Structural and Functional Investigations of *Bdellovibrio bacteriovorus* Lifecycle Proteins

Christopher John Harding

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

September 2018

Institute of Microbiology and Infection
School of Biosciences
University of Birmingham
September 2018

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This unpublished thesis/dissertation is copyright of the author and/or third parties.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

This thesis documents the findings undertaken on a diverse set of proteins that operate at distinct stages of the predatory lifecycle of *Bdellovibrio bacteriovorus*. In general, this work explores three major predatory lifecycle events: growth & metabolism, development & signalling and finally bdelloplast lysis & exit. In particular this thesis characterises four individual proteins (Bd2924, Bd1833, Bd0967 and Bd0314) that perform very different roles.

In this thesis the three-dimensional structure of Bd2924 (a putative acyl-CoA dehydrogenase) has been solved demonstrating an adaptation of the common acyl-CoA dehydrogenase fold, suggesting an important role beyond fatty acid β-oxidation. This study also presents the first structure of a multi-modular malic enzyme (Bd1833) and through extensive mutational and biochemical investigations, have unpicked the mechanism behind allosteric regulation by acetyl-CoA. In addition the three-dimensional structure of Bd0967 (a C-terminal processing protease) has been solved, revealing a divergent fold that forms a large self-compartmentalised cavity. Furthermore, Bd0967 is demonstrated to be involved in a developmental pathway that is interconnected with flagellin assembly. Bd0967 targets and processively degrades the flagellin Bd0606, which is mediated by interactions with the PDZ domain. Finally this study determines how *Bdellovibrio* exits the consumed prey at the end of the predatory lifecycle that had been chemically marked for destruction. This role is provided by a repurposed lysozyme (Bd0314) remodeled into a variant that now only accepts deacetylated cell wall sugars. Enzyme assays and addition to bdelloplasts confirms the specificity for modified prey cell wall material.

Acknowledgments

First of all I would like to thank my PhD supervisor, Andrew Lovering. You have been a great support throughout my PhD, always being there to discuss ideas and offer advise. Your enthusiasm and knowledge for science has been infectious and constantly encouraged me to grow as a research scientist. I could not have asked for a better supervisor.

I would also like to thank Patrick Moynihan, for all of your help and guidance with all things biochem and cell wall. You have been a massive help and it will always be greatly appreciated.

I greatly appreciate all the collaborative work and contributions people have made towards my PhD studies, including Oliver Daubney and Charles Moore-Kelly. I would especially like to thank Prof Liz Sockett and all those in the Sockett lab for all of your contributions and hard work, as well as being very hospitable hosts on my visits to Nottingham.

I would like to specially thank Ian Cadby, you have always been there to offer a helping hand and advice; it has been a pleasure to work and learn from you. To the rest of the Lovering lab, thank you for the support over the years, especially Richard Meek for always lightening the mood with your quirky personality. I would also like to thank all those on the 7th floor, it has been a pleasure working with you all. Additionally, thank you Tim Knowles for being my secondary supervisor and a great guy.

Personally, a big thank you to all my family for always supporting me. Mum and Dad, I would not be where I am without your constant support and encouragement. Last, but certainly not least, I would like to thank my wonderful wife, Sarah. I could always rely on your love and support, especially when things got tough.

Table of Contents

Chapter 1 ..... **1**

Introduction **1**

1.1 Predatory Bacteria **1**

1.1.1 Overview 1

1.1.2 Types 2

1.1.3 Beneficial Applications 5

1.2 *Bdellovibrio bacteriovorus* **6**

1.2.1 Overview 6

1.2.2 Prey Detection and Motility 7

1.2.3 Attachment and Invasion 9

1.2.4 Niche formation 10

1.2.5 Metabolism 11

1.2.6 Growth and Division 14

1.2.7 Maturation and Lysis 14

1.2.8 Cellular Signalling and Regulation of Predation 15

1.3 Overall Research Aim Of This Study **18**

Chapter 2 ..... **20**

Materials and Methods **20**

2.1 General **20**

2.1.1 Molecular cloning 20

2.1.2 Colony PCR 22

2.1.3 Mutagenesis 22

2.1.4 Q5 SDM 23

2.1.5 Primers 24

2.1.6 Sequencing 26

2.1.7 Transformation 26

2.1.8 Culture Media 26

2.1.10 Glycerol Stocks 27

2.1.11 Expression 27

2.1.12 Purification 28

2.1.13 Dialysis 29

2.1.14 Size Exclusion Chromatography 29

2.1.15 SDS-PAGE 29

2.1.16 Protein sample concentration and quantification 30

2.1.17 Protein crystallisation 31

2.1.18 Data collection and structure solving 31

2.2 Bd0314 ..... **32**

2.2.1 Cloning, Expression and Purification 32

2.2.2 FITC Peptidoglycan digestion assay 33

2.2.3 Premature bdelloplast lysis assay 35

2.2.4 Crystallisation and Cryoprotection 35

2.2.5 Data collection and structure determination 36

2.3 Bd1413 and Bd1440 ..... **38**

2.3.1 Cloning, Expression and Purification 38

2.4 Bd2924 ..... **38**
2.4.1 Cloning, Expression and Purification 38
2.4.2 SEC 38
2.4.3 MicroScale Thermothoresis – Cyclic-di-GMP Binding Assay 39
2.4.4 Crystallisation and cryoprotection 39
2.4.5 Data collection and structure determination 40
2.5 Bd1833 ..... **42**
2.5.1 Cloning, Expression and Purification 42
2.5.2 SEC 43
2.5.3 CoA removal Wash Steps 43
2.5.4 HPLC 44
2.5.5 Enzymatic activity assay 46
2.5.6 MST Cdg Binding Assay 48
2.5.7 Isothermal Titration Calorimetry 48
2.5.8 Crystallisation and cryoprotection 49
2.5.9 Data collection and structure determination 51
2.6 Bd0967 ..... **56**
2.6.1 Cloning, Expression and Purification 56
2.6.2 SEC 56
2.6.3 Pulldown Assay 56
2.6.4 Mass-Spectroscopy 58
2.6.5 Co-purified peptide analysis (Collaboration) 59
2.6.6 Bd0967 pulldown with Bd0606 60
2.6.7 Protease assay 60
2.6.8 Protease peptide inhibition assay 61
2.6.9 Crystallisation and cryoprotection 62
2.6.10 Data collection and structure determination 63
2.6.11 MST Cdg Binding Assay 64
2.7 Bd0606 ..... **65**
2.7.1 Cloning, Expression and Purification 65
2.7.2 Circular Dichromism (CD) 65

Chapter 3 ..... **66**

Bd0314, an Adapted Lysozyme Required for Bdelloplast Escape **66**
3.1 Introduction **66**
3.1.1 Peptidoglycan Overview 67
3.1.1.1 Overview 67
3.1.1.2 Structure and Composition 68
3.1.1.3 Peptidoglycan Variation 69
3.1.1.4 Biophysical properties 70
3.1.2 Peptidoglycan modifications and pathogenesis 72
3.1.2.1 N-deacetylation 73
3.1.2.2 O-acetylation 74
3.1.2.3 Surface polymer attachment 75
3.1.3 Lysozymes 75
3.1.3.1 Overview 75
3.1.3.2 Structures 78
3.1.3.3 Catalytic mechanism 80
3.1.3.4 Lytic transglycosylases – housekeeping enzymes 82
3.1.4 *Bdellovibrio* lifecycle and PG modifications 83
3.1.5 Introduction to the Exit enzyme 88
3.1.6 Aims and Objectives 88
3.2 Results **89**

3.2.1 Purification and Crystallisation ... 89
3.2.2 The structure of Bd0314 ... 91
3.2.2.1 Overall fold ... 91
3.2.2.2 The catalytic site ... 93
3.2.2.3 Substrate selection ... 94
3.2.2.4 Bd0314 structure comparison (DALI) and homologues ... 96
3.2.3 Bd0314 activity and Specificity ... 101
3.2.3.1 Bd0314 specificity ... 101
3.2.3.2 Catalytic residues ... 103
3.2.3.3 Bd0314 homologue activities ... 105
3.2.3.4 Bd0314 premature lysis assay ... 106
3.3 Discussion ... 108
3.3.1 Comparison of structure ... 108
3.3.2 Comparison of substrate-binding and specificity ... 111
3.3.3 Likely catalytic mechanism ... 114
3.3.4 Bd0314 role as the exit enzyme ... 116
3.3.5 Roles of other Bd0314 homologues ... 120
3.3.6 Presence in other bacterial systems ... 120
3.4 Conclusion ... 121

Chapter 4 ... 122

Structure and Function of Bd2924, a Putative Acyl-CoA Dehydrogenase ... 122
4.1 Introduction ... 122
4.1.1 Overview ... 122
4.1.2 Metabolism - β-oxidation ... 122
4.1.3 β-oxidation regulation in bacteria ... 124
4.1.4 Acyl-CoA dehydrogenases ... 125
4.1.5 Structures of ACADs ... 126
4.1.6 Substrate binding ... 128
4.1.7 Catalytic mechanism ... 130
4.1.8 ETF and its interactions with ACADs ... 131
4.1.9 ACAD homologues including acyl-CoA Oxidases (ACOs) ... 132
4.1.10 ACADs with divergent functions ... 134
4.1.11 Aims and objective ... 135
4.2 Results ... 137
4.2.1 Cloning, Expression and Purification of Bd2924 ... 137
4.2.2 Obtaining crystals and data collection ... 138
4.2.3 Protein Structure / topology ... 138
4.2.4 Dimer and Tetramer Interface ... 140
4.2.5 Ligand Binding ... 141
4.2.5.1 FAD ... 141
4.2.5.2 C10-CoA ... 143
4.2.6 CdG binding assay – MST ... 145
4.2.7 Bd2924 activity ... 146
4.3 Discussion ... 147
4.3.1 Structural comparison to ACADs ... 147
4.3.2 Substrate-binding and activity ... 148
4.3.3 A functionally relevant Bd2924 interaction with CdG is questionable ... 151
4.3.4 Role of Bd2924 ... 152
4.4 Conclusion ... 154

Chapter 5 ..... **155**

Bd1833 a Multi-modular Malic Enzyme **155**
5.1 Introduction **155**
5.1.2 Malic Enzymes 156
5.1.2.1 Overview 156
5.1.2.2 Malic enzyme roles 158
5.1.2.3 Alternative metabolism and NADPH generation 158
5.1.2.4 Malic enzyme structure 159
5.1.2.5 Divalent Cation binding 161
5.1.2.6 Substrate/ inhibitor binding 162
5.1.2.7 Nucleotide Binding 163
5.1.2.8 The catalytic residues and mechanism 164
5.1.2.9 Regulation 167
5.1.2.10 Prokaryotic MEs 167
5.1.3 Phosphotransacetylases (Pta) 171
5.1.3.1 Overview 171
5.1.3.2 Pta Biochemical Analysis 172
5.1.3.3 Pta Structure 173
5.1.3.4 Substrate Binding 175
5.1.3.5 Catalytic Mechanism 178
5.1.4 Multi-modular MEs possessing Pta domains 180
5.1.5 Aims and objectives 182
5.2 Results **183**
5.2.1 Bd1833 homologues 183
5.2.2 Bd1833 purification and crystallisation 187
5.2.3 Bd1833 Structure 188
5.2.4 The structure of Bd1833 ME domain 189
5.2.4.1 ME domain structure 189
5.2.4.2 ME catalytic site 190
5.2.4.3 Changes between bound states 193
5.2.5 The structure of Bd1833 Pta domain 193
5.2.5.1 Pta domain structure 193
5.2.5.2 Dimer and tetramer interfaces 195
5.2.5.3 Acetyl-CoA binding site 199
5.2.5.4 Pta domain lacks catalytic residues 200
5.2.5.5 Pta R535E and R535A structures 201
5.2.5.6 Pta-E544R 204
5.2.5.7 Y576S mutant movement 206
5.2.5.8 Pta-Apo structure domain conformational change 208
5.2.6 The structure of full-length Bd1833 211
5.2.7 Structure of Bd1833-SKB 216
5.2.8 Kinetics 219
5.2.8.1 Bd1833 activity 219
5.2.8.2 Malate Oxidative Decarboxylation 219
5.2.8.3 Effectors 220
5.2.8.4 Effector binding sites 221
5.2.8.5 Inhibitor Potency 224
5.2.8.6 Bd1833- Δloop kinetics 226
5.2.8.7 Other Bd1833 mutants 229
5.2.8.8 Pyruvate reductive carboxylation 230
5.2.8.9 Bd1833-SKB kinetics 231
5.2.9 HPLC analysis of bound ligands 233
5.2.9.1 Bd1833 co-purification with CoA and acetyl-CoA 233
5.2.9.2 Producing Bd1833-Apo 234
5.2.9.3 HPLC results for other Bd1833 mutants 236
5.2.9.4 HPLC Bd1833-Δloop 238
5.2.9.5 HPLC results for Bd1833-SKB 238

5.2.10 Size exclusion results 239
5.2.11 ITC results 241
5.2.12 Bd1833 interaction with CdG 242
5.3 Discussion 243
5.3.1 Structure and function of multi-modular malic enzymes 243
5.3.2 Bd1833 function / role 244
5.3.3 Bd1833 regulation 246
5.3.4 Comparison to regulation in other multi-modular malic enzymes 250
5.3.5 Acetyl-CoA binding and implications for Pta catalytic mechanism 253
5.4 Conclusion 257

Chapter 6 258

Bd0967, an Important Developmental C-terminal Processing Protease 258
6.1 Introduction 258
6.1.1 Overview 258
6.1.2 The diverse roles of C-terminal Processing proteases 259
6.1.2.1 Cell wall maintenance 259
6.1.2.2 Spore formation 260
6.1.2.3 Photosynthesis 261
6.1.2.4 Virulence 261
6.1.2.5 Quality control 262
6.1.3 Structure and mechanism of C-terminal processing proteases 263
6.1.3.1 D1P 264
6.1.3.2 CtpB 265
6.1.3.3 CT441 267
6.1.3.4 Prc 267
6.1.4 PDZ domains 268
6.1.4.1 Roles 268
6.1.4.2 Fold and substrate binding 269
6.1.5 Substrate specificity and mechanism of C-terminal processing proteases 272
6.1.5.1 C-terminal processing protease PDZ specificity 272
6.1.5.2 Protease domain binding sites 274
6.1.5.3 Mechanism involving PDZ domain 274
6.1.6 Bd0967 role in predatory cycle 279
6.1.7 Aims and objectives 280
6.2 Results 281
6.2.1 Obtaining Bd0967 crystals 281
6.2.2 The overall structure of Bd0967 282
6.2.3 Bd0967 cavity 285
6.2.4 The Active site / catalytic triad / bound peptide 286
6.2.5 Bd0967 has a unique metal-binding site 289
6.2.6 Bd0967 has unique disulphide bond 290
6.2.7 Substrate recognition by the PDZ domain 290
6.2.8 A monomer in solution 294
6.2.9 Structure of Bd0967 S457A 295
6.2.10 Identifying Bd0967 substrates / interacting partners 296
6.2.11 Bd0967 fully degrades the flagellin, Bd0606 300
6.2.12 The C-terminal peptide of Bd0606 as an inhibitor of Bd0967 activity 305
6.2.13 Bd0606 is folded 306
6.2.14 Bd0967 KO phenotype 307
6.2.15 Bd0967 interaction with CdG 308
6.3 Discussion 309
6.3.1 The structure of Bd0967 309
6.3.2 Bd0967 activity towards the flagellin Bd0606 and mechanism 311

6.3.3 The role of Bd0967 316
6.4 Conclusion 318

Chapter 7 319

Summary and Future Work 319
7. 1 Summary 319
7.2 Future Work 322

Chapter 8 325

References 325

Table of Figures

Figure 1: Types of predatory bacteria 4
Figure 2: Host dependent predatory lifecycle 7
Figure 3: Proteins of interest 17
Figure 4: Bd1833 misannotation 42
Figure 5: HPLC assay standards 46
Figure 6: Predatory lifecycle, prey cell wall modifications 67
Figure 7: Peptidoglycan structure 68
Figure 8: Modifications of the peptidoglycan strand 72
Figure 9: Structures of lysozyme superfamily members 77
Figure 10: Catalytic distances 79
Figure 11: Retaining enzyme mechanism 80
Figure 12: Inverting enzyme mechanism 82
Figure 13: Lytic Transglycosylase reaction 83
Figure 14: Bd0314 purification and crystallisation 90
Figure 15: Bd0314 structure schematic 90
Figure 16: Bd0314 structure 92
Figure 17: Bd0314 catalytic mechanism 93
Figure 18: Bd0314 substrate-binding site 94
Figure 19: Substrate-binding surface view 96
Figure 20: Bd0314 conservation 98
Figure 21: Bd0314 alignment with Bdellovibrio HD100 homologues 99
Figure 22: Bd0314 alignment and residue conservation 100
Figure 23: Bd0314 specificity 102
Figure 24: Time and concentration dependent assay 103
Figure 25: Bd0314 active site residues 104
Figure 26: Bd0314 homologue activities 105
Figure 27: Bd0314 bdelloplast lysis 107
Figure 28: Bd0314 structure comparison 110
Figure 29: Bd0314 substrate binding 112
Figure 30: Comparison of Bd0314 with chitosanases 114
Figure 31: Bd0314 catalytic residue location 116
Figure 32: Role of Bd0314 119
Figure 33: β-oxidation metabolic pathway 124
Figure 34: Acyl-CoA dehydrogenases 127
Figure 35: MCAD structure 128
Figure 36: ACAD mechanism 131
Figure 37: Electron-transfer by ETF 132
Figure 38: Structure of ACO 133
Figure 39: Structure of divergent ACAD homologues 136
Figure 40: Bd2924 purification and crystallisation 137
Figure 41: Bd2924 SEC 138
Figure 42: Structure of Bd2924 139
Figure 43: Bd2924 tetramer complex 141
Figure 44: FAD hydrogen bonding network 142
Figure 45: C10-CoA binding 143

Figure 46: Binding curve with CdG **146**
Figure 47: Bd2924 structural comparison **148**
Figure 48: Bd2924 active site **150**
Figure 49: Trp428 gates fatty acyl binding tunnel **151**
Figure 50: Illustration of central metabolism (general) **157**
Figure 51: Human-ME tetramer organisation and domains **161**
Figure 52: Human-ME active site **163**
Figure 53: Malic enzyme catalytic mechanism **166**
Figure 54: Prokaryotic ME (AYWB-ME) **168**
Figure 55: Comparison between prokaryotic and eukaryotic ME **170**
Figure 56: Pta structure **174**
Figure 57: Acetyl-P binding site **176**
Figure 58: Pta CoA binding site **178**
Figure 59: Pta domain movement **179**
Figure 60: Bd1833 alignment with multi-modular ME **184**
Figure 61: Bd1833-ME alignment with prokaryotic and eukaryotic ME **185**
Figure 62: Bd1833-Pta alignment with Pta enzymes **186**
Figure 63: Bd1833 Purification and crystallisation **187**
Figure 64: Bd1833-ME structure **190**
Figure 65: Bd1833 active site **192**
Figure 66: Bd1833-Pta structure **194**
Figure 67: Bd1833-Pta dimer and acetyl-CoA binding **195**
Figure 68: The dimer interface **196**
Figure 69: Hexamer interactions **197**
Figure 70: The hexamer interface **199**
Figure 71: R535 mutation **203**
Figure 72: Pta-E544R **205**
Figure 73: Y576S mutation **207**
Figure 74: Bd1833-Pta domain movement **209**
Figure 75: Pta mutations and hexamer interface **210**
Figure 76: Bd1833 structure **212**
Figure 77: Pta and ME interface **213**
Figure 78: Bd1833 allosteric regulation **215**
Figure 79: Bd1833-SKB superposed onto Bd1833 **217**
Figure 80: Malate oxidative decarboxylation **219**
Figure 81: Identifying Bd1833 effectors **220**
Figure 82: Bd1833 inhibitors **221**
Figure 83: Identifying acetyl-CoA binding site **223**
Figure 84: Inhibitor potency **225**
Figure 85: Multi-modular ME specific loop **227**
Figure 86: Bd1833 mutants activity and sensitivity to acetyl-CoA **228**
Figure 87: Pyruvate carboxylation reaction **230**
Figure 88: Bd1833-SKB kinetic parameters **232**
Figure 89: HPLC results **234**
Figure 90: CoA and acetyl-CoA removal by denaturing wash procedure **235**
Figure 91: HPLC traces for Bd1833 and Bd1833-Pta mutants **237**
Figure 92: HPLC results for Bd1833-SKB **238**
Figure 93: Size exclusion results **240**
Figure 94: ITC results for acetyl-CoA and CoA **242**

Figure 95: Schematic of Bd1833 role in metabolism **247**
Figure 96: Cartoon of Bd1833 mechanism **249**
Figure 97: Bd1833 acetyl-CoA binding site comparison **252**
Figure 98: Pta domain movements **254**
Figure 99: Pta active site **256**
Figure 100: Structures of C-terminal processing proteases **264**
Figure 101: PDZ domain structure **271**
Figure 102: PDZ domain movement in CtpB **276**
Figure 103: C-terminal processing protease mechanism **278**
Figure 104: Bd0967 crystallisation **281**
Figure 105: Domain architecture of Bd0967 **283**
Figure 106: Bd0967 B-factors **284**
Figure 107: Bd0967 internal cavity **285**
Figure 108: Bound peptides **286**
Figure 109: Protease peptide tunnel and binding site **288**
Figure 110: Bd0967 metal-binding site **289**
Figure 111: Peptide-binding PDZ domain **292**
Figure 112: Arg299 interaction with peptide and protease cap **293**
Figure 113: SEC results for Bd0967 **294**
Figure 114: Bd0967-S457A comparison **295**
Figure 115: Bd0967 pull-down assay **297**
Figure 116: Confirm binding **301**
Figure 117: Bd0967 degrades Bd0606 **301**
Figure 118: Bd0967 activity in comparison to mutations **302**
Figure 119: Bd0967 activity towards various proteins **303**
Figure 120: Bd0606 C-terminal mutants **304**
Figure 121: Peptide inhibition assay **305**
Figure 122: Bd0606 CD spectra **306**
Figure 123: Bd0967 KO mutant *Bdellovibrio* **308**
Figure 124: Bd0967 comparison to *E. coli* Prc **310**
Figure 125: Flagellin assembly *Bdellovibrio* **312**
Figure 126: Bd0967 mechanism **315**
Figure 127: Summary of thesis findings **320**

Table of Tables

Table 1: Gene primers **25**
Table 2: SDM primers **25**
Table 3: Bd0314 data collection and refinement **37**
Table 4: Bd2924 data collection and refinement **41**
Table 5: Bd1833 crystallisation conditions **50**
Table 6: Bd1883-ME data collection and refinement **52**
Table 7: Bd1833-Pta data collection and refinement **53**
Table 8: Bd1833 full-length data collection and refinement **54**
Table 9: Bd1833-SKB data collection and refinement **55**
Table 10: Bd0967 data collection and refinement **64**
Table 11: DALI results **97**
Table 12: HD100 ACADs **153**
Table 13: Bd1833 structures **188**
Table 14: Bd1833 kinetics **218**
Table 15: PDZ proteins in *Bdellovibrio HD100* genome **279**
Table 16: Mass spectrometry results for Bd0967 pull-down assay **299**

List of Abbreviations

A280	Absorbance at a wavelength of 280nm
aa	Amino acids
ABC	Ammonium bicarbonate
ACAD	Acyl-CoA dehydrogenase
AcCoA	Acetyl-CoA
Acetyl-P	Acetyl-Phosphate
ACOs	Acyl-CoA oxidases
AP	Attack phase
ASU	Asymmetric unit
BALO	Bdellovibrio-and-like organisms
Bd0314	Deacetylated GlcNAc specific lysozyme
Bd0967	Developmental C-terminal processing protease
Bd1413	Homologue of Bd0314
Bd1440	Homologue of Bd0314
Bd1833	Multi-modular malic enzyme (full-length enzyme)
Bd1833-Apo	Ligand-free full-length Bd1833
Bd1833-ME	Bd1833 ME domain only
Bd1833-Pta	Bd1833 Pta domain only
Bd1833-SKB	Homologue of Bd1833 from *Bdellovibrio*-SKB strain
Bd1833-Δloop	Bd1833 chimera
Bd2924	Putative Acyl-CoA dehydrogenase
BLAST	Basic local alignment tool
CAZY	Carbohydrate-Active enzymes database
CdG	Cyclic di-GMP
cf.	Compare
CoA	Coenzyme A
CTD	C-terminal domain
CTP	Carboxy-terminal processing protease
CV	Column volumes
DAP	meso-2,6-Diaminopimelic acid
dNTPs	Deoxynucleotides
e.g.	For example
EDTA	Ethylenediaminetetraacetic acid
FAD	Flavin Adenine Dinucleotide
FITC	Fluorescein-5-isothiocyanate
FOM	Figure of Merit
GH	Glycoside hydrolase
GlcN	Glucosamine (*N*-deacetylated GlcNAc)
GlcNAc	*N*-acetyl glucosamine
GP	Growth phase
GUI	Graphical user interface
HD	Host dependent
HI	Host independent

HPLC	High-performance liquid chromatography
IAA	Indole-3-acetate
i.e.	In other words
IMAC	Immobilised metal affinity chromatography
IPTG	Isopropyl-β-D-thiogalactoside
ITC	Isothermal titration calorimetry
LB	Lysogeny broth
LLG	Log likelihood gain
LPS	Lipopolysaccharide
LSQ	Least squares minimisation
LT	Lytic transglycosylase
MBP	Maltose binding protein
MCP	Methyl-accepting chemotaxis protein
ME	Malic Enzyme
MR	Molecular replacement
MS	Mass Spectrometry
MST	Microscale thermophoresis
MurNAc	*N*-acetyl muramic acid
MWCO	Molecular weight cut off
NCBI	National Center for Biotechnology Information
NTD	N-terminal domain
O/N	Over night
OAA	Oxaloacetic acid
OD_{600}	Optical density at a wavelength of 600nm
PCR	Polymerase chain reaction
PDB	Protein data bank
PG	Peptidoglycan
P_i	Inorganic Phosphate
Pta	Phosphotransacetylase
Rmsd	Root-mean-square-deviation
SDM	Site directed mutagenesis
SDS-PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
SEC	Size exclusion chromatography
SSAD	Sulphur single-wavelength anomalous dispersion
SSM	Secondary structure matching
TB	Terrific broth
TBE	Tris/Borate/EDTA
TEV	Tobacco Etch virus
TFZ	Translation Function Z-score
v / v	Volume to volume
w/ v	Weight to volume
WT	Wild type

Chapter 1

Introduction

1.1 Predatory Bacteria

1.1.1 Overview

Predation is a fundamental part of life and found in all ecosystems. However, the predatory way of life varies from the smallest microbes to the largest mammals. An organism is categorised as a predator if it actively hunts, kills and consumes the nutrients of its prey to grow and survive.

Within bacterial communities, hostile interactions occur regularly in the natural fight for survival. Bacteria compete for nutrients and ecological niches and employ a range of techniques from antibiotic secretion used to lyse or limit the growth of neighbours, to bacterial predators that actively hunt and kill prey. Predatory bacteria are a diverse group of bacteria that are widespread throughout a range of environments. Interactions between predatory and non-predatory populations are an important ecological factor and proposed as an evolutionary driving force [2].

1.1.2 Types

The diversity within bacterial predators can be categorized based on the type of hunting strategy they adopt. These hunting strategies include group attack, epibiotic and endobiotic predation (Figure 1).

Group Attack:

This predatory strategy includes bacteria that are reported to hunt cooperatively in groups. Group attack predation typically involves the secretion of diffusible lytic compounds that kill and decompose a prey cell. Alternatively prey lysing agents can be embedded into the predator's cell surface and prey lysis is caused by direct contact [3]. Group predation does not limit the rewards of the kill solely to the predators and non-secreting cells can also benefit from lysed prey. Predators adopting this technique include *Lysobacter* and *Myxococcus xanthus* [4].

Myxococcus xanthus is a social predatory bacteria that moves and feeds in cooperative groups referred to as swarms. When a swarm encounters prey they secrete hydrolytic enzymes to lyse the prey and release cellular contents. The released nutrients benefit all members of the swarm [5]. Due to the non-specific nature of group attack, *Myxococcus xanthus* is able to kill a wide variety of prey including both Gram-positive and Gram-negative bacteria and also some fungi. *Myxococcus xanthus* is not an obligate predator and can grow as saphrophytes by decomposing dead organic matter [6].

Epibiotic:

This strategy requires cell-cell contact between predator and prey. The predator remains attached to the outside surface of the prey cell where they secrete degradative

enzymes directly into the prey and then extract the prey cell nutrients. Predators that adopt this strategy include: *Vampirococcus, Vampirovibrio, Micavibrio* and *Bdellovibrio exovorus* [2]. The latter two are classified as *Bdellovibrio* and like organisms (BALO), where gene loss is believed to be a contributing factor to *B. exovorus* epibiotic predatory technique [7]. *Micavibrio aeruginosavorus* is an obligate epibiotic predator that leaches nutrients from a highly specific set of prey, including *Pseudomonas aeruginosa* and *Klebsiella pneumoniae* [8].

Endobiotic strategy:

This strategy involves a predator invading a prey cell and either residing inside the prey's periplasm or the cytoplasm. *Daptobacter,* an obligate predator, is the only reported organism to invade the prey's cytoplasm. *Daptobacter* preys on a narrow range of bacteria in the genus *Chromatium*. From the cytoplasm of the prey *Daptobacter* secretes hydrolytic enzymes to degrade the prey cell contents and uptake nutrients. The prey is consumed from within, leaving only an intact cell wall after predation [9].

On the other hand invasion into the prey's periplasm is well documented and epitomised by *Bdellovibrio bacteriovorus. Bdellovibrio bacteriovorus* belongs to a larger group of BALO that includes *bacteriovorax, halobacteriovorax* and *peredibacter* species [10]**.** BALOs adopting the endobiotic strategy are obligate predators that actively seek out Gram-negative bacterial prey, so they can invade the periplasmic space and consume their prey from within. BALO are found in a wide range of environments from terrestrial to aquatic [10].

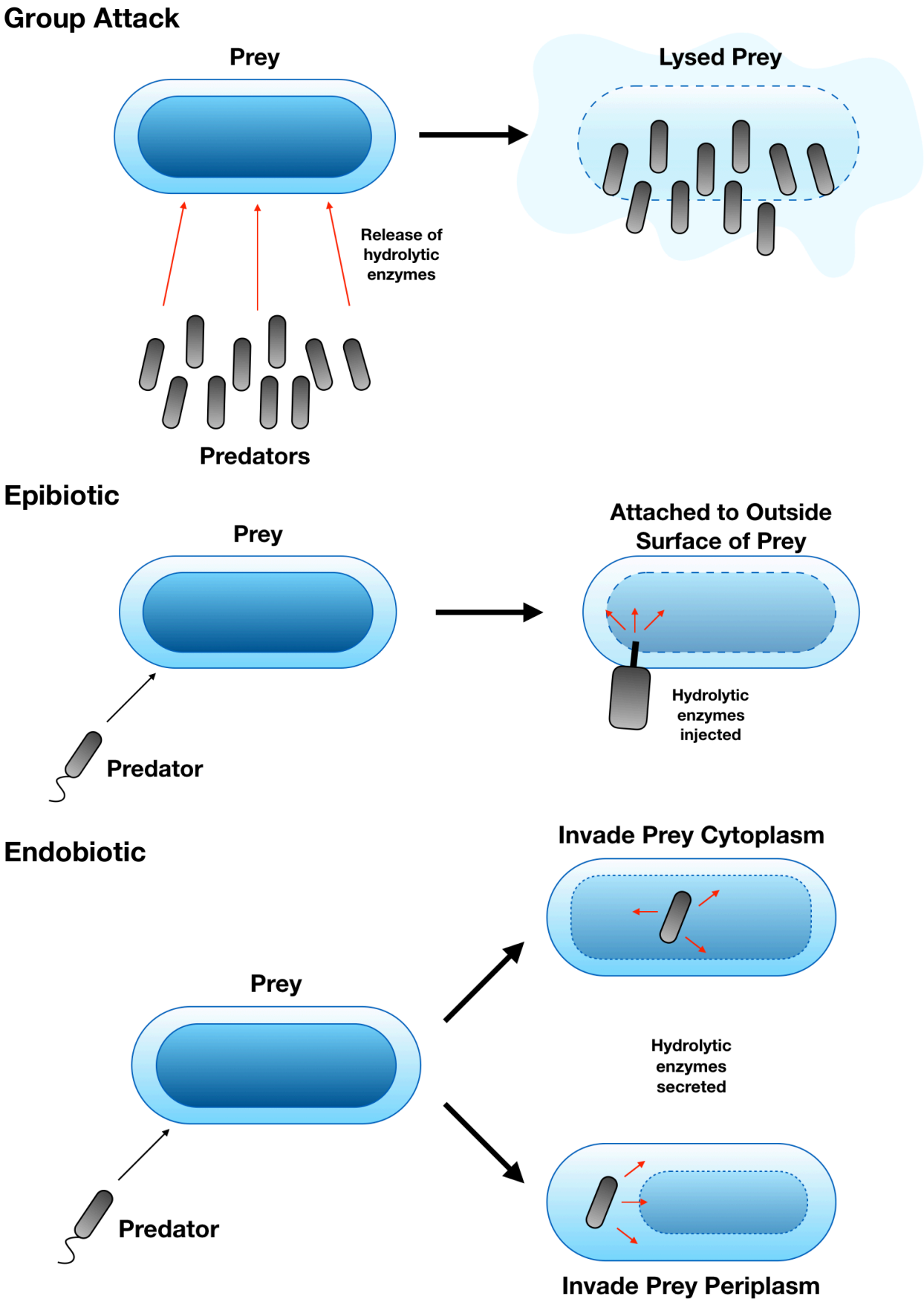


Figure 1 Types of predatory bacteria
Predatory bacteria can be categorised into three main groups based on how they kill their prey. These classes include: group attack, epibiotic and endobiotic predation. Predators are black ovals and the prey cells are blue ovals (illustration depicts Gram-negative cells). Red arrows show release of degradative hydrolytic arsenal is employed to kill the prey.

1.1.3 Beneficial Applications

Predatory bacteria have become attractive candidates for biocontrol agents with potential applications ranging from medicinal therapeutics to use in agriculture. BALOs are the most well-characterised predatory bacteria and have received the most attention, in regards to practical applications. BALOs have been proposed as potential 'living' antibiotics [11] and have been shown in multiple studies to effectively control bacterial infections. The first study using *Bdellovibrio bacteriovorus* as a potential therapeutic, investigated *Bdellovibrios* ability to control a *Salmonella* infection, which colonised the guts of chickens [12]. Promisingly, it was shown that *Bdellovibrio* significantly reduced *Salmonella* numbers and did not have any detrimental effects on the health of the infected birds. More recently *Bdellovibrio* and *Micavibrio* have been used to attenuate the burden of *Klebsiella pneumoniae* infections in rat lungs [13]. Additionally *Bdellovibrio* has been investigated as a therapeutic to treat zebrafish larvae infected with an antibiotic-resistant strain of *Shigella flexneri* [14]. Injection of *Bdellovibrio* led to increased zebrafish survival by promoting immune system activation, besides actively killing *Shigella* through predation.

There has been no reported cases of acquired resistance to predation by *Bdellovibrio*, which also makes them very attractive targets for use as a 'living' antibiotic against multidrug resistant species [15]. *Bdellovibrios* potential therapeutic abilities are limited to Gram-negative bacteria. However, a study demonstrated Gram-positive *Staphylococcus aureus* virulence was reduced by *Bdellovibrio* application due to inhibition of biofilm formation [16]. *Bdellovibrio* therefore has the potential to be used to invade deep into complex multi-organism biofilms, reducing bacterial virulence and killing bacteria where conventional therapeutics cannot penetrate [17, 18].

Beyond being used medicinally, *Bdellovibrio* has the potential to be used as an effective biocontrol agent. *Bdellovibrio* has been successfully used to improve water quality in an agricultural setting [19]. Additionally *Bdellovibrio* has also been used in agriculture and the food industry to control pests that damage crops and food produce [20-22].

1.2 *Bdellovibrio bacteriovorus*

1.2.1 Overview

Bdellovibrio bacteriovorus is a small Gram-negative δ-proteobacteria that is an obligate predator of Gram-negative bacteria. *Bdellovibrio* has an interesting biphasic lifecycle where it requires a prey cell in order to grow and replicate from within the preys periplasm. *Bdellovibrio* was first isolated by Stolp in 1962, which was the beginning of many years of intensive research targeted at developing our understanding of *Bdellovibrios* complex interactions with its prey [23]. More recently the genome of *Bdellovibrio bacteriovorus HD100* has been sequenced leading to a new wave of research aimed at revisiting the developmental biology of *Bdellovibrio* [24].

The lifecycle of *Bdellovibrio* exists as two distinct phases: Attack Phase (AP), where a highly motile flagellated *Bdellovibrio* seeks a suitable prey and a sessile Growth Phase (GP), where *Bdellovibrio* grow, replicate and mature inside the preys periplasm [25]. Typically, the predatory life cycle of *Bdellovibrio* lasts for 3-4 hours and has distinct physiological events that are shown in Figure 2.

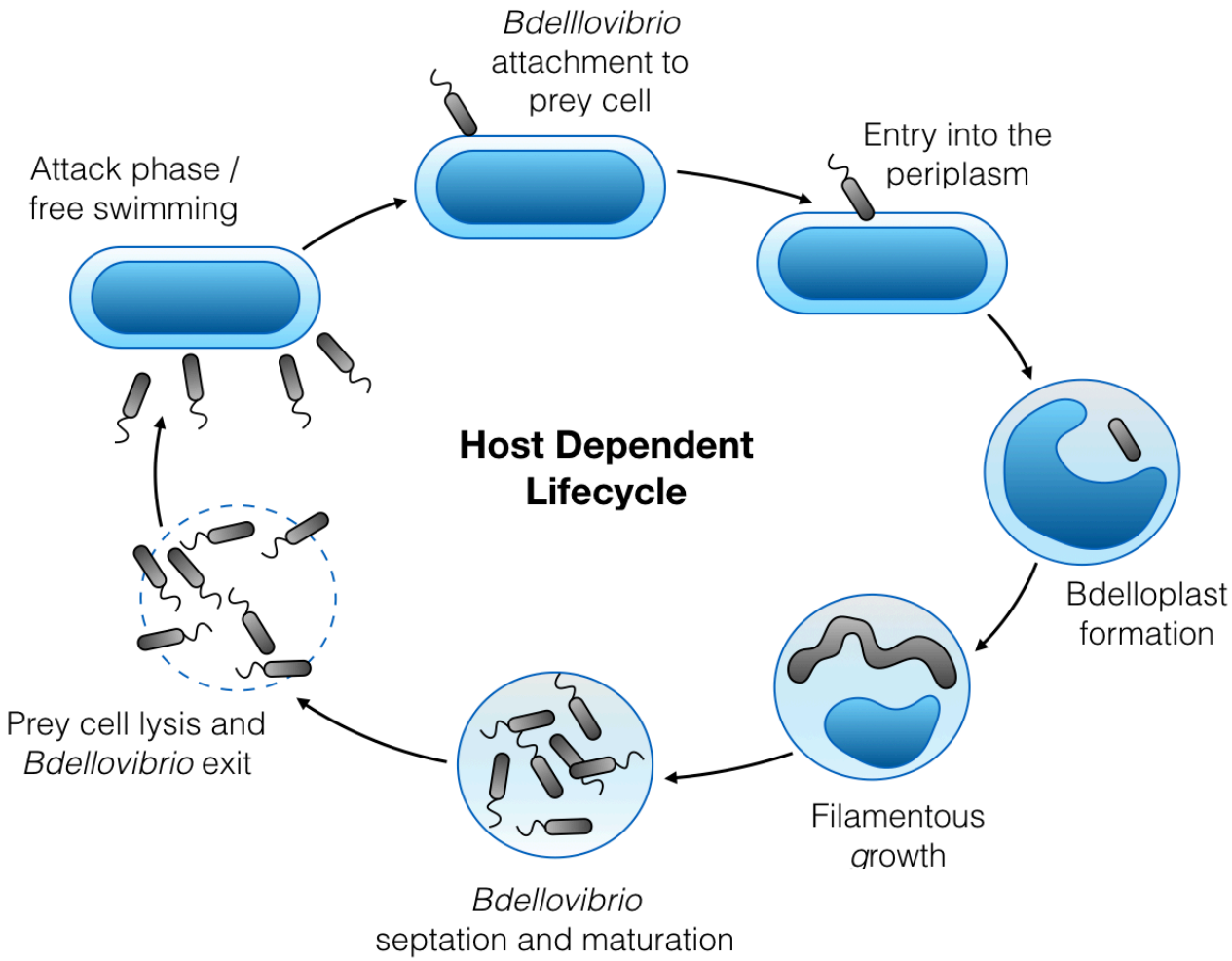


Figure 2 Host Dependent Predatory Lifecycle
The HD predatory lifecycle of *Bdellovibrio* can be divided into distinct stages. The cycle begins with free swimming and highly motile AP cells that actively seek out prey. Once located, *Bdellovibrio* attaches to the prey cell. Initially attachment is reversible (recognition period) but becomes permanent prior to invasion. Localised application of hydrolytic enzymes create a small pore in the prey cell envelope, which *Bdellovibrio* squeezes through, traversing the outer-membrane and cell wall to reside in the preys periplasm. During invasion the flagellum is shed and the entry pore is resealed. Additionally prey cell wall modifying enzymes are secreted, which remodel the prey cell wall creating an osmotically stable bdelloplast structure. Once the niche has been established *Bdellovibrio* begins the growth and replicative phase, growing filamentously. Once prey cell nutrients become depleted *Bdellovibrio* undergoes division and the progeny mature into AP cells. Mature AP progeny finally lyse the bdelloplast and begin the predatory cycle anew.

1.2.2 Prey Detection and Motility

Attack phase cells are small (0.2-0.5 µm x 0.5-2.5 µm), monoflagellate, rod shaped cells [11]. The attack phase cells are highly motile cells, swimming at speeds up to 160 µm sec^{-1} [26]. Motility is generated by a single, long, polar and sheathed flagellum that

displays a complex dampened wave form morphology [27]. The flagellum is composed of six flagellins (Fli1-6) arranged specifically from the proximal to distal end, to create a tapered flagellum filament [28]. Only FliC3, found at the proximal end, is found to be essential. It is suggested that the tapered shape may be necessary to control the optimal length of the flagellum (4 μm) and aid swimming [28]. Three sets of MotAB flagellar motor proteins contribute, to different degrees, to the rotation of a singular flagellum and provide propulsion [29].

Flagellar motility is crucial to locate prey but is not required for invasion or slow movement on surfaces such as biofilms [18, 26, 30]. Slow surface gliding of *Bdellovibrio,* is reported to occur at 16 μm $hour^{-1}$ [30], involving a similar mechanism to that employed by *Myxococcus xanthus* during group predation [31]. Interestingly *Bdellovibrio* possesses a Ras GTPase-like protein (MglA) that is an essential gliding motility protein and controls social mobility in *Myxococcus xanthus* [32]. However, in *Bdellovibrio,* MglA is not required for gliding but is a critical factor for invasion.

It is suggested that *Bdellovibrio* target prey rich regions. However, classical quorum sensing genes are absent from the genome of *Bdellovibrio HD100* so it unlikely a density dependent signal is used to guide predation [24]. The presence of three aerotaxis genes and twenty methyl-accepting chemotactic protein (MCP) genes in the genome suggests chemotaxis directs *Bdellovibrio* to prey rich regions [24]. Early experiments suggest *Bdellovibrio* is attracted towards: amino acids [33]; areas with high a concentration of prey cells (10^8 cells ml^{-1}) [34]; and areas with a high concentrations of organic and inorganic substances including acetyl-CoA, oxaloacetate, potassium and oxygen [35]. Furthermore, deletion of a single MCP gene in *Bdellovibrio* reduced the efficiency of

predation in the mutant [36]. The exact mechanism behind prey location is still disputed. However, chemotaxis likely plays a role.

1.2.3 Attachment and Invasion

Bdellovibrio is able to recognise a suitable prey cell from mixed populations of bacteria. Predation is believed to be a selective process and *Bdellovibrio* makes a preferential decision about which prey to invade at the point of attachment [37]. Membrane interactions are suggested to be important for attachment, demonstrated by *Bdellovibrios ability* to attach to ultraviolet killed *E. coli* but lose attachment competency towards *E. coli* cells killed by heat [38]. Moreover, attachment rates varied between *E. coli* and *Salmonella spp* cell wall mutants, which further demonstrates the importance of cell wall interactions [39]. The initial attachment is reversible with a 'recognition period' where *Bdellovibrio* can swim away after a few seconds before committing to predation. Once the invasion process has started *Bdellovibrio* becomes irreversibly attached to the prey cell [40].

Mounting evidence suggests the type IV pilus apparatus is required for attachment and invasion. Type IV pilus structures protrude from the 'biting' pole (non-flagellated pole) and deletion of the *pilA* gene abolishes the predatory capability of *Bdellovibrio* [41, 42]. It is believed that retraction/extrusion of type IV pilus fibres are involved during invasion and switching *Bdellovibrio* from an AP cell to a GP cell [43].

The irreversible attachment to the prey cell is succeeded by localised application of hydrolytic enzymes that create a small pore in the prey outer-membrane and cell wall.

Bdellovibrio immediately proceeds by squeezing through the small pore into the prey's periplasm, ultimately positioning between the prey's inner-membrane and cell wall [44]. The enzymes used to create the pore are still unknown. However, two enzyme activities, a glycanase and protease have been shown to be active during invasion [45]. The full invasion process typically takes between 5-20 minutes [38] and is concluded by resealing of the entry pore and shedding of the flagellum [26, 46].

1.2.4 Niche formation

During invasion *Bdellovibrio* secretes hydrolytic enzymes that remodels the prey cell creating a niche suitable for growth and replication. DD-endopeptidases are secreted, which decross-link the prey cell wall and cause the classical rounding of the prey cells (bdelloplast formation) associated with *Bdellovibrio* predation. The genes, *bd0816 and bd3459,* encode the two DD-endopeptidases required for bdelloplast formation [47]. Rounding of the prey cell creates an osmotically stable bdelloplast structure that has a more deformable cell wall, which prevents successive invasion events by other *Bdellovibrio*. Self-protection from the activity of the DD-endopeptidases is governed by an ankryin (Bd3460) that can form a complex with both Bd3459 and Bd0816 to inhibit their activity before secretion into the prey [48].

GlcNAc-deacetylases (Bd0468 and Bd3279) are also secreted into the prey cell and involved in deacetylation of *N*-acetylglucosamine residues (GlcNAc) of the prey cell wall [49]. Deacetylation of the prey cell wall renders the bdelloplast resistant to lysozyme activity and was originally believed to halt the initial pore forming glycanase activity [50]. However, deacetylation of the prey cell wall is now believed to stabilise the

bdelloplast structure and mark the prey cell wall for destruction at the end of the predatory lifecycle [49].

The prey cell wall is further modified by: removing a large proportion of Braun lipoprotein; attachment of fatty acids; and re-incorporation of DAP [51, 52]. All of the prey cell wall modifications are suggested to stabilise the bdelloplast structure and secure the intraperiplasmic niche.

During niche establishment the prey cell is killed; its cytoplasmic membrane is damaged leading to a rapid decline in the prey's ability to respire [53]. The prey's DNA is also degraded and its synthesis halted [54]. The prey's inner-membrane becomes freely permeable to small hydrophilic molecules early after invasion [55]. Release of cytoplasmic materials can be taken up by a wide variety of primary transporters in the *Bdellovibrio* genome that are capable of facilitating the transport of sugars, ions, amino acids and short chain fatty acids. Additionally carbohydrate metabolites such as glycerol 3-phosphate and di/tricarboxylates may also be transported from prey [24].

1.2.5 Metabolism

Bdellovibrio has a private source of nutrients available within the bdelloplast, which it uses for biosynthesis and reproduction. Consequently, *Bdellovibrios* metabolism is unusual and heavily reliant on the prey cell. For instance *Bdellovibrio* lacks the ability to synthesise and degrade various amino acids and riboflavin [24]. Only 11 standard amino acids can be synthesised from metabolic intermediates, even though a full

complement of tRNAs are available, suggesting amino acids acquired from the prey cell are directly used in protein synthesis.

Bdellovibrio is extremely well equipped to harness the nutrients from the prey cell and has a large hydrolytic arsenal at its disposal, including 150 proteases and peptidases, 10 glycanases, 20 deoxyribonuclease, 9 ribonucleases and 15 lipases [24]. *Bdellovibrio* is also shown to be a very energy efficient organism, aided by the direct incorporation of complex prey metabolites [56].

Up to 70% of the prey's DNA material is incorporated into the breakdown of prey DNA is tightly synchronised to nucleotide incorporation [57]. Additionally, 20-40% of prey RNA is degraded into base and ribose-1-phosphate sugars that are incorporated in *Bdellovibrio* [58]. The ribose sugars can then be further metabolised for energy. The large incorporation of prey-derived DNA and RNA material reduces the requirement of synthesis, even though a full set of genes required for purine and pyrimidine metabolism are present in the genome [24].

Bdellovibrio is reported to directly incorporate prey-derived lipids including lipopolysaccharides [59-61]. Moreover lipid metabolism is shown to be dependant on the prey they consume [62]. *Bdellovibrio* also possesses a complete LPS biosynthetic pathway, capable of synthesising LPS with a unique Lipid A molecule that is less antigenic [63].

The direct incorporation of larger membrane proteins such as OmpF was originally suggested to take place [64, 65]. However, more recently the incorporation of fully

folded outer-membrane proteins from prey is no longer believed to occur and instead *Bdellovibrio* synthesises and incorporates its own novel outer-membrane proteins [66], [67].

Bdellovibrio has a fully functioning electron-transport chain and a full set of ATP synthase genes in its genome [24, 68]. Oxidative phosphorylation is a major contributor to the electron-transport chain and energy production in *Bdellovibrio*. Studies have shown that the ATP pool is maintained during growth phase, which suggests the energy producing and energy utilising processes during this period are equal [69, 70]. Moreover, during GP, ATP can be directly taken from the prey cytoplasm [71] by a yet unknown mechanism, which further increases the energy efficiency [56]. During starvation conditions (non-growing AP cells) ATP is produced periodically as a mechanism of metabolic regulation, which makes use of their own cellular material.

Bdellovibrios metabolism relies heavily on degradation of prey molecules to respire. Both nucleotides and amino acids derived from degradation of prey material are the main sources of energy [72]. In addition prey lipids also serve as an energy-rich source for respiration [68]. Approximately 30-40% of prey-derived lipids are believed to be metabolised through fatty acid β-oxidation [52]. *Bdellovibrio* also has a fully functioning TCA cycle [72] that is central to providing energy during growth phase. TCA cycle enzyme activities have been detected with increasing activity during the growth phase [73]. The TCA cycle in *Bdellovibrio* must be able to cope with the increased metabolite influx from other pathways such as amino acid degradation. Consequently, tight control of TCA cycle enzymes and those closely associated is likely required to successfully divert flux through central metabolic pathways. Interestingly, respiration from sugar

molecules has not been detected, which is supported by the lack of sugar-metabolising enzymes in the genome [24]. Alternatively, *Bdellovibrio* may also be able to metabolise low carbon sources such as acetate (acetyl-CoA synthase gene present in genome) [68].

1.2.6 Growth and Division

Metabolically active *Bdellovibrios* grow from both poles forming a long filament [74]. During the growth phase *Bdellovibrio* begins to rapidly replicate its chromosomes into several copies that are spread along the cell filament before cell division takes place. The filamentous cell continues growing until the prey cell nutrients become depleted, at which point cell division is triggered. The filamentous cell divides synchronously to produce both odd and even numbers of progeny [74]. Typically 4-6 progeny are produced per prey cell but this is proportional to the size of the prey cell and larger prey allow more progeny to be produced [75]. Septation into progeny cells takes approximately 40 minutes to complete, which is independent of the number of progeny being produced [74].

1.2.7 Maturation and Lysis

Once division has been completed, it takes approximately 30 minutes for the progeny to mature into AP cells [74]. During maturation, flagellar synthesis is activated and the progeny cells develop a single polar flagellum. Once mature, the motile attack phase cells complete the predatory cycle by lysing the bdelloplast. Bdelloplast lysis requires a secondary burst of hydrolytic activity, directed towards solubilising the remainder of the prey cell wall [45]. The yet unknown enzyme(s) responsible for the final solubilisation of the prey cell must be able to act upon the previously deacetylated prey

cell wall (lysozyme resistant) [49, 50]. After bdelloplast lysis the free swimming AP cells go in search of new prey, starting the predatory cycle anew.

1.2.8 Cellular Signalling and Regulation of Predation

Cellular signalling and response to environmental cues is an important feature regulating transition through the predator lifecycle. The predatory lifecycle is strictly regulated and distinct lifecycle events are completed in a systematic and stringent manner [25]. Many of signalling events that elicit important physiological changes are still poorly understood; for instance, the trigger for invasion, the cause of flagellum shedding, the trigger for division and maturation and also lysis of the bdelloplast.

Our understanding of growth and development of *Bdellovibrios* Host Dependent (HD) lifestyle has been enhanced by the spontaneously arising mutant Host Independent (HI) strain, which is able to grow and replicate axenically in a rich media [76]. Spontaneous mutations have been mapped to the *'hit'* locus, typically leading to deleterious mutations of *bd0108* gene [24, 77]. The mechanism by which Bd0108 governs HI growth, although not fully understood, is believed to be involved in a regulatory mechanism associated with the type IV pilus extrusion/retraction [43, 78].

Cell cycle progression in *Bdellovibrio* has been shown to be dependent on the secondary messenger cyclic-di-GMP (CdG) [79]. CdG is an almost ubiquitous bacterial second messenger, influencing bacterial growth and a range of behaviours such as motility, virulence, biofilm formation and cell cycle progression [80]. CdG is synthesised by diguanylyl cyclases containing GGDEF domains and degraded by phosphodiesterases

containing EAL or HD-GYP domains. CdG interacts with a variety of receptors, including proteins containing PilZ domains [80]. There is an abundance of CdG related proteins in the *Bdellovibrio* genome, which includes: five GGDEF containing enzymes, one EAL and six HD-GYP containing enzymes, one putative CdG riboswitch and fifteen PilZ containing putative CdG receptors.

Each of the GGDEF containing enzymes has been found to control non-overlapping aspects of the predatory life cycle. CdG signalling in *Bdellovibrio* is implicated in controlling invasion, gliding motility and the transition from attack phase to growth phase [79]. The switch from attack phase to growth phase cells is complex and involves the differential regulation of an almost completely distinct set of genes. A CdG riboswitch is proposed to act as a CdG sink during the AP, which facilitates the rapid transition between AP and GP [81].

The large network of CdG machinery in *Bdellovibrio* is suggested to extend to a novel set of potential CdG receptors [82]. Rotem et al. identified 84 potential CdG receptors in a pull-down experiment using *Bdellovibrio* HD and HI lysates, where CdG was used as the bait molecule. 65 of the identified CdG binding proteins lacked any known CdG binding motifs [82]. Validation of the pull-down technique was conducted using microscale thermophoresis (MST) with a putative PilZ-containing protein (Bd2717), a two-component regulator (Bd2402) and a putative acyl-CoA dehydrogenase (Bd2924). Both Bd2402 and Bd2924 were validated as novel CdG-interacting proteins lacking a classical CdG-binding motif [82]. Only 3 of the 84 CdG binders identified from the study were investigated further so the risk of false positives in the dataset still remains.

Interestingly many of the novel CdG binders have putative roles in central metabolism (such as Bd2924), which suggests CdG may also regulate central metabolism. However, the reliance of co-factors (FAD, CoA, NAD) in this group of enzymes may have influenced non-specific binding. Another group of novel CdG binding proteins are components of the flagellum. CdG is known to negatively affect flagellum motility in other organisms [83]. In *Bdellovibrio* flagellar biosynthesis is coupled to CdG signalling, although no flagellar apparatus has been reported to interact directly with CdG. It is interesting to speculate that CdG interactions with flagellar components, such as flagellin (Bd0604) and hook protein (Bd3395), could inactivate flagellar activity upon invasion and its cause the flagellum to be shed [26].

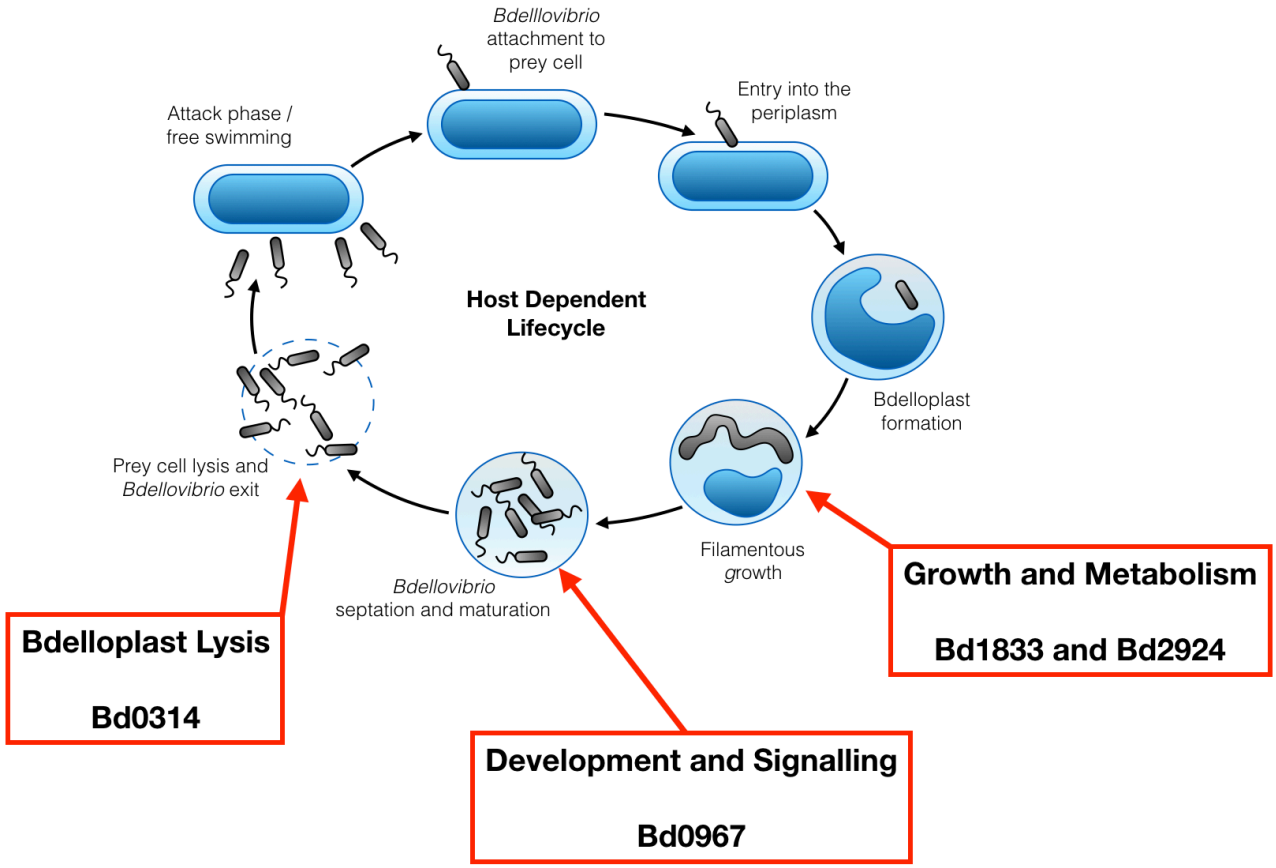


Figure 3. Proteins of interest
This Figure pinpoints the lifecycle events where the proteins of interest, for this thesis, are believed to function.

1.3 Overall Research Aim Of This Study

The overall aim of this thesis was to better understand the predatory lifecycle of *Bdellovibrio bacteriovorus.* In particular, the work carried out for this thesis was concerned with three lifecycle events:

Growth and metabolism, which has received relatively little attention in the past, even though many metabolic pathways have been adapted to facilitate the predatory lifecycle. As such this thesis has targeted two central metabolic enzymes believed to interact with the secondary messenger CdG. The first, **Bd2924**, is a putative acyl-CoA dehydrogenase (ACAD), an enzyme family that facilitates the initial step in fatty acid β-oxidation. The second, **Bd1833**, is a putative multi-modular malic enzyme (ME) possessing a presumed non-catalytic phosphotransacetylase (Pta) accessory domain that regulates ME catalytic function. MEs catalyse the reversible decarboxylation reaction of L-Malate to Pyruvate, which subverts the TCA cycle in a process linked to generating reducing power and carbon units for gluconeogenesis/lipid biosynthesis. Understanding these enzyme's roles will uncover how central metabolism is controlled during the predatory lifecycle.

Development and predator signalling in *Bdellovibrio* concerns multiple aspects that control the transition from one lifecycle stage to the next. CdG has been shown to be a master regulator of *Bdellovibrios* predatory lifecycle. Additionally, new potential CdG-sensing proteins have been identified including **Bd0967.** Bd0967 is a putative carboxy-terminal processing protease (CTP) of unknown function. In other bacterial systems

CTPs are implicated in complex signalling cascades, controlling important lifecycle events such as photosynthesis activation in *S. obliquus* [84], spore formation in *Bacillus* [85], immune system evasion by *Chlamydia* [86] and cell wall maintenance in *E. coli* [87]. Bd0967 is therefore an exciting target to investigate and likely has an important signalling role in the lifecycle of *Bdellovibrio.*

Bdelloplast lysis is the final step of the predatory lifecycle, requiring an enzyme(s) that is expressed at the end of the predatory lifecycle, whose activity is tightly controlled and is able to function upon the modified prey cell wall. The work presented herein has focussed on investigating **Bd0314**, a putative lysozyme-like enzyme, as the "exit enzyme" required to facilitate the final lytic step and lyse the bdelloplast at the end of the predatory lifecycle.

Chapter 2

Materials and Methods

2.1 General

2.1.1 Molecular cloning

A restriction free cloning method was adopted, during the course of this study, to incorporate genes of interest into bacterial expression plasmids [88]. One set of primers was sufficient to clone a gene of interest into any plasmid at a desired location without the need for restriction enzymes. Primers were designed to possess both a complementary gene flanking region and a complementary plasmid flanking region. A list of primers and plasmids used throughout this study can be found in Table 1.

Two rounds of PCR are required: the first to amplify a region from genomic DNA (and thus create a mega primer) and the second to amplify the plasmid (incorporating the mega primer into a specific site with the plasmid).

First Round:

First round PCR reaction mixture (100 µl reaction, made from the following stocks): 20 µl 5 x High-Fidelity buffer (Bioline Bio-21098), 2 µl Forward primer (10 µM stock), 2 µl Reverse primer (10 µM stock), 4 µl dNTPs (10 mM), 1 µl genomic DNA (HD100, 100 ng/µl), 69.5 µl H_2O, 1.5 µl Velocity DNA polymerase (Bioline Bio-21098, 2U/µl).

First round PCR protocol:

Step	Temperature (°C)	Time	Cycles
1	98	4 min	1
2	98	30 s	30
	55	30 s	
	72	4 min	
3	72	8 min	1
4	4	∞	1

The PCR product was run on a 1% agarose TBE (1x Tris-Borate-EDTA buffer) gel alongside a 1kb ladder to evaluate the PCR success. Products containing a single band running at the correct length were purified using a Qiagen QIAquick PCR purification kit.

Second Round:

Second round PCR reaction mixture (100 µl reaction split into 3x33 µl tubes, made from the following stocks): 20 µl 5 x High-Fidelity buffer (Bioline Bio-21098), 4 µl 1st round product, 4 µl dNTPs (10 mM), 1 µl plasmid DNA (pET28a, pET28b-MBP, pET41c) (20ng/µl), 69.5 µl H_2O, 1.5 µl velocity DNA polymerase (Bioline Bio-21098, 2U/µl).

Second round PCR protocol:

Step	Temperature (°C)	Time	Cycles
1	98	4 min	1
2	98	30 s	35
	65 - 55	30 s	
	72	10 min	
3	72	10min	1
4	4	∞	1

PCR products were then incubated with 1 µl DpnI for 1 hour at 37°C to digest methylated parental plasmid DNA. The PCR products were then purified and concentrated <5 µl using a hetovac. The concentrated PCR products were then

transformed into competent XL10-GOLD cells and plated onto LB agar plates supplemented with the antibiotic.

2.1.2 Colony PCR

Colony PCR was performed to identify colonies that had taken up plasmids containing the correct insert. Colonies were picked and spotted onto an agar plate marked with a grid and then dipped into the PCR reaction mixture.

Colony PCR reaction mixture (30 µl reaction, made from the following stocks): 1 µl pET Forward primer (10 µM), 1 µl pET Reverse primer (10 µM), 15 µl MyTaq Red mix (Bioline Bio-25043), 13 µl H_2O.

Colony PCR protocol:

Step	Temperature (°C)	Time	Cycles
1	98	5 min	1
2	98	30 s	35
	50	30 s	
	72	10 min	
3	4	∞	1

2.1.3 Mutagenesis

Quikchange PCR:

Primers were designed following the guidance of the Agilent Quikchange SDM protocol, whereby the mutation was incorporated into the centre of the primers. A list of Quikchange primers can be found in Table 2.

Quikchange Reaction Mixture (100 µl reaction, made from the following stocks): 20 µl 5x High-Fidelity buffer (Bioline Bio-21098), 2 µl Forward primer (10 µM), 2 µl Reverse primer (10 µM), 4 µl dNTPs (10 mM), 1 µl Template DNA (50 ng/µl), 69.5 µl H_2O, 1.5 µl Velocity DNA polymerase (Bioline Bio-21098, 2U/µl).

Quikchange PCR protocol:

Step	Temperature (°C)	Time	Cycles
1	98	1 min	1
2	98	30 s	18
	60	30 s	
	72	8 min	
3	4	∞	1

PCR products were incubated with 1 µl DpnI for 1 hour and finally concentrated to 1-5 µl before being transformed into XL10-GOLD competent cells. Transformed cells were plated onto antibiotic selection LB agar plates and incubated O/N at 37°C.

2.1.4 Q5 SDM

Q5 SDM (NEB) was used as the primary mutagenesis technique. The online NEBaseChanger website was used to generate primers, which are shown in Table 2. The Q5 Site-Directed Mutagenesis Kit protocol was followed as for manufacturers recommendations (NEB) with adjustments, which are outlined below.

Q5 SDM reaction mix (25 µl): 12.5 µl Q5 HotStart HF 2x Master Mix, 1.25 µl Forward primer (10 µM), 1.25 µl Reverse primer (10 µM), 1 µl Template DNA (10 ng/µl), 9 µl H_2O.

Q5 SDM PCR protocol:

Step	Temperature (°C)	Time	Cycles
1	98	30s	1
2	98	10 s	25
	Ta* (primer dependent)	30 s	
	72	4 min	
3	72	2 min	1
3	4	∞	1

A KLD (Kinase, Ligase, DpnI) reaction was performed on the PCR products. The reaction mixture included: 1 µl Q5 SDM product, 1 µl T4 Ligase Buffer, 6 µl H_2O, 0.5 µl DpnI (NEB #R0176S), 0.75 µl T4 Polynucleotide Kinase (NEB #M0201S), 0.75 µl T4 ligase (NEB #R0176S).

The reaction mixture was incubated at room temperature for 1 hour. After incubation the full 9.5 µl reaction mixture was transformed into 100 µl XL10-GOLD competent cells and plated out onto antibiotic selection LB agar plates.

2.1.5 Primers

The primers used throughout this work are displayed in Table 1 and 2. All primers were custom oligos supplied by Sigma-Aldrich.

Protein	Plasmid	Primer (Fwd) (5' - 3')	Primer (Rev) (5' - 3')
Bd0314	pET28b-MBP	GAGAACCTGTACTTCCAATCCGGATCCCGCGA AGTC ATTCCTCTTTG	GGCTTTGTTAGCAGCCGGATCTCACTTGCAAA TAGACAGATTCTGGC
Bd1413	pET28a	GCAGCGGCCTGGTGCCGCGCGGCAGCCATGG CTATAAGATGCAGGCCCTGG	CTCAGTGGTGGTGGTGGTGGTGCTCGAGCTAT TTGCACAAGGGAAGAGTCTTGACG
Bd1440	pET28a	GCAGCGGCCTGGTGCCGCGCGGCAGCCATAG CTACAAAATAGAGGCCCTGTCC	CTCAGTGGTGGTGGTGGTGGTGCTCTTACTTA CAGAAGGATAATTTTTTGACCATGCC
Bd2924	pET28a	AGCAGCGGCCTGGTGCCGCGCGGCAGCCATA TGAAAAACTTTTACCAAGACGGc	CTCAGTGGTGGTGGTGGTGGTGCTCGAGCTAT AAAGAATACTTCTTCATAGCGATACGG
Bd1833	pET28a	GCAGCGGCCTGGTGCCGCGCGGCAGCCATAT GGACAACAAAACTGAAACTAAAATTGAAC	CTCAGTGGTGGTGGTGGTGGTGCTCGAGTTAC TTTTTGCCACGGCTTTTCAG
Bd1833-Pta	pET28a	GCAGCGGCCTGGTGCCGCGCGGCAGCCATTC CAAAGTCTTCATCCGCTCTG	CTCAGTGGTGGTGGTGGTGGTGCTCGAGTTAC TTTTTGCCACGGCTTTTCAG
Bd1833-ME	pET28a	GCAGCGGCCTGGTGCCGCGCGGCAGCCATAT GGACAACAAAACTGAAACTAAAATTGAAC	CTCAGTGGTGGTGGTGGTGGTGCTCGAGCTA AGATTCGCGGTACTGATCCC
Bd1833(SKB)	pET28a	GCAGCGGCCTGGTGCCGCGCGGCAGCCATAT GCAGCGAAAGGAAATATCC	CTCAGTGGTGGTGGTGGTGGTGCTCGAGTTAT TTGGTCAGTTCCGATTTTTTAGC
Bd1833(SKB)-Pta	pET28a	GCAGCGGCCTGGTGCCGCGCGGCAGCCATTC CAAAGTCTTCATTCGCTCTGCC	CTCAGTGGTGGTGGTGGTGGTGCTCGAGTTAT TTGGTCAGTTCCGATTTTTTAGC
Bd0967	pET28a	GCAGCGGCCTGGTGCCGCGCGGCAGCCATAT GAGTTCATTCTCAAAAGGCGTTAGAGG	CTCAGTGGTGGTGGTGGTGGTGCTCGAGCTG AGGCACCAAGGATTTCCC
Bd0606	pET28a	GCAGCGGCCTGGTGCCGCGCGGCAGCCATAT GGGAATGAGAATTTCTACG	CTCAGTGGTGGTGGTGGTGGTGCTCGAGTTA ACCAATCAGTCTTAGTGC

Table 1. Gene primers
A list of primers used in restriction free cloning.

Protein	Mutation	Type	Primer (Fwd) (5' - 3')	Primer (Rev) (5' - 3')
Bd0314	E143Q	QuickChange	ATGGTCAGATTTCAGAGTAACTTTAAAACC	GGTTTTAAAGTTACTCTGAAATCTGACCAT
Bd0314	E154Q	Q5	GTCCTACACCCAGGATTTTAAC	ATTGCGGTTTTAAAGTTAC
Bd0314	Y228F	Q5	TGGCGCCCGCGCCTGGTCTGTG	CCTTTCCAGGCTCCATCC
Bd0314	M150A	Q5	TAAAACCGCAGCGTCCTACACC	AAGTTACTCTCAAATCTGAC
Bd1440	E157Q	Q5	ACGCATGCATCAAACCACCATG	GAAGTCGGGCTGTAAGCG
Bd0967	S547A	Q5	TATCTCTGCTGCCGCTTCCGA	CGGCTGGTCAATACCACC
Bd0967	R299M	Q5	CAAAAAAATCATGGGTAACAAGGGCACGAAAGTG	ACAACGTCTTTCAAATCC
Bd1833	R535A	Q5	GTATTCTTTGGCCCAGCGTAAAGG	AGCTTTTCAACGAAGCTG
Bd1833	R535E	Q5	GTATTCTTTGGAGCAGCGTAAAGGC	AGCTTTTCAACGAAGCTG
Bd1833	Y576C	Q5	TCCATCAACTGTGCCGATGCG	GGATCCACTGACCATACC
Bd1833	Y576S	Q5	TCCATCAACTCTGCCGATGCG	GGATCCACTGACCATACC
Bd1833	E544R	Q5	CAATCTGGGCAGAGCCGAGCGC	ATGCCTTTACGCTGGCGC
Bd1833	E544A	Q5	CAATCTGGGCGCAGCCGAGCGCT	ATGCCTTTACGCTGGCGC
Bd1833	N718D	Q5	GGAATCCTCCGACATCGCCTA	AGATTCGGGAAGACCAGC
Bd1833	ΔLoop	Q5	CTTAACGAGAACAACCTGATCCCGAAACCCTTC	GTCTTGCTCTTTGATCAGTTTTGCCAGTGCATG
Bd1833	R547A	Q5	CGAAGCCGAGGCCTTGATGGCGG	CCCAGATTGATGCCTTTAC
Bd1833	R749A	Q5	CGTTCTGCAAGCAACCACAACGGTTGATGGC	TTGGCAGAGCGGCGCACA
Bd1833	R581K	Q5	CGATGCGGTTAAGCCGATTCTGC	GCATAGTTGATGGAGGATC
Bd0606	LIG	Q5	TAACTCGAGCACCACCAC	TCTTAGTGCAGAGTTTGGC
Bd0606	RLIG	Q5	TAACTCGAGCACCACCAC	TAGTGCAGAGTTTGGCATC

Table 2. SDM primers
A list of primers used for SDM

2.1.6 Sequencing

Cloning success was evaluated by DNA sequencing. Samples (200 ng of plasmid and 3.2 pmol of sequencing primer in 10 µl solution) were submitted to the DNA sequencing service ran by the University of Birmingham. The plasmids were run a capillary sequencer ABI 3730 to produce read lengths up to 1000 bp. Sequencing reads in the forward and reverse directions were required to cover larger sequence reads.

2.1.7 Transformation

The Qiagen bench guide Protocol 1 (Preparation of competent *E. coli*) was followed to produce competent: XL10-GOLD cells (for high efficiency DNA amplification) and BL21*(DE3) cells (for recombinant gene expression). Commercial competent SHuffle cells (for recombinant gene expression and disulphide bond formation in the cytoplasm) were purchased from NEB.

Transformation involved incubating 1-5 µl of plasmid DNA incubated on ice for 30 minutes with 100 µL of competent cells. After incubation the cells were heat shocked at 42°C for 1 minute and then placed back on ice for a further 5 minutes. 1 ml of LB media was then added to the cell mixture and incubated at 37°C for 2 hours before being plated on the relevant antibiotic selection LB agar plate.

2.1.8 Culture Media

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

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

2.1.9 Antibiotic Stocks

Ampicillin – 100 mg/ml (in H_2O)

Kanamycin – 100 mg/ml (in H_2O)

Chloramphenicol – 25 mg/ml (in ethanol)

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

2.1.10 Glycerol Stocks

Transformed cells were plated on antibiotic selection plates and incubated O/N at 37°C. A 5 ml LB solution, supplemented with antibiotic, was inoculated with a single colony picked from the O/N plate and incubated in a shaking incubator O/N (set at 180 rpm and 37°C). 500 µl of O/N culture was extracted and mixed with 500 µl of 40% v/v glycerol solution and then stored at -80°C.

2.1.11 Expression

A sterile pipette tip was used to extract a small amount of glycerol stock, which was used to inoculate an LB media starter culture (supplemented with antibiotic). The starter culture was grown O/N in a shaking incubator (set to 180 rpm and 37°C). 10 ml of starter culture was used to inoculate 1 L of expression media supplemented with antibiotic (in a 2.5 L baffled flask). The expression media used varied between proteins (noted in individuals expression section). Inoculated expression media was incubated at 37°C and shaken at 180 rpm until OD_{600} reached 0.8. Cultures at OD_{600} 0.8 were induced with 1mM IPTG and then incubated O/N in a shaking incubator (set at 180 rpm

and 18°C). Cells were harvested by centrifugation in a JLA-8.1000 rotor at 6000 rpm (6891 RCF) for 10 minutes. The supernatant was discarded and the pellets were scraped into 50 ml Corning tubes (~20g per tube) and frozen at -20°C.

2.1.12 Purification

Cell pellets were resuspended with appropriate amounts (~40 ml per 10 g) of fresh lysis buffer (50 mM HEPES pH 7.5, 20 mM PEG pH 7.5, 300 mM NaCl, 0.05% v/v Tween20) and incubated with ~1mg/ml Lysozyme for 30minutes (at 4°C). Cells were lysed using a Fisher ScientificTM Model 505 Sonic Dismembrator (set to 40%). Seven cycles of 30s bursts separated by 2 minute rests on ice was sufficient to effectively lyse the cells. Insoluble cell debris was removed by centrifugation for 1 hour at 20000 rpm (32826 RCF) using a JA-25.50 rotor. The soluble material was decanted into a new container and stored on ice and the insoluble material was discarded. The soluble material was passed over a 1 ml or 5 ml HisTrap HP column (GE Healthcare life science), pre-equilibrated with lysis buffer. A peristaltic pump set to a 1 ml/min flow rate was used for purifications. The column was washed with 20 column volumes (CV) of lysis buffer to remove non-specific interacting proteins. The column was then eluted with 10 column volumes of elution buffer (50 mM HEPES pH 7.5, 300 mM NaCl, 400 mM PEG pH 7.5) into 2 CV fractions. The fractions were analysed by SDS-PAGE to determine purity and assess quantity of target protein in each fraction before the fractions containing the target protein were pooled (see individual protein methods section for variations on purification).

2.1.13 Dialysis

Pooled protein was transferred to 10kDa MWCO dialysis tubing and dialysed into 2L of dialysis buffer (differs for individual proteins), for 3 hours at 4°C. The dialysis tubing was then transferred to a fresh 2L of dialysis buffer, O/N at 4°C to fully equilibrate.

2.1.14 Size Exclusion Chromatography

Size exclusion chromatography (SEC) was carried out using a 26/600 Superdex 200 prep grade column (GE Healthcare) attached to an AKTA-purifier system operated using Unicorn software. A 26/600 Superdex 200 column with a volume of 320 ml was used for high resolution of proteins with molecular weights between 10 kDa and 600 kDa. The column was pre-equilibrated with 1.5 CV of protein dialysis buffer (differs for individual proteins). The protein samples were loaded onto the column and ran at a flow rate of 2 ml/min. Absorbance readings at wavelength 280 nm (A_{280}) was used to monitor elution from the column.

2.1.15 SDS-PAGE

SDS polyacrylamide gels were prepared using a BioRad gel-casting apparatus using the following reagents:

- Resolving buffer – 200 ml (18.2g Tris pH 8.5, 0.4g SDS)
- Stacking buffer - 100 ml (15.2g Tris pH 6.8, 1g SDS)
- Running buffer – 1 L (30.0g Tris pH 8.3, 144g of glycine, 10g SDS)
- Loading dye - (100 mM Tris pH 6.8, 4% w/v SDS, 0.2% w/v Bromophenol blue, 20% v/v glycerol, 200 mM β-mercaptoethanol)
- Protein marker - NEB Color Prestained Protein Standard (11–245 kDa)
- Coomassie Stain – (50% v/v methanol, 10% v/v acetic acid, 0.2% w/v Coomassie Brilliant Blue)
- Destain – (40% v/v methanol, 10% v/v acetic acid)

Resolving Gel:

- 15 ml Resolving buffer
- 15 ml 30% w/v acrylamide
- 15 µl TEMED
- 300 µl 8% w/v Ammonium persulfate (APS)

Stacking Gel:

- 15 ml H_2O
- 4.5 ml 30% w/v acrylamide
- 2.25 ml stacking buffer
- 11.5 µl TEMED
- 225 µl 8% w/v APS

Sample preparation and running the gel:

Samples were prepared by adding 20 µl 2xSDS loading dye to 20 µl protein sample and then heated at 100°C for 10minutes. 15 µl of boiled sample was loaded into the wells alongside a protein marker (NEB prestained protein marker). Gels were run at a constant 170 V until complete and then removed and stained with Coomassie stain for 20 minutes and subsequently destained in destain solution O/N. Images of the gels were taken using UVITECH gel imager.

2.1.16 Protein sample concentration and quantification

Protein samples were concentrated using Vivaspin 20 spin concentrators, centrifuged at 3220 *g* at 4°C in an Eppendorf benchtop centrifuge. Samples were concentrated to 10-20 mg/ml (protein dependent). The protein concentration was assessed using a DeNovix DS-11 FX nanodrop, measuring the A_{280}. The molecular mass and molar absorbance (extinction coefficient) used for concentration calculation was theoretically calculated using an online tool (ProtParam by Expasy).

2.1.17 Protein crystallisation

Pure protein (between 10-20 mg/ml) was used to set up crystallisation screens using the sitting drop vapour diffusion method (at 18°C). Crystallisation screening was set up by hand in 96 well CombiClover™ crystallisation plates (2 µl protein + 2 µl reservoir solution per drop) or using a Mosquito HTS liquid handling robot (300 nl protein + 300 nl reservoir solution per drop). A range of commercial crystallisation screens (purchased from Molecular Dimensions) were used to obtain initial crystallisation conditions: JCSG+ [89], BCS [90], PACT [91], Proplex [92], MIDAS [93], Morpheus [94]. A grid screen around the original condition was then used to improve crystal quality. Once crystals had developed they were transferred to cryo-protection solution (mother liquor + 20% v/v ethylene glycol or 20% v/v glycerol) and incubated for 5 minutes before snap freezing in liquid nitrogen. Cryo-cooled samples were then stored in cryo-cooled pucks.

2.1.18 Data collection and structure solving

Diffraction data were collected at Diamond light source in Oxford, UK. Throughout this study Beamlines IO3, IO4 or IO4-I were used and datasets were collected in a routine manner (images = 3600, oscillations = 0.1° Ω, transmission = 50-100%, exposure = 0.05s, wavelength = 0.9763Å). The resolution limit and beamsize was adjusted depending on crystal size and diffraction quality. The diffraction data were autoprocessed by ISPyB automated pipelines (Xia2 [95] and autoproc [96]); the datasets with best statistics (resolution, completeness, multiplicity, I/σI, Rmerge, Rpim, Rmeas, $CC_{(1/2)}$ [97-100]) were used for structure solution.

Molecular replacement (MR) was used as the principle technique to solve protein structures. Initial electron density maps and protein structure models were generated using the MR program Phaser [101]. The automated MR mode of Phaser required the following inputs: the reflection data (.mtz file), an ensemble (model of structure used in search), the sequence of protein and the number of copies to search for in the ASU (number of copies were determined using the Mathews Coefficient). Solutions with high LLG (Log likelihood gain) (LLG>40) and TFZ (translation function z-score) (TFZ>8) values were investigated manually to determine a correct solution. Initial models from molecular replacement were edited, using coot [102], to remove ambiguous regions and manually build traceable regions. In cases were poor initial maps and models were generated by MR Phenix.autobuild [103] was used to automatically improve phases and the initial model. Models were improved by alternating rounds of manual model building using coot and refinement using Phenix.refine and PDB-redo [104]. The final models were validated using MolProbity software through the Phenix GUI [105]. Validation statistics (R_{free}, R_{work}, B-factors, $RMS_{(bonds)}$, $RMS_{(angles)}$, Ramachandran plot [106, 107]) were analysed to assess the goodness of the model.

2.2 Bd0314

2.2.1 Cloning, Expression and Purification

The secreted catalytic domain of Bd0314, starting at R74 and ending at K254 was cloned into a modified pET28b-MBP expression plasmid (His_6-MBP tag), truncating the disordered N-terminal region and signal peptide. Bd0314 was placed in frame directly after a TEV protease cleavage site and MBP tag sequence. Bd0314 expression plasmid

was transformed into the commercially available *E. coli* SHuffle T7 expression strain (NEB). SHuffle cells were required to correctly form Bd0314's two disulphide bonds in the cytoplasm, minimising the toxic effects of overexpression. Bd0314 expression procedure was adjusted from the standard expression procedure outlined in Section 2.1.11. Starter cultures were supplemented with 1% glucose w/v to suppress leaky expression. TB media supplemented with 1% glucose w/v was used for expression and cultures were induced at OD_{600} 0.6 with 1mM IPTG.

Bd0314 was purified as outlined in Section 2.1.12 with an additional MBP tag removal step. Fractions containing pure MBP-tagged Bd0314 were pooled and TEV protease (prepared following Tropea et al protocol [108]) was added at a ratio of 1:50 (TEV:Bd0314). The sample was then dialysed O/N at 4°C in (20 mM HEPES pH 7.5, 300 mM NaCl), during which time the TEV protease activity worked to remove the MBP tag. After O/N dialysis and incubation with TEV, cleaved Bd0314 was separated from the MBP fusion tag and residual MBP-Bd0314 by a second passage over a 5 ml HisTrap HP column pre equilibrated in dialysis buffer. Cleavage efficiency and final purity was assessed by SDS-PAGE before the protein was concentrated to ~15 mg/ml. Expression and purification of Bd0314 mutants (E143Q, E154Q, Y228F) was performed using the same methods as Bd0314-Wt.

2.2.2 FITC Peptidoglycan digestion assay

Isolated peptidoglycan was labelled with FITC (Fluorescein-5-isothiocyanate) in a similar method to that described by H. Maeda, et al [109]. 28 mg PG was incubated with 8 mg FITC in 300 µl H_2O for 4 hours. FITC labelled PG (normal) was washed, dried,

aliquoted and stored at -20°C. A portion of FITC labelled PG was deacetylated using the HD100 GlcNAc deacetylases (Bd0468 and Bd3279). The GlcNAc deacetylation reaction was performed by resuspending FITC labelled PG in deacetylation buffer (20 mM HEPES pH 7.5, 300 mM NaCl and 1 mM $ZnCl_2$) and then incubating with 1 µM of Bd0468 and Bd3279, for 12 hours at 37°C. The deacetylated FITC labelled PG was washed, dried, aliquoted and stored at -20°C.

Assays were performed in a standard assay buffer (20 mM HEPES pH 7.5, 300 mM NaCl) for all assayed lysozymes (HEWL, Mutanolysin and Bd0314). The FITC labelled PG substrates (GlcNAc deacetylated and normal) were prepared by resuspending in assay buffer to a stock concentration of 2 mg/ml. The assay was performed by mixing: 1 µL enzyme (1 µM), 5 µL FITC labelled PG (100 µg/ml) and 94 µL assay buffer, which was incubated at 37°C for 1 hour. The FITC-PG assay is an endpoint assay, which was quenched by filtration through 0.1 micron filters, removing the insoluble PG substrate. The reaction products (liberated FITC) then had the pH adjusted by mixing 30 µL of soluble sample with 70 µL of 1M HEPES pH 8.0. The fluorescence reading for the reaction products was monitored using a PheraStar FS plate reader set to excite at 485 nm and detect at 520 nm. Liberated FITC was a measure of enzyme activity. A no enzyme control was used to determine the baseline fluorescence counts that was negligible. For the time course assay (100 µg/ml PG, 20 mM HEPES, 300 mM NaCl, 10 nM enzyme) was mixed and incubated for up to 60 minutes. Samples were taken at regular intervals and the reaction was quenched by filtration.

2.2.3 Premature bdelloplast lysis assay

The premature bdelloplast lysis assay was used to assay the activity of Bd0314 towards bdelloplasts. The assay was performed collaboratively at the University of Nottingham, under the supervision and guidance of the Sockett lab who have an expertise in manipulating *Bdellovibrio bacteriovorus*.

E. coli S17-1 (pAKF220) were grown O/N in 200 μM IPTG to induce expression of the pMal::mNeon, which fluorescently labelled the periplasm and cytoplasm green. O/N *E. coli* cultures were backdiluted to OD_{600} 1.0 in YT broth. The *E. coli* cells were also treated with 500 μM HCC-amino-D-alanine (HADA) for 30 minutes to fluorescently label their peptidoglycan blue. Labelled cells were washed in Ca/HEPES and then incubated with a 1:1 volume of 10x concentrated *Bdellovibrio bacteriovorus HD100* for 1 hour to achieve synchronous infection. Synchronous infection was required to ensure >99% of the prey rounded to form bdelloplasts.

100 μl samples of bdelloplasts were treated with 10 mM EDTA for 10 minutes, to permeabilise the bdelloplast outer-membrane and facilitate access to the PG layer. After EDTA treatment bdelloplasts were further incubated for 1 hour (control) or treated with 10 μM Bd0314 for 1 hour at room temperature. 10 μl samples were extracted and placed onto 1% agarose Ca/HEPES pads for imaging.

2.2.4 Crystallisation and Cryoprotection

Bd0314 crystals were obtained and handled as outline in Section 2.1.17. Bd0314-Wt crystals were grown in (0.1M sodium acetate pH 4.5, 50% w/v PEG 400) and (0.2M

lithium sulphate, 0.1M sodium acetate pH 4.5, 30% w/v PEG 8000) for the $P2_12_12$ and $P2_12_12_1$ forms respectively. Bd0314-E143Q and Bd0314-E154Q were crystallised in (0.1M sodium citrate pH 5.5, 20% w/v PEG 3000 and 0.2M lithium sulphate).

Extensive ligand soaking and co-crystallisation experiments were conducted with Bd0314-Wt, E143Q and E154Q, which were unsuccessful at obtaining a protein:ligand complex. Four ligands were trialled (glucosamine, chitosan dimer dihydrochloride, chitosan Low molecular weight and chitosan oligosaccharide lactate) at varying concentrations and incubation periods.

2.2.5 Data collection and structure determination

MR failed for Bd0314 structure solution so a phasing dataset was collected at a wavelength of 1.77Å to maximise the anomalous dispersive signal from native sulphur atoms (SSAD). A medium beam size (80x20 μM) and low transmission (19.92%) was used to collect two datasets of 9000 images with the Ω oscillations set to 0.1° at two distant sites on a single crystal. The large dataset was required to increase multiplicity and increase the signal to noise for anomalous signal of native sulphur atoms. The data collection statistics are shown in Table 3.

The unmerged auto-processed data from ISPyB (Multixia2) was extracted and used for experimental phasing using Phenix AutoSol [110]. The search criteria used for experimental phasing was: atoms = "S", wavelength = 1.77Å, f' = 0.36e, f'' = 0.7e, sites = 6. Experimental phasing successfully located 13 sulphur sites but the signal dropped off rapidly after the 6th site (FOM = 0.3, BAYES_CC = 27.9 ± 28.8, R-factor = 0.32, Skew = 0.06, Correlation of local RMS density = 0.74). The experimental data and heavy atom

sites were used in a round of automated model building using Phenix autobuild [111]. The auto-built model successfully traced most of the Bd0314 sequence (2 fragments, 170 residues, R-work = 0.2245, R-free = 0.2446). See Section 2.1.18 for how the initial model was improved. Datasets for Bd0314-Wt and E154Q were collected and solved in a typical process using molecular replacement as outlined in Section 2.1.17 and 2.1.18.

Data Collection				
	Bd0314 Wt (Jcsg A1)	Bd0314 Wt (Jcsg D4)	Bd0314 E143Q	Bd0314 E154Q
Spacegroup	P 2_1 2_1 2	P 2_1 2_1 2_1	P 2_1 2_1 2_1	P 2_1 2_1 2
Cell Dimensions a, b, c (Å)	84.7, 47.8, 49.2	48.9, 64.7, 82.7	48.1, 62.1, 80.9	84.7, 47.9, 49.2
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (Å)	47.818 - 1.361	64.67-1.26	62.13 - 1.82	49.21 - 1.37
Total observations	531095 (14923)	812815 (25973)	968126 (1081)	566104 (29666)
Unique reflections	42769 (1766)	71428 (3469)	19553 (149)	45432 (2275)
Rmerge(I)	0.088 (0.785)	0.0682 (1.786)	0.1612 (0.592)	0.137 (1.713)
Rpim	0.036 (0.384)	0.02 (0.692)	0.021 (0.240)	0.056 (0.694)
I/sigI	16 (2)	13.17 (1.2)	20.1 (2.1)	12.2 (2.2)
CC(1/2)	0.998 (0.804)	0.9994 (0.479)	0.9983 (0.638)	0.994 (0.617)
Completeness (%)	98.1 (82.8)	99.8 (98.4)	87.65 (13.6)	98.4 (100)
Multiplicity	12.4 (8.5)	11.4 (7.5)	49.51 (7.3)	12.5 (13)
Refinement				
R-work	0.1739	0.1699	0.1721	0.1771
R-free	0.1911	0.1940	0.2101	0.1951
RMS (bonds) (Å)	0.008	0.0160	0.006	0.005
RMS (angles) (°)	1.142	1.593	0.789	1.044
Average B-Factor (Å^2)	18.1	21.2	18.2	17.8
# of non-hydrogen	1691	1729	1650	1695
Molecules per ASU	1	1	1	1
Ramachandran Plot				
Preferred region	97.1	98.8	96.6	97.1
Allowed region	2.9	1.2	3.4	2.9

Table 3. Bd0314 data collection and refinement

2.3 Bd1413 and Bd1440

2.3.1 Cloning, Expression and Purification

Bd1413 and Bd1440 constructs were designed similarly to the construct of Bd0314, beginning after the region of disorder. The Bd1413 construct starts at G80 and ends at K271, whereas the Bd1440 construct starts at S75 and ends K266. Both Bd1413 and Bd1440 constructs were cloned into the pET28a vector and transformed into *E. coli* expression strain SHuffle cells. Expression and purification of Bd1413 and Bd1440 was performed using the same procedure as described for Bd0314.

2.4 Bd2924

2.4.1 Cloning, Expression and Purification

The full 505aa sequence of Bd2924 was cloned into a pET28a vector and transformed into *E. coli* expression strain BL21*(DE3) cells. Bd2924 was expressed in TB media as mentioned in Section 2.1.12 and purified to give a bright yellow/orange coloured solution and was dialysed into (20 mM HEPES pH 7.5, 200 mM NaCl). Bd2924 was concentrated to ~20mg/ml and snap frozen with liquid nitrogen before storage at 80°C.

2.4.2 SEC

SEC was performed on Bd2924 to assess the oligomeric state in solution. The procedure is outlined in Section 2.1.14 using (20 mM HEPES pH 7.5, 300 mM NaCl) as the SEC buffer.

2.4.3 MicroScale Thermothoresis – Cyclic-di-GMP Binding Assay

MicroScale Thermothoresis (MST) was used to investigate the binding affinity between Bd2924 and CdG, which was previously reported as 7.67 μM [82]. Bd2924 was labeled with a Monolith NT RED protein labeling kit according to the supplied labeling protocol (NanoTemper Technologies). Bd2924 buffer (20 mM HEPES pH 7.5, 200 mM NaCl) was used as the labeling buffer. Bd2924 was used at a final concentration of 50 nM in the binding experiments and CdG (InvivoGen) was used at a starting concentration of 1mM and diluted 1:1 v/v with assay buffer in a serial dilution for 16 capillaries. The experiments were performed in CdG binding buffer (20 mM HEPES pH 7.5, 300 mM NaCl) or (10 mM Tris, 5 mM $MgCl_2$, 250 mM NaCl, 10% v/v glycerol, 1 mM dithiothreitol pH 8.0 supplemented with 0.05% Tween) (conditions replicated from published studies [82]). All binding reaction mixtures were incubated for 10 minutes at room temperature before they were loaded into standard treated capillaries (NanoTemper Technologies). The measurements were done with a NanoTemper Monolith NT.115 instrument. All measurements were performed with MST power 20%, excitation power 40%, 30s laser on time and 5s laser off time. All experiments were performed with at least three experimental replicates.

2.4.4 Crystallisation and cryoprotection

Crystals of Bd2924 were grown and handled as outline in Section 2.1.17. Crystals used to solve the structure of native Bd2924 was harvested from Midas-C6 (40% v/v pentaerythritol propoxylate (5/4 PO/OH), 15% ethanol). Crystals used for Bd2924-C10CoA structure were grown in Pact-B3 (0.1M MIB (sodium malonate dibasic monohydrate, imidazole, boric acid) pH 6.0, 25% PEG 1500) and then transferred to a

solution of mother liquor supplemented with 20% v/v glycerol, 400 μM C10-CoA and 10 μM FAD (Sigma-Aldrich) and incubated for 30 minutes before harvesting.

2.4.5 Data collection and structure determination

Bd2924 datasets were collected and solved routinely; see Section 2.1.18. For MR, 1 copy of a ACDH-11 dimer (PDB:4Y9L) was searched for in the ASU, giving a solution with a TFZ of 22.2 and LLG of 656 (suggesting a correct solution). The Bd2924 C10-CoA soaked dataset was also solved by MR using the native structure as the search model.

Data Collection		
	Bd2924 native	Bd2924 (C10-CoA)
Spacegroup	C 2	C 2 2 2_1
Cell Dimensions a, b, c (Å)	137.8, 113.0, 89.0	126.3, 134.7, 133.2
α, β, γ (°)	90, 120.8, 90	90, 90, 90
Resolution (Å)	81.71 - 1.51	133.18 - 1.87
Total observations	1198412 (53307)	1226302 (51988)
Unique reflections	177144 (8571)	93777 (4648)
Rmerge(I)	0.0508 (1.2811)	0.073 (0.860)
Rpim	0.021 (0.555)	0.030 (0.375)
I/sigI	11.16 (1.12)	19.3 (2.1)
CC(1/2)	0.9986 (0.7687)	0.999 (0.833)
Completeness (%)	96.87 (93.85)	100 (99.6)
Multiplicity	6.77 (6.22)	13.1 (11.2)

Refinement		
R-work	0.1545	0.1849
R-free	0.1914	0.2167
RMS (bonds) (Å)	0.018	0.008
RMS (angles) (°)	1.79	1.222
Average B-Factor ($Å^2$)	31.0	32.1
# of non-hydrogen atoms	8572	8909
Molecules per ASU	2	2
Ramachandran Plot (%)		
Preferred region	98	98.2
Allowed region	1.5	1.3

Table 4. Bd2924 data collection and refinement

2.5 Bd1833

2.5.1 Cloning, Expression and Purification

The gene for Bd1833 had been mis-annotated on NCBI and KEGG databases, due to an error in sequencing reads, which results in a premature stop codon. Consequently Bd1833 is annotated as two gene fragments (Bd1833 and Bd1834). The construct of full-length Bd1833 was designed to start at the start codon of *bd1834* and finish at the stop codon of *bd1833*. Bd1833/Bd1834 was successfully cloned into pET28a expression vector and its nucleotide sequence was confirmed by DNA sequencing. DNA sequencing of the Bd1833/Bd1834 cloned region confirmed Bd1833 exists as a single gene and the original stop codon was incorrectly placed due to a frame shift in sequencing reads (Figure 4).



Figure 4. Bd1833 misannotation
bd1833/bd1834 gene is mis-annotated as two genes on NCBI; in fact *bd1833/bd1834* encode a single protein – Bd1833. A frameshift had occurred during the original sequencing producing a premature stop codon and annotation as two gene products. The region starting from *bd1834* start codon through to the stop codon of *bd1833* was cloned and re-sequenced. Re-sequencing revealed the correct sequence read through this region and the position of the frame shift (highlight in red), which in turn makes the stop codon (highlighted in red).

The full-length Bd1833 construct (encoding M1 – K780) was cloned into the pET28a expression plasmid with N-terminal His_6Tag and transformed in BL21* (DE3) as described in Section 2.1. Two truncation constructs of Bd1833 were created, Bd1833-ME truncation (M1 – A434) and Bd1833-Pta (S439 – K780), which were both cloned into pET28a. Expression and purification for Bd1833 and the truncation mutants was performed as described in Section 2.1 up until elution from the HisTrap column. Bd1833 and Bd1833-ME required the addition of 5 mM β-mercaptoethanol and 5 mM $MgCl_2$ to the post elution fraction to stabilize the protein and keep cysteine residues reduced. Bd1833 and Bd1833-ME were dialysed against (20 mM BisTris pH 6.0, 75 mM KCl, 5 mM $MgCl_2$, 2 mM DTT), whereas Bd1833-Pta did not require any supplementation of $MgCl_2$ or DTT and was dialysed against (20 mM HEPES pH 7.5, 300 mM NaCl). Bd1833 was concentrated to ~15 mg/ml, Bd1833-ME to ~20 mg/ml and Bd1833-Pta to 30mg/ml using spin concetrators. All Bd1833 mutants were expressed and purified as described for WT Bd1833.

2.5.2 SEC

SEC was performed for Bd1833 and mutants to polish the protein for crystallographic studies and to investigate the oligomeric composition in solution. SEC analysis was performed as described in Section 2.1.14, in a buffer matching the proteins dialysis buffer.

2.5.3 CoA removal Wash Steps

CoA and acetyl-CoA co-purify with Bd1833, tightly interacting with residues located in the Pta domain. In order to remove the ligands an additional mild denaturing wash step

was used. Two mild denaturing buffers (2 M urea, 2.5 M KCl, 50 mM HEPES pH 7.5 and 2.5 M urea, 2.5 M KBr, 50 mM HEPES pH 7.5) and two procedures were used in this investigation. The protein was bound to a 1 ml HisTrap column and then washed with 20 CV of lysis buffer. After contaminants were removed the buffer was exchanged to the mild denaturing buffer and ran at a flow rate of 1 ml/minute for 50 or 100 CV. After the denaturing wash step another 20 CV of lysis buffer was passed over the column before the column was eluted with 10 CV of elution buffer. The eluted samples were pooled and dialysed against the appropriate dialysis buffer. The efficiency of the ligand removal procedure was determined by HPLC analysis. The most efficient wash step (2.5 M urea, 2 M KBr, 50 mM HEPES pH 7.5 buffer for 100 CV) was used for all future ligand removal wash steps.

2.5.4 HPLC

In order to determine whether the co-purified and co-crystallised ligand in the Bd1833 structures was acetyl-CoA or CoA, a modified version of a previously described HPLC method for CoA and acetyl-CoA detection was used [112]. The protein samples were prepared for HPLC analysis by boiling 100 µM protein samples for 10 minutes and then centrifuged at highspeed (20,000 *g*) to remove insoluble material (denatured protein) and the supernatant transferred into HPLC micro-sampling vials (Thermo Scientific). An Ultimate3000 UHPLC (ThemoFisher Scientific) equipped with a kinetex 150 x 2.1 mm C18 column was used for HPLC analysis. The detection procedure was set up as follows: 5 µl injection, 0.175 ml/min flow rate, 20 minute run time, column oven 25°C, UV detection 259 nm. The mobile phase used was 100 mM monosodium phosphate and 75 mM sodium acetate adjusted to pH 4.6 with phosphoric acid. HPLC grade

acetonitrile was added at a ratio of 6 (acetonitrile) to 94 (mobile phase) to prepare the mobile phase before use. CoA and acetyl-CoA gave distinct and well-separated peaks using this method, eluting at ~3.5 ml and ~7 ml respectively (Figure 5). Baseline shifts were observed when new mobile phase was prepared; consequently a standard was always run with each batch of sample tested. Sample standards, ranging from 10 μM to 150 μM CoA and acetyl-CoA were ran together and separately in order to determine the location of the peak. The area under the peak for the standards was calculated to calibrate the system.

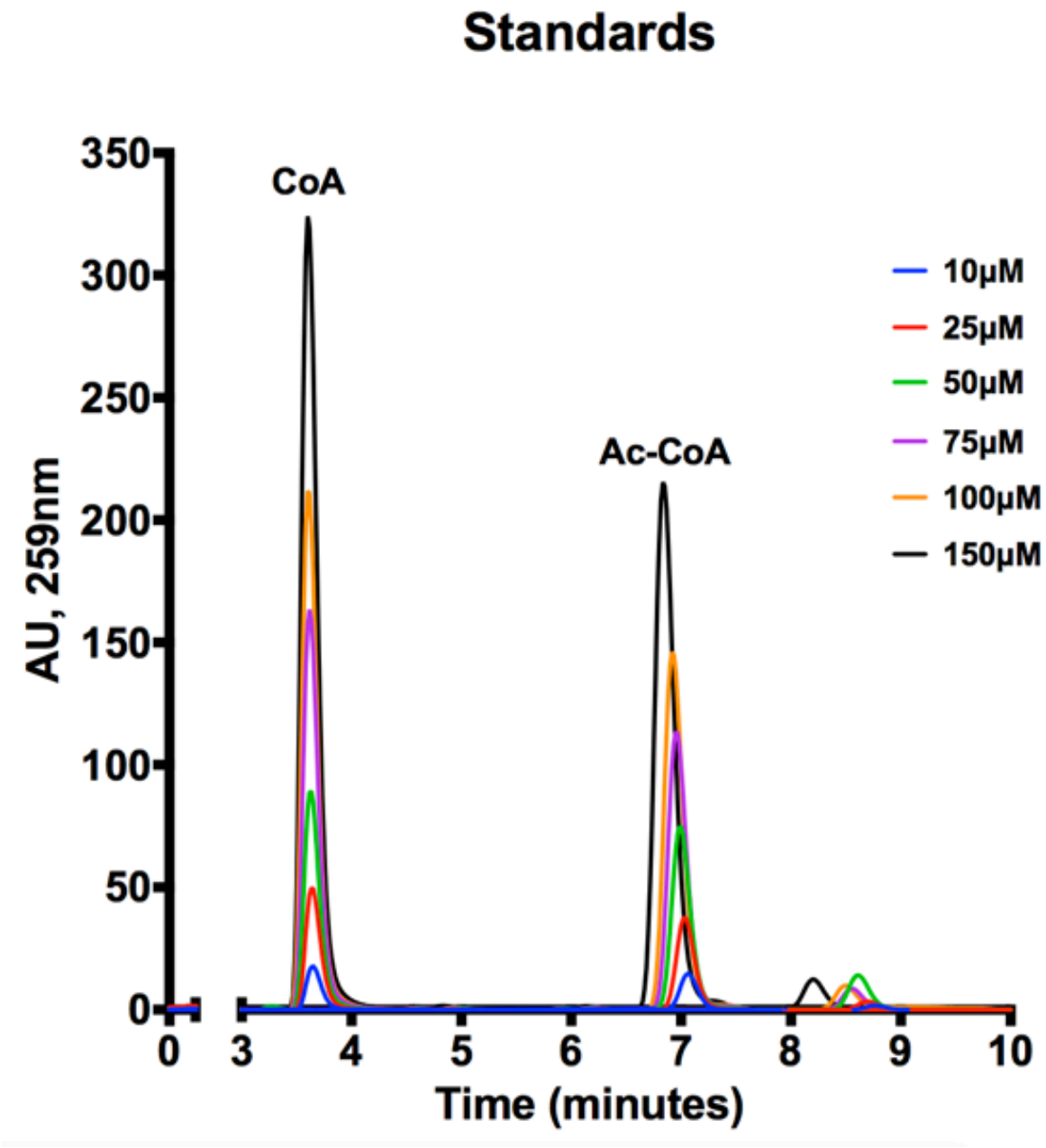


Figure 5. HPLC assay standards
CoA and acetyl-CoA had well-separated and distinct peaks, using HPLC.

2.5.5 Enzymatic activity assay

Spectrophotometric assays were used to monitor the oxidative decarboxylation of L-malate to pyruvate and the reductive carboxylation of pyruvate to L-malate using a CARY 300 Bio, UV-Visible spectrophotometer operated by CARY WinUV kinetics software. All assays recorded the absorbance at 340 nm (NADPH) for a 2 minute period

at 25°C. A molar extinction coefficient for NADPH of 6.22 mM^{-1} cm^{-1} was used for all calculations. Assays were initiated by the addition of enzyme to the otherwise complete reaction mixture.

The standard reaction mixture for malate oxidative decarboxylation was 100 mM Tris pH 7.5, 50 mM KCl, 5 mM $MnCl_2$, 5 mM L-malate, 0.5 mM $NADP^+$. The standard reaction mixture for pyruvate reductive carboxylation was 100 mM Tris pH 7.5, 50 mM KCl, 5 mM $MnCl_2$, 30 mM $NaHCO_3$, 0.1 mM NADPH, 10 mM pyruvate. The apparent Michaelis-Menten parameters were determined for substrate and cofactor by varying the concentration of one substrate while keeping the other components unchanged at saturating concentrations. A maximum concentration of 10 mM L-malate was used in assays, beyond which reduced rates were observed. A final enzyme concentration of 100 nM was used to determine $K_{M(malate)}$, whilst 50 nM was used to determine $K_{M(NADP^+)}$. The curves were fit to the Michaelis-Menten model $v = V_{max}*[malate]/(K_M + [malate])$, where v is the rate. k_{cat} was obtained by adjusting the model to compensate for enzyme concentration (Et), $v = Et*k_{cat}*[malate]/(K_M + [malate])$. All kinetic parameters were fit with non-linear regression using GraphPad Prism 7 software and presented as means of triplicate experiments ± standard error.

A range of compounds known to affect the activity of MEs was tested against Bd1833. The compounds (acetyl-P, pyruvate, succinate, fumarate, oxaloacetate, glucose-6-P, glutamate and glutamine) were tested at 2 mM and (CoA and acetyl-CoA) were tested at 5 μM. The concentration of L-malate used during these tests was kept at one-fifth the $K_{M(malate)}$ (0.5 mM). The results are presented as a percentage of the activity between the activity in the presence of effector and the activity in the absence of effector.

The mode of CoA and acetyl-CoA inhibition was investigated by measuring substrate velocity curves (L-malate and NADP+) at multiple inhibitor concentrations. Curves were fit to the non-competitive inhibition model

$v = (V_{max}/(1+[I]/K_i))*[malate]/(K_M+[malate])$ to obtain the K_i parameter. Data were also plotted as a double reciprocal plot (Lineweaver-Burk) to graphically represent the mode of inhibitor concentration. CoA or acetyl-CoA inhibitor potency was investigated by measuring the activity of Bd1833-apo (ligand free) under standard assay conditions at multiple inhibitor concentration. The data were graphically plotted Rate vs [inhibitor] and fit using the non-linear regression to the [Inhibitor] vs. response -- Variable slope model on GraphPad Prism 7 software. The IC_{50} values were then converted to a K_i using the Cheng-Prusoff equation $K_i = IC_{50}/(1+([S]/K_M))$ based on knowledge of the mode of inhibition (allosteric non-competitive).

2.5.6 MST Cdg Binding Assay

The same procedure as described for Bd2924 MST was used. 20 mM HEPES pH 7.5, 50 mM KCl was used as the MST buffer.

2.5.7 Isothermal Titration Calorimetry

Isothermal Titration Calorimetry (ITC) was used to investigate the binding behavior of CoA and acetyl-CoA ligands with Bd1833. Both CoA and acetyl-CoA are inhibitors of Bd1833 and both co-purify with the protein. To get more reliable data Bd1833-Pta that had co-purifying ligands removed during purification (Apo) was used because it was more stable than full-length Bd1833. Removal of co-purifying ligands before ITC

experiments was required to make data interpretation easier. ITC experiments were ran using a VP-ITC micro-calorimeter with the following protocol for data collection: injection volume = 6 µl, injection length = 12 s, gap between injections = 400 s, total injections = 60, cell temperature = 25°C, reference power = 10 µCal/sec, initial delay = 60 s, stirring speed = 307 rpm. Bd1833-Pta was dialysed into (100 mM Tris, 50 mM KCl) assay buffer. CoA and acetyl-CoA (Sigma) where resuspend in assay buffer to 20 mM stock solutions. The cell was loaded with 100 µM of degassed Bd1833-Pta and the syringe loaded with either 1.5 mM acetyl-CoA or 1 mM CoA. Two controls were performed: buffer injection into cell containing Bd188-Pta and ligand injections into cell containing buffer. The final heat changes were analysed using Origin software and heat changes due to dilution were subtracted from the experimental data. The data were fit to a six-site binding model; Bd1833-Pta forms a tightly associating hexamer complex were each site is identical; however, ligand-binding is known to change the proteins conformation. The ITC data obtained for Bd1833-Pta with CoA and acetyl-CoA is hard to interpret: firstly because residual amounts of co-purifying ligand remain making determining the exact concentration problematic, secondly the hexamer complex of Bd1833-Pta shows cooperative binding.

2.5.8 Crystallisation and cryoprotection

Crystals were obtained for full-length Bd1833, Bd1833-ME, Bd1833-Pta and Bd1833 mutants in a variety of different conditions and forms from commercial crystallization screens as described in Section 2.1.17. The crystallization conditions for Bd1833 crystals and the cryo-protection used are presented in Table 5. Crystals were handled as described in Section 2.1.17. Bd1833 full-length crystals were harvested two days

after the first crystals developed; reduced growth time resulted in better quality diffraction. A soaking experiment was used to get the Bd1833-ME ($NADP^+$) complex structure. Bd1833-ME crystals were transferred into a solution of mother liquor supplemented with (1mM $NADP^+$ and 20% v/v ethylene glycol) and incubated for one hour before being snap frozen in liquid nitrogen.

	Crystallisation Condition
Bd1833 (Full-length)	0.2M sodium citrate tribasic monohydrate, 0.1M BisTris Propane pH 8.5, 20% w/v PEG 3350
Bd1833-ME (native)	1) 0.2M sodium fluoride, 20% w/v PEG 3350 2) 0.1M Bicine pH 9.3, 25% v/v PEG smear medium
Bd1833-ME ($NADP^+$)	0.1M Bicine pH 9.3, 25% v/v PEG smear medium
Bd1833-Pta	0.2M TMAO, 0.1M Tris pH 8.5, 20% w/v PEG2000 MME
Bd1833-Pta (Apo)	0.1M magnesium acetate, 0.1M sodium acetate pH 4.5, 8% w/v PEG 8000
Bd1833-Pta (R535A)	0.2M magnesium chloride hexahydrate, 10% v/v ethylene glycol, 0.1M HEPES pH 7.5, 15% v/v PEG Smear Medium, 5% v/v 2-propanol
Bd1833-Pta (R535E)	0.1M magnesium chloride, 0.1M HEPES pH 7.5, 10% w/v PEG400
Bd1833-Pta (E544R)	0.2M magnesium chloride hexahydrate, 10% v/v ethylene glycol, 0.1M HEPES pH 7.5, 15% v/v PEG Smear Medium, 5% v/v 2-propanol
Bd1833-Pta (Y576S)	0.2m sodium iodide, 0.1M Bis-Tris propane pH 8.5, 20% w/v PEG 3350
Bd1833-SKB (Full)	0.1M Tris pH 8.0, 25% v/v PEG 350 MME
Bd1833-SKB (Pta)	0.005M cobalt(II) chloride hexahydrate, 0.005M cadmium chloride hemi(pentahydrate), 0.005M magnesium chloride hexahydrate, 0.005M nickle(II) chloride hexahydrate, 0.1M HEPES pH 7.5, 12% w/v PEG 3350

Table 5. Bd1833 crystallisation conditions

2.5.9 Data collection and structure determination

Datasets for Bd1833 crystals were routinely collected at Diamond, autoprocessed by ISPyB automated pipelines and solved by MR; see Section 2.1.18. Bd1833-Pta was initially solved by MR, searching for three copies of a 1TD9 dimer (edited through Sculptor, to prune the sidechains [113]) in the ASU, which gave a solution with a TFZ of 17.3 and LLG of 484. Bd1833-ME was also solved by MR: one copy of a 5CEE (AYWB-ME) dimer in the ASU was searched for, which gave a TFZ of 32.4 and LLG of 1044. Bd1833 full-length was solved by MR, searching for one copy of the Bd1833-Pta hexamer and three copies of the Bd1833-ME dimer, resulting in a solution with a TFZ of 17.9 and LLG of 16256. Finally, *Bdellovibrio*-SKB strain homologues of Bd1833 were solved using HD100 Bd1833 and Bd1833-Pta structures. Coot was used to manually place ligands, such as acetyl-CoA and $NADP^+$ into the electron density. Initial MR models were improved as described in Section 2.1.18.

Data Collection	Bd1833-ME (Bcs-B6)	Bd1833-ME (Pact-E1)	Bd1833-ME ($NADP^+$)
Spacegroup	P 2_1	P1	P 2
Cell Dimensions a, b, c (Å)	48.1, 92.7, 97.1	56.2, 100.2, 130.9	47.8, 92.2, 95.8
α, β, γ (°)	90, 91.1, 90	88.1, 83.7, 81.9	90, 91.3, 90
Resolution (Å)	43.41 - 1.67	37.07 - 2.2	95.78 - 1.67
Total observations	656294 (32403)	475063 (21446)	626225 (29211)
Unique reflections	96511 (4697)	138460 (6888)	96099 (4736)
Rmerge(I)	0.0851 (1.6100)	0.1055 (0.9044)	0.1191 (0.7974)
Rpim	0.036 (0.634)	0.068 (0.621)	0.050 (0.350)
I/sigI	10.81 (1.22)	6.58 (1.02)	6.87 (1.32)
CC(1/2)	0.9977 (0.4634)	0.9939 (0.5009)	0.9962 (0.4718)
Completeness (%)	97.80 (96.54)	97.23 (96.17)	99.90 (98.14)
Multiplicity	6.81 (6.98)	3.43 (3.11)	6.52 (6.17)
Refinement			
R-work	0.1702	0.2109	0.1799
R-free	0.2039	0.2264	0.2159
RMS (bonds) (Å)	0.0077	0.0106	0.0108
RMS (angles) (°)	1.113	1.429	1.257
Average B-Factor ($Å^2$)	37	53.3	34.5
# of non-hydrogen	6808	18553	6913
Molecules per ASU	2	6	2
Ramachandran Plot			
Preferred region	96.4	95.5	96.1
Allowed region	3.3	4.1	3.7

Table 6. Bd1833-ME data collection and refinement

Data Collection	Bd1833-Pta	Apo	R535A	R535E	E544R	Y576S
Spacegroup	$P2_1$	$P\ 2_1\ 2_1\ 2_1$	C 2	C 2	C 2	C 2
Cell Dimensions a, b, c (Å)	111.2, 60.6, 167.8	139.2, 151.3, 285.3	129.6, 183.7, 119.9	129.4, 183.4, 120.2	129.1, 182.9, 119.7	128.8, 182.4, 121.4
α, β, γ (°)	90, 91.1, 90	90, 90, 90	90, 117, 90	90, 117, 90	90, 118, 90	90
Resolution (Å)	91.86 - 1.96	142.66 - 2.72	64.50 - 2.35	107.10 - 2.39	69.22 - 2.33	95.97 - 2.38
Total observations	1083798 (55212)	2082343 (102453)	705212 (37173)	424649 (17692)	698066 (266312)	663750 (28462)
Unique reflections	160802 (7925)	162146 (7836)	103459 (5235)	96539 (4863)	104231 (5084)	98271 (4884)
Rmerge(I)	0.0589 (1.4047)	0.1294 (0.8860)	0.0867 (1.8559)	0.103 (1.975)	0.0872 (1.2654)	0.119 (1.079)
Rpim	0.025 (0.571)	0.038 (0.254)	0.036 (0.744)	0.055 (1.201)	0.036 (0.600)	0.042 (0.47)
I/sigI	15.70 (1.14)	11.40 (1.47)	10.69 (1.16)	8.5 (0.5)	12.03 (1.14)	10.8 (0.7)
CC(1/2)	0.999 (0.637)	0.998 (0.600)	0.999 (0.573)	0.998 (0.697)	0.999 (0.590)	0.999 (0.569)
Completeness (%)	99.92 (99.92)	99.95 (97.82)	99.90 (100.00)	98.4 (99.7)	99.72 (97.98)	100 (100)
Multiplicity	6.74 (6.97)	12.84 (13.07)	6.81 (7.10)	4.4 (3.6)	6.70 (5.24)	6.8 (5.8)

Refinement						
R-work	0.1829	0.2520	0.2158	0.2153	0.2069	0.2222
R-free	0.2052	0.2995	0.2368	0.2602	0.2265	0.2725
RMS (bonds) (Å)	0.009	0.012	0.0103	0.01	0.0097	0.01
RMS (angles) (°)	1.178	1.590	1.498	1.415	1.448	1.321
Average B-Factor ($Å^2$)	29.5	86.2	45.7	75.5	74.9	74.1
# of non-hydrogen	16475	31268	15572	15628	15849	15720
Molecules per ASU	6	12	6	6	6	6
Ramachandran Plot						
Preferred region	95.8	96.2	98.7	96.6	97.0	96.3
Allowed region	4.0	3.6	1.3	3.1	2.7	3.6

Table 7. Bd1833-Pta data collection and refinement

Data Collection	Bd1833 (Full-length)
Spacegroup	C 2
Cell Dimensions *a, b, c (Å)*	250.422 144.792 171.138
α, β, γ (°)	90.000 100.605 90.000
Resolution (Å)	124.80 - 2.72
Total observations	1086009 (56926)
Unique reflections	160818 (8003)
Rmerge(I)	0.1 (2.788)
Rpim	0.042 (1.125)
I/sigI	11.0 (0.6)
CC(1/2)	0.999 (0.317)
Completeness (%)	99.9 (100)
Multiplicity	6.8 (7.1)

Refinement	
R-work	0.1922
R-free	0.2277
RMS (bonds) (Å)	0.0108
RMS (angles) (°)	1.437
Average B-Factor (Å^2)	108
# of non-hydrogen atoms	34603
Molecules per ASU	6
Ramachandran Plot (%)	
Preferred region	95.5
Allowed region	4.1

Table 8. Bd1833 full-length data collection and refinement

Data Collection	Bd1833-SKB (Full-length)	Bd1833-SKB (Pta)
Spacegroup	P 4_1 2_1 2	F 2 3
Cell Dimensions a, b, c (Å)	174.62, 174,62, 203.29	307.09, 307.09, 307.09
α, β, γ (°)	90, 90, 90	90, 90, 90
Resolution (Å)	132.46 - 3.71	54.29 - 2.71
Total observations	885942 (45499)	663665 (33984)
Unique reflections	34022 (1653)	64779 (3209)
Rmerge(I)	0.195 (6.015)	0.1309 (1.8829)
Rpim	0.034 (0.620)	0.043 (0.606)
I/sigI	13.6 (0.7)	12.06 (1.25)
CC(1/2)	1.000 (0.393)	0.9981 (0.4515)
Completeness (%)	100 (100)	100.00 (99.97)
Multiplicity	26.0 (27.5)	10.25 (10.59)

Refinement		
R-work	0.2503	0.2040
R-free	0.2996	0.2392
RMS (bonds) (Å)	0.004	0.013
RMS (angles) (°)	0.783	1.43
Average B-Factor ($Å^2$)	199.1	64.5
# of non-hydrogen atoms	17163	10592
Molecules per ASU	3	3
Ramachandran Plot (%)		
Preferred region	95.1	97.0
Allowed region	4.3	3.0

Table 9. Bd1833 SKB data collection and refinement

2.6 Bd0967

2.6.1 Cloning, Expression and Purification

A DNA construct of Bd0967, encoding the region from M1 and ending Q666, was designed to incorporate the full-length protein truncated at the C-terminus to remove the predicted region of disorder. Bd0967 has a potential signal peptide (residues 1-25) that ends with the sequence LASA; the signal peptide region was also cloned into the Bd0967 construct. Bd0967 was inserted into the pET28a expression plasmid with an N-terminal His_6Tag and a C-terminal His_6Tag. The plasmid was transformed into a BL21* (DE3) *E. coli* expression strain. Expression and purification of Bd0967 was performed using the same procedures as described in Section 2.1. Bd0967 was dialysed into (20 mM HEPES pH 7.5, 300 mM NaCl, 5 mM $MgCl_2$); $MgCl_2$ was supplemented to stabilize the metal-binding site present in the protease domain.

2.6.2 SEC

SEC was performed for Bd0967 and mutants to polish the protein for crystallographic studies and to investigate the oligomeric composition in solution. SEC analysis was performed as described in Section 2.1.14, using Bd0967 dialysis buffer (20 mM HEPES pH 7.5, 300 mM NaCl, 5 mM $MgCl_2$).

2.6.3 Pulldown Assay

A pulldown assay was performed to identify proteins and, or substrates that interact with Bd0967. The assay was performed at the University of Nottingham with assistance of the Sockett lab.

Bdellovibrio bacteriovorus HD100 cells, grown to a cell density of $2.5x10^8$ cells/ml, were harvested by centrifugation 6000 RPM for 20 minutes at 4°C in a JLA-10.5 rotor. Supernatant was discarded and the cell pellet was resuspended in ice-cold lysis buffer (20 mM Na phosphate pH 7.4, 40 mM imidazole pH 8.0, 300 mM NaCl, a complete protease inhibitor tablet (Roche) and lysozyme). Resuspended cells were aliquoted into fractions of 5 ml (maximum) and lysed by sonication for 5 minutes in cycles of 20s on – 20s off. Lysed cells were centrifuged at 38000 *g* for 45 minutes at 4°C to remove insoluble material. The soluble fraction was removed and stored for pulldown assays with soluble fraction. The insoluble pellet was resuspended in ice-cold lysis buffer supplemented with 2% v/v Triton X-100 and then centrifuged at 38000 *g* for 10 minutes to remove cell debris. The supernatant was extracted and stored for pulldown assays with insoluble fraction.

The pulldown assay was prepared by gently pelleting a 200 µl slurry of Qiagen Ni-NTA beads 2500 *g* for 2 minutes at 4°C. The pellet was washed three times with wash buffer to equilibrate the beads. Purified Bd0967 was diluted using wash buffer (20 mM sodium phosphate pH 7.4, 300 mM NaCl, 40 mM imidazole) and 1 ml of 150 µg of Bd0967 was applied to the pelleted beads. The beads were incubated with Bd0967 for 30 minutes at 4°C by end-over-end rotation. A control experiment with lysis buffer only was also prepared. The incubated beads were then pelleted and unbound protein solution removed and the pellets washed three times with wash buffer.

The Ni-NTA beads bound with Bd0967 and the lysis buffer-only control were then incubated for 2 hours with 5 ml of soluble or insoluble *Bdellovibrio bacteriovorus*

HD100 cell lysates prepared earlier by end-over-end rotation at 4°C. Another control (Bd0967-bound beads incubated with lysis buffer for 2 hours) was used at this step. After incubation the beads were pelleted and washed three times with wash buffer. The pelleted beads were finally resuspended in 100 µl 1x Laemeli buffer and boiled at 100°C for 15 minutes. The boiled samples were analysed by SDS-PAGE. Bd0967 interacting proteins were detected in samples where the insoluble cell lysate fraction had been used. Gel samples were cut into lanes and frozen at -20°C.

2.6.4 Mass-Spectroscopy

Mass-Spectroscopy (MS) was used to identify interacting proteins uncovered by SDS-PAGE analysis of the pulldown assay. MS analysis was conducted at the University of Warwick Proteomics facility; samples were prepared using the standard proteomic facility protocol for in-gel protein digestion.

Three gel lanes were analysed by MS: (1) beads + Bd0967 + lysate, (2) beads + Bd0967 only control and (3) beads + lysate only control. The gels were defrosted, cut into small pieces and resuspended in destain solution (50 mM ammonium bicarbonate (ABC) and 50% ethanol) then shaken at 650 rpm for 20 minutes. The destain solution was removed and fresh distain reapplied and the process repeated until all stain was removed. Gel segments were then dehydrated with 100% ethanol until the segments turned white. The ethanol was then removed and incubated with 10 mM DTT and 50 mM ABC for 30 minutes at 56°C and shaken at 650 rpm. The samples were cooled to room temperature and the liquid removed. Cysteine residues were alkylated by incubation with 55 mM Indole-3-acetate (IAA) and 50 mM ABC for 20 minutes in the

dark at room temperature. The liquid was then removed and washed three times with 50% methanol and 50 mM ABC, shaken at 650 rpm for 20 minutes. The samples were then dehydrated with 100% ethanol, shaken at 650 rpm for 5 minutes. The liquid was then removed and the samples re-hydrated in trypsin solution containing 2.5 ng/µl for 10 minutes. Enough 50 mM ABC solution was then applied to cover the gel segments, which were incubated at 37°C O/N. The peptides were extracted from the gel segments by sonication in a sonication bath with 25% v/v acetonitrile and 5% v/v formic acid for 10 minutes and then removed into a new Eppendorf tube. The extraction procedure was repeated three times and the solutions pooled, which were then concentrated in a speed vacuum centrifuge (set at 40°C) until 20 µl of solution remained. The peptides were then resuspended in 2.5% v/v acetonitrile and 0.05% v/v trifluoroacetic acid to bring the final volume to 50 µl. A 20 µl sample of each sample (1,2,3) was run on an Ultimate 3000/Orbitrap Fusion instrumentation (Thermo Scientific) using nanoLC-ESI-MS/MS with a 30 minute LC separation on a 25 cm column. Initial analysis was also provided by Warwick Proteomic facility using the Uniprot database for *E. coli* and *B. bacteriovorus* HD100 for peptide searches.

2.6.5 Co-purified peptide analysis (Collaboration)

Unsuccessful attempts were made to identify the co-purifying and co-crystallising peptide. Kvido Stříšovský's lab at the Czech Academy of sciences collaborated to identify the bound-peptides. The Stříšovský lab analysed Bd0967-Wt, Bd0967-S457A and Bd0967-R299M by native ESI-MS and acidic denatured ESI-MS. The N-terminus of Bd0967 was unambiguously determined as QLKEGLE… demonstrating Bd0967 does have a signal peptide that is removed during expression/purification. Both native and

denaturing ESI-MS did not detect any significant protein candidates and the majority of those detected were common Ni-NTA contaminating proteins.

2.6.6 Bd0967 pulldown with Bd0606

A pulldown assay was performed with purified Bd0606 instead of *Bdellovibrio* cell lysates using the same procedure with two modifications. Firstly Bd0967 buffer (20 mM HEPES pH 7.5, 300 mM NaCl, 5 mM $MgCl_2$) was used as the binding and wash buffer. Secondly Ni-NTA beads bound with Bd0967 was incubated for two hours with 500 µl of 1 mg/ml purified Bd0606 (N-terminal His_6Tag removed) at 4°C by end or end rotation. The pulldown with Bd0606 was performed with Bd0967-Wt, Bd0967-S457A and Bd0967-R299M as bait protein and Bd0967 as prey. Binding was analysed by SDS-PAGE.

2.6.7 Protease assay

Bd0606 appeared to be a weak binder to Bd0967 so an end-point protease assay was developed to determine if Bd0606 was a potential substrate. The standard protease assay monitored the degradation of 4 mg Bd0606 in the presence of 100 µg Bd0967-Wt at 37°C (while being agitated). 20 mM HEPES pH 7.5, 300 mM NaCl, 5 mM $MgCl_2$ was used as the assay buffer throughout. The assay was initiated by the addition of Bd0967 and a sample taken immediately and then at varying time intervals up until 60 minutes. The reaction was quenched by the addition of 2xSDS loading dye and immediately boiled at 100°C for 10 minutes. Two controls were used for every protease assay: 1) Bd0967 incubated alone for 60 minute (no substrate added) and 2) Bd0606 incubated alone for 60 minutes (no protease added). The samples were analysed by SDS-PAGE.

The SDS-PAGE gel was imaged and analysed using imageJ to illustrate Bd0606 degradation graphically. This assay was also used to investigate S457A and R299M mutations and Bd0606 LIG and RLIG truncation mutants.

The assay was adapted to investigate Bd0967 ability to degrade other proteins. A selection of proteins of varying sizes, folds and C-termini sequences were chosen. The concentration of proteins (substrates) was maintained at 4mg for each reaction. Three samples were taken for protein tested: 0 min +Bd0967, 60 min +Bd0967 and 60 min -Bd0967.

2.6.8 Protease peptide inhibition assay

To investigate the role of the PDZ domain in substrate recognition, a 12aa peptide matching the C-terminus of Bd0606 (NAMPNSALRLIG), was kindly prepared by Oliver Daubney. The peptide was N-terminally synthesized and purified. The product was confirmed by MS to be >96% pure.

The peptide was resuspended in Bd0967 assay buffer (20 mM HEPES pH 7.5, 300 mM NaCl, 5 mM $MgCl_2$) to a final concentration of 10 mg/ml, which was aliquoted and frozen at -20°C. The peptide inhibition assay monitored substrate degradation in the presence of peptide inhibitor ranging from 0 mM to 8 mM. Bd0967 (100 µg) was incubated with the peptide (concentration dependent) for 30 minutes to ensure the peptide had a sufficient time to interact with Bd0967 prior to initiating the reaction. The assay was performed in Bd0967 assay buffer (20 mM HEPES pH 7.5, 300 mM NaCl, 5 mM MgCl2) and initiated by the addition of 4 mg Bd0606. Samples were incubated at

37°C, for 30 minutes, while being agitated. The reaction was quenched by the addition of 2x SDS loading dye and boiled at 100°C for 10 minutes.

2.6.9 Crystallisation and cryoprotection

Bd0967 crystals were optimized by successive rounds of random matrix microseeding [114]. Initial poor quality crystals were grown in Morpheus G4 ((0.1M Carboxylic acid mix (0.2M sodium formate, 0.2M ammonium acetate, 0.2M sodium citrate tribasic dihydrate, 0.2M sodium potassium tartrate tetrahydrate, 0.2M sodium oxamate), 0.1M Buffer system 1 (PEG , MES monohydrate) pH 6.5, Precipitant mix 4 (25% v/v MPD, v/v 25% PEG 1000, 25% w/v PEG 3350)). The crystals were crushed and extracted by adding 10 µl of mother liquor on top, mixing and removing into a cooled Eppendorf tube. This process was repeated 10 times to remove as many seeds as possible. The crystals were broken into small seeds by cycling vortexing for 30s and resting on ice for 2 minutes with a small plastic bead in the tube to aid crushing. The crushed seeds were used immediately in microseeding experiments, set up using a TTPlabtech Mosquito HTS nanolitre liquid handler. Morpheus B5 (0.09M Halogens (0.3M sodium fluoride, 03M sodium bromide, 0.3M sodium iodide), 0.1M Buffer system 2 (HEPES, MOPS) pH 7.5, 50% precipitant mix 1 (40% v/v PEG 500 MME, 20% w/v PEG 2000)), that had been seeded, gave the best quality diffracting crystals. Bd0967-S457A crystals were grown in Morpheus G12 (0.1M Carboxylic acid mix, 0.1M buffer system 3 (Tris, BICINE) pH 8.5, 50% v/v precipitant mix 4) that had been seeded. No additional cryo-protection was required and the crystals were looped and flash frozen directly from the well.

2.6.10 Data collection and structure determination

Bd0967 datasets were routinely collected and solved; see Section 2.1.18. The program, chainsaw, was used to edit the search model, chain A of PDB:4QL6 (CT441). The sidechains of residues were pruned to match those of Bd0967 in a sequence alignment [115]. An incomplete initial model for Bd0967 was found using MR (TFZ = 31, LLG = 1138.8), which was improved by Phenix.autobuild, which built 815 out of residues and 27 fragments (R-work = 0.3121, R-free = 0.3718). The model was finalized as described in Section 2.1.18.

Data Collection	Bd0967 Wt	Bd0967 S457A
Spacegroup	P1	P1
Cell Dimensions a, b, c (Å)	73.4, 95.8, 114.1	72.9, 95.0, 113.3
α, β, γ (°)	77.5, 71.3, 81.9	77.7, 71.8, 81.9
Resolution (Å)	106.26 - 2.04	38.42 - 2.58
Total observations	848316 (18620)	304174 (15674)
Unique reflections	179707 (8159)	87121 (4331)
Rmerge(I)	0.1656 (0.3533)	0.1133 (1.0666)
Rpim	0.079 (0.280)	0.071 (0.654)
I/sigI	4.78 (1.23)	7.52 (1.13)
CC(1/2)	0.9517 (0.7332)	0.9946 (0.4479)
Completeness (%)	97.84 (89.57)	98.52 (98.12)
Multiplicity	4.72 (2.28)	3.49 (3.62)
Refinement		
R-work	0.1990	0.2118
R-free	0.2399	0.2607
RMS (bonds) (Å)	0.0104	0.010
RMS (angles) (°)	1.431	1.328
Average B-Factor (Å^2)	43.1	59.9
# of non-hydrogen atoms	21634	19935
Molecules per ASU	4	4
Ramachandran Plot (%)		
Preferred region	97.6	97.0
Allowed region	1.4	2.2

Table 10. Bd0967 data collection and refinement

2.6.11 MST Cdg Binding Assay

The same procedure as described for Bd2924 MST was used. 20 mM HEPES pH 7.5, 50 mM KCl, 5 mM $MgCl_2$ was used as the MST buffer.

2.7 Bd0606

2.7.1 Cloning, Expression and Purification

Bd0606 (M1 - G277) was cloned into pET28a expression plasmid and transformed into BL21* (DE3) *E. coli* expression strain. Two C-terminal truncation mutants of Bd0606 were created: LIG – lacking three terminal amino acids and RLIG lacking four terminal amino acids. Bd0606 and truncation mutants were expressed and purified as described in Section 2.1. Bd0606 was dialysed into Bd0967 buffer (20 mM HEPES pH 7.5, 300 mM NaCl, 5 mM $MgCl_2$) and concentrated to 30 mg/ml.

The N-terminal His_6Tag was removed for a fraction of Bd0606 (use in pull-down experiments). Bd0606 was incubated with 1:100 (thrombin:Bd0606) thrombin (Novagen) for 2-3 hours at room temperature. After incubation the sample was passed over a HisTrap column in tandem with a benzamidine column (GE Healthcare) to remove un-cleaved Bd0606, His_6-tag and thrombin. The flow through contained tag-less Bd0606 and the cleavage efficiency was monitored by SDS-PAGE.

2.7.2 Circular Dichromism (CD)

CD was carried out to determine whether Bd0606 was correctly folded and not disordered. Bd0606 was prepared for CD by dilution into 10 mM sodium phosphate buffer pH 7.4 to a final concentration of 0.2 mg/ml. CD experiments were generously performed and analysed by Charles Moore-Kelly on a JAS.CO – J-1500 cd spectrometer.

Chapter 3

Bd0314, an Adapted Lysozyme Required for Bdelloplast Escape

3.1 Introduction

Bdellovibrio converts its prey into a secure niche, in part, by modifying the prey cell wall. The prey peptidoglycan layer is solubilized locally to create a small pore that *Bdellovibrio* invades through to reside between the inner membrane and peptidoglycan layer [45]. The prey cell wall is ultimately converted into a stable, rounded bdelloplast structure. The bdelloplast is osmotically stable and limits accessibility to further predatory invasion [47]. *N*-deacetylation of the prey cell wall is performed during invasion and contributes to stability and renders the bdelloplast structure insensitive to lysozyme activity [50]. Moreover, *N*-deacetylation distinguishes the PG material of "self" and "prey", which marks the prey cell wall for destruction at a later stage [49]. In order to complete the predatory lifecycle, *Bdellovibrio* must exit the fortified bdelloplast structure. Consequently this final step requires novel enzymology, targeted towards the modified prey cell peptidoglycan, deployed at the correct time. Here, we investigate the possibility that Bd0314 could perform the role of "Exit enzyme" and thus facilitate prey cell lysis at the end of the predatory lifecycle.

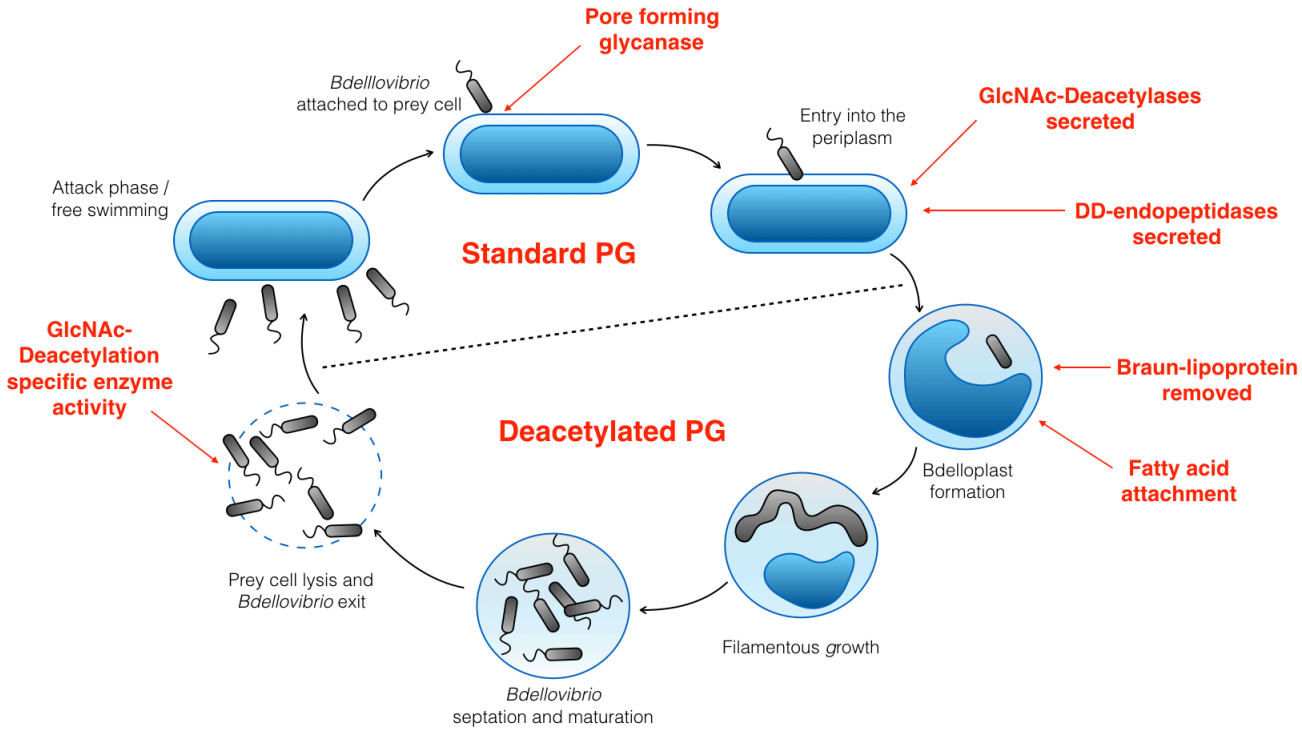


Figure 6. *Bdellovibrio* lifecycle highlighting prey cell wall modifications.
Bdellovibrio actively seeks out prey cells and attach to the outer membrane. They then secrete a localised pore forming glycanase that creates a small entry pore in the prey cell wall that Bdellovibrio squeezes through and then reseals. During invasion *Bdellovibrio* secretes deacetylases and DD-endopeptidases. GlcNAc-deacetylases (Bd0468 and Bd3279) deacetylate the prey cell wall marking 'prey' from 'self'. *N*-deacetylation of the prey cell wall makes the bdelloplast structure resistant to lysozyme activity. DD-endopeptidases (0816 and 3459) break cross-links in the prey cell wall and rounds the prey cell into the bdelloplast structure. Rounding of the prey cell acts as an 'occupied' signal preventing further invasion events. A significant proportion of the prey cell Braun-lipoprotein is then removed and fatty acid chains are attached to the prey cell wall, which stabilises the bdelloplast structure. At the end of the predatory cycle a final lytic enzyme activity is required to completely solubilise the prey cell wall and lyse the bdelloplast structure. The final lytic enzyme activity must be highly regulated and active towards the deacetylated prey cell wall.

3.1.1 Peptidoglycan Overview

3.1.1.1 Overview

Peptidoglycan is an essential feature of bacterial cell walls, located on the outside of the cytoplasmic membrane. The main function of peptidoglycan is to withstand turgor pressure and maintain the cellular integrity. If the peptidoglycan layer is compromised, by inhibiting its biosynthesis (e.g. antibiotics) or specific degradation (e.g. lysozymes), cells lyse. Peptidoglycan is directly involved in cell growth and division, while also

being required to maintain cell shape and serve as a molecular scaffold for anchoring other cell envelope components such as teichoic acids [116].

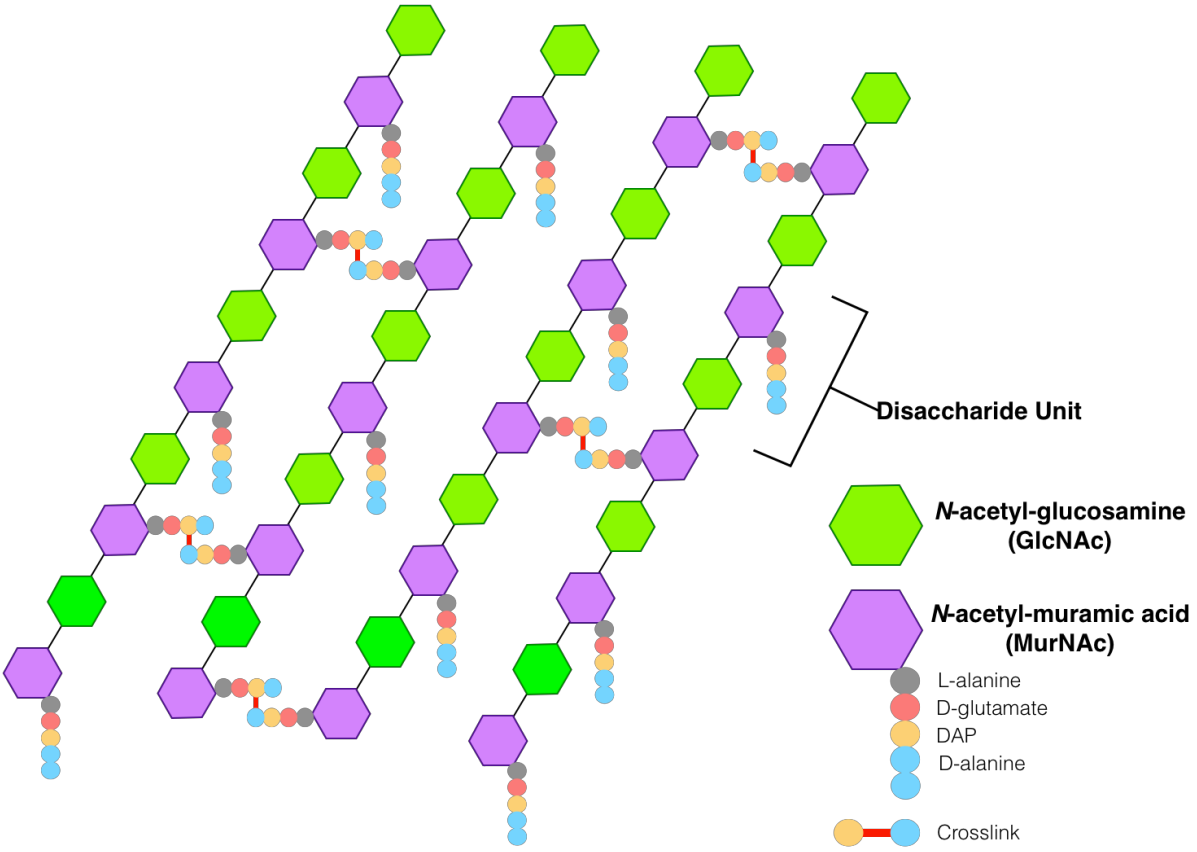


Figure 7. Peptidoglycan structure
This Figure represents the peptidoglycan structure of *E. coli*. The glycan strand is composed of alternating β-1-4 linked GlcNAc (green) and MurNAc (purple) residues. The basic disaccharide subunit (monomer) is highlighted. Components making up the *E. coli* peptide stem are highlighted. Cross-links between the tetrapeptide stems are represented by a red bond.

3.1.1.2 Structure and Composition

Peptidoglycan is comprised of a linear chain of glycan stands that are cross-linked together by short peptide strands (Figure 7). The glycan strands or made up of alternating *N*-acetylglucosamine (GlcNAc) and *N*-acetylmuramic acid (MurNAc) residues, linked via β-1-4 bonds. MurNAc contains a D-lactoyl group that is substituted

by a pentapeptide stem. The peptide stem is commonly made of L-Ala-γ-D-Glu-meso-A_2pm (or L-Lys)-D-Ala-D-Ala, where the final D-Ala residue is removed during peptide cross-linking. Cross-linking of the peptide stem occurs between the carboxyl group of D-Ala at position 4 and the amino group of the residue at position 3. Peptidoglycan can either have directly cross-linked peptide stems or peptide stem cross-linked by a short peptide bridge [116].

3.1.1.3 Peptidoglycan Variation

Peptidoglycan is structurally conserved in bacterial specie. However, a degree of variation is found in the peptide stems, glycan strands and peptide cross-links. Glycan strands, polymerised disaccharide peptide units (Lipid II), are produced by glycosyltransferases. Modification to the glycan strand such as *N*-deacetylation and O-acetylation are abundant in pathogenic bacteria and contribute to virulence. Glycan strands can either terminate with a reducing MurNAc or GlcNAc residue [116] or with a 1,6-anhydroMurNAc residue [117]. The length of the glycan strands varies between species and the average length ranges between 20 to 40 disaccharide units [116].

Variation in the peptide stem is common between bacterial species and can arise at numerous positions. The greatest variation occurs at position three where a di-amino acid such as meso-A_2pm (Diaminopimelic acid / DAP) or L-Lys is typically found. In certain other species position three can be substituted for other di-amino acids such as L-Orn or even mono-amino acids such as L-Glu. The amino acids at positions four and five are added as a dipeptide; generally D-Ala-D-Ala. However, position five can be substituted for D-Lac or D-Ser, which is linked to vancomycin resistance [116].

Furthermore, the peptide stem is the point of covalent attachment for cell envelope proteins to peptidoglycan. In Gram-negative bacteria such as *E. coli,* Braun's lipoprotein is currently the only known protein to be covalently attached to the peptidoglycan [118], which acts to stabilise the cell envelope by linking the outer-membrane to the peptidoglycan layer. Braun's lipoprotein is a 58 amino acid protein that has three fatty acid chains attached to the N-terminal region, which are buried into outer-membrane, whilst the C-terminal lysine of Braun's lipoprotein is covalently attached to meso-A_2pm [119]. Gram-positive bacteria, on the other hand, are decorated with numerous surface proteins, which are covalently attached to the peptide stems and play roles in pathogenesis [120].

Cross-linking between peptide stems is another aspect of variation between glycan strands. The peptides can either be directly cross-linked or cross-linked through an inter-peptide bridge [116]. The transpeptidase domain of penicillin binding proteins (PBPs) carries out the cross-linking reactions. PBPs are the targets of the cell wall inhibiting β-lactam antibiotics such as penicillin [121]. Moreover the degree of cross-linking between glycan strands varies between species. For instance in *E. coli* ~20% of the MurNAc residues are cross-linked, whereas in *S. aureus* >90% of the MurNAc residues are cross-linked [116].

3.1.1.4 Biophysical properties

Peptidoglycan sacculi completely surround the cytoplasmic membrane, providing structural strength and shape to the bacteria. Peptidoglycan sacculi have unique

biophysical properties; they are able to withstand large stresses, while also containing relatively large pores [116].

The thickness of peptidoglycan varies massively between Gram-positive and Gram-negative species. Gram-positive species have a peptidoglycan layer that is several layers thick that is extensively coated with cell wall anchored proteins. In comparison, Gram-negative bacteria have a relatively thin peptidoglycan layer that is shielded from the external environment by an outer membrane. Cryo-TEM has been was used to experimentally determine the thickness of Gram-negative and Gram-positive bacteria. In *E. coli* the peptidoglycan layer has a average thickness of around 2.41 nm [122], whereas in Gram-positive bacteria the peptidoglycan layer is not uniform and is organised into two different layers. In *B. subtilis* the inner wall zone is less dense (~22 nm thick), whereas the outer wall zone is denser (15-30 nm thick) and contains covalently attached teichoic acids [123].

Peptidoglycan sacculi have been shown to have elastic properties and can expand and shrink 3-fold without rupturing [124]. Besides elasticity, sacculi are also very porous structures. Sacculi pores range in sizes, but on average have diameters with a mean radius of 2.06 nm for *E. coli* and 2.12 nm for *B. subtilis* [125]. The porous nature of the sacculi is sometimes referred to as a molecular sieve, allowing the diffusion of smaller molecules but preventing larger macromolecules from penetrating it.

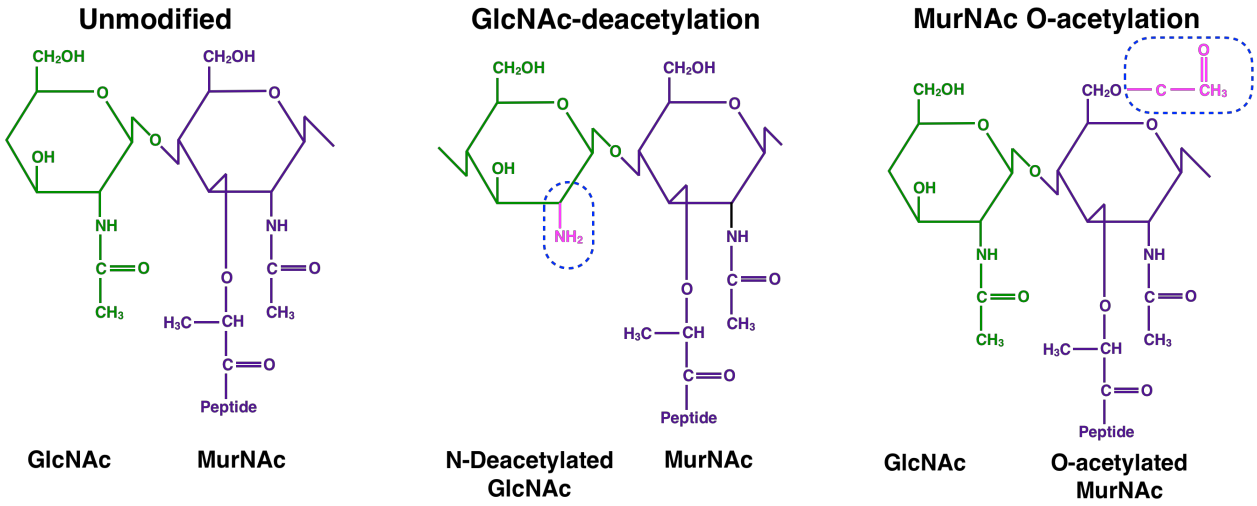


Figure 8. Modifications of the peptidoglycan strand
The unmodified PG disaccharide unit of β-1-4 linked GlcNAc (green) and MurNAc (Purple) residues is illustrated on the left. N-deacetylation of GlcNAc residues is shown in the central panel. *O*-acetylation of MurNAc residues is shown to the right. Chemical alterations are coloured in pink and circled in blue.

3.1.2 Peptidoglycan modifications and pathogenesis

As previously mentioned the fine structure of peptidoglycan varies between bacterial species. Moreover, certain modifications are common in pathogenic bacteria and are used as virulence factors to invade host defences. These modifications are typically generated in the glycan strand by post-synthetic reactions, including acetylation of the C6 hydroxyl group or deacetylation of the C2 amine group of GlcNAc and MurNAc residues. There are three main types of glycan strand modifications: *N*-deacetylation, *O*-acetylation and attachment of surface biopolymers, which will be discussed further.

3.1.2.1 *N*-deacetylation

N-deacetylation of peptidoglycan sugars has been reported in numerous bacterial species and linked to pathogenesis and resistance to lysozyme [126, 127]. Both MurNAc and GlcNAc residues can be *N*-deacetylated to produce MurN and GlcN. In certain species, mainly pathogenic Gram-positives, the majority of their PG (>80%) is deacetylated [128]. Whilst GlcNAc deacetylation is common among pathogenic bacteria, MurNAc deacetylation is less frequently reported [129].

N-deacetylation of peptidoglycan residues is performed by a specialised group of enzymes that deacetylate polysaccharides. Peptidoglycan deacetylases belong to a larger group of carbohydrate esterase family 4 (CE4) of CAZY classification [130] that posses the ability to deacetylate a wide variety of polysaccharide substrates from peptidoglycan to chitin (a polymer of GlcNAc) [130]. Typically deacetylases require a catalytic metal ion (Co^{2+} or Zn^{2+}) to cleave the acetyl-groups. The first deacetylase gene discovered was *pgdA*, a GlcNAc deacetylase from *S. pneumoniae* [129].

The biological role of PG *N*-deacetylation is mainly linked to virulence, in particular the acquired resistance to lysozyme. Several studies demonstrate that deacetylated PG is a poor substrate for lysozymes. As a result PG deacetylation is common in pathogenic bacteria such as *S. pneumoniae* [131], *L. monoctogenes* [126] and *B. anthracis* [128]. PG deacetylation is believed to have two effects, which contribute to lysozyme resistance. Firstly the removal of the acetyl group is proposed to reduce binding affinity, as lysozymes typically bind PG through interactions with the acetyl-group [132]. Secondly the free amine group is likely to be protonated (pK_a ~6.5) [133] hence there will be a net increase in positive charge on PG, which is suggested to repel the positively charged

surface of the lysozyme active site [134]. Furthermore the increased positive charge may also lead to resistance towards cationic antimicrobial peptides (important elements of the innate immune response).

N-deacetylation of MurNAc residues is less common and its biological significance is less well understood. It is less likely to perturb lysozyme binding as the *N*-acetyl group faces outwards, therefore not directly involved in conventional lysozyme substrate interactions [132]. It may, however, have a similar role to GlcNAc deacetylation by incorporating additional positive charges into the cell wall.

3.1.2.2 *O*-acetylation

O-acetylation was first detected in *M. luteus* and *S. faecalis* [135, 136]. The modification involves the addition of an acetyl-group linked to C6-OH of MurNAc residues only; *O*-acetylation of GlcNAc residues has not been observed. *O*-acetylation is a more prevalent glycan strand modification than *N*-deacetylation and is found in both Gram-positive and Gram-negative species. Certain species can have between 20-70% *O*-acetylated MurNAc residues [134]. *O*-acetylation is also a contributor to bacterial pathogenesis and a determining factor of lysozyme resistance [137]. Furthermore, the modification is implicated in autolysin control, in cells at the stationary phase of growth [138]. *O*-acetylation has also been shown to partly inhibit predatory invasion by the obligate predator *Bdellovibrio bacteriovorus [139].*

The genes required for *O*-acetylation were first discovered in 2005 and annotated as *oatA* in *S. aureus [140]*. The *oatA* gene encodes an integral transmembrane protein with

an acyltransferase domain. Gram-negative species do not encode OatA homologues but instead possess another class of enzymes referred to as PG *O*-acetyltransferase (PAT) proteins [138].

3.1.2.3 Surface polymer attachment

Cell surface polymers are implicated in virulence of certain bacterial species and are a major component of Gram-positive bacteria cell wall. Surface polymers such as the anionic teichoic acids, capsular polysaccharides and arabinogalactans are linked to the glycan strands by a phosphodiester bond to the C6-OH of MurNAc residues [134].

3.1.3 Lysozymes

3.1.3.1 Overview

Lysozyme is a broad term given to enzymes that are active upon a peptidoglycan substrate. Lysozymes belong to a larger lysozyme superfamily, including different classes of glycoside hydrolases (GH), which act on various substrates. The CAZY database classifies GH enzymes based on their primary sequence, protein fold and substrate specificity. The lysozyme superfamily includes the GH families: GH19 – plant chitinase, GH22 – C-type lysozyme (chicken-type) and I-type lysozyme (invertebrate-type), GH23 – G-type lysozyme (goose-type) and LT (bacterial lytic-transglycosylases), GH24 – V-type lysozymes (Viral-type) and GH46 – chitosanase [141].

Lysozymes hydrolyse the β-1-4 linked glycosidic linkage between alternating MurNAc and GlcNAc residues of peptidoglycan and are widespread enzymes in nature [142]. In higher organisms lysozymes are used as a defence against bacterial invasion. LTs,

though not strictly a hydrolase, have structural and functional similarities to lysozymes and provide a housekeeping role in bacteria by remodelling the cell wall during growth and division. Furthermore, LTs are required to clear gaps in the cell wall to facilitate the insertion of protein complexes that extend through the cell wall, such as pili, flagella and secretion systems [143]. Chitinases catalyse the breakdown of chitin, a GlcNAc polymer linked by β-1-4 bonds [144]. Chitinases are found in organisms that require their own chitin to be remodelled or to degrade chitin as a source of energy. Chitosanases on the other hand break down chitosan, a polymer of β-1-4 linked GlcN that retains a proportion of GlcNAc. Chitosanases are found in bacteria and multiple plant species where they provide defence against fungal species that possess a high proportion of chitosan in their cell walls [145].

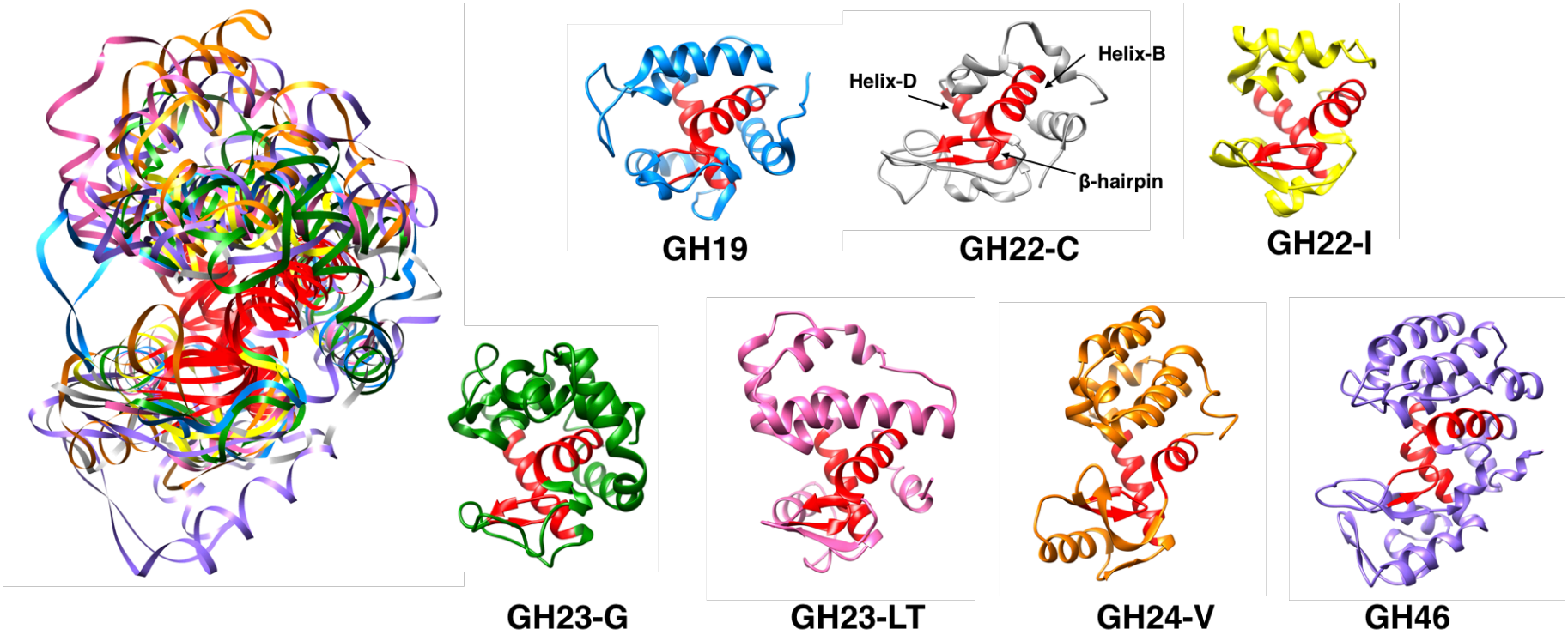


Figure 9. Structures of lysozyme superfamily members
The top left images shows the superimposition of seven representative structures of lysozyme superfamily members. Each representative is also displayed in an individual panel in the same relative orientation. The conserved core region (2 alpha helices and strict beta hairpin) is coloured in red in all images. GH19 –blue (PDB: 3W6D), GH22-C –gray (PDB: 2VB1), GH22-I – yellow (PDB:2DQA), GH23-G – green (PDB:154L), GH23-LT – pink (catalytic domain of PDB:1LSY), GH-24-V – orange (PDB:2LZM), GH46 – purple (PDB:1QGI). The structural elements of HEWL (GH22-C) are labelled for reference.

3.1.3.2 Structures

Hen egg white lysozyme (HEWL) was the first enzyme to have its three-dimensional structure solved in 1965 [146]. Since then a wealth of enzyme structures have been solved including many other lysozymes. Although the primary sequence similarities of lysozyme family members differs substantially, superimposition of their structures reveals a common fold/core. Lysozymes fold into two domains separated by a pronounced active site cleft. The larger of the two domains is formed mainly of an α-helical domain, whilst the smaller domain consists of three anti-parallel β-strands that form a single β-sheet and one wall of the substrate-binding cleft. [141]. Two α-helices (B and D – HEWL annotation) and two β-strands that form a strict β-hairpin are structurally invariant features of the five GH-families [141]. Each of the five GH families is represented in Figure 9 to highlight the variation (and similarity) in structures.

Lysozyme superfamily members possess a strictly conserved ES-motif located on the end of the central α-helix (helix-B). The catalytic glutamate on the end of the α-helix faces into the substrate-binding cleft. Typically an Asp residue (also catalytic) is located on one of the opposing β-strands, also facing into the substrate-binding cleft. The location of this second catalytic residue (strand-2 or strand-3) generally dictates the catalytic mechanism of the enzyme. The distance between the two catalytic residues suggests whether the enzyme mechanism will proceed via a retaining or an inverting mechanism (Figure 10). Retaining enzymes will typically have a distance of <6Å between the two catalytic residues, whereas inverting enzymes tend to be spaced >7Å [147]. In addition to the acid-base pair a serine residue is usually located within the active site, which is reported to position the attacking water molecule [148].

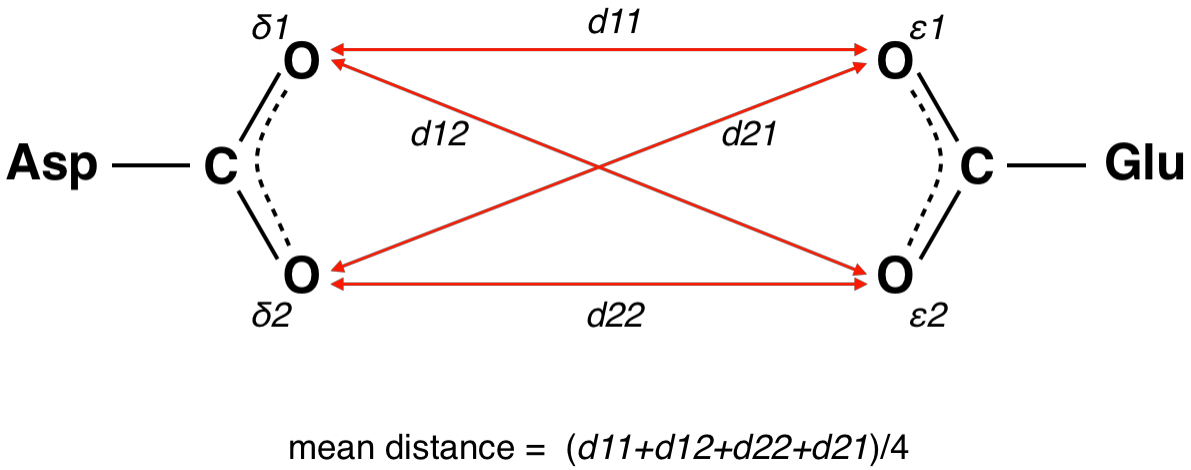


Figure 10. Catalytic distances
The average distance between the catalytic residues is calculated by taking the mean average distance between the carboxyl groups of the catalytic residues.

Beyond the active site, conserved residues contribute to the hydrophobic core of the enzymes. Bulky hydrophobic residues line helix-B, several residues towards the N-terminus of the catalytic glutamate. These hydrophobic residues form a hydrophobic pocket with conserved residues located on the invariant helix-D. Furthermore a strictly conserved glycine residue is located on strand-2, which maintains packing [141].

The stable folding pattern of lysozyme superfamily members is suggested to have originated from a common ancestor [149]. The invariant lysozyme core elements (helix-B, helix-D and two β-strands with a strict β-hairpin) alongside the conserved catalytic residues suggest the common fold did not arrive through convergent evolution. Moreover, the presence of non-essential elements that append the core structure suggest divergence from a common ancestor.

3.1.3.3 Catalytic mechanism

There are two reported catalytic mechanisms utilised by lysozymes, which have been determined using biochemical and structural data. Lysozymes and other GH family members are considered to either be retaining or inverting enzymes, depending on the product created after the glycosidic linkage is cleaved. Retaining enzymes use a double displacement mechanism that creates a covalent glycosyl-enzyme intermediate. The reaction retains the original stereochemistry at the anomeric centre. Retaining enzymes possess two acidic catalytic residues, where one carboxyl group acts as an acid/base while the other acts as a nucleophile.

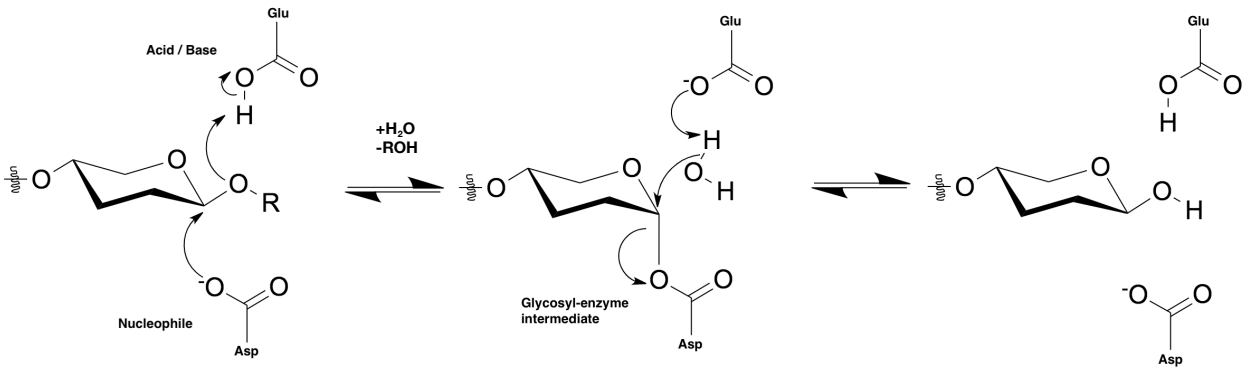


Figure 11. Retaining enzyme mechanism
Proposed mechanism for retaining lysozymes. ROH represents the leaving group.

GH-22 family member, HEWL, is the most well-characterised lysozyme in terms of structure, substrate binding and catalytic mechanism. The retaining lysozyme mechanism is exemplified by HEWL. The structure of HEWL has been solved in complex with various ligands revealing six sugar subsites, A-F [150]. HEWL possesses two catalytic residues, with E35 (helix-B) acting as a proton donor and D52 as a nucleophile [132]. It was initially suggested from structural evaluation that D52 in HEWL played a

role in stabilising an oxocarbenium ion intermediate. However, there is now substantial evidence for the formation of a covalent glycosyl enzyme intermediate [132]. The catalytic mechanism of HEWLs is therefore proposed to proceed by a classical Koshland retaining mechanism (Figure 11).

Inverting enzymes on the other hand function through a direct displacement mechanism, which leads to an inversion of stereochemistry at the anomeric centre. Inverting enzymes usually contain two catalytic residues. However, there are examples were only a single catalytic residue is required. In a double acid mechanism one carboxyl group is used as a general acid, while the other acts as a general base [151].

GH-23 member G-type lysozyme is the most well-characterised enzyme for this type of mechanism. Until recently G-type lysozymes were believe to only have a single catalytic residue. Structural studies of goose egg white lysozyme (GEWL) revealed E73 was located in the classical lysozyme ES-motif [152] and was suggested to act as a general acid/base in the absence of a second acidic residue. More recently the structure of another G-type lysozyme (from Atlantic cod) was solved, revealing D101 and D90 as two potential candidates for the second acidic residue [153]. Mutations of both Asp residues individually and together, suggest that both are able to participate in catalysis. However, D101 mutation was more detrimental, suggesting the catalytic mechanism generally proceeds via D101. D101 and D90 are spatially equivalent to D97 and D86 of GEWL. D86 of GEWL is located in a β-turn, which does not face into the active site. It has therefore been suggested that this region switches conformation upon substrate-binding to facilitate catalysis. G-type lysozymes perform catalysis though a typical inverting mechanism [153] outlined in Figure 12.

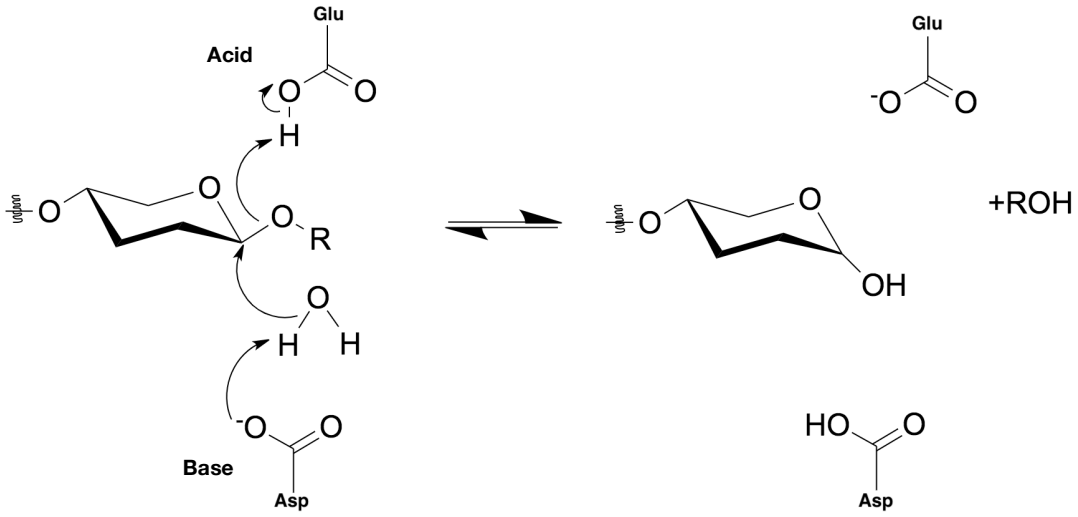


Figure 12. Inverting enzyme mechanism
Proposed mechanism for inverting lysozymes. ROH represents the leaving group.

3.1.3.4 Lytic transglycosylases – housekeeping enzymes

LTs are peptidoglycan-degrading enzymes that fulfil a housekeeping role. LTs are integral to PG maintenance, turnover and remodelling during cell division and cell growth [154]. Like classical lysozymes they cleave the β-1-4 linkage between MurNAc and GlcNAc of PG. However, they do not perform catalysis through hydrolysis of the substrate but rather use an unusual intra-molecular glycosltransferase reaction, which produces a terminal 1,6-anhydro-MurNAc residue [117]. Catalytically LTs perform quite a different reaction to classical lysozymes. LTs possess the invariant catalytic glutamate located in the ES-motif on the central α-helix. However, they lack a second catalytic residue and do not require a catalytic water molecule. Instead LTs use a substrate-assisted catalytic mechanism [155]. The catalytic glutamate is proposed to serve initially as the general acid, donating a proton to the glycosidic linkage and forming a muramoyl oxazolinium-ion intermediate. The glutamate then switches roles and acts as a base to abstract the C6 hydroxyl proton, leading to the formation of the 1,6- anhydromuramoyl product [154].

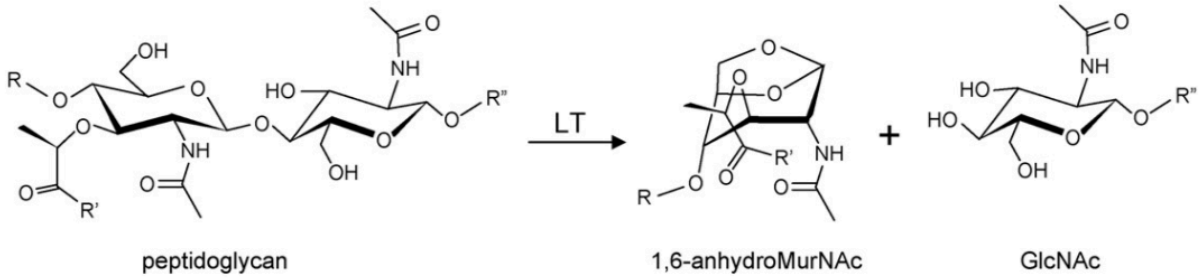


Figure 13. Lytic Transglycosylase reaction.
Lytic Transglycosylases catalyse the cleavage of the β-1-4 linkage between MurNAc and GlcNAc of PG with a single catalytic residue. Cleavage by LTs results in the formation of a 1.6-anhydroMurNAc terminal residue.

LTs can be divided into four families (1-4) and several *E. coli* LTs have been structurally and biochemically characterised for families 1 and 2 [156]. LTs are common in Gram-negative bacteria and always found in the periplasm. A small number of LTs have been reported in Gram-positive bacteria [157]. LTs can exist as soluble species or attached to the outer membrane via the N-terminus. Structures of the LT catalytic domain adopt a common fold that resembles G-type lysozymes [158].

3.1.4 *Bdellovibrio* lifecycle and PG modifications

The obligate predatory bacterium, *Bdellovibrio bacteriovorus*, has adapted PG modifying enzymes to facilitate its unusual biphasic predatory lifecycle. Early investigation into predation by *Bdellovibrio* highlighted a number of prey cell wall modifications required for successful completion of the predatory lifecycle and hinted at the enzymes responsible.

During penetration of an *E. coli* prey, *Bdellovibrio* rapidly solubilises approximately 10-15% of the prey peptidoglycan in a local, controlled manner [45]. Two enzyme

activities were identified during invasion: a glycanase and a peptidase. Shortly after invasion is complete the peptidoglycan solubilising activity of the glycanase is abruptly halted, while the peptidase activity continues to cleave DAP at a gradually diminishing rate [45]. The two enzyme activities were believed to create the initial entry pore that allowed *Bdellovibrio* to traverse the PG layer and reside in the intraperiplasmic space. The glycanase activity was associated with solubilising a localised region of prey cell wall for entry, while the peptidase activity was associated with cleaving DAP more globally. Removal of DAP annuls PG cross-linking and was suggested to be the cause of prey cell rounding and bdelloplast formation. Following the initial period of DAP cleavage is an enzyme that reincorporates DAP into the prey PG, which aids to stabilise the bdelloplast structure [51].

Later work highlighted other prey cell wall modifications made during predation. A GlcNAc-deacetylase activity was identified during the early stages of predation, which deacetylates the prey cell wall. Around 60-70% of the prey cell wall is *N*-deacetylated, converting it into a lysozyme resistant substrate early in the predatory cycle [50]. Chemical acetylation of lysozyme resistant bdelloplasts restored sensitivity to lysozyme. It was therefore hypothesised that GlcNAc-deacetylation was required to halt the initial pore forming glycanase activity and prevent premature prey cell lysis. This hypothesis was later shown to be incorrect given that a deacetylase-null *Bdellovibrio* bdelloplast does not lyse (detailed later in this section).

During predation, a large proportion of Braun lipoprotein is lost from the prey cell wall [50]. In Gram-negative cell walls the Braun lipoprotein provides structural integrity to the outer membrane. Removal of the Braun lipoprotein in bdelloplast would therefore

weaken the structural integrity of the outer membrane. A protease, able to cleave between the amino group of the C-terminal lysine residue and the D-carboxyl group of DAP, was proposed to perform the removal of the Braun lipoprotein [50].

Additionally, fatty acids are covalently attached to the prey cell wall during predation. The fatty acids were shown to be predominantly palmitic and oleic acids that were linked via an *O*-acyl group [52]. Attachment would require an enzyme possessing an acylase activity to attach the fatty acids to the prey PG. Attaching long-chain fatty acids to the prey cell wall may act to stabilise the bdelloplast structure against osmotic stress. The fatty acids could associate with the phospholipids of the outer membrane and perform a stabilising role, compensating for the removal of the Braun lipoprotein.

Finally a lytic enzyme activity was identified towards the end of the growth phase, immediately prior to prey cell lysis [45]. The prey cell wall is completely degraded upon completion of growth by a newly synthesised enzyme(s) that are able to act on the GlcNAc-deacetylated substrate [45, 52]. Furthermore the final lytic enzyme was shown to only cleave bonds in the glycan portion of PG that were GlcNAc-deacetylated, suggesting specificity.

More recently, studies have been aided by the sequenced genome of *Bdellovibrio HD100.* Consequently there has been an increase in our understanding of prey cell wall modifications during the predatory lifecycle. Firstly two genes *bd0816* and *bd3459*, encoding proteins with limited homologies to PBP4 DD-endopeptidases, are expressed 15 minutes after the predatory cell invades [47]. PBP4 proteins are implicated in cutting and remodelling peptidoglycan by cleaving the D-Ala-*meso*-DAP cross-links. As

such PBP4 proteins typically function as housekeeping enzymes during peptidoglycan growth and division. *Bdellovibrio,* however, has adapted these enzymes into a secreted predator-specific protein used to remodel the prey cell wall [47]. Bd0816 and Bd3549 are responsible for prey cell rounding and creating the osmotically stable bdelloplast niche. A double *Δbd0816 Δbd3459* mutant resulted in a number of invaded cells that did not round and retained the native prey cell shape after invasion. Besides this, Bd0816 and Bd3549 activities also enable significantly faster invasion speeds and prevent wasteful double invasion events. Bd0816 and Bd3549 are proposed to reduce the torsional stress of the prey cell wall by de-cross-linking the prey PG, thereby allowing *Bdellovibrio* to enter more quickly. Furthermore the bdelloplast cell wall is more deformable than uninvaded prey, which may impact on the ability of subsequent *Bdellovibrio* to invade via the proposed inward traction of the type IV pili, which putatively attach to the PG layer [41].

The structure of Bd3459 was solved, demonstrating divergence from a standard PBP4 fold, which facilitates Bd3459s ability to rapidly decross-link a diverse range of PG chemistries. Bd3459 lacks PBP4 domain III, which typically gates access to the active site. The lack of this regulatory domain demonstrates how a greater degree of substrate freedom is tolerated [47].

Secondly, the genes responsible for GlcNAc-deacetylation of the prey cell wall material have been identified (*bd0468* and *bd3279*) [49]. A double *Δbd0468Δbd3279* mutant was shown to leave behind a significant proportion of prey cell wall after predation. The residual prey material or ‘ghosts’ consist of undigested prey PG surrounded by the outer-membrane and small amounts of the inner membrane. The finding that stable

'ghost' cells remained in GlcNAc-deacetylation mutants contradicts Thomashow and Rittenberg's hypothesis that GlcNAc-deacetylation halts the initial pore forming glycanase activity [45, 50], as there was no documented evidence of premature prey cell lysis; instead the prey cells remained undigested upon completion of predation. This finding suggests that GlcNAc-deacetylation marks the prey cell material for digestion at a later stage by an enzyme(s) that is able to act upon the modified substrate. Furthermore the *Δbd0468Δbd3279* mutant prey cell 'ghosts' were susceptible to digestion by lysozyme [49], confirming the initial observation that bdelloplasts become resistant to lysozyme activity [50]. In addition *Δbd0468Δbd3279* mutant predators entered the prey more slowly and a large proportion of them did not cross the peptidoglycan layer and instead uncharacteristically grew and replicated between the outer membrane and PG layer. This finding suggests that GlcNAc-deacetylation of the prey cell wall actually aids the activity of the pore-forming glycanase rather than inhibiting it. *Δbd0468Δbd3279* mutant progeny were also markedly slower at exiting the prey cell and many of the predators that managed to cross the prey cell wall initially exited through a single small pore, which has been suggested to be the resealed entry pore [49].

The structure of Bd3279 was solved, showing a similar overall structure to other published GlcNAc-deacetylases such as PgaB [159]. The Bd3279 structure revealed a N-terminal active site capping domain that was suggested to be required to regulate the activity of the enzyme [49].

3.1.5 Introduction to the Exit enzyme

In order to escape/lyse the bdelloplast structure at the end of the predatory lifecycle, *Bdellovibrio* must solubilise the remaining prey cell PG. Previous studies indicated an increase in a PG solubilising/lytic enzyme towards the end of the predatory lifecycle [45]. The enzyme responsible for bdelloplast lysis would need to have activity towards the modified prey cell wall and be expressed at the correct time (immediately prior to lysis). Furthermore GlcNAc-deacetylation chemically differentiates the predator and prey cell walls; hence specificity towards prey cell wall would be required [49].

Unpublished work by the Sockett lab has indicated a possible enzyme candidate that may perform the final solubilisation of the prey cell wall. A series of putative lysozyme genes from *Bdellovibrio HD100* were analysed by qPCR to determine their expression profiles. The *bd0314* gene was discovered to encode a putative lysozyme upregulated in the later stages of the predatory lifecycle. Furthermore, a *Δbd0314 Bdellovibrio* mutant rendered replicating *Bdellovibrio* trapped within bdelloplasts (which ultimately escaped after prolonged time in bdelloplast) and left behind prey cell wall PG 'ghosts' reminiscent of the GlcNAc-deacetylase knockout mutants [49].

3.1.6 Aims and Objectives

The aim of this current study was to determine whether Bd0314 possessed the capability to perform the role of the exit enzyme and shed light on how predators exit their prey. Structural and functional studies were conducted to reveal the adapted lysozyme-like fold of Bd0314 and confirm its specificity towards a *N*-deacetylated PG substrate.

3.2 Results

3.2.1 Purification and Crystallisation

The catalytic domain of Bd0314 (R74-K254) was cloned into a modified pET28b plasmid containing a TEV cleavable, N-terminal MBP-(His_6) tag, (Figure 15). The MBP tag was required to suppress toxicity and enhance solubility. Bd0314 was expressed using the commercial SHuffle *E. coli* expression system to correctly fold disulphide bonds in the cytoplasm. Bd0314 was purified by performing two rounds of IMAC; the first to purify MBP-tagged Bd0314 and the second to remove MBP after O/N incubation with TEV. Bd0314 generally crystallised as small, thin, stacked plates that were difficult to manipulate, (Figure 14). Larger crystals were optimised and used for SSAD phasing experiments. (See Section 2.2.5 for more detail)

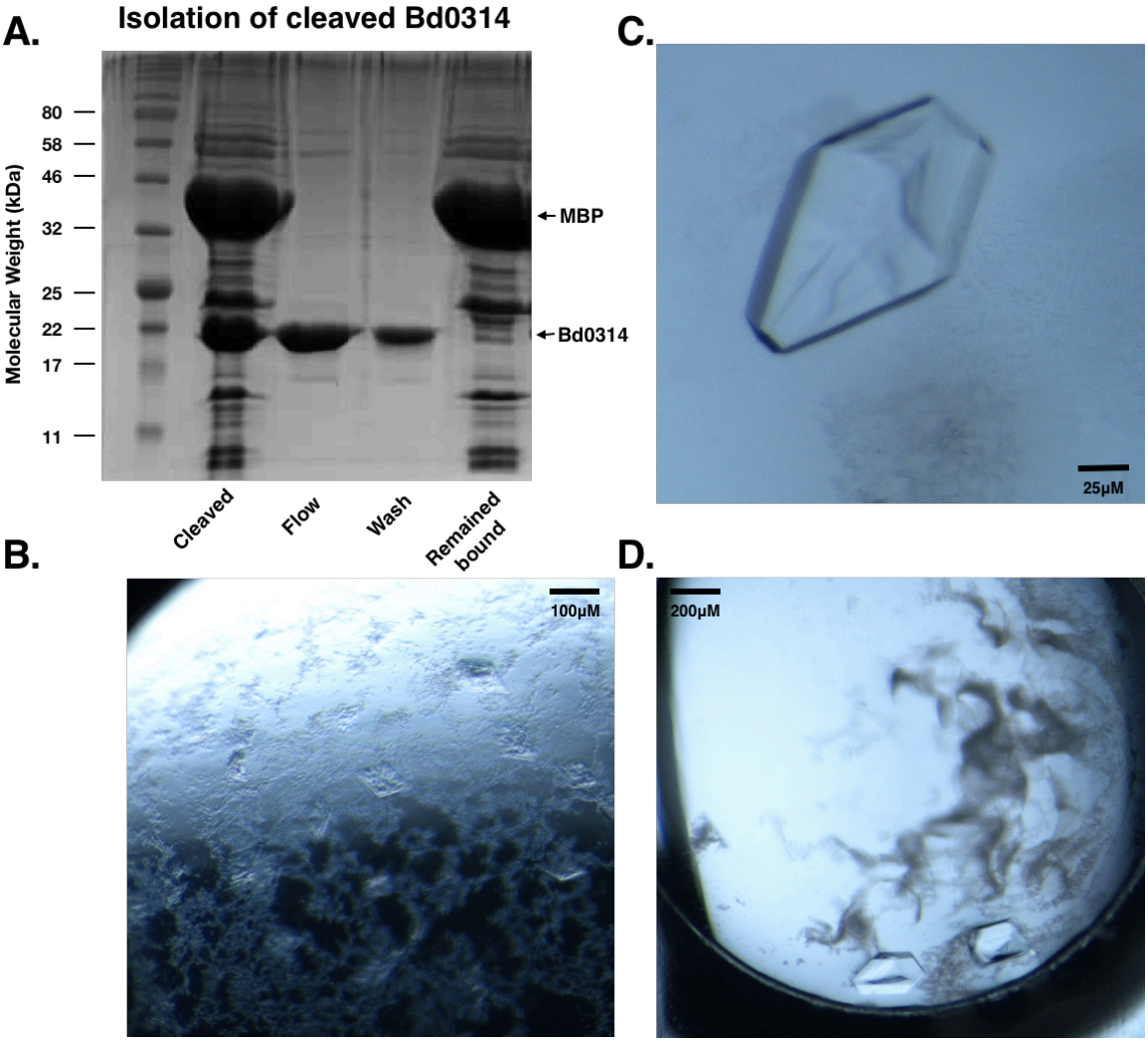


Figure 14. Bd0314 purification and crystallisation
A. An MBP tag was used to express and purify Bd0314. The MBP tag was cleaved by incubation with TEV and removed by a second passage over a HisTrap column. **B.** Bd0314-Wt crystals were small thin plates. **C.** The Bd0314-E143Q crystal used for SSAD phasing was optimised. **D.** Bd0314-Wt crystals that grew in a different space group.

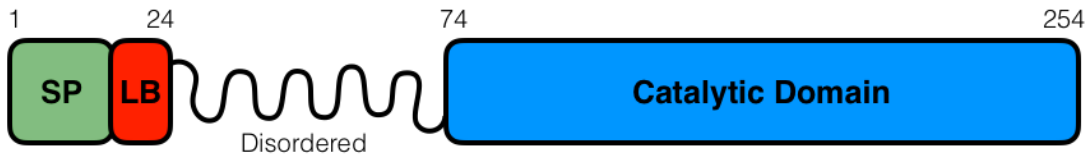


Figure 15. Bd0314 structure schematic
Bd0314 is a 254 residue protein that has a 180 residue catalytic domain. Bd0314 primary sequence suggests the presence of an N-terminal signal peptide (SP) containing a Lipobox (LB). Between the predicted lipobox and the catalytic domain is a 50 residue region of predicted disorder. Secondary structure predictions made using the XtalPred server.

3.2.2 The structure of Bd0314

3.2.2.1 Overall fold

The structure of the catalytic domain (R74-K254) of Bd0314 resembles a lysozyme-like fold, formed by two individual folding domains, separated by a large substrate-binding cleft (Figure 16). The larger core domain is mainly composed of α-helices (nine helices, A-F and I-L) with a short two-stranded anti-parallel β-sheet projection (strand-3 and strand-4). This domain contains the invariant lysozyme α-helices (helix-F and helix-I). Helix-F possesses the classical lysozyme ES-motif, with E143 (conserved catalytic residue) located on the end of the helix facing into the cleft. The smaller domain is a mixture of α-helices and β-strands (helix-G and H, strands-1 and 2). The two strands form an anti-parallel β-sheet, which forms base of the substrate-binding cleft. Strand-2 (the longer β-strand) runs continuously, whereas strand-1 has a single amino acid break in H-bonding (Asp82) that splits the strand into two parts. The kink in strand-1 projects the sidechain of E154 up into the cleft, directly opposite E143, suggesting a role as a possible second catalytic residue.

The core region of Bd0314 is stabilised by hydrophobic interactions between side chains residues of the invariant lysozyme helices (helix-F and helix-I). In addition Bd0314 contains two disulphide bond pairs; C182-C199 and C117-C253, highlighted in yellow in Figure 16, that are required for stabilisation of the protein. The C117-C253 disulphide bond ensures that the C terminus of the protein remains tightly associated with the protein core. The C182-C199 disulphide bond supports the interaction between domains and stabilises the base of the substrate-binding cleft. The Bd0314 fold is also stabilised by salt bridges between R74 and E81 and between R227 and D211. The base of the substrate-binding cleft is stabilised by a number of hydrogen

bonds. E175 located on helix G hydrogen bonds to S159, which stabilises the end of the β-sheet and turn. In addition Bd0314 may contain a third β-strand, where I172 (located immediately after the β-hairpin) and R167 of strand-2 make a β-sheet-like interaction.

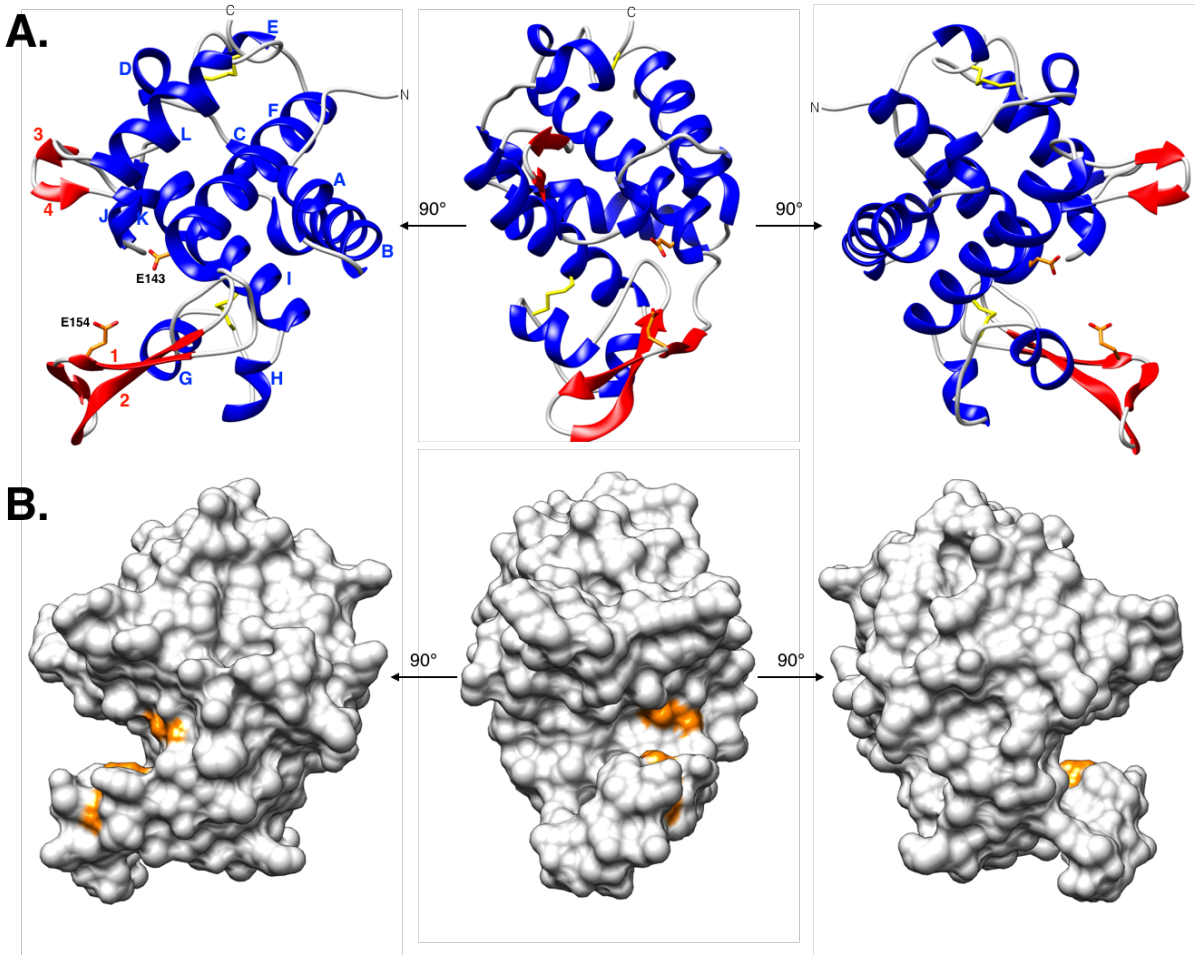


Figure 16. Bd0314 structure
A. The secondary structure of Bd0314 shown from three angles 90° apart. α-helices are coloured blue and β-strands are coloured red. Disulphide bonds are highlighted in yellow and the positions of the catalytic E143 and E154 are displayed in orange. Helices are labelled with blue letters and strands are labelled with red numbers in the first panel. **B.** Surface view of Bd0314 illustrating the active-site cleft. The location of the catalytic residues is highlighted in orange.

3.2.2.2 The catalytic site

Analysis of the Bd0314 structure revealed two potential catalytic residues. The classical ES invariant glutamate (E143) is located on the end of helix-F facing into the catalytic site. The second (E154) is located on strand-1, and is projected into the active site, directly opposite E143. The typical location of the second catalytic Asp in the lysozyme superfamily is on strand-2. However, in Bd0314 S166 occupies this location and the hydroxyl group makes a hydrogen bond to the E154 carboxyl-group, which may help to retain the orientation of E154. The presence and location of E143 and E154 suggest they act as a general acid-base pair. The distance between the two acidic residues is 8Å, suggesting that Bd0314 proceeds by an inverting catalytic mechanism [147].

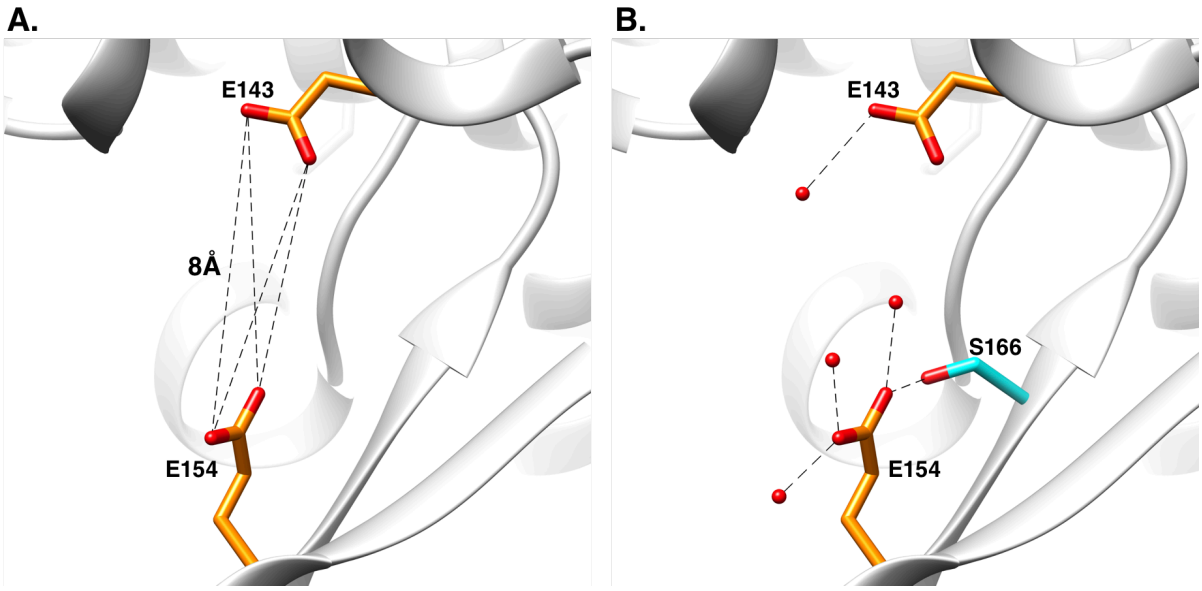


Figure 17. Bd0314 catalytic residues
A. The distance between the carbonyl groups of E143 and E154 is ~8Å, which suggests Bd0314 is an inverting lysozyme. **B.** E143 and E154 make hydrogen bonds with water molecules in the active-site cleft (potential catalytic water molecules). E154 also makes a hydrogen bond with S166, which may have stabilising effects during catalysis.

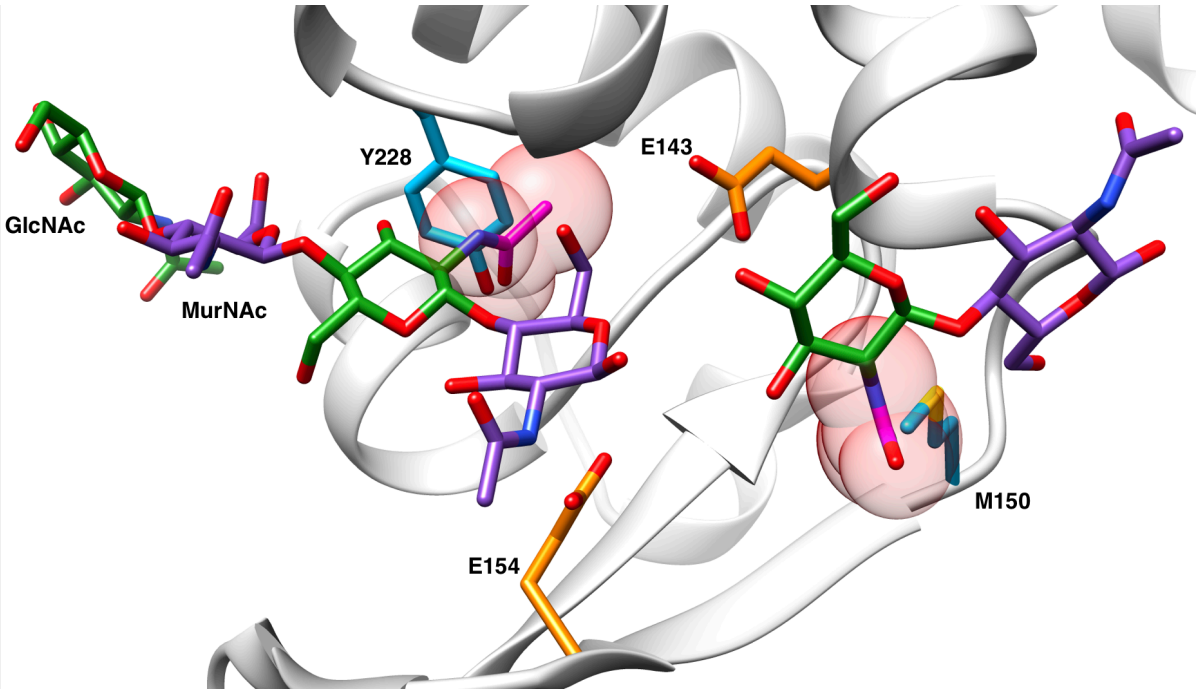


Figure 18. Bd0314 substrate-binding site
Model of Bd0314 substrate binding, generated by superimposition of a regular lysozyme (PDB:1LZS) in complex with a hexa-*N*-acetyl-chitohexaose substrate. Bd0314 was structurally aligned with PDB:1LZS by least-squares minimization of the structurally conserved regions (helix-F, helix-I and active site beta-hairpin) using the CCP4 suit superpose software (length 2474 atoms from RSMD=1.13Å). Bd0314 is displayed in ribbon form, coloured in white. The ligand is alternatively coloured green (GlcNAc) and purple (MurNAc), to represent the positions of GlcNAc and MurNAc for a peptidoglycan substrate similarly to that in Figure 8. The substrate GlcNAc residue acetyl-group faces inwards towards the cleft, whereas MurNAc residues face outwards to accommodate the peptide stem. The catalytic glutamate residues (E143 and E154) of Bd0314 are shown in orange and are correctly positioned to cleave after the MurNAc residue. The acetyl-groups of GlcNAc residues clash with residues of the peptide-binding cleft. The residues that clash with the acetyl-group of GlcNAc residues (Y228 and M150) are shown in cyan. The red van der waals surface of the acetyl groups demonstrates the degree of clash.

3.2.2.3 Substrate selection

Bd0314 has an extensive substrate-binding cleft provided by the extended β-sheet, which forms the base and the additional β-sheet region of the core domain, which creates a peak to the roof of the binding cleft. We have been unable to obtain a protein:ligand complex after many attempts with various ligands. However, numerous structures of lysozyme-superfamily members in complex with ligands have been published [153, 162, 163]. Superimposition of these ligand complex structures with Bd0314 (using LSQ superimposition of helix-F (residues 133-144), helix-I (residues

193-202) and the strict β-hairpin (residues 165-172)) places the ligand into the substrate-binding cleft in a mechanistically plausible orientation (Figure 18). The superimposed complex structure (PDB:1LZS) was used to generate a model for Bd0314-ligand interactions. Firstly the ligand placement supports the concept of E143 and E154 acting as the general acid-base pair during catalysis, cleaving after MurNAc residues. Secondly GlcNAc residues always face into the cleft, whereas MurNAc residues face outwards, presumably to accommodate the peptide stem at the O3 position. Interestingly the acetyl-groups of modelled GlcNAc residues clash with residues lining the Bd0314 substrate-binding cleft. These clashes indicate a possible mechanism for substrate specificity towards a *N*-deacetylated PG substrate. Y228 and M150 both adopt conformations that project into the substrate-binding cleft and occupy the location of the *N*-acetyl group of the GlcNAc residues (Figure 18). In Bd0314 Y228 and M150 likely act to sterically occlude acetyl-groups at this position and therefore prevent the accommodation of an acetylated PG substrate. The Y228 sidechain is more likely to be critical to selection because it is held tightly in this conformation by a hydrogen bond with S176 as well as backing onto a hydrophobic face provided by W229 and I203, (Figure 29). M150, on the other hand, has the freedom to adopt multiple conformations (two of which are can be modelled in the $2F_o$-F_c map for Bd0314). In addition to blocking the acetyl-group of GlcNAc, Y228 may also participate directly with binding the free amine group of GlcN residues. A trio of hydroxyl-groups provided by Y228, S176 and S173 create a triangle of H-bonding candidates that could co-ordinate the free amine group of GlcN. Although the free amine-group of GlcN residues is likely to be protonated (pK_a ~6.5) [133] and carry a positive charge, there are no negatively charged residues lining the substrate-binding site.

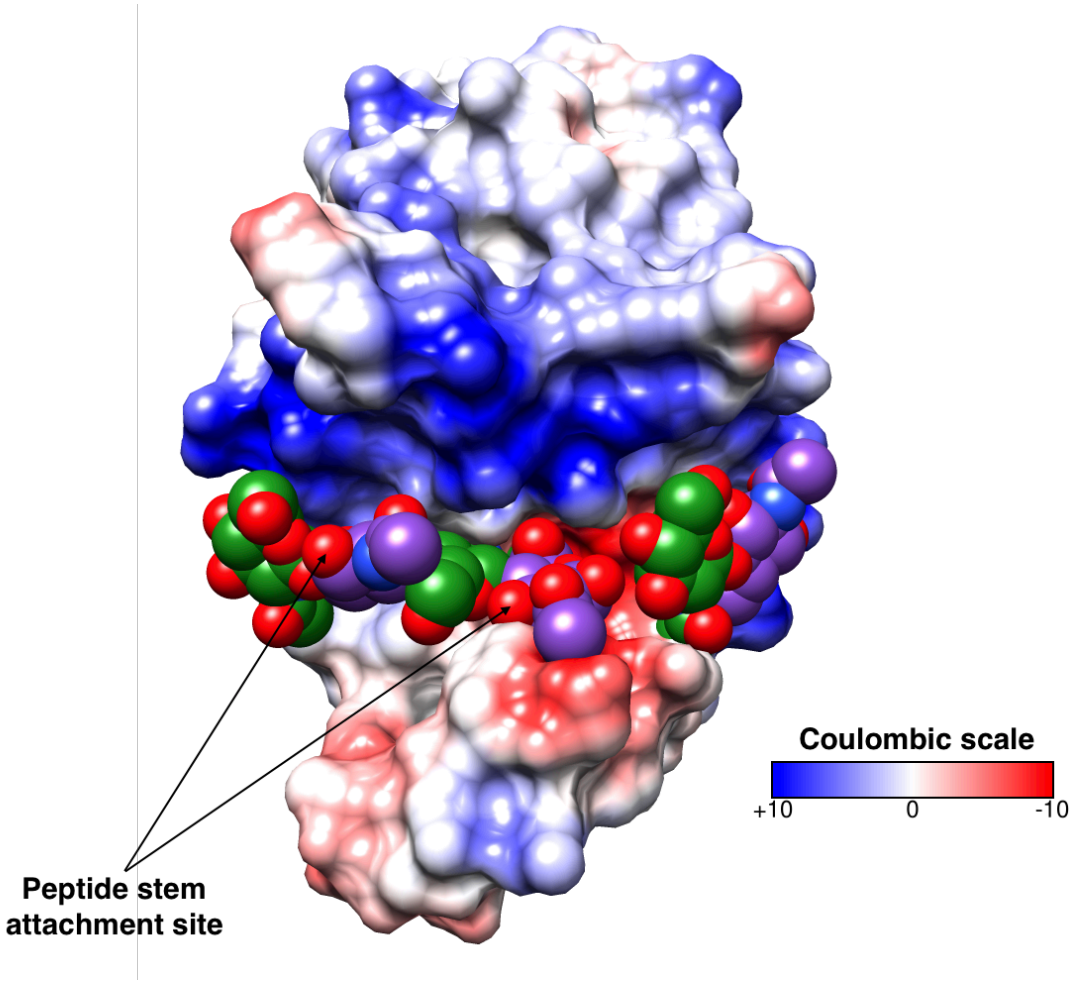


Figure 19. Substrate binding surface view
The substrate-binding model shown in the surface view demonstrates how the acetyl-groups of GlcNAc residues face inwards whereas the acetyl-group of MurNAc residues face outwards, so the peptide stem can accommodated. The roof of the substrate-binding cleft carries a positive charge. Substrate colours are the same as Figure 8.

3.2.2.4 Bd0314 structure comparison (DALI) and homologues

The DALI protein structure comparison server was used to identify proteins with similar structural elements from the PDB [160]. The top 25 matches from DALI are shown in Table 11. The greatest proportion of sequence and structural conservation for Bd0314 occurs for the protein core and the catalytic cleft. This region includes (residues 130-154) the central helix-F and part of the strand-1 and (residues 164-207) part of strand-2 up to and including central helix-I (Figure 20). Lytic transglycosylase

family proteins (GH23), are the most frequent structural matches. The catalytic core domain of Slt70 PDB: 1QTE (a soluble LT) gives the best structural match. However, the lower subdomain containing the second catalytic residue is not present in Slt70. Interestingly human lysozyme PDB:5LSH, is one of the closest structural matches and has a high Z-score, low rmsd. Four regions of Bd0314, residues 1-14, 26-54, 82-90 and 143-150, show little structural conservation to other lysozyme structures or primary sequences (Figure 28.B).

No:	Chain	Z	rmsd	lali	nres	%id	PDB	Description
1:	1qte-A	8.8	2.8	119	620	18	PDB	MOLECULE: SOLUBLE LYTIC TRANSGLYCOSYLASE SLT70;
2:	6cf9-A	8.8	2.7	112	516	13	PDB	MOLECULE: LYTIC TRANSGLYCOSYLASE;
3:	5lsh-A	8.4	2.6	103	130	15	PDB	MOLECULE: LYSOZYME C;
4:	4c5f-A	8.2	2.9	116	330	18	PDB	MOLECULE: MEMBRANE-BOUND LYTIC MUREIN
5:	4owd-A	8.1	3.2	118	429	16	PDB	MOLECULE: MEMBRANE-BOUND LYTIC MUREIN
6:	5vas-A	8.1	2.6	101	129	18	PDB	MOLECULE: LYSOZYME;
7:	6fcu-A	8.0	2.9	120	615	16	PDB	MOLECULE: SOLUBLE LYTIC MUREIN TRANSGLYCOSYLASE;
8:	4yim-B	7.9	2.8	116	586	13	PDB	MOLECULE: PUTATIVE SOLUBLE LYTIC MUREIN
9:	3bkh-A	7.8	3.0	113	268	10	PDB	MOLECULE: LYTIC TRANSGLYCOSYLASE;
10:	4dq7-A	7.7	3.1	108	168	12	PDB	MOLECULE: MEMBRANE PROTEIN PHI6 P5;
11:	3w6b-A	7.7	3.4	119	160	22	PDB	MOLECULE: LYSOZYME-LIKE CHITINOLYTIC ENZYME;
12:	2xqo-A	7.6	3.0	125	210	16	PDB	MOLECULE: CELLULOSOME ENZYME, DOCKERIN TYPE I;
13:	1lsg-A	7.4	2.6	101	144	21	PDB	MOLECULE: HEN EGG WHITE LYSOZYME;
14:	4xp8-A	7.0	2.9	84	88	19	PDB	MOLECULE: ETGA PROTEIN;
15:	1am7-A	6.8	3.6	102	154	7	PDB	MOLECULE: LYSOZYME;
16:	2dqa-A	6.7	2.9	99	123	13	PDB	MOLECULE: LYSOZYME;
17:	5e27-A	6.7	2.7	74	226	12	PDB	MOLECULE: RESUSCITATION-PROMOTING FACTOR RPFB;
18:	2dkv-A	6.0	4.0	140	286	10	PDB	MOLECULE: CHITINASE;
19:	5xsv-A	6.0	3.2	121	486	9	PDB	MOLECULE: CHITINASE;
20:	5nm7-G	5.9	3.1	117	263	12	PDB	MOLECULE: PEPTIDOGLYCAN-BINDING DOMAIN 1;
21:	4ok7-A	5.9	3.3	126	223	10	PDB	MOLECULE: ENDOLYSIN;
22:	4luq-A	5.6	3.5	125	409	10	PDB	MOLECULE: UNCHARACTERIZED PROTEIN;
23:	4g9s-A	5.4	3.2	106	187	16	PDB	MOLECULE: GOOSE-TYPE LYSOZYME;
24:	4fdy-A	4.9	2.9	98	295	10	PDB	MOLECULE: SIMILAR TO LIPOPROTEIN, NLP/P60 FAMILY;
25:	1d0k-A	4.9	3.2	111	318	14	PDB	MOLECULE: 35KD SOLUBLE LYTIC TRANSGLYCOSYLASE;

Table 11. DALI results
Top 25 matches from the DALI server for Bd0314 coordinates as the search query [160].

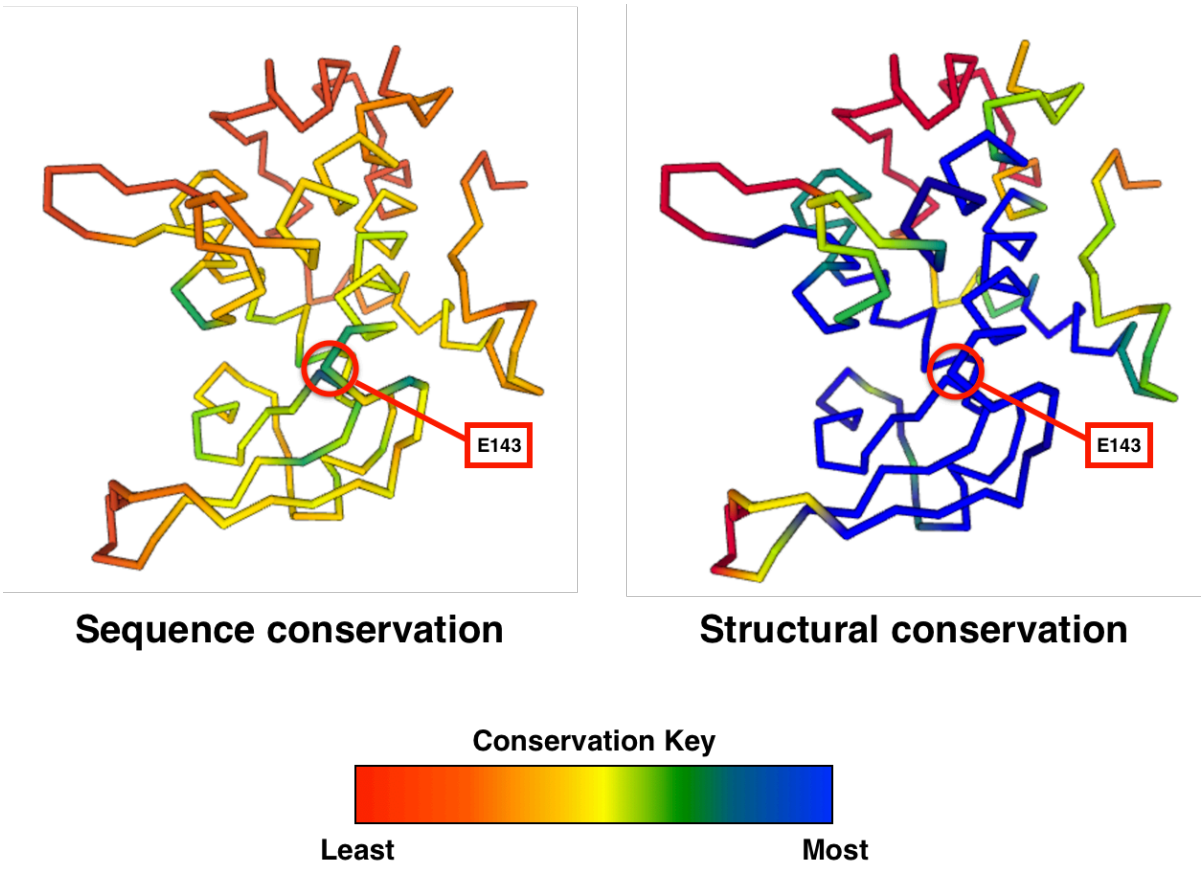


Figure 20. Bd0314 conservation
The sequence and structural conservation (coloured key) is mapped onto the Cα-trace of Bd0314 structure. Sequence conservation in Bd0314 is low, whereas the structural conservation is higher in regions. The core region of Bd0314 has the highest sequence and structural conservation. These results are obtained the DALI [160].

BLAST analysis of the Bd0314 primary sequence against a non-redundant database revealed a number of closely related homologues. Included in the results were two *Bdellovibrio bacteriovorus HD100* proteins (Bd1413 and Bd1440). Bd1413 and Bd1440 share 60% ID to each other and 33% ID to Bd0314 (Figure 21). Bd1413 and Bd1440 both possess the two catalytic residues highlighted in the structure of Bd0314 and they also possess the YW-motif (Y228) associated with GlcN selectivity. There are two regions of minor disparity in the sequence alignment between Bd0314 and Bd1413/Bd1440. Firstly Bd1413/Bd1440 has a 11aa insertion between helices F and G. Secondly the short β-sheet extension (strand-3 and 4) of the core domain exists as an 8aa insertion in the primary sequence of Bd0314.

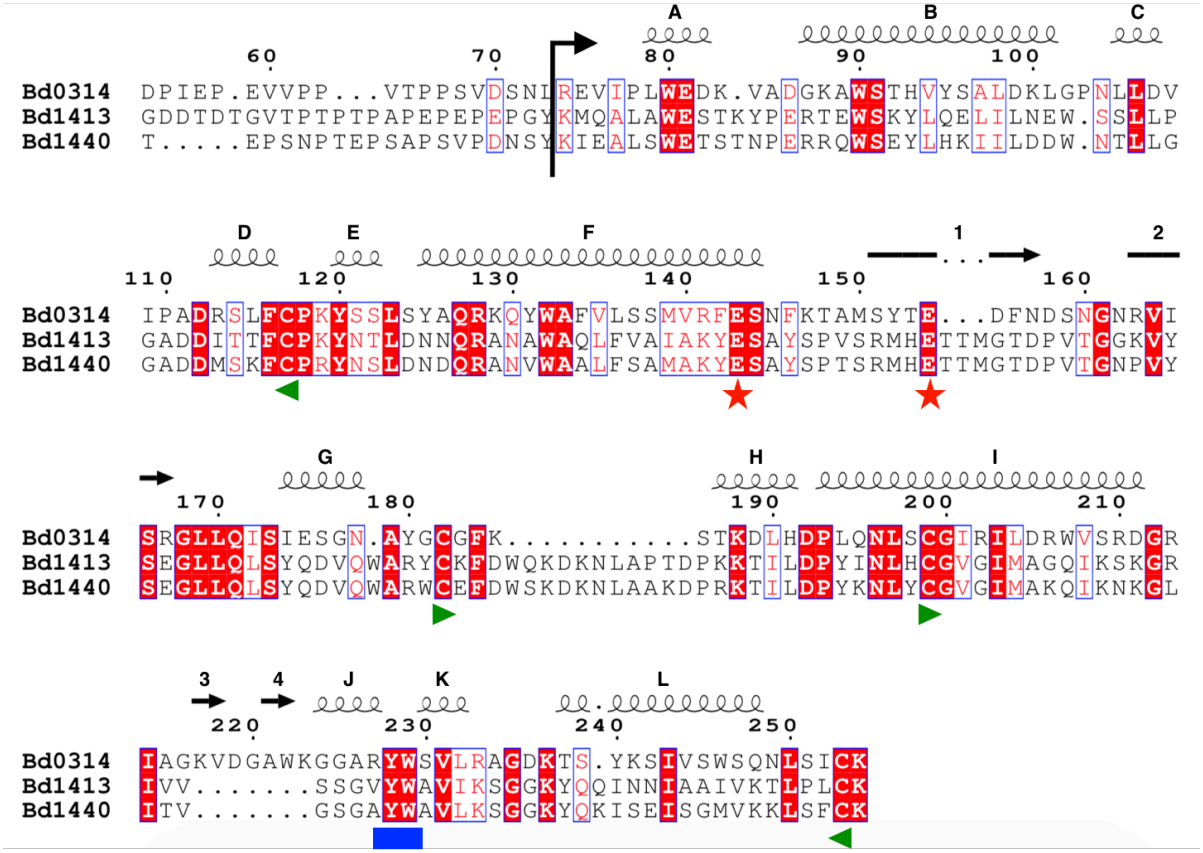


Figure 21. Bd0314 alignment with *Bdellovibrio HD100* homologues
Alignment of Bd0314 primary sequence with Bd1440 and Bd1413. Bd1440 and Bd1413 sequences are more closely related to one another (60% ID) than to Bd0314 (33% ID). Bd1440 and Bd1413 do not possess the second short β-sheet region of the α-helical domain. Bd1440 and Bd1413 also have an elongated region between helix-G and helix-H. The start of the catalytic domain is shown by a black line and arrow. The conserved cysteine residues are shown by green triangles; the direction of the green triangle represents the cysteine pairs. Red stars highlight the catalytic residues and the GlcNAc *N*-acetyl group blocking motif is shown by the blue rectangle. α-helices are represented by black hoops and β-sheets represented by black arrows. Drawn using Espript 3 [161].

BLAST analysis also revealed a number of Bd0314 homologues from other organisms that possess equivalents of both catalytic residues (E143 and E154) as well as the YW-motif (Y228). An alignment of selected homologues is shown in Figure 22. The alignment demonstrates the co-carriage of the second catalytic residue motif (TED) with the GlcN specificity motif (YW).

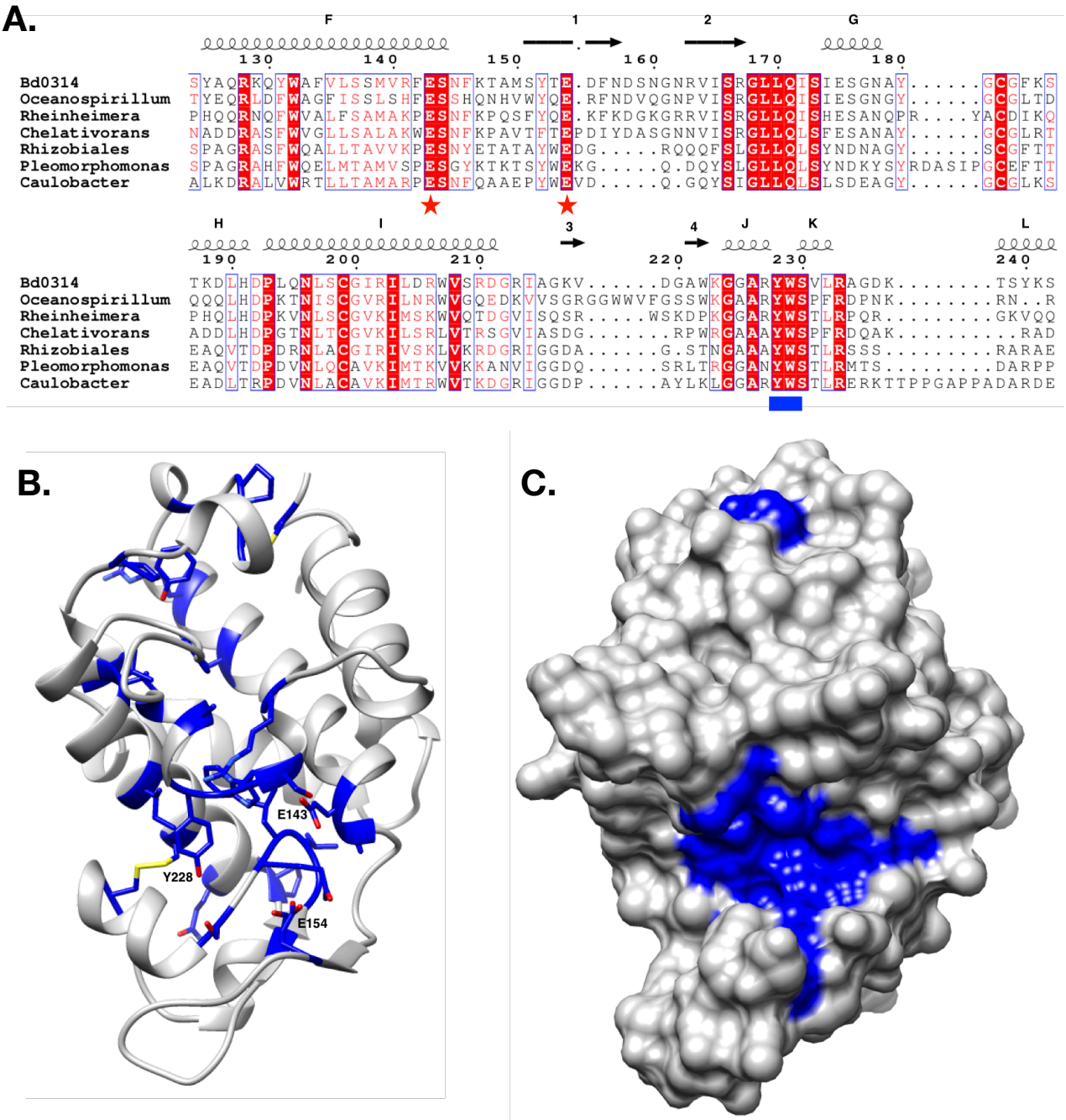


Figure 22. Bd0314 alignment and residue conservation
A. BLASTp searches reveal that homologues of Bd0314 exist in organisms other than predatory bacteria. All organisms possess CE4 deacetylases in their genomes. The presence of the second catalytic glutamate (E154) has a 100% co-carriage with the *N*-acetyl group blocking tyrosine (Y228) in these predicted *N*-deacetylation specific lysozymes. Catalytic residues are highlighted with blue stars and the *N*-deacetylation specific lysozyme motif is highlighted with a blue bar. **B.** Residue conservation (100% ID) mapped onto the secondary structure of Bd0314. **C.** Residue conservation mapped onto the surface of Bd0314.

3.2.3 Bd0314 activity and Specificity

3.2.3.1 Bd0314 specificity

We designed and implemented a new assay to determine the specificity of Bd0314 (see Section 2.2.2). Briefly, Fluorescein-5-Isothiocyanate (FITC) was used to label isolated *E. coli* PG and a subset was then *N*-deacetylated using recombinant *Bdellovibrio* GlcNAc-deacetylases (Bd0468 and Bd3279). The assay monitored the amount of liberated FITC material (solubilised after lysozyme incubation) by measuring fluorescence readings in the soluble material. Control samples with no enzyme (baseline) showed minimal fluorescence readings. Hence the enzyme activity is proportionate to the fluorescence readings (figure 24).

Figure 23, shows the results from the FITC-lysozyme assay. HEWL was used as a negative control in the assay because it is inactive towards GlcNAc-deacetylated PG. Mutanolysin was used as the positive control because it is active towards acetylated and deacetylated PG substrates.

Our assay demonstrates HEWL has diminished activity towards GlcNAc-deacetylated substrate (GlcN). The existence of residual HEWL activity towards the putative GlcN substrate suggested a proportion remained acetylated and that the GlcNAc-deacetylation reaction did not go to completion. Mutanolysin was active towards both GlcNAc and GlcN substrates to a similar extent.

The activity of Bd0314 revealed a distinct substrate preference. Bd0314 showed minimal activity towards the GlcNAc substrate, but displayed considerable activity

upon the GlcN substrate. The Bd0314 activity towards the GlcN substrate is comparable to the activities of the HEWL and Mutanolysin upon acetylated substrates. This result indicates Bd0314 possesses a high specificity towards GlcNAc-deacetylated PG substrates.

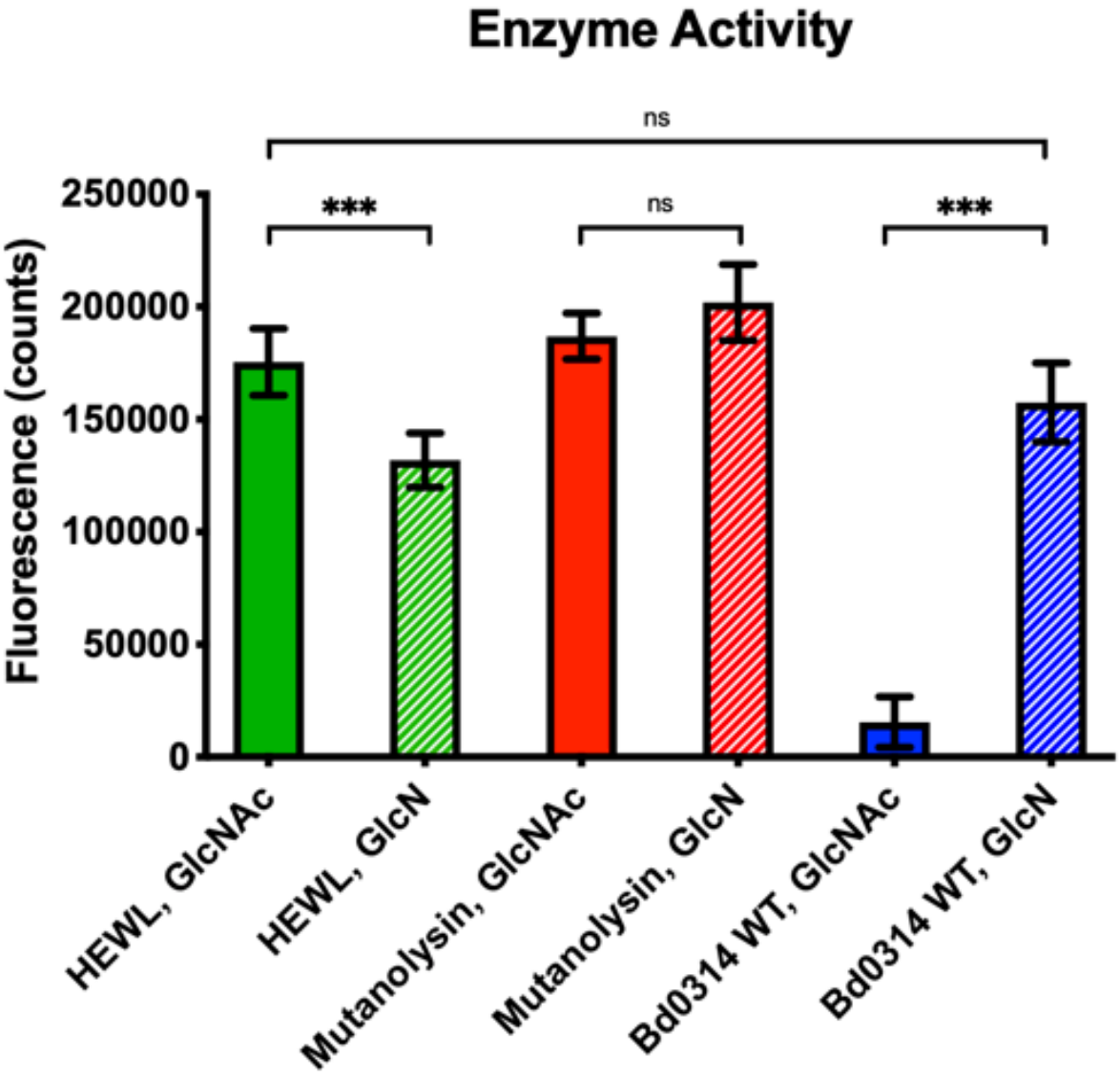


Figure 23. Bd0314 specificity
The results of the FITC-PG assay demonstrating Bd0314 specificity. Block coloured bars represent GlcNAc (acetylated PG) whilst striped bars represent GlcN (N-deacetylated PG). HEWL (Green) shows reduced activity with N-deacetylated substrate. Mutanolysin activity is similar for both substrates. Bd0314 shows specificity for the GlcN substrate. All data points were measured in at least triplicates. The error bars represent the standard deviation. Unpaired t-test was used to determine significance, *** = $p<0.0001$, ns = not significant.

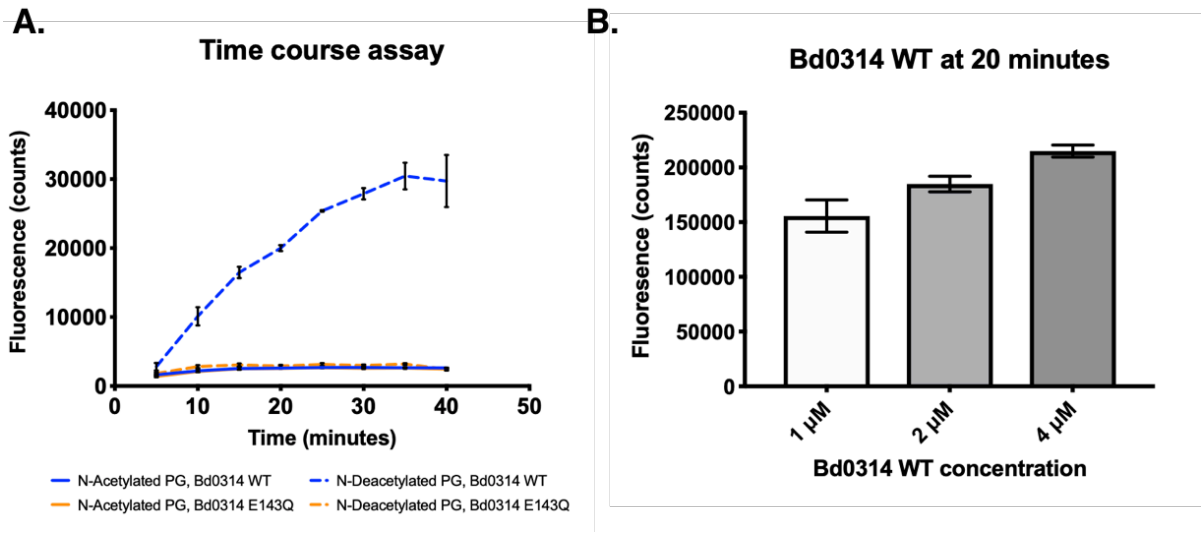


Figure 24. Time and Concentration dependent assay
A. Turnover of the FITC-PG substrate by Bd0314 is time dependent; activity is linear for the first 15 minutes. Bd0314 is only active towards the *N*-deacetylated PG substrate and shows minimal activity towards normal acetylated PG substrate. **B.** The FITC-PG assay is concentration dependent; increased concentrations of Bd0314 increased the fluorescence readings. All data points were measured in at least triplicates. The error bars represent the standard deviation.

3.2.3.2 Catalytic residues

The FITC-PG assay was used to confirm whether E143 and E154 (general acid-base pair identified from the Bd0314 structure) were required for catalysis. The glutamate residues were mutated to glutamines thereby changing the chemistry of the residue but maintaining the structure (Bd0314-E143Q and Bd0314-E154Q). Figure 25 shows that the activity of both Bd0314-E143Q and Bd0314-E154Q are greatly reduced compared to Bd0314. This finding confirms both E143 and E154 are required for catalysis.

Y228 was mutated to an alanine residue (Bd0314-Y228A) to investigate its role in substrate-binding and specificity. Bd0314-Y228A showed a reduced activity towards the GlcN substrate but no increase in activity towards the GlcNAc substrate. This

finding suggests that Y228 also participates in substrate recognition by direct interaction with the free amine group of GlcN residues (Figure 30). The retained substrate specificity of Bd0314-Y228A suggests other features also contribute to substrate selection. M150 was identified as another substrate selective residue, although we were unable to successfully express and purify a Bd0314-M150A mutant.

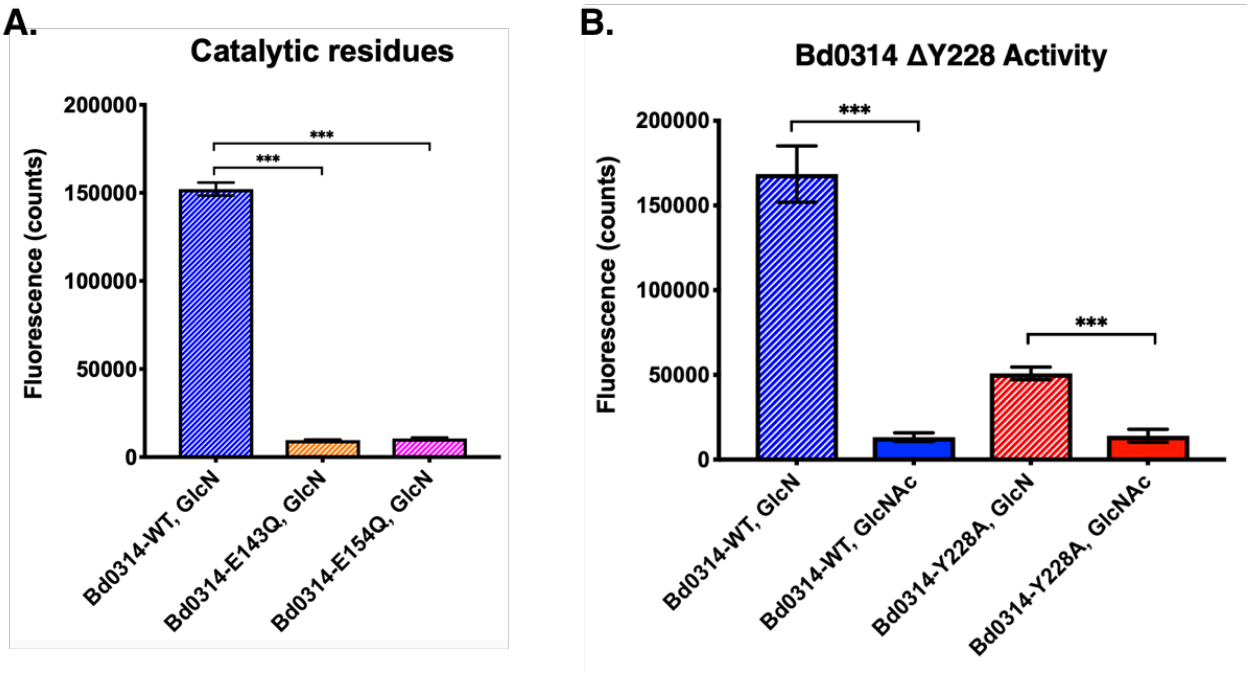


Figure 25. Bd0314 active site residues
A. Mutation of either E143 or E154 abolishes Bd0314 activity, suggesting both residues are involved in catalysis and supporting an inversion catalytic mechanism, involving two glutamate residues. **B.** Y228A mutation reduces Bd0314 activity towards *N*-deacetylated PG. This finding suggests Y228 is involved in substrate binding. Y228A activity towards GlcNAc PG does not alter. All data points were measured in at least triplicates. The error bars represent the standard deviation. Unpaired t-test was used to determine significance, *** = $p<0.0001$, ns = not significant.

3.2.3.3 Bd0314 homologue activities

Bd1440 (S75 – K266) and Bd1413 (G80 – K271) constructs were created to investigate the activities of Bd0314 *Bdellovibrio* homologues using the FITC-PG assay (Figure 26). Bd1440 and Bd1413 both possess the same high specificity towards the GlcNAc-deacetylated substrate (GlcN). Bd0314, Bd1440 and Bd1413 all liberated a similar quantity of FITC, demonstrating similar activities.

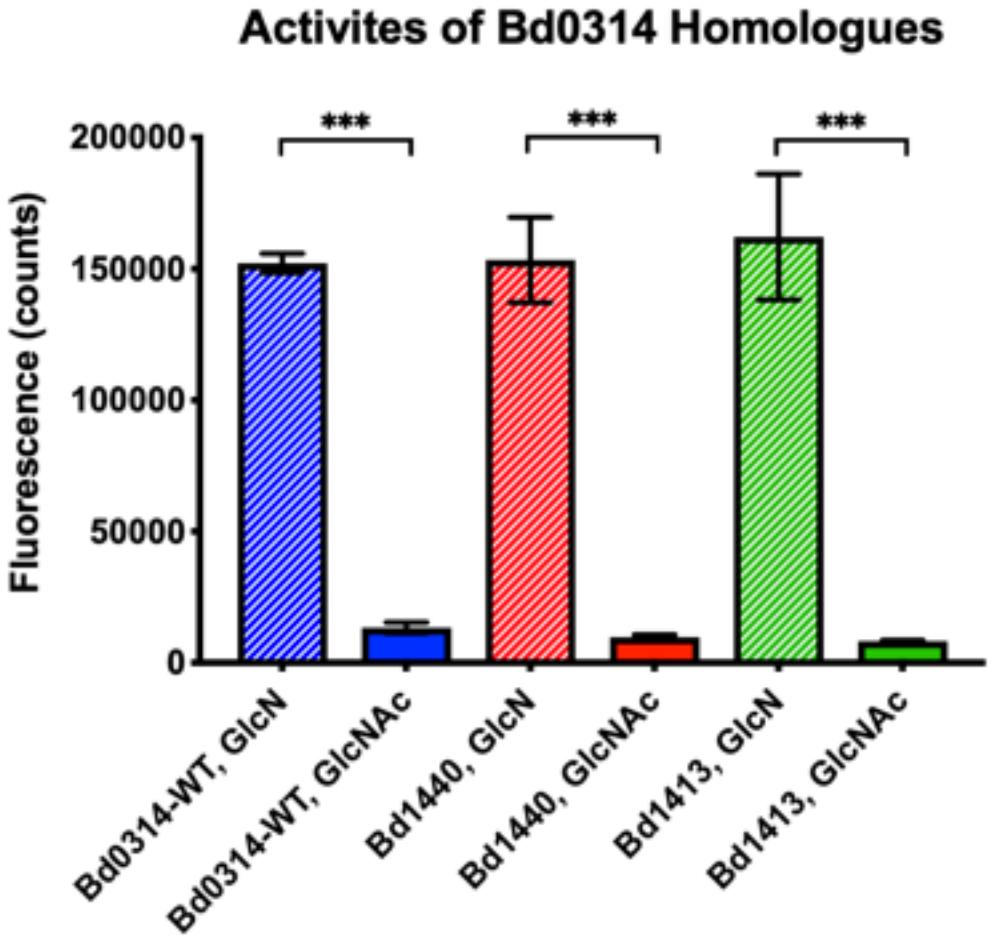


Figure 26. Bd0314 HD100 homologue activities
The activity of Bd0314 HD100 homologues (Bd1413/1440) compared to Bd0314 activity. Bd0314-WT in blue, Bd1440 in red and Bd1413 in green, with block coloured bars representing acetylated GlcNAc substrate and striped coloured bars representing *N*-deacetylated substrate (GlcN). The activities of Bd1440 and Bd1413 also demonstrate substrate specificity for GlcN substrate and have minimal activity with GlcNAc substrate. Bd1413/1440 have a similar activity as Bd0314. All data points were measured in at least triplicates. The error bars represent the standard deviation. Unpaired t-test was used to determine significance, *** = $p<0.0001$, ns = not significant.

3.2.3.4 Bd0314 premature lysis assay

Bd0314 is proposed to lyse the osmotically stable and lysozyme-resistant bdelloplast structure at the end of the predatory lifecycle. We demonstrated Bd0314 possesses high specificity *in-vitro* towards a GlcNAc-deacetylated substrate. However, we also wanted to investigate its activity towards the true *in-vivo* substrate - bdelloplasts. To investigate whether Bd0314 activity could rupture bdelloplasts and bring about premature lysis we developed a bdelloplast lysis assay (see Section 2.2.3). Briefly, bdelloplasts were incubated with EDTA to disrupt the outer membrane, allowing enzyme accessibility to the bdelloplast cell wall.

EDTA treatment alone did not affect the viability of the bdelloplasts. EDTA treatment followed by incubation with Bd0314 resulted in lysis of bdelloplasts and premature release of *Bdellovibrio* progeny (Figure 27). Bdelloplasts incubated with Bd0314 appear ruptured with free swimming progeny visible. At the same time point, EDTA-only treated bdelloplasts remained rounded and intact with progeny visible within, suggesting Bd0314 activity was responsible for release. Furthermore, prey cells were labelled with HADA for the purpose of visualising cell wall material. The level of HADA label appeared reduced in Bd0314 incubated samples compared to the control. Reduced levels of HADA label suggest bdelloplast peptidoglycan had been degraded by Bd0314 activity. pMAL::mNeon was used to label the cytoplasm and periplasm of prey cells, which gave a good indication of the bdelloplast location. The bright pMAL::mNeon label was used to confirm the identity of the misshapen artefacts (lysed bdelloplasts). These results demonstrate that Bd0314 is able to act upon the modified prey cell wall of the bdelloplast and is therefore capable of lysing bdelloplasts at the end of the predatory cycle.

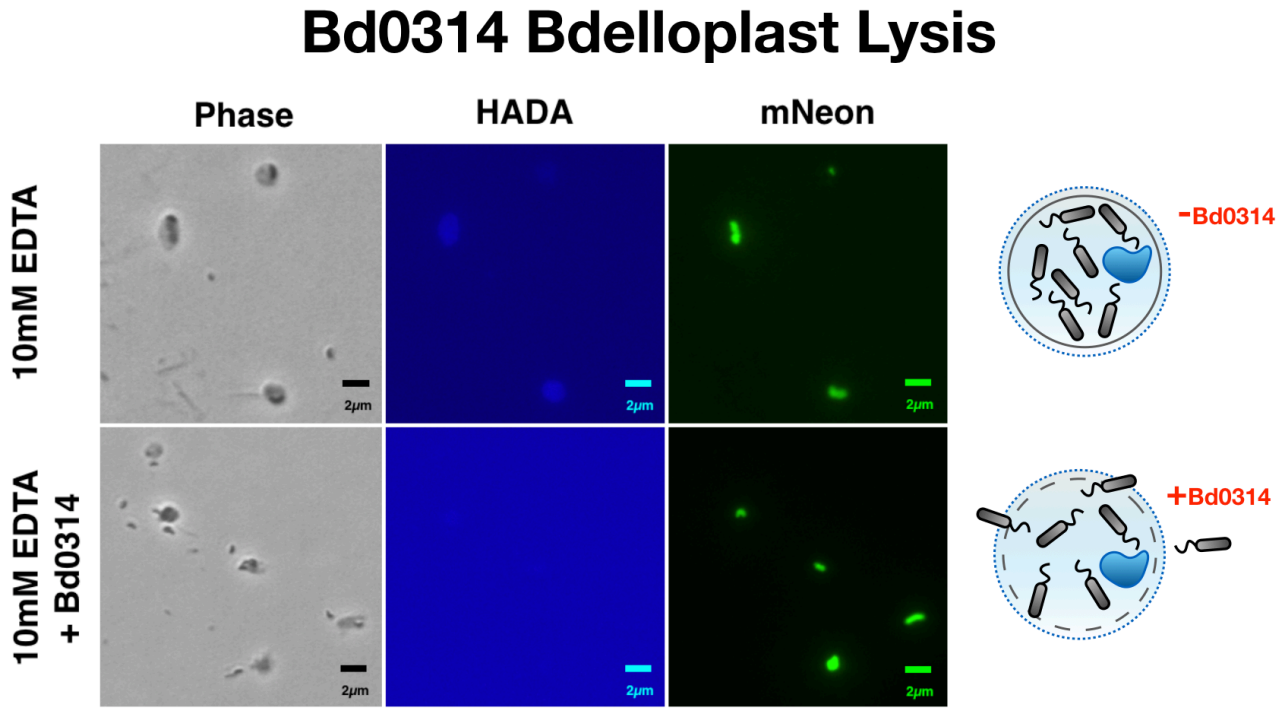


Figure 27. Bd0314 Bdelloplast Lysis
E. coli cells were incubated with *Bdellovibrio bacteriovorus* HD100 cells resulting in >99% prey rounding into bdelloplasts. Bdelloplasts were incubated for 1 hour with 10 mM EDTA +/- 10 μM Bd0314. Cells were immobilised and imaged on 1% agarose Ca/HEPES pads. HADA was used to label the cell wall. pMAL::mNeon was used to label the periplasm and cytoplasm of *E. coli.* Bdelloplasts incubated with Bd0314 lyse and release *Bdellovibrio* progeny when compared to EDTA alone control. Furthermore bdelloplasts incubated with Bd0314 have diminished levels of HADA labeled peptidoglycan, demonstrating that Bd0314 is active towards this material. pMAL::mNeon labeling allows identification of lysed bdelloplasts. The far right hand cartoon illustrates the effect Bd0314 incubation has on the bdelloplast structures.

3.3 Discussion

3.3.1 Comparison of structure

The structure of Bd0314 resembles a lysozyme-like fold, retaining the key structural elements conserved in the lysozyme super-family [141]: the two α-helices (helix-F and I) that make up the central core of the α-helical domain, with the conserved strict β-hairpin at the base of the cleft [144] (Figure 28).

Besides the core similarities, the structure of Bd0314 varies significantly from other lysozyme-superfamily folds, distinguishing it as a novel variant fold. Bd0314 contains a short β-sheet projection (strand-3 and 4), which adds a peak to the α-helical domain. The structural significance of the short β-sheet region to the function of Bd0314 is unknown. Moreover the short β-sheet is not retained in the Bd0314 *Bdellovibrio* homologues (Bd1413 & Bd1440) that also display specificity; hence it is unlikely to contribute significantly to substrate specificity. Additionally the α-helical domain of Bd0314 has a region of short α-helices (residues 26-56 helix-C, D and E), that are not structurally conserved in other superfamily members. This region adds to the bulk of α-helical domain and contains a number of charged residues on the surface. Another divergent region is the N-terminal region (residues 1-14, helix-A) and small section of the C-terminal region (177-181), which again bulk out the α-helical domain. Also the extended β-sheet region (residues 82-90 strand-1 and 2) of Bd0314 is not structurally related to other lysozymes. In other lysozymes the β-sheet extends for ~five amino acids, whereas in Bd0314 the β-sheet has a kink in strand-1 at the 5th amino acid and then continues for a further three residues. The two-part strand-1 makes a β-sheet with strand-2 that runs continuously. The kink in strand-1 is required to project E154 up

into the active site, where it acts as the second catalytic residue. The structurally unique features of the Bd0314 fold are highlighted in red in Figure 28.B.

The Bd0314 fold is distinct and contains features of multiple lysozyme superfamily members. The primary sequence conservation between Bd0314 and other lysozymes is low. However, differing classes of lysozymes typically share little primary sequence conservation [141]. Results from DALI suggest Bd0314 is more closely related to LTs. This is mainly because LTs and G-type lysozymes (like Bd0314) share bulkier α-helical domains [153, 155] (length = 5525 atoms, rmsd = 0.88Å). Bd0314 and LTs have a similar α-helix extension located between the β-hairpin and helix-I. This region in other lysozymes such as C-type, folds back towards the β-sheet extending the base of the cleft. In Bd0314 the extended β-sheet provides a similar structural feature instead but from the other direction.

GH19, GH22, GH23, GH24, GH46 are all believed to have evolved from a common ancestor protein [141]. All have unique structural elements yet maintain the invariant lysozyme core. Their unique structural elements dictate their substrate specificity, such as GH-22 for PG, GH19 for chitin and GH46 for chitosan. Bd0314 retains the invariant lysozyme core, has unique structural elements and is specific towards GlcNAc deacetylated PG suggesting Bd0314 represents a new GH class that has also evolved from the common ancestor of lysozymes.

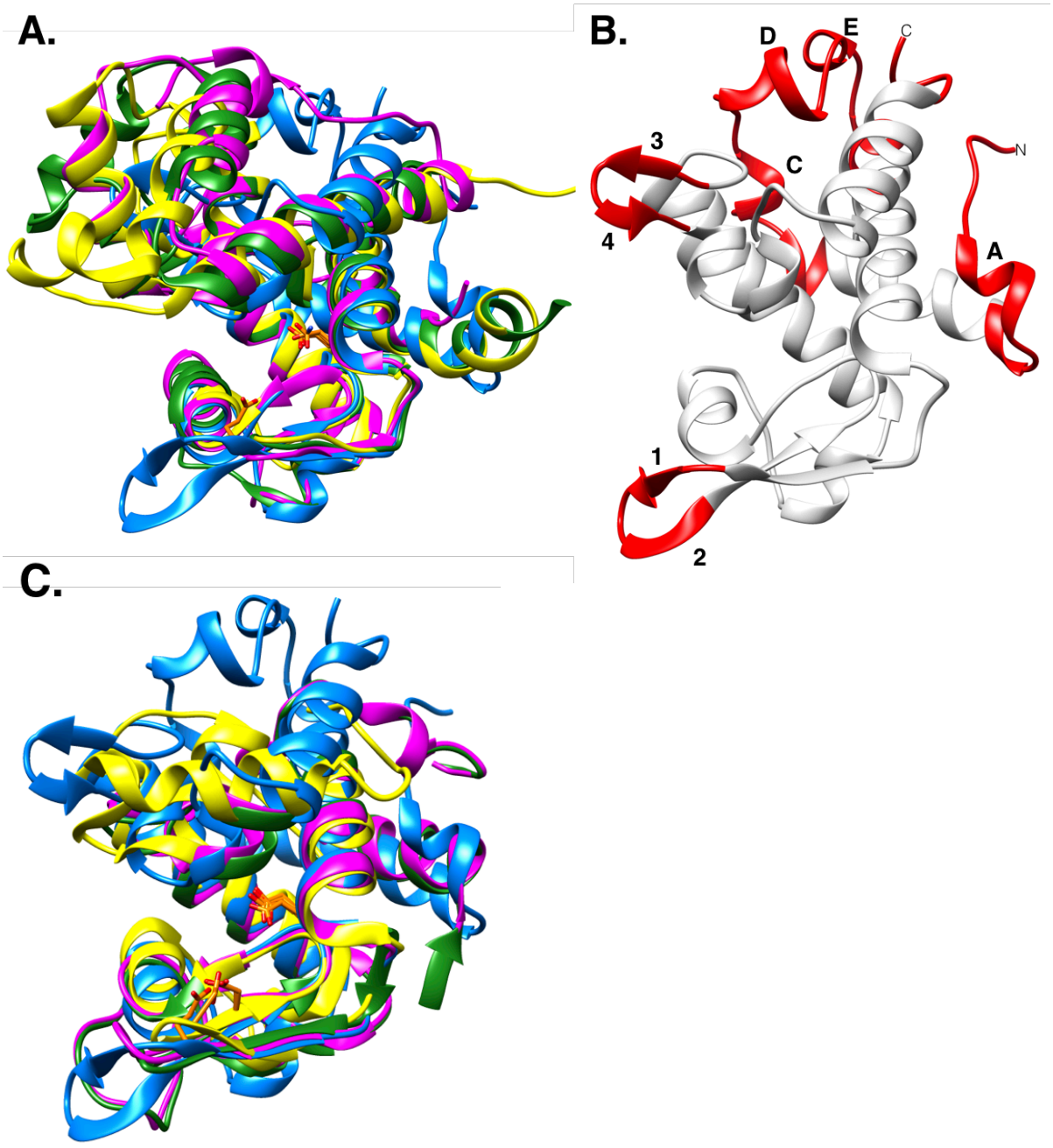


Figure 28. Bd0314 structure comparison
A. Bd0314 (blue) superimposed with the catalytic domains of lytic transglycosylases; MltC = PDB:4CFP (green, length = 5525 atoms, rmsd = 0.88Å); MltF = PDB:4P0G (yellow, length = 4162 atoms, rmsd = 0.88Å); Slt70 = PDB:1SLY (magenta, length = 5000 atoms, rmsd = 0.83Å). **B.** Unique structural features of Bd0314 are shown in red. **C.** Bd0314 (blue) superimposed with GH22 lysozymes; GH22-C HEWL = PDB:1LZS (green, length = 2474 atoms, rmsd = 1.13Å); GH22-I Tapes japonica lysozyme = PDB:2DQA (yellow, length = 2511 atoms, rmsd = 0.59Å); GH22-C Human lysozyme = PDB:5LSH (magenta, length = 1414 atoms, rmsd = 1.03Å). Superimposition was performed by least squares minimisation using the CCP4 suit superpose software (alignment of conserved secondary structure elements, helix-F, helix-I and beta-hairpin).

3.3.2 Comparison of substrate-binding and specificity

Bd0314 is specific towards a GlcNAc-deacetylated PG and displays minimal activity towards 'normal' acetylated PG. The fold of Bd0314 has been adapted to facilitate specificity. The substrate-binding cleft and the residues lining it have been modified to select against acetylated PG. Residues in the cleft are positioned to directly clash with the position of the acetyl-group of GlcNAc residues in the Bd0314 substrate bound model (superimposition of lysozyme:ligand complexes). The sidechain of Y228, located in the loop above the active site, is directed into the substrate-binding cleft. Furthermore, the loop containing Y228 is increased by a single amino acid insertion (relative to other lysozymes, R227), which pushes Y228 further into the binding cleft. The Y228 hydroxyl-group makes a hydrogen bond with S176 and a hydrophobic interaction with W229 and I203, which lock it into this conformation. In the substrate-binding models Y228 consistently clashes with the acetyl-group of GlcNAc residues (Figure 29) [153, 162, 163]. The acetyl-group of GlcNAc residues always faces into the binding cleft, whereas the acetyl-groups on MurNAc residues face outwards so that the peptide stem can be accommodated. Consequently only deacetylation of GlcNAc residues appears to be relevant to binding. A Bd0314-Y228A mutation reduced the activity of Bd0314 towards GlcN PG but did not increase the activity towards GlcNAc PG. This finding suggests that Y228 is required to bind the amine group of GlcN residues and that specificity is multifactorial.

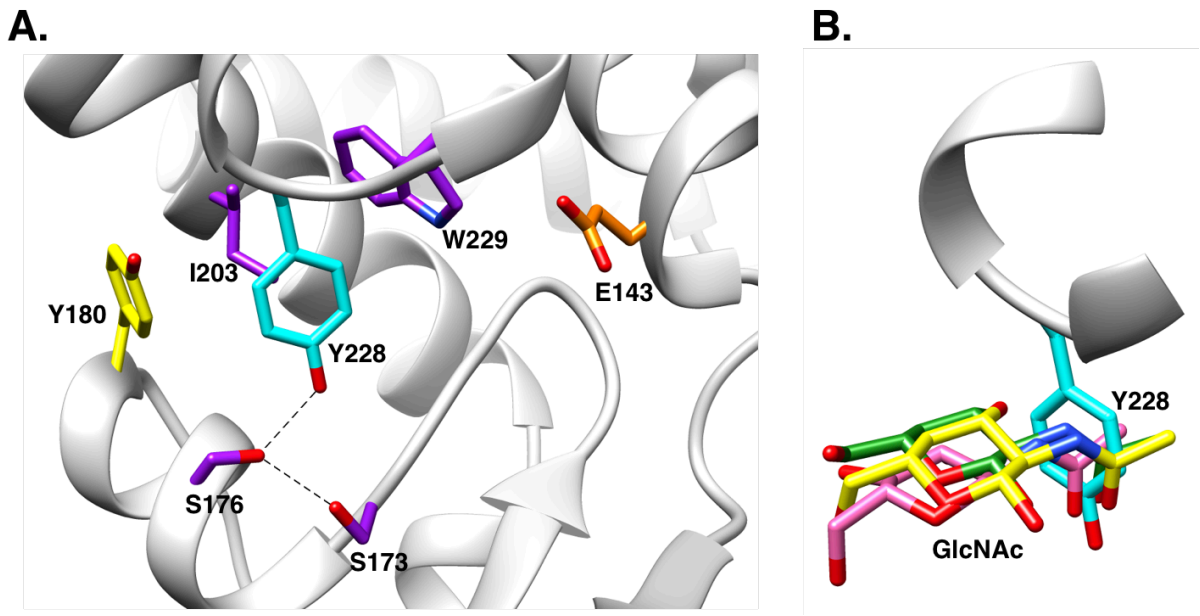


Figure 29. Bd0314 substrate binding and selection
A. Y228 (cyan) is positioned facing into the active site and held in position by a hydrogen bond with S176 (black line) and by hydrophobic interactions with W229 and I203. Y180 (yellow) may participate in substrate binding by interacting with the MurNAc hydroxyl-group. **B.** Bd0314 was superimposed with 1LZS (yellow), 3W6D (pink) and 3GXR (green) – lysozyme:ligand complex structures. The acetyl group of GlcNAc residues in these superimposed models always clashed with Y228 of Bd0314.

M150 also occupies the same space as the acetyl-group of GlcNAc residues and could also participate in selection. However, M150 is not conserved in predicted deacetylation-specific lysozymes whereas Y228 is (Figure 21). Also M150 is not locked into one conformation and in the structure of Bd0314 was observed in multiple conformations suggesting that it could easily be repositioned during substrate binding. The ligand binding models also highlight other residues that may be involved in substrate binding. For instance Y180 is positioned in close proximately to the hydroxyl group of the MurNAc residues and may make a hydrogen bond. Similarly S176 that interacts with Y228 may share the H-bond to the GlcN amine group.

Chitosanases (GH46) are a member of the lysozyme superfamily and hydrolyse the β-1-4 linkage between poly-GlcN [145]. Chitosanases are structurally divergent from

lysozymes and have a much more larger β-sheet domain, bulked with additional α-helices but retain the invariant core lysozyme fold and the catalytic glutamate in the ES-motif [141]. Chitosanases differentially select and act on the deacetylated chitin polymer (chitosan). Interestingly these enzymes select their substrates in a similar mechanism to our proposal of steric occlusion, where the *N*-acetyl groups would face steric crowding at the binding site [145]. Moreover, the polar environment of the binding site, provided in part by tyrosine residues, provides interactions to the amine group of GlcN residues [164, 165]. However, in chitosanases the free amine (protonated at neutral pH) is typically recognised by charged residues, such as aspartates. There is no evidence that Bd0314 selects deacetylated substrates via charge recognition. Superimposition of chitosanase OU01 (40LT) and chitosanase (1QGI) from *Bacillus circulans* provides a model for how the GlcN ligand would interact with Bd0314. Interestingly Y228 is positioned within hydrogen bonding distance of the GlcN amine group, similarly to how Y37, of chitosanase OU01, interacts with the amine-group of GlcN residues.

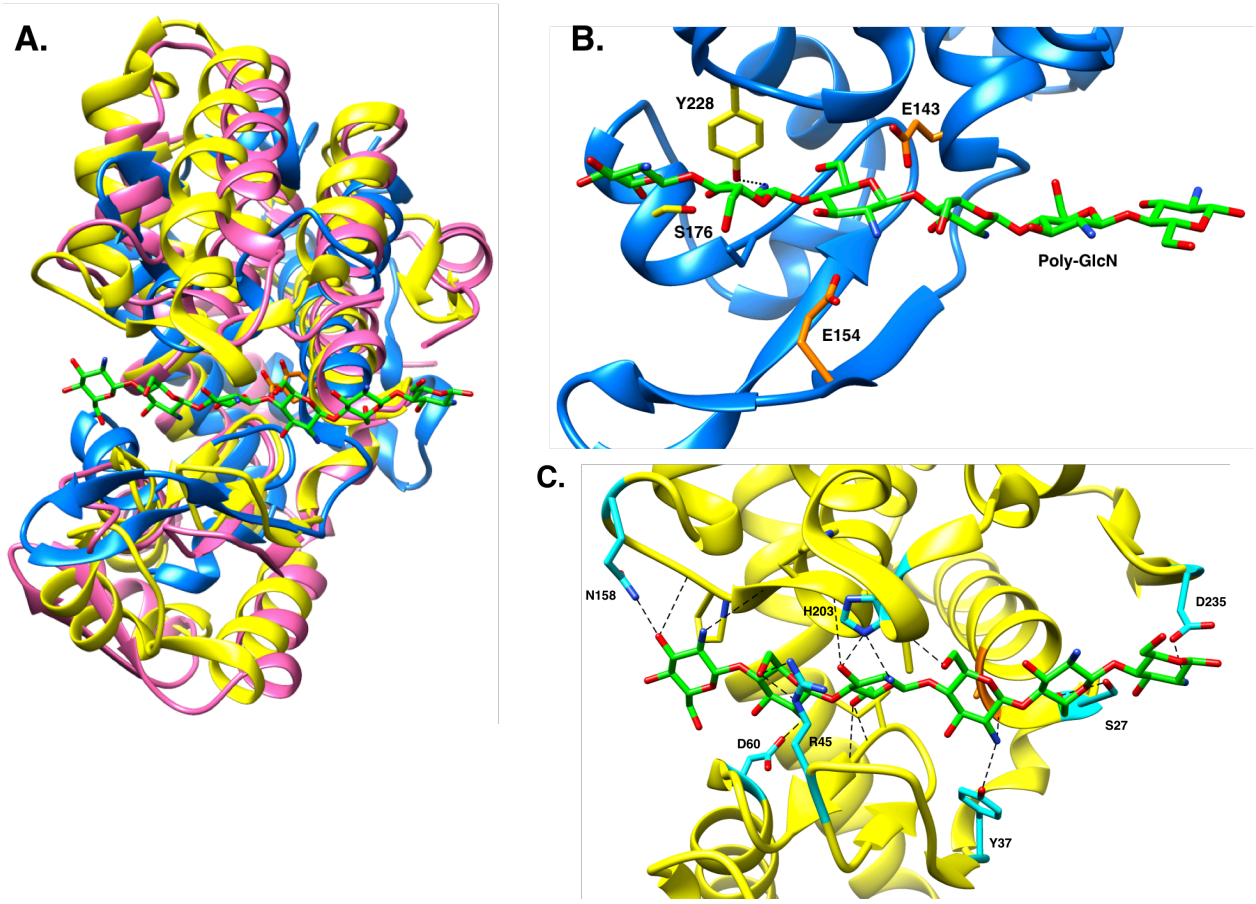


Figure 30. Comparison of Bd0314 with chitosanases.
A. Bd0314 (blue) superimposed with chitosanase-OU01 (PDB:4OLT, yellow, length = 2325 atoms, rmsd = 1.82Å) and *Bacillus circulans* chitosanase (PDB:1QGI, pink, length = 2226 atoms, rmsd = 2.57Å). **B.** Close up of the view of a poly-GlcN substrate-binding model to Bd0314 active site cleft (from superimposition with 4OLT). Y228 could potentially make a hydrogen bond contact with the amine group of GlcN residues. **C.** Substrate binding site of chitosanase-OU01 shows polar residues make hydrogen bonds to free amine group. Charged residues also interact with probable charge on GlcN. Superimposition was performed by least-squares minimisation using the CCP4 suit superpose software (alignment of conserved secondary structure elements, helix-F, helix-I and beta-hairpin).

3.3.3 Likely catalytic mechanism

The structure of Bd0314 revealed a potential secondary catalytic residue (E154) located opposite to the strictly conserved catalytic E143 residue. We demonstrated that both E154 and E143 are required for catalysis. A catalytic mechanism involving two glutamate residues is unusual for lysozyme superfamily members, which typically

utilise a Glu-Asp pair. A native Glu-Glu general acid-base pair has only been observed once before, in a chitosanase [166]. There is an average 8Å gap between the carboxyl-groups of E143 and E154, suggesting Bd0314 operates via a typical inverting mechanism [147]. A water molecule is located 2.7Å from the carboxyl-group of E154 and could represent the location of the catalytic water used in the hydrolysis reaction.

The location of the second catalytic residue differs between lysozyme superfamily members, which can either exist on the first or second strand. In Bd0314 the second catalytic residue (E154) is located immediately before the kink in strand-1. In comparison D30 of GH22-I (2DQA) is also located on the first strand [167], whereas in HEWL the second residue D53 is located on strand-2 (Figure 31).

Besides the acid-base pair, Bd0314 has a conserved serine residue (S166) that occupies the position of D53 (HEWL). Furthermore S166 makes a hydrogen bond with E154 and may have a catalytic role, presenting the catalytic water molecule. In GH19 family a conserved serine/threonine residue at an equivalent position on the β-hairpin has been associated with playing an essential role during catalysis [168] and H-bonded to the presumed catalytic water molecule [144, 169, 170].

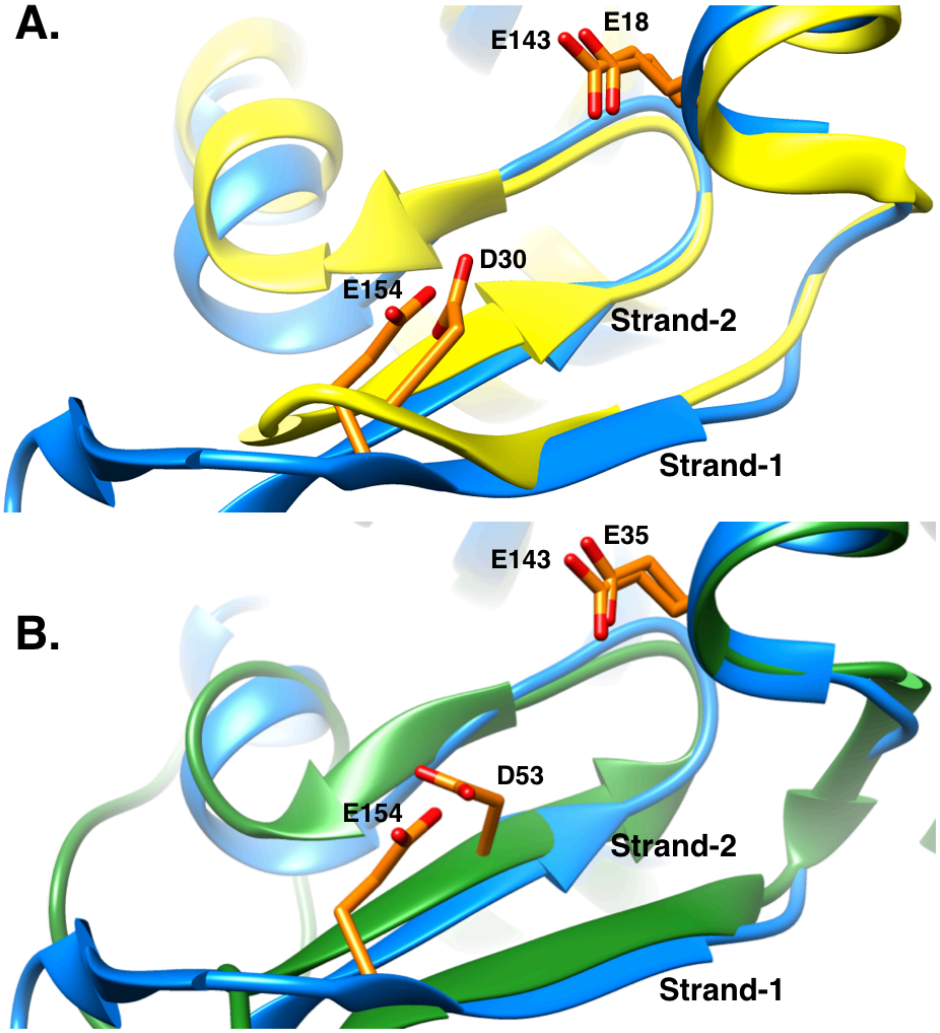


Figure 31. Bd0314 catalytic residue location.
The location of the second catalytic residue varies between lysozymes and generally dictates the catalytic mechanism, which is dependent on distance. **A.** E143 is located on the ES motif at the end of the conserved central α-helix in Bd0314 (blue). Bd0314 unusually has a second glutamate catalytic residue (E154), which is located opposite on strand-1 similar to D30 of *Tapes japonica* Lysozyme (PDB:2DQA, yellow). **B.** In HEWL (PDB:1LZS, green) the second catalytic residue, D53 is located on strand-2.

3.3.4 Bd0314 role as the exit enzyme

Early studies suggested bdelloplast lysis at the end of the predatory lifecycle was brought about by an increase in lytic enzyme activity towards the end of the predatory cycle. The lytic activity was shown to be distinct from the initial pore forming activity, which was halted after entry [45]. The enzyme(s) responsible for bdelloplast lysis was hypothesised to work specifically on the modified (lysozyme resistant) prey cell wall

[45, 50, 52]. We have shown that Bd0314 shows specificity towards GlcNAc-deacetylated PG in comparison to normal 'acetylated' PG. Bd0314 shows minimal activity towards a normal PG substrate suggesting that it is a poor substrate, which was confirmed by the structure of Bd0314 (an adapted lysozyme fold). Furthermore Bd0314 expression was shown to be upregulated immediately prior to bdelloplast lysis (unpublished work – Sockett Lab). Therefore Bd0314 possesses the desired attributes originally described to the "Exit enzyme": be able to act on the modified prey cell wall; be expressed immediately before prey cell lysis; and distinguish between "prey" and "self".

The specificity of Bd0314 activity allows the lytic activity to be directed solely towards the modified prey cell wall material and not the putatively acetylated *Bdellovibrio* cell wall. Hence, the deacetylation modification of prey cell walls not only stabilises the bdelloplast structure but also marks the prey cell wall material for destruction [49]. Furthermore Bd0314 activity agrees well with the previously reported *Δbd0468Δbd3279* mutant phenotype that left 'ghosts' (undigested prey cell wall material) and had slower exit rates [49]. Bd0314 would not be able to act as the exit enzyme if the prey cell walls had not been deacetylated, explaining why acetylated prey cell wall material was not digested upon exit and explaining why progeny cells became trapped. Furthermore, a *Δbd0314 mutant Bdellovibrio* (unpublished work – Sockett Lab) left behind similar undigested prey cell 'ghosts' and a significant proportion of *Δbd0314* mutant *Bdellovibrio* progeny remained trapped in the bdelloplasts unable to complete the predatory cycle. Together the phenotypes of these two mutants suggest Bd0314 facilitates escape, although in both cases the progeny ultimately managed to exit. This suggests that *Bdellovibrio* possesses a backup escape strategy, if something

goes wrong. *Bdellovibrio* reseals and strengthens the entry pore after invasion [46] and it has been suggested that the entry pore was used to escape in *Δbd0468Δbd3279* mutant *Bdellovibrio.* The resealed pore may be made out of a different material that a backup enzyme can solubilise.

The bdelloplast lysis assay demonstrated the ability of Bd0314 to lyse lysozyme resistant bdelloplasts, which caused the premature release of *Bdellovibrio* progeny. The first reported instance of release of premature *Bdellovibrio* progeny was performed by Ruby et al. in 1983. Interestingly these experiments used concentrated solutions of old culture medium from preceding infection experiments. The old culture medium was applied to bdelloplasts, causing them to rupture and release the progeny cells prematurely [171]. It is likely Bd0314 or other deacetylation specific lysozymes provided the lytic activity of the old culture supernatant.

The primary sequence of Bd0314 has a signal peptide and a lipobox suggesting it is secreted and attached to the outer membrane of *Bdellovibrio* (Figure 15). Between the lipobox sequence and the beginning of the catalytic domain is a region of predicted disorder. The 47aa region of disorder is mainly composed of proline, serine and threonine residues (>60%). As a result this region may act as a flexible chain tethering Bd0314 to the outer-membrane and localising the activity. Localised employment of Bd0314 would be most favourable post bdelloplast lysis and prevent promiscuous activity, which may initiate premature lysis of neighbouring bdelloplasts. Some LTs are similarly anchored to the outer membrane, which localises their activity. For example MltF is a two domain, membrane-bound protein that has a catalytic domain under the control of a regulatory domain that is able to sense muropeptides [172]. The tethering

of MltF to the outer membrane does not impede its activity but ensures localised control of its activity.

On the other hand the flexible tether of Bd0314 may be cleaved by one of the many proteases *Bdellovibrio* secretes into the prey cell during predation [24]. The disordered region would be an easily accessible substrate to a protease. If cutting the tether and freeing Bd0314 is required for bdelloplast lysis then the need for cleavage would add in an additional control step before committing to exiting the bdelloplast.

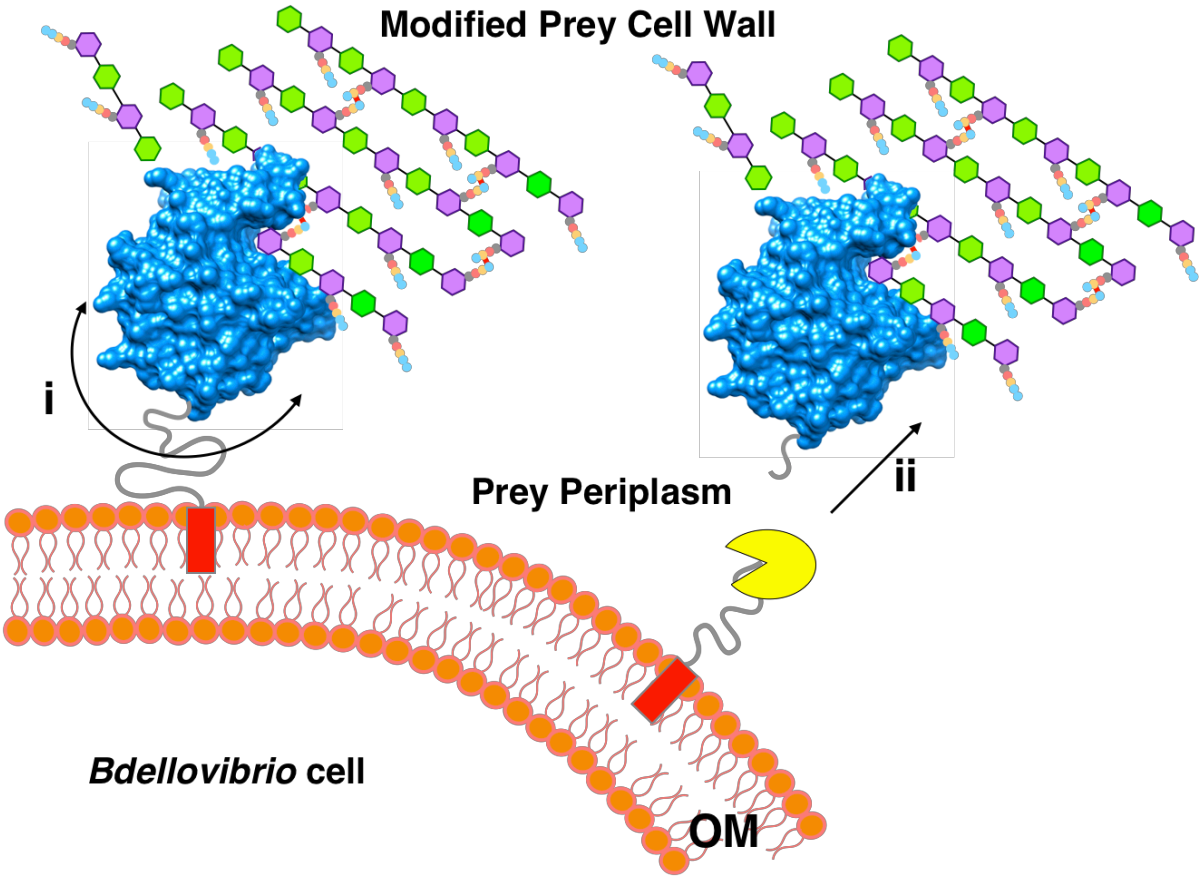


Figure 32. Role of Bd0314
Bd0314 activity is able to solubilise the modified prey cell at the end of the predatory cycle. Full-length Bd0314 possesses a signal peptide with a lipobox suggesting Bd0314 is secreted to the outer-membrane. Between the lipobox and the catalytic domain is a 47aa region of disorder. The disordered region may act as a tether, retaining Bd0314 to the outer membrane for localised prey cell wall degradation (i). Alternatively the tether may be a target of protease action and cleavage of the protease accessible disordered region would release Bd0314 into the bdelloplast (ii). The lipobox is represented by the red box. The disordered region of Bd0314 is illustrated by the grey curved line. Bd0314 is shown in blue and the prey cell wall is illustrated as in Figure 7 (green - GlcNAc/GlcN residues and purple – MurNAc). Potential protease cleavage is shown by the yellow shape.

3.3.5 Roles of other Bd0314 homologues

There are two Bd0314 homologues in the *Bdellovibrio* HD100 genome (Bd1413 and Bd1440). The roles of Bd1440 and Bd1413 are unknown but both possess the same high specificity towards a GlcNAc-deacetylated PG. The expression profiles of Bd1440 and Bd1413 are different to Bd0314. They are both expressed at low levels throughout the lifecycle and not upregulated prior to bdelloplast lysis (unpublished work – Sockett Lab). This suggests that Bd1440 and Bd1413 have a different role to Bd0314. Moreover Bd1413/1440 are more closely related to each other (60% ID) compared to Bd0314 (29-33% ID). Bd1413/1440 lack the short β-sheet of Bd0314, which may also contribute to their different roles, although the function of the short β-sheet in Bd0314 is unknown. Additionally Bd1413/1440 have an insertion between (helix G and H), which is typically associated with LT folds. Bd1413/1440 may therefore have a more general housekeeping role, similar to LT. If Bd1413/1440 are required to cleave glycan strands and create room for growth during the predatory lifecycle, their activities would need to be tightly controlled in order to not lyse the bdelloplast prematurely. On the other hand Bd1413/1440 may serve as redundant back up proteins, should the activity of Bd0314 fail.

3.3.6 Presence in other bacterial systems

Alignment of Bd0314 with Bd1413/1440 identified a number of conserved regions including the second glutamate (E154) and the acetyl-group blocking YW-motif (Y228). As these two regions were specific to Bd0314 and Bd1413/1440 and not featured in other “standard” lysozymes we presumed these two features represent a deacetylation-specific lysozyme motif. We searched the non-redundant NCBI database and identified

a number of other potential deacetylation-specific lysozymes in other organisms separate from predatory bacteria (Figure 22). Search results demonstrated a 100% correlation between possessing a glutamate at the E154 position and having a tyrosine at the Y228 position (one was never present without the other). Other organisms possessing potential deacetylation-specific lysozymes also have a CE4 GlcNAc-deacetylase in their genomes that would generate such a substrate. For example deacetylation specific lysozymes were identified in *Caulobacter* species. *Caulobacter species N*-deacetylate their PG to modify charge and increase the cohesive network required to attach *Caulobacter* to surfaces and facilitate its biphasic lifecycle [173, 174]. It is interesting to speculate that Bd0314 represents a new class of lysozymes that act solely upon GlcNAc-deacetylated PG substrates. It is yet to be determined if any of the predicted deacetylation-specific lysozymes from other organisms possess the same substrate specificity.

3.4 Conclusion

The intraperiplasmic predator *Bdellovibrio bacteriovorus* uses cell wall deacetylation as a chemical identifier to mark "self" versus "prey" during the predatory lifecycle. Such a modification requires novel enzymology to rupture the deacetylated prey cell wall material during prey cell exit. This work confirms that this function is provided by repurposed lysozymes (Bd0314 and homologues). The structure of Bd0314 reveals evolution has shaped the ancient lysozyme fold [141] to remodel it into a variant that now only accepts deacetylated cell wall sugars. Enzyme assays and addition to bdelloplasts confirm the specificity for modified cell wall material and provide a new insight into the means to which predators exit their prey.

Chapter 4

Structure and Function of Bd2924, a Putative Acyl-CoA Dehydrogenase

4.1 Introduction

4.1.1 Overview

Our initial interest in Bd2924 arose after it was identified as a putative CdG binder in Rotem et al. Bd2924 does not possess any known CdG binding motifs within its primary sequence. Therefore uncovering how Bd2924 potentially interacts with CdG and any implications on activity are of interest. Bd2924 is a putative acyl-CoA dehydrogenase (ACAD); an important class of metabolic enzymes that catalyse the initial step in fatty acid β-oxidation. Bd2924 may therefore have an important role in central metabolism of *Bdellovibrio*.

4.1.2 Metabolism - β-oxidation

ACADs catalyse the initial dehydrogenation step of β-oxidation, a major metabolic pathway metabolising fatty acids. β-oxidation is a catabolic process consisting of four steps, which operate between the $2^{nd}(C_\alpha)$ and $3^{rd}(C_\beta)$ carbons of acyl-CoA esters. β-oxidation results in shorting of the acyl chain of acyl-CoA esters, production of acetyl-CoA and transfer of electrons to the electron-transport chain [175].

Figure 33 illustrates the four steps involved in β-oxidation of fatty acids. Firstly, acyl-CoA dehydrogenases catalyse the dehydrogenation of fatty acyl-CoA to produce trans-Δ^2-enoyl-CoA using FAD as a cofactor. Trans-Δ^2-enoyl-CoA is then converted into L-β-hydroxyacyl-CoA in a hydration reaction performed by a enoyl-CoA hydratase. Following this, the hydroxyl group of L-β-hydroxyacyl-CoA is oxidised to a keto group by a 3-hydroxyacyl-CoA dehydrogenase. Lastly, a thiolysis reaction, performed by β-ketothiolase, inserts the thiol group of CoA between the C_α and C_β to cleave the bond. The thiolysis reaction produces a molecule of acetyl-CoA and shortens the fatty acyl-CoA chain by two carbon units. Cycles of β-oxidation can be repeated until the acyl-chain length is reduced to the final two carbons. Alternatively, if a fatty acyl chain with an odd number of carbons was used as substrate, propionyl-CoA (three carbons) will be the ultimate product of the reaction.

Unsaturated fatty acids can also be utilised as an energy source in β-oxidation if they are first converted to a saturated fatty acid. A 2,4-dienoyl-CoA reductase converts poly-unsaturated fatty acid into trans-Δ^2-enoyl-CoA [176].

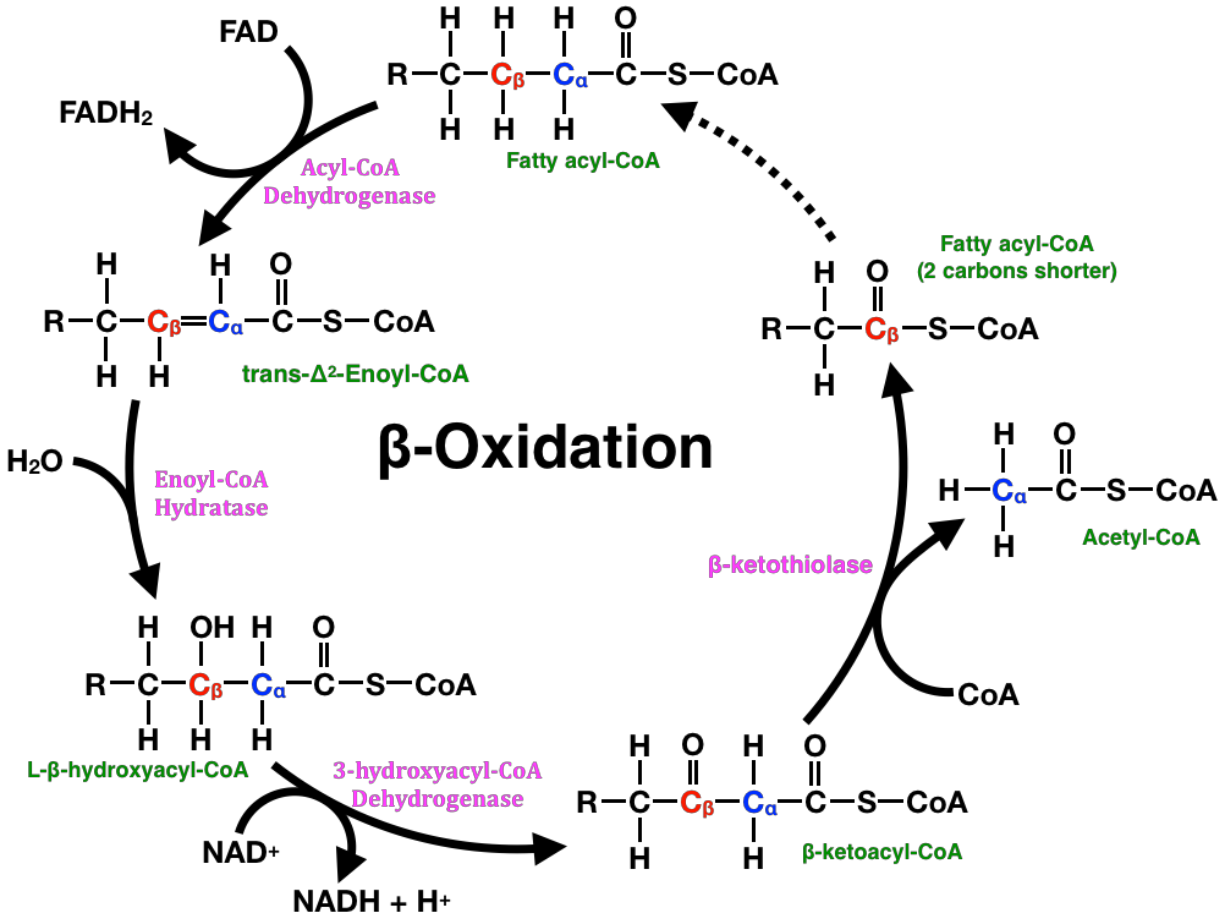


Figure 33. β-oxidation metabolic pathway
β-oxidation is a catabolic process, involving the **c**hemical modification of fatty acyl-CoA substrates at the C_α and C_β positions. Acyl-CoA dehydrogenases catalyse the initial dehydrogenation reaction of fatty acyl-CoA to form trans-Δ^2-enoyl-CoA. An enoyl-CoA hydratase converts trans-Δ^2-enoyl-CoA into L-β-hydroxyacyl-CoA. Then a 3-hydroxyacyl-CoA dehydrogenase produces β-ketoacyl-CoA with the use of an NAD cofactor. A thiolase catalyses the final thiolysis step by inserting the thiol group of CoA between the C_α and C_β positions. The β-oxidation cycle can be repeated until the acyl-chain length is reduced to the final two carbons.

4.1.3 β-oxidation regulation in bacteria

Fatty acids are an essential component of bacterial membranes and a source of metabolic energy. Regulation of fatty acid biosynthesis and degradation is tightly controlled to ensure membrane lipid homeostasis is maintained. The majority of research into β-oxidation pathway regulation has been carried out in *E. coli* [177].

E. coli use fatty acids of various chain lengths as a an energy and carbon source. The enzymes responsible for fatty acid degradation are encoded in the *fad* regulon [177]. FadL and FadD are responsible for transporting the fatty acids across the outer and inner membrane, respectively. FadD is also a CoA synthase that attaches CoA to the fatty acid chain, creating a fatty acyl-CoA ester, the precursor for β-oxidation. FadE performs the initial dehydrogenation reaction while the hydration, oxidation and thiolysis reactions are carried out by a tetrameric complex of FadB and FadA [178]. Transcriptional control of β-oxidation genes is under the regulation of FadR. FadR repression is released when fatty acyl-CoA esters of >14 carbons in length are present [179].

Bdellovibrio bacteriovorus incorporates a large amount of unaltered prey fatty acids when growing intraperiplasmically [59]. Conservation of prey lipids means that *Bdellovibrio* is less likely to require de-novo synthesis of fatty acids from acetate, which is an energy consuming process. Furthermore, *lipid* metabolism in *Bdellovibrio* is dependent upon the prey consumed [62]. *Bdellovibrio* alters, degrades and synthesises fatty acids derived from prey cells. Approximately 30-40% of prey cell lipids are degraded by the usual β-oxidation mechanism [52]. As such, lipids and amino acid derived from prey constitute a significant proportion of *Bdellovibrios* energy source [68].

4.1.4 Acyl-CoA dehydrogenases

ACADs are a large family of enzymes that are involved in the metabolism of fatty acids and branched chain amino acids. ACADs are categorized into distinct groups based on

their substrate specificity. ACADs involved in β-oxidation of fatty acids can be further split into: short-chain acyl-CoA dehydrogenase (SCAD), medium-chain acyl-CoA dehydrogenase (MCAD), long-chain acyl-CoA dehydrogenase (LCAD) and very long-chain acyl-CoA dehydrogenase (VLCAD) enzymes. ACADs involved in amino acid oxidation are: isovaleryl-CoA dehydrogenase (IVD), isobutyryl-CoA dehydrogenase (IBD), short/branched-chain acyl-CoA dehydrogenase (SBCAD) and glutaryl-CoA dehydrogenase (GD). ACADs share high sequence identities (35-45%) evenly distributed over the entire peptide span with the exception of VLCAD [180]. Besides the similarities in primary sequence ACADs also share a common fold.

4.1.5 Structures of ACADs

ACADs exist primarily as soluble homotetramers of subunits of approximately 43kDa, each associated with an FAD molecule [180]. VLCAD is an exception and exists as a homodimer with subunits of approximately 73kDa, which are anchored to the inner membrane via an additional C-terminal extension [181]. A number of ACAD family member structures have been published: Human VLCAD (PDB: 3B96), Pig MCAD (3MDE), Human SCAD (2VIG), Human IVD (1IVH), *Desulfococcus multivorans* GD (3MPI) and Human SBCAD (2JIF) [181-184], which are displayed in Figure 34.

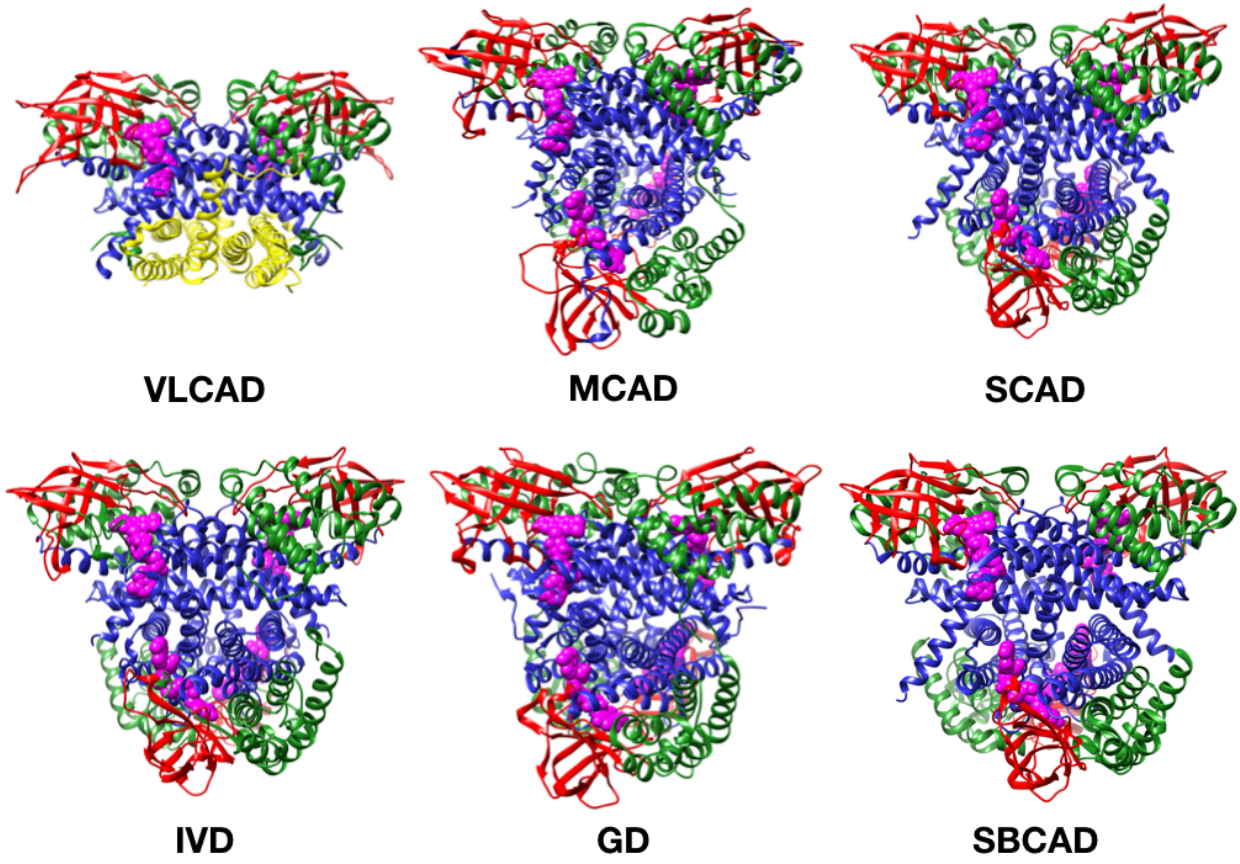


Figure 34. Acyl-CoA dehydrogenases
This Figure illustrates the three-dimensional structures of different members of the ACAD family. The structures are coloured based on their domains: N-terminal α-helical domain (green), central β-sheet domain (red), C-terminal α-helical domain (blue) and finally C-terminal extension domain (yellow) the FAD molecules are coloured magenta. ACADs form tetramers with the exception of VLCAD that forms a dimer. VLCAD – PDB:3B96, MCAD – PDB:3MDE, SCAD – PDB:2VIG, IVD – PDB:1IVH, GD – PDB:3MPI, SBCAD – PDB:2JIF.

ACAD subunits fold into three distinct domains: a N-terminal domain (green) and a C-terminal domain (blue) that are mainly comprised of α-helices and a central domain (red) that is formed of two orthogonal β-sheets (Figure 35.A). The C-terminal domain is extensively involved in forming the tetramer and dimer interfaces. The tetrameric structure is composed of a dimer of dimers in a tetrahedral arrangement. The dimer interface is extensive and includes shared FAD binding sites, whereas the tetramer interface mainly involves helix-helix interactions [180]. VLCAD is an exceptional case, which forms an identical dimer interface to MCAD. However, an additional C-terminal

domain (+180 residue) prevents the formation of the conventional tetramer interface. The additional C-terminal domain adopts a fold identical to the N and C-terminal α-helical domains (tetramer interface contacts) of the opposing MCAD dimer, hence the additional C-terminal domain mimics the tetramer interface contacts.

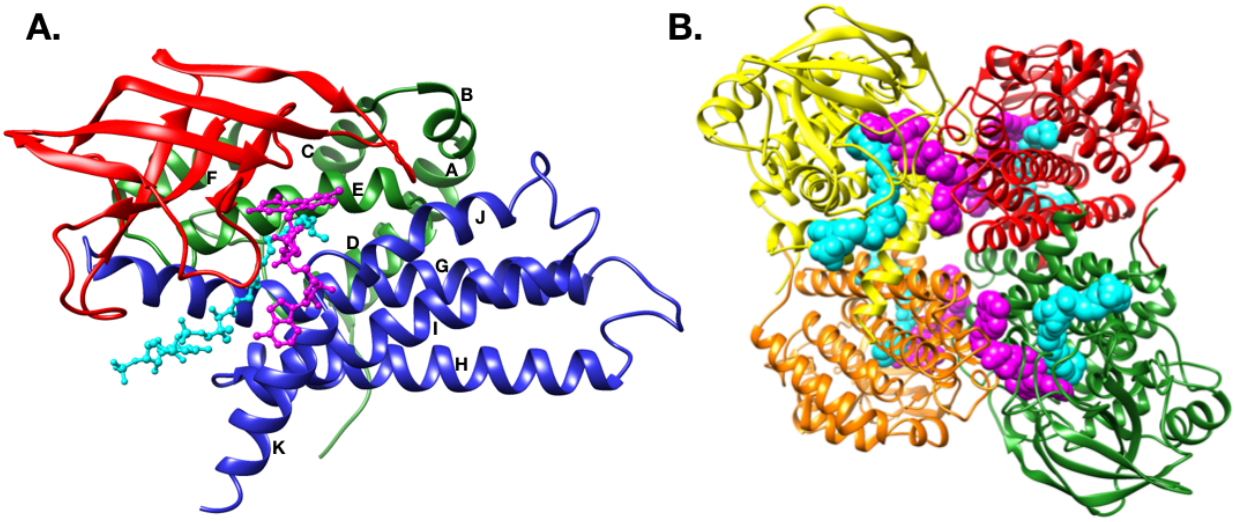


Figure 35. MCAD structure
MCAD is the most well-characterised ACAD, structurally and biochemically. **A.** MCAD can be split into three domains: an N-terminal α-helical domain (green), central β-sheet domain (red) and C-terminal α-helical domain (blue) that provides the tetramer interface. Each monomer binds an FAD molecule (magenta) and can bind a single C8-CoA ligand(cyan). Helices are labelled according to Pig MCAD, for reference to the text. **B.** MCAD forms a tetramer complex in the form of a dimer of dimers. Red + Yellow and Green + Orange monomers make a single dimer. The FAD molecules bind between dimers.

4.1.6 Substrate binding

The FAD molecule is located at the dimer interface between two monomers. The isoalloxazine ring is located in the crevice between the two α-helical domains and the β-sheet domain of one monomer, whilst the adenosine group is located at the dimer interface (Figure 35).

The fatty acyl-CoA substrate binds in the tunnel between the three domains of a single monomer. The adenosine group binds at the surface in a groove leading to the opening

of the tunnel and makes hydrogen bonds and van der waals contacts. The fatty acyl portion of the acyl-CoA inserts into the tunnel between helices G and F and the loop between helices J and K. The tunnel is just wide enough to accept the fatty acyl group, which displaces coordinated water molecules from the tunnel upon binding [180]. The precise docking of the thioester is achieved by two tight H-bonds formed between the FAD-2'OH and E376-NH with the thioester carbonyl in Pig MCAD [185]. No major conformational changes occur upon substrate binding; the most pronounced changes occur for Y375 and E99 that adopt different conformations to accommodate the substrate.

Substrate specificity for varying CoA thioesters is achieved by changes in the enzyme active site architecture and substrate-binding pocket. Different lengths of acyl-CoA thioesters are accommodated depending on the depth of the binding pocket, typically controlled by the presence of bulkier amino acids that restrict access. In SCAD (PDB: 1JQI) an isoleucine residue restricts access to the active site, thus selecting shorter acyl-CoA substrates [186]. Alternatively, in MCADs (PDB: 3MDE) the substrate-binding cavity is widened by the presence of a proline residue, changing the trajectory of helix G, which allows acyl-CoA chain lengths of between 4-16 carbons to be accommodated [187]. Bulkier residues control the depth of the cavity and in VLCAD (PDB: 3B96) these residues are replaced by glycines, which extends the tunnel and allows acyl-CoA thioesters up to 24 carbons in length to be accommodated [181]. The tunnel depth of VLCAD is calculated to be 24Å, twice that of MCADs 12Å tunnel. Branched chain substrates are accommodated by the widening of the substrate cavity. In IVD (PDB: 1IVH) a glycine residue is substituted in for the conserved tyrosine residues found in other ACADs, which prevents steric repulsion of the branched side chain. The catalytic

glutamate residue of IVD is also found 115 residues away from the location of MCAD catalytic residues, which positions it for catalysis of the branched chain substrates [183].

4.1.7 Catalytic mechanism

The catalytic mechanism of ACADs is illustrated in Figure 36. The reaction generates a trans double bond between the C_α and C_β of acyl-CoA substrates utilising an FAD molecule as a hydride acceptor [185]. A conserved catalytic glutamate residue is located opposite the re-face of FAD and acts to initiate proton abstraction from the α carbon, which results in a simultaneous hydride transfer from the β carbon to the N5 of the FAD cofactor. Structural and mutational studies confirm E376 is the conserved catalytic base in pig MCAD [187, 188]. The location of the catalytic base varies between ACADs. E376 is located on the loop between helices J and K in MCAD, whereas in IVD and LCAD, E254 and E261 respectively, are located on the adjacent helix G [183]. The catalytic residues are located in different positions in the primary sequence, although topologically they occupy the same space.

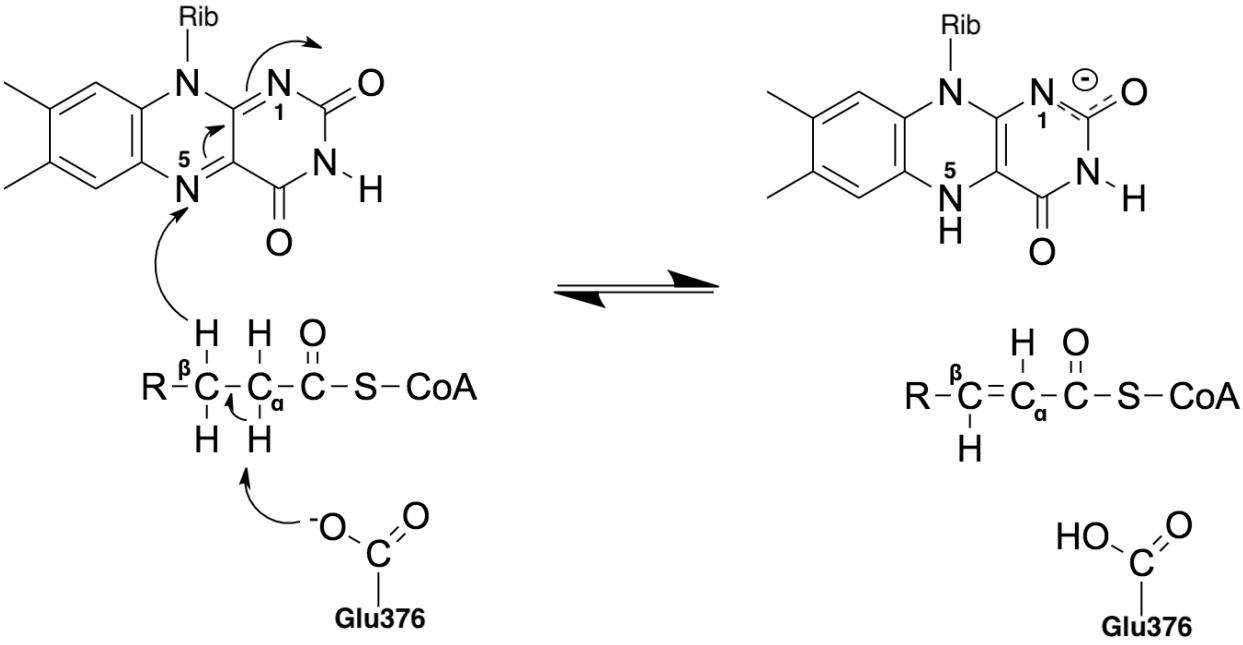


Figure 36. ACAD mechanism
This figure illustrates the catalytic mechanism of ACADs, exemplified by pig MCAD. The catalytic E376 abstracts a proton from the C_α, which synchronously causes the hydride transfer from C_β to be transferred to the N5 position of FAD as a hydride. $FADH_2$ is deoxidised to FAD in two $1e^-$ transfers to a flavin carrying ETF.

4.1.8 ETF and its interactions with ACADs

The reduced FAD molecule is re-oxidised through two sequential re-oxidation reactions with an electron transfer flavoprotein (ETF). ETFs are involved in accepting electrons from all acyl-CoA dehydrogenases shuttling them to the membrane bound ETF-ubiquinone oxidoreductase of the respiratory chain [185, 189] (Figure 37).

ETFs are unusually heterodimers, composed of α and β subunits with a single FAD molecule and a redox inactive AMP molecule [190]. The interflavin redox reaction between the FAD molecules of ACAD and ETF occurs through two successive one electron transfer steps. The FAD molecules are buried in both enzymes thus a direct hydride exchange would not be feasible. It is believed that ACADs and ETF interact weakly with each other and ETF undergoes a large dynamic conformational change in

order to interact with ACAD. A single ETF α-subunit:MCAD complex structure has been published revealing the ETF α-subunit interacts with the N-terminal α-helical domain of MCAD [191]. The interaction mainly involves hydrophobic interactions. The location of the FAD molecule in the β subunit of ETF in the resting state implies a large conformational change is required to correctly position the FAD molecule in the electron transferring state [191]. Conformational change is supported by SAXS results that confirm ETFs sample a range of conformations in solution [192].

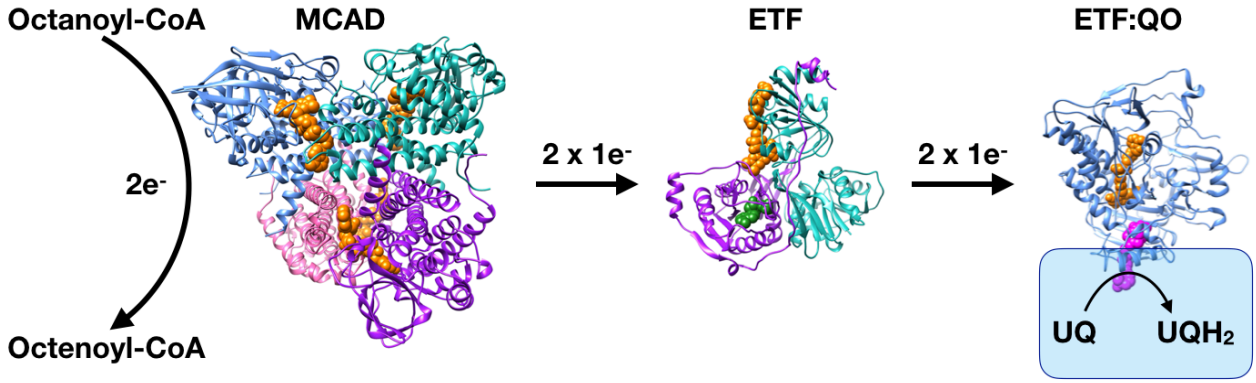


Figure 37. Electron-transfer by ETF
Octanoyl-CoA is dehydrogenated into Octenoyl-CoA transferring two electrons to the FAD of MCAD in the first step of β-Oxidation. $FADH_2$ is re-oxidised to FAD by two successive $1e^-$ transfers to an ETF. ETF then transfers the single e^- to ETF:QO which in turn transfers the electrons to ubiquitin pool. ETF can accept electron transfer from various ACAD family members. Structures: MCAD – 1T9G, ETF – 1EFP, ETF-QO - 2GMH. FAD molecules are coloured in orange, AMP molecule in green and Ubiquinone molecule in magenta.

4.1.9 ACAD homologues including acyl-CoA Oxidases (ACOs)

Homologues of the ACAD fold exist that possess divergent functions. Most divergent homologues are acyl-CoA oxidases of peroxisomal origin [180, 193], while others are involved in steroid side chain cleavage [194], antibiotic synthesis [195] and response to environmental stresses [196-198].

ACOs are highly similar to ACADs and catalyse the initial dehydrogenation reaction of β-oxidation in peroxisomes. ACOs, however, do not require an ETF to re-oxidise; instead ACO use molecular oxygen, producing hydrogen peroxide. The structures of ACOs share a common ACAD fold, more similar to VLCAD. ACOs form a soluble dimer and have an additional C-terminal α-helical domain (C-I domain, yellow) that mimics the tetramer interface of MCADs. In addition ACOs have an α-helical extension in the C-terminal domain (C-II domain, orange) that overlaps with the central β-sheet domain and reduces the accessibility of FAD to bulk solvent. The overlapping helix is also believed to be involved in blocking ETF docking and potentially changing the redox environment of the FAD molecule [180].

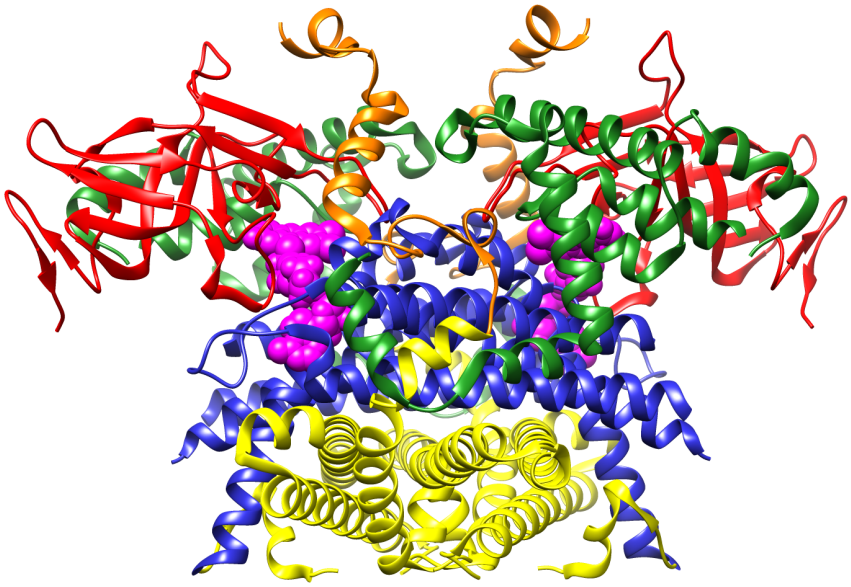


Figure 38. Structure of ACO
The structure of ACO (PDB:2DDH) dimer is coloured according ACAD domains. N-terminal domain (green), central domain (red), C-terminal domain (blue). ACOs contain an addition C-terminal α-helical domain (yellow) that is similar to the C-terminal domain in VLCAD, which mimics the MCAD tetramer interface. Additionally ACOs have an elongated α-helical region (orange) that caps the FAD accessible groove and prevents ETF access to FAD.

4.1.10 ACADs with divergent functions

ACAD homologues have also been attributed functions beyond fatty acid β-oxidation. AidB and ACDH-11 from *E. coli* and *C. elegans* respectively, are two well-characterised proteins that possess both sequence and structural homology to ACAD family members (~20% Seq ID, rmsd = ~2.8 compared to Pig MCAD). Both AidB and ACDH-11 share a similar fold that has diverged from the common ACAD fold, which is more closely comparable to VLCAD. AidB and ACDH-11 possess a C-terminal extension region similar to VLCAD that mimics the MCAD tetramer interface, yet lack the membrane associated region [197, 199]. Moreover AidB and ACDH-11 both possess an additional extended N-terminal β-sheet loop region that fashions a novel tetramer interface. Consequently AidB and ACDH-11 form a soluble tetramer complex unlike other ACADs, (Figure 39).

AidB was first discovered in early studies investigating the effects of exposure to DNA-methylating agents in *E. coli*. In a response referred to as the adaptive response, four proteins (Ada, AlkA, AlkB and AidB) were found to be upregulated to mitigate the mutagenic and cytotoxic effects of DNA-methylation [196]. Convincing DNA repair roles have been assigned to all but one of the four proteins (AidB). AidB was first reported to reduce the mutagenic effects of *N*-methyl-*N*'-nitro-*N*-nitrosoguanidine (MNNG) by utilising the FAD cofactor to detoxify damaging agents [200]. However, this role has been hard to establish with difficulties in determining how the FAD cofactor might perform such a function whilst being buried deep at the dimer interface. Because AidB was shown to bind double-stranded DNA, current reasoning suggests it is required to provide a sheath to protect DNA or to localise detoxification through direct interaction with the DNA [199]. However, the AidB DNA binding region is not required to facilitate

its protective function [201]. The exact role of AidB in the adaptive response remains unclear.

On the other hand ACDH-11 plays a role in adaptation to temperature shifts by altering the levels of lipid desaturation and membrane fluidity. ACDH-11 expression is increased at elevated temperatures. Increasing ACDH-11 reduces the levels of stearic CoA desaturase (FAT-7) by sequestering C11-C12 fatty acids. Reducing the level of C11-C12 fatty acids prevents them activating nuclear hormone receptors that in turn drive FAT-7 expression. Human diseases linked to mutations in ACAD genes cause fatty acid β-oxidation disorders that increase in severity during hyperthermia. This suggests ACAD deficient human patients are more susceptible to heat changes due to an imbalance of desaturated lipids. ACDH-11 therefore plays an important role in signalling heat changes, by sequestering C11-12 fatty acyl-CoA rather than performing a dehydrogenation reaction [197].

4.1.11 Aims and objective

The aim of this study was to characterise Bd2924, a putative acyl-CoA dehydrogenase with supposed CdG binding capabilities. The interaction between CdG and B2924 was probed using structural and biophysical methods.

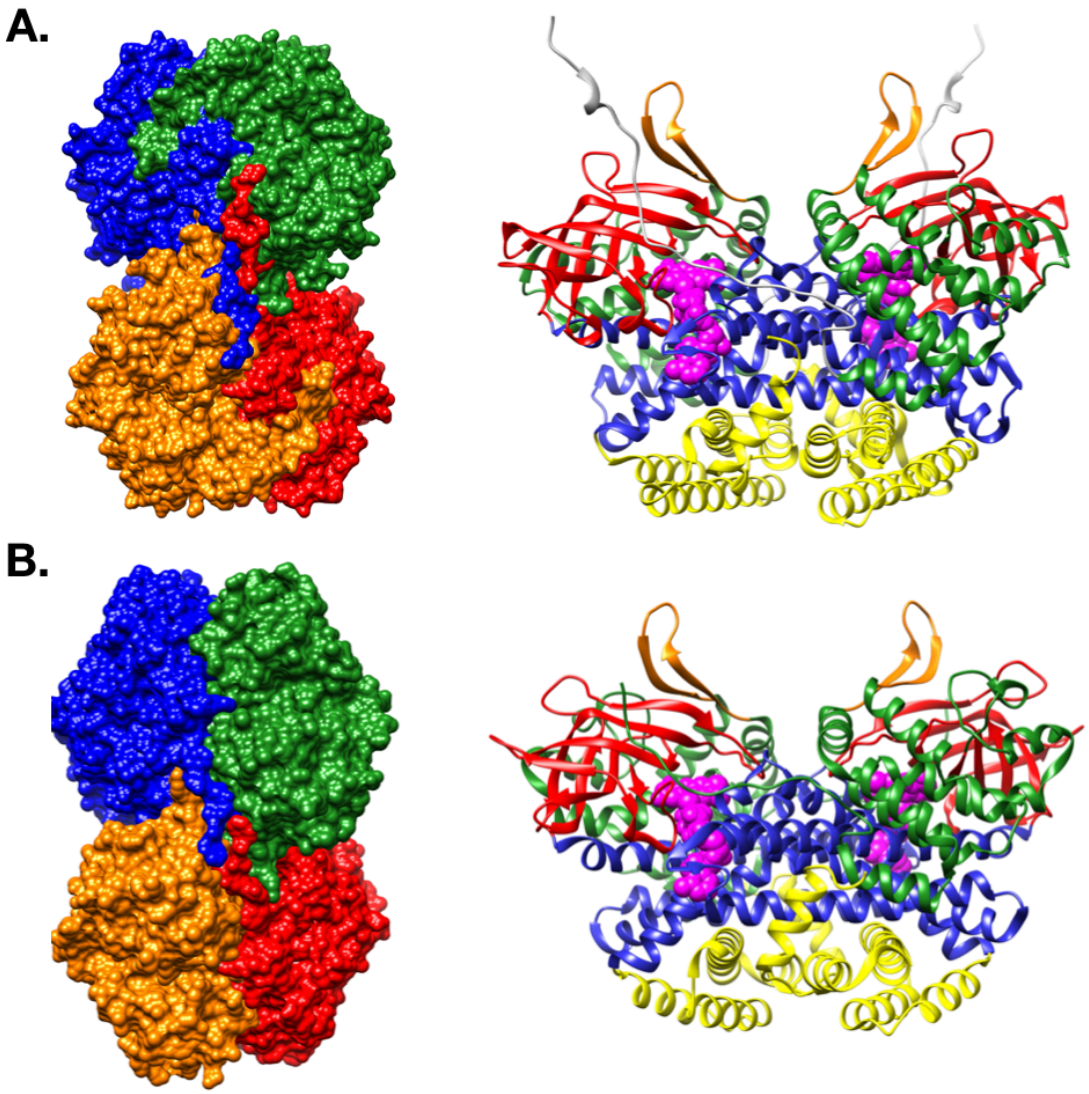


Figure 39. Structure of divergent ACAD homologues
A. ACDH-11 forms a tetramer in the form of a dimer of dimers (blue:green + red:orange), which is organised differently to ACAD tetramers. The tetramer interface is provided by a β-sheet loop in the N-terminal domain (orange). The structure of ACDH-11 domains are coloured similarly to ACAD structures, N-terminal domain (green), central domain (red), C-terminal domain (blue) and C-terminal extension domain (yellow). ACDH-11 is more closely related to the structure of VLCAD. **B.** The structure of AidB is highly similar to ACDH-11. AidB also forms a similar tetramer interface to ACDH-11, facilitated by the N-terminal domain β-sheet loop (orange).

4.2 Results

4.2.1 Cloning, Expression and Purification of Bd2924

Full-length Bd2924 (residues 1 to 505) was cloned into a pET28a vector and expressed in *E. coli* Bl21* (DE3) cells. Bd2924 was purified by IMAC to >95% purity (Figure 40). Purified Bd2924 was bright yellow in solution, indicating the presence of co-purified FAD cofactor.

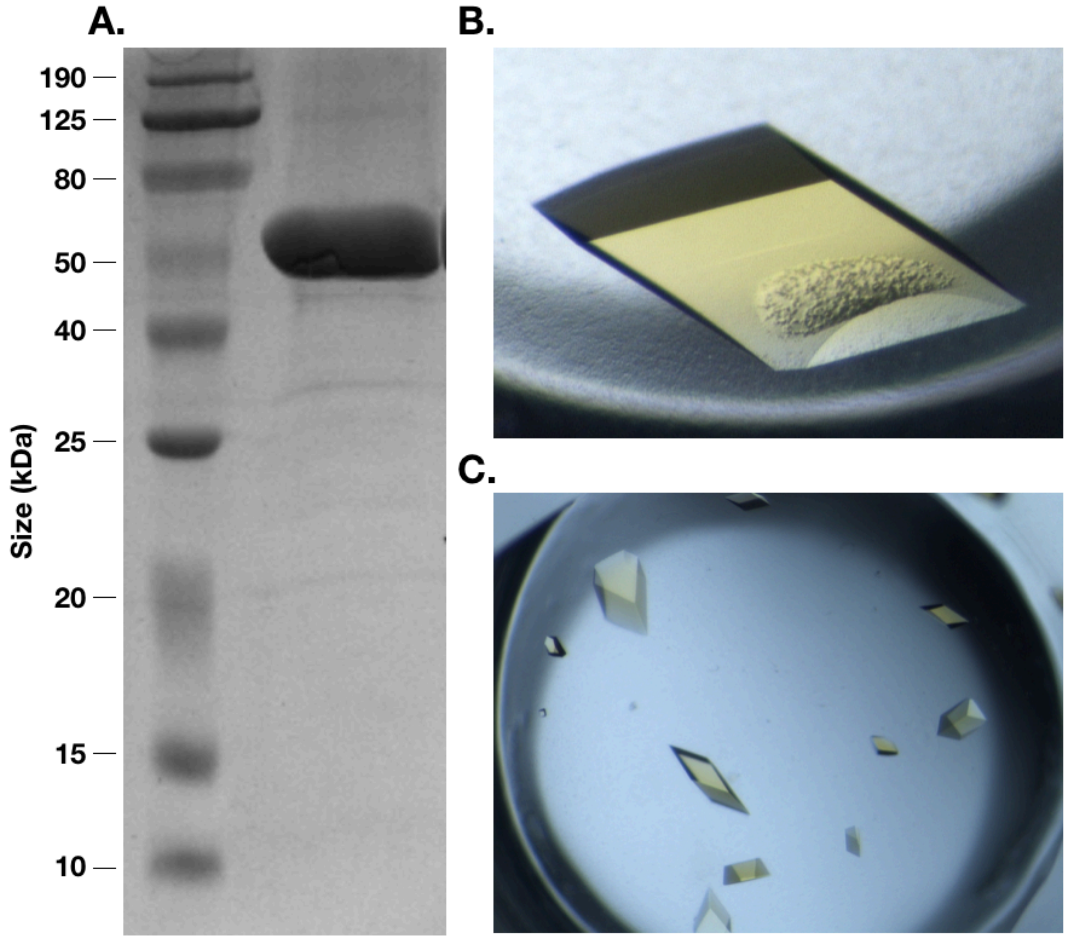


Figure 40. Bd2924 purification and crystallisation
A. Bd2924 was purified to >95% purity, judged by SDS-PAGE. **B.** Image of Bd2924 crystal **C.** Image of Bd2924 crystals

4.2.2 Obtaining crystals and data collection

Large bright yellow Bd2924 crystals were grown, which were used for data collection. The structure of Bd2924 was originally solved by molecular replacement, using PDB:4Y9J as the search model (see Section 2.4.5). We determined the structure of Bd2924 (residues 3-505) in complex with the cofactor FAD, both with and without a C10 length acetyl-CoA thioester ligand (C10-CoA) bound, to 1.51Å and 1.87Å respectively.

4.2.3 Protein Structure / topology

Overall the tertiary structure of Bd2924 is a homotetramer consisting of a dimer of dimers. Bd2924 shares a common ACAD fold, more closely related to divergent ACAD homologues such as AidB and ACDH-11. The homodimer packs across a 2-fold axis to create a likely biologically relevant crystallographic homotetramer, where the CD dimer is identical to the AB dimer. SEC confirms Bd2924 also exists as a tetramer in solution (Figure 41).

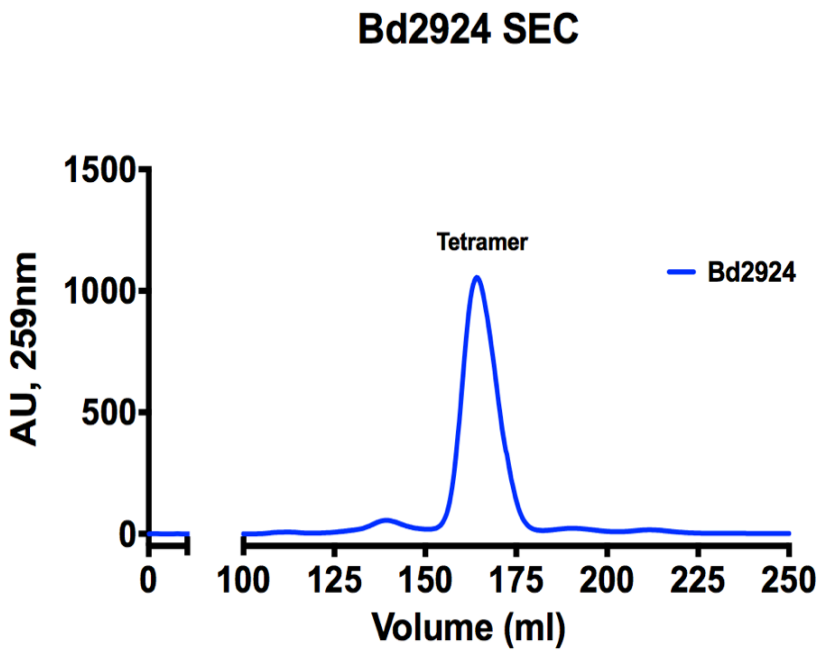


Figure 41. Bd2924 SEC
Bd2924 elutes as a single peak at 165 ml, confirming Bd2924 exists as a tetramer in solution.

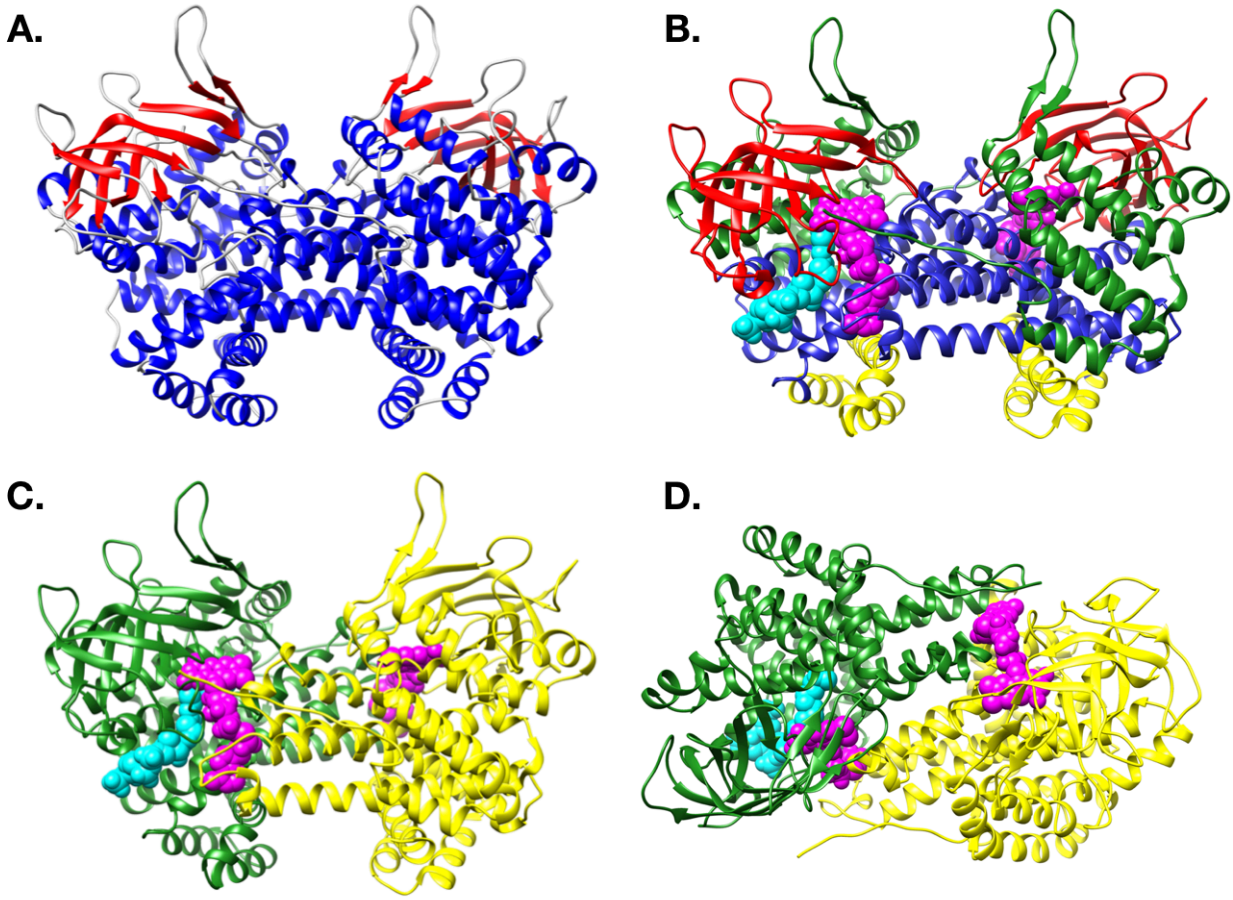


Figure 42. Structure of Bd2924
A. The secondary structure of Bd2924 dimer is mainly composed of α-helices (blue), with a region of β-strands (red). **B.** Bd2924 can be divided into four domains, similarly to other ACADs. Bd2924 has an N-terminal α-helical domain (green), a central β-sheet domain (red), a C-terminal α-helical domain (blue) and finally a C-terminal α-helical extension domain (yellow) similar to VLCAD. **C.** Bd2924 forms a tightly interacting dimer, shown here as a green monomer and yellow monomer. FAD (magenta) binds at the dimer interface. The C10-CoA binding site (Cyan) is found is located in the tunnel running beneath the FAD molecule. **D.** A 90° rotation of Bd2924 showing the dimer interface from another angle.

Bd2924 can be divided into four domains based on other ACAD structures. Bd2924 has an N-terminal α-helical domain that is interrupted by a single β-sheet projection (residues 3-168), a central β-sheet domain formed by two orthogonal four-stranded antiparallel β-sheets (residues 169-285), a central α-helical domain (residues 286-449) and a short C-terminal α-helical domain (residues 450-505). A single FAD molecule binds per monomer in a crevice at the dimer interface. The N-terminal β-sheet projection and the C-terminal α-helical domain create two large grooves in the Bd2914

dimer. The N-terminal β-sheet projections forms a groove that is approximately 30x20x20Å (WxHxD) that acts as the docking site for the CD dimer (tetramer interface), to form the tetramer complex. The C-terminal α-helical domain forms a groove of approximately 20x20x30Å; unique to Bd2924.

4.2.4 Dimer and Tetramer Interface

The Bd2924 dimer is stabilised by a large surface interface (2530Å^2), which provides numerous interactions. In total there are eight salt bridges and 32 hydrogen bonds formed at the dimer interface as well as numerous non-bonded contacts. Salt bridges exist between residues D412 + R256, E399 + K391, E402 + K391 and R485 + E399 from chains A and B. Dimers are further stabilised by the presence of the FAD cofactor that binds between Bd2924 monomers, which further increases dimer surface contacts and the amount of residues participating in dimer stabilisation.

Bd2924 forms a tetramer composed of a dimer of dimers (Figure 43); formed by contacts made between the β-sheet projections from the N-terminal domains. The tetramer interface is smaller than the dimer interface (two contacts 670Å^2 and 390 Å^2). The β-sheet projection (residue region 64 to 80) forms a claw like docking structure. In total six hydrogen bonds and multiple non-bonded interactions stabilise the tetramer interface. The major H-bond provider in this region is the R74 sidechain that interacts with the Q6 backbone carbonyl of the CD dimer while internally stabilising the β-sheet projection by making a H-bond with the D70 sidechain. At the tip of the strict β-hairpin is a tryptophan that becomes buried in a hydrophobic cleft.

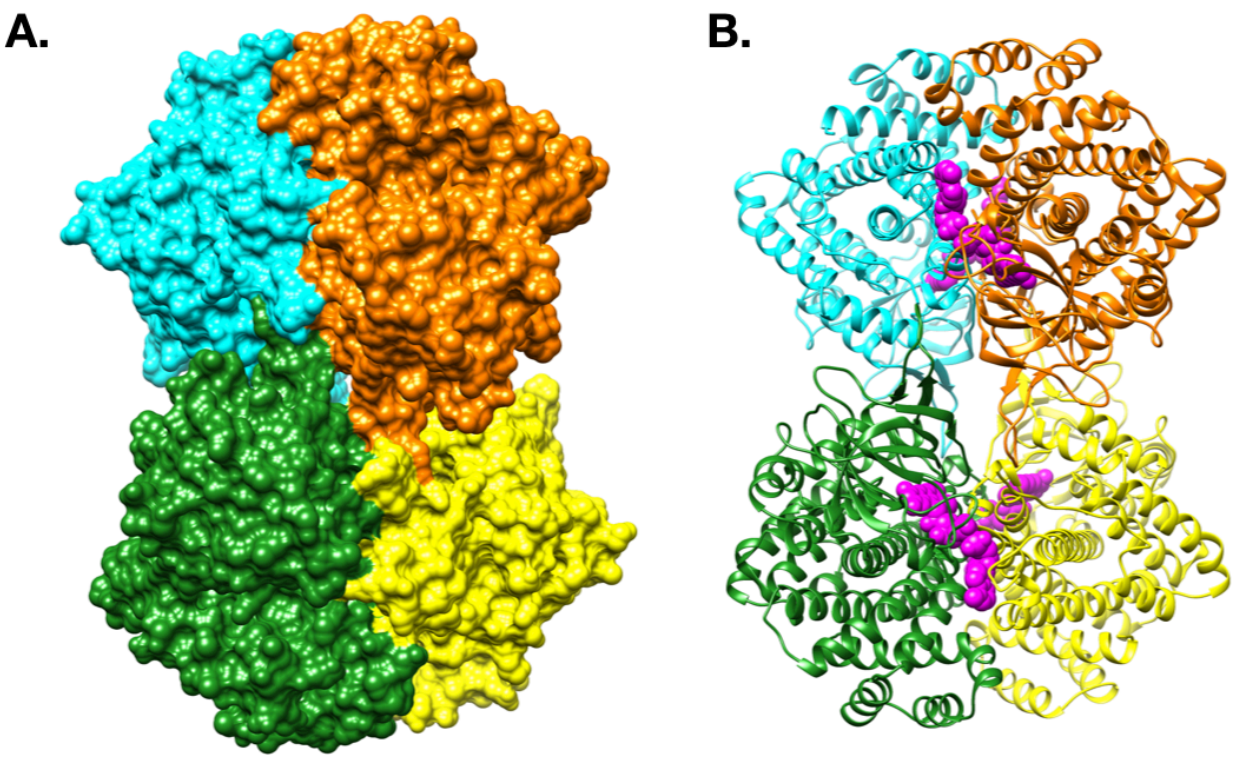


Figure 43. Bd2924 tetramer complex
Bd2924 forms a tetramer complex formed by a dimer of dimers. The tetramer interface is provided by a β-sheet loop located as an insertion into the N-terminal domain. The tetramer complex surface is shown in (**A**) and in ribbon form (**B**), which shows the location of FAD molecules (magenta).

4.2.5 Ligand Binding

4.2.5.1 FAD

FAD binds at the dimer interface and provides structural stability to the protein besides being a required catalytic cofactor. Each monomer binds a single FAD molecule in a crevice created at the dimer interface. The isoalloxazine group of FAD is located in a groove created between the N-terminal α-helical domain, the β-sheet domain and the central α-helical domain of a single monomer with a solvent accessible tunnel running beneath. The ADP region of FAD forms multiple hydrogen bonds with both chains of the AB dimer as well as multiple non-bonded interactions. Figure 44 below depicts the hydrogen-bonding network associated with FAD binding to Bd2924. The FMN region

makes H-bonds with Q171, M173 and T174 of Chain A and G406 of Chain B. The di-phosphate group makes the most intra-dimer contacts including: S180 and G179 of chain A and G406 and R324. The AMP group makes less extensive interactions: E402 and T431 H-bond with the ribose moiety whereas the adenine moiety is stabilised solely through non-bonded interactions.

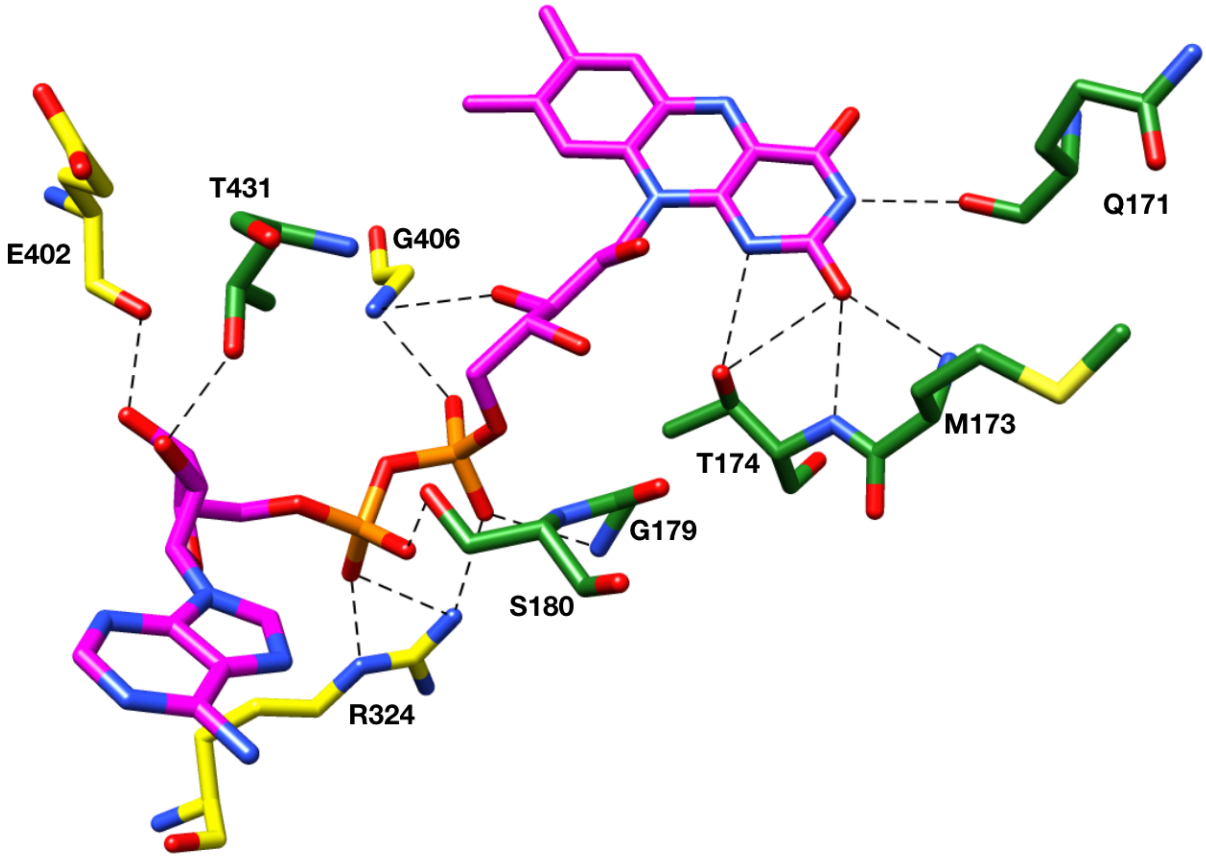


Figure 44. FAD hydrogen bonding network
FAD binds in the pocket at the dimer interface and held in place by numerous non-bonding contacts and a network of hydrogen bonds from residues from both monomers of the dimer. Residues with green coloured carbon atoms are from Chain A, whereas those in yellow are from Chain B (of an AB dimer). Hydrogen bonds are depicted by black dashed lines.

4.2.5.2 C10-CoA

A Bd2924:C10-CoA complex was solved in order to understand substrate binding. This was achieved by crystal soaking C10 and C12 chain length acyl-CoA thioesters into Bd2924 crystals. A chain length of C10-12 was chosen because the Bd2924 substrate-binding tunnel most closely resembles that of ACDH-11, which co-purified and co-crystallised with a C11-CoA thioester ligand. The structure of Bd2924 in complex with C10-CoA fatty acid revealed the interactions Bd2924 makes with substrate. In the Bd2924:C10-CoA complex structure only a single C10-CoA molecule can be confidently placed per dimer, due to weak electron density in the $2F_o$-F_c map for the second molecule.

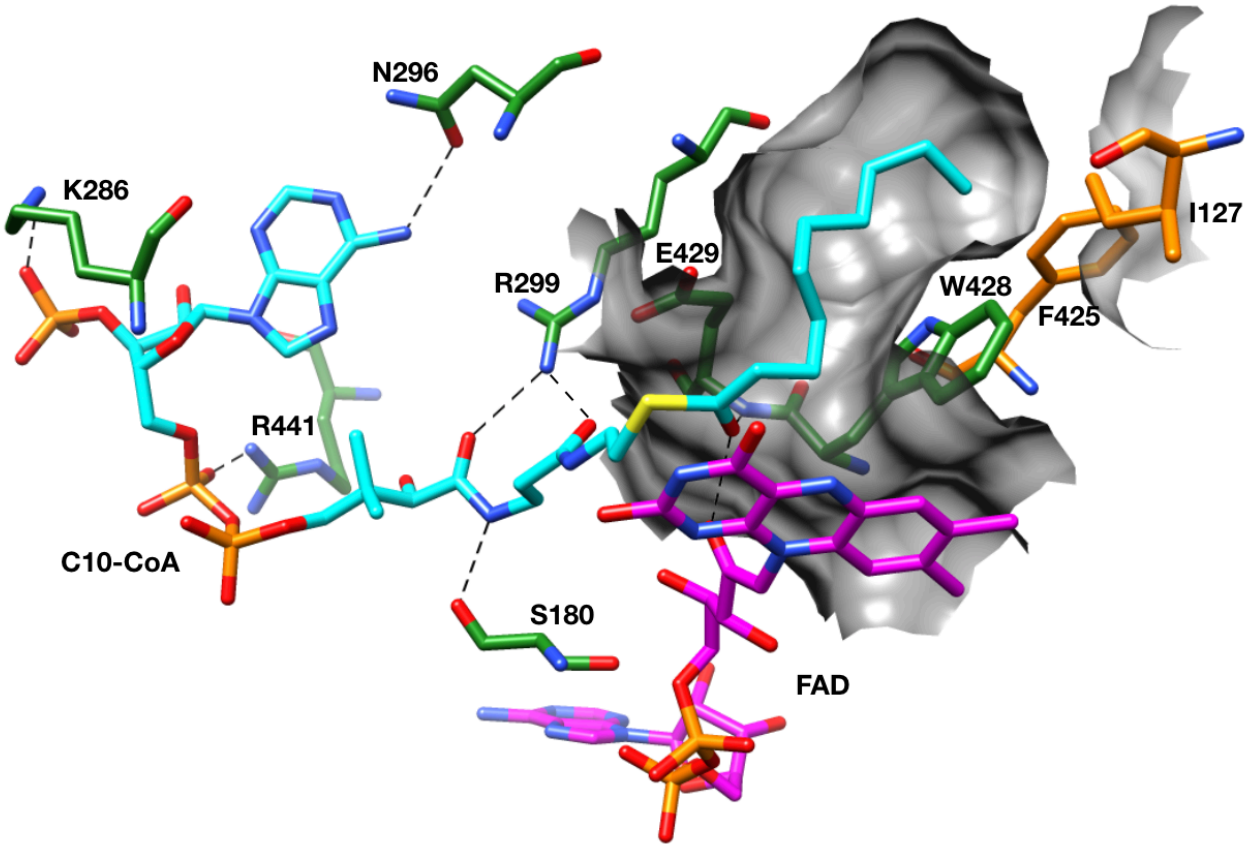


Figure 45. C10-CoA binding
The structure of Bd2924 has been solved in complex with a C10-CoA ligand (cyan). A number of hydrogen bonds (8) hold C10-CoA in place. An important hydrogen bond between the thioester carbonyl group and the backbone NH of E429 and the FAD-2'OH correctly orientates the substrate for catalysis. The depth of the substrate tunnel is limited by the sidechains of I127 and F428 (orange). Grey represents surface of pocket.

The C10-CoA binding site is located between the FAD binding site and the central α-helical domain, which forms a long narrow tunnel that runs deep into protein beneath the bound FAD molecule with a funnel shaped opening. The tunnel is narrow enough to only accept the fatty acyl portion of the C10-CoA ligand. The depth of the substrate-binding pocket appears to be determined by two bulky side-chain residues, I127 and F425, which restrict access, limiting substrates to fatty acyl chains of no more than 12 carbons in length. The adenine-portion of C10-CoA binds to the exterior of the tunnel opening. C10-CoA is bound to Bd2924 by multiple H-bonds which includes: one between N296 and the adenine group, one between K289 and the 3'-phosphate, two between the di-phosphate group and R441 and S232, two between the pantothene group and R299 and S180 backbone. The thiolate carbonyl of C10-CoA makes a H-bond with E429-NH and FAD 2'-OH. This interaction positions the fatty acyl chain flat between E429 and the re-face of FAD, poised for catalysis. It is likely from our structural analysis that E429 acts as the catalytic base in Bd2924, which is located on a loop between helix 11 and 12. The positioning of the C10-CoA ligand is similar to other ACAD:ligand complex structures, which suggests that E429 would be able to abstract a H^+ from the C_α causing a simultaneous hydride transfer to N5 of FAD from the C_β.

There is little conformational change between the C10-CoA complex and ligand-free structure of Bd2924. The sidechains of a few residues change conformation to adopt the C10-CoA ligand. R441 changes conformation in order to form a hydrogen bond with the pyrophosphate group. R379 and D438 reposition and form a salt bridge underneath the adenine group, acting as a platform. N296 also changes orientation to make a hydrogen bond with the adenine group. A conformational change in W428 is associated with allowing accessibility of the fatty acyl group to position beneath the re-face of FAD.

W428 is positioned under the C5-C10 ring of the isoalloxazine group of FAD in the Bd2924 ligand free structure and the ligand bound monomer in the Bd2924:C10-CoA complex structure. However, in the ligand free monomer of Bd2924:C10-CoA complex structure, W428 adopts two conformations; one under the C5-C10 ring and one central to the isoalloxazine group. The movement of W428 appears to gate the access of the fatty acyl group (Figure 49). The conformational shift may be an important feature of substrate displacement and access to the catalytic site, although this has not been tested further. Q171, located close to the active site adopts multiple conformations in the Bd2924:C10-CoA complex. Q171 re-orientates to make a potential H-bond with the catalytic E429 sidechain, which may alter the pK_a of E429 and decrease its proton abstracting ability.

4.2.6 CdG binding assay – MST

Bd2924 was identified as a potential novel CdG receptor and linked to the CdG regulatory network, a prominent feature of the predatory lifecycle of *Bdellovibrio bacteriovorus* [82]. To corroborate the findings of Rotem et al. we re-examined the CdG binding capability of Bd2924. MicroScale Thermothoresis (MST) and the published assay conditions were used to ensure consistency. Our investigation used a fluorescently labelled protein and label free CdG, opposed to native Bd2924 and a fluorescently tagged CdG originally used in the Rotem study. MST results reveal Bd2924 binds CdG with a K_D of ~71 μM (Figure 46). The K_D obtained from our binding assay is almost 10 fold greater than the published K_D of 7.67 μM.

Besides MST binding studies we also attempted to obtain a Bd2924:CdG complex through X-ray crystallography experiments. Co-crystallisation and ligand soaking experiments, with concentrations of CdG up to 2 mM, were carried out with little success. The failed attempts to determine the location of a binding site via crystallography alongside the relatively weak K_D obtained through our binding assay and no conventional nucleotide-binding sites in the structure, suggest CdG may bind to Bd2924 non-specifically.

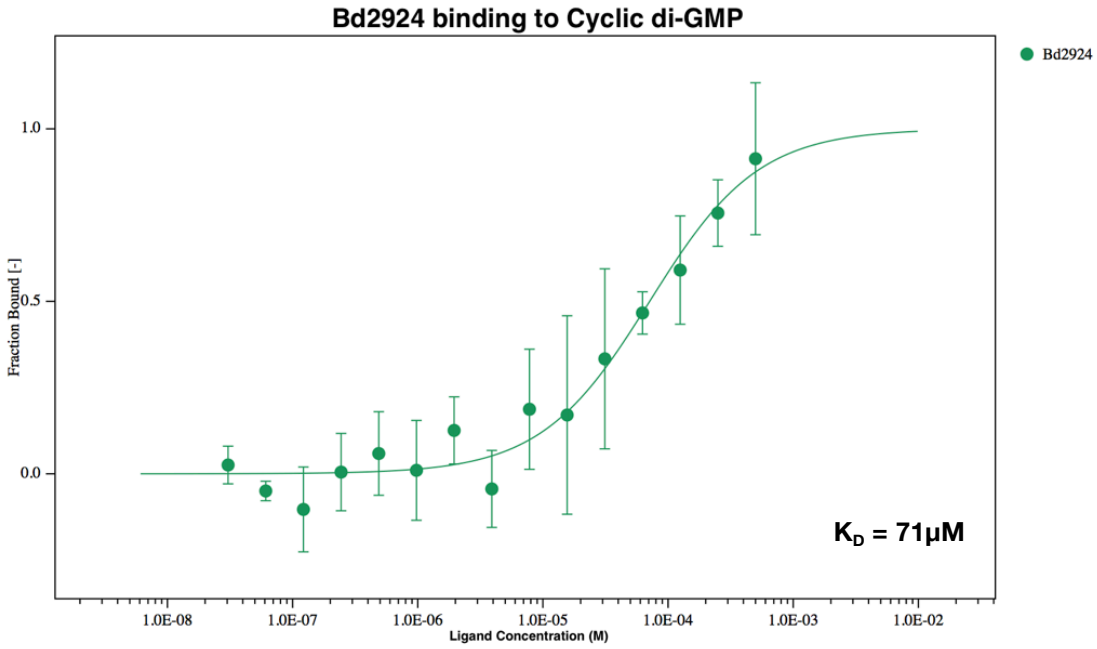


Figure 46. Binding curve with CdG
The binding interaction between Bd2924 and CdG was monitored using MST

4.2.7 Bd2924 activity

We were unable to detect any enzymatic activity for Bd2924 in two independent ACAD activity assays with either C10 or C12 fatty acyl-CoA substrates (DCPIP+PMS assay & Ferrocenium ion assay).

4.3 Discussion

4.3.1 Structural comparison to ACADs

The core structure of Bd2924 shares the common ACAD fold [180]. Bd2924 also forms a homotetramer, composed of a dimer of dimers, although Bd2924 does not form a typical ACAD tetramer. In ACADs the tetramer interface is provided by the C-terminal domains flat surface. However, in Bd2924 an additional C-terminal α-helical domain is located at this interface and mimics the C-terminal domain of the opposing dimer in ACAD tetramers [187] (Figure 47.A). The C-terminal domain of Bd2924 is similar to VLCAD that consequently forms a dimer [181] and ACAD homologues such as AidB and ACDH-11 [197, 198] that form a similar tetramer. The tetramer formed by Bd2924 is stabilised by a single β-sheet projection from the N-terminal domain that forms a claw like structure. The dimers slot together at 90° angles to one another, between the claw structures. The tetramer interface is unlike typical ACAD structures but highly similar to the divergent ACAD homologues ACDH-11 and AidB.

The Bd2924 structure most closely resembles the structures of ACDH-11 (Z-score = 28.2, rmsd = 2.8Å, number of residues = 593, %id = 43%) and AidB (Z-score = 26.2, rmsd = 2.3Å, number of residues = 538, %id = 29%). A major difference between the structure of Bd2924 and those of ACDH-11 and AidB is the extent of the C-terminal α-helical domain. In Bd2924 this domain is much shorter, lacking the final two helices. The lack of these two helices creates a large open groove between the dimers, which is a unique feature whose relevance is unknown (Figure 47.D). It may act as a docking site for protein-protein interactions.

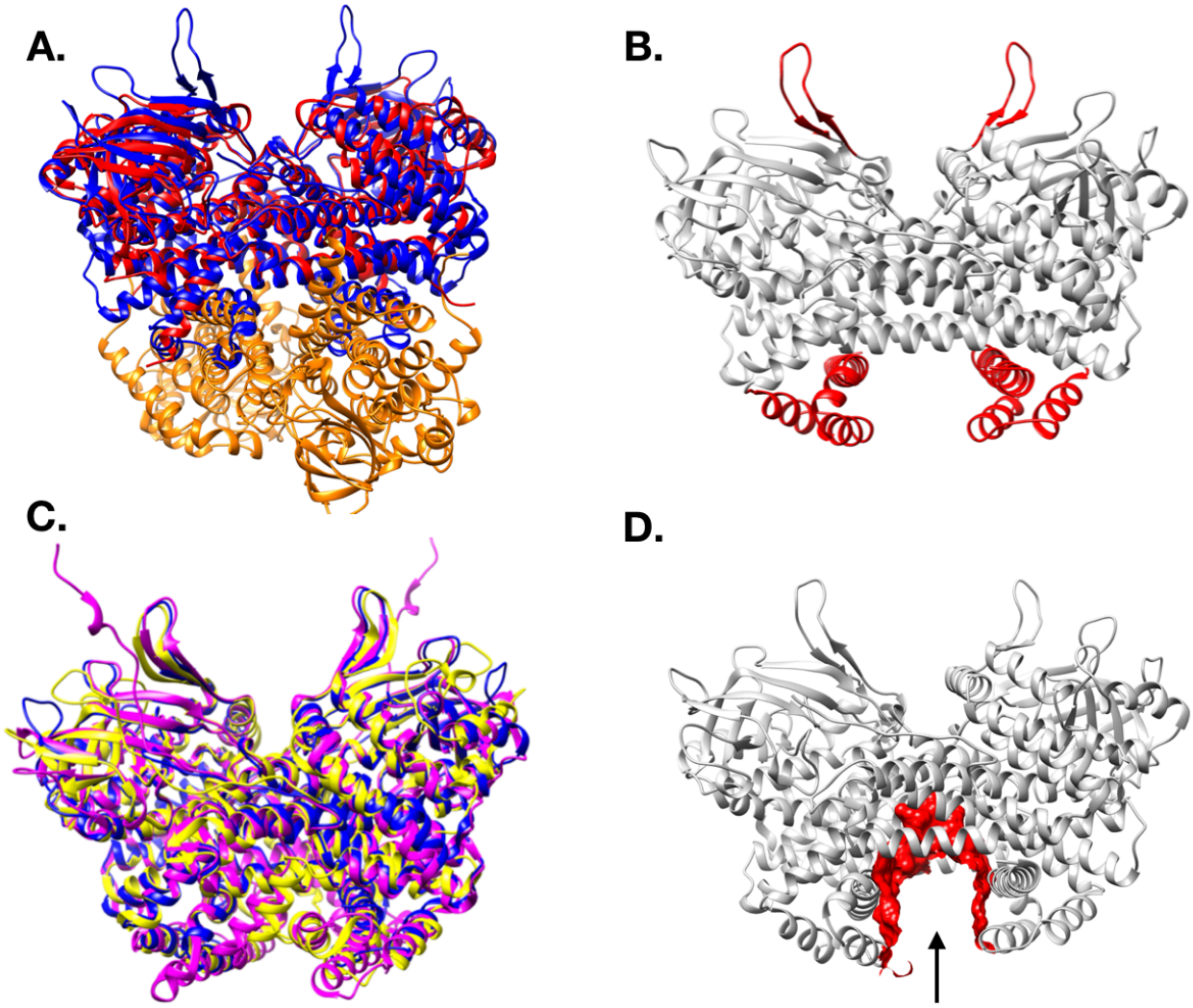


Figure 47. Bd2924 structural comparison.
A. Bd2924 dimer (blue) overlaid with MCAD tetramer (PDB:1EGD, red + orange, length = 746 residues, rmsd = 2.21Å). **B.** The major structural difference between Bd2924 and MCAD are highlighted in red on the structure on Bd2924. The N-terminal β-sheet loop and C-terminal extension domain are features of divergent ACAD homologues that form an unrelated tetramer. **C.** Bd2924 (blue) overlaid with ACDH-11 (PDB:4Y9J, magenta, 961 residues, 1.41Å) and AidB (PDB:3DJL, yellow, length = 881 residues, rsmd = 1.54Å). **D.** Bd2924 has a less extensive C-terminal extension domain and as a result has a large groove between dimers (red). Alignments were performed using the SSM superimpose tool on coot.

4.3.2 Substrate-binding and activity

Bd2924 is able to bind C10 chain-length fatty acyl-CoA, demonstrated by the Bd2924:C10-CoA complex structure. Bd2924 binds C10-CoA with more interactions (eight hydrogen bonds) than MCAD binds acyl-CoA substrates (four hydrogen bonds)

[187]. Interestingly ACDH-11 makes at least ten hydrogen bonds with C11-CoA, which is believed to facilitate the ACDH-11 role as a fatty acyl-CoA sensor [197].

We were unable to detect any dehydrogenase activity for Bd2924 (with C10 and C12-CoA substrates). Likewise no dehydrogenase activity has been reported for the divergent ACAD homologue ACDH-11, while only very weak dehydrogenase activity has been reported for AidB [198, 202]. In AidB the position of W424 is suggested to be the reason for the weak activity observed. W424 is positioned beneath the isoalloxazine group of FAD, which blocks the accessibility to the N5 position of FAD [199]. This is similar to the conformational change in W428 in our Bd2924:C10-CoA complex where W428 adopts a conformation that appears to gate access of the fatty acyl chain. Additionally, Bd2924 Q171 makes a hydrogen bond with the catalytic E429, which may impede the ability of E429 to abstract a proton from the C_α and initiate catalysis. In ACDH-11 a similar hydrogen bonding interaction occurs between N206 and the catalytic E464. In other proven catalytically active ACADs this residue is typically a non-polar residue such as phenylalanine and therefore does not affect the chemistry of the catalytic base [183] (Figure 48).

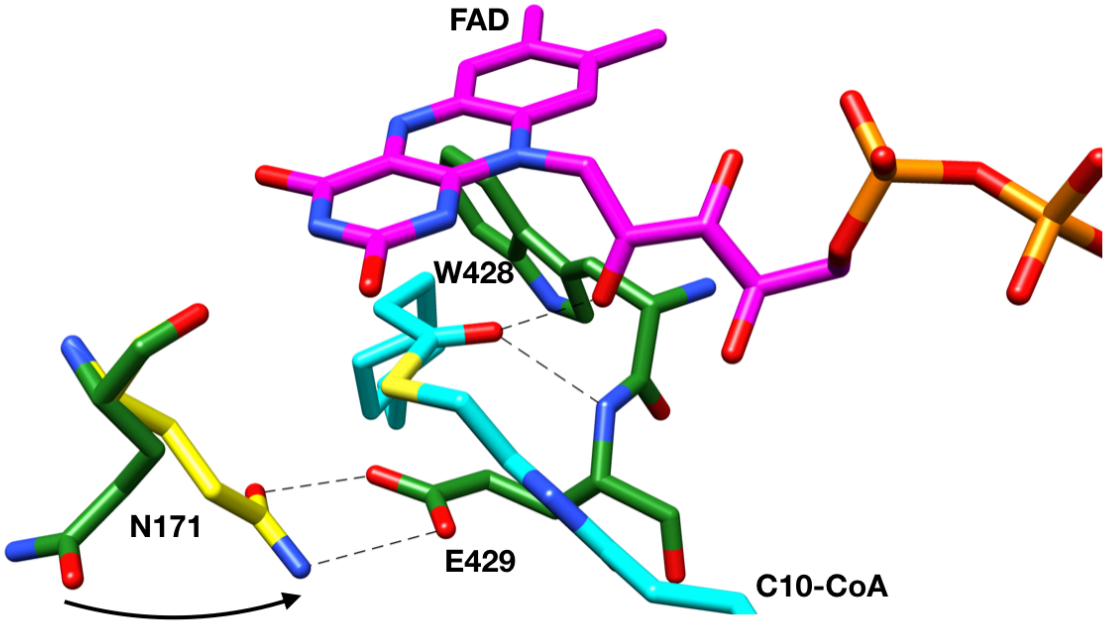


Figure 48. Bd2924 active site.
The C10-CoA ligand is bound in the active site between FAD and the catalytic base E429. In Bd2924:C10-CoA complex N171 adopts a new conformation (yellow) that forms a hydrogen bond with the sidechain of E429 and may affect its ability to abstract hydrogen from the C_α (supressing catalytic funciton).

Besides the lack of detectable dehydrogenase activity for Bd2924 (and comparison to other ACAD homologues) it seems unlikely that transfer of electrons to ETF would be possible by the currently proposed mechanism [191]. Bd2924 tetramer formation would impede access to the FAD molecule buried deep between the dimer and tetramer complex. As such for an ETF to access and re-oxidise the buried FAD molecule in Bd2924, it would have to disassemble our observed tetramer complex [203]. Blocking of ETF access is demonstrated in ACO, where a C-terminal α-helix crosses the solvent accessible groove to the si-face of FAD and thereby shields the accessibility to ETF [203]. Bd2924 could alternatively utilise molecular oxygen similar to ACOs although solvent accessibility to the FAD molecule is not as extensive as in ACOs.

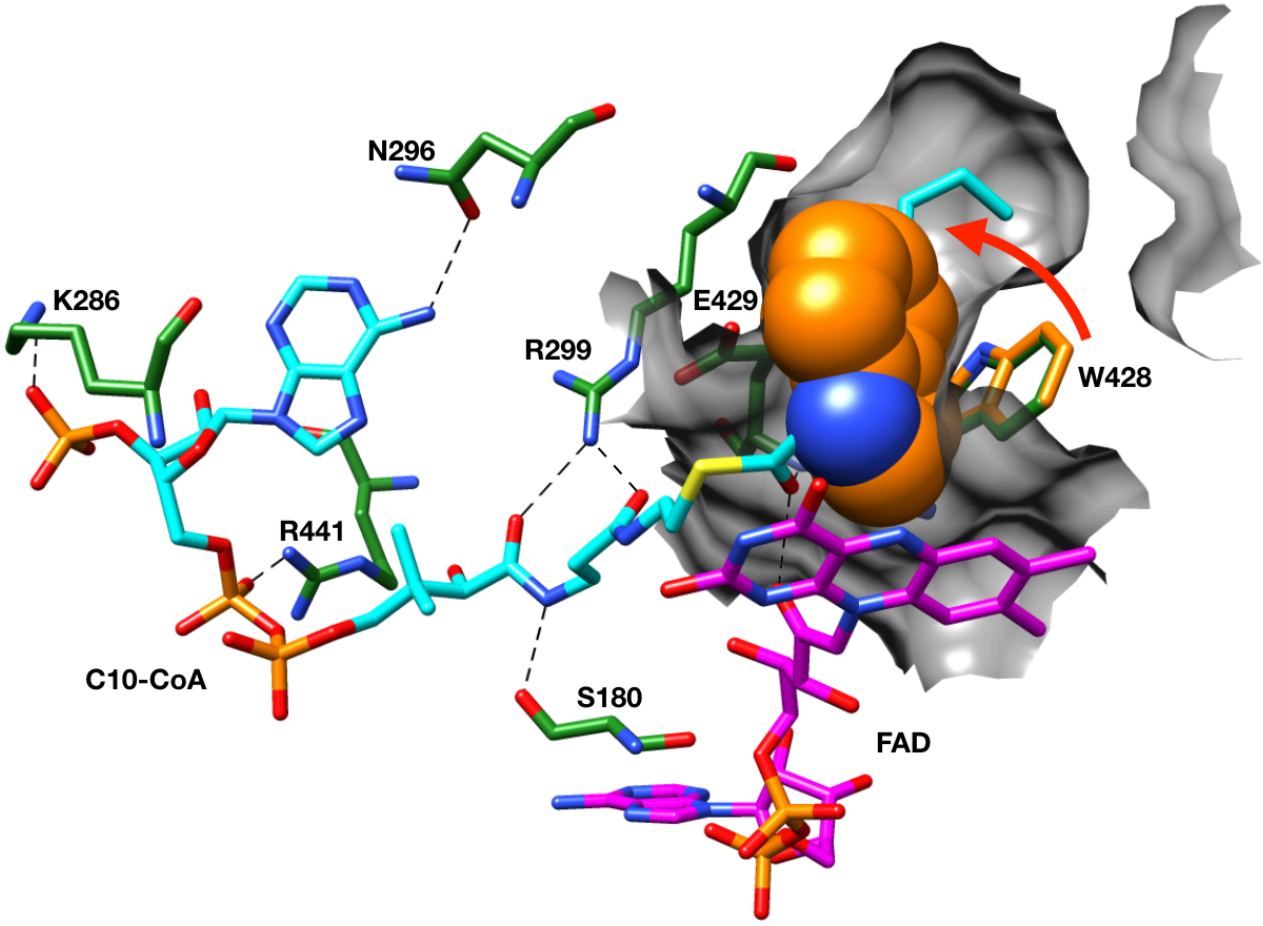


Figure 49. Trp428 gates the fatty acyl binding tunnel
The Bd2924:C10-CoA complex structure has one occupied C10-CoA binding site. In the monomer where the substrate binding site is free of ligand, W428 adopts two conformations (orange). One W428 conformation is located directly beneath the FAD N5, which blocks access to the substrate binding tunnel.

4.3.3 A functionally relevant Bd2924 interaction with CdG is questionable

We determined the binding affinity between Bd2924 and CdG to be ~71 μM, which is approximately 10 fold greater than the published value determined by Rotem et al [82]. To ensure consistency the binding buffer used by Rotem et al. was replicated, but no measurable binding events were observed with a maximum CdG concentration of 5 mM, which far exceeds the cellular concentration of CdG [79]. Binding was detected after rigorously testing multiple assay buffers and protein labelling techniques. The increased K_D value we obtained suggests that Bd2924 binds CdG in a non-specific

manner. Labelling the protein may have obscured CdG binding to Bd2924, but this is unlikely because two independent labelling procedures produced similar results. Alternatively the labelled CdG ligand used by Rotem et al. may have increased the binding affinity. Closer inspection of Rotem et al. binding data reveals a fully saturated bound state was not achieved (similar to our binding data). It is therefore possible that an incorrect extrapolation of the data may have led to an over-estimation of binding affinity.

Numerous proteins identified as novel CdG receptors by Rotem et al. appear to be associated with CoA related substrates. The list contains: acetyl-CoA acyltransferases Bd2070 and Bd1835, Bd1833 (a malic enzyme that is tightly regulated by CoA - also investigated in this thesis), Bd2174 a (butyryl-CoA dehydrogenase) and Bd2924 itself. It is therefore possible that a large proportion of the novel CdG binding proteins are artefacts (false positives) caused by non-specific binding arising from proteins that naturally interact with nucleotide containing substrates [82].

4.3.4 Role of Bd2924

Bd2924 is structurally related to divergent ACAD homologues that have cellular functions beyond β-oxidation. ACAD homologues have been linked to responding to environmental stresses such as the role of FadF in bacterial survival in carbon starvation conditions in *Salmonella typhimurium* [204]. Additionally ACDH-11 is proposed to respond to heat changes by sequestering C11/12 fatty acyl-CoA esters thereby supressing *fat-7* expression and driving heat adaptation [197]. Furthermore AidB is implicated in an adaptive response to DNA damage by alkylating agents [198].

Bd2924s role is also likely to be divergent from β-oxidation based on its structural similarities to divergent ACADs and the lack of detectable dehydrogenation activity.

The exact function of Bd2924 remains unknown and requires further exploration. However, it is interesting to speculate that Bd2924 may be involved in fatty acid detection and utilisation during the *Bdellovibrio* predatory cycle. *Bdellovibrio* utilises prey-derived long-chain fatty acids for energy through β-oxidation and for direct incorporation into biosynthetic pathways [59]. Bd2924 may therefore act as a sensor of fatty acyl chains. Its function may be to: transport fatty acyl-CoA; sequester fatty acyl-CoA as part of a signalling cascade (alike ACDH-11); or generally be involved in β-oxidation. ETF genes exist in *Bdellovibrio HD100* genome (Bd0021 and Bd0022) but these may interact with other acyl-CoA dehydrogenases, such as those named in Table 12. Close homologues of Bd2924 are found in other endo-biotic predatory bacteria (60-75% ID) and *Legionella* species (55% ID). It is therefore interesting to speculate that Bd2924s role is specific to bacteria which invade cells.

Protein	Type	Length	Up-regulated in?
Bd2174	Butyryl-CoA dehydrogenase	816 (including 284aa DUF)	growth phase
Bd1973	Butyryl-CoA dehydrogenase	388	growth phase
Bd1201	Isovaleryl-CoA dehydrogenase	428	growth phase
Bd2924	acyl-CoA dehydrogenase	505	growth phase

Table 12. Hd100 ACADs
A list of ACADs in the HD100 genome – annotated by uniprot

The shortened C-terminal α-helical domain is a unique structural feature that differentiates Bd2924 from other ACAD structures. The wide groove found between monomers may have a physiological role unique to Bd2924. The groove is highly charged with numerous basic and acidic residues lining the surface. The groove may therefore be a site of protein-protein interactions. Interactions with a binding partner may elicit Bd2924 activity or regulate the activity of the binding partner.

4.4 Conclusion

We have determined the three-dimensional structure of Bd2924 demonstrating an adaptation of the common ACAD fold. Bd2924s structure more closely resembles those with a divergent function, such as ACDH-11 and AidB. No dehydrogenation activity was detected for Bd2924 suggesting it may also possess an important role, beyond fatty acid β-oxidation, such as responding to environmental cues during the lifecycle of *Bdellovibrio*.

Chapter 5

Bd1833 a Multi-modular Malic Enzyme

5.1 Introduction

Metabolism in *Bdellovibrio* has been understudied, yet it is an important and functionally interesting stage of the predatory lifecycle. Many metabolic pathways have been adapted to manage the influx of prey-derived metabolites used for energy and growth. A study by Rotem et al. identified a number of possible CdG regulated metabolic enzymes [82]. One of these enzymes, Bd1833, is a putative multi-modular malic enzyme (ME), comprised of two domains, a Malic enzyme (ME) domain and an additional presumed non-catalytic phosphotransacetylase (Pta) like domain. Both domains have very different biological roles and catalyse unrelated reactions. MEs catalyse the reversible decarboxylation reaction of L-malate to pyruvate, while Pta enzymes reversibly catalyse the transfer of acetyl groups between CoA and inorganic phosphate. The combination and regulatory-signalling between ME and Pta domains of multi-modular MEs have remained uncharacterised.

5.1.2 Malic Enzymes

5.1.2.1 Overview

Malic enzymes catalyse the oxidative decarboxylation reaction of L-malate to pyruvate using the cofactor $NAD(P)^+$ in the presence of a divalent cation such as Mn^{2+} or Mg^{2+}.

$$\text{L-malate} + \text{NADP}^+ \rightleftharpoons \text{pyruvate} + CO_2 + \text{NADPH}$$

MEs are categorised into three classes based on coenzyme specificity and their ability to decarboxylate an oxaloacetic acid (OAA) substrate. These classes include: Non-OAA decarboxylating NAD^+ -specific, OAA decarboxylating NAD^+ -specific and OAA decarboxylating $NADP^+$ -specific. MEs are widely distributed enzymes involved in different metabolic pathways such as energy metabolism, photosynthesis and lipogenesis in organisms from bacteria to higher Eukaryotes [205]. Three isoforms of eukaryotic ME have been identified and studied (cytosolic, mitochondrial and plastid). On the other hand, prokaryotic MEs isoforms include a wider range of structural and functional diversity (including multi-modular MEs), which are comparatively understudied.

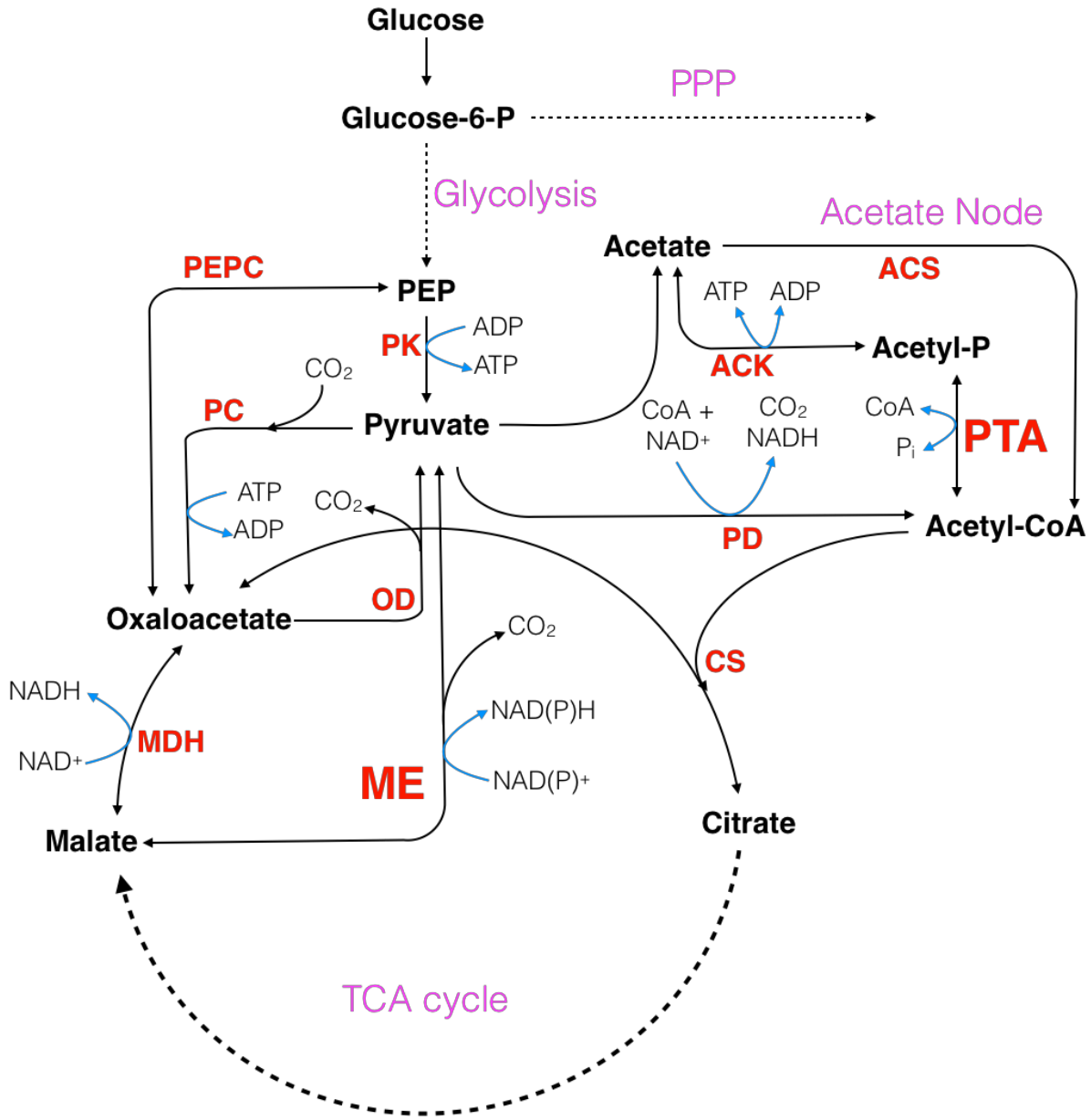


Figure 50. Illustration of central metabolism (general)
This Figures depicts a simplified illustration of a central metabolic network involving Pta and ME enzymes (based on *E. coli* metabolism). Each individual pathway is labelled in pink. The enzymes involved in the reactions are displayed in red: PK - pyruvate kinase, PD - pyruvate dehydrogenase, CS - citrate synthase, PEPC - phosphoenolpyruvate carboxylase, PC - pyruvate carboxylase, OD - oxaloacetate decarboxylase, MDH - malate dehydrogenase, ME - malic enzyme, ACS - acetyl-CoA synthase, ACK - acetate kinase, PTA - phosphotransacetylase

5.1.2.2 Malic enzyme roles

Malic enzyme isoforms are involved in a range of metabolic pathways. Generally NAD^+ dependent MEs function in anaplerotic reactions to replenish TCA cycle intermediates, whereas $NADP^+$ dependent ME function to generate NADPH for anabolic pathways.

For example, cytosolic human NADP-ME is required to provide NADPH for biosynthesis of fatty acids and steroids [206, 207]. Furthermore, MEs have been implicated in lipid storage [208]. On the other hand, human mitochondrial NAD-ME is implicated in energy metabolism of neurons and tumour cells [209]. This includes their involvement in glutaminolysis (metabolism of glutamine) to supply energy to rapidly proliferating tissues such as tumours [210, 211].

In plants some MEs have specialised functions such as providing CO_2 for fixation by RuBisCO in photosynthesis [212] [213]. Additionally NADP-MEs from plants have been implemented in plant defences and oxidative stress responses [214]. However, MEs in C4 plants generally function to replenish TCA cycle intermediates [213].

5.1.2.3 Alternative metabolism and NADPH generation

MEs are associated with the PEP-pyruvate-oxaloacetate node; the metabolic link between glycolysis/gluconeogenesis and the TCA cycle [215]. Consequently, MEs are important for growth on TCA cycle intermediates and other compounds entering the TCA cycle via acetyl-CoA.

ME positioning in the PEP-pyruvate-oxaloacetate node suggests their importance for cellular NADPH production [216]. NADPH is an important cofactor, required for many anabolic reactions, whereas NADH is generally used in catabolic reactions requiring an oxidising agent. Moreover the redox potential of NADPH is also used to control toxicity in cells.

The oxidative pentose phosphate pathway (PPP) and the isocitrate dehydrogenase step of the TCA cycle are typified as the main NADPH generating systems [216]. More recently MEs have been proposed as an alternative way to generate NADPH. ME can operate in metabolic pathways to convert NADH to NADPH (act as a kind of transhydrogenase) [217]. This pathway is referred to as the malate shunt and involves reactions catalysed by ME, MDH (malate dehydrogenase) and PC (pyruvate carboxylase). Pyruvate is converted to OAA (ATP dependent) and back to pyruvate by NADH and $NADP^+$ catalysed reactions [218]. A direct link between PPP and ME has been established between 6-phospogluconate dehydrogenase and Malic-enzyme 1 (ME1), supporting flux through the PPP and NADPH production [219].

5.1.2.4 Malic enzyme structure

MEs are a relatively new class of oxidative decarboxylases [205] that have highly conserved primary sequences and structures. Structures have been solved for eukaryotic and prokaryotic ME isoforms, although eukaryotic ME are better characterised.

Eukaryotic ME are structurally and functionally well-characterised enzymes, first detected in extracts from pigeon liver. The first eukaryotic ME structure was published for Human mitochondrial $NAD(P)^+$-dependant malic enzyme [220]. Since then, structures for pigeon liver cytosolic NADP-ME [221] and *Ascaris suum* mitochondrial NAD-ME [222] and human cytosolic NADP-ME [223] have been determined.

Eukaryotic MEs exist primarily as homotetramers (dimer of dimers) made up of 60kDa monomers. Each monomer can be divided into four domains (A-D) (Figure 51A). Domain A is a small α-helical domain that provides the dimer interface. Domain B is a much larger mixed β-sheet/α-helical domain that contains the cation-binding site. Domain C possesses a Rossmann fold [224] and is the domain associated with $NAD(P)^+$ binding (GXGXXG nucleotide-binding motif). Finally domain D is a small domain containing a single α-helix that contributes to the tetramer interface. The ME active site is located between domains B and C.

Eukaryotic MEs have been observed in two catalytic states - “open” and “closed”. In the open form (PDB:1QR6) the active site is accessible to substrates and to the nucleotide cofactor. The closed form (PDB:1EFK) on the other hand has an enclosed catalytic site, which is achieved by an 11° rigid body rotational movement of domain C relative to domain B [225] (Figure 51C).

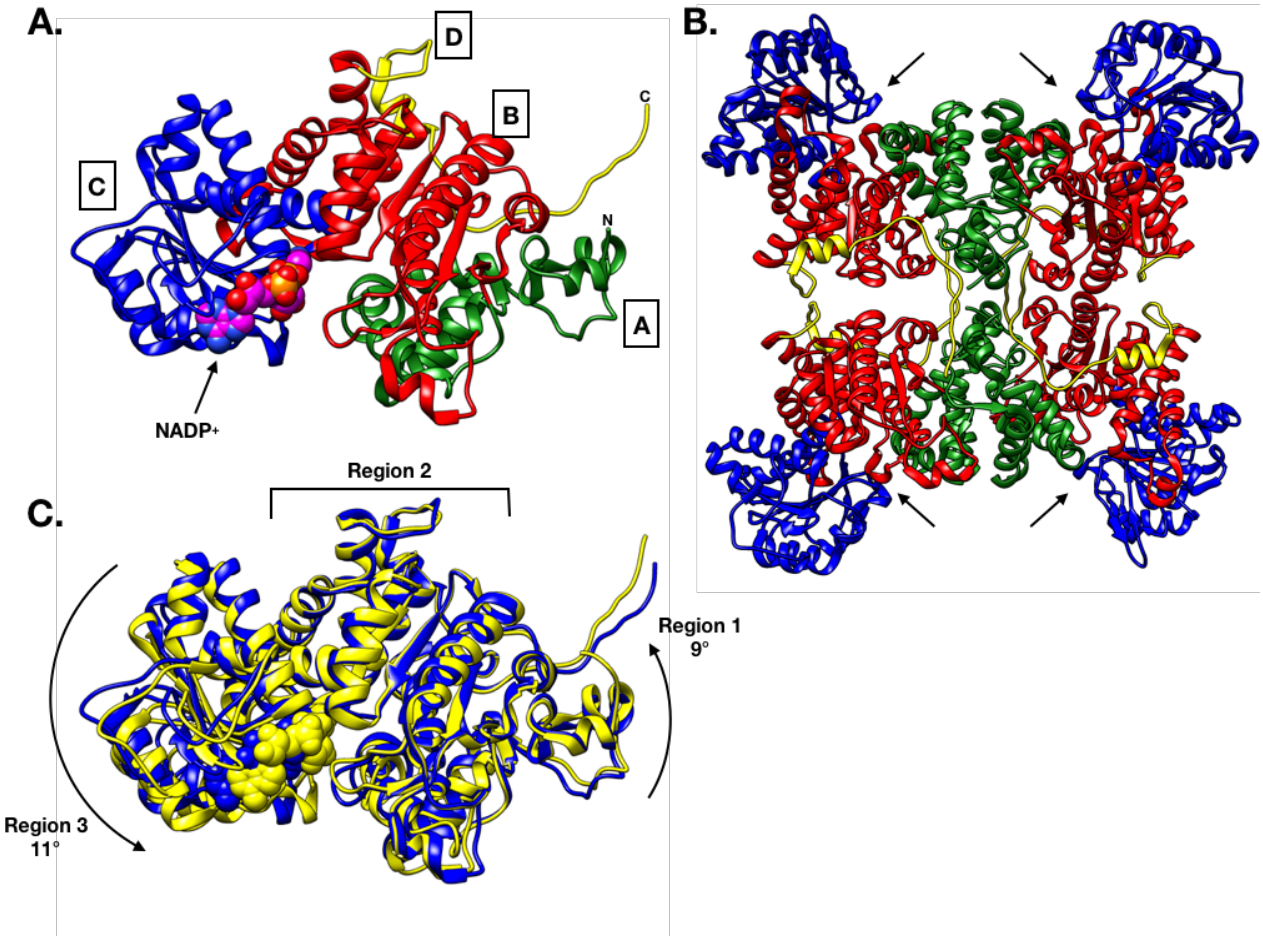


Figure 51. Human-ME tetramer organisation and domains
A. Human-ME like other eukaryotic MEs can be split into four domains: Domain A (Green) makes the dimer interface; Domain B (Red) composes the central core of the protein and cation-binding site; Domain C (Blue) contains the NAD(P)$^+$ binding site; and Domain D (yellow) is a short extension that contributes to the tetramer interface. **B.** Eukaryotic MEs arrange to form a tetramer complex – a dimer of dimers. Black arrows indicate the location of the substrate-binding cleft. **C.** The structure of human-ME has been captured in a catalytically active 'closed' conformation and a resting 'open' conformation. Region-3 including domain C (NADP$^+$ -binding domain) undergoes an 11° rigid body rotational movement that encloses the active-site cavity and positions the catalytic residues and cofactor ready for catalysis. A smaller 9° rigid body rotational movement of part of domain A (region-1) also occurs, which re-orientates the tetramer interface. Region-2 containing domain B and part of domain A maintains the same conformation. The closed form (PDB:1EFK) was superimposed onto the open form (PDB:1QR6), using LSQ superpose tool in coot to align domain B (1275 atoms, rmsd = 0.94Å).

5.1.2.5 Divalent Cation binding

ME activity is dependent on the presence of a divalent cation such as Mn^{2+} and Mg^{2+}. In human NAD(P)$^+$ -dependant-ME, Mn^{2+} is ligated by six oxygen atoms in an octahedral fashion. Three of these oxygen atoms are from acidic residue sidechains; E255, D279 and D256. When no ligand is present in the active-site water molecules provide the remaining oxygen atoms.

5.1.2.6 Substrate/ inhibitor binding

Substrate binding in eukaryotic ME has been well documented and various ME substrates/inhibitor complexes have been published. ME sequentially binds substrate after the cofactor and cation are already in place [226]. Upon binding of the substrate the enzymes undergo a large conformational change. The enzyme changes from an open state to a closed state, which shields the active site from bulk solvent. In human-ME a rigid body movement mediates the 'closed state', where domain C ($NADP^+$ -binding domain) moves relative to domain B (cation-binding domain). Domains A and D also make a rigid body movement that changes the orientation of the tetramer complex [225]. A potent lanthanide inhibitor (Lu^{3+}) was used to demonstrate that domain A and domain D conformational movements are independent of each other. These findings suggest more than one ME open form exists and cooperativity affects transition between states [227].

A variety of different substrate analogue inhibitors (oxalate, tartronate, and ketomalonate) [225] and substrates have been characterised [228]. L-malate binds deep in the active-site pocket ligated to the divalent cation by the C2 hydroxyl group, with one of the C1 hydroxyl oxygen atoms interacting with the divalent cation. Furthermore L-malate makes H-bonds with the sidechains of R165, K183, N421 and N466, and the ribose sugar of NAD^+ [228] (Figure 52).

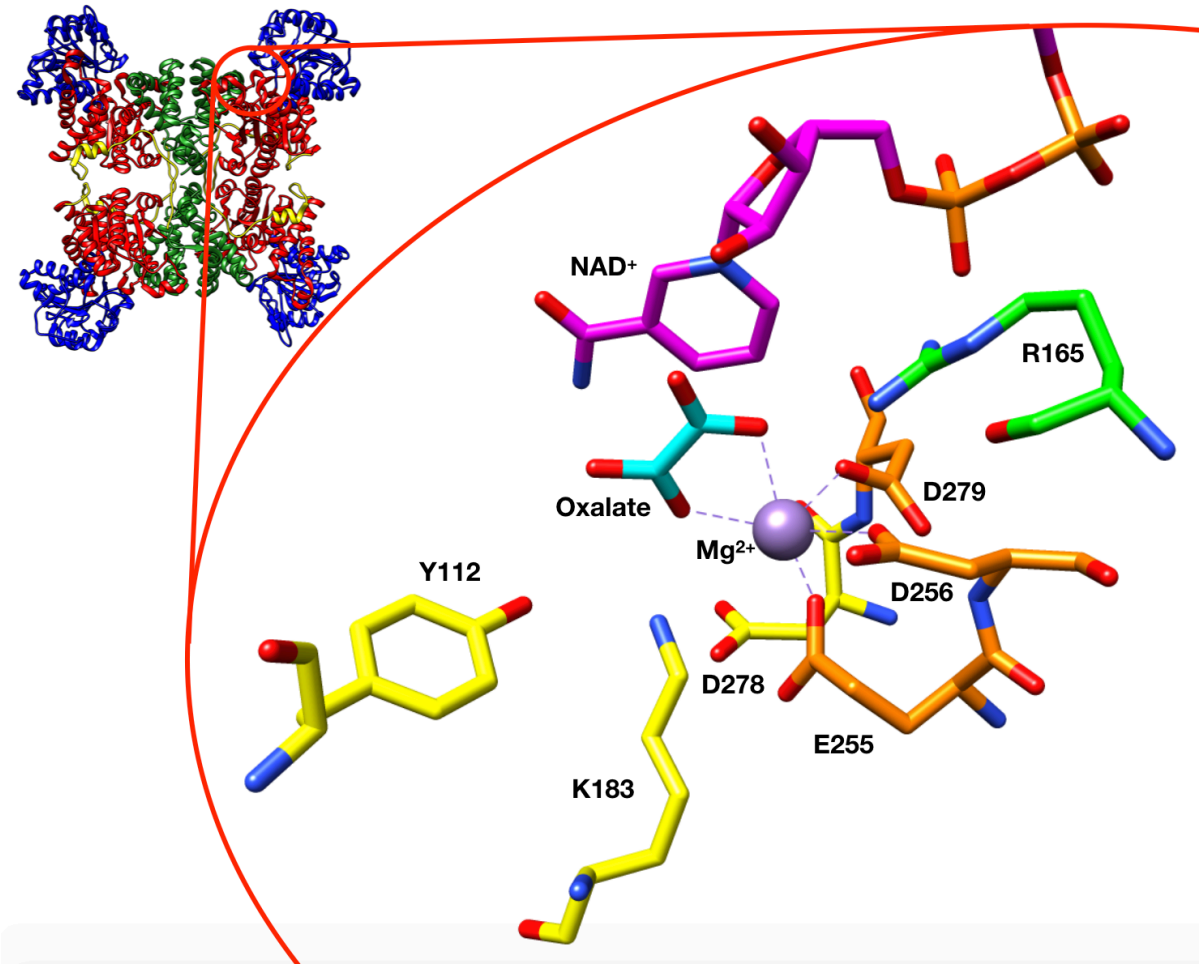


Figure 52. Human-ME active Site
A close up view of Human-ME active site in complex with: Mg^{2+} cation (purple), a malate substrate analogue – oxalate (cyan) and NAD^+ cofactor (magenta). The cation coordinating residues are shown in orange. The substrate-binding residue (R165) is shown in green. Catalytic residues (Y112, K163 and D278) are shown in yellow. PDB:1DO8

5.1.2.7 Nucleotide Binding

ME utilise a $NAD(P)^+$ cofactor to perform catalysis. Generally MEs perform the reaction specifically using either NAD^+ or $NADP^+$. However, some isoforms are able to utilise both cofactors. $NAD(P)^+$ is bound to domain C in eukaryotic MEs and associates with the Rossmann fold. The nicotinamide portion of the cofactor binds in an *anti* conformation, similarly to how other dinucleotides bind to a Rossmann fold. The

adenine group lies on the protein surface, stacked between residues 346 and 393 of human-ME but making no specific interactions.

The adenine portion of the cofactor is bound significantly different between isoforms that use either NAD^+ or $NADP^+$. Recent structures demonstrate how specificity is achieved. An aspartate residue near towards the end of Rossmann fold recognises the 2'-hydroxyl group and indicates the enzyme will be NAD^+ specific. However, many $NADP^+$ specific ME also have an Aspartate at this position; which instead makes a salt bridge with a neighbouring arginine residue thus preventing repulsion of the 2'-phosphate [221]. A generally conserved Lys residue, whose sidechain makes a H-bond with the 2'-phosphate group of $NADP^+$, confers $NADP^+$ specificity. This is also supported by mutagenesis studies that reveal an increased $K_{M(NADP)}$ in a K362A mutants, whereas the k_{cat} and K_M for malate and Mn^{2+} remained unaffected [229, 230]. The dinucleotide-binding motif also impacts on specificity, where GXGXXG is generally preferred by NAD specific ME and GXGXXA by NADP specific ME [231].

5.1.2.8 The catalytic residues and mechanism

ME catalytic mechanism is proposed to involve three steps: the dehydrogenation of L-malate to form OAA, decarboxylation of OAA to produce enolpyruvate and finally tautomerization of enolpyruvate to create pyruvate. Catalysis requires a general acid-base pair, cofactor and divalent cation.

The published catalytic mechanism of the *A. suum* ME is a model catalytic mechanism for MEs, outlined in Figure 53 [232]. MEs bind substrates and cofactors in an ordered

sequential manner starting with binding of $NAD(P)^+$ and a divalent cation (Mg^{2+}) [226, 233]. The ME:cofactor complex adopts an open conformation where the proposed general base (K199) makes a hydrogen bond to D294. L-malate binds to the open conformation interacting directly with Mg^{2+} via the C1 carboxyl group and the C2 hydroxyl group. Malate binding initiates a rigid-body conformational change that results in a closed active-site cavity, shielded from bulk solvent. Closure of the active site brings Y126 within H-bonding distance of K199 and causes a proton to be transferred to D294 as it approaches E271 (Mg^{2+} coordinated). Dehydrogenation is initiated by proton abstraction from the C2 hydroxyl group of L-malate by K199, resulting in a hydride transfer to NAD^+. Dehydrogenation is followed by the decarboxylation of the oxaloacetate intermediate to form enolpyruvate. Decarboxylation is facilitated by K199, this time acting as a Lewis acid, assisted by the bound metal ion. Finally a tautomerization step involves a general acid/base catalytic process. K199 accepts a proton from the C2 hydroxyl group, which coincides with synchronous protonation at the C3 position, performed by Y126. After catalysis NAD(P)H and pyruvate products are released, which resets the enzyme's protonation state and the conformational state returns to normal [232].

Stereospecificity of MEs for L-malate is demonstrated by L-malate and $NAD(P)^+$ enzyme complexes. The hydrogen on the C2 carbon points towards the C4 atom of the nicotinamide ring of NAD^+ (within hydride transfer distance) [220]. Such a binding mode cannot accommodate D-malate, hence, hydride transfer is prevented.

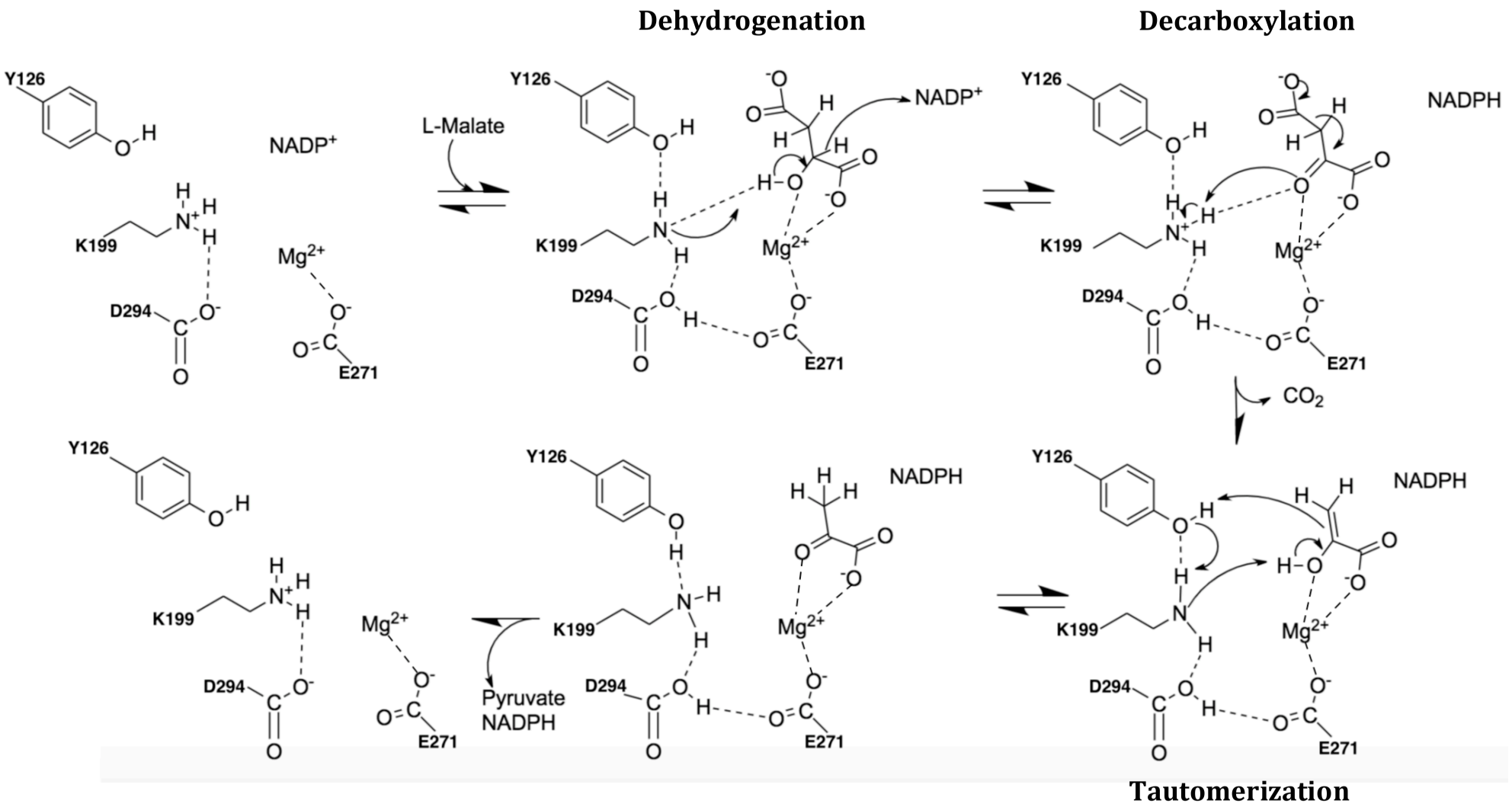


Figure 53. Malic enzyme catalytic mechanism.
The catalytic mechanism performed by malic enzymes progresses via three steps: Dehydrogenastion of L-malate to form oxaloacetate, decarboxylation of oxaloacetate to form enolpyruvate and finally tautomerization of enolpyruvate to form pyruvate. The enzyme is primed by binding Mg^{2+} and $NADP^{+}$ and upon substrate binding a conformational change repositions the catalytic residues.

5.1.2.9 Regulation

A number of compounds, many of which are associated with TCA intermediates, such as α-Ketoglutarate, citrate, lactate, glutamate, aspartate, OAA, fumarate and ATP, regulate the activity of MEs [234]. Many of these regulators display allosteric sigmoidal kinetics and bind to allosteric-binding pockets between dimer and tetramer interfaces [222, 235].

5.1.2.10 Prokaryotic MEs

Prokaryotic and eukaryotic MEs perform similar metabolic functions, while a greater diversity of prokaryotic MEs isoforms exist. Prokaryotic ME isoforms range from smaller ~40kDa subunits to much larger multi-modular 85kDa subunits. Several Gram-positive bacteria contain smaller ME isoforms, including: MEs from *Bacillus stearothermophilus* [236], *Streptococcus bovis [237]*, *Corynebacterium glutamicum* [238] and CitM from *Lactococcus lactis* [239]. Additionally in the *Bacillus subtilis* genome there are four putative ME paralogues of varying sizes that all possess ME activity to different degrees [240]. On the other hand, ME subunits that are a similar size to eukaryotic MEs have also been identified, such as the ~60kDa SfcA from *Escherichia coli* [241].

The greater structural diversity of prokaryotic ME may be related to the slightly different roles each isoform performs and the need to regulate the activity accordingly. Eukaryotic cells compartmentalise isoforms (mitochondrial and cytosolic); a physical barrier therefore reduces the need for complex regulatory mechanisms and structural diversity to distinguish cellular roles for each isoform.

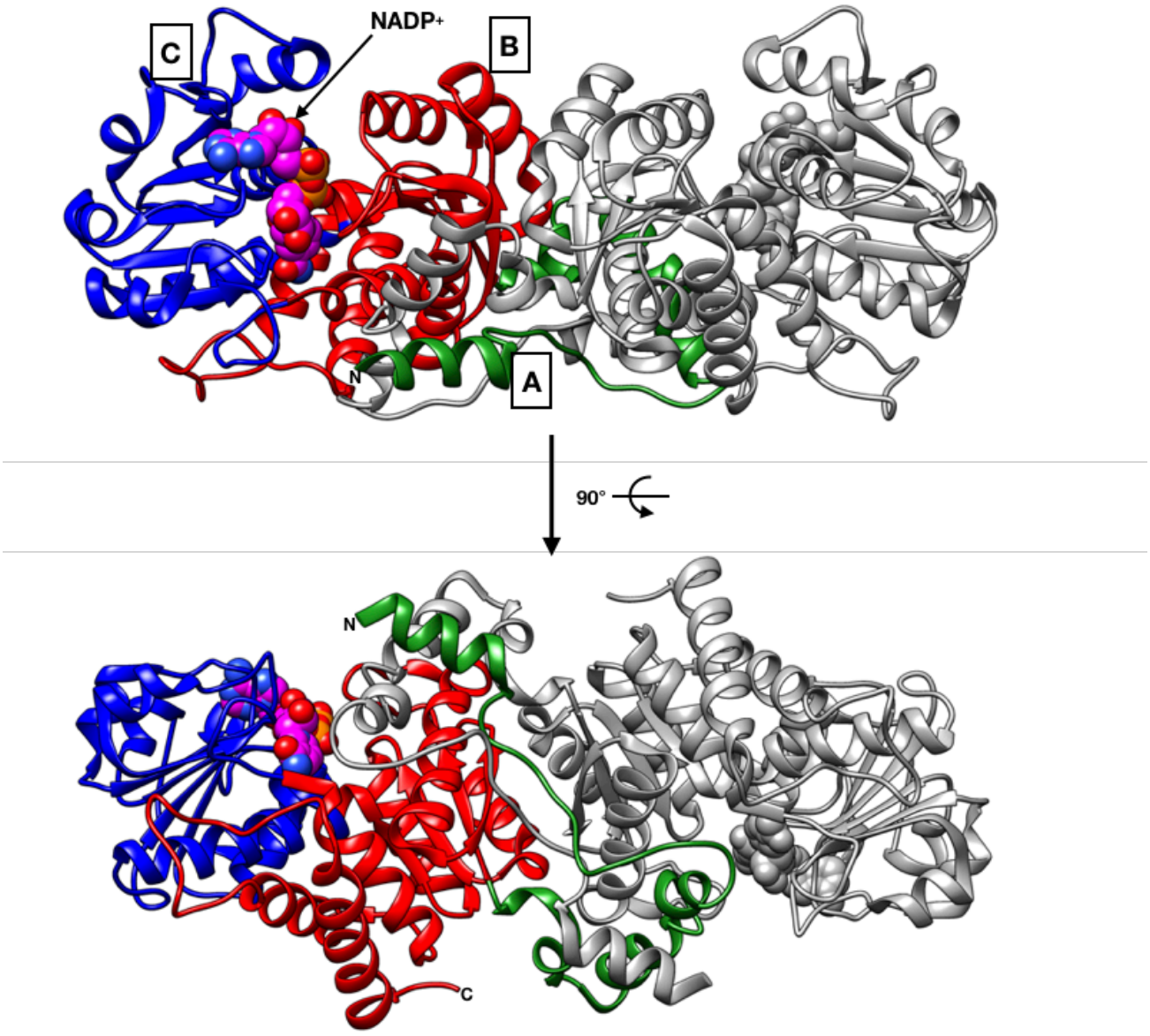


Figure 54. Prokaryotic ME (AYWB-ME)
The structure of AYWB-ME (PDB:5CEE) in two orientations. AYWB-ME has a minimal ME fold. The fold of domain A is only associated with prokaryotic ME and forms the dimer interface (green). Domain B is the core of the ME structure (Red) and domain C is the NAD⁺ binding domain (Blue). AYWB-ME forms a tightly interacting homodimer (monomer1 = coloured, monomer2 = grey) that is stabilised mainly by interactions between residues of domain A. A NAD⁺ ligand is shown bound in substrate binding cleft between domains B and C (magenta) [1].

A number of prokaryotic ME structures have been determined: *Pyrococcus horikoshii* (PDB: 2DVM); *Streptococcus pyogenes* (PDB: 2A9F); *Thermotoga maritima* (PDB: 1VL6 and 2HAE) *Entamoeba histolytica* (PDB: 3nv9) and finally AYWB-ME from *Candidatus Phytoplasma* (PDB: 5CEE) have been deposited to the PDB, whilst only the structure of AYWB-ME has been published [1].

Typically prokaryotic ME structures represent a minimal ME fold composed of two 45kDa monomers that form tightly interacting dimers, rather than a 60kDa tetramer as in eukaryotes [1]. Each monomer folds into three distinct domains: domain A, an N-terminal, alpha helical hook-shaped domain that creates an interlocking dimer interface; domain B, the central core fold containing the cation-binding residues; and domain C, a nucleotide-binding domain containing a Rossmann fold (Figure 54).

Prokaryotic ME completely lack domain D of the eukaryotic ME and has a repurposed domain A, but retains domain B and C. Importantly, the established ME catalytic residues are conserved. However, the catalytic site is composed of residues from both monomers of the dimer. For example in AYWB-ME Tyr36 is located on domain A and inserted into the active site of the adjoining monomer of the dimer complex, adopting the same orientation as in eukaryotic ME [1] (Figure 55). The close dimer interactions and active-site makeup in prokaryotic MEs likely contribute to differential regulation.

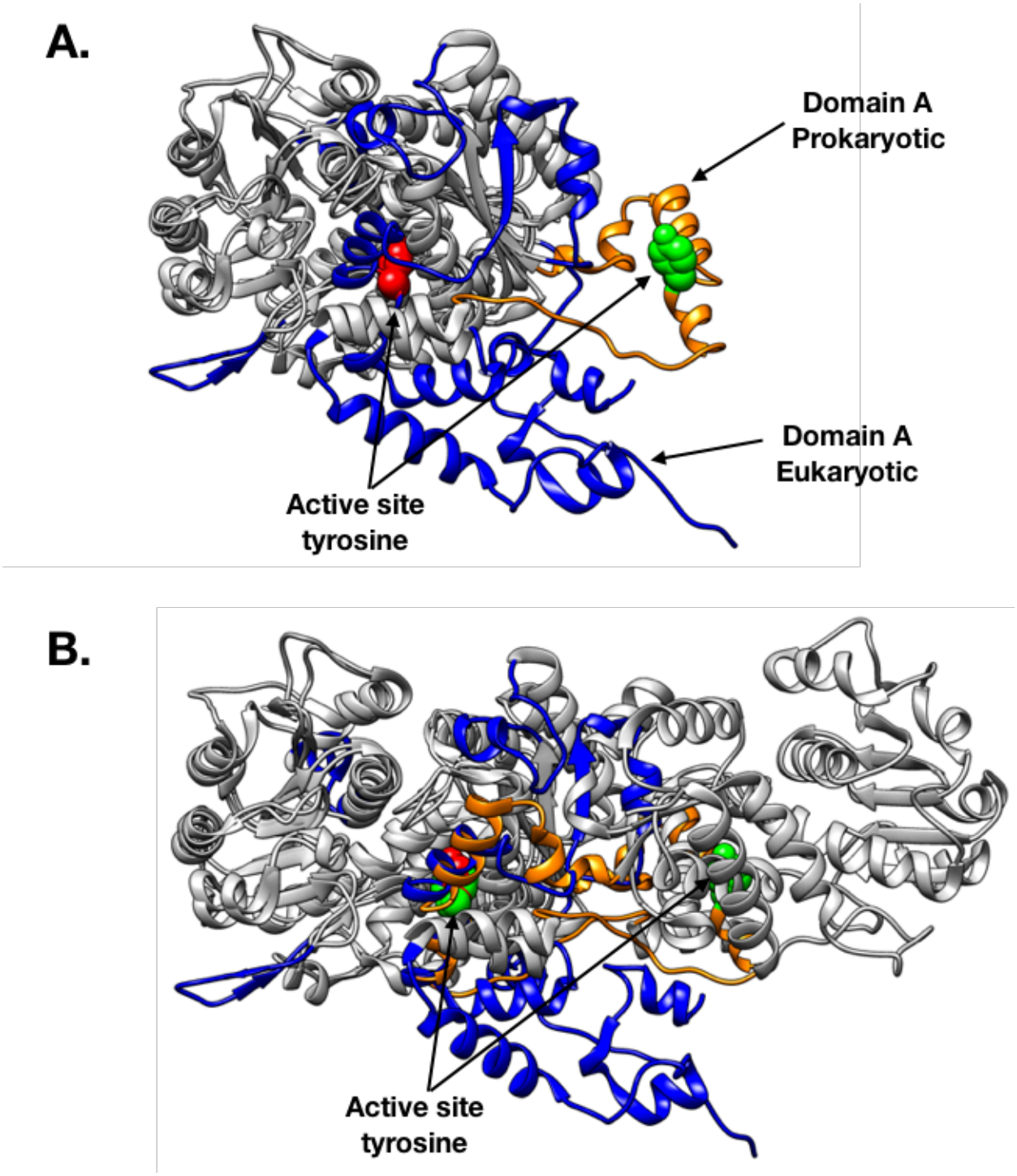


Figure 55. Comparison between prokaryotic and eukaryotic ME
The structures of AYWB-ME (PDB:5CEE) and human-ME (PDB:1QR6) have been superimposed using SSM superpose tool on coot (length = 320 residues, rmsd = 1.97Å). These two structures are typical of prokaryotic and eukaryotic ME structures respectively. **A.** The structural differences between the two structures are highlighted in orange (prokaryotic only) and blue (eukaryotic only). The core fold of the proteins is highly similar between the two MEs therefore has been coloured grey. Domain A has a significantly different fold between prokaryotic and eukaryotic ME. Furthermore the position of the active site tyrosine is located at different position between the two monomers (red = 1QR6, green= 5CEE). **B.** The active site of prokaryotic MEs is composed of residues from both monomers of the dimer complex. When the second monomer of the 5CEE dimer is displayed the location of the active site tyrosine from the other monomer is positioned into the active site (green and red tyrosine residues overlap).

5.1.3 Phosphotransacetylases (Pta)

5.1.3.1 Overview

Phosphotransacetylases (Pta) play an integral role in acetate metabolism in fermentive anaerobes. Pta catalyses the transfer of the acyl group to Coenzyme A (CoA) from acetyl-phosphate (acetyl-P) to form inorganic phosphate (Pi) and acetyl-CoA. Acetyl-CoA is a key central metabolite used directly in anabolic processes, such as the delivery of the acetyl group to oxaloacetate, forming citric acid, which is an important step in the Krebs cycle.

$$\text{Acetyl-CoA} + P_i \rightleftharpoons \text{CoA} + \text{Acetyl-phosphate}$$

Fermentive microbes utilise Pta and acetate kinases (Ack) in a metabolic pathway to generate energy. Ack reversibly catalyse the transfer of an acetyl-group from acetyl-P to ADP to create a molecule of ATP [242].

$$\text{Acetyl-phosphate} + \text{ADP} \rightleftharpoons \text{Acetate} + \text{ATP}$$

Pta enzymes play important roles in the “acetate switch”, a phenomenon where cells transit from a rapid growth state that produces acetate to one of a slower growth state that utilises acetate as an energy source. Acetate is regarded as a major intermediate in the global carbon cycle [243].

Despite the importance of Pta in anaerobic metabolism as well as their widespread distribution, their catalytic mechanism remains poorly understood. J Henkin et al.

performed the first biochemical analysis of Pta back in 1976 [244]. Pta enzymes are believed to bind substrates in a tertiary complex as apposed to the conventional ping-pong mechanism [245, 246]. Furthermore, studies suggest Pta enzymes do not form an acyl-enzyme intermediate during catalysis and a catalytic thiol group is required to perform nucleophilic attack [244].

5.1.3.2 Pta Biochemical Analysis

Extensive biochemical and mutational studies have added to these early studies into the catalytic mechanism of Pta. The majority of published work exists for *Methanosarcina thermophile* Pta (an obligatory anaerobic methano-archaeon) [247]. A number of key Pta residues have been highlighted as catalytically important, including multiple arginine and cysteine residues [248, 249].

Pta activity was shown to be sensitive *N*-ethylmaleimide, a chemical that reacts with thiols and commonly used to modify cysteines. Mutation of Pta cysteine residues indicated C159 is essential for activity. However, a C159S mutation retained full Pta activity suggesting the thiol group was not essential [249]. Furthermore, C312 was established to reside close to the active site but was not required for catalysis [249].

R310, R87 and R133 were all believed to be important for Pta activity. R310 mutants displayed significantly reduced $k_{cat(CoA)}$ and $k_{cat(acetyl\text{-}P)}$, whilst the $K_{M(CoA)}$ was unaltered, the $K_{M(acetyl\text{-}P)}$ was drastically reduced, which demonstrated R310 is important for interacting with acetyl-P during the reaction. R87 and R133 mutants both displayed a drastically reduced K_M and k_{cat} for CoA, which suggested a critical role in CoA binding

during catalysis [249]. Furthermore R87 was established to form a salt bridge with the 3'-phosphate due to lack of sensitivity between CoA and a 3'-dephospho-CoA analogue. Additionally R133 was shown to interact with the 5'-phosphate of CoA [248]. The arginine-modifying agent 2,3-butanedione inactivated Pta, suggesting a reactive arginine is present in the active site. R310 was suggested as the active-site arginine, required to correctly position acetyl-P [248].

5.1.3.3 Pta Structure

The first Pta structure was published in 2004, for the *M. thermophila* Pta (PDB:1QZT) [250]. To date a further four Pta structures have been solved: *Streptoccoccus pyogenes* (PDB:1R5J) [251], *Bacillus subtilis* (PDB:1TD9 + 1XCO) [252] and *Staphylococcus aureus* (PDB:4E4R). All structures have a highly conserved fold and primary sequence. Iyer et al. aligned 32 Pta sequences, highlighting a number of highly and strictly conserved residues (Figure 62) [250].

Pta enzymes are generally ~320aa in length and fold into two distinct domains (domain-I and domain-II). Domain-I includes the N-terminal (residues 2-148) and C-terminal (residues 300-333) regions and forms a strict parallel five-stranded β-sheet that is sandwiched between two α-helix regions, reminiscent of the Rossmann fold. Domain-II (residues 149-299) forms another α-β-α sandwich structure consisting of a mixed six-stranded β-sheet. A distinct substrate-binding groove separates the two domains (Figure 56) [250].

Pta enzymes form a dimer complex, adjoined centrally by interactions between domain-II. The dimer interface is stabilised by two helices that align almost parallel at the centre (Helix-12 from both monomers). Residues from Helix-12 also interact with residues from Helix-9 and Helix-10 of opposing monomers, further stabilising the central core of the dimer [250].

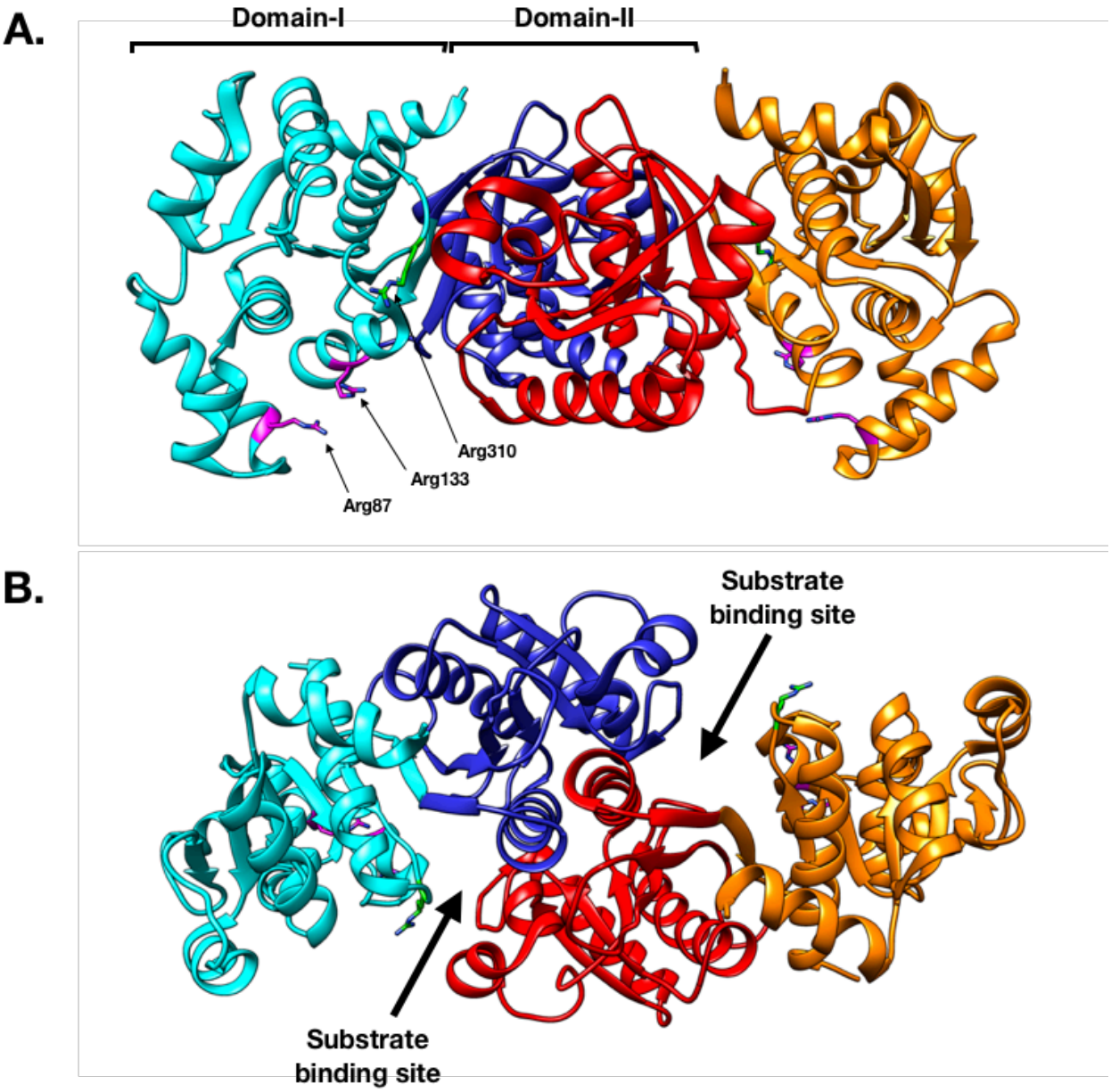


Figure 56. Pta structure
A. Pta enzymes form a bi-lobed homodimer. The dimer is held in place by interactions occurring between residues of the dimerization domain (domain-II – blue and red). The location of the conserved Pta arginine residues believed to be involved in substrate binding are displayed. Arginine residues coloured in magenta interact with CoA, whereas the arginine residue in green interacts with acetyl-P. **B.** A 90° rotation around the x-axis highlights the almost parallel alpha helix that contributes to the dimer interface. Between domain-I (cyan) and domain-II (blue) is a large substrate binding cleft – marked with black arrows.

5.1.3.4 Substrate Binding

The catalytic mechanism of Pta enzymes remains poorly understood, lacking informative enzyme substrate complexes. The structure of Pta has been shown to adopt two conformations, 'open' and 'closed', involving a rigid body rotational conformational change of domain-I relative to domain-II [250]. This conformational change is reported to facilitate catalysis (Figure 59). The Pta from *Bacillus subtilis* (PDB:1XCO) demonstrates multiple acetyl-P binding sites (Figure 57) [252]. Each monomer interacts with four acetyl-P ligands, two bound to the inter-domain cleft and two bound to R88 and R89. The two acetyl-P bound to R88 and R89 are believed to be non-specific (corresponding to sulphates in other Pta structures) and are distal from the predicted active site. The two acetyl-P molecules bound in the inter-domain cleft co-ordinate residues from domain-I or domain-II. Acetyl-P(1) forms hydrogen bonds between its phosphate group and R304 (*M. thermophile* R310 catalytic acetyl-P interacting residue) and S303 sidechains (two strictly conserved residues amongst Pta enzymes) [248]. Additionally the acetyl-group of acetyl-P(1) makes a hydrogen bond with the sidechain of T129 (a highly conserved Pta residue). Acetyl-P(2) interacts more strongly with residues from domain-II and in the reverse orientation to acetyl-P(1), making hydrogen bonds with K251, Y276, K277, N273 and Q238 sidechains (all highly/strictly conserved) [252]. The structure of 1XCO suggests the active site is located between domain-I and II, confirmed by the presence of the acetyl-P binding residue R304.

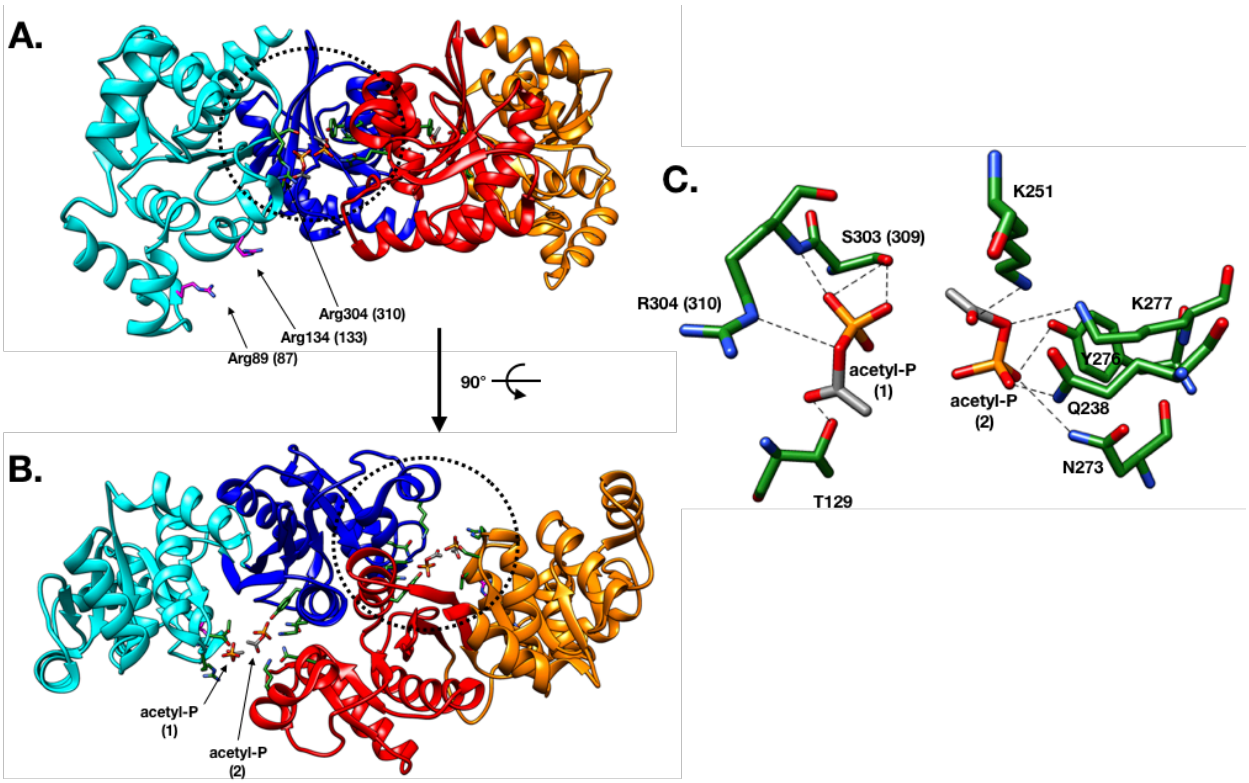


Figure 57. Acetyl-P binding site
A. The acetyl-P binding site is highlighted by a dashed black circle. **B.** The acetyl-binding site viewed from an angle 90° around the x-axis. **C.** A close up view of the residues involved in binding acetyl-P. Two acetyl-P molecules are shown to bind to the active site. The conserved arginine, R304 (R310 in *Methanosarcina thermonphila*) makes a hydrogen bond with acetyl-P(1). (PDB:1XCO)

Two Pta:CoA complex structures for *M. thermophila* Pta (PDB: 2AF3 and 2AF4) have been published [247]. CoA was observed to interact with the Pta enzyme at two distal sites (CoA^1 and CoA^2) in the co-crystallised and soaked structures (Figure 58). One CoA site (CoA^1) was subjectively chosen as the higher affinity binding site (CoA observed at this location in both ligand complex structures) and catalytic residues R310, S309 and D316 were assigned based on their proximity to the bound CoA thiol group [247]. ITC experiments confirmed two CoA binding sites with disparate K_D values of 20 µM and 1mM. The lower affinity binding site is likely non-specific as the cellular concentration of CoA is unlikely to approach concentrations above 1mM [112]. CoA^1 binding site

(higher affinity) makes thirteen hydrogen bonds, including R310 with the 3'-phosphate group of CoA. R310 interaction with the 3'-P group is contradictory to kinetic studies and the structure of 1XCO (acetyl-P complex) that demonstrates R310 is required to interact with acetyl-P during catalysis [248, 252]. CoA^1 is located in the upper portion of the inter-domain cleft with the ADP group located centrally and the pantothene group bent backward to position between S309 and D316 (conserved Pta residues). In the co-crystallised structure the thiol group of CoA has reacted to form a covalent bond with C312 (a residue believed to be close to the active site though not essential for catalysis). CoA^2 (lower affinity site) makes fewer H-bonds but a bidentate salt bridge between R87 and the 3'-P group and electrostatic interactions between R133 and the 5'-phosphate are present [247]. The R87 and R133 interactions with CoA at the CoA^2 site agree with earlier biochemical and mutational studies [248, 249]. The CoA^2 binding site is occupied in both monomers in the soaked structure (2AF3), whereas the CoA^1 site is only occupied in one of the dimers due to a conformational change that restricts the substrate-binding groove.

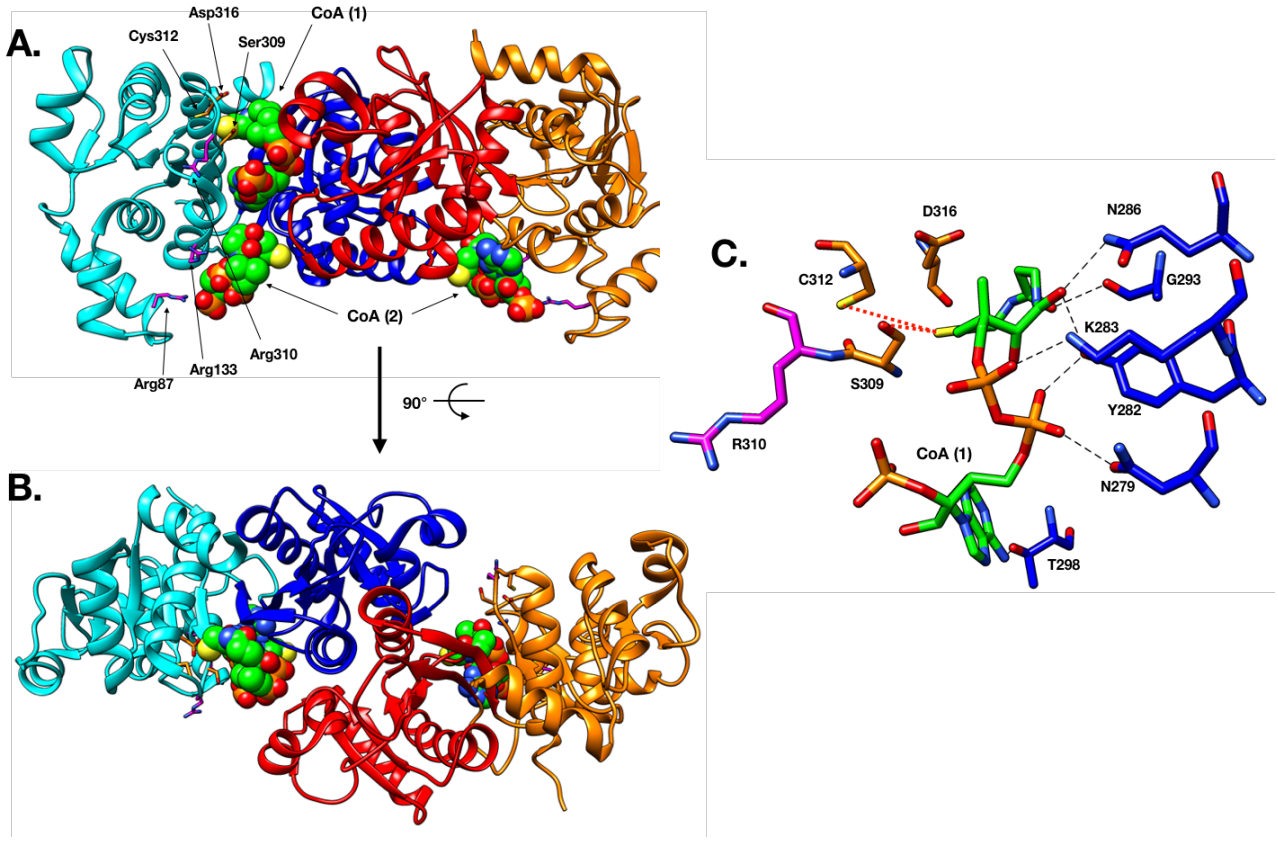


Figure 58. Pta CoA binding site
A. Structure of Pta dimer showing the location of CoA ligands (green) bound in the substrate binding groove. Two CoA ligands are shown to bind to the first monomer and one CoA ligand to the second monomer of the dimer complex. CoA1 and CoA2 sites are indicated. The location of the active-site residues and the conserved arginine residues that are shown to interact with CoA (R67, R133) are also highlighted (magenta). **B.** Rotation of image A by 90° around the x-axis. **C.** A close up view of the residues involved in the binding site of CoA1. Residues in blue are involved in binding the CoA substrate and the H-bond interactions highlighted by black dashed lines. Active-site residues are shown in orange and the interactions (>3.2Å) made with the thiolate group of CoA are shown in red. The conserved R310 that is believed to position acetyl-P is also shown. (PDB:2AF3)

5.1.3.5 Catalytic Mechanism

The currently proposed catalytic mechanism is based on the CoA bound structures. D312 is proposed to act as a general base and abstract the sulfhydryl proton from CoA, forming a negatively charged transition state that is stabilised by S309. The activated transition state then attacks the acetyl-P substrate, which is held in position by R310 [247]. Although the proposed mechanism is plausible it is not in agreement with biochemical and mutational studies. The mechanism does not address how R133 and R87 affect the $K_{M(CoA)}$ if they do not facilitate binding. Nor does this address how R310

binds and orientates acetyl-P if it makes hydrogen bonds to the 3'-phosphate group of CoA or why R310 mutation only displays a decrease in $K_{M(acetyl-P)}$. Furthermore R310 is orientated incorrectly to deliver acetyl-P to the active site (a concern of the authors in Lawrence et al.) [247].

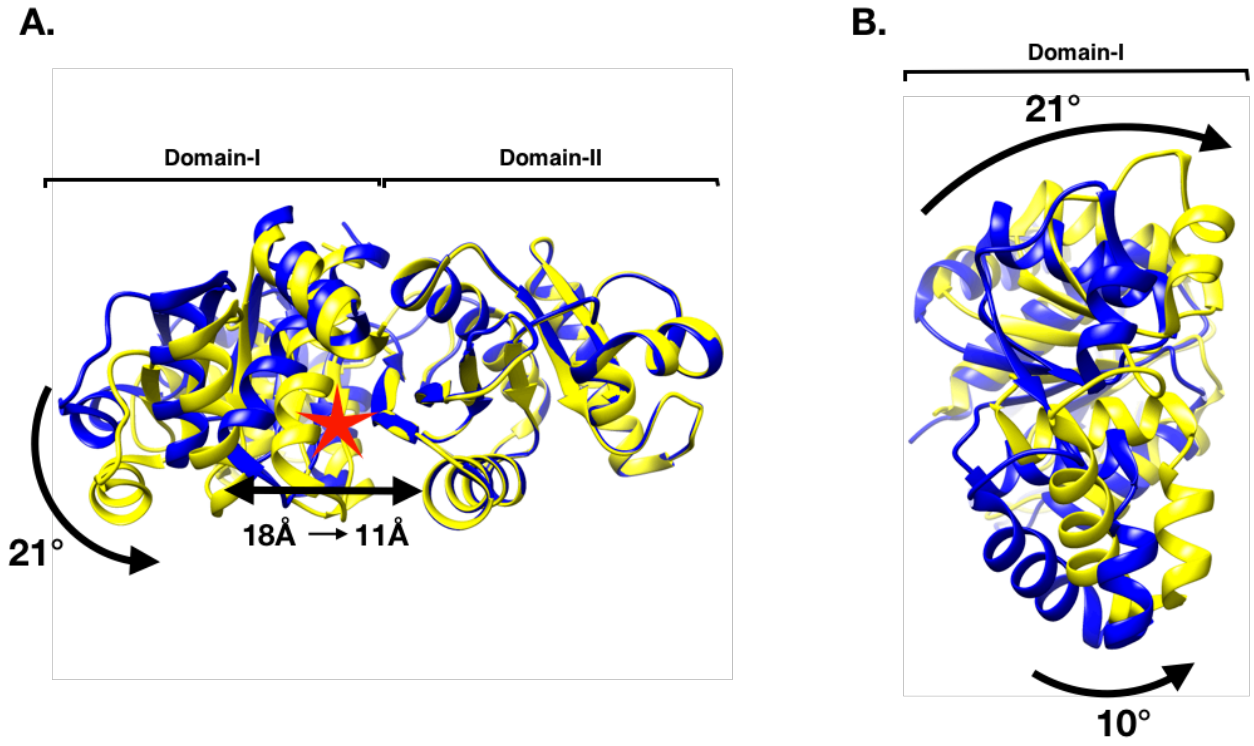


Figure 59. Pta domain movement
A. Domain I adopts multiple conformations in Pta structures, rotating as a rigid body to open and close the substrate binding cleft, located between domain I and II. A 21° swing of domain-I constricts the substrate binding groove from 18Å to 11Å. The movement is believed to be a mechanistic feature of substrate binding. **B.** The same rigid body rotational conformational change in domain-I viewed from the side. The upper portion of the domain undergoes a larger rigid body movement than the lower portion. The LSQ superpose tool in coot was used to align domain II (residues 149-299) of PDB:1QZT (length = 1114 atoms, rmsd = 1.97 Å)

5.1.4 Multi-modular MEs possessing Pta domains

The larger multi-modular ME isoforms found in prokaryotes possess a C-terminal region homologous to Pta enzymes. Multi-modular MEs are widespread in prokaryotes (easily determined by primary sequence) but are mentioned sparingly in the literature. To date three multi-modular ME have been characterised kinetically: DME and TME from *Sinorhizobium meliloti* [253] and MaeB from *Escherichia coli* [254].

MaeB activity is associated with connecting gluconeogenesis to the TCA cycle as part of the phosphoenolpyruvate (PEP)-pyruvate-oxaloacetate node. This process is particularly relevant for organisms that grow on C4 TCA cycle intermediates or substrates that enter the TCA via acetyl-CoA, such as acetate. The role of MaeB is proposed to be in the generation of NADPH and acetyl-CoA, required for anabolic reactions [254].

MaeB is reported to be $NADP^+$ -specific with no detectable activity in the presence of an NAD^+ cofactor. Additionally MaeB was shown to require the use of a divalent cation, such as Mg^{2+} or Mn^{2+} similarly to other eukaryotic ME isoforms. MaeB activity favours the pyruvate forming direction (turnover 200 times quicker). MaeB activity is regulated by multiple effectors such as glutamate, aspartate, glucose-6P, and acetyl-P (which are activators) and inactivated by fumarate, OAA, acetyl-CoA [254]. Furthermore high concentrations of L-malate also inhibit the activity of MaeB.

Truncation mutants of MaeB (Pta domain truncated) revealed that many of the activators/inhibitors were allosteric effectors that interacted with the Pta domain. This

is interesting given that the allosteric inhibitor acetyl-CoA is a substrate/product of Pta enzymes. ME truncations retained activity, albeit reduced compared to the full-length enzyme. Moreover the Pta domain was also shown to be involved in oligomerisation to form catalytically active octomer units [254], suggesting higher order oligomers are required for efficient catalysis in multi-modular ME. No Pta enzyme activity was detected for MaeB similarly to other multi-modular MEs [255]. This suggests that the Pta domain in multi-modular ME isoforms is non-catalytic but important for allostery and oligomer formation.

Rhizobium meliloti, a nitrogen fixing Gram-negative bacterium, has two multi-modular ME isoforms, DME and TME. DME and TME were shown to be $NAD(P)^+$ and $NADP^+$ -dependent, respectively [256]. DME plays an important role in symbiotic nitrogen fixation as a result of growth in the nodule where C4 di-carboxylates are a major carbon and energy source [257]. TME on the other hand was suggested to function as a generator of NADPH like other $NADP^+$ -specific malic enzymes [255].

Multiple metabolic effectors have been shown to affect the activity of DME including the potent allosteric effector acetyl-CoA. In addition DME was also inhibited by other CoA derivatives such as butyryl-CoA and propionyl-CoA. Besides these, OAA and NADH also negatively affected DME activity. The TCA cycle intermediates Fumarate and Succinate positively regulated DME activity [256]. The strong inhibitory effects displayed by acetyl-CoA and OAA suggests that they are end products of the metabolic pathway. On the other hand TME had no detectable sensitivity towards acetyl-CoA, CoA or any other metabolic effector and as a result the cellular role of TME is less clear [256].

Both DME and TME C-terminal truncations revealed the Pta domain plays a role in oligomerisation. Size exclusion demonstrated that DME and TME both form homo-octomers in solution, whereas C-terminal truncation mutants formed homo-dimers. The ΔDME and ΔTME homo-dimers displayed decreased ME activity compared to full-length, in keeping with MaeB [253]. Consequently it is believed that the Pta domain of DME and TME facilitates oligomerisation, which impacts on enzyme activity.

It is interesting to speculate that larger multi-modular ME isoforms have adopted a C-terminal extension with a Pta domain fold in order to allosterically respond to Pta enzyme substrates. Because prokaryotes lack the ability to compartmentalise ME isoforms like eukaryotes additional regulatory machinery have been incorporated into the enzymes. The structure and function underpinning multi-modular ME activity remains elusive.

5.1.5 Aims and objectives

This study aims to structurally and functionally characterise the multi-modular ME, Bd1833 and determine the mechanism behind allosteric inhibition by acetyl-CoA. Extensive structural investigations alongside in depth mutational and enzymatic studies were used to confirm the combination of two unrelated enzymes and reveal how the catalytically inactive Pta domain acts as a sensor of acetyl-CoA to regulate the activity of the ME domain.

5.2 Results

5.2.1 Bd1833 homologues

Bd1833 is multi-modular ME that is composed of an N-terminal ME domain and a C-terminal Pta domain. Bd1833 is homologous to other multi-modular MEs that have been previously characterised, such as: MaeB (sequence identity = 49.5%), DME (45.6%) and TME (48.5%). Figure 60 displays the sequence alignment of these proteins, which reveals that sequence conservation at the ME domain is far greater than in the Pta. This is highlighted by the sequence identity scores for separate ME and Pta domains of Bd1833 and MaeB. The ME are 61% identical, whereas the Pta domains display 35% sequence identity.

Alignment of the ME domain with other sequences from prokaryotic MEs whose structures have been determined reveals that the sequences are highly conserved (identity between 44 and 48%). However, there is a 17aa region that is unique for Bd1833 (Figure 61). This insertion is also present in other multi-modular ME, suggesting its insertion is relevant to the function of multi-modular MEs in particular.

Alignment of the Bd1833 Pta domain with other characterised Pta domains (Figure 62) reveals that many of the key catalytic residues are missing. However, all of residues that have been implicated in substrate binding (R87, R133, R310 *Methanosarcina thermophile* nomenclature) are present. Bd1833 Pta has sequence identity of 32.7% with 1QZT, 35.2% with 1R5J and 34.4% with 1XCO.

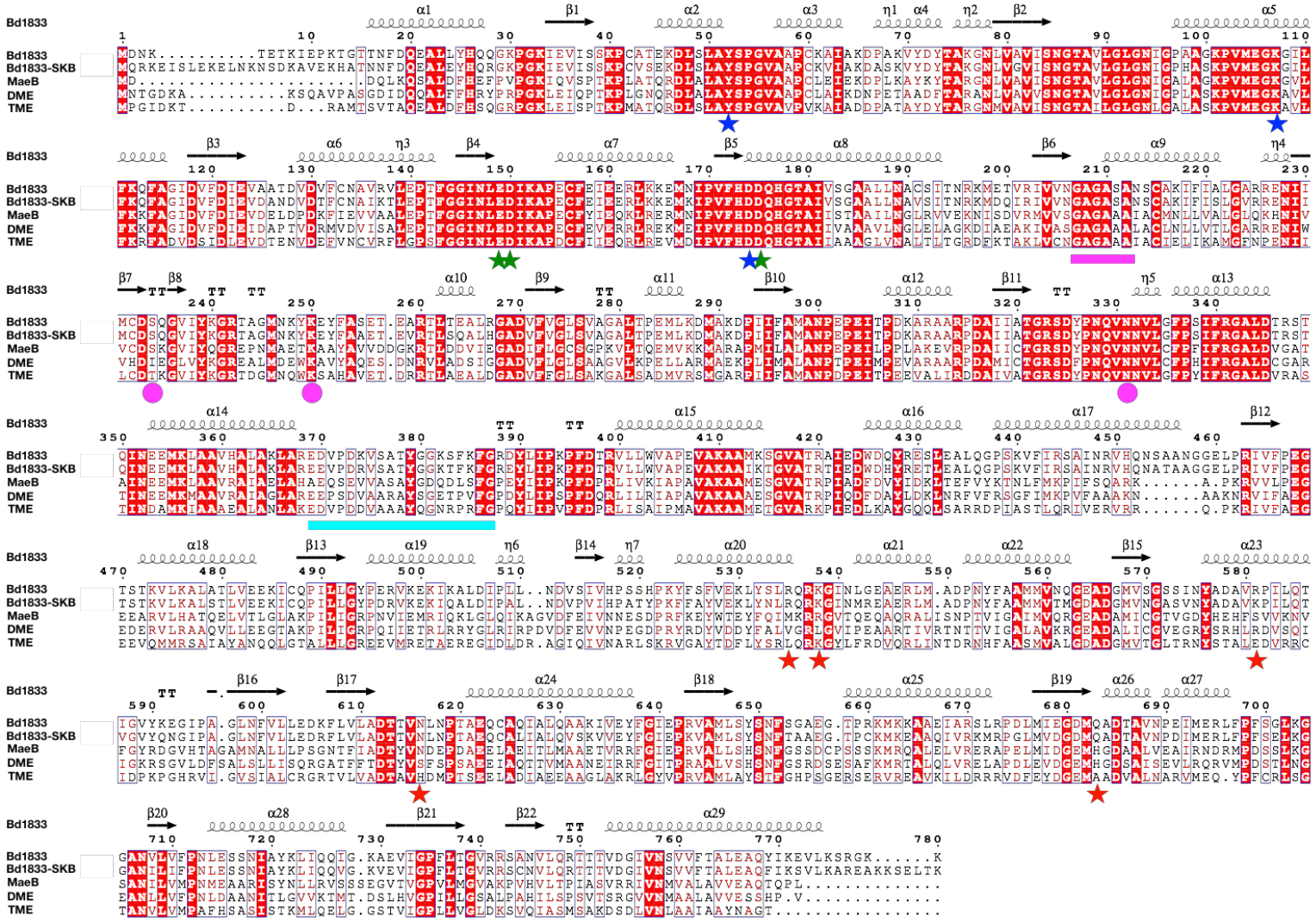


Figure 60. Bd1833 alignment with multi-modular malic enzymes

Catalytic residues are highlighted with blue stars. Cation-binding residues highlighted with green stars. Residues that hydrogen bond with $NADP^+$ are highlighted with pink dots and the GXGXXG motif with a pink bar. Residues with side chains that hydrogen bond with acetyl-CoA are highlighted with red stars. The cyan bar shows the location of subdomain D (multi-modular ME specific loop)

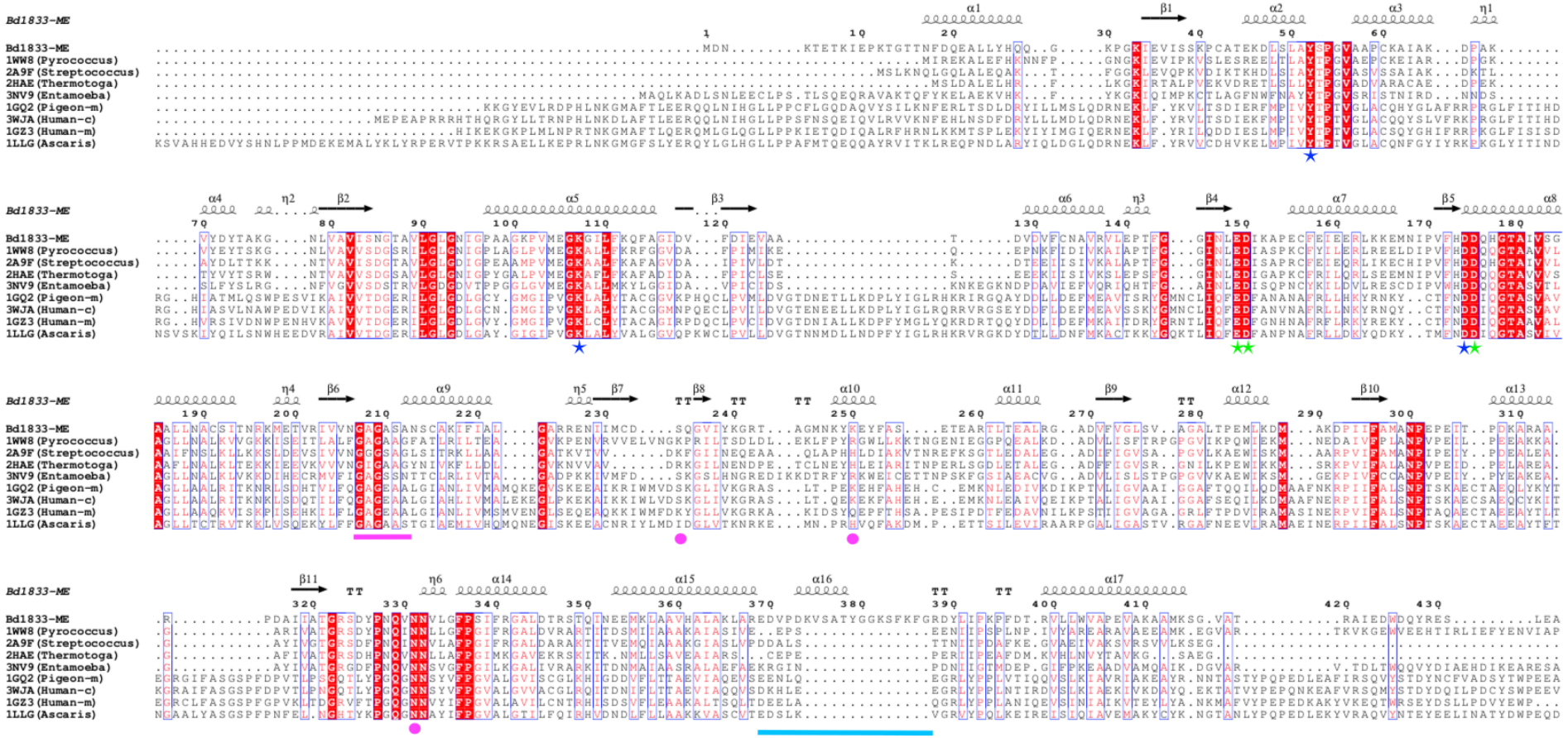


Figure 61. Bd1833-ME alignment with prokaryotic and eukaryotic ME
Catalytic residues are highlighted with blue stars. Cation-binding residues highlighted with green stars. Residues that hydrogen bond with NADP⁺ are highlighted with pink dots and the GXGXXG motif with a pink bar. The cyan bar highlights subdomain-D (specific for multi-modular malic enzymes).

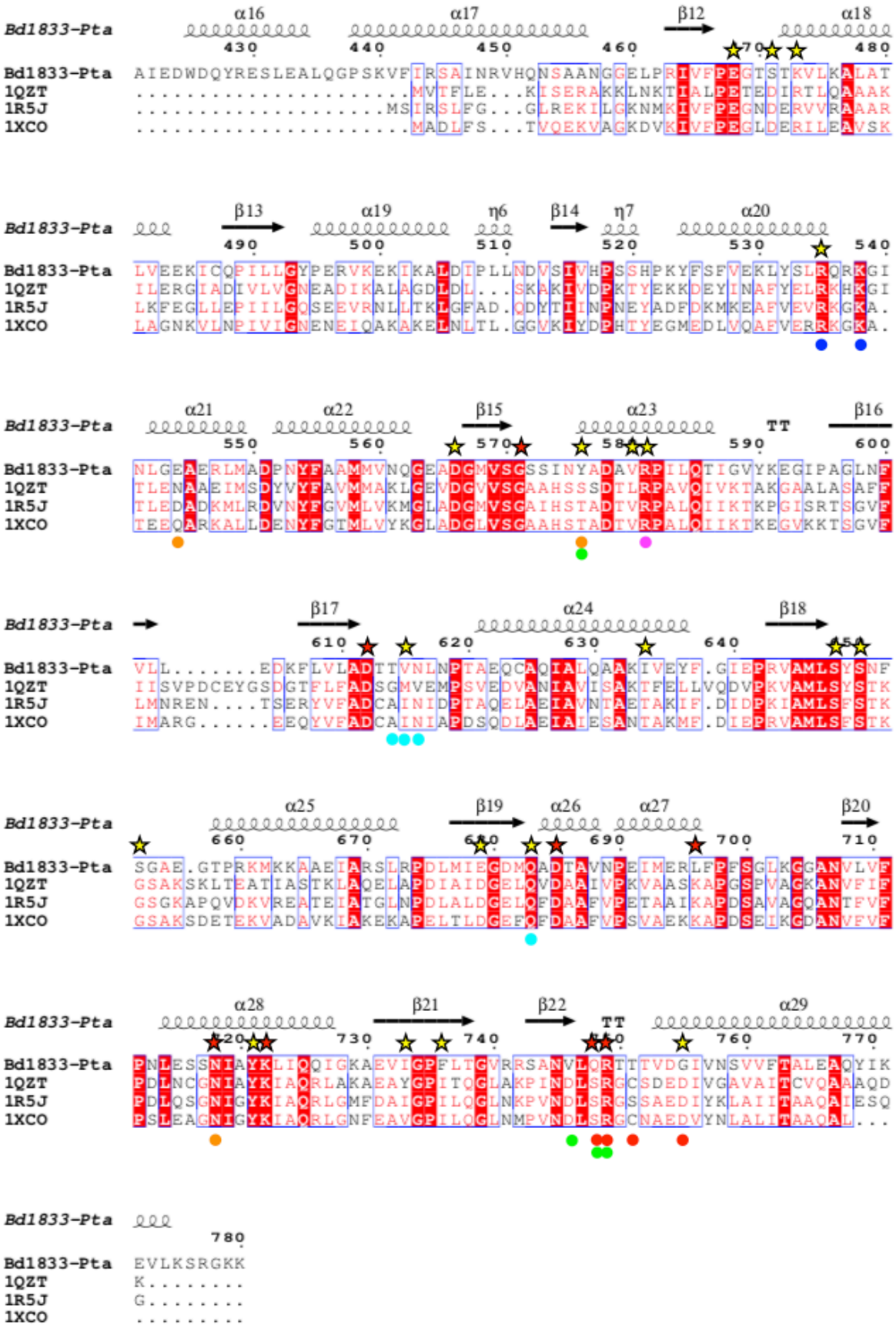


Figure 62. Bd1833-Pta alignment with Pta enzymes
The highly conserved and strictly conserved residues associated with Pta enzymes are highlighted by the yellow and red stars respectively. Dark blue circle – residues interacting with CoA 3'phosphate group; pink circle – residues interacting with CoA pyrophosphate group; light blue circle – other residues hydrogen bonding to CoA; orange circle – key additional residues highlighted from Bd1833 studies. Red circles are the previously proposed catalytic residues and the green circles are the newly proposed catalytic residues.

5.2.2 Bd1833 purification and crystallisation

Three constructs were designed for Bd1833: full-length (Bd1833), ME domain only (Bd1833-ME) and Pta domain only (Bd1833-Pta). Each construct of Bd1833 was successfully cloned, purified, crystallised and the structures solved. Figure 63 shows a representative purification gel and crystals, which were used to solve each construct.

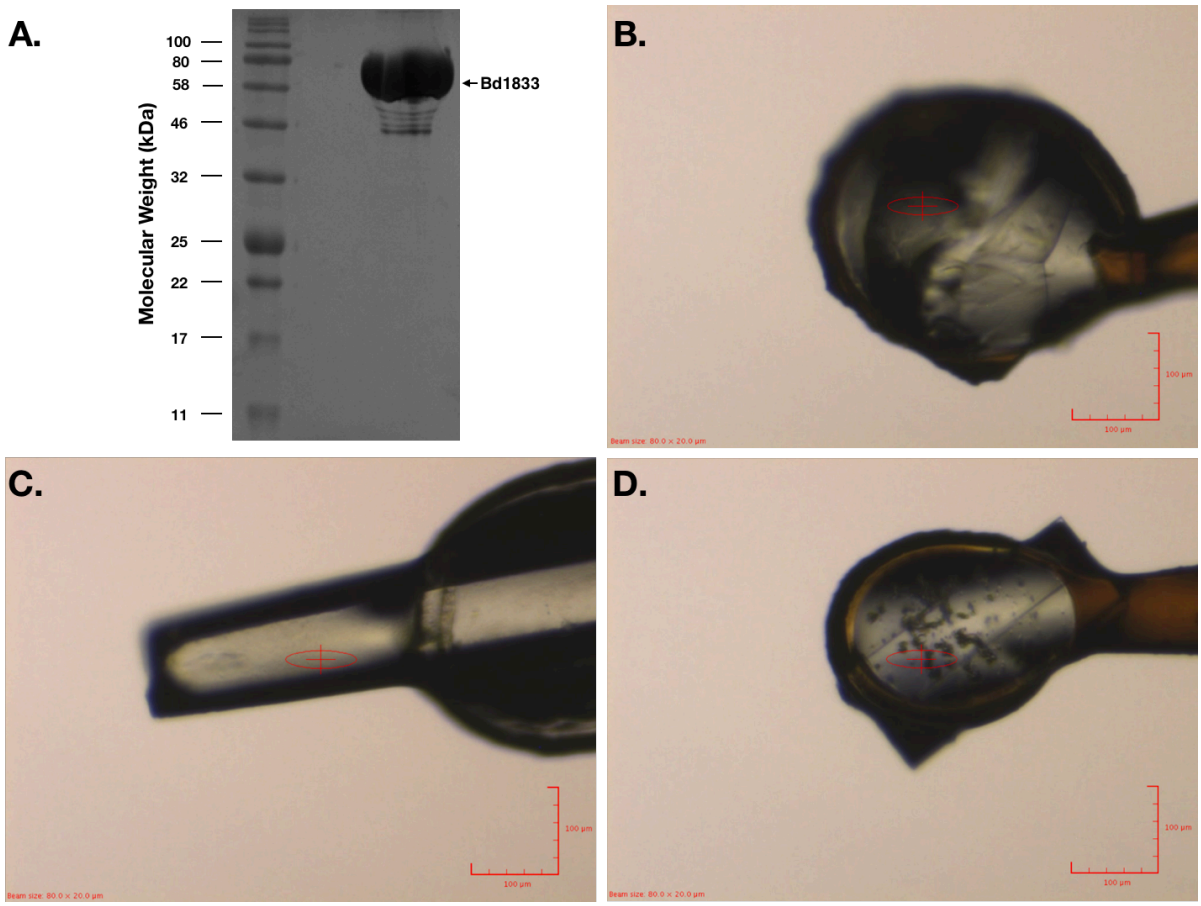


Figure 63. Bd1833 Purification and Crystallisation
A. SDS PAGE gel showing purified full-length Bd1833 **B.** Crystal of Bd1833 used for data collection **C.** Crystal of Bd1833-Pta used for data collection. **D.** Crystal of Bd1833-ME used for data collection

5.2.3 Bd1833 Structure

We have solved the three-dimensional structure of full-length Bd1833, its constituent domains individually, mutant forms and ligand complexes to aid in the determination of function. A list of Bd1833 structures solved and an associated brief description is presented in Table 13, which is followed by more in depth structural analysis in the following individual sections.

	Construct	Amino acids traced	Spacegroup	Resolution (Å)	Copies in asymmetic unit	R-work	R-free	Oligomeric state	Ligand?	Description/Rationale
Bd1833-ME (Bcs-B6)	1- 434	15-422	P 2_1	1.67	2	0.1702	0.2039	Dimer	$Mg2^+$	High resolution structure of ME domain
Bd1833-ME (Pact-E1)	1- 434	16-427	P 1	2.2	6	0.2109	0.2264	Dimer	-	Lacking divalent cation
Bd1833-ME ($NADP^+$)	1- 434	15-425	P 2	1.67	2	0.1799	0.2159	Dimer	$Mg2^+$ $NADP^+$	NADP+ complex
Bd1833-Pta	439- 780	439- 776	P 2_1	1.96	6	0.1829	0.2052	Hexamer	acetyl-CoA	High resolution structure of Pta domain, inhibited state
Pta R535A	439- 780	439- 776	C 2	2.35	6	0.2158	0.2368	Hexamer	-	Ligand free structure, 3'-P mutant
Pta R535E	439- 780	439- 776	C 2	2.39	6	0.2153	0.2602	Hexamer	-	Ligand free structure, 3'-P mutant
Pta E544R	439- 780	439- 776	C 2	2.33	6	0.2069	0.2265	Hexamer	acetyl-CoA	Interupt hexamer interface
Pta Y576S	439- 780	439- 776	C 2	2.38	6	0.2222	0.2725	Hexamer	-	acetyl-CoA binding pocket distrupted
Pta-Apo	439-780	439-776	P 2_1 2_1 2_1	2.72	12	0.2520	0.2995	Hexamer	-	WT Ligand free structure that shows rigid body conformational change
Bd1833 (Full-length)	1-780	15-776	C 2	2.72	6	0.1922	0.2277	Hexamer	acetyl-CoA	Full length protein
Bd1833-SKB (full-length)	1-796	16-776	P 4_1 2_1 2	3.71	3	0.2503	0.2996	Hexamer	-	Full length of closely related homologue
Bd1833-SKB (Pta)	449-796	449-786	F 2 3	2.71	4	0.2040	0.2392	Hexamer	acetyl-CoA	Homologue Pta domain, CoA binding site differs

Table 13. Bd1833 structures
This table displays a brief structural analysis of each Bd1833 structure reported in this thesis and the rationale behind the structure.

5.2.4 The structure of Bd1833 ME domain

5.2.4.1 ME domain structure

The structure of Bd1833-ME closely resembles that of the expected small prokaryotic ME fold. Bd1833-ME forms a tightly interacting dimer that can be divided into four functional subdomains, displayed in Figure 64. Subdomain-A, residues 16- 78 (green), is composed of five α-helices and a single β-strand that are organised to resemble a hook-like structure. Two protomers of subdomain-A from the dimer interface. The single β-strands from subdomain-A interact to form a two-stranded β-sheet at the centre of the dimer interface. PISA calculates the total surface area of the dimer complex to be 19508Å^2 and the dimer interface to be 4135Å^2. Subdomain-B, comprised of residues 79-173, 332-368 and 387-420, forms the central core of the protein. Subdomain-B is composed of a central four-stranded β-sheet sandwiched between two α-helical regions on each side and two additional α-helices below. Subdomain-C, formed from residues 174-331 contains the nucleotide-binding site and resembles a Rossmann fold, formed of a five-stranded β-sheet sandwiched between three α-helices on each face. Finally Bd1833-ME has a small loop region located as a relative insertion (c.f other ME) in subdomain-B. This 17aa loop region has a single short α-helix at the tip and has no structural equivalent in other ME folds. Furthermore, the loop region is easily discerned as an insert in primary sequence (Figure 61). As such the loop region is described as subdomain-D, a multi-modular ME-specific feature.

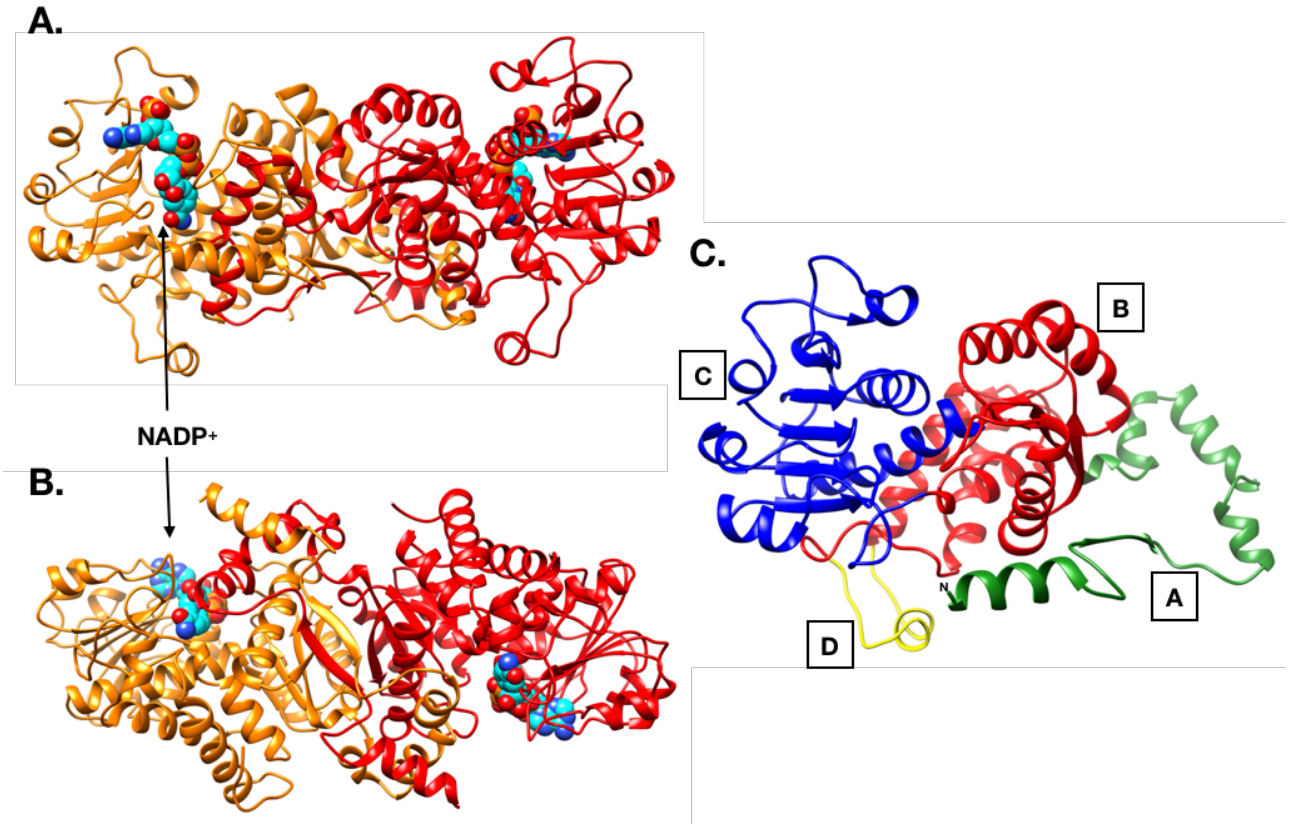


Figure 64. Bd1833-ME structure
A. Bd1833-ME forms a biological dimer (orange and red monomers). The active site is located in the groove between subdomains B and C, which is highlighted by the $NADP^+$ molecule (cyan) bound to subdomain-C. **B.** Illustrates the tightly interacting dimer interface, viewed from a 90° rotation of A around the x-axis. **C.** Illustrates the subdomains that make up Bd1833-ME. Subdomain-A, providing the dimer interface, residues 16-78 (green). Subdomain-B, residues 79-173, 332-368 and 387-420 (red) form the core of the protein containing the divalent cation-binding site. Subdomain-C, residues 174-331 (blue) forms the $NADP^+$ -binding site possessing a Rossmann fold. Bd1833-ME has a small loop specific to multi-modular malic enzymes, subdomain D 369-386 (yellow).

5.2.4.2 ME catalytic site

The catalytic site of Bd1833-ME is positioned between subdomains-B and C, which is highlighted by the position of $NADP^+$ in Figure 64. The catalytic pocket of Bd1833 is composed of a divalent cation-binding motif, a nucleotide-binding site and the substrate-binding site. The divalent cation-binding motif is central to the catalytic site and six oxygen atoms coordinate Mg^{2+} in an octahedral fashion (Figure 65). The coordinating oxygen atoms are provided by the sidechain carboxylate groups of E143

and D175, two water molecules in the equatorial plane, the sidechain carboxylate group of E150 and an additional water molecule in the axial plane.

Bd1833-ME has a nucleotide-binding motif (207-GAGASA-212) that is located in subdomain-C and provides the floor of the adenine-binding groove similarly to other nucleotide-binding proteins possessing a Rossmann fold. The Bd1833-ME:$NADP^+$ complex reveals the interactions $NADP^+$ makes with Bd1833-ME subdomain-C (Figure 65). The adenine group of $NADP^+$ sits in the upper portion of subdomain-C, slotting vertically between a coil and a β-strand region. The adenine-binding groove is capped by N206. The adenine group is held in place solely through non-bonding interactions with residues of the adenine-binding groove. The 2'-phosphate group makes hydrogen bonds to the sidechains of S234 and K250. $NADP^+$ -specificity is likely contributed by the interaction K250 makes with the 2'-phosphate group [230]. The pyrophosphate portion sits on top of the nucleotide-binding motif making numerous H-bond interactions. The nicotinamide portion of $NADP^+$ is positioned <7Å away from the bound cation, leaving space for the substrate to enter. The nicotinamide group is bound by two hydrogen bonds with the backbone carboxyl group of V330 and the sidechain of N300.

Bd1833-ME has a conserved set of ME catalytic residues, including: E149, the metal ion coordinating proton donor; D174, that hydrogen bonds to the catalytic lysine; K107, the general acid and base during catalysis; and Y52 the proton donor in the final tautomerisation step. The catalytic residues of Bd1833-ME adopt similar orientations to their counterparts in the structure of human mitochondrial NAD(P)-ME open form [220]. A key difference in the catalytic residue makeup is the presence of catalytic

residues from both monomers of the dimer complex. The catalytic residue, Y52, located on subdomain-A, is provided by the other monomer. A single catalytic site containing residues from both monomers is typical of prokaryotic MEs [1] but not of eukaryotic ME. Moreover the proposed substrate-binding site residues are also slightly different to eukaryotic MEs, for example lacking the active-site arginine that assists in substrate binding [225].

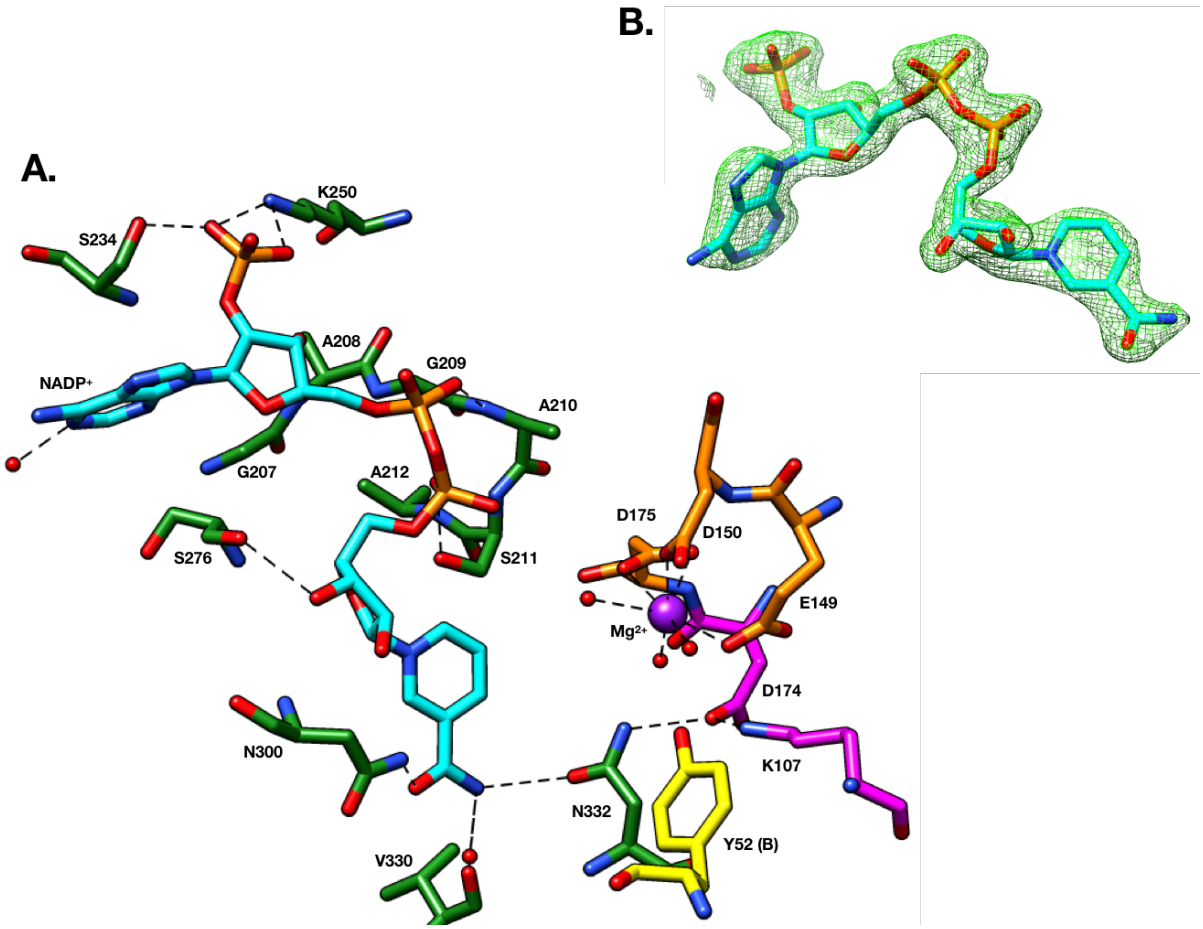


Figure 65. Bd1833 active site
A. The active site of Bd1833-ME is made up of residues from both monomers. The $NADP^+$ -binding groove located between subdomains B and C. Subdomain-C contains the Rossmann fold and GXGXXG nucleotide-binding motif. The residues that make hydrogen bond contacts with $NADP^+$ (cyan) are shown in green. The active-site residues are shown with magenta coloured carbons, while Y52 provided by the other monomer is shown with yellow carbons. The metal coordinating residues are shown in orange, bound to the divalent cation (Mg^{2+}). All residues shown, except for Y52, belong to a single monomer (chain A). All hydrogen bonds illustrated by dashed black lines. **B.** The electron density observed for $NADP^+$ in the Bd1833-ME: $NADP^+$ soaked crystal structure. ($2F_o$-F_c map, radius =2Å, σ = 1)

5.2.4.3 Changes between bound states

$NADP^+$ binding does not cause large conformational changes in the structure of Bd1833-ME, although small changes in the orientation of local residues are observed. N300 repositions to face into the active site and interact with the nicotinamide group, and S234 changes conformation to interact with the 2'-phosphate group. This finding is in agreement with previous findings that demonstrate conformational rearrangement of ME domains is only observed when all substrates and cofactors are bound [225].

5.2.5 The structure of Bd1833 Pta domain

5.2.5.1 Pta domain structure

We solved the three-dimensional structure of Bd1833-Pta, which co-purified/co-crystallised with acetyl-CoA. The secondary structure of Bd1833-Pta is highly similar to other published Pta enzymes and folds into two distinct subdomains: subdomain-I and subdomain-II. Bd1833-Pta assembles into a large hexamer complex composed of three homo-dimers, arranged into an offset triangular shape (Figure 66). The dimers are arranged similarly to other Pta enzymes, whilst the hexamer complex is unique.

Pta secondary structure consists of a mixture of α-helices and β-strand regions folding into two distinct subdomains. Subdomain-I comprises the N and C-terminal regions (residues 441- 595 and 738-776) of the Pta domain, which fold into α-β-α region composed of a five-stranded parallel β-sheet sandwiched between four α-helices on each side. Two α-helices protrude from the core of subdomain-I, forming a flap that contains the conserved CoA binding residue, R535 (R87 *Methanosarcina thermophile*). This region also contributes to the hexamer interface. Subdomain-II (residues 596 and

737) folds into a six-stranded mixed β-sheet that is sandwiched between two α-helices on either side. Subdomain-II is positioned centrally in the Pta dimer arrangement, forming the dimer interface. A shallow groove splits subdomain-I and II and is the location of the substrate-binding pocket where the co-purified and co-crystallised acetyl-CoA is located (Figure 67).

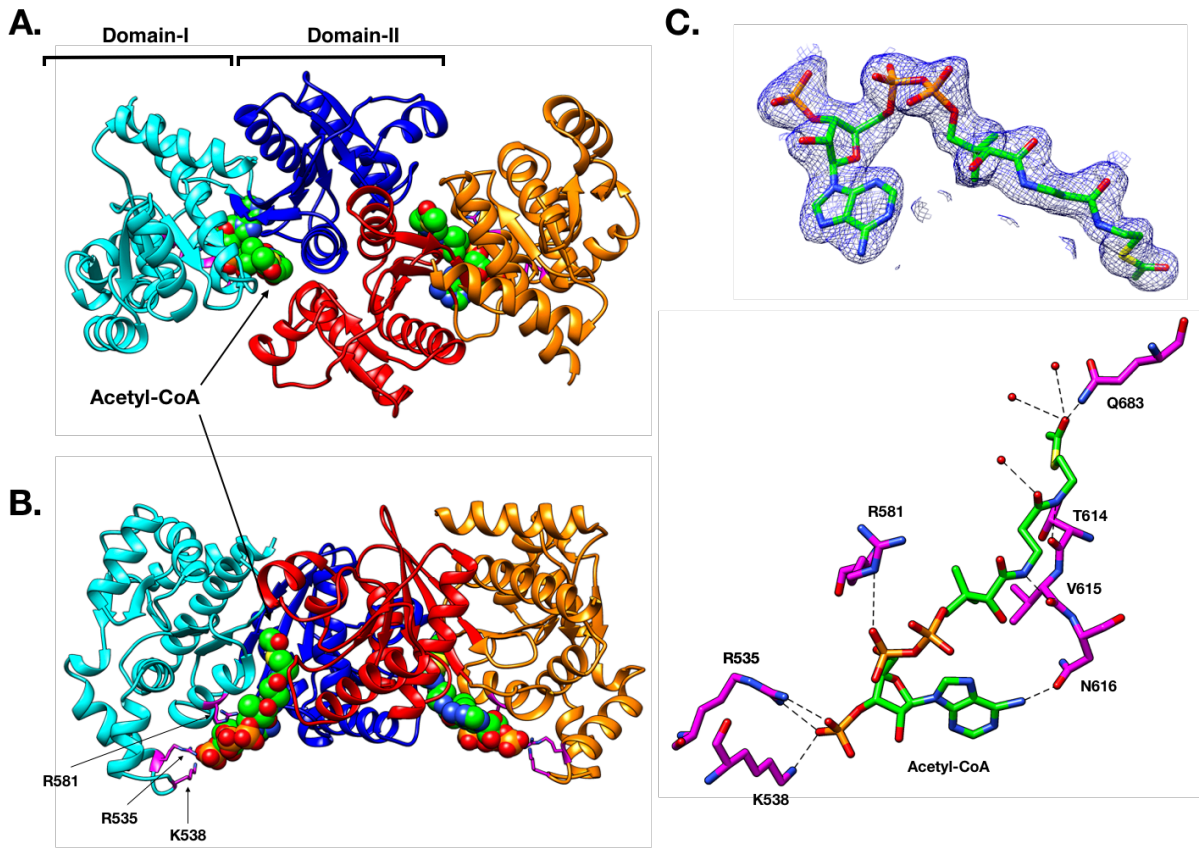


Figure 67. Bd1833-Pta dimer and acetyl-CoA binding
A. Bd1833-Pta forms a hexamer (trimer of dimers) where the dimers are structurally similarly to other previously solved Pta structures represented here as a blue:cyan and a red:orange dimer. Bd1833-Pta monomer can be split into two subdomains: subdomain-I, residues 440-595 and 734-776 (cyan) and the central subdomain-II, 596-733 residues (blue). Central to the dimer is an almost parallel alpha helix that is the main contributor to the dimer interface. Acetyl-CoA (green) binds in the substrate-binding groove located between subdomains-I and II. The conserved CoA binding residues (Pta enzymes) are highlighted in magenta making hydrogen bonds to acetyl-CoA. **B.** Rotation of image A 90° around the x-axis. **C.** The electron density observed for acetyl-CoA ($2F_o$-F_c map, radius=2Å, σ=1). **D.** Illustration of the hydrogen-bonding network associated with acetyl-CoA binding.

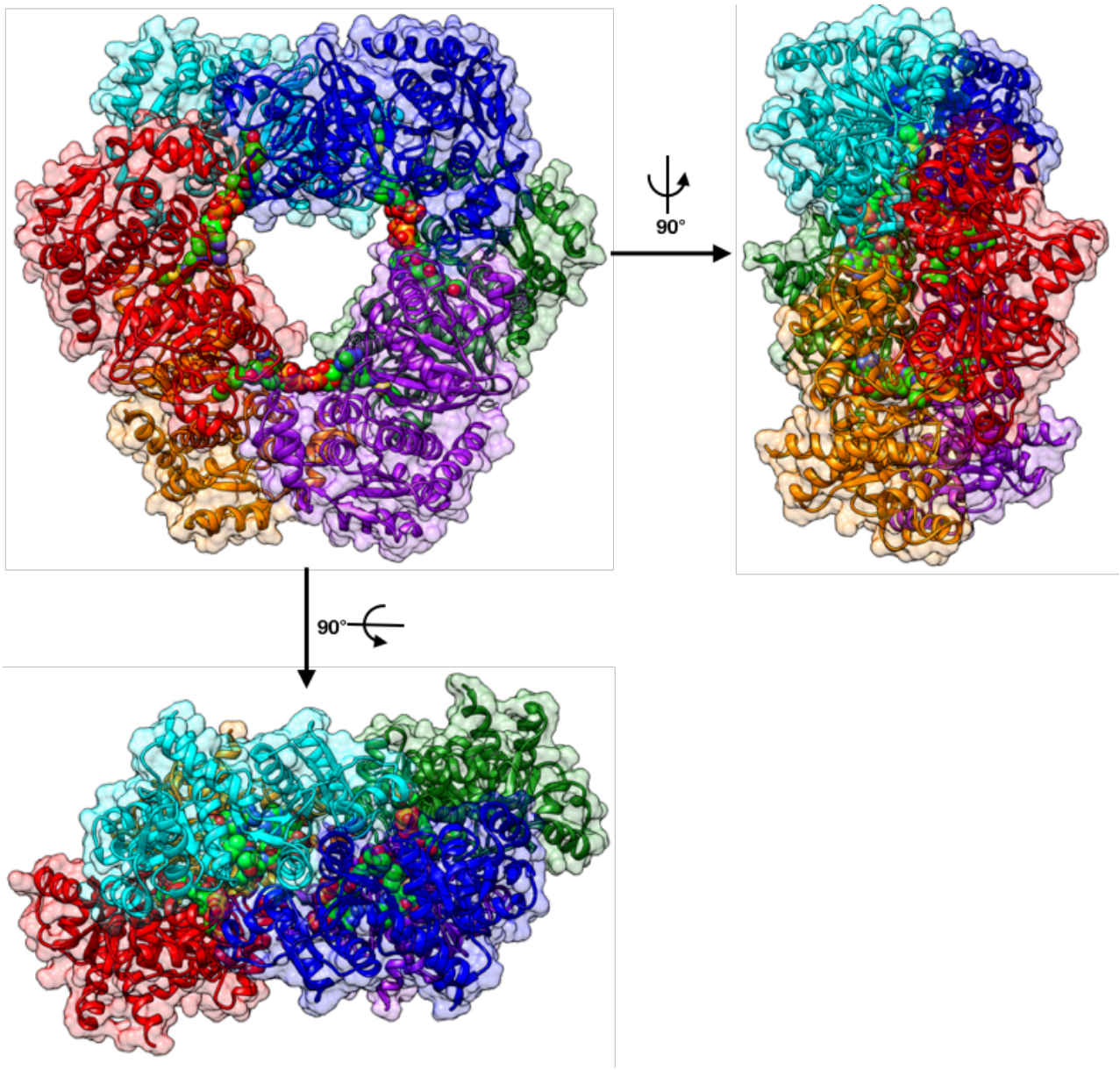


Figure 66. Bd1833-Pta structure
Bd1833-Pta forms a unique hexamer complex made of a trimer of dimers. The hexamer is organised into an offset triangle shape where the dimers form the flat sides. Dimers are coloured – blue:cyan, red:orange and green:purple. Each monomer makes contacts to four other monomers. The hexamer complex has six acetyl-CoA ligands bound, located on the inside face at the corners of the triangle formation. The hexamer complex is displayed in three orientations, rotated by 90°.

5.2.5.2 Dimer and tetramer interfaces

The dimer interface, 1707Å^2 buried surface area, is composed mainly of residues from subdomain-II. In particular two parallel α-helices (713-730) are positioned centrally to the dimer interface. Each monomer of the dimer provides a single α-helix to form the

parallel α-helix pair that makes substantial hydrophobic interactions to stabilise the dimer interface. In addition H-bonds further stabilise the dimer interface (Figure 68).

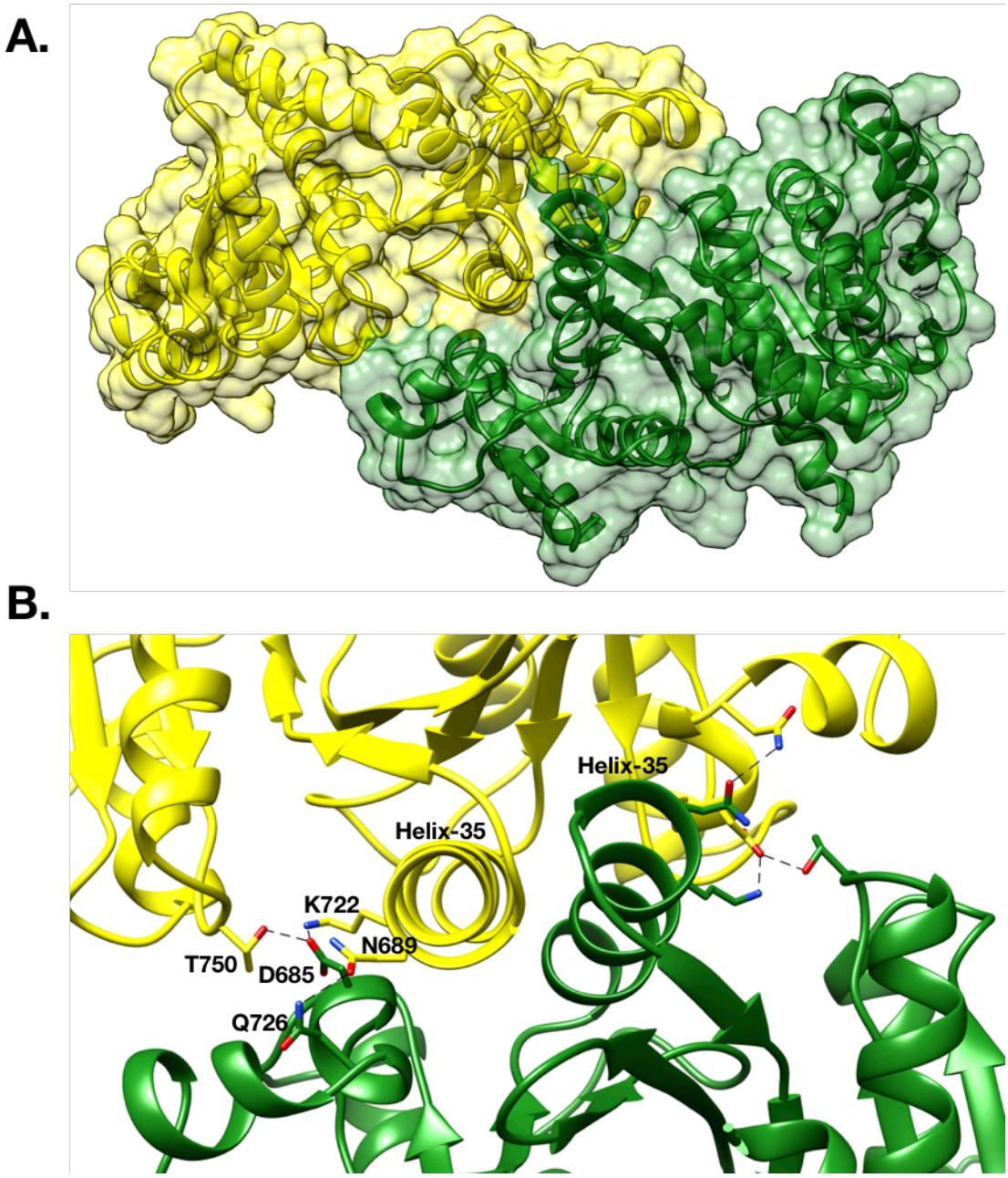


Figure 68. The dimer interface
A. Surface view of dimer interface. **B.** View of dimer interactions. The dimer interface is mainly formed by interaction between subdomain-II; a central alpha helix pair (helix-35) runs almost parallel and creates a large buried hydrophobic core. A number of hydrogen bonds are formed between residues from the central α-helix and other residues of subdomain I and II.

The hexamer complex, made of a trimer of dimers, is organised into an offset triangular shape, with dihedral D3 symmetry (3-fold symmetry with addition 2-fold axis). An extensive hexamer interface exists at the meeting point of two dimers at the triangle's vertices. Consequently a single Pta domain monomer makes interactions to four other monomers through dimer and hexamer interfaces. The hexamer interface covers numerous overlapping interactions at the triangle vertex, which extends up to the midpoint (Figure 69). The larger hexamer interface, at the vertex point (hexamer vertex contact) has a buried surface area of 644Å^2. The hexamer interface also includes two smaller contacts at the midpoint (hexamer midpoint contact) that has a buried surface area of 322Å^2 (buried surface area calculated by PDB:sum).

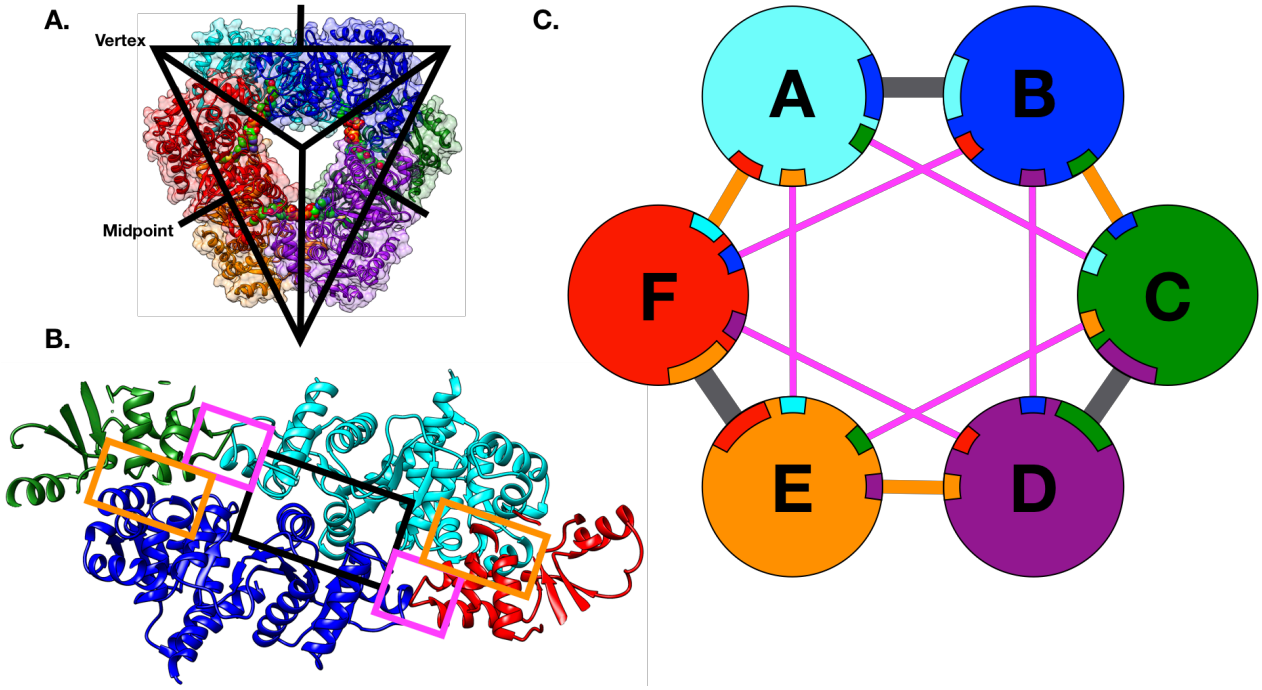


Figure 69. Hexamer interactions
A. Bd1833-Pta hexamer forms a triangular shape with dihedral D3 symmetry. **B.** Pta dimer viewed from the inside face demonstrating the hexamer surface contacts. The contacts in coloured boxes match the interactions shown in (C): Orange = larger hexamer vertex contact. Pink = smaller hexamer midpoint contact. Grey = dimeric contacts. **C.** A single monomer makes contacts to four other monomers through interactions at the dimer interface and hexamer interface. The hexamer interface is composed of four interacting monomers and two different surface contacts. The larger hexamer vertex contacts (orange line) are located at the hexamer verteces. The smaller hexamer midpoint contacts (pink lines) are made between the tips of subdomain-I helical flap region and the edge of subdomain-II.

The hexamer interface is stabilised by numerous non-bonding interactions as well as numerous salt bridges and hydrogen bonds. In total four pairs of ionic interactions occur at the hexamer interface, where one monomer of the dimer interacts with both monomers of the opposing dimer. Figure 70 illustrates the extent of the hexamer interface, where the AB dimer meets with the CD dimer. E544 (chain B) makes numerous interactions: a salt bridge with R749 and hydrogen bonds to the side of N574 from chain C; the backbone of N541 of chain B which further H-bonds to L696 from chain D. The hexamer interface undergoes the largest conformational change upon ligand binding, which is described in more detail in the following sections.

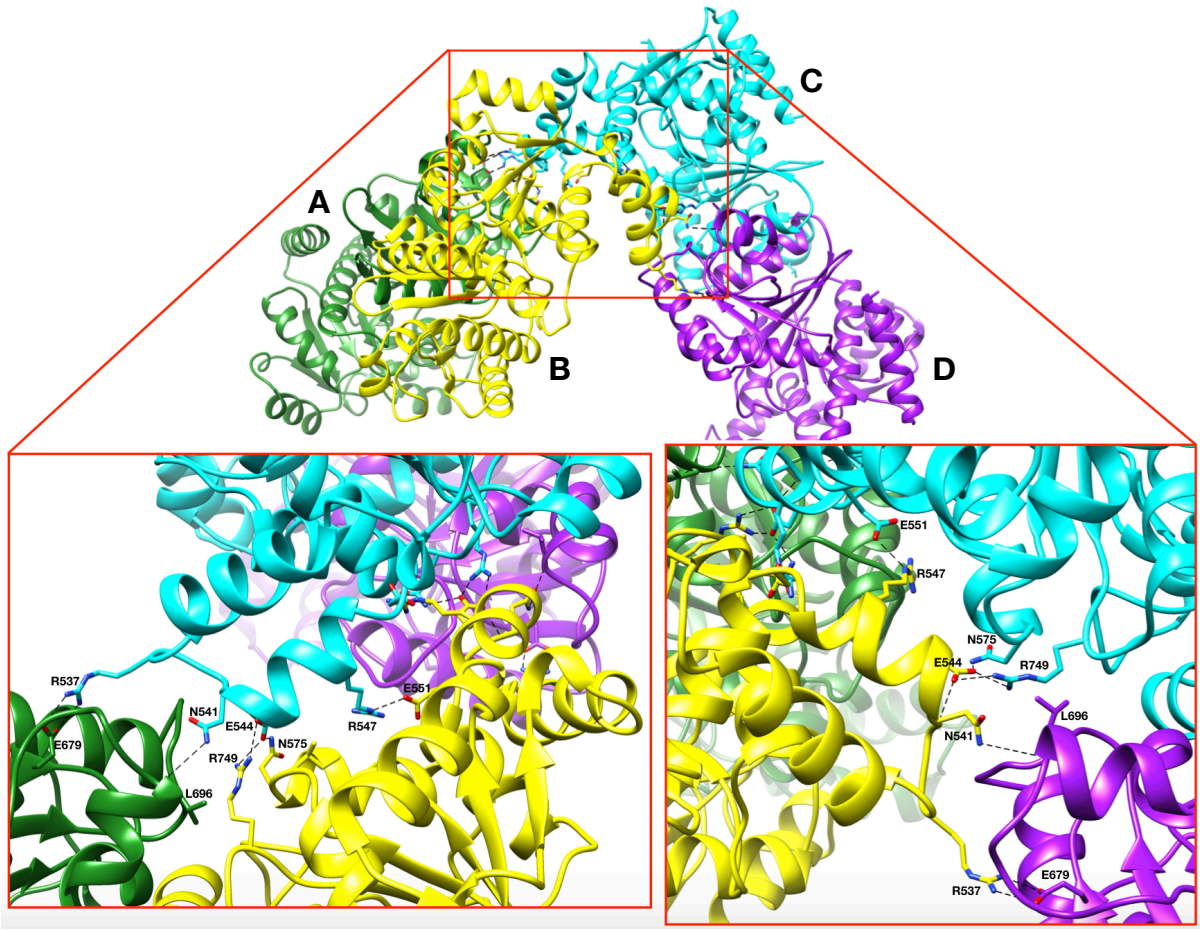


Figure 70. The hexamer interface
Bd1833-Pta forms a hexamer and each monomer forms a series of interactions with thee other monomers at the hexamer interface. This Figure illustrates a number of hydrogen bonds and ionic interactions that occur at the hexamer interface. The upper panel shows the overall hexamer interface. The left and right lower panels show a zoomed in view of the hydrogen bond network at the hexamer interface from two viewpoints. Chain A (green), Chain B (yellow), Chain C (cyan) and Chain (D) (purple). Chains A & B form a dimer and chain C & D form the other dimer that meet at the hexamer interface.

5.2.5.3 Acetyl-CoA binding site

Bd1833-Pta co-purifies and co-crystallises with acetyl-CoA, which binds in the shallow substrate-binding groove separating subdomain-I and II. One molecule of acetyl-CoA binds per monomer, hence the Pta hexamer complex has six molecules of acetyl-CoA ligand bound, located on the interior face at the triangle vertices. A number of non-bonding (vdw) interactions as well as specific hydrogen bonds are involved in acetyl-CoA binding, illustrated in Figure 67. The adenine group of acetyl-CoA makes a single hydrogen bond with the N616 sidechain. The side chain of R535 makes an important

bidentate ionic bond with the 3'-phosphate group and K538 side chain also makes a H-bond to the 3' phosphate. The pyrophosphate group is positioned parallel to the side chain of R581, which makes mainly electrostatic interactions. The guanidinium group of R581 is too far away from the acetyl-CoA pyrophosphate region (<3.2Å), therefore unable to make a H-bond. The tail of acetyl-CoA (pantothene group and acetyl group) is inserted into the binding pocket located between subdomain-I and II of one monomer and part of domain II of the other monomer of the dimer arrangement. The pantothene group makes hydrogen bond interactions with the backbone carbonyl atoms of Thr614 and V615. The acetyl group is situated at the end of the binding pocket with the carbonyl orientated towards a polar face, made by N718, K722, Y721 and Q683. The acetyl group makes a direct hydrogen bond with Q683 from the opposing monomer of the dimer and a highly coordinated water molecule.

5.2.5.4 Pta domain lacks catalytic residues

An alignment of the primary sequence of Bd1833 Pta domain with experimentally proven active Pta enzymes revealed many of the residues required to perform catalysis are not conserved. However, residues proposed to bind substrate in Pta enzymes are conserved in Bd1833 Pta domain (Figure 62). For example, the strictly conserved active-site arginine (R310 *Methanosarcina thermophile*) is represented by R749 in Bd1833-Pta. In the structure of Bd1833-Pta R749 is located near the acetyl group binding pocket, confirming speculation about involvement in binding and orientating acetyl-P in active Pta enzymes. However, in Bd1833-Pta R749 makes an important ionic interaction with E544 at the hexamer interface, suggesting a repurposed role. However, proposed Pta catalytic residues are not conserved in Bd1833, such as the active-site

serine residue (S309 *M. thermophile*), which is substituted to N748 in Bd1833. In addition other residues that line the acetyl-CoA binding pocket are not conserved in Bd1833-Pta: these include a consensus Asp (Pta enzymes) substituted to V746 and the highly conserved serine/threonine (Pta enzymes) substituted to Y576. Substitution of the strictly/highly conserved residues within the active site of Bd1833-Pta and lack of detectable enzymatic activity, suggests that the Pta domain is non-catalytic.

5.2.5.5 Pta R535E and R535A structures

R535 makes a significant bidentate interaction with the 3'-phosphate group of acetyl-CoA in the structure of Bd1833-Pta. Mutation of R535 to alanine or glutamate abolished acetyl-CoA binding (confirmed by HPLC analysis, shown in Section 5.2.9.1) and alleviates the inhibitory affects of acetyl-CoA (confirmed by kinetic analysis, shown in Section 5.2.8.4). The structures of both R535A and R535E mutants were solved to unpick the mechanism behind acetyl-CoA mediated inhibition. Pta-R535E and Pta-R535A both assemble into the same hexamer complex observed for WT Bd1833-Pta domain. A small, localised rigid-body conformational change is observed in the α-helix flap region (residues 529-548) of subdomain-I, which contains R535 at its tip. The flap region undergoes a ~5° outwards rotational swing, which pushes the loop ~1.5Å away from its original position. This conformational change suggests that the R535 interaction with the 3'-P group pulls this flap region of subdomain-I inward. Although the localised conformational change observed in the structures of Pta-R535A and Pta-R535E is small, there is a larger global affect on the hexamer arrangement (i.e. this change is propagated through all interfaces). A number of interactions at the hexamer interface are lost (cf. wild-type), such as the important E544 interaction with R749 and

N575. In addition the salt bridge formed between R547 and D551 is also lost. The loss of these interactions has a major impact of the organisation of the hexamer, changing the relative location of the interacting dimers by ~6Å and a 13° rigid body rotation of the interacting dimer. Consequently most of lost hexamer contacts are located within the larger hexamer vertex contact (Figure 71). The relatively small local conformational change is therefore amplified globally throughout the hexamer complex, altering the contacts made at the hexamer interface.

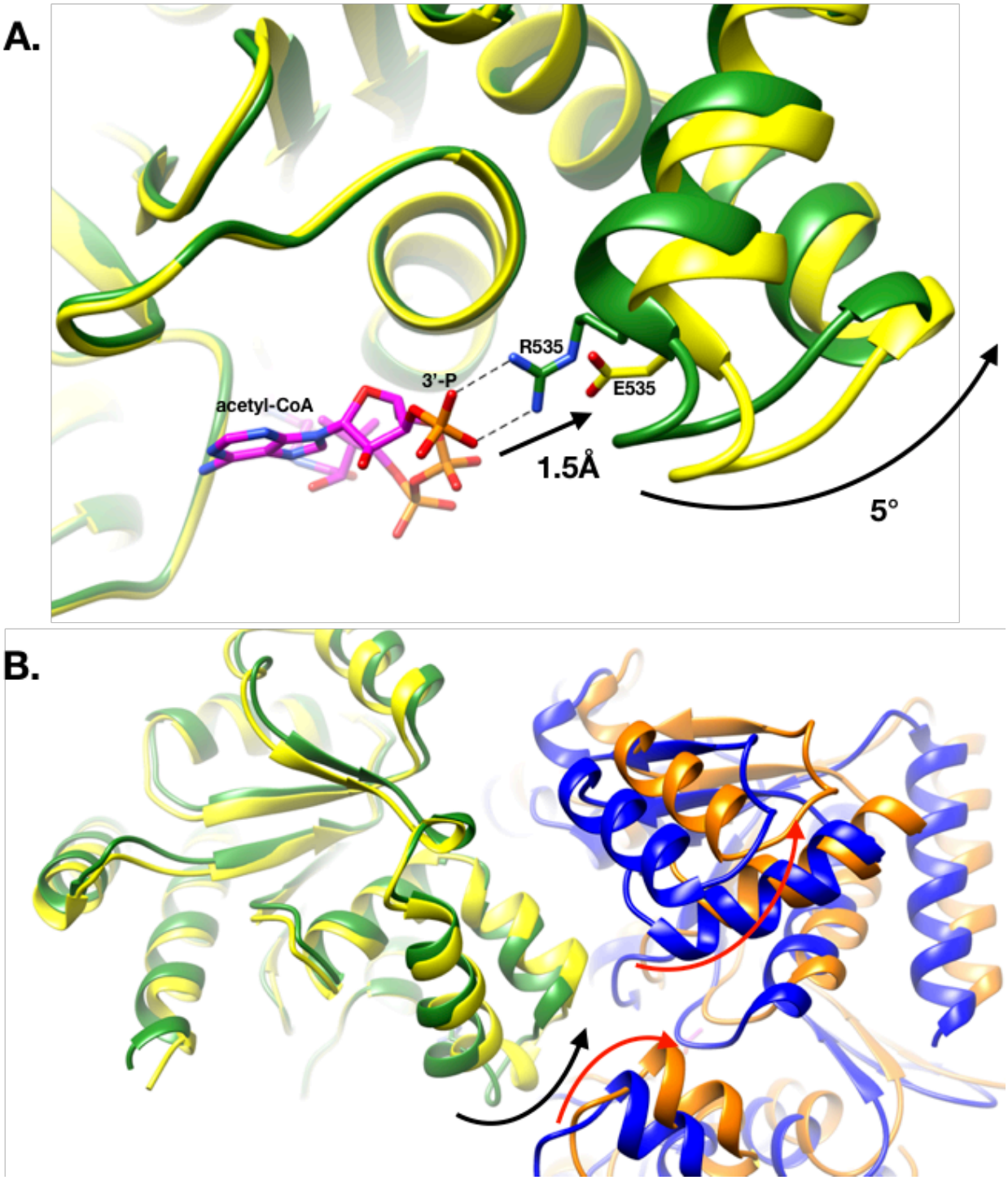


Figure 71. R535 mutation
A. Mutating R535 to a glutamate residue abolishes acetyl-CoA binding by repelling the 3'-phosphate group. R535E (yellow) mutation causes the lower loop (helix 27 and 28) of Bd1833-Pta to swing outwards when compared to WT- acetyl-CoA bound structure (green). **B.** The outwards swing (black arrow) causes a larger adjustment of the hexamer interface interactions (mainly lost from the vertex interface), shown by the red arrows. Blue is the position of hexamer interface in Bd1833-Pta (acetyl-CoA bound) and orange is the position of the hexamer interface in Pta-R535E mutant. A similar movement is observed for Bd1833-Pta R535A. The SSM superpose tool in coot was used to align chain A from Bd1833-Pta and Pta-R535E (length = 336 residues, rmsd = 0.67Å).

5.2.5.6 Pta-E544R

To investigate hexamer assembly and Pta:acetyl-CoA interactions further we designed a single mutation to destabilise the hexamer interface. E544 makes an important salt bridge with R749 in the structure of Bd1833-Pta, therefore E544 was mutated to an arginine or alanine residue (E544R / E544A) to abolish the interaction. Bd1833-E544R and Pta-E544R mutants were created in both the full-length and the Pta domain to evaluate mutant effects catalytically (full-length) and structurally (Pta). Interestingly the E544R mutation behaved differently in the two forms. Bd1833-Pta E544R did not co-purify with acetyl-CoA, was not inhibited and formed a dimer in solution (described in following sections). Pta-E544R, however, co-purified and co-crystallised with acetyl-CoA and remained a hexamer in solution and in the solved structure.

The structure of Pta-E544R reveals that the global hexamer complex adopts a similar arrangement to that observed for the Pta-R535E and Pta-R535A structures (less extensive hexamer interface). Additionally the acetyl-CoA binding flap adopts the same position as Bd1833-Pta structure, demonstrating that the movement in this region is associated with interactions with the 3'-phosphate group. Locally E544R is folded flat against the surface of the hexamer interface, repelled away from R749, which also adopts an outward facing conformation. In full-length Bd1833-E544R (dimer) the weakened hexamer interface may have been enough to fully destabilise the hexamer complex; however, in the Pta domain the E544R mutation was not sufficient. The Pta-E544R mutation demonstrates that a tightly packed hexamer arrangement is not required for acetyl-CoA binding. Additionally the mutation confirms the rearrangement of the hexamer complex observed in Pta-R535E and R535A structures is caused by destabilisation of the hexamer interface.

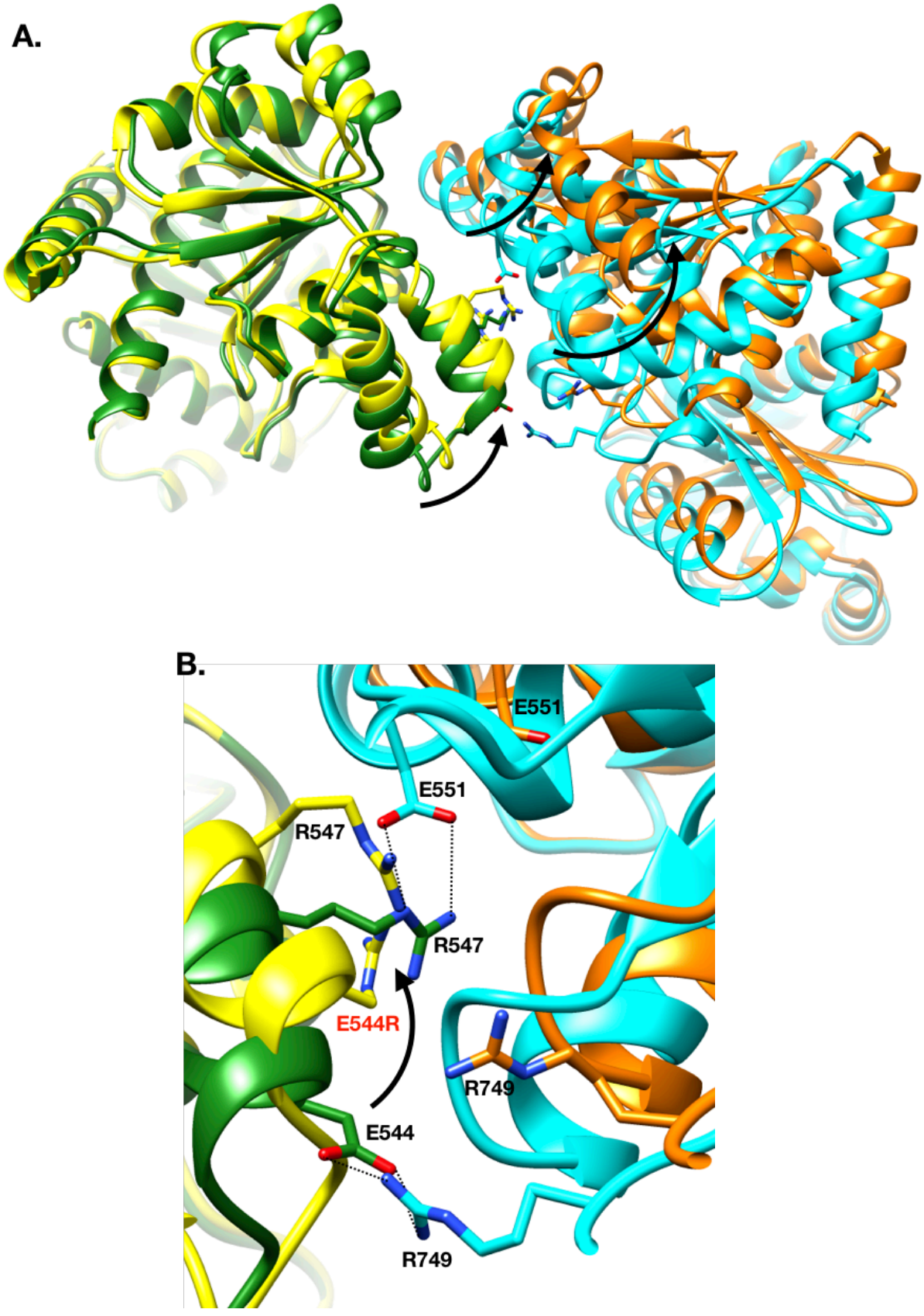


Figure 72. Pta-E544R
A. The hexamer interface of E544R is destabilised and causes a conformational reorientation between interacting dimers. Bd1833-Pta (green and cyan) and Bd1833-E544R (yellow and orange). **B.** E544R mutation is repelled by R749 but folds back inwards adopting a similar location as R547. Superimposition was performed by least-squares minimisation using the CCP4 suit superpose software. Alignment was performed for Pta dimers using subdomain-II (residues 596-738) from both protomers (rmsd = 0.55Å).

5.2.5.7 Y576S mutant movement

Y576 is located deep in the acetyl-CoA binding groove, where the Y576 C_α is located 3.5Å from the thiol group. In catalytically active Pta enzymes Y576 is a conserved small polar residue (serine or threonine) thus in a primary location to participate in catalysis. We therefore designed a mutant (Pta-Y576S) to convert Y576 back into a consensus Pta, to investigate the affects on acetyl-CoA interaction. Interestingly Pta-Y576S selectively co-purifies with CoA and not acetyl-CoA (HPLC result). We then solved the structure of Pta-Y576S; however, no density was visible for a CoA ligand in the refined $2F_o$-F_c map. Subdomain-II (residues 604-729) superimpose very well between Bd1833-Pta and Pta-Y576S structures (Figure 73). However, subdomain-I shows a small rigid-body rotational movement (4°), which slightly opens the acetyl-CoA binding groove between the two domains. Moreover, interactions at the hexamer interface are also disrupted, such as the E544 interaction with R749 and the R547 interaction with D551. The extent of hexamer disruption is similar to those observed for other Pta mutations (R535R, R535A, E544R), wherein the larger hexamer interface is reduced, accompanied by a rotational rigid-body conformational change of the interacting dimers and the hexamer interface. The structure of Y576S supports the concept of acetyl-CoA binding impacting on hexamer arrangement.

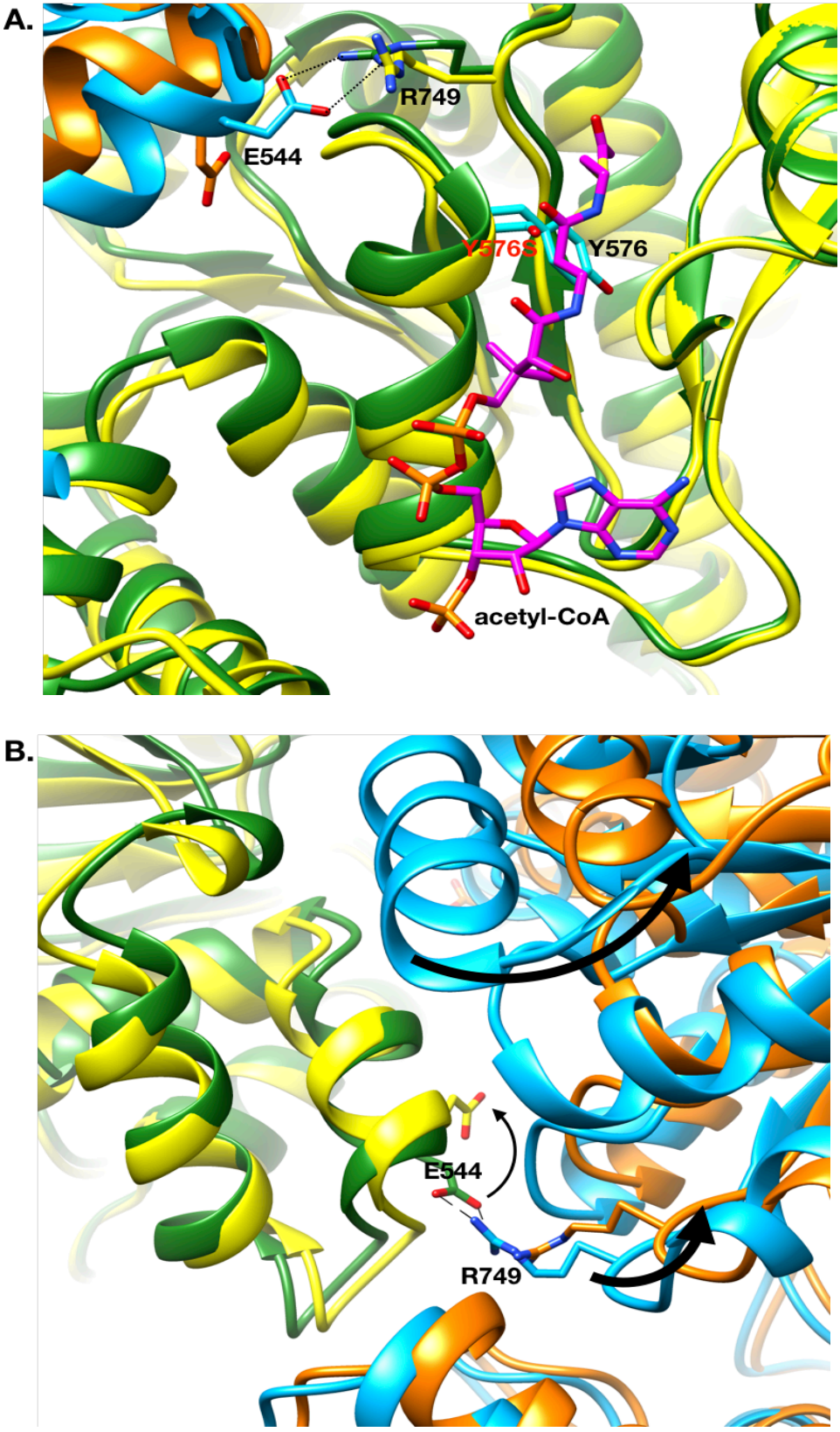


Figure 73. Y576S mutation
A. Y576 is located close to the thiol group of acetyl-CoA and is a highly conserved serine residue in Pta enzymes. Y576S selectively prevents acetyl-CoA binding but allows CoA binding. The structure of Pta-Y576S (yellow and orange) superimposed with Bd1833-Pta (green and cyan). **B.** Y576S mutation selectively prevents acetyl-CoA binding, which changes the conformation of subdomain-I and consequently destabilises the hexamer interface. Superimposition was performed by least-squares minimisation using the CCP4 suit superpose software. Alignment was performed for Pta dimers using subdomain-II (residues 596-738) from both protomers (rmsd = 0.72Å).

5.2.5.8 Pta-Apo structure domain conformational change

In order to assess the effect of acetyl-CoA binding on the structure of Bd1833-Pta, without introducing mutations, we needed to remove the co-purifying ligand. To remove ligand an additional mild denaturing wash step (2.5M urea, 2M KBr) was introduced into the purification procedure. HPLC was used to determine whether the wash step had reduced the amount of co-purified ligand to <95% of that found in Bd1833-Pta (see HPLC results).

The structure of Pta-Apo was solved revealing a rigid-body rotational conformational change associated with subdomain-I. When the structure of Bd1833-Pta is superimposed onto the structure of Pta-Apo, the central subdomain-II forming the dimer interface remains similar; however, the relative position of subdomain-I is altered. Subdomain-I rotates away from subdomain-II, creating a larger substrate-binding groove between the two subdomains. Figure 74 demonstrates the conformational change observed in Pta-Apo, wherein the upper region of the protein undergoes an 11° rigid-body rotation compared to the lower region that rotates 1°. This finding demonstrates that Bd1833 Pta domain changes conformation upon acetyl-CoA binding, and subdomain-I repositions in response to acetyl-CoA interaction. The movement observed suggests that upon acetyl-CoA binding, subdomain-I interacts more closely with subdomain-II, reducing the extent of the substrate-binding groove between the two domains and encasing the bound acetyl-CoA ligand. Furthermore, the conformational change in subdomain-I also impacts in the global arrangement of the hexamer complex, similarly to that observed in R535E, Q576S and E544R structures (Figure 75).

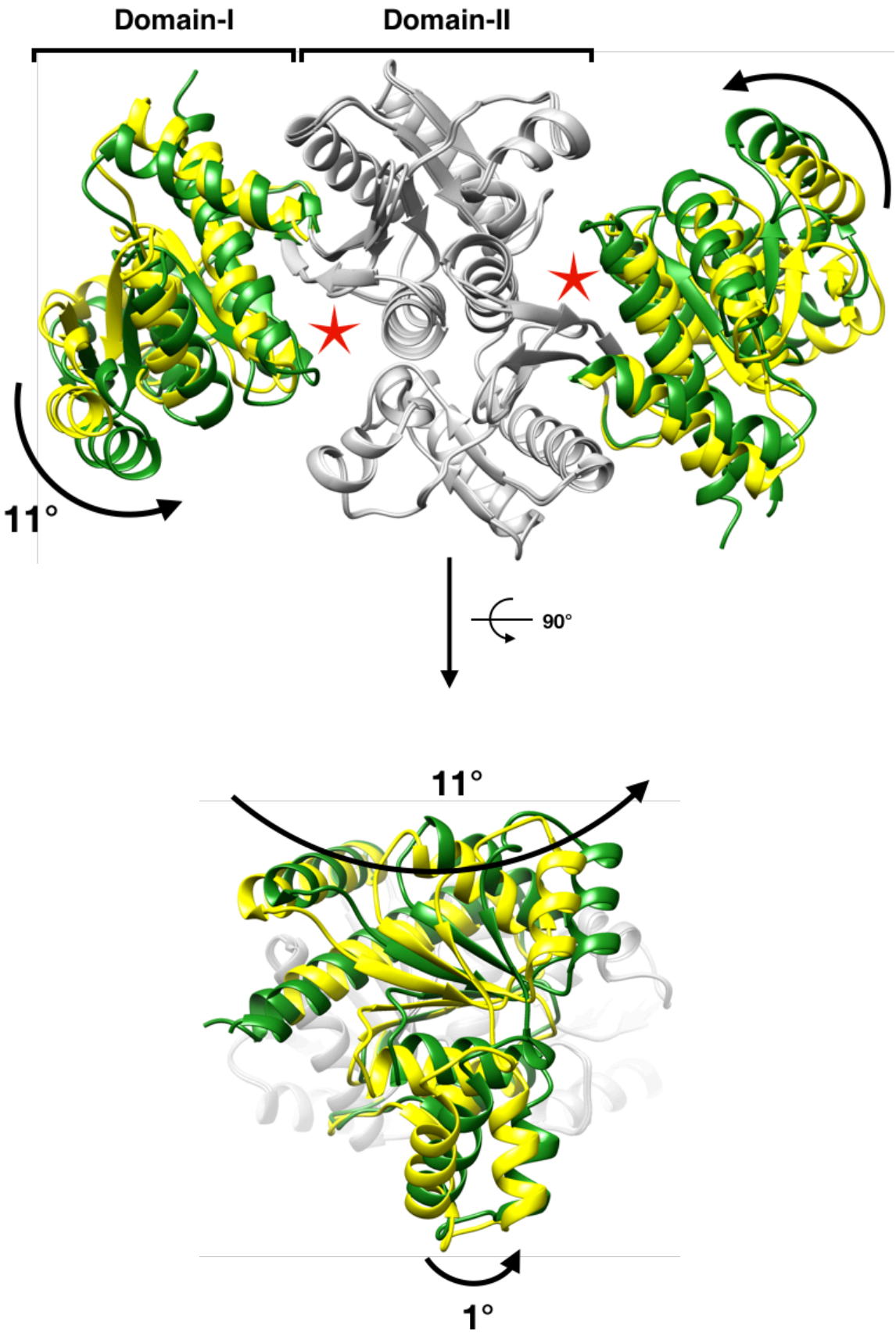


Figure 74. Bd1833-Pta domain movement
Bd1833-Pta (green) and Pta-Apo (yellow) structures are shown, superimposed by subdomain-II (grey). A red star marks the substrate binding sites. Subdomain-I undergoes an 11° rotational rigid body conformational change, indicated by the black curved arrows. The movement changes accessibility to the substrate binding groove located between subdomain-I and II. A sideview of the same rotational rigid body movement demonstrates the upper region undergoes a larger rotational change compared to the lower region. Superimposition was performed by least-squares minimisation using the CCP4 suit superpose software. Alignment was performed for Pta dimers using subdomain-II (residues 596-738) from both protomers (rmsd = 0.82Å).

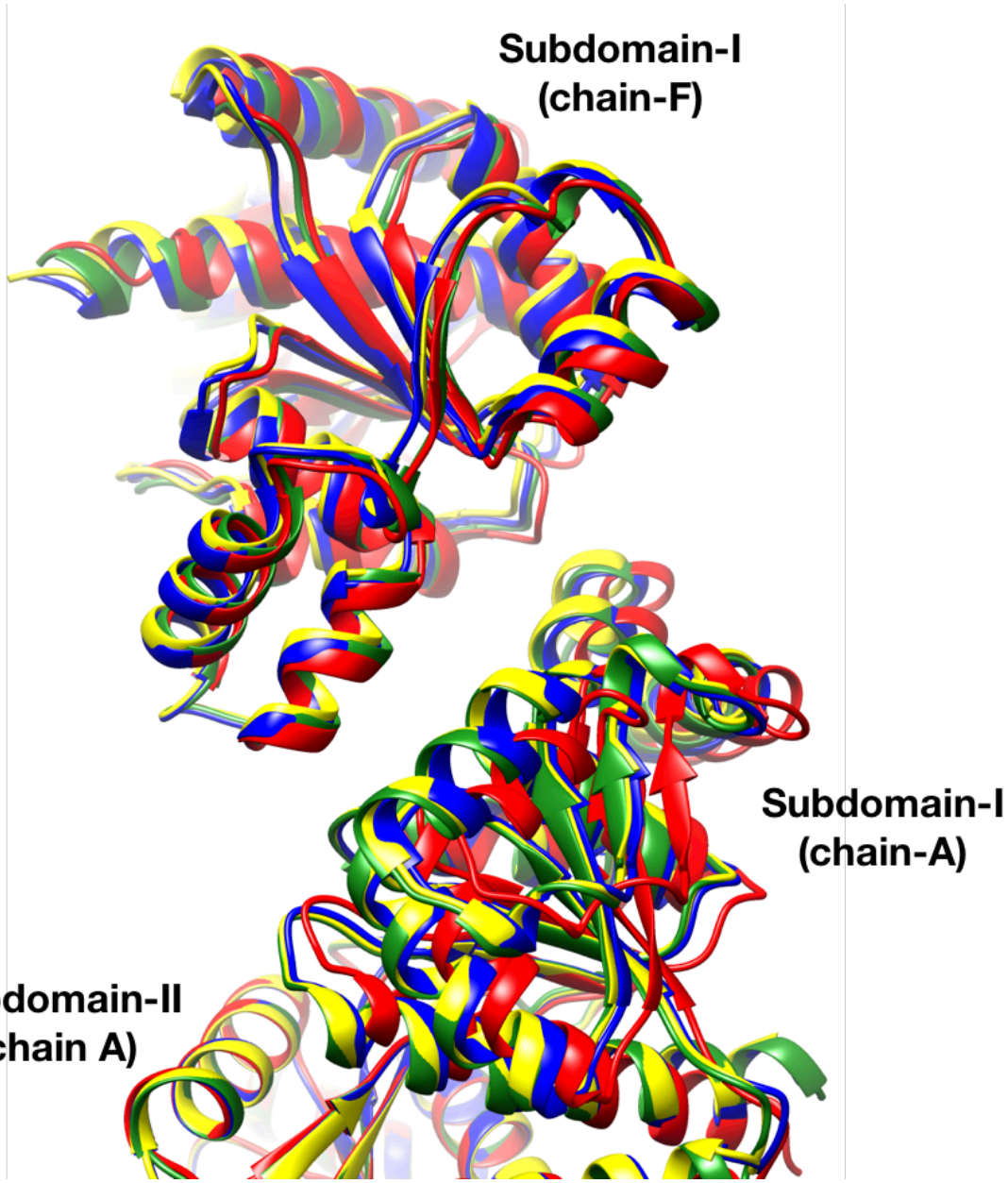


Figure 75. Pta mutations and hexamer interface
The hexamer vertex interface between the AB dimer and EF dimers is shown for R535E (blue), R544R (yellow) and Q576S (green) mutants aligned with Pta-Apo (red). The hexamer is aligned around subdomain-II in chain-A. The degree of hexamer rearrangement affecting the vertex interface is similar in all mutations and Apo form. The vertex hexamer interface is reduced from 644Å^2 in Bd1833-Pta in complex with acetyl-CoA to ~240Å^2 in ligand free Bd1833-Pta structures, which is more than a two-fold decrease.

5.2.6 The structure of full-length Bd1833

We have solved the structure of full-length Bd1833 (residues 15-776) in complex with acetyl-CoA revealing the multi-modular domain arrangement (Figure 76). The Pta domain (residues 438-777) and ME domain (residues 16-420) oligomer interfaces of the full-length Bd1833 protein are identical to those observed in the isolated domain structures. Bd1833 forms a large hexamer complex (trimer of dimers) formed by interactions between Pta domains. The hexamer forms an offset triangle, where the Pta domain provides the base and the ME domain is above the flat Pta dimer sides, offset by ~40°. Linking the two domains is a 17aa region containing a 13aa alpha helix (residues 421-437), referred to hereafter as the linker-helix.

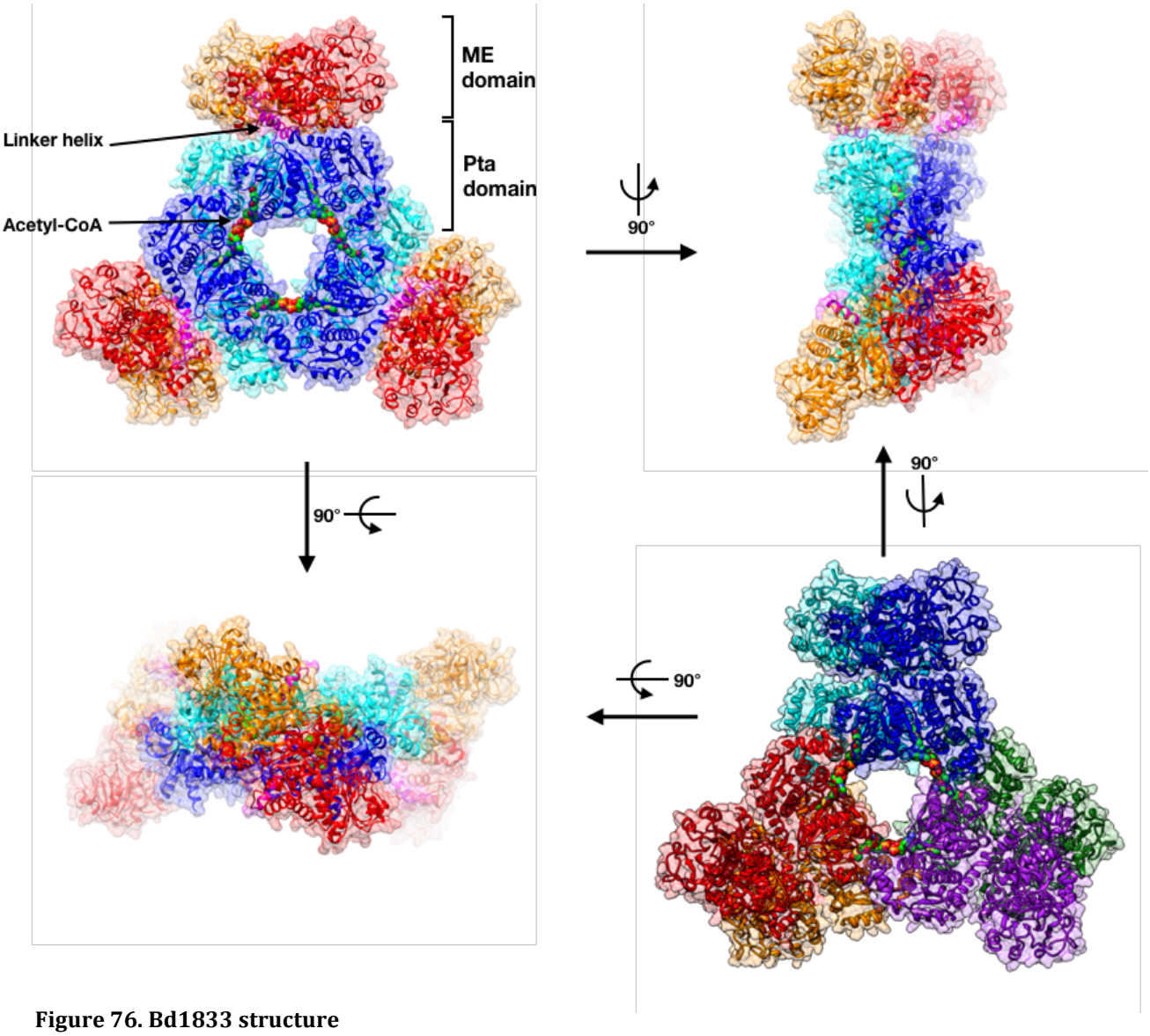


Figure 76. Bd1833 structure
Full-length Bd1833 protein forms a larger hexamer complex. The top left panel shows the hexameric structure of Bd1833 coloured according to the domains (red/orange = ME and blue/cyan = Pta). The linker helix is shown in magenta and acetyl-CoA ligand in green. Top right and bottom left panels show the hexameric structure in two different orientations rotated by 90° demonstrating the twisted nature of the hexamer complex and how the ME domain is positioned on top of the Pta domain, offset by ~40°. Finally the bottom right panel shows Bd1833 hexamer coloured according to the protomers that form the hexamer complex.

The Pta and ME domains are held together by the linker helix which is located on either side of the Bd1833 dimer (Figure 77). There are minimal interactions between the Pta surface and ME surface and in the central region of the dimer a 6Å gap exists between the surfaces of both domains. A few specific H-bonds are formed between the Pta domain and the ME domain. These include: R444 with F396 backbone carbonyl and

K440 with K112 backbone carbonyl. The majority of the surface interactions between Pta and ME domains are electrostatic, involving charged and polar residues (Figure 77).

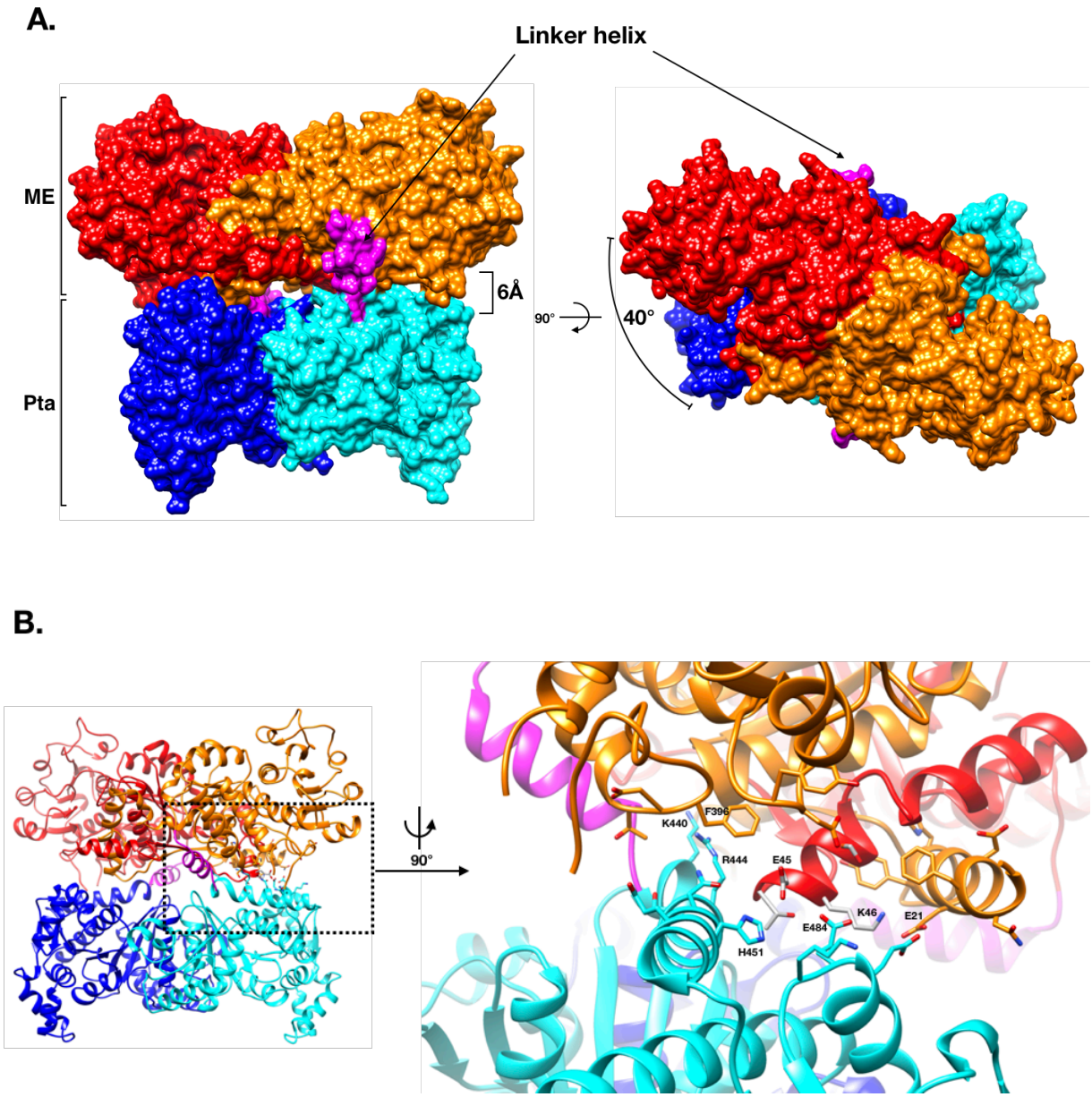


Figure 77. Pta and ME interface
A. The Pta and ME domains in full-length Bd1833 are linked by a single linker helix, located on either side of the Bd1833 dimers. Little surface contact is made between the two domains that are offset by ~40° relative to one another. In the central regions a gap up to 6Å exists between the domain surfaces. **B.** Minimal interactions exist between the Pta and ME domain. The interactions that are present are mainly electrostatic and formed between subdomain-A/C of the ME domain and subdomain-I of the Pta domain. The residues contributing to the minimal interface between the Pta and ME domains are displayed.

Although the structures of the individual domains closely resemble those of the full-length structure, some key differences are observed. The structure of full-length Bd1833 is complexed to acetyl-CoA, therefore representing an inhibited state. The ME domain of full length Bd1833 adopts a different conformation in comparison to Bd1833-ME (non-inhibited state). There is a rigid body conformational change in ME subdomain-C ($NADP^+$ -binding domain). This rigid body rotation of subdomain-C, by ~9°, restricts access to the substrate/co-factor binding site and therefore represents a closed/inhibited state. Another obvious structural difference between Bd1833 and Bd1833-ME is the complete lack of electron density for subdomain-D (the multi-modular ME specific loop (residues 371-382)). When the structure of Bd1833-ME is superimposed onto full-length Bd1833, subdomain-D clashes with the Pta domains N-terminal helix, which adjoins directly to the linker-helix. If the structure of Bd1833-Pta is superimposed onto Bd1833, there are few conformational differences. However, when Pta-Apo (non-inhibited state) is superimposed onto Bd1833 conformational changes in Pta subdomain-I are observed. Interestingly, the conformational changes in subdomain-I demonstrate how the movement in this domain is allosterically conveyed to a conformational change in ME subdomain-C (Figure 78).

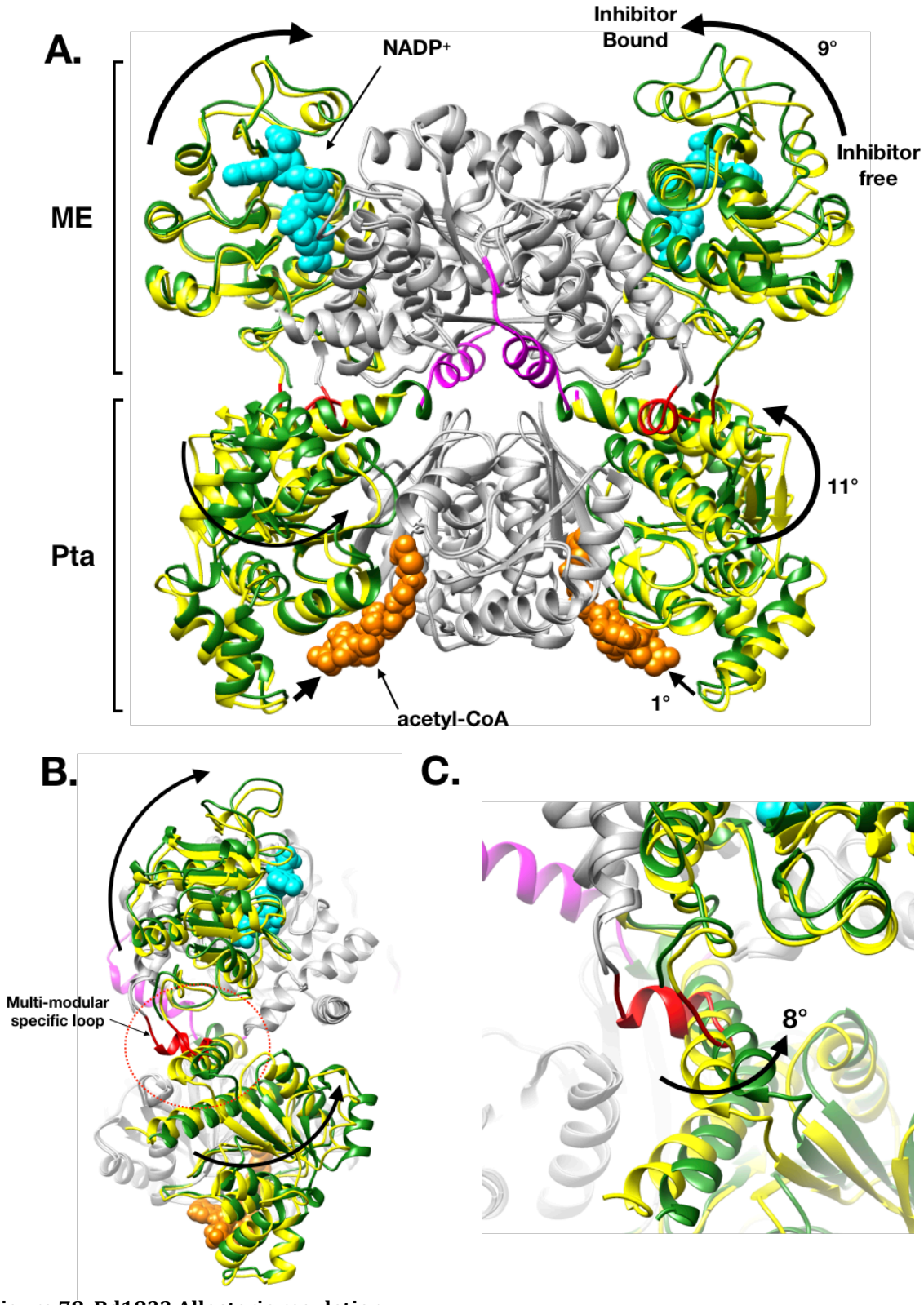


Figure 78. Bd1833 Allosteric regulation

A. The structure of full-length Bd1833 in complex with acetyl-CoA (green) is superimposed with Pta-Apo (yellow, length = 1090 atoms, rmsd = 0.86Å) and Bd1833-ME-NADP (yellow, length = 6670 atoms, rmsd = 0.47Å). The core region of Bd1833 (ME subdomain A and B and Pta subdomain-II) does not alter conformation between structures and is coloured in grey for clarity. The linker helix is shown in magenta and the multi-modular specific loop region (ME subdomain-D) is shown in red. The location of Pta and ME binding sites are represented by $NADP^+$ (cyan) and acetyl-CoA (orange) to gauge the distance between binding sites. An 11° rigid body rotational movement of Pta subdomain-I allosterically causes a 9° rigid body rotational movement in ME subdomain-C, restricting access to the active site. **B.** Demonstrates the conformational changes in full-length Bd1833, viewed from the side. This image demonstrates the clash with Bd1833-ME subdomain-D (red). **C.** A close up view of the clash with subdomain-D; subdomain-D which is not traceable in the structure of full-length Bd1833. Superimposition was performed by least-squares minimisation using the CCP4 suit superpose software. Alignment was performed for Bd1833 dimers using residues 34-174 & 596-738 from both protomers.

5.2.7 Structure of Bd1833-SKB

Bd1833-SKB is a homologue of Bd1833 (HD100) from a strain of *Bdellovibrio bacteriovorus* SKB1291214 (sharing 86.41% sequence identity). Bd1833-SKB does not co-purify with ligand (HPLC results), possibly due to a lysine residue substitution at residue 581 (R581 in Bd1833), which interacts with the pyrophosphate group of acetyl-CoA in Bd1833 structures. We have solved the structure of Bd1833-SKB to 3.7Å and although the resolution is low, the structural information gained is valuable. Bd1833-SKB crystallises with three copies in the asymmetric unit, which forms a hexamer complex, similar to Bd1833, when crystallographic symmetry factors are applied. Consequently the structure of Bd1833-SKB likely represents a non-inhibited full-length structure of Bd1833.

The structure of Bd1833-SKB is highly similar to the structure of Bd1833 in that only the ME subdomain-D is unstructured. Importantly, conformational changes in ME subdomain-C and Pta subdomain-I are observed when Bd1833 and Bd1833-SKB are superimposed (Figure 79). This finding is supportive of acetyl-CoA mediated conformational change in Pta domain conveying a conformational change in the ME domain. However, the conformational changes observed in Bd1833-SKB compared to Bd1833 are relatively small suggesting large conformational changes may not be required for acetyl-CoA to inhibit activity.

The Pta domain only, truncation mutant of Bd1833-SKB (Pta-SKB) co-purified and co-crystallises with acetyl-CoA, unlike full-length Bd1833-SKB. We solved the structure of Pta-SKB to higher resolution (2.7Å), uncovering the binding site similarities and

differences. K581 represents the only distinctive change, which does not appear to make an interaction with the pyrophosphate group of acetyl-CoA (Figure 79.B).

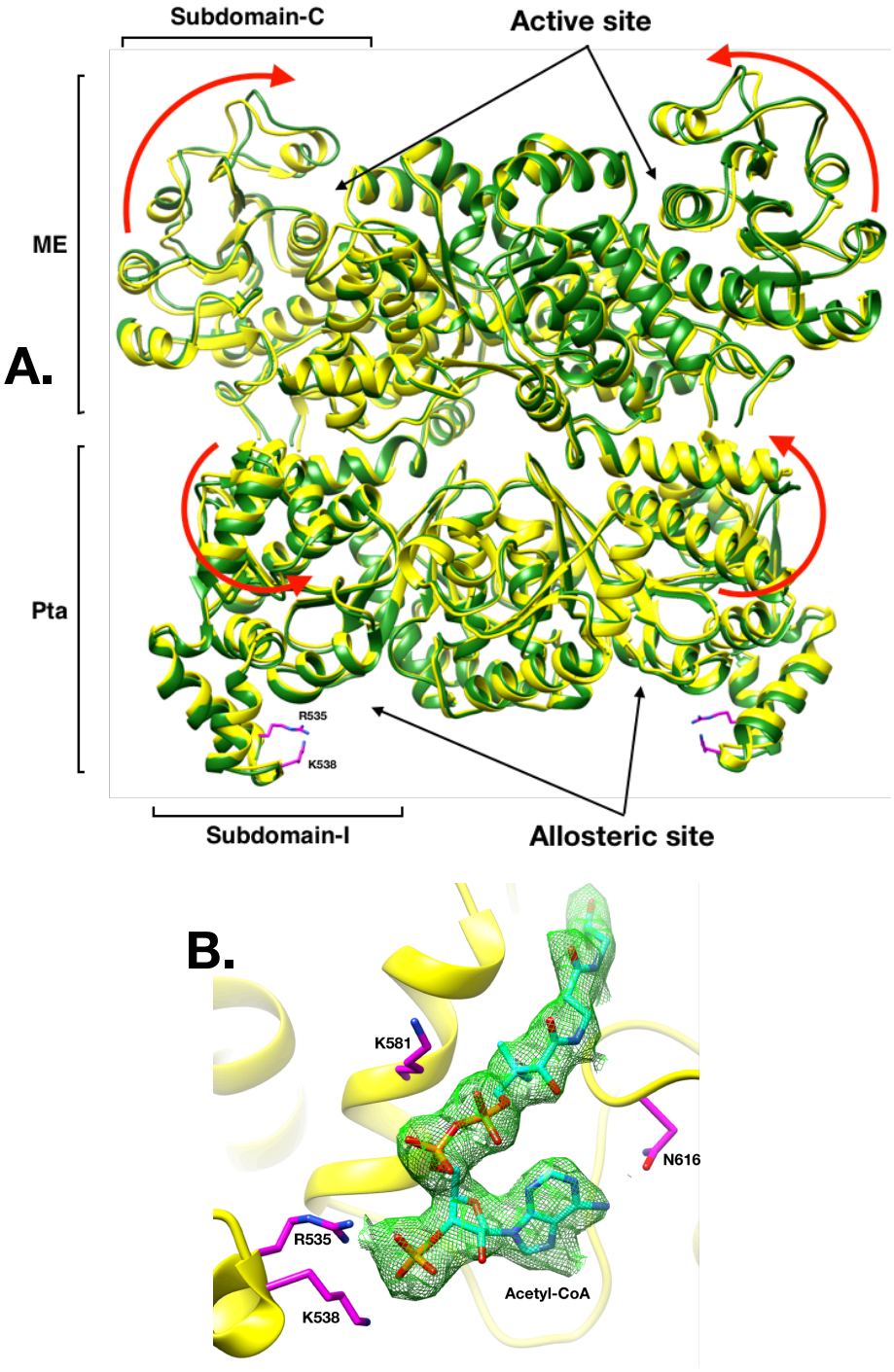


Figure 79. B1833-SKB structure
A. The structure of Bd1833-SKB (Yellow) is superimposed with the structure of Bd1833 (Green)(length = 23032 atoms, rmsd 0.55Å). Bd1833-SKB does not co-purify with ligand and likely represents a non-inhibited full-length conformation. The ME domain active site and Pta allosteric site are highlighted. Red arrows demonstrate the rigid body movements in subdomain-I and subdomain-C. Superimposition was performed by least-squares minimisation using the CCP4 suit superpose software. Alignment was performed for Bd1833 dimers using residues 34-174 & 596-738 from both protomers. **B.** The acetyl-CoA interaction with Pta-SKB (truncation mutant of Bd1833-SKB that co-purifies with ligand). K581 does not interact with the pyrophosphate group. ($2F_o$-F_c map, radius=2Å, σ=1).

Bd1833 species	Kinetic Parameters (malate) K_M (mM)	Kinetic Parameters (malate) k_{cat} (s^{-1})	Kinetic Parameters ($NADP^+$) K_M (µM)	Kinetic Parameters ($NADP^+$) k_{cat} (s^{-1})	Oligomeric State	Acetyl-CoA Inhibition	Co-purified Ligand?	Reasoning
Bd1833	2.61 ± 0.28	26.1 ± 1.2	21.7 ± 2.3	21.1 ± 0.13	Hexamer	Yes	CoA + AcCoA	
Bd1833-Apo	3.02 ± 0.31	35.2 ± 1.4	43.1 ± 5.2	33.5 ± 1.3	Hexamer	Yes	None	Ligand free
Bd1833-ME	4.15 ± 0.50	3.28 ± 0.17			Dimer	No	None	ME domain only
Bd1833-Δloop	12.5 ± 1.0	1.62 ± 0.09			ND	No	CoA + AcCoA	Chimera lacking multi-modular specific loop
Bd1833-R535E	3.17 ± 0.15	19.4 ± 0.36	41.1 ± 4.0	16.8 ± 0.54	Hexamer	No	None	Prevent CoA binding (3'-P)
Bd1833-R535A	3.48 ± 0.23	20.3 ± 0.53			Hexamer	No	None	Prevent CoA binding (3'-P)
Bd1833-N718D	3.35 ± 0.33	41.8 ± 1.6			Hexamer	No	None	Prevent CoA binding (acetyl-group)
Bd1833-E544R	12.6 ± 2.0	18.9 ± 1.9			Dimer	No	None	Destabilise hexamer
Bd1833-E544A	4.81 ± 0.44	38.7 ± 1.6			Hexamer	Yes	CoA + AcCoA	Destabilise hexamer
Bd1833-Y576S					Hexamer	No	CoA	Convert back to conserved Serine residue
Bd1833-Y576C					ND	Yes	CoA	Interact with thiol group
Bd1833-R749A					ND	Yes	CoA + AcCoA	Destabilise hexamer
Bd1833-R547A					ND	Yes	CoA + AcCoA	Destabilise hexamer
Bd1833-R581K					ND	Yes	CoA + AcCoA	Pyrophosphate interaction
Bd1833-SKB	1.96 ± 0.24	35.0 ± 1.3	45.4 ± 6.6	39.3 ± 1.8	Hexamer	Yes	None	Close homologue

Table 14. Bd1833 kinetics
This table includes the values obtained from our kinetic studies using malate as the substrate, the oligomeric state and whether inhibition by acetyl-CoA was detected for each Bd1833 species. ND = not determined. Values presented as a mean of triplicates alongside the standard deviation.

5.2.8 Kinetics

5.2.8.1 Bd1833 activity

A detailed kinetic analysis of Bd1833 has been performed. ME activity was detected for both malate oxidative decarboxylation and pyruvate reductive carboxylation reactions; however, no Pta activity was detected. Table 14, displays the kinetic parameters obtained for Bd1833 and Bd1833-mutant species.

5.2.8.2 Malate Oxidative Decarboxylation

Bd1833 required a divalent cation such as Mn^{2+} to be catalytically active. Bd1833 also demonstrates specificity for $NADP^+$ as a cofactor, showing no activity in the presence of NAD^+. Kinetic parameters were determined using L-malate as the substrate, at a concentration between 0.1 and 10 mM; reduced rates were observed beyond 10 mM L-malate (Figure 80).

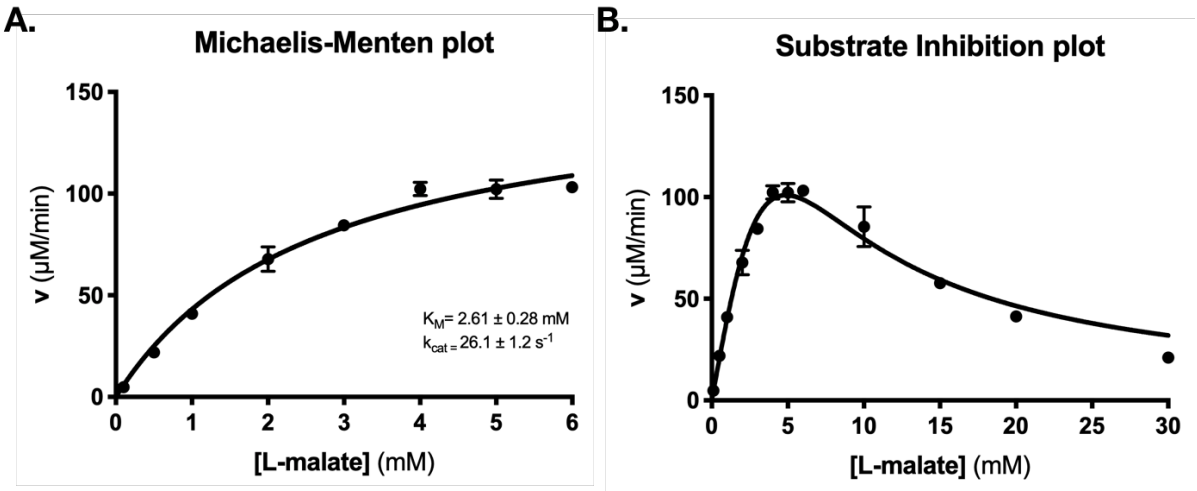


Figure 80. Malate oxidative decarboxylation
Kinetic data shown for full-length Bd1833. **A.** The data for Bd1833 fit to a Michaelis-Menten plot. **B.** At concentrations of malate above 8 mM, reduced rates are observed. The data has been fit to substrate inhibition model $Y=V_{max}*[\text{L-malate}]/(K_M + [\text{L-malate}]*(1+[\text{L-malate}]/K_i))$ using prism7 software to explain the reduced rates. The fit is "ambiguous" suggesting the model is not ideal to explain the data, suggesting other factors such as ΔpH or ionic strength may also contribute. All data points were measured in at least triplicates. The error bars represent the standard deviation.

5.2.8.3 Effectors

A series of potential effectors, published for other malic enzymes, were tested as potential regulators of Bd1833. The activity of Bd1833 in the presence of effectors is presented as a percentage of the activity of Bd1833 without effector (Figure 81). Acetyl-CoA and OAA are shown to be potent inhibitors of Bd1833 activity and reduce the activity to below 10% of the un-inhibited activity. CoA and acetyl-P reduced the activity to ~50% suggesting a weaker mode of inhibition. No activators of Bd1833 activity were identified.

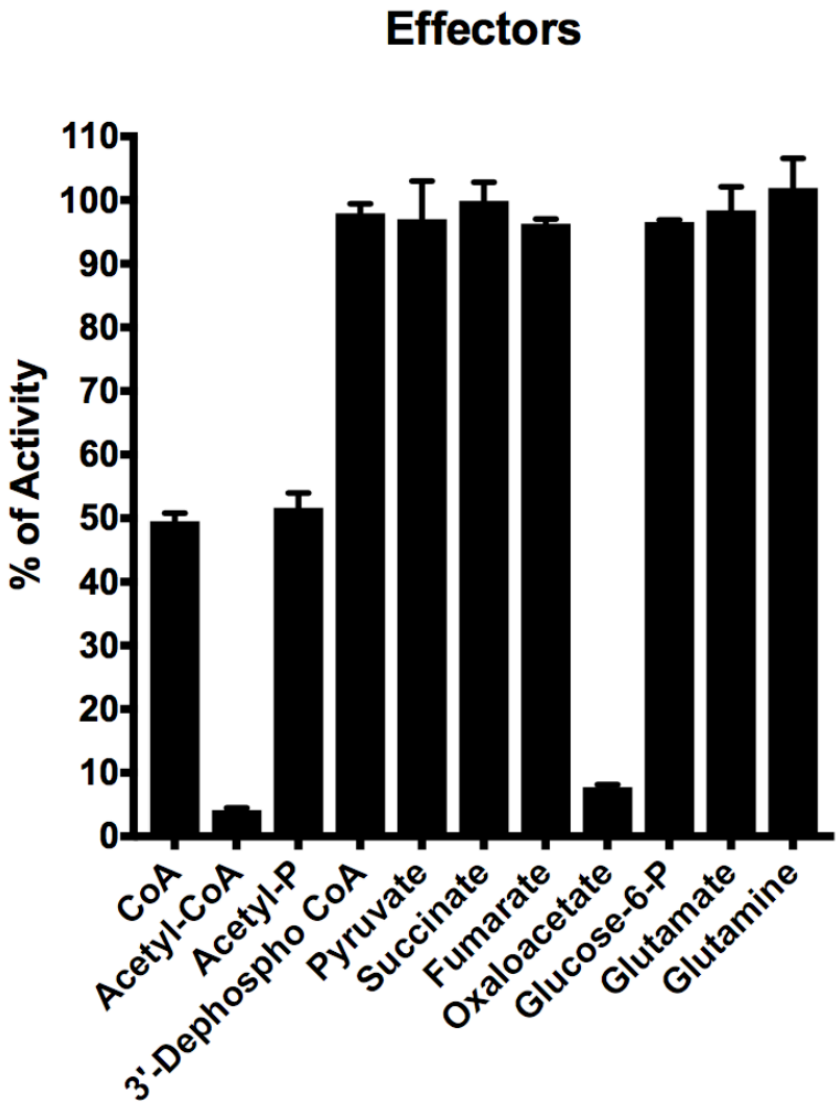


Figure 81. Identifying Bd1833 effectors
A range of malic enzyme effectors were tested for their effects on Bd1833 activity. Bd1833 activity is reduced to ~50% in the presence of CoA and acetyl-P. Acetyl-CoA and Oxaloacetate are potent inhibitors of Bd1833 activity, reducing the activity to below 10% of Bd1833 activity in standard assay conditions. All data points were measured in at least triplicates. The error bars represent the standard deviation.

5.2.8.4 Effector binding sites

To investigate how these effectors regulate the activity of Bd1833 we created a number of Bd1833 mutants. The truncation mutant (Bd1833-ME) activity was measured in the presence of the effectors and compared to the activity without (Figure 82). Acetyl-CoA, CoA and acetyl-P showed no inhibitory effects towards truncated Bd1833-ME; however, OAA remained an inhibitor. This result confirms that acetyl-CoA, CoA and acetyl-P interact with the Pta domain and allosterically regulate ME activity, whereas OAA interacts directly with the ME domain, which supports our structural data. Interestingly, the activity of Bd1833-ME was reduced in comparison to full-length Bd1833, suggesting communication with the Pta domain is important for catalysis.

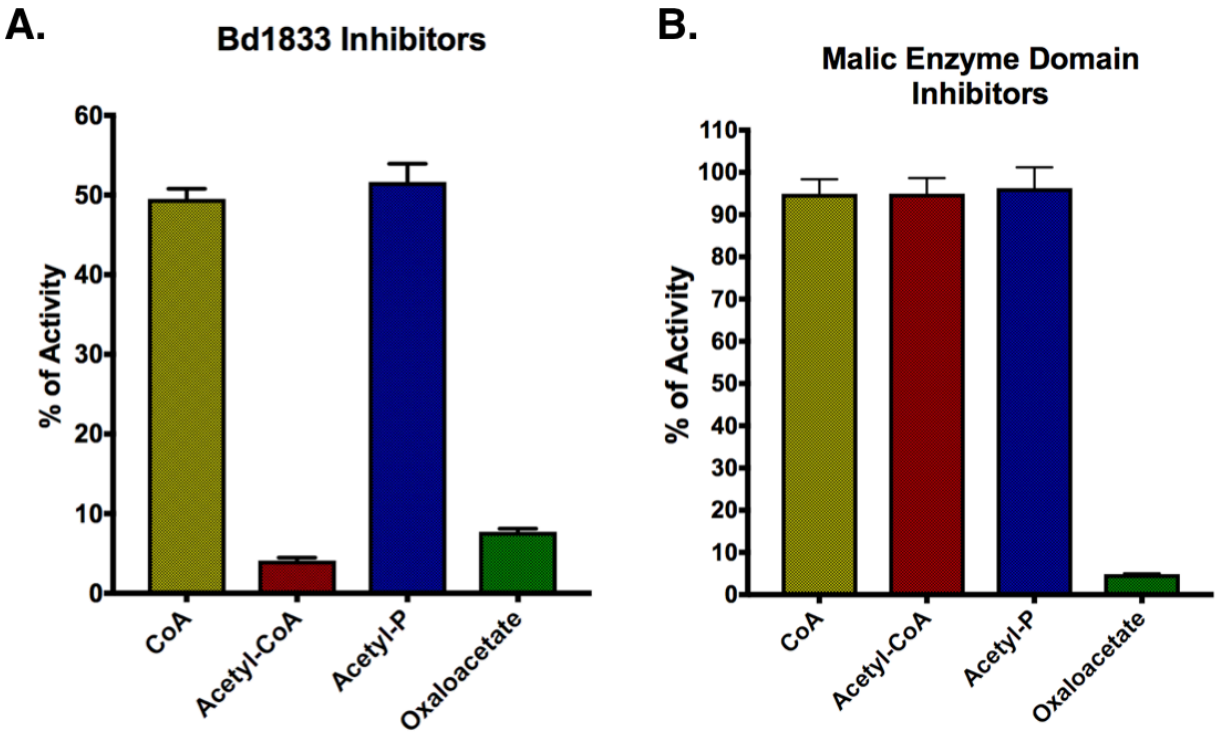


Figure 82. Bd1833 inhibitors
A. The activity of Bd1833 in the presence of inhibitors. **B.** The activity of Bd1833-ME truncation mutant in the presence of the same effectors. Three of the effectors, including acetyl-CoA, do not inhibit Bd1833-ME whereas oxaloacetate still inhibits. This finding indicates which domains the inhibitors interact with. All data points were measured in at least triplicates. The error bars represent the standard deviation.

To confirm the inhibitor-binding site, we created mutations of R535 (the highly conserved residue responsible for binding the 3'-phosphate group of acetyl-CoA in the structure of Bd1833). The kinetic parameters for two Bd1833 mutants (Bd1833-R535A and Bd1833-R535E) were determined, revealing both mutants had a reduced k_{cat} compared to Bd1833. However, Bd1833-R535E and Bd1833-R535A were both insensitive to acetyl-CoA and CoA (Figure 83). Furthermore, a CoA analogue lacking the 3'-phosphate group (3'-Dephospho-CoA) did not inhibit the activity of Bd1833. These results, together with structural data, suggest R535 and its interaction with the 3'-phosphate group of CoA is important for inhibitor binding and initiating allosteric regulation.

Another mutant, Bd1833-N718D was designed to block acetyl-CoA binding by interfering with the acetyl-group binding pocket. N718 is located within 3.5Å of the thiol group of acetyl-CoA and contributes to the polar face of the acetyl-group binding pocket. Importantly, the Bd1833-N718D mutant was almost twice as active as Bd1833, with a k_{cat} of 41.8 ± 1.6s^{-1}. Bd1833-N718D was also insensitive to inhibition by acetyl-CoA and CoA (Figure 83).

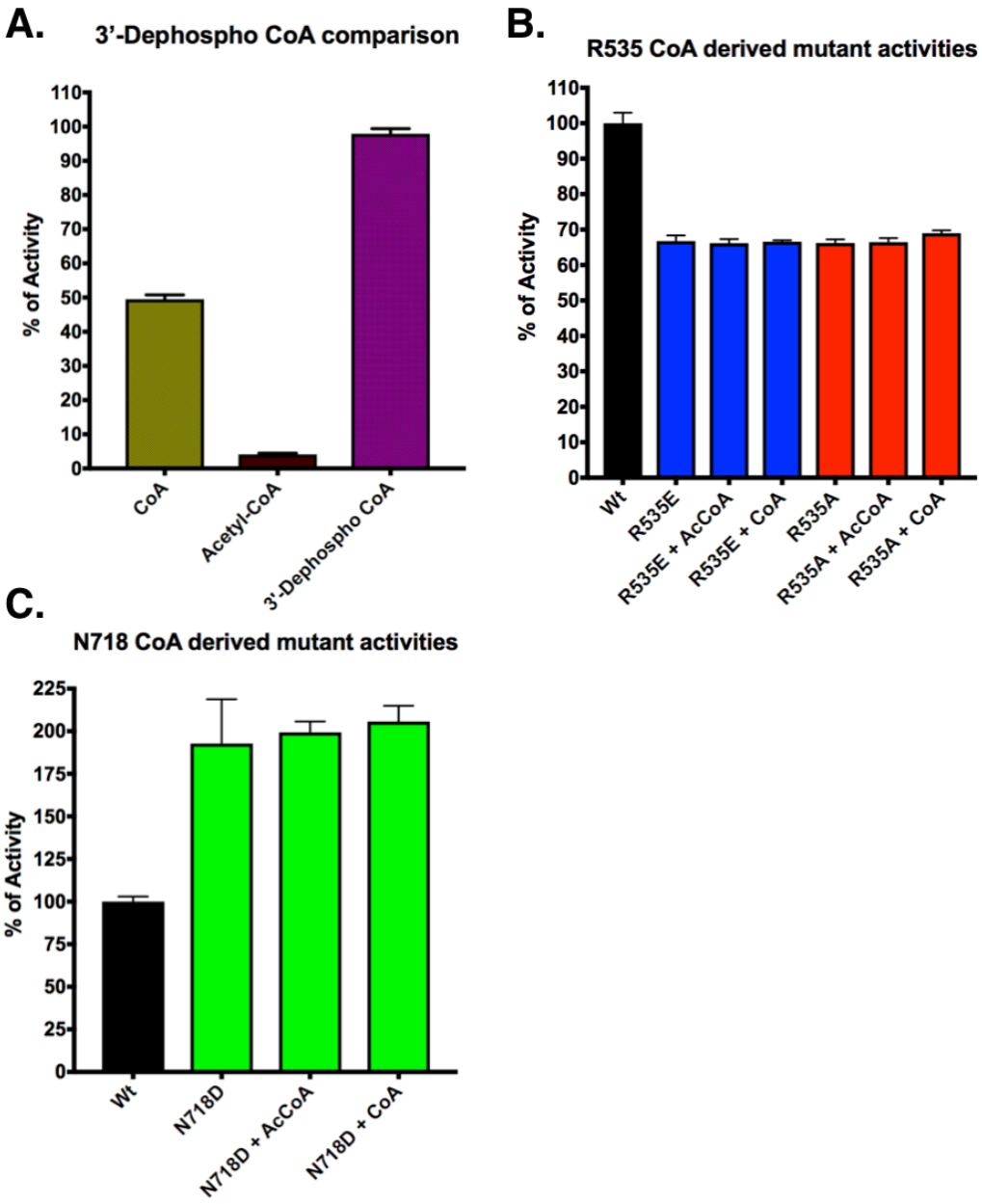


Figure 83. Identifying acetyl-CoA binding site
A. CoA and acetyl-CoA both inhibit the activity of Bd1833; however, a CoA analogue lacking the 3'-phosphate group shows no inhibition. **B.** Mutation of R535 that interacts with the 3'-phophate group alleviates the inhibitory affects of both acetyl-CoA and CoA. The rate of R535A and R535E is reduced compared to WT. **C.** N718D prevents acetyl-CoA and CoA at the thiol-group binding site. As a result CoA and acetyl-CoA no longer inhibit and the activity of Bd1833-R718D is increased compared to Wt. All data points were measured in at least triplicates. The error bars represent the standard deviation.

5.2.8.5 Inhibitor Potency

The inhibitory type and potency of acetyl-CoA and CoA was determined using inhibitor assays performed using Bd1833-Apo (urea washed Bd1833). The turnover of $NADP^+$ at a fixed substrate/co-factor and enzyme concentration while varying inhibitor concentration was used to determine the IC_{50} value, which was converted into K_i. Michaelis-Menten parameters for $NADP^+$ and L-malate substrates were also determined for Bd1833 at increasing concentrations of [inhibitor].

The mechanism of enzyme inhibition was probed by measuring Michaelis-Menten parameters for both substrates (L-malate and $NADP^+$) at increasing concentrations of inhibitor (acetyl-CoA or CoA). Saturation of Michaelis-Menten plots was not feasible when investigating the effects of inhibitor with variable L-malate substrate concentrations (reduced rates beyond 10 mM L-malate), therefore an accurate K_i could not be obtained. However, Lineweaver-Burk plots demonstrated noncompetitive inhibition. Moreover, noncompetitive inhibition was observed when $NADP^+$ was investigated as the varied substrate with increasing concentrations of acetyl-CoA (the V_{max} decreased and the apparent K_M was unaltered), with $K_{i\ (acetyl\text{-}CoA)} = 0.703 \pm 0.071$ μM. (Figure 84).

The IC_{50} values of acetyl-CoA and CoA were determined as a measure of potency of the inhibitors (Figure 84). The IC_{50} were then converted into K_i values using knowledge of inhibition type. The K_i value obtained for acetyl-CoA (0.453 ± 0.019 μM) is 45 times smaller than the K_i determined for CoA (20.2 ± 0.89 μM). This finding reveals that Bd1833 activity is far more sensitive to changes in the concentration of acetyl-CoA than CoA. At low concentrations of inhibitor (2 μM ≤ [I] < 10 μM) Bd1833 is almost

completely inhibited in the presence of acetyl-CoA yet has almost maximal activity in the presence of CoA. Importantly the K_i values obtained for acetyl-CoA from the two assays are relatively similar.

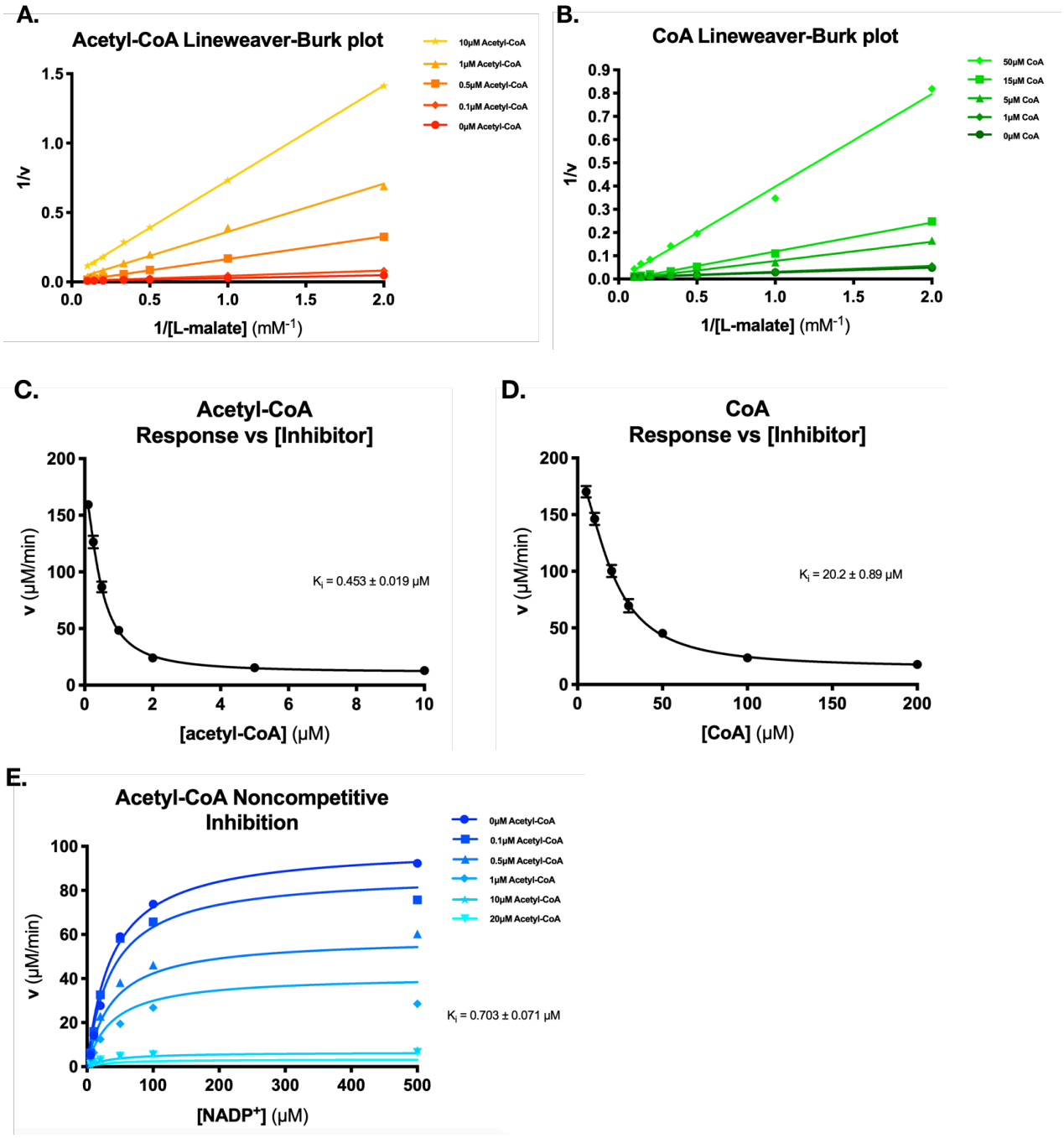


Figure 84. Inhibitor Potency
A. Lineweaver-Burk plot demonstrating the effect of acetyl-CoA on Bd1833 Michaelis-Menten kinetics **B.** Lineweaver-Burk plot demonstrating the effect of CoA on Bd1833 Michaelis-Menten kinetics **C.** Plot of response vs [inhibitor], demonstrating the effect of acetyl-CoA concentration on the activity of Bd1833 at a fixed substrate concentration (IC_{50}) **D.** Plot of response vs [inhibitor], demonstrating the effect of CoA concentration of the activity of Bd1833 at a fixed substrate concentration (IC_{50}) **E.** Plot of Michaelis-Menten data for Bd1833 ($NADP^+$ varied substrate) in the presence of increasing [acetyl-CoA]. The data were fit to the noncompetitive inhibition model. All data points were measured in at least triplicates. The error bars represent the standard deviation.

5.2.8.6 Bd1833- Δloop kinetics

The structure of Bd1833-ME revealed a 17aa loop region (subdomain-D), located beneath subdomain-B (catalytic core region). However, in structures of full-length Bd1833 the location of the subdomain-D cannot be determined and is replaced by helix-21 (N-terminal helix of Pta domain). Subdomain-D appears as a relative insertion only in the primary sequences of multi-modular ME. We therefore suspected that the loop is important for regulating multi-modular ME activity. Consequently we designed and created a chimera of Bd1833 (Bd1833-Δloop = lacking subdomain-D), which was replaced by the equivalent region from AYWB-ME (PDB:5CEE), (Figure 85).

Bd1833-Δloop had a considerably reduced catalytic activity compared to Bd1833, demonstrated by the decreased k_{cat} (Table 14). Moreover, the $K_{M(Malate)}$ was also increased 4-fold for Bd1833-Δloop. These results suggest that subdomain-D is important for Bd1833 activity. Interestingly, Bd1833-Δloop was insensitive to acetyl-CoA and CoA inhibition, suggesting subdomain-D participates in conveying allosteric messages between ME and Pta domains (Figure 86).

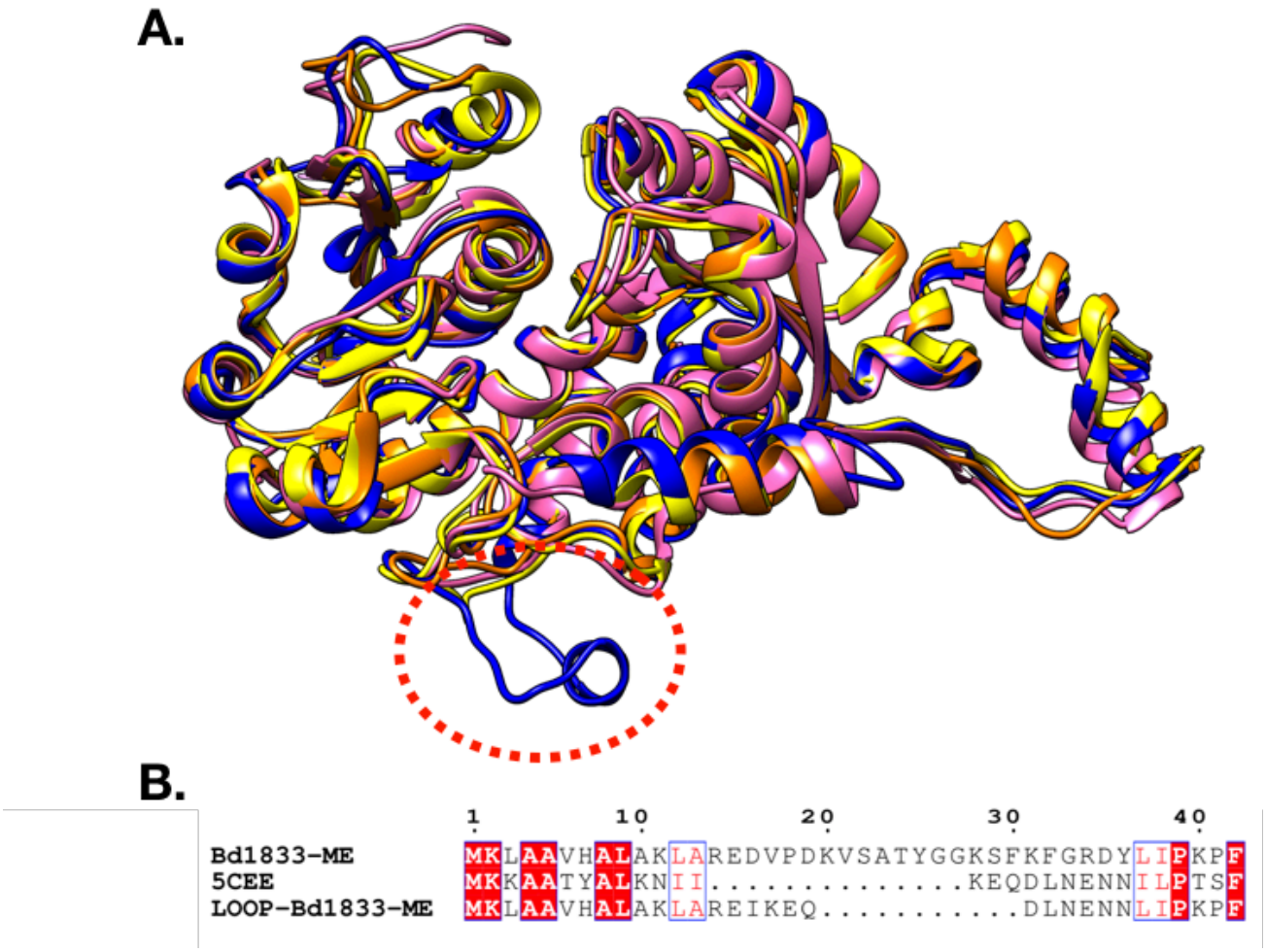


Figure 85. Multi-modular ME specific loop
A. The structure of Bd1833-ME (blue) superimposed with the prokaryotic MEs (pink, PDB:2A9F, length = 337 residues, rmsd = 1.78Å), (yellow, PDB:5CEE, length = 379 residues, rmsd = 1.09Å) and (orange, PDB:2HAE, length = 348 residues, rmsd = 1.58Å) using SSM superpose tool in coot. The loop region highlighted by the red dashed circle is specific to Bd1833 and other multi-modular malic enzymes. **B.** An alignment of the multi-modular specific loop region, indicating the sequence used to create the chimera of Bd1833-Δloop.

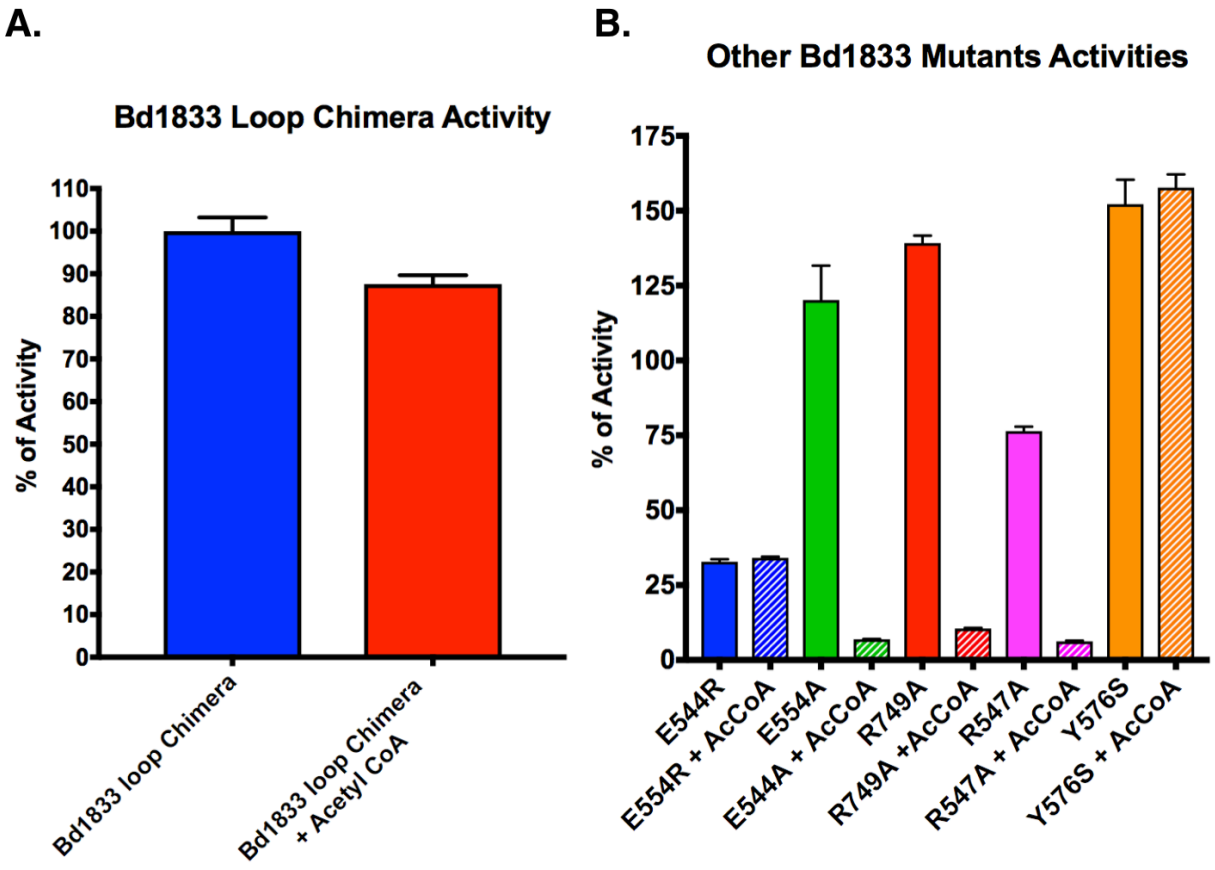


Figure 86. Bd1833 mutants activity and sensitivity to acetyl-CoA
A. The activity of Bd188-Δloop with no acetyl-CoA and in the presence of acetyl-CoA. **B.** This Figure presents the activities of a series of Bd1833 mutants and their sensitivity to acetyl-CoA. All activities are displayed as a percentage of the Bd1833 Wt. E544R shows a decreased activity but is insensitive to acetyl-CoA; however, E544A remains sensitive an has an unaffected activity. Y576S is insensitive to acetyl-CoA and has an increased activity. Furthermore, R749 and R547 mutants located at the hexamer interface remain sensitive to acetyl-CoA. All data points were measured in at least triplicates. The error bars represent the standard deviation.

5.2.8.7 Other Bd1833 mutants

A series of Bd1833 mutants were designed to unpick the mechanism behind allosteric regulation by acetyl-CoA and probe the importance of the hexamer interface for activity. The activity of Bd1833-mutants were determined in standard assay buffer with and without the addition of acetyl-CoA. The activities were compared to Bd1833 and displayed as a percentage of Bd1833 activity in standard assay buffer (Figure 86).

Bd1833-E544R (mutation designed to interfere with hexamer formation) had reduced activity compared to Bd1833. However; Bd1833-E544R was also insensitive to inhibition by acetyl-CoA. The kinetic parameters for Bd1833-E544R were determined; demonstrating a reduced k_{cat} (similar to Bd1833-R535E and Bd1833-R535A) and increased $K_{M(Malate)}$ (similar to Bd1833-Δloop). These results suggest that hexamer formation is important for Bd1833 activity. Furthermore, replacement of E544 with an alanine residue (Bd1833-E544A), which was ineffective at destabilising the hexamer interface (SEC results), maintains catalytic activity and sensitivity to acetyl-CoA inhibition. Moreover, Bd1833-R547A (residue at hexamer interface) has a slightly reduced catalytic activity but remains sensitive to acetyl-CoA, and Bd1833-R749A (located at hexamer interface, interacts with E544 and is a strictly conserved Pta residue, involved in acetyl-phosphate binding) has increased catalytic activity and remains sensitive to acetyl-CoA. Bd1833-Y576S (converted back to Pta consensus, located close to acetyl-CoA thiol group) had increased catalytic activity and was also insensitive to acetyl-CoA inhibition.

5.2.8.8 Pyruvate reductive carboxylation

We investigated the reverse reaction of Bd1833 (pyruvate reductive carboxylation) to confirm the direction of the most favourable reaction. The parameters obtained from this experiment are displayed in Figure 87. The rates obtained for pyruvate reductive carboxylation were >120 times lower than those observed from the malate oxidative decarboxylation reaction suggesting that Bd1833 strongly favours the forward reaction where pyruvate and NADPH are the products.

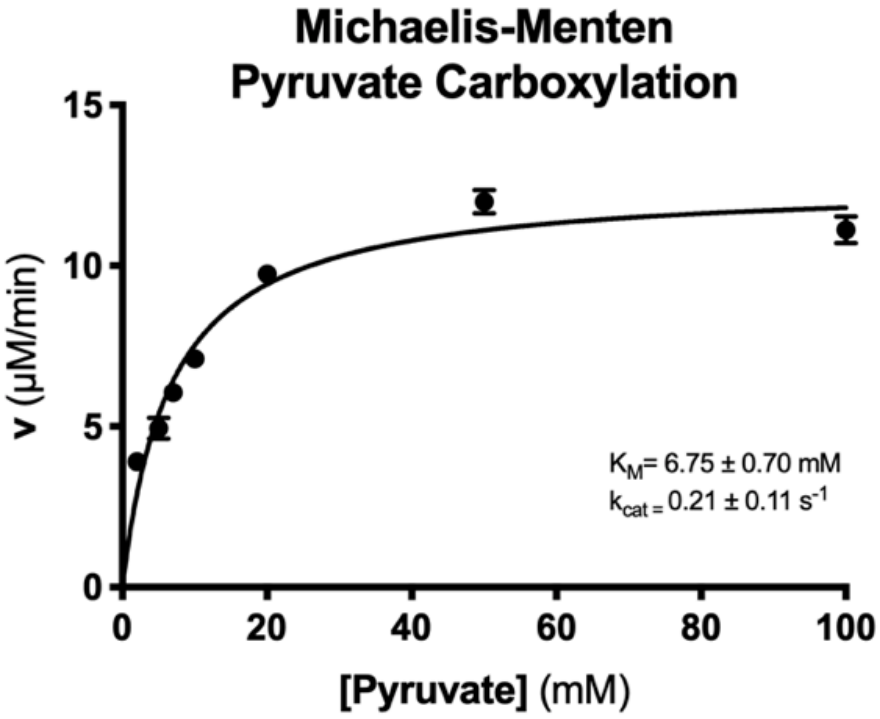


Figure 87. Pyruvate carboxylation reaction
Kinetic data for Bd1833 performing the reverse reaction (pyruvate carboxylation) fit to the Michaelis-Menten equation. Activity is reduced in the reverse direction, which suggests that Bd1833 favours the forward reaction and formation of pyruvate. All data points were measured in at least triplicates. The error bars represent the standard deviation.

5.2.8.9 Bd1833-SKB kinetics

Bd1833-SKB possesses similar kinetic parameters for the malate oxidative decarboxylation reaction when compared to Bd1833-Apo (urea washed Bd1833), (displayed in Figure 88 and Table 14). However, Bd1833-SKB differs in its sensitivity to CoA and acetyl-CoA inhibition. Bd1833-SKB activity is unaffected by a CoA concentration that reduced the activity of Bd1833 by ~50%. Moreover, Bd1833-SKB retains sensitivity to acetyl-CoA inhibition, albeit to a lesser degree compared to Bd1833. The K_i of acetyl-CoA (21.3 ± 0.84 μM) for Bd1833-SKB is ~50-fold higher than the K_i of acetyl-CoA for Bd1833 (0.453 ± 0.019 μM). The K_i of acetyl-CoA for Bd1833-SKB is similar to the Ki of CoA for Bd1833 (20.2 ± 0.89 μM) demonstrating Bd1833 interacts more tightly with CoA and acetyl-CoA. The increase in K_i value of acetyl-CoA can be partly explained by the substitution of R581K in the primary sequence of Bd1833-SKB; resulting in loss of interaction with the pyrophosphate group of CoA.

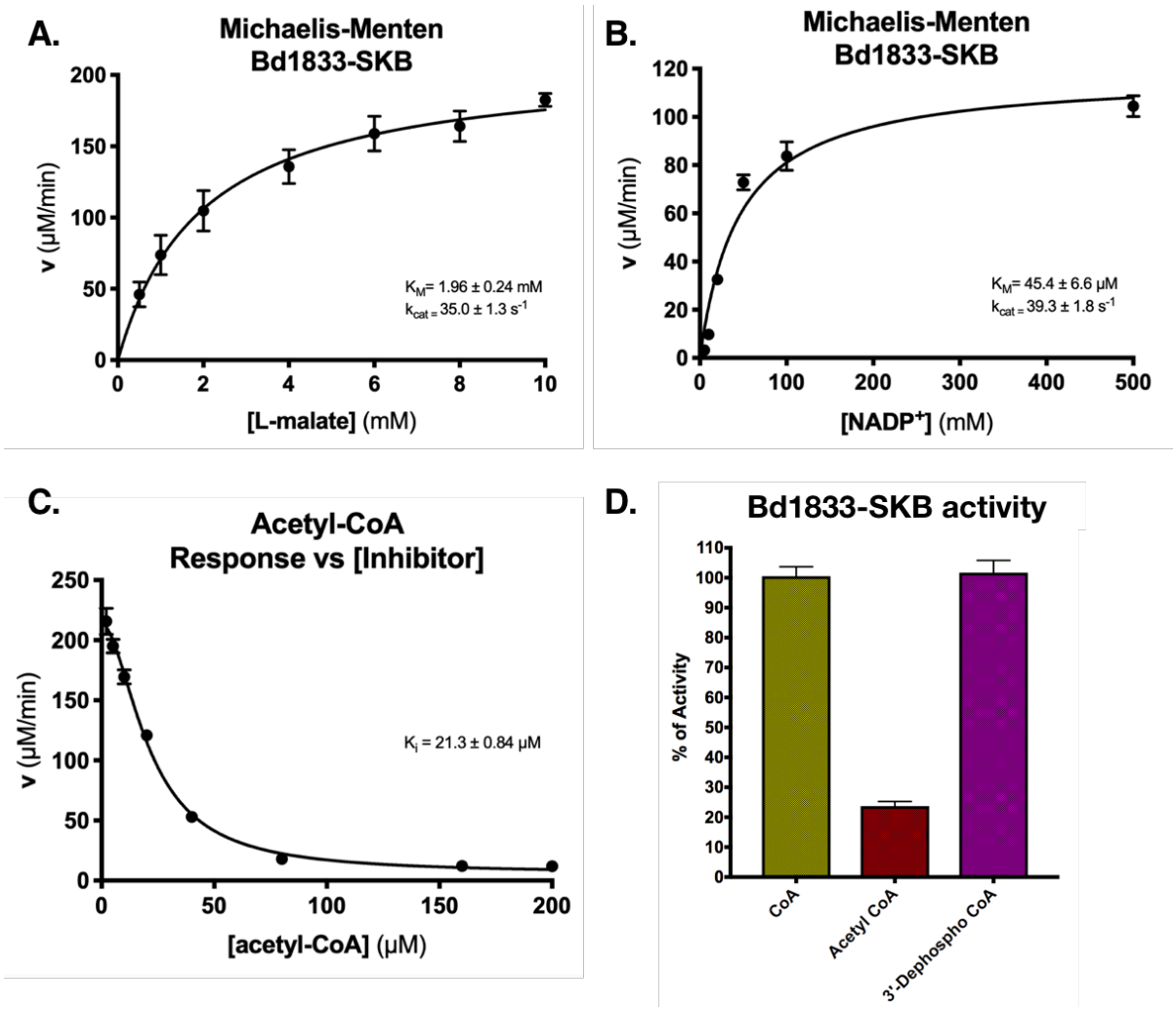


Figure 88. Bd1833-SKB kinetic parameters
A. Oxidative decarboxylation reaction with malate as substrate fit to Michaelis-Menten equation. **B.** Oxidative decarboxylation reaction with $NADP^+$ as substrate fit to Michaelis-Menten equation. **C.** Response vs [acetyl-CoA] data were measured for Bd1833-SKB activity. The K_i value obtained was increased compared to Bd1833 **D.** Bd1833-SKB is not inhibited by CoA at physiological concentrations, whereas acetyl-CoA is still an inhibitor. All data points were measured in at least triplicates. The error bars represent the standard deviation.

5.2.9 HPLC analysis of bound ligands

5.2.9.1 Bd1833 co-purification with CoA and acetyl-CoA

HPLC was used to confirm the identity of co-purifying and co-crystallising ligands observed in Bd1833 structures. The co-purified ligands were found to be a mixture of CoA and acetyl-CoA, with acetyl-CoA the more abundant ligand (Figure 89). CoA and acetyl-CoA standards were used to determine an approximation of the concentration of co-purified ligands. Approximately 45 μM acetyl-CoA and 15 μM CoA was detected from a 100 μM purified sample of Bd1833. This value differs in Bd1833-Pta samples, wherein 90 μM of acetyl-CoA and 10 μM CoA are present in a 100 μM sample.

The three Bd1833 mutants designed to abolish CoA binding (R535A, R535E and N718D) were analysed by HPLC. All three Bd1833 mutations resulted in diminished levels of CoA and acetyl-CoA (<5 μM).

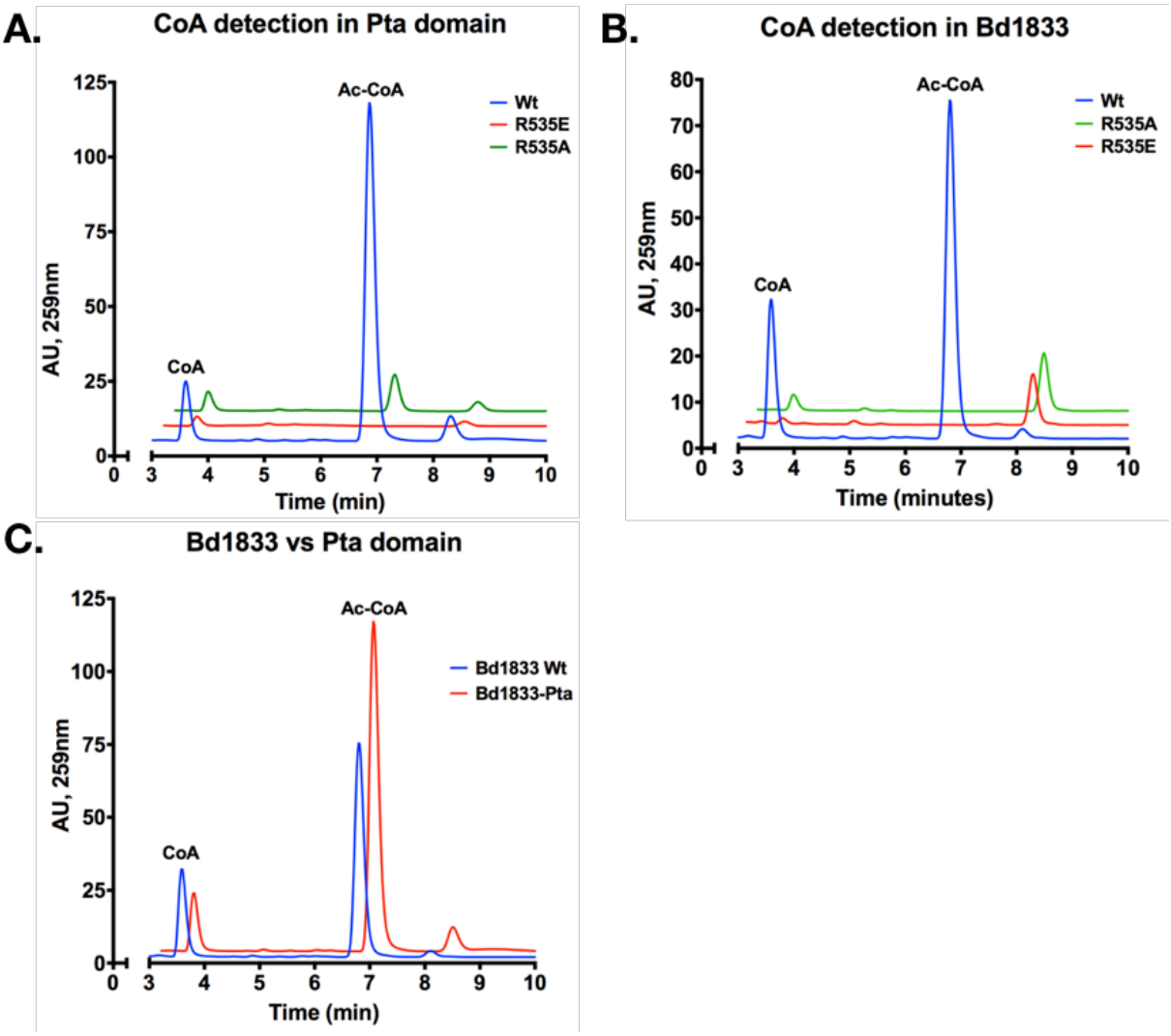


Figure 89. HPLC results
A. Bd1833-Pta CoA and acetyl-CoA HPLC traces in comparison to R535A and R535E mutants. **B.** Bd1833 full-length CoA and acetyl-CoA HPLC traces in comparison to R535A and R535E mutants. **C.** Comparison of Bd1833 and Bd1833-Pta HPLC traces demonstrate the ratio difference between co-purifying ligands. Bd1833-Pta co-purifies with more acetyl-CoA and less CoA than full-length Bd1833.

5.2.9.2 Producing Bd1833-Apo

A ligand free version of Bd1833 (Bd1833-Apo) was required to investigate allosteric regulation by acetyl-CoA without complications from co-purifying ligands. To produce Bd1833-Apo an additional mild denaturing wash step was added to the purification procedure of Bd1833. Illustrated in Figure 90 are the results from two different wash procedures performed on Bd1833 and Bd1833-Pta purification. The most effective

wash procedure was to pass >50CV of (50 mM HEPES pH 7.5, 300 mM NaCl, 2.5M Urea, 2M KBr) over immobilised Bd1833, which resulted in <5 µM of co-purified ligand in a 100 µM sample (>95% of protein is thus assumed to be ligand free).

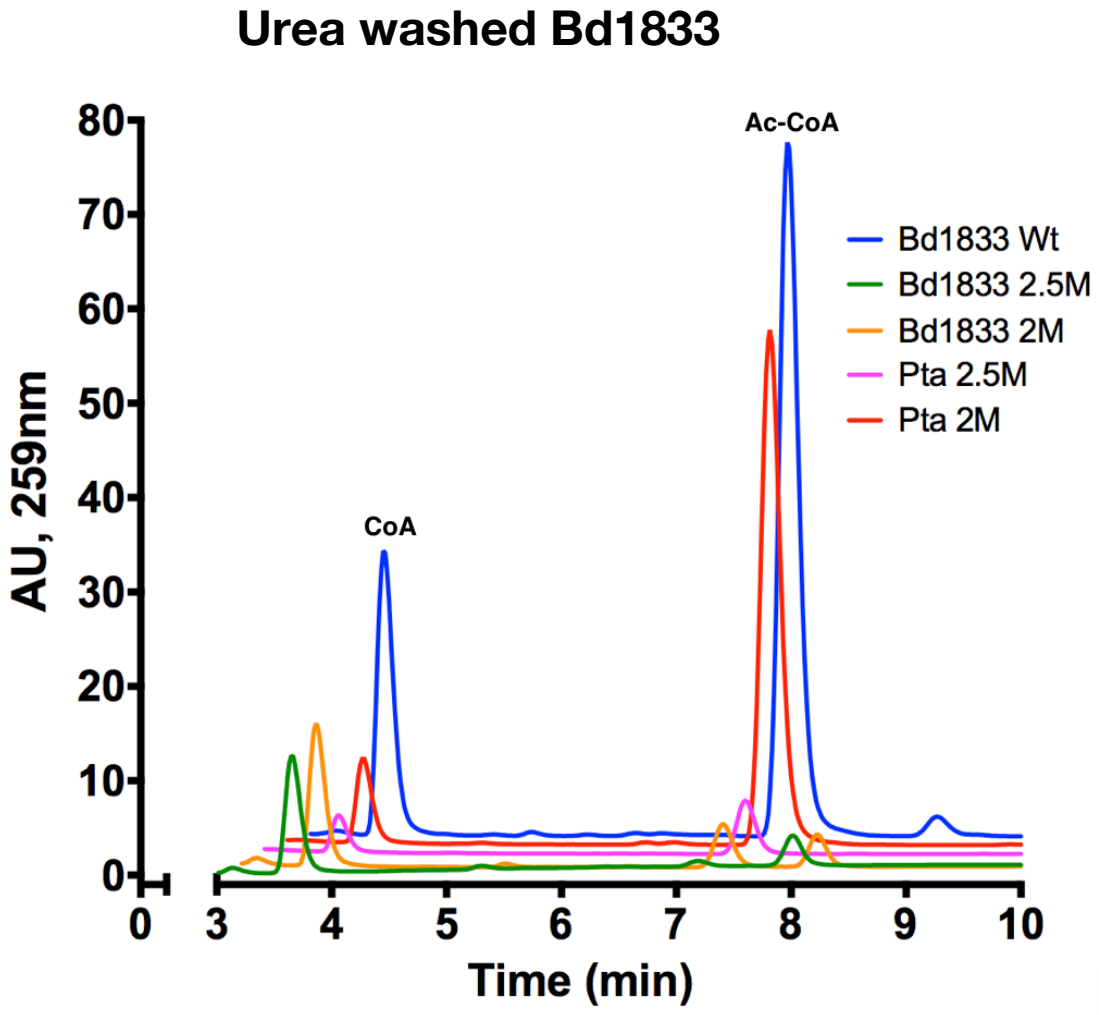


Figure 90. CoA and acetyl-CoA removal by denaturing wash procedure
Removal of CoA and acetyl-CoA from Bd1833 and Bd1833-Pta was achieved by incorporating an addition denaturing wash procedure to the purification protocol. Best results were achieved with 2.5M urea buffer (green and pink traces).

5.2.9.3 HPLC results for other Bd1833 mutants

HPLC was used to determine the CoA:acetyl-CoA profile for each Bd1833 mutation used in this investigation (Figure 91). HPLC analysis revealed R749A and R547 mutations (at hexamer interface) had slightly reduced levels of acetyl-CoA compared to Bd1833. Bd1833-E544R (disrupts hexamer) co-purified with minimal amounts of CoA and acetyl-CoA, whereas Bd1833-E544A (disrupts hexamer) co-purified with levels of ligand comparable to Bd1833. Furthermore the same mutation in Pta domain only (Pta-E544R) also co-purified with levels of ligand comparable to Bd1833.

Bd1833-Y576S and Bd1833-Y576C mutants (designed to interact with the CoA thiol group) co-purified selectively with CoA. These two mutants demonstrate that selectivity between CoA and acetyl-CoA can be achieved if mutations within thiol/acetyl group binding pocket are introduced.

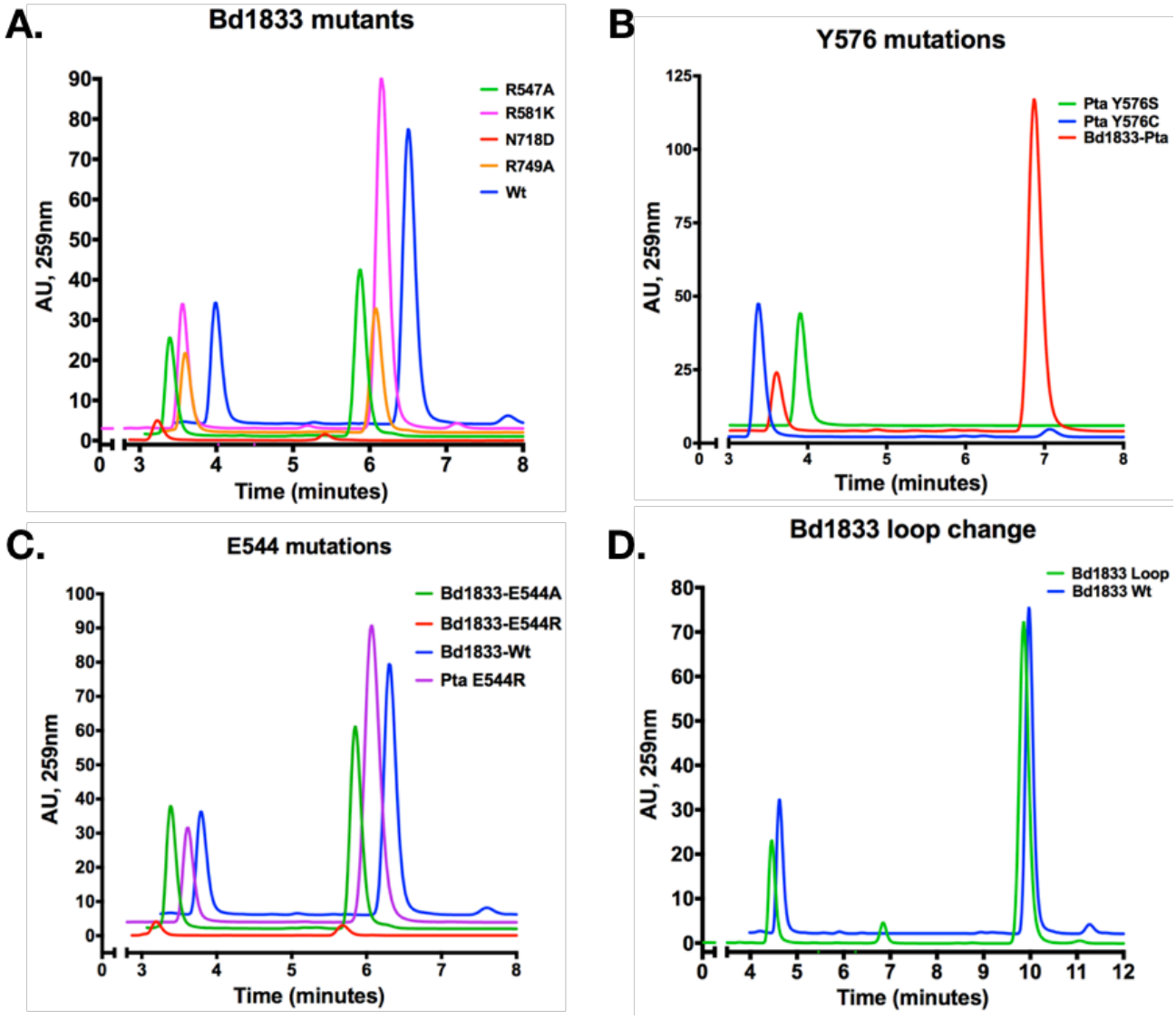


Figure 91. HPLC traces for Bd1833 and Bd1833-Pta mutants
A. Traces of Bd1833 mutants revealed all co-purify with CoA and acetyl-CoA except for N718D. **B.** Bd1833-Pta Y576S and Y576C selectively co-purify with CoA suggesting acetyl-CoA binding is unfavoured. **C.** Bd1833-E544R does not co-purify with ligand; however, Bd1833-E544A and Pta-E544R do co-purify with ligand. **D.** Bd1833-Δloop co-purifies with both ligands. This is the only case of a mutant that is insensitive to acetyl-CoA inhibition and still co-purifies with acetyl-CoA.

5.2.9.4 HPLC Bd1833-Δloop

Bd1833-Δloop co-purified with comparable amounts of CoA and acetyl-CoA as Bd1833 (Figure 91). This finding demonstrates the first of our Bd1833 mutations that co-purify with ligand but are insensitive to acetyl-CoA inhibition.

5.2.9.5 HPLC results for Bd1833-SKB

Bd1833-SKB did not co-purify with CoA or acetyl-CoA, which is strikingly different to Bd1833. Interestingly Pta-SKB (truncation of Bd1833-SKB) co-purified with levels of CoA and acetyl-CoA comparable to Bd1833-Pta.

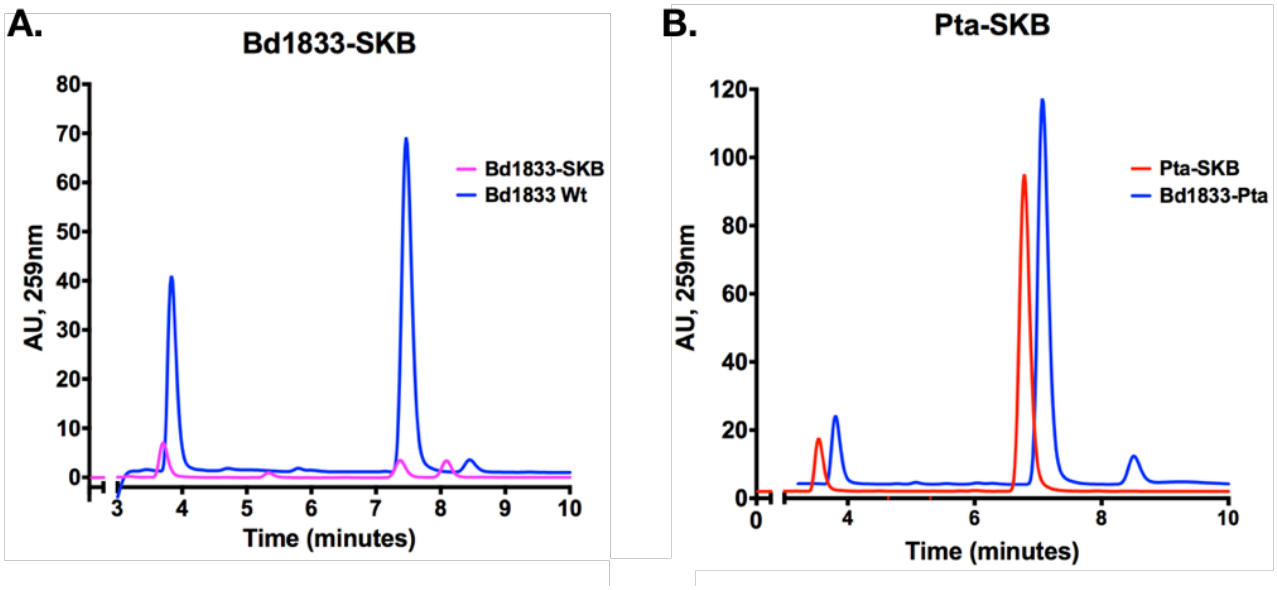


Figure 92. HPLC results for Bd1833-SKB and Bd1833-SKB-Pta
A. Bd1833-SKB did not co-purify with CoA or acetyl-CoA. **B.** Pta-SKB co-purified with CoA and acetyl-CoA.

5.2.10 Size exclusion results

Size exclusion (SEC) was used to determine the oligomeric state of Bd1833 and its mutations in solution, (Figure 93). SEC results demonstrate full-length Bd1833 forms a hexamer in solution (peak at 130 ml), which is also the case for Bd1833-Pta (peak at 155 ml); however, Bd1833-ME forms a dimer in solution (peak at 185 ml). These results confirm the oligomeric complexes observed in the crystal structures.

Bd1833 mutations that disrupted the CoA binding groove in the Pta domain (R535E, Y576S and N718D) were also evaluated using SEC. These CoA binding site interfering mutants were all observed to exist as hexamers in solution.

The hexamer destabilising Bd1833 mutants (E544R and E544A) were also analysed using SEC to confirm that the hexamer complex had been disrupted significantly enough to prevent hexamer formation. Bd1833-E544R elutes as a dimer (peak at 160 ml) confirming hexamer formation is prevented by this mutation. However, Bd1833-E544A elutes as a hexamer (peak at 130 ml) suggesting the mutation does not destabilise the hexamer interface sufficiently to prevent hexamer formation. Furthermore, the same E544R mutation in truncated Bd1833 (Pta-E554R) forms a hexamer in solution. This finding was unexpected and suggests the hexamer interface is under more mechanical stress in the full-length protein therefore requiring increased interactions to stabilise the hexamer complex.

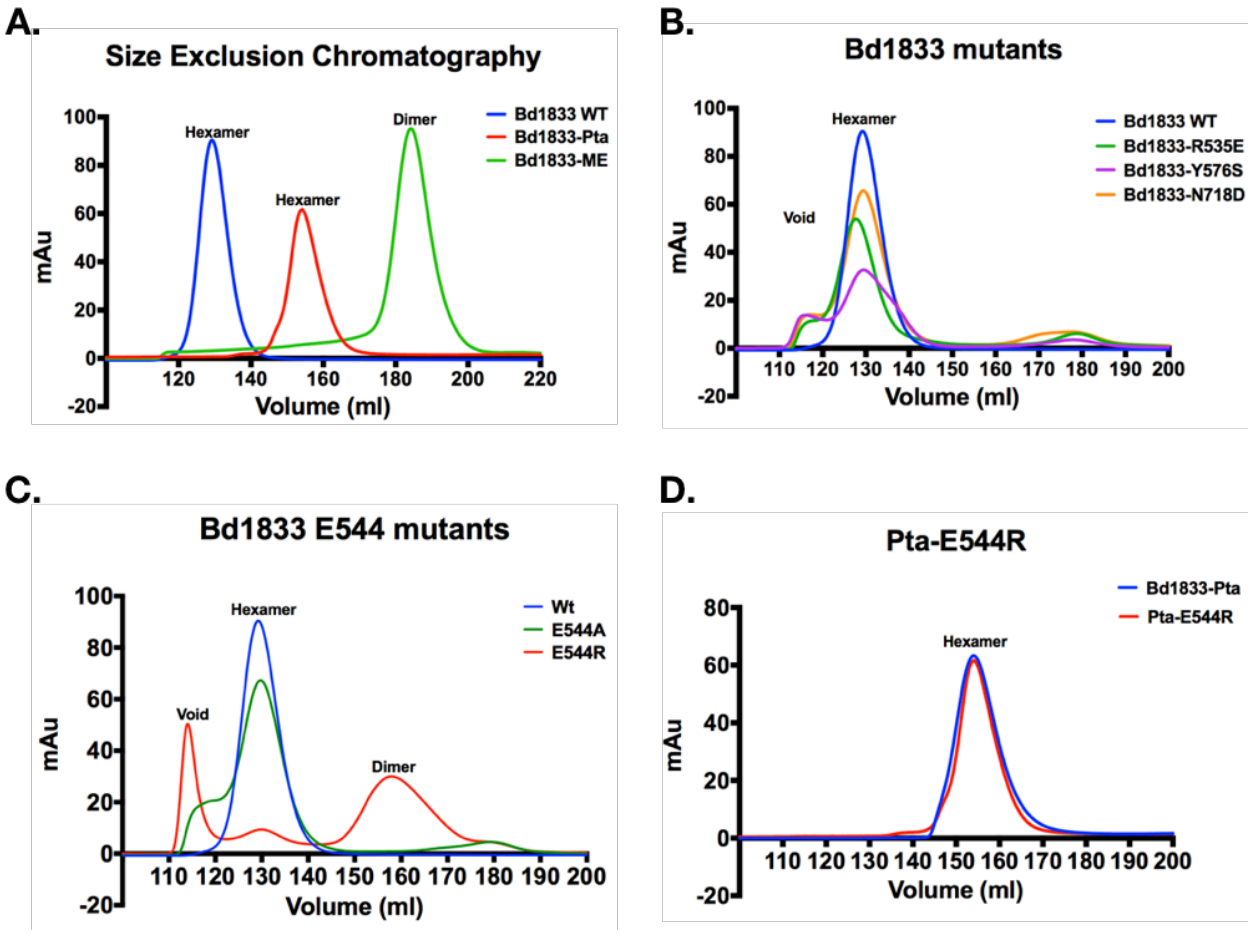
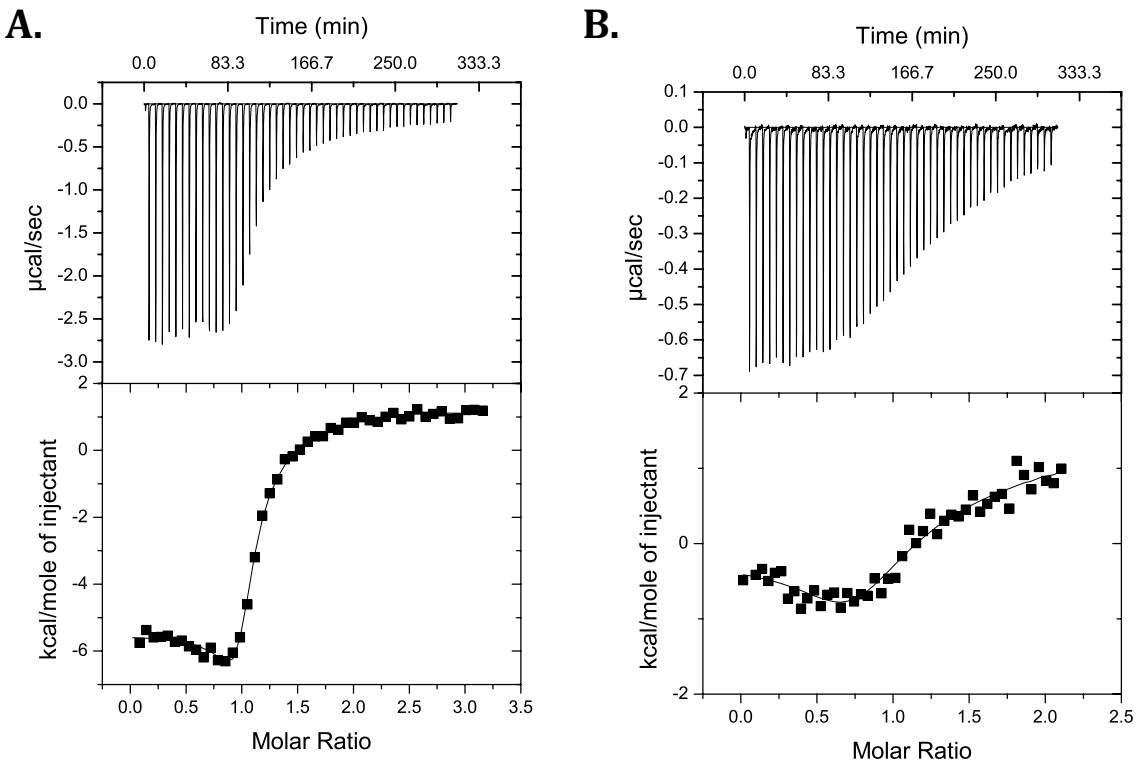


Figure 93. Size exclusion results
A. Bd1833 and Bd1833-Pta exist as a hexamer in solution, while Bd1833-ME exists as a dimer. **B.** Bd1833 mutants that we identified to affect hexamer interface also exist as hexamers in solution. **C.** Bd1833-E544R disrupts the hexamer interface to produce a mutant that forms a dimer in solution. Bd1833-E544A does not destabilise the hexamer interface sufficiently and still forms a hexamer. **D.** The same E544R mutation in the Pta domain alone (Pta-E544R) does not result in dimer formation.

5.2.11 ITC results

ITC was used to investigate the nature of acetyl-CoA and CoA binding to Bd1833-Pta. The ITC data collected for Bd1833-Pta interaction with CoA and acetyl-CoA are difficult to interpret; complicated by residual bound ligands after purification and the complex network of interactions in the hexamer complex. The data show cooperative binding and suggests more than one acetyl-CoA binding site is required to fit the data. Structural evidence shows that Bd1833-Pta has six identical acetyl-CoA binding sites and acetyl-CoA binding is accompanied by a conformational change that affects the hexamer interface (which can propagate to other protomers). In order to fit the data a six site binding model was used. Fitting the ITC data to a six site sequential binding model gives a good fit (Chi^2/DoF = $1.311e^4$ for acetyl-CoA and Chi^2/DoF =$2.436e^4$); however, this could be a result of increasing the amount of parameters. The data reveals that acetyl-CoA interaction with Bd1833-Pta is more energetically favourable than CoA, due to the larger ΔH values. Acetyl-CoA also reaches saturation at a lower concentration than CoA. The ITC titration curves (upper panel) and binding isotherms (lower panel) are displayed in Figure 94.



C.

Site	Acetyl-CoA K (M)	Acetyl-CoA ΔH (J)	Acetyl-CoA ΔS (J mol⁻¹ K⁻¹)	CoA K (M)	CoA ΔH (J)	CoA ΔS (J mol⁻¹ K⁻¹)
1	5.7e5 ± 6.6e5	-5600 ± 88	7.6	1.8e5 ± 1.1e5	-440 ± 160	23
2	7.9e3 ± 1.6e3	-2.3E4 ± 8.2e3	-58	1.9e4 ± 1.1e4	-550 ± 1.1e4	18
3	2.9e3 ± 4.8e2	1.6e4 ± 5.2e5	71	1.2e5 ± 6.3e4	-7800 ± 7.4e4	-3.0
4	4.7e3 ± 8.4e3	3.9e4 ± 1.5e6	150	4.9e4 ± 2.0e4	1200 ± 3.8e5	25
5	1.5e5 ± 2.6e4	3.8e4 ± 1.3e6	150	1.6e5 ± 1.0e5	1.1e4 ± 4.9e5	61
6	5.4e5 ± 1.0e5	-6.9e4 ± 3.1e5	-210	1.0e5 ± 9.0e4	9500 ± 2.3e5	55

Figure 94. ITC results for Acetyl-CoA and CoA binding to Bd1833-Pta Apo
Upper panels represent the ITC titration curve and the lower panels represent the binding isotherm **A.** Acetyl-CoA binding. **B.** CoA binding. All data are fit to a sequential six-site model. **C.** A tabulated list of values obtained from the fit.

5.2.12 Bd1833 interaction with CdG

MST was used to determine whether Bd1833 interacted with CdG. A K_D value of 26 µM was obtained, suggesting weak non-specific binding. Furthermore CdG did not regulate the activity of Bd1833 in enzyme assays.

5.3 Discussion

5.3.1 Structure and function of multi-modular malic enzymes

The structure of Bd1833 confirms that two distinct domains (ME and Pta domains) compose the full-length structure. Each domain has expected physical similarities to previously isolated prokaryotic ME and Pta enzymes [1, 250]. Bd1833 forms a large hexamer complex, formed by interactions between Pta domains, which arrange into a central triangle shaped hub. Atop of the Pta domains on the flat sides of the dimer is the ME domain linked by a single linker helix. Other multi-modular ME, such as DME and MaeB, are suggested to form higher order oligomers (octomers) and the Pta domain was suspected to be important for their formation [253, 254]. These findings confirm that the Pta domain is central to the formation of higher-order oligomer formation (and suggest, via strong sequence identity, that the octomeric assembly for others was presumably an over-estimation).

This study has identified that the ME domain of Bd1833 has a unique small loop (subdomain-D), which is a relative insertional mutation, found only in multi-modular ME (Figure 61). Furthermore, subdomain-D is demonstrated to be an important feature regulating Bd1833 activity. Bd1833-Δloop has a reduced k_{cat} compared to Bd1833 suggesting it has an important role in activity. Moreover, the Bd1833-Δloop variant is insensitive to inhibition by acetyl-CoA, but still able to bind the ligand (HPLC), indicating that the allosteric communication behind acetyl-CoA inhibition has been disrupted.

Unlike other published Pta enzyme structures, Bd1833-Pta forms a hexamer complex (a trimer of Pta dimers), which forms the central hub of the full-length Bd1833 structure. Bd1833-Pta dimers form an identical interface to published Pta enzymes. The additional hexamer interface is provided by residues of subdomain-I, not conserved in dimeric Pta enzymes, such as E544-R749 and R537-E679 that form important salt bridges. The hexamer configuration allows a single Bd1833-Pta monomer to make contacts with four other Bd1833-Pta monomers (Figure 69) producing a complex network of contacts and interactions that assist allosteric regulation. Notably the Pta domain appears to be non-catalytic, lacking key residues. However, the residues involved in substrate binding (acetyl-CoA, acetyl-P, CoA) are all conserved [247]. This implies the Pta domain of Bd1833 is required as a sensor of Pta substrates/products, regulating the activity of the ME domain accordingly via allostery.

5.3.2 Bd1833 function / role

The *Bdellovibrio bacteriovorus* lifecycle includes a highly efficient growth phase inside the periplasm of a prey cell. Many of the nutrients required for growth are derived from prey cell material during intraperisplamic growth. A vast hydrolytic arsenal is harnessed by *Bdellovibrio* to provide material for energy generation and growth. Consequently the flux of metabolites entering central metabolic pathways of *Bdellovibrio* is different to non-predatory bacteria. For instance there is an increase in the amount of prey-derived amino acids and nucleotides being used for catabolism [68, 72]. Furthermore *Bdellovibrio* has a full functioning TCA cycle [73] but the lack of upstream sugar-metabolising enzymes, suggests that the TCA cycle is fed by pathways other than conventional glycolysis [24].

The role of Bd1833 in *Bdellovibrio* metabolism is presumed to be similar to MEs in other organisms: subverting the flow of TCA cycle intermediates and producing NADPH, which is used in anabolic reactions such as lipid biosynthesis. Interestingly *Bdellovibrio* lacks a transhydrogenase; an enzyme found in many organisms that is responsible for interconversion of NADH and NADPH, which in turn facilitates anabolic growth [258]. It has been suggested previously that $NADP^+$ -specific MEs have a major role in NADPH generation for use in anabolic pathways [216]. The role of Bd1833 may thus be to contribute to the pool of NADPH required for anabolic reactions and control of redox potential. Furthermore the neighbouring gene of *bd1833, bd1832*, is a chromate transporter. Chromate is a toxic substance that damages cells. Reduction of chromate reduces its toxic effects and has been linked partly to NADPH production by a ME in a chromate hyper-resistant strain of *Pseudomonas corrugate* [259].

Additionally Bd1833 is well positioned to divert the flow of TCA cycle intermediates to pyruvate and ultimately acetyl-CoA that are fed by the increased flux of amino acids and other prey-derived metabolites. There are no oxaloacetate dehydrogenase (OD) or phosphoenolpyruvate carboxylase (PEPC) genes identified in the genome of *Bdellovibrio HD100* or sugar-metabolising enzymes, which are typically used to generate pyruvate/acetyl-CoA in other organisms. Consequently Bd1833 may be integral for anabolic growth in *Bdellovibrio* by providing a metabolic route to acetyl-CoA from prey-derived amino acids that enter the TCA cycle, while simultaneously producing NADPH (Figure 95). This is similar to cancer cells that metabolise glutamine as a major source of energy when glucose is restricted (utilising malic enzyme 1 -ME1) [211, 260]. ME1 in cancer cells therefore provides a useful enzymatic pathway for the

creation of NADPH and pyruvate when glycolysis and PPP pathways are attenuated. Similarly ME1 has been shown to be important for NADPH production and pyruvate cycling in islet β-cells and ultimately controls the secretion of insulin [261]. Acetyl-CoA and OAA are the end products of the reaction, which interestingly we show to also be potent inhibitors of Bd1833 activity.

5.3.3 Bd1833 regulation

Bd1833 activity is potently negatively regulated by acetyl-CoA and OAA, whereas CoA and acetyl-P inhibit to a lesser extent. Acetyl-CoA is the product of the forward reaction catalysed by Pta enzymes, while CoA and acetyl-P are the substrates. These findings therefore suggest that the Pta domain senses the ratio between substrate and products of conventional Pta reactions and regulates the activity of the ME domain accordingly.
The structure of Bd1833 alongside kinetic and mutational studies demonstrates acetyl-CoA interacts at a distal site ~50Å away from the ME active site. Allosteric inhibition by acetyl-CoA is complex, involving a large conformational change in the Pta domain that is conveyed to the ME domain via the multi-modular specific loop (subdomain-D).

The mechanism behind allosteric regulation is initiated by acetyl-CoA binding, which causes a rigid body movement in subdomain-I that rotates the subdomain closer to sub-domain II, restricting the ligand-binding groove and encasing acetyl-CoA. R535 is crucial for acetyl-CoA binding and initiating the conformational change, demonstrated by R535 mutants that are insensitive to inhibition by acetyl-CoA and prevent co-purification of Bd1833 with acetyl-CoA ligand. Moreover the structures of R535 mutants show a localised conformational change in the 3'-phosphate-binding region,

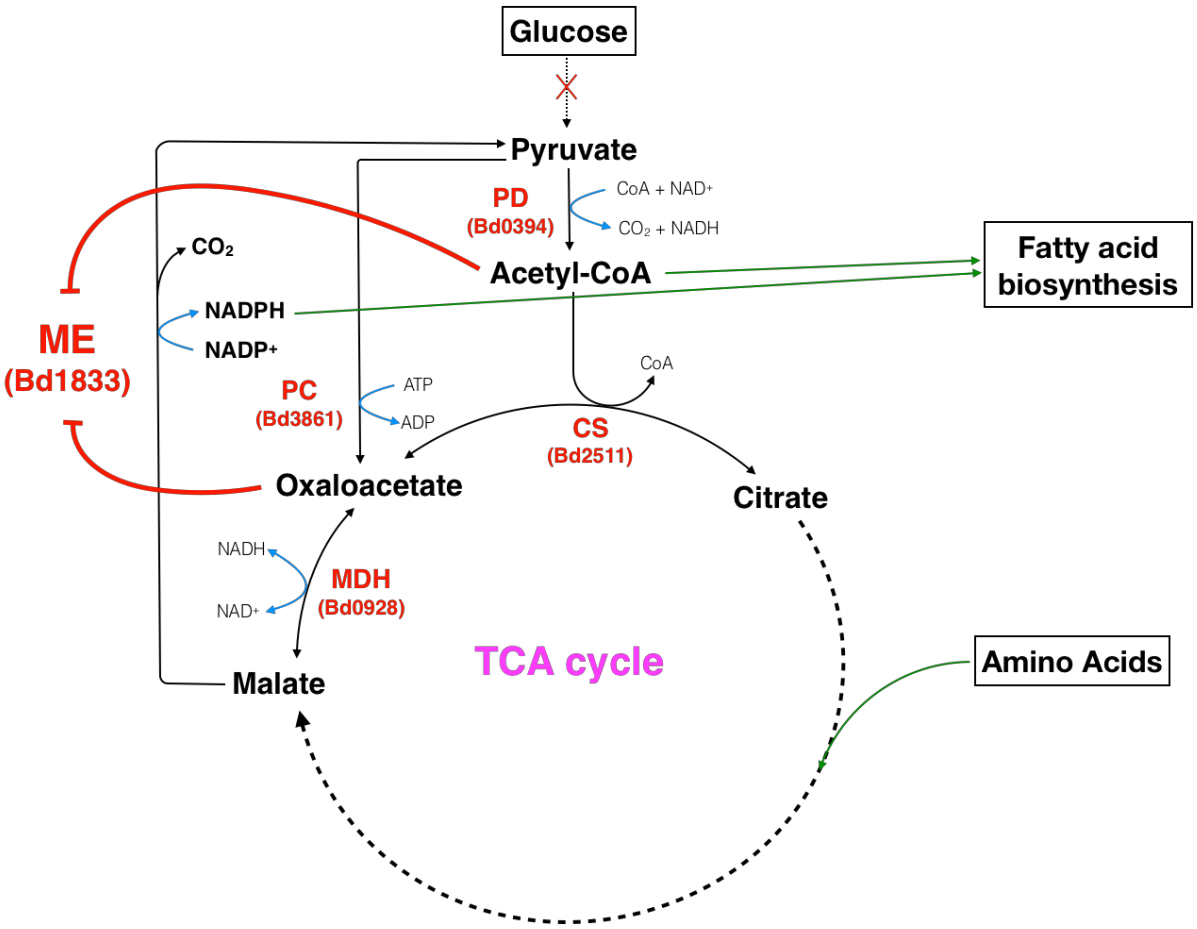


Figure 95. Schematic of Bd1833 role in metabolism
Bdellovibrio lacks sugar-metabolising pathways that are conventionally used to derive pyruvate. An increase flux of TCA cycle intermediates from prey-derived amino acids fuel *Bdellovibrio* metabolism. Bd1833 is central to subverting TCA intermediates to produce pyruvate and NADPH, which is ultimately used in anabolic reactions such as fatty acid synthesis. ME – malic enzyme, MDH – malate dehydrogenase PC – pyruvate carboxylase, PD, pyruvate dehydrogenase, CS – citrate synthase

which destabilises interactions at the hexamer interface such as E544:R749 salt bridge. The N-terminal helix of the Pta domain connects directly to the single linker helix that connects the Pta to the ME domain above. During the conformational change the N-terminal helix of the Pta rotates inwards by 8°, an action that likely impacts on the position of the multi-modular specific loop (subdomain-D). Subdomain-D could not be traced in the structure of full-length Bd1833; however, superimposition of Bd1833-ME (subdomain readily traceable) demonstrates that subdomain-D folds beneath the ME

domain, a position that clashes with the Pta domains N-terminal helix in full-length Bd1833 (Figure 78). Importantly Bd1833-Δloop (chimera deletion of subdomain-D) is not inhibited by acetyl-CoA but still retains the ability to bind ligand (HPLC). These results demonstrate that subdomain-D is important for allostery and that it is able to adopt multiple conformations, only one of which has been captured. The movement of subdomain-D is suspected to sequentially causes the conformational change observed in the ME domain in full-length Bd1833 structures. The structure of full-length Bd1833 (in complex with acetyl-CoA) thus represents an inhibited form. Superimposition of Bd1833-ME with full-length Bd1833 demonstrates a 9° rigid body rotational conformational change in ME subdomain-C ($NADP^+$-binding domain), which alters the active site. Similar ME conformational changes have been reported in human-ME (despite completely different appended regulatory domains), where domain-C ($NADP^+$-binding domain) undergoes an 11° rigid body movement, repositioning the cofactor and active-site residues, priming it for catalysis [225]. The ME domain conformational change likely reduces the accessibility of substrate to the active site; however, the reverse (products become trapped) is also plausible. Kinetic data suggest acetyl-CoA obeys noncompetitive allosteric inhibition and does not discriminate between free enzyme and enzyme substrate complex states, demonstrated by reduced k_{cat} but unaffected K_M in the presence of inhibitor.

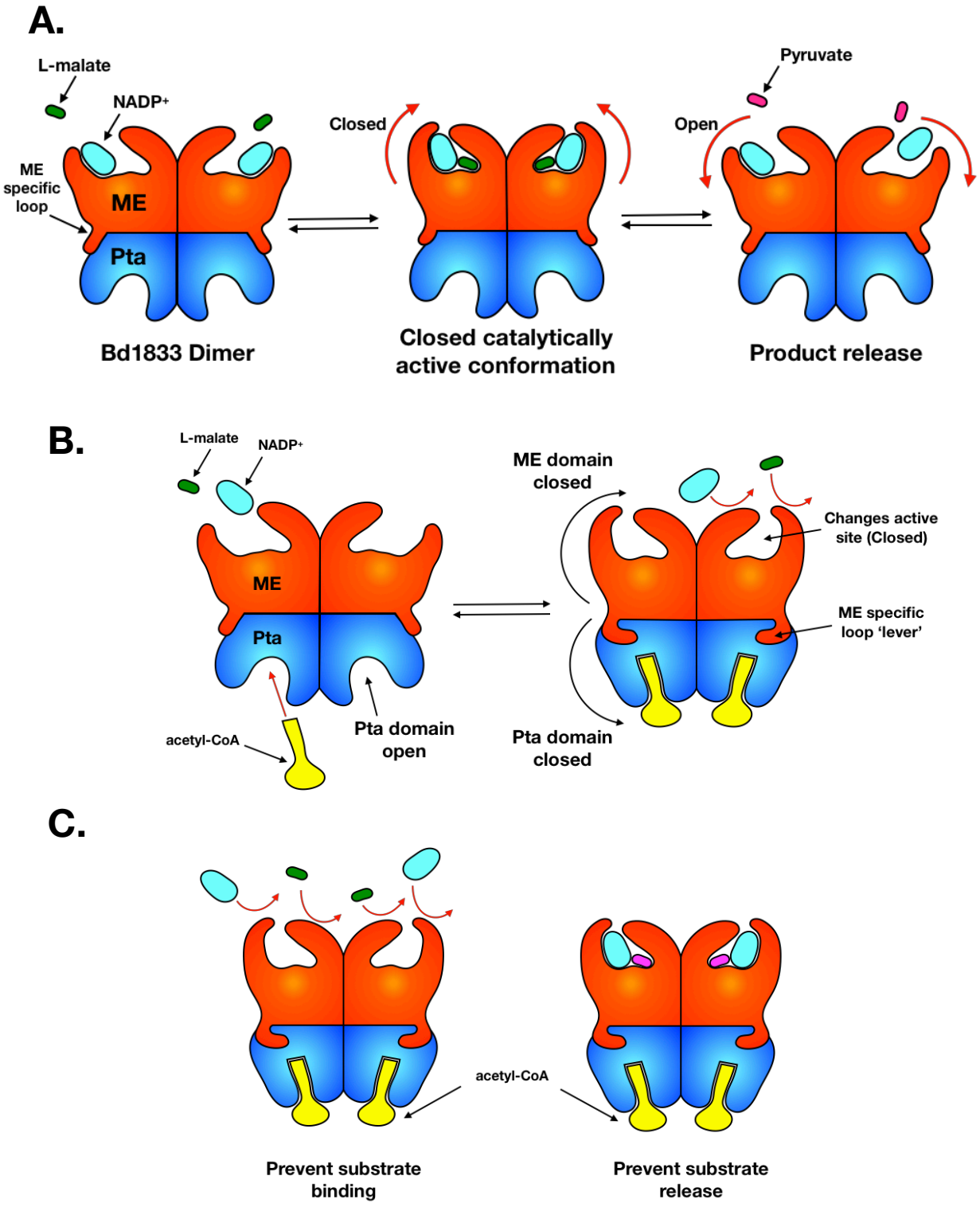


Figure 96. Cartoon of Bd1833 mechanism
A. Mechanism of Bd1833 ME activity. $NADP^+$ binds first then L-malate binds which initiates a conformational change and closes the active site. After the reaction is complete the ME domain opens and releases the products. **B.** Inhibition by acetyl-CoA. Acetyl-CoA binding initiates a conformational change in the Pta domain, which is conveyed to a conformational change in the ME domain via movement of the multi-modular specific loop. **C.** Acetyl-CoA shows non-competitive inhibition and is equally likely to bind the enzyme:substrate complex and free enzyme.

In addition to an acetyl-CoA induced conformational change of the active site, the hexamer complex is also affected. The nine structures of Bd1833 (including truncations, mutants and homologues) all assemble into crystallographic hexamers with dihedral D3 symmetry (also hexamers in solution). In each of these structures contacts at the hexamer interface are variable, with the most extensive hexamer interface observed for Bd1833-Pta (acetyl-CoA bound, ~600Å). In structures lacking acetyl-CoA (Apo and acetyl-CoA binding site mutations) the hexamer interface is reduced (mainly the vertex interface ~300Å) and many interactions are lost including salt bridges between E544 & R749 and D551 & R547. Interestingly, mutations that prevent hexamer formation such as Bd1833-E544R and Bd1833-ME have reduced k_{cat}. Furthermore, acetyl-CoA does not inhibit Bd1833-E544R that forms a dimer. This suggests that hexamer formation is not only important for activity but also for allostery to be properly propagated. Interestingly, allostery is common in other proteins that form dihedral D3 symmetrical homomers [262]. It has been proposed that the isologous interfaces formed in dihedral D3 homomers provide a simple physical mechanism to propagate conformational change [263]. Moreover this type of regulation is enriched in metabolic enzymes that form large homomers. For example, human ribonucleotide reductase is allosterically regulated by induction into a hexameric complex of dihedral D3 symmetry [264].

5.3.4 Comparison to regulation in other multi-modular malic enzymes

Bd1833 shares homology to other multi-modular MEs such as MaeB, DME and TME. MaeB, DME and TME are the only cases to date of multi-modular MEs that have been investigated kinetically. A conclusive cellular function for these enzymes has yet to be

established, though they are believed to be required for anaplerotic synthesis of pyruvate and generation of NADPH. Interestingly the activities of these enzymes have been shown to respond to similar effectors such as acetyl-CoA [254, 256]. DME is even inhibited by acetyl-, propionyl- & butyryl-CoA. TME is an unusual multi-modular ME and does not appear to be inhibited by either CoA or acetyl-CoA. Bd1833, however, unusually displays inhibition by both CoA and acetyl-CoA, though acetyl-CoA inhibition is far more potent (K_i acetyl-CoA ~45 fold less than K_i CoA). This is also supported by Bd1833 co-purification with both CoA and acetyl-CoA ligands, where acetyl-CoA is more favourably bound. Moreover ITC confirms that acetyl-CoA binding is more energetically favourable (larger ΔH values) and saturation is reached at a lower concentration. These findings can be explained by the additional interactions acetyl-CoA makes at the binding site in comparison to those of CoA. The acetate group makes a H-bond with Q683 and is accommodated into a polar, hydrated pocket.

Comparison of the binding-site pocket residues between multi-modular MEs can partly explain the difference in CoA/acetyl-CoA sensitivity (Figure 97). The Bd1833 acetyl-CoA binding pocket is more closely related to catalytic Pta enzymes, whereas MaeB, DME and TME show more substitution in key residues. TME has a relative mutation in residue R581 (Bd1833 numbering), which has been exchanged for a glutamate. In Bd1833 R581 interacts with the pyrophosphate group, hence the negative charge of glutamate would be adjacent to the negatively charged pyrophosphate group and prevent CoA binding [256]. R535 (3'-phosphate binding) has been established as important for acetyl-CoA binding to Bd1833; however, in MaeB, DME and TME, R535 has been exchanged for a non-polar residues. K538 is also exchanged to a leucine in DME but an arginine in MaeB, which may partly compensate for the R535 substitution.

Furthermore, R581 is also altered in MaeB to a serine. Interestingly Bd1833-SKB also differs at R581 (lysine residue). Mutation of R581K could explain why Bd1833-SKB is not inhibited by CoA and inhibited less by acetyl-CoA compared to Bd1833 ($K_{i(acetyl\text{-}CoA)}$ 0.453 ± 0.019 µM for Bd1833 and $K_{i(acetyl\text{-}CoA)}$ 21.3 ± 0.84 µM for Bd1833-SKB). The acetyl-group binding pocket also differs between Bd1833 and other multi-modular ME. Q683 (H-bond acetyl-group) is not conserved and is exchanged to a His in MaeB and DME, which may explain why acetyl-CoA is selectively inhibitory (weaker 3'-P and pyrophosphate binding sites and stronger acetyl-group interactions). Together these observations demonstrate that Bd1833 likely exhibits the highest affinity CoA binding site explaining why Bd1833 activity displays inhibition to both CoA and acetyl-CoA and the additional interaction with the acetyl-Group explains why acetyl-CoA is more potent.

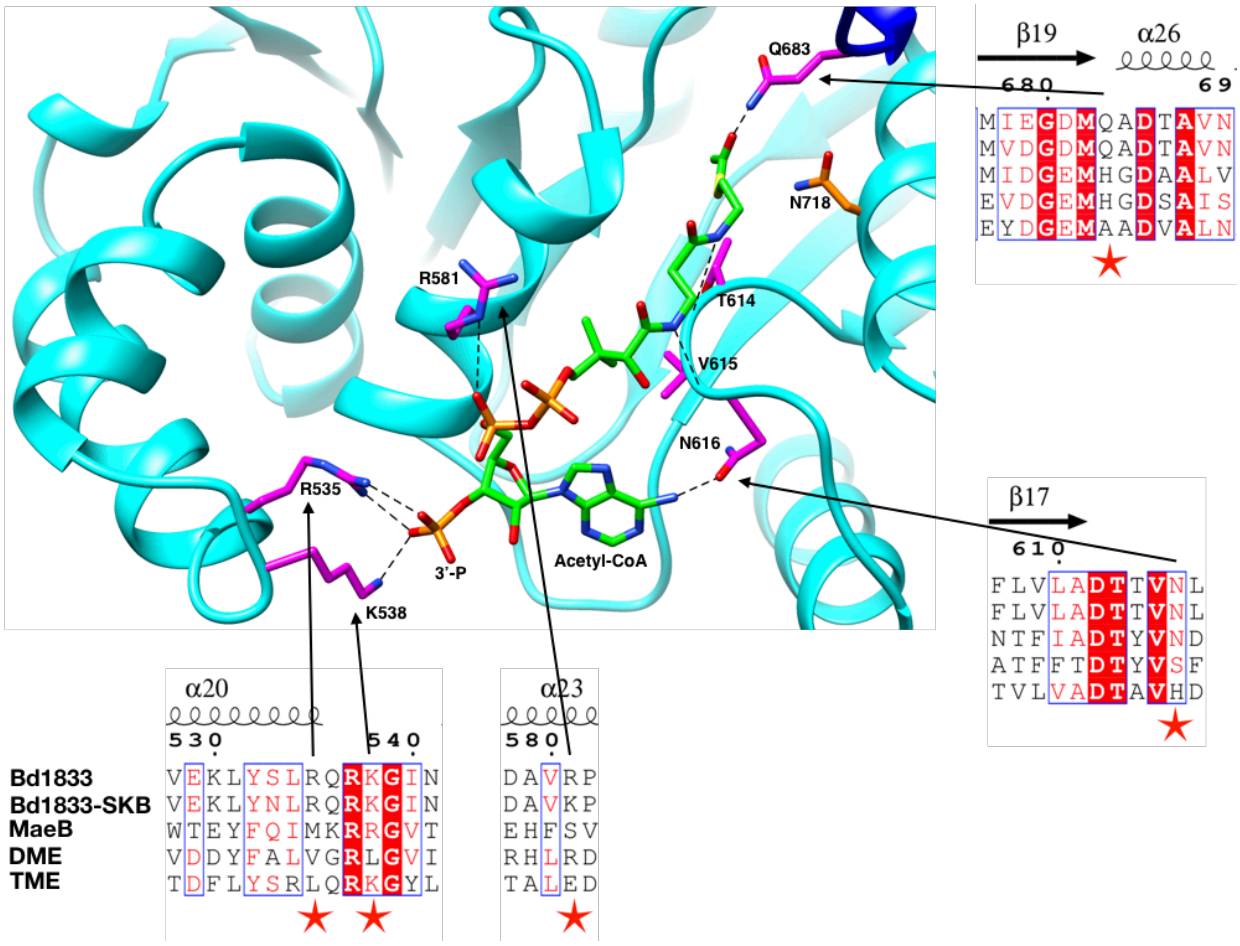


Figure 97. Bd1833 acetyl-CoA binding site comparison
The binding site of acetyl-CoA with the corresponding sections of sequence alignment of multi-modular ME, demonstrating the mutations at these locations, that likely contribute to the different binding affinities for CoA / acetyl-CoA.

5.3.5 Acetyl-CoA binding and implications for Pta catalytic mechanism

The structure of Bd1833-Pta demonstrates the binding site of acetyl-CoA to be located in the substrate-binding groove between subdomains-I and II. Furthermore the structure of Bd1833-Pta highlights how the highly conserved residues (R535, K538 and R581) are involved in binding the acetyl-CoA ligand, and these sidechains adopt similar conformations to those observed in previously published Pta structures [247]. Ligand binding in Bd1833-Pta is in excellent agreement with the previously published kinetic investigation into Pta enzymes [248]. Moreover ligand binding to Bd1833-Pta highlights problems with the currently proposed Pta catalytic mechanism and ligand binding site [247]. Two CoA binding sites are currently proposed for *Methanosarcina thermophila* Pta (CoA1 and CoA2). However, closer inspection of the F_o-F_c electron density map for PDB:2AF3 (CoA ligand soak - resolution: 2.6 Å, R-Free: 0.290, R-Work: 0.215) suggests that the CoA ligands may have been incorrectly modelled into the electron density. Importantly the binding site of CoA2 (proposed weaker interaction) interacts with R87, K90 and R133 (equivalent to R535, K538 and R581 in Bd1833).

The proposed CoA binding site for Pta enzymes disagrees with biochemical studies, whereas in Bd1833-Pta the observed acetyl-CoA binding site agrees with previous biochemical studies [247, 248]. The conformational change observed in Bd1833 upon acetyl-CoA binding is similar to that observed in other Pta enzymes, which has been suggested to be important for catalysis [250]. Comparison of the conformational changes observed in subdomain-I with those of Pta enzymes reveals the transition between Bd1833-Pta and Pta-Apo represents another closed state (Figure 98). Importantly the structure of Bd1833-Pta demonstrates a closed form bound to acetyl-CoA, thus in a catalytically competent state (were it to be an enzyme).

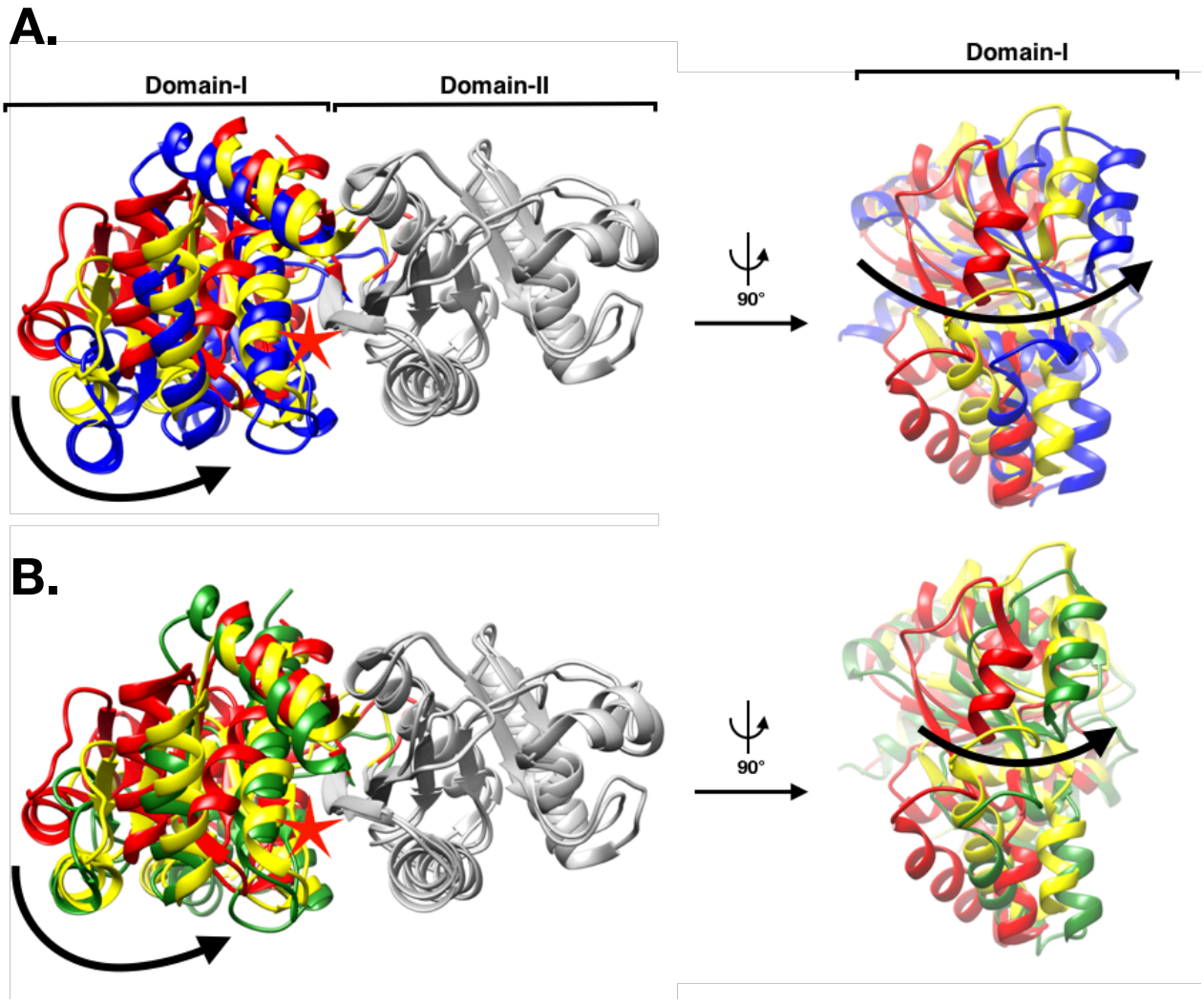


Figure 98. Pta domain movements
A. The conformational change in Bd1833-Pta subdomain-I (blue) superimposed with PDB:1QZT in the open conformation (red) and closed conformation (yellow). Bd1833-Pta adopts a different closed state. **B.** The conformation of Pta-Apo subdomain-I (green) superimposed with PDB:1QZT in open (red) and closed (yellow) conformations. Pta-Apo adopts a similar conformation as PDB:1QZT in the closed state. Superimposition was performed by aligning subdomain-II (residues 596-737) using the LSQ superpose tool in coot. (Length = 118 residues, rmsd 4.97Å).

Though Bd1833-Pta is non-catalytic, the acetyl-CoA binding site provides structural evidence for an alternative catalytic mechanism. *Methanosarcina thermophila* Pta (PDB:1QZT) can be superimposed onto the structure of Bd1833-Pta revealing the location of potential catalytic residues. R310 (R749-Bd1833) is currently proposed to interact with and position acetyl-P during catalysis [220]. R749 is observed close to the acetyl-CoA acetyl-group in Bd1833 although it is involved forming an important salt bridge with E544 at the hexamer interface, suggesting a repurposed role. The currently proposed Pta catalytic mechanism requires R310 to undergo a large conformational change, which would require a complete reorganisation of the protein in order to deliver the acetyl-P molecule to the active site [249]. However, in the structure of Bd1833-Pta R749 (R310) is orientated correctly, relative to the acetyl group to facilitate catalysis. This is further supported by the structure of *Bacillus subtilis* Pta in complex with acetyl-P (PDB:1XCO), where R304 (R749) is bound to acetyl-P, adopting a similar conformation [252]. Thus the acetyl-P binding site present in 1XCO and the acetyl-CoA binding site present in Bd1833-Pta are in agreement, suggesting the residues close to the acetyl-CoA thiol group in Bd1833-Pta may constitute the catalytic residues (were it to be an enzyme) (Figure 99).

The highly conserved S128 (*Methanosarcina)* (Y576 in Bd1833) would be within hydrogen bonding distance of the CoA thiol group and within hydrogen bonding distance of D307 (V746 in Bd1833) (a highly conserved Pta residue). D307 could potentially perform the role of catalytic base and abstract the proton from the thiol group, a required step of catalysis [265]. The thiolate ion would then in turn perform nucleophilic attack of acetyl-P. In Bd1833 the strictly conserved N718 (N279 in *Methanosarcina*), proposed to stabilise the catalytic intermediate is located ~3.5Å away

from the thiol group of acetyl-CoA demonstrating its potential ability to fulfil this role [250].

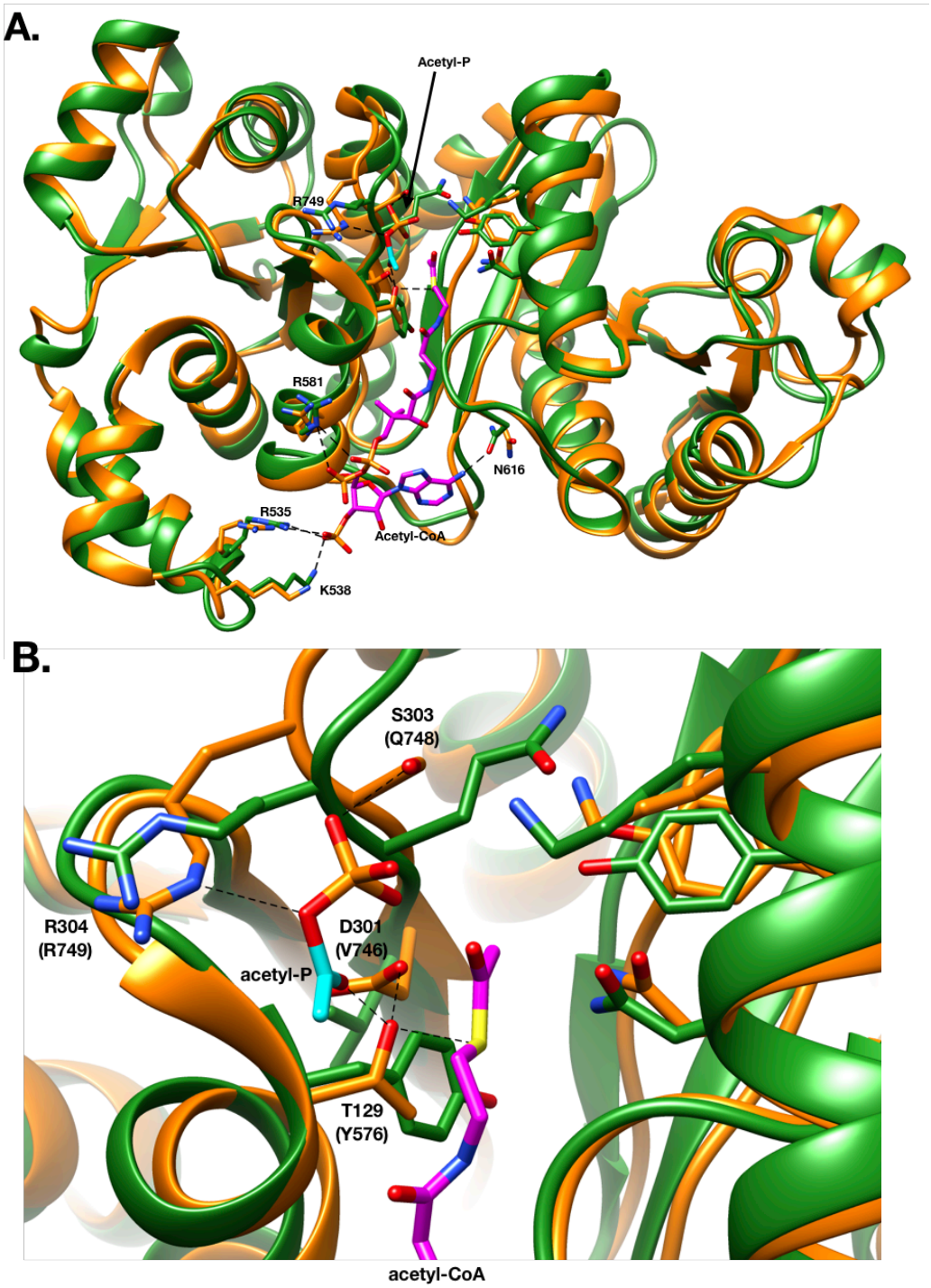


Figure 99. Pta active site
A. Bd1833-Pta monomer (green) superimposed with PDB:1XCO Pta monomer (orange). Superimposition was performed by individually aligning Bd1833-Pta subdomain-1 and II with the individual domains of PDB:1XCO. Bd1833-Pta subdomain-I (residues 460-587) with PDB:1XCO domain-I (residues 14-297) (length =174 residues, rmsd = 1.43Å). Bd1833-Pta subdomain-II (residues 596-737Å) with PDB:1XCO domain-II (residues 150-292) (length = 144 residues, rmsd = 1.57). The active site is likely located in the upper portion of the substrate-binding groove **B.** A close up view of proposed active site. The catalytic residues of 1XCO are labelled alongside the comparative Bd1833-Pta residues, demonstrating substitutions. The thiol group is within H-bonding distance of T129. T129 is in H-bond distance of D301 and acetyl-P.

5.4 Conclusion

We have solved the first structure of a multi-modular malic enzyme and developed an understanding of the mechanism behind allosteric regulation by acetyl-CoA. The structure of Bd1833 reveals that the Pta domain is central to hexamer complex formation and binding of acetyl-CoA. We show that R535 acts to sense acetyl-CoA via interaction with the 3' phosphate group; this sensing results in a conformational change communicated using the 17aa loop specific to multi-modular MEs. In total, the binding of acetyl-CoA acts to close the ME active-site cleft, acting as a negative regulator of enzyme activity. Our work provides a structural basis for understanding allosteric regulation of multi-modular MEs, which will assist in analysis of their roles in central metabolism [254, 266].

Chapter 6

Bd0967, an Important Developmental C-terminal Processing Protease

6.1 Introduction

6.1.1 Overview

Signalling events are crucial for the correct completion of lifecycle stages in *Bdellovibrio*. A C-terminal processing protease (CTP), Bd0967, has been identified as being part of a network of cellular signalling events in *Bdellovibrio* [82]; however, very little is known about the role or function of this protein.

CTPs have been documented to have essential roles in complex developmental signalling cascades that influence varying organism lifestyles. Though important, CTPs have been understudied and there is a lack of structural and functional data regarding these enzymes. CTPs are a diverse family of proteases whose activity is regulated by a PDZ domain. The roles of CTPs have been implicated in developmental pathways and signalling during: spore formation in *Bacillus subtilis* [85]; immune system evasion by *Chamydia thrachomatis* [86]; cell wall maintenance in *E. coli* [87] and photosynthesis activation in *S. obliquus* [84]. Hence, Bd0967 is likely to be involved in a developmental

signalling cascade that underpins the complex lifecycle of *Bdellovibrio*. Uncovering its role will help reveal how intricate lifecycle progressing events may be regulated in *Bdellovibrio* and also broaden our understanding of this important and diverse set of proteases.

6.1.2 The diverse roles of C-terminal Processing proteases

CTPs have been implicated in a wide range of roles in different organisms; many of these roles are central to important lifestyle events. CTPs have been shown to perform either regulatory roles by cleaving a portion of peptide from the C-terminus of precursor proteins or as housekeeping enzymes that degrade proteins tagged by the SsrA system due to incomplete synthesis [267]. Below are a few examples of the roles of CTPs.

6.1.2.1 Cell wall maintenance

In *E. coli* the tail-specific protease (Tsp), is multi-functional. Firstly Tsp has a role in degrading proteins with SsrA tags. SsrA tags are added to proteins that are unsuccessfully synthesised at the ribosome, which drastically reduces the protein half-life. SsrA tags tend to contain small non-polar residues and are recognised and targeted by degradative proteases [267].

Tsp is also referred to as Prc in the literature because of its ability to cleave the C-terminus of the precursor form of penicillin binding protein 3 (PBP-3) [268]. However, the C-terminal sequence is not suggestive of recognition by Tsp PDZ domain, implying that an additional recognition mode may be used [269].

More recently it has been demonstrated that a Prc also has a key role in regulating the levels of MepS [87], an outer-membrane lipoprotein that specifically hydrolyses D-Ala-meso-DAP cross-links, and therefore plays an important role in cell wall maintenance during growth and division. By controlling the levels of MepS, Prc contributes to controlling cell wall cross-linking levels and cell wall viability. It has been shown that Prc requires an adaptor protein NlpI, an outer-membrane lipoprotein with tetratricopeptide repeats (TPRs), in order to recognise MepS as a substrate. NlpI interacts with Prc and facilitates degradation of MepS by providing a docking site for MepS alongside PrC. Prc processively degrades MepS and does not perform a single cleavage event as is the case for many other CTPs [87].

6.1.2.2 Spore formation

In *Bacillus subtilis* a CTP (CtpB) is integral to synchronisation of mother-cell and forespore development [85]. Endospores are produced in *B. subtilis* to cope with surviving harsh environmental stresses. A coordinated intracellular signalling cascade controlled by cell-type-specific sigma factors tightly regulates control of endospore formation. The final step of spore formation is initiated by sigma factor σ^k. Sigma factor σ^k is synthesised as an inactive membrane-bound precursor that is activated by regulated intra-membrane proteolysis (RIP). The SpoIV signalling cascade that leads to maturation of sigma factor σ^k is complex and tightly regulated. The RIP pathway contains three proteases (4B, CtpB and 4FB) that act sequentially. Sigma factor σ^k is released from a precursor state by proteolysis catalysed by 4FB. 4FB activity is suppressed by the presence of an intra-membrane regulator (4FA) that must be

specifically degraded to release its inhibition of 4FB. CtpB performs a specific cleavage of 4FA to remove a peptide segment that inhibits 4FB activity. Initially a C-terminal folded region protects 4FA from proteolytic cleavage by CtpB, which is removed by 4B. The truncated 4FA C-terminus is recognised by CtpB and binding activates CtpB, which proceeds to remove the inhibitory portion of 4FA and releases repression of 4FB activity. Active 4FB cleaves the membrane-associated sigma factor σ^k precursor, releasing the product into the cytosol, which causes sporulation to proceed [85].

6.1.2.3 Photosynthesis

In *Scenedesmus obliquus* a CTP has a critical role during the activation of photosystem II reaction centre protein D1. D1 is inserted into the thylakoid membrane as an inactive precursor and activation is achieved by the removal of a short C-terminal extension from the precursor pD1 by the CTP, D1P [84]. Proteolytic processing is required to allow the light driven assembly of the tetranuclear manganese cluster, which is the site of photosynthetic water oxidation [270].

6.1.2.4 Virulence

CTPs also have a role in virulence, such as CT441 from *Chlamydia trachomatis*. CT441 is proposed to be an important factor of pathogenesis and proposed to interfere with host antimicrobial and inflammatory responses; however, cleavage of NF-κB during the *Chlamydia* infection is now being questioned. CT441 is documented to modulate the estrogen-signalling pathway of host cells by interactions with the steroid receptor RNA activator protein 1 (SRAP1) [86]. CT441 has been shown to degrade SRAP1 *in vitro*, which leads to the complete substrate degradation within a 1 hour time period. This

finding indicates that CT441 fully degrades SRAP1 in a manner similar to Tsp with ssRNA-tagged proteins, rather than cleave a portion like CtpB and D1P. Furthermore, additional chaperone activity has been assigned to CT441, similar to that of other PDZ containing proteases belonging to the HtrA family that are responsible for protein quality control in the periplasm [271].

In *Staphylococcus aureus* a lone CTP, CtpA, is suspected of promoting virulence. CtpA localises to the cell wall where it contributes to virulence by maintaining cell wall integrity [272].

In *Vibrio cholerae* Tsp activity has been associated with the regulation of the virulence activator TcpP via intramembrane proteolysis. Tsp activity is shown to process the C-terminus of TcpP (subsequently marking TcpP for cleavage by a second protease, YaeL). This process leads to TcpP removal from the membrane, where it can then act as a transcriptional regulator for the expression of *cholera* toxin and toxin-co-regulated pilus [273].

6.1.2.5 Quality control

Quality control in cells is typically performed by the High temperature requirements A (HtrA) family of serine proteases that share a similar structural arrangement to CTPs. HtrA proteases utilise either one or two PDZ domains that facilitate substrate recognition (via the C-terminus), which activates catalysis (similarly to CTP). Although these two families of proteases share a similar domain makeup and arrangement they are structurally distinct and possess different active site chemistries. The HtrA protease

domain resembles a chymotrypsin like fold, containing two distinct six-stranded β-barrel subdomains with a central peptide-binding cleft [274]. HtrAs have a typical serine protease catalytic triad (Ser, His, Asp), whereas CTP have a Ser, Lys, Gln triad. HtrA proteases are mainly involved in protein quality contro; however, a small proportion have been implicated in triggering signalling pathways [271]. HtrAs form large oligomeric structures (like the 24mer DegP), which act as molecular shredders [275].

6.1.3 Structure and mechanism of C-terminal processing proteases

To date relatively few structures have been documented for this diverse class of proteases. Figure 100 illustrates the structural diversity of CTPs. CTPs share a conserved central protease domain (belonging to the S41 class of proteases) and a hinged PDZ domain, which regulates protease activity. N-terminal and C-terminal domains of CTPs vary in structure, promoting the diverse structures and functions associated with this class of proteases.

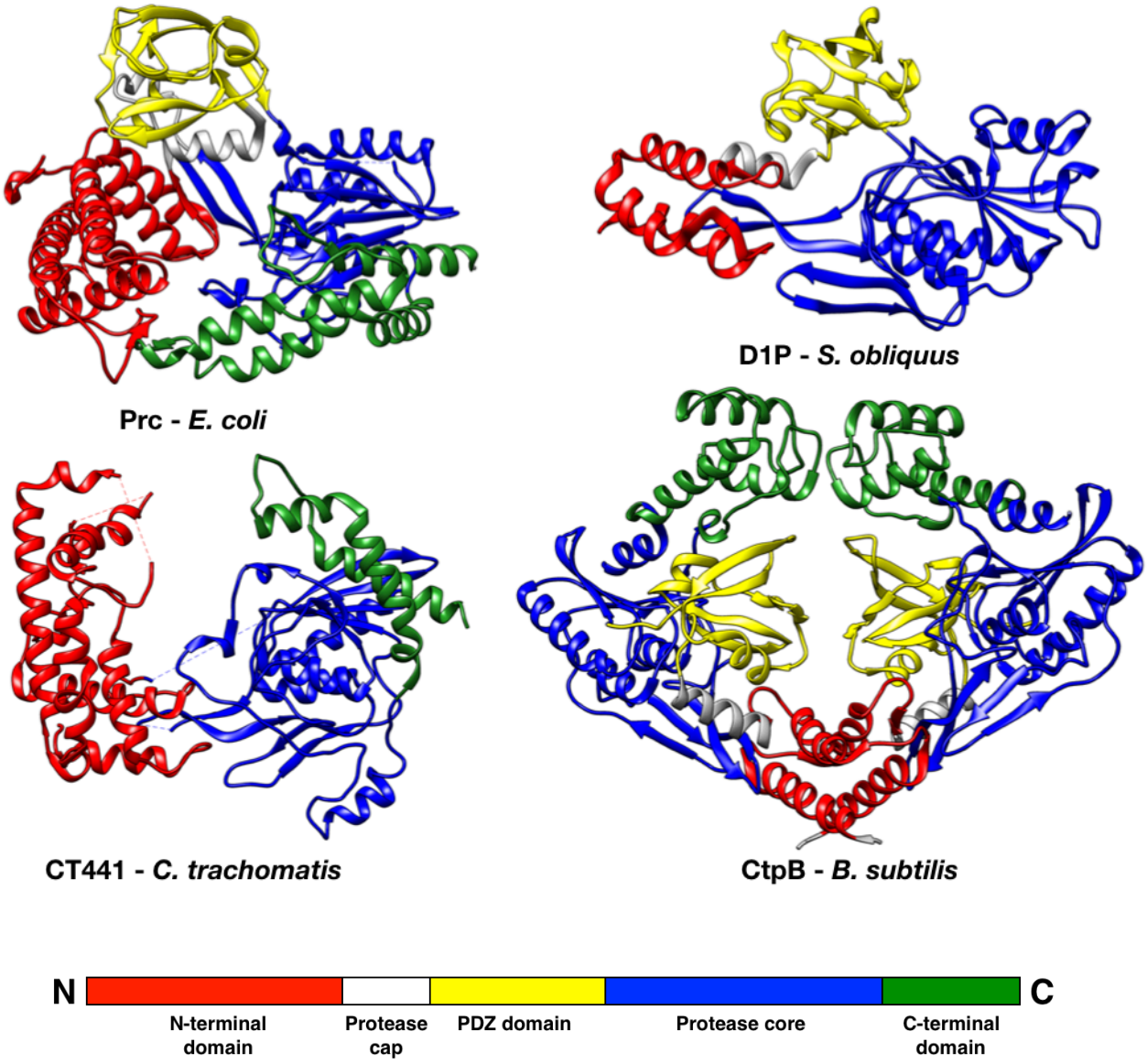


Figure 100. Structures of C-terminal processing proteases
The four solved structures of C-terminal processing proteases are presented and coloured by their domains. NTD in red, PDZ domain in yellow, protease domain in blue, protease cap helix in grey and CTD in green. Prc – 5WQL, D1P – 1FC7, CT441 – 4QL6, CtpB – 4C2C.

6.1.3.1 D1P

The first three-dimensional structure solved for a CTP was the structure of D1P, which consists of three loosely packed, distinct domains: an α-helical N-terminal domain, a central PDZ domain and a larger C-terminal protease domain. The PDZ domain is composed predominantly of a twisted five-stranded β-sheet with two α-helices. The protease domain is folded into two β-sheet regions. The larger, mixed five-stranded β-

sheet region is packed between an α-helix on one side and three α-helices on the other. The second, smaller four-stranded antiparallel β-sheet region loosely associates with the three central α-helices that pack against the larger β-sheet. The smaller β-sheet has two strands that project away from the core to form an integral short three-stranded β-sheet with the N-terminal domain [84]. The structure of D1P represents a minimal fold for CTPs.

The catalytic site of D1P is proposed to consist of a catalytic dyad of S372 acting as the nucleophile and K397 acting as the general base [84]. However, more recent studies suggest that this family of proteases actually use a catalytic triad where a glutamine residue is required to orientate the catalytic lysine residue. D1P does possess a conserved glutamine residue (positioned 5Å away from the lysine), unable to participate in catalysis via hydrogen bonding, although active site dynamics imposed by PDZ domain conformational changes may reposition this residue in the active conformation.

6.1.3.2 CtpB

The second structure determined from this family was that of CtpB, a CTP with a divergent three-dimensional structure that expands on the core fold of D1P. CtpB is a 480aa protein that forms a functionally relevant homodimer, creating an enclosed ring-like structure. The overall structure consists of a N-terminal domain required for dimerization, a central PDZ domain, a core protease domain with a subdomain referred to as the protease cap and an additional C-terminal domain that also participates in dimerisation [85]. CtpB structures have been solved for both active and resting

conformations, revealing a rigid-body conformational swing of the PDZ domain, which gates the activity of the protease domain by permitting accessibility to the active site and rearranging the catalytic residues. The conformational change is initiated by the substrate binding to the PDZ domain.

The protease core and the protease cap make a narrow tunnel about 7Å wide that could potentially accept the passage of an unfolded protein segment. Buried deep inside the tunnel, towards the interior of the protein are the active-site residues. The peptide binding cleft of the PDZ domain is positioned at the opening of the tunnel in the resting state (closed state). The active-site residues of CtpB consist of the active-site nucleophile S309, which is in hydrogen bonding distance of K334 (presumed to participate in an acid-base catalytic mechanism). Q228 completes the catalytic triad by orientating K334. The crystal structure of CtpB revealed a peptide bound in the active-site binding site both as a covalently attached, trapped catalytic intermediate and as a free peptide chain. This observation demonstrates how the peptide is accommodated in the P2, P3 and P4 pockets whereby the peptide interacts closely with a β-strand found close to the active site [85].

Density for an additional peptide fragment bound to the PDZ peptide-binding groove was also observed in CtpB. Structural comparisons between the 'resting' state (where no peptides are bound and the 'active state' (where peptides are bound to the PDZ domain and active-site peptide-binding clefts) reveals striking conformational rearrangements in the structure of CtpB. Upon peptide binding the PDZ domain switched from blocking access to the protease tunnel in the 'resting' state to a distant outwardly swung location. The PDZ domain swing opens access to the protease tunnel

and reshapes the active site, bringing K344 into hydrogen bonding distance of S309 [85]. Peptide binding to the PDZ groove is believed to initiate the swing, which is coupled to a mechanistic pulling force upon the bound peptide, which pulls the peptide through the tunnel and positions the substrate for catalysis.

6.1.3.3 CT441

A partial structure of CT441 has been solved, revealing another divergent structure for a CTP. CT441 is comprised of four distinct domains: an expanded N-terminal scaffold domain, a central PDZ domain (not traceable in structure), a core protease domain and a short C-terminal domain. The N-terminal domain of CT441 is formed from a large bundle of α-helices and is considerably larger than the N-terminal domains found in CptB and D1P. The position of the PDZ was not resolved for CT441 suggesting that this domain is highly mobile and able to adopt multiple conformations [86]. Additionally, CT441 has an extended C-terminal region that can be considered an individual, separately folding domain, which elongates the C-terminus creating a cup-like structure with the N-terminal domain. CT441 also contains a catalytic triad of S455, K481 and Q485 located in the protease core domain, similarly to CtpB.

6.1.3.4 Prc

The most-recent CTP structure to be solved is the structure of Prc (also known as Tsp). Prc forms a monomeric, self-compartmentalised bowl shaped protein with a lid-like PDZ domain, whose structure is distinct from other CTPs. The Prc structure can be divided into five domains; the C-terminal and N-terminal domains form the sides of the bowl whilst the protease core and vault domains make up the base, with the PDZ

domain is positioned at the lid of the structure, attached by a double hinge. The protease domain and PDZ domain share a common fold with other CTPs. The fold of the N and C-terminal domains is unique to Prc. They are extended in comparison to other CTPs, which allows them to interact at their tips to form the self-compartmentalised ring shape. Prc is reported to have a catalytic dyad consisting of K477 and S452, located at the end of the peptide tunnel, which runs through the protease domain [87]. The comparative glutamine residue of Prc appears to be too far away from K477 to contribute to catalysis like in CtpB and CT441.

Prc was also solved in complex with an unidentified co-purified ligand located at the active site and PDZ peptide-binding site, similarly to CtpB. Bound peptides suggest that the structure of Prc is in an active conformation. Moreover the PDZ domain adopts a similar pose to that of CtpB in the 'active state'. This structure therefore provides more evidence that the PDZ domains are able to adopt multiple conformations, which in turn gates the protease activity.

6.1.4 PDZ domains

6.1.4.1 Roles

PDZ domains (postsynaptic density of 95 kDa, discs large, and zonula occludens) are abundant protein-protein interacting modules, found in various species, from prokaryotes to eukaryotes [276]. PDZ domain containing proteins have diverse cellular and biological roles; commonly they are involved in signalling cascades, where the PDZ domain is required to mediate protein-protein interactions.

For instance in the mouse genome there are 328 recognised PDZ containing proteins, which are found in modular proteins with a diverse set of other domains e.g. PDZ domains are found in complex with: nitric oxide synthase (NOS) domains, protease domains, guanylate kinase-like (GK) domains and Src homology 3 (SH3) domains [84, 277-280]. Multiple copies of PDZ domains are found in some proteins such as glutamate receptor interaction protein that has six PDZ domain copies [281]. The sheer diversity and abundance of PDZ containing proteins suggests their importance in cellular function and modulating protein interactions.

6.1.4.2 Fold and substrate binding

PDZ domains are a small (80-100aa), compact domain that share a common fold. PDZ domains usually contain five or six β-strands and two or three α-helices [282] that form a highly conserved fold (consisting of a central, five or six-stranded, anti-parallel β-sheet). The folded β-sheet is enclosed at each end by one short α-helix and one longer α-helix [283]. PDZ-like domains of CTPs and HtrA proteases possess a similar fold but typically only have a five-stranded β-sheet and two additional β-strands at the N and C-terminus of the domain [84, 271].

PDZ domains bind to and recognise the terminal carboxylate group of peptides and proteins [284]. An array of structural information has provided insights into the binding specificity and promiscuity of PDZ protein-protein interactions. In addition peptide libraries and protein arrays have also been used to probe PDZ binding properties [285]. The PDZ peptide-binding site is located in a groove between the longer α-helix and the first β-strand of the shorter β-sheet. The binding groove contains

a highly conserved carboxylate-binding motif (R/K-XXX-G-Φ-G-Φ, where X is any amino acid and Φ is a hydrophobic residue) [285-287].

The peptide substrate binds in the PDZ peptide binding groove in a β-augmentation manner via the first β-strand [276, 286], (Figure 101). PDZ domains recognise a specific peptide sequence at the C-terminus of their substrates. Each residue surrounding the PDZ peptide binding groove can interact with the PDZ ligand and contribute to binding. Variation of residues (sidechains) at the PDZ peptide binding groove alter the character of each pocket and dictate which ligands can be accommodated. The PDZ peptide binding groove binds the C-terminal four amino acids of the PDZ, where P(0) is the C-terminus. Specificity for the sidechain of P(0) is usually dictated by the sidechains of the first β-strand (β1); bulkier P(0) sidechains are accommodated when isoleucine/leucine are the first residue of β1 (β1-1) [288]. The sidechain of P(-1) faces outwards, towards solvent and was initially believed to not be selected for. However, mounting evidence suggests bulky hydrophobic residues are preferred at the P(-1) position [288-291]. The P(-2) residue can interact with residues of the longer α-helix, which is known to be an important feature of specificity in PDZ domains [286-288]. Preference for the P(-3) residue is harder to determine and polar residues can be accepted as well as short and long sidechains, depending on the residues located at β1-4 and β1-5 positions [285].

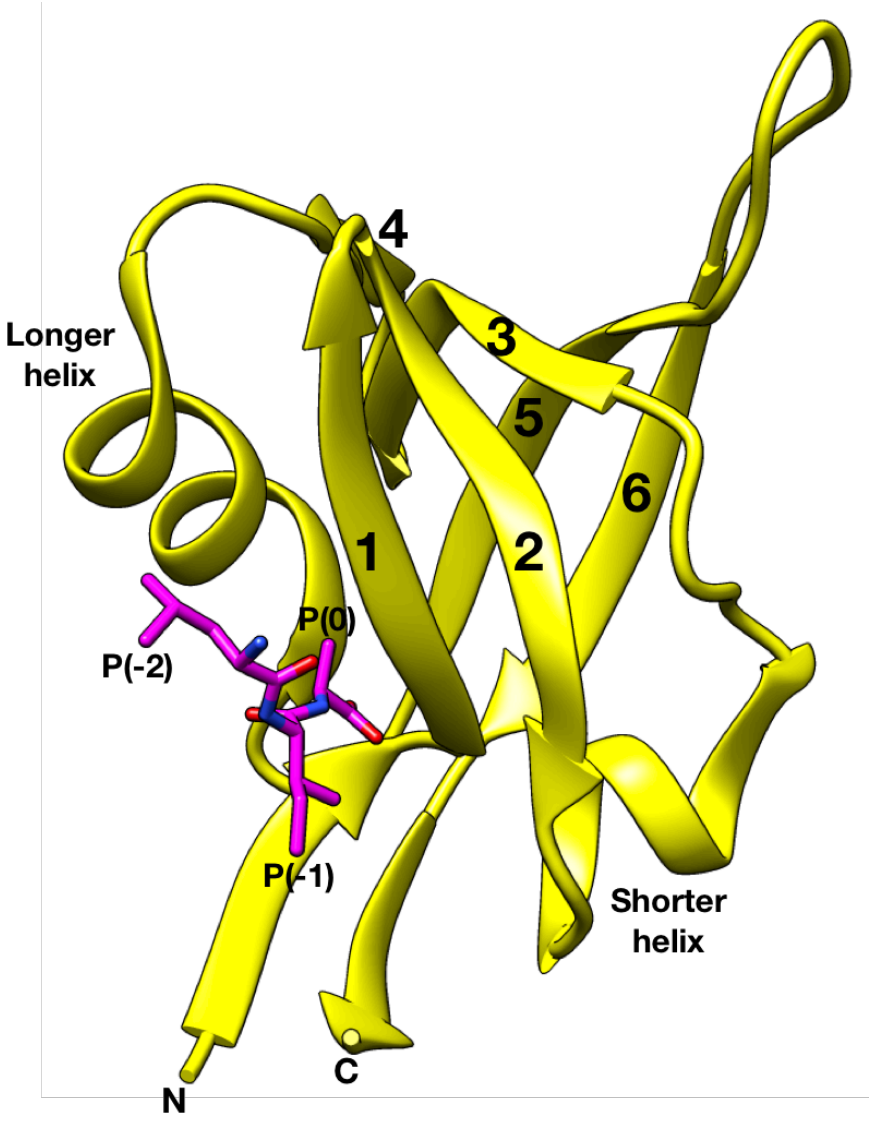


Figure 101. PDZ domain structure
The structure of a C-terminal processing protease PDZ domain is shown here bound to a 3aa peptide substrate. The peptide binds in a β-augmentation manner with strand-1. The peptide residues are numbered according to the text. The N and C-terminal β-strands are specific to CTPs. (PDB: 6BXG)

6.1.5 Substrate specificity and mechanism of C-terminal processing proteases

6.1.5.1 C-terminal processing protease PDZ specificity

For CTPs the PDZ domain facilitates substrate recognition by interacting with the C-terminus of the substrate. For the PDZ domain to bind the substrate C-terminus it must firstly be accessible. In the case of Prc and CtpB the substrate C-terminus must be disordered enough to fit through the protease domain peptide tunnel in order to interact with the PDZ peptide-binding groove.

Prc will rapidly degrade proteins that possess a hydrophobic or nonpolar C-terminal sequence, such as WVAAA [269]. Additionally alanine is preferred at the very C-terminal residue, with alanine or tyrosine penultimate and alanine or leucine at the third position from the end. Amidation of the α-carboxylate group protects peptides from degradation by Prc suggesting that this group is involved in substrate selection [269].

The structure of Prc revealed a co-purified and co-crystallised peptide fragment bound to the PDZ peptide-binding groove with a carboxylate recognition motif of G-I-G-A. The P(1) binding pocket is small and hydrophobic, lined with hydrophobic residues (isoleucine, leucine and valine), likely preferring peptide fragments with small non-polar residues at the P(1) position. The P(2) binding pocket is similarly lined with hydrophobic residues although it is slightly more solvent accessible. The peptide docked in the PDZ peptide-binding groove was modelled as LSRS-COOH, where the

penultimate arginine residue makes contact with D342 of the PDZ hinge region [87]. The PDZ and protease domain peptide binding sites have similar properties, which probably facilitates the processive nature of substrate degradation by Prc, whereby the peptide fragment produced upon proteolytic cleavage is a ligand for the PDZ domain.

Similarly to Prc, the structure of CtpB had a co-purified and co-crystallised peptide fragment bound, although its sequence could not be determined by mass spectrometry. The peptide fragment interacts with the PDZ peptide binding groove, where the carboxylate group is anchored by a consensus carboxylate binding loop (G-I-G-A). The peptide binds in a β-augmented fashion, similarly to other PDZ domains. Peptide library analysis revealed CtpB preferentially binds peptides with Y/F/W-A/V at their C-terminus. This finding is consistent with other PDZ domains that prefer bulkier hydrophobic residues at the P(1) position, which fit into the larger hydrophobic pocket [85].

A high-resolution structure of *Vibrio cholerae* Tsp PDZ domain (PDB:6BXG) has been solved in complex with a peptide fragment revealing the binding mode to its substrate (Figure 101). Although this work is yet to be published it reveals the same mode of binding as demonstrated by the CtpB and Prc PDZ domains. The bound peptide of LIA, at high resolution (1.45Å), demonstrates how the hydrophobic peptide sidechains interact with the PDZ domain residues surrounding the PDZ peptide-binding groove.

6.1.5.2 Protease domain binding sites

The specificity of Prc cleavage sites has been mapped, showing that the P1 position is selective to small uncharged amino acids such as alanine, serine or valine. However, there seems to be less selectivity towards small uncharged amino acids at the P1' site as this was shown to allow larger amino acids, such as methionine and tyrosine [269]. The structure of Prc agrees with biochemical data, revealing Prc has a short hydrophobic proteolytic binding pocket. The P1 binding pocket is a small hydrophobic pocket, lined with residues from the protease domain and the protease cap, whereas the P2 site is partially solvent exposed and formed primarily by bulkier residues from the protease domain. The binding sites for other peptide residues are less defined and more exposed to the solvent [87].

In CtpB, the peptide binding specificity at the protease domain primarily depends on the P1 and P2 residues. The P1 sidechain is accommodated in a small hydrophobic pocket, whereas the P2 sidechain protrudes into a larger hydrophobic pocket formed by residues of the protease cap [85]. Biochemical data, consistent with the structural data, suggest CtpB preferentially selects substrates with residues P1-(alanine and valine) and P2-(threonine, leucine and isoleucine) [85].

6.1.5.3 Mechanism involving PDZ domain

Substrate recognition by the PDZ domain, in CTPs, regulates proteolytic activity. This has been demonstrated best for CtpB, where the PDZ domain gates access through the protease channel and upon substrate binding the PDZ relocates to an active conformation, which pulls the substrate through the protease channel and repositions

the catalytic residues for protease activation. The PDB currently contains three unique structures of CTPs with unambiguous location of the PDZ domain. There are the two unique CtpB structures where the PDZ domain is adopting a 'resting state' and an 'active state'. Superimposition of D1P and Prc structures with CtpB reveals the PDZ domain of D1P adopts a 'resting state' whereas the PDZ domain of Prc adopts an 'active state'. The two examples of PDZ domains in a 'resting state' and two examples of PDZ domains in an 'active state' suggests that PDZ relocation is an importantly conserved mechanistic feature which governs protease activity in this family of proteases.

The PDZ domain conformational switch between a 'resting' state to an 'active' state in CtpB requires a remarkable conformation change in the position. The PDZ domain, pivots about the flexible hinge, pulled away from the protease core, as demonstrated in Figure 102 [85].

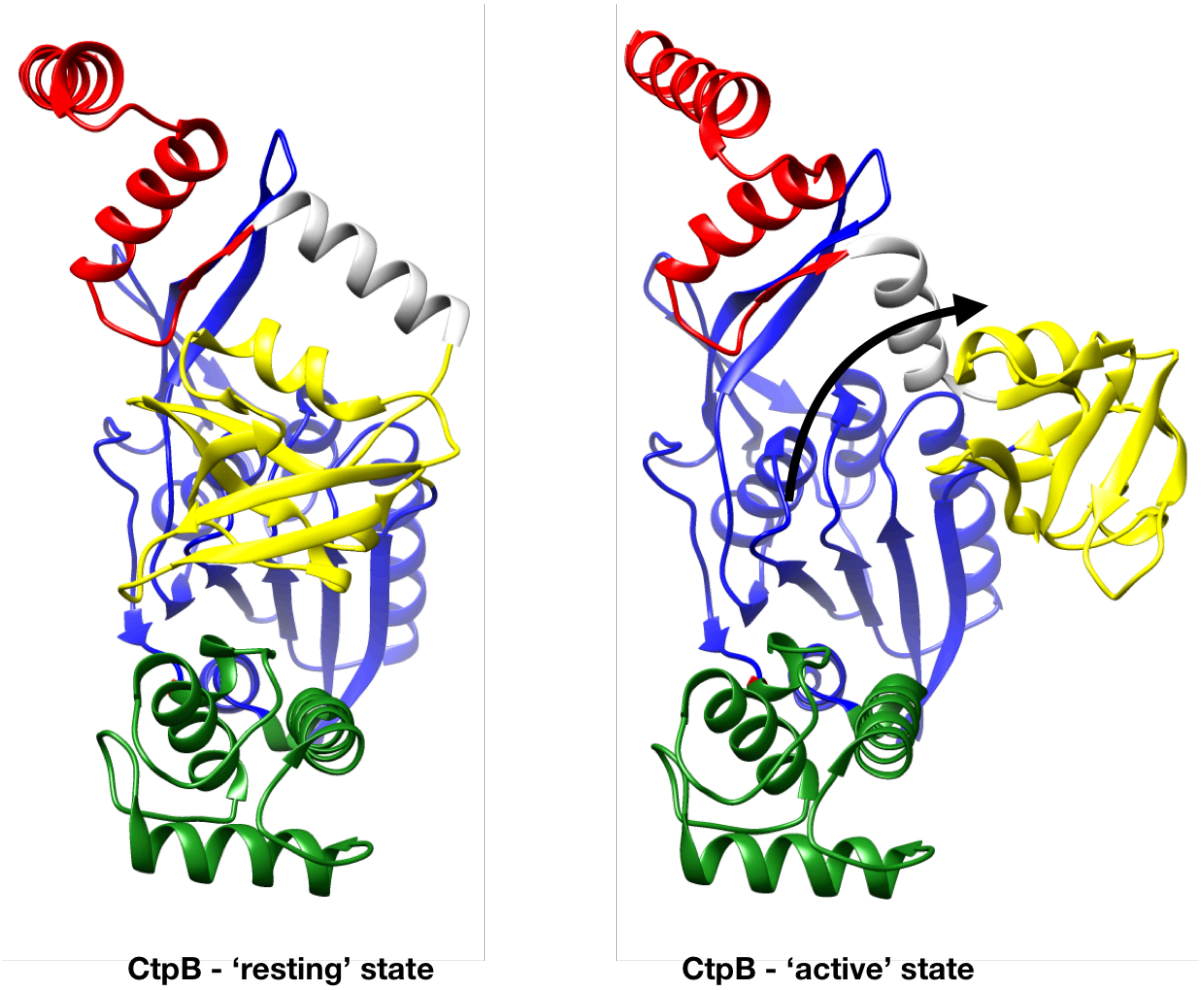


Figure 102. PDZ domain movement in CtpB
The mechanism of activation and regulation of C-terminal processing protease by the PDZ domain has been well documented in CtpB. The PDZ domain coloured (yellow), has been observed in multiple conformations and samples between a 'resting' and an 'active ' state. In the 'resting' state the PDZ domain is tightly tucked into the protease domain (blue). Whereas in the 'active' state (peptide bound) the PDZ domain pulls away from the protease domain. The black arrow represents the movement. The switch also repositions the protease cap (grey). PDB:4C2C & 4C2E

R168 was shown to be a key residue facilitating the sensing of peptide binding at the PDZ binding cleft and transferring the signal allosterically to bring about the conformational switch of CtpB. In the 'resting' state the sidechain of R168 faces away from the PDZ domain, making contacts with S106 located on protease cap. In the 'active' state the sidechain of R168 repositions, aligning with the bound peptide but still making contacts with S106. By maintaining these inter-domain contacts in the bound

and unbound states, R168 is ideally placed to function as the peptide sensor, stabilising the active conformation when a peptide is bound. Mutation of R168 produced a less active protease that bound its substrate with lower affinity. Furthermore, deletion of the PDZ domain produced a protease that was more active and had less substrate specificity. This suggests a mechanism by which CtpB is able to sample between a 'resting' and 'active' state. Upon substrate binding at the PDZ domain a conformational switch occurs, stabilising the active conformation and facilitating proteolytic activity [85].

Studies have demonstrated Prc can only degrade MepS when in complex with NlpI, a finding that has not been associated with any other Prc to date. Furthermore, the PDZ domain of Prc was shown to be absolutely required for activity, unlike CtpB. Deletion of the PDZ yielded inactive Prc; however, deletion of the PDZ domain in CtpB yielded constitutively active unregulated protease activity. This finding suggests that the PDZ domain in Prc is essential for degradative activity, possibly because it is required to sense substrate; the movement behind the sensing is then able produce a mechanical pulling force enabling complete substrate degradation. PDZ domain movement is believed to be aided by two hinged residues L245 and L340 that can slide over each other in a 'greasy' motion [87, 292].

Similarly in HtrA proteases the PDZ domain regulates activity. The PDZ domains are connected to the protease domains by a flexible region that allows considerable mobility. The PDZ domains not only control access to the proteolytic chamber but also provide the protease activation mechanism. Three distinct mechanisms have been outlined for this family of proteases, two of which can be applied to CTPs. The first is

the molecular ruler, where the distance between PDZ domain and proteolytic chamber is a measured uniform distance. The second is a processive cleavage mechanism, where the cleavage of the peptide chain at the proteolytic chamber creates a new substrate C-terminus that is subsequently recognised by the PDZ domain. The peptide-binding sites at the PDZ domain and protease domain would thus have to be highly similar. This mechanism would operate in a ratchet like manner and allow continued proteolytic processing of the substrate. The other possible mechanism involves PDZ imposed oligomer reorganisation by a conformational state change (not a proposed mechanism for CTPs) [271]. Figure 103 illustrates a cartoon representation for PDZ recognition and substrate presentation in CTPs.

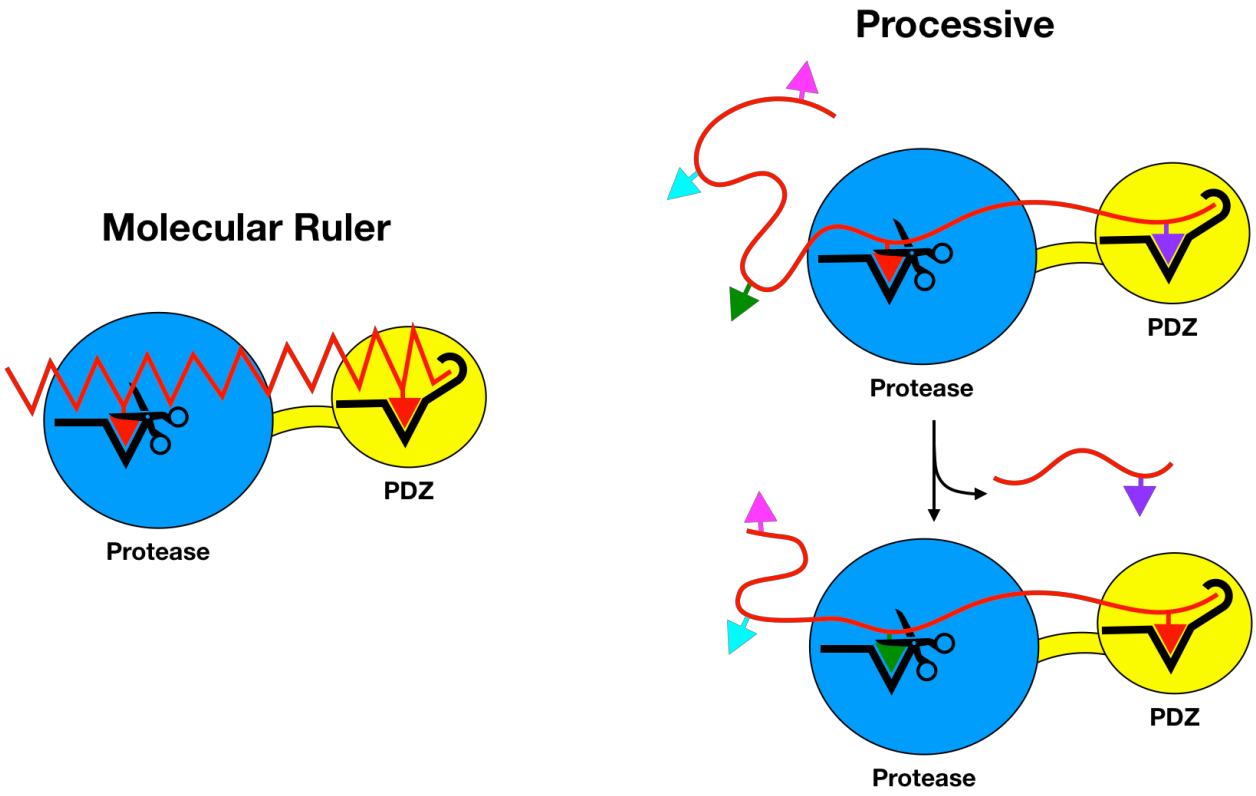


Figure 103. C-terminal processing protease mechanism
Two mechanisms are proposed for C-terminal processing proteases. **A.** A molecular ruler mechanism, where the distance between PDZ domain and proteolytic site is set. **B.** A processive mechanism, where the protease and PDZ binding sites are similar. The C-terminus of the new peptide becomes a ligand for the PDZ domain. Red line depicts the substrate and the coloured triangles demonstrates the recognition sequences (different colours demonstrate a different amino acid sequence.

6.1.6 Bd0967 role in predatory cycle

Bd0967 has been linked to a larger network of cellular signalling events in *Bdellovibrio,* possibly controlled by CdG [82]. Moreover, Bd0967 is a putative CTP, containing a PDZ domain for possible co-ordination of protein-protein interactions. The precise role of Bd0967 during the lifecycle of *Bdellovibrio* is currently unknown; however, an important role during cellular signalling and lifecycle progression is highly likely, due to our current knowledge of cellular and biological roles for CTPs in other organisms.

In addition to Bd0967, numerous other PDZ-containing proteins are found in the genome of *Bdellovibrio HD100*, which are listed in Table 15. Seven of the nine proteins identified as PDZ-containing proteins are proteases, one of which (Bd3534) has been shown to be an essential predation protein [293].

Uniprot ID	Gene name	Protein type	Length
Q6MGY2	Bd3787	Hypothetical zinc metalloprotease	557
Q6MHK7	Bd3534	Carboxyl-terminal protease	526
Q6MJH7	Bd2800	Serine protease MucD	474
Q6MP94	Bd0967	Tail-specific protease	673
Q6MRC0	Bd0169	Carboxyl-terminal processing protease	458
Q6MNK9	Bd1239	Carboxy-terminal processing protease	401
Q6MMM7	Bd1598	General secretion pathway protein C	304
Q6MPD5	Bd0922	Periplasmic serine protease	476
Q6MLF8	Bd2053	Component of the Tol biopolymer transport system	974

Table 15. PDZ proteins in *Bdellovibrio HD100* genome
A list of PDZ containing proteins found in the genome of *Bdellovibrio HD100.* The majority of the proteins are proteases. Bd3534 (green) has been identified as an essential predation protein.

6.1.7 Aims and objectives

This study aims to elucidate the importance of Bd0967, a C-terminal processing protease (understudied class of proteases). Structural and functional investigations were used to probe Bd0967 role as a potentially important developmental protease. The potential substrates of Bd0967 were identified from pull-down assays and Mass-Spectroscopy, which were later confirmed by protease digestion assays.

6.2 Results

6.2.1 Obtaining Bd0967 crystals

A construct of Bd0967 was created to encompass the full-length protein with the exclusion of the seven C-terminal amino acids that were predicted to be disordered. The protein was tagged with a His_6-tag at the N and C terminus. Closer inspection of the amino acid sequence revealed that Bd0967 possessed a probable signal peptide with a signal peptide cleavage site (LASA) located 24aa from the N-terminus. Thus Bd0967 is likely to be trafficked to the periplasm where it is functionally active. An unambiguous N-terminal sequence of QLKEGLE… was determined by ESI-MS, confirming purified Bd0967 had been processed to remove the 24aa signal peptide (ESI-MS was performed by the Stříšovský lab).

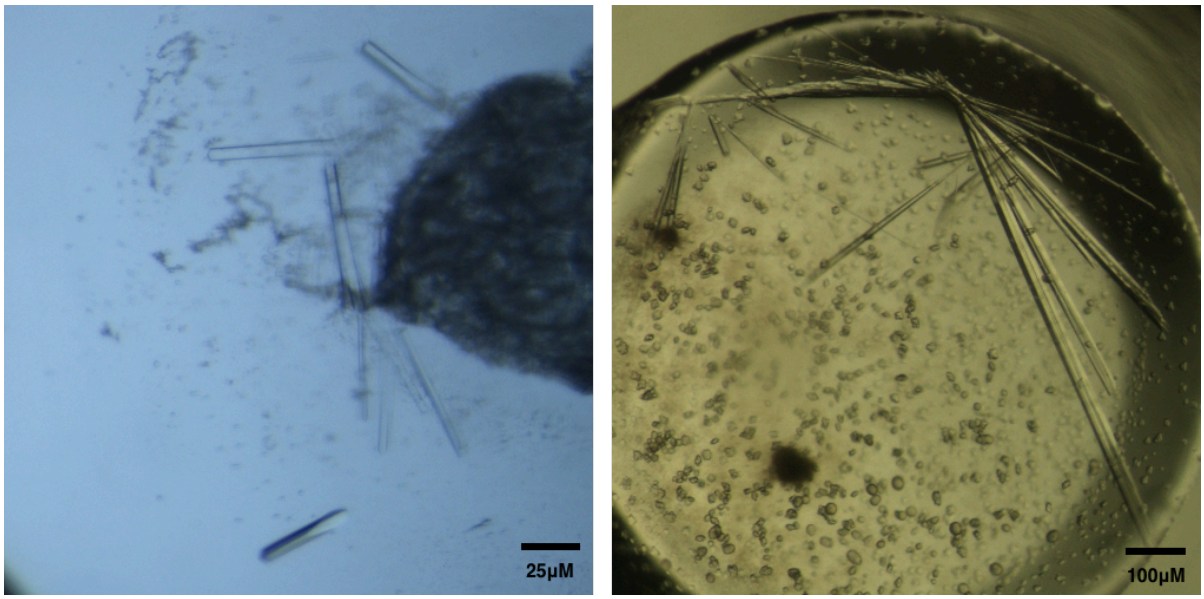


Figure 104. Bd0967 crystallisation
Images of Bd0967 crystals. Higher quality crystals were obtained by using random matrix microseeding as described in Section 2.6.9.

Bd0967 was purified by Immobilised Metal Affinity Chromatography (IMAC) and then polished for crystallography experiments using SEC. Initial crystal hits for Bd0967 were small thin needles. Random Microseed Matrix Screening (rMMS) was implemented [114], which considerably improved the size and quality of the crystals (Figure 104).

6.2.2 The overall structure of Bd0967

Overall Bd0967 resembles a monomeric self-compartmentalised bowl with a hinged lid that creates a large cavity. Bd0967 can be split into four distinct domains, which are represented in Figure 105. Bd0967 has a large α-helix dominated N-terminal domain (NTD), that is comprised of eight bundled α-helices and a single short β-strand at the tip. The C-terminal domain (CTD) of Bd0967 adopts a unique fold compared to other CTPs and is formed of five α-helices and a single β-strand, resembling a paddle-like structure. The NTD and CTD are mainly α-helical domains, and contact each other at their tips (furthest region from protease domain), via a composite β-sheet formed from single β-strands originating from each domain. The more buried NTD provides a scaffold for the CTD to attach to, completing the bowl shaped structure. The CTD of Bd0967 is a unique feature for this class of proteases and acts to self-compartmentalise the protein. The protease domain provides the core of the protein and is of a similar fold to other protease domains of CTPs. The protease domain is made up of two β-sheet regions; a five-stranded mixed β-sheet, and a five-stranded β-sheet formed around a single extended β-strand with a 90° twist. Separating the two β-sheet regions are three α-helices. The protease domain has an additional subdomain (protease cap) that forms the peptide tunnel between the protease domain. The PDZ domain (formed of a five-

stranded, anti-parallel, β-sheet) is attached between the protease domain and the protease cap via a flexible hinge.

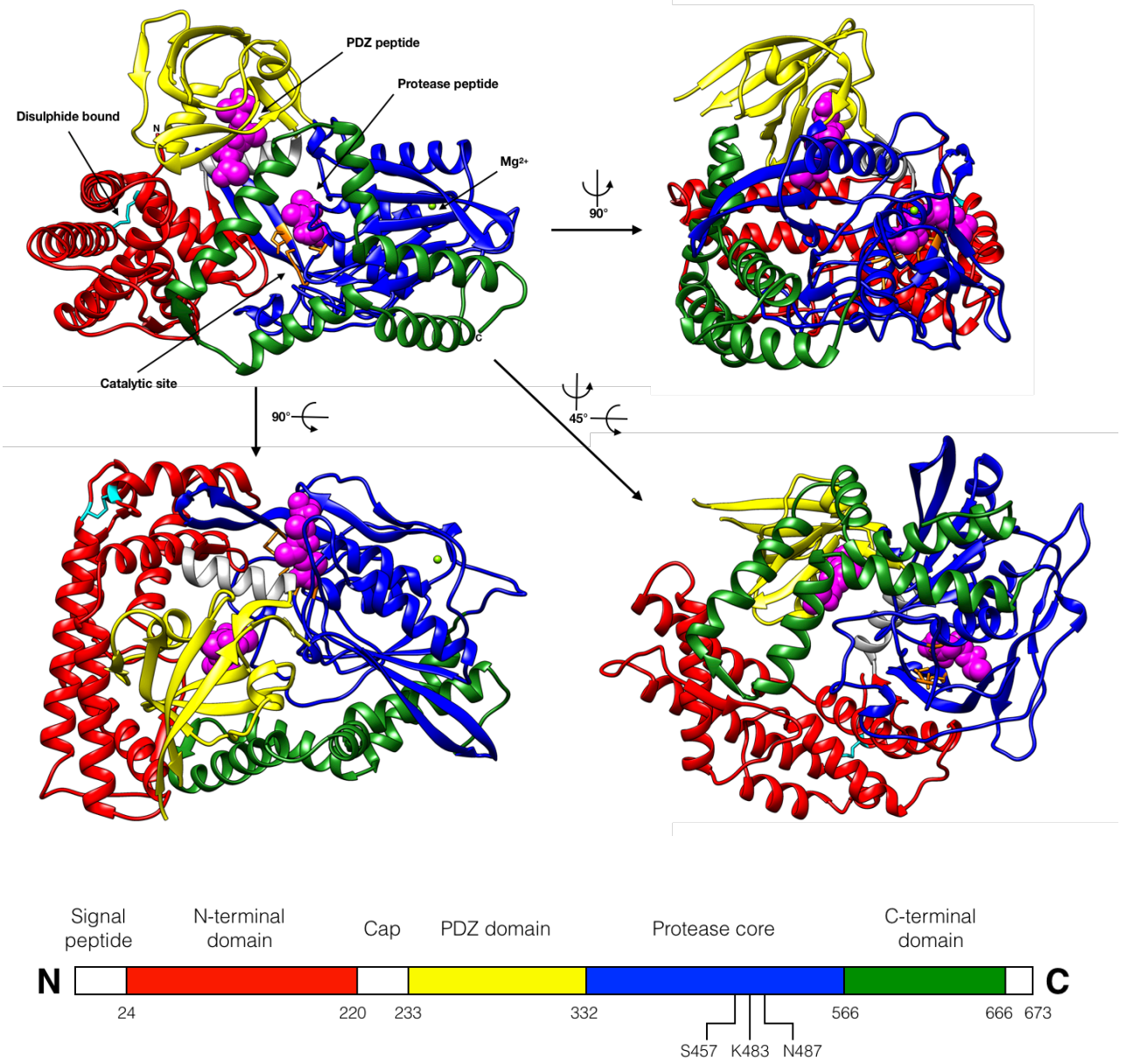


Figure 105. Domain architecture of Bd0967
Bd0967 can be divided into four distinct domains and a subdomain, which are represented by the different colours in this image. The scaffold-like NTD (red), the protease cap helix (grey), the PDZ domain responsible for peptide recognition (yellow), the protease domain (blue) and finally the paddle-like CTD (green). Two peptide fragments (magenta) are bound to the PDZ domain and protease domain. The active-site residues are highlighted in orange, the NTD disulphide bond is shown in cyan and the metal ion bound to the protease domain is represented as a green sphere.

Larger B-factor values are associated with the CTD domain, which suggests that it is more flexible and likely alters conformation to facilitate the movement associated with PDZ domains in this family of enzymes. Higher B-factors are also associated with the PDZ domain, which is known to adopt multiple conformations in other CTPs. A map of B-factors for Bd0967 is shown in Figure 106, which illustrates that the base of the protein (protease and lower regions of NTD and CTD) has lower B-factor values, whereas the upper regions of the protein have higher B-factor value suggesting these are more flexible.

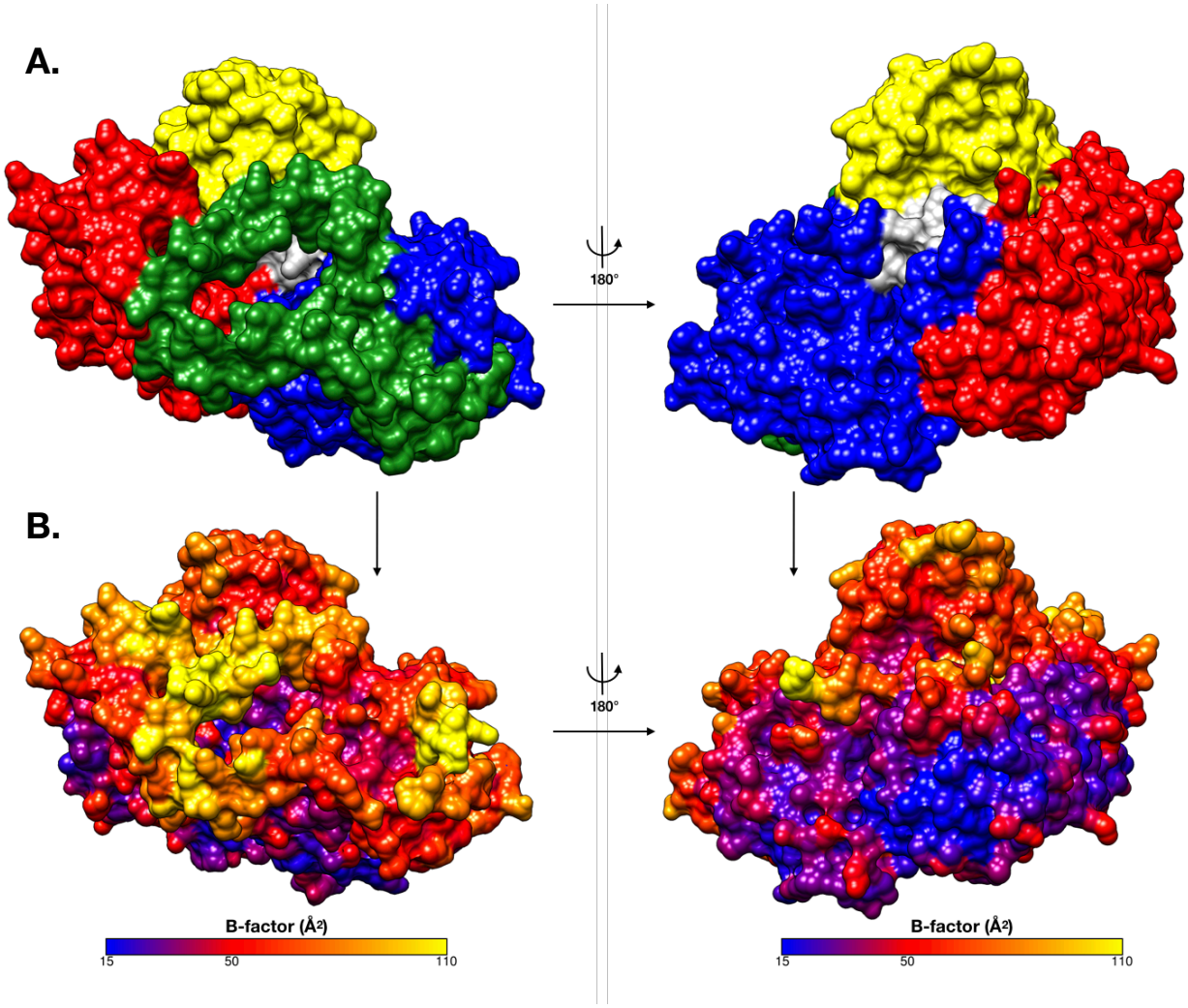


Figure 106. Bd0967 B-factors
A. Represents Bd0967 orientation (front and back), coloured according to domains. **B.** Bd0967 in the same orientation as panel A, coloured according to B-factor. Bd0967 structure has regions of contrasting B-factor values suggesting certain regions such as the CTD and PDZ are more dynamic, whereas the protease domain is more static. Average B-factor values per residue have been coloured using a gradient from low B-factor (blue) to high B-factor (yellow).

6.2.3 Bd0967 cavity

The structure of Bd0967 creates a large self-compartmentalised cavity. The volume of the cavity was accessed using CASTp, which calculates the MS volume to be 19500Å^3 with a pocket surface area of 7200Å^2 [294]. The dimensions of the interior cavity have been calculated roughly as 30Å x 30Å x 30Å (WxHxD). The surface area of the cavity space is represented in Figure 107, coloured in magenta. The cavity would be able to accommodate small unfolded proteins and may provide an environment shielded from bulk solvent.

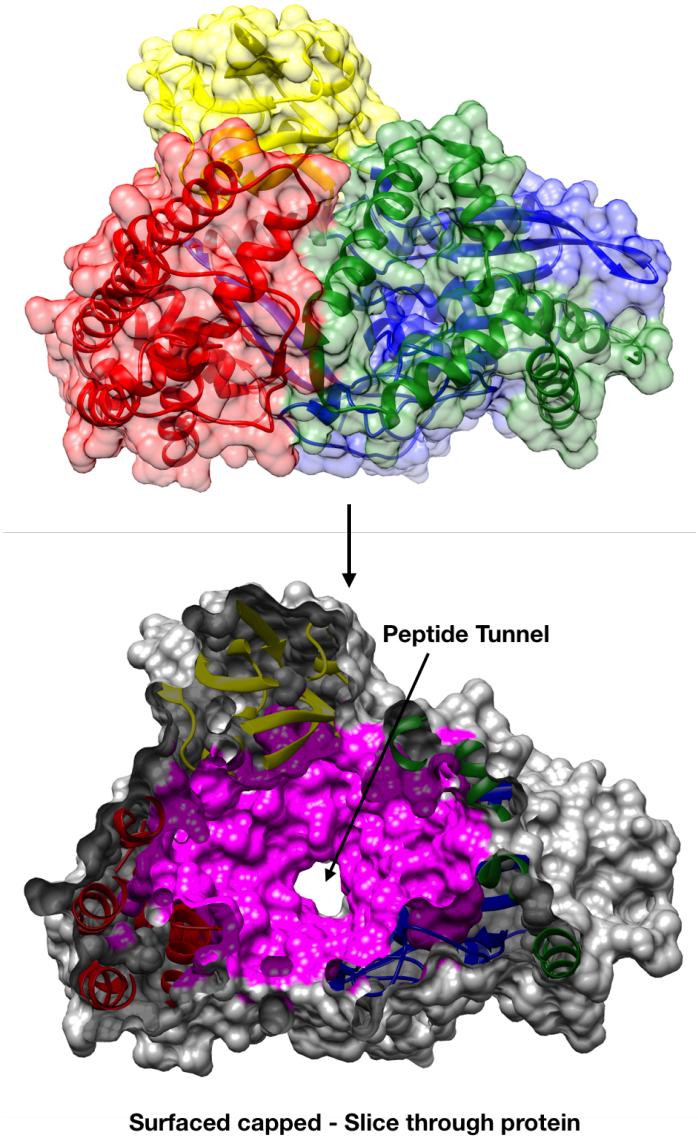


Figure 107. Bd0967 internal cavity
Bd0967 domain organisation creates a self-compartmentalised enzyme with a large internal cavity. The top image shows the exterior surface of Bd0967 coloured according to the domains. The bottom image shows a dissection of Bd0967 (orientation is the same) revealing the internal cavity surface, calculated using CASTp, shown in magenta. The peptide tunnel leading into the internal cavity is labelled.

6.2.4 The Active site / catalytic triad / bound peptide

The active site of Bd0967 is located in the protease domain with the catalytic triad positioned at the end of a shallow tunnel leading into the cavity. The peptide-binding tunnel is formed between the protease domain and protease cap subdomain. The catalytic triad is comprised of the proposed active-site nucleophile S457, the catalytic base K433, and Q487 that makes a hydrogen bond with K433 to orientate it correctly for catalysis similarly to other CTPs.

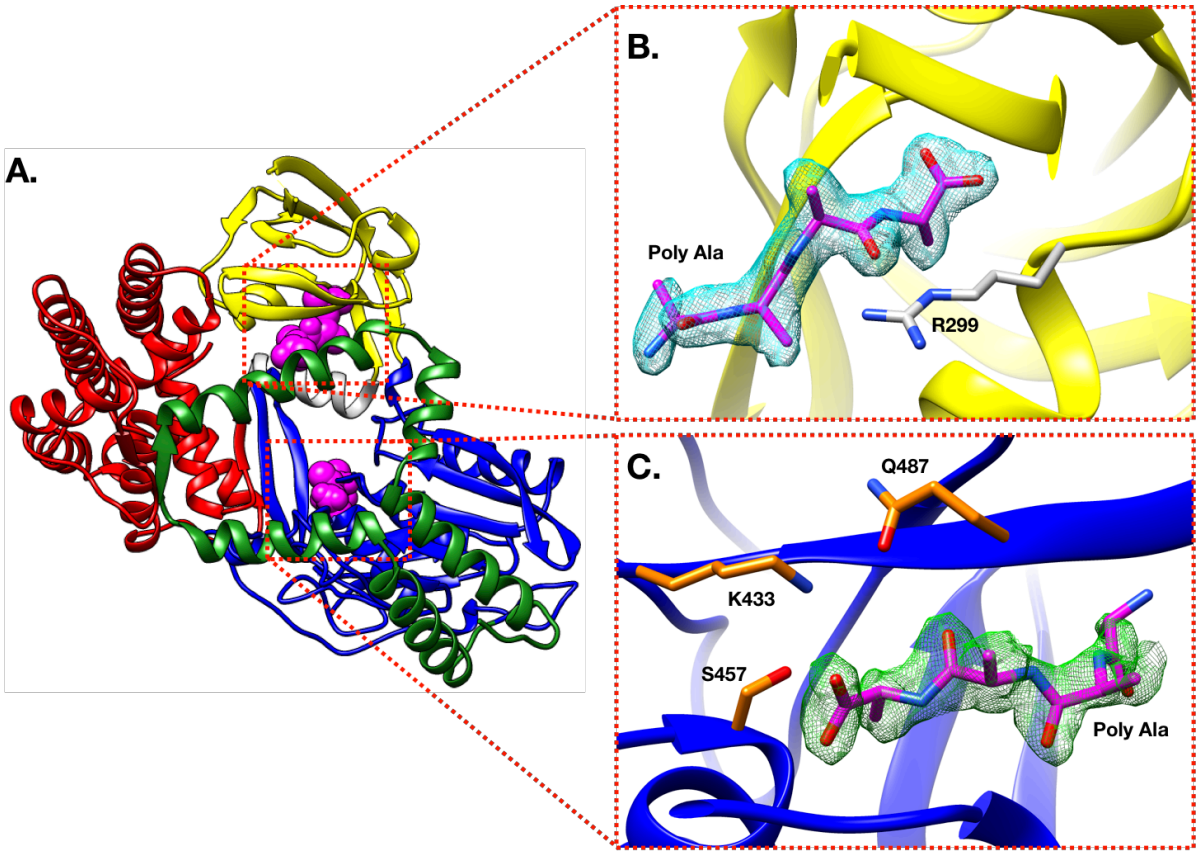


Figure 108. Bound peptides of Bd0967
Bd0967 co-purifies and co-crystallises with two peptide fragments that bind to the PDZ peptide binding site and the active-site peptide binding site. **A.** The peptides coloured magenta in space filling shown to display there positions within the structure of Bd0967. **B.** Shows the electron density map (cyan) of the bound peptide (magenta) at the PDZ domain. R299 is shown in grey for reference. **C.** Shows the electron density map (green) of the bound peptide (magenta) at the protease domain. The active-site residues are shown in orange as a reference. ($2F_o$-F_c map, radius=1.5Å, σ=1.3)

The structure of Bd0967 reveals the presence of two unidentified co-purified and co-crystallised peptide fragments located in both the catalytic site and PDZ peptide-binding groove. Positive F_o-F_c density for the sidechains of the peptide fragments is visible; however, the exact nature of the sidechains cannot be distinguished at this resolution. As a result the peptide fragments have been modelled as a poly-alanine chains (Figure 108).

The peptide fragment at the protease domain makes extensive hydrogen bonding contacts with backbone atoms in the active site. The peptide stacks between a β-strand and the start of a coil immediately prior to a α-helix, and makes an augmented β-sheet interaction (Figure 109). Residues at the P1 position would likely need to be small hydrophobic or uncharged residues because the S1 site is lined with the nonpolar residues L403, I461, V486 and F506. The S2 site, lined with F220, F228, V489 and K500 is slightly more solvent exposed and could easily accept bulkier residues in the P2 position. The P3 could potentially prefer a small polar residue that could interact with the S488 sidechain and V489 backbone atoms.

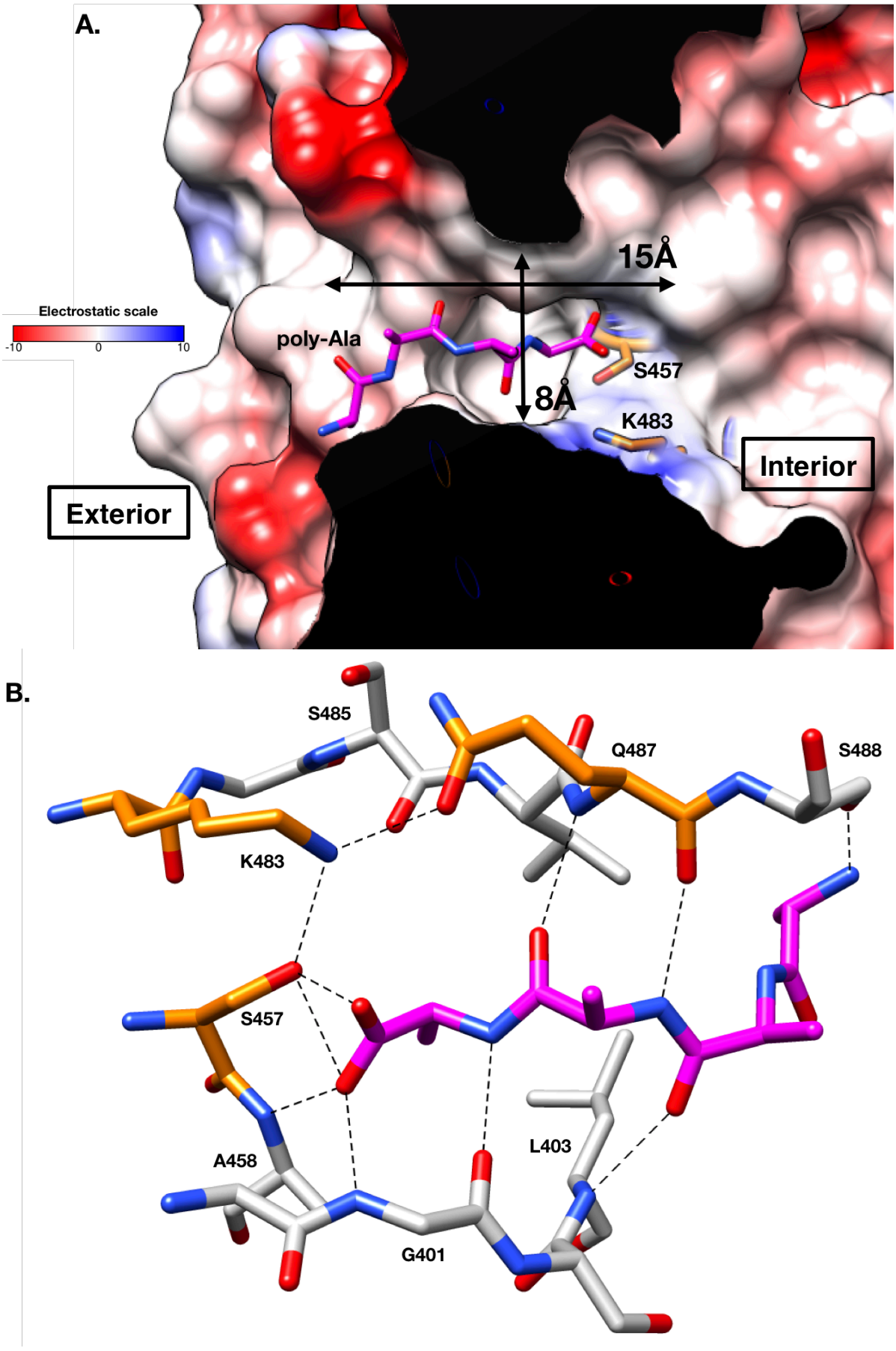


Figure 109. Protease peptide tunnel and binding site
A. Bd0967 has a shallow tunnel leading to the interior cavity, formed between the protease domain and the protease cap subdomain. The tunnel is mainly lined with hydrophobic residues. The active site residues are located towards the end of the tunnel near the interior opening. **B.** Binding of the peptide at the active site is facilitated by multiple hydrogen bonds. The peptide, coloured in magenta, is bound in a beta-augmented fashion. The active-site residues are coloured in orange and other residues participating in the peptide binding are in grey. Black lines depict the hydrogen bonds.

6.2.5 Bd0967 has a unique metal-binding site

Bd0967 possesses a metal-binding site that is unique to this family of CTPs. The metal-binding site is located in a peripheral loop region of the protease domain (Figure 110). The metal is co-ordinated by seven atoms: a water molecule; a single side chain residue (D436) and five backbone carbonyl groups (G411, F414, Q415, D438 and M440). Binding of the metal ion is facilitated by the placement of P439 that allows the specific orientation of backbone carbonyl groups in this region. The binding of a metal ion at this site likely provides a structural role by stabilising the otherwise flexible coiled region. A Mg^{2+} ion has been modelled in this position due to its presence in the crystallisation conditions and is in agreement with the observed diffraction data.

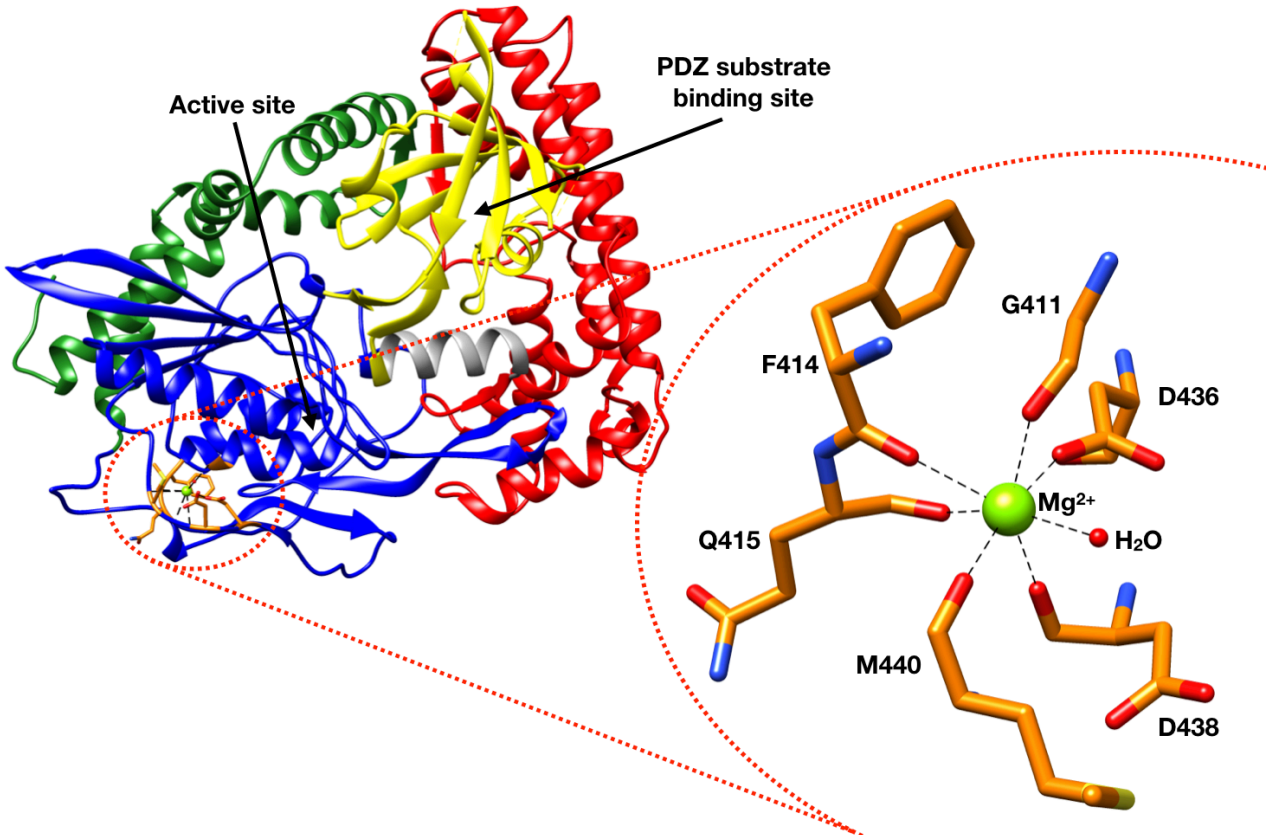


Figure 110. Bd0967 metal-binding site
The metal-binding site of Bd0967 is located in a loop at the base of the protease domain (blue).The residues involved in binding the metal ion are coloured in orange. The Mg^{2+} metal ion is coloured green and makes seven coordinated bonds to five backbone carbonyl groups the carboxylate group of D436 side chain and also to a coordinated water molecule.

6.2.6 Bd0967 has unique disulphide bond

Bd0967 has a disulphide bond located in the NTD and is not a feature previously documented for CTP (Figure 105). The disulphide bond is formed between Cys33 and Cys100 and acts to stabilise the base of the α-helical scaffold domain (NTD). Cys33 is located 4aa from the N-terminal peptide resolved in the structure of Bd0967, beyond this point no electron density is observed. The disulphide bond likely acts to stabilise the N-terminus and therefore the whole NTD, which provides the scaffold for the structure of Bd0967. The existence of a bona fide disulphide bond adds further evidence that Bd0967 possesses a signal peptide and is likely to be transported to the periplasm for its specific function.

6.2.7 Substrate recognition by the PDZ domain

The PDZ domain of Bd0967 is attached to main body of Bd0967 on a flexible hinge. At the base of the hinge are two hydrophobic residues, L234 and L332 that could easily slide past each other, in a greased motion, to facilitate the movement of the PDZ domain.

An additional peptide fragment is bound in the PDZ peptide binding groove, which is believed to originate from the same protein as the peptide fragment bound to the protease domain. We have modelled the peptide fragment, bound at the PDZ binding groove, as a four amino acid poly-alanine chain. Although positive density in the refined F_o-F_c electron density map exists for residue sidechains, at the resolution obtained the sequence register is ambiguous, furthermore MS was unable to identifying the peptide fragment.

The PDZ peptide binding groove is located between helix-12 (larger PDZ helix) and strand-3 of the PDZ domain (Figure 111). The peptide C-terminal carboxylate group is anchored by the carboxylate-binding loop (G238-I239-G240-A241) similarly to other PDZ-Peptide complexes. The peptide chain is bound in a β-augmented fashion, hydrogen bonding to the backbone of β-strand 4. I239, G240, A241, T242 and L243 all make direct hydrogen bonds to the peptide via their backbone atoms.

The PDZ peptide binding groove is mainly lined with hydrophobic residues. The binding pocket for the sidechain of the P(0) residue is mainly lined with hydrophobic residues such as I298, I239 and L243, which suggests a smaller hydrophobic residue would be preferred at the P(0) position. However, the presence of A241 at the start of strand-4 allows access to a deeper hydrophobic pocket, lined with leucine and isoleucine residues, which may allow bulkier residues to be accommodated at the P(0) position. The penultimate residue, P(-1) would preferentially be a non-polar residue. The binding pocket for P(-1) residue is also lined with hydrophobic residues (I230, L332 and L236), although it is more solvent exposed. The P(-2) and P(-3) binding pockets are solvent exposed, making it difficult to structurally determine a binding preference. However, the P(-2) sidechain is orientated towards the residues L291, V295 and K292, suggesting polar residues may also be tolerated in this open binding pocket. The PDZ-binding groove and protease binding site are highly similar, suggesting they are able to accommodate similar ligands.

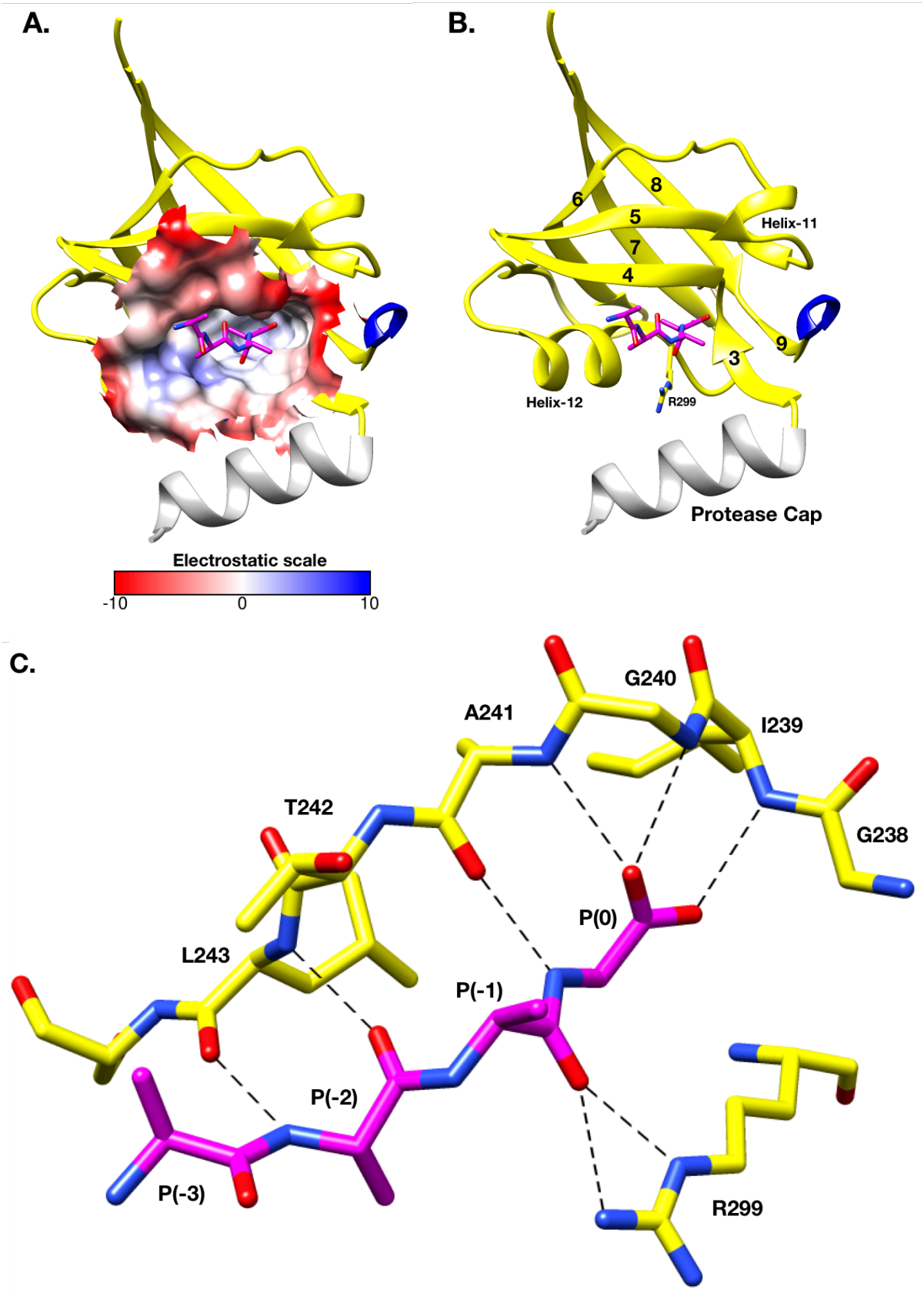


Figure 111. Peptide binding PDZ domain
A. The peptide binding groove of the PDZ domain is highly hydrophobic. The surface of the peptide binding groove is shown coloured according to the coulombic scale. **B.** Image of PDZ domain binding to peptide in same orientation as (A) to highlight the relative position of R299. The strands and helices are labelled for reference to text. **C.** Hydrogen bond network involved in peptide binding at the PDZ domain.

In addition the sidechain of R299 makes likely important hydrogen bonds to the peptide chain. The R299 sidechain aligns parallel to the peptide chain making a β-sheet like contact. R299 sidechain makes additional hydrogen bonding contacts with E226 and D227 of the protease cap (Figure 112). The H-bonds R299 makes are likely to be important for substrate sensing and activation of the protease domain, similarly to the observed conformational change observed between the 'resting' and 'active' states of CtpB [85].

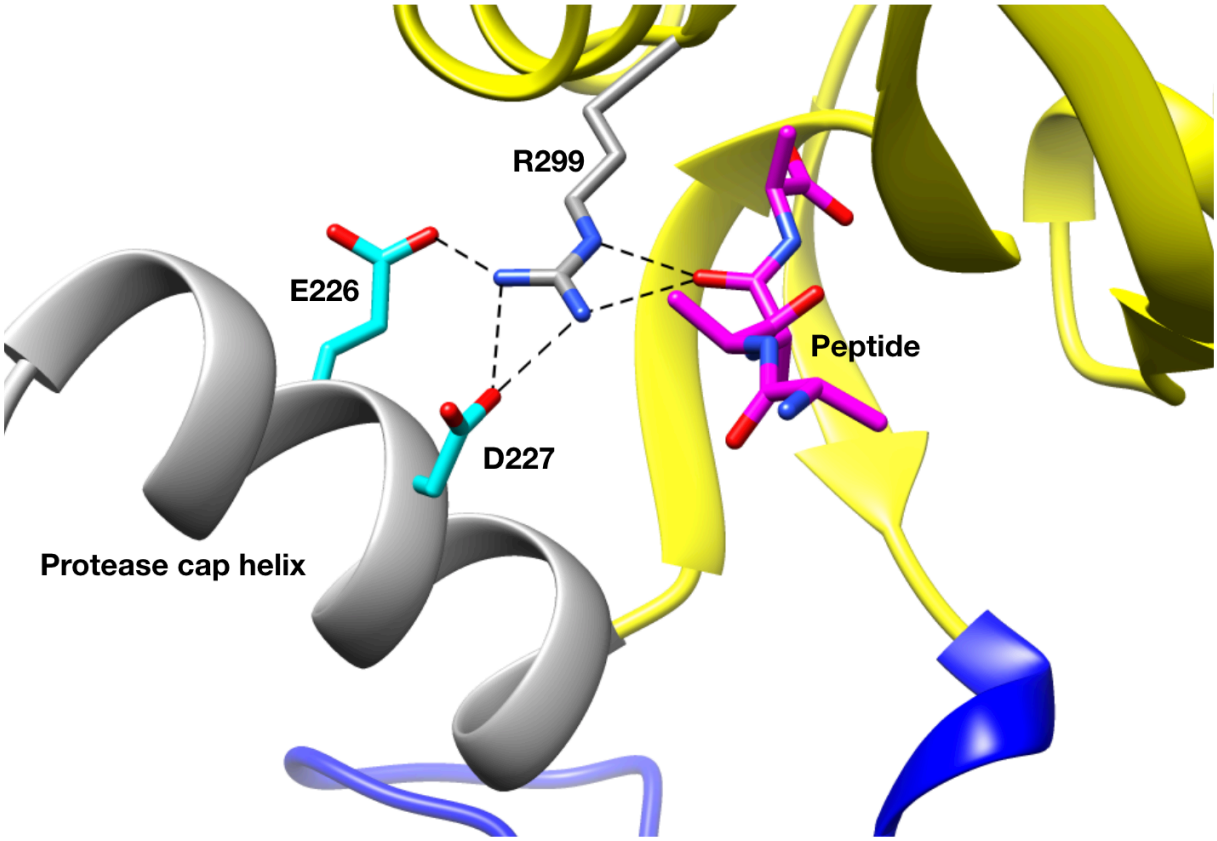


Figure 112. Arg299 interaction with peptide and protease cap helix
Peptide binding at the PDZ domain is facilitated by R299 by imitating a beta strand. R299 makes similar hydrogen bonding contacts with the protease cap helix that are responsible for switching the protease domain on and off in CtpB. R299 hydrogen bonds to two acidic residues E226 and D227 (cyan) as well as maintaining hydrogen bonds with the bound peptide (magenta).

6.2.8 A monomer in solution

SEC was used to determine the oligomeric state of Bd0967 in solution. The crystal structures also suggest Bd0967 is a monomer; however, homologues of Bd0967 have been shown to exist as functionally active dimers. SEC of Bd0967 (Wt, S457A and R199M constructs) revealed all exist in a monomeric form; however, a sizeable void peak was present for all constructs.

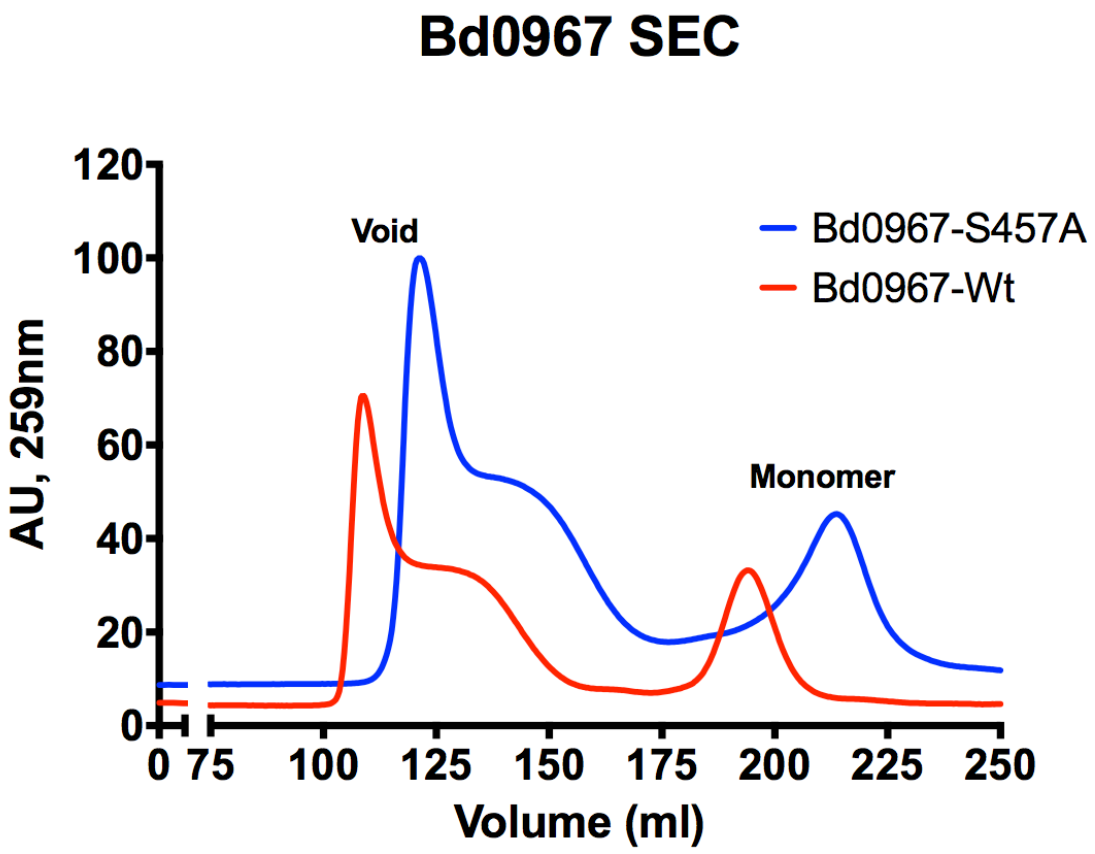


Figure 113. SEC results for Bd0967
The size exclusion traces for Bd0967 WT and S457A are shown. Monomeric Bd0967 elutes at approximately 190 ml. A large void peak is observed for Bd0967.

6.2.9 Structure of Bd0967 S457A

The structure of Bd0967-S457A is highly similar to the structure of Bd0967-WT, adopting the same fold and domain arrangement. Little conformational change is observed between WT and S457A structures with the exception of a slight rigid body rotational movement of the PDZ domain (6°). Although this slight conformational change is not seen for all chains (two out of four), suggesting conformational sampling of the PDZ domain. Furthermore, bound peptides are located at the PDZ peptide binding site and active site of Bd0967-S457A demonstrating substrates are still able to bind even though catalysis cannot be performed.

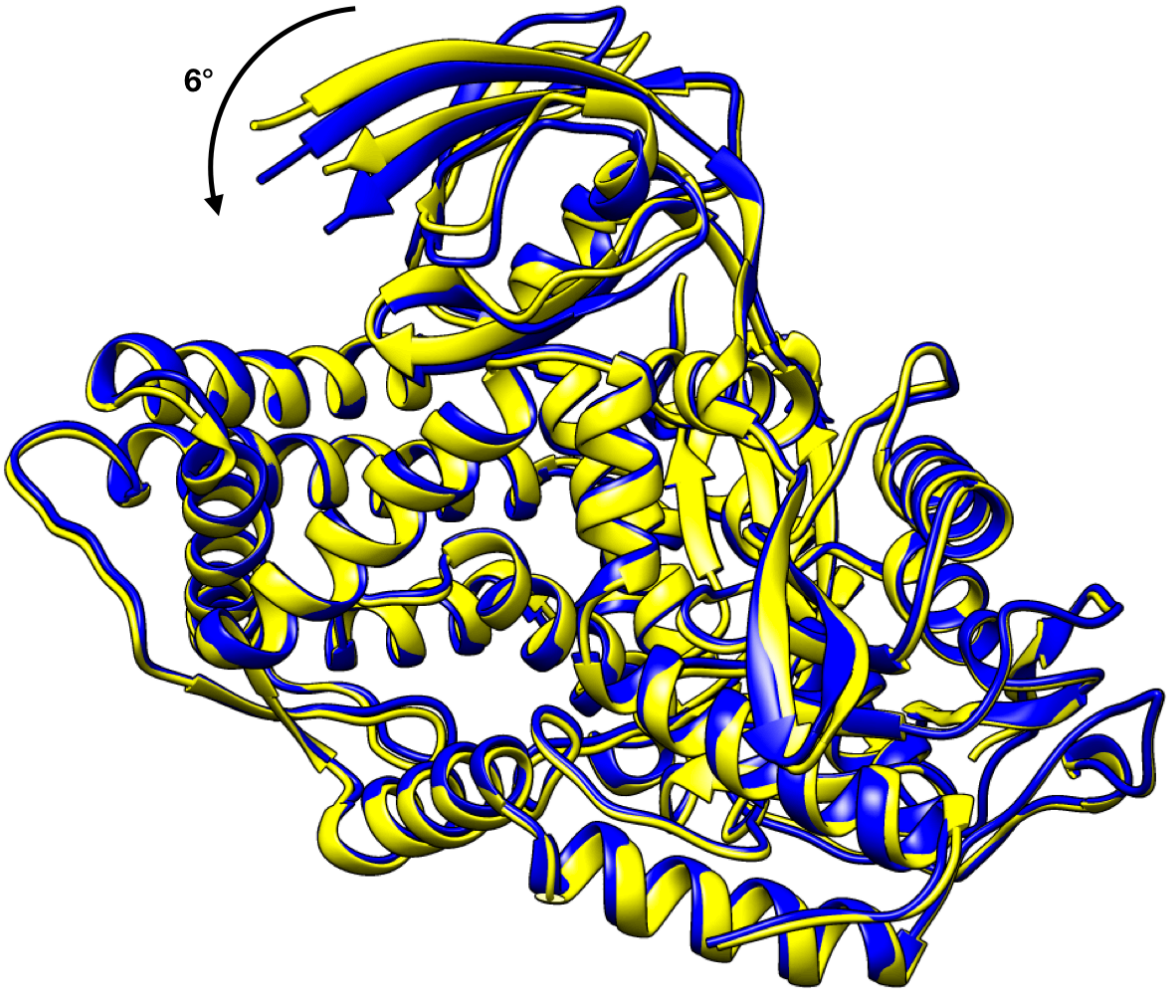


Figure 114. Structure of Bd0967-S457A
Bd0967-S457A (yellow) superimposed onto the structure of Bd0967-WT (blue) demonstrates the slight conformational change observed for the PDZ domain of Bd0967-S457A. Superimposition was performed using the SSM superpose tool in coot (length = 634 residues, 1.13Å).

6.2.10 Identifying Bd0967 substrates / interacting partners

MS and Edman degradation experiments were performed to identify the co-purified and co-crystallised peptides (collaboration with the Stříšovský lab, Czech Academy of sciences). These techniques were used comparatively with Bd0967-Wt, S457A and R299M mutants. Bd0967-R299M mutant was designed to prevent peptide binding at the PDZ peptide-binding groove, thereby prevent substrate recognition. However, identification of the bound peptide has been unsuccessful. Furthermore, electron density present in the $2F_o$-F_c map is not easily interpretable and ambiguous for the sidechains of the bound peptide. This may be because the density is representative of multiple peptide fragments and is therefore an average of these. Bd0967 may be promiscuous, accommodating multiple ligands in the absence of the true (HD100) substrate.

To uncover the substrate of Bd0967 (or a binding partner, like Prc and NlpI), we performed a pull-down assay with *Bdellovibrio bacteriovorus HD100* lysate. Pull-down assays were performed with both soluble and insoluble protein fractions, as described in Section 2.6.3. Bd0967-WT was used as the bait protein to detect interacting proteins. Interacting partners were only detected when the insoluble fraction of the cell lysate (resuspended in 2% Triton X-100) was used. Two distinct interacting proteins bands are visible on the SDS-PAGE for the Bd0967 sample lane, which are not visible in either control samples (Figure 115).

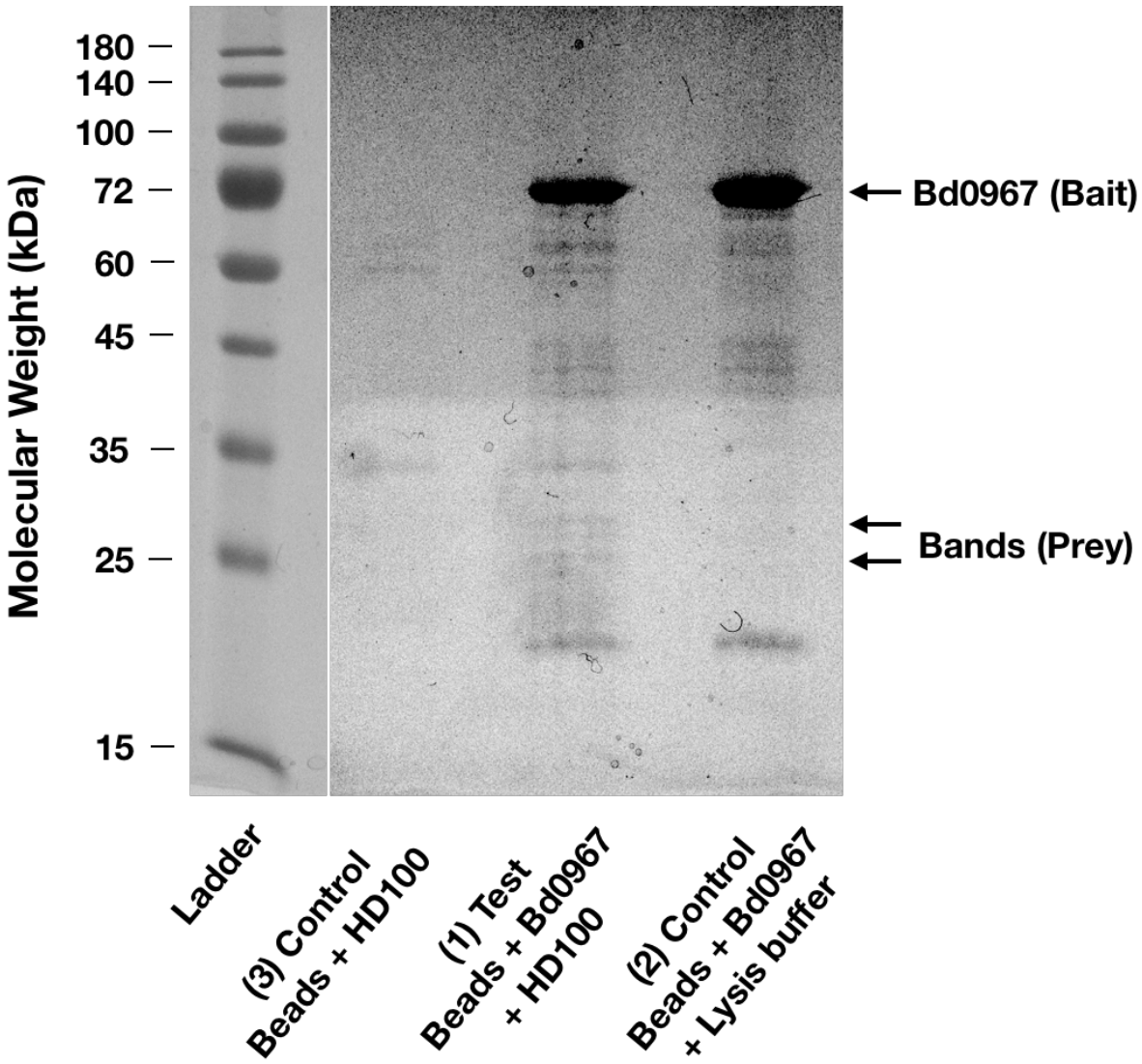


Figure 115. Bd0967 pull-down assay
The SDS-PAGE of the pull-down assay with Bd0967 as the bait, shown with the upper black arrow. The two faint bands in the test sample (1) are highlighted by the two lower black arrows.

MS was carried out to identify the interacting proteins detected from the pull-down assay. Whole lane, trypsin-digested samples were analysed for: (1) beads + Bd0967 + HD100 lysate test sample; (2) Beads + Bd0967 + lysis buffer only control sample and (3) Beads + HD100 lysate control sample. MS results indicated that several proteins were more abundant in the Bd0967 test sample (1) compared to the controls. Moreover, several proteins were only detected in the Bd0967 test sample (1). The stand out candidate for the (potential) interacting partner of Bd0967 was the flagella sheath protein Bd0606, only identified in sample (1) with a normalised spectrum of 33, which was considerably higher than other identities (normalised spectra of ~4). Other candidates include Bd0604, a flagellin that was present in samples (1) and (3) with a normalised spectra ratio of 20:4 and Bd3052, a flagellin that was present in (1) and (3) with a normalised spectra ratio of 37:5. A list of the top five proteins found in the three samples is shown in Table 16.

All of the candidates identified through MS from the pull-down assay had very similar C-terminal sequences, which was intriguing because Bd0967 likely recognises its substrate via direct interaction with the C-terminus. The most promising substrates Bd0606, Bd0604 and Bd3052 had highly similar C-terminal sequences, which are shown in Figure 126.

Identified in Samples:	Organism	Protein Name	Uniprot accession code	Gene Name	Protein Size	Normalised total spectra			Terminal residues
						(1)	(2)	(3)	
(1)	*Bdellovibrio HD100*	Flagellin	Q6MQ75_BDEBA	Bd0606	29 kDa	33	0	0	ALRLIG
	Bdellovibrio HD100	Serine protein kinase	Q6MR70_BDEBA	Bd0228	80 kDa	9	0	0	GDVKNK
	Bdellovibrio HD100	Methyltransferase	Q6MP37_BDEBA	Bd1029	31 kDa	4	0	0	QARART
	Bdellovibrio HD100	ATP synthase subunit beta	ATPB_BDEBA	Bd3897	51 kDa	4	0	0	AKKLQA
	Bdellovibrio HD100	Probable signale peptide	Q6MK10_BDEBA	Bd2606	22 kDa	4	0	0	LIKFIP
(1) + (2)	*E. coli Strain K12*	cAMP-activated global transcriptional regulator CRP	CRP_ECOLI	crp	24 kDa	33	37	0	VVYGTR
	E. coli Strain K13	ATP-dependent RNA helicase Rhl	RHLE_ECOLI	rhlE	50 kDa	12	21	0	KPAAAQ
	E. coli Strain K14	3-ketoacyl-CoA thiolase	FADA_ECOLI	fadA	41 kDa	7	9	0	TVFERV
	E. coli Strain K15	Succinylornithine transaminase	ASTC_ECOLI	astC	44 kDa	7	9	0	VSRGSS
	E. coli Strain K16	tRNA (guanosine(18)-2'-O)-methyltransferase	TRMH_ECOLI	trmH	25 kDa	6	9	0	TMQAAG
(1) + (3)	*Bdellovibrio HD100*	Inosine-5'-monophosphate dehydrogenase	Q6MLD1_BDEBA	Bd2081	52 kDa	33	0	37	NYRIES
	Bdellovibrio HD100	Octaprenyl diphosphate synthase	Q6MQB9_BDEBA	Bd0557	37 kDa	26	0	39	WRRKPS
	Bdellovibrio HD100	Flagellin	Q6MIU6_BDEBA	Bd0352	29 kDa	37	0	5	ALKLIG
	Bdellovibrio HD100	Thioredoxin reductase	Q6MQT7_BDEBA	Bd0373	34 kDa	20	0	23	EAQEAH
	Bdellovibrio HD100	Flagellin	Q6MQ77_BDEBA	Bd0604	29 kDa	20	0	4	VAMKLV
(1) + (2) + (3)	*Bdellovibrio HD100*	Tail-specific protease	Q6MP94_BDEBA	Bd0967	75 kDa	315	376	14	KQANAK

Table 16. Mass Spectrometry results for Bd0967 pull-down assay
The top five proteins identified for each category is displayed in this table. Significantly increased normalised spectra (for proteins identified from sample (1) = beads + Bd0967 + Lysate is highlighted in green. Yellow indicates the low normalised spectra for proteins in control sample that had increased normalised spectra in sample (1).

6.2.11 Bd0967 fully degrades the flagellin, Bd0606

To evaluate the MS results for the Bd0967 pull-down assay, N-terminally tagged constructs of Bd0606 and Bd3052 retaining the native C-terminal sequence were created. Bd0606 was successfully expressed and purified whereas Bd3502 appeared to form inclusion bodies. Consequently only Bd0606 was used in further experiments with Bd0967.

A pull-down experiment, similar to the one that identified Bd0606 as an interacting partner, was carried out to validate the MS results. The N-terminal His-tag was removed from Bd0606 before being used in the pull-down assay, by pre-incubation with thrombin and then re-purified. Bd0606 was shown to interact with Bd0967 Wt and S457A, but not with R299M, albeit not in a 1:1 stoichiometry (Figure 116). However, this may have been due to a number of reasons: such as the binding of Bd0967 to the resin may have prevented the mobility required to facilitate interaction; the binding conditions were not optimised; and also that Bd0967 may be turning over the Bd0606 substrate. Furthermore, 1:1 stoichiometry is not necessary in an enzyme:substrate relationship.

To examine whether Bd0606 was a substrate for Bd0967, a protease digestion assay (wherein Bd0606 was incubated at 37°C with Bd0967 for 1 hour) was carried out. The abundance of Bd0606 after incubation was analysed by SDS-PAGE to determine whether degradation had occurred. In a 1 hour time period Bd0606 is almost completely degraded when incubated with Bd0967-Wt, whereas the abundance of Bd0606 in the control sample (Bd0606 incubation without the addition of Bd0967) remained constant (Figure 117).

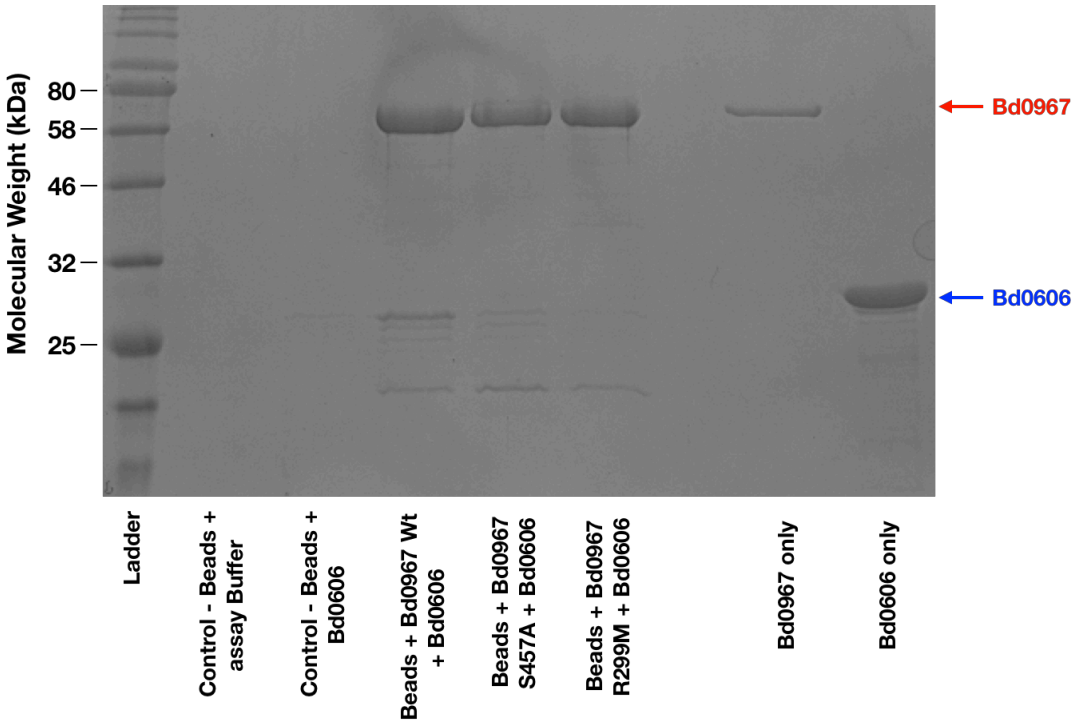


Figure 116. Confirm binding
A similar pull-down assay was re-performed with purified Bd0606 to evaluate the MS result. Bd0606 binds to Bd0967-Wt and to a lesser extent Bd0967-S547A. Bd0606 does not bind to the PDZ binding mutant Bd0967-R299M. The binding is weak (not 1:1 stoichiometry) under the assay conditions.

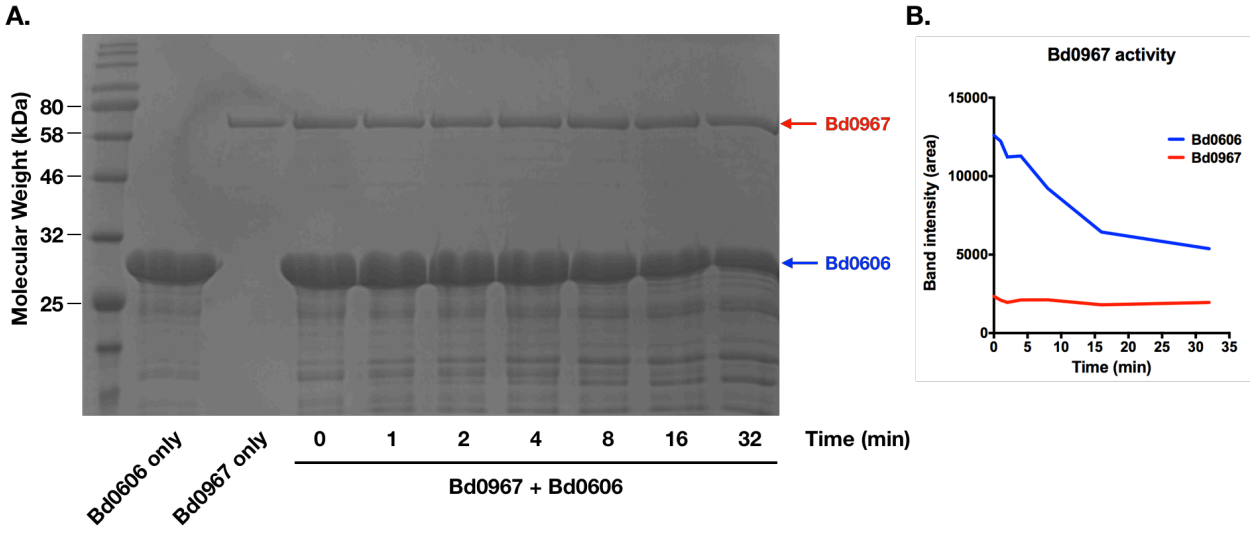


Figure 117. Bd0967 degrades Bd0606
A. Bd0967-Wt degrades Bd0606 in a processive manner, which was monitored over a time period. The red arrow highlights Bd0967 and the blue arrow highlights Bd0606. **B.** Graphical illustration of Bd0606 being degraded over time (band intensity) alongside Bd0967.

A similar result was observed when Bd0606 was incubated with the catalytic mutant Bd0967-S457A. Bd0606 remained at a similar level throughout the 1 hour incubation when incubated with Bd0967-S457A mutant suggesting that degradation is being performed by Bd0967 in the Wt sample. Furthermore, a Bd0967-R299M (PDZ peptide binding mutant) was also incapable of degrading Bd0606. This finding supports the importance of the PDZ domain interaction with Bd0606 (substrates) to facilitate catalysis.

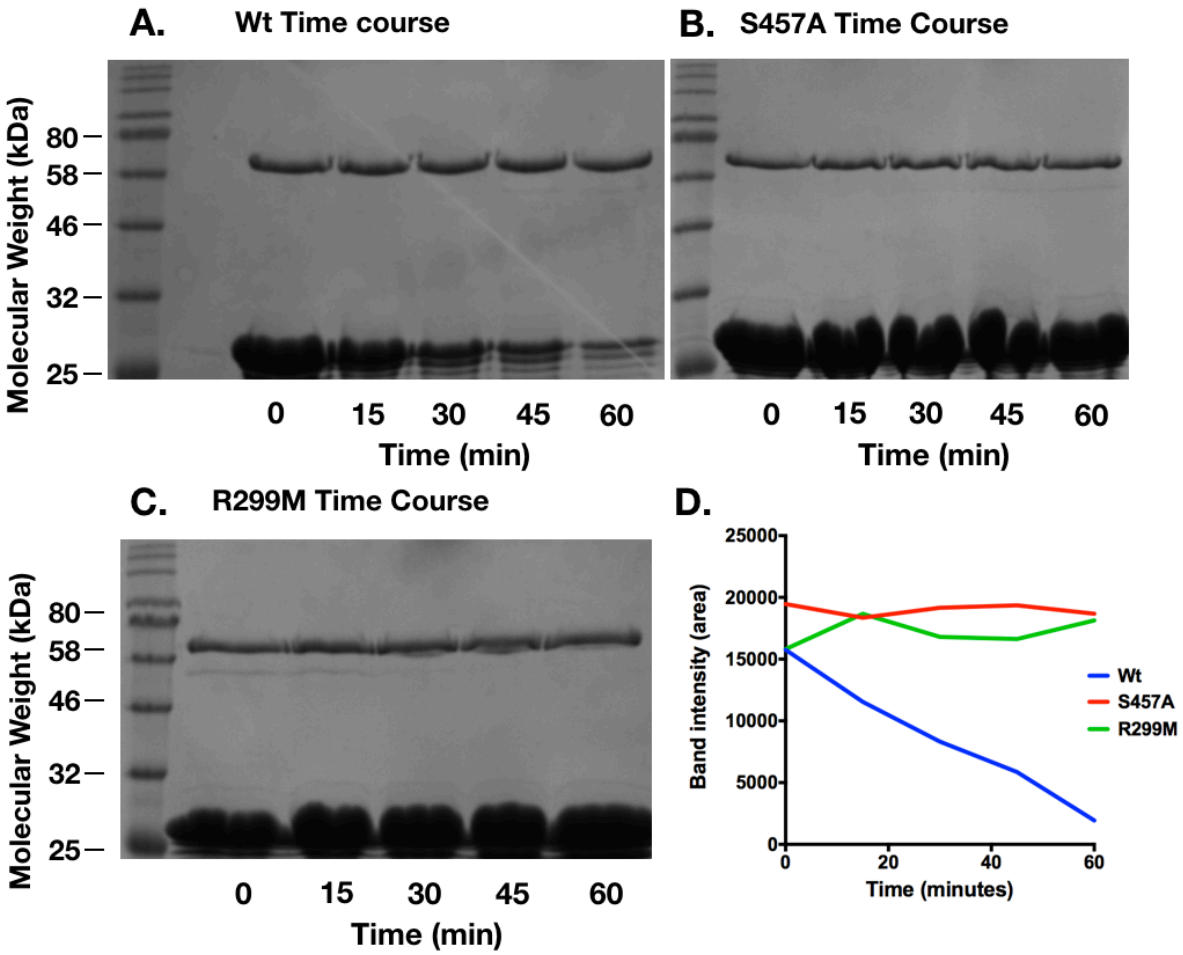


Figure 118. Bd0967 activity in comparison to mutations
A. Bd0967 Wt degradative activity towards Bd0606 in a 60 minute time period with samples taken at regular time intervals. **B.** S457A mutant does not show any activity. **C.** R299M also lacks the ability to degrade Bd0606 efficiently due to a non-functional PDZ domain. **D.** A graphical representation showing the degradation of Bd0606 when incubated with Bd0967-Wt, S457A and R299M mutants. Image J was used to analyse the intensity of gel bands.

To determine whether Bd0967 was specifically targeting Bd0606 for degradation we also investigated the activity of Bd0967 towards a variety of other *Bdellovibrio* proteins of varying size and C-terminal sequences (Figure 119). This assay did not reveal any evidence that Bd0967 was active towards any of these other proteins. Furthermore Bd0108 (a small intrinsically disordered protein) was also not a substrate for Bd0967 suggesting that a disordered C-terminal sequence is not the only prerequisite to binding and catalysis. Additionally, Bd0967 showed no degradative activity towards Bd1996 and Bd0314, which are both similar molecular weights to Bd0606.

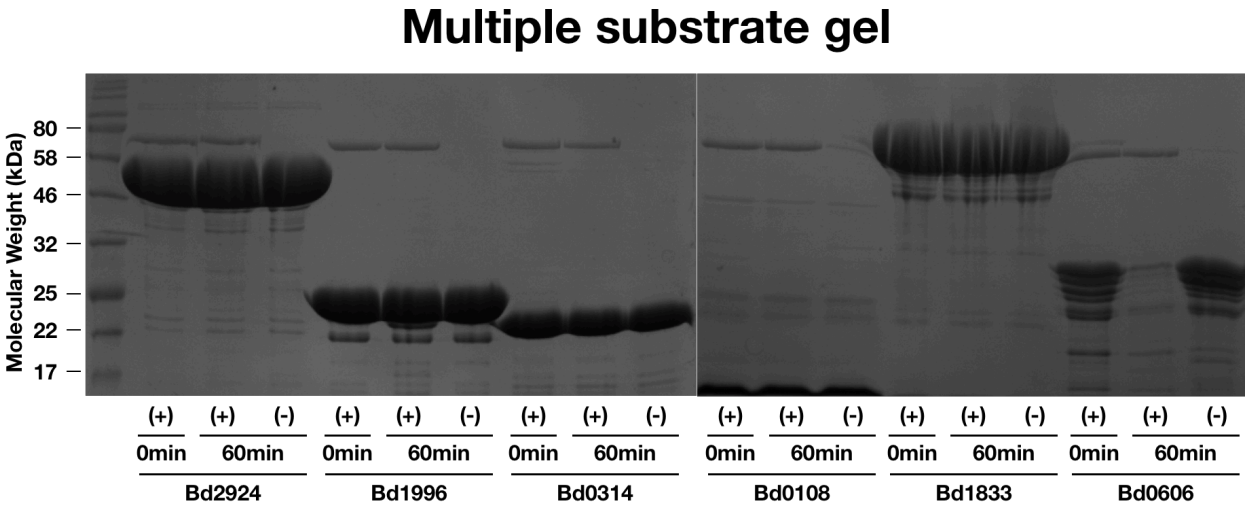


Figure 119. Bd0967 activity towards various proteins
Bd0967 degradative activity has been analysed against *Bdellovibrio HD100* proteins of varying sizes and C-terminal sequences. Bd0967 showed no activity towards any of the proteins trialled in this assay. Bd0108 is a small intrinsically disordered protein that remained un-digested suggesting the folded state of the proteins is not a factor determining Bd0967 degradative ability.

To investigate the mechanism behind Bd0606 recognition and degradation by Bd0967, we produced C-terminal truncation mutants of Bd0606. These mutants, ΔLIG and ΔRLIG, lack the terminal three and four amino acids, which typically interact with the PDZ domain of CTPs. To our surprise Bd0967 was able to degrade these truncation mutants of Bd0606 with similar efficiency to WT Bd0606 (Figure 120). This finding

along with the Bd0967-R299M mutant finding suggests that although a functioning PDZ domain is absolutely required for recognition of the C-terminal sequence and initiating catalysis; Bd0967 also has a degree of promiscuity and tolerance in peptide binding at the PDZ domain. On the other hand, Bd0967 may also adopt an additional mechanism of substrate recognition, which could also explain these results.

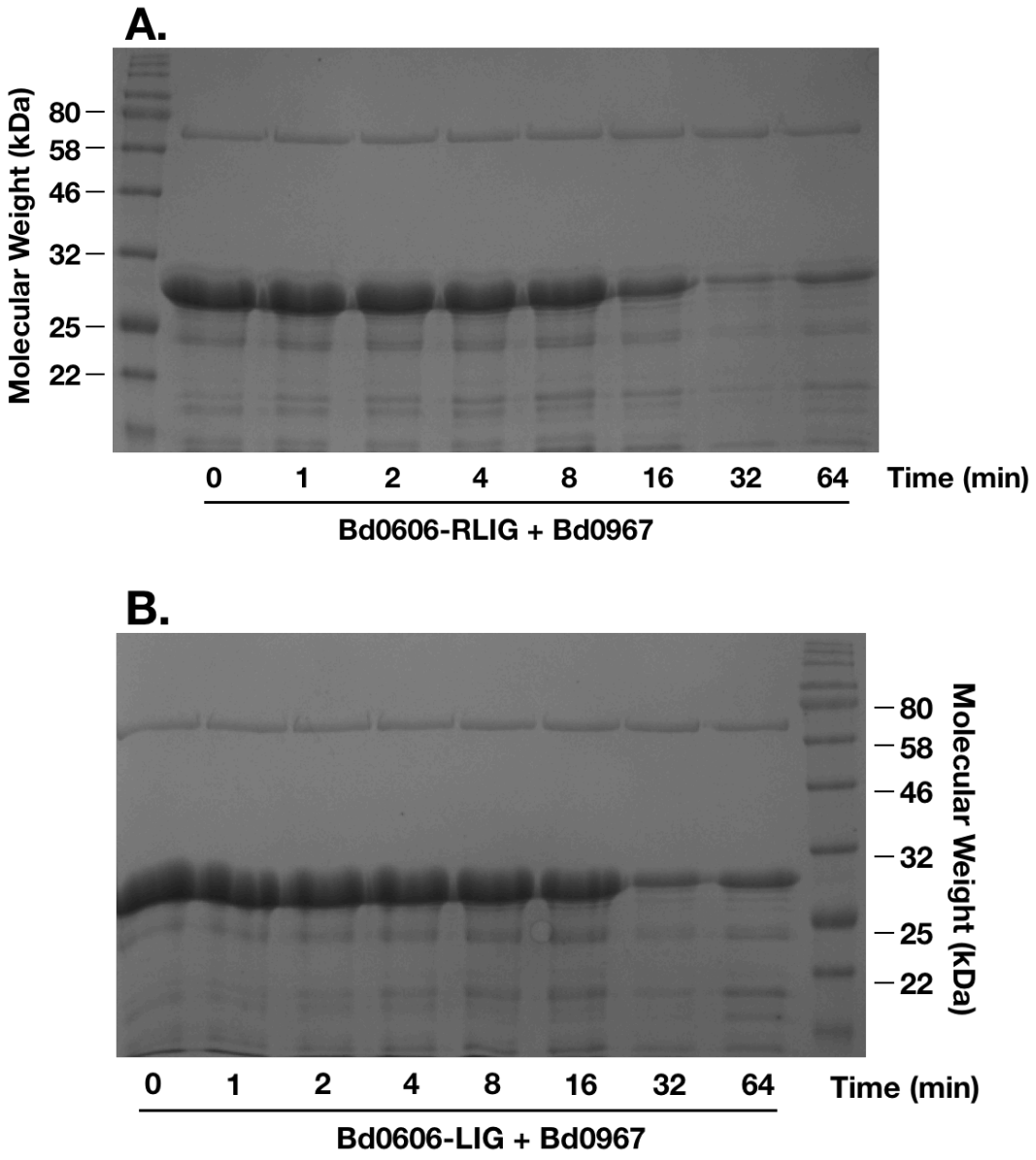


Figure 120. Bd0606 C-terminal mutants
A. Bd0606 mutant RLIG lacks the C-terminal four amino acids, which bind to the PDZ domain. **B.** Bd0606 mutant LIG lack the three C-terminal amino acids. Bd0967-Wt is able to degraded both RLIG and LIG mutants to a similar degree as WT Bd0606 substrate. This finding suggests Bd0967 PDZ binding site allows promiscuity and also suggest Bd0606 may be recognised via an alternative mechanism distinct from direct interaction with the C-terminal.

6.2.12 The C-terminal peptide of Bd0606 as an inhibitor of Bd0967 activity

We designed and created a 12aa peptide matching the C-terminal sequence of Bd0606 (NAMPNSALRLIG) in order to investigate whether C-terminal binding was required for Bd0967 degradative activity. This 12aa peptide of Bd0606 was kindly prepared by Oliver Daubney by N-terminal peptide synthesis to ensure that the α-carboxylate was present. Bd0606 was incubated with Bd0967-Wt in buffer containing increasing concentration of peptide. The results of this assay show that increasing the concentrations of peptide reduces the degradation of Bd0606 (Figure 121). At low concentrations of peptide Bd0967 remains active and the levels of Bd0606 are reduced, whereas at a high peptide concentration the degradative activity of Bd0967 is reduced. This result demonstrate that the peptide fragment can act as an inhibitor of Bd0967, presumably by interacting with and saturating binding of the PDZ domain, which disables the activity of Bd0967.

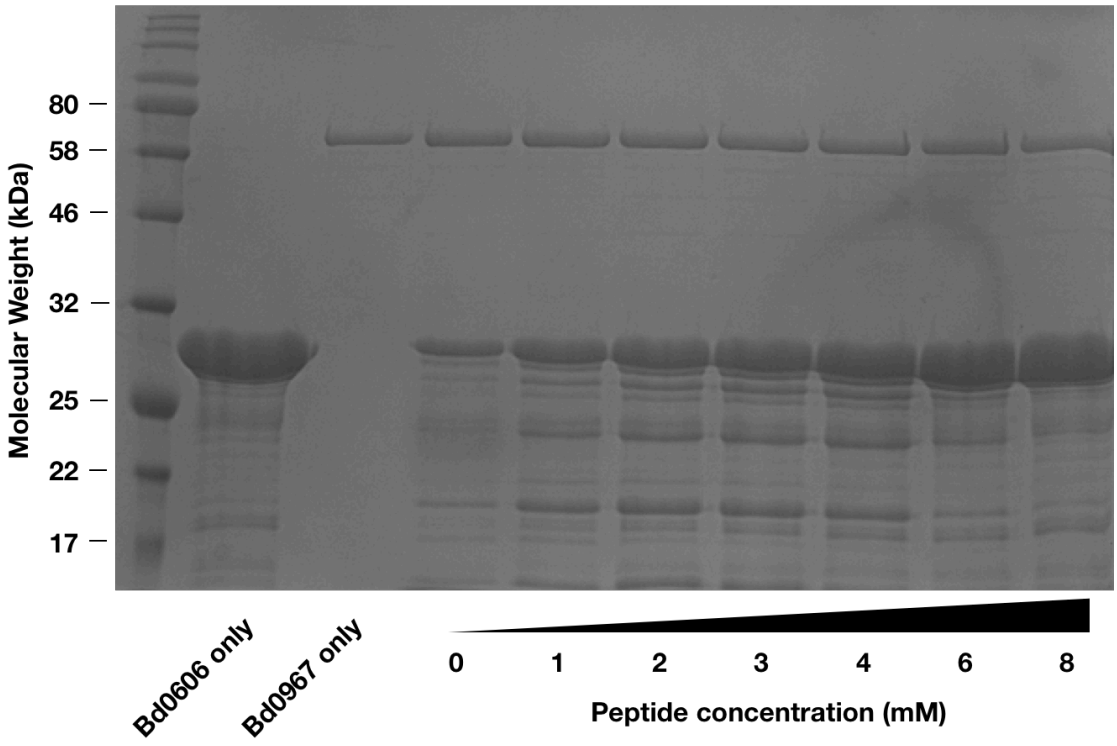


Figure 121. Peptide inhibition assay
Bd0967 activity was measured in the presence of an increasing concentration of a 12aa peptide fragment, synthesised to match the C-terminus of Bd0606. At high peptide concentrations Bd0967 degradative activity towards Bd0606 was reduced indicating the PDZ domain is important for substrate recognition.

6.2.13 Bd0606 is folded

Circular Dichroism (CD) was performed on Bd0606 to determine whether it was correctly folded. CD is a good biophysical technique for determining the proportion and nature of a folded protein based on a distinct UV spectrum for α-helix and β-sheet. The CD spectrum revealed that Bd0606 is folded, with α-helices being the major secondary structure element, which is expected for a flagellin protein based on previous structural information [295]. This finding supports Bd0606 being correctly folded and Bd0967 is not degrading unfolded proteins in a quality control degradative manner.

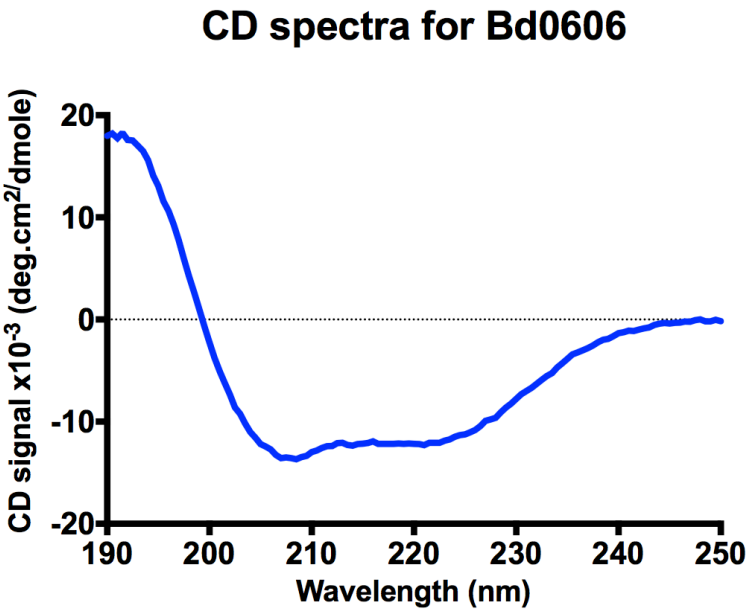


Figure 122. Bd0606 CD spectra
CD spectra for Bd0606 reveals Bd0606 is folded and is formed predominantly of α-helices. Minima peaks at ~208 nm and 222 nm and a maxima peak at 193 nm exemplify a α-helix folded proteins.

6.2.14 Bd0967 KO phenotype

A *Bdellovibrio* mutant lacking Bd0967 (*Δbd0967*) was created by the Sockett lab (a difficult process, for which we extend our gratitude). *Δbd0967* has a striking, developmentally impaired phenotype, wherein these mutant *Bdellovibrio* grow and divide at a considerably slower rate compared to Wt. *Δbd0967 Bdellovibrio* remained in the bdelloplast >7 and even up to >24 hours post invasion in comparison to WT cells (that exit the cell fully developed after approximately 3.5 hours). This finding suggests that *Δbd0967* interferes with a signalling cascade that allows the successful progression through the predatory lifecycle.

The Sockett lab also used electron microscopy (EM) to analyse *Δbd0967*, which revealed a pleiotropic phenotype. There was an abundance of *Δbd0967* cells that had erroneously formed flagellar that were far greater in length (>15 μM) when compared to WT (4 μM) and also had a thicker sheath (Figure 123). As well as the flagellar phenotypes multiple *Δbd0967* progeny appeared to be stalled in growth phase, with many filamentous progeny not dividing correctly. These findings suggest that Bd0967 has a cellular role that regulates the levels of flagellin sheath protein, which in turn may act as a novel signalling factor during cell division in *Bdellovibrio*.

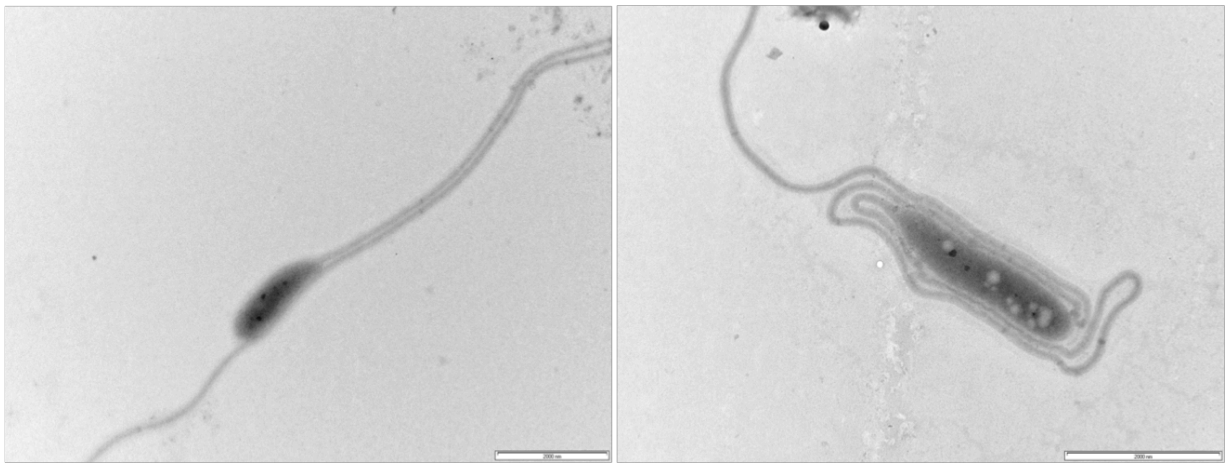


Figure 123. Bd0967 KO mutant *Bdellovibrio*
Bd0967 KO mutants have a pleiotropic developmental phenotype that involves the flagellum assembly. The flagellum of Bd0967 KO *Bdellovibrios* have been seen to grow to ~5 times the length of Wt, suggesting flagellin assembly is unregulated. These two images show the large mutant flagellar. **(Electron Micrograph images kindly provided by the Sockett lab)**

6.2.15 Bd0967 interaction with CdG

MST was used to determine whether Bd0967 interacted with CdG. A K_D value of 22 μM was obtained, suggesting weak non-specific binding. Furthermore, Bd0967 has a signal-peptide, suggesting transport into the periplasm where CdG is unlikely to be present.

6.3 Discussion

6.3.1 The structure of Bd0967

The structure of Bd0967 (a hollow sphere, containing a self-contained cavity) represents another divergent CTP, whose fold has diverged from a minimal CTP fold alike that of D1P [84]. The protease domain and PDZ domains are structurally conserved in CTPs; however, the NTD and CTD of CTPs are structurally divergent, which impacts on the proteins function. For example in CtpB, the NTD and CTD act as dimerization domains creating an enclosed ring shaped protease [85], whereas in Prc the NTD and CTD create the rims of a bowl shape [87].

The active-site residues of Bd0967 correspond to those found in CtpB and CT441, where a catalytic triad has been established [85, 86]. In contrast in Prc and D1P a catalytic dyad has been proposed. Although the suspected third catalytic residue (Asn) of triad is conserved in sequence alignment, the position it adopts in the structure suggests it plays no role during catalysis [84, 87].

Bd0967 possesses two unique structural features. Unlike other CTPs Bd0967 possesses a disulphide bond located close to the N-terminus that provides additional structural and stabilising support to the NTD. Moreover the formation of a disulphide bond supports the prediction that Bd0967 is localised to the periplasm, similarly to other CTPs [296]. Additionally Bd0967 has a metal-binding site that is located in the protease domain. The metal-binding site appears to provide structural stability to the protease domain and not believed to regulate proteolytic activity.

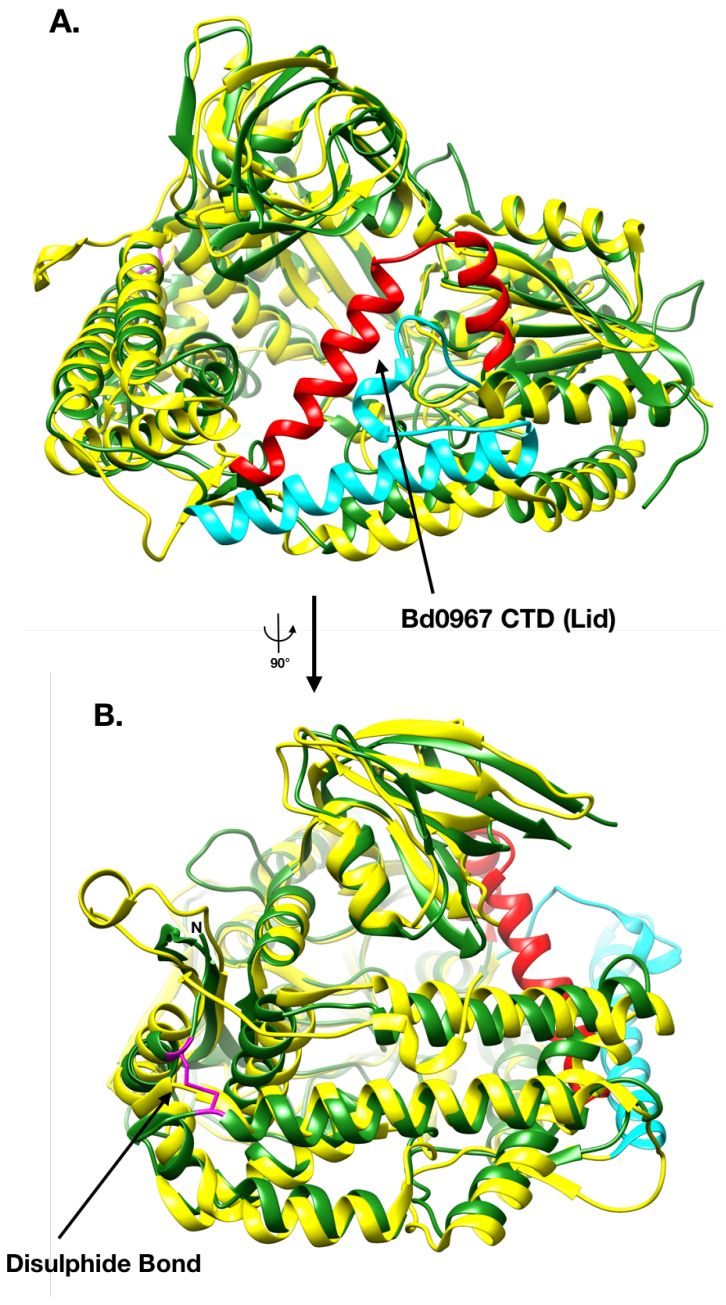


Figure 124. Bd0967 comparison to *E. coli* Prc
A. The structure of Bd0967 (green) is superposed with *E. coli* Prc (yellow, PDB:5WQL). Superimposition was performed using the SSM superpose tool in coot (length = 536 residues, rmsd = 2.01Å). The proteins three-dimensional structures are highly similar and overlap well in all domain except from the CTD. The CTD of Bd0967 is shown in red, while the CTD of Prc is shown in Cyan. The CTD of Bd0967 forms more of a paddle-like structure creating a cavity shielded from bulk solvent, whereas Prc CTD forms one wall of a bowl shape. **B.** Bd0967 also has a disulphide bond (magenta) located in the NTD, which is not found in other C-terminal processing proteases.

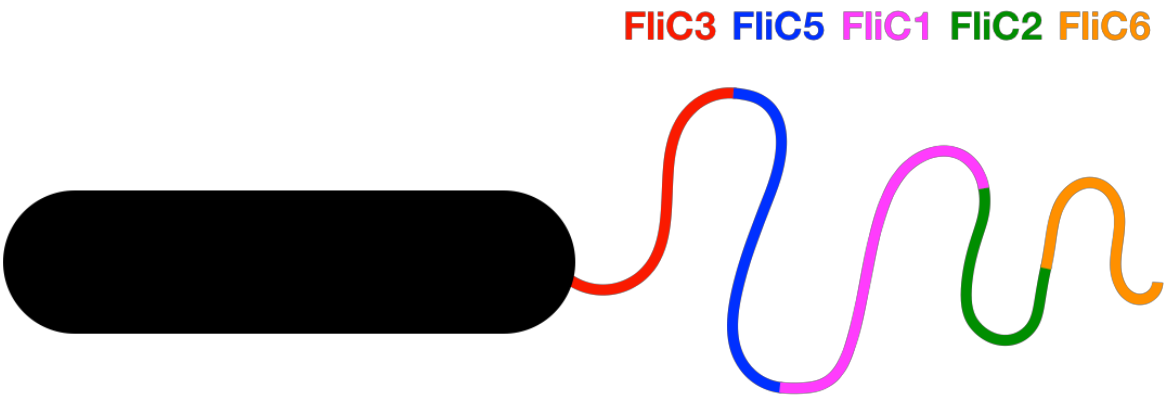
6.3.2 Bd0967 activity towards the flagellin Bd0606 and mechanism

The activity of Bd0967 is unlike that of D1P and CtpB that cleave a specific segment from the C-terminus of substrates to process them. Instead we have demonstrated Bd0967 fully degrades its substrate, Bd0606, in a processive manner similarly to the activity described for Prc and CT441.

Bd0606 is a flagellin, a major subunit component that forms the helical structure of bacterial flagella. Flagella are required for bacterial motility and in *Bdellovibrio* the flagellum is required to encounter prey. Flagella are assembled in a specific type III secretion system where the axial proteins such as flagellin are secreted [297]. Flagellin are synthesised in the cytoplasm as monomers that have a disordered termini [298]. Chaperones are used to retain monomeric flagellins and traffic them for export [299]. Flagellins polymerise only when they encounter a template, hence they self-assemble at the distal end of the growing filament [297].

These results have shown by mutatingR299 (the residue associated with peptide binding in the PDZ domain) and by using a 12aa C-terminal peptide inhibitor (sequence of Bd0606 C-terminus) that the PDZ domain is required for the substrate recognition and proteolytic activity. Bd0967 is also demonstrated to be promiscuous in its activity towards its substrate Bd0606 and is still able to degrade the ΔLIG and ΔRLIG Bd0606 truncation mutants. Surprisingly in the ΔLIG truncation mutant an arginine residue would be located at the C-terminus, which is suspected to be unfavourable due to the nature of the PDZ peptide binding pocket at the P(0) position. Thus Bd0967 may also have a secondary recognition mechanism, similarly to Prc wherein the adaptor protein (NlpI) is required for efficient proteolytic activity [87]. Bd0606 has a disordered and

hydrophobic C-terminal sequence making it a suitable substrate for Bd0967. Moreover other *Bdellovibrio* flagellins have highly similar C-terminal sequences (Figure 125), suggesting they may also be potential substrates, which requires experimental verification.



Flagellin	Size (aa)	C-terminal sequence
FliC1 (Bd0604)	274	ANQLPQVAMKLV
FliC2 (Bd0606)	277	NAMPNSALRLIG
FliC3 (Bd0408)	282	VALGLLNKSFQG
FliC4 (Bd0410)	277	NQSQQLALKLLG
FliC5 (Bd3052)	277	NQSNQLALKLIG
FliC6 (Bd3342)	277	NQANQLALKLIG

Figure 125. Flagellin assembly *Bdellovibrio*
Bdellovibrios flagellum is composed of six flagellins in an ordered arrangement from the proximal end to the distal end, producing a tapered wave formation. The location of each flagellin species, within the flagellum, is illustrated above in the cartoon representation. Only FliC4 location has not been determined. The table highlights the similarities of flagellin C-terminal sequences.

We speculate Bd0967 operates similarly to CtpB, whereby the PDZ domain conformational change, upon substrate binding, switches between 'resting' and 'active' states. In the 'resting' state the PDZ domain adopts a position close to the protease domain (observed in CtpB and D1P). In the 'active' state the PDZ domain adopts a position away from the protease domain (observed in CtpB and Prc). In Bd0967 the PDZ domain also adopts a position away from the protease domain, hence presumed the structure of Bd0967 is presumed to be in the 'active' state. Figure 126.A demonstrates the comparative positions of the PDZ domain from Bd0967, CtpB, D1P and Prc.

In Bd0967 the PDZ domain makes many H-bonds with residues from both the NTD and CTD. These H-bond contacts are not observed in CtpB but are observed in PRC, suggesting the PDZ domains may function differently in Bd0967 and Prc. The PDZ domain in Prc is required for activity, whereas in CtpB the PDZ domain acts as a regulator of activity [85, 87]. The PDZ domain in Bd0967 is believed to be able to undergo the same dynamic PDZ domain movement associated with other CTPs and relocate from the 'resting' state to 'active' state upon peptide binding. This movement would require a large conformational shift, breaking and reforming of a number of non-covalent bonds and requiring flexibility in other domains. High B-Factors in the CTD and PDZ domain, as well as the observed slight conformational sampling of the PDZ domain between WT and S457A structures suggest PDZ rearrangement, similarly to CtpB, is feasible. Furthermore the sizeable internal cavity of Bd0967 could accommodate the folded PDZ domain.

Outlined in Figure 126.B is a cartoon mechanism for Bd0967 activity. Bd0967 would start in a 'resting' state, where the PDZ domain is tucked inside the bowl, with the peptide binding groove located adjacent to the peptide tunnel. The disordered tail of its substrate, Bd0606, would thread through the peptide tunnel and interact with the PDZ domain. Peptide binding would trigger a conformational change in the PDZ domain to an outward swung position, creating a mechanical pulling force that pulls the substrate through the protease tunnel, unfolding the substrate in the process. The conformational change likely repositions the catalytic residues into catalytically competent residues, in a process similar to that observed in CtpB [85]. When the correct sequence is registered in the catalytic peptide binding groove, peptide cleavage can occur. The protease and PDZ binding sites are highly similar and can likely accommodate similar peptide sequences. As a result the new C-terminus of the cleaved substrate would also be a ligand for the PDZ domain and would explain the processive degradation mechanism observed for Bd0967.

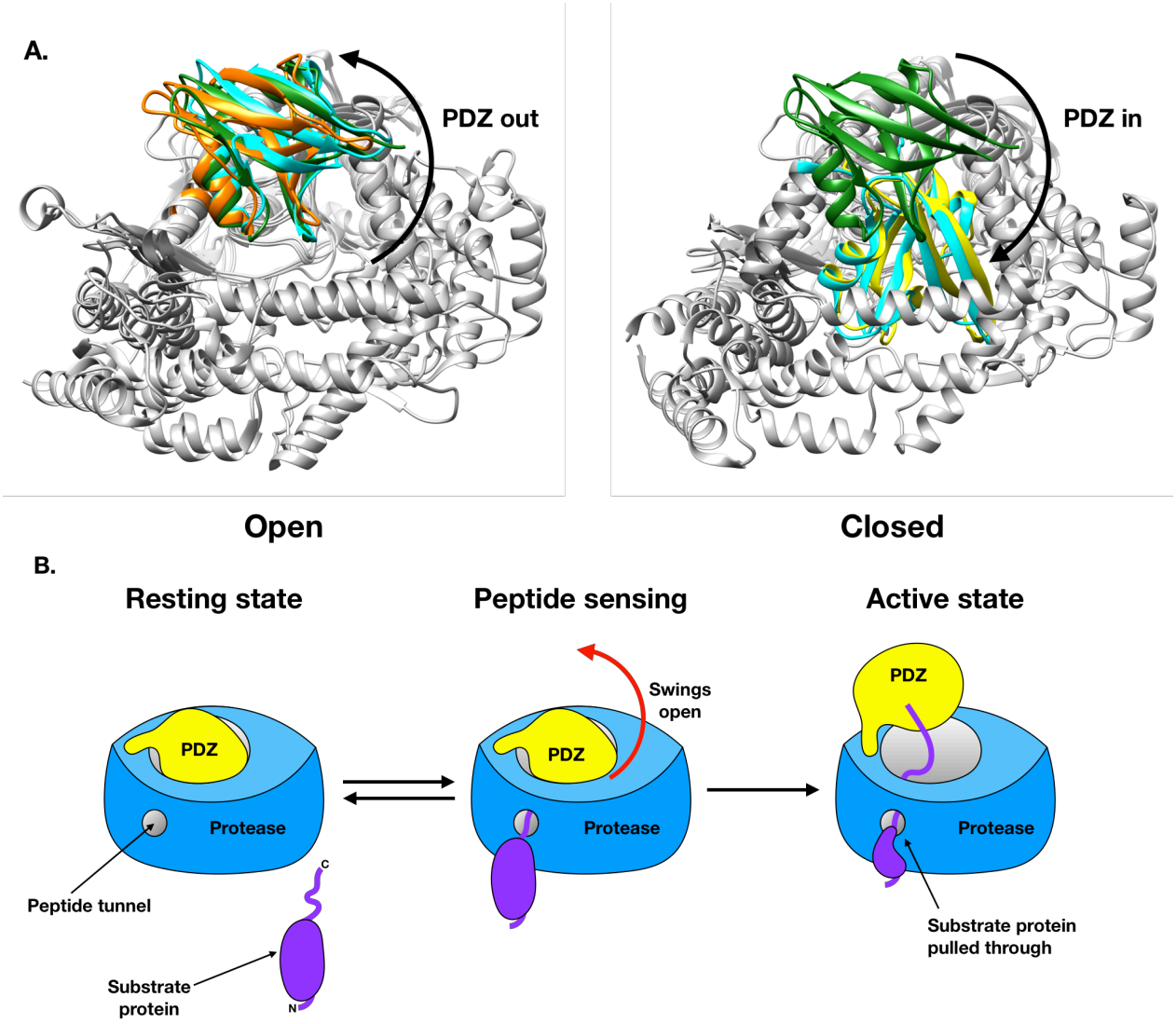


Figure 126. Bd0967 mechanism
A. The movement of the PDZ domains in C-terminal processing proteases is associated with regulating protease activity. Bd0967 PDZ domain (green) is in an outward conformation and its location is compared, relative to other PDZ domains of C-terminal processing proteases. Bd0967 PDZ also aligns well with CtpB-active (cyan, PDB:4C2C, length = 302 residues, rmsd = 2.37Å) and Prc (orange, PDB:5WQL, length = 536 residues, rmsd = 2.01Å) PDZ domains in the active "out" position. The PDZ domains of CtpB-resting (cyan, PDB:4C2E, length = 222 residues, rmsd = 3.40Å) and D1P (yellow, PDB:1FC7, length = 234 residues, rmsd = 3.03Å) have been shown to adopt an inward position (resting state), fitting close to the protease peptide binding tunnel. Superimposition was performed using the SSM superpose tool in coot. **B.** A cartoon illustration of the proposed mechanism of Bd0967. Bd0967 exists in the resting state with the PDZ domain tucked into the bowl of the protein. When the PDZ domain binds / recognises the C-terminal peptide sequence of its substrate the PDZ domain undergoes a conformational change and swings into an out position, activating the protease domain. The PDZ domain movement pulls the substrate through the peptide tunnel past the catalytic site residues for degradation (unfolding in the process).

6.3.3 The role of Bd0967

We have demonstrated that Bd0967 targets and degrades flagellin proteins such as Bd0606 and that *Δbd0967 Bdellovibrio* mutants have a severe pleiotropic developmental phenotype involving flagellar assembly and cell development. Additionally Bd0967 has previously been identified as a component of *Bdellovibrio* flagellar (unpublished study carried out by the Socket lab, which analysed all the proteins associated with sheared and purified flagellar). Together our data suggests that Bd0967 is involved in the regulation/control of the levels of flagellins and consequently a novel signalling pathway responding to the levels of flagellin.

The developmental lifecycle of *Bdellovibrio* is complex and involves many simultaneously operating signalling systems to successfully progress through the lifecycle in a coordinated and timely manner. A prominent physiological feature of *Bdellovibrio* is the large single, polar flagellum that is required to actively seek out prey cells. The flagellum is shed upon invasion of a prey cell and then re-assembled during the development stage of *Bdellovibrios* lifecycle once progeny cells have divided [26]. Flagella shedding and disassembly is not a common physiological event in other organism because flagella assembly is an energetically costly process.

Flagella are assembled into a multi-protein complex spanning the cytoplasmic and outer membranes. The flagella filament is made up of thousands of flagellin subunits that are polymerised at the growing end (distal end) of the filament. In *Bdellovibrio* the flagellar is composed of six flagellins; FliC1-6 [28]. Each flagellin is highly similar, especially at the C-terminal end that is responsible for self-assembly to form the helical flagellar structure (Figure 125). In *Bdellovibrio* each flagellin contributes differently, in

localisation and abundance, which produces a tapered wave shaped flagellum. Bd0606 (FliC2) is only used in small amounts as the penultimate flagellin immediately before a large amount of FliC6. Bd3052 (FliC5) and Bd0604 (FliC1), also identified through out MS analysis, are used in much larger quantities to form the central section of the flagellum [28].

The flagellum of *Bdellovibrio* also contains a membranous sheath coating. Sheathed flagella are associated with many marine Vibrio species and also gastric pathogens and is believed to play roles in virulence and pathogenesis [300]. The sheath has been associated with reducing immune responses caused by the shedding of pro-inflammatory agents (flagellins) [301]. The role of the sheath has also been proposed to aid the assembly of the filament by retaining the filament monomers in its structure [302].

How *Bdellovibrio* sheds and then re-assembles its flagellar has long been a mystery. We therefore propose that Bd0967 is integral to this complex developmental process and our data suggest it has a role in regulating the levels of flagellins, which may be a developmental cue. The *Δbd0967* mutant lacks this control mechanism, which is why it has such a detrimental pleiotropic developmental phenotype. The exact biological role of Bd0967 and its activity towards Bd0606 is hard to discern. One possibility is, post invasion, once the flagellum has been shed, flagellin continue to be exported through the flagellar base. The flagellins will ultimately be retained by the sheath but not assembled due to the lack of growing filament end. As a result the flagellins will start to accumulate in the periplasm of *Bdellovibrio*. Therefore Bd0967s role would be to target

and degrade the miss-localised flagellin. On the other hand Bd0967 may actively participate in the flagellum shedding process by degrading polymerised flagellin.

PDZ containing proteases have been widely reported to have developmental phenotypes and play important roles in cellular signalling and biological processes. Interestingly there is published evidence for a link between CTPs and flagellins. For example comparison between the ΔPrc (Prc knockout) and Wt cells from *Xanthomonas oryzae pv. Oryzae* revealed that Prc played a role in virulence and also robustness to environmental stresses. The study revealed several proteins were more abundant between ΔPrc and Wt cells, including a flagellin, which had already been shown experimentally to play a role in virulence [303]. Furthermore, another study examined the effects of a CtpA knockout mutant in *Borrelia burgdorferii,* which resulted in developmental pleiotropic effects on *Borrelia* protein synthesis [304]. Finally the periplasmic protease, AlgW from *P. syringae pv. Maculicola* has been identified as a key negative regulator of flagellin abundance [305].

6.4 Conclusion

In this study we have solved the three-dimensional structure of Bd0967, a CTP from *Bdellovibrio bacteriovorus* with a distinct structure forming a large self-compartmentalised cavity. Bd0967 possess two unique structural features: a disulphide bond and a metal-binding site that provides additional structural support. We also demonstrate Bd0967 is involved in a developmental pathway that is interconnected with flagellin assembly and show Bd0967 targets and processively degrades the flagellin Bd0606, which is mediated by interactions with the PDZ domain.

Chapter 7

Summary and Future Work

7. 1 Summary

The overall aim for this thesis was to better understand the proteins that are involved in key lifecycle events of the predatory bacterium *Bdellovibrio bacteriovorus.* The work undertaken for this thesis encompasses a diverse set of proteins that operate at distinct stages of the predatory lifecycle. In general, this work has improved our understanding of three major predatory lifecycle events: growth & metabolism, development & signalling and finally bdelloplast lysis & exit. As a result of the work carried out for this thesis there is now a greater understanding, structurally and functionally, of each protein that was investigated (Bd2924, Bd1833, Bd0967 and Bd0314). Figure 125 summarises the findings of this thesis and illustrates the stage in the predatory lifecycle each protein functions.

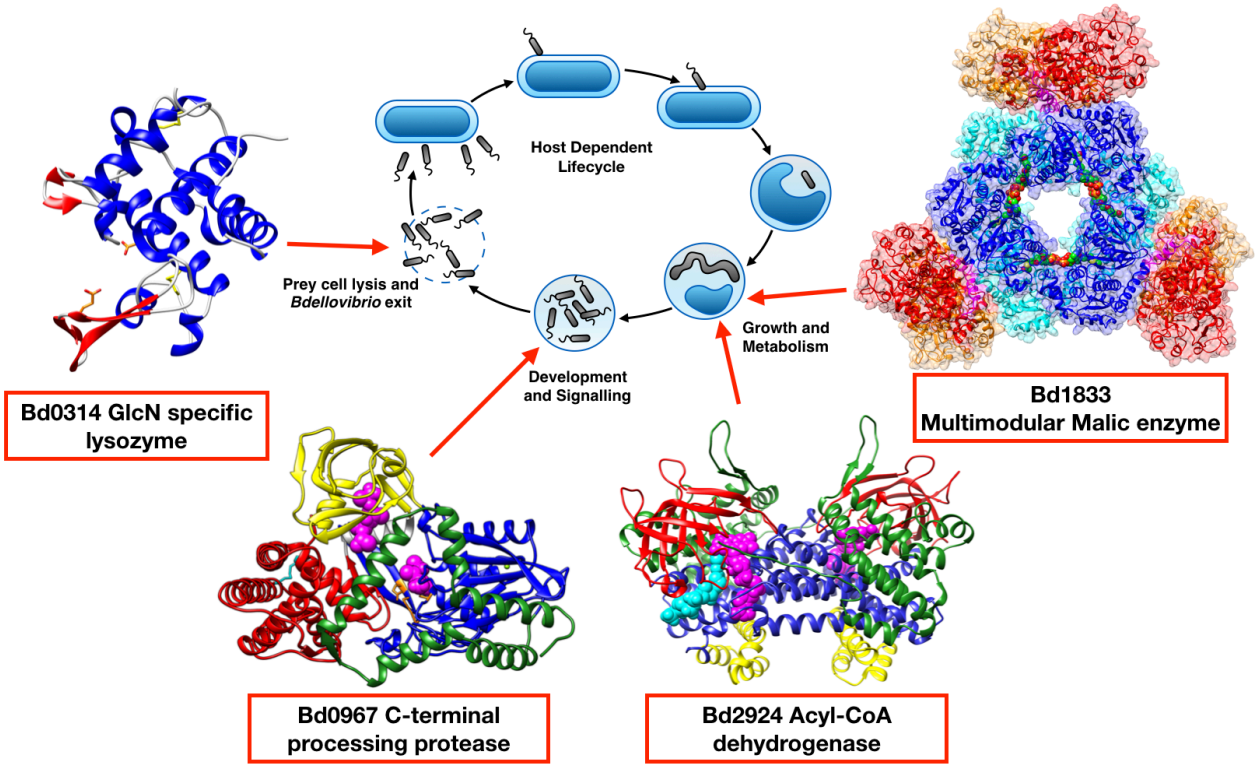


Figure 127. Summary of thesis findings
This thesis has structurally and functionally investigated four proteins with diverse roles within the predatory lifecycle. Bd1833 is a multi-modular malic enzyme involved in an alternative metabolic pathway, diverting TCA cycle intermediates and switching to a carbon storage pathway that is allosterically inhibited by acetyl-CoA. Bd2924 is a divergent putative acyl-CoA dehydrogenase potentially involved in metabolism of prey cell lipids. Bd0967 is a C-terminal processing protease involved in a critical cell developmental pathway involving flagellin degradation. Bd0314 is a modified lysozyme that has been specifically adapted to exit the invaded prey cell, acting specifically towards N-deacetylated peptidoglycan cell walls.

Two of the proteins investigated in this study (Bd2924 and Bd1833) are enzymes that have roles in central metabolism. Central metabolism has been an understudied area of research in the predatory lifecycle of *Bdellovibrio*. Many of the metabolic pathways have been adapted to deal with the needs of a predator's diet, such as solely using the mevalonate pathway for isoprenoid biosynthesis [306]. Consequently the findings presented in this thesis represent the first structurally and functionally characterised central metabolic enzymes from *Bdellovibrio*. We reveal Bd2924s three-dimensional structure is related to ACADs with divergent functions, suggesting that Bd2924 has a

role beyond fatty acid β-oxidation. The structure of Bd1833 represents the first structure of a multi-modular ME. Multi-modular MEs have been shown to be functionally important for subverting the flow of TCA cycle intermediates into carbon storage pathways such as gluconeogenesis and lipid biosynthesis. Carbon/energy storage, possibly into the observed cellular granules, is required to survive between kills when *Bdellovibrio* enters a starvation period [307]. Multi-modular MEs have been less well-characterised than the single component MEs found in eukaryotes; however, previous research had suggested acetyl-CoA was an effector of multi-modular ME activity. We have further developed this finding by uncovering the mechanism behind the allosteric regulation by acetyl-CoA. The structure of Bd1833 reveals that the non-catalytic Pta domain is central to hexamer complex formation and binding of acetyl-CoA. We show that R535 acts to sense acetyl-CoA via interactions with the 3'-phosphate group; this sensing results in a conformational change communicated using a 17aa loop specific to the ME domain. The binding of acetyl-CoA acts to close the ME active-site cleft, acting as a negative regulator of enzyme activity. Our work provides a structural basis for understanding allosteric regulation of multi-modular MEs, which will assist in analysis of their roles in central metabolism.

This study also uncovers an important role in development and signalling for Bd0967, which is linked to its ability to target flagellins such as Bd0606 for destruction. Furthermore, we reveal the structure of Bd0967, a CTP that forms a self-compartmentalised cavity. The structure of Bd0967 is divergent from other published CTPs and contains two unique features: a disulphide bond and a metal-binding site. The structure of Bd0967 provides even more diversity to the already extremely diverse family of CTPs that have been characterised thus far.

Finally, we present Bd0314 as an adapted lysozyme that has been repurposed to facilitate bdelloplast lysis and cell exit at the end of predatory lifecycle. Our work demonstrates that Bd0314 is specific towards *N*-deacetylated PG substrates and has a modified active site, which prevents GlcNAc residues being accommodated. This work builds on previous investigations into *Bdellovibrio bacteriovorus* adaptation of the prey cell wall, to form an intraperiplasmic niche and marking "prey" from "self" by *N*-deacetylating the prey cell wall. We demonstrate *Bdellovibrio* uses Bd0314 to specifically target the modified prey PG at the end of the predatory lifecycle.

7.2 Future Work

Understanding the interesting and unusual predatory lifecycle of *Bdellovibrio* is a huge task. The work presented in this thesis has developed our understanding of the proteins involved in key predatory lifecycle stages; however, it has also provided many new avenues to be explored for future work.

Bd1833 is structurally and mechanistically a very interesting protein, operating as a TCA cycle subverting enzyme providing a pivotal role in the diversion of intermediates into carbon storage pathways. It is not fully understood why Bd1833 retains such a strong affinity for acetyl-CoA and CoA if not required for other multi-modular ME. Further mutational studies of acetyl-CoA interacting residues such as exchanging Q683 to a His, whilst simultaneously exchanging R535 for a hydrophobic residue such as V or L may result in only inhibition by acetyl-CoA as is the case for other multi-modular ME. Furthermore, HPLC results revealed the ratio of co-purified CoA and acetyl-CoA differed between full-length Bd1833 and Bd1833-Pta. It is not fully understood why the

ratio of co-purified ligands is affected by the truncation, but conformational sampling of the ME domain (multiple open forms observed in eukaryotic ME) is suspected, which would affect acetyl-CoA binding in the Pta domain. It would therefore be interesting to evaluate whether Bd1833 catalysis affects ligand co-purification, whereby the ME domain can feedback to the Pta domain causing a conformational change, which reduces the affinity for acetyl-CoA.

Additionally our Bd1833-Pta:acetyl-CoA complex contradicts the current catalytic mechanism proposed for Pta enzymes [247], yet supports the earlier biochemical and mutational Pta studies [248]. It would be interesting to mutate the residues surrounding the acetyl-group binding pocket in catalytically active Pta enzymes to determine the true catalytic mechanism. For example, mutations of S128(Y576) or D307(V746) (*M. thermophila* Pta vs Bd1833 numbering) have not been investigated in the existing Pta literature.

Our work demonstrates Bd0314 possesses the ability to specifically degrade GlcNAc-deacetylated PG. Although this work has uncovered the mechanism behind specificity (Y228 steric clash with the acetyl group of GlcNAc residues) it does not determine how Bd0314 binds GlcNAc-deacetylated PG. The free amine group of GlcN residues is presumed to make interactions with the hydroxyl group of Y228 and potentially with nearby serine residues, similarly to that observed for Chitosanses. A Bd0314:ligand complex for Bd0314 would provide detailed evidence of protein and ligand interactions, rather than relying on those documented for “regular” lysozymes and chitosanases.

Although our results suggest Bd0314 is responsible for aiding prey cell lysis at the end of the predatory lifecycle, we also reveal *Bdellovibrio* has two other deacetylation specific lysozymes, Bd1413 and Bd1440. As a result we were able to present a "deacetylation specific lysozyme motif" and suggest deacetylation specific mutants may exist in other organisms such as *Caulobacter* species. It would be interesting to evaluate the activity of other potential "deacetylation specific" lysozymes using our FITC-PG assay and confirm whether "deacetylation specific" lysozymes are a widely distributed class of enzymes.

Finally we revealed that Bd0967 is a CTP that processively degrades the flagellin Bd0606. It would be interesting to determine whether Bd0967 can degrade any of the other *Bdellovibrio* flagellins (all possess a highly similar C-terminal sequence). Furthermore, we have not tested whether Bd0967 is able to degrade polymerised flagellins. We could test this by sheering the flagellum from *Bdellovibrio* and incubating with Bd0967 to determine whether the flagellum is depolymerised.

Mechanistically it would be interesting to obtain a structure of Bd0967 in the 'resting' state documented for CtpB. This would help us understand how proteolysis is gated in CTPs. Presently only CtpB has been observed in the 'resting' and 'active' conformations. It would also be valuable to determine the specificity at the protease cleavage site by using substrate peptide libraries. Furthermore, we could also use peptide libraries to probe the specificity of the PDZ domain and see if the results concur with the C-terminus of Bd0606.

Chapter 8

References

1. Alvarez, C.E., et al., The crystal structure of the malic enzyme from Candidatus Phytoplasma reveals the minimal structural determinants for a malic enzyme. Acta Crystallographica Section D-Structural Biology, 2018. **74**: p. 332-340.
2. Perez, J., et al., Bacterial predation: 75 years and counting! Environmental Microbiology, 2016. **18**(3): p. 766-779.
3. Velicer, G.J. and H. Mendes-Soares, Bacterial predators. Current Biology, 2009. **19**(2): p. R55-R56.
4. Berleman, J.E. and J.R. Kirby, Deciphering the hunting strategy of a bacterial wolfpack. Fems Microbiology Reviews, 2009. **33**(5): p. 942-957.
5. Munoz-Dorado, J., et al., Myxobacteria: Moving, Killing, Feeding, and Surviving Together. Frontiers in Microbiology, 2016. **7**.
6. Konovalova, A., T. Petters, and L. Sogaard-Andersen, Extracellular biology of Myxococcus xanthus. Fems Microbiology Reviews, 2010. **34**(2): p. 89-106.
7. Pasternak, Z., et al., In and out: an analysis of epibiotic vs periplasmic bacterial predators. Isme Journal, 2014. **8**(3): p. 625-635.
8. Wang, Z., D.E. Kadouri, and M. Wu, Genomic insights into an obligate epibiotic bacterial predator: Micavibrio aeruginosavorus ARL-13. Bmc Genomics, 2011. **12**.
9. Guerrero, R., et al., Predatory prokaryotes - predation and primary consumption evolved in bacteria. Proceedings of the National Academy of Sciences of the United States of America, 1986. **83**(7): p. 2138-2142.
10. Davidov, Y. 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 Bacteriovoracaceae fam. Nov. International Journal of Systematic and Evolutionary Microbiology, 2004. **54**: p. 1439-1452.
11. Sockett, R.E. and C. Lambert, Bdellovibrio as therapeutic agents: a predatory renaissance? Nature Reviews Microbiology, 2004. **2**(8): p. 669-675.
12. Atterbury, R.J., et al., Effects of Orally Administered Bdellovibrio bacteriovorus on the Well-Being and Salmonella Colonization of Young Chicks. Applied and Environmental Microbiology, 2011. **77**(16): p. 5794-5803.
13. Shatzkes, K., et al., Predatory Bacteria Attenuate Klebsiella pneumoniae Burden in Rat Lungs. Mbio, 2016. **7**(6): p. 9.
14. Willis, A.R., et al., Injections of Predatory Bacteria Work Alongside Host Immune Cells to Treat Shigella Infection in Zebrafish Larvae. Current Biology, 2016. **26**(24): p. 3343-3351.
15. Dwidar, M., A.K. Monnappa, and R.J. Mitchell, The dual probiotic and antibiotic nature of Bdellovibrio bacteriovorus. Bmb Reports, 2012. **45**(2): p. 71-78.
16. Monnappa, A.K., et al., Bdellovibrio bacteriovorus inhibits Staphylococcus aureus biofilm formation and invasion into human epithelial cells. Sci Rep, 2014. **4**: p. 3811.

17. Dashiff, A., et al., Predation of human pathogens by the predatory bacteria Micavibrio aeruginosavorus and Bdellovibrio bacteriovorus. Journal of Applied Microbiology, 2011. **110**(2): p. 431-444.
18. Kadouri, D. And G.A. O'Toole, Susceptibility of biofilms to Bdellovibrio bacteriovorus attack. Applied and Environmental Microbiology, 2005. **71**(7): p. 4044-4051.
19. Qi, Z.Z., et al., Probiotics in aquaculture of China - Current state, problems and prospect. Aquaculture, 2009. **290**(1-2): p. 15-21.
20. Saxon, E.B., et al., Bdellovibrio bacteriovorus HD100 guards against Pseudomonas tolaasii brown-blotch lesions on the surface of post-harvest Agaricus bisporus supermarket mushrooms. Bmc Microbiology, 2014. **14**.
21. Song, W.Y., Identification and characterization of Bdellovibrio bacteriovorus, a predator of Burkholderia glumae. Journal of Microbiology and Biotechnology, 2004. **14**(1): p. 48-55.
22. Scherff, R.H., Control of bacterial blight of soybean by Bdellovibrio-bacteriovorus. Phytopathology, 1973. **63**(3): p. 400-402.
23. Stolp, H. And M.P. Starr, Bdellovibrio bacteriovorus gen. et sp. n., a predatory, ectoparasitic, and bacteriolytic microorganism. Antonie Van Leeuwenhoek Journal of Microbiology and Serology, 1963. **29**(3): p. 217.
24. Rendulic, S., et al., A predator unmasked: Life cycle of Bdellovibrio bacteriovorus from a genomic perspective. Science, 2004. **303**(5658): p. 689-692.
25. Lambert, C., et al., Bdellovibrio: growth and development during the predatory cycle. Current Opinion in Microbiology, 2006. **9**(6): p. 639-644.
26. Lambert, C., et al., Characterizing the flagellar filament and the role of motility in bacterial prey-penetration by Bdellovibrio bacteriovorus. Mol Microbiol, 2006. **60**(2): p. 274-86.
27. Thomashow, L.S. and S.C. Rittenberg, Waveform analysis and structure of flagella and basal complexes from Bdellovibrio-bacteriovorus-109J. Journal of Bacteriology, 1985. **163**(3): p. 1038-1046.
28. Iida, Y., et al., Roles of multiple flagellins in flagellar formation and flagellar growth post bdelloplast lysis in Bdellovibrio bacteriovorus. J Mol Biol, 2009. **394**(5): p. 1011-21.
29. Morehouse, K.A., et al., Three motab Stator Gene Products in Bdellovibrio bacteriovorus Contribute to Motility of a Single Flagellum during Predatory and Prey-Independent Growth. Journal of Bacteriology, 2011. **193**(4): p. 932-943.
30. Lambert, C., et al., Predatory Bdellovibrio Bacteria Use Gliding Motility To Scout for Prey on Surfaces. Journal of Bacteriology, 2011. **193**(12): p. 3139-3141.
31. Mauriello, E.M.F., et al., Gliding Motility Revisited: How Do the Myxobacteria Move without Flagella? Microbiology and Molecular Biology Reviews, 2010. **74**(2): p. 229.
32. Milner, D.S., et al., Ras gtpase-Like Protein mgla, a Controller of Bacterial Social-Motility in Myxobacteria, Has Evolved to Control Bacterial Predation by Bdellovibrio. Plos Genetics, 2014. **10**(4): p. 13.
33. Lamarre, A.G., S.C. Straley, and S.F. Conti, Chemotaxis toward amino-acids by Bdellovibrio-bacteriovorus. Journal of Bacteriology, 1977. **131**(1): p. 201-207.
34. Straley, S.C. and S.F. Conti, Chemotaxis by Bdellovibrio-bacteriovorus toward prey. Journal of Bacteriology, 1977. **132**(2): p. 628-640.
35. Straley, S.C., et al., Chemotaxis of Bdellovibrio-bacteriovorus toward pure compounds. Journal of Bacteriology, 1979. **140**(2): p. 634-642.

36. Lambert, C., 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. Environmental Microbiology, 2003. **5**(2): p. 127-132.
37. Rogosky, A.M., P.L. Moak, and E.A.B. Emmert, Differential predation by Bdellovibrio bacteriovorus 109J. Current Microbiology, 2006. **52**(2): p. 81-85.
38. Varon, M. And M. Shilo, Interaction of bdellovibrio bacteriovorus and host bacteria .i. kinetic studies of attachment and invasion of Escherichia coli B by Bdellovibrio bacteriovorus. Journal of Bacteriology, 1968. **95**(3): p. 744.
39. Varon, M. And M. Shilo, Attachment of Bdellovibrio bacteriovorus to cell wall mutants of Salmonella spp and Escherichia coli. Journal of Bacteriology, 1969. **97**(2): p. 977.
40. Burnham, J.C., T. Hashimoto, and S.F. Conti, Electron microscopic observations on penetration of Bdellovibrio bacteriovorus into gram-negative bacterial hosts. Journal of Bacteriology, 1968. **96**(4): p. 1366.
41. Evans, K.J., C. Lambert, and R.E. Sockett, Predation by Bdellovibrio bacteriovorus HD100 requires type IV pili. Journal of Bacteriology, 2007. **189**(13): p. 4850-4859.
42. Mahmoud, K.K. and S.F. Koval, Characterization of type IV pili in the life cycle of the predator bacterium Bdellovibrio. Microbiology-Sgm, 2010. **156**: p. 1040-1051.
43. Capeness, M.J., 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, 2013. **8**(11): p. 22.
44. Abram, D., Castroem.J, and D. Chou, Penetration of Bdellovibrio-bacteriovorus into host-cells. Journal of Bacteriology, 1974. **118**(2): p. 663-680.
45. Thomashow, M.F. and S.C. Rittenberg, Intraperiplasmic growth of Bdellovibrio-bacteriovorus 109J - solubilization of Escherichia-coli peptidoglycan. Journal of Bacteriology, 1978. **135**(3): p. 998-1007.
46. Kuru, E., et al., Fluorescent D-amino-acids reveal bi-cellular cell wall modifications important for Bdellovibrio bacteriovorus predation. Nature Microbiology, 2017. **2**(12): p. 1648-1657.
47. Lerner, T.R., et al., Specialized peptidoglycan hydrolases sculpt the intra-bacterial niche of predatory Bdellovibrio and increase population fitness. Plos Pathog, 2012. **8**(2): p. E1002524.
48. Lambert, C., et al., Ankyrin-mediated self-protection during cell invasion by the bacterial predator Bdellovibrio bacteriovorus. Nature Communications, 2016. **7**: p. 1.
49. Lambert, C., et al., Interrupting peptidoglycan deacetylation during Bdellovibrio predator-prey interaction prevents ultimate destruction of prey wall, liberating bacterial-ghosts. Scientific Reports, 2016. **6**: p. 19.
50. Thomashow, M.F. and S.C. Rittenberg, Intraperiplasmic growth of Bdellovibrio-bacteriovorus 109J - N-deacetylation of Escherichia-coli peptidoglycan amino-sugars. Journal of Bacteriology, 1978. **135**(3): p. 1008-1014.
51. Ruby, E.G. and S.C. Rittenberg, attachment of diaminopimelic acid to bdelloplast peptidoglycan during intraperiplasmic growth of Bdellovibrio-bacteriovorus 109J. Journal of Bacteriology, 1984. **158**(2): p. 597-602.

52. Thomashow, M.F. and S.C. Rittenberg, Intraperiplasmic growth of Bdellovibrio-bacteriovorus 109J - attachment of long-chain fatty-acids to Escherichia-coli peptidoglycan. Journal of Bacteriology, 1978. **135**(3): p. 1015-1023.
53. Rittenberg, S.C. and M. Shilo, Early host damage in infection cycle of Bdellovibrio-bacteriovorus. Journal of Bacteriology, 1970. **102**(1): p. 149-160.
54. Matin, A., Kinetics of deoxyribonucleic acid destruction and synthesis during growth of Bdellovibrio-bacteriovorus on Pseudomonas-putida and Escherichia-coli. Antonie Van Leeuwenhoek Journal of Microbiology, 1972. **38**(4): p. 630-630.
55. Cover, W.H., R.J. Martinez, and S.C. Rittenberg, Permeability of the boundary-layers of Bdellovibrio-bacteriovorus-109J and its bdelloplasts to small hydrophilic molecules. Journal of Bacteriology, 1984. **157**(2): p. 385-390.
56. Rittenberg, S.C. and R.B. Hespell, Energy efficiency of intraperiplasmic growth of Bdellovibrio-bacteriovorus. Journal of Bacteriology, 1975. **121**(3): p. 1158-1165.
57. Rosson, R.A. and S.C. Rittenberg, Regulated breakdown of Escherichia-coli deoxyribonucleic-acid during intraperiplasmic growth of Bdellovibrio-bacteriovorus-109J. Journal of Bacteriology, 1979. **140**(2): p. 620-633.
58. Hespell, R.B. and D.A. Odelson, Metabolism of rna-ribose by Bdellovibrio-bacteriovorus during intraperiplasmic growth on Escherichia-coli. Journal of Bacteriology, 1978. **136**(3): p. 936-946.
59. Kuenen, J.G. and S.C. Rittenberg, Incorporation of long-chain fatty-acids of substrate organism by Bdellovibrio-bacteriovorus during intraperiplasmic growth. Journal of Bacteriology, 1975. **121**(3): p. 1145-1157.
60. Stein, M.A., et al., Acquisition of apparently intact and unmodified lipopolysaccharides from Escherichia-coli by Bdellovibrio-bacteriovorus. Journal of Bacteriology, 1992. **174**(9): p. 2858-2864.
61. Nelson, D.R. and S.C. Rittenberg, Incorporation of substrate cell lipid-a components into the lipopolysaccharide of intraperiplasmically grown Bdellovibrio-bacteriovorus. Journal of Bacteriology, 1981. **147**(3): p. 860-868.
62. Andreev, L.V., et al., Peculiarities of the fatty-acid composition of Bdellovibrio. Folia Microbiologica, 1983. **28**(1): p. 28-35.
63. Schwudke, D., et al., The obligate predatory Bdellovibrio bacteriovorus possesses a neutral lipid A containing alpha-D-mannoses that replace phosphate residues - Similarities and differences between the lipid As and the lipopolysaccharides of the wild type strain B-bacteriovorus HD100 and its host-independent derivative HI100. Journal of Biological Chemistry, 2003. **278**(30): p. 27502-27512.
64. Diedrich, D.L., C.A. Portnoy, and S.F. Conti, Bdellovibrio possesses a prey-derived ompf protein in its outer-membrane. Current Microbiology, 1983. **8**(1): p. 51-56.
65. Talley, B.G., R.L. mcdade, and D.L. Diedrich, Verification of the protein in the outer-membrane of Bdellovibrio-bacteriovorus as the ompf protein of its escherichia-coli prey. Journal of Bacteriology, 1987. **169**(2): p. 694-698.
66. Barel, G., et al., Fate of predator and prey proteins during growth of Bdellovibrio bacteriovorus on Escherichia coli and Pseudomonas syringae prey. Journal of Bacteriology, 2005. **187**(1): p. 329-335.
67. Beck, S., et al., Bdellovibrio bacteriovorus strains produce a novel major outer membrane protein during predacious growth in the periplasm of prey bacteria. Journal of Bacteriology, 2004. **186**(9): p. 2766-2773.

68. Hespell, R.B., et al., Respiration of Bdellovibrio-bacteriovorus strain 109J and its energy substrates for intraperiplasmic growth. Journal of Bacteriology, 1973. **113**(3): p. 1280-1288.
69. Gadkari, D. And H. Stolp, Energy metabolism of Bdellovibrio-bacteriovorus .1. energy production, atp pool, energy charge. Archives of Microbiology, 1975. **102**(3): p. 179-185.
70. Gadkari, D. And H. Stolp, Energy-metabolism of Bdellovibrio-bacteriovorus .2. p-o ratio and atp pool turnover rate. Archives of Microbiology, 1976. **108**(1): p. 125-132.
71. Ruby, E.G. and J.B. mccabe, An atp transport-system in the intracellular bacterium, Bdellovibrio-bacteriovorus 109J. Journal of Bacteriology, 1986. **167**(3): p. 1066-1070.
72. Hespell, R.B., Glycolytic and tricarboxylic-acid cycle enzyme-activities during intraperiplasmic growth of Bdellovibrio-bacteriovorus on Escherichia-coli. Journal of Bacteriology, 1976. **128**(2): p. 677-680.
73. Simpson, F.J. and J. Robinson, Some energy-producing systems in Bdellovibrio bacteriovorus strain 6-5-S. Canadian Journal of Biochemistry, 1968. **46**(8): p. 865-873.
74. Fenton, A.K., et al., Shadowing the Actions of a Predator: Backlit Fluorescent Microscopy Reveals Synchronous Nonbinary Septation of Predatory Bdellovibrio inside Prey and Exit through Discrete Bdelloplast Pores. Journal of Bacteriology, 2010. **192**(24): p. 6329-6335.
75. Kessel, M. And M. Shilo, Relationship of Bdellovibrio elongation and fission to host-cell size. Journal of Bacteriology, 1976. **128**(1): p. 477-480.
76. Seidler, R.J. and M.P. Starr, Isolation and characterization of host-independent Bdellovibrios. Journal of Bacteriology, 1969. **100**(2): p. 769.
77. Cotter, T.W. and M.F. Thomashow, Identification of a Bdellovibrio-bacteriovorus genetic-locus, hit, associated with the host-independent phenotype. Journal of Bacteriology, 1992. **174**(19): p. 6018-6024.
78. Prehna, G., B.E. Ramirez, and A.L. Lovering, The Lifestyle Switch Protein Bd0108 of Bdellovibrio bacteriovorus Is an Intrinsically Disordered Protein. Plos One, 2014. **9**(12).
79. Hobley, L., et al., Discrete Cyclic di-GMP-Dependent Control of Bacterial Predation versus Axenic Growth in Bdellovibrio bacteriovorus. Plos Pathogens, 2012. **8**(2): p. 13.
80. Jenal, U., A. Reinders, and C. Lori, Cyclic di-GMP: second messenger extraordinaire. Nature Reviews Microbiology, 2017. **15**(5): p. 271-284.
81. Karunker, I., et al., A global transcriptional switch between the attack and growth forms of Bdellovibrio bacteriovorus. Plos One, 2013. **8**(4): p. E61850.
82. Rotem, O., et al., An Extended Cyclic Di-GMP Network in the Predatory Bacterium Bdellovibrio bacteriovorus. J Bacteriol, 2016. **198**(1): p. 127-37.
83. Paul, K., et al., The c-di-GMP Binding Protein ycgr Controls Flagellar Motor Direction and Speed to Affect Chemotaxis by a "Backstop Brake" Mechanism. Molecular Cell, 2010. **38**(1): p. 128-139.
84. Liao, D.I., et al., Crystal structures of the photosystem II D1 C-terminal processing protease. Nature Structural Biology, 2000. **7**(9): p. 749-753.
85. Mastny, M., et al., ctpb assembles a gated protease tunnel regulating cell-cell signaling during spore formation in Bacillus subtilis. Cell, 2013. **155**(3): p. 647-58.

86. Kohlmann, F., et al., Structural basis of the proteolytic and chaperone activity of Chlamydia trachomatis CT441. J Bacteriol, 2015. **197**(1): p. 211-8.
87. Su, M.Y., et al., Structural basis of adaptor-mediated protein degradation by the tail-specific PDZ-protease Prc. Nat Commun, 2017. **8**(1): p. 1516.
88. Van den Ent, F. And J. Lowe, RF cloning: A restriction-free method for inserting target genes into plasmids. Journal of Biochemical and Biophysical Methods, 2006. **67**(1): p. 67-74.
89. Page, R., et al., Shotgun crystallization strategy for structural genomics: an optimized two-tiered crystallization screen against the Thermotoga maritima proteome. Acta Crystallographica Section D-Biological Crystallography, 2003. **59**: p. 1028-1037.
90. Chaikuad, A., S. Knapp, and F. Von Delft, Defined PEG smears as an alternative approach to enhance the search for crystallization conditions and crystal-quality improvement in reduced screens. Acta Crystallographica Section D-Structural Biology, 2015. **71**: p. 1627-1639.
91. Newman, J., et al., Towards rationalization of crystallization screening for small- to medium-sized academic laboratories: the PACT/JCSG plus strategy. Acta Crystallographica Section D-Structural Biology, 2005. **61**: p. 1426-1431.
92. Radaev, S. And P.D. Sun, Crystallization of protein-protein complexes. Journal of Applied Crystallography, 2002. **35**: p. 674-676.
93. Grimm, C., et al., A crystallization screen based on alternative polymeric precipitants. Acta Crystallographica Section D-Biological Crystallography, 2010. **66**: p. 685-697.
94. Gorrec, F., The MORPHEUS protein crystallization screen. Journal of Applied Crystallography, 2009. **42**: p. 1035-1042.
95. Winter, G., xia2: an expert system for macromolecular crystallography data reduction. Journal of Applied Crystallography, 2010. **43**: p. 186-190.
96. Vonrhein, C., et al., Data processing and analysis with the autoproc toolbox. Acta Crystallographica Section D-Biological Crystallography, 2011. **67**: p. 293-302.
97. Weiss, M.S., Global indicators of X-ray data quality. Journal of Applied Crystallography, 2001. **34**: p. 130-135.
98. Weiss, M.S. and R. Hilgenfeld, On the use of the merging R factor as a quality indicator for X-ray data. Journal of Applied Crystallography, 1997. **30**: p. 203-205.
99. Diederichs, K. And P.A. Karplus, Improved R-factors for diffraction data analysis in macromolecular crystallography. Nature Structural Biology, 1997. **4**(4): p. 269-275.
100. Karplus, P.A. and K. Diederichs, Linking Crystallographic Model and Data Quality. Science, 2012. **336**(6084): p. 1030-1033.
101. Mccoy, A.J., et al., Phaser crystallographic software. Journal of Applied Crystallography, 2007. **40**: p. 658-674.
102. Emsley, P., et al., Features and development of Coot. Acta Crystallographica Section D-Biological Crystallography, 2010. **66**: p. 486-501.
103. Afonine, P.V., et al., Towards automated crystallographic structure refinement with phenix.refine. Acta Crystallographica Section D-Structural Biology, 2012. **68**: p. 352-367.
104. Joosten, R.P., et al., The PDB_REDO server for macromolecular structure model optimization. Iucrj, 2014. **1**: p. 213-220.

105. Chen, V.B., et al., molprobity: all-atom structure validation for macromolecular crystallography. Acta Crystallographica Section D-Structural Biology, 2010. **66**: p. 12-21.
106. Lovell, S.C., et al., Structure validation by C alpha geometry: phi,psi and C beta deviation. Proteins-Structure Function and Genetics, 2003. **50**(3): p. 437-450.
107. Kleywegt, G.J. and T.A. Jones, Model building and refinement practice. Macromolecular Crystallography, Pt B, 1997. **277**: p. 208-230.
108. Tropea, J.E., Expression and Purification of Soluble His6-Tagged TEV Protease. Methods in Molecular Biology: High Throughput Protein Expression and Purification, 2009. **498**.
109. Maeda, H., A new lysozyme assay based on fluorescence polarization or fluorescence intensity utilizing a fluorescent peptidoglycan substrate. Journal of Biochemistry, 1980. **88**(4): p. 1185-1191.
110. Adams, P.D., et al., PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallographica Section D-Biological Crystallography, 2010. **66**: p. 213-221.
111. Terwilliger, T.C., et al., Iterative model building, structure refinement and density modification with the PHENIX autobuild wizard. Acta Crystallographica Section D-Biological Crystallography, 2008. **64**: p. 61-69.
112. Shurubor, Y.I., et al., Determination of Coenzyme A and Acetyl-Coenzyme A in Biological Samples Using HPLC with UV Detection. Molecules, 2017. **22**(9).
113. Bunkoczi, G. And R.J. Read, Improvement of molecular-replacement models with Sculptor. Acta Crystallographica Section D-Biological Crystallography, 2011. **67**: p. 303-312.
114. Stewart, P.D.S., et al., Random Microseeding: A Theoretical and Practical Exploration of Seed Stability and Seeding Techniques for Successful Protein Crystallization. Crystal Growth & Design, 2011. **11**(8): p. 3432-3441.
115. Stein, N., CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. Journal of Applied Crystallography, 2008. **41**: p. 641-643.
116. Vollmer, W., D. Blanot, and M.A. de Pedro, Peptidoglycan structure and architecture. Fems Microbiology Reviews, 2008. **32**(2): p. 149-167.
117. Holtje, J.V., et al., Novel type of murein transglycosylase in Escherichia-coli. Journal of Bacteriology, 1975. **124**(3): p. 1067-1076.
118. Braun, V. And K. Rehn, Chemical characterization, spatial distribution and function of a lipoprotein (murein-lipoprotein) of e coli cell wall - specific effect of trypsin on membrane structure. European Journal of Biochemistry, 1969. **10**(3): p. 426-438.
119. Magnet, S., et al., Identification of the L,D-transpeptidases responsible for attachment of the Braun lipoprotein to Escherichia coli peptidoglycan. Journal of Bacteriology, 2007. **189**(10): p. 3927-3931.
120. Dramsi, S., et al., Covalent attachment of proteins to peptidoglycan. Fems Microbiology Reviews, 2008. **32**(2): p. 307-320.
121. Sauvage, E., et al., The penicillin-binding proteins: structure and role in peptidoglycan biosynthesis. Fems Microbiology Reviews, 2008. **32**(2): p. 234-258.
122. Matias, V.R.F., et al., Cryo-transmission electron Microscopy of frozen-hydrated sections of Escherichia coli and Pseudomonas aeruginosa. Journal of Bacteriology, 2003. **185**(20): p. 6112-6118.

123. Matias, V.R.F. and T.J. Beveridge, Cryo-electron microscopy reveals native polymeric cell wall structure in Bacillus subtilis 168 and the existence of a periplasmic space. Molecular Microbiology, 2005. **56**(1): p. 240-251.
124. Koch, A.L. and S. Woeste, ELASTICITY OF THE SACCULUS OF ESCHERICHIA-COLI. Journal of Bacteriology, 1992. **174**(14): p. 4811-4819.
125. Demchick, P. And A.L. Koch, The permeability of the wall fabric of Escherichia coli and Bacillus subtilis. Journal of Bacteriology, 1996. **178**(3): p. 768-773.
126. Boneca, I.G., et al., A critical role for peptidoglycan N-deacetylation in Listeria evasion from the host innate immune system. Proceedings of the National Academy of Sciences of the United States of America, 2007. **104**(3): p. 997-1002.
127. Hebert, L., et al., Enterococcus faecalis constitutes an unusual bacterial model in lysozyme resistance. Infection and Immunity, 2007. **75**(11): p. 5390-5398.
128. Zipperle, G.F., J.W. Ezzell, and R.J. Doyle, Glucosamine substitution and muramidase susceptibility in Bacillus-anthracis. Canadian Journal of Microbiology, 1984. **30**(5): p. 553-559.
129. Vollmer, W. And A. Tomasz, The pgda gene encodes for a peptidoglycan N-acetylglucosamine deacetylase in Streptococcus pneumoniae. Journal of Biological Chemistry, 2000. **275**(27): p. 20496-20501.
130. Coutinho, P.M. and B. Henrissat, Carbohydrate-active enzymes: An integrated database approach. Recent Advances in Carbohydrate Bioengineering, 1999(246): p. 3-12.
131. Vollmer, W. And A. Tomasz, Peptidoglycan N-acetylglucosamine deacetylase, a putative virulence factor in Streptococcus pneumoniae. Infection and Immunity, 2002. **70**(12): p. 7176-7178.
132. Vocadlo, D.J., et al., Catalysis by hen egg-white lysozyme proceeds via a covalent intermediate. Nature, 2001. **412**(6849): p. 835-838.
133. Lim, C., et al., Contact time- and ph-dependent adhesion and cohesion of low molecular weight chitosan coated surfaces. Carbohydrate Polymers, 2015. **117**: p. 887-894.
134. Vollmer, W., Structural variation in the glycan strands of bacterial peptidoglycan. Fems Microbiology Reviews, 2008. **32**(2): p. 287-306.
135. Abrams, A., O-acetyl groups in the cell wall of Streptococcus faecalis. Journal of Biological Chemistry, 1958. **230**(2): p. 949-959.
136. Brumfitt, W., A.C. Wardlaw, and J.T. Park, Development of lysozyme-resistance in Micrococcus-lysodiekticus and its association with an increased O-acetyl content of the cell wall. Nature, 1958. **181**(4626): p. 1783-1784.
137. Bera, A., et al., The presence of peptidoglycan O-acetyltransferase in various staphylococcal species correlates with lysozyme resistance and pathogenicity. Infection and Immunity, 2006. **74**(8): p. 4598-4604.
138. Pfeffer, J.M., et al., Peptidoglycan O acetylation and autolysin profile of Enterococcus faecalis in the viable but nonculturable state. Journal of Bacteriology, 2006. **188**(3): p. 902-908.
139. Szmiett, D.C., Role of peptidoglycan modifications in predation by Bdellovibrio bacteriovorus 109J. Western University Electronic Thesis and Dissertation Repository, 2015. **2655**.
140. Bera, A., et al., Why are pathogenic staphylococci so lysozyme resistant? The peptidoglycan O-acetyltransferase oata is the major determinant for lysozyme resistance of Staphylococcus aureus. Molecular Microbiology, 2005. **55**(3): p. 778-787.

141. Monzingo, A.F., et al., Chitinases, chitosanases, and lysozymes can be divided into procaryotic and eucaryotic families sharing a conserved core. Nature Structural Biology, 1996. **3**(2): p. 133-140.
142. Ragland, S.A. and A.K. Criss, From bacterial killing to immune modulation: Recent insights into the functions of lysozyme. Plos Pathogens, 2017. **13**(9): p. 22.
143. Koraimann, G., Lytic transglycosylases in macromolecular transport systems of Gram-negative bacteria. Cellular and Molecular Life Sciences, 2003. **60**(11): p. 2371-2388.
144. Wohlkonig, A., et al., Structural Relationships in the Lysozyme Superfamily: Significant Evidence for Glycoside Hydrolase Signature Motifs. Plos One, 2010. **5**(11): p. 10.
145. Marcotte, E.M., et al., X-ray structure of an anti-fungal chitosanase from Streptomyces N174. Nature Structural Biology, 1996. **3**(2): p. 155-162.
146. Blake, C.C.F., et al., Structure of hen egg-white lysozyme - a 3-dimensional fourier synthesis at 2A resolution. Nature, 1965. **206**(4986): p. 757-761.
147. Mhlongo, N.N., et al., A critical survey of average distances between catalytic carboxyl groups in glycoside hydrolases. Proteins-Structure Function and Bioinformatics, 2014. **82**(9): p. 1747-1755.
148. Hart, P.J., et al., Crystal-structure of an endochitinase from Hordeum-vulgare l seeds. Journal of Molecular Biology, 1993. **229**(1): p. 189-193.
149. Matthews, B.W., et al., Common precursor of lysozymes of hen egg-white and bacteriophage-T4. Nature, 1981. **290**(5804): p. 334-335.
150. Strynadka, N.C.J. and M.N.G. James, Lysozyme revisited - crystallographic evidence for distortion of an N-acetylmuramic acid residue bound in site-D. Journal of Molecular Biology, 1991. **220**(2): p. 401-424.
151. Kuroki, R., L.H. Weaver, and B.W. Matthews, Structural basis of the conversion of T4 lysozyme into a transglycosidase by reengineering the active site. Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(16): p. 8949-8954.
152. Weaver, L.H., M.G. Grutter, and B.W. Matthews, The refined structures of goose lysozyme and its complex with a bound trisaccharide show that the goose-type lysozymes lack a catalytic aspartate residue. Journal of Molecular Biology, 1995. **245**(1): p. 54-68.
153. Helland, R., et al., Crystal structures of g-type lysozyme from Atlantic cod shed new light on substrate binding and the catalytic mechanism. Cellular and Molecular Life Sciences, 2009. **66**(15): p. 2585-2598.
154. Scheurwater, E., C.W. Reid, and A.J. Clarke, Lytic transglycosylases: Bacterial space-making autolysins. International Journal of Biochemistry & Cell Biology, 2008. **40**(4): p. 586-591.
155. Thunnissen, A., et al., Structure of the 70-kDa soluble lytic transglycosylase complexed with bulgecin-A - implications for the enzymatic mechanism. Biochemistry, 1995. **34**(39): p. 12729-12737.
156. Blackburn, N.T. and A.J. Clarke, Identification of four families of peptidoglycan lytic transglycosylases. Journal of Molecular Evolution, 2001. **52**(1): p. 78-84.
157. Fibriansah, G., F.I. Gliubich, and A. Thunnissen, On the Mechanism of Peptidoglycan Binding and Cleavage by the endo-Specific Lytic Transglycosylase mlte from Escherichia coli. Biochemistry, 2012. **51**(45): p. 9164-9177.

158. Thunnissen, A., N.W. Isaacs, and B.W. Dijkstra, The catalytic domain of a bacterial lytic transglycosylase defines a novel class of lysozymes. Proteins-Structure Function and Bioinformatics, 1995. **22**(3): p. 245-258.
159. Little, D.J., et al., The Structure- and Metal-dependent Activity of Escherichia coli pgab Provides Insight into the Partial De-N-acetylation of Poly-beta-1,6-N-acetyl-D-glucosamine. Journal of Biological Chemistry, 2012. **287**(37): p. 31126-31137.
160. Holm, L. And L.M. Laakso, Dali server update. Nucleic Acids Research, 2016. **44**(W1): p. W351-W355.
161. Robert, X. And P. Gouet, Deciphering key features in protein structures with the new endscript server. Nucleic Acids Research, 2014. **42**(W1): p. W320-W324.
162. Song, H.W., et al., Structural-changes of active-site cleft and different saccharide binding modes in human lysozyme co-crystallized with hexa-N-acetyl-chitohexaose at pH 4.0. Journal of Molecular Biology, 1994. **244**(5): p. 522-540.
163. Arimori, T., et al., Crystal Structures of the Catalytic Domain of a Novel Glycohydrolase Family 23 Chitinase from Ralstonia sp A-471 Reveals a Unique Arrangement of the Catalytic Residues for Inverting Chitin Hydrolysis. Journal of Biological Chemistry, 2013. **288**(26): p. 18696-18706.
164. Saito, J., et al., Crystal structure of chitosanase from Bacillus circulans MH-K1 at 1.6-angstrom resolution and its substrate recognition mechanism. Journal of Biological Chemistry, 1999. **274**(43): p. 30818-30825.
165. Lyu, Q.Q., et al., Structural insights into the substrate-binding mechanism for a novel chitosanase. Biochemical Journal, 2014. **461**: p. 335-345.
166. Yorinaga, Y., et al., Crystal structure of a family 80 chitosanase from Mitsuaria chitosanitabida. Febs Letters, 2017. **591**(3): p. 540-547.
167. Goto, T., et al., Crystal structure of tapes japonica lysozyme with substrate analogue - Structural basis of the catalytic mechanism and manifestation of its chitinase activity accompanied by quaternary structural change. Journal of Biological Chemistry, 2007. **282**(37): p. 27459-27467.
168. Garcia-Casado, G., et al., Site-directed mutagenesis of active site residues in a class I endochitinase from chestnut seeds. Glycobiology, 1998. **8**(10): p. 1021-1028.
169. Hahn, M., et al., Structure of jack bean chitinase. Acta Crystallographica Section D-Biological Crystallography, 2000. **56**: p. 1096-1099.
170. Huet, J., et al., X-ray structure of papaya chitinase reveals the substrate binding mode of glycosyl hydrolase family 19 chitinases. Biochemistry, 2008. **47**(32): p. 8283-8291.
171. Ruby, E.G. and S.C. Rittenberg, Differentiation after premature release of intraperiplasmically growing Bdellovibrio-bacteriovorous. Journal of Bacteriology, 1983. **154**(1): p. 32-40.
172. Dominguez-Gil, T., et al., Activation by Allostery in Cell-Wall Remodeling by a Modular Membrane-Bound lytictransglycosylase from Pseudomonas aeruginosa. Structure, 2016. **24**(10): p. 1729-1741.
173. Wan, Z., et al., The adhesive and cohesive properties of a bacterial polysaccharide adhesin are modulated by a deacetylase. Molecular Microbiology, 2013. **88**(3): p. 486-500.
174. Berne, C., et al., Physiochemical Properties of Caulobacter crescentus Holdfast: A Localized Bacterial Adhesive. Journal of Physical Chemistry B, 2013. **117**(36): p. 10492-10503.

175. Hiltunen, J.K. and Y.M. Qin, beta-Oxidation - strategies for the metabolism of a wide variety of acyl-coa esters. Biochimica Et Biophysica Acta-Molecular and Cell Biology of Lipids, 2000. **1484**(2-3): p. 117-128.
176. Mizugaki, M., et al., Studies on the metabolism of unsaturated fatty-acids .12. reaction catalyzed by 2,4-dienoyl-coa reductase of Escherichia-coli. Journal of Biochemistry, 1983. **94**(2): p. 409-413.
177. Fujita, Y., H. Matsuoka, and K. Hirooka, Regulation of fatty acid metabolism in bacteria. Molecular Microbiology, 2007. **66**(4): p. 829-839.
178. Binstock, J.F., A. Pramanik, and H. Schulz, Isolation of a multienzyme complex of fatty-acid oxidation from Escherichia-coli. Proceedings of the National Academy of Sciences of the United States of America, 1977. **74**(2): p. 492-495.
179. Henry, M.F. and J.E. Cronan, Escherichia-coli transcription factor that both activates fatty-acid synthesis and represses fatty-acid degradation. Journal of Molecular Biology, 1991. **222**(4): p. 843-849.
180. Kim, J.J.P. and R. Miura, Acyl-coa dehydrogenases and acyl-coa oxidases - Structural basis for mechanistic similarities and differences. European Journal of Biochemistry, 2004. **271**(3): p. 483-493.
181. Mcandrew, R.P., et al., Structural basis for substrate fatty acyl chain specificity - Crystal structure of human very-long-chain acyl-coa dehydrogenase. Journal of Biological Chemistry, 2008. **283**(14): p. 9435-9443.
182. Lee, H.J., et al., Crystal structures of the wild type and the Glu376Gly/Thr255Glu mutant of human medium-chain acyl-coa dehydrogenase: Influence of the location of the catalytic base on substrate specificity. Biochemistry, 1996. **35**(38): p. 12412-12420.
183. Tiffany, K.A., et al., Structure of human isovaleryl-coa dehydrogenase at 2.6 angstrom resolution: Basis for substrate specificity. Biochemistry, 1997. **36**(28): p. 8455-8464.
184. Wischgoll, S., et al., Structural Basis for Promoting and Preventing Decarboxylation in Glutaryl-Coenzyme A Dehydrogenases. Biochemistry, 2010. **49**(25): p. 5350-5357.
185. Ghisla, S. And C. Thorpe, Acyl-coa dehydrogenases - A mechanistic overview. European Journal of Biochemistry, 2004. **271**(3): p. 494-508.
186. Battaile, K.P., et al., Crystal structure of rat short chain acyl-coa dehydrogenase complexed with acetoacetyl-coa - Comparison with other acyl-coa dehydrogenases. Journal of Biological Chemistry, 2002. **277**(14): p. 12200-12207.
187. Kim, J.J.P., M. Wang, and R. Paschke, Crystal-structures of medium-chain acyl-coa dehydrogenase from pig-liver mitochondria with and without substrate. Proceedings of the National Academy of Sciences of the United States of America, 1993. **90**(16): p. 7523-7527.
188. Bross, P., et al., Characterization of wild-type and an active-site mutant of human medium chain acyl-coa dehydrogenase after expression in Escherichia-coli. Journal of Biological Chemistry, 1990. **265**(13): p. 7116-7119.
189. Watmough, N.J. and F.E. Frerman, The electron transfer flavoprotein Ubiquinone oxidoreductases. Biochimica Et Biophysica Acta-Bioenergetics, 2010. **1797**(12): p. 1910-1916.
190. Sato, K., Y. Nishina, and K. Shiga, Electron-transferring flavoprotein has an AMP-binding site in addition to the FAD-binding site. Journal of Biochemistry, 1993. **114**(2): p. 215-222.

191. Toogood, H.S., et al., Extensive domain motion and electron transfer in the human electron transferring flavoprotein center dot medium chain acyl-coa dehydrogenase complex. Journal of Biological Chemistry, 2004. **279**(31): p. 32904-32912.
192. Chohan, K.K., et al., Protein dynamics enhance electronic coupling in electron transfer complexes. Journal of Biological Chemistry, 2001. **276**(36): p. 34142-34147.
193. Pedersen, J.I., Peroxisomal oxidation of the steroid side-chain in bile-acid formation. Biochimie, 1993. **75**(3-4): p. 159-165.
194. Pedersen, J.I., T. Veggan, and I. Bjorkhem, Substrate stereospecificity in oxidation of (25S)-3 alpha,7 alpha,12 alpha-trihydroxy-5 beta-cholestanoyl-coa by peroxisomal trihydroxy-5 beta-cholestanoyl-coa oxidase. Biochemical and Biophysical Research Communications, 1996. **224**(1): p. 37-42.
195. Zhang, Y.X., et al., Genes encoding acyl-coa dehydrogenase (acdh) homologues from Streptomyces coelicolor and Streptomyces avermitilis provide insights into the metabolism of small branched-chain fatty acids and macrolide antibiotic production. Microbiology-Sgm, 1999. **145**: p. 2323-2334.
196. Krasko, A., et al., Identification and expression of the SOS response, aidb-like, gene in the marine sponge Geodia cydonium: Implication for the phylogenetic relationships of metazoan acyl-coa dehydrogenases and acyl-coa oxidases. Journal of Molecular Evolution, 1998. **47**(3): p. 343-352.
197. Ma, D.K.K., et al., Acyl-coa Dehydrogenase Drives Heat Adaptation by Sequestering Fatty Acids. Cell, 2015. **161**(5): p. 1152-1163.
198. Rohankhedkar, M.S., et al., The aidb component of the Escherichia coli adaptive response to alkylating agents is a flavin-containing, DNA-binding protein. Journal of Bacteriology, 2006. **188**(1): p. 223-230.
199. Bowles, T., et al., Structure and DNA binding of alkylation response protein aidb. Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(40): p. 15299-15304.
200. Volkert, M.R., L.I. Hajec, and P. Landini, Overexpression of the Escherichia-coli AidB gene results in increased MNNG resistance. Journal of Cellular Biochemistry, 1995: p. 350-350.
201. Rippa, V., et al., Preferential DNA damage prevention by the E-coli aidb gene: A new mechanism for the protection of specific genes. DNA Repair, 2011. **10**(9): p. 934-941.
202. Landini, P., L.I. Hajec, and M.R. Volkert, Structure and transcriptional regulation of the Escherichia-coli adaptive response gene AidB. Journal of Bacteriology, 1994. **176**(21): p. 6583-6589.
203. Nakajima, Y., et al., Three-dimensional structure of the flavoenzyme Acyl-coa oxidase-II from rat liver, the peroxisomal counterpart of mitochondrial Acyl-coa dehydrogenase. Journal of Biochemistry, 2002. **131**(3): p. 365-374.
204. Spector, M.P., et al., The medium-/long-chain fatty acyl-coa dehydrogenase (fadf) gene of Salmonella typhimurium is a phase 1 starvation-stress response (SSR) locus. Microbiology-Sgm, 1999. **145**: p. 15-31.
205. Chang, G.G. and L. Tong, Structure and function of malic enzymes, a new class of oxidative decarboxylases. Biochemistry, 2003. **42**(44): p. 12721-12733.
206. Hsu, R.Y., Pigeon liver malic enzyme. Molecular and Cellular Biochemistry, 1982. **43**(1): p. 3-26.

207. Goodridge, A.G., et al., Nutritional and hormonal regulation of expression of the gene for malic enzyme. Progress in Nucleic Acid Research and Molecular Biology, Vol 52, 1996. **52**: p. 89-122.
208. Wynn, J.P. and C. Ratledge, Malic enzyme is a major source of NADPH for lipid accumulation by Aspergillus nidulans. Microbiology-Sgm, 1997. **143**: p. 253-257.
209. Moreadith, R.W. and A.L. Lehninger, The pathways of glutamate and glutamine oxidation by tumor-cell mitochondria - role of mitochondrial NAD(P)+-dependent malic enzyme. Journal of Biological Chemistry, 1984. **259**(10): p. 6215-6221.
210. Baggetto, L.G., Deviant energetic metabolism of glycolytic cancer-cells. Biochimie, 1992. **74**(11): p. 959-974.
211. Yang, C.D., et al., Glutamine Oxidation Maintains the TCA Cycle and Cell Survival during Impaired Mitochondrial Pyruvate Transport. Molecular Cell, 2014. **56**(3): p. 414-424.
212. Hibberd, J.M. and W.P. Quick, Characteristics of C-4 photosynthesis in stems and petioles of C-3 flowering plants. Nature, 2002. **415**(6870): p. 451-454.
213. Drincovich, M.F., P. Casati, and C.S. Andreo, NADP-malic enzyme from plants: a ubiquitous enzyme involved in different metabolic pathways. Febs Letters, 2001. **490**(1-2): p. 1-6.
214. Casati, P., et al., Malate metabolism by NADP-malic enzyme in plant defense. Photosynthesis Research, 1999. **61**(2): p. 99-105.
215. Sauer, U. And B.J. Eikmanns, The PEP-pyruvate-oxaloacetate node as the switch point for carbon flux distribution in bacteria. Fems Microbiology Reviews, 2005. **29**(4): p. 765-794.
216. Spaans, S.K., et al., NADPH-generating systems in bacteria and archaea. Front Microbiol, 2015. **6**: p. 742.
217. Taillefer, M., et al., Reassessment of the Transhydrogenase/Malate Shunt Pathway in Clostridium thermocellum ATCC 27405 through Kinetic Characterization of Malic Enzyme and Malate Dehydrogenase. Applied and Environmental Microbiology, 2015. **81**(7): p. 2423-2432.
218. Zhou, J.L., et al., Atypical Glycolysis in Clostridium thermocellum. Applied and Environmental Microbiology, 2013. **79**(9): p. 3000-3008.
219. Yao, P., et al., Evidence for a direct cross-talk between malic enzyme and the pentose phosphate pathway via structural interactions. J Biol Chem, 2017. **292**(41): p. 17113-17120.
220. Xu, Y.W., et al., Crystal structure of human mitochondrial NAD(P)(+)-dependent malic enzyme: a new class of oxidative decarboxylases. Structure with Folding & Design, 1999. **7**(8): p. 877-889.
221. Yang, Z., et al., Structural studies of the pigeon cytosolic NADP(+)-dependent malic enzyme. Protein Sci, 2002. **11**(2): p. 332-41.
222. Rao, G.S.J., et al., Crystallographic studies on Ascaris suum NAD-malic enzyme bound to reduced cofactor and identification of an effector site. Journal of Biological Chemistry, 2003. **278**(39): p. 38051-38058.
223. Hsieh, J.Y., et al., Structural characteristics of the nonallosteric human cytosolic malic enzyme. Biochim Biophys Acta, 2014. **1844**(10): p. 1773-83.
224. Rao, S.T. and M.G. Rossmann, Comparison of super-secondary structures in proteins. Journal of Molecular Biology, 1973. **76**(2): p. 241-250.

225. Yang, Z.R., et al., Structure of a closed form of human malic enzyme and implications for catalytic mechanism. Nature Structural Biology, 2000. **7**(3): p. 251-257.
226. Chang, G.G., et al., Kinetic mechanism of the cytosolic malic enzyme from human breast-cancer cell-line. Archives of Biochemistry and Biophysics, 1992. **296**(2): p. 468-473.
227. Kuo, C.W., et al., Metal-induced reversible structural interconversion of human mitochondrial NAD(P)(+)-dependent malic enzyme. Proteins-Structure Function and Genetics, 2004. **54**(3): p. 404-411.
228. Tao, X., Z. Yang, and L. Tong, Crystal Structures of Substrate Complexes of Malic Enzyme and Insights into the Catalytic Mechanism. Structure, 2003. **11**(9): p. 1141-1150.
229. Kuo, C.C., et al., Lysine residues 162 and 340 are involved in the catalysis and coenzyme binding of NADP(+)-dependent malic enzyme from pigeon. Biochemical and Biophysical Research Communications, 2000. **270**(3): p. 821-825.
230. Hsieh, J.Y., M.C. Chen, and H.C. Hung, Determinants of Nucleotide-Binding Selectivity of Malic Enzyme. Plos One, 2011. **6**(9): p. 9.
231. Scrutton, N.S., A. Berry, and R.N. Perham, Redesign of the coenzyme specificity of a dehydrogenase by protein engineering. Nature, 1990. **343**(6253): p. 38-43.
232. Karsten, W.E., et al., A catalytic triad is responsible for acid-base chemistry in the Ascaris suum NAD-Malic enzyme. Biochemistry, 2005. **44**(9): p. 3626-3635.
233. Tronconi, M.A., et al., NAD-malic enzymes of Arabidopsis thaliana display distinct kinetic mechanisms that support differences in physiological control. Biochemical Journal, 2010. **430**: p. 295-303.
234. Mckenna, M.C., et al., Regulation of mitochondrial and cytosolic malic enzymes from cultured rat brain astrocytes. Neurochemical Research, 1995. **20**(12): p. 1491-1501.
235. Yang, Z.R., C.W. Lanks, and L. Tong, Molecular mechanism for the regulation of human mitochondrial NAD(P)(+)-dependent malic enzyme by ATP and fumarate. Structure, 2002. **10**(7): p. 951-960.
236. Kobayashi, K., et al., Structure and properties of malic enzyme from Bacillus-stearothermophilus. Journal of Biological Chemistry, 1989. **264**(6): p. 3200-3205.
237. Kawai, S., et al., Purification and characterization of a malic enzyme from the ruminal bacterium Streptococcus bovis ATCC 15352 and cloning and sequencing of its gene. Applied and Environmental Microbiology, 1996. **62**(8): p. 2692-2700.
238. Gourdon, P., et al., Cloning of the malic enzyme gene from Corynebacterium glutamicum and role of the enzyme in lactate metabolism. Applied and Environmental Microbiology, 2000. **66**(7): p. 2981-2987.
239. Sender, P.D., et al., Characterization of an oxaloacetate decarboxylase that belongs to the malic enzyme family. Febs Letters, 2004. **570**(1-3): p. 217-222.
240. Lerondel, G., et al., ytsj has the major physiological role of the four paralogous malic enzyme isoforms in Bacillus subtilis. J Bacteriol, 2006. **188**(13): p. 4727-36.
241. Stols, L. And M.I. Donnelly, Production of succinic acid through overexpression of NAD(+)-dependent malic enzyme in an Escherichia coli mutant. Applied and Environmental Microbiology, 1997. **63**(7): p. 2695-2701.

242. Castano-Cerezo, S., et al., An insight into the role of phosphotransacetylase (pta) and the acetate/acetyl-coa node in Escherichia coli. Microb Cell Fact, 2009. **8**: p. 54.
243. Ferry, J.G., Methane from acetate. Journal of Bacteriology, 1992. **174**(17): p. 5489-5495.
244. Henkin, J. And R.H. Abeles, Evidence against an acyl-enzyme intermediate in reaction catalyzed by Clostridial phosphotransacetylase. Biochemistry, 1976. **15**(16): p. 3472-3479.
245. Kyrtopoulos, S.A. and D.P. Satchell, Kinetic studies with phosphotransacetylase .4. inhibition by products. Biochimica Et Biophysica Acta, 1972. **276**(2): p. 383-391.
246. Lawrence, S.H. and J.G. Ferry, Steady-state kinetic analysis of phosphotransacetylase from Methanosarcina thermophila. J Bacteriol, 2006. **188**(3): p. 1155-8.
247. Lawrence, S.H., et al., Structural and functional studies suggest a catalytic mechanism for the phosphotransacetylase from Methanosarcina thermophila. Journal of Bacteriology, 2006. **188**(3): p. 1143-1154.
248. Iyer, P.P. and J.G. Ferry, Role of arginines in coenzyme A binding and catalysis by the phosphotransacetylase from Methanosarcina thermophila. J Bacteriol, 2001. **183**(14): p. 4244-50.
249. Rasche, M.E., K.S. Smith, and J.G. Ferry, Identification of cysteine and arginine residues essential for the phosphotransacetylase from Methanosarcina thermophila. Journal of Bacteriology, 1997. **179**(24): p. 7712-7717.
250. Iyer, P.P., et al., Crystal structure of phosphotransacetylase from the methanogenic archaeon Methanosarcina thermophila. Structure, 2004. **12**(4): p. 559-67.
251. Xu, Q.S., et al., Crystal structure of a phosphotransacetylase from Streptococcus pyogenes. Proteins, 2004. **55**(2): p. 479-81.
252. Xu, Q.S., et al., Crystal structures of a phosphotransacetylase from Bacillus subtilis and its complex with acetyl phosphate. J Struct Funct Genomics, 2005. **6**(4): p. 269-79.
253. Mitsch, M.J., et al., Chimeric structure of the NAD(P)(+)- and NADP(+)-dependent malic enzymes of Rhizobium (Sinorhizobium) meliloti. Journal of Biological Chemistry, 1998. **273**(15): p. 9330-9336.
254. Bologna, F.P., C.S. Andreo, and M.F. Drincovich, Escherichia coli malic enzymes: two isoforms with substantial differences in kinetic properties, metabolic regulation, and structure. J Bacteriol, 2007. **189**(16): p. 5937-46.
255. Mitsch, M.J., A. Cowie, and T.M. Finan, Malic enzyme cofactor and domain requirements for symbiotic N-2 fixation by Sinorhizobium meliloti. Journal of Bacteriology, 2007. **189**(1): p. 160-168.
256. Voegele, R.T., M.J. Mitsch, and T.M. Finan, Characterization of two members of a novel malic enzyme class. Biochimica Et Biophysica Acta-Protein Structure and Molecular Enzymology, 1999. **1432**(2): p. 275-285.
257. Driscoll, B.T. and T.M. Finan, NAD+-dependent malic enzyme of Rhizobium-meliloti is required for symbiotic nitrogen-fixation. Molecular Microbiology, 1993. **7**(6): p. 865-873.
258. Bizouarn, T., et al., Proton translocating nicotinamide nucleotide transhydrogenase from E-coli. Mechanism of action deduced from its structural

and catalytic properties. Biochimica Et Biophysica Acta-Bioenergetics, 2000. **1457**(3): p. 211-228.
259. Decorosi, F., et al., Characterization of two genes involved in chromate resistance in a Cr(VI)-hyper-resistant bacterium. Extremophiles, 2009. **13**(6): p. 917-923.
260. Murai, S., et al., Inhibition of malic enzyme 1 disrupts cellular metabolism and leads to vulnerability in cancer cells in glucose-restricted conditions. Oncogenesis, 2017. **6**.
261. Pongratz, R.L., et al., Cytosolic and mitochondrial malic enzyme isoforms differentially control insulin secretion. Journal of Biological Chemistry, 2007. **282**(1): p. 200-207.
262. Goodsell, D.S. and A.J. Olson, Structural symmetry and protein function. Annual Review of Biophysics and Biomolecular Structure, 2000. **29**: p. 105-153.
263. Bergendahl, L.T. and J.A. Marsh, Functional determinants of protein assembly into homomeric complexes. Scientific Reports, 2017. **7**.
264. Fairman, J.W., et al., Structural basis for allosteric regulation of human ribonucleotide reductase by nucleotide-induced oligomerization. Nature Structural & Molecular Biology, 2011. **18**(3): p. 316-U102.
265. Kyrtopoulos, S.A. and D.P. Satchell, Kinetic studies with phosphotransacetylase .3. acylation of phosphate ions by acetyl coenzyme-A. Biochimica Et Biophysica Acta, 1972. **276**(2): p. 376-382.
266. Mitsch, M.J., A. Cowie, and T.M. Finan, Malic enzyme cofactor and domain requirements for symbiotic N2 fixation by Sinorhizobium meliloti. J Bacteriol, 2007. **189**(1): p. 160-8.
267. Keiler, K.C., P.R.H. Waller, and R.T. Sauer, Role of a peptide tagging system in degradation of proteins synthesized from damaged messenger RNA. Science, 1996. **271**(5251): p. 990-993.
268. Hara, H., et al., Cloning, mapping, and characterization of the Escherichia-coli prc gene, which is involved in c-terminal processing of penicillin-binding protein-3. Journal of Bacteriology, 1991. **173**(15): p. 4799-4813.
269. Keiler, K.C., et al., C-terminal specific protein-degradation - activity and substrate-specificity of the tsp protease. Protein Science, 1995. **4**(8): p. 1507-1515.
270. Che, Y., et al., C-terminal processing of reaction center protein D1 is essential for the function and assembly of photosystem II in Arabidopsis. Proc Natl Acad Sci U S A, 2013. **110**(40): p. 16247-52.
271. Clausen, T., et al., HTRA proteases: regulated proteolysis in protein quality control. Nat Rev Mol Cell Biol, 2011. **12**(3): p. 152-62.
272. Carroll, R.K., et al., The lone S41 family C-terminal processing protease in Staphylococcus aureus is localized to the cell wall and contributes to virulence. Microbiology, 2014. **160**(Pt 8): p. 1737-48.
273. Teoh, W.P., J.S. Matson, and V.J. dirita, Regulated intramembrane proteolysis of the virulence activator tcpp in Vibrio cholerae is initiated by the tail-specific protease (Tsp). Mol Microbiol, 2015. **97**(5): p. 822-31.
274. Clausen, T., C. Southan, and M. Ehrmann, The htra family of proteases: Implications for protein composition and cell fate. Molecular Cell, 2002. **10**(3): p. 443-455.
275. Sawa, J., et al., Molecular adaptation of the degq protease to exert protein quality control in the bacterial cell envelope. J Biol Chem, 2011. **286**(35): p. 30680-90.

276. Kennedy, M.B., Origin of pdz (dhr, glgf) domains. Trends in Biochemical Sciences, 1995. **20**(9): p. 350-350.
277. Hillier, B.J., et al., Unexpected modes of PDZ domain scaffolding revealed by structure of nnos-syntrophin complex. Science, 1999. **284**(5415): p. 812-815.
278. Rogelj, B., et al., The X11/Mint family of adaptor proteins. Brain Research Reviews, 2006. **52**(2): p. 305-315.
279. Utepbergenov, D.I., A.S. Fanning, and J.M. Anderson, Dimerization of the scaffolding protein ZO-1 through the second PDZ domain. Journal of Biological Chemistry, 2006. **281**(34): p. 24671-24677.
280. Long, J.F., et al., Supramodular structure and synergistic target binding of the N-terminal tandem PDZ domains of PSD-95. Journal of Molecular Biology, 2003. **327**(1): p. 203-214.
281. Im, Y.J., et al., Crystal structure of GRIP1 PDZ6-peptide complex reveals the structural basis for class IIPDZ target recognition and PDZ domain-mediated multimerization. Journal of Biological Chemistry, 2003. **278**(10): p. 8501-8507.
282. Fanning, A.S. and J.M. Anderson, Protein-protein interactions: PDZ domain networks. Current Biology, 1996. **6**(11): p. 1385-1388.
283. Cabral, J.H.M., et al., Crystal structure of a PDZ domain. Nature, 1996. **382**(6592): p. 649-652.
284. Saras, J. And C.H. Heldin, PDZ domains bind carboxy-terminal sequences of target proteins. Trends in Biochemical Sciences, 1996. **21**(12): p. 455-458.
285. Lee, H.J. and J.J. Zheng, PDZ domains and their binding partners: structure, specificity, and modification. Cell Communication and Signaling, 2010. **8**.
286. Doyle, D.A., et al., Crystal structures of a complexed and peptide-free membrane protein-binding domain: Molecular basis of peptide recognition by PDZ. Cell, 1996. **85**(7): p. 1067-1076.
287. Songyang, Z., et al., Recognition of unique carboxyl-terminal motifs by distinct PDZ domains. Science, 1997. **275**(5296): p. 73-77.
288. Tonikian, R., et al., A specificity map for the PDZ domain family. Plos Biology, 2008. **6**(9): p. 2043-2059.
289. Runyon, S.T., et al., Structural and functional analysis of the PDZ domains of human htra1 and htra3. Protein Science, 2007. **16**(11): p. 2454-2471.
290. Appleton, B.A., et al., Comparative structural analysis of the Erbin PDZ domain and the first PDZ domain of ZO-1 - Insights into determinants of PDZ domain specificity. Journal of Biological Chemistry, 2006. **281**(31): p. 22312-22320.
291. Zhang, Y.N., et al., Structural and functional analysis of the ligand specificity of the htra2/Omi PDZ domain. Protein Science, 2007. **16**(8): p. 1738-1750.
292. Singh, S.K., et al., Regulated proteolysis of a cross-link-specific peptidoglycan hydrolase contributes to bacterial morphogenesis. Proc Natl Acad Sci U S A, 2015. **112**(35): p. 10956-61.
293. Tudor, J.J., et al., Isolation of predation-deficient mutants of Bdellovibrio bacteriovorus by using transposon mutagenesis. Appl Environ Microbiol, 2008. **74**(17): p. 5436-43.
294. Tian, W., et al., castp 3.0: computed atlas of surface topography of proteins. Nucleic Acids Research, 2018. **46**(W1): p. W363-W367.
295. Song, W.S. and S.I. Yoon, Crystal structure of flic flagellin from Pseudomonas aeruginosa and its implication in TLR5 binding and formation of the flagellar filament. Biochemical and Biophysical Research Communications, 2014. **444**(2): p. 109-115.

296. Hoge, R., et al., The subcellular localization of a C-terminal processing protease in Pseudomonas aeruginosa. Fems Microbiology Letters, 2011. **316**(1): p. 23-30.
297. Aldridge, P. And K.T. Hughes, Regulation of flagellar assembly. Current Opinion in Microbiology, 2002. **5**(2): p. 160-165.
298. Vonderviszt, F., et al., Terminal regions of flagellin are disordered in solution. Journal of Molecular Biology, 1989. **209**(1): p. 127-133.
299. Auvray, F., et al., Flagellin polymerisation control by a cytosolic export chaperone. Journal of Molecular Biology, 2001. **308**(2): p. 221-229.
300. Sjoblad, R.D., C.W. Emala, and R.N. Doetsch, Bacterial flagellar sheaths - structures in search of a function. Cell Motility and the Cytoskeleton, 1983. **3**(1): p. 93-103.
301. Gewirtz, A.T., et al., Helicobacter pylori flagellin evades toll-like receptor 5-mediated innate immunity. Journal of Infectious Diseases, 2004. **189**(10): p. 1914-1920.
302. Mccarter, L.L., Genetic and molecular characterization of the polar flagellum of Vibrio-parahaemolyticus. Journal of Bacteriology, 1995. **177**(6): p. 1595-1609.
303. Deng, C.Y., et al., The periplasmic PDZ domain-containing protein Prc modulates full virulence, envelops stress responses, and directly interacts with dipeptidyl peptidase of Xanthomonas oryzae pv. Oryzae. Mol Plant Microbe Interact, 2014. **27**(2): p. 101-12.
304. Ostberg, Y., et al., Pleiotropic Effects of Inactivating a Carboxyl-Terminal Protease, ctpa, in Borrelia burgdorferi. Journal of Bacteriology, 2004. **186**(7): p. 2074-2084.
305. Schreiber, K.J. and D. Desveaux, algw regulates multiple Pseudomonas syringae virulence strategies. Mol Microbiol, 2011. **80**(2): p. 364-77.
306. Pasternak, Z., et al., By their genes ye shall know them: genomic signatures of predatory bacteria. Isme Journal, 2013. **7**(4): p. 756-769.
307. Borgnia, M.J., S. Subramaniam, and J.L.S. Milne, Three-dimensional imaging of the highly bent architecture of Bdellovibrio bacteriovorus by using cryo-electron tomography. Journal of Bacteriology, 2008. **190**(7): p. 2588-2596.

Abstract

This thesis documents the findings undertaken on a diverse set of proteins that operate at distinct stages of the predatory lifecycle of *Bdellovibrio bacteriovorus*. In general, this work explores three major predatory lifecycle events: growth & metabolism, development & signalling and finally bdelloplast lysis & exit. In particular this thesis characterises four individual proteins (Bd2924, Bd1833, Bd0967 and Bd0314) that perform very different roles.

In this thesis the three-dimensional structure of Bd2924 (a putative acyl-CoA dehydrogenase) has been solved demonstrating an adaptation of the common acyl-CoA dehydrogenase fold, suggesting an important role beyond fatty acid β-oxidation. This study also presents the first structure of a multi-modular malic enzyme (Bd1833) and through extensive mutational and biochemical investigations, have unpicked the mechanism behind allosteric regulation by acetyl-CoA. In addition the three-dimensional structure of Bd0967 (a C-terminal processing protease) has been solved, revealing a divergent fold that forms a large self-compartmentalised cavity. Furthermore, Bd0967 is demonstrated to be involved in a developmental pathway that is interconnected with flagellin assembly. Bd0967 targets and processively degrades the flagellin Bd0606, which is mediated by interactions with the PDZ domain. Finally this study determines how *Bdellovibrio* exits the consumed prey at the end of the predatory lifecycle that had been chemically marked for destruction. This role is provided by a repurposed lysozyme (Bd0314) remodeled into a variant that now only accepts deacetylated cell wall sugars. Enzyme assays and addition to bdelloplasts confirms the specificity for modified prey cell wall material.

Acknowledgments

First of all I would like to thank my PhD supervisor, Andrew Lovering. You have been a great support throughout my PhD, always being there to discuss ideas and offer advise. Your enthusiasm and knowledge for science has been infectious and constantly encouraged me to grow as a research scientist. I could not have asked for a better supervisor.

I would also like to thank Patrick Moynihan, for all of your help and guidance with all things biochem and cell wall. You have been a massive help and it will always be greatly appreciated.

I greatly appreciate all the collaborative work and contributions people have made towards my PhD studies, including Oliver Daubney and Charles Moore-Kelly. I would especially like to thank Prof Liz Sockett and all those in the Sockett lab for all of your contributions and hard work, as well as being very hospitable hosts on my visits to Nottingham.

I would like to specially thank Ian Cadby, you have always been there to offer a helping hand and advice; it has been a pleasure to work and learn from you. To the rest of the Lovering lab, thank you for the support over the years, especially Richard Meek for always lightening the mood with your quirky personality. I would also like to thank all those on the 7th floor, it has been a pleasure working with you all. Additionally, thank you Tim Knowles for being my secondary supervisor and a great guy.

Personally, a big thank you to all my family for always supporting me. Mum and Dad, I would not be where I am without your constant support and encouragement. Last, but certainly not least, I would like to thank my wonderful wife, Sarah. I could always rely on your love and support, especially when things got tough.

Table of Contents

Chapter 1 ..... **1**

Introduction **1**

1.1 Predatory Bacteria **1**

1.1.1 Overview 1

1.1.2 Types 2

1.1.3 Beneficial Applications 5

1.2 *Bdellovibrio bacteriovorus* **6**

1.2.1 Overview 6

1.2.2 Prey Detection and Motility 7

1.2.3 Attachment and Invasion 9

1.2.4 Niche formation 10

1.2.5 Metabolism 11

1.2.6 Growth and Division 14

1.2.7 Maturation and Lysis 14

1.2.8 Cellular Signalling and Regulation of Predation 15

1.3 Overall Research Aim Of This Study **18**

Chapter 2 ..... **20**

Materials and Methods **20**

2.1 General **20**

2.1.1 Molecular cloning 20

2.1.2 Colony PCR 22

2.1.3 Mutagenesis 22

2.1.4 Q5 SDM 23

2.1.5 Primers 24

2.1.6 Sequencing 26

2.1.7 Transformation 26

2.1.8 Culture Media 26

2.1.10 Glycerol Stocks 27

2.1.11 Expression 27

2.1.12 Purification 28

2.1.13 Dialysis 29

2.1.14 Size Exclusion Chromatography 29

2.1.15 SDS-PAGE 29

2.1.16 Protein sample concentration and quantification 30

2.1.17 Protein crystallisation 31

2.1.18 Data collection and structure solving 31

2.2 Bd0314 ..... **32**

2.2.1 Cloning, Expression and Purification 32

2.2.2 FITC Peptidoglycan digestion assay 33

2.2.3 Premature bdelloplast lysis assay 35

2.2.4 Crystallisation and Cryoprotection 35

2.2.5 Data collection and structure determination 36

2.3 Bd1413 and Bd1440 ..... **38**

2.3.1 Cloning, Expression and Purification 38

Table of Figures

Figure 1: Types of predatory bacteria 4
Figure 2: Host dependent predatory lifecycle 7
Figure 3: Proteins of interest 17
Figure 4: Bd1833 misannotation 42
Figure 5: HPLC assay standards 46
Figure 6: Predatory lifecycle, prey cell wall modifications 67
Figure 7: Peptidoglycan structure 68
Figure 8: Modifications of the peptidoglycan strand 72
Figure 9: Structures of lysozyme superfamily members 77
Figure 10: Catalytic distances 79
Figure 11: Retaining enzyme mechanism 80
Figure 12: Inverting enzyme mechanism 82
Figure 13: Lytic Transglycosylase reaction 83
Figure 14: Bd0314 purification and crystallisation 90
Figure 15: Bd0314 structure schematic 90
Figure 16: Bd0314 structure 92
Figure 17: Bd0314 catalytic mechanism 93
Figure 18: Bd0314 substrate-binding site 94
Figure 19: Substrate-binding surface view 96
Figure 20: Bd0314 conservation 98
Figure 21: Bd0314 alignment with Bdellovibrio HD100 homologues 99
Figure 22: Bd0314 alignment and residue conservation 100
Figure 23: Bd0314 specificity 102
Figure 24: Time and concentration dependent assay 103
Figure 25: Bd0314 active site residues 104
Figure 26: Bd0314 homologue activities 105
Figure 27: Bd0314 bdelloplast lysis 107
Figure 28: Bd0314 structure comparison 110
Figure 29: Bd0314 substrate binding 112
Figure 30: Comparison of Bd0314 with chitosanases 114
Figure 31: Bd0314 catalytic residue location 116
Figure 32: Role of Bd0314 119
Figure 33: β-oxidation metabolic pathway 124
Figure 34: Acyl-CoA dehydrogenases 127
Figure 35: MCAD structure 128
Figure 36: ACAD mechanism 131
Figure 37: Electron-transfer by ETF 132
Figure 38: Structure of ACO 133
Figure 39: Structure of divergent ACAD homologues 136
Figure 40: Bd2924 purification and crystallisation 137
Figure 41: Bd2924 SEC 138
Figure 42: Structure of Bd2924 139
Figure 43: Bd2924 tetramer complex 141
Figure 44: FAD hydrogen bonding network 142
Figure 45: C10-CoA binding 143

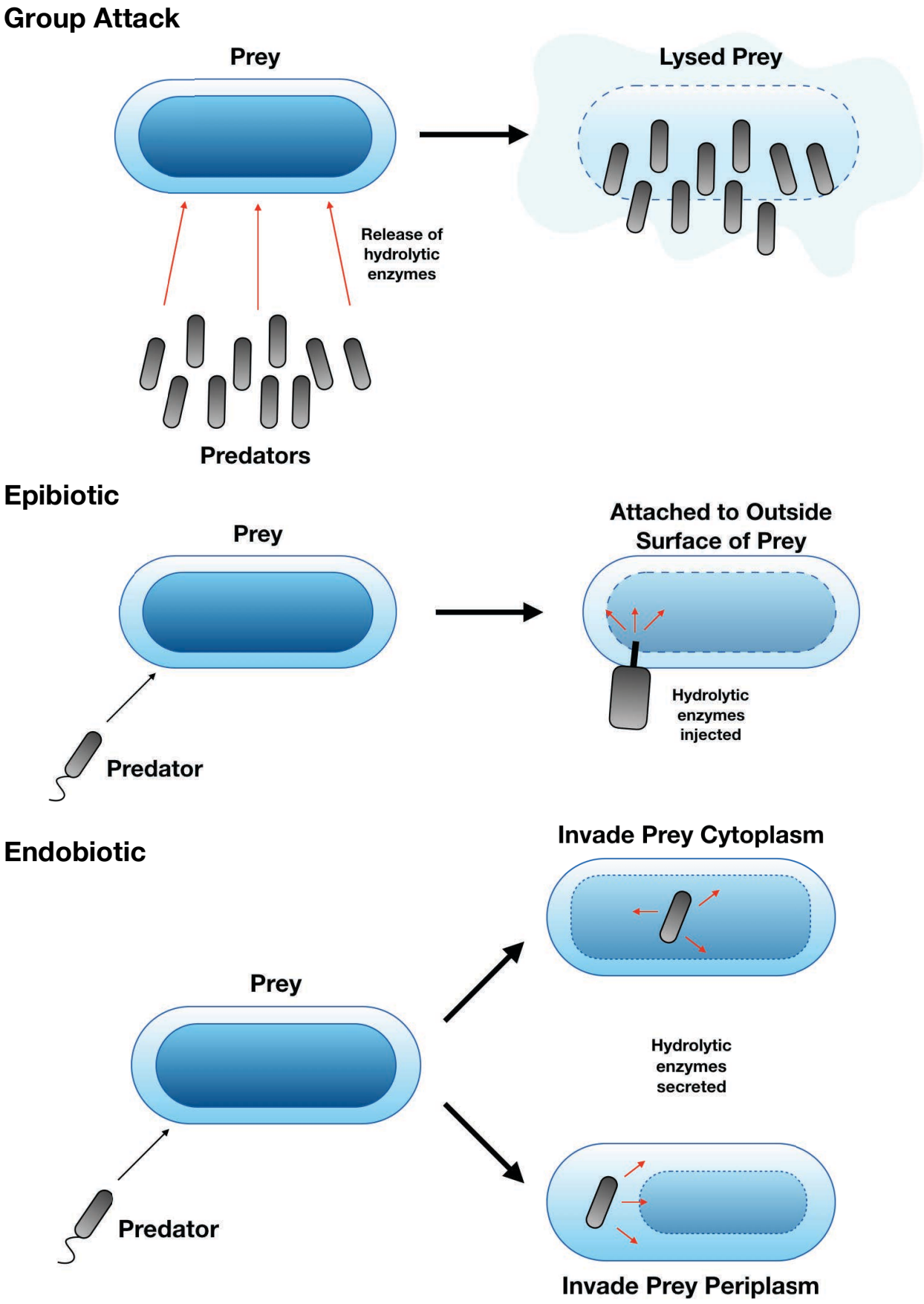


Figure 1 Types of predatory bacteria
Predatory bacteria can be categorised into three main groups based on how they kill their prey. These classes include: group attack, epibiotic and endobiotic predation. Predators are black ovals and the prey cells are blue ovals (illustration depicts Gram-negative cells). Red arrows show release of degradative hydrolytic arsenal is employed to kill the prey.

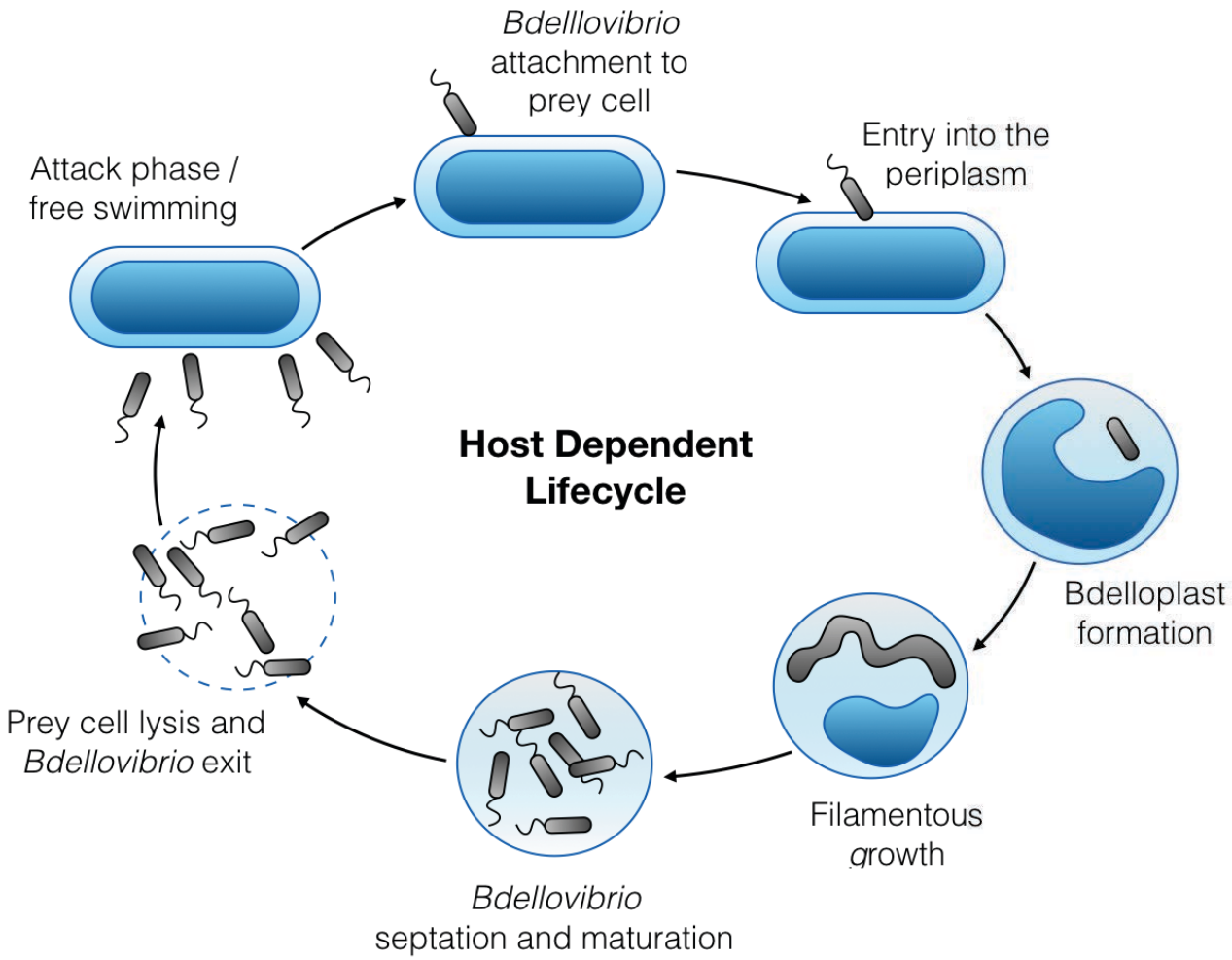


Figure 2 Host Dependent Predatory Lifecycle
The HD predatory lifecycle of *Bdellovibrio* can be divided into distinct stages. The cycle begins with free swimming and highly motile AP cells that actively seek out prey. Once located, *Bdellovibrio* attaches to the prey cell. Initially attachment is reversible (recognition period) but becomes permanent prior to invasion. Localised application of hydrolytic enzymes create a small pore in the prey cell envelope, which *Bdellovibrio* squeezes through, traversing the outer-membrane and cell wall to reside in the preys periplasm. During invasion the flagellum is shed and the entry pore is resealed. Additionally prey cell wall modifying enzymes are secreted, which remodel the prey cell wall creating an osmotically stable bdelloplast structure. Once the niche has been established *Bdellovibrio* begins the growth and replicative phase, growing filamentously. Once prey cell nutrients become depleted *Bdellovibrio* undergoes division and the progeny mature into AP cells. Mature AP progeny finally lyse the bdelloplast and begin the predatory cycle anew.

1.2.2 Prey Detection and Motility

Attack phase cells are small (0.2-0.5 µm x 0.5-2.5 µm), monoflagellate, rod shaped cells [11]. The attack phase cells are highly motile cells, swimming at speeds up to 160 µm sec^{-1} [26]. Motility is generated by a single, long, polar and sheathed flagellum that

Interestingly many of the novel CdG binders have putative roles in central metabolism (such as Bd2924), which suggests CdG may also regulate central metabolism. However, the reliance of co-factors (FAD, CoA, NAD) in this group of enzymes may have influenced non-specific binding. Another group of novel CdG binding proteins are components of the flagellum. CdG is known to negatively affect flagellum motility in other organisms [83]. In *Bdellovibrio* flagellar biosynthesis is coupled to CdG signalling, although no flagellar apparatus has been reported to interact directly with CdG. It is interesting to speculate that CdG interactions with flagellar components, such as flagellin (Bd0604) and hook protein (Bd3395), could inactivate flagellar activity upon invasion and its cause the flagellum to be shed [26].

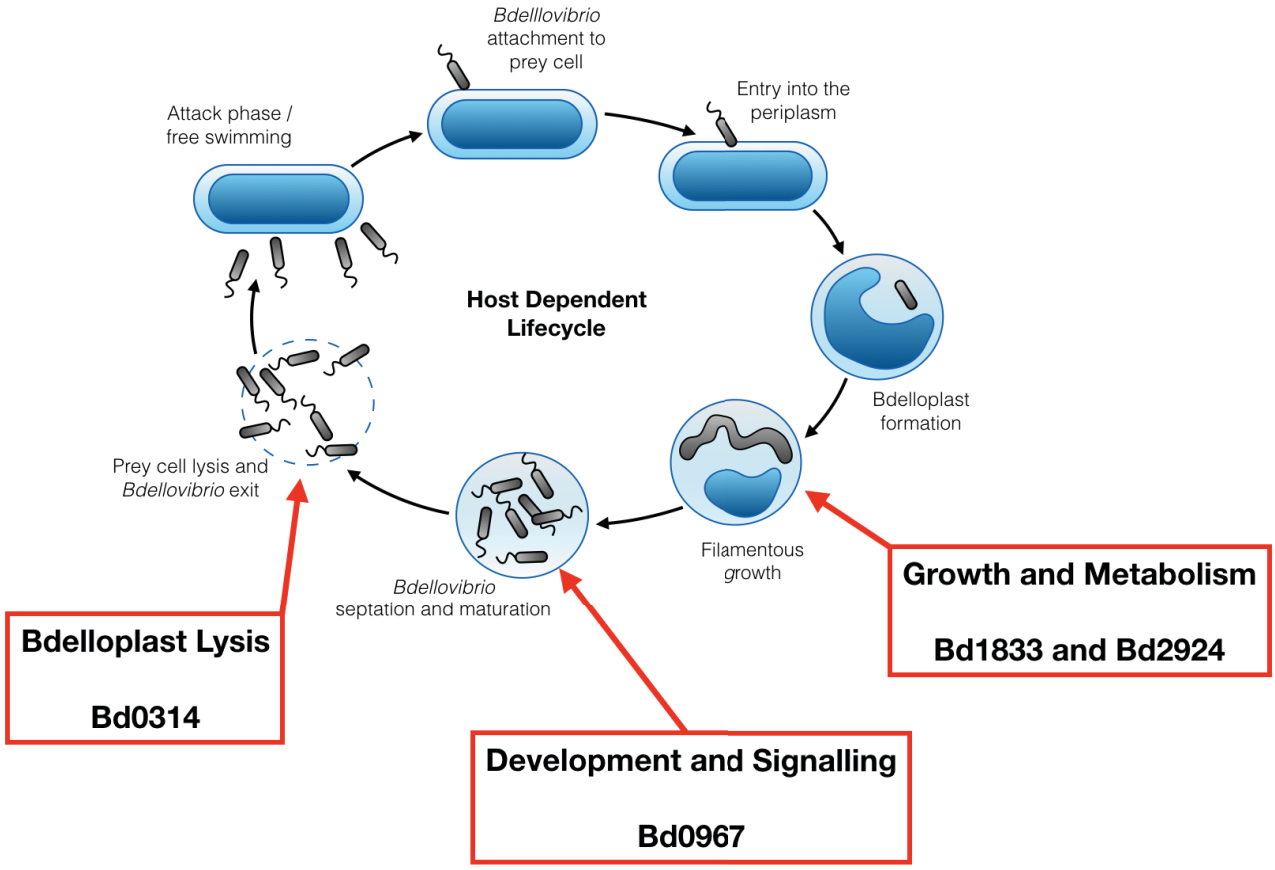


Figure 3. Proteins of interest
This Figure pinpoints the lifecycle events where the proteins of interest, for this thesis, are believed to function.

Protein	Plasmid	Primer (Fwd) (5' - 3')	Primer (Rev) (5' - 3')
Bd0314	pET28b-MBP	GAGAACCTGTACTTCCAATCCGGATCCCGCGA AGTC ATTCCTCTTTG	GGCTTTGTTAGCAGCCGGATCTCACTTGCAAA TAGACAGATTCTGGC
Bd1413	pET28a	GCAGCGGCCTGGTGCCGCGCGGCAGCCATGG CTATAAGATGCAGGCCCTGG	CTCAGTGGTGGTGGTGGTGGTGCTCGAGCTAT TTGCACAAGGGAAGAGTCTTGACG
Bd1440	pET28a	GCAGCGGCCTGGTGCCGCGCGGCAGCCATAG CTACAAAATAGAGGCCCTGTCC	CTCAGTGGTGGTGGTGGTGGTGCTCTTACTTA CAGAAGGATAATTTTTTGACCATGCC
Bd2924	pET28a	AGCAGCGGCCTGGTGCCGCGCGGCAGCCATA TGAAAAACTTTTACCAAGACGGc	CTCAGTGGTGGTGGTGGTGGTGCTCGAGCTAT AAAGAATACTTCTTCATAGCGATACGG
Bd1833	pET28a	GCAGCGGCCTGGTGCCGCGCGGCAGCCATAT GGACAACAAAACTGAAACTAAAATTGAAC	CTCAGTGGTGGTGGTGGTGGTGCTCGAGTTAC TTTTTGCCACGGCTTTTCAG
Bd1833-Pta	pET28a	GCAGCGGCCTGGTGCCGCGCGGCAGCCATTC CAAAGTCTTCATCCGCTCTG	CTCAGTGGTGGTGGTGGTGGTGCTCGAGTTAC TTTTTGCCACGGCTTTTCAG
Bd1833-ME	pET28a	GCAGCGGCCTGGTGCCGCGCGGCAGCCATAT GGACAACAAAACTGAAACTAAAATTGAAC	CTCAGTGGTGGTGGTGGTGGTGCTCGAGCTA AGATTCGCGGTACTGATCCC
Bd1833(SKB)	pET28a	GCAGCGGCCTGGTGCCGCGCGGCAGCCATAT GCAGCGAAAGGAAATATCC	CTCAGTGGTGGTGGTGGTGGTGCTCGAGTTAT TTGGTCAGTTCCGATTTTTTAGC
Bd1833(SKB)-Pta	pET28a	GCAGCGGCCTGGTGCCGCGCGGCAGCCATTC CAAAGTCTTCATTCGCTCTGCC	CTCAGTGGTGGTGGTGGTGGTGCTCGAGTTAT TTGGTCAGTTCCGATTTTTTAGC
Bd0967	pET28a	GCAGCGGCCTGGTGCCGCGCGGCAGCCATAT GAGTTCATTCTCAAAAGGCGTTAGAGG	CTCAGTGGTGGTGGTGGTGGTGCTCGAGCTG AGGCACCAAGGATTTCCC
Bd0606	pET28a	GCAGCGGCCTGGTGCCGCGCGGCAGCCATAT GGGAATGAGAATTTCTACG	CTCAGTGGTGGTGGTGGTGGTGCTCGAGTTA ACCAATCAGTCTTAGTGC

Table 1. Gene primers
A list of primers used in restriction free cloning.

Protein	Mutation	Type	Primer (Fwd) (5' - 3')	Primer (Rev) (5' - 3')
Bd0314	E143Q	QuickChange	ATGGTCAGATTTCAGAGTAACTTTAAAACC	GGTTTTAAAGTTACTCTGAAATCTGACCAT
Bd0314	E154Q	Q5	GTCCTACACCCAGGATTTTAAC	ATTGCGGTTTTAAAGTTAC
Bd0314	Y228F	Q5	TGGCGCCCGCGCCTGGTCTGTG	CCTTTCCAGGCTCCATCC
Bd0314	M150A	Q5	TAAAACCGCAGCGTCCTACACC	AAGTTACTCTCAAATCTGAC
Bd1440	E157Q	Q5	ACGCATGCATCAAACCACCATG	GAAGTCGGGCTGTAAGCG
Bd0967	S547A	Q5	TATCTCTGCTGCCGCTTCCGA	CGGCTGGTCAATACCACC
Bd0967	R299M	Q5	CAAAAAAATCATGGGTAACAAGGGCACGAAAGTG	ACAACGTCTTTCAAATCC
Bd1833	R535A	Q5	GTATTCTTTGGCCCAGCGTAAAGG	AGCTTTTCAACGAAGCTG
Bd1833	R535E	Q5	GTATTCTTTGGAGCAGCGTAAAGGC	AGCTTTTCAACGAAGCTG
Bd1833	Y576C	Q5	TCCATCAACTGTGCCGATGCG	GGATCCACTGACCATACC
Bd1833	Y576S	Q5	TCCATCAACTCTGCCGATGCG	GGATCCACTGACCATACC
Bd1833	E544R	Q5	CAATCTGGGCAGAGCCGAGCGC	ATGCCTTTACGCTGGCGC
Bd1833	E544A	Q5	CAATCTGGGCGCAGCCGAGCGCT	ATGCCTTTACGCTGGCGC
Bd1833	N718D	Q5	GGAATCCTCCGACATCGCCTA	AGATTCGGGAAGACCAGC
Bd1833	ΔLoop	Q5	CTTAACGAGAACAACCTGATCCCGAAACCCTTC	GTCTTGCTCTTTGATCAGTTTTGCCAGTGCATG
Bd1833	R547A	Q5	CGAAGCCGAGGCCTTGATGGCGG	CCCAGATTGATGCCTTTAC
Bd1833	R749A	Q5	CGTTCTGCAAGCAACCACAACGGTTGATGGC	TTGGCAGAGCGGCGCACA
Bd1833	R581K	Q5	CGATGCGGTTAAGCCGATTCTGC	GCATAGTTGATGGAGGATC
Bd0606	LIG	Q5	TAACTCGAGCACCACCAC	TCTTAGTGCAGAGTTTGGC
Bd0606	RLIG	Q5	TAACTCGAGCACCACCAC	TAGTGCAGAGTTTGGCATC

Table 2. SDM primers
A list of primers used for SDM

sites were used in a round of automated model building using Phenix autobuild [111]. The auto-built model successfully traced most of the Bd0314 sequence (2 fragments, 170 residues, R-work = 0.2245, R-free = 0.2446). See Section 2.1.18 for how the initial model was improved. Datasets for Bd0314-Wt and E154Q were collected and solved in a typical process using molecular replacement as outlined in Section 2.1.17 and 2.1.18.

Data Collection				
	Bd0314 Wt (Jcsg A1)	Bd0314 Wt (Jcsg D4)	Bd0314 E143Q	Bd0314 E154Q
Spacegroup	P 2_1 2_1 2	P 2_1 2_1 2_1	P 2_1 2_1 2_1	P 2_1 2_1 2
Cell Dimensions a, b, c (Å)	84.7, 47.8, 49.2	48.9, 64.7, 82.7	48.1, 62.1, 80.9	84.7, 47.9, 49.2
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (Å)	47.818 - 1.361	64.67-1.26	62.13 - 1.82	49.21 - 1.37
Total observations	531095 (14923)	812815 (25973)	968126 (1081)	566104 (29666)
Unique reflections	42769 (1766)	71428 (3469)	19553 (149)	45432 (2275)
Rmerge(I)	0.088 (0.785)	0.0682 (1.786)	0.1612 (0.592)	0.137 (1.713)
Rpim	0.036 (0.384)	0.02 (0.692)	0.021 (0.240)	0.056 (0.694)
I/sigI	16 (2)	13.17 (1.2)	20.1 (2.1)	12.2 (2.2)
CC(1/2)	0.998 (0.804)	0.9994 (0.479)	0.9983 (0.638)	0.994 (0.617)
Completeness (%)	98.1 (82.8)	99.8 (98.4)	87.65 (13.6)	98.4 (100)
Multiplicity	12.4 (8.5)	11.4 (7.5)	49.51 (7.3)	12.5 (13)
Refinement				
R-work	0.1739	0.1699	0.1721	0.1771
R-free	0.1911	0.1940	0.2101	0.1951
RMS (bonds) (Å)	0.008	0.0160	0.006	0.005
RMS (angles) (°)	1.142	1.593	0.789	1.044
Average B-Factor ($Å^2$)	18.1	21.2	18.2	17.8
# of non-hydrogen	1691	1729	1650	1695
Molecules per ASU	1	1	1	1
Ramachandran Plot				
Preferred region	97.1	98.8	96.6	97.1
Allowed region	2.9	1.2	3.4	2.9

Table 3. Bd0314 data collection and refinement

Data Collection	Bd2924 native	Bd2924 (C10-CoA)
Spacegroup	C 2	C 2 2 2_1
Cell Dimensions a, b, c (Å)	137.8, 113.0, 89.0	126.3, 134.7, 133.2
α, β, γ (°)	90, 120.8, 90	90, 90, 90
Resolution (Å)	81.71 - 1.51	133.18 - 1.87
Total observations	1198412 (53307)	1226302 (51988)
Unique reflections	177144 (8571)	93777 (4648)
Rmerge(I)	0.0508 (1.2811)	0.073 (0.860)
Rpim	0.021 (0.555)	0.030 (0.375)
I/sigI	11.16 (1.12)	19.3 (2.1)
CC(1/2)	0.9986 (0.7687)	0.999 (0.833)
Completeness (%)	96.87 (93.85)	100 (99.6)
Multiplicity	6.77 (6.22)	13.1 (11.2)

Refinement		
R-work	0.1545	0.1849
R-free	0.1914	0.2167
RMS (bonds) (Å)	0.018	0.008
RMS (angles) (°)	1.79	1.222
Average B-Factor ($Å^2$)	31.0	32.1
# of non-hydrogen atoms	8572	8909
Molecules per ASU	2	2
Ramachandran Plot (%)		
Preferred region	98	98.2
Allowed region	1.5	1.3

Table 4. Bd2924 data collection and refinement

2.5 Bd1833

2.5.1 Cloning, Expression and Purification

The gene for Bd1833 had been mis-annotated on NCBI and KEGG databases, due to an error in sequencing reads, which results in a premature stop codon. Consequently Bd1833 is annotated as two gene fragments (Bd1833 and Bd1834). The construct of full-length Bd1833 was designed to start at the start codon of *bd1834* and finish at the stop codon of *bd1833*. Bd1833/Bd1834 was successfully cloned into pET28a expression vector and its nucleotide sequence was confirmed by DNA sequencing. DNA sequencing of the Bd1833/Bd1834 cloned region confirmed Bd1833 exists as a single gene and the original stop codon was incorrectly placed due to a frame shift in sequencing reads (Figure 4).



Figure 4. Bd1833 misannotation
bd1833/bd1834 gene is mis-annotated as two genes on NCBI; in fact *bd1833/bd1834* encode a single protein – Bd1833. A frameshift had occurred during the original sequencing producing a premature stop codon and annotation as two gene products. The region starting from *bd1834* start codon through to the stop codon of *bd1833* was cloned and re-sequenced. Re-sequencing revealed the correct sequence read through this region and the position of the frame shift (highlight in red), which in turn makes the stop codon (highlighted in red).

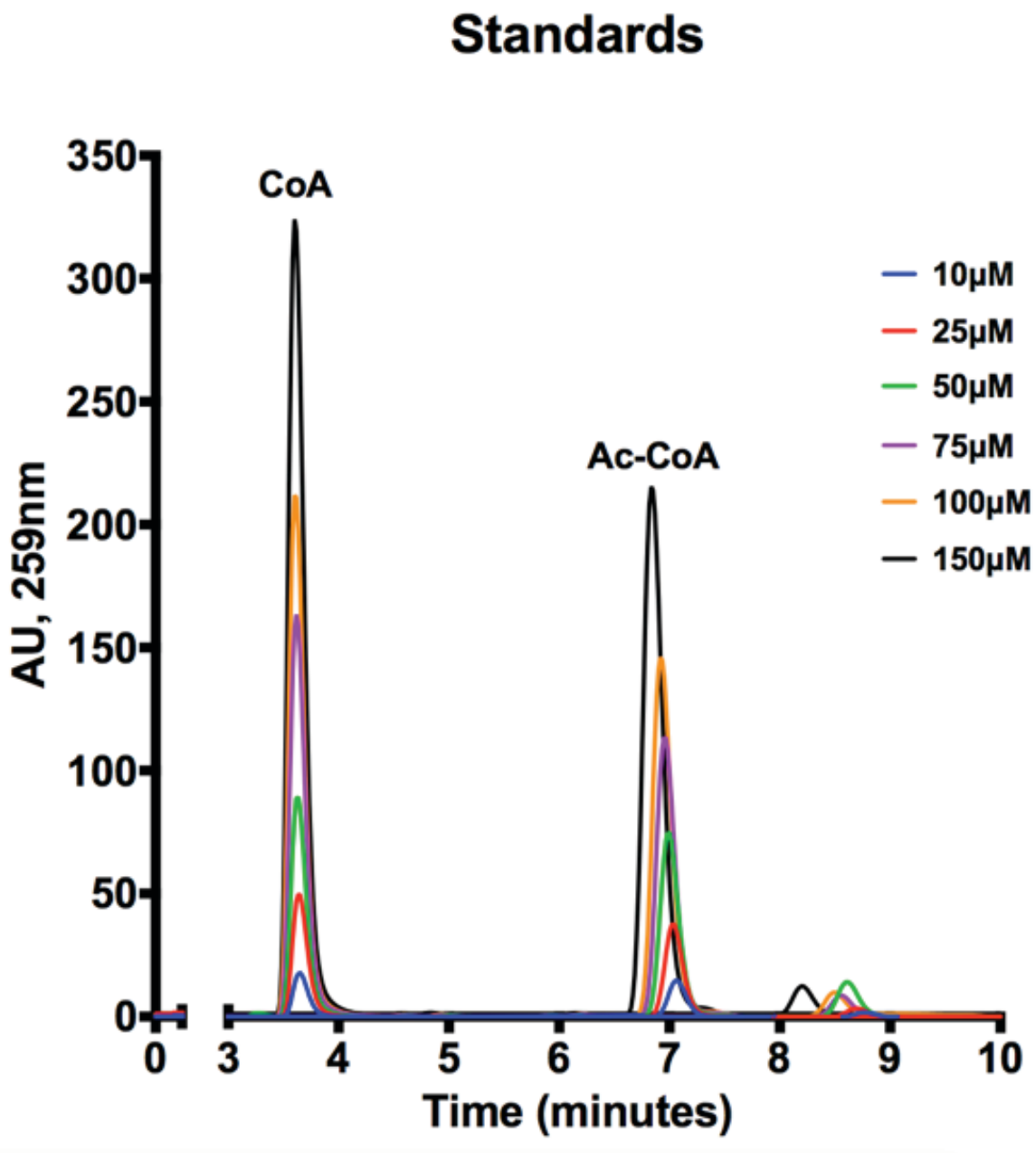


Figure 5. HPLC assay standards
CoA and acetyl-CoA had well-separated and distinct peaks, using HPLC.

2.5.5 Enzymatic activity assay

Spectrophotometric assays were used to monitor the oxidative decarboxylation of L-malate to pyruvate and the reductive carboxylation of pyruvate to L-malate using a CARY 300 Bio, UV-Visible spectrophotometer operated by CARY WinUV kinetics software. All assays recorded the absorbance at 340 nm (NADPH) for a 2 minute period

after the first crystals developed; reduced growth time resulted in better quality diffraction. A soaking experiment was used to get the Bd1833-ME ($NADP^+$) complex structure. Bd1833-ME crystals were transferred into a solution of mother liquor supplemented with (1mM $NADP^+$ and 20% v/v ethylene glycol) and incubated for one hour before being snap frozen in liquid nitrogen.

	Crystallisation Condition
Bd1833 (Full-length)	0.2M sodium citrate tribasic monohydrate, 0.1M BisTris Propane pH 8.5, 20% w/v PEG 3350
Bd1833-ME (native)	1) 0.2M sodium fluoride, 20% w/v PEG 3350 2) 0.1M Bicine pH 9.3, 25% v/v PEG smear medium
Bd1833-ME ($NADP^+$)	0.1M Bicine pH 9.3, 25% v/v PEG smear medium
Bd1833-Pta	0.2M TMAO, 0.1M Tris pH 8.5, 20% w/v PEG2000 MME
Bd1833-Pta (Apo)	0.1M magnesium acetate, 0.1M sodium acetate pH 4.5, 8% w/v PEG 8000
Bd1833-Pta (R535A)	0.2M magnesium chloride hexahydrate, 10% v/v ethylene glycol, 0.1M HEPES pH 7.5, 15% v/v PEG Smear Medium, 5% v/v 2-propanol
Bd1833-Pta (R535E)	0.1M magnesium chloride, 0.1M HEPES pH 7.5, 10% w/v PEG400
Bd1833-Pta (E544R)	0.2M magnesium chloride hexahydrate, 10% v/v ethylene glycol, 0.1M HEPES pH 7.5, 15% v/v PEG Smear Medium, 5% v/v 2-propanol
Bd1833-Pta (Y576S)	0.2m sodium iodide, 0.1M Bis-Tris propane pH 8.5, 20% w/v PEG 3350
Bd1833-SKB (Full)	0.1M Tris pH 8.0, 25% v/v PEG 350 MME
Bd1833-SKB (Pta)	0.005M cobalt(II) chloride hexahydrate, 0.005M cadmium chloride hemi(pentahydrate), 0.005M magnesium chloride hexahydrate, 0.005M nickle(II) chloride hexahydrate, 0.1M HEPES pH 7.5, 12% w/v PEG 3350

Table 5. Bd1833 crystallisation conditions

Data Collection	Bd1833-ME (Bcs-B6)	Bd1833-ME (Pact-E1)	Bd1833-ME ($NADP^+$)
Spacegroup	P 2_1	P1	P 2
Cell Dimensions a, b, c (Å)	48.1, 92.7, 97.1	56.2, 100.2, 130.9	47.8, 92.2, 95.8
α, β, γ (°)	90, 91.1, 90	88.1, 83.7, 81.9	90, 91.3, 90
Resolution (Å)	43.41 - 1.67	37.07 - 2.2	95.78 - 1.67
Total observations	656294 (32403)	475063 (21446)	626225 (29211)
Unique reflections	96511 (4697)	138460 (6888)	96099 (4736)
Rmerge(I)	0.0851 (1.6100)	0.1055 (0.9044)	0.1191 (0.7974)
Rpim	0.036 (0.634)	0.068 (0.621)	0.050 (0.350)
I/sigI	10.81 (1.22)	6.58 (1.02)	6.87 (1.32)
CC(1/2)	0.9977 (0.4634)	0.9939 (0.5009)	0.9962 (0.4718)
Completeness (%)	97.80 (96.54)	97.23 (96.17)	99.90 (98.14)
Multiplicity	6.81 (6.98)	3.43 (3.11)	6.52 (6.17)
Refinement			
R-work	0.1702	0.2109	0.1799
R-free	0.2039	0.2264	0.2159
RMS (bonds) (Å)	0.0077	0.0106	0.0108
RMS (angles) (°)	1.113	1.429	1.257
Average B-Factor ($Å^2$)	37	53.3	34.5
# of non-hydrogen	6808	18553	6913
Molecules per ASU	2	6	2
Ramachandran Plot			
Preferred region	96.4	95.5	96.1
Allowed region	3.3	4.1	3.7

Table 6. Bd1833-ME data collection and refinement

Data Collection	Bd1833-Pta	Apo	R535A	R535E	E544R	Y576S
Spacegroup	$P2_1$	$P\ 2_1\ 2_1\ 2_1$	C 2	C 2	C 2	C 2
Cell Dimensions a, b, c (Å)	111.2, 60.6, 167.8	139.2, 151.3, 285.3	129.6, 183.7, 119.9	129.4, 183.4, 120.2	129.1, 182.9, 119.7	128.8, 182.4, 121.4
α, β, γ (°)	90, 91.1, 90	90, 90, 90	90, 117, 90	90, 117, 90	90, 118, 90	90
Resolution (Å)	91.86 - 1.96	142.66 - 2.72	64.50 - 2.35	107.10 - 2.39	69.22 - 2.33	95.97 - 2.38
Total observations	1083798 (55212)	2082343 (102453)	705212 (37173)	424649 (17692)	698066 (266312)	663750 (28462)
Unique reflections	160802 (7925)	162146 (7836)	103459 (5235)	96539 (4863)	104231 (5084)	98271 (4884)
Rmerge(I)	0.0589 (1.4047)	0.1294 (0.8860)	0.0867 (1.8559)	0.103 (1.975)	0.0872 (1.2654)	0.119 (1.079)
Rpim	0.025 (0.571)	0.038 (0.254)	0.036 (0.744)	0.055 (1.201)	0.036 (0.600)	0.042 (0.47)
I/sigI	15.70 (1.14)	11.40 (1.47)	10.69 (1.16)	8.5 (0.5)	12.03 (1.14)	10.8 (0.7)
CC(1/2)	0.999 (0.637)	0.998 (0.600)	0.999 (0.573)	0.998 (0.697)	0.999 (0.590)	0.999 (0.569)
Completeness (%)	99.92 (99.92)	99.95 (97.82)	99.90 (100.00)	98.4 (99.7)	99.72 (97.98)	100 (100)
Multiplicity	6.74 (6.97)	12.84 (13.07)	6.81 (7.10)	4.4 (3.6)	6.70 (5.24)	6.8 (5.8)

Refinement						
R-work	0.1829	0.2520	0.2158	0.2153	0.2069	0.2222
R-free	0.2052	0.2995	0.2368	0.2602	0.2265	0.2725
RMS (bonds) (Å)	0.009	0.012	0.0103	0.01	0.0097	0.01
RMS (angles) (°)	1.178	1.590	1.498	1.415	1.448	1.321
Average B-Factor ($Å^2$)	29.5	86.2	45.7	75.5	74.9	74.1
# of non-hydrogen	16475	31268	15572	15628	15849	15720
Molecules per ASU	6	12	6	6	6	6
Ramachandran Plot						
Preferred region	95.8	96.2	98.7	96.6	97.0	96.3
Allowed region	4.0	3.6	1.3	3.1	2.7	3.6

Table 7. Bd1833-Pta data collection and refinement

Data Collection	Bd1833 (Full-length)
Spacegroup	C 2
Cell Dimensions *a, b, c (Å)*	250.422 144.792 171.138
α, β, γ (°)	90.000 100.605 90.000
Resolution (Å)	124.80 - 2.72
Total observations	1086009 (56926)
Unique reflections	160818 (8003)
Rmerge(I)	0.1 (2.788)
Rpim	0.042 (1.125)
I/sigI	11.0 (0.6)
CC(1/2)	0.999 (0.317)
Completeness (%)	99.9 (100)
Multiplicity	6.8 (7.1)

Refinement	
R-work	0.1922
R-free	0.2277
RMS (bonds) (Å)	0.0108
RMS (angles) (°)	1.437
Average B-Factor (Å^2)	108
# of non-hydrogen atoms	34603
Molecules per ASU	6
Ramachandran Plot (%)	
Preferred region	95.5
Allowed region	4.1

Table 8. Bd1833 full-length data collection and refinement

Data Collection	Bd1833-SKB (Full-length)	Bd1833-SKB (Pta)
Spacegroup	P 4_1 2_1 2	F 2 3
Cell Dimensions a, b, c (Å)	174.62, 174,62, 203.29	307.09, 307.09, 307.09
α, β, γ (°)	90, 90, 90	90, 90, 90
Resolution (Å)	132.46 - 3.71	54.29 - 2.71
Total observations	885942 (45499)	663665 (33984)
Unique reflections	34022 (1653)	64779 (3209)
Rmerge(I)	0.195 (6.015)	0.1309 (1.8829)
Rpim	0.034 (0.620)	0.043 (0.606)
I/sigI	13.6 (0.7)	12.06 (1.25)
CC(1/2)	1.000 (0.393)	0.9981 (0.4515)
Completeness (%)	100 (100)	100.00 (99.97)
Multiplicity	26.0 (27.5)	10.25 (10.59)

Refinement		
R-work	0.2503	0.2040
R-free	0.2996	0.2392
RMS (bonds) (Å)	0.004	0.013
RMS (angles) (°)	0.783	1.43
Average B-Factor ($Å^2$)	199.1	64.5
# of non-hydrogen atoms	17163	10592
Molecules per ASU	3	3
Ramachandran Plot (%)		
Preferred region	95.1	97.0
Allowed region	4.3	3.0

Table 9. Bd1833 SKB data collection and refinement

Data Collection	Bd0967 Wt	Bd0967 S457A
Spacegroup	P1	P1
Cell Dimensions a, b, c (Å)	73.4, 95.8, 114.1	72.9, 95.0, 113.3
α, β, γ (°)	77.5, 71.3, 81.9	77.7, 71.8, 81.9
Resolution (Å)	106.26 - 2.04	38.42 - 2.58
Total observations	848316 (18620)	304174 (15674)
Unique reflections	179707 (8159)	87121 (4331)
Rmerge(I)	0.1656 (0.3533)	0.1133 (1.0666)
Rpim	0.079 (0.280)	0.071 (0.654)
I/sigI	4.78 (1.23)	7.52 (1.13)
CC(1/2)	0.9517 (0.7332)	0.9946 (0.4479)
Completeness (%)	97.84 (89.57)	98.52 (98.12)
Multiplicity	4.72 (2.28)	3.49 (3.62)
Refinement		
R-work	0.1990	0.2118
R-free	0.2399	0.2607
RMS (bonds) (Å)	0.0104	0.010
RMS (angles) (°)	1.431	1.328
Average B-Factor (Å^2)	43.1	59.9
# of non-hydrogen atoms	21634	19935
Molecules per ASU	4	4
Ramachandran Plot (%)		
Preferred region	97.6	97.0
Allowed region	1.4	2.2

Table 10. Bd0967 data collection and refinement

2.6.11 MST Cdg Binding Assay

The same procedure as described for Bd2924 MST was used. 20 mM HEPES pH 7.5, 50 mM KCl, 5 mM $MgCl_2$ was used as the MST buffer.

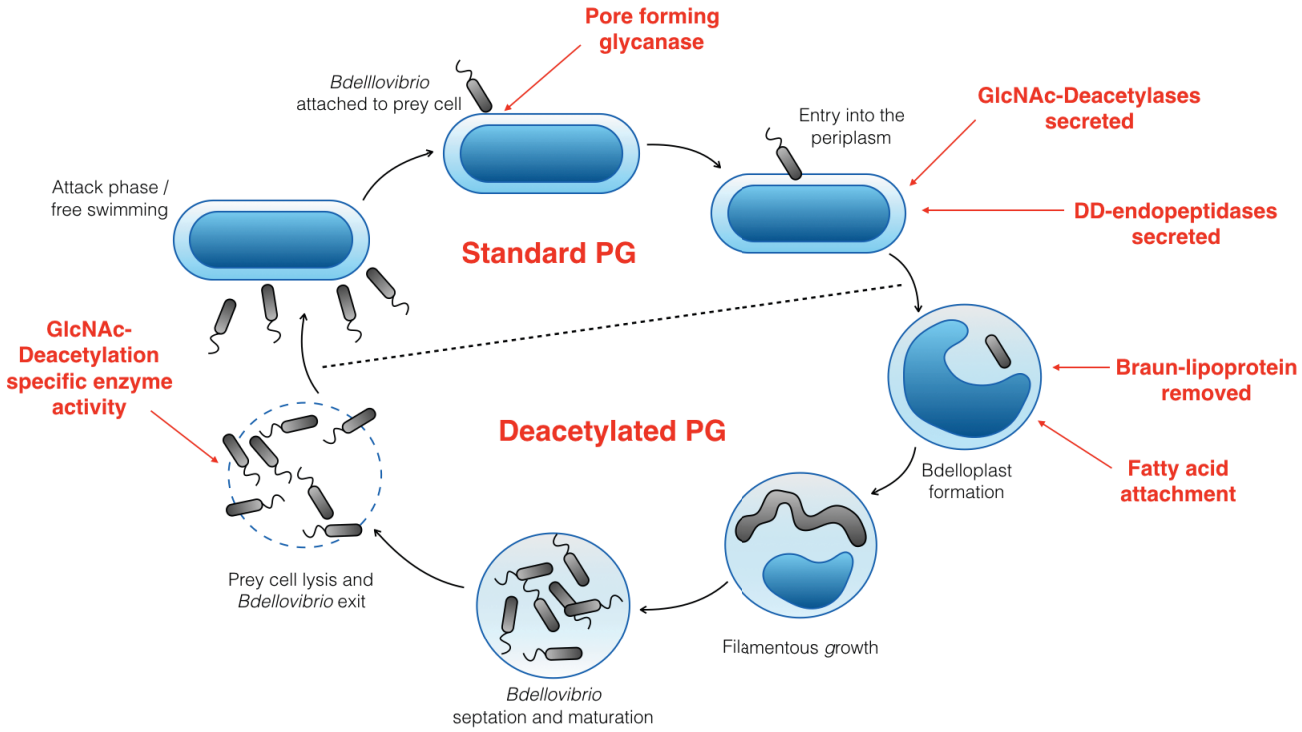


Figure 6. *Bdellovibrio* lifecycle highlighting prey cell wall modifications.
Bdellovibrio actively seeks out prey cells and attach to the outer membrane. They then secrete a localised pore forming glycanase that creates a small entry pore in the prey cell wall that Bdellovibrio squeezes through and then reseals. During invasion *Bdellovibrio* secretes deacetylases and DD-endopeptidases. GlcNAc-deacetylases (Bd0468 and Bd3279) deacetylate the prey cell wall marking 'prey' from 'self'. *N*-deacetylation of the prey cell wall makes the bdelloplast structure resistant to lysozyme activity. DD-endopeptidases (0816 and 3459) break cross-links in the prey cell wall and rounds the prey cell into the bdelloplast structure. Rounding of the prey cell acts as an 'occupied' signal preventing further invasion events. A significant proportion of the prey cell Braun-lipoprotein is then removed and fatty acid chains are attached to the prey cell wall, which stabilises the bdelloplast structure. At the end of the predatory cycle a final lytic enzyme activity is required to completely solubilise the prey cell wall and lyse the bdelloplast structure. The final lytic enzyme activity must be highly regulated and active towards the deacetylated prey cell wall.

3.1.1 Peptidoglycan Overview

3.1.1.1 Overview

Peptidoglycan is an essential feature of bacterial cell walls, located on the outside of the cytoplasmic membrane. The main function of peptidoglycan is to withstand turgor pressure and maintain the cellular integrity. If the peptidoglycan layer is compromised, by inhibiting its biosynthesis (e.g. antibiotics) or specific degradation (e.g. lysozymes), cells lyse. Peptidoglycan is directly involved in cell growth and division, while also

being required to maintain cell shape and serve as a molecular scaffold for anchoring other cell envelope components such as teichoic acids [116].

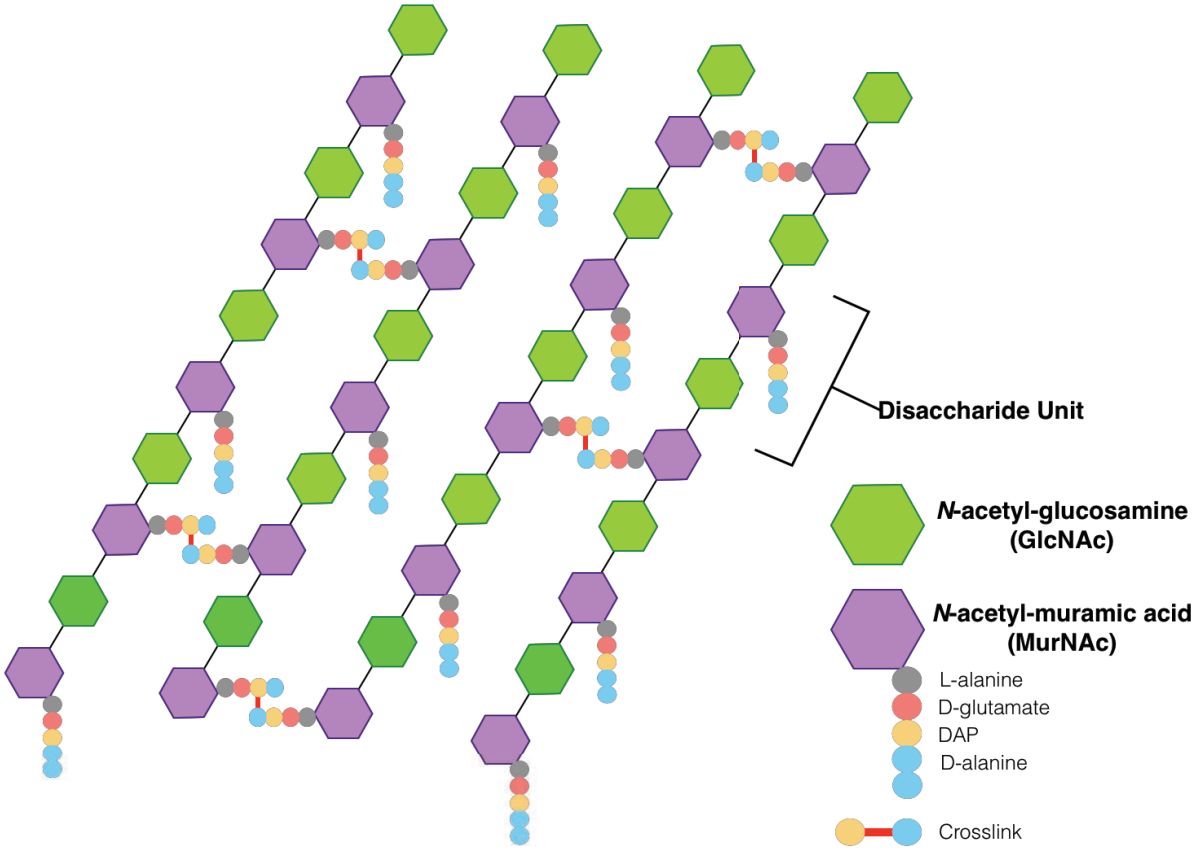


Figure 7. Peptidoglycan structure
This Figure represents the peptidoglycan structure of *E. coli*. The glycan strand is composed of alternating β-1-4 linked GlcNAc (green) and MurNAc (purple) residues. The basic disaccharide subunit (monomer) is highlighted. Components making up the *E. coli* peptide stem are highlighted. Cross-links between the tetrapeptide stems are represented by a red bond.

3.1.1.2 Structure and Composition

Peptidoglycan is comprised of a linear chain of glycan stands that are cross-linked together by short peptide strands (Figure 7). The glycan strands or made up of alternating *N*-acetylglucosamine (GlcNAc) and *N*-acetylmuramic acid (MurNAc) residues, linked via β-1-4 bonds. MurNAc contains a D-lactoyl group that is substituted

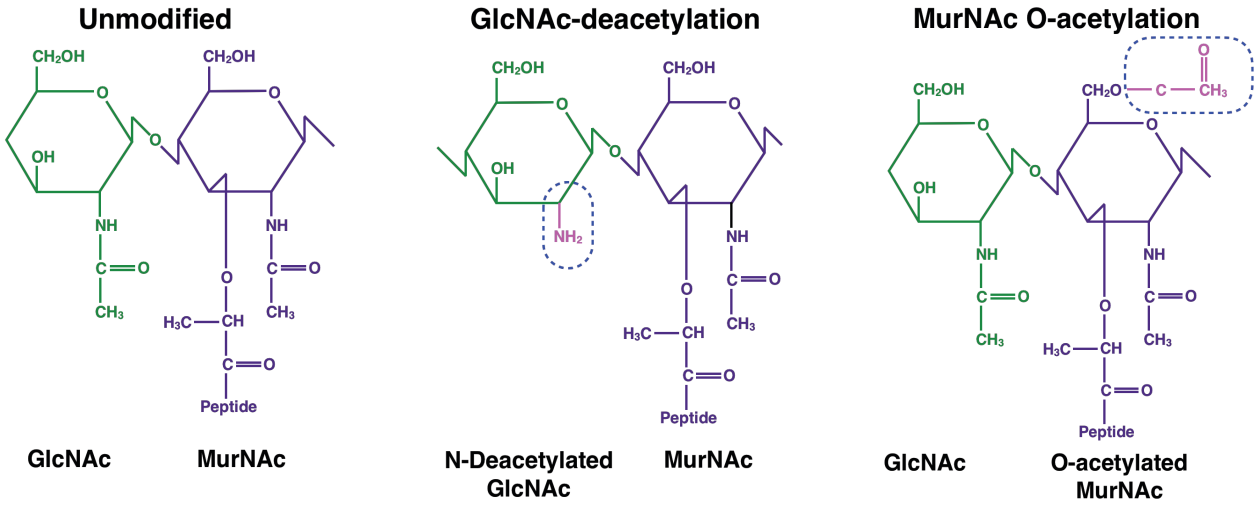


Figure 8. Modifications of the peptidoglycan strand
The unmodified PG disaccharide unit of β-1-4 linked GlcNAc (green) and MurNAc (Purple) residues is illustrated on the left. N-deacetylation of GlcNAc residues is shown in the central panel. *O*-acetylation of MurNAc residues is shown to the right. Chemical alterations are coloured in pink and circled in blue.

3.1.2 Peptidoglycan modifications and pathogenesis

As previously mentioned the fine structure of peptidoglycan varies between bacterial species. Moreover, certain modifications are common in pathogenic bacteria and are used as virulence factors to invade host defences. These modifications are typically generated in the glycan strand by post-synthetic reactions, including acetylation of the C6 hydroxyl group or deacetylation of the C2 amine group of GlcNAc and MurNAc residues. There are three main types of glycan strand modifications: *N*-deacetylation, *O*-acetylation and attachment of surface biopolymers, which will be discussed further.

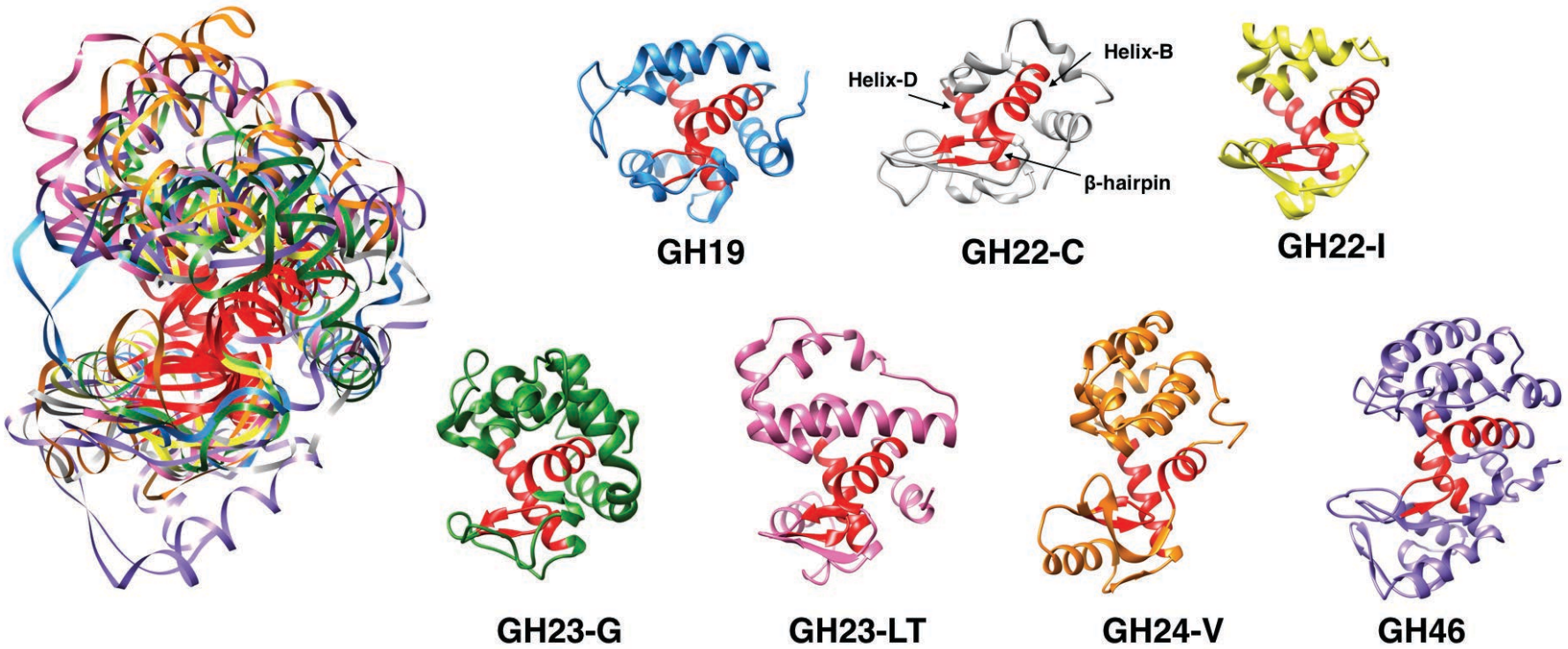


Figure 9. Structures of lysozyme superfamily members
The top left images shows the superimposition of seven representative structures of lysozyme superfamily members. Each representative is also displayed in an individual panel in the same relative orientation. The conserved core region (2 alpha helices and strict beta hairpin) is coloured in red in all images. GH19 –blue (PDB: 3W6D), GH22-C –gray (PDB: 2VB1), GH22-I – yellow (PDB:2DQA), GH23-G – green (PDB:154L), GH23-LT – pink (catalytic domain of PDB:1LSY), GH-24-V – orange (PDB:2LZM), GH46 – purple (PDB:1QGI). The structural elements of HEWL (GH22-C) are labelled for reference.

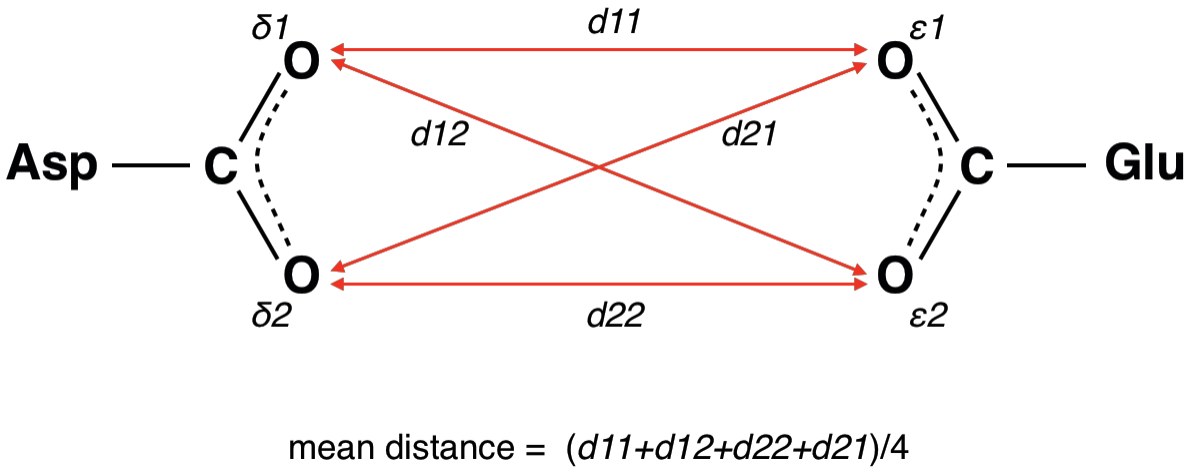


Figure 10. Catalytic distances
The average distance between the catalytic residues is calculated by taking the mean average distance between the carboxyl groups of the catalytic residues.

Beyond the active site, conserved residues contribute to the hydrophobic core of the enzymes. Bulky hydrophobic residues line helix-B, several residues towards the N-terminus of the catalytic glutamate. These hydrophobic residues form a hydrophobic pocket with conserved residues located on the invariant helix-D. Furthermore a strictly conserved glycine residue is located on strand-2, which maintains packing [141].

The stable folding pattern of lysozyme superfamily members is suggested to have originated from a common ancestor [149]. The invariant lysozyme core elements (helix-B, helix-D and two β-strands with a strict β-hairpin) alongside the conserved catalytic residues suggest the common fold did not arrive through convergent evolution. Moreover, the presence of non-essential elements that append the core structure suggest divergence from a common ancestor.

3.1.3.3 Catalytic mechanism

There are two reported catalytic mechanisms utilised by lysozymes, which have been determined using biochemical and structural data. Lysozymes and other GH family members are considered to either be retaining or inverting enzymes, depending on the product created after the glycosidic linkage is cleaved. Retaining enzymes use a double displacement mechanism that creates a covalent glycosyl-enzyme intermediate. The reaction retains the original stereochemistry at the anomeric centre. Retaining enzymes possess two acidic catalytic residues, where one carboxyl group acts as an acid/base while the other acts as a nucleophile.

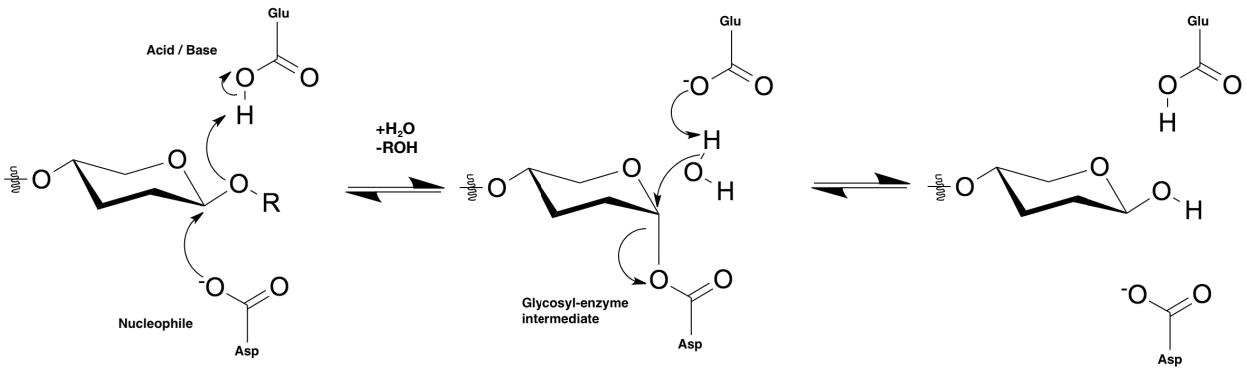


Figure 11. Retaining enzyme mechanism
Proposed mechanism for retaining lysozymes. ROH represents the leaving group.

GH-22 family member, HEWL, is the most well-characterised lysozyme in terms of structure, substrate binding and catalytic mechanism. The retaining lysozyme mechanism is exemplified by HEWL. The structure of HEWL has been solved in complex with various ligands revealing six sugar subsites, A-F [150]. HEWL possesses two catalytic residues, with E35 (helix-B) acting as a proton donor and D52 as a nucleophile [132]. It was initially suggested from structural evaluation that D52 in HEWL played a

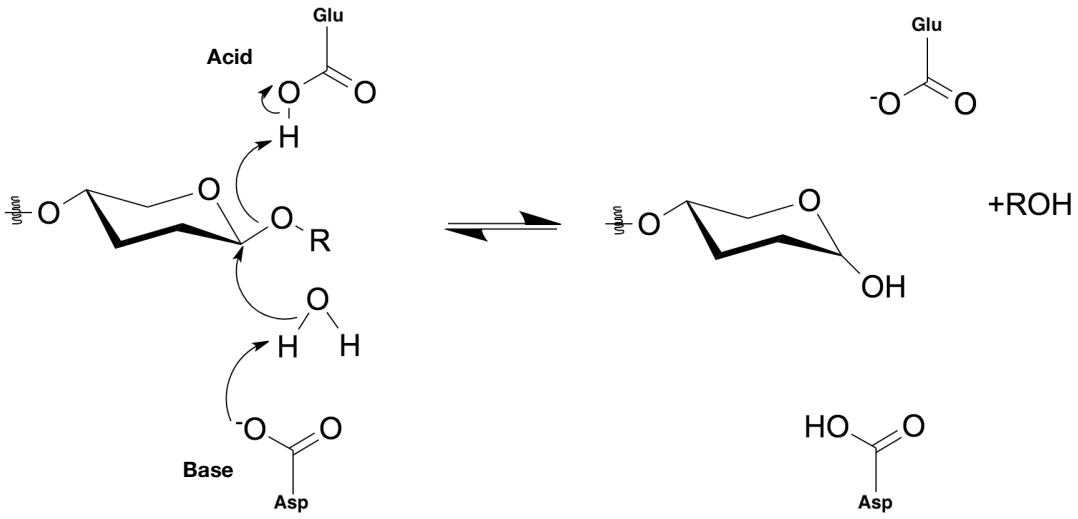


Figure 12. Inverting enzyme mechanism
Proposed mechanism for inverting lysozymes. ROH represents the leaving group.

3.1.3.4 Lytic transglycosylases – housekeeping enzymes

LTs are peptidoglycan-degrading enzymes that fulfil a housekeeping role. LTs are integral to PG maintenance, turnover and remodelling during cell division and cell growth [154]. Like classical lysozymes they cleave the β-1-4 linkage between MurNAc and GlcNAc of PG. However, they do not perform catalysis through hydrolysis of the substrate but rather use an unusual intra-molecular glycosltransferase reaction, which produces a terminal 1,6-anhydro-MurNAc residue [117]. Catalytically LTs perform quite a different reaction to classical lysozymes. LTs possess the invariant catalytic glutamate located in the ES-motif on the central α-helix. However, they lack a second catalytic residue and do not require a catalytic water molecule. Instead LTs use a substrate-assisted catalytic mechanism [155]. The catalytic glutamate is proposed to serve initially as the general acid, donating a proton to the glycosidic linkage and forming a muramoyl oxazolinium-ion intermediate. The glutamate then switches roles and acts as a base to abstract the C6 hydroxyl proton, leading to the formation of the 1,6- anhydromuramoyl product [154].

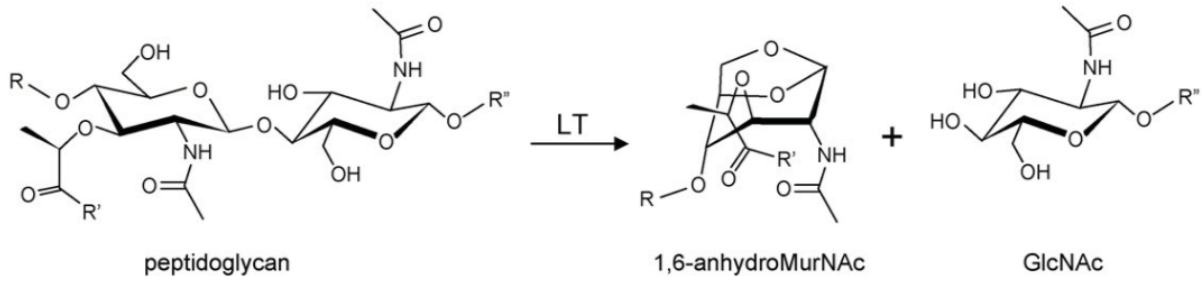


Figure 13. Lytic Transglycosylase reaction.
Lytic Transglycosylases catalyse the cleavage of the β-1-4 linkage between MurNAc and GlcNAc of PG with a single catalytic residue. Cleavage by LTs results in the formation of a 1.6-anhydroMurNAc terminal residue.

LTs can be divided into four families (1-4) and several *E. coli* LTs have been structurally and biochemically characterised for families 1 and 2 [156]. LTs are common in Gram-negative bacteria and always found in the periplasm. A small number of LTs have been reported in Gram-positive bacteria [157]. LTs can exist as soluble species or attached to the outer membrane via the N-terminus. Structures of the LT catalytic domain adopt a common fold that resembles G-type lysozymes [158].

3.1.4 *Bdellovibrio* lifecycle and PG modifications

The obligate predatory bacterium, *Bdellovibrio bacteriovorus*, has adapted PG modifying enzymes to facilitate its unusual biphasic predatory lifecycle. Early investigation into predation by *Bdellovibrio* highlighted a number of prey cell wall modifications required for successful completion of the predatory lifecycle and hinted at the enzymes responsible.

During penetration of an *E. coli* prey, *Bdellovibrio* rapidly solubilises approximately 10-15% of the prey peptidoglycan in a local, controlled manner [45]. Two enzyme

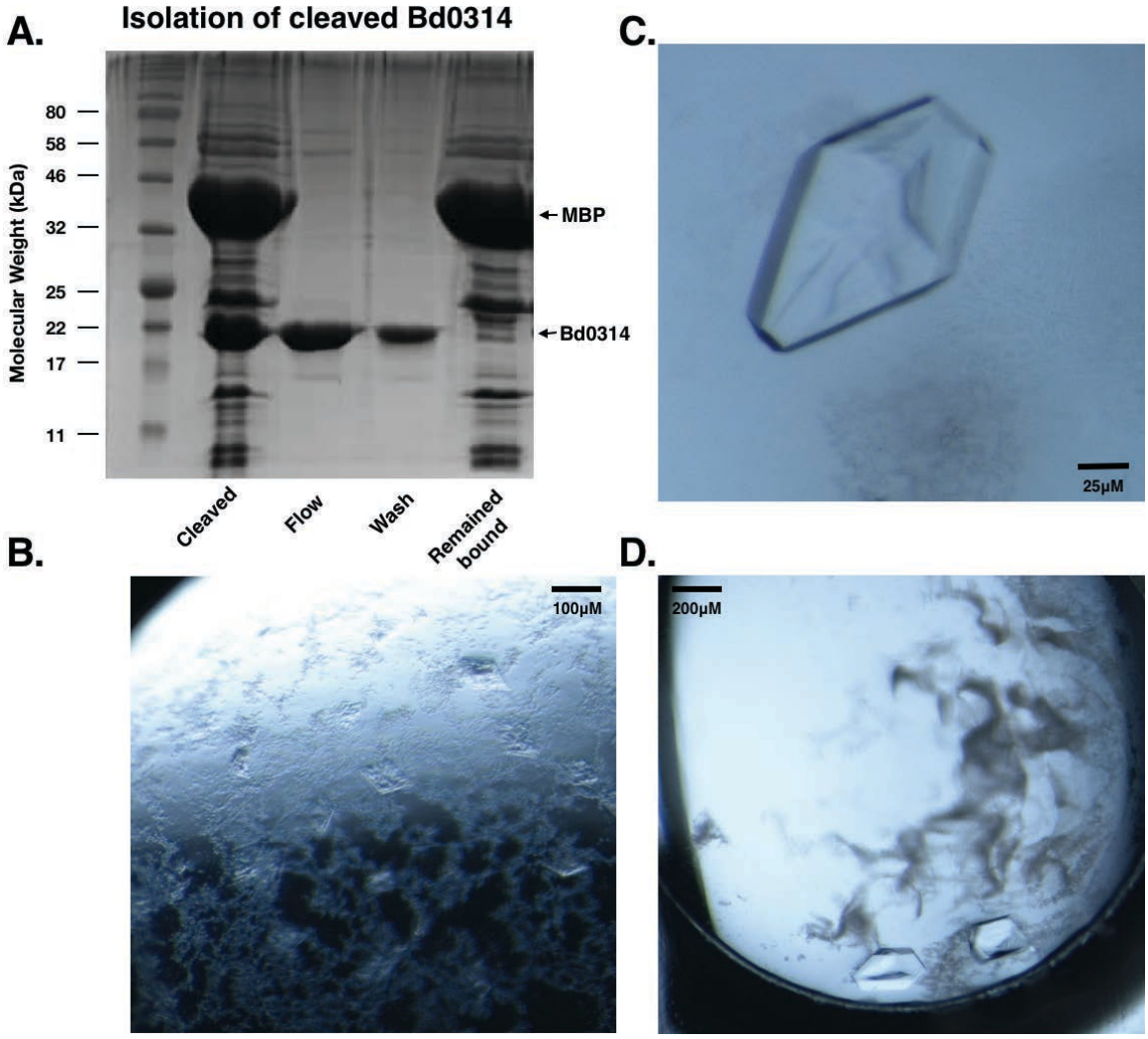


Figure 14. Bd0314 purification and crystallisation
A. An MBP tag was used to express and purify Bd0314. The MBP tag was cleaved by incubation with TEV and removed by a second passage over a HisTrap column. **B.** Bd0314-Wt crystals were small thin plates. **C.** The Bd0314-E143Q crystal used for SSAD phasing was optimised. **D.** Bd0314-Wt crystals that grew in a different space group.



Figure 15. Bd0314 structure schematic
Bd0314 is a 254 residue protein that has a 180 residue catalytic domain. Bd0314 primary sequence suggests the presence of an N-terminal signal peptide (SP) containing a Lipobox (LB). Between the predicted lipobox and the catalytic domain is a 50 residue region of predicted disorder. Secondary structure predictions made using the XtalPred server.

bonds. E175 located on helix G hydrogen bonds to S159, which stabilises the end of the β-sheet and turn. In addition Bd0314 may contain a third β-strand, where I172 (located immediately after the β-hairpin) and R167 of strand-2 make a β-sheet-like interaction.

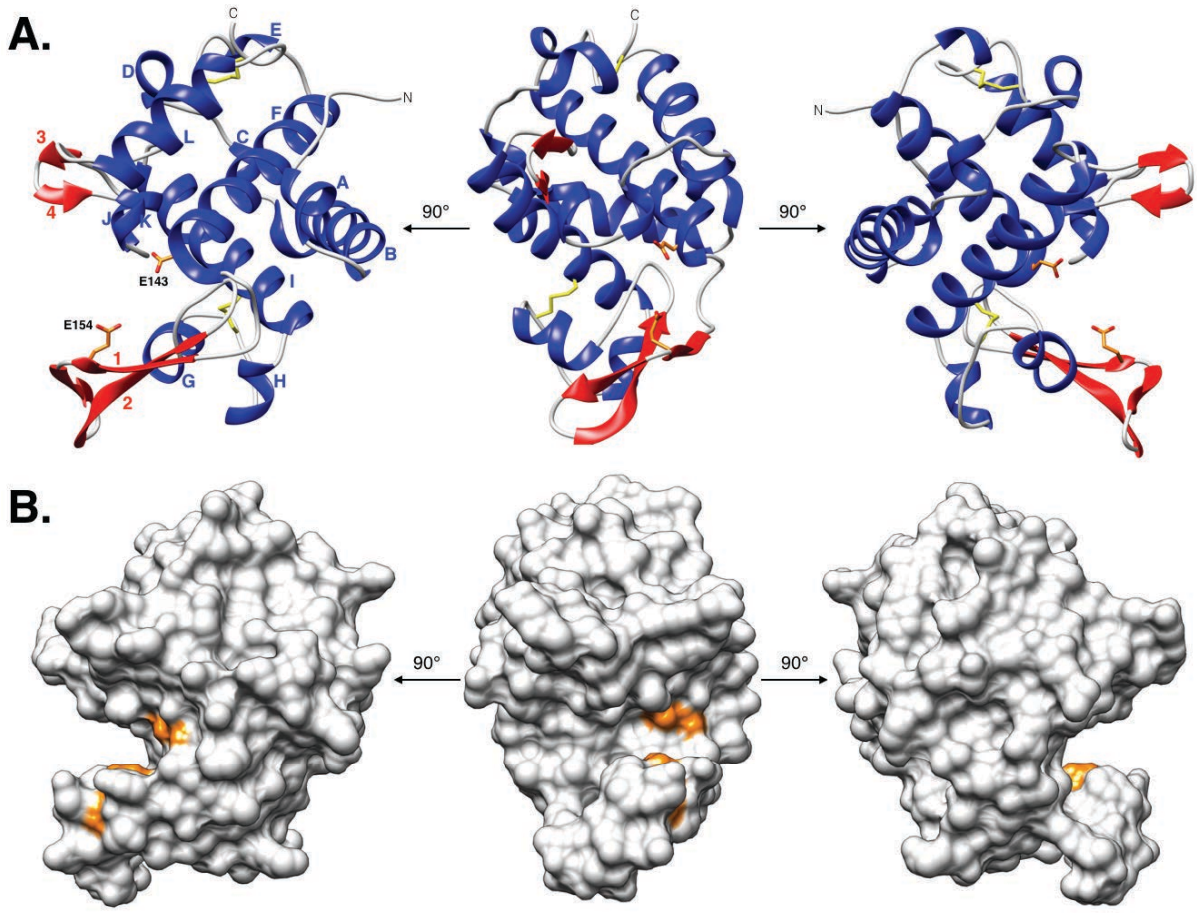


Figure 16. Bd0314 structure
A. The secondary structure of Bd0314 shown from three angles 90° apart. α-helices are coloured blue and β-strands are coloured red. Disulphide bonds are highlighted in yellow and the positions of the catalytic E143 and E154 are displayed in orange. Helices are labelled with blue letters and strands are labelled with red numbers in the first panel. **B.** Surface view of Bd0314 illustrating the active-site cleft. The location of the catalytic residues is highlighted in orange.

3.2.2.2 The catalytic site

Analysis of the Bd0314 structure revealed two potential catalytic residues. The classical ES invariant glutamate (E143) is located on the end of helix-F facing into the catalytic site. The second (E154) is located on strand-1, and is projected into the active site, directly opposite E143. The typical location of the second catalytic Asp in the lysozyme superfamily is on strand-2. However, in Bd0314 S166 occupies this location and the hydroxyl group makes a hydrogen bond to the E154 carboxyl-group, which may help to retain the orientation of E154. The presence and location of E143 and E154 suggest they act as a general acid-base pair. The distance between the two acidic residues is 8Å, suggesting that Bd0314 proceeds by an inverting catalytic mechanism [147].

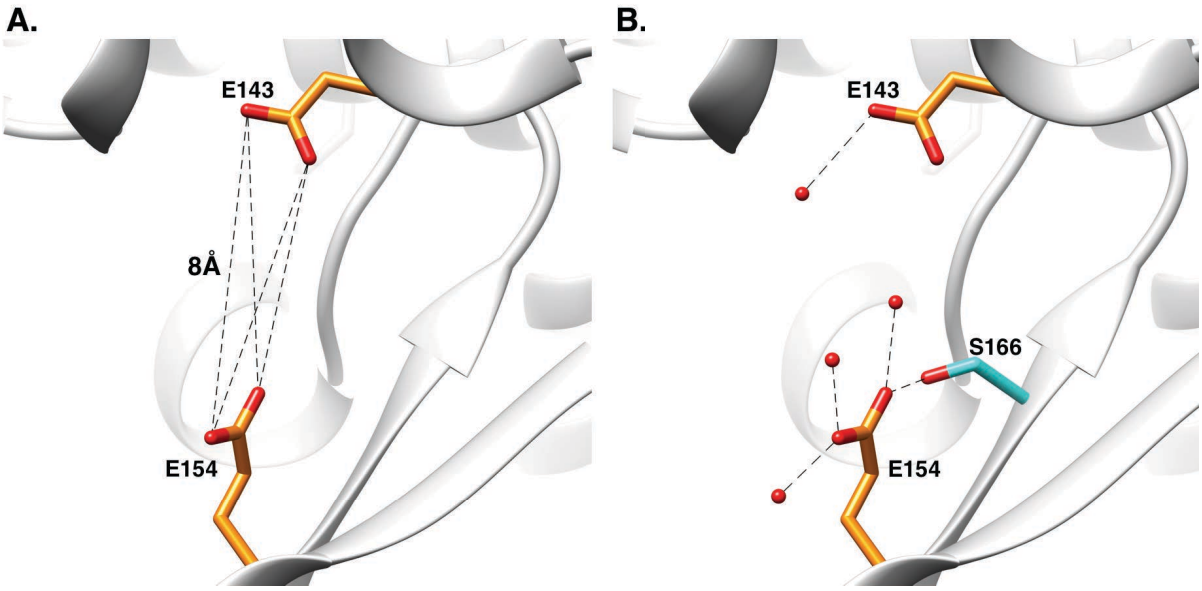


Figure 17. Bd0314 catalytic residues
A. The distance between the carbonyl groups of E143 and E154 is ~8Å, which suggests Bd0314 is an inverting lysozyme. **B.** E143 and E154 make hydrogen bonds with water molecules in the active-site cleft (potential catalytic water molecules). E154 also makes a hydrogen bond with S166, which may have stabilising effects during catalysis.

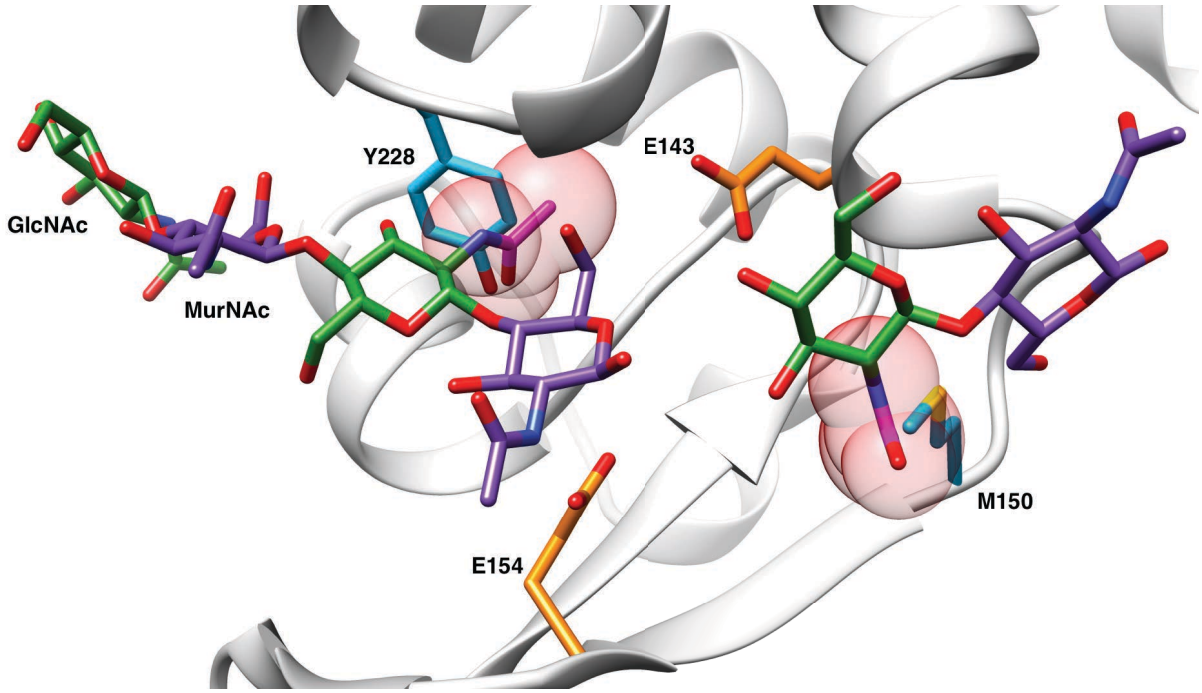


Figure 18. Bd0314 substrate-binding site
Model of Bd0314 substrate binding, generated by superimposition of a regular lysozyme (PDB:1LZS) in complex with a hexa-*N*-acetyl-chitohexaose substrate. Bd0314 was structurally aligned with PDB:1LZS by least-squares minimization of the structurally conserved regions (helix-F, helix-I and active site beta-hairpin) using the CCP4 suit superpose software (length 2474 atoms from RSMD=1.13Å). Bd0314 is displayed in ribbon form, coloured in white. The ligand is alternatively coloured green (GlcNAc) and purple (MurNAc), to represent the positions of GlcNAc and MurNAc for a peptidoglycan substrate similarly to that in Figure 8. The substrate GlcNAc residue acetyl-group faces inwards towards the cleft, whereas MurNAc residues face outwards to accommodate the peptide stem. The catalytic glutamate residues (E143 and E154) of Bd0314 are shown in orange and are correctly positioned to cleave after the MurNAc residue. The acetyl-groups of GlcNAc residues clash with residues of the peptide-binding cleft. The residues that clash with the acetyl-group of GlcNAc residues (Y228 and M150) are shown in cyan. The red van der waals surface of the acetyl groups demonstrates the degree of clash.

3.2.2.3 Substrate selection

Bd0314 has an extensive substrate-binding cleft provided by the extended β-sheet, which forms the base and the additional β-sheet region of the core domain, which creates a peak to the roof of the binding cleft. We have been unable to obtain a protein:ligand complex after many attempts with various ligands. However, numerous structures of lysozyme-superfamily members in complex with ligands have been published [153, 162, 163]. Superimposition of these ligand complex structures with Bd0314 (using LSQ superimposition of helix-F (residues 133-144), helix-I (residues

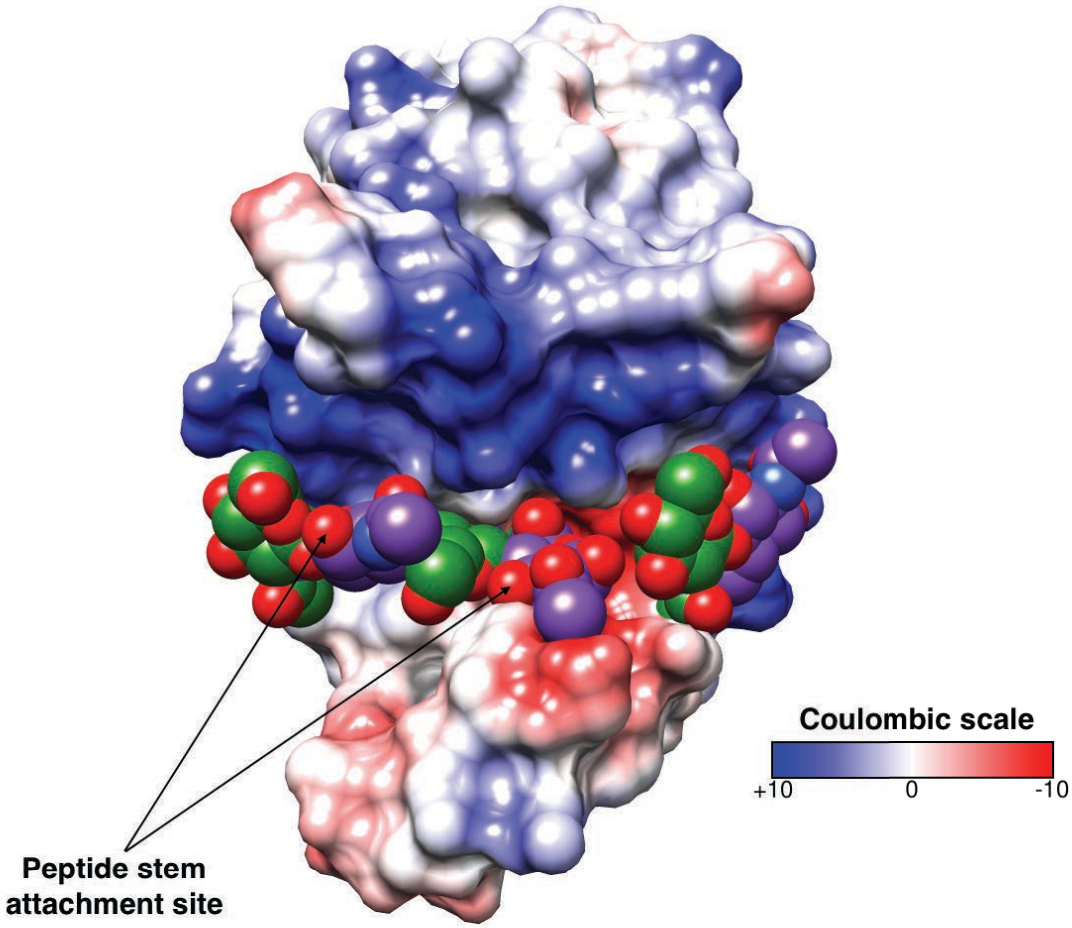


Figure 19. Substrate binding surface view
The substrate-binding model shown in the surface view demonstrates how the acetyl-groups of GlcNAc residues face inwards whereas the acetyl-group of MurNAc residues face outwards, so the peptide stem can accommodated. The roof of the substrate-binding cleft carries a positive charge. Substrate colours are the same as Figure 8.

3.2.2.4 Bd0314 structure comparison (DALI) and homologues

The DALI protein structure comparison server was used to identify proteins with similar structural elements from the PDB [160]. The top 25 matches from DALI are shown in Table 11. The greatest proportion of sequence and structural conservation for Bd0314 occurs for the protein core and the catalytic cleft. This region includes (residues 130-154) the central helix-F and part of the strand-1 and (residues 164-207) part of strand-2 up to and including central helix-I (Figure 20). Lytic transglycosylase

family proteins (GH23), are the most frequent structural matches. The catalytic core domain of Slt70 PDB: 1QTE (a soluble LT) gives the best structural match. However, the lower subdomain containing the second catalytic residue is not present in Slt70. Interestingly human lysozyme PDB:5LSH, is one of the closest structural matches and has a high Z-score, low rmsd. Four regions of Bd0314, residues 1-14, 26-54, 82-90 and 143-150, show little structural conservation to other lysozyme structures or primary sequences (Figure 28.B).

No:	Chain	Z	rmsd	lali	nres	%id	PDB	Description
1:	1qte-A	8.8	2.8	119	620	18	PDB	MOLECULE: SOLUBLE LYTIC TRANSGLYCOSYLASE SLT70;
2:	6cf9-A	8.8	2.7	112	516	13	PDB	MOLECULE: LYTIC TRANSGLYCOSYLASE;
3:	5lsh-A	8.4	2.6	103	130	15	PDB	MOLECULE: LYSOZYME C;
4:	4c5f-A	8.2	2.9	116	330	18	PDB	MOLECULE: MEMBRANE-BOUND LYTIC MUREIN
5:	4owd-A	8.1	3.2	118	429	16	PDB	MOLECULE: MEMBRANE-BOUND LYTIC MUREIN
6:	5vas-A	8.1	2.6	101	129	18	PDB	MOLECULE: LYSOZYME;
7:	6fcu-A	8.0	2.9	120	615	16	PDB	MOLECULE: SOLUBLE LYTIC MUREIN TRANSGLYCOSYLASE;
8:	4yim-B	7.9	2.8	116	586	13	PDB	MOLECULE: PUTATIVE SOLUBLE LYTIC MUREIN
9:	3bkh-A	7.8	3.0	113	268	10	PDB	MOLECULE: LYTIC TRANSGLYCOSYLASE;
10:	4dq7-A	7.7	3.1	108	168	12	PDB	MOLECULE: MEMBRANE PROTEIN PHI6 P5;
11:	3w6b-A	7.7	3.4	119	160	22	PDB	MOLECULE: LYSOZYME-LIKE CHITINOLYTIC ENZYME;
12:	2xqo-A	7.6	3.0	125	210	16	PDB	MOLECULE: CELLULOSOME ENZYME, DOCKERIN TYPE I;
13:	1lsg-A	7.4	2.6	101	144	21	PDB	MOLECULE: HEN EGG WHITE LYSOZYME;
14:	4xp8-A	7.0	2.9	84	88	19	PDB	MOLECULE: ETGA PROTEIN;
15:	1am7-A	6.8	3.6	102	154	7	PDB	MOLECULE: LYSOZYME;
16:	2dqa-A	6.7	2.9	99	123	13	PDB	MOLECULE: LYSOZYME;
17:	5e27-A	6.7	2.7	74	226	12	PDB	MOLECULE: RESUSCITATION-PROMOTING FACTOR RPFB;
18:	2dkv-A	6.0	4.0	140	286	10	PDB	MOLECULE: CHITINASE;
19:	5xsv-A	6.0	3.2	121	486	9	PDB	MOLECULE: CHITINASE;
20:	5nm7-G	5.9	3.1	117	263	12	PDB	MOLECULE: PEPTIDOGLYCAN-BINDING DOMAIN 1;
21:	4ok7-A	5.9	3.3	126	223	10	PDB	MOLECULE: ENDOLYSIN;
22:	4luq-A	5.6	3.5	125	409	10	PDB	MOLECULE: UNCHARACTERIZED PROTEIN;
23:	4g9s-A	5.4	3.2	106	187	16	PDB	MOLECULE: GOOSE-TYPE LYSOZYME;
24:	4fdy-A	4.9	2.9	98	295	10	PDB	MOLECULE: SIMILAR TO LIPOPROTEIN, NLP/P60 FAMILY;
25:	1d0k-A	4.9	3.2	111	318	14	PDB	MOLECULE: 35KD SOLUBLE LYTIC TRANSGLYCOSYLASE;

Table 11. DALI results
Top 25 matches from the DALI server for Bd0314 coordinates as the search query [160].

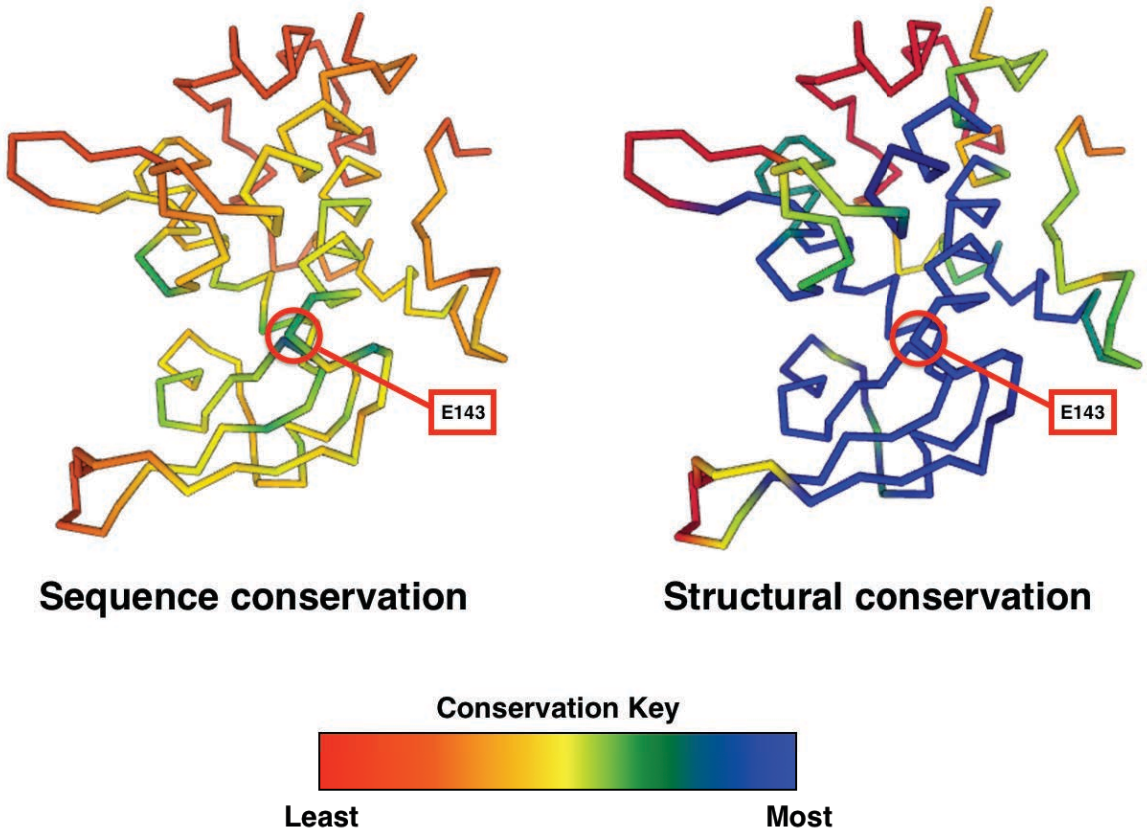


Figure 20. Bd0314 conservation
The sequence and structural conservation (coloured key) is mapped onto the Cα-trace of Bd0314 structure. Sequence conservation in Bd0314 is low, whereas the structural conservation is higher in regions. The core region of Bd0314 has the highest sequence and structural conservation. These results are obtained the DALI [160].

BLAST analysis of the Bd0314 primary sequence against a non-redundant database revealed a number of closely related homologues. Included in the results were two *Bdellovibrio bacteriovorus HD100* proteins (Bd1413 and Bd1440). Bd1413 and Bd1440 share 60% ID to each other and 33% ID to Bd0314 (Figure 21). Bd1413 and Bd1440 both possess the two catalytic residues highlighted in the structure of Bd0314 and they also possess the YW-motif (Y228) associated with GlcN selectivity. There are two regions of minor disparity in the sequence alignment between Bd0314 and Bd1413/Bd1440. Firstly Bd1413/Bd1440 has a 11aa insertion between helices F and G. Secondly the short β-sheet extension (strand-3 and 4) of the core domain exists as an 8aa insertion in the primary sequence of Bd0314.

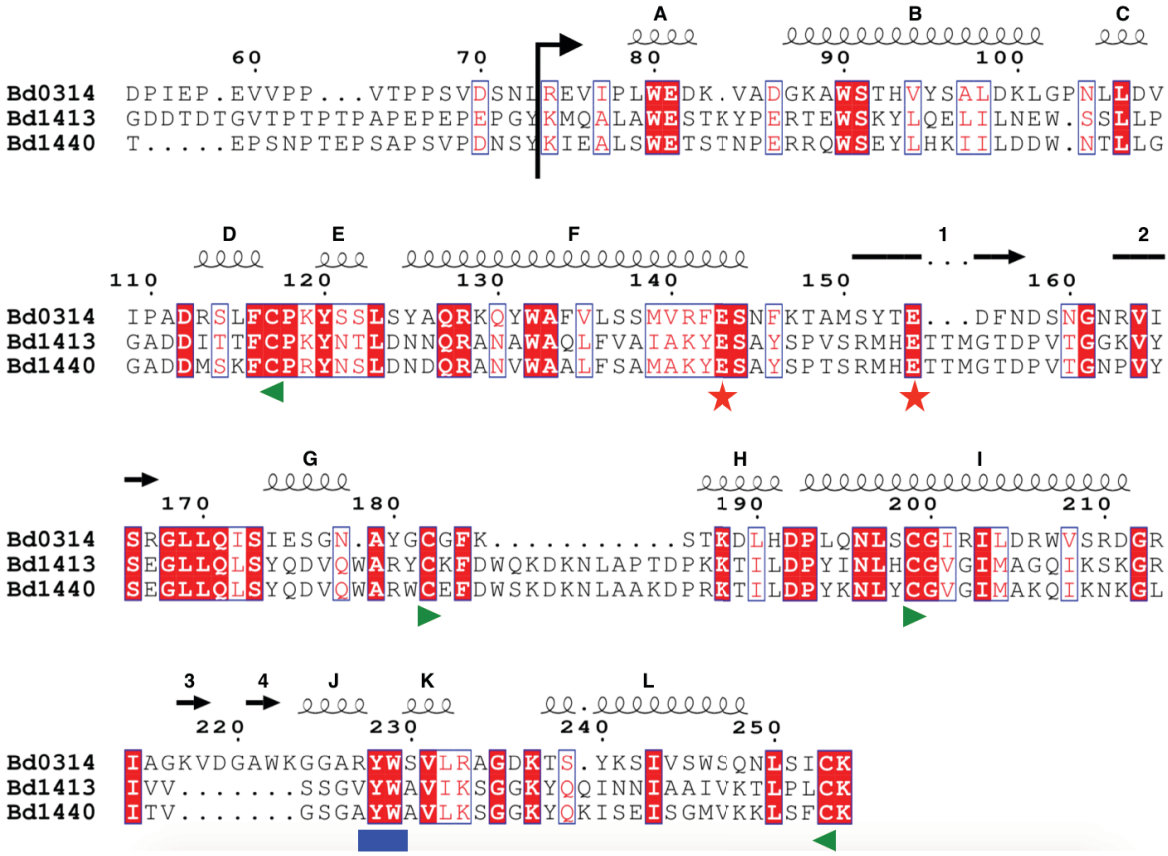


Figure 21. Bd0314 alignment with *Bdellovibrio HD100* homologues
Alignment of Bd0314 primary sequence with Bd1440 and Bd1413. Bd1440 and Bd1413 sequences are more closely related to one another (60% ID) than to Bd0314 (33% ID). Bd1440 and Bd1413 do not possess the second short β-sheet region of the α-helical domain. Bd1440 and Bd1413 also have an elongated region between helix-G and helix-H. The start of the catalytic domain is shown by a black line and arrow. The conserved cysteine residues are shown by green triangles; the direction of the green triangle represents the cysteine pairs. Red stars highlight the catalytic residues and the GlcNAc *N*-acetyl group blocking motif is shown by the blue rectangle. α-helices are represented by black hoops and β-sheets represented by black arrows. Drawn using Espript 3 [161].

BLAST analysis also revealed a number of Bd0314 homologues from other organisms that possess equivalents of both catalytic residues (E143 and E154) as well as the YW-motif (Y228). An alignment of selected homologues is shown in Figure 22. The alignment demonstrates the co-carriage of the second catalytic residue motif (TED) with the GlcN specificity motif (YW).

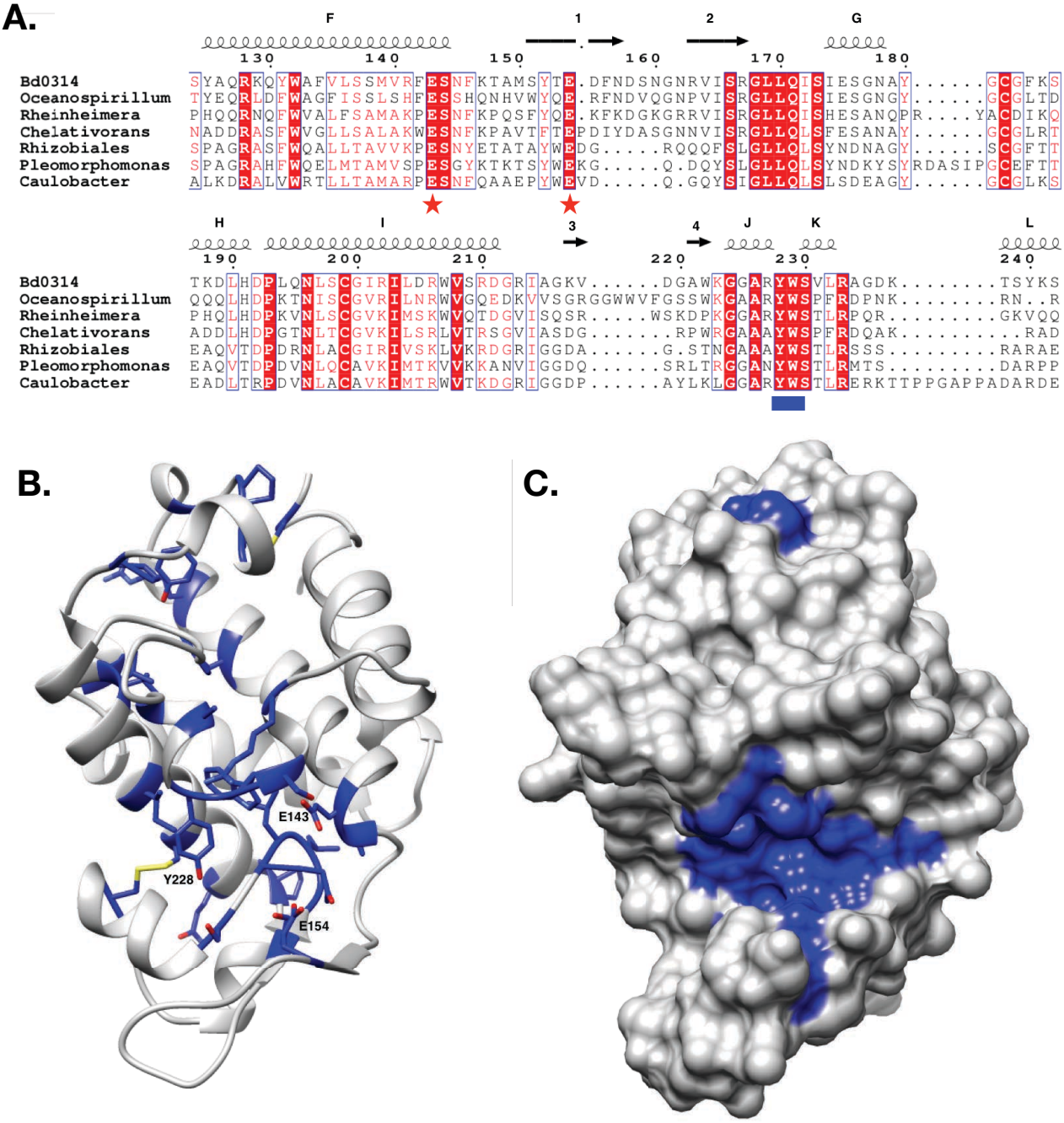


Figure 22. Bd0314 alignment and residue conservation
A. BLASTp searches reveal that homologues of Bd0314 exist in organisms other than predatory bacteria. All organisms possess CE4 deacetylases in their genomes. The presence of the second catalytic glutamate (E154) has a 100% co-carriage with the *N*-acetyl group blocking tyrosine (Y228) in these predicted *N*-deacetylation specific lysozymes. Catalytic residues are highlighted with blue stars and the *N*-deacetylation specific lysozyme motif is highlighted with a blue bar. **B.** Residue conservation (100% ID) mapped onto the secondary structure of Bd0314. **C.** Residue conservation mapped onto the surface of Bd0314.

upon the GlcN substrate. The Bd0314 activity towards the GlcN substrate is comparable to the activities of the HEWL and Mutanolysin upon acetylated substrates. This result indicates Bd0314 possesses a high specificity towards GlcNAc-deacetylated PG substrates.

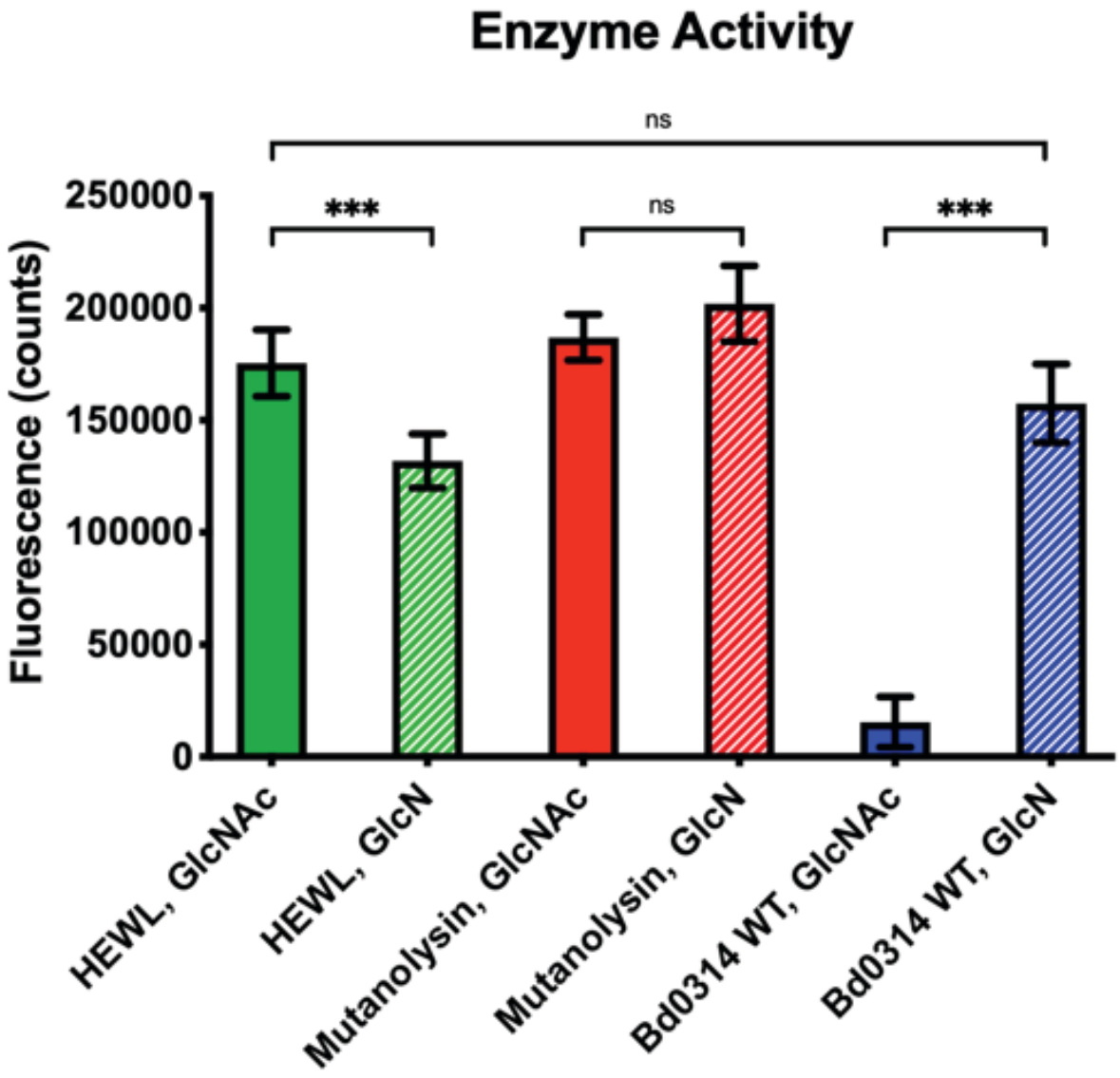


Figure 23. Bd0314 specificity
The results of the FITC-PG assay demonstrating Bd0314 specificity. Block coloured bars represent GlcNAc (acetylated PG) whilst striped bars represent GlcN (N-deacetylated PG). HEWL (Green) shows reduced activity with N-deacetylated substrate. Mutanolysin activity is similar for both substrates. Bd0314 shows specificity for the GlcN substrate. All data points were measured in at least triplicates. The error bars represent the standard deviation. Unpaired t-test was used to determine significance, *** = $p<0.0001$, ns = not significant.

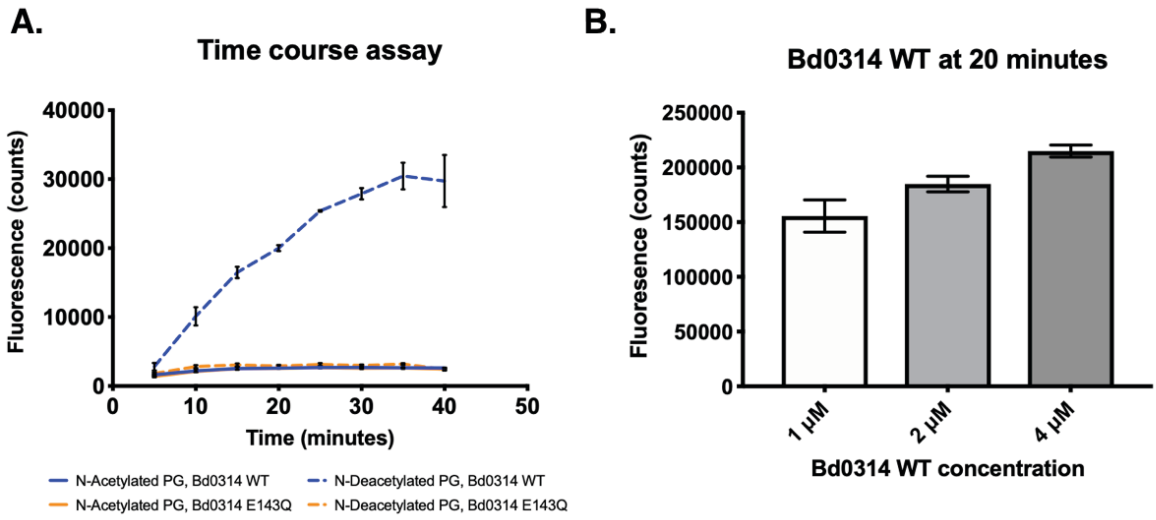


Figure 24. Time and Concentration dependent assay
A. Turnover of the FITC-PG substrate by Bd0314 is time dependent; activity is linear for the first 15 minutes. Bd0314 is only active towards the *N*-deacetylated PG substrate and shows minimal activity towards normal acetylated PG substrate. **B.** The FITC-PG assay is concentration dependent; increased concentrations of Bd0314 increased the fluorescence readings. All data points were measured in at least triplicates. The error bars represent the standard deviation.

3.2.3.2 Catalytic residues

The FITC-PG assay was used to confirm whether E143 and E154 (general acid-base pair identified from the Bd0314 structure) were required for catalysis. The glutamate residues were mutated to glutamines thereby changing the chemistry of the residue but maintaining the structure (Bd0314-E143Q and Bd0314-E154Q). Figure 25 shows that the activity of both Bd0314-E143Q and Bd0314-E154Q are greatly reduced compared to Bd0314. This finding confirms both E143 and E154 are required for catalysis.

Y228 was mutated to an alanine residue (Bd0314-Y228A) to investigate its role in substrate-binding and specificity. Bd0314-Y228A showed a reduced activity towards the GlcN substrate but no increase in activity towards the GlcNAc substrate. This

finding suggests that Y228 also participates in substrate recognition by direct interaction with the free amine group of GlcN residues (Figure 30). The retained substrate specificity of Bd0314-Y228A suggests other features also contribute to substrate selection. M150 was identified as another substrate selective residue, although we were unable to successfully express and purify a Bd0314-M150A mutant.

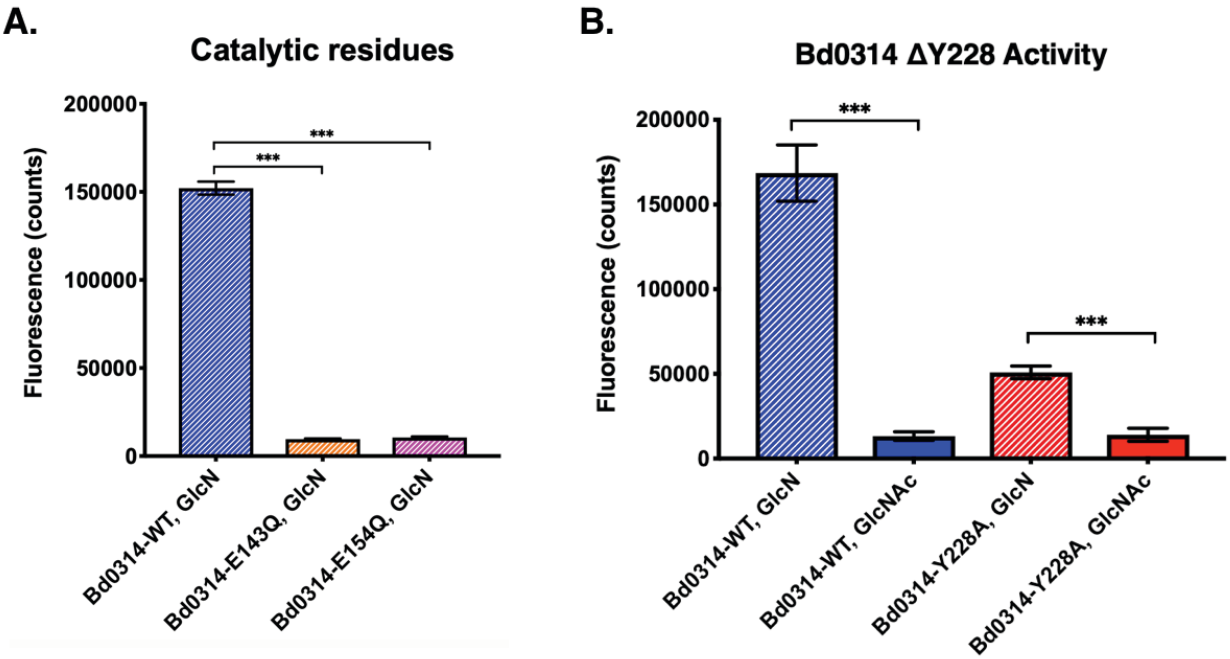


Figure 25. Bd0314 active site residues
A. Mutation of either E143 or E154 abolishes Bd0314 activity, suggesting both residues are involved in catalysis and supporting an inversion catalytic mechanism, involving two glutamate residues. **B.** Y228A mutation reduces Bd0314 activity towards *N*-deacetylated PG. This finding suggests Y228 is involved in substrate binding. Y228A activity towards GlcNAc PG does not alter. All data points were measured in at least triplicates. The error bars represent the standard deviation. Unpaired t-test was used to determine significance, *** = $p<0.0001$, ns = not significant.

3.2.3.3 Bd0314 homologue activities

Bd1440 (S75 – K266) and Bd1413 (G80 – K271) constructs were created to investigate the activities of Bd0314 *Bdellovibrio* homologues using the FITC-PG assay (Figure 26). Bd1440 and Bd1413 both possess the same high specificity towards the GlcNAc-deacetylated substrate (GlcN). Bd0314, Bd1440 and Bd1413 all liberated a similar quantity of FITC, demonstrating similar activities.

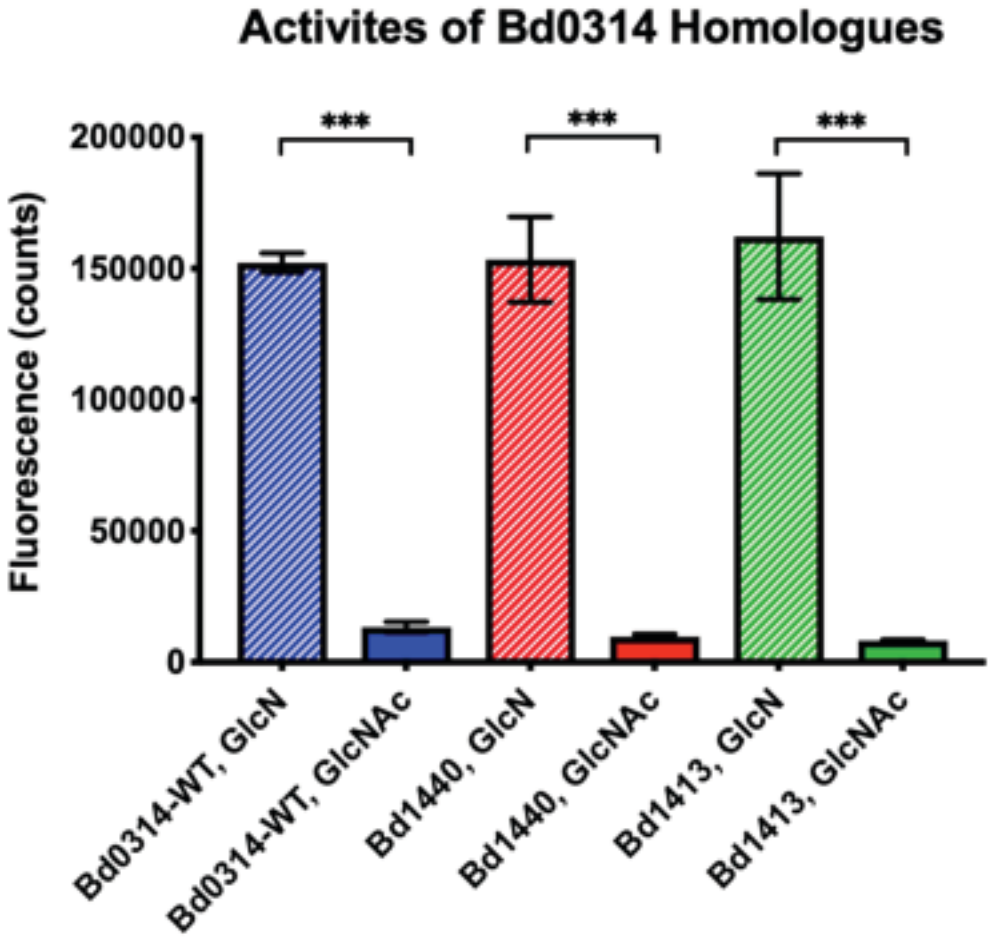


Figure 26. Bd0314 HD100 homologue activities
The activity of Bd0314 HD100 homologues (Bd1413/1440) compared to Bd0314 activity. Bd0314-WT in blue, Bd1440 in red and Bd1413 in green, with block coloured bars representing acetylated GlcNAc substrate and striped coloured bars representing *N*-deacetylated substrate (GlcN). The activities of Bd1440 and Bd1413 also demonstrate substrate specificity for GlcN substrate and have minimal activity with GlcNAc substrate. Bd1413/1440 have a similar activity as Bd0314. All data points were measured in at least triplicates. The error bars represent the standard deviation. Unpaired t-test was used to determine significance, *** = $p<0.0001$, ns = not significant.

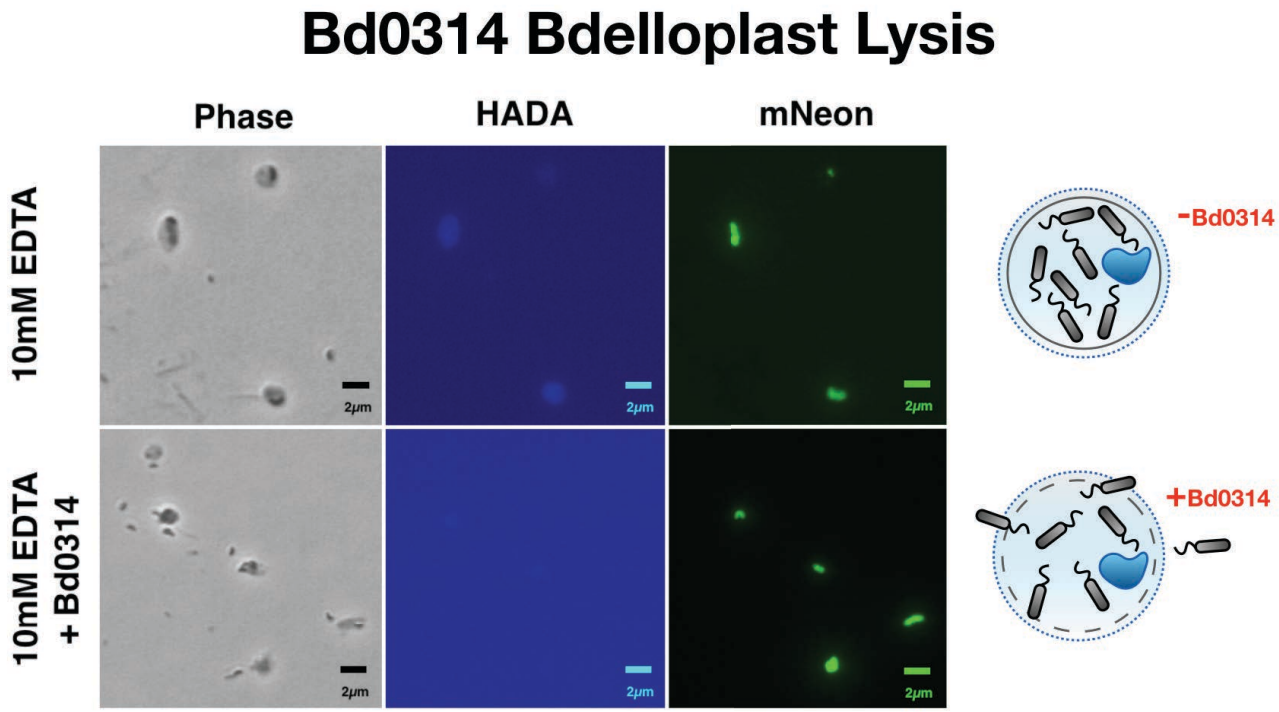


Figure 27. Bd0314 Bdelloplast Lysis
E. coli cells were incubated with *Bdellovibrio bacteriovorus* HD100 cells resulting in >99% prey rounding into bdelloplasts. Bdelloplasts were incubated for 1 hour with 10 mM EDTA +/- 10 μM Bd0314. Cells were immobilised and imaged on 1% agarose Ca/HEPES pads. HADA was used to label the cell wall. pMAL::mNeon was used to label the periplasm and cytoplasm of *E. coli.* Bdelloplasts incubated with Bd0314 lyse and release *Bdellovibrio* progeny when compared to EDTA alone control. Furthermore bdelloplasts incubated with Bd0314 have diminished levels of HADA labeled peptidoglycan, demonstrating that Bd0314 is active towards this material. pMAL::mNeon labeling allows identification of lysed bdelloplasts. The far right hand cartoon illustrates the effect Bd0314 incubation has on the bdelloplast structures.

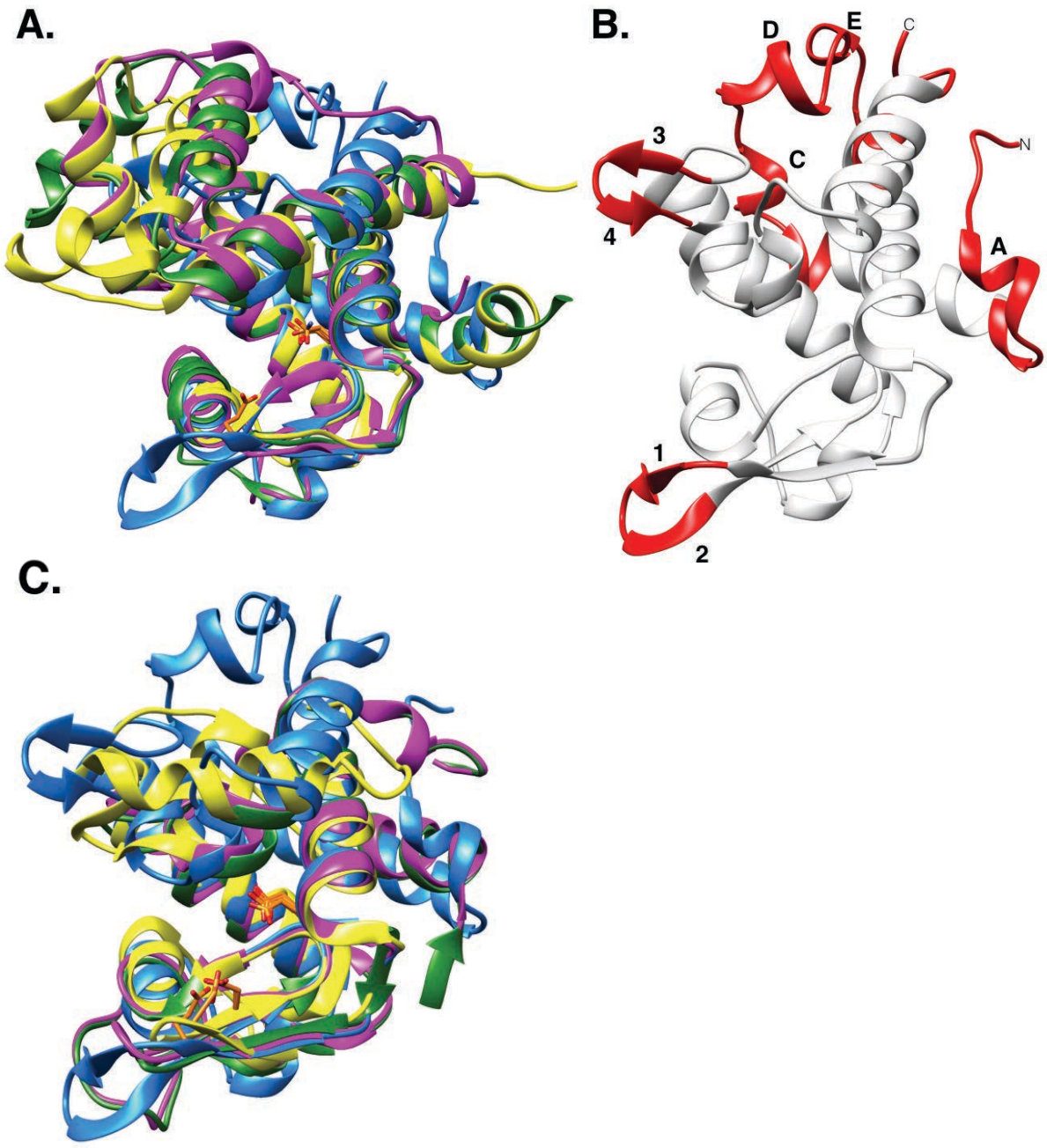


Figure 28. Bd0314 structure comparison
A. Bd0314 (blue) superimposed with the catalytic domains of lytic transglycosylases; MltC = PDB:4CFP (green, length = 5525 atoms, rmsd = 0.88Å); MltF = PDB:4P0G (yellow, length = 4162 atoms, rmsd = 0.88Å); Slt70 = PDB:1SLY (magenta, length = 5000 atoms, rmsd = 0.83Å). **B.** Unique structural features of Bd0314 are shown in red. **C.** Bd0314 (blue) superimposed with GH22 lysozymes; GH22-C HEWL = PDB:1LZS (green, length = 2474 atoms, rmsd = 1.13Å); GH22-I Tapes japonica lysozyme = PDB:2DQA (yellow, length = 2511 atoms, rmsd = 0.59Å); GH22-C Human lysozyme = PDB:5LSH (magenta, length = 1414 atoms, rmsd = 1.03Å). Superimposition was performed by least squares minimisation using the CCP4 suit superpose software (alignment of conserved secondary structure elements, helix-F, helix-I and beta-hairpin).

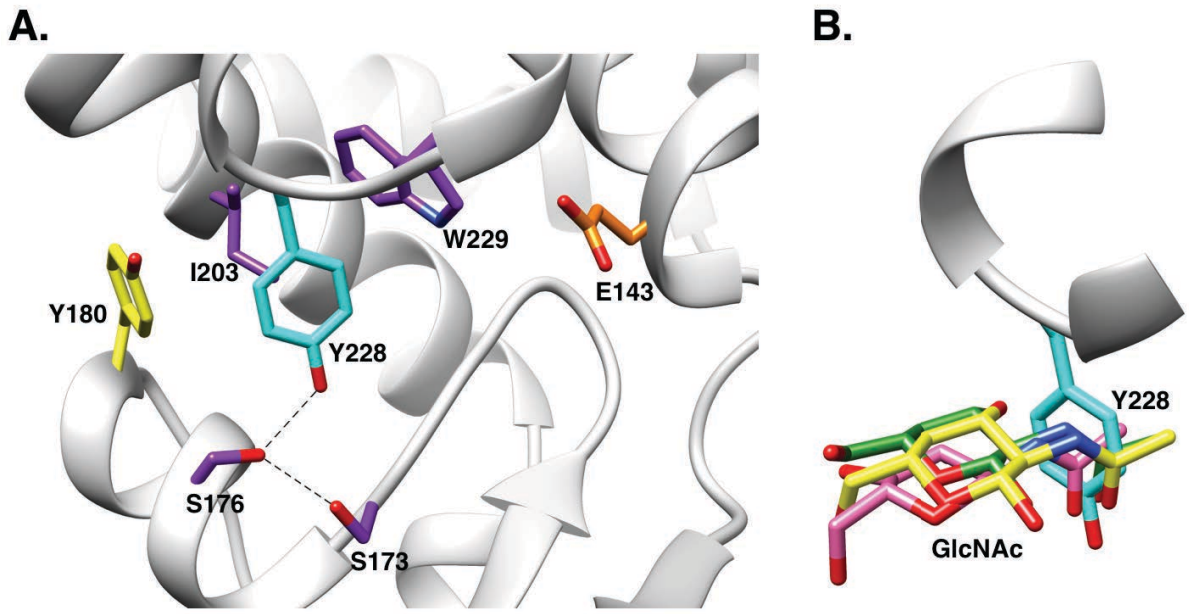


Figure 29. Bd0314 substrate binding and selection
A. Y228 (cyan) is positioned facing into the active site and held in position by a hydrogen bond with S176 (black line) and by hydrophobic interactions with W229 and I203. Y180 (yellow) may participate in substrate binding by interacting with the MurNAc hydroxyl-group. **B.** Bd0314 was superimposed with 1LZS (yellow), 3W6D (pink) and 3GXR (green) – lysozyme:ligand complex structures. The acetyl group of GlcNAc residues in these superimposed models always clashed with Y228 of Bd0314.

M150 also occupies the same space as the acetyl-group of GlcNAc residues and could also participate in selection. However, M150 is not conserved in predicted deacetylation-specific lysozymes whereas Y228 is (Figure 21). Also M150 is not locked into one conformation and in the structure of Bd0314 was observed in multiple conformations suggesting that it could easily be repositioned during substrate binding. The ligand binding models also highlight other residues that may be involved in substrate binding. For instance Y180 is positioned in close proximately to the hydroxyl group of the MurNAc residues and may make a hydrogen bond. Similarly S176 that interacts with Y228 may share the H-bond to the GlcN amine group.

Chitosanases (GH46) are a member of the lysozyme superfamily and hydrolyse the β-1-4 linkage between poly-GlcN [145]. Chitosanases are structurally divergent from

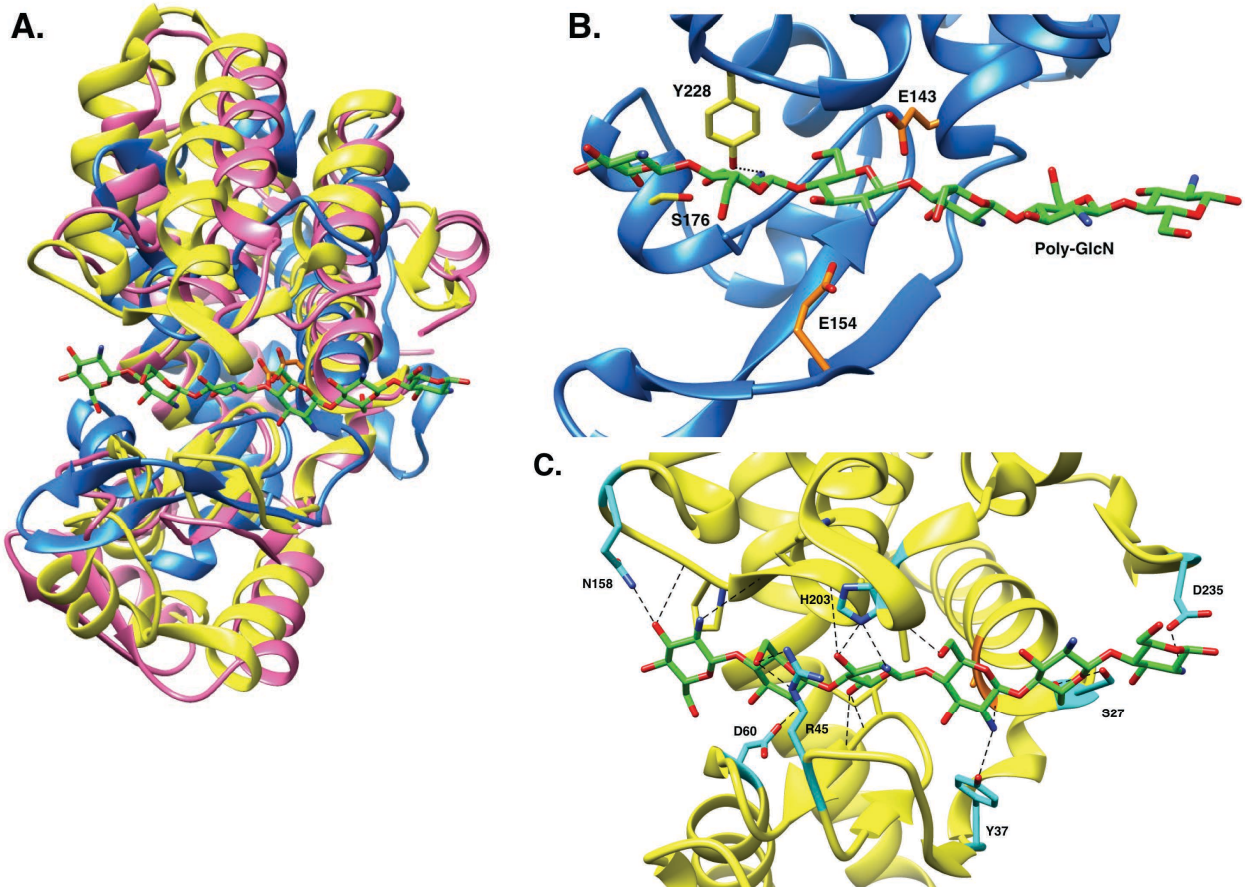


Figure 30. Comparison of Bd0314 with chitosanases.
A. Bd0314 (blue) superimposed with chitosanase-OU01 (PDB:4OLT, yellow, length = 2325 atoms, rmsd = 1.82Å) and *Bacillus circulans* chitosanase (PDB:1QGI, pink, length = 2226 atoms, rmsd = 2.57Å). **B.** Close up of the view of a poly-GlcN substrate-binding model to Bd0314 active site cleft (from superimposition with 4OLT). Y228 could potentially make a hydrogen bond contact with the amine group of GlcN residues. **C.** Substrate binding site of chitosanase-OU01 shows polar residues make hydrogen bonds to free amine group. Charged residues also interact with probable charge on GlcN. Superimposition was performed by least-squares minimisation using the CCP4 suit superpose software (alignment of conserved secondary structure elements, helix-F, helix-I and beta-hairpin).

3.3.3 Likely catalytic mechanism

The structure of Bd0314 revealed a potential secondary catalytic residue (E154) located opposite to the strictly conserved catalytic E143 residue. We demonstrated that both E154 and E143 are required for catalysis. A catalytic mechanism involving two glutamate residues is unusual for lysozyme superfamily members, which typically

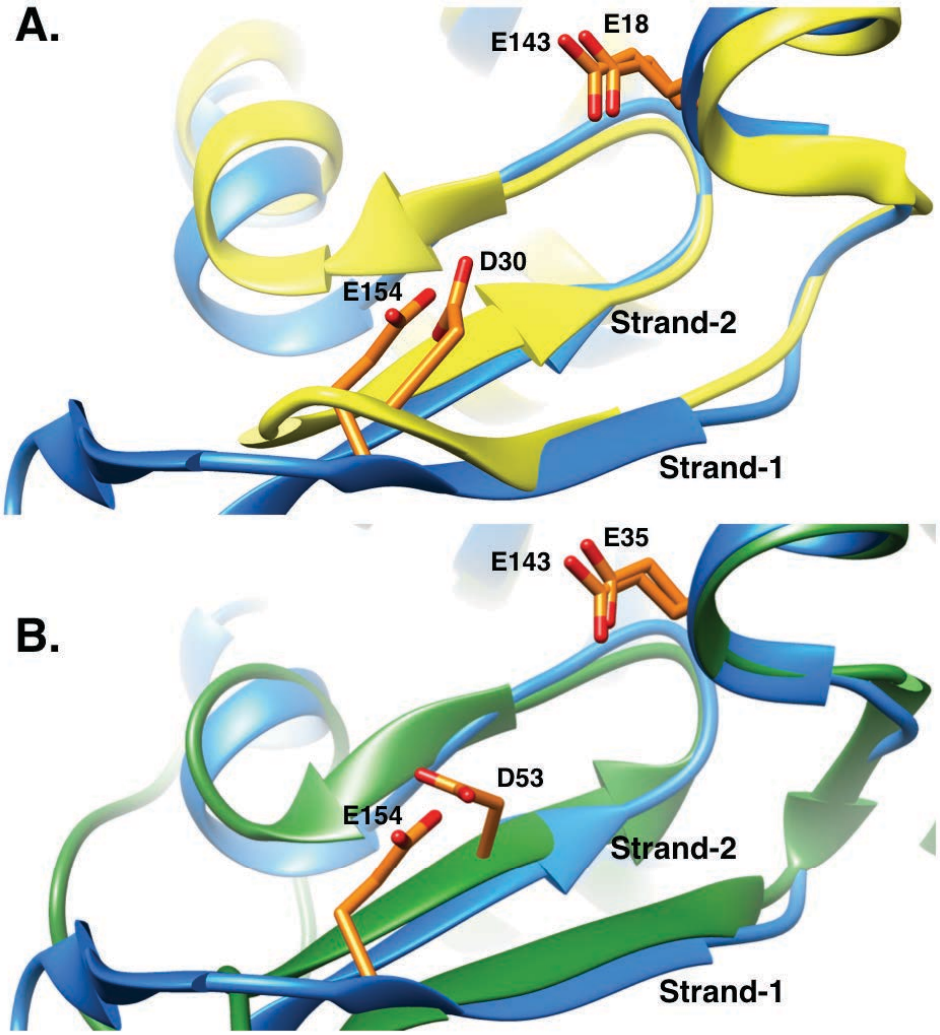


Figure 31. Bd0314 catalytic residue location.
The location of the second catalytic residue varies between lysozymes and generally dictates the catalytic mechanism, which is dependent on distance. **A.** E143 is located on the ES motif at the end of the conserved central α-helix in Bd0314 (blue). Bd0314 unusually has a second glutamate catalytic residue (E154), which is located opposite on strand-1 similar to D30 of *Tapes japonica* Lysozyme (PDB:2DQA, yellow). **B.** In HEWL (PDB:1LZS, green) the second catalytic residue, D53 is located on strand-2.

3.3.4 Bd0314 role as the exit enzyme

Early studies suggested bdelloplast lysis at the end of the predatory lifecycle was brought about by an increase in lytic enzyme activity towards the end of the predatory cycle. The lytic activity was shown to be distinct from the initial pore forming activity, which was halted after entry [45]. The enzyme(s) responsible for bdelloplast lysis was hypothesised to work specifically on the modified (lysozyme resistant) prey cell wall

of MltF to the outer membrane does not impede its activity but ensures localised control of its activity.

On the other hand the flexible tether of Bd0314 may be cleaved by one of the many proteases *Bdellovibrio* secretes into the prey cell during predation [24]. The disordered region would be an easily accessible substrate to a protease. If cutting the tether and freeing Bd0314 is required for bdelloplast lysis then the need for cleavage would add in an additional control step before committing to exiting the bdelloplast.

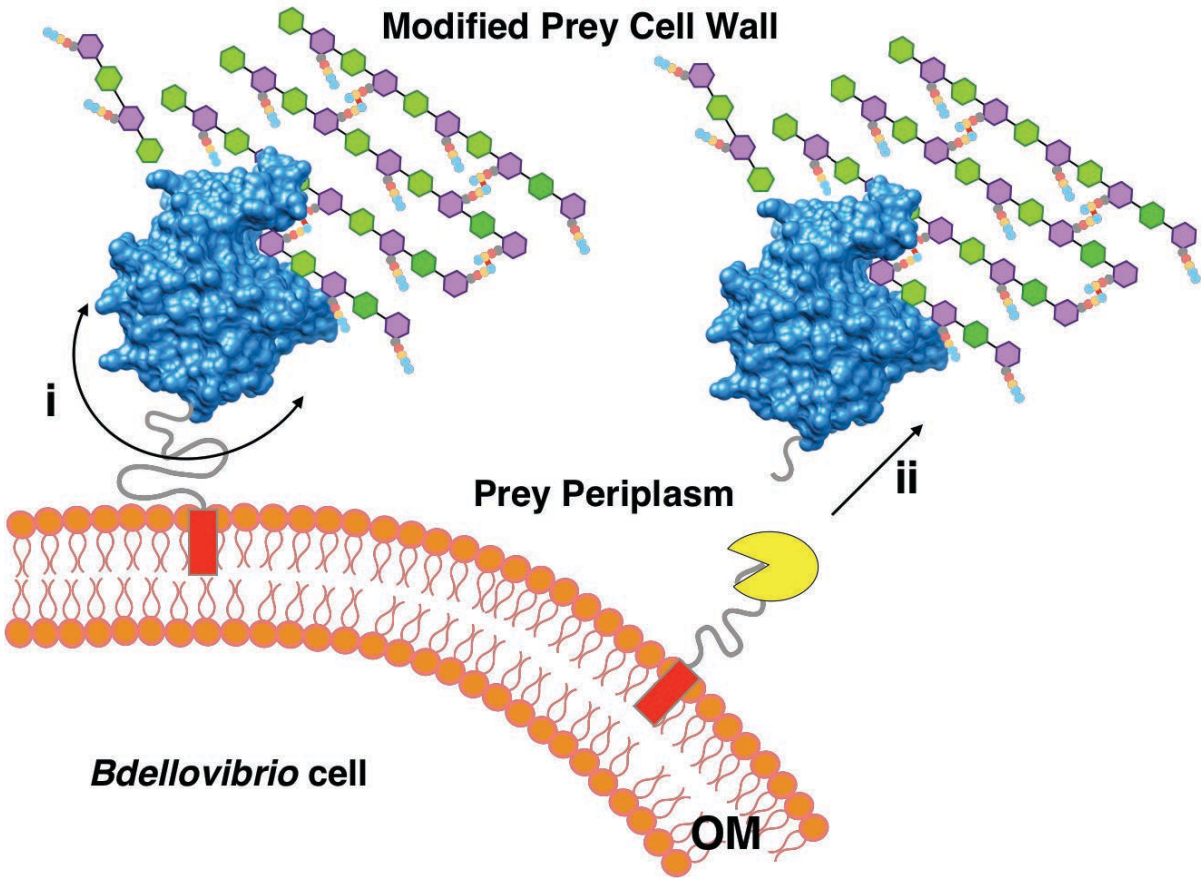


Figure 32. Role of Bd0314
Bd0314 activity is able to solubilise the modified prey cell at the end of the predatory cycle. Full-length Bd0314 possesses a signal peptide with a lipobox suggesting Bd0314 is secreted to the outer-membrane. Between the lipobox and the catalytic domain is a 47aa region of disorder. The disordered region may act as a tether, retaining Bd0314 to the outer membrane for localised prey cell wall degradation (i). Alternatively the tether may be a target of protease action and cleavage of the protease accessible disordered region would release Bd0314 into the bdelloplast (ii). The lipobox is represented by the red box. The disordered region of Bd0314 is illustrated by the grey curved line. Bd0314 is shown in blue and the prey cell wall is illustrated as in Figure 7 (green - GlcNAc/GlcN residues and purple – MurNAc). Potential protease cleavage is shown by the yellow shape.

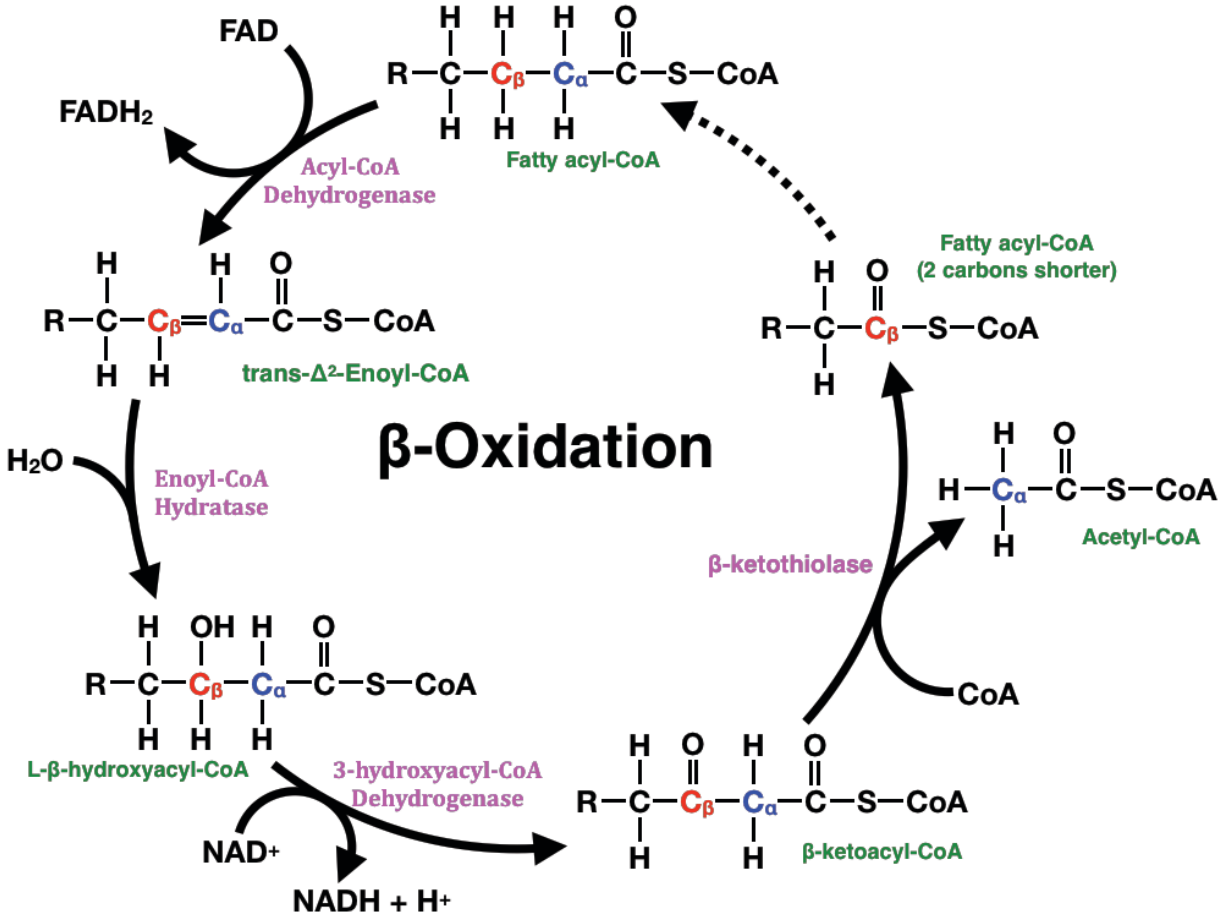


Figure 33. β-oxidation metabolic pathway
β-oxidation is a catabolic process, involving the **c**hemical modification of fatty acyl-CoA substrates at the C_α and C_β positions. Acyl-CoA dehydrogenases catalyse the initial dehydrogenation reaction of fatty acyl-CoA to form trans-Δ^2-enoyl-CoA. An enoyl-CoA hydratase converts trans-Δ^2-enoyl-CoA into L-β-hydroxyacyl-CoA. Then a 3-hydroxyacyl-CoA dehydrogenase produces β-ketoacyl-CoA with the use of an NAD cofactor. A thiolase catalyses the final thiolysis step by inserting the thiol group of CoA between the C_α and C_β positions. The β-oxidation cycle can be repeated until the acyl-chain length is reduced to the final two carbons.

4.1.3 β-oxidation regulation in bacteria

Fatty acids are an essential component of bacterial membranes and a source of metabolic energy. Regulation of fatty acid biosynthesis and degradation is tightly controlled to ensure membrane lipid homeostasis is maintained. The majority of research into β-oxidation pathway regulation has been carried out in *E. coli* [177].

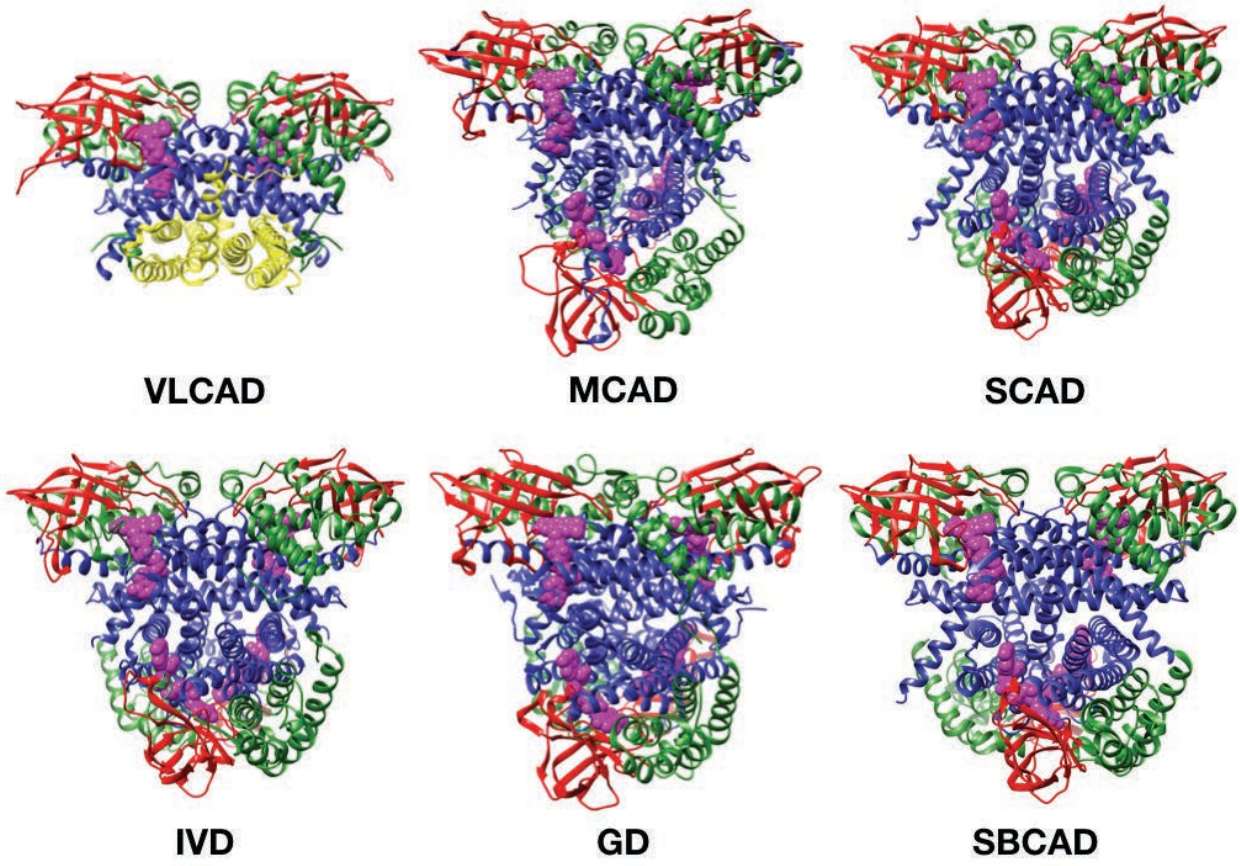


Figure 34. Acyl-CoA dehydrogenases
This Figure illustrates the three-dimensional structures of different members of the ACAD family. The structures are coloured based on their domains: N-terminal α-helical domain (green), central β-sheet domain (red), C-terminal α-helical domain (blue) and finally C-terminal extension domain (yellow) the FAD molecules are coloured magenta. ACADs form tetramers with the exception of VLCAD that forms a dimer. VLCAD – PDB:3B96, MCAD – PDB:3MDE, SCAD – PDB:2VIG, IVD – PDB:1IVH, GD – PDB:3MPI, SBCAD – PDB:2JIF.

ACAD subunits fold into three distinct domains: a N-terminal domain (green) and a C-terminal domain (blue) that are mainly comprised of α-helices and a central domain (red) that is formed of two orthogonal β-sheets (Figure 35.A). The C-terminal domain is extensively involved in forming the tetramer and dimer interfaces. The tetrameric structure is composed of a dimer of dimers in a tetrahedral arrangement. The dimer interface is extensive and includes shared FAD binding sites, whereas the tetramer interface mainly involves helix-helix interactions [180]. VLCAD is an exceptional case, which forms an identical dimer interface to MCAD. However, an additional C-terminal

domain (+180 residue) prevents the formation of the conventional tetramer interface. The additional C-terminal domain adopts a fold identical to the N and C-terminal α-helical domains (tetramer interface contacts) of the opposing MCAD dimer, hence the additional C-terminal domain mimics the tetramer interface contacts.

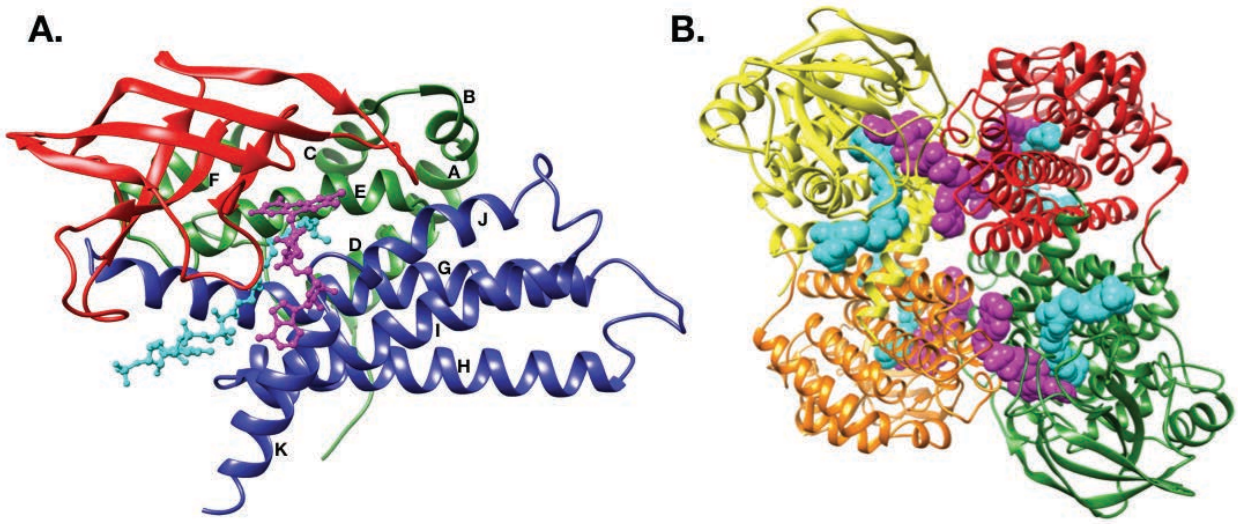


Figure 35. MCAD structure
MCAD is the most well-characterised ACAD, structurally and biochemically. **A.** MCAD can be split into three domains: an N-terminal α-helical domain (green), central β-sheet domain (red) and C-terminal α-helical domain (blue) that provides the tetramer interface. Each monomer binds an FAD molecule (magenta) and can bind a single C8-CoA ligand(cyan). Helices are labelled according to Pig MCAD, for reference to the text. **B.** MCAD forms a tetramer complex in the form of a dimer of dimers. Red + Yellow and Green + Orange monomers make a single dimer. The FAD molecules bind between dimers.

4.1.6 Substrate binding

The FAD molecule is located at the dimer interface between two monomers. The isoalloxazine ring is located in the crevice between the two α-helical domains and the β-sheet domain of one monomer, whilst the adenosine group is located at the dimer interface (Figure 35).

The fatty acyl-CoA substrate binds in the tunnel between the three domains of a single monomer. The adenosine group binds at the surface in a groove leading to the opening

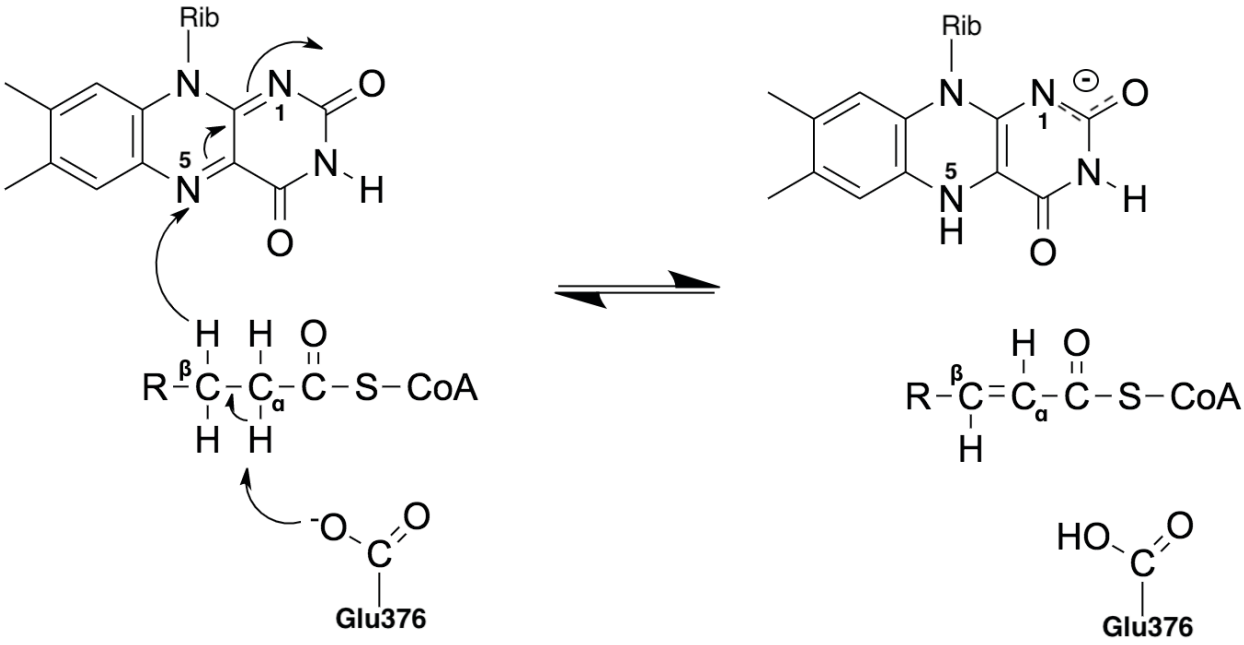


Figure 36. ACAD mechanism
This figure illustrates the catalytic mechanism of ACADs, exemplified by pig MCAD. The catalytic E376 abstracts a proton from the C_α, which synchronously causes the hydride transfer from C_β to be transferred to the N5 position of FAD as a hydride. $FADH_2$ is deoxidised to FAD in two $1e^-$ transfers to a flavin carrying ETF.

4.1.8 ETF and its interactions with ACADs

The reduced FAD molecule is re-oxidised through two sequential re-oxidation reactions with an electron transfer flavoprotein (ETF). ETFs are involved in accepting electrons from all acyl-CoA dehydrogenases shuttling them to the membrane bound ETF-ubiquinone oxidoreductase of the respiratory chain [185, 189] (Figure 37).

ETFs are unusually heterodimers, composed of α and β subunits with a single FAD molecule and a redox inactive AMP molecule [190]. The interflavin redox reaction between the FAD molecules of ACAD and ETF occurs through two successive one electron transfer steps. The FAD molecules are buried in both enzymes thus a direct hydride exchange would not be feasible. It is believed that ACADs and ETF interact weakly with each other and ETF undergoes a large dynamic conformational change in

order to interact with ACAD. A single ETF α-subunit:MCAD complex structure has been published revealing the ETF α-subunit interacts with the N-terminal α-helical domain of MCAD [191]. The interaction mainly involves hydrophobic interactions. The location of the FAD molecule in the β subunit of ETF in the resting state implies a large conformational change is required to correctly position the FAD molecule in the electron transferring state [191]. Conformational change is supported by SAXS results that confirm ETFs sample a range of conformations in solution [192].

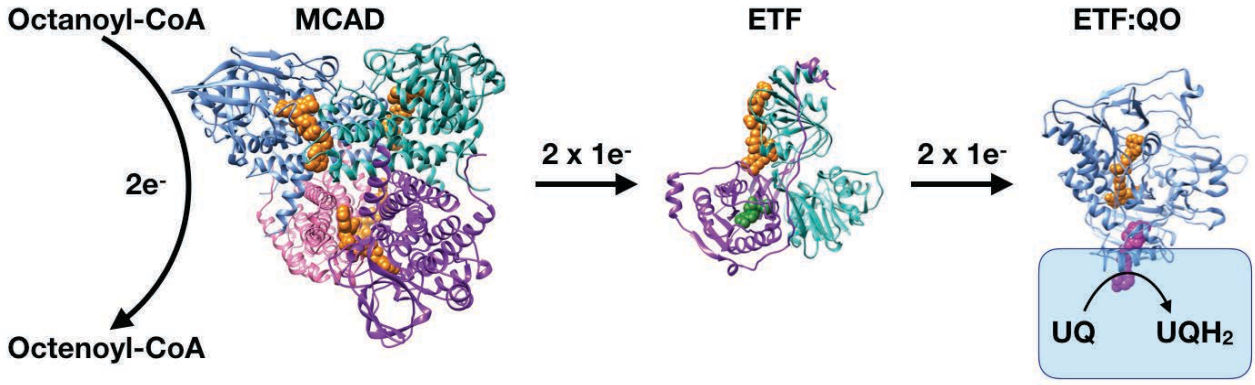


Figure 37. Electron-transfer by ETF
Octanoyl-CoA is dehydrogenated into Octenoyl-CoA transferring two electrons to the FAD of MCAD in the first step of β-Oxidation. $FADH_2$ is re-oxidised to FAD by two successive $1e^-$ transfers to an ETF. ETF then transfers the single e^- to ETF:QO which in turn transfers the electrons to ubiquitin pool. ETF can accept electron transfer from various ACAD family members. Structures: MCAD – 1T9G, ETF – 1EFP, ETF-QO - 2GMH. FAD molecules are coloured in orange, AMP molecule in green and Ubiquinone molecule in magenta.

4.1.9 ACAD homologues including acyl-CoA Oxidases (ACOs)

Homologues of the ACAD fold exist that possess divergent functions. Most divergent homologues are acyl-CoA oxidases of peroxisomal origin [180, 193], while others are involved in steroid side chain cleavage [194], antibiotic synthesis [195] and response to environmental stresses [196-198].

ACOs are highly similar to ACADs and catalyse the initial dehydrogenation reaction of β-oxidation in peroxisomes. ACOs, however, do not require an ETF to re-oxidise; instead ACO use molecular oxygen, producing hydrogen peroxide. The structures of ACOs share a common ACAD fold, more similar to VLCAD. ACOs form a soluble dimer and have an additional C-terminal α-helical domain (C-I domain, yellow) that mimics the tetramer interface of MCADs. In addition ACOs have an α-helical extension in the C-terminal domain (C-II domain, orange) that overlaps with the central β-sheet domain and reduces the accessibility of FAD to bulk solvent. The overlapping helix is also believed to be involved in blocking ETF docking and potentially changing the redox environment of the FAD molecule [180].

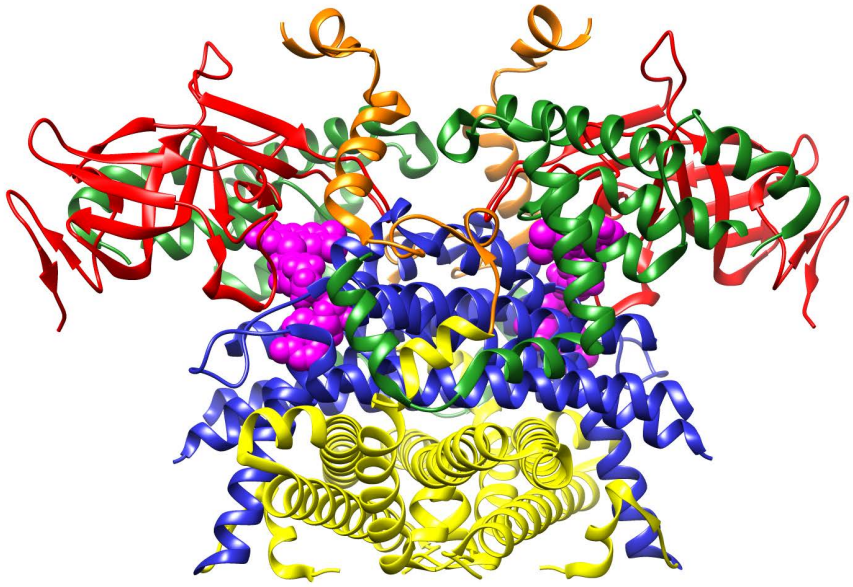


Figure 38. Structure of ACO
The structure of ACO (PDB:2DDH) dimer is coloured according ACAD domains. N-terminal domain (green), central domain (red), C-terminal domain (blue). ACOs contain an addition C-terminal α-helical domain (yellow) that is similar to the C-terminal domain in VLCAD, which mimics the MCAD tetramer interface. Additionally ACOs have an elongated α-helical region (orange) that caps the FAD accessible groove and prevents ETF access to FAD.

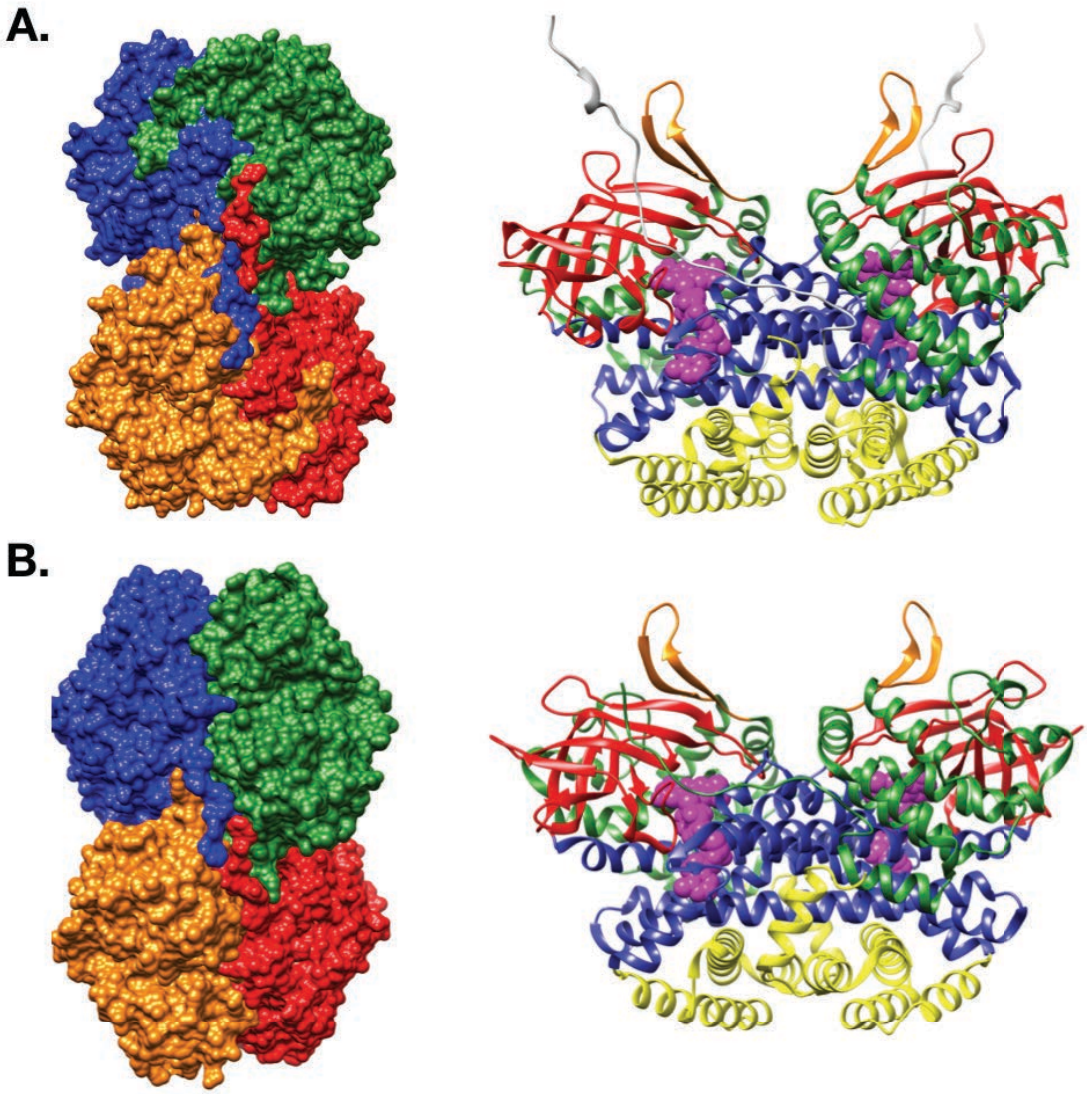


Figure 39. Structure of divergent ACAD homologues
A. ACDH-11 forms a tetramer in the form of a dimer of dimers (blue:green + red:orange), which is organised differently to ACAD tetramers. The tetramer interface is provided by a β-sheet loop in the N-terminal domain (orange). The structure of ACDH-11 domains are coloured similarly to ACAD structures, N-terminal domain (green), central domain (red), C-terminal domain (blue) and C-terminal extension domain (yellow). ACDH-11 is more closely related to the structure of VLCAD. **B.** The structure of AidB is highly similar to ACDH-11. AidB also forms a similar tetramer interface to ACDH-11, facilitated by the N-terminal domain β-sheet loop (orange).

4.2 Results

4.2.1 Cloning, Expression and Purification of Bd2924

Full-length Bd2924 (residues 1 to 505) was cloned into a pET28a vector and expressed in *E. coli* Bl21* (DE3) cells. Bd2924 was purified by IMAC to >95% purity (Figure 40). Purified Bd2924 was bright yellow in solution, indicating the presence of co-purified FAD cofactor.

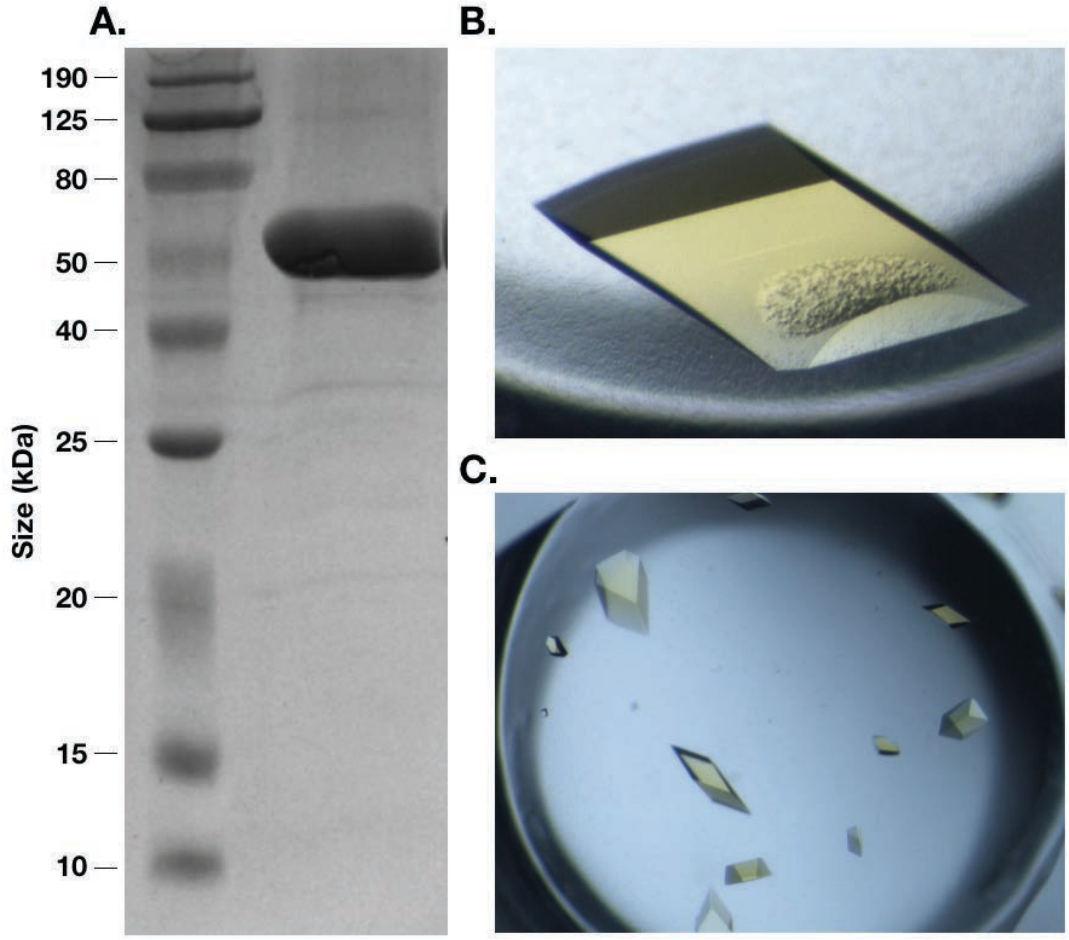


Figure 40. Bd2924 purification and crystallisation
A. Bd2924 was purified to >95% purity, judged by SDS-PAGE. **B.** Image of Bd2924 crystal **C.** Image of Bd2924 crystals

4.2.2 Obtaining crystals and data collection

Large bright yellow Bd2924 crystals were grown, which were used for data collection. The structure of Bd2924 was originally solved by molecular replacement, using PDB:4Y9J as the search model (see Section 2.4.5). We determined the structure of Bd2924 (residues 3-505) in complex with the cofactor FAD, both with and without a C10 length acetyl-CoA thioester ligand (C10-CoA) bound, to 1.51Å and 1.87Å respectively.

4.2.3 Protein Structure / topology

Overall the tertiary structure of Bd2924 is a homotetramer consisting of a dimer of dimers. Bd2924 shares a common ACAD fold, more closely related to divergent ACAD homologues such as AidB and ACDH-11. The homodimer packs across a 2-fold axis to create a likely biologically relevant crystallographic homotetramer, where the CD dimer is identical to the AB dimer. SEC confirms Bd2924 also exists as a tetramer in solution (Figure 41).

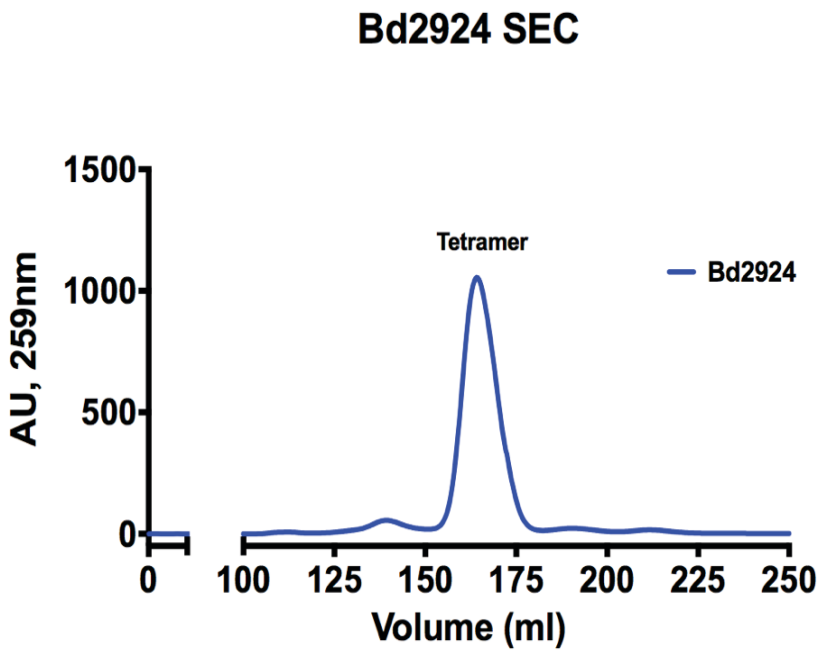


Figure 41. Bd2924 SEC
Bd2924 elutes as a single peak at 165 ml, confirming Bd2924 exists as a tetramer in solution.

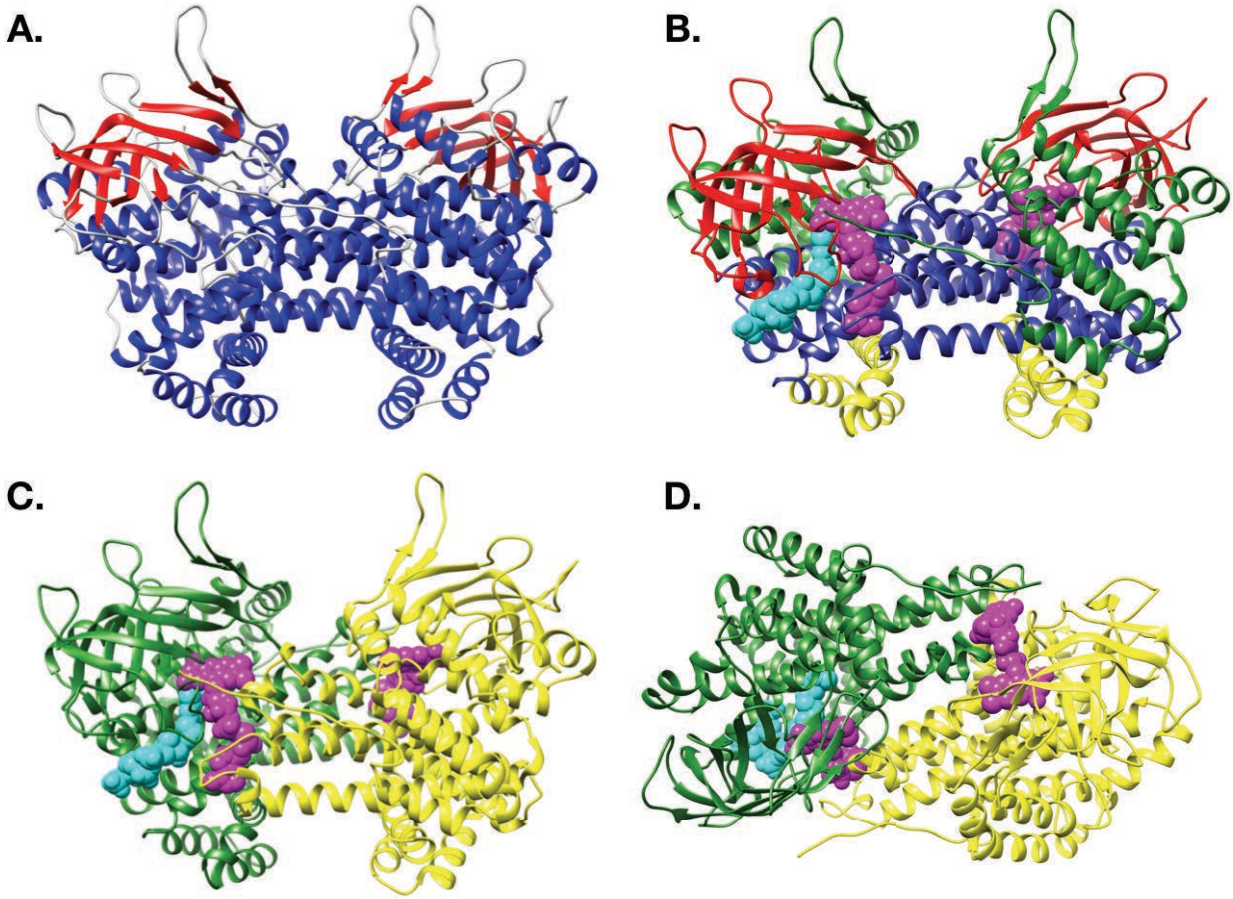


Figure 42. Structure of Bd2924
A. The secondary structure of Bd2924 dimer is mainly composed of α-helices (blue), with a region of β-strands (red). **B.** Bd2924 can be divided into four domains, similarly to other ACADs. Bd2924 has an N-terminal α-helical domain (green), a central β-sheet domain (red), a C-terminal α-helical domain (blue) and finally a C-terminal α-helical extension domain (yellow) similar to VLCAD. **C.** Bd2924 forms a tightly interacting dimer, shown here as a green monomer and yellow monomer. FAD (magenta) binds at the dimer interface. The C10-CoA binding site (Cyan) is found is located in the tunnel running beneath the FAD molecule. **D.** A 90° rotation of Bd2924 showing the dimer interface from another angle.

Bd2924 can be divided into four domains based on other ACAD structures. Bd2924 has an N-terminal α-helical domain that is interrupted by a single β-sheet projection (residues 3-168), a central β-sheet domain formed by two orthogonal four-stranded antiparallel β-sheets (residues 169-285), a central α-helical domain (residues 286-449) and a short C-terminal α-helical domain (residues 450-505). A single FAD molecule binds per monomer in a crevice at the dimer interface. The N-terminal β-sheet projection and the C-terminal α-helical domain create two large grooves in the Bd2914

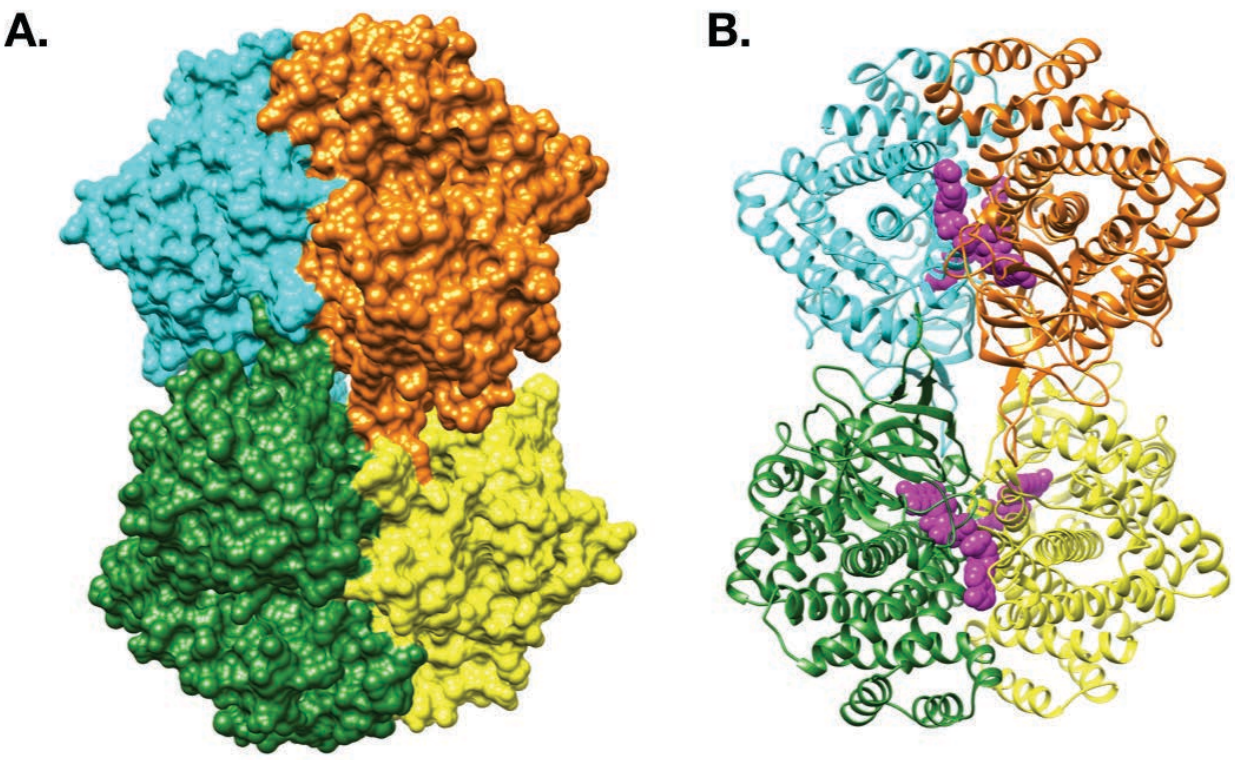


Figure 43. Bd2924 tetramer complex
Bd2924 forms a tetramer complex formed by a dimer of dimers. The tetramer interface is provided by a β-sheet loop located as an insertion into the N-terminal domain. The tetramer complex surface is shown in (**A**) and in ribbon form (**B**), which shows the location of FAD molecules (magenta).

4.2.5 Ligand Binding

4.2.5.1 FAD

FAD binds at the dimer interface and provides structural stability to the protein besides being a required catalytic cofactor. Each monomer binds a single FAD molecule in a crevice created at the dimer interface. The isoalloxazine group of FAD is located in a groove created between the N-terminal α-helical domain, the β-sheet domain and the central α-helical domain of a single monomer with a solvent accessible tunnel running beneath. The ADP region of FAD forms multiple hydrogen bonds with both chains of the AB dimer as well as multiple non-bonded interactions. Figure 44 below depicts the hydrogen-bonding network associated with FAD binding to Bd2924. The FMN region

makes H-bonds with Q171, M173 and T174 of Chain A and G406 of Chain B. The di-phosphate group makes the most intra-dimer contacts including: S180 and G179 of chain A and G406 and R324. The AMP group makes less extensive interactions: E402 and T431 H-bond with the ribose moiety whereas the adenine moiety is stabilised solely through non-bonded interactions.

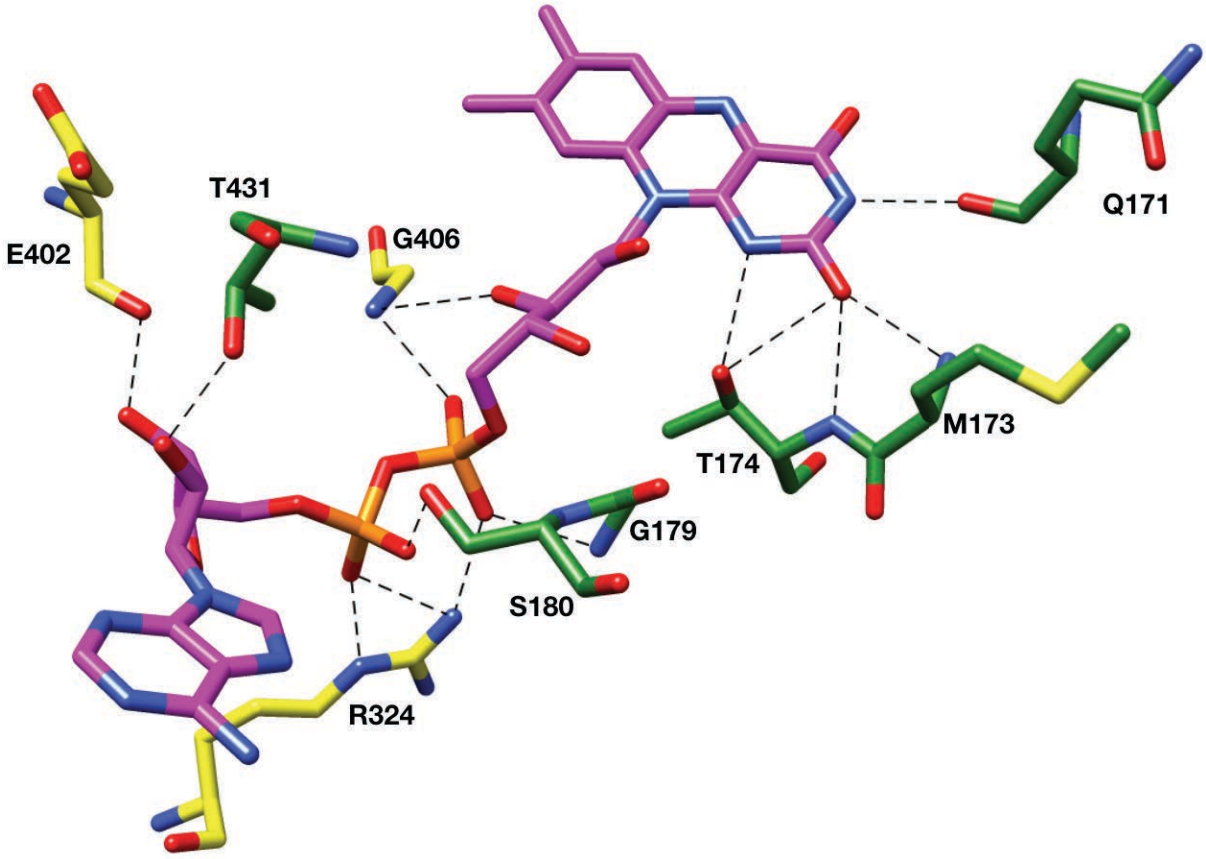


Figure 44. FAD hydrogen bonding network
FAD binds in the pocket at the dimer interface and held in place by numerous non-bonding contacts and a network of hydrogen bonds from residues from both monomers of the dimer. Residues with green coloured carbon atoms are from Chain A, whereas those in yellow are from Chain B (of an AB dimer). Hydrogen bonds are depicted by black dashed lines.

4.2.5.2 C10-CoA

A Bd2924:C10-CoA complex was solved in order to understand substrate binding. This was achieved by crystal soaking C10 and C12 chain length acyl-CoA thioesters into Bd2924 crystals. A chain length of C10-12 was chosen because the Bd2924 substrate-binding tunnel most closely resembles that of ACDH-11, which co-purified and co-crystallised with a C11-CoA thioester ligand. The structure of Bd2924 in complex with C10-CoA fatty acid revealed the interactions Bd2924 makes with substrate. In the Bd2924:C10-CoA complex structure only a single C10-CoA molecule can be confidently placed per dimer, due to weak electron density in the $2F_o$-F_c map for the second molecule.

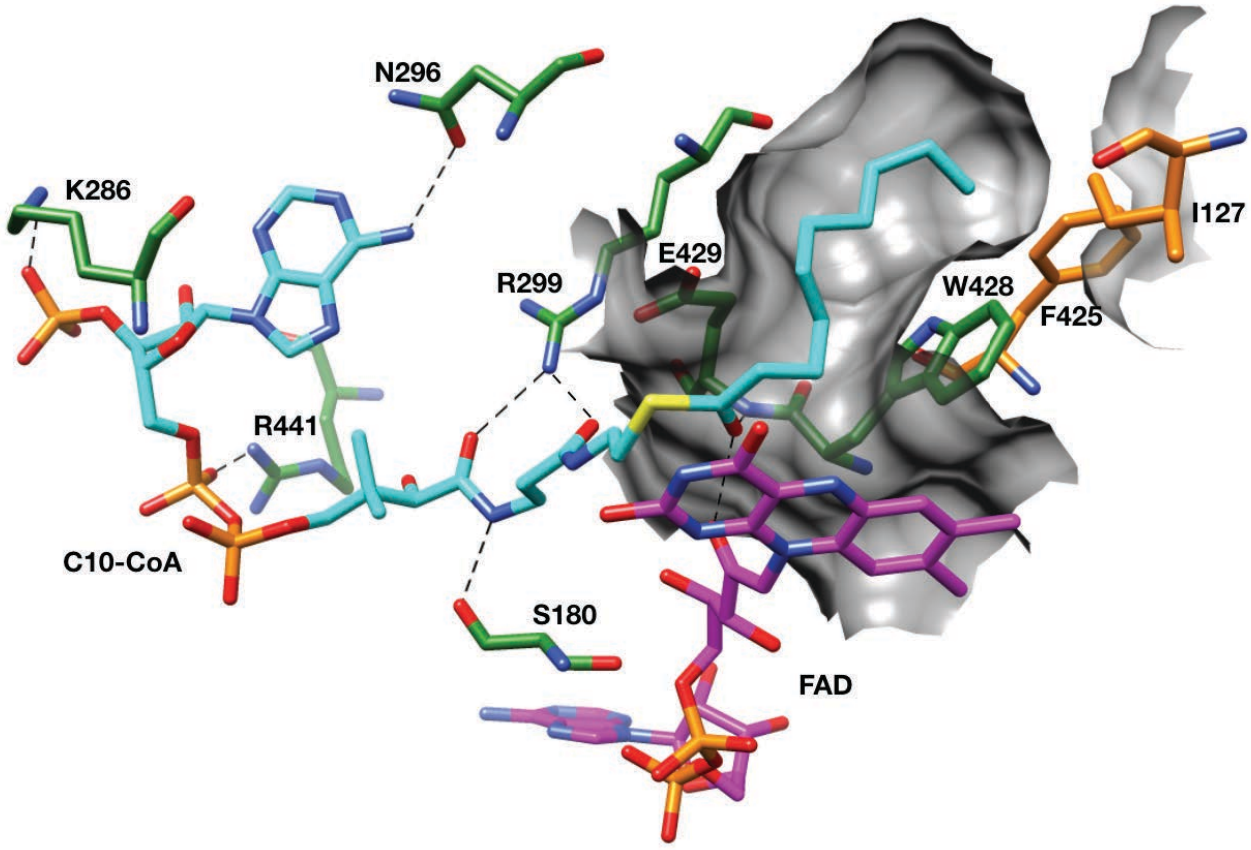


Figure 45. C10-CoA binding
The structure of Bd2924 has been solved in complex with a C10-CoA ligand (cyan). A number of hydrogen bonds (8) hold C10-CoA in place. An important hydrogen bond between the thioester carbonyl group and the backbone NH of E429 and the FAD-2'OH correctly orientates the substrate for catalysis. The depth of the substrate tunnel is limited by the sidechains of I127 and F428 (orange). Grey represents surface of pocket.

Besides MST binding studies we also attempted to obtain a Bd2924:CdG complex through X-ray crystallography experiments. Co-crystallisation and ligand soaking experiments, with concentrations of CdG up to 2 mM, were carried out with little success. The failed attempts to determine the location of a binding site via crystallography alongside the relatively weak K_D obtained through our binding assay and no conventional nucleotide-binding sites in the structure, suggest CdG may bind to Bd2924 non-specifically.

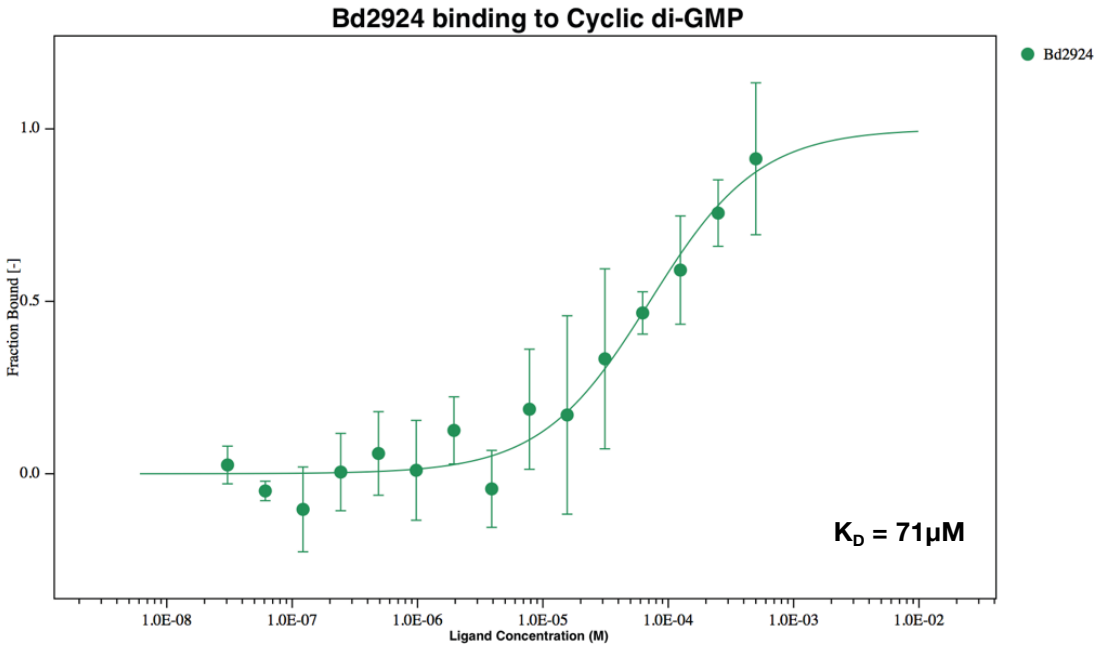


Figure 46. Binding curve with CdG
The binding interaction between Bd2924 and CdG was monitored using MST

4.2.7 Bd2924 activity

We were unable to detect any enzymatic activity for Bd2924 in two independent ACAD activity assays with either C10 or C12 fatty acyl-CoA substrates (DCPIP+PMS assay & Ferrocenium ion assay).

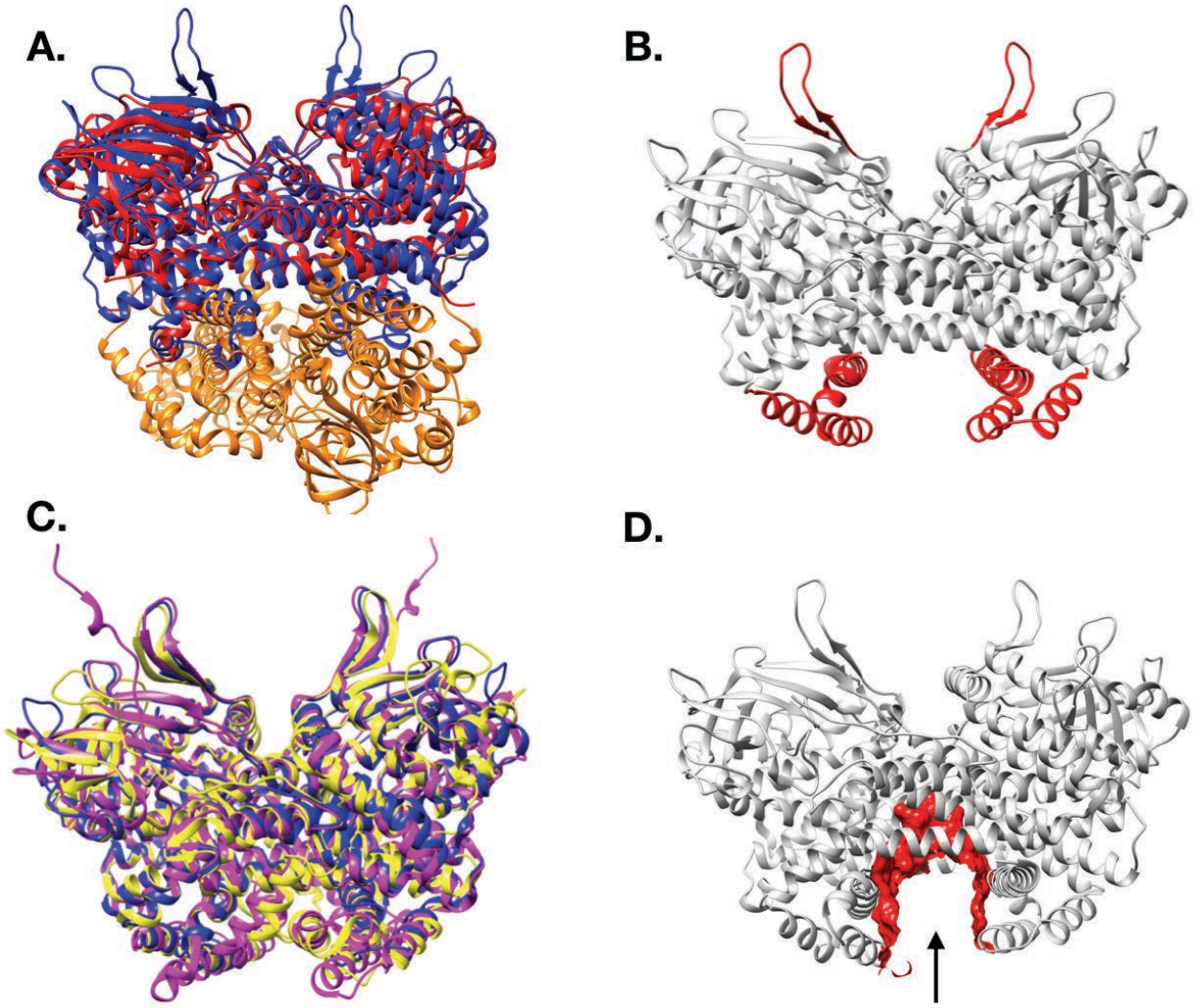


Figure 47. Bd2924 structural comparison.
A. Bd2924 dimer (blue) overlaid with MCAD tetramer (PDB:1EGD, red + orange, length = 746 residues, rmsd = 2.21Å). **B.** The major structural difference between Bd2924 and MCAD are highlighted in red on the structure on Bd2924. The N-terminal β-sheet loop and C-terminal extension domain are features of divergent ACAD homologues that form an unrelated tetramer. **C.** Bd2924 (blue) overlaid with ACDH-11 (PDB:4Y9J, magenta, 961 residues, 1.41Å) and AidB (PDB:3DJL, yellow, length = 881 residues, rsmd = 1.54Å). **D.** Bd2924 has a less extensive C-terminal extension domain and as a result has a large groove between dimers (red). Alignments were performed using the SSM superimpose tool on coot.

4.3.2 Substrate-binding and activity

Bd2924 is able to bind C10 chain-length fatty acyl-CoA, demonstrated by the Bd2924:C10-CoA complex structure. Bd2924 binds C10-CoA with more interactions (eight hydrogen bonds) than MCAD binds acyl-CoA substrates (four hydrogen bonds)

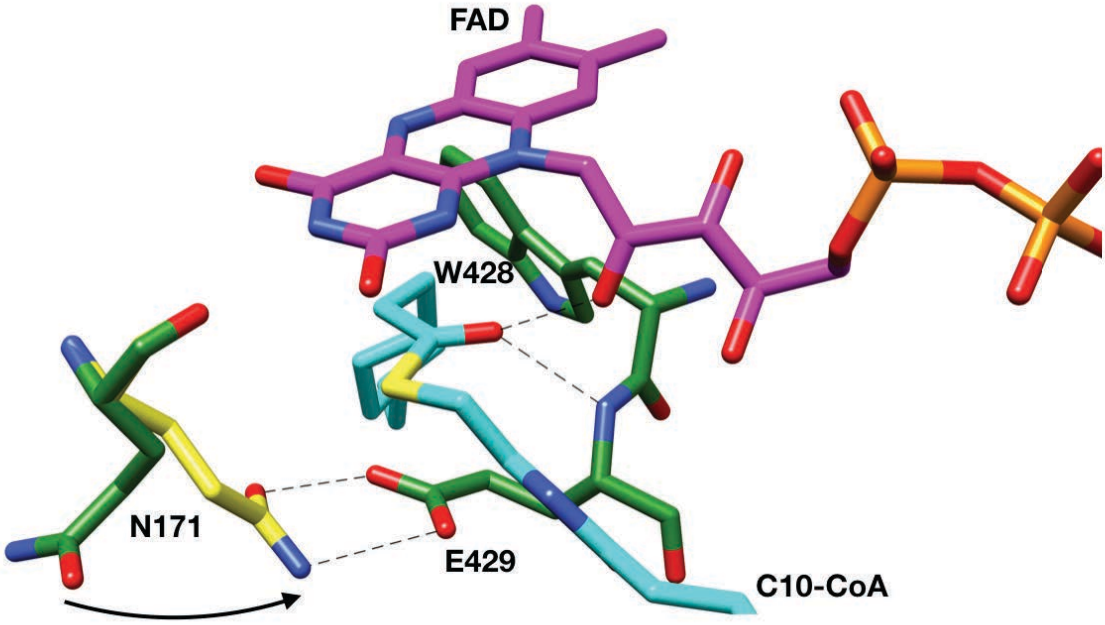


Figure 48. Bd2924 active site.
The C10-CoA ligand is bound in the active site between FAD and the catalytic base E429. In Bd2924:C10-CoA complex N171 adopts a new conformation (yellow) that forms a hydrogen bond with the sidechain of E429 and may affect its ability to abstract hydrogen from the C_α (supressing catalytic funciton).

Besides the lack of detectable dehydrogenase activity for Bd2924 (and comparison to other ACAD homologues) it seems unlikely that transfer of electrons to ETF would be possible by the currently proposed mechanism [191]. Bd2924 tetramer formation would impede access to the FAD molecule buried deep between the dimer and tetramer complex. As such for an ETF to access and re-oxidise the buried FAD molecule in Bd2924, it would have to disassemble our observed tetramer complex [203]. Blocking of ETF access is demonstrated in ACO, where a C-terminal α-helix crosses the solvent accessible groove to the si-face of FAD and thereby shields the accessibility to ETF [203]. Bd2924 could alternatively utilise molecular oxygen similar to ACOs although solvent accessibility to the FAD molecule is not as extensive as in ACOs.

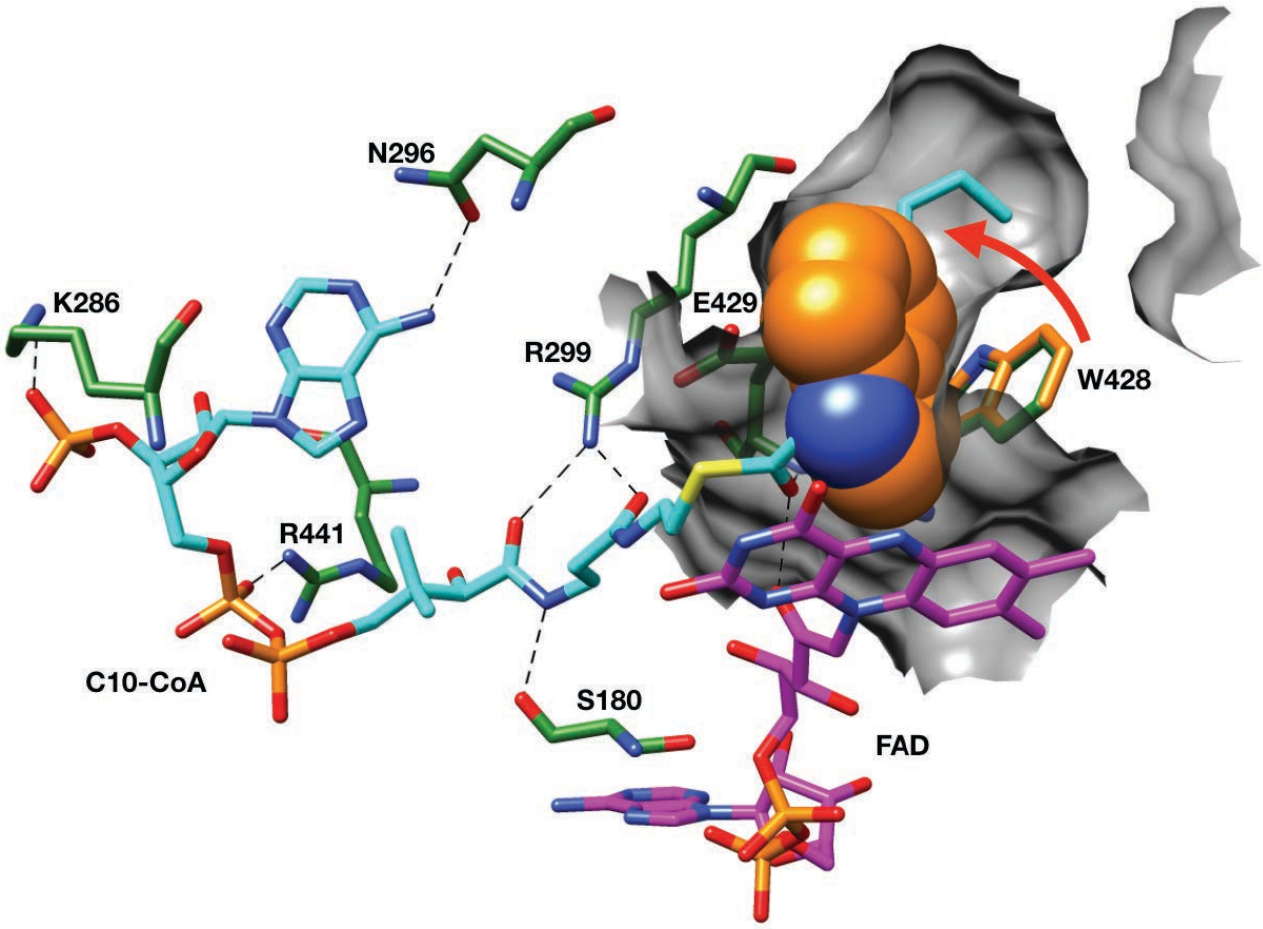


Figure 49. Trp428 gates the fatty acyl binding tunnel
The Bd2924:C10-CoA complex structure has one occupied C10-CoA binding site. In the monomer where the substrate binding site is free of ligand, W428 adopts two conformations (orange). One W428 conformation is located directly beneath the FAD N5, which blocks access to the substrate binding tunnel.

4.3.3 A functionally relevant Bd2924 interaction with CdG is questionable

We determined the binding affinity between Bd2924 and CdG to be ~71 μM, which is approximately 10 fold greater than the published value determined by Rotem et al [82]. To ensure consistency the binding buffer used by Rotem et al. was replicated, but no measurable binding events were observed with a maximum CdG concentration of 5 mM, which far exceeds the cellular concentration of CdG [79]. Binding was detected after rigorously testing multiple assay buffers and protein labelling techniques. The increased K_D value we obtained suggests that Bd2924 binds CdG in a non-specific

Bd2924s role is also likely to be divergent from β-oxidation based on its structural similarities to divergent ACADs and the lack of detectable dehydrogenation activity.

The exact function of Bd2924 remains unknown and requires further exploration. However, it is interesting to speculate that Bd2924 may be involved in fatty acid detection and utilisation during the *Bdellovibrio* predatory cycle. *Bdellovibrio* utilises prey-derived long-chain fatty acids for energy through β-oxidation and for direct incorporation into biosynthetic pathways [59]. Bd2924 may therefore act as a sensor of fatty acyl chains. Its function may be to: transport fatty acyl-CoA; sequester fatty acyl-CoA as part of a signalling cascade (alike ACDH-11); or generally be involved in β-oxidation. ETF genes exist in *Bdellovibrio HD100* genome (Bd0021 and Bd0022) but these may interact with other acyl-CoA dehydrogenases, such as those named in Table 12. Close homologues of Bd2924 are found in other endo-biotic predatory bacteria (60-75% ID) and *Legionella* species (55% ID). It is therefore interesting to speculate that Bd2924s role is specific to bacteria which invade cells.

Protein	Type	Length	Up-regulated in?
Bd2174	Butyryl-CoA dehydrogenase	816 (including 284aa DUF)	growth phase
Bd1973	Butyryl-CoA dehydrogenase	388	growth phase
Bd1201	Isovaleryl-CoA dehydrogenase	428	growth phase
Bd2924	acyl-CoA dehydrogenase	505	growth phase

Table 12. Hd100 ACADs
A list of ACADs in the HD100 genome – annotated by uniprot

5.1.2 Malic Enzymes

5.1.2.1 Overview

Malic enzymes catalyse the oxidative decarboxylation reaction of L-malate to pyruvate using the cofactor $NAD(P)^+$ in the presence of a divalent cation such as Mn^{2+} or Mg^{2+}.

$$\text{L-malate} + \text{NADP}^+ \rightleftharpoons \text{pyruvate} + CO_2 + \text{NADPH}$$

MEs are categorised into three classes based on coenzyme specificity and their ability to decarboxylate an oxaloacetic acid (OAA) substrate. These classes include: Non-OAA decarboxylating NAD^+ -specific, OAA decarboxylating NAD^+ -specific and OAA decarboxylating $NADP^+$ -specific. MEs are widely distributed enzymes involved in different metabolic pathways such as energy metabolism, photosynthesis and lipogenesis in organisms from bacteria to higher Eukaryotes [205]. Three isoforms of eukaryotic ME have been identified and studied (cytosolic, mitochondrial and plastid). On the other hand, prokaryotic MEs isoforms include a wider range of structural and functional diversity (including multi-modular MEs), which are comparatively understudied.

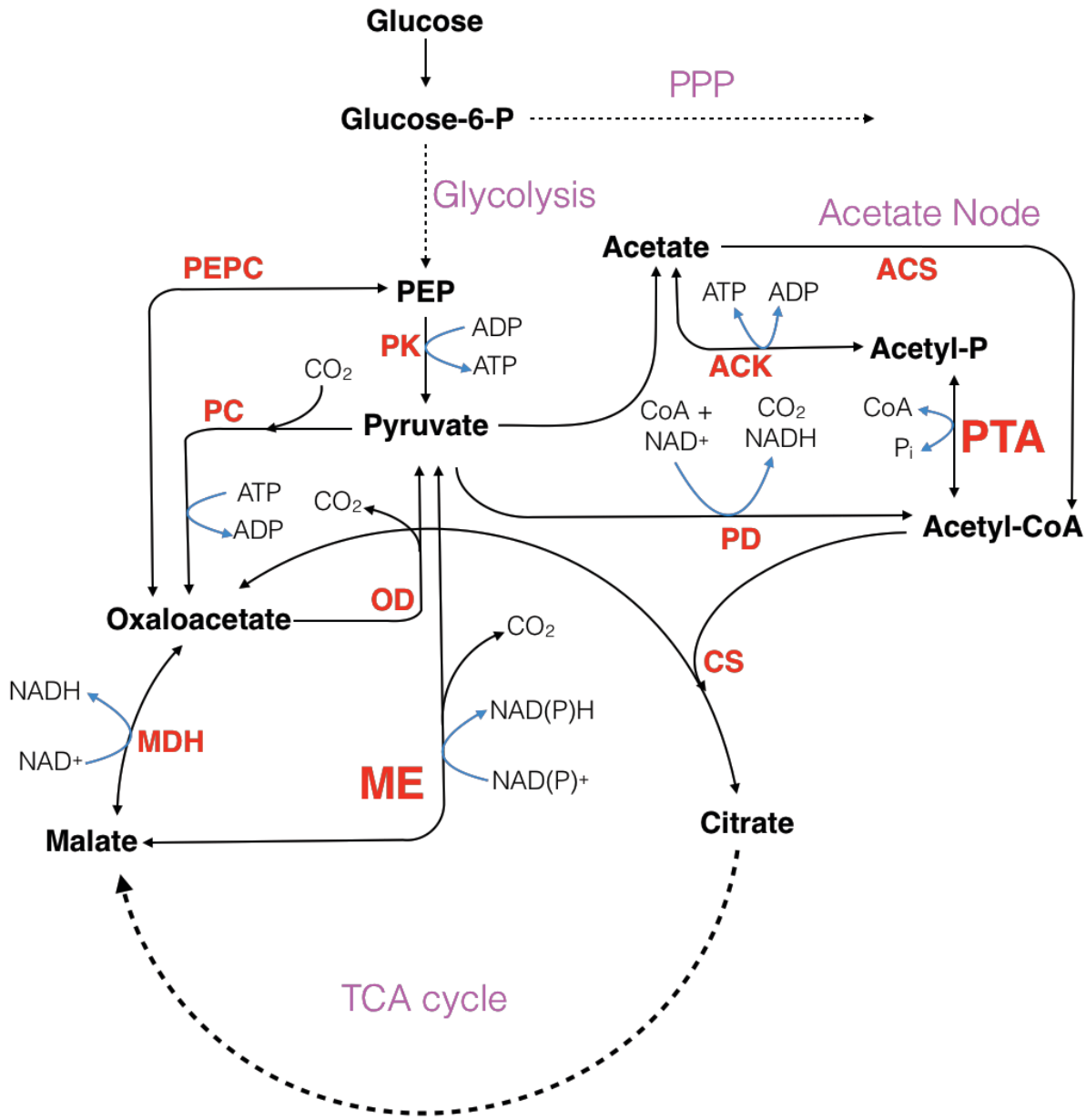


Figure 50. Illustration of central metabolism (general)
This Figures depicts a simplified illustration of a central metabolic network involving Pta and ME enzymes (based on *E. coli* metabolism). Each individual pathway is labelled in pink. The enzymes involved in the reactions are displayed in red: PK - pyruvate kinase, PD - pyruvate dehydrogenase, CS - citrate synthase, PEPC - phosphoenolpyruvate carboxylase, PC - pyruvate carboxylase, OD - oxaloacetate decarboxylase, MDH - malate dehydrogenase, ME - malic enzyme, ACS - acetyl-CoA synthase, ACK - acetate kinase, PTA - phosphotransacetylase

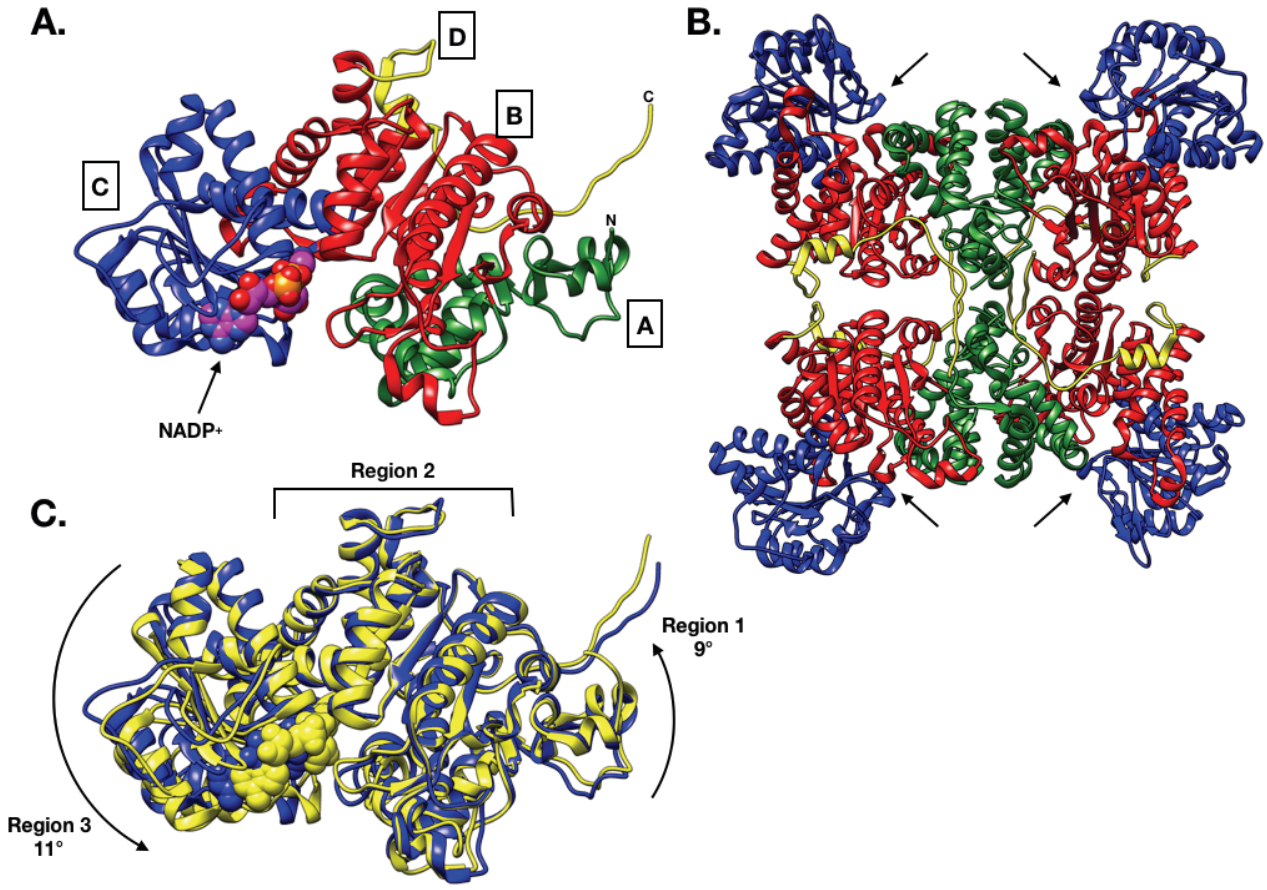


Figure 51. Human-ME tetramer organisation and domains
A. Human-ME like other eukaryotic MEs can be split into four domains: Domain A (Green) makes the dimer interface; Domain B (Red) composes the central core of the protein and cation-binding site; Domain C (Blue) contains the NAD(P)$^+$ binding site; and Domain D (yellow) is a short extension that contributes to the tetramer interface. **B.** Eukaryotic MEs arrange to form a tetramer complex – a dimer of dimers. Black arrows indicate the location of the substrate-binding cleft. **C.** The structure of human-ME has been captured in a catalytically active 'closed' conformation and a resting 'open' conformation. Region-3 including domain C (NADP$^+$ -binding domain) undergoes an 11° rigid body rotational movement that encloses the active-site cavity and positions the catalytic residues and cofactor ready for catalysis. A smaller 9° rigid body rotational movement of part of domain A (region-1) also occurs, which re-orientates the tetramer interface. Region-2 containing domain B and part of domain A maintains the same conformation. The closed form (PDB:1EFK) was superimposed onto the open form (PDB:1QR6), using LSQ superpose tool in coot to align domain B (1275 atoms, rmsd = 0.94Å).

5.1.2.5 Divalent Cation binding

ME activity is dependent on the presence of a divalent cation such as Mn^{2+} and Mg^{2+}. In human NAD(P)$^+$ -dependant-ME, Mn^{2+} is ligated by six oxygen atoms in an octahedral fashion. Three of these oxygen atoms are from acidic residue sidechains; E255, D279 and D256. When no ligand is present in the active-site water molecules provide the remaining oxygen atoms.

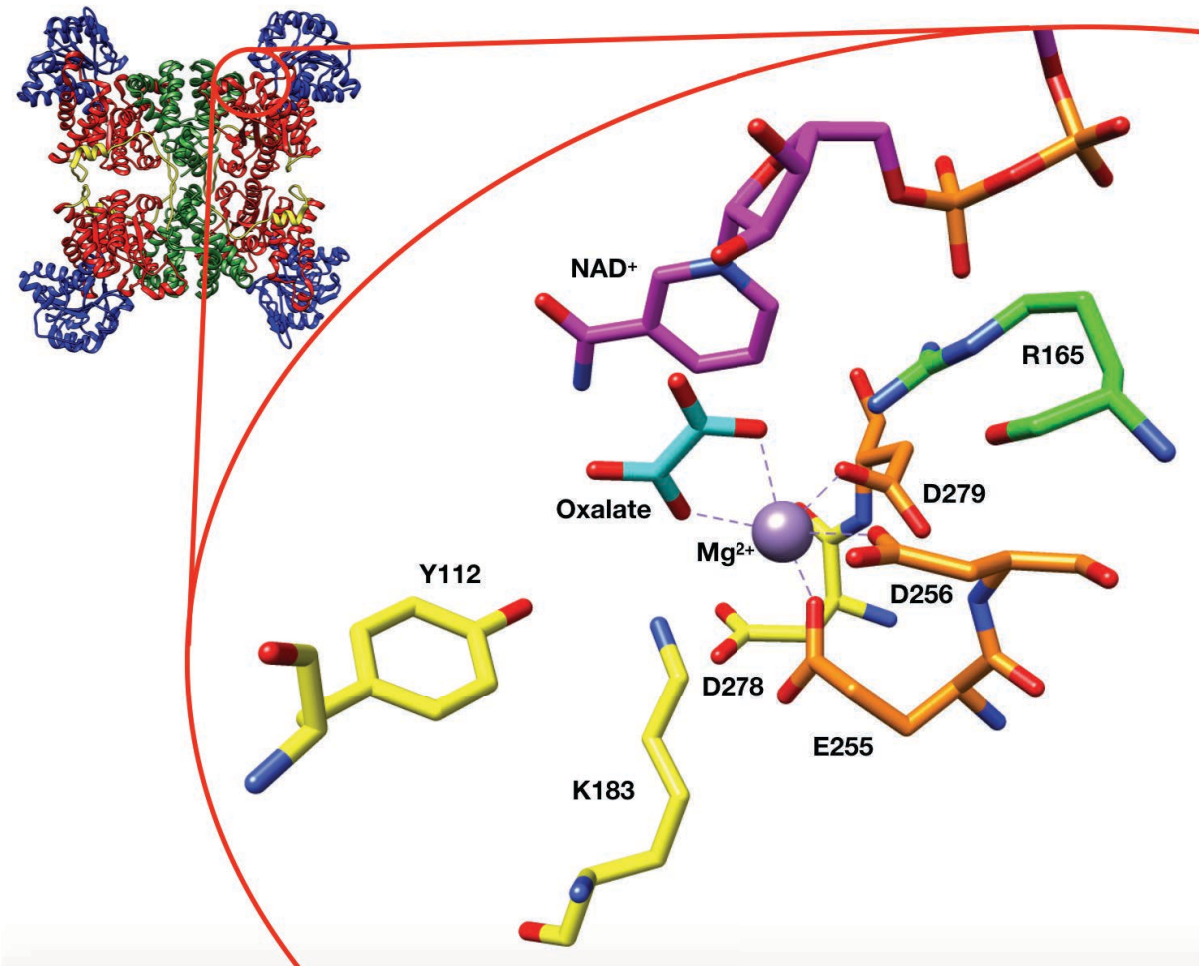


Figure 52. Human-ME active Site
A close up view of Human-ME active site in complex with: Mg^{2+} cation (purple), a malate substrate analogue – oxalate (cyan) and NAD^+ cofactor (magenta). The cation coordinating residues are shown in orange. The substrate-binding residue (R165) is shown in green. Catalytic residues (Y112, K163 and D278) are shown in yellow. PDB:1DO8

5.1.2.7 Nucleotide Binding

ME utilise a $NAD(P)^+$ cofactor to perform catalysis. Generally MEs perform the reaction specifically using either NAD^+ or $NADP^+$. However, some isoforms are able to utilise both cofactors. $NAD(P)^+$ is bound to domain C in eukaryotic MEs and associates with the Rossmann fold. The nicotinamide portion of the cofactor binds in an *anti* conformation, similarly to how other dinucleotides bind to a Rossmann fold. The

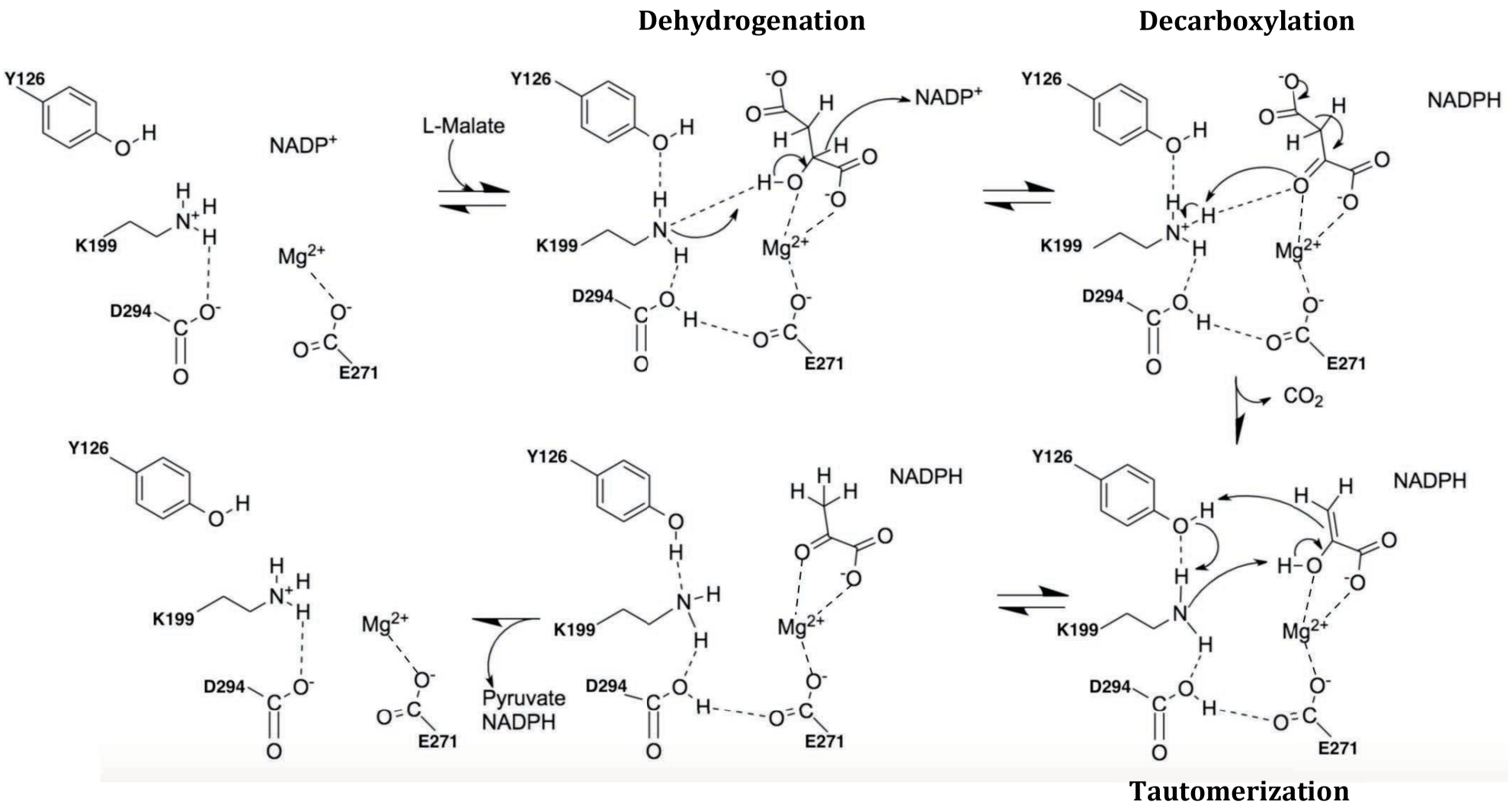


Figure 53. Malic enzyme catalytic mechanism.
The catalytic mechanism performed by malic enzymes progresses via three steps: Dehydrogenastion of L-malate to form oxaloacetate, decarboxylation of oxaloacetate to form enolpyruvate and finally tautomerization of enolpyruvate to form pyruvate. The enzyme is primed by binding Mg^{2+} and $NADP^{+}$ and upon substrate binding a conformational change repositions the catalytic residues.

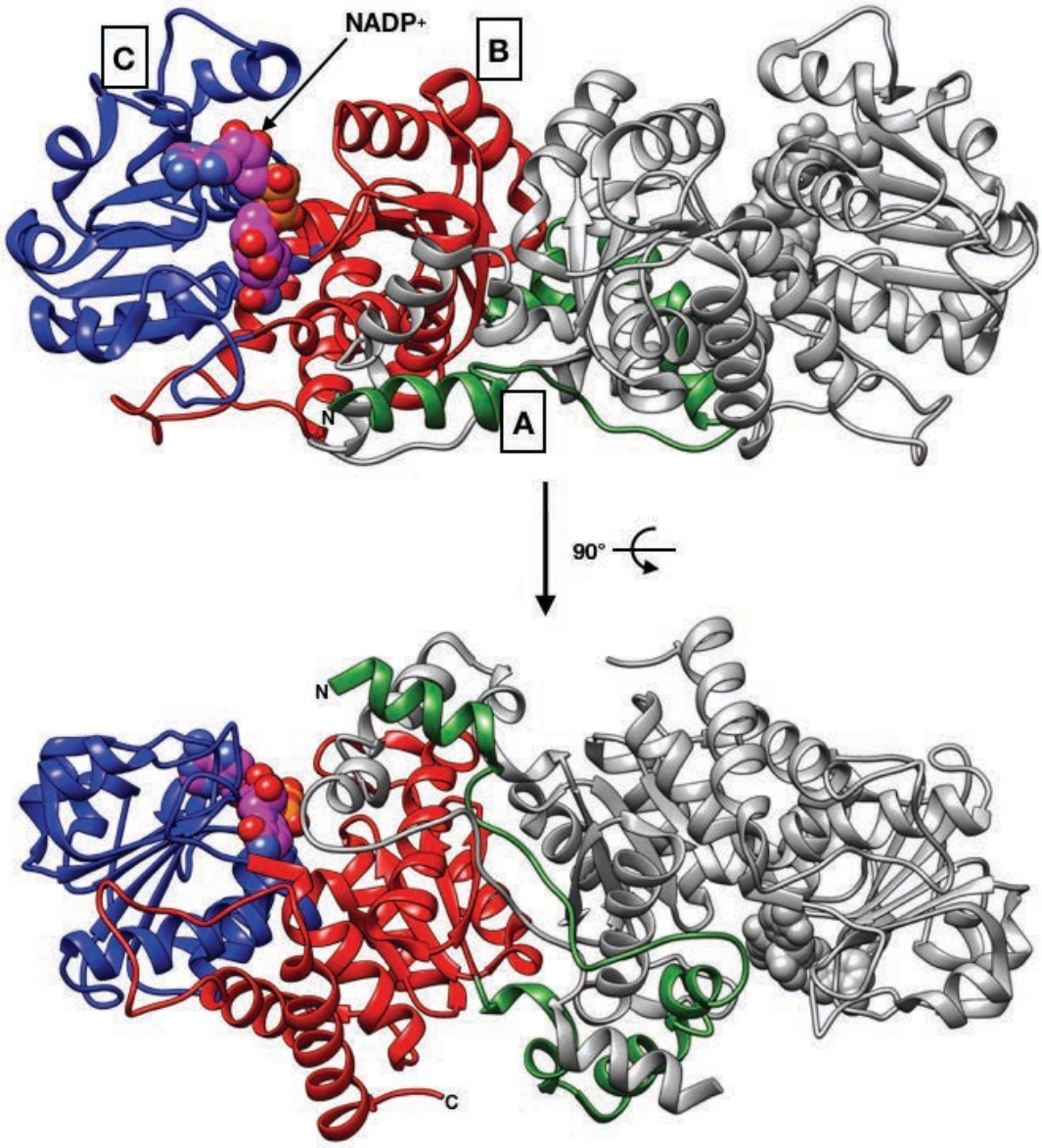


Figure 54. Prokaryotic ME (AYWB-ME)
The structure of AYWB-ME (PDB:5CEE) in two orientations. AYWB-ME has a minimal ME fold. The fold of domain A is only associated with prokaryotic ME and forms the dimer interface (green). Domain B is the core of the ME structure (Red) and domain C is the NAD+ binding domain (Blue). AYWB-ME forms a tightly interacting homodimer (monomer1 = coloured, monomer2 = grey) that is stabilised mainly by interactions between residues of domain A. A NAD+ ligand is shown bound in substrate binding cleft between domains B and C (magenta) [1].

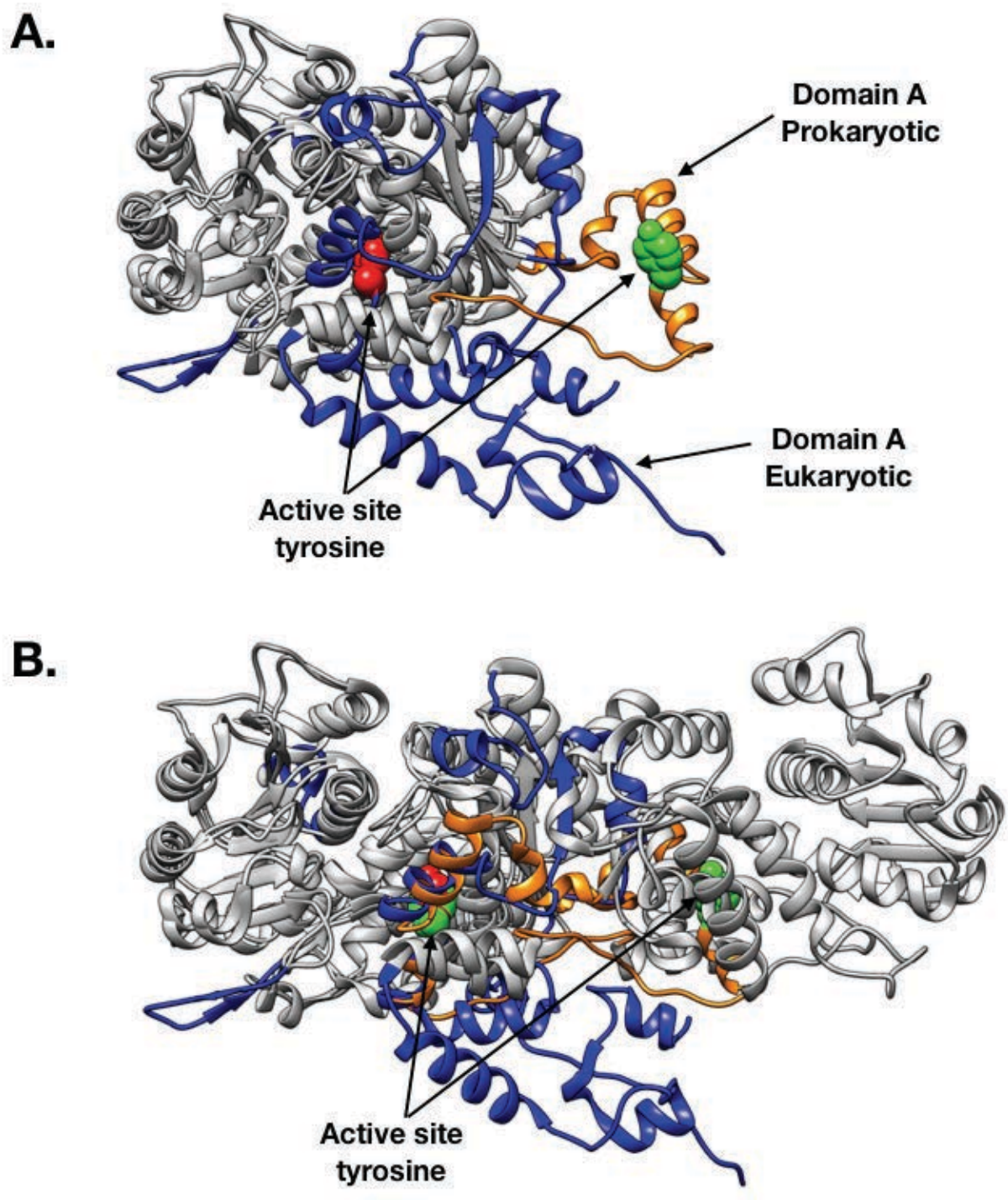


Figure 55. Comparison between prokaryotic and eukaryotic ME
The structures of AYWB-ME (PDB:5CEE) and human-ME (PDB:1QR6) have been superimposed using SSM superpose tool on coot (length = 320 residues, rmsd = 1.97Å). These two structures are typical of prokaryotic and eukaryotic ME structures respectively. **A.** The structural differences between the two structures are highlighted in orange (prokaryotic only) and blue (eukaryotic only). The core fold of the proteins is highly similar between the two MEs therefore has been coloured grey. Domain A has a significantly different fold between prokaryotic and eukaryotic ME. Furthermore the position of the active site tyrosine is located at different position between the two monomers (red = 1QR6, green= 5CEE). **B.** The active site of prokaryotic MEs is composed of residues from both monomers of the dimer complex. When the second monomer of the 5CEE dimer is displayed the location of the active site tyrosine from the other monomer is positioned into the active site (green and red tyrosine residues overlap).

5.1.3 Phosphotransacetylases (Pta)

5.1.3.1 Overview

Phosphotransacetylases (Pta) play an integral role in acetate metabolism in fermentive anaerobes. Pta catalyses the transfer of the acyl group to Coenzyme A (CoA) from acetyl-phosphate (acetyl-P) to form inorganic phosphate (Pi) and acetyl-CoA. Acetyl-CoA is a key central metabolite used directly in anabolic processes, such as the delivery of the acetyl group to oxaloacetate, forming citric acid, which is an important step in the Krebs cycle.

$$\text{Acetyl-CoA} + P_i \rightleftharpoons \text{CoA} + \text{Acetyl-phosphate}$$

Fermentive microbes utilise Pta and acetate kinases (Ack) in a metabolic pathway to generate energy. Ack reversibly catalyse the transfer of an acetyl-group from acetyl-P to ADP to create a molecule of ATP [242].

$$\text{Acetyl-phosphate} + \text{ADP} \rightleftharpoons \text{Acetate} + \text{ATP}$$

Pta enzymes play important roles in the “acetate switch”, a phenomenon where cells transit from a rapid growth state that produces acetate to one of a slower growth state that utilises acetate as an energy source. Acetate is regarded as a major intermediate in the global carbon cycle [243].

Despite the importance of Pta in anaerobic metabolism as well as their widespread distribution, their catalytic mechanism remains poorly understood. J Henkin et al.

Pta enzymes form a dimer complex, adjoined centrally by interactions between domain-II. The dimer interface is stabilised by two helices that align almost parallel at the centre (Helix-12 from both monomers). Residues from Helix-12 also interact with residues from Helix-9 and Helix-10 of opposing monomers, further stabilising the central core of the dimer [250].

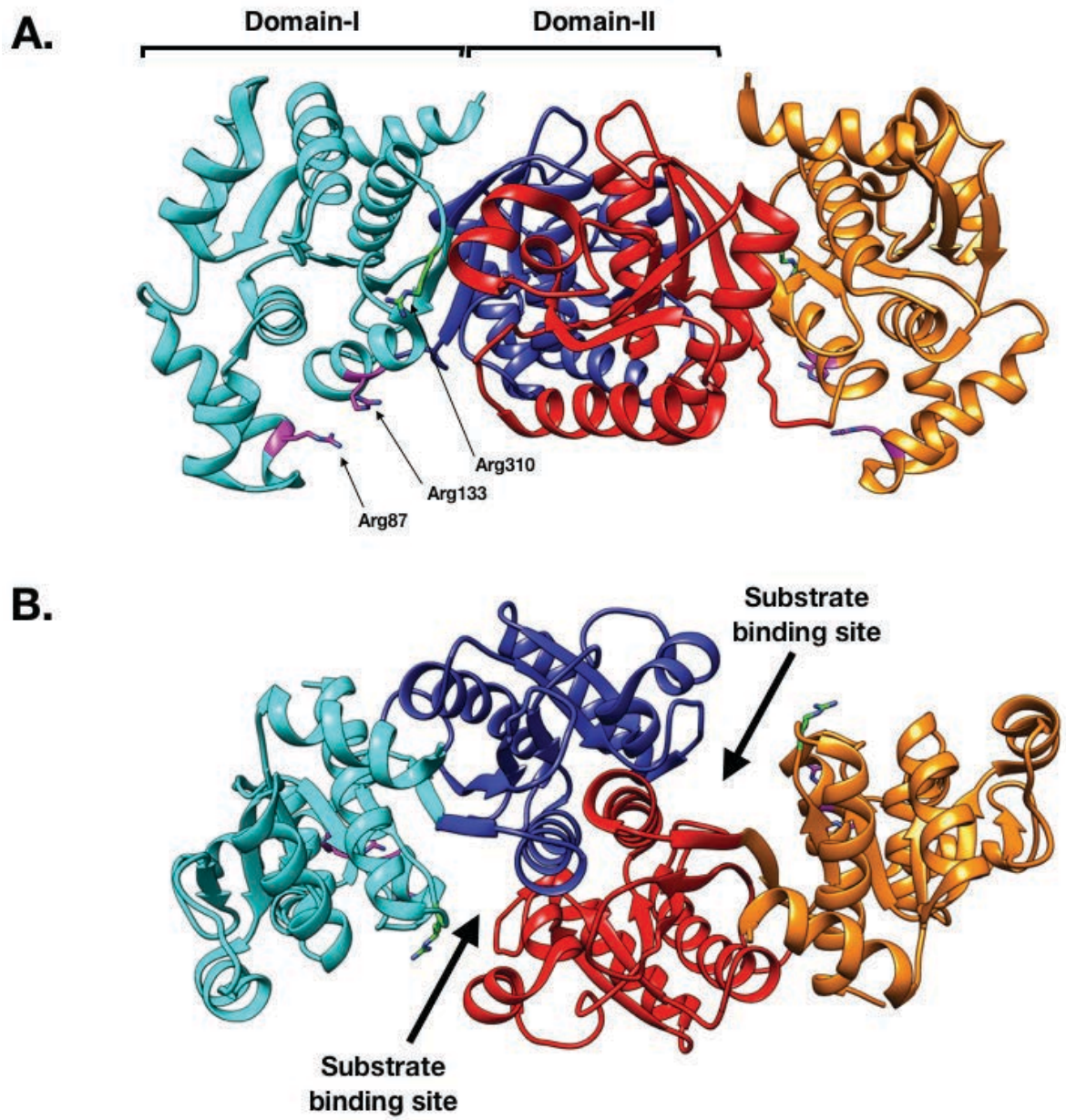


Figure 56. Pta structure
A. Pta enzymes form a bi-lobed homodimer. The dimer is held in place by interactions occurring between residues of the dimerization domain (domain-II – blue and red). The location of the conserved Pta arginine residues believed to be involved in substrate binding are displayed. Arginine residues coloured in magenta interact with CoA, whereas the arginine residue in green interacts with acetyl-P. **B.** A 90° rotation around the x-axis highlights the almost parallel alpha helix that contributes to the dimer interface. Between domain-I (cyan) and domain-II (blue) is a large substrate binding cleft – marked with black arrows.

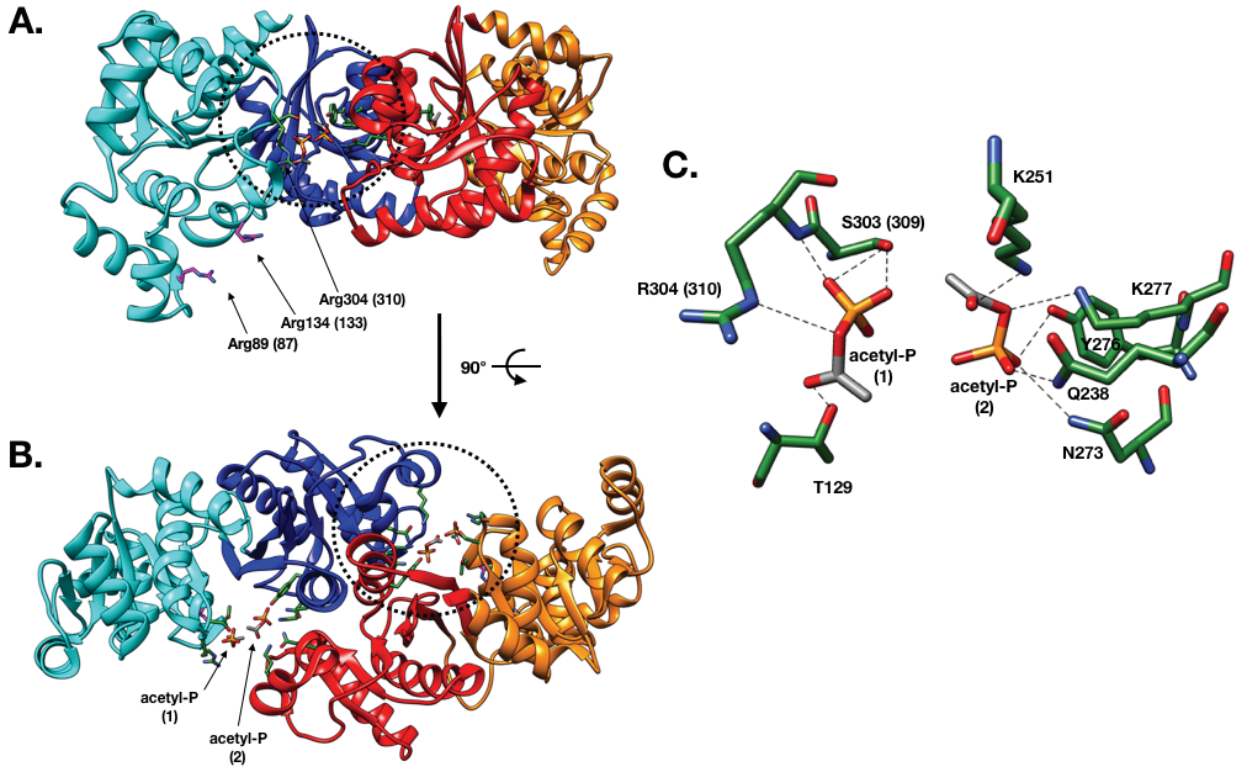


Figure 57. Acetyl-P binding site
A. The acetyl-P binding site is highlighted by a dashed black circle. **B.** The acetyl-binding site viewed from an angle 90° around the x-axis. **C.** A close up view of the residues involved in binding acetyl-P. Two acetyl-P molecules are shown to bind to the active site. The conserved arginine, R304 (R310 in *Methanosarcina thermonphila*) makes a hydrogen bond with acetyl-P(1). (PDB:1XCO)

Two Pta:CoA complex structures for *M. thermophila* Pta (PDB: 2AF3 and 2AF4) have been published [247]. CoA was observed to interact with the Pta enzyme at two distal sites (CoA^1 and CoA^2) in the co-crystallised and soaked structures (Figure 58). One CoA site (CoA^1) was subjectively chosen as the higher affinity binding site (CoA observed at this location in both ligand complex structures) and catalytic residues R310, S309 and D316 were assigned based on their proximity to the bound CoA thiol group [247]. ITC experiments confirmed two CoA binding sites with disparate K_D values of 20 µM and 1mM. The lower affinity binding site is likely non-specific as the cellular concentration of CoA is unlikely to approach concentrations above 1mM [112]. CoA^1 binding site

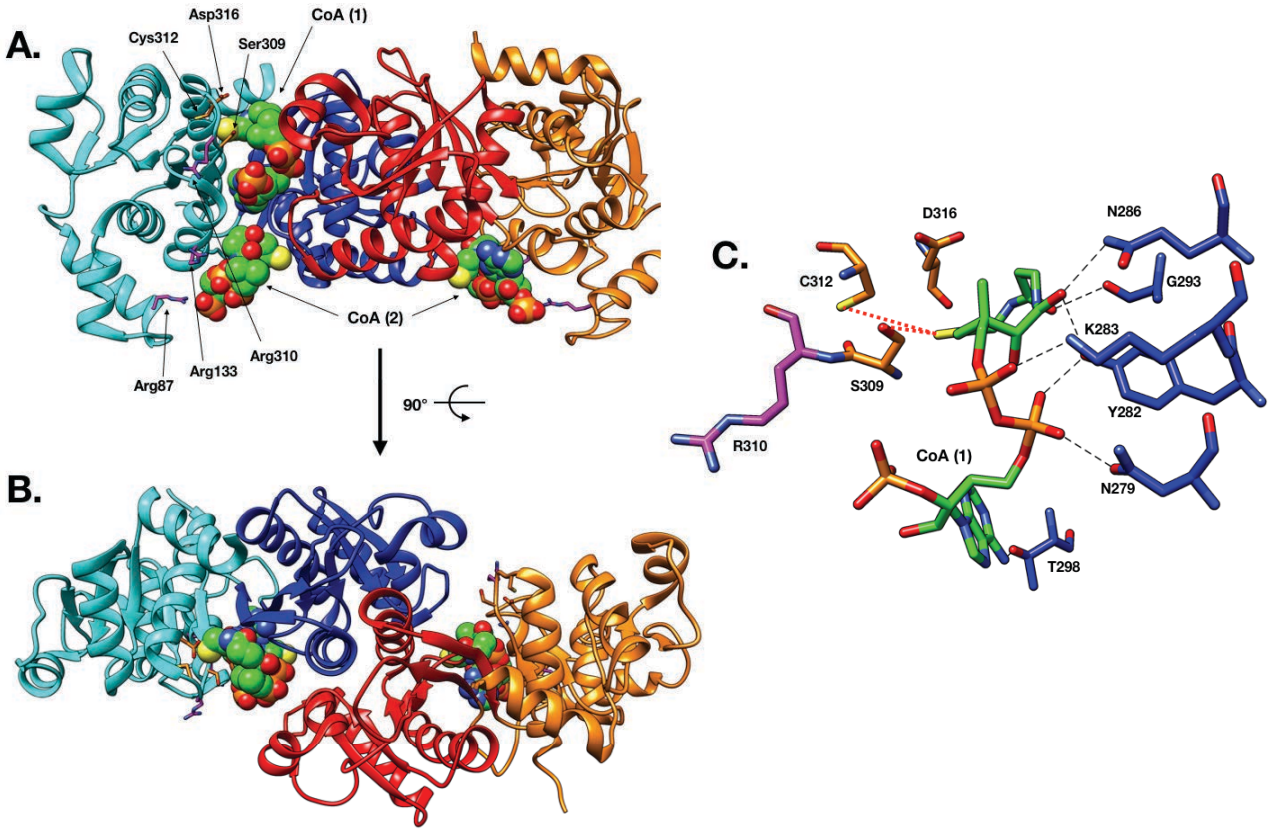


Figure 58. Pta CoA binding site
A. Structure of Pta dimer showing the location of CoA ligands (green) bound in the substrate binding groove. Two CoA ligands are shown to bind to the first monomer and one CoA ligand to the second monomer of the dimer complex. CoA1 and CoA2 sites are indicated. The location of the active-site residues and the conserved arginine residues that are shown to interact with CoA (R67, R133) are also highlighted (magenta). **B.** Rotation of image A by 90° around the x-axis. **C.** A close up view of the residues involved in the binding site of CoA1. Residues in blue are involved in binding the CoA substrate and the H-bond interactions highlighted by black dashed lines. Active-site residues are shown in orange and the interactions (>3.2Å) made with the thiolate group of CoA are shown in red. The conserved R310 that is believed to position acetyl-P is also shown. (PDB:2AF3)

5.1.3.5 Catalytic Mechanism

The currently proposed catalytic mechanism is based on the CoA bound structures. D312 is proposed to act as a general base and abstract the sulfhydryl proton from CoA, forming a negatively charged transition state that is stabilised by S309. The activated transition state then attacks the acetyl-P substrate, which is held in position by R310 [247]. Although the proposed mechanism is plausible it is not in agreement with biochemical and mutational studies. The mechanism does not address how R133 and R87 affect the $K_{M(CoA)}$ if they do not facilitate binding. Nor does this address how R310

binds and orientates acetyl-P if it makes hydrogen bonds to the 3'-phosphate group of CoA or why R310 mutation only displays a decrease in $K_{M(acetyl-P)}$. Furthermore R310 is orientated incorrectly to deliver acetyl-P to the active site (a concern of the authors in Lawrence et al.) [247].

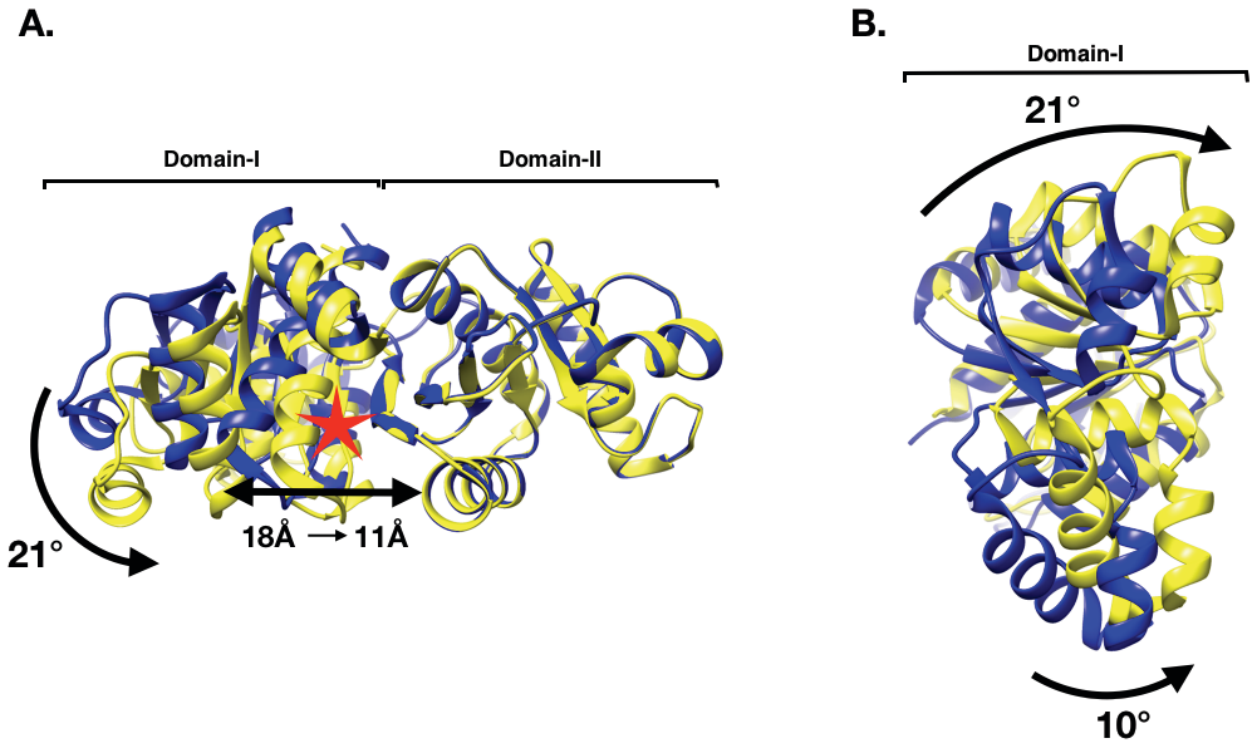


Figure 59. Pta domain movement
A. Domain I adopts multiple conformations in Pta structures, rotating as a rigid body to open and close the substrate binding cleft, located between domain I and II. A 21° swing of domain-I constricts the substrate binding groove from 18Å to 11Å. The movement is believed to be a mechanistic feature of substrate binding. **B.** The same rigid body rotational conformational change in domain-I viewed from the side. The upper portion of the domain undergoes a larger rigid body movement than the lower portion. The LSQ superpose tool in coot was used to align domain II (residues 149-299) of PDB:1QZT (length = 1114 atoms, rmsd = 1.97 Å)

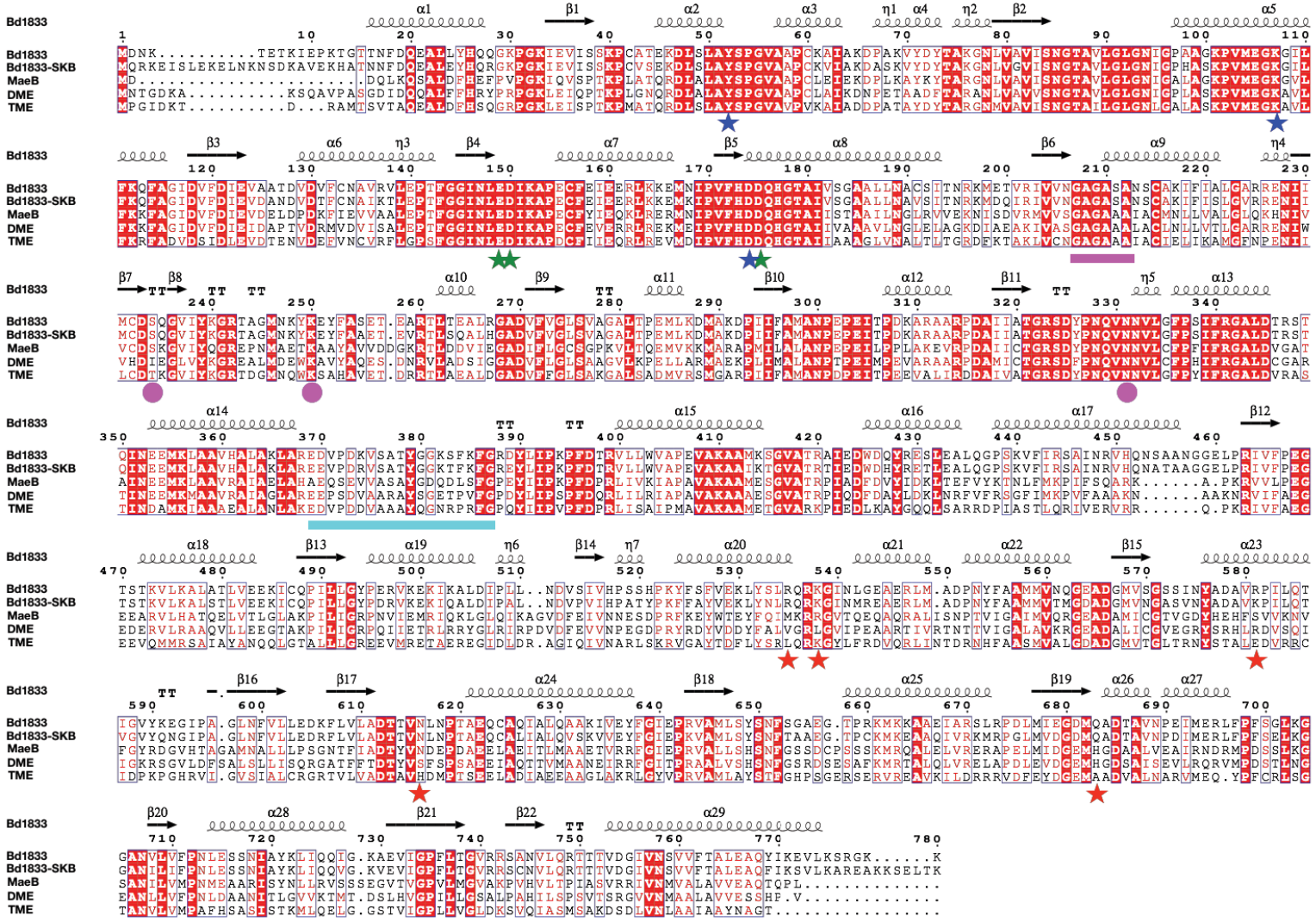


Figure 60. Bd1833 alignment with multi-modular malic enzymes

Catalytic residues are highlighted with blue stars. Cation-binding residues highlighted with green stars. Residues that hydrogen bond with $NADP^+$ are highlighted with pink dots and the GXGXXG motif with a pink bar. Residues with side chains that hydrogen bond with acetyl-CoA are highlighted with red stars. The cyan bar shows the location of subdomain D (multi-modular ME specific loop)

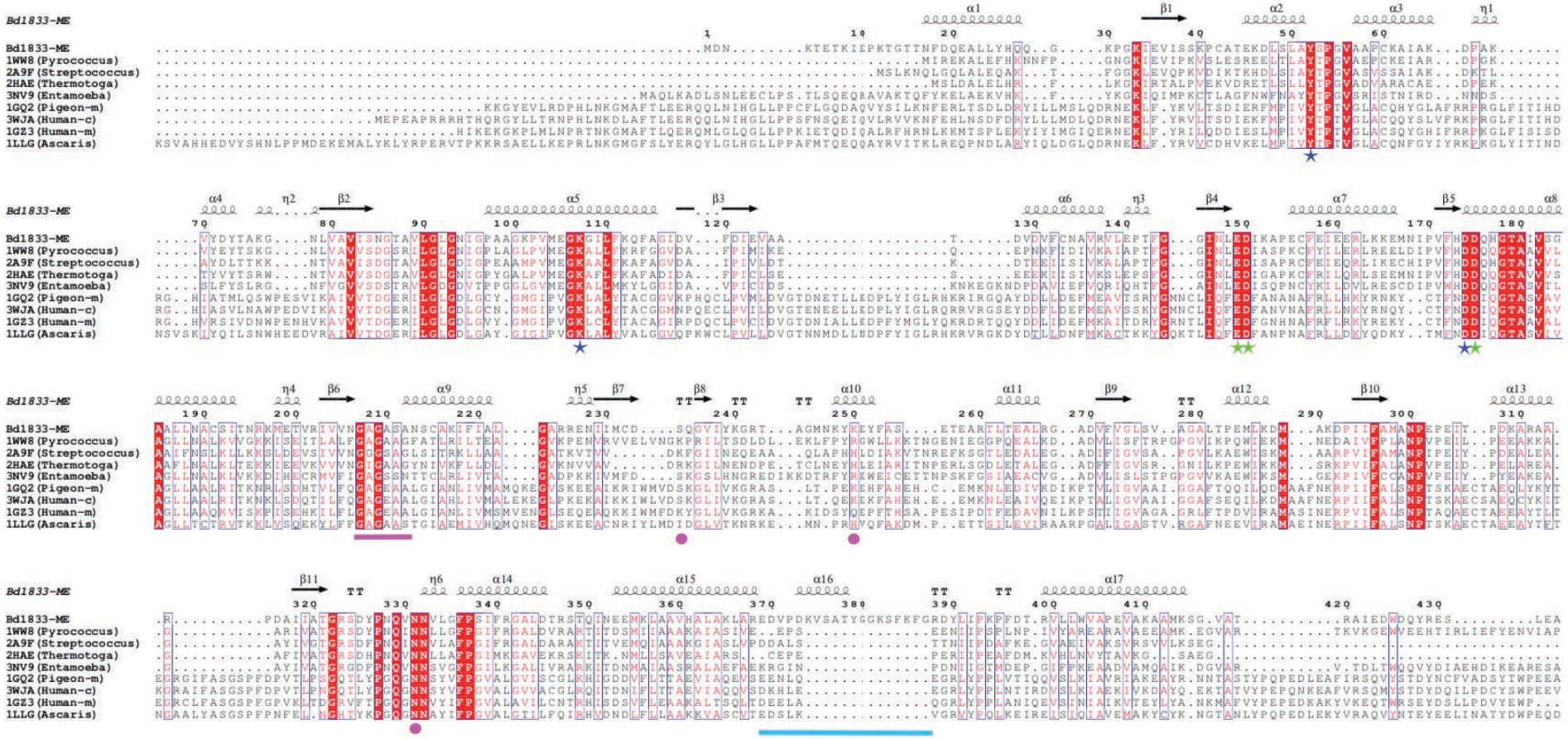


Figure 61. Bd1833-ME alignment with prokaryotic and eukaryotic ME
Catalytic residues are highlighted with blue stars. Cation-binding residues highlighted with green stars. Residues that hydrogen bond with NADP⁺ are highlighted with pink dots and the GXGXXG motif with a pink bar. The cyan bar highlights subdomain-D (specific for multi-modular malic enzymes).

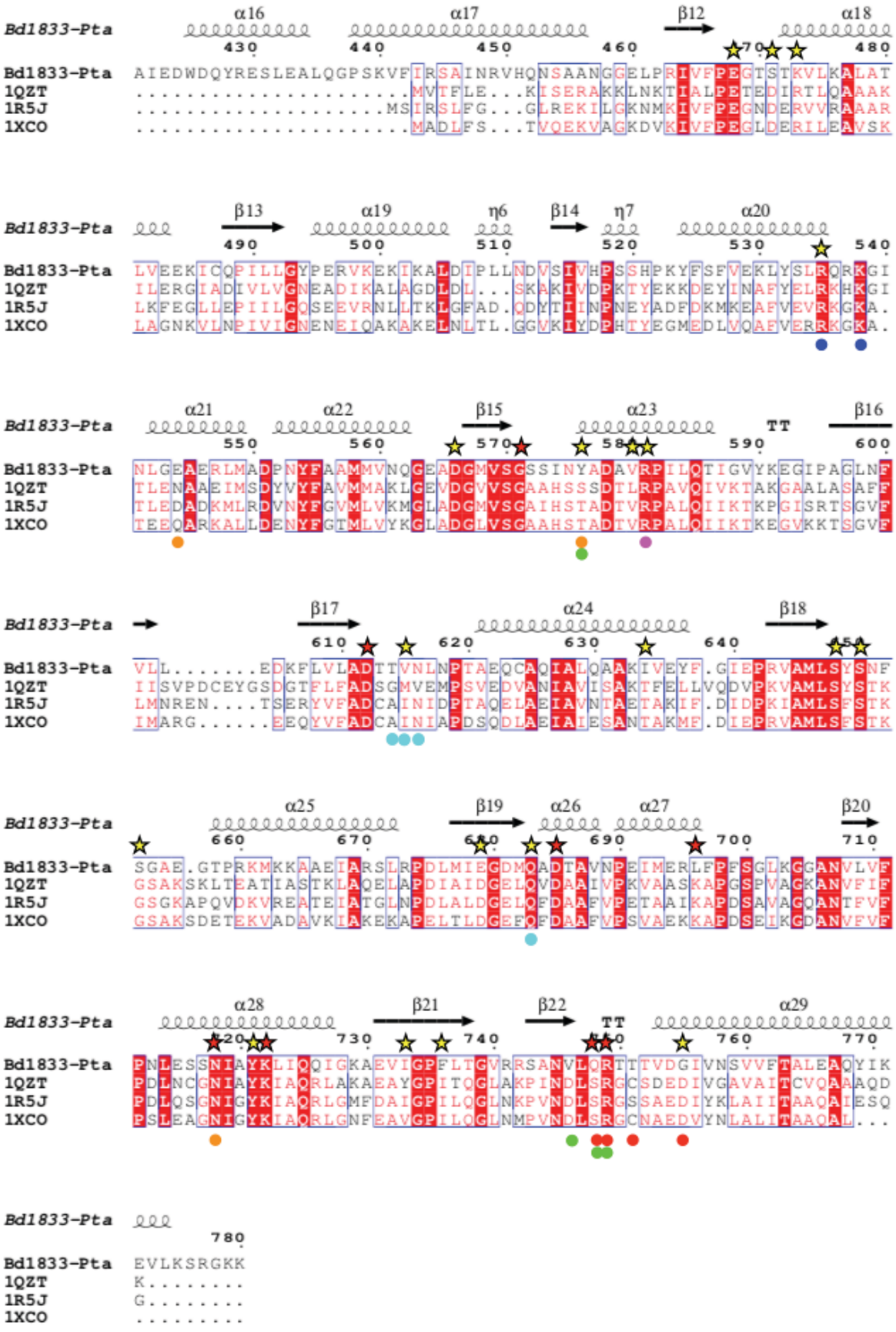


Figure 62. Bd1833-Pta alignment with Pta enzymes
The highly conserved and strictly conserved residues associated with Pta enzymes are highlighted by the yellow and red stars respectively. Dark blue circle – residues interacting with CoA 3'phosphate group; pink circle – residues interacting with CoA pyrophosphate group; light blue circle – other residues hydrogen bonding to CoA; orange circle – key additional residues highlighted from Bd1833 studies. Red circles are the previously proposed catalytic residues and the green circles are the newly proposed catalytic residues.

5.2.2 Bd1833 purification and crystallisation

Three constructs were designed for Bd1833: full-length (Bd1833), ME domain only (Bd1833-ME) and Pta domain only (Bd1833-Pta). Each construct of Bd1833 was successfully cloned, purified, crystallised and the structures solved. Figure 63 shows a representative purification gel and crystals, which were used to solve each construct.

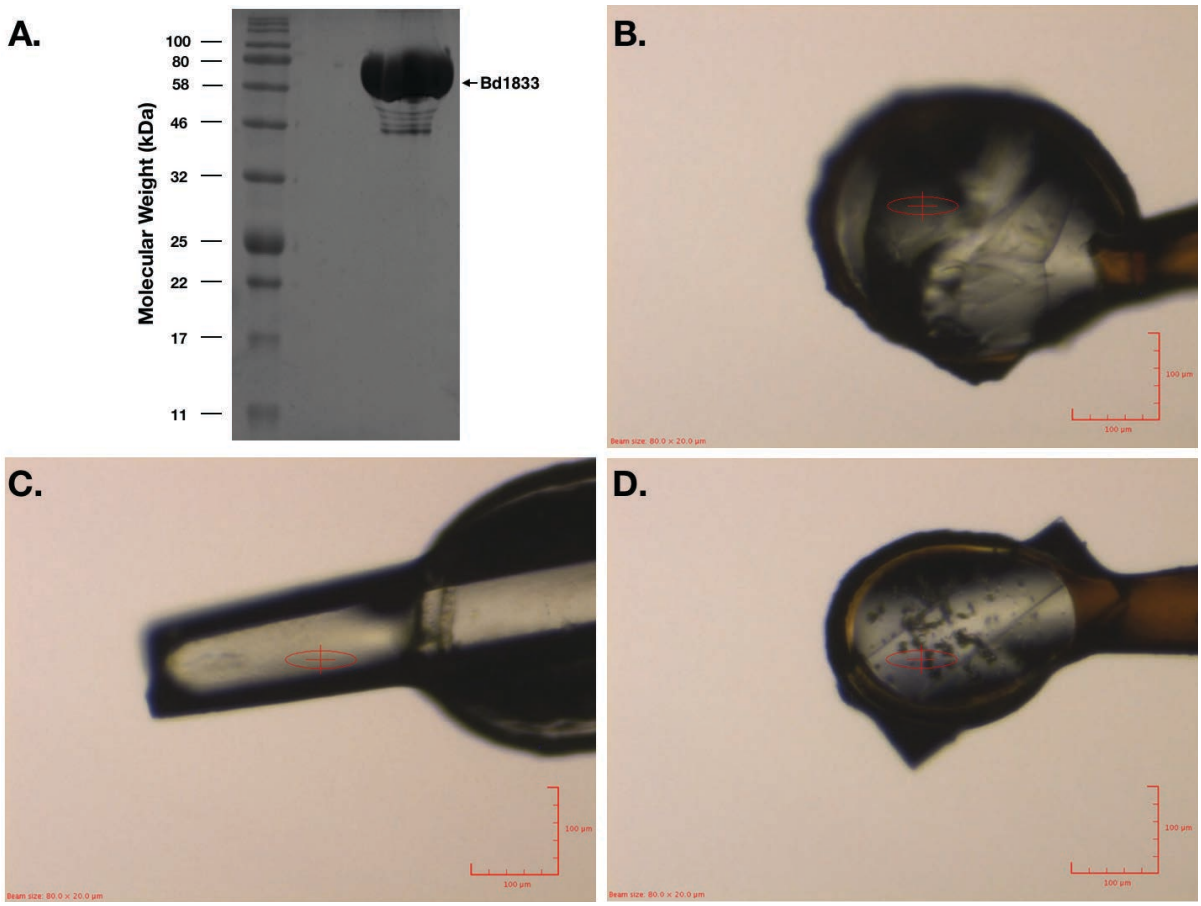


Figure 63. Bd1833 Purification and Crystallisation
A. SDS PAGE gel showing purified full-length Bd1833 **B.** Crystal of Bd1833 used for data collection **C.** Crystal of Bd1833-Pta used for data collection. **D.** Crystal of Bd1833-ME used for data collection

5.2.3 Bd1833 Structure

We have solved the three-dimensional structure of full-length Bd1833, its constituent domains individually, mutant forms and ligand complexes to aid in the determination of function. A list of Bd1833 structures solved and an associated brief description is presented in Table 13, which is followed by more in depth structural analysis in the following individual sections.

	Construct	Amino acids traced	Spacegroup	Resolution (Å)	Copies in asymmetic unit	R-work	R-free	Oligomeric state	Ligand?	Description/Rationale
Bd1833-ME (Bcs-B6)	1- 434	15-422	P 2_1	1.67	2	0.1702	0.2039	Dimer	$Mg2^+$	High resolution structure of ME domain
Bd1833-ME (Pact-E1)	1- 434	16-427	P 1	2.2	6	0.2109	0.2264	Dimer	-	Lacking divalent cation
Bd1833-ME ($NADP^+$)	1- 434	15-425	P 2	1.67	2	0.1799	0.2159	Dimer	$Mg2^+$ $NADP^+$	NADP+ complex
Bd1833-Pta	439- 780	439- 776	P 2_1	1.96	6	0.1829	0.2052	Hexamer	acetyl-CoA	High resolution structure of Pta domain, inhibited state
Pta R535A	439- 780	439- 776	C 2	2.35	6	0.2158	0.2368	Hexamer	-	Ligand free structure, 3'-P mutant
Pta R535E	439- 780	439- 776	C 2	2.39	6	0.2153	0.2602	Hexamer	-	Ligand free structure, 3'-P mutant
Pta E544R	439- 780	439- 776	C 2	2.33	6	0.2069	0.2265	Hexamer	acetyl-CoA	Interupt hexamer interface
Pta Y576S	439- 780	439- 776	C 2	2.38	6	0.2222	0.2725	Hexamer	-	acetyl-CoA binding pocket distrupted
Pta-Apo	439-780	439-776	P 2_1 2_1 2_1	2.72	12	0.2520	0.2995	Hexamer	-	WT Ligand free structure that shows rigid body conformational change
Bd1833 (Full-length)	1-780	15-776	C 2	2.72	6	0.1922	0.2277	Hexamer	acetyl-CoA	Full length protein
Bd1833-SKB (full-length)	1-796	16-776	P 4_1 2_1 2	3.71	3	0.2503	0.2996	Hexamer	-	Full length of closely related homologue
Bd1833-SKB (Pta)	449-796	449-786	F 2 3	2.71	4	0.2040	0.2392	Hexamer	acetyl-CoA	Homologue Pta domain, CoA binding site differs

Table 13. Bd1833 structures
This table displays a brief structural analysis of each Bd1833 structure reported in this thesis and the rationale behind the structure.

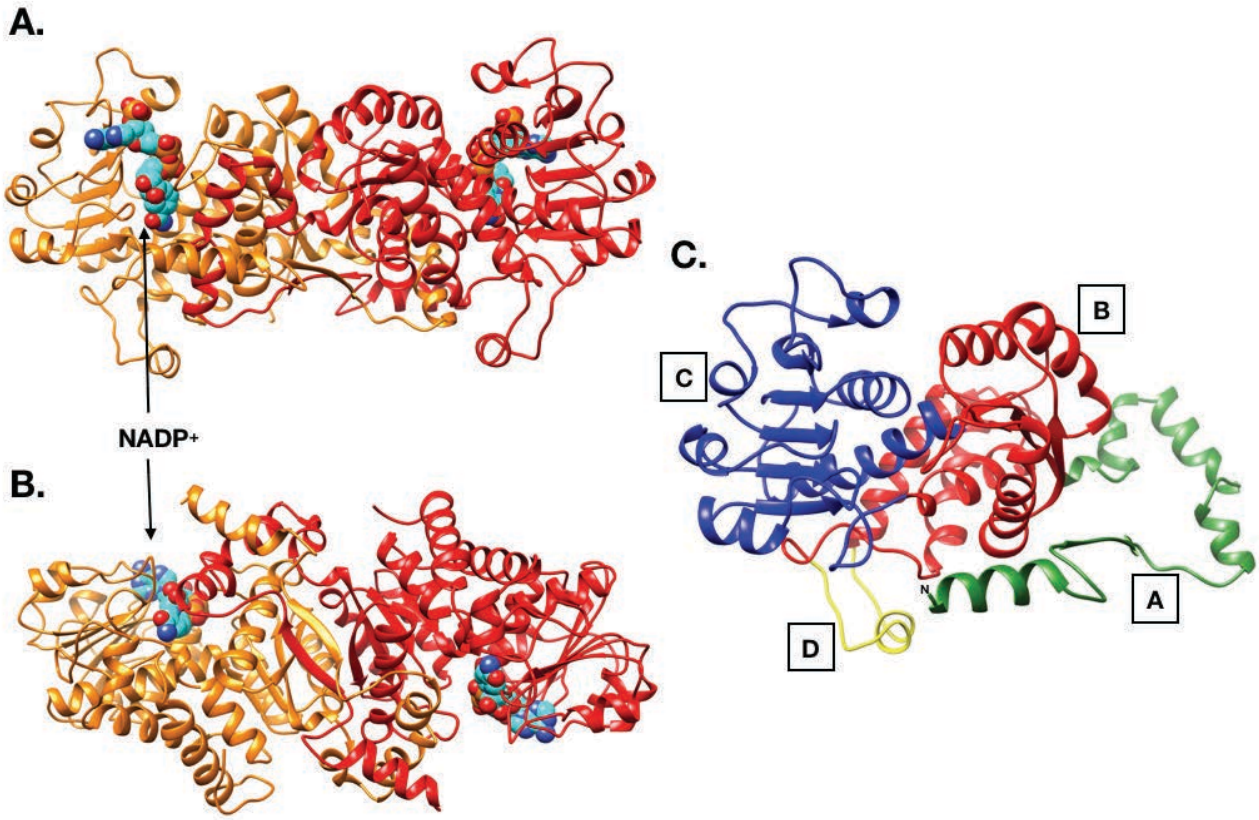


Figure 64. Bd1833-ME structure
A. Bd1833-ME forms a biological dimer (orange and red monomers). The active site is located in the groove between subdomains B and C, which is highlighted by the $NADP^+$ molecule (cyan) bound to subdomain-C. **B.** Illustrates the tightly interacting dimer interface, viewed from a 90° rotation of A around the x-axis. **C.** Illustrates the subdomains that make up Bd1833-ME. Subdomain-A, providing the dimer interface, residues 16-78 (green). Subdomain-B, residues 79-173, 332-368 and 387-420 (red) form the core of the protein containing the divalent cation-binding site. Subdomain-C, residues 174-331 (blue) forms the $NADP^+$ -binding site possessing a Rossmann fold. Bd1833-ME has a small loop specific to multi-modular malic enzymes, subdomain D 369-386 (yellow).

5.2.4.2 ME catalytic site

The catalytic site of Bd1833-ME is positioned between subdomains-B and C, which is highlighted by the position of $NADP^+$ in Figure 64. The catalytic pocket of Bd1833 is composed of a divalent cation-binding motif, a nucleotide-binding site and the substrate-binding site. The divalent cation-binding motif is central to the catalytic site and six oxygen atoms coordinate Mg^{2+} in an octahedral fashion (Figure 65). The coordinating oxygen atoms are provided by the sidechain carboxylate groups of E143

residues from both monomers of the dimer complex. The catalytic residue, Y52, located on subdomain-A, is provided by the other monomer. A single catalytic site containing residues from both monomers is typical of prokaryotic MEs [1] but not of eukaryotic ME. Moreover the proposed substrate-binding site residues are also slightly different to eukaryotic MEs, for example lacking the active-site arginine that assists in substrate binding [225].

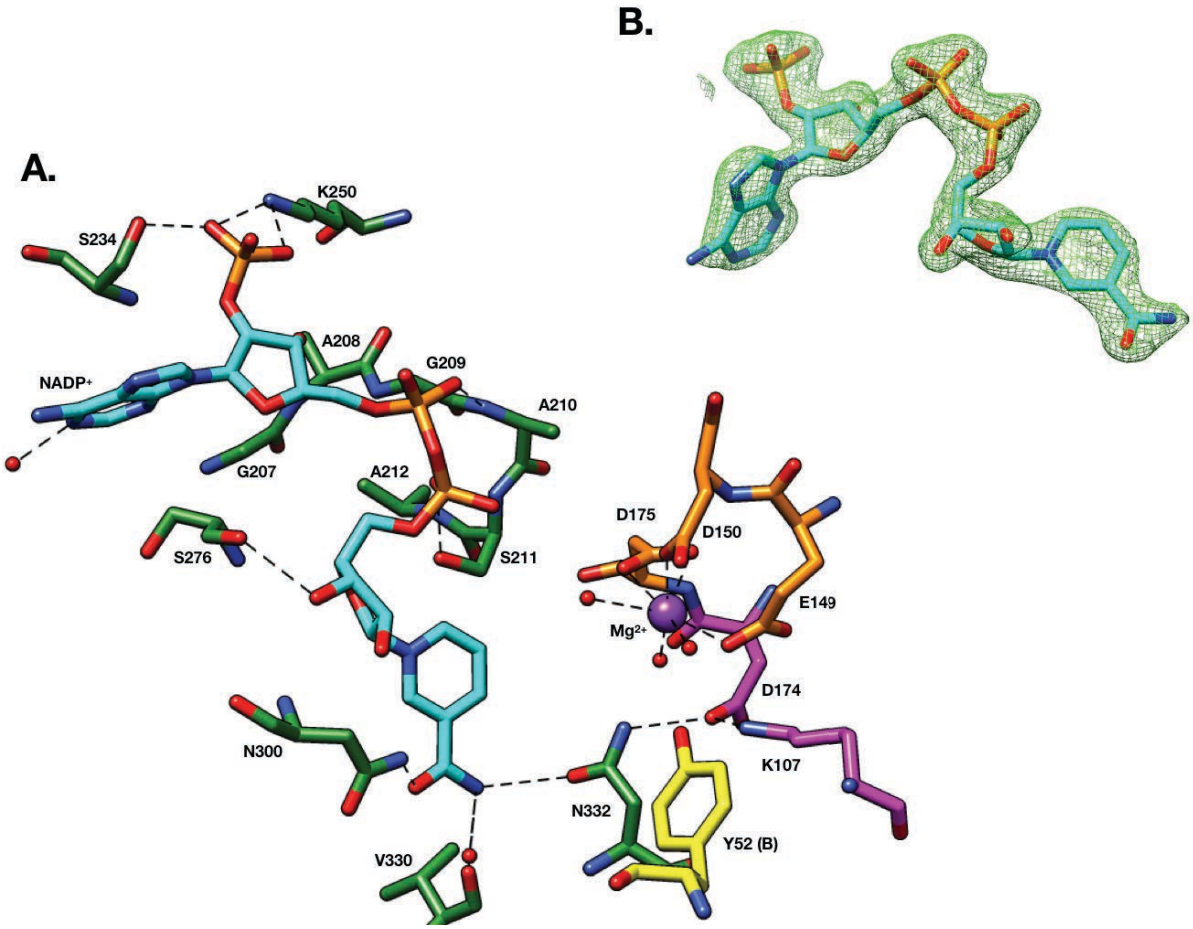


Figure 65. Bd1833 active site
A. The active site of Bd1833-ME is made up of residues from both monomers. The $NADP^+$ -binding groove located between subdomains B and C. Subdomain-C contains the Rossmann fold and GXGXXG nucleotide-binding motif. The residues that make hydrogen bond contacts with $NADP^+$ (cyan) are shown in green. The active-site residues are shown with magenta coloured carbons, while Y52 provided by the other monomer is shown with yellow carbons. The metal coordinating residues are shown in orange, bound to the divalent cation (Mg^{2+}). All residues shown, except for Y52, belong to a single monomer (chain A). All hydrogen bonds illustrated by dashed black lines. **B.** The electron density observed for $NADP^+$ in the Bd1833-ME: $NADP^+$ soaked crystal structure. ($2F_o$-F_c map, radius =2Å, σ = 1)

737) folds into a six-stranded mixed β-sheet that is sandwiched between two α-helices on either side. Subdomain-II is positioned centrally in the Pta dimer arrangement, forming the dimer interface. A shallow groove splits subdomain-I and II and is the location of the substrate-binding pocket where the co-purified and co-crystallised acetyl-CoA is located (Figure 67).

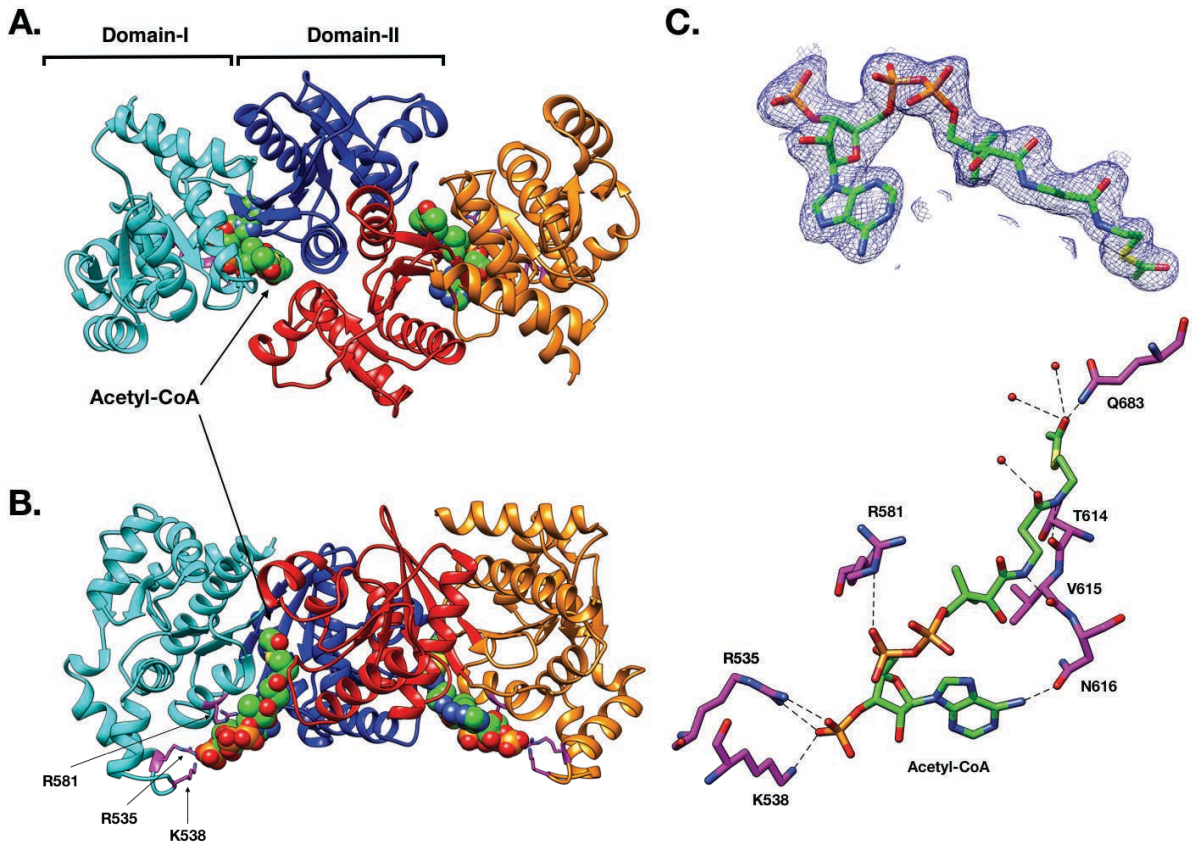


Figure 67. Bd1833-Pta dimer and acetyl-CoA binding
A. Bd1833-Pta forms a hexamer (trimer of dimers) where the dimers are structurally similarly to other previously solved Pta structures represented here as a blue:cyan and a red:orange dimer. Bd1833-Pta monomer can be split into two subdomains: subdomain-I, residues 440-595 and 734-776 (cyan) and the central subdomain-II, 596-733 residues (blue). Central to the dimer is an almost parallel alpha helix that is the main contributor to the dimer interface. Acetyl-CoA (green) binds in the substrate-binding groove located between subdomains-I and II. The conserved CoA binding residues (Pta enzymes) are highlighted in magenta making hydrogen bonds to acetyl-CoA. **B.** Rotation of image A 90° around the x-axis. **C.** The electron density observed for acetyl-CoA ($2F_o$-F_c map, radius=2Å, σ=1). **D.** Illustration of the hydrogen-bonding network associated with acetyl-CoA binding.

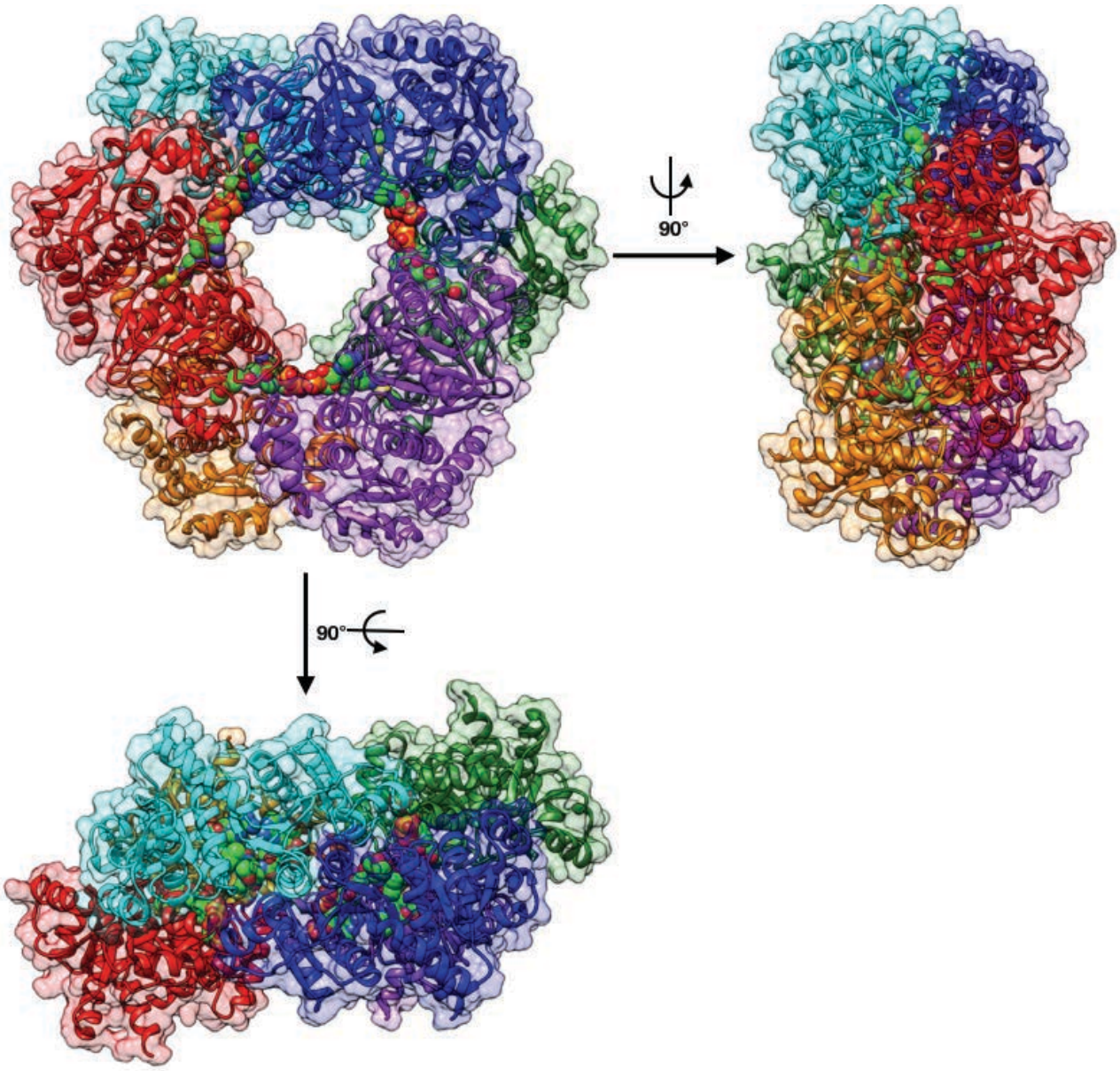


Figure 66. Bd1833-Pta structure
Bd1833-Pta forms a unique hexamer complex made of a trimer of dimers. The hexamer is organised into an offset triangle shape where the dimers form the flat sides. Dimers are coloured – blue:cyan, red:orange and green:purple. Each monomer makes contacts to four other monomers. The hexamer complex has six acetyl-CoA ligands bound, located on the inside face at the corners of the triangle formation. The hexamer complex is displayed in three orientations, rotated by 90°.

5.2.5.2 Dimer and tetramer interfaces

The dimer interface, 1707Å^2 buried surface area, is composed mainly of residues from subdomain-II. In particular two parallel α-helices (713-730) are positioned centrally to the dimer interface. Each monomer of the dimer provides a single α-helix to form the

parallel α-helix pair that makes substantial hydrophobic interactions to stabilise the dimer interface. In addition H-bonds further stabilise the dimer interface (Figure 68).

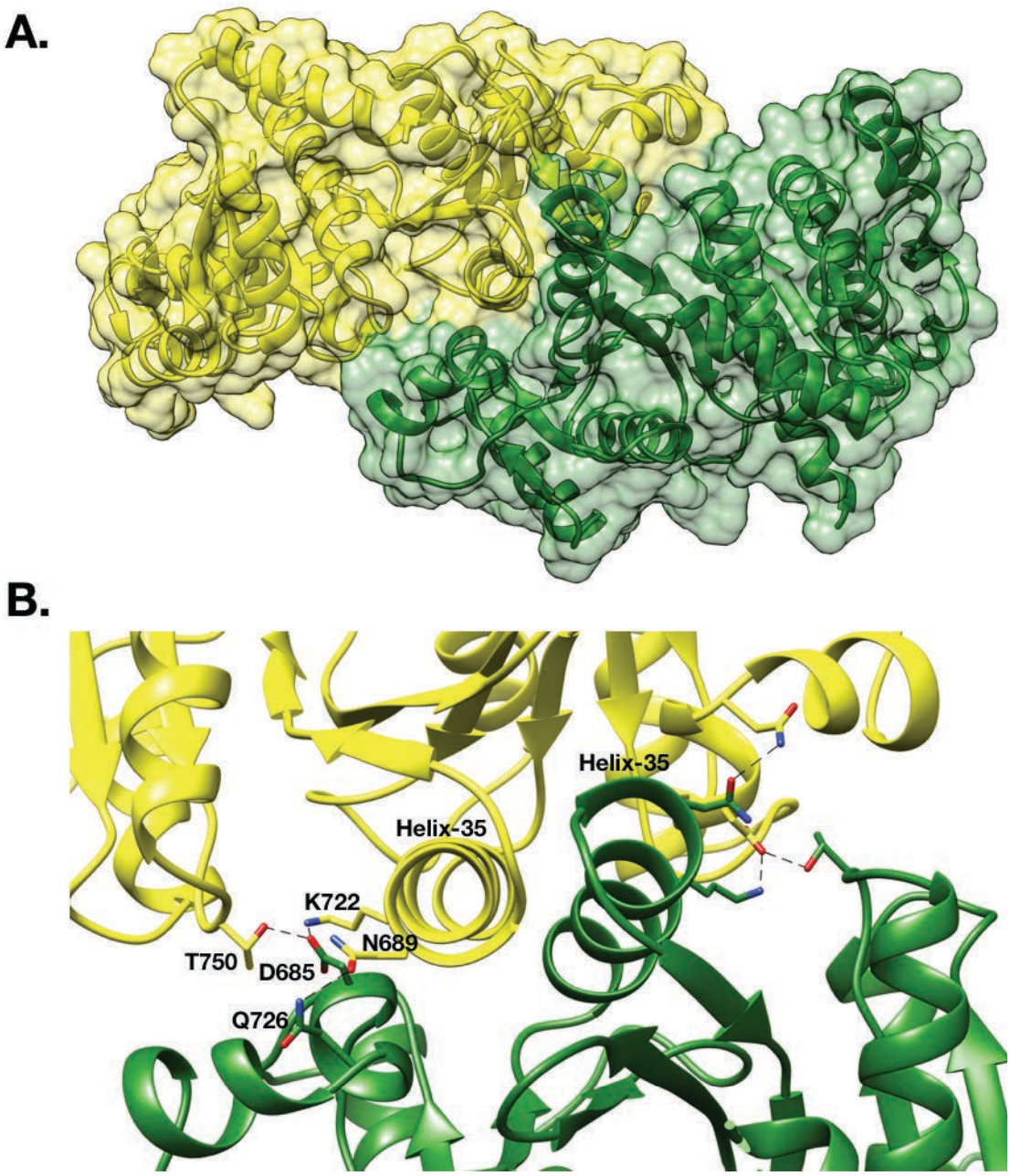


Figure 68. The dimer interface
A. Surface view of dimer interface. **B.** View of dimer interactions. The dimer interface is mainly formed by interaction between subdomain-II; a central alpha helix pair (helix-35) runs almost parallel and creates a large buried hydrophobic core. A number of hydrogen bonds are formed between residues from the central α-helix and other residues of subdomain I and II.

The hexamer complex, made of a trimer of dimers, is organised into an offset triangular shape, with dihedral D3 symmetry (3-fold symmetry with addition 2-fold axis). An extensive hexamer interface exists at the meeting point of two dimers at the triangle's vertices. Consequently a single Pta domain monomer makes interactions to four other monomers through dimer and hexamer interfaces. The hexamer interface covers numerous overlapping interactions at the triangle vertex, which extends up to the midpoint (Figure 69). The larger hexamer interface, at the vertex point (hexamer vertex contact) has a buried surface area of 644Å^2. The hexamer interface also includes two smaller contacts at the midpoint (hexamer midpoint contact) that has a buried surface area of 322Å^2 (buried surface area calculated by PDB:sum).

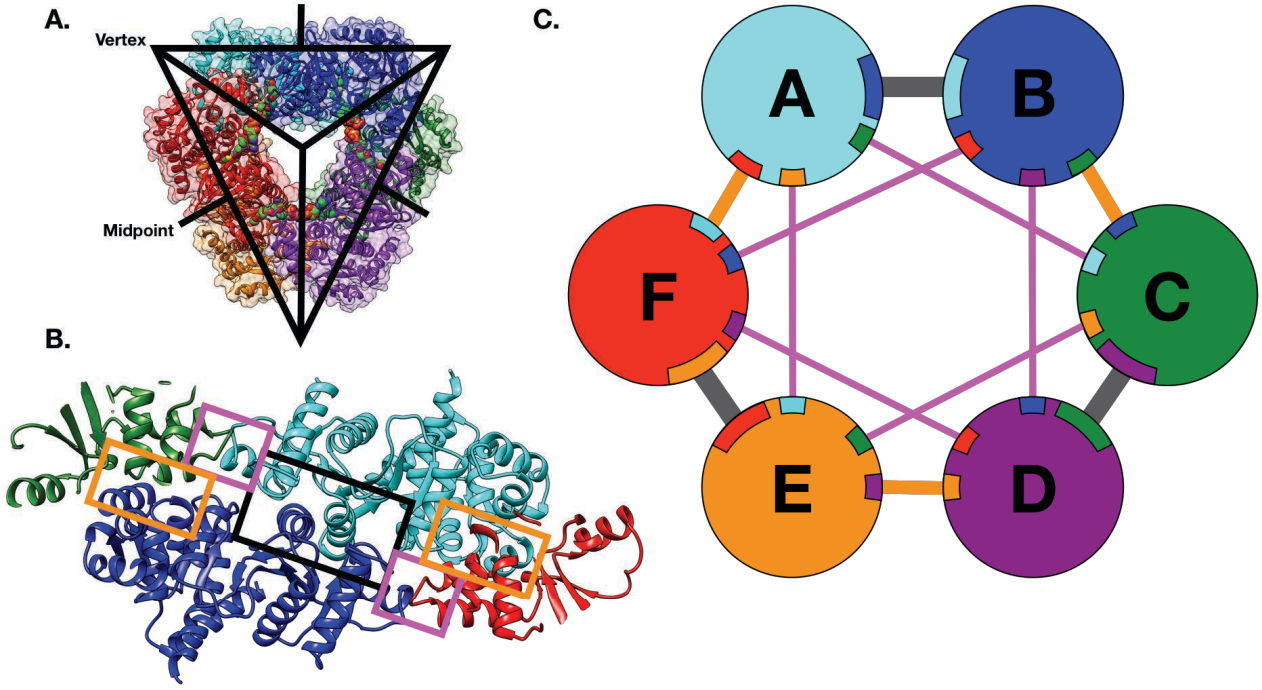


Figure 69. Hexamer interactions
A. Bd1833-Pta hexamer forms a triangular shape with dihedral D3 symmetry. **B.** Pta dimer viewed from the inside face demonstrating the hexamer surface contacts. The contacts in coloured boxes match the interactions shown in (C): Orange = larger hexamer vertex contact. Pink = smaller hexamer midpoint contact. Grey = dimeric contacts. **C.** A single monomer makes contacts to four other monomers through interactions at the dimer interface and hexamer interface. The hexamer interface is composed of four interacting monomers and two different surface contacts. The larger hexamer vertex contacts (orange line) are located at the hexamer verteces. The smaller hexamer midpoint contacts (pink lines) are made between the tips of subdomain-I helical flap region and the edge of subdomain-II.

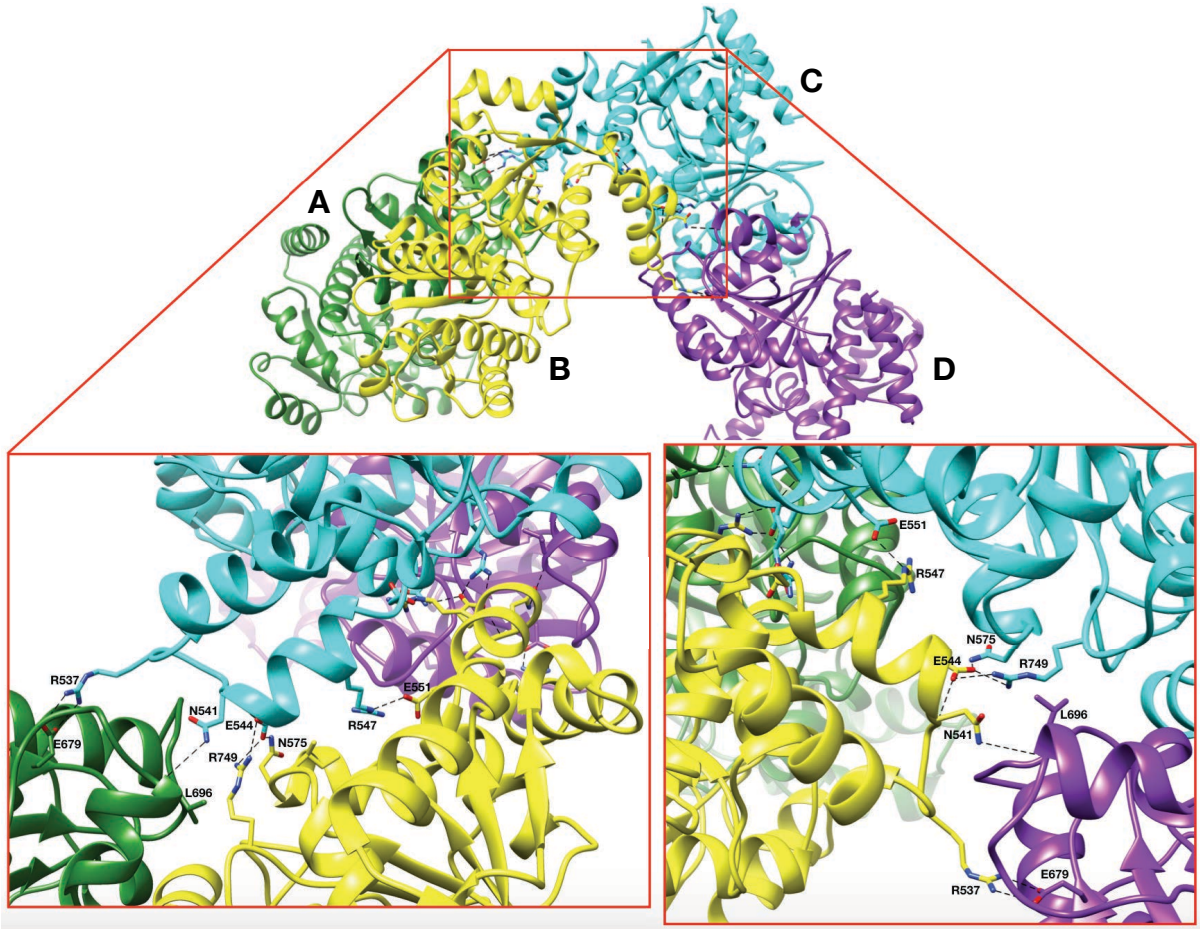


Figure 70. The hexamer interface
Bd1833-Pta forms a hexamer and each monomer forms a series of interactions with thee other monomers at the hexamer interface. This Figure illustrates a number of hydrogen bonds and ionic interactions that occur at the hexamer interface. The upper panel shows the overall hexamer interface. The left and right lower panels show a zoomed in view of the hydrogen bond network at the hexamer interface from two viewpoints. Chain A (green), Chain B (yellow), Chain C (cyan) and Chain (D) (purple). Chains A & B form a dimer and chain C & D form the other dimer that meet at the hexamer interface.

5.2.5.3 Acetyl-CoA binding site

Bd1833-Pta co-purifies and co-crystallises with acetyl-CoA, which binds in the shallow substrate-binding groove separating subdomain-I and II. One molecule of acetyl-CoA binds per monomer, hence the Pta hexamer complex has six molecules of acetyl-CoA ligand bound, located on the interior face at the triangle vertices. A number of non-bonding (vdw) interactions as well as specific hydrogen bonds are involved in acetyl-CoA binding, illustrated in Figure 67. The adenine group of acetyl-CoA makes a single hydrogen bond with the N616 sidechain. The side chain of R535 makes an important

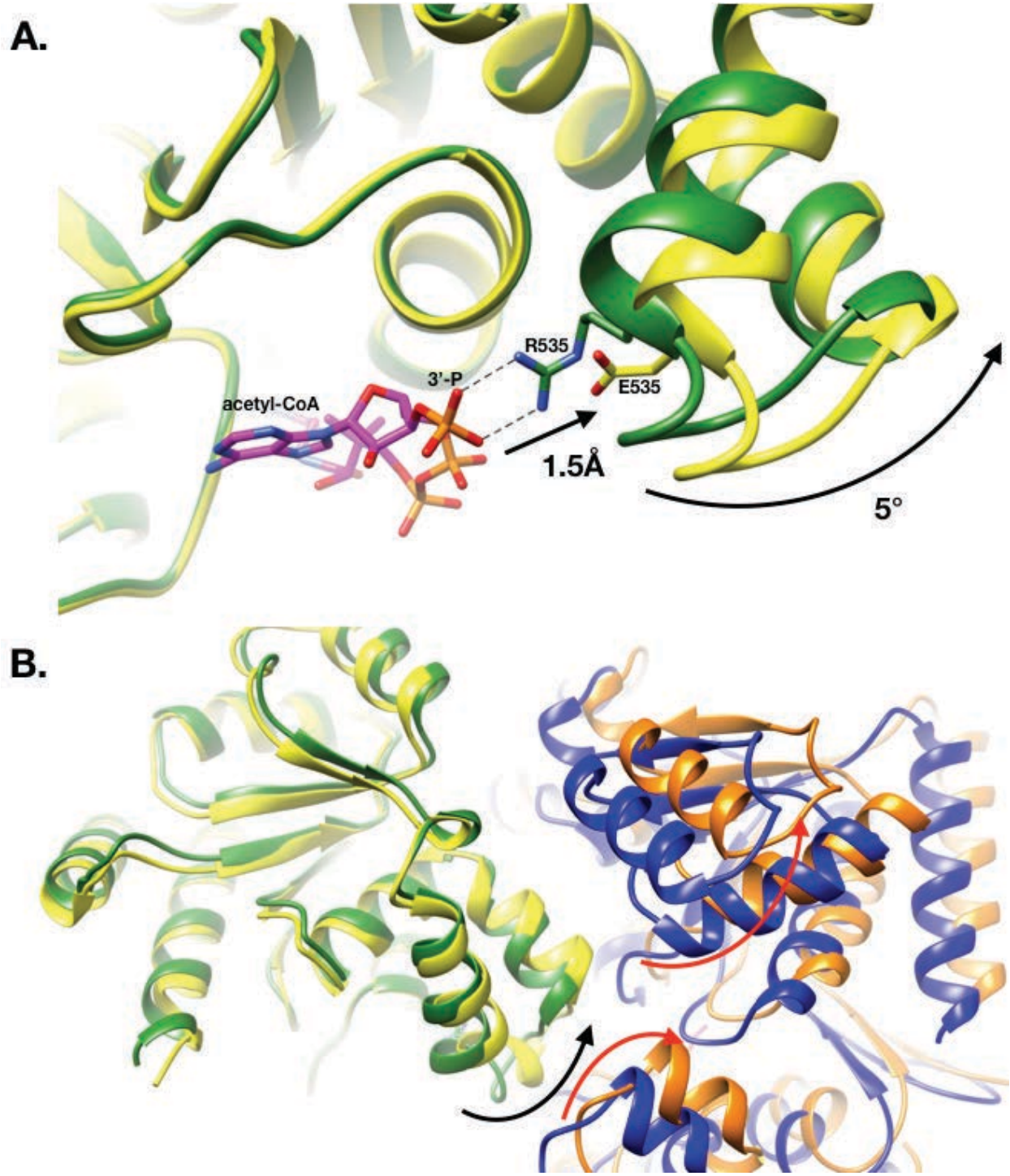


Figure 71. R535 mutation
A. Mutating R535 to a glutamate residue abolishes acetyl-CoA binding by repelling the 3'-phosphate group. R535E (yellow) mutation causes the lower loop (helix 27 and 28) of Bd1833-Pta to swing outwards when compared to WT- acetyl-CoA bound structure (green). **B.** The outwards swing (black arrow) causes a larger adjustment of the hexamer interface interactions (mainly lost from the vertex interface), shown by the red arrows. Blue is the position of hexamer interface in Bd1833-Pta (acetyl-CoA bound) and orange is the position of the hexamer interface in Pta-R535E mutant. A similar movement is observed for Bd1833-Pta R535A. The SSM superpose tool in coot was used to align chain A from Bd1833-Pta and Pta-R535E (length = 336 residues, rmsd = 0.67Å).

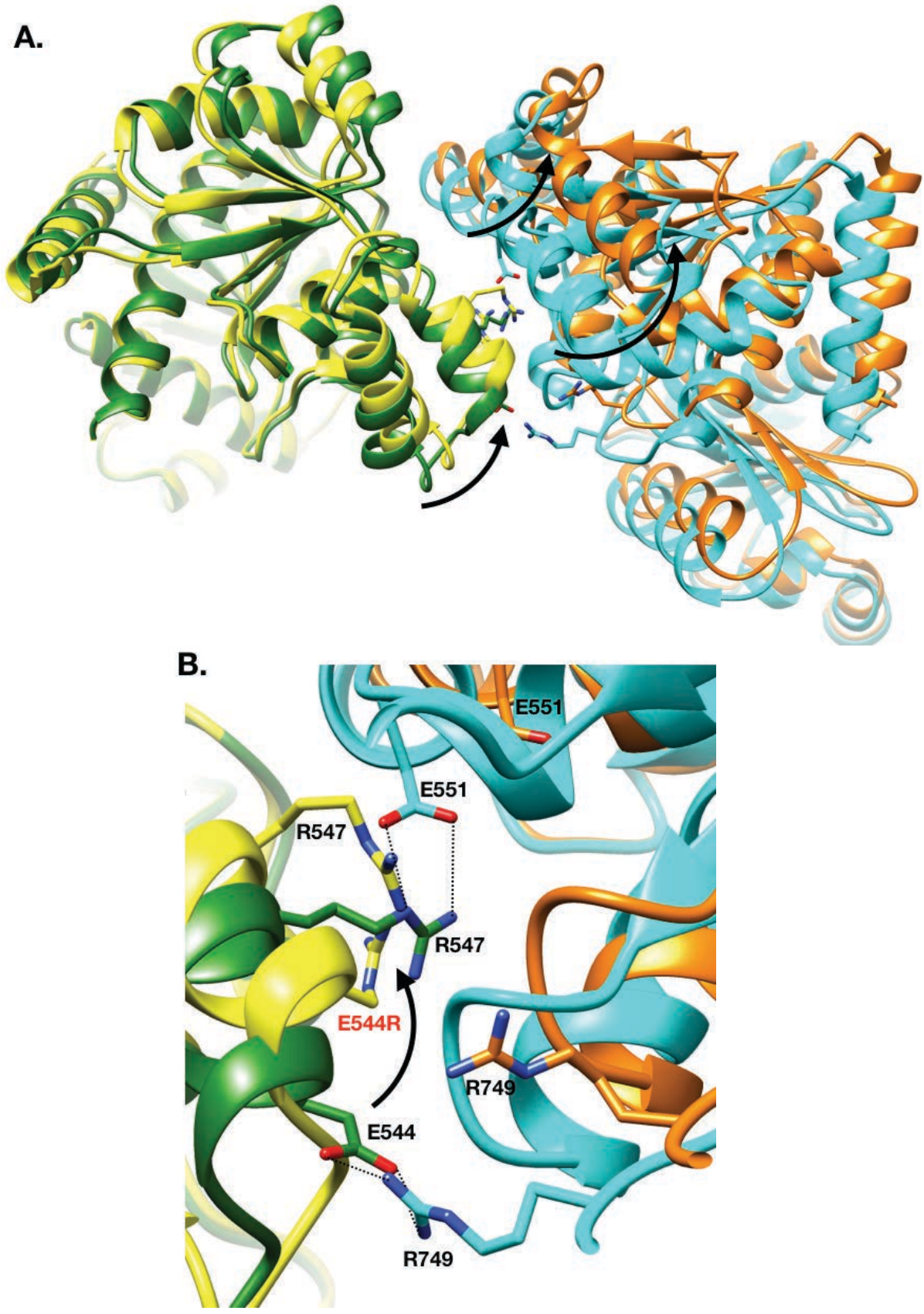


Figure 72. Pta-E544R
A. The hexamer interface of E544R is destabilised and causes a conformational reorientation between interacting dimers. Bd1833-Pta (green and cyan) and Bd1833-E544R (yellow and orange). **B.** E544R mutation is repelled by R749 but folds back inwards adopting a similar location as R547. Superimposition was performed by least-squares minimisation using the CCP4 suit superpose software. Alignment was performed for Pta dimers using subdomain-II (residues 596-738) from both protomers (rmsd = 0.55Å).

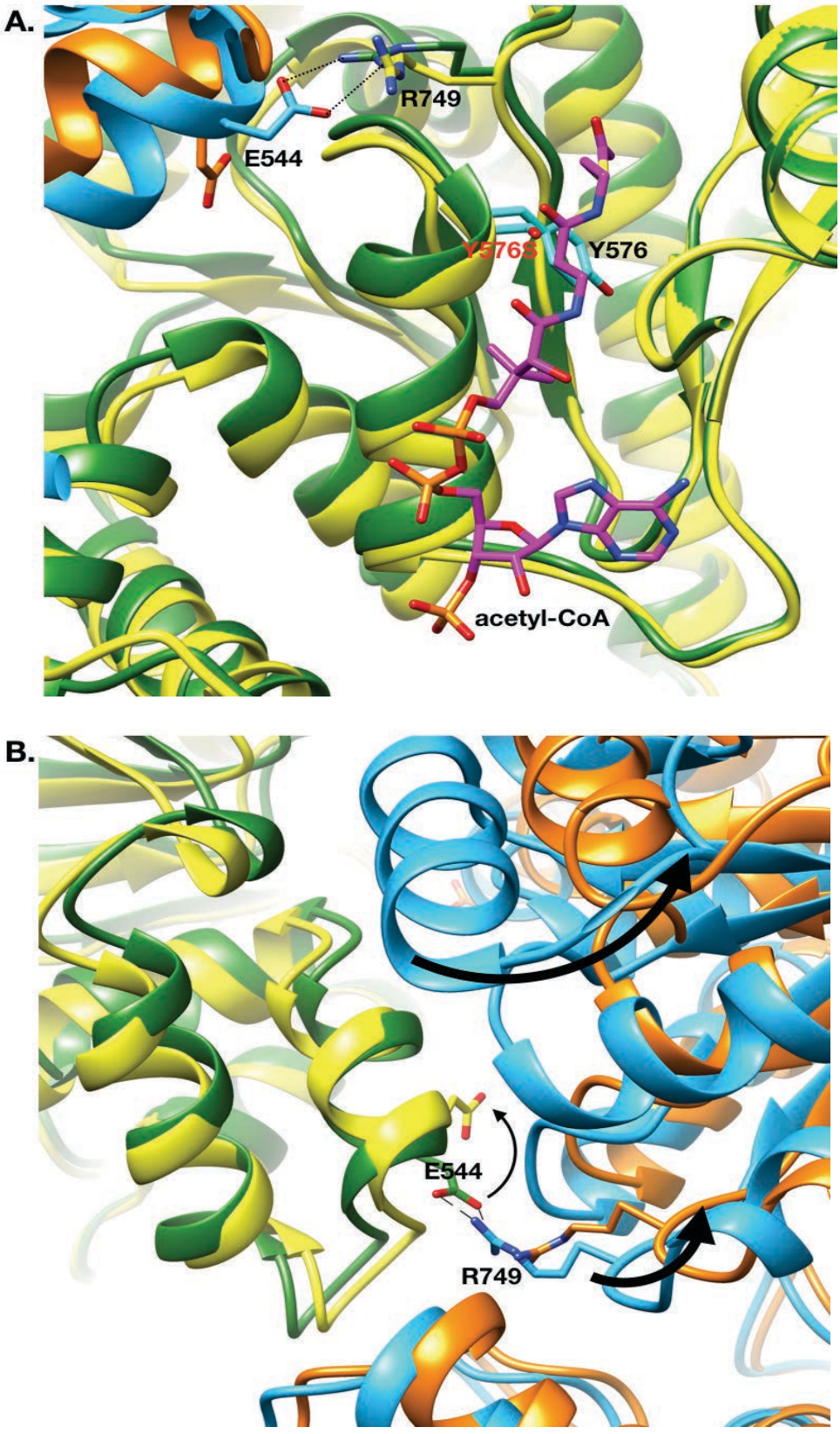


Figure 73. Y576S mutation
A. Y576 is located close to the thiol group of acetyl-CoA and is a highly conserved serine residue in Pta enzymes. Y576S selectively prevents acetyl-CoA binding but allows CoA binding. The structure of Pta-Y576S (yellow and orange) superimposed with Bd1833-Pta (green and cyan). **B.** Y576S mutation selectively prevents acetyl-CoA binding, which changes the conformation of subdomain-I and consequently destabilises the hexamer interface. Superimposition was performed by least-squares minimisation using the CCP4 suit superpose software. Alignment was performed for Pta dimers using subdomain-II (residues 596-738) from both protomers (rmsd = 0.72Å).

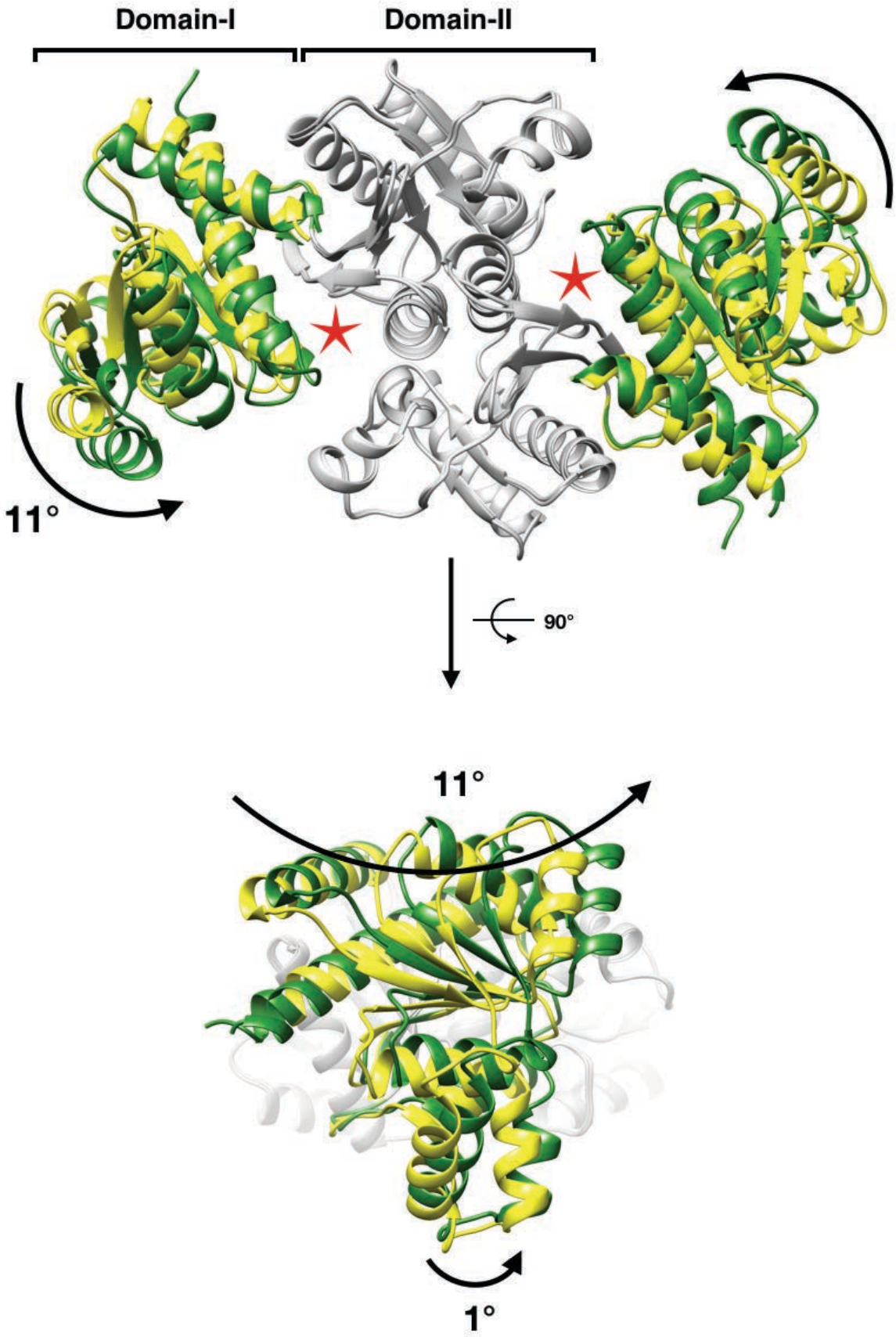


Figure 74. Bd1833-Pta domain movement
Bd1833-Pta (green) and Pta-Apo (yellow) structures are shown, superimposed by subdomain-II (grey). A red star marks the substrate binding sites. Subdomain-I undergoes an 11° rotational rigid body conformational change, indicated by the black curved arrows. The movement changes accessibility to the substrate binding groove located between subdomain-I and II. A sideview of the same rotational rigid body movement demonstrates the upper region undergoes a larger rotational change compared to the lower region. Superimposition was performed by least-squares minimisation using the CCP4 suit superpose software. Alignment was performed for Pta dimers using subdomain-II (residues 596-738) from both protomers (rmsd = 0.82Å).

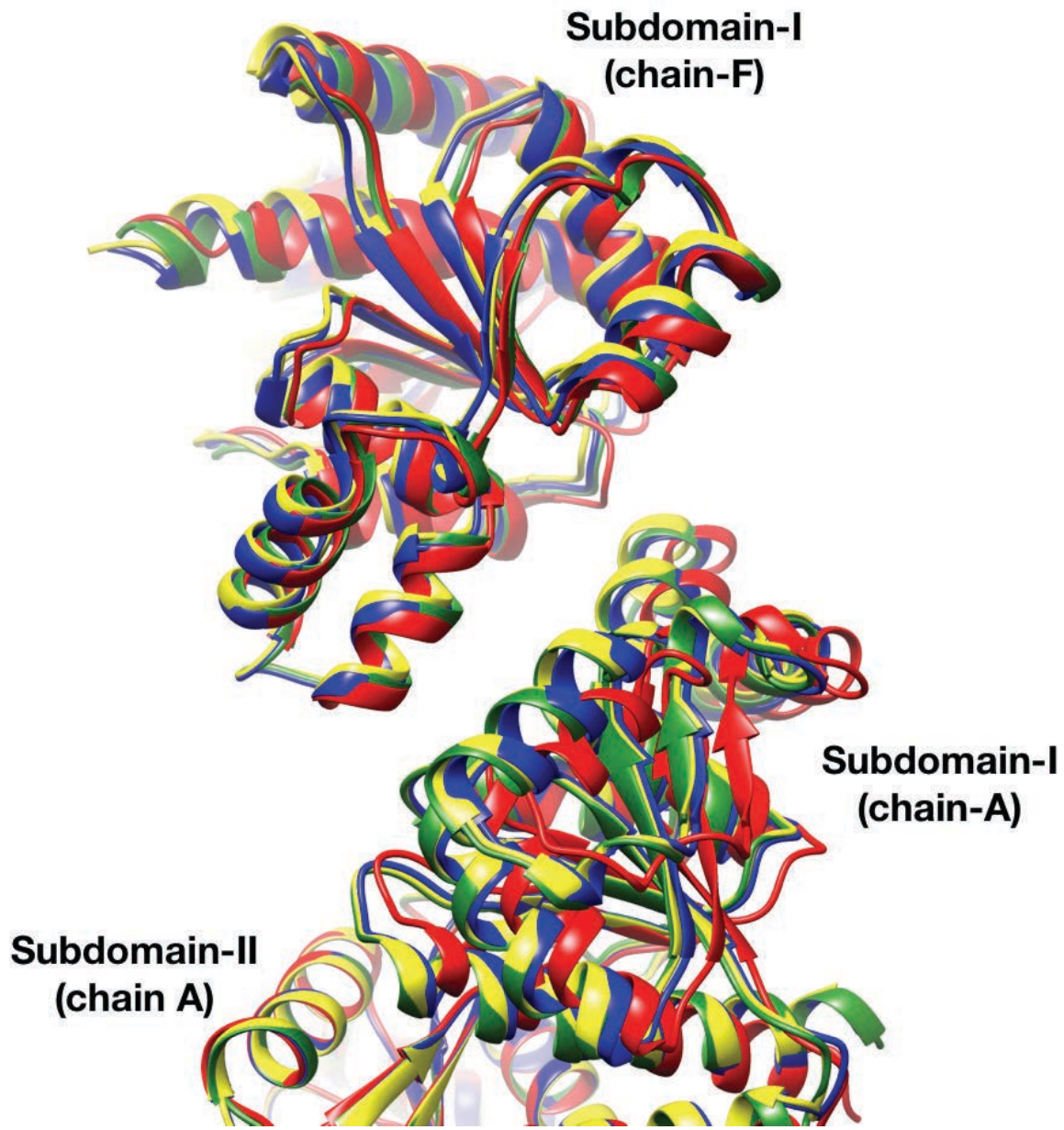


Figure 75. Pta mutations and hexamer interface
The hexamer vertex interface between the AB dimer and EF dimers is shown for R535E (blue), R544R (yellow) and Q576S (green) mutants aligned with Pta-Apo (red). The hexamer is aligned around subdomain-II in chain-A. The degree of hexamer rearrangement affecting the vertex interface is similar in all mutations and Apo form. The vertex hexamer interface is reduced from 644Å^2 in Bd1833-Pta in complex with acetyl-CoA to ~240Å^2 in ligand free Bd1833-Pta structures, which is more than a two-fold decrease.

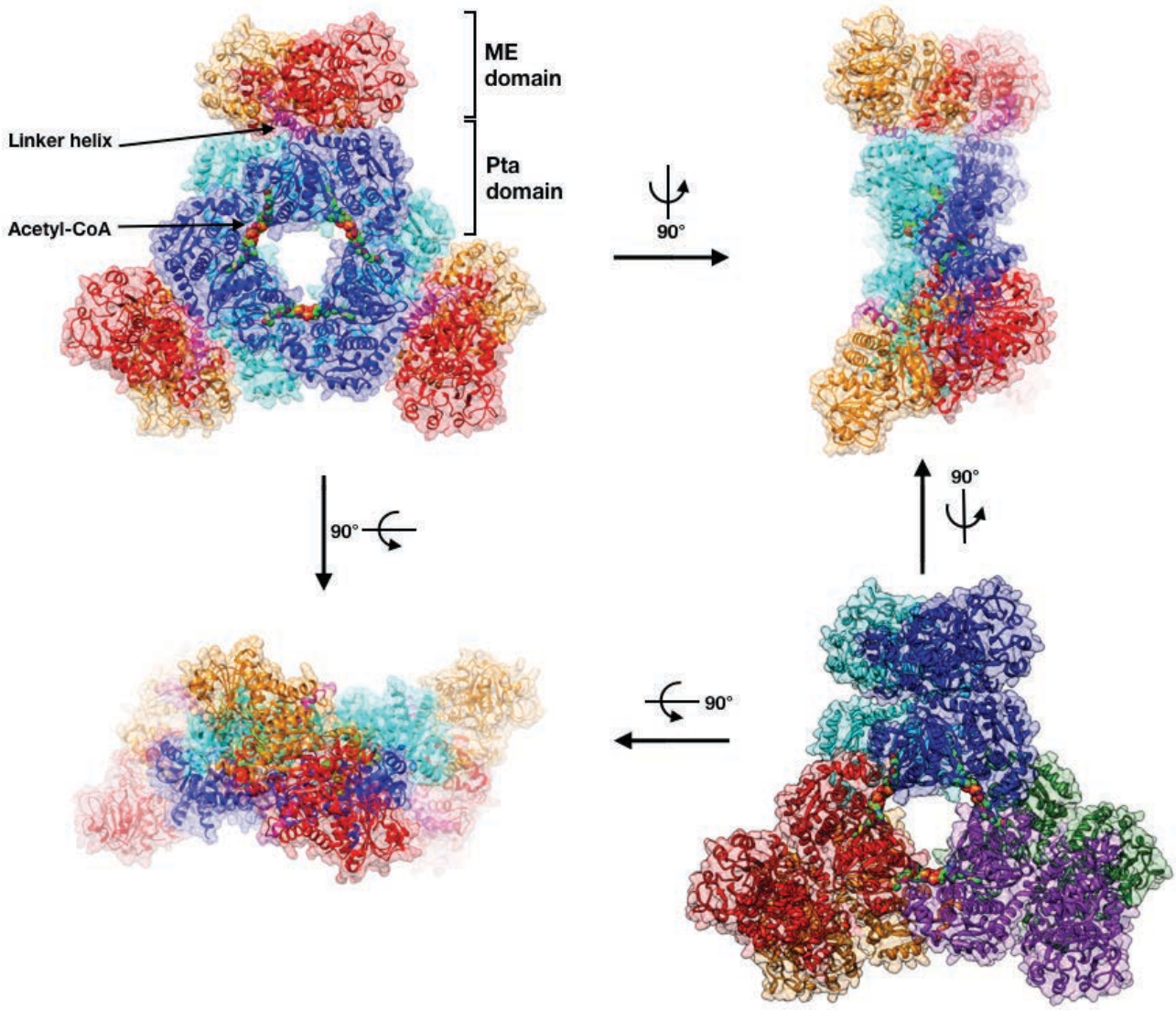


Figure 76. Bd1833 structure
Full-length Bd1833 protein forms a larger hexamer complex. The top left panel shows the hexameric structure of Bd1833 coloured according to the domains (red/orange = ME and blue/cyan = Pta). The linker helix is shown in magenta and acetyl-CoA ligand in green. Top right and bottom left panels show the hexameric structure in two different orientations rotated by 90° demonstrating the twisted nature of the hexamer complex and how the ME domain is positioned on top of the Pta domain, offset by ~40°. Finally the bottom right panel shows Bd1833 hexamer coloured according to the protomers that form the hexamer complex.

The Pta and ME domains are held together by the linker helix which is located on either side of the Bd1833 dimer (Figure 77). There are minimal interactions between the Pta surface and ME surface and in the central region of the dimer a 6Å gap exists between the surfaces of both domains. A few specific H-bonds are formed between the Pta domain and the ME domain. These include: R444 with F396 backbone carbonyl and

K440 with K112 backbone carbonyl. The majority of the surface interactions between Pta and ME domains are electrostatic, involving charged and polar residues (Figure 77).

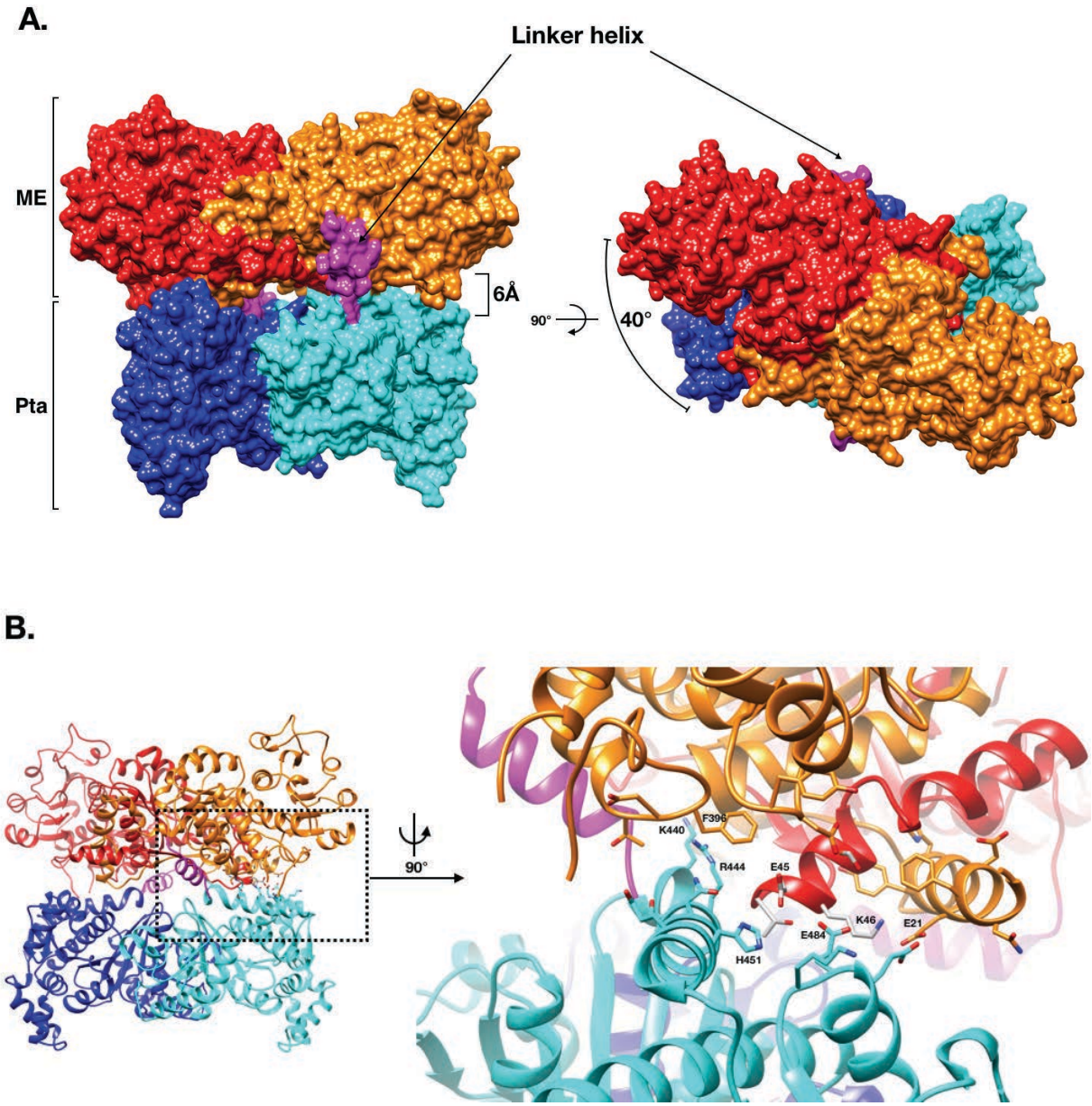


Figure 77. Pta and ME interface
A. The Pta and ME domains in full-length Bd1833 are linked by a single linker helix, located on either side of the Bd1833 dimers. Little surface contact is made between the two domains that are offset by ~40° relative to one another. In the central regions a gap up to 6Å exists between the domain surfaces. **B.** Minimal interactions exist between the Pta and ME domain. The interactions that are present are mainly electrostatic and formed between subdomain-A/C of the ME domain and subdomain-I of the Pta domain. The residues contributing to the minimal interface between the Pta and ME domains are displayed.

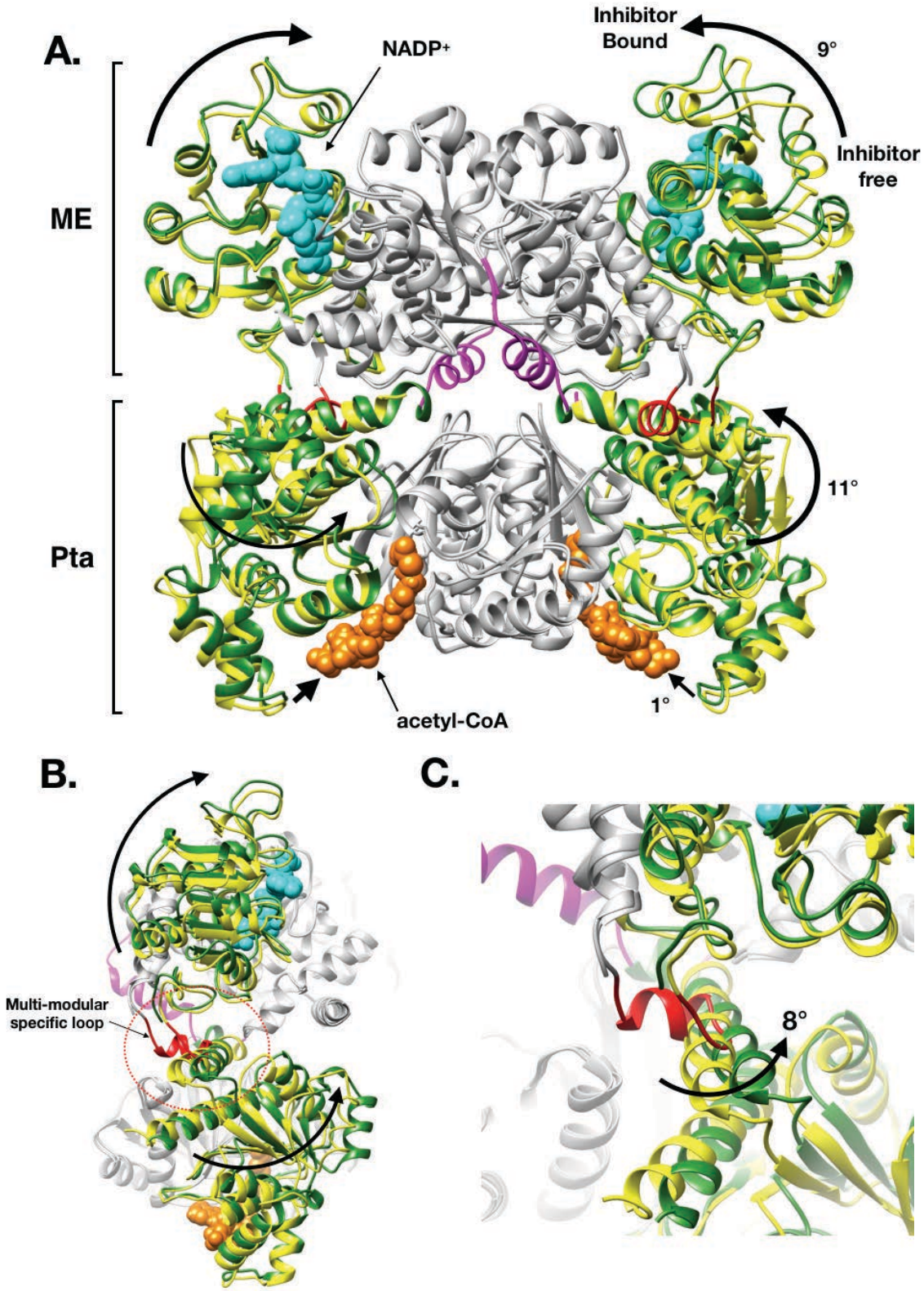


Figure 78. Bd1833 Allosteric regulation

A. The structure of full-length Bd1833 in complex with acetyl-CoA (green) is superimposed with Pta-Apo (yellow, length = 1090 atoms, rmsd = 0.86Å) and Bd1833-ME-NADP (yellow, length = 6670 atoms, rmsd = 0.47Å). The core region of Bd1833 (ME subdomain A and B and Pta subdomain-II) does not alter conformation between structures and is coloured in grey for clarity. The linker helix is shown in magenta and the multi-modular specific loop region (ME subdomain-D) is shown in red. The location of Pta and ME binding sites are represented by $NADP^+$ (cyan) and acetyl-CoA (orange) to gauge the distance between binding sites. An 11° rigid body rotational movement of Pta subdomain-I allosterically causes a 9° rigid body rotational movement in ME subdomain-C, restricting access to the active site. **B.** Demonstrates the conformational changes in full-length Bd1833, viewed from the side. This image demonstrates the clash with Bd1833-ME subdomain-D (red). **C.** A close up view of the clash with subdomain-D; subdomain-D which is not traceable in the structure of full-length Bd1833. Superimposition was performed by least-squares minimisation using the CCP4 suit superpose software. Alignment was performed for Bd1833 dimers using residues 34-174 & 596-738 from both protomers.

differences. K581 represents the only distinctive change, which does not appear to make an interaction with the pyrophosphate group of acetyl-CoA (Figure 79.B).

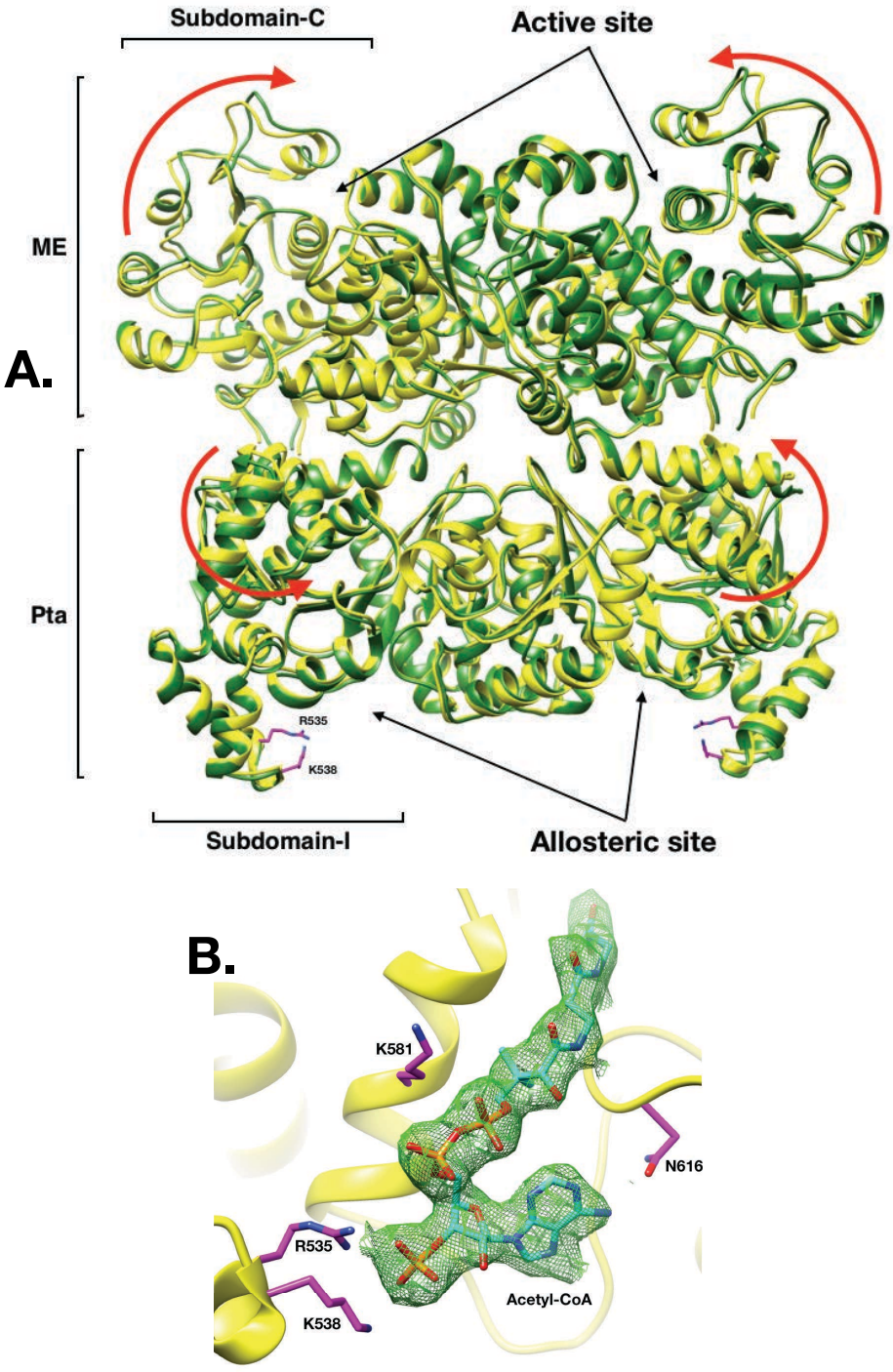


Figure 79. B1833-SKB structure
A. The structure of Bd1833-SKB (Yellow) is superimposed with the structure of Bd1833 (Green)(length = 23032 atoms, rmsd 0.55Å). Bd1833-SKB does not co-purify with ligand and likely represents a non-inhibited full-length conformation. The ME domain active site and Pta allosteric site are highlighted. Red arrows demonstrate the rigid body movements in subdomain-I and subdomain-C. Superimposition was performed by least-squares minimisation using the CCP4 suit superpose software. Alignment was performed for Bd1833 dimers using residues 34-174 & 596-738 from both protomers. **B.** The acetyl-CoA interaction with Pta-SKB (truncation mutant of Bd1833-SKB that co-purifies with ligand). K581 does not interact with the pyrophosphate group. ($2F_o$-F_c map, radius=2Å, σ=1).

Bd1833 species	Kinetic Parameters (malate)		Kinetic Parameters (NADP+)		Oligomeric State	Acetyl-CoA Inhibition	Co-purified Ligand?	Reasoning
	K_M (mM)	k_{cat} (s^{-1})	K_M (µM)	k_{cat} (s^{-1})				
Bd1833	2.61 ± 0.28	26.1 ± 1.2	21.7 ± 2.3	21.1 ± 0.13	Hexamer	Yes	CoA + AcCoA	
Bd1833-Apo	3.02 ± 0.31	35.2 ± 1.4	43.1 ± 5.2	33.5 ± 1.3	Hexamer	Yes	None	Ligand free
Bd1833-ME	4.15 ± 0.50	3.28 ± 0.17			Dimer	No	None	ME domain only
Bd1833-Δloop	12.5 ± 1.0	1.62 ± 0.09			ND	No	CoA + AcCoA	Chimera lacking multi-modular specific loop
Bd1833-R535E	3.17 ± 0.15	19.4 ± 0.36	41.1 ± 4.0	16.8 ± 0.54	Hexamer	No	None	Prevent CoA binding (3'-P)
Bd1833-R535A	3.48 ± 0.23	20.3 ± 0.53			Hexamer	No	None	Prevent CoA binding (3'-P)
Bd1833-N718D	3.35 ± 0.33	41.8 ± 1.6			Hexamer	No	None	Prevent CoA binding (acetyl-group)
Bd1833-E544R	12.6 ± 2.0	18.9 ± 1.9			Dimer	No	None	Destabilise hexamer
Bd1833-E544A	4.81 ± 0.44	38.7 ± 1.6			Hexamer	Yes	CoA + AcCoA	Destabilise hexamer
Bd1833-Y576S					Hexamer	No	CoA	Convert back to conserved Serine residue
Bd1833-Y576C					ND	Yes	CoA	Interact with thiol group
Bd1833-R749A					ND	Yes	CoA + AcCoA	Destabilise hexamer
Bd1833-R547A					ND	Yes	CoA + AcCoA	Destabilise hexamer
Bd1833-R581K					ND	Yes	CoA + AcCoA	Pyrophosphate interaction
Bd1833-SKB	1.96 ± 0.24	35.0 ± 1.3	45.4 ± 6.6	39.3 ± 1.8	Hexamer	Yes	None	Close homologue

Table 14. Bd1833 kinetics
This table includes the values obtained from our kinetic studies using malate as the substrate, the oligomeric state and whether inhibition by acetyl-CoA was detected for each Bd1833 species. ND = not determined. Values presented as a mean of triplicates alongside the standard deviation.

5.2.8 Kinetics

5.2.8.1 Bd1833 activity

A detailed kinetic analysis of Bd1833 has been performed. ME activity was detected for both malate oxidative decarboxylation and pyruvate reductive carboxylation reactions; however, no Pta activity was detected. Table 14, displays the kinetic parameters obtained for Bd1833 and Bd1833-mutant species.

5.2.8.2 Malate Oxidative Decarboxylation

Bd1833 required a divalent cation such as Mn^{2+} to be catalytically active. Bd1833 also demonstrates specificity for $NADP^+$ as a cofactor, showing no activity in the presence of NAD^+. Kinetic parameters were determined using L-malate as the substrate, at a concentration between 0.1 and 10 mM; reduced rates were observed beyond 10 mM L-malate (Figure 80).

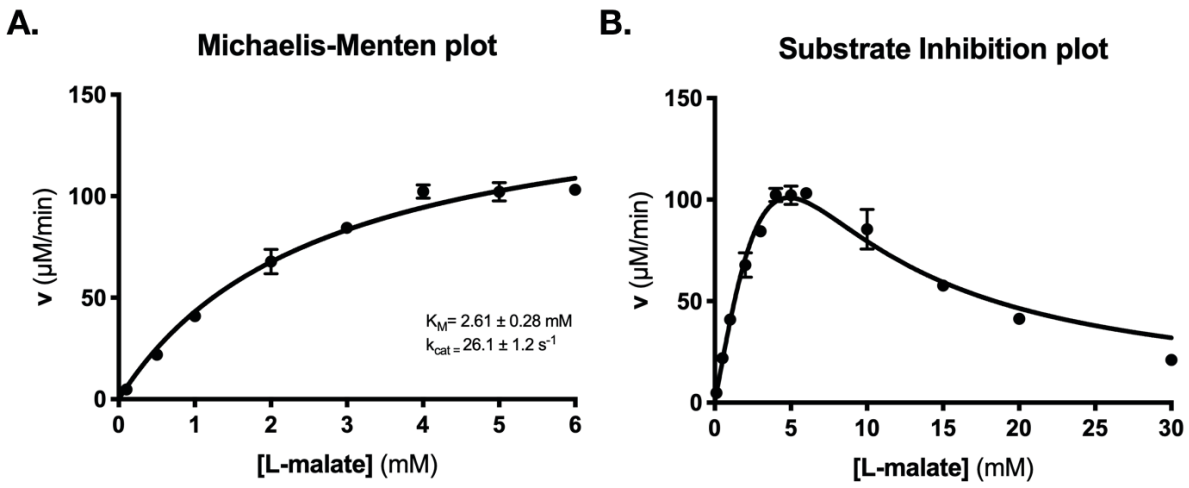


Figure 80. Malate oxidative decarboxylation
Kinetic data shown for full-length Bd1833. **A.** The data for Bd1833 fit to a Michaelis-Menten plot. **B.** At concentrations of malate above 8 mM, reduced rates are observed. The data has been fit to substrate inhibition model $Y=V_{max}*[\text{L-malate}]/(K_M + [\text{L-malate}]*(1+[\text{L-malate}]/K_i))$ using prism7 software to explain the reduced rates. The fit is "ambiguous" suggesting the model is not ideal to explain the data, suggesting other factors such as ΔpH or ionic strength may also contribute. All data points were measured in at least triplicates. The error bars represent the standard deviation.

5.2.8.3 Effectors

A series of potential effectors, published for other malic enzymes, were tested as potential regulators of Bd1833. The activity of Bd1833 in the presence of effectors is presented as a percentage of the activity of Bd1833 without effector (Figure 81). Acetyl-CoA and OAA are shown to be potent inhibitors of Bd1833 activity and reduce the activity to below 10% of the un-inhibited activity. CoA and acetyl-P reduced the activity to ~50% suggesting a weaker mode of inhibition. No activators of Bd1833 activity were identified.

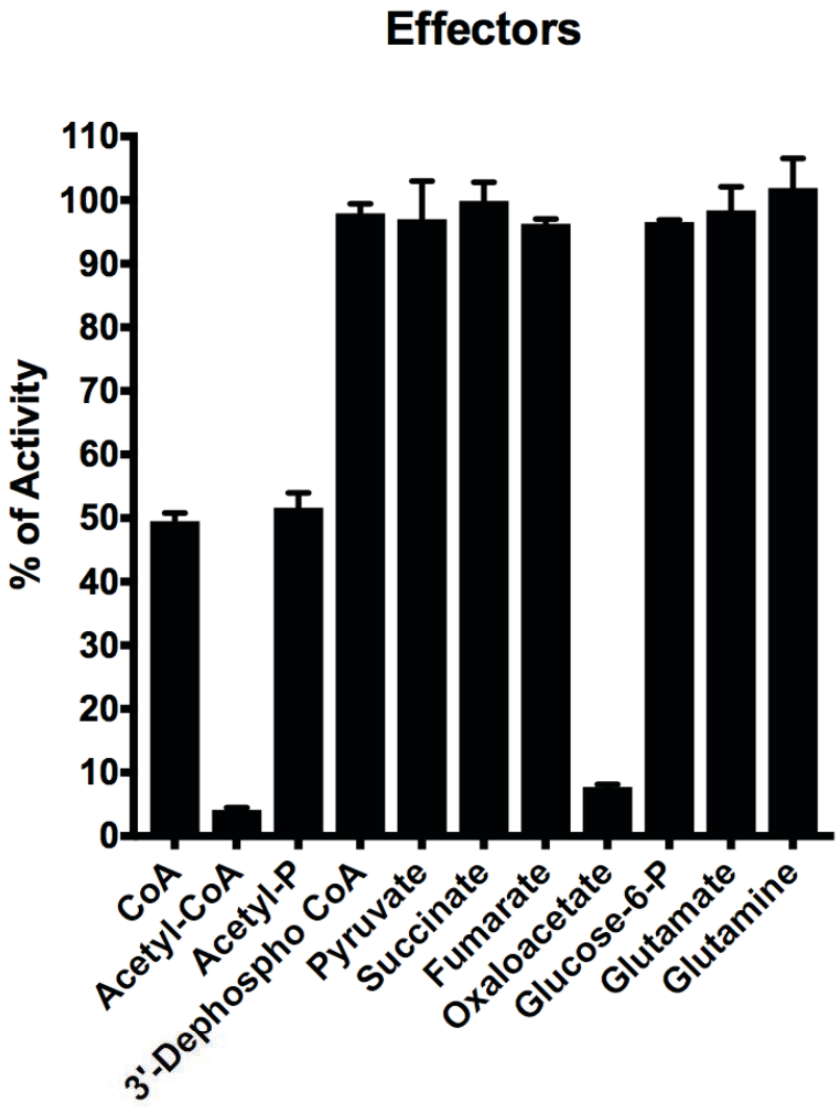


Figure 81. Identifying Bd1833 effectors
A range of malic enzyme effectors were tested for their effects on Bd1833 activity. Bd1833 activity is reduced to ~50% in the presence of CoA and acetyl-P. Acetyl-CoA and Oxaloacetate are potent inhibitors of Bd1833 activity, reducing the activity to below 10% of Bd1833 activity in standard assay conditions. All data points were measured in at least triplicates. The error bars represent the standard deviation.

5.2.8.4 Effector binding sites

To investigate how these effectors regulate the activity of Bd1833 we created a number of Bd1833 mutants. The truncation mutant (Bd1833-ME) activity was measured in the presence of the effectors and compared to the activity without (Figure 82). Acetyl-CoA, CoA and acetyl-P showed no inhibitory effects towards truncated Bd1833-ME; however, OAA remained an inhibitor. This result confirms that acetyl-CoA, CoA and acetyl-P interact with the Pta domain and allosterically regulate ME activity, whereas OAA interacts directly with the ME domain, which supports our structural data. Interestingly, the activity of Bd1833-ME was reduced in comparison to full-length Bd1833, suggesting communication with the Pta domain is important for catalysis.

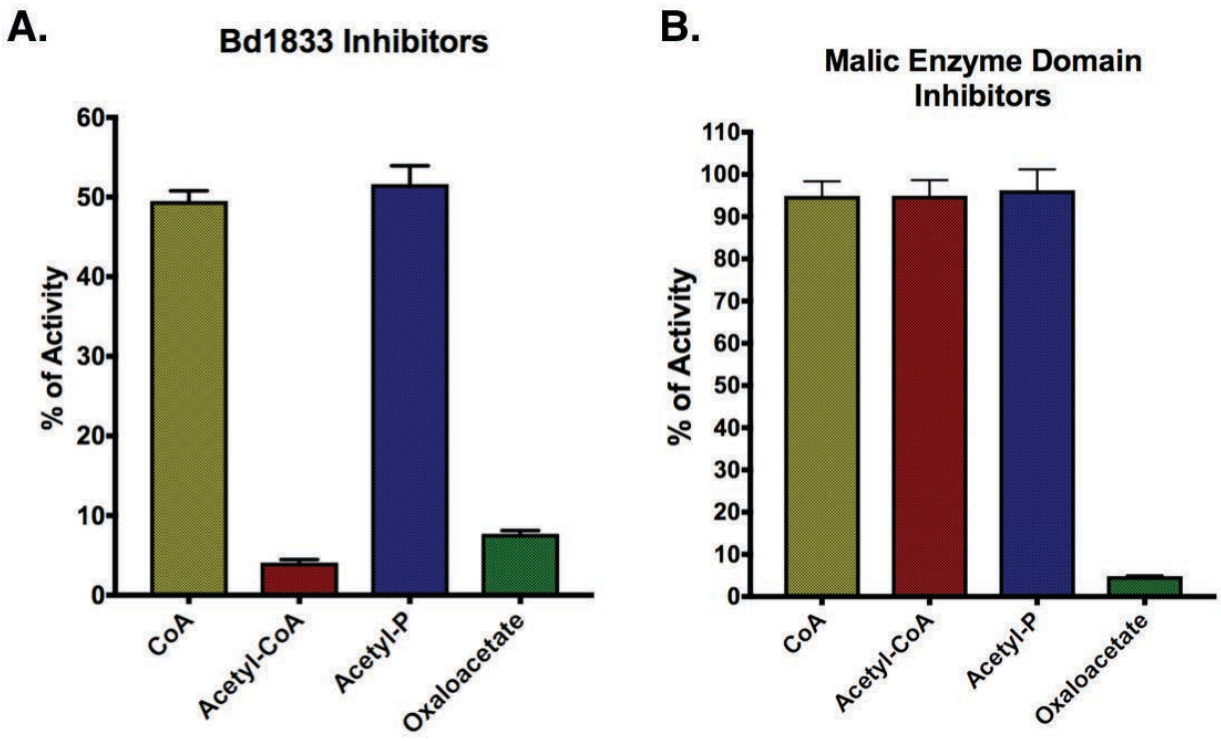


Figure 82. Bd1833 inhibitors
A. The activity of Bd1833 in the presence of inhibitors. **B.** The activity of Bd1833-ME truncation mutant in the presence of the same effectors. Three of the effectors, including acetyl-CoA, do not inhibit Bd1833-ME whereas oxaloacetate still inhibits. This finding indicates which domains the inhibitors interact with. All data points were measured in at least triplicates. The error bars represent the standard deviation.

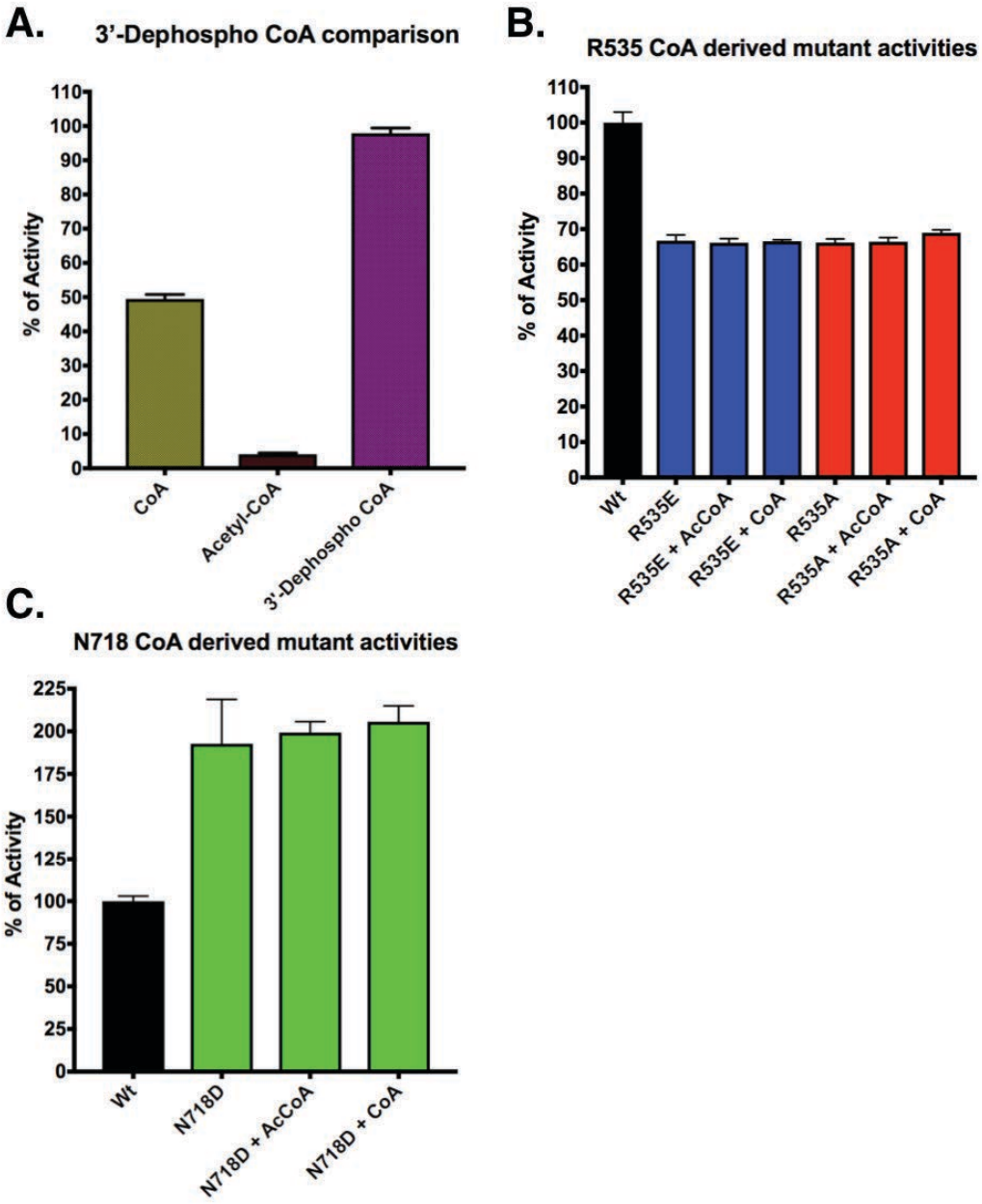


Figure 83. Identifying acetyl-CoA binding site
A. CoA and acetyl-CoA both inhibit the activity of Bd1833; however, a CoA analogue lacking the 3'-phosphate group shows no inhibition. **B.** Mutation of R535 that interacts with the 3'-phophate group alleviates the inhibitory affects of both acetyl-CoA and CoA. The rate of R535A and R535E is reduced compared to WT. **C.** N718D prevents acetyl-CoA and CoA at the thiol-group binding site. As a result CoA and acetyl-CoA no longer inhibit and the activity of Bd1833-R718D is increased compared to Wt. All data points were measured in at least triplicates. The error bars represent the standard deviation.

completely inhibited in the presence of acetyl-CoA yet has almost maximal activity in the presence of CoA. Importantly the K_i values obtained for acetyl-CoA from the two assays are relatively similar.

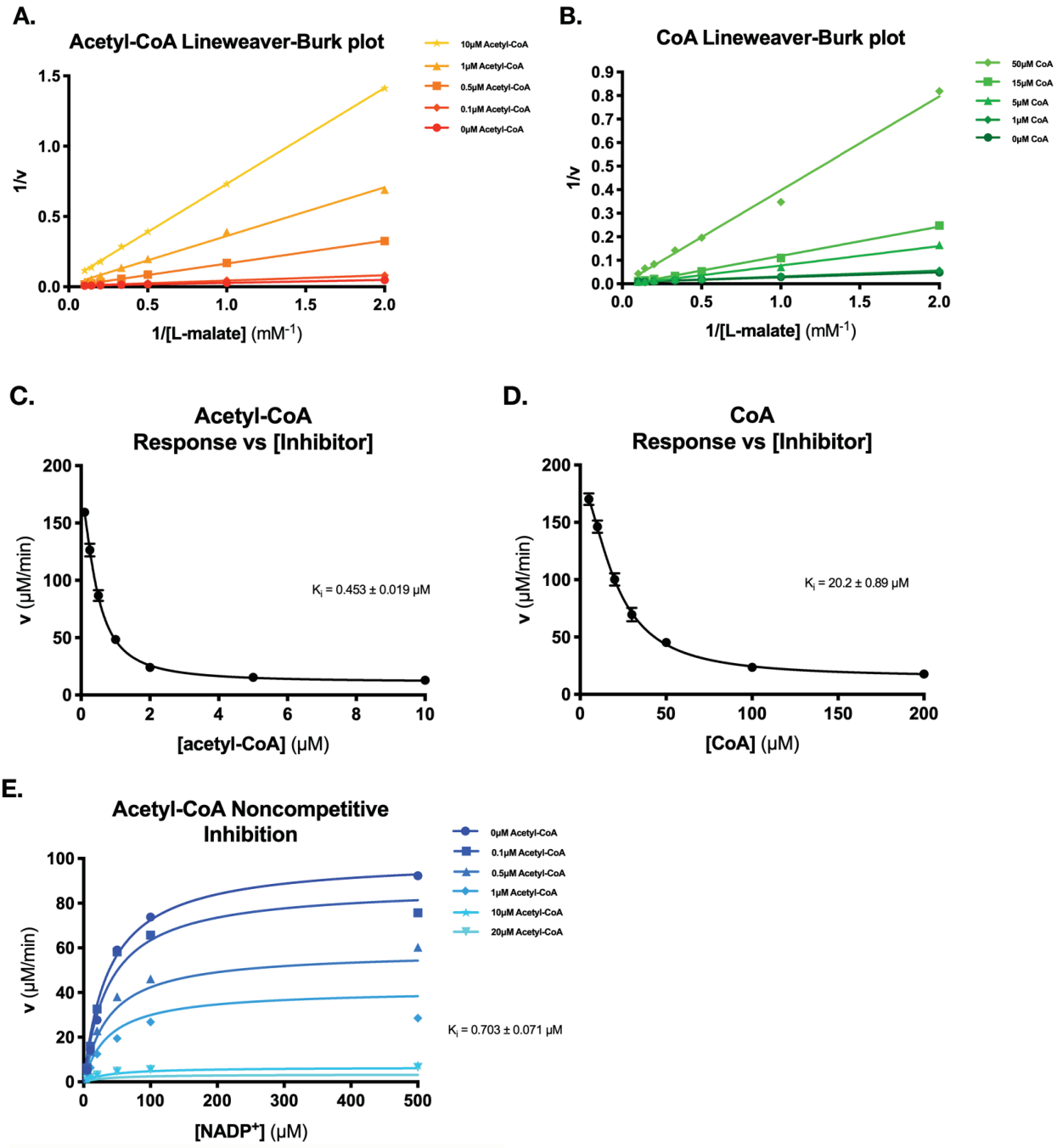


Figure 84. Inhibitor Potency
A. Lineweaver-Burk plot demonstrating the effect of acetyl-CoA on Bd1833 Michaelis-Menten kinetics **B.** Lineweaver-Burk plot demonstrating the effect of CoA on Bd1833 Michaelis-Menten kinetics **C.** Plot of response vs [inhibitor], demonstrating the effect of acetyl-CoA concentration on the activity of Bd1833 at a fixed substrate concentration (IC_{50}) **D.** Plot of response vs [inhibitor], demonstrating the effect of CoA concentration of the activity of Bd1833 at a fixed substrate concentration (IC_{50}) **E.** Plot of Michaelis-Menten data for Bd1833 ($NADP^+$ varied substrate) in the presence of increasing [acetyl-CoA]. The data were fit to the noncompetitive inhibition model. All data points were measured in at least triplicates. The error bars represent the standard deviation.

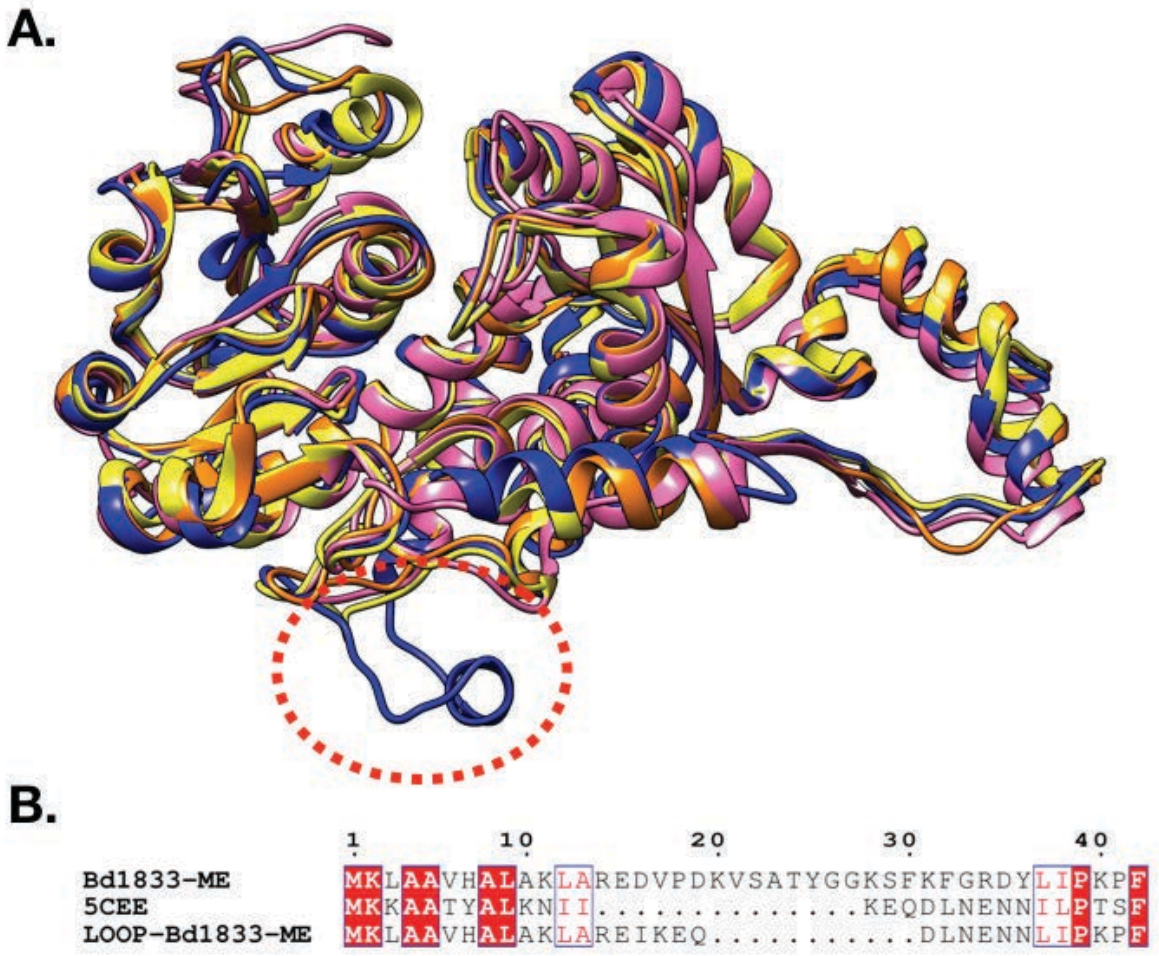


Figure 85. Multi-modular ME specific loop
A. The structure of Bd1833-ME (blue) superimposed with the prokaryotic MEs (pink, PDB:2A9F, length = 337 residues, rmsd = 1.78Å), (yellow, PDB:5CEE, length = 379 residues, rmsd = 1.09Å) and (orange, PDB:2HAE, length = 348 residues, rmsd = 1.58Å) using SSM superpose tool in coot. The loop region highlighted by the red dashed circle is specific to Bd1833 and other multi-modular malic enzymes. **B.** An alignment of the multi-modular specific loop region, indicating the sequence used to create the chimera of Bd1833-Δloop.

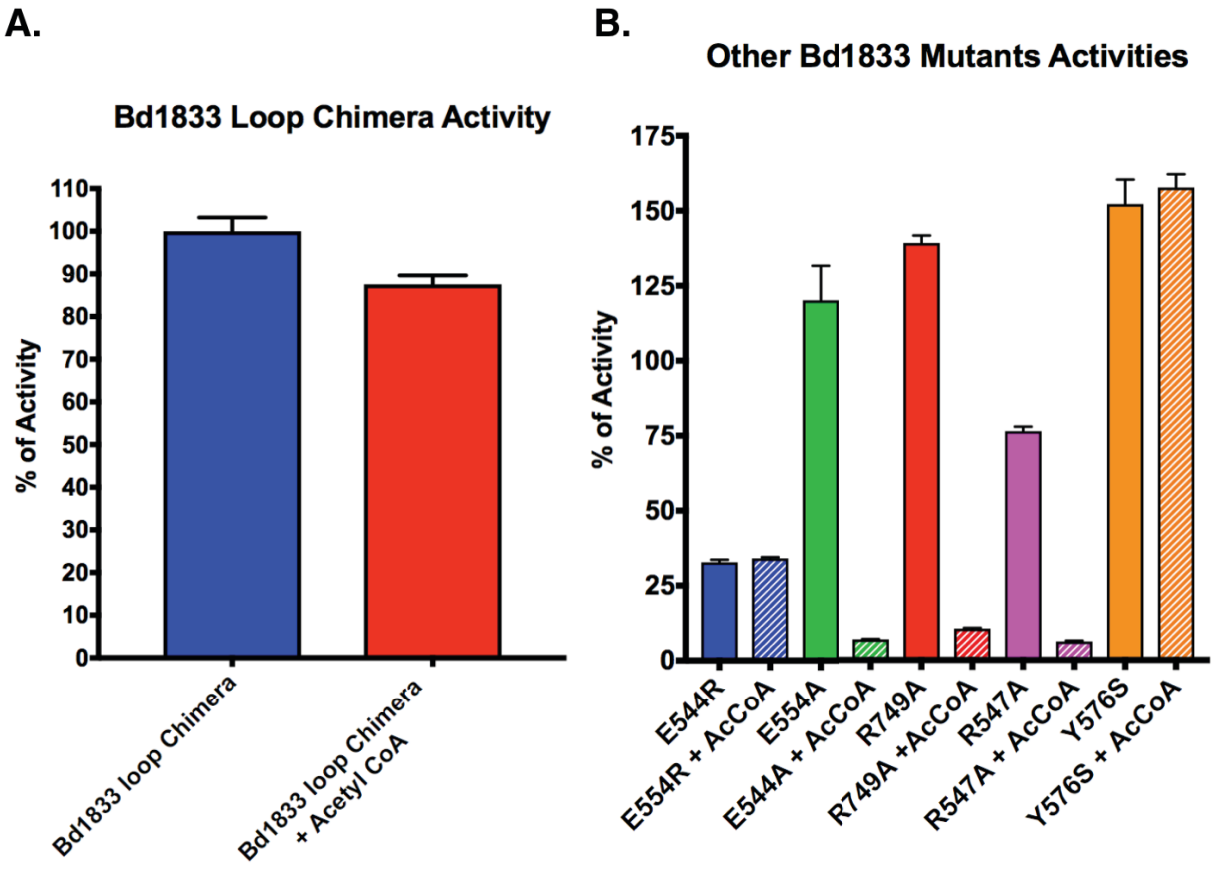


Figure 86. Bd1833 mutants activity and sensitivity to acetyl-CoA
A. The activity of Bd188-Δloop with no acetyl-CoA and in the presence of acetyl-CoA. **B.** This Figure presents the activities of a series of Bd1833 mutants and their sensitivity to acetyl-CoA. All activities are displayed as a percentage of the Bd1833 Wt. E544R shows a decreased activity but is insensitive to acetyl-CoA; however, E544A remains sensitive an has an unaffected activity. Y576S is insensitive to acetyl-CoA and has an increased activity. Furthermore, R749 and R547 mutants located at the hexamer interface remain sensitive to acetyl-CoA. All data points were measured in at least triplicates. The error bars represent the standard deviation.

5.2.8.8 Pyruvate reductive carboxylation

We investigated the reverse reaction of Bd1833 (pyruvate reductive carboxylation) to confirm the direction of the most favourable reaction. The parameters obtained from this experiment are displayed in Figure 87. The rates obtained for pyruvate reductive carboxylation were >120 times lower than those observed from the malate oxidative decarboxylation reaction suggesting that Bd1833 strongly favours the forward reaction where pyruvate and NADPH are the products.

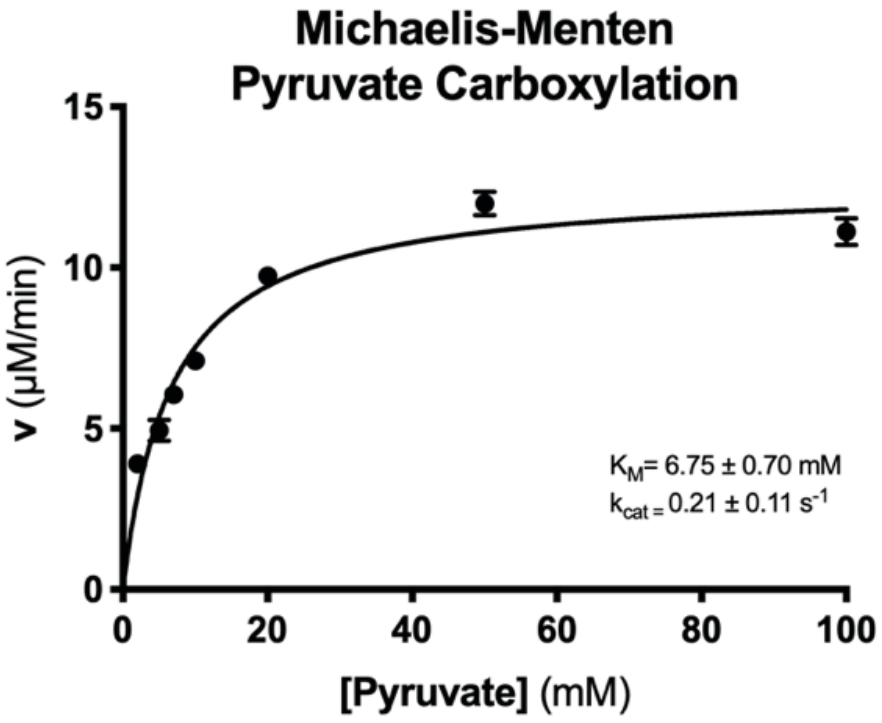


Figure 87. Pyruvate carboxylation reaction
Kinetic data for Bd1833 performing the reverse reaction (pyruvate carboxylation) fit to the Michaelis-Menten equation. Activity is reduced in the reverse direction, which suggests that Bd1833 favours the forward reaction and formation of pyruvate. All data points were measured in at least triplicates. The error bars represent the standard deviation.

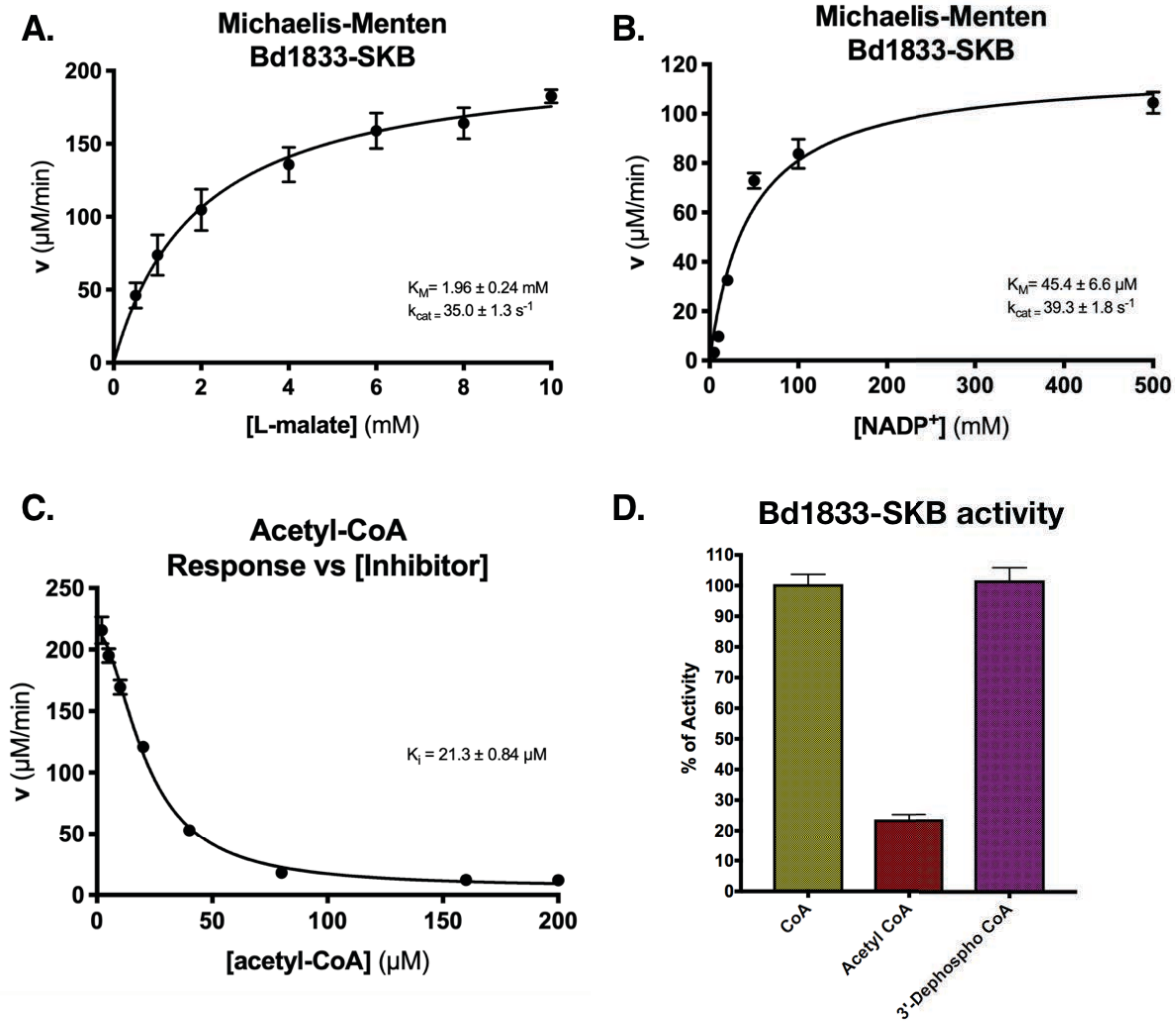


Figure 88. Bd1833-SKB kinetic parameters
A. Oxidative decarboxylation reaction with malate as substrate fit to Michaelis-Menten equation. **B.** Oxidative decarboxylation reaction with $NADP^+$ as substrate fit to Michaelis-Menten equation. **C.** Response vs [acetyl-CoA] data were measured for Bd1833-SKB activity. The K_i value obtained was increased compared to Bd1833 **D.** Bd1833-SKB is not inhibited by CoA at physiological concentrations, whereas acetyl-CoA is still an inhibitor. All data points were measured in at least triplicates. The error bars represent the standard deviation.

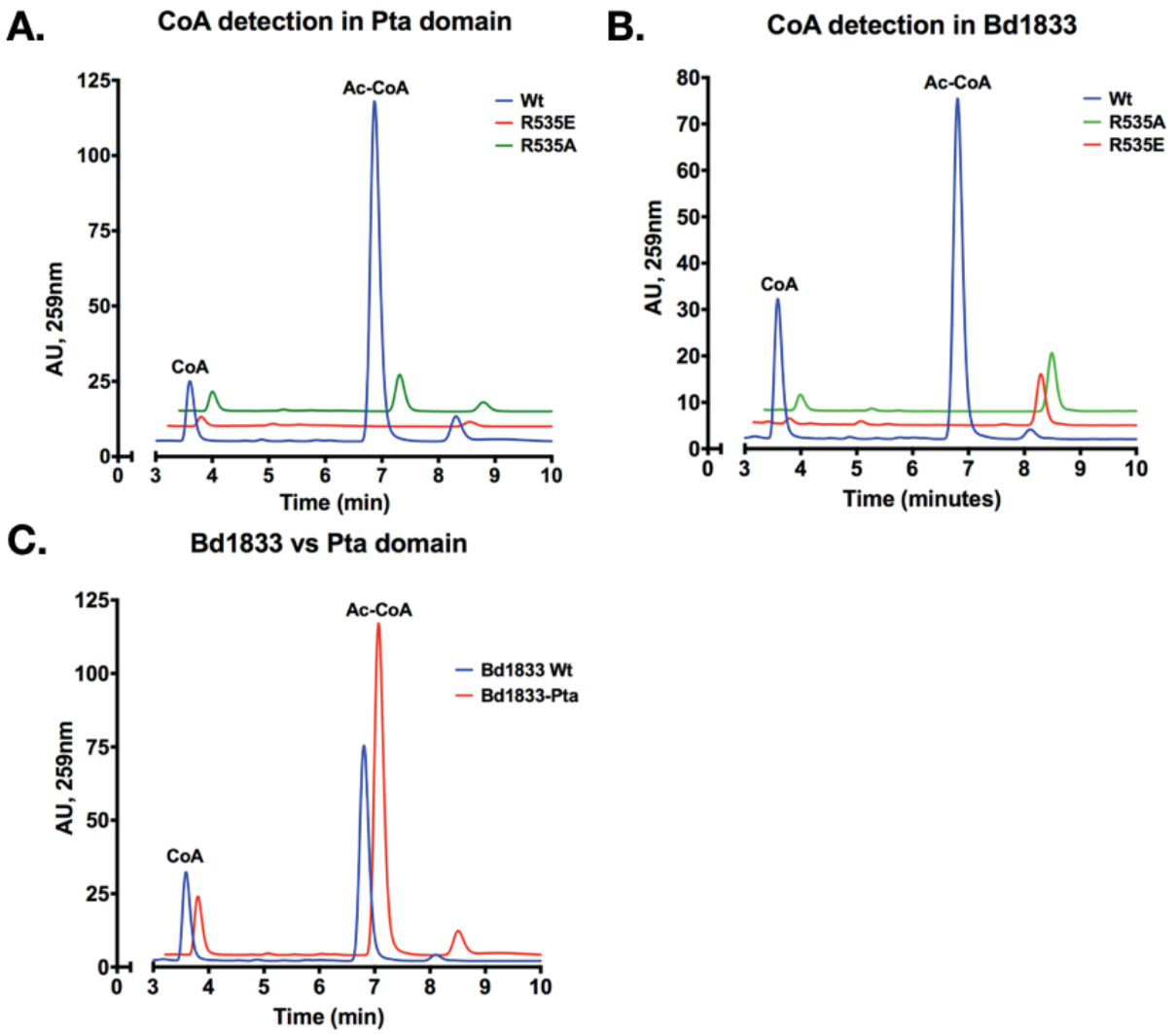


Figure 89. HPLC results
A. Bd1833-Pta CoA and acetyl-CoA HPLC traces in comparison to R535A and R535E mutants. **B.** Bd1833 full-length CoA and acetyl-CoA HPLC traces in comparison to R535A and R535E mutants. **C.** Comparison of Bd1833 and Bd1833-Pta HPLC traces demonstrate the ratio difference between co-purifying ligands. Bd1833-Pta co-purifies with more acetyl-CoA and less CoA than full-length Bd1833.

5.2.9.2 Producing Bd1833-Apo

A ligand free version of Bd1833 (Bd1833-Apo) was required to investigate allosteric regulation by acetyl-CoA without complications from co-purifying ligands. To produce Bd1833-Apo an additional mild denaturing wash step was added to the purification procedure of Bd1833. Illustrated in Figure 90 are the results from two different wash procedures performed on Bd1833 and Bd1833-Pta purification. The most effective

wash procedure was to pass >50CV of (50 mM HEPES pH 7.5, 300 mM NaCl, 2.5M Urea, 2M KBr) over immobilised Bd1833, which resulted in <5 µM of co-purified ligand in a 100 µM sample (>95% of protein is thus assumed to be ligand free).

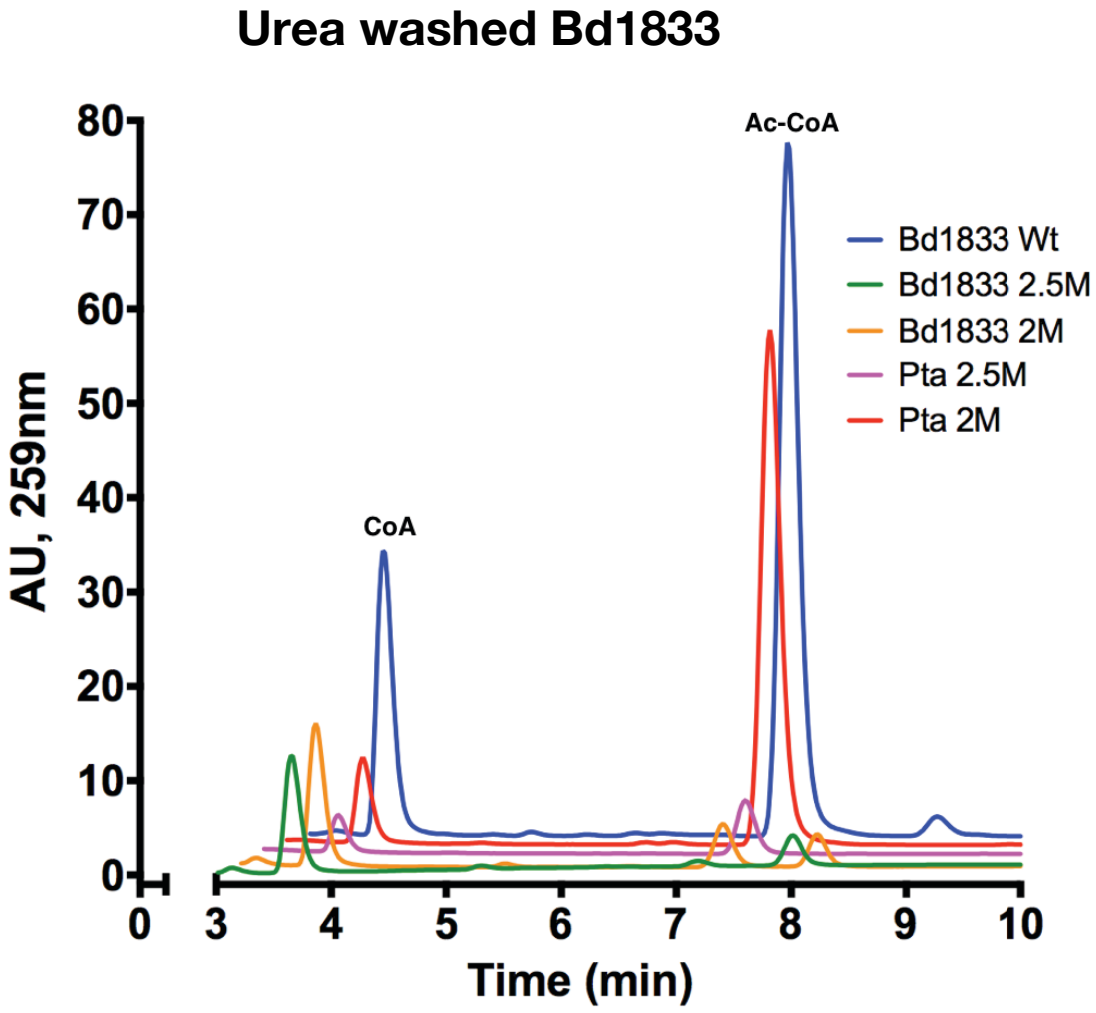


Figure 90. CoA and acetyl-CoA removal by denaturing wash procedure
Removal of CoA and acetyl-CoA from Bd1833 and Bd1833-Pta was achieved by incorporating an addition denaturing wash procedure to the purification protocol. Best results were achieved with 2.5M urea buffer (green and pink traces).

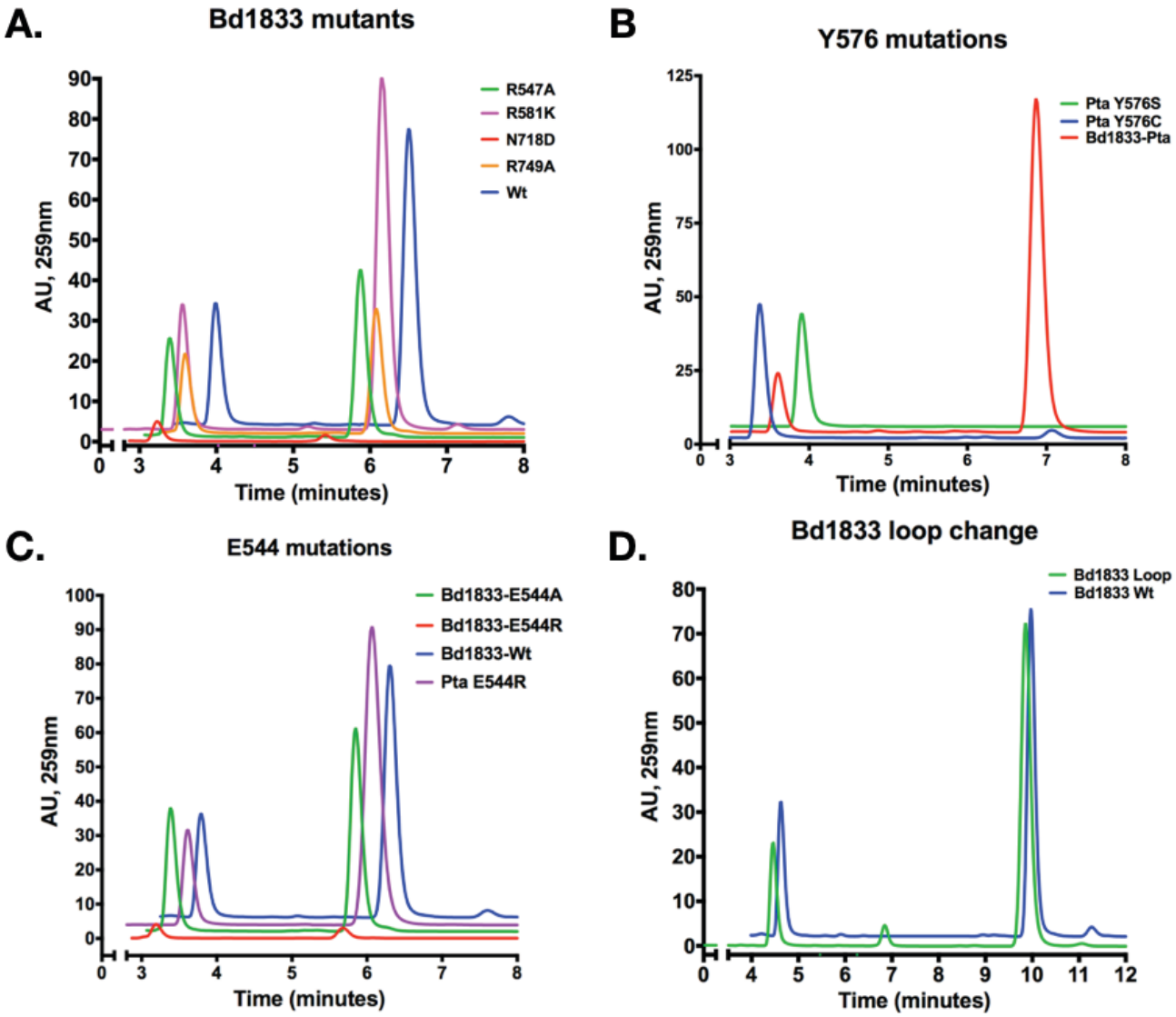


Figure 91. HPLC traces for Bd1833 and Bd1833-Pta mutants
A. Traces of Bd1833 mutants revealed all co-purify with CoA and acetyl-CoA except for N718D. **B.** Bd1833-Pta Y576S and Y576C selectively co-purify with CoA suggesting acetyl-CoA binding is unfavoured. **C.** Bd1833-E544R does not co-purify with ligand; however, Bd1833-E544A and Pta-E544R do co-purify with ligand. **D.** Bd1833-Δloop co-purifies with both ligands. This is the only case of a mutant that is insensitive to acetyl-CoA inhibition and still co-purifies with acetyl-CoA.

5.2.9.4 HPLC Bd1833-Δloop

Bd1833-Δloop co-purified with comparable amounts of CoA and acetyl-CoA as Bd1833 (Figure 91). This finding demonstrates the first of our Bd1833 mutations that co-purify with ligand but are insensitive to acetyl-CoA inhibition.

5.2.9.5 HPLC results for Bd1833-SKB

Bd1833-SKB did not co-purify with CoA or acetyl-CoA, which is strikingly different to Bd1833. Interestingly Pta-SKB (truncation of Bd1833-SKB) co-purified with levels of CoA and acetyl-CoA comparable to Bd1833-Pta.

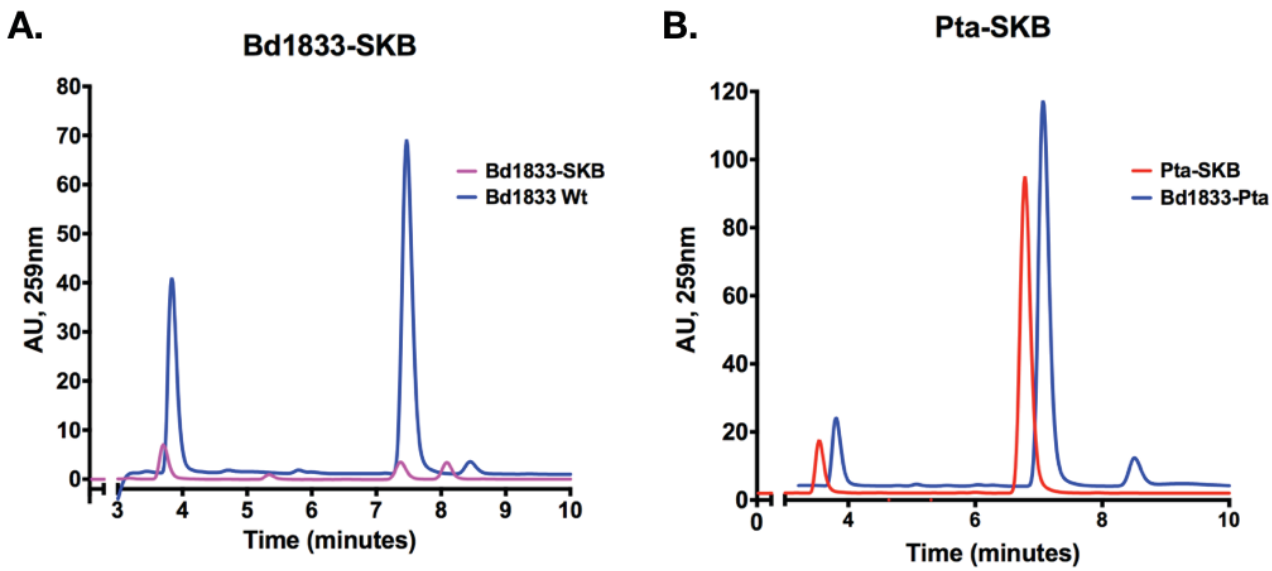


Figure 92. HPLC results for Bd1833-SKB and Bd1833-SKB-Pta
A. Bd1833-SKB did not co-purify with CoA or acetyl-CoA. **B.** Pta-SKB co-purified with CoA and acetyl-CoA.

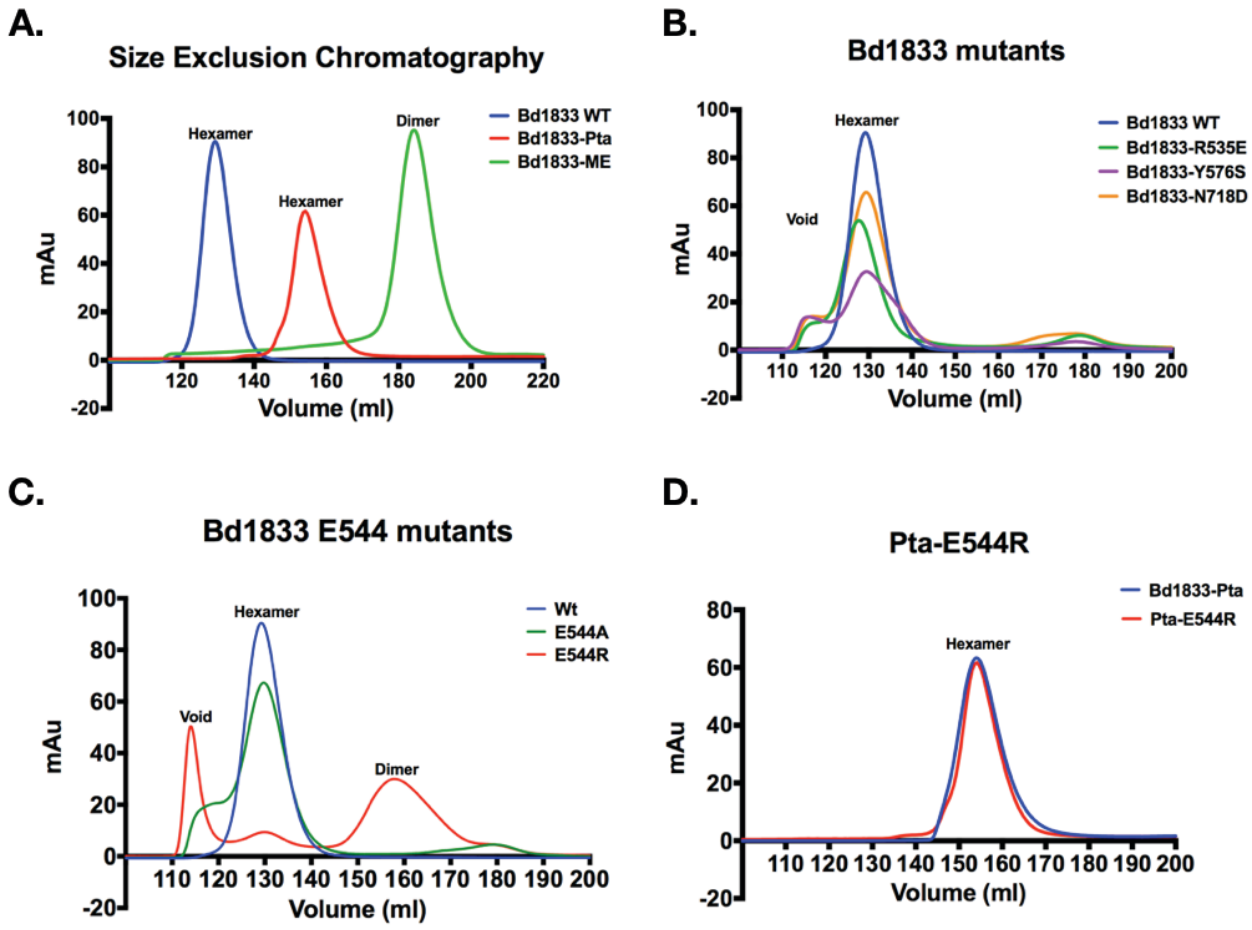
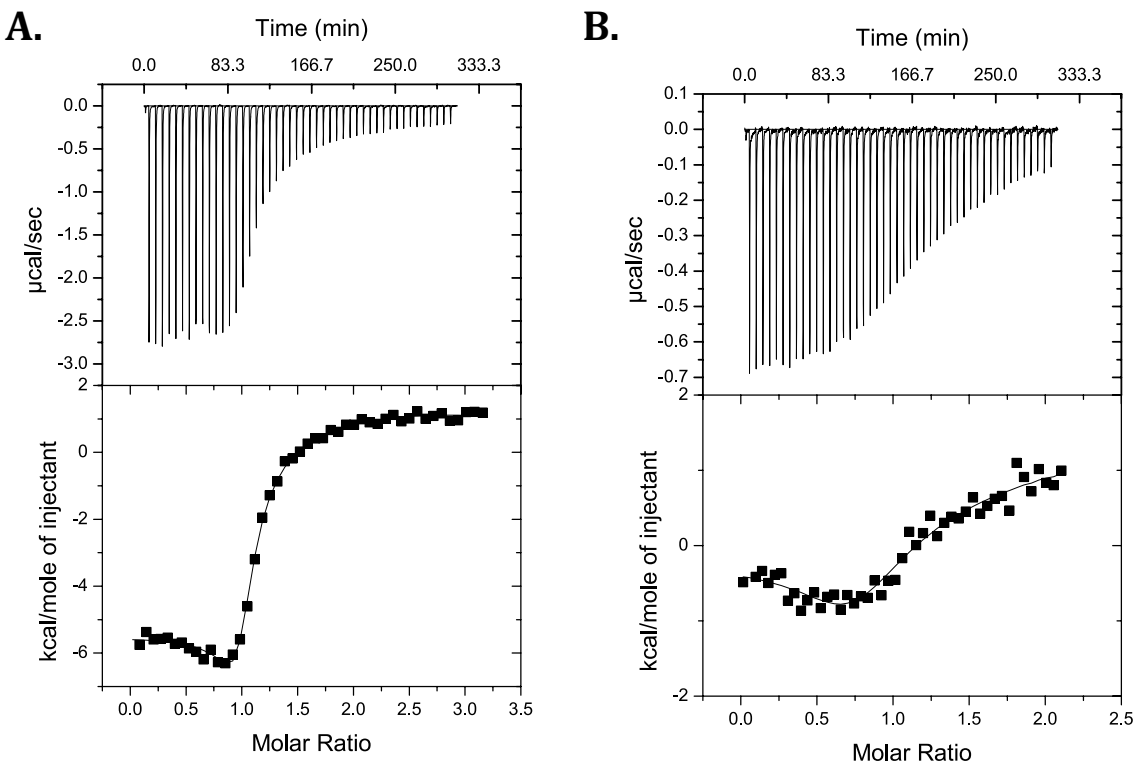


Figure 93. Size exclusion results
A. Bd1833 and Bd1833-Pta exist as a hexamer in solution, while Bd1833-ME exists as a dimer. **B.** Bd1833 mutants that we identified to affect hexamer interface also exist as hexamers in solution. **C.** Bd1833-E544R disrupts the hexamer interface to produce a mutant that forms a dimer in solution. Bd1833-E544A does not destabilise the hexamer interface sufficiently and still forms a hexamer. **D.** The same E544R mutation in the Pta domain alone (Pta-E544R) does not result in dimer formation.



C.

Site	Acetyl-CoA K (M)	Acetyl-CoA ΔH (J)	Acetyl-CoA ΔS (J mol⁻¹ K⁻¹)	CoA K (M)	CoA ΔH (J)	CoA ΔS (J mol⁻¹ K⁻¹)
1	5.7e5 ± 6.6e5	-5600 ± 88	7.6	1.8e5 ± 1.1e5	-440 ± 160	23
2	7.9e3 ± 1.6e3	-2.3E4 ± 8.2e3	-58	1.9e4 ± 1.1e4	-550 ± 1.1e4	18
3	2.9e3 ± 4.8e2	1.6e4 ± 5.2e5	71	1.2e5 ± 6.3e4	-7800 ± 7.4e4	-3.0
4	4.7e3 ± 8.4e3	3.9e4 ± 1.5e6	150	4.9e4 ± 2.0e4	1200 ± 3.8e5	25
5	1.5e5 ± 2.6e4	3.8e4 ± 1.3e6	150	1.6e5 ± 1.0e5	1.1e4 ± 4.9e5	61
6	5.4e5 ± 1.0e5	-6.9e4 ± 3.1e5	-210	1.0e5 ± 9.0e4	9500 ± 2.3e5	55

Figure 94. ITC results for Acetyl-CoA and CoA binding to Bd1833-Pta Apo
Upper panels represent the ITC titration curve and the lower panels represent the binding isotherm **A.** Acetyl-CoA binding. **B.** CoA binding. All data are fit to a sequential six-site model. **C.** A tabulated list of values obtained from the fit.

5.2.12 Bd1833 interaction with CdG

MST was used to determine whether Bd1833 interacted with CdG. A K_D value of 26 µM was obtained, suggesting weak non-specific binding. Furthermore CdG did not regulate the activity of Bd1833 in enzyme assays.

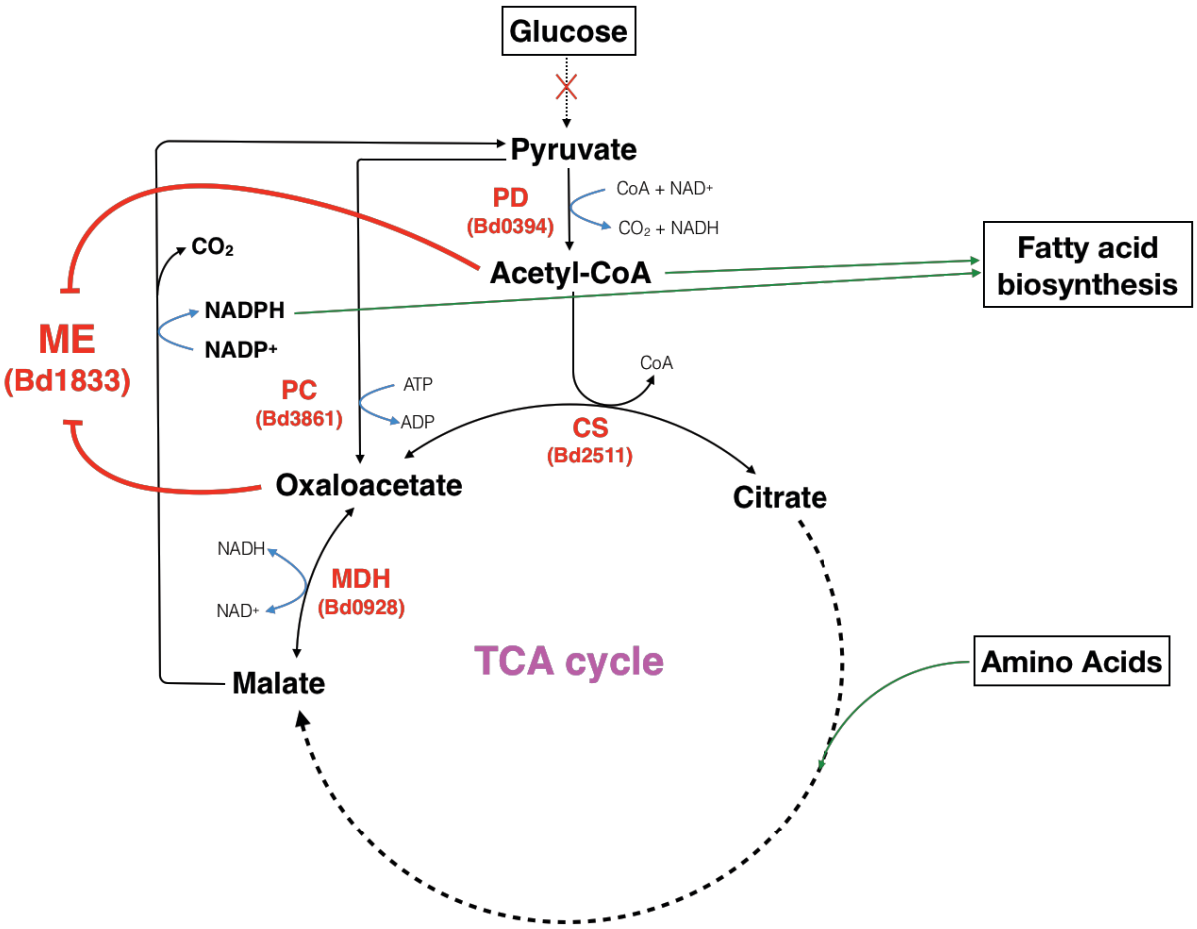


Figure 95. Schematic of Bd1833 role in metabolism
Bdellovibrio lacks sugar-metabolising pathways that are conventionally used to derive pyruvate. An increase flux of TCA cycle intermediates from prey-derived amino acids fuel *Bdellovibrio* metabolism. Bd1833 is central to subverting TCA intermediates to produce pyruvate and NADPH, which is ultimately used in anabolic reactions such as fatty acid synthesis. ME – malic enzyme, MDH – malate dehydrogenase PC – pyruvate carboxylase, PD, pyruvate dehydrogenase, CS – citrate synthase

which destabilises interactions at the hexamer interface such as E544:R749 salt bridge. The N-terminal helix of the Pta domain connects directly to the single linker helix that connects the Pta to the ME domain above. During the conformational change the N-terminal helix of the Pta rotates inwards by 8°, an action that likely impacts on the position of the multi-modular specific loop (subdomain-D). Subdomain-D could not be traced in the structure of full-length Bd1833; however, superimposition of Bd1833-ME (subdomain readily traceable) demonstrates that subdomain-D folds beneath the ME

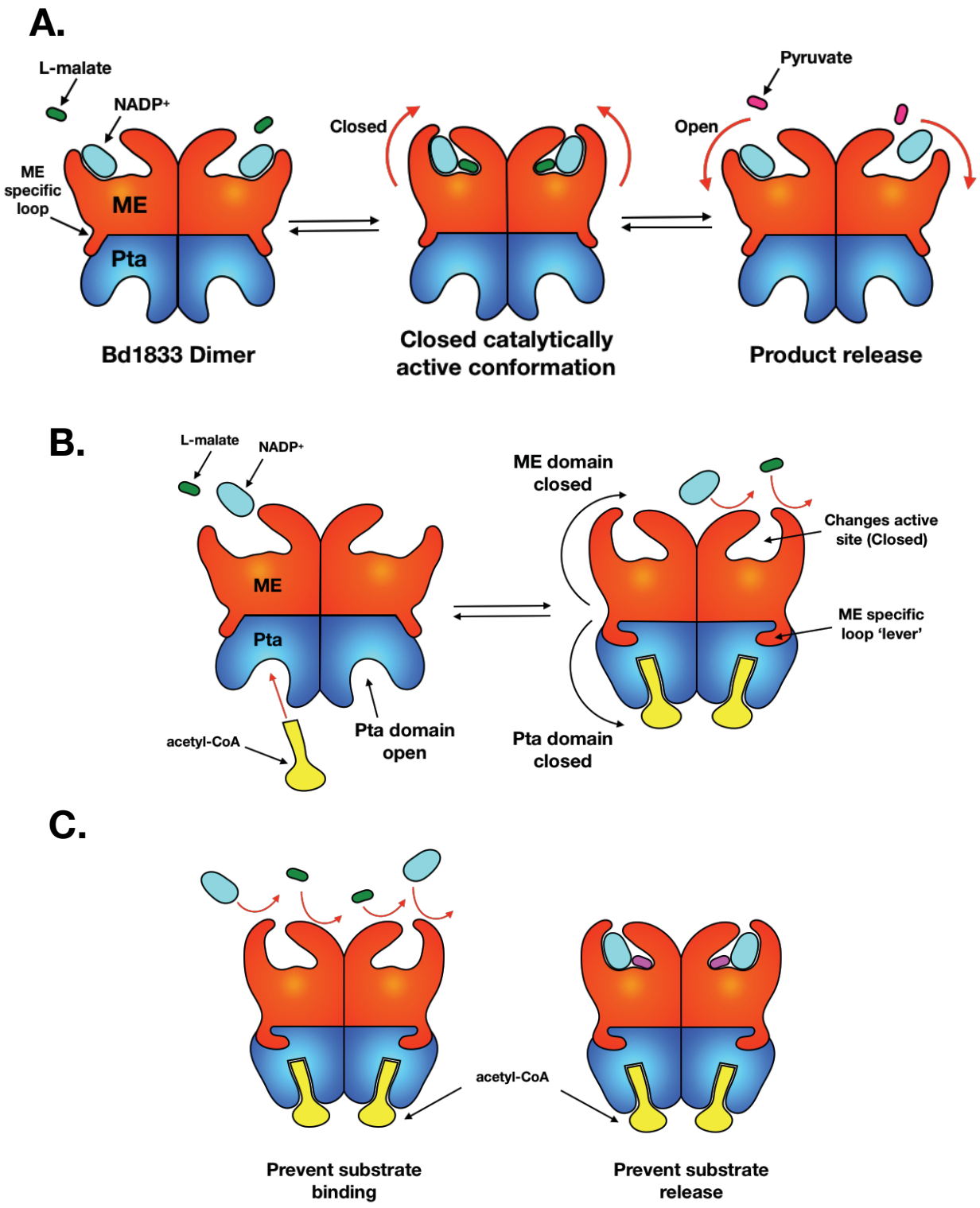


Figure 96. Cartoon of Bd1833 mechanism
A. Mechanism of Bd1833 ME activity. $NADP^+$ binds first then L-malate binds which initiates a conformational change and closes the active site. After the reaction is complete the ME domain opens and releases the products. **B.** Inhibition by acetyl-CoA. Acetyl-CoA binding initiates a conformational change in the Pta domain, which is conveyed to a conformational change in the ME domain via movement of the multi-modular specific loop. **C.** Acetyl-CoA shows non-competitive inhibition and is equally likely to bind the enzyme:substrate complex and free enzyme.

Furthermore, R581 is also altered in MaeB to a serine. Interestingly Bd1833-SKB also differs at R581 (lysine residue). Mutation of R581K could explain why Bd1833-SKB is not inhibited by CoA and inhibited less by acetyl-CoA compared to Bd1833 ($K_{i(acetyl\text{-}CoA)}$ 0.453 ± 0.019 µM for Bd1833 and $K_{i(acetyl\text{-}CoA)}$ 21.3 ± 0.84 µM for Bd1833-SKB). The acetyl-group binding pocket also differs between Bd1833 and other multi-modular ME. Q683 (H-bond acetyl-group) is not conserved and is exchanged to a His in MaeB and DME, which may explain why acetyl-CoA is selectively inhibitory (weaker 3'-P and pyrophosphate binding sites and stronger acetyl-group interactions). Together these observations demonstrate that Bd1833 likely exhibits the highest affinity CoA binding site explaining why Bd1833 activity displays inhibition to both CoA and acetyl-CoA and the additional interaction with the acetyl-Group explains why acetyl-CoA is more potent.

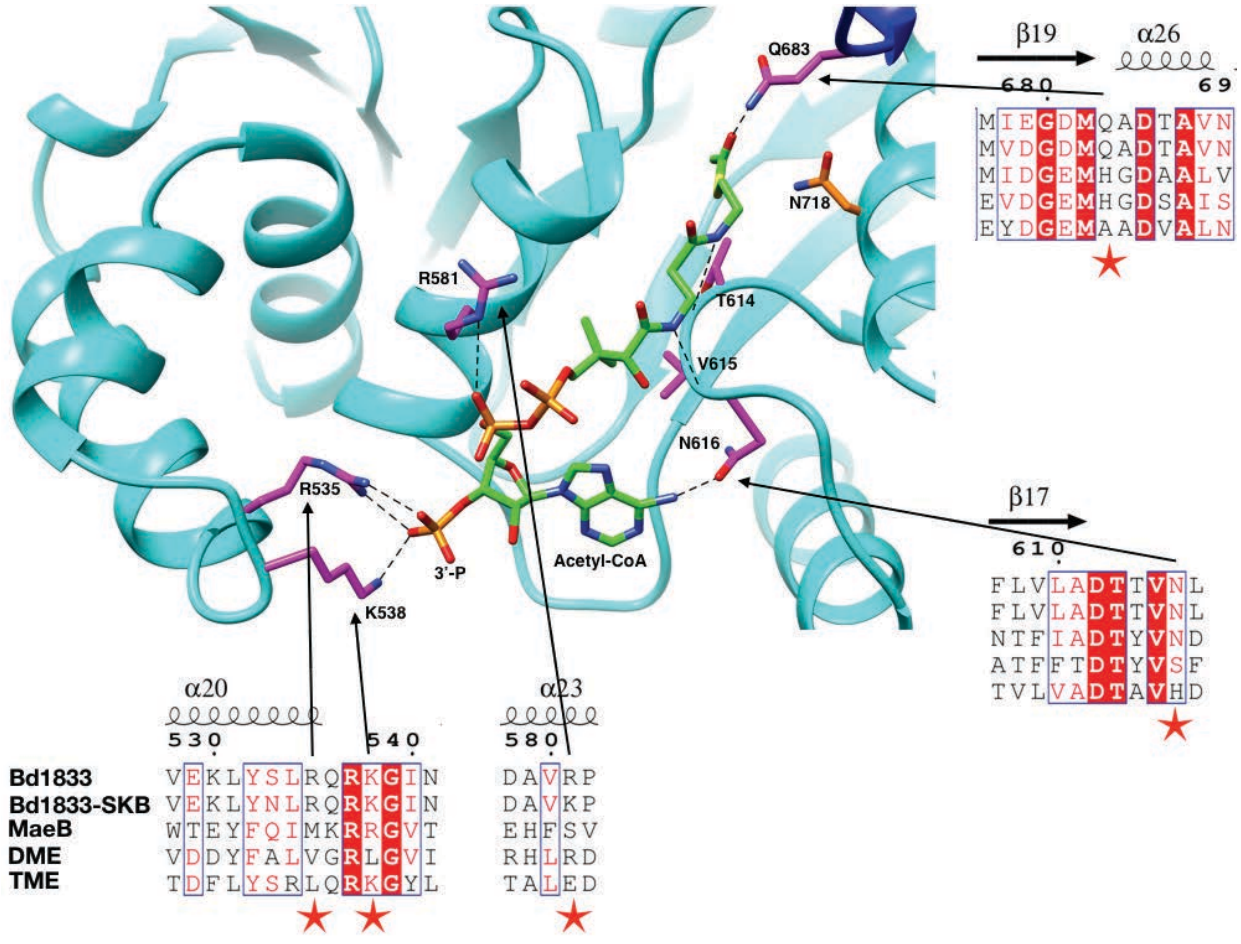


Figure 97. Bd1833 acetyl-CoA binding site comparison
The binding site of acetyl-CoA with the corresponding sections of sequence alignment of multi-modular ME, demonstrating the mutations at these locations, that likely contribute to the different binding affinities for CoA / acetyl-CoA.

The proposed CoA binding site for Pta enzymes disagrees with biochemical studies, whereas in Bd1833-Pta the observed acetyl-CoA binding site agrees with previous biochemical studies [247, 248]. The conformational change observed in Bd1833 upon acetyl-CoA binding is similar to that observed in other Pta enzymes, which has been suggested to be important for catalysis [250]. Comparison of the conformational changes observed in subdomain-I with those of Pta enzymes reveals the transition between Bd1833-Pta and Pta-Apo represents another closed state (Figure 98). Importantly the structure of Bd1833-Pta demonstrates a closed form bound to acetyl-CoA, thus in a catalytically competent state (were it to be an enzyme).

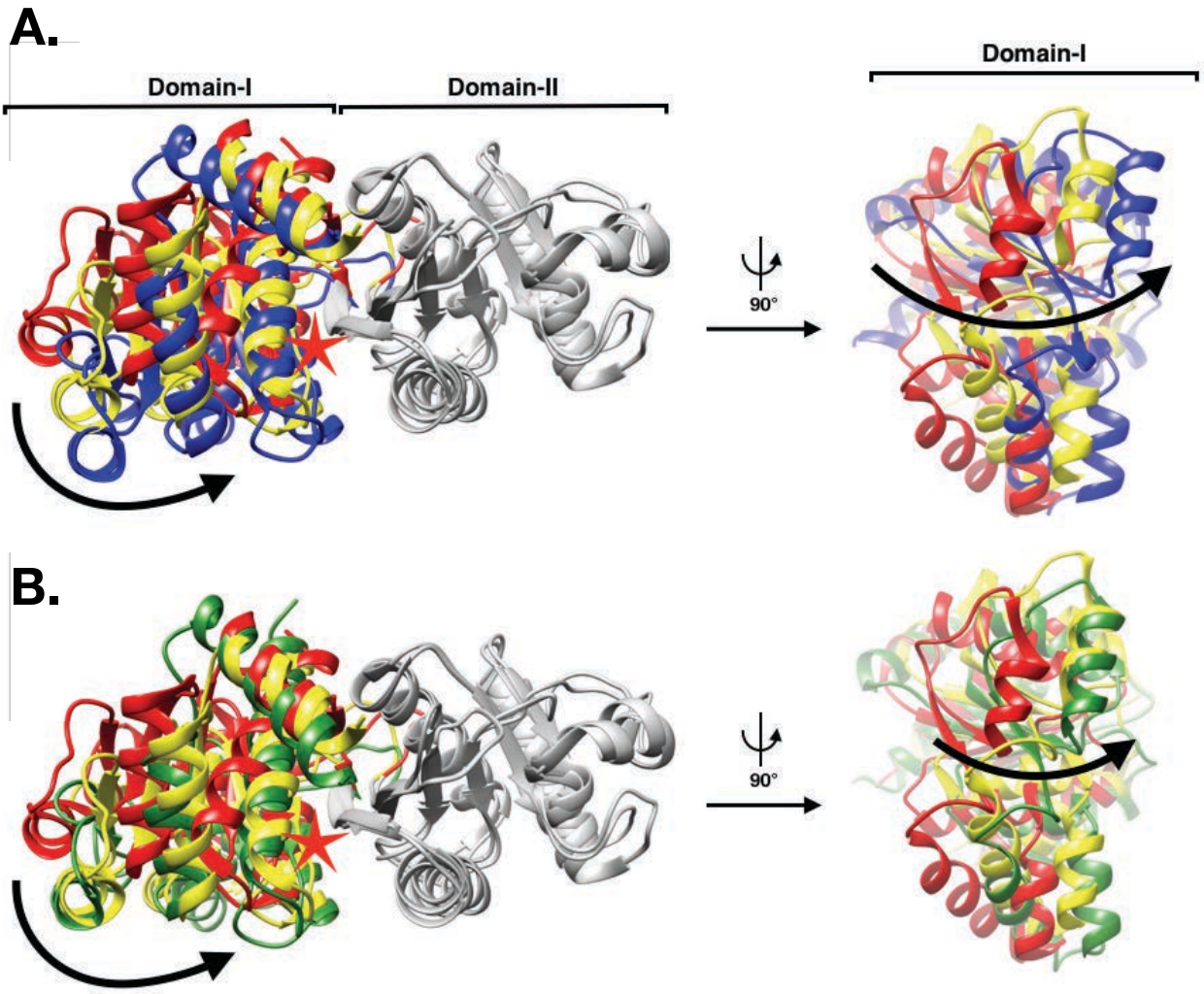


Figure 98. Pta domain movements
A. The conformational change in Bd1833-Pta subdomain-I (blue) superimposed with PDB:1QZT in the open conformation (red) and closed conformation (yellow). Bd1833-Pta adopts a different closed state. **B.** The conformation of Pta-Apo subdomain-I (green) superimposed with PDB:1QZT in open (red) and closed (yellow) conformations. Pta-Apo adopts a similar conformation as PDB:1QZT in the closed state. Superimposition was performed by aligning subdomain-II (residues 596-737) using the LSQ superpose tool in coot. (Length = 118 residues, rmsd 4.97Å).

from the thiol group of acetyl-CoA demonstrating its potential ability to fulfil this role [250].

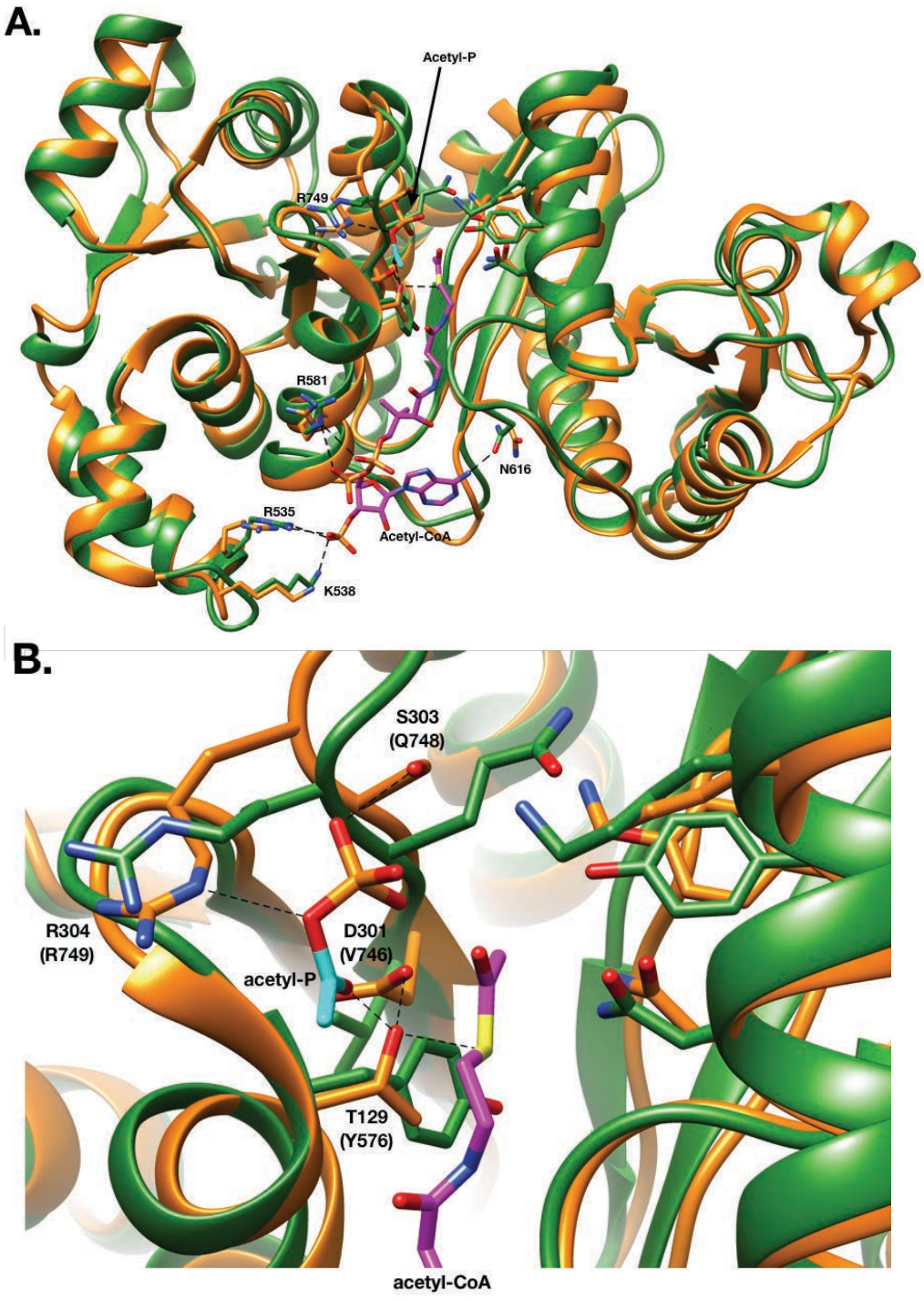


Figure 99. Pta active site
A. Bd1833-Pta monomer (green) superimposed with PDB:1XCO Pta monomer (orange). Superimposition was performed by individually aligning Bd1833-Pta subdomain-1 and II with the individual domains of PDB:1XCO. Bd1833-Pta subdomain-I (residues 460-587) with PDB:1XCO domain-I (residues 14-297) (length =174 residues, rmsd = 1.43Å). Bd1833-Pta subdomain-II (residues 596-737Å) with PDB:1XCO domain-II (residues 150-292) (length = 144 residues, rmsd = 1.57). The active site is likely located in the upper portion of the substrate-binding groove **B.** A close up view of proposed active site. The catalytic residues of 1XCO are labelled alongside the comparative Bd1833-Pta residues, demonstrating substitutions. The thiol group is within H-bonding distance of T129. T129 is in H-bond distance of D301 and acetyl-P.

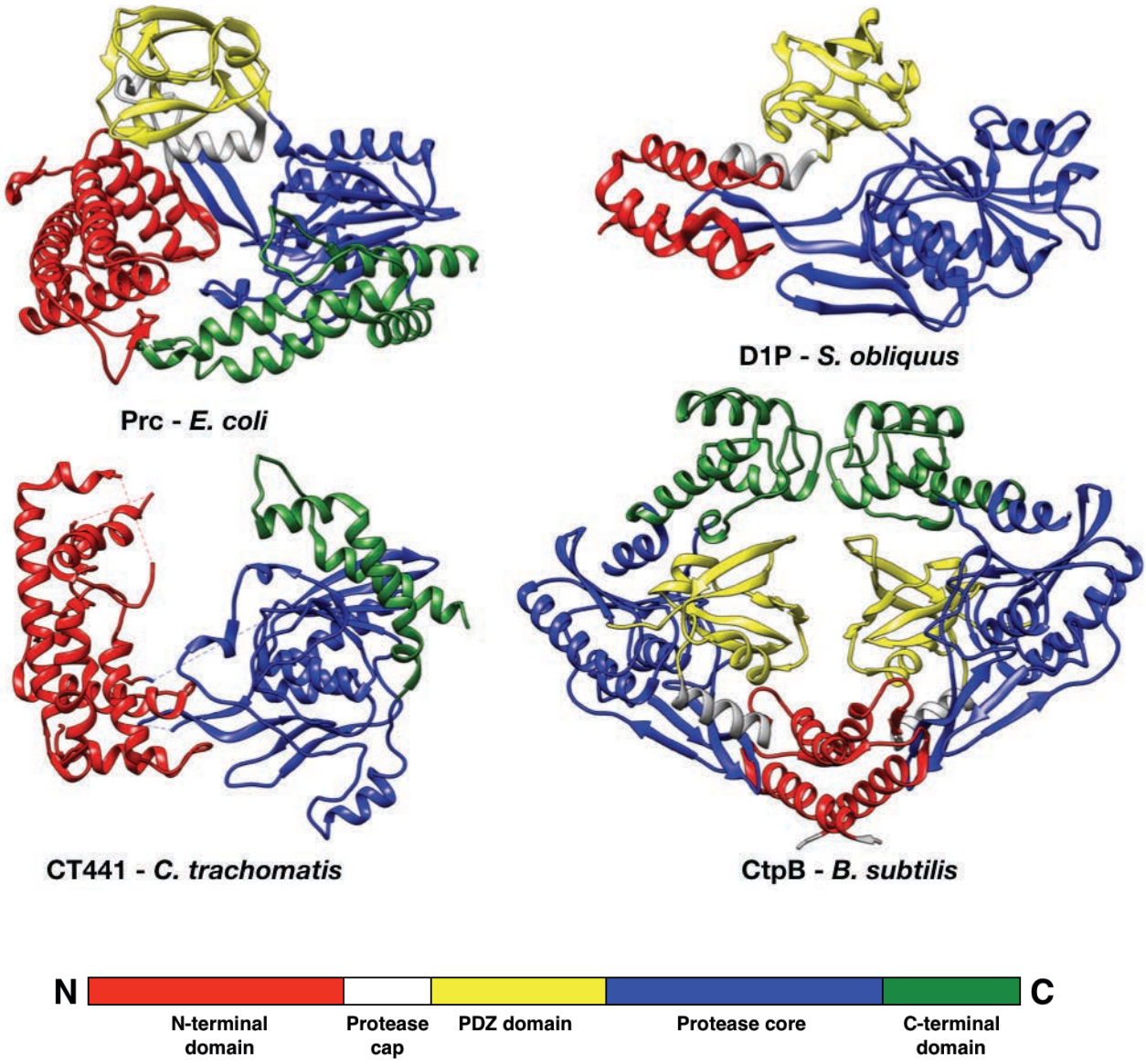


Figure 100. Structures of C-terminal processing proteases
The four solved structures of C-terminal processing proteases are presented and coloured by their domains. NTD in red, PDZ domain in yellow, protease domain in blue, protease cap helix in grey and CTD in green. Prc – 5WQL, D1P – 1FC7, CT441 – 4QL6, CtpB – 4C2C.

6.1.3.1 D1P

The first three-dimensional structure solved for a CTP was the structure of D1P, which consists of three loosely packed, distinct domains: an α-helical N-terminal domain, a central PDZ domain and a larger C-terminal protease domain. The PDZ domain is composed predominantly of a twisted five-stranded β-sheet with two α-helices. The protease domain is folded into two β-sheet regions. The larger, mixed five-stranded β-

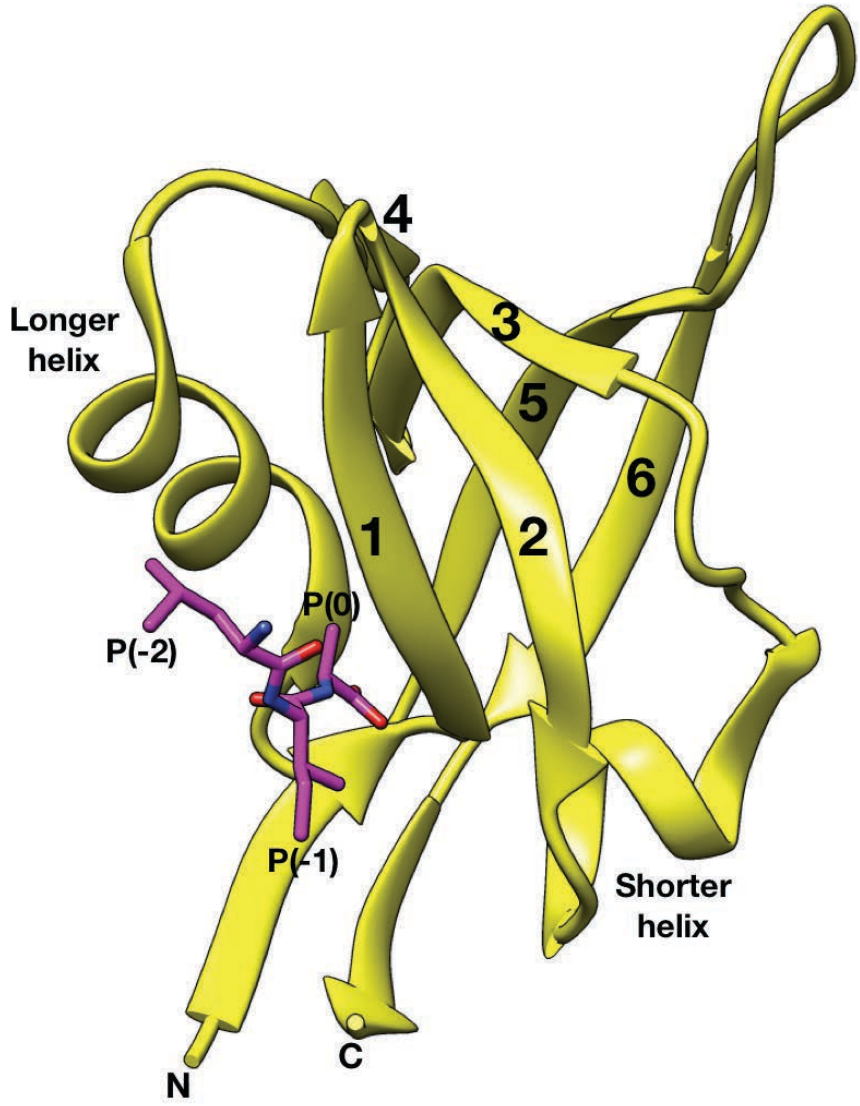


Figure 101. PDZ domain structure
The structure of a C-terminal processing protease PDZ domain is shown here bound to a 3aa peptide substrate. The peptide binds in a β-augmentation manner with strand-1. The peptide residues are numbered according to the text. The N and C-terminal β-strands are specific to CTPs. (PDB: 6BXG)

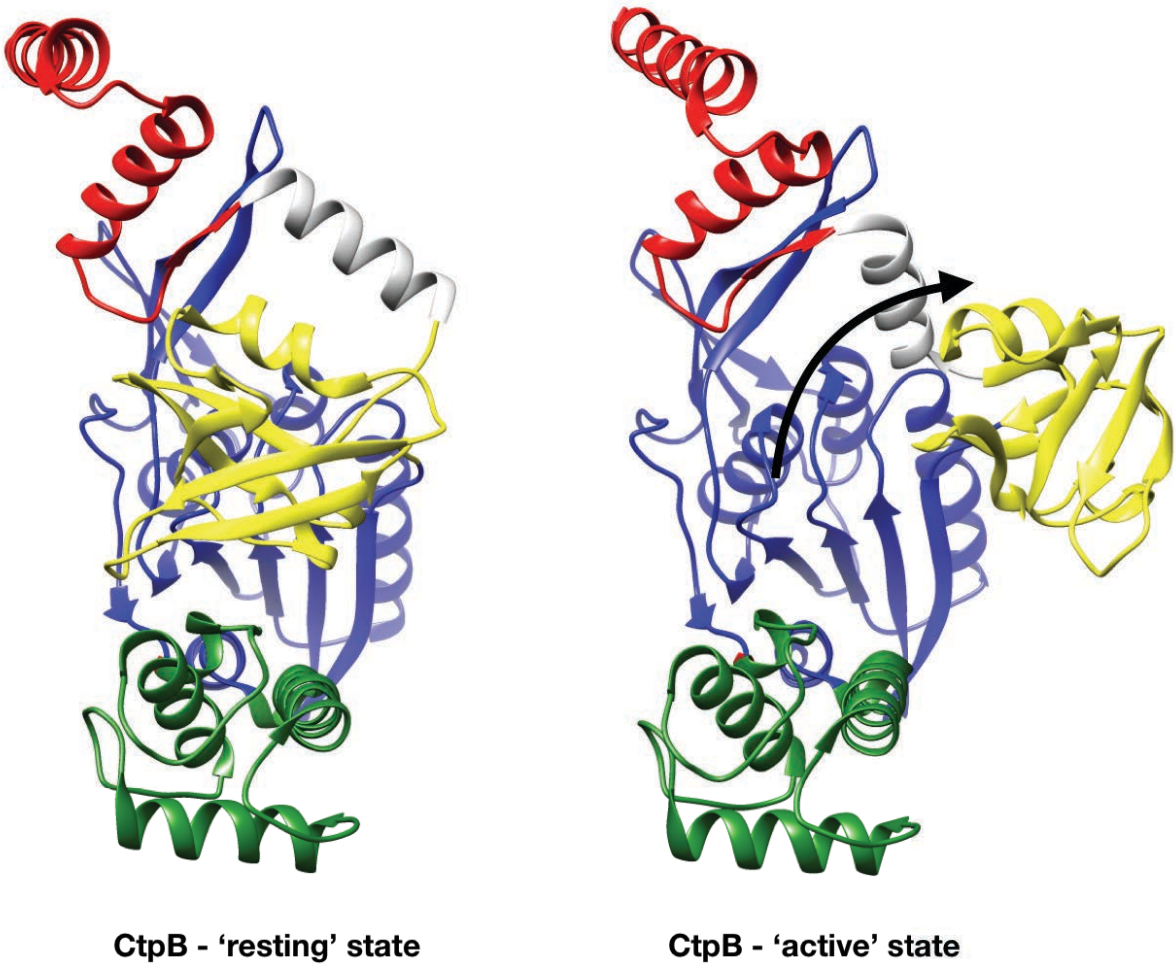


Figure 102. PDZ domain movement in CtpB
The mechanism of activation and regulation of C-terminal processing protease by the PDZ domain has been well documented in CtpB. The PDZ domain coloured (yellow), has been observed in multiple conformations and samples between a 'resting' and an 'active ' state. In the 'resting' state the PDZ domain is tightly tucked into the protease domain (blue). Whereas in the 'active' state (peptide bound) the PDZ domain pulls away from the protease domain. The black arrow represents the movement. The switch also repositions the protease cap (grey). PDB:4C2C & 4C2E

R168 was shown to be a key residue facilitating the sensing of peptide binding at the PDZ binding cleft and transferring the signal allosterically to bring about the conformational switch of CtpB. In the 'resting' state the sidechain of R168 faces away from the PDZ domain, making contacts with S106 located on protease cap. In the 'active' state the sidechain of R168 repositions, aligning with the bound peptide but still making contacts with S106. By maintaining these inter-domain contacts in the bound

the molecular ruler, where the distance between PDZ domain and proteolytic chamber is a measured uniform distance. The second is a processive cleavage mechanism, where the cleavage of the peptide chain at the proteolytic chamber creates a new substrate C-terminus that is subsequently recognised by the PDZ domain. The peptide-binding sites at the PDZ domain and protease domain would thus have to be highly similar. This mechanism would operate in a ratchet like manner and allow continued proteolytic processing of the substrate. The other possible mechanism involves PDZ imposed oligomer reorganisation by a conformational state change (not a proposed mechanism for CTPs) [271]. Figure 103 illustrates a cartoon representation for PDZ recognition and substrate presentation in CTPs.

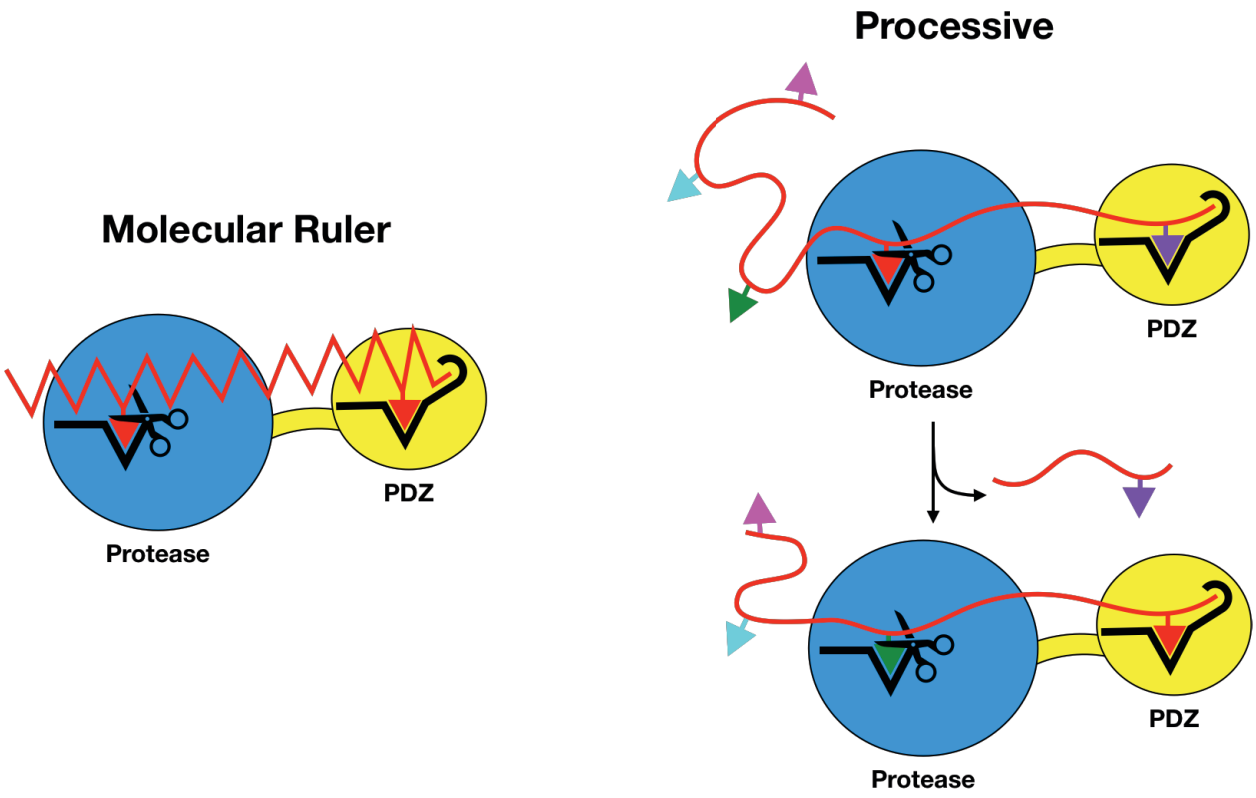


Figure 103. C-terminal processing protease mechanism
Two mechanisms are proposed for C-terminal processing proteases. **A.** A molecular ruler mechanism, where the distance between PDZ domain and proteolytic site is set. **B.** A processive mechanism, where the protease and PDZ binding sites are similar. The C-terminus of the new peptide becomes a ligand for the PDZ domain. Red line depicts the substrate and the coloured triangles demonstrates the recognition sequences (different colours demonstrate a different amino acid sequence.

6.1.6 Bd0967 role in predatory cycle

Bd0967 has been linked to a larger network of cellular signalling events in *Bdellovibrio,* possibly controlled by CdG [82]. Moreover, Bd0967 is a putative CTP, containing a PDZ domain for possible co-ordination of protein-protein interactions. The precise role of Bd0967 during the lifecycle of *Bdellovibrio* is currently unknown; however, an important role during cellular signalling and lifecycle progression is highly likely, due to our current knowledge of cellular and biological roles for CTPs in other organisms.

In addition to Bd0967, numerous other PDZ-containing proteins are found in the genome of *Bdellovibrio HD100*, which are listed in Table 15. Seven of the nine proteins identified as PDZ-containing proteins are proteases, one of which (Bd3534) has been shown to be an essential predation protein [293].

Uniprot ID	Gene name	Protein type	Length
Q6MGY2	Bd3787	Hypothetical zinc metalloprotease	557
Q6MHK7	Bd3534	Carboxyl-terminal protease	526
Q6MJH7	Bd2800	Serine protease MucD	474
Q6MP94	Bd0967	Tail-specific protease	673
Q6MRC0	Bd0169	Carboxyl-terminal processing protease	458
Q6MNK9	Bd1239	Carboxy-terminal processing protease	401
Q6MMM7	Bd1598	General secretion pathway protein C	304
Q6MPD5	Bd0922	Periplasmic serine protease	476
Q6MLF8	Bd2053	Component of the Tol biopolymer transport system	974

Table 15. PDZ proteins in *Bdellovibrio HD100* genome
A list of PDZ containing proteins found in the genome of *Bdellovibrio HD100.* The majority of the proteins are proteases. Bd3534 (green) has been identified as an essential predation protein.

6.2 Results

6.2.1 Obtaining Bd0967 crystals

A construct of Bd0967 was created to encompass the full-length protein with the exclusion of the seven C-terminal amino acids that were predicted to be disordered. The protein was tagged with a His_6-tag at the N and C terminus. Closer inspection of the amino acid sequence revealed that Bd0967 possessed a probable signal peptide with a signal peptide cleavage site (LASA) located 24aa from the N-terminus. Thus Bd0967 is likely to be trafficked to the periplasm where it is functionally active. An unambiguous N-terminal sequence of QLKEGLE… was determined by ESI-MS, confirming purified Bd0967 had been processed to remove the 24aa signal peptide (ESI-MS was performed by the Stříšovský lab).

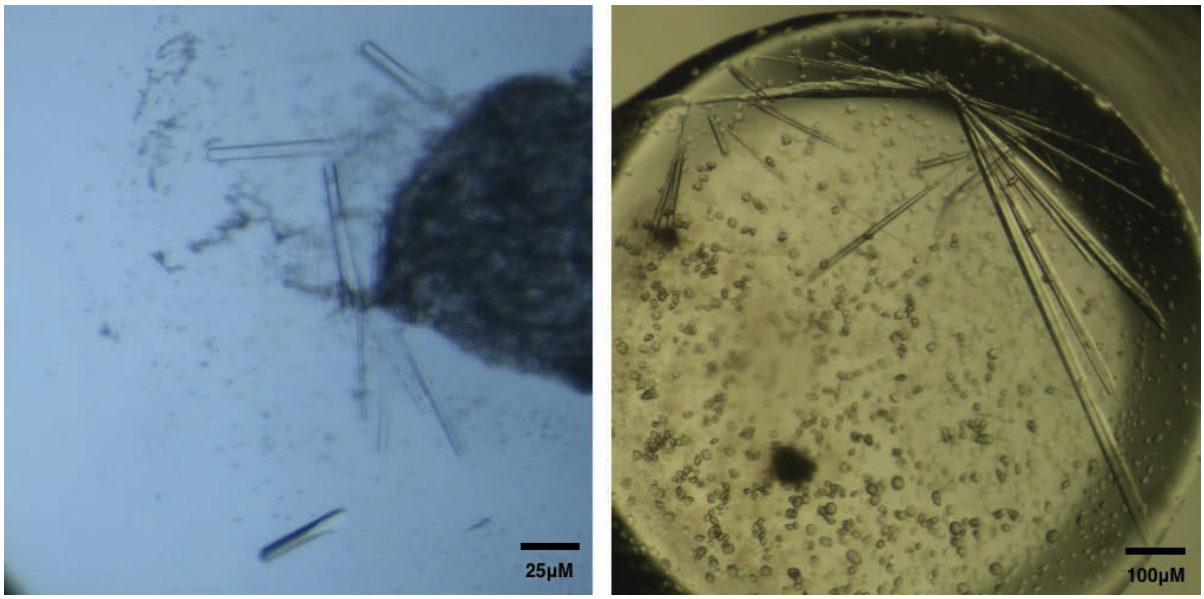


Figure 104. Bd0967 crystallisation
Images of Bd0967 crystals. Higher quality crystals were obtained by using random matrix microseeding as described in Section 2.6.9.

stranded, anti-parallel, β-sheet) is attached between the protease domain and the protease cap via a flexible hinge.

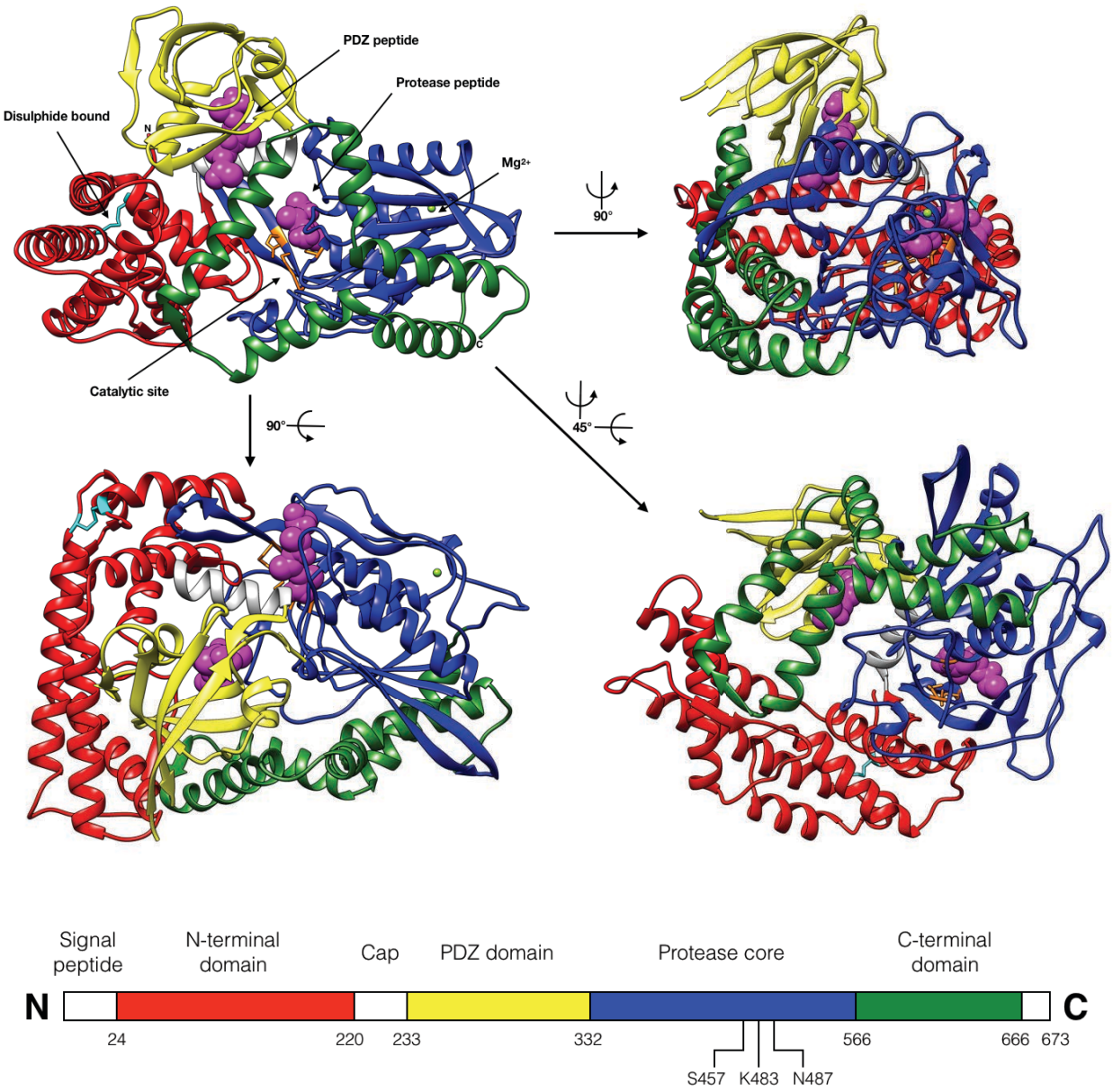


Figure 105. Domain architecture of Bd0967
Bd0967 can be divided into four distinct domains and a subdomain, which are represented by the different colours in this image. The scaffold-like NTD (red), the protease cap helix (grey), the PDZ domain responsible for peptide recognition (yellow), the protease domain (blue) and finally the paddle-like CTD (green). Two peptide fragments (magenta) are bound to the PDZ domain and protease domain. The active-site residues are highlighted in orange, the NTD disulphide bond is shown in cyan and the metal ion bound to the protease domain is represented as a green sphere.

Larger B-factor values are associated with the CTD domain, which suggests that it is more flexible and likely alters conformation to facilitate the movement associated with PDZ domains in this family of enzymes. Higher B-factors are also associated with the PDZ domain, which is known to adopt multiple conformations in other CTPs. A map of B-factors for Bd0967 is shown in Figure 106, which illustrates that the base of the protein (protease and lower regions of NTD and CTD) has lower B-factor values, whereas the upper regions of the protein have higher B-factor value suggesting these are more flexible.

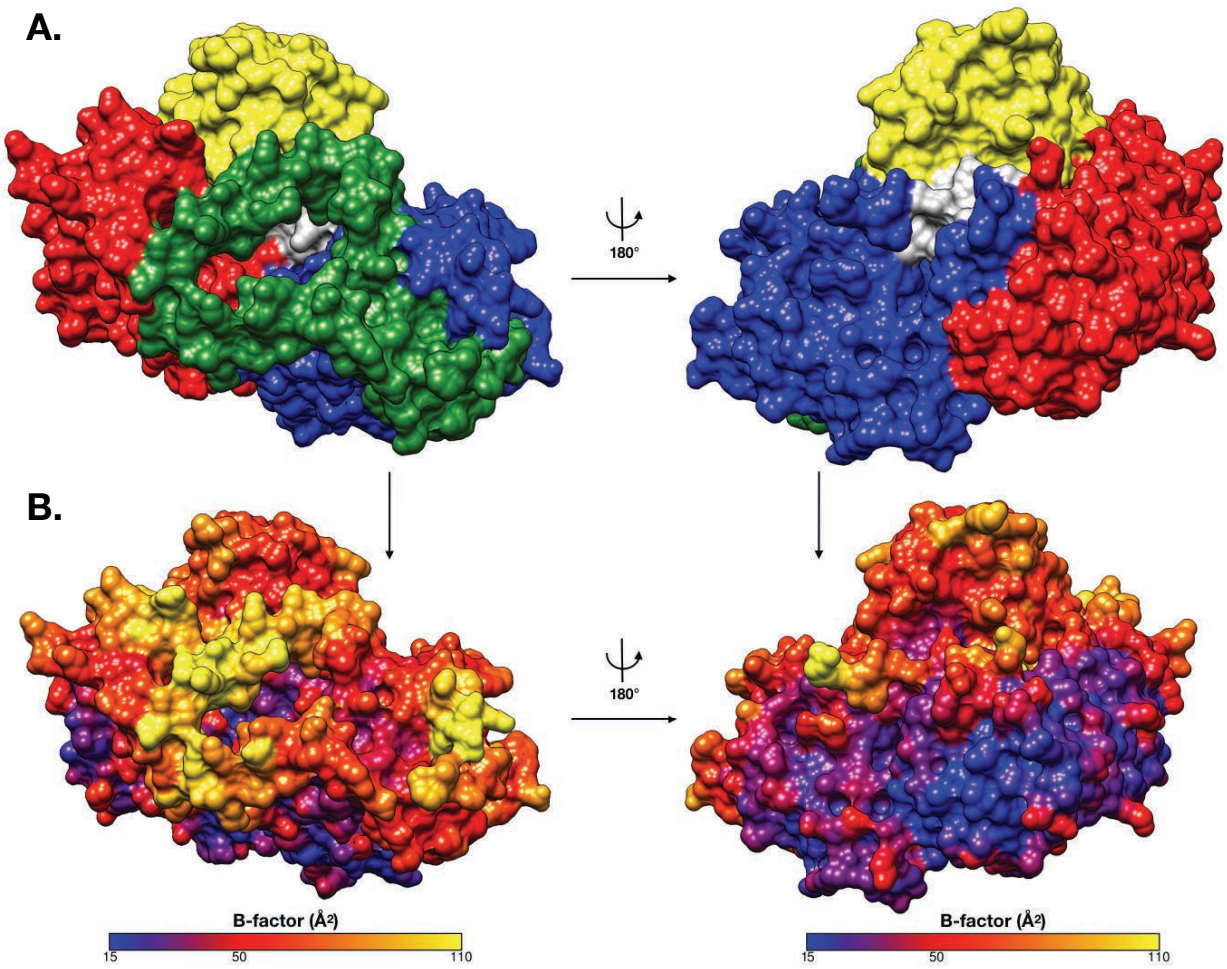


Figure 106. Bd0967 B-factors
A. Represents Bd0967 orientation (front and back), coloured according to domains. **B.** Bd0967 in the same orientation as panel A, coloured according to B-factor. Bd0967 structure has regions of contrasting B-factor values suggesting certain regions such as the CTD and PDZ are more dynamic, whereas the protease domain is more static. Average B-factor values per residue have been coloured using a gradient from low B-factor (blue) to high B-factor (yellow).

6.2.3 Bd0967 cavity

The structure of Bd0967 creates a large self-compartmentalised cavity. The volume of the cavity was accessed using CASTp, which calculates the MS volume to be 19500Å^3 with a pocket surface area of 7200Å^2 [294]. The dimensions of the interior cavity have been calculated roughly as 30Å x 30Å x 30Å (WxHxD). The surface area of the cavity space is represented in Figure 107, coloured in magenta. The cavity would be able to accommodate small unfolded proteins and may provide an environment shielded from bulk solvent.

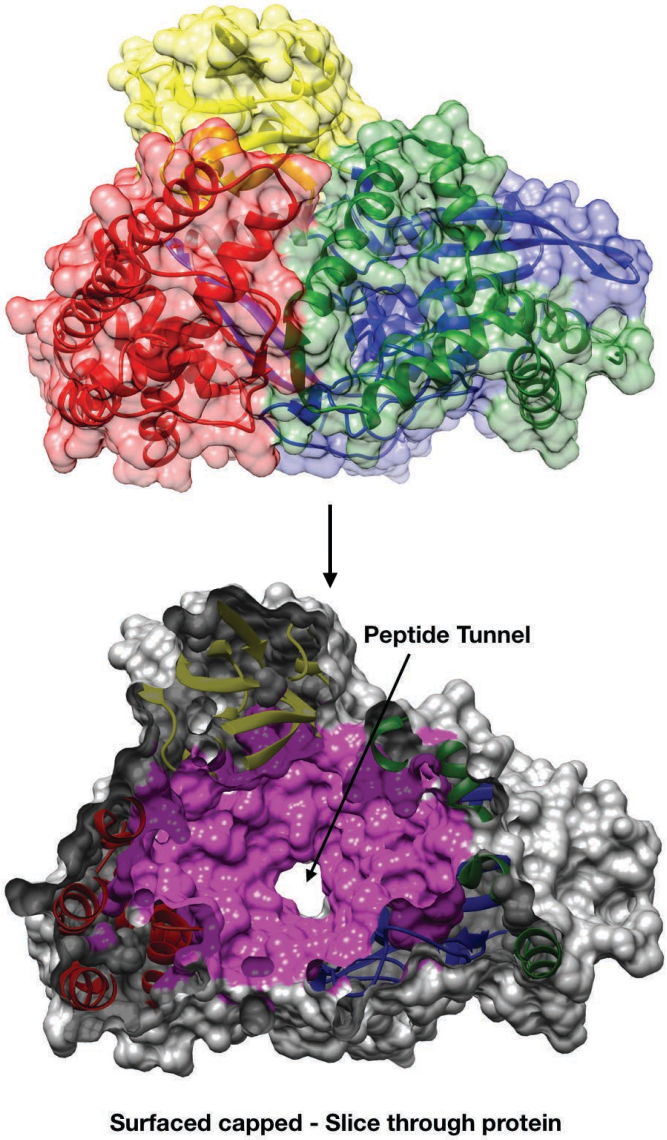


Figure 107. Bd0967 internal cavity
Bd0967 domain organisation creates a self-compartmentalised enzyme with a large internal cavity. The top image shows the exterior surface of Bd0967 coloured according to the domains. The bottom image shows a dissection of Bd0967 (orientation is the same) revealing the internal cavity surface, calculated using CASTp, shown in magenta. The peptide tunnel leading into the internal cavity is labelled.

6.2.4 The Active site / catalytic triad / bound peptide

The active site of Bd0967 is located in the protease domain with the catalytic triad positioned at the end of a shallow tunnel leading into the cavity. The peptide-binding tunnel is formed between the protease domain and protease cap subdomain. The catalytic triad is comprised of the proposed active-site nucleophile S457, the catalytic base K433, and Q487 that makes a hydrogen bond with K433 to orientate it correctly for catalysis similarly to other CTPs.

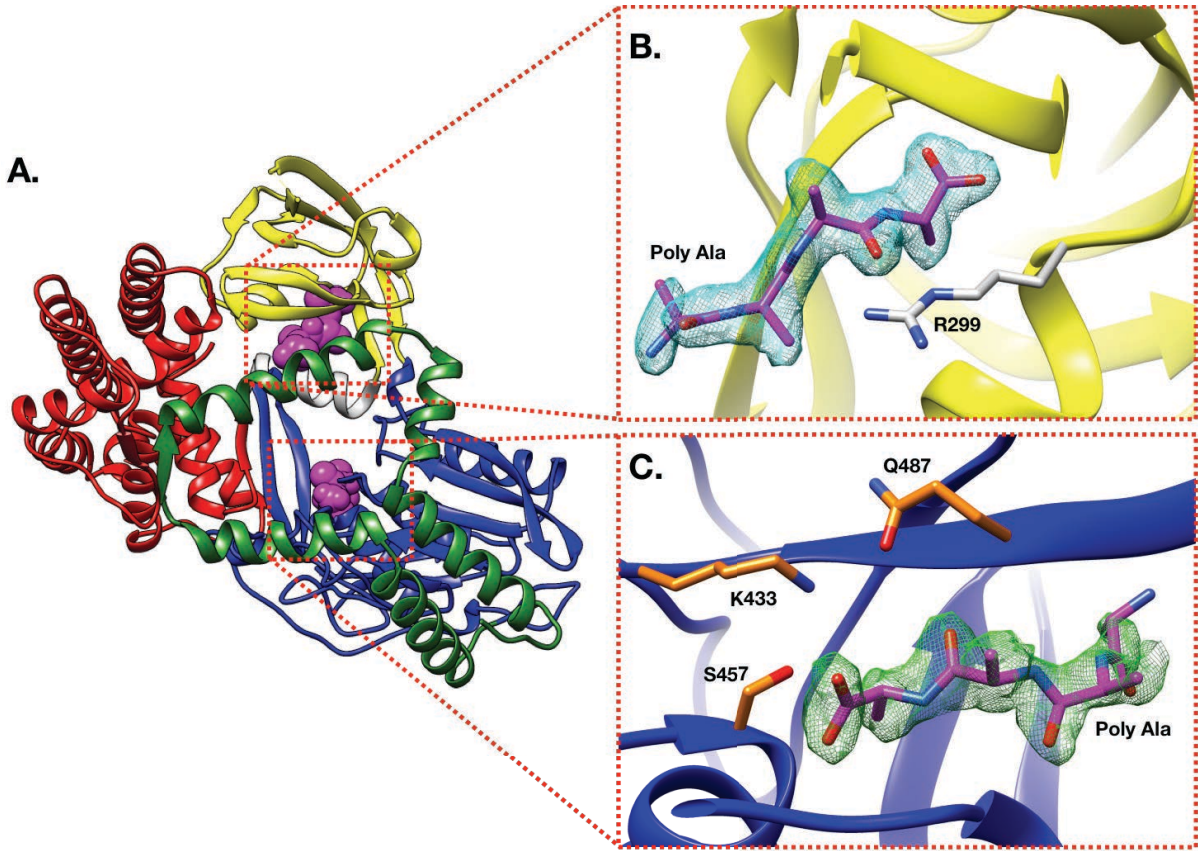


Figure 108. Bound peptides of Bd0967
Bd0967 co-purifies and co-crystallises with two peptide fragments that bind to the PDZ peptide binding site and the active-site peptide binding site. **A.** The peptides coloured magenta in space filling shown to display there positions within the structure of Bd0967. **B.** Shows the electron density map (cyan) of the bound peptide (magenta) at the PDZ domain. R299 is shown in grey for reference. **C.** Shows the electron density map (green) of the bound peptide (magenta) at the protease domain. The active-site residues are shown in orange as a reference. ($2F_o$-F_c map, radius=1.5Å, σ=1.3)

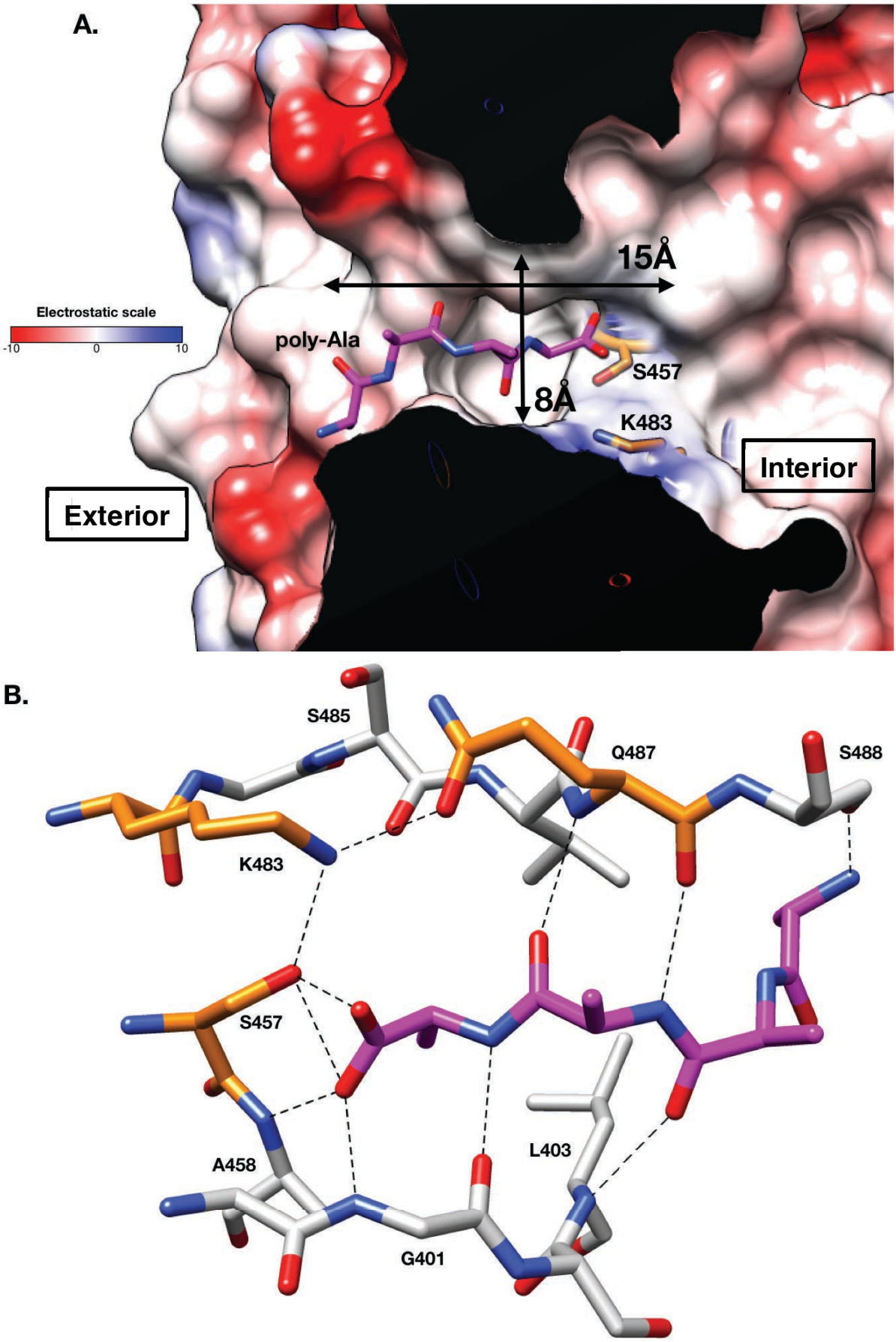


Figure 109. Protease peptide tunnel and binding site
A. Bd0967 has a shallow tunnel leading to the interior cavity, formed between the protease domain and the protease cap subdomain. The tunnel is mainly lined with hydrophobic residues. The active site residues are located towards the end of the tunnel near the interior opening. **B.** Binding of the peptide at the active site is facilitated by multiple hydrogen bonds. The peptide, coloured in magenta, is bound in a beta-augmented fashion. The active-site residues are coloured in orange and other residues participating in the peptide binding are in grey. Black lines depict the hydrogen bonds.

6.2.5 Bd0967 has a unique metal-binding site

Bd0967 possesses a metal-binding site that is unique to this family of CTPs. The metal-binding site is located in a peripheral loop region of the protease domain (Figure 110). The metal is co-ordinated by seven atoms: a water molecule; a single side chain residue (D436) and five backbone carbonyl groups (G411, F414, Q415, D438 and M440). Binding of the metal ion is facilitated by the placement of P439 that allows the specific orientation of backbone carbonyl groups in this region. The binding of a metal ion at this site likely provides a structural role by stabilising the otherwise flexible coiled region. A Mg^{2+} ion has been modelled in this position due to its presence in the crystallisation conditions and is in agreement with the observed diffraction data.

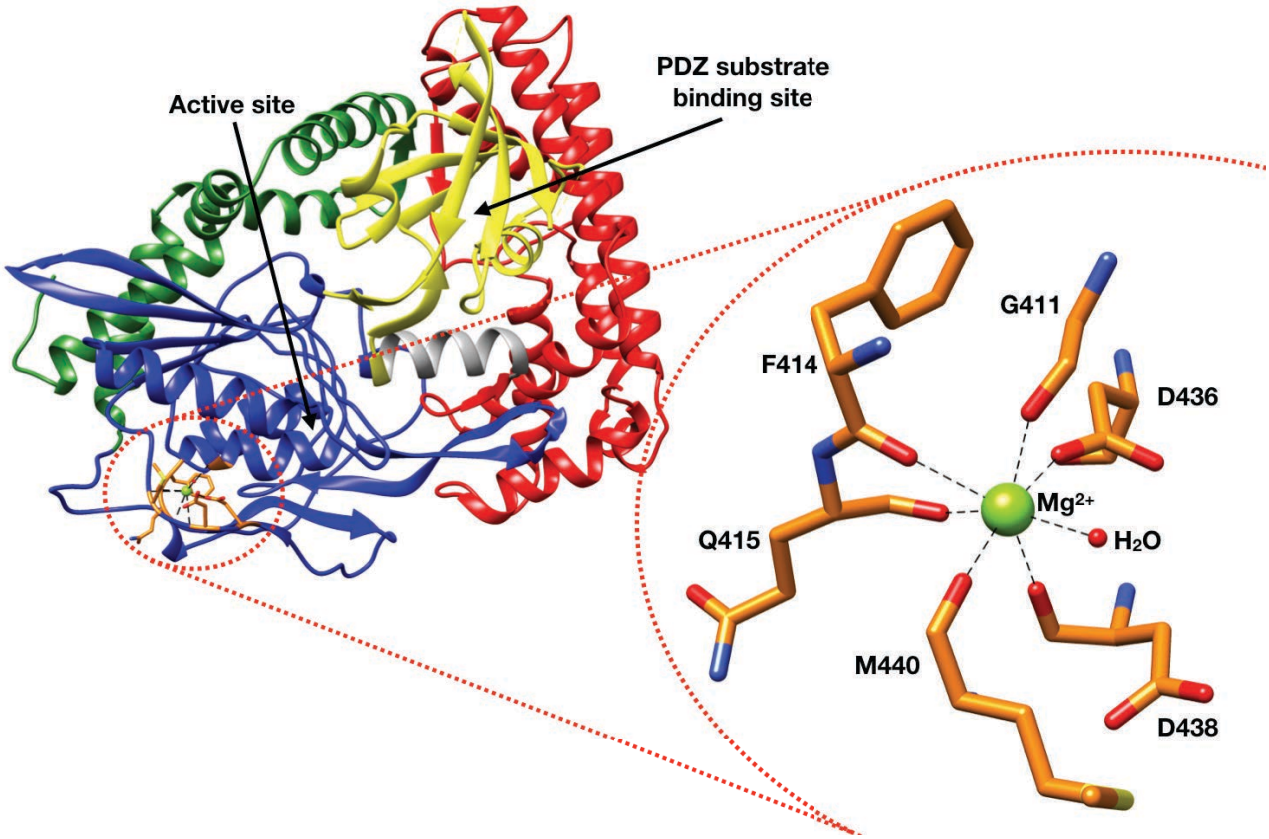


Figure 110. Bd0967 metal-binding site
The metal-binding site of Bd0967 is located in a loop at the base of the protease domain (blue).The residues involved in binding the metal ion are coloured in orange. The Mg^{2+} metal ion is coloured green and makes seven coordinated bonds to five backbone carbonyl groups the carboxylate group of D436 side chain and also to a coordinated water molecule.

6.2.6 Bd0967 has unique disulphide bond

Bd0967 has a disulphide bond located in the NTD and is not a feature previously documented for CTP (Figure 105). The disulphide bond is formed between Cys33 and Cys100 and acts to stabilise the base of the α-helical scaffold domain (NTD). Cys33 is located 4aa from the N-terminal peptide resolved in the structure of Bd0967, beyond this point no electron density is observed. The disulphide bond likely acts to stabilise the N-terminus and therefore the whole NTD, which provides the scaffold for the structure of Bd0967. The existence of a bona fide disulphide bond adds further evidence that Bd0967 possesses a signal peptide and is likely to be transported to the periplasm for its specific function.

6.2.7 Substrate recognition by the PDZ domain

The PDZ domain of Bd0967 is attached to main body of Bd0967 on a flexible hinge. At the base of the hinge are two hydrophobic residues, L234 and L332 that could easily slide past each other, in a greased motion, to facilitate the movement of the PDZ domain.

An additional peptide fragment is bound in the PDZ peptide binding groove, which is believed to originate from the same protein as the peptide fragment bound to the protease domain. We have modelled the peptide fragment, bound at the PDZ binding groove, as a four amino acid poly-alanine chain. Although positive density in the refined F_o-F_c electron density map exists for residue sidechains, at the resolution obtained the sequence register is ambiguous, furthermore MS was unable to identifying the peptide fragment.

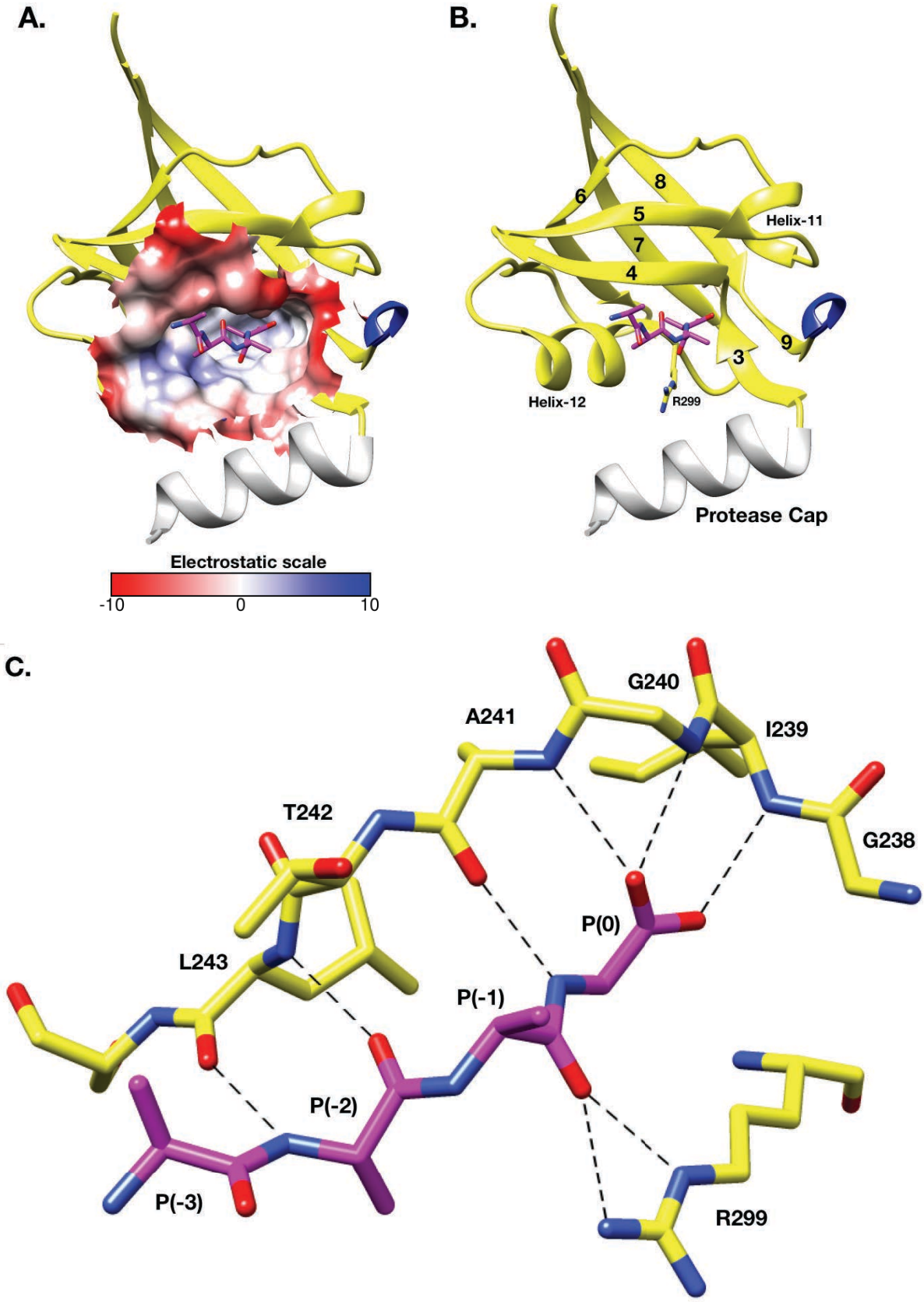


Figure 111. Peptide binding PDZ domain
A. The peptide binding groove of the PDZ domain is highly hydrophobic. The surface of the peptide binding groove is shown coloured according to the coulombic scale. **B.** Image of PDZ domain binding to peptide in same orientation as (A) to highlight the relative position of R299. The strands and helices are labelled for reference to text. **C.** Hydrogen bond network involved in peptide binding at the PDZ domain.

In addition the sidechain of R299 makes likely important hydrogen bonds to the peptide chain. The R299 sidechain aligns parallel to the peptide chain making a β-sheet like contact. R299 sidechain makes additional hydrogen bonding contacts with E226 and D227 of the protease cap (Figure 112). The H-bonds R299 makes are likely to be important for substrate sensing and activation of the protease domain, similarly to the observed conformational change observed between the 'resting' and 'active' states of CtpB [85].

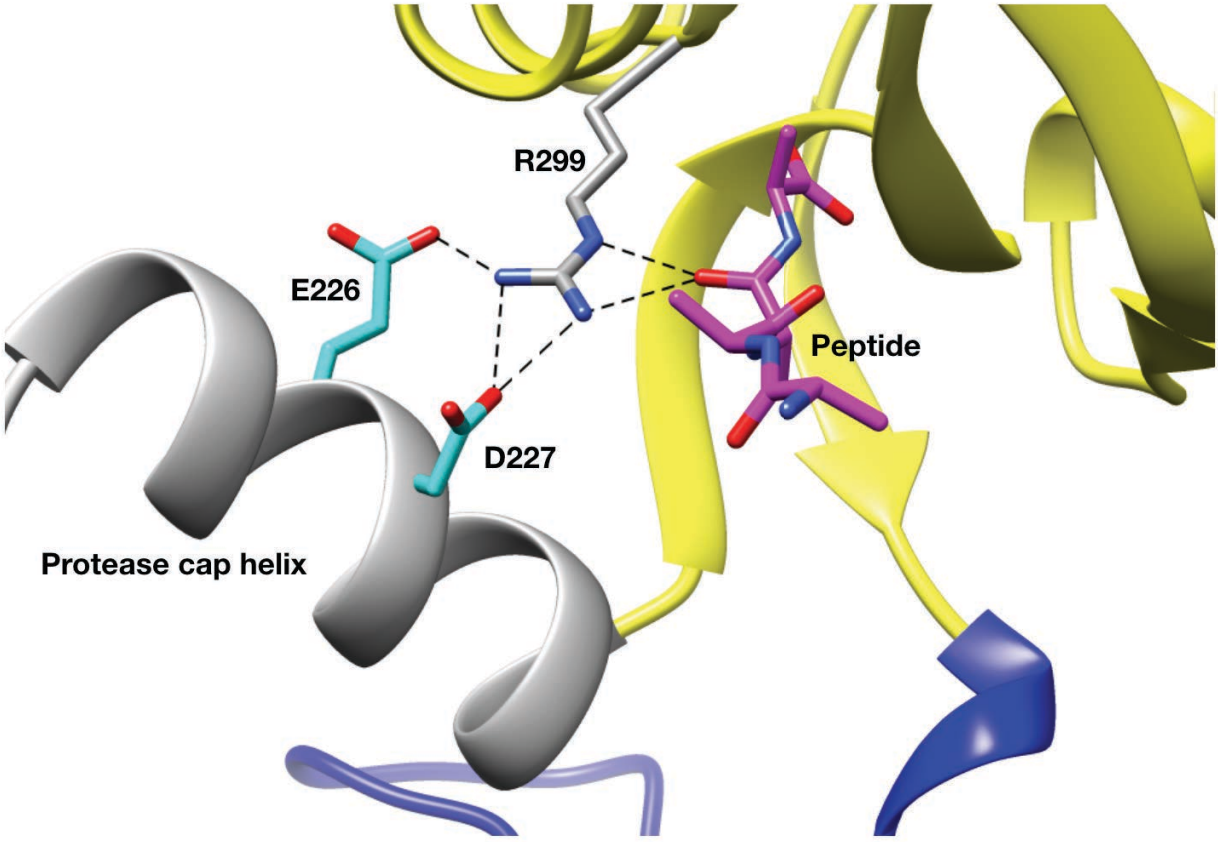


Figure 112. Arg299 interaction with peptide and protease cap helix
Peptide binding at the PDZ domain is facilitated by R299 by imitating a beta strand. R299 makes similar hydrogen bonding contacts with the protease cap helix that are responsible for switching the protease domain on and off in CtpB. R299 hydrogen bonds to two acidic residues E226 and D227 (cyan) as well as maintaining hydrogen bonds with the bound peptide (magenta).

6.2.8 A monomer in solution

SEC was used to determine the oligomeric state of Bd0967 in solution. The crystal structures also suggest Bd0967 is a monomer; however, homologues of Bd0967 have been shown to exist as functionally active dimers. SEC of Bd0967 (Wt, S457A and R199M constructs) revealed all exist in a monomeric form; however, a sizeable void peak was present for all constructs.

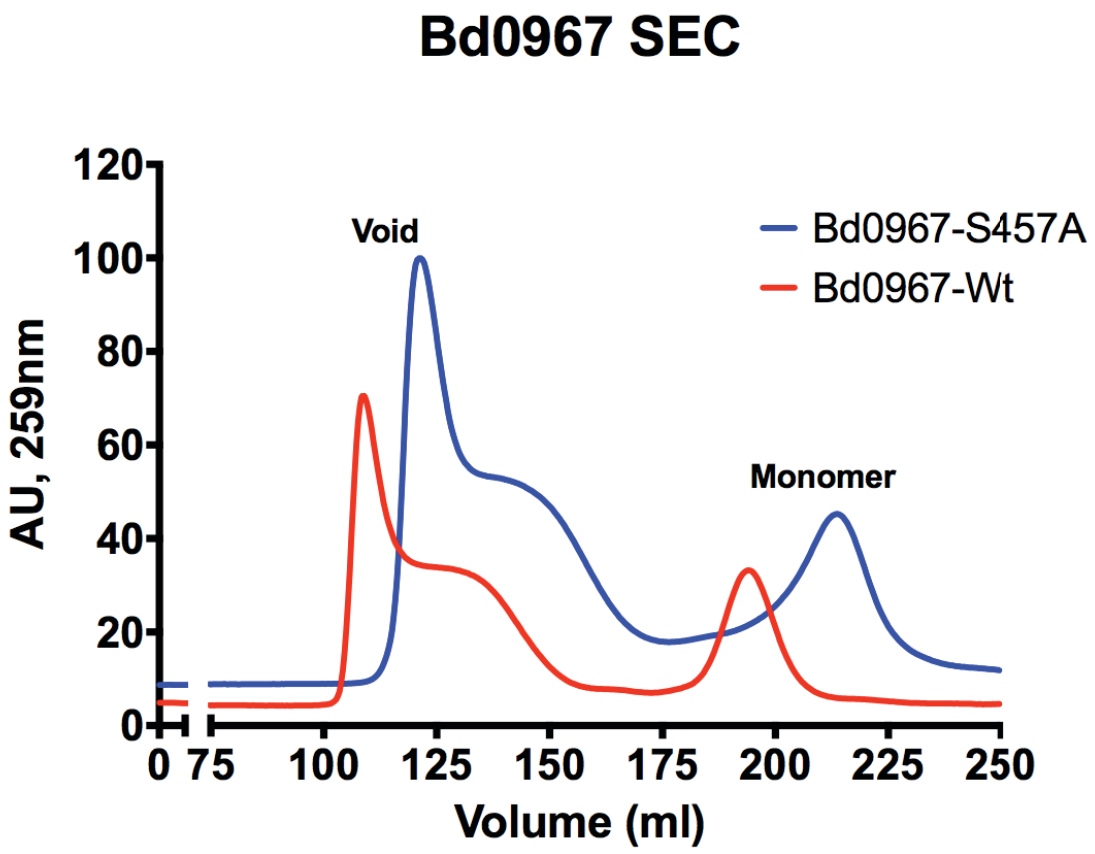


Figure 113. SEC results for Bd0967
The size exclusion traces for Bd0967 WT and S457A are shown. Monomeric Bd0967 elutes at approximately 190 ml. A large void peak is observed for Bd0967.

6.2.9 Structure of Bd0967 S457A

The structure of Bd0967-S457A is highly similar to the structure of Bd0967-WT, adopting the same fold and domain arrangement. Little conformational change is observed between WT and S457A structures with the exception of a slight rigid body rotational movement of the PDZ domain (6°). Although this slight conformational change is not seen for all chains (two out of four), suggesting conformational sampling of the PDZ domain. Furthermore, bound peptides are located at the PDZ peptide binding site and active site of Bd0967-S457A demonstrating substrates are still able to bind even though catalysis cannot be performed.

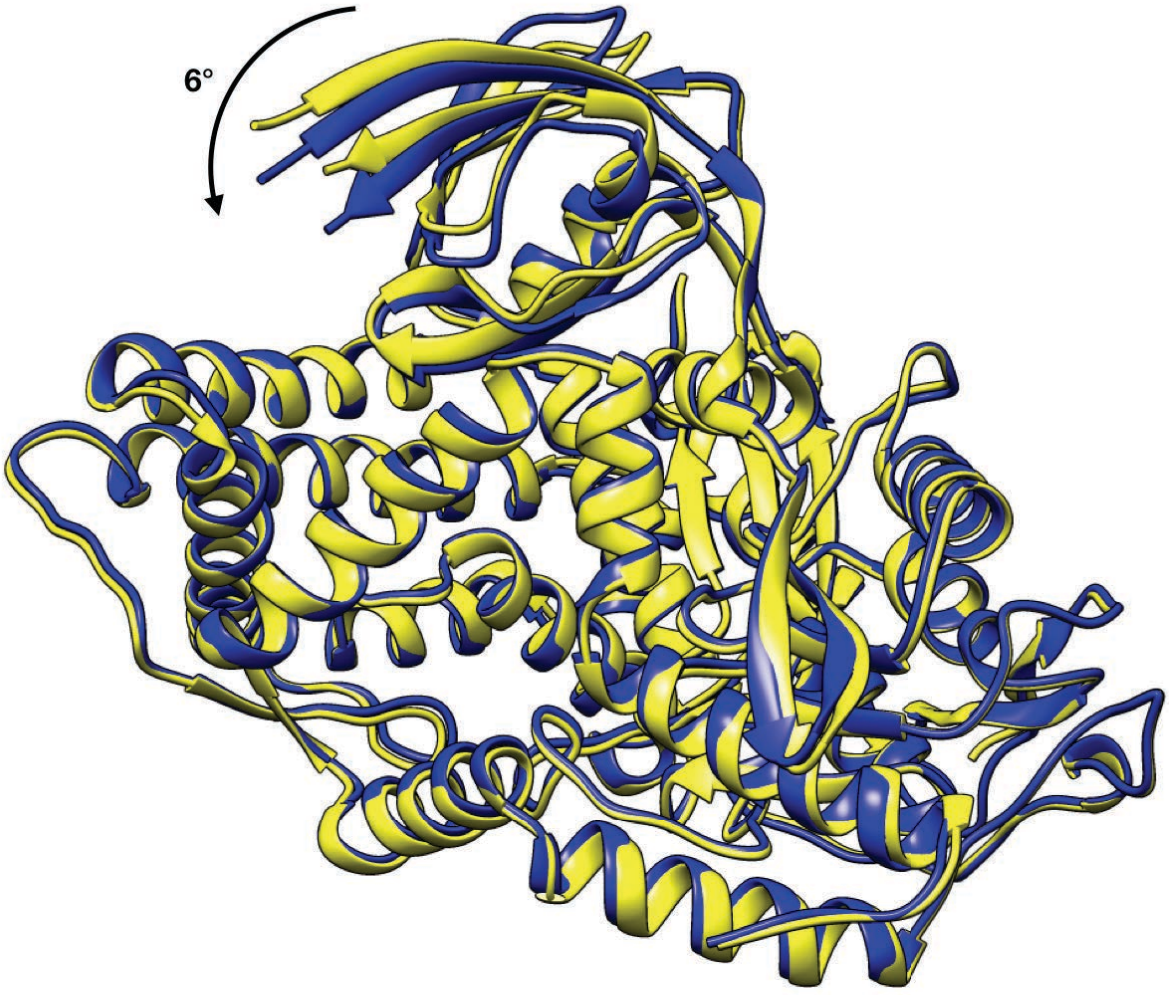


Figure 114. Structure of Bd0967-S457A
Bd0967-S457A (yellow) superimposed onto the structure of Bd0967-WT (blue) demonstrates the slight conformational change observed for the PDZ domain of Bd0967-S457A. Superimposition was performed using the SSM superpose tool in coot (length = 634 residues, 1.13Å).

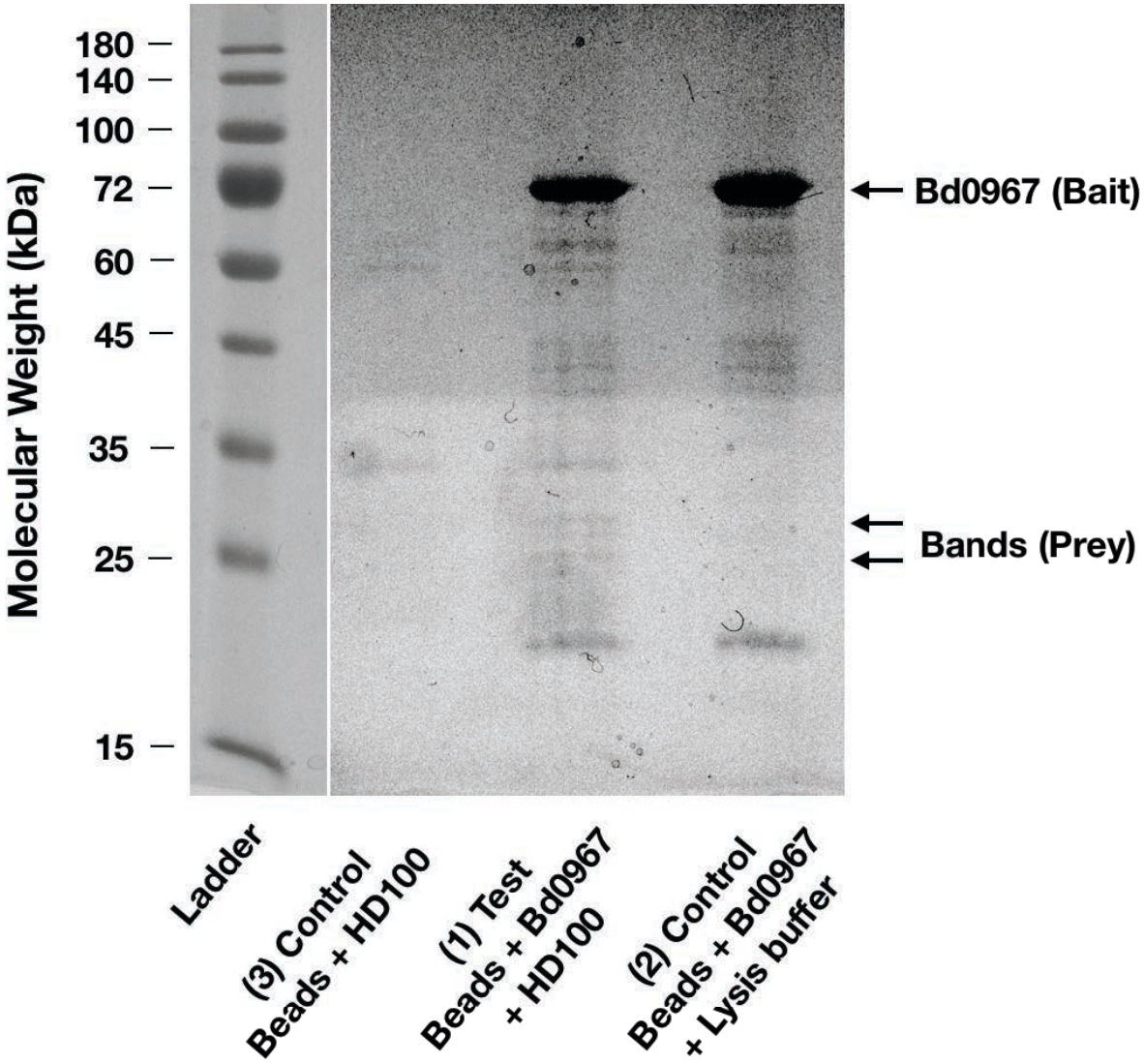


Figure 115. Bd0967 pull-down assay
The SDS-PAGE of the pull-down assay with Bd0967 as the bait, shown with the upper black arrow. The two faint bands in the test sample (1) are highlighted by the two lower black arrows.

Identified in Samples:	Organism	Protein Name	Uniprot accession code	Gene Name	Protein Size	Normalised total spectra			Terminal residues
						(1)	(2)	(3)	
(1)	*Bdellovibrio HD100*	Flagellin	Q6MQ75_BDEBA	Bd0606	29 kDa	33	0	0	ALRLIG
	Bdellovibrio HD100	Serine protein kinase	Q6MR70_BDEBA	Bd0228	80 kDa	9	0	0	GDVKNK
	Bdellovibrio HD100	Methyltransferase	Q6MP37_BDEBA	Bd1029	31 kDa	4	0	0	QARART
	Bdellovibrio HD100	ATP synthase subunit beta	ATPB_BDEBA	Bd3897	51 kDa	4	0	0	AKKLQA
	Bdellovibrio HD100	Probable signale peptide	Q6MK10_BDEBA	Bd2606	22 kDa	4	0	0	LIKFIP
(1) + (2)	*E. coli Strain K12*	cAMP-activated global transcriptional regulator CRP	CRP_ECOLI	crp	24 kDa	33	37	0	VVYGTR
	E. coli Strain K13	ATP-dependent RNA helicase RhlE	RHLE_ECOLI	rhlE	50 kDa	12	21	0	KPAAAQ
	E. coli Strain K14	3-ketoacyl-CoA thiolase	FADA_ECOLI	fadA	41 kDa	7	9	0	TVFERV
	E. coli Strain K15	Succinylornithine transaminase	ASTC_ECOLI	astC	44 kDa	7	9	0	VSRGSS
	E. coli Strain K16	tRNA (guanosine(18)-2'-O)-methyltransferase	TRMH_ECOLI	trmH	25 kDa	6	9	0	TMQAAG
(1) + (3)	*Bdellovibrio HD100*	Inosine-5'-monophosphate dehydrogenase	Q6MLD1_BDEBA	Bd2081	52 kDa	33	0	37	NYRIES
	Bdellovibrio HD100	Octaprenyl diphosphate synthase	Q6MQB9_BDEBA	Bd0557	37 kDa	26	0	39	WRRKPS
	Bdellovibrio HD100	Flagellin	Q6MIU6_BDEBA	Bd0352	29 kDa	37	0	5	ALKLIG
	Bdellovibrio HD100	Thioredoxin reductase	Q6MQT7_BDEBA	Bd0373	34 kDa	20	0	23	EAQEAH
	Bdellovibrio HD100	Flagellin	Q6MQ77_BDEBA	Bd0604	29 kDa	20	0	4	VAMKLV
(1) + (2) + (3)	*Bdellovibrio HD100*	Tail-specific protease	Q6MP94_BDEBA	Bd0967	75 kDa	315	376	14	KQANAK

Table 16. Mass Spectrometry results for Bd0967 pull-down assay
The top five proteins identified for each category is displayed in this table. Significantly increased normalised spectra (for proteins identified from sample (1) = beads + Bd0967 + Lysate is highlighted in green. Yellow indicates the low normalised spectra for proteins in control sample that had increased normalised spectra in sample (1).

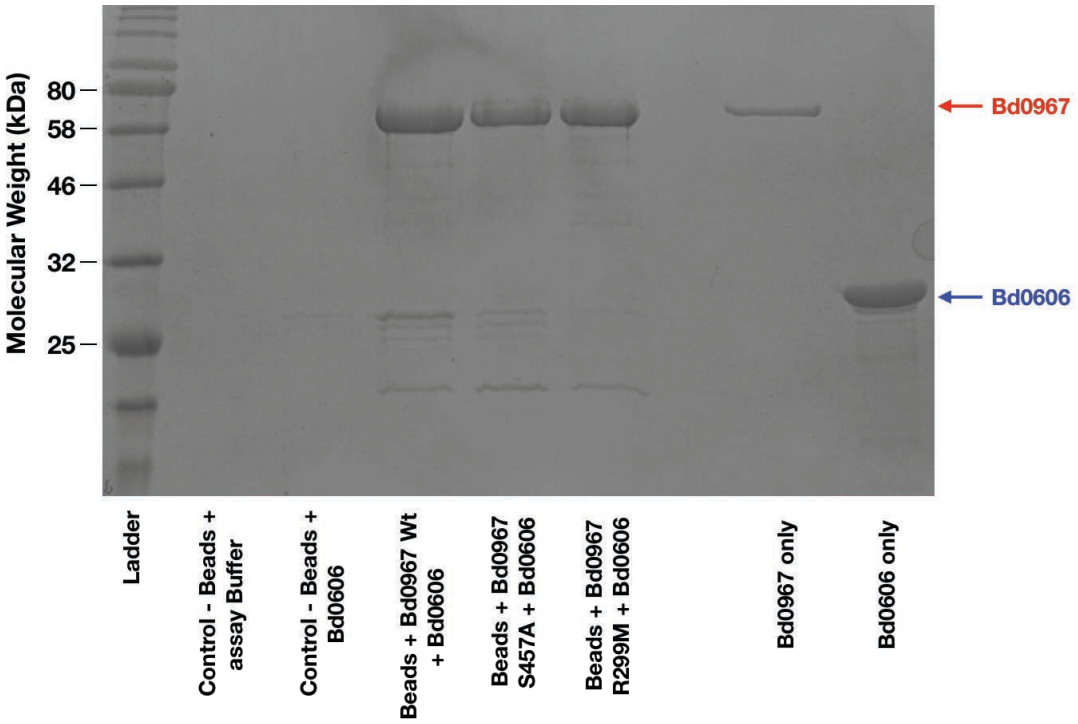


Figure 116. Confirm binding
A similar pull-down assay was re-performed with purified Bd0606 to evaluate the MS result. Bd0606 binds to Bd0967-Wt and to a lesser extent Bd0967-S547A. Bd0606 does not bind to the PDZ binding mutant Bd0967-R299M. The binding is weak (not 1:1 stoichiometry) under the assay conditions.

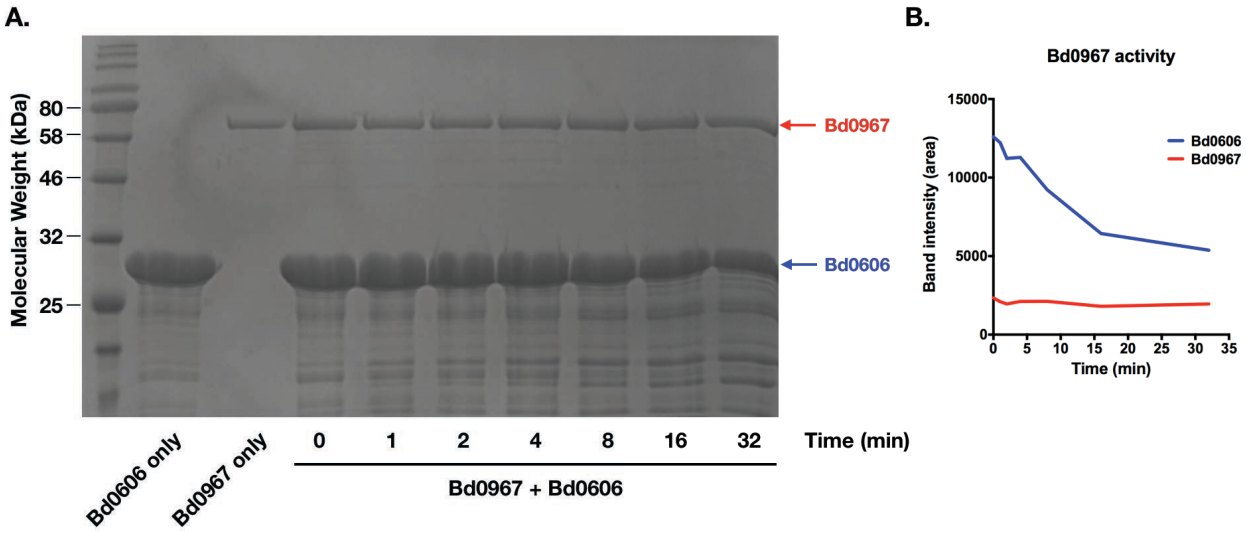


Figure 117. Bd0967 degrades Bd0606
A. Bd0967-Wt degrades Bd0606 in a processive manner, which was monitored over a time period. The red arrow highlights Bd0967 and the blue arrow highlights Bd0606. **B.** Graphical illustration of Bd0606 being degraded over time (band intensity) alongside Bd0967.

A similar result was observed when Bd0606 was incubated with the catalytic mutant Bd0967-S457A. Bd0606 remained at a similar level throughout the 1 hour incubation when incubated with Bd0967-S457A mutant suggesting that degradation is being performed by Bd0967 in the Wt sample. Furthermore, a Bd0967-R299M (PDZ peptide binding mutant) was also incapable of degrading Bd0606. This finding supports the importance of the PDZ domain interaction with Bd0606 (substrates) to facilitate catalysis.

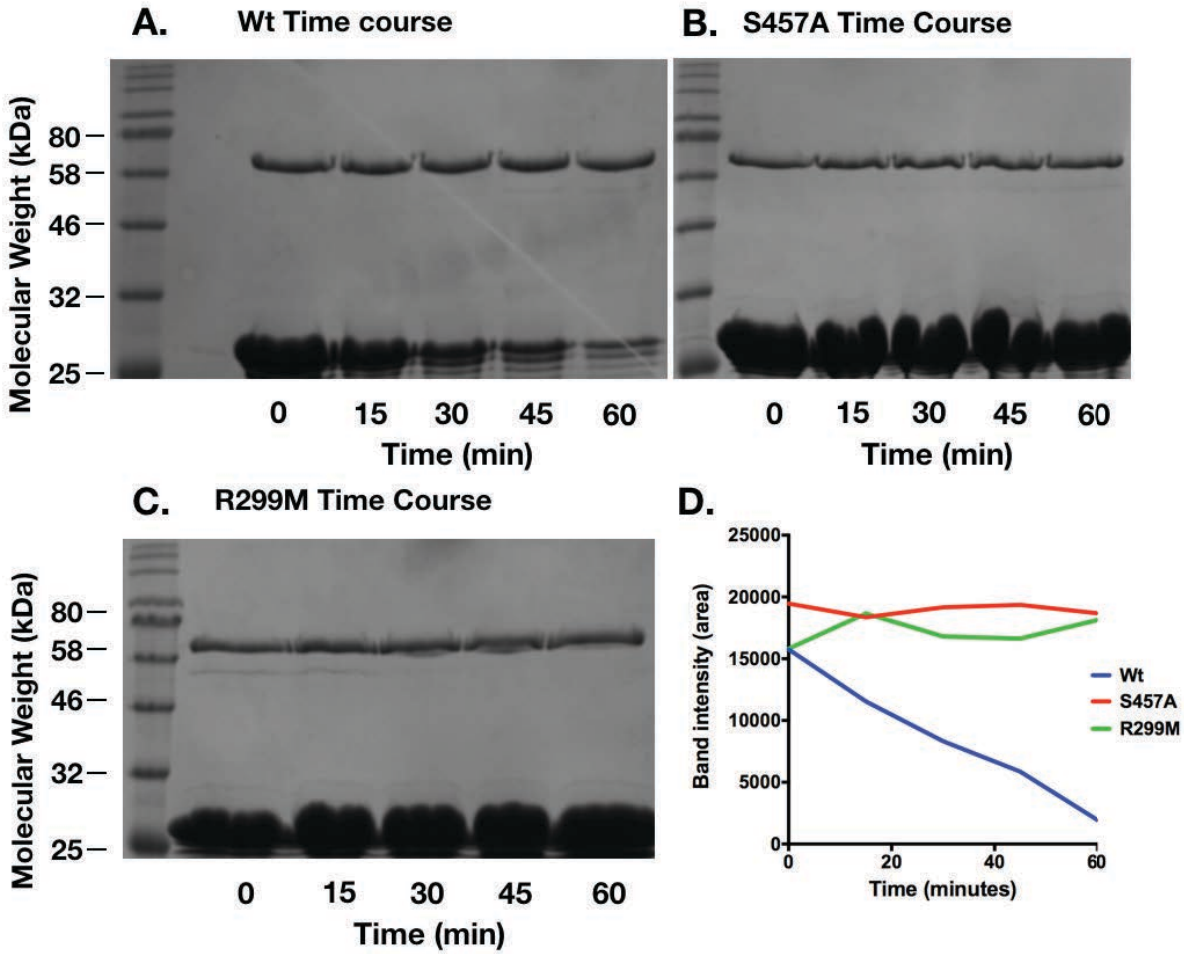


Figure 118. Bd0967 activity in comparison to mutations
A. Bd0967 Wt degradative activity towards Bd0606 in a 60 minute time period with samples taken at regular time intervals. **B.** S457A mutant does not show any activity. **C.** R299M also lacks the ability to degrade Bd0606 efficiently due to a non-functional PDZ domain. **D.** A graphical representation showing the degradation of Bd0606 when incubated with Bd0967-Wt, S457A and R299M mutants. Image J was used to analyse the intensity of gel bands.

To determine whether Bd0967 was specifically targeting Bd0606 for degradation we also investigated the activity of Bd0967 towards a variety of other *Bdellovibrio* proteins of varying size and C-terminal sequences (Figure 119). This assay did not reveal any evidence that Bd0967 was active towards any of these other proteins. Furthermore Bd0108 (a small intrinsically disordered protein) was also not a substrate for Bd0967 suggesting that a disordered C-terminal sequence is not the only prerequisite to binding and catalysis. Additionally, Bd0967 showed no degradative activity towards Bd1996 and Bd0314, which are both similar molecular weights to Bd0606.

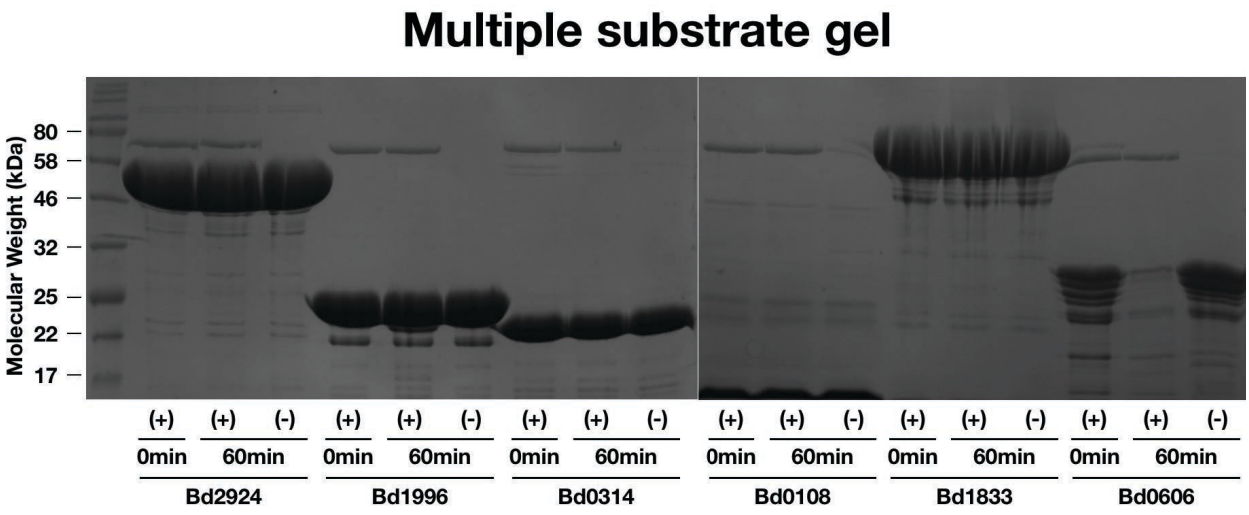


Figure 119. Bd0967 activity towards various proteins
Bd0967 degradative activity has been analysed against *Bdellovibrio HD100* proteins of varying sizes and C-terminal sequences. Bd0967 showed no activity towards any of the proteins trialled in this assay. Bd0108 is a small intrinsically disordered protein that remained un-digested suggesting the folded state of the proteins is not a factor determining Bd0967 degradative ability.

To investigate the mechanism behind Bd0606 recognition and degradation by Bd0967, we produced C-terminal truncation mutants of Bd0606. These mutants, ΔLIG and ΔRLIG, lack the terminal three and four amino acids, which typically interact with the PDZ domain of CTPs. To our surprise Bd0967 was able to degrade these truncation mutants of Bd0606 with similar efficiency to WT Bd0606 (Figure 120). This finding

along with the Bd0967-R299M mutant finding suggests that although a functioning PDZ domain is absolutely required for recognition of the C-terminal sequence and initiating catalysis; Bd0967 also has a degree of promiscuity and tolerance in peptide binding at the PDZ domain. On the other hand, Bd0967 may also adopt an additional mechanism of substrate recognition, which could also explain these results.

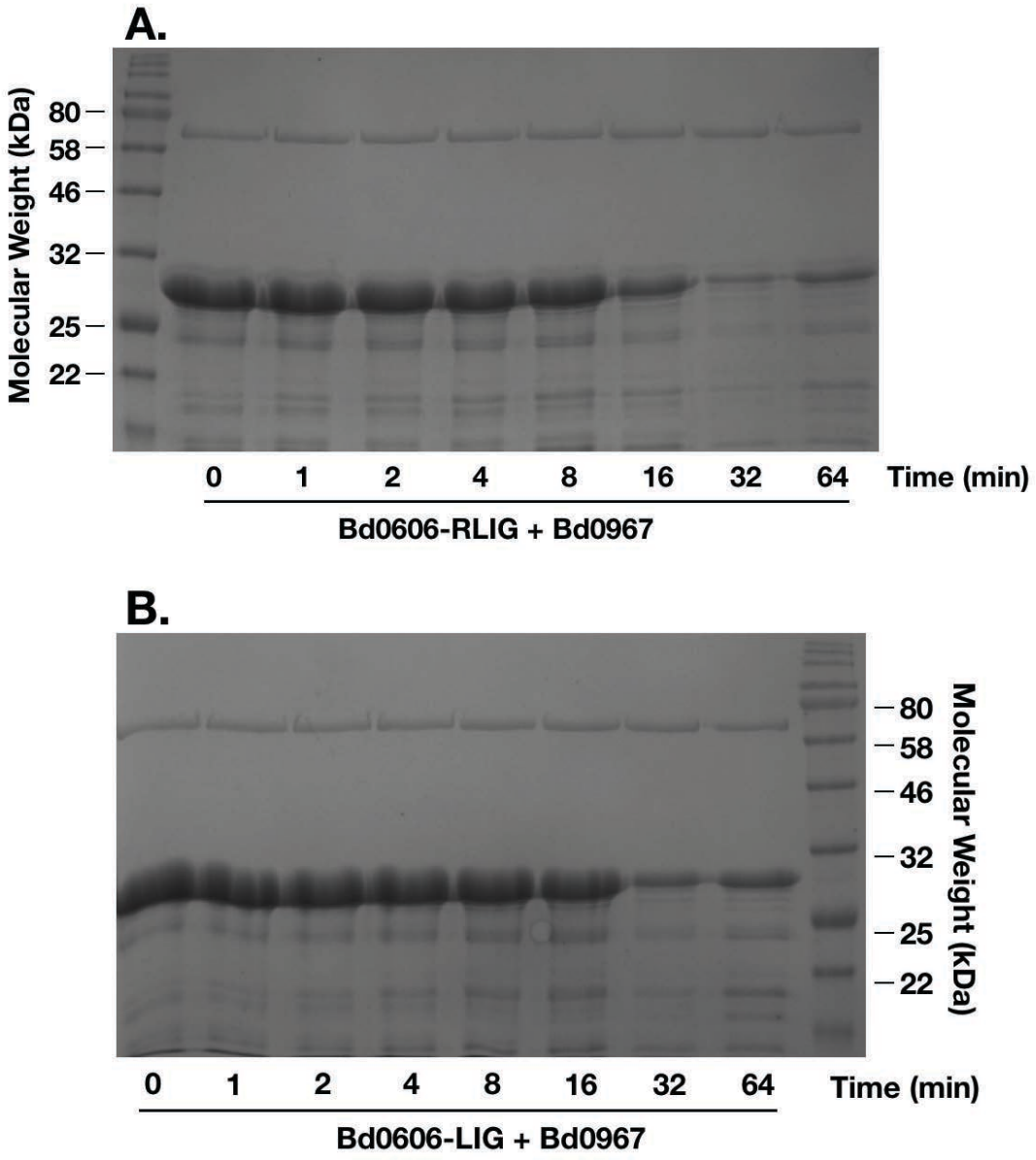


Figure 120. Bd0606 C-terminal mutants
A. Bd0606 mutant RLIG lacks the C-terminal four amino acids, which bind to the PDZ domain. **B.** Bd0606 mutant LIG lack the three C-terminal amino acids. Bd0967-Wt is able to degraded both RLIG and LIG mutants to a similar degree as WT Bd0606 substrate. This finding suggests Bd0967 PDZ binding site allows promiscuity and also suggest Bd0606 may be recognised via an alternative mechanism distinct from direct interaction with the C-terminal.

6.2.12 The C-terminal peptide of Bd0606 as an inhibitor of Bd0967 activity

We designed and created a 12aa peptide matching the C-terminal sequence of Bd0606 (NAMPNSALRLIG) in order to investigate whether C-terminal binding was required for Bd0967 degradative activity. This 12aa peptide of Bd0606 was kindly prepared by Oliver Daubney by N-terminal peptide synthesis to ensure that the α-carboxylate was present. Bd0606 was incubated with Bd0967-Wt in buffer containing increasing concentration of peptide. The results of this assay show that increasing the concentrations of peptide reduces the degradation of Bd0606 (Figure 121). At low concentrations of peptide Bd0967 remains active and the levels of Bd0606 are reduced, whereas at a high peptide concentration the degradative activity of Bd0967 is reduced. This result demonstrate that the peptide fragment can act as an inhibitor of Bd0967, presumably by interacting with and saturating binding of the PDZ domain, which disables the activity of Bd0967.

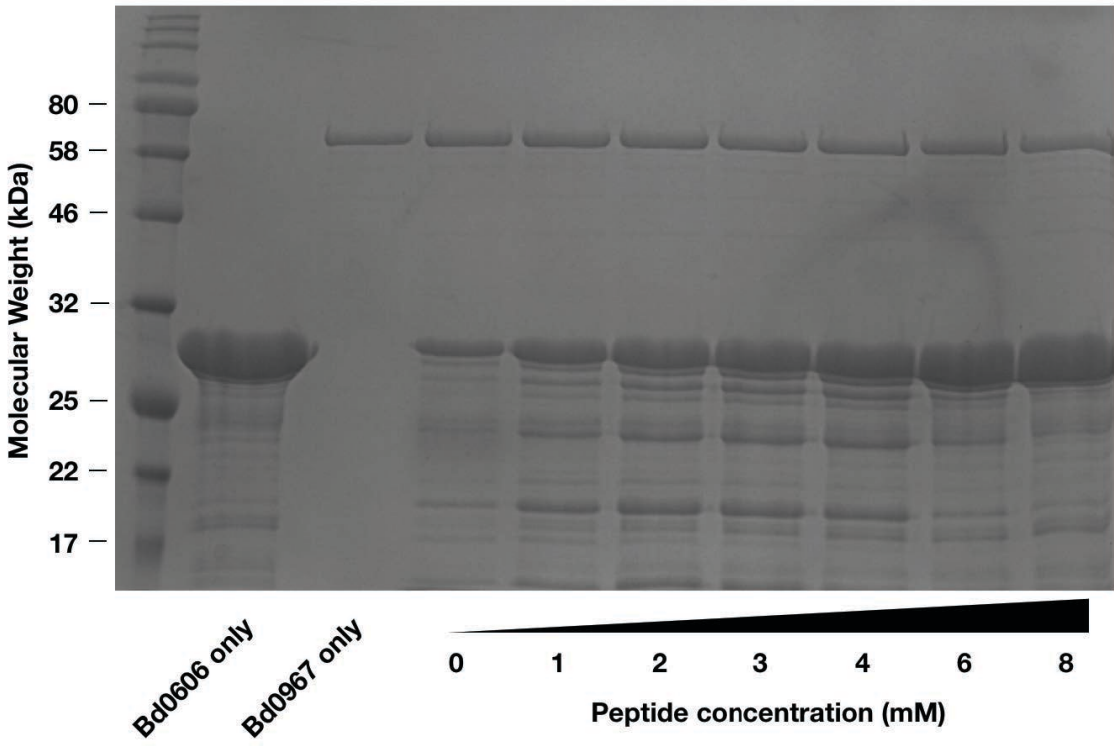


Figure 121. Peptide inhibition assay
Bd0967 activity was measured in the presence of an increasing concentration of a 12aa peptide fragment, synthesised to match the C-terminus of Bd0606. At high peptide concentrations Bd0967 degradative activity towards Bd0606 was reduced indicating the PDZ domain is important for substrate recognition.

6.2.13 Bd0606 is folded

Circular Dichroism (CD) was performed on Bd0606 to determine whether it was correctly folded. CD is a good biophysical technique for determining the proportion and nature of a folded protein based on a distinct UV spectrum for α-helix and β-sheet. The CD spectrum revealed that Bd0606 is folded, with α-helices being the major secondary structure element, which is expected for a flagellin protein based on previous structural information [295]. This finding supports Bd0606 being correctly folded and Bd0967 is not degrading unfolded proteins in a quality control degradative manner.

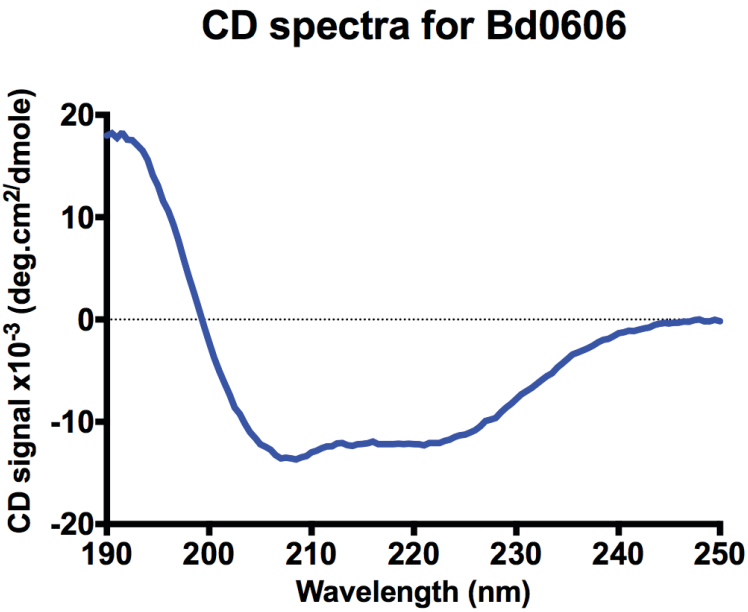


Figure 122. Bd0606 CD spectra
CD spectra for Bd0606 reveals Bd0606 is folded and is formed predominantly of α-helices. Minima peaks at ~208 nm and 222 nm and a maxima peak at 193 nm exemplify a α-helix folded proteins.

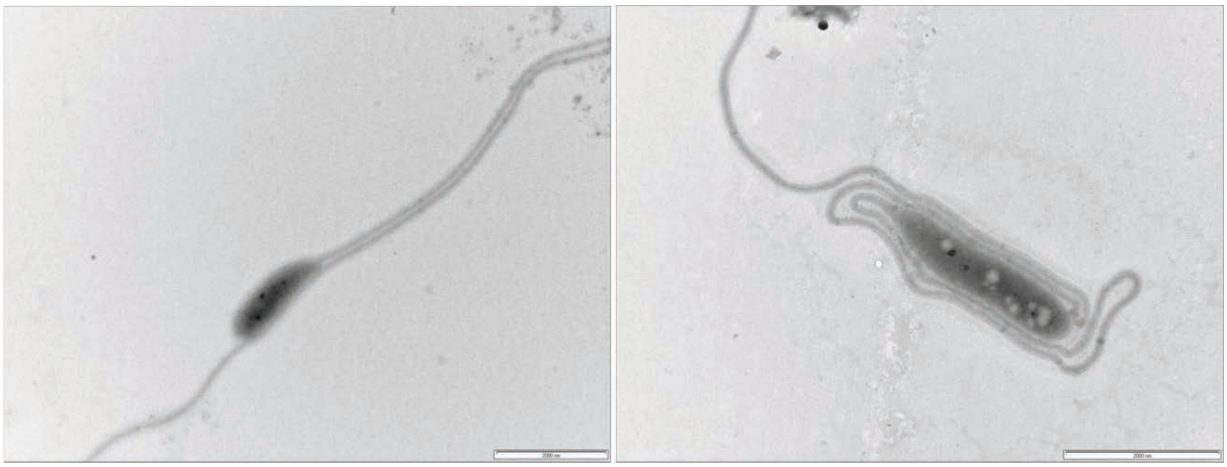


Figure 123. Bd0967 KO mutant *Bdellovibrio*
Bd0967 KO mutants have a pleiotropic developmental phenotype that involves the flagellum assembly. The flagellum of Bd0967 KO *Bdellovibrios* have been seen to grow to ~5 times the length of Wt, suggesting flagellin assembly is unregulated. These two images show the large mutant flagellar. **(Electron Micrograph images kindly provided by the Sockett lab)**

6.2.15 Bd0967 interaction with CdG

MST was used to determine whether Bd0967 interacted with CdG. A K_D value of 22 μM was obtained, suggesting weak non-specific binding. Furthermore, Bd0967 has a signal-peptide, suggesting transport into the periplasm where CdG is unlikely to be present.

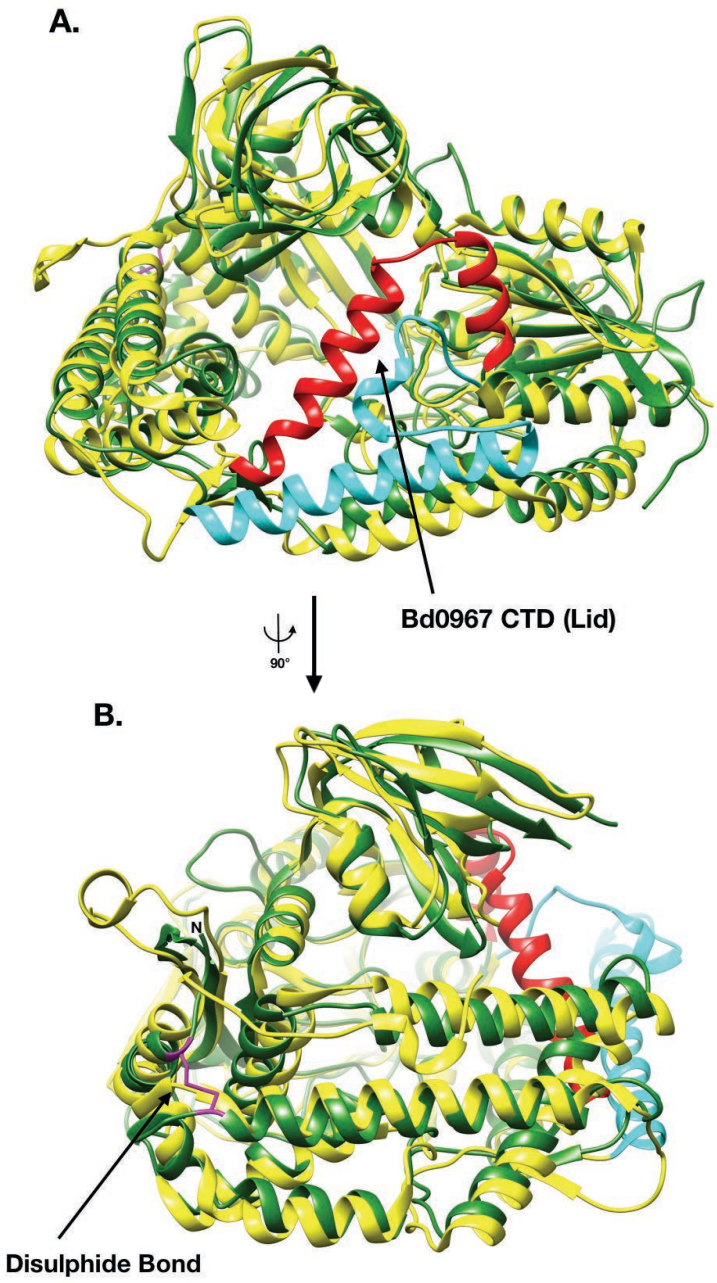
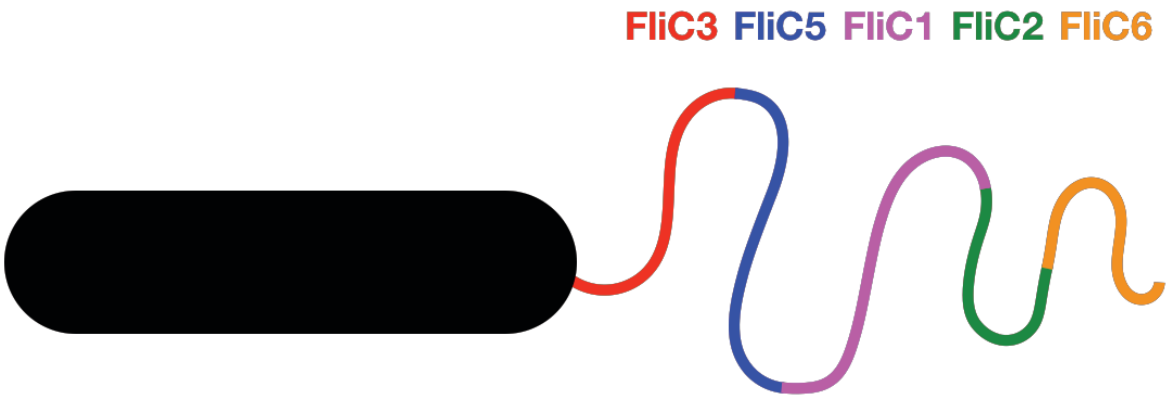


Figure 124. Bd0967 comparison to *E. coli* Prc
A. The structure of Bd0967 (green) is superposed with *E. coli* Prc (yellow, PDB:5WQL). Superimposition was performed using the SSM superpose tool in coot (length = 536 residues, rmsd = 2.01Å). The proteins three-dimensional structures are highly similar and overlap well in all domain except from the CTD. The CTD of Bd0967 is shown in red, while the CTD of Prc is shown in Cyan. The CTD of Bd0967 forms more of a paddle-like structure creating a cavity shielded from bulk solvent, whereas Prc CTD forms one wall of a bowl shape. **B.** Bd0967 also has a disulphide bond (magenta) located in the NTD, which is not found in other C-terminal processing proteases.

hydrophobic C-terminal sequence making it a suitable substrate for Bd0967. Moreover other *Bdellovibrio* flagellins have highly similar C-terminal sequences (Figure 125), suggesting they may also be potential substrates, which requires experimental verification.



Flagellin	Size (aa)	C-terminal sequence
FliC1 (Bd0604)	274	ANQLPQVAMKLV
FliC2 (Bd0606)	277	NAMPNSALRLIG
FliC3 (Bd0408)	282	VALGLLNKSFQG
FliC4 (Bd0410)	277	NQSQQLALKLLG
FliC5 (Bd3052)	277	NQSNQLALKLIG
FliC6 (Bd3342)	277	NQANQLALKLIG

Figure 125. Flagellin assembly *Bdellovibrio*
Bdellovibrios flagellum is composed of six flagellins in an ordered arrangement from the proximal end to the distal end, producing a tapered wave formation. The location of each flagellin species, within the flagellum, is illustrated above in the cartoon representation. Only FliC4 location has not been determined. The table highlights the similarities of flagellin C-terminal sequences.

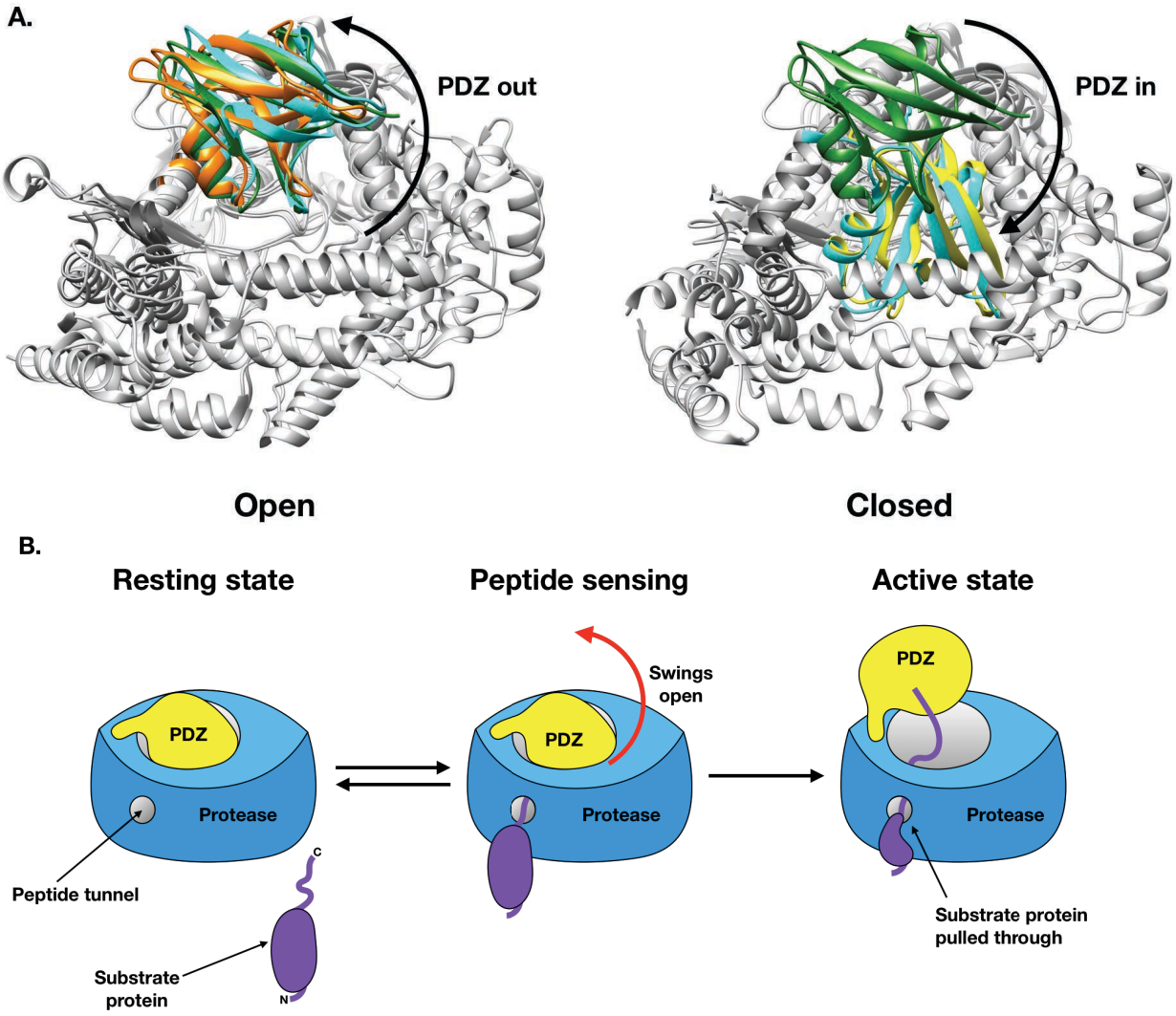


Figure 126. Bd0967 mechanism
A. The movement of the PDZ domains in C-terminal processing proteases is associated with regulating protease activity. Bd0967 PDZ domain (green) is in an outward conformation and its location is compared, relative to other PDZ domains of C-terminal processing proteases. Bd0967 PDZ also aligns well with CtpB-active (cyan, PDB:4C2C, length = 302 residues, rmsd = 2.37Å) and Prc (orange, PDB:5WQL, length = 536 residues, rmsd = 2.01Å) PDZ domains in the active "out" position. The PDZ domains of CtpB-resting (cyan, PDB:4C2E, length = 222 residues, rmsd = 3.40Å) and D1P (yellow, PDB:1FC7, length = 234 residues, rmsd = 3.03Å) have been shown to adopt an inward position (resting state), fitting close to the protease peptide binding tunnel. Superimposition was performed using the SSM superpose tool in coot. **B.** A cartoon illustration of the proposed mechanism of Bd0967. Bd0967 exists in the resting state with the PDZ domain tucked into the bowl of the protein. When the PDZ domain binds / recognises the C-terminal peptide sequence of its substrate the PDZ domain undergoes a conformational change and swings into an out position, activating the protease domain. The PDZ domain movement pulls the substrate through the peptide tunnel past the catalytic site residues for degradation (unfolding in the process).

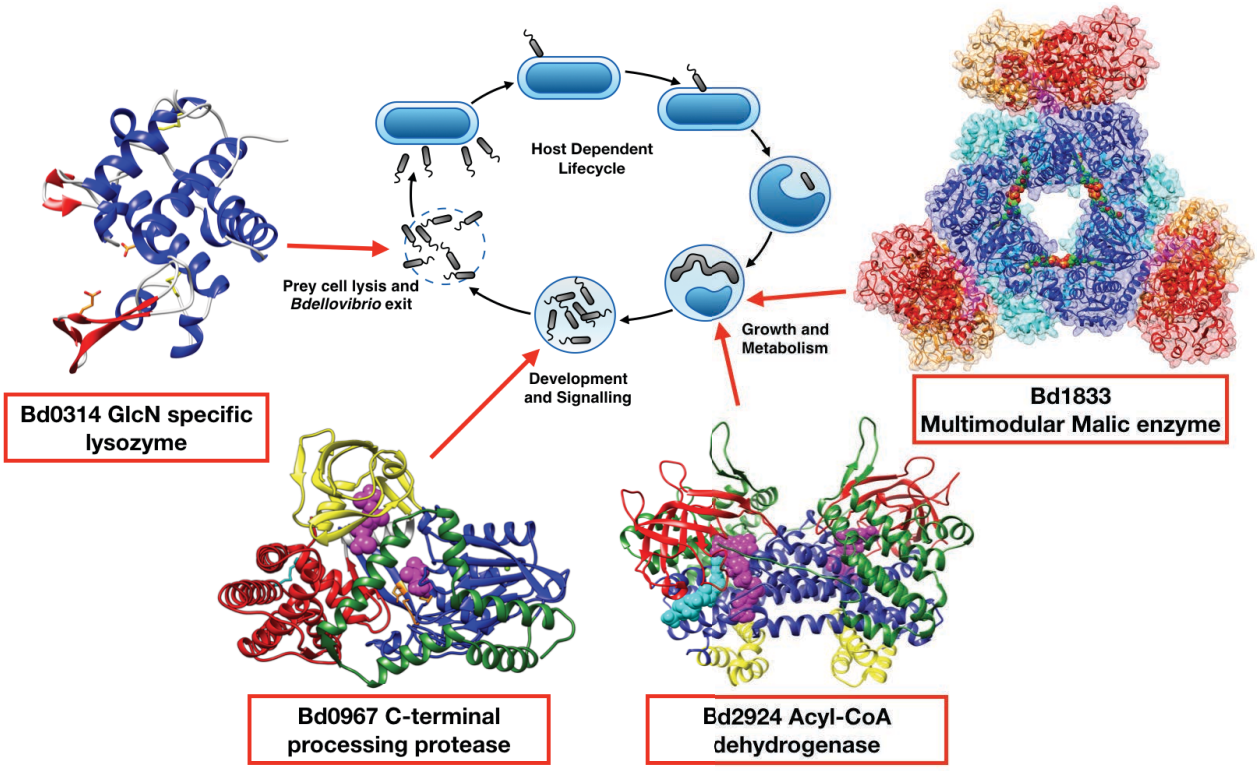


Figure 127. Summary of thesis findings
This thesis has structurally and functionally investigated four proteins with diverse roles within the predatory lifecycle. Bd1833 is a multi-modular malic enzyme involved in an alternative metabolic pathway, diverting TCA cycle intermediates and switching to a carbon storage pathway that is allosterically inhibited by acetyl-CoA. Bd2924 is a divergent putative acyl-CoA dehydrogenase potentially involved in metabolism of prey cell lipids. Bd0967 is a C-terminal processing protease involved in a critical cell developmental pathway involving flagellin degradation. Bd0314 is a modified lysozyme that has been specifically adapted to exit the invaded prey cell, acting specifically towards N-deacetylated peptidoglycan cell walls.

Two of the proteins investigated in this study (Bd2924 and Bd1833) are enzymes that have roles in central metabolism. Central metabolism has been an understudied area of research in the predatory lifecycle of *Bdellovibrio*. Many of the metabolic pathways have been adapted to deal with the needs of a predator's diet, such as solely using the mevalonate pathway for isoprenoid biosynthesis [306]. Consequently the findings presented in this thesis represent the first structurally and functionally characterised central metabolic enzymes from *Bdellovibrio*. We reveal Bd2924s three-dimensional structure is related to ACADs with divergent functions, suggesting that Bd2924 has a

Abstract

This thesis documents the findings undertaken on a diverse set of proteins that operate at distinct stages of the predatory lifecycle of *Bdellovibrio bacteriovorus*. In general, this work explores three major predatory lifecycle events: growth & metabolism, development & signalling and finally bdelloplast lysis & exit. In particular this thesis characterises four individual proteins (Bd2924, Bd1833, Bd0967 and Bd0314) that perform very different roles.

In this thesis the three-dimensional structure of Bd2924 (a putative acyl-CoA dehydrogenase) has been solved demonstrating an adaptation of the common acyl-CoA dehydrogenase fold, suggesting an important role beyond fatty acid β-oxidation. This study also presents the first structure of a multi-modular malic enzyme (Bd1833) and through extensive mutational and biochemical investigations, have unpicked the mechanism behind allosteric regulation by acetyl-CoA. In addition the three-dimensional structure of Bd0967 (a C-terminal processing protease) has been solved, revealing a divergent fold that forms a large self-compartmentalised cavity. Furthermore, Bd0967 is demonstrated to be involved in a developmental pathway that is interconnected with flagellin assembly. Bd0967 targets and processively degrades the flagellin Bd0606, which is mediated by interactions with the PDZ domain. Finally this study determines how *Bdellovibrio* exits the consumed prey at the end of the predatory lifecycle that had been chemically marked for destruction. This role is provided by a repurposed lysozyme (Bd0314) remodeled into a variant that now only accepts deacetylated cell wall sugars. Enzyme assays and addition to bdelloplasts confirms the specificity for modified prey cell wall material.

Acknowledgments

First of all I would like to thank my PhD supervisor, Andrew Lovering. You have been a great support throughout my PhD, always being there to discuss ideas and offer advise. Your enthusiasm and knowledge for science has been infectious and constantly encouraged me to grow as a research scientist. I could not have asked for a better supervisor.

I would also like to thank Patrick Moynihan, for all of your help and guidance with all things biochem and cell wall. You have been a massive help and it will always be greatly appreciated.

I greatly appreciate all the collaborative work and contributions people have made towards my PhD studies, including Oliver Daubney and Charles Moore-Kelly. I would especially like to thank Prof Liz Sockett and all those in the Sockett lab for all of your contributions and hard work, as well as being very hospitable hosts on my visits to Nottingham.

I would like to specially thank Ian Cadby, you have always been there to offer a helping hand and advice; it has been a pleasure to work and learn from you. To the rest of the Lovering lab, thank you for the support over the years, especially Richard Meek for always lightening the mood with your quirky personality. I would also like to thank all those on the 7th floor, it has been a pleasure working with you all. Additionally, thank you Tim Knowles for being my secondary supervisor and a great guy.

Personally, a big thank you to all my family for always supporting me. Mum and Dad, I would not be where I am without your constant support and encouragement. Last, but certainly not least, I would like to thank my wonderful wife, Sarah. I could always rely on your love and support, especially when things got tough.

Table of Contents

Chapter 1 ..... **1**

Introduction **1**

1.1 Predatory Bacteria **1**

1.1.1 Overview 1

1.1.2 Types 2

1.1.3 Beneficial Applications 5

1.2 *Bdellovibrio bacteriovorus* **6**

1.2.1 Overview 6

1.2.2 Prey Detection and Motility 7

1.2.3 Attachment and Invasion 9

1.2.4 Niche formation 10

1.2.5 Metabolism 11

1.2.6 Growth and Division 14

1.2.7 Maturation and Lysis 14

1.2.8 Cellular Signalling and Regulation of Predation 15

1.3 Overall Research Aim Of This Study **18**

Chapter 2 ..... **20**

Materials and Methods **20**

2.1 General **20**

2.1.1 Molecular cloning 20

2.1.2 Colony PCR 22

2.1.3 Mutagenesis 22

2.1.4 Q5 SDM 23

2.1.5 Primers 24

2.1.6 Sequencing 26

2.1.7 Transformation 26

2.1.8 Culture Media 26

2.1.10 Glycerol Stocks 27

2.1.11 Expression 27

2.1.12 Purification 28

2.1.13 Dialysis 29

2.1.14 Size Exclusion Chromatography 29

2.1.15 SDS-PAGE 29

2.1.16 Protein sample concentration and quantification 30

2.1.17 Protein crystallisation 31

2.1.18 Data collection and structure solving 31

2.2 Bd0314 ..... **32**

2.2.1 Cloning, Expression and Purification 32

2.2.2 FITC Peptidoglycan digestion assay 33

2.2.3 Premature bdelloplast lysis assay 35

2.2.4 Crystallisation and Cryoprotection 35

2.2.5 Data collection and structure determination 36

2.3 Bd1413 and Bd1440 ..... **38**

2.3.1 Cloning, Expression and Purification 38

Table of Figures

Figure 1: Types of predatory bacteria 4
Figure 2: Host dependent predatory lifecycle 7
Figure 3: Proteins of interest 17
Figure 4: Bd1833 misannotation 42
Figure 5: HPLC assay standards 46
Figure 6: Predatory lifecycle, prey cell wall modifications 67
Figure 7: Peptidoglycan structure 68
Figure 8: Modifications of the peptidoglycan strand 72
Figure 9: Structures of lysozyme superfamily members 77
Figure 10: Catalytic distances 79
Figure 11: Retaining enzyme mechanism 80
Figure 12: Inverting enzyme mechanism 82
Figure 13: Lytic Transglycosylase reaction 83
Figure 14: Bd0314 purification and crystallisation 90
Figure 15: Bd0314 structure schematic 90
Figure 16: Bd0314 structure 92
Figure 17: Bd0314 catalytic mechanism 93
Figure 18: Bd0314 substrate-binding site 94
Figure 19: Substrate-binding surface view 96
Figure 20: Bd0314 conservation 98
Figure 21: Bd0314 alignment with Bdellovibrio HD100 homologues 99
Figure 22: Bd0314 alignment and residue conservation 100
Figure 23: Bd0314 specificity 102
Figure 24: Time and concentration dependent assay 103
Figure 25: Bd0314 active site residues 104
Figure 26: Bd0314 homologue activities 105
Figure 27: Bd0314 bdelloplast lysis 107
Figure 28: Bd0314 structure comparison 110
Figure 29: Bd0314 substrate binding 112
Figure 30: Comparison of Bd0314 with chitosanases 114
Figure 31: Bd0314 catalytic residue location 116
Figure 32: Role of Bd0314 119
Figure 33: β-oxidation metabolic pathway 124
Figure 34: Acyl-CoA dehydrogenases 127
Figure 35: MCAD structure 128
Figure 36: ACAD mechanism 131
Figure 37: Electron-transfer by ETF 132
Figure 38: Structure of ACO 133
Figure 39: Structure of divergent ACAD homologues 136
Figure 40: Bd2924 purification and crystallisation 137
Figure 41: Bd2924 SEC 138
Figure 42: Structure of Bd2924 139
Figure 43: Bd2924 tetramer complex 141
Figure 44: FAD hydrogen bonding network 142
Figure 45: C10-CoA binding 143

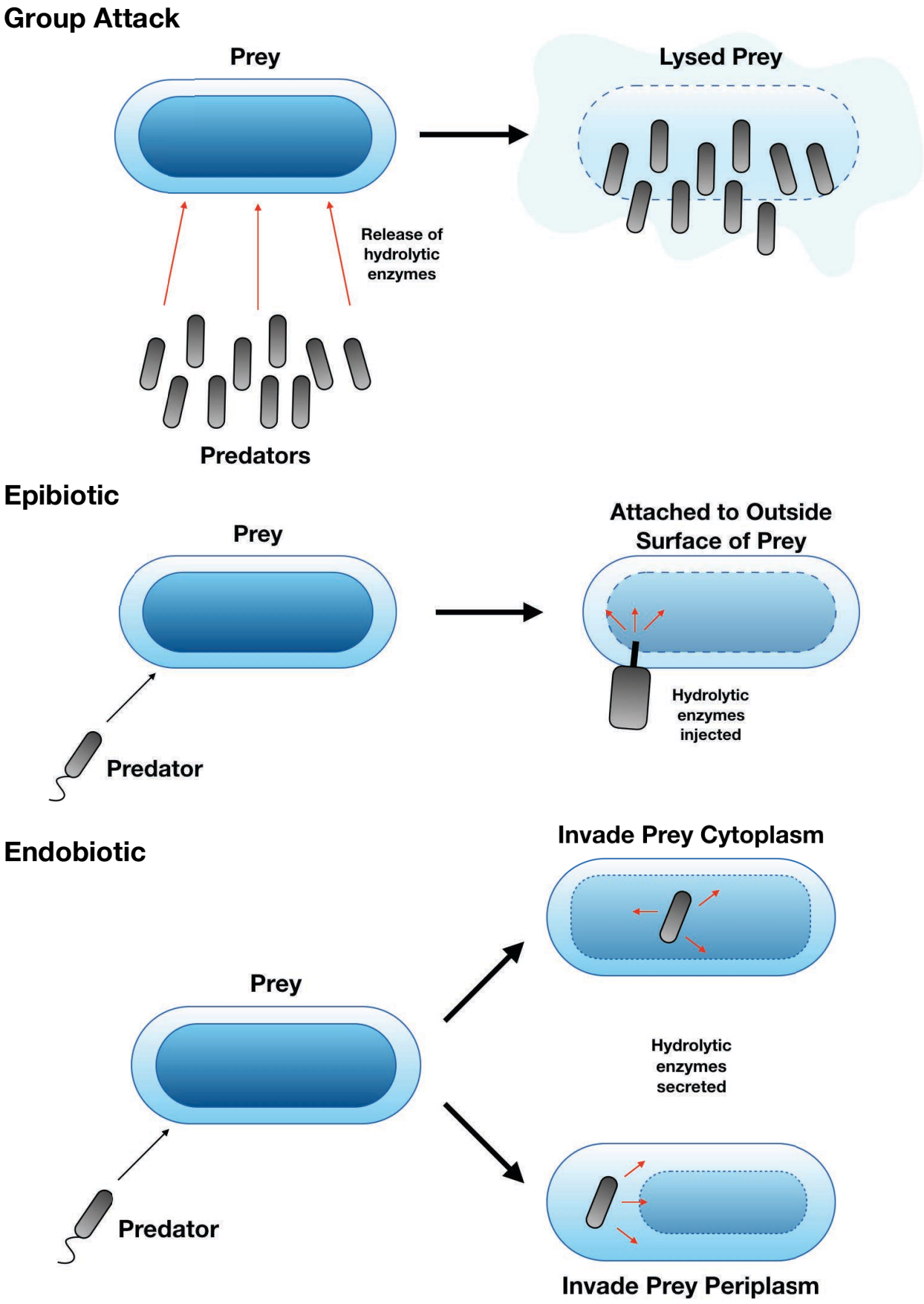


Figure 1 Types of predatory bacteria
Predatory bacteria can be categorised into three main groups based on how they kill their prey. These classes include: group attack, epibiotic and endobiotic predation. Predators are black ovals and the prey cells are blue ovals (illustration depicts Gram-negative cells). Red arrows show release of degradative hydrolytic arsenal is employed to kill the prey.

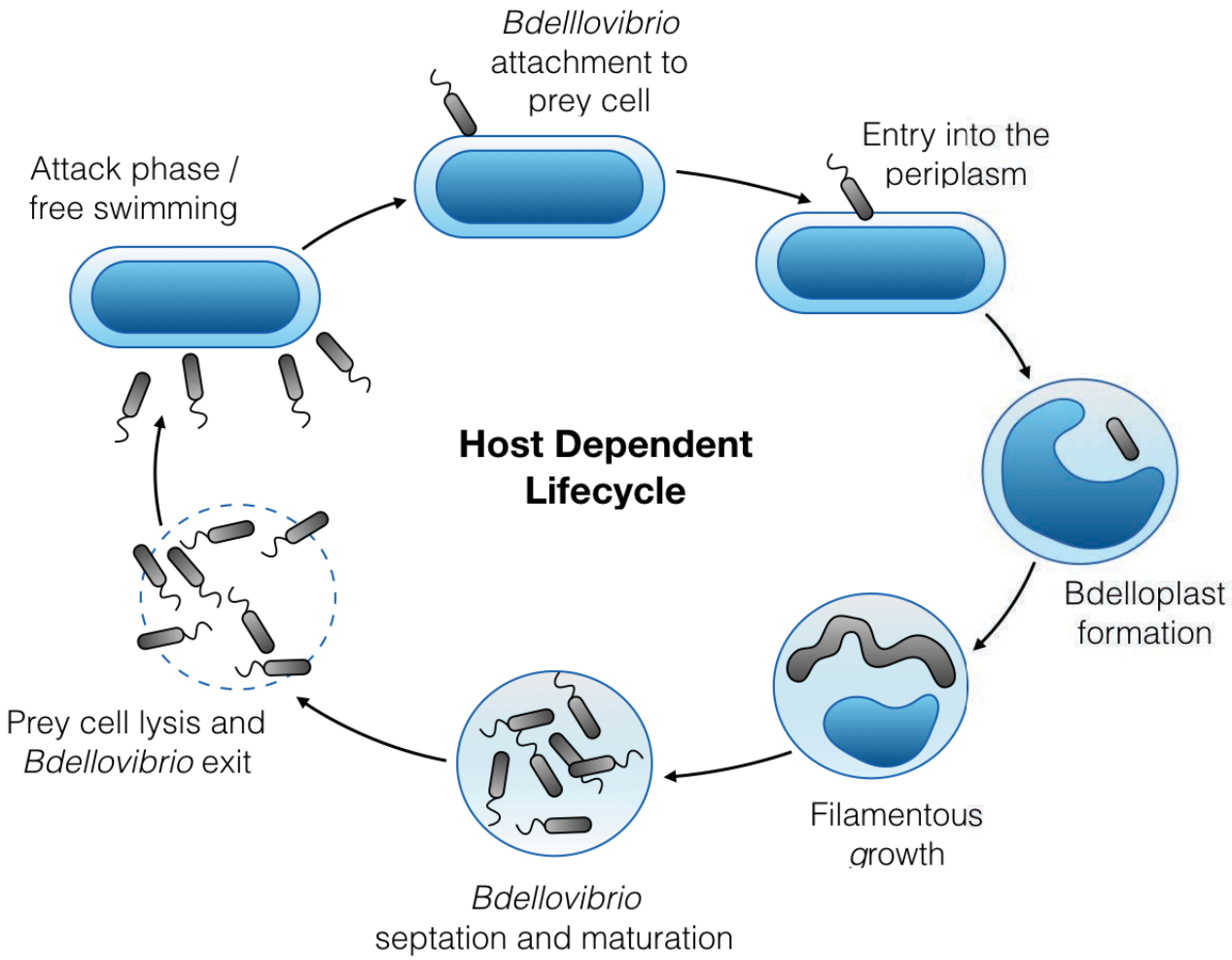


Figure 2 Host Dependent Predatory Lifecycle
The HD predatory lifecycle of *Bdellovibrio* can be divided into distinct stages. The cycle begins with free swimming and highly motile AP cells that actively seek out prey. Once located, *Bdellovibrio* attaches to the prey cell. Initially attachment is reversible (recognition period) but becomes permanent prior to invasion. Localised application of hydrolytic enzymes create a small pore in the prey cell envelope, which *Bdellovibrio* squeezes through, traversing the outer-membrane and cell wall to reside in the preys periplasm. During invasion the flagellum is shed and the entry pore is resealed. Additionally prey cell wall modifying enzymes are secreted, which remodel the prey cell wall creating an osmotically stable bdelloplast structure. Once the niche has been established *Bdellovibrio* begins the growth and replicative phase, growing filamentously. Once prey cell nutrients become depleted *Bdellovibrio* undergoes division and the progeny mature into AP cells. Mature AP progeny finally lyse the bdelloplast and begin the predatory cycle anew.

1.2.2 Prey Detection and Motility

Attack phase cells are small (0.2-0.5 µm x 0.5-2.5 µm), monoflagellate, rod shaped cells [11]. The attack phase cells are highly motile cells, swimming at speeds up to 160 µm sec^{-1} [26]. Motility is generated by a single, long, polar and sheathed flagellum that

Interestingly many of the novel CdG binders have putative roles in central metabolism (such as Bd2924), which suggests CdG may also regulate central metabolism. However, the reliance of co-factors (FAD, CoA, NAD) in this group of enzymes may have influenced non-specific binding. Another group of novel CdG binding proteins are components of the flagellum. CdG is known to negatively affect flagellum motility in other organisms [83]. In *Bdellovibrio* flagellar biosynthesis is coupled to CdG signalling, although no flagellar apparatus has been reported to interact directly with CdG. It is interesting to speculate that CdG interactions with flagellar components, such as flagellin (Bd0604) and hook protein (Bd3395), could inactivate flagellar activity upon invasion and its cause the flagellum to be shed [26].

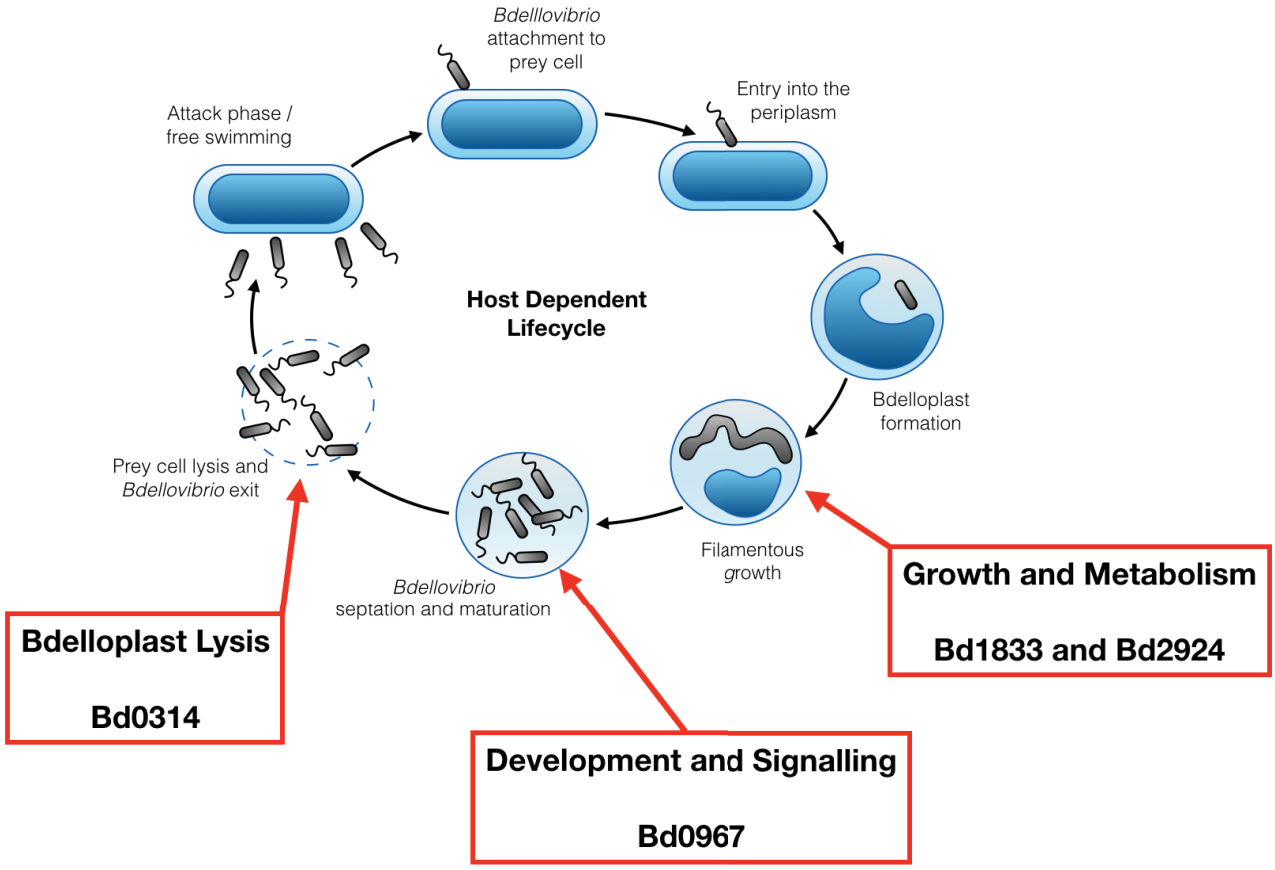


Figure 3. Proteins of interest
This Figure pinpoints the lifecycle events where the proteins of interest, for this thesis, are believed to function.

Protein	Plasmid	Primer (Fwd) (5' - 3')	Primer (Rev) (5' - 3')
Bd0314	pET28b-MBP	GAGAACCTGTACTTCCAATCCGGATCCCGCGA AGTC ATTCCTCTTTG	GGCTTTGTTAGCAGCCGGATCTCACTTGCAAA TAGACAGATTCTGGC
Bd1413	pET28a	GCAGCGGCCTGGTGCCGCGCGGCAGCCATGG CTATAAGATGCAGGCCCTGG	CTCAGTGGTGGTGGTGGTGGTGCTCGAGCTAT TTGCACAAGGGAAGAGTCTTGACG
Bd1440	pET28a	GCAGCGGCCTGGTGCCGCGCGGCAGCCATAG CTACAAAATAGAGGCCCTGTCC	CTCAGTGGTGGTGGTGGTGGTGCTCTTACTTA CAGAAGGATAATTTTTTGACCATGCC
Bd2924	pET28a	AGCAGCGGCCTGGTGCCGCGCGGCAGCCATA TGAAAAACTTTTACCAAGACGGc	CTCAGTGGTGGTGGTGGTGGTGCTCGAGCTAT AAAGAATACTTCTTCATAGCGATACGG
Bd1833	pET28a	GCAGCGGCCTGGTGCCGCGCGGCAGCCATAT GGACAACAAAACTGAAACTAAAATTGAAC	CTCAGTGGTGGTGGTGGTGGTGCTCGAGTTAC TTTTTGCCACGGCTTTTCAG
Bd1833-Pta	pET28a	GCAGCGGCCTGGTGCCGCGCGGCAGCCATTC CAAAGTCTTCATCCGCTCTG	CTCAGTGGTGGTGGTGGTGGTGCTCGAGTTAC TTTTTGCCACGGCTTTTCAG
Bd1833-ME	pET28a	GCAGCGGCCTGGTGCCGCGCGGCAGCCATAT GGACAACAAAACTGAAACTAAAATTGAAC	CTCAGTGGTGGTGGTGGTGGTGCTCGAGCTA AGATTCGCGGTACTGATCCC
Bd1833(SKB)	pET28a	GCAGCGGCCTGGTGCCGCGCGGCAGCCATAT GCAGCGAAAGGAAATATCC	CTCAGTGGTGGTGGTGGTGGTGCTCGAGTTAT TTGGTCAGTTCCGATTTTTTAGC
Bd1833(SKB)-Pta	pET28a	GCAGCGGCCTGGTGCCGCGCGGCAGCCATTC CAAAGTCTTCATTCGCTCTGCC	CTCAGTGGTGGTGGTGGTGGTGCTCGAGTTAT TTGGTCAGTTCCGATTTTTTAGC
Bd0967	pET28a	GCAGCGGCCTGGTGCCGCGCGGCAGCCATAT GAGTTCATTCTCAAAAGGCGTTAGAGG	CTCAGTGGTGGTGGTGGTGGTGCTCGAGCTG AGGCACCAAGGATTTCCC
Bd0606	pET28a	GCAGCGGCCTGGTGCCGCGCGGCAGCCATAT GGGAATGAGAATTTCTACG	CTCAGTGGTGGTGGTGGTGGTGCTCGAGTTA ACCAATCAGTCTTAGTGC

Table 1. Gene primers
A list of primers used in restriction free cloning.

Protein	Mutation	Type	Primer (Fwd) (5' - 3')	Primer (Rev) (5' - 3')
Bd0314	E143Q	QuickChange	ATGGTCAGATTTCAGAGTAACTTTAAAACC	GGTTTTAAAGTTACTCTGAAATCTGACCAT
Bd0314	E154Q	Q5	GTCCTACACCCAGGATTTTAAC	ATTGCGGTTTTAAAGTTAC
Bd0314	Y228F	Q5	TGGCGCCCGCGCCTGGTCTGTG	CCTTTCCAGGCTCCATCC
Bd0314	M150A	Q5	TAAAACCGCAGCGTCCTACACC	AAGTTACTCTCAAATCTGAC
Bd1440	E157Q	Q5	ACGCATGCATCAAACCACCATG	GAAGTCGGGCTGTAAGCG
Bd0967	S547A	Q5	TATCTCTGCTGCCGCTTCCGA	CGGCTGGTCAATACCACC
Bd0967	R299M	Q5	CAAAAAAATCATGGGTAACAAGGGCACGAAAGTG	ACAACGTCTTTCAAATCC
Bd1833	R535A	Q5	GTATTCTTTGGCCCAGCGTAAAGG	AGCTTTTCAACGAAGCTG
Bd1833	R535E	Q5	GTATTCTTTGGAGCAGCGTAAAGGC	AGCTTTTCAACGAAGCTG
Bd1833	Y576C	Q5	TCCATCAACTGTGCCGATGCG	GGATCCACTGACCATACC
Bd1833	Y576S	Q5	TCCATCAACTCTGCCGATGCG	GGATCCACTGACCATACC
Bd1833	E544R	Q5	CAATCTGGGCAGAGCCGAGCGC	ATGCCTTTACGCTGGCGC
Bd1833	E544A	Q5	CAATCTGGGCGCAGCCGAGCGCT	ATGCCTTTACGCTGGCGC
Bd1833	N718D	Q5	GGAATCCTCCGACATCGCCTA	AGATTCGGGAAGACCAGC
Bd1833	ΔLoop	Q5	CTTAACGAGAACAACCTGATCCCGAAACCCTTC	GTCTTGCTCTTTGATCAGTTTTGCCAGTGCATG
Bd1833	R547A	Q5	CGAAGCCGAGGCCTTGATGGCGG	CCCAGATTGATGCCTTTAC
Bd1833	R749A	Q5	CGTTCTGCAAGCAACCACAACGGTTGATGGC	TTGGCAGAGCGGCGCACA
Bd1833	R581K	Q5	CGATGCGGTTAAGCCGATTCTGC	GCATAGTTGATGGAGGATC
Bd0606	LIG	Q5	TAACTCGAGCACCACCAC	TCTTAGTGCAGAGTTTGGC
Bd0606	RLIG	Q5	TAACTCGAGCACCACCAC	TAGTGCAGAGTTTGGCATC

Table 2. SDM primers
A list of primers used for SDM

sites were used in a round of automated model building using Phenix autobuild [111]. The auto-built model successfully traced most of the Bd0314 sequence (2 fragments, 170 residues, R-work = 0.2245, R-free = 0.2446). See Section 2.1.18 for how the initial model was improved. Datasets for Bd0314-Wt and E154Q were collected and solved in a typical process using molecular replacement as outlined in Section 2.1.17 and 2.1.18.

Data Collection				
	Bd0314 Wt (Jcsg A1)	Bd0314 Wt (Jcsg D4)	Bd0314 E143Q	Bd0314 E154Q
Spacegroup	P 2_1 2_1 2	P 2_1 2_1 2_1	P 2_1 2_1 2_1	P 2_1 2_1 2
Cell Dimensions a, b, c (Å)	84.7, 47.8, 49.2	48.9, 64.7, 82.7	48.1, 62.1, 80.9	84.7, 47.9, 49.2
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (Å)	47.818 - 1.361	64.67-1.26	62.13 - 1.82	49.21 - 1.37
Total observations	531095 (14923)	812815 (25973)	968126 (1081)	566104 (29666)
Unique reflections	42769 (1766)	71428 (3469)	19553 (149)	45432 (2275)
Rmerge(I)	0.088 (0.785)	0.0682 (1.786)	0.1612 (0.592)	0.137 (1.713)
Rpim	0.036 (0.384)	0.02 (0.692)	0.021 (0.240)	0.056 (0.694)
I/sigI	16 (2)	13.17 (1.2)	20.1 (2.1)	12.2 (2.2)
CC(1/2)	0.998 (0.804)	0.9994 (0.479)	0.9983 (0.638)	0.994 (0.617)
Completeness (%)	98.1 (82.8)	99.8 (98.4)	87.65 (13.6)	98.4 (100)
Multiplicity	12.4 (8.5)	11.4 (7.5)	49.51 (7.3)	12.5 (13)
Refinement				
R-work	0.1739	0.1699	0.1721	0.1771
R-free	0.1911	0.1940	0.2101	0.1951
RMS (bonds) (Å)	0.008	0.0160	0.006	0.005
RMS (angles) (°)	1.142	1.593	0.789	1.044
Average B-Factor ($Å^2$)	18.1	21.2	18.2	17.8
# of non-hydrogen	1691	1729	1650	1695
Molecules per ASU	1	1	1	1
Ramachandran Plot				
Preferred region	97.1	98.8	96.6	97.1
Allowed region	2.9	1.2	3.4	2.9

Table 3. Bd0314 data collection and refinement

Data Collection	Bd2924 native	Bd2924 (C10-CoA)
Spacegroup	C 2	C 2 2 2_1
Cell Dimensions a, b, c (Å)	137.8, 113.0, 89.0	126.3, 134.7, 133.2
α, β, γ (°)	90, 120.8, 90	90, 90, 90
Resolution (Å)	81.71 - 1.51	133.18 - 1.87
Total observations	1198412 (53307)	1226302 (51988)
Unique reflections	177144 (8571)	93777 (4648)
Rmerge(I)	0.0508 (1.2811)	0.073 (0.860)
Rpim	0.021 (0.555)	0.030 (0.375)
I/sigI	11.16 (1.12)	19.3 (2.1)
CC(1/2)	0.9986 (0.7687)	0.999 (0.833)
Completeness (%)	96.87 (93.85)	100 (99.6)
Multiplicity	6.77 (6.22)	13.1 (11.2)

Refinement		
R-work	0.1545	0.1849
R-free	0.1914	0.2167
RMS (bonds) (Å)	0.018	0.008
RMS (angles) (°)	1.79	1.222
Average B-Factor ($Å^2$)	31.0	32.1
# of non-hydrogen atoms	8572	8909
Molecules per ASU	2	2
Ramachandran Plot (%)		
Preferred region	98	98.2
Allowed region	1.5	1.3

Table 4. Bd2924 data collection and refinement

2.5 Bd1833

2.5.1 Cloning, Expression and Purification

The gene for Bd1833 had been mis-annotated on NCBI and KEGG databases, due to an error in sequencing reads, which results in a premature stop codon. Consequently Bd1833 is annotated as two gene fragments (Bd1833 and Bd1834). The construct of full-length Bd1833 was designed to start at the start codon of *bd1834* and finish at the stop codon of *bd1833*. Bd1833/Bd1834 was successfully cloned into pET28a expression vector and its nucleotide sequence was confirmed by DNA sequencing. DNA sequencing of the Bd1833/Bd1834 cloned region confirmed Bd1833 exists as a single gene and the original stop codon was incorrectly placed due to a frame shift in sequencing reads (Figure 4).



Figure 4. Bd1833 misannotation
bd1833/bd1834 gene is mis-annotated as two genes on NCBI; in fact *bd1833/bd1834* encode a single protein – Bd1833. A frameshift had occurred during the original sequencing producing a premature stop codon and annotation as two gene products. The region starting from *bd1834* start codon through to the stop codon of *bd1833* was cloned and re-sequenced. Re-sequencing revealed the correct sequence read through this region and the position of the frame shift (highlight in red), which in turn makes the stop codon (highlighted in red).

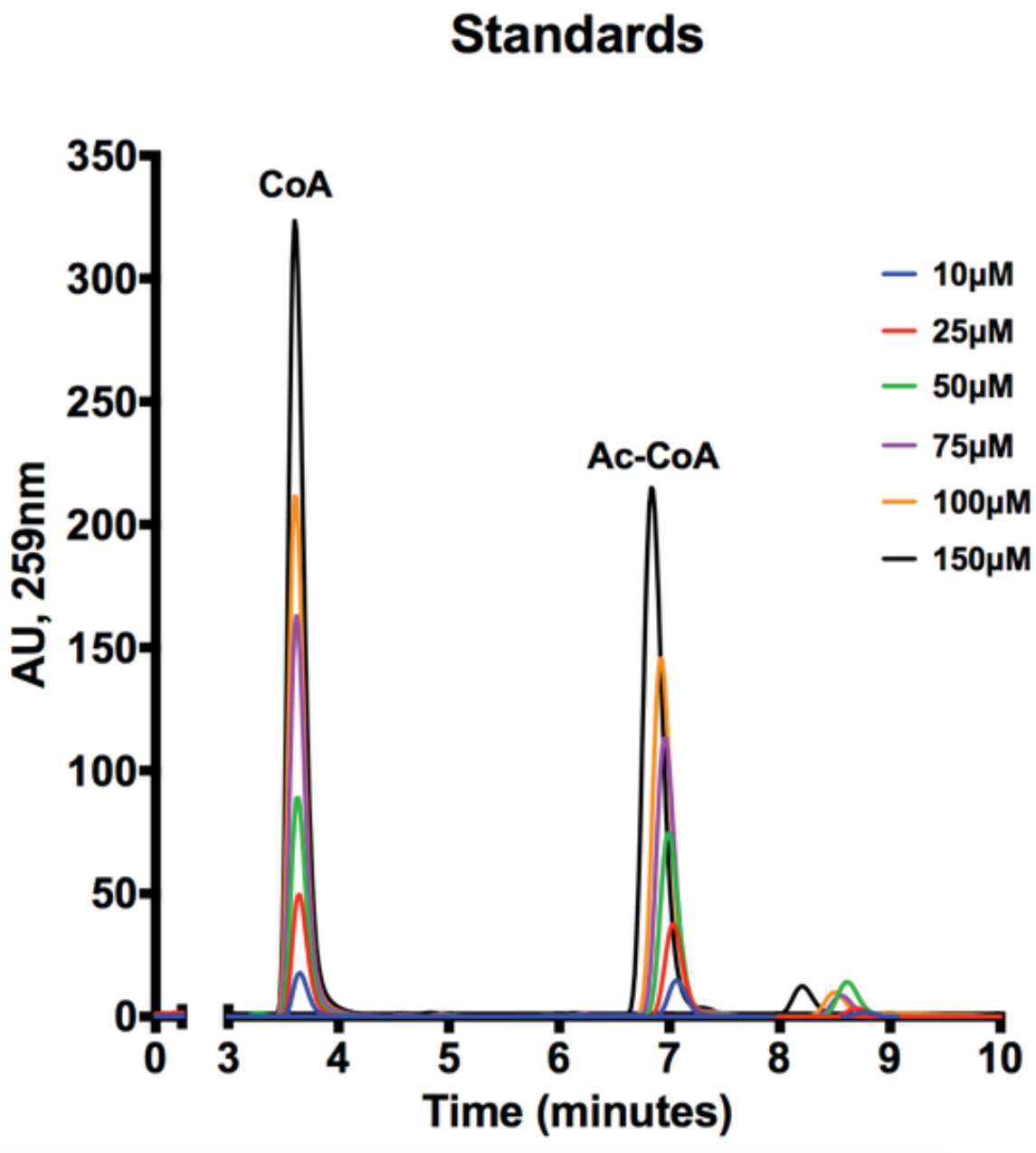


Figure 5. HPLC assay standards
CoA and acetyl-CoA had well-separated and distinct peaks, using HPLC.

2.5.5 Enzymatic activity assay

Spectrophotometric assays were used to monitor the oxidative decarboxylation of L-malate to pyruvate and the reductive carboxylation of pyruvate to L-malate using a CARY 300 Bio, UV-Visible spectrophotometer operated by CARY WinUV kinetics software. All assays recorded the absorbance at 340 nm (NADPH) for a 2 minute period

after the first crystals developed; reduced growth time resulted in better quality diffraction. A soaking experiment was used to get the Bd1833-ME ($NADP^+$) complex structure. Bd1833-ME crystals were transferred into a solution of mother liquor supplemented with (1mM $NADP^+$ and 20% v/v ethylene glycol) and incubated for one hour before being snap frozen in liquid nitrogen.

	Crystallisation Condition
Bd1833 (Full-length)	0.2M sodium citrate tribasic monohydrate, 0.1M BisTris Propane pH 8.5, 20% w/v PEG 3350
Bd1833-ME (native)	1) 0.2M sodium fluoride, 20% w/v PEG 3350 2) 0.1M Bicine pH 9.3, 25% v/v PEG smear medium
Bd1833-ME ($NADP^+$)	0.1M Bicine pH 9.3, 25% v/v PEG smear medium
Bd1833-Pta	0.2M TMAO, 0.1M Tris pH 8.5, 20% w/v PEG2000 MME
Bd1833-Pta (Apo)	0.1M magnesium acetate, 0.1M sodium acetate pH 4.5, 8% w/v PEG 8000
Bd1833-Pta (R535A)	0.2M magnesium chloride hexahydrate, 10% v/v ethylene glycol, 0.1M HEPES pH 7.5, 15% v/v PEG Smear Medium, 5% v/v 2-propanol
Bd1833-Pta (R535E)	0.1M magnesium chloride, 0.1M HEPES pH 7.5, 10% w/v PEG400
Bd1833-Pta (E544R)	0.2M magnesium chloride hexahydrate, 10% v/v ethylene glycol, 0.1M HEPES pH 7.5, 15% v/v PEG Smear Medium, 5% v/v 2-propanol
Bd1833-Pta (Y576S)	0.2m sodium iodide, 0.1M Bis-Tris propane pH 8.5, 20% w/v PEG 3350
Bd1833-SKB (Full)	0.1M Tris pH 8.0, 25% v/v PEG 350 MME
Bd1833-SKB (Pta)	0.005M cobalt(II) chloride hexahydrate, 0.005M cadmium chloride hemi(pentahydrate), 0.005M magnesium chloride hexahydrate, 0.005M nickle(II) chloride hexahydrate, 0.1M HEPES pH 7.5, 12% w/v PEG 3350

Table 5. Bd1833 crystallisation conditions

Data Collection	Bd1833-ME (Bcs-B6)	Bd1833-ME (Pact-E1)	Bd1833-ME ($NADP^+$)
Spacegroup	P 2_1	P1	P 2
Cell Dimensions a, b, c (Å)	48.1, 92.7, 97.1	56.2, 100.2, 130.9	47.8, 92.2, 95.8
α, β, γ (°)	90, 91.1, 90	88.1, 83.7, 81.9	90, 91.3, 90
Resolution (Å)	43.41 - 1.67	37.07 - 2.2	95.78 - 1.67
Total observations	656294 (32403)	475063 (21446)	626225 (29211)
Unique reflections	96511 (4697)	138460 (6888)	96099 (4736)
Rmerge(I)	0.0851 (1.6100)	0.1055 (0.9044)	0.1191 (0.7974)
Rpim	0.036 (0.634)	0.068 (0.621)	0.050 (0.350)
I/sigI	10.81 (1.22)	6.58 (1.02)	6.87 (1.32)
CC(1/2)	0.9977 (0.4634)	0.9939 (0.5009)	0.9962 (0.4718)
Completeness (%)	97.80 (96.54)	97.23 (96.17)	99.90 (98.14)
Multiplicity	6.81 (6.98)	3.43 (3.11)	6.52 (6.17)
Refinement			
R-work	0.1702	0.2109	0.1799
R-free	0.2039	0.2264	0.2159
RMS (bonds) (Å)	0.0077	0.0106	0.0108
RMS (angles) (°)	1.113	1.429	1.257
Average B-Factor ($Å^2$)	37	53.3	34.5
# of non-hydrogen	6808	18553	6913
Molecules per ASU	2	6	2
Ramachandran Plot			
Preferred region	96.4	95.5	96.1
Allowed region	3.3	4.1	3.7

Table 6. Bd1833-ME data collection and refinement

Data Collection	Bd1833-Pta	Apo	R535A	R535E	E544R	Y576S
Spacegroup	$P2_1$	$P\ 2_1\ 2_1\ 2_1$	C 2	C 2	C 2	C 2
Cell Dimensions a, b, c (Å)	111.2, 60.6, 167.8	139.2, 151.3, 285.3	129.6, 183.7, 119.9	129.4, 183.4, 120.2	129.1, 182.9, 119.7	128.8, 182.4, 121.4
α, β, γ (°)	90, 91.1, 90	90, 90, 90	90, 117, 90	90, 117, 90	90, 118, 90	90
Resolution (Å)	91.86 - 1.96	142.66 - 2.72	64.50 - 2.35	107.10 - 2.39	69.22 - 2.33	95.97 - 2.38
Total observations	1083798 (55212)	2082343 (102453)	705212 (37173)	424649 (17692)	698066 (266312)	663750 (28462)
Unique reflections	160802 (7925)	162146 (7836)	103459 (5235)	96539 (4863)	104231 (5084)	98271 (4884)
Rmerge(I)	0.0589 (1.4047)	0.1294 (0.8860)	0.0867 (1.8559)	0.103 (1.975)	0.0872 (1.2654)	0.119 (1.079)
Rpim	0.025 (0.571)	0.038 (0.254)	0.036 (0.744)	0.055 (1.201)	0.036 (0.600)	0.042 (0.47)
I/sigI	15.70 (1.14)	11.40 (1.47)	10.69 (1.16)	8.5 (0.5)	12.03 (1.14)	10.8 (0.7)
CC(1/2)	0.999 (0.637)	0.998 (0.600)	0.999 (0.573)	0.998 (0.697)	0.999 (0.590)	0.999 (0.569)
Completeness (%)	99.92 (99.92)	99.95 (97.82)	99.90 (100.00)	98.4 (99.7)	99.72 (97.98)	100 (100)
Multiplicity	6.74 (6.97)	12.84 (13.07)	6.81 (7.10)	4.4 (3.6)	6.70 (5.24)	6.8 (5.8)

Refinement						
R-work	0.1829	0.2520	0.2158	0.2153	0.2069	0.2222
R-free	0.2052	0.2995	0.2368	0.2602	0.2265	0.2725
RMS (bonds) (Å)	0.009	0.012	0.0103	0.01	0.0097	0.01
RMS (angles) (°)	1.178	1.590	1.498	1.415	1.448	1.321
Average B-Factor ($Å^2$)	29.5	86.2	45.7	75.5	74.9	74.1
# of non-hydrogen	16475	31268	15572	15628	15849	15720
Molecules per ASU	6	12	6	6	6	6
Ramachandran Plot						
Preferred region	95.8	96.2	98.7	96.6	97.0	96.3
Allowed region	4.0	3.6	1.3	3.1	2.7	3.6

Table 7. Bd1833-Pta data collection and refinement

Data Collection	Bd1833 (Full-length)
Spacegroup	C 2
Cell Dimensions *a, b, c (Å)*	250.422 144.792 171.138
α, β, γ (°)	90.000 100.605 90.000
Resolution (Å)	124.80 - 2.72
Total observations	1086009 (56926)
Unique reflections	160818 (8003)
Rmerge(I)	0.1 (2.788)
Rpim	0.042 (1.125)
I/sigI	11.0 (0.6)
CC(1/2)	0.999 (0.317)
Completeness (%)	99.9 (100)
Multiplicity	6.8 (7.1)

Refinement	
R-work	0.1922
R-free	0.2277
RMS (bonds) (Å)	0.0108
RMS (angles) (°)	1.437
Average B-Factor (Å^2)	108
# of non-hydrogen atoms	34603
Molecules per ASU	6
Ramachandran Plot (%)	
Preferred region	95.5
Allowed region	4.1

Table 8. Bd1833 full-length data collection and refinement

Data Collection	Bd1833-SKB (Full-length)	Bd1833-SKB (Pta)
Spacegroup	P 4_1 2_1 2	F 2 3
Cell Dimensions a, b, c (Å)	174.62, 174,62, 203.29	307.09, 307.09, 307.09
α, β, γ (°)	90, 90, 90	90, 90, 90
Resolution (Å)	132.46 - 3.71	54.29 - 2.71
Total observations	885942 (45499)	663665 (33984)
Unique reflections	34022 (1653)	64779 (3209)
Rmerge(I)	0.195 (6.015)	0.1309 (1.8829)
Rpim	0.034 (0.620)	0.043 (0.606)
I/sigI	13.6 (0.7)	12.06 (1.25)
CC(1/2)	1.000 (0.393)	0.9981 (0.4515)
Completeness (%)	100 (100)	100.00 (99.97)
Multiplicity	26.0 (27.5)	10.25 (10.59)

Refinement		
R-work	0.2503	0.2040
R-free	0.2996	0.2392
RMS (bonds) (Å)	0.004	0.013
RMS (angles) (°)	0.783	1.43
Average B-Factor ($Å^2$)	199.1	64.5
# of non-hydrogen atoms	17163	10592
Molecules per ASU	3	3
Ramachandran Plot (%)		
Preferred region	95.1	97.0
Allowed region	4.3	3.0

Table 9. Bd1833 SKB data collection and refinement

Data Collection	Bd0967 Wt	Bd0967 S457A
Spacegroup	P1	P1
Cell Dimensions a, b, c (Å)	73.4, 95.8, 114.1	72.9, 95.0, 113.3
α, β, γ (°)	77.5, 71.3, 81.9	77.7, 71.8, 81.9
Resolution (Å)	106.26 - 2.04	38.42 - 2.58
Total observations	848316 (18620)	304174 (15674)
Unique reflections	179707 (8159)	87121 (4331)
Rmerge(I)	0.1656 (0.3533)	0.1133 (1.0666)
Rpim	0.079 (0.280)	0.071 (0.654)
I/sigI	4.78 (1.23)	7.52 (1.13)
CC(1/2)	0.9517 (0.7332)	0.9946 (0.4479)
Completeness (%)	97.84 (89.57)	98.52 (98.12)
Multiplicity	4.72 (2.28)	3.49 (3.62)
Refinement		
R-work	0.1990	0.2118
R-free	0.2399	0.2607
RMS (bonds) (Å)	0.0104	0.010
RMS (angles) (°)	1.431	1.328
Average B-Factor (Å^2)	43.1	59.9
# of non-hydrogen atoms	21634	19935
Molecules per ASU	4	4
Ramachandran Plot (%)		
Preferred region	97.6	97.0
Allowed region	1.4	2.2

Table 10. Bd0967 data collection and refinement

2.6.11 MST Cdg Binding Assay

The same procedure as described for Bd2924 MST was used. 20 mM HEPES pH 7.5, 50 mM KCl, 5 mM $MgCl_2$ was used as the MST buffer.

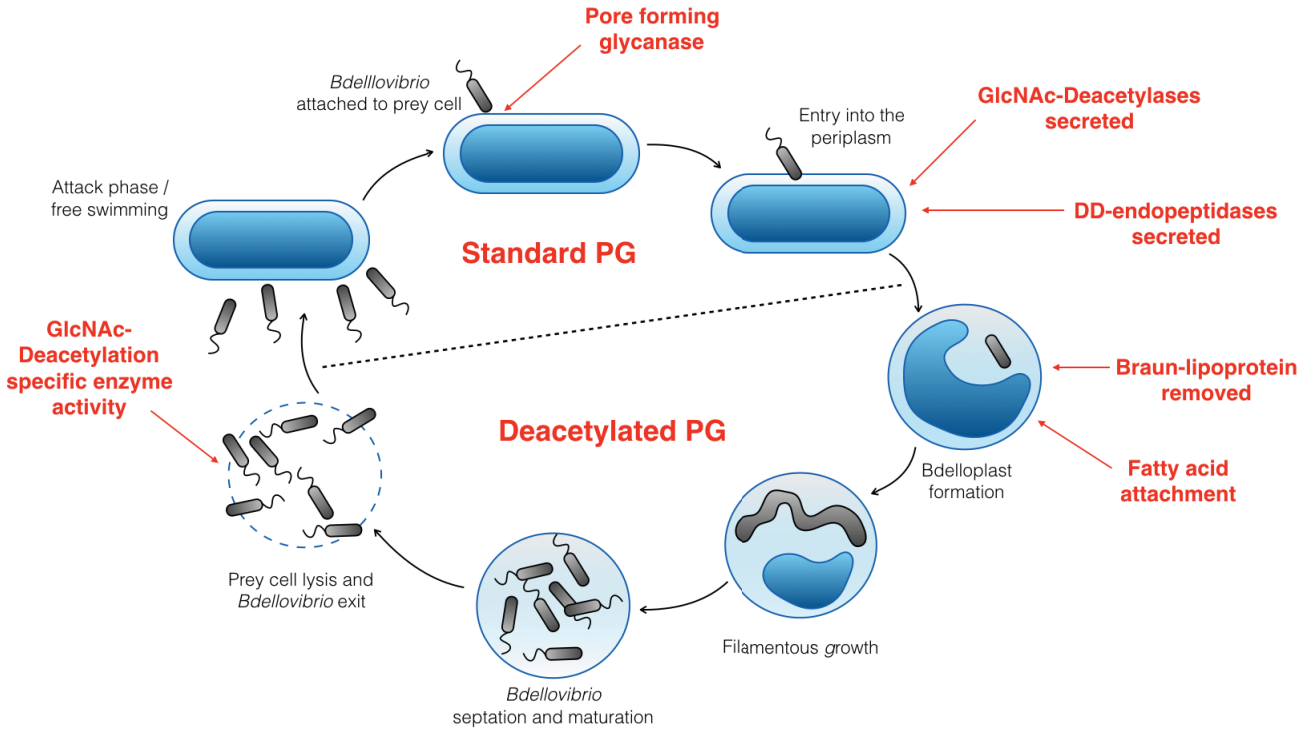


Figure 6. *Bdellovibrio* lifecycle highlighting prey cell wall modifications.
Bdellovibrio actively seeks out prey cells and attach to the outer membrane. They then secrete a localised pore forming glycanase that creates a small entry pore in the prey cell wall that Bdellovibrio squeezes through and then reseals. During invasion *Bdellovibrio* secretes deacetylases and DD-endopeptidases. GlcNAc-deacetylases (Bd0468 and Bd3279) deacetylate the prey cell wall marking 'prey' from 'self'. *N*-deacetylation of the prey cell wall makes the bdelloplast structure resistant to lysozyme activity. DD-endopeptidases (0816 and 3459) break cross-links in the prey cell wall and rounds the prey cell into the bdelloplast structure. Rounding of the prey cell acts as an 'occupied' signal preventing further invasion events. A significant proportion of the prey cell Braun-lipoprotein is then removed and fatty acid chains are attached to the prey cell wall, which stabilises the bdelloplast structure. At the end of the predatory cycle a final lytic enzyme activity is required to completely solubilise the prey cell wall and lyse the bdelloplast structure. The final lytic enzyme activity must be highly regulated and active towards the deacetylated prey cell wall.

3.1.1 Peptidoglycan Overview

3.1.1.1 Overview

Peptidoglycan is an essential feature of bacterial cell walls, located on the outside of the cytoplasmic membrane. The main function of peptidoglycan is to withstand turgor pressure and maintain the cellular integrity. If the peptidoglycan layer is compromised, by inhibiting its biosynthesis (e.g. antibiotics) or specific degradation (e.g. lysozymes), cells lyse. Peptidoglycan is directly involved in cell growth and division, while also

being required to maintain cell shape and serve as a molecular scaffold for anchoring other cell envelope components such as teichoic acids [116].

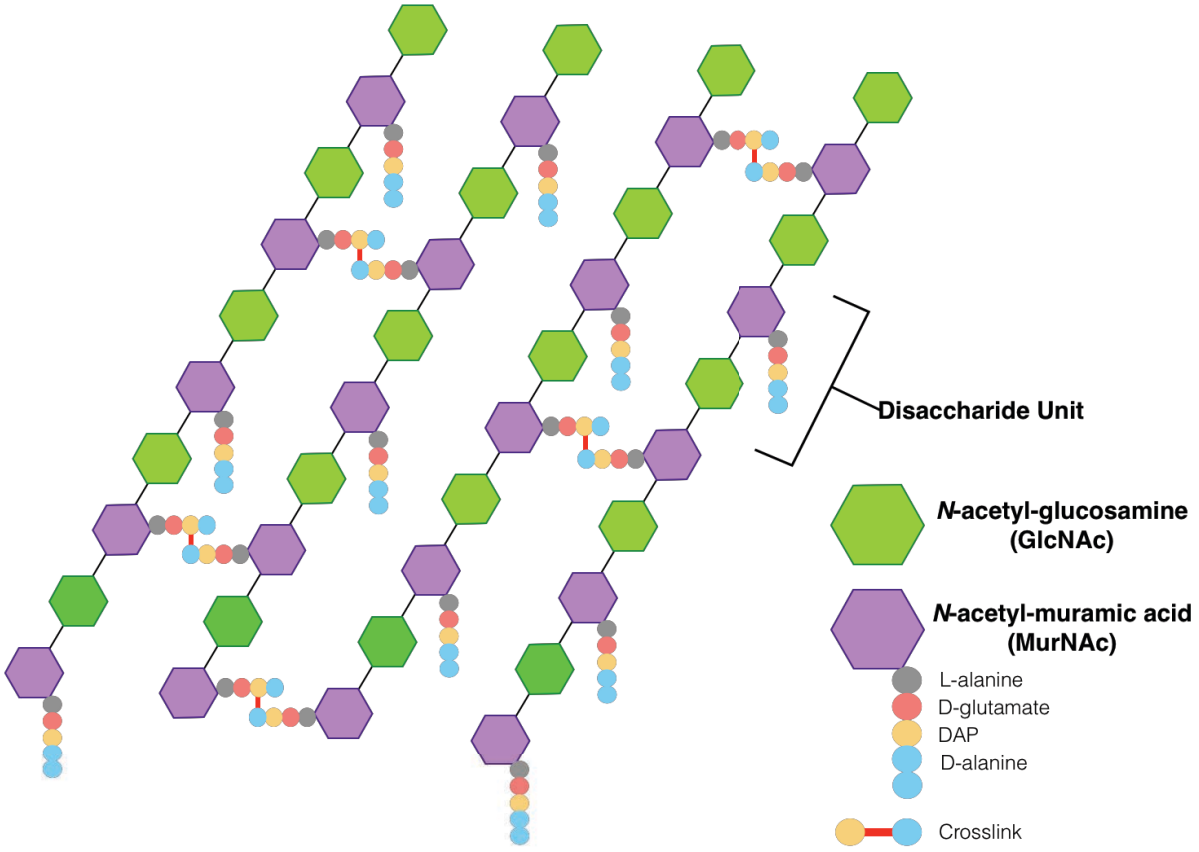


Figure 7. Peptidoglycan structure
This Figure represents the peptidoglycan structure of *E. coli*. The glycan strand is composed of alternating β-1-4 linked GlcNAc (green) and MurNAc (purple) residues. The basic disaccharide subunit (monomer) is highlighted. Components making up the *E. coli* peptide stem are highlighted. Cross-links between the tetrapeptide stems are represented by a red bond.

3.1.1.2 Structure and Composition

Peptidoglycan is comprised of a linear chain of glycan stands that are cross-linked together by short peptide strands (Figure 7). The glycan strands or made up of alternating *N*-acetylglucosamine (GlcNAc) and *N*-acetylmuramic acid (MurNAc) residues, linked via β-1-4 bonds. MurNAc contains a D-lactoyl group that is substituted

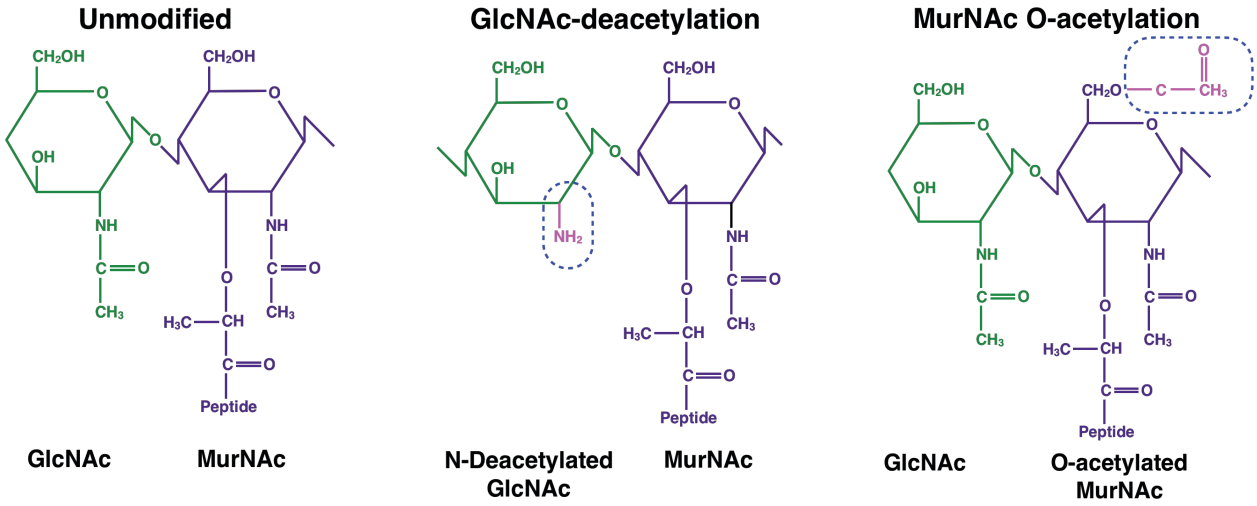


Figure 8. Modifications of the peptidoglycan strand
The unmodified PG disaccharide unit of β-1-4 linked GlcNAc (green) and MurNAc (Purple) residues is illustrated on the left. N-deacetylation of GlcNAc residues is shown in the central panel. *O*-acetylation of MurNAc residues is shown to the right. Chemical alterations are coloured in pink and circled in blue.

3.1.2 Peptidoglycan modifications and pathogenesis

As previously mentioned the fine structure of peptidoglycan varies between bacterial species. Moreover, certain modifications are common in pathogenic bacteria and are used as virulence factors to invade host defences. These modifications are typically generated in the glycan strand by post-synthetic reactions, including acetylation of the C6 hydroxyl group or deacetylation of the C2 amine group of GlcNAc and MurNAc residues. There are three main types of glycan strand modifications: *N*-deacetylation, *O*-acetylation and attachment of surface biopolymers, which will be discussed further.

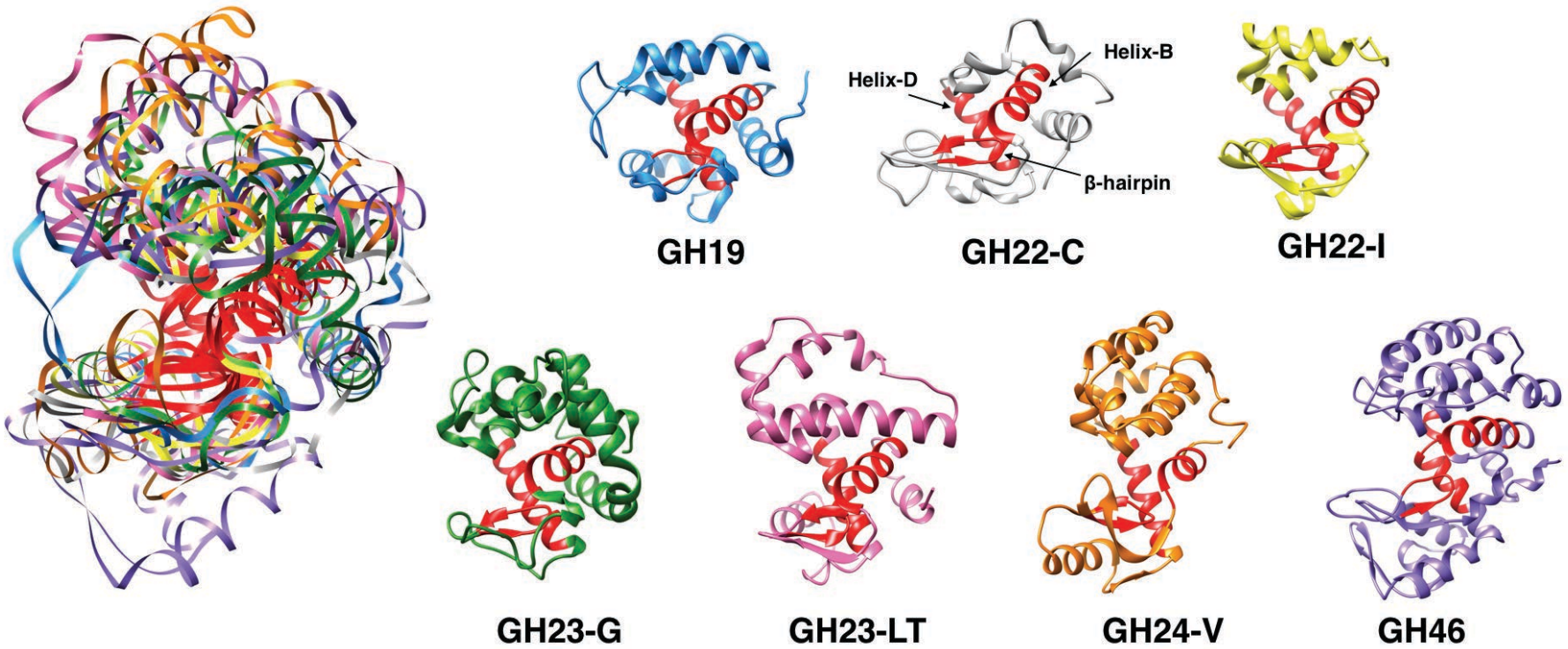


Figure 9. Structures of lysozyme superfamily members
The top left images shows the superimposition of seven representative structures of lysozyme superfamily members. Each representative is also displayed in an individual panel in the same relative orientation. The conserved core region (2 alpha helices and strict beta hairpin) is coloured in red in all images. GH19 –blue (PDB: 3W6D), GH22-C –gray (PDB: 2VB1), GH22-I – yellow (PDB:2DQA), GH23-G – green (PDB:154L), GH23-LT – pink (catalytic domain of PDB:1LSY), GH-24-V – orange (PDB:2LZM), GH46 – purple (PDB:1QGI). The structural elements of HEWL (GH22-C) are labelled for reference.

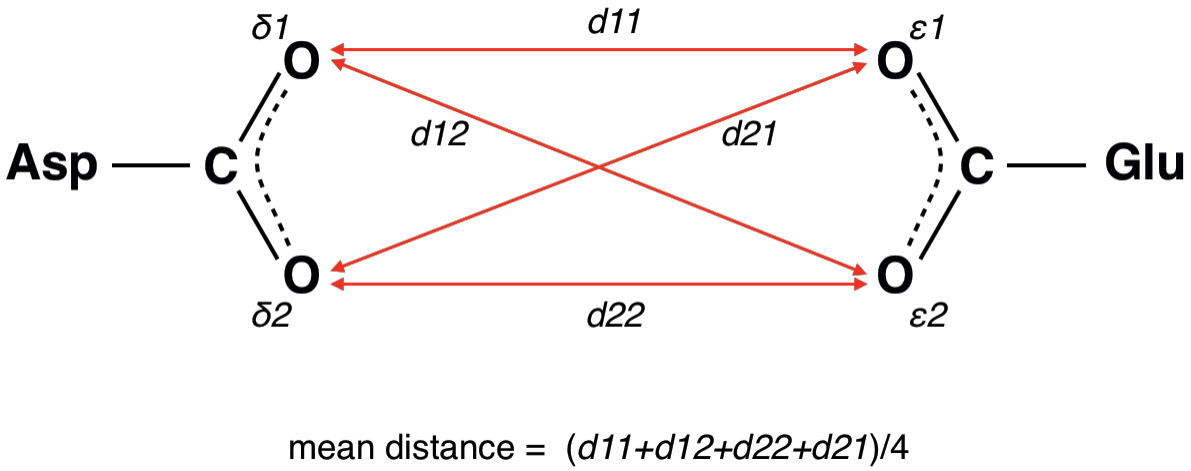


Figure 10. Catalytic distances
The average distance between the catalytic residues is calculated by taking the mean average distance between the carboxyl groups of the catalytic residues.

Beyond the active site, conserved residues contribute to the hydrophobic core of the enzymes. Bulky hydrophobic residues line helix-B, several residues towards the N-terminus of the catalytic glutamate. These hydrophobic residues form a hydrophobic pocket with conserved residues located on the invariant helix-D. Furthermore a strictly conserved glycine residue is located on strand-2, which maintains packing [141].

The stable folding pattern of lysozyme superfamily members is suggested to have originated from a common ancestor [149]. The invariant lysozyme core elements (helix-B, helix-D and two β-strands with a strict β-hairpin) alongside the conserved catalytic residues suggest the common fold did not arrive through convergent evolution. Moreover, the presence of non-essential elements that append the core structure suggest divergence from a common ancestor.

3.1.3.3 Catalytic mechanism

There are two reported catalytic mechanisms utilised by lysozymes, which have been determined using biochemical and structural data. Lysozymes and other GH family members are considered to either be retaining or inverting enzymes, depending on the product created after the glycosidic linkage is cleaved. Retaining enzymes use a double displacement mechanism that creates a covalent glycosyl-enzyme intermediate. The reaction retains the original stereochemistry at the anomeric centre. Retaining enzymes possess two acidic catalytic residues, where one carboxyl group acts as an acid/base while the other acts as a nucleophile.

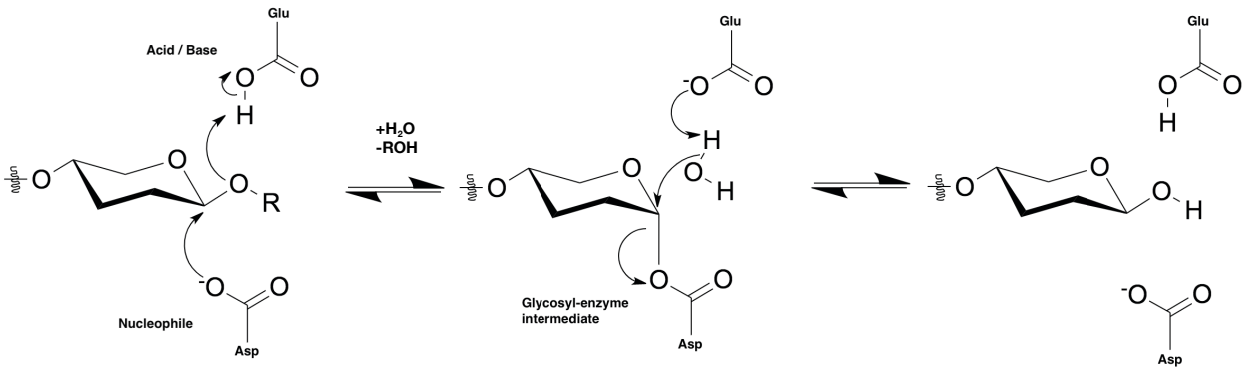


Figure 11. Retaining enzyme mechanism
Proposed mechanism for retaining lysozymes. ROH represents the leaving group.

GH-22 family member, HEWL, is the most well-characterised lysozyme in terms of structure, substrate binding and catalytic mechanism. The retaining lysozyme mechanism is exemplified by HEWL. The structure of HEWL has been solved in complex with various ligands revealing six sugar subsites, A-F [150]. HEWL possesses two catalytic residues, with E35 (helix-B) acting as a proton donor and D52 as a nucleophile [132]. It was initially suggested from structural evaluation that D52 in HEWL played a

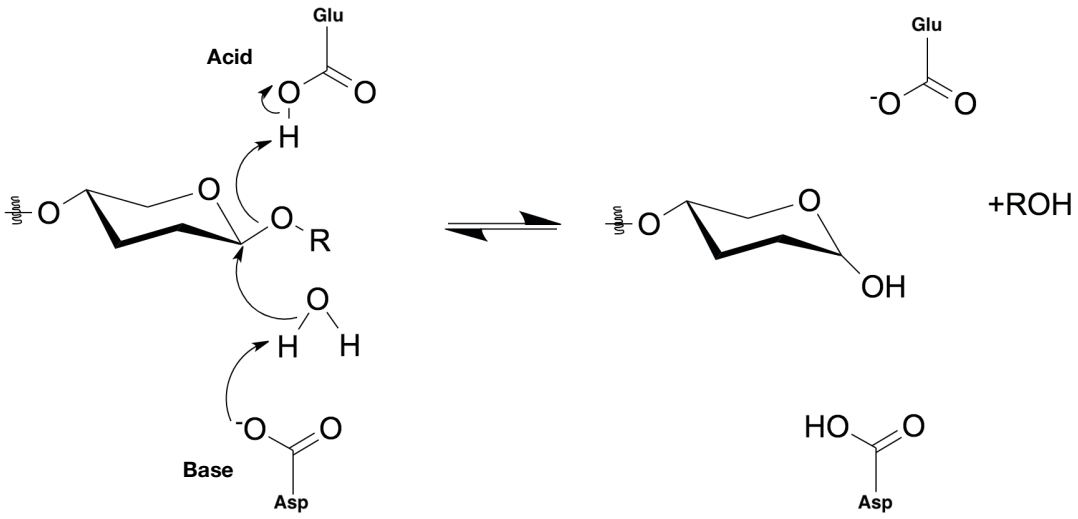


Figure 12. Inverting enzyme mechanism
Proposed mechanism for inverting lysozymes. ROH represents the leaving group.

3.1.3.4 Lytic transglycosylases – housekeeping enzymes

LTs are peptidoglycan-degrading enzymes that fulfil a housekeeping role. LTs are integral to PG maintenance, turnover and remodelling during cell division and cell growth [154]. Like classical lysozymes they cleave the β-1-4 linkage between MurNAc and GlcNAc of PG. However, they do not perform catalysis through hydrolysis of the substrate but rather use an unusual intra-molecular glycosltransferase reaction, which produces a terminal 1,6-anhydro-MurNAc residue [117]. Catalytically LTs perform quite a different reaction to classical lysozymes. LTs possess the invariant catalytic glutamate located in the ES-motif on the central α-helix. However, they lack a second catalytic residue and do not require a catalytic water molecule. Instead LTs use a substrate-assisted catalytic mechanism [155]. The catalytic glutamate is proposed to serve initially as the general acid, donating a proton to the glycosidic linkage and forming a muramoyl oxazolinium-ion intermediate. The glutamate then switches roles and acts as a base to abstract the C6 hydroxyl proton, leading to the formation of the 1,6- anhydromuramoyl product [154].

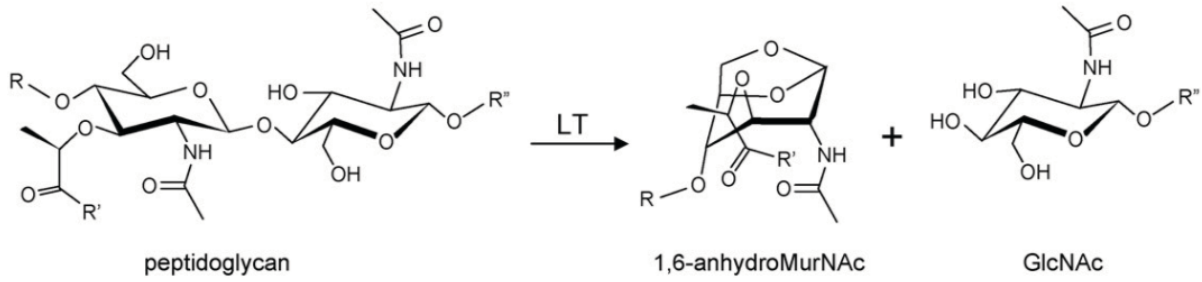


Figure 13. Lytic Transglycosylase reaction.
Lytic Transglycosylases catalyse the cleavage of the β-1-4 linkage between MurNAc and GlcNAc of PG with a single catalytic residue. Cleavage by LTs results in the formation of a 1.6-anhydroMurNAc terminal residue.

LTs can be divided into four families (1-4) and several *E. coli* LTs have been structurally and biochemically characterised for families 1 and 2 [156]. LTs are common in Gram-negative bacteria and always found in the periplasm. A small number of LTs have been reported in Gram-positive bacteria [157]. LTs can exist as soluble species or attached to the outer membrane via the N-terminus. Structures of the LT catalytic domain adopt a common fold that resembles G-type lysozymes [158].

3.1.4 *Bdellovibrio* lifecycle and PG modifications

The obligate predatory bacterium, *Bdellovibrio bacteriovorus*, has adapted PG modifying enzymes to facilitate its unusual biphasic predatory lifecycle. Early investigation into predation by *Bdellovibrio* highlighted a number of prey cell wall modifications required for successful completion of the predatory lifecycle and hinted at the enzymes responsible.

During penetration of an *E. coli* prey, *Bdellovibrio* rapidly solubilises approximately 10-15% of the prey peptidoglycan in a local, controlled manner [45]. Two enzyme

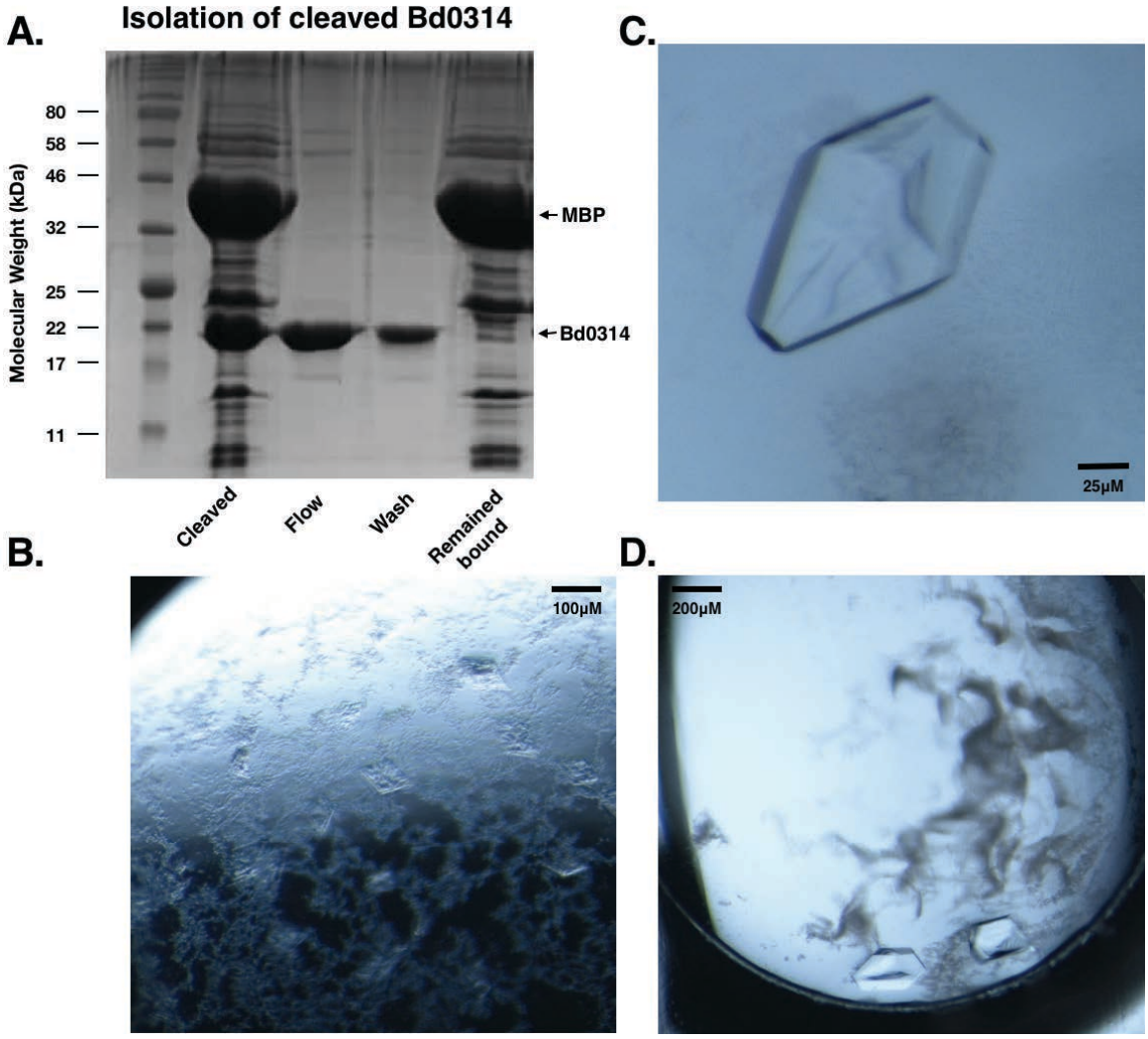


Figure 14. Bd0314 purification and crystallisation
A. An MBP tag was used to express and purify Bd0314. The MBP tag was cleaved by incubation with TEV and removed by a second passage over a HisTrap column. **B.** Bd0314-Wt crystals were small thin plates. **C.** The Bd0314-E143Q crystal used for SSAD phasing was optimised. **D.** Bd0314-Wt crystals that grew in a different space group.



Figure 15. Bd0314 structure schematic
Bd0314 is a 254 residue protein that has a 180 residue catalytic domain. Bd0314 primary sequence suggests the presence of an N-terminal signal peptide (SP) containing a Lipobox (LB). Between the predicted lipobox and the catalytic domain is a 50 residue region of predicted disorder. Secondary structure predictions made using the XtalPred server.

bonds. E175 located on helix G hydrogen bonds to S159, which stabilises the end of the β-sheet and turn. In addition Bd0314 may contain a third β-strand, where I172 (located immediately after the β-hairpin) and R167 of strand-2 make a β-sheet-like interaction.

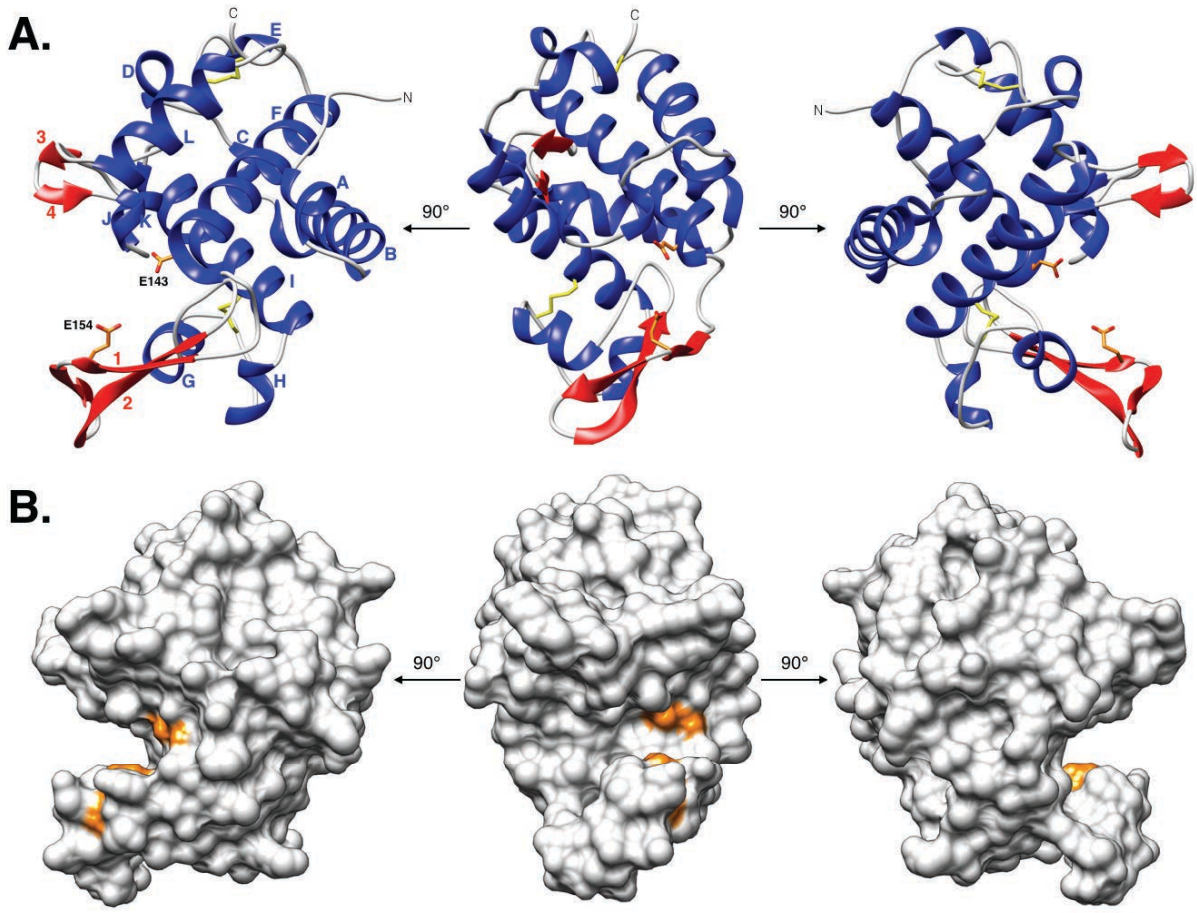


Figure 16. Bd0314 structure
A. The secondary structure of Bd0314 shown from three angles 90° apart. α-helices are coloured blue and β-strands are coloured red. Disulphide bonds are highlighted in yellow and the positions of the catalytic E143 and E154 are displayed in orange. Helices are labelled with blue letters and strands are labelled with red numbers in the first panel. **B.** Surface view of Bd0314 illustrating the active-site cleft. The location of the catalytic residues is highlighted in orange.

3.2.2.2 The catalytic site

Analysis of the Bd0314 structure revealed two potential catalytic residues. The classical ES invariant glutamate (E143) is located on the end of helix-F facing into the catalytic site. The second (E154) is located on strand-1, and is projected into the active site, directly opposite E143. The typical location of the second catalytic Asp in the lysozyme superfamily is on strand-2. However, in Bd0314 S166 occupies this location and the hydroxyl group makes a hydrogen bond to the E154 carboxyl-group, which may help to retain the orientation of E154. The presence and location of E143 and E154 suggest they act as a general acid-base pair. The distance between the two acidic residues is 8Å, suggesting that Bd0314 proceeds by an inverting catalytic mechanism [147].

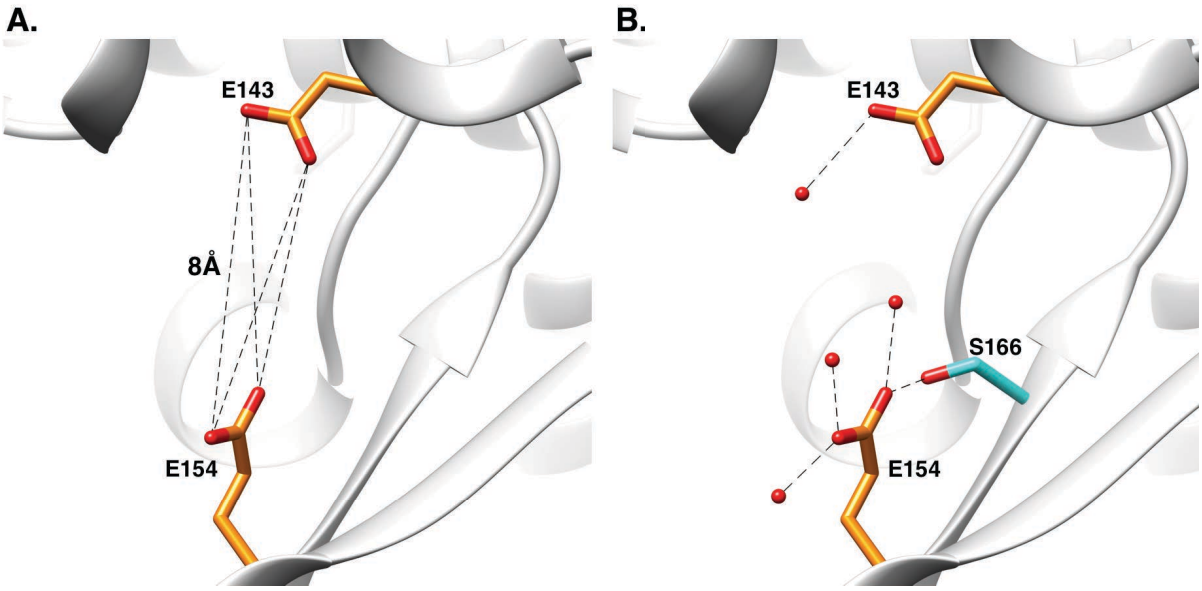


Figure 17. Bd0314 catalytic residues
A. The distance between the carbonyl groups of E143 and E154 is ~8Å, which suggests Bd0314 is an inverting lysozyme. **B.** E143 and E154 make hydrogen bonds with water molecules in the active-site cleft (potential catalytic water molecules). E154 also makes a hydrogen bond with S166, which may have stabilising effects during catalysis.

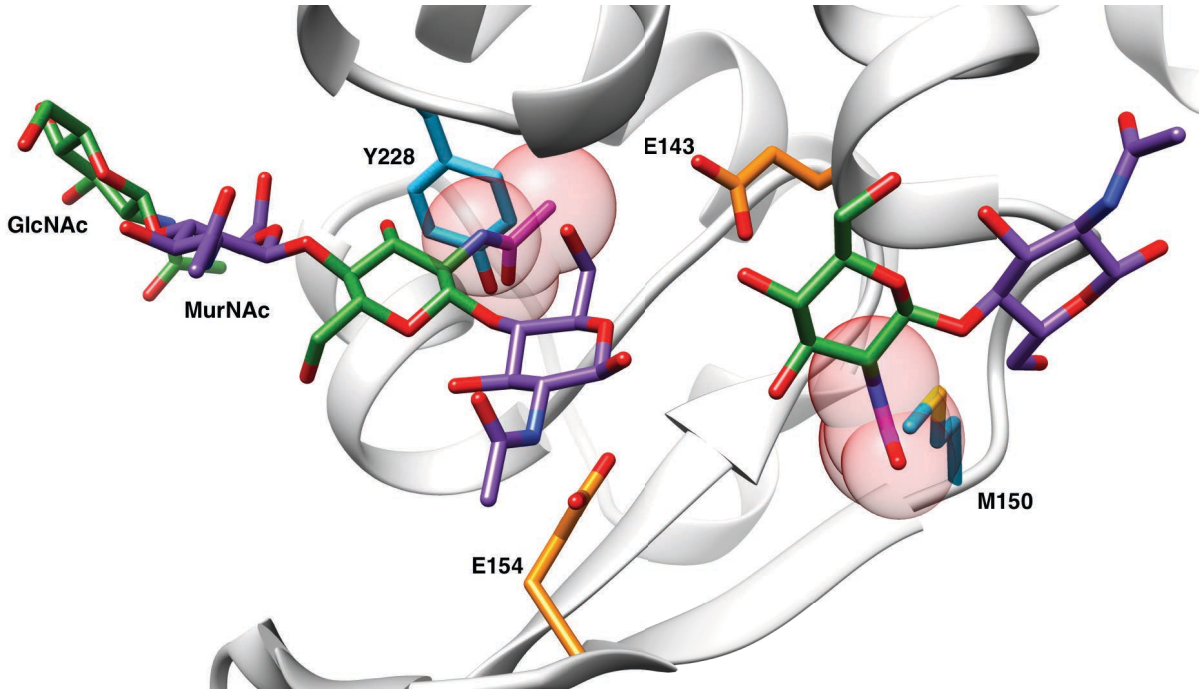


Figure 18. Bd0314 substrate-binding site
Model of Bd0314 substrate binding, generated by superimposition of a regular lysozyme (PDB:1LZS) in complex with a hexa-*N*-acetyl-chitohexaose substrate. Bd0314 was structurally aligned with PDB:1LZS by least-squares minimization of the structurally conserved regions (helix-F, helix-I and active site beta-hairpin) using the CCP4 suit superpose software (length 2474 atoms from RSMD=1.13Å). Bd0314 is displayed in ribbon form, coloured in white. The ligand is alternatively coloured green (GlcNAc) and purple (MurNAc), to represent the positions of GlcNAc and MurNAc for a peptidoglycan substrate similarly to that in Figure 8. The substrate GlcNAc residue acetyl-group faces inwards towards the cleft, whereas MurNAc residues face outwards to accommodate the peptide stem. The catalytic glutamate residues (E143 and E154) of Bd0314 are shown in orange and are correctly positioned to cleave after the MurNAc residue. The acetyl-groups of GlcNAc residues clash with residues of the peptide-binding cleft. The residues that clash with the acetyl-group of GlcNAc residues (Y228 and M150) are shown in cyan. The red van der waals surface of the acetyl groups demonstrates the degree of clash.

3.2.2.3 Substrate selection

Bd0314 has an extensive substrate-binding cleft provided by the extended β-sheet, which forms the base and the additional β-sheet region of the core domain, which creates a peak to the roof of the binding cleft. We have been unable to obtain a protein:ligand complex after many attempts with various ligands. However, numerous structures of lysozyme-superfamily members in complex with ligands have been published [153, 162, 163]. Superimposition of these ligand complex structures with Bd0314 (using LSQ superimposition of helix-F (residues 133-144), helix-I (residues

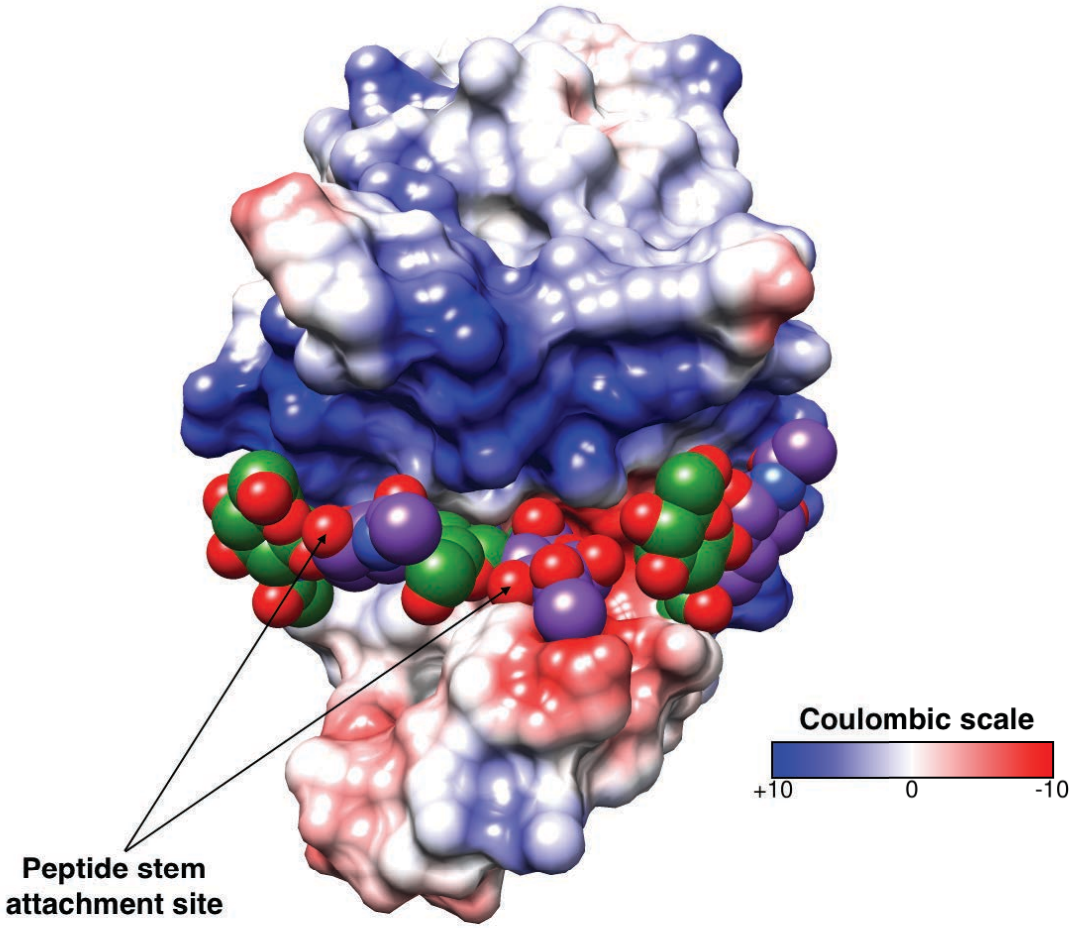


Figure 19. Substrate binding surface view
The substrate-binding model shown in the surface view demonstrates how the acetyl-groups of GlcNAc residues face inwards whereas the acetyl-group of MurNAc residues face outwards, so the peptide stem can accommodated. The roof of the substrate-binding cleft carries a positive charge. Substrate colours are the same as Figure 8.

3.2.2.4 Bd0314 structure comparison (DALI) and homologues

The DALI protein structure comparison server was used to identify proteins with similar structural elements from the PDB [160]. The top 25 matches from DALI are shown in Table 11. The greatest proportion of sequence and structural conservation for Bd0314 occurs for the protein core and the catalytic cleft. This region includes (residues 130-154) the central helix-F and part of the strand-1 and (residues 164-207) part of strand-2 up to and including central helix-I (Figure 20). Lytic transglycosylase

family proteins (GH23), are the most frequent structural matches. The catalytic core domain of Slt70 PDB: 1QTE (a soluble LT) gives the best structural match. However, the lower subdomain containing the second catalytic residue is not present in Slt70. Interestingly human lysozyme PDB:5LSH, is one of the closest structural matches and has a high Z-score, low rmsd. Four regions of Bd0314, residues 1-14, 26-54, 82-90 and 143-150, show little structural conservation to other lysozyme structures or primary sequences (Figure 28.B).

No:	Chain	Z	rmsd	lali	nres	%id	PDB	Description
1:	1qte-A	8.8	2.8	119	620	18	PDB	MOLECULE: SOLUBLE LYTIC TRANSGLYCOSYLASE SLT70;
2:	6cf9-A	8.8	2.7	112	516	13	PDB	MOLECULE: LYTIC TRANSGLYCOSYLASE;
3:	5lsh-A	8.4	2.6	103	130	15	PDB	MOLECULE: LYSOZYME C;
4:	4c5f-A	8.2	2.9	116	330	18	PDB	MOLECULE: MEMBRANE-BOUND LYTIC MUREIN
5:	4owd-A	8.1	3.2	118	429	16	PDB	MOLECULE: MEMBRANE-BOUND LYTIC MUREIN
6:	5vas-A	8.1	2.6	101	129	18	PDB	MOLECULE: LYSOZYME;
7:	6fcu-A	8.0	2.9	120	615	16	PDB	MOLECULE: SOLUBLE LYTIC MUREIN TRANSGLYCOSYLASE;
8:	4yim-B	7.9	2.8	116	586	13	PDB	MOLECULE: PUTATIVE SOLUBLE LYTIC MUREIN
9:	3bkh-A	7.8	3.0	113	268	10	PDB	MOLECULE: LYTIC TRANSGLYCOSYLASE;
10:	4dq7-A	7.7	3.1	108	168	12	PDB	MOLECULE: MEMBRANE PROTEIN PHI6 P5;
11:	3w6b-A	7.7	3.4	119	160	22	PDB	MOLECULE: LYSOZYME-LIKE CHITINOLYTIC ENZYME;
12:	2xqo-A	7.6	3.0	125	210	16	PDB	MOLECULE: CELLULOSOME ENZYME, DOCKERIN TYPE I;
13:	1lsg-A	7.4	2.6	101	144	21	PDB	MOLECULE: HEN EGG WHITE LYSOZYME;
14:	4xp8-A	7.0	2.9	84	88	19	PDB	MOLECULE: ETGA PROTEIN;
15:	1am7-A	6.8	3.6	102	154	7	PDB	MOLECULE: LYSOZYME;
16:	2dqa-A	6.7	2.9	99	123	13	PDB	MOLECULE: LYSOZYME;
17:	5e27-A	6.7	2.7	74	226	12	PDB	MOLECULE: RESUSCITATION-PROMOTING FACTOR RPFB;
18:	2dkv-A	6.0	4.0	140	286	10	PDB	MOLECULE: CHITINASE;
19:	5xsv-A	6.0	3.2	121	486	9	PDB	MOLECULE: CHITINASE;
20:	5nm7-G	5.9	3.1	117	263	12	PDB	MOLECULE: PEPTIDOGLYCAN-BINDING DOMAIN 1;
21:	4ok7-A	5.9	3.3	126	223	10	PDB	MOLECULE: ENDOLYSIN;
22:	4luq-A	5.6	3.5	125	409	10	PDB	MOLECULE: UNCHARACTERIZED PROTEIN;
23:	4g9s-A	5.4	3.2	106	187	16	PDB	MOLECULE: GOOSE-TYPE LYSOZYME;
24:	4fdy-A	4.9	2.9	98	295	10	PDB	MOLECULE: SIMILAR TO LIPOPROTEIN, NLP/P60 FAMILY;
25:	1d0k-A	4.9	3.2	111	318	14	PDB	MOLECULE: 35KD SOLUBLE LYTIC TRANSGLYCOSYLASE;

Table 11. DALI results
Top 25 matches from the DALI server for Bd0314 coordinates as the search query [160].

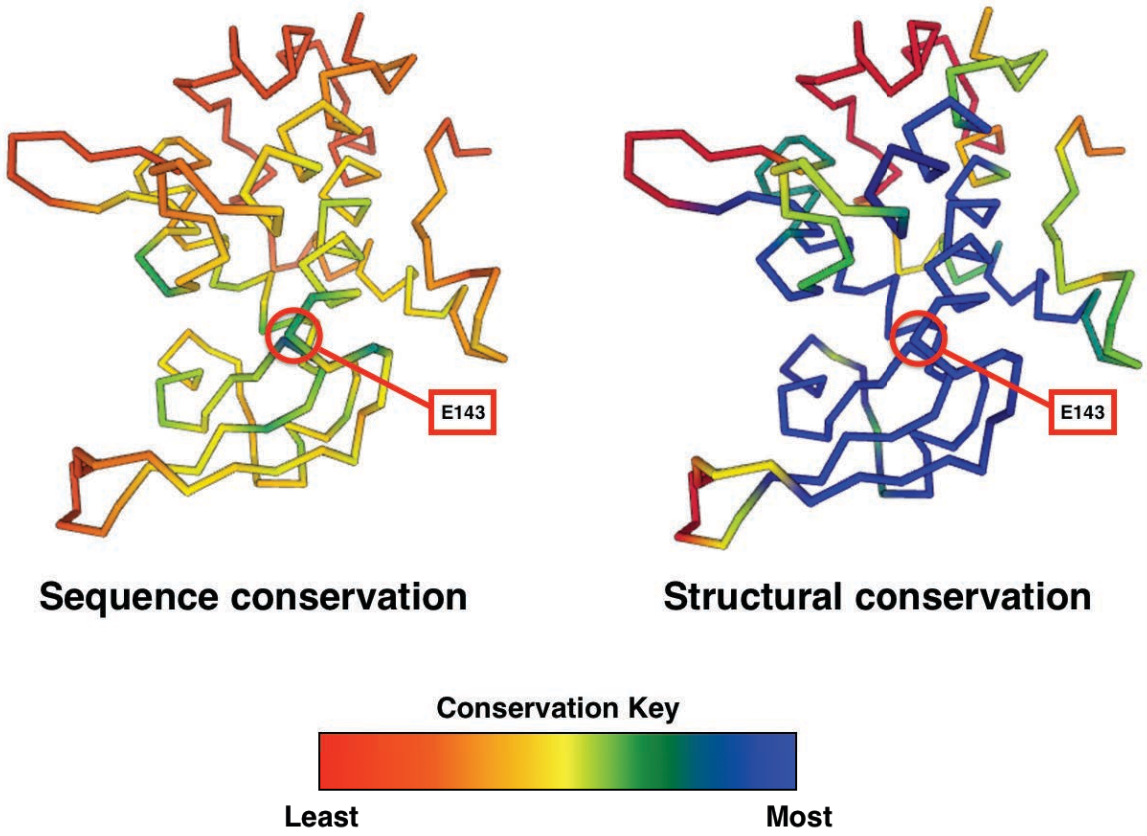


Figure 20. Bd0314 conservation
The sequence and structural conservation (coloured key) is mapped onto the Cα-trace of Bd0314 structure. Sequence conservation in Bd0314 is low, whereas the structural conservation is higher in regions. The core region of Bd0314 has the highest sequence and structural conservation. These results are obtained the DALI [160].

BLAST analysis of the Bd0314 primary sequence against a non-redundant database revealed a number of closely related homologues. Included in the results were two *Bdellovibrio bacteriovorus HD100* proteins (Bd1413 and Bd1440). Bd1413 and Bd1440 share 60% ID to each other and 33% ID to Bd0314 (Figure 21). Bd1413 and Bd1440 both possess the two catalytic residues highlighted in the structure of Bd0314 and they also possess the YW-motif (Y228) associated with GlcN selectivity. There are two regions of minor disparity in the sequence alignment between Bd0314 and Bd1413/Bd1440. Firstly Bd1413/Bd1440 has a 11aa insertion between helices F and G. Secondly the short β-sheet extension (strand-3 and 4) of the core domain exists as an 8aa insertion in the primary sequence of Bd0314.

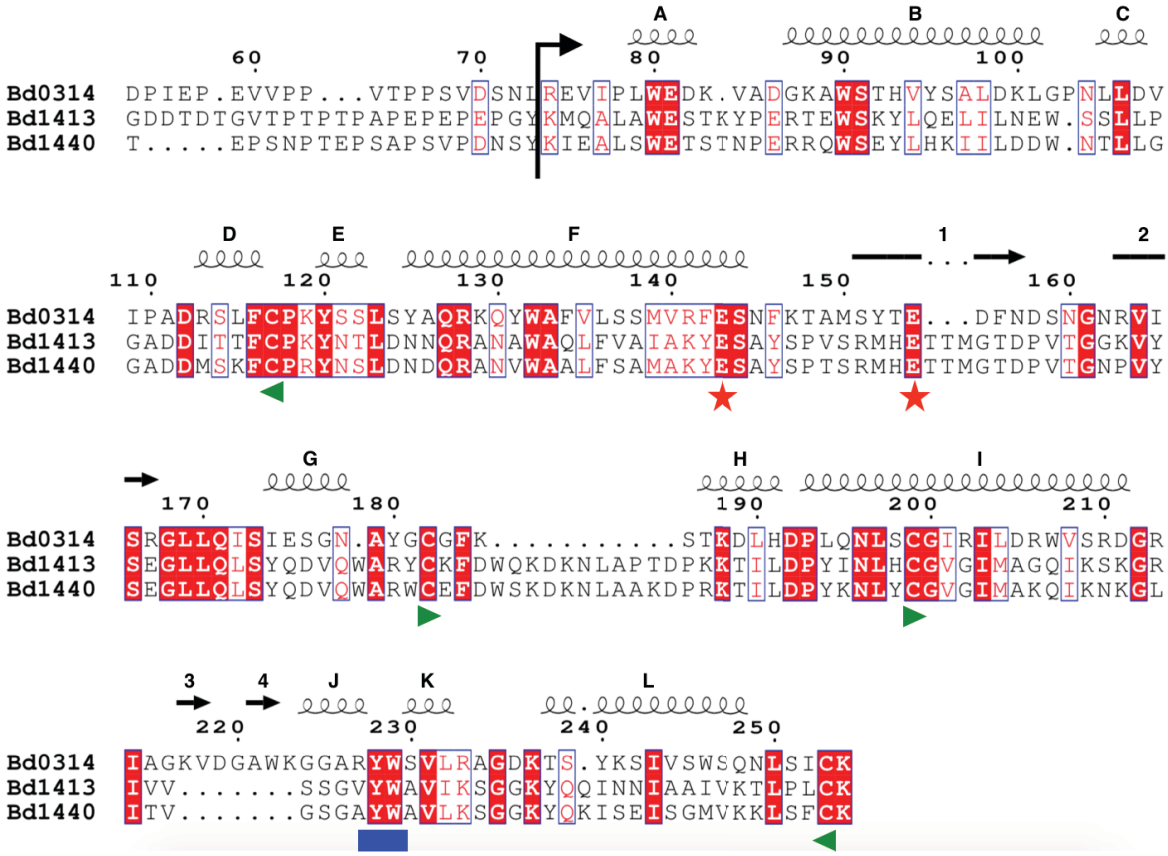


Figure 21. Bd0314 alignment with *Bdellovibrio HD100* homologues
Alignment of Bd0314 primary sequence with Bd1440 and Bd1413. Bd1440 and Bd1413 sequences are more closely related to one another (60% ID) than to Bd0314 (33% ID). Bd1440 and Bd1413 do not possess the second short β-sheet region of the α-helical domain. Bd1440 and Bd1413 also have an elongated region between helix-G and helix-H. The start of the catalytic domain is shown by a black line and arrow. The conserved cysteine residues are shown by green triangles; the direction of the green triangle represents the cysteine pairs. Red stars highlight the catalytic residues and the GlcNAc *N*-acetyl group blocking motif is shown by the blue rectangle. α-helices are represented by black hoops and β-sheets represented by black arrows. Drawn using Espript 3 [161].

BLAST analysis also revealed a number of Bd0314 homologues from other organisms that possess equivalents of both catalytic residues (E143 and E154) as well as the YW-motif (Y228). An alignment of selected homologues is shown in Figure 22. The alignment demonstrates the co-carriage of the second catalytic residue motif (TED) with the GlcN specificity motif (YW).

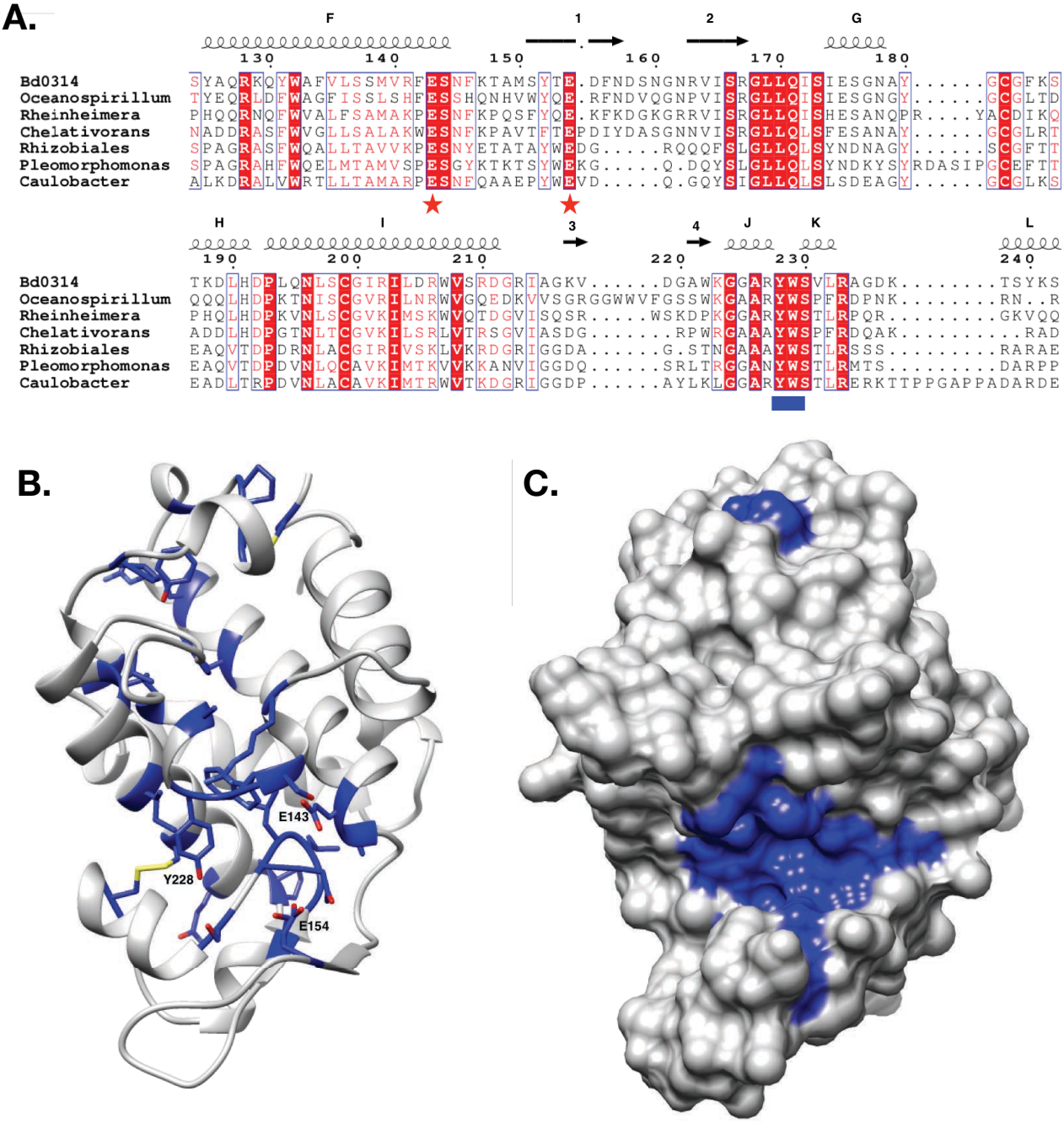


Figure 22. Bd0314 alignment and residue conservation
A. BLASTp searches reveal that homologues of Bd0314 exist in organisms other than predatory bacteria. All organisms possess CE4 deacetylases in their genomes. The presence of the second catalytic glutamate (E154) has a 100% co-carriage with the *N*-acetyl group blocking tyrosine (Y228) in these predicted *N*-deacetylation specific lysozymes. Catalytic residues are highlighted with blue stars and the *N*-deacetylation specific lysozyme motif is highlighted with a blue bar. **B.** Residue conservation (100% ID) mapped onto the secondary structure of Bd0314. **C.** Residue conservation mapped onto the surface of Bd0314.

upon the GlcN substrate. The Bd0314 activity towards the GlcN substrate is comparable to the activities of the HEWL and Mutanolysin upon acetylated substrates. This result indicates Bd0314 possesses a high specificity towards GlcNAc-deacetylated PG substrates.

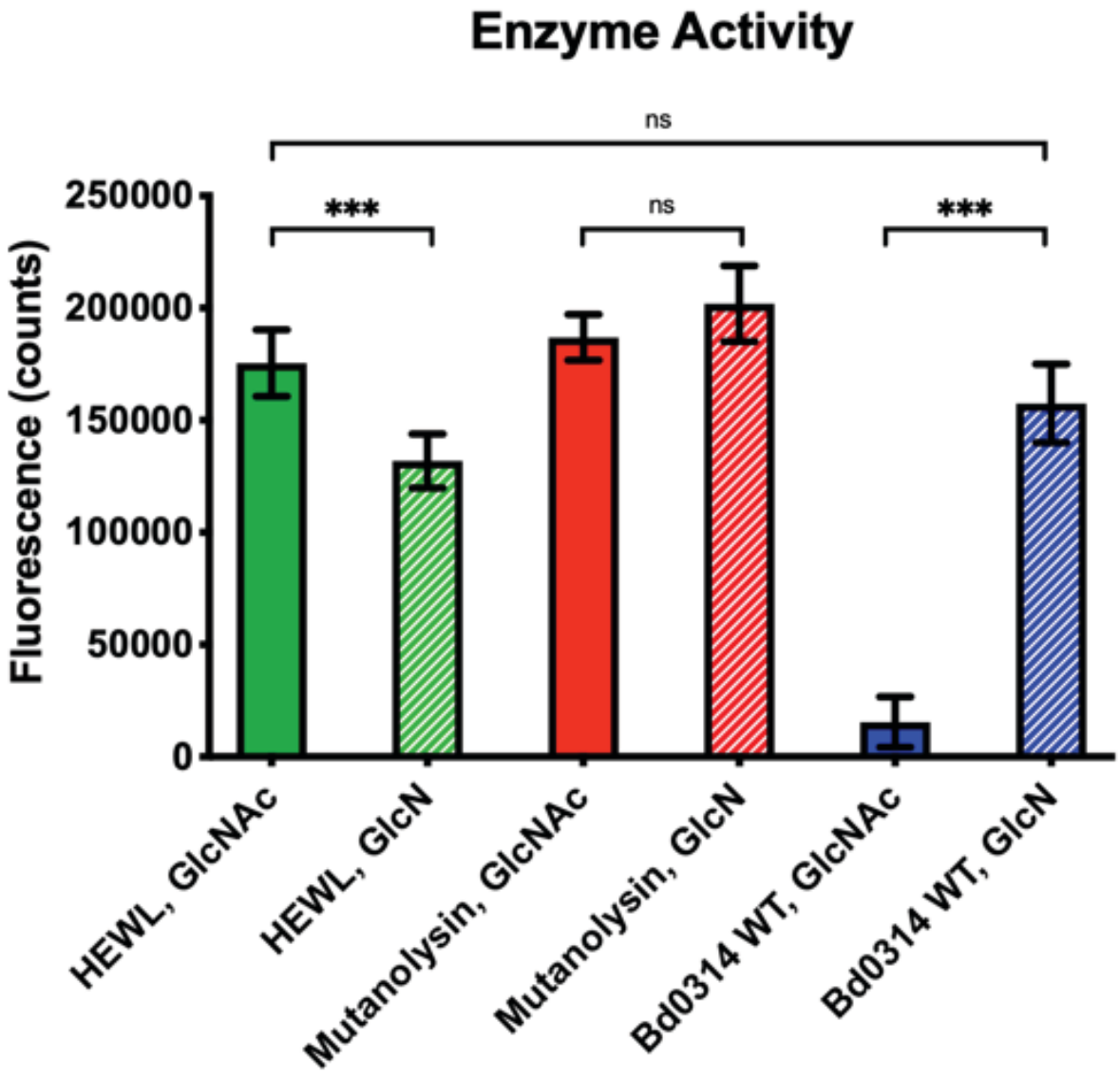


Figure 23. Bd0314 specificity
The results of the FITC-PG assay demonstrating Bd0314 specificity. Block coloured bars represent GlcNAc (acetylated PG) whilst striped bars represent GlcN (N-deacetylated PG). HEWL (Green) shows reduced activity with N-deacetylated substrate. Mutanolysin activity is similar for both substrates. Bd0314 shows specificity for the GlcN substrate. All data points were measured in at least triplicates. The error bars represent the standard deviation. Unpaired t-test was used to determine significance, *** = $p<0.0001$, ns = not significant.

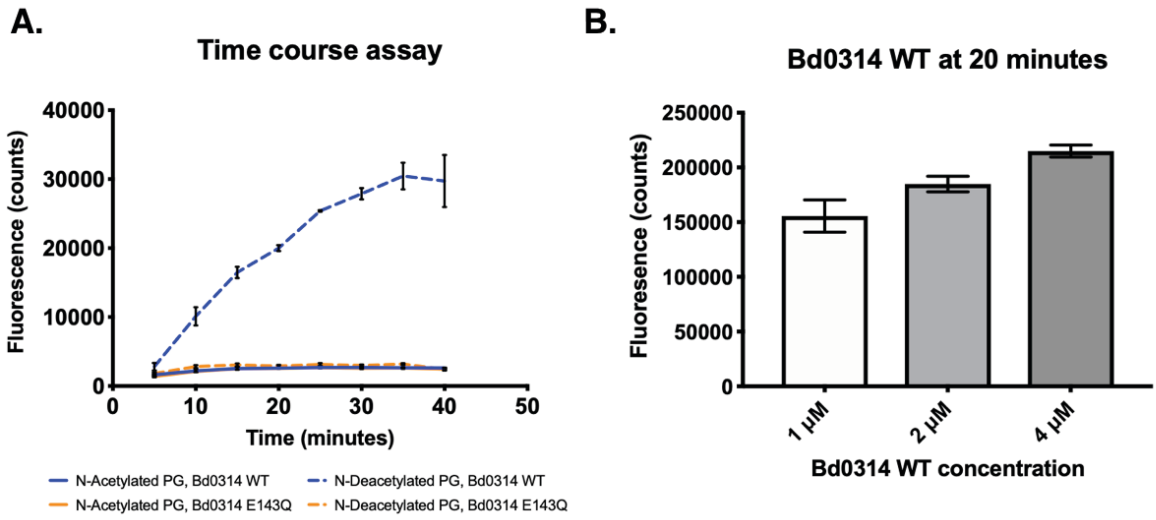


Figure 24. Time and Concentration dependent assay
A. Turnover of the FITC-PG substrate by Bd0314 is time dependent; activity is linear for the first 15 minutes. Bd0314 is only active towards the *N*-deacetylated PG substrate and shows minimal activity towards normal acetylated PG substrate. **B.** The FITC-PG assay is concentration dependent; increased concentrations of Bd0314 increased the fluorescence readings. All data points were measured in at least triplicates. The error bars represent the standard deviation.

3.2.3.2 Catalytic residues

The FITC-PG assay was used to confirm whether E143 and E154 (general acid-base pair identified from the Bd0314 structure) were required for catalysis. The glutamate residues were mutated to glutamines thereby changing the chemistry of the residue but maintaining the structure (Bd0314-E143Q and Bd0314-E154Q). Figure 25 shows that the activity of both Bd0314-E143Q and Bd0314-E154Q are greatly reduced compared to Bd0314. This finding confirms both E143 and E154 are required for catalysis.

Y228 was mutated to an alanine residue (Bd0314-Y228A) to investigate its role in substrate-binding and specificity. Bd0314-Y228A showed a reduced activity towards the GlcN substrate but no increase in activity towards the GlcNAc substrate. This

finding suggests that Y228 also participates in substrate recognition by direct interaction with the free amine group of GlcN residues (Figure 30). The retained substrate specificity of Bd0314-Y228A suggests other features also contribute to substrate selection. M150 was identified as another substrate selective residue, although we were unable to successfully express and purify a Bd0314-M150A mutant.

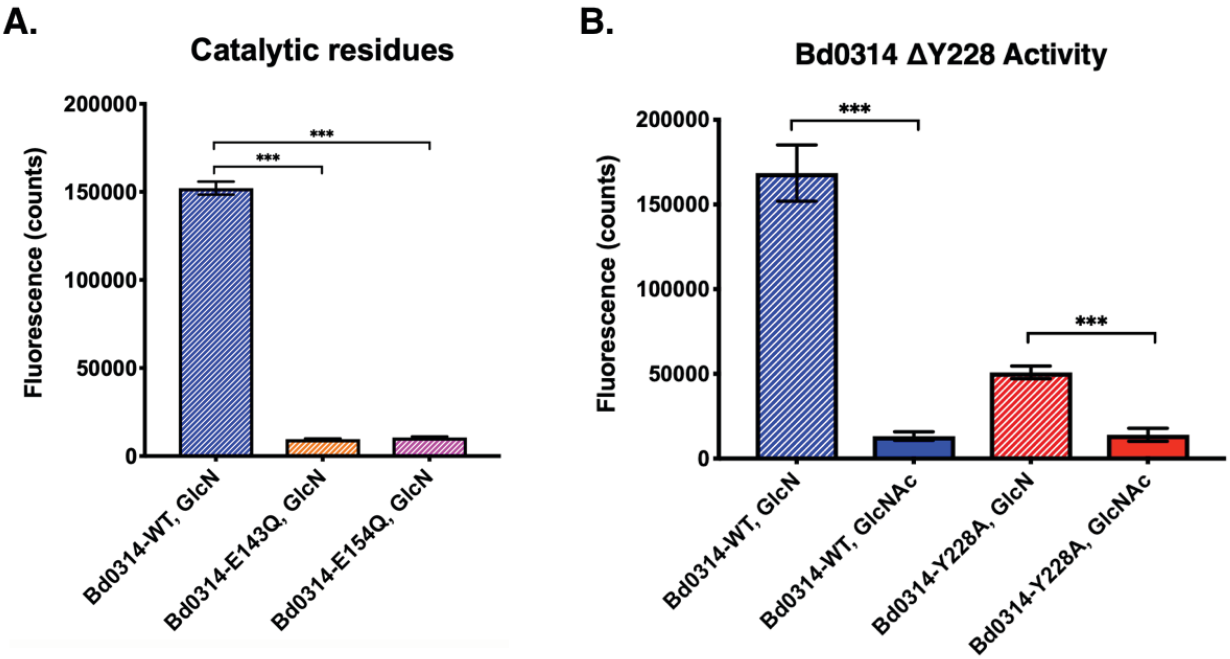


Figure 25. Bd0314 active site residues
A. Mutation of either E143 or E154 abolishes Bd0314 activity, suggesting both residues are involved in catalysis and supporting an inversion catalytic mechanism, involving two glutamate residues. **B.** Y228A mutation reduces Bd0314 activity towards *N*-deacetylated PG. This finding suggests Y228 is involved in substrate binding. Y228A activity towards GlcNAc PG does not alter. All data points were measured in at least triplicates. The error bars represent the standard deviation. Unpaired t-test was used to determine significance, *** = $p<0.0001$, ns = not significant.

3.2.3.3 Bd0314 homologue activities

Bd1440 (S75 – K266) and Bd1413 (G80 – K271) constructs were created to investigate the activities of Bd0314 *Bdellovibrio* homologues using the FITC-PG assay (Figure 26). Bd1440 and Bd1413 both possess the same high specificity towards the GlcNAc-deacetylated substrate (GlcN). Bd0314, Bd1440 and Bd1413 all liberated a similar quantity of FITC, demonstrating similar activities.

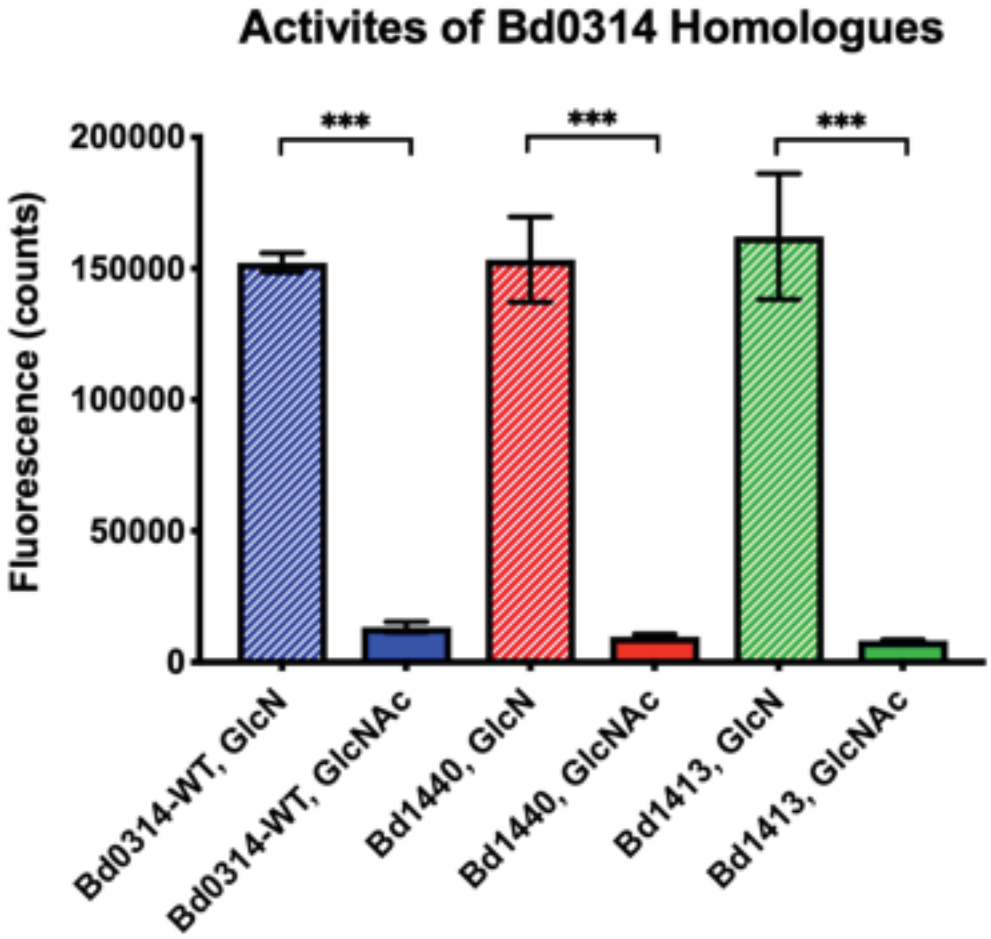


Figure 26. Bd0314 HD100 homologue activities
The activity of Bd0314 HD100 homologues (Bd1413/1440) compared to Bd0314 activity. Bd0314-WT in blue, Bd1440 in red and Bd1413 in green, with block coloured bars representing acetylated GlcNAc substrate and striped coloured bars representing *N*-deacetylated substrate (GlcN). The activities of Bd1440 and Bd1413 also demonstrate substrate specificity for GlcN substrate and have minimal activity with GlcNAc substrate. Bd1413/1440 have a similar activity as Bd0314. All data points were measured in at least triplicates. The error bars represent the standard deviation. Unpaired t-test was used to determine significance, *** = $p<0.0001$, ns = not significant.

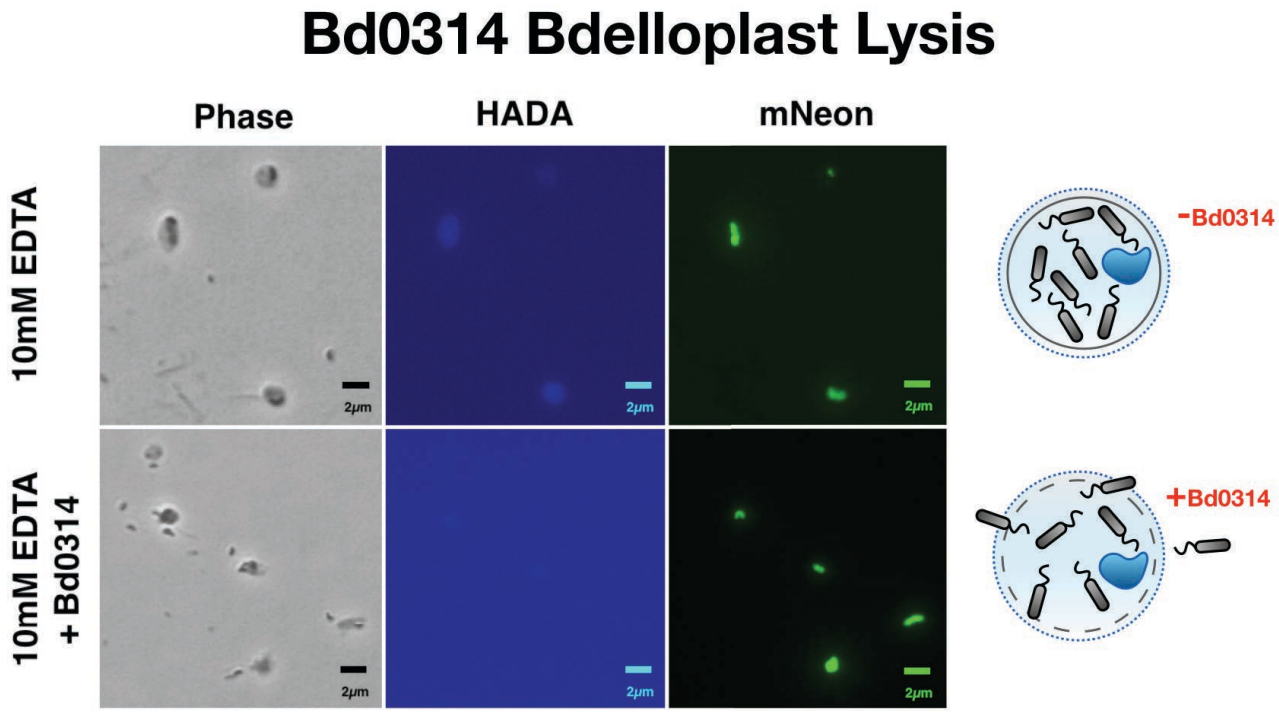


Figure 27. Bd0314 Bdelloplast Lysis
E. coli cells were incubated with *Bdellovibrio bacteriovorus* HD100 cells resulting in >99% prey rounding into bdelloplasts. Bdelloplasts were incubated for 1 hour with 10 mM EDTA +/- 10 μM Bd0314. Cells were immobilised and imaged on 1% agarose Ca/HEPES pads. HADA was used to label the cell wall. pMAL::mNeon was used to label the periplasm and cytoplasm of *E. coli.* Bdelloplasts incubated with Bd0314 lyse and release *Bdellovibrio* progeny when compared to EDTA alone control. Furthermore bdelloplasts incubated with Bd0314 have diminished levels of HADA labeled peptidoglycan, demonstrating that Bd0314 is active towards this material. pMAL::mNeon labeling allows identification of lysed bdelloplasts. The far right hand cartoon illustrates the effect Bd0314 incubation has on the bdelloplast structures.

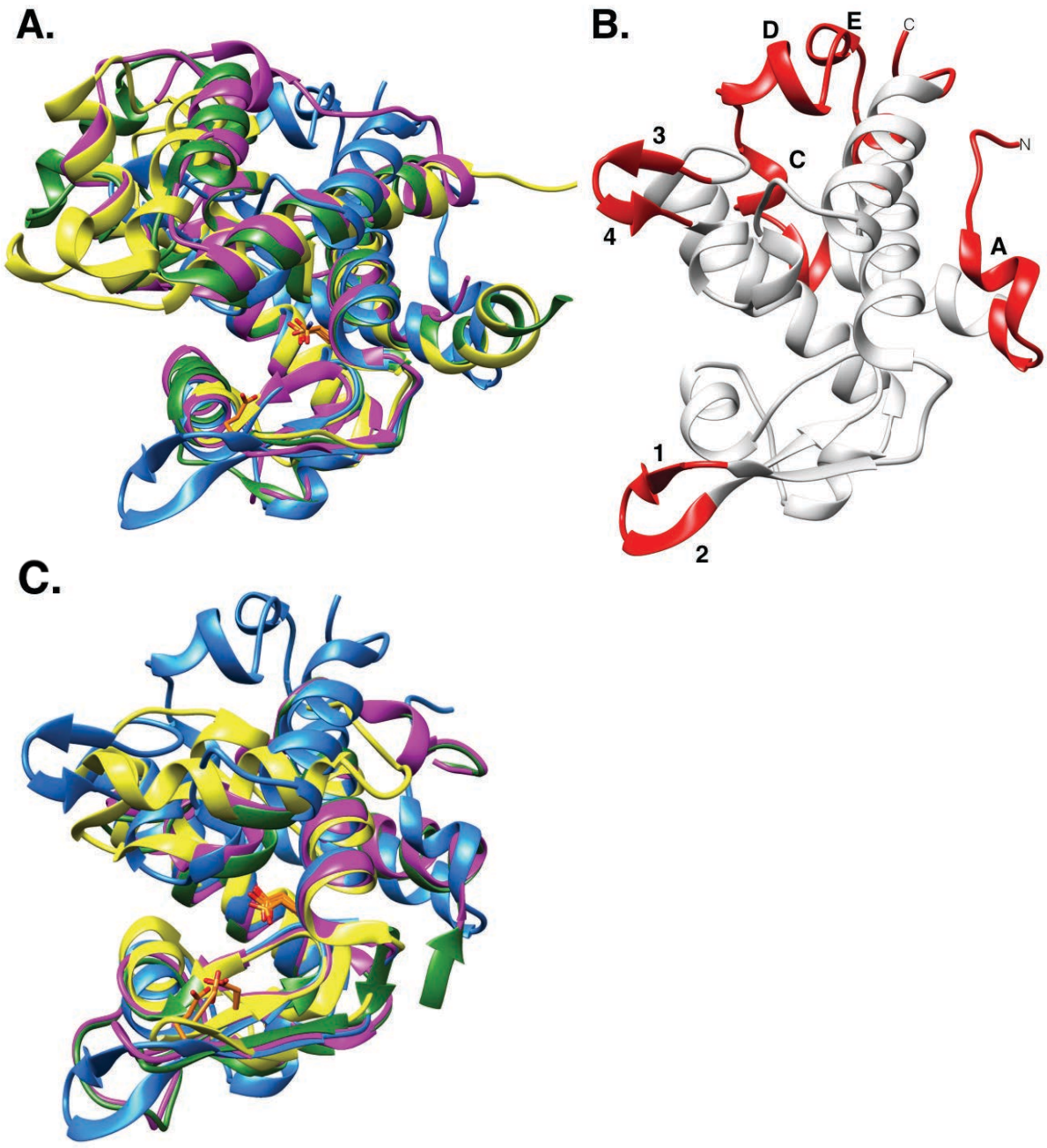


Figure 28. Bd0314 structure comparison
A. Bd0314 (blue) superimposed with the catalytic domains of lytic transglycosylases; MltC = PDB:4CFP (green, length = 5525 atoms, rmsd = 0.88Å); MltF = PDB:4P0G (yellow, length = 4162 atoms, rmsd = 0.88Å); Slt70 = PDB:1SLY (magenta, length = 5000 atoms, rmsd = 0.83Å). **B.** Unique structural features of Bd0314 are shown in red. **C.** Bd0314 (blue) superimposed with GH22 lysozymes; GH22-C HEWL = PDB:1LZS (green, length = 2474 atoms, rmsd = 1.13Å); GH22-I Tapes japonica lysozyme = PDB:2DQA (yellow, length = 2511 atoms, rmsd = 0.59Å); GH22-C Human lysozyme = PDB:5LSH (magenta, length = 1414 atoms, rmsd = 1.03Å). Superimposition was performed by least squares minimisation using the CCP4 suit superpose software (alignment of conserved secondary structure elements, helix-F, helix-I and beta-hairpin).

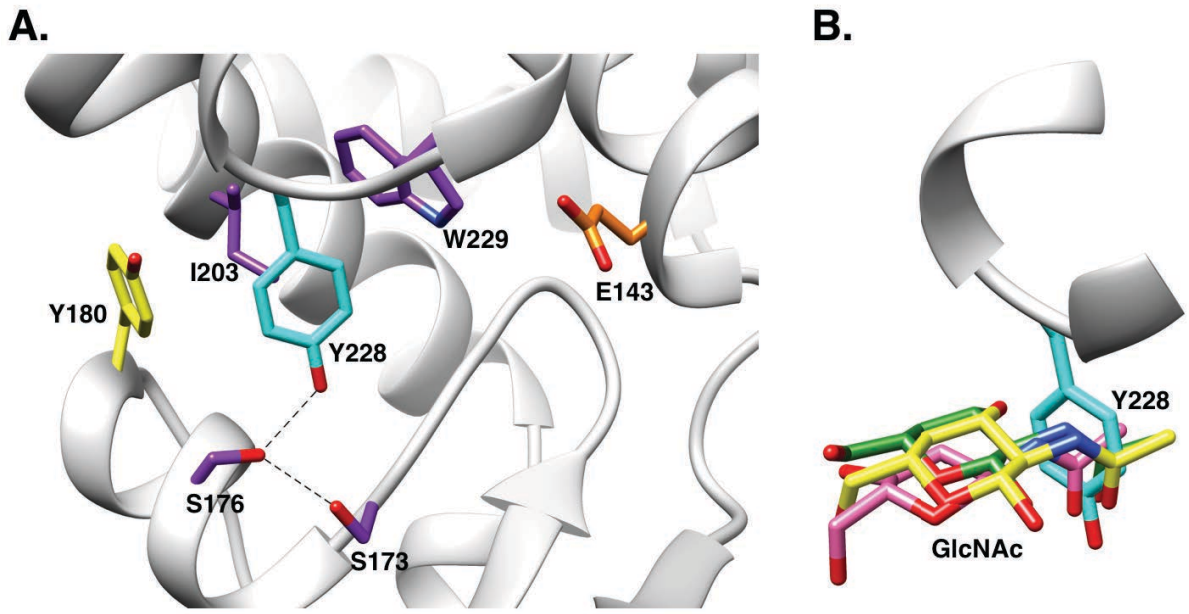


Figure 29. Bd0314 substrate binding and selection
A. Y228 (cyan) is positioned facing into the active site and held in position by a hydrogen bond with S176 (black line) and by hydrophobic interactions with W229 and I203. Y180 (yellow) may participate in substrate binding by interacting with the MurNAc hydroxyl-group. **B.** Bd0314 was superimposed with 1LZS (yellow), 3W6D (pink) and 3GXR (green) – lysozyme:ligand complex structures. The acetyl group of GlcNAc residues in these superimposed models always clashed with Y228 of Bd0314.

M150 also occupies the same space as the acetyl-group of GlcNAc residues and could also participate in selection. However, M150 is not conserved in predicted deacetylation-specific lysozymes whereas Y228 is (Figure 21). Also M150 is not locked into one conformation and in the structure of Bd0314 was observed in multiple conformations suggesting that it could easily be repositioned during substrate binding. The ligand binding models also highlight other residues that may be involved in substrate binding. For instance Y180 is positioned in close proximately to the hydroxyl group of the MurNAc residues and may make a hydrogen bond. Similarly S176 that interacts with Y228 may share the H-bond to the GlcN amine group.

Chitosanases (GH46) are a member of the lysozyme superfamily and hydrolyse the β-1-4 linkage between poly-GlcN [145]. Chitosanases are structurally divergent from

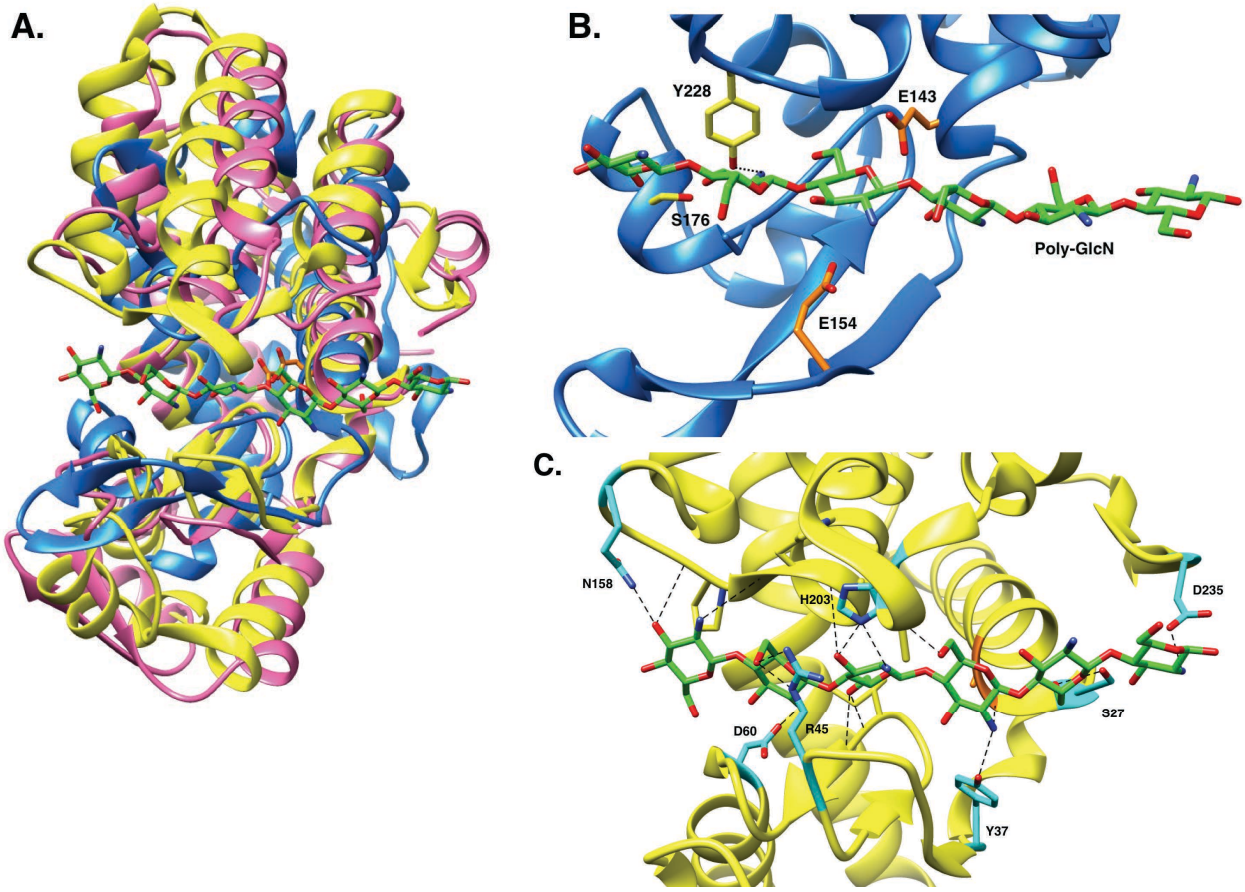


Figure 30. Comparison of Bd0314 with chitosanases.
A. Bd0314 (blue) superimposed with chitosanase-OU01 (PDB:4OLT, yellow, length = 2325 atoms, rmsd = 1.82Å) and *Bacillus circulans* chitosanase (PDB:1QGI, pink, length = 2226 atoms, rmsd = 2.57Å). **B.** Close up of the view of a poly-GlcN substrate-binding model to Bd0314 active site cleft (from superimposition with 4OLT). Y228 could potentially make a hydrogen bond contact with the amine group of GlcN residues. **C.** Substrate binding site of chitosanase-OU01 shows polar residues make hydrogen bonds to free amine group. Charged residues also interact with probable charge on GlcN. Superimposition was performed by least-squares minimisation using the CCP4 suit superpose software (alignment of conserved secondary structure elements, helix-F, helix-I and beta-hairpin).

3.3.3 Likely catalytic mechanism

The structure of Bd0314 revealed a potential secondary catalytic residue (E154) located opposite to the strictly conserved catalytic E143 residue. We demonstrated that both E154 and E143 are required for catalysis. A catalytic mechanism involving two glutamate residues is unusual for lysozyme superfamily members, which typically

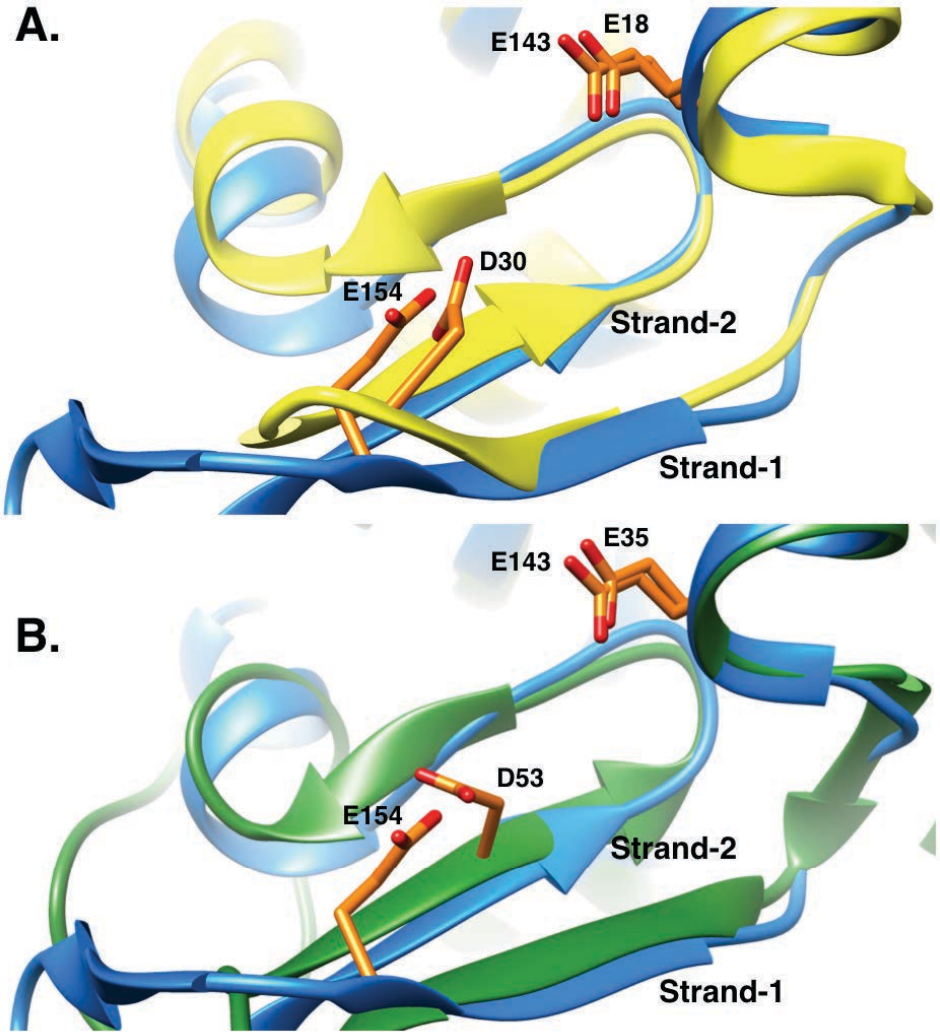


Figure 31. Bd0314 catalytic residue location.
The location of the second catalytic residue varies between lysozymes and generally dictates the catalytic mechanism, which is dependent on distance. **A.** E143 is located on the ES motif at the end of the conserved central α-helix in Bd0314 (blue). Bd0314 unusually has a second glutamate catalytic residue (E154), which is located opposite on strand-1 similar to D30 of *Tapes japonica* Lysozyme (PDB:2DQA, yellow). **B.** In HEWL (PDB:1LZS, green) the second catalytic residue, D53 is located on strand-2.

3.3.4 Bd0314 role as the exit enzyme

Early studies suggested bdelloplast lysis at the end of the predatory lifecycle was brought about by an increase in lytic enzyme activity towards the end of the predatory cycle. The lytic activity was shown to be distinct from the initial pore forming activity, which was halted after entry [45]. The enzyme(s) responsible for bdelloplast lysis was hypothesised to work specifically on the modified (lysozyme resistant) prey cell wall

of MltF to the outer membrane does not impede its activity but ensures localised control of its activity.

On the other hand the flexible tether of Bd0314 may be cleaved by one of the many proteases *Bdellovibrio* secretes into the prey cell during predation [24]. The disordered region would be an easily accessible substrate to a protease. If cutting the tether and freeing Bd0314 is required for bdelloplast lysis then the need for cleavage would add in an additional control step before committing to exiting the bdelloplast.

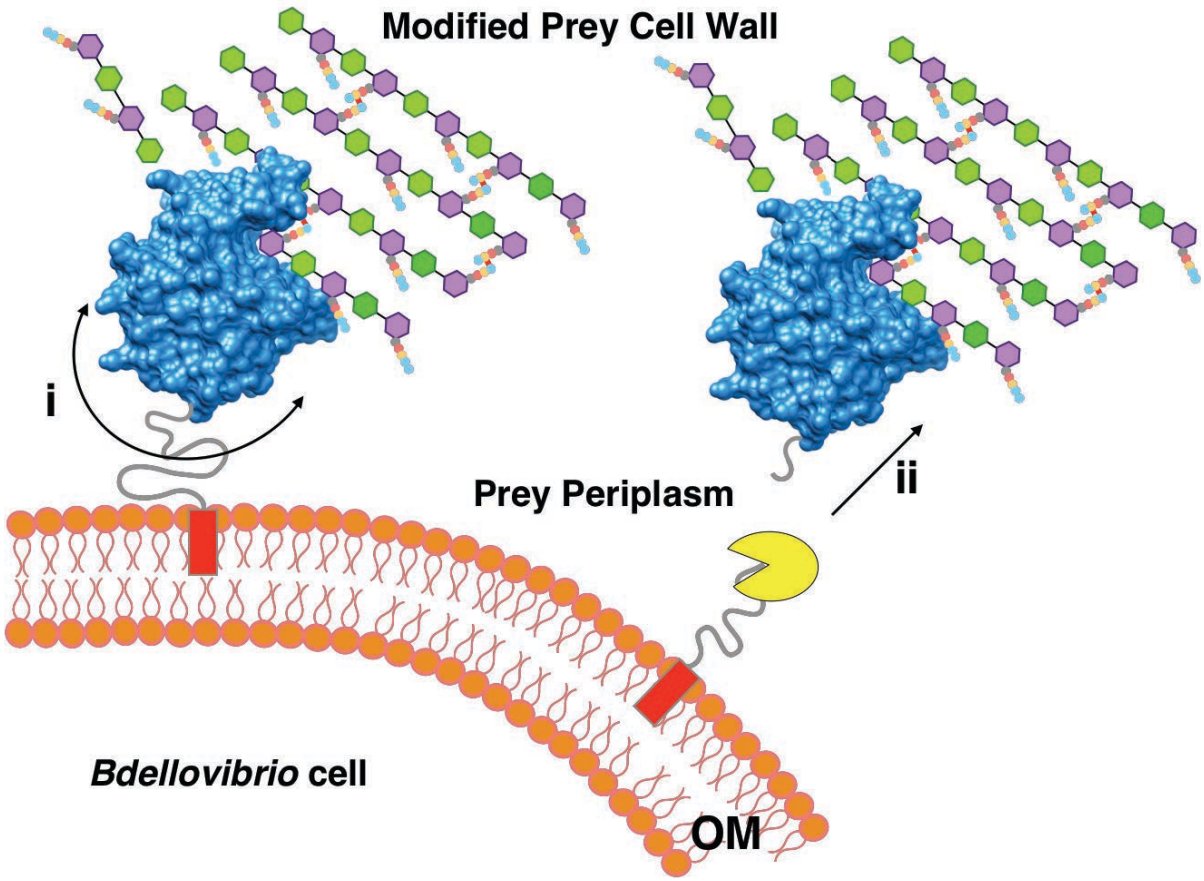


Figure 32. Role of Bd0314
Bd0314 activity is able to solubilise the modified prey cell at the end of the predatory cycle. Full-length Bd0314 possesses a signal peptide with a lipobox suggesting Bd0314 is secreted to the outer-membrane. Between the lipobox and the catalytic domain is a 47aa region of disorder. The disordered region may act as a tether, retaining Bd0314 to the outer membrane for localised prey cell wall degradation (i). Alternatively the tether may be a target of protease action and cleavage of the protease accessible disordered region would release Bd0314 into the bdelloplast (ii). The lipobox is represented by the red box. The disordered region of Bd0314 is illustrated by the grey curved line. Bd0314 is shown in blue and the prey cell wall is illustrated as in Figure 7 (green - GlcNAc/GlcN residues and purple – MurNAc). Potential protease cleavage is shown by the yellow shape.

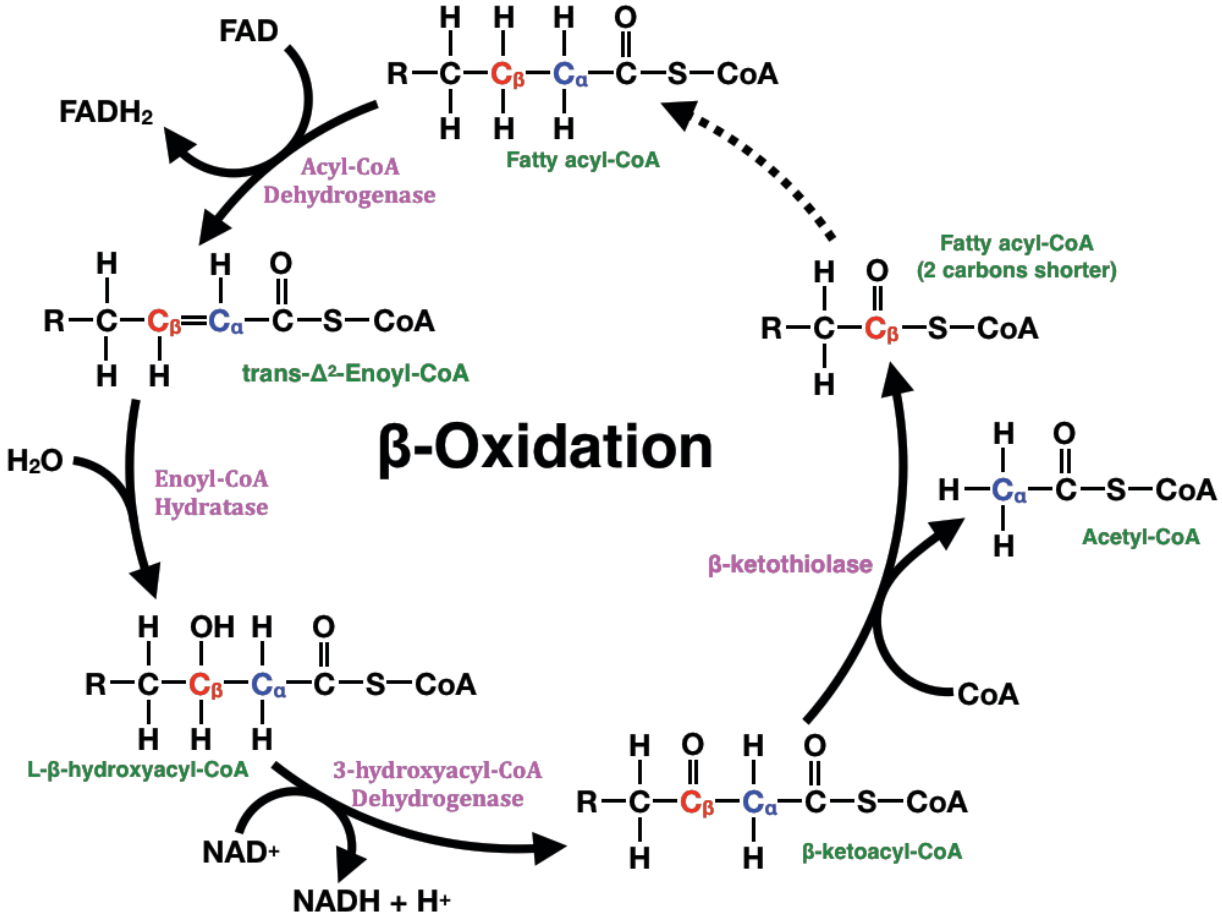


Figure 33. β-oxidation metabolic pathway
β-oxidation is a catabolic process, involving the **c**hemical modification of fatty acyl-CoA substrates at the C_α and C_β positions. Acyl-CoA dehydrogenases catalyse the initial dehydrogenation reaction of fatty acyl-CoA to form trans-Δ^2-enoyl-CoA. An enoyl-CoA hydratase converts trans-Δ^2-enoyl-CoA into L-β-hydroxyacyl-CoA. Then a 3-hydroxyacyl-CoA dehydrogenase produces β-ketoacyl-CoA with the use of an NAD cofactor. A thiolase catalyses the final thiolysis step by inserting the thiol group of CoA between the C_α and C_β positions. The β-oxidation cycle can be repeated until the acyl-chain length is reduced to the final two carbons.

4.1.3 β-oxidation regulation in bacteria

Fatty acids are an essential component of bacterial membranes and a source of metabolic energy. Regulation of fatty acid biosynthesis and degradation is tightly controlled to ensure membrane lipid homeostasis is maintained. The majority of research into β-oxidation pathway regulation has been carried out in *E. coli* [177].

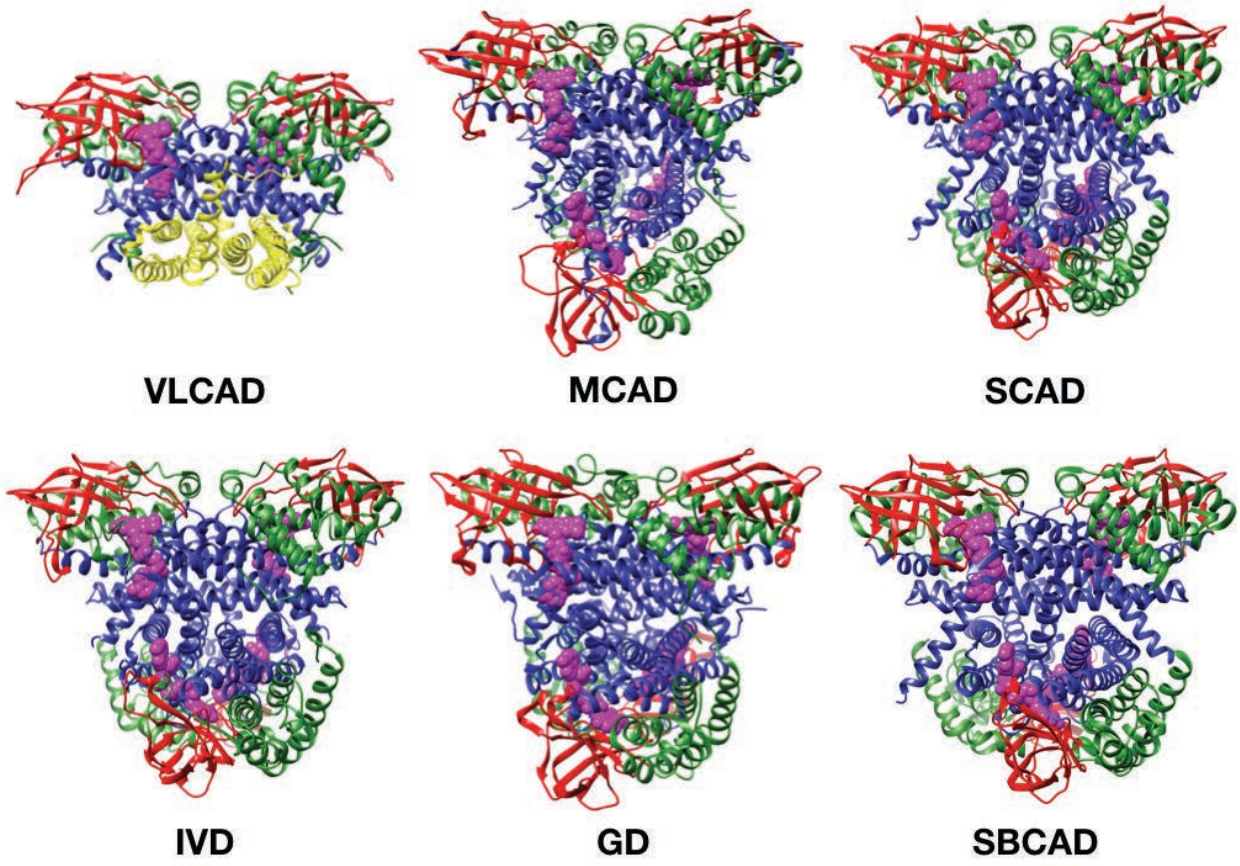


Figure 34. Acyl-CoA dehydrogenases
This Figure illustrates the three-dimensional structures of different members of the ACAD family. The structures are coloured based on their domains: N-terminal α-helical domain (green), central β-sheet domain (red), C-terminal α-helical domain (blue) and finally C-terminal extension domain (yellow) the FAD molecules are coloured magenta. ACADs form tetramers with the exception of VLCAD that forms a dimer. VLCAD – PDB:3B96, MCAD – PDB:3MDE, SCAD – PDB:2VIG, IVD – PDB:1IVH, GD – PDB:3MPI, SBCAD – PDB:2JIF.

ACAD subunits fold into three distinct domains: a N-terminal domain (green) and a C-terminal domain (blue) that are mainly comprised of α-helices and a central domain (red) that is formed of two orthogonal β-sheets (Figure 35.A). The C-terminal domain is extensively involved in forming the tetramer and dimer interfaces. The tetrameric structure is composed of a dimer of dimers in a tetrahedral arrangement. The dimer interface is extensive and includes shared FAD binding sites, whereas the tetramer interface mainly involves helix-helix interactions [180]. VLCAD is an exceptional case, which forms an identical dimer interface to MCAD. However, an additional C-terminal

domain (+180 residue) prevents the formation of the conventional tetramer interface. The additional C-terminal domain adopts a fold identical to the N and C-terminal α-helical domains (tetramer interface contacts) of the opposing MCAD dimer, hence the additional C-terminal domain mimics the tetramer interface contacts.

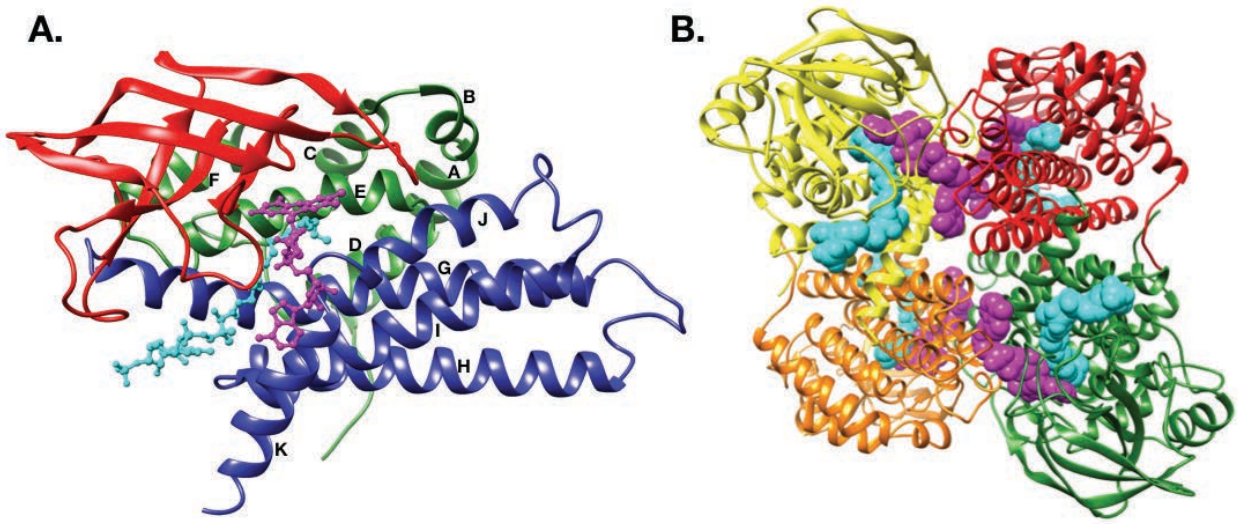


Figure 35. MCAD structure
MCAD is the most well-characterised ACAD, structurally and biochemically. **A.** MCAD can be split into three domains: an N-terminal α-helical domain (green), central β-sheet domain (red) and C-terminal α-helical domain (blue) that provides the tetramer interface. Each monomer binds an FAD molecule (magenta) and can bind a single C8-CoA ligand(cyan). Helices are labelled according to Pig MCAD, for reference to the text. **B.** MCAD forms a tetramer complex in the form of a dimer of dimers. Red + Yellow and Green + Orange monomers make a single dimer. The FAD molecules bind between dimers.

4.1.6 Substrate binding

The FAD molecule is located at the dimer interface between two monomers. The isoalloxazine ring is located in the crevice between the two α-helical domains and the β-sheet domain of one monomer, whilst the adenosine group is located at the dimer interface (Figure 35).

The fatty acyl-CoA substrate binds in the tunnel between the three domains of a single monomer. The adenosine group binds at the surface in a groove leading to the opening

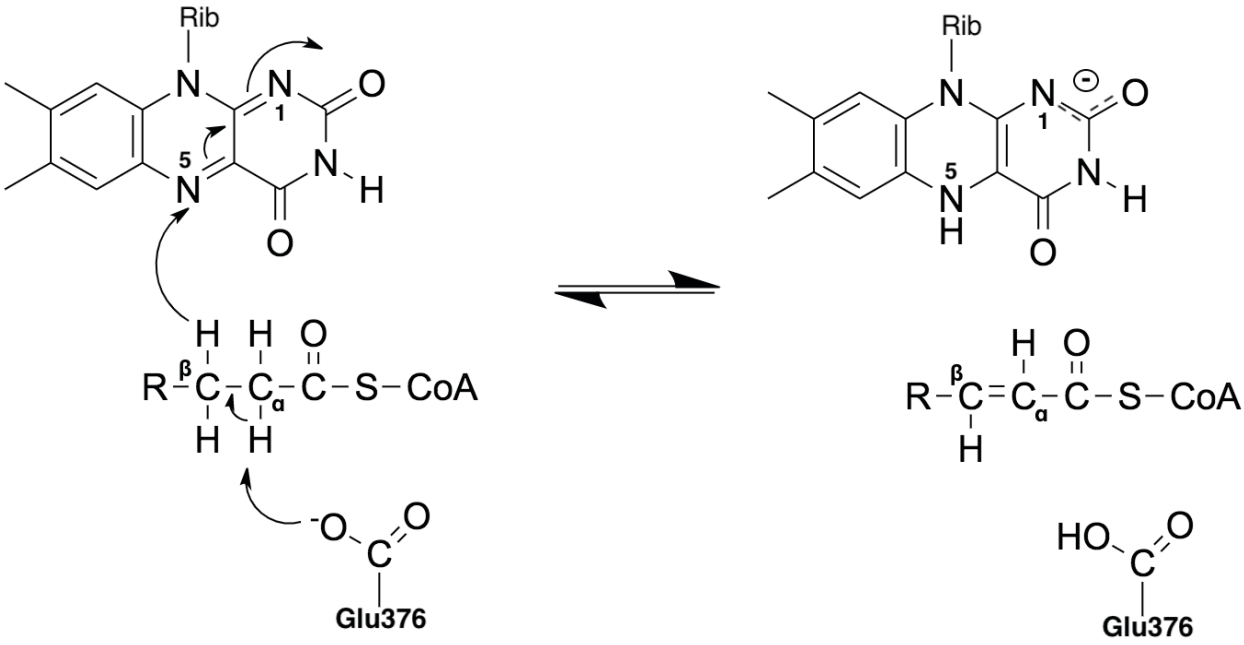


Figure 36. ACAD mechanism
This figure illustrates the catalytic mechanism of ACADs, exemplified by pig MCAD. The catalytic E376 abstracts a proton from the C_α, which synchronously causes the hydride transfer from C_β to be transferred to the N5 position of FAD as a hydride. $FADH_2$ is deoxidised to FAD in two $1e^-$ transfers to a flavin carrying ETF.

4.1.8 ETF and its interactions with ACADs

The reduced FAD molecule is re-oxidised through two sequential re-oxidation reactions with an electron transfer flavoprotein (ETF). ETFs are involved in accepting electrons from all acyl-CoA dehydrogenases shuttling them to the membrane bound ETF-ubiquinone oxidoreductase of the respiratory chain [185, 189] (Figure 37).

ETFs are unusually heterodimers, composed of α and β subunits with a single FAD molecule and a redox inactive AMP molecule [190]. The interflavin redox reaction between the FAD molecules of ACAD and ETF occurs through two successive one electron transfer steps. The FAD molecules are buried in both enzymes thus a direct hydride exchange would not be feasible. It is believed that ACADs and ETF interact weakly with each other and ETF undergoes a large dynamic conformational change in

order to interact with ACAD. A single ETF α-subunit:MCAD complex structure has been published revealing the ETF α-subunit interacts with the N-terminal α-helical domain of MCAD [191]. The interaction mainly involves hydrophobic interactions. The location of the FAD molecule in the β subunit of ETF in the resting state implies a large conformational change is required to correctly position the FAD molecule in the electron transferring state [191]. Conformational change is supported by SAXS results that confirm ETFs sample a range of conformations in solution [192].

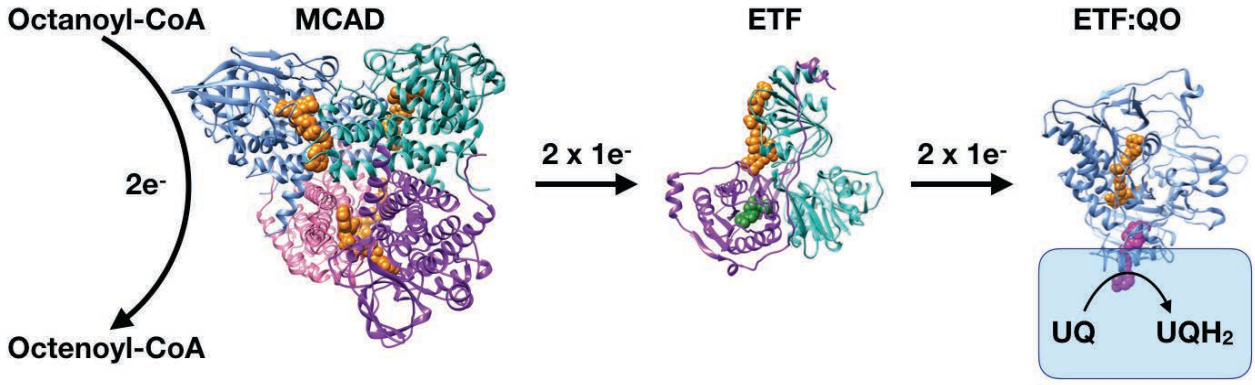


Figure 37. Electron-transfer by ETF
Octanoyl-CoA is dehydrogenated into Octenoyl-CoA transferring two electrons to the FAD of MCAD in the first step of β-Oxidation. $FADH_2$ is re-oxidised to FAD by two successive $1e^-$ transfers to an ETF. ETF then transfers the single e^- to ETF:QO which in turn transfers the electrons to ubiquitin pool. ETF can accept electron transfer from various ACAD family members. Structures: MCAD – 1T9G, ETF – 1EFP, ETF-QO - 2GMH. FAD molecules are coloured in orange, AMP molecule in green and Ubiquinone molecule in magenta.

4.1.9 ACAD homologues including acyl-CoA Oxidases (ACOs)

Homologues of the ACAD fold exist that possess divergent functions. Most divergent homologues are acyl-CoA oxidases of peroxisomal origin [180, 193], while others are involved in steroid side chain cleavage [194], antibiotic synthesis [195] and response to environmental stresses [196-198].

ACOs are highly similar to ACADs and catalyse the initial dehydrogenation reaction of β-oxidation in peroxisomes. ACOs, however, do not require an ETF to re-oxidise; instead ACO use molecular oxygen, producing hydrogen peroxide. The structures of ACOs share a common ACAD fold, more similar to VLCAD. ACOs form a soluble dimer and have an additional C-terminal α-helical domain (C-I domain, yellow) that mimics the tetramer interface of MCADs. In addition ACOs have an α-helical extension in the C-terminal domain (C-II domain, orange) that overlaps with the central β-sheet domain and reduces the accessibility of FAD to bulk solvent. The overlapping helix is also believed to be involved in blocking ETF docking and potentially changing the redox environment of the FAD molecule [180].

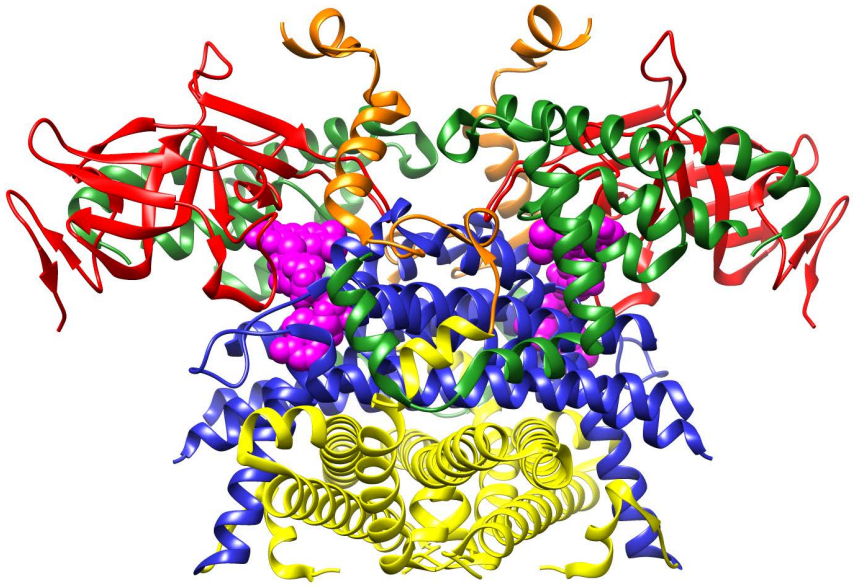


Figure 38. Structure of ACO
The structure of ACO (PDB:2DDH) dimer is coloured according ACAD domains. N-terminal domain (green), central domain (red), C-terminal domain (blue). ACOs contain an addition C-terminal α-helical domain (yellow) that is similar to the C-terminal domain in VLCAD, which mimics the MCAD tetramer interface. Additionally ACOs have an elongated α-helical region (orange) that caps the FAD accessible groove and prevents ETF access to FAD.

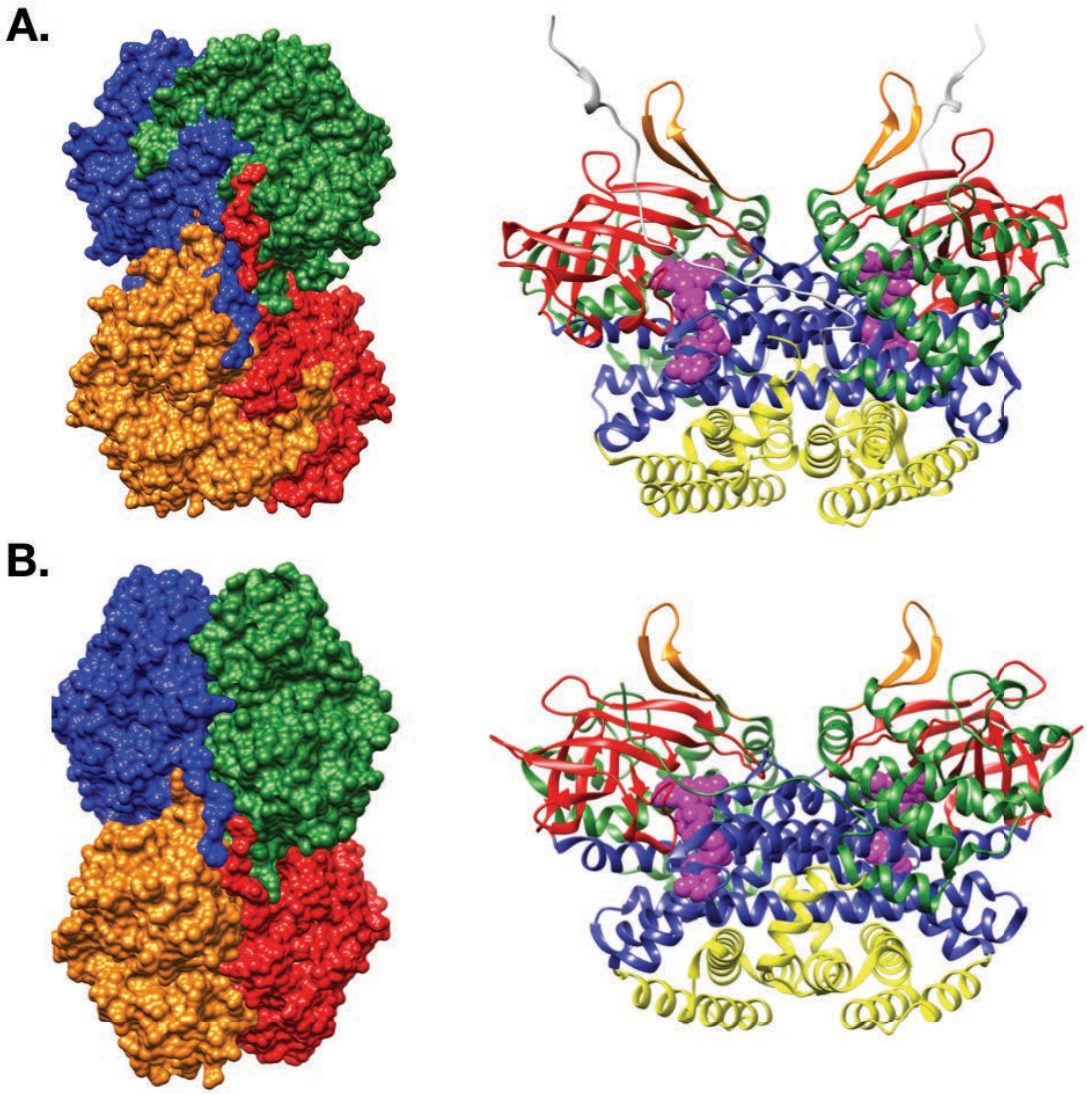


Figure 39. Structure of divergent ACAD homologues
A. ACDH-11 forms a tetramer in the form of a dimer of dimers (blue:green + red:orange), which is organised differently to ACAD tetramers. The tetramer interface is provided by a β-sheet loop in the N-terminal domain (orange). The structure of ACDH-11 domains are coloured similarly to ACAD structures, N-terminal domain (green), central domain (red), C-terminal domain (blue) and C-terminal extension domain (yellow). ACDH-11 is more closely related to the structure of VLCAD. **B.** The structure of AidB is highly similar to ACDH-11. AidB also forms a similar tetramer interface to ACDH-11, facilitated by the N-terminal domain β-sheet loop (orange).

4.2 Results

4.2.1 Cloning, Expression and Purification of Bd2924

Full-length Bd2924 (residues 1 to 505) was cloned into a pET28a vector and expressed in *E. coli* Bl21* (DE3) cells. Bd2924 was purified by IMAC to >95% purity (Figure 40). Purified Bd2924 was bright yellow in solution, indicating the presence of co-purified FAD cofactor.

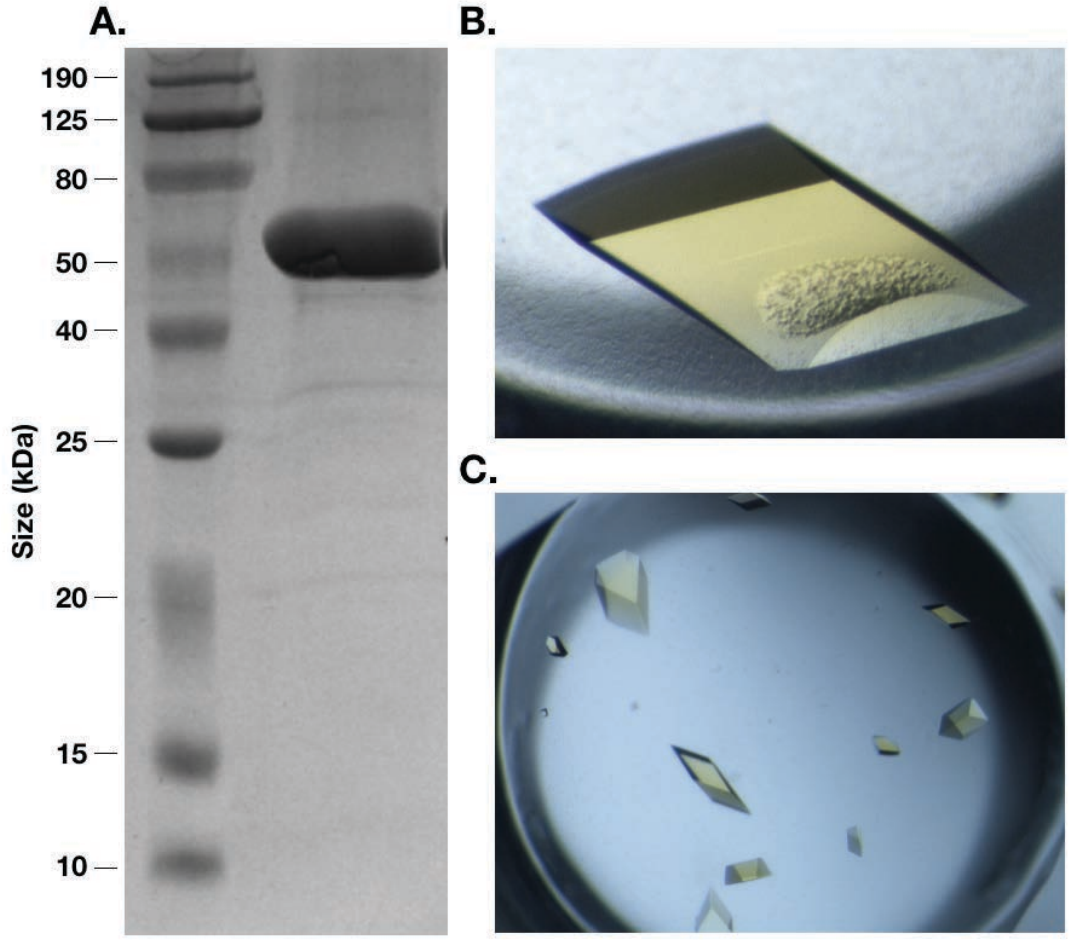


Figure 40. Bd2924 purification and crystallisation
A. Bd2924 was purified to >95% purity, judged by SDS-PAGE. **B.** Image of Bd2924 crystal **C.** Image of Bd2924 crystals

4.2.2 Obtaining crystals and data collection

Large bright yellow Bd2924 crystals were grown, which were used for data collection. The structure of Bd2924 was originally solved by molecular replacement, using PDB:4Y9J as the search model (see Section 2.4.5). We determined the structure of Bd2924 (residues 3-505) in complex with the cofactor FAD, both with and without a C10 length acetyl-CoA thioester ligand (C10-CoA) bound, to 1.51Å and 1.87Å respectively.

4.2.3 Protein Structure / topology

Overall the tertiary structure of Bd2924 is a homotetramer consisting of a dimer of dimers. Bd2924 shares a common ACAD fold, more closely related to divergent ACAD homologues such as AidB and ACDH-11. The homodimer packs across a 2-fold axis to create a likely biologically relevant crystallographic homotetramer, where the CD dimer is identical to the AB dimer. SEC confirms Bd2924 also exists as a tetramer in solution (Figure 41).

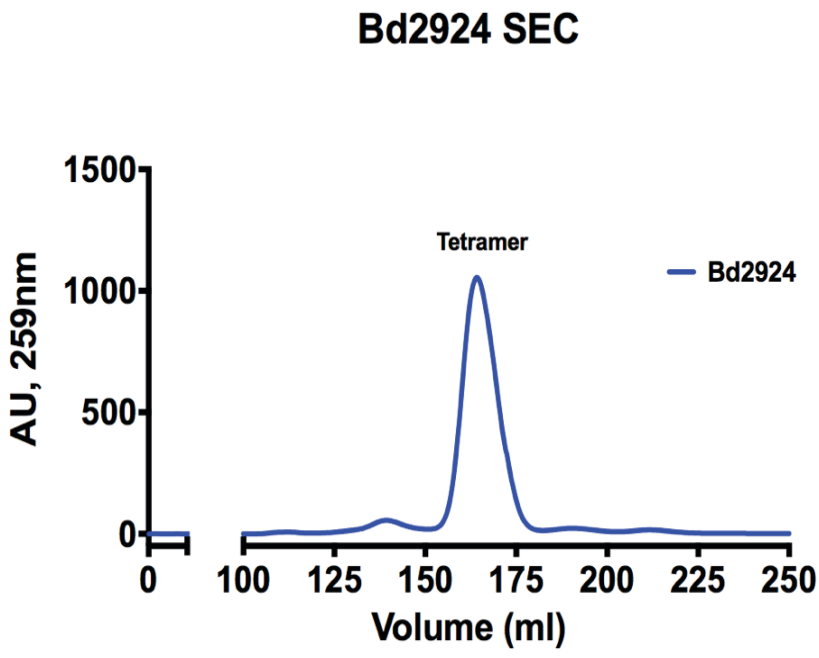


Figure 41. Bd2924 SEC
Bd2924 elutes as a single peak at 165 ml, confirming Bd2924 exists as a tetramer in solution.

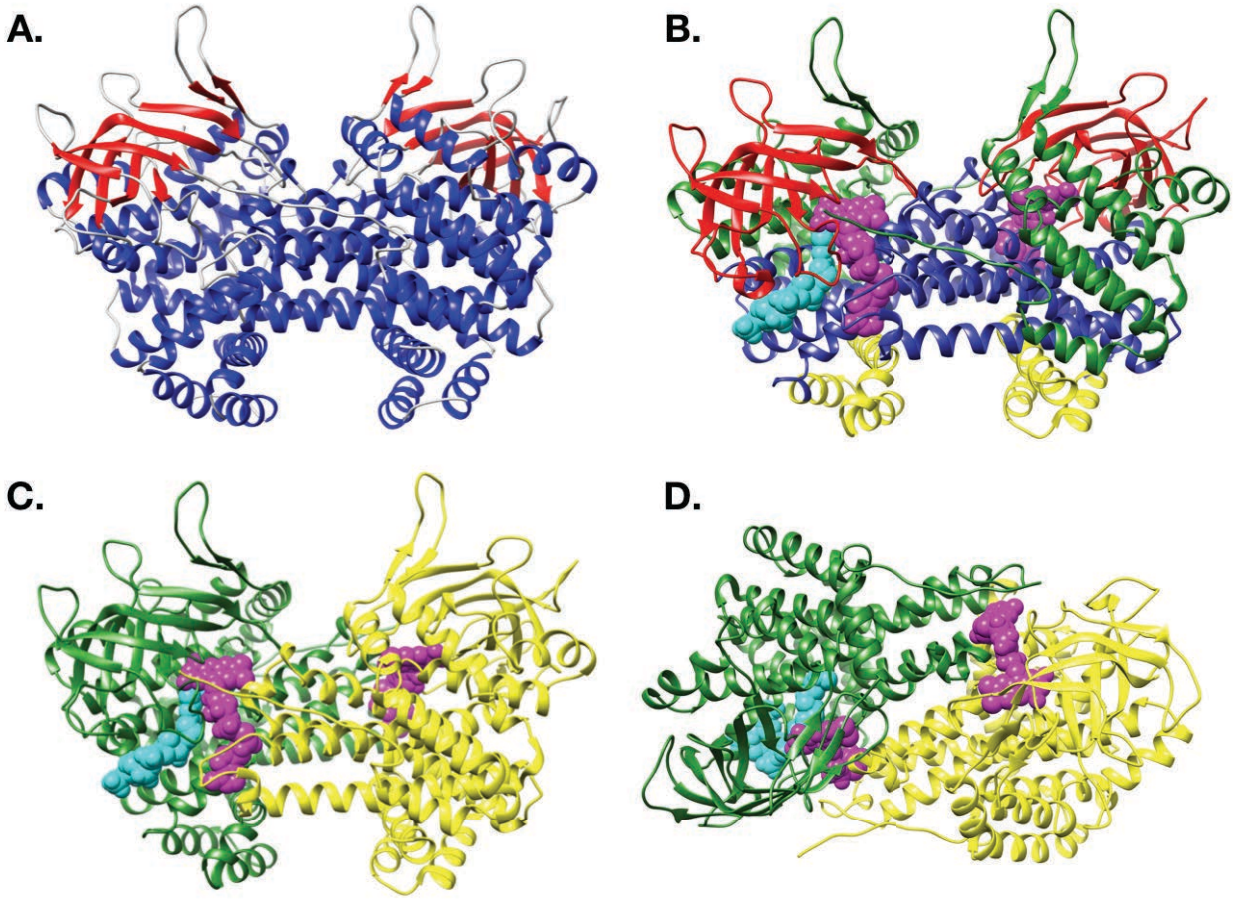


Figure 42. Structure of Bd2924
A. The secondary structure of Bd2924 dimer is mainly composed of α-helices (blue), with a region of β-strands (red). **B.** Bd2924 can be divided into four domains, similarly to other ACADs. Bd2924 has an N-terminal α-helical domain (green), a central β-sheet domain (red), a C-terminal α-helical domain (blue) and finally a C-terminal α-helical extension domain (yellow) similar to VLCAD. **C.** Bd2924 forms a tightly interacting dimer, shown here as a green monomer and yellow monomer. FAD (magenta) binds at the dimer interface. The C10-CoA binding site (Cyan) is found is located in the tunnel running beneath the FAD molecule. **D.** A 90° rotation of Bd2924 showing the dimer interface from another angle.

Bd2924 can be divided into four domains based on other ACAD structures. Bd2924 has an N-terminal α-helical domain that is interrupted by a single β-sheet projection (residues 3-168), a central β-sheet domain formed by two orthogonal four-stranded antiparallel β-sheets (residues 169-285), a central α-helical domain (residues 286-449) and a short C-terminal α-helical domain (residues 450-505). A single FAD molecule binds per monomer in a crevice at the dimer interface. The N-terminal β-sheet projection and the C-terminal α-helical domain create two large grooves in the Bd2914

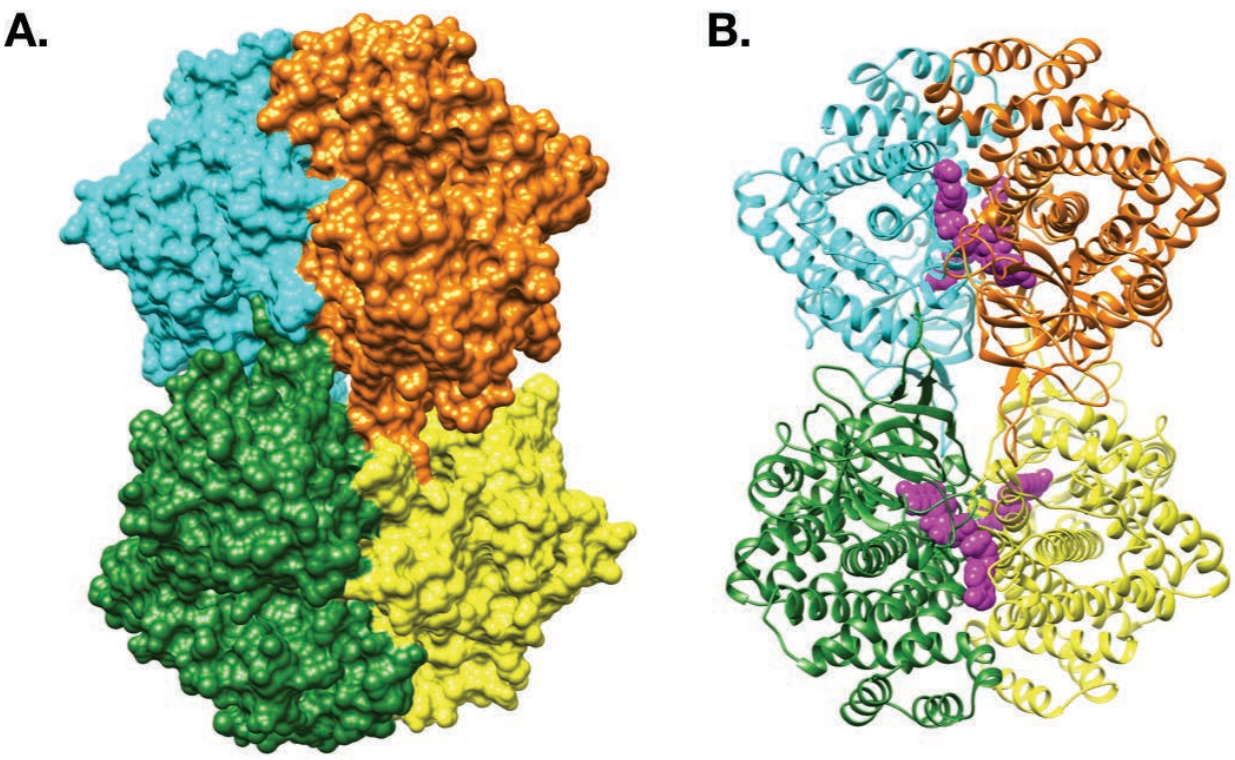


Figure 43. Bd2924 tetramer complex
Bd2924 forms a tetramer complex formed by a dimer of dimers. The tetramer interface is provided by a β-sheet loop located as an insertion into the N-terminal domain. The tetramer complex surface is shown in (**A**) and in ribbon form (**B**), which shows the location of FAD molecules (magenta).

4.2.5 Ligand Binding

4.2.5.1 FAD

FAD binds at the dimer interface and provides structural stability to the protein besides being a required catalytic cofactor. Each monomer binds a single FAD molecule in a crevice created at the dimer interface. The isoalloxazine group of FAD is located in a groove created between the N-terminal α-helical domain, the β-sheet domain and the central α-helical domain of a single monomer with a solvent accessible tunnel running beneath. The ADP region of FAD forms multiple hydrogen bonds with both chains of the AB dimer as well as multiple non-bonded interactions. Figure 44 below depicts the hydrogen-bonding network associated with FAD binding to Bd2924. The FMN region

makes H-bonds with Q171, M173 and T174 of Chain A and G406 of Chain B. The di-phosphate group makes the most intra-dimer contacts including: S180 and G179 of chain A and G406 and R324. The AMP group makes less extensive interactions: E402 and T431 H-bond with the ribose moiety whereas the adenine moiety is stabilised solely through non-bonded interactions.

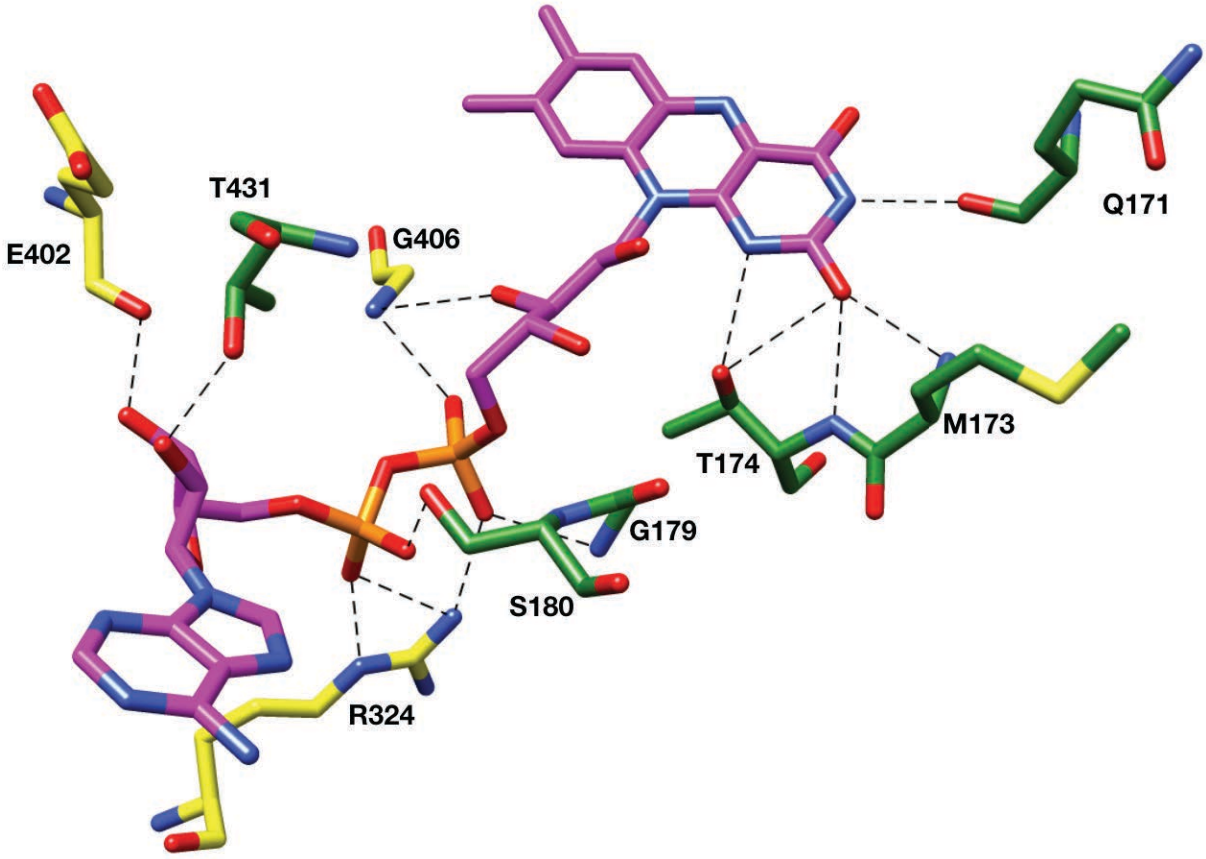


Figure 44. FAD hydrogen bonding network
FAD binds in the pocket at the dimer interface and held in place by numerous non-bonding contacts and a network of hydrogen bonds from residues from both monomers of the dimer. Residues with green coloured carbon atoms are from Chain A, whereas those in yellow are from Chain B (of an AB dimer). Hydrogen bonds are depicted by black dashed lines.

4.2.5.2 C10-CoA

A Bd2924:C10-CoA complex was solved in order to understand substrate binding. This was achieved by crystal soaking C10 and C12 chain length acyl-CoA thioesters into Bd2924 crystals. A chain length of C10-12 was chosen because the Bd2924 substrate-binding tunnel most closely resembles that of ACDH-11, which co-purified and co-crystallised with a C11-CoA thioester ligand. The structure of Bd2924 in complex with C10-CoA fatty acid revealed the interactions Bd2924 makes with substrate. In the Bd2924:C10-CoA complex structure only a single C10-CoA molecule can be confidently placed per dimer, due to weak electron density in the $2F_o$-F_c map for the second molecule.

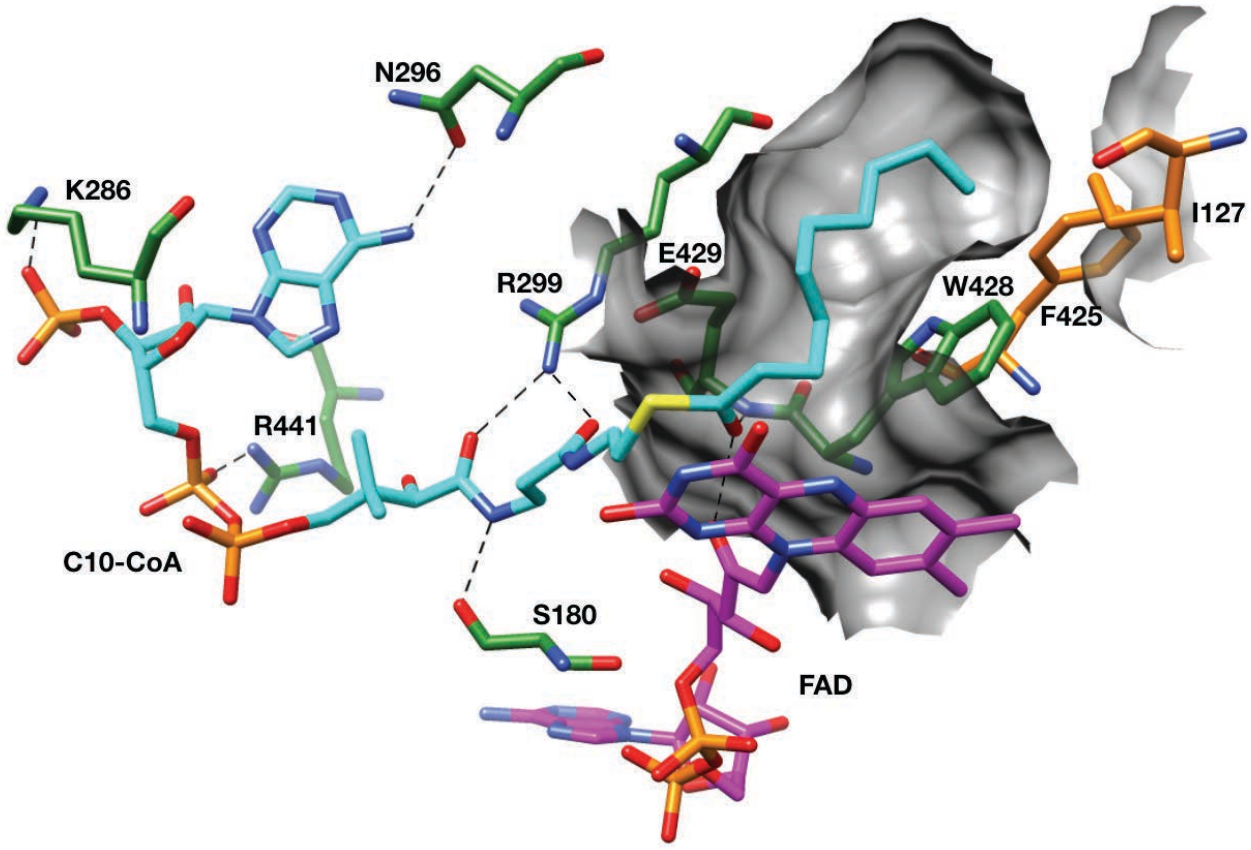


Figure 45. C10-CoA binding
The structure of Bd2924 has been solved in complex with a C10-CoA ligand (cyan). A number of hydrogen bonds (8) hold C10-CoA in place. An important hydrogen bond between the thioester carbonyl group and the backbone NH of E429 and the FAD-2'OH correctly orientates the substrate for catalysis. The depth of the substrate tunnel is limited by the sidechains of I127 and F428 (orange). Grey represents surface of pocket.

Besides MST binding studies we also attempted to obtain a Bd2924:CdG complex through X-ray crystallography experiments. Co-crystallisation and ligand soaking experiments, with concentrations of CdG up to 2 mM, were carried out with little success. The failed attempts to determine the location of a binding site via crystallography alongside the relatively weak K_D obtained through our binding assay and no conventional nucleotide-binding sites in the structure, suggest CdG may bind to Bd2924 non-specifically.

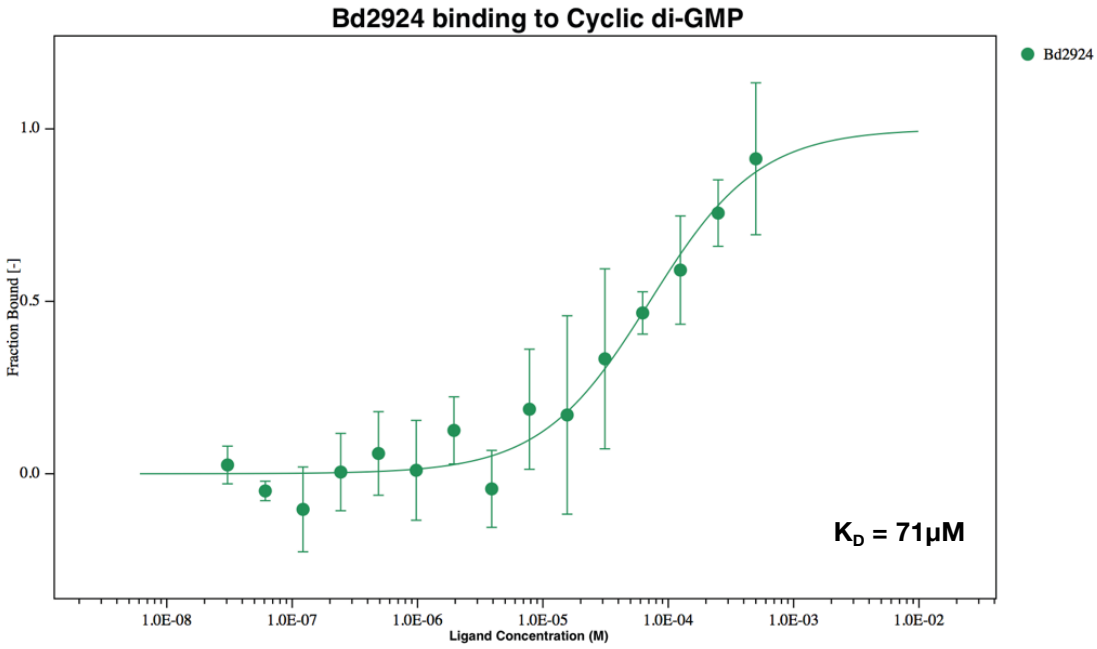


Figure 46. Binding curve with CdG
The binding interaction between Bd2924 and CdG was monitored using MST

4.2.7 Bd2924 activity

We were unable to detect any enzymatic activity for Bd2924 in two independent ACAD activity assays with either C10 or C12 fatty acyl-CoA substrates (DCPIP+PMS assay & Ferrocenium ion assay).

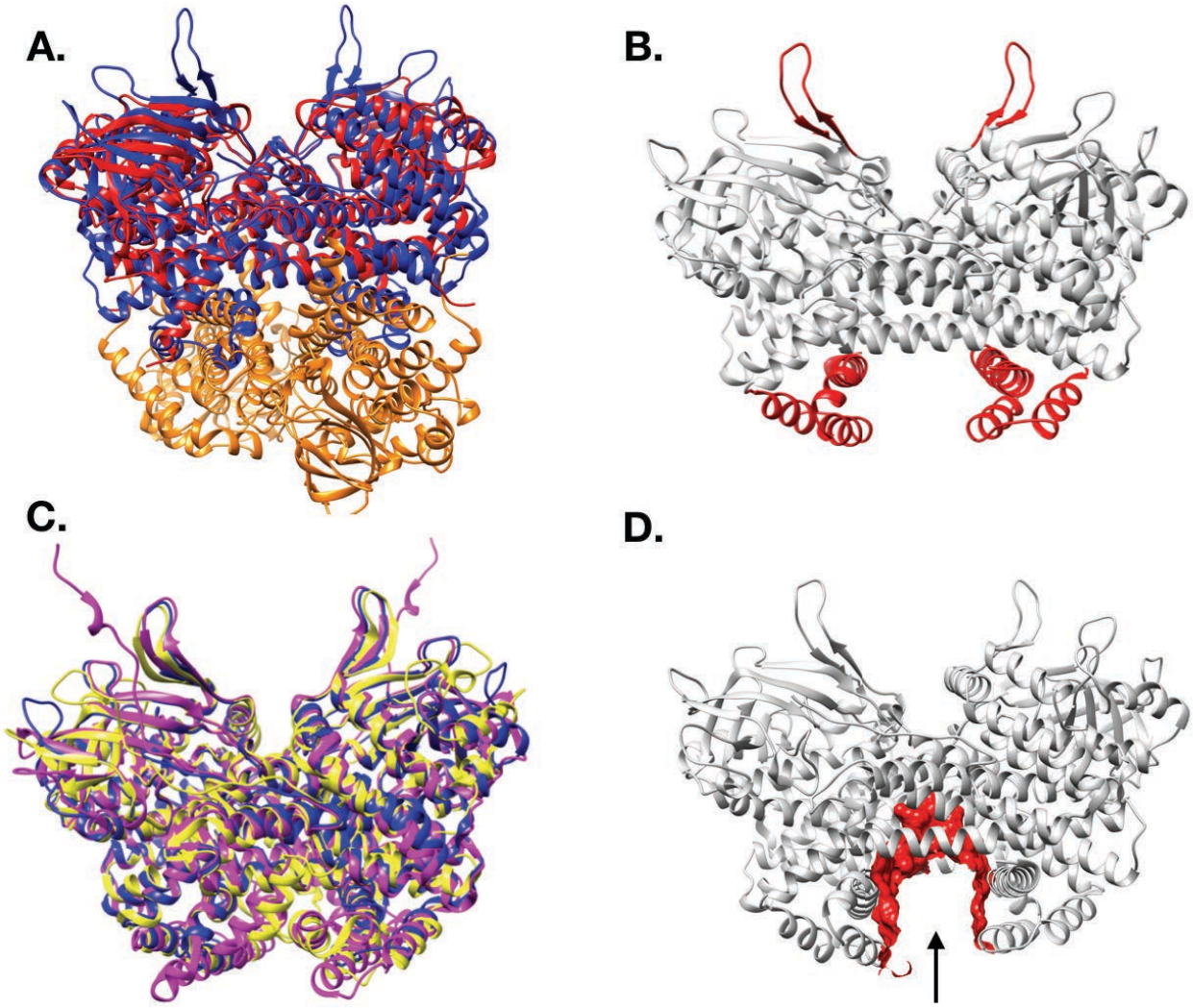


Figure 47. Bd2924 structural comparison.
A. Bd2924 dimer (blue) overlaid with MCAD tetramer (PDB:1EGD, red + orange, length = 746 residues, rmsd = 2.21Å). **B.** The major structural difference between Bd2924 and MCAD are highlighted in red on the structure on Bd2924. The N-terminal β-sheet loop and C-terminal extension domain are features of divergent ACAD homologues that form an unrelated tetramer. **C.** Bd2924 (blue) overlaid with ACDH-11 (PDB:4Y9J, magenta, 961 residues, 1.41Å) and AidB (PDB:3DJL, yellow, length = 881 residues, rsmd = 1.54Å). **D.** Bd2924 has a less extensive C-terminal extension domain and as a result has a large groove between dimers (red). Alignments were performed using the SSM superimpose tool on coot.

4.3.2 Substrate-binding and activity

Bd2924 is able to bind C10 chain-length fatty acyl-CoA, demonstrated by the Bd2924:C10-CoA complex structure. Bd2924 binds C10-CoA with more interactions (eight hydrogen bonds) than MCAD binds acyl-CoA substrates (four hydrogen bonds)

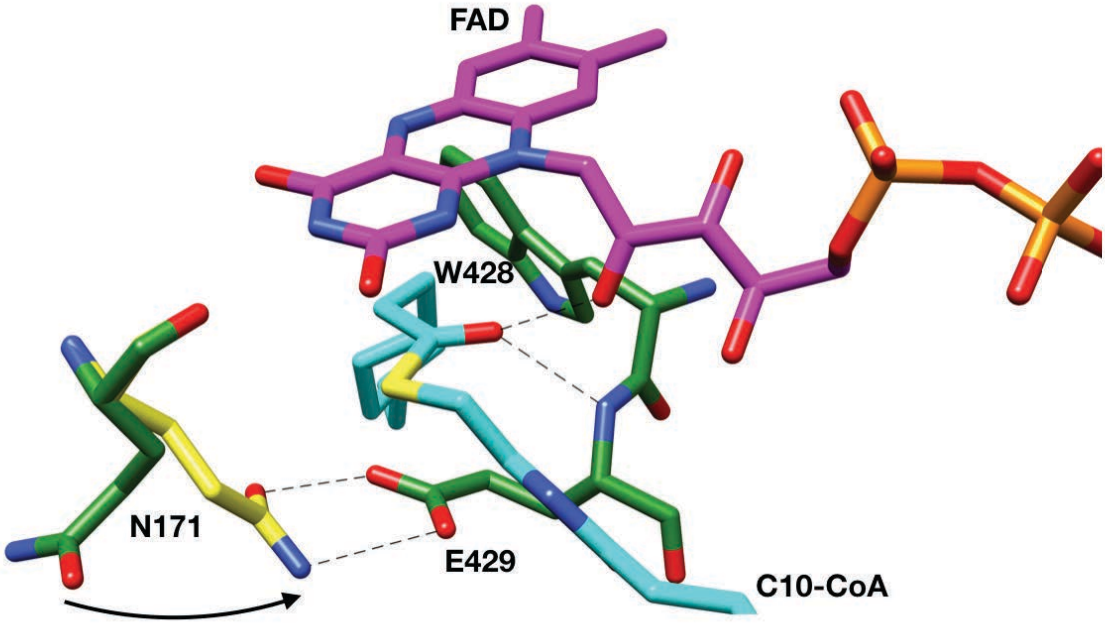


Figure 48. Bd2924 active site.
The C10-CoA ligand is bound in the active site between FAD and the catalytic base E429. In Bd2924:C10-CoA complex N171 adopts a new conformation (yellow) that forms a hydrogen bond with the sidechain of E429 and may affect its ability to abstract hydrogen from the C_α (supressing catalytic funciton).

Besides the lack of detectable dehydrogenase activity for Bd2924 (and comparison to other ACAD homologues) it seems unlikely that transfer of electrons to ETF would be possible by the currently proposed mechanism [191]. Bd2924 tetramer formation would impede access to the FAD molecule buried deep between the dimer and tetramer complex. As such for an ETF to access and re-oxidise the buried FAD molecule in Bd2924, it would have to disassemble our observed tetramer complex [203]. Blocking of ETF access is demonstrated in ACO, where a C-terminal α-helix crosses the solvent accessible groove to the si-face of FAD and thereby shields the accessibility to ETF [203]. Bd2924 could alternatively utilise molecular oxygen similar to ACOs although solvent accessibility to the FAD molecule is not as extensive as in ACOs.

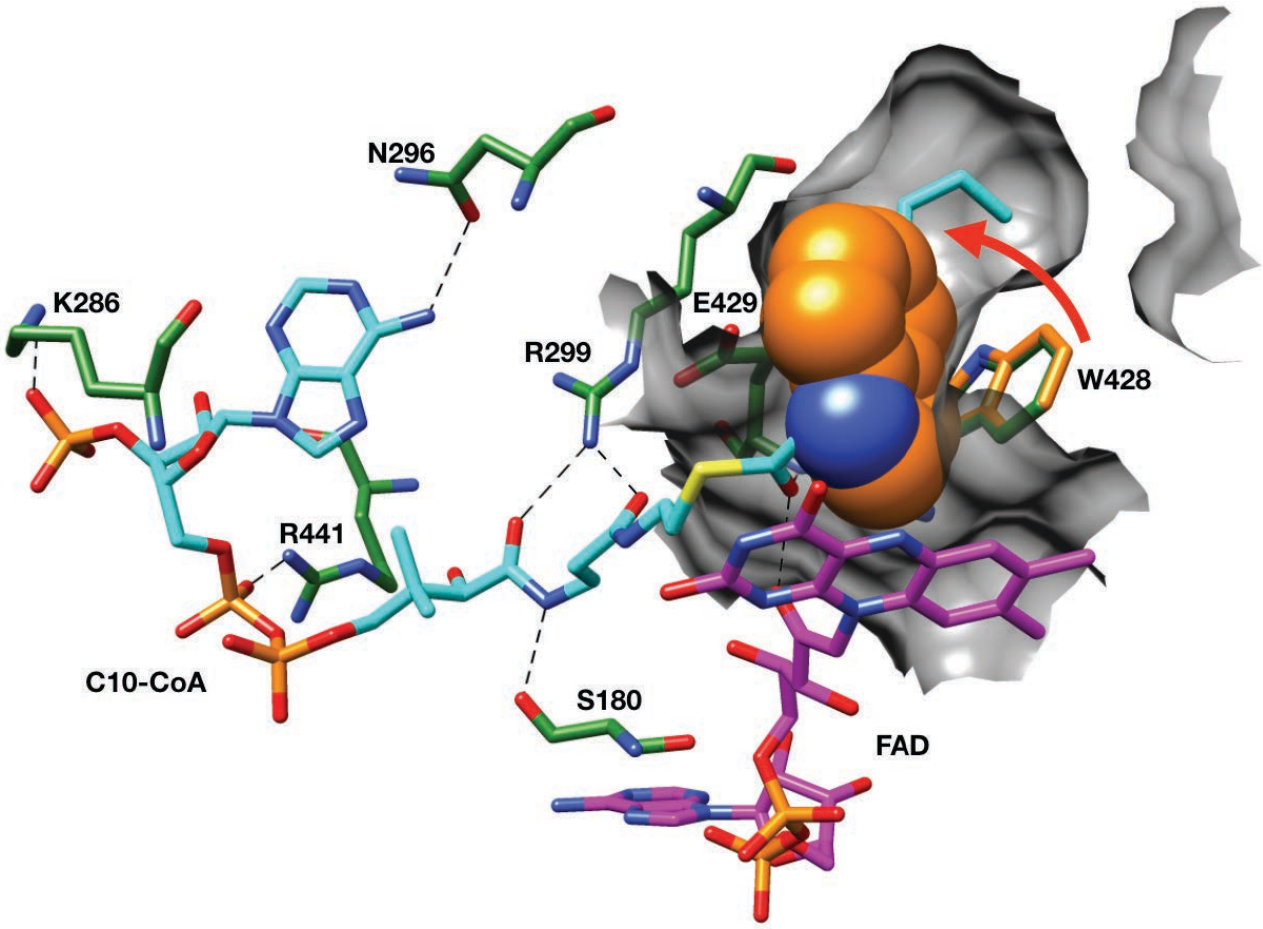


Figure 49. Trp428 gates the fatty acyl binding tunnel
The Bd2924:C10-CoA complex structure has one occupied C10-CoA binding site. In the monomer where the substrate binding site is free of ligand, W428 adopts two conformations (orange). One W428 conformation is located directly beneath the FAD N5, which blocks access to the substrate binding tunnel.

4.3.3 A functionally relevant Bd2924 interaction with CdG is questionable

We determined the binding affinity between Bd2924 and CdG to be ~71 μM, which is approximately 10 fold greater than the published value determined by Rotem et al [82]. To ensure consistency the binding buffer used by Rotem et al. was replicated, but no measurable binding events were observed with a maximum CdG concentration of 5 mM, which far exceeds the cellular concentration of CdG [79]. Binding was detected after rigorously testing multiple assay buffers and protein labelling techniques. The increased K_D value we obtained suggests that Bd2924 binds CdG in a non-specific

Bd2924s role is also likely to be divergent from β-oxidation based on its structural similarities to divergent ACADs and the lack of detectable dehydrogenation activity.

The exact function of Bd2924 remains unknown and requires further exploration. However, it is interesting to speculate that Bd2924 may be involved in fatty acid detection and utilisation during the *Bdellovibrio* predatory cycle. *Bdellovibrio* utilises prey-derived long-chain fatty acids for energy through β-oxidation and for direct incorporation into biosynthetic pathways [59]. Bd2924 may therefore act as a sensor of fatty acyl chains. Its function may be to: transport fatty acyl-CoA; sequester fatty acyl-CoA as part of a signalling cascade (alike ACDH-11); or generally be involved in β-oxidation. ETF genes exist in *Bdellovibrio HD100* genome (Bd0021 and Bd0022) but these may interact with other acyl-CoA dehydrogenases, such as those named in Table 12. Close homologues of Bd2924 are found in other endo-biotic predatory bacteria (60-75% ID) and *Legionella* species (55% ID). It is therefore interesting to speculate that Bd2924s role is specific to bacteria which invade cells.

Protein	Type	Length	Up-regulated in?
Bd2174	Butyryl-CoA dehydrogenase	816 (including 284aa DUF)	growth phase
Bd1973	Butyryl-CoA dehydrogenase	388	growth phase
Bd1201	Isovaleryl-CoA dehydrogenase	428	growth phase
Bd2924	acyl-CoA dehydrogenase	505	growth phase

Table 12. Hd100 ACADs
A list of ACADs in the HD100 genome – annotated by uniprot

5.1.2 Malic Enzymes

5.1.2.1 Overview

Malic enzymes catalyse the oxidative decarboxylation reaction of L-malate to pyruvate using the cofactor $NAD(P)^+$ in the presence of a divalent cation such as Mn^{2+} or Mg^{2+}.

$$\text{L-malate} + \text{NADP}^+ \rightleftharpoons \text{pyruvate} + CO_2 + \text{NADPH}$$

MEs are categorised into three classes based on coenzyme specificity and their ability to decarboxylate an oxaloacetic acid (OAA) substrate. These classes include: Non-OAA decarboxylating NAD^+ -specific, OAA decarboxylating NAD^+ -specific and OAA decarboxylating $NADP^+$ -specific. MEs are widely distributed enzymes involved in different metabolic pathways such as energy metabolism, photosynthesis and lipogenesis in organisms from bacteria to higher Eukaryotes [205]. Three isoforms of eukaryotic ME have been identified and studied (cytosolic, mitochondrial and plastid). On the other hand, prokaryotic MEs isoforms include a wider range of structural and functional diversity (including multi-modular MEs), which are comparatively understudied.

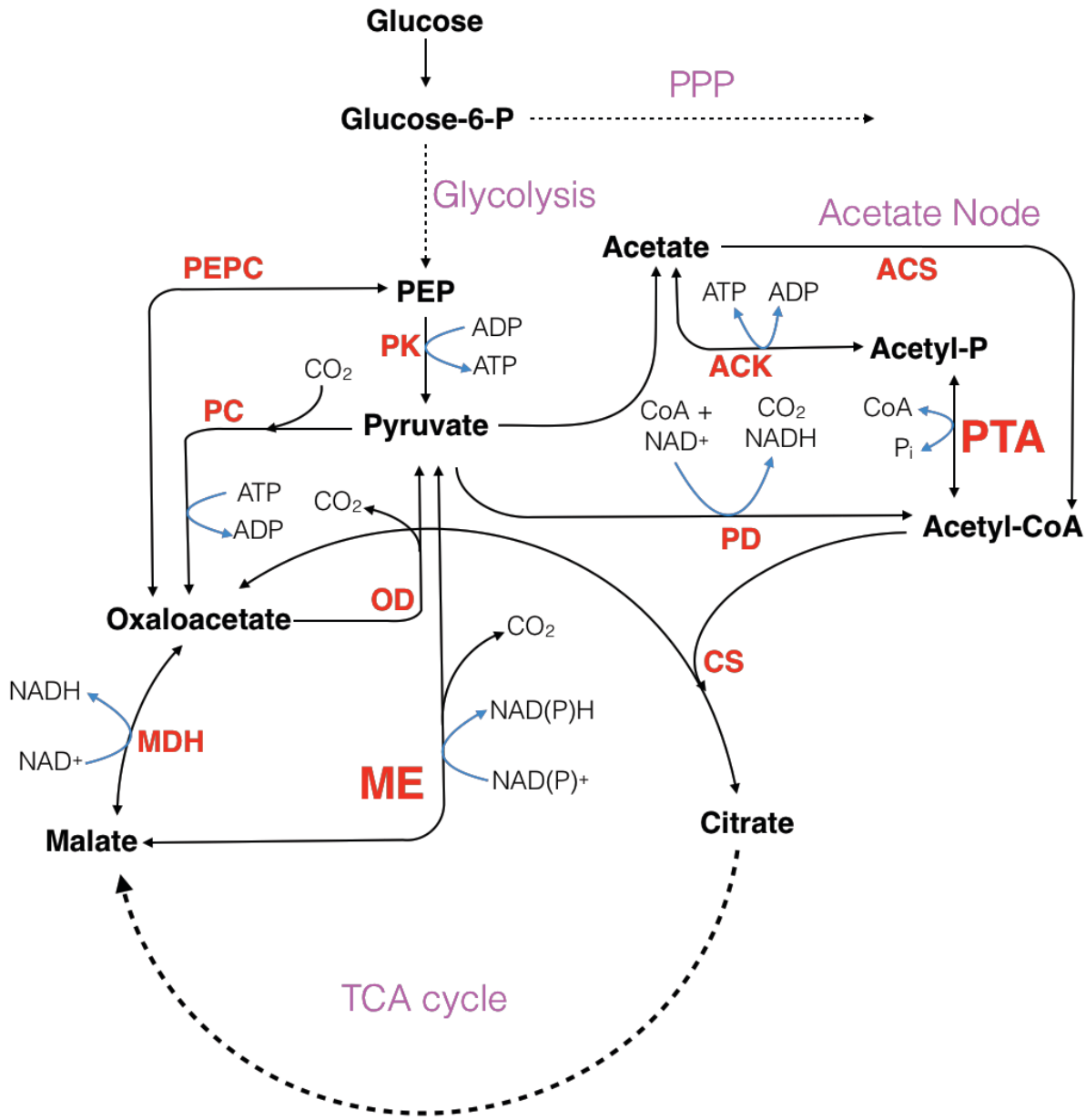


Figure 50. Illustration of central metabolism (general)
This Figures depicts a simplified illustration of a central metabolic network involving Pta and ME enzymes (based on *E. coli* metabolism). Each individual pathway is labelled in pink. The enzymes involved in the reactions are displayed in red: PK - pyruvate kinase, PD - pyruvate dehydrogenase, CS - citrate synthase, PEPC - phosphoenolpyruvate carboxylase, PC - pyruvate carboxylase, OD - oxaloacetate decarboxylase, MDH - malate dehydrogenase, ME - malic enzyme, ACS - acetyl-CoA synthase, ACK - acetate kinase, PTA - phosphotransacetylase

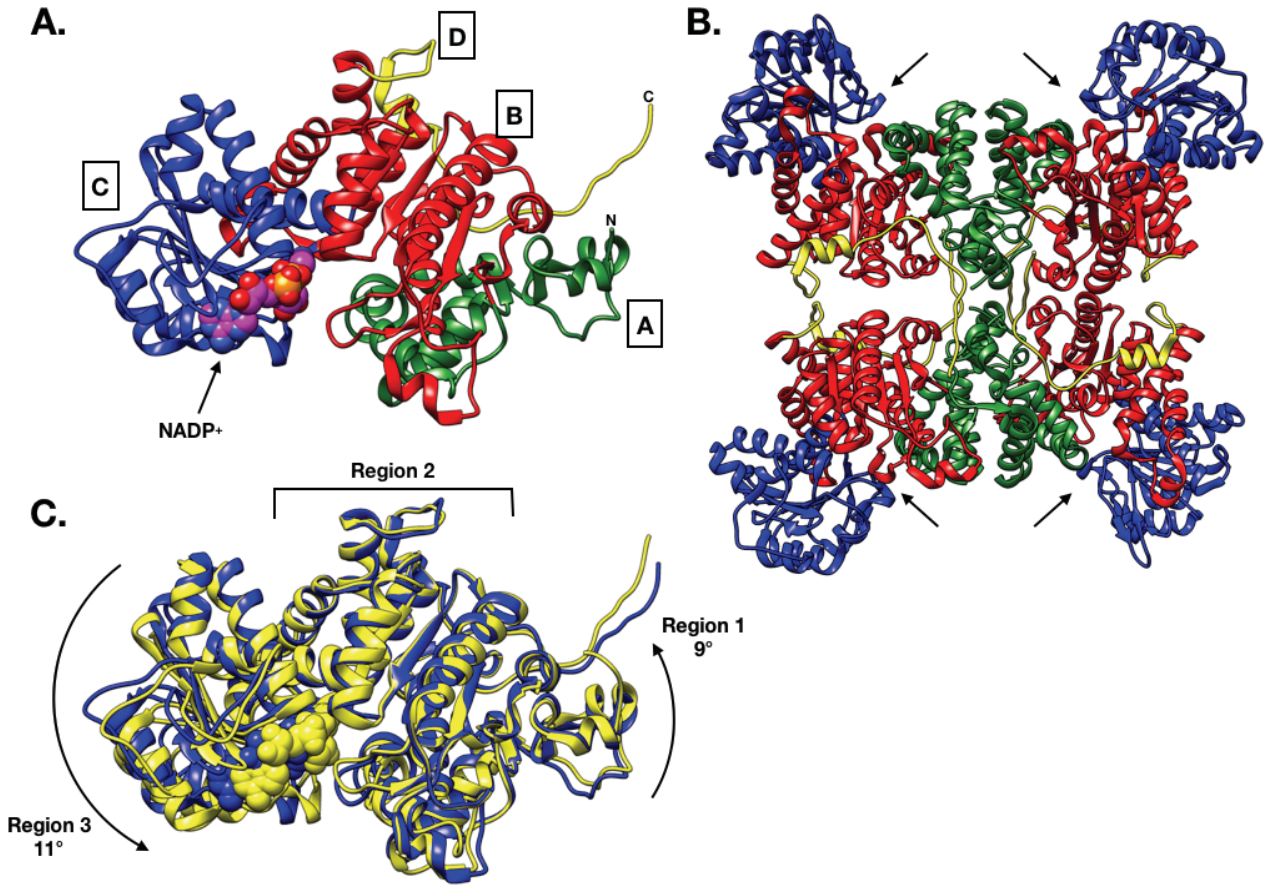


Figure 51. Human-ME tetramer organisation and domains
A. Human-ME like other eukaryotic MEs can be split into four domains: Domain A (Green) makes the dimer interface; Domain B (Red) composes the central core of the protein and cation-binding site; Domain C (Blue) contains the NAD(P)$^+$ binding site; and Domain D (yellow) is a short extension that contributes to the tetramer interface. **B.** Eukaryotic MEs arrange to form a tetramer complex – a dimer of dimers. Black arrows indicate the location of the substrate-binding cleft. **C.** The structure of human-ME has been captured in a catalytically active 'closed' conformation and a resting 'open' conformation. Region-3 including domain C (NADP$^+$ -binding domain) undergoes an 11° rigid body rotational movement that encloses the active-site cavity and positions the catalytic residues and cofactor ready for catalysis. A smaller 9° rigid body rotational movement of part of domain A (region-1) also occurs, which re-orientates the tetramer interface. Region-2 containing domain B and part of domain A maintains the same conformation. The closed form (PDB:1EFK) was superimposed onto the open form (PDB:1QR6), using LSQ superpose tool in coot to align domain B (1275 atoms, rmsd = 0.94Å).

5.1.2.5 Divalent Cation binding

ME activity is dependent on the presence of a divalent cation such as Mn^{2+} and Mg^{2+}. In human NAD(P)$^+$ -dependant-ME, Mn^{2+} is ligated by six oxygen atoms in an octahedral fashion. Three of these oxygen atoms are from acidic residue sidechains; E255, D279 and D256. When no ligand is present in the active-site water molecules provide the remaining oxygen atoms.

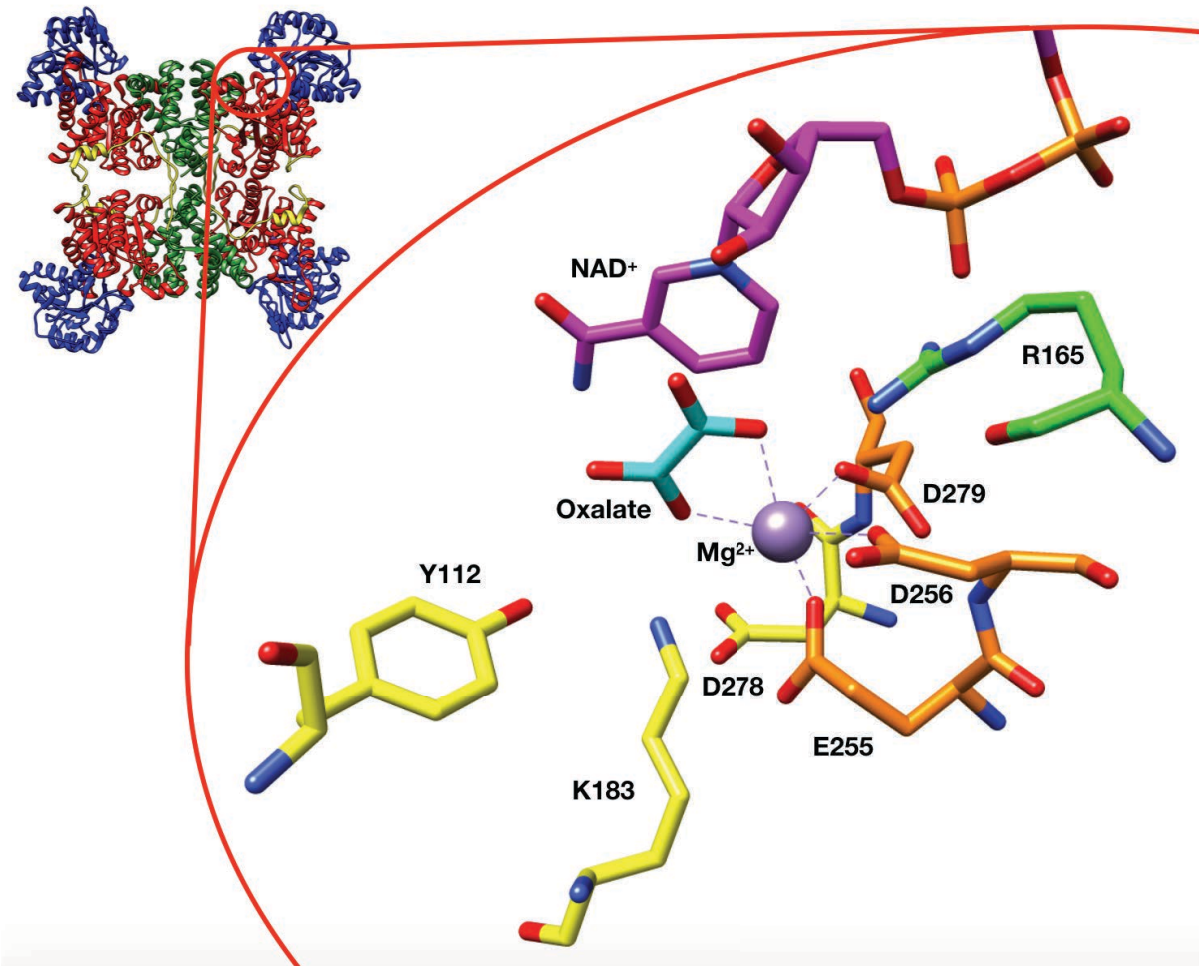


Figure 52. Human-ME active Site
A close up view of Human-ME active site in complex with: Mg^{2+} cation (purple), a malate substrate analogue – oxalate (cyan) and NAD^+ cofactor (magenta). The cation coordinating residues are shown in orange. The substrate-binding residue (R165) is shown in green. Catalytic residues (Y112, K163 and D278) are shown in yellow. PDB:1DO8

5.1.2.7 Nucleotide Binding

ME utilise a $NAD(P)^+$ cofactor to perform catalysis. Generally MEs perform the reaction specifically using either NAD^+ or $NADP^+$. However, some isoforms are able to utilise both cofactors. $NAD(P)^+$ is bound to domain C in eukaryotic MEs and associates with the Rossmann fold. The nicotinamide portion of the cofactor binds in an *anti* conformation, similarly to how other dinucleotides bind to a Rossmann fold. The

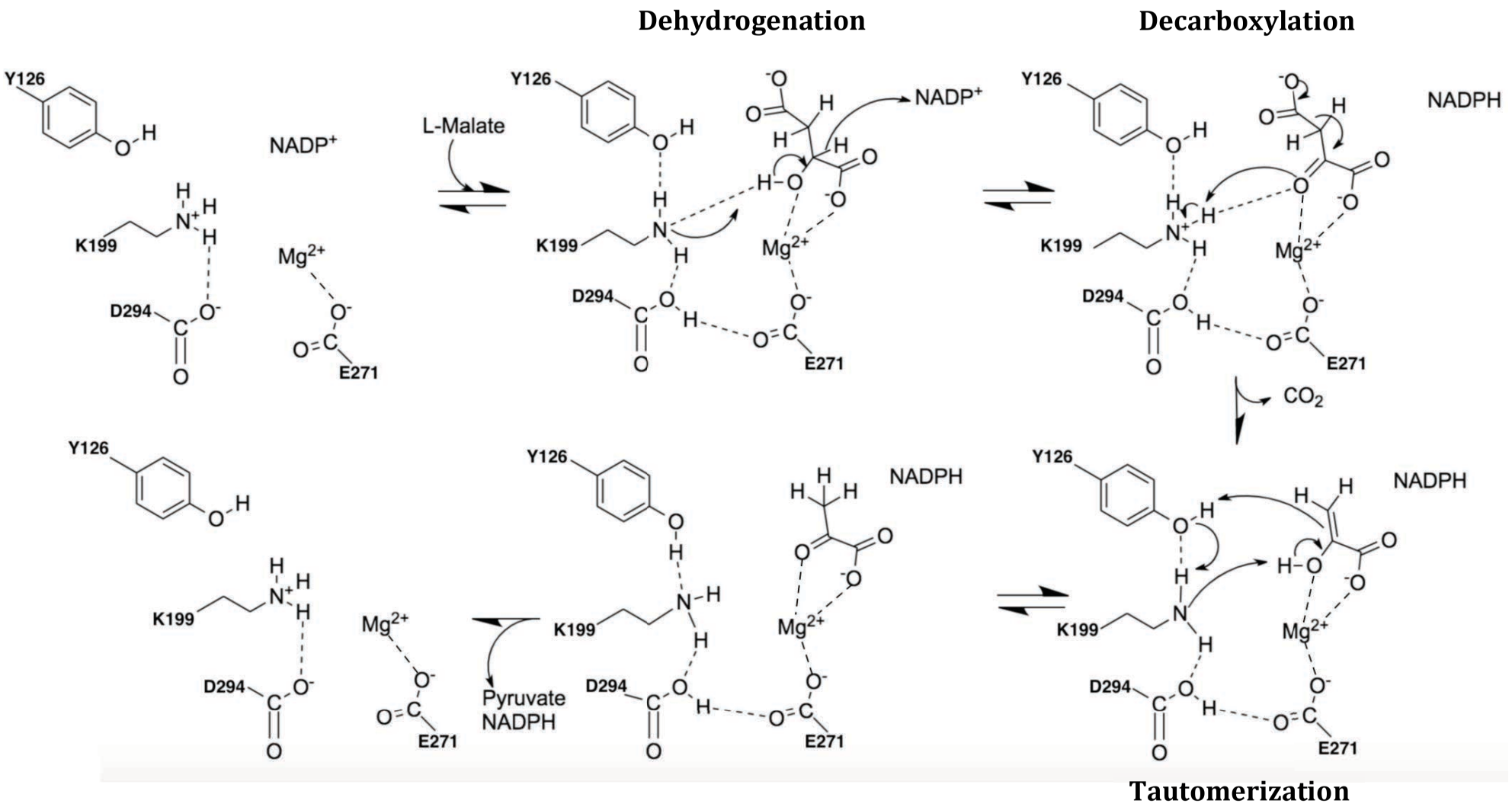


Figure 53. Malic enzyme catalytic mechanism.
The catalytic mechanism performed by malic enzymes progresses via three steps: Dehydrogenastion of L-malate to form oxaloacetate, decarboxylation of oxaloacetate to form enolpyruvate and finally tautomerization of enolpyruvate to form pyruvate. The enzyme is primed by binding Mg^{2+} and $NADP^{+}$ and upon substrate binding a conformational change repositions the catalytic residues.

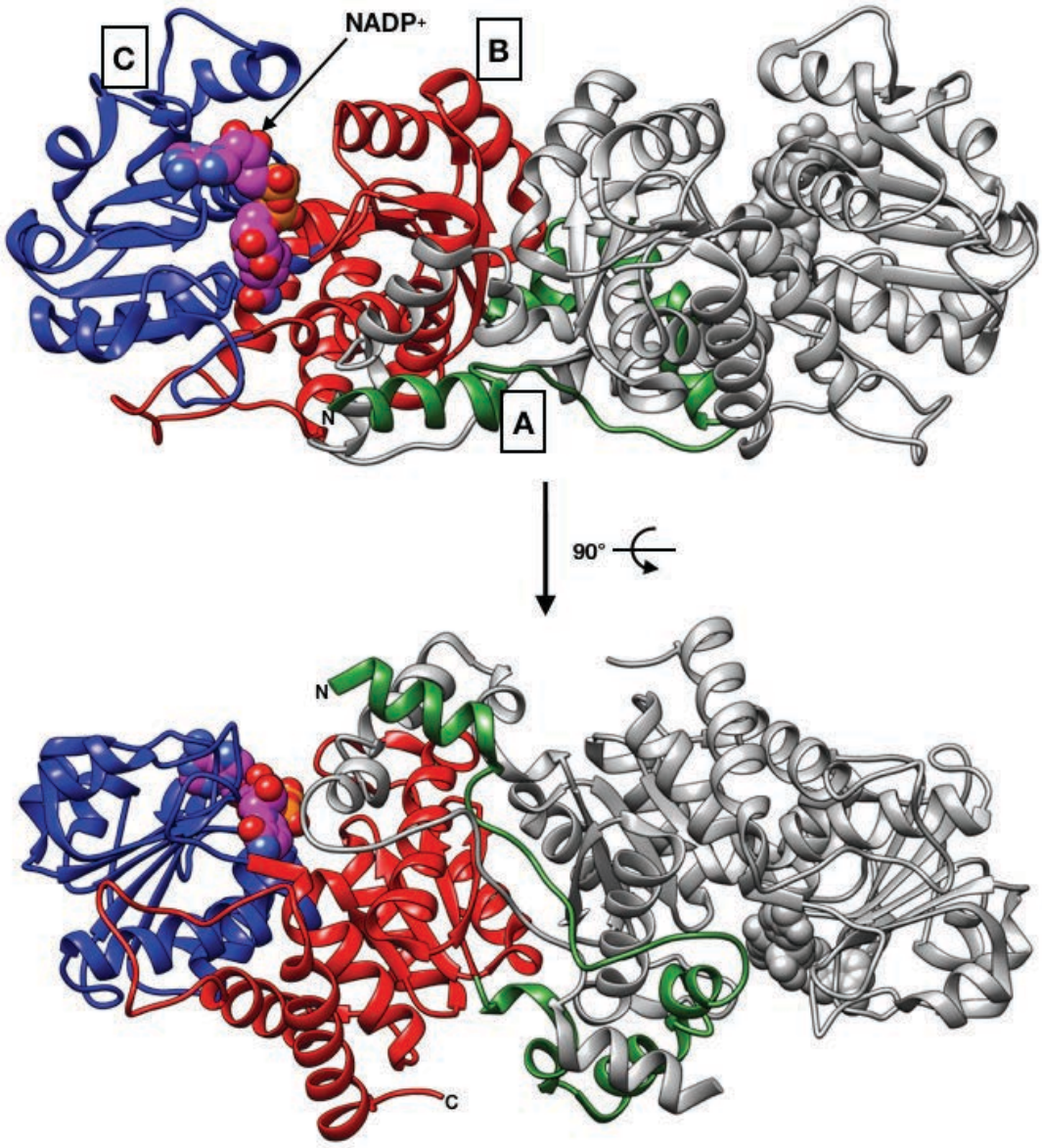


Figure 54. Prokaryotic ME (AYWB-ME)
The structure of AYWB-ME (PDB:5CEE) in two orientations. AYWB-ME has a minimal ME fold. The fold of domain A is only associated with prokaryotic ME and forms the dimer interface (green). Domain B is the core of the ME structure (Red) and domain C is the NAD+ binding domain (Blue). AYWB-ME forms a tightly interacting homodimer (monomer1 = coloured, monomer2 = grey) that is stabilised mainly by interactions between residues of domain A. A NAD+ ligand is shown bound in substrate binding cleft between domains B and C (magenta) [1].

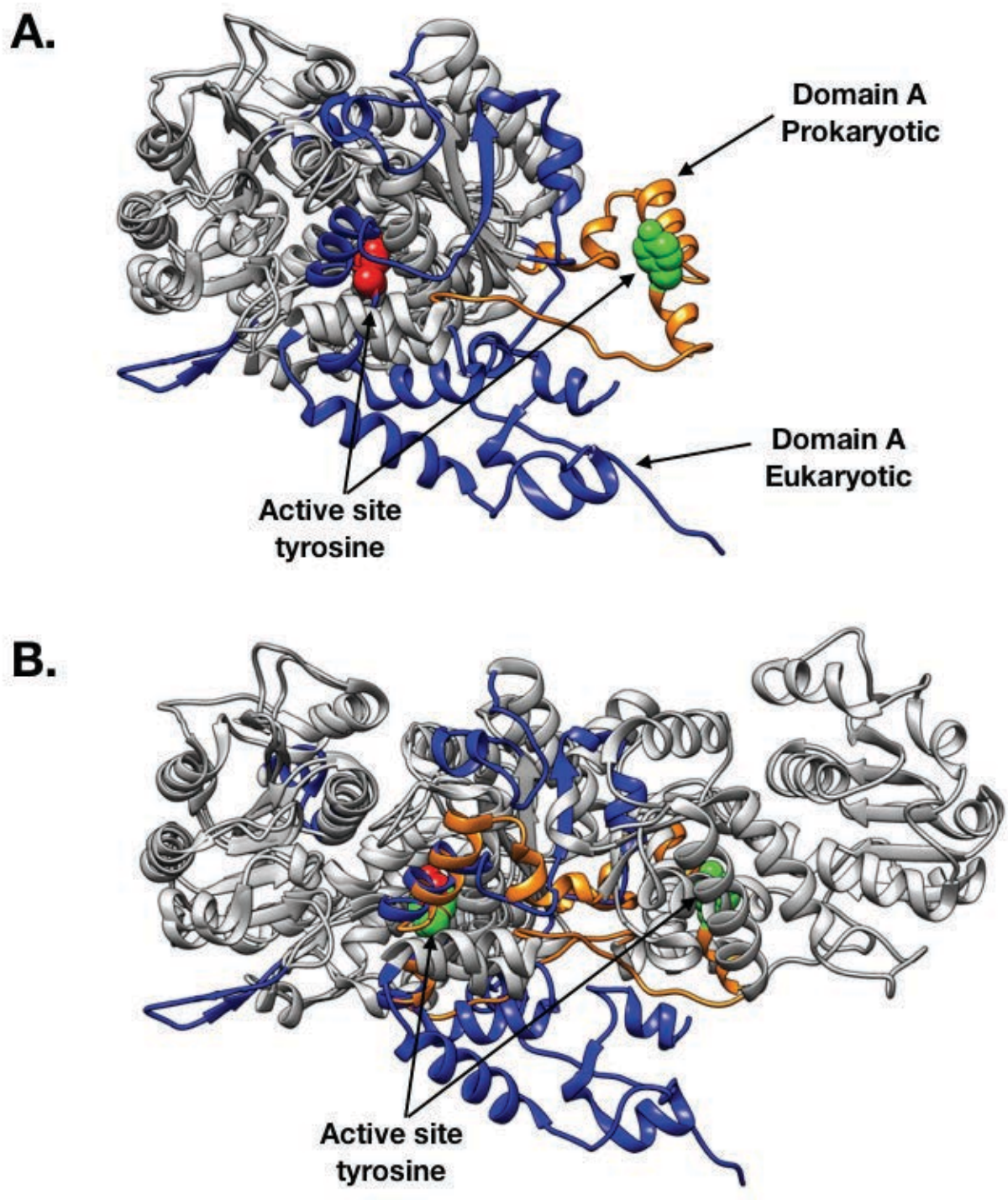


Figure 55. Comparison between prokaryotic and eukaryotic ME
The structures of AYWB-ME (PDB:5CEE) and human-ME (PDB:1QR6) have been superimposed using SSM superpose tool on coot (length = 320 residues, rmsd = 1.97Å). These two structures are typical of prokaryotic and eukaryotic ME structures respectively. **A.** The structural differences between the two structures are highlighted in orange (prokaryotic only) and blue (eukaryotic only). The core fold of the proteins is highly similar between the two MEs therefore has been coloured grey. Domain A has a significantly different fold between prokaryotic and eukaryotic ME. Furthermore the position of the active site tyrosine is located at different position between the two monomers (red = 1QR6, green= 5CEE). **B.** The active site of prokaryotic MEs is composed of residues from both monomers of the dimer complex. When the second monomer of the 5CEE dimer is displayed the location of the active site tyrosine from the other monomer is positioned into the active site (green and red tyrosine residues overlap).

5.1.3 Phosphotransacetylases (Pta)

5.1.3.1 Overview

Phosphotransacetylases (Pta) play an integral role in acetate metabolism in fermentive anaerobes. Pta catalyses the transfer of the acyl group to Coenzyme A (CoA) from acetyl-phosphate (acetyl-P) to form inorganic phosphate (Pi) and acetyl-CoA. Acetyl-CoA is a key central metabolite used directly in anabolic processes, such as the delivery of the acetyl group to oxaloacetate, forming citric acid, which is an important step in the Krebs cycle.

$$\text{Acetyl-CoA} + P_i \rightleftharpoons \text{CoA} + \text{Acetyl-phosphate}$$

Fermentive microbes utilise Pta and acetate kinases (Ack) in a metabolic pathway to generate energy. Ack reversibly catalyse the transfer of an acetyl-group from acetyl-P to ADP to create a molecule of ATP [242].

$$\text{Acetyl-phosphate} + \text{ADP} \rightleftharpoons \text{Acetate} + \text{ATP}$$

Pta enzymes play important roles in the “acetate switch”, a phenomenon where cells transit from a rapid growth state that produces acetate to one of a slower growth state that utilises acetate as an energy source. Acetate is regarded as a major intermediate in the global carbon cycle [243].

Despite the importance of Pta in anaerobic metabolism as well as their widespread distribution, their catalytic mechanism remains poorly understood. J Henkin et al.

Pta enzymes form a dimer complex, adjoined centrally by interactions between domain-II. The dimer interface is stabilised by two helices that align almost parallel at the centre (Helix-12 from both monomers). Residues from Helix-12 also interact with residues from Helix-9 and Helix-10 of opposing monomers, further stabilising the central core of the dimer [250].

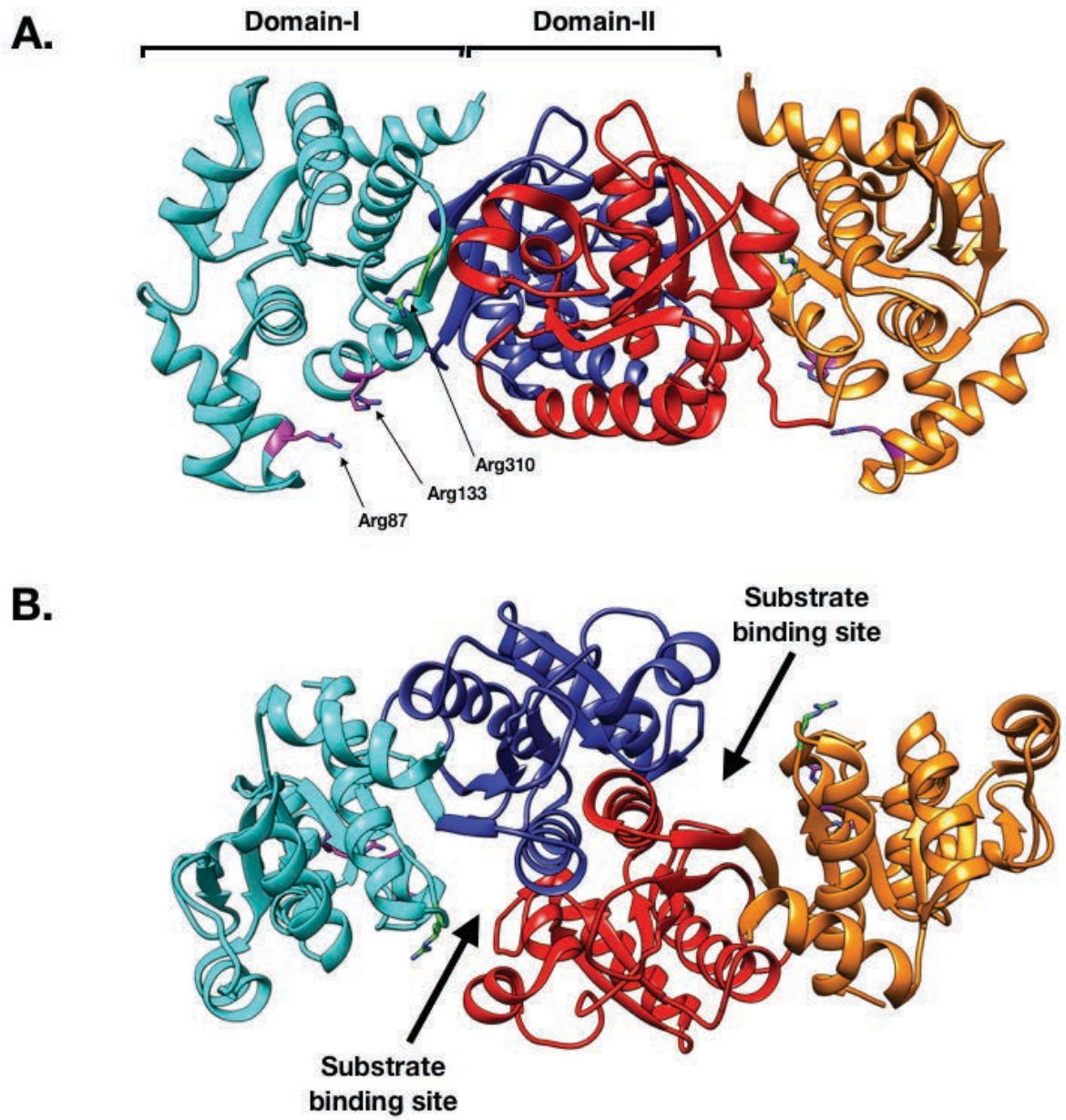


Figure 56. Pta structure
A. Pta enzymes form a bi-lobed homodimer. The dimer is held in place by interactions occurring between residues of the dimerization domain (domain-II – blue and red). The location of the conserved Pta arginine residues believed to be involved in substrate binding are displayed. Arginine residues coloured in magenta interact with CoA, whereas the arginine residue in green interacts with acetyl-P. **B.** A 90° rotation around the x-axis highlights the almost parallel alpha helix that contributes to the dimer interface. Between domain-I (cyan) and domain-II (blue) is a large substrate binding cleft – marked with black arrows.

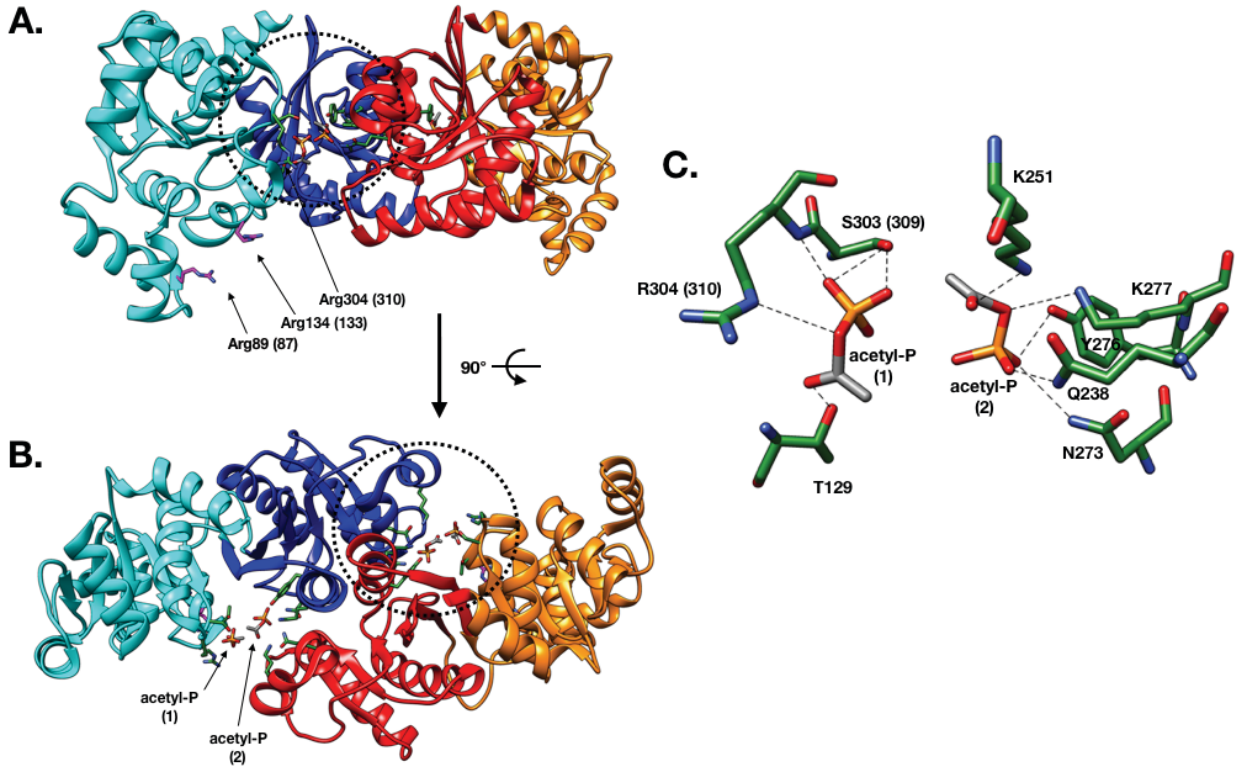


Figure 57. Acetyl-P binding site
A. The acetyl-P binding site is highlighted by a dashed black circle. **B.** The acetyl-binding site viewed from an angle 90° around the x-axis. **C.** A close up view of the residues involved in binding acetyl-P. Two acetyl-P molecules are shown to bind to the active site. The conserved arginine, R304 (R310 in *Methanosarcina thermonphila*) makes a hydrogen bond with acetyl-P(1). (PDB:1XCO)

Two Pta:CoA complex structures for *M. thermophila* Pta (PDB: 2AF3 and 2AF4) have been published [247]. CoA was observed to interact with the Pta enzyme at two distal sites (CoA^1 and CoA^2) in the co-crystallised and soaked structures (Figure 58). One CoA site (CoA^1) was subjectively chosen as the higher affinity binding site (CoA observed at this location in both ligand complex structures) and catalytic residues R310, S309 and D316 were assigned based on their proximity to the bound CoA thiol group [247]. ITC experiments confirmed two CoA binding sites with disparate K_D values of 20 µM and 1mM. The lower affinity binding site is likely non-specific as the cellular concentration of CoA is unlikely to approach concentrations above 1mM [112]. CoA^1 binding site

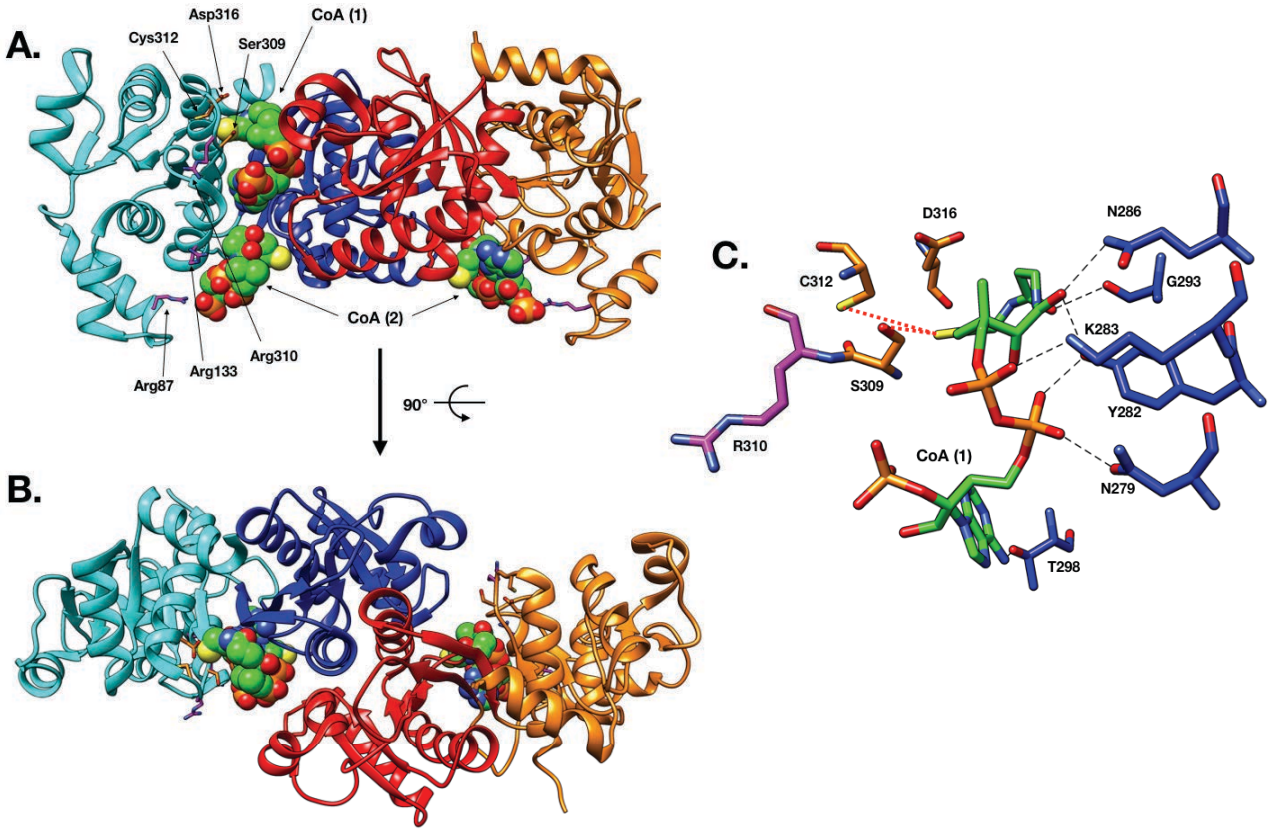


Figure 58. Pta CoA binding site
A. Structure of Pta dimer showing the location of CoA ligands (green) bound in the substrate binding groove. Two CoA ligands are shown to bind to the first monomer and one CoA ligand to the second monomer of the dimer complex. CoA1 and CoA2 sites are indicated. The location of the active-site residues and the conserved arginine residues that are shown to interact with CoA (R67, R133) are also highlighted (magenta). **B.** Rotation of image A by 90° around the x-axis. **C.** A close up view of the residues involved in the binding site of CoA1. Residues in blue are involved in binding the CoA substrate and the H-bond interactions highlighted by black dashed lines. Active-site residues are shown in orange and the interactions (>3.2Å) made with the thiolate group of CoA are shown in red. The conserved R310 that is believed to position acetyl-P is also shown. (PDB:2AF3)

5.1.3.5 Catalytic Mechanism

The currently proposed catalytic mechanism is based on the CoA bound structures. D312 is proposed to act as a general base and abstract the sulfhydryl proton from CoA, forming a negatively charged transition state that is stabilised by S309. The activated transition state then attacks the acetyl-P substrate, which is held in position by R310 [247]. Although the proposed mechanism is plausible it is not in agreement with biochemical and mutational studies. The mechanism does not address how R133 and R87 affect the $K_{M(CoA)}$ if they do not facilitate binding. Nor does this address how R310

binds and orientates acetyl-P if it makes hydrogen bonds to the 3'-phosphate group of CoA or why R310 mutation only displays a decrease in $K_{M(acetyl-P)}$. Furthermore R310 is orientated incorrectly to deliver acetyl-P to the active site (a concern of the authors in Lawrence et al.) [247].

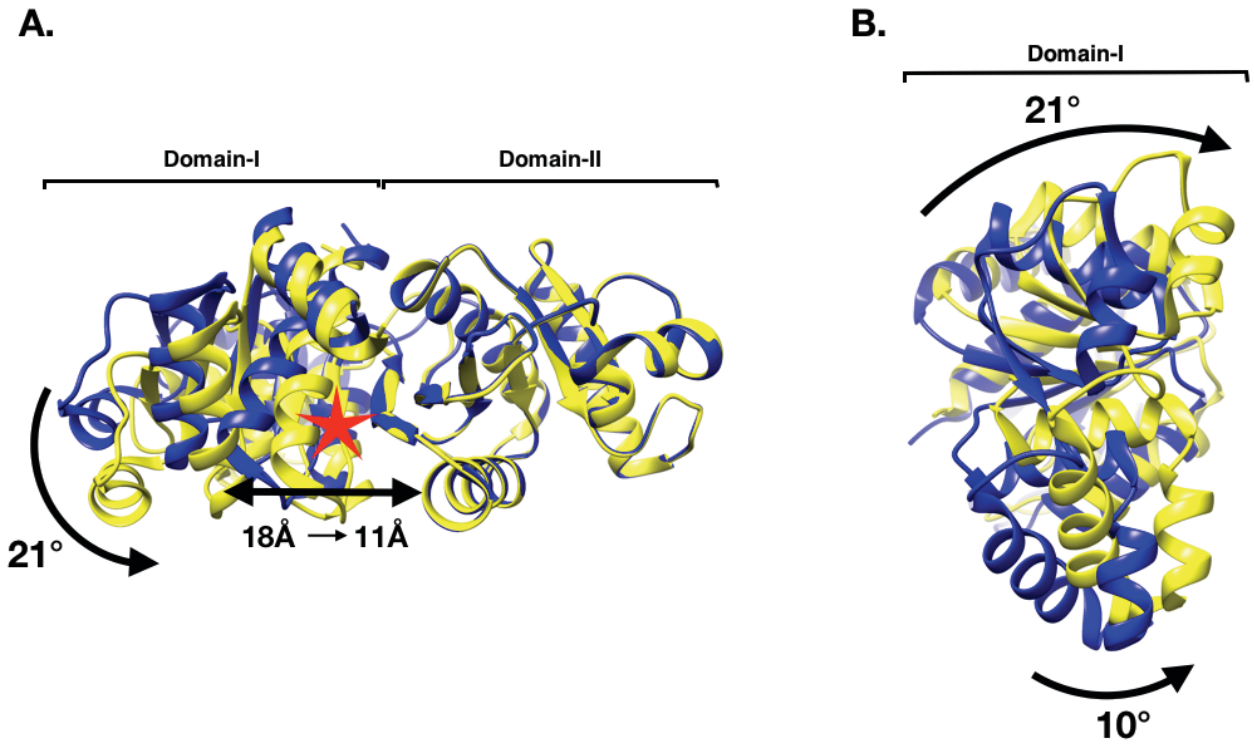


Figure 59. Pta domain movement
A. Domain I adopts multiple conformations in Pta structures, rotating as a rigid body to open and close the substrate binding cleft, located between domain I and II. A 21° swing of domain-I constricts the substrate binding groove from 18Å to 11Å. The movement is believed to be a mechanistic feature of substrate binding. **B.** The same rigid body rotational conformational change in domain-I viewed from the side. The upper portion of the domain undergoes a larger rigid body movement than the lower portion. The LSQ superpose tool in coot was used to align domain II (residues 149-299) of PDB:1QZT (length = 1114 atoms, rmsd = 1.97 Å)

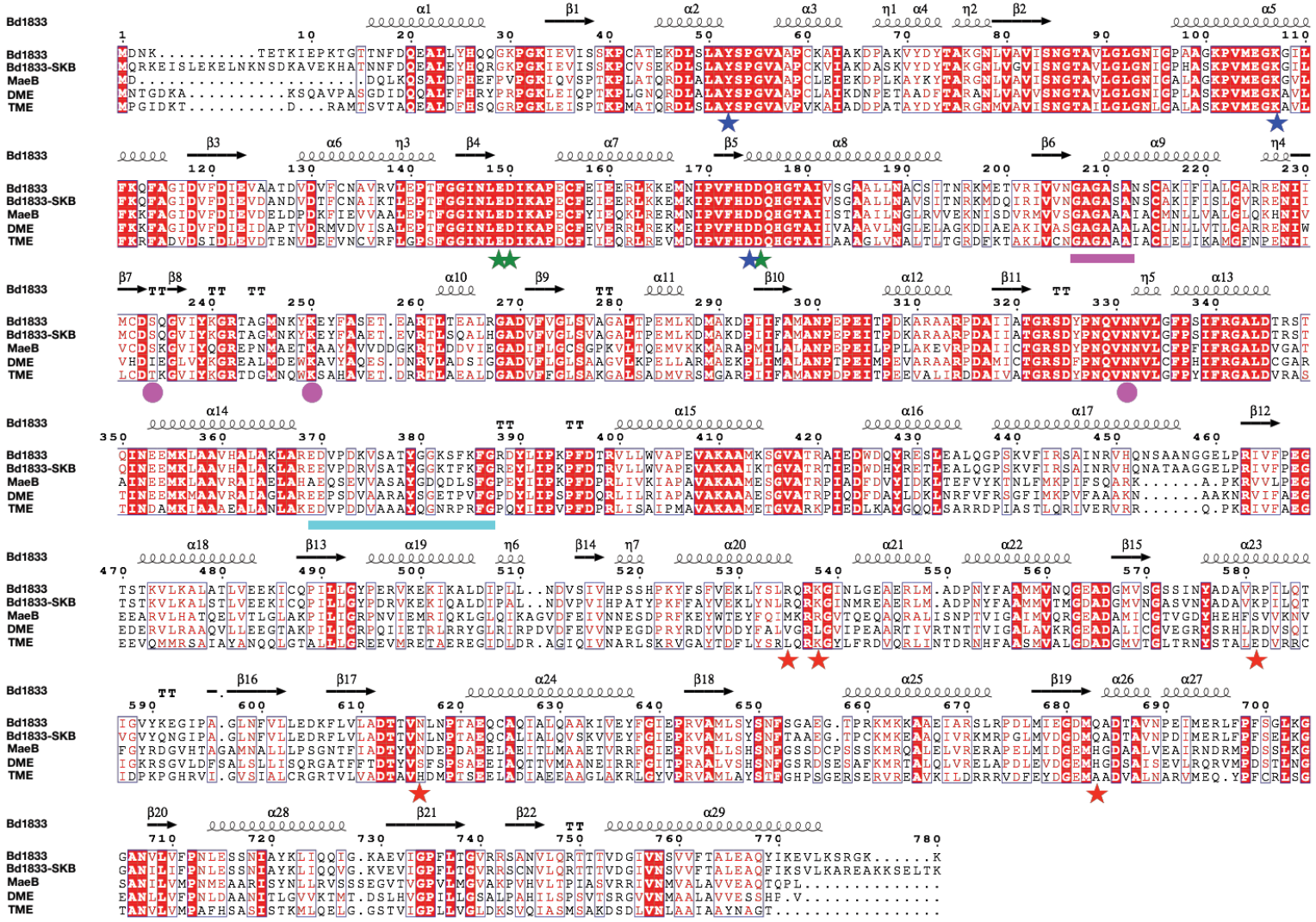


Figure 60. Bd1833 alignment with multi-modular malic enzymes

Catalytic residues are highlighted with blue stars. Cation-binding residues highlighted with green stars. Residues that hydrogen bond with $NADP^+$ are highlighted with pink dots and the GXGXXG motif with a pink bar. Residues with side chains that hydrogen bond with acetyl-CoA are highlighted with red stars. The cyan bar shows the location of subdomain D (multi-modular ME specific loop)

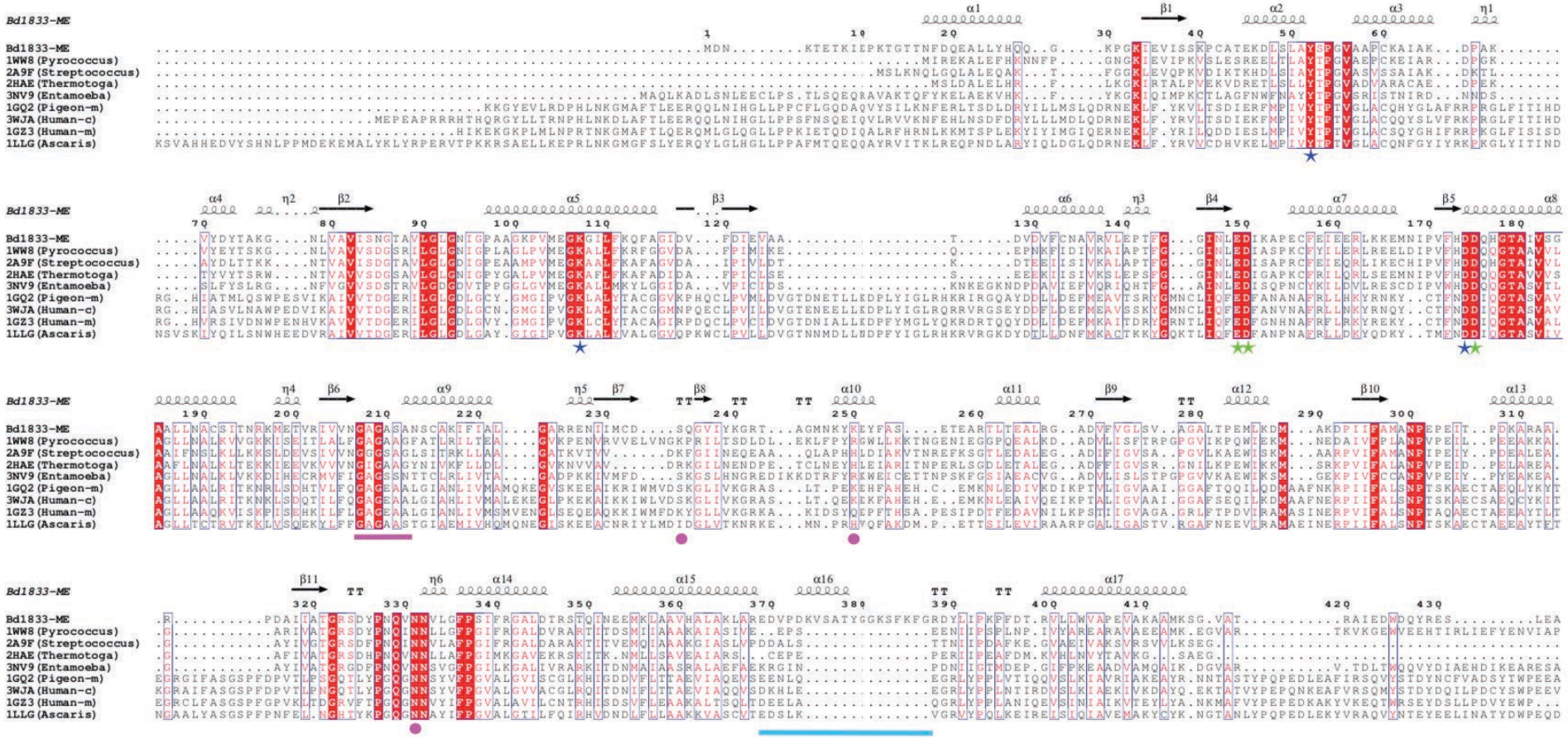


Figure 61. Bd1833-ME alignment with prokaryotic and eukaryotic ME
Catalytic residues are highlighted with blue stars. Cation-binding residues highlighted with green stars. Residues that hydrogen bond with NADP⁺ are highlighted with pink dots and the GXGXXG motif with a pink bar. The cyan bar highlights subdomain-D (specific for multi-modular malic enzymes).

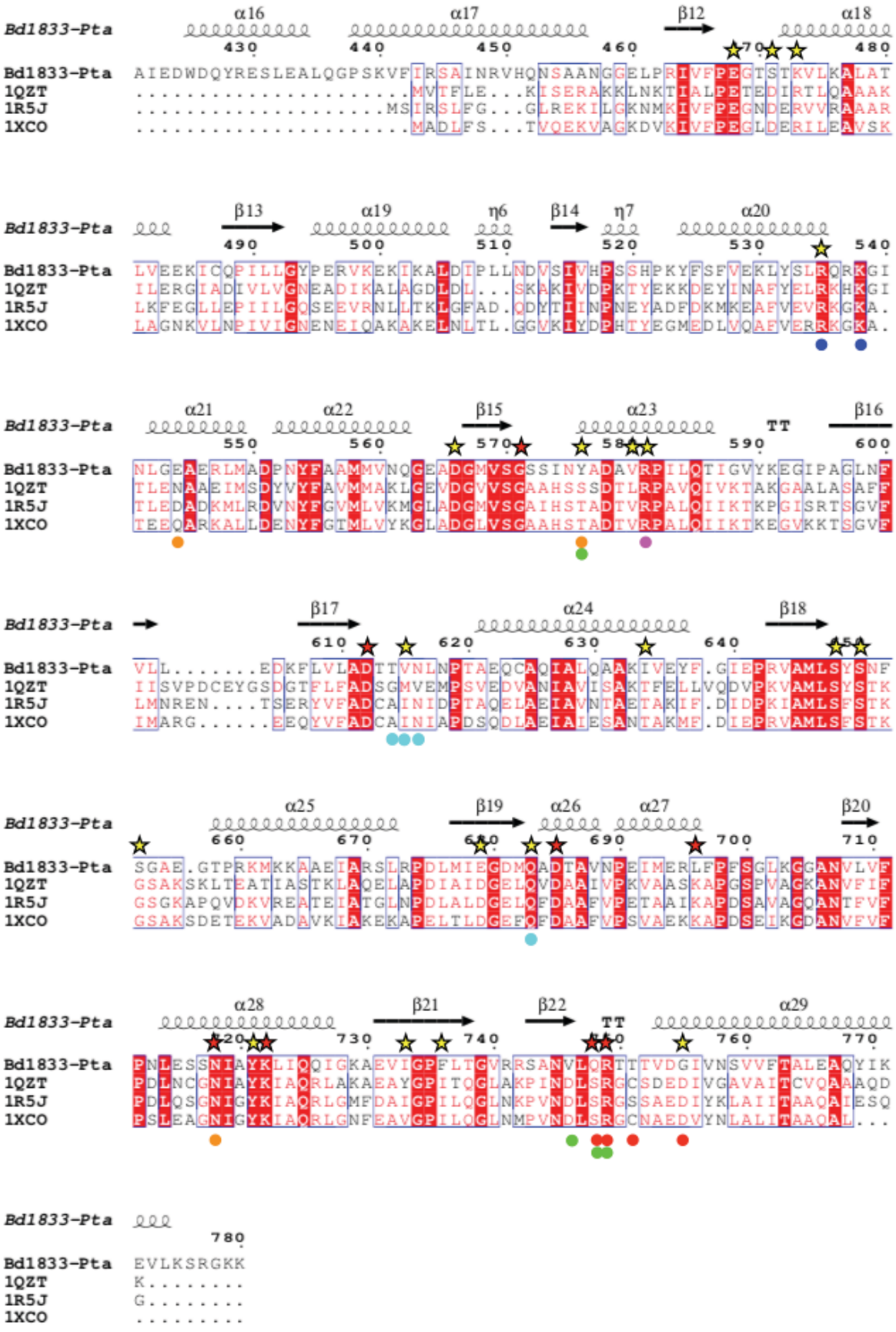


Figure 62. Bd1833-Pta alignment with Pta enzymes
The highly conserved and strictly conserved residues associated with Pta enzymes are highlighted by the yellow and red stars respectively. Dark blue circle – residues interacting with CoA 3'phosphate group; pink circle – residues interacting with CoA pyrophosphate group; light blue circle – other residues hydrogen bonding to CoA; orange circle – key additional residues highlighted from Bd1833 studies. Red circles are the previously proposed catalytic residues and the green circles are the newly proposed catalytic residues.

5.2.2 Bd1833 purification and crystallisation

Three constructs were designed for Bd1833: full-length (Bd1833), ME domain only (Bd1833-ME) and Pta domain only (Bd1833-Pta). Each construct of Bd1833 was successfully cloned, purified, crystallised and the structures solved. Figure 63 shows a representative purification gel and crystals, which were used to solve each construct.

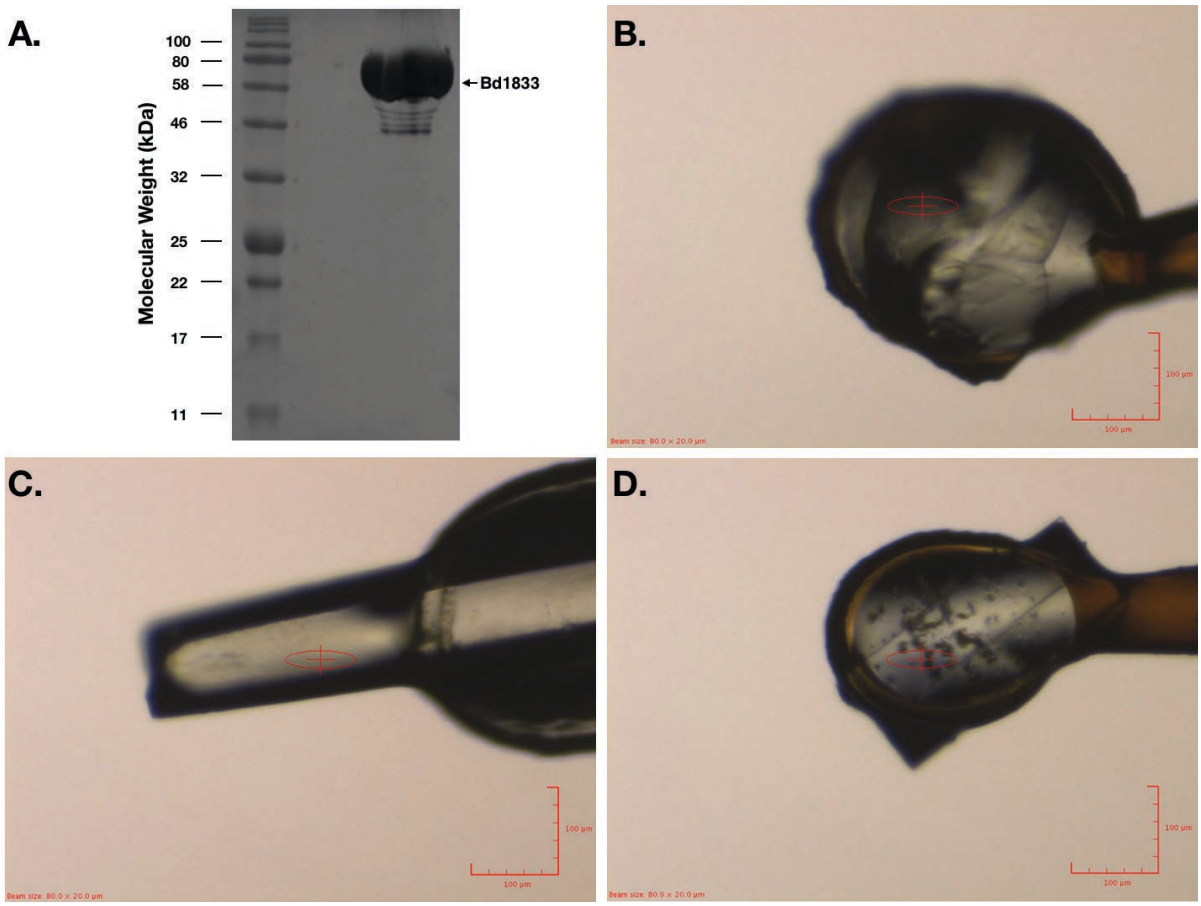


Figure 63. Bd1833 Purification and Crystallisation
A. SDS PAGE gel showing purified full-length Bd1833 **B.** Crystal of Bd1833 used for data collection **C.** Crystal of Bd1833-Pta used for data collection. **D.** Crystal of Bd1833-ME used for data collection

5.2.3 Bd1833 Structure

We have solved the three-dimensional structure of full-length Bd1833, its constituent domains individually, mutant forms and ligand complexes to aid in the determination of function. A list of Bd1833 structures solved and an associated brief description is presented in Table 13, which is followed by more in depth structural analysis in the following individual sections.

	Construct	Amino acids traced	Spacegroup	Resolution (Å)	Copies in asymmetic unit	R-work	R-free	Oligomeric state	Ligand?	Description/Rationale
Bd1833-ME (Bcs-B6)	1- 434	15-422	P 2_1	1.67	2	0.1702	0.2039	Dimer	$Mg2^+$	High resolution structure of ME domain
Bd1833-ME (Pact-E1)	1- 434	16-427	P 1	2.2	6	0.2109	0.2264	Dimer	-	Lacking divalent cation
Bd1833-ME ($NADP^+$)	1- 434	15-425	P 2	1.67	2	0.1799	0.2159	Dimer	$Mg2^+$ $NADP^+$	NADP+ complex
Bd1833-Pta	439- 780	439- 776	P 2_1	1.96	6	0.1829	0.2052	Hexamer	acetyl-CoA	High resolution structure of Pta domain, inhibited state
Pta R535A	439- 780	439- 776	C 2	2.35	6	0.2158	0.2368	Hexamer	-	Ligand free structure, 3'-P mutant
Pta R535E	439- 780	439- 776	C 2	2.39	6	0.2153	0.2602	Hexamer	-	Ligand free structure, 3'-P mutant
Pta E544R	439- 780	439- 776	C 2	2.33	6	0.2069	0.2265	Hexamer	acetyl-CoA	Interupt hexamer interface
Pta Y576S	439- 780	439- 776	C 2	2.38	6	0.2222	0.2725	Hexamer	-	acetyl-CoA binding pocket distrupted
Pta-Apo	439-780	439-776	P 2_1 2_1 2_1	2.72	12	0.2520	0.2995	Hexamer	-	WT Ligand free structure that shows rigid body conformational change
Bd1833 (Full-length)	1-780	15-776	C 2	2.72	6	0.1922	0.2277	Hexamer	acetyl-CoA	Full length protein
Bd1833-SKB (full-length)	1-796	16-776	P 4_1 2_1 2	3.71	3	0.2503	0.2996	Hexamer	-	Full length of closely related homologue
Bd1833-SKB (Pta)	449-796	449-786	F 2 3	2.71	4	0.2040	0.2392	Hexamer	acetyl-CoA	Homologue Pta domain, CoA binding site differs

Table 13. Bd1833 structures
This table displays a brief structural analysis of each Bd1833 structure reported in this thesis and the rationale behind the structure.

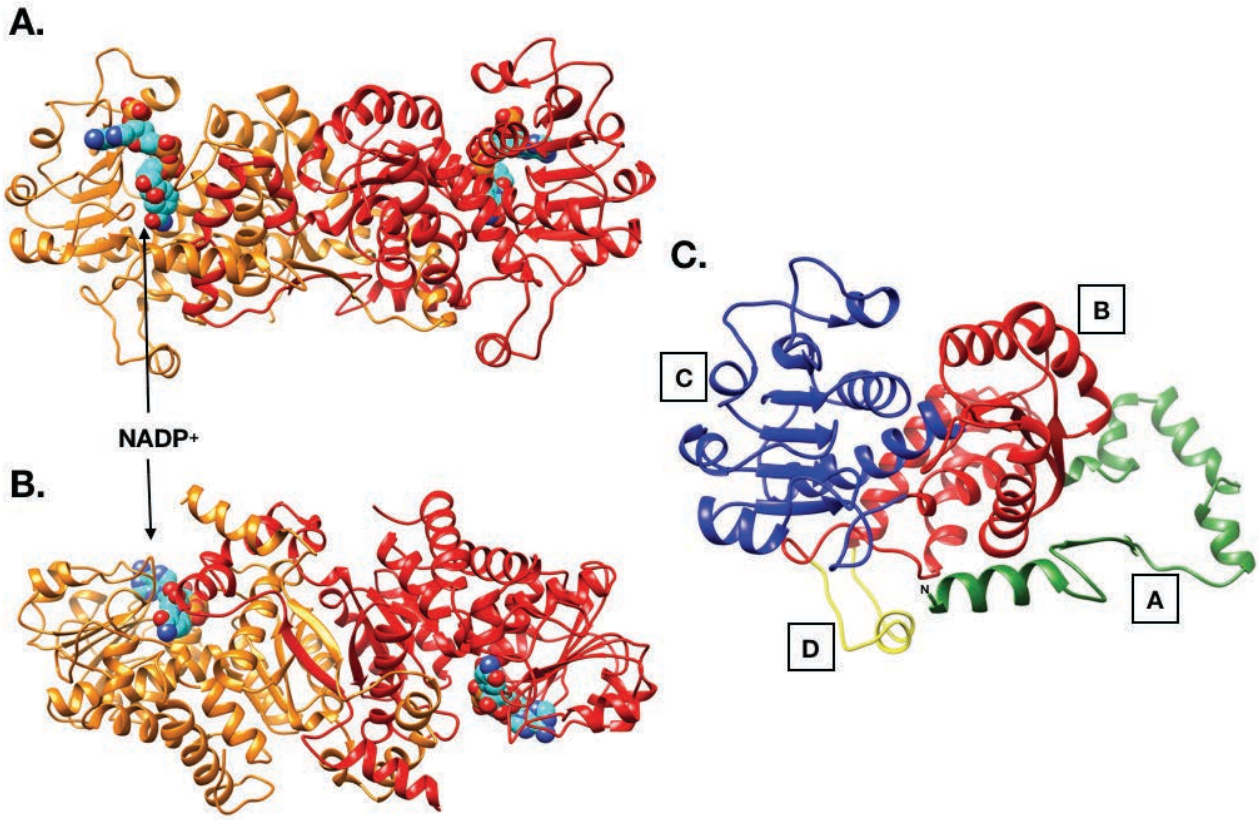


Figure 64. Bd1833-ME structure
A. Bd1833-ME forms a biological dimer (orange and red monomers). The active site is located in the groove between subdomains B and C, which is highlighted by the $NADP^+$ molecule (cyan) bound to subdomain-C. **B.** Illustrates the tightly interacting dimer interface, viewed from a 90° rotation of A around the x-axis. **C.** Illustrates the subdomains that make up Bd1833-ME. Subdomain-A, providing the dimer interface, residues 16-78 (green). Subdomain-B, residues 79-173, 332-368 and 387-420 (red) form the core of the protein containing the divalent cation-binding site. Subdomain-C, residues 174-331 (blue) forms the $NADP^+$ -binding site possessing a Rossmann fold. Bd1833-ME has a small loop specific to multi-modular malic enzymes, subdomain D 369-386 (yellow).

5.2.4.2 ME catalytic site

The catalytic site of Bd1833-ME is positioned between subdomains-B and C, which is highlighted by the position of $NADP^+$ in Figure 64. The catalytic pocket of Bd1833 is composed of a divalent cation-binding motif, a nucleotide-binding site and the substrate-binding site. The divalent cation-binding motif is central to the catalytic site and six oxygen atoms coordinate Mg^{2+} in an octahedral fashion (Figure 65). The coordinating oxygen atoms are provided by the sidechain carboxylate groups of E143

residues from both monomers of the dimer complex. The catalytic residue, Y52, located on subdomain-A, is provided by the other monomer. A single catalytic site containing residues from both monomers is typical of prokaryotic MEs [1] but not of eukaryotic ME. Moreover the proposed substrate-binding site residues are also slightly different to eukaryotic MEs, for example lacking the active-site arginine that assists in substrate binding [225].

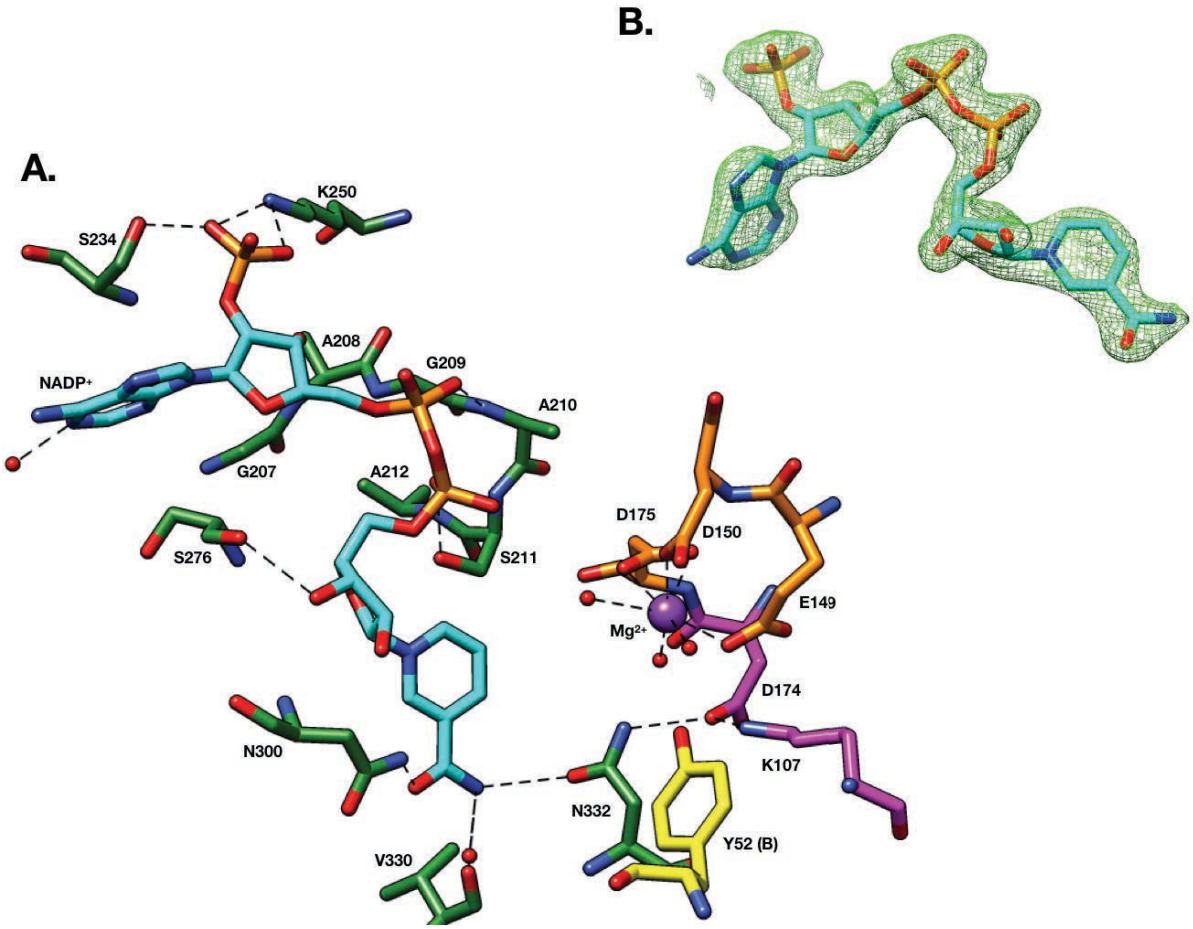


Figure 65. Bd1833 active site
A. The active site of Bd1833-ME is made up of residues from both monomers. The $NADP^+$ -binding groove located between subdomains B and C. Subdomain-C contains the Rossmann fold and GXGXXG nucleotide-binding motif. The residues that make hydrogen bond contacts with $NADP^+$ (cyan) are shown in green. The active-site residues are shown with magenta coloured carbons, while Y52 provided by the other monomer is shown with yellow carbons. The metal coordinating residues are shown in orange, bound to the divalent cation (Mg^{2+}). All residues shown, except for Y52, belong to a single monomer (chain A). All hydrogen bonds illustrated by dashed black lines. **B.** The electron density observed for $NADP^+$ in the Bd1833-ME: $NADP^+$ soaked crystal structure. ($2F_o$-F_c map, radius =2Å, σ = 1)

737) folds into a six-stranded mixed β-sheet that is sandwiched between two α-helices on either side. Subdomain-II is positioned centrally in the Pta dimer arrangement, forming the dimer interface. A shallow groove splits subdomain-I and II and is the location of the substrate-binding pocket where the co-purified and co-crystallised acetyl-CoA is located (Figure 67).

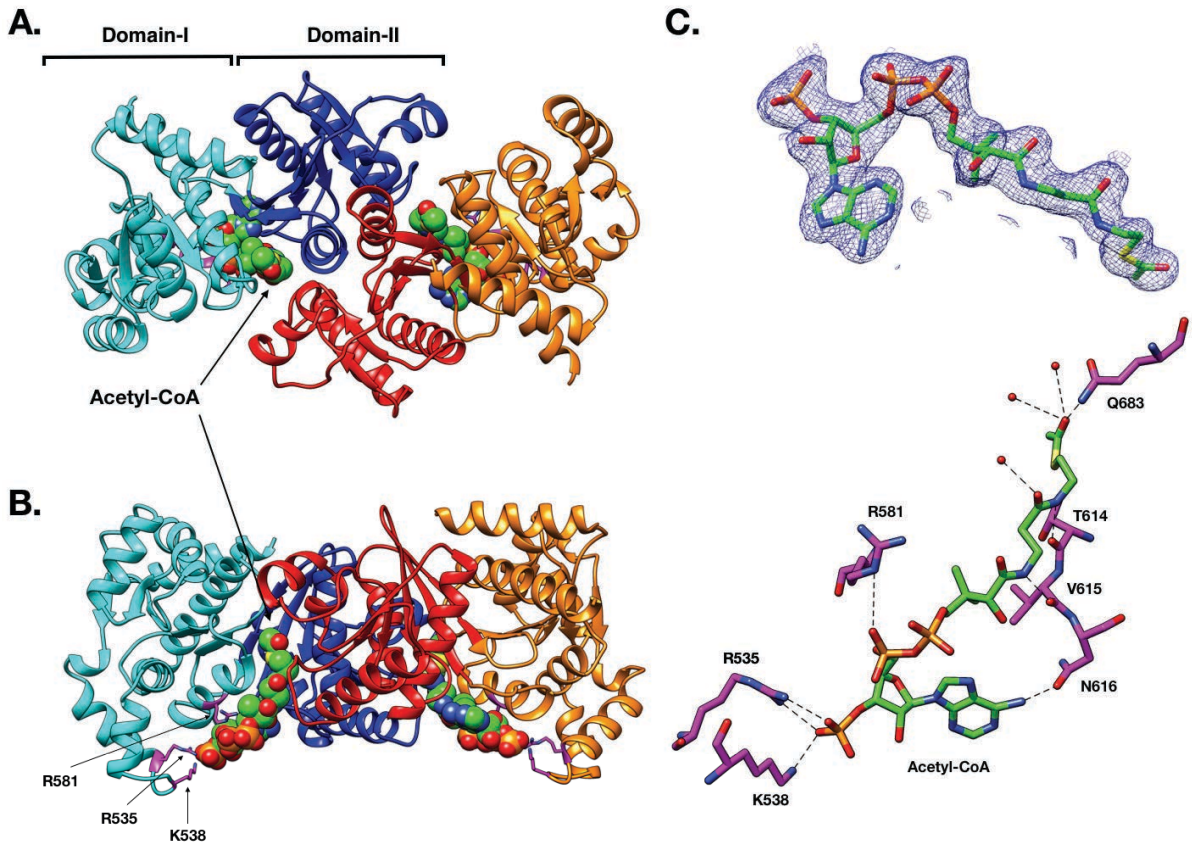


Figure 67. Bd1833-Pta dimer and acetyl-CoA binding
A. Bd1833-Pta forms a hexamer (trimer of dimers) where the dimers are structurally similarly to other previously solved Pta structures represented here as a blue:cyan and a red:orange dimer. Bd1833-Pta monomer can be split into two subdomains: subdomain-I, residues 440-595 and 734-776 (cyan) and the central subdomain-II, 596-733 residues (blue). Central to the dimer is an almost parallel alpha helix that is the main contributor to the dimer interface. Acetyl-CoA (green) binds in the substrate-binding groove located between subdomains-I and II. The conserved CoA binding residues (Pta enzymes) are highlighted in magenta making hydrogen bonds to acetyl-CoA. **B.** Rotation of image A 90° around the x-axis. **C.** The electron density observed for acetyl-CoA ($2F_o$-F_c map, radius=2Å, σ=1). **D.** Illustration of the hydrogen-bonding network associated with acetyl-CoA binding.

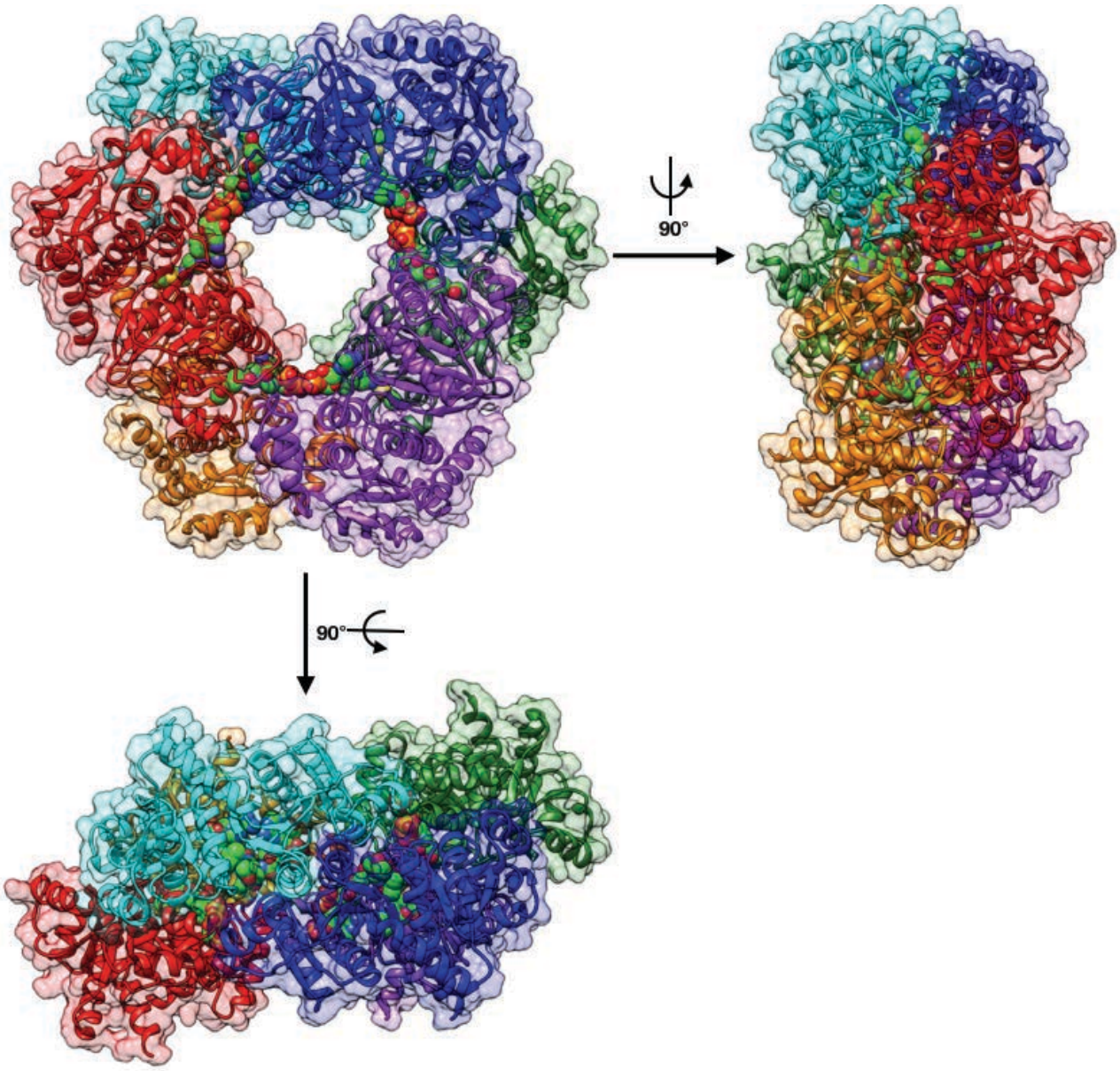


Figure 66. Bd1833-Pta structure
Bd1833-Pta forms a unique hexamer complex made of a trimer of dimers. The hexamer is organised into an offset triangle shape where the dimers form the flat sides. Dimers are coloured – blue:cyan, red:orange and green:purple. Each monomer makes contacts to four other monomers. The hexamer complex has six acetyl-CoA ligands bound, located on the inside face at the corners of the triangle formation. The hexamer complex is displayed in three orientations, rotated by 90°.

5.2.5.2 Dimer and tetramer interfaces

The dimer interface, 1707Å^2 buried surface area, is composed mainly of residues from subdomain-II. In particular two parallel α-helices (713-730) are positioned centrally to the dimer interface. Each monomer of the dimer provides a single α-helix to form the

parallel α-helix pair that makes substantial hydrophobic interactions to stabilise the dimer interface. In addition H-bonds further stabilise the dimer interface (Figure 68).

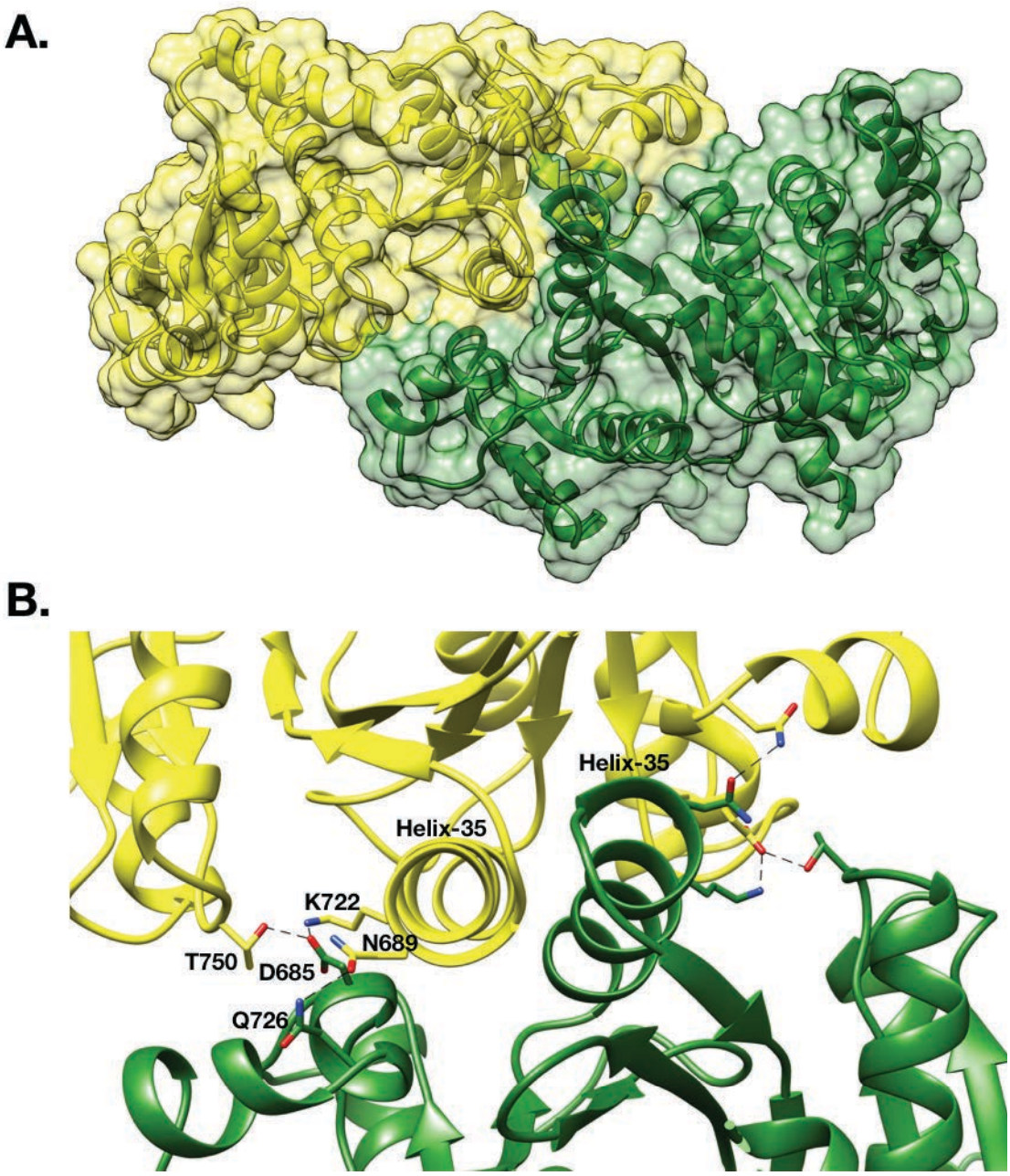


Figure 68. The dimer interface
A. Surface view of dimer interface. **B.** View of dimer interactions. The dimer interface is mainly formed by interaction between subdomain-II; a central alpha helix pair (helix-35) runs almost parallel and creates a large buried hydrophobic core. A number of hydrogen bonds are formed between residues from the central α-helix and other residues of subdomain I and II.

The hexamer complex, made of a trimer of dimers, is organised into an offset triangular shape, with dihedral D3 symmetry (3-fold symmetry with addition 2-fold axis). An extensive hexamer interface exists at the meeting point of two dimers at the triangle's vertices. Consequently a single Pta domain monomer makes interactions to four other monomers through dimer and hexamer interfaces. The hexamer interface covers numerous overlapping interactions at the triangle vertex, which extends up to the midpoint (Figure 69). The larger hexamer interface, at the vertex point (hexamer vertex contact) has a buried surface area of 644Å^2. The hexamer interface also includes two smaller contacts at the midpoint (hexamer midpoint contact) that has a buried surface area of 322Å^2 (buried surface area calculated by PDB:sum).

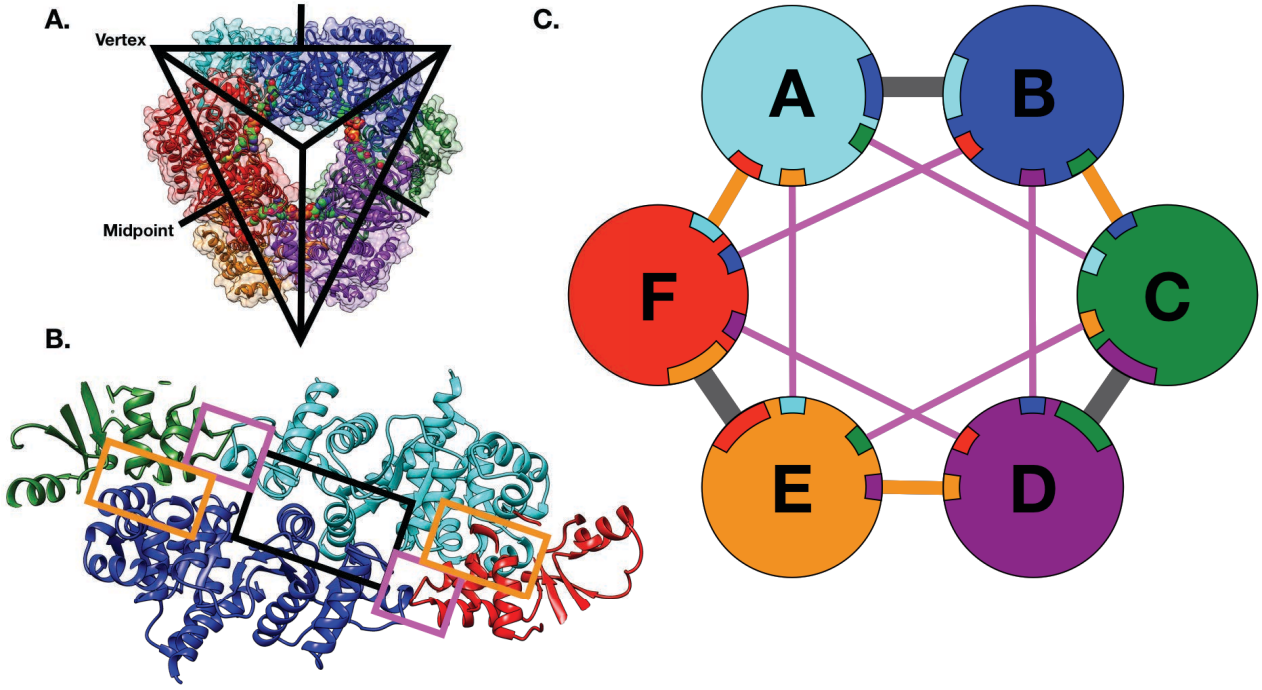


Figure 69. Hexamer interactions
A. Bd1833-Pta hexamer forms a triangular shape with dihedral D3 symmetry. **B.** Pta dimer viewed from the inside face demonstrating the hexamer surface contacts. The contacts in coloured boxes match the interactions shown in (C): Orange = larger hexamer vertex contact. Pink = smaller hexamer midpoint contact. Grey = dimeric contacts. **C.** A single monomer makes contacts to four other monomers through interactions at the dimer interface and hexamer interface. The hexamer interface is composed of four interacting monomers and two different surface contacts. The larger hexamer vertex contacts (orange line) are located at the hexamer verteces. The smaller hexamer midpoint contacts (pink lines) are made between the tips of subdomain-I helical flap region and the edge of subdomain-II.

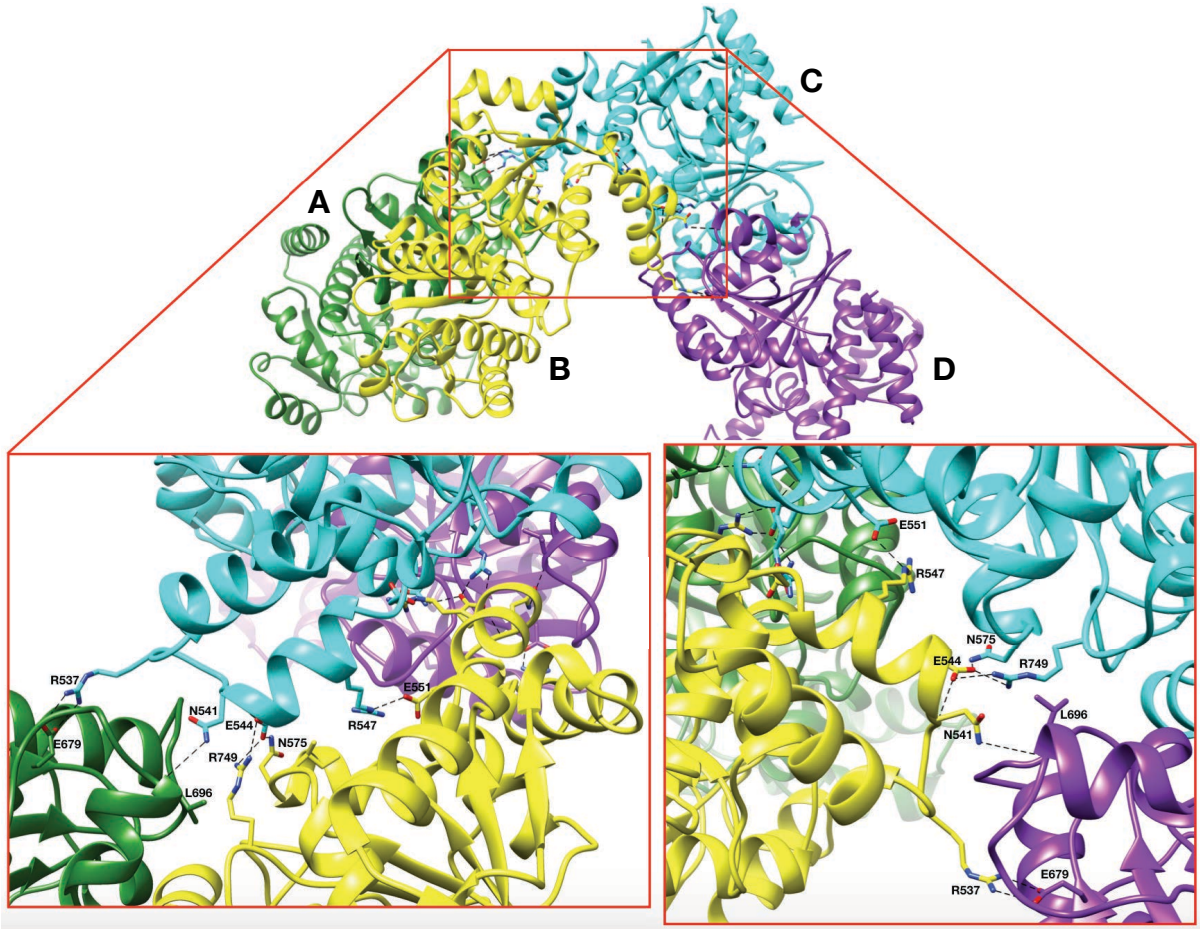


Figure 70. The hexamer interface
Bd1833-Pta forms a hexamer and each monomer forms a series of interactions with thee other monomers at the hexamer interface. This Figure illustrates a number of hydrogen bonds and ionic interactions that occur at the hexamer interface. The upper panel shows the overall hexamer interface. The left and right lower panels show a zoomed in view of the hydrogen bond network at the hexamer interface from two viewpoints. Chain A (green), Chain B (yellow), Chain C (cyan) and Chain (D) (purple). Chains A & B form a dimer and chain C & D form the other dimer that meet at the hexamer interface.

5.2.5.3 Acetyl-CoA binding site

Bd1833-Pta co-purifies and co-crystallises with acetyl-CoA, which binds in the shallow substrate-binding groove separating subdomain-I and II. One molecule of acetyl-CoA binds per monomer, hence the Pta hexamer complex has six molecules of acetyl-CoA ligand bound, located on the interior face at the triangle vertices. A number of non-bonding (vdw) interactions as well as specific hydrogen bonds are involved in acetyl-CoA binding, illustrated in Figure 67. The adenine group of acetyl-CoA makes a single hydrogen bond with the N616 sidechain. The side chain of R535 makes an important

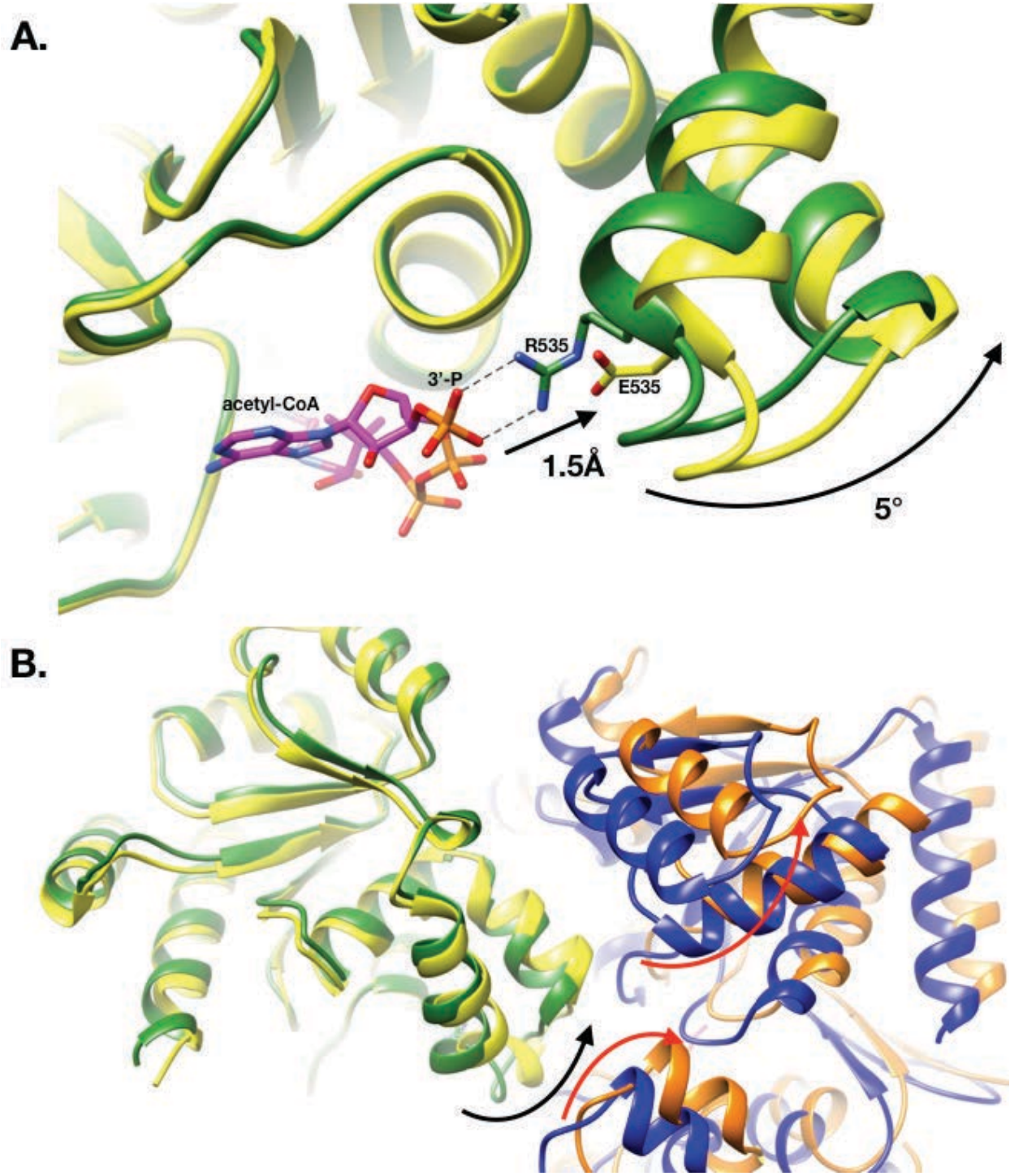


Figure 71. R535 mutation
A. Mutating R535 to a glutamate residue abolishes acetyl-CoA binding by repelling the 3'-phosphate group. R535E (yellow) mutation causes the lower loop (helix 27 and 28) of Bd1833-Pta to swing outwards when compared to WT- acetyl-CoA bound structure (green). **B.** The outwards swing (black arrow) causes a larger adjustment of the hexamer interface interactions (mainly lost from the vertex interface), shown by the red arrows. Blue is the position of hexamer interface in Bd1833-Pta (acetyl-CoA bound) and orange is the position of the hexamer interface in Pta-R535E mutant. A similar movement is observed for Bd1833-Pta R535A. The SSM superpose tool in coot was used to align chain A from Bd1833-Pta and Pta-R535E (length = 336 residues, rmsd = 0.67Å).

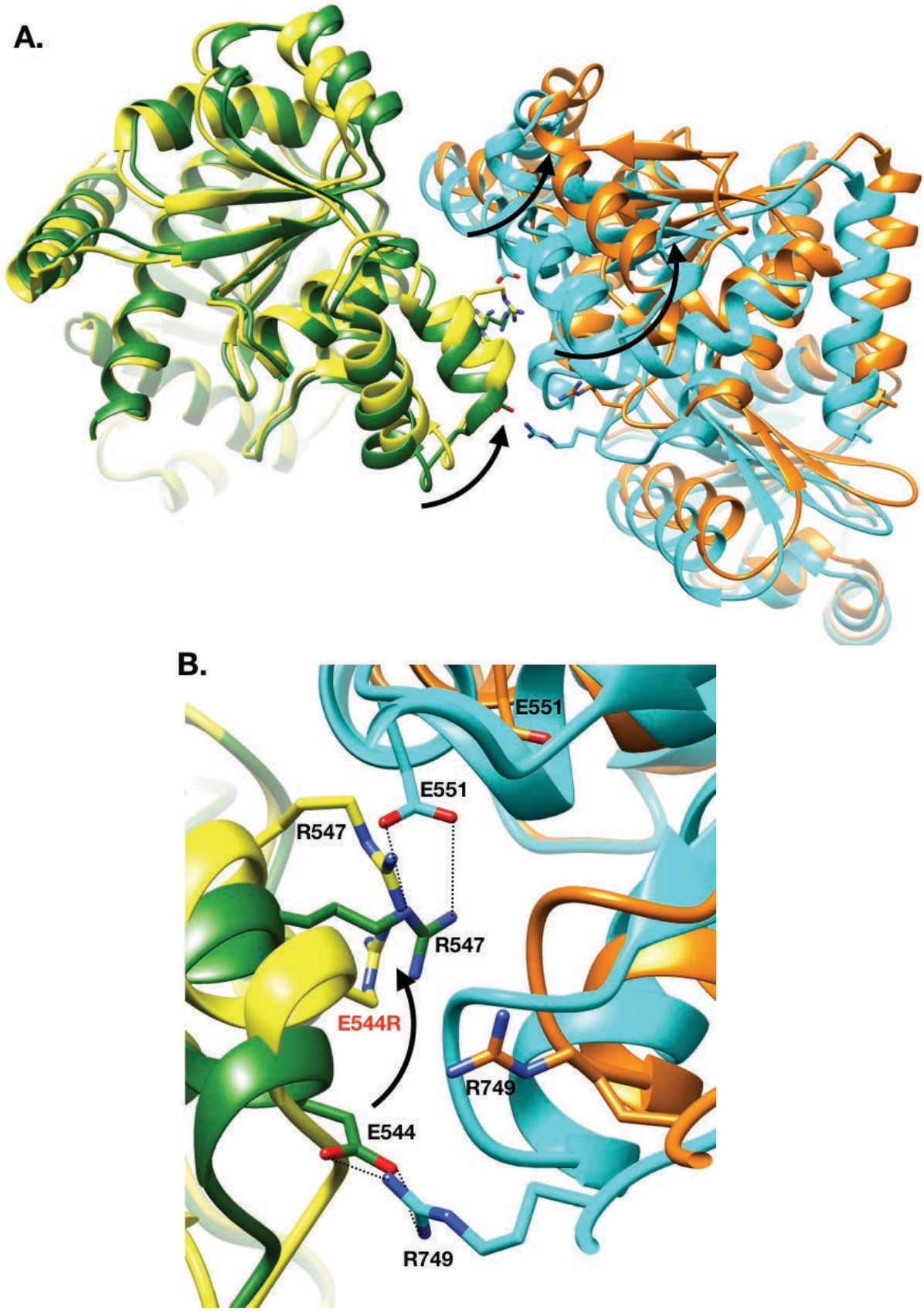


Figure 72. Pta-E544R
A. The hexamer interface of E544R is destabilised and causes a conformational reorientation between interacting dimers. Bd1833-Pta (green and cyan) and Bd1833-E544R (yellow and orange). **B.** E544R mutation is repelled by R749 but folds back inwards adopting a similar location as R547. Superimposition was performed by least-squares minimisation using the CCP4 suit superpose software. Alignment was performed for Pta dimers using subdomain-II (residues 596-738) from both protomers (rmsd = 0.55Å).

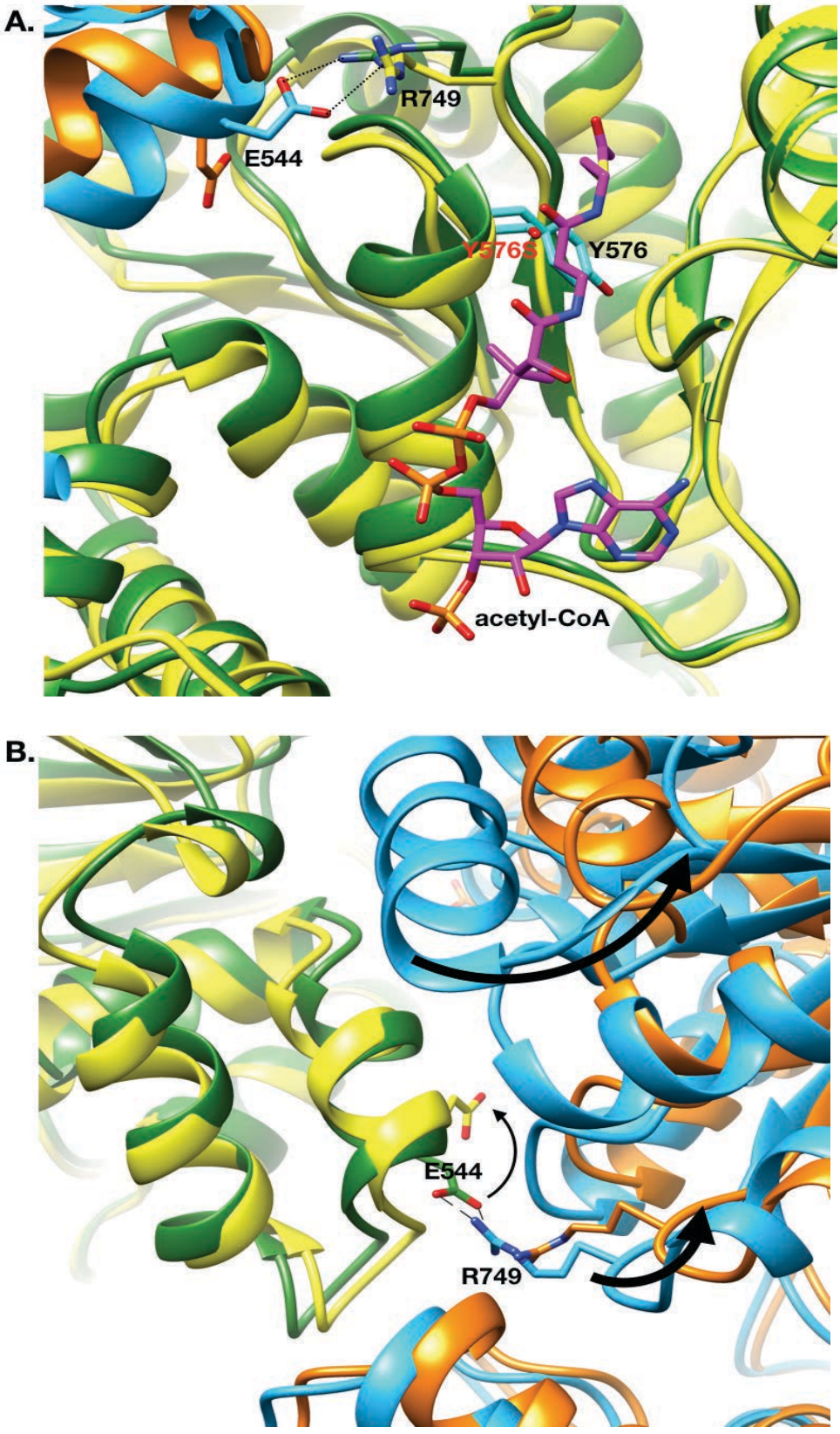


Figure 73. Y576S mutation
A. Y576 is located close to the thiol group of acetyl-CoA and is a highly conserved serine residue in Pta enzymes. Y576S selectively prevents acetyl-CoA binding but allows CoA binding. The structure of Pta-Y576S (yellow and orange) superimposed with Bd1833-Pta (green and cyan). **B.** Y576S mutation selectively prevents acetyl-CoA binding, which changes the conformation of subdomain-I and consequently destabilises the hexamer interface. Superimposition was performed by least-squares minimisation using the CCP4 suit superpose software. Alignment was performed for Pta dimers using subdomain-II (residues 596-738) from both protomers (rmsd = 0.72Å).

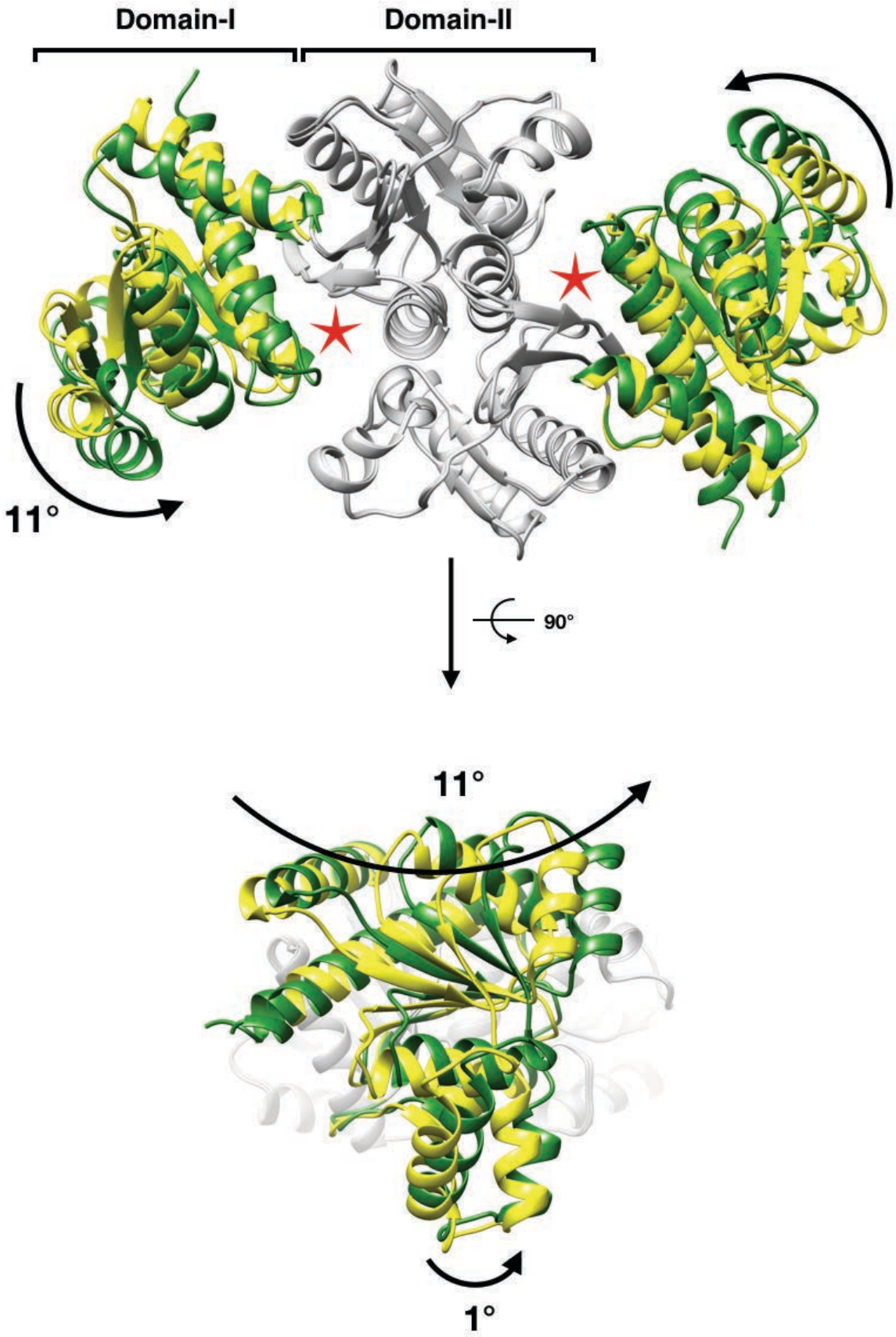


Figure 74. Bd1833-Pta domain movement
Bd1833-Pta (green) and Pta-Apo (yellow) structures are shown, superimposed by subdomain-II (grey). A red star marks the substrate binding sites. Subdomain-I undergoes an 11° rotational rigid body conformational change, indicated by the black curved arrows. The movement changes accessibility to the substrate binding groove located between subdomain-I and II. A sideview of the same rotational rigid body movement demonstrates the upper region undergoes a larger rotational change compared to the lower region. Superimposition was performed by least-squares minimisation using the CCP4 suit superpose software. Alignment was performed for Pta dimers using subdomain-II (residues 596-738) from both protomers (rmsd = 0.82Å).

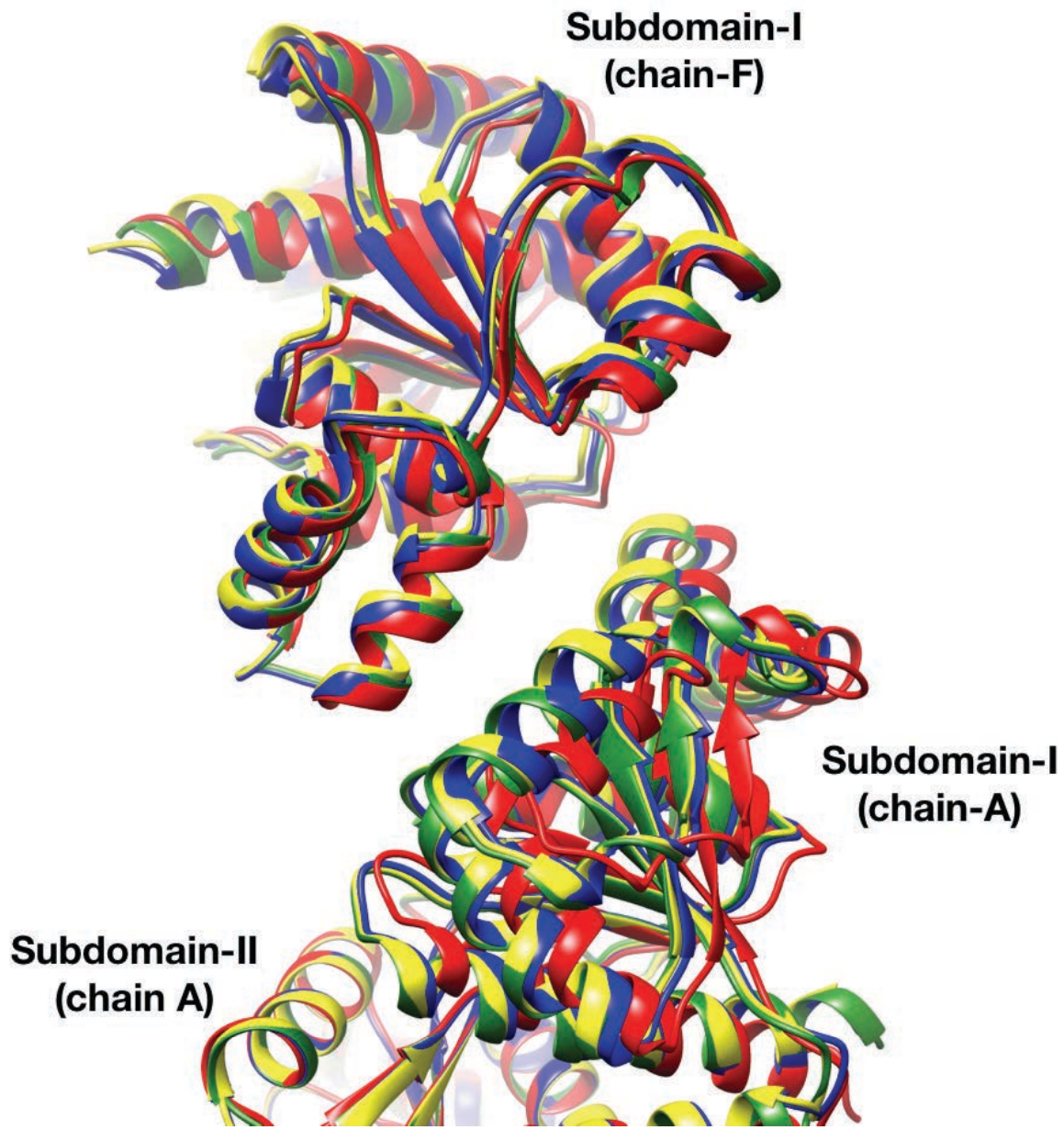


Figure 75. Pta mutations and hexamer interface
The hexamer vertex interface between the AB dimer and EF dimers is shown for R535E (blue), R544R (yellow) and Q576S (green) mutants aligned with Pta-Apo (red). The hexamer is aligned around subdomain-II in chain-A. The degree of hexamer rearrangement affecting the vertex interface is similar in all mutations and Apo form. The vertex hexamer interface is reduced from 644Å^2 in Bd1833-Pta in complex with acetyl-CoA to ~240Å^2 in ligand free Bd1833-Pta structures, which is more than a two-fold decrease.

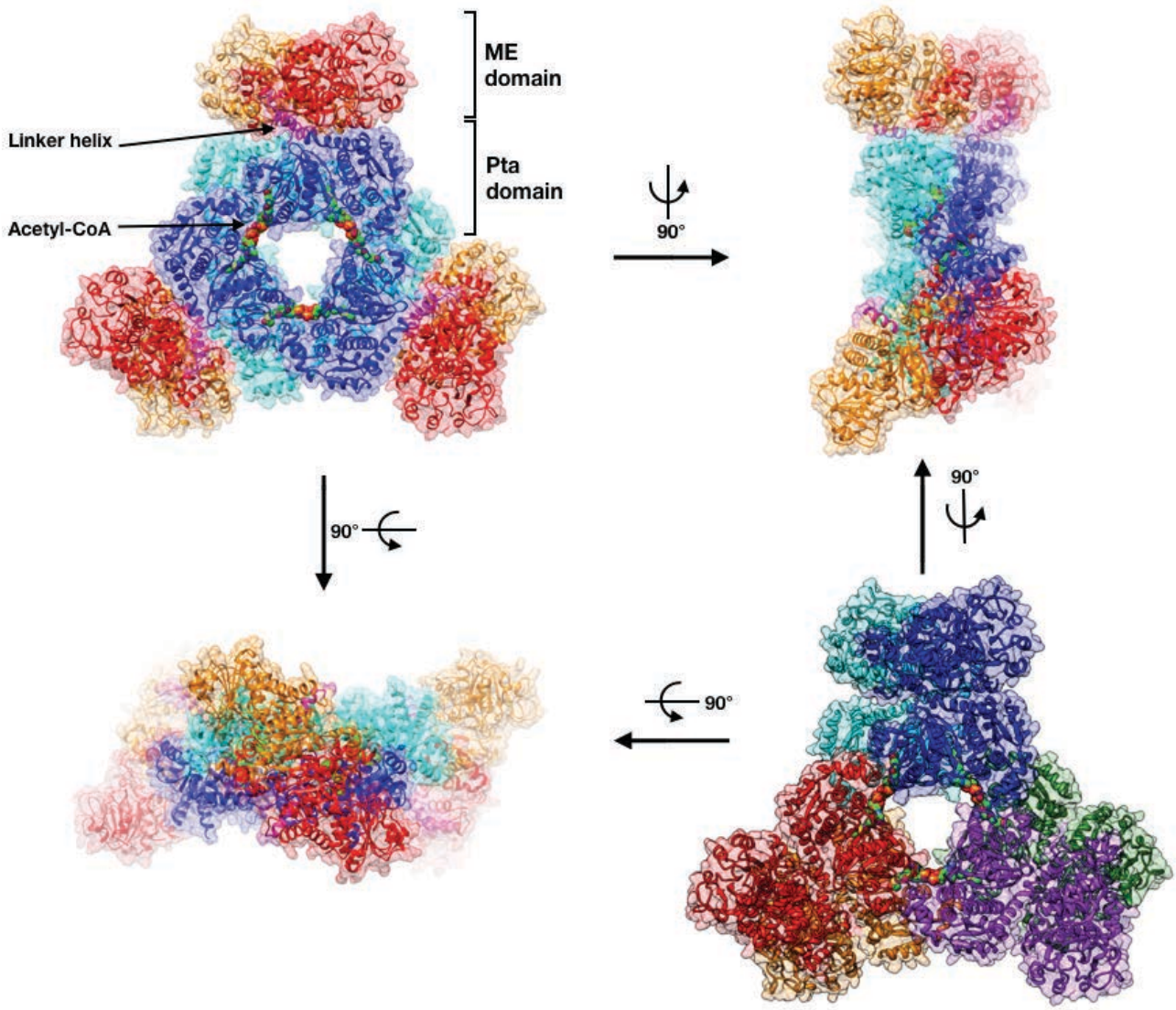


Figure 76. Bd1833 structure
Full-length Bd1833 protein forms a larger hexamer complex. The top left panel shows the hexameric structure of Bd1833 coloured according to the domains (red/orange = ME and blue/cyan = Pta). The linker helix is shown in magenta and acetyl-CoA ligand in green. Top right and bottom left panels show the hexameric structure in two different orientations rotated by 90° demonstrating the twisted nature of the hexamer complex and how the ME domain is positioned on top of the Pta domain, offset by ~40°. Finally the bottom right panel shows Bd1833 hexamer coloured according to the protomers that form the hexamer complex.

The Pta and ME domains are held together by the linker helix which is located on either side of the Bd1833 dimer (Figure 77). There are minimal interactions between the Pta surface and ME surface and in the central region of the dimer a 6Å gap exists between the surfaces of both domains. A few specific H-bonds are formed between the Pta domain and the ME domain. These include: R444 with F396 backbone carbonyl and

K440 with K112 backbone carbonyl. The majority of the surface interactions between Pta and ME domains are electrostatic, involving charged and polar residues (Figure 77).

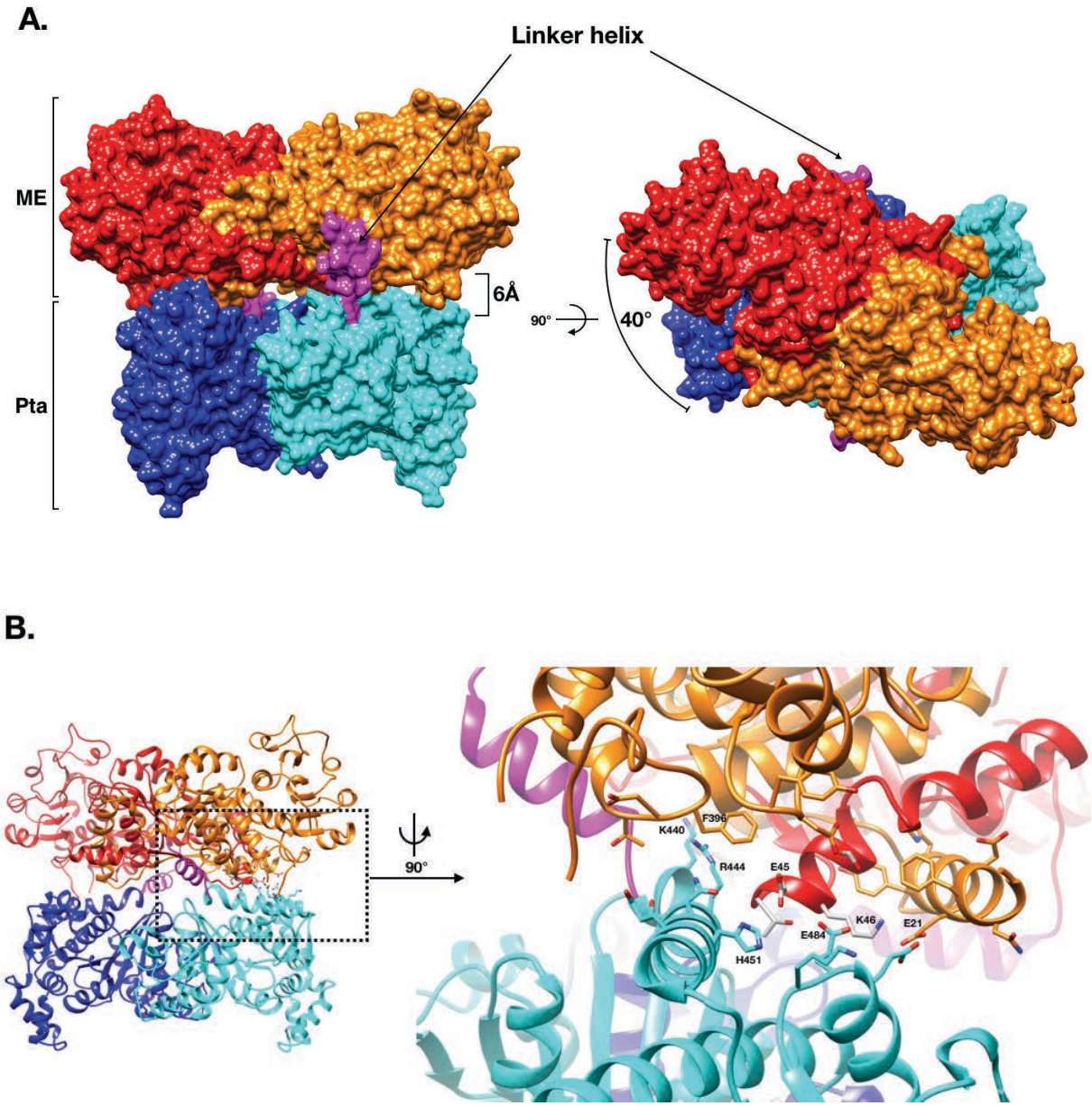


Figure 77. Pta and ME interface
A. The Pta and ME domains in full-length Bd1833 are linked by a single linker helix, located on either side of the Bd1833 dimers. Little surface contact is made between the two domains that are offset by ~40° relative to one another. In the central regions a gap up to 6Å exists between the domain surfaces. **B.** Minimal interactions exist between the Pta and ME domain. The interactions that are present are mainly electrostatic and formed between subdomain-A/C of the ME domain and subdomain-I of the Pta domain. The residues contributing to the minimal interface between the Pta and ME domains are displayed.

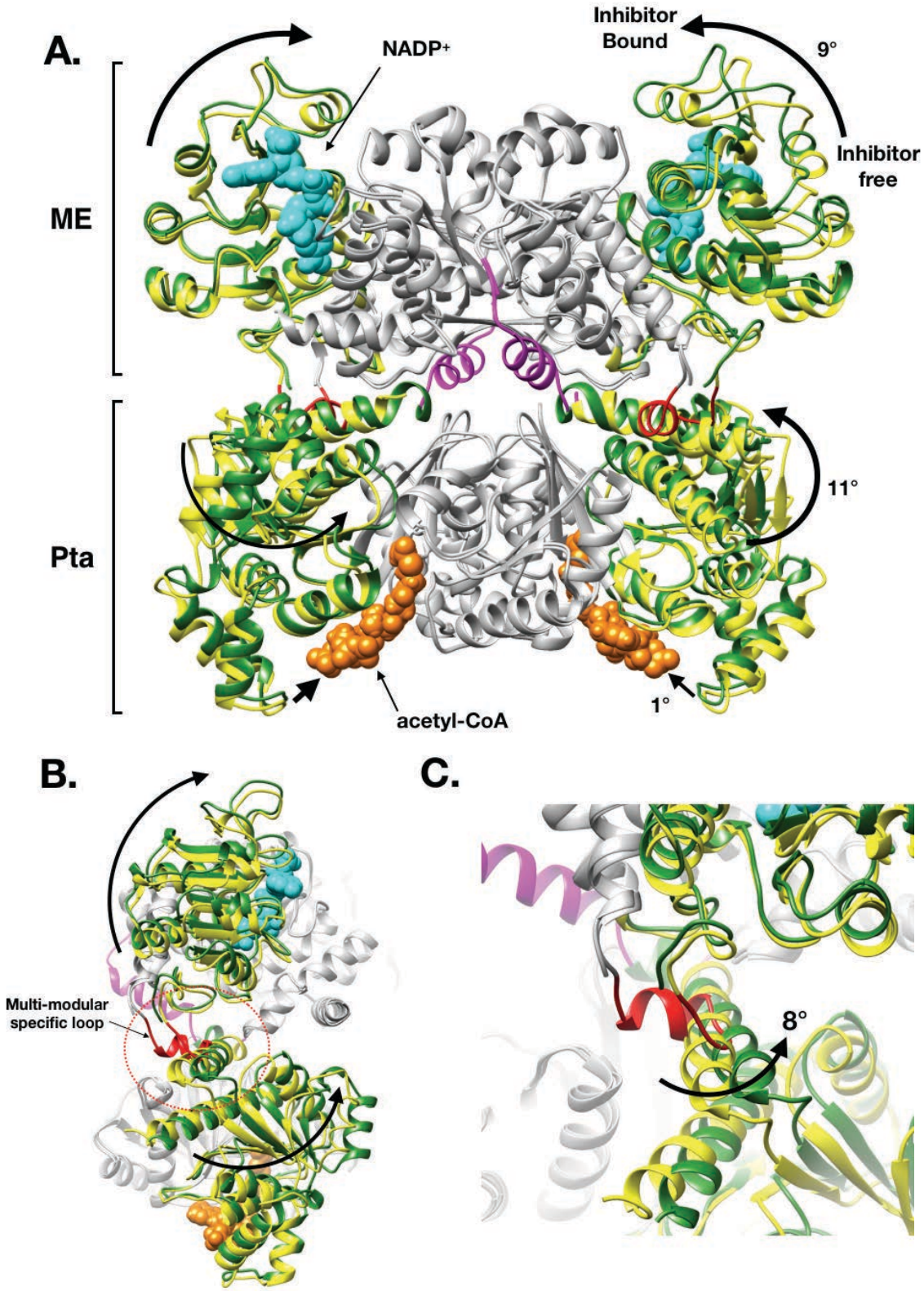


Figure 78. Bd1833 Allosteric regulation

A. The structure of full-length Bd1833 in complex with acetyl-CoA (green) is superimposed with Pta-Apo (yellow, length = 1090 atoms, rmsd = 0.86Å) and Bd1833-ME-NADP (yellow, length = 6670 atoms, rmsd = 0.47Å). The core region of Bd1833 (ME subdomain A and B and Pta subdomain-II) does not alter conformation between structures and is coloured in grey for clarity. The linker helix is shown in magenta and the multi-modular specific loop region (ME subdomain-D) is shown in red. The location of Pta and ME binding sites are represented by $NADP^+$ (cyan) and acetyl-CoA (orange) to gauge the distance between binding sites. An 11° rigid body rotational movement of Pta subdomain-I allosterically causes a 9° rigid body rotational movement in ME subdomain-C, restricting access to the active site. **B.** Demonstrates the conformational changes in full-length Bd1833, viewed from the side. This image demonstrates the clash with Bd1833-ME subdomain-D (red). **C.** A close up view of the clash with subdomain-D; subdomain-D which is not traceable in the structure of full-length Bd1833. Superimposition was performed by least-squares minimisation using the CCP4 suit superpose software. Alignment was performed for Bd1833 dimers using residues 34-174 & 596-738 from both protomers.

differences. K581 represents the only distinctive change, which does not appear to make an interaction with the pyrophosphate group of acetyl-CoA (Figure 79.B).

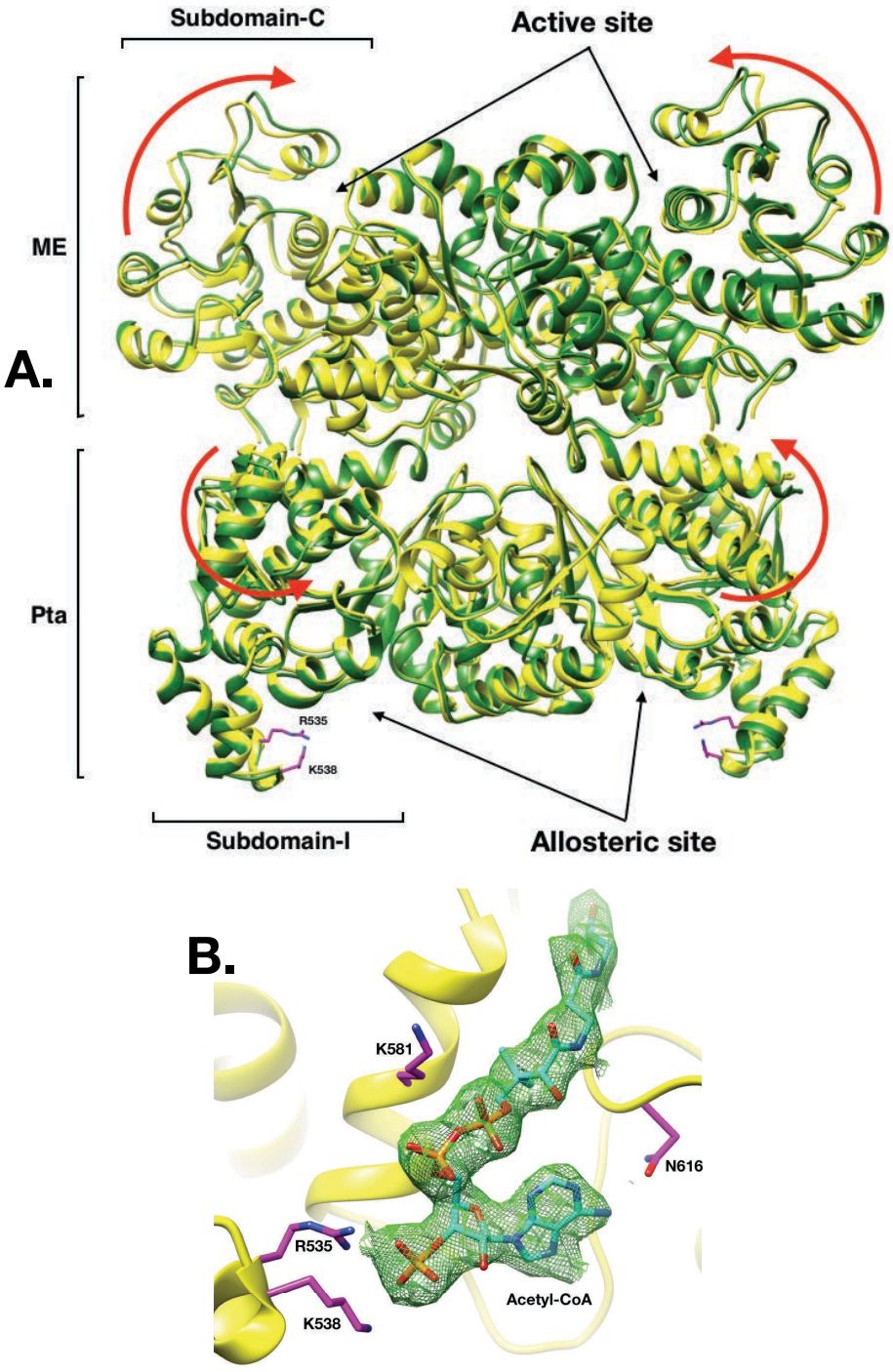


Figure 79. B1833-SKB structure
A. The structure of Bd1833-SKB (Yellow) is superimposed with the structure of Bd1833 (Green)(length = 23032 atoms, rmsd 0.55Å). Bd1833-SKB does not co-purify with ligand and likely represents a non-inhibited full-length conformation. The ME domain active site and Pta allosteric site are highlighted. Red arrows demonstrate the rigid body movements in subdomain-I and subdomain-C. Superimposition was performed by least-squares minimisation using the CCP4 suit superpose software. Alignment was performed for Bd1833 dimers using residues 34-174 & 596-738 from both protomers. **B.** The acetyl-CoA interaction with Pta-SKB (truncation mutant of Bd1833-SKB that co-purifies with ligand). K581 does not interact with the pyrophosphate group. ($2F_o$-F_c map, radius=2Å, σ=1).

Bd1833 species	Kinetic Parameters (malate)		Kinetic Parameters (NADP+)		Oligomeric State	Acetyl-CoA Inhibition	Co-purified Ligand?	Reasoning
	K_M (mM)	k_{cat} (s^{-1})	K_M (µM)	k_{cat} (s^{-1})				
Bd1833	2.61 ± 0.28	26.1 ± 1.2	21.7 ± 2.3	21.1 ± 0.13	Hexamer	Yes	CoA + AcCoA	
Bd1833-Apo	3.02 ± 0.31	35.2 ± 1.4	43.1 ± 5.2	33.5 ± 1.3	Hexamer	Yes	None	Ligand free
Bd1833-ME	4.15 ± 0.50	3.28 ± 0.17			Dimer	No	None	ME domain only
Bd1833-Δloop	12.5 ± 1.0	1.62 ± 0.09			ND	No	CoA + AcCoA	Chimera lacking multi-modular specific loop
Bd1833-R535E	3.17 ± 0.15	19.4 ± 0.36	41.1 ± 4.0	16.8 ± 0.54	Hexamer	No	None	Prevent CoA binding (3'-P)
Bd1833-R535A	3.48 ± 0.23	20.3 ± 0.53			Hexamer	No	None	Prevent CoA binding (3'-P)
Bd1833-N718D	3.35 ± 0.33	41.8 ± 1.6			Hexamer	No	None	Prevent CoA binding (acetyl-group)
Bd1833-E544R	12.6 ± 2.0	18.9 ± 1.9			Dimer	No	None	Destabilise hexamer
Bd1833-E544A	4.81 ± 0.44	38.7 ± 1.6			Hexamer	Yes	CoA + AcCoA	Destabilise hexamer
Bd1833-Y576S					Hexamer	No	CoA	Convert back to conserved Serine residue
Bd1833-Y576C					ND	Yes	CoA	Interact with thiol group
Bd1833-R749A					ND	Yes	CoA + AcCoA	Destabilise hexamer
Bd1833-R547A					ND	Yes	CoA + AcCoA	Destabilise hexamer
Bd1833-R581K					ND	Yes	CoA + AcCoA	Pyrophosphate interaction
Bd1833-SKB	1.96 ± 0.24	35.0 ± 1.3	45.4 ± 6.6	39.3 ± 1.8	Hexamer	Yes	None	Close homologue

Table 14. Bd1833 kinetics
This table includes the values obtained from our kinetic studies using malate as the substrate, the oligomeric state and whether inhibition by acetyl-CoA was detected for each Bd1833 species. ND = not determined. Values presented as a mean of triplicates alongside the standard deviation.

5.2.8 Kinetics

5.2.8.1 Bd1833 activity

A detailed kinetic analysis of Bd1833 has been performed. ME activity was detected for both malate oxidative decarboxylation and pyruvate reductive carboxylation reactions; however, no Pta activity was detected. Table 14, displays the kinetic parameters obtained for Bd1833 and Bd1833-mutant species.

5.2.8.2 Malate Oxidative Decarboxylation

Bd1833 required a divalent cation such as Mn^{2+} to be catalytically active. Bd1833 also demonstrates specificity for $NADP^+$ as a cofactor, showing no activity in the presence of NAD^+. Kinetic parameters were determined using L-malate as the substrate, at a concentration between 0.1 and 10 mM; reduced rates were observed beyond 10 mM L-malate (Figure 80).

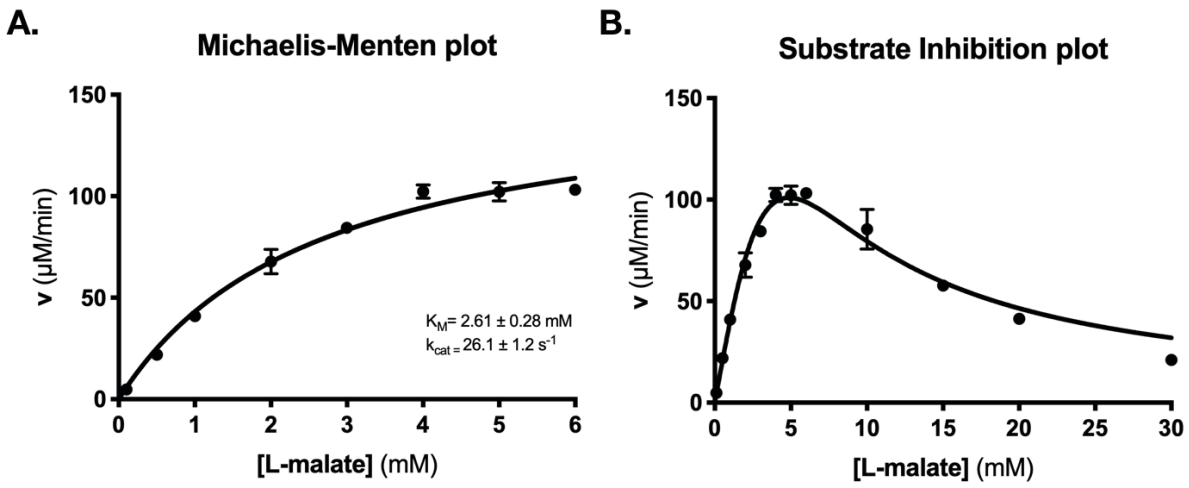


Figure 80. Malate oxidative decarboxylation
Kinetic data shown for full-length Bd1833. **A.** The data for Bd1833 fit to a Michaelis-Menten plot. **B.** At concentrations of malate above 8 mM, reduced rates are observed. The data has been fit to substrate inhibition model $Y=V_{max}*[\text{L-malate}]/(K_M + [\text{L-malate}]*(1+[\text{L-malate}]/K_i))$ using prism7 software to explain the reduced rates. The fit is "ambiguous" suggesting the model is not ideal to explain the data, suggesting other factors such as ΔpH or ionic strength may also contribute. All data points were measured in at least triplicates. The error bars represent the standard deviation.

5.2.8.3 Effectors

A series of potential effectors, published for other malic enzymes, were tested as potential regulators of Bd1833. The activity of Bd1833 in the presence of effectors is presented as a percentage of the activity of Bd1833 without effector (Figure 81). Acetyl-CoA and OAA are shown to be potent inhibitors of Bd1833 activity and reduce the activity to below 10% of the un-inhibited activity. CoA and acetyl-P reduced the activity to ~50% suggesting a weaker mode of inhibition. No activators of Bd1833 activity were identified.

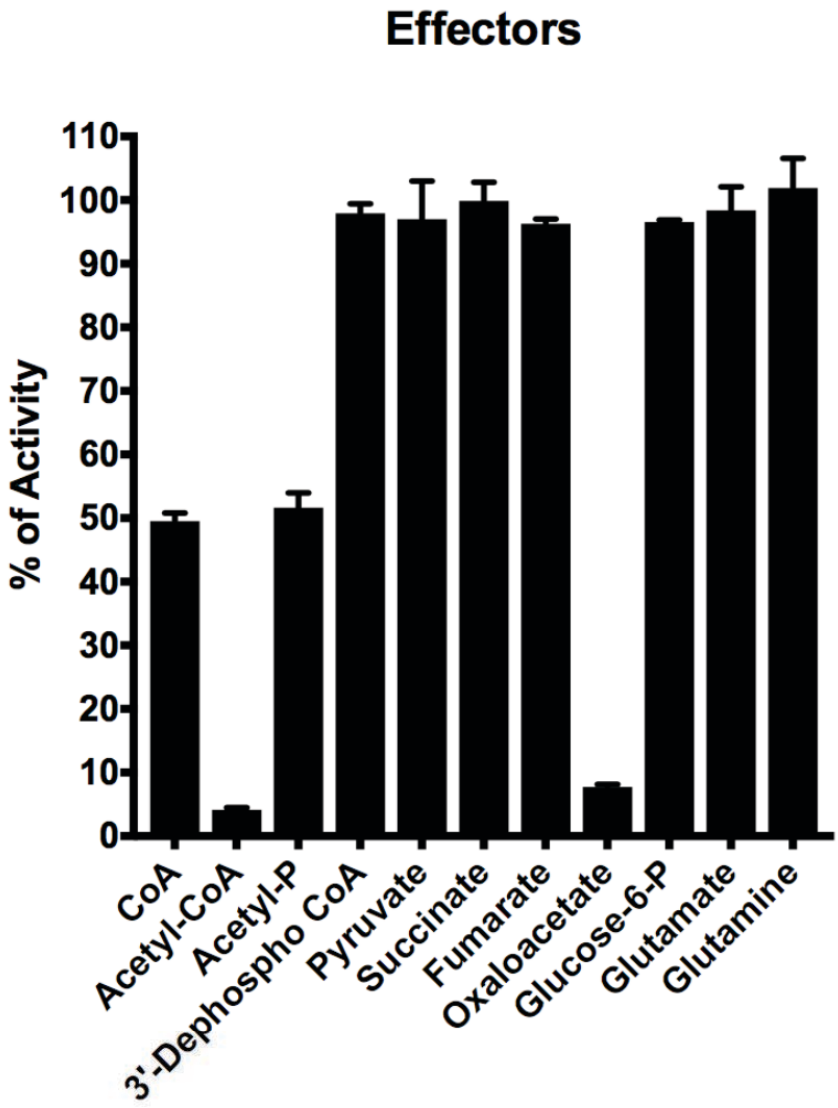


Figure 81. Identifying Bd1833 effectors
A range of malic enzyme effectors were tested for their effects on Bd1833 activity. Bd1833 activity is reduced to ~50% in the presence of CoA and acetyl-P. Acetyl-CoA and Oxaloacetate are potent inhibitors of Bd1833 activity, reducing the activity to below 10% of Bd1833 activity in standard assay conditions. All data points were measured in at least triplicates. The error bars represent the standard deviation.

5.2.8.4 Effector binding sites

To investigate how these effectors regulate the activity of Bd1833 we created a number of Bd1833 mutants. The truncation mutant (Bd1833-ME) activity was measured in the presence of the effectors and compared to the activity without (Figure 82). Acetyl-CoA, CoA and acetyl-P showed no inhibitory effects towards truncated Bd1833-ME; however, OAA remained an inhibitor. This result confirms that acetyl-CoA, CoA and acetyl-P interact with the Pta domain and allosterically regulate ME activity, whereas OAA interacts directly with the ME domain, which supports our structural data. Interestingly, the activity of Bd1833-ME was reduced in comparison to full-length Bd1833, suggesting communication with the Pta domain is important for catalysis.

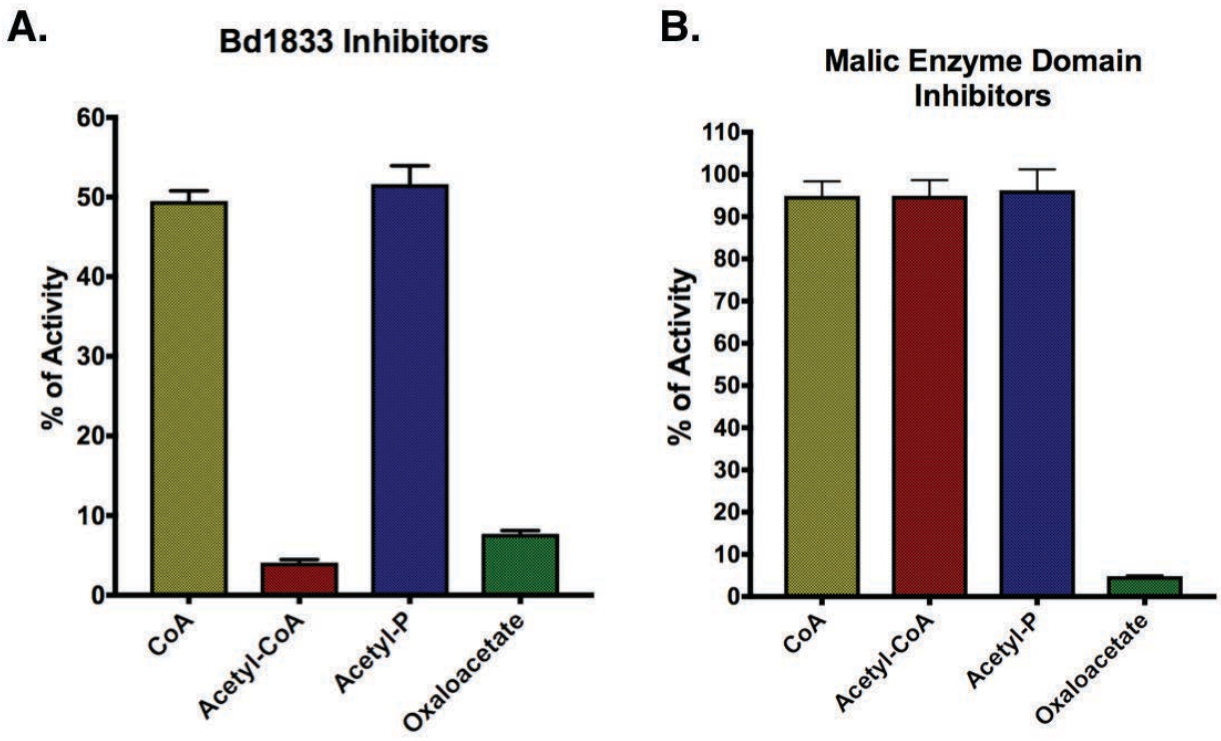


Figure 82. Bd1833 inhibitors
A. The activity of Bd1833 in the presence of inhibitors. **B.** The activity of Bd1833-ME truncation mutant in the presence of the same effectors. Three of the effectors, including acetyl-CoA, do not inhibit Bd1833-ME whereas oxaloacetate still inhibits. This finding indicates which domains the inhibitors interact with. All data points were measured in at least triplicates. The error bars represent the standard deviation.

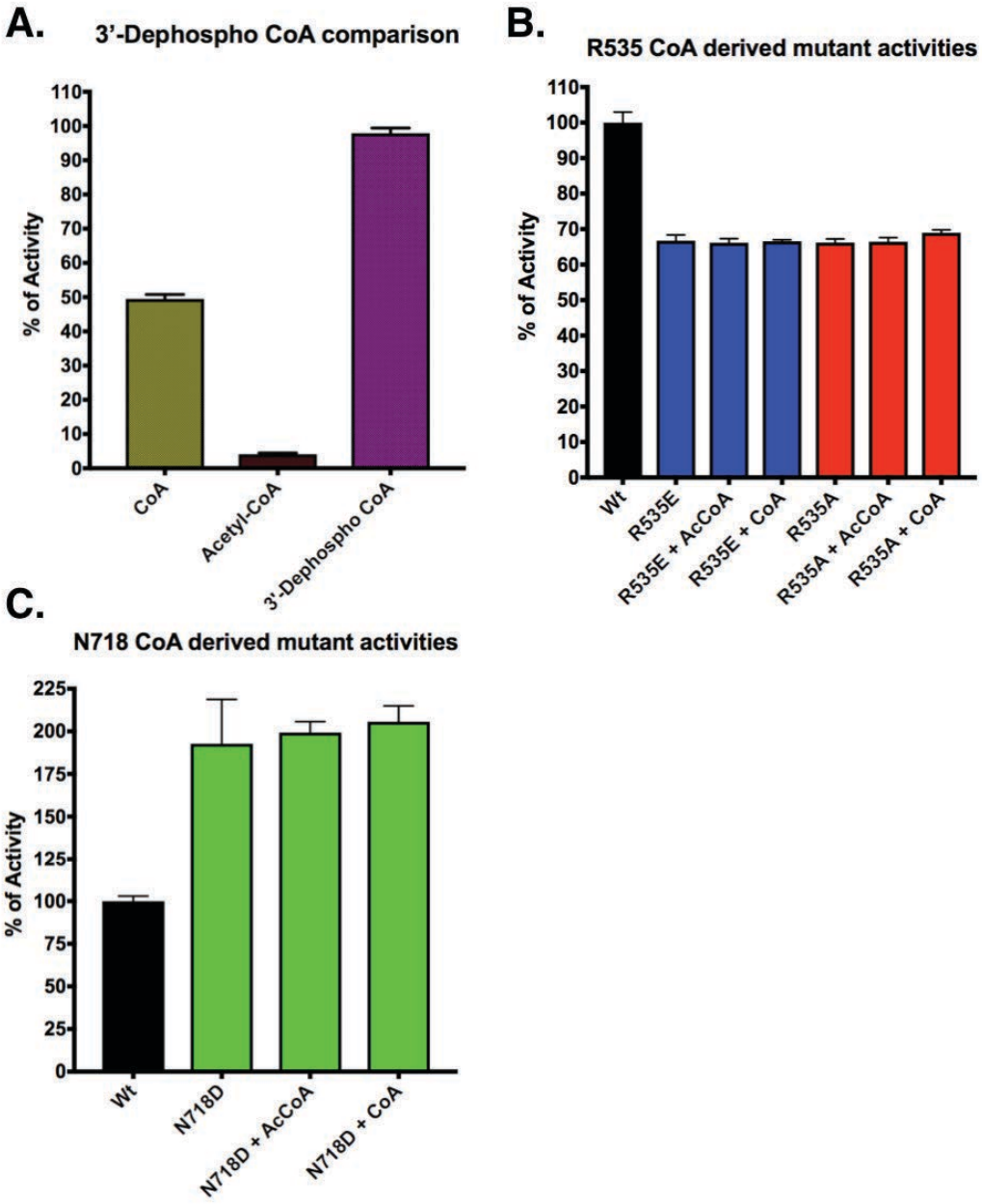


Figure 83. Identifying acetyl-CoA binding site
A. CoA and acetyl-CoA both inhibit the activity of Bd1833; however, a CoA analogue lacking the 3'-phosphate group shows no inhibition. **B.** Mutation of R535 that interacts with the 3'-phophate group alleviates the inhibitory affects of both acetyl-CoA and CoA. The rate of R535A and R535E is reduced compared to WT. **C.** N718D prevents acetyl-CoA and CoA at the thiol-group binding site. As a result CoA and acetyl-CoA no longer inhibit and the activity of Bd1833-R718D is increased compared to Wt. All data points were measured in at least triplicates. The error bars represent the standard deviation.

completely inhibited in the presence of acetyl-CoA yet has almost maximal activity in the presence of CoA. Importantly the K_i values obtained for acetyl-CoA from the two assays are relatively similar.

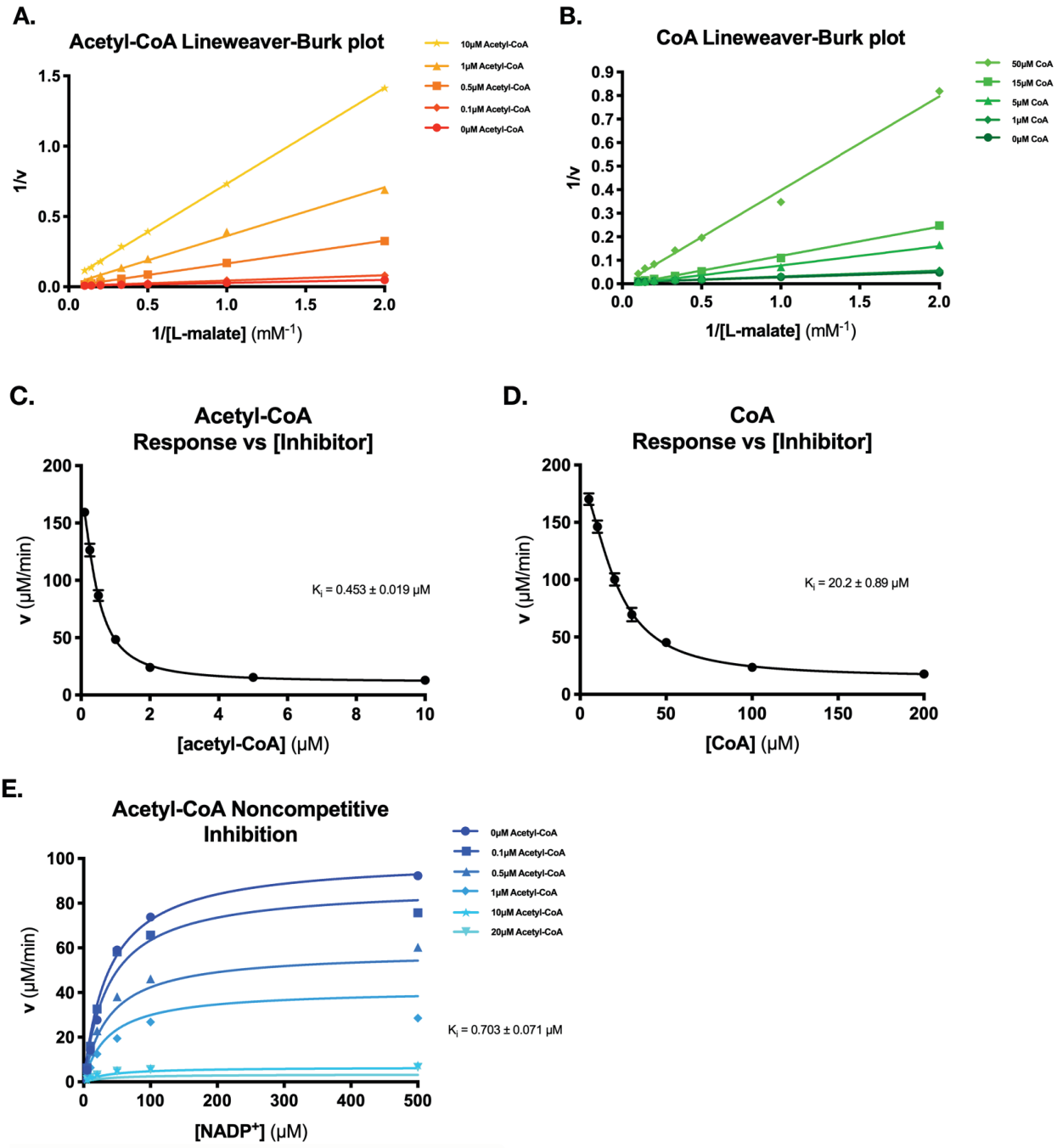


Figure 84. Inhibitor Potency
A. Lineweaver-Burk plot demonstrating the effect of acetyl-CoA on Bd1833 Michaelis-Menten kinetics **B.** Lineweaver-Burk plot demonstrating the effect of CoA on Bd1833 Michaelis-Menten kinetics **C.** Plot of response vs [inhibitor], demonstrating the effect of acetyl-CoA concentration on the activity of Bd1833 at a fixed substrate concentration (IC_{50}) **D.** Plot of response vs [inhibitor], demonstrating the effect of CoA concentration of the activity of Bd1833 at a fixed substrate concentration (IC_{50}) **E.** Plot of Michaelis-Menten data for Bd1833 ($NADP^+$ varied substrate) in the presence of increasing [acetyl-CoA]. The data were fit to the noncompetitive inhibition model. All data points were measured in at least triplicates. The error bars represent the standard deviation.

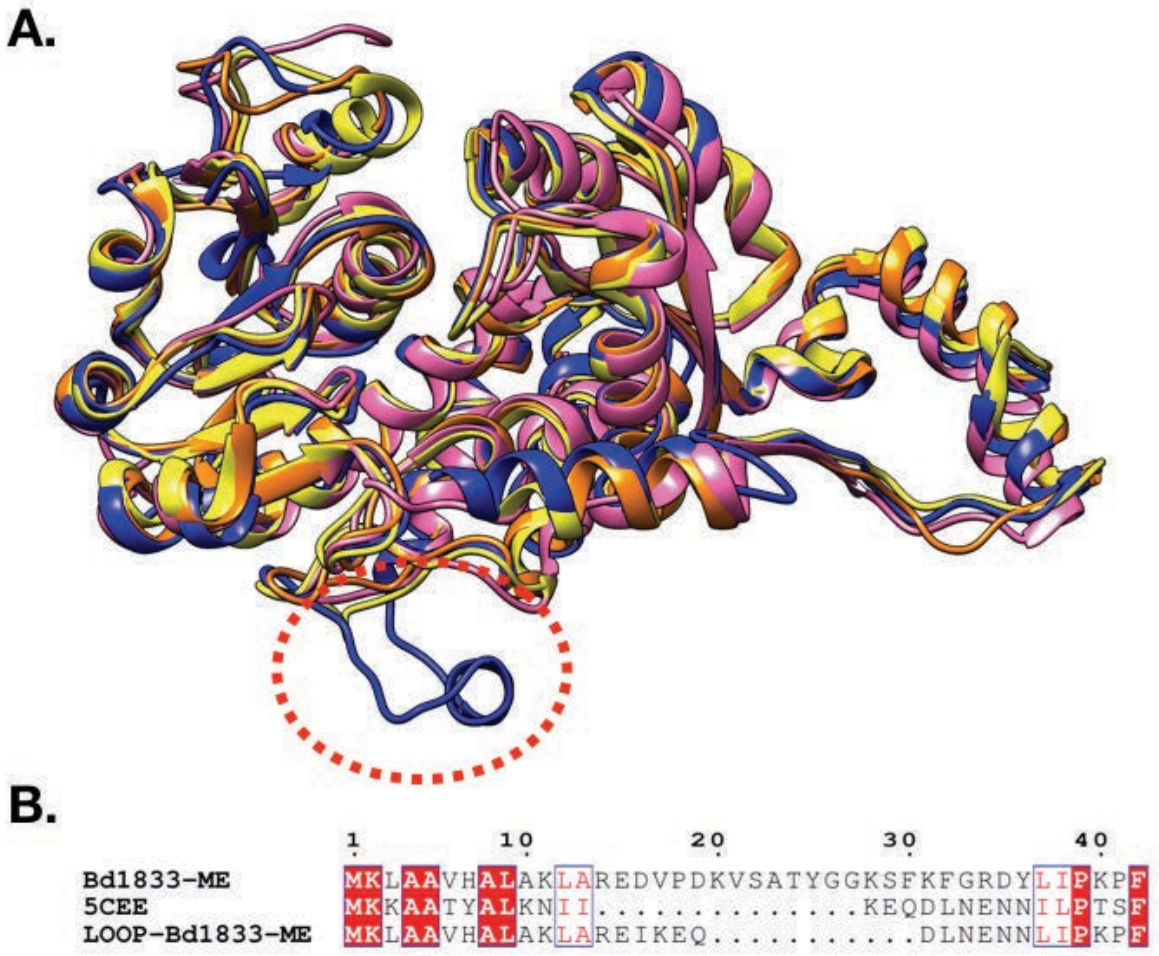


Figure 85. Multi-modular ME specific loop
A. The structure of Bd1833-ME (blue) superimposed with the prokaryotic MEs (pink, PDB:2A9F, length = 337 residues, rmsd = 1.78Å), (yellow, PDB:5CEE, length = 379 residues, rmsd = 1.09Å) and (orange, PDB:2HAE, length = 348 residues, rmsd = 1.58Å) using SSM superpose tool in coot. The loop region highlighted by the red dashed circle is specific to Bd1833 and other multi-modular malic enzymes. **B.** An alignment of the multi-modular specific loop region, indicating the sequence used to create the chimera of Bd1833-Δloop.

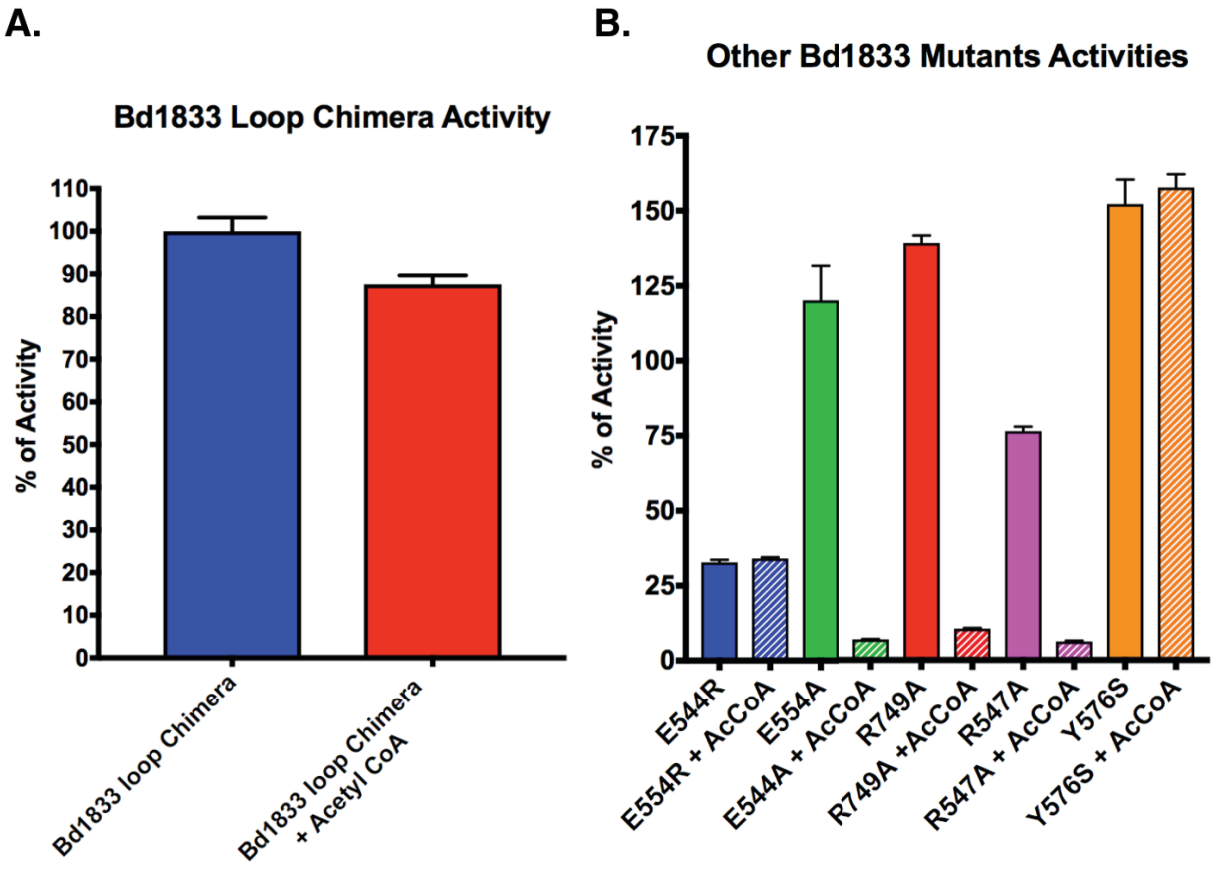


Figure 86. Bd1833 mutants activity and sensitivity to acetyl-CoA
A. The activity of Bd188-Δloop with no acetyl-CoA and in the presence of acetyl-CoA. **B.** This Figure presents the activities of a series of Bd1833 mutants and their sensitivity to acetyl-CoA. All activities are displayed as a percentage of the Bd1833 Wt. E544R shows a decreased activity but is insensitive to acetyl-CoA; however, E544A remains sensitive an has an unaffected activity. Y576S is insensitive to acetyl-CoA and has an increased activity. Furthermore, R749 and R547 mutants located at the hexamer interface remain sensitive to acetyl-CoA. All data points were measured in at least triplicates. The error bars represent the standard deviation.

5.2.8.8 Pyruvate reductive carboxylation

We investigated the reverse reaction of Bd1833 (pyruvate reductive carboxylation) to confirm the direction of the most favourable reaction. The parameters obtained from this experiment are displayed in Figure 87. The rates obtained for pyruvate reductive carboxylation were >120 times lower than those observed from the malate oxidative decarboxylation reaction suggesting that Bd1833 strongly favours the forward reaction where pyruvate and NADPH are the products.

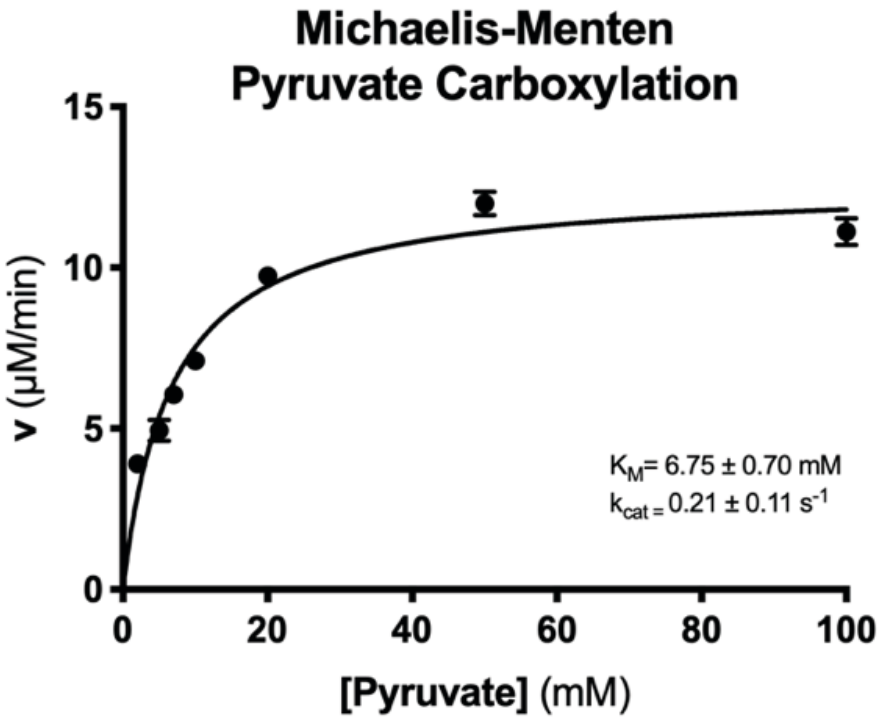


Figure 87. Pyruvate carboxylation reaction
Kinetic data for Bd1833 performing the reverse reaction (pyruvate carboxylation) fit to the Michaelis-Menten equation. Activity is reduced in the reverse direction, which suggests that Bd1833 favours the forward reaction and formation of pyruvate. All data points were measured in at least triplicates. The error bars represent the standard deviation.

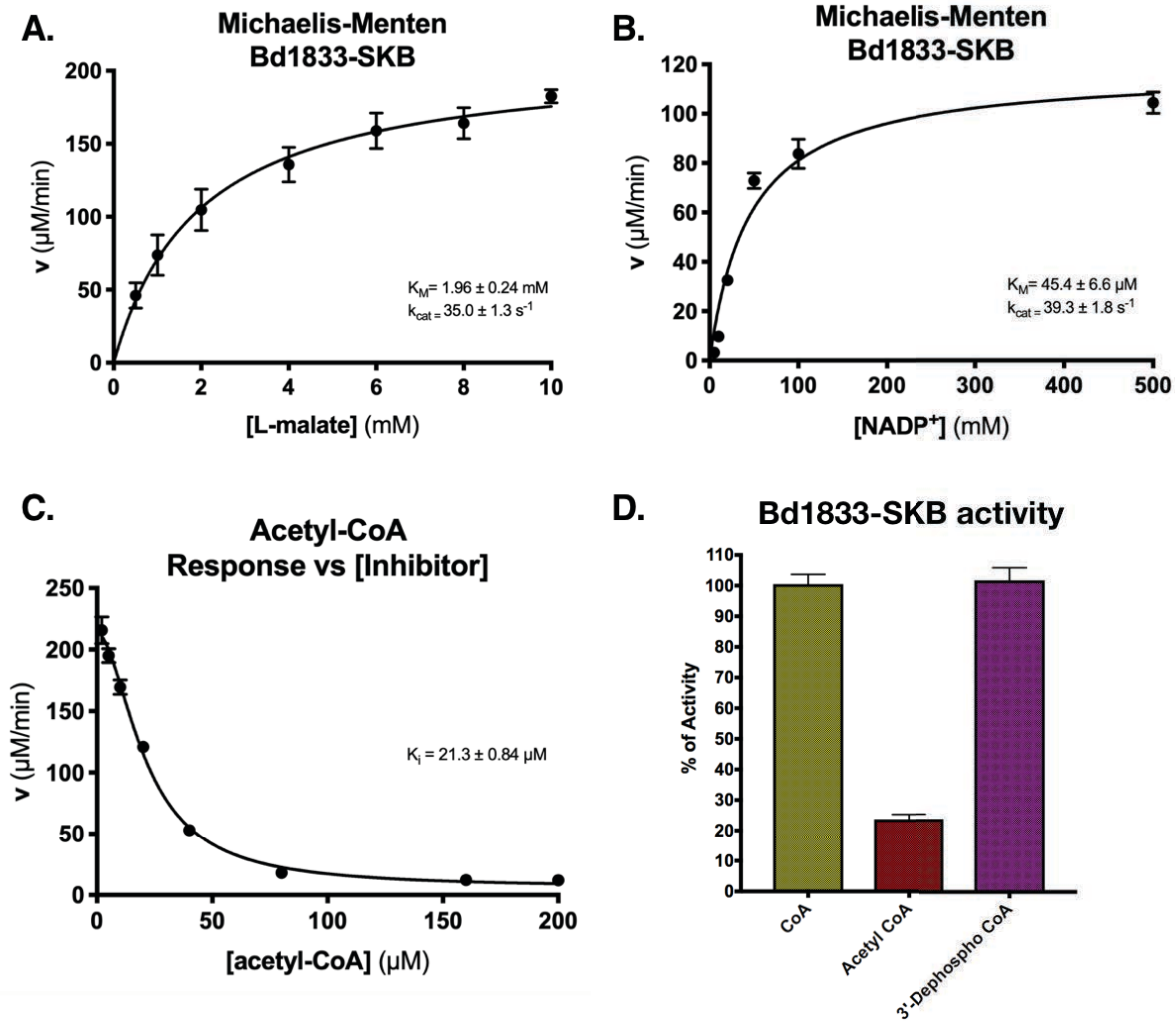


Figure 88. Bd1833-SKB kinetic parameters
A. Oxidative decarboxylation reaction with malate as substrate fit to Michaelis-Menten equation. **B.** Oxidative decarboxylation reaction with $NADP^+$ as substrate fit to Michaelis-Menten equation. **C.** Response vs [acetyl-CoA] data were measured for Bd1833-SKB activity. The K_i value obtained was increased compared to Bd1833 **D.** Bd1833-SKB is not inhibited by CoA at physiological concentrations, whereas acetyl-CoA is still an inhibitor. All data points were measured in at least triplicates. The error bars represent the standard deviation.

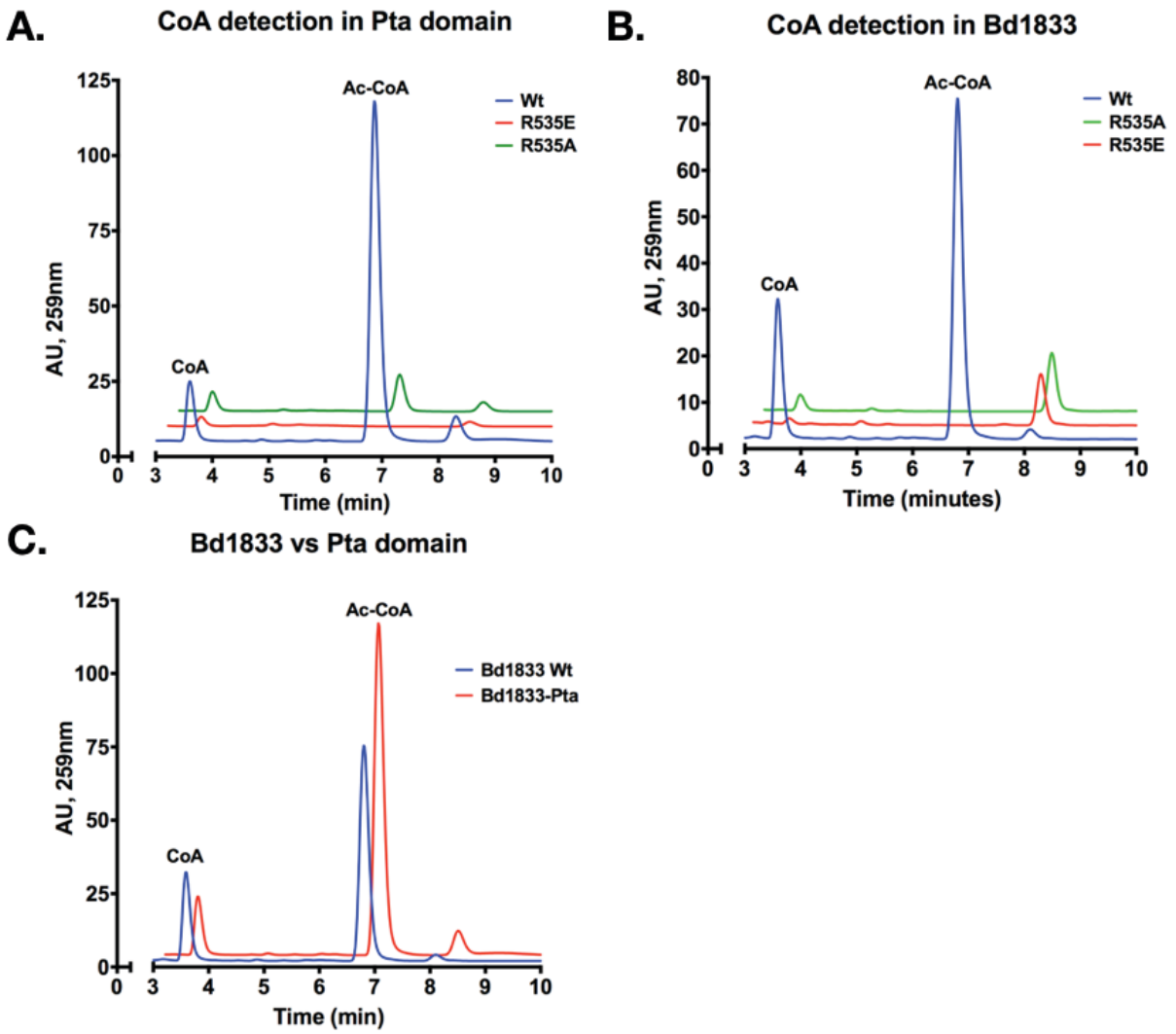


Figure 89. HPLC results
A. Bd1833-Pta CoA and acetyl-CoA HPLC traces in comparison to R535A and R535E mutants. **B.** Bd1833 full-length CoA and acetyl-CoA HPLC traces in comparison to R535A and R535E mutants. **C.** Comparison of Bd1833 and Bd1833-Pta HPLC traces demonstrate the ratio difference between co-purifying ligands. Bd1833-Pta co-purifies with more acetyl-CoA and less CoA than full-length Bd1833.

5.2.9.2 Producing Bd1833-Apo

A ligand free version of Bd1833 (Bd1833-Apo) was required to investigate allosteric regulation by acetyl-CoA without complications from co-purifying ligands. To produce Bd1833-Apo an additional mild denaturing wash step was added to the purification procedure of Bd1833. Illustrated in Figure 90 are the results from two different wash procedures performed on Bd1833 and Bd1833-Pta purification. The most effective

wash procedure was to pass >50CV of (50 mM HEPES pH 7.5, 300 mM NaCl, 2.5M Urea, 2M KBr) over immobilised Bd1833, which resulted in <5 µM of co-purified ligand in a 100 µM sample (>95% of protein is thus assumed to be ligand free).

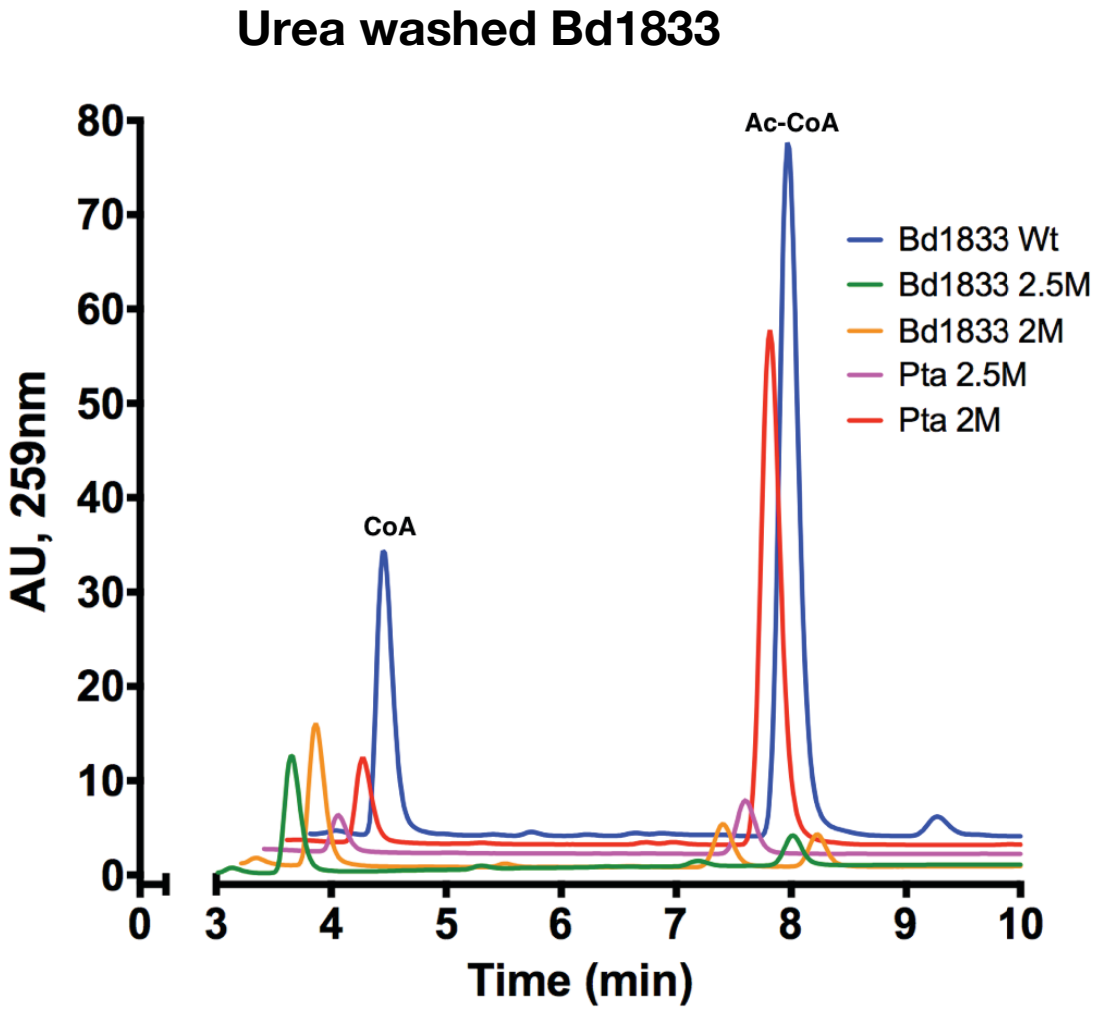


Figure 90. CoA and acetyl-CoA removal by denaturing wash procedure
Removal of CoA and acetyl-CoA from Bd1833 and Bd1833-Pta was achieved by incorporating an addition denaturing wash procedure to the purification protocol. Best results were achieved with 2.5M urea buffer (green and pink traces).

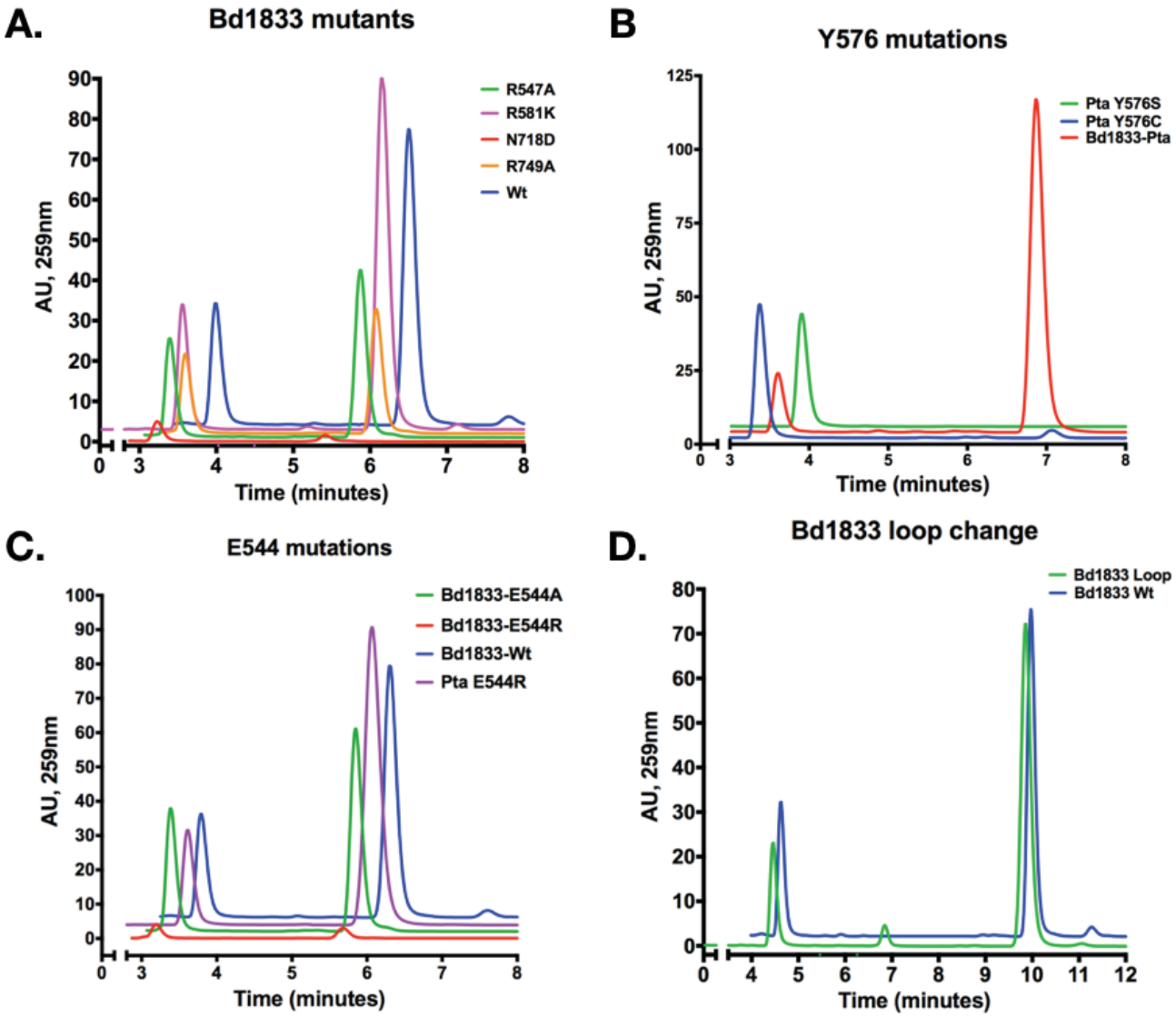


Figure 91. HPLC traces for Bd1833 and Bd1833-Pta mutants
A. Traces of Bd1833 mutants revealed all co-purify with CoA and acetyl-CoA except for N718D. **B.** Bd1833-Pta Y576S and Y576C selectively co-purify with CoA suggesting acetyl-CoA binding is unfavoured. **C.** Bd1833-E544R does not co-purify with ligand; however, Bd1833-E544A and Pta-E544R do co-purify with ligand. **D.** Bd1833-Δloop co-purifies with both ligands. This is the only case of a mutant that is insensitive to acetyl-CoA inhibition and still co-purifies with acetyl-CoA.

5.2.9.4 HPLC Bd1833-Δloop

Bd1833-Δloop co-purified with comparable amounts of CoA and acetyl-CoA as Bd1833 (Figure 91). This finding demonstrates the first of our Bd1833 mutations that co-purify with ligand but are insensitive to acetyl-CoA inhibition.

5.2.9.5 HPLC results for Bd1833-SKB

Bd1833-SKB did not co-purify with CoA or acetyl-CoA, which is strikingly different to Bd1833. Interestingly Pta-SKB (truncation of Bd1833-SKB) co-purified with levels of CoA and acetyl-CoA comparable to Bd1833-Pta.

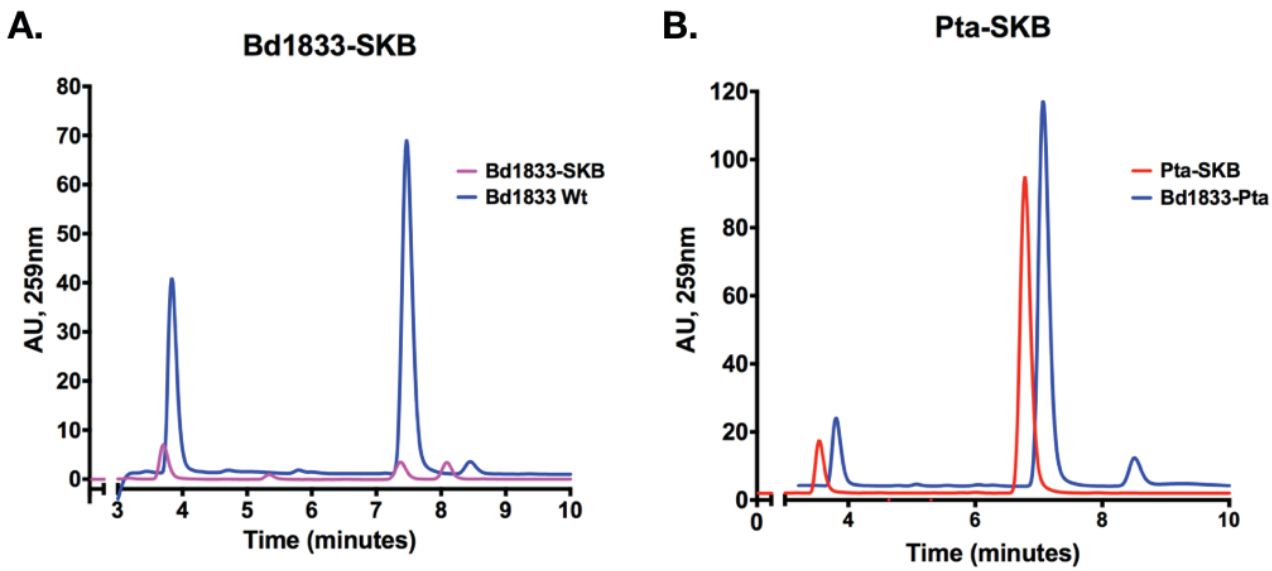


Figure 92. HPLC results for Bd1833-SKB and Bd1833-SKB-Pta
A. Bd1833-SKB did not co-purify with CoA or acetyl-CoA. **B.** Pta-SKB co-purified with CoA and acetyl-CoA.

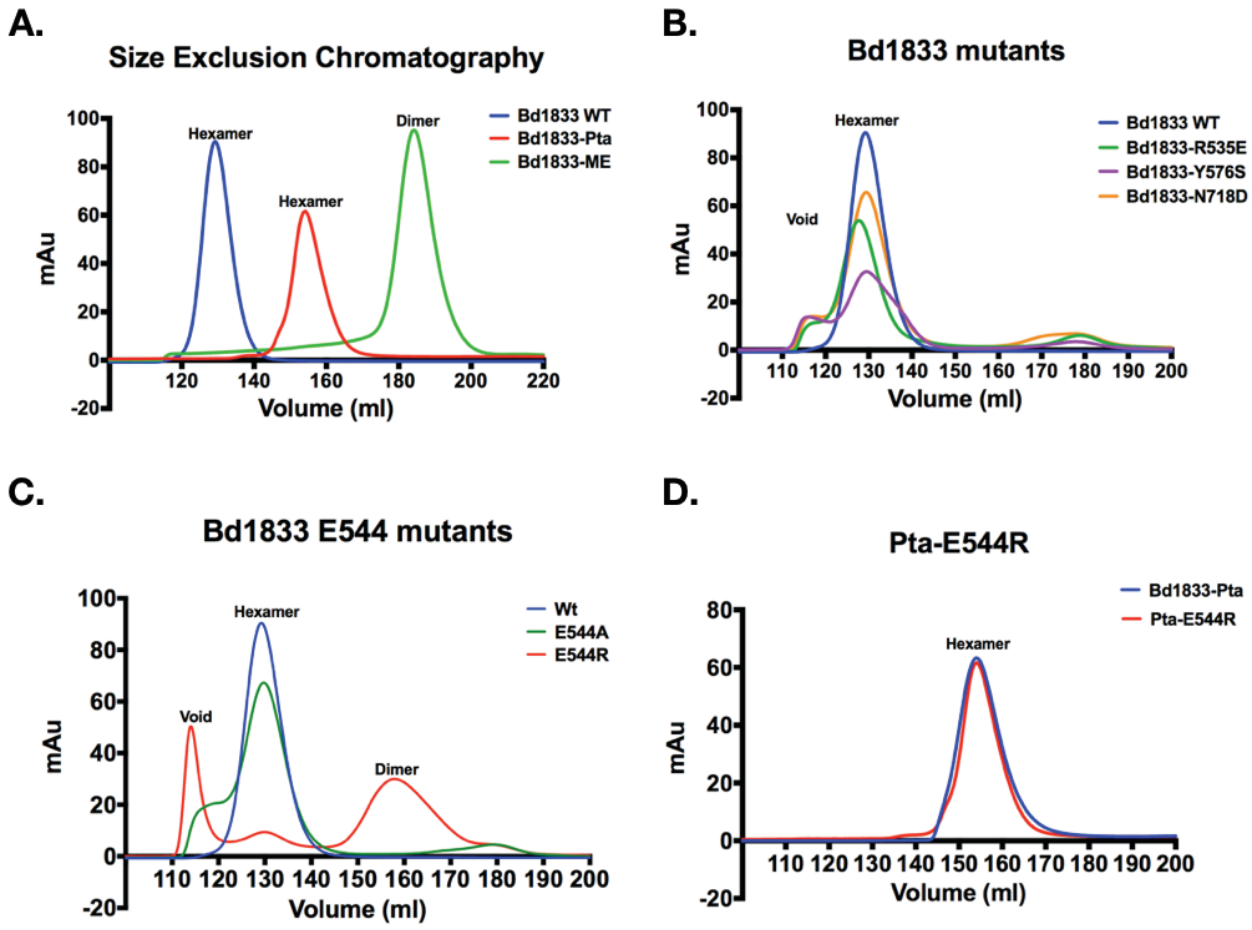
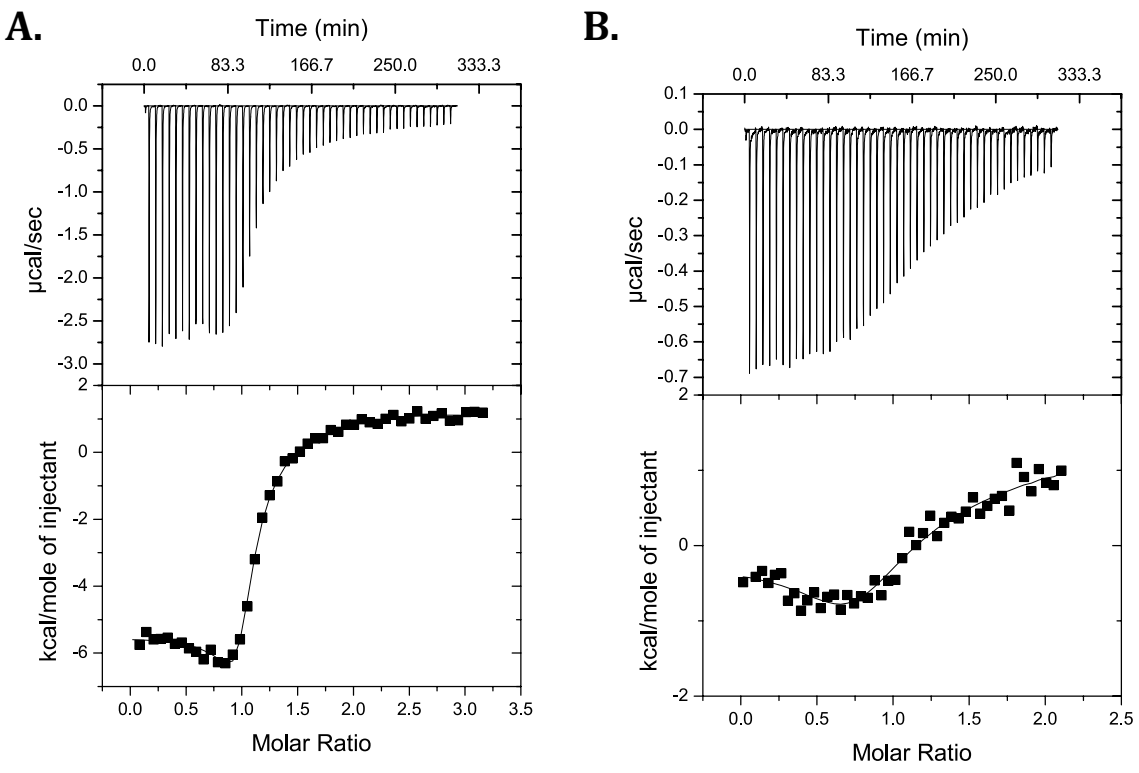


Figure 93. Size exclusion results
A. Bd1833 and Bd1833-Pta exist as a hexamer in solution, while Bd1833-ME exists as a dimer. **B.** Bd1833 mutants that we identified to affect hexamer interface also exist as hexamers in solution. **C.** Bd1833-E544R disrupts the hexamer interface to produce a mutant that forms a dimer in solution. Bd1833-E544A does not destabilise the hexamer interface sufficiently and still forms a hexamer. **D.** The same E544R mutation in the Pta domain alone (Pta-E544R) does not result in dimer formation.



C.

Site	Acetyl-CoA K (M)	Acetyl-CoA ΔH (J)	Acetyl-CoA ΔS (J mol⁻¹ K⁻¹)	CoA K (M)	CoA ΔH (J)	CoA ΔS (J mol⁻¹ K⁻¹)
1	5.7e5 ± 6.6e5	-5600 ± 88	7.6	1.8e5 ± 1.1e5	-440 ± 160	23
2	7.9e3 ± 1.6e3	-2.3E4 ± 8.2e3	-58	1.9e4 ± 1.1e4	-550 ± 1.1e4	18
3	2.9e3 ± 4.8e2	1.6e4 ± 5.2e5	71	1.2e5 ± 6.3e4	-7800 ± 7.4e4	-3.0
4	4.7e3 ± 8.4e3	3.9e4 ± 1.5e6	150	4.9e4 ± 2.0e4	1200 ± 3.8e5	25
5	1.5e5 ± 2.6e4	3.8e4 ± 1.3e6	150	1.6e5 ± 1.0e5	1.1e4 ± 4.9e5	61
6	5.4e5 ± 1.0e5	-6.9e4 ± 3.1e5	-210	1.0e5 ± 9.0e4	9500 ± 2.3e5	55

Figure 94. ITC results for Acetyl-CoA and CoA binding to Bd1833-Pta Apo
Upper panels represent the ITC titration curve and the lower panels represent the binding isotherm **A.** Acetyl-CoA binding. **B.** CoA binding. All data are fit to a sequential six-site model. **C.** A tabulated list of values obtained from the fit.

5.2.12 Bd1833 interaction with CdG

MST was used to determine whether Bd1833 interacted with CdG. A K_D value of 26 µM was obtained, suggesting weak non-specific binding. Furthermore CdG did not regulate the activity of Bd1833 in enzyme assays.

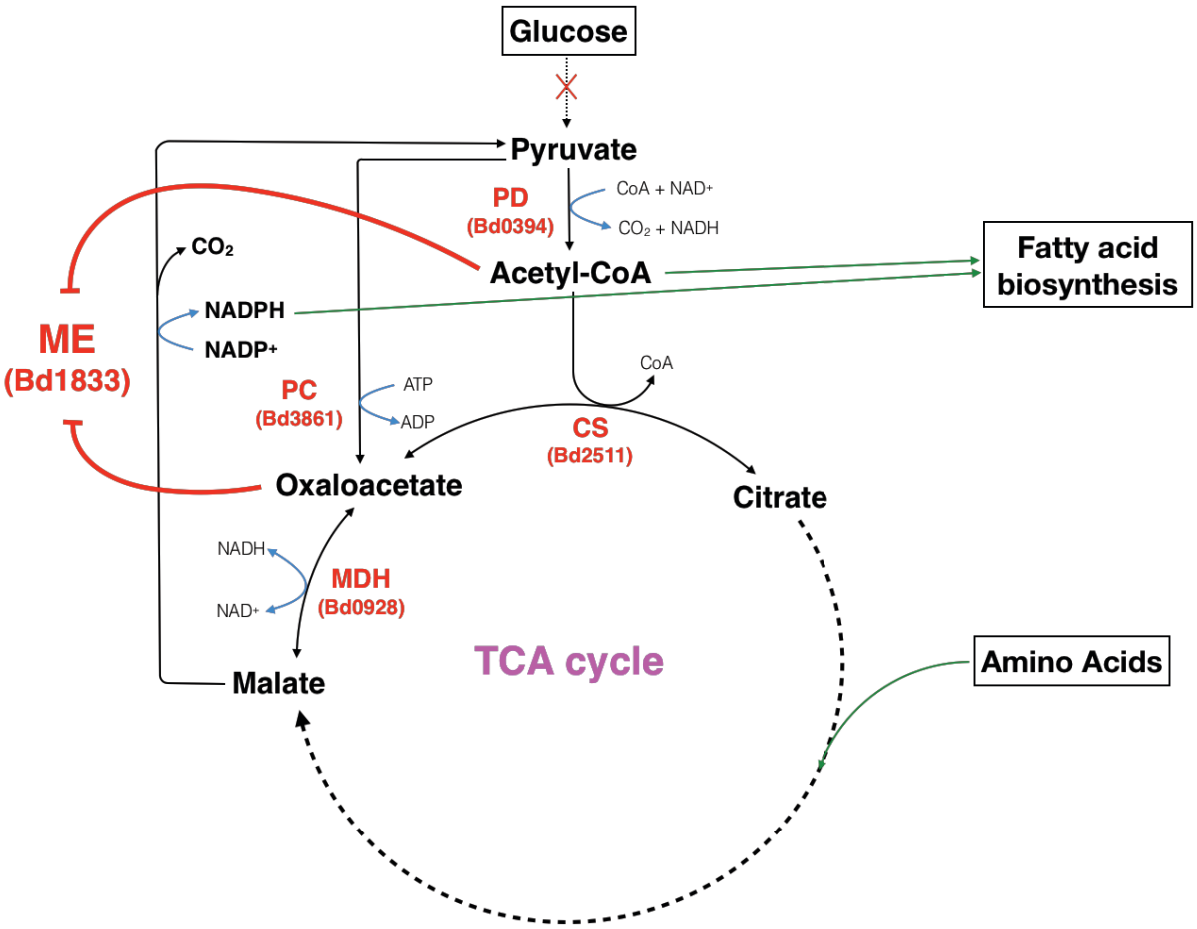


Figure 95. Schematic of Bd1833 role in metabolism
Bdellovibrio lacks sugar-metabolising pathways that are conventionally used to derive pyruvate. An increase flux of TCA cycle intermediates from prey-derived amino acids fuel *Bdellovibrio* metabolism. Bd1833 is central to subverting TCA intermediates to produce pyruvate and NADPH, which is ultimately used in anabolic reactions such as fatty acid synthesis. ME – malic enzyme, MDH – malate dehydrogenase PC – pyruvate carboxylase, PD, pyruvate dehydrogenase, CS – citrate synthase

which destabilises interactions at the hexamer interface such as E544:R749 salt bridge. The N-terminal helix of the Pta domain connects directly to the single linker helix that connects the Pta to the ME domain above. During the conformational change the N-terminal helix of the Pta rotates inwards by 8°, an action that likely impacts on the position of the multi-modular specific loop (subdomain-D). Subdomain-D could not be traced in the structure of full-length Bd1833; however, superimposition of Bd1833-ME (subdomain readily traceable) demonstrates that subdomain-D folds beneath the ME

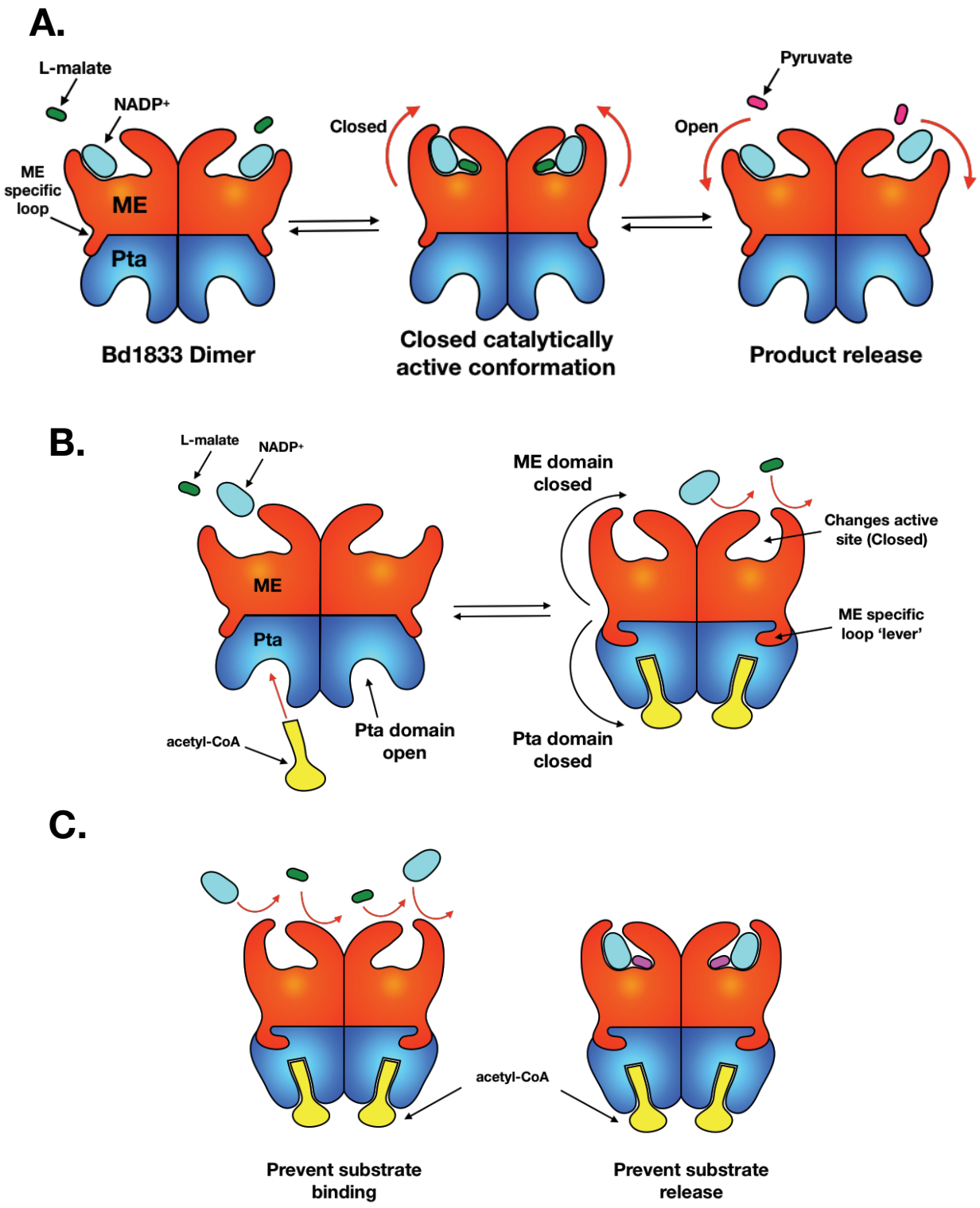


Figure 96. Cartoon of Bd1833 mechanism
A. Mechanism of Bd1833 ME activity. $NADP^+$ binds first then L-malate binds which initiates a conformational change and closes the active site. After the reaction is complete the ME domain opens and releases the products. **B.** Inhibition by acetyl-CoA. Acetyl-CoA binding initiates a conformational change in the Pta domain, which is conveyed to a conformational change in the ME domain via movement of the multi-modular specific loop. **C.** Acetyl-CoA shows non-competitive inhibition and is equally likely to bind the enzyme:substrate complex and free enzyme.

Furthermore, R581 is also altered in MaeB to a serine. Interestingly Bd1833-SKB also differs at R581 (lysine residue). Mutation of R581K could explain why Bd1833-SKB is not inhibited by CoA and inhibited less by acetyl-CoA compared to Bd1833 ($K_{i(acetyl\text{-}CoA)}$ 0.453 ± 0.019 µM for Bd1833 and $K_{i(acetyl\text{-}CoA)}$ 21.3 ± 0.84 µM for Bd1833-SKB). The acetyl-group binding pocket also differs between Bd1833 and other multi-modular ME. Q683 (H-bond acetyl-group) is not conserved and is exchanged to a His in MaeB and DME, which may explain why acetyl-CoA is selectively inhibitory (weaker 3'-P and pyrophosphate binding sites and stronger acetyl-group interactions). Together these observations demonstrate that Bd1833 likely exhibits the highest affinity CoA binding site explaining why Bd1833 activity displays inhibition to both CoA and acetyl-CoA and the additional interaction with the acetyl-Group explains why acetyl-CoA is more potent.

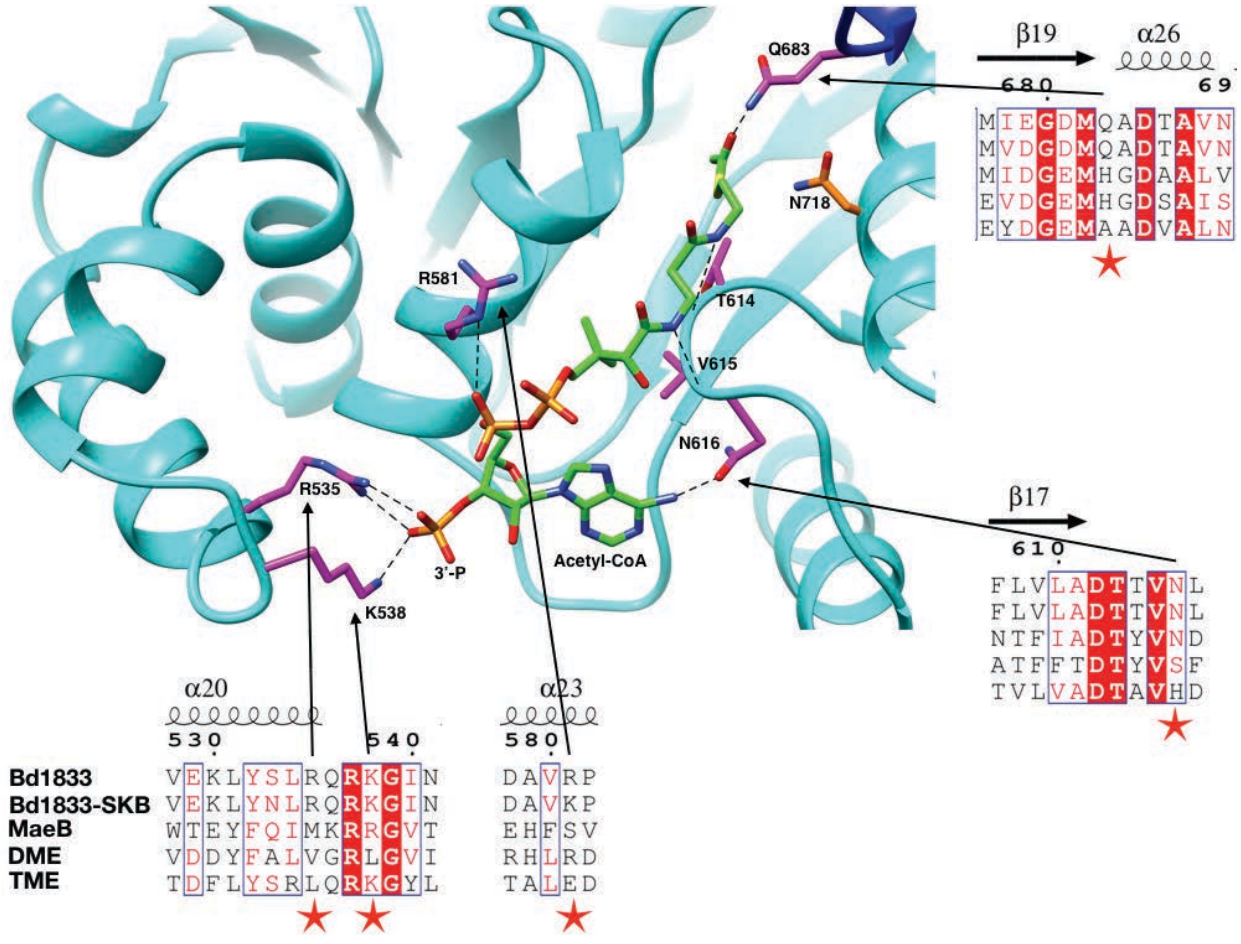


Figure 97. Bd1833 acetyl-CoA binding site comparison
The binding site of acetyl-CoA with the corresponding sections of sequence alignment of multi-modular ME, demonstrating the mutations at these locations, that likely contribute to the different binding affinities for CoA / acetyl-CoA.

The proposed CoA binding site for Pta enzymes disagrees with biochemical studies, whereas in Bd1833-Pta the observed acetyl-CoA binding site agrees with previous biochemical studies [247, 248]. The conformational change observed in Bd1833 upon acetyl-CoA binding is similar to that observed in other Pta enzymes, which has been suggested to be important for catalysis [250]. Comparison of the conformational changes observed in subdomain-I with those of Pta enzymes reveals the transition between Bd1833-Pta and Pta-Apo represents another closed state (Figure 98). Importantly the structure of Bd1833-Pta demonstrates a closed form bound to acetyl-CoA, thus in a catalytically competent state (were it to be an enzyme).

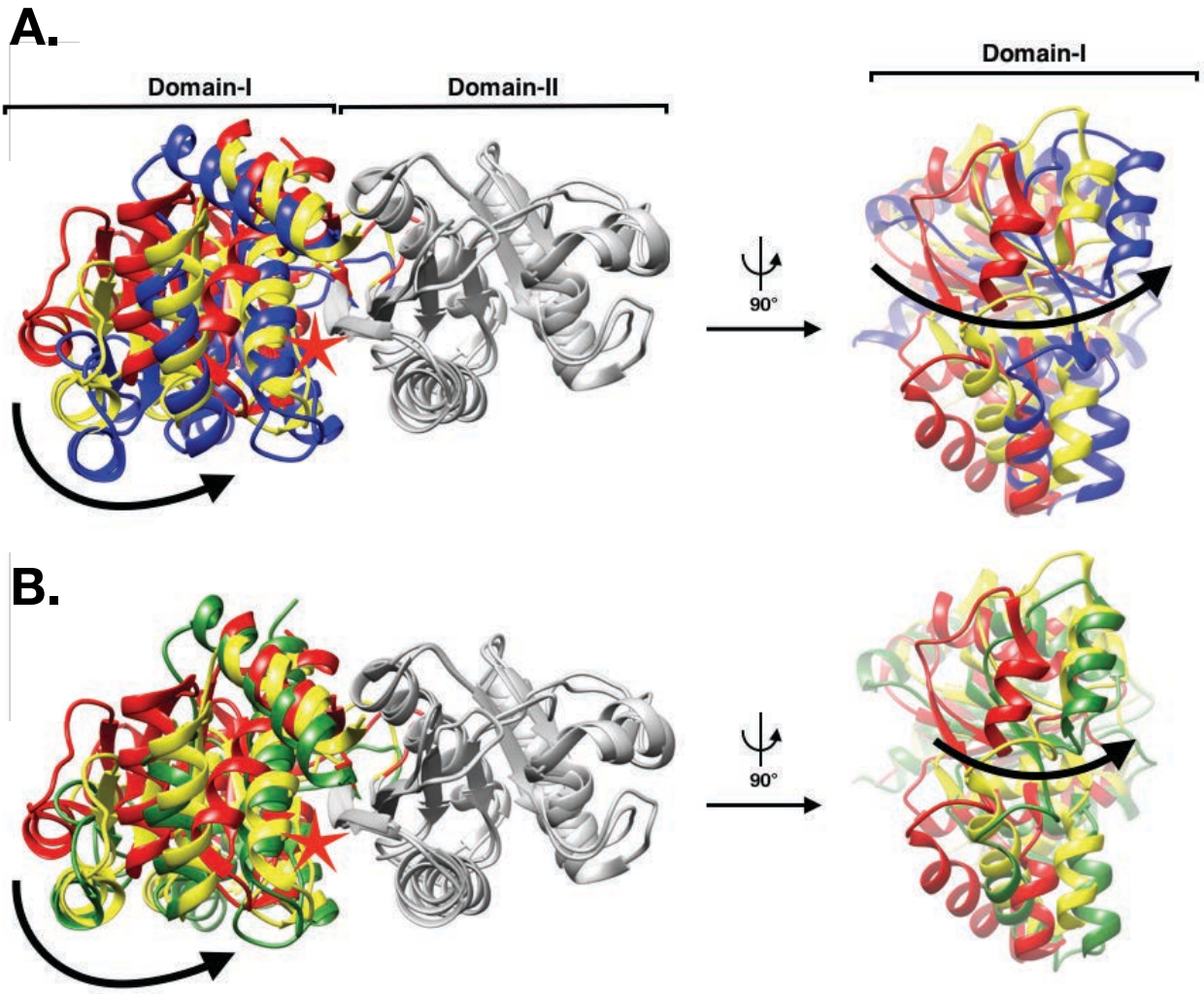


Figure 98. Pta domain movements
A. The conformational change in Bd1833-Pta subdomain-I (blue) superimposed with PDB:1QZT in the open conformation (red) and closed conformation (yellow). Bd1833-Pta adopts a different closed state. **B.** The conformation of Pta-Apo subdomain-I (green) superimposed with PDB:1QZT in open (red) and closed (yellow) conformations. Pta-Apo adopts a similar conformation as PDB:1QZT in the closed state. Superimposition was performed by aligning subdomain-II (residues 596-737) using the LSQ superpose tool in coot. (Length = 118 residues, rmsd 4.97Å).

from the thiol group of acetyl-CoA demonstrating its potential ability to fulfil this role [250].

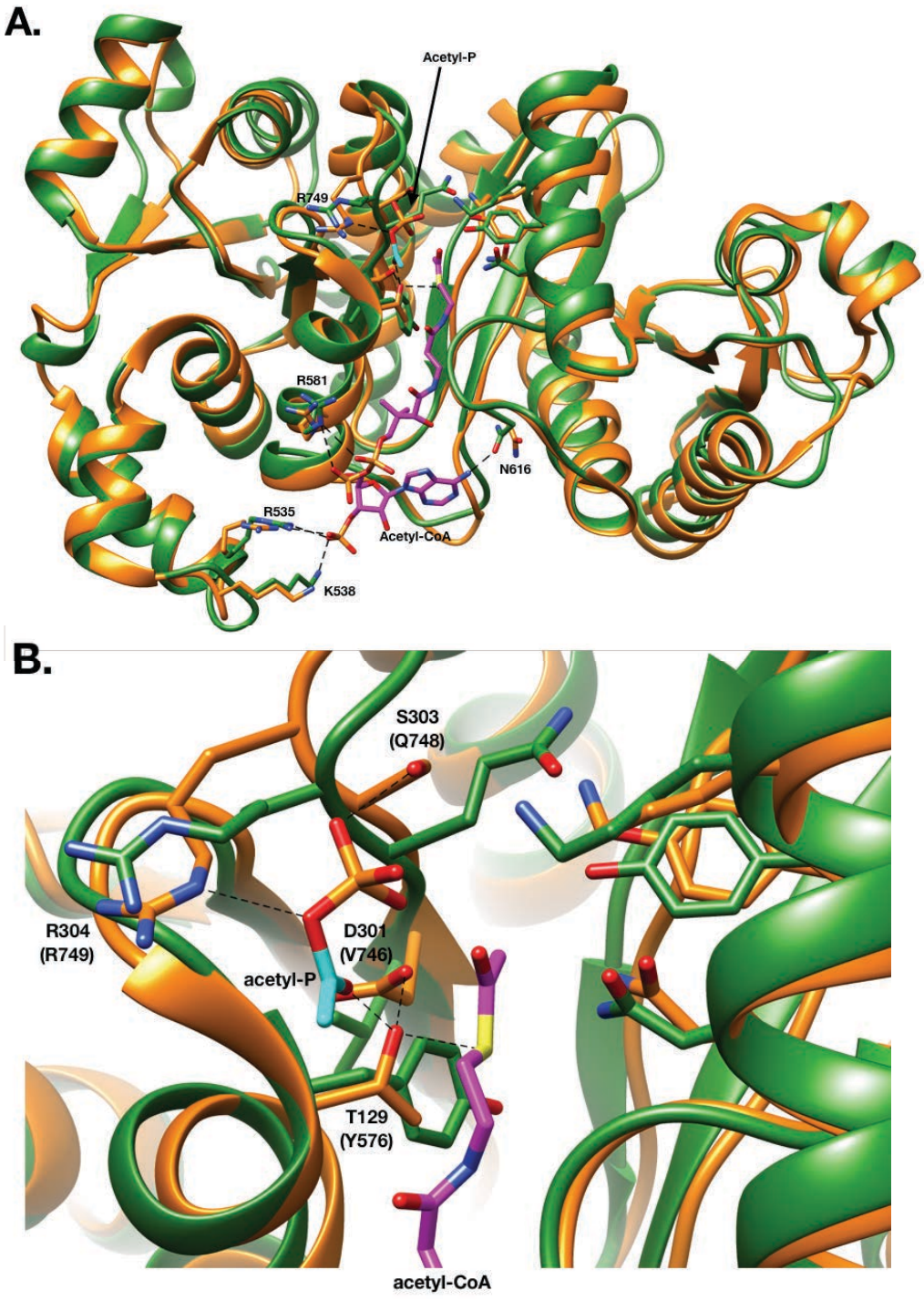


Figure 99. Pta active site
A. Bd1833-Pta monomer (green) superimposed with PDB:1XCO Pta monomer (orange). Superimposition was performed by individually aligning Bd1833-Pta subdomain-1 and II with the individual domains of PDB:1XCO. Bd1833-Pta subdomain-I (residues 460-587) with PDB:1XCO domain-I (residues 14-297) (length =174 residues, rmsd = 1.43Å). Bd1833-Pta subdomain-II (residues 596-737Å) with PDB:1XCO domain-II (residues 150-292) (length = 144 residues, rmsd = 1.57). The active site is likely located in the upper portion of the substrate-binding groove **B.** A close up view of proposed active site. The catalytic residues of 1XCO are labelled alongside the comparative Bd1833-Pta residues, demonstrating substitutions. The thiol group is within H-bonding distance of T129. T129 is in H-bond distance of D301 and acetyl-P.

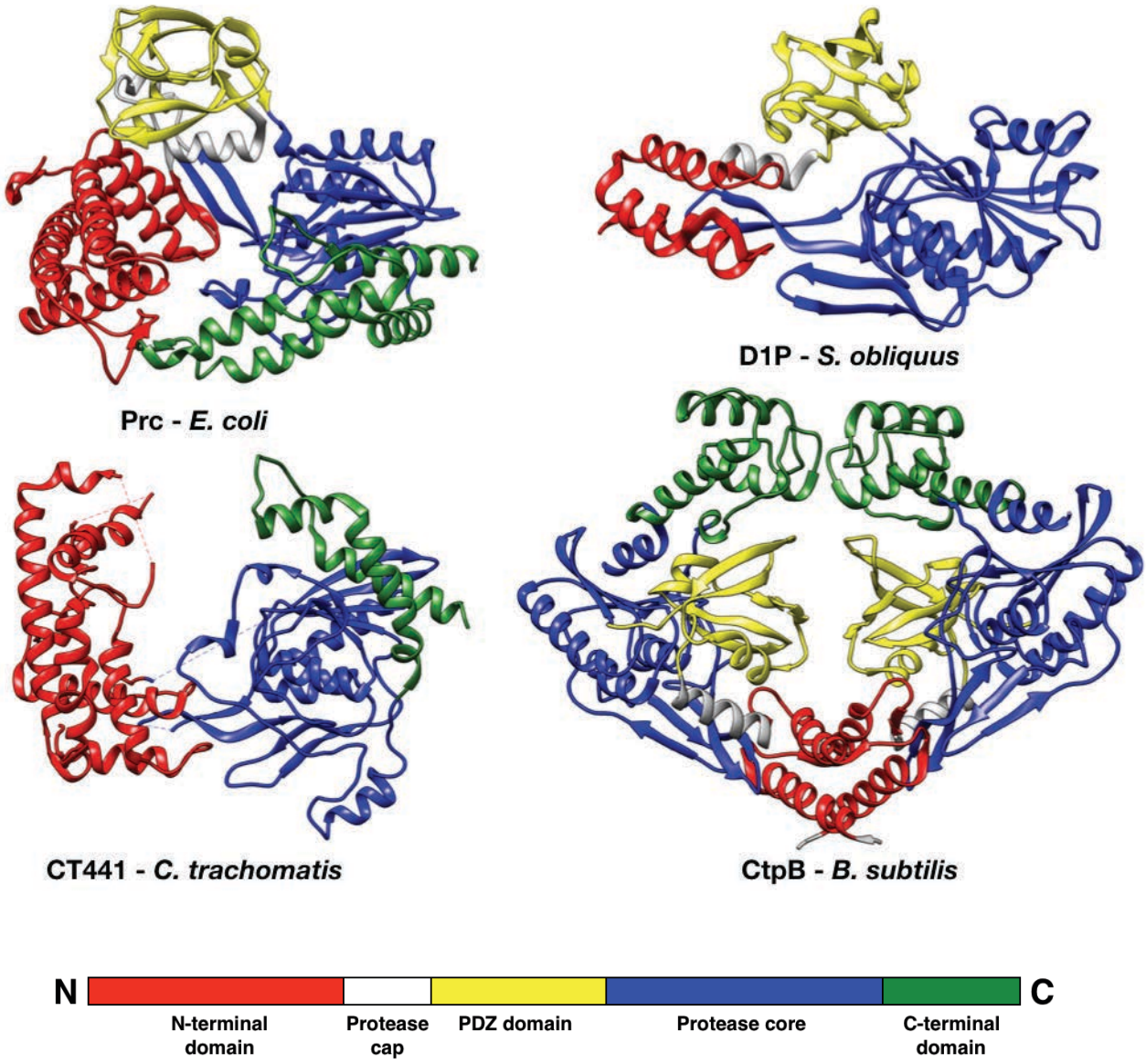


Figure 100. Structures of C-terminal processing proteases
The four solved structures of C-terminal processing proteases are presented and coloured by their domains. NTD in red, PDZ domain in yellow, protease domain in blue, protease cap helix in grey and CTD in green. Prc – 5WQL, D1P – 1FC7, CT441 – 4QL6, CtpB – 4C2C.

6.1.3.1 D1P

The first three-dimensional structure solved for a CTP was the structure of D1P, which consists of three loosely packed, distinct domains: an α-helical N-terminal domain, a central PDZ domain and a larger C-terminal protease domain. The PDZ domain is composed predominantly of a twisted five-stranded β-sheet with two α-helices. The protease domain is folded into two β-sheet regions. The larger, mixed five-stranded β-

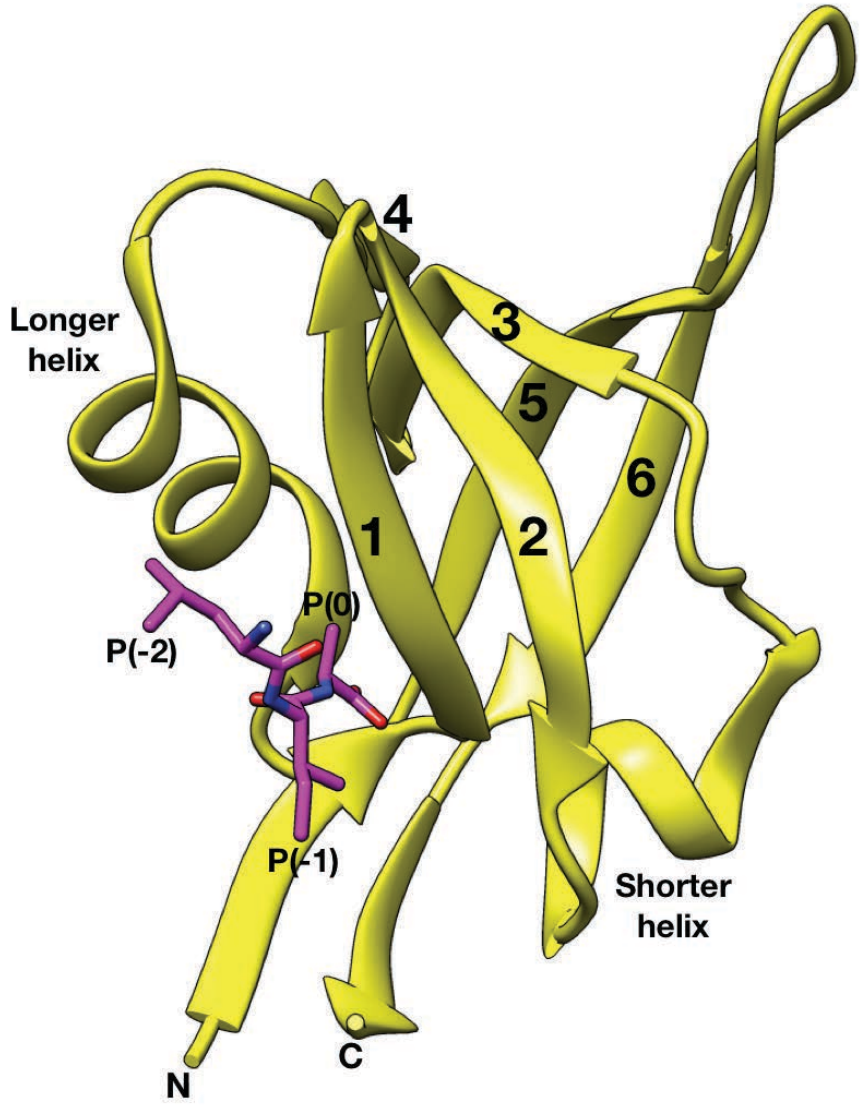


Figure 101. PDZ domain structure
The structure of a C-terminal processing protease PDZ domain is shown here bound to a 3aa peptide substrate. The peptide binds in a β-augmentation manner with strand-1. The peptide residues are numbered according to the text. The N and C-terminal β-strands are specific to CTPs. (PDB: 6BXG)

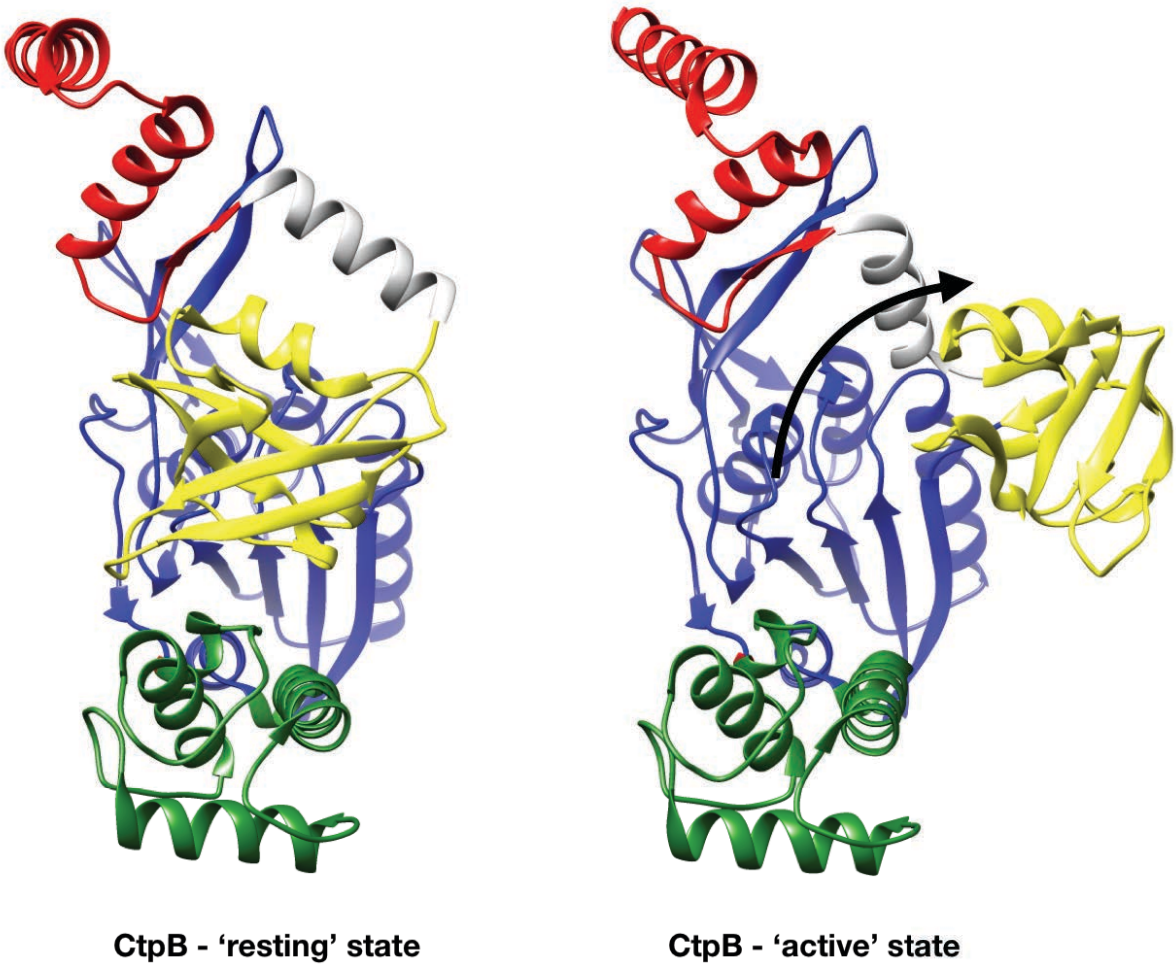


Figure 102. PDZ domain movement in CtpB
The mechanism of activation and regulation of C-terminal processing protease by the PDZ domain has been well documented in CtpB. The PDZ domain coloured (yellow), has been observed in multiple conformations and samples between a 'resting' and an 'active ' state. In the 'resting' state the PDZ domain is tightly tucked into the protease domain (blue). Whereas in the 'active' state (peptide bound) the PDZ domain pulls away from the protease domain. The black arrow represents the movement. The switch also repositions the protease cap (grey). PDB:4C2C & 4C2E

R168 was shown to be a key residue facilitating the sensing of peptide binding at the PDZ binding cleft and transferring the signal allosterically to bring about the conformational switch of CtpB. In the 'resting' state the sidechain of R168 faces away from the PDZ domain, making contacts with S106 located on protease cap. In the 'active' state the sidechain of R168 repositions, aligning with the bound peptide but still making contacts with S106. By maintaining these inter-domain contacts in the bound

the molecular ruler, where the distance between PDZ domain and proteolytic chamber is a measured uniform distance. The second is a processive cleavage mechanism, where the cleavage of the peptide chain at the proteolytic chamber creates a new substrate C-terminus that is subsequently recognised by the PDZ domain. The peptide-binding sites at the PDZ domain and protease domain would thus have to be highly similar. This mechanism would operate in a ratchet like manner and allow continued proteolytic processing of the substrate. The other possible mechanism involves PDZ imposed oligomer reorganisation by a conformational state change (not a proposed mechanism for CTPs) [271]. Figure 103 illustrates a cartoon representation for PDZ recognition and substrate presentation in CTPs.

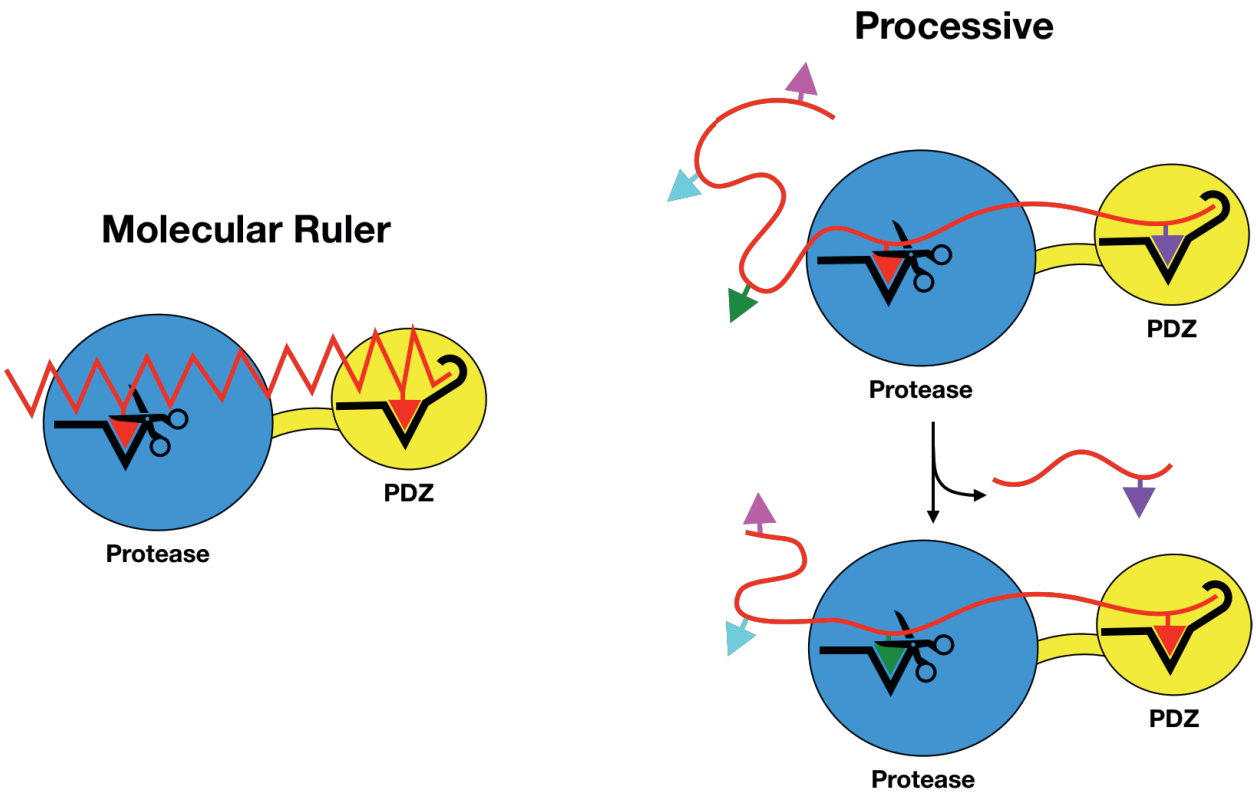


Figure 103. C-terminal processing protease mechanism
Two mechanisms are proposed for C-terminal processing proteases. **A.** A molecular ruler mechanism, where the distance between PDZ domain and proteolytic site is set. **B.** A processive mechanism, where the protease and PDZ binding sites are similar. The C-terminus of the new peptide becomes a ligand for the PDZ domain. Red line depicts the substrate and the coloured triangles demonstrates the recognition sequences (different colours demonstrate a different amino acid sequence.

6.1.6 Bd0967 role in predatory cycle

Bd0967 has been linked to a larger network of cellular signalling events in *Bdellovibrio,* possibly controlled by CdG [82]. Moreover, Bd0967 is a putative CTP, containing a PDZ domain for possible co-ordination of protein-protein interactions. The precise role of Bd0967 during the lifecycle of *Bdellovibrio* is currently unknown; however, an important role during cellular signalling and lifecycle progression is highly likely, due to our current knowledge of cellular and biological roles for CTPs in other organisms.

In addition to Bd0967, numerous other PDZ-containing proteins are found in the genome of *Bdellovibrio HD100*, which are listed in Table 15. Seven of the nine proteins identified as PDZ-containing proteins are proteases, one of which (Bd3534) has been shown to be an essential predation protein [293].

Uniprot ID	Gene name	Protein type	Length
Q6MGY2	Bd3787	Hypothetical zinc metalloprotease	557
Q6MHK7	Bd3534	Carboxyl-terminal protease	526
Q6MJH7	Bd2800	Serine protease MucD	474
Q6MP94	Bd0967	Tail-specific protease	673
Q6MRC0	Bd0169	Carboxyl-terminal processing protease	458
Q6MNK9	Bd1239	Carboxy-terminal processing protease	401
Q6MMM7	Bd1598	General secretion pathway protein C	304
Q6MPD5	Bd0922	Periplasmic serine protease	476
Q6MLF8	Bd2053	Component of the Tol biopolymer transport system	974

Table 15. PDZ proteins in *Bdellovibrio HD100* genome
A list of PDZ containing proteins found in the genome of *Bdellovibrio HD100.* The majority of the proteins are proteases. Bd3534 (green) has been identified as an essential predation protein.

6.2 Results

6.2.1 Obtaining Bd0967 crystals

A construct of Bd0967 was created to encompass the full-length protein with the exclusion of the seven C-terminal amino acids that were predicted to be disordered. The protein was tagged with a His_6-tag at the N and C terminus. Closer inspection of the amino acid sequence revealed that Bd0967 possessed a probable signal peptide with a signal peptide cleavage site (LASA) located 24aa from the N-terminus. Thus Bd0967 is likely to be trafficked to the periplasm where it is functionally active. An unambiguous N-terminal sequence of QLKEGLE… was determined by ESI-MS, confirming purified Bd0967 had been processed to remove the 24aa signal peptide (ESI-MS was performed by the Stříšovský lab).

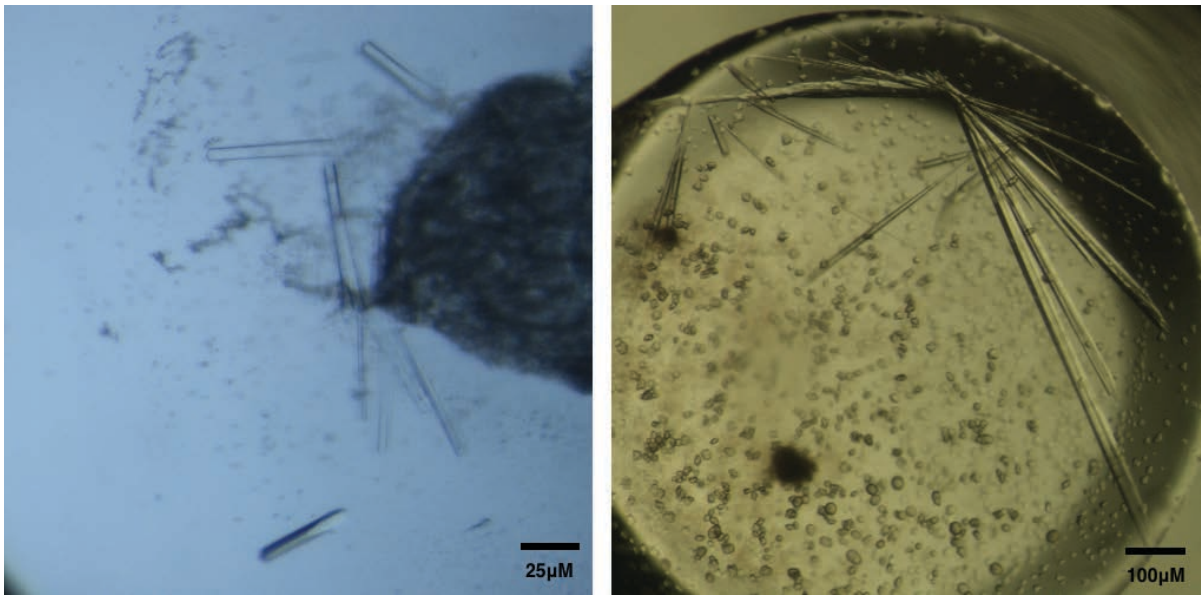


Figure 104. Bd0967 crystallisation
Images of Bd0967 crystals. Higher quality crystals were obtained by using random matrix microseeding as described in Section 2.6.9.

stranded, anti-parallel, β-sheet) is attached between the protease domain and the protease cap via a flexible hinge.

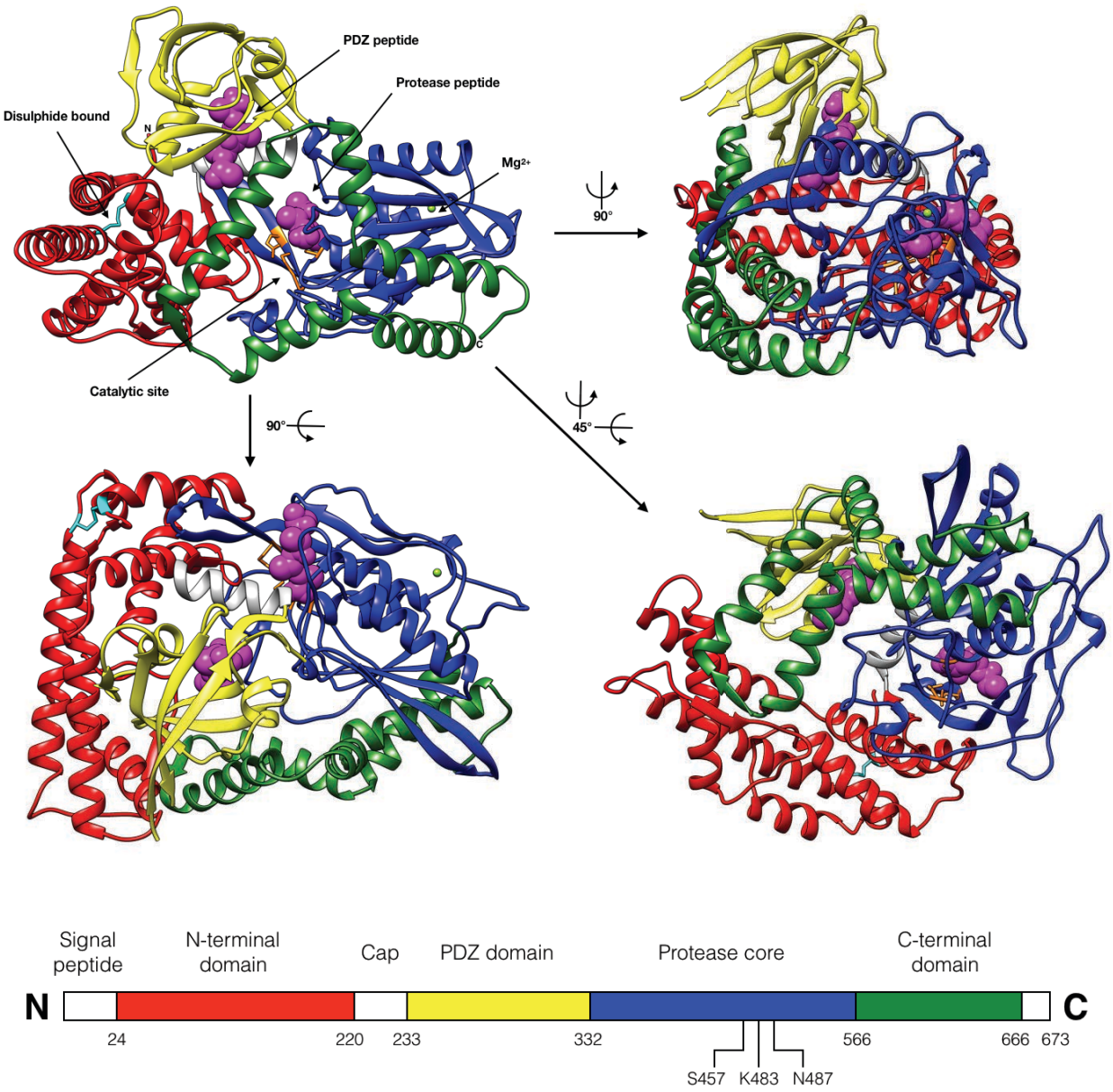


Figure 105. Domain architecture of Bd0967
Bd0967 can be divided into four distinct domains and a subdomain, which are represented by the different colours in this image. The scaffold-like NTD (red), the protease cap helix (grey), the PDZ domain responsible for peptide recognition (yellow), the protease domain (blue) and finally the paddle-like CTD (green). Two peptide fragments (magenta) are bound to the PDZ domain and protease domain. The active-site residues are highlighted in orange, the NTD disulphide bond is shown in cyan and the metal ion bound to the protease domain is represented as a green sphere.

Larger B-factor values are associated with the CTD domain, which suggests that it is more flexible and likely alters conformation to facilitate the movement associated with PDZ domains in this family of enzymes. Higher B-factors are also associated with the PDZ domain, which is known to adopt multiple conformations in other CTPs. A map of B-factors for Bd0967 is shown in Figure 106, which illustrates that the base of the protein (protease and lower regions of NTD and CTD) has lower B-factor values, whereas the upper regions of the protein have higher B-factor value suggesting these are more flexible.

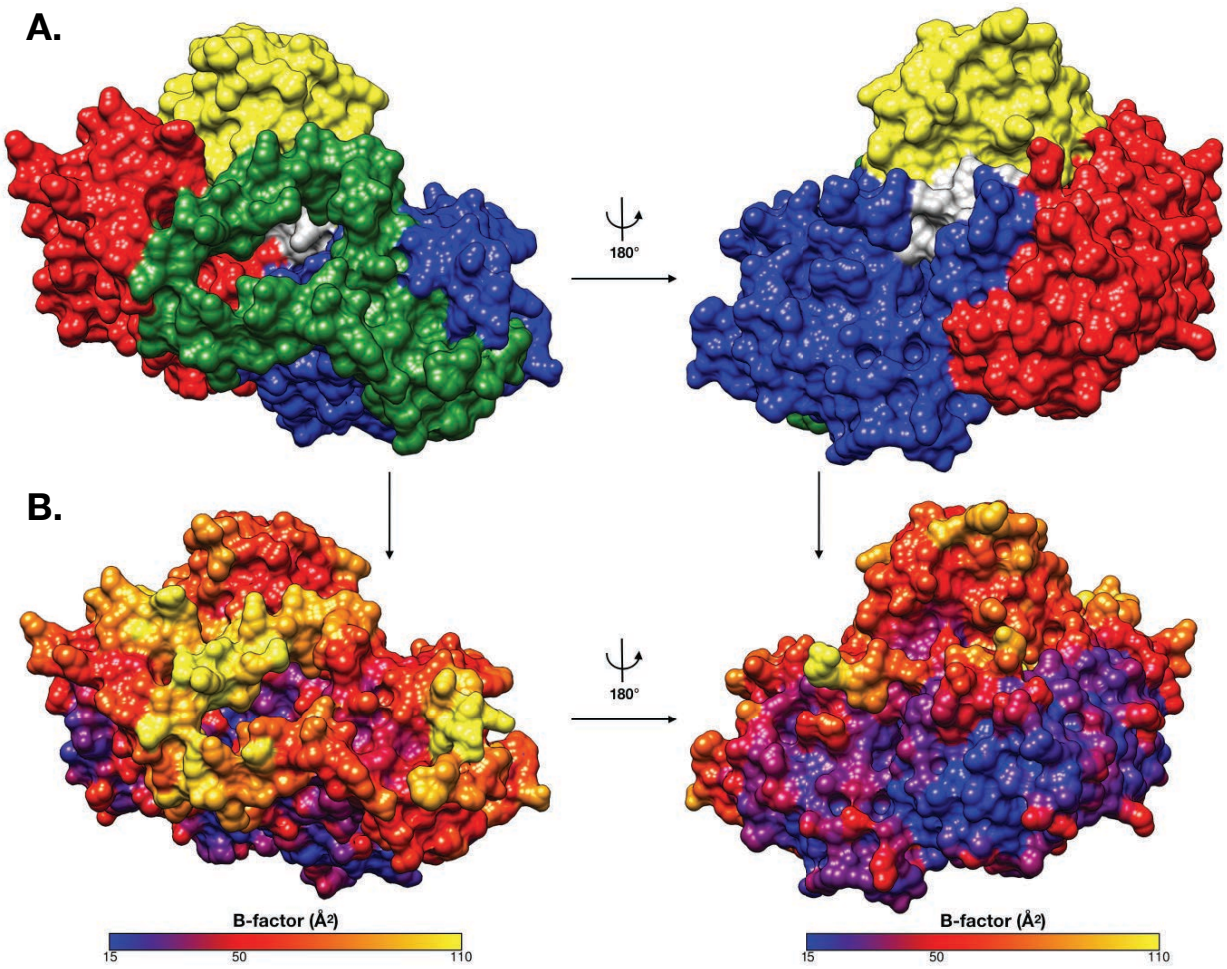


Figure 106. Bd0967 B-factors
A. Represents Bd0967 orientation (front and back), coloured according to domains. **B.** Bd0967 in the same orientation as panel A, coloured according to B-factor. Bd0967 structure has regions of contrasting B-factor values suggesting certain regions such as the CTD and PDZ are more dynamic, whereas the protease domain is more static. Average B-factor values per residue have been coloured using a gradient from low B-factor (blue) to high B-factor (yellow).

6.2.3 Bd0967 cavity

The structure of Bd0967 creates a large self-compartmentalised cavity. The volume of the cavity was accessed using CASTp, which calculates the MS volume to be 19500Å^3 with a pocket surface area of 7200Å^2 [294]. The dimensions of the interior cavity have been calculated roughly as 30Å x 30Å x 30Å (WxHxD). The surface area of the cavity space is represented in Figure 107, coloured in magenta. The cavity would be able to accommodate small unfolded proteins and may provide an environment shielded from bulk solvent.

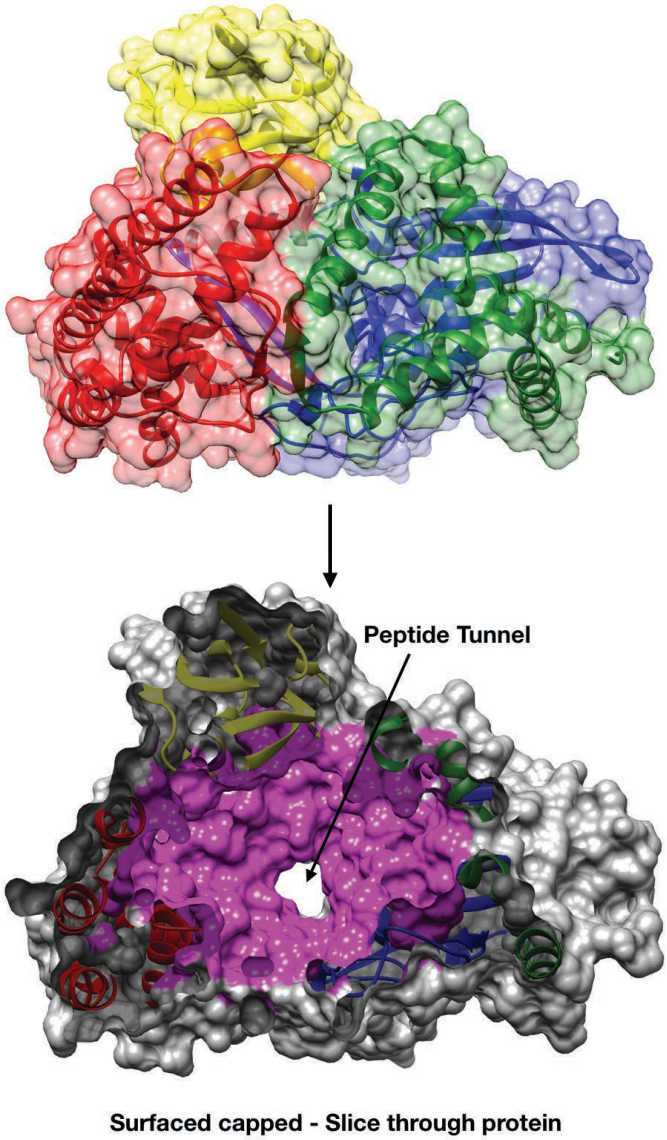


Figure 107. Bd0967 internal cavity
Bd0967 domain organisation creates a self-compartmentalised enzyme with a large internal cavity. The top image shows the exterior surface of Bd0967 coloured according to the domains. The bottom image shows a dissection of Bd0967 (orientation is the same) revealing the internal cavity surface, calculated using CASTp, shown in magenta. The peptide tunnel leading into the internal cavity is labelled.

6.2.4 The Active site / catalytic triad / bound peptide

The active site of Bd0967 is located in the protease domain with the catalytic triad positioned at the end of a shallow tunnel leading into the cavity. The peptide-binding tunnel is formed between the protease domain and protease cap subdomain. The catalytic triad is comprised of the proposed active-site nucleophile S457, the catalytic base K433, and Q487 that makes a hydrogen bond with K433 to orientate it correctly for catalysis similarly to other CTPs.

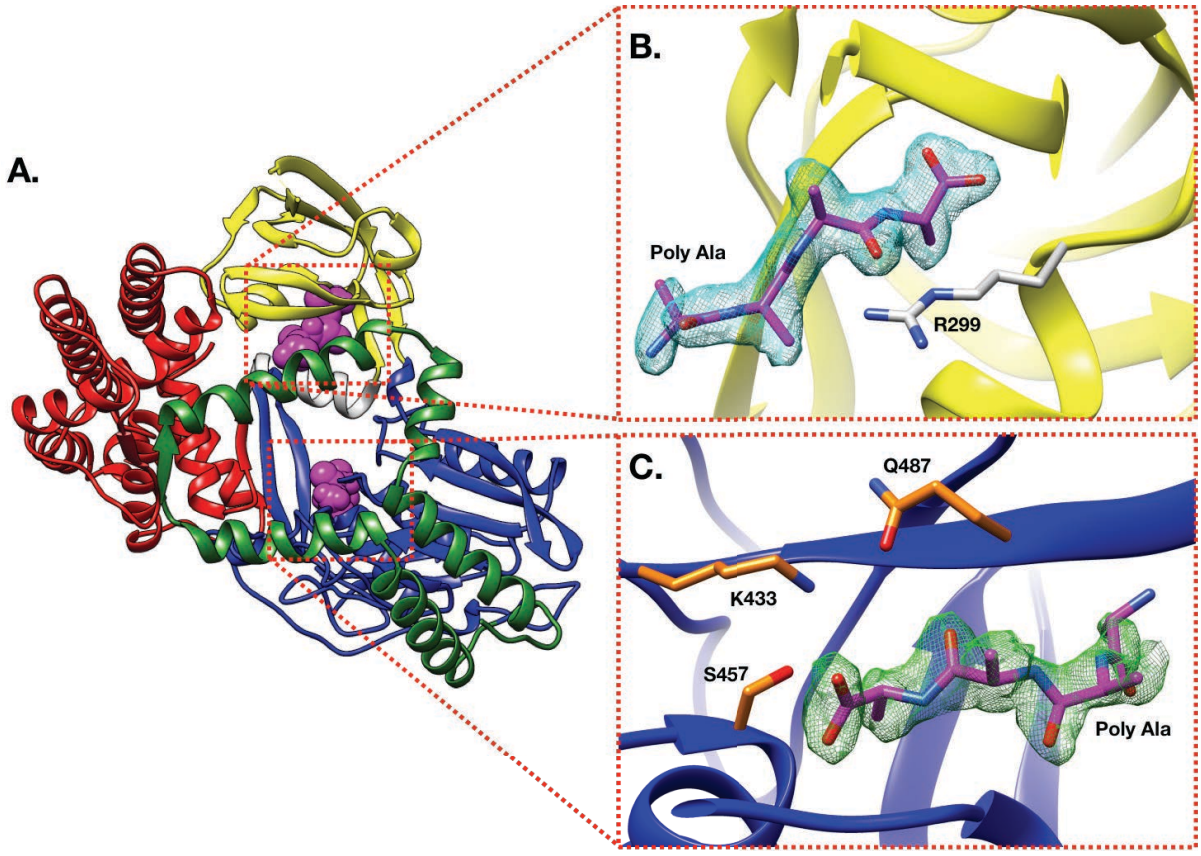


Figure 108. Bound peptides of Bd0967
Bd0967 co-purifies and co-crystallises with two peptide fragments that bind to the PDZ peptide binding site and the active-site peptide binding site. **A.** The peptides coloured magenta in space filling shown to display there positions within the structure of Bd0967. **B.** Shows the electron density map (cyan) of the bound peptide (magenta) at the PDZ domain. R299 is shown in grey for reference. **C.** Shows the electron density map (green) of the bound peptide (magenta) at the protease domain. The active-site residues are shown in orange as a reference. ($2F_o$-F_c map, radius=1.5Å, σ=1.3)

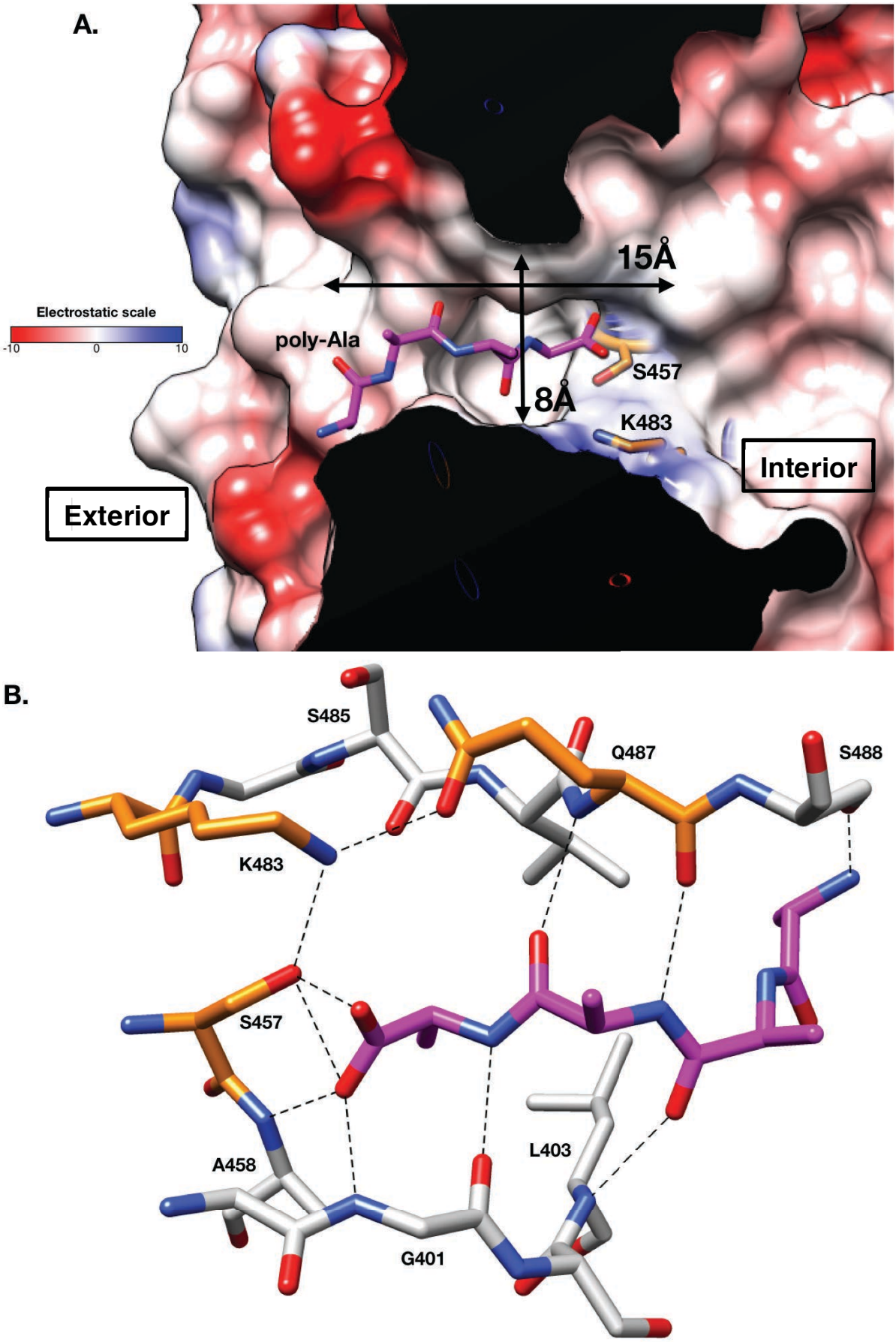


Figure 109. Protease peptide tunnel and binding site
A. Bd0967 has a shallow tunnel leading to the interior cavity, formed between the protease domain and the protease cap subdomain. The tunnel is mainly lined with hydrophobic residues. The active site residues are located towards the end of the tunnel near the interior opening. **B.** Binding of the peptide at the active site is facilitated by multiple hydrogen bonds. The peptide, coloured in magenta, is bound in a beta-augmented fashion. The active-site residues are coloured in orange and other residues participating in the peptide binding are in grey. Black lines depict the hydrogen bonds.

6.2.5 Bd0967 has a unique metal-binding site

Bd0967 possesses a metal-binding site that is unique to this family of CTPs. The metal-binding site is located in a peripheral loop region of the protease domain (Figure 110). The metal is co-ordinated by seven atoms: a water molecule; a single side chain residue (D436) and five backbone carbonyl groups (G411, F414, Q415, D438 and M440). Binding of the metal ion is facilitated by the placement of P439 that allows the specific orientation of backbone carbonyl groups in this region. The binding of a metal ion at this site likely provides a structural role by stabilising the otherwise flexible coiled region. A Mg^{2+} ion has been modelled in this position due to its presence in the crystallisation conditions and is in agreement with the observed diffraction data.

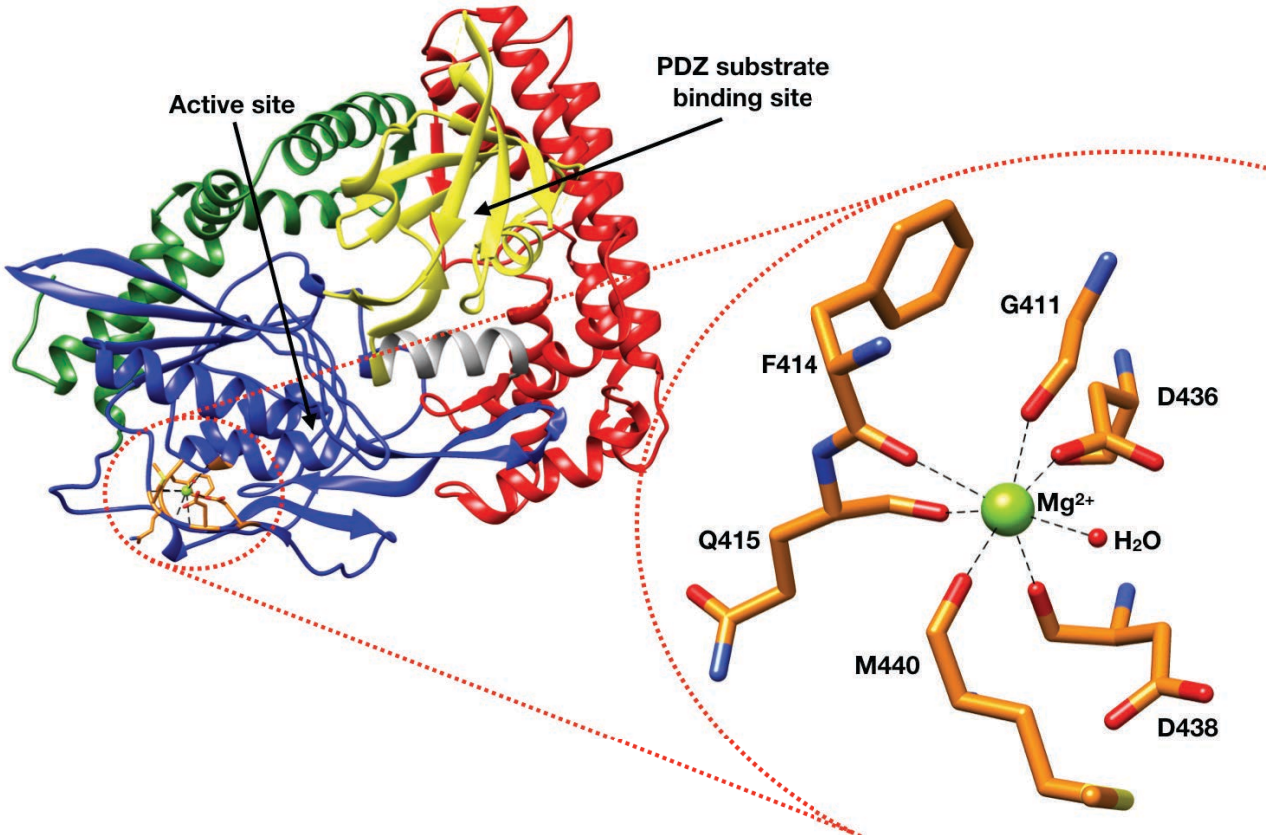


Figure 110. Bd0967 metal-binding site
The metal-binding site of Bd0967 is located in a loop at the base of the protease domain (blue).The residues involved in binding the metal ion are coloured in orange. The Mg^{2+} metal ion is coloured green and makes seven coordinated bonds to five backbone carbonyl groups the carboxylate group of D436 side chain and also to a coordinated water molecule.

6.2.6 Bd0967 has unique disulphide bond

Bd0967 has a disulphide bond located in the NTD and is not a feature previously documented for CTP (Figure 105). The disulphide bond is formed between Cys33 and Cys100 and acts to stabilise the base of the α-helical scaffold domain (NTD). Cys33 is located 4aa from the N-terminal peptide resolved in the structure of Bd0967, beyond this point no electron density is observed. The disulphide bond likely acts to stabilise the N-terminus and therefore the whole NTD, which provides the scaffold for the structure of Bd0967. The existence of a bona fide disulphide bond adds further evidence that Bd0967 possesses a signal peptide and is likely to be transported to the periplasm for its specific function.

6.2.7 Substrate recognition by the PDZ domain

The PDZ domain of Bd0967 is attached to main body of Bd0967 on a flexible hinge. At the base of the hinge are two hydrophobic residues, L234 and L332 that could easily slide past each other, in a greased motion, to facilitate the movement of the PDZ domain.

An additional peptide fragment is bound in the PDZ peptide binding groove, which is believed to originate from the same protein as the peptide fragment bound to the protease domain. We have modelled the peptide fragment, bound at the PDZ binding groove, as a four amino acid poly-alanine chain. Although positive density in the refined F_o-F_c electron density map exists for residue sidechains, at the resolution obtained the sequence register is ambiguous, furthermore MS was unable to identifying the peptide fragment.

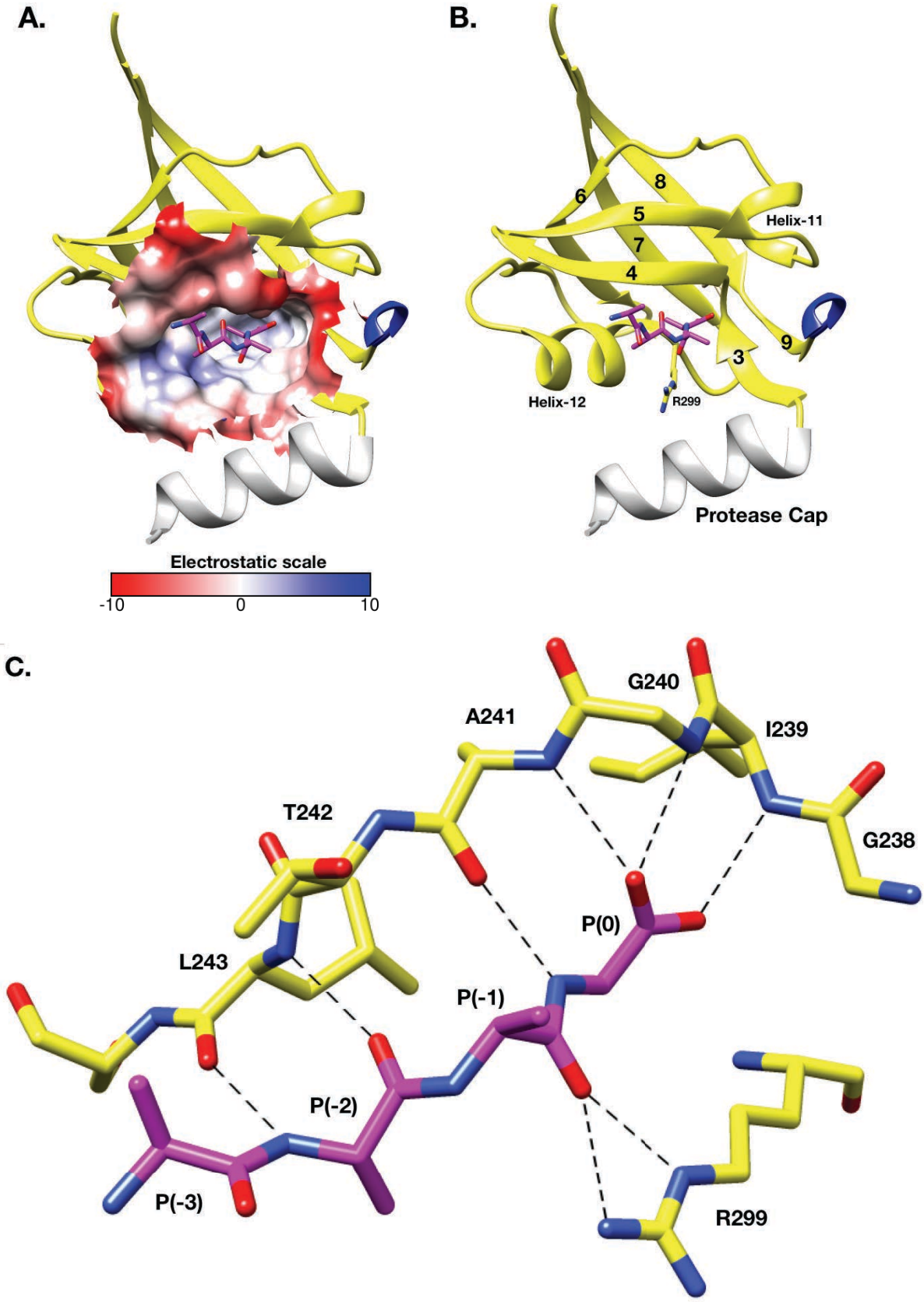


Figure 111. Peptide binding PDZ domain
A. The peptide binding groove of the PDZ domain is highly hydrophobic. The surface of the peptide binding groove is shown coloured according to the coulombic scale. **B.** Image of PDZ domain binding to peptide in same orientation as (A) to highlight the relative position of R299. The strands and helices are labelled for reference to text. **C.** Hydrogen bond network involved in peptide binding at the PDZ domain.

In addition the sidechain of R299 makes likely important hydrogen bonds to the peptide chain. The R299 sidechain aligns parallel to the peptide chain making a β-sheet like contact. R299 sidechain makes additional hydrogen bonding contacts with E226 and D227 of the protease cap (Figure 112). The H-bonds R299 makes are likely to be important for substrate sensing and activation of the protease domain, similarly to the observed conformational change observed between the 'resting' and 'active' states of CtpB [85].

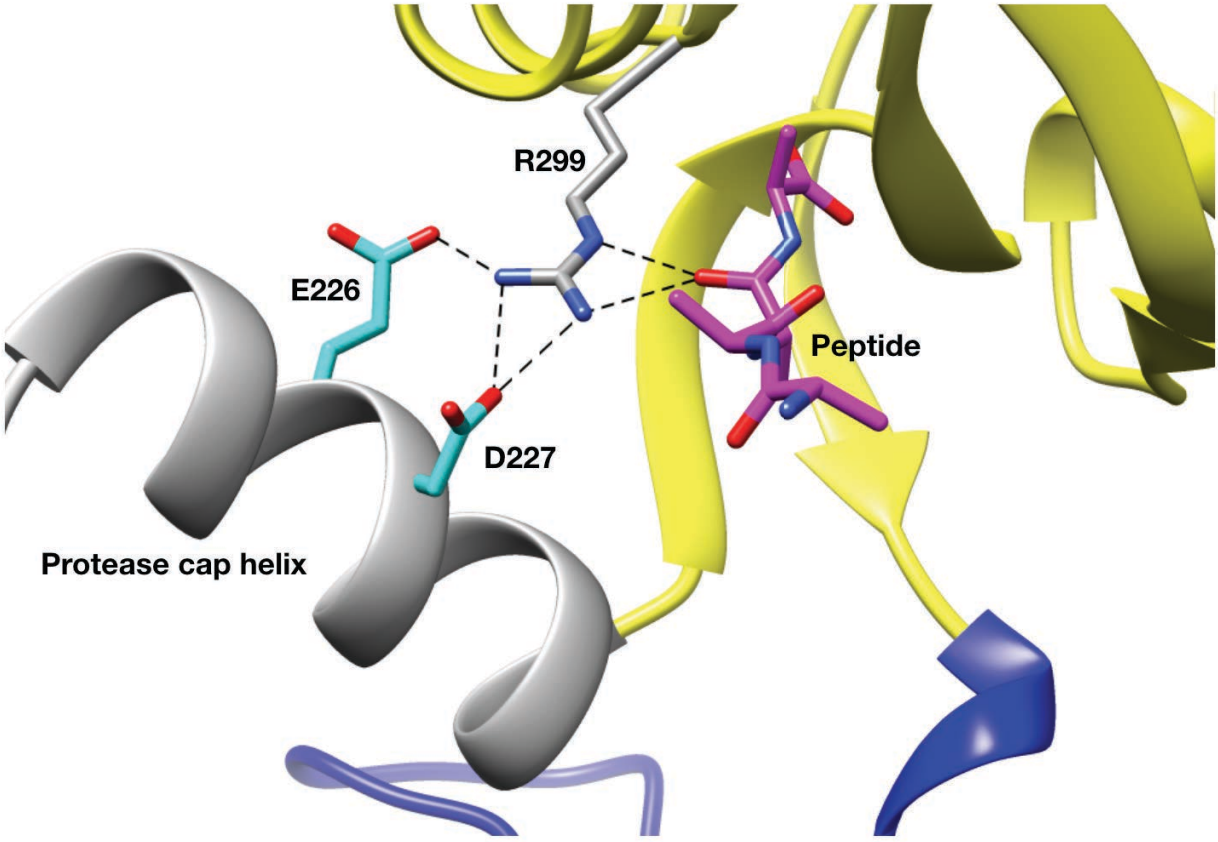


Figure 112. Arg299 interaction with peptide and protease cap helix
Peptide binding at the PDZ domain is facilitated by R299 by imitating a beta strand. R299 makes similar hydrogen bonding contacts with the protease cap helix that are responsible for switching the protease domain on and off in CtpB. R299 hydrogen bonds to two acidic residues E226 and D227 (cyan) as well as maintaining hydrogen bonds with the bound peptide (magenta).

6.2.8 A monomer in solution

SEC was used to determine the oligomeric state of Bd0967 in solution. The crystal structures also suggest Bd0967 is a monomer; however, homologues of Bd0967 have been shown to exist as functionally active dimers. SEC of Bd0967 (Wt, S457A and R199M constructs) revealed all exist in a monomeric form; however, a sizeable void peak was present for all constructs.

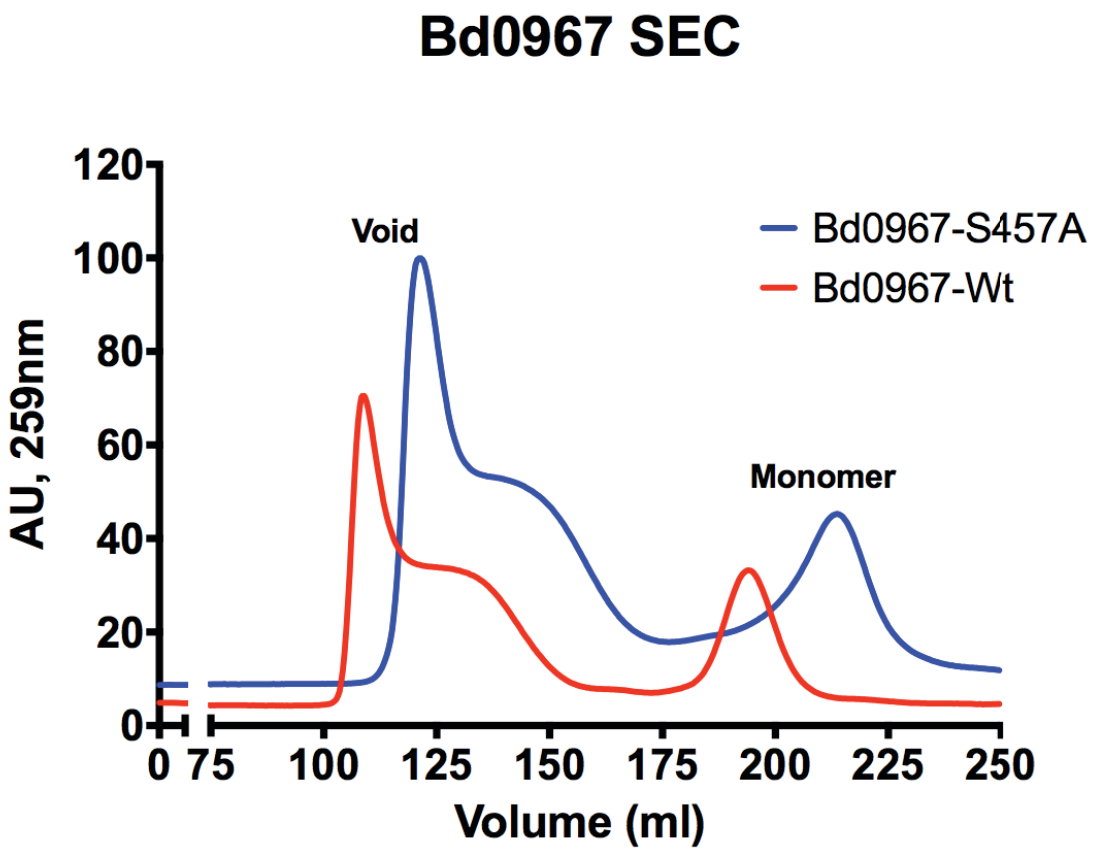


Figure 113. SEC results for Bd0967
The size exclusion traces for Bd0967 WT and S457A are shown. Monomeric Bd0967 elutes at approximately 190 ml. A large void peak is observed for Bd0967.

6.2.9 Structure of Bd0967 S457A

The structure of Bd0967-S457A is highly similar to the structure of Bd0967-WT, adopting the same fold and domain arrangement. Little conformational change is observed between WT and S457A structures with the exception of a slight rigid body rotational movement of the PDZ domain (6°). Although this slight conformational change is not seen for all chains (two out of four), suggesting conformational sampling of the PDZ domain. Furthermore, bound peptides are located at the PDZ peptide binding site and active site of Bd0967-S457A demonstrating substrates are still able to bind even though catalysis cannot be performed.

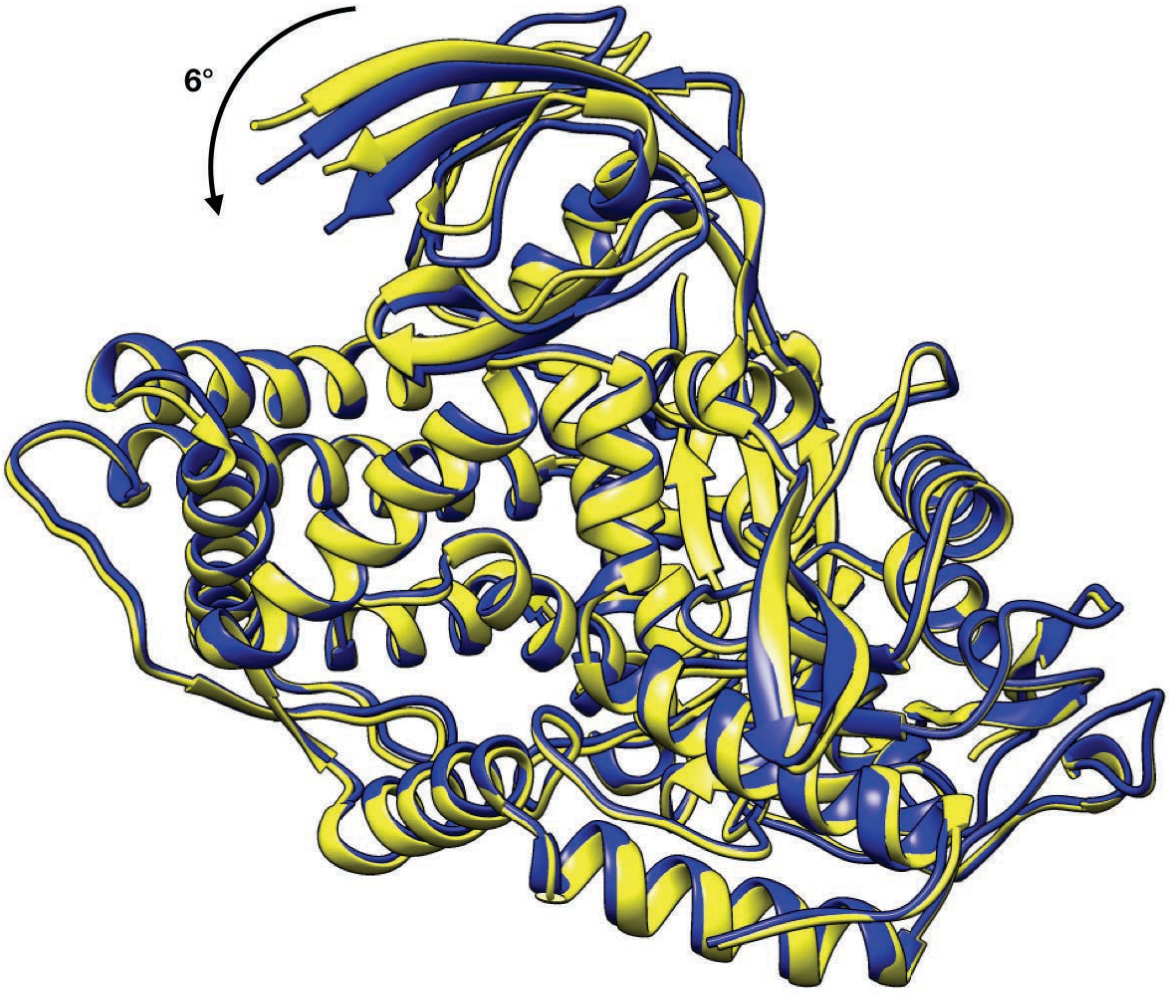


Figure 114. Structure of Bd0967-S457A
Bd0967-S457A (yellow) superimposed onto the structure of Bd0967-WT (blue) demonstrates the slight conformational change observed for the PDZ domain of Bd0967-S457A. Superimposition was performed using the SSM superpose tool in coot (length = 634 residues, 1.13Å).

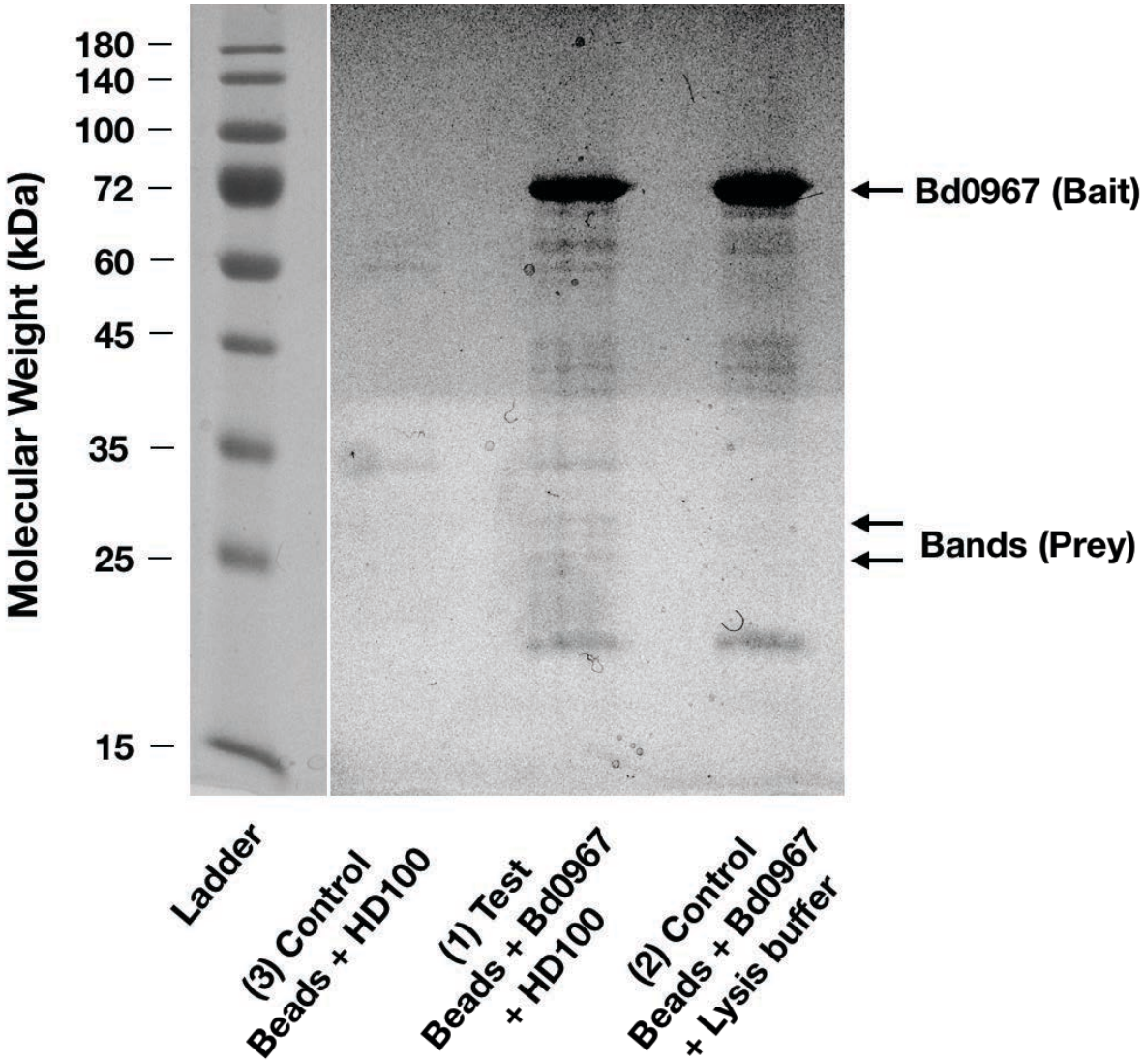


Figure 115. Bd0967 pull-down assay
The SDS-PAGE of the pull-down assay with Bd0967 as the bait, shown with the upper black arrow. The two faint bands in the test sample (1) are highlighted by the two lower black arrows.

Identified in Samples:	Organism	Protein Name	Uniprot accession code	Gene Name	Protein Size	Normalised total spectra			Terminal residues
						(1)	(2)	(3)	
(1)	*Bdellovibrio HD100*	Flagellin	Q6MQ75_BDEBA	Bd0606	29 kDa	33	0	0	ALRLIG
	Bdellovibrio HD100	Serine protein kinase	Q6MR70_BDEBA	Bd0228	80 kDa	9	0	0	GDVKNK
	Bdellovibrio HD100	Methyltransferase	Q6MP37_BDEBA	Bd1029	31 kDa	4	0	0	QARART
	Bdellovibrio HD100	ATP synthase subunit beta	ATPB_BDEBA	Bd3897	51 kDa	4	0	0	AKKLQA
	Bdellovibrio HD100	Probable signale peptide	Q6MK10_BDEBA	Bd2606	22 kDa	4	0	0	LIKFIP
(1) + (2)	*E. coli Strain K12*	cAMP-activated global transcriptional regulator CRP	CRP_ECOLI	crp	24 kDa	33	37	0	VVYGTR
	E. coli Strain K13	ATP-dependent RNA helicase RhlE	RHLE_ECOLI	rhlE	50 kDa	12	21	0	KPAAAQ
	E. coli Strain K14	3-ketoacyl-CoA thiolase	FADA_ECOLI	fadA	41 kDa	7	9	0	TVFERV
	E. coli Strain K15	Succinylornithine transaminase	ASTC_ECOLI	astC	44 kDa	7	9	0	VSRGSS
	E. coli Strain K16	tRNA (guanosine(18)-2'-O)-methyltransferase	TRMH_ECOLI	trmH	25 kDa	6	9	0	TMQAAG
(1) + (3)	*Bdellovibrio HD100*	Inosine-5'-monophosphate dehydrogenase	Q6MLD1_BDEBA	Bd2081	52 kDa	33	0	37	NYRIES
	Bdellovibrio HD100	Octaprenyl diphosphate synthase	Q6MQB9_BDEBA	Bd0557	37 kDa	26	0	39	WRRKPS
	Bdellovibrio HD100	Flagellin	Q6MIU6_BDEBA	Bd0352	29 kDa	37	0	5	ALKLIG
	Bdellovibrio HD100	Thioredoxin reductase	Q6MQT7_BDEBA	Bd0373	34 kDa	20	0	23	EAQEAH
	Bdellovibrio HD100	Flagellin	Q6MQ77_BDEBA	Bd0604	29 kDa	20	0	4	VAMKLV
(1) + (2) + (3)	*Bdellovibrio HD100*	Tail-specific protease	Q6MP94_BDEBA	Bd0967	75 kDa	315	376	14	KQANAK

Table 16. Mass Spectrometry results for Bd0967 pull-down assay
The top five proteins identified for each category is displayed in this table. Significantly increased normalised spectra (for proteins identified from sample (1) = beads + Bd0967 + Lysate is highlighted in green. Yellow indicates the low normalised spectra for proteins in control sample that had increased normalised spectra in sample (1).

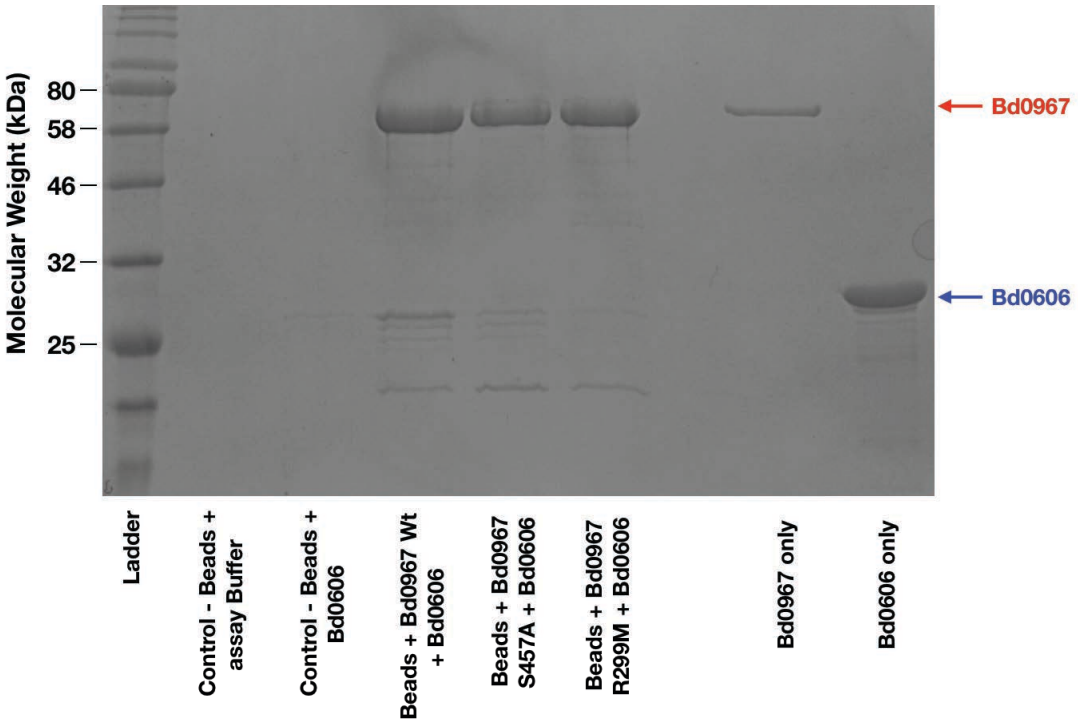


Figure 116. Confirm binding
A similar pull-down assay was re-performed with purified Bd0606 to evaluate the MS result. Bd0606 binds to Bd0967-Wt and to a lesser extent Bd0967-S547A. Bd0606 does not bind to the PDZ binding mutant Bd0967-R299M. The binding is weak (not 1:1 stoichiometry) under the assay conditions.

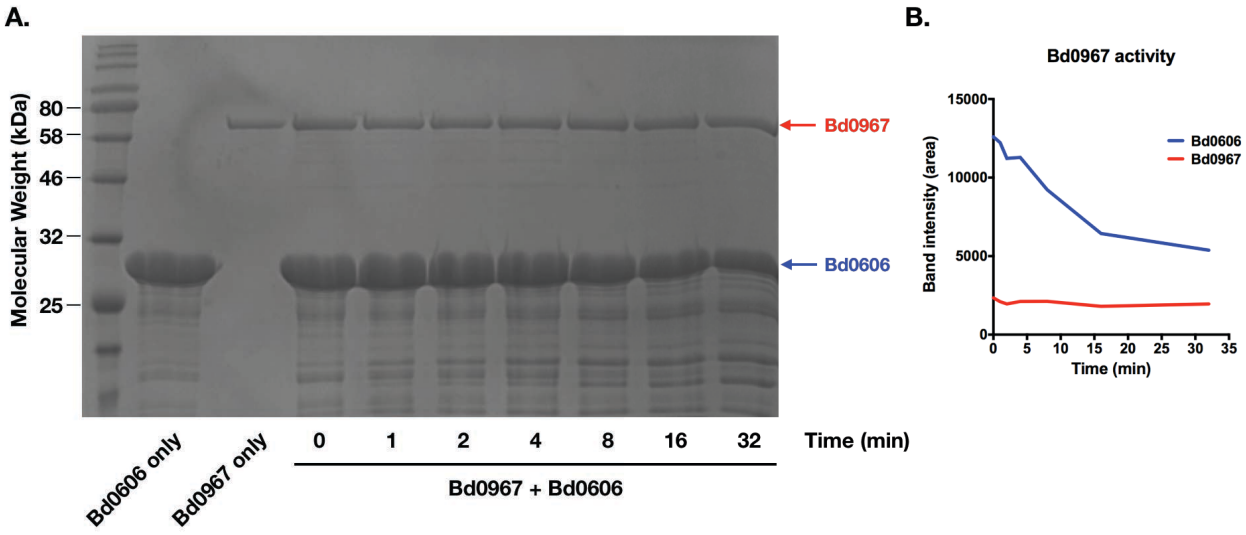


Figure 117. Bd0967 degrades Bd0606
A. Bd0967-Wt degrades Bd0606 in a processive manner, which was monitored over a time period. The red arrow highlights Bd0967 and the blue arrow highlights Bd0606. **B.** Graphical illustration of Bd0606 being degraded over time (band intensity) alongside Bd0967.

A similar result was observed when Bd0606 was incubated with the catalytic mutant Bd0967-S457A. Bd0606 remained at a similar level throughout the 1 hour incubation when incubated with Bd0967-S457A mutant suggesting that degradation is being performed by Bd0967 in the Wt sample. Furthermore, a Bd0967-R299M (PDZ peptide binding mutant) was also incapable of degrading Bd0606. This finding supports the importance of the PDZ domain interaction with Bd0606 (substrates) to facilitate catalysis.

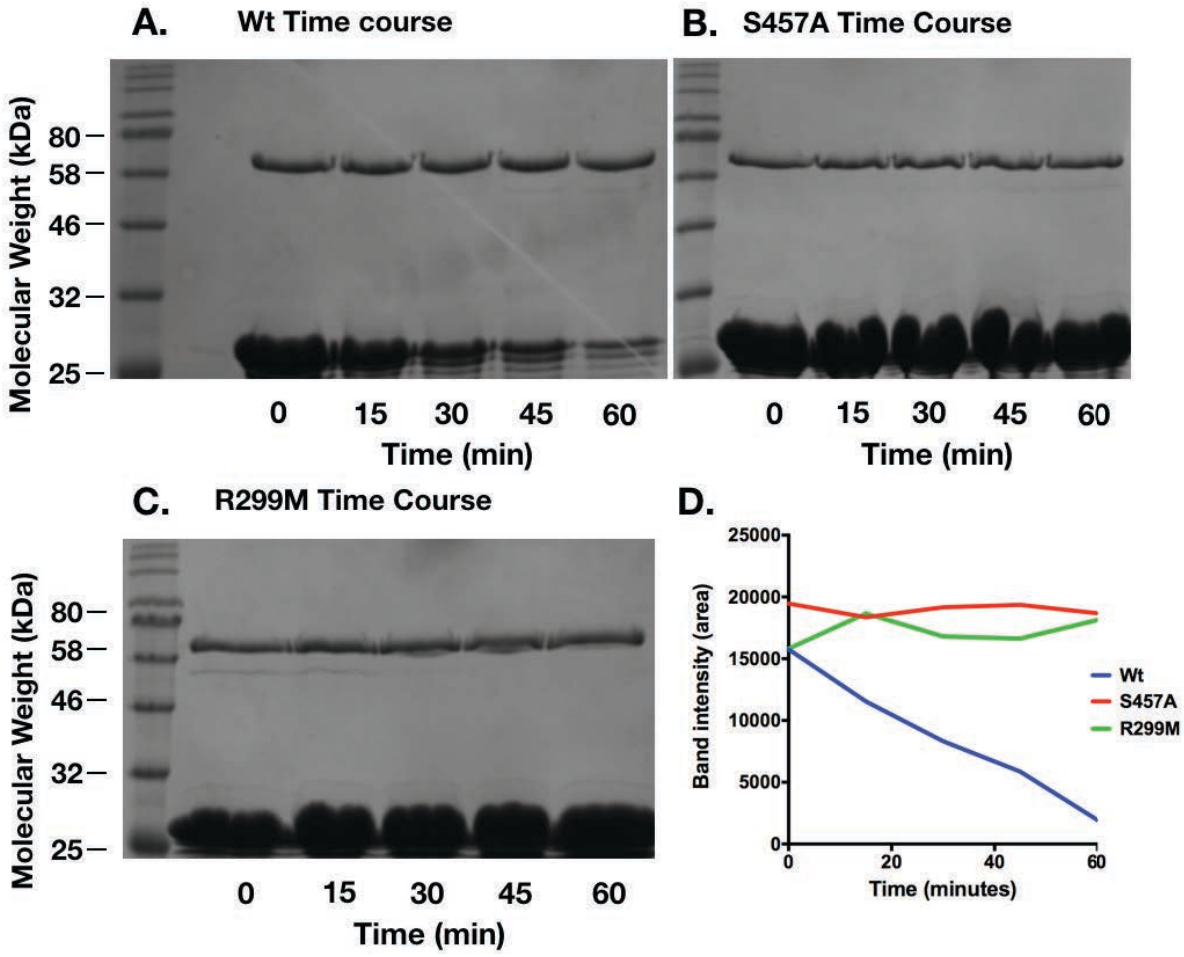


Figure 118. Bd0967 activity in comparison to mutations
A. Bd0967 Wt degradative activity towards Bd0606 in a 60 minute time period with samples taken at regular time intervals. **B.** S457A mutant does not show any activity. **C.** R299M also lacks the ability to degrade Bd0606 efficiently due to a non-functional PDZ domain. **D.** A graphical representation showing the degradation of Bd0606 when incubated with Bd0967-Wt, S457A and R299M mutants. Image J was used to analyse the intensity of gel bands.

To determine whether Bd0967 was specifically targeting Bd0606 for degradation we also investigated the activity of Bd0967 towards a variety of other *Bdellovibrio* proteins of varying size and C-terminal sequences (Figure 119). This assay did not reveal any evidence that Bd0967 was active towards any of these other proteins. Furthermore Bd0108 (a small intrinsically disordered protein) was also not a substrate for Bd0967 suggesting that a disordered C-terminal sequence is not the only prerequisite to binding and catalysis. Additionally, Bd0967 showed no degradative activity towards Bd1996 and Bd0314, which are both similar molecular weights to Bd0606.

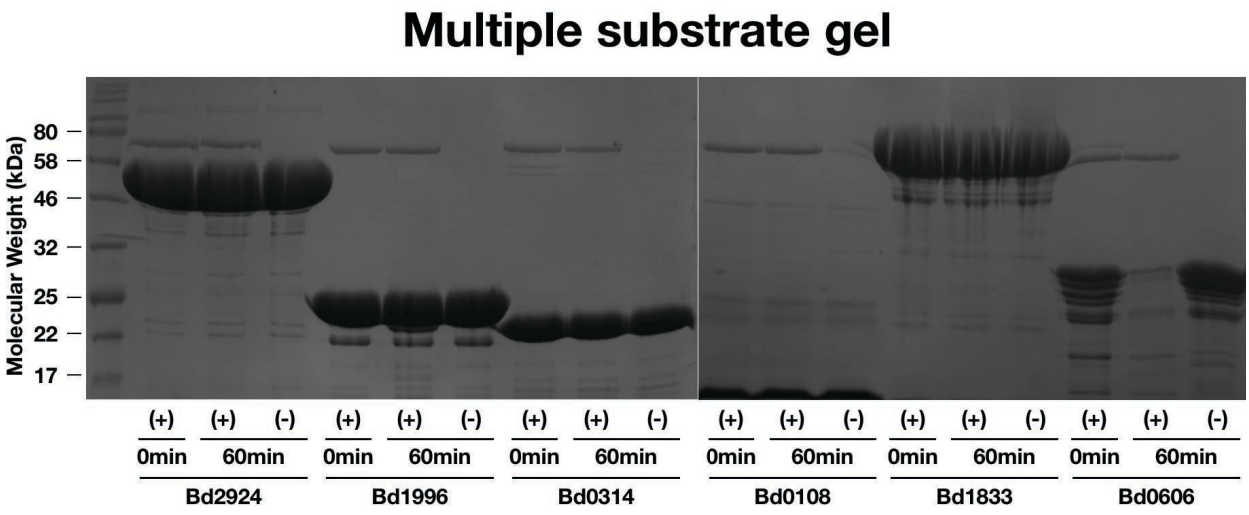


Figure 119. Bd0967 activity towards various proteins
Bd0967 degradative activity has been analysed against *Bdellovibrio HD100* proteins of varying sizes and C-terminal sequences. Bd0967 showed no activity towards any of the proteins trialled in this assay. Bd0108 is a small intrinsically disordered protein that remained un-digested suggesting the folded state of the proteins is not a factor determining Bd0967 degradative ability.

To investigate the mechanism behind Bd0606 recognition and degradation by Bd0967, we produced C-terminal truncation mutants of Bd0606. These mutants, ΔLIG and ΔRLIG, lack the terminal three and four amino acids, which typically interact with the PDZ domain of CTPs. To our surprise Bd0967 was able to degrade these truncation mutants of Bd0606 with similar efficiency to WT Bd0606 (Figure 120). This finding

along with the Bd0967-R299M mutant finding suggests that although a functioning PDZ domain is absolutely required for recognition of the C-terminal sequence and initiating catalysis; Bd0967 also has a degree of promiscuity and tolerance in peptide binding at the PDZ domain. On the other hand, Bd0967 may also adopt an additional mechanism of substrate recognition, which could also explain these results.

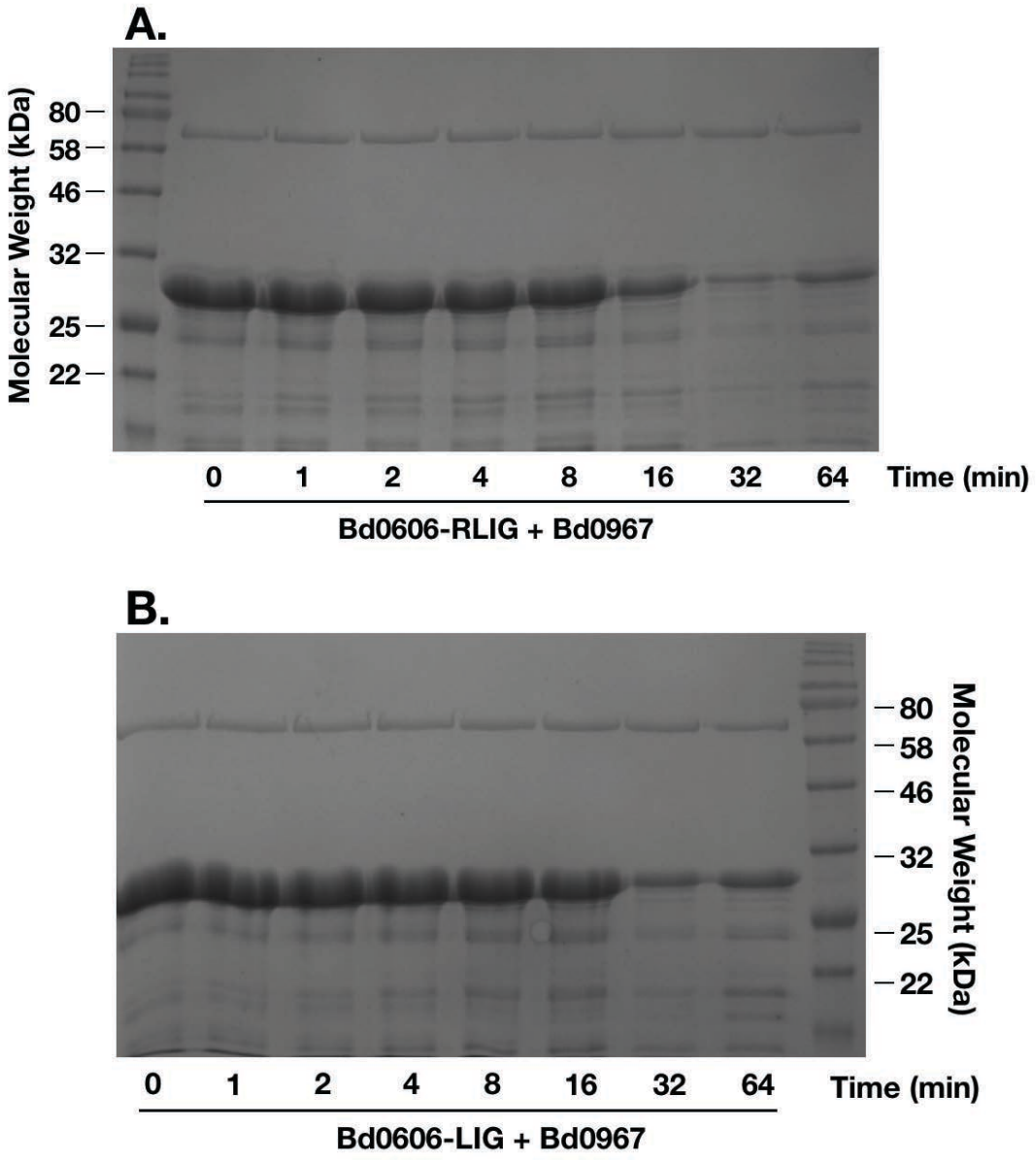


Figure 120. Bd0606 C-terminal mutants
A. Bd0606 mutant RLIG lacks the C-terminal four amino acids, which bind to the PDZ domain. **B.** Bd0606 mutant LIG lack the three C-terminal amino acids. Bd0967-Wt is able to degraded both RLIG and LIG mutants to a similar degree as WT Bd0606 substrate. This finding suggests Bd0967 PDZ binding site allows promiscuity and also suggest Bd0606 may be recognised via an alternative mechanism distinct from direct interaction with the C-terminal.

6.2.12 The C-terminal peptide of Bd0606 as an inhibitor of Bd0967 activity

We designed and created a 12aa peptide matching the C-terminal sequence of Bd0606 (NAMPNSALRLIG) in order to investigate whether C-terminal binding was required for Bd0967 degradative activity. This 12aa peptide of Bd0606 was kindly prepared by Oliver Daubney by N-terminal peptide synthesis to ensure that the α-carboxylate was present. Bd0606 was incubated with Bd0967-Wt in buffer containing increasing concentration of peptide. The results of this assay show that increasing the concentrations of peptide reduces the degradation of Bd0606 (Figure 121). At low concentrations of peptide Bd0967 remains active and the levels of Bd0606 are reduced, whereas at a high peptide concentration the degradative activity of Bd0967 is reduced. This result demonstrate that the peptide fragment can act as an inhibitor of Bd0967, presumably by interacting with and saturating binding of the PDZ domain, which disables the activity of Bd0967.

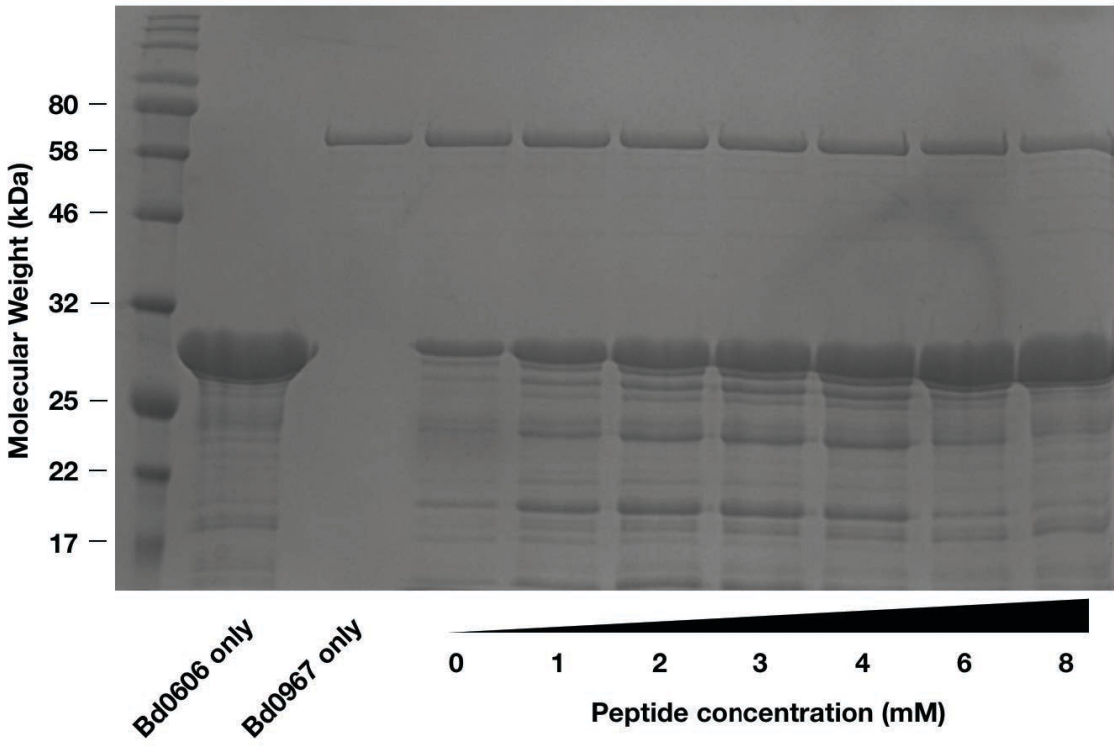


Figure 121. Peptide inhibition assay
Bd0967 activity was measured in the presence of an increasing concentration of a 12aa peptide fragment, synthesised to match the C-terminus of Bd0606. At high peptide concentrations Bd0967 degradative activity towards Bd0606 was reduced indicating the PDZ domain is important for substrate recognition.

6.2.13 Bd0606 is folded

Circular Dichroism (CD) was performed on Bd0606 to determine whether it was correctly folded. CD is a good biophysical technique for determining the proportion and nature of a folded protein based on a distinct UV spectrum for α-helix and β-sheet. The CD spectrum revealed that Bd0606 is folded, with α-helices being the major secondary structure element, which is expected for a flagellin protein based on previous structural information [295]. This finding supports Bd0606 being correctly folded and Bd0967 is not degrading unfolded proteins in a quality control degradative manner.

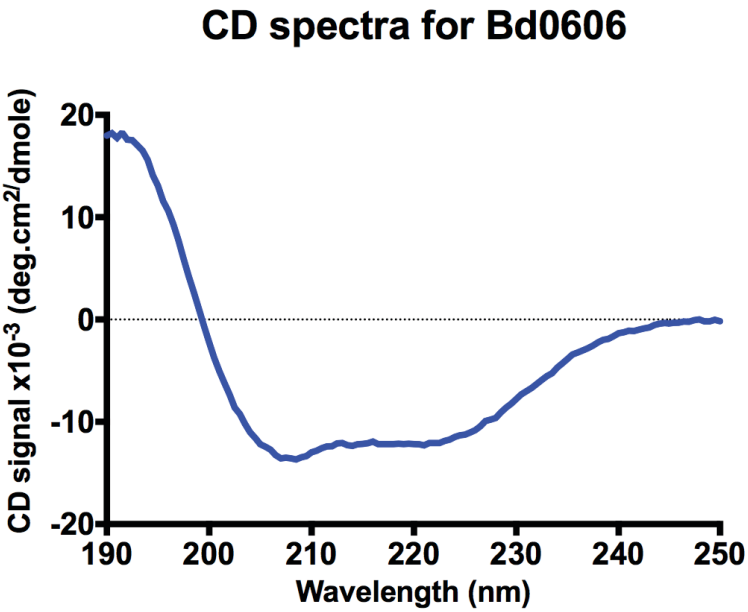


Figure 122. Bd0606 CD spectra
CD spectra for Bd0606 reveals Bd0606 is folded and is formed predominantly of α-helices. Minima peaks at ~208 nm and 222 nm and a maxima peak at 193 nm exemplify a α-helix folded proteins.

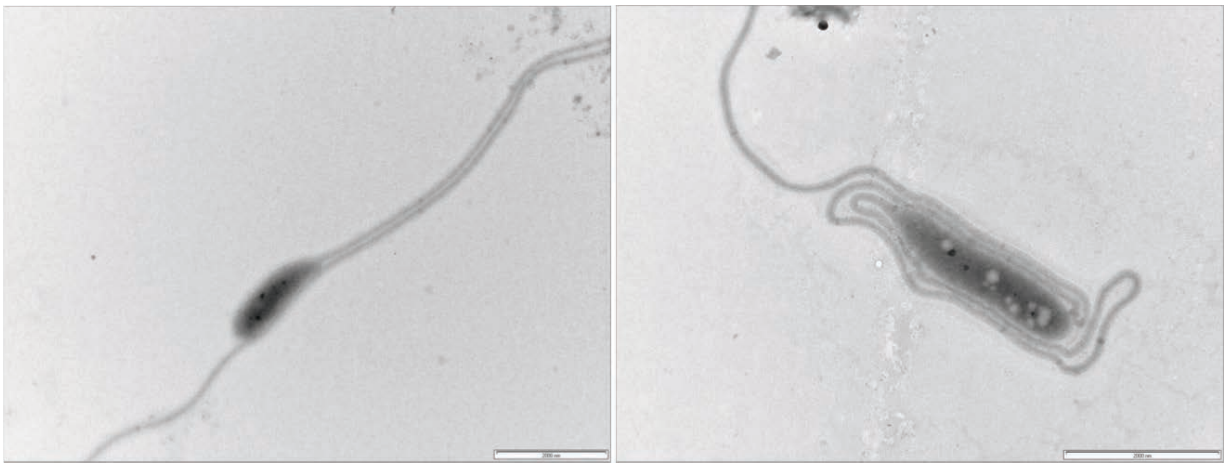


Figure 123. Bd0967 KO mutant *Bdellovibrio*
Bd0967 KO mutants have a pleiotropic developmental phenotype that involves the flagellum assembly. The flagellum of Bd0967 KO *Bdellovibrios* have been seen to grow to ~5 times the length of Wt, suggesting flagellin assembly is unregulated. These two images show the large mutant flagellar. **(Electron Micrograph images kindly provided by the Sockett lab)**

6.2.15 Bd0967 interaction with CdG

MST was used to determine whether Bd0967 interacted with CdG. A K_D value of 22 μM was obtained, suggesting weak non-specific binding. Furthermore, Bd0967 has a signal-peptide, suggesting transport into the periplasm where CdG is unlikely to be present.

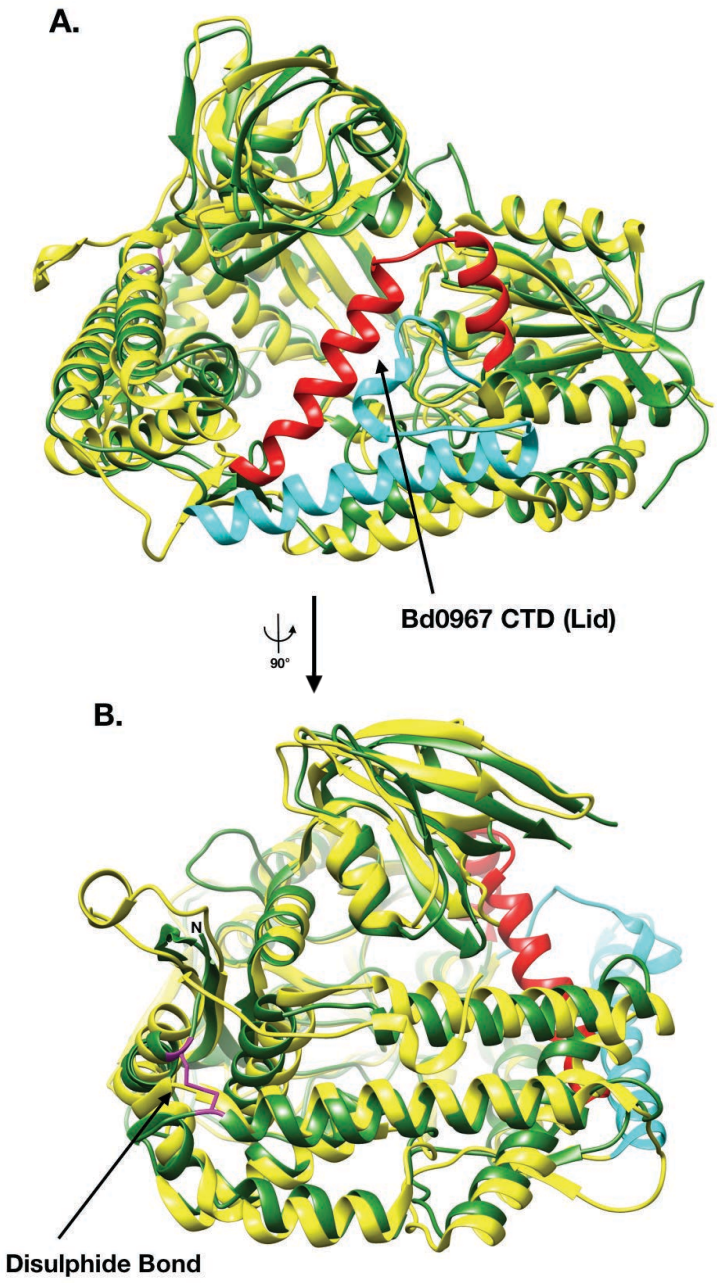
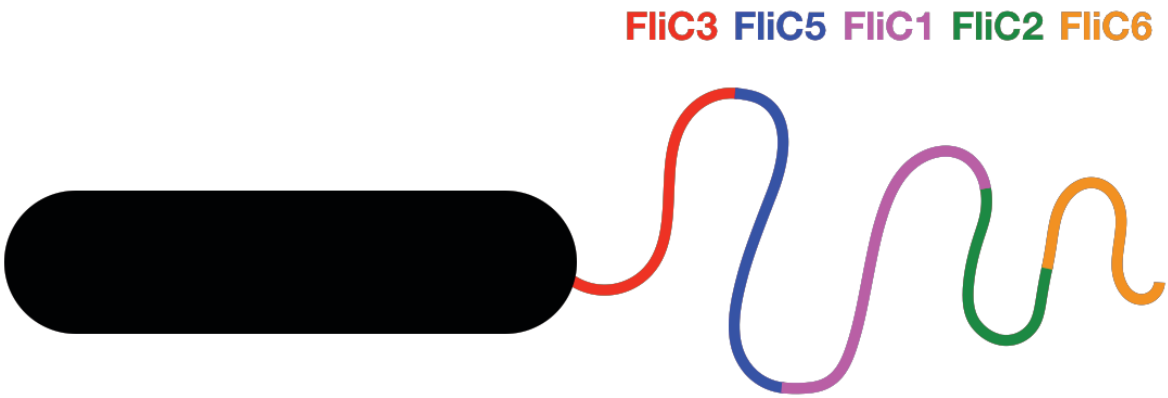


Figure 124. Bd0967 comparison to *E. coli* Prc
A. The structure of Bd0967 (green) is superposed with *E. coli* Prc (yellow, PDB:5WQL). Superimposition was performed using the SSM superpose tool in coot (length = 536 residues, rmsd = 2.01Å). The proteins three-dimensional structures are highly similar and overlap well in all domain except from the CTD. The CTD of Bd0967 is shown in red, while the CTD of Prc is shown in Cyan. The CTD of Bd0967 forms more of a paddle-like structure creating a cavity shielded from bulk solvent, whereas Prc CTD forms one wall of a bowl shape. **B.** Bd0967 also has a disulphide bond (magenta) located in the NTD, which is not found in other C-terminal processing proteases.

hydrophobic C-terminal sequence making it a suitable substrate for Bd0967. Moreover other *Bdellovibrio* flagellins have highly similar C-terminal sequences (Figure 125), suggesting they may also be potential substrates, which requires experimental verification.



Flagellin	Size (aa)	C-terminal sequence
FliC1 (Bd0604)	274	ANQLPQVAMKLV
FliC2 (Bd0606)	277	NAMPNSALRLIG
FliC3 (Bd0408)	282	VALGLLNKSFQG
FliC4 (Bd0410)	277	NQSQQLALKLLG
FliC5 (Bd3052)	277	NQSNQLALKLIG
FliC6 (Bd3342)	277	NQANQLALKLIG

Figure 125. Flagellin assembly *Bdellovibrio*
Bdellovibrios flagellum is composed of six flagellins in an ordered arrangement from the proximal end to the distal end, producing a tapered wave formation. The location of each flagellin species, within the flagellum, is illustrated above in the cartoon representation. Only FliC4 location has not been determined. The table highlights the similarities of flagellin C-terminal sequences.

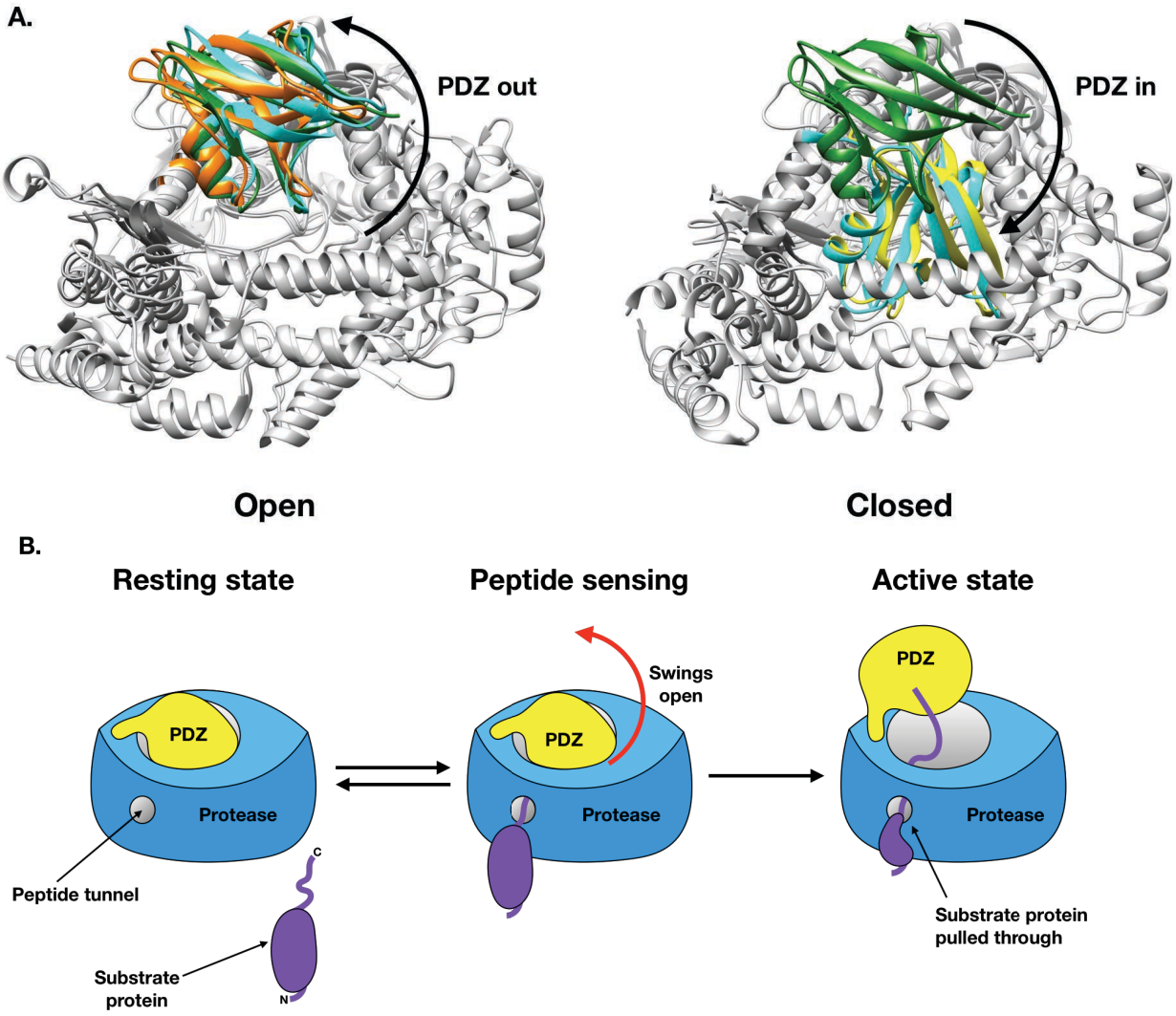


Figure 126. Bd0967 mechanism
A. The movement of the PDZ domains in C-terminal processing proteases is associated with regulating protease activity. Bd0967 PDZ domain (green) is in an outward conformation and its location is compared, relative to other PDZ domains of C-terminal processing proteases. Bd0967 PDZ also aligns well with CtpB-active (cyan, PDB:4C2C, length = 302 residues, rmsd = 2.37Å) and Prc (orange, PDB:5WQL, length = 536 residues, rmsd = 2.01Å) PDZ domains in the active "out" position. The PDZ domains of CtpB-resting (cyan, PDB:4C2E, length = 222 residues, rmsd = 3.40Å) and D1P (yellow, PDB:1FC7, length = 234 residues, rmsd = 3.03Å) have been shown to adopt an inward position (resting state), fitting close to the protease peptide binding tunnel. Superimposition was performed using the SSM superpose tool in coot. **B.** A cartoon illustration of the proposed mechanism of Bd0967. Bd0967 exists in the resting state with the PDZ domain tucked into the bowl of the protein. When the PDZ domain binds / recognises the C-terminal peptide sequence of its substrate the PDZ domain undergoes a conformational change and swings into an out position, activating the protease domain. The PDZ domain movement pulls the substrate through the peptide tunnel past the catalytic site residues for degradation (unfolding in the process).

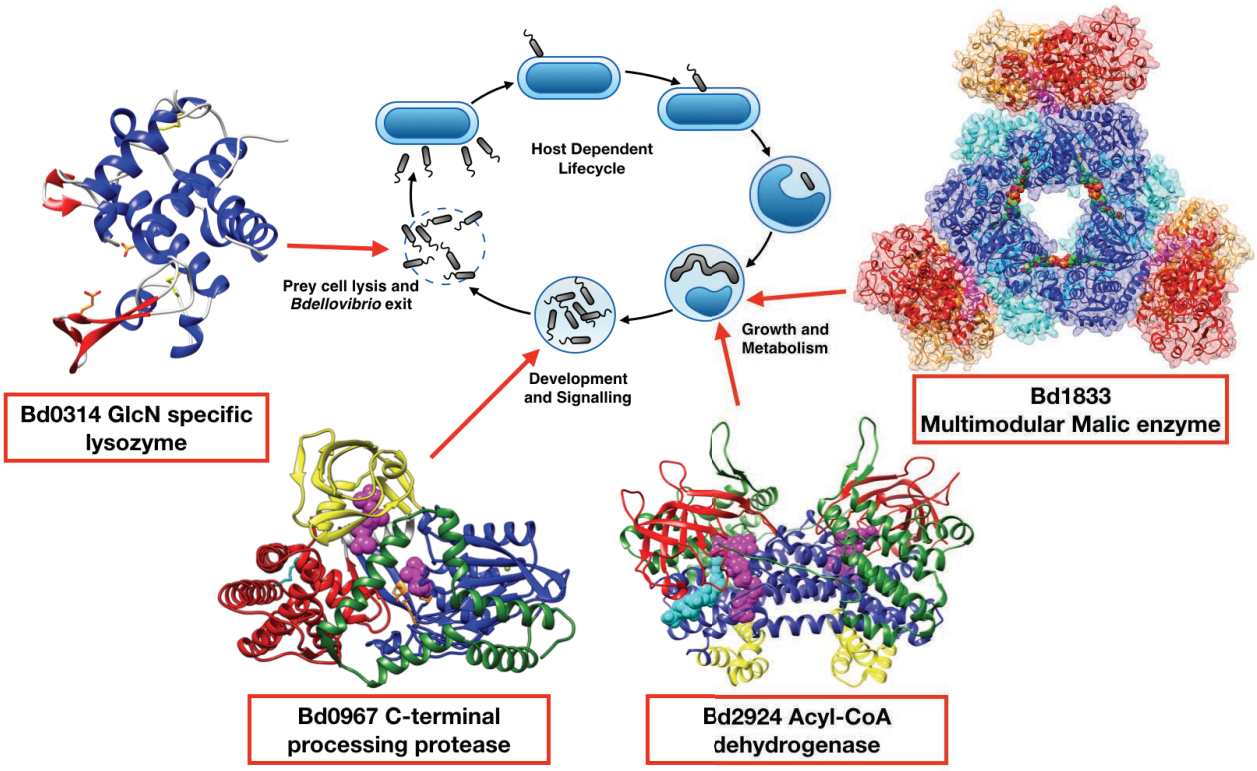


Figure 127. Summary of thesis findings
This thesis has structurally and functionally investigated four proteins with diverse roles within the predatory lifecycle. Bd1833 is a multi-modular malic enzyme involved in an alternative metabolic pathway, diverting TCA cycle intermediates and switching to a carbon storage pathway that is allosterically inhibited by acetyl-CoA. Bd2924 is a divergent putative acyl-CoA dehydrogenase potentially involved in metabolism of prey cell lipids. Bd0967 is a C-terminal processing protease involved in a critical cell developmental pathway involving flagellin degradation. Bd0314 is a modified lysozyme that has been specifically adapted to exit the invaded prey cell, acting specifically towards N-deacetylated peptidoglycan cell walls.

Two of the proteins investigated in this study (Bd2924 and Bd1833) are enzymes that have roles in central metabolism. Central metabolism has been an understudied area of research in the predatory lifecycle of *Bdellovibrio*. Many of the metabolic pathways have been adapted to deal with the needs of a predator's diet, such as solely using the mevalonate pathway for isoprenoid biosynthesis [306]. Consequently the findings presented in this thesis represent the first structurally and functionally characterised central metabolic enzymes from *Bdellovibrio*. We reveal Bd2924s three-dimensional structure is related to ACADs with divergent functions, suggesting that Bd2924 has a