron density; and a cryptic domain of unknown function. **C)** 180° angles apart views of Bd3741 coulombic potential surface representation from +10 to -10 Kcal/mol (blue to red, respectively).

In addition to the signature LDT catalytic core domain, two additional regions are distinguishable in the Bd3741 structure. First, a possible capping subdomain formed by the 11 residues of the missing flexible loop (residues 105 to 116) connecting β-strands 4 and 5, along with α-helix 1, which may interact with the aforementioned missing loop. As similarly presumed for Bd0566, this capping subdomain may determine substrate binding

and specificity for each particular LDT group. This hypothesis is supported by noticing that the positively charged α-helix 1, even though traceable in the electron density map, is less defined than the rest of the structure, suggesting a certain degree of flexibility of this portion of the protein. Moreover, the average B-factor values of the protein are around 20 Å^2 for most of the structure but as high as 140 Å^2 for α-helix 1 (**Figure 26**), and thus would form part of a hypothetical flexible capping subdomain might participate in substrate selection.

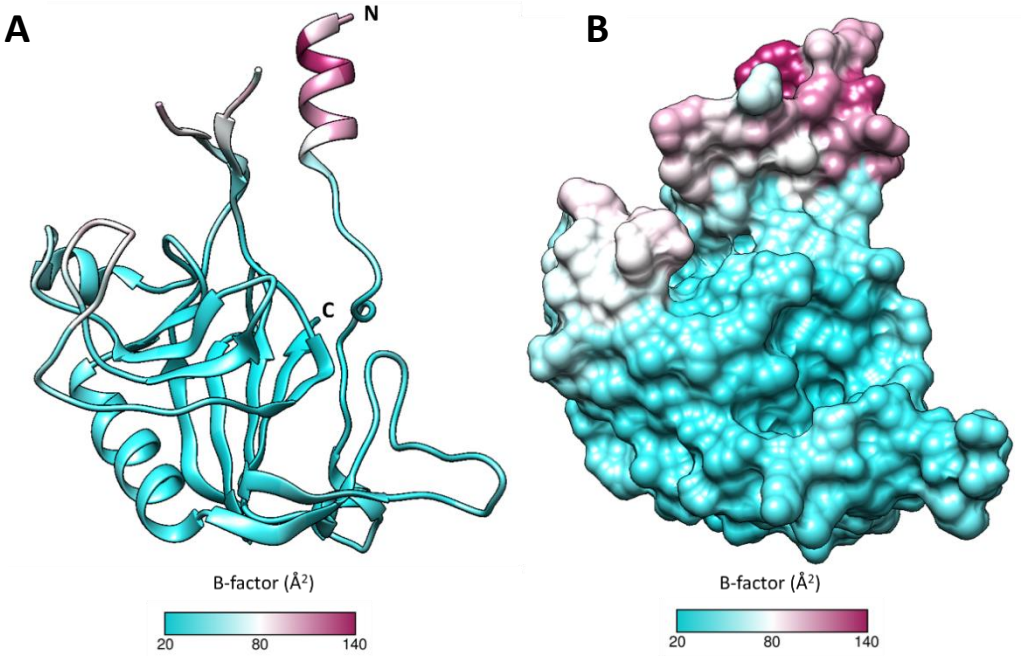


Figure 26: B-factor colouring of Bd3741. B-factor colouring from 20 to 140 Å^2 (cyan to magenta, respectively) of the ribbon (**A**) and surface (**B**) representations of Bd3741. B-factor analysis suggests that less defined electron density for Bd3741 is localised towards the capping subdomain (α-helix 1, β-strands 4 and 5 and the missing capping loop) and loops flanking the substrate binding cleft, suggestive of a conformational change in these elements upon substrate binding.

Secondly, a cryptic domain unique to Bd3741 is present in the structure. This unprecedented domain is formed by β-strands 10 and 11 connected by a triangular loop forming a hook-like structure, as shown in **Figure 25A**. The β-strands of this triangular hook region are positioned by its interaction with the two short α-helices 2 and 3.

As seen in the topological diagram in **Figure 25B**, the folding of Bd3741 probably is highly cooperative due to its re-entrant nature. In fact, the structural elements of each domain are spread discontinuously along the sequence: α-helix 1 (residues 24 to 33), most of the LDT catalytic core (residues 88 to 208) and β-strand 12 (residues 239 to 244) which closes

the funnel-like β-barrel of the catalytic domain; and α-helices 2 and 3 (residues 49 to 61) and β-strands 10 and 11 along with the triangular hook (residues 210 to 237) forming the cryptic domain. This topological arrangement suggests that catalysis and the unknown function of the cryptic domain are intrinsically linked.

Finally, inspection of coulombic surface charge of Bd3741 reveals that the protein surface is mostly positively charged favouring interaction of the enzyme with the negatively charged stem peptides of the PG cell wall, in accordance with what has been reported for most LDTs (**Figure 25C**).

3.2.1.4. Bd3741 catalytic domain and active site

The core of the catalytic domain of Bd3741 comprised the LDT β-funnel wherein residues of the catalytic dyad are localised in the loop connecting α-helix 4 with β-strand 9 for C190 and in β-strand 8 for H165 (**Figure 27A**).

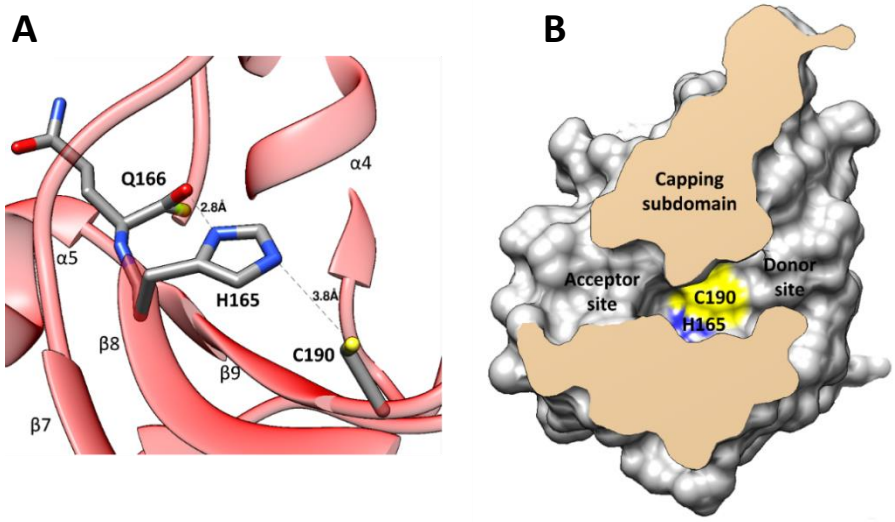


Figure 27: Bd3741 active site organisation. A) Bd3741 catalytic dyad comprises residues H165 and C190 which act as general base and nucleophile, respectively. The $N^{\varepsilon 2}$ of the H142 is positioned at 3.8 Å from the sulphur atom of C190 while the imidazole ring is further positioned by the interaction between $N^{\delta 1}$ of H165 with the backbone carbonyl of Q166 at 2.8 Å of distance. The secondary structure elements present in the active site are indicated. **B)** Top view of the surface representation of the Bd3741 active site. The surface of sidechains of the C190 (in yellow) and H165 (in blue) are located towards the acyl donor binding site (PG cell wall binding site), separated from the acyl acceptor site by the proposed capping subdomain.

In the crystalline structure, the $N^{\varepsilon 2}$ of H165 is positioned 3.8 Å from the sulphur atom of C190, polarising the residue for it to be a competent nucleophile. The polarised state of the imidazole ring of H165 is then stabilised by the backbone carboxyl group of Q166, which is situated at 2.8 Å from the $N^{\delta 1}$ of H165. As observed for Bd0566, the distance

between H165 and C190 is greater than the 2.5-3.5 Å range typical for hydrogen bonding, and therefore C190 is possibly re-oriented upon substrate binding.

Bd3741 is part of the predatory LDTs grouped as ErfK/Ybis related enzymes. In *E. coli*, the ErfK/Ybis LDTs are involved in anchoring the PG cell wall to the outer membrane by crosslinking the terminal lysine of the Braun lipoprotein (Lpp1) to the mDAP of the stem tetrapeptide by using the D-Ala at position 4 of the stem peptide as leaving group [61], [84]. Thus, the PG tetrapeptide acts as acyl donor while the Braun lipoprotein acts as acyl acceptor during catalysis, yielding a PG tripeptide crosslinked to the C-terminal of Lpp1 as product.

Surface analysis of Bd3741 defines two sites for ligand binding (**Figure 27B**). First, a donor site which is hypothesised to be occupied by the PG tetrapeptide, is defined by the common features previously reported by LDTs, including the catalytic dyad C190-H165 and the oxyanion hole formed by the backbone nitrogen of A163 and C190 that potentially stabilised the oxycarbanion tetrahedral intermediate previous to the acyl-enzyme intermediate formation.

Based on the exposed amino acid sidechains on the LDT substrate binding cleft, possible substrate binding residues are highlighted in **Figure 28**. Considering the orientation of C190, residues R135, H137, Y141, S143, M150, A153, F155 and E158 are proposed to be part of the acyl donor site of Bd3741. On the other hand, S146, F148, P168, L171, S188 and R192 are located on what is putatively assigned as the acyl acceptor site of the LDTs. Residue R192 is conserved for the (S/T)XGCϕRϕ motif, suggesting that Bd3741 is part of the YkuD-like LDT crosslinkers/NCDAA exchangers [118], [142].

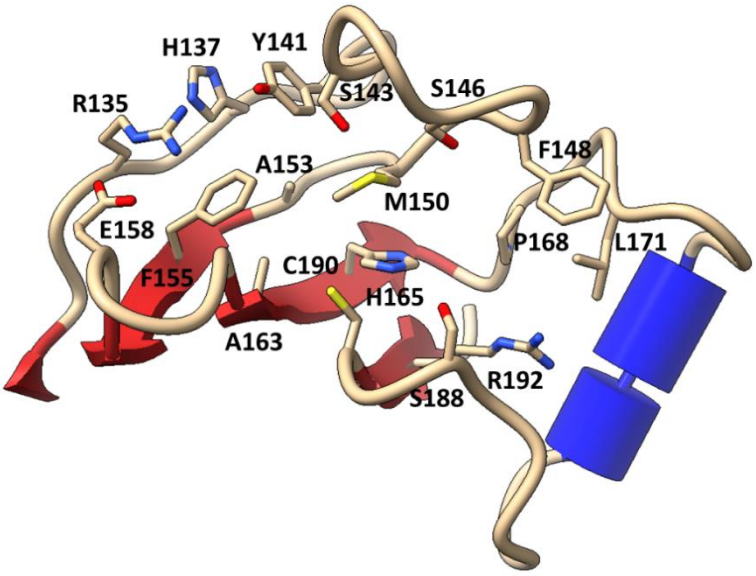


Figure 28: Bd3741 substrate binding residue identification. Residues in proximity to the C190-H165 catalytic dyad are shown in stick representation. R135, H137, Y141, S143, M150, A153, F155 and E158 are proposed to be part of the acyl donor site whereas S146, F148, P168, L171, S188 and R192 are located on what is putatively assigned as the acyl acceptor site. Residue R192 is part of the (S/T)XGCϕRϕ crosslinker/exchanging defining motif.

To probe the conservation of the proposed substrate binding residues of Bd3741, the structure was superimposed with the available AlphaFold2 in silico model of *E. coli* ErfK (pLDDT= 93.5) [61], [89], [145]. Superimposition was limited to the LDT catalytic core due to ErfK and Bd3741 possessing accessory divergent domains with an RMSD of 0.88 Å (**Figure 29A**).

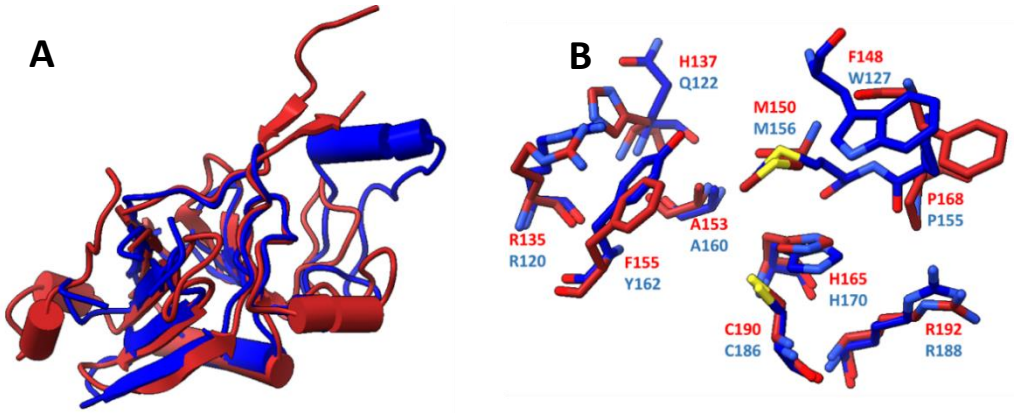


Figure 29: Structural conservation of Bd3741 active site. A) Bd3741 and *E. coli* ErfK structural comparison. Bd3741 crystal structure (in red) and an ErfK AlphaFold2 model (in blue) were superimposed with an overall RMSD of 0.88Å for the YkuD-like domains of each protein. **B)** Conserved active site residues between Bd3741 (in red) and ErfK (in blue) are shown in stick representation.

The architecture of the LDT catalytic core is unsurprisingly conserved between both proteins, however, the re-entrant topology of the domain is unique to Bd3741. Both LDT cores show possible capping regions, however in different positions and with different secondary structure contents. For the ErfK model, the capping portion of the domain is a α-helix loop connecting the β-strands corresponding to 6 and 7 (in Bd3741 coordinates) of the LDT core, in contrast to the β-strand loop connecting β-strands 3 and 6 in Bd3741.

Comparison of the active site sidechains between ErfK and Bd3741 reveals a similar arrangement of the catalytic dyad between both proteins (see **Figure 29B**). Nevertheless, of the proposed Bd3741 substrate binding residues only R135, H137, F148, M150, A153, F155, P168 and R192 are conserved in ErfK, which suggests that they are critical in substrate binding either at the acyl donor or acceptor sites.

3.2.1.5. Bd3741 cryptic triangular hook insertion

Bd3741 depicts a triangular hook insertion between residues 209 and 238. As shown in **Figure 30A**, the three-dimensional disposition of the triangular loop is maintained by hydrogen bonding between conserved residues within the loop and by hydrophobic interactions with other secondary structure elements of the putative cryptic domain. Particularly, residue L216 of the loop interacts with L59 of the α-helix 3 while Y58 of α-helix 3 interacts with both P223 and Y225 of the triangular loop, forming a hydrophobic patch. On the other hand, the three-dimensional geometry of the loop is maintained by intra-loop interaction between residues Q227-N228, P223-Y225-V231 and D217-R219. Further, residues G221 and G229 are located in the curvatures in the corners of the loop, allowing it to acquire the angles to form a triangular geometry.

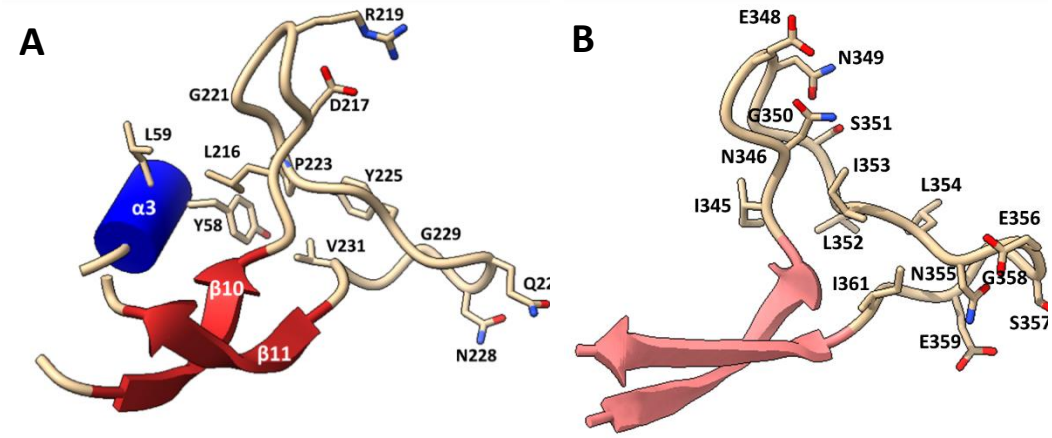


Figure 30: **Triangular loop of Bd3741 cryptic domain**. **A)** The main feature of the Bd3741 cryptic domain corresponds to a triangular hook spanning from residues 209 to 238. Sidechains of residues responsible for positioning the loop and maintaining the triangular geometry are indicated. **B)** Triangular loop of Ypr118w (residues 339-366) and residues possibly involved in adopting this fold are shown. This domain has been shown to participate in the dimerization of Ypr118w.

To propose a function for this domain, similar structures to Bd3741 were searched in the PDB using the Dali server, finding LDTs of different organisms. Even though most of the significant matches share the central LDT core domain with Bd3741, none of the found structure showed any of the accessory features of this protein (data not shown).

Remarkably, the methylthioribose-1-phosphate isomerase from *S. cerevisiae* Ypr118w (PDB ID: 1W2W) adopts a similar fold between residues 339 and 366. The triangular conformation of the loop is possibly achieved by the interaction of residues N346, E348, N349, E356, N355 and E359, in addition to the location of residues G350 and G358 in the corners of the loop, similarly to Bd3741 (**Figure 30B**). This region of the enzyme is unrelated by sequence to Bd3741 and has been shown to participate in the formation of the Ypr118w homodimer structure [146]. Nonetheless, the monomeric state of Bd3741 suggests a repurposing of this structural feature. Therefore, additional experiments are required to address the function of this cryptic domain and how it specifies the activity of this group of LDTs.

3.2.2. Group 4: Bd0566

3.2.2.1. Bd0566 purification and crystallisation

Bd0566 mature protein sequence (residues 22-505, omitting the 21 residues signal peptide sequence) was fused to a C-terminal His_6-tag, expressed in AI media, and purified by nickel IMAC followed by SEC on a Superdex® 75 26/60 column (**Figure 31A**), as

indicated in Materials and Methods. Bd0566 was concentrated to 20 mg/mL by centrifugal ultrafiltration in a final buffer comprised by 25 mM tri-sodium citrate pH 6.0, 200 mM NaCl and 2% glycerol. From the SEC profile, Bd0566 is monomeric in solution (**Figure 31B**) and requires the supplementation of 5% ethylene glycol as cryoprotectant to be stored frozen at -80°C and avoid precipitation after thawing.

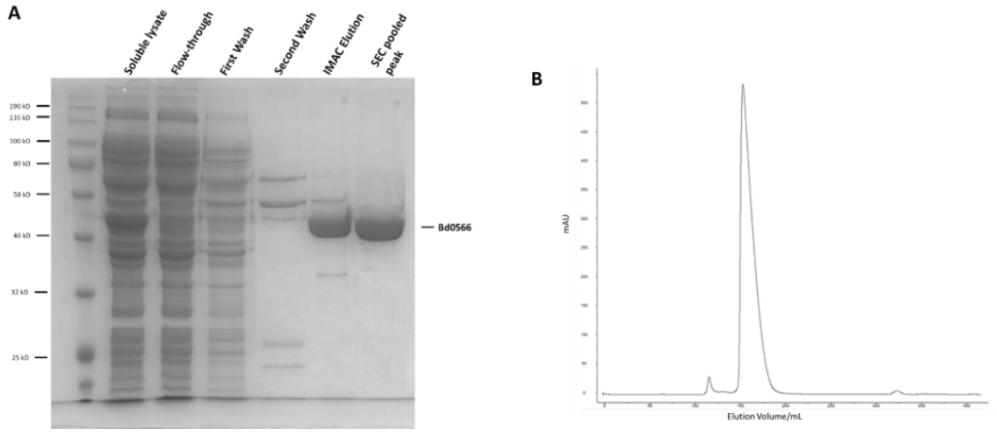


Figure 31: Bd0566 purification. Bd0566 was expressed in AI media and purified by nickel IMAC and SEC. **A)** 15% SDS-PAGE of the purified Bd0566. First and second wash contain 20 mM and 50 mM of imidazole, respectively. **B)** SEC profile of Bd0566 showing a monomeric peak for the protein. The run was performed at room temperature and flow rate of 2 mL/minute in a buffer composed by 25 mM tri-sodium citrate pH 6.0, 200 mM NaCl and 2% glycerol. The protein was concentrated to 20 mg/mL and supplemented with 5% ethylene glycol before storage at -80°C.

Bd0566 was screened for crystallisation conditions obtaining the best diffracting crystals in the Proplex® and MidasPlus® suites (Molecular Dimensions©). Specifically, crystals used for data collection were grown in MidasPlus® B1 (0.1 M sodium formate and 20% Sokalan® CP 45) and Proplex® condition F12 (0.1 M Na•HEPES pH 7.0 and 1.5 M $(NH_4)_2SO_4$), obtaining a 2.45 and 2.22 Å datasets, respectively (**Figure 32A and 32B**, respectively).

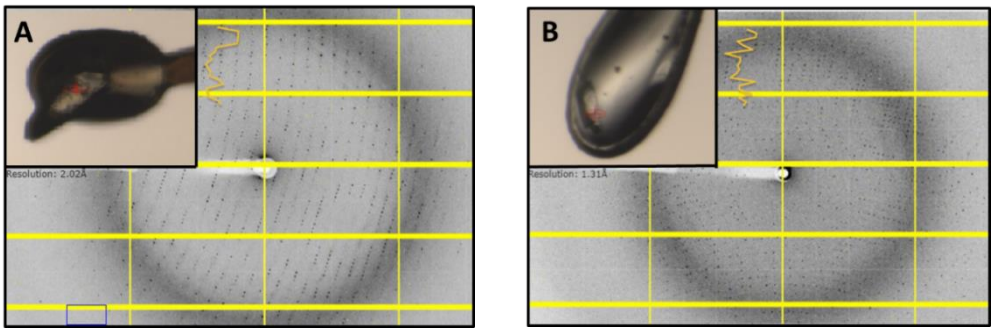


Figure 32: Crystal and diffraction pattern for the two Bd0566 datasets. **A)** Crystal and diffraction pattern obtained at a resolution of 2.45 Å; the spacegroup for this dataset is $P6_5$. **B)** Crystal and diffraction pattern for the 2.22 Å dataset; the spacegroup for this dataset is I 222.

3.2.2.2. Bd0566 data collection and structure solving

Data was autoprocessed by the Xia2-3dii automated ISPyB pipeline for both datasets and solved by Molecular Replacement (MR) using the reported structure for the *E. coli* YcbB (PDB ID: 6NTW) [83], which has an overall sequence identity of 28% with Bd0566.

The first dataset obtained for Bd0566 was the 2.45 Å resolution form and was solved by using a two fragments ensemble of *E. coli* YcbB. The final crystallographic structure traced 481 of the 483 residues of the mature protein (residues 24 to 505).

The resulting 2.45 Å structure for Bd0566 was used for MR solving of the 2.22 Å dataset in the I 222 spacegroup and built in Phenix Autobuild, finding two copies of Bd0566 in the ASU. The subdomain comprising residues 316-376 was manually built, obtaining an equivalent solved structure for both datasets. Statistics for data collection and refinement are shown in **Table 9**.

Considering the higher resolution of the second Bd0566 crystal form, subsequent structural analysis will refer to this crystallographic structure of the protein.

Table 9: Data collection, phasing, and refinement statistics for Bd0566. Statistics for data collection and refinement of Bd0566 crystal forms 1 and 2, obtained at 2.45 and 2.22 Å, respectively.

	Bd0566 crystal form 1	Bd0566 crystal form 2
Data Collection		
Resolution range (Å)	38.07 - 2.45	93.56 – 2.22
Space group	P 6_5	I 222
Cell dimensions		
a, b, c (Å)	157.17, 157.17, 56.65	110.4, 150.0, 187.1
α, β, γ (°)	90, 90, 120	90, 90, 90
Total reflections	861163 (84568)	998204 (87663)
Unique reflections	43652 (4338)	75983 (7287)
Completeness (%)	98.29 (99.88)	99.05 (93.75)
Redundancy	19.7 (19.5)	13.1 (12.0)
<I>/σI	12.00 (1.36)	8.19 (3.52)
R_{meas}	0.1245 (3.975)	0.7102 (10.96)
R_{pim}	0.02834 (0.8978)	0.205 (3.408)
$CC_{1/2}$	0.998 (0.757)	0.831 (0.34)
Refinement		
Maximum resolution (Å)	2.45	2.22
Number of reflections used	42960 (4333)	75532 (7063)
Number of reflections for R_{free}	2166 (184)	3777 (338)
R_{work}	0.1821	0.1874
R_{free}	0.2147	0.2239
RMSD		
Bonds (Å)	0.0086	0.010
Angles (°)	0.956	1.07
Number of non-hydrogen atoms		
Protein	3967	7937
Ligand/ion	16	80
Water	34	431
B-factors		
Protein	82.20	41.96
Ligand/ion	84.84	55.88
Water	65.16	42.56
Ramachandran favoured (%)	96.06	98.24
Ramachandran allowed (%)	3.94	1.76
Ramachandran outliers (%)	0.00	0.00
Rotamer outliers (%)	0.00	0.57

*Values in parenthesis correspond to the highest-resolution shell

3.2.2.3. Bd0566 overall structure and domains

Bd0566 crystallographic structure grossly resembles the overall fold of the *E. coli* YcbB (PDB ID: 6NTW), composed of three domains: a helical scaffolding domain, a putative PG binding domain (PGBD) and the catalytic domain or LDT domain, as shown in **Figure 33A** and **B**.

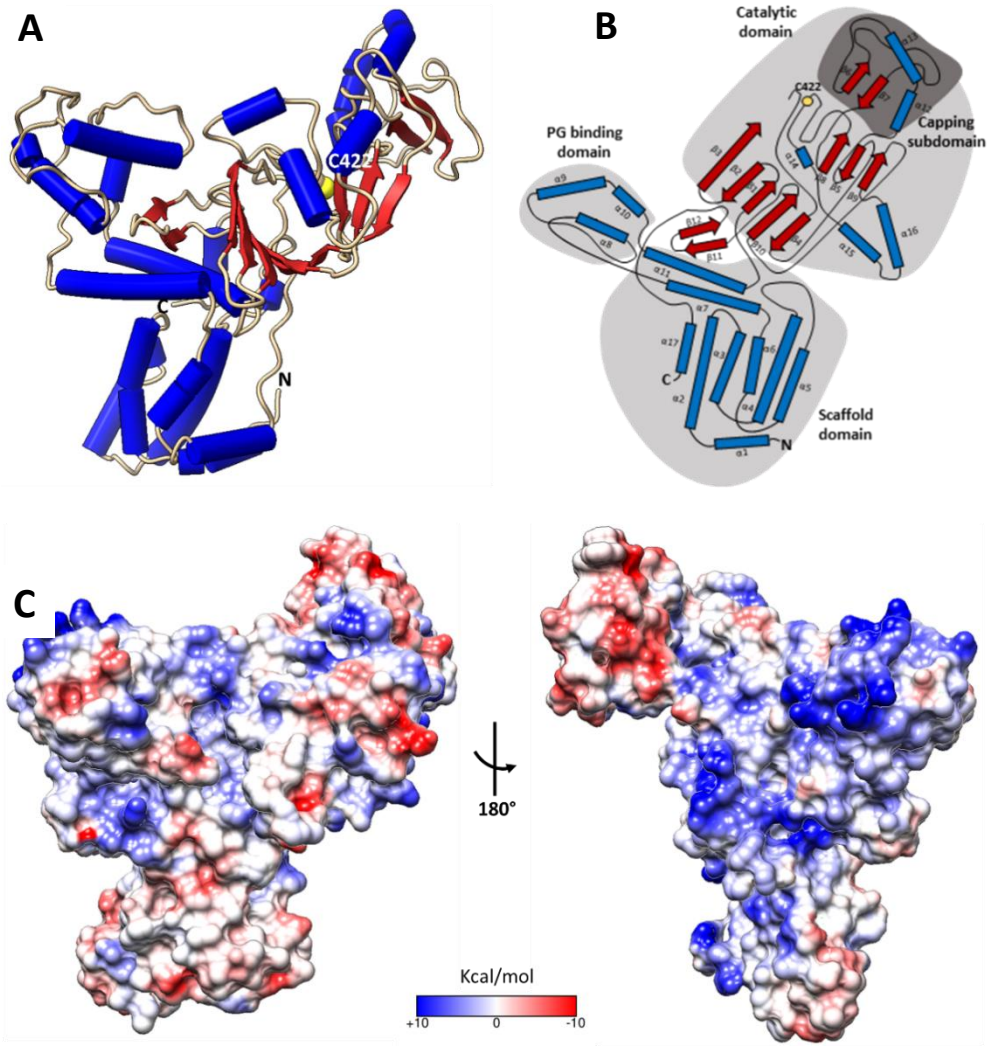


Figure 33: Bd0566 structure and coulombic surface potential. A) Cartoon depiction of the crystallographic structure of Bd0566. α-helix are depicted in blue, and β-strands are in red. Catalytic cysteine C422 is highlighted in yellow **B)** Topology diagram of Bd0566. Domains and the capping subdomain are indicated, α-helix and β-strands numbered from 1-17 and 1-12, respectively. The catalytic C422 is highlighted in yellow. **C)** Coulombic surface potential of Bd0566 in a scale of +10 to -10 Kcal/mol in 90° angles apart dispositions.

Bd0566 topology is re-entrant, wherein the scaffolding domain spans between residues 22-157, 237-253 and 496-505, whereas the PGBD and the LDT are independent and comprised by residues 177-234 and 261-475, respectively.

The scaffolding domain of Bd0566 is formed by nine α-helices with fold that differs from the ten α-helices of the YcbB scaffolding domain [83]. The core of the scaffolding domain is a three α-helical bundle formed by helices 2-4 which is stabilised by hydrophobic interactions with α-helix 1 and the perpendicular arrangement of α-helices 5 and 6. The scaffolding domain core is covered by the disulphide-stabilised crossed helices 7 and 11,

and by the re-entrant α-helix 17. The disulphide bond between helices 7 and 11 is formed by residues C155 and C237, respectively (which replaces the C229-Q354 pair present in YcbB), connecting this domain with the PGBD.

The PGBD displays a structure of a closed three helical bundle with a left-handed twist motif that is conserved in a variety of PG modifying proteins, mainly amidases and phage autolysins, as well as for YcbB. This domain differs from the lysine PG binding motifs (LysM) present in many structurally characterised bacterial LDTs which is formed by two antiparallel β-strands connected by two or more short α-helices, or to the open β-barrel fold of bacterial SH3b domains, which have been proved to bind PG by the stem peptide portion of the molecule [123], [124].

Finally, the LDT catalytic domain of Bd0566 displays the archetypical β-barrel architecture capped by two consecutive α-helices, wherein the active site is exposed to the solvent and located on the side of the core β-barrel. Additionally, Bd0566 possesses an accessory capping subdomain that could participate in substrate selection. This capping subdomain is formed by α-helices 12 and 13 and β-strands 6 and 7, which are interconnected by long loops of around 15 residues, defining a cleft between the β-barrel core architecture and the capping region where catalysis occurs. Additionally, β-strands 11 and 12 form a β-hairpin connecting the catalytic domain of the enzyme with the re-entrant helix 17 of the scaffolding domain and helices 7 and 11.

Coulombic surface potential analysis of the Bd0566 surface depicts a mostly positive surface potential distribution, which is in accordance with PG binding proteins and LDTs in particular, suggestive of an extensive interaction of the protein with the mostly negatively charged PG cell wall stem peptides (**Figure 33C**). However, this surface potential distribution is negative in the capping subdomain of Bd0566 in contrast to the positively charged capping subdomain of YcbB, which suggests that the substrate specificity and/or recognition mechanisms of both enzymes differ.

3.2.2.4. Bd0566 catalytic domain and active site

The catalytic core domain of the enzyme adopts the signature YkuD-like transpeptidase core, wherein the β-strands of the central barrel are curved forming a funnel-like structure which is further capped by two α-helices (**Figure 34A**) [83], [144]. Specifically, β-strands 1 and 2 interact with the β-hairpin formed by strands 11 and 12, while β-strands 5, 8 and 9

form part of the substrate binding cleft in the active site. Both sides of the central barrel are connected by β-strands 3 and 4, which may transmit conformational changes between the active site and the rest of the protein.

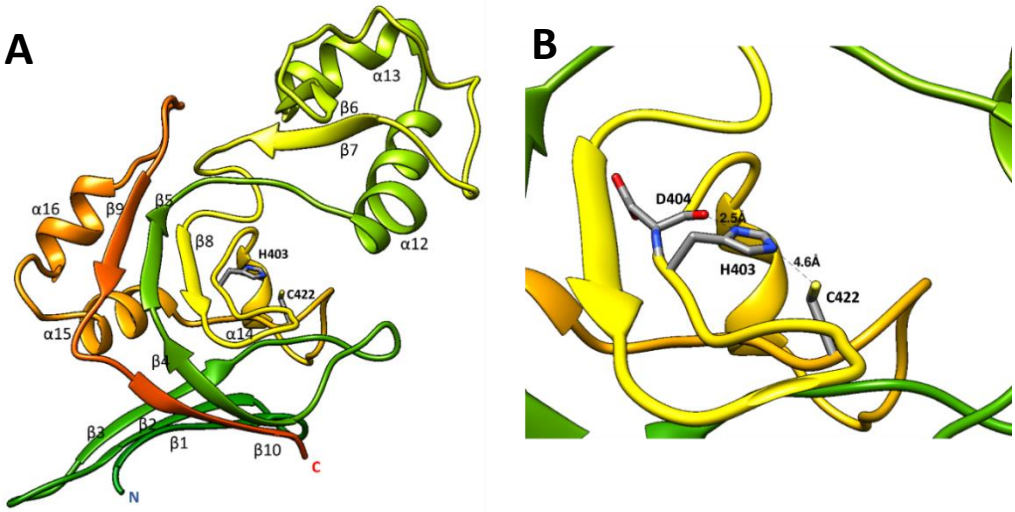


Figure 34: Bd0566 catalytic domain and active site. A) The LDT catalytic domain of Bd0566 (residues 261 to 475) are depicted in ribbon representation and coloured by Jones' rainbow, highlighting sidechains of the catalytic residues H03 and C422, located in the binding cleft defined by the core of the domain and the capping subdomain. **B)** Zoom-in of the active site of Bd0566. The nucleophilic residue C422 is at a distance of 4.6 Å of the $N^{\varepsilon 2}$ of H403. The $N^{\delta 1}$ of H403 is at hydrogen bonding distance of 2.5 Å from the carbonyl backbone of D404, stabilising and position H403 to act as general base.

The Bd0566 substrate binding cleft sits in the interface between the side of the barrel formed by β-strands 5, 8 and 9 and α-helix 12 of the capping subdomain. The active site of the protein displays an archetypical catalytic LDT dyad formed by the general base H403 and the nucleophile C422, catalysing the ping-pong 3-3 PG crosslinking reaction (**Figure 34B**). More precisely, the $N^{\varepsilon 2}$ of H03, which is positioned at 4.6 Å from the thiol group of C422, polarises the sulfhydryl group of C422 for it to be a competent nucleophile. The polarising imidazole ring of H403 is positioned and stabilised by hydrogen bonding of the $N^{\delta 1}$ of H403 with the backbone carboxyl of D404 at 2.5 Å distance. It is worth noting that the H403-C422 distance is longer than the standard hydrogen bond length of 1.5-2.5 Å and therefore, for deprotonation of C422 to occur, the active site must rearrange upon substrate binding in the acyl donor site, probably by repositioning of the C422 sidechain.

Bd0566 active site is flanked by a capping subdomain spanning from residues 316 to 376, formed by α-helices 12 and 13 and β-strands 6 and 7, which is proposed to participate in substrate selection. Specifically, α-helices 12 and 13 are perpendicularly arranged and

sandwiched against an antiparallel β-sheet formed by β-strands 6 and 7 (**Figure 35A**). The capping subdomain of Bd0566 divides the substrate binding cleft in two sites: the acyl donor site and the acyl acceptor site. Both sites of the substrate binding cleft are communicated by a polar channel formed by the interaction between the capping subdomain and the catalytic LDT core. This interaction is mainly comprised by the 2.7 Å hydrogen bonding interaction between residues E323 of the capping subdomain and R296 of the rest of the catalytic domain (**Figure 35B**).

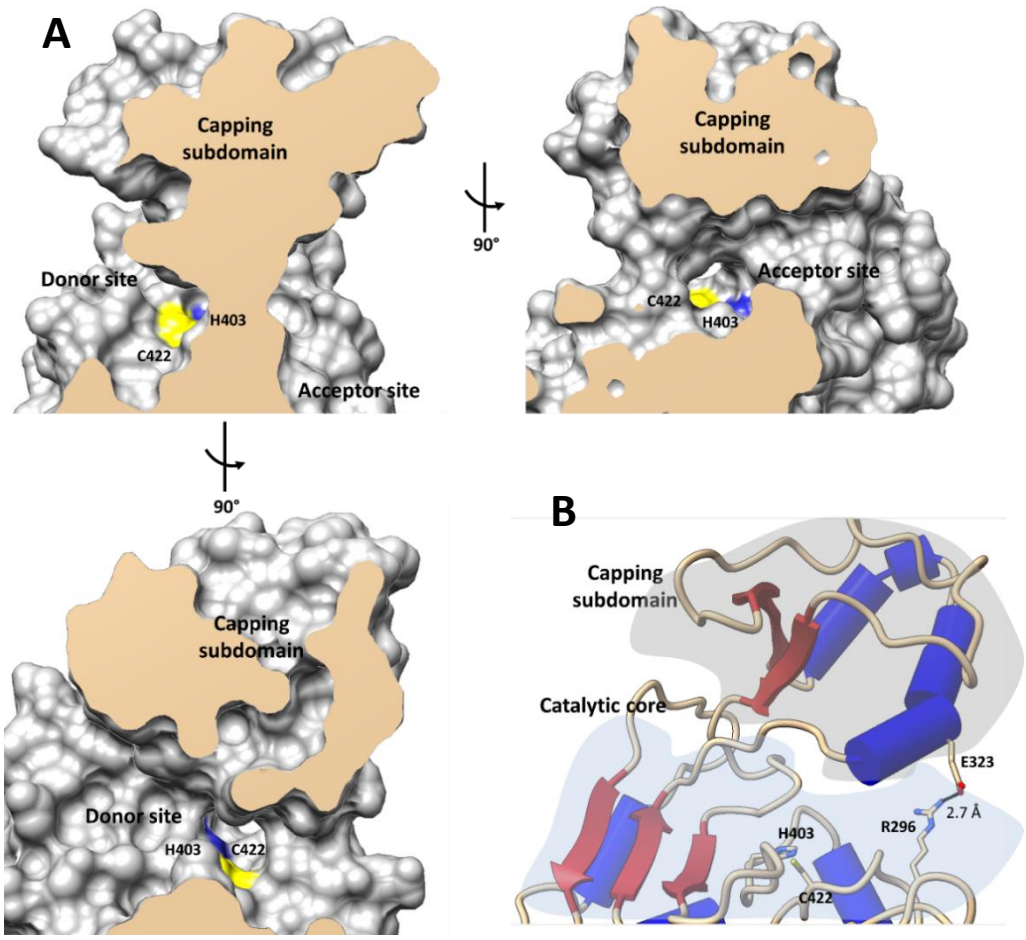


Figure 35: Bd0566 substrate binding sites and catalytic domain interactions. **A)** 90° angles apart surface views of the Bd0566 substrate binding sites. The catalytic dyad H403 and C422 sidechains are oriented towards the acyl donor site of the enzyme. The capping subdomain separates the acyl donor and acceptor sites. **B)** The capping subdomain interacts with the rest of the LDT catalytic core by hydrogen bonding interactions between the highlighted residues. Interactions of Bd0566 with PG and/or with other molecules may affect the conformation of the capping subdomain and in consequence, determine substrate selection and binding.

Even though the capping subdomain closes the substrate binding cleft, average B-factor analysis of the Bd0566 surface reveals that highest uncertainty values (coloured in scale from cyan to magenta, from 30 to 120 Å^2, respectively) are localised towards the capping subdomain of Bd0566, with values tending to 120 Å^2 (**Figure 36**). These values suggest that this region of the enzyme is particularly flexible, regardless of the hydrogen bonding interaction of the capping subdomain with the rest of the catalytic core. In this sense, the 25 residues long loop connecting β-strands 6 and 7, may contribute to the intrinsic flexibility of this region of the protein.

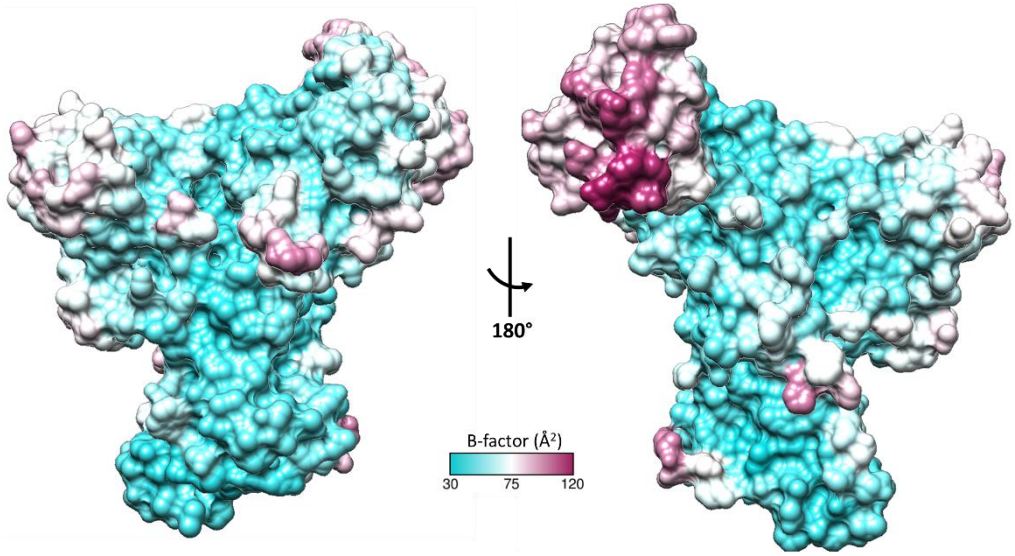


Figure 36: B-factor colouring of Bd0566. B-factor analysis of 180° angle apart surface representation of Bd0566. B-factors range from 30 to 120 Å^2 for Bd0566 (cyan to magenta, respectively), with the highest values corresponding to the capping subdomain, suggesting that this part of the enzyme is flexible and might be susceptible to conformational changes upon substrate binding.

Structural alignment of Bd0566 and *E. coli* YcbB reveals a high degree of sequence and structural conservation in the catalytic domain of the enzyme. Besides the catalytic dyad H403 and C422, and the stabilising residue D404, residues W314, P317, K391, Y401 and R424 are localised in the substrate binding cleft and appear to actively participate in substrate binding (**Figure 37**).

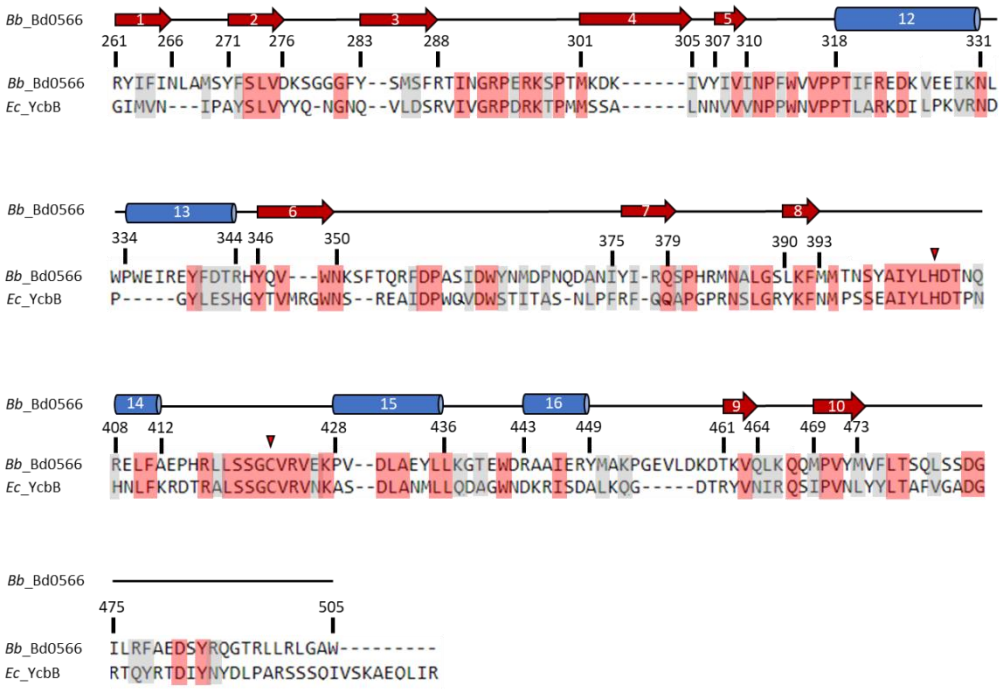


Figure 37: Sequence alignment and structural context of Bd0566 and YcbB LDT domain. Bd0566 and YcbB protein sequences were aligned using Clustal Ω and compared in view of the Bd0566 determined structure. Residues numbers are in reference to Bd0566. Identical residues between Bd0566 and YcbB are highlighted in red while similar residues are indicated in grey. The catalytic histidine and cysteine are indicated with red arrow heads.

Particularly, K391 and Y401 are part of the acyl donor site, while R424 form part of the acyl acceptor site. Interestingly, W314 and P317, which are part of the capping subdomain, localise between both acyl substrate sites. This in view of the structural information obtained for Bd0566, indicates that positively charged residues K391 and R424 may act as stem peptide stabilising groups, either on the donor or acceptor site. On the other hand, the aromatic residues Y401 and W314 and the cyclic residue P317 are commonly associated with stabilising the chair conformation of the sugar rings of the PG polyglycan chain in both donor and acceptor sites, respectively [83].

Since the YcbB structure was solved in a Meropenem-bound state, possibly mimicking the acyl-enzyme intermediate conformation, the Bd0566 apo structure was superimposed to identify residues that might be actively involved in the transpeptidation mechanism, as shown in **Figure 38**.

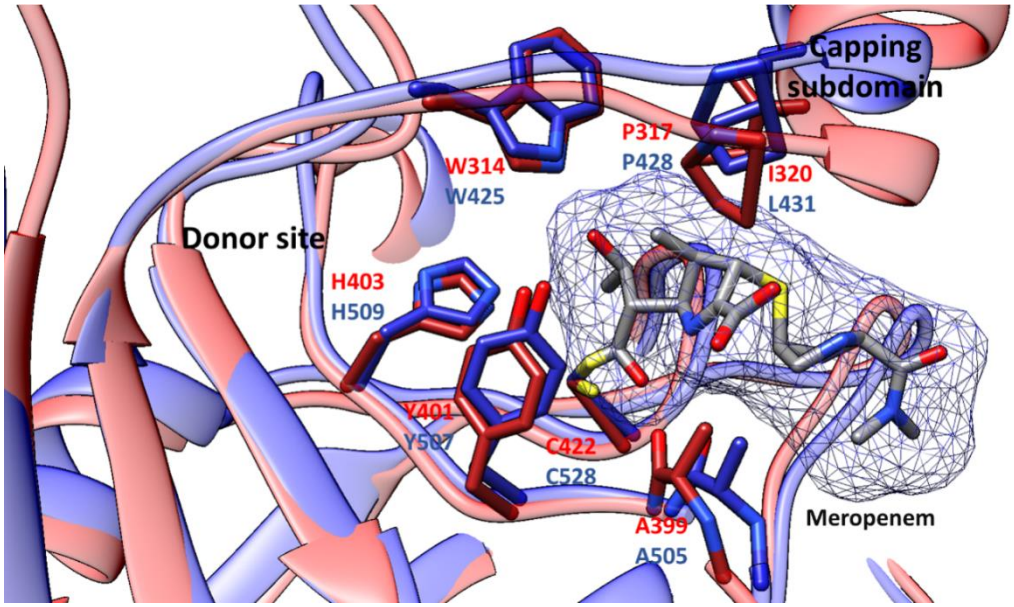


Figure 38: Superimposition of the Meropenem binding site of YcbB with Bd0566 structure. The active site of Bd0566 in its apo form (in red) is compared with the YcbB acyl-enzyme intermediate trapped state (in blue) with Meropenem (grey). Structural comparison indicates that both active sites are in similar conformations and therefore, structural rearrangement induced by substrate binding occurs before formation of the acyl-LDT intermediate.

As shown in the YcbB and Bd0566 sequence alignment, residues W314 and P317 of Bd0566 (W425 and P428 in YcbB, respectively) form a hydrophobic pocket that possibly stabilises the backbone of the PG stem peptide and/or the sugar rings of the glycan portion of the substrate [83], [147]. Additionally, Bd0566 I320, which substitutes YcbB L431, appears to also participate in forming the aforementioned hydrophobic pocket.

On the other hand, Bd0566 Y401 (Y507 in YcbB) is suggested to form an oxyanion hole responsible for stabilising the tetrahedral oxycarbanion intermediate formed during catalysis. Finally, as discussed by Caveney *et. al.* (2016) [83], residue A399 of Bd0566 (A505 in YcbB) would contribute to the stabilisation of the Meropenem molecule by hydrogen bonding with the backbone carbonyl of the residue. Thus, most active site residues and their orientations are conserved in both structures suggesting that any conformational change must occur upon substrate binding and before formation of the acyl-enzyme intermediate.

The Bd0566 second crystal form crystallised with a citrate molecule bound in the active site, in close proximity to the catalytic dyad C422-H403 (**Figure 39A**, $2F_0$-F_0 electron density map, σ= 1.50). It has been found that for PG modifying enzymes, citrate molecules

partially mimic the binding chemistry of the mDAP residue of the PG stem peptide, being useful for mapping substrate binding sites [148]–[150]. Residues surrounding the citrate molecule modelled in the Bd0566 structure (< 5Å linear distance) are depicted in **Figure 39B**.

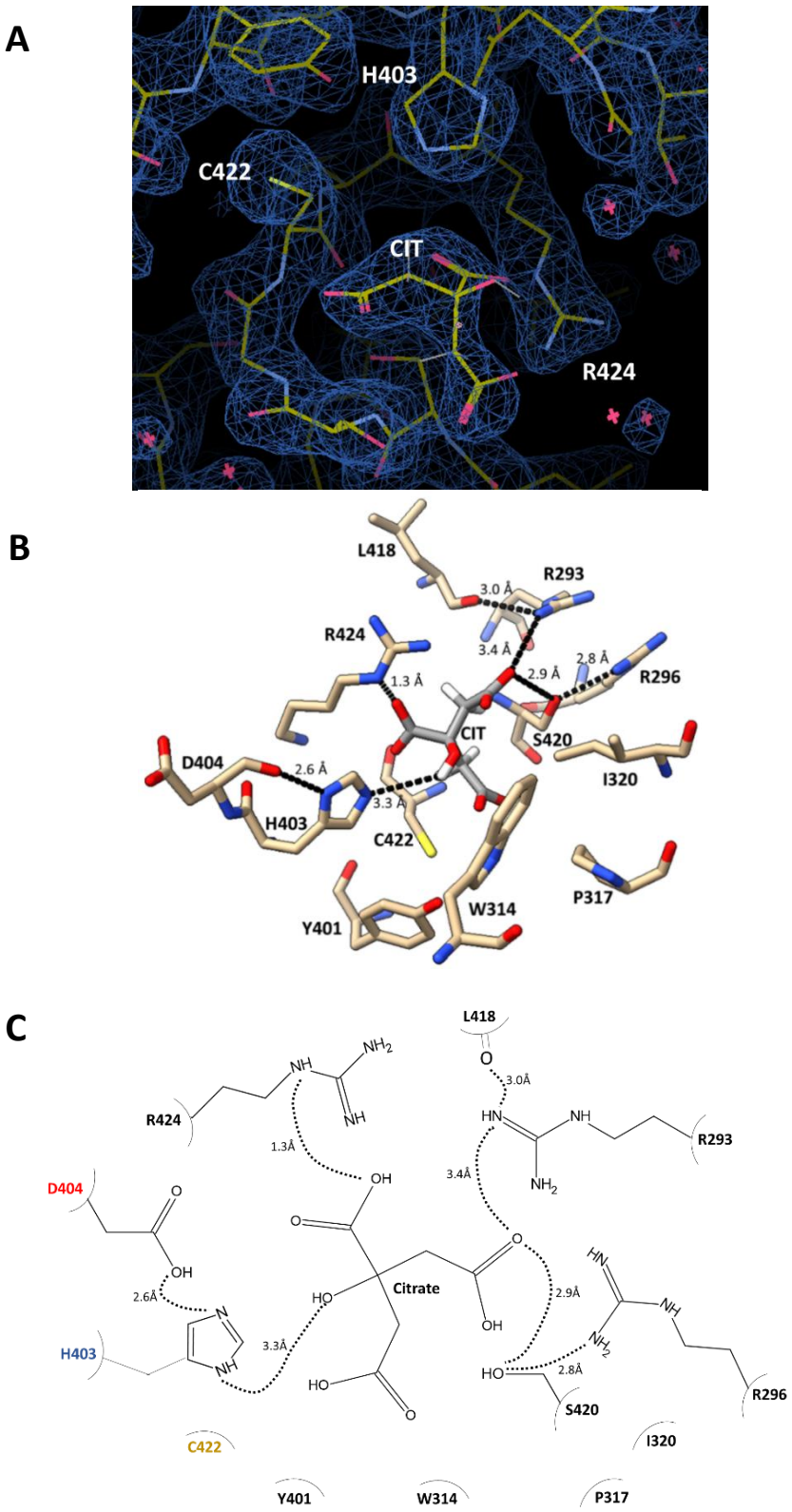


Figure 39: Citrate binding in Bd0566 active site. A) $2F_0$-F_0 electron density map of citrate in the Bd0566 solved structure (σ=1.50). **B)** Stick representation of the citrate molecule bound in Bd0566 active site. Interacting residues are indicated with their corresponding distance, which are in the hydrogen bonding range. These residues are partially shared with the meropenem superimposed Bd0566/YcbB structures. **C)** Schematic representation of citrate contacts in panel **B**.

The citrate bound molecule (in grey) locates approximately in the meropenem mapped pocket of Bd0566. Two of the carboxylate groups of the citrate molecule are stabilised by interacting with residues R293, S420 and R424, which are conserved between Bd0566 and YcbB. Additionally, H403 interacts with the hydroxyl group of the bound citrate. The aforementioned interactions are schematised in **Figure 39C**.

It is worth noting that that deletion of the YcbB capping subdomain abolishes the 3-3 PG crosslinking pathway in *E. coli*, generating a β-lactam sensitive phenotype [83]. Thus, the capping subdomain may additionally regulate catalysis in Bd0566.

These overall results give structural insights into possible transpeptidation mechanisms associated with Bd0566 LDT, including possible substrate binding sites and selection by this group of transpeptidases.

3.2.3. Unique group 5: Bd3334

3.2.3.1. Bd3334 expression and purification

Bd3334 mature protein sequence (residues 23-261) was cloned in the pET28a(+) vector fused to an N-terminal His_6-tag. Bd3334 was expressed in AI media supplemented with standard Sugar mix N as indicated in section 2.1.3 and purified by nickel IMAC and SEC (**Figure 40A** and **B**, respectively).

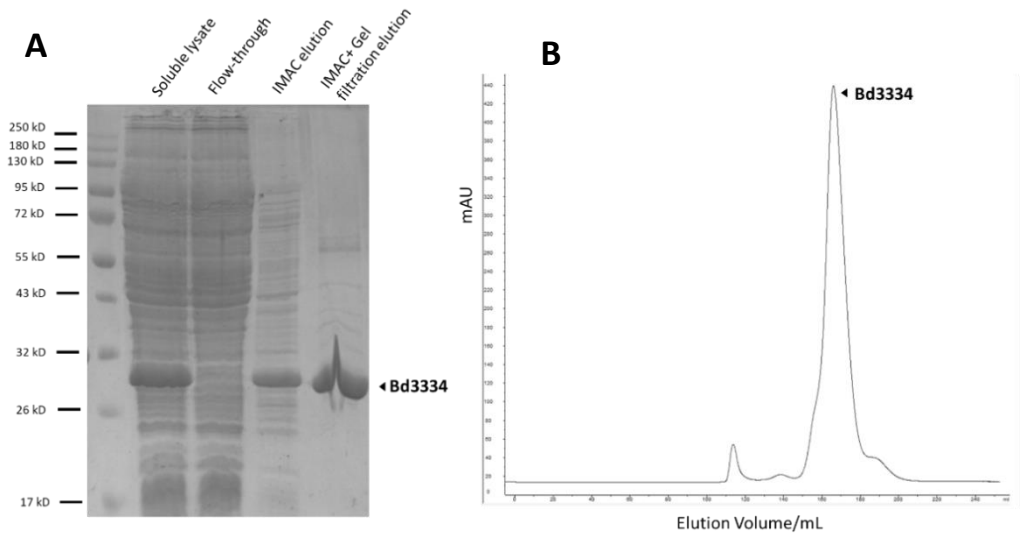


Figure 40: Bd3334 purification and gel filtration profile. A) Bd3334 was purified by nickel IMAC and gel filtration chromatography. Homogeneity of purified Bd3334 was assessed by SDS-PAGE in 15% acrylamide gels. **B)** Gel filtration elution profile of Bd3334 on a Superdex® 75 26/60 column. A monodisperse peak corresponding to Bd3334 elutes at approximately 170 mL, corresponding to monomeric Bd3334.

The hydrodynamic radius of Bd3334 determined from the Superdex® 75 26/60 column corresponds to a 25 kD globular protein, which suggests that in contrast to Bd1402, which is also a member of group 5 of LDTs, Bd3334 is monomeric in solution.

Purified Bd3334 was concentrated to 30 mg/mL and screened for crystallisation conditions. Crystals were obtained for condition B2 of the JCSG-Plus™ commercial screen (0.2M NaSCN and 20% PEG3350), and cryoprotected with mother liquor supplemented with 25% of ethylene glycol (**Figure 41**).



Figure 41: Bd3334 protein crystals. Bd3334 was manually screened for crystallisation conditions at 25 mg/mL protein concentration and 18°C temperature. Crystals were obtained in condition B2 of the JCSG-Plus™ commercial screen (0.2M NaSCN and 20% PEG3350), and cryoprotected with mother liquor supplemented with 25% of ethylene glycol prior to crystal harvesting.

3.2.3.2. Bd3334 data collection and structure solving

Diffraction experiments were carried out at the IO4 beamline as indicated in section 2.1.11, collecting native Bd3334 datasets at a maximum resolution of 1.7 Å.

Diffraction data was autoprocessed by the automated pipeline Xias2-dials and used for structure solving. Specifically, Bd3334 dataset was solved in the P $2_12_12_1$ spacegroup using the programme Phaser-MR of the Phenix suite by deploying an ensemble of the Bd1402 LDT core (residues 125-163 and 170-264) as search model, finding a single copy of the ensemble in the ASU (TFZ= 22.0, LLG= 354.896). Statistical results for data collection and refinement are shown in **Table 10**.

Table 10: Data collection and refinement statistics for Bd3334.

	Bd3334
Data Collection	
Resolution range (Å)	45.97 - 1.70
Space group	P $2_12_12_1$
Cell dimensions	
a, b, c (Å)	56.69, 59.75, 71.97
α, β, γ (°)	90, 90, 90
Total reflections	350399 (27595)
Unique reflections	27556 (2721)
Completeness (%)	99.99 (100.00)
Redundancy	12.7 (10.1)
<I>/σI	28.79 (4.57)
R_{meas}	0.05945 (0.632)
R_{pim}	0.01654 (0.1959)
$CC_{1/2}$	1.00 (0.939)
Refinement	
Maximum resolution (Å)	1.70
Number of reflections used	27556 (2721)
Number of reflections for R_{free}	1332 (122)
R_{work}	0.1865
R_{free}	0.2120
RMSD	
Bonds (Å)	0.011
Angles (°)	1.19
Number of non-hydrogen atoms	
Protein	1873
Ligand/ion	62
Water	199
B-factors	
Protein	26.73
Ligand/ion	41.83
Water	35.30
Ramachandran favoured (%)	96.55
Ramachandran allowed (%)	3.02
Ramachandran outliers (%)	0.43
Rotamer outliers (%)	0.99

*Values in parenthesis correspond to the highest-resolution shell

3.2.3.3. Bd3334 overall structure and domains

Bd3334 crystallographic structure traces 233 of the 238 residues of the mature protein (residues 26 to 259) and is depicted in **Figure 42A**.

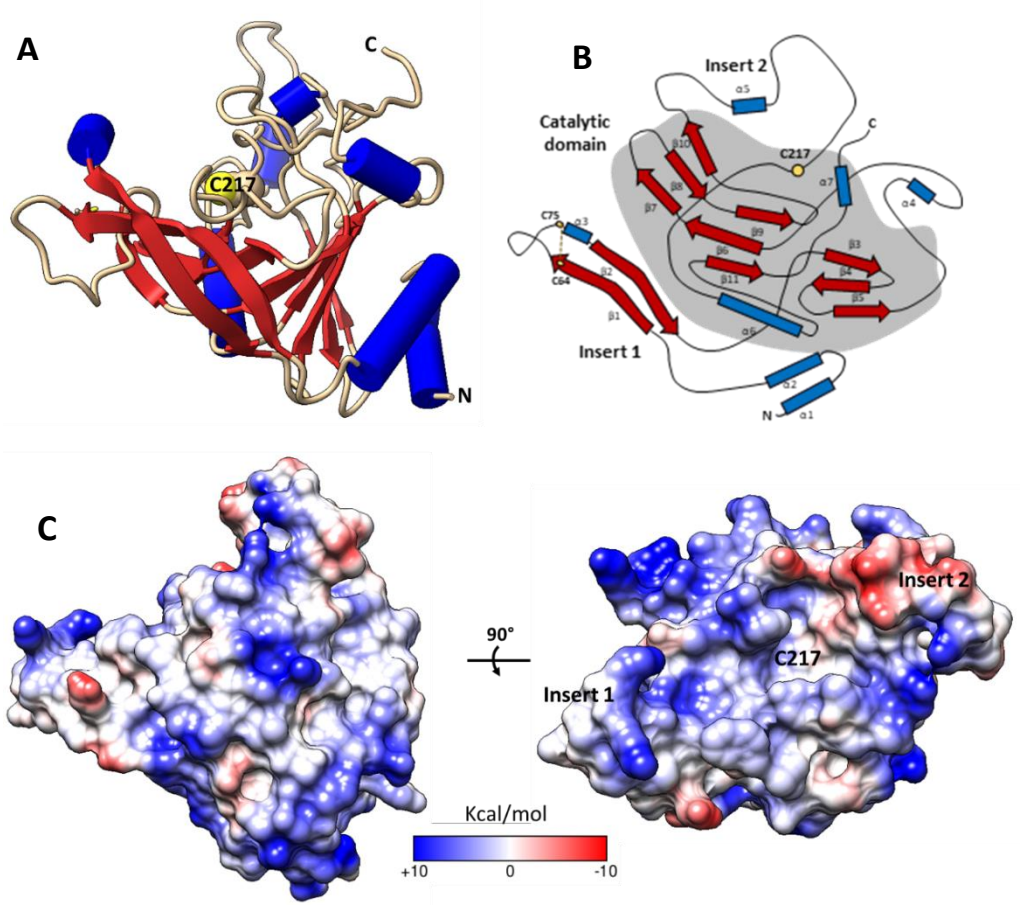


Figure 42: Bd3334 crystallographic structure, topology, and surface representation. A) Cartoon representation of Bd3334 structure coloured by secondary structure (alpha helices in blue and beta strands in red). Catalytic C217 is shown in yellow sphere format. Bd3334 comprises the archetypical YkuD2-like domain fold with a N-terminal insertion stabilised by the C64-C75 disulphide bond. **B)** Topology diagram of Bd3334 depicting the YkuD2-like domain of Bd3334 and two structural insertions (insert 1 and 2) that are exclusive for this enzyme. Catalytic C217 is highlighted in yellow. **C)** Surface electrostatic potential representation of Bd3334 in two 90° angle apart views. Surface shows that Bd3334 is mostly basic as for other PG interacting proteins, with exception of the acidic surface of insert 2.

The Bd3334 structure has a re-entrant topology including the archetypical LDT core (residues 94-122, 151-187 and 214-241) formed by a funnel-like beta barrel comprising by nine beta strands capped by an alpha helix (α6, **Figure 42B**).

The Bd3334 active site is located in the cleft formed by beta strands β7, β8 and β10 is composed by the C217-H187 catalytic dyad. In addition to the LDT core, two structural insertions are observed flanking the putative substrate binding cleft and are unique to Bd3334.

The first structural insertion (insert 1, residues 27-88) comprised of alpha helices α1-3 and the two long beta strands β1 and β2 forming an arm-like extension that projects α3 and

the loop connecting α3 and β1 towards the substrate binding cleft, specifically on the acyl donor site. This insertion is stabilised by the disulphide bond formed between residues C64 and C75. The second structural insertion (insertion 2) includes residues 193 to 212, corresponding to a flexible loop and alpha helix α5, which delimits the acyl acceptor site of LDT catalytic domain.

As shown in **Figure 42C**, Bd3334 electrostatic surface charge is mainly basic, which agrees with most LDTs for their interaction with the primarily acidic PG cell wall. A top view of the Bd3334 substrate binding cleft surface additionally shows that inserts 1 and two flanking the active site are distinctively charged positively and negatively, respectively. Further, B-factor analysis of the Bd3334 structure reveals that the potentially most flexible region of the protein is the structural insert 1 (values around 63 $Å^2$), in contrast to the rest of the structure (values approaching 15 $Å^2$) (**Figure 43**). These results suggest that the unique structural insertions 1 and 2 of Bd3334 define the substrate binding region of the enzyme and therefore might determine substrate specificity.

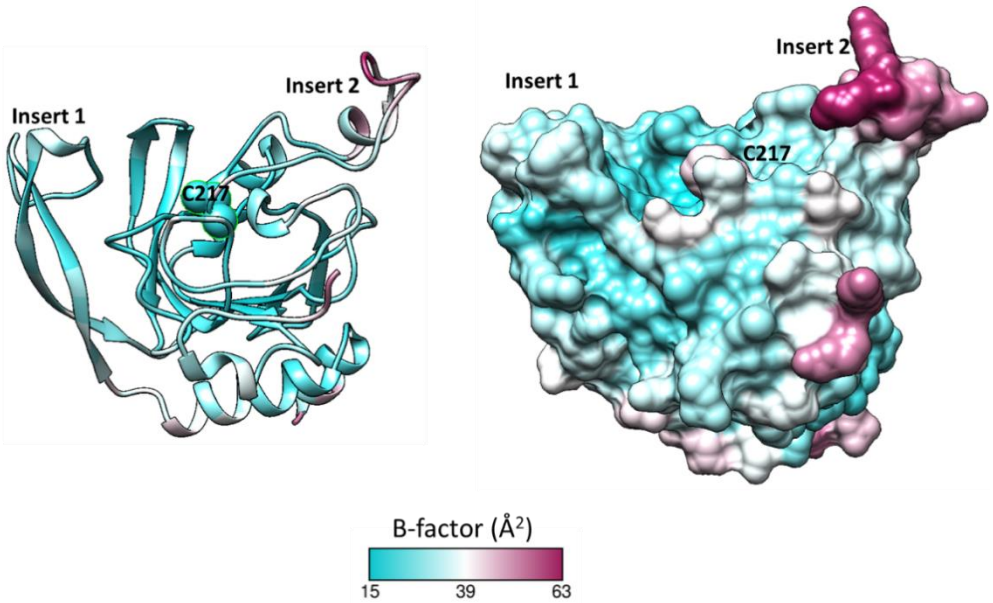


Figure 43: B-factor colouring of Bd3334 cartoon and surface representation. Cartoon and surface representation of Bd3334 were rendered by B-factor value using Chimera® software. B-factor analysis suggests that Bd3334 most mobile region corresponds to structural insert 1.

3.2.3.4. Bd3334 active site and substrate binding cleft

The catalytic dyad C217-H187 of Bd3334 is depicted in **Figure 44A**. H187 is oriented towards C217 by the 2.7 Å hydrogen bonding interaction between $N^{\delta 1}$ of H187 and the carboxylate backbone group of G188. As observed for Bd0566, the sulfhydryl group of

C217 is at a 4.3 Å distance to $N^{\varepsilon 2}$ of H187, which suggests that C217 might be reoriented upon substrate binding.

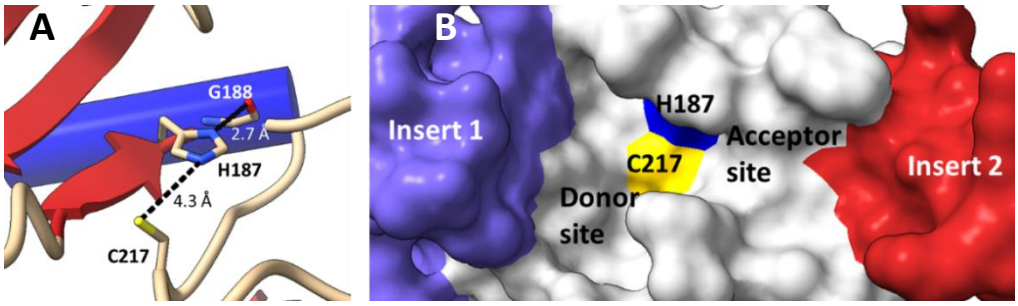


Figure 44: Bd3334 active site. A) Bd3334 catalytic dyad comprises residues C217 and H187 which are at a distance of 4.3 Å between the thiol group and the $N^{\varepsilon 2}$ of the imidazole ring. The imidazole group of 187 is oriented by its 2.7Å hydrogen bonding interaction with the backbone carbonyl group of G188. **B)** Surface representation overview of Bd3334 active site. The hourglass-shaped active site is flanked by insertions 1 and 2, suggesting they might determine substrate specificity of Bd3334.

The hourglass shaped substrate binding cleft of Bd3334 is solvent accessible, with no capping subdomains (**Figure 44B**). Structural inserts 1 and 2 delimit the substrate binding site and might determine Bd3334 specificity. Interestingly, open substrate binding sites have been associated with endopeptidase activity rather than transpeptidation activity in PG modifying enzymes, as described for Acd24020 from *C. difficile* [151]. The top Dali server matches of Bd3334 related structures are shown in **Table 11**.

Table 11: Dali results for Bd3334 structural search. Dali matches for PDB deposited structures are organised according to their Z-score.

№	PDB ID-Chain	Z score	RMSD	lali	residues	% identity	Description
1	4y4v-A	11.8	3.1	132	318	11	Conserved Hypothetical Secreted Protein
2	4a1i-A	11.8	3.0	122	165	14	Putative L,D-transpeptidase YKUD
3	5bmq-A	10.5	2.4	110	205	17	Erfk/Ybis/Ycfs/Ynhg family protein
4	4z7a-A	10.2	3.2	112	355	20	Mycobacterium tuberculosis (3,3) L,D-transpeptidase
5	6fj1-A	9.9	2.9	118	254	13	L,D-transpeptidase
6	4xxt-A	9.5	3.0	114	221	18	Fusion of predicted Zn-dependent amidase/peptidase
7	4lpq-A	8.2	2.6	109	204	21	Erfk/Ybis/Ycfs/Ynhg family protein
8	7kgm-A	7.7	3.1	114	512	16	YcbB family protein
9	4lzh-A	5.8	3.3	114	278	18	L,D-transpeptidase

Structure conservation is concentrated on the LDT domain. Even though the LDT catalytic core is structurally conserved, sequence conservation of residues oriented towards the substrate binding cleft, outside the corresponding C217-H187 catalytic dyad, is poor.

Then, possible Bd3334 substrate binding residues are proposed based on spatial proximity to C217. Proposed sidechains are highlighted in **Figure 45**.

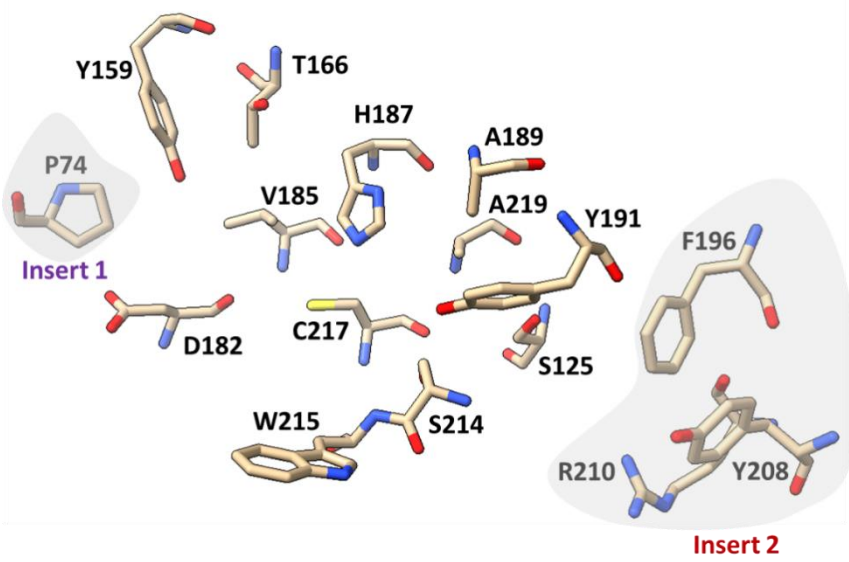


Figure 45: Active site and possible substrate binding residues of Bd3334. Residues in close proximity to the C217-187 catalytic dyad are shown in stick representation. Residue A219 corresponds to the alanine in the proposed hydrolytic SXGCXAϕ LDT motif, suggesting that Bd3334 is a peptidase rather than a transpeptidase. Residues P74 from insert 1 and F196, Y208 and R210 of insert 2 are unique features to Bd3334 and are proposed to partially determine substrate specificity,

As proposed in **Figure 44B**, residues from the unique structural inserts 1 and 2 appear to define the boundaries of the Bd3334 binding cleft. Specifically, residues P74 from structural insert 1 and F196, Y208 and R210 form the structural insert 2 are spatially oriented towards the cleft defined by the active site, suggesting that these specific Bd3334 structural variation contribute to substrate specificity. Further residue A219 corresponds to the conserved alanine of the hydrolytic YkuD2-like SXGCXAϕ motif, suggesting that Bd3334 would have endo and/or carboxypeptidase activity rather than transpeptidase activity [118], [142].

3.2.4. Group 6: Bd1075

3.2.4.1. Bd1075 expression, purification, and crystallisation

The nucleotide sequence encoding Bd1075 mature protein sequence (residues 20-332) cloned in the pET41c(+) expression vector was expressed in AI media supplemented with sugar mix N. Bd1075 was purified to homogeneity by IMAC and SEC as detailed in section 2.1.3 (**Figure 46A**).

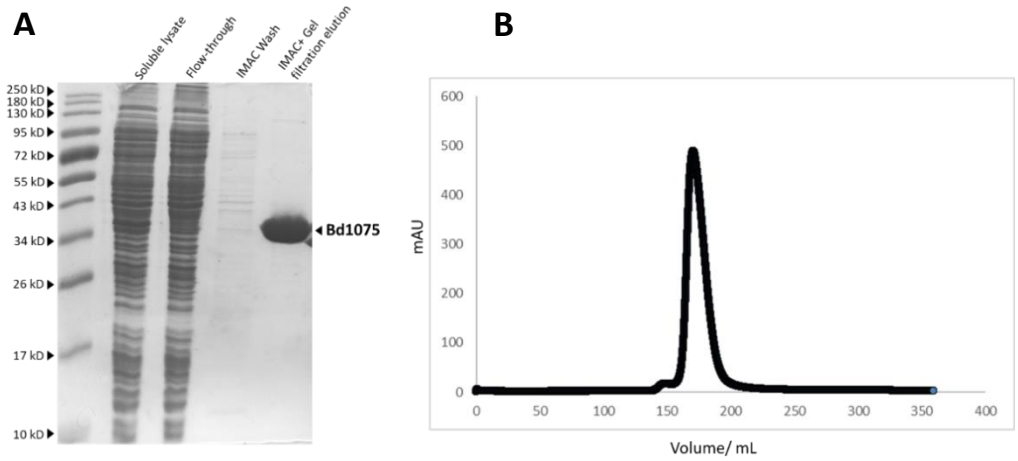


Figure 46: Bd1075 purification. **A)** Purification of recombinant Bd1075. Recombinant Bd1075 (35 kD) was purified by IMAC and Gel filtration chromatography. The purity of the obtained sample was evaluated by 15% SDS-PAGE and Coomassie blue stain visualisation. **B)** Bd1075 gel filtration elution profile on a Superdex™ 75 26/60 column. The profile depicts a monodisperse peak at 170 mL, corresponding to monomeric Bd1075. This monomeric peak of Bd1075 was pooled for crystallization and structure determination.

Bd1075 gel filtration elution volume of 170 mL from a Superdex™ 75 26/60 column shows that Bd1075 is monodisperse and monomeric in solution, with a hydrodynamic radius corresponding to a 35 kD globular protein (**Figure 46B**).

Pure Bd1075 SEC fractions were pooled and concentrated to a final concentration of 25 mg/mL and screened for crystallisation conditions. Bd1075 protein crystals were obtained in condition H8 of the BCS™ crystallisation screen (0.1 M Tris pH 7.8, 25% PEG Smear Broad, 0.1M KSCN and 0.1M NaBr) (**Figure 47**). Bd1075 crystals were cryoprotected with mother liquor supplemented with 25% of ethylene glycol and harvested for data collection.



Figure 47: Bd1075 protein crystals. Bd1075 was manually screened for crystallisation conditions at 25 mg/mL protein concentration and 18°C temperature. Crystals were obtained in condition H8 of the JCSG-Plus™ commercial screen (0.1 M Tris pH 7.8, 25% PEG Smear Broad, 0.1M KSCN and 0.1M NaBr), and cryoprotected with mother liquor supplemented with 25% of ethylene glycol prior to crystal harvesting for data collection.

3.2.4.2. Bd1075 data collection and structure solving

Diffraction experiments were carried out at the IO4 beamline as indicated in section 2.1.11. Initially, MR solving of the native diffraction data autoprocessed by the automated pipeline Xias2-dials was attempted with no positive outcomes, therefore, bromide derivatised crystallographic data (Br-SAD) was collected due to presence of sodium bromide in the crystallisation condition of Bd1075.

Bd1075 Br-SAD data was autoprocessed by the Xia2-dials automated pipeline and solved using CCP4 online CRANK2, which located seven bromide sites with an initial FOM of 0.14. Bd1075 was then built by iterative cycles of building and model-based phasing improvement. The mapped bromide sites and the corresponding density modified map ($2F_0$-F_0, σ=1.50) are depicted in **Figure 48 A** and **B**, respectively.

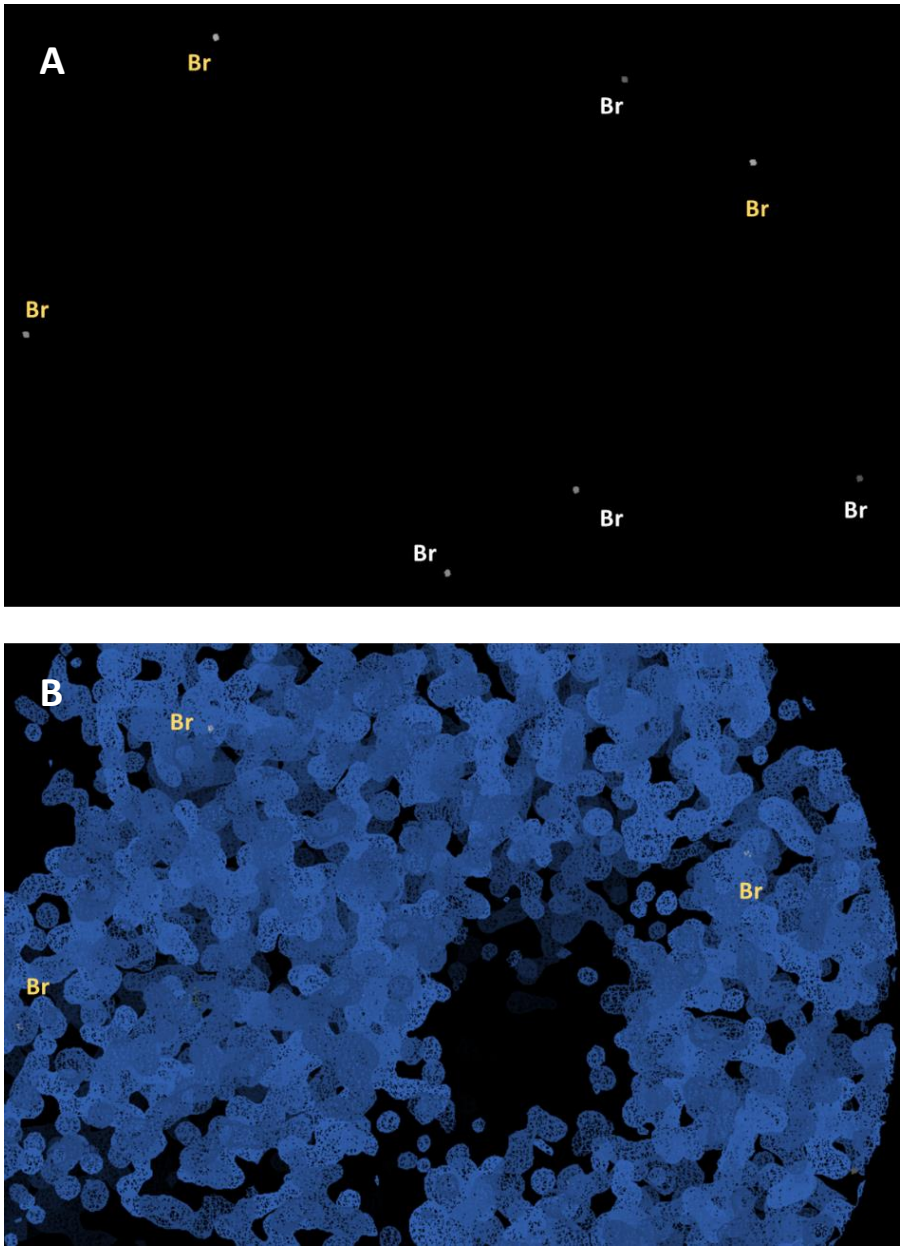


Figure 48: Bd1075 bromide phasing sites. A) CRANK2 bromide mapped sites (in white). Correspondence with the electron density mapped is showed for the yellow-coloured bromide sites. **B)** Electron density map ($2F_0$-F_0, σ= 1.50) for the bromide sites.

The obtained Br-SAD model was further built and modified using WinCoot, with cycles of refinement in PHENIX Refine. The final Bd1075 Br-SAD model comprises two protein molecules in the ASU, solved in the P 2_1 spacegroup.

Bd105 Br-SAD was used as search model for MR solving of Bd1075 native datasets at a maximum resolution of 1.55 Å. This dataset was autoprocessed by the Xia2-dials automated pipeline and solved in the P 2_1 spacegroup with two protein molecules in the

ASU (TFZ= 148, LLG= 29047). The obtained model was then subjected to multiple cycles of manual building in WinCoot and refinement using Phenix Refine. Statistical results for data collection and refinement of the bromide phased and native Bd1075 are shown in **Table 12**. The solved structure is publicly available in the PDB under the code 7O21.

Table 12: Data collection, phasing, and refinement statistics for Bd1075. Statistics for BrSAD and native Bd1075 crystallographic data are shown.

	Bd1075 BrSAD	**Bd1075 native**
Data Collection		
Resolution range (Å)	40.19 - 1.39	43.92 - 1.55
Space group	P 2_1	P 2_1
Cell dimensions		
a, b, c (Å)	49.11, 100.39, 62.28	49.19, 100.63, 62.28
α, β, γ (°)	90, 108, 90	90, 108.07, 90
Total reflections	593918 (14215)	558998 (54137)
Unique reflections	94881 (4170)	81050 (7945)
Completeness (%)	82.47 (36.26)	97.19 (95.56)
Redundancy	6.3 (3.4)	6.9 (6.8)
<I>/σI	20.58 (0.72)	26.11 (2.25)
R_{meas}	0.04208 (1.62)	0.04925 (0.5175)
R_{merge}	0.03875 (1.39)	0.04537 (0.4779)
$CC_{1/2}$	1.00 (0.36)	0.999 (0.907)
CC_{anom}	0.3 (0.0)	
Refinement		
Maximum resolution (Å)		1.55
Number of reflections used		81025 (7944)
Number of reflections for R_{free}		3999 (370)
R_{work}		0.1948
R_{free}		0.2153
RMSD		
Bonds (Å)		0.013
Angles (°)		1.18
Number of non-hydrogen atoms		
Protein		4629
Ligand/ion		43
Water		536
B-factors		
Protein		23.16
Ligand/ion		37.48
Water		33.64
Ramachandran favoured (%)		97.96
Ramachandran allowed (%)		2.04
Ramachandran outliers (%)		0.00
Rotamer outliers (%)		0.39

*Values in parenthesis correspond to the highest-resolution shell

3.2.4.3. Bd1075 overall structure

Bd1075 crystal structure comprises two protein chains in the ASU (chain A and chain B). The solved structure traces 274 and 268 of the 312 residues-long mature Bd1075 for chain A and B, respectively. Since the overall fold and conformation of both Bd1075 chains is nearly equivalent, description and analysis of the structure will be referred to chain A for being the most complete.

Bd1075 structure comprises a N-terminal LDT catalytic domain (residues 32-185) and a C-terminal nuclear transport factor 2-like domain (NTF2-like domain, residues 191-312) connected by a six residues-long linker (**Figure 49A**). The missing N-terminal residues 20-31 and C-terminal residues 313-332 are predicted to be disordered, while residues 87-91 are part of a predicted flexible loop neighbouring the catalytic C159 and are referred as the active site lid.

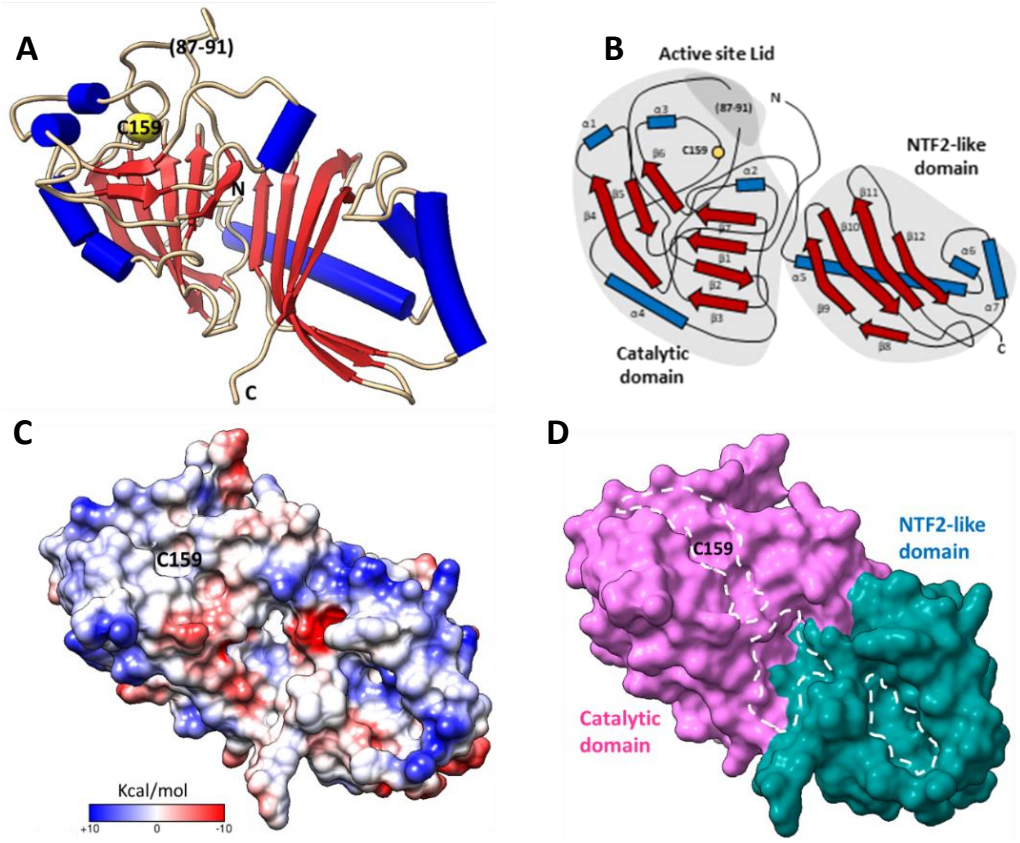


Figure 49: Bd1075 structure, topology, electrostatic surface potential and possible ligand binding pockets. A) Cartoon representation of Bd1075 crystallographic structure coloured by secondary structure (alpha helices in blue, beta strands in red). Bd1075 comprises a N-terminal YkuD-like LDT domain with a conserved catalytic C159 (yellow sphere representation) and a C-terminal NTF2-like domain connected by a six residues-long linker. A predicted four residues-long loop is not traceable in the crystallographic structure (residues 87-91) and is proposed to act as an active site lid. **B)** Topological diagram of Bd1075 (alpha helices in blue, beta strands in red; catalytic C159 is highlighted in yellow). Bd1075 fold comprises the catalytic YkuD-like domain and a NTF2-like domain. **C)** Surface electrostatic potential representation of Bd1075. Bd1075 displays a majorly basic surface charge in the proposed PG binding regions. **D)** Surface representation of Bd1075 rendered by domain. YkuD-like domain and NTF2-like domain are coloured in pink and green, respectively. Three possible pockets are highlighted in dashed white lines. The first pocket corresponds to a shallow groove centred on the nucleophilic C159 and is proposed to accommodate PG stem peptides. A second possible pocket is located in the interface between the catalytic and NTF2-like domains. The third proposed pocket is located in the NTF2-like domain suggesting that this domain participates in substrate recognition or enzyme localisation.

As illustrated in **Figure 49B**, the LDT catalytic core is formed by four alpha helices and seven beta strands adopting the beta-barrel funnel fold characteristic of these enzymes (helices α1-4 and strands β1-7) centred on the catalytic C159. The LDT domain of Bd1075 corresponds to the only experimentally determined *B. bacteriovorus* LDT that does not have a re-entrant topology.

The NTF2-like domain of Bd1075 is formed by the main helix α5 enveloped by a concave beta sheet formed by beta strands β8-12. These two main components of the NTF2-like domain are connected by an arrangement of the two parallel alpha helices α6 and α7.

As shared by other LDTs and PG modifying enzymes, the surface electrostatic potential of Bd1075 is generally basic in order to bind the mostly negatively charged PG (**Figure 49C**).

Interestingly, the top view of the Bd1075 surface shows the presence of three discrete proximal pockets that might be involved in substrate recognition (**Figure 49D**). The first pocket is part of relatively shallow groove surrounding the catalytic C159, possibly binding the stem peptide portion of the PG substrate, formed by alpha helices α1 and α3 and beta strand β6, in addition to the loops connecting such elements and the proposed active site lid. A second pocket is located in the interface between the LDT (alpha helix α2 and beta strand β4) and the NTF2-like domain (beta strands β10-12). The third pocket is located in the NTF2-like domain and is formed mainly by the alpha helix α7 and the inward face of beta strands β11 and β12. The presence of the latter two pockets suggests that the NTF2-like domain of Bd1075 might regulate substrate specificity, subcellular localisation and/or association of this enzyme with other cellular components.

Similar structures to Bd1075 were searched using Dali, finding three significant matches listed in **Table 13**. Identified matches 1 and 2 correspond to *H. pylori* Csd6 in an apo and D-Ala bound state (PDB ID: 4XZZ and 4Y4V, respectively), whereas the third most significant match is the *Campylobacter jejuni* cell shape determinant Pgp2 (PDB ID: 6XJ7) [68], [152].

Table 13: **Dali results for Bd structural search.** Dali matches for PDB deposited structures are organised according to their Z-score.

№	PDB ID-Chain	Z score	RMSD	lali	residues	% identity	Description
1	4xzz-A	23.4	3.1	246	316	23	Csd6 PG CPase (apoenzyme)
2	4y4v-A	22.4	3.2	249	318	23	Csd6 PG CPase (D-Ala bound)
3	6xj7-A	21.7	3.6	248	293	27	Pgp2 K307A mutant PG CPase

Bd1075 has been shown to be a CPase responsible for modulating the curvature of *B. bacteriovorus* [125]. In this same sense, the three matching structures that correspond to PG CPases are identified as cell shape determinants comprising an N-terminal LDT domain

and a C-terminal NTF2-like domain wherein structural differences with Bd1075 are distributed along the whole structure, including a longer active site lid in Csd6 and Pgp2 but shorter NTF2-like domains respect to Bd1075 (**Figure 50**, Bd1075 in red, Csd6 in grey and Pgp2 in white).

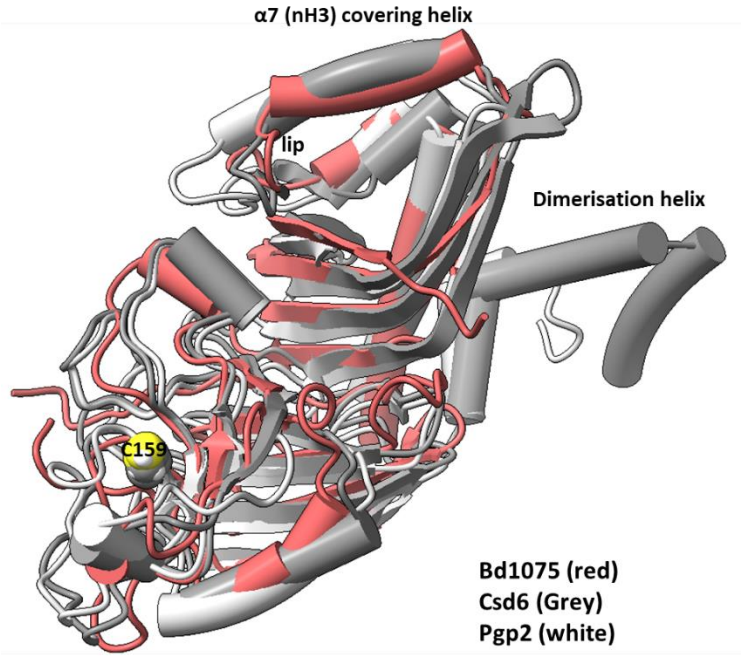


Figure 50: Structural variation of Bd1075 compared to Csd6 and Pgp2. Bd1075 crystallographic structure (PDB ID: 7O21, in red) was superimposed with the closest Dali matches Csd6 from *H. pylori* (PDB ID: 4XZZ, in grey) and Pgp2 from *C. jejuni* (PDB ID: 6XJ7, in white). As shown, structural differences are distributed along the whole Bd1075 model such as a longer active site lid in Csd6 and Pgp2 but shorter NTF2-like domains respect to Bd1075. Moreover, Csd6 dimerization helix is absent in Bd1075 and Pgp2. Further, alpha helix α7 of Bd1075 (nH3) is closer to the putative ligand binding pocket of the NTF2-like domain compared to Csd6 and Pgp2.

Remarkably, Csd6 has been shown to dimerise using three N-terminal helices that have been replaced by a predicted twelve residues-long unstructured region in Bd1075 and a truncated N-terminal helix in Pgp2.

Additionally, alpha helix α7 of Bd1075 (the third alpha helix of the NTF-2like domain, also referred as nH3) is closer to the putative ligand binding pocket of the NTF2-like domain (shown in **Figure 49D**) along the corresponding connecting loop, labelled as “lip”.

3.2.4.4. Bd1075 LDT domain and active site

The active site of Bd1075 consists of the catalytic dyad C159-H144, wherein the imidazole ring of H144 is positioned by the 2.7Å hydrogen bonding interaction of its $N^{\delta 1}$ with the backbone carboxylate of A145. As seen for the other LDTs, the 4.0Å between the

sulfhydryl group of C159 and the $N^{\varepsilon 2}$ of H144 suggests that Bd1075 C159 reorientates for catalysis, possibly upon substrate recognition/binding (**Figure 51A**).

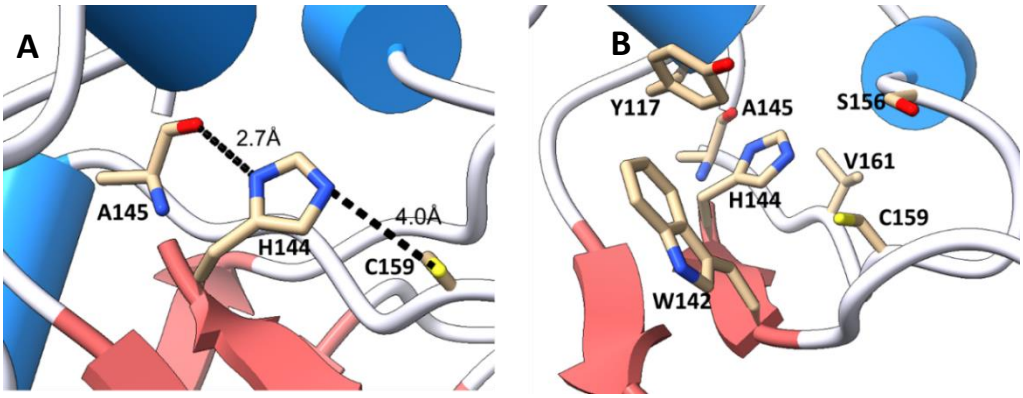


Figure 51: Bd1075 active site and substrate binding residues. A) The Bd1075 catalytic dyad comprises residues C159 and H144 which are at a distance of 4.0Å. Orientation of the imidazole ring of H148 is generated by its 2.7Å hydrogen bonding interaction with the backbone carbonyl group of A145. **B)** Residues in proximity to the C159-H144 catalytic dyad are shown in stick representation. Particularly residue V161 corresponds substitutes the conserved alanine residue in the YkuD-like domain hydrolytic motif SXGCXAϕ, in contrast to Csd6 and Pgp2, which both retain the alanine residue in this motif.

Residues proximal to the catalytic C159 are shown in **Figure 51B**. Residues Y117, W142, S156 and V161 form a mostly hydrophobic pocket surrounding the active site that could potentially accommodate the stem peptide of PG molecules.

When comparing the Bd1075 structure with the apo and D-Ala bound Csd6, and with Pgp2, it is noticeable that the highlighted active site residues of Bd1075 are conserved with exception of V161, which substitutes the conserved alanine residue of the YkuD- and YkuD2-like L,D-hydrolases conserved motif SXGCXAϕ [118], [142]. Both Csd6 and Pgp2 retain the aforementioned alanine residue, suggesting that Bd1075 is the outlier of the cell-shape LDT determinants.

3.2.4.5. Bd1075 NTF2-like domain

The Bd1075 NTF2-like domain has been shown to localise this enzyme to what becomes the outer convex cell face of *B. bacteriovorus* and is necessary for the cell to acquire its curvature, which is related to competent predation [125].

The ASU of Bd1075 is composed of two protein molecules. Remarkably, residues P110 and K111 form part of a flexible loop of the LDT of one of the protein chains that interacts with the putative ligand binding pocket of the NTF2-like domain of the second molecule in the

ASU, possibly mimicking ligand binding and allowing mapping of residues involved in Bd1075 localisation, specifically, residues F234, 242, L246, Y250, H282, D284 and W306 (**Figure 52A**).

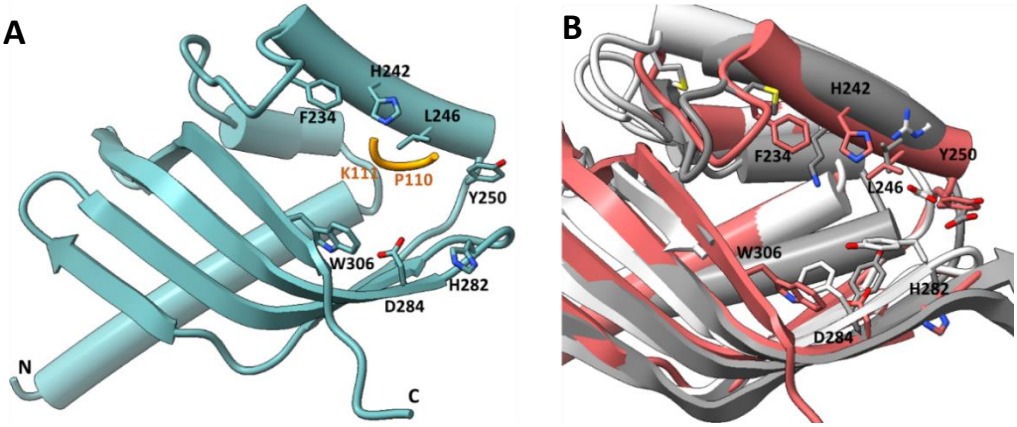


Figure 52: Bd1075 NTF2-like domain ligand binding pocket. A) The putative ligand binding pocket of the NTF2-like domain is identified by the interaction between the two Bd1075 molecules in the crystal ASU. Residues P110 and K111 of chain B interact with residues F234, H242, L246, Y250, H282, D284 and W306. **B)** Structural conservation of the NTF2-like domain ligand binding pocket residues. Bd1075 residues (PDB ID: 7O21, in red) are significantly conserved in Csd6 (PDB ID: 4XZZ, in grey) and Pgp2 (PDB ID: 6XJ7, in white), with the exception of W306, which has been shown to be critical for Bd1075 subcellular targeting.

As shown in **Figure 52B**, most of the proposed ligand binding residues of the NTF2-like pocket are somewhat conserved in Csd6 and Pgp2, however, residues F234 and H242 which are part of the "lip" and the nH3 helix, respectively, are closer to the inside of the pocket in Bd1075 when compared to Csd6 and Pgp2.

Interestingly, residue W306 is highly conserved among *Bdellovibrio* species, but is absent in other Csd6-like proteins [125]. Specifically, beta strand β12 of Bd1075 makes this NTF2-like domain longer than in Csd6 and Pgp2. In this sense, it has been shown that Bd1075 truncated at residue E305 fails to localise to the outer curve of *B. bacteriovorus* nor complements a *Δbd1075* strain, in contrast to the wild-type version of Bd1075 and a mutant truncated at position W306, suggesting that W306 is critical for ligand recognition and consequent localisation of Bd1075 [125].

3.3. Discussion

The crystallographic structures of four additional *B. bacteriovorus* LDTs, Bd0566, Bd3741, Bd3334 and Bd1075, have been determined, which in addition to the three unpublished structures for Bd0553, Bd0886 and Bd1402 make a panel of seven LDT structures. This

structural collection includes one representative of each LDT group identified in *B. bacteriovorus* HD100, wherein both Bd1402 and Bd3334 structures belong to group 5 of LDTs.

As with the previous data, each novel LDT structure presented in the result section depicts a relatively well conserved YkuD- or YkuD2-like domain that has diversified and acquire certain structural variations and/or domains specific to each LDT group. These variations will be analysed and discussed in further detail within and between each group of LDTs in the Chapter 4 of this work.

3.3.1. Bd3741 is a possible LDT with crosslinking/NCDAA-exchanging activity

Bd3741 proved to be the only soluble member of the group 2 of LDTs. This protein has been proposed to be the closest to the Lpp1-PG crosslinker ErfK in *E. coli*, nevertheless showing significant sequence and structural differences outside the YkuD-like domain [61], [89].

The overall Bd3741 fold depicts a N-terminal α-helix that, due to its proximity to what is proposed to be an eleven residues-long flexible linker, may act as a capping subdomain for the enzyme. On the other hand, a cryptic domain, named as triangular hook, is located at the opposite site of the YkuD-like β-barrel central core. Both of these structures appear to be exclusive for Bd3741, however, a similarly shaped triangular loop has been shown to be part of the dimerization domain of Ypr118w, however, the monomeric state of Bd3741 both in the crystal structure and in solution suggests a different function for this domain [146].

When comparing the experimentally solved Bd3741 structure to the predicted AlphaFold2 models available for ErfK and the endopeptidase counterpart DpaA (formerly LdtF/YafK), substantial differences are observed with models that are predicted to a high confidence [145]. However, as shown in **Figure 53**, both Bd3741 (in red) and ErfK (AlphaFold2 model, in grey) appear to have some type of capping subdomain surrounding the active site.

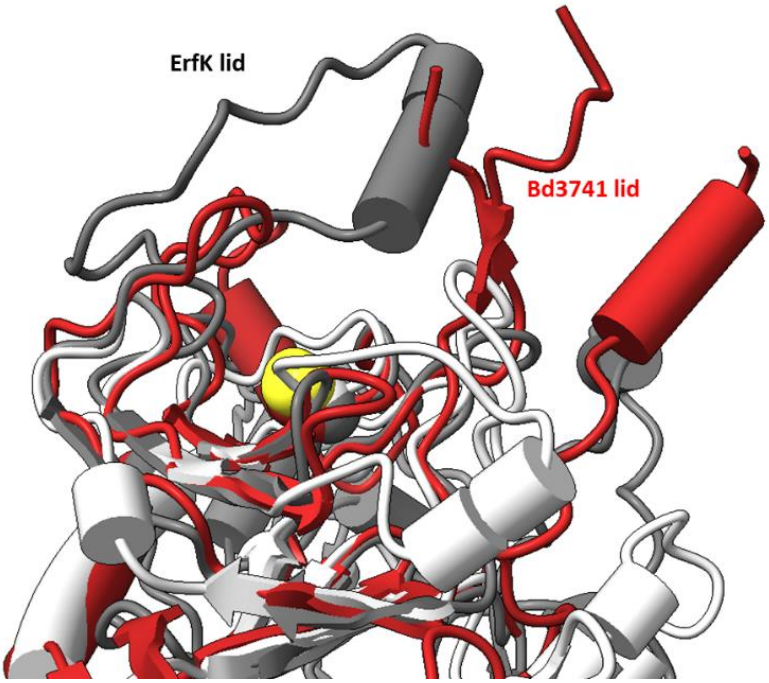


Figure 53: **Structural comparison between Bd3741, ErfK and DpaA.** The crystallographic structure of Bd3741 (in red) was superimposed with the AlphaFold2 predicted models of ErfK (in grey) and DpaA (in white). Besides the apparent differences between the three, Bd3741 and ErfK are shown to present a lid-like structure capping the active site (catalytic cysteines highlighted as yellow spheres). In contrast the predicted structure of DpaA shows a completely open active site, which is concordant with a hydrolytic second step of the LDT ping-pong mechanism.

In contrast, the DpaA structural model (in white), is predicted to lack structural insertions surrounding the active site of the enzyme, suggesting it is more solvent accessible compared to Bd3741 and ErfK. This observation agrees with the proposed crosslinking/exchanger motif (S/T)XGCϕRϕ, which is present in Bd3741 and ErfK, whereas DpaA aligns with the experimentally demonstrated peptidase motif SXGCXAϕ [118], [142]. Altogether, the structural comparison of Bd3741 with ErfK and DpaA besides suggesting that Bd3741 acts as a LDT crosslinker, appears to also be correlated with the presence of a lid structure surrounding the substrate binding cleft for LDT crosslinkers. This may be required possibly for the stabilisation of the water-labile acyl-enzyme reaction intermediate in order for the acyl-acceptor substrate (Lpp1 for ErfK) to attack in a second reaction step [61], [89]. In contrast, DpaA active site is predicted to be solvent accessible, which agrees with the second step of the reaction being an acid-base hydrolysis rather than a nucleophilic addition-elimination as in crosslinking/NCDAA-exchange reactions.

3.3.2. Bd0566 is a YcbB-related LDT with a predation-specialised structure

Bd0566 is part of the YcbB-like (or LdtA-like) group 4 of *B. bacteriovorus* LDTs. Of this group, only Bd0566 and Bd1181 were stable in solution, wherein only Bd0566 crystallised during the course of this work.

Bd0566 crystal structure depicts the most conserved fold when compared to its *E. coli* counterpart, both sharing the signature crosslinking motif (S/T)XGCϕRϕ. Nevertheless, domain specific analysis shows that Bd0566 fold and electrostatic surface charge differs from YcbB mainly in the "scaffolding" domain and the PGBD, which may impact Bd0566 specialisation in predation.

Firstly, the scaffolding domain of Bd0566 retains the α-helical content of YcbB with no evident difference in coulombic surface potential (**Figure 54**).

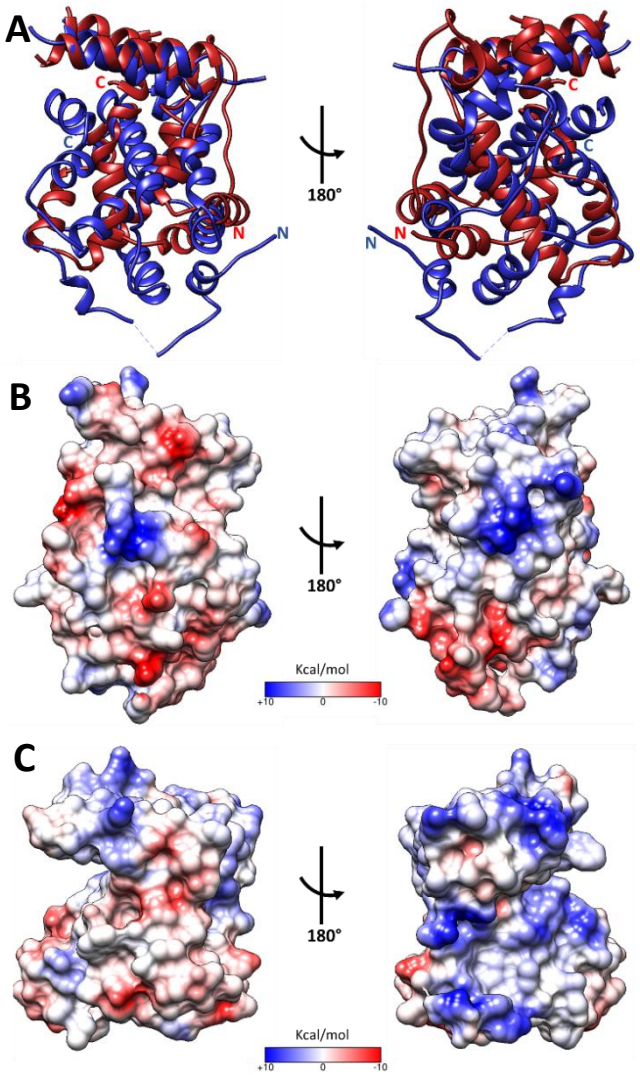


Figure 54: Structural comparison of Bd0566 and YcbB scaffolding domains. A) Structure superimposition of YcbB (in blue) and Bd0566 (in red) scaffolding domains in 180° angle apart orientations. It is visible that most structural conservation is centred on the hinge formed by the top two α-helices, whereas the rest of the domain has been reoriented and minimised in Bd0566. **B)** Coulombic surface potential of YcbB scaffolding domain in a scale of +10 to -10 Kcal/mol in 180° angles apart dispositions. **C)** Coulombic surface potential of Bd0566 scaffolding domain in a scale of +10 to -10 Kcal/mol in 180° angles apart dispositions.

Nevertheless, the equivalent "scaffolding" domain of Bd0566 has been rearranged with helices being reoriented, repositioned, or eliminated compared to YcbB (RMSD= 2.5 Å, all atoms, valid for all RMSD values unless otherwise stated). More precisely, the Bd0566 (in red) and YcbB (in blue) scaffolding domains share the two pivotal α-helix 7 and 11, which connect the scaffolding domain with the PGBD and the LDT domain. This α-helix pair has the most conserved amino acid sequence between both proteins and is held in a hinge position by the disulphide bond formed between C134-C216 in Bd0566, which has

replaced the hydrogen bonding pair C229-Q354 in YcbB. However, even though the rest of α-helices forming the core helical bundle of the scaffolding domain are present on both proteins, these have been spatially reorganised, which along with the omission of the outermost α-helices present in YcbB, give raise to a more compact scaffolding domain in Bd0566 (**Figure 54A**).

Moreover, when considering the coulombic surface potential of YcbB and Bd0566 scaffolding domains (**Figure 54B**, and **54C**, respectively), differences between both enzymes become more evident. The scaffolding domain of YcbB has a more expanded surface with an anterior negatively charged face and a posterior dipolar surface. In contrast to YcbB, Bd0566 putative scaffolding domain defines a much more compact fold, with an anterior negatively charged face as in YcbB, but with a positively charged posterior face.

All the above in addition to alarmone production in response to β-lactams not being related to the predatory cycle of *B. bacteriovorus*, makes it feasible to suggest that the scaffolding domain of Bd0566 has evolved specifically for predation. Thus, the repurposed scaffolding domain might recognise other PG modifications carried out by the predator, such as the identified mDAP terminal additions or lipid attachment to the stem peptide [33], [39], [153]. A similar interdependence has been observed between prey's PG deacetylation and the exit of predator's offspring at the end of *B. bacteriovorus*' life cycle [116], [117].

The PGBD of Bd0566 is topologically similar and highly superimposable with its YcbB counterpart (RMSD= 1.6 Å), wherein most conserved residues are localised in the hydrophobic core of the domain, in addition to R244 and D337 of YcbB (K171 and D220 in Bd0566, respectively), which are proposed to participate in PG binding (**Figure 55A**).

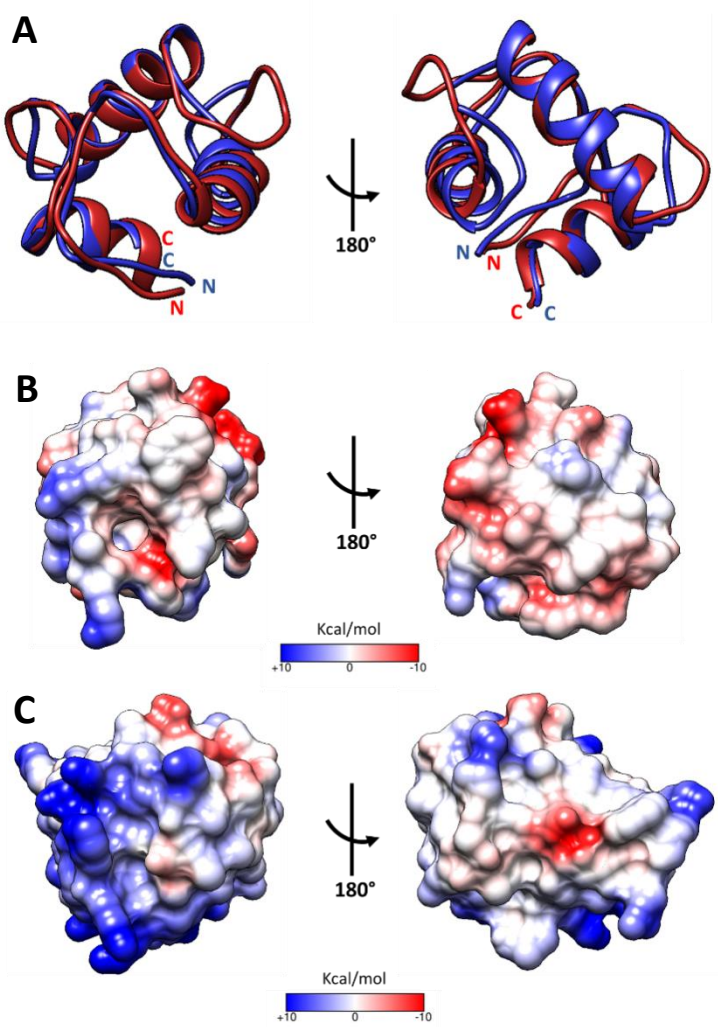


Figure 55: Structural comparison of Bd0566 and YcbB PGBDs. A) Structure superimposition of YcbB (in blue) and Bd0566 (in red) PGBDs in 180° angle apart orientations. The secondary structure of this domain is highly conserved in both enzymes. **B)** Coulombic surface potential of YcbB PGBD in a scale of +10 to -10 Kcal/mol in 180° angles apart dispositions. **C)** Coulombic surface potential of Bd0566 PGBD in a scale of +10 to -10 Kcal/mol in 180° angles apart dispositions.

Yet, differences between both domains are apparent when analysing the coulombic surface potential of Bd0566 and YcbB PGBDs. Specifically, the YcbB PGBD is mainly acidic on its surface and lacking any prominent groove that may accommodate the stem peptide of a PG chain, suggesting that this domain interacts with the sugar portion of the PG mesh, stabilising the sugar rings by hydrogen bonding with the carboxylate groups of the PGBD residues (**Figure 55B**) [83]. In contrast, Bd0566 putative PGBD presents a prominent basic patch and possible binding groups, suggesting that PG chains potentially interact with Bd0566 via the stem peptide portion of the molecule (**Figure 55C**).

It has been determined that a D337A YcbB mutant produced a β-lactam sensitive phenotype in *E. coli*, in contrast to the wild-type and a R244A mutant, indicating that this residue, and in consequence PG binding by this accessory domain, are necessary for *in vivo* functioning of the enzyme [83]. Is was then hypothesised that this domain orientates the PG mesh towards the acyl acceptor site of the LDT catalytic domain [83]. For Bd0566, a similar contribution of PGBD to catalysis might be expected, however, the consequences of a possible difference in PG binding by Bd0566 PGBD are unknown and need further study. A first approach would be studying the impact of mutating Bd0566 D220 (D337 in YcbB) on the 3-3 crosslinking activity of the enzyme.

3.3.3. Bd3334 is a possible predatory YkuD2-like LDT with proposed peptidase activity

Bd3334 is part of group 5 of *B. bacteriovorus* LDTs, as well as Bd1402, that in contrast has been determined to be monomeric in the crystal structure and in solution.

Bd3334 crystallographic structure depicts a more open active site, which along the hydrolytic YkuD2-like motif SXGCXAϕ, suggests that this enzyme has peptidase activity over crosslinking and/or exchanging activity [118], [142], [151]. Additional structural features surrounding the substrate binding cleft (insert 1 and 2) are observed in the determined structure, however, this information by itself is insufficient to determine its possible targets and substrate specificity. Thus, further comparative analysis within this group and with the rest of the LDTs will be presented in the following chapter.

3.3.4. Bd1075 is a L,D-carboxypeptidase that regulates the cell shape of B. *bacteriovorus*

Bd1075 (group 6) is the only *B. bacteriovorus* LDT that has been proven to stay in the predator rather than being translocated to the bdelloplast. Moreover, it has been shown that Bd1075 is in fact a L,D-carboxypeptidase that regulates the curvature of *B. bacteriovorus*, indirectly affecting its predation fitness [125]. Interestingly, the hydrolytic motif SXGCXAϕ has diverged to SXGCXVϕ in Bd1075, which is relatively anomalous compared to other L,D-hydrolases.

In this sense, the Bd1075 crystallographic structure depicts a C-terminal NTF2-like domain in addition to the conserved YkuD-like catalytic core, which has been shown to determine the subcellular localisation of Bd1075.

As supported by the crystal lattice interaction between the two molecules of Bd1075 in the ASU, residues F234, 242, L246, Y250, H282, D284 and W306 are part of a pocket that is suggested to be responsible for enzyme localisation by interacting with the PG mesh in *B. bacteriovorus*. Furthermore, residue W306 has been shown to be pivotal for the NTF2-like domain targeting function, which is a novel feature not observed in the closest available structural relatives Csd6 and Pgp2 [68], [152].

Thus, the presented Bd1075 structural information along with functional data demonstrate that this particular enzyme affects *B. bacteriovorus*' capacity to predate effectively, even though it might not be considered a "predatory" LDT like the ones from groups 1 to 5. In this regard, it is possible to propose that hydrolysis of 3-3 PG crosslinks by the targeted activity of Bd1075 could diminish the surface tension of one part of the cell wall, allowing the cell to be deformed by the internal turgor pressure, generating the convex form of *B. bacteriovorus*.

3.4. Conclusion

The crystal structures of two proposed LDT crosslinkers, Bd3741 and Bd0566, and two LDT peptidases, Bd3334 and Bd1075, have been determined. These results complete a panel of at least one representative structure of each LDT subfamily in *B. bacteriovorus* that will allow the integrated analysis within and between each group to probe their possible role in predation.

CHAPTER 4

4. COMPARATIVE ANALYSIS OF *B. bacteriovorus* L,D-transpeptidases

4.1. Introduction

Seven out of the 19 putative *B. bacteriovorus* HD100 LDT structures have been experimentally determined, with at least one representative structure per group (groups 1 to 6), with the exception of group 5, wherein two structures have been obtained.

Therefore, evolutionary diversification of this family of enzymes is the subject of analysis of this chapter. Particularly, acquisition of additional domains, amino acid and structural diversification within each group, possible substrate recognition, and divergence between each group of LDTs will be analysed and discussed to better understand their specific role in the *B. bacteriovorus* predatory cycle. Structural AlphaFold2 predictions of non-crystallised LDTs quality was analysed based on pLDDT scores (percentage of Local Distance Difference Test). pLDDT score plots are shown in the Appendix section.

4.2. Results

4.2.1. Clustering of *B. bacteriovorus* HD100 LDTs

To stablish possible relationships within and between LDT groups, protein sequences of the 19 predicted LDTs encoded in the *B. bacteriovorus* HD100 genome were compared using the MEGA®11 software, with 500 iterations for the generation of a dendrogram using the Neighbor-Joining algorithm (**Figure 56**).

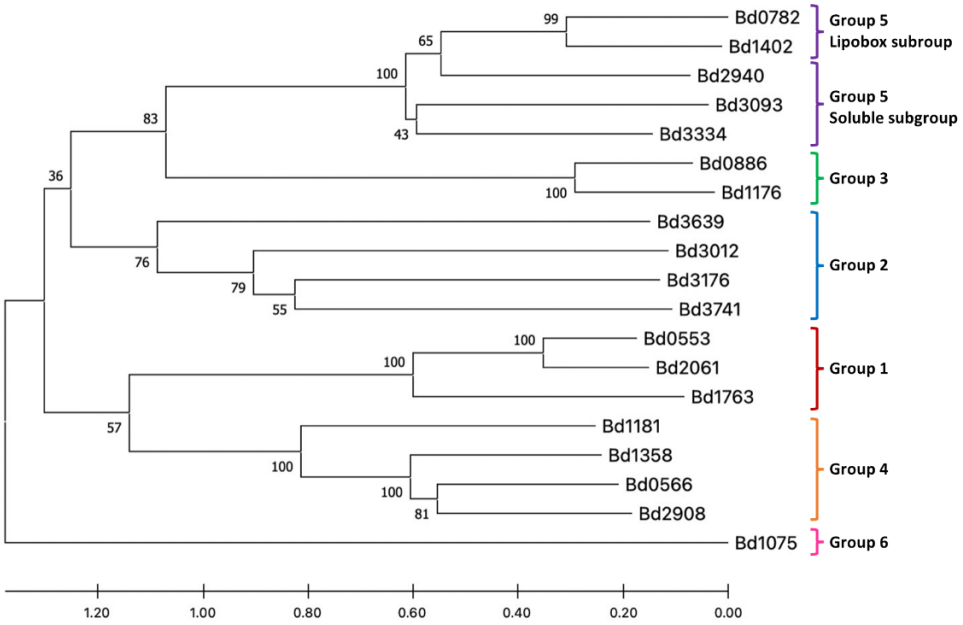


Figure 56: Dendrogram analysis of *B. bacteriovorus* HD100 LDTs. 19 mature LDT protein sequences were aligned and analysed using the MEGA11® software, with 500 iterations for the generation of the dendrogram. Bootstrapping values correspond to percentage of iterations yielding the presented results. Scaled distance is also presented. *B. bacteriovorus* HD100 LDTs cluster in six distinctive groups as previously determined.

As previously stated, the 19 LDTs cluster in six groups, with Bd1075 branching out from the rest of LDTs. This correlates with Bd1075 being linked to *B. bacteriovorus* HD100 cell shape determination instead of possible bdelloplast modification [125].

Group 1 shows a differentiation of two clades from a common ancestor: the BBd0553/Bd2061 clade and the Bd1763 clade. This separation is due to the diversification of the N-terminus of Bd1763 compared to Bd0553 and Bd2061 and is further discussed in section 4.2.2.1.

In group 2, Bd3639 is the first LDT to separate from the rest, defining it as the possible outlier of this group. On the other hand, Bd3012 shows the same origin as the Bd3176/Bd3741 clade, suggesting that they might retain similar key structural features necessary for their folding and/or function. This is further discussed in section 4.2.2.2.

Group 3 comprises only two LDTs, Bd0886 and Bd1176, which are almost identical in their amino acid sequences, forming a single clade.

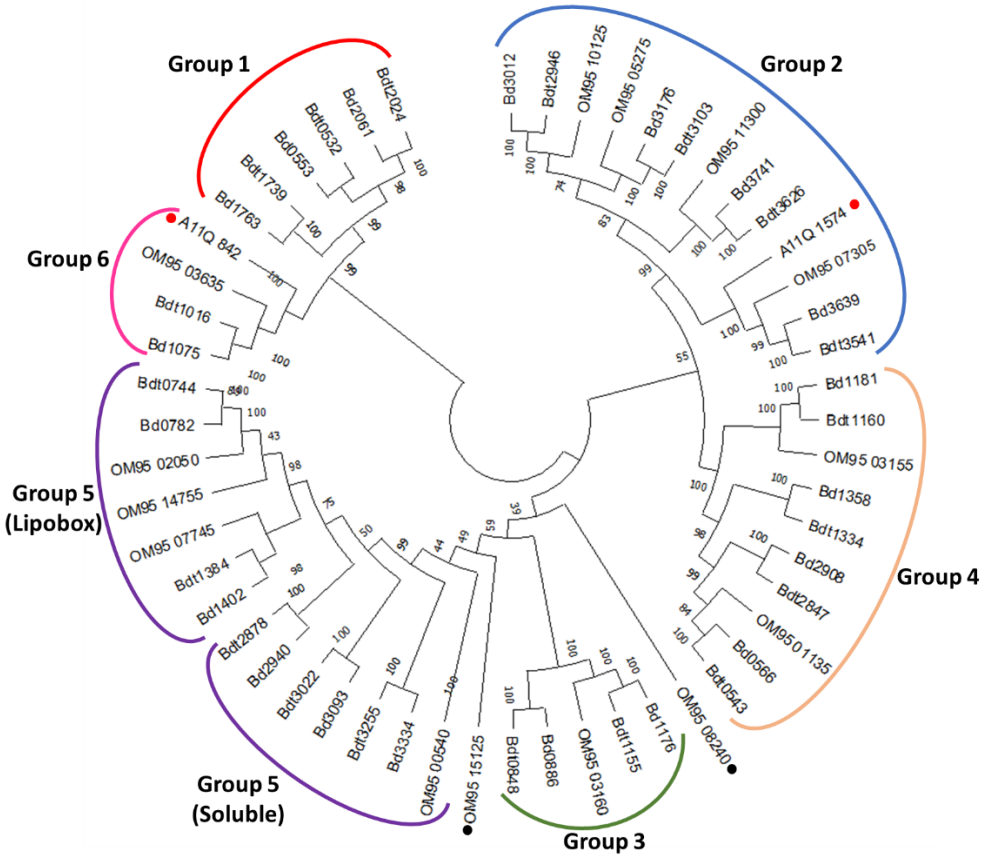
Group 4 depicts a distribution that is similar to group 2, with Bd1181 separating early from Bd0566, Bd1358 and Bd2908, suggesting possible structure variations between Bd1181 and the rest of these LDTs.

Group 5 members are defined by the YkuD2-like LDT domain and can be initially subdivided into lipobox LDTs and soluble LDTs based on their signal peptide sequence. Bd2940, Bd3093 and Bd3334 have a type I SP, being secreted to the periplasm (or other subcellular locations). In contrast, Bd0782 and Bd1402 have a type II SP, with a lipobox signal that will anchor the mature protein to the cell membrane (and/or restrict it to vesicles).

Figure 56 shows that group 5 forms three clades: Bd0782/Bd1402, Bd3093/Bd3334 and Bd2940 only. The Bd2940 clade is closer to the Bd0782/Bd1402 lipobox SP clade than to the soluble Bd3093/Bd3334 clade. This result suggests that in *B. bacteriovorus* HD100, the common group 5 ancestor could have separated, originating the Bd3093/Bd3334 and Bd0782/Bd1402/Bd2940 subgroups. Then, a later acquisition of the lipobox SP could have primed the separation of Bd2940 from Bd0782 and Bd1402. Another possible scenario is that the original Bd0782/Bd1402/Bd2940 subgroup separated due to the loss of the lipobox SP from a prototypical Bd2940.

LDT presence and distribution was further analysed in three fully sequenced BALOs: *B. bacteriovorus* HD100, *B. bacteriovorus* Tiberius, *Bdellovibrio* ArHS and *B. exovorus* JSS (*Pseudobdellovibrio exovorus* JSS). *B. bacteriovorus* Tiberius was chosen due to its genome being extensively syntenic with *B. bacteriovorus* HD100, however it naturally exhibits simultaneous HD and HI lifecycles [154]. *Bdellovibrio* ArHS is a thermophilic BALO [155]. Finally, *B. exovorus* JSS, which is a fresh water epibiotic predator, was used as outgroup for the analysis [10].

Homologues for the 19 *B. bacteriovorus* HD100 LDTs were searched in each genome using NCBI BLAST and aligned using Clustal Ω. Aligned sequence were then analysed using the MEGA®11 software, with 500 iterations for the generation of a dendrogram using the Neighbor-Joining algorithm (**Figure 57**).



<u>Figure 57</u>: LDT distribution in *Bdellovibrio*-related predators. Mature LDT protein sequences from *B. bacteriovorus* HD100 (Bd proteins), *B. bacteriovorus* Tiberius (Bdt proteins), *Bdellovibrio* ArHS (OM95 proteins) and *B. exovorus* JSS (A11Q proteins) were aligned and analysed using the MEGA11® software, with 500 iterations for the generation of the dendrogram. Bootstrapping values correspond to percentage of iterations yielding the presented results. The six LDT groups are present in *B. bacteriovorus* HD100 and Tiberius, however group 1 is absent in *Bdellovibrio* ArHS. *B. exovorus* has only two LDT-related sequences (indicated in red dots), A11Q 842 from group 6 and A11Q 1574 from group 2. Two sequences from *Bdellovibrio* ArHS that were initially predicted to be LDTs (indicated in black dots) proved to correspond to non-related protein.

As expected for the two syntenic strains of *B. bacteriovorus*, all six LDT groups are present in HD100 and Tiberius (Bd versus Bdt proteins).

Interestingly, group 1, which comprises predicted YkuD-like hydrolases, is absent in *Bdellovibrio* ArHS (OM95 proteins). In contrast, *Bdellovibrio* ArHS encodes three lipobox (OM95 02050, OM95 07745 and OM95 14755) and one soluble (OM95 00540) YkuD2-like hydrolases that cluster as group 5 of LDTs. This result suggests that there is a potential overlap between the PG hydrolysis carried-out by groups 1 and 5. On the other hand, it is possible that substrate availability and specificity differs in the thermophilic environment of *Bdellovibrio* ArHS when compared to *B. bacteriovorus* HD100 and Tiberius.

The other LDTs from groups 2-6 are represented in *Bdellovibrio* ArHS even though with a lower number of members per group, indicating that even in extreme environments, these groups of LDTs are required for endobiotic predation. Proteins OM95 15125 and OM95 08240 (indicated with black dots), were initially predicted to be part of groups 2 and 4, respectively, however, they correspond to a non-related acetyl transferase and MagC1-domain protein, branching out of the dendrogram.

Figure 57 also shows that when more sequences are included in the analysis, Bd2940 (which grouped closer to the Bd0782/Bd1402 clade in *B. bacteriovorus* HD100), is more closely associated with the soluble LDTs of group 5.

Finally, only groups 2 and 6 are represented in *B. exovorus* JSS (indicated in red dots, proteins A11Q 1574 and A11Q 842, respectively). This suggests that the crosslinking activity of the ErfK/Ybis-like group is required even in epibiotic predation. This further suggests that groups 1, 3, 4 and 5 are potentially linked to prey's periplasm colonization and/or bdelloplast formation, which is not observed for this epibiotic predator.

4.2.2. Individual LDT group analysis

4.2.2.1. LDT group 1

The crystallised member of group 1, Bd0553, was determined to share a sequence identity of 68% with Bd2061 and 38% with Bd1763, suggesting that Bd1763 is a probably the relative outlier within this group.

As shown in **Figure 58**, most of the sequence differences between Bd0553 and Bd2061 with Bd1763 are located towards the N-terminal region.

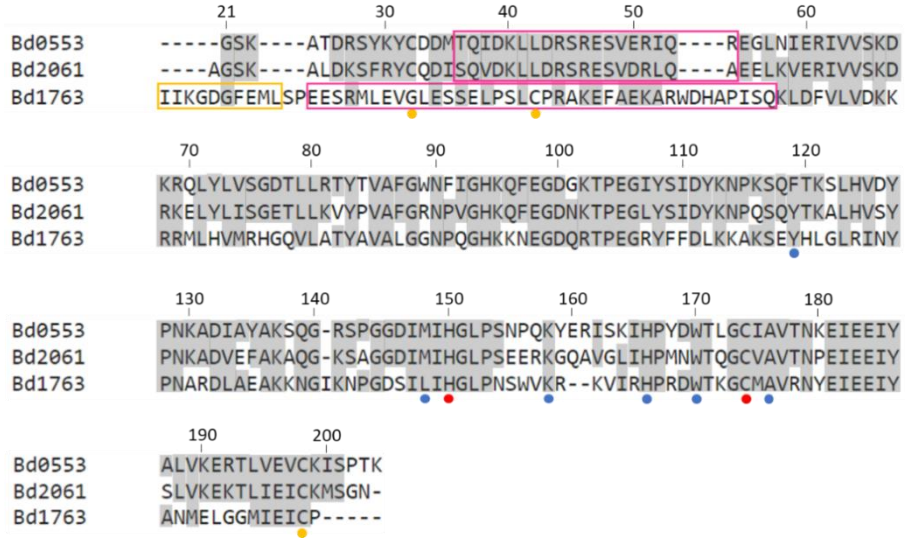


Figure 58: LDT group 1 sequence alignment. Amino acid sequence alignment of Bd0553, Bd1763 and Bd2061 is shown with similar residues highlighted in grey. Catalytic, substrate binding residues and disulphide forming cysteines are indicated with red, blue, and yellow dots, respectively. The predicted alpha-helix rich domain is highlighted in magenta for each protein. A unique N-terminal extension of Bd1763 is indicated with a yellow box. Amino acid numbering is relative to Bd0553 sequence.

Specifically, Bd1763 diverges significantly from Bd0553 and Bd2061 in the predicted alpha-helix rich domain (magenta-coloured boxes), including repositioning of one of the disulphide-forming cysteines in Bd1763 by nine residues. Furthermore, a ten residues-long insertion is present in the N-terminus of Bd1763 (yellow-coloured box).

In contrast, the three proteins are highly similar in the YkuD-like domain, showing complete conservation of not only the catalytic residues (H150 and C174 in Bd0553, red dots), but also of the proposed substrate binding residues presented in section 3.1.2 (**Figure 19**), including A176 of Bd0553, which corresponds to the L,D-hydrolase defining residue of the SXGCXAϕ motif [118], [142].

To analyse the possible structural impact of the sequence diversification of group 1, particularly of the divergent N-terminus of Bd1763, the experimentally determined structure of Bd0553 was compared with the AlphaFold2 models of Bd2061 and Bd1763 (**Figure 59**, pLDDT score values of 94.2 and 88.2, respectively).

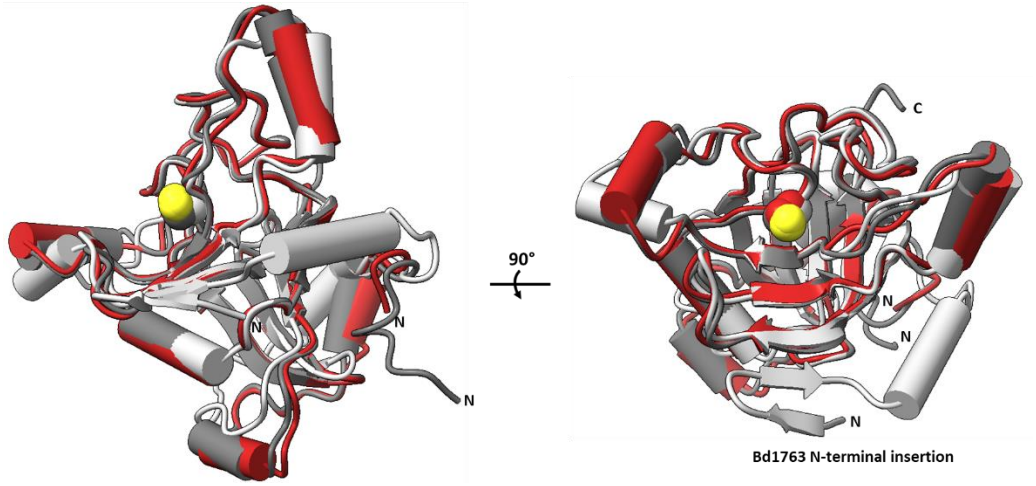


Figure 59: LDT group 1 structural divergence. Bd0553 crystal structure (in red) was superimposed with the Alphafold2 predicted structures of Bd1763 and Bd2061 (grey and white, respectively). The resulting ensemble is shown in two 90° angle views. Catalytic cysteine is depicted in yellow sphere representation.

As predicted by the amino acid sequences, Bd0553 shows the highest structural similarities with Bd2061, with an overall RMSD of 0.90 Å in contrast to the 2.94 Å RMSD of Bd1763 with respect to Bd0553.

Even though the Alphafold2 predicted structures of Bd1763 and Bd2061 retain a fold similar to the experimentally determined structure of Bd0553, significant differences are found in the N-terminus of Bd1763.

Particularly, the N-terminal extension of Bd1763 in **Figure 59** adopts an antiparallel beta sheet topology that interacts with the substrate binding side of the LDT core beta barrel (predicted with a confidence value over 80%). Moreover, the predicted first helix of the helix rich domain of Bd1763 is longer and repositioned differently with respect to Bd0553 and Bd2061. These structural differences may account for different roles of the alpha-helix rich domain.

As shown in **Figure 60A**, the N-terminal extension of Bd1763 (in red) potentially forms a groove that is absent in Bd0553 and Bd2061. Particularly, residues S33, E37, F120 and D122 of Bd1763 form a discrete pocket that could potentially bind a ligand and have no equivalents in Bd0553 crystal structure (**Figure 60B**). This pocket is not conserved in either Bd0553 nor Bd2061, with the exception of residue E37 (D111 in Bd0553), suggesting that Bd1763 could have a potentially altered substrate specificity or be targeted to a different subcellular localisation inside the bdelloplast.

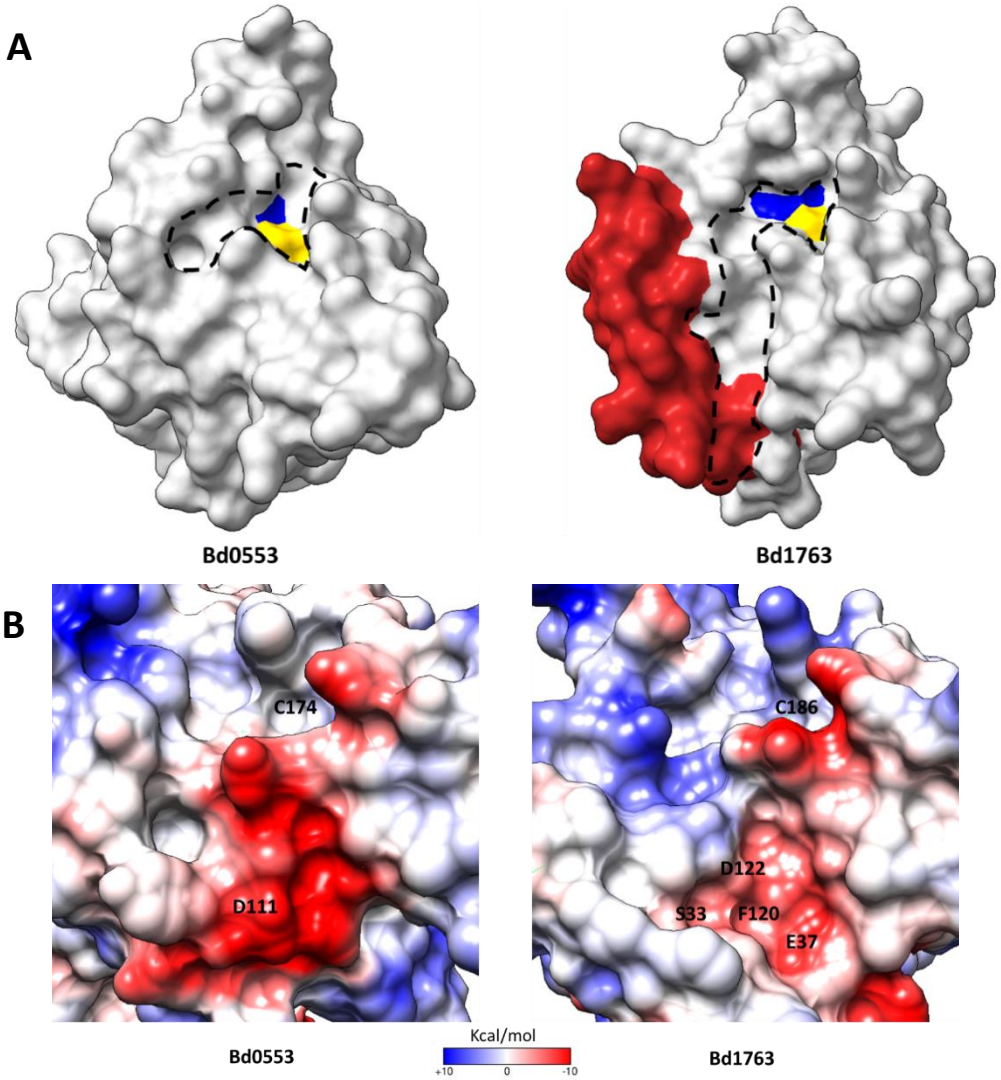


Figure 60: Bd0553 and Bd1763 surface comparison. A) Surface representation of Bd0553 and Bd1763. The catalytic C-H dyad for each enzyme is indicated in yellow and blue, respectively. The N-terminal Bd1763 unique extension is depicted in red. This structural feature appears to form a more extensive groove in Bd1763 which is absent in Bd0553 (dashed lines). **B)** Electrostatic surface potential representation of Bd0553 and Bd1763. An acidic pocket is observed only in Bd1763, connected to the putative substrate binding groove suggesting an altered specificity of this LDT with respect to Bd0553 and Bd2061. Surface potential values vary between 10 and -10 Kcal/mol (blue to red, respectively).

4.2.2.2. LDT group 2

Group 2 comprises Bd3012, Bd3176, Bd3639 and the crystallised Bd3741. In contrast to group 1, group 2 LDTs share a limited sequence identity with respect to Bd3741, being 28% for Bd3012, 30% for Bd3176 and 26% for Bd3639. Sequence alignment is shown in **Figure 61**.



Figure 61: LDT group 2 sequence alignment. Amino acid sequence alignment of Bd3012, Bd3176, Bd3639 and Bd3741 is shown with similar residues highlighted in grey. Catalytic, substrate binding residues and disulphide forming cysteines are indicated with red, blue, and yellow dots, respectively. The N-terminal and flexible lid forming part of the capping subdomain in Bd3741 experimental structure are indicated with a purple and orange box, respectively. The residues forming part of the Bd3741 triangular hook are indicated with a green box. Amino acid numbering is relative to Bd3741 sequence.

As expected, residue conservation between Bd3741 and the other members of group 2 is mostly concentrated in the YkuD-like domain. Particularly, proposed Bd3741 substrate binding residues M150, F155, A163 and R192 (part of the (S/T)XGCϕRϕ YkuD-like crosslinking/exchanger motif) are conserved in the four members of this group. Further, Y141 appears to be a conserved position for an aromatic residue, being replaced by a histidine in Bd3176. Residue S143 is also conserved except in the most divergent member of the group, Bd3639.

The three characteristic regions identified in the Bd3741 structure, N-terminal-helix, active site lid and triangular hook are remarkably the most divergent sequence-wise. The N-terminus and active site lid (framed in purple and orange, respectively) are specific to each LDT showing little to no conservation in their sequences. Interestingly, Bd3176 possesses two non-catalytic cysteines: C37 in the N-terminus and C130 in the active site lid (indicated in yellow dots), which potentially form a disulphide bond, suggesting that both the N-terminus and the active site proposed lid form a capping subdomain for the enzyme, as was suggested for Bd3741 in Chapter 3.

The distinctive triangular hook of Bd3741 (framed in green), whose geometry is defined by the two glycine residues G221 and G229, appears to be conserved in Bd3176 and Bd3012, suggesting that these proteins could share this structural feature. In contrast, Bd3639 lacks an equivalent region in its amino acid sequence, reinforcing the hypothesis of Bd3639 being the outlier LDT of group 2.

To determine the structural impact of sequence divergence in group 2, the experimentally solved structure of Bd3741 was compared to the AlphaFold2 predicted structural models for Bd3176, Bd3012 and Bd3639 (**Figure 62A**, in red, cyan, grey, and magenta, respectively, all predicted as monomers with pLDDT scores of 85.1, 86.9 and 87.2, respectively).

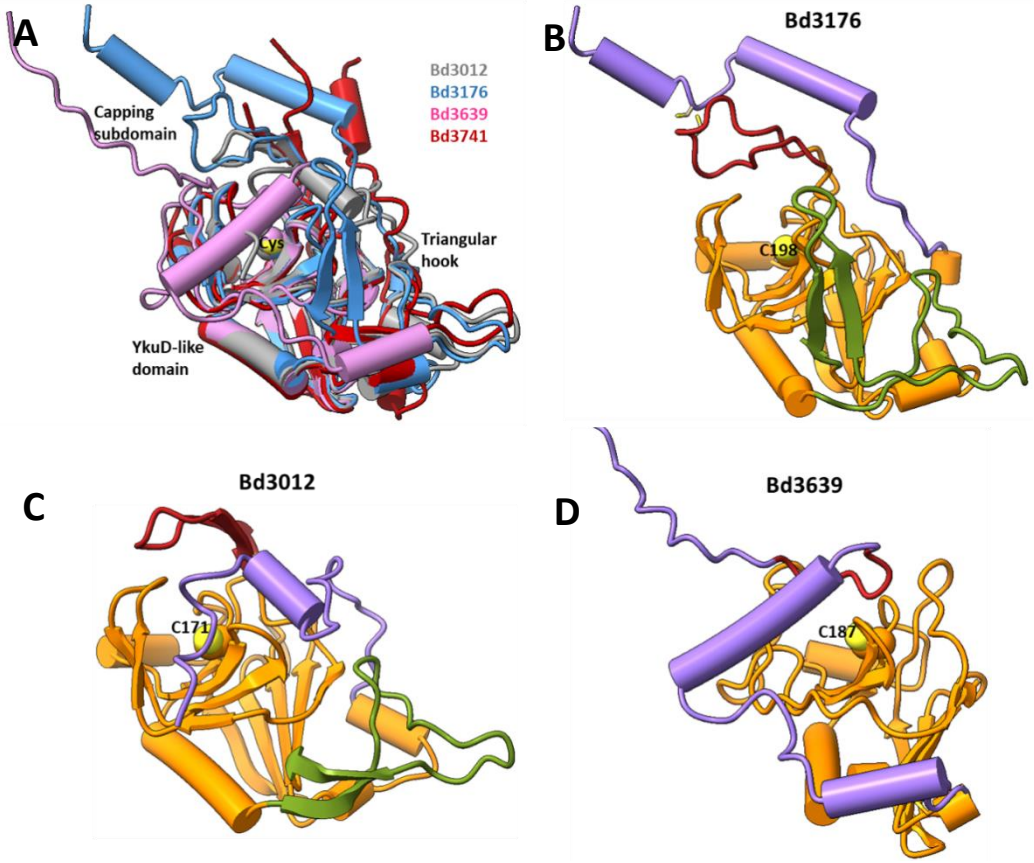


Figure 62: Structural comparison of group 2 of LDTs. A) Superimposition of the experimental structure of Bd3741 (in red) and the predicted structural models of Bd3012 (grey), Bd3176 (cyan) and Bd3639 (magenta). The YkuD-like domain corresponds to the most conserved portion of the structure among group 2. The substrate binding cleft is defined by the catalytic cysteine residue (yellow sphere) and is covered by a divergent capping subdomain in each member of this group. A triangular hook is depicted in models for Bd3012 and Bd3176 but not for Bd3639. **B)** Bd3176 AlphaFold2 predicted model. The YkuD-like domain is coloured in orange. The substrate binding cleft is centred around C198 and is capped by a predicted unstructured linker (in red). This lid interacts with Bd3176 N-terminal helical domain (in purple) via the C37-C130 disulphide bond. A triangular hook is predicted to form as in Bd3741 (in green), including a beta hairpin insertion between residues 243-259. **C)** Bd3012 AlphaFold2 structural model. Besides the conserved YkuD-like catalytic domain (in orange), a lid (in red) and a unique N-terminal (in purple) are observed, including the presence of a triangular hook (in green). **D)** Bd3639 structural model prediction by AlphaFold2. A unique N-terminal region (in purple) is depicted in the model. The substrate binding cleft of the YkuD-like domain (in orange) is capped by a loop (in red) and by the N-terminal region of the enzyme. The triangular hook depicted in the other members of this group is absent in Bd3639 structural prediction, which agrees with the amino acid sequence alignment in group 2.

Despite the comparably high overall RMSD values of superimposing Bd3012, Bd3176 and Bd3639 to Bd3741 (6.35, 4.76 and 16.1 Å, respectively), a similar YkuD-like fold is observed across all members of group 2. Moreover, the presence of a lid-like subdomain structure covering the substrate binding cleft is depicted in all models, even though they diverge between each protein. Interestingly, the distinctive triangular hook of Bd3741 is shared

by Bd3012 and Bd3176 despite the poor conservation of its amino acid sequence with exception of the equivalents of Bd3741 G221 and G229, all with pLDDT scores over 80, which is considered a reliable structural prediction [145]. As predicted by its sequence (**Figure 61**), Bd3639 lacks the triangular hook structure and could be then considered as the structural outlier of group 2.

Bd3176 is sequence-wise the closest member to Bd3741, which agrees with the predicted structure of this protein having the smaller overall RMSD value (**Figure 62B**). The predicted structure comprises a helical N-terminal structure (residues 25-63, in purple) which is covalently coupled to the active site lid (residues 122-138, in red) by the C37-C130 disulphide bond. This suggests that the N-terminus of the enzyme acts with the active site lid as a capping subdomain, as proposed for Bd3741.

The Bd3176 triangular loop comprises residues 220 to 242 (in green), adopting a similar geometry to the one observed in Bd3741. Interestingly, Bd3176 possesses an insertion between residues 243 and 259 before the final beta strand closing the re-entrant YkuD-like domain beta barrel. This insertion adopts an antiparallel two-stranded beta sheet topology and is exclusive to Bd3176.

Bd3012 corresponds to the second group 2 member most similar to Bd3741 (**Figure 62C**). The N-terminus of Bd3012 is predicted to be highly disordered with exception of a single alpha helix (residues 29 to 35). Residues 19-25 are modelled blocking the substrate binding cleft of the YkuD-like domain along with the predicted active site lid (residues 103-114, in red). The triangular loop is also present in Bd3012 (residues 196-213, in green) whose geometry appears to be defined by the two conserved glycine residues G201 and G209.

In this sense, structural comparison of Bd3012, Bd3176, Bd3741 and the triangular loop depicted in Ypr118w (PDB ID: 1W2W, residues 339-366) appear to indicate that an eight residues-long spacer between two glycine residues is at least one of the requirements to acquire the triangular geometry. In fact, despite their sequence divergence, a GX_4N/DX_2G motif is shared between the aforementioned structures.

Bd3639 could potentially be considered the structural outlier of group 2 due to its lack of the triangular hook domain (**Figure 62D**). Bd3639 includes an extensive N-terminal region (residues 25-85, in purple) that includes the active site lid (residues 39-45, in red) and an

alpha helix blocking access to the substrate binding cleft (residues 49-63). Despite the lack of the triangular hook domain, a second N-terminal alpha helix (residues 73-81) is predicted to be positioned on the same side of the YkuD-like beta barrel.

Surface analysis of group 2 of LDTs is shown in **Figure 63**. As seen in the cartoon depiction of the structural models, the substrate binding cleft of the enzymes is partially or completely blocked by the active site lid (in red) (**Figure 63A** and **B**) and additionally by the N-terminus alpha helices of Bd3012 and Bd3639 (in purple) (**Figure 63C** and **D**, respectively). The electrostatic surface potential depiction of each model differs for each enzyme, being mostly basic for Bd3741 and acidic for Bd3176 and Bd3012.

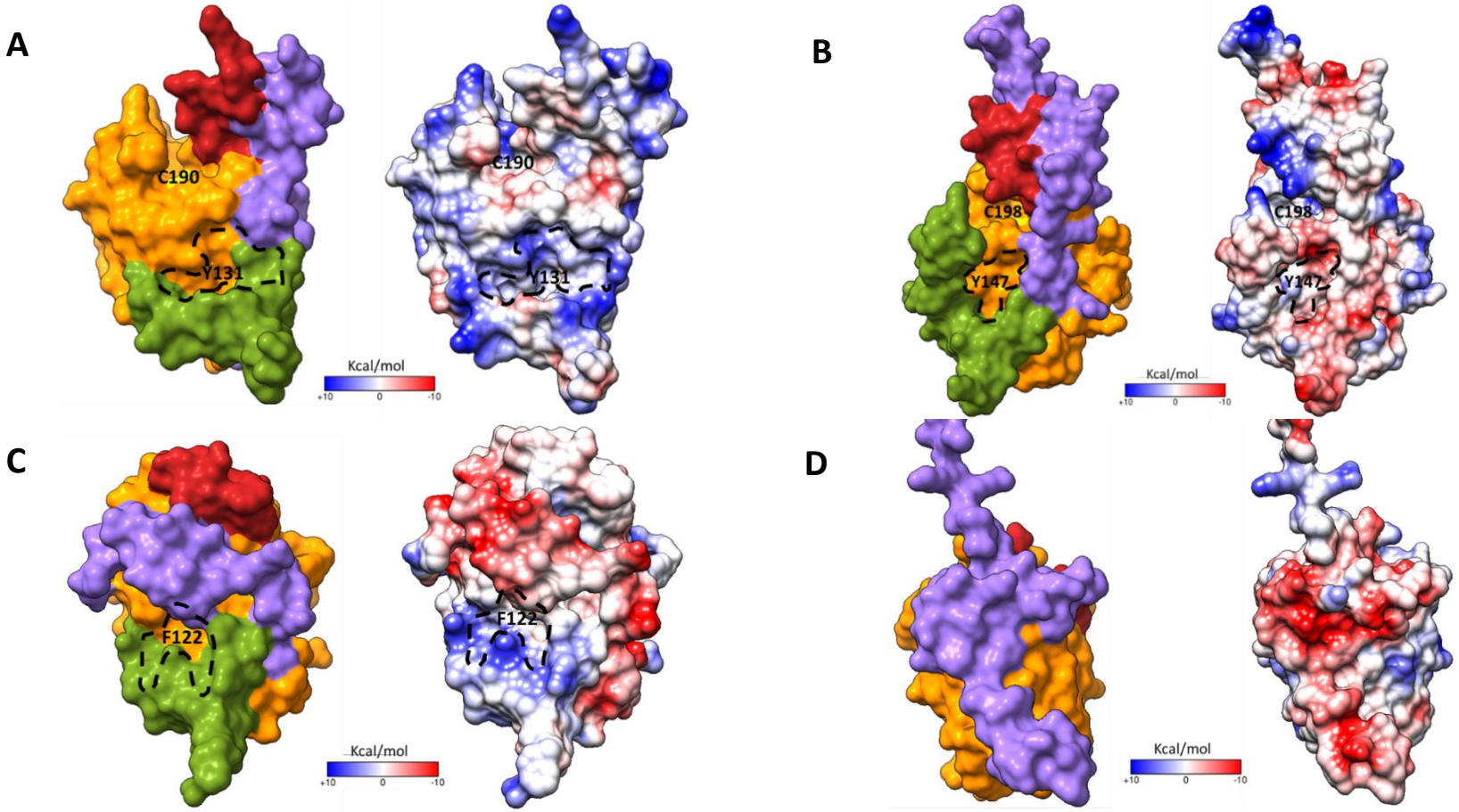


Figure 63: Surface analysis of group 2. Group 2 LDTs are depicted in surface representation and coloured either by structural domains (N-terminal in purple, active site lid in red, triangular hook domain in green and YkuD-like domain in orange) or by coulombic surface potential. **A)** Bd3741, **B)** Bd3176, **C)** Bd3012 and **D)** Bd3639. The triangular loop of Bd3741, Bd3176 and Bd3012 form pocket in the interface with the YkuD-like domain beta barrel around a conserved aromatic residue (Y131, Y147 and F122, respectively).

Remarkably, the triangular hook domain containing enzymes Bd3741, Bd3176 and Bd3012 (in green), show that a hydrophobic pocket is formed in the interface between the aforementioned domain and the side of the YkuD-like domain beta barrel. Furthermore, despite the poor conservation of the triangular hook domain residues, the observed interface pockets are centred around a tyrosine or phenylalanine residue (Y131, Y147 and F122 in Bd3741, Bd3176 and Bd3012, respectively) that is absent in the structural outlier Bd3639. The Bd3741 FTSite binding site prediction server of this pocket is shown **Figure 64** (https://ftsite.bu.edu/) [156].

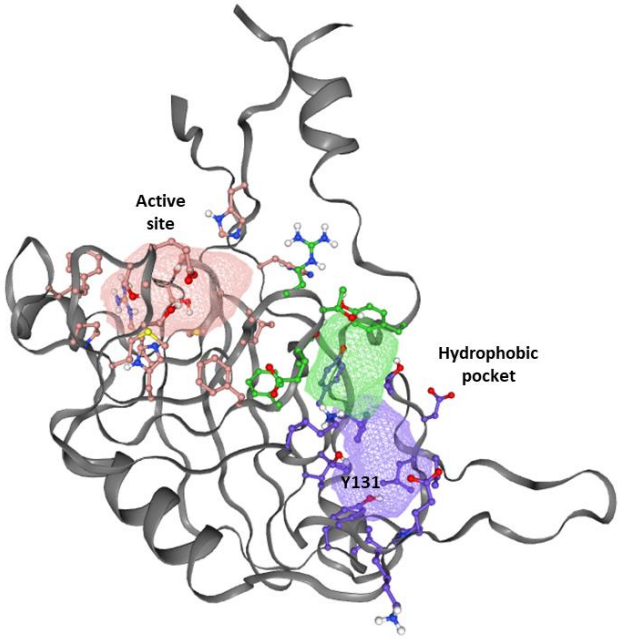


Figure 64: Predicted hydrophobic pocket of Bd3741. The proposed hydrophobic pocket centred on Y131 formed by the triangular hook of Bd3741 was predicted by the FTSite binding site prediction server. FTSite additionally identifies the active site of the enzyme.

In view of the evidence provided by the conserved W306 in the NTF2-like domain pocket of Bd1075, it is possible for this triangular hook domain and the pocket formed with the YkuD-like domain to participate in ligand and/or substrate binding. A first approach to assign a function to this novel domain would be identifying the impact of deleting the triangular hook on the activity of Bd3741 and related enzymes.

4.2.2.3. LDT group 3

Group 3 comprises the SH3b-domain LDTs Bd0886 and Bd1176, which share a sequence identity of 60.5%, suggesting they adopt an almost identical fold.

Sequence alignment in **Figure 65** shows that these enzymes are most similar in the SH3b and YkuD-like domains (framed in green and orange, respectively), wherein variations are mainly observed in the longer N- and C-terminuses of Bd1176 respect to Bd0886. As indicated, catalytic residues (red dots), proposed substrate binding residues (blue dots) and disulphide forming cysteine residues (yellow dots) are conserved between both proteins, including R303 of Bd0886, which is part of the PG transpeptidase/exchanger motif (S/T)XGCϕRϕ.

The dimerization residues observed in the Bd0886 crystallographic packing (framed in purple) are partially conserved in Bd1176, indicating that it could also form a tail-to-tail dimer. Interestingly, no confident dimer prediction was reported for Bd0886 nor Bd1176, (AlphaFold-multimer-v2 scores below 40%).

An insertion between positions 336 and 337 relative to Bd0886 is observed in Bd1176. Further discrete one to four residues long discrepancies are observed throughout the Bd1176 sequence with respect to Bd0886, corresponding mainly due to unstructured or flexible loops in the Bd0886 experimental structure.

The putative Bd0886 active site lid (residues 280-299) is almost completely conserved in Bd1176. This in addition to the conservation of the SH3b binding domain suggests that both members of group 3 recognise a similar if not the same type of substrate. Interestingly, residues 122-130 of Bd0886, which connect the SH3b domain to the YkuD-like domain are only partially conserved, including the substitution of residue P126 for an alanine in Bd1176, potentially increasing the conformational freedom of the SH3b-YkuD-like domains connecting loop in Bd1176.

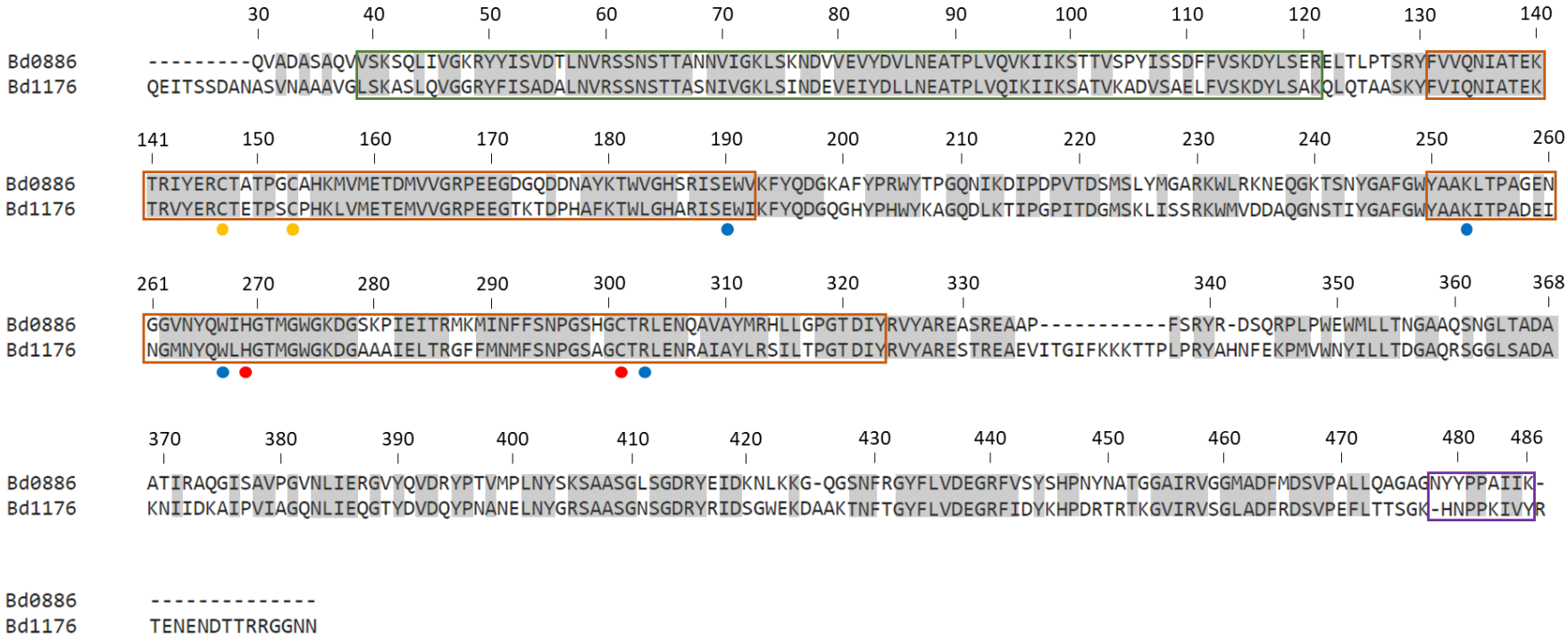


Figure 65: LDT group 3 sequence alignment. Amino acid sequence alignment of Bd0886 and Bd1176 is shown with similar residues highlighted in grey. Catalytic, substrate binding residues and disulphide forming cysteines are indicated with red, blue, and yellow dots, respectively. SH3b domain residues are framed in green. YkuD-like domain residues are framed in orange. Dimerization residues observed in the Bd0886 ASU are framed in purple. Non-framed residues correspond to structural insertions exclusive to group 3 of LDTs. Amino acid numbering is relative to Bd0886 sequence.

Structural superimposition of the experimental Bd0886 structure (red) with the AlphaFold2 model of Bd1176 (grey, overall pLDDT score of 85.8) is in accordance with the sequence alignment information, depicting a conserved general fold with an overall RMSD of 3.79 Å (**Figure 66A**).

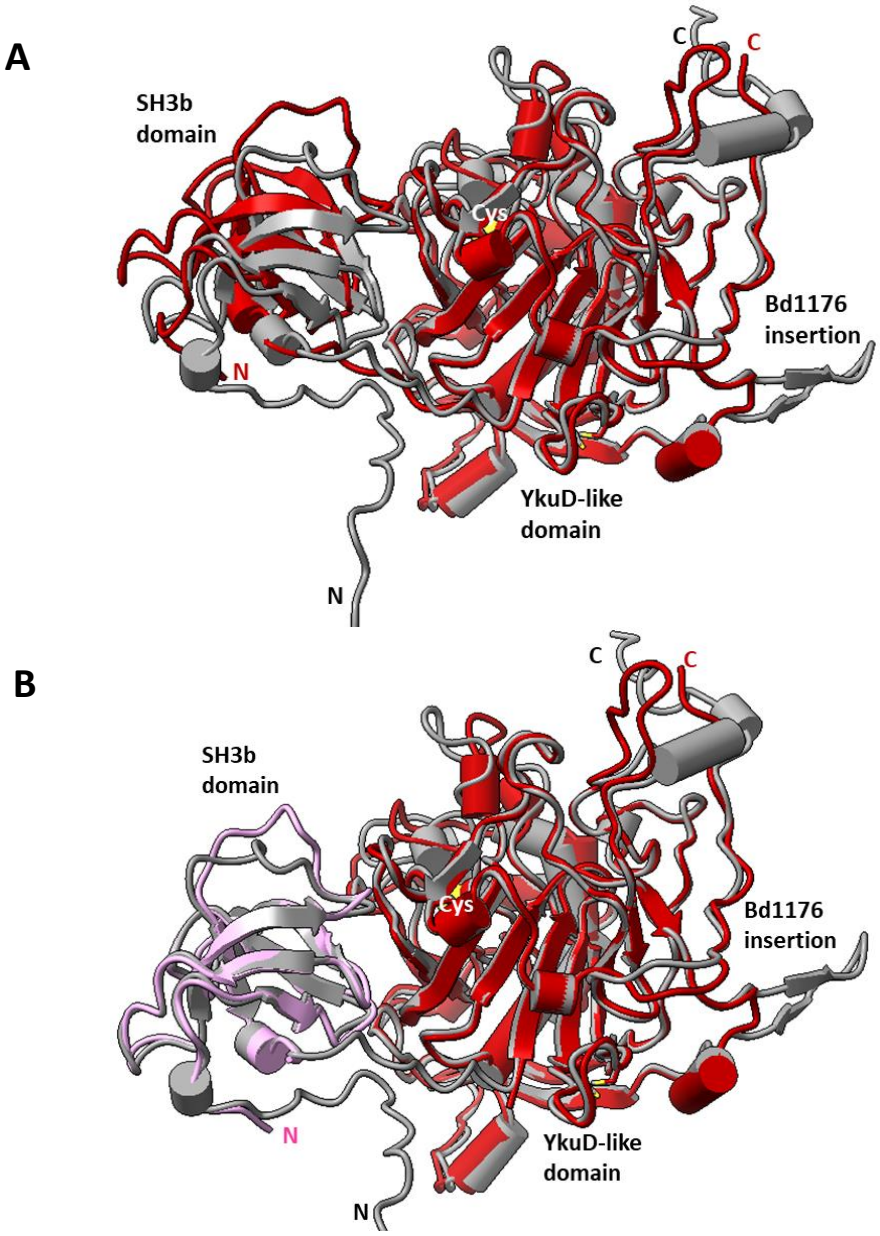


Figure 66: Structural comparison of group 3 of LDTs. A) Structural superimposition of Bd0886 (red) and Bd1176 (grey). SH3b and YkuD-like domains are indicated. Bd1176 differs from Bd0886 crystallographic structure in the SH3b spatial orientation and possessing a beta hairpin structural insertion. Catalytic cysteines are depicted in yellow sphere representation. **B)** Individual structural superimposition of the SH3b and YkuD-like domains of Bd0886 (pink and red, respectively) and the Bd1176 predicted structure (grey). Orientation differences of the SH3b domain of each protein are due to the conformational freedom of the loop connecting the aforementioned domain with the rest of the structure. Catalytic cysteines are depicted in yellow sphere representation.

Structural differences of Bd1176 with respect to Bd0886 are observed mainly in the unstructured loops, in contrast to the high structural conservation of the YkuD-like domain of the enzyme. Remarkable differences include the longer N- and C-terminuses of Bd1176 and a beta hairpin insertion between residues 336 and 337 relative to Bd0886.

Of particular interest is the reorientation of the SH3b domain in Bd1176 with respect to Bd0886. This alteration arises from modelling the loop connecting the SH3b with the rest of Bd1176 in a different conformation compared to the Bd0886 crystallographic structure (pLDDT score for the Bd1176 loop prediction < 50). As shown in **Figure 66B**, no structural differences are observed between the SH3b domains of each enzyme (pink and grey for Bd0886 and Bd1176, respectively), which are superimposable with an RMSD of 1.60 Å.

SH3b domains have been implicated in PG binding in cell wall hydrolases [106], [122], [124], [157]. Most biochemical data on PG binding by SH3b domains is referred to lysostaphins, and are proposed to work as targeting domains to localise the hydrolytic activity of the enzyme [122], [124], [157], [158]. Nevertheless, for the PG cysteine endopeptidases (PCPs) of *Anabaena variabilis* and *Nostoc punctiforme*, it has been suggested that the SH3b domain has a role in substrate specificity due to it forming a continuous pocket with the AvPCP and NpPCP active sites regardless of the possible conformational fluctuations of the loop connecting the catalytic domain of the enzyme with the SH3b domain[159].

As shown in **Figure 67A**, the SH3b domain sequence conservation is limited to the *Bdellovibrio* genus, suggesting that this domain potentially recognises a PG modification generated by the predator. Alignment shows that besides the hydrophobic core, polar surface exposed residues are conserved in Bd1176 (*B. bacteriovorus* HD100), Bdt1155 (*B. bacteriovorus* Tiberius), AZI87 01895 (unclassified *B. bacteriovorus*), DOE51 05470 (*Bdellovibrio* sp. NC01) and DOM22 14485 (*Bdellovibrio* sp. ZAP7).

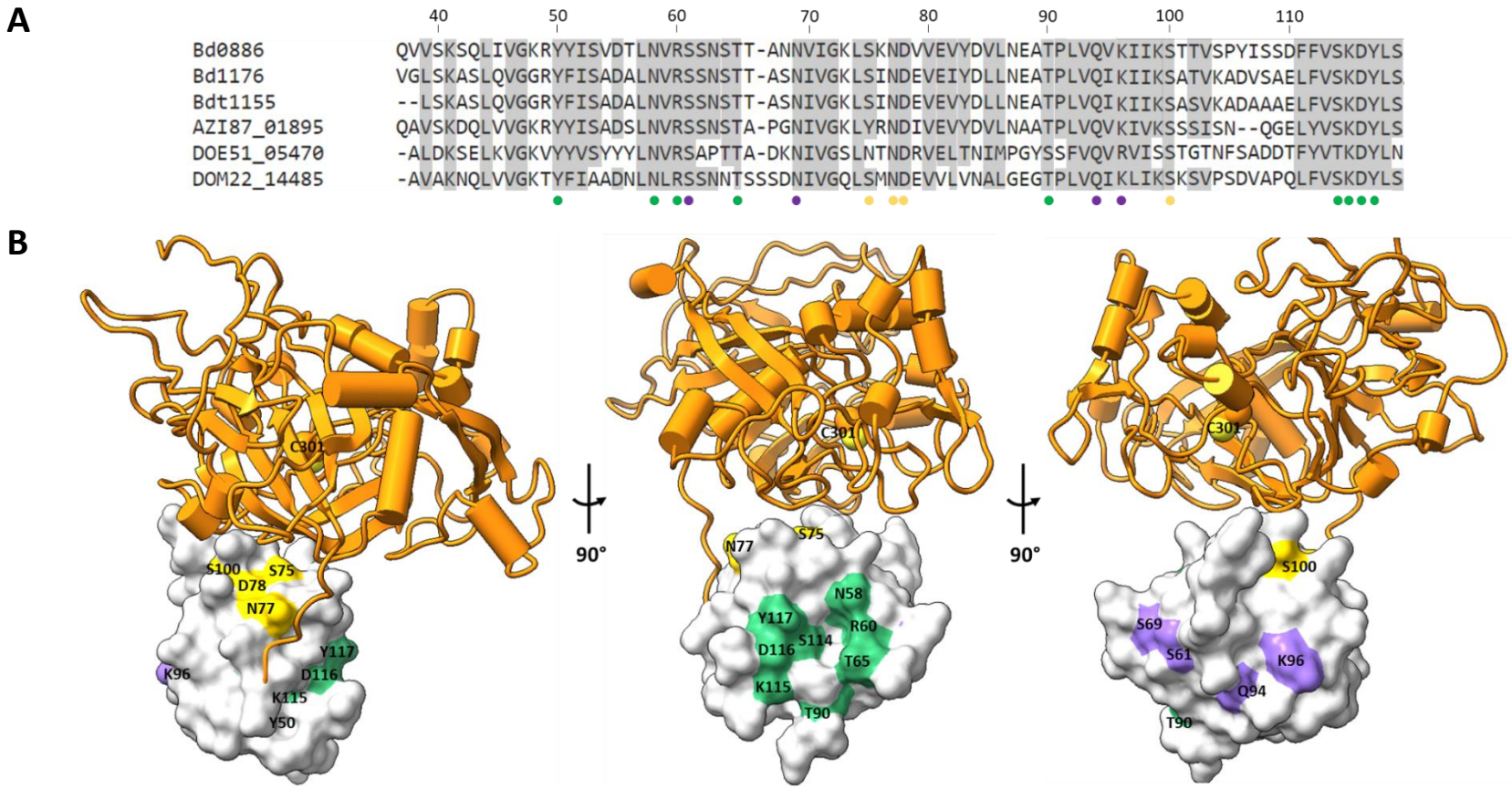


Figure 67: SH3b domain conservation and possible ligand binding pockets. A) SH3b sequence alignment between Bd0886 homologues, Bd1176 (*B. bacteriovorus* HD100), Bdt1155 (*B. bacteriovorus* Tiberius), AZI87 01895 (unclassified *B. bacteriovorus*), DOE51 05470 (*Bdellovibrio* sp. NC01) and DOM22 14485 (*Bdellovibrio* sp. ZAP7). Conserved residues are highlighted in grey. Conserved residues forming part of possible ligand binding pockets are indicated in yellow, green, and purple dots, depending on each pocket. Amino acid numbering is respect to Bd0886 sequence. **B)** Tree 90-degree apart views of SH3b surface representation of possible ligand binding pockets in Bd0886. Conserved residues indicated in panel (A) are coloured for each proposed pocket. The YkuD-like domain of Bd0886 is depicted in cartoon representation.

Conserved polar residues form possible ligand binding pockets on the surface of the SH3b domain (**Figure 67B**). Residues Y50, N58, R60, T65, T90, S114, K115, D116 and Y117 form the most extensive groove on the SH3b domain surface (in green), suggesting its potential for PG binding.

Two secondary potential ligand binding sites are formed by S75, N77, D78 and S100 (in yellow) and S61, S69, Q94 and K96 (in purple), a feature that has been observed in other SH3b domains despite their lack of binding sites similarity [124], [157], [160]. Interestingly, no conserved residues are located in the SH3b domain and YkuD-like domain interface, including in the substrate binding cleft of the enzyme, possibly suggesting that the SH3b binding domain may bind more distant regions of the putative PG substrate, as observed for the YcbB PGBD, or determine enzyme localisation [83].

4.2.2.4. LDT group 4

Group 4 includes the YcbB-like LDTs Bd0566, Bd1181, Bd1358 and Bd2908. Sequence identities respect to Bd0566 of each member are 29.26% for Bd1181, 46.15% for Bd1358 and 48.55% for Bd2908, suggesting that Bd1181 is the outlier of the group.

As shown in **Figure 67**, Bd1181 and Bd1358 have a longer N-terminus with respect to Bd0566 and Bd2908 (42 and 13 residues, respectively). Interestingly, the Bd1358 N-terminus is predicted to be disordered in contrast to the longer Bd1181 N-terminus that may potentially acquire a defined three-dimensional structure.

The scaffolding domain of group 4 (framed in black) is the most variable portion of the sequence alignment between the different members, however hydrophobic residue conservation suggests that this domain possibly adopts a similar fold in all four members. Nevertheless, Bd1181 has a 42 residues-long N-terminal extension between positions 27 and 49 relative to Bd0566 sequence numbering.

Interestingly, the disulphide bond between helices 7 and 11 of Bd0566 (C155-C237) is only conserved in the closest member of group 4, Bd2908, whereas it's substituted by a leucine-leucine hydrophobic contact in Bd1358 and by a glutamine-valine pair in Bd1181.

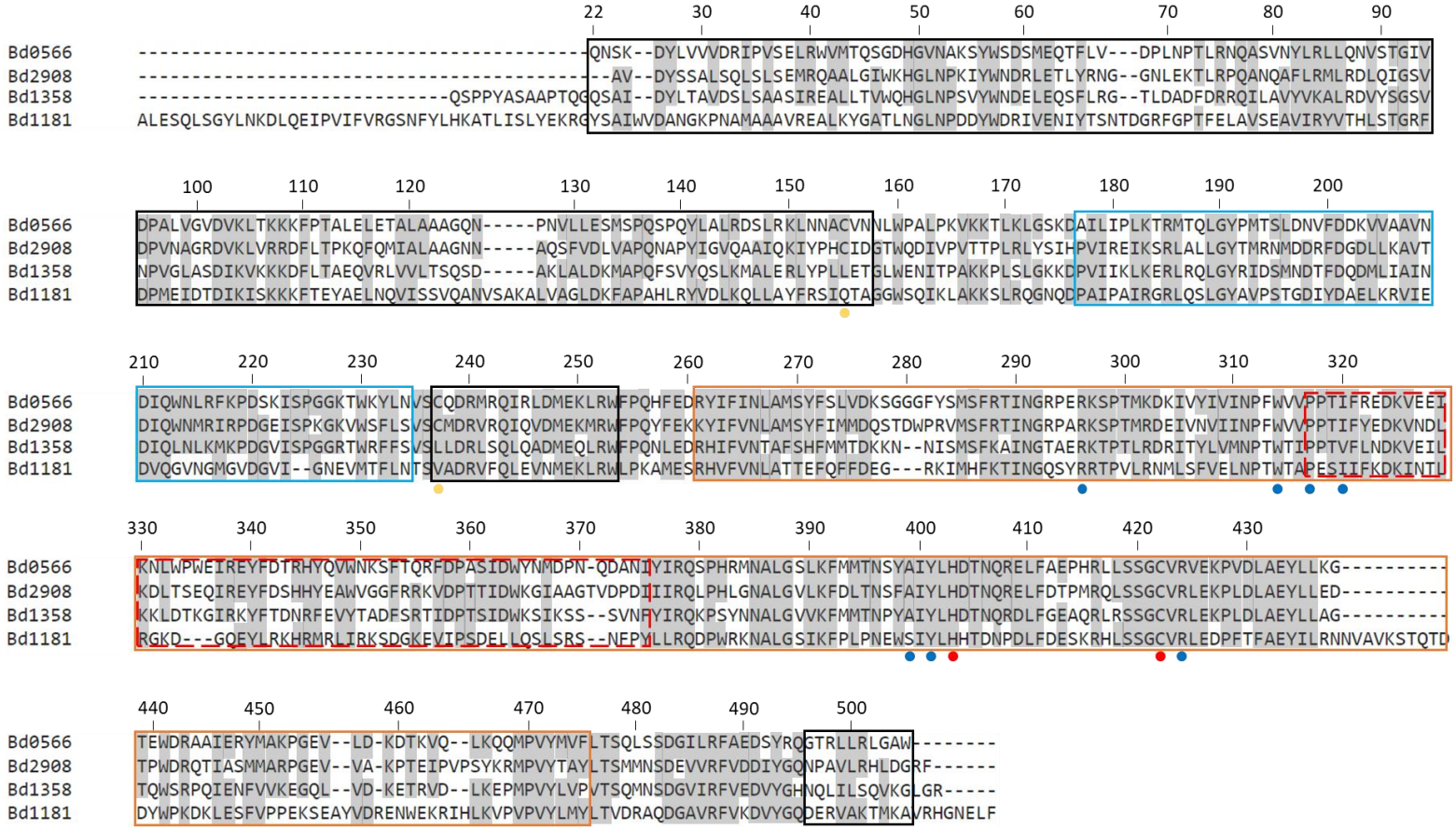


Figure 68: LDT group 4 sequence alignment. Amino acid sequence alignment of Bd0566, Bd2908, Bd1358 and Bd1181 is shown with similar residues highlighted in grey. Catalytic, substrate binding residues and disulphide forming cysteines are indicated with red, blue, and yellow dots, respectively. Scaffolding domain residues are framed in black. YkuD-like domain residues are framed in orange, wherein the capping subdomain is further framed with red dashed lines. PGBD residues are framed in cyan. Amino acid numbering is relative to Bd0566 sequence.

The group 4 PGBD (framed in cyan) is mostly conserved in all members, excepting the two unstructured loops between residues 213-216 and 227-229 of Bd0566, which have a different amino acid composition in Bd1181. Residue conservation extends to residues K171 and D220, which are hypothesised to participate in PG binding [83]. D220 in particular has been shown to be critical for *in vivo* activity of *E. coli* YcbB, residue that is conserved in all members of group 4, suggesting a similar role for this domain in *B. bacteriovorus* YcbB-like LDTs.

The YkuD-like domain of group 4 (framed in orange) is mostly conserved, including the proposed substrate binding residues (indicated in blue dots) and catalytic residues (indicated in red dots), with exception of the unstructured loops present in this domain, especially for Bd1181 which possesses two sequence divergent and longer loops, the first between positions 437-440, and the second between 451-464 relative to Bd0566 amino acid sequence. Nonetheless, all four members of group 4 possess the proposed crosslinking motif (S/T)XGCϕRϕ [118], [142].

Remarkably, the capping subdomain (framed in red dashed lines) is particularly variable for the unstructured regions between positions 333-336, 350-356 and 368-374 relative to Bd0566, however overall sequence identity is retained for the secondary structural elements of this region. This extends to the highly conserved residues D324 and K325, whose sidechains are projected towards the Bd0566 substrate binding groove and could potentially participate in substrate recognition. Overall, sequence identity suggests a similar structure and catalytic mechanism for all members of group 4 of LDTs.

To further compare LDTs of group 4, the crystallographic structure of Bd0566 was compared to the AlphaFold2 structural predictions of Bd2908, Bd1358 and Bd1181 (**Figure 69A**, Bd2908 in grey, Bd1358 in white and Bd1181 in cyan, with pLDDT scores of 85.4, 82.7 and 83.9, respectively).

Predicted models are highly superimposable to the experimentally determined structure of Bd0566, with overall RMSD values of 1.79, 2.23 and 3.66 Å for Bd2908, Bd1358 and Bd1181, respectively. As hypothesised based on the sequence alignment of group 4, the general Bd0566 fold is retained in the other members of this group.

Nonetheless, Bd1181 is the most structurally divergent member of the group, particularly in the scaffolding domain organisation.

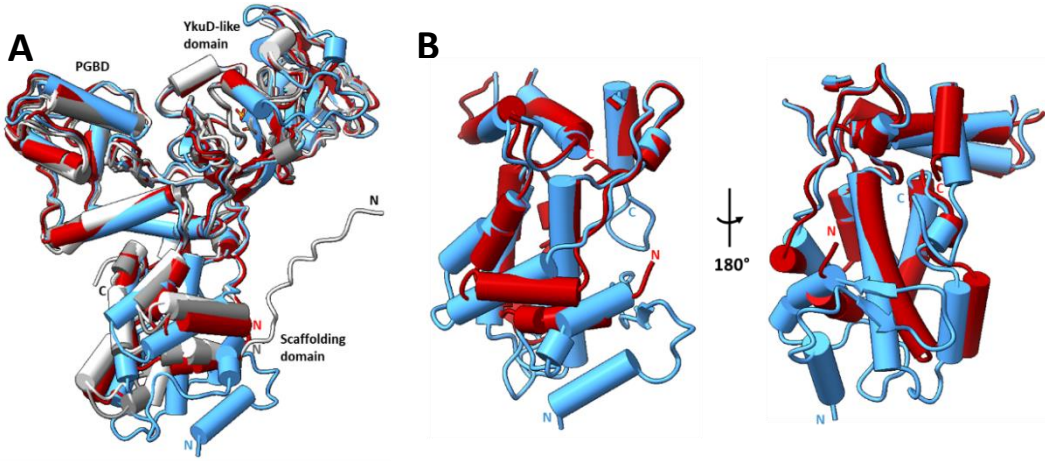


Figure 69: Structural comparison of group 4 of LDTs. A) Structural superimposition of the Bd0566 crystallographic structure (red) and the Alphafold2 predicted models of Bd2908 (grey), Bd1358 (white) and Bd1181 (cyan). **B)** Two 180° apart views of the superimposed scaffolding domains of Bd0566 (red) and Bd1181 (cyan). The N-terminus of Bd1181 is 42 residues longer than Bd0566 and is predicted to fold as a N-terminal helix connected to a beta hairpin.

The Bd1181 scaffolding domain predicted structure depicts a unique N-terminal extension comprising an N-terminal alpha helix followed by a beta hairpin structure (**Figure 69B**, Bd0566 in red, Bd1181 in cyan). Further, the first alpha helix of Bd0566 (residues 24-34) has a different amino acid composition in Bd1181, resulting in a different orientation of the equivalent helix in the Bd1181 predicted structure. In this regard, Bd1181 interacting partners could potentially differ from Bd0566, Bd1358 and Bd2908.

4.2.2.5. LDT group 5

LDT group 5 corresponds to the YkuD2-like domain comprising the lipobox enzymes Bd0782 and Bd1402, and the soluble enzymes Bd2940, Bd3093 and Bd3334. Sequence alignments were performed independently for the lipobox and soluble subgroups.

The lipobox Bd0782 and Bd1402 LDTs are the most similar in their amino acid sequence, sharing 55.5% sequence identity. As observed for other LDTs, the highest similarity is shared between the YkuD2-like domains of each enzyme (framed in orange), including the catalytic (red dots) and predicted substrate binding residues (blue dots) based on the Bd1402 crystallographic structure (**Figure 70A**).

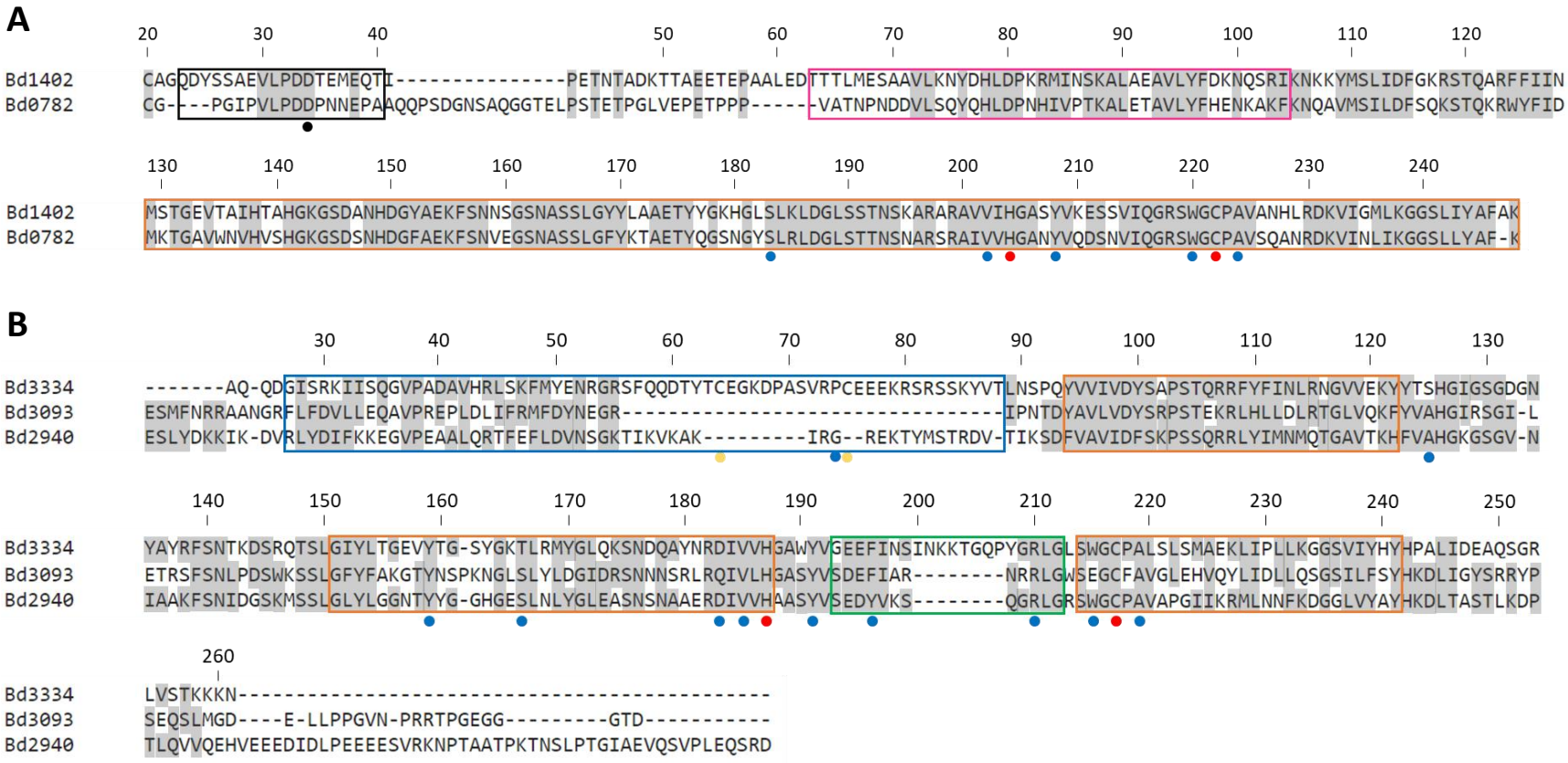


Figure 70: LDT group 5 sequence alignment. Similar residues are highlighted in grey. Catalytic, substrate binding residues and disulphide-forming cysteines are indicated with red, blue, and yellow dots, respectively. **A)** Amino acid sequence alignment of membrane bound LDTs Bd0782 and Bd1402 is shown. Active site lid residues are framed in black, with substrate specificity related residue D34 indicated with a black dot. Dimerization domain residues are framed in magenta. YkuD2-like domain residues are framed in orange. Amino acid numbering is relative to Bd1402 sequence. **B)** Amino acid sequence alignment of soluble LDTs Bd2940, Bd3093 and Bd3334 is shown. Structural inserts 1 and 2 are framed in blue and green, respectively. YkuD2-like domain residues are framed in orange. Amino acid numbering is relative to Bd3334 sequence.

The active site lid depicted in the Bd1402 structure (framed in black) shows a partial conservation of residues in Bd0782, particularly for D34 in Bd1402 (black dot), which orientates towards the substrate binding cleft and is hypothesised to impact substrate specificity.

On the other hand, the least conserved portion of the lipobox subgroup corresponds to the unstructured loop connecting the active site lid to the dimerization helix (position 41 to 62, relative to Bd1402). This loop is longer for Bd0782, comprising 32 residues in contrast to the 21 residues of Bd1402.

The dimerization domain of Bd1402 (framed in magenta) appears to be conserved in Bd0782, however, room temperature gel filtration data has shown that Bd0782 is monomeric in solution (20 mM Tricine pH 8.0 and 200 mM NaCl, data not shown). In this sense, the Bd1402 crystallographic structure shows that dimerization is achieved mainly by the hydrophobic interaction between L66 of one Bd1402 with L73 of the second chain. Sequence alignment shows only L73 is conserved in Bd0782. Furthermore, the equivalent helix in Bd0782 appears to be more charged, having two aspartic acid residues in positions 70 and 71 relative to Bd1402 amino acid sequence, which might further contribute to Bd0782 presenting as a monomer.

Nevertheless, dimerization might be dependent on the lipobox presence limiting diffusion by membrane anchoring to two dimensions. This effect might be potentially mimicked in by the higher effective concentration of protein in the crystal lattice compared to solution.

In contrast to the lipobox subgroup, the soluble subgroup of group 5 LDTs share a lower sequence identity relative to Bd3334, being 36.7% and 36.55% for Bd2940 and Bd3093, respectively.

Similar to other LDTs, the YkuD2-like domain of this subgroup (framed in orange) is the most conserved part of the sequence, including most of the structure-based predicted catalytic and substrate binding residues (red and blue dots, respectively) (**Figure 70B**). Bd2940 and Bd3093 have an eight-residue extension at the N-terminus for both enzymes, and significantly longer C-terminuses compared to Bd3334 (20 and 41 residues long for Bd3093 and Bd2940, respectively), which are both predicted to be disordered.

Sequence conservation analysis of the structural insertion 1 defined for Bd3334 (framed in blue) shows that this feature is partially conserved in Bd2940 and Bd3093. Particularly,

residues 27-54 of Bd3334 that fold as alpha helices α1 and α2 are conserved in Bd2940 and Bd3093 suggesting that this part of insert 1 is maintained in this subgroup. However, residues 55-88 that form the beta arm of Bd3334 are variable in Bd2940 and absent in Bd3093, including the proposed substrate binding residue P74 and the disulphide forming cysteine residues C64 and C75.

Structural insert 2 (framed in green) is also partially conserved in this subgroup, wherein a longer loop between positions 200 and 208 is present only in Bd3334. Nonetheless, predicted Bd3334 substrate binding residues F196 and specially R210 are conserved across this subgroup.

Finally, regardless of each enzyme being part of either the soluble or lipobox subgroup, all their YkuD2-like domains share the active site motif SXGCXAϕ, suggesting that all five members of this group have carboxypeptidase and/or endopeptidase activity.

To analyse the structural differences between the lipobox and soluble LDTs in group 5, the experimentally determined structures of Bd1402 (red) and Bd3334 (grey) were compared (**Figure 71**).

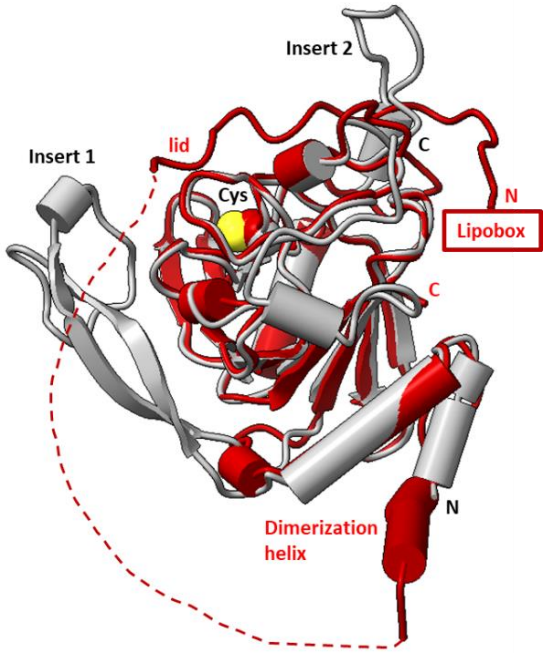


Figure 71: Structural comparison between lipobox Bd1402 and soluble Bd3334. The experimentally determined structures of Bd1402 (red) and Bd3334 (grey) were superimposed to analyse structural differences. Structural inserts 1 and 2 are unique to Bd3334 whereas the active site lid and dimerization helix are present only in Bd1402. The dashed line corresponds to the 23 residues-long missing loop in Bd1402.

Bd1402 and Bd3334 are superimposable with RMSD values of 0.68 Å limited to the YkuD2-like core and 6.76 Å across the whole Bd1402 structure. Key differences between both proteins correspond to the structural insertions 1 and 2 of Bd3334, and the lid and dimerization helix of Bd1402.

The Bd3334 active site lacks the lid depicted in the Bd1402 structure, making the substrate binding cleft more solvent accessible when compared to Bd1402. On the other hand, the Bd1402 dimerization helix is absent in Bd3334, which further supports the hypothesis that the three soluble LDTs from group 5 are intrinsically monomeric.

To analyse the impact of sequence divergence in the structure of LDTs of each subgroup, the experimentally determined structures of Bd1402 and Bd3334 were compared with the respective AlphaFold2 models of membrane bound and soluble LDTs.

As suggested by the significant sequence identity between Bd0782 and Bd1402, their respective predicted and experimentally determined structures are highly superimposable, with an RMSD of 0.51 Å when omitting the N-terminal loop of Bd0782 (which equivalent is not traceable in the Bd1402 structure) (**Figure 72A**).

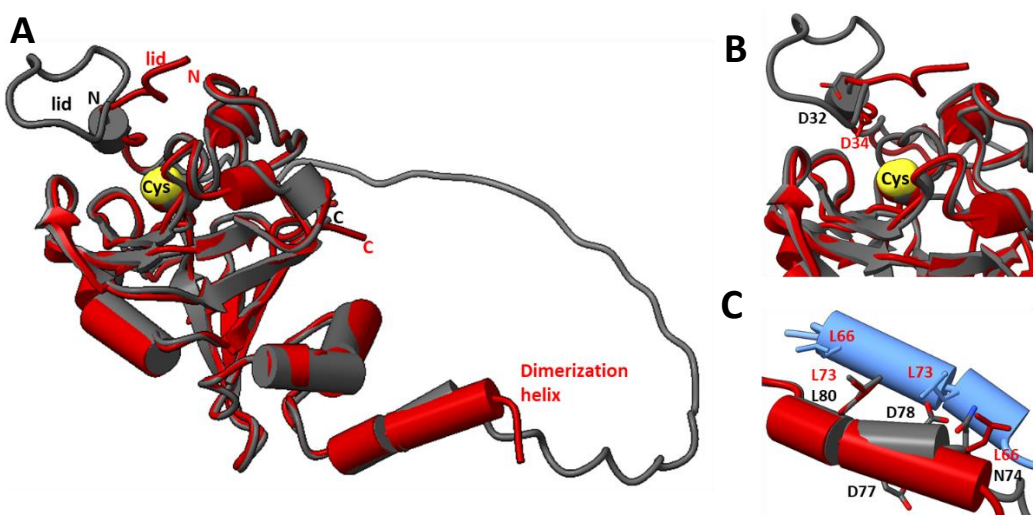


Figure 72: Bd0782 and Bd1402 structural comparison. A) Bd0782 (grey) and Bd1402 (red) structural superimposition. Active site and lid of both proteins and Bd1402 dimerization helix are indicated. **B)** Bd0782 and Bd1402 active site lids. Conserved D32 (Bd0782) and D34 (Bd1402) are indicated. **C)** Dimerization helix interface representation (Bd0782 in grey, Bd1402 chain A and B in red and blue, respectively). L66 which is part of the L66-L73 dimerization interaction in Bd1402 is substituted by residue N74. Non-conserved D78 in Bd0782 is oriented towards the helix-helix interface, possibly avoiding the formation of the Bd0782 dimer due to electrostatic repulsion.

Bd0782 (grey, overall pLDDT of 80.9, <60 for active site lid, >90 for YkuD2-like domain) and Bd1402 (red) structures comprise the active site lid connected to the helix responsible for dimerization in Bd1402 by a long unstructured loop. The first residue of the active site lid corresponds to the membrane anchoring cysteine residue in each enzyme (C22 and C20 in Bd0782 and Bd1402, respectively), suggesting that the substrate binding cleft is oriented towards the outer membrane inner leaflet, further limiting substrate accessibility. Bd0782 structural predictions are inaccurate in determining the most probable conformation of the active site lid, however, residue D32 (corresponding to D34 in Bd1402) is predicted to be oriented towards the active site (**Figure 72B**).

Main structural differences between Bd0782 and Bd1402 correspond to the dimerization helix (**Figure 72C**). Specifically, residues L66 and L73 have been proposed as the dimerization contacts in Bd1402, however, L66 is substituted for N74 in the Bd0782 predicted structure. Moreover, as suggested based on the sequence conservation analysis of both proteins, the two acidic residues D77 and D78 might contribute to further repulsion between the corresponding Bd0782 helix. In fact, Bd0782 D78 is predicted to be situated in the equivalent Bd1402 dimerization interface, favouring the monomeric state of Bd0782 over a homodimer, as observed for Bd1402.

The soluble subgroup is also significantly similar, with overall RMSD values of 2.59 and 2.33 Å respect to Bd3334 for Bd2940 and Bd3093, respectively. Structural superimposition is shown in **Figure 73A**.

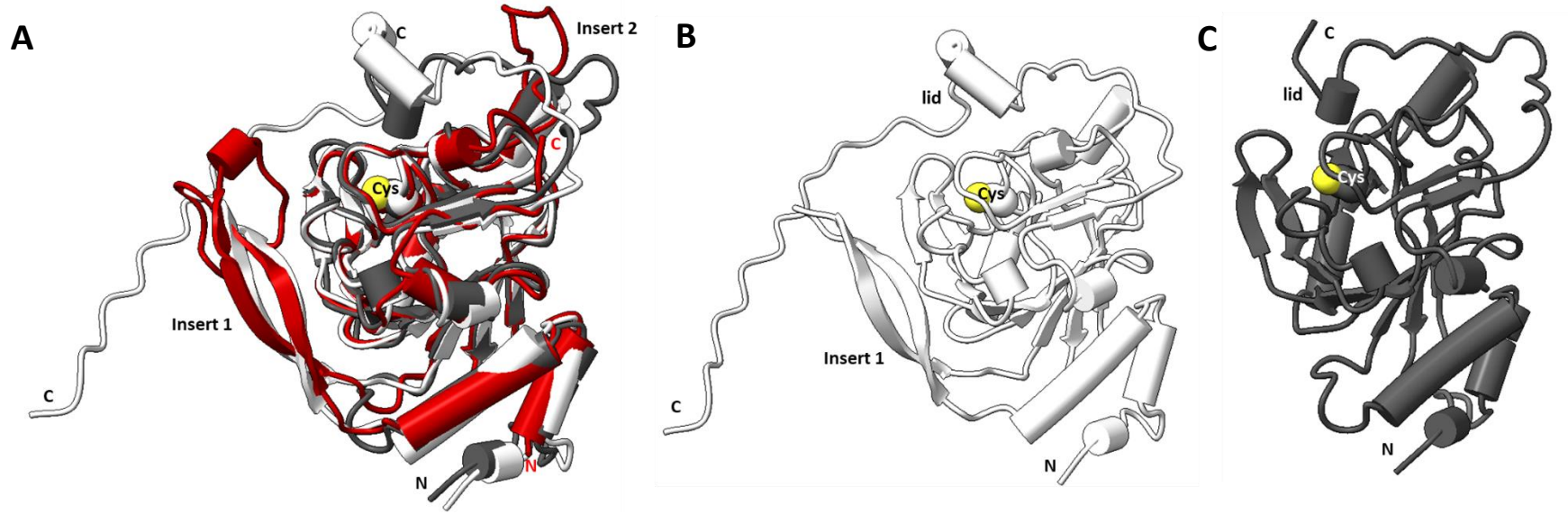


Figure 73: Bd2940, Bd3093 and Bd3334 structural comparison. A) Superimposition of predicted structures of Bd2940 (white) and Bd3093 (grey) to the experimentally determined structure of Bd3334 (red). Structural insert 1 is present in Bd2940 and Bd3334. Longer N- and C-terminuses are observed for Bd2940 and Bd3093 respect to Bd3334. **B)** Bd2940 predicted structure. An insert 1 equivalent to Bd3334 is depicted in the model, with a long C-terminus threading back to cover the substrate binding cleft, possibly working as a lid. **C)** Bd3093 predicted structure. No conserved Bd3334 structural insertions are observed in the model. The longer C-terminus threads back to cover the active site forming a lid, as observed in Bd2940 predicted structure.

As proposed based on the sequence alignment, Bd2940 (white) and Bd3093 (grey) possess longer N-terminuses compared to Bd3334 (red), predicted to acquire an alpha helix structure.

Noticeably, both Bd2940 and Bd3093 possess a long and unstructured C-terminus that is predicted to thread back toward (covering) the active site, which is reflected in pLDDT scores for the Bd2940 and Bd3093 AlphaFold2 models being >90 for most of the model, excepting the unstructured loops, with scores <60. The proposed Bd2940 and Bd3093 active site lids differ in amino acid sequence, being distinctively acidic for Bd2940 (**Figure 70**).

Regarding the distinctive structural features defined for Bd3334, structural insert 1 is partially conserved in Bd2940 and Bd3093, particularly the first two alpha helices of this structural feature (residues 27-54 in Bd3334). In this regard, the beta arm observed in the Bd3334 crystallographic structure (residues 55-88) is structurally conserved only in Bd2940, being absent in Bd3093 (**Figure 73 B** and **C**, respectively).

The structural conservation of the beta arm in Bd2940 and Bd3334 might suggest that both enzymes share a similar substrate specificity. Nevertheless, the different amino acid sequences of the beta arm of both proteins suggests that this feature determines specificity by sterically limiting access to the active site of the enzyme rather than by specific substrate-side chain interactions.

Finally, the Bd3334 basic loop of structural insert 2 is not conserved in either Bd2940 or Bd3093, as predicted based on the sequence alignment of this subgroup.

4.2.3. LDT intergroup analysis

4.2.3.1. YkuD-like domains of L,D-transferases

B. bacteriovorus HD100 LDTs may be divided according to their proposed general activity based on the active site motif, as either L,D-transferases (transpeptidases and NCDAA exchangers) or L,D-hydrolases (endopeptidases and carboxypeptidases).

Sequence alignment of the YkuD-like domains of the L,D-transferases Bd0566, Bd0886 and Bd3741 shows that sequence conservation is poor outside the signature motif (S/T)XGCϕRϕ (**Figure 74A**). This is indicative of different substrate specificity for each group of L,D-transferases, despite sharing a structurally similar catalytic domain.

Remarkably, an additional arginine residue different from the one contained in the L,D-transferase motif is conserved in group 2 Bd3741, group 3 Bd0886, and group 4 Bd0566 (R98, R167 and R293, respectively), indicated with a blue arrowhead). This conserved residue is exclusive to L,D-transferases, thus it might participate in binding the acyl acceptor substrate.

Structural superimposition of the Bd3741, Bd0886 and Bd0566 YkuD-like domains reveals that the abovementioned arginine residues are structurally conserved in L,D-transferases (**Figure 74B**). Further, these conserved arginine residues are indeed oriented towards the putative acyl acceptor site, suggesting that they participate in selecting a nucleophile as second substrate in contrast to the water molecule utilised as base by L,D-hydrolases.

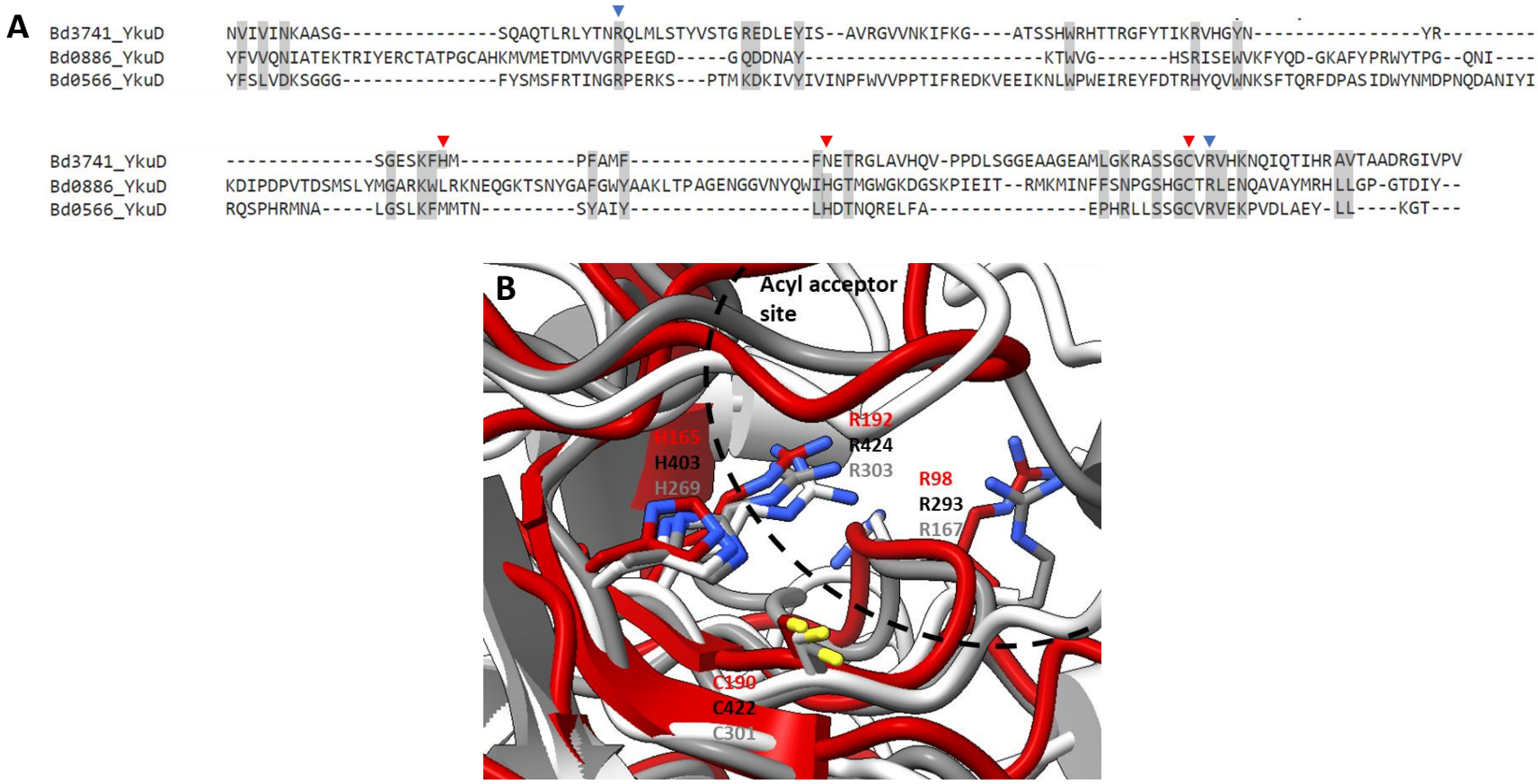


Figure 74: YkuD-like L,D-transferases active site comparison. A) Sequence alignment between the YkuD-like domains of L,D-transferases group 2 Bd3741, group 3 Bd0886, and group 4 Bd0566. Catalytic residues are indicated with red arrow heads. Conserved arginine residues are indicated with blue arrow heads. **B)** Structural superimposition of YkuD-like domains of Bd0566 (grey), Bd0886 (white) and Bd3741 (red). Conserved residues corresponding to the L,D-transferase motif and the additional arginine are depicted in stick representation (R98, R293 and R167 for Bd3741, Bd0566 and Bd0886, respectively). Both conserved arginine residues are oriented towards the acyl acceptor site of the substrate binding cleft.

Additionally to the conserved arginine residues, experimentally determined structures and AlphaFold2 model predictions show that every *B. bacteriovorus* HD100 L,D-transferase possess an active site lid with variable amino acid sequences and structures. Specifically, group 4 of LDTs possess a capping subdomain that is required for *in vivo* activity in the *E. coli* YcbB ortholog [83].

Active site lids of the *M. tuberculosis* Ldt_{Mt1} and Ldt_{Mt2} have been implicated in the transpeptidation mechanism [76]–[79]. Specifically, catalytic cysteine inactivation with carbapenem β-lactams, such as meropenem, or with 2-phenyl-1,2-benzisoselenazol-3(2H)-one (ebselen) traps the enzyme in the probable acyl-enzyme intermediate state, wherein the active site lid displays extensive hydrophobic interactions with the drug adduct, making the active site inaccessible to solvent. Thus, it is possible that active site lids protect the acyl-enzyme intermediate from hydrolysis by excluding water molecules from the substrate binding pocket.

Thus, the two conserved acyl-acceptor site arginine residues in addition to the presence of a lid are possibly the main features determining the putative L,D-transferase activity of groups 2, 3 and 4 of LDTs, while substrate specificity is potentially defined by the group-specific binding residues in the substrate binding cleft and active site lid sequence and structure.

4.2.3.2. YkuD- and YkuD2-like domain of L,D-hydrolases

The *B. bacteriovorus* HD100 L,D-hydrolases are putatively identified by the hydrolytic motif SXGCXAϕ, wherein the signature L,D-transferase arginine residue is replaced by an alanine residue. This substitution possibly abrogates binding of a nucleophilic acyl acceptor substrate, favouring the hydrolysis of the acyl-enzyme intermediate instead.

Group 1 and group 6 of the LDTs are classified as YkuD-like L,D-hydrolases, wherein group 6 Bd1075 corresponds to the only non-predatory L,D-carboxypeptidase. Group 5 of the LDTs corresponds to the soluble and membrane-bound YkuD2-like L,D-hydrolases.

As shown in **Figure 75**, YkuD- and YkuD2-like domains differ in primary sequence despite retaining an almost identical fold, suggesting possible diversification in substrate specificity. Interestingly, overall alignment suggests higher similarity between the YkuD-like Bd0553 and the YkuD2-like Bd1402 and Bd3334 rather than with the YkuD-like Bd1075.

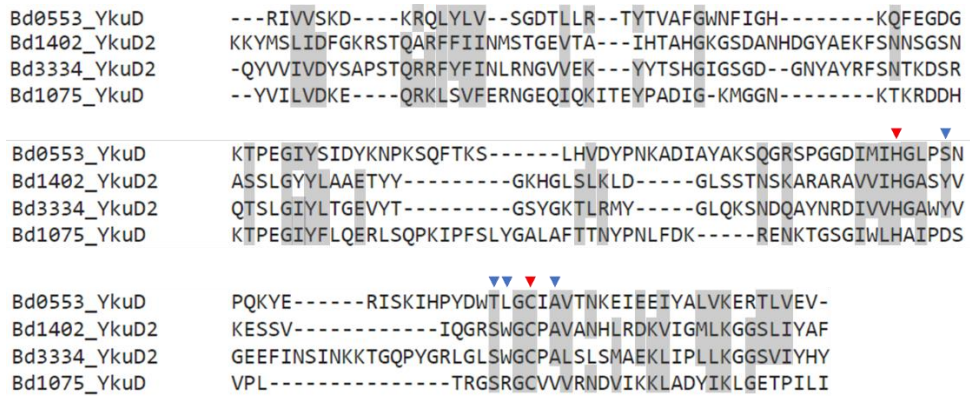


Figure 75: L,D-hydrolase catalytic domain sequence comparison. Sequence alignment between the YkuD-like domains of Bd0553 (predatory) and Bd1075 (non-predatory), and the YkuD2-like domains of Bd1402 (membrane-associated) and Bd3334 (soluble). Catalytic residues are indicated with red arrow heads. Signature substrate binding residues are indicated with blue arrow heads.

Furthermore, the signature alanine residue of the L,D-hydrolase motif is replaced by a valine residue in the non-predatory Bd1075 YkuD-like domain, which is a unique feature even among other Csd6-like cell shape determinants [125].

Due to the high sequence divergence between L,D-hydrolases, experimentally determined structures were superimposed to identify shared features between the YkuD- and YkuD2-like catalytic cores (**Figure 76**).

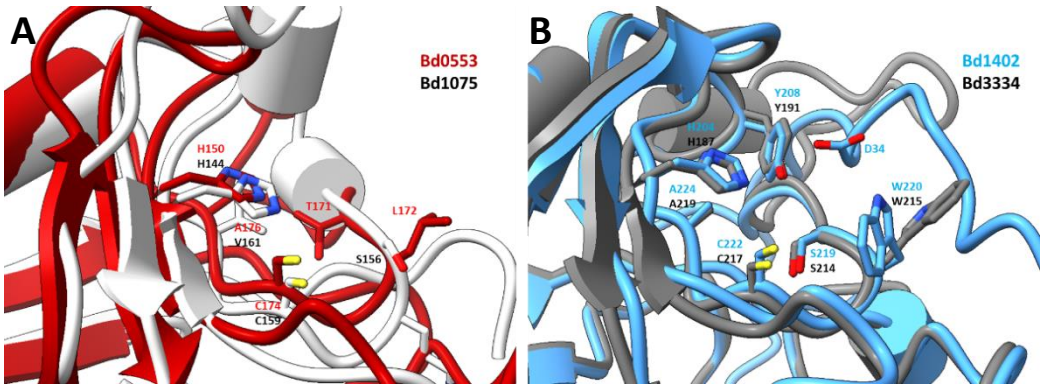


Figure 76: Structural comparison between L,D-hydrolases. A) Structural comparison between the YkuD-like domains of predatory Bd0553 (red) and non-predatory Bd1075 (white). Conserved residues are shown in stick representation and indicated in the structure. **B)** Structural comparison between the YkuD2-like domains of membrane-associated Bd1402 (blue) and soluble Bd3334 (grey). Conserved residues are shown in stick representation and labelled accordingly.

The YkuD-like domains of Bd0553 (red) and Bd1075 (white) overall fold is conserved in both enzymes, with exception of the signature alanine residue of the L,D-hydrolase motif, which is substituted for a valine residue in Bd1075. (**Figure 76A**).

It has been shown that Bd1075 has L,D-carboxypeptidase activity [125], while there is experimental evidence showing the L,D-carboxypeptidase and L,D-endopeptidase activity of Bd0553 (unpublished data). This information correlates with the open Bd0553 active site in contrast to the predicted lid present in Bd1075, allowing Bd0553 to accommodate longer stem peptide regions for hydrolysing internal peptide bonds.

For the YkuD2-like domains of Bd1402 (blue) and Bd334 (grey), sequence and structural similarities are more evident than for the YkuD-like L,D-hydrolases (**Figure 76B**).

Of particular interest are the conserved residues Y208 and W220 (Bd1402), and Y191 and W215 (Bd3334), which are exclusive to the YkuD2-like domain. These two aromatic residues are partially occupying the substrate binding cleft of each enzyme, suggesting their participation in substrate selection, possibly recognising substrates that are different from the Bd0553 ones.

Remarkably, the Bd1402 lid (which is also possibly present in Bd0782) indicates that lipobox L,D-hydrolases have a different substrate recognition and/or binding mechanism when compared to soluble YkuD2-like L,D-hydrolases such as Bd3334, requiring displacement of the active site lid. The Bd1402 lid might additionally limit the hydrolytic activity to the carboxy terminus of the PG stem peptides, as hypothesised for Bd1075.

These results then suggest that active site open L,D-hydrolases of group 1 and group 5 (YkuD2-like Bd3334) have endopeptidase and carboxypeptidase activity, that might recognise chemically divergent PG crosslinked substrates.

4.3. Discussion

4.3.1. *B. bacteriovorus* HD100 LDTs cluster in six specialised groups

The 19 predicted LDTs were grouped according to their sequence similarity, with Bd1075 being the only LDT clustering in group 6. This is coincidental with the non-predatory role described for Bd1075, and with it being the most distant sequence-wise to the rest of the enzymes [125].

Interestingly, not every defined group is homogenous regarding the sequence similarity of their members, despite a clear conservation in their catalytic YkuD- or YkuD2-like domains. The only observed exception corresponds to group 3 of LDTs, wherein the SH3b-containing enzymes Bd0886 and Bd1176 are almost identical in their primary sequences.

In group 1 for instance, Bd1763 is the most distant member when compared to Bd0553 and Bd2061, especially in the alpha-helix rich domain defining this group. This observation suggests that even though certain functional redundancy is expected between enzymes, localisation and substrate or ligand binding specificities may differ.

Group 2 shows high variability, wherein Bd3176 and Bd3741 are the most similar respect to their amino acid sequence, grouping in the same clade. On the contrary, Bd3639 is the first group 2 LDT to diverge, which agrees with missing key structural features in the AlphaFold2 model depicted in **Figure 62**, particularly the signature triangular hook domain present even in Bd3012, which is in-between Bd3639 and the Bd3176-Bd3741 clade.

Group 4 comprises four highly conserved members, with Bd1181 being the most distant to the other three due to it having an insertion between positions 27 and 49 (relative to Bd0566 sequence numbering). Interestingly, most differences in this group are localised in the N-terminal domain that has been characterised as a protein-protein interaction region in YcbB, increasing the number of L,D-crosslinks upon beta-lactam stress [83], [147].

The L,D-hydrolases of group 5 on the other hand are particularly heterogenous, considering their divergent YkuD2-like catalytic domain, which is not shared by any other LDT group. Remarkably, based on their signal peptide sequence, this group further splits into lipid-linked Ykud2-like hydrolases (Bd0782 and Bd1402), and soluble YkuD2-like hydrolases (Bd2940, Bd3093 and Bd3334). Notwithstanding the significant sequence similarity between the lipobox subgroup (over 60% sequence), similarity is much more marginal for the soluble subgroup (around 30%), suggesting their possible divergence in substrate recognition and/or enzyme catalysis parameters.

The LDT group distribution appears to be conserved in endobiotic predatory *Bdellovibrio*, as shown in **Figure 56**. Interestingly, *Bdellovibrio* ArHS lacks LDTs from group 1 suggesting that this predator has a more restricted prey scope when compared to *B. bacteriovorus* HD100 or Tiberius, or there is functional redundancy between L,D-hydrolases of group 1

and group 5. In this same sense, *B. exovorus* JSS only predatory LDT, A11Q_1574, clusters as part of group 2, suggesting that LDTs from groups 1 and 3-5 are important for the intra-periplasmic predatory cycle.

4.3.2. *B. bacteriovorus* HD100 LDTs accessory domains as targeting and substrate specificity determinants

Each LDT group is characterised by at least one accessory domain, which might be involved in substrate specification, protein-protein interactions and/or subcellular localisation. Nevertheless, biochemical studies directed at understating the interaction between PG substrates with LDTs is remarkably scarce and is mainly focused on their inactivation by small molecules as possible antibiotics [76]–[79], [89], [80].

Group 1 of YkuD-like hydrolases accessory domain corresponds to an initially predicted coiled-coiled domain, which was proven to actually correspond to an alpha helix rich domain with a different topology in the Bd0553 experimental structure.

Protein BLAST analysis has shown that this domain is exclusive to *B. bacteriovorus*, thus further studies are required to establish its role in group 1 function. However, due to its distance to the active site of the enzyme in addition to the minimal conformational fluctuations predicted from the experimental B-factors of the Bd0553 structure, one possible hypothesis is that this domain mediates the interaction of the enzyme with PG cell wall. Since acidic clefts are formed by this domain, it could interact with the glycan portion of PG, while the stem peptide is accommodated in the active site for hydrolysis. Another option is for this domain to participate in protein complex formation.

Group 2 corresponds to the ErfK-like LDTs with a signature triangular hook domain that is structurally conserved but not in amino acid sequence. This domain is conserved in Bd3012, Bd3176 and Bd3741, forming an aromatic residue-centred pocket proximal to the substrate binding region of the enzyme, suggesting its participation in substrate binding. Interestingly, this domain is absent in the predicted structure of Bd3639, which could potentially lead to this LDT having a different crosslinking activity.

Even though the amino acid sequences of group 2 relate these enzymes with the Erfk PG-Lpp1 crosslinker, it has been proposed that bdelloplast formation requires the hydrolysis of this linkage[118], [142], [161], [162]. It has been shown that fatty acid acylation of the

PG sacculi occurs in the bdelloplast and could functionally replace the PG-Lpp1 linkage [39].

Experimental evidence suggests that fatty acid modification of PG occurs via a O-acylation reaction, condensing a carboxylic acid ester that is resistant to trypsin digestion [39]. This observation implies that esterification occurs in the glycan portion of PG and not in the stem peptide or a PG-crosslinked protein.

In this case, group 2 might participate in forming 3-3 crosslinks between lipid-esterified PG molecules, possibly stabilising the bdelloplast. This might explain the sequence diversity found in this group, in order to recognise chemically varied lipid molecules.

Regarding group 3, these enzymes have been shown to participate in NCDAA incorporation into bdelloplast PG molecules, strengthening and stabilising this structure [33]. Moreover, it has been shown that one of the bdelloplast PG modifications is the incorporation of terminal mDAP residues, which might potentially be catalysed by this group of LDTs [153].

With respect to the SH3b domain, as depicted in **Figure 66**, the inherent flexibility of the loop connecting the SH3b domain with the YkuD-like catalytic domain of Bd0886 suggests that the SH3b domain might act as a targeting domain rather than a substrate selection domain, as has been hypothesised for SH3b-containing lysostaphins [124], [157], [158]. Functional experiments comparing the NCDAA exchanging activities of these enzymes to mutants lacking the SH3b domain are necessary to discard their contribution to direct substrate binding.

Group 4 is the most conserved structure-wise when compared to the *E. coli* YcbB counterpart, suggesting they share similar enzymatic activities [83]. In fact, unpublished data from our group shows that Bd1181 has 3-3 PG crosslinking activity.

However, even though the PGBD and the N-terminal protein-protein interaction domains of YcbB are present in the experimentally determined structure of Bd0566 and the AlphaFold2 models of Bd1181, Bd1358 and Bd2908, the N-terminal domain of the enzyme is structurally and chemically divergent between the two species.

L,D-crosslinks are not prevalent in the *E. coli* cell wall unless the D,D-transpeptidation route is stalled by an stressor agent such as penicillin-related beta lactams, leading to

increased alarmone concentrations [83], [89], [147]. Consequently, YcbB is recruited to PBP transpeptidation/transglycosylation machinery, replacing the normally occurring 3-4 crosslinks for 3-3 crosslinks. Since this scenario is not emulated by *B. bacteriovorus* invasion of prey and bdelloplast formation, one possible hypothesis is that the modified N-terminal domain of the group 4 L,D-transferases allows them to replace the 3-4 prey's crosslinks for 3-3 crosslinks even if PBPs are still active. This enrichment in 3-3 crosslinks alongside the deacetylation of glycan chains, NCDAA incorporation and fatty acid modification might be related to bdelloplast structural stability, and even be part of mechanisms differentiating prey from predator and/or indicating that the current cell has already been invaded [39], [153].

The final group of predatory LDTs, group 5, poses the most challenging to analyse due to the considerable divergence in sequence, including the differentiation between soluble and lipid-linked L,D-hydrolases. Nonetheless, the YkuD2-like domain, whose amino acid sequence differs from the YkuD-like domain of the rest of the LDTs, defines this as a standalone group. Particularly, two aromatic residues, a tyrosine, and a tryptophan residue, are part of the substrate binding region of the YkuD2-like domain of both type of enzymes, suggesting a different substrate specificity for this group of hydrolases.

Bd0782 and Bd1402 are membrane associated putative L,D-hydrolases. As proposed by the analysis of the crystallographic structure of Bd1402, the active site lid region, which is also connected to the dimerization helix of the enzyme, might be involved in substrate selection, particularly by a conserved aspartate residue proximal to the catalytic dyad. Further, activity might be restricted to membrane-linked PG substrates due to presence of the lipobox. The potential substrates might include PG-anchoring proteins as Lpp1 or OMPs, potentially contributing to bdelloplast formation in different preys [118], [142], [162].

On the other hand, it was possible to solve the structure of Bd3334 as representative of the soluble subgroup, depicting an open active site similar to what is observed for group 1. However, AlphaFold2 predictions of Bd2940 and Bd3093 suggest the presence of unstructured C-terminal lids which are not conserved between these two proteins or in the lipobox subgroup. These results suggest that this subgroup potentially recognises different substrates or PG modifications compared to group 1, possibly in different preys or during different stages of the life cycle. In this sense, it has been shown that PG is

extensively deacetylated in the bdelloplast, which might affect the interaction of the LDTs with the prey's cell wall [38], [116], [117]. Nonetheless, since the glycan portion of the PG molecule is not the target of the LDTs activities, activity and binding experiments contrasting deacetylated and acetylated PG are needed.

4.3.3. Active site lids and residue conservation are the features defining either the L,D-transferases or L,D-hydrolases preference of each enzyme

When comparing the YkuD-like domains of the L,D-transferase Bd0566 (white) with the YkuD- and YkuD2-like domains of the L,D-hydrolases Bd0553 (red) and Bd3334 (grey), respectively, main differences are related to the higher availability of hydrogen bonding sidechains in Bd0566 acyl acceptor site (**Figure 77**). In particular, the transferase signature arginine residue (R424 in Bd0566) might aid the selection of a nucleophilic second substrate over a hydrolytic water molecule.

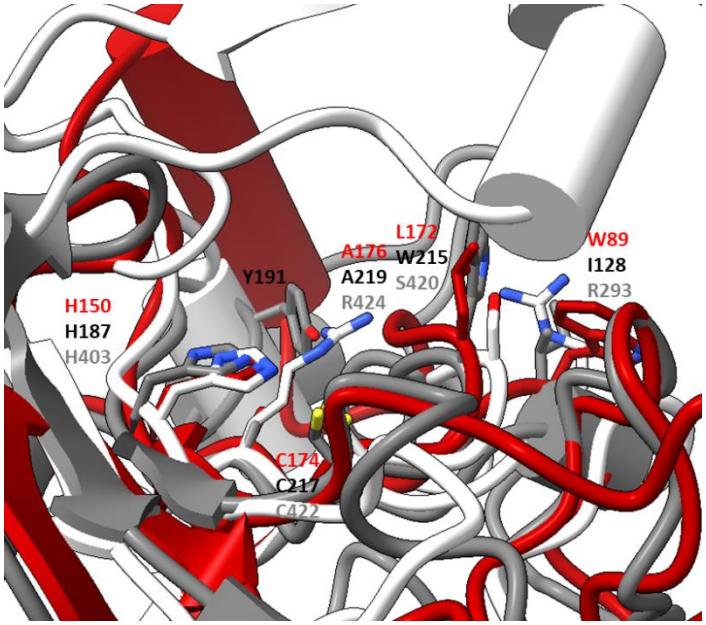


Figure 77: Structural comparison between L,D-transferases and L,D- hydrolases active site. YkuD-like domains of Bd0566 (white) and Bd0553 (red), and the YkuD2-like domain of Bd3334 (grey) were superimposed depicting characteristic residues in stick representation. The identity of the residue two positions after the catalytic cysteine residue appears to be the major contributor to determining either transferase or hydrolase activity.

Remarkably, the major structural feature distinguishing both types of enzymes is the presence an active site lid that has been proposed in other models to stabilise the acyl-enzyme intermediate [78], [79], [163].

It has been observed that adding a lid to the *B. subtilis* GGT promotes selection of shorter glutamic acid molecules as acyl donors [163], [164]. In this context, active site lids might sterically select the terminal region of the PG stem peptides either for NCDAA transferase (as proposed for group 3) reaction or carboxypeptidation of terminal residues (as observed for Bd0553).

Additionally, the active site lid of the γ-glutamyl transferase (GGT) of *E. coli* has been shown to not only define substrate specificity, but to stabilise the acyl-enzyme intermediate when inserted in the lid-less *B. subtilis* orthologue, favouring the transpeptidation activity over hydrolysis by excluding solvent molecules from the active site [163]. This makes possible to hypothesise that regions such as the capping subdomain of Bd0566 not only participates in substrate binding but promotes stabilisation of the acyl-enzyme intermediate by de-solvating the active site.

Thus, hydrolase/transferase determination in *B. bacteriovorus* HD100 LDTs results possibly from the combination between the chemical identity of residues in the acyl acceptor site (arginine against alanine/valine residue) and the lid-driven stability of the reaction intermediate.

4.4. Conclusion

The structure of seven representative *B. bacteriovorus* HD100 LDTs were determined, identifying accessory domains and diverse substrate binding sites.

In combination with sequence variation analysis and structural predictions for the remaining members of each group, these results contrast the different groups of LDTs to assign possible roles in predation.

L,D-transferases of group 2 (with exception of Bd3639) are proposed to attach the PG cell wall to membrane by direct acylation of the stem peptide by a lipid molecule, while group 3 and 4 are proposed to direct NCDAA attachment and 3-3 stem peptide crosslink formation, respectively.

On the other hand, L,D-hydrolases of group 1 and group 5 could be associated with the outer membrane-cell wall protein link hydrolysis, particularly influenced by the membrane anchoring signal in Bd0782 and Bd1402. Additionally, both groups could be involved in 3-3 crosslink hydrolysis and/or carboxypeptidation reactions.

Comparative analysis between the hydrolases of group 1 (YkuD-like) and group 5 (YkuD2-like) shows that the active site of both groups diverges, with the presence of two exclusive residues (tyrosine and tryptophan residues) in YkuD2-like active sites. This characteristic along with the presence of an active site lid in the experimental structure of Bd1402 and the predicted models of Bd0782, Bd2940 and Bd3093 suggests that both groups could recognise chemically diverse substrates in potentially different types of prey or during different stages of the predatory cycle.

Finally, based on predicted and experimental activity data, in view of the structural information presented in chapters 3 and 4, it has been hypothesised that the presence of an active site lid or capping subdomain, in addition to the residues present at the acyl acceptor site of the active site determine the transferase or hydrolase identity of the different LDTs.

CHAPTER 5

5. STRUCTURAL CHARACTERISATION OF *B. bacteriovorus* PUTATIVE HYDROLASES Bd0601 AND Bd1177

5.1. Introduction

CWPs (EC 3.4) and CWA (EC 3.5.1) modify the PG cell wall by hydrolysing amide bonds between amino acids or between the MurNAc and the stem peptide, respectively [165]–[167].

CWAs have N-acetylmuramoyl-L-alanine amidohydrolase (NALAA) catalytic domains that have been classified in three groups in bacteria based on sequence similarity: i) NALAA-2 (IPR002502), (ii) NALAA-3 (IPR002508), and (iii) NALAA-5 (IPR008044).

For all NALAA domain containing enzymes, it has been shown that catalysis is dependent on a Zn(II) ion and produces tri and tetrapeptides derived from the hydrolysis of the 1,6-anhydro-MurNAc-L-Ala and/or MurNAc-L-Ala bonds [96], [168]. NALAA domains possess a three-layer α/β/α sandwich architecture, wherein the catalytic Zn(II) ion is coordinated in the ligand binding cleft delimiting the stem peptide and glycan binding pockets of the enzyme [169]–[171].

CWPs hydrolyse peptide bonds and are divided between carboxypeptidases, which remove the C-terminal residue from the stem peptide, or endopeptidases, which in turn cleave peptide bonds within the stem peptide [94]. The activity of these enzymes may be further described by considering the chirality (D- or L-) between the corresponding amino acids, distinguishing between D,D- or L,D- peptidases [94], [172]. At least ten types of CWP domains have been reported, including the transpeptidase (TP, IPR005490), CHAP (IPR007921) and NlpC/P60 (IPR000064) domains.

CHAP domains are distinctive in that they may act as peptidases by hydrolysing stem peptide crosslinks as well as amidases, releasing the stem peptide from the glycan portion of the PG molecule [101], [102]. CHAP domains are involved in several bacterial processes, including cell division and virulence, in which their activity is regulated by their association to precisely localised subcellular partners such as ActH, EnvC and FtsX [167], [173]. Binding partners are proposed to produce a conformational change that releases the autoinhibited CHAP activity, allowing localised PG hydrolysis [167], [173].

Catalytic mechanisms of CHAP hydrolases are varied. For instance, the papain-like CHAP fold of CWPs LysK and LysGH15 possess a Ca(II) ion binding motif that is required for catalysis, whereas PlyC lacks this feature and has been proven to be calcium-independent for its activity [103], [108], [110], [174], [175]. Nevertheless, CHAP domains are usually part of multidomain enzymes which tends to obscure their individual contribution to the overall catalytic activity of the same [103], [110], [175]. For the *Streptococcus dysgalactiae* phage endolysin PlySK1249, an N-terminal amidase NALAA-3 domain acts synergistically with a non-bacteriolytic C-terminal CHAP domain [176], [177]. In this case, the CHAP domain of PlySK1249 hydrolyses PG crosslinks, which are then further cleaved to tri- and tetrapeptides by the amidase activity of the N-terminal NALAA-3 domain [176], [177].

NlpC/P60 domains adopt a papain-like fold and share many similarities to CHAP domains, however, have been mostly identified as calcium-independent cysteine proteases [94], [130]. For NlpC/P60 proteases such as YaeF, it has been determined that the characteristic fold is circularly permutated, reversing the Cys-His-His catalytic triad position in the amino acid sequence, jeopardising the identification of putative NlpC/P60 domains by conservation of their amino acid sequence [130], [178].

As for their biological importance, the NlpC/P60 peptidase RipA has been identified as responsible for septum degradation in mycobacterial cytokinesis [114]. NlpC/P60s have also been implicated in cell separation in *Corynebacterium glutamicum* and as part of the antibiotic resistance conferred by transposon Tn916, which encodes a bifunctional GH23 glycosidase-NlpC/P60 endopeptidase from *S. aureus* and *C. difficile* [128]. For the latter bifunctional NlpC/P60 domain, active site accessibility is suggested to be regulated by a side chain above the nucleophilic cysteine, switching the enzyme from a closed to an open conformation upon substrate binding [128].

B. bacteriovorus has been shown to encode a wide repertoire of PG hydrolytic enzymes, including hypothetical CHAP and NlpC/P60 domain containing enzymes that have diversified their primary sequence by acquiring novel and even cryptic domains [21], [179]. Thus, using X-ray crystallography we have determined the structure of two proposed CWPs and/or CWAs, Bd601 and Bd1177. Bd0601 is predicted to belong to the DUF1460 family of NlpC/P60 enzymes, while Bd1177 is a poorly predicted CHAP or NlpC/P60 that might have undergone multiple rearrangements and permutations during *B. bacteriovorus* evolution.

5.2. Results

5.2.1. Bd0601

5.2.1.1. Bd0601 purification and crystallisation

Bd0601 is encoded in the operon *bd0601-0603* as shown in **Figure 16**. As a first approach, the full-length operon *bd0601-0603* was codon optimised and cloned in the pET21d(+) vector. This was done due to the precedent that the predicted ankyrin Bd0602 may act as an immunity protein for Bd0601, similar to what was observed for DDTs Bd0816 and Bd3459, and the ankyrin Bd3460 [35], [36]. Specifically, the predicted *B. bacteriovorus* signal peptide SecI for Bd0601 was substituted by a pelB sequence. Additionally, the predicted SecII/lipobox signal peptides for Bd0602 and Bd0603, including the lipid-anchoring cysteine residue, were exchanged for an optimised pelB sequence for periplasmic expression.

No soluble expression was observed for the full-length operon construct, or for individual components expressed separately in the periplasm under the different tested conditions. Nevertheless, a catalytically null mutant for Bd0601, Bd0601 C94A, was successfully expressed and purified either from the *bd0601-0603* operon construct or by itself in AI media supplemented with sugar mix A as inducer.

Bd0601 C94A was purified to homogeneity **(Figure 78A**) and produced a monodisperse peak on gel filtration chromatography on a Superdex™ 75 26/60 column (**Figure 78B**), showing that the observed doublet band corresponds to an artifact that is not representative of the homogeneity of the sample. The elution profile corresponds to a globular protein of approximately 30 kD, suggesting that Bd0601 C94A is monomeric in solution, which agrees with what has been observed for other DUF1460 enzymes [113].

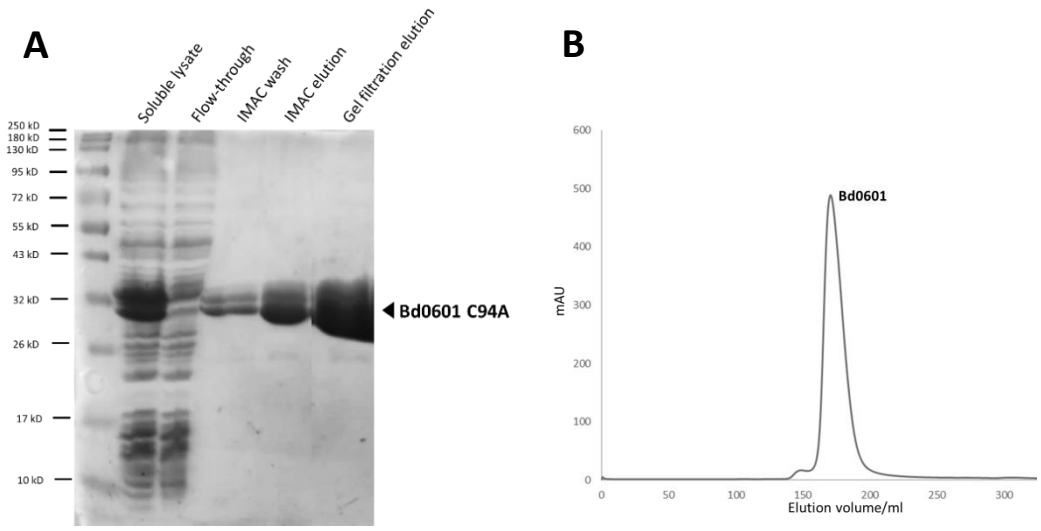


Figure 78: Bd0601 C94A purification and gel filtration profile. Bd0601 C94A was expressed in AI media supplemented with sugar mix A and purified by nickel IMAC and gel filtration chromatography. **A)** 15% SDS-PAGE purification gel for Bd0601 C94A. IMAC wash indicates washing with lysis buffer containing 20 mM imidazole. **B)** Gel filtration profile of Bd0601 C94A showing a monodisperse peak corresponding to the hydrodynamic radius of a globular 30 kD protein. The run was performed at room temperature and flow rate of 2 mL/minute in a buffer composed by 15 mM HEPES pH 7.5 and 200 mM NaCl.

Bd0601 C94A was concentrated and screened for crystallisation conditions as indicated in section 2.3.9 and 2.3.10, respectively. Crystals for Bd0601 C94A were obtained at a concentration of 35 mg/mL in MIDAS-Plus™ condition G3 (40 % glycerol ethoxylate) and Morpheus™ condition F10 (Buffer system 3 pH 8.5, 0.12 M monosaccharides and 30% ethylene glycol- PEG 8000 precipitant mix), which were harvested and subjected to data collection (**Figure 79A** and **B** for each crystal form, respectively).

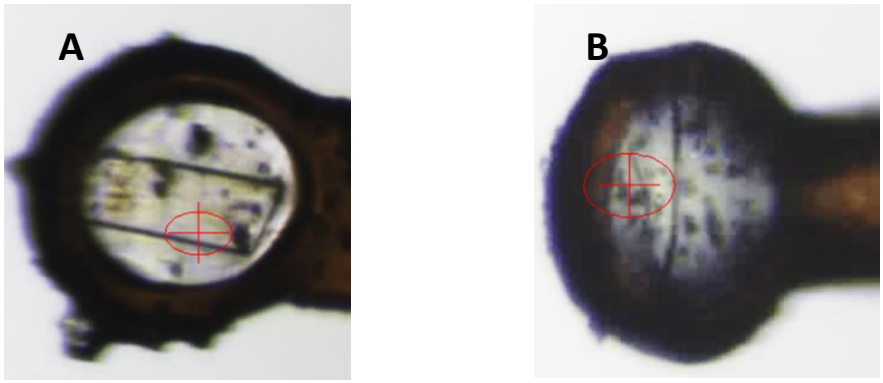


Figure 79: Bd0601 C94A crystals. Bd0601 C94A crystallised at the same concentration (35 mg/mL) and temperature (18°C) in two different conditions, which are defined as crystal form 1 and 2. **A)** Crystal form 1 of Bd0601 C94A obtained in 40 % glycerol ethoxylate. **B)** Crystal form 2 of Bd0601 C94A obtained in Morpheus™ Buffer system 3 pH 8.5, 0.12 M monosaccharides and 30% ethylene glycol- PEG 8000 precipitant mix.

5.2.1.2. Bd0601 C94A structure solving and refinement

The diffraction experiment was carried out at the IO4 beamline as indicated in section 2.1.11, collecting native datasets of Bd0601 C94A at a resolution of 3.95 Å and 2.05 Å for the MIDAS-Plus™ G3 and Morpheus™ F10 crystals, respectively.

Diffraction data was autoprocessed by the Xia2-multiplex and Xia-3dii automated pipelines for the MIDAS-Plus™ G3 and Morpheus™ F10 crystals, respectively. Crystallographic data was solved using the Phaser-MR programme of the Phenix suite and LPG0564 from *Legionella pneumophila* (PDB ID: 2IM9) as search model. Collected data was solved in the $P2_1$ spacegroup, finding two copies of Bd0601 C94A in the ASU for both datasets, subjected to automatic building using Phenix Autobuild and refined with Phenix Refine. Statistical results for data collection and refinement are shown in **Table 14**.

Table 14: **Data collection and refinement statistics for apo- and GlcNAc-Bd0601 C94A**. Statistics for apo- and GlcNAc-Bd0601 C94A crystallographic data obtained at 3.95 and 2.05 Å, respectively, are shown.

	Apo-Bd00601 C94A	GlcNAc-Bd0601 C94A
Data Collection		
Resolution range (Å)	71.48 - 3.95	46.4 - 2.05
Space group	P 2_1	P 2_1
Cell dimensions		
a, b, c (Å)	42.30, 142.95, 50.12	43.27, 143.19, 50.13
α, β, γ (°)	90, 101.67, 90	90, 101.91, 90
Total reflections	67744 (6867)	262224 (26394)
Unique reflections	5152 (510)	37335 (3693)
Completeness (%)	99.77 (99.22)	99.95 (99.84)
Redundancy	13.1 (13.5)	7.0 (7.1)
<I>/σI	5.15 (2.05)	12.12 (2.24)
R_{meas}	0.2581 (0.5386)	0.1166 (1.033)
R_{pim}	0.07092 (0.1446)	0.04365 (0.3841)
$CC_{1/2}$	0.993 (0.978)	0.998 (0.822)
Refinement		
Maximum resolution (Å)	3.95	2.05
Number of reflections used	5143 (506)	37323 (3689)
Number of reflections for R_{free}	515 (50)	1809 (172)
R_{work}	0.1892	0.1818
R_{free}	0.1917	0.2216
RMSD		
Bonds (Å)	0.005	0.010
Angles (°)	0.80	1.27
Number of non-hydrogen atoms		
Protein	4460	4490
Ligand/ion	0	30
Water	18	266
B-factors		
Protein	92.30	36.29
Ligand/ion	-	34.81
Water	76.84	41.64
Ramachandran favoured (%)	96.21	97.47
Ramachandran allowed (%)	3.61	1.99
Ramachandran outliers (%)	0.18	0.54
Rotamer outliers (%)	0.62	1.24

*Values in parenthesis correspond to the highest-resolution shell

Electron density map analysis and further refinement revealed the presence of GlcNAc in the active site of the Bd0601 C94A Morpheus™ F10 dataset (**Figure80A**). This ligand possibly comes from the monosaccharides mix of the Morpheus™ F10 condition, whereas the MIDAS-Plus™ G3 dataset does not depict any relevant non-assigned electron density (**Figure 80B**). Thus, Bd0601 C94A datasets will be referred as apo-Bd0601 C94A and GlcNAc-Bd0601 C94A for crystal forms 1 and 2, respectively.

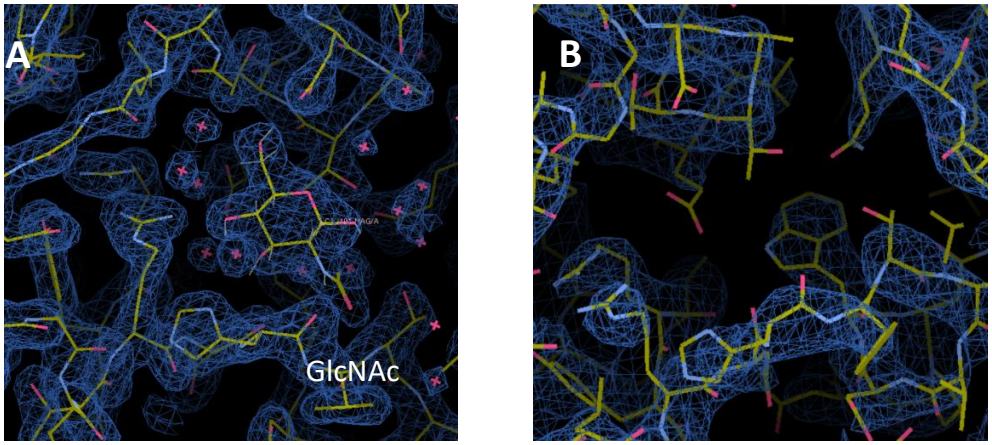


Figure 80: **Electron density contour map of GlcNAc- and apo-Bd0601 C94A.** Electron density map are depicted contoured at 1.4 σ. **A)** Electron density map ($2F_0$-F_0) corresponding to the solved Bd0601 C94A crystal form 2, wherein the GlcNAc molecule is present in the active site. **B)** Electron density map ($2F_0$-F_0) of Bd0601 C94A crystal form 1. No unassigned density is present in the active site of the apoenzyme.

5.2.1.3. Bd0601 C94A structure overview

Apo- and GlcNAc-Bd0601 C94A crystal structures trace 279 of the 287 residues of the mature protein, with the same overall fold in both crystallographic models.

As depicted in **Figure 81A**, Bd0601 C94A adopts a papain-like fold characteristic of the NlpC/P60 CWPs, with a central alpha/half beta barrel sandwich formed by five alpha helices (in blue) and six beta strands (in red).

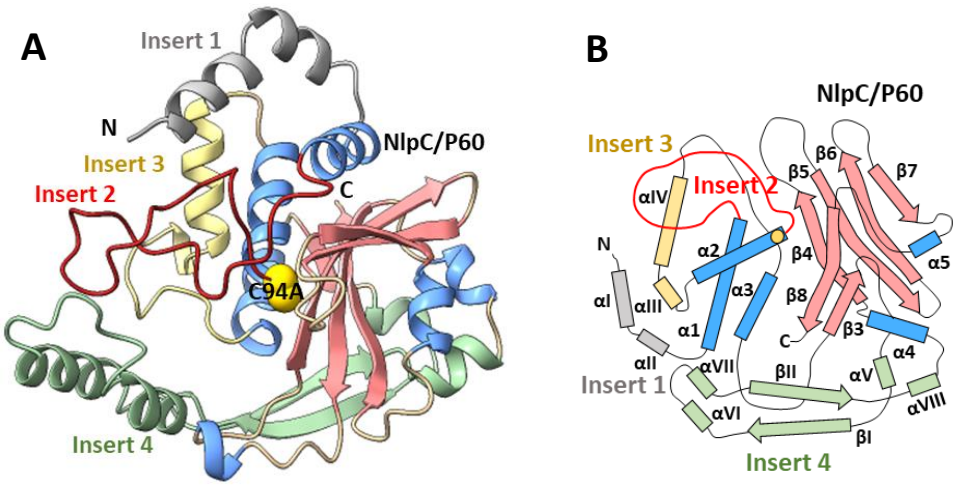


Figure 81: **Bd0601 C94A overall fold and topology diagram.** Representations are colour the DUF1460 structural insertions: insert 1 (residues 29-51, in grey), insert 2 (63-93, in red), insert 3 (residues 109-135, in yellow) and insert 4 (residues 154-220, in green). The NlpC/P60 core of the structure is coloured by secondary structure (alpha helices in blue and beta strands in red). **A)** Cartoon representation of Bd0601 C94A crystallographic structure. The C94A mutation defining the putative active site of the enzyme is shown in yellow sphere representation. **B)** Topology diagram of Bd0601 C94A structure. C94A mutation is highlighted in yellow.

Bd0601 C94A fold has four structural insertions with respect to the NlpC/P60 papain-like fold, which have been described for the DUF1460 CWPs, with exception of two N-terminal alpha helices (αI and αII, residues 29-51, in grey) that are absent in the most well-characterised DUF1460, AmiA, from *Bacteroides uniformis* (PDB ID: 4H4J) [113]. The remaining three structural insertions are part of DUF1460 amidases/peptidases (**Figure 81A** and **B**), consisting of a 30-residues long loop insertion between helices α1 and α2 of the catalytic core (residues 63-93, in red), alpha helices αIII and αIV (residues 109-135, in yellow), between alpha helices α2 and α3 of the DUF1460 papain-like domain, and the large third insert (residues 154-220, in green) between strand β3 and helix α4 of the catalytic core.

Bd0601 C94A active site comprises the catalytic triad C94A-H255-H270 and a conserved tyrosine residue, Y68, which is proposed to stabilise the tetrahedral PG-enzyme intermediate in DUF1460 amidases/peptidases [113], [130], [178]. The nucleophilic C94 (C94A in the crystallised Bd0601 mutant) is located in α2 similarly to other DUF1460 and papain-like enzymes.

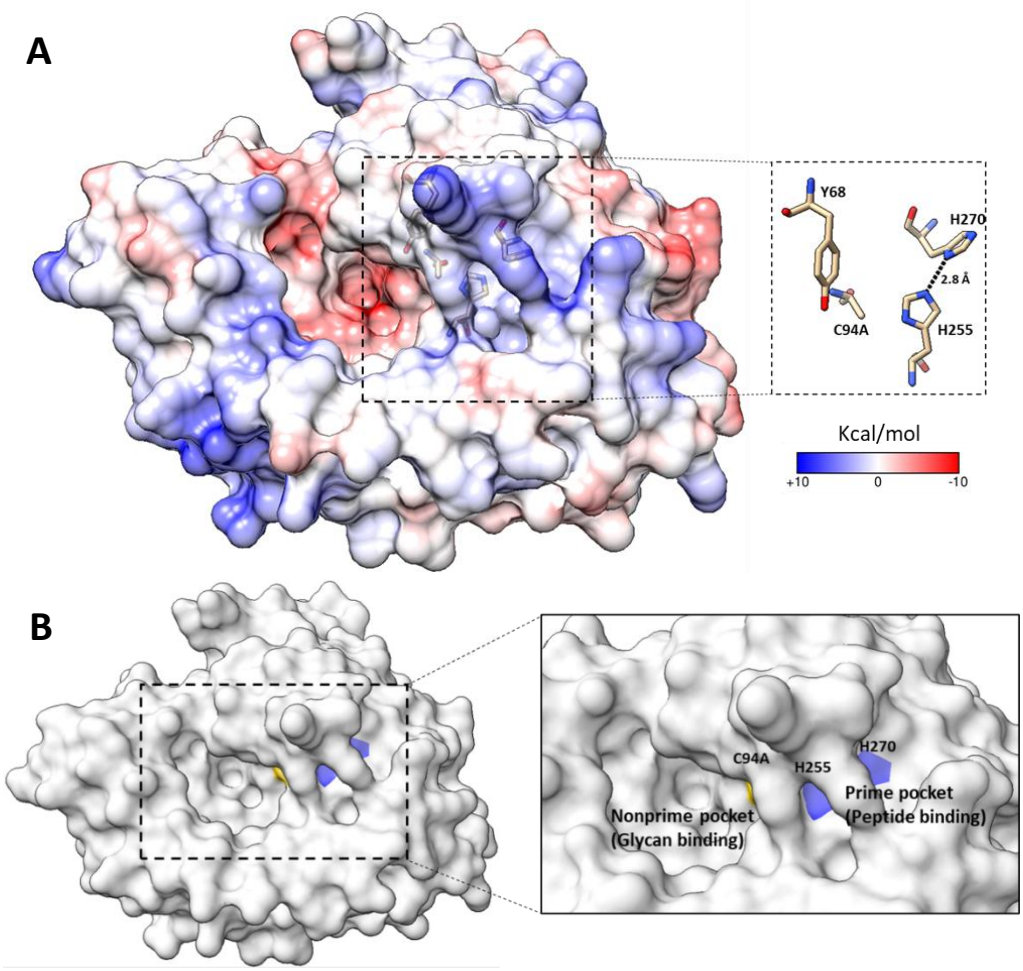


Figure 82: Electrostatic surface potential and surface representation of substrate binding pockets of Bd0601 C94A. **A)** Electrostatic surface potential of Bd0601 C94A. An acidic pocket is observed at the left side of the predicted catalytic triad C94-H255-H270 of Bd0601, that is predicted to bind the sugar portion of the PG molecule. The oppositely located basic groove is predicted to accommodate the stem peptide portion of PG molecules. **B)** Bd0601 C94A surface representation. Location of the predicted catalytic triad C94-H255-H270 defines a nonprime pocket for PG sugar binding and a prime pocket for P peptide binding.

The structural inserts around the DUF1460 catalytic core of Bd0601 C94A define a large pocket surrounding the catalytic triad C94A-H255-H270 with an acidic region towards the N-terminus of the enzyme and a basic groove towards the C-terminal region of the papain-like fold (**Figure 81A**). These pockets, in a similar manner to the *B. uniformis* AmiA, are suggested to work as nonprime side and prime side pockets, respectively. The acidic nonprime pocket would putatively bind the glycan portion of the PG molecule, while the prime pocket would accommodate the stem peptide of the PG (**Figure 81B**).

5.2.1.4. Bd0601 C94A sugar binding pocket

A portion of the substrate binding residues of Bd0601 C94A are identifiable on the GlcNAc-Bd0601 C94A crystallographic structure (**Figure 83A**).

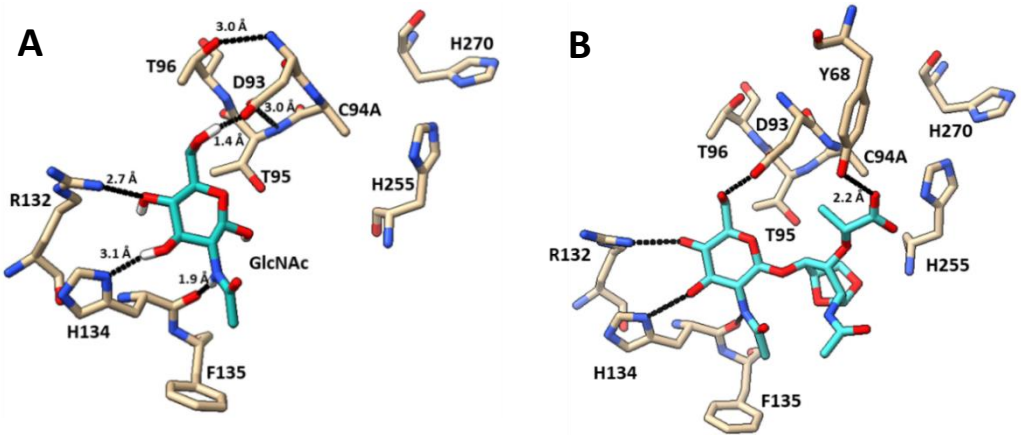


Figure 83: Bd0601 C94A glycan binding pocket. A) Stick representation of the GlcNAc bound structure of Bd0601 C94A. Residues D93, R132 and H134 are shown to stabilise the polar groups of GlcNAc by hydrogen bonding interactions. **B)** Stick representation of the GlcNAc-Bd0601 C94A structure superimposed on the anhMurNAc-GlcNAc-AmiA structure showing that only Y68 is potentially involved in disaccharide binding.

Sidechains of residues D93, R132 and H134 stabilise the hydroxyl groups of GlcNAc while the amide nitrogen of the sugar is stabilised by the carboxyl of 134 backbone. Interestingly, in contrast to what was shown for the GlcNAc bound form of AmiA (PDB ID: 4Q68), which superimposes with an overall RMSD of 1.04 Å to GlcNAc-Bd0601 C94A, the acetyl group of the GlcNAc ligand does not interact with a stabilising sidechain (Y152 in AmiA) [113].

Superimposition of the anhMurNAc-GlcNAc-AmiA (PDB ID: 4Q5K) on GlcNAc-Bd0601 C94A reveals high quality overlapping of the GlcNAc portion of both ligands (RMSD= 1.03 Å for the overall structure). The superimposition of the anhMurNAc-GlcNAc disaccharide on Bd0601 C94A depicts the hydrogen bonding between the hydroxyl group of Y68 with the carboxylate of the anhMurNAc molecule, which agrees with what has been observed for AmiA, and with most members of the DUF1460 family (**Figure 83B**) [113].

As shown in **Figure 84**, residues responsible for direct sugar binding are mostly invariable in the three publicly available DUF1460 structures: AmiA from *B. uniformis* (PDB ID: 4H4J), BF2036 from *B. fragilis* (PDB ID: 2P1G) and LPG0564 from *L. pneumophila* (PDB ID: 2IM9), and the non-solved DEC5E from *E. coli*. Equivalent sugar binding residues Y68, D93, R132

and H134 are present through these DUF1460 family members. The motif RXH is present in all DUF1460, wherein residues R (R132 in Bd0601) and H (H134 in Bd0601) stabilise the hydroxyls in positions 3 and 4 of the sugar ring (at 3.1 and 2.7 Å distance, respectively). Additionally, the backbone carboxyl group of H134 interacts with the amide radical of GlcNAc, at 1.9 Å distance. This suggests that this family of enzymes has optimised these residues as initially proposed for AmiA [113].

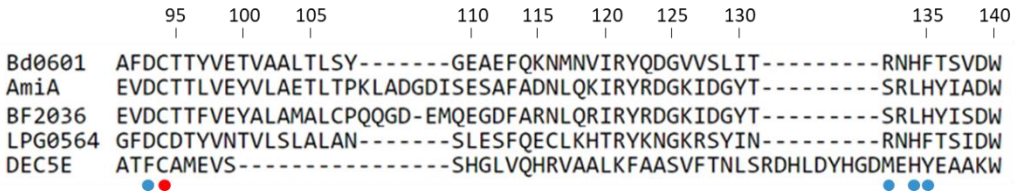


Figure 84: Sequence alignment between Bd0601 and available DUF1460 structures. Bd0601 sequence was aligned with AmiA, BF2036, LPG0564 and DEC5E to identify sugar binding residues (highlighted with blue dots). GlcNAc and anhMurNAc-GlcNAc binding residues are conserved along the crystallised DUF1460 members. Catalytic cysteine is indicated with a red dot. Residue numbering is relative to Bd0601.

Non-priming pockets of Bd0601 C94A, AmiA, BF2036 and LPG0564 possess a similar solvent excluded surface, ranging between 1350 and 1660 $Å^3$, respectively, suggesting that the different DUF1460 enzymes bind the glycan portion of PG in a similar way.

Interestingly, no residue is observed interacting with the acetyl group of GlcNAc in **Figure 83**, suggesting that Bd0601 could accept acetylated or deacetylated PG molecules indistinctly, in contrast to the deacetylation specificity reported for DslA (Bd0314,) which localises two steric blocking residues in the active site Y228 and M150 [38], [116], [117].

5.2.1.5. Bd0601 C94A stem peptide binding pocket

The Bd0601 C94A predicted stem peptide binding pocket was compared with those of the three DUF1460 protein structures deposited in the PDB: AmiA from *B. uniformis* (PDB ID: 4H4J, 4Q68 and 4Q5K, for apo-, GlcNAc- and AnhMurNAc-GlcNAc- bound structures, respectively), BF2036 xylanase from *B. fragilis* (PDB ID: 2P1G), and LPG0564 from *Legionella pneumophila* (PDB ID: 2IM9).

In contrast to the non-prime pocket, the putative stem peptide binding pocket of DUF1460 enzymes has been poorly characterised, partially due to its high sequence variability. Residues located on the putative prime binding pocket surface of Bd0601 C94A are depicted in **Figure 85A**.

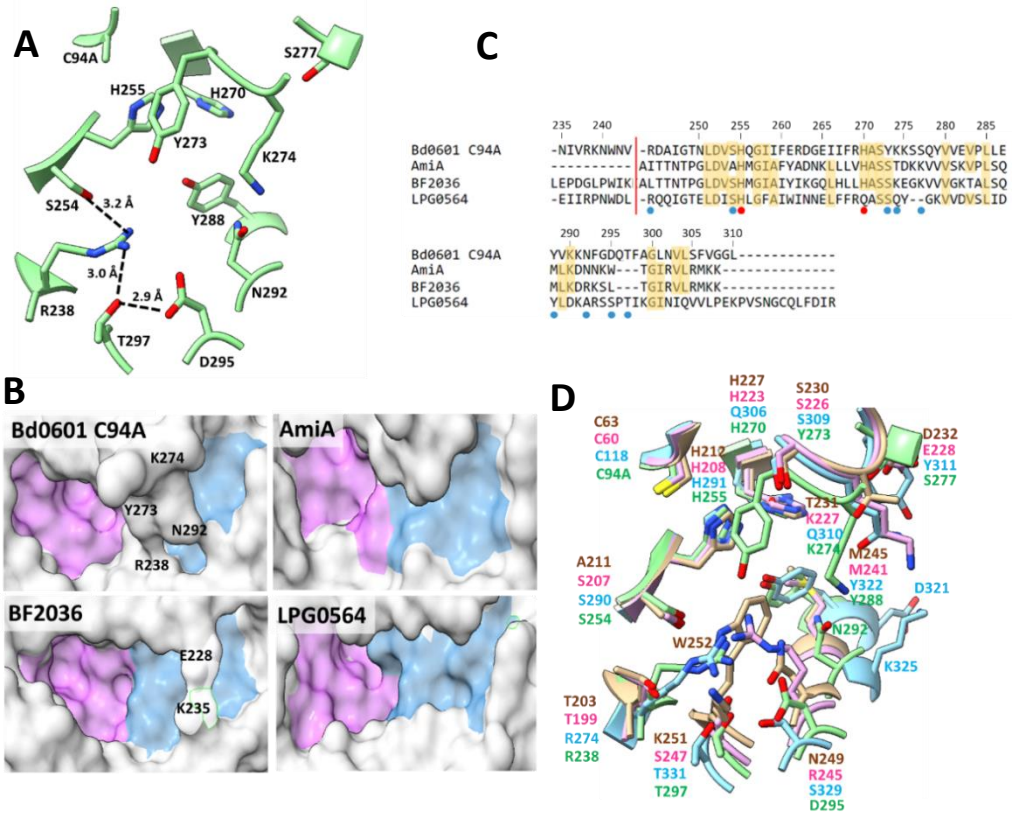


Figure 85: Stem peptide binding groove of Bd0601 C94A. A) Putative Bd0601 C94A stem peptide binding residues. Residues oriented towards the stem peptide binding groove and in proximity to the C94-H255-H270 catalytic triad are shown in stick representation. **B)** Surface representation of DUF1460 enzymes substrate binding pockets. Non-prime pocket is coloured in magenta and prime pockets in blue. The stem peptide binding groove of Bd0601 C94A is partially blocked by Y273 in contrast to the continuous non-prime/prime substrate binding region of the other DUF1460 enzymes. **C)** Sequence alignment of Bd0601 proposed stem peptide binding pocket. Catalytic residues are indicated in red dots, whereas proposed substrate binding residues are indicated in blue dots. Identical residues are highlighted in yellow. **D)** Structural superimposition of predicted stem peptide binding grooves of the DUF1460 available structures (Bd0601 C94A in green, AmiA in brown, BF2036 in magenta and LPG0564 in cyan). Bulky residue Y273 is a unique structural feature of Bd0601 C94A absent in other Duf1460s.

Putative stem peptide binding residues are located in the vicinity of the C94-H255-H270 catalytic triad of Bd0601, with sidechains oriented towards the groove proposed to bind the stem peptide of the PG molecule. Residues S254 and T297 are at hydrogen bonding distance to the terminal nitrogen groups of R238 (3.2 and 3.0 Å, respectively), orienting R238 for its possible interaction with the acidic portion of the L-Ala or γ-D-Glu of the stem peptide. T297 hydroxyl group appears to be correctly positioned via its hydrogen bonding with the D295 carboxylate group. Residues K274, S277 and N292 are located in two loops surrounding the stem peptide binding cleft and due to their location relative to the catalytic Cys-His-His triad, are proposed to interact with the mDAP of the stem peptide.

Coincidentally, the Y288 aromatic ring is partially oriented towards the cleft, which might stabilise the carbon chain of the mDAP residue. The hydroxyl group of Y288 is also exposed and possibly available for stem peptide binding.

Interestingly, as depicted in **Figure 85A** and **B**, the Bd0601 C94A prime pocket appears to be at least partially blocked by Y273, in contrast to what is observed for AmiA, BF2036 and LPG0564 (**Figure 85B**). For BF2036, residues E228 and K235 protrude out of the binding cleft, similarly to Bd0601 C94A K274 and N292, whose sidechains prolong out of the pocket.

In this regard, the sequence alignment limited to the stem peptide binding regions of Bd0601 C94A, AmiA, BF2036 and LPG0564 shown in **Figure 85C** suggests that the proposed stem peptide binding residues are partially conserved between these different DUF1460 enzymes (blue dots), at least in the vicinity of the catalytic H255 and H270 (red dots, this latter being replaced by a Q in LPG0564).

Due to the poor conservation of the DUF1460 stem peptide binding pocket sequences, structures of AmiA (brown), BF2036 (magenta), LPG0564 (cyan) and Bd0601 C94A (green) were superimposed based on their secondary structure, and equivalent residues to those proposed for Bd0601 C94A were analysed (**Figure 85D**).

As predicted by the sequence alignment, the highest conservation is observed for residues near the catalytic triad. Interestingly, Bd0601 C94A appears to be the most similar to LPG0564, sharing equivalent R238/R274 and Y288/Y322 residues, which are replaced by T203/T199 and M245/M241 in AmiA and BF2036, respectively. The basic qualities of R238 appear to be substituted by residues W252 and R245 of AmiA and BF2036, respectively.

Remarkably, no equivalents to Y273 are present in any of the DUF1460 available structures which agrees with only the Bd0601 C94A prime pocket appearing blocked. Moreover, as shown by the sequence alignment in **Figure 85C** and structural superimposition in **Figure 85D**, this position is occupied by a serine residue in the other DUF1460s.

Protein sequence alignment shows that only *Bdellovibrio* sp. ArHS and the archaea *Legionellales bacterium* RIFCSPHIGHO2 DUF1460 homologues (OM95 01295 and A3F18 02985, respectively) possess a conservative variation in the corresponding sequence position, replacing Bd0601 Y273 for a phenylalanine residue (**Figure 86**). Thus, Y273 appears to be a defining characteristic of Bd0601.



Figure 86: Bd0601 stem peptide binding groove sequence alignment. Bd0601 stem peptide binding residues (indicated in blue dots) were compared with DUF1460 orthologs OM95 01295 (*Bdellovibrio* sp. ArHS), A3F18 02985 (*Legionellales bacterium* RIFCSPHIGHO2) alongside the structurally characterised DUF1460s AmiA, BF2036 and LPG0564. Specifically, partial conservation of a bulky aromatic residue in position 273 is observed only in OM95 01295 and A3F18 02985.

5.2.1.6. Bd0601 C94A docking of PG model substrates

To further probe the hypothesis of the stem peptide binding pocket of Bd0601 C94A being blocked by Y273, docking experiments were carried out using Autodock Vina as described in section 2.3.13, using a GlcNAc-anhMurNAc-tetrapeptide substrate (PDB ID: MLD), as performed for AmiA [113].

As shown in **Figure 87A**, it is possible to fit MLD with relative confidence in the non-prime pocket of Bd0601 C94A (ΔG°= -6.9 Kcal mol^{-1}). However, it was not possible to accommodate the stem peptide of MLD on the predicted pocket for Bd0601 C94A, with all possible orientations of the peptide extending out of the surface of the enzyme.

To test the contribution of Y273 possible blocking of the stem peptide binding pocket, an *in silico* Y273S mutation was introduced to Bd0601 C94A and used as receptor for docking experiments. As depicted in **Figure 87B** and **C**, the theoretical Y273S mutation allows to partially accommodate the peptide portion of MLD in the peptide binding pocket of Bd0601 C94A/Y273S (ΔG°= -7.7 Kcal mol^{-1}), while maintaining a relatively experimentally supported orientation of the disaccharide in the non-prime pocket.

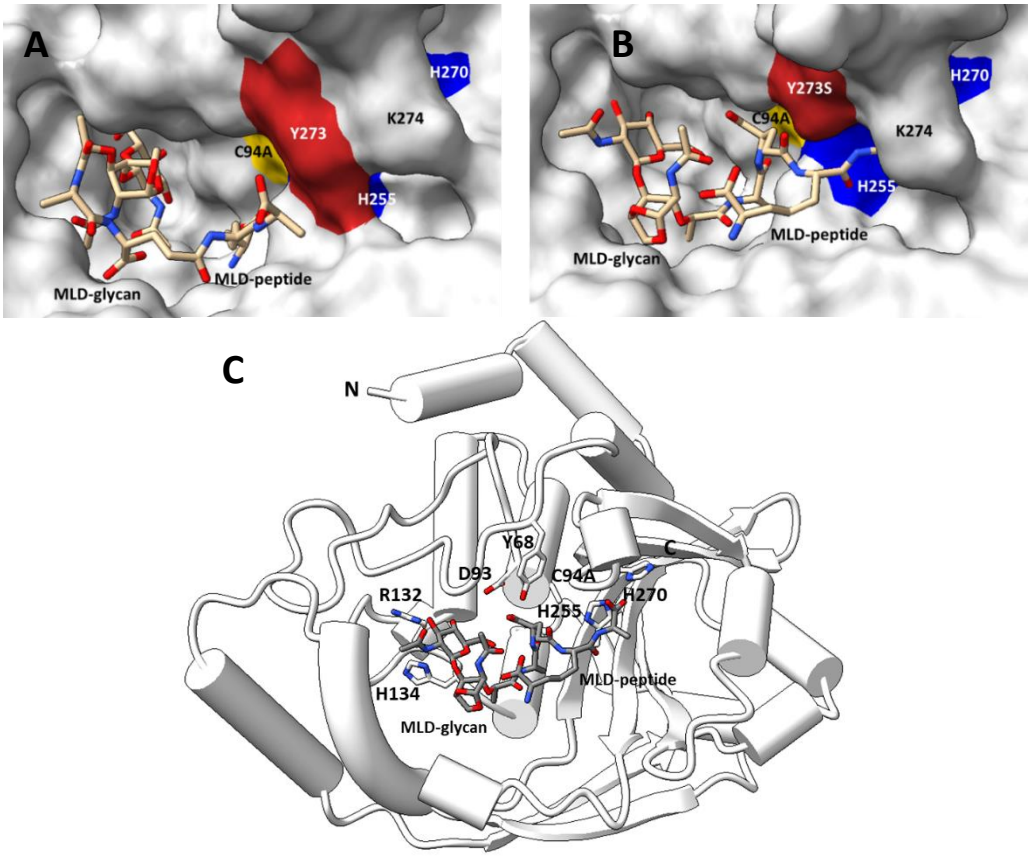


Figure 87: PG model substrate docking on Bd0601 substrate binding site. MLD stem tetrapeptide was used as model substrate for docking simulations using Autodock Vina. The catalytic triad C94A-H255-H270 surface representation is coloured in yellow and blue for the cysteine and histidine residues, respectively. **A)** MLD was docked on the predicted DUF1460 binding pocket of Bd0601 C94A structure. Y273 (in red) blocks access to the putative stem peptide binding groove of Bd0601 C94A. **B)** MLD was docked on the predicted DUF1460 binding pocket of an *in silico* double mutant Bd0601 C94A/Y273S. Substitution of Y273 for a serine, as is mostly conserved in other DUF1460, partially restores the connection between the non-prime and prime sites of the substrate binding pocket. **C)** MLD docking result of the Y273S substitution in panel B, depicting Bd0601 in cartoon representation (white) and the MLD docked ligand in stick representation (grey).

Interestingly, not fully docked MLD substrates were obtained even when the proposed obstruction of the prime pocket was released by the Y273S mutation. This could be potentially due to the presence of other sidechains partially occupying the prime pocket of Bd0601 C94A, as R238, K274 and N292, which might re-accommodate upon substrate binding. This agrees with a relatively smaller SES for the peptide binding groove of Bd0601 C94A (433 $Å^3$) compared to the average of the other solved DUF1460 enzymes (over 500 $Å^3$, on average).

5.2.2. Bd1177

5.2.2.1. Bd1177 purification and crystallisation

The nucleotide fragment of *bd1177* encoding the mature protein Bd1177 was cloned in the pET28a(+) vector for cytoplasmic expression, purifying this protein to homogeneity as detailed in section 2.3.5 (**Figure 88A**). The elution profile of Bd1177 from a Superdex™ 75 26/60 column correlates with a hydrodynamic radius of an approximately 42 kD globular protein (**Figure 88B**).

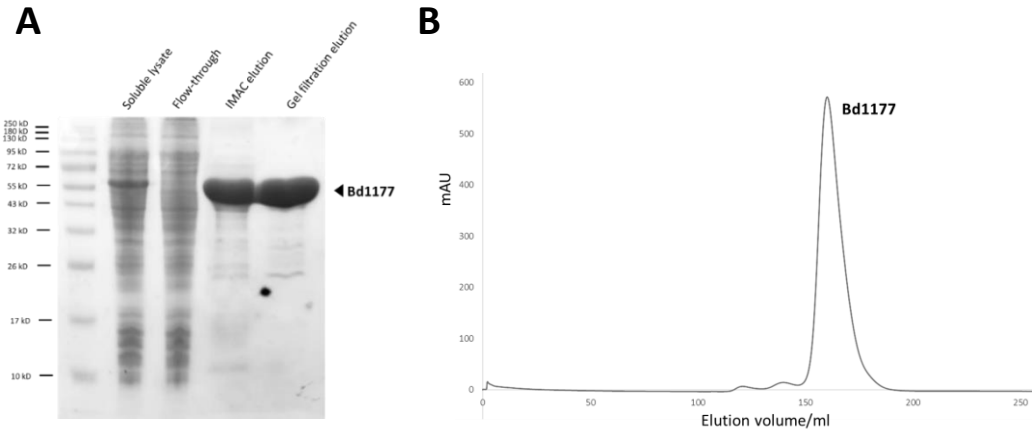


Figure 88: Bd1177 purification and gel filtration profile. Bd1177 was expressed in TB media, induced with IPTG, and purified by nickel IMAC and gel filtration chromatography. **A)** 15% SDS-PAGE purification gel for Bd1177. **B)** Gel filtration profile of Bd1177 showing a monodisperse peak corresponding to the hydrodynamic radius of a globular 42 kD protein (160 mL, approximately). The run was performed at room temperature and flow rate of 2 mL/minute in a buffer composed by 15 mM MES pH 6.0 and 200 mM NaCl.

Purified Bd1177 was concentrated as indicated in section 2.3.9 to a final concentration of 20 mg/mL and screened for crystallisation. Crystals were obtained in condition F1 of the ProPlex™ screen (0.1M Na•HEPES pH 7.0 and 20% PEG 8000) and are shown in **Figure 89**. Native and crystals derivatised with K_2PtCl_4 were harvested for data collection as detailed in section 2.3.10.

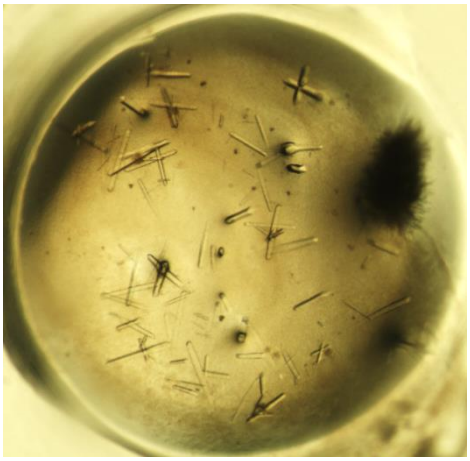


Figure 89: Bd1177 crystals. Bd1177 crystals were obtained in condition F1 of the ProPlex™ screen (0.1M Na•HEPES pH 7.0 and 20% PEG 8000) and cryoprotected with mother liquor supplemented with 30% glycerol before collection or K_2PtCl_4 derivatisation.

5.2.2.2. Bd1177 data collection and refinement

Diffraction data was collected at the IO4 beamline as indicated in section 2.1.11, collecting native and platinum derivatised datasets of Bd1177. Collected data was autoprocessed by the Xia-3dii automated pipelines for the ProPlex™ F1 crystals.

Crystallographic data was solved using Phenix Autosol in the $P4_32_12$ spacegroup for the platinum-phased dataset, finding one copy of Bd1177 in the ASU and seven heavy atom sites (**Figure 90**).

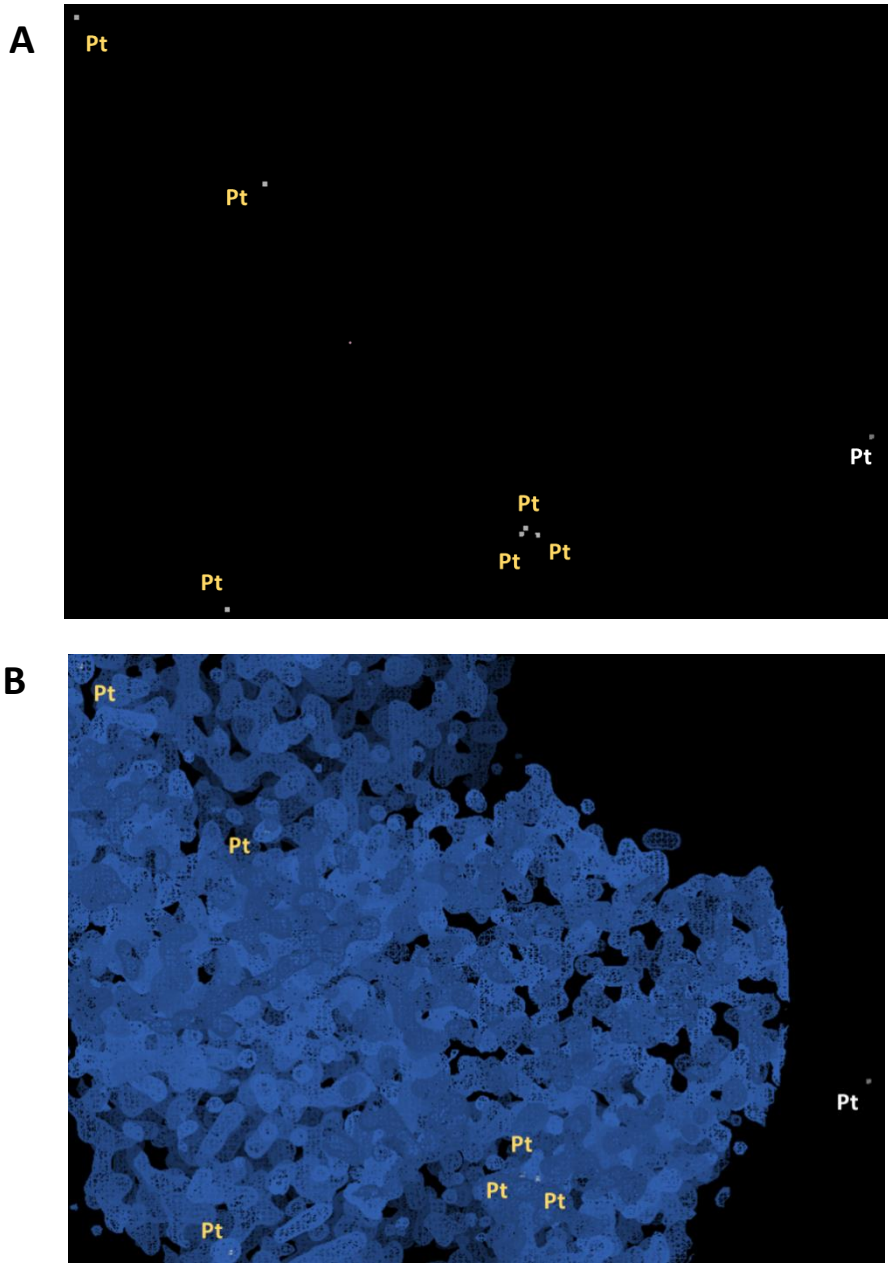


Figure 90: Bd1177 platinum phasing sites. A) Phenix Autosol platinum mapped sites (in white). Correspondence with the electron density mapped is showed for the yellow-coloured platinum sites. **B)** Electron density map ($2F_0$-F_0, σ= 1.50) for the platinum sites.

The PtSAD model was used for MR solving of the native dataset in the $P4_32_12$ spacegroup, finding one copy of Bd1177 in the ASU. Crystallographic data collection and refinement statistics are shown in **Table 15**.

Table 15: **Data collection and refinement statistics for PtSAD and native Bd1177 C-terminus**. Statistics for PtSAD and native Bd1177 C-terminus crystallographic data obtained at 2.00 and 1.70 Å, respectively, are shown.

	Bd1177 C-terminus PtSAD	Bd1177 C-terminus Native
Data Collection		
Resolution range (Å)	50.96 – 2.00	59.61 - 1.70
Space group	P 4_3 2_1 2	P 4_3 2_1 2
Cell dimensions		
a, b, c (Å)	84.25, 84.25, 98.41	84.30, 84.30, 98.82
α, β, γ (°)	90, 90, 90	90, 90, 90
Total reflections	631012 (58219)	1013088 (94489)
Unique reflections	24576 (2405)	39802 (3917)
Completeness (%)	99.92 (99.83)	99.96 (99.90)
Redundancy	25.7 (24.2)	25.5 (24.1)
<I>/σI	17.92 (3.21)	27.39 (2.88)
R_{meas}	0.131 (0.626)	0.0724 (0.657)
$CC_{1/2}$	0.999 (0.969)	1 (0.97)
CC_{anom}	0.3 (0.1)	
Refinement		
Maximum resolution (Å)		1.70
Number of reflections used		39787 (3913)
Number of reflections for R_{free}		2079 (183)
R_{work}		0.1729
R_{free}		0.1907
RMSD		
Bonds (Å)		0.008
Angles (°)		1.07
Number of atoms		
Protein		1845
Ligand/ion		20
Water		264
B-factors		
Protein		25.19
Ligand/ion		37.74
Water		37.70
Ramachandran favoured (%)		97.29
Ramachandran allowed (%)		2.71
Ramachandran outliers (%)		0.00
Rotamer outliers (%)		0.51

*Values in parenthesis correspond to the highest-resolution shell

Nevertheless, data solving revealed that Bd1177 diffraction corresponds to a proteolyzed fragment of Bd1177, specifically, the C-terminus of Bd1177. This obtained structure will be referred as Bd1177 C-terminus from here onwards.

5.2.2.3. Bd1177 partially proteolyzed C-terminal structure

Bd1177 solved structure traces 222 of the 524 residues of the mature protein, specifically, residues 322 to 544, corresponding to the C-terminus of the protein (**Figure 91**).

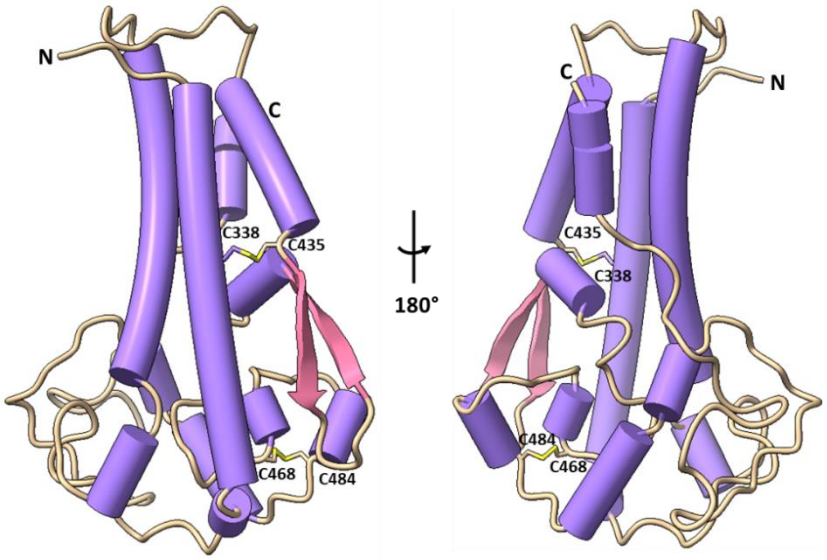


Figure 91: Bd1177 proteolyzed C-terminal structure. Cartoon representation of the native Bd1177 proteolyzed C-terminal structure shown in two 180° views apart. Alpha helices and beta strands are coloured in purple and pink, respectively. The C-terminal half of Bd1177 adopts a bulb-like structure that is connected to the proteolyzed N-terminal by a precited flexible linker. The C-terminal fold is stabilised by two disulphide bonds, C338-C435 and C468-C484.

The C-terminal fragment of Bd1177 adopts a bulb-like topology consisting of a central three helical bundle packed against shorter α-helices (in purple) and two antiparallel β-strands (in pink). The whole fold is further stabilised by two disulphide bonds, between C338-C435 and C468-C484.

Submission of the solved C-terminal fragment of Bd1177 to the Dali server does not reveal any significant matches with publicly available protein structures in the PDB, suggesting that this portion of the protein corresponds, at least partially, to a novel domain.

To stop proteolysis of Bd1177, a protease cocktail was added during Bd1177 purification and crystal screening, however, only systematic crystallisation of Bd1177 C-terminal fragment was observed.

5.2.2.4. Bd1177 C86A and Bd1177 C86A/Y124C/Y508C design

Considering the lack of information provided by the proteolyzed fragment of Bd1177, a structural prediction for Bd1177 by RosettaFold was carried out [180]. The obtained structural model is shown in **Figure 92**.

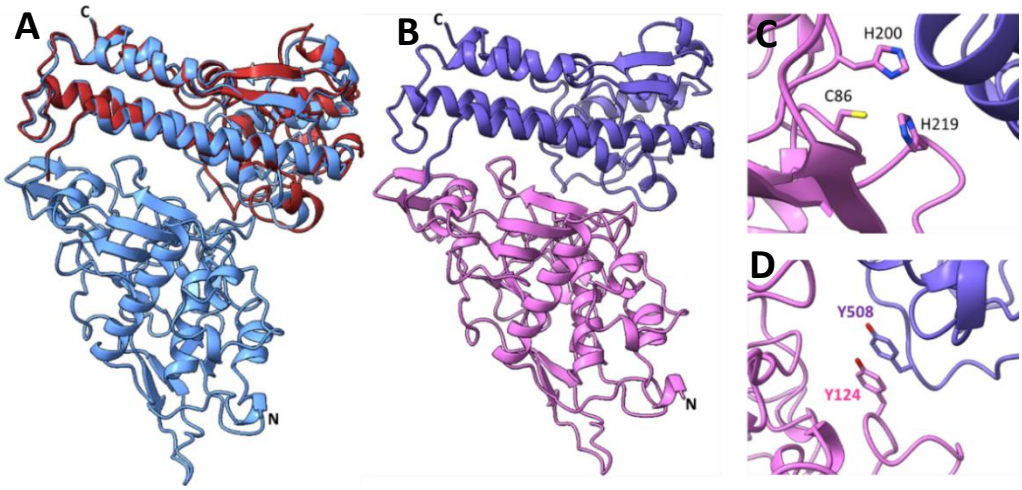


Figure 92: Bd1177 predicted structural model. A) Structural superimposition of the Bd1177 C-terminal crystal structure (in red) and the RosettaFold full-length predicted model (in blue). **B)** The predicted full-length structure of Bd1177 adopts a bi-lobulated fold (N-terminal domain in pink and C-terminal domain in purple). **C)** Predicted active site of Bd1177 situated in the N-terminal domain, centred on the nucleophilic C86. **D)** Aromatic stacking between residues Y124 and Y508 is shown as one of the interacting points between the N-terminal and C-terminal domain of Bd1177 in the RosettaFold predicted model.

As seen in **Figure 92A**, the full-length structural prediction of Bd1177 (in blue) superimposes reliably with the crystallised C-terminal fragment of Bd1177 (in red) with an overall RMSD of 2.59 Å, suggesting that the obtained model could be used for construct design to obtain the full-length experimental structure of Bd1177.

The full-length predicted structure of Bd1177 is formed by two discrete structural domains, the presumable N-terminal catalytic domain (in pink) and the cryptic crystallised C-terminal proteolyzed fragment of Bd1177 (in purple, from here onwards, C-terminal domain of Bd1177), connected by a disordered loop comprising residues 316 to 325 (**Figure 92B**).

Figure 92C shows the putative active site of Bd1177 which coincides with a cysteine peptidase/amidase consisting of a catalytic triad C86-H200-H219. Thus, as a first approach, Bd1177 truncated constructs (residues 20-321) were designed to purify and crystallise the N-terminal portion of Bd1177, however, no soluble protein was obtained for these Bd1177 constructs.

Considering this, a catalytically null mutant, Bd1177 C86A, was designed to produce a more stable version of Bd1177 lacking a free cysteine residue and to avoid any possible self-proteolysis. Additionally, analysis of the Bd1177 full-length predicted structure using the Disulfide by Design 2.0 web server (http://cptweb.cpt.wayne.edu/DbD2/) reveals that

the interaction between the two domains of Bd1177 might be mediated, at least partially, by the aromatic stacking interaction between Y124 (from the N-terminal domain) and Y508 (from the C-terminal domain) (**Figure 92D**) [181]. Thus, a disulphide locked and catalytically null triple mutant Bd1177 C86A/Y124C/Y508C was designed to promote crystallisation of the full-length structure of Bd1177.

5.2.2.5. Bd1177 C86A and Bd1177 C86A/Y124C/Y508C purification and crystallisation

Soluble protein was successfully purified for both mutants Bd1177 C86A and Bd1177 C86A/Y124C/Y508C in a similar fashion to native Bd1177 (**Figure 93A**). Moreover, gel filtration elution profiles for both mutants are identical between each other and with native Bd1177, suggesting that mutation does not impact drastically on Bd1177 three-dimensional structure (**Figure 93B**).

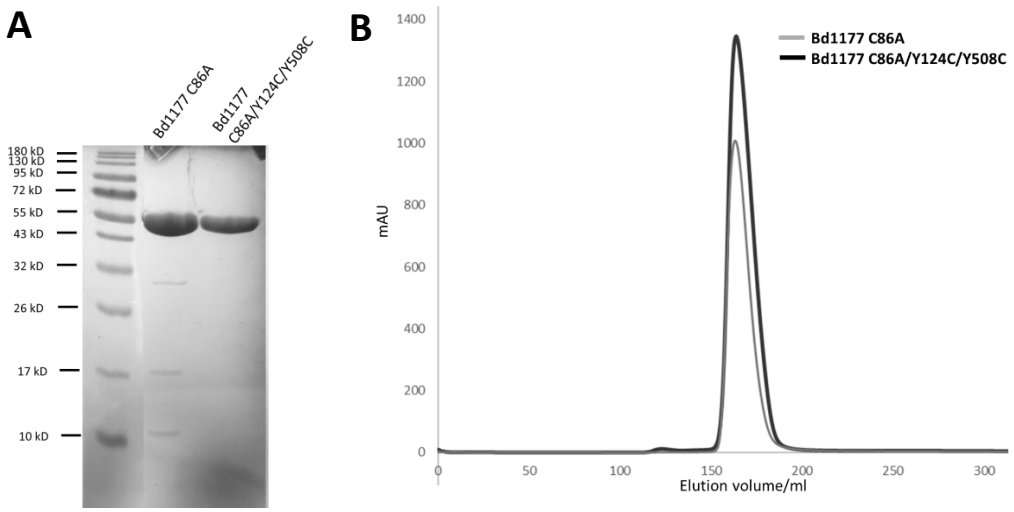


Figure 93: Purification and gel filtration profile of Bd1177 C86A and Bd1177 C86A/Y124C/Y508C. Bd1177 C86A and Bd1177 C86A/Y124C/Y508C were expressed in TB media, induced with IPTG, and purified by nickel IMAC and gel filtration chromatography analogously to wild-type Bd1177. **A)** 15% SDS-PAGE gel showing purity of Bd1177 C86A and Bd1177 C86A/Y124C/Y508C after IMAC and gel filtration chromatography. **B)** Gel filtration profile of Bd1177, Bd1177 C86A and Bd1177 C86A/Y124C/Y508C showing monodisperse peaks corresponding to the hydrodynamic radius of globular 42 kD proteins (160 mL, approximately). Equivalence between elution profiles suggests that Bd1177 mutants adopt a conformation similar to Bd1177 in solution. The run was performed at room temperature and flow rate of 2 mL/minute in a buffer composed by 15 mM MES pH 6.0 and 200 mM NaCl.

Bd1177 C86A and Bd1177 C86A/Y124C/Y508C were concentrated to 20 mg/mL and screened for crystallisation conditions. Crystal formation was only observed for Bd1177 C86A/Y124C/Y508C in JCSG-plus™ condition A7 (0.1M CHES pH 9.5 and 20% PEG 8000)

(**Figure 94**). These crystals were cryoprotected with mother liquor supplemented with 30% of glycerol and harvested for data collection.

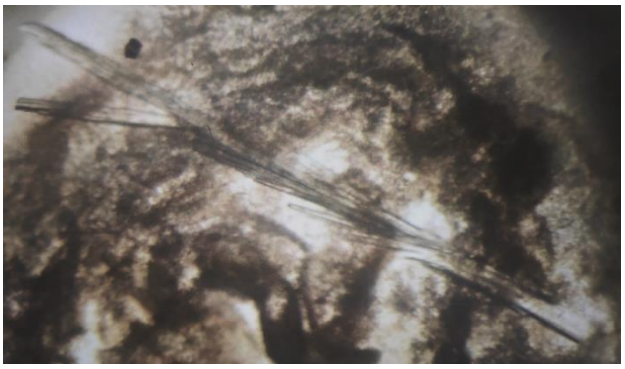


Figure 94: Bd1177 C86A/Y124C/Y508C crystals. Crystals were obtained in condition A7 of the JCSG-plus™ screen (0.1M CHES pH 9.5 and 20% PEG 8000).

5.2.2.6. Bd1177 C86A and Bd1177 C86A/Y124C/Y508C data collection and refinement

Diffraction data was collected at the IO4 beamline as indicated in section 2.1.9, collecting native datasets of Bd1177 /Y124C/Y508C.

Diffraction data for the Bd1177 C86A/Y124C/Y508C triple mutant was autoprocessed by the Xia2-3dii automated pipeline and solved by molecular replacement at 2.2 Å resolution in the P1 spacegroup with four copies of the protein in the ASU. Crystallographic data collection and refinement statistics are shown in **Table 16**.

Table 16: Data collection and refinement statistics for Bd1177 C86A/Y124C/Y508C. Crystallographic data was obtained at 2.2 Å maximum resolution.

	Bd1177 C86A/Y124C/Y508C
Data Collection	
Resolution range (Å)	53.4 - 2.2
Space group	P 1
Cell dimensions	
a, b, c (Å)	45.46, 110.83, 116.94
α, β, γ (°)	101.68, 90.02, 90.02
Total reflections	352512 (16992)
Unique reflections	100717 (5435)
Completeness (%)	88.56 (47.48)
Redundancy	3.5 (3.1)
<I>/σI	12.83 (3.41)
R_{meas}	0.06632 (0.3082)
R_{pim}	0.03497 (0.1681)
$CC_{1/2}$	0.998 (0.953)
Refinement	
Maximum resolution (Å)	2.20
Number of reflections used	100528 (5426)
Number of reflections for R_{free}	4920 (288)
R_{work}	0.1940
R_{free}	0.2331
RMSD	
Bonds (Å)	0.018
Angles (°)	1.72
Number of non-hydrogen atoms	
Protein	16626
Ligand/ion	104
Water	1323
B-factors	
Protein	47.29
Ligand/ion	46.00
Water	44.03
Ramachandran favoured (%)	96.50
Ramachandran allowed (%)	3.12
Ramachandran outliers (%)	0.38
Rotamer outliers (%)	5.73

*Values in parenthesis correspond to the highest-resolution shell

5.2.2.7. Bd1177 C86A/Y124C/Y508C overall structure and N-terminal domain

The obtained crystallographic structure of Bd1177 C86A/Y124C/Y508C traces the 524 residues of the mature protein (residues 21-544) and is depicted in **Figure 95A**, in two 90-degree views.

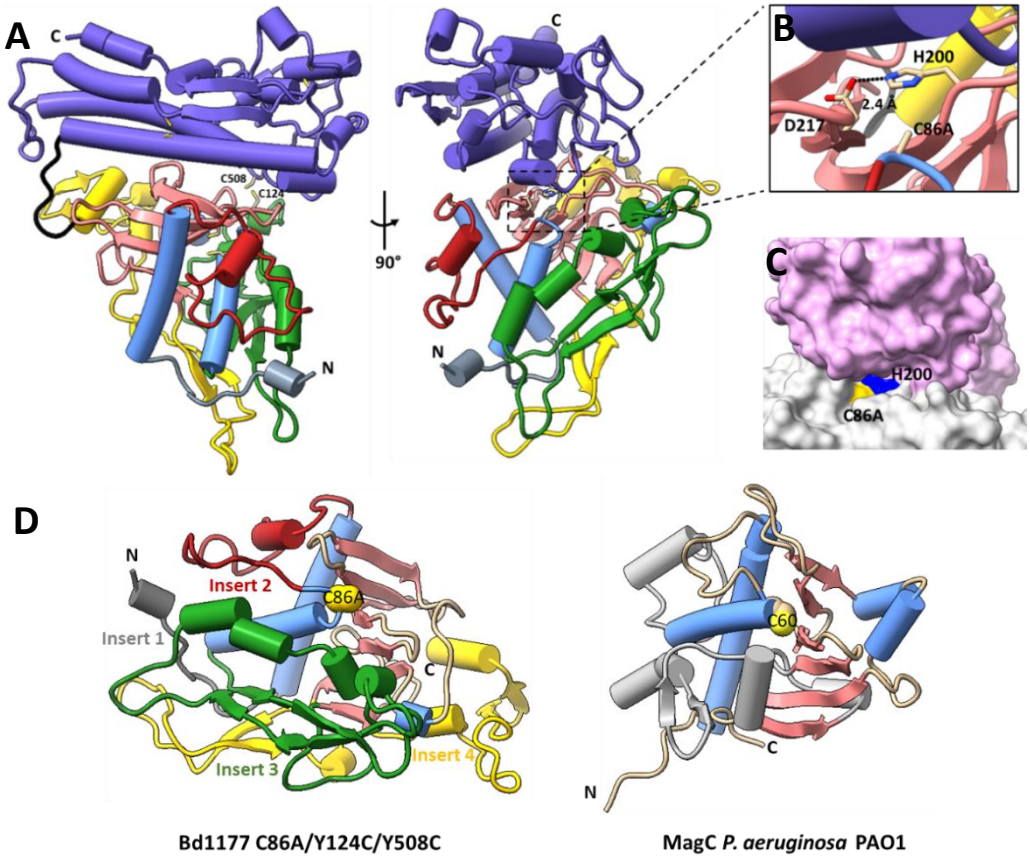


Figure 95: Bd1177 C86A/Y124C/Y508C crystal structure, active site, and putative catalytic core. A) Bd1177 C86A/Y124C/Y508C depicted in cartoon representation, in two 90° angle views. The structure confirms a bi-lobulated fold connected by a linker, wherein the N-terminal domain incorporates four structural insertions (in grey, red, green, and yellow) surrounding a NlpC/P60-like core (coloured by secondary structure, alpha helices in blue and beta strands in pink) and is connected to the previously crystallised C-terminal domain (in purple). **B)** Bd1177 C86A/Y124C/Y508C putative active site formed by the catalytic triad C86-H200-D217. **C)** The Bd1177 C86A/Y124C/Y508C putative active site is located in the interface between the N-terminal and C-terminal domain (in grey and pink, respectively). **D)** NlpC/P60 Bd1177 C86A/Y124C/Y508C central core depicted with structural insertions 1 to 4 and compared to the NlpC/P60 core of MagC (PDB ID: 7BK8).

As predicted by the RosettaFold model of native Bd1177, the Bd1177 disulphide-locked triple mutant depicts a bilobular three-dimensional structure consisting of the previously non-crystallisable N-terminal domain (residues 21-317) and the cryptic C-terminal domain (residues 327-544, in purple) connected by a flexible linker (residues 318-326). The overall fold of Bd1177 is further stabilised by three disulphide bonds, C64-C224 in the N-terminal domain, and C338-C435 and C468-C484 in the C-terminal domain. As designed, both lobes of the protein are linked by the introduced disulphide bond C124-C508.

The N-terminal domain of Bd1177 comprises a NlpC/P60 putative catalytic core with a canonical papain-like fold. The NlpC/P60 core of Bd1177 is formed by a six-stranded half

beta barrel sandwiched against two long alpha helices (coloured by secondary structure, beta strands in pink and alpha helices in blue). The putative active site of Bd1177 is comprised by the nucleophile C86 (C86A in the crystallised protein), and the two conserved acid-base residues H200 and D217 (**Figure 95B**). The NlpC/P60 core of Bd1177 is structurally related to the NlpC/P60s MagC from *P. aeruginosa* PAO1 (PDB ID: 7BK8), which superimposes with Bd1177 on the papain-like core with an RMSD of 3.1 Å [182]. The side chains of the catalytic triad C86-H200-D217 extend towards the interface between the N-terminal and C-terminal domain, suggesting that the C-terminal domain might participate in substrate selection and/or binding by acting as a capping domain (**Figure 95C**).

Four structural insertions of the NlpC/P60 core are exclusive to Bd1177: insert 1 (grey, residues 21-35), insert 2 (red, residues 55-84), insert 3 (green, residues 102-177) and insert 4 (yellow, residues 248-317) (**Figure 95D**). As opposed to the DUF1460 insertions in Bd0601, the structural insertions of Bd1177 do not appear to define a discrete substrate binding pocket, however they might participate in the subcellular localisation of the enzyme, protein-protein interactions, or substrate selection by modulating the enzyme conformation enabling substrate binding.

5.2.2.8. Structural interaction between N- and C-terminal domains of Bd1177 C86A/Y124C/Y508C

The C-terminal capping domain of Bd1177 interacts with the N-terminal domain by residues mainly located in the first alpha helix of the C-terminal domain (residues 326-353). Other than the introduced Y124C and Y508C mutated pair, the strongest interaction is the 2.6 Å salt bridge between residues R314 (in insert 4 of the N-terminal domain) and D329 (fist helix of the C-terminal domain), which is located towards the hinge region defined by the flexible linker connecting both Bd1177 domains (**Figure 96A**). In close proximity to this salt bridge, Q332 of the C-terminal domain forms a 3.3 Å hydrogen bond with R212 and a 3.0 Å hydrogen bond with D210. The middle point of the interacting interface is comparatively less extensive, with a sole 3.1 Å hydrogen bond between K232 and D336 (**Figure 96B**). The rest of the stabilising interactions between N-terminal and C-terminal domain surround the putative catalytic triad of Bd1177: a 2.6 Å hydrogen bonding between T223 and R384, and the 3.6 Å cation-pi interaction between R387 and H200 of the catalytic triad, respectively.

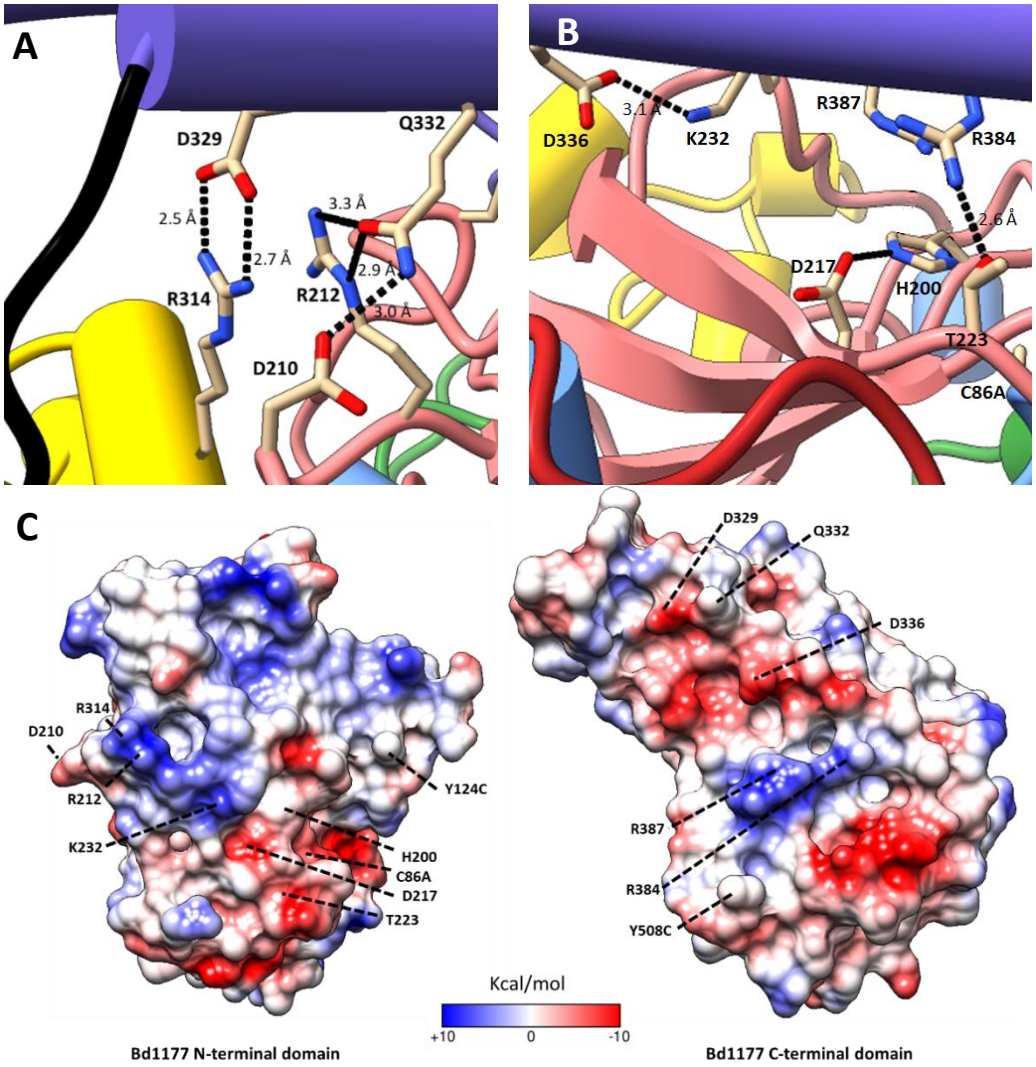


Figure 96: Analysis of the interacting interface between the N- and C-terminal domains of Bd1177 C86A/Y124C/Y508C. A) Interacting sidechains in proximity to the hinge region connecting the N- and C-terminal domains of Bd1177 C86A/Y124C/Y508C. The most remarkable interaction corresponds to a salt bridge formed between residues R314 (N-terminal domain) and D329 (C-terminal domain). **B)** Interacting interface proximal to the catalytic triad C86-H200-D217. Conserved H200 and the closely located T223 appear at hydrogen bonding distance to R387 and R384 of the C-terminal domain, respectively. **C)** Interface view of the electrostatic surface potential of the N- and C-terminal domains of the Bd1177 triple mutant. Residues visualized in panels A and B are indicated.

As shown in **Figure 96C**, the putative C86-H200-D217 active site is mostly inaccessible due to it being covered by the C-terminal domain. It is possible that the introduction of three mutations, particularly the Y124C-Y508C disulphide bond, might affect the three-dimensional conformation of Bd1177, however the gel filtration elution profiles of the mutants are similar to the native protein. This, in addition to the RosettaFold predicted model and an AlphaFold2 prediction for Bd1177 depicting a similar conformation to the

experimentally determined structure, suggests that the mutant adopts a near-native conformation.

Additionally, analysis of the average B-factors per residue in the Bd1177 triple mutant structure suggests that the linker connecting the N- and C-terminal domains of Bd1177 is not highly mobile in the crystal arrangement (**Figure 97**). These results support the hypothesis that the C124-C508 would restrict conformational fluctuations of Bd1177 without altering its overall conformation.

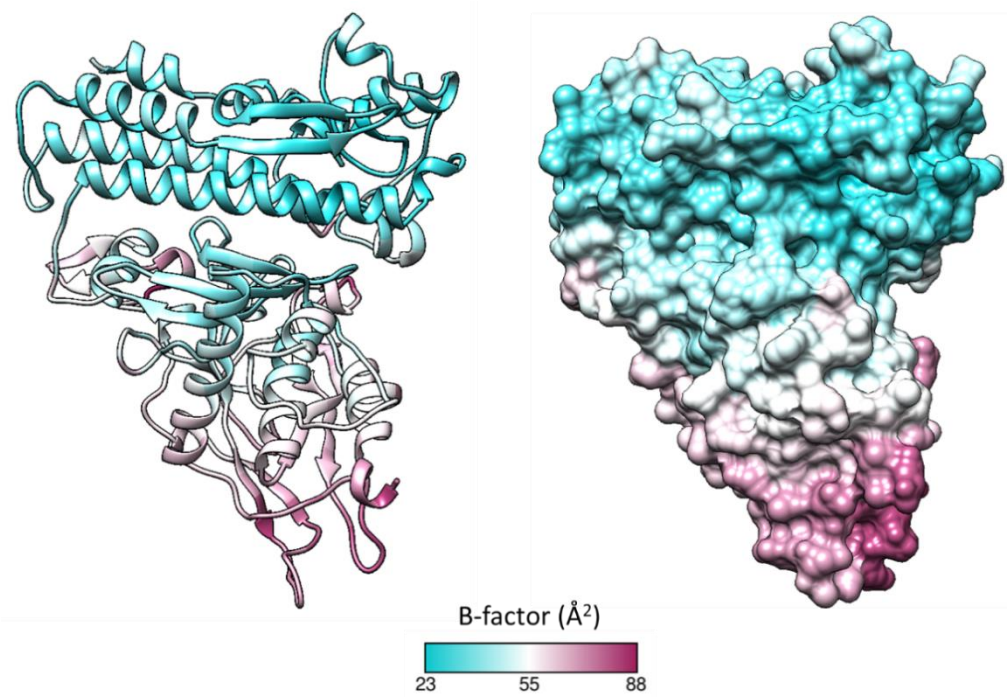


Figure 97: B-factor colouring of Bd1177 C86A/Y124C/Y508C. B-factor analysis of cartoon and surface representation of Bd1177 triple mutant. B-factors range from 23 to 88 Å^2 (cyan to magenta, respectively), with the highest values corresponding to the N-terminal NlpC/P60-like domain. Lowest B-factor values are generally seen for the hinge loop and the C-terminal domain, suggesting minimal fluctuations of Bd1177 C86A/Y124C/Y508C in the crystal lattice.

Structural data was used for calculating low frequency normal modes to identify possible conformational changes of Bd1177 using the *elNémo* (Elastic Network Model, ENM) webserver (http://www.sciences.univ-nantes.fr/elnemo/). The initial and final conformations of the calculated trajectory are shown in **Figure 98** in red and blue, respectively.

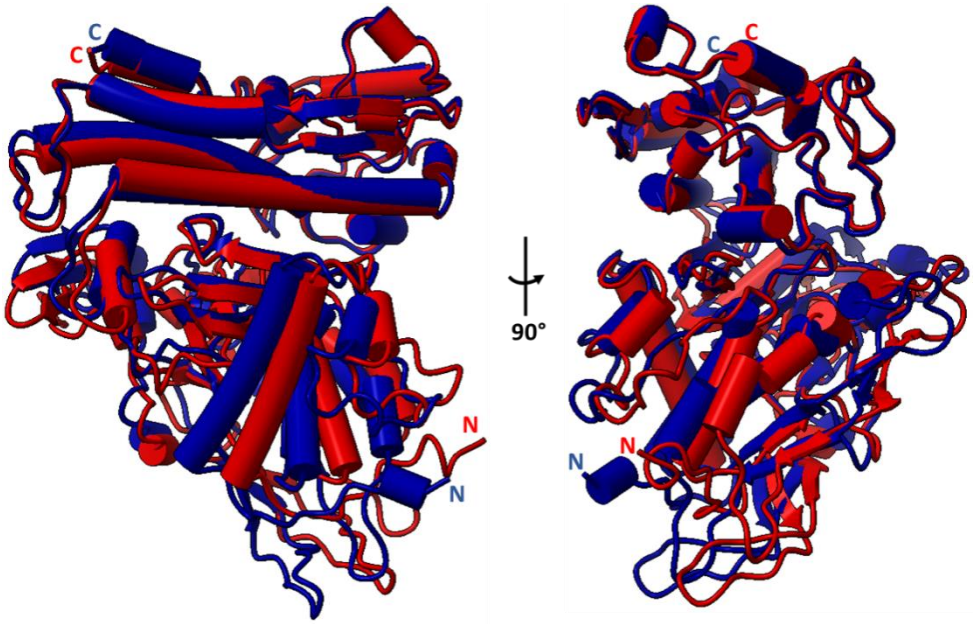


Figure 98: Simulated conformational change of Bd1177. Bd1177 possible conformational opening is calculated based on the crystallised structure using the ENM approximation and is shown in cartoon representation in two 90° angle apart views. Crystallised closed state of Bd1177 is shown in red, while the calculated final conformation is depicted in blue.

A 4.15 Å overall shift in conformation is observed from the crystallised closed state (in red) to a relatively more open conformation (in blue). Apart from the loop regions (with intrinsically higher B-factor values), the two 90° views of Bd1177 show structural fluctuations mainly in the C-terminal domain, which displaces in a rigid body-like fashion. This model correlates with an opening and closing conformational change mechanism of Bd1177 to accommodate possible PG substrates. This proposed conformational change possibly requires disengaging the hydrogen bonding interactions stabilising the N-terminal/C-terminal domain interface, which could be potentially induced upon substrate recognition and binding.

To analyse the conservation of residues proposed to participate in the interaction between N- and C-terminal domains of Bd1177, a BLAST search was carried out first in *B. bacteriovorus* HD100, identifying only protein Bd2480 as a significant homologous. Bd2480 residues are mostly identical or conserved in their chemistry when compared to Bd1177, with few exceptions of loop and surface exposed residues (68.99% sequence identity).

Thus, a broader Bd1177 BLAST search was carried out across different organisms, including other *Bdellovibrio* species and strains. Seven representative Bd1177

homologues are shown in **Figure 99**: Bd2480 from *B. bacteriovorus* HD100, A0A514XLV5 from *Bdellovibrio* sp. Zap7, A0A254QLJ5 from *Bdellovibrio* sp. SKB1291214, Bdt2422 from *B. bacteriovorus* Tiberius, H7Z71_07925 from *Moraxellaceae bacterium*, H7301_03220 from *Cryobacterium* sp. and WP_167082726 from *Rhizomicrobium palustre*. Overall protein sequence is highly conserved, with over 60% sequence identity across the different *Bdellovibrio* species, and over 45% for non-*Bdellovibrio* bacteria, suggesting they adopt a similar three-dimensional structure.

Residues proposed to be involved in the interaction between the N- and C-terminal domains of Bd1177 are mostly conserved. The salt bridge pair R314-D329 is conserved across the different homologues, but interestingly, the aspartic acid residue is replaced by a glutamic acid in non *Bdellovibrio* bacteria. On the contrary, the interaction between residues D210-Q332-R212 is specific for *B. bacteriovorus*.

Further, residues K232-D336 are mostly conserved in the different homologues except *Cryobacterium* sp. and *R. palustre*, wherein the aspartic acid is replaced by a glycine or a serine, respectively. Due to the high conservation of the basic K232 (which is replaced by an arginine residue in *R. palustre*), it is possible for this residue to participate in substrate binding.

Moreover, the H200-R387 and T223-R384 pairs are conserved in these homologues, which in addition to their close proximity to the catalytic C86 suggests that these residues might participate in substrate recognition or binding.

Finally, residues Y124 and Y508, which are not shown in the sequence alignment, are conserved in all Bd1177 homologues, indicating that the aromatic ring stacking possibly occurs in the closed state of all these enzymes.

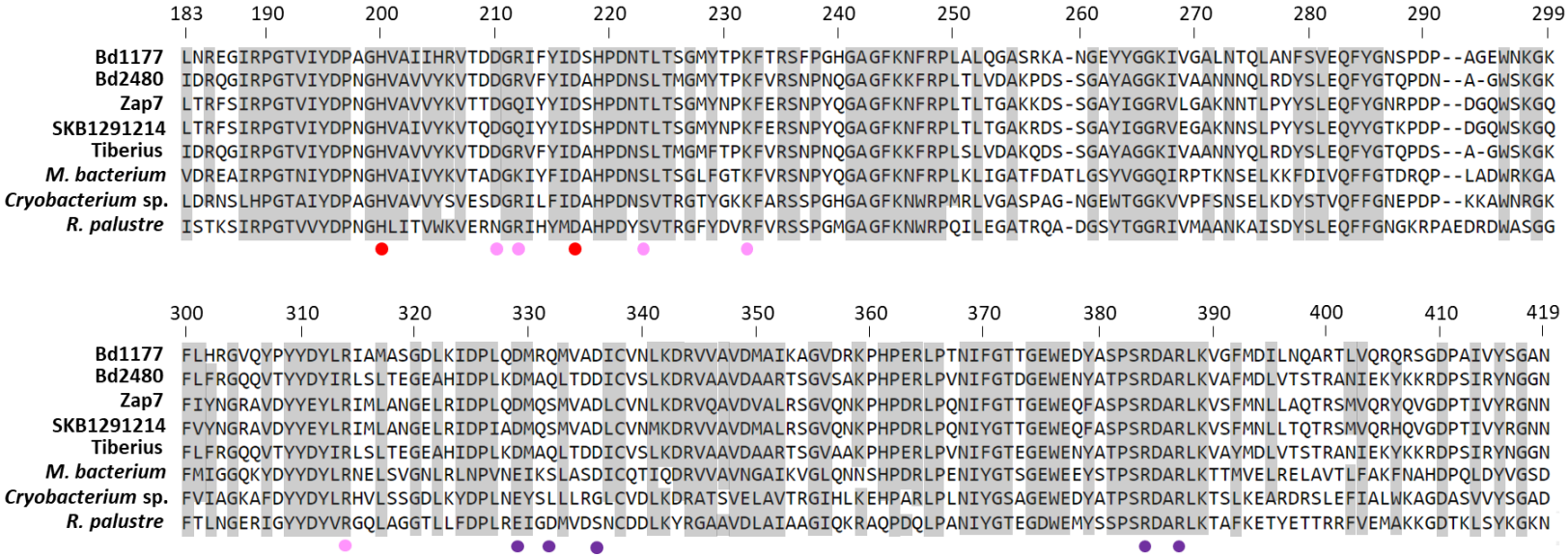


Figure 99: Conservation of N- and C-terminal interface interacting residues in Bd1177 homologues. Seven representative Bd1177 homologues were identified by BLAST, including *B. bacteriovorus* Bd2480 (strain HD100) and Bdt2422 (strain Tiberius) and homologues from *Bdellovibrio* sp. Zap7 and SKB12911214. Additionally, non *Bdellovibrio* homologues from *Moraxellaceae bacterium*, *Cryobacterium* sp. and *Rhizomicrobium palustre* were also included. Residue conservation is shown in grey. Catalytic residues, N-terminal interacting residues and C-terminal interacting residues are indicated with red, pink, and purple dots, respectively. Residue numbering is respect to Bd1177 protein sequence.

5.2.2.9. Active site analysis and possible substrate binding residues of Bd1177 C86A/Y124C/Y508C

Analysis of the residues surrounding the C86-H200-D217 catalytic triad (<10 Å linear distance) shows that residues D85, D217, H219, D221, T225, E377, S383 and R506 may potentially participate in substrate binding upon displacement of the C-terminal domain (**Figure 100**).

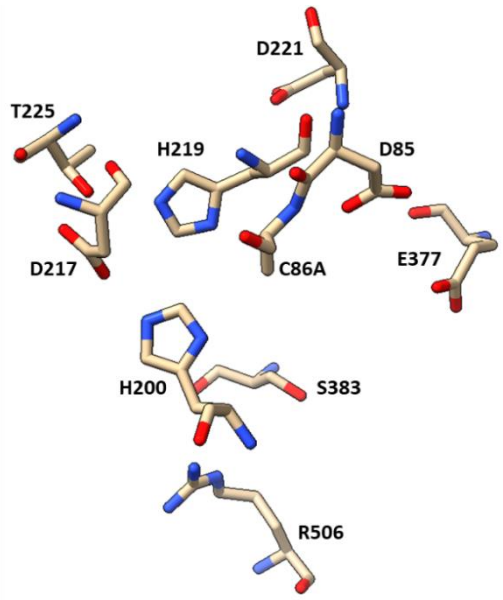


Figure 100: Potential Bd1177 substrate binding residues. Residues in proximity to the catalytic triad C86-H200-D217 are depicted in stick representation. Residues are highly conserved in different Bd1177 homologues, including non *Bdellovibrio* organisms.

Comparative structural search using the Dali webserver reveals poor conservation of the overall structure, only finding significant matches for the NlpC/P60 catalytic core of Bd1177.

Specifically, only MagC from *P. aeruginosa* (PDB ID: 7BK8, RMSD= 3.1 Å, Z= 6.8), COG0791 from *Nostoc punctiforme* (PDB ID: 2EVR, RMSD= 3.8 Å, Z= 6.7) and Acd24020 from *C. difficile* (PDB ID: 7CFL, RMSD= 2.9 Å, Z= 6.3) may be considered significant matches, with no results for the C-terminal domain of Bd1177. Of these, both COG0791 and Acd24020 have been proven to have PG endopeptidase activity, whereas MagC has been shown to be inactive, acting as a macroglobulin scaffolding protein instead [151], [159], [182].

As expected for NlpC/P60 CWHs, even though the catalytic core retains a similar papain-like fold, residue conservation outside the catalytic triad is poor, limiting prediction of

ligand binding sidechains. Coincidentally, none of the polar residues in proximity to the proposed Bd1177 catalytic triad C86-H200-D217 are conserved among the identified NlpC/P60 structures with exception of D85.

It is not possible to align the amino acid sequence of Bd1177 and the identified NlpC/P60s due to their divergence even when limiting the alignment only to the catalytic domain. Moreover, in the Dali-identified NlpC/P60 structures, the catalytic triad of the enzymes deploy a histidine as the third residue, whereas in Bd1177 it has been replaced with an aspartate.

Altogether, these results suggest that Bd1177 defines a new family of NlpC/P60 peptidoglycan endopeptidases with a unique C-terminal capping domain, different from the DUF1460 Bd0601 amidohydrolase.

5.3. Discussion

5.3.1. Bd0601 C94A

5.3.1.1. Bd0601 C94A express only as a catalytically null mutant

Bd0601 is a DUF1460 amidase encoded in the *bd0601-0603* operon and targeted for periplasm secretion by a SecI signal peptide. Three DUF1460 ortholog structures have been solved and are publicly available in the PDB: AmiA, BF2036 and LPG0564, defining a family of tetra- and tripeptide PG amidohydrolases in apo and ligand-bound states [113].

Remarkably, even though different constructs and expression conditions were tested, only the codon-optimised catalytically null mutant Bd0601 C94A was solubly expressed, which was not necessary for the other crystallised DUF1460s. Native Bd0601 proved to be toxic for *E. coli* when overexpressed either in the cytoplasm or the periplasm, which could be due to its proposed hydrolytic activity directed towards *E. coli* PG and PG precursors.

Attempts to co-express the native or catalytically null C94A mutant of Bd0601 with Bd0602 and B0603, with the working hypothesis that these may act as immunity proteins protecting the predator from Bd0601 hydrolytic activity, proved unsuccessful due to the latter proteins being insoluble under all conditions tested. Thus, it is not possible to discard that native Bd0601 would have been successfully expressed in combination with soluble constructs of Bd0602 and/or Bd0603, or if Bd0602 and Bd0603 are solubilised and purified as membrane-bound proteins.

5.3.1.2. Bd0601 C94A is a DUF1460 amidase with a conserved structure but divergent sequence

Bd0601 C94A crystallised in apo and GlcNAc-bound forms which were solved at 3.95 and 2.05 Å resolution, respectively. No significant conformational changes were observed upon GlcNAc binding, as for the GlcNAc- and anhMurNAc-GlcNAc-bound forms of AmiA [113].

The three-dimensional structure of Bd0601 C94A depicts a central papain-like fold with a catalytic C94-H255-H270 triad and four structural insertions that surround the catalytic core. These DUF1460 specific structural insertions define two discreet ligand-binding pockets: a prime pocket, which binds the sugar portion of PG molecules, and a non-prime pocket, which accommodates the stem peptide region of PG on a shallow surface groove.

The prime pocket of DUF1460 amidohydrolases is the most conserved sequence wise. Sugar binding residues are mostly unchanged with respect to other DUF1460s, wherein D93, R132 and H134 stabilise the chair conformation of the GlcNAc ring and Y68 would potentially stabilise the carboxylate group of anhMurNAc.

In contrast to what was found for the sugar biding pocket of DUF1460s, the non-prime pocket has been poorly analysed and even overlooked due to its high sequence variability. For Bd0601, general features such as the placement of R238 as a possible mDAP binding residue are, to some extent, conserved among the solved DUF1460s. Nonetheless, the most outstanding feature of Bd0601 C94A compared to the available solved structures of other DUF1460 orthologs is the partial blocking of the stem peptide binding groove of the enzyme by Y273, which substitutes the nearly invariant serine residue in this position.

As proposed by surface volume inspection of Bd0601 and preliminary docking studies, Y273 blocks the stem peptide binding groove of the enzyme, which would potentially impair catalysis. This positioning of an aromatic residue is almost exclusive to *B. bacteriovorus* HD100, with few exceptions. Since GlcNAc binding to the prime pocket does not alter the enzyme conformation, reorientation of Y273 would likely require either recognising a specific portion of the peptide or interaction of Bd0601 with an activating partner (or a binding of a substantially different peptide substrate).

It has been observed that in *S. aureus* the amidase LytH coordinates sacculi synthesis with cell division [115], [167]. The activity of LytH is triggered by its association with the

transmembrane tetratricopeptide repeat (TPR) containing protein ActH, limiting LytH hydrolytic activity to the poles of the dividing cell. On the other hand, *E. coli* AmiA and AmiB, two NALAA CWAs, are controlled via binding to the LytM motif of periplasmic cell division regulator EnvC [173]. Other amidases are naturally in s self-inhibited state, such as the *Streptococcus pneumoniae* PcsB CHAP [183]. This inhibition is caused by the catalytic domain interacting with the enzyme's coiled-coil domain and is relieved by the ATP-dependent conformational change of its interacting partner, FtsEX [183].

DUF1460s are part of a group of NlpC/P60s that has been poorly characterised, nevertheless, the limited available information indicates that the stem peptide binding groove is accessible, and no partner is required for their hydrolytic activity. Thus, Bd0601 potentially requiring activation is unprecedented for this family of enzymes [113].

5.3.1.3. Bd0602 and Bd0603 are membrane-linked proteins that may associate with Bd0601

In *B. bacteriovorus* HD100, the mRNA levels of the *bd0601-0603* operon are significantly increased during predation, suggesting that Bd0601 is required either for PG modification in *B. bacteriovorus* in order to predate, or for prey cell wall hydrolysis [21]. In any case, Bd0601 might be synthesised in an inhibited state that could be relieved by its interaction with a specific partner, targeting particular subcellular locations for PG hydrolysis.

The *bd0601-0603* operon encodes the soluble NlpC/P60 Bd0601 and two lipobox membrane-linked proteins, Bd0602 and Bd0603. Bd0602 is predicted to have at least one ankyrin-like repeat (residues 44 to 104) whereas Bd0603 has no significant putative domain prediction resulting from its amino acid sequence.

Since it was not possible to produce soluble constructs for Bd0602 nor Bd0603, predicting the three-dimensional structure of both proteins was attempted using RosettaFold and AlphaFold2 [145], [180]. Unfortunately, no model prediction for any of the two proteins reached at least 50% of confidence in either algorithm, yielding poor predicted structures even for the putative ankyrin repeat of Bd0602, with no discernible compact fold. Not even simultaneous folding and complex formation with Bd0601 or between Bd0602 and Bd0603 improved the models, jeopardising construct design or proposing a functional role for these proteins.

Since the ankyrin Bd3460 has been shown to interact and inhibit the DDTs Bd0816 and Bd3459, conferring protection of their hydrolytic activity to the predator until the enzyme is secreted to the prey's periplasm, the first working hypothesis was that Bd0602 could act as an immunity protein for Bd0601 [36].

B. bacteriovorus, specifically strain HD100, is the only known case of a (*Bdellovibrio*) DUF1460 encoded by an operon, wherein even the closely related *B. bacteriovorus* Tiberius strain homologue *bdt0574* has diverged as a single transcriptional unit. Bdt0574 lacks an equivalent for Y273 in Bd0601, replacing it with a serine residue as observed in most DUF1460s. In contrast, other organisms encode DUF1460 hydrolases in the same transcriptional unit as oligopeptide transporters and glycosyl hydrolases, suggesting that these enzymes are deployed as part of the PG recycling machinery [113], [184], [185].

Therefore, it is not unreasonable to hypothesise that Bd0602 and Bd0603 are related to Bd0601 function. Since both these proteins are covalently anchored to the membrane, one of the possibilities is that they localise and/or activate Bd0601 to particular subcellular localisations in the predator, as observed for other membrane-linked interacting partners of different CWAs [115], [167], [186]. Further functional and structural studies are required to probe this hypothesis.

5.3.2. Bd1177 and Bd1177 C86A/Y124C/Y508C

5.3.2.1. Bd1177 is expressed but crystallises only as a C-terminal fragment

Bd1177 was predicted to be a poorly conserved CWA or CWP of the CHAP or a related family. Wild-type Bd1177 was successfully expressed and purified to homogeneity for crystallisation screening and structure determination, in order to characterise this divergent group of CWHs. However, only a 24.5 kD C-terminal fragment of the protein (residues 322 to 544) crystallised in a variety of conditions and commercial crystal screening months after the crystallisation experiments started. Addition of protease inhibitors during purification and crystallisation, as well as decreasing temperature to 4°C and even seeding to accelerate the nucleation process did not allow crystallisation of the full length Bd1177.

The structure of the C-terminal fragment of Bd1177 was solved at 1.70 Å resolution showing that this fragment adopts a compact and mostly helical lightbulb-like structure stabilised by two disulphide bonds. No significant structural matches were found in the

PDB after performing a Dali search FoldSeek search of predicted structural models, suggesting that this fragment adopts a novel fold.

Considering the lack of information provided by the C-terminal fragment of Bd1177 and the failed attempts to crystallise the full-length protein, different truncated N-terminal constructs for Bd1177 were designed based on the start of the crystallised C-terminal fragment (residue 322). No soluble protein was expressed for the designed N-terminal Bd1177 constructs, indicating that the interaction of the N-terminal part of the protein with the crystallisable C-terminal fragment was likely necessary for protein folding.

Bd1177 was predicted with reasonable confidence depicting a bilobular protein fold, one of them corresponding to the experimentally-determined C-terminal fragment [145], [180]. In fact, the overall predicted structure of the C-terminal of Bd1177 mostly coincides with the experimental model, supporting the use of these models for construct design to obtain the full-length Bd1177.

Taking this into consideration, it was found that residues Y124 (of the N-terminal half) and Y508 (of the C-terminal half) of the RosettaFold *in silico* model were oriented forming an aromatic stacking interaction between N- and C-terminal halves of Bd1177. Therefore, a disulphide bond was engineered by introducing the point mutations Y124C and Y508C [181]. Additionally, residue C86 was predicted to be the putative catalytic cysteine of Bd1177. Thus, in order to improve protein stability and promote crystallisation of the full-length Bd1177, a third point mutation, C86A was designed. Two mutant versions of Bd1177 were then successfully expressed and purified, Bd1177 C86A and Bd1177 C86A/Y124C/Y508C. Of these two proteins, only the triple mutant Bd1177 C86A/Y124C/Y508C successfully crystallised and was subjected to X-ray diffraction and data collection.

5.3.2.2. Bd1177 C86A/Y124C/Y508C is a bilobular protein with a divergent NlpC/P60 fold and a novel capping domain

The 2.2 Å crystal structure of Bd1177 C86A/Y124C/Y508C was determined tracing the 524 residues (residue 21 to 544) of the mature protein in two of the four copies of the macromolecule in the ASU.

The general fold of the crystallographic model depicts a bilobular structure comprising the C-terminal crystallised fragment, which now clearly forms a discrete domain of the

protein, connected by an eight-residue linker to the N-terminal half of the protein. Interestingly, considering that the general fold of the protein is bilobular and that the C86A substitution was not enough for Bd1177 to crystallise, it is possible that the introduced C124-C508 disulphide bond stabilises the closed conformation of Bd1177 diminishing conformational fluctuation and thus promoting crystallisation.

The overall obtained structure is relatively similar to the predicted RosettaFold and AlphaFold2 model of Bd1177 (RMSD= 4.00 and 1.17 Å, respectively), which in addition to the hydrodynamic radius observed for the wild-type Bd1177 and both designed mutants suggests that no major structure alteration was conferred upon mutation.

The N-terminal half of Bd1177 forms a possible domain comprised by a NlpC/P60 papain-like fold that has undergone sequence diversification and acquired several structural insertions surrounding the catalytic triad C86-H200-D217. The predicted catalytic triad is located in the interface between the N-terminal and C-terminal domains of Bd1177, suggesting that the C-terminal domain participates in substrate selection by acting as a capping domain.

In contrast to what is observed for Bd0601, the Bd1177 structural insertions do not define a discrete substrate binding region surrounding the NlpC/P60 catalytic core, and therefore could regulate Bd1177 potential enzymatic activity by participating in the possible re-accommodation of the C-terminal capping domain, accommodating part of the substrate, or controlling subcellular localisation of the protein.

The C-terminal capping domain of Bd1177 adopts a novel lightbulb-like fold that interacts with the N-terminal NlpC/P60 through several residues. The most remarkable interactions correspond to the R314-D329 salt bridge near the hinge region where the linker connecting both domains is located, and the possible aromatic stacking Y124-Y508 that has been substituted in the crystallised protein for the C124-C508 disulphide bond.

5.3.2.3. Bd1177 C86A/Y124C/Y508C might undergo a substrate induced conformational change

The average B-factor values per residue of the solved Bd1177 structure interestingly suggest that the connecting linker conformation does not fluctuate in conformation. This could be due to the introduction of the C124-C508 disulphide bond stabilising the closed conformation of the enzyme. Since structural predictions depict a similar closed

conformation for the native Bd1177, it is proposed that the active site would be inherently occluded and would require a considerable conformation change to displace the C-terminal capping domain.

To identify possible conformational change mechanisms of NlpC/P60 proteins related to Bd1177 as well as to identify possible substrate binding residues, a Dali search was conducted but only yielded results with a relatively good conservation of the papain-like fold.

The first of the Dali-identified proteins, MagC has been determined to be a non-hydrolytic NlpC/P60 component of the α-2-macroglobulin (A2M) Mag complex of *P. aeruginosa* PAO1, which links the Mag complex with the PG cell wall [182], [187], [188]. The MagC active site is catalytically disengaged with the acid-base catalytic residues H153 and H165 oriented away from the nucleophilic C60, a feature that is not shared by the solved Bd1177 triple mutant structure. The hydrogen bonding distance between H200 and D217 indicates that at least these two residues are engaged in the archetypical proton relay mechanism to deprotonate the nucleophilic C86, which is absent in the catalytically null determined structure. Thus, Bd1177 might be active, however, further experiments are necessary to assess this hypothesis.

The second Dali related structure, COG0791 from *N. punctiforme*, depicts a SH3b N-terminal domain in which the catalytic triad and most NlpC/P60 conserved residues are clustered in the interface between the SH3b and the NlpC/P60 domain, suggesting that the accessory SH3b domain participates in substrate selection [159].

On the other hand, the third significant Dali match, Acd24020 from *C. difficile*, has been shown to be a PG endopeptidase [151]. It is proposed that the conserved structure of the hourglass-like substrate binding groove accommodates 3-4 crosslinked PG molecules for their hydrolysis.

Unlike the abovementioned NlpC/P60s, the Bd1177 active site is inaccessible due to the C-terminal domain interacting and blocking access to the putative NlpC/P60 catalytic domain. The C-terminal domain residues R384 and R387 are spatially the closest to the catalytic triad, interacting with T223 and the catalytic H200, respectively. Moreover, residues Y124 and Y508 are likely in proximity as observed in the RosettaFold model and in the formation of a disulphide bond between the cysteine residues introduced in

positions 124 and 508. Further interactions such as the R314-D329 salt bridge are observed in the structure, closer to the hinge region defined by the linker connecting the N-terminal and C-terminal domains of Bd1177.

The similarity between the *in-silico* models and the solved structure of the Bd1177 triple mutant suggests that Bd1177 is naturally in a closed state. It is then feasible that the C-terminal domain is displaced upon substrate recognition, wherein at least part of the residues mediating the interaction between the N- and the C-terminal domain are re-oriented for substrate binding.

In a similar way, it is possible that the interaction of Bd1177 with a cognate partner either in *B. bacteriovorus* or in the prey shifts the position of the C-terminal domain, releasing the interface interacting residues enabling substrate binding and catalysis. This is further suggested by the open Bd1177 conformation calculated using the ENM approximation, showing that the N-terminal/C-terminal domain interface of Bd1177 could potentially fluctuate to be solvent accessible.

Considering the structural divergence of Bd1177 with respect to the currently characterised NlpC/P60s, further experiments are required not only to assess the conformational change mechanism of Bd1177, but for identifying the reaction catalysed by this predicted enzyme. A first approach may involve incubating generic PG sacculi with Bd1177, determining if the presence of partner is needed for enzymatic activity, and consequently identifying hydrolysis products by mass-spectrometry.

5.4. Conclusions

The structure of two *B. bacteriovorus'* NlpC/P60 putative enzymes, Bd0601 and Bd1177, has been determined, defining two separate families of possible CWHs.

Bd0601 C94A was crystallised in its apo state and in presence of GlcNAc, with no significant conformational changes driven by ligand binding. The Bd0601 structure shows that this putative enzyme is part of the DUF1460 family of NlpC/P60s, which have been determined to have amidohydrolase activity, catalysing the release of PG stem peptides from the glycan portion of the molecule.

Structural data shows that even though the general architecture of a priming and non-priming pocket of the DUF1460s is retained in Bd0601, ligand binding residues, specifically

on the stem peptide binding pocket, are poorly conserved, and thus are not identifiable by protein sequence analysis alone. This suggests either a higher tolerance in varying dimensions and composition of the stem peptide binding groove with respect to the glycan binding pocket of the enzyme, or adaptation of each DUF1460 to different PG molecules.

Moreover, Bd0601 C94A was shown to place a bulky aromatic residue, Y273, blocking access to the stem peptide binding groove. This feature is only present in the unclassified *Bdellovibrio sp. ArHS* and the archaea *L. bacterium* RIFCSPHIGHO2 orthologs. The inability to dock model PG substrates in the Bd0601 structure suggests that reorienting of the Y273 sidechain is required. This could be achieved either by substrate recognition itself or by association of Bd0601 with a binding partner. Further experiments are required to test these hypotheses, especially since the two other components of the *bd0601-0603* operon, Bd0602 and Bd0603, were not expressed in soluble form and no satisfactory *in silico* model was obtained for these proteins.

On the other hand, Bd1177 was crystallised as a catalytically inactive and disulphide locked triple mutant (Bd1177 C86A/Y124C/Y508C). This protein structure depicts a novel bilobular fold consisting of a highly divergent NlpC/P60 N-terminal domain and a cryptic C-terminal domain acting as a lid, blocking access to the predicted ligand binding groove. Both domains are connected by a nine long residue unstructured linker.

The N-terminal domain of Bd1177 consists of a NlpC/P60 papain-like core that has acquired multiple structural insertions. Unlike Bd0601, these insertions do not define a discrete substrate binding pocket.

The C-terminal domain adopts a lightbulb-like structure that is unique for Bd1177 homologues, interacting with the N-terminal domain via several hydrogen bonding and cation-pi stacking interactions.

The unique fold of Bd1177 and its poor sequence conservation suggests that this protein defines a novel family of putative CWHs different from Bd0601 or other previously characterised NlpC/P60 proteins. Further experiments are required to determine the reaction catalysed by this putative enzyme and how the C-terminal domain participates in substrate selection.

CHAPTER 6

6. STRUCTURAL CHARACTERISATION OF *B. bacteriovorus* CRYPTIC SECRETED PROTEIN Bd0675

6.1. Introduction

Structure-based assignment of molecular functions for hypothetical proteins has been used successfully in cases when putative prediction of conserved domains and/or sequence similarity analysis is ineffective.

MJ0577 from *Methanococcus jannaschii* (PDB ID: 1MJH), for instance, has been proven to be an ATP-binding protein depicting nucleotide binding motifs that are conserved in structure but not in amino acid sequence [189]. Furthermore, in some cases, structure determination has defined novel domain families, as shown for the RNA binding protein TTHB192 from *Thermus thermophilus* HB8 (PDB ID: 1J9) [190].

Besides being significantly upregulated during predation, Bd0675 from *B. bacteriovorus* HD100 does not (obviously) encode a conserved predicted domain, jeopardising the understanding of its role in the predatory life cycle [21]. Results by Lambert *et al*. (2011). show that *bd0675* is the 36th most upregulated gene in predation, which is comparable to genes with predicted functions such as gliding motility (*bd01416* and *bd0419*) and PG-related genes (*bd0886* and *bd1358*).

Here we have determined the crystallographic structure of Bd0675 and propose a possible ligand-binding function of this protein. Ligand identification attempts based on structural and sequence conservation are further discussed in this chapter. Additionally, considering the novel topology of this protein, Bd0675 was used as target to assess the performance of prediction algorithms participating in the Critical Assessment of Techniques for Protein Structure Prediction 14 (CASP14) experiment, and analysed as a target highlight in a corresponding publication [191].

6.2. Results

6.2.1. Bd0675 purification and crystallisation

Bd0675 was cloned fused to an N-terminal or C-terminal His_6-tag in either pET28a(+) or pET41c(+), respectively, as indicated in Materials and Methods. However, both tagged versions of His-tagged Bd0675 proved to be toxic for the cells when overexpressed in native state either in AI or TB media. Specifically, after initiating the induction of Bd0675, the cell cultures stop growing and in consequence will not produce the recombinant protein.

Therefore, N-terminal His-tagged Bd0675 (here onwards Bd0675) was expressed in inclusion bodies, purified, and refolded on-column during the IMAC purification step as indicated in Materials and Methods. The IMAC eluted pooled fractions were concentrated and loaded onto a Superdex® 75 26/600 size exclusion column connected to an AKTA™ Pure system for separating the refolded His-Bd0675 from possible soluble aggregates and other contaminants. The column was equilibrated and run with a buffer composed of 25 mM HEPES pH 7.5, 200 mM NaCl and 2.5% glycerol.

As shown in **Figure 101**, Bd0675 refolds with an efficiency of more than 80%. Furthermore, Bd0675 elutes with a hydrodynamic radius corresponding to a 12-15 kD spherical protein, suggesting that it is monomer in solution. It has also been noted that Bd0675 reversibly precipitates at 4°C and thus is best stored at room temperature.

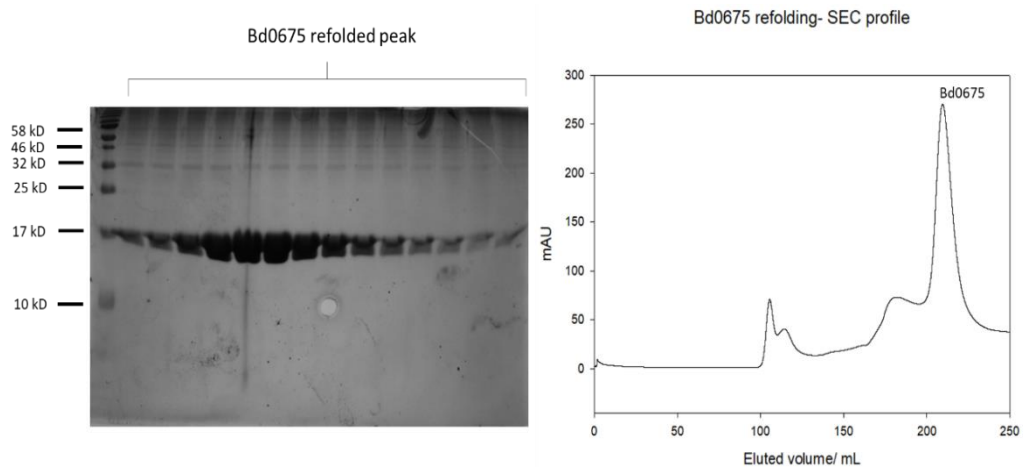


Figure 101: Purification and refolding of Bd0675. Bd0675 was expressed in inclusion bodies, purified, and refolded on the Ni-NTA column as indicated. The eluted protein was then loaded on a Superdex® 75 26/600 column to isolate the refolded protein from remaining contaminants and Bd0675 soluble aggregates. The purity of the fractions was assessed by SDS-PAGE, pooled, and concentrated by ultrafiltration at room temperature. Bd0675 elutes with a hydrodynamic radius corresponding to a 12-15 kD monomeric protein.

SEC fractions containing the purified and refolded Bd0675 were pooled, concentrated by ultrafiltration, and screened for crystallisation conditions as indicated in Materials and Methods. Microcrystals were obtained at 35 mg/mL in Proplex® H9 (0.1 M Imidazole pH 7.0 and 50% (±)-2-Methyl-2,4-pentanediol or MPD), JBScreen® Classic 7 condition D6 (50% MPD), JBScreen® Classic 8 conditions A1 (50% MPD, 15% ethanol and 10 mM sodium acetate) and A4 (55% MPD).

The obtained crystalline needles were optimised by MMS screening as detailed in Materials and Methods by using crystals obtained in JBScreen Classic 7 D6 as seed (**Figure 102A**), re-screening Bd0675 at 35 mg/mL in the MPD based screens JBScreen Classic 7 and 8 using both 1:10 and 1:100 seed dilutions. Due to the high redundancy of the MPD based screens, crystals were obtained for any condition with MPD concentrations of over 30% in either 1:10 or 1:100 seed dilutions. Representative optimised crystal morphologies are presented in **Figure 102B** and **C**. Further, since crystals were obtained in 30% or more MPD, no cryoprotectant supplementation was needed before flash-cooling and data collection.

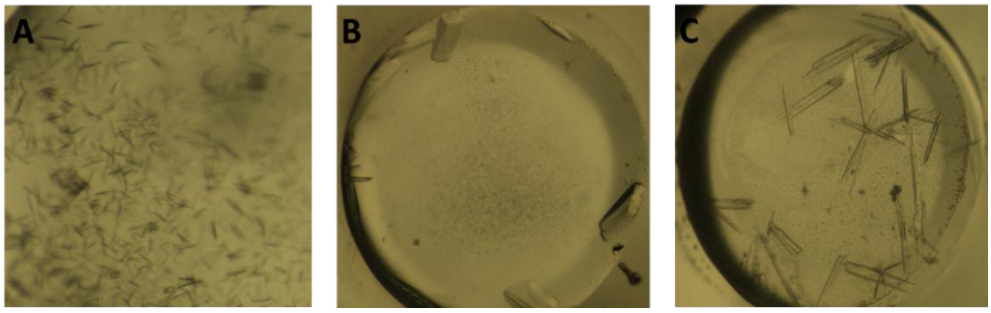


Figure 102: Crystals of Bd0675. **A)** Bd0675 crystals at 35 mg/mL and 18°C in JBScreen Classic 7 condition D6 (50% MPD) before MMS screening optimisation. **B)** MMS optimised crystals of Bd0675 JBScreen Classic 7 D5 (47% MPD and 2% tert-butanol) using crystals in **panel A** as seed in a 1:10 dilution. **C)** MMS optimised crystals of Bd0675 in JBScreen Classic 8 condition A1 (50% MPD, 15% ethanol and 10 mM NaOAc), using crystals in **panel A** as seed in a 1:100 dilution.

Bd0675 protein sequence was compared with structures deposited and available in the PDB, not finding any significant match. Thus, an experimental phasing strategy was adopted by soaking part of the Bd0675 optimised crystals in mother liquor supplemented with 10 mM K_2PtCl_4 for one hour at 18°C prior to crystal collection. Non-derivatised Bd0675 datasets were also obtained in order to solve the native structure of the protein.

6.2.2. Bd0675 structure determination and overall fold

The diffraction experiment was carried out at the IO4 beamline as indicated in section 2.1.9, collecting both platinum- phased and native Bd0675 datasets (0.9100 Å wavelength collection for both datasets) for optimised crystals in condition D5 of the JBScreen Classic 7 using a 1:10 seed dilution of the Bd0675 microcrystals obtained for condition D6 of JBScreen Classic 7.

Diffraction data was autoprocessed by the Xias2-dials or Xias2-3dii automated pipelines and used for structure solving using programmes Autosol and Phaser-MR of the Phenix suite, for platinum phased and native datasets, respectively. Statistical results for data collection, phasing and refinement are shown in **Table 17**.

Table 17: Data collection, phasing, and refinement statistics for Bd0675. Statistics for PtSAD and native Bd0675 crystallographic data are shown.

	Bd0675 PtSAD	Bd0675 Native
Data Collection		
Resolution range (Å)	42.11 - 1.83	42.43 - 1.50
Space group	C 2 2 2_1	C 2 2 2_1
Cell dimensions		
α, b, c (Å)	53.41, 68.44, 86.90	53.49, 69.67, 85.85
α, β, γ (°)	90, 90, 90	90, 90, 90
Total reflections	168267 (12375)	337475 (33375)
Unique reflections	14370 (1416)	25800 (2533)
Completeness (%)	99.78 (99.09)	98.97 (98.25)
Redundancy	11.7 (8.7)	13.1 (13.2)
<I>/σI	21.32 (1.22)	24.27 (1.44)
R_{meas}	0.0581 (1.621)	0.04357 (1.962)
$CC_{1/2}$	0.999 (0.491)	0.999 (0.638)
CC_{anom}	0.6 (0.1)	
Refinement		
Maximum resolution (Å)	1.83	1.50
Number of reflections used	14369 (1416)	25800 (2533)
Number of reflections for R_{free}	704 (60)	1260 (125)
R_{work}	0.2324	0.1780
R_{free}	0.2895	0.1943
RMSD		
Bonds (Å)	0.007	0.012
Angles (°)	1.28	1.59
Number of atoms		
Protein	1040	1043
Ligand/ion	1	34
Water	46	73
B-factors		
Protein	57.27	52.95
Ligand/ion	63.27	63.53
Water	53.52	48.07
Ramachandran favoured (%)	96.18	95.42
Ramachandran allowed (%)	3.05	4.58
Ramachandran outliers (%)	0.76	0.00
Rotamer outliers (%)	0.00	0.00

*Values in parenthesis correspond to the highest-resolution shell

The Bd0675 structure was initially solved from the PtSAD dataset experiment at a 1.83 Å resolution in the $C222_1$ spacegroup, finding one protein molecule in the ASU. The PtSAD solved and refined model was used for solving the native dataset.

The 1.50 Å structure of Bd0675 has been solved in space group $C222_1$, finding a single copy of the protein in the asymmetric unit. The determined structural model of Bd0675 is depicted in **Figure 103A** and spans from amino acid 2 to 134 (133 residues of 134 total).

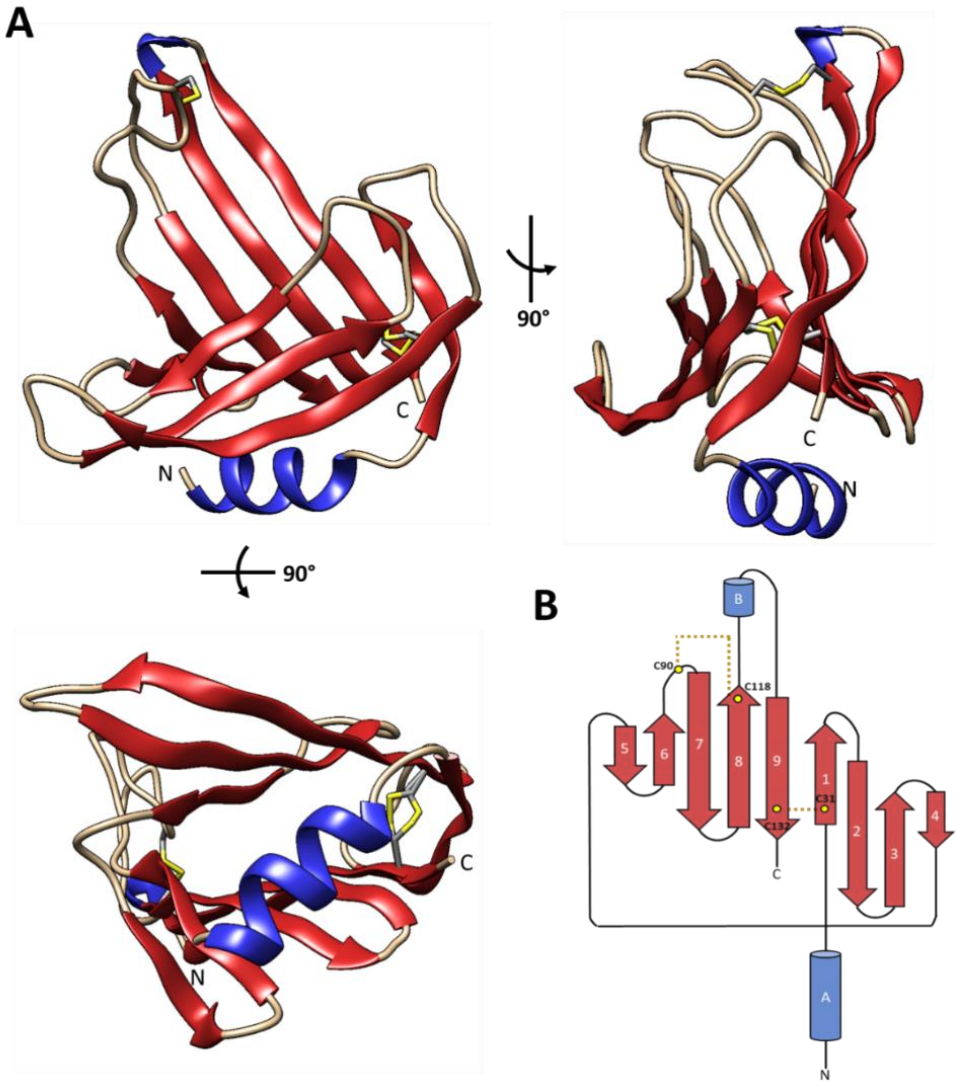


Figure 103: Bd0675 structural model and topological organisation. A) Ribbon diagram of Bd0675 structure, wherein β-strands are coloured in red and α-helices in blue, with disulphide bonds highlighted in yellow. 90° angles apart views of Bd0675 are depicted. **B)** Topology diagram of Bd0675. α-helices and β-strands are listed A and B, and 1-9, respectively, from N- to C-terminus of the protein. The C31-C132 disulphide bond is formed between β-strands 1 and 9, whereas the C90-C118 bond is formed between the β-strands 6 and 7 interloop and β-strand 8.

Bd0675 has a β-roll-like distorted architecture containing two α-helices (**Figure 103B**, helices A and B, in blue) and nine β-strands (**Figure 103B**, strands 1 to 9, in red). α-helix A is part of the N-terminal region of the protein and interacts directly with the β-roll,

whereas α-helix B is part of the linker element between β-strands 8 and 9. The overall β-roll fold part of the structure is formed by two β-leaflets, one comprised by β-strands 1 to 4 (β-leaflet 1), which is connected to a second leaflet comprising β-strands 5 to 9 (β-leaflet 2).

6.2.3. Bd0675 hydrophobic core, disulphide bonding and coulombic surface potential

The Bd0675 three component architecture is stabilised by hydrophobic interactions between both β-leaflets and the amphipathic α-helix A, which closes the hydrophobic core.

The hydrophobic core of the protein is formed by residues F5, V6, L9, K10 and L12 of α-helix A; I19 of β-strand 1; L33 and L35 of β-strand 2; L47, V49 and L51 of β-strand 3; T54, M56 and P60 of connecting loop between strands 3 and 4; A62 and L64 of β-strand 4; T68 of connecting loop between strands 4 and 5; L70 of β-strand 5; A76 and V78 of β-strand 6; L103 and L105 of β-strand 7; V110 and I112 of β-strand 8; and T130 of β-strand 9.

Residue K10 interacts with the core via the alkyl chain of the sidechain while exposing the ε-amine group to the solvent. Moreover, the sulphur group of M56 is stabilised by interacting with the hydroxyl group of both T56 and T130, residues that are further stabilised by interaction of the methyl group of the sidechain with I19. These interactions are depicted in **Figure 104**. Nevertheless, is noticeable that this hydrophobic core is partially exposed at the region delimited by the N-terminal of the protein and the loop connecting β-strands 4 and 5.

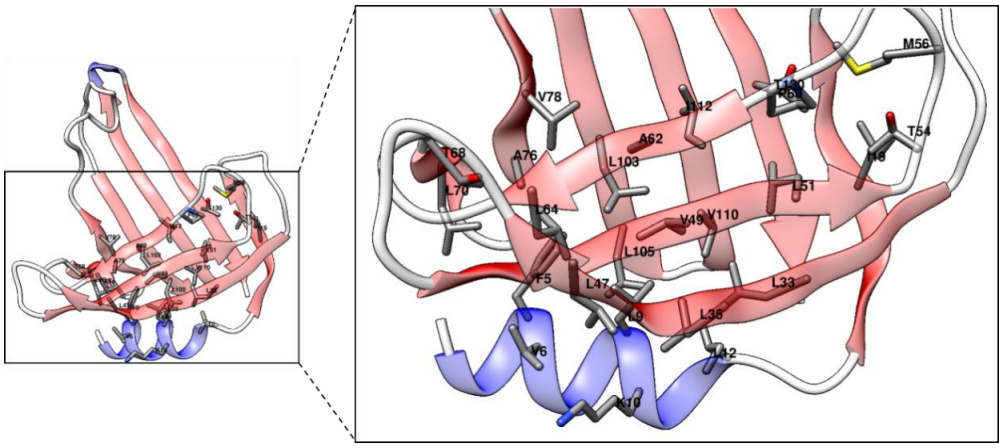


Figure 104: Hydrophobic core of Bd0675. The overall three component fold of Bd0675 is stabilised by hydrophobic interactions between both β-leaflets and the amphipathic α-helix. The location and residues forming part of the hydrophobic core are highlighted and labelled.

Further, the Bd0675 three component fold is held together by disulphide bonding between highly conserved cysteine residues. The disulphide bond C31-C132 appears to adopt two possible conformations and links β-strands 1 and 9, stabilising the β-leaflet interaction of the structure. On the other hand, the disulphide bond C90-C118 links the 19-residue loop between β-strands 6 and 7 with β-strand 8, suggesting that correct positioning of this loop is relevant for Bd0675 function.

Inspection of the coulombic surface potential shows that Bd0675 possesses a hand-like shape with a prominent cleft in the middle of the protein that could potentially act as a ligand binding region (**Figure 105A**). The groove is formed by the outward facing sidechains of residues of β-leaflet 1 and the non-buried inward facing β-leaflet 2 residues and flanked by the β-strand connecting loops (**Figure 105B**).

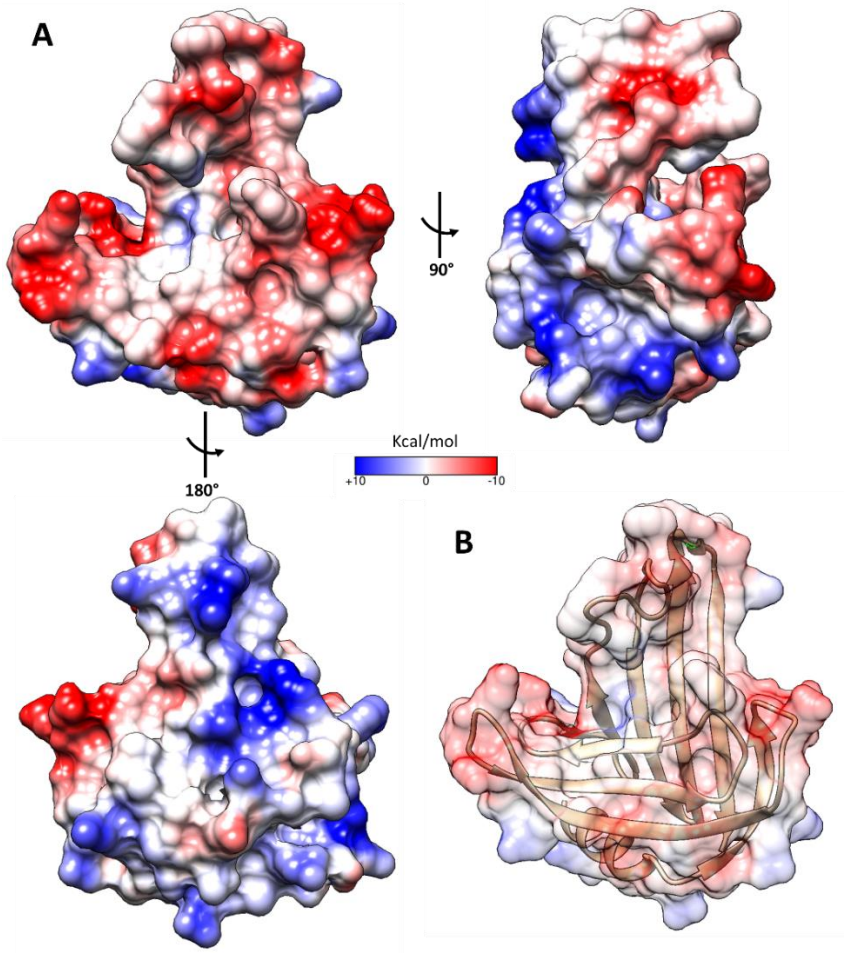


Figure 105: Electrostatic surface potential of Bd0675. A) The surface structure of Bd0675 was analysed by coulombic charge distribution in a scale of +10 to -10 Kcal/mol and displayed in 90° and 180° dispositions, showing a hand shape architecture with a negatively delimited cleft. **B)** The surface of Bd0675 shows an extensive cleft delimited by the negatively charged connecting loops between β-strands 1 and 2, 2 and 3, 4 and 5, and 6 and 7, which could act as a ligand binding site.

6.2.4. Bd0675 proposed ligand binding cleft definition and function elucidation

The proposed binding cleft of Bd0675 comprises the negatively charged connecting loops between β-strands 1 and 2, 2 and 3, 4 and 5, and 6 and 7, wherein the latter is fixed in position by the C90-C118 disulphide bond. The cleft further consists of residues from β-strands 4 and 6-8, creating a less negatively charged central region delimited by the negatively charged connecting loops.

To define the residues forming part of the proposed binding cleft, amino acid conservation of Bd0675 homologues in predator species *B. bacteriovorus* strains HD100 and Tiberius, *Halobacteriovorax spp* and *Bacteriovorax spp* were analysed (see **Figure 106**).

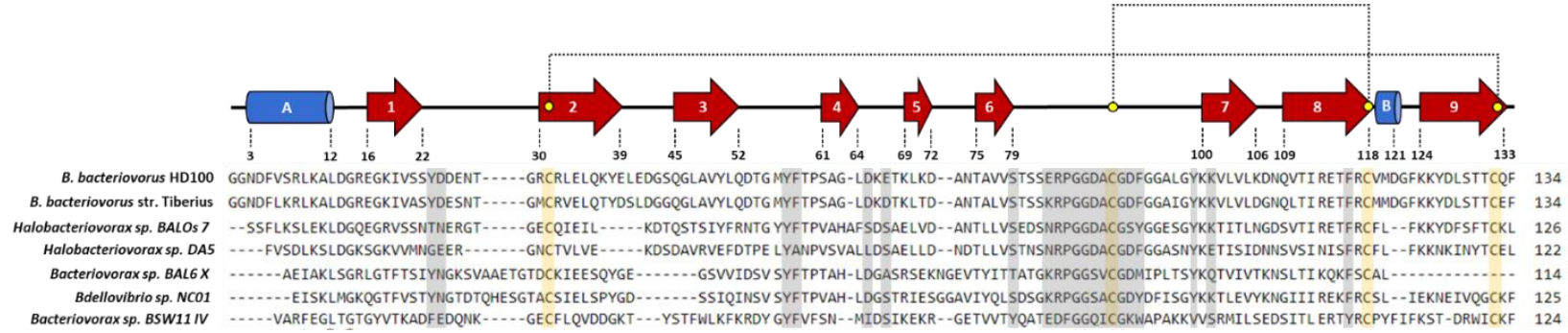


Figure 106: Amino acid conservation of Bd0675 across predators. Protein sequences of Bd0675 homologues in *B. bacteriovorus* strains HD100 and Tiberius, *Halobacteriovorax spp* and *Bacteriovorax spp* were aligned using Clustal Ω and compared in the context of Bd0675 solved structure (secondary structure elements are numbered in reference to Bd0675 sequence). Conserved cysteine residues are highlighted in yellow, being only absent in one of the *Bacterivorax* species probably for lack of coverage of the full sequence. The conserved residues situated in the proposed Bd0675 binding cleft are highlighted in grey, wherein most of the conservation is localised in the loop between β-strands 6 and 7, which is fixed in position in the structure by the disulphide bond C90-C118.

Besides residues forming part of the hydrophobic core of the protein and the four disulphide forming cysteines (in yellow), with the exception of C90, most conserved residues of Bd0675 are located in the proposed binding cleft of the protein.

Specifically, residues Y57, F58, G86, G87, C90, G91, Y99, F116 and C118 are mostly invariant in every Bd0675 related protein, while residues Y23, D24, D65, D67, S79, E83, E84, E85, D88, A89, D92, F93 and K101 are conserved in polarity or aromaticity in at least five of the seven compared homologues. Furthermore, most of the conserved residues are located in the disulphide bonded loop connecting β-strands 6 and 7, forming part of a (R/D)PGGXϕCGXΩX_5Y motif, wherein X is any amino acid, ϕ is a hydrophobic residue and Ω is an aromatic residue.

These conserved amino acids are depicted in context of the coulombic surface potential of Bd0675 (**Figure 107**). Residues Y23 and D24 of loop connecting β-strands 1 and 2, D65, E67 of loop connecting β-strands 4 and 5, Y57 and F58 loop between β-strands 3 and 4, and residues 83 to 93, which are part of β-strands 6 and 7 connecting loop, form lobes limiting the centre of the proposed ligand binding site of Bd0675. The limiting lobes are mostly composed by negatively charged sidechains and aromatic chains whereas, residues S79, Y99 and particularly K101 are located towards the centre of the cleft.

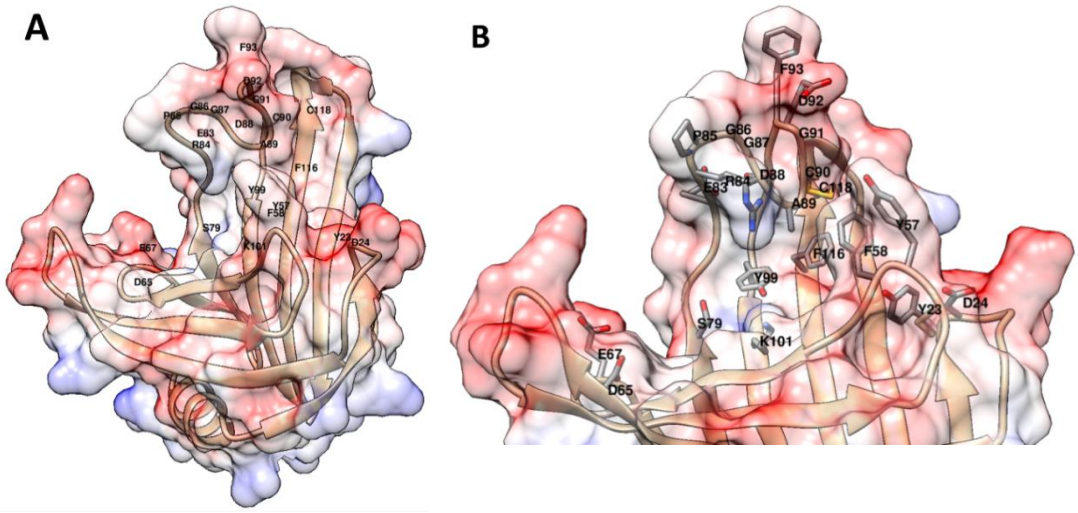


Figure 107: Conserved residues in Bd0675 proposed ligand binding cleft. A) Distribution of conserved residues at the hypothesised binding cleft of Bd0675. Majority of conserved residues localised to the top lobe formed by the loop connecting β-strands 6 and 7. **B)** Sidechain chemistry and orientation of conserved residues at the Bd0675 hypothetical binding cleft. The cleft is defined by a central region and three negatively charged limiting lobes where most of conserved residues localise. The surface coulombic potential is coded from red (-10 Kcal/mol) to blue (+10 Kcal/mol).

Conserved residues of the (R/D)PGGXϕCGXΩX_5Y motif in the disulphide stabilised loop interact with residues in β-strands 7 and 8. For instance, the sidechain of the R84 of the motif directly delimits the putative binding grove while the ϕ residue, corresponding to A89 in Bd0675, interacts with the aromatic ring of the highly conserved F116, positioning the aromatic group towards the putative binding cleft and thus is potentially a direct participant of ligand binding. In contrast, the highly conserved Y99 of the motif is at hydrogen bonding distance to the ε-amino group of the K101 positioned towards the centre of the cleft. It is worth noting that the aforementioned Y99 and K101 residues are substituted only in the *Bacterivorax sp.* BSW11 IV sequence for valine and serine residues which could potentially replace the hydrogen bonding interaction with a hydrophobic interaction.

Further, besides L12 which forms part of the hydrophobic core of Bd0675 and G14 that is located in the loop connecting α-helix A with β-strand 1, and the four cysteines, the only invariable residues of Bd0675 are Y57, G86, G87 and G91. Glycine residues are located in the disulphide stabilised loop probably for it to acquire the correct conformation, delimiting the centre of the binding cleft or for positioning of residues of the conserved loop for ligand binding.

On the other hand, Y57 localises towards one of the delimiting loops of putative groove, forming an acidic pocket with Y23, Y53, F58 and L128 (**Figure 108**). All Y57 interacting residues are conserved in every homologue at least in polarity or aromaticity, suggesting that this dielectric pocket could be critical in ligand binding and thus in Bd0675 function.

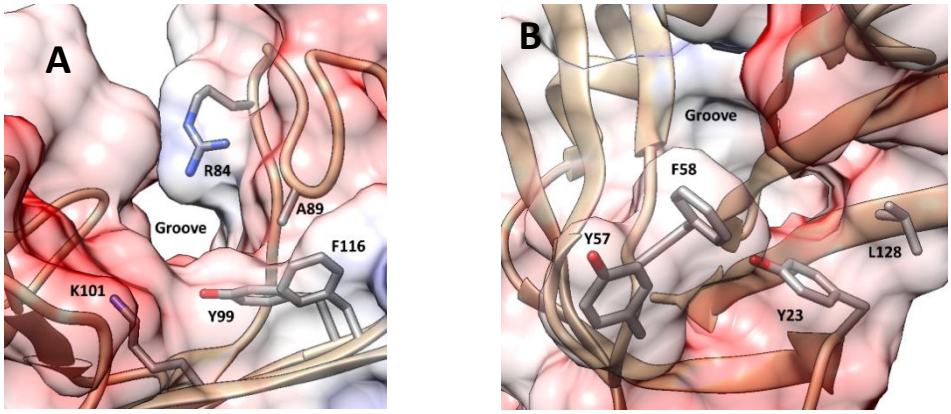


Figure 108: Conserved binding pockets of Bd0675. A) Binding pocket defined by the highly conserved disulphide stabilised loop, located towards the centre of the Bd0675 cleft. Residues R84, Y99, K101 and F116 sidechains project towards the cleft while A189 stabilised Y99 and F116 conformations. **B)** Aromatic interacting pocket conformed by the invariable residue Y57 along with highly conserved sidechains Y23, F58 and L128. Coulombic surface potential is scaled from red (-10 Kcal/mol) to blue (+10 Kcal/mol).

The refined native structure of Bd0675 has components of the crystal screen associated with the protein. The electron density shows four tert-butanols, one MPD and one glycerol molecules, whose interactions with Bd0675 structure are depicted in **Figure 109A**.

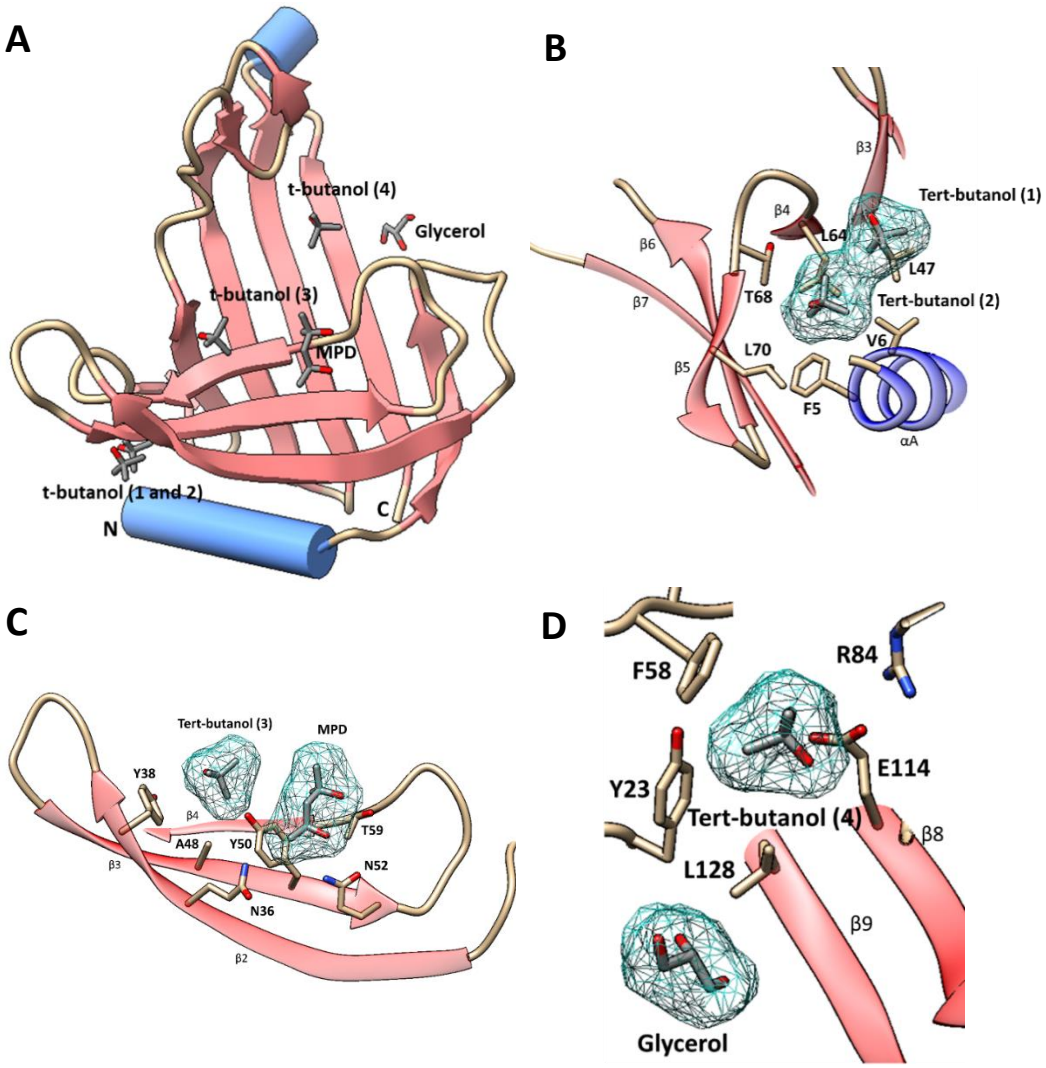


Figure 109: Crystallisation ligands interactions with Bd0675. A) Four tert-butanols, one glycerol and one MPD molecules interact with Bd0675 in the crystal structure. Density corresponds to $2F_0$-F_0 map at 1.5 σ. **B)** Residues of the partially exposed hydrophobic core of Bd0675 interact with two tert-butanol molecules. **C)** An MPD and a tert-butanol molecule interact with residues on the surface of Bd0675. This interaction is located away from the hypothesised ligand binding cleft and show no significant conservation with Bd0675 homologues in other predatory bacteria. **D)** Tert-butanol and glycerol molecules interact with highly conserved residues located in the putative binding cleft of Bd0675.

As illustrated in **Figure 109B**, two tert-butanol molecules interact with the partially exposed hydrophobic core of Bd0675, specifically, with residues F5, V6, L47, L64, T68 and L70, which are located in the interface between α-helix A and β-leaflets 1 and 2. Further, as depicted in **Figure 109C**, a third tert-butanol and an MPD molecule are found interacting with the solvent exposed residues N36, Y38, A48, Y50, N52 and T59, which are part of the β-leaflet 1, however these residues are not conserved in Bd0675 homologues and are not oriented towards the hypothesised ligand binding cleft of the protein.

Nevertheless, a glycerol and a fourth tert-butanol interact with conserved residues located at the cleft of Bd0675 (**Figure 109D**). Particularly, the tert-butanol molecule interacts with Y23, F58, R84 and E114, while the glycerol molecule is stabilised by hydrophobic interactions with L128. Residues interacting with this tert-butanol molecule are highly conserved with exception of E114, and part of the putative binding pocket formed by β-strands 8 and 9 and connecting loops between strands 1 and 2, 3 and 4, and 6 and 7.

Additionally, the PtSAD solved structure shows that the platinum ion binds between the crystal contacts of two Bd0675 symmetry copies in the crystal lattice (**Figure 110A**).

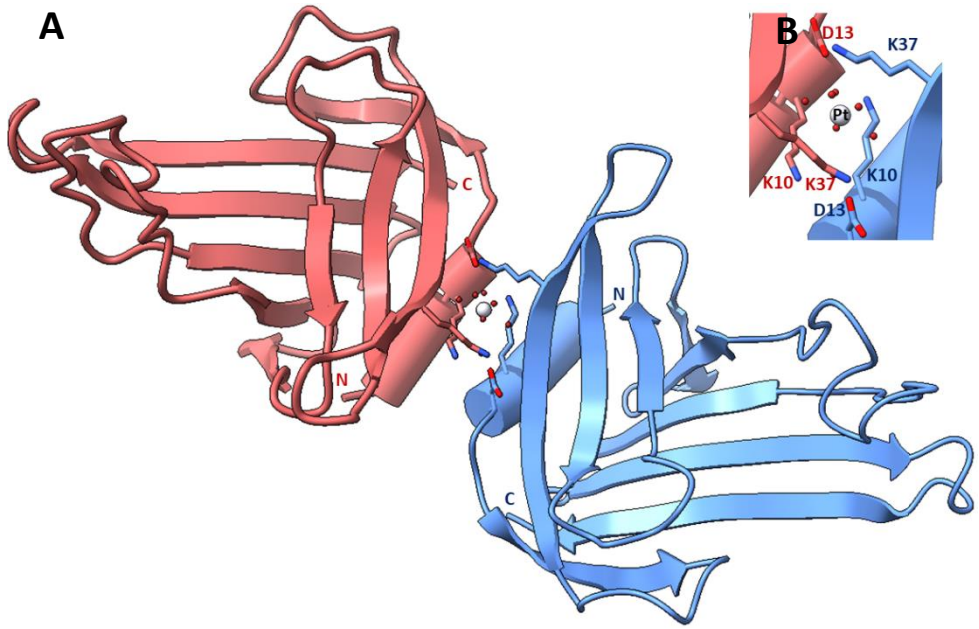


Figure 110: Bd0675 platinum interacting site. A) Overview of the Bd0675 platinum (white sphere) interacting site located in the interface of two Bd0675 crystal symmetry copies (pink and blue, respectively). **B)** The platinum interacting interface between the two copies of Bd0675 is defined by residues K10, D13 and K37 which are part of alpha helix α1 and beta strand β2.

The two-fold symmetry axis contacts between residues K10 and D13 of alpha helix α1, and K37 from beta strand β2 generate a solvent pocket that is occupied by the platinum ion in the solved structure. The poor conservation of these residues along with their location in the structure suggests that this platinum binding pocket is particular to the crystallisation and experiment and do not have a direct role in Bd0675 physiological function.

Altogether, these results along with the conservation analysis of Bd0675 is inconclusive in proposing a ligand for Bd0675 due to the complex nature of the proposed binding cleft

and the limited information related to Bd0675 homologues. Therefore, proteins with similar topologies were searched for elucidating Bd0675 function.

6.2.5. Bd0675 structural conservation

The structure of Bd0675 was submitted to the Dali server to find related structures deposited in the PDB. The top 25 matches from Dali are shown in **Table 18**.

Table 18: Dali results for Bd0675 structural comparison. Dali matches for PDB deposited structures are organised according to the Z-score.

№	PDB ID-Chain	Z score	RMSD	lali	residues	% identity	Description
1	1l3a-C	6.2	4.7	92	182	9	P24: Plant Transcriptional Regulator PBF-2
2	6nnx-A	6.1	3.2	92	137	12	Cellular Retinoic Acid-Binding Protein 2
3	3elz-B	5.9	3.6	88	135	9	Ileal Bile Acid-Binding Protein
4	3ia8-A	5.8	3.8	104	162	6	THAP Domain-containing Protein 4
5	2rcq-A	5.8	3.4	93	141	3	Retinol-Binding Protein II, cellular
6	6ljv-A	5.7	3.5	92	136	9	Fatty Acid-Binding Protein, adipocyte
7	6xu5-A	5.7	3.3	86	133	9	Myelin P2 Protein
8	2fwv-A	5.5	3.9	107	190	7	Hypothetical Protein MTUBF_01000852
9	5l8i-B	5.4	3.5	82	126	9	Gastrotropin
10	3os7-A	5.4	4.3	78	338	6	Galactose Mutarotase-like Protein
11	6ofr-A	5.4	4.2	68	758	3	TONB-dependent Outer Membrane Receptor
12	5m9b-A	5.3	3.5	72	707	3	Ferric Enterobactin Receptor
13	6qvi-A	5.2	3.7	69	518	13	COMZ
14	5vzt-B	5.2	3.1	83	418	8	S-phase Kinase-associated protein 1
15	4rdr-A	5.1	4.7	87	706	2	ZNUD
16	3kvn-X	5.1	3.9	87	628	5	Esterase ESTA
17	5nup-A	5.1	4.1	84	344	8	OMPK36
18	4tq2-A	5.0	3.5	98	177	8	Putative Phycoerythrin Lyase
19	2mo5-A	5.0	3.9	87	142	10	Fatty Acid-Binding Protein, intestinal
20	5h3z-A	4.9	3.2	76	1113	9	Uncharacterized Protein
21	5yvf-A	4.8	3.6	101	322	7	BFA1
22	1k1x-A	4.8	3.5	79	636	8	4-Alpha-Glucanotransferase
23	5dl5-A	4.8	3.8	66	416	11	Membrane Protein
24	4pr7-A	4.7	3.5	64	171	5	Oligogalacturonate-specific Porin KDGM
25	5ur9-F	4.7	3.2	85	138	13	Fatty Acid-Binding Protein, epidermal

In general, the Bd0675 sequence identity to Dali matches is poor, which is reflected in high overall values of RMSD and modest Z-scores. As it is observed in **Figure 111**, the sequence and structural conservation relative to the top 25 Dali matches of Bd0675, is concentrated in β-leaflet 2, with exception of the loops, particularly the disulphide bond stabilised loop spanning from residues 80-99 which is structurally pivotal for Bd0675 central cleft architecture.

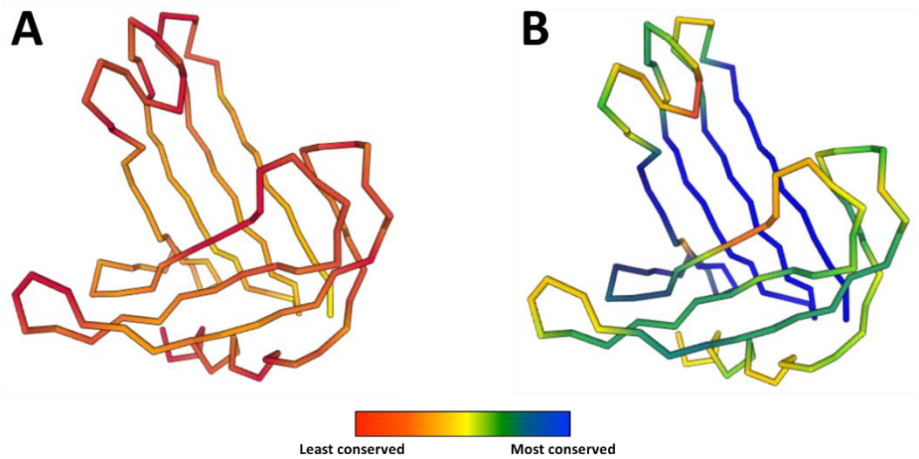


Figure 111: Bd0675 structural and sequence conservation against the top 25 Dali matches of the PDB. A) Sequence conservation of Bd0675 with the top 25 Dali matches shows poor similarity, which is mainly concentrated in the β-leaflet 2. No conservation of Bd0675 disulphide loop is found. **B)** Structure conservation of Bd0675 respect to the top 25 Dali matches. Highest structural conservation is concentrated in β-leaflet 2, however overall fold conservation is low.

As shown in **Table 18**, most of the matches are proteins with lipid binding functions and β-barrel like folds , with exception of the highest Z-score match PBF2, which is a β-roll transcription factor of the Whirly family found commonly in chloroplast and mitochondria [192]. The structural alignment with both type of topologies is represented in **Figure 112**.

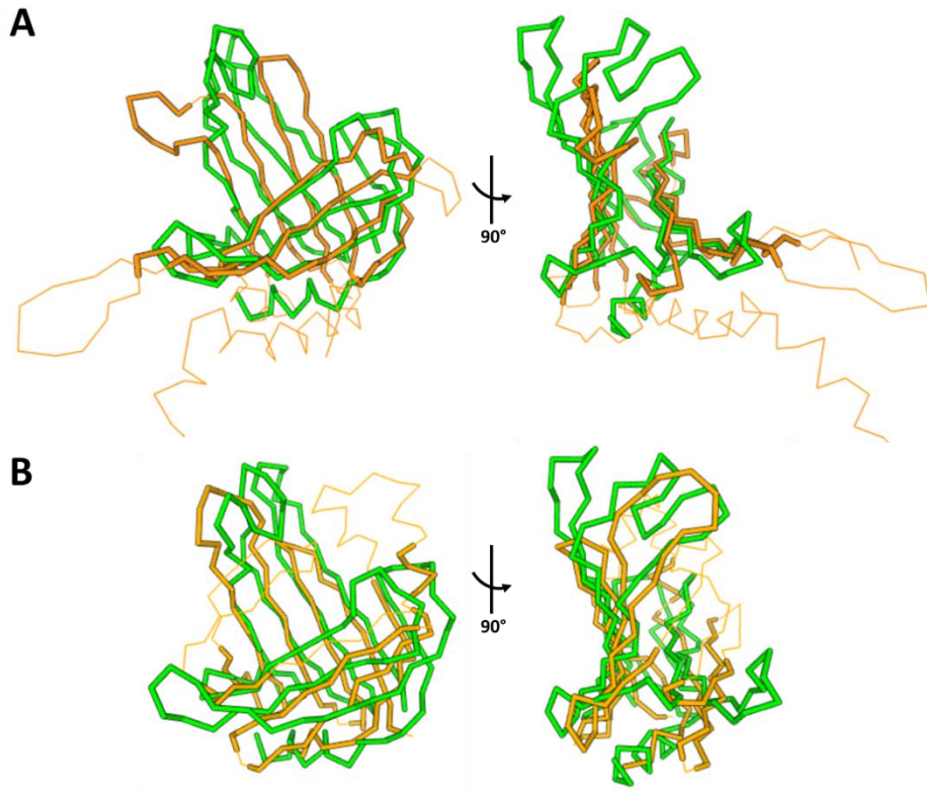


Figure 112: Structural comparison of Bd0675 with the two main folds found by Dali. Bd0675 backbone trace is represented in green whereas aligned Dali matches are represented in orange. **A)** Bd0675 structural comparison with PBF2 transcription factor (N°1 Dali match, PDB ID: 1l3a). Bd0675 preserves a β-roll like fold, however, major helical structures of P24 are missing and the conserved loops defining Bd0675 central cleft are absent in P24. **B)** Bd0675 structural comparison with the Cellular Retinoic Acid-Binding Protein 2 (cRABP2, PDB ID: 6nnx). The β-barrel fold of cRABP2 is not part of Bd0675 structure and thus, the central lumen of the β-barrel is absent in Bd0675.

As seen in **Figure 112A**, PBF2 (PDB ID: 1l3a) has the most similar topology to Bd0675, with a β-roll architecture and similarity on the core β-strands positioning. In contrast, the second type of fold found by Dali is represented in **Figure 112B** by the Cellular Retinoic Acid-Binding Protein 2 (cRABP2, PDB ID: 6nnx), which possesses a β-barrel architecture and a central lumen where ligands are bound, which is absent in the Bd0675 structure.

The PBF2 transcription factor is a 125 kD homotetramer of the Whirly family that binds to ssDNA (**Figure 113A**). PBF2 binds ssDNA in a basic cleft formed between the two sides of the β-roll fold, which interacts with the phosphodiester backbone of the nucleic acid (**Figure 113B**).

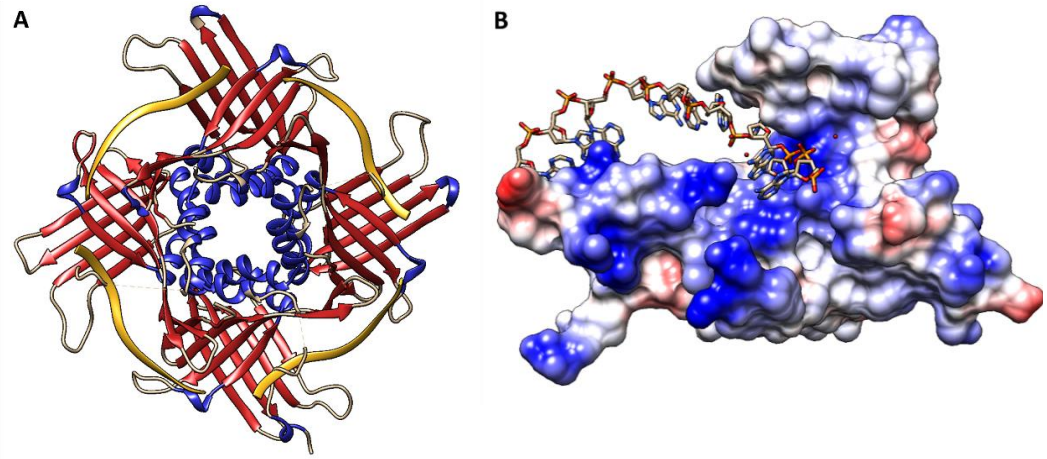


Figure 113: PBF2 structure and ssDNA binding. A) PBF2 tetrameric structure (α-helix in blue and β-strands in red) with bound ssDNA molecule (in yellow). **B)** Surface representation of PBF2 bound to a 32 mer oligo dT. The ssDNA binds to the basic cleft formed between the β-leaflets forming the β-roll fold of the protein. The surface coulombic potential is coded from red (-10 Kcal/mol) to blue (+10 Kcal/mol).

While the secondary structure of PBF2 bears the greatest similarity to Bd0675, critical elements of the Whirly family are missing in Bd0675. Firstly, the three α-helices mediating the tetramerization of PBF2 are absent in Bd0675, explaining the oligomerisation state differences between both proteins (**Figure 113A**). Secondly, the ligand binding cleft of both proteins is constituted by different arrangements of β-strand structures and amino acid sidechains. Specifically, a 17 residue-long positively charged structural element formed by two β-strands that is pivotal for ssDNA binding by PBF2 (**Figure 113B**), is not present in Bd0675. In fact, the ligand binding cleft of PBF2 is mainly positively charged in order to accommodate the phosphodiester backbone of the ssDNA molecule, whereas the central pocket of Bd0675 is mainly hydrophobic or acidic, delimited by negatively charged loops. Moreover, the disulphide bonded loop that contains most of the conserved residues of Bd0675 and defines the possible ligand binding specificity of Bd0675 is lacking in PBF2, suggesting that both proteins do not share a similar function.

The second type of fold represented by cRABP2 is a β-barrel, which are associated with lipid binding functions (PDB ID:6NNX, **Figure 114A**). cRABP2 possesses a central binding pocket, which is positively charged at the most central part of the pocket to stabilise the carboxyl group of lipids, and more hydrophobic towards the surface of the protein to accommodate the aliphatic chains of the ligand (**Figure 114B**).

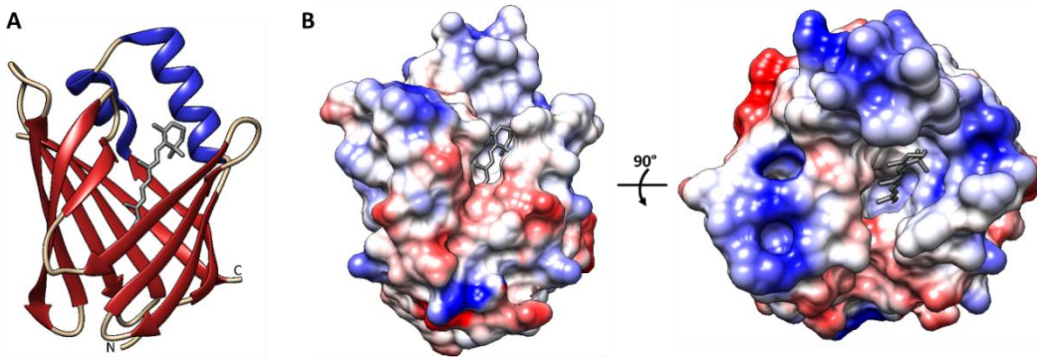


Figure 114: cRABP2 structure and lipid binding. A) cRABP2 secondary structure (α-helices in blue and β-strands in red) forming a β-barrel fold, which accommodates a retinoic acid molecule (grey) in the lumen of the protein. **B)** Coulombic potential surface representation with 90° angle differences of cRABP2 showing a central binding basic and hydrophobic pocket. The surface coulombic potential is coded from red (-10 Kcal/mol) to blue (+10 Kcal/mol).

The absence of the aforementioned features on Bd0675, plus the lack of conservation of the critical residues located at the putative ligand binding cleft of Bd0675 would suggest that Bd0675 is not a lipid binding protein.

6.2.6. Bd0675 ligand identification analyses

Even though at close inspection no match detected by Dali shows a similar fold or structural features to Bd0675, a possible ssDNA binding activity of the protein was studied by electrophoretic mobility shift assay (EMSA). Briefly, oligo dT probes of 15, 25, 35, 45 and 55 mer were incubated for 1 hour at room temperature with increasing amounts of purified Bd0675 in presence of magnesium and run on 12% acrylamide gels. A representative result is shown in **Figure 115** for the 35 mer oligo dT, nevertheless similar results were obtained for the rest of the tested oligonucleotides.

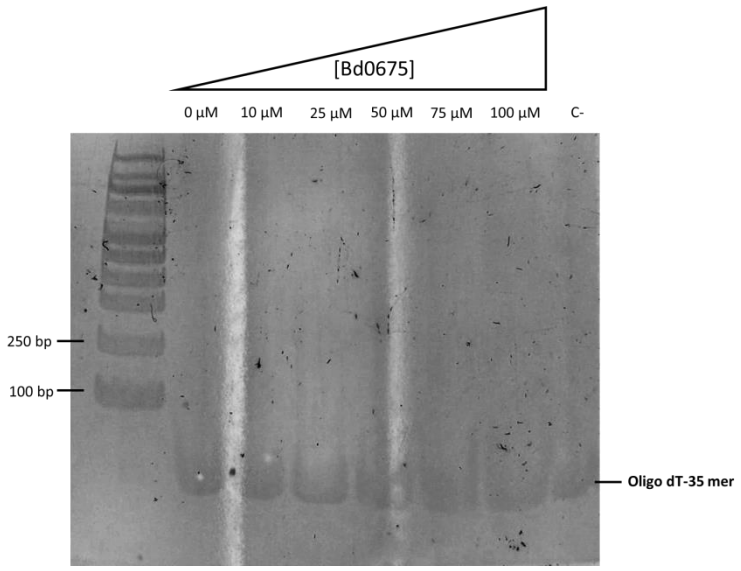


Figure 115: Electrophoretic mobility shift assay (EMSA) of Bd0675 using a 35 mer Oligo dT. 100 ng of oligo dT 35 mer were incubated with 0 to 100 µM of purified Bd0675, using BSA at 100 µM as negative control. No detectable shift in electrophoretic mobility was detected for the 35 mer oligo dT independent on the concentration of Bd0675.Similar results were obtained for 15, 25, 45 and 55 mer oligo dT.

No shift in electrophoretic mobility was detected for any of the tested oligonucleotides even at the highest concentration of Bd0675.

Furthermore, soaking experiments with oligonucleotides of six, ten and 15 mer were made, but no ligand molecule density was detected. These results support the hypothesis that Bd0675 does not bind to ssDNA, which agrees with the crucial differences between the ligand binding sites of both proteins.

Further co-crystallisation and crystal soaking experiments were performed using ligands such as glucose and glucosamine in case Bd0675 shows sugar binding functions, however no ligand binding density was observed in any X-ray diffraction dataset.

6.2.7. Structural predictions of Bd0675

In view of the unique fold adopted by Bd0675, the experimentally determined structure was submitted as a CASP14 target and contrasted with models predicted by different research groups [191].

The Bd0675 structure (blue, PDB ID: 7OC9, CASP14 target ID: T1074) was compared with the three best scoring predicted models: TS427 (red, AlphaFold2, GDT-TS = 91.10), TS352

(yellow, Laufer group, GDT-TS = 60.61) and TS140 (green, Yang group, GDT-TS = 56.44) (**Figure 116**).

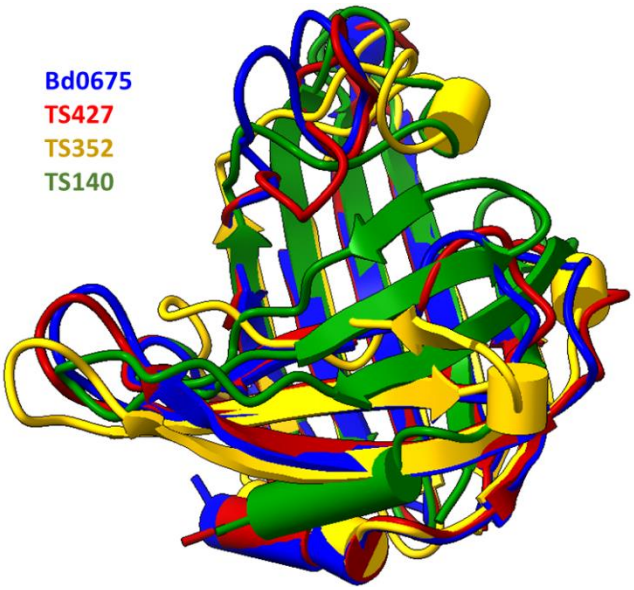


Figure 116: Structural prediction of Bd0675. The experimentally solved Bd0675 structure (blue) was superimposed with the CASP14 best scoring models TS427 (AlphaFold2, red), TS352 (Laufer group, yellow) and TS140 (Yang group, green) and depicted in cartoon representation. The overall structural features of the protein are significantly predicted by the different algorithms, with the best prediction corresponding to TS427.

These models predict the main fold features of the Bd0675 experimentally solved structure despite the lack of related sequences and homologous structures available in the PDB. Prediction quality GDT-TS values for each model correlate with RMSD values with respect to the Bd0675 crystallographic structure of 1.98, 4.14 and 5.65 Å for TS427, TS352 and TS140, respectively.

Analysis of the residues proposed to be significant for Bd0675 cryptic function were also compared between the experimental and predicted Bd0675 models.

As shown in **Figure 117**, only TS427 accurately predicts the C90-C118 disulphide bond that is proposed to position the Bd0675 conserved loop.

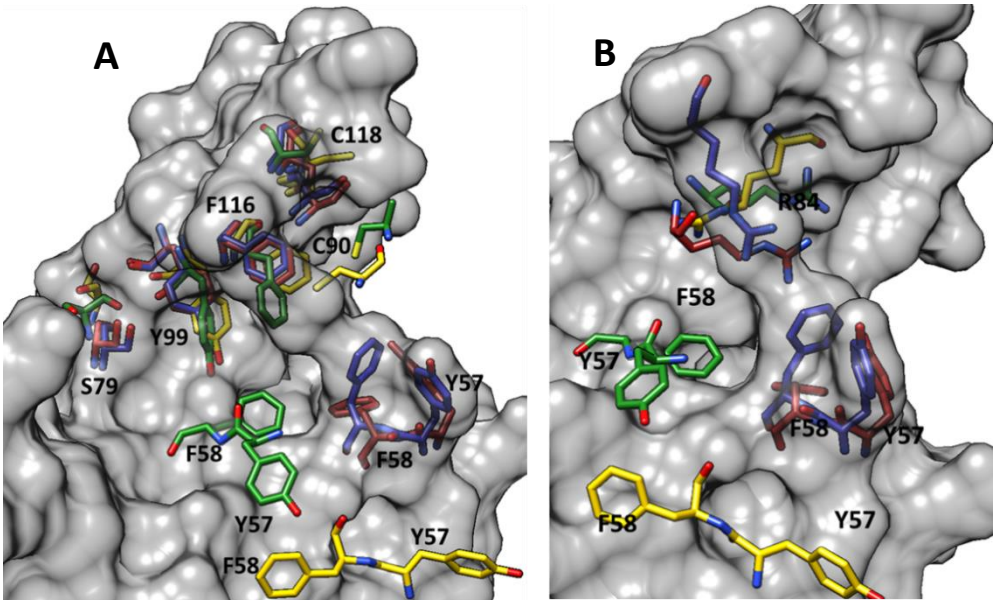


Figure 117: Comparison of the proposed ligand binding site residues of Bd0675 structure and predicted models. Two views, **A** and **B**, depict the proposed residues participating ligand binding are depicted in stick representation on Bd0675 surface representation (Bd0675 blue, TS427 red, TS352 yellow and TS140 green).

Sidechains of S79, Y99, and F116, which are oriented toward the proposed ligand binding groove of Bd0675, are depicted in different conformations for TS352 and TS140, whereas TS427 models these residues in the same orientation observed in the Bd0675 experimental structure (**Figure 117A**). However, none of the CASP14 best scoring predictions model the experimentally determined orientation of Y57 and F58. These two residues are shown to be invariant in Bd0675 orthologs across different predators suggesting their importance in Bd0675 function (**Figure 106**).

Moreover, residue R84 is modelled in different conformations for each CASP14 Bd0675 prediction (**Figure 117B**). This residue is observed forming a cation-pi interaction with F58 in the Bd0675 crystallographic structure, which is suggested to participate in positioning the Bd0675 disulphide-stabilised loop. As previously pointed, F58 is not correctly oriented in any of the CASP14 models, which may explain this discrepancy.

Therefore, even though structure predicting algorithms are successful in predicting the main structural elements of Bd0675, particularly TS427 by the AlphaFold2 algorithm, key specific elements such as residue orientation and sidechain interactions are fields to improve upon.

6.3. Discussion

6.3.1. Bd0675 expression, purification, and crystallisation

Bd0675 was successfully cloned fused to an N-terminal His_6 tag, however its expression under native conditions stopped culture growth and in consequence had to be produced in inclusion bodies and refolded. If Bd0675 function is connected to this observation remains to be studied.

Further, Bd0675 was effectively refolded on column and behaves as a monomeric spherical protein of the predicted molecular weight, which only crystallised in alcohol and MPD centred crystallisation screens, suggesting the existence of dielectric pockets in Bd0675 which are further discussed in context of the solved structure.

6.3.2. Bd0675 structure comprises a conserved predicted ligand binding cleft

The topological arrangement of Bd0675 is stabilised by the conserved disulphide bonds C31-C132 and C90-C118, wherein the latter keeps the conserved loop connecting beta strands β6 and β7 in a mostly fixed position. This fold defines an extensive cleft in the middle of the protein that is further evident when visualising the surface representation of the same.

The cleft consists of a central groove formed mainly by hydrophobic sidechains of beta strands β4 and β6-8, which is further delimited by three acidic lobes formed by the connecting loops between beta strands β1 and β2, β4 and β5, β3 and β4, and β6 and β7.

It is noticeable that most of the conserved residues localise towards the central cleft, surprisingly concentrated in the connecting loops of the protein. The major conserved motif (R/D)PGGXϕCGXΩX5Y is part of the disulphide loop connecting beta strands β6 and β7.

Amusingly, the C90 of the motif that forms the disulphide bond with C118 of β-strand 8 does not appear to be crucial for maintenance of the Bd0675 fold. The C90-C118 disulphide appears to be holding this conserved loop into place rather than further stabilising the overall structure, and when observed in context of the central cleft, seems to be pivotal on defining ligand binding function or specificity.

The properties of the conserved amino acids located in the central cleft of the structure suggest that the putative ligand for Bd0675 is mainly stabilised by hydrogen bonding and

hydrophobic interactions. These interactions would involve conserved aromatic and hydrophobic residues Y23, Y53, Y57, F58, Y99, F116 and L128, and the slightly less conserved basic sidechains of R84 and K101.

Even though the abovementioned residues define well conserved interacting pockets, this does not make obvious the nature of a possible Bd0675 ligand. The chemistry of the pockets suggests a molecule containing positively charged or hydrogen-bonding groups such as amines, carbonyls, and oxygen groups, along with extensive hydrophobic or aromatic groups that could interact with the conserved aromatic residues, particularly Y57, F58, Y99 and F116.

6.3.3. Bd0675 structure diverges significantly from ssDNA and lipid binding proteins

To further analyse the proposed ligand binding cleft, the solved Bd0675 structure was compared with publicly available structures deposited in the PDB using the Dali server.

As shown in **Table 18**, and illustrated in **Figures 109-111**, the structural similarity with the best Dali matches is modest at best (in case of PBF2) and poor when it comes to comparison with β-barrel-like folds.

The β-roll fold of PBF2 is marginally similar to the Bd0675 architecture. The overall fold of Bd0675 is three-dimensionally distorted compared to PBF2. This fold variation seems to be mainly stabilised by the C31-C132 disulphide bond connecting both beta leaflets of the protein. Moreover, the alpha helices mediating the physiological oligomerisation and ssDNA binding of PBF2 are absent in Bd0675. Furthermore, the ssDNA binding pocket of PBF2 agrees with general basic pockets observed in different nucleic acid binding proteins [192]. On the contrary, the observed Bd0675 cleft comprises mostly aromatic and acidic residues suggesting that they bind chemically divergent ligands. This hypothesis is further supported by the EMSA results showing no sequence-independent binding of 15-55 bp oligonucleotides.

On the other hand, the β-barrel Dali matches are lipid binding proteins, which bury the hydrophobic ligand in the lumen of the protein [193]. However, since the putative binding cleft of Bd0675 is solvent accessible and on the surface of the protein instead of being buried in the core of the same would make lipid binding entropically unfavourable.

Since one of the main targets of the *B. bacteriovorus* hydrolytic machinery is the prey's PG cell wall and that Bd0675 is one of the most upregulated genes in early phases of predation, the possibility of the glycan portion of the PG cell wall being the Bd0675 ligand was considered.

In this scenario, co-crystallisation and crystal soaking experiments were tested with glucosamine and glucose, however no ligand electron density was identified in the collected datasets. Moreover, preliminary isothermal titration calorimetry (ITC) experiments using glucosamine as ligand showed no detectable interaction between this sugar and Bd0675. If Bd0675 binds chemically modified sugars, such as those generated by deacetylation, remains unexplored [117].

Therefore, additional experiments such as pull-downs using prey's and/or predator's lysate, binding experiments with other candidate molecules and complimentary information such as localisation experiments or characterisation of knock-out predation phenotypes are necessary to identify the biological function of Bd0675.

6.3.4. Bd0675 structure prediction by CASP14 algorithms

Bd0675 was submitted for CASP14 prediction experiments considering the unique fold depicted by the experimentally determined structure. As discussed in the corresponding publication by Alexander *et al.* (2021), AlphaFold2 is the best performing algorithm with all highlighted targets, including Bd0675 [191].

Remarkably, AlphaFold2 is particularly accurate in predicting the Bd0675 overall fold despite the lack of available structural templates deposited in the PDB. Nonetheless, there is potential for improvement even for small proteins such as Bd0675, a feature that has been shown to positively correlate with structural prediction quality [145], [180].

Particularly, conserved residues that are proposed to be key for Bd0675 function are improperly oriented. This is further evidenced by the incorrect modelling of the Bd0675 disulphide-stabilised loop, which is the most conserved part of the protein sequence and delimits the proposed ligand binding cleft of the protein.

Notwithstanding their limitations, these prediction algorithms may work synergistically with experimental structure elucidation, by providing MR models, aiding in construct

design as shown for Bd1177 in section 5.2.2.4 or even proposing possible conformational fluctuations with respect to the experimentally solved structures [194].

6.4. Conclusion

B. bacteriovorus deploys an arsenal of novel proteins and enzymes to support the progression of an endobiotic predatory lifecycle. An important portion of the predatosome proteins do not possess putative domains with known functions, hindering the understanding of predation at the molecular level. One of the most upregulated genes upon predation is Bd0675, which does not contain any readily identifiable putative domain conserved in its sequence.

The crystallographic structure of Bd0675 has been solved at 1.50 Å, showing that the protein possesses a novel β-rich topology stabilised by two highly conserved disulphide bonds, with a central cleft that may act as a potential ligand binding site. By analysing Bd0675 homologues, it has been determined that most of the sequence conservation is located towards the central groove of the protein, further supporting its role in Bd0675 function.

Comparing the structure of Bd0675 with available solved structures in the PBD does not reveal a close similarity with any available fold. Bd0675 closest structures correspond to the PBF2 ssDNA binding protein and β-barrel lipid carrier proteins. Both folds and ligand binding pockets differ considerably from the putative ligand binding site of Bd0675 suggesting that this protein would not act as either a ssDNA binding protein or lipid carrier, hypothesis supported by EMSA results showing no binding to ssDNA oligonucleotides. Other possible ligands such as sugars were also tested but with no conclusive results. Thus, further experiments are needed to define the ligand for Bd0675 and propose a function for this protein in the predatory lifecycle of *B. bacteriovorus*.

Regarding the CASP14 structural predictions, the AlphaFold2 algorithm exceeded other tools, but with limitations mainly on sidechain orientation and loop modelling. Nevertheless, the fast-paced improvement of prediction methods proves useful to structural experimentalists, providing complimentary approaches and information on the subject proteins.

CHAPTER 7

7. SUMMARY AND FUTURE WORK

7.1. Summary

This work was mainly focused on the structural dissection of the *B. bacteriovorus* predatosome, particularly the PG modifying machinery and novel hypothetical proteins [21].

Four novel LDT structures have been solved, completing a panel of seven enzymes representing all six groups encoded by *B. bacteriovorus*. Each group is characterised by different accessory domains and substrate binding sites, indicating that they fulfil specific roles in predation.

Additionally, the structure of the putative DUF1460 amidohydrolase Bd0601 has been determined. This structure shows conservation of critical features of the predicted enzyme and acquisition of possible predatory adaptations.

Further, the structure of Bd1177 has been solved, which is predicted as an evolutionary divergent NlpC/P60. This new structure comprises a predicted regulatory C-terminal domain unprecedented for this group of enzymes.

Finally, a structural-based assignment approach was taken to study the cryptic secreted protein Bd0675. Bd0675 structure depicts a unique fold with possible ligand binding functions to be determined.

7.1.1. L,D-transpeptidases

LDTs share a YkuD- or YkuD2-like catalytic core and may be further classified based on their active site into L,D-transferases (groups 2-4) and L,D-hydrolases (YkuD-like groups 1 and 6, and YkuD2-like group 5). **Figure 118** summarises the total of solved LDT structures and putative overall activities. The LDTs solved in this thesis are indicated with an asterisk (*).

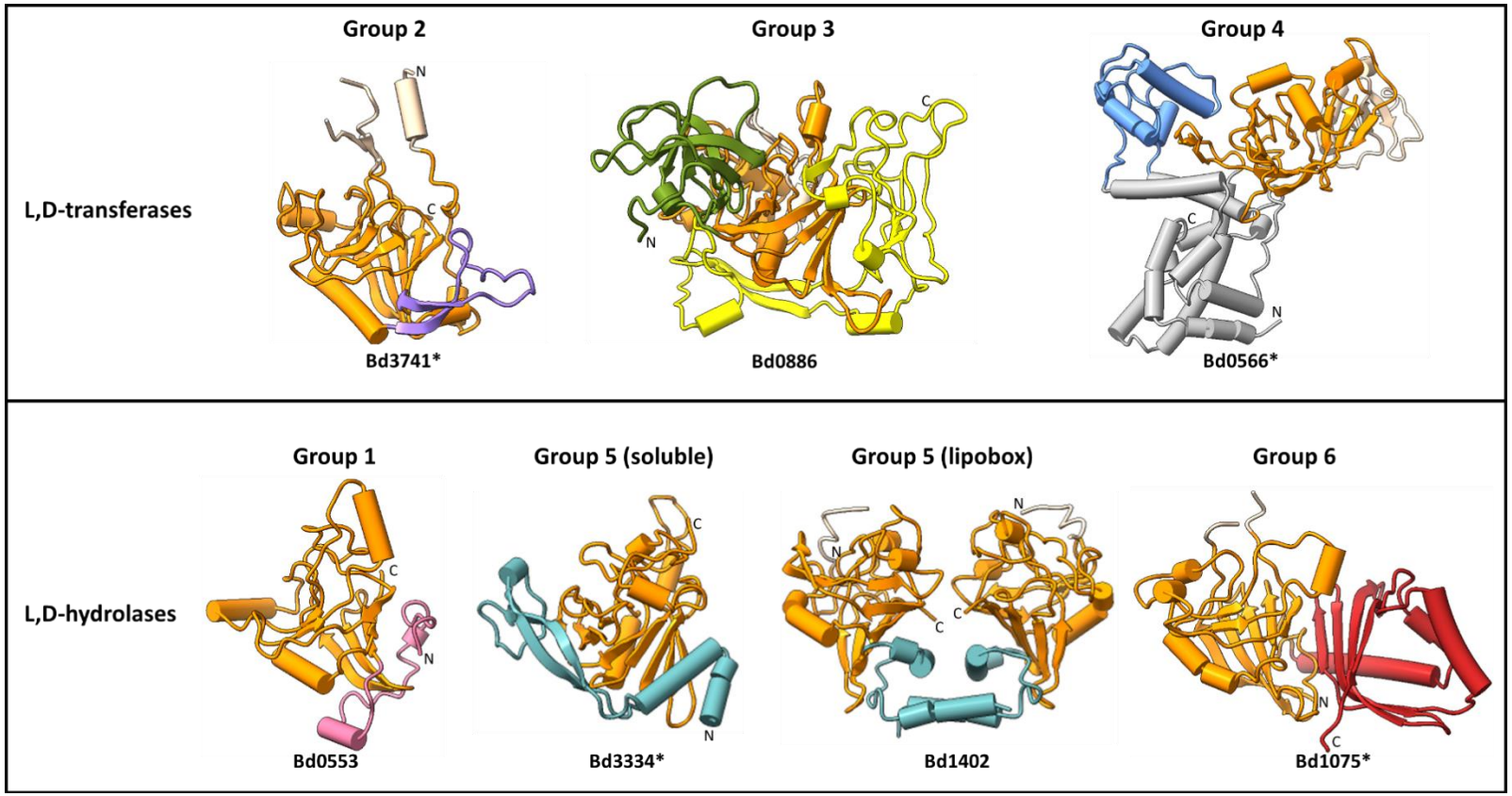


Figure 118: Summary of *B. bacteriovorus* HD100 LDT structures. *B. bacteriovorus* HD100 LDTs can be divided into six groups which can be further classified as L,D-transferases or L,D-hydrolases. (*) indicates LDT structures determined in this thesis. Experimentally determined structures for each group are depicted in cartoon format, wherein the YkuD- or YkuD2-like catalytic domain is coloured in orange. Characteristic accessory domains of each group are coloured as follows: helical-rich domain in pink (group 1), triangular hook in purple (group 2), SH3b domain and cryptic domain in green and yellow, respectively (group 3), scaffolding domain and PGBD in grey and blue, respectively (group 4), unique insertions in teal (group 5 soluble and lipobox), and NTF2-like domain in red (group 6, non-predatory). Bd1402 is represented as a dimer.

Based on the structural information and the reported PG modifications occurring during bdelloplast formation, we have proposed that L,D-transferases from groups 2-4 modify the PG cell wall by crosslinking phospholipids, adding terminal DAP or other NCDAA, and condensing 3-3 stem peptide crosslinks, respectively [39], [89]. It is suggested that substrate specificity is defined, at least partially, by the accessory domains of each LDT group, for instance, by the triangular hook domain and active site lid of Bd3741, and the capping subdomain and putative PGBD of Bd0566.

On the other hand, predatory L,D-hydrolases are represented by group 1 and group 5 (YkuD- and Ykud2-like hydrolases, respectively). Structural analyses reveal that group 1, and the soluble hydrolases of group 5 are characterised by an accessible active site. Thus, these enzymes are suggested to have both carboxypeptidase and endopeptidase activities, which is supported by Bd0553 activity data (unpublished). Further, membrane-bound L,D-hydrolases of group 5, represented by Bd1402, are shown to have an active site lid, which has been observed to limit substrate size in other models [163], [164].

Additionally, the non-predatory L,D-carboxypeptidase Bd1075 (group 6) structure is analysed in view of functional and biological data, showing that this enzyme regulates *B. bacteriovorus* curvature [125]. In this case, the NTF2-like domain defines the subcellular localisation of the enzyme by means of a unique C-terminal PG-binding pocket.

Finally, we have proposed that the presence and complexity of an active site-lid, along with the chemistry of the acyl-acceptor site residues, partially define the transferase or hydrolase preference of each LDT group.

Altogether, these results provide the structural framework for analysing cell wall modifications occurring during the *B. bacteriovorus* predatory life cycle.

7.1.2. NlpC/P60 amidohydrolases and peptidases

The structure of two NlpC/P60 amidohydrolases/peptidases, Bd0601 and Bd1177, are presented in this work.

Bd0601 is a putative DUF1460 amidohydrolase encoded by the tripartite operon *bd0601-0603*. This operon additionally encoded two membrane-anchored cryptic proteins, Bd0602 and Bd0603.

The Bd0601 structure adopts a papain-like fold with structural insertions that are characteristic of the DUF1460 family of enzymes. We have shown that regardless of the structural conservation of Bd0601, key differences in the stem peptide binding groove may affect the substrate specificity, particularly the steric effect of the Y273 sidechain that is only present in *B. bacteriovorus*.

On the other hand, Bd1177 is a structurally divergent NlpC/P60, comprising the putative catalytic N-terminal domain and a novel C-terminal domain. The Bd1177 N-terminal domain depicts a papain-like core surrounded by several insertions not reported for NlpC/P60s, including Bd0601. Further, analysis of the interdomain interface of Bd1177 suggests that substrate binding requires displacement of the C-terminal domain, suggesting it may participate in substrate selection.

Both structures provide evidence of specialised amidohydrolases/peptidases having a role in *B. bacteriovorus* specific PG modifications during predation.

7.1.3. Hypothetical cryptic secreted protein Bd0675

bd0675 encodes a hypothetical protein from *B. bacteriovorus* HD100 observed to be upregulated during predation [21]. We have determined the structure of Bd0675, which adopts a novel β-roll-like architecture stabilised by two conserved disulphide bonds. The structure defines a prominent central acidic/hydrophobic groove in the structure suggesting that it acts as a ligand binding protein.

Even though our efforts in identifying a putative ligand were inconclusive, the solved structure proved useful as a CASP14 target, including structure prediction algorithms such as the recently developed AlphaFold2 [145], [191]. AlphaFold2 proved to be the most accurate prediction tool even with the lack of Bd0675-related structures and sequences available. Nevertheless, refinement of sidechain orientations and loop conformations is required, especially considering that this elements form part of the proposed ligand binding cleft of this protein.

7.2. Discussion and future work

7.2.1. L,D-transpeptidases

The solved LDT structures provide insight in the evolution and specialisation of these enzymes. Nevertheless, biochemical, and functional information are necessary for further understating their role in the *B. bacteriovorus* predatory cycle.

There is preliminary evidence supporting the endopeptidase and carboxypeptidase activity of group 1 Bd0553 against 3-3 crosslinked stem peptides. Nonetheless, hydrolytic activity against OM-linked PG substrates should be tested as done for *E. coli* DpaA [118], [142].

The enzymatic activities and substrate specificities of the putative group 5 YkuD-2 like L,D-hydrolases (Bd1402 and Bd3334) should also be determined. Bd1402 and Bd3334 active sites include two aromatic residues that are absent in YkuD-like LDTs (Y208 and W220 in Bd1402, and Y191 and W215 in Bd3334). This suggests that YkuD- and YkuD2-like L,D-hydrolases might have different substrate specificities, which could be tested using 3-4, 3-3, OM-linked and non-crosslinked PG substrates and detecting the hydrolysis products by mass spectrometry or high performance liquid chromatography (HPLC).

On the other hand, 3-3 crosslinking activity of Bd1181 (Bd0566 homologue) has been observed, which correlates with structural features shared between group 4 of LDTs and *E. coli* and *S. enterica* YcbB [83], [147]. Crosslinking activity of groups 2 and 3 should be determined to hypothesise about the structural divergence of these L,D-transferases.

Previous studies have stablished that the bdelloplast PG modifications include fatty acids and DAP attachment to the stem peptide [39], [153]. Additionally, it has been determined that a double in frame deletion of group 3 of LDTs (*Δbd0886Δ1176*) reduces the incorporation of NCDAA in bdelloplast PG [33].

Therefore, it is necessary to determine if Bd0886 (group 3) is able to incorporate DAP, and possibly other NCDAA, to prey's PG. In this same sense, it would prove interesting to test this possible NCDAA-exchanging activity using Bd0886 truncated mutants wherein the SH3b domain has been deleted. With these experiments, it could be possible to propose if this accessory domain participates in substrate binding or in enzyme localisation.

Group 2 is predicted to have L,D-transferase activity based on active site sequences [142]. One hypothesis is that group 2 LDTs, such as Bd3741, catalyse the 3-3 crosslinking of fatty acid-esterified PG molecules or stem peptides with a terminal DAP added [39]. This could be tested by using bdelloplast acylated PG sacculi, which is separable from non-acylated PG, and measuring 3-3 crosslinking activity of the enzyme by mass spectrometry or HPLC [39].

In this same sense, 3-3 crosslinking activity of acylated or terminal DAP PG could be measure for group 4 of LDTs (Bd0566 and Bd1181) to determine the difference in substrate specificity of these two groups. Moreover, it would prove interesting to test truncated Bd3741 mutants (group 2), wherein the triangular hook has been deleted to dissect the role of this domain in substrate selection.

Additional experiments such as PG sacculi pull downs using domain truncation mutants of the solved LDTs could aid in stablishing if the LDT accessory domains participate directly in substrate selection or are required for other events, such as enzyme localisation or protein-protein interactions. Similarly, any co-dependence of the different PG modifications between LDTs and even with other enzymes such as DDTs, TGases and deacetylases is a new area to consider, for instance, if the LDT reactions require a terminal NCDAA different from D-Ala [153].

On the other hand, further studies to understand the impact of active site lids are a particularly interesting subject.

M. tuberculosis LDTs Ldt_{Mt1}, Ldt_{Mt2}, Ldt_{Mt3} and Ldt_{Mt5} structures have been solved, showing that the active site of these enzymes is capped by a lid [75], [78]–[80]. In Ldt_{Mt2}, it has been determined that upon carbapenem acylation of the catalytic cysteine, the active site lid moves, closing the active site [79]. This conformational shift is not observed in Ldt_{Mt1} or Ldt_{Mt3}, wherein the apo and faropenem inactivated enzymes are depicted in the same closed conformation [78], [80].

On the other hand, Ldt_{Mt5}, which is a non-redundant paralogue of Ldt_{Mt2}, is observed to shift from a closed to a more solvent accessible conformation of the active site by displacing the lid upon substrate binding by an unknown mechanism [75]. This conformational change is proposed to orientate the active site residues of Ldt_{Mt5} in a catalytically competent conformation [75].

Thus, the structural evidence suggests that the mechanism defining the acyl-acceptor site by the movement of the active site lid is different depending on the LDT.

Therefore, active site lids could have different roles depending on the LDT. As an example, the lipobox LDT Bd1402 is the only predatory LDT with an active site lid. In this case, generating a solvent accessible active site mutant or varying the length of the lid could potentially modify its carboxypeptidase/endopeptidase preference or modify substrate specificity, as observed in GGTs and suggested for *M. tuberculosis* LDTs [75], [163], [164].

7.2.2. NlpC/P60 amidohydrolases and peptidases

The role of Bd0601 in predation has not been explored. Further, new conditions to produce a catalytically active Bd0601 should be tested, even though structural conservation of the DUF1460 domain features suggests that Bd0601 is possibly an amidohydrolase. In this sense, a catalytically active Bd0601 could make possible to determine if Y273 is determinant for its enzymatic activity or specificity.

Bd0602 and Bd0603 are encoded in the same operon as Bd0601. These two proteins predicted to have a lipobox SP and putative ankyrin-like repeats in their sequences. These features have been observed in toxin-antitoxin protection strategies deployed by other bacteria, exposing the membrane-bound immunity factors while the toxin or virulence factor is secreted or translocated [36], [129], [195], [196].

Since no soluble protein was obtained from the tested constructs, different domain boundaries could be thoroughly tested to express soluble truncated Bd0602 and Bd0603 mutants to study their association with Bd0601. On the other hand, these proteins could be expressed as membrane-bound proteins retaining a lipobox SP, alone or in combination with Bd0601, which could improve their stability.

The second studied NlpC/P60 Bd1177 defines a novel group of enzymes that are open to biochemical characterisation. Substrate specificity, deletion phenotypes and subcellular localisation of Bd1177 remain unexplored.

Once the enzymatic activity of Bd1177 has been stablished (for example as an endopeptidase or amidohydrolase), it would prove possible to study the role of the proposed conformational change of Bd1177 in catalysis. We have proposed that to

accommodate a PG substrate, Bd1177 must transition from a closed to an open state, by displacing the putative regulatory C-terminal domain.

Besides structural approaches aimed to crystallise a ligand-bound form of Bd1177, biochemical and biophysical experiments could support a possible conformational change of the putative enzyme.

For example, determining the enzymatic activity effect of generating permanently open or closed Bd1177 mutants by mutating the Y124/Y508 interacting pair, as proven for the disulphide locked mutant Y124C/Y508C.

Opening and closing of the enzyme could be monitored by fluorescently labelling each half of the enzyme and monitoring the Förster resonance energy transfer (FRET) of between each fluorescent probe. Fluorescence quenching of intrinsic tryptophanes in the interdomain interface of Bd1177 is also an alternative to determine their solvent accessibility upon substrate binding.

7.2.3. Hypothetical cryptic secreted protein Bd0675

Finally, for the cryptic protein Bd0675, further functional and biochemical experiments are needed. For instance, by determining the localisation of this protein (predator or bdelloplast, for example), it could be possible to perform pull-down experiments and identify possible bound molecules by mass spectrometry. Furthermore, other major metabolites such as PG stem peptides, phospholipids and polyamines could be explored as prospective ligands for this protein.

7.2.4. *B. bacteriovorus* and possible applications of PG modifying enzymes

B. bacteriovorus predation has been exploited to control pathogenic bacteria such as those of the *Pseudomonas*, *Acinetobacter* and *Klebsiella* genera [197], [198]. *B. bacteriovorus* has also been used as food and crop pathogen bactericide and as food preservative [199]–[201]. In clinical settings, *B. bacteriovorus* has been used in the topical treatment of ocular infections and cystic fibrosis isolates of *P. aeruginosa* and *S. aureus*, among other examples [202], [203]. Further, a patent application related to compositions claiming improvement of neurodegenerative diseases as Amyotrophic Lateral Sclerosis (ALS) has been published [204].

At the enzymatic level, *B. bacteriovorus* has been used for modulating and mapping therapeutic antibodies [205]. For example, the serine protease BspK (Bd1962) has been shown to hydrolyse IgG_1 at the hinge region, enabling its mass spectrometry mapping [206]. Similarly, the elastase-like serine protease BspE (Bd2692) specifically hydrolyses the Fc-tail from IgA [207].

It has also been determined that host-independent *B. bacteriovorus* produce extracellular enzymes that are efficacious against *S. aureus* biofilms, despite it being a Gram-positive bacterium [208]. The identified enzymes included the serine proteases BspK and Bd2269, the carboxypeptidase Bd0306 and unspecified DNases [208].

Additionally, *B. bacteriovorus* has been used for recovering intracellular bio-products as polyhydroxyalkanoate (PHA) in a more efficient, cost-effective and environmentally-safe when compared to classical physical and chemical methods of cell disruption [209]. In general terms, transpeptidases have been used to produce block polymers with different potential applications, such as producing thermoplastic elastomers and polymer blends (Patent application EP3652199A1) [210].

Further, the bacteriolytic activity of bacteriophage amidohydrolases has been proven useful in controlling pathogen agents (US2013336954A1), multi-drug resistant bacteria (US2010158886A1) and treating sludge (WO2020173473A1) [211]–[213].

Thus, considering the possible hydrolytic activities of the LDTs and putative NlpC/P60 endopeptidases discussed in this thesis, it could be possible to formulate bacteriolytic compositions comprising these enzymes. These enzymatic compositions could be used as pathogen control formulation or for recovery of intracellular biopolymers and are potentially easier to implement when compared to whole cells. Moreover, in view of the possible novel specificities and domains of *B. bacteriovorus* LDTs, these enzymes could be used for producing chemically diverse biopolymers.

Altogether, the novel substrate specificities and activities of *B. bacteriovorus* LDTs, NlpC/P60 enzymes and cryptic proteins could generate novel complex molecules and biological formulations with applications in biomedicine and industry.

REFERENCES

[1] O. Toussaint and E. D. Schneider, “The thermodynamics and evolution of complexity in biological systems,” *Comp. Biochem. Physiol.*, vol. 120, pp. 3–9, 1998.

[2] R. Guerrero, C. Pedrs-ali, I. Esteve, J. Mas, and D. Chaset, “Predatory prokaryotes : Predation and primary consumption evolved in bacteria,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 83, no. April, pp. 2138–2142, 1986.

[3] E. Jurkevitch, “Predatory Behaviors in Bacteria-Diversity and Transitions,” *Microbe-American Soc. Microbiol.*, vol. 2, no. 2, pp. 67–73, 2007.

[4] J. Pérez, A. Moraleda-muñoz, F. J. Marcos-torres, and J. Muñoz-dorado, “Bacterial predation : 75 years and counting !,” *Environ. Microbiol.*, vol. 18, no. 3, pp. 766–779, 2016.

[5] J. Muñoz-dorado, F. J. Marcos-torres, E. García-bravo, A. Moraleda-muñoz, and J. Pérez, “Myxobacteria : Moving , Killing , Feeding , and Surviving Together,” *Front. Microbiol.*, vol. 7, no. May, pp. 1–18, 2016.

[6] W. Zhang *et al.*, “Dynamics of solitary predation by Myxococcus xanthus on Escherichia coli observed at the single-cell level,” *Appl. Environ. Microbiol.*, vol. 86, no. 3, pp. 1–13, 2020.

[7] R. C. Marshall and D. E. Whitworth, “Is ‘ Wolf - Pack ’ Predation by Antimicrobial Bacteria Cooperative ? Cell Behaviour and Predatory Mechanisms Indicate Profound Sel fi shness , Even when Working Alongside Kin,” *Bioessays*, vol. 41, no. 4, p. 1800247, 2019.

[8] B. A. Hungate *et al.*, “The Functional Significance of Bacterial Predators,” *MBio*, vol. 12, no. 2, pp. e00466-21, 2021.

[9] Y. Davidov and E. Jurkevitch, “Diversity and evolution of Bdellovibrio -and-like organisms (BALOs), reclassification of Bacteriovorax starrii as Peredibacter starrii gen . nov ., comb . nov ., and description of the Bacteriovorax – Peredibacter clade as Printed in Great Britain,” *Int. J. Syst. Evol. Microbiol.*, vol. 54, no. 5, pp. 1439–1452, 2004.

[10] Z. Pasternak *et al.*, “In and out : an analysis of epibiotic vs periplasmic bacterial predators,” *ISME J.*, vol. 8, no. 3, pp. 625–635, 2013.

[11] Z. Wang, D. E. Kadouri, and M. Wu, “Genomic insights into an obligate epibiotic bacterial predator : Micavibrio aeruginosavorus Genomic insights into an obligate epibiotic bacterial predator : Micavibrio aeruginosavorus,” *BMC Genomics*, vol. 453, no. September, p. 453, 2011.

[12] S. F. Koval, S. H. Hynes, R. S. Flannagan, Z. Pasternak, Y. Davidov, and E. Jurkevitch, “Bdellovibrio exovorus sp . nov ., a novel predator of Caulobacter crescentus,” *Int. J. Syst. Evol. Microbiol.*, vol. 63, no. 1, pp. 146–151, 2013.

[13] R. J. Atterbury *et al.*, “Effects of Orally Administered Bdellovibrio bacteriovorus on the Well-Being and Salmonella Colonization of Young Chicks □ †,” *Appl. Environ. Microbiol.*, vol. 77, no. 16, pp. 5794–5803, 2011.

[14] D. E. Shatzkes, K., Singleton, E., Tang, C., Zuena, M., Shukla, S., Gupta, S., Dharani, S., Onyile, O., Rinaggio, J., Connell, N. D., & Kadouri, "Predatory Bacteria Attenuate Klebsiella pneumoniae Burden in Rat lungs," *MBio*, vol. 7, no. 6, pp. 1–9, 2016.

[15] A. R. Willis *et al.*, "Injections of Predatory Bacteria Work Alongside Host Immune Cells to Treat Shigella Infection in Zebrafish Larvae Report Injections of Predatory Bacteria Work Alongside Host Immune Cells to Treat Shigella Infection in Zebrafish Larvae," *Curr. Biol.*, vol. 26, no. 24, pp. 3343–3351, 2016.

[16] H. Stolp and M. P. Starr, "Bdellovibrio bacteriovorus gen. etsp. n., a predatory, ectoparasitic, and bacteriolytic microorganism," *Antonie Van Leeuwenhoek*, vol. 29, pp. 217–248, 1963.

[17] S. Rendulic *et al.*, "A Predator Unmasked : Life Cycle of Bdellovibrio bacteriovorus from a Genomic Perspective," *Science (80-.).*, vol. 303, no. January, pp. 689–693, 2004.

[18] R. E. Sockett, "Predatory Lifestyle of Bdellovibrio bacteriovorus," *Annu. Rev. Microbiol.*, vol. 63, no. 1, pp. 523–539, 2009.

[19] N. Said, A. Chatzinotas, and M. Schmidt, " Have an Ion on It: The Life-Cycle of Bdellovibrio bacteriovorus Viewed by Helium-Ion Microscopy ," *Adv. Biosyst.*, vol. 3, no. 1, p. 1800250, 2018.

[20] T. W. Cotter and M. F. Thomashowl, "Identification of a Bdellovibrio bacteriovorus Genetic Locus , hit , Associated with the Host-Independent Phenotype," *J. Bacteriol.*, vol. 174, no. 19, pp. 6018–6024, 1992.

[21] C. Lambert, C. Y. Chang, M. J. Capeness, and R. E. Sockett, "The first bite - Profiling the predatosome in the bacterial pathogen Bdellovibrio," *PLoS One*, vol. 5, no. 1, 2010.

[22] M. J. Capeness *et al.*, "Activity of Bdellovibrio Hit Locus Proteins , Bd0108 and Bd0109 , Links Type IVa Pilus Extrusion / Retraction Status to Prey-Independent Growth Signalling," *PLoS One*, vol. 8, no. 11, p. e79759, 2013.

[23] G. Prehna, B. E. Ramirez, and A. L. Lovering, "The Lifestyle Switch Protein Bd0108 of Bdellovibrio bacteriovorus Is an Intrinsically Disordered Protein," *PLoS One*, vol. 9, no. 12, p. e115390, 2014.

[24] S. C. Straley and S. F. Conti, "Chemotaxis by Bdellovibrio bacteriovorus toward prey," *J. Bacteriol.*, vol. 132, no. 2, pp. 628–640, 1977.

[25] S. C. Straley, A. G. LaMarre, L. J. Lawrence, and S. F. Conti, "Chemotaxis of Bdellovibrio bacteriovorus toward pure compounds," *J. Bacteriol.*, vol. 140, no. 2, pp. 634–642, 1979.

[26] C. Lambert, M. C. M. Smith, and R. E. Sockett, "A novel assay to monitor predator-prey interactions for Bdellovibrio bacteriovorus 109 J reveals a role for methyl-accepting chemotaxis proteins in predation," *Environ. Microbiol.*, vol. 5, no. 2, pp. 127–132, 2003.

[27] S. G. Caulton and A. L. Lovering, “Bacterial invasion and killing by predatory Bdellovibrio primed by predator prey cell recognition and self protection,” *Curr. Opin. Microbiol.*, vol. 56, pp. 74–80, 2020.

[28] R. W. Meek, I. T. Cadby, P. J. Moynihan, and A. L. Lovering, “Structural basis for activation of a diguanylate cyclase required for bacterial predation in Bdellovibrio,” *Nat. Commun.*, vol. 10, no. 1, 2019.

[29] U. Jenal, A. Reinders, and C. Lori, “Cyclic di-GMP: Second messenger extraordinaire,” *Nat. Rev. Microbiol.*, vol. 15, no. 5, pp. 271–284, 2017.

[30] M. C. Duncan *et al.*, “High-throughput analysis of gene function in the bacterial predator bdellovibrio bacteriovorus,” *MBio*, vol. 10, no. 3, 2019.

[31] H. Im, H. Kwon, G. Cho, J. Kwon, S. Y. Choi, and R. J. Mitchell, “Viscosity has dichotomous effects on Bdellovibrio bacteriovorus HD100 predation,” *Environ. Microbiol.*, vol. 21, no. 12, pp. 4675–4684, 2019.

[32] M. Varon and M. Shilo, “Interaction of Bdellovibrio bacteriovorus and Host Bacteria,” *J. Bacteriol.*, vol. 99, no. 1, pp. 136–141, 1969.

[33] E. Kuru *et al.*, “Fluorescent D-amino-acids reveal bi-cellular cell wall modifications important for Bdellovibrio bacteriovorus predation,” *Nat. Microbiol.*, vol. 2, no. 12, pp. 1648–1657, 2017.

[34] K. J. Evans, C. Lambert, and R. E. Sockett, “Predation by Bdellovibrio bacteriovorus HD100 requires type IV pili,” *J. Bacteriol.*, vol. 189, no. 13, pp. 4850–4859, 2007.

[35] A. L. Lovering *et al.*, “Specialized Peptidoglycan Hydrolases Sculpt the Intra-bacterial Niche of Predatory Bdellovibrio and Increase Population Fitness,” *PLoS Pathog.*, vol. 8, no. 2, p. e1002524, 2012.

[36] C. Lambert *et al.*, “Ankyrin-mediated self-protection during cell invasion by the bacterial predator Bdellovibrio bacteriovorus,” *Nat. Commun.*, vol. 6, pp. 1–9, 2015.

[37] M. F. Thomashow and S. C. Rittenberg, “Intraperiplasmic growth of Bdellovibrio bacteriovorus 109J: Solubilization of Escherichia coli Peptidoglycan,” *J. Bacteriol.*, vol. 135, no. 3, pp. 998–1007, 1978.

[38] M. F. Thomashow and S. C. Rittenberg, “Intraperiplasmic growth of Bdellovibrio bacteriovorus 109J: N-Deacetylation of Escherichia coli Peptidoglycan Amino Sugar,” *J. Bacteriol.*, vol. 135, no. 3, pp. 1008–1014, 1978.

[39] M. F. Thomashow and S. C. Rittenberg, “Intraperiplasmic growth of Bdellovibrio bacteriovorus 109J: Attachment of long-chain fatty acids to Escherichia coli peptidoglycan,” *J. Bacteriol.*, vol. 135, no. 3, pp. 1015–1023, 1978.

[40] A. Typas, M. Banzhaf, C. A. Gross, and W. Vollmer, “From the regulation of peptidoglycan synthesis to bacterial growth and morphology,” *Nat. Rev. Microbiol.*, vol. 10, pp. 123–136, 2012.

[41] A. L. Lovering, S. S. Safadi, and N. C. J. Strynadka, “Structural Perspective of Peptidoglycan Biosynthesis and Assembly,” *Annu. Rev. Biochem.*, vol. 81, pp. 451–478, 2012.

[42] T. J. Wyckoff, J. A. Taylor, and N. R. Salama, “Beyond growth: Novel functions for bacterial cell wall hydrolases,” *Trends Microbiol.*, vol. 20, no. 11, pp. 540–547, 2012.

[43] A. K. Yadav, A. Espaillat, and F. Cava, “Bacterial strategies to preserve cell wall integrity against environmental threats,” *Front. Microbiol.*, vol. 9, no. AUG, pp. 1–9, 2018.

[44] E. D. Brown, E. I. Vivas, and C. T. Walsh, “MurA (MurZ), the Enzyme That Catalyzes the First Committed Step in Peptidoglycan Biosynthesis , Is Essential in Escherichia coli,” *J. Bacteriol.*, vol. 177, no. 14, pp. 4194–4197, 1995.

[45] M. J. Pucci, L. F. Discotto, and T. J. Dougherty, “Cloning and Identification of the Escherichia coli murB DNA Sequence , Which Encodes UDP-N-Acetylenolpyruvoylglucosamine Reductase,” *J. Bacteriol.*, vol. 174, no. 5, pp. 1690–1693, 1992.

[46] M. Ikeda, M. Wachil, H. K. Jung, F. Ishino, and M. Matsuhashi, “Nucleotide sequence involving murG and murC in the mra gene cluster region of Escherichia coli,” *Nucleic Acids Res.*, vol. 18, no. 13, p. 4014, 1990.

[47] M. Ikeda, M., Wachi, M., Jung, H. K., Ishino, F., & Matsuhashi, “Homology among MurC, MurD, MurE and MurF proteins in Escherichia coli and that between E. coli MurG and a possible MurG protein in Bacillus subtilis,” *J. Gen. Appl. Microbiol.*, vol. 36, no. 3, pp. 179–187, 1990.

[48] D. M. Lecreulx *et al.*, “Organization of the murE-murG Region of Escherichia coli : Identification of the murD Gene Encoding the D-Glutamic-Acid-Adding Enzyme,” *J. Bacteriol.*, vol. 171, no. 11, pp. 6126–6134, 1989.

[49] E. E. Tao, J. S., & Ishiguro, “Nucleotide sequence of the murE gene of Escherichia coli,” *Can. J. Microbiol.*, vol. 35, no. 11, pp. 1051–1054, 1989.

[50] I. N. Maruyama, A. H. Yamamoto, and Y. Hirota, “Determination of Gene Products and Coding Regions from the murE-murF Region of Escherichia coli,” *J. Bacteriol.*, vol. 170, no. 8, pp. 3786–3788, 1988.

[51] S. J. Kim, J. Chang, and M. Singh, “Peptidoglycan architecture of Gram-positive bacteria by solid-state NMR,” *Biochim. Biophys. Acta - Biomembr.*, vol. 1848, no. 1, pp. 350–362, 2015.

[52] M. Ikeda, M., Wachi, M., Jung, H. K., Ishino, F., & Matsuhashi, “The Escherichia coli mraY Gene Encoding UDP-N-AcetylmuramoylPentapeptide:Undecaprenyl-Phosphate Phospho-NAcetylmuramoyl-Pentapeptide Transferase,” *J. Bacteriol.*, vol. 173, no. 3, pp. 1021–1026, 1991.

[53] D. M. Lecreulx, L. Texier, M. Rousseau, and J. V. A. N. Heijenoort, “The murG gene of Escherichia coli codes for the UDP-N-acetylglucosamine: N-acetylmuramyl-(pentapeptide) pyrophosphoryl-undecaprenol N-acetylglucosamine transferase involved in the membrane steps of peptidoglycan synthesis.,” *J. Bacteriol.*, vol. 173, no. 15, pp. 4625–4636, 1991.

[54] N. Ruiz, “Bioinformatics identification of MurJ (MviN) as the peptidoglycan lipid II flippase in Escherichia coli,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 105, no. 40, pp. 15553–15557, 2008.

[55] N. S. Gray *et al.*, “MurJ is the flippase of lipid-linked precursors for peptidoglycan biogenesis,” *Science (80-.).*, vol. 345, no. 6193, pp. 220–222, 2014.

[56] A. J. F. Egan, J. Errington, and W. Vollmer, “Regulation of peptidoglycan synthesis and remodelling,” *Nat. Rev. Microbiol.*, pp. 1–15, 2020.

[57] E. Sauvage, M. Terrak, J. A. Ayala, and P. Charlier, “The penicillin-binding proteins : structure and role in peptidoglycan biosynthesis,” *FEMS Microbiol. Rev.*, vol. 32, no. 2, pp. 234–258, 2008.

[58] M. Sung, Y. Lai, C. Huang, L. Chou, H. Shih, and W. Cheng, “Crystal structure of the membrane-bound bifunctional transglycosylase PBP1b from Escherichia coli,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 106, no. 22, pp. 8824–8829, 2009.

[59] L. Potluri *et al.*, “Septal and lateral wall localization of PBP5, the major D,D-carboxypeptidase of Escherichia coli, requires substrate recognition and membrane attachment,” *Mol. Microbiol.*, vol. 77, no. 2, pp. 300–323, 2010.

[60] M. Sjodt *et al.*, “Structural coordination of polymerization and crosslinking by a SEDS–bPBP peptidoglycan synthase complex,” *Nat. Microbiol.*, vol. 5, no. 6, pp. 813–820, 2020.

[61] S. Magnet, L. Dubost, A. Marie, M. Arthur, and L. Gutmann, “Identification of the L,D-transpeptidases for peptidoglycan cross-linking in Escherichia coli,” *J. Bacteriol.*, vol. 190, no. 13, pp. 4782–4785, 2008.

[62] V. Dubée, S. Triboulet, J. Mainardi, M. Ethève-quelquejeu, and L. Gutmann, “Inactivation of Mycobacterium tuberculosis L, D-transpeptidase LdtMt1 by carbapenems and cephalosporins,” *Antimicrob. Agents Chemother.*, vol. 56, no. 8, pp. 4189–4195, 2012.

[63] J. Peltier, P. Courtin, I. El Meouche, and L. Leme, “Clostridium difficile Has an Original Peptidoglycan Structure with a High Level of N -Acetylglucosamine Deacetylation and,” *J. Biol. Chem.*, vol. 286, no. 33, pp. 29053–29062, 2011.

[64] J. Mainardi, R. Legrand, M. Arthur, B. Schoot, J. Van Heijenoort, and L. Gutmann, “Novel Mechanism of β-Lactam Resistance Due to Bypass of DD-Transpeptidation in Enterococcus faecium,” *J. Biol. Chem.*, vol. 275, no. 22, pp. 16490–16496, 2000.

[65] T. A. Zandi, R. L. Marshburn, P. K. Stateler, and L. A. B. Basta, “Phylogenetic and Biochemical Analyses of Mycobacterial l,d-Transpeptidases Reveal a Distinct Enzyme Class That Is Preferentially Acylated by Meropenem,” *ACS Infect. Dis.*, 2019.

[66] T. A. Van Laar, T. Chen, T. You, and P. Leung, “Sublethal Concentrations of Carbapenems Alter Cell Morphology and Genomic Expression of Klebsiella pneumoniae Biofilms,” *Antimicrob. Agents Chemother.*, vol. 59, no. 3, pp. 1707–1717, 2015.

[67] S. Magnet *et al.*, “Specificity of L,D -Transpeptidases from Gram-positive Bacteria Producing Different Peptidoglycan Chemotypes,” *J. Biol. Chem.*, vol. 282, no. 18, pp. 13151–13159, 2007.

[68] H. S. Kim *et al.*, “The Cell Shape-determining Csd6 Protein from Helicobacter pylori Constitutes a New Family of L , D -Carboxypeptidase *,” *J. Biol. Chem.*, vol. 290, no. 41, pp. 25103–25117, 2015.

[69] S. Biarrotte-sorin *et al.*, “Crystal structure of a novel β-lactam-insensitive peptidoglycan transpeptidase,” *J. Mol. Biol.*, vol. 359, no. 3, pp. 533–538, 2006.

[70] G. Minasov *et al.*, “Structure of the LdcB LD-carboxypeptidase reveals the molecular basis of peptidoglycan recognition,” *Structure*, vol. 22, no. 7, pp. 949–960, 2014.

[71] R. Grangeon, J. R. Zupan, J. Anderson-furgeson, and P. C. Zambryski, “PopZ identifies the new pole, and PodJ identifies the old pole during polar growth in Agrobacterium tumefaciens,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 112, no. 37, pp. 11666–11671, 2015.

[72] T. A. Cameron, J. Anderson-furgeson, J. R. Zupan, J. J. Zik, and P. C. Zambryski, “Peptidoglycan synthesis machinery in Agrobacterium tumefaciens during unipolar growth and cell division,” *MBio*, vol. 5, no. 3, pp. 1–10, 2014.

[73] K. J. Kieser and E. J. Rubin, “How sisters grow apart: Mycobacterial growth and division,” *Nat. Rev. Microbiol.*, vol. 12, no. 8, pp. 550–562, 2014.

[74] M. K. Schoonmaker, W. R. Bishai, and G. Lamichhanea, “Nonclassical transpeptidases of Mycobacterium tuberculosis alter cell size, morphology, the cytosolic matrix, protein localization, virulence, and resistance to β-lactams,” *J. Bacteriol.*, vol. 196, no. 7, pp. 1394–1402, 2014.

[75] L. A. B. Basta *et al.*, “Loss of a functionally and structurally distinct LD-transpeptidase, LdtMt5, compromises cell wall integrity in mycobacterium tuberculosis,” *J. Biol. Chem.*, vol. 290, no. 42, pp. 25670–25685, 2015.

[76] E. M. Steiner, G. Schneider, and R. Schnell, “Binding and processing of β-lactam antibiotics by the transpeptidase LdtMt2 from Mycobacterium tuberculosis,” *FEBS J.*, vol. 284, no. 5, pp. 725–741, 2017.

[77] M. De Munnik *et al.*, “Targeting the Mycobacterium tuberculosis transpeptidase LdtMt2 with cysteine-reactive inhibitors including ebselen,” *Chem. Commun.*, vol. 55, no. 69, pp. 10214–10217, 2019.

[78] H. S. Kim *et al.*, “Structural basis for the inhibition of Mycobacterium tuberculosis L,D-transpeptidase by meropenem, a drug effective against extensively drug-resistant strains,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 69, no. 3, pp. 420–431, 2013.

[79] S. Correale, A. Ruggiero, R. Capparelli, E. Pedone, and R. Berisio, “Structures of free and inhibited forms of the L,D-transpeptidase LdtMt1 from Mycobacterium tuberculosis,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 69, no. 9, pp. 1697–1706, 2013.

[80] G. A. Libreros-Zúniga, C. Dos Santos Silva, R. Salgado Ferreira, and M. V. B. Dias, “Structural Basis for the Interaction and Processing of β-Lactam Antibiotics by l, d -Transpeptidase 3 (Ldt Mt3) from Mycobacterium tuberculosis,” *ACS Infect. Dis.*, vol. 5, no. 2, pp. 260–271, 2019.

[81] T. Geiger, M. Pazos, M. Lara-tejero, W. Vollmer, and J. E. Galán, “Peptidoglycan editing by a specific LD-transpeptidase controls the muramidase-dependent secretion of typhoid toxin,” *Nat. Microbiol.*, vol. 3, no. 11, pp. 1243–1254, 2018.

[82] J. Hugonnet *et al.*, “Factors essential for L,D-transpeptidasemediated peptidoglycan cross-linking and b-lactam resistance in Escherichia coli,” *Elife*, pp. 1–22, 2016.

[83] N. A. Caveney *et al.*, “Structural insight into YcbB-mediated beta-lactam resistance in Escherichia coli,” *Nat. Commun.*, vol. 10, no. 1, pp. 1–11, 2019.

[84] S. Magnet *et al.*, “Identification of the L,D -Transpeptidases Responsible for Attachment of the Braun Lipoprotein to Escherichia coli Peptidoglycan ,” *J. Bacteriol.*, vol. 189, no. 10, pp. 3927–3931, 2007.

[85] J. F. Asmar, A. T., & Collet, “Lpp , the Braun lipoprotein , turns 50 — major achievements and remaining issues,” *FEMS Microbiol. Lett.*, vol. 365, no. 18, p. fny199, 2018.

[86] Y. Hirota, H. Suzuki, Y. Nishimura, and S. Yasuda, “On the process of cellular division in Escherichia coli : A mutant of E . coli lacking a murein-lipoprotein,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 74, no. 4, pp. 1417–1420, 1977.

[87] D. W. Yem and H. C. Wu, “Physiological Characterization of an Escherichia coli Mutant Altered in the Structure of Murein Lipoprotein,” *J. Bacteriol.*, vol. 133, no. 3, pp. 1419–1426, 1978.

[88] D. Hoekstra, J. A. N. W. L. V. A. N. D. E. R. Laan, L. O. E. D. E. Leij, and B. Wit-, “Release of outer membrane fragments from normally growing Escherichia coli,” *Biochim. Biophys. Acta (BBA)-Biomembranes*, vol. 455, no. 3, pp. 889–899, 1976.

[89] A. N. Sanders and M. S. Pavelka, “Phenotypic analysis of Eschericia coli mutants lacking L, D-transpeptidases,” *Microbiology*, vol. 159, no. 9, pp. 1842–1852, 2013.

[90] S. Cho *et al.*, “Detecting envelope stress by monitoring β-barrel assembly,” *Cell*, vol. 159, no. 7, pp. 1652–1664, 2014.

[91] K. Braun, V., & Rehn, “Chemical characterization, spatial distribution and function of a lipoprotein (murein-lipoprotein) of the E. coli cell wall: the specific effect of trypsin on the membrane structure,” *Eur. J. Biochem.*, vol. 10, no. 3, pp. 426–438, 1969.

[92] W. Shu, J. Liu, H. Ji, and M. Lu, “Core structure of the outer membrane lipoprotein from Escherichia coli at 1.9 Å resolution,” *J. Mol. Biol.*, vol. 299, no. 4, pp. 1101–1112, 2000.

[93] J. Liu, W. Cao, and M. Lu, “Core side-chain packing and backbone conformation in Lpp-56 coiled-coil mutants,” *J. Mol. Biol.*, vol. 318, no. 3, pp. 877–888, 2002.

[94] A. Vermassen, S. Leroy, R. Talon, C. Provot, M. Popowska, and M. Desvaux, “Cell wall hydrolases in bacteria: insight on the diversity of cell wall amidases, glycosidases and peptidases toward peptidoglycan,” *Front. Microbiol.*, vol. 10, p. 331, 2019.

[95] W. Vollmer, B. Joris, P. Charlier, and S. Foster, “Bacterial peptidoglycan (murein) hydrolases,” *FEMS Microbiol. Rev.*, vol. 32, no. 2, pp. 259–286, 2008.

[96] J. V. Höltje, “From growth to autolysis: the murein hydrolases inEscherichia coli,” *Arch. Microbiol.*, vol. 164, no. 4, pp. 243–254, 1995.

[97] Y. Karamanos, “Endo-N-acetyl-β-D-glucosaminidases and their potential substrates: structure/function relationships,” *Res. Microbiol.*, vol. 148, no. 8, pp. 661–671, 1997.

[98] S. Litzinger, S. Fischer, P. Polzer, K. Diederichs, W. Welte, and C. Mayer, “Structural and kinetic analysis of Bacillus subtilis N-acetylglucosaminidase reveals a unique Asp-His dyad mechanism,” *J. Biol. Chem.*, vol. 285, no. 46, pp. 35675–35684, 2010.

[99] X. Bai *et al.*, “Structure of pneumococcal peptidoglycan hydrolase LytB reveals insights into the bacterial cell wall remodeling and pathogenesis,” *J. Biol. Chem.*, vol. 289, no. 34, pp. 23403–23416, 2014.

[100] F. A. Herlihey and A. J. Clarke, “Controlling autolysis during flagella insertion in Gram-negative bacteria,” *Protein Rev.*, pp. 41–56, 2016.

[101] A. Bateman and N. D. Rawlings, “The CHAP domain: a large family of amidases including GSP amidase and peptidoglycan hydrolases,” *TRENDS Biochem. Sci.*, vol. 28, no. 5, pp. 234–237, 2003.

[102] D. J. Rigden, M. J. Jedrzejas, and M. Y. Galperin, “Amidase domains from bacterial and phage autolysins define a family of γ-d,l-glutamate-specific amidohydrolases,” *TRENDS Biochem. Sci.*, vol. 28, no. 5, pp. 230–234, 2003.

[103] J. Gu *et al.*, “Structural and biochemical characterization reveals LysGH15 as an unprecedented ‘EF-hand-like’ calcium-binding phage lysin,” *PLoS Pathog.*, vol. 10, no. 5, p. e1004109, 2014.

[104] G. Buist, A. Steen, J. Kok, and O. P. Kuipers, “LysM, a widely distributed protein motif for binding to (peptido) glycans,” *Mol. Microbiol.*, vol. 68, no. 4, pp. 838–847, 2008.

[105] M. Rocaboy *et al.*, “The crystal structure of the cell division amidase AmiC reveals the fold of the AMIN domain, a new peptidoglycan binding domain,” *Mol. Microbiol.*, vol. 90, no. 2, pp. 267–277, 2013.

[106] S. C. Becker *et al.*, “Lytic activity of the staphylolytic Twort phage endolysin CHAP domain is enhanced by the SH3b cell wall binding domain,” *FEMS Microbiol. Lett.*, vol. 362, no. 1, pp. 1–8, 2015.

[107] Q. Xu *et al.*, “Insights into Substrate Specificity of NlpC/P60 Cell Wall Hydrolases Containing Bacterial SH3 Domains,” *Am. Soc. Microbiol.*, vol. 6, no. 5, pp. 1–12, 2015.

[108] M. J. Sanz-Gaitero, M., Keary, R., Garcia-Doval, C., Coffey, A., & van Raaij, “Crystal structure of the lytic CHAP K domain of the endolysin LysK from Staphylococcus aureus bacteriophage K,” *Virol. J.*, vol. 11, no. 1, p. 133, 2014.

[109] S. S. Broendum, A. M. Buckle, and S. Mcgowan, “Catalytic diversity and cell wall binding repeats in the phage-encoded endolysins,” *Mol. Microbiol.*, vol. 110, no. 6, pp. 879–896, 2018.

[110] S. C. Becker, S. Dong, J. R. Baker, J. Foster-frey, D. G. Pritchard, and D. M. Donovan, “LysK CHAP endopeptidase domain is required for lysis of live staphylococcal cells,” *FEMS Microbiol. Lett.*, vol. 294, no. 1, pp. 52–60, 2009.

[111] S. Layec *et al.*, “The CHAP domain of Cse functions as an endopeptidase that acts at mature septa to promote Streptococcus thermophilus cell separation,” *Mol. Microbiol.*, vol. 71, no. 5, pp. 1205–1217, 2009.

[112] B. Kim *et al.*, “Enterococcus faecium secreted antigen A generates muropeptides to enhance host immunity and limit bacterial pathogenesis,” *Elife*, vol. 8, p. e45343, 2019.

[113] Q. Xu *et al.*, “Structure-Guided Functional Characterization of DUF1460 Reveals a Highly Specific NlpC/P60 Amidase Family,” *Structure*, vol. 22, no. 2, pp. 1799–1809, 2014.

[114] F. Squeglia, M. Moreira, A. Ruggiero, and R. Berisio, “The Cell Wall Hydrolytic NlpC/P60 Endopeptidases in Mycobacterial Cytokinesis: A Structural Perspective,” *Cells*, vol. 8, no. 6, p. 609, 2019.

[115] M. C. Duchêne *et al.*, “Distinct and specific role of NlpC/P60 endopeptidases LytA and LytB in cell elongation and division of lactobacillus plantarum,” *Front. Microbiol.*, vol. 10, no. APR, pp. 1–18, 2019.

[116] C. Lambert *et al.*, “Interrupting peptidoglycan deacetylation during Bdellovibrio predator-prey interaction prevents ultimate destruction of prey wall, liberating bacterial-ghosts,” *Sci. Rep.*, vol. 6, no. May, pp. 1–19, 2016.

[117] C. J. Harding *et al.*, “A lysozyme with altered substrate specificity facilitates prey cell exit by the periplasmic predator Bdellovibrio bacteriovorus,” *Nat. Commun.*, vol. 11, no. 1, 2020.

[118] R. Bahadur, P. K. Chodisetti, and M. Reddy, “Cleavage of Braun’s lipoprotein Lpp from the bacterial peptidoglycan by a paralog of l,d-transpeptidases, LdtF,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 118, no. 19, pp. 1–7, 2021.

[119] R. Hurme, K. D. Berndt, E. Namork, and M. Rhen, “DNA Binding Exerted by a Bacterial Gene Regulator with an Extensive Coiled-coil Domain,” *J. Biol. Chem.*, vol. 271, no. 21, pp. 12626–12631, 1996.

[120] S. Lobedanz, E. Bokma, M. F. Symmons, E. Koronakis, C. Hughes, and V. Koronakis, “A periplasmic coiled-coil interface underlying TolC recruitment and the assembly of bacterial drug efflux pumps,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 104, no. 11, pp. 4612–4617, 2007.

[121] M. G. Surette and J. B. Stock, “Role of a-Helical Coiled-coil Interactions in Receptor Dimerization, Signaling, and Adaptation during Bacterial Chemotaxis,” *J. Biol. Chem.*, vol. 271, no. 30, pp. 17966–17973, 1996.

[122] S. C. Becker, J. Foster-frey, A. J. Stodola, D. Anacker, and D. M. Donovan, “Differentially conserved staphylococcal SH3b_5 cell wall binding domains confer increased staphylolytic and streptolytic activity to a streptococcal prophage endolysin domain,” *Gene*, vol. 443, no. 1–2, pp. 32–41, 2009.

[123] Y. Shen *et al.*, “Molecular basis for recognition of Listeria cell wall teichoic acid by the pseudo-symmetric SH3b-like repeats of a bacteriophage endolysin,” *bioRxiv*, 2020.

[124] L. S. Gonzalez-Delgado *et al.*, “Two-site recognition of Staphylococcus aureus peptidoglycan by lysostaphin SH3b,” *Nat. Chem. Biol.*, vol. 16, no. 1, pp. 24–30, 2020.

[125] E. J. Banks *et al.*, “Asymmetric peptidoglycan editing generates cell curvature in Bdellovibrio predatory bacteria,” *Nat. Commun.*, vol. 13, no. 1, pp. 1–15, 2022.

[126] W. Vollmer, D. Blanot, and M. A. De Pedro, “Peptidoglycan structure and architecture,” *FEMS Microbiol. Rev.*, vol. 32, no. 2, pp. 149–167, 2008.

[127] H. N. Williams and H. Chen, “Environmental Regulation of the Distribution and Ecology of Bdellovibrio and Like Organisms,” *Front. Microbiol.*, vol. 11, pp. 1–19, 2020.

[128] Q. Xu *et al.*, “Structures of a bifunctional cell wall hydrolase CwlT containing a novel bacterial lysozyme and an NlpC/P60 DL-endopeptidase,” *J. Mol. Biol.*, vol. 426, no. 1, pp. 169–184, 2014.

[129] S. Al-khodor, C. T. Price, A. Kalia, and Y. A. Kwaik, “Functional diversity of ankyrin repeats in microbial proteins,” *Trends Microbiol.*, vol. 18, no. 3, pp. 132–139, 2009.

[130] V. Anantharaman and L. Aravind, “Evolutionary history, structural features and biochemical diversity of the NlpC/P60 superfamily of enzymes.,” *Genome Biol.*, vol. 4, no. 2, pp. 1–12, 2003.

[131] H. N. Williams *et al.*, “Halobacteriovorax, an underestimated predator on bacteria: potential impact relative to viruses on bacterial mortality,” *ISME J.*, vol. 10, no. 2, pp. 491–499, 2015.

[132] F. Van Den Ent and J. Lo, “Short note RF cloning : A restriction-free method for inserting target genes into plasmids,” *J. Biochem. Biophys. Methods*, vol. 67, pp. 67–74, 2006.

[133] P. A. Karplus and K. Diederichs, “Assessing and maximizing data quality in macromolecular crystallography,” *Curr. Opin. Struct. Biol.*, vol. 34, pp. 60–68, 2015.

[134] E. F. Pettersen *et al.*, “UCSF Chimera - A visualization system for exploratory research and analysis,” *J. Comput. Chem.*, vol. 25, no. 13, pp. 1605–1612, 2004.

[135] E. F. Pettersen *et al.*, “UCSF ChimeraX: Structure visualization for researchers, educators, and developers,” *Protein Sci.*, vol. 30, no. 1, pp. 70–82, 2021.

[136] A. Allouche, “AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading,” *J. Comput. Chem.*, vol. 31, no. 2, pp. 455–461, 2010.

[137] S. J. De Vries, M. Van Dijk, and A. M. J. J. Bonvin, “The HADDOCK web server for data-driven biomolecular docking,” *Nat. Protoc.*, vol. 5, no. 5, pp. 883–897, 2010.

[138] G. Winter, “Xia2: An expert system for macromolecular crystallography data reduction,” *J. Appl. Crystallogr.*, vol. 43, no. 1, pp. 186–190, 2010.

[139] P. H. Zwart *et al.*, “Automated structure solution with the PHENIX suite,” *Struct. Proteomics*, vol. 426, pp. 419–435, 2008.

[140] P. Emsley and K. Cowtan, “Coot: Model-building tools for molecular graphics,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 60, no. 12 I, pp. 2126–2132, 2004.

[141] P. Skubák and N. S. Pannu, “Automatic protein structure solution from weak X-ray data,” *Nat. Commun.*, vol. 4, no. 1, pp. 1–6, 2013.

[142] M. Winkle *et al.*, “Dpaa detaches braun’s lipoprotein from peptidoglycan,” *MBio*, vol. 12, no. 3, 2021.

[143] F. Cava, M. A. De Pedro, H. Lam, B. M. Davis, and M. K. Waldor, “Distinct pathways for modification of the bacterial cell wall by non-canonical D-amino acids,” *EMBO J.*, vol. 30, no. 16, pp. 3442–3453, 2011.

[144] J. Bielnicki, Y. Devedjiev, U. Derewenda, Z. Dauter, A. Joachimiak, and Z. S. Derewenda, “B. subtilis ykuD protein at 2.0 Å resolution: Insights into the structure and function of a novel, ubiquitous family of bacterial enzymes,” *Proteins Struct. Funct. Genet.*, vol. 62, no. 1, pp. 144–151, 2006.

[145] J. Jumper *et al.*, “Highly accurate protein structure prediction with AlphaFold,” *Nature*, vol. 596, no. 7873, pp. 583–589, 2021.

[146] M. Bumann *et al.*, “Crystal structure of yeast Ypr118w, a methylthioribose-1-phosphate isomerase related to regulatory eIF2B subunits,” *J. Biol. Chem.*, vol. 279, no. 35, pp. 37087–37094, 2004.

[147] N. C. J. S. N. A. Caveney, A. Serapio-Palacios, S. E. Woodward, T. Bozorgmehr, G. Caballero, M. Vuckovic, W. Deng, B. B. Finlay, “Structural and Cellular Insights into the l,d-Transpeptidase YcbB as a Therapeutic Target in Citrobacter rodentium, Salmonella Typhimurium, and Salmonella Typhi Infections,” *Antimicrob. Agents Chemother.*, vol. 65, no. 2, pp. e01592-20, 2021.

[148] D. M. Prigozhin *et al.*, “Structural and genetic analyses of the mycobacterium tuberculosis protein kinase B sensor domain identify a potential ligand-binding site,” *J. Biol. Chem.*, vol. 291, no. 44, pp. 22961–22969, 2016.

[149] W. C. Lee, A. Jang, J. Y. Lee, and Y. Kim, “Structural implication of substrate binding by peptidoglycan remodeling enzyme MepS,” *Biochem. Biophys. Res. Commun.*, vol. 583, pp. 178–183, 2021.

[150] K. Min *et al.*, “Peptidoglycan reshaping by a noncanonical peptidase for helical cell shape in Campylobacter jejuni,” *Nat. Commun.*, vol. 11, no. 1, pp. 1–11, 2020.

[151] H. Sekiya, E. Tamai, J. Kawasaki, K. Murakami, and S. Kamitori, “Structural and biochemical characterizations of the novel autolysin Acd24020 from Clostridioides difficile and its full-function catalytic domain as a lytic enzyme,” *Mol. Microbiol.*, vol. 115, no. 4, pp. 684–698, 2021.

[152] C. S. H. Lin *et al.*, “Peptidoglycan binding by a pocket on the accessory NTF2-domain of Pgp2 directs helical cell shape of Campylobacter jejuni,” *J. Biol. Chem.*, vol. 296, no. 13, p. 100528, 2021.

[153] E. G. Ruby and S. C. Rittenberg, “Attachment of diaminopimelic acid to bdelloplast peptidoglycan during intraperiplasmic growth of Bdellovibrio bacteriovorus 109J,” *J. Bacteriol.*, vol. 158, no. 2, pp. 597–602, 1984.

[154] L. Hobley *et al.*, “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*, vol. 13, no. 1, pp. 1–13, 2012.

[155] N. Sangwan *et al.*, “Arsenic rich Himalayan hot spring metagenomics reveal genetically novel predator-prey genotypes,” *Environ. Microbiol. Rep.*, vol. 7, no. 6, pp. 812–823, 2015.

[156] C. H. Ngan, D. R. Hall, B. Zerbe, L. E. Grove, D. Kozakov, and S. Vajda, “FtSite: High accuracy detection of ligand binding sites on unbound protein structures,” *Bioinformatics*, vol. 28, no. 2, pp. 286–287, 2012.

[157] P. Mitkowski *et al.*, “Structural bases of peptidoglycan recognition by lysostaphin SH3b domain,” *Sci. Rep.*, vol. 9, no. 1, pp. 1–14, 2019.

[158] I. Sabala *et al.*, “Crystal structure of the antimicrobial peptidase lysostaphin from Staphylococcus simulans,” *FEBS J.*, vol. 281, no. 18, pp. 4112–4122, 2014.

[159] Q. Xu *et al.*, “Structural Basis of Murein Peptide Specificity of a γ-D-Glutamyl-L-Diamino Acid Endopeptidase,” *Structure*, vol. 17, no. 2, pp. 303–313, 2009.

[160] Q. Xu *et al.*, “Insights into substrate specificity of NlpC/P60 cell wall hydrolases containing bacterial SH3 domains,” *MBio*, vol. 6, no. 5, pp. 1–12, 2015.

[161] W. H. Cover, R. J. Martinez, and S. C. Rittenberg, “Permeability of the boundary layers of Bdellovibrio bacteriovorus 109J and its bdelloplasts to small hydrophilic molecules,” *J. Bacteriol.*, vol. 157, no. 2, pp. 385–390, 1984.

[162] P. Godessart *et al.*, “β-Barrels covalently link peptidoglycan and the outer membrane in the α-proteobacterium Brucella abortus,” *Nat. Microbiol.*, vol. 6, no. 1, pp. 27–33, 2021.

[163] C. Calvio, F. Romagnuolo, F. Vulcano, G. Speranza, and C. F. Morelli, “Evidences on the role of the lid loop of γ-glutamyltransferases (GGT) in substrate selection,” *Enzyme Microb. Technol.*, vol. 114, pp. 55–62, 2018.

[164] X. Hu *et al.*, “Probing the donor and acceptor substrate specificity of the γ-glutamyl transpeptidase,” *Biochemistry*, vol. 51, no. 6, pp. 1199–1212, 2012.

[165] S. Chou *et al.*, “Structure of a Peptidoglycan Amidase Effector Targeted to Gram-Negative Bacteria by the Type VI Secretion System,” *Cell Rep.*, vol. 1, no. 6, pp. 656–664, 2012.

[166] R. Priyadarshini, M. A. De Pedro, and K. D. Young, “Role of peptidoglycan amidases in the development and morphology of the division septum in Escherichia coli,” *J. Bacteriol.*, vol. 189, no. 14, pp. 5334–5347, 2007.

[167] T. Do *et al.*, “Staphylococcus aureus cell growth and division are regulated by an amidase that trims peptides from uncrosslinked peptidoglycan,” *Nat. Microbiol.*, vol. 5, no. 2, pp. 291–303, 2020.

[168] F. M. Büttner, S. Zoll, M. Nega, F. Götz, and T. Stehle, “Structure-function analysis of Staphylococcus aureus amidase reveals the determinants of peptidoglycan recognition and cleavage,” *J. Biol. Chem.*, vol. 289, no. 16, pp. 11083–11094, 2014.

[169] S. Martínez-Caballero *et al.*, “Reaction products and the X-ray structure of AmpDh2, a virulence determinant of Pseudomonas aeruginosa,” *J. Am. Chem. Soc.*, vol. 135, no. 28, pp. 10318–10321, 2013.

[170] M. Lee *et al.*, “Cell-wall remodeling by the zinc-protease AmpDh3 from pseudomonas aeruginosa,” *J. Am. Chem. Soc.*, vol. 135, no. 34, pp. 12604–12607, 2013.

[171] C. Carrasco-López *et al.*, “Crystal structures of bacterial peptidoglycan amidase AmpD and an unprecedented activation mechanism,” *J. Biol. Chem.*, vol. 286, no. 36, pp. 31714–31722, 2011.

[172] T. J. Smith, S. A. Blackman, and S. J. Foster, “Autolysins of Bacillus subtilis: Multiple enzymes with multiple functions,” *Microbiology*, vol. 146, no. 2, pp. 249–262, 2000.

[173] J. Cook, T. C. Baverstock, M. B. L. McAndrew, P. J. Stansfeld, D. I. Roper, and A. Crow, “Insights into bacterial cell division from a structure of EnvC bound to the FtsX periplasmic domain,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 117, no. 45, pp. 28355–28365, 2020.

[174] D. Nelson, R. Schuch, P. Chahales, S. Zhu, and V. A. Fischetti, “PlyC: A multimeric bacteriophage lysin,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 103, no. 28, pp. 10765–10770, 2006.

[175] S. McGowan *et al.*, “X-ray crystal structure of the streptococcal specific phage lysin PlyC,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 109, no. 31, pp. 12752–12757, 2012.

[176] F. Oechslin, J. Daraspe, M. Giddey, P. Moreillon, and G. Resch, “In vitro characterization of PlySK1249, a novel phage lysin, and assessment of its antibacterial activity in a mouse model of Streptococcus agalactiae bacteremia,” *Antimicrob. Agents Chemother.*, vol. 57, no. 12, pp. 6276–6283, 2013.

[177] F. Oechslin, C. Menzi, P. Moreillon, and G. Resch, “The multidomain architecture of a bacteriophage endolysin enables intramolecular synergism and regulation of bacterial lysis,” *J. Biol. Chem.*, vol. 296, p. 100639, 2021.

[178] Q. Xu *et al.*, “Structural analysis of papain-like NlpC/p60 superfamily enzymes with a circularly permuted topology reveals potential lipid binding sites,” *PLoS One*, vol. 6, no. 7, p. e22013, 2011.

[179] I. Karunker, O. Rotem, M. Dori-Bachash, E. Jurkevitch, and R. Sorek, “A Global Transcriptional Switch between the Attack and Growth Forms of Bdellovibrio bacteriovorus,” *PLoS One*, vol. 8, no. 4, 2013.

[180] M. Baek *et al.*, “Accurate prediction of protein structures and interactions using a three-track neural network,” *Science (80-.).*, vol. 373, no. 6557, pp. 871–876, 2021.

[181] D. B. Craig and A. A. Dombkowski, “Disulfide by Design 2.0: A web-based tool for disulfide engineering in proteins,” *BMC Bioinformatics*, vol. 14, no. 1, pp. 1–7, 2013.

[182] S. Zouhir, C. Contreras-Martel, D. Maragno Trindade, I. Attrée, A. Dessen, and P. Macheboeuf, *MagC is a NplC/P60-like member of the α-2-macroglobulin Mag complex of Pseudomonas aeruginosa that interacts with peptidoglycan*, vol. 595, no. 15. 2021.

[183] R. Bajaj, K. E. Bruce, A. L. Davidson, B. E. Rued, C. V Stauffacher, and M. E. Winkler, “Biochemical characterization of essential cell division proteins FtsX and FtsE that mediate peptidoglycan hydrolysis by PcsB in Streptococcus pneumoniae,” *Microbiol. Open*, vol. 5, no. 5, pp. 738–752, 2016.

[184] V. R. Correa *et al.*, “The bacterium Pantoea stewartii uses two different type III secretion systems to colonize its plant host and insect vector,” *Appl. Environ. Microbiol.*, vol. 78, no. 17, pp. 6327–6336, 2012.

[185] D. E. Fouts *et al.*, “Genomewide identification of Pseudomonas syringae pv. tomato DC3000 promoters controlled by the HrpL alternative sigma factor,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 99, no. 4, pp. 2275–2280, 2002.

[186] M. Billini, J. Biboy, J. Kühn, W. Vollmer, and M. Thanbichler, “A specialized MreB-dependent cell wall biosynthetic complex mediates the formation of stalk-specific peptidoglycan in Caulobacter crescentus,” *PLOS Genet.*, vol. 15, no. 2, p. e1007897, Feb. 2019.

[187] I. Robert-Genthon, M., Casabona, M. G., Neves, D., Couté, Y., Cicéron, F., Elsen, S., ... & Attrée, “Unique Features of a Pseudomonas aeruginosa α2-Macroglobulin Homolog,” *MBio*, vol. 4, no. 4, pp. e00309-13, 2013.

[188] S. Zouhir *et al.*, “Assembly of an atypical α-macroglobulin complex from Pseudomonas aeruginosa,” *Sci. Rep.*, vol. 8, no. 1, pp. 1–14, 2018.

[189] T. I. Zarembinski *et al.*, “Structure-based assignment of the biochemical function of a hypothetical protein: A test case of structural genomics,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 95, no. 26, pp. 15189–15193, 1998.

[190] A. Ebihara, M. Yao, R. Masui, I. Tanaka, S. Yokoyama, and S. Kuramitsu, “ Crystal structure of hypothetical protein TTHB192 from Thermus thermophilus HB8 reveals a new protein family with an RNA recognition motif-like domain ,” *Protein Sci.*, vol. 15, no. 6, pp. 1494–1499, 2006.

[191] L. T. Alexander *et al.*, “Target highlights in CASP14: Analysis of models by structure providers,” *Proteins Struct. Funct. Bioinforma.*, vol. 89, no. 12, pp. 1647–1672, 2021.

[192] D. Desveaux, J. Allard, N. Brisson, and J. Sygusch, “A new family of plant transcription factors displays a novel ssDNA-binding surface,” *Nat. Struct. Biol.*, vol. 9, no. 7, pp. 512–517, 2002.

[193] R. J. Sessler and N. Noy, “A ligand-activated nuclear localization signal in cellular retinoic acid binding protein-II,” *Mol. Cell*, vol. 18, no. 3, pp. 343–353, 2005.

[194] C. Millán *et al.*, “Assessing the utility of CASP14 models for molecular replacement,” *Proteins Struct. Funct. Bioinforma.*, vol. 89, no. 12, pp. 1752–1769, 2021.

[195] D. Zhang, R. F. de Souza, V. Anantharaman, L. M. Iyer, and L. Aravind, “Polymorphic toxin systems: Comprehensive characterization of trafficking modes, processing, mechanisms of action, immunity and ecology using comparative genomics,” *Biol. Direct*, vol. 7, 2012.

[196] A. Rocker, M. Peschke, T. Kittilä, R. Sakson, C. Brieke, and A. Meinhart, “The ng-ζ1 toxin of the gonococcal epsilon/zeta toxin/antitoxin system drains precursors for cell wall synthesis,” *Nat. Commun.*, vol. 9, no. 1, pp. 1–11, 2018.

[197] A. Dashiff, R. A. Junka, M. Libera, and D. E. Kadouri, “Predation of human pathogens by the predatory bacteria Micavibrio aeruginosavorus and Bdellovibrio bacteriovorus,” *J. Appl. Microbiol.*, vol. 110, no. 2, pp. 431–444, 2011.

[198] S. Dharani, D. H. Kim, R. M. Q. Shanks, Y. Doi, and D. E. Kadouri, “Susceptibility of colistin-resistant pathogens to predatory bacteria,” *Res. Microbiol.*, vol. 169, no. 1, pp. 52–55, 2018.

[199] P. M. Fratamico and R. C. Whiting, “Ability of bdellovibrio bacteriovorus 109j to lyse gram-negative food-borne pathogenic and spoilage bacteria,” *J. Food Prot.*, vol. 58, no. 2, pp. 160–164, 1995.

[200] P. M. Fratamico and P. H. Cooke, “Isolation of Bdellovibrios that prey on Escherichia coli O157:H7 and Salmonella species and application for removal of prey from stainless steel surfaces,” *J. Food Saf.*, vol. 16, no. 2, pp. 161–173, 1996.

[201] D. Ottaviani *et al.*, “Bdellovibrio bacteriovorus to control Escherichia coli on meat matrices,” *Int. J. Food Sci. Technol.*, vol. 55, no. 3, pp. 988–994, 2020.

[202] R. M. Q. Shanks *et al.*, “An Eye to a Kill: Using Predatory Bacteria to Control Gram-Negative Pathogens Associated with Ocular Infections,” *PLoS One*, vol. 8, no. 6, pp. 1–7, 2013.

[203] V. Iebba *et al.*, “Bdellovibrio bacteriovorus directly attacks Pseudomonas aeruginosa and Staphylococcus aureus cystic fibrosis isolates,” *Front. Microbiol.*, vol. 5, no. JUN, pp. 1–9, 2014.

[204] S. W. BOARD OF REGENTS, THE UNIVERSITY OF TEXAS SYSTEM, PICKERING, Andrew M.; KIM, “WO2021108195A1-BDELLOVIBRIO TREATMENT FOR AMYOTROPHIC LATERAL SCLEROSIS.” 2021.

[205] E. Bratanis, T. Andersson, R. Lood, and E. Bukowska-Faniband, “Biotechnological Potential of Bdellovibrio and Like Organisms and Their Secreted Enzymes,” *Front. Microbiol.*, vol. 11, no. April, pp. 1–14, 2020.

[206] R. Bratanis, E., Molina, H., Naegeli, A., Collin, M., & Lood, “BspK, a Serine Protease from the Predatory Bacterium Bdellovibrio bacteriovorus with Utility for Analysis of Therapeutic Antibodies,” *Appl. Environ. Microbiol.*, vol. 83, no. 4, pp. e03037-16, 2017.

[207] E. Bratanis and R. Lood, “A novel broad-spectrum elastase-like serine protease from the predatory bacterium Bdellovibrio bacteriovorus facilitates elucidation of site-specific IgA glycosylation pattern,” *Front. Microbiol.*, vol. 10, p. 971, 2019.

[208] A. K. Monnappa, M. Dwidar, J. K. Seo, J. H. Hur, and R. J. Mitchell, “Bdellovibrio bacteriovorus inhibits staphylococcus aureus biofilm formation and invasion into human epithelial cells,” *Sci. Rep.*, vol. 4, no. 1, pp. 1–8, 2014.

[209] V. Martínez, C. Herencias, E. Jurkevitch, and M. A. Prieto, “Engineering a predatory bacterium as a proficient killer agent for intracellular bio-products recovery: The case of the polyhydroxyalkanoates,” *Sci. Rep.*, vol. 6, no. 1, pp. 1–12, 2016.

[210] Ulrich GLEBE, Xiaolin Dai, Alexander BÖKER, Martin Schmidt, Ulrich Schwaneberg, Felix Jakob, Diana MATE, Tayebeh Mirzaei, GARAKANI, “WO2019011922A1-Method for producing block polymers by means of linking blocks by a transpeptidase, and block polymers obtained by transpeptidase linking.” 2019.

[211] "US2013336954A1-Staphylococcal Phage2638A Endolysin Amidase Domain Is Lytic for Staphylococcus aureus.” 2013.

[212] DONOVAN DAVID M; BECKER STEPHEN C and US AGRICULTURE, “US2010158886A1- Triple acting antimicrobials that are refractory to resistance development.” 2010.

[213] M. NOVOZYMES A/S; ZHANG, “WO2020173473A1- POLYPEPTIDES WITH CHAP DOMAIN AND THEIR USE FOR TREATING SLUDGE.” 2020.

APPENDIX I. ELECTRON DENSITY MAPS

A1. Bd3741 (group 2)

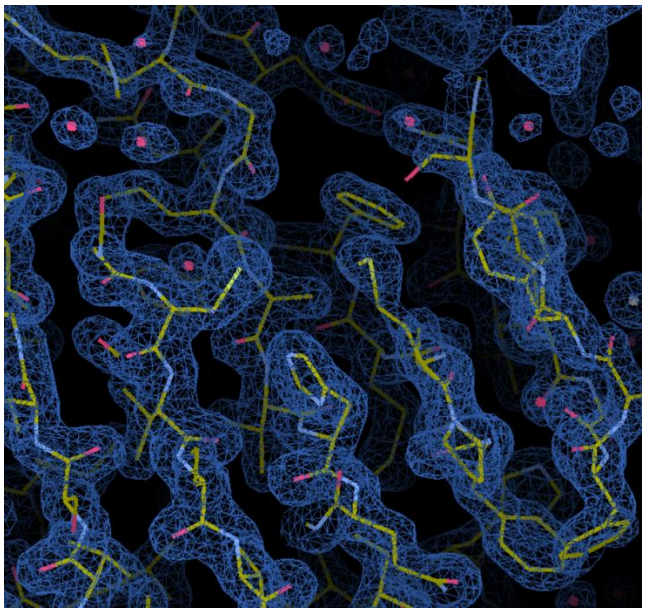


Figure A1: Bd3741 electron density map. Native electron density map ($2F_0$-F_0, blue mesh) of Bd3741 contoured at σ= 1.50. Protein model is depicted in stick representation.

A2. Bd0566 (group 4)

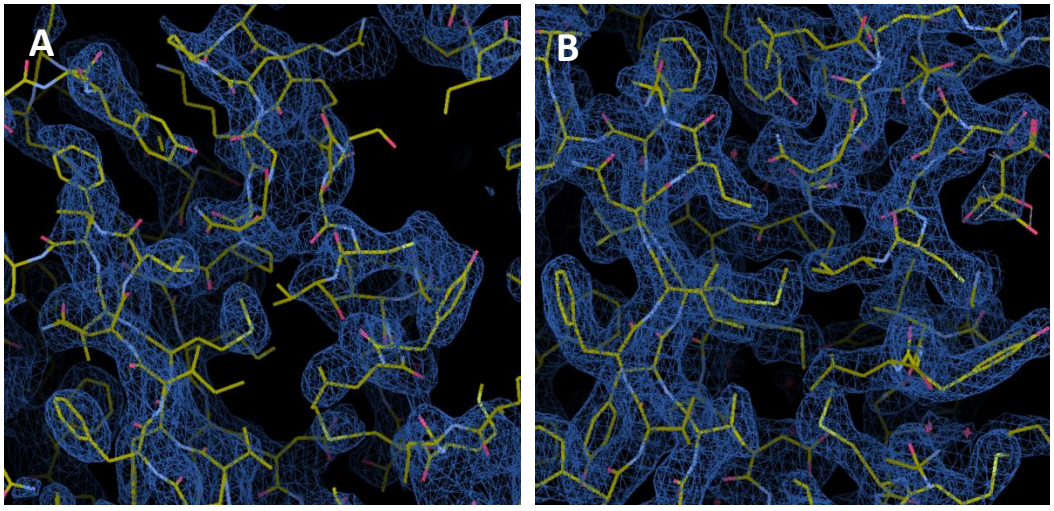


Figure A2: Bd0566 electron density map. Native electron density map ($2F_0$-F_0, blue mesh) of Bd3741 contoured at σ= 1.50. Protein model is depicted in stick representation. **A)** Density map corresponding to the 2.45 Å Bd0566 dataset. **B)** Density map corresponding to the 2.22 Å Bd0566 dataset.

A3. Bd3334 (group 5, soluble)

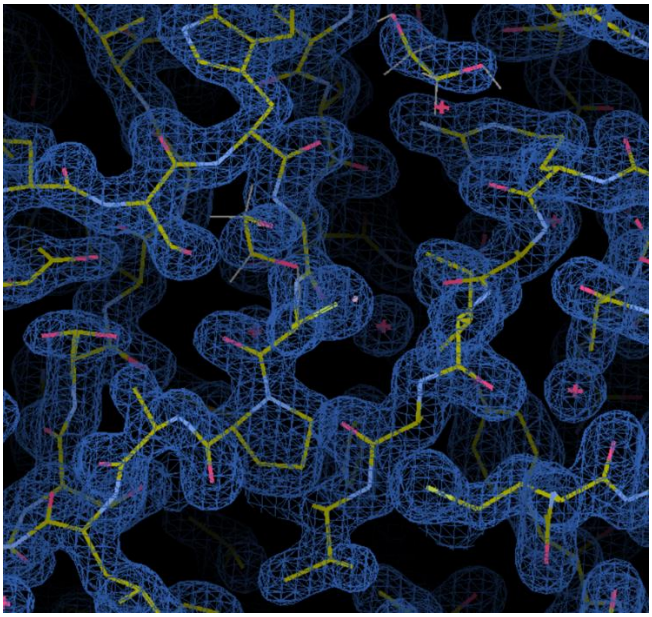


Figure A3: Bd3334 electron density map. Native electron density map ($2F_O$-F_O, blue mesh) of Bd3334 contoured at σ= 1.50. Protein model is depicted in stick representation.

A4. Bd1075 (group 6)

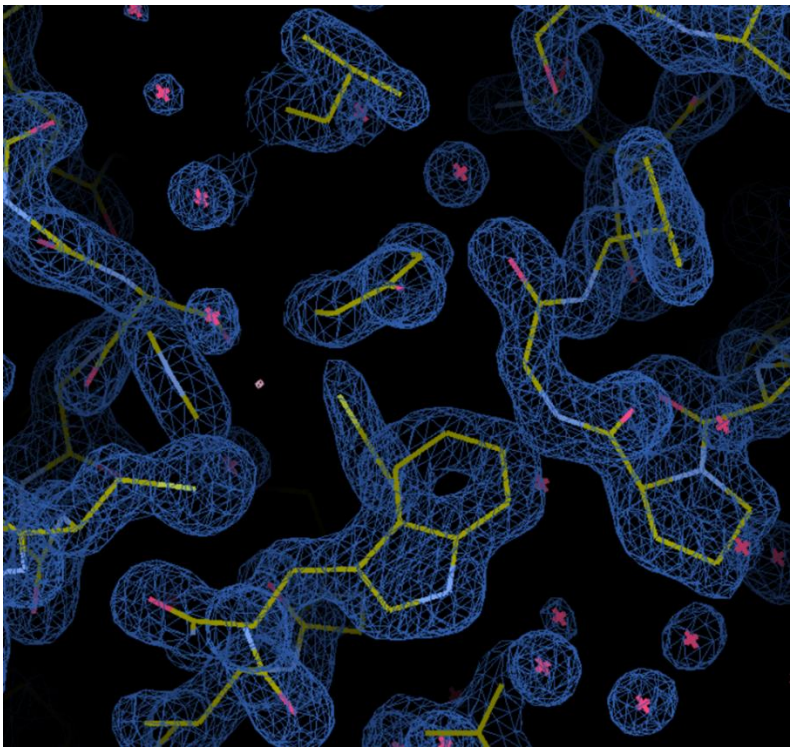


Figure A4: Bd1075 electron density map. Native electron density map ($2F_O$-F_O, blue mesh) of Bd1075 contoured at σ= 1.50. Protein model is depicted in stick representation.

A5. Bd1177 C-terminal domain

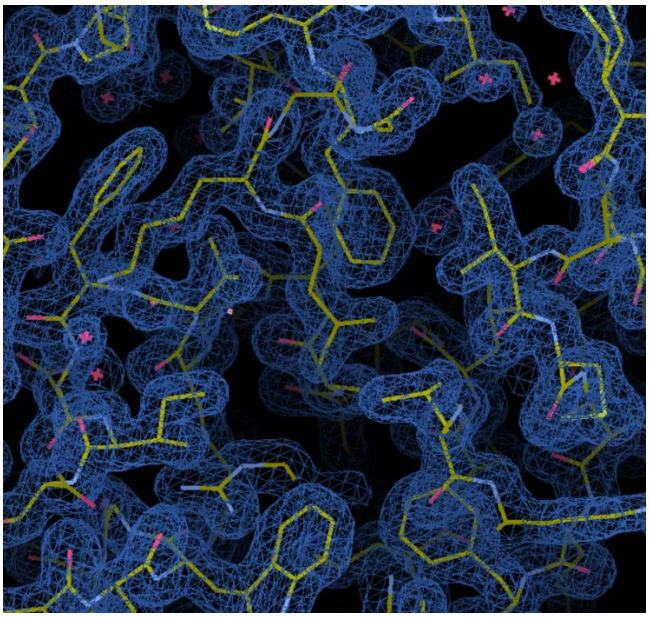


Figure A5: Bd1177 C-terminal domain electron density map. Native electron density map ($2F_O$-F_O, blue mesh) of Bd1177 C-terminal domain contoured at σ= 1.50. Protein model is depicted in stick representation.

A6. Bd1177 C86A/Y124C/Y508C

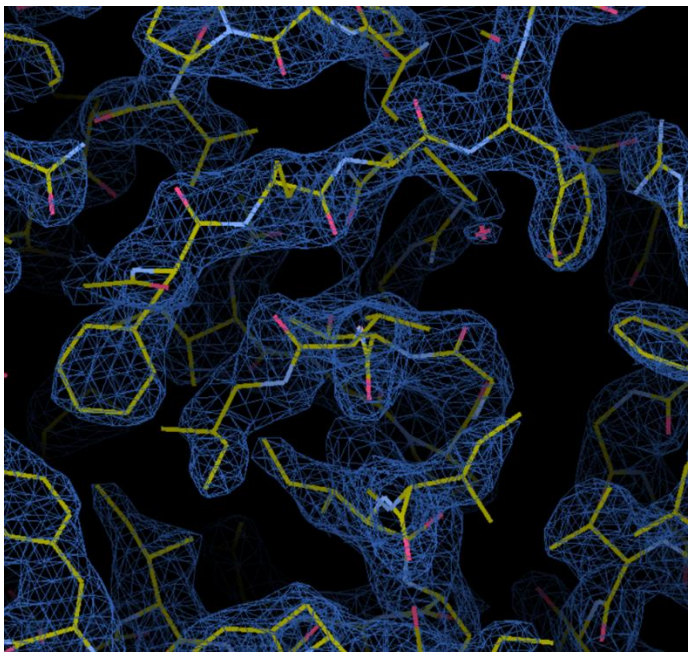


Figure A6: Bd1177 disulphide-locked mutant electron density map. Native electron density map ($2F_O$-F_O, blue mesh) of Bd1177 triple mutant contoured at σ= 1.50. Protein model is depicted in stick representation.

A7. Bd0675

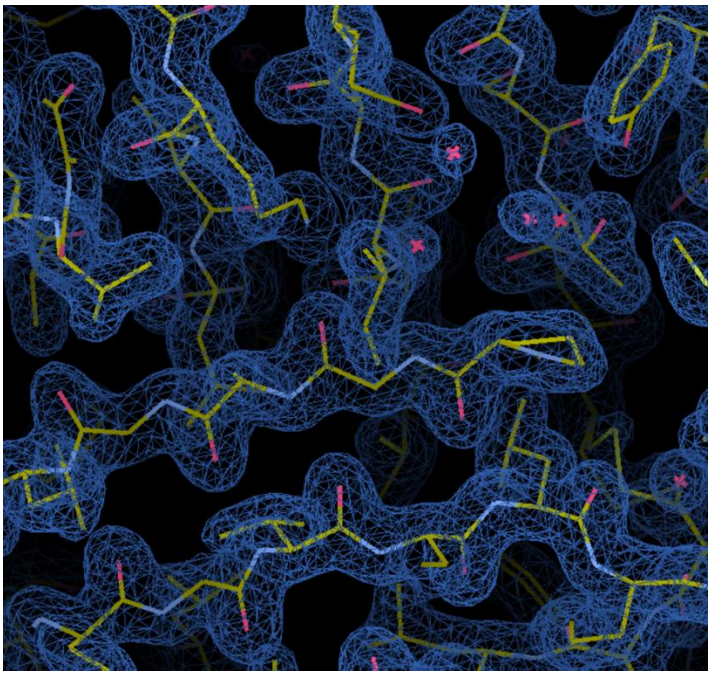


Figure A7: Bd0675 electron density map. Native electron density map ($2F_0$-F_0, blue mesh) of Bd0675 contoured at σ= 1.50. Protein model is depicted in stick representation.

APPENDIX II. pLDDT PLOTS FOR AlphaFold2 MODELS

A8. ErfK

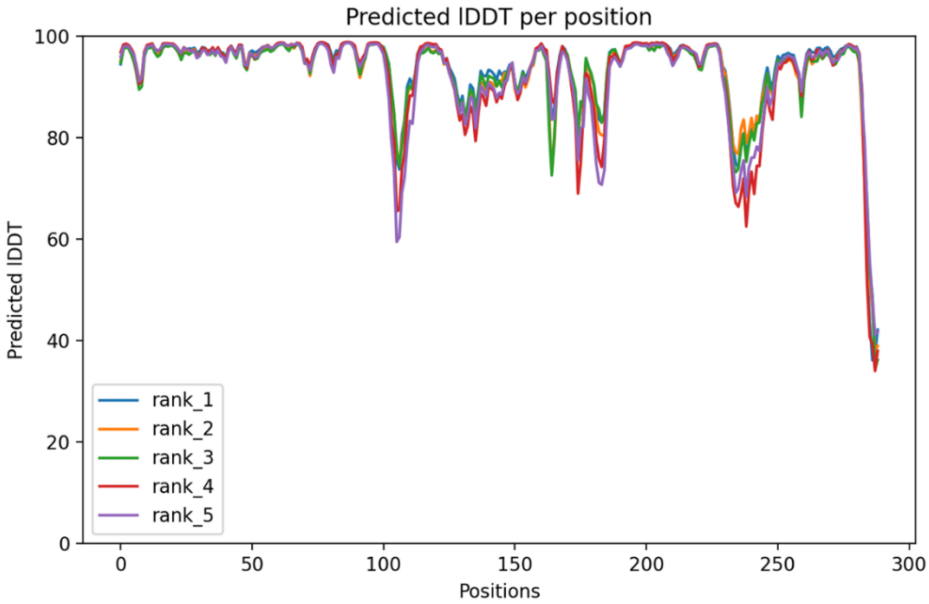


Figure A8: ErfK pLDDT plot. Percentage of Local Distance Difference Test (pLDDT) of ErfK AlphaFold2 models (rank 1 to 5) against amino acid position of the mature protein.

A9. Bd1763 (group 1)

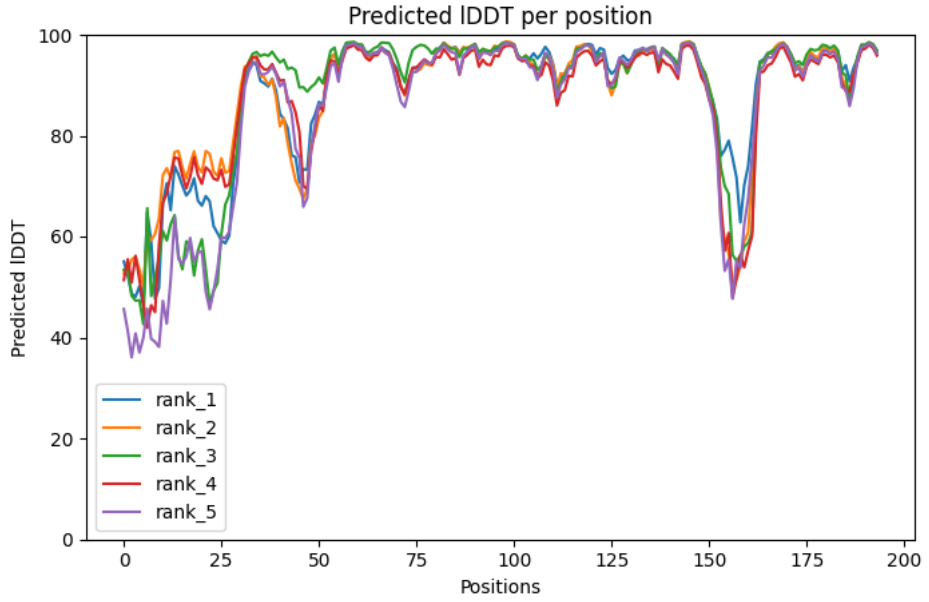


Figure A9: Bd1763 (group 1) pLDDT plot. Percentage of Local Distance Difference Test (pLDDT) of Bd1763 AlphaFold2 models (rank 1 to 5) against amino acid position of the mature protein.

A10. Bd2061 (group 1)

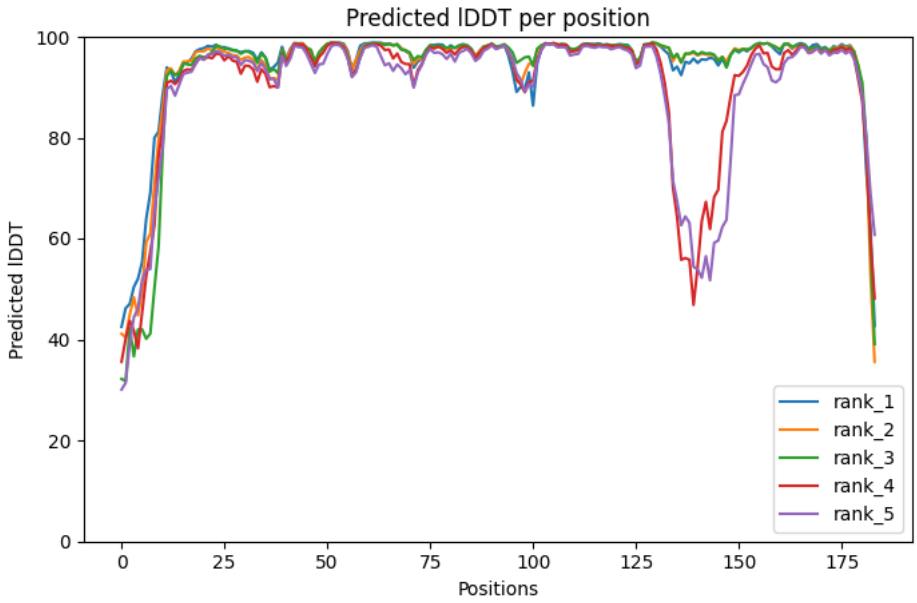


Figure A10: Bd2061 (group 1) pLDDT plot. Percentage of Local Distance Difference Test (pLDDT) of Bd2061 AlphaFold2 models (rank 1 to 5) against amino acid position of the mature protein.

A11. Bd3012 (group 2)

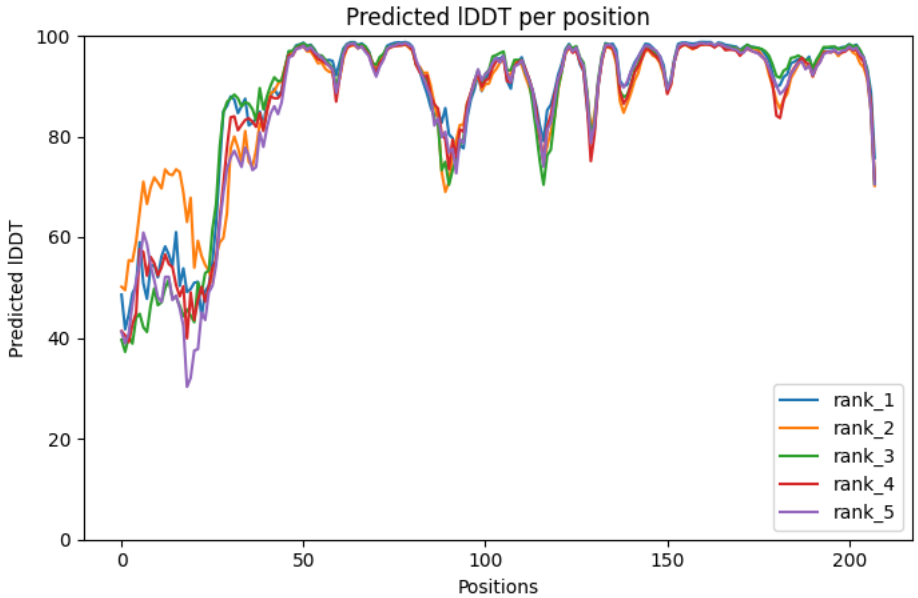


Figure A11: Bd3012 (group 2) pLDDT plot. Percentage of Local Distance Difference Test (pLDDT) of Bd3012 AlphaFold2 models (rank 1 to 5) against amino acid position of the mature protein.

A12. Bd3176 (group 2)

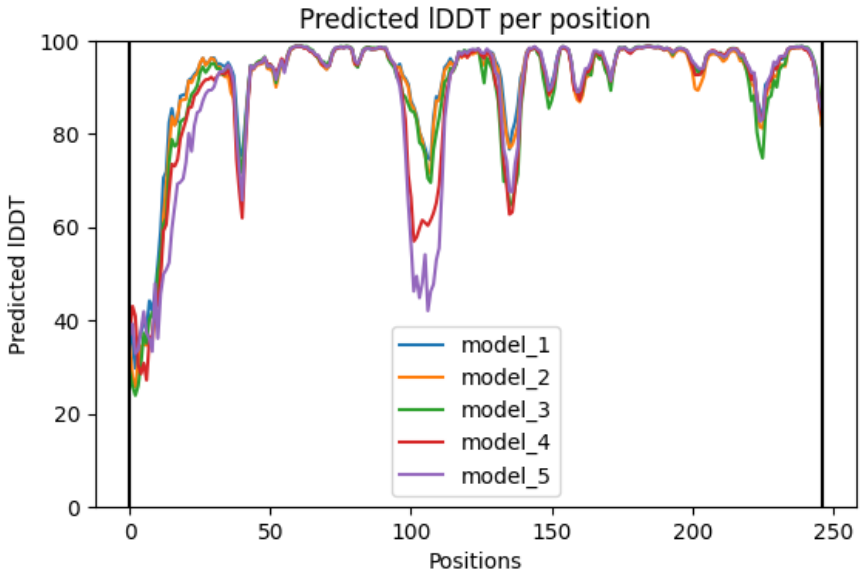


Figure A12: Bd3176 (group 2) pLDDT plot. Percentage of Local Distance Difference Test (pLDDT) of Bd3176 AlphaFold2 models (rank 1 to 5) against amino acid position of the mature protein.

A13. Bd3639 (group 2)

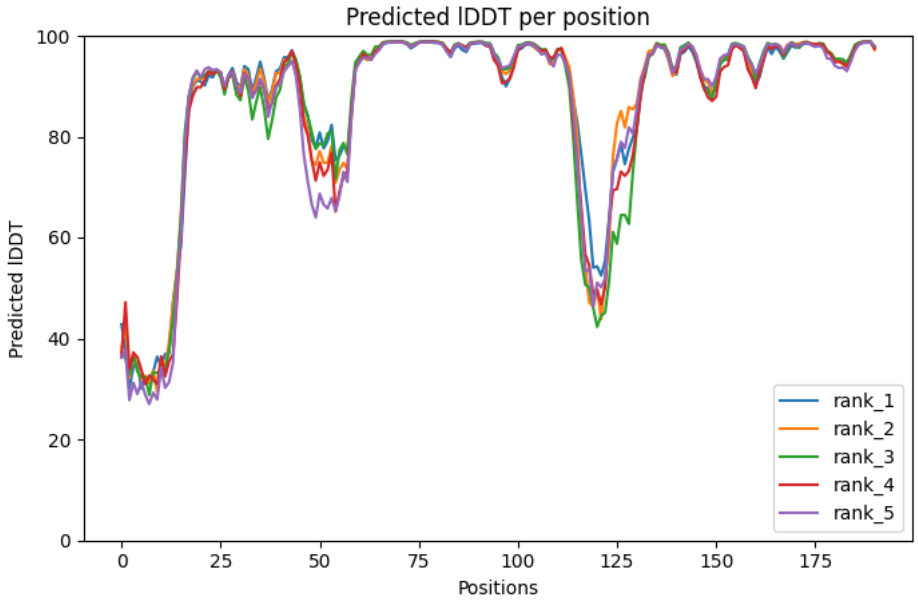


Figure A13: Bd3639 (group 2) pLDDT plot. Percentage of Local Distance Difference Test (pLDDT) of Bd3639 AlphaFold2 models (rank 1 to 5) against amino acid position of the mature protein.

A14. Bd1176 (group 3)

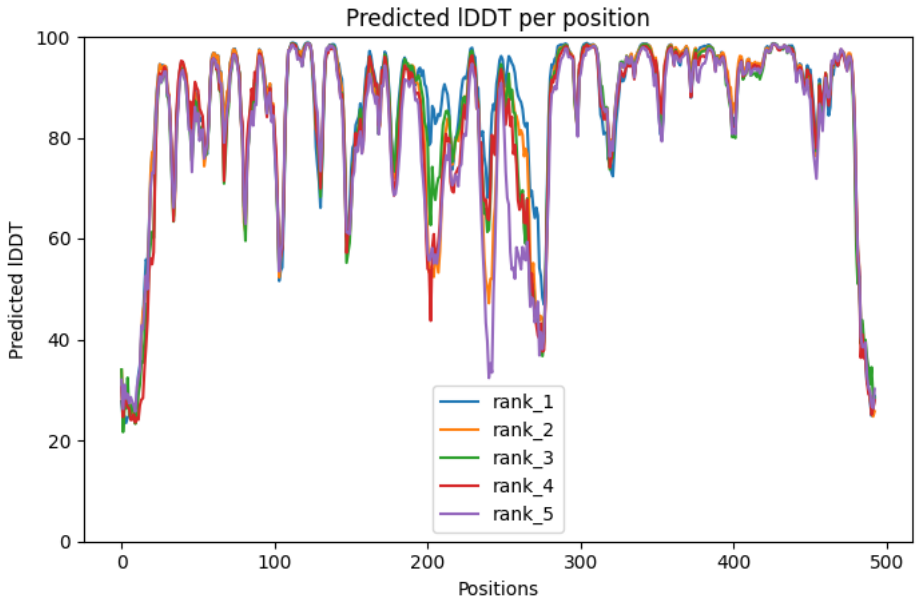


Figure A14: Bd1176 (group 3) pLDDT plot. Percentage of Local Distance Difference Test (pLDDT) of Bd1176 AlphaFold2 models (rank 1 to 5) against amino acid position of the mature protein.

A15. Bd1181 (group 4)

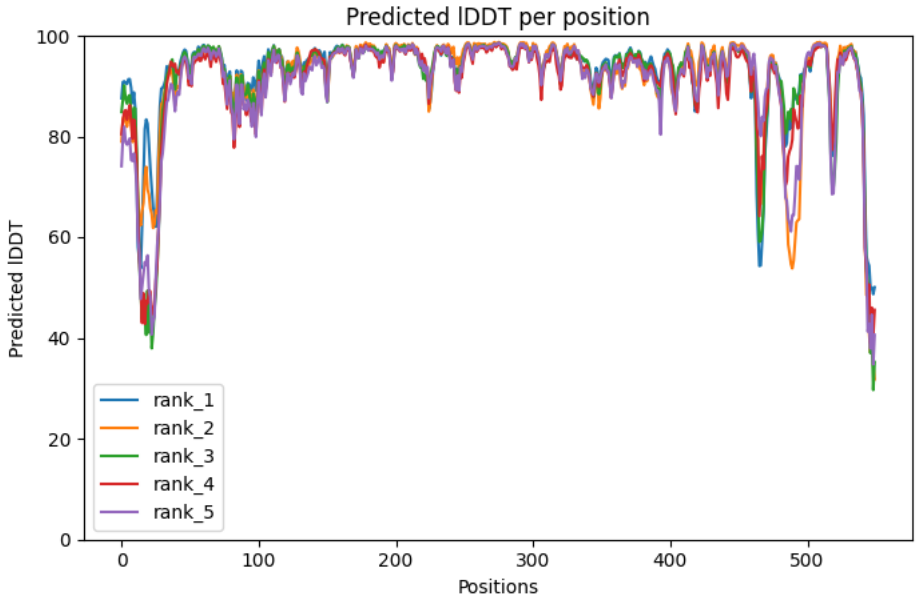


Figure A15: Bd1181 (group 4) pLDDT plot. Percentage of Local Distance Difference Test (pLDDT) of Bd1181 AlphaFold2 models (rank 1 to 5) against amino acid position of the mature protein.

A16. Bd1358 (group 4)

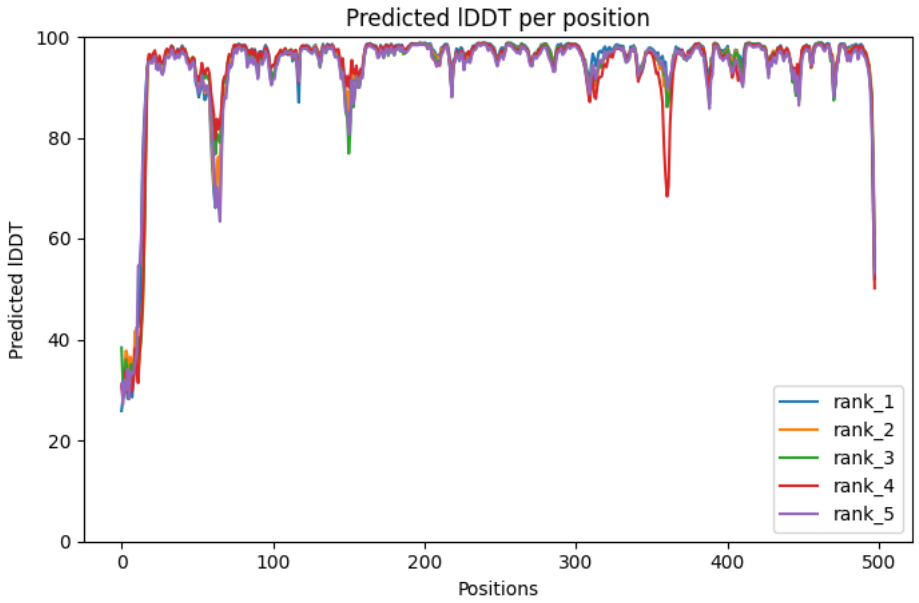


Figure A16: Bd1358 (group 4) pLDDT plot. Percentage of Local Distance Difference Test (pLDDT) of Bd1358 AlphaFold2 models (rank 1 to 5) against amino acid position of the mature protein.

A17. Bd2908 (group 4)

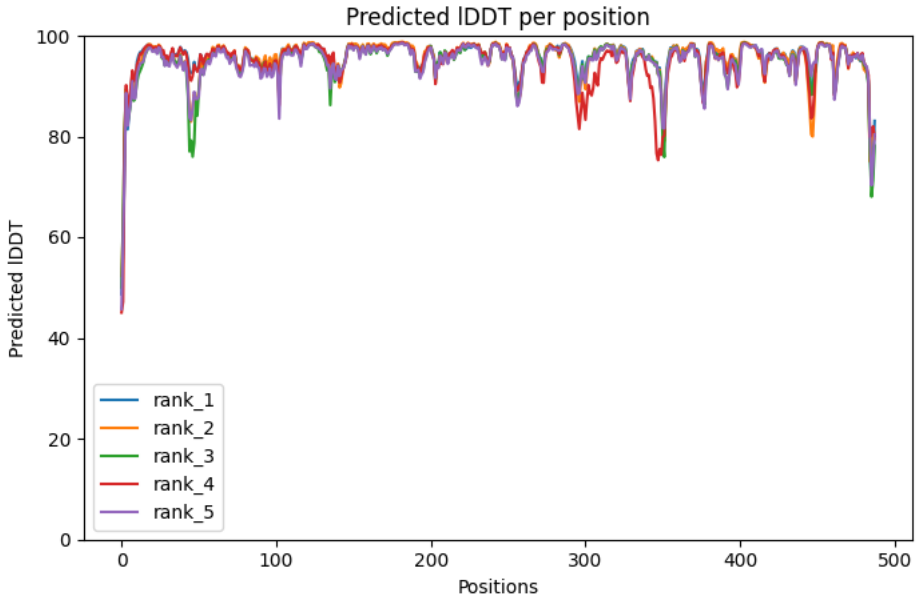


Figure A17: Bd2908 (group 4) pLDDT plot. Percentage of Local Distance Difference Test (pLDDT) of Bd2908 AlphaFold2 models (rank 1 to 5) against amino acid position of the mature protein.

A18. Bd0782 (group 5, lipobox SP)

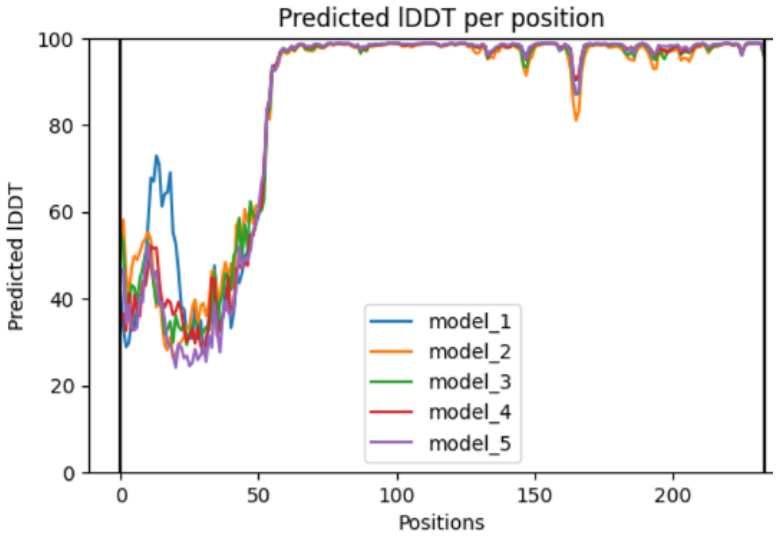


Figure A18: Bd0782 (group 5, lipobox SP) pLDDT plot. Percentage of Local Distance Difference Test (pLDDT) of Bd0782 AlphaFold2 models (rank 1 to 5) against amino acid position of the mature protein.

A19. Bd2940 (group 5, soluble)

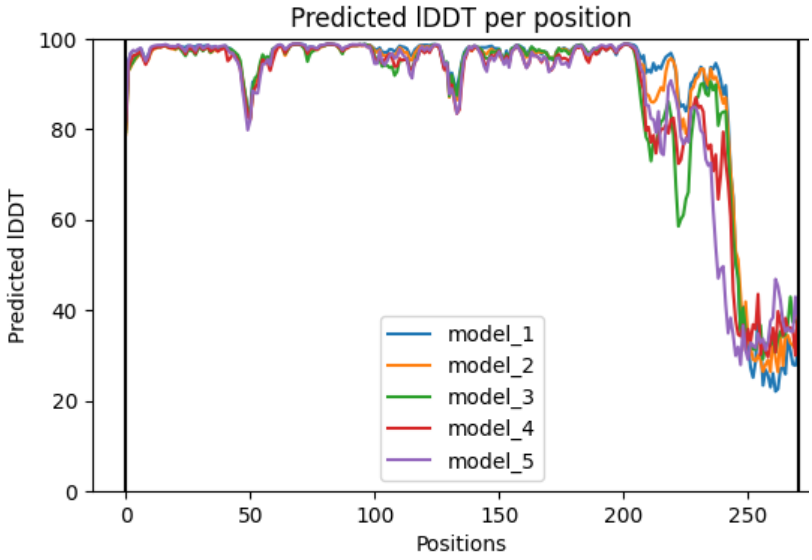


Figure A19: Bd2940 (group 5, soluble) pLDDT plot. Percentage of Local Distance Difference Test (pLDDT) of Bd2940 AlphaFold2 models (rank 1 to 5) against amino acid position of the mature protein.

A20. Bd3093 (group 5, soluble)

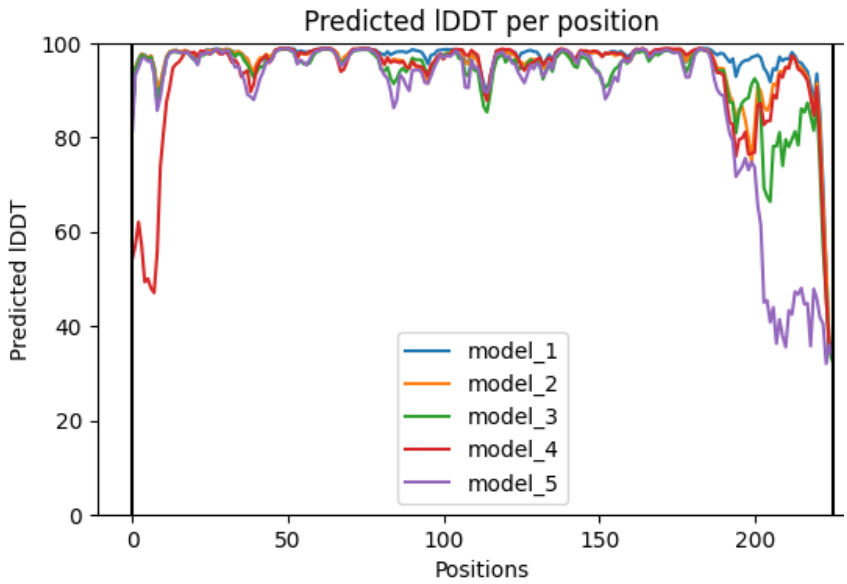


Figure A20: Bd3093 (group 5, soluble) pLDDT plot. Percentage of Local Distance Difference Test (pLDDT) of Bd3093 AlphaFold2 models (rank 1 to 5) against amino acid position of the mature protein.

APPENDIX III. RELATED PUBLICATIONS

This thesis work has contributed to the following publications:

- Banks, E. J., Valdivia-Delgado, M., Biboy, J., Wilson, A., Cadby, I. T., Vollmer, W., ... & Sockett, R. E. (2022). Asymmetric peptidoglycan editing generates cell curvature in Bdellovibrio predatory bacteria. Nature communications, 13, 1-15. (Bd1075 PDB ID: 7O21)

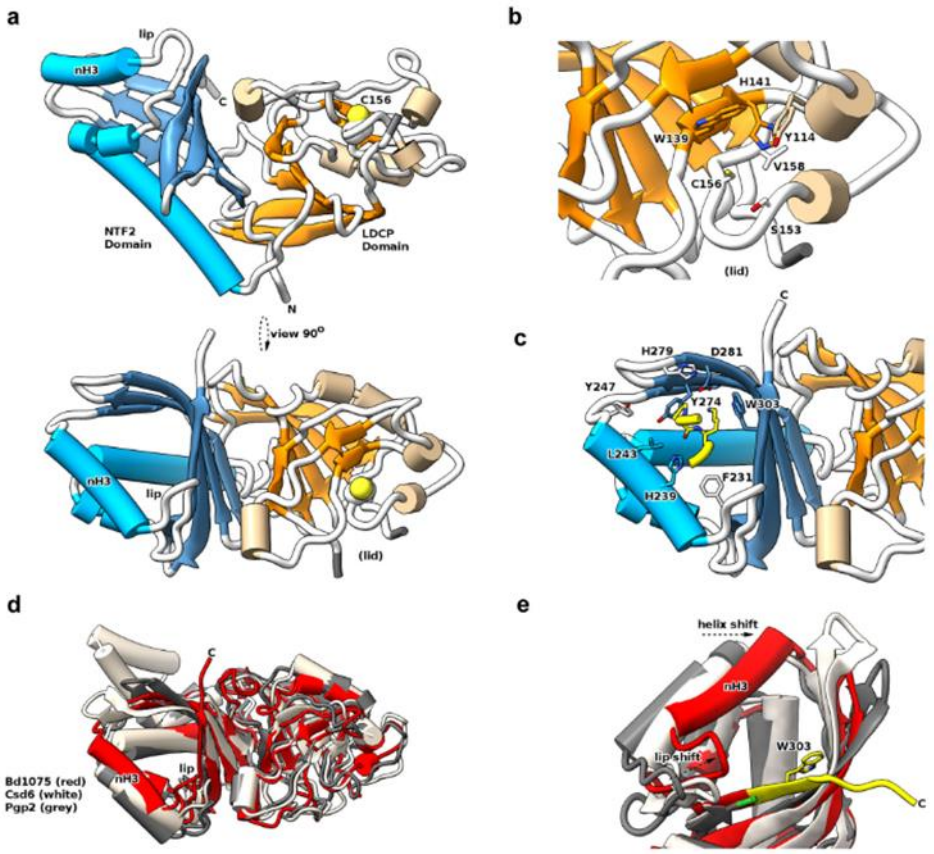


Fig. 5 Structure of Bd1075 and features different to other characterized LD-CPase enzymes. a Two orthogonal views of the Bd1075 fold, with catalytic residue C156 in space-fill form and features labeled. **b** Close-up view of the Bd1075 LD-CPase catalytic domain with selected residues that form the active site pocket displayed in stick form. **c** Close-up view of the Bd1075 NTF2 pocket, demonstrating complexation of a loop (residues 106–109 colored yellow, P107 and K108 in stick form) from a neighboring molecule in the crystal lattice. **d** Comparison of Bd1075 (red, 7O21: https://www.rcsb.org/structure/7O21), Csd6 (white, 4XZZ: https://www.rcsb.org/structure/4XZZ), and Pgp2 (gray, 6XJ6: https://www.rcsb.org/structure/6XJ6) structures. Helix 3 of the Bd1075 NTF2 domain (labeled 'nH3') and the associated loop ('lip') are relatively closer to the NTF2 pocket than the respective features of Csd6/Pgp2. **e** Close-up of the NTF2 terminus from structural alignment in **d**, demonstrating the relative extension of the Bd1075 C-terminus (colored yellow, includes NTF2 pocket-forming residue W303) in comparison to the shorter Csd6/Pgp2 termini (end residue colored green). The relative shifts of the nH3 helix and lip loop to constrict the NTF2 pocket are denoted by dashed arrows.

- Alexander, L. T., Lepore, R., Kryshtafovych, A., Adamopoulos, A., Alahuhta, M., Arvin, A. M., ..., Valdivia-Delgado, M., & Schwede, T. (2021). Target highlights in CASP14: Analysis of models by structure providers. Proteins: Structure, Function, and Bioinformatics, 89, 1647-1672. (Bd0675 PDB ID: 7OC9)

While a few methods in CASP14 predicted well the relative orientation of the two longest helices of the fold (α1, α2) and some predicted fairly well (albeit not accurately) also the relative orientation of the support (α3) and recognition helices (α4), they all failed to model accurately the N-terminal region and the orientation of the C-terminal α5 helix. The highest GDT-TS score for these methods was 61.03, with a total of 30 methods produced models with scores higher than 55.0. Rather remarkably, the score of the AlphaFold2 model was 96.09. AlpaFold2 correctly predicted the different placement of both terminal regions in relationship to the rest of the fold, but also modeled with remarkable accuracy the relative placement of the α1, α2 pair of helices in relation to the α3, α4 pair. Importantly, this could be independent evidence that the relative placement of the N-terminal region and helix α5 were not an artifact of crystal packing, but a real feature of the structure of JBP3 J-DBD. Finally, the AlphaFold2 model, but not other models we examined, is sufficiently accurate to easily phase the crystallographic data using a standard version of PHASER.[92]

2.8 | A cryptic predatory secreted protein, Bd0675, from *B. bacteriovorus* (CASP: T1074, PDB: 7OC9). Provided by Mauricio Valdivia-Delgado and Andrew L. Lovering

B. bacteriovorus are ubiquitous predatory Gram-negative δ-proteobacteria which predate on other Gram-negative bacteria.[93] To succeed as a micro-predator, *B. bacteriovorus* has developed a life-cycle consisting of location of the prey and initiation of the attack phase, attachment, and entry of prey, invasion of periplasm, bdelloplast formation, filamentous growth, exhaustion of prey cell resources, septation, and bdelloplast lysis and release of progeny.[93,94] The transcriptomic analysis of *B. bacteriovorus* strain HD100 has shown the upregulation of ~240 genes during predation (the predatosome), and the roles of most are cryptic and require further examination.[95]

We obtained the 1.50 Å structure of Bd0675 (target T1074), a 14 kDa secreted cryptic predatosome protein with no discernible domain annotation. Proteins similar to Bd0675 are found in different *B. bacteriovorus* strains and in other predators such as *Halobacteriovorax* spp.,[96] but no information regarding their function is available.

The determined structure, covering 133 of the 134 residues of the mature protein, forms a β-roll-like distorted architecture containing two α-helices and nine β-strands (Figure 10A). The overall β-roll fold part of the structure is formed by two β-sheets, one comprised by β-strands 1–4, which is connected to a second sheet, comprising β-strands 5–9, via a disulphide bond formed between residues C31 and C132, which appears to adopt two alternative conformations. Additionally, a disulphide bond C90–C118 links the 19 residues loop between β-strands 6 and 7 with β-strand 8, suggesting that correct positioning of this loop is relevant for Bd0675 function. All cysteine residues are conserved in predatory homologs.

An electrostatic surface potential shows that Bd0675 possesses a hand-like shape with a potential binding cleft, which is mainly negatively charged, situated in the middle of the protein formed by the connecting loops between β-strand C-termini ends, where the outermost loop is fixed in position by the C90–C118 disulphide bond, forming part of a (R/D)PGGXφCGXΩX_5Y motif, where X is any amino acid, φ is a hydrophobic residue and Ω is an aromatic residue. Residues Y23 and D24 (loop β1–2), Y57 and F58 (loop β3–4), D65, E67 (loop β4–5), and residues 83–93 (β6–7), form lobes limiting the width of the proposed ligand binding site of Bd0675 (Figure 9C). The limiting lobes are mostly composed of negatively charged and aromatic side chains in addition to residues S79, Y99, and K101 which are located toward the center of the cleft. Besides L12 (which forms part of the hydrophobic core) and G14 (located in the loop connecting α-helix A with β-strand 1), and the four cysteines, the only invariable residues of Bd0675 are Y57, G86, G87, and G91. Y57 localizes