Microbial Pectin Recognition and Utilization in the Mammalian Gastrointestinal Tract

Ana Sofia de Jesus Vaz Luís

A thesis submitted for the degree of Doctor of Philosophy

2013 – 2016

Institute for Cell and Molecular Biosciences
Faculty of Medical Sciences
Newcastle University

Abstract

Plant cell wall polysaccharides represent one of the major nutrients available to the human microbiota, which has a significant impact on host nutrition and health. Pectins are one of the major plant cell wall components. Understanding the mechanism of protein recognition and enzymatic degradation of these polysaccharides can have significant implications, not only in promotion of human health, but also in an industrial context. In this thesis the founding member of a carbohydrate binding module (CBM) family that targets homogalacturonan (HG) was characterized. The mechanism of degradation of rhamnogalacturonan-I and II (RG-I and RG-II) by *Bacteroides thetaiotaomicron*, a member of human microbiota, was also explored.

Efficient plant cell wall degradation requires a close association between catalytic modules with CBMs. In Chapter 3, the characterization of *Ruminococcus flavefaciens* $CBM77_{PL1/9}$ revealed that this CBM specifically binds to non-methylated HG. The structural characterization disclosed a new binding mechanism where the positively charged residues (lysines) interact with the negative carboxyl groups present in the ligand. Additionally, a dimer of this CBM has shown to be a suitable probe to target pectins in a plant cell wall context.

B. thetaiotaomicron is able to grow on pectins. In this study, a model for RG-I and RG-II degradation was described. The cleavage of RG-II is achieved by a hierchical exo mode of degradation. Chapter 4 showed that the degradation of Chain B requires eight enzymes where the α-L-rhamnosidases, BT0986 and BT1019, specifically target the linkages α1,2 and α1,3-arabinopyranose, respectively. It was shown that the linkage between L-rhamnose and D-apiose is an α-linkage, likely α1,3'. BT1001, which targets the rhamnose-apiose glycosidic bond requires the cleavage of side chain and backbone, suggesting that this is the last linkage cleaved in RG-II degradation. The structural and biochemical characterization of a GH106 enzyme (BT0986) revealed a calcium dependent enzyme with a catalytic apparatus consistent with an inverting mechanism. The structure of a GH127 enzyme (BT1003) involved in RG-II showed that the aceric acidase lacks the residue that is the proposed catalytic acid/base in GH127 β-L-arabinofuranosidases. This may indicate that the aceric acidase cleaves glycosidic bonds by a mechanism that differs from other enzymes in the GH127 family.

In Chapter 5, key enzymes in the RG-I degradation system were characterized. Three polysaccharide lyases (PLs) (BT4170, BT4175 and BT4183) revealed complementary specificities. BT4170, identified on the cell surface, was shown to be essential for RG-I backbone utilization. Four GH28 enzymes were characterized: BT4149 and BT4153 target the RG-I backbone displaying distinct substrate specificities based on the degree of polymerization of the oligosaccharide. BT4155 removes the HG remnants attached to RG-I and BT4146 is specific for a RG-I backbone disaccharide. The surface glycan binding protein BT4167 has shown to recognize RG-I. The structural characterization of BT4170 (PL9) revealed that the residues implicated in substrate recognition in this rhamnogalacturonan lyase are not conserved in PL9 pectate lyases. The proposed model of how 13 enzymes encoded by *B. thetaiotaomicron* mediate the degradation of RG-I contributes to our knowledge of pectin degradation.

Acknowledgments

I would like to thank my supervisor Professor Harry Gilbert for giving me the opportunity to do this PhD. I would like to thank his guidance, tutoring and support during this work.

I would also like to thank Professor Carlos Fontes, who always supported me and believed in my qualities even when I doubted about them.

To all my colleagues in Lab M2035 for their advice and all the good moments that we shared. This work would not be possible without the collaboration with Arnaud Basle for the X-ray crystallography.

Finally, I would like to thank to my family and friends for their support.

Contents

Abstract iii
Acknowledgments v
Contents vii
List of Figures xiii
List of Tables xvii
Abbreviations xix
Journal articles xxi
Chapter 1. General introduction 1
1.1 Plant cell wall 1
1.1.1 Cellulose and hemicelluloses 2
1.1.2 Pectins 3
1.1.2.1 Homogalacturonan 4
1.1.2.2 Rhamnogalacturonan-I 4
1.1.2.3 Rhamnogalacturonan-II 5
1.1.2.4 Pectin organization in plant cell wall 7
1.2 Plant cell wall degradation 8
1.2.1 Enzymes 8
1.2.1.1 Glycoside hydrolases 8
1.2.1.1.1 Classification 9
1.2.1.1.2 Main catalytic mechanisms 11
1.2.1.2 Polysaccharide lyases 13
1.2.1.3 Carbohydrate esterases 16
1.2.1.4 Subsite classification 17
1.2.2 Carbohydrate-binding modules 18
1.2.2.1 Classification 18
1.2.2.2 CBM functions 21
1.2.3 Pectin degrading enzymes 22
1.2.3.1 Pectin lyase, pectate lyase and rhamnogalacturonan lyase 23
1.2.3.2 GH28 enzymes 24
1.2.3.3 GH105 enzymes 25
1.3 Polysaccharide degradation by human gut microbiota 26
1.3.1 Microbiota 26

1.3.2 *Bacteroides thetaiotaomicron* 29
1.3.2.1 Polysaccharide utilization loci (PULs) 30
1.4 Cellulolytic bacteria *Ruminococcus flavefaciens* 34
1.4.1 Cellulosome 34
1.5 Objectives 37
Chapter 2. Material and Methods 39
2.1 Molecular biology 39
2.1.1 Chemicals, commercial kit and water 39
2.1.2 Bacterial strains and plasmid 39
2.1.3 Sterilization 40
2.1.4 Growth media 40
2.1.5 Centrifugation 42
2.1.6 Storage of DNA and bacteria 42
2.1.7 Transformation of chemical competent *E. coli* 43
2.1.8 Growth conditions for propagation of bacteria 43
2.1.9 DNA isolation 43
2.1.10 Quantification of DNA 44
2.1.11 Polymerase Chain Reaction (PCR) 44
2.1.11.1 Standard and site direct mutagenesis PCR 45
2.1.12 Agarose gel electrophoresis 46
2.1.13 Purification of DNA fragments 47
2.1.14 Restriction digest of DNA 48
2.1.15 Ligation of insert and vector DNA 48
2.1.16 Automated DNA sequencing 48
2.1.17 Recombinant protein production in *E. coli* 49
2.1.18 Preparation of cell free extracts 49
2.1.19 Purification of proteins 50
2.1.19.1 Immobilized metal affinity chromatography (IMAC) 50
2.1.19.2 Ion exchange 51
2.1.19.3 Gel Filtration chromatography 51
2.1.20 Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE)… 52
2.1.21 Quantification of proteins 53
2.1.22 Concentration and buffer exchange of proteins 54
2.1.23 *B. thetaiotaomicro*n counter-selectable gene deletion or FLAG-Tag 54

2.1.23.1 Preparation of plasmids for genomic deletion or FLAG-Tag 54
2.1.23.2 Conjugation into *B. thetaiotaomicron* 56
2.1.24 Growth of *B. thetaiotaomicron* 59
2.2 Biochemistry 59
2.2.1 Affinity gel electrophoresis (AEG) 59
2.2.2 Isothermal titration calorimetry (ITC) 60
2.2.3 Thin layer chromatography (TLC) 61
2.2.4 High-performance anion exchange chromatography (HPAEC) 62
2.2.5 Enzymatic assays 63
2.2.5.1 Unsaturated double bond detection at 235 nm 64
2.2.5.2 L-Rhamnose and D-galacturonic acid dehydrogenase assay 65
2.2.5.3 Substrate depletion method by HPAEC 65
2.2.6 Preparation of substrates and oligosaccharides 66
2.2.6.1 Purification of rhamnogalacturonan-I backbone (AM-RG-I) 66
2.2.6.2 Production of rhamnogalacturonan-II oligosaccharides 67
2.2.7 Purification of oligosaccharides by size exclusion chromatography 67
2.2.8 Concentrating of samples by freeze drying 68
2.2.9 Mass spectrometry (MS) 68
2.2.10 Determination of enzymatic mechanism of catalysis by ^{1}H-NMR 69
2.2.11 Assays in different cell context 69
2.2.12 Cell localization of *B. thetaiotaomicron* proteins 70
2.2.12.1 Western blot 71
2.2.13 Enzyme-linked immunosorbent assay (ELISA) 72
2.2.14 Plant cell wall immunolabelling using monoclonal antibodies 73
2.2.14.1 Quantification of enzymatic activity against plant cell wall 74
2.3 Crystallography 75
2.3.1 Preparation of selenomethionine derivatives 75
2.3.2 Protein crystallisation 76
2.3.3 Visualization of structures 77
2.4 Bioinformatics tools 77
Chapter 3. Characterization of a novel carbohydrate binding module 79
3.1 Introduction 79
3.2 Objectives 80
3.3 Results 81
3.3.1 Cloning, expression and purification 81

3.3.2 CBM77$_{PL1/9}$ binds to pectins with low degree of esterification 85
3.3.3 The structure of CBM77$_{PL1/9}$ 88
3.3.3.1 Protein crystallisation and structure 89
3.3.3.2 Probing the location of the ligand binding site in CBM77$_{PL1/9}$ 91
3.3.3.3 Different approaches explored to obtain a CBM77$_{PL1/9}$ ligand complex….. 94
3.3.4 Phylogeny of CBM77 homologues proteins 95
3.3.5 CBM77$_{PL1/9}$ associated polysaccharide lyases activity 101
3.3.6 The role of CBM77$_{PL1/9}$ in plant cell wall degradation 106
3.3.7 Exploring CBM77$_{PL1/9}$ as a probe to label pectins in plant cell wall 108
3.4 Discussion 112
3.5 Future work 117
Chapter 4. Exploring the enzymatic degradation of rhamnogalacturonan-II by *Bacteroides thetaiotaomicron* 119
4.1 Introduction 119
4.2 Objectives 122
4.3 Results 123
4.3.1 Protein expression and purification 123
4.3.2 Define the BT0986 and BT1019 linkage specificity 125
4.3.3 Production and purification of oligosaccharides generated by mutants of *B. thetaiotaomicron* 127
4.3.4 RG-II degradation by Chain B active enzymes 129
4.3.5 Refinement of RG-II structure: L-Rhamnose-D-Apiose linkage 133
4.3.6 The influence of the RG-II backbone on cleavage of L-Rhamnose-α1,3'-D-Apiose by BT1001 136
4.3.7 Specificity of enzymes acting on RG-II 139
4.3.7.1 α-L-rhamnosidases: BT1001, BT0986, BT1019 and BT1013 140
4.3.7.2 β-L-arabinofuranosidases: BT0996 and BT1020 142
4.3.7.3 α-L-arabinopyranosidase: BT0983 145
4.3.7.4 β-D-galactosidase: BT0993 146
4.3.8 Structural characterization of BT0986 148
4.3.8.1 Protein crystallisation 148
4.3.8.2 Protein structure 150
4.3.8.3 The BT0986 active site 153
4.3.8.3.1 Site direct mutagenesis of BT0986 active site residues 155

4.3.8.4 Phylogeny of BT0986 157
4.3.9 Structural characterization of the aceric acidase BT1003 161
4.3.9.1 Protein crystallization and structure 161
4.3.9.2 Characterization of BT1003 active site 164
4.3.9.3 Site-direct mutagenesis of BT1003 active site residues 167
4.3.9.4 Phylogeny of BT1003 168
4.4 Discussion 173
4.5 Future work 183
Chapter 5. Rhamnogalacturonan-I utilization by *Bacteroides thetaiotaomicron* 185
5.1 Introduction 185
5.2 Objectives 188
5.3 Results 189
5.3.1 Cloning, expression and purification 189
5.3.2 Biochemical characterization of RG-I locus polysaccharide lyases: PL9s (BT4170, BT4183) and PL11 (BT4175) 191
5.3.3 α-L-rhamnosidase BT4145 from RG-I utilization locus 196
5.3.3.1 Biochemical characterization 196
5.3.3.2 Mechanism of catalysis by ^{1}H-NMR 198
5.3.3.3 BT4145 proposed active site 199
5.3.4 The four GH28 enzymes from RG-I locus: BT4146, BT4149, BT4153, BT4155… 206
5.3.4.1 Production and purification of oligosaccharides 206
5.3.4.2 BT4146, BT4153 and BT4155: Substrate specificity 209
5.3.4.3 Catalytic activity of BT4146, BT4153 and BT4155 212
5.3.4.4 BT4149 sequence and protein homologues 213
5.3.4.4.1 BT4149 protein homologue 216
5.3.5 Exploring the SGBP and SusD-like proteins of RG-I PUL 217
5.3.6 Growth curves of B. thetaiotaomicron mutants on different substrates 219
5.3.7 Assays in different cell context 221
5.3.7.1 BT4170 is essential to produce AM-RG-I oligosaccharides at cell surface…. 222
5.3.7.1.1 Mass spectrometry of supernatant protein 225
5.3.8 Cell localization of RG-I PUL encoded proteins 226
5.3.9 The structure of BT4170 228
5.3.9.1 Protein crystallisation 228

5.3.9.2 BT4170 structure 229
5.3.9.3 Complex of BT4170 with ligand bound in the negative subsites 231
5.3.9.4 Complex of BT4170 with ligand bound in the negative and positive subsites… 232
5.3.9.5 Substrate recognition 234
5.3.9.5.1 Substrate binding in the distal subsites 235
5.3.9.5.2 Substrate binding in the active site 236
5.3.9.6 Potential mechanism of BT4170 238
5.3.9.7 Phylogeny of BT4170 239
5.4 Discussion 244
5.5 Future work 255
Chapter 6. Final discussion 257
References 262
Appendix A 281
Appendix B 283
Appendix C 284
Appendix D 288
Appendix E 291
Appendix F 292
Appendix G 294
Appendix H 299
Appendix I 300
Appendix J 301
Appendix K 303

List of Figures

Figure 1.1 Schematic representation of plant cell wall composition 2
Figure 1.2 Schematic representation of cellulose and hemicelluloses 3
Figure 1.3 Schematic representation of HG and RG-I 5
Figure 1.4 Schematic representations of RG-II 6
Figure 1.5 Monosaccharides present in RG-II and borate cross-linkage 7
Figure 1.6 Pectin organization 7
Figure 1.7 Schematic representation of GH structures for each clan 10
Figure 1.8 Topology of GH active sites 11
Figure 1.9 Schematic representation of the classical retention mechanism of a β-glycosidase 12
Figure 1.10 Schematic representation of the inversion mechanism of a glycosidase 13
Figure 1.11 Schematic representation of PL fold and strucutres 14
Figure 1.12 Schematic representations of PL general mechanisms *syn* and *anti* β-elimination 16
Figure 1.13 Catalytic mechanism of a carbohydrate esterase 17
Figure 1.14 Schematic representation of the enzyme subsites 18
Figure 1.15 Schematic representation of CBM folds 19
Figure 1.16 Carbohydrate binding modules Type A, B and C 21
Figure 1.17 Structure of a GH28 family enzyme 25
Figure 1.18 Schematic representation of GH105 structure and proposed mechanism 26
Figure 1.19 Human microbiota 29
Figure 1.20 Schematic representation of *B. thetaiotaomicron* starch utilization system 31
Figure 1.21 *Ruminococcus flavefacie*ns FD-1 cellulosomal system 36
Figure 2.1 Sewing PCR to generate the gene deletion fragment 55
Figure 2.2 Sewing PCR to generate the FLAG-tag fragment 56
Figure 2.3 Conjugation in *B. thetaiotaomicron* 58
Figure 2.4 HPAEC programs and buffers 62
Figure 2.5 Protein crystallization by vapour diffusion method 77
Figure 3.1 Glycan microarray binding profiles of six new CBM families 81
Figure 3.2 Molecular architecture of different constructs used in this study 82

Figure 3.3 Example of PCR-derived fragments of different constructs analysed by agarose gel electrophoresis 83
Figure 3.4 Examples of SDS-PAGE gels of IMAC and gel filtration purifications 85
Figure 3.5 CBM77$_{PL1/9}$ affinity gel electrophoresis 86
Figure 3.6 Examples of the ITC titration curves of CBM77$_{PL1/9}$ against different pectins 88
Figure 3.7 CBM77$_{PL1/9}$ crystal structure 90
Figure 3.8 Agarose gel electrophoresis of CBM77$_{PL1/9}$ site-direct mutagenesis PCR products 92
Figure 3.9 Cartoon showing the location of the CBM77$_{PL1/9}$ mutations and affinity gel electrophoresis 93
Figure 3.10 Alignment of proteins in CBM77 family 99
Figure 3.11 Phylogenetic tree of proteins in CBM77 family 99
Figure 3.12 Molecular architecture of proteins containing a CBM77 101
Figure 3.13 pH profile of RfPel1, RfPel9 and RfPel1/9 102
Figure 3.14 PLs products on pectins with different degree of esterification 103
Figure 3.15 Graphical representation of kinetics data for PLs with and without CBM77$_{PL1/9}$ 105
Figure 3.16 Activity of RfPel1/9 and RfPel1/9 +CBM77$_{PL1/9}$ against plant cell wall . 107
Figure 3.17 CBM77$_{PL1/9}$_CBM77$_{PL1/9}$ dimer binding thermodynamic parameters ... 109
Figure 3.18 ELISA of CBM77$_{PL1/9}$, CBM77 $_{PL1/9}$ dimer and LM19 binding to pectins 110
Figure 3.19 Pectin labelling in *Tobacco* and *Miscanthus* sections 111
Figure 4.1 Schematic representation of RG-II 119
Figure 4.2 *B. thetaiotaomicron* RG-II utilization PULs 121
Figure 4.3 Schematic representation of RG-II degradation model 122
Figure 4.4 Examples of SDS-PAGE gels of IMAC 124
Figure 4.5 Quantification of L-Rha and L-Ara released by BT0996, BT1019 and BT0986 by HPAEC 127
Figure 4.6 Oligosaccharides generated by mutants of *B. thetaiotaomicron* 128
Figure 4.7 Mechanism by which *B. thetaiotaomicron* depolymerizes Chain B of RG-II 130
Figure 4.8 Activity of Chain B enzymes against RG-II 132
Figure 4.9 BT1001 activity against different linkages of L-Rhamnose-D-Apiose 135
Figure 4.10 Bt1001 activity against *Δbt1017* oligosaccharide 137

Figure 4.11 Proposed exo-active model of degradation of *Δbt1017* oligosaccharide 139
Figure 4.12 Examples of representative α-L-rhamnosidases activity detection by HPAEC and TLC 142
Figure 4.13 Example of β-arabinofuranosidase activity detection by HPAEC 144
Figure 4.14 Example of α-L-arabinopyranosidase activity detection by HPAEC 146
Figure 4.15 Example of β-D-galactosidase activity detection by HPAEC 148
Figure 4.16 BT0986 crystals.....149
Figure 4.17 BT0986 structure.....151
Figure 4.18 BT0986 overlay to structural homologues 152
Figure 4.19 BT0986 active site.....154
Figure 4.20 Alignment of proteins in GH106 family 160
Figure 4.21 BT1003 crystal structure 163
Figure 4.22 BT1003 overlay to structural homologues 164
Figure 4.23 BT1003 overlay with GH127 structures.....166
Figure 4.24 Activity of BT1003 mutants.....168
Figure 4.25 Alignment of proteins in GH127 family 172
Figure 4.26 GH78 structures 175
Figure 4.27 Proposed model of RG-II degradation by *B. thetaiotaomicron* 180
Figure 4.28 BT0996 structure in complex with Chain B.....183
Figure 5.1 Schematic representation of the structures of homogalacturonan and rhamnogalacturonan-I 186
Figure 5.2 Rhamnogalacturonan-I polysaccharide utilization loci (PUL) 187
Figure 5.3 Examples of SDS-PAGE after proteins purification.....190
Figure 5.4 pH profile of BT4170, BT4175 and BT4183 192
Figure 5.5 Graphical representation of kinetic data for the PLs against different substrates.....193
Figure 5.6 PLs products on different substrates 195
Figure 5.7 TLC showing the BT4170, BT4175 and BT4183 endo-activity.....195
Figure 5.8 Catalytic activity of BT4145 197
Figure 5.9 BT4145 ion dependence 198
Figure 5.10 Partial ^{1}H-NMR of α-L-Rha-α1,4-D-GalA treated with BT4175 199
Figure 5.11 BT4145 and BT0986 alignment.....204
Figure 5.12 GH106 protein structure 205
Figure 5.13 Production of D-Gal-α1,2-L-Rha oligosaccharides.....207

Figure 5.14 Purification and LC-MS analysis of oligosaccharides 208
Figure 5.15 BT4146, BT4153 and BT4155 activity against α1,4-D-GalA-α1,2-L-Rha oligosaccharides 210
Figure 5.16 BT4146, BT4153 and BT4155 activity against D-GalA-α1,4-D-GalA oligosaccharides 211
Figure 5.17 Activity of BT4146 and BT4155 against polysaccharides 212
Figure 5.18 Comparison between the BT4149 sequences 214
Figure 5.19 Alignment of BT4149 215
Figure 5.20 SGBP and SusD affinity gel electrophoresis 218
Figure 5.21 BT4167 binding thermodynamics parameters 219
Figure 5.22 Growth curves of *B. thetaiotaomicron* KO mutants on different substrates 221
Figure 5.23 Assays in different cell context against AM-RG-I 224
Figure 5.24 BT4170 identification by mass spectrometry 225
Figure 5.25 Detection of FLAG-Tag proteins by Western blot 227
Figure 5.26 Analysis of BT4170 reaction product by TLC and protein crystal 229
Figure 5.27 Schematic representation of BT4170 structure 230
Figure 5.28 BT4170 structure superimpose to Pel9A 231
Figure 5.29 Structure of Bt4170 in complex with L-Rha-α1,4-D-GalA-α1,2-L-Rha .. 232
Figure 5.30 BT4170 structure in complex with two trisaccharides 234
Figure 5.31 BT4170 active site 237
Figure 5.32 PL9 proteins alignment 243
Figure 5.33 BT4170 and Pel9A active site and protein surface 244
Figure 5.34 Comparison between HG and AM-RG-I oligosaccharides 250
Figure 5.35 Proposed model of RG-I utilization by *B. thetaiotaomicron* 254

List of Tables

Table 1.1 Anaerobic bacteria and fungi producing cellulosomes.....35
Table 2.1 Bacterial strains.....39
Table 2.2 Plasmids.....40
Table 2.3 Growth media.....41
Table 2.4 Antibiotics.....42
Table 2.5 Typical PCR.....46
Table 2.6 Typical PCR conditions.....46
Table 2.7 Preparation of SDS-PAGE gel.....53
Table 2.8 SDS-PAGE buffers.....53
Table 2.9 Composition of a typical reaction for a linked assay.....65
Table 2.10 Bioinformatics tools.....78
Table 3.1 Primers designed to clone different constructs utilized in this chapter.....83
Table 3.2 Details of the different constructs cloned into pET28a.....84
Table 3.3 Thermodynamic parameters of the binding of $CBM77_{PL1/9}$ to different ligands.....87
Table 3.4 Structural homologs of $CBM77_{PL1/9}$.....91
Table 3.5 ITC of the binding of $CBM77_{PL1/9}$ mutants against PGA orange.....94
Table 3.6 Catalytic activity of PL proteins with and without $CBM77_{PL1/9}$.....104
Table 4.1 The details of proteins investigated in this chapter.....125
Table 4.2 Identification of *Δbt1017* oligosaccharide reaction products.....138
Table 4.3 BT1001, BT0986, BT1019 and BT1013 α-L-rhamnosidases specificities.....141
Table 4.4 BT0996 and BT1020 β-L-arabinofuranosidase specificity.....144
Table 4.5 BT0983 α-arabinopyranosidase specificity.....145
Table 4.6 BT0993 β-D-galactosidase specificity.....147
Table 4.7 Structural homologs of BT0986.....152
Table 4.8 Catalytic activity of BT0986 mutants.....156
Table 4.9 Structural homologs of BT1003.....163
Table 5.1 *B. thethaiotaomicron* PULs upregulated in response to pectins.....187
Table 5.2 The details of the proteins mention in this chapter.....191
Table 5.3 Activity of BT4170, BT4175 and BT4183.....193
Table 5.4 Catalytic activity of BT4145 mutants.....205
Table 5.5 Catalytic activity of BT4146, BT4153 and BT4155.....213

Table 5.6 Activity of BT4149 homologue 217
Table 5.7 Catalytic activity of BT4170 mutants 238

Abbreviations

Δ4,5-GalA	4,5-unsaturated galacturonic acid
^{1}H-NMR	Proton nuclear magnetic resonance
A_{235}	Absorbance at a wavelength of 235 nm
A_{340}	Absorbance at a wavelength of 340 nm
A_{600}	Absorbance at a wavelength of 600 nm
AceA	Aceric acid
AceAase	Aceric acidase
AGE	Affinity gel electrophoresis
AGP	Arabinogalactan
AM-RG-I	Arabidopsis - Rhamnogalacturonan-I (RG-I backbone)
Api	Apiose
Ara*f*	Arabinofuranose
Ara*f*ase	Arabinofuranosidase
Ara*p*	Arabinopyranose
BLAST	Basic Local Alignment Search Tool
BSA	Bovine serum albumin
CAZy	Carbohydrate-active enzymes database
CBM	Carbohydrate-binding modules
CE	Carbohydrate esterase
DE	Degree of esterification
Dha	3-deoxy-D-lyxo-2-heptulosaric acid
DP	Degree of polymerization
EDTA	Ethylenediaminetetraacetic acid
ELISA	Enzyme-linked immunosorbent assay
FITC	Fluorescein isothiocyanate
FPLC	Fast protein liquid chromatography
Fuc	Fucose
Gal	Galactose
GalA	Galacturonic acid
GH	Glycoside hydrolase
GlcA	Glucuronic acid
HG	Homogalacturonan
HGM	Human gut microbiota
His	Polyhistidine tag
HPAEC	High-performance anion-exchange chromatography
HRP	Horseradish peroxidase
HTCS	hybrid two component system
IMAC	Immobilized metal-ion affinity chromatography
ITC	Isothermal titration calorimetry
Kdo	3-deoxy-D-manno-2-octulosonic acid or Ketodeoxyoctonic acid
LB	Luria-Bertani broth
LC-MS	Liquid chromatography–mass spectrometry

MM	Minimal media
MS	Mass spectrometry
MUSCLE	Multiple Sequence Comparison by Log-Expectation
NCBI	National Center for Biotechnology Information
ORF	Open reading frame
PAD	Pulsed amperometric detection
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PDB	Protein Data Base
PEG	Polyethylene glycol
PGA	Polygalacturonic acid
PL	Polysaccharide lyase
P-RG-I	Potato rhamnogalacturonan-I
PSI-BLAST	Position-Specific Iterated BLAST
PUL	Polysaccharide utilization loci
RG-I	Rhamnogalacturonan-I
RG-II	Rhamnogalacturonan-II
Rha	Rhamnose
RMSD	Root mean square deviation
RT	Room temperature
SAD	Single Anomalous Diffraction
SDS-PAGE	Sodium dodecyl sulfate - polyacrylamide gel electrophoresis
SeMet	Selenomethionine
SGBP	Surface glycan binding proteins
SP	Signal peptide
tdk	Thymidine kinase
TEMED	N,N,N′,N′-Tetramethylethylenediamine
TLC	Thin layer chromatography
TYG	Tryptone yeast extract glucose
UV	Ultraviolet
Xyl	Xylose

Journal articles

One article was published containing work included in this thesis:

Venditto, I.*, **Luis, A.S**.*, Rydahl, M.*, Schückel, J.*, Fernandes, V.O., Vidal-Melgosa, S., Bule, P., Goyal, A., Pires, V.M., Dourado, C.G., Ferreira, L.M., Coutinho, P.M., Henrissat, B., Knox, J.P., Basle, A., Najmudin, S., Gilbert, H.J., Willats, W.G. and Fontes, C.M. 2016. Complexity of the *Ruminococcus flavefaciens* cellulosome reflects an expansion in glycan recognition. *Proc Natl Acad Sci U S A* 113 (26):7136-7141 (*author with equal contribution)

Additional publications containing work conducted during the course of this investigation are still in preparation.

Other publications containing work conducted by the author, which are not included in this thesis:

Venditto, I., Najmudin, S., **Luís, A.S**., Ferreira, L.M., Sakka, K., Knox, J.P., Gilbert, H.J. and Fontes, C.M. 2015. Family 46 carbohydrate-binding modules contribute to the enzymatic hydrolysis of xyloglucan and beta-1,3-1,4-glucans through distinct mechanisms. *J Biol Chem* 290 (17):10572-10586.

Chapter 1. General introduction

1.1 Plant cell wall

The walls that surround plant cells confer the necessary rigidity and strength in order to withstand the osmotic turgor pressure. These walls are also involved in plant signalling that influence growth, cell shape modulation and differentiation (Lodish *et al*. 2000). The plant cell wall is composed of glycans and is the most abundant source of terrestrial biomass.

The major plant cell wall glycans are cellulose, hemicelluloses and pectins (Figure 1.1). Lignin, which is present in secondary walls, contributes to the rigidity of these composite structures. The organization and composition of these polysaccharides varies with the plant type (monocotyledon or dicotyledon), the tissue and the developmental stage (Naran *et al*. 2008; Taiz and Zeiger 2010). Plant cell wall polysaccharides represent one of the major nutrients available to the microbial community in the human gut, which has a significant impact on host nutrition and health (reviewed by Lattimer and Haub 2010). These glycans also have an industrial significance, particularly in bioenergy and bioprocessing sectors, where they provide a renewable and thus sustainable substrates (reviewed by Ragauskas *et al*. 2006).

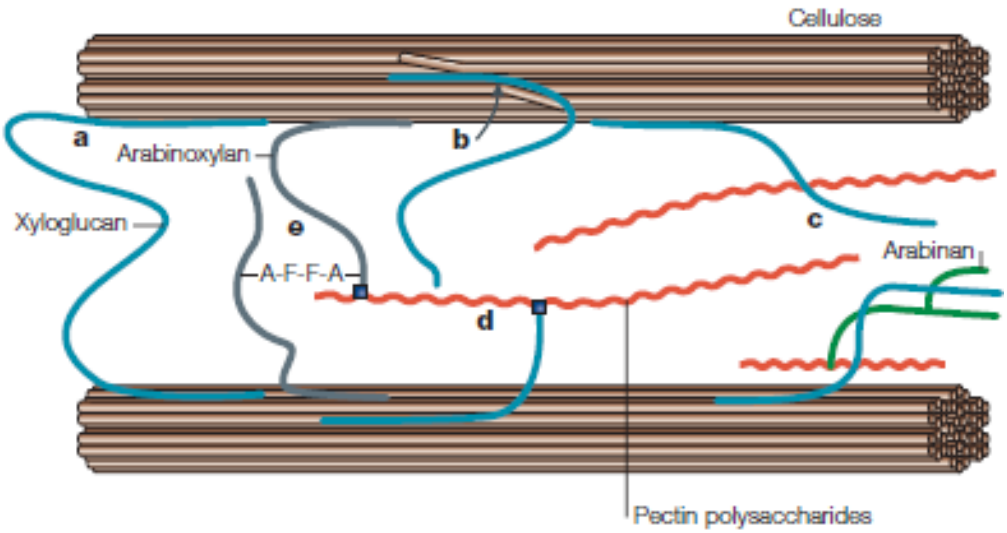


Figure 1.1 Schematic representation of plant cell wall composition

Cellulose is organized in microfibrils that are connected by branched polysaccharides with a β-1,4 backbone, referred to as the hemicelluloses. This network is embedded in a complex matrix of proteins and pectins that can be decorated with highly complex branched glycans, such as arabinan. Figure adapted from (Cosgrove 2005).

1.1.1 *Cellulose and hemicelluloses*

Cellulose, a β1,4-D-glucose (D-Glc) linear polysaccharide (Figure 1.2), is the most abundant source of organic carbon on earth. In the plant cell wall, these linear chains are linked together in microfibrils stabilized by several intra and inter chain hydrogen bonds. These microfibrils are responsible for the cell wall strength and are linked to hemicelluloses by hydrogen bonds (reviewed by Cosgrove 2005).

The backbone of plant hemicelluloses contains primarily β1,4 linked sugars, although β(1,3)(1,4)- and β1,3-glucans are also components of this group of polysaccharides. The sugars in the backbone of hemicelluloses are D-Glc, D-mannose or D-xylose (D-Xyl). Figure 1.2 shows a schematic representation of some of the hemicellulosic polysaccharides.

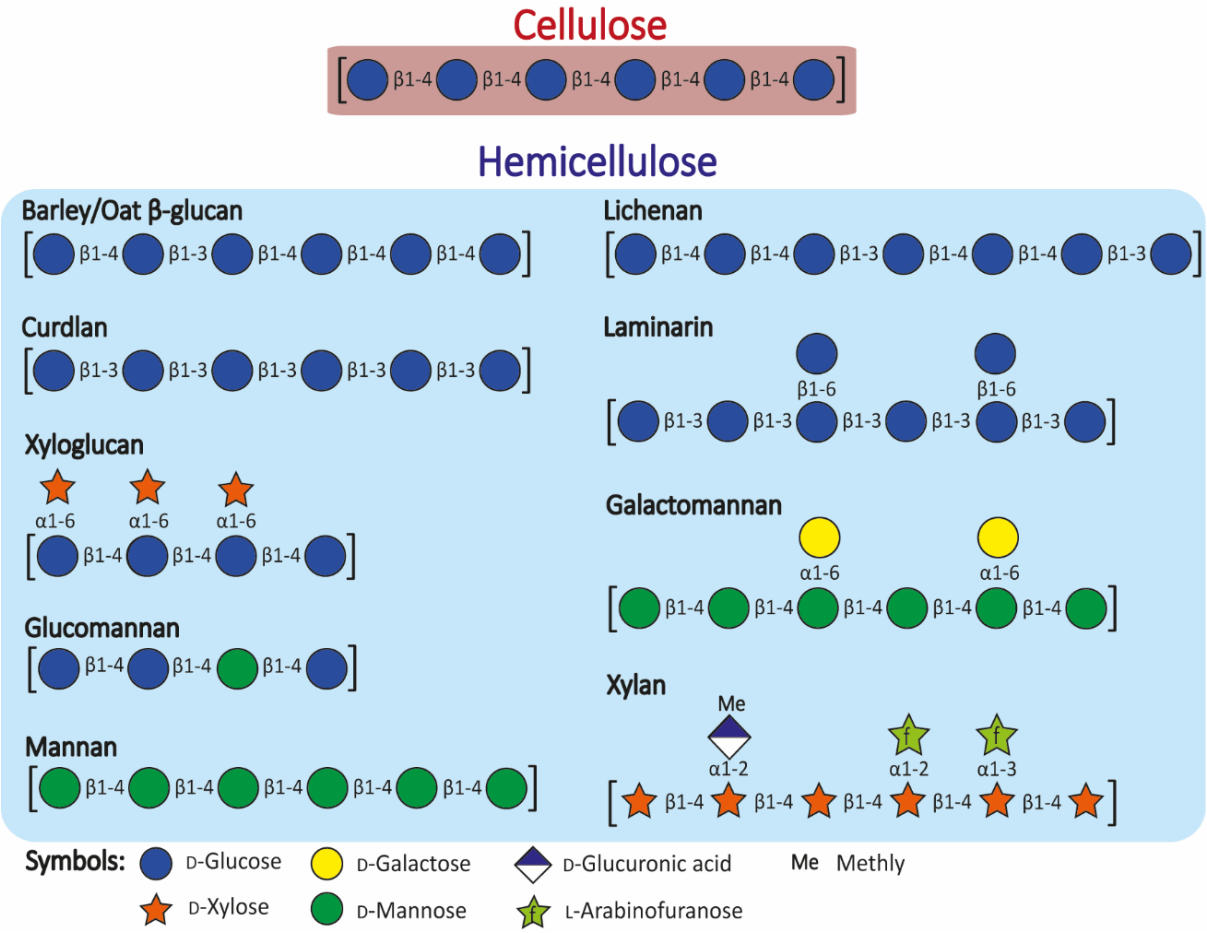


Figure 1.2 Schematic representation of cellulose and hemicelluloses

Information about the different linkages and side chains composition is represented in the figure.

1.1.2 *Pectins*

Pectins are the major constituent of the primary plant cell wall. These glycans form a network that can influence features such as, porosity, rigidity and plant development (Oomen *et al*. 2002; Ryden *et al*. 2003; Barbacci *et al*. 2013). Pectins are found in a high percentage in some plants exemplified by citrus fruit (oranges pulp 12,4-28 %), apples (0,5–1.6%), carrots (6,9-18,6%) and tomatoes (2.4-4.6 %) (reviewed by Jayani *et al*. 2005). These polysaccharides also account for a significant percentage of the human diet but have a minimal direct impact in human nutrition as they are not degraded in the small intestine. The intake of pectins, however, has been associated with a positive impact on human health, such as reducing blood cholesterol levels,

inducing apoptosis of cancer cells and stimulating the immune system (Inngjerdingen *et al*. 2007; Jackson *et al*. 2007; Brouns *et al*. 2012; Maxwell *et al*. 2016). Additionally, these complex polysaccharides also have important biotechnological applications in the pharmaceutical industry, and stabilizing and gelling polymers in the food and cosmetic industries (Lin *et al*. 2015; Delphi and Sepehri 2016).

Pectins are negatively charged carbohydrates with a huge variability of sugars presenting a complex structure. However, these structural polysaccharides share a common feature: the rich content of the acidic sugar D-galacturonic acid (D-GalA) in the backbone. D-GalA is responsible for the negative charge of pectins (reviewed by Willats *et al*. 2001). The major pectins are homogalacturonan, rhamnogalacturonan-I and rhamnogalacturonan-II.

1.1.2.1 Homogalacturonan

Homogalacturonan (HG) (Figure 1.3) is a linear polymer of ~100 α1,4-D-GalA residues that can be partially methylated at C6 (70-80 %) and *O*-acetylated at O2 and O3. The methylation of the carboxyl group leads to a negative charge reduction, preventing the interaction of Ca^{2+} with pectin and, as a consequence, a reduction of the pectin gelling properties (reviewed by Mohnen 2008).

1.1.2.2 Rhamnogalacturonan-I

The rhamnogalacturonan-I (RG-I) (Figure 1.3) backbone contains alternating α1,2-L-Rha and α1,4-D-GalA residues, where the O2 and O3 of galacturonic residues can be acetylated. The O4 in rhamnose residues can be extensively decorated (~ 25-80 %) with neutral sugars (galactose in galactan and arabinose in arabinan). The size and

sugar composition of these side chains is not a conserved feature but dependent on the cell type and the stage of development (reviewed by Willats *et al*. 2001; Atmodjo *et al*. 2013).

Arabinans are characterized by a α1,5-L-Arabinofuranose (L-Ara*f*) backbone with side chains of α1,2-L-Ara*f*, α1,3-L-Ara*f* or double α-L-Ara*f* substitutions at O2 and O3. Pectic galactan consist primarily of chains of β1,4-D-Galactose (D-Gal) with a degree of polymerization (DP) of 43 to 47 (reviewed by Mohnen 2008). Arabinogalactan pectin type I (AGP-I) is the most common arabinogalactan linked to RG-I. AGP-I present a β1,4-D-Gal backbone decorated with short chains of α1,3-L-Ara*f* (reviewed by Caffall and Mohnen 2009). The RG-I side chains can also be decorated with sugars other than L-Ara*f* and D-Gal*p*, such as, β-D-glucuronic acid (β-D-GlcA), 4-O-methly-β-D-GlcA and α-L-fucose (L-Fuc) (reviewed by Atmodjo *et al*. 2013).

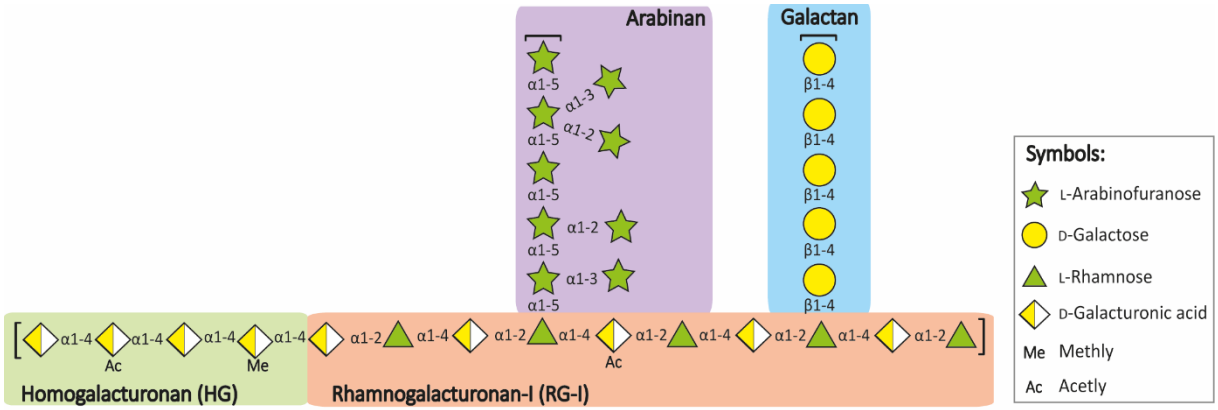


Figure 1.3 Schematic representation of HG and RG-I

Information about the different linkages and side chains composition is represented in the figure.

1.1.2.3 Rhamnogalacturonan-II

Rhamnogalacturonan-II (RG-II) (Figure 1.4) is a less common but highly complex polysaccharide. RG-II presents a backbone of seven to nine α1,4-D-GalA units that is substituted with six side chains (A to F) containing unusual sugars such as 2-O-methyl-L-fucose, 2-O-methyl-D-xylose, D-apiose (D-Api), L-galactose, 3-C-carboxy-5-deoxy-L-

xylose (L-aceric acid), 3-deoxy-D-manno-2-octulosonic acid (D-Kdo), and 3-deoxy-D-lyxo-2-heptulosaric acid (D-Dha) (York *et al*. 1985; Stevenson *et al*. 1988; Vidal *et al*. 2000; Glushka *et al*. 2003; Pabst *et al*. 2013). Despite the complexity of this highly branched polysaccharide, comprising 13 different sugars (Figure 1.5) and 21 different linkages, this glycan is highly conserved among plants (reviewed by O'Neill *et al*. 2004; Bar-Peled *et al*. 2012). RG-II is a dimer in primary walls, which is mediated by a borate diester linkages between the D-Api residues in Chain A (Figure 1.5) (Ishii and Ono 1999). This borate cross-linkage creates a glycan network important in the porosity of the cell wall and is required for normal plant growth (Fleischer *et al*. 1999; O'Neill *et al*. 2001).

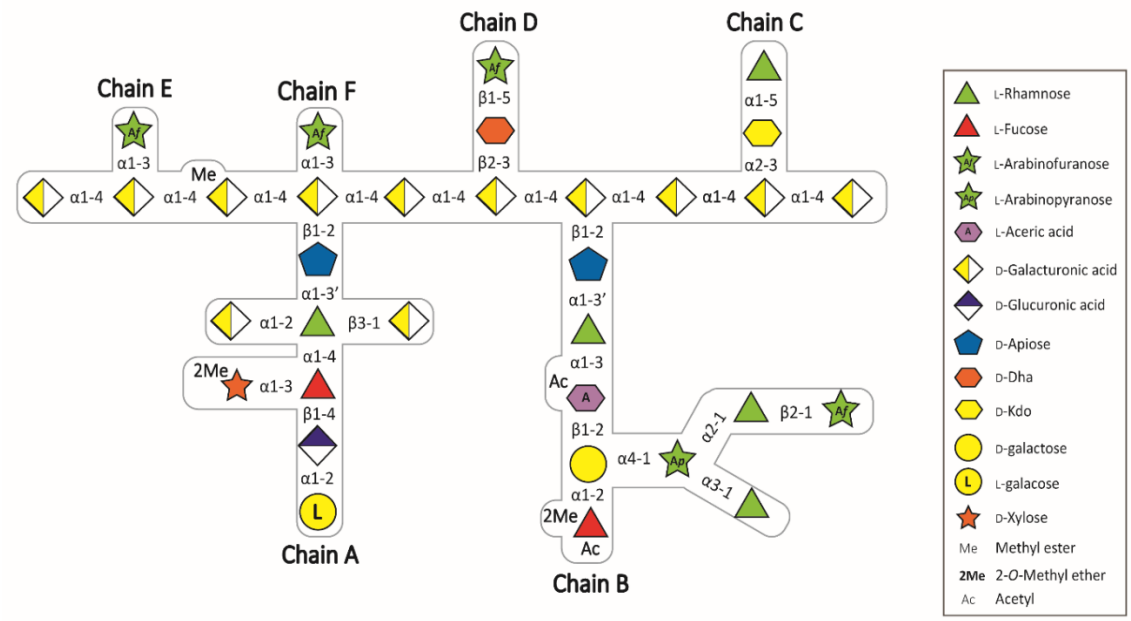


Figure 1.4 Schematic representations of RG-II

Information about the different linkages and side chains composition is represented in the figure. Figure modified with the courtesy of Dr Artur Rogowski (ICaMB, Newcastle University).

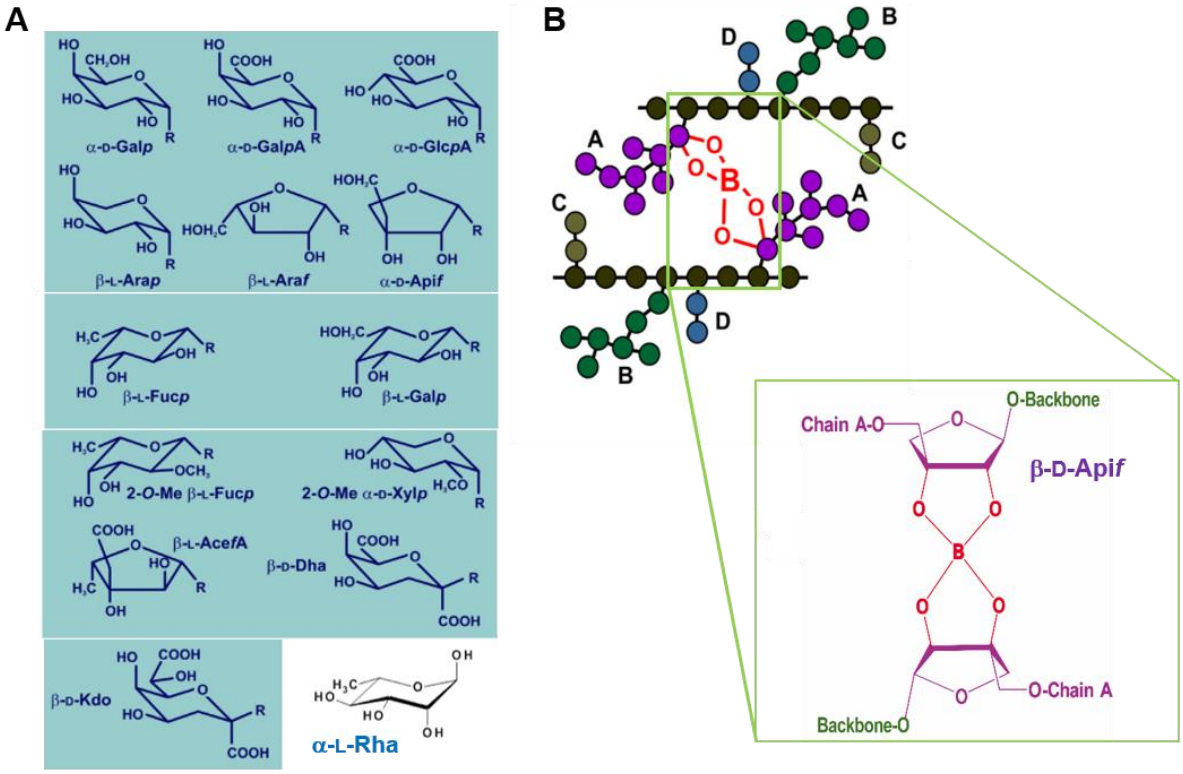


Figure 1.5 Monosaccharides present in RG-II and borate cross-linkage

A. Different monosaccharides present in RG-II. R can be substituted with a hydroxyl or involved in the glyosidic bond. **B.** RG-II cross-linkage mediated by borate between to D-Api*f*. D-galactose (D-Gal), D-galacturonic acid (D-GalA), D-glucuronic acid (D-GlcA), L-arabinopyranose (L-Ara*p*), L-arabinofuranose (L-Ara*f*), D-apiose (D-Api), L-fucose (L-Fuc), L-galactose (L-Gal), D-xylose (D-Xyl), L-aceric acid (L-Ace), L-Rha (L-Rha). "*f*" represents a furanose and "*p*" a pyranose. Figure adapted from (Bar-Peled *et al*. 2012)

1.1.2.4 Pectin organization in plant cell wall

In the plant cell wall, pectins are linked via their backbones forming an interconnected pectin structure (Figure 1.6). The DP of each pectin, however, is unknown, and the specific order of the different polysaccharides within the pectin network is opaque (reviewed by Atmodjo *et al*. 2013).

Figure 1.6 Pectin organization

(RG-I) rhamnogalacturonan-I; (HG) homogalacturonan; (RG-II) rhamnogalacturonan-II; (AGP) arabinogalactan protein; (AG) arabinogalactan.

1.2 Plant cell wall degradation

The plant cell wall represents an extremely complex biological structure. In nature, different microorganisms (bacteria and fungi) have evolved to degrade this recalcitrant substrate. The degradation of the array of polysaccharides that comprise the plant cell wall is possible through the combination of enzymes with different specificities, typically glycoside hydrolases (GHs), polysaccharide lyases (PLs), carbohydrate esterases (CEs) and, more recently, lytic polysaccharide monooxygenases or LPMOs (reviewed by Gilbert 2010; Johansen 2016). These enzymes can be associated with non-catalytic carbohydrate-binding modules (CBMs) that facilitate enzymatic degradation by promoting intimate contact between substrate and enzyme (Herve *et al*. 2010; Venditto *et al*. 2015).

1.2.1 *Enzymes*

Carbohydrate active enzymes or CAZymes are classified into sequence-based families on the continuously updated CAZy database (Cantarel *et al*. 2009; Lombard *et al*. 2014). At least one member of each family has been characterized.

1.2.1.1 Glycoside hydrolases

Glycoside hydrolases (GHs) catalyse the hydrolysis of diverse glycoside bonds (between carbohydrates, or carbohydrate and a different moiety).

1.2.1.1.1 Classification

In 1991 Henrissat proposed the classification of GHs into families based on sequence similarity (Henrissat 1991; Henrissat and Bairoch 1993, 1996). There are currently (October, 2016) 135 GH families listed in the CAZy database (Cantarel *et al*. 2009; Lombard *et al*. 2014). The members of a GH family display the same fold and, usually, the catalytic apparatus and mechanism is similarly conserved (examples of exceptions are GH8 and GH97 that contain enzymes with two different mechanisms of action). This classification does not take in account the substrate specificity within the enzymes of the same family. In some families, such as GH10, GH11 and GH78, substrate specificity is conserved. In other GH families, such as GH2 and GH43, the target substrate can be highly variable, suggesting that in these families substrate specificity is driven by active site topology (Cantarel *et al*. 2009; Lombard *et al*. 2014). Some families, such as GH13, are divided into subfamilies that are characterized by sharing the same specificity and being phylogenetically related (Stam *et al*. 2006)

The structurally characterized GH families display one of seven folds: $(\beta/\alpha)_8$, β-jelly roll, six-bladed β-propeller, five-bladed β-propeller, $(\alpha/\alpha)_6$, $\alpha+\beta$ and β-helix (reviewed by Henrissat and Davies 1997). The GH families displaying the same fold, mechanism and catalytic residues are grouped into clans. However, the families in the same clan show low sequence similarity. There are currently 14 clans, GH-A to GH-N (Figure 1.7). The largest clan, GH-A, includes 19 families and display a $(\beta/\alpha)_8$ TIM-barrel fold. Enzymes in this family display a “retaining” mechanism (see below) and the catalytic apparatus comprises two glutamates located in the C-terminus of β-stands 4 and 7 (Henrissat *et al*. 1995).

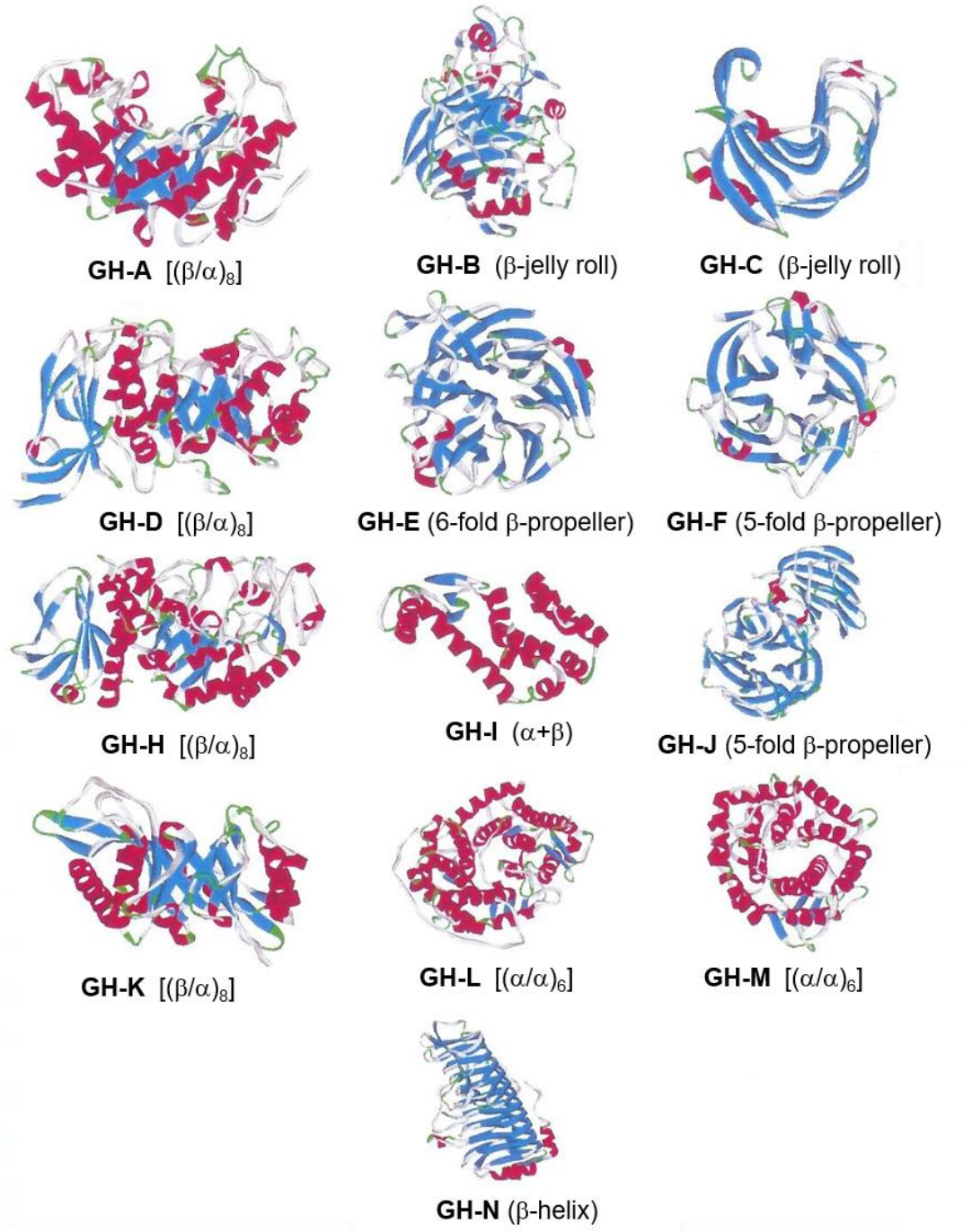


Figure 1.7 Schematic representation of GH structures for each clan

Secondary structural elements are coloured in red (α-helices) and blue (β-sheet). The representative enzymes displayed are as follows for the different Clans: GH-A *Cellvibrio japonicus* xylanase Xyn10C (GH10); GH-B, *Hypocrea jecorina* cellobiohydrolase I (GH7); GH-C, *Aspergillus niger* XynA (GH11); GH-D, *H. jecorina* RuC-30 α-galactosidase 1 (GH27); GH-E, *Salmonella typhimurium* TA262 sialidase (GH33); GH-F, *C. japonicus* arabinanase Arb43A (GH43A); GH-H, *A. niger* α-amylase (GH13); GH-I, Bacteriophage T4 lysozyme (GH24); GH-J, *Thermotoga maritima* invertase (GH32); GH-K, *Flavobacterium meningospticum endo*-β-N-acetylglucosaminidase F1 (GH18); GH-L, *A. awamori* var. X-100 glucoamylase (GH15); GH-M, *Clostridium thermocellum endo*-β-1,4-glucanase; GH-N, *A. aculeatus* rhamnogalacturonase A (GH28). Figure adapted from (Pell 2004).

The typology of GHs is related to their endo/exo mode of action (Figure 1.8). Endo-acting enzymes (cleave internal linkages within the substrate) have an open cleft whereas exo-acting hydrolases (cleave at end of the chain) present an active site pocket that only accommodates the terminal sugars. The tunnel topography, typical of cellobiohydrolases, is consistent with the proposed processive exo-mode of action of these enzymes (Figure 1.8C) (reviewed by Davies and Henrissat 1995; McKee *et al*. 2012).

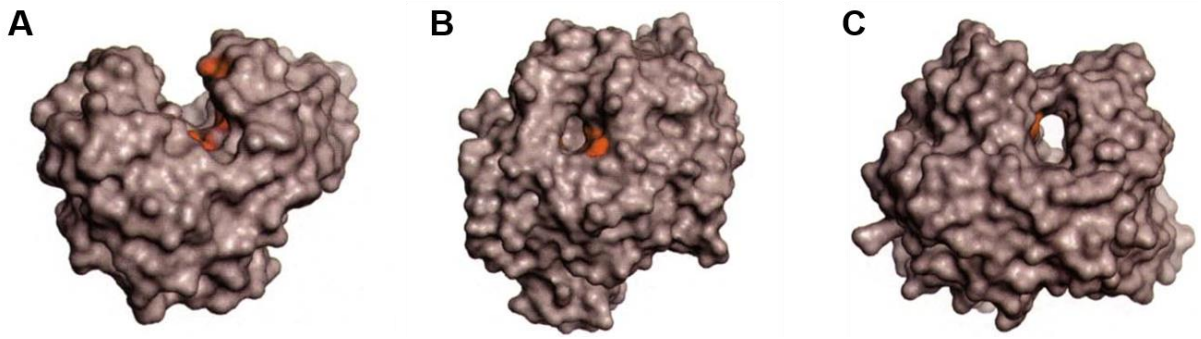


Figure 1.8 Topology of GH active sites

Panel A shows the open cleft topography. Panel B presents the pocket topography. Panel C displays the tunnel to topography. Figure adapted from (reviewed by Davies and Henrissat 1995).

1.2.1.1.2 Main catalytic mechanisms

GHs cleave glycosidic bonds through two main catalytic mechanisms classified according to the stereochemistry of the anomeric carbon after hydrolysis. In the retention mechanism the conformation of the anomeric carbon will be maintained (α to α or β to β) while in the inverting mechanism the configuration of this chiral centre is inverted (α to β or β to α) (reviewed by Withers 2001).

The retention mechanism (Figure 1.9) is achieved by a double-displacement mechanism involving a covalent glycosyl-enzyme intermediate with an opposite anomeric conformation to the substrate. The two amino acids function as the acid/base and nucleophile (usually a glutamate or aspartate). Briefly, in the glycosylation step the

nucleophile (carboxylic acid group) attacks the anomeric centre leading to cleavage of the glycosidic bond in concert with protonation of the glycosidic oxygen by the acid-base residue to promote leaving group departure. The resultant glycosyl-enzyme intermediate is hydrolysed (deglycosylation step) by the catalytic acid/base amino acid, which deprotonates a water molecule and the resultant hydroxyl ion attacks the anomeric centre releasing the reaction product with the initial conformation of the anomeric carbon retained (reviewed by McCarter and Withers 1994; White and Rose 1997).

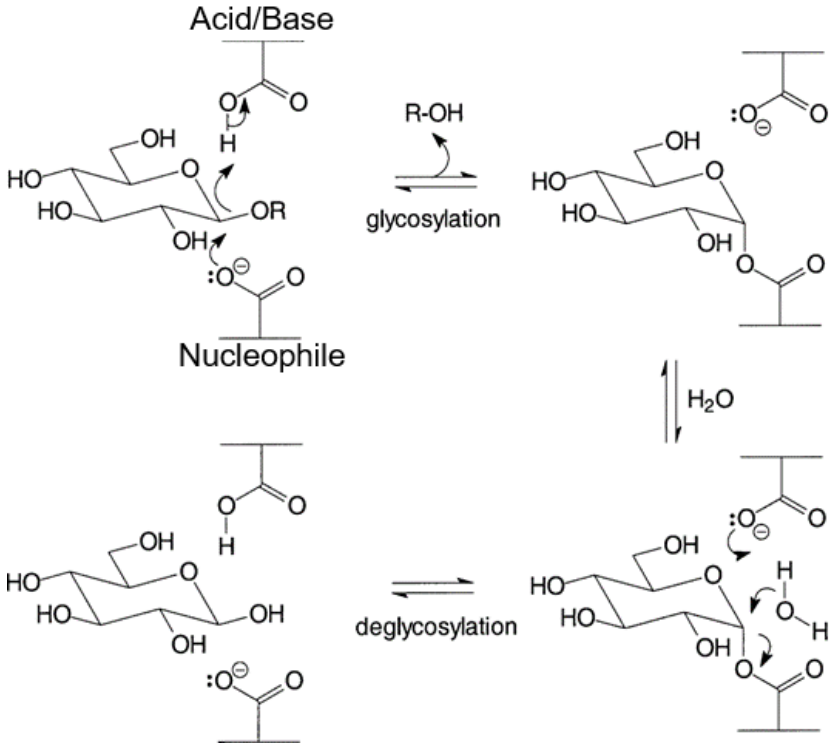


Figure 1.9 Schematic representation of the classical retention mechanism of a β-glycosidase
The hydrolysis by a retention mechanism of the anomeric centre involved a double-displacement reaction, with a glycosyl-enzyme intermediate. Glycoside hydrolases that hydrolyse glycoside bonds by retention mechanism present the two carboxylic residues separated by ~5 Å. Figure adapted from (Paal *et al*. 2004)

The hydrolysis by an inversion mechanism of the anomeric centre involves a one-step reaction in which the transition state, in common with the retaining mechanism, is an oxocarbenium ion (Figure 1.10). The reaction is driven by two amino acids (glutamate

or aspartate) whose carboxyl groups are separated by ~10 Å (to allow the placement of a water molecule). The catalytic base deprotonates the water, resulting in the generation of a hydroxyl ion that attacks the anomeric carbon leading to glycosidic bond cleavage, again encouraged by glycosidic bond protonation by the catalytic acid, which facilitates leaving group departure (reviewed by Collins *et al*. 2005). Different alternative mechanisms occur in a minority of GHs but will not be discussed further here.

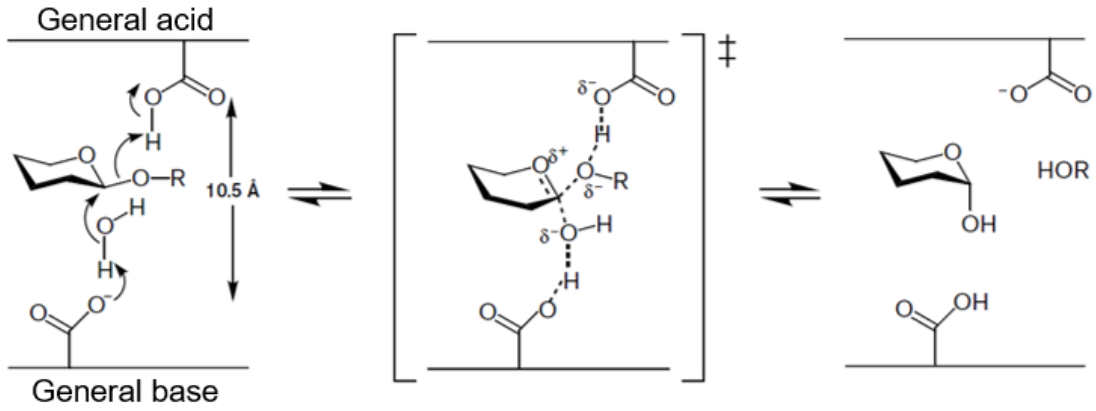


Figure 1.10 Schematic representation of the inversion mechanism of a glycosidase
The hydrolysis by an inversion mechanism of the anomeric centre involves a one-step reaction, with an oxocarbenium ion-like transition state resulting in a product presenting an inversion of C1 conformation. Figure adapted from (Rye and Withers 2000).

1.2.1.2 Polysaccharide lyases

Polysaccharide lyases (PLs) are non-hydrolytic enzymes that cleave sugar polymers that contain uronic acids such as pectins (polygalacturonates), alginates (mannuronate), heparin and chondroitin (glucuronic/iduronic acid), via a β-elimination mechanism yielding an unsaturated uronic acid at the non-reducing end (reviewed by Garron and Cygler 2010). For example, the cleavage of an unmethylated α1,4-D-GalA polysaccharide by pectate lyases generates 4-deoxy-α-D-galact-4-enuronosyl acid (Δ4,5-unsaturated galacturonic acid) (Soriano *et al*. 2006).

These enzymes are classified into PL families based on amino acid sequence similarities, reflecting structural similarities, such as the conservation of catalytic residues and the protein fold. Currently (October, 2016), there are 24 PL families described in the CAZy database (Cantarel *et al*. 2009; Lombard *et al*. 2014). Additional classification of PLs into subfamilies allows substrate specificity extrapolation, although some of these subfamilies are polyspecific, such PL1_5, PL9_1 and PL14_3 (reviewed by Lombard *et al*. 2010). Similar to GHs, PL families present a diversity of protein folds: parallel β-helix, β-propeller, $(\alpha/\alpha)_6$ barrel, $(\alpha/\alpha)_7$ barrel, $(\alpha/\alpha)_3$ barrel, β-sandwich+ β-sheet, β-jelly roll, $(\alpha/\alpha)_6$ barrel+antiparallel β-sheet, and triple strand β-helix (Figure 1.11) (reviewed by Garron and Cygler 2010; Lombard *et al*. 2010). However, families with different folds can present a similar active site implying convergent evolution between at least some PL families (Charnock *et al*. 2002).

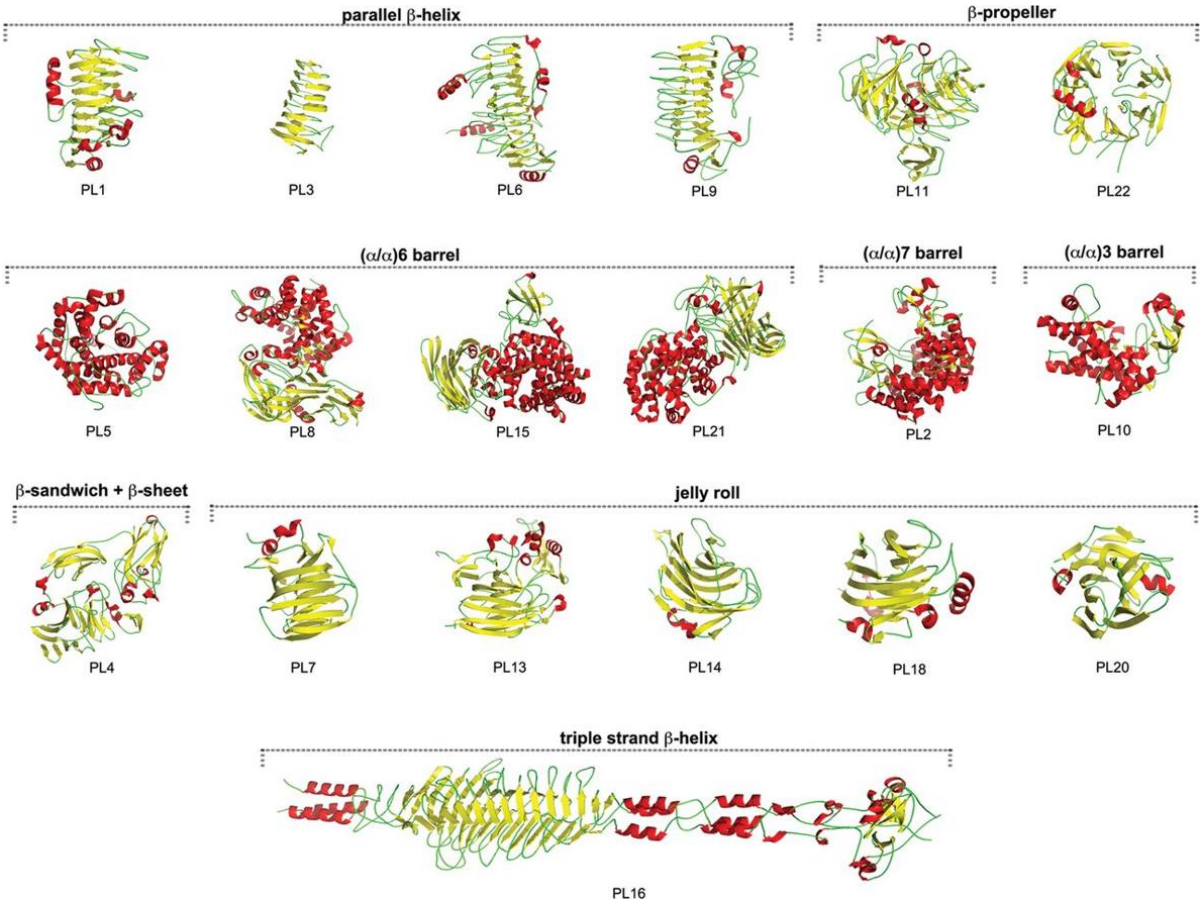


Figure 1.11 Schematic representation of PL fold and strucutres

The families are organized by protein fold. The α-helix and β-strands are shown in red and yellow, respectively. Figure adapted from (Lombard *et al*. 2010)

PLs catalyse the non-hydrolytic cleavage of uronic acid bonds. Like GHs, these enzymes can be endo- or exo-acting and the substrate target, combined with the catalytic site organization, determines the *syn*- or *anti*-β-elimination (Figure 1.12) (reviewed by Garron and Cygler 2010). The catalytic mechanism can be "metal assisted" (galacturonan lyases) or thyrosine/histidine (alginate and glycosaminoglycan lyases) (reviewed by Garron and Cygler 2014). In galacturonan lyases the catalytic mechanism is initiated with C5 proton abstraction by a basic amino acid (Figure 1.12B), usually an arginine (PL1 and PL2) or a lysine (PL3 or PL9) (reviewed by Garron and Cygler 2014). In this step the H5 acidification occurs due to the presence of divalent metals or a positive amino acid (lysine) present in the active site. This positive amino acid or metal interacts with the C5 carboxyl group drawing the charge of this centre and increasing H5 acidity, facilitating the deprotonation reaction (Seyedarabi *et al*. 2010). This charge delocalization is also important in intermediate product stabilization. The shift of the double bond from C5-C6 to C4-C5 leads to the elimination of the glycan sugar linked to C4. The leaving group (glycosidic oxygen) is protonated by the catalytic acid (reviewed by Garron and Cygler 2010). Due to the molecular restrictions of the double bond the pyranosyl ring C2, C3, C4 and C5 are in the same plane (Abbott *et al*. 2010).

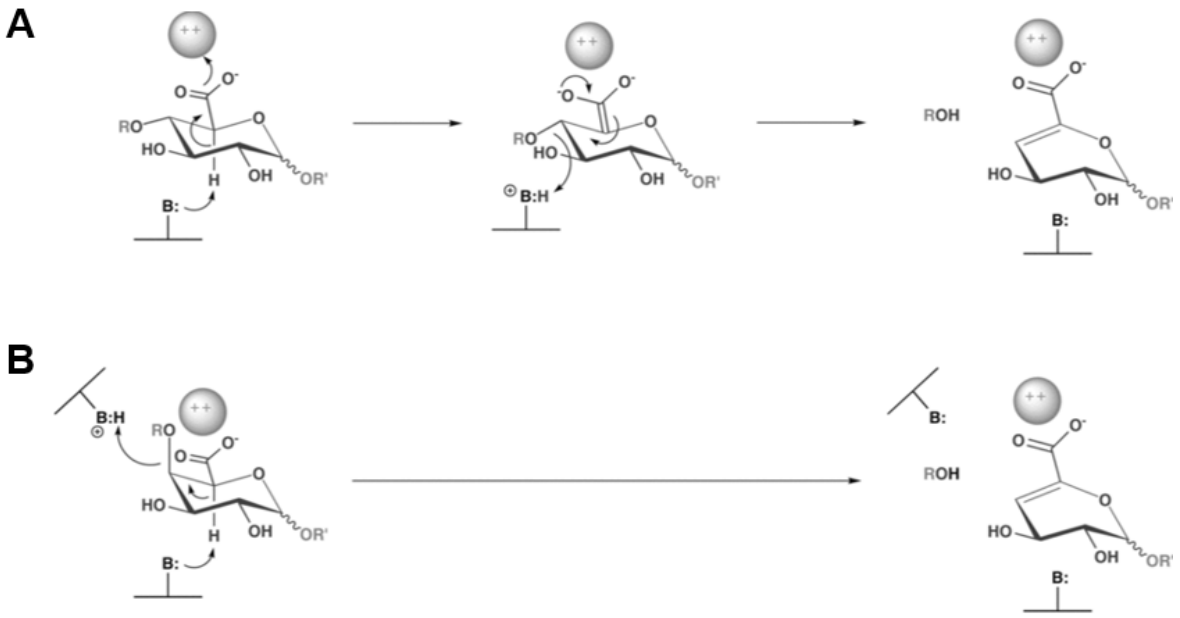


Figure 1.12 Schematic representations of PL general mechanisms *syn* and *anti* β-elimination
Polysaccharide lyases cleave the O-4:C-4 bond yielding 4,5-unsaturated hexenuronic acid and a new reducing end (ROH) by one of two β-elimination mechanisms that differ in the localization of the leaving groups: In *syn*-β-elimination (panel **A**), H5 and O4 are in same side of the sugar ring. A tyrosine that acts first as base ("B") to abstract H5 and after as acid ("B:H") to protonate the scissile glycosidic oxygen; In *anti*-β-elimination (panel **B**) both leaving groups are in opposite sides of the sugar ring. A positively charged acts as base ("B") and a histidine (PL4) or a tyrosine (glycosaminoglycan lyases) protonates the leaving group (Garron and Cygler 2014). The sphere with positive charges represents a polar amino acid or an ion. Figure adapted from (Lombard *et al*. 2010).

1.2.1.3 Carbohydrate esterases

Carbohydrate esterases (CE) catalyse the deacetylation of substituted polysaccharides. In the CAZy database there are 16 sequence-based families that display six folds: $\alpha/\beta/\alpha$-sandwich, $(\beta/\alpha)_7$-barrel, β-helix, $(\beta/\alpha)_8$-barrel, 2-layer sandwich and (α/β)-fold (Cantarel *et al*. 2009; Lombard *et al*. 2014). Most of the CEs are serine esterases that utilize a catalytic triad Asp-His-Ser (Figure 1.13). The Asp acts as pKa modulator of the His that activates the Ser. The nucleophilic attack on the acetyl group generates a tetrahedral oxyanion transition state that is stabilized through interactions with the enzyme oxyanion hole. The His activates a water molecule that hydrolyses the acyl-enzyme intermediate (Hakulinen *et al*. 2000). However, in same families,

catalysis requires only a His-Ser diad or an ion such as Zn^{2+} (Adesioye *et al*. 2016). The bulk of CEs are acetyl esterases that range greatly in specificity. However, there are families with tightly specificity, such as, CE8 that only contains pectin methylesterases (Boraston and Abbott 2012).

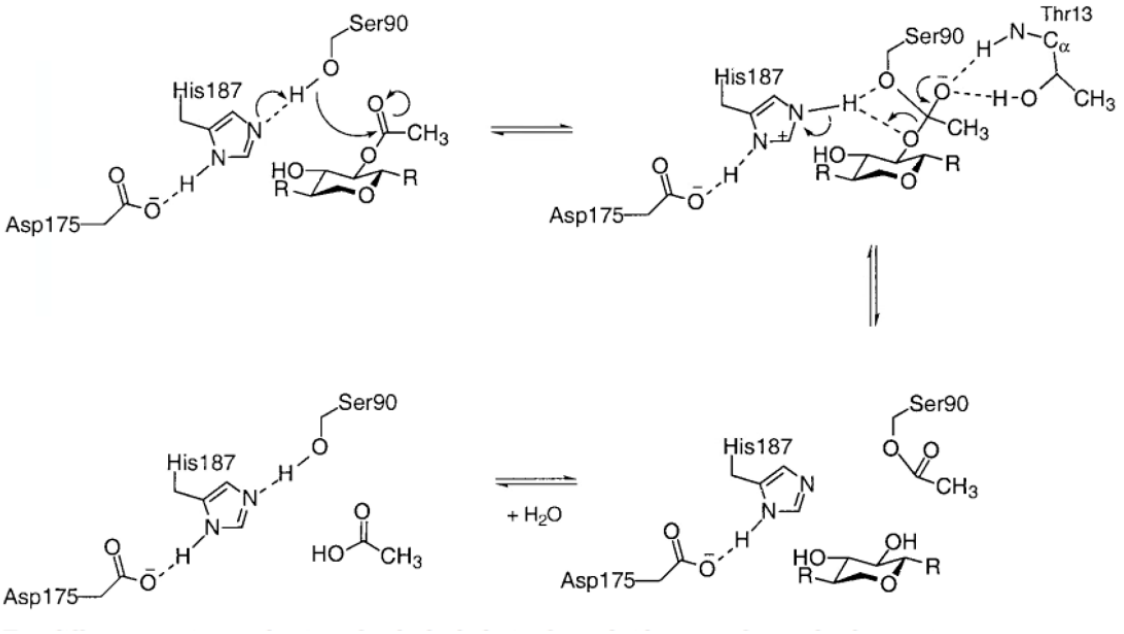


Figure 1.13 Catalytic mechanism of a carbohydrate esterase

Reaction mechanism for a CE that requires the triad Asp-His-Ser. Figure adapted from (Hakulinen *et al*. 2000)

1.2.1.4 Subsite classification

The substrate binding regions of GHs and PLs are divided into subsites. A subsite is characterized by the amino acids that are interacting with a single sugar (Figure 1.14). In an endo-acting enzyme the subsites are located on both sides of the scissile bond. Subsites are numbered from – n (non-reducing end) to + n (reducing end) where -1 represents the active site and the scissle bond is between the sugar bound at -1 and +1 (Davies *et al*. 1997). The number and topology of subsites varies from enzyme to enzyme.

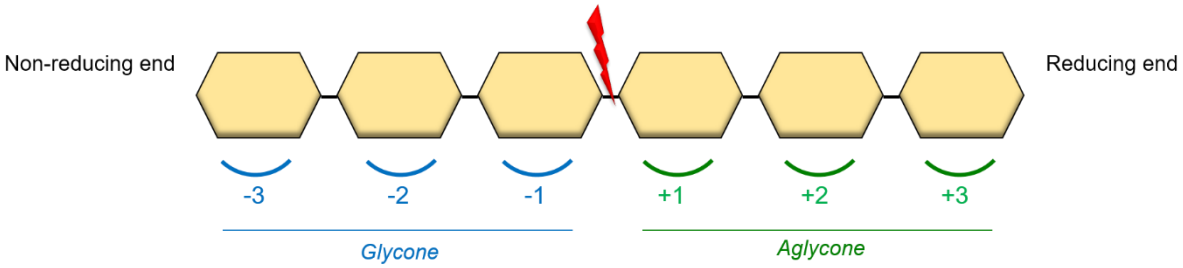


Figure 1.14 Schematic representation of the enzyme subsites

The hydrolysis occurs between -1 and +1 (red). Positive subsites (green) are located in the leaving group region of the substrate binding region and negative subsites (blue) are located in the glycone.

Despite the huge number of enzymes identified and characterized in the CAZy database (Cantarel *et al*. 2009; Lombard *et al*. 2014), understanding how these enzymes work together to achieve the degradation of complex polysaccharides remains limited. From a biotechnological perspective, it is essential to understand the mechanisms of enzymatic regulation and link this information in a rational way that mimics polysaccharides degradation in nature.

1.2.2 *Carbohydrate-binding modules*

Efficient plant cell wall degradation requires physical linkage, or at last close association, between catalytic modules (GHs or PLs) with non-catalytic carbohydrate-binding modules (CBMs). CBMs were first found to bind crystalline cellulose (Gilkes *et al*. 1988). Currently, CBM specificity reveals a diversity of ligands (Cantarel *et al*. 2009; Lombard *et al*. 2014).

1.2.2.1 Classification

CBMs are classified into sequence-based families in the CAZy database. So far, 81 CBM families have been identified (October, 2016) (Cantarel *et al*. 2009; Lombard *et*

al. 2014). This classification highlights similarities in proteins fold between families. However, the classification does not always provide insight into specificity as there are several polyspecific CBM families. For example, CBM family 4 contains members that bind to amorphous cellulose, xylan and different β-glucans (Abou Hachem *et al*. 2000). However, in some families, such as CBM1, CBM5, CBM10 and CBM77, substrate recognition is specific (Raghothama *et al*. 2000; Venditto *et al*. 2016).

The three-dimensional structure of members of 54 CBM families has been determined using initially NMR but latterly X-ray crystallography. The data reveal “seven different folds” (Figure 1.15): β-sandwich, β-trefoil, cysteine knot, OB-fold, hevein fold, and hevein-like fold (reviewed by Boraston *et al*. 2004). The β-sandwich, which comprises two β-sheets, is the dominant fold that is adopted by >20 CBM families (Fujimoto 2013).

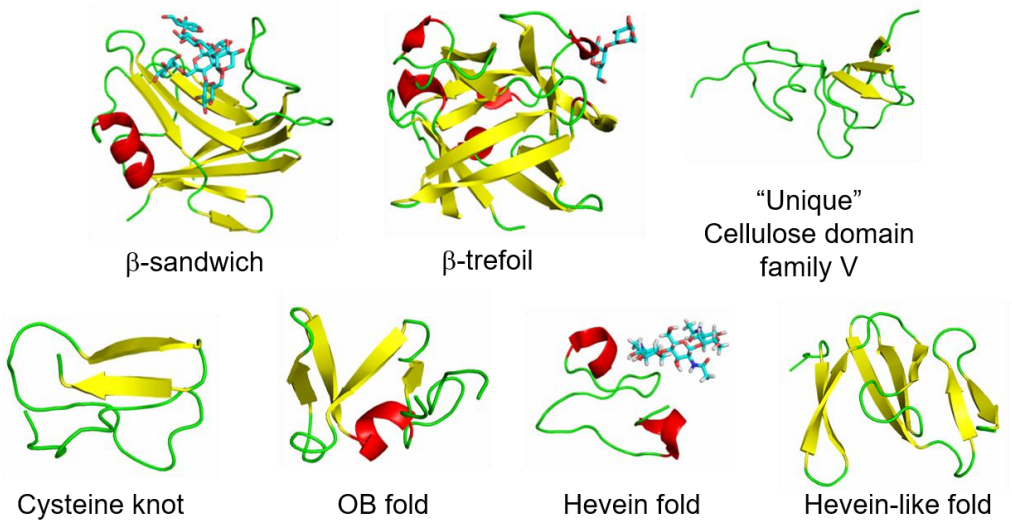


Figure 1.15 Schematic representation of CBM folds

Secondary structural elements are coloured in red (α-helices), yellow (β-sheet) and green (loops). The ligands are coloured in cyan (carbon) and red (oxygen). The representative CBMs displayed are as follows: (β-sandwich) CBM65 from *Eubacterium cellulosolvans* 5 in complex with XXXG (PDB 2YPJ); (β-trefoil) CBM13 from *Clostridium thermocellum* in complex with di galactose (PDB 3VSY); (Cellulose domain) CBM5 from *Pyrococcus furiosus* DSM3658 (PDB 2RTS); (cysteine knot) CBM1 from *Thichoderma reesei* (PDB 1AZ6); (OB fold) CBM10 from *Cellvibrio japonicus* (PDB 1E8R); (Hevein fold) CBM18 from *Hevea brasiliensis* in complex with N,N,N-triacetlyglucosamina (PDB 1TOW); (Hevein-like fold) CBM14 from *Blomia tropicalis* (PDB 2MFK). The figure was generated in PyMOL.

CBMs families can also be grouped into different types reflecting their mode of ligand recognition: “surface-binding CBMs” (type A), “*endo*-type CBMs” (type B), and “*exo*-type CBMs” (type C) (Figure 1.16) (reviewed by Gilbert *et al*. 2013). This classification reflects the complementarity between the CBM binding site and the conformation of the ligand. In type A, the planar hydrophobic surface of the CBM binding site interacts with the flat surface of crystalline polysaccharides such as cellulose or chitin. Type A CBMs include, among others, CBM families 1, 2, 5, 10 and 63 (reviewed by Gilbert *et al*. 2013). Type B CBMs include a large variety of CBM families that bind to internal polysaccharide chains (reviewed by Gilbert *et al*. 2013). These *endo*-type CBMs have a ligand-binding site that displays a groove or cleft topology that is optimised to linear, helical or branched glycan chains (Najmudin *et al*. 2006). CBM65 is a type B CBM that binds preferentially to xyloglucan but the recognition is specifically mediated by the ligand side chains (Luis *et al*. 2013). Type C CBMs or *exo*-type are characterized by a pocket-like binding site. This CBM type include different families, such as 9, 6, and 35 families (reviewed by Gilbert *et al*. 2013). There are examples of CBM families containing members of more than one type, or even two different binding sites in the same protein. For example, CBM6 from *Cellvibrio mixtus* has a type B and C binding site with different specificities. Cleft A (type C) binds the reducing end sugar of β1,3 and β1,4-glucans, while cleft B (type B) recognizes mixed linked β(1,3)(1,4)-glucans (Pires *et al*. 2004). Additionally, CBM32 has members displaying a cleft that can accommodate polysaccharides (type B), and also proteins that contain a pocket that binds only the reducing end sugar (type C) (Abbott *et al*. 2007; Ficko-Blean *et al*. 2012). Some CBMs families only contain type B (CBM36) protein modules, while others have only type C CBMs members (CBM62) (Jamal-Talabani *et al*. 2004; Montanier *et al*. 2011).

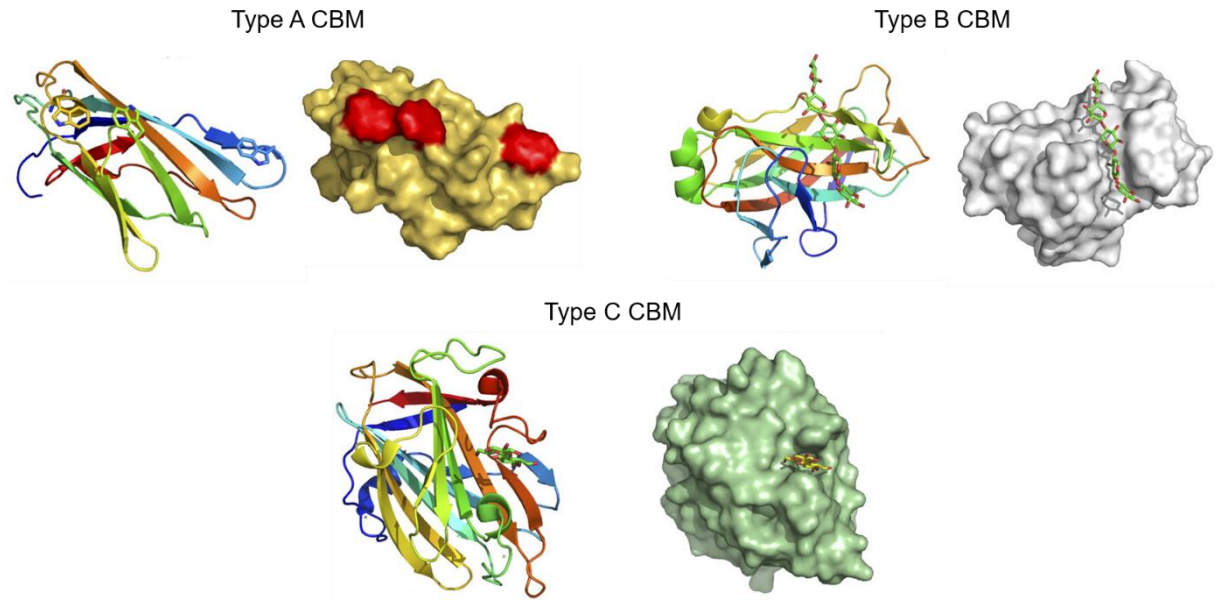


Figure 1.16 Carbohydrate binding modules Type A, B and C

For each CBM type is show the surface and the secondary structure cartoon ramped coloured from blue (N terminus) to red (C terminus). Type A corresponds to CBM2a from the *Cellulomonas fimi* xylanase Xyn10A (PDB 1XG); Type B shows CBM15 from *C. japonicas* xylanase Xyn10C (PDB 1GNY) and; Type C presents the CBM9 from a *Thermotoga maritima* GH10 xylanase (PDB 1I82). The folds are ramped from blue (N terminus) to red (C terminus). The three aromatic residues that form a ligand-binding apolar surface in the CBM2a module are colored red and are shown in stick format in the respective structure. Figure adapted from (Gilbert 2010). Bound ligand is shown in stick format with carbons coloured green (type B) or yellow (type C).

1.2.2.2 CBM functions

It is widely accepted that CBMs can have an important role in enhancing the efficiency of enzymatic breakdown of plant structural polysaccharide (Herve *et al*. 2010). Different functional roles have been associated with these non-catalytic modules. Usually, CBM specificity reflects the substrate targeted by the appended enzyme. Through this association of specificities, CBMs can target substrates accessible to degradation by the cognate enzymes (CBM targeting effect) (McCartney *et al*. 2006). CBMs can also potentiate the interaction of the enzyme and substrate, increasing the local concentration of the hydrolase leading to enhanced catalytic efficiency through a "proximity effect" (Bolam *et al*. 1998; Herve *et al*. 2010). More controversial is the ability of some CBMs to apparently disrupt crystalline structures increasing substrate

accessibility and enhancing the degradation efficiency; the “CBM disruption effect” (Din *et al*. 1994). Furthermore, some CBMs can, apparently, direct enzymes to regions of the plant cell wall that undergoing rapid degradation and are thus highly accessible to hydrolase attack; the “enzymatic cell wall adhesion effect” (Montanier *et al*. 2009). Additionally, some studies provide evidence that CBMs influence enzymatic specificity exemplified by a CBM66 that confers exo-levanase activity on a non-specific fructosidase (Cuskin *et al*. 2012).

The diverse specificities exhibited by CBMs characterized to date can be explained by the complex composition of plant cell walls (Blake *et al*. 2006; McCartney *et al*. 2006). However, until now, CBM characterization has been dominated by binding and structural studies with purified ligands. Since CBMs are linked to the catalytic modules of enzymes, from a biotechnological viewpoint, it is essential to understand how these non-catalytic modules contribute to the plant cell wall degrading capacity of specific microorganisms. Such knowledge will have a major impact in designing enzyme-directed bioengineering strategies to improve the efficient plant cell wall degradation.

1.2.3 *Pectin degrading enzymes*

The degradation of different polysaccharides requires a specialized enzymatic cocktail able to targ*et al*l the different linkages in the plant cell wall. Pectin degradation is achieved through the combination of multiple enzymes specialized in side chain removal (galactosidases and arabinofuranosidases), methylesterases responsible for cleaving the methoxyl groups, and PLs and GHs that target the backbone of the various polysaccharides (reviewed by Gilbert 2010). In this section, only the enzymes required to depolymerize the pectin backbone will be described.

1.2.3.1 Pectin lyase, pectate lyase and rhamnogalacturonan lyase

PLs active on homogalacturonan (HG) cleave the α1,4-linkage between two D-GalA units, and are dominated by pectin lyases, if active on methylated substrates, or pectate lyases for PLs that require the de-esterification of C6 of D-GalA. These enzymes are members of families PL1, PL2, PL3, PL9 and PL10 (reviewed by Lombard *et al*. 2010). The RG-I modifying PLs, rhamnogalacturonan lyases, cleave α-L-Rha-α1,4-D-GalA linkages and are located in families PL4 and PL11 (Lombard *et al*. 2010). Despite the different folds displayed by PLs (see Figure 1.11), the topology of the catalytic apparatus of PL10 and PL1 are conserved providing an example of convergent evolution (Charnock *et al*. 2002). The PL mechanism of cleavage occurs by *anti* β-elimination (see Figure 1.12) where the catalytic base is an arginine (PL1, PL2, PL10 and PL11) or a lysine (PL3, PL4 and PL9). However, lyases do not present an obvious residue that can act as proton donors and it is suggested that a water molecule is implicated in protonation of the leaving group (reviewed by Garron and Cygler 2010; Charnock *et al*. 2002). The only exception is PL4 enzymes where a histidine was identified as the proton donor (Jensen *et al*. 2010). Typically, the lyase mechanism requires a divalent ion (calcium or, manganese in PL2 family). As mentioned before, this positive charged ion interacts with the C6 carboxyl increasing the C5-proton acidification and facilitates its abstraction by the catalytic base (Charnock *et al*. 2002; Jenkins *et al*. 2004; Abbott and Boraston 2007a; Ochiai *et al*. 2007; Creze *et al*. 2008; Jensen *et al*. 2010). However, some PLs, such as pectin lyases (from PL1 family) and PL4 family members do not have a function divalent ion (Jensen *et al*. 2010). In PL4 family, the lack of calcium at +1 is compensated by solvent-mediated hydrogen bounds between an aspartate and the C6 carboxyl group (Jensen *et al*. 2010). In pectin lyases, the presence of the methyl group at C6 will mask the negative charge and the interactions with calcium. In this lyases, an arginine interacts

with the carboxylate to increase H5 acidification, Indeed, pectin lyases present an active site more hydrophobic and negatively charged compared to pectate lyases. This allows the specific recognition of the more positively methylated pectin. In pectate lyases, the charged residues and calcium present a positively charged active site that binds to the negatively charged demethylated pectate substrates (Mayans *et al*. 1997).

1.2.3.2 GH28 enzymes

The GH28 family include exo and endo polygalacturonases that cleave the HG linkage α-D-GalA-α1,4-D-GalA at the non-reducing end, or within the polysaccharide chain, respectively (Pickersgill *et al*. 1998; Abbott and Boraston 2007b). This family also includes rhamnogalacturonases that cleave α-D-GalA linked α1,4-linked to L-Rha present in the RG-I backbone (Markovic and Janecek 2001). Rhamnohydrolases and rhamnogalacturonan galacturonohydrolase are also reported in this family. These enzymes are exo active and release L-Rha and D-GalA, respectively, from the non-reducing end of RG-I (Mutter *et al*. 1994 ; Mutter *et al*. 1998). GH28 enzymes display a parallel β-helix fold similar to PL1, PL3 and PL9 enzymes (Figure 1.17). Galacturonases present four β-sheets and ten turns, whereas rhamnogalacturonases have a longer protein sequence and show eleven turns (Pickersgill *et al*. 1998). These hydrolases act by a single displacement mechanism with inversion of the anomeric carbon from α to β (Biely *et al*. 1996). However, GH28 do not present a standard inverting mechanism where the general acid and base are separated by 10 Å to allow the correct positioning of a water molecule (see Figure 1.10). In this family two aspartates act as a general base and are responsible for activating the water molecule. These residues are ~5 Å from the general acid (aspartate). Another unusual feature of

these enzymes is that the general base and acid attack the substrate from the same side (van Santen *et al*. 1999).

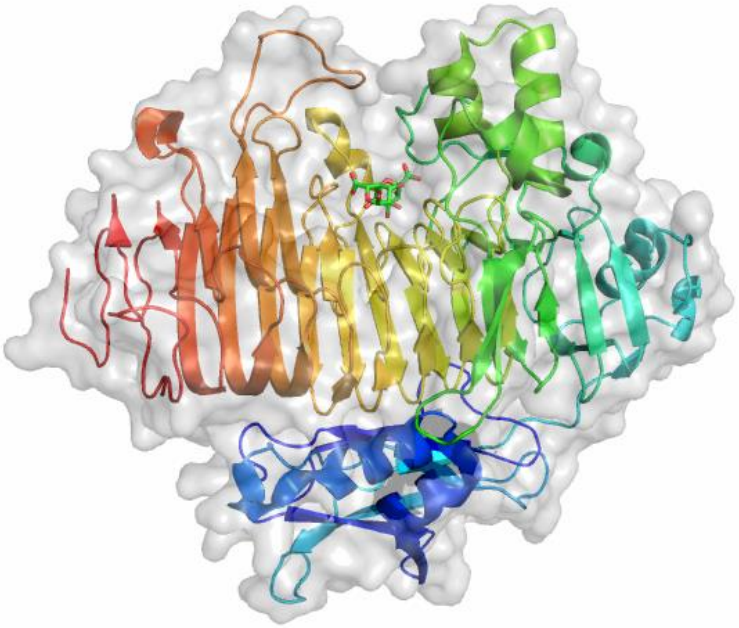


Figure 1.17 Structure of a GH28 family enzyme

Exopolygalacturonase from *Yersinia enterocolitica* (PDB 2UVF) in complex with digalacturonic acid. Cartoon is colour ramped from blue (N-terminus) to red (C-terminus). Ligand is shown as sticks with carbon coloured in green. Figure created in PyMOL.

1.2.3.3 GH105 enzymes

The GH105 family contains unsaturated rhamnogalacturonyl hydrolases that act after the lyases releasing unsaturated galacturonic acid. The active site of these enzymes is located in a pocket of the $(\alpha/\alpha)_6$-barrel (Figure 1.18A) (Itoh *et al*. 2006b). GH105 enzymes are structurally similar to unsaturated GH88 glucuronyl hydrolases (Itoh *et al*. 2006a). These enzymes do not have a conventional GH mechanism but utilize a vinyl ether hydration mechanism (Figure 1.18B). An aspartate is the proposed general acid/base (Itoh *et al*. 2006a). The reaction is initiated with the protonation at C4 by the general acid/base, producing an oxocarbenium ion-like intermediate between C5 and the endocyclic oxygen. The general acid/base activates a water molecule that attacks C5 generating a hemiketal. This product is unsaturable and can spontaneously convert

to an intermediate hemiacetal leading to cleavage of the glycosidic bond. The unsaturated product can be protonated generating cyclic products (Jongkees and Withers 2011).

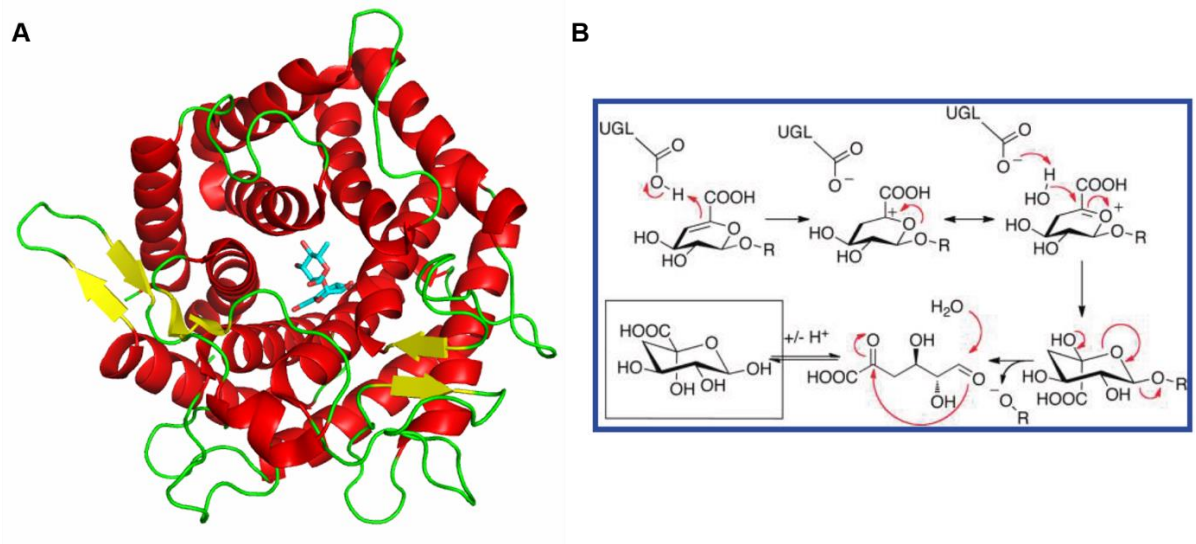


Figure 1.18 Schematic representation of GH105 structure and proposed mechanism

A. Structure of a rhamnogalacturonyl hydrolase from *Bacillus subtilis* in complex with unsaturated galacturonic acid α1,2-linked to L-rhamnose (PDB 2GH4). Secondary structural elements are coloured in red (α-helices), yellow (β-sheet) and green (loops). The ligands are coloured in cyan (carbon) and red (oxygen). Figure created in PyMOL. **B.** Schematic representation of the vinyl ether hydration mechanism. Figure adapted from (Jongkees and Withers 2011).

1.3 Polysaccharide degradation by human gut microbiota

1.3.1 *Microbiota*

The human intestine is colonized by an extremely dense and complex bacterial community (10^{11} - 10^{12} bacteria/ml), the microbiota. These bacteria establish one of the most interesting symbiotic relationships with the human host resulting in mutual benefits. The human gastrointestinal tract provides a controlled environment rich in nutrients for the microbiota, while metabolism of this ecosystem contributes to the nutrition and physiology of the mammalian host. (Backhed *et al*. 2005).

The major source of energy for the microbiota comprises a huge range of complex plant carbohydrates (plant cell wall polysaccharides and resistant starch) that are not accessible to human digestive enzymes and are thus presented to the large bowel. Host glycans (also provided by the animal component of the human diet) such as mucins, (*O*-glycans), *N*-glycans and glucosaminoglycans (GAGs), are additional substrates used by the microbiota (Sonnenburg *et al*. 2005; Bjursell *et al*. 2006; reviewed by Koropatkin *et al*. 2012). The anaerobic fermentation of these non-digestible polysaccharides produce short-chain fatty acids (SCFA) (together with CO_2, H_2 and CH_4) that can be absorbed in the intestine. In humans, the SCFAs represent ~ 10% of the energy balance, however, in herbivorous this fermentation process is responsible for an impressive 70% of the maintenance energy (McNeil 1984).

The close relationship established between microbiota and host has a major impact in human health (Figure 1.19A). These gut bacteria are implicated in a reduction in pathogen colonization and modulation of the local and systemic immune system (Schachtschneider *et al*. 2013). Moreover, several studies have proposed that alterations in microbiota composition and function (dysbiosis) can have an important role in diseases such as, cancer, diabetes, obesity and inflammatory bowel diseases (Crohn's disease and ulcerative colitis) (An *et al*. 2007; reviewed by Clarke *et al*. 2012; reviewed by Tremaroli and Backhed 2012; Midtvedt *et al*. 2013; Marietta *et al*. 2013). For example, in obesity several studies established an interesting relationship between the changes in microbiota composition and this physiological state. Microbiota that contain a high proportion of bacteria highly efficient in energy extraction (Firmicutes), can lead to an increased risk of obesity (Ley *et al*. 2005). Additionally, the short chain fatty acid butyrate has been associated with anti-proliferation in colon cancer cell lines (Hu *et al*. 2011).

The dietary intake of different polysaccharides that are not degraded by human intestinal enzymes can have a major impact on the structure of the microbiota. Therefore, dietary manipulation can represent a powerful approach to modifying the microbiota to enhance the proliferation of beneficial bacteria (reviewed by Flint 2012). Indeed, to improve human health the intestinal microflora has been manipulated by the intake of probiotics (beneficial bacteria that promote health of the host) or prebiotics, such as, inulin and other fructooligosaccharides (non-digestible polysaccharides that promote the growth of beneficial groups of bacteria, like *Bifidobacteria*) (Duncan *et al*. 2003; reviewed by Guinane and Cotter 2013).

Recently, important developments in genomics and proteomics have contributed to a better understanding of microbiota complexity (Turnbaugh *et al*. 2007; Caporaso *et al*. 2011). Interestingly, despite the enormous variability of strains in the human microbiota, more than 90% of these bacteria are included in just two of the ~70 known bacterial phyla: Firmicutes and Bacteroidetes (Figure 1.19B). Remarkably, 95% of Firmicutes were identified as *Clostridia* members, characterized as butyrate-producing bacteria. The Bacteroidetes phyla include, among others, *Bacteroides thetaiotaomicron (Eckburg et al. 2005; Ley et al. 2008)*. A recent study proposed that the human microbiota is divided into three enterotypes that are characterized by a microbial ecosystem that share a common functional role and may have evolved to function optimally as a community inside the cluster (Figure 1.19C). It is proposed that enterotype 1 is a *Bacteroides* dominant cluster, characterized by the use of polysaccharides and proteins as the energy source (characteristic of a western diet) (Arumugam *et al*. 2011; Hold 2014). Enterotype 2 is characterized by *Prevotella* sp. and is associated with a diet rich in carbohydrates characteristic of an agrarian society (Africa) (Wu *et al*. 2011). Enterotype 3 is dominated by *Rumincoccus* and species involved in mucin degradation such as *Akkermansia* (Arumugam *et al*. 2011).

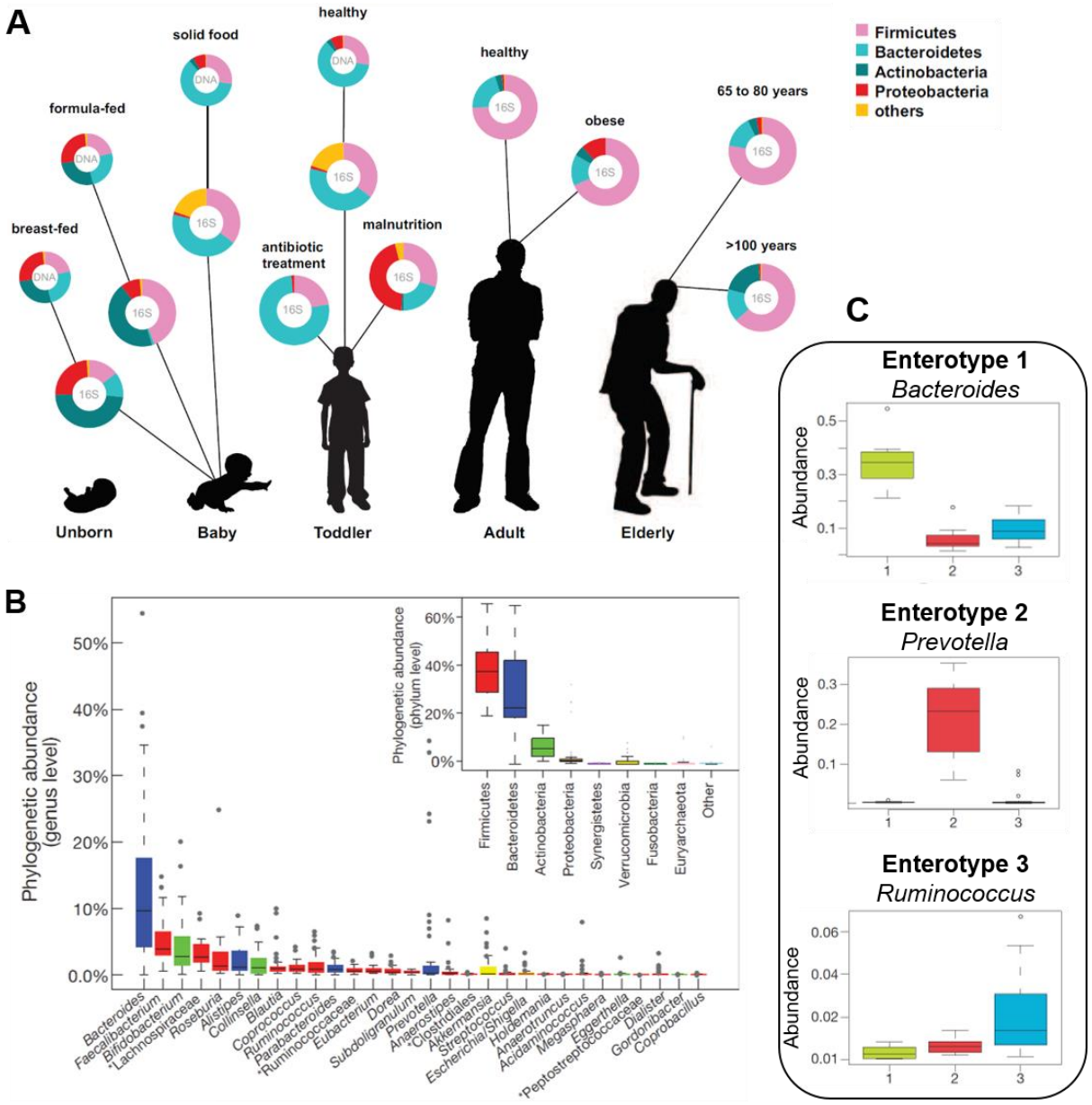


Figure 1.19 Human microbiota

A. Relative abundance of the different phyla of the human microbiota in different stages of life by measuring either 16S RNA or metagenomic approaches (DNA). The panel shows the impact of the diet or human health on microbiota composition. **B.** Abundance of different phylum and genus in the human microbiota (healthy adult). **C.** Abundance of the main contributors of each enterotype where 1, 2, and 3 represents Bacteroides, Prevotella and Ruminococcus. Panel A adapted form (Ottman *et al*. 2012) and Panels B and C adapted from (Arumugam *et al*. 2011).

1.3.2 *Bacteroides thetaiotaomicron*

B. thetaiotaomicron is a Gram-negative bacterium that can comprise 12% of the phyla Bacteroidetes, and 6% of the bacteria, in the human gut (Ley *et al*. 2006). This

generalist bacterium is highly adapted to the gut microenvironment due to its ability to utilize a huge range of dietary polysaccharides as energy sources and, in their absence, rapidly shift to host *O*-glycan utilization (Salyers *et al*. 1977; Sonnenburg *et al*. 2005; Bjursell *et al*. 2006). In *B. thetaiotaomicron,* like almost all the bacteria included in Bacteroidetes phyla, the genes encoding all the required "machinery" for the recognition, uptake and degradation of a specific glycan are physically linked and co-regulated in regions of the chromosome known as polysaccharide utilization loci (PULs).

1.3.2.1 Polysaccharide utilization loci (PULs)

Salyers and colleagues identified and characterized the first PUL in the *B. thetaiotaomicron* genome. This locus, which orchestrates starch metabolism contains eight genes (*sus*RABCDEFG), and is known as the starch-utilization system (Sus) (Figure 1.20) (Anderson and Salyers 1989a, 1989b; Tancula *et al*. 1992; Reeves *et al*. 1997). SusG, an extracellular neopullulanase, catalyses the initial degradation of starch to small oligosaccharides, which are then imported into the periplasm by SusC, a porin that derives the energy required for active transport from the TonB-ExbBD complex (Reeves *et al*. 1996; Shipman *et al*. 1999; reviewed by Koropatkin and Smith 2010). The SusC protein interacts with SusD, a protein that binds to maltooligosaccharides and directs these sugar polymers into the porin (Cho and Salyers 2001; Koropatkin *et al*. 2008). The role of SusF and SusE is less clear, but these outer membrane proteins bind to starch and play a role in presenting the polysaccharide to the surface amylase, SusG (Shipman *et al*. 2000). In the periplasm, SusA and SusB, a neopullulanase and α-glycosidase, respectively, complete the degradation of the oligosaccharides to simple sugars (D'Elia and Salyers 1996a). The glucose released in the periplasm is transported across the inner membrane by an

undefined permease and used in fermentation pathways. SusR is the sensor that controls the expression of the starch PUL. In the presence of starch the maltose released in the periplasm binds to the SusR periplasmic sensor domain upregulating the expression of the other Sus proteins (D'Elia and Salyers 1996b). A common feature of the glycan degradation apparatus is the association of different proteins with different functions (enzymes and binding modules) in order to achieve the complete degradation of the substrates to simple sugars that can be imported into the cell and used as energy source (reviewed by Martens *et al*. 2009a).

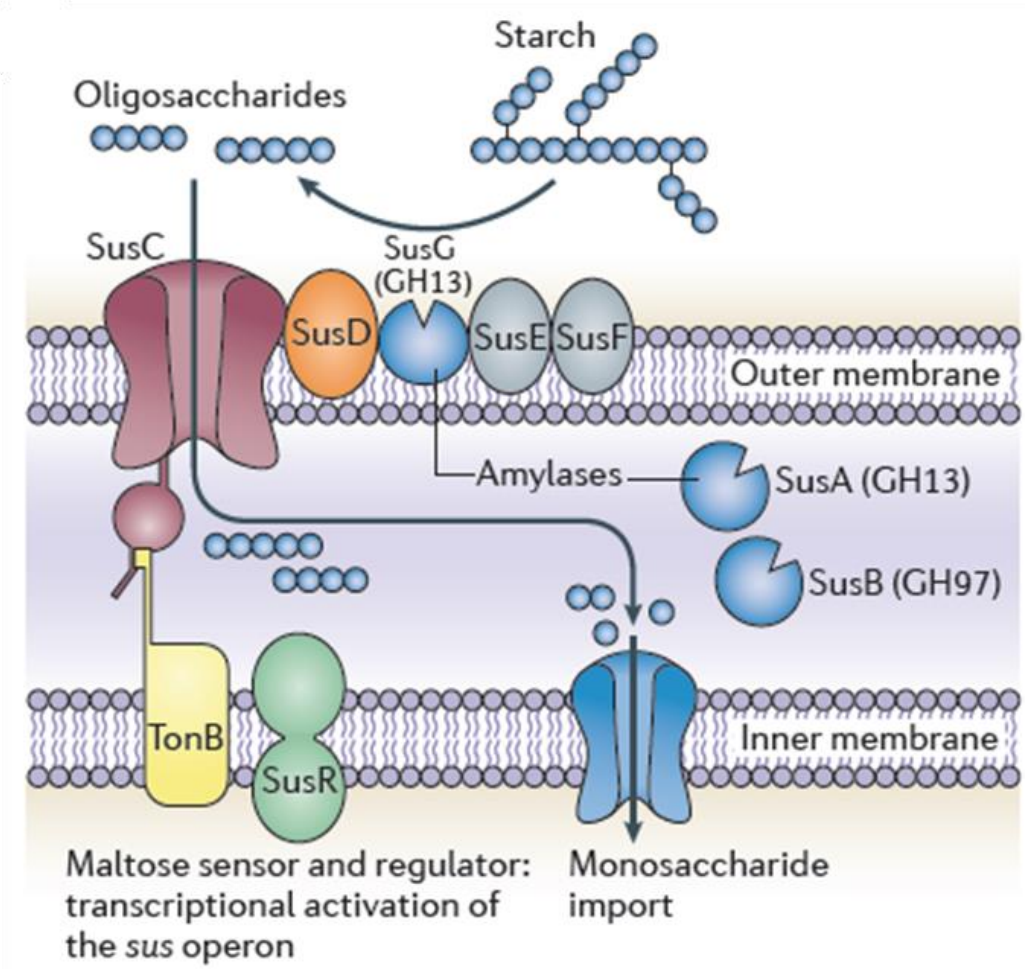


Figure 1.20 Schematic representation of *B. thetaiotaomicron* starch utilization system

The extracellular proteins SusG, SusD, SusE and SusF are responsible for initial starch degradation the products of which are import through the porin SusC. In the periplasm SusA and SusB complete breakdown the imported oligosaccharides, releasing glucose that is imported through a permease located in the inner membrane. SusR is a periplasmic sensor implicated in regulation of the PUL. (GH) Glycoside hydrolase. Adapted from (Koropatkin *et al*. 2012).

Since the first description of the Sus-system, different PULs containing SusC/SusD-like proteins have been identified in *Bacteroides sp*. The *B. thetaiotaomicron* genome encodes 106 and 57 paralogues of SusC and SusD, respectively, most of them organized in SusC/SusD-like pairs in one of 88 PULs. Each of these PULs also encode a variable number of enzymes responsible for the degradation of the polysaccharide targeted by the specific locus (Xu *et al*. 2003). The regulation of these PULs is generally through one of two systems: extra-cytoplasmic function (ECF) σ-factor/anti-σ-factor (in mucin *O*-glycan degradation) or a hybrid two-component system (HTCS) (implicated in plant cell wall polysaccharides and *O*-linked GAGs utilization) (reviewed by Xu *et al*. 2004; Martens *et al*. 2009b). Like SusR, these regulators are specific for each PUL and allow the bacteria to sense the polysaccharides available in the surrounding microenvironment and activate transcription of the locus to achieve the degradation of these glycans (Sonnenburg *et al*. 2006; Martens *et al*. 2011).

The genomic organization of PULs reflects how these bacteria evolved to survive in the highly competitive gut environment (reviewed by Koropatkin *et al*. 2012). In *B. thetaiotaomicron* strain *VPI-5482* PULs comprise 18% of the genome. These PULs encode an unprecedented number of proteins related to polysaccharide degradation (Xu *et al*. 2003). It is predicted that this microorganism expresses 269 GHs from 47 different families. The *B. thetaiotaomicron* genome also encodes 16 PLs derived from nine families, seventeen CEs from nine families and 32 different CBMs from nine families (Cantarel *et al*. 2009; Lombard *et al*. 2014). The organization of a large number of these genes within the PULs enables *B. thetaiotaomicron* to use a wide range of polysaccharides (Xu *et al*. 2007).

Recent studies have combined genomic and *in vivo* experiments to uncover the features leading to the successful adaptation of *B. thetaiotaomicron* to different gut environments, and improve knowledge of the nutrient drivers that define the structure

of the microbiota. In the absence of polysaccharides in the gut, such as during the suckling period, *B. thetaiotaomicron* survives by using host glycans (mainly *O*-glycans) (Martens *et al*. 2008). Another study indicates that *B. thetaiotaomicron* and *B. ovatus* are able to use almost all the common plant cell wall polysaccharides (except cellulose). Martens *et al*. also disclosed the specific PULs that mediated the degradation of each different polysaccharide. For example, in *B. thetaiotaomicron* the PUL *bt4145-83* is upregulated in response to RG-I. Moreover, this study also suggests that these two closely related bacteria evolved to target different polysaccharides as an adaptation mechanism to their different gut niches. For example, *B. thetaiotaomicron* was shown to be associated with the utilization of accessible plant cell wall polysaccharides, such as pectins. However, *B. ovatus* evolved to use less soluble plant cell polysaccharides, such as hemicelluloses (Martens *et al*. 2011). A recent study showed that when presented with a range of polysaccharides *B. thetaiotaomicron* displays a clear hierarchy in glycan utilization. In this study, pectic galactan, heparin, chondroitin sulphate and homogalacturonan were preferentially metabolized in a mixture containing α-mannan, arabinan, arabinogalactan, RG-I, mucin-O-linked glycans, dextran and levan (Rogers *et al*. 2013). The degradation of highy abundant gycans, such as xyloglucan, is achived by niche species (related to *B. ovatus*) suggesting that alterations in microbiota composition can have a major impact in human health and nutrion (Larsbrink *et al*. 2014).

The high complexity of the microbiota is revealed at the functional level where different bacteria have evolved to use divergent systems in order to ensure an efficient degradation of all available polysaccharides. These differences are evident in the two most common microbiota phyla. *Firmicutes* are highly specialized in the degradation of recalcitrant substrates, such as resistant starch (Ze *et al*. 2013). However, the glycan degrading systems identified to date in several *Bacteroides* reveal that these bacteria

can degrade all the available glycans, with the exception of cellulose (reviewed by Koropatkin *et al*. 2012). Additionally, the degradation of some polysaccharides also depends on synergistic interaction between bacteria. For example, some microbiota members can release polysaccharide breakdown products that can be utilized by other bacteria that are unable to degrade this polysaccharide (sharing mechanism) (Rakoff-Nahoum *et al*. 2014). However, with respect to complex glycan utilization, a selfish mechanism of polysaccharide utilization was recently described in *B. thetaiotaomicron*. The selfish mechanism can result of the presence specific enzymes to degrade the polysaccharide at the cell surface (corn xylan) or from a degradation that occurs mainly in the periplasm (in yeast mannan) (Cuskin *et al*. 2015; Rogowski *et al*. 2015). These studies underpin the complexity of the mechanisms underlying polysaccharide utilization by *B. thetaiotaomicron* in different physiological environments within the human gut.

1.4 Cellulolytic bacteria *Ruminococcus flavefaciens*

The anaerobic Gram-positive bacterium *Ruminococcus flavefaciens* is able to degrade highly recalcitrant substrates, such as cellulose (crystalline and amorphous), a major component of the plant cell wall (Pettipher and Latham 1979; Krause *et al*. 1999). In this microorganism, the highly efficient degradation of recalcitrant polysaccharides is achieved by the presence of a multi-enzymatic complex, the cellulosome.

1.4.1 *Cellulosome*

A cellulosome is an extracellular multi-modular protein complex which was first described in the cellulolytic bacterium *Clostridium thermocellum* in the 1980s (Bayer

and Lamed 1986; Lamed *et al*. 1987). Subsequently, this highly efficient cellulose-degrading system has been identified in several anaerobic bacteria and fungi (Table 1.1).

Table 1.1 Anaerobic bacteria and fungi producing cellulosomes

Species	Reference
Bacteria	
Acetivibro cellulolyticus	(Dassa *et al*. 2012)
Bacteroides cellulosolvens	(Ding *et al*. 2000)
Clostridium acetobutylicum	(Sabathé *et al*. 2002)
Clostridium cellulovorans	(Tamaru *et al*. 2010)
Clostridium cellulolyticum	(Blouzard *et al*. 2010)
Clostridium josui	(Kakiuchi *et al*. 1998)
Clostridium papyrosolvens	(Pohlschröder *et al*. 1994)
Clostridium thermocellum	(Lamed *et al*. 1987)
Ruminococcus champanellensis	(Ben David *et al*. 2015)
Ruminococcus flavefaciens	(Ding *et al*. 2001)
Fungi	
Neocallimaxtic patriciarum	(Fanutti *et al*. 1995)
Orpinomyces sp	(Li *et al*. 1997)
Piromyces sp.	(Fanutti *et al*. 1995)

R. flavefaciens synthesises the most complex cellulosome described so far (reviewed by Flint and Bayer 2008). The cellulosome basic modular structure consist of catalytic subunits (enzymes) associated with a structural non-catalytic protein (scaffoldin) through dockerin-cohesin interactions. Briefly, the scaffoldin comprises several non-catalytic domains, termed cohesions that bind specifically to dockerins present in catalytic modules (reviewed by Fontes and Gilbert 2010).

The *R. flavefaciens* strain FD-1 cellulosome has four scaffoldins (ScaA, B, C and E) that can interact with each other or the bacterial cell surface (Figure 1.21) (Jindou *et al*. 2006). Cellulosome attachment to the bacterial cell wall is through the interaction of a cohesin in a sortase-mediated surface bound ScaE with the cognate dockerin domains in ScaB and in the multi-modular proteins R_flaF_05439 (of unknown

function) and CttA (Rincon *et al*. 2005; Salama-Alber *et al*. 2013). CttA is composed of two cellulose-binding modules that are members of a putative novel CBM family. In contrast with the other cellulosomes studied until now, the *Ruminococcus* cellulosome does not contain a CBM in any of its structural scaffoldins. Instead, CttA mediates adhesion between the plant cell wall and the bacterial surface. (Rincon *et al*. 2007). In addition to CttA, it is assumed that different CBMs appended to catalytic modules also are implicated in the interaction of the cellulosome with its substrate (Rincon *et al*. 2001).

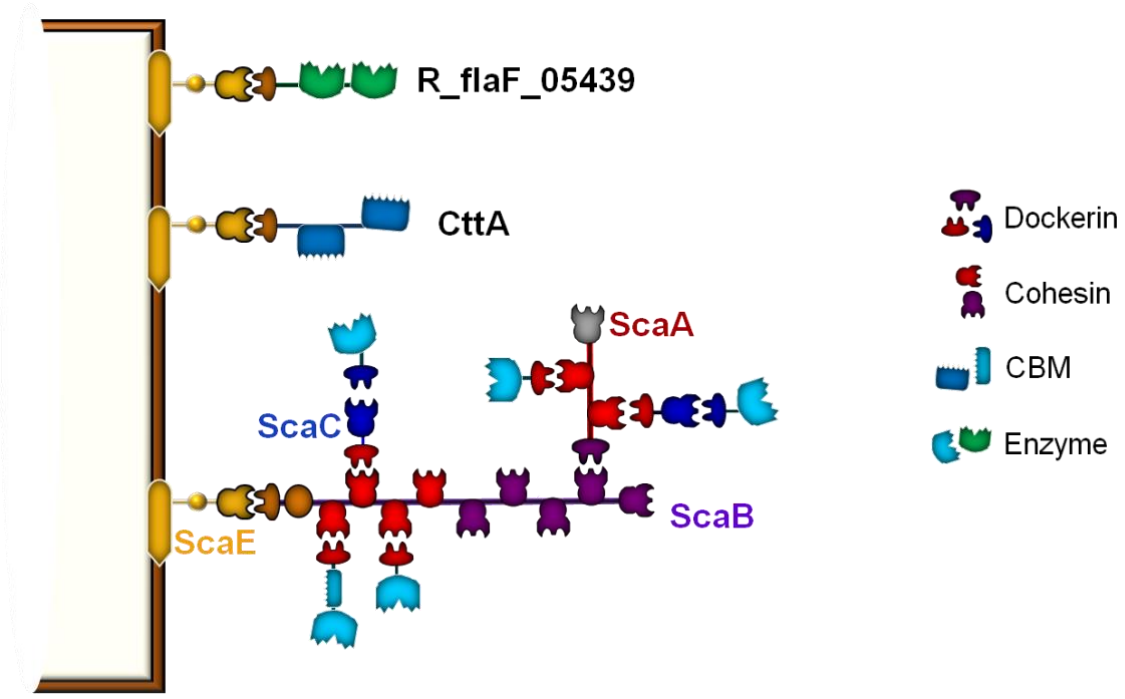


Figure 1.21 *Ruminococcus flavefacien*s FD-1 cellulosomal system

R. flavefaciens cellulosome presents three scaffoldins (ScaA, ScaB and ScaC) harbouring three different types of cohesins (in red, purple and grey) that recognize dockerins present in catalytic modules or other scaffoldin proteins. This multimodular complex is attached to cell wall through the interaction with ScaE. This ScaE also recognize the CttA, comprising two cellulose binding modules, and the modular protein R_flaF_05439, of unknown function.

Despite the huge complexity of the *R. flavefaciens* FD-1 cellulosome, the recent genome sequencing of this microorganism reveals further complex unexpected features. The *R. flavefaciens* genome encodes the largest number of predicted

catalytic modules (GHs, PLs and CEs) described to date in a cellulosomal bacterium. It was predicted that this microorganism encodes 225 open reading frames (ORF) containing dockerins domains, in contrast, *C. thermocellum* contain 76 dockerin-containing ORFs. However, only 33% of *R. flavefaciens* dockerins encoding ORFs are predicted to comprise CAZyme catalytic modules (Berg Miller *et al*. 2009). Bioinformatic analysis identified 64 enzymes from 17 GH families that contain dockerins (Berg Miller *et al*. 2009). In the eight cellulosome-related CBM families identified, half of the 54 CBMs are members of CBM22 and CBM35, which bind to xylan (both families), β-glucans (CBM 22) mannan and uronic acids (CBM 35) (Berg Miller *et al*. 2009). In addition, 12 PLs, located in PL1, PL9 and PL11, have also been shown to contain dockerin domains (Berg Miller *et al*. 2009). Finally, 19 CE modules grouped into seven families also contained dockerin domains and were thus components of the cellulosome. Of particular Interest is that 21% of cellulosome proteins have no domains corresponding to known CAZymes or scaffoldins. These non-CAZyme proteins include transglutaminases, serine protease inhibitors and leucine-rich repeats (Berg Miller *et al*. 2009; Rincon *et al*. 2010).

1.5 Objectives

Plant cell walls consist of polysaccharides organized in a complex structure. The mammalian gut microbiota produces an enormous number of enzymes that allow the degradation of these macromolecular structures. Understanding the degradation mechanisms developed by the different bacteria can have implications, not only in an industrial context, but also in promotion of human or animal health. This study is designed to investigate how pectins, complex polysaccharides of D-GalA, are degraded and targeted by different gut bacteria.

To achieve this aim, this study was divided into two main goals:

i) Characterization of new CBM families displaying binding to pectins

CBMs have an essential role in polysaccharide degradation by allowing close contact between substrate and enzyme. Despite the extensive characterization of CBMs that bind to cellulose and hemicellulose, there is a lack of information on CBMs that bind to pectins. Since pectins are one of the major components of the plant cell wall there is an urgent need to understand the role of CBMs in the degradation of this polysaccharide. This goal is designed to increase our knowledge of pectin-specific CBMs.

ii) Understand how *B. thetaiotaomicron* degrades and utilizes RG-II and I

The identification of the PULs related to the degradation of specific glycans allows the rational understanding of the mechanism by which dietary and host glycans are utilized by the human microbiota. RG-II is the most complex plant polysaccharide known. Until now, there have been no reports on how this glycan is degraded. The activities of enzymes acting on Chain B of RG-II will be explored to contribute to the emerging model of RG-II degradation by *B. thetaiotaomicron*. This work will also focus in RG-I utilization by this gut bacterium. In order to create a complete model, the activities of both PLs and GHs that target the backbone of RG-I will be characterized.

Additionally, the structural characterization of key proteins related to this complex system is essential to understand the general mechanisms by which polysaccharides are degraded by the *Bacteroidetes*. This study is designed to study the structure of several components of a PUL encoded system that degrades RG-I and RG-II.

Chapter 2. Material and Methods

2.1 Molecular biology

2.1.1 *Chemicals, commercial kit and water*

All experiments were made using 18.2Ω H_2O produced by a Millipore Milli-RO 10 Plus Water Purification System. Water was used as the solvent unless stated otherwise. All the chemicals, enzymes and media, which are not specifically listed in the text, were from Sigma Chemical Company, Megazyme or BDH Laboratories Ltd.

2.1.2 *Bacterial strains and plasmid*

The different *Escherichia coli* strains and bacterial plasmids used during this study are listed in Table 2.1 and

Table 2.2, respectively.

Table 2.1 Bacterial strains

Strain	Genotype	Use	Reference
BL21(DE3)	F- *dcm ompT hsdS*(r_B- m_B-) *gal* (DE3)	Protein expression	(Studier and Moffatt 1986)
Tuner™(DE3)	F– *ompT hsd*SB (r_B– m_B–) *gal dcm lac*Y1 (DE3)	Protein expression	Novagen
B834	F– *ompT hsd*S_B (r_B– m_B–) *gal dcm met* (DE3)	Seleno-methionine protein expression	Novagen
One Shot™TOP10	F- *mcrA* Δ(*mrr-hsd*RMS-*mcr*BC) ϕ80*lacZ*ΔM15 Δ*lac*X74 *rec*A1 *ara*D139 Δ(*ara, leu*)7697 *gal*U *gal*K *rps*L (STR^R) *end*A1 *nup*G λ-	DNA ligation and replication	Invitrogen
CC118 λ-pir	Δ(*ara-leu*) *ara*D Δ*lac*X74 *gal*E *gal*K *pho*A20 *thi*-1 *rps*E *rpo*B *arg*E (Am) *rec*A1 λ*pir*	Gene deletion (plasmid propagation)	(Herrero *et al*. 1990)
S17.1 λ-pir	*rec*A *pro hsd*R RP4-2 (Tc::Mu;Km::Tn7) (λ pir)	Gene deletion	(Skorupski and Taylor 1996)

Table 2.2 Plasmids

Plasmid	Size (kb)	Features	Reference
pET28a	5.37	kan^r, T7 *lac* $lacI^q$	Novagen
pExchange-*tdk*	4.2	amp^r	(Koropatkin *et al*. 2008)

Different plasmid multiple cloning sites are shown in Appendix A.

2.1.3 *Sterilization*

All solutions, media and glassware were sterilised by autoclaving using either an Astell Hearson 2000 Series Autoclave or a Prestige© Medical Series 2100 Clinical Autoclave at 121 ^oC, 32 or 15 lb / $inch^{-2}$ for 20 min. When necessary, solutions were filter sterilized using sterile Milipore filter disc (0.22 µm pore size) (Stupor® Acrodisc® 3.2 Gelman Sciences) and a suitable sterile syringe (Plastipak®, Becton Dickinson).

2.1.4 *Growth media*

Media were used alone as broth or as plates with addition of 2% (w/v) of Bacteriological agar. The media utilized in this study are shown in (Table 2.3). *E. coli* strains were grown in sterile Luria-Bertani Broth (LB) or agar plates. *Bacteroides thetaiotaomicron* were cultured anaerobically in sterile Tryptone/Yeast/Glucose (TYG) medium and Minimal Medium (MM). Selenomethionine medium was used for protein utilized in crystallography to solve the phase problem. Brain-heart infusion (BHI) plates were rich media utilized in conjugation protocol (see Section 2.1.23.2). When necessary, after autoclaving the selective antibiotics were added (Table 2.4). Histidine-Hemantin (His-Hem) was add to *Bacteroides* media prior to inoculation (improve the growth). This solution was prepared by dissolving hematin (1.2 g/l) in 0.2 M Histidine-HCl (0.42 g/l) pH 8.0. After preparation His-Hem was stored at 4 °C.

Table 2.3 Growth media

Media	Component	Amount per liter	Details
Luria-Bertani Broth (LB)	LB powder (Sigma-Aldrich)	25 g	Component dissolved in water and pH adjusted to 7.2 with NaOH (before autoclaving).
Tryptone Yeast Extract Glucose (TYG)	Tryptone Peptone	10 g	TYG media was used to obtain a high density of cells. MM was required in conjugations protocol (see Section 2.1.23.2) and to monitor the growth of *B. thetaiotaomicron* in different carbon sources.
	Bacto Yeast Extract	5 g	
	Glucose	2 g	
	Cysteine (Free base)	0.5 g	
	1 M KPO_4 pH 7.2	100 ml	
	Vitamin K solution, 1 mg/ml	1 ml	
	TYG salts ($MgSO_4$ 0.5 g/l, $NaHCO_3$ 10 g/l, NaCl 2 g/l)	40 ml	
	0.8% $CaCl_2$	1 ml	
	$FeSO_4$, 0.4 mg/ml	1 ml	
	Resazurin, 0.25 mg/ml	4ml	
Minimal Medium (MM)	NH_4SO_4	1 g	Component dissolved in water and volume adjusted to 1 l. After autoclaving, medium was allowed to cool and was added His-Hem (1:1000).
	Na_2CO_3	1 g	
	Cysteine (Free base)	0.5 g	
	1 M KPO_4 pH 7.2	100 ml	
	Vitamin K solution, 1 mg/ml	1 ml	
	$FeSO_4$, 0.4 mg/ml	10 ml	
	Vitamin B12, 0.01 mg/ml	0.5 ml	
	Mineral Salts for defined medium (NaCl 18 g/l,$CaCl_2$ 0.53 g/l, $MgCl_2$ 0.4 g/l, $MnCl_2$ 0.2 g/l, $CoCl_2$ 0.2 g/l)	50 ml	
	Resazurin, 0.25 mg/ml	4ml	
Brain-Heart Infusion (BHI)	BHI powder (Sigma-Aldrich)	37.5 g	After autoclaving added His-Hem (1:1000).
Seleno-methionine	SelenoMet Medium Base®	21.6 g	Media components supplied by Molecular dimension. Medium base was prepared and autoclaved. Nutrient Mix was dissolved and sterile filter. SelenoMethionine were added to base medium prior to inoculation.
	SelenoMet Nutrient Mix® (5.1 g)	50 ml	
	SelenoMethionine	4 ml	

Table 2.4 Antibiotics

Antibiotic	Working concentration	Storage	Description (use with)
Ampicillin	50 μg/ml	4 ºC for less than 4 days	pExchange-*tdk*
Kanamycin	50 μg/ml	4 ºC for less than 4 days	pET28a
Gentamycin	200 μg/ml	Prepared when necessary	pExchange-*tdk*
Erythromycin	25 μg/ml	Prepared when necessary	pExchange-*tdk*
5-fluoro-2'-deoxyuridine (FUdR)	200 μg/ml	Prepared when necessary	Conjugation (see Section 2.1.23.2)

All antibiotics were prepared as a stock concentration 1000 x the working concentration.

2.1.5 *Centrifugation*

Volumes higher than 100 ml were centrifuged at 5000 revolution per minute (rpm) for 10 minutes (at 4 ºC), using a Beckman J2-21 centrifuge with a JA-10 rotor. Volumes between 50 and 100 ml were centrifuged at 25000 rpm for 30 minutes (at 4 ºC) using the same centrifuge and a JA25-50 rotor with 50 ml centrifuge tubes (Nalgene). Volumes ranging from 2-50 ml were harvested by centrifugation at 5000 rpm, in 25 ml universal (Sterilin) tubes in a Hettich Mikro 220R benchtop centrifuge with fixed angle rotor. Volumes < 2 ml were centrifuged in Eppendorf tubes (2 or 1.5 ml) at 13000 rpm using a Heraeus Instruments Biofuge pico benchtop centrifuge. For concentration a Harrier 18/80R centrifuge with swing out rotor was used at 4000 g (4 ºC). The centrifugation time was dependent of the sample to concentrate. Before all centrifugation steps all tubes were balanced.

2.1.6 *Storage of DNA and bacteria*

Stocks of different DNA plasmids and genomic DNAs were stored at -20 °C and 4 °C, respectively. Stocks of all bacterial strains were stored as glycerol stocks (25 % (v/v) glycerol) at -80 °C in cryovials.

2.1.7 *Transformation of chemical competent E. coli*

Competent *E. coli* were prepared by Mr. Carl Morland and stored at -80 °C for up to a year. A 100 µl aliquot was thawed and then mixed with 200 ng of plasmid DNA and incubated on ice for 30 minutes. Cells were then heat-shocked at 42 °C for 2 minutes, followed by incubation on ice for 2 minutes. Cells were mixed with 500 µl of LB (see Table 2.3) and incubated 37 °C, 180 rpm for 1 h. Cells were then centrifuged at 5000 rpm and 500 µl of supernatant was removed and the pellet was re-suspended in the remaining 100 µl of LB and plated onto plates containing an appropriate antibiotic. Plates were incubated in an inverted position at 37 °C for maximum 16 h using a static incubator (LTE Ltd).

2.1.8 *Growth conditions for propagation of bacteria*

Stocks of plasmid DNA were made by transformation of the recombinant plasmid into Top10 (for pET28a vector) or CC118 (for pExchange-*tdk* vector). From transformation plates (see Section 2.1.7) single colonies were grown in 5 ml of LB with antibiotic (see Section 2.1.4) 37 ºC, 180 rpm for 16 h. For *B. thetaiotaomicron* a 2.5 - 5 ml culture in TYG was grown (see Section 2.1.24)

2.1.9 *DNA isolation*

Plasmid DNA was purified from 5 ml cultures using the QIAspin Miniprep Kit (QIAGEN) according to the manufacturer's instructions. *B. thetaiotaomicron* genomic DNA was extracted for 1.0 ml cultures using the Sigma GenElute™ Bacterial Genomic DNA Kit (Sigma-Aldrich) according to the manufacturer's instructions.

2.1.10 *Quantification of DNA*

The concentration of DNA was determined by measuring the absorbance at 260 nm using a NanoDrop 2000 UV-Vis spectrophotometer (Thermo Fisher Scientific Inc, USA). The sample purity was estimated based on ratio between absorbance at 260 nm and 280 nm (260/280). For pure DNA samples this value should be ~ 1.8. For a pure sample, the ratio between absorbance at 260 nm and 230 nm (260/230) was expected in the range of 2.0 – 2.2. Values below were indicative of contaminants present in the sample.

2.1.11 *Polymerase Chain Reaction (PCR)*

The polymerase chain reaction (PCR) developed by Mullis and Faloona (Mullis and Faloona 1987) was used to amplify genes from genomic or plasmid DNA, to mutate recombinant genes already cloned into vector plasmids (site direct mutagenesis) and to construct the fragments required to build the gene Knockout construct (this PCR is described in details in Section 2.1.23.1). This technique requires a thermostable DNA polymerase, a DNA template and two oligonucleotide primers that are complementary to each DNA strand and flank the region to be amplified. The polymerase enzyme catalyzes the synthesis of the complementary DNA strand in presence of dNTPs. The reactions were performed using a PHC-3 thermocylcer (Biorad).

All primers used in this study were produced by MWG-Biotech AG (Germany) and supplied lyophilized. Primers were resuspended in sterile double distilled water to 100 µM (stock solution). When necessary, primers were diluted 20x to 5 µM. Stock and working solution were stored at – 20 °C.

For standard PCR, primers were designed to flank the side of the gene to be amplified. The forward primer is complimentary to the 5' end of the forward strand and the reverse to the 5' end of the reverse strand. Primers were constructed with approximately 20-25 bases in length, a ~50 % GC content. To increase the PCR efficiency, when possible, primers started and ended with G's or C's. The melting temperature (Tm) was calculated using the online tool "Oligonucleotide Properties Calculator" (Kibbe 2007). To each pair of primer, the Tm was > 50 °C with a maximum of 2 °C between each primer. When necessary, restriction sites were added to 5' end of primers. In this case, an additional six nucleotides (CACACA) were also added upstream of the restriction site to allow efficient cutting of the PCR product by endonucleases. Primers were checked for internal restriction sites using the online tool NEBcutter (Vincze *et al*. 2003).

For site direct mutagenesis, primers were constructed containing the desired mutation flanked by 10-15 nucleotides and Tm ≥ 78 °C. The mutagenesis was carried out by modifying a single codon in order to introduce the required nucleotides to specifically codify the new amino acid.

2.1.11.1 Standard and site direct mutagenesis PCR

Standard PCR was used to amplify genes from genomic or plasmid DNA. For all PCRs negative controls were set-up without DNA template. The composition of a typical reaction is shown in Table 2.5. Unless stated otherwise, the standard PCR cycle used is shown in Table 2.6.

For site direct mutagenesis a typical reaction (Table 2.5) was utilised the appropriated recombinant plasmid DNA was used as template. After PCR, the DNA amplified was purified (see Section 2.1.13) and eluted in 30 μl of Elution buffer (EB). A reaction with

10 μl of fragment and 1 μl of *DpnI* (Termo Fisher Scientific) (see Section 2.1.14) in final volume of 20 μl was incubated 2 h at 37 °C to allow the digestion of parental methylated DNA (template in PCR reaction). The reaction was incubated 5 min on ice and 2 μl were transformed into *E. coli* Top10^{TM} cells (see Section 2.1.7).

Table 2.5 Typical PCR

Component	Volume
Novagen® KOD Hot start polymerase[1]	1 μl
10 × Buffer for KOD DNA Polymerase[1]	5 μl
$MgCl_2$ (25 mM)[1]	2.5 μl
dNTPs (2 mM each)[1]	5 μl
Oligonucleotide primer forward (5 μM)	5 μl
Oligonucleotide primer reverse (5 μM)	5 μl
DMSO (100 %)	4 μl
DNA template	~ 100 ng
PCR Grade Water	Up to 50 μl

[1]From Novagen Hot Start KOD polymerase kit.

Table 2.6 Typical PCR conditions

Temperature	Time	Number of cycles	Description
95 °C	2 min	1	Initial denaturation
95 °C	30 sec	35*	Denaturation
55 °C	1 min		Annealing
68 °C	1 min/kb fragment size		Elongation
68 °C	10 min	1	Final Enlogantion
4 °C	∞		Storage

* For site direct mutagenesis PCR was performed only with 16 cycles

2.1.12 *Agarose gel electrophoresis*

To confirm the correct size of the linearized DNA, the separation by size and visualization were carried out by electrophoresis in submerged horizontal agarose gels (Brody and Kern 2004). Agarose gels were made by dissolving (by boiling) 0.4 – 1 g (0.8 – 2 %) of agarose in 50 ml of 1x Tris/Borato/EDTA (TBE) buffer (89 mM Tris-borate

and 2 mM EDTA, pH 8.3). After cooling ethidium bromide (1 μg/ml) was added to allow the DNA visualization by UV light. Gels were cast in mini-gel trays (Applied Biosystems). Once set, the gel was submerged in TBE buffer and the samples were loaded (1 μl of 10x loading buffer (0.25 % bromofenol blue, 50 % v/v glycerol, 10x TBE) was added to 9 μl of DNA). An appropriated DNA standard (Hyperladder™ I or IV, Bioline) was also applied (10 μl). DNA was electrophoresed at 70 volts for 1 hour. DNA fragments migrate through the gel matrix at a rate inversely proportional to the Log 10 of the size of the nucleic acid. So, the size of DNA fragments was determined by comparing their electrophoretic mobility to the known DNA standards.

Gels were visualised using a gel documentation system (Bio-Rad Gel Doc 1000, Molecular AnalystTM/PC windows Software). Photographs were printed with linked Mitubishi Video Copy Processor P68B with Mitsubishi thermal paper. Gel images were archived digitally and in the respective lab book for the duration of this project.

2.1.13 *Purification of DNA fragments*

When the PCR products were pure the purification was performed using a QIAquick PCR Purification Kit according the manufacturer' instructions (QIAGEN). When the PCR fragments were impure and after DNA (vector or insert) digestion the samples were separated by size in an agarose gel electrophoresis (see 2.1.12) and the required band was excised. The DNA was purified using a Gel Extraction Kit (QIAGEN) as described in the manufacturer' instructions.

2.1.14 *Restriction digest of DNA*

The required amount of DNA (1 – 2 μg) was mixed with 10 x concentrated reaction buffer according the manufacturer's (Thermo Scientific) instructions. For a 20 μl reaction 1U of endonuclease was used. Reactions were incubated at 37 °C for 1 h.

2.1.15 *Ligation of insert and vector DNA*

Ligations were performed with digested DNA fragment and vectors containing matched cohesive ends. The T4 ligase kit (Invitrogen) was used according the manufacturer' instructions. The molar ratio of 3:1 insert:vector added per reaction was calculated using the formula:

$$Mass\ of\ insert\ (ng) = 3\ \times \left(\frac{Size\ of\ insert\ (bp)}{Size\ of\ vector\ (bp)}\right) \times Mass\ of\ vector\ (ng)$$

2.1.16 *Automated DNA sequencing*

Automated DNA sequencing was performed by MWG- Biotech (value read service). 100 ng of DNA in a 1.5 ml Eppendorf was sent to the company by postal service. For pET28a the standard sequencing primers T7 (TAATACGACTCACTATAGGG) and T7term (CTAGTTATTGCTCAGCGGT) were selected to perform the sequencing. When necessary, custom primers were designed to amplify the region of interest. The sequencing results were compared to the expected sequences using the online alignment tool MUSCLE (Edgar 2004a; Edgar 2004b).

2.1.17 *Recombinant protein production in E. coli*

To protein production were used the plasmids and cells described in Section 2.1.2. From the transformation plate (see Section 2.1.7) a single colony was picked and grown in 5 ml of LB broth containing the appropriated antibiotic (see Section 2.1.4) for 16 h, with 180 rpm, at 37 °C using an orbital incubator (Sanyo Biomedical). This pre-inoculum was used to inoculate 1 l of LB broth (containing antibiotic). Culture was grown for 4 h at 37 °C until optical density (OD) 0.4-0.8 at 600 nm. Conical baffle flasks two times the volume of the media were used to provide sufficient aeration by rotary shaking (180 rpm). Cultures were cooled and isopropylthio-β-D-galactoside (IPTG) was added to strains carrying a *lacI*^{q} (on plasmids or genome) to induce gene expression controlled by *lacO*. To BL21(DE3) and B834 strains was added 1 mM as final concentration. To Tuner strains an IPTG final concentration of 0.2 mM was used. Induced flasks were incubated 16h at 19 °C (140 rpm). Cells were harvested by centrifugation (see Section 2.1.5) and re-suspended in TALON (50 mM Tris-HCl, 300 mM NaCl, pH8.0) buffer. For 1 l of cell culture was used 10 ml of buffer. This suspension was used to prepare cell free extracts or could be stored at – 20 °C for up to 2 months.

2.1.18 *Preparation of cell free extracts*

Cell suspension containing the target protein was lysed by sonication for 2 minutes using a B. Braun Labsonic U sonicator set at low intensity (~42 watts and 0.5 second cycling). The soluble supernatant (cell free extract) was isolated by centrifugation (see Section 2.1.5) and retained for future use. The insoluble fraction (cell debris) was re-suspended in TALON (50 mM Tris-HCl, 300 mM NaCl, pH8.0) buffer and, when necessary, a sample was visualised by SDS-PAGE (see Section 2.1.20) alongside a

sample of cell free extract to ensure the target protein was present in the soluble fraction.

2.1.19 *Purification of proteins*

2.1.19.1 Immobilized metal affinity chromatography (IMAC)

His-tagged proteins were purified using Immobilised Metal Affinity Chromatography (IMAC). The purification was carried out in TALON buffer (50 mM Tris-HCl, 300 mM NaCl, pH8.0). At this pH histidine, which has a pKa of 6.0, is deprotonated and interacts with electropositive transition metals immobilized in a column. The interaction can be dissociated by competing off the tag with high concentrations of imidazole.

Columns were prepared by allowing 3 cm^3 of TALON^{TM} resin containing cobalt ions (Clontech Laboratories Inc.) settle in a gravity flow column. Column volumes were increased if a high protein yield was expected. Columns were equilibrated with 3 column volumes of TALON buffer. Cell free extracts (5-20 ml) were applied to the column and the eluent was collected. Unless stated otherwise, columns were washed with 10 ml of TALON buffer (remove unbound proteins) and two concentrations of imidazole (10 mM or 100 mM) in TALON buffer (to elute weakly bound proteins prior to tightly bound proteins). The washes eluent was collected and retained for SDS-PAGE analysis. When necessary the imidazole gradient was adjusted in order to increase the protein purity. Each fraction (cell free extract eluent, TALON buffer wash, and imidazole eluent fractions) was analysed by SDS-PAGE (see Section 2.1.20).

2.1.19.2 Ion exchange

Previously to anion exchange purification, the proteins were dialysed (see Section 2.1.22) into 10 mM Tris-HCl, pH 8.0 (Buffer A) at 4 ºC. Proteins purified by Fast Protein Liquid Chromatography (FPLC) using a column HIScreen Capto^{TM} Q (GE Healthcare) in conjunction with an ÄKTA pure 25 chromatography system (GE Healthcare). The colum was washed with five column volume (CV) of 10 mM Tris-HCl buffer, 1M NaCl, pH 8.0 (Buffer B) and equilibrated with two CV of Buffer A. Protein was loaded with a clean loop (1 ml or 40 ml dependent of sample volume). To remove unbounded proteins, the column washed with five CV Buffer A. Protein was eluted with a liner gradient of Buffer B from 0 to 500 mM of NaCl (25 CV) and collected using a fraction collector F9-R (GE Healthcare). If necessary, a step gradient was designed to improve the separation of proteins. In all steps a flow of 2.7 ml/minute was used. Fractions were visualised using SDS-PAGE gels (see Section 2.1.20) and those containing high yields of pure protein were pooled and run in gel filtration chromatography.

2.1.19.3 Gel Filtration chromatography

Proteins were purified by gel filtration chromatography using FPLC with a HiLoadTM 16/60 SuperdexTM 200 Prep grade column (GE Healthcare) in conjunction with an ÄKTA pure 25 chromatography system (GE Healthcare). The column was equilibrated with two CV of buffer (20 mM Tris-HCl, 150 mM NaCl, pH 8.0). Protein was concentrated (see Section 2.1.22) to a volume lower than 5 ml and loaded with a clean loop (1 ml or 5 ml dependent of sample volume). Proteins were separated on the column using the respective buffer (isocratic elution with two CV) and collected using a fraction collector F9-R (GE Healthcare). In all steps a flow of 1 ml/minute was used.

Fractions were visualised using SDS-PAGE gels (see Section 2.1.20) and those containing high yields of pure protein were pooled for further use.

2.1.20 *Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE)*

Visualization of proteins by SDS-PAGE SDS-PAGE as described by (Laemmli 1970) was performed to determine the size and relative purity of samples. The gels were made using the AE-6450 apparatus from ATTO Corporation (Genetic Research Instruments) with 12 cm x 10 cm glass plates sealed by a rubber gasket. The list of required volumes to prepare a 12.5 % acrylamide gel and respective buffers are shown in Table 2.7 and Table 2.8, respectively. The resolving gel was prepared, poured into the plates and covered with water. After polymerization the water was removed and the stacking gel poured on its top and a comb was added. Before use, the comb and rubber seal were removed and the gel affixed within the gel tank, which was filled with running buffer. The loading buffer was added to samples (1:2) and boiled 5 minutes to denature proteins. Samples and a suitable molecular weight standard (SigmaMarker^{TM} wide range, Sigma-Aldrich) were loaded and a current of 35 A (per gel).

After electrophoresis, protein bands were visualized by soaking the gel in InstantBlueTM stain (Expedeon) for at least 15 minutes followed of washing with distilled water. Protein molecular weight was estimated by comparing their electrophoretic mobility with standard. Gels were photographed (Bio-Rad Gel Doc 1000, Molecular AnalystTM/PC windows Software). Photographs were printed with linked Mitubishi Video Copy Processor P68B with Mitsubishi thermal paper. Gel images were archived digitally and in the respective lab book.

Table 2.7 Preparation of SDS-PAGE gel

Component	Reagent	Volume per gel
Resolving gel	0.75 M Tris-HCl, pH 8.8, 0.2 % SDS	2.35 ml
	40 % Acrylamide*	1.45 ml
	d. d. H_2O	0.875 ml
	10 % (w/v) Ammonium persulfate	22.5 µl
	TEMED	7.5 µl
Stacking gel	0.25 M Tris-HCl, pH 8.8, 0.2 % SDS	0.938 ml
	40 % Acrylamide*	0.188 ml
	d. d. H_2O	0.75 ml
	10 % (w/v) Ammonium persulfate	15 µl
	TEMED	5 µl

* BDH Electran acrylamide, 3 % (w/v) bisacrylamide

Table 2.8 SDS-PAGE buffers

Component	Reagent	Volume or concentration
Running Buffer (for 1 l)	32 mM Tris/190 mM glycine, pH 8.3	350 ml
	SDS	0.1 %
Loading buffer (for 10 ml)	SDS	10 % (w/v)
	0.25 M Tris-HCl, pH 8.8	5 ml
	Glycerol	25 % (w/v)
	β-mercaptoethanol	2.5 ml
	Bromophenol blue dye	0.1 % (v/v)

2.1.21 *Quantification of proteins*

The concentration of protein in samples was routinely measures using a NanoDrop 2000 UV-Vis spectrophotometer (Thermo Fisher Scientific) and the Beer-Lambert equation: $A=\varepsilon lC$, where A = absorbance 280 nm, ε = extinction coefficient ($M^{-1}cm^{-1}$), l = length of light path (cm), and C = concentration of sample (M).

2.1.22 *Concentration and buffer exchange of proteins*

Vivaspin® 20 concentrator tubes (20 ml) (Sartorius Stedim Biotech) were used to concentrate protein by centrifugation (see Section 2.1.5). The molecular weight cut-off (10, 30, 50 or 100 kDa) was dependent of the sample to concentrate.

The buffer exchange of proteins could be performed by concentration of the protein and dilution with the required buffer followed of a new concentration step (two times). When necessary PD-10 desalting columns were utilized according manufacture' instructions (GE Healthcare). The buffer exchange of larger volumes was achieved by dialysis. A dialysis tubing (cut-off of 13.5 kDa) was used to contain the sample submerged in 4 l of the appropriated buffer for 16 h at 4 °C (with stirring).

2.1.23 *B. thetaiotaomicron counter-selectable gene deletion or FLAG-Tag*

2.1.23.1 Preparation of plasmids for genomic deletion or FLAG-Tag

To delete a gene in *B. thetaiotaomicron* a plasmid carrying the flanking sequences of the target gene was required. The DNA fragment was obtained by PCR of the homologous region of 1000 base pairs (bp) upstream and 1000 bp downstream of the gene to be removed. A combination of four primers was required to obtain this fragment. Primer Fw1 (upstream) and Rv2 (downstream) were designed as standard PCR primers (see Section 2.1.11) containing the appropriate restriction sites. Internal primers Rv1 and Fw2 were designed selecting the 20 – 25 bp upstream of the start codon and downstream of stop codon and adding this together (Figure 2.1). In the first PCR step two fragments (1000 bp) were generated. For that, two standard PCRs (see Section 2.1.11.1) were performed using primers Fw1 + Rv1 (upstream fragment) and

Fw2 + Rv2 (downstream fragment) (Figure 2.1A). In a second step a sewing PCR allow to join the two fragments in a single construct with 2000 bp (Figure 2.1B). This fragments share some sequence complementarity and they will act as primer and template to each other. The PCR performed with an equal concentration (200 ng) of both fragments and Primers Fw1 + Rv2 allowed the amplification of the final fragment which lacks the target gene (Figure 2.1C). The list of primers utilized to generate the different mutants are listed in Appendix J.

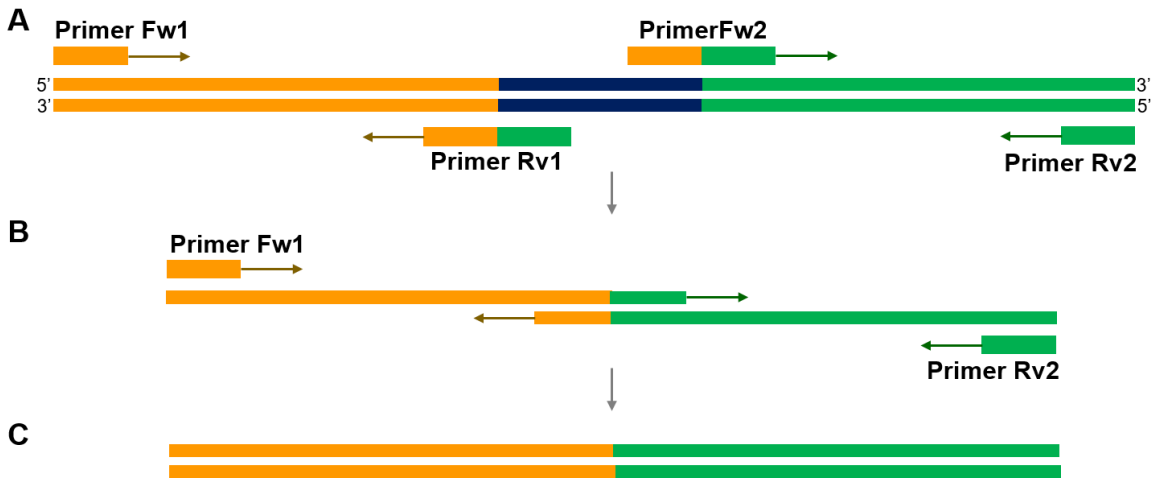


Figure 2.1 Sewing PCR to generate the gene deletion fragment

A. Two independent PCRs are performed in order to obtain an upstream (orange) and a downstream (green) fragments amplified with primers Fw1 + Rv1 and Fw2 + Rv2, respectively. The gene to be detected is represented in dark blue. **B.** PCR sewing with both fragments as template and primers Fw1 + Rv2 generates **C**. a single fragment which lack the target gene to deleted (dark blue). Examples of agarose gels of the different PCR products are presented in Appendix J,

To FLAG-tag a specific gene, the PCR was performed with primers that allow the introduction of the FLAG sequence (DYKDDDDK) at the C-terminal region. In this case, the plasmid carrying the flanking sequence of the C-terminal region was required for introduce the FLAG-Tag in *B. thetaiotaomicron*. A similar approach to the one described to gene deletion was utilized (see above) however the primer design is different (Figure 2.2). Primer Fw1 (upstream) and Rv2 (downstream) where designed as standard primers containing the appropriated restriction enzymes (see Section

2.1.11). Primers Fw1 and Rw2 are 1000 bp upstream or downstream, respectively, of the stop codon of the target protein. Primer Fw2 contains from the 5' to 3': 20 nucleotides upstream to stop codon; the DNA sequence encoding FLAG-tag (DYKDDDDK); stop codon and 20 nucleotides downstream. Primer Rv1 is the reverse complement sequence of Fw2. The primers utilized to generate the FLAG-Tag fragments are listed in Appendix K.

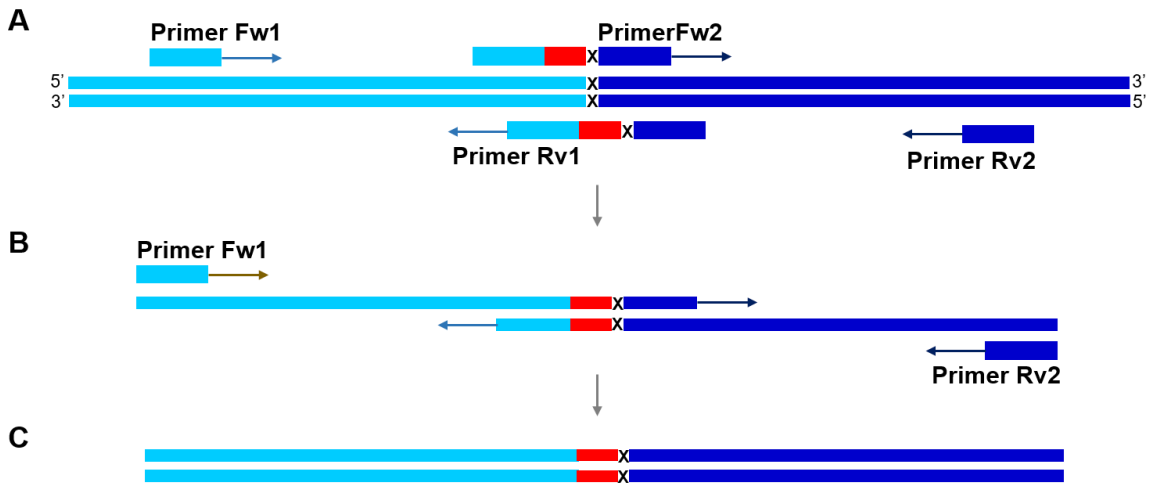


Figure 2.2 Sewing PCR to generate the FLAG-tag fragment

A. Two independent PCRs are performed in order to obtain an upstream (light blue) and a downstream (dark blue) fragments amplified with primers Fw1 + Rv1 and Fw2 + Rv2, respectively. In light blue is represented the target gene. Primers Fw2 and Rv1 were designed in order to introduce a FLAG-tag sequence DYKDDDDK (red) upstream of the stop codon (X). **B.** PCR sewing with both fragments as template and primers Fw1 + Rv2 generates **C**. a single fragment containing a C-terminal FLAG-tag sequence. Examples of agarose gels showing the different PCR products are presented in Appendix K.

The DNA fragments required for gene deletion or FLAG-Tag were cloned into pExchange-*tdk* (see Section 2.1.2) and the resulting plasmids were propagated in *E. coli* CC118 λ pir (see Sections 2.1.8).

2.1.23.2 Conjugation into *B. thetaiotaomicron*

The previously described pExchange-*tdk* plasmids were transformed into S17 λ *pir* *E.coli* cells (see Section 2.1.2), referred to as the "donor" strain. *B. thetaiotaomicron*

tdk- is the “recipient” strain (Figure 2.3A). The donor and recipient strains were cultured (5 ml) for 4 h to roughly equivalent cell densities in LB broth and TYG media, respectively (see Section 2.1.8). Cells were harvested by centrifugation (see Section 2.1.5) and washed once in TYG medium. Equal sized cell pellets were then re-suspended in 1 ml TYG and spread evenly on the surface of BHI plates with no antibiotic (see Section 2.1.4). These plates were incubated agar side down at 37 °C using a static incubator (LTE Ltd) and grown for 16-24 hours. This allowed the formation of a thick layer: *E. coli* grows first, creating an anaerobic environment underneath in which *B. thetaiotaomicron* can thrive, providing the necessary conditions for plasmid conjugation from the donor to the recipient strains. This biomass was scraped from the plate and re-suspended in 5 ml TYG medium. 100 µl of this solution and a 1:10 dilution, both were plated onto BHI + gentamycin + erythromycin plates (see Section 2.1.4). These antibiotics select for the recipient strain and the pExchange-*tdk* plasmid, thus colonies represent single recombinant where the pExchange-*tdk* has recombined with the genomic DNA via one of the flanks (Figure 2.3B). These plates were incubated anaerobically for 2 days, then 10 colonies were picked and re-streaked onto fresh BHI + gentamycin + erythromycin plates to minimise wild type contamination. 10 colonies were cultured anaerobically overnight in TYG medium (see Section 2.1.8), 1 ml of each culture was taken and a pooled stock created**.** The pooled stock and a 1:10 dilution were plated upon BHI + FUdR (see Section 2.1.4) and allowed to grow anaerobically for 2 days. FUdR is toxic to strains able to synthesise thymidine. The recipient strain lacks the *tdk* gene, but this has been complemented within the pExchange-*tdk* plasmid, in this manner FUdR selects for the second recombination event, whereby the second flank incorporates into the genome and the pExchange-*tdk* sequence is eliminated (Figure 2.3C). 10 FUdR resistant colonies were re-streaked onto fresh BHI + FUdR plates to minimise wild-type contamination. These re-streaked

colonies were picked and cultured in 5 ml of TYG so that genomic DNA could be extracted (see Section 2.1.9) and glycerol stocks could be made (see Section 2.1.6).

Isolated DNA was screened by PCR for successful knockout mutations or FLAG-tag genes using PCR downstream and upstream primers used to create the plasmid (primer Fw1 + Rv2, see Section 2.1.23.1), using wild-type *B. thetaiotaomicron* as control. The wild-type strain will produce a fragment which is the length of the target gene (500-2000 bp), plus the length of both flanks (1000 bp each). Any successful knockouts will lack the target gene, yielding a fragment of 2000 bp. The FLAG-Tag gene was screened by PCR using an Fw primer that only hybridize with the FLAG-Tag sequence (FLAG_Fw: gac tac aaa gac gat gac gac aaa) and the upstream primers used to create the plasmid (primer Rv2, see Section 2.1.23.1). The successful FLAG-Tag genes will produce a fragment with 1000 bp and no amplification should be observed with wild-type *B. thetaiotaomicron*. Clones which appeared successful after screening were then sequenced using external primers to ensure the correct mutation had taken place.

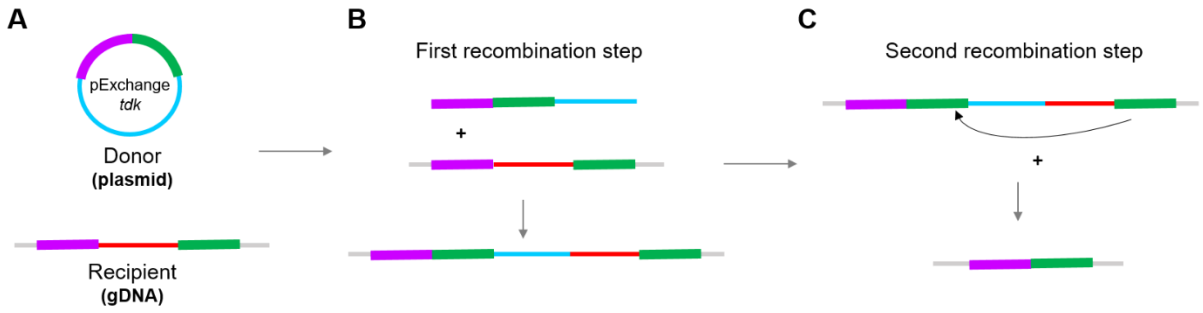


Figure 2.3 Conjugation in *B. thetaiotaomicron*

A. Donor and recipient DNA. The target gene to be removed of FLAG-Tag is represented in red. The upstream and downstream flanks are shown in purple and green, respectively. **B.** pExchange-*tdk* is inserted in the recipient genome by homologous recombination with one of the flanking regions (purple). **C.** pExchange-*tdk* and target gene are eliminated of the recipient genomic DNA by homologous recombination by the second flank (green).

2.1.24 *Growth of B. thetaiotaomicron*

In this study TYG and MM (see Section 2.1.4) were used as culture medium for *B. thetaiotaomicron*. An aliquot of glycerol stock was gently thawed on ice for 5 min. TYG cultures (2.5 – 5 ml) were set up in glass test tubes by inoculation with 1:100 dilution of glycerol stock. Bacterial cultures were grown for 16 h at controlled anaerobic conditions at 37 °C in an anaerobic chamber (Don Whitely Scientific). These cultures were after used to inoculate MM medium. Volumes of 2.5 – 5 ml (glass test tubes) were inoculated with 1:100 dilutions of TYG pre-culture and volumes of 2 ml to 200 µl (plates) were inoculated with 1:25 dilutions of cells. MM cultures were grown in same conditions that TYG cultures. When required Bacterial cultures could be monitored directly throughout growth using a 96 well or 24 well corning® costar® culture plate (Sigma-Aldrich) in conjunction with an Epoch microplate spectrometer (Biotek Instruments Ltd.) inside of an anaerobic chamber. The plate reader measured and recorded the OD at 600 nm of each well at 15 minute intervals. Each well was prepared in triplicate. Data was manipulated in Gen5 2.05 software and the mean and respective standard error of the mean (SEM) were plotted using Prism 6.0 (GraphPad). A control containing media without bacterial inoculum was run to ensure no contamination has occurred throughout the growth period.

2.2 Biochemistry

2.2.1 *Affinity gel electrophoresis (AEG)*

Affinity gel electrophoresis was used to screen or confirm proteins that bound to soluble polysaccharides. A continuous gel system was applied with the same gel apparatus (ATTO Corporation) as described for SDS-PAGE (see Section 2.1.20) however this

method uses native polyacrylamide gels containing soluble ligand sugars. Gels contained 7.5 % (w/v) acrylamide (Acrylogel 3; BDH ElectranTM) in 25 mM Tris, 250 mM glycine buffer, pH 8.3, which comprised the Running and Sample buffers. For ligand-containing gels, appropriate polysaccharides were added at 0.001 - 0.1 % (w/v) final concentration prior to polymerisation. Pure proteins (6 µg) in 7 µl of loading buffer, (5 % (v/v) glycerol and 0.0025 % Bromophenol, Blue), were electrophoresed at 10 mA/gel at room temperature. Gels with and without ligands were run in the same gel box with identical samples loaded on each. BSA (15 µg) was used as a negative, non-interacting control. Proteins were stained and visualised as described for SDS-PAGE (see Section 2.1.20).

2.2.2 *Isothermal titration calorimetry (ITC)*

The thermodynamic parameters of glycan binding to proteins were quantified by ITC using a MicroCalTM VP-Isothermal Titration Calorimeter. Proteins were dialysed into 50 mM HEPES, 200 mM NaCl, pH 7.5 (see Section 2.1.22) and the ligands were dissolved in the same dialysis buffer. Titrations were performed at 25 °C and 307 rpm with 28 injections (at 200 second intervals) of 10 µl of ligand into the cell containing the protein. The concentrations of ligand and protein were optimized for each reaction and are stated in respective legends. Ligand-protein binding results in the release (exothermic) or absorbance (endothermic) of heat, and the machine applies sufficient energy such that the temperature of the reaction cell is maintained relative to a reference cell. The amount of energy used to maintain the temperature of the reaction cell provides a readout of the amount of heat released or absorbed. The thermodynamic parameters, stoichiometry of reaction (*n*), association constant (K_a) and binding enthalpy (ΔH), were calculated from the titration curve using MicroCal

Origin 7.0 software. The polysaccharide concentrations were converted into molarity that gave n=1 (assuming the protein contains a single ligand binding site), to determine the molar concentration of protein binding sites on the polymer. The equation $-RTlnK_a = \Delta G = \Delta H - T\Delta S$, where R represents the gas constant (1.99 cal.K^{-1}.mol^{-1}) and T the absolute temperature, was used to calculate the Gibbs energy change (ΔG) and the entropy change (ΔS). For each dataset the values were estimated from the fitted model.

2.2.3 *Thin layer chromatography (TLC)*

Thin layer chromatography (TLC) uses a solvent system to separate and analyse glycans. Different sugars will migrate with the solvent at different rates; this is due to differences in solubilisation and attraction to the adsorbent layer. Generally, smaller compounds will migrate more rapidly than larger compounds, but this is not always the case (charged sugars migrate rapidly than non-charged sugars). A solution 1-butanol/acetic acid/water at 2:1:1 (mobile phase or solvent) was added to a glass chromatography tank (23 cm x 23 cm x 7.5 cm). The tank covered tightly and vapours were allowed to equilibrate for at least 2 hours prior to use. A foil backed silica plate (Silicagel 60, 20 x 20, Merck) was used as stationary phase. The plate was cut to an appropriate size with a height of at least 10 cm and a line drawn 1 cm from the bottom edge across the plate. A small dot was drawn every 1 cm along this line to serve as a guide to loaded (3-6 μl) of sample. The TLC plate was then left into the tank to allow the samples to migrate up the plate, via capillary action, until the solvent front was 1 cm from the top of the plate. Plates were then removed and dried. Visualisation of the migrated sugars was achieved by immersing the TLC plate in developer (sulphuric acid/ethanol/water 3:70:20 v/v, orcinol 1 %) for 1-2 seconds. Plates were again dried

and developed by heating between 80 and 100 °C. When possible, appropriated standards were run to identify the separated sugars.

2.2.4 *High-performance anion exchange chromatography (HPAEC)*

Sugars were separated by High Performance Anion Exchange Chromatography with Pulsed Amperometric Detection (HPAEC-PAD). In case analysis of unsaturated sugars was used an Ultraviolet Detection (HPAEC-UV). The fully automated system (ICS-3000 gradient pump, detector compartment, detector, auto sampler) had a loop size of 100 µl and a CARBOPACTM PA-1 anion exchange column (Dionex) equipped with a CARBOPACTM PA-1 guard column. The PAD detector settings were E1 = +0.05, E2 = +0.6 and E3 = -0.6. The UV detector data was collected at wavelength 235nm. A gradient of sodium acetate associated or an isocratic elution were used to elute the different sugars. Figure 2.4 shows the two programs and respective solutions used in this study.

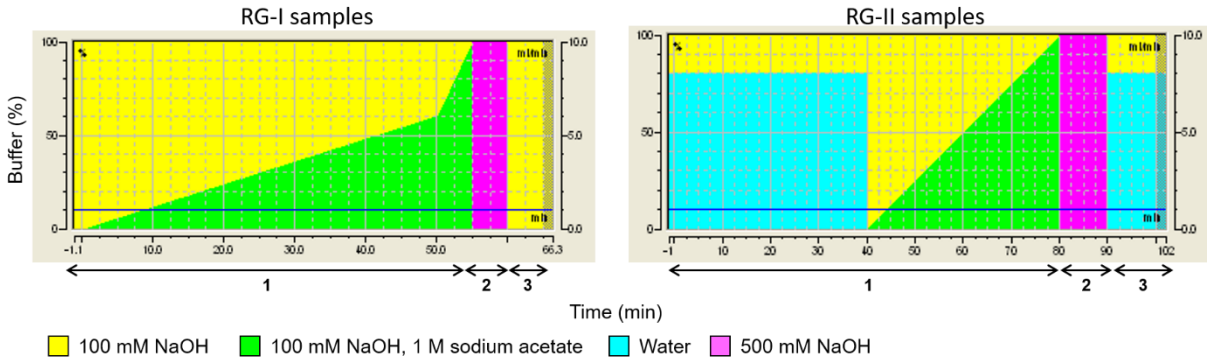


Figure 2.4 HPAEC programs and buffers

Programs used in Chapter 5 (RG-I oligosaccharides, left panel) and Chapter 4 (RG-II oligosaccharides, right panel). The sugars were eluted (1) with a gradient of sodium acetate (RG-I) or an isocratic flow followed by a gradient (RG-II). After elution the column was washed (2) and equilibrated (3) before the next sample injection.

Appropriate standards were run for comparison. A glucose standard (50 µg/ml) was run before and after data collection to control for any loss of detection or variance in elution time. Data were collected and manipulated using ChromeleonTM Chromatography Management System V.6.8 (Dionex). Final graphs were drawn with Prism 6.0 (GraphPad).

2.2.5 *Enzymatic assays*

All enzymatic assays were performed at standard conditions with 1 µM of enzyme, 50 mM sodium phosphate pH 7.5, at 37 °C for 16h (unless stated otherwise).

For catalytic activity assays, all reactions were pre-warmed to 37 °C and initiated by the addition of appropriately diluted enzyme, also pre-warmed to 37 °C. The same set of volumetric pipettes was used throughout. All reactions were performed in at least three technical triplicates.

The values of K_M and k_{cat} were determined, when possible, by using eight substrate concentrations that straddled the K_M using the formula:

$$V_0 = \frac{V_{max}\,[S]}{K_M + [S]}$$

Where V_0 is the initial velocity, V_{max} means number of substrate molecules hydrolysed at infinite substrate concentration, K_M is the concentration at which the reaction rate is half V_{max}, and [S] indicates the substrate concentration. Graphs were plotted in Prism 6.0 (GraphPad) and the kinetic constants were determined by non-linear regression of the plots of V_0 against [S].

When stated, k_{cat}/K_M was determined using a single substrate concentration considerably below the K_M where:

$$V_0 = \frac{k_{cat}}{K_M} \times [S] \times [E]$$

Where k_{cat} is the number of substrate molecules turned over per molecule of enzyme and [E] means enzyme concentration. The validity of the above assumption was tested by carrying out the assay at two different substrate concentrations and comparing the results obtained; they remain constant when both assays obey the above condition.

2.2.5.1 Unsaturated double bond detection at 235 nm

Polysaccharide lyases cleave glycosidic bonds by a β-elimination mechanism generating an unsaturated product containing a double bond between *C-4* and *C-5* of D-galacturonic acid (D-GalA). The double bond absorbs at A_{235nm} and thus its synthesis can be directly monitored. An extinction coefficient (ε) of 4600 M^{-1} cm^{-1} was used to calculate the product concentration. All reactions were performed at 37 °C with quartz cuvettes using a Biotech Ultrospec 4000, UV/vis spectrophotometer. A typical reaction includes the substrate, the enzyme (appropriated diluted), 100-200 mM of respective buffer, 2 mM $CaCl_2$ and sterile water (up to 450 μl). Additionally, the polysaccharide concentration in molar was calculated by adding an excess of enzyme to a low concentration of substrate. The increase of absorbance at 235 nm represent the substrate available bonds for the tested enzyme against a substrate concentration. This value was converted to molar using the $\varepsilon_{\text{double bound}}$ (4600 M^{-1} cm^{-1}) and then used to extrapolate the remain substrate concentrations in molar. For assessing the enzyme pH optimum, the assays were carried out as described with the exception that a range of different buffers were used, corresponding to the pH to be tested (information

provided in the respective experiments). Some pH values could not be overlapped due to inhibition of the enzyme by the buffer.

2.2.5.2 L-Rhamnose and D-galacturonic acid dehydrogenase assay

The release of L-rhamnose and D-galacturonic acid was monitored through the use of a linked assay utilising L-rhamnose dehydrogenase and uronate dehydrogenase, respectively. These dehydrogenases catalyse the oxidation of the L-rhamnose/D-galacturonic acid and reduces NAD^+ to generate NADH in a 1:1 molar ratio. NADH absorbs at A_{340nm} and thus its synthesis can be directly monitored. An extinction coefficient of 6230 M^{-1} cm^{-1} was used to calculate the NADH concentration. All reactions were carried out at 37 °C in the specified buffer. Glass cuvettes were used and NADH release monitored using a Biotech Ultrospec 4000, UV/vis spectrophotometer. Table 2.9 lists the reaction composition for the linked assay.

Table 2.9 Composition of a typical reaction for a linked assay

Component	Volume
Enzyme (appropriately diluted)	50 µl
500 mM Buffer	50 µl
10 mM NAD^+	50 µl
Substrate	50-345 µl
L-rhamnose/uronate dehydrogenase 500 U/ml (1 U = oxidation of 1 µmol galactose min^{-1})	5 µl
Sterile water	50-345 µl

NAD^+ and L-rhamnose/uronate dehydrogenase provided in respective L-Rhamnose or D-Galacturonic acid detection kits (Megazyme).

2.2.5.3 Substrate depletion method by HPAEC

This method was used in Chapter 5.3.3.1 to calculate k_{cat}/K_M for (α-L-Rha-α1,4-D-GalA)$_n$ oligosaccharides where the amount of tetrasaccharide (n=4) and longer

oligosaccharides were limited. In this case, the accurate oligosaccharide quantification by HPAEC (see Section 2.2.4) was used to characterise the catalytic efficiency of BT4145 against various length oligosaccharides. The rate of the reaction can be determined by plotting the amount of substrate remaining over 50 % of the reaction against time using the following equation:

$$k \times t = ln\left(\frac{[S_0]}{[S_t]}\right) \qquad \text{where} \qquad k = \frac{k_{cat}}{K_M} \times [E]$$

t means time and [S_0] and [S_t], represent the substrate concentrations at time 0 and time t respectively. This calculation of k_{cat}/K_M is only valid when [E] << [S] << K_M (Matsui *et al*. 1991). The validity was tested by doubling the substrate concentration and observing an appropriate increase in rate. Enzyme and substrate concentration used in k_{cat}/K_M determinations are given in the appropriated results text.

The assays were performed at 37 °C with D-fructose (0.25 mM) as internal standard. The peak area of each standard sugar determined, and this was used to calculate sample sugar concentrations, thereby eliminating error due to minor fluctuation over time in the detector sensitivity.

2.2.6 *Preparation of substrates and oligosaccharides*

2.2.6.1 Purification of rhamnogalacturonan-I backbone (AM-RG-I)

Arabidopsis thaliana seeds were used to extracted undecorated rhamnogalacturonan-I backbone (AM-RG-I). The method used was as follows: 80 g of seeds were resuspended in 80 ml of distilled water and stirred at ~15 rpm 16 h at 4 °C. The suspension was then centrifuged (see Section 2.1.5) and the supernatant was filtered through a G1 glass filter (pore size: 15–40 μm), Whatman filter paper. The solution as

then dialyzed (see Section 2.1.22) at least twice against 40 volumes of water before freeze drying the glycan. From 80 g of seeds the yield of AM-RG-I was typically 1 g.

2.2.6.2 Production of rhamnogalacturonan-II oligosaccharides

The enzymes acting on rhamnogalacturonan-II (RG-II) indicated an exo-acting model of degradation. By using *B. thetaiotaomicron* strains where the genes encoding specific enzymes involved in RG-II degradation were deleted different oligosaccharides were produced (see Chapter 4.3.3 for more details). The deletion strains were grown on 5 ml MM supplemented with 0.5 % RG-II until late stationary phase (see Section 2.1.24). To identify the oligosaccharides generated by the different mutants, a *B. thetaiotaomicron* wild type (*btWT*) culture was set up in the same conditions. The cultures supernatants of deletion strains and were analysed by thin chromatography plate. The culture was centrifuged 13.000 rpm, 10 minutes and the supernatant was stored at – 20 °C. The generated oligosaccharides were purified by size exclusion chromatography (see Section 2.2.7).

2.2.7 *Purification of oligosaccharides by size exclusion chromatography*

Oligosaccharides were separated by size exclusion chromatography using P2 Bio-gel (Bio-Rad) matrix packed in 3 Glass Econo-Columns™ (2.5 cm × 80 cm) with a flow adaptor (Bio-Rad) at 0.7 ml/min. Columns were equilibrated with degassed 50 mM acetic acid. Fully digested and freeze dried samples were diluted and loaded directly onto the columns, which was then run with 50 mM acetic acid as the mobile phase at 0.5 ml/min using a peristaltic pump (LKB Bromma 2132 microperpex™). The 2 ml

fractions were collected continuously 20 h after loading for 39 h using a Bio-Rad model 2110 fraction collector. The collected fractions were then tested by TLC (see Section 2.2.3) for the presence of oligosaccharides. The fractions containing the pure oligosaccharides were pooled together, freeze dried (see Section 2.2.8) and stored at -20 °C.

2.2.8 *Concentrating of samples by freeze drying*

Samples, such as purified oligosaccharides, were frozen to -80 °C and then lyophilised in a Christ Alpha 1-2 Freeze Drier at -60 °C.

2.2.9 *Mass spectrometry (MS)*

Mass spectrometry (MS) analysis of oligosaccharides and protein presented in this study was performed at Pinnacle (mass spectrometry facilities at ICaMB, Newcastle University).

Oligosaccharide samples were diluted and analysed by liquid chromatography-mass spectrometry (LC-MS). Nanoelectrospray data was acquired using an LTQ-FT mass spectrometer (Thermo).

For protein analysis by MS the sample was run in SDS-PAGE (see Section 2.1.20). The protein band was excised and sent to Pinnacle. After protein extraction from the gel, the sample was trypsin digested and run in Matrix-assisted laser desorption/ionization (MALDI)-MS. The peptide fingerprint was searched in an appropriated base in order to identify the protein(s) present in the tested sample.

2.2.10 *Determination of enzymatic mechanism of catalysis by ^{1}H-NMR*

The analyses of anomeric configuration of the newly created reducing end from enzymatic hydrolysis by proton nuclear magnetic resonance (^{1}H-NMR) as performed at NMR facilities (Newcastle University). Enzyme (BT4145), substrate (α-L-Rha-α1,4-D-GalA) and buffer (20 mM Tris-HCl, 500 mM NaCl, pH 7.5) were freeze dried (see Section 2.2.8) and resuspended in deuterium oxide. To completely remove the water from samples, this step was repeated 3 times. An initial ^{1}H-NRM spectrum of the substrate was obtained (between 4 to 5.5 ppm). The enzyme was added and the course of reaction was followed after 1 minute, 11 minutes and 18.5 h. A standard of L-Rha (20 mM) was run in the same conditions. The peaks for α-L-Rha and β-L-Rha were assigned based on the literature (Pitson *et al*. 1998).

2.2.11 *Assays in different cell context*

Assays in different cell contexts were performed to explore enzymatic activity at the cell surface, in the culture supernatant and in all biological context. In presence of oxygen whole cells are metabolically inactive but retain structural integrity. Thus, proteins which do not require ATP (such as CAZymes) presented at the cell surface remain active and the activity can be observed. The sonicated cells are structural disrupted and the activity of all enzymes present on the cell surface, periplasm and cytoplasm can be explored.

B. thetaiotaomicron 5 ml cultures in MM (see Section 2.1.4) supplemented with polysaccharide were grown to mid-exponential phase (see Section 2.1.24). The cultures were gently harvested by centrifugation (5,000 *g*). The supernatant was

concentrated by centrifugation (see Section 2.1.5) using Vivaspin® 20 ml concentrator tubes (Sartorius Stedim Biotech) with 30 kDa cut-off (to ensure that proteins will be retained during concentration). Supernatant was washed once with 20 ml of Phosphate Buffered Saline Buffer (PBS) and concentrated until 1 ml and stored on ice. The cell pellet was washed twice with 20 ml of PBS to remove the growth media and contaminants. After this washing steps, the pellet was re-suspended in 2 ml PBS and split in two aliquots of 1 ml. One aliquot was kept on ice (whole cells) and the second aliquot was sonicated for 1 minutes using a B. Braun Labsonic U sonicator set at low intensity (~42 watts and 0.5 second cycling). Assays were performed with 500 μl cells or supernatant and 500 μl of 1 % glycan solution in PBS or PBS (negative control). A substrate control with PBS and tested glycan was also performed in the same condition. All reactions were incubated at 37 °C in water bath. To stop this reaction, samples were centrifuged (to remove cells) (see Section 2.1.5) and the supernatant boiled 5 minutes. These samples were analysed by TLC (see Section 2.2.3) and HPAEC (see Section 2.2.4). If necessary, samples were stored at - 20 °C for months.

2.2.12 *Cell localization of B. thetaiotaomicron proteins*

B. thetaiotaomicron strains presenting the targed protein FLAG-tag (see Section 2.1.23.2) were used to explore the cell localization of tagged proteins. A 5 ml *B. thetaiotaomicron* culture was gown in MM with 0.5 % potato RG-I (Megazyme) until mid-exponential phase (see Section 2.1.24). Cells were harvested by centrifugation (5000 *g*), washed in 5 ml PBS and resuspended in 5 ml of PBS. The cells were then split into two 2 ml aliquots that were incubated at 37 °C for 16 h with Proteinase K (Sigma-Aldrich) 2 mg/ml in PBS or PBS (control). In this step Proteinase K degrades the surface proteins. Samples were centrifuged (5000 *g* for 10 min) and the

supernatant discarded. The cell pellets were resuspended in 1ml PBS and the proteins precipitated by the addition of 200 µl trichloroacetic acid 100% (inhibits the Proteinase K) and incubation on ice for 30 min. The precipitated proteins were pelleted by centrifugation (5000 *g* for 10 min) and washed 2 times in 1 ml ice-cold acetone. The protein pellets were resuspended in 400 µl PBS, centrifuged and the supernatant discarded. The treatment with Proteinase K degrades the surface proteins and prevents antibody detection), where periplasmic proteins will not be digested and the antibody detection will be similar to one detected in control sample. The pellets are then resuspended in 50 µl PBS. Samples stored at – 20 °C and used to detect protein by Western blot.

2.2.12.1 Western blot

The FLAG-Tag proteins present in the samples prepared in the previously Section were detected by Western blot. To 5 µl of each sample (Proteinase K treated or control), 5 µl of SDS loading buffer (see Table 2.8) was added. Samples are boiled and loaded on a 10% SDS-PAGE gel (see Section 2.1.20). The gel was run and proteins were transfer to nitrocellulose membrane using a Biorad Protean western blot system. The transfer layers were stacked onto the back side of the cassette as follows: bottom pad, two filter paper, SDS-PAGE gel, nitrocellulose membrane, two filter paper and top pad. The cassette was then closed and loaded into the tank. The transfer buffer (25 mM Tris-HCl, 192 mM glycine, 10% methanol) was added and ran at 80 V at 4 °C for 90 minutes. The membrane was removed from the cassette and blocked for 1 h at room temperature (RT) or 16 h at 4 °C with blocking solution (5 % (w/v) Milk, 0.5 % Tween 20 in PBS). Membrane was incubated with primary antibody rabbit ant-FLAG (Sigma-Aldrich) diluted 1:2000 in blocking solution for 1 h at RT. Membrane was then washed 3 times (for 3 minutes) with 0.5 % Tween 20 in PBS (washing set). Membrane

was then incubated with goat ant-rabbit-HRP (Santa Cruz) at a 1:5000 dilution in blocking solution for 1 h at RT. After a new washing step antibodies were detected by chemi-luminescence using Bio-Rad Clarity™ western ECL substrate (Biorad). BT4661 was used as internal control because it has been shown to be present on the surface of *B. thetaiotaomicron* and displays basal expression (personal communication, Dr Lis Lowe). In this case the primary antibody rabbit anti-BT4661 diluted 1:100 in blocking solution was used.

2.2.13 *Enzyme-linked immunosorbent assay (ELISA)*

Enzyme-linked immunosorbent assay (ELISA) was used to test the binding of $CBM77_{PL1/9}$ or LM19, a rat monoclonal antibody (Verhertbruggen *et al*. 2009), to pectins with different degree of esterification. Polystyrene plates (96-well) were coated with substrate (antigen) (50 µg/ml in PBS) for 16 h at 4 °C. The wells were blocked with 5% (w/v) milk (Sigma-Aldrich) in PBS (Milk/PBS) for 2 h at RT. After washing the wells (nine times) with PBS. Serial dilutions (1:5) of the $CBM77_{PL1/9}$ or LM19 (primary antibody) in 5% Milk/PBS were incubated with the plate. For each substrate, five replicates of same concentrations were used. After 1.5 h at RT the solution was removed and the wells were washed nine times with PBS. The secundary antibody diluted 1:1000 in 5% Milk/PBS was incubated 1.5 h at RT. For $CBM77_{PL1/9}$ and LM19 were used the antibody anti-His-horseradish peroxidase (HRP) (Sigma-Aldrich) and anti-rat IgG (whole molecule)-HRP (Sigma-Aldrich), respectively. Plates were washed 5 times with PBS and the TMB ELISA Substrate (Abcam) was added. This contains 3,3',5,5'-tetramethylbenzidine in a mildly acidic buffer that reacted with peroxidase and a soluble blue reaction product is obtained. After adding the stop solution, a yellow colour is developed which is read at 450 nm

using an Epoch microplate spectrometer (Biotek Instruments Ltd.). Two negative controls with no antigen and no antibody were performed to remove the effect of any unspecific binding. To remove esterification, before the blocking step, the plates were incubated with 0.1 M sodium carbonate pH 11.4 for 1 h at RT. The data was exported to Microsoft Excel and the mean and respective standard error of the mean (SEM) were plotted using Prism 6.0 (GraphPad).

2.2.14 *Plant cell wall immunolabelling using monoclonal antibodies*

Binding to plant cell wall homogalacturonan (HG) was tested in equivalent sections of *Tobacco* stem or *Miscanthus*. The dewaxed sections and LM19 antibody were provided by Prof. Paul Knox (University of Leeds). Section were incubated 1 h with 0.1M sodium carbonate pH 11.4 (to remove esterification). After washing two times, an incubation with a PL enzyme in Tris pH8.0 + 2mM $CaCl_2$ was added to remove the HG (PL concentrations specified in the respective experiments). Sections were washed three times. Sections were incubated with Proteinase K (5 µg/mL in PBS) for 20 minutes. After washing three times the sections were blocked with 5% Milk (Sigma) in PBS for 30 minutes. Sections were rinse off with PBS and incubated with primary antibody LM19 (1:5 dilution) or $CBM77_{PL1/9}$ (0.8 µM) in 5% Milk/PBS for 90 minutes. Sections were washed three times and incubated with the secondary antibody. For sections labelled with LM19 was used the antibody anti-rat-Fluorescein Isothiocyanate (FITC) (Sigma) diluted 1:100 in 5% Milk/PBS. In case of $CBM77_{PL1/9}$ sections an antibody mouse anti-His at 1:1000 was used. The incubation was performed for 60 minutes in the dark as FITC is light sensitive and sections were washed three times. Sections labelled with $CBM77_{PL1/9}$ were incubated with antibody anti-mouse-FITC

(Sigma-Aldrich) at 1:50 dilution in 5% Milk/PBS for 1 hour. To stain cellulose in section, calcofluor (Fluorescent Brightner 28, Sigma-Aldrich) was incubated for 5 minutes at 0.25mg/ml in PBS. Additionally, *Miscanthus* sections were incubated with 0.1% Toluidine Blue O (pH 5.5, 0.2 M sodium phosphate buffer) to remove auto-florescence. Sections were washed and Citifluor in PBS/glycerol (AFI, Agar Scientific) was added to prevent fluorescence fading. The coverslip was careful placed over the section and the slides were kept in the dark at 4°C except for viewing. In this technique, all the incubations were performed at RT and the washes were done with PBS during 5 minutes. Sections were analysed with an Olympus BX-61 microscope equipped with epifluorescence irradiation, and all micrographs were captured with an ORCA 285 camera (Hamamatsu) using Volocity software (PerkinElmer Life Sciences). For each section five micrographs of different regions of pith parenchyma were recorded.

2.2.14.1 Quantification of enzymatic activity against plant cell wall

Tobacco sections were incubated with RfPel1/9 or RfPel1/9+CBM77$_{PL1/9}$ and the pectin was detected by immunolabelling (as described above) with LM19 (monoclonal antibody that specifically binds to HG) (Verhertbruggen *et al*. 2009). The micrographs were analysed with Volocity software. The pixels of each micrograph were count in order to quantify the enzyme effect. For each micrographs analysis the following parameters were set for quantification: find object by intensity, 50-400; exclude objects by size < 5 μM^2; colour objects, uniform, orange; measure objective, by intensity and volume measurements, centroid, perimeter; the different measurements were merged in a single object and a single pixel count was obtained. For each condition tested, the pixel values of four replicates of two independent equivalent sections were plotted

using Prism 6.0 (GraphPad) as mean and respective standard error of the mean (SEM).

2.3 Crystallography

2.3.1 *Preparation of selenomethionine derivatives*

This heavy atom derivative was used to perform SAD experiments. In order to express Selenomethionine derived forms of protein *E. coli* strain B834 (see Table 2.1), a methionine auxotroph, was used. A specific medium was prepared comprising a SelenoMet Medium Base™, a SelenoMet™ Nutrient Mix and a SelenoMethionine Solution (see Table 2.3). *E. coli* strain B834 was transformed with the appropriate plasmid as described in Section 2.1.7. From the transformation plate, a single colony was grown in 5 ml of LB at 37 °C, 180 rpm for 16 h in the presence of the appropriate antibiotic (see Table 2.4). This culture was then inoculated into a 500 ml unbaffled flask containing 100 ml of LB supplemented with the appropriate antibiotic. The culture was grown to an optical density of A_{600nm} = 0.2. Cells were then harvested by centrifugation at 5000 x g for 5 min, the supernatant decanted and the cell pellet re-suspended in 10 ml of SelenoMet Medium Base™, this step was repeated 3 times to remove any trace of LB media. The final suspension of cells was used to inoculate 1 litre of SelenoMet Medium Base™ in a 2 litre baffled flask. Then 50 ml of SelenoMet™ Nutrient Mix was added followed by 4 ml of a Selenomethionine Solution (10 mg/ml). Prior to the addition of the Selenomethionine Solution 10 ml of culture was removed to act as a control; the culture should not grow without the addition of Selenomethionine Solution. Overexpression and purification of the selenomethionine protein was then carried out as described in Section 2.1.17 and 2.1.19, respectively.

2.3.2 *Protein crystallisation*

Proteins solution (native and selenomethionine derivatives) for crystallisation were prepared by IMAC followed by ion exchange and gel filtration purification steps (see Section 2.1.19), unless stated otherwise. Proteins were concentrated (see Section 2.1.22) between 10 mg/ml and 20 mg/ml in water ($CBM77_{PL1/9}$ and BT4170 without ligand) or buffer 10 mM Tris-HCl, 75 mM NaCl, pH8.0 (remain proteins). Protein crystallization was performed by sitting-drop vapour-phase diffusion method (Figure 2.5A) using automated Mosquito^{R} nanodrop dispensing system (TTP Labtech) (Jenkins and Cook 2004). Screens were conducted on MRC 96 well crystallisation plates (Molecular Dimensions) using 100 nl:100 nl and 200 nl:100 nl of protein:crystallisation condition. Different commercially available screening kits were used in initial screens: Index (Hampton Research), JCSG core (I to IV) Suite and $AmSO_4$ Suite (QIAGEN) and, PACT, Morpheus and JCSG Plus (Molecular Dimensions). Co-crystallisations were performed by incubating protein solutions with required ligand concentration prior to crystallisation. Plates were examined using a Leica MZ-6 crystallisation microscope (Leica MicroSystems). Protein cystallization by hanging-drop vapour-phase diffusion method (Figure 2.5B) was performed manually using 24 well Linbro plates (Molecular Dimensions). In this case the drop volumes (μl) were 1:1, 2:1 and 1:2 of protein: crystallisation condition. In this case the crystallisation condition was prepared according with the condition to be optimized by modifying the concentration of precipitant or salt or the pH of the buffer.

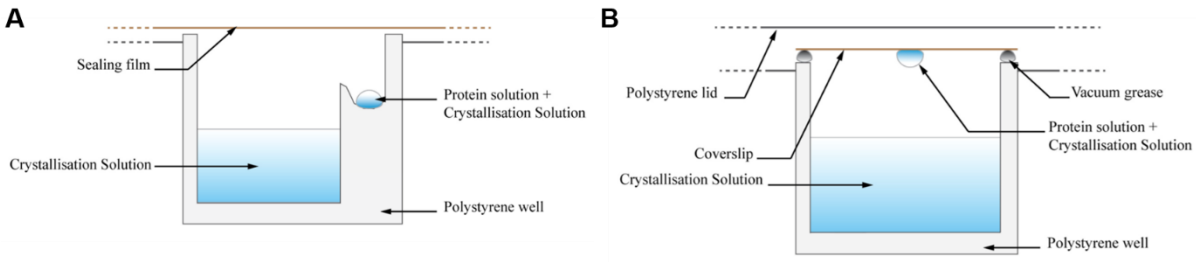


Figure 2.5 Protein crystallization by vapour diffusion method

A. Sitting-drop crystallisation where drops sits upon a shelf above the reservoir containing the crystallisation solution. **B.** Hanging-drop crystallisation where drops are in an inverted coverslip above the reservoir. Figure adapted from (Shapiro 2015).

Crystal handling, data collection, structure solution, model building and validation were performed Dr. Arnaud Basle (ICaMB, Newcastle University).

2.3.3 *Visualization of structures*

Protein structures were visualized and manipulated using PyMOL software. The superimposition of structural homologues was performed using the Secondary Structure Matching tool in software WinCoot. The coordinates were saved and opened in PyMOL software.

2.4 Bioinformatics tools

The different bioinformatics tools used in this study are shown in Table 2.10.

Table 2.10 Bioinformatics tools

Bioinformatic Tool	Server	Description
Prediction of procariotic signal peptides	LipoP 1.0	Sequence base prediction of lipoproteins (signal peptide type II) or periplasmic (signal peptide type I) in Gram-negative bacteria (Juncker *et al*. 2003)
	Signal P 4.1	Prediction of signal peptides in Gram-negative or Gram-positive bacteria (Petersen *et al*. 2011)
Protein parameters	ProtParam	Calculates protein parameters (molecular weight, isoelectric point and extinction coefficient) based on protein sequence (Gasteiger *et al*. 2005)
Primers design	Oligocalc	Calculates oligonucleotide parameters based on single-stand DNA sequence (Kibbe 2007)
	NEBcutter	Find restriction sites in a DNA sequences (Vincze *et al*. 2003)
Genetic and proteomics tools	BLAST	Finds proteins with conserved regions (homologues). The query sequence is compared to a database of non-redundant protein sequences (Johnson *et al*. 2008)
	Pfam 30.0	Database of proteins domains (Finn *et al*. 2016)
	PSI-BLAST	Improved version of the BLAST search that automatically uses the high-score sequences of a normal BLAST to perform a new search until no longer find new sequences (Altschul *et al*. 1997)
	KEGG	Collection of databases of different biological systems used to search *B. thetaiotaomicron* DNA and protein sequences (Kanehisa and Goto 2000)
Phylogenetic tools	MUSCLE	Multiple alignment of protein sequences (Edgar 2004a; Edgar 2004b)
	ESPript 3	Used to visualize the MUSCLE alignments and generate the alignment figures (Robert and Gouet 2014)
	Phylogeny.fr	Drawing of phylogenetic trees (Dereeper *et al*. 2008; Dereeper *et al*. 2010)
Structures	PDB	Data base of structures used to access the coordinates of the several structures presented in this study (www.rcsb.org) (Berman *et al*. 2000)
	PDBsum	For analysis of protein secondary structure and to generate the list of interactions proteins with ligands or ions (de Beer *et al*. 2014)
	DaliLite v. 3	Identification of structural homologues (Holm and Rosenstrom 2010)

(BLAST) Basic Local Alignment Search Tool; (PSI-BLAST) Position-Specific Iterated BLAST; (KEGG) Kyoto Encyclopedia of Genes and Genomes; (MUSCLE) Multiple Sequence Comparison by Log-Expectation; (ESPript) Easy Sequencing in PostScript; (PDB) Protein Data Bank.

Chapter 3. Characterization of a novel carbohydrate binding module

3.1 Introduction

Efficient plant cell wall degradation requires a close association between catalytic modules [glycoside hydrolases (GHs) or polysaccharide lyases (PLs)] with non-catalytic carbohydrate binding modules (CBMs). Pectin is one of the major plant cell wall components and is characterized by a high content of the acidic sugar D-galacturonic acid (D-GalA). Recent studies have contributed to a better understanding of how these glycans are degraded (reviewed by Senechal *et al*. 2014; Silva *et al*. 2016). There is, however, a paucity of information on the contribution of CBMs to pectin degradation. Of the 74 CBM families listed in the CAZy database (prior to publication of the paper associated with this thesis chapter (Venditto *et al*. 2015) only two, CBM32 and CBM35 have members with affinity for pectins (Abbott *et al*. 2007; Montanier *et al*. 2009; Lombard *et al*. 2014). The degradation of these highly complex glycans requires a combination of several different enzymes but, due to the lack of characterized CBMs that bind to pectins, the role of these non-catalytic modules in this process is unknown.

In several anaerobic bacteria the highly efficient degradation of recalcitrant polysaccharides is achieved through a multi-enzymatic complex, the cellulosome (reviewed by Bayer *et al*. 2004). The cellulolytic bacterium *Ruminococcus flavefaciens* FD-1 synthesises the most complex cellulosome described so far. It was predicted that this microorganism encodes 225 proteins containing dockerin domains, a signature for cellulosome incorporation (Berg Miller *et al*. 2009). Unexpectedly, a large number of the modules present in dockerin-containing proteins are of unknown function. These

protein modules may be new families of plant cell wall degrading enzymes (GHs and PLs) or CBMs.

To understand the contribution that CBMs play in glycan degradation by *R. flavefaciens*, Professor Fontes and colleagues at the Technical University of Lisbon, Portugal, identified several novel CBM families present in the cellulosome of this microorganism (Venditto *et al*. 2016). Briefly, 237 genes encoding cellulosomal proteins or protein modules with unknown function were cloned and expressed in *Escherichia coli* and purified using high-throughput technology. From the initial 237 protein targets it was possible to purify 177 in soluble form. Using high-throughput technology, these proteins were screened against a glycan library printed microarray. This technique allowed the identification of six putative novel CBM families, including one with apparent affinity for pectins (defined as $CBM77_{PL1/9}$ and the founding member of family CBM77) (Figure 3.1).

3.2 Objectives

The objectives of this chapter are:

i) The functional and structural characterization of $CBM77_{PL1/9}$, a representative member of CBM77

ii) Explore the *in vitro* and *in vivo* role of $CBM77_{PL1/9}$ in pectin degradation

iii) Investigate the use of the pectin binding $CBM77_{PL1/9}$ as a probe to target homogalacturonan in the context of plant cell wall

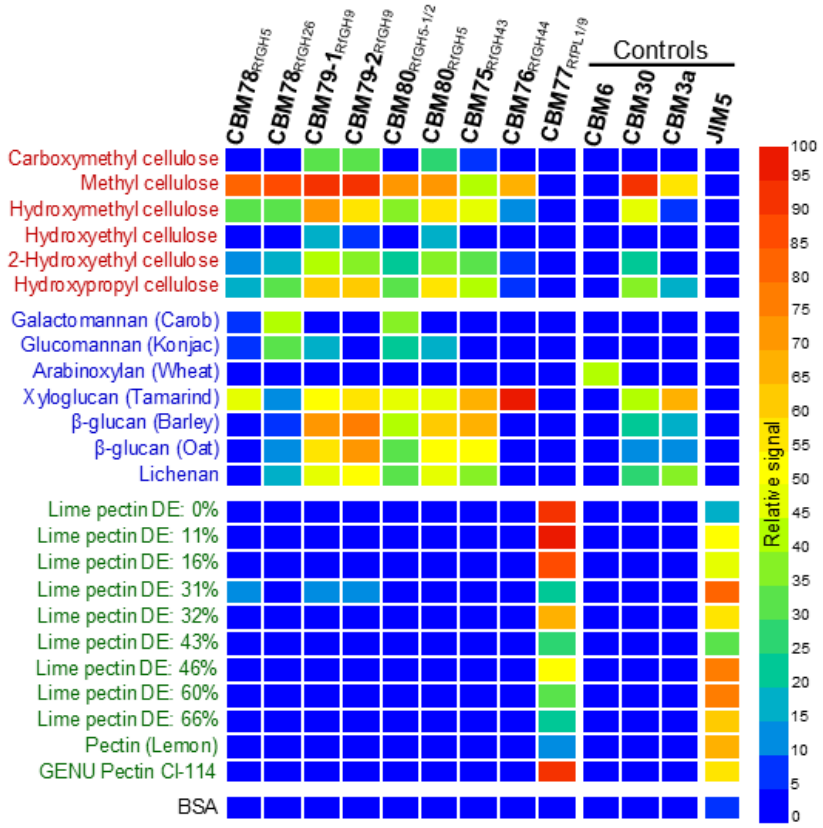


Figure 3.1 Glycan microarray binding profiles of six new CBM families

Controls were characterized CBMs (CBM6, CBM30 and CBM3a) or JIM5, a monoclonal antibody that recognizes homogalacturonan (Venditto *et al*. 2016).

3.3 Results

3.3.1 *Cloning, expression and purification*

$CBM77_{PL1/9}$ is of a component of a cellulosomal (dockerin containing) protein that contains two PLs, RfPel1 and RfPel9, belonging to PL1 and PL9, respectively. To explore the impact of $CBM77_{PL1/9}$ on pectin degradation truncated forms of the enzyme were designed (Figure 3.2).

Figure 3.2 Molecular architecture of different constructs used in this study

A. CBM77$_{PL1/9}$ containing protein presents an N-terminal signal peptide (SP) and two catalytic modules, belonging to PL1 module (orange) and PL9 subfamily 1 (blue). CBM77 module (green) is flanked by two small linker sequences (grey line). A C-terminal dockerin (Doc) module (purple) allows the attachment of the protein to a cellulosomal scaffoldin. The boundaries of the modules in the full-length sequence of the enzymes are indicated. **B.** Different constructs designed in this study: *CBM77$_{PL1/9}$*, CBM77 module; *RfPel1/9+CBM77$_{PL1/9}$*, full length protein without SP and dockerin sequences; *RfPel1*, PL1 module; *RfPel1+CBM77$_{PL1/9}$*, PL1 construct followed by a linker and CBM77$_{PL1/9}$; *RfPel9*, represents the PL9_1 sequence; *RfPel9+CBM77$_{PL1/9}$*, proteins PL9 and CBM77$_{PL1/9}$ separated by a linker.

R. flavefaciens genomic DNA was used as the template to amplify the DNA sequences using PCR. The primer engineered restriction sites are listed in Table 3.1 and an example of an agarose gel with the PCR products is presented in Figure 3.3. DNA inserts were cloned into pET28a using the restriction sites *NcoI*/*XhoI*, except for *rfpel1*, which was cloned into *NcoI*/*BamHI*. *rfpel1+cbm77$_{PL1/9}$* was obtained by amplifying the DNA sequence encoding *cbm$_{PL1/9}$* and the upstream linker. This DNA was ligated into *rfpel1* vector digested with *BamHI*/*XhoI* (Table 3.2). All the proteins generated from these constructs contain a C-terminal His$_6$-tag encoded by the vector.

Table 3.1 Primers designed to clone different constructs utilized in this chapter

Primer name	Primer sequence (5' to 3')	Restriction site
CBM77$_{PL1/9}$_Fw	cac acc atg ggc cct gta gca gga aac tat g	*NcoI*
CBM77$_{PL1/9}$_Rv_XhoI	cac act cg agt gct gca tat tcc atg tag	*XhoI*
RfPel1_Fw_NcoI	cac acc atg ggc gca cac ttt cag gca c	*NcoI*
RfPel1_Rv_ BamHI	cac agg atc cca att atc ttc ctt aaa acc gc	*BamHI*
RfPel9_Fw_NcoI	cac acc atg ggc gcc gca cctg ttac ag	*NcoI*
RfPel9_Rv_XhoI	cac act cga gag tga ggc ttg tga agt g	*XhoI*
CBM77$_{PL1/9}$_Linker_Fw_BamHI	cac agg atc cac caa gca cac ctg tta caa c	*BamHI*

Restriction sites are shown underlined; Fw (forward primer); Rv (reverse primer).

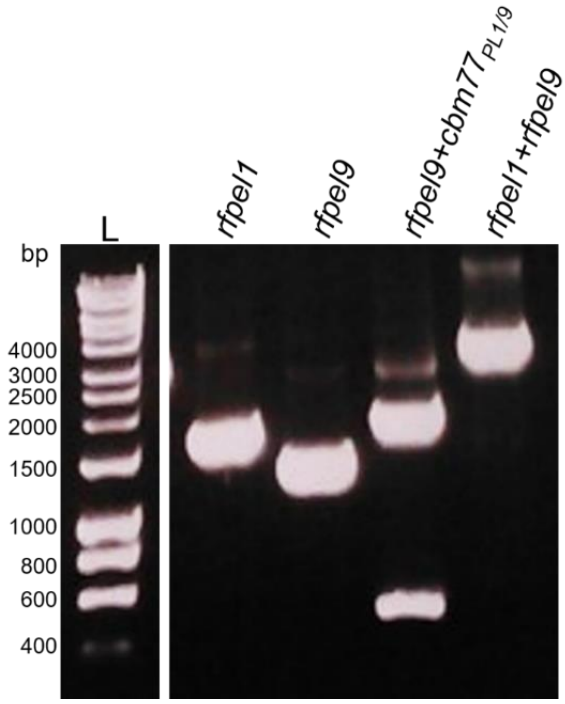


Figure 3.3 Example of PCR-derived fragments of different constructs analysed by agarose gel electrophoresis

After the PCR (Chapter 2.1.11), 5 µl of products with 2 µl of loading dye were electrophoresed in a 1% agarose gel (Chapter 2.1.12). The respective DNAs were labelled and L represents the size marker in base pairs (bp). The expected sizes were 1290 bp (*rfpel1*), 1290 bp (*rfpel9*), 1746 bp (rfpel9+cbm77$_{PL1/9}$) and 2998 bp (*rfpel1/9*).

Table 3.2 Details of the different constructs cloned into pET28a

Construct name	Primers used to amplify the DNA fragments	Cloning enzymes	Cloning vector
$CBM77_{PL1/9}$	$CBM77_{PL1/9}$Fw + $CBM77_{PL1/9}$_Rv_XhoI	*Nco I/ Xho I*	pET28a
RfPel1/9+$CBM77_{PL1/9}$	RfPel1_Fw_NcoI + $CBM77_{PL1/9}$_Rv_XhoI	*Nco I/ Xho I*	pET28a
RfPel1	RfPel1_Fw_NcoI + RfPel1_Rv_ BamHI	*NcoI/ BamHI*	pET28a
RfPel1+$CBM77_{PL1/9}$	$CBM77_{PL1/9}$_Linker_Fw_BamHI + $CBM77_{PL1/9}$_Rv_XhoI	*BamHI/ XhoI*	RfPel1
RfPel9	RfPel9_Fw_NcoI + RfPel9_Rv_XhoI	*Nco I/ Xho I*	pET28a
RfPel9+$CBM77_{PL1/9}$	RfPel9_Fw_NcoI + $CBM77_{PL1/9}$_Rv_XhoI	*Nco I/ Xho I*	pET28a
RfPel1/9	RfPel1_Fw_NcoI + RfPel9_Rv_XhoI	*Nco I/ Xho I*	pET28a

Fw (forward primer); Rv (reverse primer).

To explore $CBM77_{PL1/9}$ as a probe to target pectin, a dimer of this protein was used. The gene containing two $CBM77_{PL1/9}$ molecules connected by a linker was chemically synthesised (see protein sequence in Appendix B, Figure B.1). The gene was designed to contain the restriction sites *NheI*/*XhoI* and a stop codon after the *XhoI* restriction site and the codons were optimized to *E. coli* protein expression. This allowed the further cloning of the dimer gene into pET28a in fusion with a N-terminal His_6-tag.

Recombinant proteins were expressed in *E. coli* strain TUNER and purified by immobilised metal ion affinity chromatography (IMAC). When IMAC purification was not sufficient to obtain proteins with the required purity, size exclusion chromatography was also performed. The purity of protein samples was assessed by SDS-PAGE. Different examples of SDS-PAGE analysis of protein expression and purification are shown in Figure 3.4.

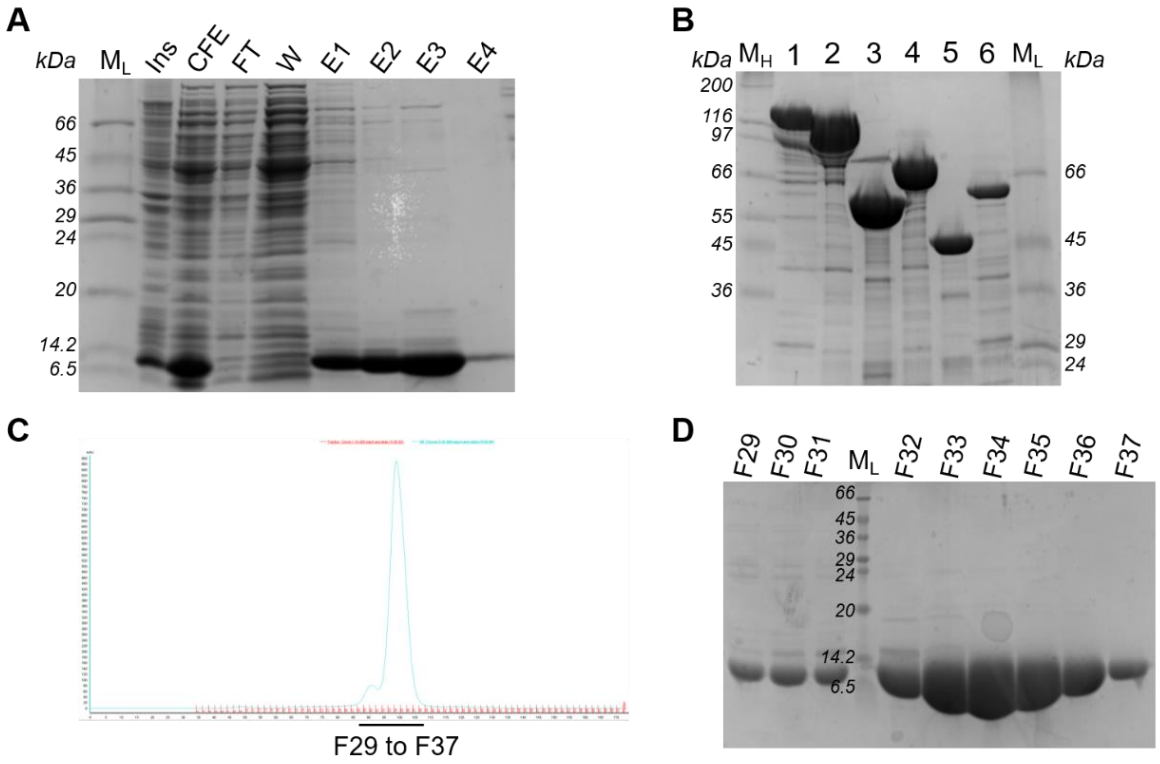


Figure 3.4 Examples of SDS-PAGE gels of IMAC and gel filtration purifications

A. Purification of CBM77$_{PL1/9}$ by IMAC (Chapter 2.1.19.1). The protein expected size is 12.6 kDa; Ins – Insoluble fraction; CFE – Cell free extract; FT – Flow through; W- Wash with Talon buffer; E1 and E2 – protein elution with 10 mM imidazole in Talon buffer; E3 and E4 – elution with 100 mM imidazole in Talon Buffer (Chapter 2.1.19.1). **B**. SDS-PAGE (Chapter 2.1.20) of the proteins purified by IMAC and gel filtration: 1 – RfPel1/9+CBM77$_{PL1/9}$ (123 kDa); 2 - RfPel1/9 (107 kDa); 3 - RfPel1 (59 kDa); 4 - RfPel1/9+CBM77$_{PL1/9}$ (74 kDa); 5 - RfPel9 (47 kDa); 6 - RfPel9+CBM77$_{PL1/9}$ (62 kDa). **C.** Example of a gel filtration chromatogram (Chapter 2.1.19.3) for CBM77$_{PL1/9}$ showing the protein elution peak. **D.** Gel of CBM77$_{PL1/9}$ purification by gel filtration. F29 to F37 correspond to different factions eluted (see Panel C). M_L and M_H correspond to different molecular weight (kDa) markers (L – low and H – high).

3.3.2 *CBM77$_{PL1/9}$ binds to pectins with low degree of esterification*

The microarrays allowed the identification of a putative novel CBM displaying affinity for pectin (Figure 3.1). To confirm the microarray data, binding was assessed by affinity gel electrophoresis (AGE) and isothermal titration calorimetry (ITC). In AGE the equivalent amount of protein was applied to control and different ligand containing gels (Figure 3.5). CBM77$_{PL1/9}$ showed a slow migration in the control gel, which made the results difficult to interpret. However, the protein appeared to display a retardation in

electrophoretic migration compared to the control gel containing no pectin. This suggests that CBM77$_{PL1/9}$ may be binding to the pectin polysaccharides tested.

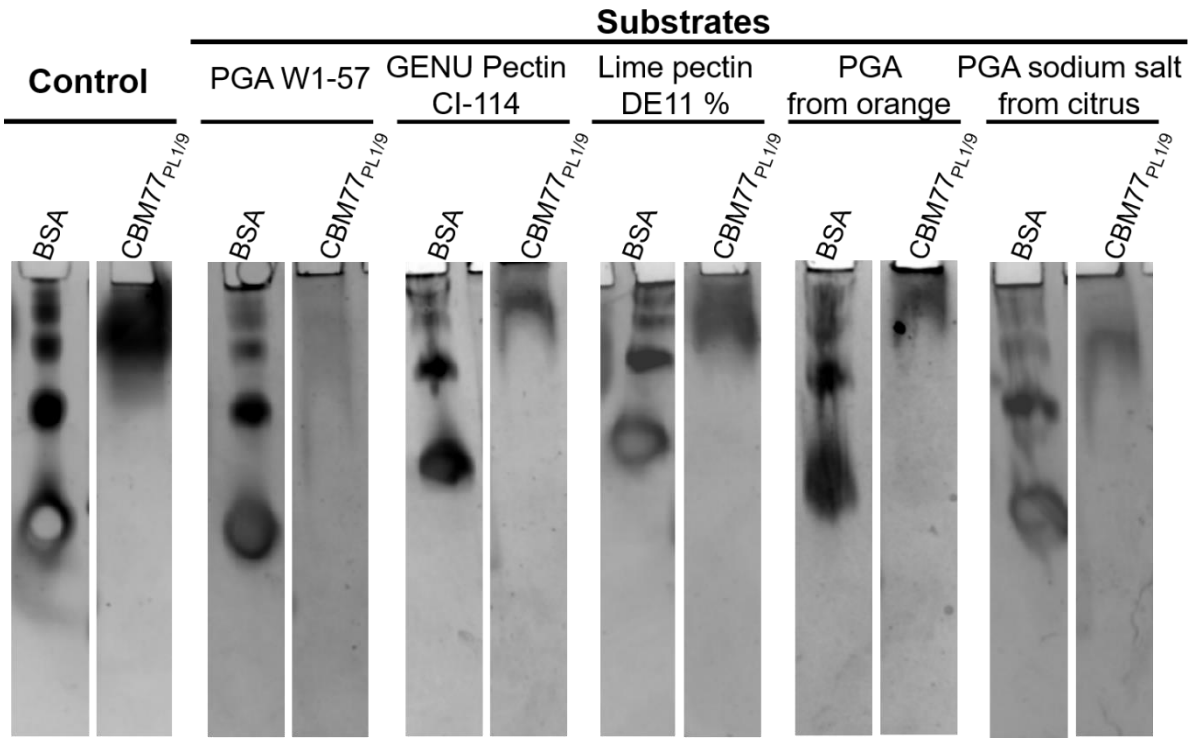


Figure 3.5 CBM77$_{PL1/9}$ affinity gel electrophoresis

AGE (Chapter 2.2.1) of CBM77$_{PL1/9}$ was performed in non-denaturing conditions using 10% gels. Ligand gels contained 0.2% polysaccharide, while the control gel lacked polysaccharide. BSA and CBM77$_{PL1/9}$ (5 µg) were applied to all gels. (PGA) Polygalacturonic acid; (DE) Degree of esterification.

The thermodynamic parameters of the binding of CBM77$_{PL1/9}$ to its ligands were assessed by ITC and are shown in Table 3.3. Examples of ITC titrations are represented in Figure 3.6 The data confirmed the binding of CBM77$_{PL1/9}$ to different pectins. CBM77$_{PL1/9}$ displayed highest affinity for polysaccharides with a low DE such as lime DE 11% and citrus pectin DE 30% (K_A 1.8 x 10^5 M^{-1} and 1.1 x 10^4 M^{-1}, respectively). The affinity for GENU pectin CI-114 (K_A 4.2 x 10^5 M^{-1}) was similar to lime pectin. However, this polysaccharide was provided by Professor William G. Willats and the DE was not specified. The binding module displayed a reduction in affinity as the DE of the ligand increased. Indeed CBM77$_{PL1/9}$ exhibited no binding to polysaccharides with a DE ≥80%. These data are consistent with the microarrays where increased

ligand esterification led to a marked reduction in the binding of CBM77$_{PL1/9}$ (Figure 3.1). CBM77$_{PL1/9}$ displayed optimum binding to oligosaccharides of α1,4-D-GalA with a degree of polymerization (DP) of 7 to 8. The presence of a chelating agent (EDTA) in ITC titrations did not influence affinity indicating a metal independent binding mechanism. This is in contrast with pectate lyases, where calcium plays a important role in substrate recognition (Herron *et al*. 2003; Garron and Cygler 2010). No binding was observed to AM-RG-I (contains the backbone of rhamnogalacturonan-I, a repeating unit of α1,2-D-GalA-α1,4-L-Rha) revealing that CBM77$_{PL1/9}$ binds specifically to homopolymers of D-GalA.

Table 3.3 Thermodynamic parameters of the binding of CBM77$_{PL1/9}$ to different ligands

Ligand	K_A	ΔG	ΔH	TΔS	n
	M^{-1}	*kcal $mole^{-1}$*	*kcal $mole^{-1}$*	*kcal $mole^{-1}$*	
Lime Pectin DE 11% (0.05%)	1.8 (±0.3) x 10^5	-7.11	-20.7 (±0.7)	-13.54	0.95
GENU pectin CI-114 (0.1%)	4.2 (±0.2) x 10^5	-7.64	-17.3 (±0.2)	-9.66	0.95
PGA from orange (0.4%)	1.2 (±0.0) x 10^4	-5.56	-15.0 (±0.5)	-9.45	1.0
PGA from orange (0.4%) + EDTA (5 mM)	1.1 (±0.0) x 10^4	-5.53	-11.0 (±1.3)	-5.49	1.0
Citrus pectin DE 30% (1%)	1.1 (±0.0) x 10^4	-5.49	-7.6 (±0.5)	-2.06	1.0
Citrus pectin DE 60% (0.4%)	Binding too weak to quantify				
Citrus pectin DE ≥ 80% (0.4%)	No binding				
D-GalA DP3/DP4 (10 mM)	Binding too weak to quantify				
D-GalA DP7/DP8 (6.3 mM)	1.2 (±0.2) x 10^5	1.02	-3.0 (±0.2)	-4.03	1.0
AM-RG-I $(Rha\text{-}GalA)_n$ (0.2%)	No binding				

The errors correspond to the fitting of the ITC curve using the software MicroCal Origin7. The thermodynamic parameters were as follows K_A is the association constant, ΔG is the change in free enthalpy, ΔH is the enthalpy of binding, T is the absolute temperature and ΔS is the entropy of binding. For details see Chapter 2.2.2. ITC titrations were performed with 50 µM CBM77$_{PL1/9}$ (cell) and an adequate concentration of substrate (syringe) that allowed the complete titration of the protein (showed in brackets in the table). The titrations were performed in 50 mM HEPES, 200 mM NaCl pH 7.0. (DE) Degree of esterification; (PGA) Polygalacturonic acid; (DP) Degree of polymerization, (Rha) L-rhamnose; (D-GalA) D-galacturonic acid.

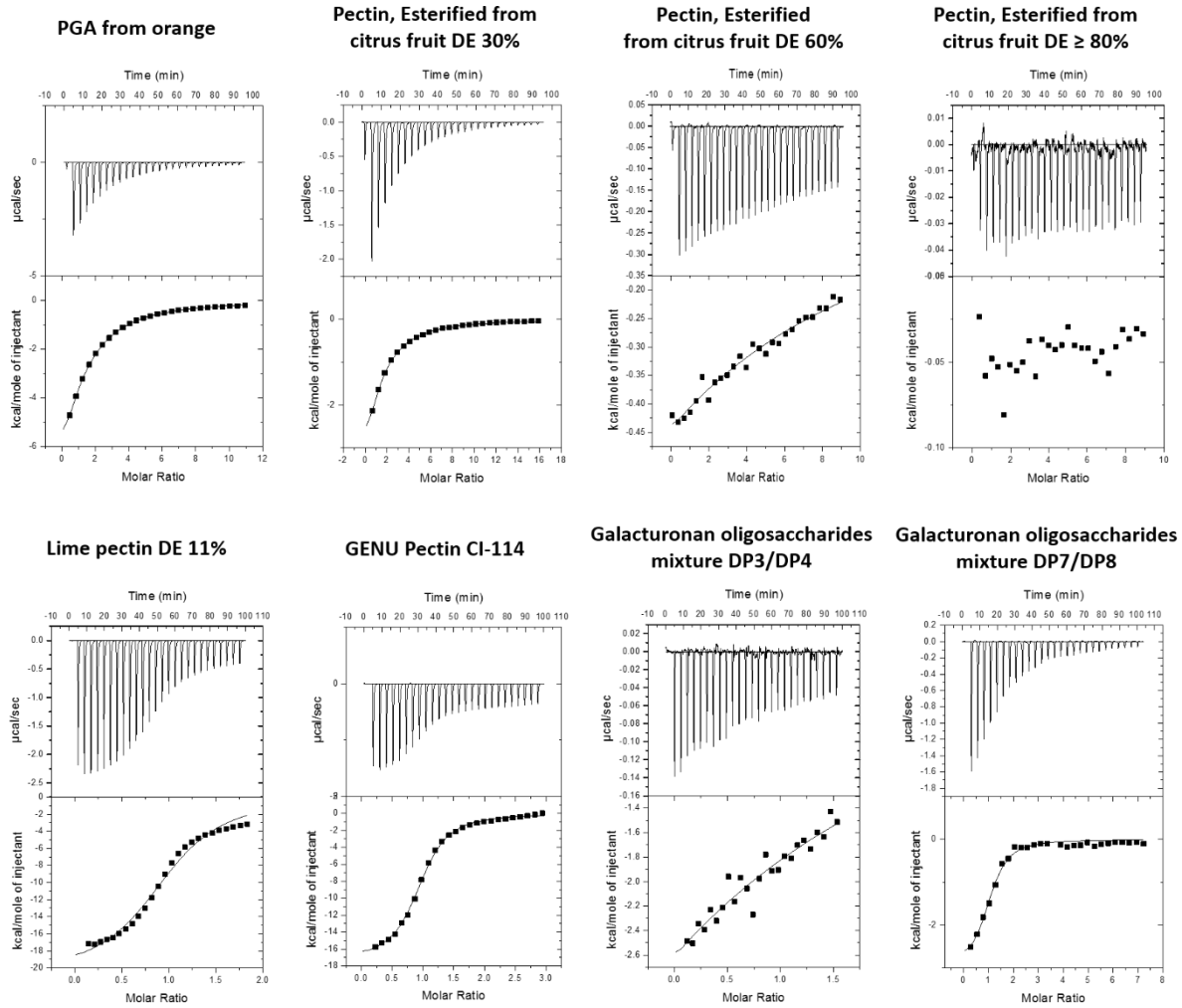


Figure 3.6 Examples of the ITC titration curves of CBM77$_{PL1/9}$ against different pectins

ITC titrations were performed as described in Chapter 2.2.2. For details see legend on Table 3.3. The effect of heat of dilution (release of heat when ligand is titrated into buffer) was removed in each experiment. The top part of each panel corresponds to raw ITC data and the bottom part represents the integration of the peak areas, fitted using the one site binding model using MicroCal Origin7 software. (PGA) Polygalacturonic acid; (DE) Degree of esterification; (DP) Degree of polymerization.

3.3.3 *The structure of CBM77$_{PL1/9}$*

To understand the mechanism of ligand recognition CBM77$_{PL1/9}$ was crystallised and the structure was solved by Dr. Arnaud Basle (ICaMB, Newcastle University).

3.3.3.1 Protein crystallisation and structure

$CBM77_{PL1/9}$ was crystallized by sitting-drop vapour-phase diffusion method using the automated Mosquito^{R} nanodrop dispensing system (see Chapter 2.3.2). An equal volume (100 nl) of protein and reservoir solutions were mixed and crystals were grown at 20 ºC. $CBM77_{PL1/9}$ hexagonal crystals were obtained with 17 mg/ml in 0.1 M citric acid pH 3.5, 2 M ammonium sulphate after 16 h (Figure 3.7A). The fishing of crystals, data collection and structure solution and refinement were performed by Dr. Arnaud Basle (ICaMB, Newcastle University). Briefly, crystals were frozen in saturated ammonium sulphate solution. Crystallographic data were collected in-house (Rigaku Micromax 007) or at Diamond Light Source Ltd, UK, beamline I04-1. The phase problem was solved in-house by the SAD-sulphur method to generate the model that was further used in molecular replacement (Sarma and Karplus 2006). The structure of $CBM77_{PL1/9}$ was solved at 1.5 Å. Crystal data collection statistics and refinement are reported in Appendix C (Table C.1).

$CBM77_{PL1/9}$ revealed a β-sandwich fold, typical of the majority of CBMs, (Figure 3.7B). The structure contains 13 antiparallel β-strands organized in two β-sheets (defined as 1 and 2) and a small helix. β-sheet 1 (represented in dark blue in Figure 3.7C) comprises four β-strands (β-1, β-8, β-11 and β-13), while β-sheet 2 (in green, Figure 3.7C) includes five β-strands: β-2, β-7, β-9, β-10 and β-12. These two β-sheets are connected by short loops. A small α-helix (in orange, Figure 3.7) is present between β-strands 1 and 2. Additionally, four small anti-parallel β-strands (light blue, Figure 3.7C) β-3, β-4, β-5 and β-6 were found between β-strands 2 and 7.

The structural alignment program DaliLite version 3 (Holm and Rosenstrom 2010) revealed that $CBM77_{PL1/9}$ displayed same structure similarity to members of families CBM35 and CBM22 (Table 3.4). The sequence similarity between $CBM77_{PL1/9}$ and

these structural homologs, however, is very low (9-5%) and there is no conservation in the ligand binding residues (see below).

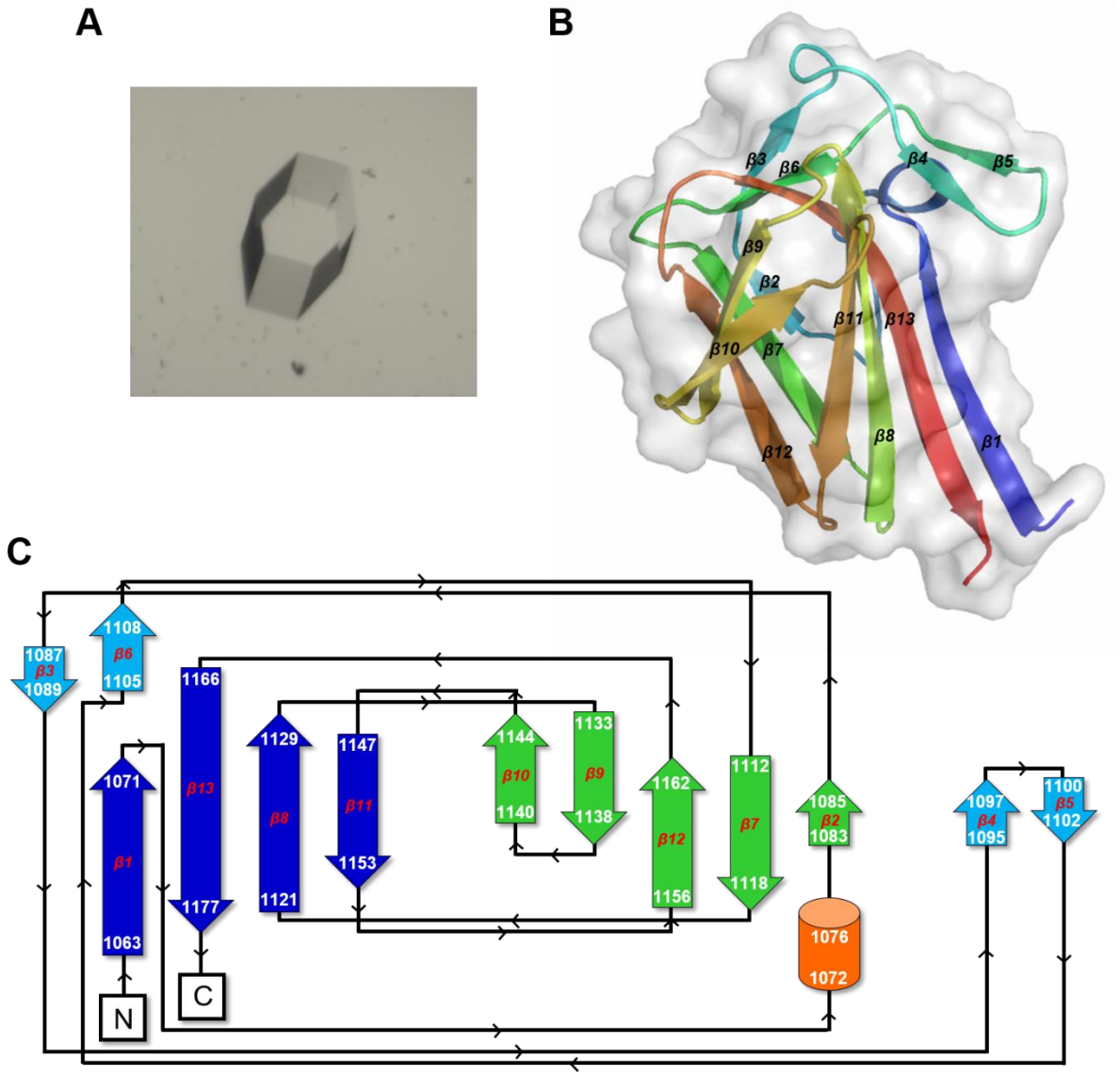


Figure 3.7 CBM77$_{PL1/9}$ crystal structure

A. The crystal of CBM77$_{PL1/9}$ was obtained in 0.1 M citric acid pH 3.5, 2 M ammonium sulphate. **B**. Cartoon of CBM77$_{PL1/9}$ tertiary structure colour ramped from blue (N-terminal) to red (C-terminal), embedded in the surface representation of the protein. Panel was created utilizing PyMOL. **C.** Schematic representation of CBM77$_{PL1/9}$ secondary structure determined using the online software PDBsum (de Beer *et al*. 2014). β-sheets 1 and 2 are represented in dark blue and green, respectively. α-helix is shown in orange and additional small β-strands are coloured in light blue. The β-strands and α-helix amino acid numbers corresponding to the CBM77PL1/PL9 sequence are shown in the figure. In Panels B and C, the β-strands are numbered from the N-terminal (β1) to the C-terminal (β13).

Table 3.4 Structural homologs of CBM77$_{PL1/9}$

CBM family of homologue	PDB code	Z-score	RMSD (Å)	Residues aligned	% identity
CBM35	3ZM8	8.1	3.0	96	8
CBM35	2W87	8.0	2.9	99	9
CBM22	2WYS	7.2	2.9	99	5

Z-score (measure of similarity where a value above 2 means a significant similarity and usually corresponds to a similar fold); (RMSD) Root mean square deviation; (% identity) percentage of identity for the structurally aligned region.

3.3.3.2 Probing the location of the ligand binding site in CBM77$_{PL1/9}$

Site-directed mutagenesis was performed to try to identify the ligand binding site in CBM77$_{PL1/9}$. The primers used to generate the different mutants are shown in Appendix D (Table D.1) and an example of the agarose gel electrophoresis of the PCR products is presented in Figure 3.8. The rationale for the design of the mutants was as follows: 1) Usually, the CBM binding is mediated by aromatic residues that establish hydrophobic interactions with the ligand sugar rings (Gilbert *et al*. 2013). Thus, the two aromatic residues exposed at the protein surface, Y1097 and Y1170, were mutated (Figure 3.9A). 2) Additionally, due to the acidic character of the substrate, the positively charged solvent exposed amino acids (lysines) were also substituted with alanine (Figure 3.9A). Two of these lysines, K1100 and K1123, were presented at the concave surface close to the aromatic residues; the remaining lysines (K1092, K1107, K1136, K1141 and K1162) were grouped on the opposite convex surface.

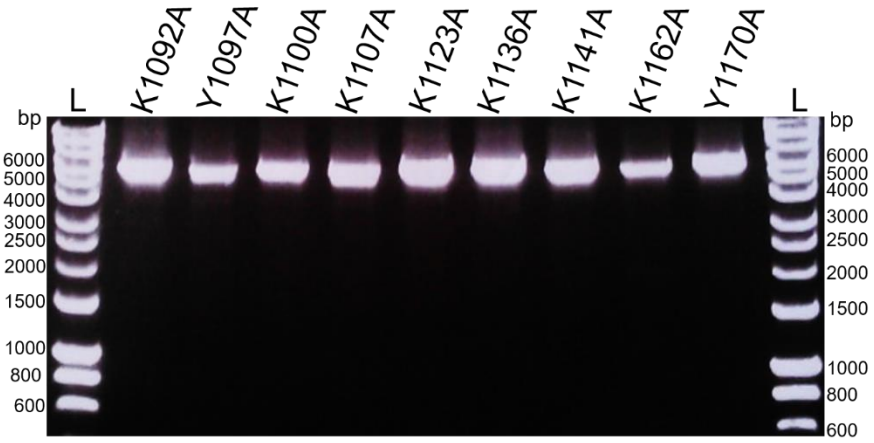


Figure 3.8 Agarose gel electrophoresis of CBM77$_{PL1/9}$ site-direct mutagenesis PCR products
After PCR (Chapter 2.1.11) and *DpnI* digestion, 5 µl of products with 2 µl of loading dye were run in a 1% agarose gel (Chapter 2.1.12). The respective mutants were labelled at the figure top and L represent the size marker in base pairs (bp). The expected sizes for all fragments is 5700 bp.

The AGE of wild type and variants of CBM77$_{PL1/9}$ revealed that the mutants K1092A, K1107A and K1162A displayed no electrophoretic retardation of migration in the ligand-containing gels, and thus no binding to pectin (Figure 3.9B). The mutants K1136A and K1141A showed a modest reduction in migration in the PGA-containing gel, suggesting weak binding to the polysaccharide. Mutation of the residues on the concave surface (Y1097, Y1170, K1100 and K1123) did not alter the migration profile compared to the wild type protein, suggesting no loss in pectin binding.

The affinity of the different mutants against PGA orange was quantified by ITC Table 3.5). As observed in the AGE experiments, the mutations K1092A, K1107A and K1162A abrogated binding to pectins indicating that these lysine residues comprise the ligand binding site. Interestingly, the single mutants K1136A and K1141A displayed only a modest reduction in affinity for PGA orange (K_A 5 x 10^3 M^{-1} and 4.2 x 10^3 M^{-1}, respectively) when compared with the wild type protein (K_A = 1.2 x 10^4 M^{-1}). However, the alanine substitution of both residues in a single protein (K1136A/K1141A) resulted in complete loss of binding indicating that these lysines display functional redundancy.

Based on the mutant binding data, the CBM77$_{PL1/9}$ ligand recognition surface is ~ 25 Å. Pectins adopt a compressed “accordion-like” structure in which a disaccharide subunit spans a distance of 8 Å (Abbott *et al*. 2007), suggesting that the binding site can accommodate a hexasaccharide, as reported before (Table 3.3).

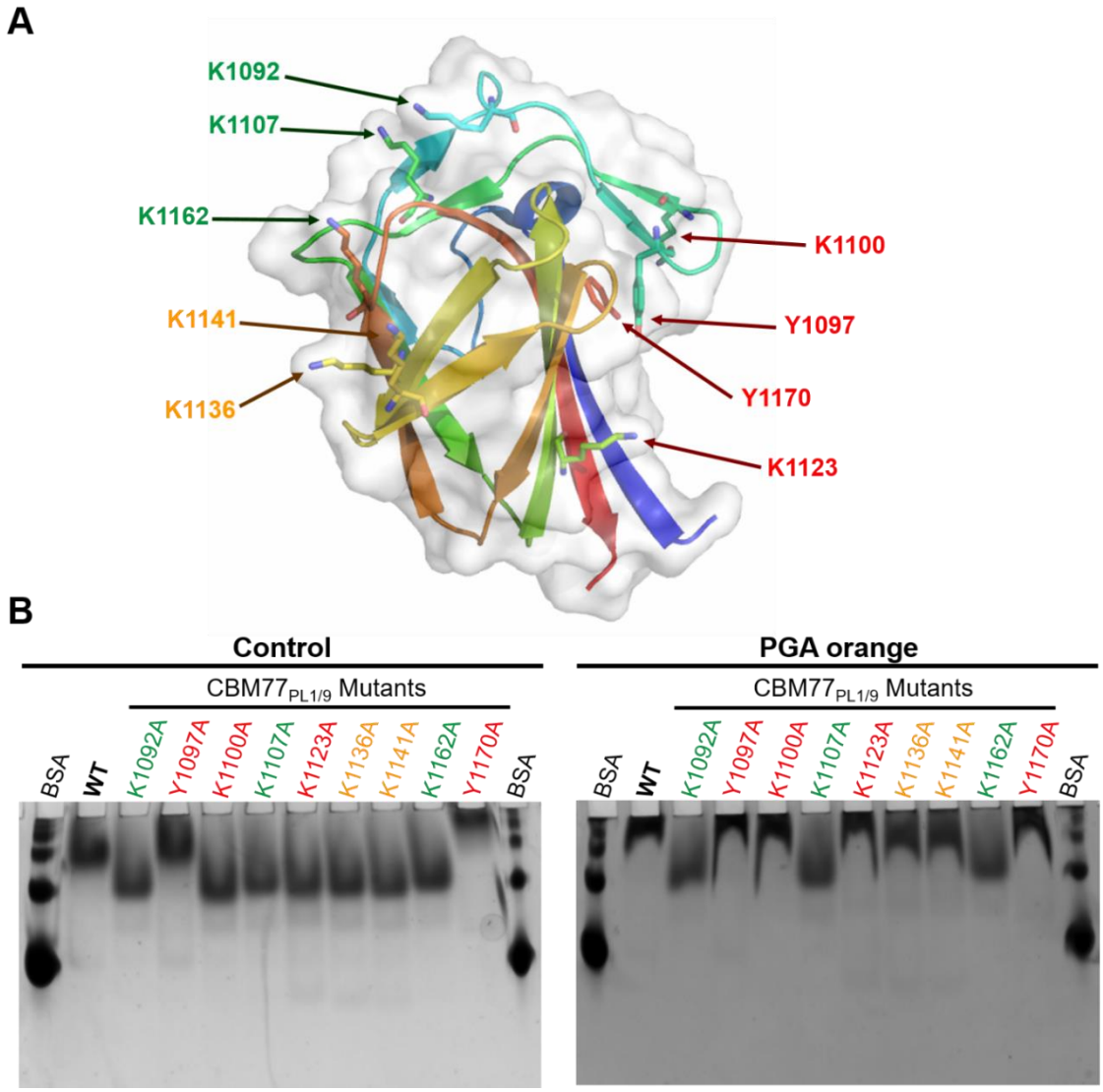


Figure 3.9 Cartoon showing the location of the CBM77$_{PL1/9}$ mutations and affinity gel electrophoresis

A. Cartoon of CBM77$_{PL1/9}$ showing the different residues mutated. Basic residues that contribute to binding are represented in green when a single mutation resulted in loss of binding. In orange are residues that can also contribute to pectin recognition but only the double mutant K1136A/K1141A had an effect on pectin binding. Other resides mutated that are not implicated in ligand recognition are represented in red. All the residues indicated were mutated to alanine. Panel was constructed using PyMOL software. **B.** AGE (Chapter 2.2.1) of CBM77$_{PL1/9}$ wild-type (WT) and mutants without ligand (control) and in presence of 0.1% ligand (PGA orange). Each protein (5 µg) was applied to the native gels and electrophoresis was performed in non-denaturing conditions. The amino acid colour is as described in Panel A. (PGA) Polygalacturonic acid.

Table 3.5 ITC of the binding of CBM77$_{PL1/9}$ mutants against PGA orange

	Protein	K_A M^{-1}	ΔG *kcal mole^{-1}*	ΔH *kcal mole^{-1}*	TΔS *kcal mole^{-1}*	n
	CBM77$_{PL1/9}$	1.2 (±0.0) x 10^4	-5.56	-15.0 (±0.5)	-9.45	1.0
CBM77$_{PL1/9}$ Mutant	K1092A	No binding				
	Y1097A	9.1 (±0.3) x 10^3	-5.4	-12.9 (±0.2)	-7.5	1.0
	K1100A	9.4 (±0.3) x 10^3	-5.4	-11.6 (±0.2)	-6.2	1.0
	K1107A	No binding				
	K1123A	9.4 (±0.4) x 10^3	-5.4	-9.1 (±0.2)	-3.7	1.0
	K1136A	5.0 (±0.8) x 10^3	-5.0	-7.6 (±0.7)	-2.6	1.0
	K1141A	4.2 (±0.4) x 10^3	-4.9	-9.1 (±0.4)	-4.2	1.0
	K1162A	No binding				
	Y1170A	7.2 (±0.2) x 10^3	-5.2	-12.2 (±0.1)	-7.0	1.0
	K1136A/K1141A	Binding too weak to quantify				

The errors correspond to the fitting of the ITC curve using the software MicroCal Origin7. The thermodynamic parameters were as follows K_A is the association constant, ΔG is the change in free enthalpy, ΔH is the enthalpy of binding, T is the temperature in degree absolute and ΔS is the entropy of the binding. The titrations were performed in 50 mM HEPES, 200 mM NaCl pH 7.0 with 50 μM of protein (cell), 0.4% of substrate (syringe). For details about ITC see Chapter 2 2.2.

3.3.3.3 Different approaches explored to obtain a CBM77$_{PL1/9}$ ligand complex

The structure of CBM77$_{PL1/9}$ in complex with ligand would have disclosed the interactions involved in ligand recognition. Different approaches were used to try to obtain this complex: CBM77$_{PL1/9}$ was co-crystallised in the presence of i) 30 mM trigalacturonic acid, ii) 20 mM of tri and tetragalacturonic acid or iii) 20 mM hepta- and octagalacturonic acid. Additionally, several crystals were soaked with 20 mM of hepta- and octagalacturonic acid (Dr. Arnaud Basle performed the soaking technique). Despite these different approaches, it was not possible obtain a structure with ligand bound.

3.3.4 *Phylogeny of CBM77 homologues proteins*

The CAZy database (Cantarel *et al*. 2009; Lombard *et al*. 2014) has, currently, 23 members listed in CBM family 77. This database only presents proteins identified in organisms for which genome sequencing is completed and published. However, the protein BLAST (see chapter 2.4) (Johnson *et al*. 2008) of $CBM77_{PL1/9}$ sequence revealed that the protein has 101 homologues with sequence similarity higher than 45%. To explore the conservation of amino acids implicated in ligand recognition, all the 102 sequences were aligned (Figure 3.10) using the online tool MUSCLE (Edgar 2004a; Edgar 2004b). Two of the three key ligand binding residues (K1107 and K1162) were invariant. The additional key residue, K1092, was highly conserved but not invariant (9 sequences did not have this residue). As K1136 and K1141 displayed functional redundancy, it was also important to establish the conservation of these binding residues. The alignment showed that K1136 was highly conserved in CBM77. Indeed, only 8 sequences lacked this amino acid. In these 8 sequences K1141 was not conserved. In contrast, only 19 out of 102 sequences (20%) had K1141. The lack of the key residue K1092 (in 9 sequences) and both of the functional redundant amino acids K1136/K1141 (in 8 sequences) indicates that 17 members of CBM77 may not retain a pectin binding function.

A phylogenetic tree for all the 102 proteins was generated (Figure 3.11) utilizing the online tool Phylogeny.fr (Dereeper *et al*. 2008; Dereeper *et al*. 2010). To explore the conservation of the key ligand binding residues from an evolutionary perspective, the different branches were analysed. The CBM77 homologs evolved from a common branch. One branch showed all the proteins without the key binding residue K1092 originated from the same internal node (Figure 3.11, red). The branch containing $CBM77_{PL1/9}$ was highlighted in green (Figure 3.11). This green branch also contains

the majority of the proteins (17 out of 19 sequences) in which the five residues implicated in pectin binding are conserved (K1092, K1107, K1136, K1141 and K1162) (circled in purple). Regarding the functionally redundant residues, K1136 and K1141, the lack of both amino acids appeared to be random, these proteins do not segregate to a common branch (asterisk, Figure 3.11). Thus, it was not possible to identify a common ancestor for these putative non-pectin binding protein modules. Proteins that were not coloured red or highlighted (circle or asterisk) are CBM77 homologs in which K1141 is not conserved. However, all these proteins showed complete conservation of the remaining four residues (K1092, K1107, K1136 and K1162) so, due to functional redundancy between K1136 and K1141, they are expected to bind to PGA with low DE.


1070
1080
1090
1100
1110
1120
gi|524417993|
gi|523997677|
gi|987335348|
gi|492699252|
gi|524072348|
gi|494743614|
gi|494748257|
gi|488622263|
gi|501155016|
gi|657705528|
gi|916589867|
gi|656287948|
gi|918009241|
gi|302396200|
gi|916919308|
gi|551035491|
gi|916938738|
gi|524733776|
gi|524646331|
gi|1011203728|
gi|280977867|
gi|938964799|
gi|916849385|
gi|916929760|
CBM77PL1/9
gi|915417709|
gi|916899756|
gi|640298674|
gi|503264543|
gi|515522873|
gi|515522873|
gi|917020771|
gi|1044457998|
gi|941238041|
gi|495741518|
gi|947364632|
gi|947415005|
gi|736706489|
gi|746322763|
gi|835291328|
gi|1039211638|
gi|524217644|
gi|972462257|
gi|500931641|
gi|949903819|
gi|737717831|
gi|947543022|
gi|924284237|
gi|755725553|
gi|973371377|
gi|753211454|
gi|1044455067|
gi|947580726|
gi|941248251|
gi|655152621|
gi|655108143|
gi|947963332|
gi|1033306847|
gi|950272975|
gi|918362507|
gi|916487846|
gi|916487846|
gi|916358126|
gi|972462226|
gi|764886895|
gi|738800228|
gi|738836611|
gi|939707331|
gi|754809106|
gi|754841217|
gi|748610912|
gi|974715421|
gi|754871028|
gi|746248469|
gi|754849217|
gi|918369344|
gi|754821533|
gi|573548264|
gi|740782856|
gi|985639211|
gi|1011676659|
gi|985646027|
gi|521095222|
gi|494218073|
gi|502570193|
gi|738771650|
gi|497279179|
gi|918362507|
gi|647225100|
gi|517716831|
gi|738732908|
gi|494670221|
gi|504045213|
gi|498034491|
gi|497762333|
gi|697197441|
gi|916948254|
gi|917030410|
gi|932999421|
gi|502237142|
gi|495394767|
gi|495394768|


1130
1140
1150
1160
1170
gi|524417993|
gi|523997677|
gi|987335348|
gi|492699252|
gi|524072348|
gi|494743614|
gi|494748257|
gi|488622263|
gi|501155016|
gi|657705528|
gi|916589867|
gi|656287948|
gi|918009241|
gi|302396200|
gi|916919308|
gi|551035491|
gi|916938738|
gi|524733776|
gi|524646331|
gi|1011203728|
gi|280977867|
gi|938964799|
gi|916849385|
gi|916929760|
CBM77PL1/9
gi|915417709|
gi|916899756|
gi|640298674|
gi|503264543|
gi|515522873|
gi|515522873|
gi|917020771|
gi|1044457998|
gi|941238041|
gi|495741518|
gi|947364632|
gi|947415005|
gi|736706489|
gi|746322763|
gi|835291328|
gi|1039211638|
gi|524217644|
gi|972462257|
gi|500931641|
gi|949903819|
gi|737717831|
gi|947543022|
gi|924284237|
gi|755725553|
gi|973371377|
gi|753211454|
gi|1044455067|
gi|947580726|
gi|941248251|
gi|655152621|
gi|655108143|
gi|947963332|
gi|1033306847|
gi|950272975|
gi|918362507|
gi|916487846|
gi|916487846|
gi|916358126|
gi|972462226|
gi|764886895|
gi|738800228|
gi|738836611|
gi|939707331|
gi|754809106|
gi|754841217|
gi|748610912|
gi|974715421|
gi|754871028|
gi|746248469|
gi|754849217|
gi|918369344|
gi|754821533|
gi|573548264|
gi|740782856|
gi|985639211|
gi|1011676659|
gi|985646027|
gi|521095222|
gi|494218073|
gi|502570193|
gi|738771650|
gi|497279179|
gi|918362507|
gi|647225100|
gi|517716831|
gi|738732908|
gi|494670221|
gi|504045213|
gi|498034491|
gi|497762333|
gi|697197441|
gi|916948254|
gi|917030410|
gi|932999421|
gi|502237142|
gi|495394767|
gi|495394768|

Figure 3.10 Alignment of proteins in CBM77 family

NCBI reference codes for each protein are shown left. Sequences were aligned in MUSCLE (Edgar 2004a; Edgar 2004b) and visualized using ESPrit 3.0 (Robert and Gouet 2014). White circle identifies the binding residues (K1107, K1162) invariant in the CBM77 family. Black circle indicates other residues that are implicated in CBM77 binding but are not fully conserved within the family (K1092, K1141 and K1136). Amino acids with 100% of conservation highlighted in a red background and residues that are functionally highly conserved are boxed and coloured red. The residues numeration is according with $CBM77_{PL1/9}$ sequence.

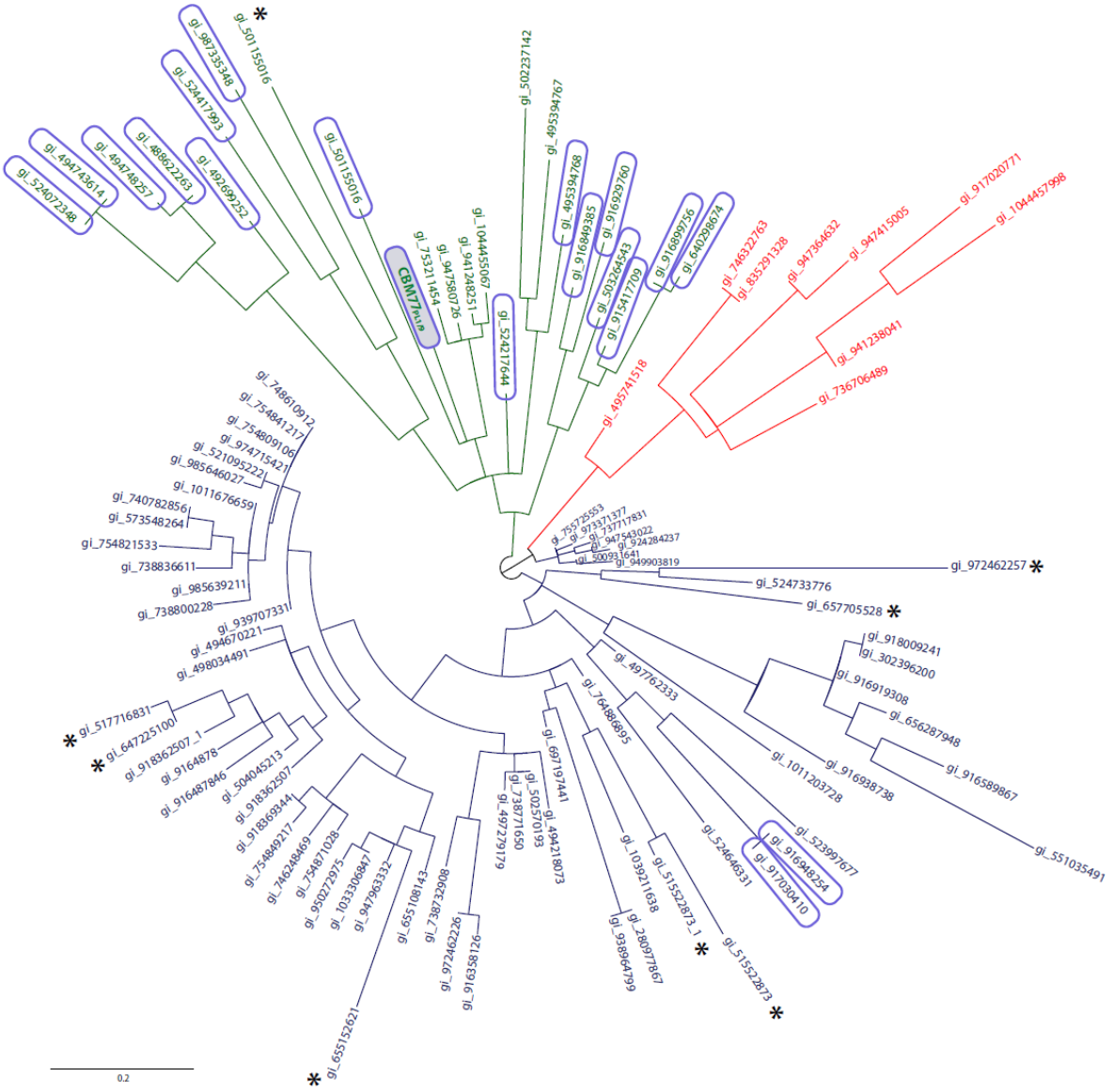


Figure 3.11 Phylogenetic tree of proteins in CBM77 family

The phylogenetic circle tree was constructed using the alignment file generated in Figure 3.10 and the online tool Phylogeny.fr (Dereeper *et al*. 2008; Dereeper *et al*. 2010). Proteins are indicated with the respective NCBI reference code. Proteins sharing a common ancestry with $CBM77_{PL1/9}$ are shown in green. Proteins without the key ligand binding lysine K1092 are shown in red. The remaining proteins are shown in blue. Proteins with K1136/K1146 conserved are circled in purple. * denotes proteins that lack an equivalent basic residue to the amino acids K1136 or K1146. The branch length is proportional to the number of substitutions per site.

CBM77$_{PL1/9}$ is a component of a large *R. flavefaciens* cellulosomal protein that contains two protein modules located in pectate lyase families PL1 and PL9 (Figure 3.2). As shown before, this CBM family is present in ~100 proteins from different bacteria. To explore if the molecular architecture of enzymes containing members of CBM77 are conserved the different bacterial species containing the pectin CBM were identified using the BLAST tool (Johnson *et al*. 2008). The data revealed 11 bacterial species containing proteins where a CBM77 member was a component. Representatives for each protein were analysed. From the different species, the protein sequence with highest identity (to CBM77$_{PL1/9}$) was selected for further analysis. The signal peptide was predicted using the online tool Signal P 4.1 Server (Petersen *et al*. 2011) and the different modules were identified using Pfam 30.0 (Finn *et al*. 2016), protein BLAST and the protein alignment tool MUSCLE (Edgar 2004a; Edgar 2004b). The schematic representation of the molecular architecture of proteins containing a CBM77 is shown in Figure 3.12. The majority of the proteins display at least one CBM77 appended to a PL1 catalytic module. Similar to *R. flavefaciens*, in the *Eubacterium* and *Catenovulum* proteins the pectin binding CBM was also associated with a PL9 catalytic module. In the *Clostridium cellulovorans* protein the CBM77 was not associated with any catalytic module but with a β-helix domain of unknown function. Since the PLs display a typical β-helix domain this module may be a functional PL from a new family.

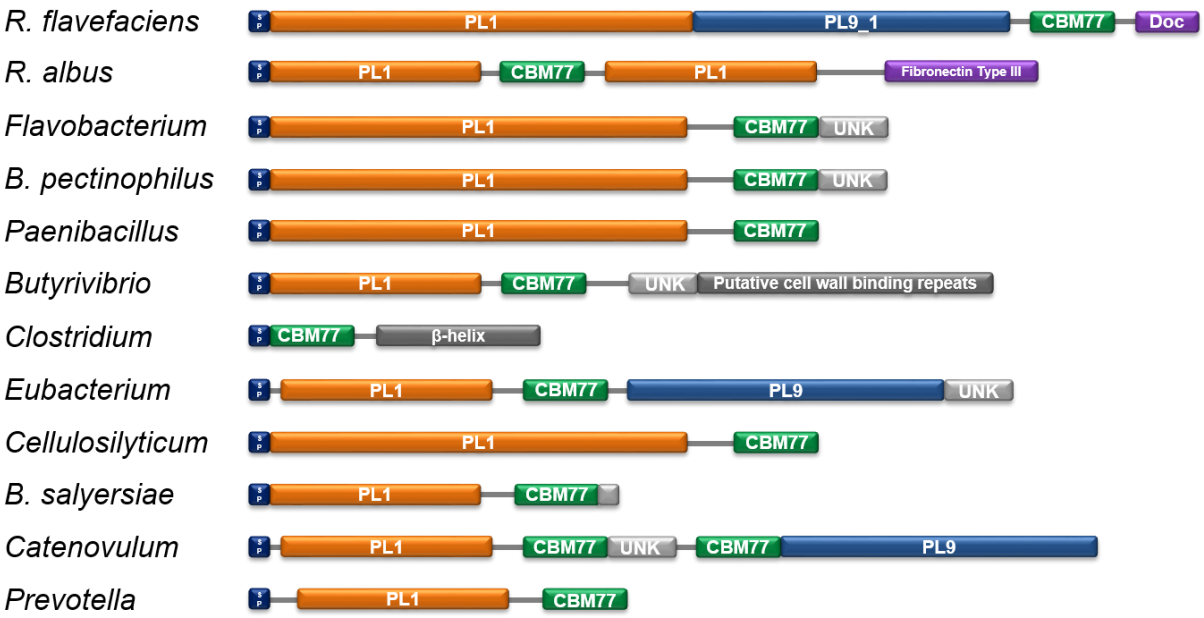


Figure 3.12 Molecular architecture of proteins containing a CBM77

CBM77 modules are represented in green and PL1 and PL9 catalytic modules in orange and blue, respectively. All proteins have an N-terminal signal peptide (SP). In *R. flavefaciens* a C-terminal dockerin module (Doc) is represented in purple. UNK (light grey) refers to modules of unknown function. According with Pfam, modules with a putative fold or domain are represented in dark grey. Linkers are shown as a solid line. The species represented in this figure and the respective NCBI reference code are: *R. flavefaciens* 17 (WP_009983557.1); *R.* albus 7 (WP_013499204.1); *Flavobacterium johnsoniae UW101 (WP_102023313); B. pectinophilus* (WP_008119467.1)*; Paenibacillus sp.* Y412MC10 (CDD55579.1); *Butyryvibrio proteoclasticus B316* (ADL3515.1); *Clostridium cellulovorans* (WP_010076517.1)*; Eubacterium sp. CAG:86* (CCX83499); *Cellulosilyticum rummicola* (AC298651.1); *B. salyersiae* (WP_007479022.1); *Catenovulum agarivorans* (WP_016956127.1)*;* and *Prevotella bergensis* (WP_007175372.1). (*R.*) *Ruminococcus*; (*B.*) *Bacteroides*.

3.3.5 *CBM77$_{PL1/9}$ associated polysaccharide lyases activity*

To understand the contribution of CBM77$_{PL1/9}$ in homogalacturonan degradation different constructs were generated in which the CBM was appended to the PL1 (RfPel1) and PL9 (RfPel9) catalytic modules, or enzymes without the binding module were generated (see Figure 3.2). The different constructs were successfully expressed in soluble form (see Figure 3.4) and characterized.

The pH profiles of the three enzymes lacking the CBM were determined by measuring the formation of unsaturated products at 235 nm. RfPel9 and RfPel1/9 were active in a wide pH range, with an optimum at 8.5 in Tris-HCl buffer (Figure 3.13). The similar pH profile for both constructs reflects the dominance of the more active RfPel9 catalytic module compared to the RfPel1 pectate lyase (see below). Unexpectedly, the pH optimum of RfPel1 was 7.0 (in Tris-HCl buffer), significantly lower than RfPel9 (Figure 3.13).

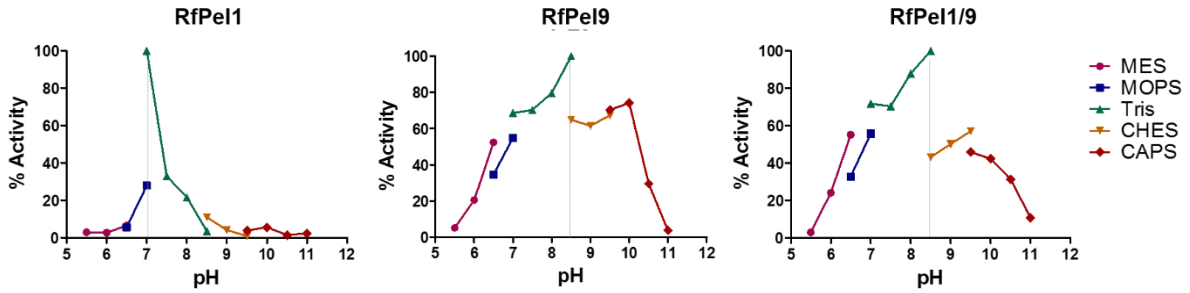


Figure 3.13 pH profile of RfPel1, RfPel9 and RfPel1/9

The product formation was measured at 235 nm (see Chapter 2.2.5.1). All reactions were performed at 37 °C using the substrate pectin DE 30% at 0.2%, 2 mM $CaCl_2$ and 200 mM of the respective buffer. The enzyme concentrations were 10 nM for RfPel1 and RfPel9 and 5 nM for RfPel1/9. The % of activity was calculated based on the highest enzyme rate defined as 100%.

The final (limit) products produced by the various enzymes from pectins with different DEs were determined by TLC (Figure 3.14). Against pectin DE 30%, RfPel1 produced two major products while RfPel9 generated an additional longer oligosaccharide. The association of both catalytic modules produced the same product profile as RfPel9. Moreover, the association of the lyases with the $CBM77_{PL1/9}$ did not influence the nature of the limit reaction products for any of the substrates tested. Both lyases irrespective of the presence of $CBM77_{PL1/9}$ displayed specificity for pectins with a low DE, consistent with the binding profile of $CBM77_{PL1/9}$.

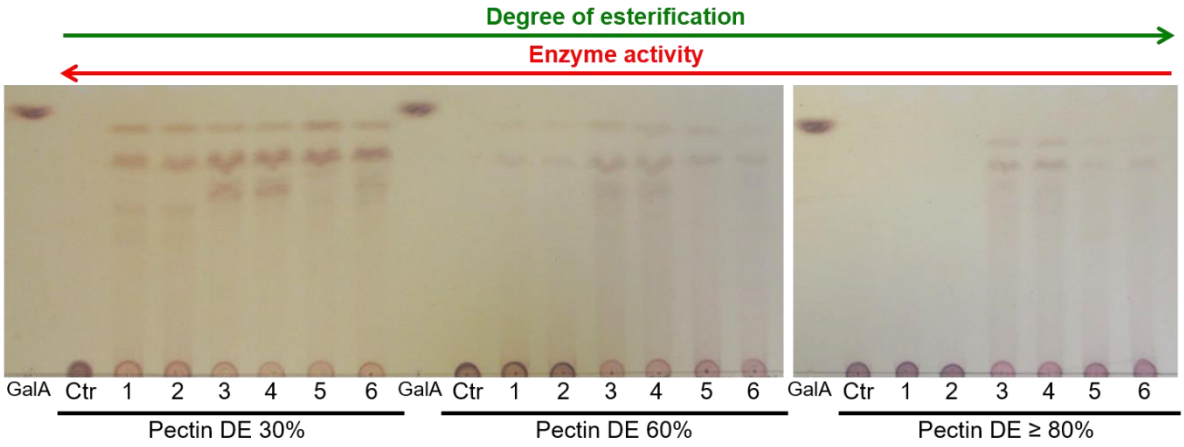


Figure 3.14 PLs products on pectins with different degree of esterification

All reactions were performed with 100 nM of enzyme, 0.2% substrate, 2 mM $CaCl_2$ and 100 mM Tris-HCl pH 7.5 (RfPel1 constructs) or pH 8.5 (RfPel9 and RfPel1/9 constructs). Reactions were incubated at 37 °C for 16 h, stopped by heat inactivation and then subjected to TLC analysis (Chapter 2.2.3). Ctr – control reaction where substrate was incubated with buffer but no enzyme 1 – RfPel1; 2 – RfPel1+CBM77$_{PL1/9}$; 3 – RfPel9; 4 – RfPel9+CBM77$_{PL1/9}$; 5 – RfPel1/9; 6 – RfPel1/9+CBM77$_{PL1/9}$; (GalA) D-Galacturonic acid standard; (DE) degree of esterification.

The catalytic activity of the enzymes against the different substrates was measured at 235 nm, reflecting the double bond formed between C4 and C5 during the β-elimination reaction. The kinetic values for each construct are shown in Table 3.6 and the respective kinetic curves presented in Figure 3.15. Against pectin DE 30%, RfPel9 was 25-fold more active than RfPel1. The difference in catalytic efficiency of these enzymes is explained by variation in the K_M while the k_{cat} values were similar. The lower K_M of RfPel9 suggests that the enzyme presents a higher affinity for pectins when compared with RfPel1, assuming that k_2 is much lower than k_{-1}. The association of both enzymes (RfPel1/9) resulted in a k_{cat}/K_M 1.6 x 10^5 min^{-1} mg^{-1} ml against Pectin DE 30%. This enzyme construct was 15-fold more active that the RfPel1 lyase alone but 1.8-fold slower than the RfPel9 enzyme. So, it is possible to speculate that the activity of RfPel1/9 is mainly driven by the RfPel9 catalytic module. As expected, the activity of the enzymes decreased ~10 and ~140 fold when the substrate DE increased from 30% to 60% or ≥80%, respectively. This decrease in activity is driven by an elevated K_M,

suggesting that affinity for the substrate decreases when DE of the pectin increases. The specificity of the RfPel1 and RfPel9 lyases for substrates with a low DE reflects the binding profile of the appended CBM77$_{PL1/9}$. The association of CBM77$_{PL1/9}$ with the lyases, however, did not result in a significant increase in k_{cat}/K_M for any of the different constructs. Indeed, the fusion of CBM77$_{PL1/9}$ to different enzymes led to a modest decrease in both K_M and k_{cat}.

Table 3.6 Catalytic activity of PL proteins with and without CBM77$_{PL1/9}$

Substrate	Recombinant protein	k_{cat} (min^{-1})	K_M (mg/ml)	k_{cat}/K_M (min^{-1} mg^{-1} ml)
Pectin DE 30%	RfPel1	5.9 x 10^3 ± 351	0.547 ± 0.080	1.1 x 10^4 ± 2.2 x 10^3
	RfPel1+CBM77$_{PL1/9}$	158 ± 7.5	0.050 ± 0.001	3.1 x 10^3 ± 774
	RfPel9	4.6 x 10^3 ± 100	0.015 ± 0.001	2.8 x 10^5 ± 3.2 x 10^4
	RfPel9+CBM77$_{PL1/9}$	3.4 x 10^3 ± 109	0.009 ± 0.001	3.7 x 10^5 ± 6.9 x 10^4
	RfPel1/9	1.8 x 10^3 ± 68	0.011 ± 0.001	1.6 x 10^5 ± 2.6 x 10^4
	RfPel1/9+CBM77$_{PL1/9}$	1.5 x 10^3 ± 68	0.007 ± 0.001	2.1 x 10^5 ± 4.2 x 10^4
Pectin DE 60%	RfPel1/9	3.0 x 10^3 ± 482	0.180 ± 0.05	1.6 x 10^4 ± 7.2 x 10^3
	RfPel1/9+CBM77$_{PL1/9}$	1.5 x 10^3 ± 81	0.068 ± 0.012	2.2 x 10^4 ± 5.2 x 10^3
Pectin DE ≥ 80%	RfPel1/9	2.2 x 10^3 ± 417	3.27 ± 1.20	662 ± 370
	RfPel1/9+CBM77$_{PL1/9}$	1.4 x 10^3 ± 417	2.70 ± 0.64	485 ± 168

All data was generated using a non-regression analysis. The standard errors were generated from technical triplicates. For further details see Figure 3.15.

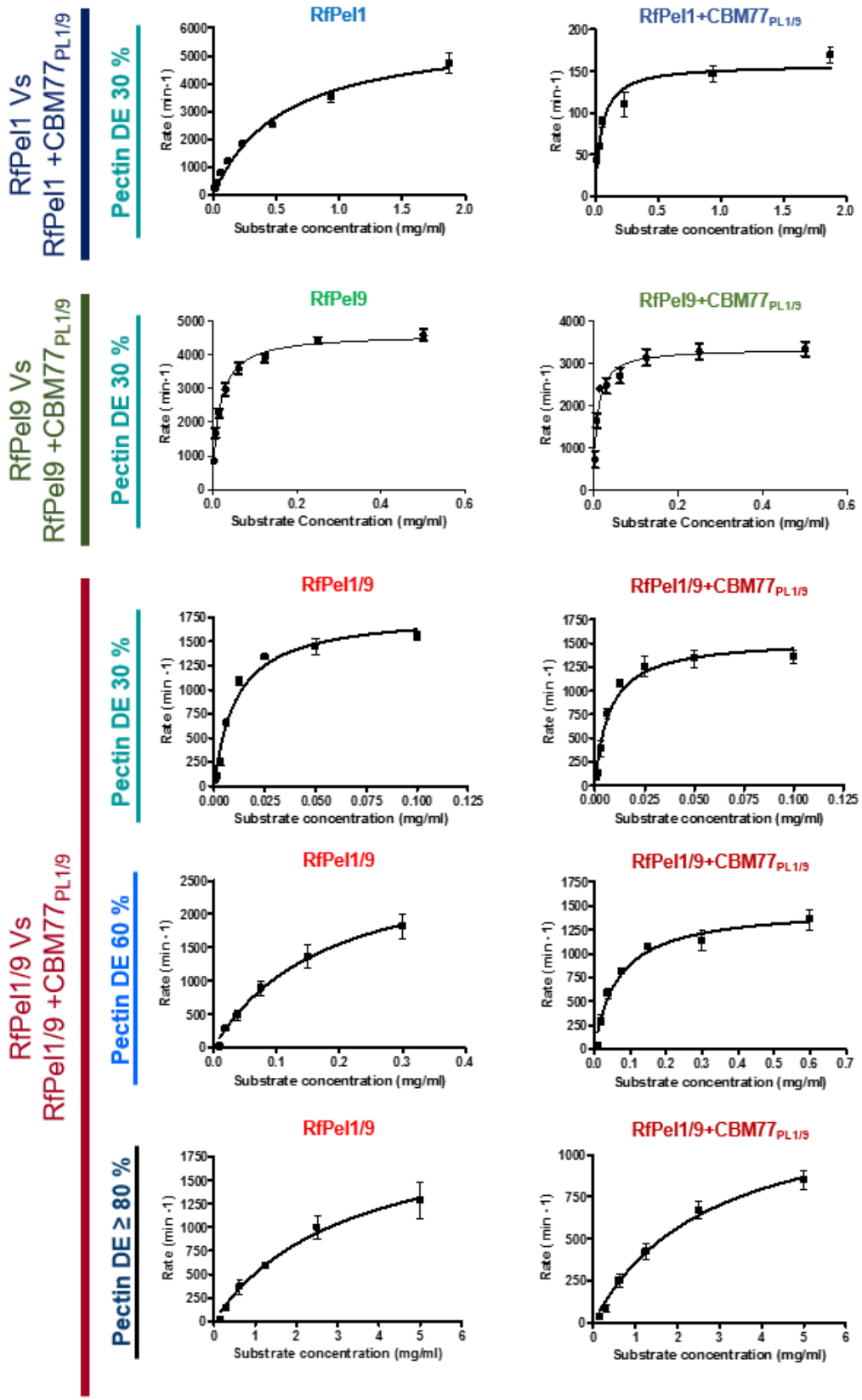


Figure 3.15 Graphical representation of kinetics data for PLs with and without CBM77$_{PL1/9}$

All reactions were performed at 37 °C with 2 mM $CaCl_2$ and 200 mM Tris-HCl pH 7.5 (RfPel1 constructs) or pH 8.0 (RfPel9 and RfPel1/9 constructs). Initial rates were measured at 235 nm and plotted in GraphPad Prism 5.0 software using non-regression analysis (see Chapter 2.2.5.1). The standard errors were generated from technical triplicates.

3.3.6 *The role of CBM77$_{PL1/9}$ in plant cell wall degradation*

To understand the *in vivo* role of CBM77$_{PL1/9}$ in pectin degradation, sections of *Tobacco* were incubated with RfPel1/9 or RfPel1/9+CBM77$_{PL1/9}$ and pectin was detected by immunolabelling with LM19 (monoclonal antibody that specifically binds to PGA) (Verhertbruggen *et al*. 2009). The level of pectin degradation was quantified indirectly by measuring the intensity of LM19 signal in pixels (Figure 3.16). As expected, the increase in enzyme concentration resulted in a significant decrease in pectin present in the plant cell wall and, as evidenced by a reduction in the LM19 signal. Incubation with 84 nM of either enzyme for 2 h resulted in complete degradation of the pectin present in these sections. The level of pectin degradation was not directly proportional to enzyme concentration. Indeed, in sections treated between 8.4 nM and 42 nM RfPel1/9+CBM77$_{PL1/9}$ only a small decrease of LM19 signal was detected, suggesting that there is a small proportion of pectin that was recalcitrant to degradation by the PL. As with purified substrate, CBM77$_{PL1/9}$ did not have a significant effect on pectin degradation. As observed previously, CBM77$_{PL1/9}$ did not potentiate the activity of the two lyases against, at least, tobacco cell walls.

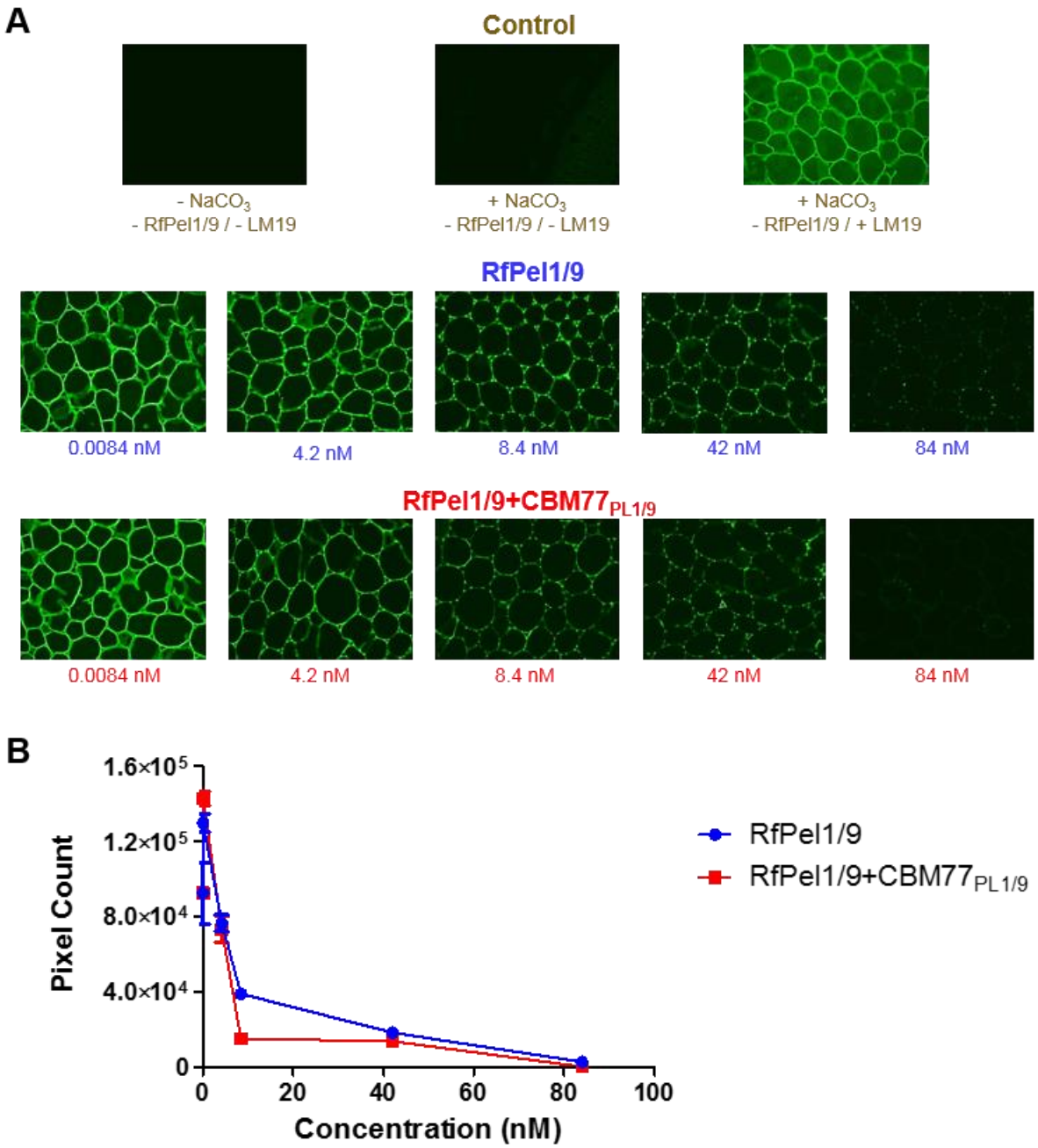


Figure 3.16 Activity of RfPel1/9 and RfPel1/9 +CBM77$_{PL1/9}$ against plant cell wall

A. Example of equivalent sections of *Tobacco* stem pith parenchyma labelled with monoclonal antibody (LM19) against pectins after incubation for 2 h with different concentrations of enzyme. To explore possible fluorescence artefacts, control sections were incubated (+) or (-) with NaCO3, RfPel1/9 enzyme and LM19. **B.** Quantification of the signal intensity of antibody LM19 after plant cell wall treatment with increasing concentration of enzymes (see Panel A). The software Volovity was used to calculate the signal intensity in pixels and the values were plotted in GraphPad Prism 5.0. The standard errors were generated from four replicates of two independent equivalent sections. Complete methods for these experiments are described in Chapter 2.2.14.

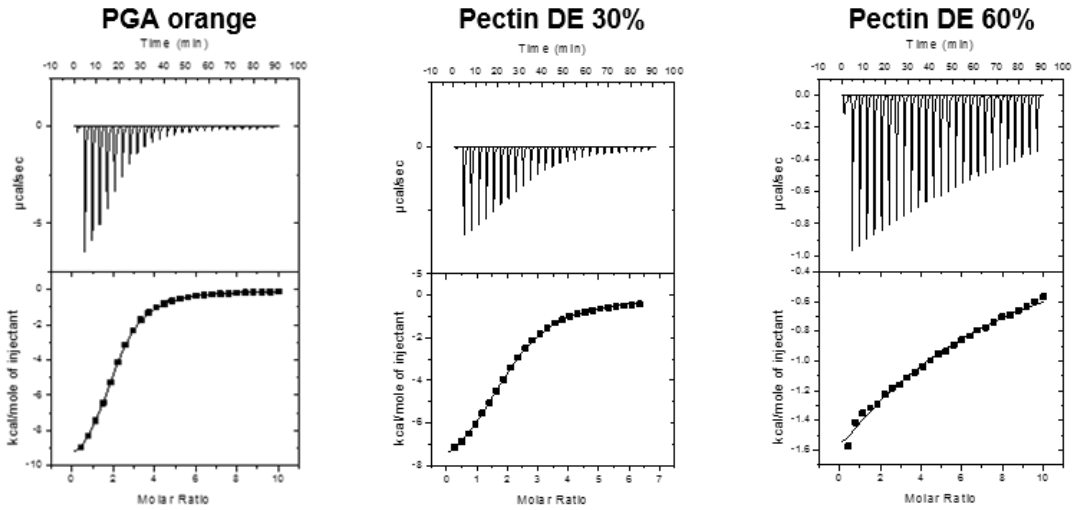
3.3.7 *Exploring CBM77$_{PL1/9}$ as a probe to label pectins in plant cell wall*

Reflecting the apparent lack of CBMs that recognise pectins, CBM77$_{PL1/9}$ was tested as a potential probe for homogalacturonan in the context of plant cell wall. To explore this hypothesis two different proteins containing CBM77$_{PL1/9}$ were used: CBM77$_{PL1/9}$ monomer and a dimer containing two molecules of the CBM separated by a linker containing Pro-Thr repeats (CBM77$_{PL1/9}$_CBM77$_{PL1/9}$). The binding of different constructs to PGA orange was tested by ITC. The data (Figure 3.17) showed that the CBM77$_{PL1/9}$ dimer displayed the same binding preference for pectins with a low DE described previously for the monomer (Figure 3.6 and Table 3.3). Surprisingly, the affinity of CBM77$_{PL1/9}$_CBM77$_{PL1/9}$ for pectin DE 30% (K_A 4.2 x 10^4 M^{-1}) was only ~four-fold higher than the monomer (K_A 1.1 x 10^4 M^{-1}, Table 3.3).

To explore if CBM77$_{PL1/9}$ is useful as a pectin probe, the binding of different proteins was compared with LM19, a well characterized monoclonal antibody that recognizes homogalacturonan with low DE (Verhertbruggen *et al*. 2009). By enzyme-linked immunosorbent assay (ELISA), CBM77$_{PL1/9}$, CBM77$_{PL1/9}$ dimer and LM19 revealed high affinity to pectin with DE 30% (Figure 3.18). As expected, none of these proteins were able to bind to pectins with a high DE (≥ 80%), but the de-esterification of this polysaccharide with $NaCO_3$ restored the binding of CBM77$_{PL1/9}$ and LM19. This proves the specificity of these proteins for non-esterified ligands. Interestingly, in pectin with a lower DE (30%) CBM77$_{PL1/9}$_CBM77$_{PL1/9}$ displayed tighter binding compared to the LM19 antibody. The increase in esterification from 30 to 60% led to a 5-fold decrease in dimer affinity while the binding strength of LM19 affinity was virtually unaffected, revealing that CBM77$_{PL1/9}$_CBM77$_{PL1/9}$ might be more sensitive to differences in ligand esterification. Additionally, despite the similar values K_A of CBM77$_{PL1/9}$ and

$CBM77_{PL1/9}$_$CBM77_{PL1/9}$ for Pectin DE 30%, determined by ITC (Table 3.3 and Figure 3.17), in ELISA assays the dimer bound 125-fold tighter to the polysaccharide compared to the monomer (Figure 3.18). Indeed, when the pectin DE increased to 60% the monomer binding was markedly reduced and the binding was not detectable upon a five-fold dilution of the protein. In contrast increasing the DE of the ligand to 60% had less influence on dimer binding.

A



B

Ligand	K_A M^{-1}	ΔG *kcal mole*$^{-1}$	ΔH *kcal mole*$^{-1}$	TΔS *kcal mole*$^{-1}$	n
PGA orange	5.9 (±0.1) x 10^4	-6.50	-10.79 (±0.0)	-4.29	2.00 ± 0.005
Pectin DE 30%	4.2 (±0.1) x 10^4	-6.30	-9.19 (±0.1)	-2.88	2.00 ± 0.014
Pectin DE 60%	Binding is too weak to quantify thermodynamics				

Figure 3.17 $CBM77_{PL1/9}$_$CBM77_{PL1/9}$ dimer binding thermodynamic parameters

A. ITC curves of CBM77_CBM77 binding to different pectins. ITC titrations were carried out as described in Chapter 2.2.2 with 50 μM of protein and 0.4% substrate in 50 mM HEPES, 200 mM NaCl pH 7.0. The top part of each panel corresponds to raw ITC data and the bottom part represents the integration of the area peaks using MicroCal Origin 7 software. **B.** Determination of thermodynamic parameters and affinity constant of the $CBM77_{PL1/9}$_$CBM77_{PL1/9}$ dimer for different ligands (for details see Chapter 2.2.2). The thermodynamic parameters were as follows K_A is the association constant, ΔG is the change in free enthalpy, ΔH is the enthalpy of binding, T is the temperature in degree absolute and ΔS is the entropy of the binding. The errors correspond to the fitting of the ITC curve using the software MicroCal Origin7. (DE) Degree of esterification; (PGA) polygalacturonic acid.

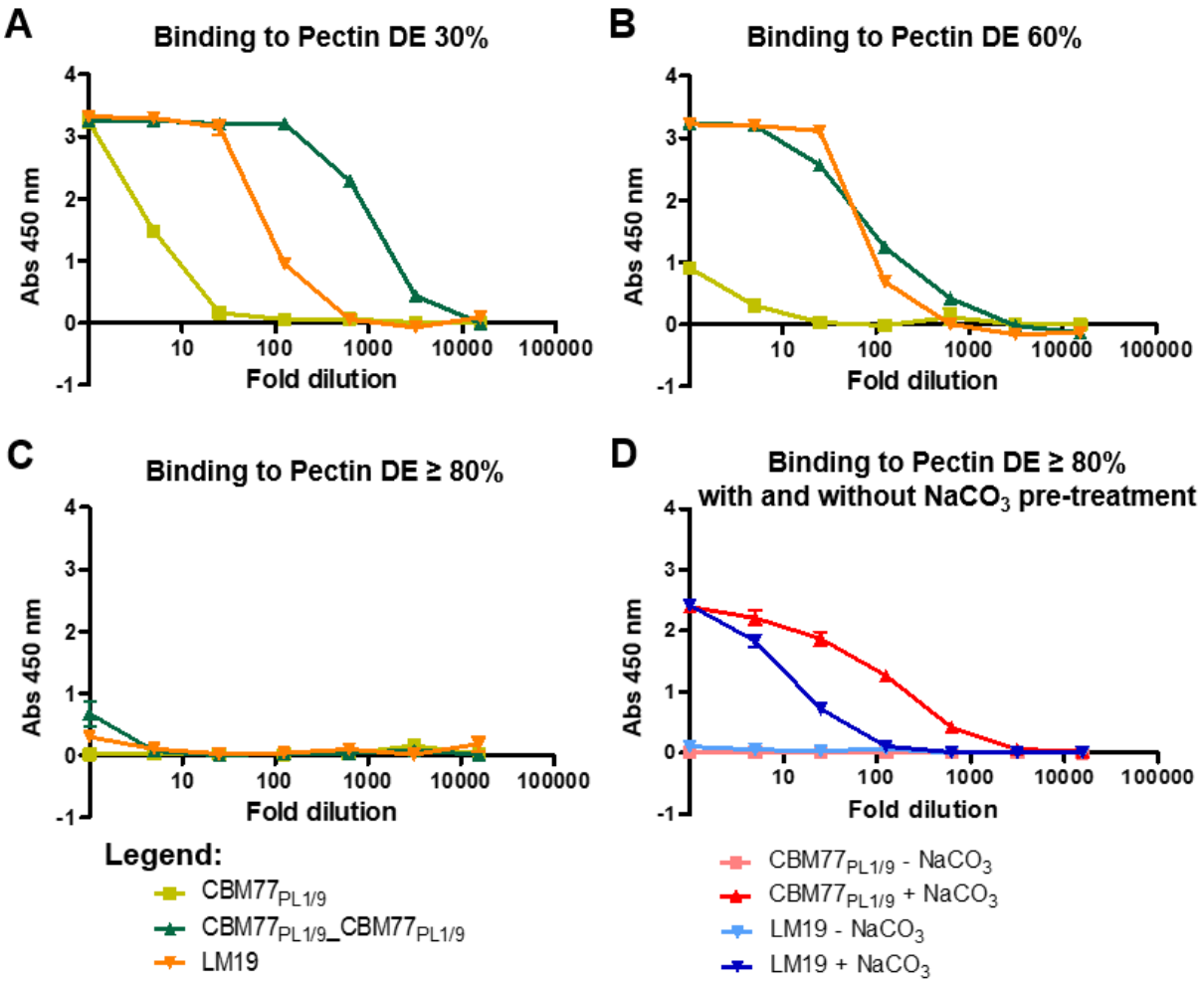


Figure 3.18 ELISA of CBM77$_{PL1/9}$, CBM77 $_{PL1/9}$ dimer and LM19 binding to pectins

The binding was measured by reading the intensity of signal al 450 nm and plotting this value vs the fold of proteins dilution (GraphPad Prism 5.0 software). Plates were coated with 50 µg/ml ligand: pectin DE 30% (A), pectin DE 60% (B) and pectin DE ≥ 80% (C and D). In ELISA D, prior to protein incubation, the ligand was pre-treated (+) or not (-) with 0.1 M sodium carbonate pH 11.4 ($NaCO_3$). ELISAs A, B and C were performed with an initial concentration of 0.8 µM CBM77$_{PL1/9}$ and CBM77$_{PL1/9}$_CBM77$_{PL1/9}$. In ELISA D the initial concentration of CBM77$_{PL1/9}$ was 20 µM. For LM19 the initial concentration was 1:10 dilution of antibody stock supplied by Professor Paul Know (Leeds University, UK). In all ELISAs 1:5 dilutions of CBMs or LM19 antibody were tested against the different ligand. CBMs and LM19 binding were detected with anti-His-HRP or anti-rat IgG-HRP, respectively. The standard errors were generated from five technical replicates. Details of the method were described in Chapter 2.2.13. (DE) Degree of esterification.

To test the hypothesis that CBM77$_{PL1/9}$ can be used as a suitable probe to target pectins in the context of plant cell wall, *Tobacco* and *Miscanthus* sections were labelled with CBM77$_{PL1/9}$, as a monomer or dimer, and LM19 (Figure 3.19). LM19 and CBM77$_{PL1/9}$_CBM77$_{PL1/9}$ bound to pectins in the cell wall of both plants. For *Tobacco*

sections, the pre-treatment with pectate lyases to remove the homogalacturonan resulted in the loss of binding for LM19 and the $CBM77_{PL1/9}$ dimer, indicating that both proteins bind specifically to this polysaccharide. The sections pre-treated with sodium carbonate to remove pectin esterification resulted in an apparent increase in the binding signal of LM19 and $CBM77_{PL1/9}$_$CBM77_{PL1/9}$ showing the preference of these proteins for un-esterified ligands. Unexpectedly, $CBM77_{PL1/9}$ (monomer) did not bind to pectins in the plant cells indicating that this protein is not a suitable probe for this polysaccharide. The labelling in *Miscanthus* showed very similar results to those obtained with *Tobacco* sections. Once more, LM19 and the $CBM77_{PL1/9}$ dimer bound to pectins, where the monomer did not show binding to this polysaccharide. This result suggests that $CBM77_{PL1/9}$_$CBM77_{PL1/9}$ can be explored as a probe to label pectins in tissues of different plants.

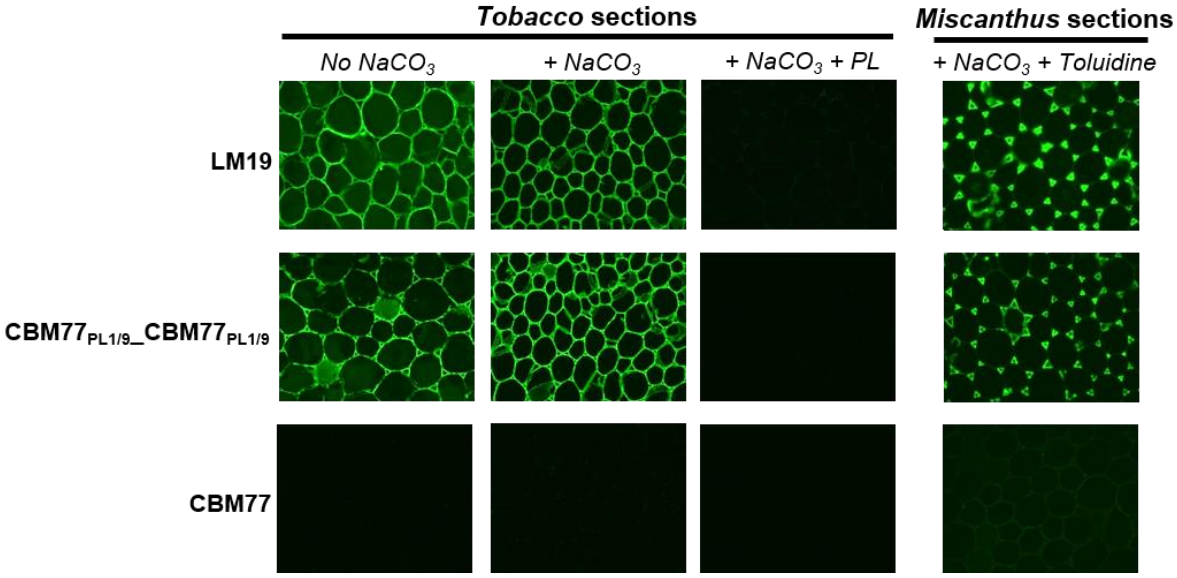


Figure 3.19 Pectin labelling in *Tobacco* and *Miscanthus* sections

Examples of equivalent sections of *Tobacco* and *Miscanthus* labelled with the monoclonal antibody LM19, $CBM77_{PL1/9}$ and the dimer $CBM77_{PL1/9}$_$CBM77_{PL1/9}$. The concentrations of proteins were 0.8 µM and 1:5 dilution of the stock for the different CBM proteins and LM19, respectively. Sections were pre-treated with PBS (no $NaCO_3$) or 0.1 M sodium carbonate pH 11.4 (+ $NaCO_3$) and, when specified, with 0.8 µM RfPel1/9 in Tris-HCl, pH 8.0, with 2 mM $CaCl_2$ (+ PL). Miscanthus sections were treated with toluidine to remove the tissue auto-florescence interference. For more details about the immunolabelling see Chapter 2.2.14.

3.4 Discussion

In this chapter CBM77$_{PL1/9}$, the founding member of a new CBM family, was characterized at a functional and structural level. This CBM77 binds to homogalacturonan, a major component of pectins present in plant cell walls. To date, only two other CBM families in the CAZy database contain members that bind to pectins. A CBM32 from *Yersinia*, a periplasmic binding protein (contains no catalytic module), has been shown to bind polygalacturonic acid (Abbott *et al*. 2007). Additionally, members of family CBM35 have also been shown to bind unsaturated oligosaccharides derived from homogalacturonan by PL action. This type-C exo-acting CBM required the PL pre-treatment because its binding site pocket recognizes Δ(4,5)-GalA but not saturated GalA (Montanier *et al*. 2009).

CBM77$_{PL1/9}$ presented preferential binding to polygalacturonic acid with a low DE. ITC showed that the CBM77 displayed optimum binding to oligosacchachides containing 7 to 8 sugar units, and that ligand recognition was metal-independent. This unusual binding to relatively long galacturonooligosaccharides was reported previously for a CBM32 that displays optimum binding to 10 galacturonic acid units (Abbott *et al*. 2007). Unlike CBM77$_{PL1/9}$, the *Yersinia* CBM was able to bind to short oligosaccharides with a DP ≥ 3. Additionally, this CBM has a calcium binding site that is conserved among members of family CBM32 (Abbott *et al*. 2007; Ficko-Blean *et al*. 2012). However, the role of metals in pectin recognition by the *Yersinia* CBM is unknown, it was speculated that, like other members of CBM32, the metal-binding site is only structural and does not contribute to ligand recognition (Abbott *et al*. 2007). However, in contrast to CBM77$_{PL1/9}$ and CBM32, calcium was shown to play a key role in CBM35 binding to unsaturated pectins and D-glucuronic acid (D-GlcA) (Montanier *et al*. 2009).

The structure of CBM77$_{PL1/9}$ was solved and the key residues implicated in substrate binding were determined by site-direct mutagenesis. The unusual binding site of this CBM was identified at the convex surface. A unexpected feature since in endo-active Type B CBMs the canonical ligand binding site is at the concave surface (Gilbert *et al*. 2013). Additionally, in CBM77$_{PL1/9}$ ligand recognition is mediated by basic polar residues (lysines). This ligand binding mode is distinct from other CBMs reported to date, where aromatic residues play a central role in glycan recognition (Boraston *et al*. 2004; Gilbert *et al*. 2013). However, the mechanism of pectin recognition displayed by CBM77$_{PL1/9}$ is similar to glycosaminoglycan binding proteins where glycan binding is also mediated by basic residues (arginines and lysines) that interact with the highly negative polysaccharides heparin and heparan sulfated (Hileman *et al*. 1998). The importance of these positively charged lysine residues explains why metal ions do not play a direct role in ligand recognition. As reported before when ligand esterification increases, binding of the CBM77 decreases. In homogalacturonan the esterification of galacturonic acid carboxyl group leads to a decrease of PGA acidity. The methyl group, but more likely the loss of negative charge of PGA, has a major impact on CBM77$_{PL1/9}$ binding. It is possible to speculate that in CBM77$_{PL1/9}$ the positive residues (K1092, K1107 and K1162) that play an integral role in ligand recognition, establish ionic interactions with the negatively charged carboxyl groups in PGA. Therefore, when the pectin is esterified the ester groups will mask the negative charge of the polysaccharide and hence CBM77$_{PL1/9}$ will not be able to make ionic interactions with its ligand. This hypothesis, however remains speculative as crystals of CBM77$_{PL1/9}$-ligand complexes could not be generated. Additionally, consistent with Type B CBMs, the thermodynamics of ligand recognition by CBM77$_{PL1/9}$ was driven by changes in enthalpy and the decrease of entropy had a negative impact on affinity (Boraston *et al*. 2004). This enthalpically driven binding results from the predicted polar interactions

established between the negatively charged carboxyl group in PGA (substrate) and the positively charged amino group from the lysines (CBM). The loss of the conformation freedom of the ligand, however, leads to unfavourable changes in entropy (García-Hernández and Hernández-Arana 1999; Flint *et al*. 2005; Georgelis *et al*. 2012).

The Dali search revelead that CBM77$_{PL1/9}$ displays low similarity with CBM35 from *Podospora anserina* (3ZM8) (Couturier *et al*. 2013) and *Cellvibrio japonicus* (2W87) (Montanier *et al*. 2009) and CBM22 from *Clostridium thermocellum* (2WYS) (Najmudin *et al*. 2010). The ligands targeted by these CBMs are different from the PGA recognized by CBM77$_{PL1/9}$: the *Podospora* CBM binds to mannose, the *Cellvibrio* CBM35 to D-GlcA and unsaturated D-GalA, and CBM22 to xylan (Montanier *et al*. 2009; Najmudin *et al*. 2010; Couturier *et al*. 2013). However, these four structures present the same β-sandwich fold. Similar to CBM77 $_{PL1/9}$, both CBM35 structures present two β-sheets with 4 and 5 β-strands (Montanier *et al*. 2009; Couturier *et al*. 2013). The *Clostridium* CBM also comprises two β-sheets, however, they consist of 5 and 6 β-strands (Najmudin *et al*. 2010). None of these three structures contain the additional 4 β-strands and α-helix that is present in CBM77$_{PL1/9}$. As mentioned before, the pectin binding by the CBM is independent of calcium and no additional ion was detected in the structure. In contrast, the CBM35s and CBM22 contain a structural calcium ion, and the *Cellvibrio* CBM35 presents a second calcium ion that has a direct role in ligand binding (Montanier *et al*. 2009; Najmudin *et al*. 2010). These structural differences associated with the major differences in ligand recognition (mediated by basic residues in CBM77) contributes to a lower similarity between the pectin CBM and the CBMs previously characterized.

The phylogenetics of CBM77 members showed that 17 of the 102 protein sequences (~ 17%) miss one of the key residues implicated in binding (K1092) or the lysine pair (K1136/K1141) identified has having functional redundancy. These proteins evolved from ancestral proteins contaning all the key binding amino acids. It is predicted that these amino acid modifications resulted in the loss of binding to PGA. It is also possible that these proteins evolved to target a different substrate in the context of the plant cell wall. Indeed, all these 17 putative CBMs are modules of proteins containing a PL catalytic module indicating that these binding modules might be functional. However, none of these protein homologues were tested in order to prove the loss of functionality or binding to a diferent substrate. Additionally, members of CBM77 are usually appended to PL1 enzymes in bacteria adapted to different environments. For example, *Ruminococcus* and *Cellulosilyticum* are present in the rumen, *Favobacterium* and *Paenibacillus* were isolated from soil samples and *Bacteroides* and *Eubacterium* are present in the human gut. The CBM77 family is also not specific to the cellulosome. Indeed, *Bacteroides* and *Flavobacterium* are Gram-negative bacteria that are characterized by presenting a Sus-like system in which CAZymes are located on the bacterial surface, together with a range of glycan binding proteins, and in the periplasm (Anderson and Salyers 1989; McBride *et al*. 2009). These results suggest that CBM77 members appear to play a more general role in pectin degradation that is not specific to ruminococci or cellulosome organization.

Usually, CBMs display ligand specificity that reflects the substrate attacked by the appended enzymes, and thus these accessory modules have an important role in enhancing enzymatic efficiency (Venditto *et al*. 2015). As expected, $CBM77_{PL1/9}$, displays the same binding specifcity as the PLs appended to the CBM. The activity of the two catalytic modules and the CBM decreases with increased esterification. $CBM77_{PL1/9}$, however, did not potentiate the *in vitro* activity of the cognate enzymes.

Interestingly, a recent study showed that a CBM46 only had an impact on enzyme activity against xyloglucan in the context of plant cell wall (Venditto *et al*. 2015). The CBM77 was also tested against plant cell walls but no significant potentiation in activity was observed. This result indicates that the role of this ruminoccocal protein may only be evident in the context of the cellulosome. The *R. flavefaciens* cellulosome does not contain a typical Type A CBM3a in the scaffoldins (Rincon *et al*. 2001), and thus lacks the ability to bind to crystalline cellulose. Additionally, several dockerin-containing proteins displayed binding modules (Berg Miller *et al*. 2009). So, in *R. flavefaciens*, the absence of a CBM3a can be compensated by additional CBM modules present in the cellulosomal proteins. Since $CBM77_{PL1/9}$ is a module of a cellulosomal protein, it is possible that this CBM has a major role in promoting close contact between the extracellular complex and the plant cell wall. Additional, it is also important to note that in the context of plant cell wall this CBM can also have a major role directing these PL enzymes to targeting more recalcitrant pectin structures in plant tissues that may be absent in the tissues analyzed here.

The CBM77 was further tested as a possible probe to label homogalacturonan. The detection of this polysaccharide in the context of plant cell walls is essential to explore their biological role. During the last few years several antibodies against pectins have been characterized (Verhertbruggen *et al*. 2009). However, the production and purification of these monoclonal antibodies is expensive and time-consuming. So the $CBM77_{PL1/9}$ characterized in this Chapter presents the possibility of exploring this binding module as a probe to specifically target un-esterified homogalacturonan. A monomer and a dimer of $CBM77_{PL1/9}$ were utilized to test this hypothesis. The rationale for the dimer construct was based on previous reports showing a significant increase in affinity when CBMs are tethered together (Freelove *et al*. 2001). Unexpectedly, by ITC, the dimer only presented a slight increase in affinity compared to monomer.

However, by ELISA the dimer revealed a 125-fold increase in affinity compared to the monomer. A possible explanation for this difference in binding might be due to the different context of the ligand in these two techniques. It is possible that when the substrate is immobilized (ELISA) the dimer is able to establish stronger interactions because both binding modules of this protein are able to bind to the same or different D-GalA chains. When the ligand is solubilized and the reaction is under agitation (ITC), only one of the dimer binding modules binds to PGA, while the second one failed to interact with the ligand. Additional protein steric constraints which are not overcome by the flexible linker can be also an explanation for why the CBM monomer and dimer present similar affinities by ITC. In the context of the plant cell wall $CBM77_{PL1/9}$ in its monomeric form failed to detect pectin. However, a dimer of this CBM was shown to be a suitable probe to detect homogalacturonan. Indeed, the avidity effect of the $CBM77_{PL1/9}$ dimer resulted in stronger binding to PGA in ELISA assays than the monomeric form of the protein module. Additionally, the binding of the $CBM77_{PL1/9}$ dimer to plant cell walls was abolished when PGA was enzymatically cleaved proving specific targeting for homogalacturonan with a low DE. Since the production and purification of this dimer is relatively easy and less expensive than LM19, the $CBM77_{PL1/9}$_$CBM77_{PL1/9}$ dimer can be an appropriate probe for low DE pectins within the plant cell wall.

3.5 Future work

In this chapter $CBM77_{PL1/9}$ was extensively characterised, however, a structure in complex with ligand is crucial to understand the molecular interactions underlying pectin recognition. Additionally, since, a completely new binding site and novel mechanism of binding was described for $CBM77_{PL1/9}$ the structure of a complex would

disclose not only important features in this CBM family but will also improve knowledge of the mechanisms that mediate protein-glycan interactions.

In order to establish $CBM77_{PL1/9}$_$CBM77_{PL1/9}$ as a probe to detect homogalacturonan with a low DE, further tests should be performed. To understand how specific this protein is, it would be useful to assay dimer binding (by ELISA) against non-pectin negatively charged polysaccharides, such as alginate. Additionally, it would be interestingly to quantify the difference in affinity of the CBM dimer and LM19. By ELISA, $CBM77_{PL1/9}$_$CBM77_{PL1/9}$ displayed an apparent high affinity for pectins with a lower DE. However, further studies on plant cell wall sections are required to establish this dimer as a probe. Indeed, the quantification of the dimer binding vs LM19 binding in the context of plant cell wall would be important to establish if $CBM77_{PL1/9}$_$CBM77_{PL1/9}$ displays the enhanced affinity observed by ELISA.

Chapter 4. Exploring the enzymatic degradation of rhamnogalacturonan-II by *Bacteroides thetaiotaomicron*

4.1 Introduction

Rhamnogalacturonan-II (RG-II) is an extremely complex pectin. In RG-II the D-galacturonic acid (D-GalA) backbone is substituted with six side chains (A, B, C, D, E and F) (Figure 4.1). The glycan contains 13 different sugars and 21 distinct glycosidic linkages (reviewed by O'Neill *et al*. 2004).

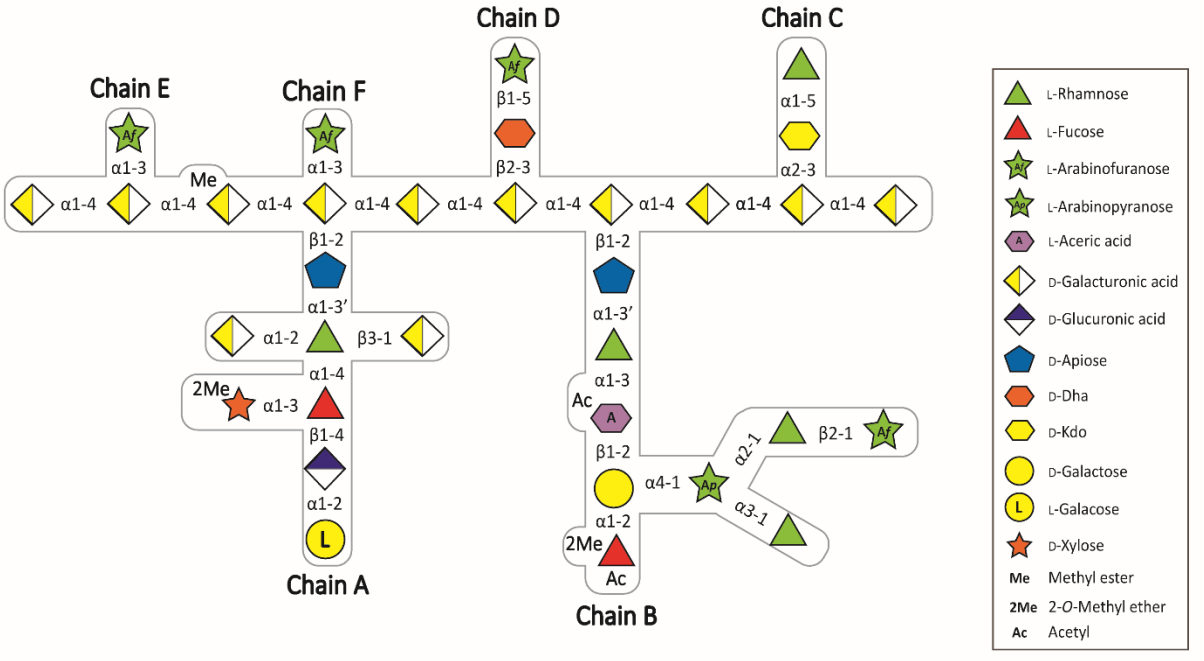


Figure 4.1 Schematic representation of RG-II
Information about the different linkages and side chain composition is represented in the figure. Figure modified with the courtesy of Dr Artur Rogowski (ICaMB, Newcastle University).

Despite the complexity of RG-II, the polysaccharide is highly conserved among plants and is essential in growth and development (O'Neill *et al*. 2001; Matsunaga *et al*. 2004).

Indeed, RG-II dimerization through the borate di-ester cross linkage between the D-apiose (D-Api) at the base of Chain A was shown to have a major impact on porosity and other physical properties of the plant cell wall (Fleischer *et al*. 1999; Iwai *et al*. 2002; Ryden *et al*. 2003).

RG-II is a minor component of plant cell walls but is present in high concentrations in red wine (100 – 400 mg/L) and fruit juices (Pellerin *et al*. 1996). In wine, this occurs because RG-II is recalcitrant to microbial degradation during the fermentation process. Interestingly, there is no evidence that RG-II accumulates in nature indicating that this polysaccharide can be degraded and used as a carbon source by some organisms. Nevertheless, until now, there has been no report of enzymes able to attack this polysaccharide. However, recently, *Bacteroides thetaiotaomicron*, a member of the human microbiota, was shown to grow on RG-II as the sole carbon source (Martens *et al*. 2011). RG-II upregulated three *B. thetaiotaomicron* polysaccharide utilization loci (PULs) encoding a total of 44 proteins, 18 of these proteins had been predicted as members of CAZy families (Figure 4.2) and thus likely to comprise glycan modifying enzymes. This suggests that *B. thetaiotaomicron* is able to degrade RG-II at least to some extent.

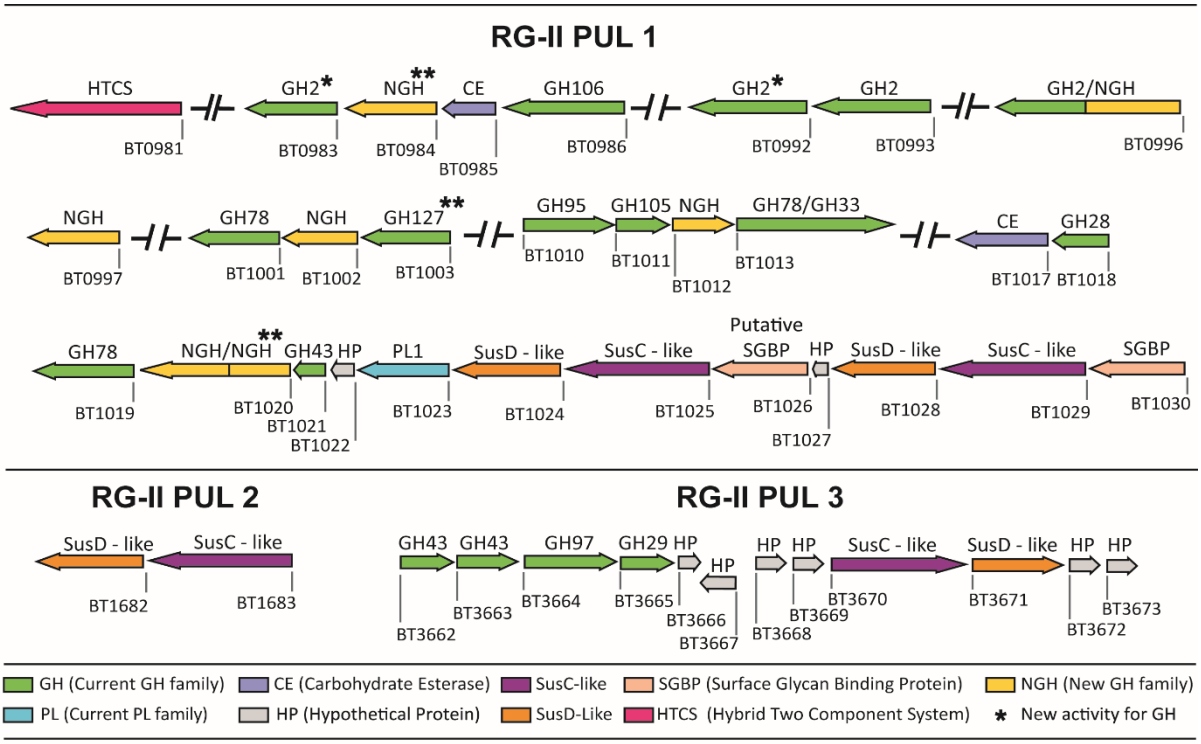


Figure 4.2 *B. thetaiotaomicron* RG-II utilization PULs

The genes encoding predicted enzymes annotated on the CAZy database are represented in different colours with the respective family indicated. Enzymes that display activities not previously reported in the literature are denoted by **. In the schematic of the PULs each gene is drawn to scale as a rectangle with its orientation indicated by the closed triangle. Figure modified with the courtesy of Artur Rogowski.

During the last two years a challenging project started by Drs Artur Rogowski and Didier A. Ndeh (both at ICaMB, Newcastle University) established a model of RG-II degradation by *B. thetaiotaomicron* (Figure 4.3). This project showed that of the 22 distinct glycosidic linkages present in RG-II only one, 2-*O*-Methy-D-Xylose-α1,3-L-Fucose (2-*O*-Me-D-Xyl-α1,3-L-Fuc), was not cleaved by the bacterium (personal statement, Harry Gilbert). Additionally, this human gut bacterium is able to utilize all the sugars of RG-II except D-Kdo, D-Api, L-aceric acid (L-AceA), 2-*O*-Methyl-L-Fucose (2-*O*-Me-L-Fuc) and the disaccharide 2-*O*-Me-D-Xyl-α1,3-L-Fuc. Interestingly, both the side chains and the backbone of RG-II are depolymerized primarily through a hierarchical exo process. Together these data provide the evidence that a single organism is able to almost fully degrade and utilize RG-II.

Previous studies provided initial insights of how *B. thetaiotaomicron* degrades RG-II. There remained, however, several issues with respect to the depolymerisation of the glycan that required clarification. The aim of this chapter is to clarify these ambiguities.

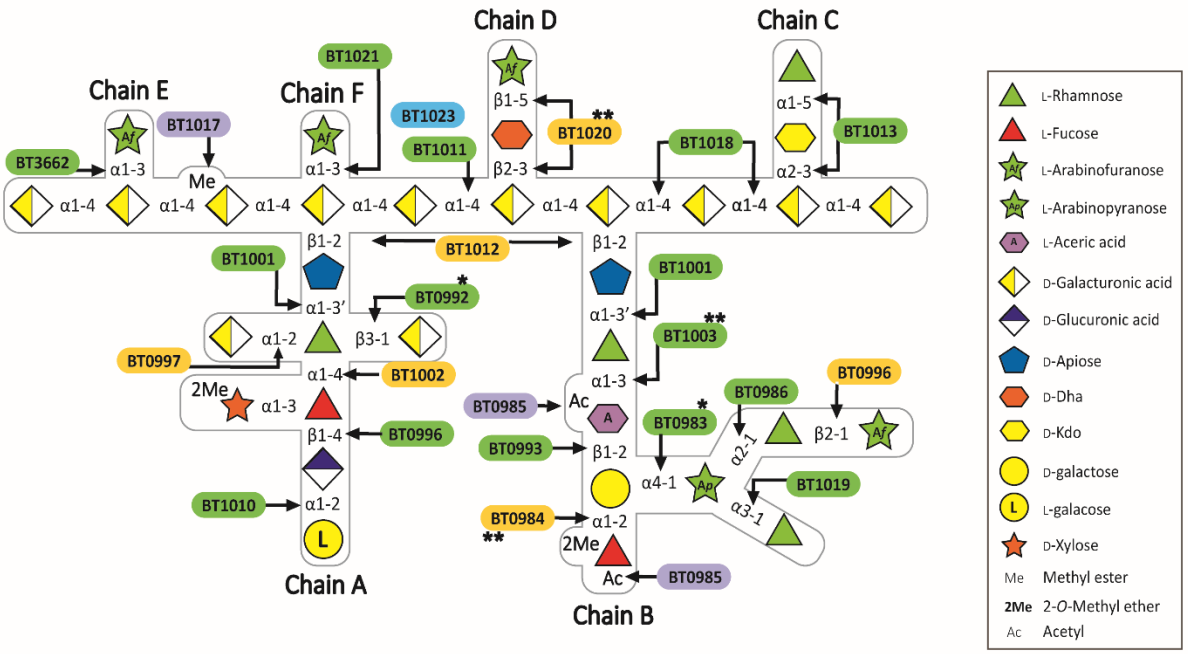


Figure 4.3 Schematic representation of RG-II degradation model

The linkages cleaved by each enzyme are indicated by an arrow. Glycoside hydrolases annotated on the CAZy database are represented in green. Enzymes that present new activities within current CAZy families are labelled with *. Enzymes that have activities never described before in the literature are signed with **. Proteins that are founding members of new CAZy families are represented with a yellow background. Polysaccharide lyase is shown in light blue and the different carbohydrate esterases are represented in purple. Figure modified with the courtesy of Artur Rogowski.

4.2 Objectives

The objectives of this chapter are:

i) Define the linkage specificities of α-L-rhamnosidases BT0986 and BT1019

ii) Test the hierarchical exo model of Chain B degradation of RG-II

iii) Identify the linkage between L-rhamnose (L-Rha) and D-apiose (D-Api)

iv) Establish the sequential order of the cleavage of the L-Rha-D-Api and D-Api-D-GalA linkages at the base of Chain A

v) Explore the specificity of additional RG-II enzymes

vi) Structural characterization of a member of GH106 family (BT0986) and an aceric acidase (BT1003)

4.3 Results

4.3.1 *Protein expression and purification*

All the genes expressed in this chapter were previously cloned into pET28a fused with an N-terminal His_6-tag. The different recombinant proteins were expressed in *Escherichia coli* strain TUNER and purified by Immobilized Metal Ion Affinity Chromatography (IMAC). When BT0986 and BT1003 were used in crystal trials the proteins were further purified by anion exchange chromatography followed by size exclusion chromatography. The purity of protein samples was assessed visually by SDS-PAGE after protein staining. Examples of SDS-PAGE gels of the IMAC purification of selected enzymes are shown in Figure 4.4. Table 4.1 shows the expected size, extinction coefficient, predicted CAZy family and identified enzymatic activity for all proteins mentioned in this chapter.

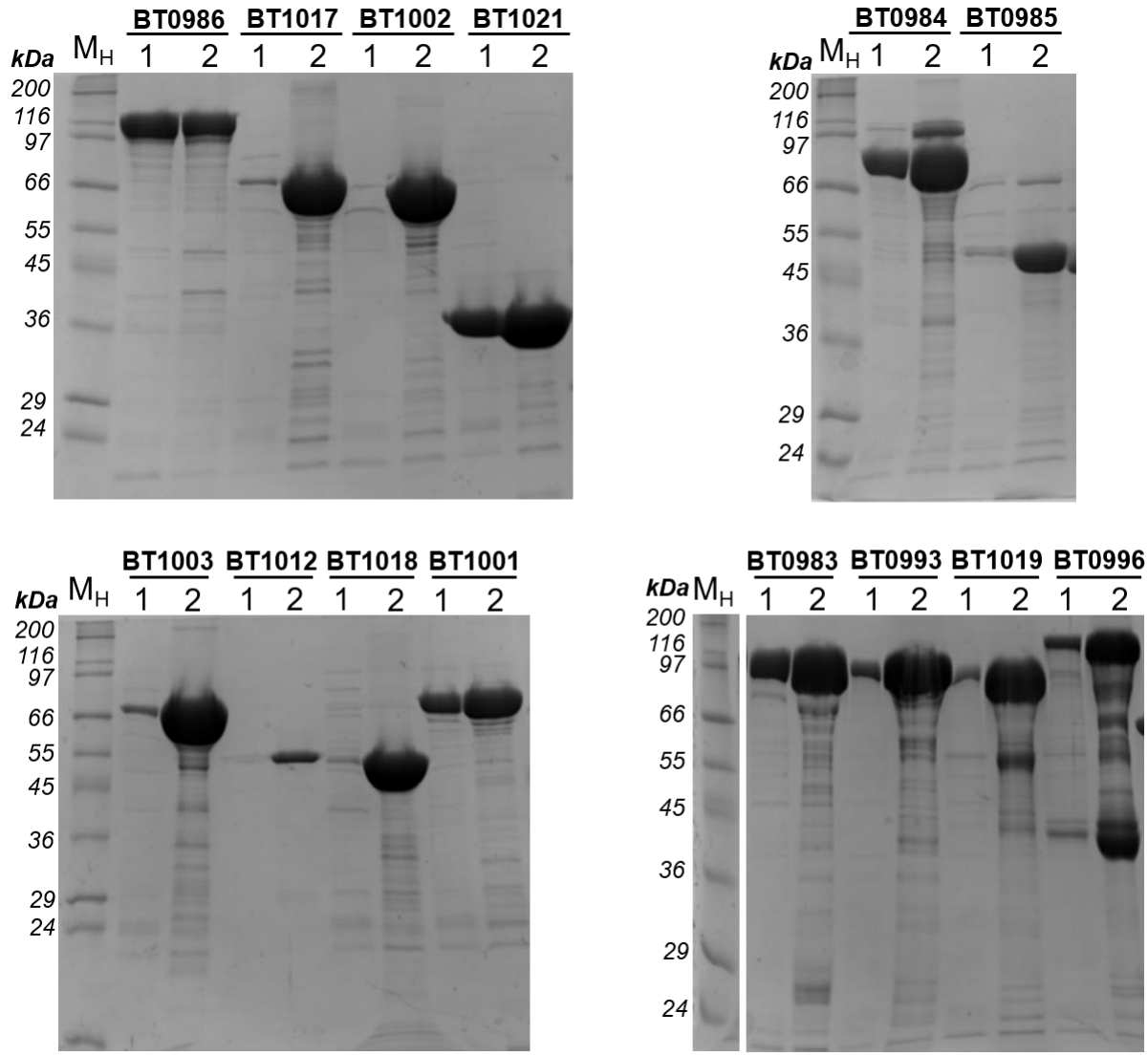


Figure 4.4 Examples of SDS-PAGE gels of IMAC

Example of SDS-PAGE gels (Chapter 2.1.20) of some of the proteins mentioned during this chapter. The proteins are labelled in the figure and the expected sizes are reported in Table 4.1. Lanes labelled 1 and 2 are fractions eluted with 10 mM imidazole or 500 mM imidazole, respectively in Talon Buffer (Chapter 2.1.19.1). The IMAC method is described in Chapter 2.1.19.1. M_r corresponds to molecular weight markers whose sizes are indicated in kDa.

Table 4.1 The details of proteins investigated in this chapter

Protein name	**Expected size** (kDa)	**Extinction coefficient** (M^{-1} cm^{-1})	**CAZy family**	**Enzymatic activity**
BT0983	98	158 765	GH2*****	α-L-arabinopyranosidase
BT0984	90	200 815	NF	α-2-*O*-Me-L-fucosidase
BT0985	51	111 060	NF	*O*-Acetylesterase
BT0986	123	219 975	GH106	α-L-rhamnosidase
BT0993	122	220 520	GH2	β-D-galactosidase
BT0996	160	340 310	NF/GH2	β-L-arabinofuranosidase/ β-D-glucuronidase
BT1001	82	190 040	GH78	α-L-rhamnosidase
BT1002	68	128 855	NF	α-L-fucosidase
BT1003	75	138 660	GH127******	Aceric acidase
BT1012	54	133 285	NF**	Apiosidase
BT1013	145	315 470	GH78/GH33	α-L-rhamnosidase/ 3-deoxy-D-manno-octulosonic-acid (D-Kdo) hydrolase
BT1017	72	75 555	NF	Methyl esterase
BT1018	47	33 055	GH28	Polygalacturonase
BT1019	105	182 480	GH78	α-L-rhamnosidase
BT1020	125	268 310	NF/NF**	β-L-arabinofuranosidase/ 3-deoxy-D-lyxo-2-heptulosaric acid (D-Dha) hydrolase
BT1021	33	62 340	GH43	α-L-arabinofuranosidase

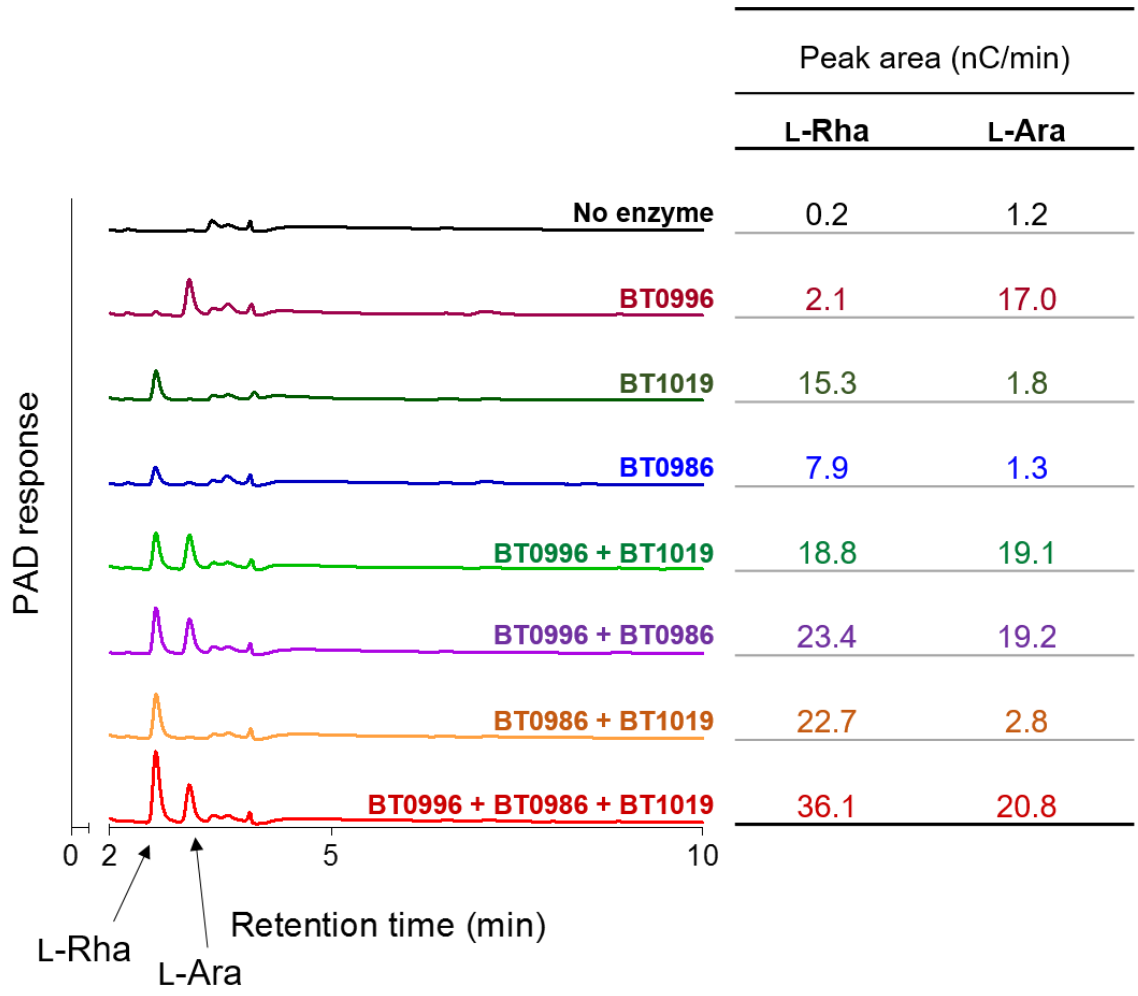
NF - New CAZy family; ***** - New activity inside CAZy family; ** - Enzymes that have activities never described before in the literature

4.3.2 *Define the BT0986 and BT1019 linkage specificity*

BT1019 (GH78) and BT0986 (GH106) are two α-rhamnosidases that were shown to be active against Chain B. Since the non-reducing terminal region of Chain B contains two L-Rha residues linked α-1,2 and α-1,3 to the same L-arabinopyranose (L-Ara*p*), it is important to clarify the linkage specificity of the two enzymes. To explore this question, RG-II was incubated with different enzyme mixtures containing these

rhamnosidases but also BT0996, the β-L-arabinofuranosidase that cleaves the terminal L-arabinofuranose (L-Ara*f*) linked to the L-Rha that is appended to O2 of the L-Ara*p*. The products of these different reactions were analysed by high-performance anion-exchange chromatography (HPAEC). The amount of L-Rha and L-arabinose (L-Ara) released in each reaction was estimated based on the peak area obtained in the HPAEC chromatogram using the software Chromeleon (Figure 4.5).

When incubated with RG-II BT1019 generated a higher amount of L-Rha compared to BT0986. This indicates that BT1019 is active on the L-Rha-α1,3-L-Ara*p* linkage (exposed L-Rha). BT0986 targets L-Rha-α1,2-L-Ara*p*, as most of the L-Rha-α1,2 is capped with β–L-Ara*f* and thus not available to the exo-acting rhamnosidase. To confirm the specificity of BT0986 the activity of the enzyme in the presence and absence of the L-β-arabinofuranosidase BT0996 was determined. The data showed that when the two enzymes were used in combination the amount of L-Rha generated was significantly higher than when only BT0986 was incubated with RG-II. These data indicate that BT0986 activity against L-Rha-α1,2-L-Ara*p* requires the BT0996 cleavage of the L-Ara*f* capping the target linkage. Additional, the total amount L-Rha released by the combination of the three enzymes was similar to the the sum of L-Rha released by BT0996+BT0986 and BT1019, suggesting that BT0986 can tolerate the α1,3-L-Ara*p* linked to L-Rha. Thus, BT0986 targets L-Rha-α1,2-L-Ara*p* and BT1019 cleaves L-Rha-α1,3-L-Ara*p*. Once the capping β–L-Ara*f* has been removed both of the rhamnosidases, individually, can attack Chain B indicating there appears to be no specific order in which BT1019 and BT0986 are active.



	Peak area (nC/min)	
	L-Rha	L-Ara
No enzyme	0.2	1.2
BT0996	2.1	17.0
BT1019	15.3	1.8
BT0986	7.9	1.3
BT0996 + BT1019	18.8	19.1
BT0996 + BT0986	23.4	19.2
BT0986 + BT1019	22.7	2.8
BT0996 + BT0986 + BT1019	36.1	20.8

Figure 4.5 Quantification of L-Rha and L-Ara released by BT0996, BT1019 and BT0986 by HPAEC
The left panel shows the HPAEC chromatogram (Chapter 2.2.4) obtained for each reaction. The sugar detected were identified using appropriated standards indicated with arrows (see Appendix E). The table on right presents the respective peak area for the reaction products L-rhamnose (L-Rha) and L-arabinose (L-Ara). Reactions were performed under standard conditions (Chapter 2.2.5) with RG-II 0.5%, 0.1 mg/ml BSA and 10 mM sodium phosphate pH 7.0.

4.3.3 *Production and purification of oligosaccharides generated by mutants of B. thetaiotaomicron*

The identification of enzymes acting on RG-II indicated an exo-acting model of degradation where the activity of a downstream enzyme required the previous activity of an upstream enzyme. The establishment of this exo-model of degradation was exploited by Didier A. Ndeh who constructed several *B. thetaiotaomicron* strains where the genes encoding specific enzymes involved in RG-II degradation were deleted.

During this thesis these strains will be defined as *Δbtxxxx*, where *Δ* means deletion of the gene *btxxxx*. When grown on RG-II these mutants were predicted to be unable to cleave the linkage targeted by the enzyme encoded by the deleted gene. This led to the accumulation of a specific oligosaccharide. Using this approach and the deletion strains constructed by Didier A. Ndeh the oligosaccharides generated by *Δbt0986* (α-L-rhamnosidase active on Chain B), *Δbt1003* (AceAase active on Chain B) and *Δbt1017* (methyl esterase active on backbone) were produced and purified from the culture supernatant. Figure 4.6 indicates the linkages targeted by the enzymes encoded by the deleted genes and shows the oligosaccharides generated by these strains.

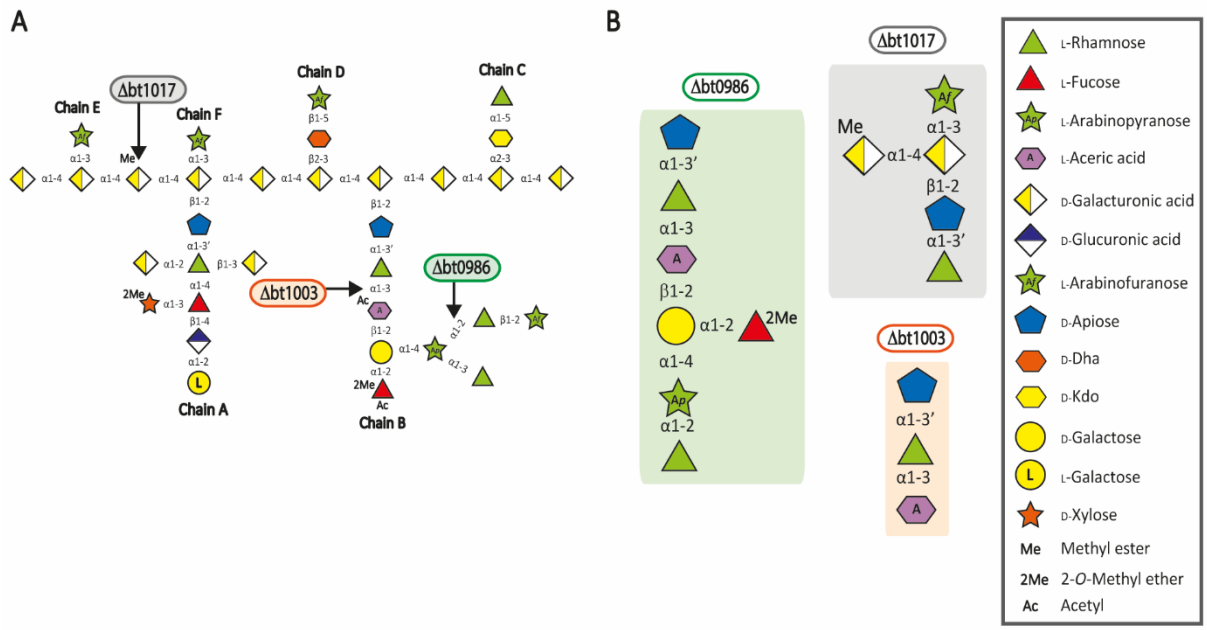


Figure 4.6 Oligosaccharides generated by mutants of *B. thetaiotaomicron*

A. The arrow points the site of action of the enzyme that was deleted in the corresponding mutant. **B.** Schematic representation of the structure of the oligosaccharides generated during this study. Details of the production and purification of these oligosaccharides are described in Chapter 2.2.6.2.

Here the cognate oligosaccharides were purified by size exclusion chromatography (see Chapter 2.2.7) from stationary phase cultures of the deletion strains *Δbt0986*,

Δbt1003 and *Δbt1017* grown on minimal media supplemented with RG-II (see Chapter 2.2.6.2).

4.3.4 *RG-II degradation by Chain B active enzymes*

Previous to this work, Didier A. Ndeh and Artur Rogowski identified the cocktail of nine enzymes required to degrade Chain B (Figure 4.3). To explore whether degradation was a hierarchical exo process the nine enzymes were sequentially incubated with Chain B (isolated by mild acid treatment of RG-II and supplied by collaborators in France). The data (Figure 4.7, HPAEC chromatograms 1 to 9) showed that the N-terminal β-L-arabinofuranosidase removed the terminal L-arabinofuranose capping the L-Rha-α1,2-L-Ara*p*. The GH78 α-L-rhamnosidase BT1019 cleaved the L-Rha-α-1,3-L-Ara*p* linkage and BT0986 (GH106) released the α1,2-L-Rha (data shown in Section 4.3.2). BT0983 cleaved L-Ara*p*-α1,4-D-Gal linkage, representing the first report of a GH2 enzyme with α-L-arabinopyranosidase activity. BT0984 removed the 2-*O*-Me-L-Fuc exposing the D-Gal that was cleaved by the GH2 enzyme BT0993. The L-AceA was cleaved by the GH127 enzyme BT1003. The last linkage in Chain B was attacked by BT1001. Additionally, Dr Rogowski showed that incubation of Chain B with the acetyl esterase BT0985 resulted in an increase of monosaccharide release. This suggests that this enzyme removes the acetylation of 2-*O*-Me-L-Fuc and L-AceA (Pabst *et al*. 2013).

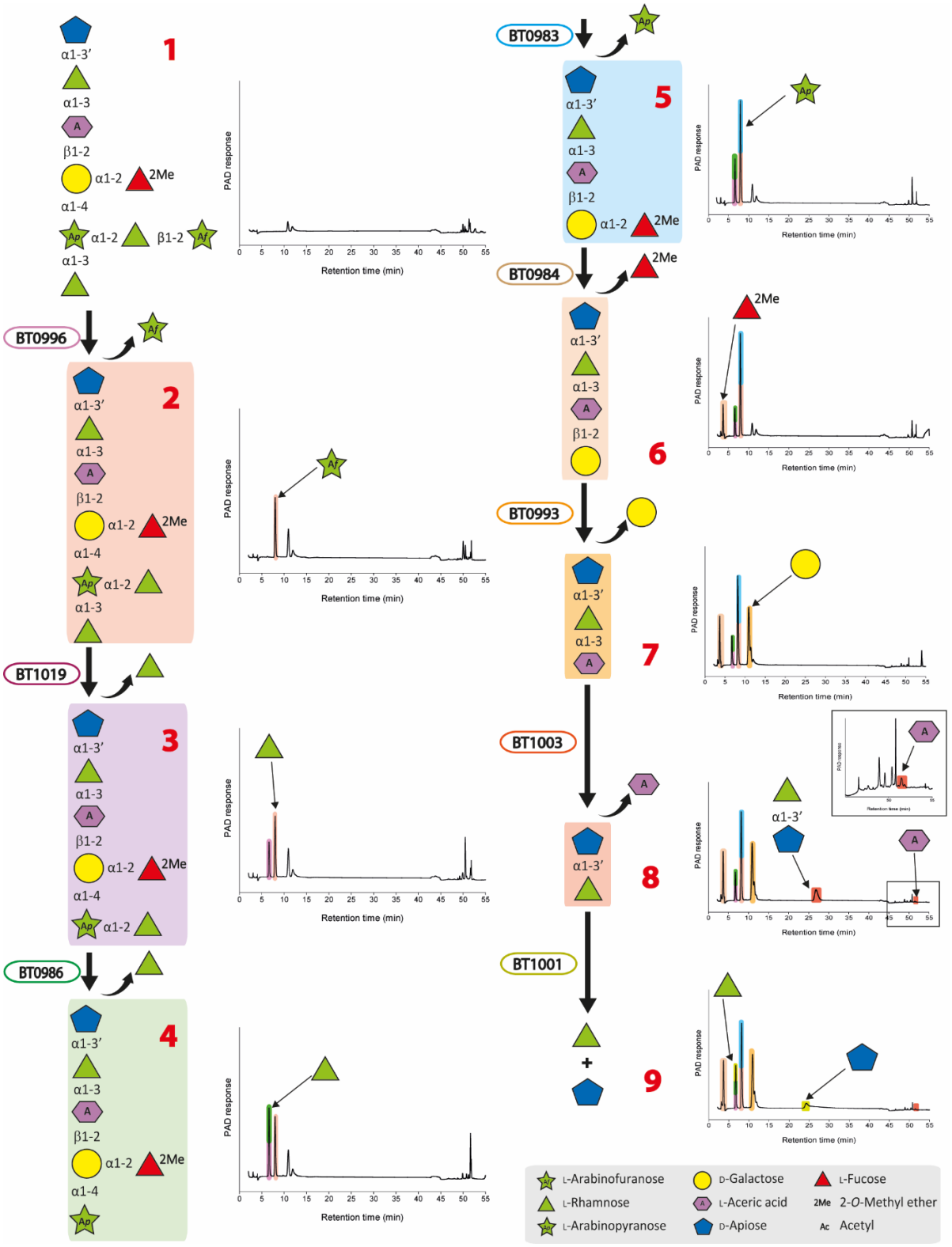


Figure 4.7 Mechanism by which *B. thetaiotaomicron* depolymerizes Chain B of RG-II

Each panel shows the schematic representation of the structure of Chain B and the sugar released by each enzyme (left) detected by HPAEC (right). The colour of the enzyme matches the colour of peak highlighted in the HPAEC chromatogram. The sugar detected were identified using appropriated standards (Appendix E). Reactions were carried out under standard conditions (Chapter 2.2.5) with 0.5% Chain B and 1 μM of enzyme, except for BT0984 (2 μM) and BT1001 (0.2 μM). For HPAEC details see Chapter 2.2.4.

The RG-II exo-active mode of degradation raised the question of whether the enzymes described in previously section were capable of degrading Chain B without its prior release from the backbone. To address this question, the degradation of Chain B in intact RG-II was explored. The polysaccharide was incubated with the nine enzyme cocktail required to degrade Chain B (BT0996, BT1019, BT0986, BT0983, BT0985, BT0984, BT0993, BT1003 and BT1001). The products of this reaction were identified by HPAEC (Figure 4.8). The chromatogram showed L-Rha, L-Ara, 2-*O*-Me-L-Fuc and D-Gal were generated. However, L-AceA and D-Api were not detected indicating that BT1003 and the downstream enzymes cannot access the linkages at the base of Chain B in RG-II.

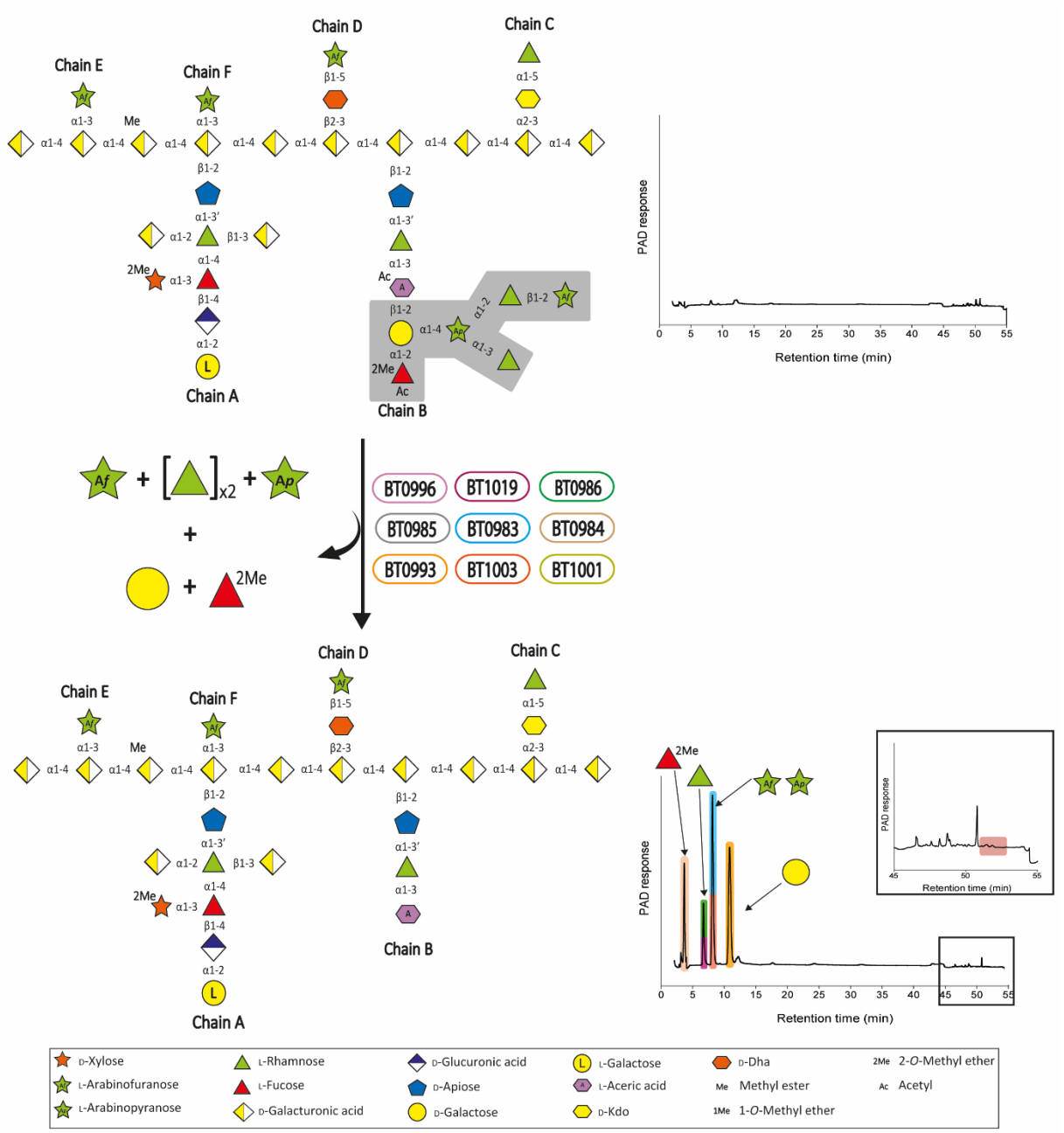


Figure 4.8 Activity of Chain B enzymes against RG-II

Top Panel shows a schematic representation of the structure of RG-II incubated with the enzymes that target Chain B. In grey are highlighted the sugars that were cleaved. Bottom Panel presents the HPAEC chromatograms (Chapter 2.2.4) for RG-II (left) and after treatment of the polysaccharide with the enzymes (right). The sugars detected were identified using appropriate standards (Appendix E) and the peaks are highlighted with the colour that matches the enzyme in the top Panel. The chromatogram region where the L-AceA peak should appear is highlighted inside the black box. The reaction was performed under standard conditions (Chapter 2.2.5) with 2% (w/v) RG-II, 1 μM of enzymes except for BT0983 (2.6 μM), BT0993 (1.8 μM) and BT1001 (0.1 μM).

4.3.5 *Refinement of RG-II structure: L-Rhamnose-D-Apiose linkage*

In RG-II, the terminal linkage of Chains A and B is L-Rha-D-Api. This linkage is cleaved by BT1001, a GH78 α-L-rhamnosidase. In the literature this linkage is described as L-Rha-β1,3'-D-Api (Spellman *et al*. 1983). However, during this project several results suggested that this linkage description was not correct and was indeed L-Rha-α1,3'-D-Api. In this section the evidence for this statement will be provided.

To explore this hypothesis two chemically synthesized oligosaccharides, provided by Dr Sergey Nepododiev (John Innes Center, Norwich) were used. When BT1001 was incubated with chemically synthesized L-Rha-β1,3'-D-Api or L-Rha-α1,3'-D-Api only cleavage of the α-linked disaccharide was observed (Figure 4.9A and B). The catalytic efficiency of BT1001 against L-Rha-α1,3'-D-Api was 7.92 (± 0.1) x 10^6 min^{-1} M^{-1}. Additionally, qualitative data, revealed by the rapid appearance of the yellow coloured 4-nitrophenolate, demonstrated that BT1001 was also active against 4-nitrophenyl-α-L-rhamnopyranose (PNP-α-L-Rha).

Since the previous experiment was performed with chemically synthesized oligosaccharides it was important to confirm that BT1001 was active in the context of RG-II. Thus, the activity of BT1001 against oligosaccharides derived from Chain A and B was evaluated. For the Chain A experiment Didier A. Ndeh kindly provided the purified oligosaccharide 2-O-Me-D-Xyl-α1,3-L-Fuc-α1,4-L-Rha-α1,3'-D-Api generated using the mutant *Δbt1002* (see Figure 4.3 for BT1002 activity against RG-II). The incubation of this oligosaccharide with BT1002 produced two products: 2-*O*-Me-D-Xyl-α1,3-L-Fuc and L-Rha-α1,3'-D-Api. Data presented in Figure 4.9C showed that when these two disaccharides were incubated with BT1001 L-Rha-α1,3'-D-Api was hydrolyzed to L-Rha and D-Api. It is important to note that in the HPAEC chromatogram

the L-Rha peak is not obvious because it virtually co-migrates with 2-*O*-Me-D-Xyl-α1,3-L-Fuc, however, the appearance of the D-Api peak after BT1001 treatment indicated that L-Rha-α1,3'-D-Api was cleaved by this enzyme. Additional data obtained with the *Δbt1017* mutant oligosaccharide (see Section 4.3.6) that contains L-Rha-α1,3'-D-Api (see Figure 4.6) also confirmed that BT1001 is able to cleave this linkage and release L-Rha and D-Api.

To test the Chain B linkage, the supernatant of a culture of the *Δbt1003* mutant was used, which contains the trisaccharide L-AceA-α1,3-L-Rha-α1,3'-D-Api (Figure 4.6) and the limit products not utilized by *B. thetaiotaomicron* 2-*O*-Me-D-Xyl-α1,3-L-Fuc, 2-*O*-Me-L-Fuc and D-Api. The incubation of the trisaccharide with BT1003 generated L-AceA and the disaccharide L-Rha-α1,3'-D-Api. Once more, the presence of BT1003 and BT1001 in the reaction mixture led to the complete degradation of the oligosaccharide to its respective monosaccharides L-AceA, L-Rha and D-Api (Figure 4.9D). It is important to note that L-AceA has a very low PAD response. However, the clear peak for the BT1003 reaction product L-Rha-α1,3'-D-Api indicates that this enzyme hydrolyses the L-AceA-α1,3-L-Rha-α1,3'-D-Api to L-AceA and disaccharide.

In addition, HPAEC analysis showed that chemically synthesized L-Rha-α1,3'-D-Api co-migrated with the disaccharides generated from Chain A (see Figure 4.11.5) and Chain B (Figure 4.9D). These disaccharides presented a retention time of 25.5 min whereas L-Rha-β1,3'-D-Api had a retention time of 22.4 min. Together, these results provide strong evidence that the linkage is L-Rha-α1,3'-D-Api and not L-Rha-β1,3'-D-Api, as reported previously (Spellman *et al*. 1983). It is also important to notice that in the oligosaccharides from Chains A and B, BT1001 only acted downstream to BT1002 or BT1003, respectively. This is entirely consistent with the proposed exo-active model of RG-II degradation.

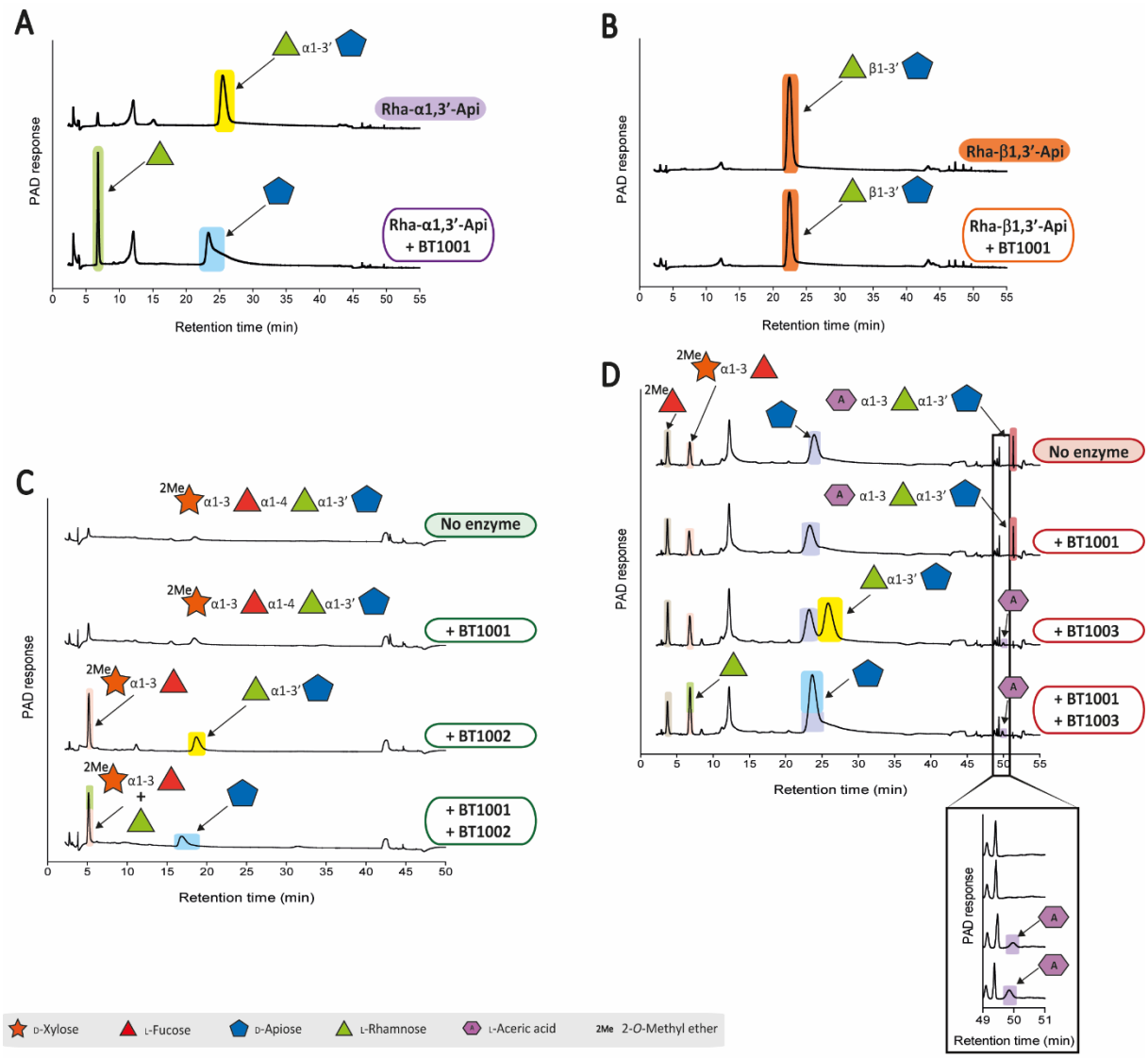


Figure 4.9 BT1001 activity against different linkages of L-Rhamnose-D-Apiose

The figure shows the HPAEC chromatogram for the reactions: (**A**) L-Rha-α1,3'-D-Api incubated with BT1001; (**B**) L-Rha-β1,3'-D-Api incubated with BT1001; (**C**) 2-O-Me-D-Xyl-α1,3-L-Fuc-α1,4-L-Rha-α1,3'-D-Api incubated with BT1001 and BT1002 and; (**D**) L-AceA-α1,3-L-Rha-α1,3'-D-Api with BT001 and BT1003. The L-AceA peak area is highlighted inside the black box. All reactions were carried out under standard conditions (Chapter 2.2.5). For reactions in A and B, 1 mM of substrate was utilized. The concentration of the substrate utilized in C was unknown. The reactions present in D were performed with 50% (v/v) of *Δbt1003* supernatant (see Section 4.3.3) and 1.3 μM of BT1003. The product profiles were analyzed by HPAEC (Chapter 2.2.4) and the sugars detected were identified using appropriated standards (Appendix E).

4.3.6 *The influence of the RG-II backbone on cleavage of L-Rhamnose-α1,3'-D-Apiose by BT1001*

As shown above recombinant BT1001 hydrolyses the terminal L-Rha-α1,3'-D-Api linkage in Chain A and B. This glycoside hydrolase only acts after the side chains upstream of the disaccharide had been removed. It is not known, however, whether the enzyme tolerates the D-GalA appended to the apiose at a potential +2 subsite. To address this question the purified *Δbt1017* oligosaccharide (see Section 4.3.3) was used. When BT1001 alone was incubated with the oligosaccharide no rhamnose was observed by HPAEC and the mass of the oligosaccharide remained unchanged (Figure 4.10B). This suggests that BT1001 was unable to cleave the L-Rha-α1,3'-D-Api linkage in the *Δbt1017* oligosaccharide. The enzyme combination BT1017, BT1018, BT1021 and BT1012 generated D-GalA, L-Ara and L-Rha-α1,3'-D-Api (Figure 4.10C). When BT1001 was added to the enzyme mixture, the disaccharide L-Rha-α1,3-D'-Api was cleaved and L-Rha and D-Api were generated (Figure 4.10D). These data suggest that BT1001 activity requires the prior action of BT1017, BT1018, BT1021 and BT1012. However, this initial experiment does not define the precise components of the *Δbt1017* oligosaccharide that need to be removed before BT1001 can cleave its target linkage.

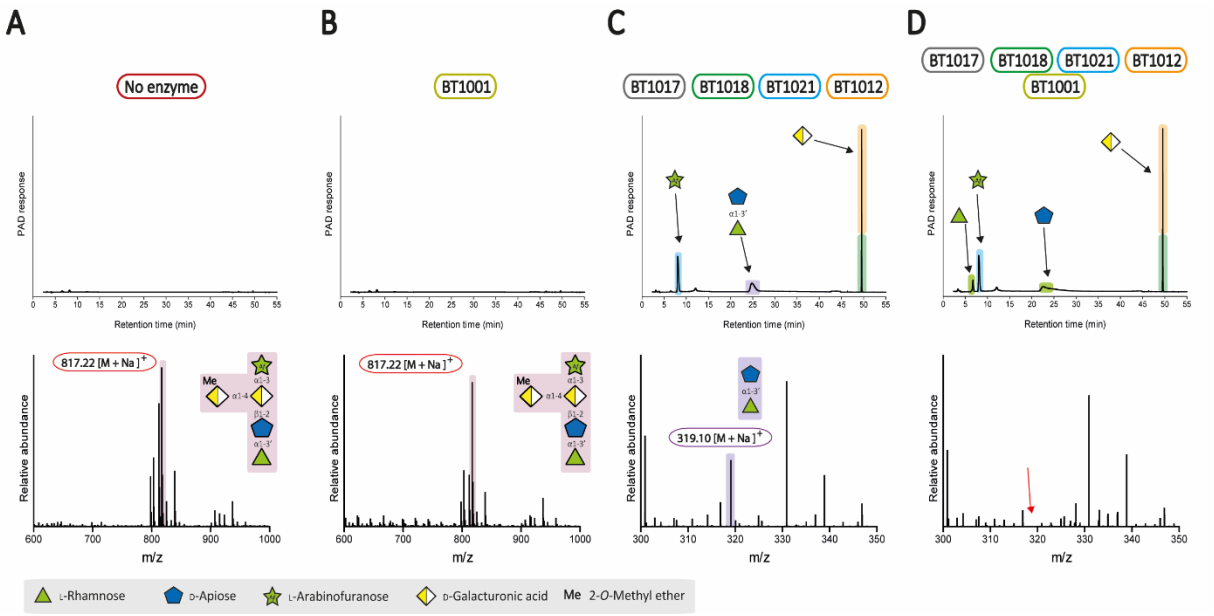


Figure 4.10 Bt1001 activity against *Δbt1017* oligosaccharide

Analysis by HPAEC (top panels) and LC-MS (bottom panels) of the reaction products generated with different enzyme combinations (B, C and D) and the oligosaccharide generated by the *Δbt1017* mutant. Panel A presents the negative control of the reaction. In HPAEC (Chapter 2.2.4) the products of the reactions were identified using appropriated standards. For the LC-MS (Chapter 2.2.9) the m/z of the highlighted peaks represent the sodium adduct of the oligosaccharide detected in positive mode. The oligosaccharide sugar composition is shown for each peak. Except in Panel D, where a red arrow indicates a *m/z* of 319.1(L-Rha-α1,3'-D-Api) was not detected. All reactions were carried out under standard conditions (Chapter 2.2.5) with 2 mM of substrate.

In order to answer the previous question and establish the sequential order of *Δbt1017* oligosaccharide degradation, a matrix of different enzymes combinations was set up in which BT1001 was added or omitted. The products of reactions were analysed by TLC and HPAEC and the results are summarized in Table 4.2. The data showed that BT1017 was required to remove the methyl ester linkage before BT1018 was able to hydrolyse the D-GalA-α1,4-D-GalA linkage. The GH43 BT1021 α-L-arabinofuranosidase was only able to remove the L-arabinofuranose attached to D-GalA after BT1017 and BT1018 had cleaved their target glycosidic bonds. Only then did the apiosidase BT1012 hydrolyse the D-Api-β1,2-D-GalA linkage to generate the disaccharide L-Rha-α1,3'-D-Api. Indeed, if BT1021 was not included in the enzyme

mixture the only reaction product detected by HPAEC was D-GalA (released by BT1018). When BT1001 was added to the various enzyme combinations, the only reaction where the α-L-rhamnosidase showed activity was one that contained BT1012, BT1017, BT1018 and BT1021. This indicates that BT1001 only released L-Rha after BT1012 generated the disaccharide L-Rha-α1,3'-D-Api. This result demonstrated that the *Δbt1017* oligosaccharide is degraded sequentially and enabled a model to be generated in which the five enzymes (BT1017, BT1018, BT1021, BT1012 and BT1001) act in a hierarchical exo-active mode to cleave the five linkages present in this oligosaccharide (Figure 4.11). Critically, the data demonstrate that the backbone D-GalA presents a steric block that prevents BT1001 from cleaving the L-Rha-α1,3'-D-Api linkage.

Table 4.2 Identification of *Δbt1017* oligosaccharide reaction products

Enzymes in reaction mix	Reaction products detected by HPAEC and TLC				
	D-GalA	**L-Ara**	**L-Rha-α1,3'-D-Api**	**L-Rha**	**D-Api**
BT1001	-	-	-	-	-
BT1018	-	-	-	-	-
BT1017 + BT1018	+	-	-	-	-
BT1017 + BT1018 + **BT1001**	+	-	-	-	-
BT1017 + BT1018 + BT1021	+	+	-	-	-
BT1017 + BT1018 + BT1021 + **BT1001**	+	+	-	-	-
BT1017 + BT1018 + BT1012	+	-	-	-	-
BT1017 + BT1018 + BT1012 + **BT1001**	+	-	-	-	-
BT1017 + BT1018 + BT1021 + BT1012	++	+	+	-	-
BT1017 + BT1018 + BT1021 + BT1012 + **BT1001**	++	+	-	+	+
No enzyme control	-	-	-	-	-

(+) means detected; (++) means an increase of peak area in HPAEC; (-) Not detectable. All reactions were performed under standard conditions with 2 mM of the oligosaccharide. The products of the reactions were analysed by TLC (Chapter 2.2.3) and HPAEC (Chapter 2.2.4). (D-GalA) D-Galacturonic acid; (L-Ara) L-Arabinose; (L-Rha) L-Rhamnose; (D-Api) D-Apiose.

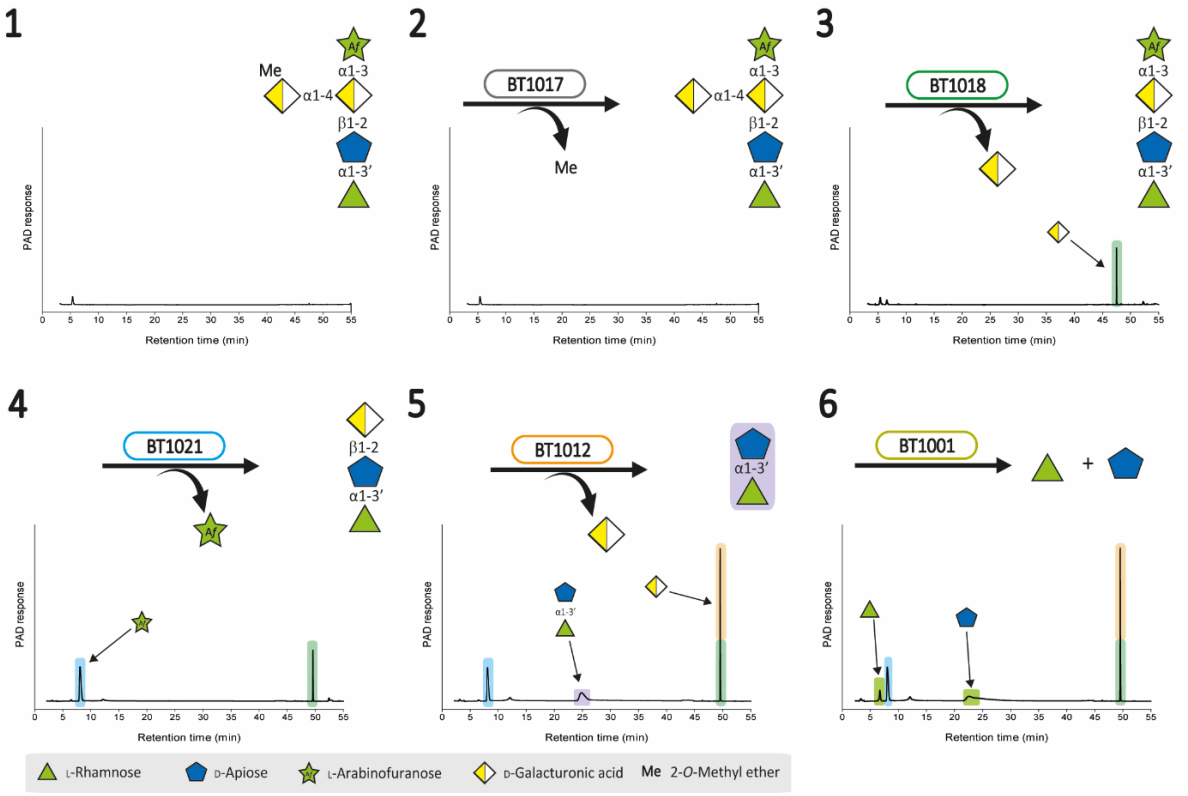


Figure 4.11 Proposed exo-active model of degradation of *Δbt1017* oligosaccharide

Each panel shows the schematic representation of the enzymatic reaction and the products generated by each enzyme (top) detected by HPAEC (bottom). The colour of the enzyme match the colour of peak highlighted in the HPAEC chromatogram. The numbers indicate the order of the reactions. All reactions were performed according to standard conditions (Chapter 2.2.5) with 2 mM of substrate. The sugar detected were identified using appropriated standards (Appendix E). For HAPEC details see Chapter 2.2.4.

4.3.7 *Specificity of enzymes acting on RG-II*

RG-II degradation by *B. thetaiotaomicron* requires the expression of several proteins with similar activities (for example, four α-L-rhamnosidases, two β-L-arabinofuranosidases and two α-L-arabinofuranosidases). These RG-II PULs also encode an α-L-arabinopyranosidase and a β-D-galactosidase that can be active on different substrates/linkages. In an *in vivo* context this bacterium will be present in an environment where different glycans (with different linkages) will be also available. So it is important to establish the specificity of these enzymes not only against specific linkages within RG-II but also against other polysaccharides.

4.3.7.1 α-L-rhamnosidases: BT1001, BT0986, BT1019 and BT1013

RG-II degradation requires the activity of four α-L-rhamnosidases: BT1001, previously described (Section 4.3.5), acts on L-Rha-α1,3'-D-Api at the base of Chain A and B; BT0986 and BT1019 are active at the terminal region of Chain B targeting L-Rha-α1,2-L-Ara*p* and L-Rha-α1,3-L-Ara*p*, respectively (see Section 4.3.2); BT1013 contains N-terminal GH78 and C-terminal GH33 catalytic modules, which are active on L-Rha-α1,5- Kdo and Kdo-α2,3-D-GalA, respectively, present in Chain C (see Figure 4.3). In this thesis the specificities of these enzymes were tested against the different linkages in RG-II and also against L-Rha-α1,4-D-GalA (RG-I backbone) and arabinogalactan (AGP) Gum Arabic that contains the linkages L-Rha-α1,4-D-glucuronic acid (GlcA) and L-Rha-α1,6-D-Gal (Clarke *et al*. 1979). For all reactions the L-Rha released was assessed by TLC and HPAEC. The activities found are summarized in Table 4.3 and details of these experiments will be discussed in this subsection.

To test the activity of the α-L-rhamnosidases against L-Rha-α1,3'-D-Api the chemically synthesized disaccharide was used. The data (Table 4.3 and Figure 4.12) showed that only BT1001 was able to cleave this linkage releasing L-Rha and D-Api. To test the linkage L-Rha-α1,2-L-Ara*p* the oligosaccharide purified from *Δbt0986* was used (see Section 4.3.3). Once more, only BT0986 was able to cleave this linkage releasing L-Rha (Table 4.3 and Figure 4.12). The specificity against the linkage L-Rha-α1,3-L-Ara*p* in Chain B was assessed using intact RG-II. Only BT1019 generated a significant quantity of L-Rha (Table 4.3). To evaluate the capacity of the rhamnosidases to cleave L-Rha-α1,5-Kdo (Chain C) a combination of enzymes was utilized. Previously, Artur Rogowski characterized a catalytic mutant of BT1013 (BT1013 E489A) that did not release L-Rha from RG-II. However, because the mutation was in the rhamnosidase module (GH78), the mutant enzyme should be active against Kdo-α2,3-D-GalA

linkage. So, it was speculated that if a rhamnosidase was non-specific and cleaved the linkage L-Rha-α1,5-Kdo, the BT1013 mutant E489A would be able to hydrolyse the Kdo-α2,3-D-GalA releasing D-Kdo as the reaction product. Thus, a series of reactions were carried out in which each rhamnosidase was combined with wild type BT1013 or with the BT1013 mutant E489A and the release of D-Kdo was assessed by HPAEC. Only when wild type BT1013 was included in the reactions was the D-Kdo peak evident (Table 4.3), and thus only this enzyme was active against the L-Rha-α1,5-Kdo linkage.

The α-rhamnosidases encoded by RG-II PUL1 were also tested against RG-I backbone L-Rha-α1,4-D-GalA oligosaccharides (di- and tetra-oligosaccharide) and against arabinogalactan (AGP) from Gum Arabic, which also contain terminal rhamnose units linked to D-Gal and D-GlcA. For all the four enzymes no L-Rha was detected by TLC or HPAEC (Table 4.3 and Figure 4.12). These results indicate that RG-II α-rhamnosidases BT1001, BT0986, BT1013 and BT1019 target specific linkages in RG-II.

Table 4.3 BT1001, BT0986, BT1019 and BT1013 α-L-rhamnosidases specificities

Enzyme	Linkage and substrate tested					
	L-Rha-α1,3'-L-Api (RG-II Chains A and B)	**L-Rha-α1,2-L-Ara*p*** (RG-II Chain B)	**L-Rha-L-α1,3-Ara*p*** (RG-II Chain B)	**L-Rha-α1,5-D-Kdo** (RG-II Chain C)	**(L-Rha-α1,4-D-GalA)$_n$** (RG-I backbone)	AGP Gum Arabic
BT1001	**+**	**-**	**-**	**-**	**-**	**-**
BT0986	**-**	**+**	**-**	**-**	**-**	**-**
BT1019	**-**	**-**	**+**	**-**	**-**	**-**
BT1013	**-**	**-**	**-**	**+**	**-**	**-**

(+) L-rhamnose (L-Rha) was detected as a reaction product; (-) no L-Rha detected. All reactions were carried out at 37 °C for 16 h (standard conditions) (Chapter 2.2.5) with 1 mM or 1% of oligosaccharides and polysaccharides, respectively. For BT1013 the final concentration was 3.6 μM. The sugar detected by TLC and HPAEC were identified using appropriated standards (Appendix E). The tested linkage is indicated in the table, except for arabinogalactan (AGP) Gum arabic. This polysaccharide contains (L-Rha linked 1,4 to D-glucuronic acid and 1,6 to D-galactose. (L-Api) D-Apiose; (L-Ara*p*) L-Arabinopyranose; (D-Kdo) Ketodeoxyoctonic acid; (D-GalA) D-Galacturonic acid; (D-GlcA) D-Glucuronic acid.

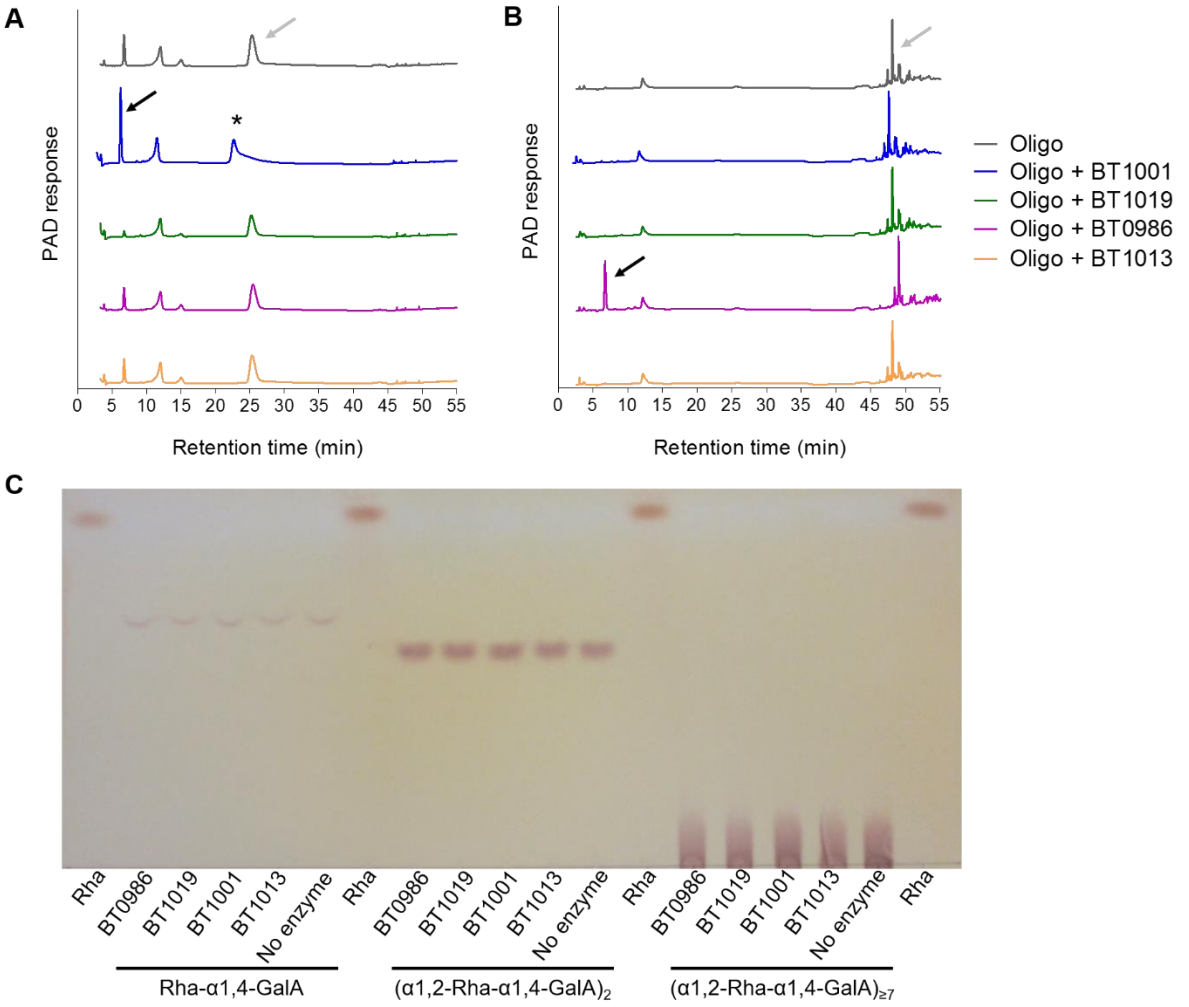


Figure 4.12 Examples of representative α-L-rhamnosidases activity detection by HPAEC and TLC

A. HPAEC of the products generated by the α-L-rhamnosidases BT1001, BT1019, BT0986 and BT1013 against L-Rha-α1,3'-D-Api. **B.** Analysis of the products of the reaction from the oligosaccharide generated by the mutant *Δbt0986* (see Figure 4.6) and the different α-rhamnosidases. **C.** TLC of reactions against RG-I oligosaccharides presenting L-rhamnose at the non-reducing end (see Chapter 5.3.3.1). For **A** and **B** the key is shown at the right of the chromatograms, where Oligo means the substrate and the Oligo reaction (in grey) represents the no enzyme control reaction. In the HPAEC chromatograms the initial substrate is indicated with a grey arrow and the back arrow points to the L-rhamnose released in the enzymatic reaction. In **A**, * indicates the D-apiose peak. For more details about TLC and HPAEC see Chapter 2.2.3 and Chapter 2.2.4, respectively. For more details about these experiments see Table 4.3 legend. (Rha) L-Rhamnose; (GalA) D-Galacturonic acid

4.3.7.2 β-L-arabinofuranosidases: BT0996 and BT1020

In the RG-II model two β-L-arabinofuranosidases were identified: the N-terminal domain of BT0996 (new CAZy family), which removes the non-reducing terminal L-

arabinofuranose in Chain B (L-Ara*f*-β1,2-L-Rha), and BT1020 that contains two catalytic modules active against Chain D. Both modules of BT1020 are founding members of new CAZy families. The N-terminal module of BT1020 comprises a β-L-arabinofuranosidase that cleaves L-Ara*f*-β1,5-D-Dha, while the C-terminal module hydrolyses D-Dha-β2,3-D-GalA. In this experiment an enzyme encoded by the AGP PULs of *B. thetaiotaomicron*, was used as positive control. This enzyme, BT3674, is a GH127 β-L-arabinofuranosidase active against L-Ara*f*-β1,6-D-Gal from Larch Wood AGP. In all experiments TLC and HPAEC was used to detect L-Ara. In reactions where L-Ara was released the enzymes were considered to be active L-arabinofuranosidases. The results are summarized in Table 4.4 and are discussed in detail below.

To explore BT0996 and BT1020 specificity, these enzymes were incubated with RG-II and the amount of L-Ara released was assessed by HPAEC. It was hypothesized that if these enzymes were specific for a different linkage, the L-Ara generated by BT0996 and BT1020 would be additive. The data presented in Figure 4.13 showed the amount of L-Ara released by the BT0996 + BT1020 combination was twice the quantity of the pentose sugar released by BT0996 and BT1020 individually. This result indicated that within the context of RG-II BT0996 and BT1020 target distinct linkages. No activity was detected when these enzymes were incubated with AGP Larch Wood, while BT3674 (encoded by an AGP regulated PUL) released L-Ara from the polysaccharide (Table 4.4). These data show that BT0996 and BT1020 are β-L-arabinofuranosidases that display tight specificity for linkages in RGII.

Table 4.4 BT0996 and BT1020 β-L-arabinofuranosidase specificity

Enzyme	Linkage and substrate tested		
	L-Ara*f*-β1,2-L-Rha (RG-II Chain B)	**L-Ara*f*-β1,5-D-Dha** (RG-II Chain D)	**L-Ara*f*-β1,6-D-Gal** (AGP Larch Wood)
BT0996	**+**	**-**	**-**
BT1020	**-**	**+**	**-**
Positive control BT3674	**NA**	**NA**	**+**

"+" indicates that L-arabinose was detected in the reactions, "-" represents no activity and NA means not assayed. All reactions were performed under standard conditions (Chapter 2.2.5) with 2% (w/v) for RG-II or 0.5% for AGP larch. For BT1020 the final concentration was 4 μM. The L-arabinose detected was identified using appropriated standards (Appendix E). For the different substrates the tested linkage is indicated in the table. (L-Rha) L-Rhamnose; (L-Ara*f*) L-Arabinofuranose; (D-Dha) 3-Deoxy-lyxo-heptulosaric acid; (D-Gal) D-galactose; (AGP) Arabinogalactan.

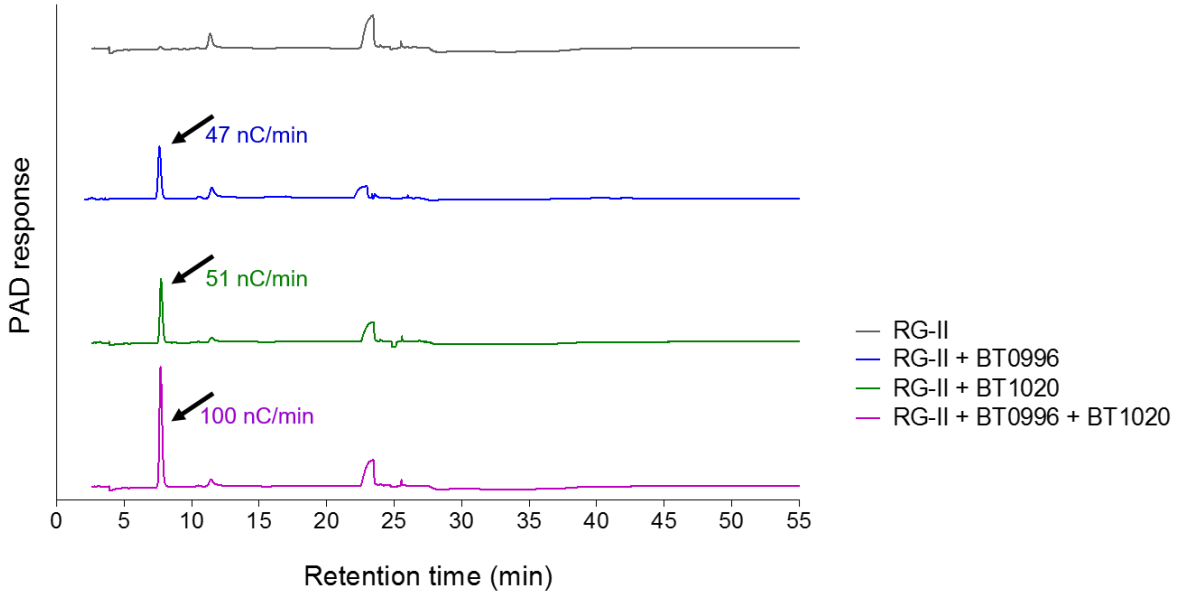


Figure 4.13 Example of β-arabinofuranosidase activity detection by HPAEC

The figure represents the HAPEC chromatogram (Chapter 2.2.4) of the reaction products of BT0996 and BT1020 against rhamnogalacturonan-II (RG-II). The arrow points to the L-arabinose peak and the respective peak area in nC/min is specified for each reaction. In the key RG-II means the negative control reaction and "RG-II +" represents the enzymatic reaction with the respective enzyme. For details of this experiment see legend of Table 4.4.

4.3.7.3 α-L-arabinopyranosidase: BT0983

B. thetaiotaomicron RG-II PUL1 encodes BT0983, a GH2 arabinopyranosidase that is active on the Chain B L-Ara*p*-α1,4-D-Gal linkage. In RG-II there is no additional L-arabinopyranose units. However, AGP Larch Wood contains L-arabinopyranose linked to L-arabinofuranose by an α1,5 linkage. This substrate was utilized to establish if BT0983 is specific for the α-linkage present in RG-II. In this experiment BT3679, an α-arabinopyranosidase encoded by the AGP PUL, was used as a positive control to confirm that L-Ara*p* was released from AGP Larch Wood (Table 4.5). This enzyme was previously characterized by Jose Munoz and is the founding member of a new family. To test the enzymes against RG-II, the glycan was pre-treated with BT0996, BT1019 and BT0986 (see Section 4.3.4) to remove the L-Ara*f* and L-Rha that cap the L-Ara*p* in Chain B (Figure 4.14). When BT0983 was added to this enzyme "mix-1" an increase of L-Ara was detected (Figure 4.14). No L-Ara was detected when BT0983 was incubated with AGP Larch Wood, although BT3679 released the pentose sugar from the glycan (Table 4.5). These data demonstrate that BT0983 is a arabinopyranosidase that specifically targets Chain B of RG-II.

Table 4.5 BT0983 α-arabinopyranosidase specificity

Enzyme		Linkage and polysaccharide tested	
		L-Ara*p*-α1,4-D-Gal (RG-II Chain B)	**L-Ara*p*-α1,5-L-Ara*f*** (AGP Larch Wood)
BT0983		**+**	**-**
Positive control	BT3679	**NA**	**+**

L-Arabinose detection is represented by "+", no L-Arabinose release is indicated with "-" and NA signifies not assayed. Reactions were performed under standard conditions (Chapter 2.2.5) with 2% (w/v) for RG-II or 0.5% for AGP larch wood. The L-arabinose was detected using HPAEC deploying appropriate standards (Appendix E). For HPAEC method see Chapter 2.2.4. (L-Ara*p*) L-Arabinopyranose; (L-Ara*f*) L-Arabinofuranose; (D-Gal) D-Galactose; (AGP) Arabinogalactan.

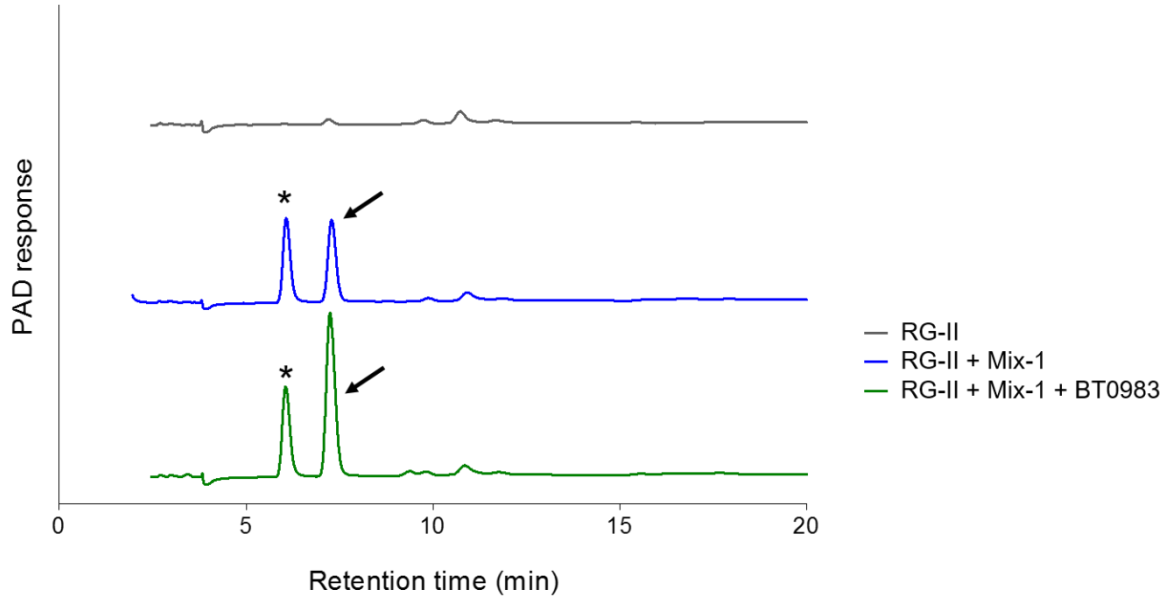


Figure 4.14 Example of α-L-arabinopyranosidase activity detection by HPAEC

The figure shows the HAPEC chromatogram (Chapter 2.2.4) of the products of reaction when RG-II was incubated with BT0996, BT1019 and BT0986 (defined as "Mix-1") without and with BT0983, respectively. The arrow points to the L-arabinose peak and * indicates the L-rhamnose peak. For details of the experiment see legend in Table 4.5.

4.3.7.4 β-D-galactosidase: BT0993

BT0993 is a GH2 D-galactosidase that hydrolyses the D-Gal-β1,2-L-AceA linkage in RG-II Chain B (see Section 4.3.4). To test the specificity of the enzyme, BT0993 was incubated with different oligosaccharides containing D-Gal β-linked to D-Gal or D-glucose (D-Glc). The enzyme was considered active when D-Gal could be detected as a reaction product by HPAEC. The results are summarized in Table 4.6 and examples of HPAEC data are shown in Figure 4.15.

To be active on RG-II, BT0993 requires pre-treatment with the upstream enzymes BT0996, BT1019, BT0986, BT0983, BT0984 and BT0985 (defined as **"**Mix-2**"**; see Section 4.3.4). As expected these enzymes removed all the sugars capping of the D-Gal (Figure 4.15): 2-*O*-Me-L-Fuc, L-Rha and L-Ara. Indeed, Figure 4.15 shows that BT0993 did not release any sugars from RG-II that had not been incubated with Mix-

2. When RG-II was incubated with Mix-2 and BT0993, HPAEC analysis showed that D-Gal was also produced. These data demonstrate, as expected, that the β-D-galactosidase BT0993 cleaves the D-Gal-β1,2-L-AceA linkage once it had been exposed by the action of the upstream enzymes.

BT0993 was also tested against lactose (β-D-Gal-β1,4-D-Glc) and galacto-oligosaccharides with different linkages (β1,3, β1,4 and β1,6) (Table 4.6 and Figure 4.15). The enzyme released D-Gal from all the oligosaccharides tested with exception of the D-Gal-β1,4-D-Gal. This qualitative data indicates that the enzyme could be classified as a non-specific β-D-galactosidase (see Section 4.4 for further details).

Table 4.6 BT0993 β-D-galactosidase specificity

Enzyme	Linkage and substrate tested				
	D-Gal-β1,2-L-AceA (Chain B)	**D-Gal-β1,4-D-Glc** (Lactose)	**D-Gal-β1,3-D-Gal** (Disaccharide)	**D-Gal-β1,4-D-Gal** (Disaccharide or tetrasaccharide)	**D-Gal-β1,6-D-Gal** (Disaccharide)
BT0993	**+**	**+**	**+**	**-**	**+**

(+) D-galactose detected by HPAEC (Chapter 2.2.4); (–) no D-galactose detected. All reactions were performed under standard conditions (Chapter 2.2.5) with RG-II 2% (w/v) or 1 mM of the disaccharides. For reactions using Chain B as the substrate the concentrations of the enzymes were 1 μM for BT0996, BT1019, BT0986 and BT0983 and 5 μM for BT0984. In all reactions 1.8 μM of BT0993 was used. The sugars generated were identified using appropriated standards (Appendix E). (D-Gal) D-Galactose; (L-AceA) L-Aceric acid; (D-Glc) D-Glucose.

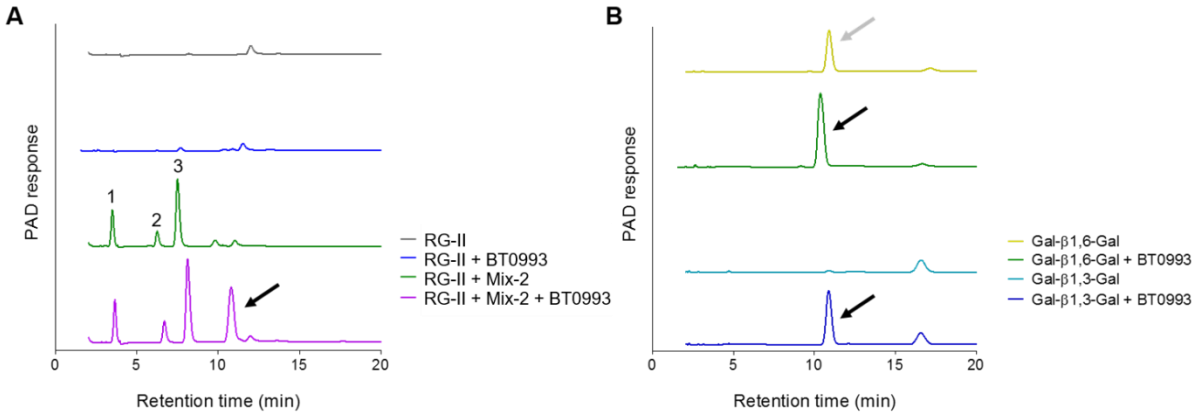


Figure 4.15 Example of β-D-galactosidase activity detection by HPAEC

The figure presents the detection of D-galactose (arrow) by HPAEC when BT0993 was incubated with RG-II (A) or D-galactose (Gal) disaccharides β1,6 and β1,3 (B). In Panel A Mix-2 represents the incubation with BT0996, BT1019, BT0986, BT0983, BT0984 and BT0985. The other sugars detected by HPAEC were 2-*O*-Methyl-L-Fucose (1), L-rhamnose (2) and L-arabinose (3). For further details about these experiments see legend of Table 4.6.

4.3.8 *Structural characterization of BT0986*

BT0986 is an α-L-rhamnosidase that cleaves the L-Rha-α-1,2-L-Ara*p* linkage in the terminal region of Chain B (see Section 4.3.2). This protein is a member of GH106 for which no mechanistic or structural information is available. To overcome this lack of information, BT0986 was crystallised and the structures (apo and complex with D-rhamnopyranose tetrazole) were solved by Dr. Arnaud Basle (ICaMB, Newcastle University).

4.3.8.1 Protein crystallisation

BT0986 apo and selenomethionine (SeMet) were crystallised by sitting-drop vapour-phase diffusion method using the automated Mosquito^{R} nanodrop dispensing system (see Chapter 2.3.2). An equal volume (0.1 µl) of protein and reservoir solutions was mixed and crystals were grown at 20 ºC. The apo crystal was obtained after 4 days

with 10 mg/ml of BT0986 in the commercial Morpheus screen® (Molecular dimensions) condition: 0.1 M amino acids, 50 mM HEPES and 50 mM MOPS pH 7.5 and 15% (w/v) PEGMME 550, 15% (w/v) PEG 20000 and 250 mM of L-Rha (Figure 4.16A). The SeMet crystal was obtained with 10 mg/ml of BT0986 in 0.2 M sodium sulphate, 0.1 M Bis-Tris propane pH 6.5, 20% (w/v) PEG 3350 and 250 mM of L-Rha. This condition was further optimized manually by hanging drop vapour-phase diffusion method (see Chapter 2.3.2). The structure in complex was obtained by soaking the crystals in the optimized condition (Figure 4.16B) with 6 mM of D-rhamnopyranose tetrazole. The soaking technique and further crystal fishing, data collection and structure solution and refinement were performed by Dr. Arnaud Basle (ICaMB, Newcastle University). Briefly, crystals were frozen in cryo-protectant PEG 400 20% (v/v) (apo crystals did not require additional cryo-protection). Crystallographic data were collected at Diamond Light Source Ltd, UK, beamline IO2 for apo and BT0986 with D-rhamnopyranose and beamline IO4 for SeMet. The phase problem was solved by the single anomalous dispersion (SAD) method using the SeMet protein to generate the model that was further used in molecular replacement. BT0986 apo and complex structure were solved at 1.9 Å and 2.1 Å, respectively. Crystal data collection statistics and refinement are reported in Appendix C, Table C.2.

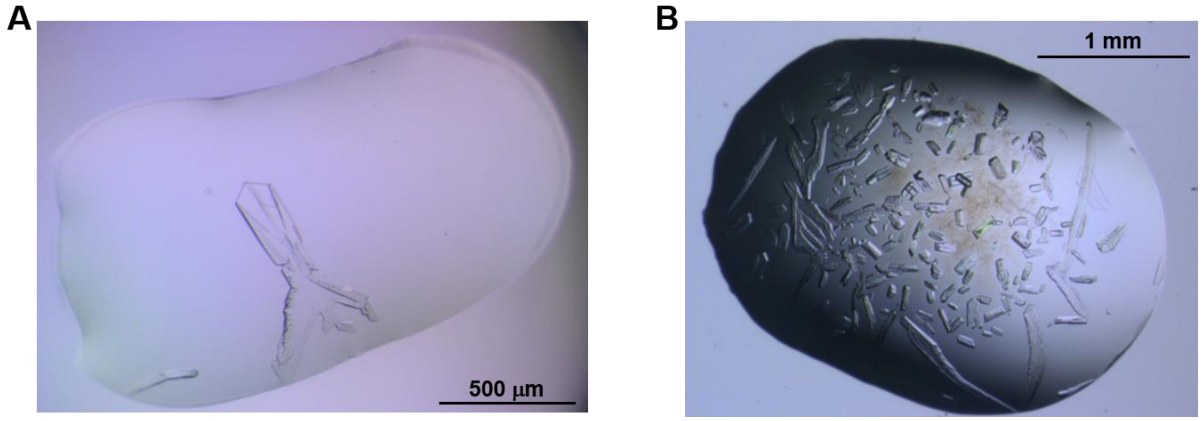


Figure 4.16 BT0986 crystals

A. BT0986 apo crystal obtained by sitting drop vapour-phase diffusion method (see Chapter 2.3.2). **B.** BT0986 crystals obtained by hanging drop vapour-phase diffusion method (see Chapter 2.3.2).

4.3.8.2 Protein structure

The crystal structure of BT0986, comprising ~ 1100 residues, displays six domains labelled from A (N-terminal) to F (C-terminal) (Figure 4.17). Domain A (yellow), extending from A30 to Q676 displays a $(\beta/\alpha)_8$-barrel fold that is interrupted by other structural elements. Thus extending from strand three of the β-barrel is a small domain (B, purple) comprising five antiparallel β-strands (Q141 to S145; T148 to D151; E171 to V180 I367 to L369; L394 to T405), which is followed by a β-sandwich domain (C, blue) comprising Q191 to H328. The two β-sheets in this domain contain five and three antiparallel β-strands. C-terminal to the $(\beta/\alpha)_8$-barrel domain are three predominantly β-sandwich domains defined as D, E and F. Domain D (grey), extending from D683 to G834 comprises two β-sheets each containing three (parallel and antiparallel) β-strands. Seven α-helices and two antiparallel β-strands are in the loops connecting the structural elements of the β-sandwich. In the β-sandwich domain E (green), extending from D838 to A922 with a short β-strand between H679 and V681, the two β-sheets comprises four and six antiparallel β-strands, respectively. C-terminal domain F (pink), comprising T954 to A1101 also displays a β-sandwich domain in which the two β-sheets contain five and six antiparallel β-strands, respectively. Four α-helices are also present in the loops of domain E. Additionally, it is important to notice that it was not possible to build the protein residues between F373 and T385 (domain B) and T947 and N1102). For more details about the protein topology see Appendix F, Figure F.2.

Electron density for D-rhamnopyranose tetrazole, a transition state mimic for α-L-rhamnosidases (Davis *et al*. 1999), was detected in a deep pocket at the centre of the $(\beta/\alpha)_8$-barrel domain, which is thus the candidate active site. An additional ion,

modelled as calcium, is found in the protein proposed active site. This suggests that BT0986 may be an ion-dependent enzyme.

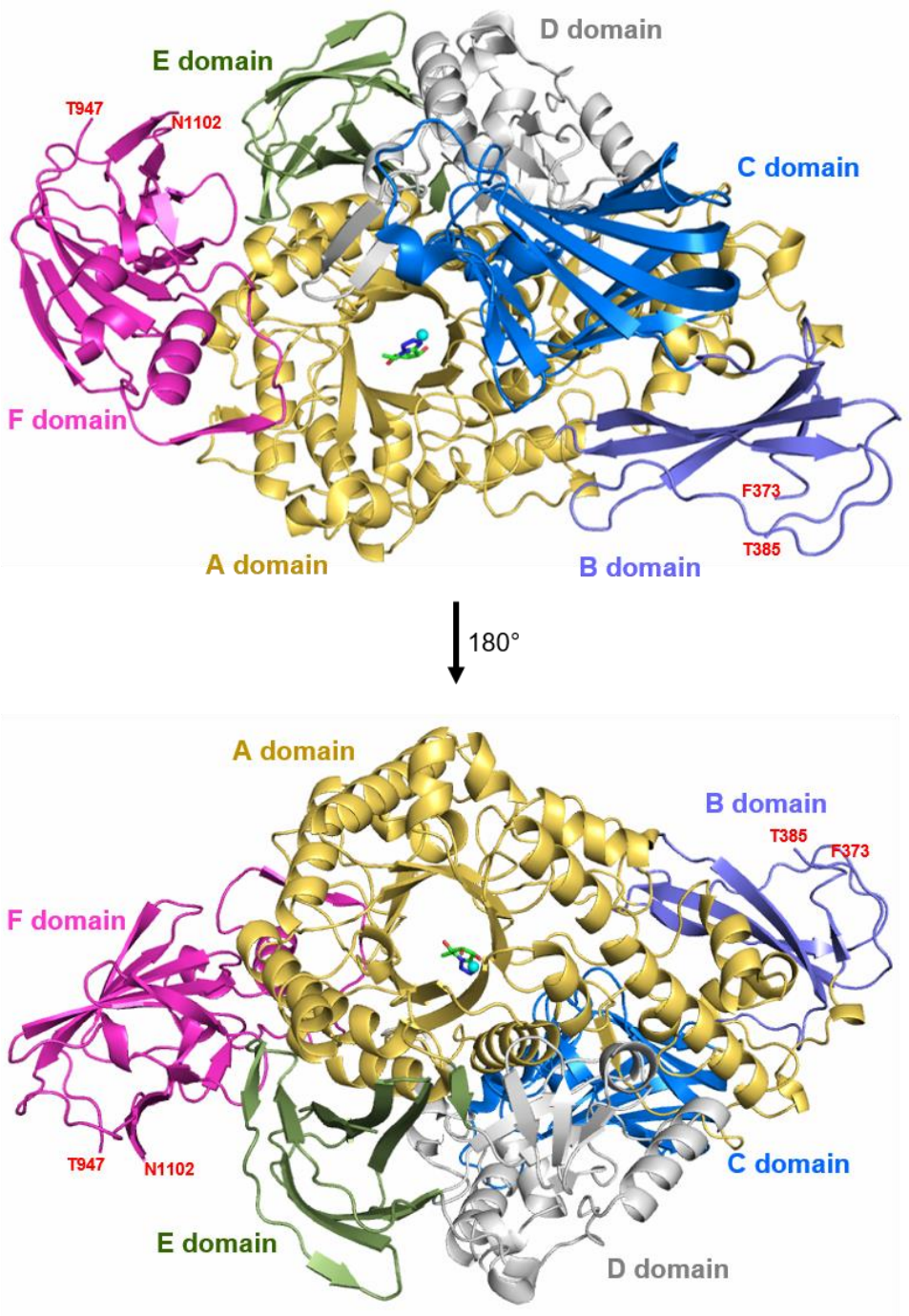


Figure 4.17 BT0986 structure

The figure shows the $(\beta/\alpha)_8$-barrel domain (A, yellow) and the additional two β-stranded domains: B (purple). Three β-sandwich domains C (grey), D (green) and F (pink) are presented in protein C-terminus. The D-rhamnopyranose tetrazole represented as sticks with carbon in green and the calcium ion (cyan sphere) are shown at the centre of the $(\beta/\alpha)_8$-barrel. The bottom structure corresponds to the 180° rotation of the top representation. The amino acids labelled in red indicate the limits of the missing structural information. The figure was generated in PyMOL.

The structural alignment of BT0986 was performed using the online program DaliLite version 3 (Holm and Rosenstrom 2010) and is shown in Table 4.7. The BT0986 structural homologues were members of GH families displaying a catalytic $(\beta/\alpha)_8$-barrel domain. The closest homologue was 3TTY, a β-galactosidase from GH42, with a Z-score of 17.6 and 467 residues aligned of 675 possible residues. However, sequence identity was very low at 12%. The BT0986 structure was superimposed with the structural homologues using the Secondary Structure Matching tool in software WinCoot (Figure 4.18). In all these structural homologues, the active site is located at the centre of the $(\beta/\alpha)_8$-barrel domain (Hidaka *et al*. 2009; Maksimainen *et al*. 2012; Couturier *et al*. 2013), suggesting that in BT0986 the active site will be also located at this location in this domain.

Table 4.7 Structural homologs of BT0986

Domain	GH family of homologue	PDB code	Z-score	RMSD (Å)	Residues aligned	Total residues	% identity
BT0986	GH42	3TTY	17.6	4.2	467	675	12
	GH112	2ZUS	14.2	6.8	419	749	12
	GH5	3ZIZ	14.0	3.7	273	355	12

Z-score (measure of similarity where a value above 2 means a significant similarity and usually corresponds to a similar fold); (RMSD) Root mean square deviation; (% identity) percentage of identity for the structurally aligned region.

Figure 4.18 BT0986 overlay to structural homologues

Overlay of BT0986 structure with the structural homologues. BT0986 is coloured light yellow, GH42 from *Bacillus circulans subsp. alkalophilus* (PDB code 3TTY) in purple, GH112 from *Bifidobacterium longum* (PDB code 2ZUS) in blue, GH5 from Podospora Anserina (PDB code 3ZIZ) in pink. The D-rhamnopyranose tetrazole represented as sticks with carbon in green and the calcium ion (cyan sphere) are shown at the centre of the $(\beta/\alpha)_8$-barrel.

4.3.8.3 The BT0986 active site

Soaking crystals of BT0986 with D-rhamnopyranose tetrazole, an L-rhamnose (6-deoxy-L-mannose) transition state mimic (Davis *et al*. 1999), revealed electron density for the ligand within a deep pocket at the centre of the $(\beta/\alpha)_8$-barrel domain supporting its identity as the active site. The ligand adopted an approximate ^{2,5}Boat ($^{2,5}B$) conformation suggesting that this transition state is required to overcome the destabilizing 'Δ2 effect' (see Section 4.4 for more details).

The BT0986 crystal structure in complex with D-rhamnopyranose tetrazole provides insight into the interactions that lead to substrate recognition in the active site (-1 subsite) (Figure 4.19A and C). Y49 and H627 form hydrogen bonds with O4 of the ligand. S459, which is coordinated with calcium also makes polar contact with the O2 hydroxyl of the rhamnose. An aromatic residue F126 stacks against the sugar ring establishing hydrophobic interactions with C2, C3 and C5. D458, W563 and E593 may also make apolar contacts with the ligand (see Appendix G, Figure G.1 for more details about the protein-ligand interaction). In the active site a calcium ion, which is coordinated with S459, D458, E538, E561 and E593, makes polar contacts with O2 and O3 of the D-rhamnopyranose ring (Figure 4.19A) (see Appendix G, Figure G.2 for more details about the protein-metal interactions).

The glutamates E461 and E593 are positioned close to C1 of rhamnotetrazole (equivalent to the anomeric carbon of rhamnose). Indeed, E461 and E593 are separated by 8.0 Å (Figure 4.19B), a distance for the catalytic residues of an enzyme displaying an inverting mechanism (GH106 mechanism details are shown in Chapter 5.3.3.2). The distances of these two glutamates to the anomeric carbon were also different; 3.4 Å and 5.8 Å for E461 and E593, respectively. Additionally, E593 Oε1 is also interacting with calcium increasing the negative charge of the carboxyl group. This

suggests that E593 is ideally positioned to act as general base and E461 is the general acid. Indeed, E461 is the only candidate that is positionated in a suitable distance to protonate the glycosidic oxygen.

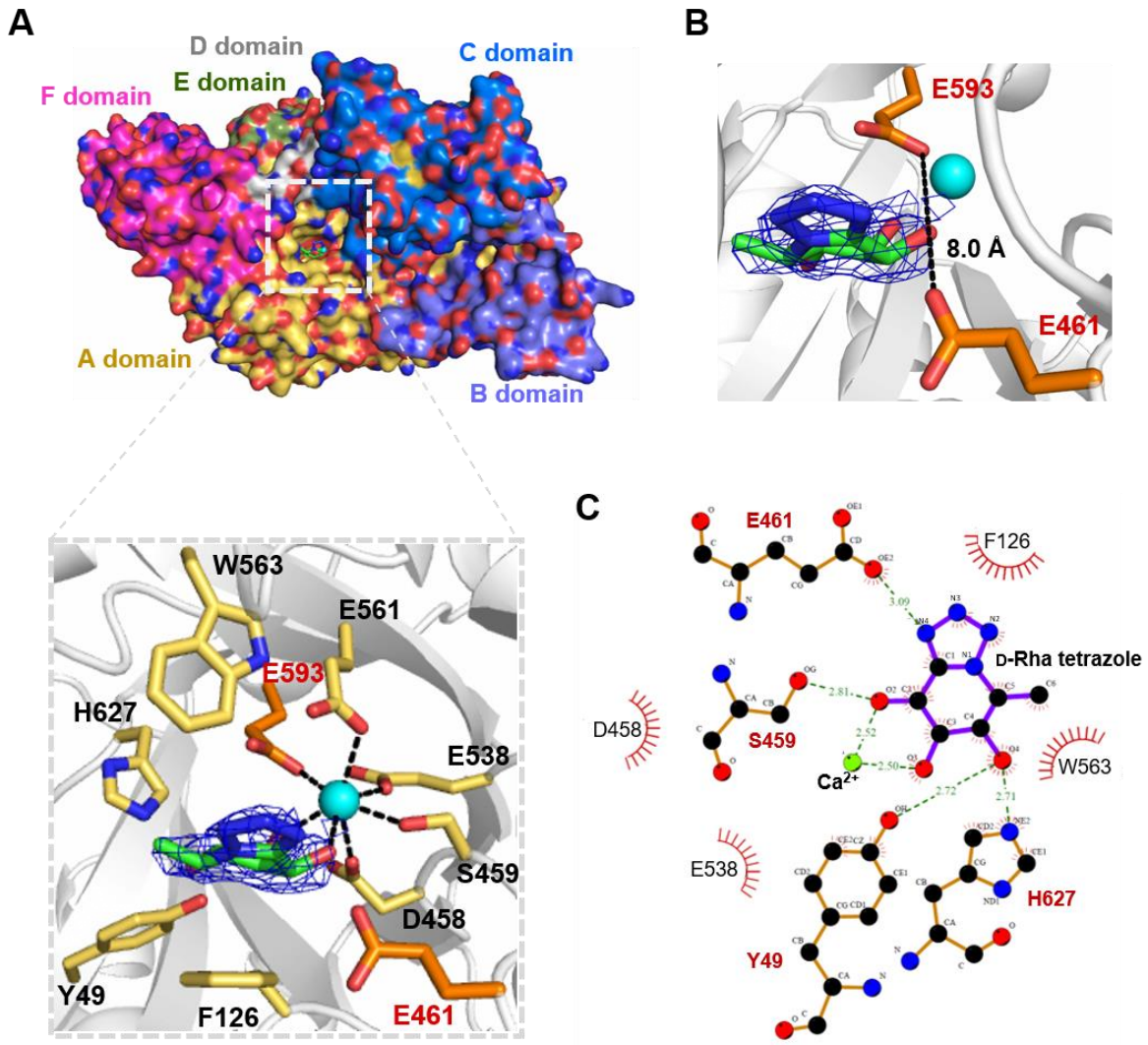


Figure 4.19 BT0986 active site

A. Surface representation of BT0986 showing the different modules (top panel). The active site is highlighted by the grey square (top panel). The catalytic amino acids are shown in orange and yellow are the remaining key residues present in the active site. The calcium ion in the active site is shown as a cyan sphere and its polar contacts indicated by black dashed lines. D-rhamnopyranose tetrazole is represented in green. The blue mesh represents the 2Fo-Fc electron density map (2.1 Å resolution) at 1.5 σ. **B**. Distance between the catalytic residues. **C.** Schematic representation of the interactions involving the ligand D-rhamnopyranose tetrazole (D-rha tetrazole). The green dashes represent hydrogen bonds and the red semi-circles indicate hydrophobic interactions. The diagram was generated using the online software PDBsum (de Beer *et al*. 2014). Panels A, B and C were generated in PyMOl.

4.3.8.3.1 Site direct mutagenesis of BT0986 active site residues

To explore the role of BT0986 residues implicated in substrate recognition in the active site (Figure 4.19) appropriate amino acids were mutated and the enzymatic activity determined using the oligosaccharide generated by the *Δbt0986* mutant. These experiments were carried out by Dr. Aurore Labourel (ICaMB, Newcastle University) and are presented in Table 4.8.The catalytic activity of BT0986 was 1.68 x 10^6 min^{-1} M^{-1}. Alanine substitution of all the amino acids identified in Figure 4.19A, resulted in loss of activity, except for E538A, which was ~170-fold less active than the wild type enzyme. This suggests that this residue is important but not essential for calcium coordination. The mutation of Y49 to phenylalanine in only a 13-fold decrease in activity, however, the alanine mutant was inactive. This indicates that this tyrosine is a key residue in stablishing hydrophobic interactions with the substrate at -1 subsite. However, the hydrogen bond between Y49 and L-Rha O4 is not essential for ligand recognition. Mutation of the majority of the calcium binding residues abolished enzymatic activity. A similar result was obtained when the wild type enzyme was assayed in the presence of EDTA (Table 4.8) suggesting that BT0986 is an ion-dependent enzyme. The loss in activity when the proposed catalytic residues (E461 and E593) were mutated is consistent with their suggested roles as the general acid (E461) and general base (E593).

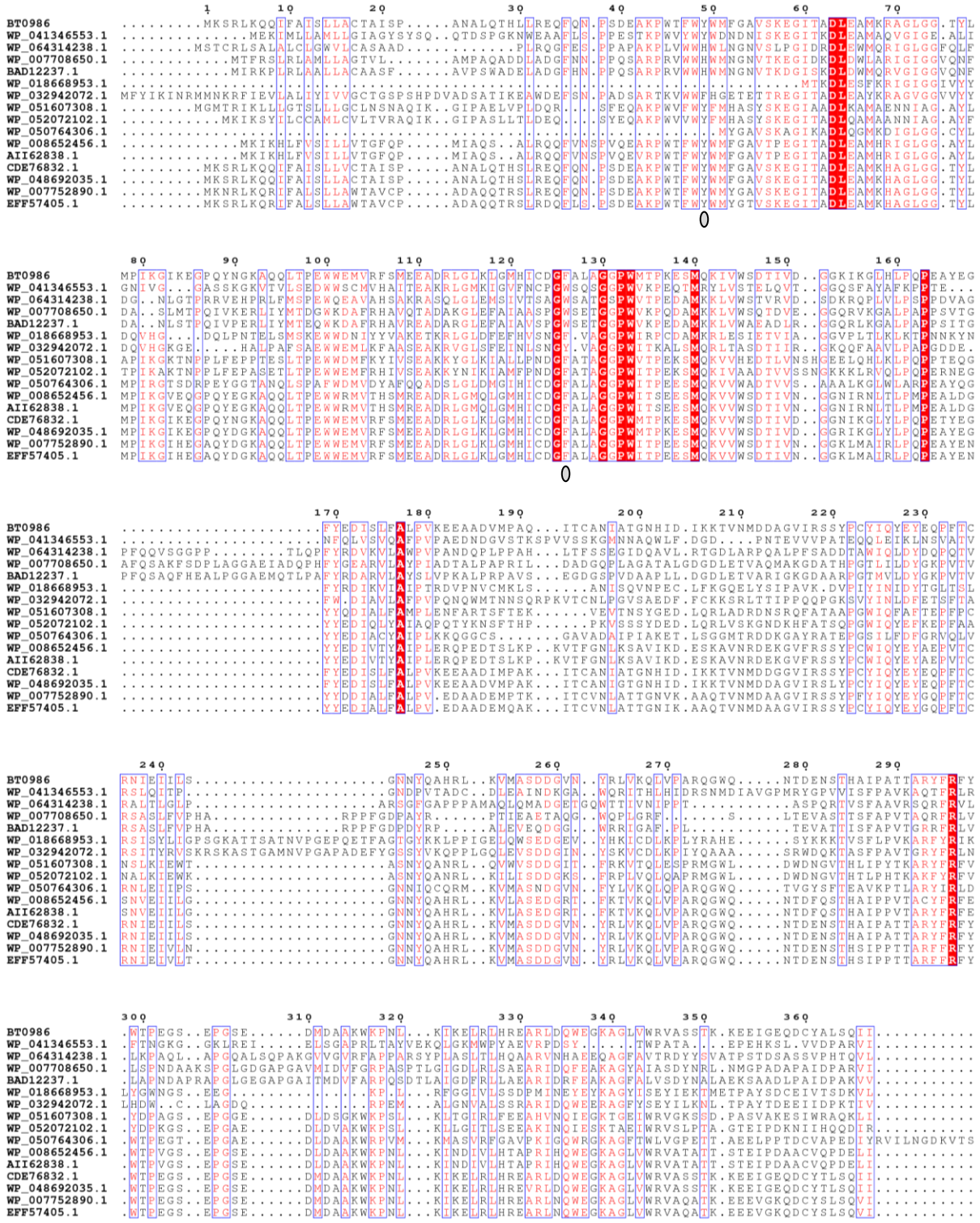
Table 4.8 Catalytic activity of BT0986 mutants

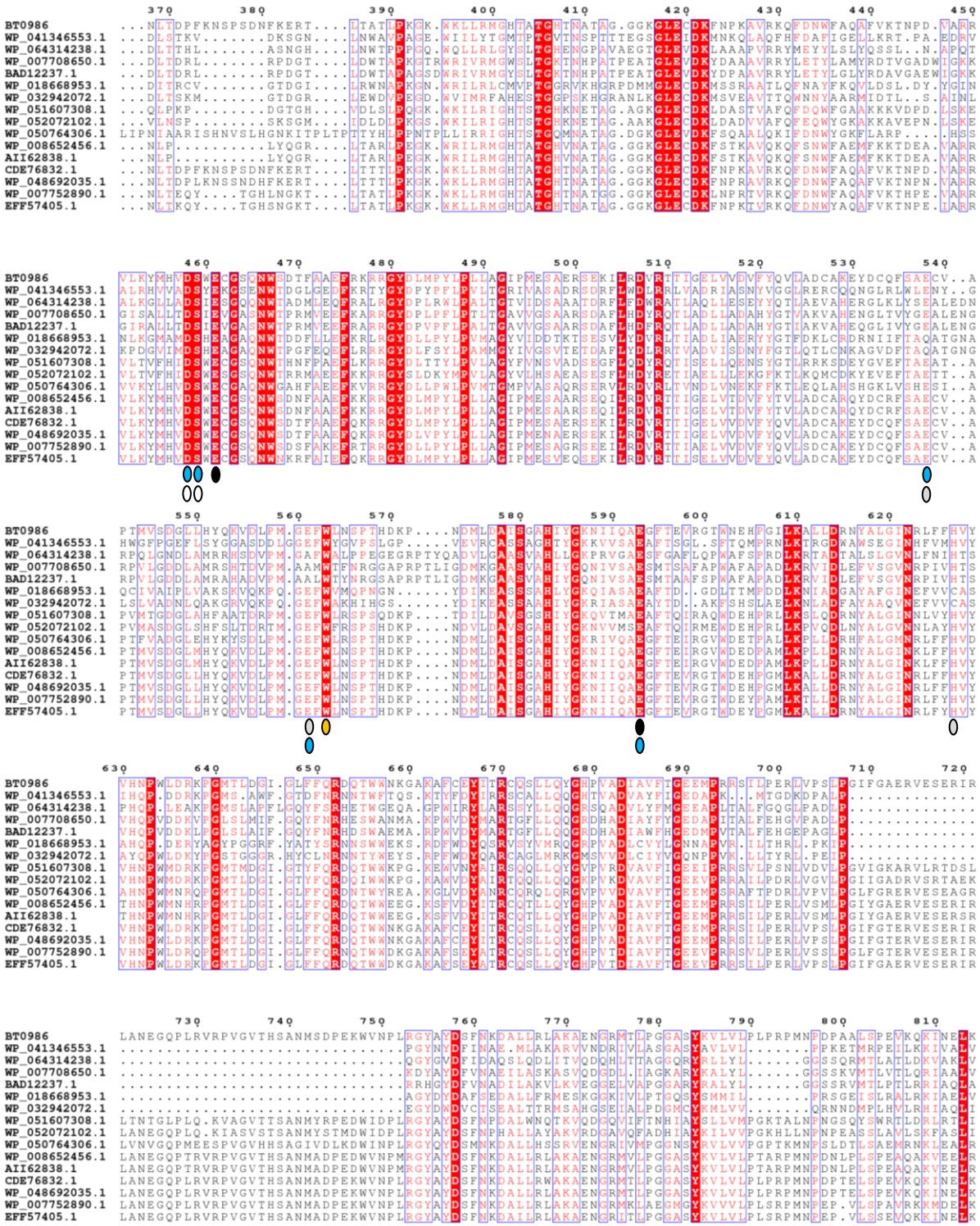
Recombinant protein		k_{cat}/K_M (min^{-1} M^{-1})
Wild type	BT0986	$1.68 \times 10^6 \pm 1.07 \times 10^5$
Mutant	Y49A	NA
	Y49F	$1.29 \times 10^5 \pm 5.62 \times 10^3$
	F126A	NA
	D458A	NA
	S459A	NA
	E461A	NA
	E538A	$1.14 \times 10^4 \pm 6.22 \times 10^2$
	E561A	NA
	W563A	NA
	E593A	NA
	H627A	NA
Wild type + 5 mM EDTA		NA

Catalytic activity of wild type BT0986 and different mutants against *Δbt0986* oligosaccharide (see Section 4.3.3) in 20 mM sodium phosphate buffer pH 7.5 and BSA 0.1 mg/ml. Activities determined by Dr. Aurore Labourel (ICaMB, Newcastle University). NA (not active).

4.3.8.4 Phylogeny of BT0986

In the previous section the active site of BT0986 was described in detail. It is expected that this catalytic apparatus will be highly conserved among members of GH106. To test this hypothesis the sequence of BT0986 was aligned with other members of GH106. The BT0986 sequence was subjected to PSI-BLAST (Altschul *et al*. 1997) analysis to identify sequences with distant evolutionary relationship. The 500 sequences were aligned in MUSCLE (Edgar 2004a; Edgar 2004b). However, in order to simplify only 15 sequences aligned will be shown. These sequences were selected based on the percentage of identity: five sequences below 30 %, three sequences with less than 50% and 6 sequences between 70% and 97%. The alignment file was visualized using the online tool ESPrit 3.0 (Robert and Gouet 2014) and is presented in Figure 4.20. The putative catalytic residues E461 and E593 were invariant in the 500 sequences aligned (black circle). Of the five calcium binding residues (blue circle) described above, only D458, S459 and the catalytic E593 were fully conserved. E538 is a glutamine in two sequences, indicating that the coordination with calcium can occur in these two proteins (see Chapter 5.3.3.3). The key residue E561 was an alanine in two sequences suggesting that these proteins may not be active (as the E561A mutant of BT0986 was inactive) or could target a different substrate. Additionally, Y49, F126, W563 and H627, residues implicated in substrate recognition are not conserved among GH106 proteins. Indeed, only W563 was highly conserved but not invariant in the 500 sequences aligned. This suggests that in some GH106 members substrate recognition might be driven by different residues or these enzymes display substrate specificities that are different to BT0986.





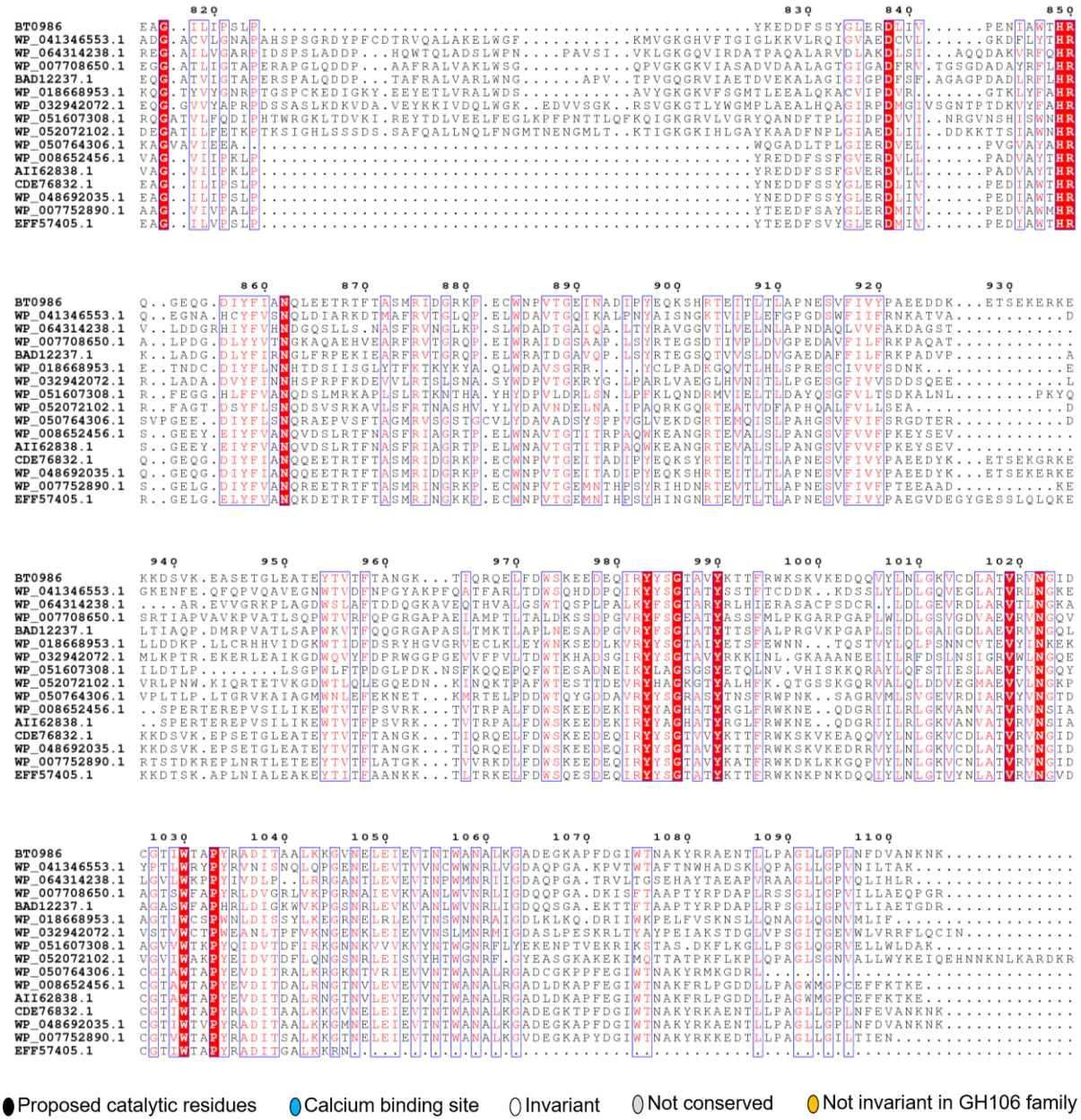


Figure 4.20 Alignment of proteins in GH106 family

NCBI reference codes for each protein are shown at the left. BT0986 represents the α-L-rhamnosidase from *B. thetaiotaomicron*. Sequences were aligned in MUSCLE (Edgar 2004a; Edgar 2004b) and visualized using ESPrit 3.0 (Robert and Gouet 2014). For residues that are conserved in this short alignment however they are not invariant in GH106 family (orange circle) see Chapter 5.3.3.3. Amino acids with 100% of conservation in this alignment are highlighted in a red background and residues that are functionally highly conserved are boxed and coloured red. The residues are numbered according to the BT0986 sequence.

4.3.9 *Structural characterization of the aceric acidase BT1003*

BT1003 is an aceric acidase (AceAase) that cleaves the L-AceA-α1,3-L-Rha linkage (see Section 4.3.4). This enzyme is a member of GH127 that, to date, contains only β-L-arabinofuranosidases (Ito *et al*. 2014). BT1003 not only represents a new activity in GH127 family but also represents the first report in the literature of an AceAase. To try to understand the mechanism of AceA recognition BT1003 was crystallised and the structure was solved by Dr. Arnaud Basle (ICaMB, Newcastle University).

4.3.9.1 Protein crystallization and structure

BT1003 was crystallised using the same method described in Section 4.3.8.1. After 3 days a single crystal was obtained with 10 mg/ml of BT1003 in 1.6 M sodium citrate pH 6.5 and 250 mM of L-Rha (Figure 4.21A). Crystal fishing, data collection and structure solution and refinement were performed by Dr. Arnaud Basle (ICaMB, Newcastle University). Briefly, crystals were frozen in oil. Crystallographic data was collected at Diamond Light Source Ltd, UK, beamline I03. The phase problem was solved by molecular replacement using the coordinates of Protein Data Bank (PDB) 3WKX and 4QJY as models. The BT1003 structure was solved at 2.5 Å. Crystal data collection statistics and refinement are reported in Appendix C, Table C.3. Despite the inclusion of L-Rha in the crystallization condition, no density was detected at the active site and the structure was solved in the apo form.

BT1003 comprises three distinct domains (Figure 4.21B), similar to the crystal structure of the GH127 β-L-arabinofuranosidases from *Bifidobacterium longum* (PBD code 3WRF) (Ito *et al*. 2014). The N-terminal domain (defined as A) comprised residues G49 to Y472 displays a $(\alpha/\alpha)_6$ barrel fold (in green, Figure 4.21B) and, by analogy to the *Bifidobacterium longum* enzyme, is the likely catalytic domain. Three

short α-helixes (H144 to V148, A236 to I240 and P286 to Q290) are present in the loops of the $(\alpha/\alpha)_6$ barrel. In addition, two β-strands (L344 to S349 and Q361 to S366) connected by a loop and a short α-helix (E356 to I360) were present between α-helix 12 and 14. When BT1003 was overlaid with other GH127 structures these β-strands and α-helix were only present in the AceAase (Figure 4.22). In the C-terminal of BT1003 are two β-sandwich domains, defined as B and C. Domain B (in orange, Figure 4.21B) extending from A37 to T47 and A473 to M579 comprises two β-sheets with five and seven anti-parallel β-strands. The C-terminal β-sandwich (in purple, Figure 4.21B) comprises the residues L433 to L435 and P580 to R698. This domain C has two β-sheets with four and six anti-parallel β-strands. For more details about BT1003 topology see Appendix F, Figure F.1. Additionally, it is important to mention that it was not possible to completely determine the structure of two loops/α-helices comprising T69 to D100 and E150 to F189, reflecting very weak density (see Appendix H). To improve structural resolution, several attempts to optimize the crystallisation condition were made. Currently it has not been possible to reproduce or obtain new crystals.

Structural alignment of BT1003 was performed using the online program DaliLite version 3 (Holm and Rosenstrom 2010). Table 4.9 shows the GH family, PBD code and the respective statistics for the three BT1003 structural homologs with a high Z-score. As expected, BT1003 displayed structure similarity to the previously described GH127 structures, 4QJY and 3WRF with a Z-score of 40.3 and 37.9 and sequence identity of 35 and 31%, respectively. Additionally, the AceAase showed structural similarity with a member of GH76 family (4VIR) that presents the same $(\alpha/\alpha)_6$ barrel fold (Thompson *et al*. 2015) (Figure 4.22). However, the sequence similarity to this homologue was only 9% with Z-scores lower than 16.2 and root-mean-square of 3.5 indicating low structural similarity between the two enzymes.

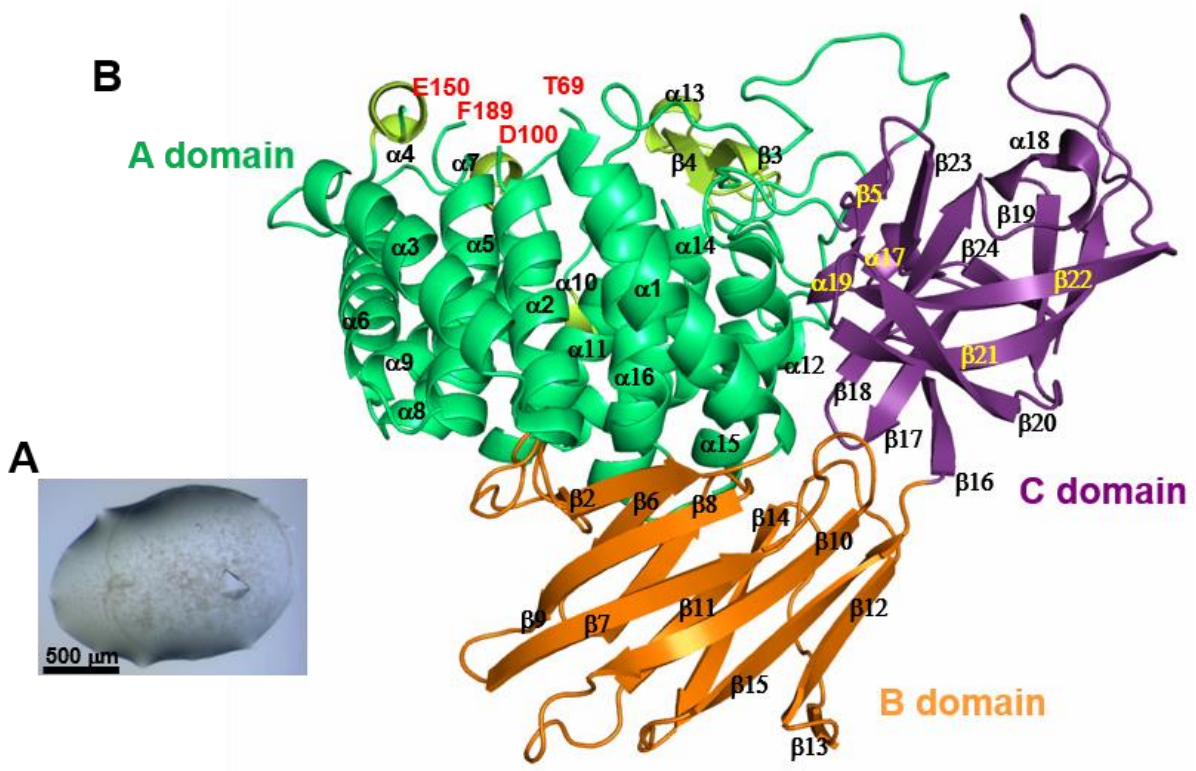


Figure 4.21 BT1003 crystal structure

A. The crystal of BT1003 was obtained in 1.6 M sodium citrate pH 6.5 **B**. Cartoon of BT1003 tertiary structure showing the different domains: A domain (green), B domain (orange) and C domain (purple). The α-helixes and β-strands present in the loops of the $(\alpha/\alpha)_6$ barrel are coloured in lemon. The terminal amino acids of the missing loops are labelled in red. The β-strands and α-helixes are numbered conforming to the BT1003 sequence from the N-terminal (β1 or α1) to the C-terminal (β24 or α19). For more details about BT1003 topology see Appendix F, Figure F.1. The panel was created utilizing PyMOL.

Table 4.9 Structural homologs of BT1003

GH family of homologue	PDB code	Z-score	RMSD (Å)	Residues aligned	Total residues	% identity
GH127	4QJY	40.3	2.1	555	637	35
GH127	3WRF	37.9	2.3	549	639	31
GH76	4V1R	16.2	3.5	288	361	9

Z-score (measure of similarity where a value above 2 means a significant similarity and usually corresponds to a similar fold); (RMSD) Root mean square deviation; (% identity) percentage of identity for the structurally aligned region.

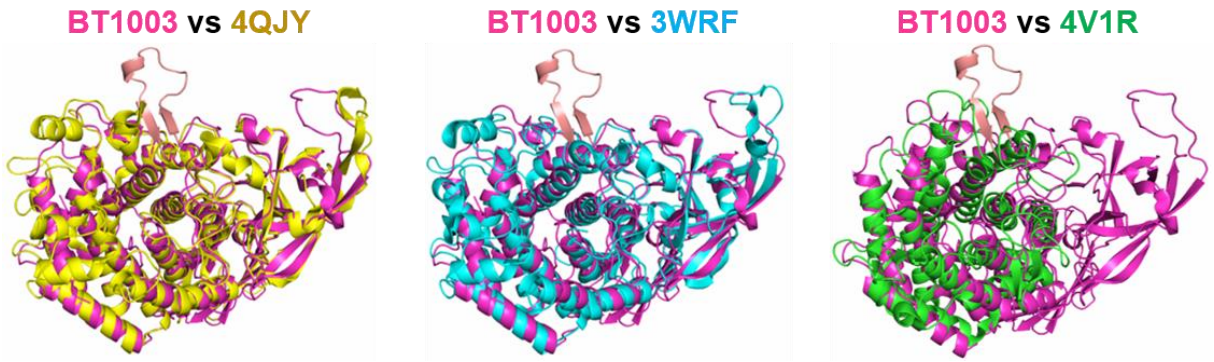


Figure 4.22 BT1003 overlay to structural homologues

BT1003 is coloured in pink, GH127 from *Geobacillus stearothermophilus* T6 (PDB code 4QJY) in light yellow, GH127 from *Bifidobacterium longum* (PDB code 3WRF) in blue, GH76 from *B. thetaiotaomicron* (PDB code 4VIR) in green. In BT1003 structure, the β-strands and α-helix, that are not present in homologue structures are coloured in salmon. BT1003 structure was superposed with the homologue structures using the Secondary Structure Matching tool in software WinCoot. Panel was created in PyMOL.

4.3.9.2 Characterization of BT1003 active site

To explore the conservation of the active site among members of GH127, the BT1003 structure was superposed with the available structures of other GH127 enzymes using the Secondary Structure Matching tool in software WinCoot. The coordinates of the superposed structures were analysed using PyMol (Figure 4.23).

The active site of HypBA1 from *B. longum* was identified by the location of the bound arabinose and the observation that mutation of residues that interact with the pentose sugar led to the loss of catalytic activity (Ito *et al*. 2014) Reflecting the exo-activity of HypBA1 the active site comprises a deep pocket at the centre of the $(\alpha/\alpha)_6$ barrel. An equivalent pocket was located in the *Geobacillus* enzyme (PDB code 4QJY) and BT1003 and was presumed to comprise the active site in these two enzymes (Figure 4.23A). Although the active site location is conserved in BT1003, this AceAase displays differences to the catalytic centre of the two β-L-arabinofuranosidases. The proposed catalytic acid-base is a glutamate in the two β-L-arabinofuranosidase (E322 in HypBA1

and E311 in 4QJY) (Ito *et al*. 2014), however, the equivalent residue in BT1003 is the glutamine Q365 (Figure 4.23A). Another major difference was the absence on zinc (Zn^{2+}) in the active site of BT1003. In HypBA1-L-Ara*f*, the Zn^{2+} coordinates with three cysteines (C340, C417 and C418) and glutamate 338; C417 is the proposed catalytic nucleophile of this "retaining" enzyme (Ito *et al*. 2014). These residues were conserved in BT1003 sequence. However, in this *Bacteroides* structure C457 and C458 (equivalent to C417 and C418, respectively) were shifted away from the active site. As a consequence, in BT1003, the proposed nucleophile C457 presented a different orientation in the active site (Figure 4.23A). Indeed, a similar shift of the proposed nucleophile was reported in the HypBA1 apo structure (Figure 4.23B) (Huang *et al*. 2014). When the *Geobacillus* apo structure (4QJY) was overlaid with HypBA1-L-Ara*f* the conserved cysteines were also in a different location (Figure 4.23B). Additionally, the overlay of these two apo structures with BT1003 failed to co-locate the conserved cysteines in the same region of the active site (Figure 4.23B). These data suggest that the cysteines are located in a flexible loop (B-factor of loop I453 to C458 is 118 $Å^2$) and that ligand binding can induce a conformational modification to the active site. Additionally, in none of the apo structures was it possible to detect a metal ion in the active site, indicating that the substrate might be required to stabilize metal binding. H142, H194, H270 and C415 were identified as key substrate (arabinose) binding residues in HypBA1-L-Ara*f* active site. C415 was conserved in all the GH127 structures (Figure 4.23B). The three histidines were conserved in the *Geobacillus* β-L-arabinofuranosidase. However, BT1003 only contained one of these imidazole amino acids (H296) equivalent to H270 in HypBA1 (Figure 4.23B). Additional, the pKa modulator (E381) was also conserved between the three GH127 structures. These results suggest that in the AceAase substrate recognition is driven by different residues

and the catalytic apparatus may be different from the two GH127 β-L-arabinofuranosidases.

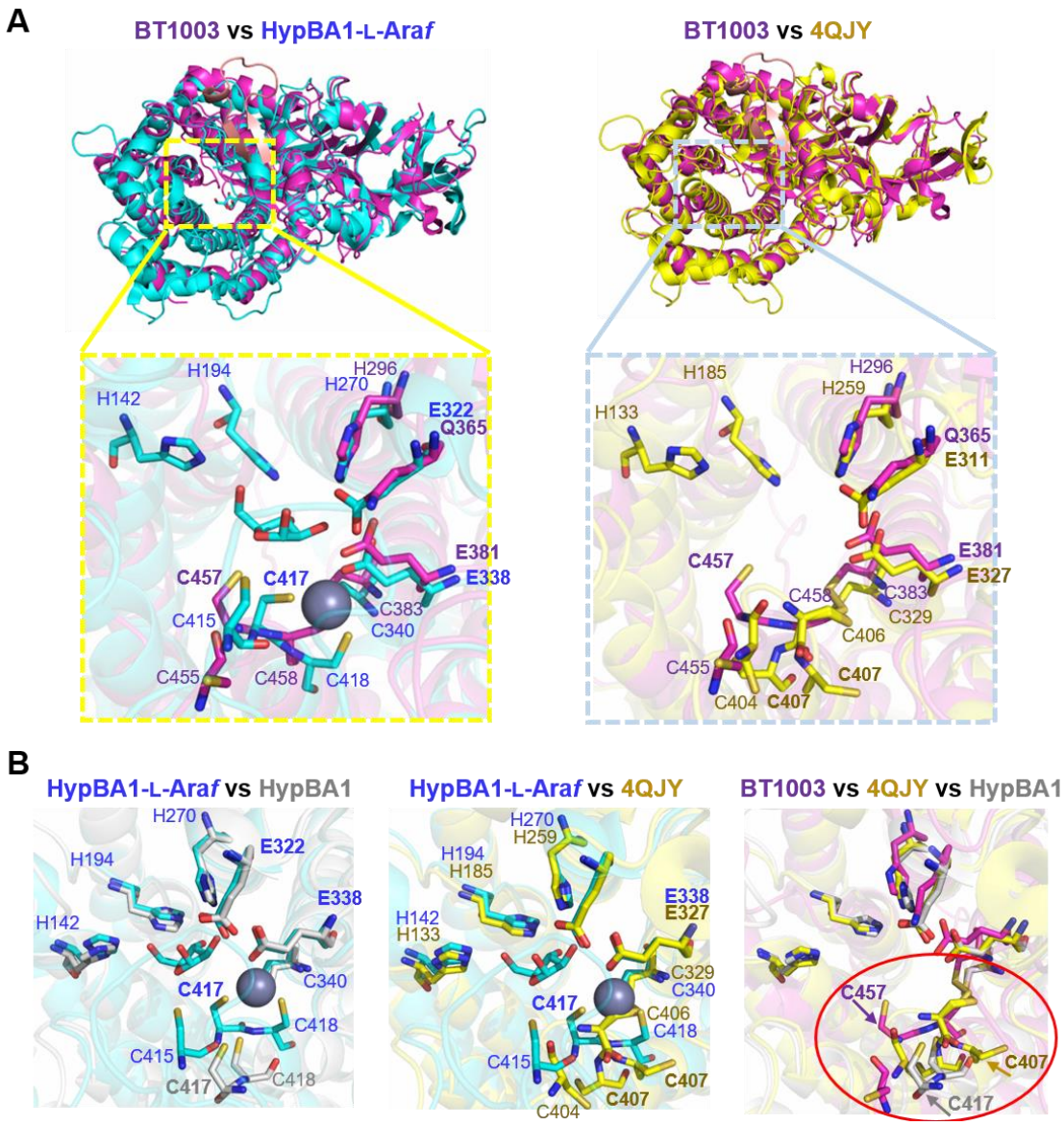


Figure 4.23 BT1003 overlay with GH127 structures

A. Overlay of tertiary structures and key residues present in the active site. **B.** Overlay of the GH127 active site residues. The red circle is highlighting the different positions of the cysteine residues and the proposed nucleophile is labelled. In all panels BT1003 structure and residues are presented in pink. The *B. longum* in complex, HypBA1-L-Ara*f* (PBD 3WKX) and, the apo structure, HypBA1(PBD 3WRF) are represented in blue and grey, respectively. The Geobacillus structure labelled with the respective PDB code 4QJY is shown in yellow. The Zn^{2+} ion (in dark purple) and the β-L-arabinofuranose (in blue) is derived from the HypBA1-L-Ara*f* structure (PBD 3WKX). The panels were created using PyMOL.

4.3.9.3 Site-direct mutagenesis of BT1003 active site residues

Based on the structural overlays described above site-direct mutagenesis was performed to explore the role of BT1003 active site residues in the activity of this enzyme (see Appendix D, Table D.2 for primers and Chapter 2.1.11 for details). The activities of the mutants were assayed qualitatively using the oligosaccharide produced by the *Δbt1003* mutant (see Section 4.3.3). Briefly, BT1003 wild-type ($BT1003^{WT}$) and different mutants were incubated with the *Δbt1003* mutant culture supernatant and the reaction products were analysed by TLC and HPAEC (Figure 4.24). The trisaccharide was cleaved by $BT1003^{WT}$ and C455A releasing L-AceA and L-Rha-α1,3'-D-Api. Additionally, the mutant H296A also displayed low residual activity generating a small amount of L-AceA and the disaccharide. Mutation of the Zn^{2+} binding amino acids (C383 and C458) led to a complete loss of activity. Similarly, mutation of the proposed pKa modulator (E381) and nucleophile (C457) also generated inactive BT1003, indicating that these residues are important in the catalytic mechanism. Interestingly, mutation of Q365 to alanine and glutamate also resulted in the loss of enzyme activity. This residue is a glutamate (general acid/base) in β-L-arabinofuranosidase members of GH127. However, this result suggests that AceAase members of GH127 require a glutamine in this position. These results indicate that the mechanism of glycosidic bond hydrolysis and substrate recognition in the active site of the AceAase are different from the previously described GH127 β-L-arabinofuranosidases.

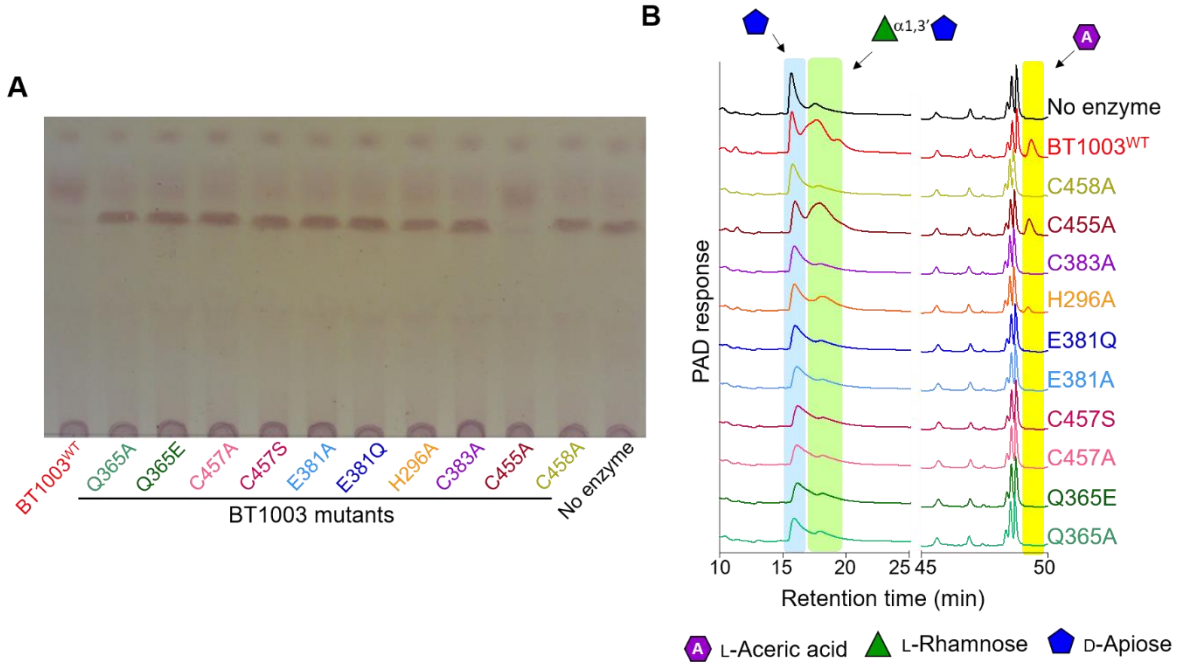


Figure 4.24 Activity of BT1003 mutants

The panels show the reaction products analysed by TLC (A) and HPAEC (B). Reaction were performed with 50% (v/v) Δ*bt1003* mutant culture supernatant, 1 μM of protein under standard conditions (see Chapter 2.2.5). The sugar detected were identified using appropriated standards (Appendix E). For TLC and HAPEC details see Chapter 2.2.3 and Chapter 2.3.4, respectively.

4.3.9.4 Phylogeny of BT1003

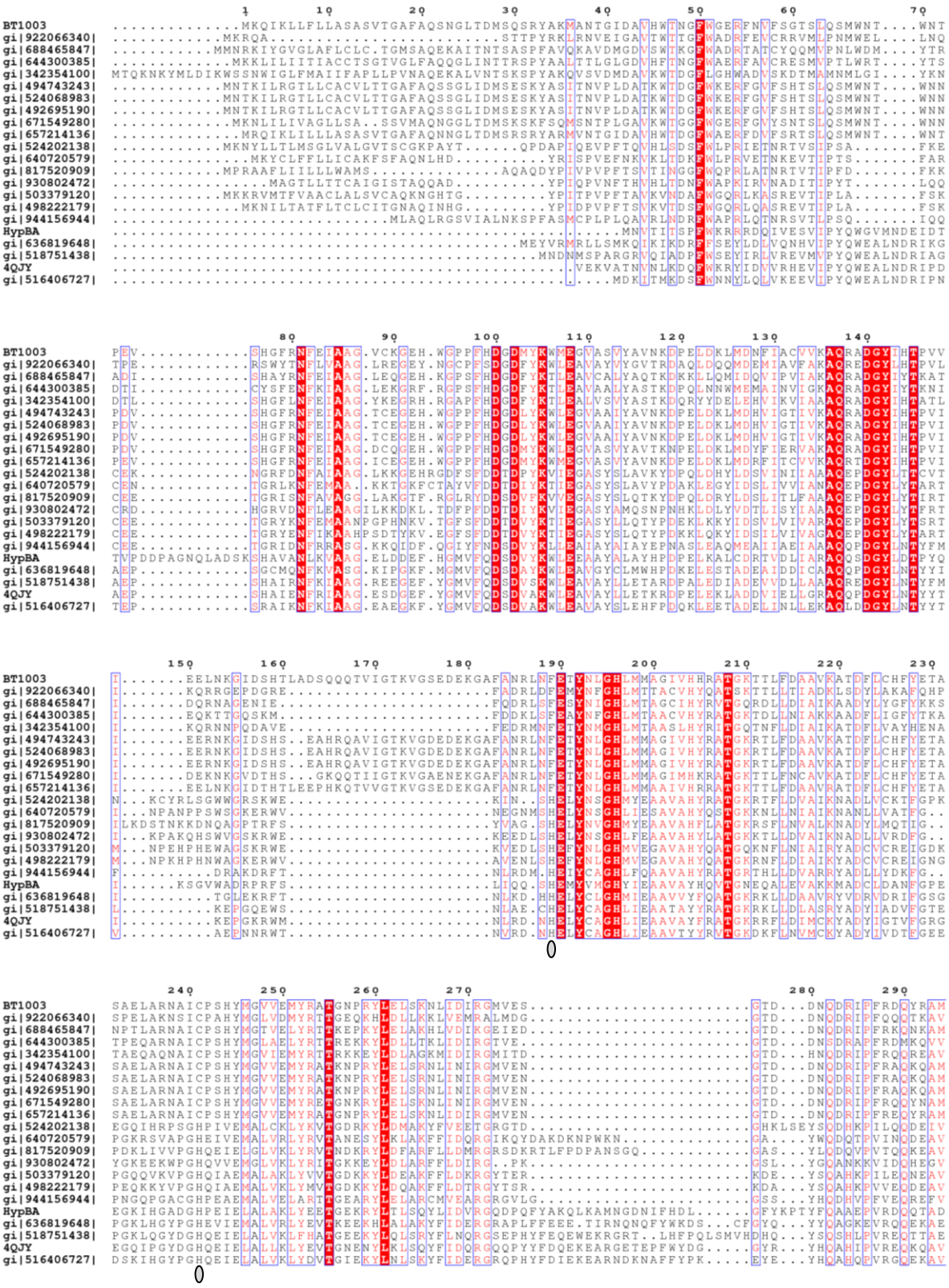
BT1003 represents the first member of GH127 that displays AceAase activity. Differences between the active sites of the GH127 AceAase and the GH127 β-L-arabinofuranosidases suggest that these two catalytic species, although sharing a common ancestor, have evolved to recognize and targed different linkages. In order to test the conservation of the active site residues, sequences from GH127 members were aligned.

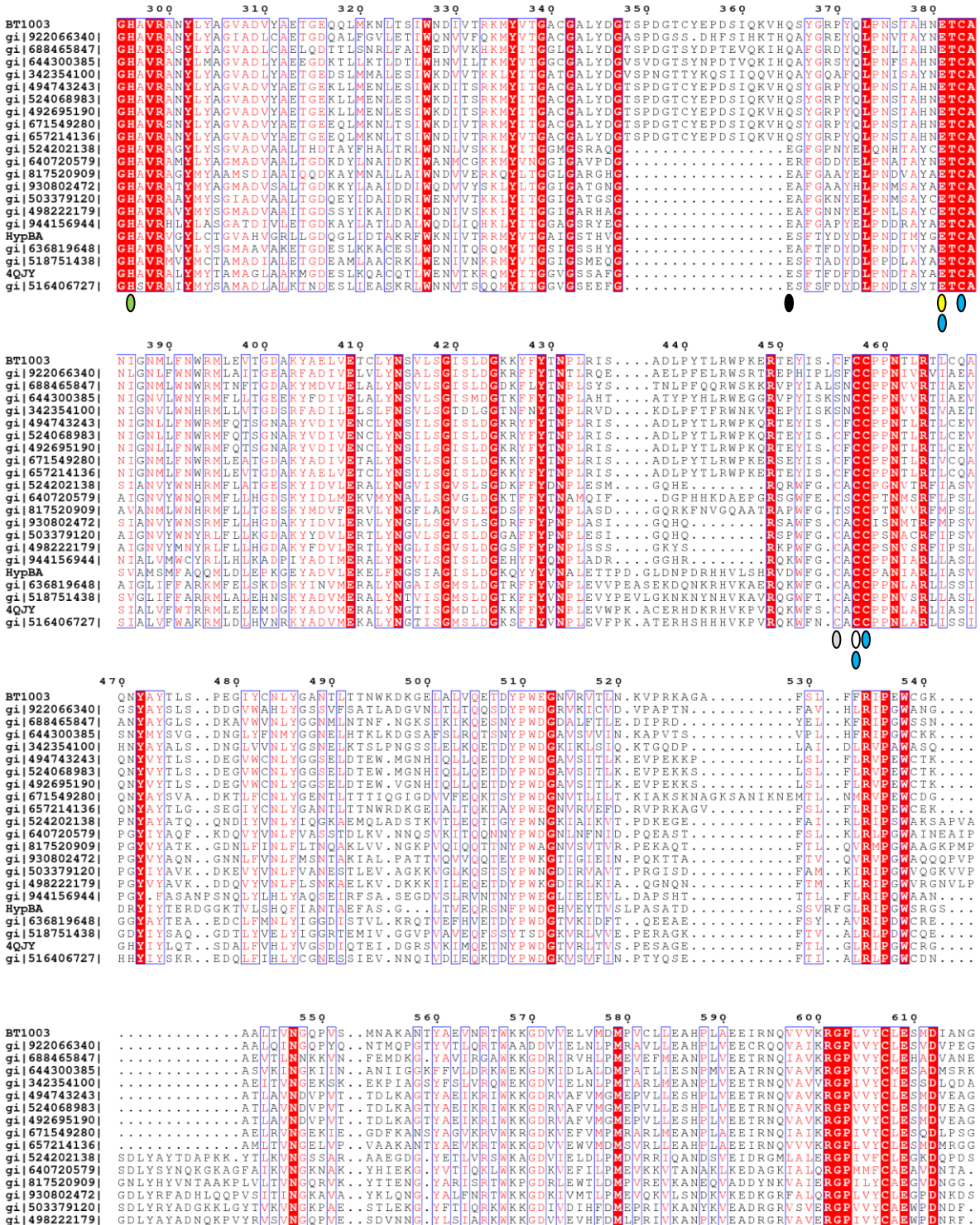
In the CAZy database GH127 comprises 1178 members, 1065 in the Bacteria domain (Cantarel *et al*. 2009; Lombard *et al*. 2014). Due this large number of members in the family the alignment was restricted to a lower number of sequences. Briefly, the BT1003 sequence was subjected to PSI-BLAST with a defined maximum of 500 target

sequences (Altschul *et al*. 1997). The sequences were aligned utilizing the online tool MUSCLE (Edgar 2004a; Edgar 2004b). The sequence hit with low homology revealed 31% sequence similarity. These 500 sequences were analysed however, in order to simplify, only twenty sequences will be shown. These sequences were selected based on the percentage of identity to BT1003: ten sequences with 31 to 32%, four sequences with 50 % and 75 %, one with 90%. The alignment was performed with HypBA1 and the *G. stearothermophilus* β-L-arabinofuranosidase (PDB code 4QJY). The alignment file was visualized using the online tool ESPrit 3.0 (Robert and Gouet 2014) and is shown at Figure 4.25.

The alignment showed that the proposed catalytic nucleophile (C457 in BT1003), the Zn^{2+} binding site (residues E381, C383, C458 and C457 in BT1003) and the active site residue H296 were invariant in GH127 members (in white, blue and green, respectively; Figure 4.25). Although, C455 (active site residue) was not conserved (in grey, Figure 4.25). These amino acids were previously described as important in ligand recognition in the β-L-arabinofuranosidase HypBA1 (Ito *et al*. 2014). However, mutagenesis data presented here show that C455 is not essential for BT1003 enzymatic activity (see Figure 4.24). Additionally, the glutamate that is the proposed catalytic acid-base in HypBA1 (Ito *et al*. 2014) was only conserved in 4QJY and the 10 sequences with low identity compared with BT1003. In the remaining proteins this glutamate was replaced by a glutamine (in black, Figure 4.25). A similar result was observed with two of the active site histidines (H142 and H194 in HypBA1) that were only conserved in GH127 members with less than 50% of sequence identity to BT1003 (in grey, Figure 4.25). In BT1003 and sequences with ≥ 50% identity to the *Bacteroides* enzyme, these histidines were substituted by a phenylalanine (F189) and a cysteine (C241) (in grey, Figure 4.25). This phylogenetic data suggests that GH127 members in which the catalytic apparatus of HypBA1 is conserved function as β-L-

arabinofuranosidases, while those that contain active site residues similar to BT1003 are putative AceAases.





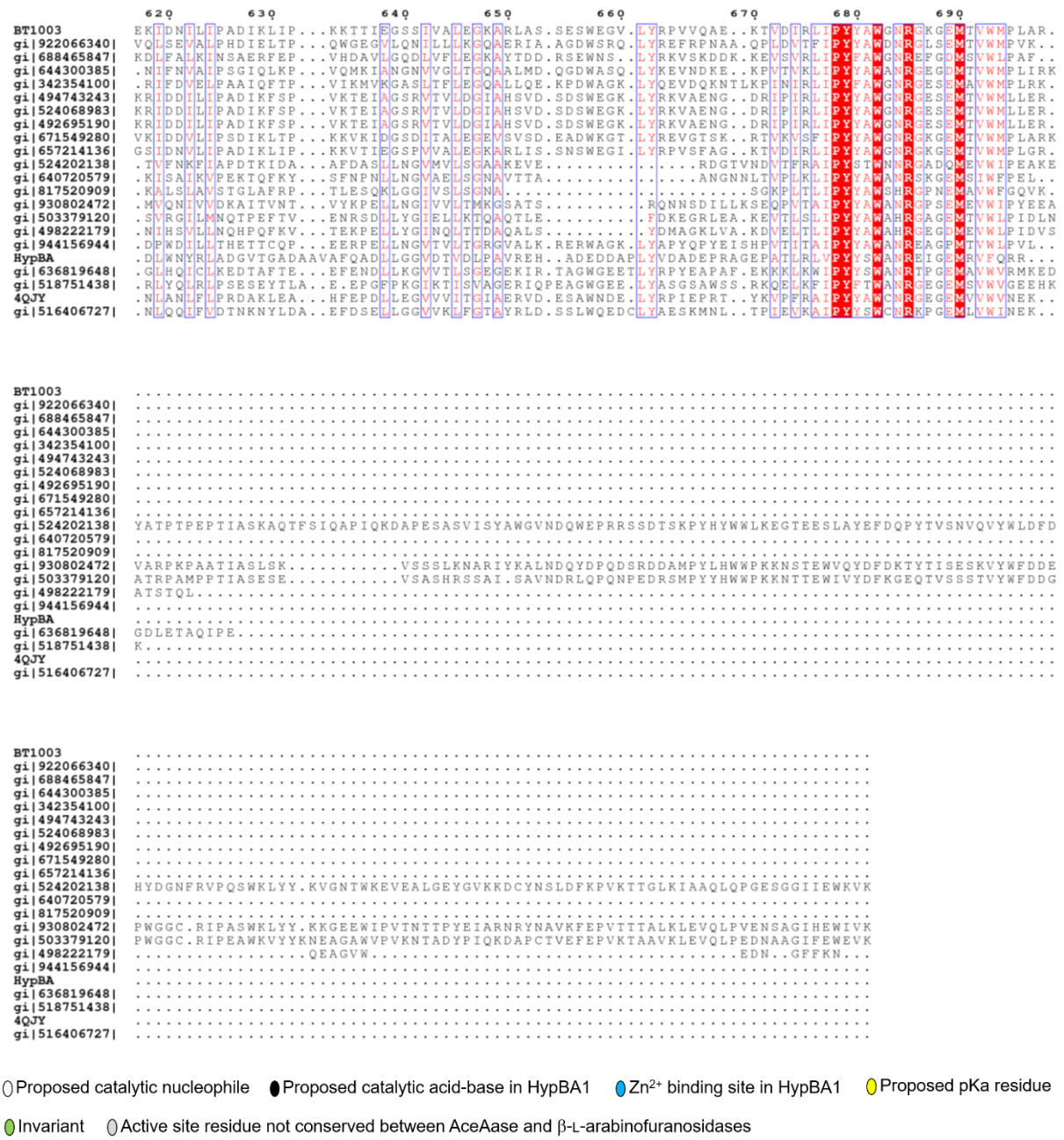


Figure 4.25 Alignment of proteins in GH127 family

NCBI reference codes for each protein are shown on the left. BT1003 represents the AceAase from *B. thetaiotaomicron*. HypBA1 and 4QJY are the GH127 β-L-arabinofuranosidases from *B. longum* and G. *stearothermophilus*, respectively. Sequences were aligned in MUSCLE (Edgar 2004a; Edgar 2004b) and visualized using ESPrit 3.0 (Robert and Gouet 2014). Amino acids with 100% conservation in this alignment are highlighted in a red background and residues that are functionally highly conserved are boxed and coloured red. The residues are numbered according to the BT1003 sequence.

4.4 Discussion

In this chapter the model of RG-II degradation by *B. thetaiotaomicron* was explored in detail. The degradation model for Chain B was presented and the linkage specificity of the RG-II enzymes was studied. This work also contributed to the refinement of the structure of the pectic polysaccharide. Additionally, two proteins were structurally characterized: BT0986, the first structure of a GH106 family member and, BT1003, the first enzyme described as displaying AceAase activity.

RG-II is the most complex polysaccharide in nature and, until now, the mechanism by which it is depolymerized was unknown. This work showed that to degrade the nine linkages of Chain B, *B. thetaiotaomicron* uses nine enzymes each targeting a specific linkage. These enzymes act through a hierarchical exo-model of degradation where a downstream enzyme requires the activity of upstream glycoside hydrolase(s) positioning the target sugar at the non-reducing end of the glycan. So, in order to establish the full RG-II degradation process it is important to clarify the linkage hydrolyzed by the different enzymes, as well determine the order of sequential cleavage. In this chapter the linkage specificity of BT0986 and BT1019 was studied in detail and is an illustration of the complexity of the RG-II degradative process. These two α-L-rhamnosidases cleave the two α-L-Rha units attached by different linkages to L-Ara*p*. BT0986 specifically cleaves the L-Rha-α1,2-L-Ara*p* linkage and BT1019 is only active against L-Rha-α1,3-L-Ara*p.* These differences in activity can be explained by steric constraints in these enzymes that result in the cleavage of a single linkage.

Another example of the complexity of the RG-II degradation model is provided by BT1001. This α-L-rhamnosidase is active against L-Rha-α1,3'-D-Api present in Chains A and B. In this chapter the activity of this enzyme was explored and, as expected from

the proposed exo-active model of degradation, the GH78 α-L-rhamnosidase requires the activity of the upstream enzymes (BT1002 in Chain A and BT1003 in Chain B), but also the downstream apiosidase. Thus, BT1001 is only active against the disaccharide L-Rha-α1,3'-L-Api. The structure of this protein is available (PDB code 3CIH). There are three additional structures for members of family GH78: *Bacillus* sp. GL1 (PDB code 20KX) (Cui *et al*. 2007), *Klebsiella oxytoca* KCTC 1686 NRRL B-199 (PDB code 4XHC) (O'Neill *et al*. 2015) and *Streptomyces avermitilis* MA-4680 (PDB code 3W5N) (Fujimoto *et al*. 2013). There is an apparent variability of substrates recognized by these enzymes. All the enzymes show activity for PNP-α-L-Rha but only the *Streptomyces* enzyme is able to recognize a polysaccharide (AGP from Gum Arabic) (Cui *et al*. 2007; Fujimoto *et al*. 2013; O'Neill *et al*. 2015). The *Bacillus* GH78 targets the disaccharide L-Rha-β1,3-D-Glc (gellan) (Cui *et al*. 2007). In order to understand if the structure of BT1001 presents steric constrains that explain the specificity for the disaccharide, the structure of the *B. thetaiotaomicron* enzyme was overlaid with the other GH78 structures using using the Secondary Structure Matching tool in software WinCoot. A schematic representation of BT1001 structure is presented in Figure 4.26A and, similar to the other GH78 structures, the active site is predicated to be in a cleft in the center of the $(\alpha/\alpha)_6$-barrel. The overlay of GH78 structures also revealed a complete conservation of the catalytic apparatus (Figure 4.26B). In this family the general catalytic base is a glutamate but the general catalytic acid can be a glutamate (*Streptomyces and Bacillus* proteins) or an aspartate (*Klebsiella* and BT1001 proteins). The proposed pKa modulator (glutamate) was also conserved in the four GH78 structures. The analysis of the BT1001 active site (Figure 4.26C) showed that the deep pocket can accommodate a sugar in +1 (D-Api) and it is formally possible that D-GalA would not be fully solvent exposed making steric clashes with the enzyme. It is also possible that O1 of the D-Api at the +1 subsite points at the enzyme preventing

extension of the reducing end of the L-Rha-α1,3'-L-Api disaccharide within the confines of the substrate binding region of BT1001.

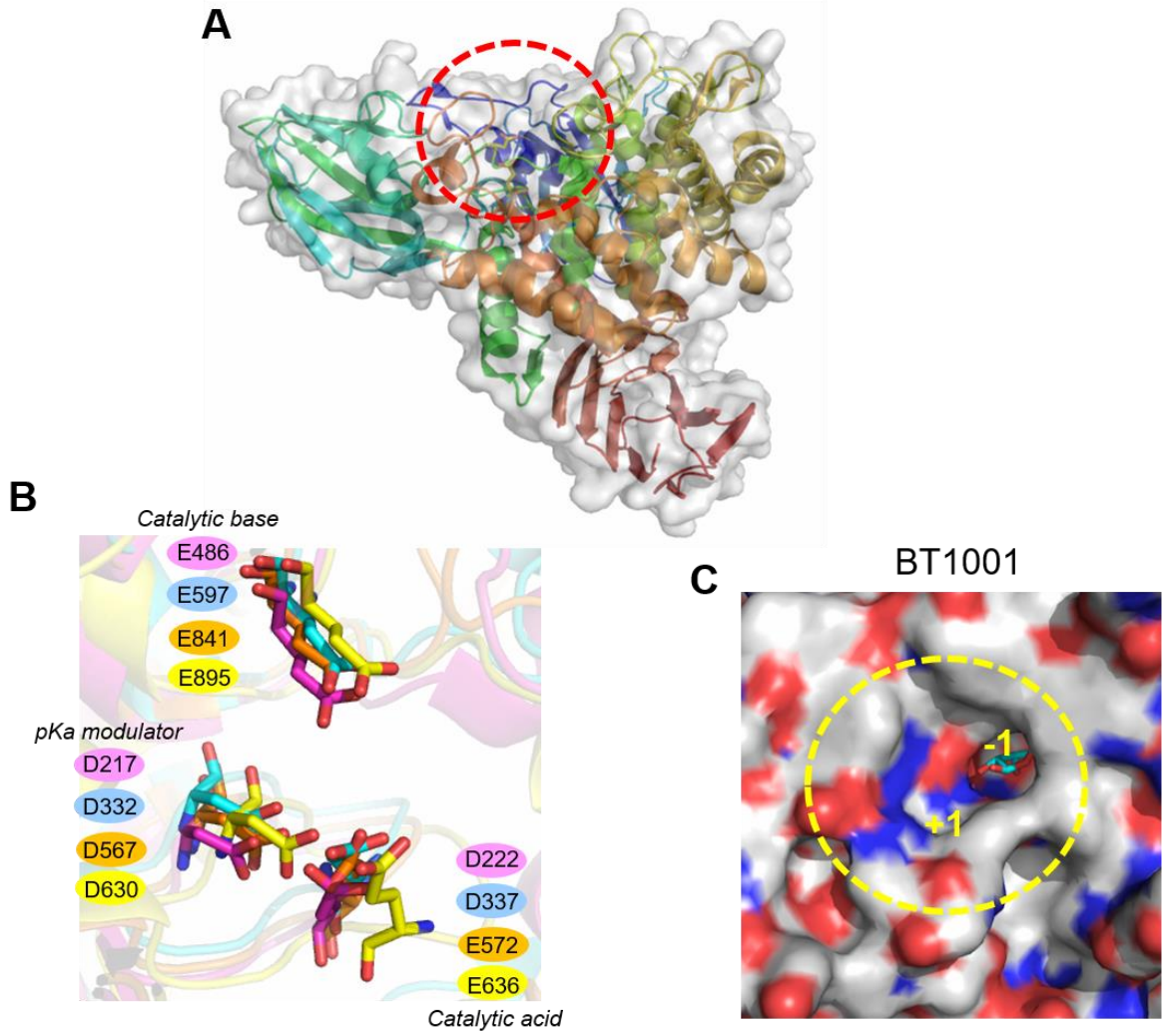


Figure 4.26 GH78 structures

A. Schematic representation of the structure of BT1001 rainbow coloured with N-terminal in blue and C-terminal in red. The red circle indicates the location of the predictive active site located in the centre of the $(\alpha/\alpha)_6$-barel catalytic domain. **B.** Overlay of the proposed catalytic apparatus. BT1001 (3CIH) is coloured light blue, *S. avermitilis* (3W5N) in yellow, *Bacillus* (2OKX) in orange and *Klebsiella* (4XHC) in pink. The amino acid background colour matches the protein structure colour. **C.** Displays the surface representation of the active site BT1001 active site. The L-rhamnose (blue sticks) in BT1001 results from the overlay with *S. avermitilis* structure. All figures were generated using PyMOL software.

To degrade RG-II *B. thetaiotaomicron* encodes four α-L-rhamnosidases and two β-L-arabinofuranosidases that are linkage and substrate specific. Only the D-galactosidase BT0993 was shown to be, apparently, non-specific. By HPAEC it was possible to detect

galactose when BT0993 was incubated with RG-II, lactose, D-Gal-β1,3-D-Gal and D-Gal-β1,6-D-Gal. The actual catalytic efficiency (k_{cat}/K_M) for the different substrates was determined by Dr Alan Cartmell (ICaMB, Newcastle University). Against the substrate D-Gal-β1,2-L-AceA-α1,3-Rha-α1,3'-L-Api (truncated Chain B oligosaccharide), BT0993 had a catalytic efficiency of 2.79 (± 0.17) x 10^6 min^{-1} M^{-1}. However, against lactose and D-Gal-β1,3-D-Gal the enzyme was substantially less active ~10000- and 5000-fold, respectively. Against D-Gal-β1,6-D-Gal the activity was too low to quantify. These results suggest that, despite BT0993 being able to cleave different linkages, the enzyme displays a high level of specificity for its target linkage in RG-II. As was mentioned before, *B. thetaiotaomicron* expresses several enzymes displaying the same activity (for example, four α-L-rhamnosidases) that target unique linkages in RG-II. This expansion of proteins with the same activities represents an additional energetic cost to the bacterium. However, it is possible to speculate that the expansion of the *B. thetaiotaomicron* enzymes results in the evolution of glycoside hydrolases that are optimized for a specific linkage. These bespoke enzymes would thus contribute to a highly efficient RG-II degrading system. Thus, the energetic cost of expressing several proteins, when a non-specific glycanase would suffice, would be reduced by the increased efficiency of the system and hence the requirement for high level expression of the individual enzymes.

The structural work presented in this chapter contributes to the GH106 family characterization for which no mechanistic or structural information was previously available. BT0986 displays a catalytic domain that presents a $(\beta/\alpha)_8$-barrel fold with several appended several β-stranded domains. The active site is located in a narrow cleft at the center of the $(\beta/\alpha)_8$-barrel where the ligand D-rhamnopyranose tetrazole was bound. The ligand adopted an approximate $^{2,5}B$ conformation, which is reciprocal

to the configuration adopted by D-mannosides in the active site of mannosidases ($B_{2,5}$ in GH2, GH26, GH38, GH92 and GH113 enzymes) (Ducros *et al*. 2002; Tailford *et al*. 2008; Zhu *et al*. 2010; Speciale *et al*. 2014). The conformations adopted by L- and D-mannosides are required to overcome the steric constraints imposed by *syn*-diaxial orientations at the transition state for the C2-hydroxyl and the leaving group/incoming nucleophiles, referred to as the 'Δ2 effect' (Zhu *et al*. 2010). Furthermore, within this site are two highly conserved carboxylate residues that were essential for activity, which likely comprise the catalytic apparatus. The E461 and E593 are separated by 8.0 Å, a distance between catalytic residues consistent with an inverting mechanism (see Chapter 5.3.3.2 for GH106 mechanism) (Rye and Withers 2000). Additionally, E593, which is 5.8 Å from the anomeric carbon likely activates a water molecule that mounts a nucleophilic attack at C1, thus the glutamate is predicted to function as the catalytic general base. E461 is likely to be within hydrogen bonding distance of the glycosidic oxygen and thus is a candidate for the catalytic general acid. It should also be noted that four active site carboxylate residues, three of which are conserved in the GH106 family, coordinate a calcium (with serine), and the divalent cation makes polar interactions with O2 and O3 of the transition state analog rhamnotetrazole bound in the active site. The carboxylate E538 is a glutamine in the RG-I homologue (see Chapter 5.3.3.3). The alanine mutagenesis of this residue has an impact on the enzymatic activity, however, did not completely inactivate the enzyme suggesting that this residue is not essential. The importance of calcium is illustrated by the loss of activity when three of the glutamates and a serine that coordinate the metal were mutated (D458A, S459A, E561A and E593A) and the observation that BT0986 displays no activity in the presence of EDTA. The importance of calcium has considerable resonance with other inverting enzymes that target α-manno-configured linkages, which often exploit divalent metal ions in substrate binding (GH47 and GH92

families) (Speciale *et al*. 2014). It has been proposed that the more flexible coordination geometry of calcium (which may be octa- or hepta-coordinate) not only allows for easier ground-state to transition-state migrations but also more easily accommodates the loss of the ligation to the nucleophilic water once the first-formed product departs the active centre (Zhu *et al*. 2010). In addition, it is also important to note that in the GH106 family there is low conservation of the residues implicated in substrate recognition at -1 subsite. This suggests that members of this family can utilize different residues to recognize L-Rha or target different substrates (see Section 5.3.3).

BT1003 cleaves L-AceA representing the first report in the literature of an AceAase. The crystal structure of this enzyme revealed an $(\alpha/\alpha)_6$ fold typical of other GH127 enzymes (Ito *et al*. 2014). While there is much conservation in the active site of the GH127 arabinofuranosidases (Ara*f*ases) and BT1003, one notable difference is that the proposed catalytic acid-base, glutamate in the Ara*f*ases, is replaced by a glutamine in the AceAase, precluding a role in the protonation of the glycosidic oxygen during glycosylation, or activating the catalytic water during deglycosylation. Mutagenesis data of the AceAase indicated that the proposed catalytic nucleophile, C457, the pKa modulator E381 and Q365**,** which replaces the catalytic acid-base in the Ara*f*ase, are all essential for activity. A candidate for the catalytic acid-base is not obvious, although it is formerly possible that the carboxylate of the L-AceA moiety of the substrate may itself participate in the catalytic mechanism. Of potential significance is that bioinformatic analysis of GH127 proteins revealed that this family can harbour two enzyme types (Ara*f*ases and AceAase) that present distinct active site residues. This result can contribute to a better prediction of enzymatic activity in family GH127 based on protein sequences.

The results presented in this chapter in combination with additional data generated in Gilbert's lab allowed the generation of a proposed model for RG-II degradation by *B. thetaiotaomicron* (Figure 4.27). In RG-II, side chains can be extensively degraded before backbone cleavage. Indeed, this chapter showed that it was possible to degrade six of the nine Chain B linkages (until the L-AceA) (Figure 4.8). Additional data from colleagues also showed that it is possible degrade one of the two linkages in Chain D, the two linkages in Chain C and the terminal L-galactosidic and D-glucuronosidic linkages in Chain A without the need for backbone degradation (Figure 4.27). A possible explanation for these differences in chain access and cleavage is the borate cross-linkage between D-Api units in Chain A (Ishii *et al*. 1999). This could modify the structure of RG-II and restrict enzymatic access to this chain. Indeed, the apiosidase (BT1012) activity was shown to be sensitive to borate when cleaving Chain A. Thus, the cross-link must be destabilized to release Chain A from the backbone. It is proposed that the pectate lyase BT1023 is responsible for the backbone degradation required to overcome the steric constrains imposed by the RG-II structure. This cleavage of backbone allows the activity of the remaining enzymes and the complete degradation of RG-II. BT1011 cleaves the non-reducing end Δ4,5-anhydrogalacturonic acid generated by BT1023 and BT1017 removes the D-GalA esterification. The D-Dha (Chain D) is removed by BT1020 and the remaining backbone is degraded by BT1018 (galacturonidase). Both Chain A and B are then degraded further by a combination of six enzymes. The cleavage of L-Rha-α1,3'-D-Api linkage by BT1001 is the final step in the depolymerisation process.

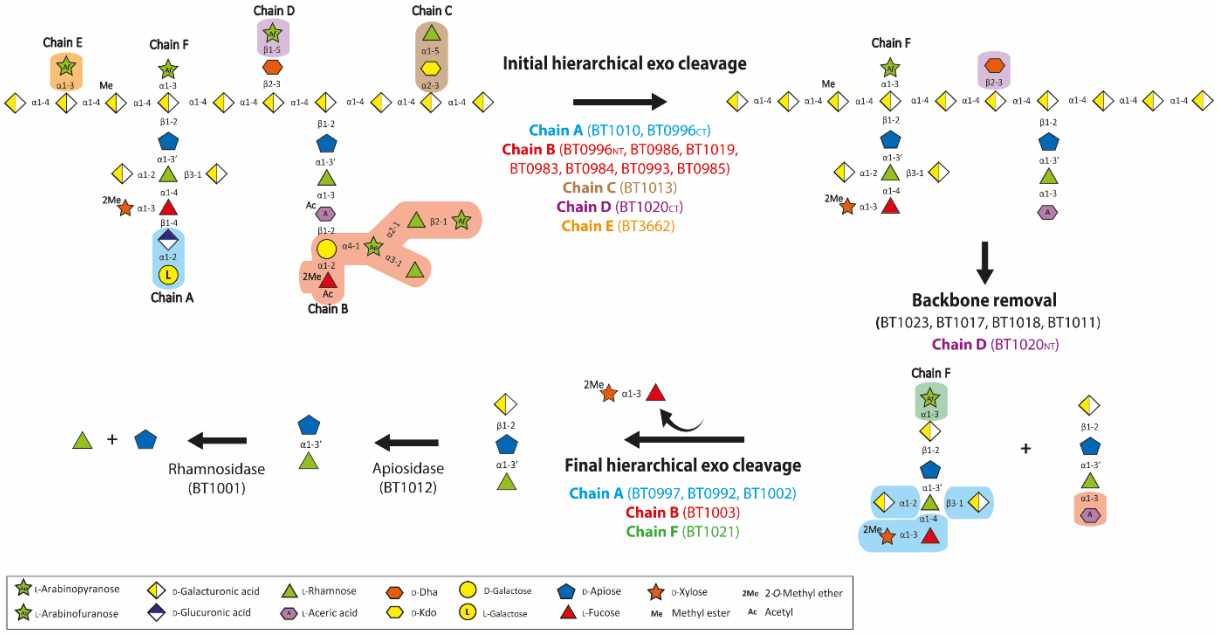


Figure 4.27 Proposed model of RG-II degradation by *B. thetaiotaomicron*

In each step of degradation of RG-II, the target sugars are highlighted with boxes that match the colour of the enzymes.

Elucidating the mechanism of RG-II degradation revealed new insights into the structure of the polysaccharide. The L-Rha-D-Api linkage configuration was revealed to be an α- and not a β-linkage, as reported in the literature (Spellman *et al*. 1983). Additionally, the characterization of the oligosaccharide produced by the *Δbt1017* mutant revealed an L-Ara*f* unit on the D-GalA backbone, which was removed through the action of BT1021. To identify the location of this L-Ara*f*, further studies were performed by Didier A. Ndeh. Briefly, strains carrying the double mutation of *bt1021* and *bt1010* (prevents degradation of Chain A) or *bt1021* and *bt0986* (prevents substantial degradation of Chain B) were grown in RG-II and the oligosaccharides released were isolated, purified and analysed by LC-MS. When *Δbt1021* and *Δbt0986* were combined in the same bacterial strain the oligosaccharides generated by the *Δbt0986* and *Δbt1021* mutation were observed. However, if the mutant strain carrying both the *Δbt1021* and *Δbt1010* mutations was grown on RG-II, a single oligosaccharide was generated comprising Chain A attached to a backbone GalA that was also

appended with L-Ara*f.* These results provide evidence that the L-Ara*f* unit, defined here as Chain F, is a conserved feature of the site of Chain A decoration of the D-GalA backbone and, as such, is a novel feature of the structure of RG-II. Furthermore, the mutant *Δbt1017* always generates a unique and well defined oligosaccharide. So, the C6 methylation of the D-GalA attached to the non-reducing GalA decorated with Chain A and Chain F is also a conserved feature of the structure of RG-II that was never been reported in literature. However, the biological role of these new features of RG-II is currently unknown.

Additionally, during the time of writing this thesis, the structure of BT0996 (L-arabinofuranosidase domain) in complex with Chain B was obtained by Alan Cartmell (Figure 4.28). This represents the first report of the 3D structure of Chain B where it is possible to observe eight of the nine sugars that comprise Chain B (only the terminal L-Ara*f* is missing) or complete Chain B in the BT0996 inactive mutant (contains the E240Q mutation in the -1 subsite or active site). The structure of Chain B was not stabilized by extensive interactions with the enzyme but through interactions of the L-Rha, linked α-1,3 to the L-Ara*p*, with D-Gal-β1,2-L-AceA-α1,3-L-Rha. The structure supports the biochemical data presented in this chapter showing that the linkage between L-Rha and D-Api is indeed α. The structure also shows that the Ara*p* is in the highly unusual $^{1}C_{4}$ conformation. This places O1, O2 and O3 axial with only O4 equatorial with the pyranose ring. This is an extremely unstable conformation for a pyranose sugar, and likely reflects the interaction of the O3-linked rhamnose within the chain. This forces the C6 deoxy-sugar into an axial orientation, which can only occur if the L-Ara*p* adopts a $^{1}C_{4}$ conformation. Encouragingly, this configuration of the pentose sugar is consistent with previous NMR data that also suggested the arabinopyranose was in a $^{1}C_{4}$ conformation (Glushka *et al*. 2003). A worrying feature of the structure, however, was that in the crystal structure the rhamnose appears to be linked to O3 and

not O3'. It was not possible for Cartmell to build an O3'-linked apiose into the structure but the observed electron density was consistent with the structure of apiose. The electron density of the reducing end sugar was more consistent with D-Api is linked O3 and not O3' (Figure 4.28C) as reported in the literature (Vidal *et al*. 2000; Glushka *et al*. 2003). If the O3 linkage is correct it may offer an alternative explanation for why borate cross links Chain A but not Chain B. The borate requires two adjacent *cis* hydroxyls, which would not be present in the O3-linked sugar. However, further studies are required to provide definitive evidence that the D-Api linkage is different in Chain A and Chain B. Furthermore, BT1001 was active against L-Rha-α1,3'-D-Api, although, given its substantial activity against 4-nitrophenyl-α-L-rhamnopyranose, it is possible that specificity at the +1 subsite is flexible. It is also possible that the crystallographic data are misleading. The electron density for the rhamnose linked to D-Api was weak and thus the location of the L-Rha-D-Api linkage provided by the crystal structure must be viewed with considerable caution.

Overall, the work present in this chapter contributed to our understanding of the mechanism by which *B. thetaiotaomicron* degrades RG-II and has improved our knowledge of the structure of the pectic polysaccharide.

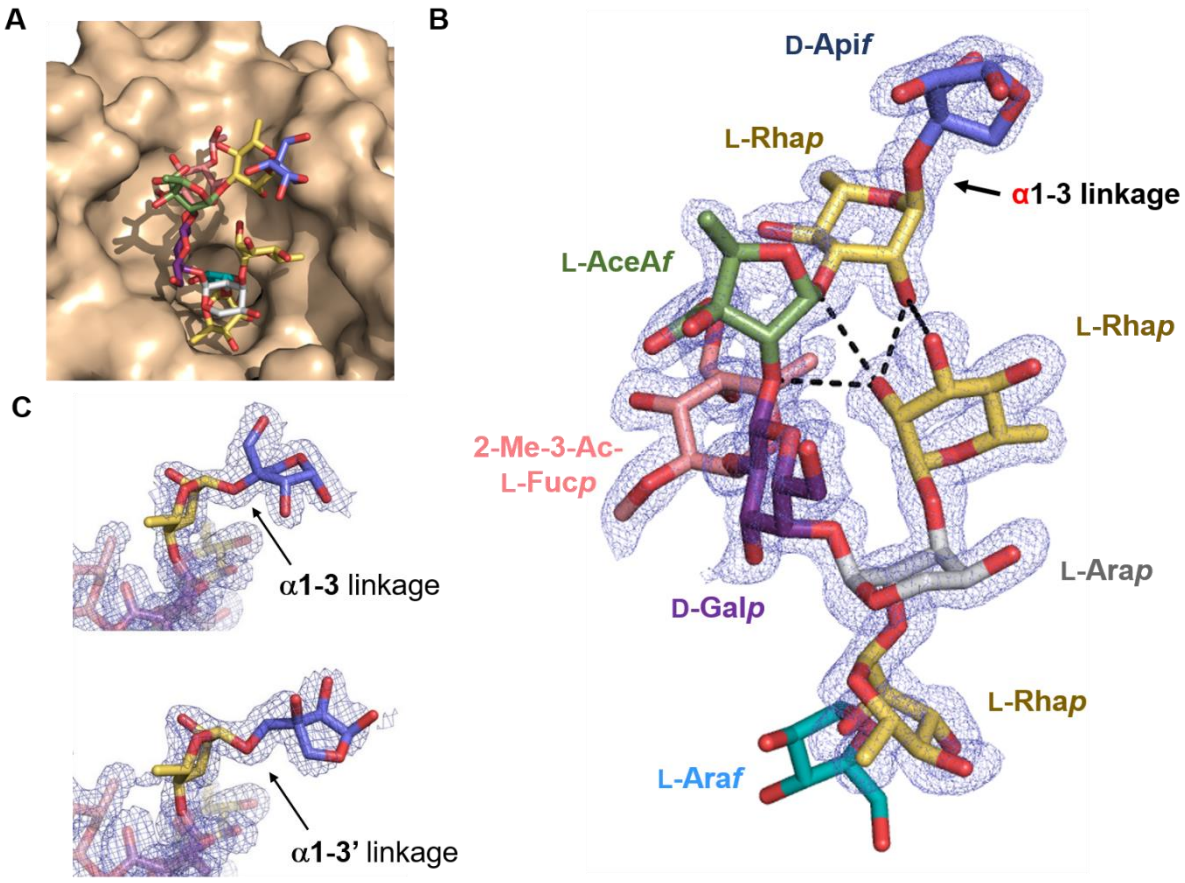


Figure 4.28 BT0996 structure in complex with Chain B

A. Surface representation of the inactive mutant (E240Q) of the L-arabinofuranosidase domain of BT0996 in complex with Chain B. Chain B is an overlay of Chain B derived from the wild type (L-Ara*f,* in blue) and the catalytic mutant complex (remaining sugars) **B.** Chain B sugar composition and linkages. The important findings in this study are pointed by arrows: the α-linkage between L-rhamnose-D-apiose. The different sugars are labelled and represented in different colours as follow L-rhamnose (L-Rha) in yellow, L-arabinofuranose in light blue, D-galactose in purple, 2-*O*-methyl-L-Fucose (2-*O*-Me-L-Fuc) in pink; L-aceric acid (L-AceA) in green and D-apiose (D-Api) in dark blue. **C.** Distal region of Chain B bound to the E240Q mutant of BT0996. The carbons of L-Rha and D-Api are in yellow and dark blue, respectively. The blue mesh represents the 2Fo-Fc electron density map (1.3 Å resolution) at 1.5 σ. In Ci and Cii the glycosidic bond has been modelled as a α1,3 and α1,3' linkage, respectively. The figure was created in PyMOL using the pdb file generated by Alan Cartmell.

4.5 Future work

In this chapter a complete model of RG-II degradation was presented, however, the *in vivo* degradation in *B. thetaiotaomicron* remains uncertain. Future studies on the cell localization of RG-II-PUL encoded enzymes in an *in vivo* context will be critical to

understand if the degradation of the polysaccharide occurs mainly at the cell surface or in the periplasm. This will allow us to build an *in vivo* model of RG-II degradation in *B. thetaiotaomicron* and evaluate whether RG-II fragments can participate in cross-feeding with other microorganisms in the human microbiota. Additionally, the BT1003 structural studies should be continued to obtain a structure of the enzyme in complex with the ligand. This will provide essential information about the catalytic apparatus of this enzyme. Furthermore, a more extensive bioinformatics analysis of the GH127 family is required to test the hypothesis of this family being populated only with Ara*f*ases and AceAase or whether it contains additional enzyme activities.

Chapter 5. Rhamnogalacturonan-I utilization by *Bacteroides thetaiotaomicron*

5.1 Introduction

The gastrointestinal tract is colonized by a community of microbes, the microbiota, which has a significant impact on human health and nutrition (Ley *et al*. 2005; Marietta *et al*. 2013). Plant cell wall polysaccharides (hemicelluloses and pectins) and host glycans (mucins and glycosaminoglycans) are the major nutrients available to the microbiota (Sonnenburg *et al*. 2005; Martens *et al*. 2011). Pectins, galacturonic acid containing plant polysaccharides, are a significant component of the human diet. These polysaccharides are found in a high percentage in plants such as citrus fruit (12 - 28%), apples (0.5 - 1.6%) and potato (1.8 - 3.3%) (reviewed by Jayani *et al*. 2005). The major pectins are homogalacturonan (HG), rhamnogalacturonan-I (RG-I) and rhamnogalacturonan-II (RG-II). A schematic representation of the HG and RG-I structures is presented in Figure 5.1. RG-II was described in detail in Chapter 4 and will not be explored further in this chapter. HG is a linear polysaccharide comprising a backbone of α1,4-D-galacturonic acid (D-GalA) residues, where O2 and O3 of D-GalA can be acetylated and C6 methylated (Mohnen 2008). RG-I is a branched pectin that contains a backbone of a repeating disaccharide of α1,2-L-rhamnose-(L-Rha)-α1,4-D-GalA. The O2 and O3 of D-GalA can be acetylated and the O4 of L-Rha can be decorated with arabinan and galactan side chains. The size and composition of RG-I side chains is dependent on the plant and the development stage. In a very simplistic description, arabinan side chains are characterized by a α1,5-L-arabinofuranose (L-Ara*f*) backbone with short α1,2-L-Ara*f* or α1,3-L-Ara*f* side chains. Galactan presents a β1,4-D-galactose (D-Gal) backbone (reviewed by Mohnen 2008; Caffall and Mohnen

2009; Atmodjo *et al*. 2013). In order to clarify the terms used in this this chapter, "AM-RG-I" refers to the RG-I extracted from *Arabidopsis* seeds (see Chapter 2.2.6.1) and "P-RG-I" is the polysaccharide from potato (Megazyme®). AM-RG-I contains no side chains, while the backbone of P-RG-I is decorated with truncated galactan side chains. Additionally, sugar beet arabinan (Megazyme®) is characterized as having a short RG-I backbone that is highly decorated with arabinan side chains.

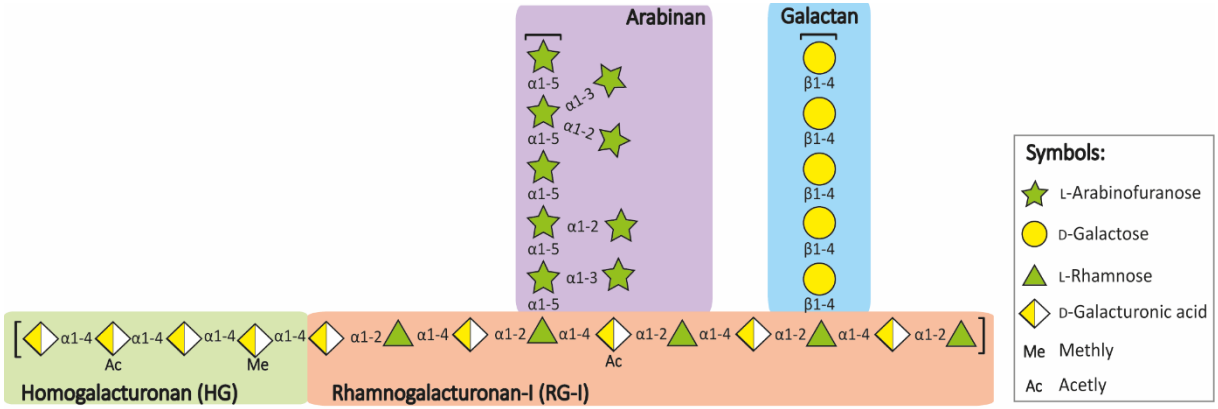


Figure 5.1 Schematic representation of the structures of homogalacturonan and rhamnogalacturonan-I

The linkages and sugar compositions are represented in the figure. The different polysaccharides are labelled with different coloured backgrounds.

Bacteroides thetaiotaomicron is a dominant member of the human microbiota. In this bacterium genes encoding specific glycan degrading systems are grouped into genetic loci (Polysaccharide Utilization Loci or PULs) (Xu *et al*. 2003). A typical PUL encodes the enzymes required to degrade the target polysaccharide; extracellular surface glycan binding proteins (SGBP); SusD- and SusC-like proteins that import glycans into the periplasm; a sensor, such as a hybrid two component system (HTCS), that regulates the expression of each PUL (Martens *et al*. 2009).

A recent study has shown that *B. thetaiotaomicron* grows on all the major pectins (Martens *et al*. 2011). Transcriptomic data identified the *B. thetaiotaomicron* PULs upregulated by the different pectins, Table 5.1. A schematic of the PUL upregulated by

the RG-I backbone is shown in Figure 5.2 (Martens *et al*. 2011). The arabinan and galactan PULs were explored by Jonathan Briggs (ICaMB, Newcastle University) and the HG PUL was characterized by Dr Wade Abbott (Lethbridge Research Center, Canada).

Table 5.1 *B. thethaiotaomicron* PULs upregulated in response to pectins

Substrate	**PUL**
Arabinan	*bt0348 - bt0369*
HG	*bt4108 – bt4123*
RG-I	*bt4145 – bt4183*
Pectic galactan	*bt4667 – bt4672*

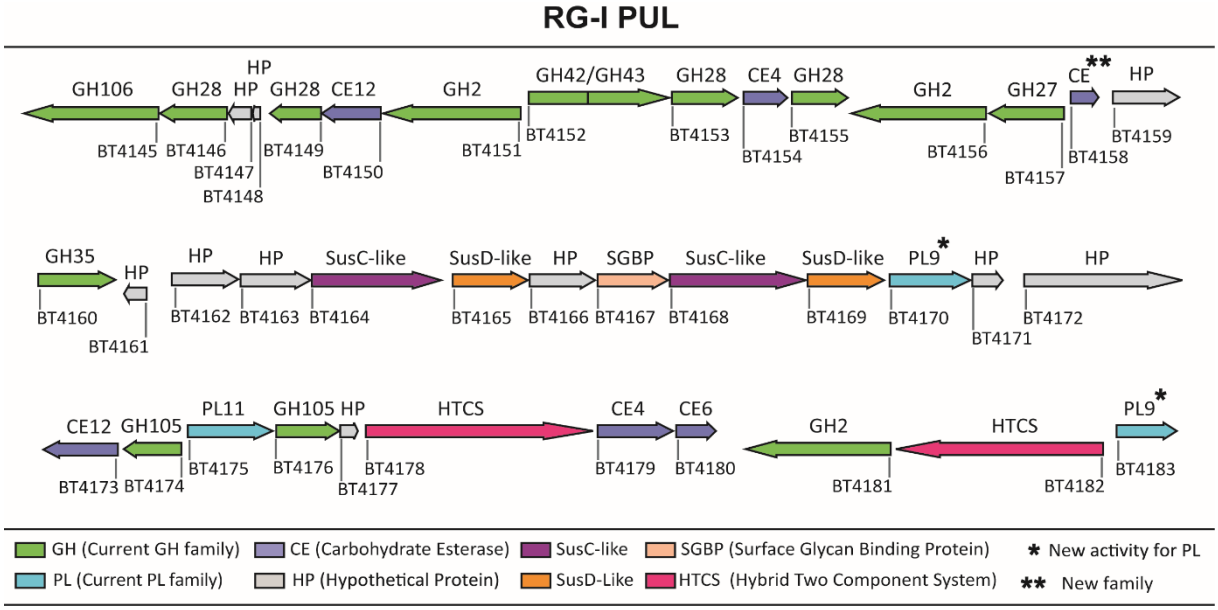


Figure 5.2 Rhamnogalacturonan-I polysaccharide utilization loci (PUL)
In the schematic of the PULs each gene is drawn to scale as a rectangle with its orientation indicated by the closed triangle.

The *B. thetaiotaomicron* RG-I PUL encodes twelve glycoside hydrolases (GH), three polysaccharide lyases (PL) and six carbohydrate esterases (CE). Previous work by Xiaoyang Zhang during his PhD showed that the GH106 α-rhamnosidase BT4145 and the GH28 rhamnogalacturonase BT4153 can cleave AM-RG-I in a sequential order releasing L-Rha and D-GalA. A founding member of a new carbohydrate esterase

family, BT4158, presented an essential role in the deacetylation of D-GalA allowing BT4145 and BT4153 access to the respective target bonds. The three PLs (BT4170, BT4175 and BT4183) are active on AM-RG-I releasing oligosaccharides containing Δ4,5-anhydrogalacturonic acid (Δ4,5-GalA) at the non-reducing terminus. The unsaturated tetrasaccharide (Δ4,5-GalA-α1,2-L-Rha-α1,4-D-GalA-α1,2-L-Rha) was identified as the signal molecule that binds to the HTCS (BT4178). Two GH105 unsaturated rhamnogalacturonyl hydrolases displayed substrate selectively. BT4176 is specific for the unsaturated disaccharide (Δ4,5-GalA-α1,2-L-Rha) while BT4174 showed a preference for the unsaturated tetrasaccharide. Additionally, BT4155 (GH28) released D-GalA from HG and P-RG-I (Zhang 2015). These preliminary data indicate that *B. thetaiotaomicron* requires several enzymes to degrade RG-I. However, to complete the model for the utilization of this polysaccharide it is essential to determine the kinetics and precise specificity of the different enzymes, determine the cellular location of the key components of the catabolic apparatus and use gene deletions to explore the functional importance of selected biocatalysts.

5.2 Objectives

The objectives of this chapter are:

iv) Characterize in detail the RG-I PUL encoded PLs: BT4170, BT4175 and BT4183

v) Explore the crystal structure of BT4170

vi) Define the specificity and the mechanism of catalysis of the α-rhamnosidase BT4145

vii) Explore the substrate specificity of the four GH28 enzymes: BT4146, BT4149, BT4153 and BT4155

viii) Identify the functional role of putative surface glycan binding proteins (SGBPs) encoded by RG-I PUL

ix) Establish the activities that occur at the cell surface and in the periplasm

x) Structural characterization of a rhamnogalacturonan lyase (BT4170) from PL9 family

5.3 Results

5.3.1 *Cloning, expression and purification*

DNA encoding the proteins characterized in this chapter were previously cloned into pET28a such that they contain an N-terminal His_6-tag. The different recombinant proteins were expressed in *Escherichia coli* strain TUNER and purified by Immobilized Metal Ion Affinity Chromatography (IMAC). When IMAC purification was not sufficient to obtain proteins with the required purity, size exclusion chromatography was also performed. When BT4170 was used in crystal trails the protein was further purified by anion exchange chromatography followed by size exclusion chromatography. The purity of protein samples was assessed visually by SDS-PAGE after protein staining. Examples of SDS-PAGE gels of the IMAC purification of selected enzymes are shown in Figure 5.3. The expected size, extinction coefficient, predicted signal peptide, the CAZy family and the activity for all proteins mentioned in this chapter are shown in Table 5.2.

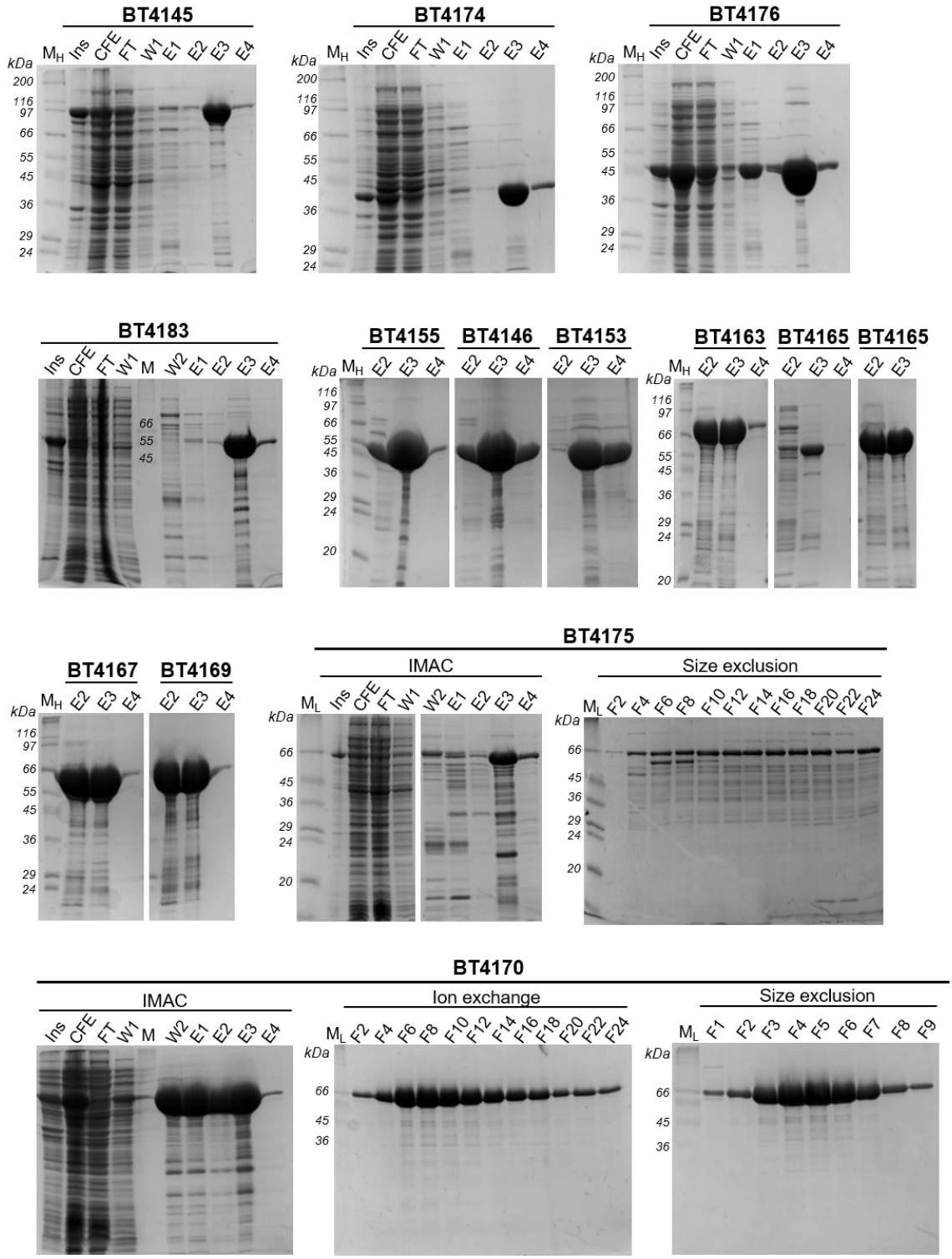


Figure 5.3 Examples of SDS-PAGE after proteins purification

Example of SDS-PAGE gels (Chapter 2.1.20) of the proteins mentioned during this chapter. The proteins are labelled in the figure and their experimentally determined sizes are consistent with the predicted molecular weights reported in Table 5.2. Ins – Insoluble fraction; CFE – Cell free extract; FT – Flow through; W1 and W2 - Wash with Talon buffer; E1 and E2 – protein elution with 10 mM imidazole in Talon buffer; E3 and E4 – elution with 100 mM imidazole in Talon Buffer (Chapter 2.1.19.1). The IMAC, ion exchange and size exclusion purification methods are described in Chapter 2.1.19.1, 2.1.19.2 and 2.1.19.3, respectively. M_L and M_H correspond to different molecular weight (kDa) markers (L – low and H – high).

Table 5.2 The details of the proteins mention in this chapter

Protein name	Expected size (kDa)	Extinction coefficient (M^{-1} cm^{-1})	CAZy family	Expected activity	Predicted signal peptide
BT4145	104	189 120	GH106	α-L-rhamnosidase	SP I
BT4146	52	61 810	GH28	Rhamnogalacturonase	SP I/SP II
BT4149	37	54 025	GH28	Rhamnogalacturonase	SP I/SP II
BT4153	49	54 610	GH28	Rhamnogalacturonase	SP I
BT4155	51	61 475	GH28	Exo-polygalacturonase	SP I
BT4163	84	134 095	NA	HP	
BT4165	59	131 460	NA	SusD-like	SP II
BT4166	65	78 660	NA	HP	SP II
BT4167	67	116 005	NA	SGBP	SP II
BT4169	67	124 360	NA	SusD-like	SP II
BT4170	58	52 175	PL9*	Rhamnogalacturonan lyase	SP I
BT4174	45	119 095	GH105	Unsaturated rhamnogalacturonyl hydrolase	SP I
BT4175	68	111 325	PL11	Rhamnogalacturonan lyase	SP I
BT4176	53	90 105	GH105	Unsaturated rhamnogalacturonyl hydrolase	SP I/SP II
BT4183	48	54 250	PL9*	Rhamnogalacturonan lyase	SP I/SP II

The expected activities were previously reported on (Zhang 2015) except for BT4149 (GH28) and BT4167 (SGBP) that were described during this PhD. HP means hypothetical protein and * signifies a new activity for a CAZy family. Signal peptide was predicted using Lipo 1.0 server (Juncker *et al*. 2003). SP I is a signal peptide sequence recognized by signal peptidase I and SP II indicates a lipoprotein signal peptide cleaved by signal peptidase II.

5.3.2 *Biochemical characterization of RG-I locus polysaccharide lyases: PL9s (BT4170, BT4183) and PL11 (BT4175)*

B. thetaiotaomicron RG-I PUL encodes three PLs, BT4170 and BT4183 belong to family PL9 and BT4175 to PL11. Previous work showed that these enzymes were active on AM-RG-I corresponding to the first report of a rhamnogalacturonan lyase in PL9, a family populated with polygalacturonan lyases (Cantarel *et al*. 2009; Lombard *et al*. 2014).

These PLs cleave glycosidic bonds by a β-elimination mechanism generating an unsaturated product containing a double bond between *C-4* and *C-5* of GalA (represented in this chapter as Δ4,5-GalA), leading to an increase in ultraviolet absorbance at 235 nm. The pH profile was determined by measuring the product formation at 235 nm and is shown in Figure 5.4. BT4170 and BT4183 were active between pH 8 and pH 10 with an optimum at pH 9.0 (in Tris-HCl buffer). However, BT4183 displayed maximum activity at pH of 7.5 (in Tris-HCl buffer).

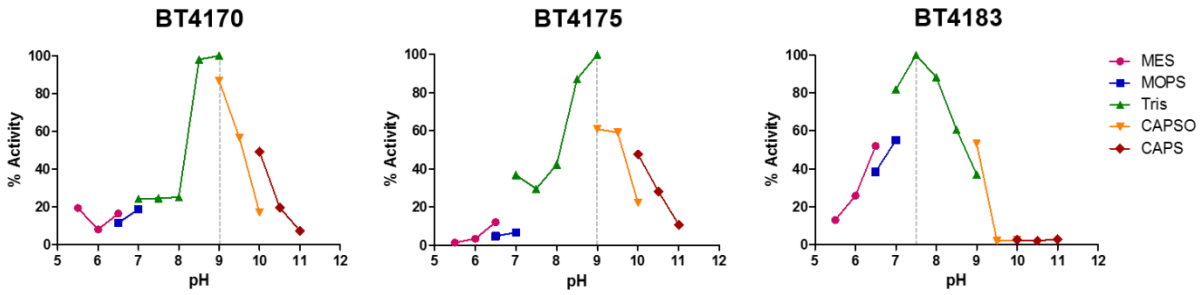


Figure 5.4 pH profile of BT4170, BT4175 and BT4183

The product formation was measured at 235 nm (see Chapter 2.2.5.1). All reactions were performed at 37 °C using the substrate sugar beet arabinan at 0.5%, 2 mM $CaCl_2$ and 100 mM of the respective buffer. The enzyme concentrations were 5 nM for BT4170, 2,5 nM for BT4175 and 20 nM for BT4183. The % of activity was calculated based on the highest enzyme rate defined as 100%.

The kinetic parameters for each enzyme were determined by measuring double bond formation at 235 nm, Table 5.3. The respective kinetics curves are presented in Figure 5.5. The molar concentration was determined based on amount of limit product generated using an excess of enzyme (for details, see Chapter 2.2.5.1). All enzymes were active on AM-RG-I and the branched substrate sugar beet arabinan. While BT4170 and BT4175 were significantly more active on AM-RG-I than sugar beet arabinan, BT4183 displayed similar activity against the two polysaccharides. The most active enzyme was BT4170 with a k_{cat}/K_M of 5.5 x 10^4 min^{-1} mM^{-1} against AM-RG-I. BT4175 was the only enzyme able to efficiently cleave P-RG-I, which contains extensive “truncated” galactan side chains. Additionally, it is important to notice that

after gel filtration BT4175 purity was low. So, the rates determined for this enzyme reported in Table 5.3 are likely underestimates.

Table 5.3 Activity of BT4170, BT4175 and BT4183

Recombinant protein	Substrate	k_{cat} (min^{-1})	K_M (mM)	k_{cat}/K_M (min^{-1} mM^{-1})
BT4170	AM-RG-I	$1.7 \times 10^3 \pm 62.2$	0.030 ± 0.004	$5.5 \times 10^4 \pm 8.7 \times 10^3$
	Sugar beet	$1.2 \times 10^3 \pm 32.4$	0.257 ± 0.022	$5.2 \times 10^3 \pm 598$
	P-RG-I		ND	
BT4175	AM-RG-I	658 ± 26.4	0.064 ± 0.007	$1.0 \times 10^4 \pm 1.4 \times 10^3$
	Sugar beet	$1.1 \times 10^3 \pm 42.5$	0.230 ± 0.029	$4.9 \times 10^3 \pm 804$
	P-RG-I	222.8 ± 9.1	0.079 ± 0.013	$2.8 \times 10^3 \pm 569$
BT4183	AM-RG-I	174.1 ± 5.0	0.020 ± 0.002	$8.8 \times 10^3 \pm 1.2 \times 10^3$
	Sugar beet	350.1 ± 10.2	0.059 ± 0.005	$6.0 \times 10^3 \pm 706$
	P-RG-I		NA	

For experimental details see legend of Figure 5.5. (ND) Activity too low to quantify; (NA) Not active.

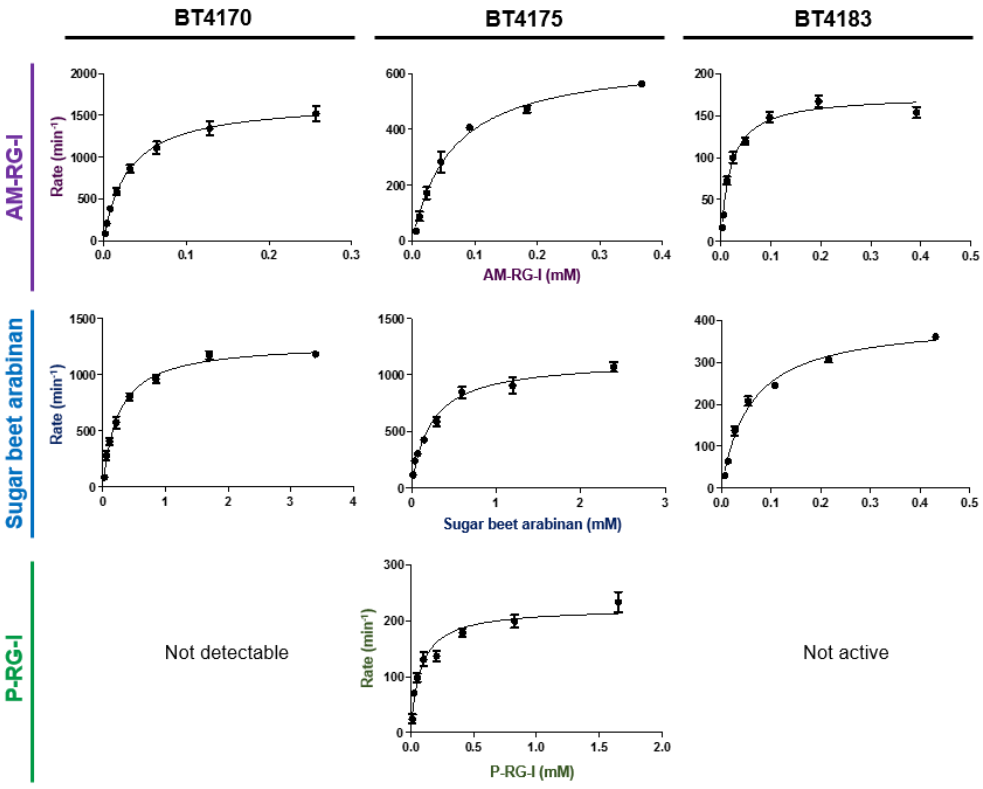


Figure 5.5 Graphical representation of kinetic data for the PLs against different substrates

All reactions were performed at 37 °C with 2 mM $CaCl_2$ and 100 mM CAPSO pH 9.0 (BT4170 and BT4175) or 100 mM Tris-HCl pH 7.5 (BT4183). Initial rates were measured at 235 nm (Chapter 2.2.5.1) and plotted in GraphPad Prism 5.0 software using using a non-regression analysis (Chapter 2.2.5). The standard errors of the mean were generated from technical triplicates.

The profile of the products produced by these enzymes from the different substrates was observed by thin layer chromatography (TLC) (Figure 5.6). Against AM-RG-I, BT4170 generated two main products: unsaturated disaccharide (Δ4,5-GalA-α1,2-L-Rha) and unsaturated tetrasaccharide (Δ4,5-GalA-α1,2-L-Rha-α1,4-D-GalA α1,2-L-Rha). BT4175 also released the di- and tetra-unsaturated oligosaccharides but, in addition, produced unsaturated hexaoligosaccharide as a main product. These end products were previously identified and the size was confirmed by mass spectrometry by Xiaoyang Zhang (Zhang 2015). Against AM-RG-I BT4183 generated four main unsaturated oligosaccharides with a degree of polymerization (DP) ≥ tetra. All three PLs released a range of long oligosaccharides from sugar beet arabinan evident by the smear of products observed on TLC (Figure 5.6), although BT4175 produced significantly more product than the other two enzymes including some smaller saccharides. BT4175 was the only enzyme active on P-RG-I. When the products of this reaction were analyzed by TLC a smear was observed indicating the release of a range of long oligosaccharides.

The PLs encoded by RG-I PUL displayed an endo mode of action (Figure 5.7). When incubated with AM-RG-I BT4170, BT4175 and BT41783 produced long oligosaccharides, not a single reaction product (characteristic of the exo acting enzymes). As the reactions progressed, the long oligosaccharides were cleaved and products with a low DP accumulated.

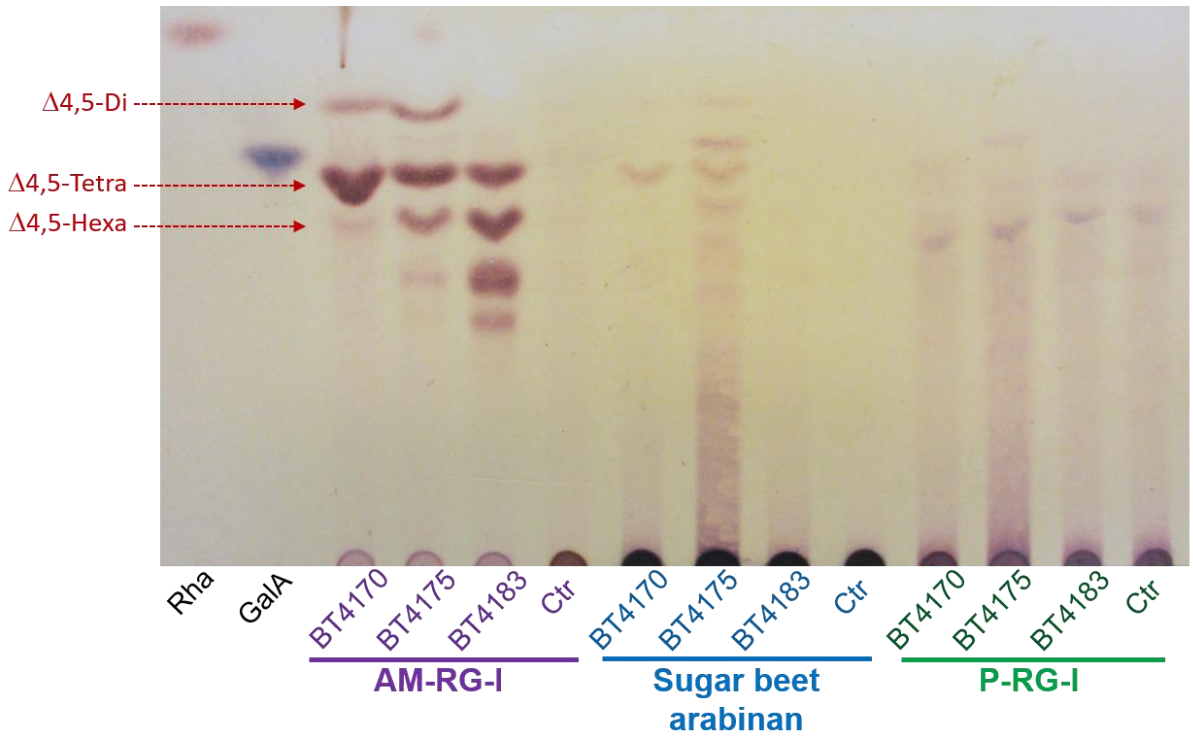


Figure 5.6 PLs products on different substrates

All reactions were performed with 500 nM of enzyme, 0.5% of substrate, 2 mM $CaCl_2$ and 100 mM Tris-HCl pH 7.5 (BT4183) or CAPSO pH 9.0 (BT4170 and BT4175). Reactions were incubated a 37 °C for 16 h. The (limit) reaction products were analysed by TLC (Chapter 2.2.3) and identified based on the migration profile compared with previously results obtained by (Zhang 2015). The BT4170 and BT4175 end products against AM-RG-I were previously identified and characterized by (Zhang 2015). (Rha) L-Rhamnose; (GalA) D-Galacturonic acid; (Ctr) Control of reaction containing substrate but not enzyme.

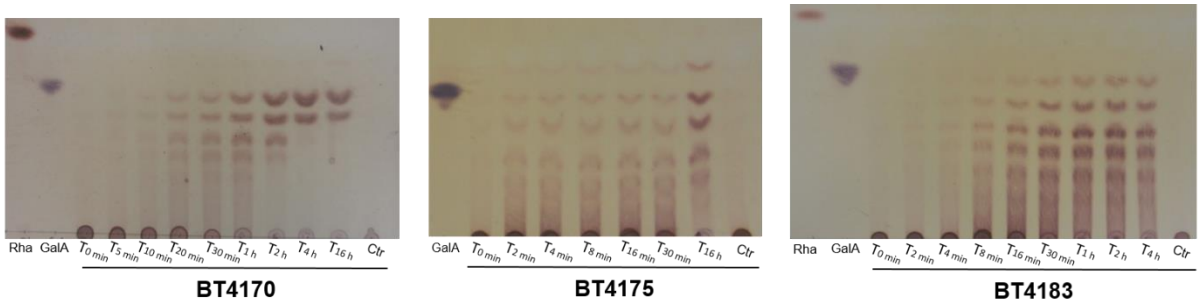


Figure 5.7 TLC showing the BT4170, BT4175 and BT4183 endo-activity

All reactions were performed at 37 °C with AM-RG-I 0.5%, 2 mM CaCl2 and 100 mM CAPSO pH 9.0 (BT4170 and BT4175) or Tris-HCl pH 7.5 (BT4183). The enzyme concentration was 30 nM for BT4170 and 300 nM for BT4175 and BT4183. Aliquots were taken at time intervals and reactions were stopped by heat inactivation and subject to TLC analysis (Chapter 2.2.3). (Rha) L-Rhamnose (GalA) D-Galacturonic acid standard; (Ctr) control reaction where substrate was incubated with buffer but no enzyme added.

5.3.3 *α-L-rhamnosidase BT4145 from RG-I utilization locus*

The RG-I PUL encodes a GH106 α-L-rhamnosidase, BT4145 that cleaves L-Rha at the non-reducing end of AM-RG-I (Zhang 2015). GH106 family contains 223 members but only one has been previously characterized (Miyata *et al*. 2005; Cantarel *et al*. 2009; Lombard *et al*. 2014). The mechanism of catalysis remains unknown in family GH106. The crystal structure and substrate specificity of another GH106 rhamnosidase active on RG-II was also determined in this thesis (BT0986, see Chapter 4.3.8). BT4145 was thus characterized and compared with BT0986 in order to explore the differences in specificity between these GH106 hydrolases.

5.3.3.1 Biochemical characterization

The catalytic activity of BT4145 was tested against different oligosaccharides containing L-Rha at the non-reducing end and D-GalA at the reducing end. These oligosaccharides were enzymatically produced and characterized by a previously PhD student (Zhang 2015). Due to the limited amount of long oligosaccharides the catalytic activity was determined by substrate depletion by high performance anion-exchange chromatography (HPAEC). For each oligosaccharide two different substrate concentrations were used and the rates obtained were constant proving that the experiment was performed below the K_M (Figure 5.8). Against the disaccharide L-Rha-α1,4-D-GalA, BT4145 displayed a k_{cat}/K_M of 3.38×10^7 min^{-1} M^{-1} (Figure 5.8). Increasing the size of the oligosaccharide did not have an impact on activity indicating that this enzyme only has two subsites.

A

Oligosacharide	Name	k_{cat}/K_M (min^{-1} M^{-1})
(Rha-GalA)	Di	3.28 x 10^7
$(Rha\text{-}GalA)_2$	Tetra	8.70 x 10^6
$(Rha\text{-}GalA)_3$	Hexa	1.47 x10^7

B

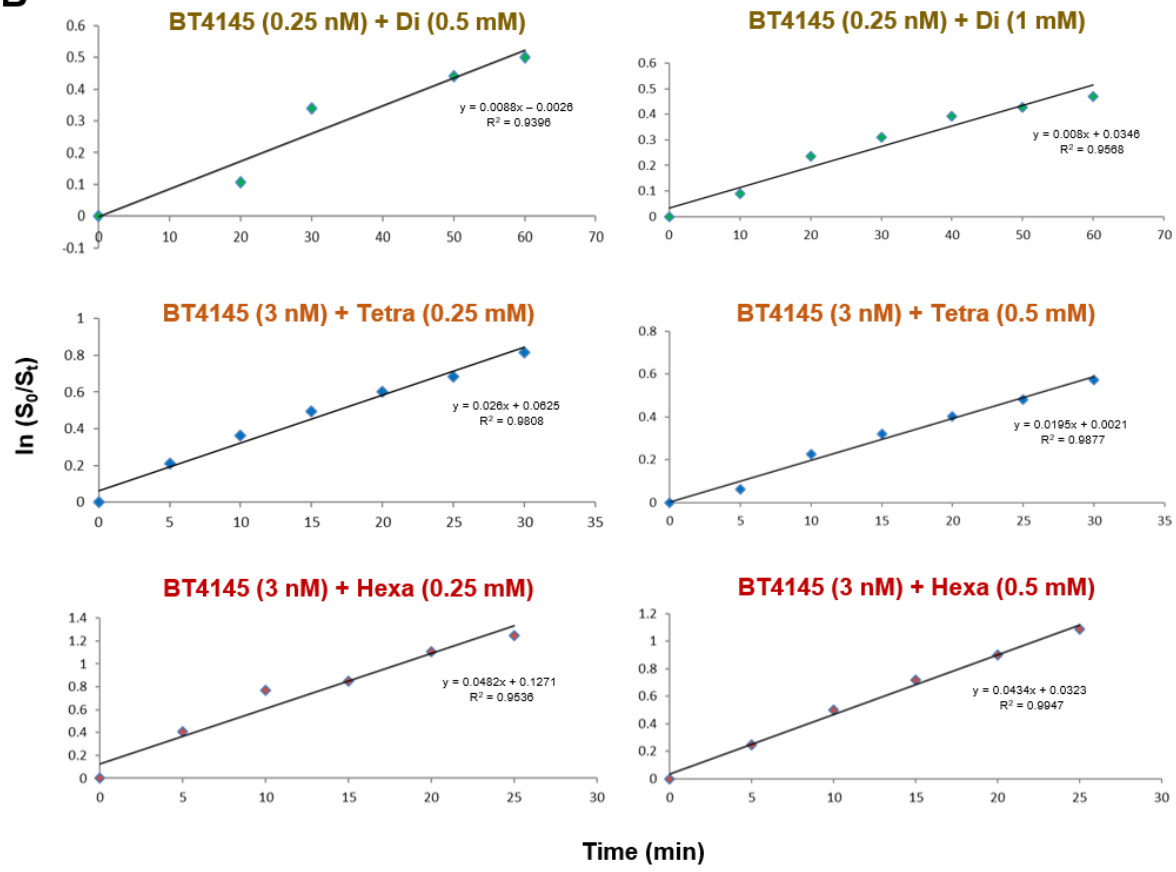


Figure 5.8 Catalytic activity of BT4145

A. Activity of BT4145 against different oligosaccharides in 100 mM sodium phosphate pH 7.0 and BSA 0.1 mg/ml. **B.** Graphical representation of the rates calculated using the substrate depletion method (see Chapter 2.2.5.3). The graphs were plotted in Microsoft Exel using regression linear. The oligosaccharide basic unit is (L-Rha-α1,4-D-GalA)$_n$ where n=1 (Di), 2 (tetra) or 3 (hexa).

The α-L-rhamnosidase was incubated with the disaccharide L-Rha-α1,4-D-GalA in the presence and absence of EDTA to test ion dependence of this enzyme (Figure 5.9). In the absence of EDTA, BT4145 cleaved the substrate in L-Rha and D-GalA. A very limited activity was observed when the reaction was performed in the presence of

EDTA. This result indicates that BT4145 is an ion depend enzyme. A similar result was shown for the GH106 homologue BT0986 (see Section 4.3.8.3).

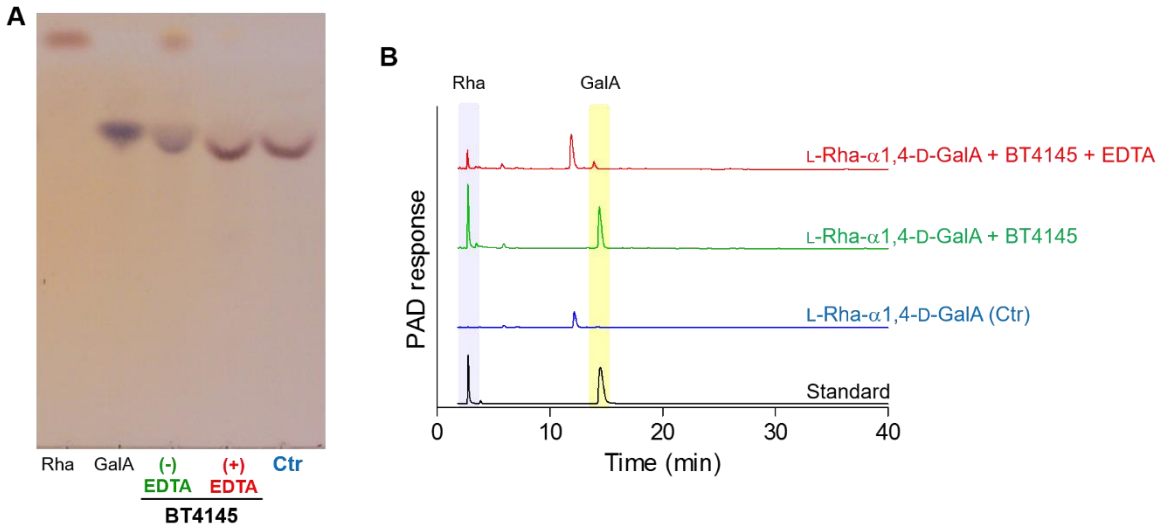


Figure 5.9 BT4145 ion dependence

Panel A and B show the reaction products were analysed by TLC (Chapter 2.2.3) and HPAEC (Chapter 2.2.4), respectively. The reactions were performed with 0.1 μM of BT4145, 10 mM of disaccharide and 5 mM of EDTA in 100 mM sodium phosphate pH 7.5. Reactions were incubated a 37 °C for 16 h. (Rha) L-Rhamnose; (GalA) D-Galacturonic acid; (Ctr) Control of the reaction; substrate without enzyme incubated in the same conditions.

5.3.3.2 Mechanism of catalysis by ^{1}H-NMR

The stereochemical course of the reaction catalyzed by BT4145 against α-L-Rha-α 1,4-D-GalA was analysed by ^{1}H-NMR (Figure 5.10). A L-Rha standard was run in the same conditions. After addition of the enzyme a peak appeared with a chemical shift of 4.84 ppm, characteristic of β-L-Rha (Pitson *et al*. 1998). These data indicate that this enzyme acts by inverting the anomeric configuration of the glycone sugar participating in the scissle glycosidic linkage (inverting mechanism), and thus mediates bond cleave through a single displacement mechanism. After 18.5 h an additional peak at 5.07 ppm, characteristic of α-L-Rha, was observed and is indicative of mutarotation and thus equilibrium between the two anomeric forms of the L-Rha reaction product (Pitson *et*

al. 1998). Additionally, the α and β peaks of L-Rha also match the peaks of this sugar standard suggesting a positive assignment of peaks resonances.

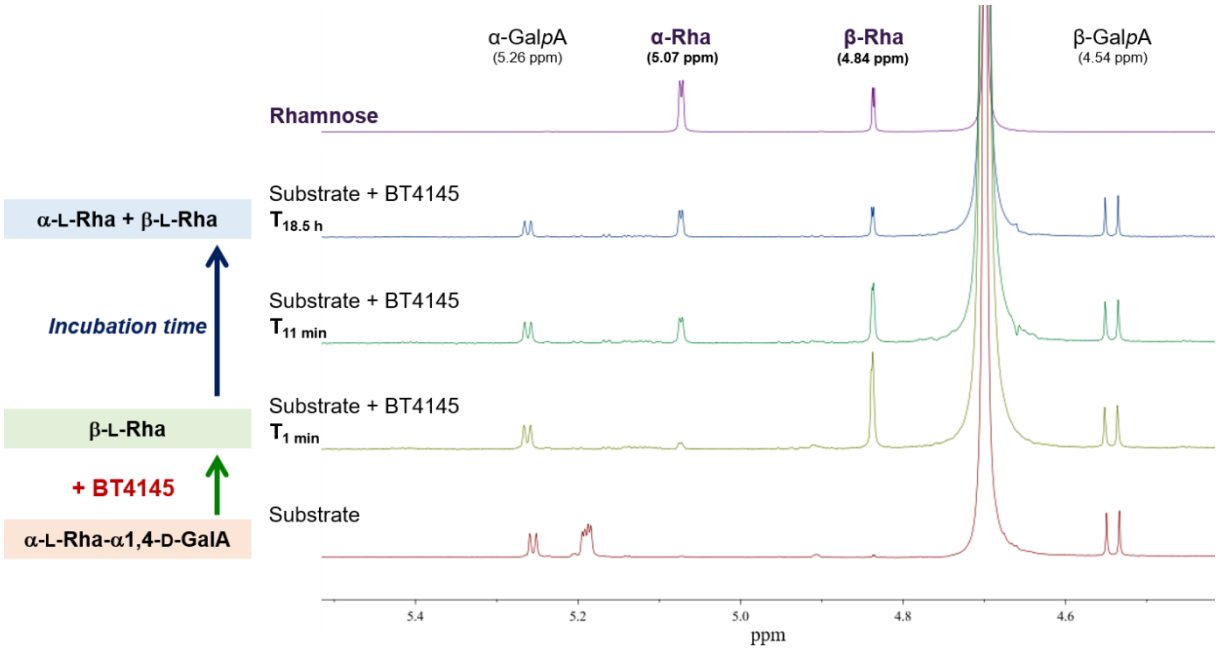


Figure 5.10 Partial ^{1}H-NMR of α-L-Rha-α1,4-D-GalA treated with BT4175

The reaction was performed with 20 mM of disaccharide and 35.5 µM of BT4145 in 20 mM Tris-HCl, 500 mM, pH7.5. All the components of the reaction were dissolved in deuterium oxide. The course of reaction was followed by ^{1}H-NMR (see Chapter 2.2.10). The peaks were assigned based on the literature (Pitson *et al*. 1998). For L-Rha a standard of this sugar (20 mM) was run in same conditions.

5.3.3.3 BT4145 proposed active site

In Chapter 4.3.8, the structure of a GH106 family member (BT0986) was described in detail. The structure of BT0896 in complex with a D-rhamnopyranose tetrazole and mutagenesis studies identified two glutamate residues, E461 and E593, as putative catalytic residues and the amino acids (D458, S459, E538, E561 and E593) that comprised the calcium binding site. In BT0986 the residues implicated in substrate recognition at the -1 subsite (Y49, F126, W563 and H627) were also described (see Figure 4.19 for details about BT0986 active site). In order to explore possible active site conservation within GH106 enzymes, the sequences of BT4145 and BT0986 were aligned (Figure 5.11), together with GH106 members from *Bacteroides* species that

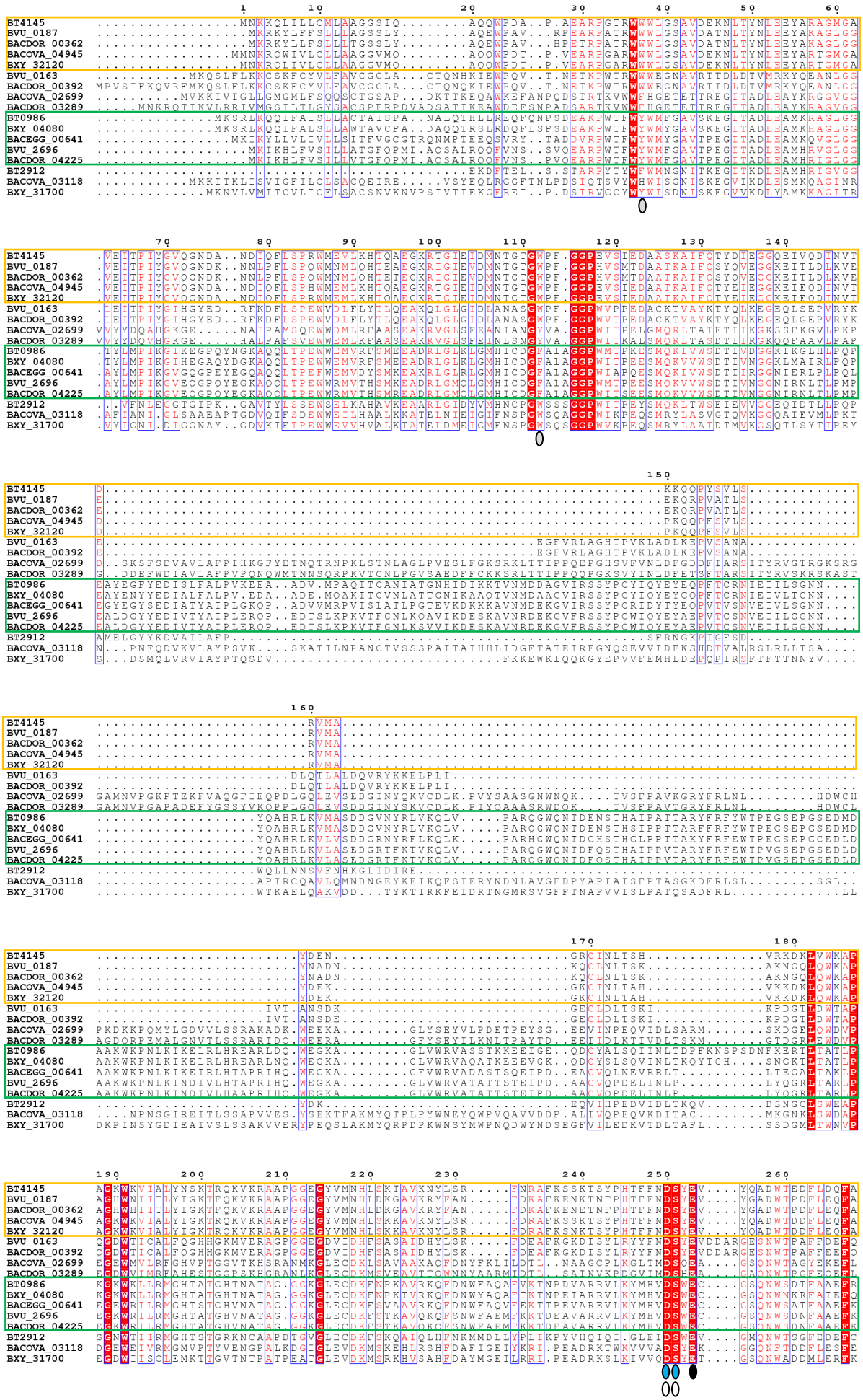
grew on RG-II (see Appendix I). The selected sequences correspond to PUL-encoded GH106 enzymes [PULs were either experimentally identified or predicted using the PULDB database (Terrapon *et al*. 2015)] and for *B. xylanisolvens* it was included an additional protein annotated in CAZy GH106 family (BXY_04080) (Cantarel *et al*. 2009; Lombard *et al*. 2014). Although BT4145 and BT0986 have only 23% sequence identity the proposed general catalytic acid/base pair is conserved in these two enzymes (Figure 5.11). The residues that contribute to the calcium binding site are also conserved with the exception of E538 in BT0986, which corresponds to Q328 in BT4145. The histidine involved in rhamnotetrazole recognition in BT0986 is also conserved in BT4145 (H429). The aromatic residues in the -1 subsite of BT0986 (Y49, F126 and W563 in BT0986) were structurally conserved in BT4145 (W37, W112 and F351, respectively).

The GH106 proteins BT0986 and BT4145 target L-Rha-α1,2-L-Ara*p* in Chain B of RG-II and L-Rha-α1,4-D-GalA in the backbone of RG-I, respectively. It is expected, therefore, that these enzymes will have different +1 subsites The alignment on Figure 5.11 showed that BT4145 does not contain a sequence equivalent to P179 to S364 of BT0986. In the crystal structure this sequence corresponds to the C domain (pink, Figure 5.12), which contributes to the leaving group (positive) subsites of the enzyme. The lack of this domain in BT4145 is consistent with the different specificities for the component of the substrate bound in the positive subsites.

The alignment of the 17 *Bacteroides* GH106 enzymes allowed the identification of four protein sequences sharing a high degree of sequence identity (70 to 89%) with BT4145 and four that are similar to BT0986 (65 to 84% sequence identity) (Figure 5.11). This may provide insight into the specificity of the homologues to the two *B. thetaiotaomicron* enzymes. Indeed, BXY_32120 and BACOVA_04945 (BT4145

homologues) were previously described as members of PULs that target RG-I (Martens *et al*. 2011; Despres *et al*. 2016). It is proposed that a careful alignment of GH106 family members can contribute to predicting the substrates targeted by these enzymes.

Site-directed mutagenesis was performed to test the importance of active site residues on the activity of BT4145. The proposed general acid/base pair (E253/E395) and the calcium binding residues (D250, Q328 and E394) were mutated to alanine (see Appendix D, Table D.3 for primers). The catalytic efficiency (k_{cat}/K_M) of BT4145 wild type and mutants were assessed spectrophotometrically against the disaccharide L-Rha-α1,4-D-GalA as described in Chapter 2.2.5.2 and are presented in Table 5.4. The wild type BT4145 showed a k_{cat}/K_M of 7.13 x 10^7 min^{-1} M^{-1}, a similar value to the one obtained by substrate depletion HPAEC (see Figure 5.8). All the mutants were inactive against the disaccharide, suggesting that these amino acids are essential for the catalysis process. This result is similar to that obtained for BT0986 by Dr. Aurore Labourel (see Table 4.8), confirming that the predicted catalytic apparatus is conserved between proteins and suggesting that the calcium ion has a major role in enzyme activity.



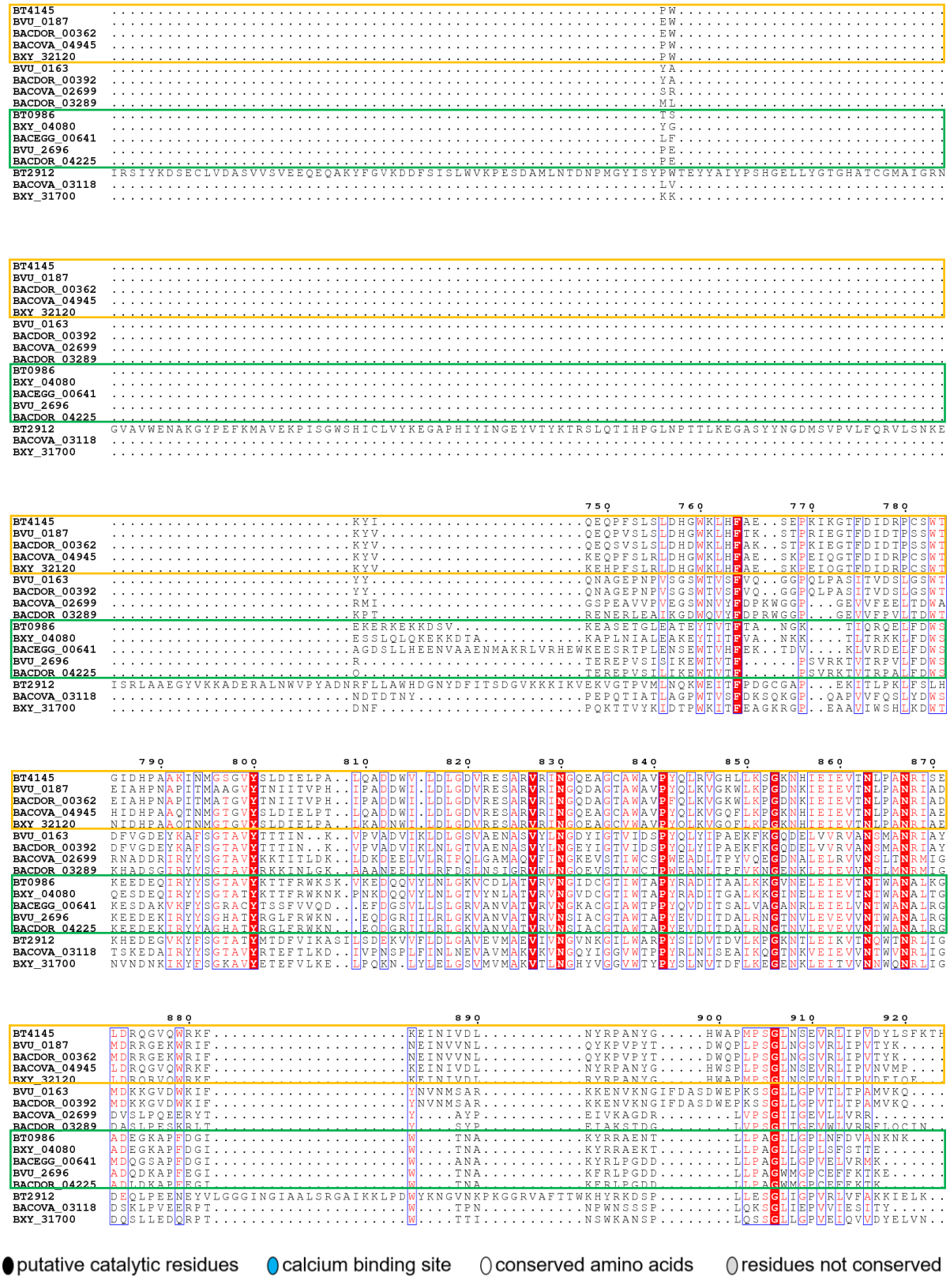


Figure 5.11 BT4145 and BT0986 alignment

The proteins were aligned in MUSCLE (Edgar 2004a) and visualized using ESPrit 3.0 (Robert and Gouet 2014). Orange and green boxes indicate the protein sequences similar to BT4145 and BT096, respectively. Amino acids with 100% of conservation are highlighted in a red background and residues that are functionally highly conserved are boxed and coloured red. The residues numeration is according with BT4145 sequence (Cantarel *et al*. 2009; Lombard *et al*. 2014).

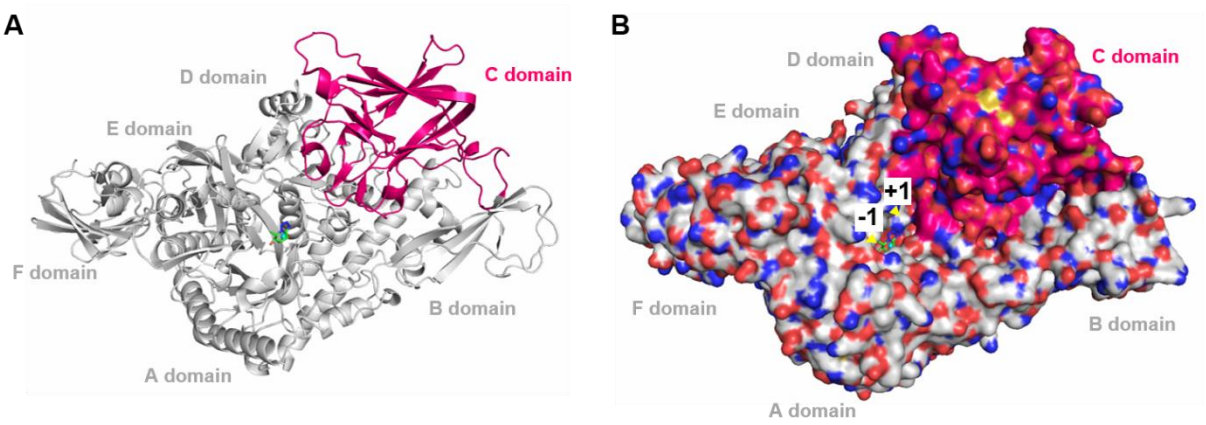


Figure 5.12 GH106 protein structure

The figure shows BT0986 cartoon (Panel A) and surface (Panel B). The domain colour in pink indicates the sequence that is missing in BT4145. The enzymes subsites -1 and +1 are labelled in Panel B. The D-rhamnopyranose tetrazole coloured in green is present in the active site and the calcium ion in the active site is shown as a cyan sphere. The figures were generated in PyMOL using the PDB of BT0986 (Chapter 4.3.8).

Table 5.4 Catalytic activity of BT4145 mutants

Recombinant protein		k_{cat}/K_M (min^{-1} M^{-1})
Wild type	BT4145	$7.13 \times 10^7 \pm 2.68 \times 10^5$
Mutant	D250A	No activity
	E253A	No activity
	Q328E	No activity
	E349A	No activity
	E395A	No activity

Catalytic activity of BT4115 and different mutants against L-Rha-α1,4-D-GalA in 20 mM sodium phosphate buffer pH 7.5 and BSA 0.1 mg/ml. The catalytic efficiency was calculated directly using a single substrate concentration (see chapter 2.2.5.2) and standard errors of the mean were generated from at least three technical triplicates.

5.3.4 *The four GH28 enzymes from RG-I locus: BT4146, BT4149, BT4153, BT4155*

It is predicted that RG-I PUL encodes four GH28 enzymes. The activities for BT4153 and BT4155 were identified by Zhang. BT4153 is a exo-rhamnogalacturonase that cleaves D-GalA from the non-reducing end of the (α1,4-D-GalA-α1,2-L-Rha)$_n$ chains. BT4155 is active on P-RG-I and HG generating D-GalA (Zhang 2015). In order to understand the mechanism of RG-I utilization by *B. thetaiotaomicron*, the specificity of the different GH28s enzymes are explored in this Section.

5.3.4.1 Production and purification of oligosaccharides

To characterize the GH28 enzymes it was necessary to produce oligosaccharides containing D-GalA at the non-reducing end. Taking advantage of the characterized enzymes encoded by the RG-I PUL, the substrate AM-RG-I was digested to generate the required D-GalA oligosaccharides (Figure 5.13). In a first reaction AM-RG-I was digested with the rhamnogalacturonan lyase BT4183 to completion. This reaction produced four oligosaccharides containing Δ4,5-GalA at the non-reducing end. The digestion of these oligosaccharides with the GH105, BT4174, cleaved the unsaturated Δ4,5-GalA and exposed L-Rha at the non-reducing end. These oligosaccharides were then digested with the α-rhamnosidase, BT4145, which removed the rhamnose generating oligosaccharides (di-, tetra-, hexa- and octa-) containing D-GalA at the non-reducing end.

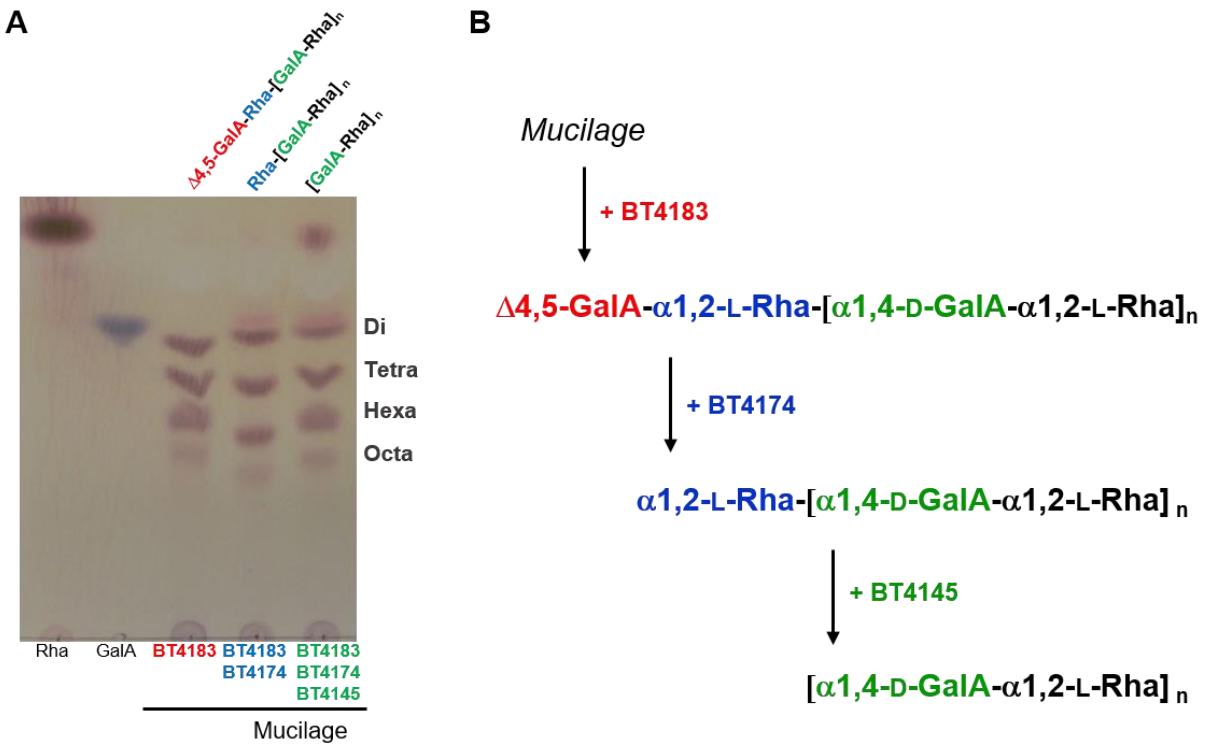


Figure 5.13 Production of D-Gal-α1,2-L-Rha oligosaccharides

A. Reaction products of each reaction analysed by TLC (see Chapter 2.2.3). **B.** Schematic representation of the reactions. For oligosaccharides di, tetra, hexa and octa "n" means 1, 2, 3 or 4 respectively. The reactions were performed with 1 μM of enzyme, 0.5% AM-RG-I, 100 mM Tris-HCl pH 7.5 and 2 mM $CaCl_2$. All reactions were performed at 37 °C for 16 h. (Rha) L-Rhamnose; (GalA) D-Galacturonic acid.

The oligosaccharides produced were resolved by size exclusion. Figure 5.14 shows the purity of the oligosaccharides in the different fractions collected during the purification process. The size of all oligosaccharides was confirmed by LC-MS Figure 5.14B). Further experiments with these oligosaccharides will be discussed in the present Section 5.3.4.

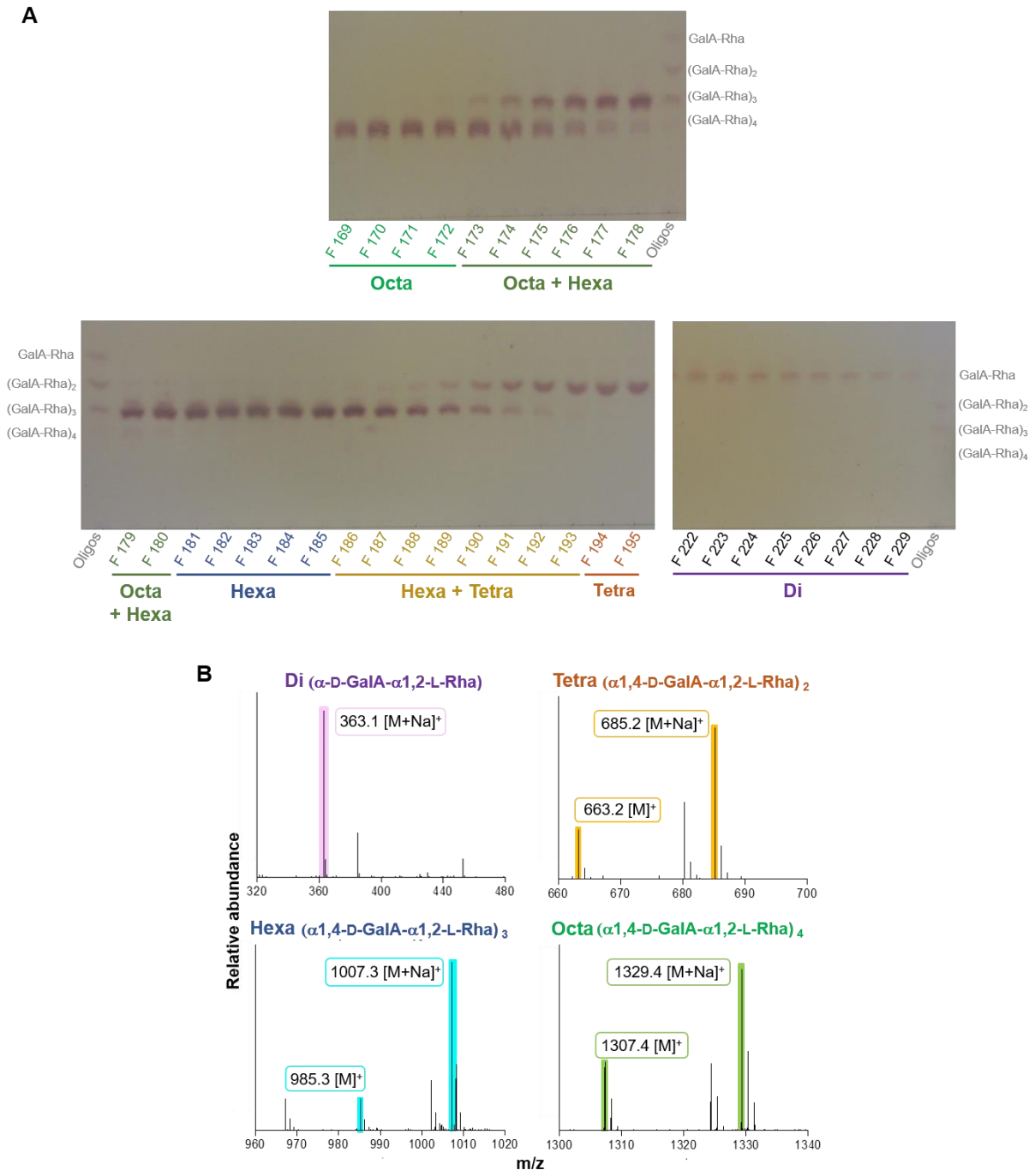


Figure 5.14 Purification and LC-MS analysis of oligosaccharides

A. Analysis by TLC (Chapter 2.2.3) of the different oligosaccharides purified by size exclusion chromatography. F correspond to a 2 ml fraction collected during purification. The size exclusion chromatography was performed in 50 mM acetic acid as buffer (Chapter 2.2.7). **B.** Analysis of oligosaccharides by LC-MS. The peak labelled in the LC-MS chromatograph corresponds to the mass (M) of the oligosaccharide and mass plus sodium adduct (Na = 23 Da) in positive mode. (GalA) D-Galacturonic acid; (Rha) L-Rhamnose.

5.3.4.2 BT4146, BT4153 and BT4155: Substrate specificity

The GH28 enzymes, BT4146, BT4153 and BT4155 were tested against different oligosaccharides and polysaccharides containing D-GalA at the non-reducing end.

The three enzymes were incubated with AM-RG-I oligosaccharides (see Section 5.3.4.1) and the release of D-GalA evaluated by TLC and HPAEC (Figure 5.15). BT4153 was the only enzyme active against all (α1,4-D-GalA-α1,2-L-Rha)$_n$ oligosaccharides tested (di, tetra, hexa and octa). BT4155 and BT4146 were only active against the disaccharide α1,4-D-GalA-α1,2-L-Rha generating D-GalA and L-Rha.

BT4155 hydrolysed commercially available HG-derived oligosaccharides D-GalA-α1,4-D-GalA (di) and D-GalA-α1,4-D-GalA-α1,4-D-GalA (tri) generating a single product, D-GalA (Figure 5.16). Additionally, BT4155 also released D-GalA when incubated with R-RG-I and Pectin (HG polysaccharide) (Figure 5.17). The release of D-GalA when this enzyme is incubated with P-RG-I can be explained by the presence of HG fragments attached to this polysaccharide (see Section 5.4). BT4146 and BT4153 failed to cleave HG oligosaccharides (Figure 5.16) and, for BT4146, no activity was observed against HG or RG-I polysaccharides (Figure 5.17).

These results sugest that BT4153 can cleave D-GalA from the non-reducing end of AM-RG-I. BT4146 only recognizes the disaccharide D-GalA-α1,2-L-Rha. BT4155 is active against HG releasing D-GalA but is also able to cleave D-GalA-α1,2-L-Rha.

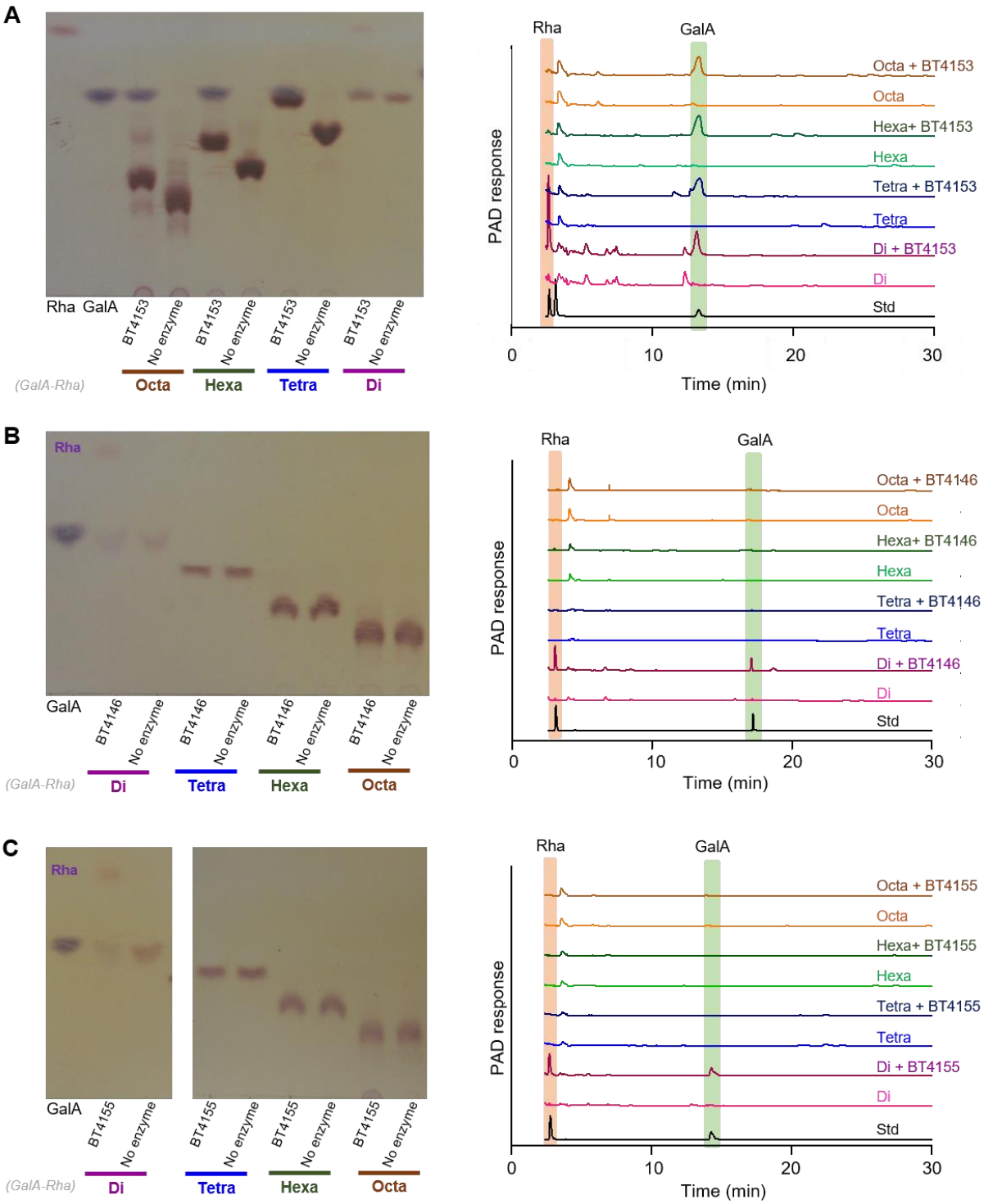


Figure 5.15 BT4146, BT4153 and BT4155 activity against α1,4-D-GalA-α1,2-L-Rha oligosaccharides

The figure shows the reaction products for BT4153 (A), BT4146 (B) and BT4155 (C) analysed by TLC and HPAEC. All reactions were carried out under standard conditions (Chapter 2.2.5) with 5 mM (Panel A) or 1 mM (Panels B and C) of oligosaccharides, 0.1 mg/ml of BSA in 100 mM of sodium phosphate pH 7.5 (BT4153) or MES pH 6.0 for (BT4146 and BT4155). No enzyme reaction represents the control performed in the same condition. For TLC and HPAEC methods see Chapter 2.2.3 and Chapter 2.2.4, respectively. (Rha) L-Rhamnose; (GalA) D-Galacturonic acid; (Std) Standard of Rha and GalA.

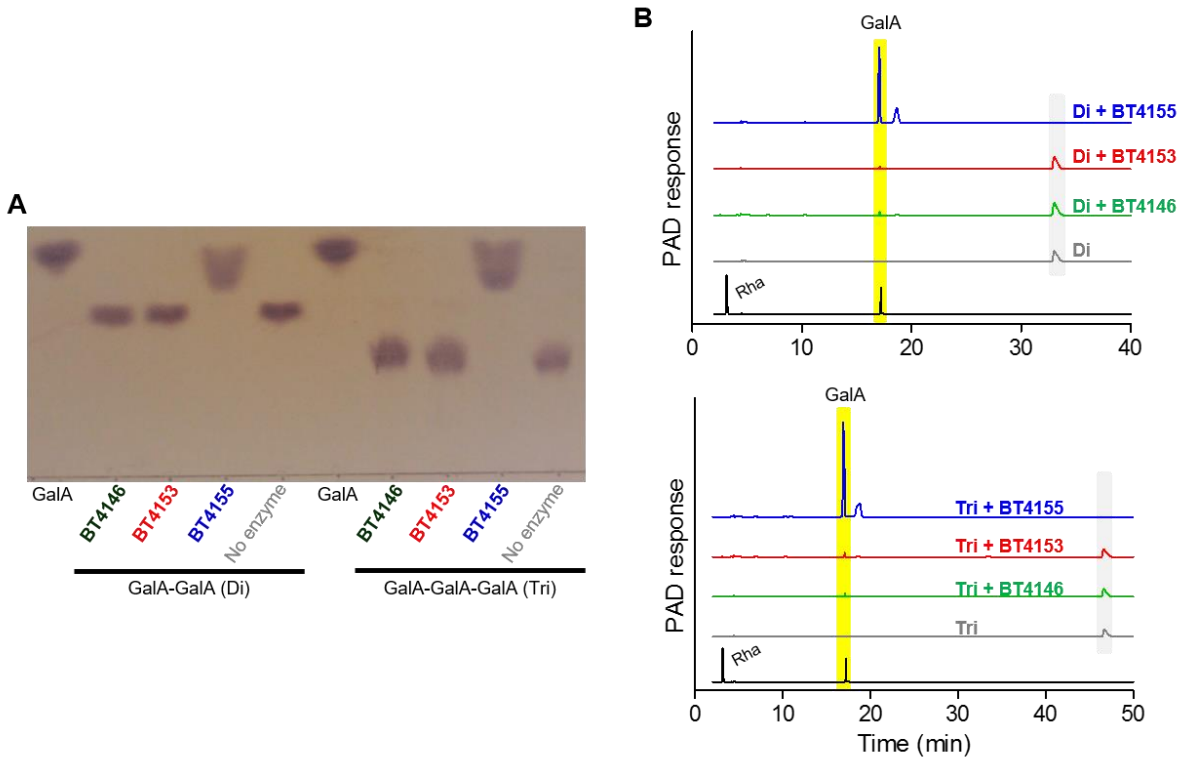


Figure 5.16 BT4146, BT4153 and BT4155 activity against D-GalA-α1,4-D-GalA oligosaccharides

The figure shows the reaction products analysed by TLC (A) and HPAEC (B). All reactions were performed under standard conditions (Chapter 2.2.5) with 2 mM of oligosaccharides, 0.1 mg/ml of BSA in 100 mM MES pH 6.0. No enzyme reaction (grey) represents the control performed in the same condition. For TLC and HPAEC methods see Chapter 2.2.3 and Chapter 2.2.4, respectively. The HPAEC products (highlighted in yellow) were identified using a standard (black line in chromatograms). The peaks of the initial oligosaccharides are highlighted in grey. (GalA) D-Galacturonic acid; (Rha) L-Rhamnose.

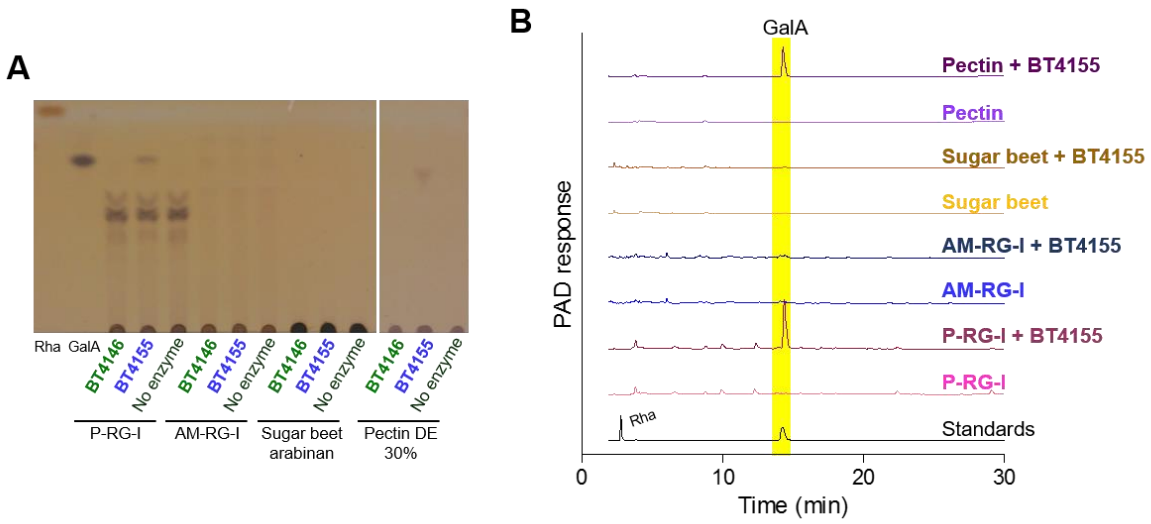


Figure 5.17 Activity of BT4146 and BT4155 against polysaccharides

A. Reaction products of BT4155 and BT4146 analysed by TLC (Chapter 2.2.3). **B.** HPAEC analyses of BT4155 reactions. (Chapter 2.2.4) All reactions were performed with 1 μM of BT4155 or 10 μM of BT4146 and, 0.5% of polysaccharide, 0.1 mg/ml of BSA in 100 mM of buffer MES pH 6.0. Reactions were incubated at 37 °C for 16 h. No enzyme reaction represents the control performed in the same condition without enzyme. (Rha) L-Rhamnose; (GalA) D-Galacturonic acid; (AM-RG-I) *Arabidopsis* rhamnogalacturonan-I (P-RG-I) Potato rhamnogalacturonan-I.

5.3.4.3 Catalytic activity of BT4146, BT4153 and BT4155

In order to clarify the specificity of each enzyme for their respective substrates the catalytic activity for BT4146, BT4153 and BT4155 was determined and is shown in Table 5.5. The enzymes displayed a similar activity against D-GalA-α1,2-L-Rha. The activity of BT4153 against the tetrasaccharide and disaccharide were similar indicating that the enzyme only has two subsites. The k_{cat}/K_M of BT4155 for D-GalA-α1,2-L-Rha (9.23×10^5 $min^{-1}M^{-1}$) was only 3 fold higher than D-GalA-α1,4-D-GalA (3.16×10^5 $min^{-1}M^{-1}$), indicating that the enzyme is able to cleave both disaccharides with similar efficiency. The increase in DP of the oligosaccharide did not have a major impact on the catalytic efficiency of BT4155 suggesting that the enzyme only has two subsites.

Table 5.5 Catalytic activity of BT4146, BT4153 and BT4155

Recombinant protein	Substrate	k_{cat}/K_M (min^{-1} M^{-1})
BT4146	GalA-Rha	$2.58 \times 10^5 \pm 2.35 \times 10^4$
BT4153	GalA-Rha	$1.96 \times 10^6 \pm 1.41 \times 10^5$
	GalA-Rha-GalA-Rha	$4.49 \times 10^6 \pm 3.09 \times 10^4$
BT4155	GalA-GalA	$3.16 \times 10^5 \pm 2.21 \times 10^4$
	GalA-GalA-GalA	$2.53 \times 10^5 \pm 1.16 \times 10^4$
	GalA-Rha	$9.23 \times 10^5 \pm 6.77 \times 10^3$
	GalA-Rha-GalA-Rha	No activity
	Pectin DE 30%	$6.93 \times 10^5 \pm 2.41 \times 10^4$
	P-RG-I	$1.11 \times 10^5 \pm 2.41 \times 10^4$

Catalytic activity of the GH28 enzymes against different oligosaccharides in 100 mM MES pH 6.0 and 0.1 mg/ml BSA. The k_{cat}/K_M calculated using a single substrate concentration below K_M as described in Chapter 2.2.5.2. The standard errors were generated from technical triplicates. (GalA) D-Galacturonic acid; (Rha) L-Rhamnose; (DE) Degree of esterification; (P-RG-I) Potato rhamnogalacturonan-I.

5.3.4.4 BT4149 sequence and protein homologues

The additional GH28 protein encoded by RG-I PUL is BT4149. In order to characterize this protein, it was necessary to clone *bt4149* into an expression vector. However, when *bt4149* was cloned from the *B. thetaiotaomicron* genome it contained a single nucleotide insertion close to the 3' end compared to the published gene sequence. This results in a different C-terminal amino acid sequence (Figure 5.18). It was hypothesized that this nucleotide insertion was present in the *B. thetaiotaomicron* genome used as a template to amplify this gene and not an artefact introduced during PCR. The protein sequence of BT4149 based on the published sequence of the *B. thetaiotaomicron* genome and the new proposed sequence for this protein (H_BT4149) were aligned with the ten closet protein homologues (Figure 5.19) using MUSCLE (Edgar 2004a; Edgar 2004b). The alignment showed that H_BT4149 had a size similar to the homologues, while BT4149 was shorter. Furthermore, the extended C-terminus

in H_BT4179 (generated from the nucleotide insertion) showed very high sequence identity with the homologues. This indicates that H_BT4179 is the correct sequence.

Figure 5.18 Comparison between the BT4149 sequences

The DNA sequences and respective translation are shown. BT4149 represents the sequence available in the KEGG database. The proposed sequence for BT4179 (H_BT4149) results from a nucleotide insertion (highlighted in red). The protein sequence conserved between proteins is underlined in green. The protein sequence underlined in yellow and red are only present in the database or in the proposed sequence, respectively. The translation of the nucleotide sequence was performed with the online ExPASy translate tool (http://web.expasy.org/translate/).

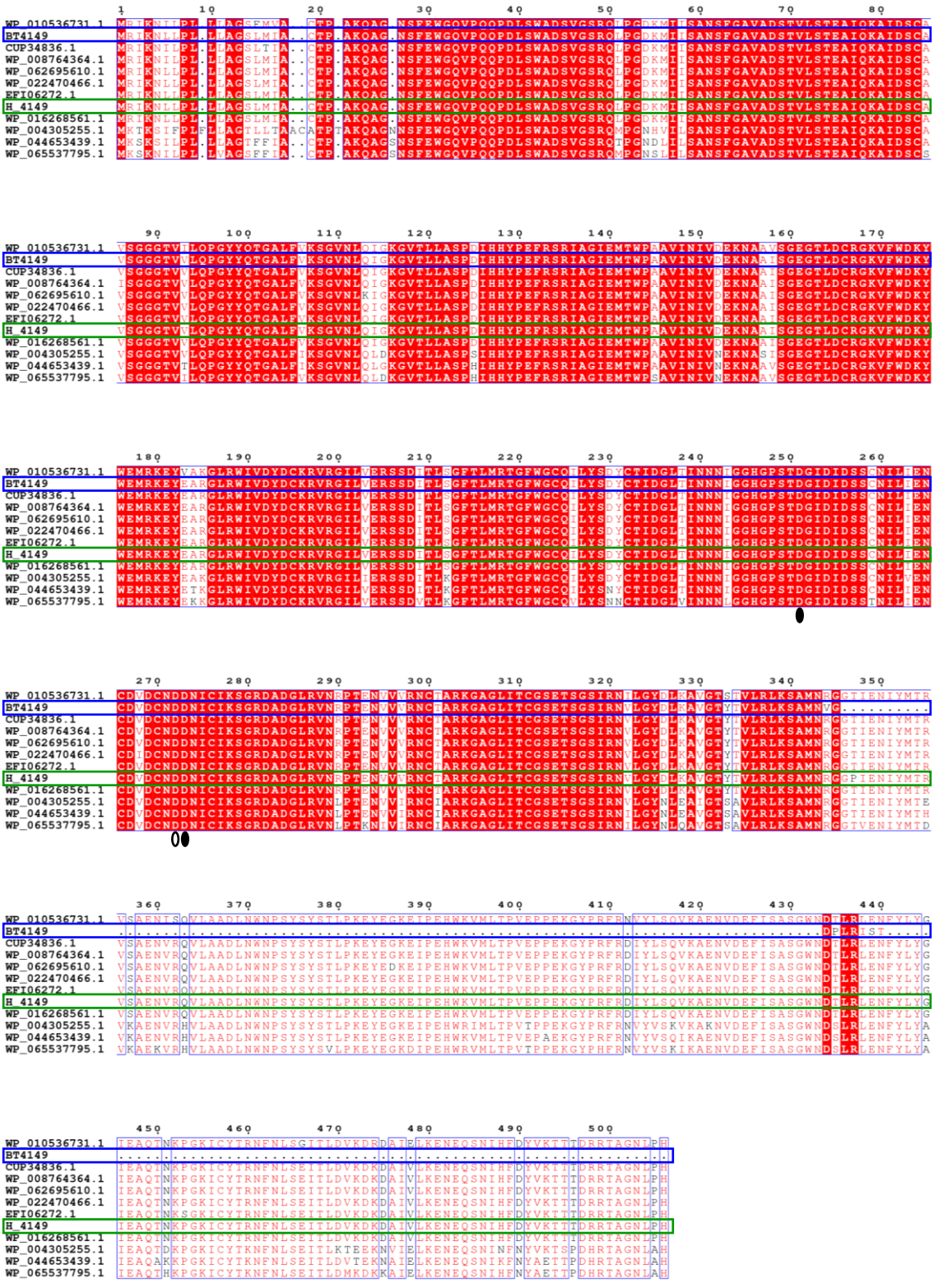


Figure 5.19 Alignment of BT4149

On the left are shown the NCBI reference codes for BT4179 homologues while BT4149 (blue box) represents the sequence available in the database and H_BT4149 (green box) the proposed sequence (see Figure 5.18). Sequences were aligned in MUSCLE (Edgar 2004a) and visualized using ESPrit 3.0 (Robert and Gouet 2014). The invariant catalytic apparatus of a GH28 family member is indicated with a black circle for general base and a white circle for the general acid (van Santen *et al*. 1999). Amino acids with 100% are conservation highlighted in a red background and residues that are functionally highly conserved are boxed and coloured red.

5.3.4.4.1 BT4149 protein homologue

To identify the activity of BT4149 a close homologue (WP_010536731.1 presenting 97% identity with H_BT4149, Figure 5.19) was characterized. The enzyme, defined as Bf4149 is derived from *B. faecis* a member of the human microbiota. The Bf4149 gene was chemically synthesized (see protein sequence in Appendix B, Figure B.2) and the subsequent cloning, protein expression and characterization was performed by Dr. Immacolata Venditto (ICaMB, Newcastle University). The enzyme was not active against HG oligosaccharides but hydrolyzed AM-RG-I oligosaccharides (Table 5.6). Bf4149 was 10,000 fold more active against the tetrasaccharide than the disaccharide but showed a modest reduction in activity (compared to the tetrasaccharide) against the hexa- and octa-oligosaccharides. Additional experiments showed that when Bf4149 was incubated with AM-RG-I in the presence of BT4145, the polysaccharide was cleaved to D-GalA and L-Rha (personal communication, Harry Gilbert). A similar result was obtained by Xiaoyang Zhang when BT4153 and BT4145 were incubated with AM-RG-I (Zhang 2015), suggesting that Bf4149 and BT4153 display overlapping activities. However, Bf4149 has a preference for the tetrasaccharide or longer oligosaccharides, while BT4153 is optimized to cleave the disaccharide D-GalA-α1,2-L-Rha. It is predicted that BT4149 displays the same enzymatic specificity described for the closer homologue Bf4149.

Table 5.6 Activity of BT4149 homologue

Substrate		Activity	k_{cat}/K_M (min^{-1} M^{-1})
D-GalA-α1,2-L-Rha Oligosaccharides	Di	+	$2.8 \times 10^4 \pm 1.3 \times 10^3$
	Tetra	+	$2.0 \times 10^8 \pm 2.8 \times 10^7$
	Hexa	+	$5.2 \times 10^7 \pm 7.2 \times 10^6$
	Octa	+	$4.8 \times 10^7 \pm 7.8 \times 10^6$
D-GalA-α1,4-D-GalA Oligosaccharides	Di	-	NA
	Tri	-	NA
Polysaccharides	P-RG-I	-	NA
	AM-RG-I	-	NA
	Sugar beet arabinan	-	NA
	PGA orange	-	NA
	Pectin DE 30%	-	NA

(+) D-GalA was detected by TLC and HPAEC; (-) no reaction products were detected; (NA) No activity detected. Activities determined by Immacolata Venditto (ICaMB, Newcastle University) in 100 mM MES pH 6.0 and 0.1 mg/ml BSA.

5.3.5 *Exploring the SGBP and SusD-like proteins of RG-I PUL*

The RG-I PUL encodes two SusD-like proteins, BT4165 and BT4169 that, potentially, bind RG-I or oligosaccharides that will then be imported by the respective SusC-like proteins (BT4164 and BT4168). *B. thetaiotaomicron* can also contain additionally SGBPs located on the outer membrane. The SGBP genes are usually positioned close to the *susC/D*-like sequences. Based on these criteria the RG-I PUL encode three potential SGBPs. The binding capacity of the two SusD-like proteins and three putative SGBPs were screened against different polysaccharides by affinity gel electrophoresis (AGE) (Figure 5.20). One of the hypothetical SGBPs, BT4167, showed electrophoretic retardation in gels containing AM-RG-I, P-RG-I and sugar beet arabinan, glycans that all contain the repeating L-Rha-α1,4-D-GalA backbone. Surprisingly, BT4167 also appeared to bind HG and RG-II that contain a backbone of α1,4-linked D-GalA residues. Indeed, the only polysaccharide that did not influence electrophoretic

migration was HG with DE 85%. This suggests that the esterification of the ligand can have a major impact on BT4167 binding. AGE revealed no binding for the SusD-like proteins or the other two hypothetical SGBPs. It is possible, however, that these putative proteins might recognize oligosaccharides produced by enzymatic cleavage at the cell surface or additional RG-I structures not present in the polysaccharides tested.

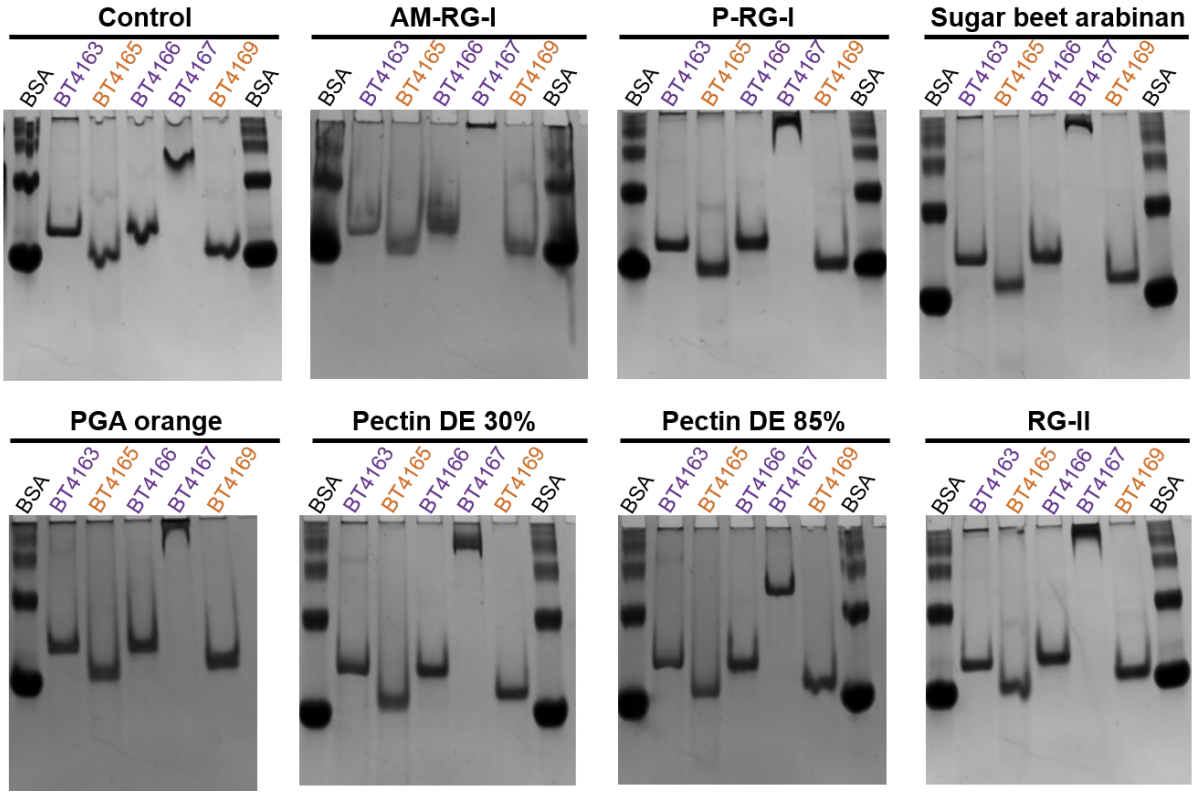


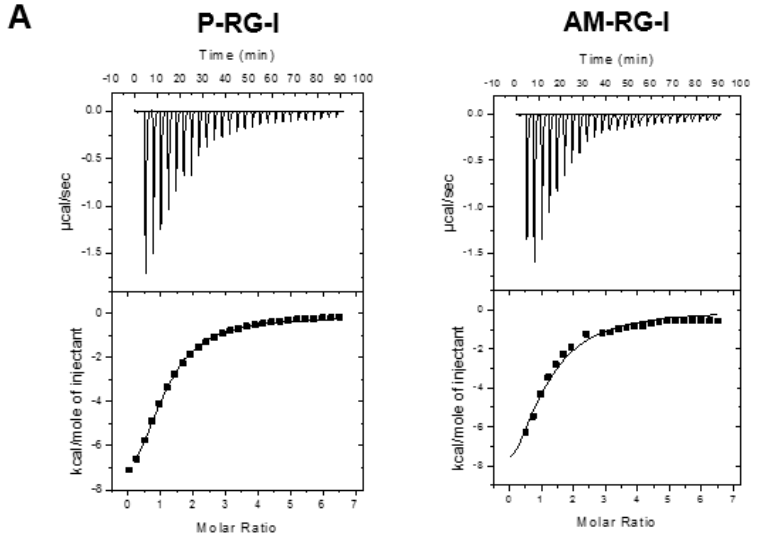
Figure 5.20 SGBP and SusD affinity gel electrophoresis

AGE (Chapter 2.2.1) was performed in non-denaturing conditions using 10% gels. Ligand-gels contained 0.2% polysaccharide, while the control gel lacked polysaccharide. BSA and different proteins (5 µg) were applied to all gels. (AM-RG-I) *Arabidopsis* rhamnogalacturonan-I (P-RG-I) Potato rhamnogalacturonan-I; PGA (Polygalacturonic acid), DE (degree of esterification).

The thermodynamics of the binding of BT4167 to its polysaccharide ligands was determined by isothermal titration calorimetry (ITC). The data, Figure 5.21, showed that BT4167 displayed similar affinity for P-RG-I and AM-RG-I with a K_A of 5.7 x 10^4 M^{-1} and 4.3 x 10^4 M^{-1}, respectively. The binding of BT4167 to PGA orange, however, was

too weak to quantify (data not shown). This indicates that this SGBP preferentially recognizes α1,4-D-GalA-α1,2-L-Rha chains tolerating substitutions at O4 of L-Rha.

A



B

Ligand	K_A M^{-1}	ΔG *kcal mole*$^{-1}$	ΔH *kcal mole*$^{-1}$	TΔS *kcal mole*$^{-1}$	n
P-RG-I	5.7 (±0.3) x 10^4	-6.48	-11.37 (±0.4)	-4.89	1.02 ± 0.031
AM-RG-I	4.3 (±0.4) x 10^4	-6.32	-13.51 (±0.5)	-7.19	1.00 ± 0

Figure 5.21 BT4167 binding thermodynamics parameters

A. ITC curves of BT4167 binding to different pectins. ITC titrations were carried out as described in Chapter 2.2.2 with 30 μM of BT4167 (cell) and the substrate concentration (syringe) was 0.5% of AM-RG-I or 1% of P-RG-I in 50 mM HEPES, 200 mM NaCl pH 7.0. The top part of each panel corresponds to raw ITC data and the bottom part represents the integration of the area peaks using MicroCal Origin 7 software. **B.** Determination of thermodynamic parameters and affinity constant of the BT4167 for different ligands (for details see Chapter 2.2.2). The thermodynamic parameters were as follows K_A is the association constant, ΔG is the change in free enthalpy, ΔH is the enthalpy of binding, T is the temperature in degree absolute and ΔS is the entropy of the binding. The errors correspond to the fitting of the ITC curve using the software MicroCal Origin7. (DE) Degree of esterification.

5.3.6 *Growth curves of B. thetaiotaomicron mutants on different substrates*

In order to explore the significance of the different enzymes in pectins utilization, the genes encoding the proteins of interest were deleted form the *B. thetaiotaomicron*

genome and the growth of these knockout (KO) strains was tested on different polysaccharides. The deletions were created using the allelic exchange method described in Chapter 2.1.23. Briefly, in frame deletions were generated by cloning the flanking sequence of the target gene into the suicide vector pExchange-*tdk*. The plasmid was conjugated into a Δ*tdk* mutant of *B. thetaiotaomicron* and transconjugants were selected using different resistance markers. The genomic modifications were confirmed by sequencing of PCR products. The primers utilized to generate the mutants and examples of agarose gels presenting the different PCR products are presented in Appendix J.

The growth curves for the KO strains are presented in Figure 5.22. In all of the recombinant mutants the gene deletion did not have an effect on growth on glucose indicating that none of these genes display a vital role. The mutant *Δbt4170* displayed no growth on AM-RG-I. The mutant, however, grew similar to the wild type bacterium on sugar beet arabinan, HG and P-RG-I (blue box, Figure 5.22), which contain metabolisable glycans other than the RG-I backbone (Martens *et al*. 2011). These data suggest that BT4170 has an essential role in the utilization of the RG-I backbone (see below Section 5.3.7.1 for further details).

The RG-I PUL encodes two pairs of SusC/SusD-like proteins. It was hypothesized that these proteins could have a different role in importing oligosaccharides into the periplasm. The mutant strains of these two SusC/D-like pairs were generated in order to explore their functional significance. The data, Figure 5.22 (purple box), showed that both mutants showed a significant decrease in growth on AM-RG-I indicating that both SusC/D-like proteins are important in RG-I utilization, and thus do not display functional redundancy. Recent data showed that SusC/D-like outer membrane importers dimerize (Glenwright *et al*. 2016). In the RG-I PUL it is possible that the SusC/D-like

pairs are interacting with each other creating heterodimers that require the presence of both pairs to be stable.

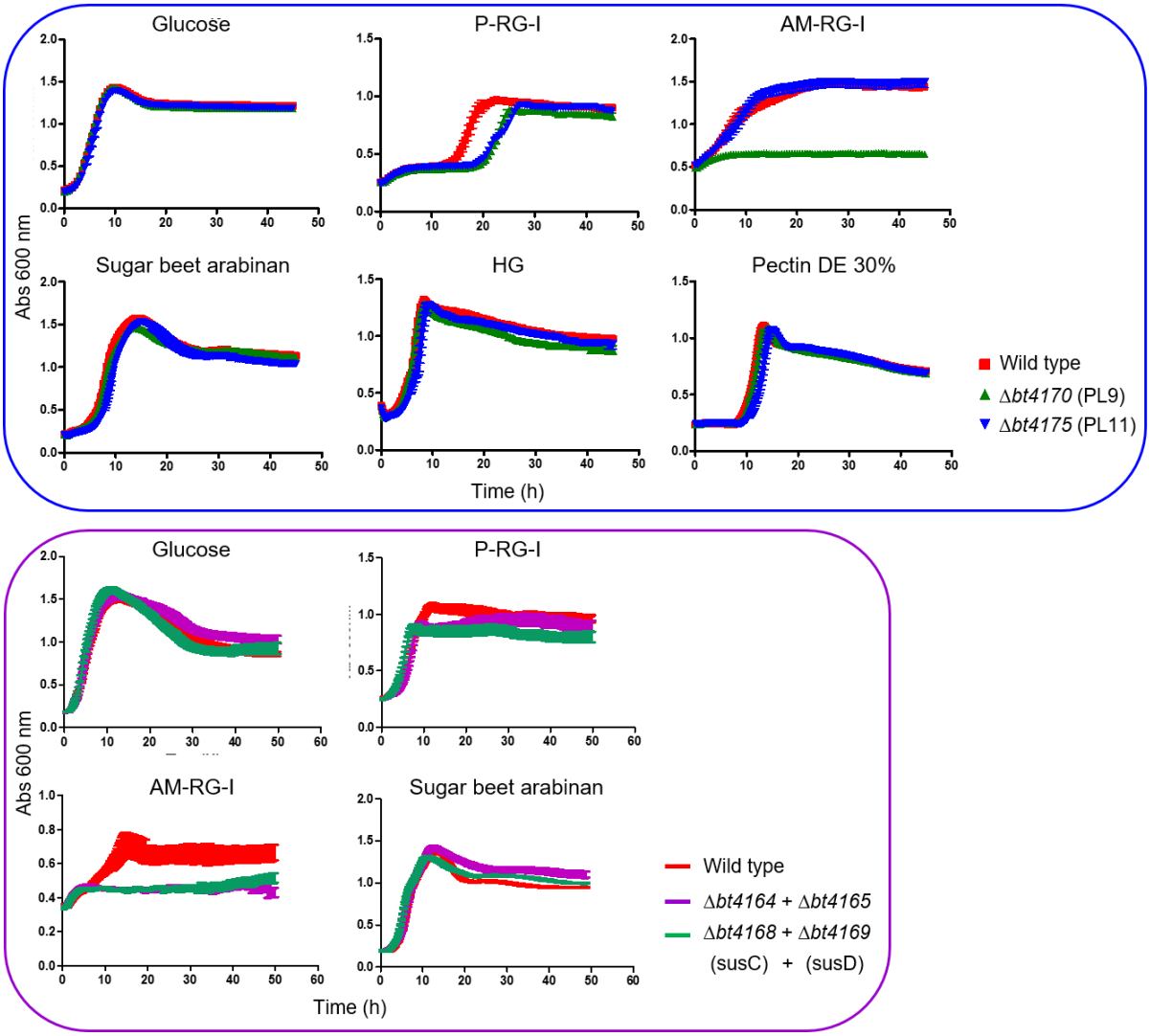


Figure 5.22 Growth curves of *B. thetaiotaomicron* KO mutants on different substrates

Each growth curve representation represents the average of four individual replicate cultures. Error bars represent the standard deviation of each averaged value between the biological replicates. For details of growth curves details see Chapter 2.1.14.

5.3.7 *Assays in different cell context*

Assays in different cell contexts were performed to explore enzymatic activity at the cell surface and in the periplasm. Briefly, *B. thetaiotaomicron* was anaerobically grown until mid-exponential on minimal medium (MM) supplemented with P-RG-I in order to up-regulate expression of the respective PUL. The bacterial cells were then incubated

aerobically with AM-RG-I or with phosphate buffered saline (PBS) (control reactions). This aerobic condition prevented glycan import into the periplasm, and thus allowed investigation of enzymatic activities at the cell outer membrane (whole cells). The sonicated cells gave the activity of all enzymes present on the cell surface, periplasm and cytoplasm. Additionally, the culture supernatant was also tested for enzymatic activity. In order to investigate the functional role of BT4170 in RG-I utilization, the recombinant mutant strain lacking the encoding gene (*Δbt4170*) was also grown and assayed in the same conditions as the wild-type strain (*btWT*).

5.3.7.1 BT4170 is essential to produce AM-RG-I oligosaccharides at cell surface

In AM-RG-I reactions the products generated were analyzed by TLC and HPAEC with electrochemical (PAD) and UV detector (Figure 5.23). The lyase β-elimination generates a double bound between C-4 and C-5 of D-GalA, resulting in an increase of absorbance at the ultraviolet range that allows the detection of unsaturated oligosaccharides by HPAEC (with UV detector).

In *btWT* and *Δbt4170* sonicated cells incubation with AM-RG-I was observed released of L-Rha and D-GalA, indicating the cleavage of this polysaccharide. When *btWT* whole cells and the culture supernatant were incubated with AM-RG-I oligosaccharides were detected using UV detector. This suggests that these products contain a double bond and result from PL-mediated β-elimination reactions. In contrast, no degradation of AM-RG-I at the cell surface or in the culture supernatant was observed when the *Δbt4170* mutant (grown on P-RG-I) was exposed to the AM-RG-I. This suggests that BT4170 is on the outer membrane and in the supernatant, and it is this enzyme that

cleaves the AM-RG-I backbone into oligosaccharides, which is required for import into the periplasm.

To understand the contribution of BT4170 at the cell surface and in the culture supernatant to AM-RG-I degradation time courses of glycan depolymerization were carried out. *btWT* whole cells and supernatant were incubated with AM-RG-I and at different times points the reactions were analysed by HPAEC, Figure 5.23C. In whole cell reactions the unsaturated oligosaccharides were produced after 30 min, while these products were only evident in the supernatant reactions after 16 h of incubation. This suggests that the PL activity at the cell surface makes a major contribution to polysaccharide degradation. However, the lyase present in the supernatant can represent an adaptive mechanism that allows *B. thetaiotaomicron* to access substrates available in the surrounding environment or allow other bacteria access to oligosaccharides released by *Bacteroides* enzymes.

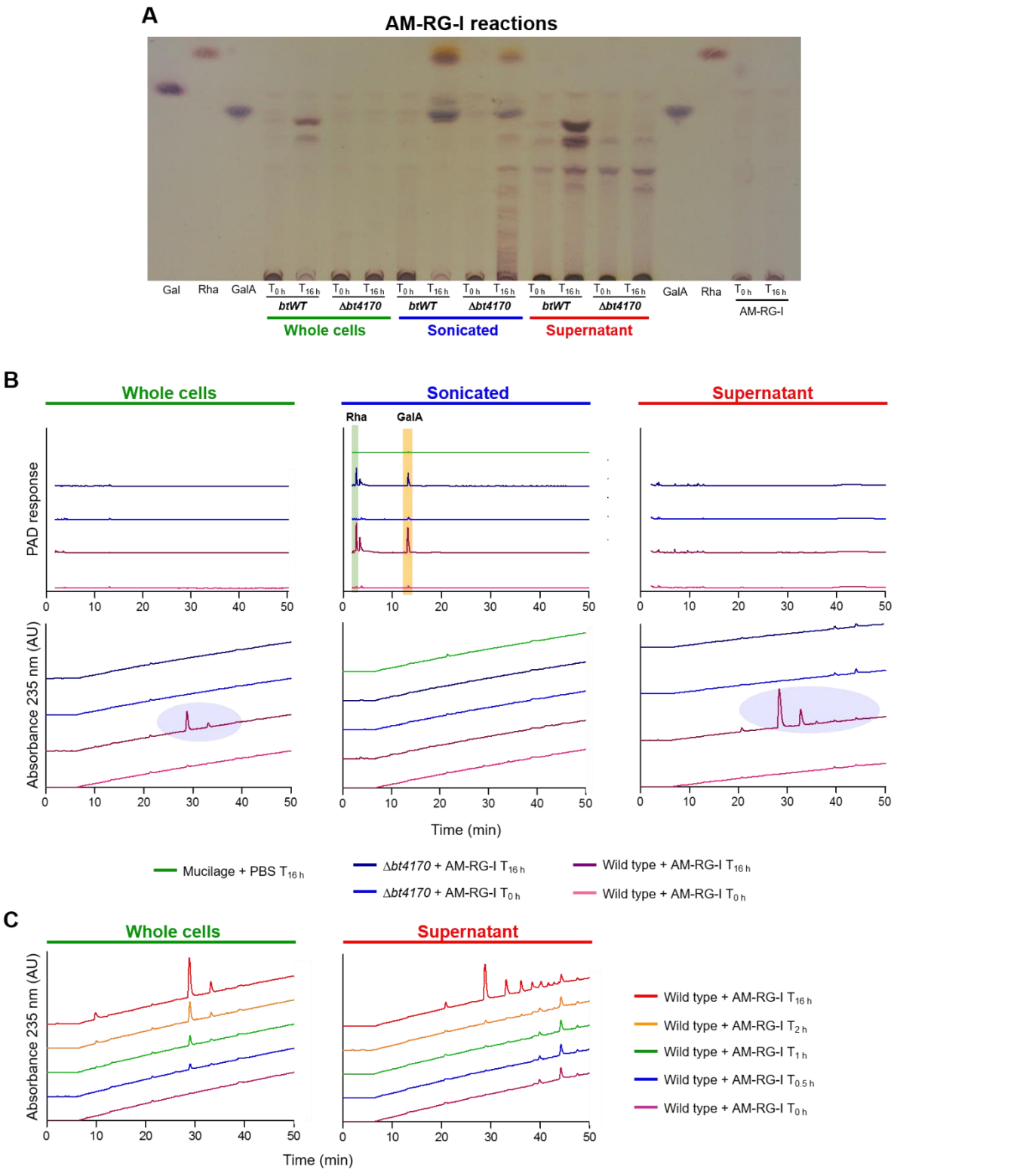


Figure 5.23 Assays in different cell context against AM-RG-I

The figure shows the reactions of the assays in different cell context analysed by TLC (Panel A) and HPAEC (Pane B). The *B. thetaiotaomicron* wild type (*btWT*) and the strain containing a genomic deletion (*Δbt4170*) were grown in MM with 0.5% P-RG-I until mid-exponential. The whole cells and culture supernatant were incubated with 0.5% of AM-RG-I at 37 °C for 16 h. Panel C presents the analysis by HPAEC of unsaturated oligosaccharides using *uv* detection. For details about assays, TLC and HPAEC see Chapter 2.2.11, Chapter 2.2.3 and Chapter 2.2.4, respectively. (Gal) D-Galactose; (Rha) L-Rhamnose; (GalA) D-Galacturonic acid.

5.3.7.1.1 Mass spectrometry of supernatant protein

In order to confirm the presence of BT4170 in the *B. thetaiotaomicron* supernatant samples, ~ 50 ml of culture supernatant media was concentrated 50 times and subjected to SDS-PAGE. The detected smear was cut and sent for analysis at Pinnacle (mass spectrometry facilities at ICaMB, Newcastle University). Briefly, the sample was trypsin digested and the peptides generated determined by mass spectrometry. The peptide finger print was used to search the *B. thetaiotaomicron* protein database. In total 18 hits were identified (Figure 5.24). One of the proteins identified by this search was BT4170, indicating that this protein is present in the culture supernatant.

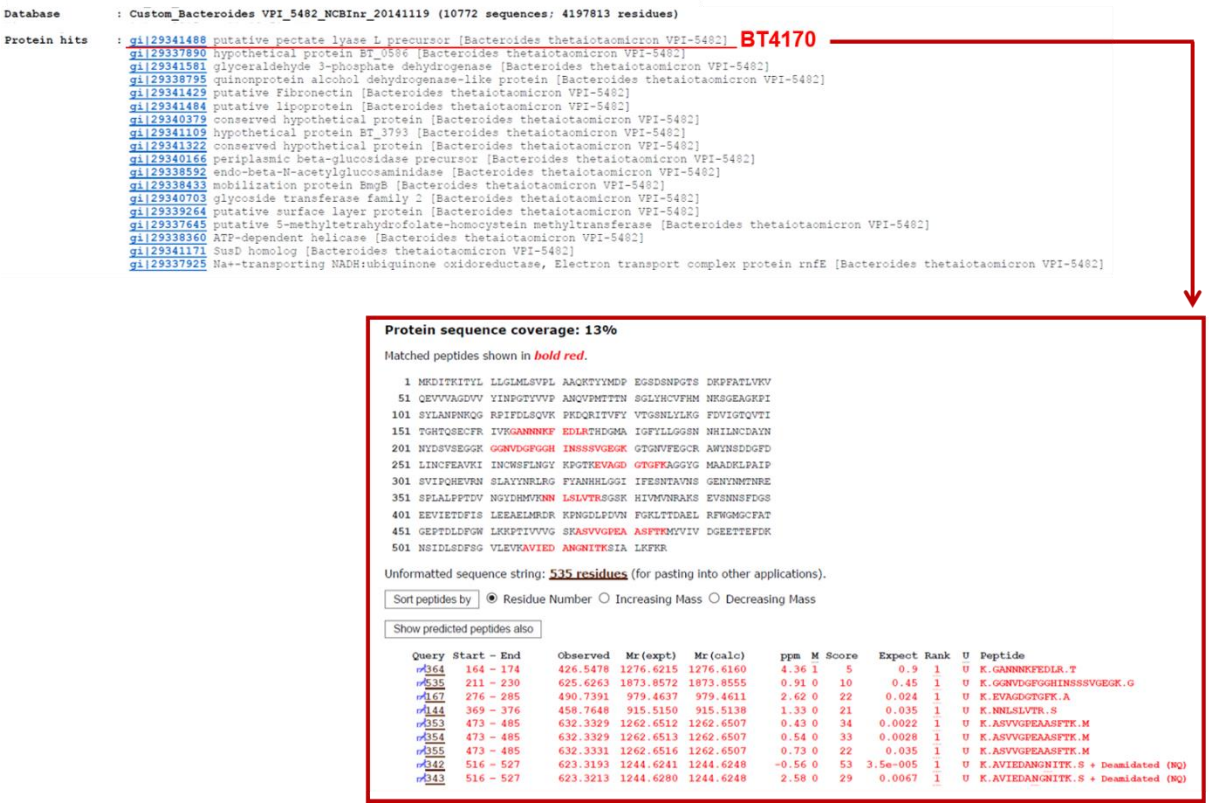


Figure 5.24 BT4170 identification by mass spectrometry

The figure shows the putative proteins identified in *B. thetaiotaomicron* culture supernatant. The box presents the peptides (in red) that match the BT4170 sequence. For details see Chapter 2.2.9.

5.3.8 *Cell localization of RG-I PUL encoded proteins*

To generate a complete model of RG-I utilization it is essential to understand where the enzymatic activities are located. The signal peptide (SP) can be used to predict the localization of specific proteins as those with a SP type I (SP I) should be in the periplasm and SP II is characteristic of lipoproteins (Zuckert 2014). The signal peptide for each protein studied in this chapter is shown in Table 5.2. However, data presented in the chapter indicate that BT4170, which contains an SP I, is at the cell surface. To clarify the cellular localization of BT4170, the protein was FLAG-tagged enabling detection using an anti-FLAG antibody. Additionally, the cellular localization PL11 BT4175 and the GH105 BT4176, which are predicted to be periplasmic, were also explored. Briefly, the primers utilized to generate the PCR fragments (Appendix K) allowed the introduction of a C-terminal FLAG-Tag sequence (DYKDDDDK) in the target gene (see Chapter 2.XXX for more details). The 1kb fragments upstream and downstream the C-terminal sequence of each gene were amplified by PCR and cloned into pExchange-*tdk*. The FLAG-tag strains were created using the allelic exchange method described in Chapter 2.1.23. An example of the PCR products visualized in agarose gel is present in Appendix K.

The FLAG-Tag strains were grown in P-RG-I until mid-exponential phase to induce expression of the RG-I PUL proteins. The bacteria were incubated with proteinase K (which degrades the surface proteins) or PBS at 37 ˚C for 16 h, and then subjected to Western blotting using anti-Flag-Tag antibody. The proteinase treatment prevents antibody detection of surface but not periplasmic proteins. The SusD-like protein BT4661 (encoded by the heparin PUL) was used as an internal control because it has been shown to be present on the surface of *B. thetaiotaomicron* and displays basal expression (personal communication, Dr Elisabeth Lowe). The data (Figure 5.25)

showed that the BT4175 (PL11) and GH BT4176 (GH105) were not degraded by proteinase K confirming the predicted periplasmic location. In these samples the internal control (BT4661) showed a decrease of signal when treated with proteinase K indicating that the proteinase successfully cleaved the surface proteins. In contrast, the FLAG-tagged BT4170 was degraded when cells were treated with proteinase K, indicating that this protein is present in the extracellular leaflet of the outer membrane. Additionally, this protein was also detected in the culture supernatant. However, the levels of the expression in supernatant were only detected in the late exponential phase of the growth (Figure 5.25B). These results are in accord with the data described above suggesting that BT4170 is present at the cell surface and in the culture supernatant.

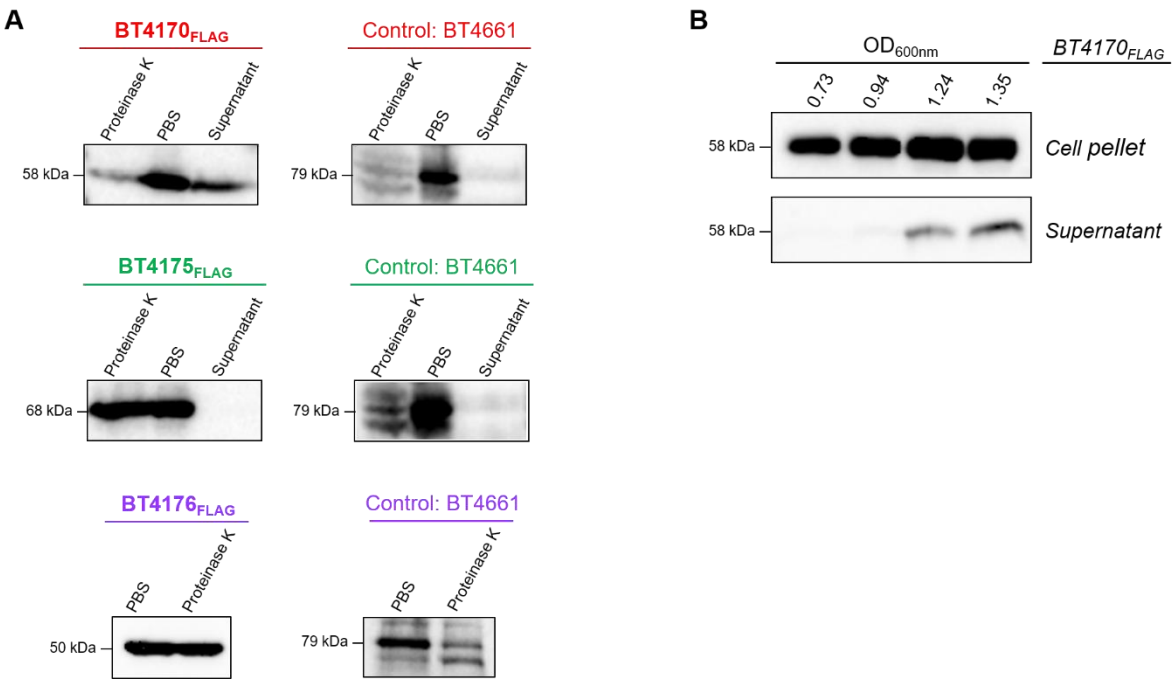


Figure 5.25 Detection of FLAG-Tag proteins by Western blot

A. Detection of FLAG-Tag proteins encoded by the RG-I PUL in the supernatant and in cells treated for 16 h with proteinase K or PBS. **B.** Expression levels of BT4170 in cells and in the media of cultures at different stages of growth monitored at OD_{600nm}. The proteins were detected using anti-FLAG or anti-BT4661 antibodies (control). For clarity only regions of the gel where the cognate protein migrates are shown. See Chapter 2.2.12 for details.

5.3.9 *The structure of BT4170*

The PLs BT4170 and BT4183 represent the first members of PL9 displaying rhamnogalacturonan lyase activity. To understand substrate recognition by these enzymes, BT4140 was crystallised in the apo form and in complex with the ligands L-Rha-α1,4-D-GalA-α1,2-L-Rha and Δ4,5-GalA-α1,2-L-Rha-α1,4-D-GalA and the structure was solved by Dr. Arnaud Basle (ICaMB, Newcastle University).

5.3.9.1 Protein crystallisation

BT1470 was crystallised by sitting-drop vapour-phase diffusion method using the automated Mosquito^{R} nanodrop dispensing system (see Chapter 2.3.2). An equal volume (100 nl) of protein and reservoir solutions were mixed and crystals were grown at 20 ºC. BT4170 apo crystals were obtained with 12.4 mg/ml in 0.02 M sodium phosphate, 20% (w/v) PEG 3350 after 16 h. The first BT4170 structure was obtained in complex with the reaction products from AM-RG-I. Briefly, after the purification the protein (9.7 mg/ml) was incubated with 0.5% AM-RG-I for 16 h at 37 °C. The reaction analysed by TLC (Figure 5.26A) showed that the substrate was cleaved to the limit-products of unsaturated di- and tetra-saccharide. The reaction was included in the crystallization trials of BT4170 as described above. Crystals were obtained after 13 days in 100 mM Bis-Tris pH6.5, 20% (w/v) PEG 3350. Figure 5.26B shows a picture on one crystal obtained by this approach.

The fishing of crystals, data collection and structure solution and refinement were performed by Dr. Arnaud Basle (ICaMB, Newcastle University). Briefly, crystals were frozen in 20% (w/v) PEG 400. Crystallographic data were collected at Diamond Light Source Ltd, UK, beamline I04. The apo structure was solved by molecular replacement

using the coordinates of Protein Data Bank (PDB) 1RU4 model. The apo and ligand complex of wild type BT4170 apo were solved at 1.48 Å and 1.07 Å, respectively. Crystal data collection statistics and refinement are reported in Appendix C (Table C.4).

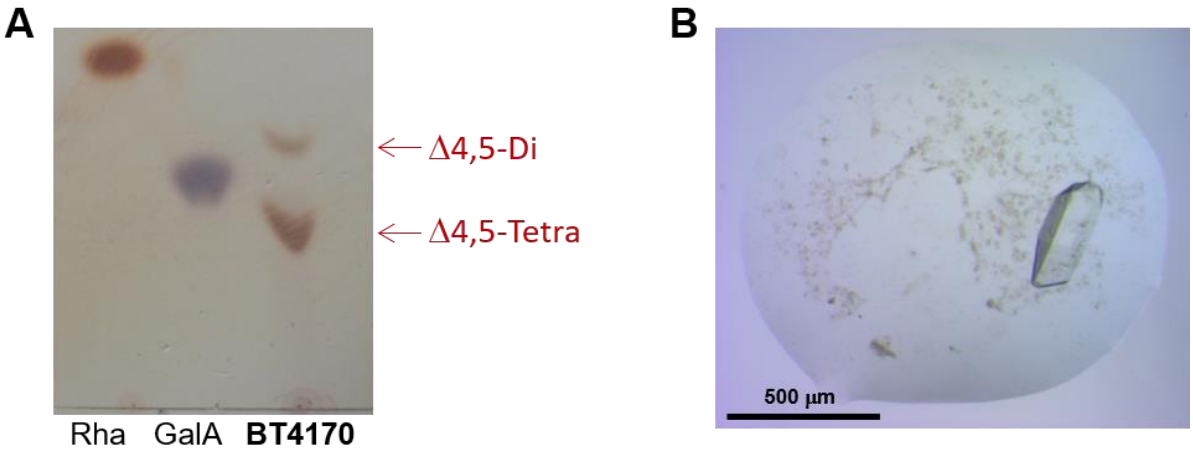


Figure 5.26 Analysis of BT4170 reaction product by TLC and protein crystal

A. Reaction products generated from AM-RG-I by BT4170 analysed by TLC (see Chapter2.2.3). The reaction was performed in 10 mM Tris-HCl, 75 mM NaCl, pH 8.0. For more details see text. **B.** Crystal obtained by co-crystallization of wild type BT4170 with the reaction products obtained in Panel A. The crystal grew in 0.2M ammonium sulphate, 0.1M HEPES pH7.5, 25% (w/v) PEG 3350.

5.3.9.2 BT4170 structure

BT4170 presented a typical PL9 right-handed parallel β-helix fold with three short parallel β-sheets (PB1, PB2 and PB3) and 10 turns (Figure 5.27). Additionally, the BT4170 β-helix is interrupted by two short antiparallel β-strands between β28 and β31 and three α-helixes between β1 and β2 (α1), β15 and β16 (α2) and, β16 and β17 (α3). The C-terminal region presents three α-helixes (α4, α5 and α6).

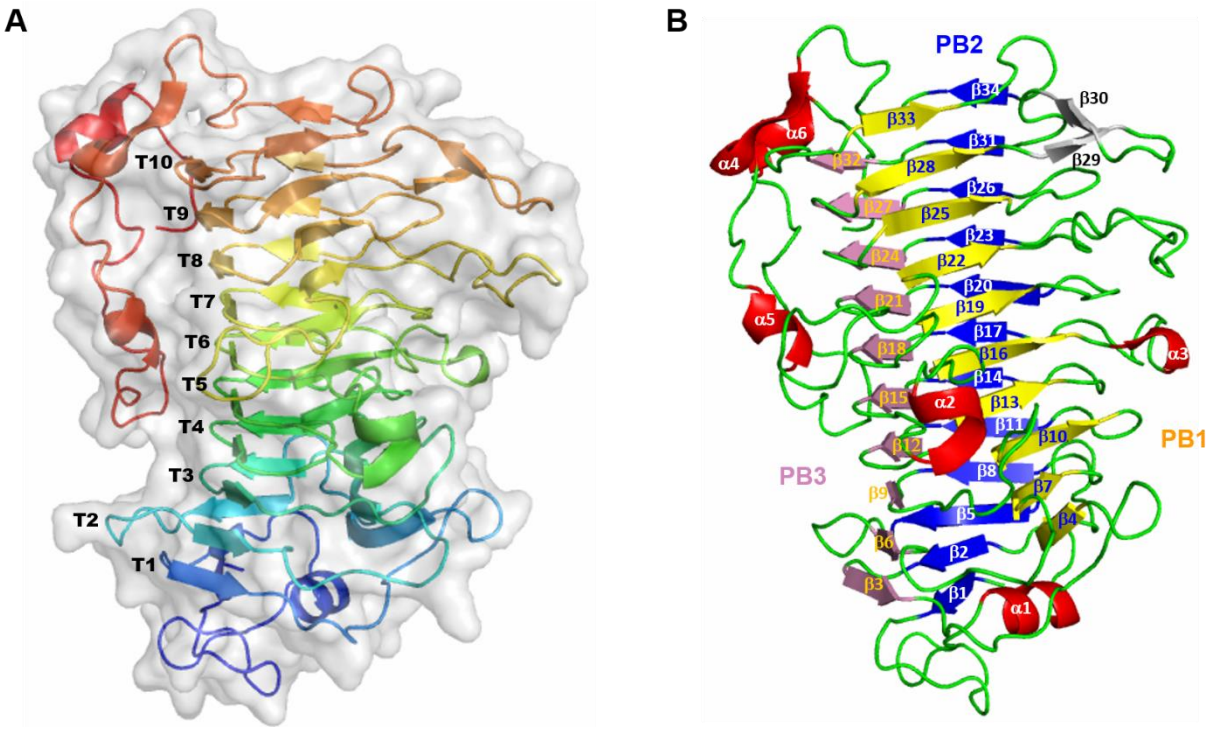


Figure 5.27 Schematic representation of BT4170 structure

A. Schematic representation of BT4170 parallel β-helix fold colour ramped from blue (N-terminal) to red (C-terminal), embedded in the surface representation of the protein. The turns (T) are numbered from the N (T1) to C-terminal (T10). **B.** Presents the three parallel β-sheet coloured in yellow (PB1), blue (PB2) and pink (PB3). The α-helix are shown in red and the additional β-strands in grey. The α-helix and β-strands are numbered from the N-terminal (α1 or β1) to the C-terminal (α6 or β34). Figure was created utilizing PyMOL

The analysis of structural alignment utilizing the program DaliLite version 3 (Holm and Rosenstrom 2010) showed that BT4170 closer structural homologue is Pel9A (PBD code 1RU4) from *Erwinia chrysanthemi* (Z-score of 43.3, root mean square deviation 2.2 Å over 353 matched Cα position showing 35% of percentage identity). Pel9A is a member of the PL9 family and it is the only structure available in the literature for this lyase family (Jenkins *et al*. 2004).

In order to explore the putative active site of BT4170 the structure was superposed with Pel9A (Figure 5.28A). Jenkins *et al.* proposed that the key components of the active site of Pel9A comprised a calcium binding site on coils 5 and 6, and a K273 on coil 7, which functions as the catalytic base abstracting a proton from C5. Both the

Pel9A calcium binding site (D215, D246, D247 and D250) and catalytic base (K285) are conserved in BT4170 (Figure 5.28B). In Pel9A, asparagine 268 can, in addition to calcium, interact with the carboxylate of GalA at the +1 subsite (Jenkins *et al*. 2004). However, in BT4170 this amino acid is substituted by an aspartate (D280), which is a component of a second calcium binding site described below.

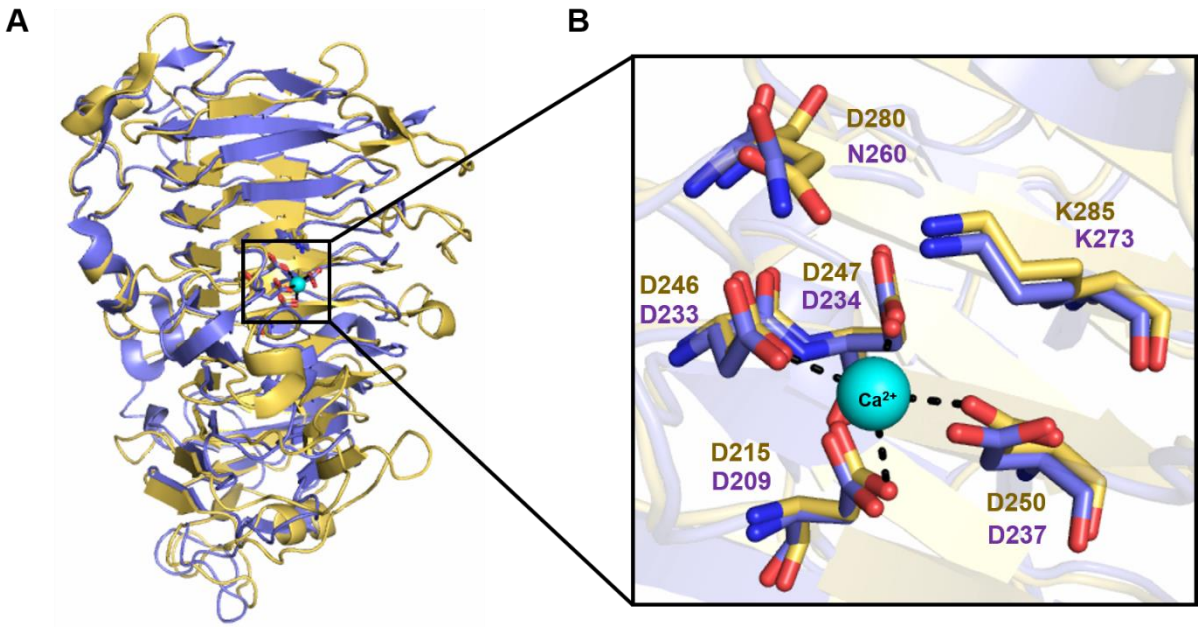


Figure 5.28 BT4170 structure superimpose to Pel9A

A. BT4170 (yellow) and Pel9A (PDB code 1Ru4) (purple) superimposed using Secondary Structure Matching tool in software WinCoot. The active site is represented as sticks and highlighted inside the black box. The calcium is represented as sphere (cyan). **B.** Blow up of the active site overlay. Figure was created utilizing PyMOL

5.3.9.3 Complex of BT4170 with ligand bound in the negative subsites

To disclose the different interactions established at the active site, BT4170 was co-crystallised in the presence of the products of the reaction with AM-RG-I (Δ4,5-D-GalA-α1,2-L-Rha and Δ4,5-D-GalA-α1,2-L-Rha-α1,4-D-GalA-α1,2-L-Rha). However, the structure only revealed the trisaccharide L-Rha-α1,4-D-GalA-α1,2-L-Rha bound to the enzyme in the cleft formed by the surface of the β-sheets and adjacent loops. The position of the proposed catalytic base (K285), that will interact with D-GalA at +1

subsite, suggests that the trisaccharide occupies the subsites -1 to -3 (Figure 5.29). A second calcium ion (Ca-2) coordinated with G212, D246 and D280 (see Section 5.3.9.4 for more details) and a phosphate ion were also detected in the active site. It is hypothesized that this additional calcium ion will be present in the active site stabilized by interactions D-GalA. Unfortunately, a phosphate, interacting with calcium and K285 was bound in the +1 active site preventing the binding of ligand to the positive subsites.

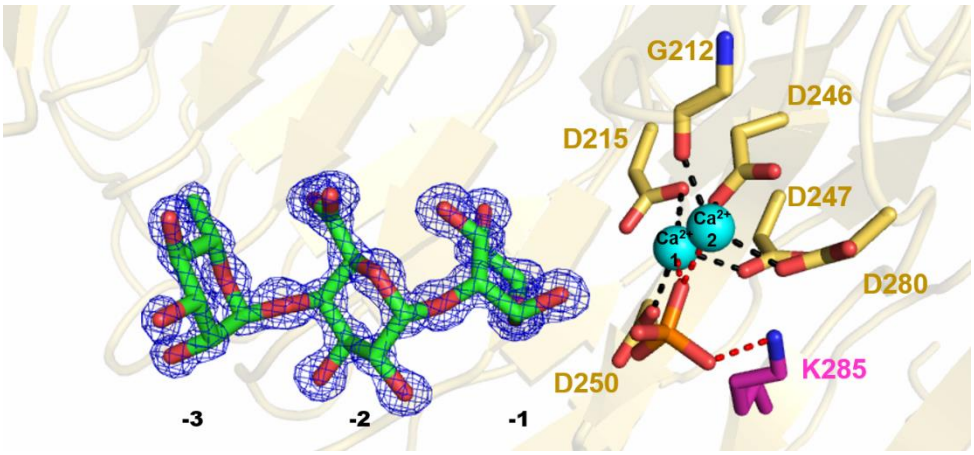


Figure 5.29 Structure of Bt4170 in complex with L-Rha-α1,4-D-GalA-α1,2-L-Rha
Schematic representation of the active site. Ligand is represented in green, calcium ions are shown as a cyan sphere and its polar contacts indicated by black dashed lines. The residues interacting with calcium and the proposed catalytic base are represented as stick in yellow and pink, respectively. The phosphate is shown orange (phosphorus) and red (oxygen). The interaction with phosphate are represented as red dashed lines. The blue mesh represents the 2Fo-Fc electron density map (1.1 Å resolution) at 2 σ. Figure created utilizing PyMOL.

5.3.9.4 Complex of BT4170 with ligand bound in the negative and positive subsites

To obtain a complex with ligand bound to the active site and the distal positive subsites the catalytic mutant K285A (the proposed catalytic base) of BT4170 was generated by site-direct mutagenesis. The primers used to generate the alanine mutant are shown in Appendix D, Table D.1 (for details see Chapter 2.1.11). To remove the phosphate occluding the active site from ligand in the wild type structures described above, after

the purification step the protein was incubated with EDTA to extract the calcium ion from the enzyme, which may then release the phosphate. After buffer exchange, the protein, loaded with calcium, was co-crystallized in the presence of the hexasaccharide (α1,2-L-Rha-α1,4-L-GalA)$_3$ prepared previously by Xiaoyang Zhang (Zhang 2015). Crystals obtained after four months in 0.2 M sodium chloride, 20% (w/v) PEG 3350 were used to solve the structure of the enzyme by molecular replacement using apo BT4170 as the search model to a resolution of 2.2 Å. Crystal data collection statistics and refinement are reported in Appendix C (Table C.4). Electron density for two oligosaccharides were present in the substrate binding site (Figure 5.30A). The density in the negative subsites (-3 to -1) was built as α-L-Rha-α1,4-D-GalA-α1,2-L-Rha and in the positive subsites (+1 to +3) as Δ4,5-GalA-α1,2-L-Rha-α1,4-D-GalA (Figure 5.30B). The ligand occupying -2 and -1 overlapped perfectly with the trisaccharide bound to the wild enzyme lyase. However the position of the L-Rha at -3, which does not interact with the protein (see Section 5.3.9.5.2), was different in the two complexes (Figure 5.30C). The rationale for building Δ4,5-GalA at the +1 subsite is as follows: There was no density for a glycosidic bond between the L-Rha at -1 and GalA at +1. Although the conformation of the sugars adopted a classical $^1C_4$ for L-Rha at -3, -1, +2 and $^4C_1$ for D-GalA at -2 and +3, the density at +1 only allowed a sugar to be built in which C3, C4 and C5 were in the same plane with no hydroxyl at C4. This is entirely consistent with Δ4,5-GalA (Figure 5.30D). A D-GalA in its relaxed $^4C_1$ conformation could not satisfactorily be fitted into the density in the +1 subsite. The origin of the two trisaccharides in BT4170 co-crystallized with an apparent hexasaccharide is discussed in Section 5.4.

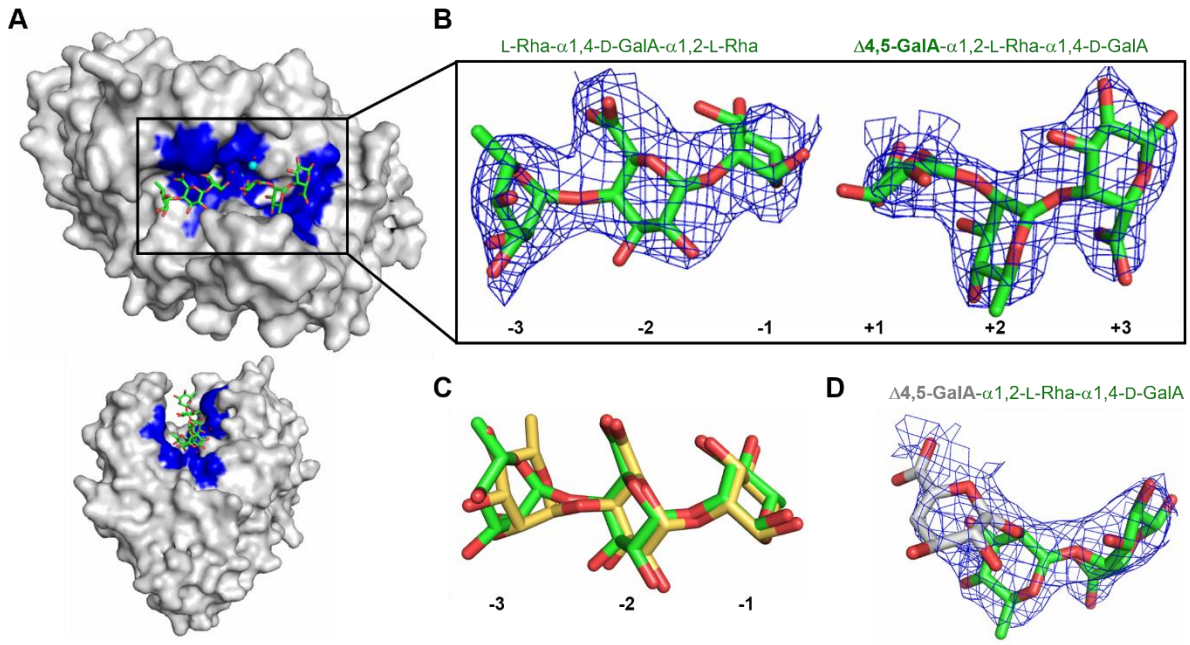


Figure 5.30 BT4170 structure in complex with two trisaccharides

A. Surface representation of BT4170 showing the active site interaction in dark blue and the ligand as sticks in green. The ligand is highlighted by the black box and shown in **B**. The subsite occupied by each sugar is indicated. **C.** Overlay of ligand in negative subsites. The ligand present in the structure in complex with the trisaccharide (see Figure 5.29) is shown in yellow and the ligand in structure with two trisaccharides is presented in green. **D.** Ligand detected in positive subsites build as Δ4,5-GalA-α1,2-L-Rha-α1,4-D-GalA. The blue mesh represents the 2Fo-Fc electron density map (2.2 Å resolution) at 1.5 σ. Panels were created utilizing PyMOL.

5.3.9.5 Substrate recognition

In this subsection the interactions BT4170-ligand will be described. To explore the role of BT4170 residues identified as implicated in substrate recognition the amino acids were mutated and the enzymatic activity against AM-RG-I was determined. Briefly, by site direct mutagenesis, the tested amino acids were mutated to alanine (see Chapter 2.2.11). The primers utilized to generate these mutants are shown in Appendix D (Table D.1). The k_{cat}/K_M for BT4170 and different mutants was determined by measuring the product formation at 235 nm (see Chapter 2.2.5.1).

5.3.9.5.1 Substrate binding in the distal subsites

Substrate recognition at the negative subsites was mediated by hydrogen bond between Q155 and O3 and O4 of L-Rha (-1 subsite) and R125 and H153 with the carboxyl group of the D-GalA at (-2 subsite) (Figure 5.31A and B). Three residues (Y184, N213 and H220) established hydrophobic interactions with the trisaccharide. The L-Rha present at the non-reducing end did not interact with the protein suggesting it does not contribute to substrate recognition and this protein only presents two negative subsites. Mutation to alanine of the residues that interacted with ligand in subsites -1 to -3 showed that R125A and H220A were completely inactive (Table 5.7). The mutation of the additional residues resulted in reduction of less than 20-fold in enzymatic activity (Table 5.7). This indicates that the substrate recognition in the negative subsites is mediated by the polar interactions between R125 and the D-GalA carboxyl group at -2 and the hydrophobic contact between the histidine imidazole ring and C6 of L-Rha at -1.

Substrate recognition at the distal positive subsites was through hydrogen bonds between N324 and O4 of L-Rha (+2 subsite), R319 and O5 of D-GalA and Y322 and H325 with O6 of D-GalA in the +3 subsite (Figure 5.31A and B). The sugars in these subsites also made apolar contacts with M291, H325 and L318. Mutation of these residues resulted in ~10-fold reductions in activity suggesting that they make a minor contribution to substrate binding (Table 5.7).

5.3.9.5.2 Substrate binding in the active site

A central feature of substrate recognition in the active site (+1 subsite) are two calcium ions, defined as Ca-1 and Ca-2, that make salt bridges with Oδ2 of the carboxylate of the bound Δ4,5-GalA (Figure 5.31A and C). Only the Ca-1 binding site is conserved in PL9 lyases. The Ca-1 binding site in BT4170 comprises four aspartate residues (D215, D246, D247 and D250) with the carboxylate side chains interacting with the metal ion (see Figure 5.28). The Ca-2 binding site comprises G212, D250 and D280. Again the metal makes polar contacts with the carboxylate with the two aspartates and the carbonyl of the glycine. I252 makes apolar contacts with Δ4,5-GalA (Figure 5.31A and B). The I252A substitution decreased activity by ~500-fold (Table 5.7) suggesting that the isoleucine contributes to substrate binding to the active site. The aliphatic residue makes apolar contacts with both L-Rha at +2 and D-GalA in the +1 active site, and, intuitively, interaction with the uronic acid would be expected to be more catalytically relevant. Substitution of any of the carboxylates that form the Ca-1 binding site completely inactivates BT4170 (Table 5.7), indicating that this calcium plays an essential role in substrate binding and/or catalysis. The G212A mutation had very little impact on catalytic activity (Table 5.7) suggesting that it makes a minor contribution to Ca-2 binding. The 100-fold reduction in activity displayed by the D280 mutant (Table 5.7) suggests that this residue does contribute to Ca-binding and that this second calcium ion contributes to substrate binding or catalysis.

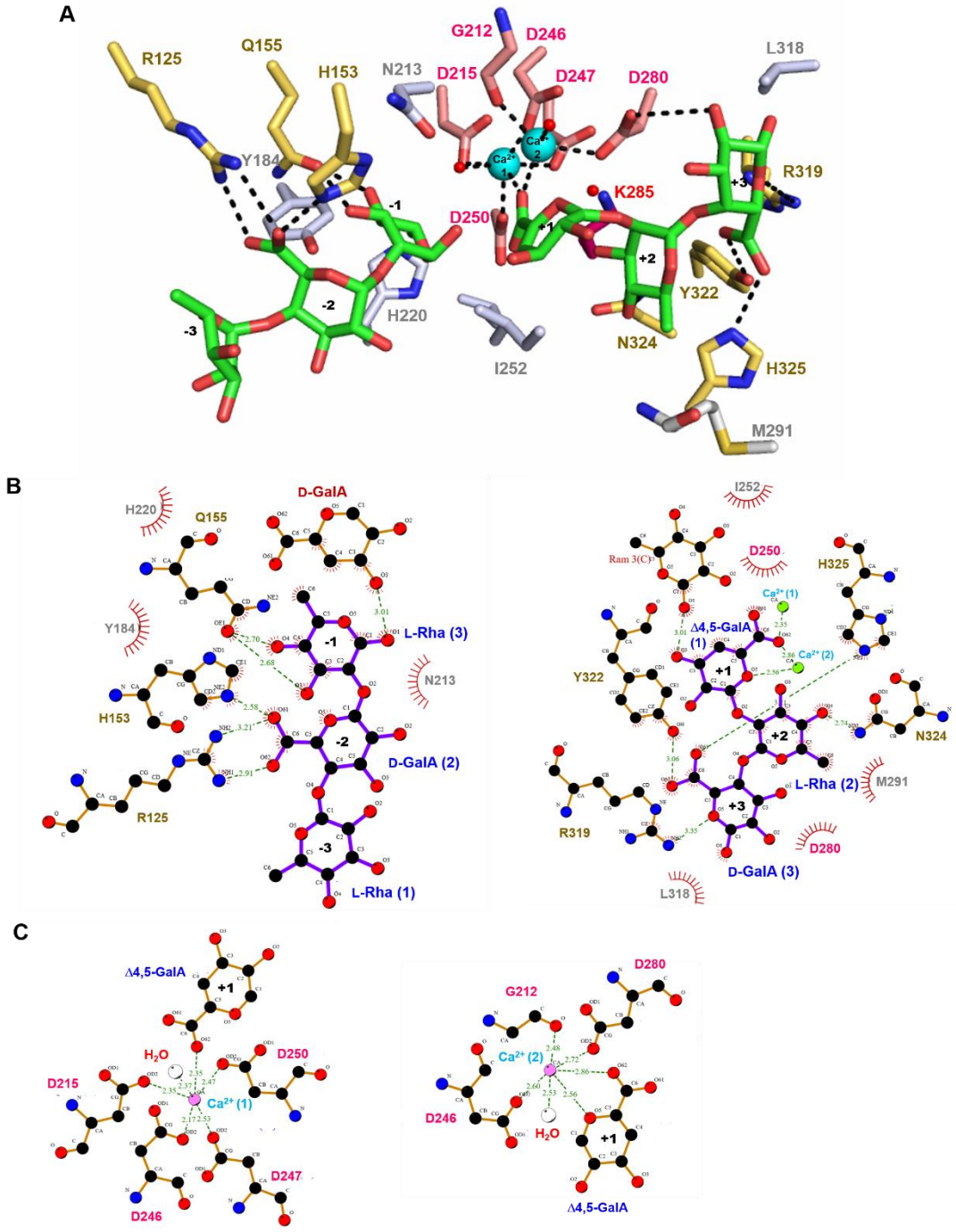


Figure 5.31 BT4170 active site

A. BT4170 active site interactions with ligand in green. The residues implicated in hydrogen bond and hydrophobic interactions are shown in yellow and grey, respectively. The calcium binding residues are present in light pink. The catalytic base K285 (dark pink) is present in the apo structure superimposed to the structure in complex. The calcium ions are shown as a cyan sphere. The polar contacts are indicated by black dashed lines. **B.** Schematic representation of the interactions with the ligands α-L-Rha-α1,4-D-GalA-α1,2 and Δ4,5-GalA-α1,2-L-Rha-α1,4-D-GalA **C.** Schematic representation of the interactions involving the calcium ions 1 and 2. The green dashes represent hydrogen bonds and the red semi-circles indicate hydrophobic interactions. For list of interactions see Appendix G, Figures G.3 to G.5. Panels A was created utilizing PyMOL. The diagrams present in B and C were generated using the online software PDBsum (de Beer *et al*. 2014).

Table 5.7 Catalytic activity of BT4170 mutants

Recombinant protein		k_{cat}/K_M (min^{-1} M^{-1})
Wild type	BT4170	$3.63 \times 10^4 \pm 521$
Mutant	**R125A**	No activity
	H153A	$7.26 \times 10^3 \pm 234$
	Q155A	$1.93 \times 10^4 \pm 434$
	Y184A	$8.72 \times 10^3 \pm 130$
	G212A	$1.77 \times 10^4 \pm 261$
	N213A	$1.66 \times 10^3 \pm 60$
	D215A	No activity
	H220A	No activity
	D246A	No activity
	D247A	No activity
	D250A	No activity
	I252A	70.5 ± 2.8
	N253A	$1.85 \times 10^4 \pm 451$
	D280A	376.0 ± 26.0
	K285A	No activity
	L318A	$1.92 \times 10^4 \pm 226$
	R319A	$1.92 \times 10^3 \pm 130$
	Y322A	$3.46 \times 10^3 \pm 86$
	N324A	$3.68 \times 10^3 \pm 162$
	H325A	$1.22 \times 10^3 \pm 26$

Catalytic activity at 235 nm (see chapter 2.2.5.1) of BT4170 and different mutants against AM-RG-I in 100 mM CAPSO pH 9.0 and 2mM $CaCl_2$. k_{cat}/K_M calculated using a single substrate concentration below K_M as described in Chapter 2.2.5. The standard errors were generated from technical triplicates. The amino acids are coloured in yellow and grey for residues implicated in hydrogen bond and hydrophobic interactions, respectively. The calcium binding residues are shown in light pink and the catalytic base is represented in dark pink.

5.3.9.6 Potential mechanism of BT4170

All PLs including those in the PL9 family are predicted to cleave its target glycosidic bond through a β-elimination reaction. A catalytic base extracts a proton from C5 with stabilization of the resulting anion by charge delocalization at the C-6 carbonyl group

through polar interactions with calcium. This leads to lytic cleavage of the O-4:C-4 bonding that is facilitated by proton donation from a catalytic acid to the glycosidic oxygen (O-4). In BT4170 the only candidate general base that can extract the H5 proton is K285. Lysine residues fulfill this function in many pectate lyases, reflecting the alkaline optimum of these enzymes, pH 9.0 in the case of BT4170. K285, however, is 4.4 Å from the C5 of Δ4,5-GalA making it difficult to visualize direct proton abstraction, which is thus likely to occur through a water molecule. In any event the complete loss of activity by alanine substitution of K285, Table 5.8, supports its proposed role as the catalytic base. Additionally, the equivalent residue Pel9A (K273) was previously proposed as the catalytic base in PL9 family (Jenkins *et al*. 2004). The resulting anion is likely stabilized by Ca-1 although Ca-2 may also contribute. The identity of the catalytic acid that protonates O-4, which is required for leaving group departure is unclear.

5.3.9.7 Phylogeny of BT4170

In order to explore the conservation of the BT4170 active site residues (Figure 5.31) the protein sequence was aligned to different homologues. BT4183, the second PL9 present in the RG-I PUL, and Pel9A, the only PL9 enzyme structurally characterized, were also included in the alignment. Additionally, these enzymes display different substrate specificities. BT4170 and BT4183 are active on AM-RG-I and Pel9A cleaves HG (Jenkins *et al*. 2004). Briefly, sequences of these three proteins were individually subjected to a BLAST search against a non-redundant protein database (Dereeper *et al*. 2010). The 100 sequences of each BLAST search were aligned in MUSCLE (Edgar 2004a; Edgar 2004b) and analyzed. In order to simplify, only a representative number of sequences showing the lower percentage of similarity to the protein blasted were

selected: ten in BT4170 (41 to 43% of identity), five in BT4183 (43 to 53% of identity) and nine sequences in Pel9A (36 to 40% of identity). These selected sequences were aligned in MUSCLE (Edgar 2004a; Edgar 2004b). The alignment file was visualized using the online tool ESPrit 3.0 (Robert and Gouet 2014) and is present at Figure 5.32. The predicted catalytic base, K285, and the residues implicated in Ca^{2+} 1 binding site (D215, D246, D247 and D250) were invariant in these sequences. However, from the residues predicted to interact with substrate in the distal subsites of BT4170 (see Figure 5.31) only N324 was invariant. This residue is implicated in a hydrogen bond with O4 of L-Rha (+2). However, in Pel9A this subsite will be occupied by D-GalA, suggesting that N324 is also able to establish an interaction with this sugar. The residues implicated in second calcium binding site (G212 and D280) were conserved but not invariant, suggesting that in some proteins the second calcium can be interacting with different residues or may not exist at all.

Additionally, the protein sequences of homologues of BT4170 and BT4183, both acting on AM-RG-I, shared a high conservation substrate binding residues. R125 (-2 subsite), Q155, N213 and H220 (-1 subsite), I252 (+1 subsite), Y322 and H325 (+3 subsite) are conserved in 200 homologues of these RG-I proteins. This suggests, that the residues implicated in AM-RG-I binding are conserved between rhamnogalacturonan lyases, in PL9 family. However, substrate recognition of PL9 pectate lyases, such as Pel9A, is mediated by more charged residues (Figure 5.33), suggesting that pectate lyases are adapted to accommodate the high negatively charged substrate (non-methylated D-GalA). In BT4170, a more hydrophobic active site allows the interaction with a less charged substrate (AM-RG-I).

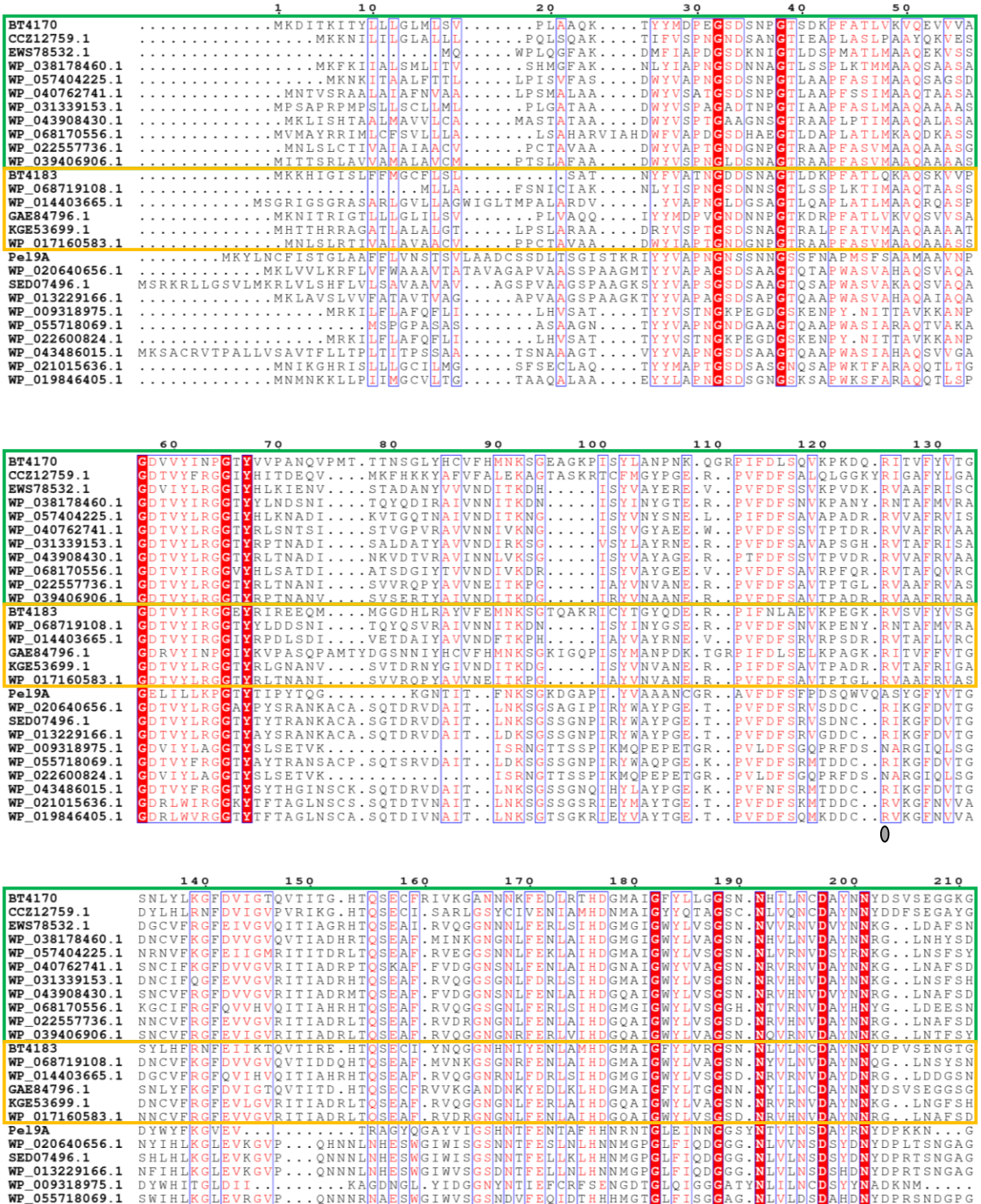


Figure 5.32 PL9 proteins alignment

NCBI reference codes for each protein are shown at the left. BT4170 and BT4183 represents the PL9 enzymes from *B. thetaiotaomicron* PL9 and Pel9A is the only PL9 structurally characterized in the literature (Jenkins *et al*. 2004). Protein sequences were aligned in MUSCLE (Edgar 2004a; Edgar 2004b) and visualized using ESPrit 3.0 (Robert and Gouet 2014). Sequences inside the green or orange box are homologues from BT4170 or BT4183, respectively. Sequences outside the boxes are homologues of Pel9A. In the shown sequences, amino acids with 100% of conservation are highlighted in a red background and residues that are functionally highly conserved are boxed and coloured red. The residues are numbered according to the BT4170 sequence.

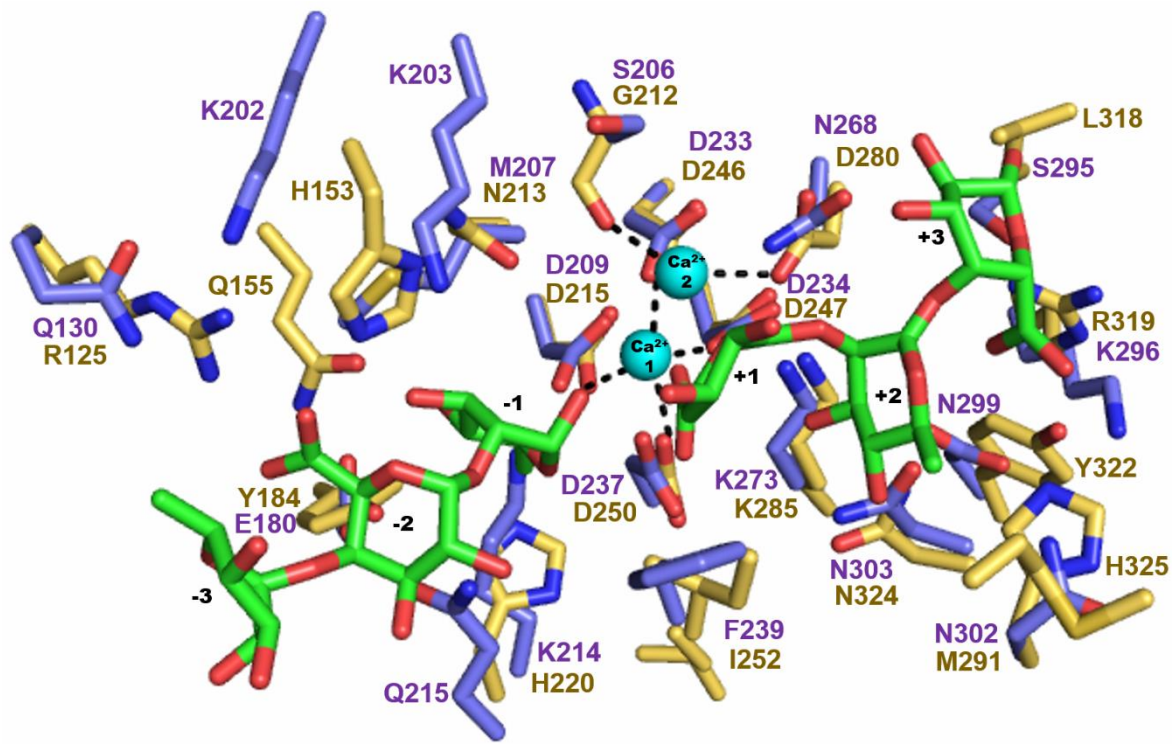


Figure 5.33 BT4170 and Pel9A active site and protein surface

A. Superimposed active sites of BT4170 (yellow) and Pel9A (PDB code 1RU4) (purple). The ligand is and calcium ions are shown in green and cyan (sphere), respectively. Figure created utilizing PyMOL

5.4 Discussion

In this chapter several enzymes encoded by RG-I PUL were characterized and the biological function in *B. thetaiotaomicron* was explored in order to propose a model for RG-I utilization by this bacterium.

RG-I PUL encodes three PLs: BT4170, BT4175 and BT4183. The expansion of lyases can be explained by differences in enzymes specificities and cell localization. The N-terminal signal peptide of a protein defines its cell localization. A protein containing a SP I, which is characterized by a hydrophobic region followed by a cleavage site, will be located in the cell periplasm. The SP II proteins contain a typical lipobox motif followed with a cysteine after the cleavage point and are predicted to be attached to the cell membrane (Juncker *et al*. 2003; Auclair *et al*. 2012; Zuckert 2014). All the PLs were initially predicted as periplasmic enzymes (SP I signal peptide). However, cell

localization studies showed that BT4170 is present at the cell surface. It is known that positively charged proteins are attracted to a negatively charged membranes by non-specific electrostatic interactions (Whited and Johs 2015). However, the predicted low isoelectric point (pI) for BT4170 (5.33) excludes any possible electrostatic interactions with the negatively charged membrane (at pH ≥ 5.33). So, the mechanism of protein interaction with the outer membrane remains unknown. Additionally, BT4170 was detected in the culture supernatant. This can result of cell lysis explaining why the lyase accumulates in the supernatant in the late phase of growth. In this stage bacteria have less nutrient sources and some lysis can occur. In this case, the periplasmic enzymes would be present in this sample and it was expected that the incubation of the supernatant with AM-RG-I would generate L-Rha and D-GalA. However, the only products detected in this reaction were unsaturated oligosaccharides that result from the activity of PL9 BT4170 suggesting that the lysis hypothesis is incorrect. It is also possible that this protein is weakly attached to the outer membrane surface to the hydrophobic interactions. In this case, an alteration in pH could result in the release of the protein. Additionally, this can be due the presence of outer membrane vesicles (OMVs) that are released into culture supernatant. Indeed, it was previously reported that when *Bacteroides* was grown in glucose it is able to secret OMVs containing several hydrolases and proteases (Elhenawy *et al*. 2014). A common feature of these OMVs is the high content of acidic proteins, suggesting that the packing of the enzymes is a pI-dependent process (Elhenawy *et al*. 2014). BT4170 is an acidic protein so it is possible to speculate that this protein can be packed into OMVs. The role of these OMVs is not clear, however, they may allow a sharing mechanism where *B. thetaiotaomicron* provides the required enzymes to release oligosaccharides that can be utilized by other bacteria present in the gut.

The PL9 BT4170 was identified in this study as the key enzyme in RG-I utilization system. The major role of this lyase was in cleaving the RG-I backbone and generating oligosaccharides that will be imported into the periplasm by the SusC/SusD-like proteins. When this gene was deleted *B. thetaiotaomicron* lost the ability to grow on AM-RG-I. Similar results had been previously reported in xylan and xyloglucan utilization by *B. ovatus*, showing that the deletion of the enzyme responsible for generating the oligosaccharides at the cell surface results in loss of ability to grow on the target substrate (Larsbrink *et al*. 2014; Rogowski *et al*. 2015). Additionally, BT4170 and BT4183 are the first rhamnogalacturonan lyase reported in PL9 family. The structural characterization of the BT4170 in complex with L-Rha-α1,4-D-GalA-α1,2-L-Rha and Δ4,5-GalA-α1,2-L-Rha-α1,4-D-GalA revealed important features. This crystal was obtained with the inactive mutant of BT4170 the detection of an unsaturated oligo could be explain by a residual enzymatic activity. However, this hypothesis is unlikely to occur in a PL enzyme where the mechanism only requires a catalytic base (usually a lysine or an arginine) to abstract the C5 proton (Garron and Cygler 2010) and, BT4170 does not present any additional candidates in the active site. A second explanation and, a more likely one, is the cross-contamination of the BT4170 K285A protein with the wild type that can cleave the hexasaccharide originating the two detected trisaccharides. The active site of PL9s enzymes was structurally conserved. A lysine likely acts as catalytic base and four aspartates interact with a calcium ion that assists in acidifying H5. A second calcium was detected between the substrate and the protein. This additional calcium interacted with the carboxyl group of D-GalA (+1 subsite), which may contribute to H5 acidification. The structure in complex allowed the identification of the three central subsites (-2, -1 and +1). However, when the structure of this rhamnogalacturonan lyase was superimposed with Pel9A structure, a pectate lyase of PL9 family, the BT4170 residues interacting with ligand were not

conserved in Pel9A. Indeed, this pectate lyase presents a substrate binding region that is more charged comparative to BT4170. These differences can be explained by difference in the charge of the target substrate. The non-methylated (HG) polysaccharide targeted by the pectate lyase, is a highly negatively charged substrate that will bind preferentially to a charged active site, such as, the one in Pel9A. While, enzymes that target AM-RG-I, such as BT4170, will present a more hydrophobic active site because the presence of a neutral sugar (L-Rha), with a deoxy C6 group, intercalated with D-GalA. A similar adaptation of the active sites was previously described in PL1, where pectate lyases present a positively charged substrate binding region to accommodate the non-methylated D-GalA substrate, while pectin lyases have a more hydrophobic active site because they recognize methyl esterified D-GalA, which carries no charge (not charged) (Mayans *et al*. 1997).

In the CAZy database there is only two other PL families, PL4 and PL11, that contain rhamnogalacturonan lyases (Cantarel *et al*. 2009; Lombard *et al*. 2014). In PL4 lyases (β-sandwich+β-sheet fold) a lysine and a histidine act as catalytic base and acid, respectively. Additionally, PL4 enzymes are the only enzymes active in pectins that do not require a divalent metal for activity. In these lyases the carboxyl group of D-GalA (+1 subsite) interacts with an aspartate and a water molecule facilitating the H5 proton abstraction (Jensen *et al*. 2010). In PL11 enzymes (β-propeller fold) an arginine is proposed to act as catalytic base. Similar to BT4170, in PL11 lyases a calcium ion can contribute to H5 acidification (Ochiai *et al*. 2007). These differences in the catalytic apparatus compared to what has previously described in BT4170 suggest there is no conservation in rhamnogalacturonan lyases active site.

The additional two PLs encoded by RG-I PUL, BT4175 and BT4183, also have an SP I signal peptide indicating that they will be located in the periplasm. The PL11 lyase,

BT4175, was more active than BT4183 (PL9) against all the substrates tested. It was mentioned before that the degree of purity of BT4175 was low. This had an impact on the rates since the concentration of the enzyme was overestimated. So, it is expected that the activity of BT4175 is higher than measured here. Activity against branched polysaccharides is also different in the two PLs. The PL11 lyase BT4175 was the only enzyme able to efficiently cleave P-RG-I suggesting that this enzyme can tolerate the galactan side chains, consistent with previous reports (McKie *et al*. 2001; Pagès *et al*. 2003; Ochiai *et al*. 2007). BT4183 (PL9) does not tolerate the galactan side chains, but its limited activity against the conformationally flexible sugar beet arabinan suggests that this enzyme can tolerate arabinan side chains. The limit products generated by these lyases are also different. Against AM-RG-I, BT4175 generated three main unsaturated oligosaccharides (di, tetra and hexa), while BT4183 produced longer unsaturated oligosaccharides (≥ that tetrasaccharide). This indicates that these two enzymes display different substrates specificities. Furthermore, the different pH optima for these two PLs indicate that the enzymes could have evolved to act in different cellular microenvironments.

Backbone cleavage of RG-I requires the action of an α-rhamnosidase, BT4145, which is a member of the poorly characterized GH106 family. The enzyme cleaves the L-Rha form the non-reducing end of oligosaccharides derived from RG-I. BT4145 displayed an ion-dependent activity and the mechanism of catalysis determined by ^{1}H-NMR showed that this enzyme hydrolyses glycosidic linkages through a single displacement mechanism leading to inversion of anomeric configuration. In such enzymes the general base and acid are predicted to be separated by 10 Å. The structural characterization of BT0986, a BT4145 homologue involved in RG-II degradation, allowed the identification of the general acid/base pair (E461/E593) that are fully conserved in BT4145 and other members of GH106. In the structure of BT0986 (see

Chapter 4.3.8.3) the general acid (E461) is 8 Å from the predicted general base (E593) (see Figure 4.19). This structural data agrees with the previous proposed inverting mechanism for GH106 members. Inactivation of BT4145 by EDTA and the functional significance of residues equivalent to those that comprise the calcium binding site in BT0986, suggest that the divalent metal ion plays an important role in the catalytic mechanism of both GH106 enzymes. The importance of calcium was reported previously in α-mannosidases and it was proposed that, by interacting with O2 and O3 of the bound mannose the metal ion stabilizes the transition-sate and the departure of the leaving group (Zhu *et al*. 2010; Speciale *et al*. 2014).

The RG-I PUL encodes four GH28 exo-acting galacturonidases that release D-GalA from the non-reducing end of the RG-I backbone or HG. The expansion of proteins belonging to the same family is explained by differences in specificity. BT4153 was active on AM-RG-I oligosaccharides but not HG and is thus specific for RG-I. The activity against oligosaccharides suggest that BT4153 contains only a -1 subsite (binds D-GalA) and a +1 subsite that interacts with L-Rha. The activity determined for the closest homologue of BT4149 suggests that this enzyme is also a RG-I specific galacturonidase, however, the enzyme has a strong preference for the tetrasaccharide and the activity did not increase with the hexasaccharide, indicating that this GH28 enzyme requires L-Rha bound at the +3 subsite for maximum activity. The GH28 enzyme BT4155 was shown to be an exo-acting galacturonase that cleaved α1,4 linkages between D-GalA units in HG. The catalytic activity of the enzyme did not increase with the size of the target oligosaccharide indicating that this enzyme only contains two subsites. BT4155, however, also cleaved the RG-I disaccharide D-GalA-α1,2-L-Rha. No activity was found for AM-RG-I oligosaccharides with a DP >2. This indicates that BT4155 can bind to L-Rha at +1, but cannot then tolerate D-GalA and L-Rha at the +2 and +3 subsites, respectively. The structures of the oligosaccharides of

HG and AM-RG-I are presented in Figure 5.34. The α1,4-D-GalA oligosaccharide present a linear structure, whereas the D-α1,4-GalA-α1,2-L-Rha oligosaccharide is more helical. These differences in substrates conformation can be one explanation why BT4155 is only able to tolerate a α1,2-L-Rha at +1. It is possible that longer oligosaccharides of GalA-Rha clash with the enzyme surface. However, solving the BT4155 in complex with both disaccharides is essential to disclose the mechanism of ligand recognition that drive the specificity of these two enzymes. BT4146 was specific for the disaccharide D-GalA-α1,2-L-Rha. However, it is possible that this enzyme is also be able to target different oligosaccharides that where not tested in this thesis work (see below).

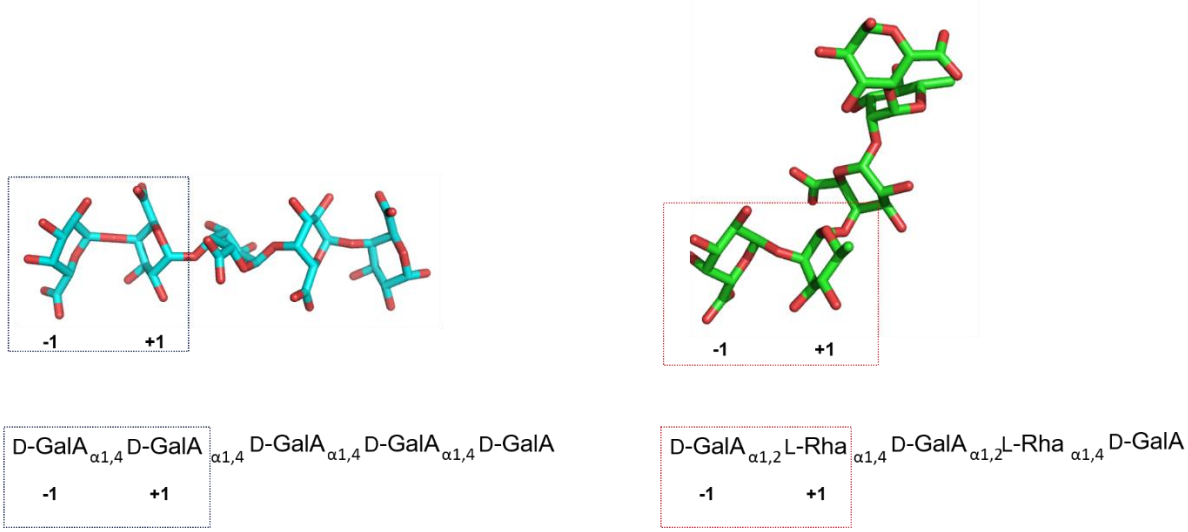


Figure 5.34 Comparison between HG and AM-RG-I oligosaccharides

The α1,4-D-GalA (HG pentasaccharide) and the α1,4-D-GalA-α1,2-L-Rha (RG-I pentasaccharide) are represented in blue and green, respectively. The sugars recognized in -1 and +1 subsites are highlighted inside the boxes. The 3-D representation for the HG and RG-I oligosaccharides was taken from the structures of the PL1 (PDB code 2NZM PDB) and PL4 (PDB code 3NJV) lyases, respectively. The figure was generated in PyMOL.

A distinctive feature of RG-I PUL is the expansion of enzymatic activities encoded by this locus. The backbone of this polysaccharide only contains two linkages that could, potentially, be degraded by two enzymes: a rhamnosidase, a rhamnogalacturonase. Instead, *B. thetaiotaomicron* encodes eleven enzymes in order to achieve the complete

degradation of this polysaccharide backbone. This extended repertoire of enzymes reflects different specificities. The lyases are the only endo-active enzymes encoded in RG-I PUL generating several oligosaccharides that can be used as substrate for the exo-active enzymes. The expansion of enzymes can represent a bacterial adaptation that mediates efficient degradation of the RG-I backbone, where the PLs generate short oligosaccharides that will be cleaved by a synergistic action of exo-active enzymes. The GH28 enzymes encoded in RG-I PUL represent an example of how the expansion of enzymatic activities is driven by different specificities. BT4153 cleaves D-GalA-α1,2-L-Rha, while BT4149 has a preference for a tetrasaccharide. A similar enzymatic specificity was described for the RG-I GH105 enzymes that cleave the unsaturated bonds Δ4,5-GalA-α1,2-L-Rha. BT4176 has a strong preference for the disaccharide, while BT4174 targets the tetrasaccharide Δ4,5-GalA-α1,2-L-Rha-α1,4-D-GalA-α1,2-L-Rha (Zhang 2015). Despite this expansion of enzymatic activities, RG-I PUL only encodes a single α-rhamnosidase, BT4145. This can be explained by the high activity of this GH106 independent of the oligosaccharide size. Additionally, in RG-I the L-Rha is always α1,4-linked to D-GalA so, in this case, it is not necessary for an additional α-rhamnosidase to overcome specificity issues. It is also possible that this PUL encodes an additional α-rhamnosidase. However, this hypothesis can be excluded because the genomic deletion of *bt4145* in *B. thetaiotaomicron* (performed by Dr. Aurore Labourel) resulted in no release of L-Rha when the sonicated cells of the mutant were incubated with P-RG-I and AM-RG-I, suggesting that this PUL only encodes a single α-rhamnosidase.

It was also evident during this work that RG-I contains some α1,4-D-GalA content. It has been reported that RG-I can be linked to HG in *Arabidopsis* and the linkage α-L-Rha-α1,4-D-GalA-α1,4-D-GalA-α1,4-D-GalA was detected in this structure (Tan *et al*.

2013). Additionally, some RG-I can be linked or intercalated with HG backbone (Atmodjo *et al*. 2013). Indeed, BT4155 is a exo-galacturonase and when incubated with P-RG-I releases D-GalA, suggesting that this polysaccharide contains some structures where HG is linked to RG-I. In this case, the oligosaccharide α-D-GalA-α1,4-D-GalA-[α1,2-L-Rha-α1,4-D-GalA]$_n$ would be present and, the GalA-GalA linkages (HG) would be cleaved by BT4155. The identity of the enzyme that cleaves the linkage at the boundary of HG and RG-I is unclear. The GH28 exo-galacturonase BT4146 is able to cleave the disaccharide D-GalA-α1,2-L-Rha. This enzyme is inactive against the AM-RG-I tetrasaccharide suggesting that this GH cannot tolerate L-Rha at the +3 subsite. However, it is not clear if the enzyme can tolerate D-GalA at +2 and +3 subsites and cleave of the RG-I-HG attachment linkage α1,4-D-GalA-α1,2-L-Rha-[α1,4-D-GalA-α1,4-D-GalA]$_n$. Additionally, BT4155 cleaves the HG linkage α1,4-D-GalA-α1,4-D-GalA. *B. thetaiotaomicron* encodes a HG PUL (Martens *et al*. 2011) so, it is unlikely that the enzyme targets this polysaccharide. However, the presence of *bt4155* in the RG-I PUL reflects the requirement to remove small chains of intercalated HG in order to release this blockage and allow the access of the rhamnosidase (BT4145) and additional exo-galacturonases (BT4149 and BT4153).

The work present in this chapter in combination with previous studies by Xiaoyang Zhang (Zhang 2015) and Jonathan Briggs (both from ICaMB, Newcastle University) allowed a model for RG-I utilization by *B. thetaiotaomicron* to be proposed (Figure 5.35). In this model, the galactan and arabinan present in RG-I are removed by enzymes encoded in two different PULs. These two systems were explored by Jonathan Briggs and will not be described in this thesis. However, it is important to notice that the enzymes encoded by the galactan PUL are unable to remove the last few D-Gal residues attached to L-Rha in the RG-I backbone (Briggs 2016). The cleavage of this short D-Gal side chains is mediated by enzymes encoded in RG-I PUL.

Degradation of RG-I is likely initiated by the binding of the polysaccharide to the SGBP BT4167 at the cell surface. Here BT4170 cleaves the backbone generating the oligosaccharides that are imported into the periplasm by the SusC/D-like system. The other enzymes, which act in synergy to depolymerize the backbone, are predicted to be in the periplasm and thus the bulk of the degradative process occurs at this cellular location. The truncated galactan side chains are removed by GH35 (BT4160) and GH2 (BT4151 and BT4156) β-D-galactosidases. The final decoration, acetyl groups linked at O2 or O3 of the D-GalA residues, is removed by the carbohydrate esterase BT4158. These enzymes release the steric constraints that would otherwise limit the myriad of exo-acting enzymes that act in synergy to degrade the RG-I backbone. The endo-acting PLs, BT4175 and BT4183 generate unsaturated oligosaccharides with limit DPs ranging from 2 to 10. One of the oligosaccharides generated is the signal molecule (Δ4,5-GalA-α1,2-L-Rha-α1,4-D-GalA-α1,2-L-Rha) that binds to the HTCS (BT4178) (Zhang 2015). The unsaturated Δ4,5-GalA from the signal molecule is specifically cleaved by the GH105 BT4174. The second GH105 encoded by the RG-I PUL, BT4176, targets the disaccharide Δ4,5-GalA-α1,2-L-Rha. BT4145 is the only α-rhamnosidase encoded in this PUL. This exo-acting GH106 enzyme cleaves the terminal L-Rha generating a new non-reducing end containing D-GalA. This acidic sugar is cleaved by the GH28 exo-galacturonases BT4149 and BT4153. The sequential backbone degradation by BT4145 and these GH28 enzymes allows the complete depolymerisation of this polysaccharide to D-GalA and L-Rha. An additional GH28, BT4146, also contributes for the backbone degradation by cleaving the disaccharide α1,4-D-GalA-α1,2-L-Rha. Additionally, the galacturonase BT4155 is implicated in cleaving the chains of HG attached to RG-I.

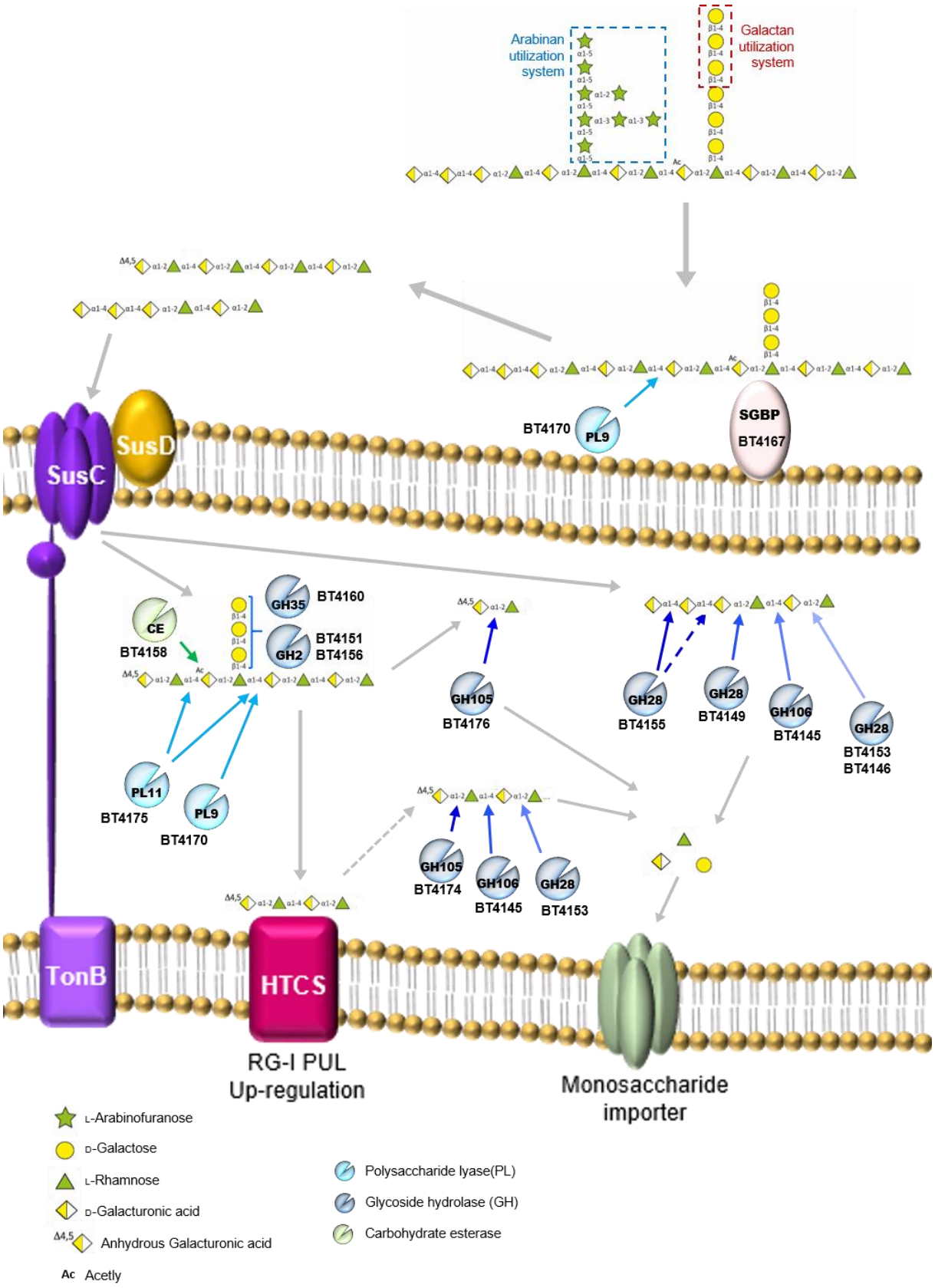


Figure 5.35 Proposed model of RG-I utilization by *B. thetaiotaomicron*

The linkages target by each enzyme are indicated with arrows. An interrupted arrow represents an activity not confirmed. Where the arrows are coloured in a hierarchical shades of blue indicates that this enzymes act in a sequential way. For details see text above.

5.5 Future work

The activities described in this chapter lead to a proposed model of RG-I degradation by *B. thetaiotaomicron*. However, this model requires to be tested *in vitro*. An experiment where all the enzymes are incubated with RG-I will represent a final confirmation of the proposed model. It is also important to notice that the RG-I PUL encodes twelve proteins of unknown function. An example is BT4152 that encodes a putative GH42 and GH43 catalytic module for which no activity was identified. BT4172 and other proteins of unknown function may also comprise members of new GH or PLs families. A screen of activities for these enzymes should be performed utilizing different RG-I sources. This will allow us to explore the activity against linkages present in different substrates such as, soy-bean or flax seeds.

In this chapter a SGBP was characterized as recognizing AM-RG-I. BT4167 displayed weak non-specific binding to HG suggesting that D-GalA contributes to ligand recognition. Since the O6 esterification of D-GalA in HG impacts on ligand binding, it is possible to speculate that basic amino acids form ionic interacts with the carboxylate of the GalA residue, similar to pectin recognition by $CBM77_{PL1/9}$ described in Chapter 3. The crystal structure of BT4167 would provide value information to test these hypotheses and disclose the interaction between the protein and its ligands. Additionally, since the PL9 cleaves the polysaccharide generating unsaturated oligosaccharides, it is important to understand if the other putative SGBP and SusD-like proteins are able to bind to PL9 products.

The characterization of the GH28 galacturonidases of RG-I PUL is presented in this chapter. However, it is essential to understand if BT4179 displays the same activity as the homologue studied here and also clarify the correct sequence of *bt4179* in the genome of *B. thetaiotaomicron*. The *B. thetaiotaomicron* VPI-5482 bacterial strain

should be re-ordered from the culture collection and the *bt4179* should be sequenced in order to confirm that the nucleotide insertion was not a result of a bacterial mutation.

In order to understand the substrate specificity of the galacturonidase BT4155 it is important to obtain the crystal structure of this protein in complex with ligands, such as D-GalA-α1,2-L-Rha and D-GalA-α1,4-D-GalA. This would allow disclosure of the interactions at the +1 subsite and understand how this protein can tolerate D-GalA and L-Rha at this location.

Additionally, it is important understand the significance of these GH28 enzymes in RG-I degradation. For example, exploring the role of BT4155 in the degradation mechanism may contribute to better characterization of the location of the short-chains of HG connected to RG-I, improving the general knowledge of pectin structural organization. It was speculated that BT4146 can be also implicated in the degradation of the RG-I-HG interface. It is important explore is this enzyme can tolerate D-GalA at +2 and +3 in order to complete the model of utilization of RG-I.

Chapter 6. Final discussion

Pectins are one of the major components of plant cell wall. These negatively charged polysaccharides are linear (homogalacturonan, HG) or highly branched (rhamnogalacturonan-I and II, RG-I and RG-II, respectively) (Atmodjo *et al*. 2013). This chapter will focus on how the work presented in this thesis contributed to the characterization of a new carbohydrate binding module (CBM) family that targets HG, and in our understanding of the mechanism by which RG-I and RG-II are degraded by *Bacteroides thetaiotaomicron*, a member of the human gut microbiota (HGM).

CBMs are appended to enzymes, such as glycoside hydrolases (GHs) and polysaccharide lyases (PLs), and contribute to the degradation of the complex plant cell wall structure by promoting enzyme-substrate contact (Gilbert 2010). Despite the increasing number of CBM families described in the CAZy database (Cantarel *et al*. 2009; Lombard *et al*. 2014), there is a lack of binding modules that target pectins. In Chapter 3 the founding member of the first CBM family ($CBM77_{PL1/9}$) that specifically binds to non-methylated HG was described. The structural characterization of $CBM77_{PL1/9}$ revealed a new mechanism by which CBMs target plant cell wall polysaccharides. Here ligand binding is meditated by the formation of salt bridges between the protein and unmethylated HG. This ionic binding mode is in sharp contrast to hydrophobic interactions that dominate CBM-ligand interactions (Boraston *et al*. 2004). Surprisingly $CBM77_{PL1/9}$ showed an apparent lack of potentiation of enzymatic activity against soluble substrates and pectins within plant cell walls. Indeed, one might argue why there is a need for a pectin-specific CBM family. Unlike the hemicellulose and cellulose components of plant cell walls, pectin is highly accessible to PLs and GHs and thus the requirement for CBMs that target HG is unclear, which is further emphasised by the plant cell wall studies reported in Chapter 3.3.6. However, the

presence of several CBM77 members in different biological contexts (soil, rumen and human gut), and their association with PLs in pectate/pectin lyase specific families suggests that this CBM family can contribute to the degradation of this acidic polysaccharide in a range of environments. It is possible that in certain cell walls $CBM77_{PL1/9}$ will target a component of the pectin that is not readily accessible to PLs and GHs. This hypothesis needs to be tested by exploring the contribution of $CBM77_{PL1/9}$ in a wide range of plant cell walls.

HGM plays an important role in the health and physiology of the human host. The major nutrients available to this ecosystem are host and dietary glycans. Therefore, the use of dietary manipulation to enhance the proliferation of beneficial bacteria can represent a powerful approach to improve human health (Hooper *et al*. 2002). To optimize this "dietary" approach it is critical to understand the mechanism by which the major glycans in the human diet are degraded by the HGM. Pectins account for a high percentage of the human diet (Jayani *et al*. 2005) and as such is a major nutrient for the HGM. Thus, exploring how pectins are degraded in the HGM may be relevant to the prebiotic and probiotic sectors of the food industry. Recently, *B. thetaiotaomicron*, a dominant member of human microbiota, was shown to utilize pectins (Martens *et al*. 2011). In this thesis the mechanisms by which *B. thetaiotaomicron* degraded RG-II (Chapter 4) and RG-I (Chapter 5) were explored.

RG-II is the most complex polysaccharide in nature. The model of degradation presented in Chapter 4 represents the first report of how this complex pectin can be depolymerized. Indeed, *B. thetaiotaomicron* can cleave 20 of the 21 linkages on RG-II. The extensive utilization of such a complex polysaccharide is achieved by a hierarchical exo-mode of degradation where the activity of an upstream enzyme is required before downstream GHs can access their target linkage. This exo-mode of degradation was exemplified in Chapter 4 by the model of Chain B degradation. It is

interesting that RG-II depolymerisation is mediated by enzymes that are linkage specific. For example, *B. thetaiotaomicron* expresses four α-L-rhamnosidases with each enzyme hydrolysing a specific rhamnosidic linkage in RG-II. This enzyme specificity contrasts with the RG-I degradation mechanism described in Chapter 5. To cleave the two linkages presented in the RG-I backbone, *B. thetaiotaomicron* encodes eleven enzymes, many with similar or overlapping specificities. The requirement for such an array of enzymes may reflect the cell localization of the GHs and PLs and the variable structure of RG-I. For example, the RG-I PUL encodes three PLs. One of these lyases (BT4170) was present on the cell surface and is essential for the degradation of the RG-I backbone. This lyase produces oligosaccharides that are imported into the periplasm. The two additional periplasmic PLs are required to endo-cleave the imported oligosaccharides producing the substrate for the exo-active enzymes. The requirement for two periplasmic PLs reflects the heterogeneity of the backbone decoration of RG-I. Additionally, in the RG-I degradation system enzymes displayed a preference for long (BT4149 and BT4174) or short oligosaccharides (BT4153 and BT4176) suggesting that this system is adapted to efficiently target the different sizes of oligosaccharides produced by the endo-active lyases. Thus, the apparent redundancy in the RG-I system likely reflects the heterogeneity of available substrate within the plant and generated by endo-acting enzymes. The structure of RG-II is highly conserved in *planta* and during the degradative process hence the deployment of enzymes that are bespoke for a specific linkage.

The structural data presented in Chapters 4 and 5 contributed to the understanding of the catalytic mechanism and substrate recognition by GHs and PLs. In Chapter 4 the first structure of a GH106 enzyme (BT0986) was presented leading to the identification of the catalytic apparatus. Alignment of the BT0986 sequence with the homologue acting on RG-I (BT4145), showed that the catalytic apparatus was conserved but little

further conservation was evident. This may suggest that BT4145 and BT0986 could be members of two different subfamilies. The structural characterization of BT1003 showed that the active site of this AceAse is different to other members of GH127. Indeed, BT1003 and respective homologues lacked the apparent catalytic acid/base residue previously identified in GH127 β-L-arabinofuranosidases. The low sequence similarity between β-L-arabinofuranosidases and AceAses suggests that these proteins can also be members of different GH127 subfamilies. In Chapter 5, the first structure of a PL9 rhamnogalacturonan lyase (BT4170) was described. The structure of BT4170 in complex with reaction products contributed to a better understanding of the β-elimination mechanism mediated by PL9 enzymes. When BT4170 was compared with a PL9 pectate lyase, the active site of both proteins show conservation of the catalytic apparatus. However, a low conservation was observed in residues implicated in substrate recognition. This suggests that the active site in PL9 enzymes evolved to specifically recognize the negatively charged polygalacturonic acid (pectate lyases) or the less charged RG-I backbone (rhamnogalacturonan lyases).

Overall, this thesis describes a CBM that has the potential of being used as probe to target HG in the plant cell walls, and contributed to a better understanding of RG-I and RG-II degradation by *B. thetaiotaomicron*. Additionally, pectins in the human diet are significant nutrients for the HGM. These acidic polysaccharides are also thought to have a positive impact on human health such as reducing blood cholesterol levels (Brouns *et al*. 2012). The mechanisms behind these beneficial properties of pectins is unknown. However, this study showed that these polysaccharides can be used by members of the HGM. It is possible that pectin utilization by some bacteria favour the proliferation of “beneficial” species present in human gut. Further studies should be done to understand the impact of these dietary polysaccharides in HGM composition.

Understanding how pectins are degraded and utilized by human gut bacteria could contribute to the development of a “dietary approach” to improve human health.

References

Abbott, D.W. and Boraston, A.B. 2007a. A family 2 pectate lyase displays a rare fold and transition metal-assisted beta-elimination. *J Biol Chem* 282 (48):35328-35336.

Abbott, D.W. and Boraston, A.B. 2007b. The structural basis for exopolygalacturonase activity in a family 28 glycoside hydrolase. *J Mol Biol* 368 (5):1215-1222.

Abbott, D.W., Gilbert, H.J., Boraston, A.B. 2010. The active site of oligogalacturonate lyase provides unique insights into cytoplasmic oligogalacturonate beta-elimination. *J Biol Chem* 285 (50):39029-39038.

Abbott, D.W., Hrynuik, S. and Boraston, A.B. 2007. Identification and characterization of a novel periplasmic polygalacturonic acid binding protein from *Yersinia enterolitica*. *J Mol Biol* 367 (4):1023-1033.

Abou Hachem, M., Nordberg Karlsson, E., Bartonek-Roxâ, E., Raghothama, S., Simpson, P.J., Gilbert, H.J., Williamson, M.P. and Holst, O. 2000. Carbohydrate-binding modules from a thermostable *Rhodothermus marinus* xylanase: cloning, expression and binding studies. Biochem J 345 (Pt 1):53-60.

Adesioye, F.A., Makhalanyane, T.P., Biely, P. and Cowan, D.A. 2016. Phylogeny, classification and metagenomic bioprospecting of microbial acetyl xylan esterases. Enzyme Microb Technol 93-94:79-91.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25 (17):3389-3402.

An, G., Wei, B., Xia, B., McDaniel, J.M., Ju, T., Cummings, R.D., Braun, J. and Xia, L. 2007. Increased susceptibility to colitis and colorectal tumors in mice lacking core 3-derived O-glycans. *J Exp Med* 204 (6):1417-1429.

Anderson, K.L. and Salyers, A.A. 1989. Biochemical evidence that starch breakdown by *Bacteroides thetaiotaomicron* involves outer membrane starch-binding sites and periplasmic starch-degrading enzymes. *J Bacteriol* 171 (6):3192-3198.

Anderson, K.L. and Salyers, A.A. 1989b. Genetic evidence that outer membrane binding of starch is required for starch utilization by *Bacteroides thetaiotaomicron*. *J Bacteriol* 171 (6):3199-3204.

Arumugam, M., Raes, J., Pelletier, E., Le Paslier, D., Yamada, T., Mende, D.R., Fernandes, G.R., Tap, J., Bruls, T., Batto, J.M., Bertalan, M., Borruel, N., Casellas, F., Fernandez, L., Gautier, L., Hansen, T., Hattori, M., Hayashi, T., Kleerebezem, M., Kurokawa, K., Leclerc, M., Levenez, F., Manichanh, C., Nielsen, H.B., Nielsen, T., Pons, N., Poulain, J., Qin, J., Sicheritz-Ponten, T., Tims, S., Torrents, D., Ugarte, E., Zoetendal, E.G., Wang, J., Guarner, F., Pedersen, O., de Vos, W.M., Brunak, S., Dore, J., Meta, H.I.T.C., Antolin, M., Artiguenave, F., Blottiere, H.M., Almeida, M., Brechot, C., Cara, C., Chervaux, C., Cultrone, A., Delorme, C., Denariaz, G., Dervyn, R., Foerstner, K.U., Friss, C., van de Guchte, M., Guedon, E., Haimet, F., Huber, W., van Hylckama-Vlieg, J., Jamet, A., Juste, C., Kaci, G., Knol, J., Lakhdari, O., Layec, S., Le Roux, K., Maguin, E., Merieux, A., Melo Minardi, R., M'Rini, C., Muller, J., Oozeer, R., Parkhill, J., Renault, P., Rescigno, M., Sanchez, N., Sunagawa, S., Torrejon, A., Turner, K., Vandemeulebrouck, G., Varela, E., Winogradsky, Y., Zeller, G., Weissenbach, J., Ehrlich, S.D. and Bork, P. 2011. Enterotypes of the human gut microbiome. *Nature* 473 (7346):174-180.

Atmodjo, M.A., Hao, Z. and Mohnen, D. 2013. Evolving views of pectin biosynthesis. *Annu Rev Plant Biol* 64:747-779.

Auclair, S.M., Bhanu, M.K. and Kendall, D.A. 2012. Signal peptidase I: cleaving the way to mature proteins. *Protein Sci* 21 (1):13-25.

Backhed, F., Ley,R.E., Sonnenburg, J.L., Peterson, D.A. and Gordon, J.I. 2005. Host-bacterial mutualism in the human intestine. *Science* 307 (5717):1915-1920.

Barbacci, A., Lahaye, M. and Magnenet, V. 2013. Another brick in the cell wall: biosynthesis dependent growth model. *PLoS One* 8 (9):e74400.

Bar-Peled, M., Urbanowicz, B.R. and O'Neill, M.A. 2012. The synthesis and origin of the pectic polysaccharide rhamnogalacturonan II - Insights from nucleotide sugar formation and diversity. *Front Plant Sci* 3:92.

Bayer, E.A. and Lamed, R. 1986. Ultrastructure of the cell surface cellulosome of *Clostridium thermocellum* and its interaction with cellulose. *J Bacteriol* 167 (3):828-836.

Bayer, E.A., Belaich, J.P., Shoham, Y. and Lamed, R. 2004. The cellulosomes: multienzyme machines for degradation of plant cell wall polysaccharides. *Annu Rev Microbiol* 58:521-554.

Ben David, Y., Dassa, B., Borovok, I., Lamed, R., Koropatkin, N.M., Martens, E.C., White. B.A., Bernalier-Donadille, A., Duncan, S.H., Flint, H.J., Bayer, E.A. and Moraïs, S. 2015. Ruminococcal cellulosome systems from rumen to human. *Environ Microbiol* 17 (9):3407-3426.

Berg Miller, M.E., Antonopoulos, D.A., Rincon, M.T., Band, M., Bari, A., Akraiko, T., Hernandez, A., Thimmapuram, J., Henrissat, B., Coutinho, P.M., Borovok, I., Jindou, S., Lamed, R., Flint, H.J., Bayer, E.A. and White, B.A. 2009. Diversity and strain specificity of plant cell wall degrading enzymes revealed by the draft genome of *Ruminococcus flavefaciens* FD-1. *PLoS One* 4 (8):e6650.

Berman, H.M., Westbrook, J.,Feng, Z., Gilliland, G., Bhat, T.N.,Weissig, H.,Shindyalov, I.N. and Bourne, P.E. 2000. The Protein Data Bank. *Nucleic Acids Res* 28:235-242.

Biely, P., Benen, J., Heinrichová, K., Kester, H.C. and Visser, J. 1996. Inversion of configuration during hydrolysis of alpha-1,4-galacturonidic linkage by three *Aspergillus* polygalacturonases. *FEBS Lett* 382 (3):249-255.

Bjursell, M.K.,Martens, E.C. and Gordon, J.I. 2006. Functional genomic and metabolic studies of the adaptations of a prominent adult human gut symbiont, *Bacteroides thetaiotaomicron*, to the suckling period. *J Biol Chem* 281 (47):36269-36279.

Blake, A.W., McCartney, L., Flint, J.E., Bolam, D.N., Boraston, A.B., Gilbert, H.J. and Knox, J.P. 2006. Understanding the biological rationale for the diversity of cellulose-directed carbohydrate-binding modules in prokaryotic enzymes. *J Biol Chem* 281 (39):29321-29329.

Blouzard, J.C., Coutinho, P.M., Fierobe, H.P., Henrissat, B., Lignon, S., Tardif, C., Pages, S. and de Philip, P. 2010. Modulation of cellulosome composition in Clostridium cellulolyticum: adaptation to the polysaccharide environment revealed by proteomic and carbohydrate-active enzyme analyses. *Proteomics* 10 (3):541-554.

Bolam, D.N., Ciruela, A., McQueen-Mason, S., Simpson, P., Williamson, M.P., Rixon, J.E., Boraston, A., Hazlewood, G.P. and Gilbert, H.J. 1998. *Pseudomonas* cellulose-binding domains mediate their effects by increasing enzyme substrate proximity. *Biochem J* 331 (Pt 3):775-781.

Boraston, A.B. and Abbott, D.W. 2012. Structure of a pectin methylesterase from *Yersinia enterocolitica*. *Acta Crystallogr Sect F Struct Biol Cryst Commun* 68 (Pt 2):129-133.

Boraston, A.B., Bolam, D.N., Gilbert, H.J. and Davies, G.J. 2004. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem J* 82 (Pt 3):769-781.

Briggs, J.A. 2016. Glycan utilisation and resource allocation by prominent members of the human gut microbiota, Institute for Cell and Molecular Biosciences, Newcastle University.

Brody, J.R. and Kern, S.E. 2004. History and principles of conductive media for standard DNA electrophoresis. *Anal Biochem* 333 (1):1-13.

Brouns, F., Theuwissen, E., Adam, A., Bell, M., Berger, A. and Mensink, R.P. 2012. Cholesterol-lowering properties of different pectin types in mildly hyper-cholesterolemic men and women. *Eur J Clin Nutr* 66 (5):591-599.

Caffall, K.H. and Mohnen, D. 2009. The structure, function, and biosynthesis of plant cell wall pectic polysaccharides. *Carbohydr Res* 344 (14):1879-1900.

Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V. and Henrissat, B. 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res* 37 (Database issue):D233-238.

Caporaso, J.G., Lauber, C.L., Costello, E.K., Berg-Lyons, D., Gonzalez, A., Stombaugh, J., Knights, D., Gajer, P., Ravel, J., Fierer, N., Gordon, J.I. and Knight, R. 2011. Moving pictures of the human microbiome. *Genome Biol* 12 (5):R50.

Charnock, S.J., Brown, I.E., Turkenburg, J.P., Black, G.W. and Davies, G.J. 2002. Convergent evolution sheds light on the anti-beta-elimination mechanism common to family 1 and 10 polysaccharide lyases. *Proc Natl Acad Sci U S A* 99 (19):12067-12072.

Cho, K.H. and Salyers, A.A. 2001. Biochemical analysis of interactions between outer membrane proteins that contribute to starch utilization by *Bacteroides thetaiotaomicron. J Bacteriol* 183 (24):7224-7230.

Clarke, A.E., Anderson, R.L. and Stone,B.A. 1979. Form and function of arabinogalactans and arabinogalactan-proteins. *Phytochemistry* 18 (4):521-540.

Clarke, S.F., Murphy, E.F., Nilaweera, K., Ross, P.R., Shanahan, F., O'Toole, P.W. and Cotter, P.D. 2012. The gut microbiota and its relationship to diet and obesity: new insights. *Gut Microbes* 3 (3):186-202.

Collins, T., Gerday, C. and Feller, G. 2005. Xylanases, xylanase families and extremophilic xylanases. *FEMS Microbiol Rev* 29 (1):3-23.

Cosgrove, D.J. 2005. Growth of the plant cell wall. *Nat Rev Mol Cell Biol* 6 (11):850-861.

Couturier, M., Roussel,A., Rosengren, A., Leone, P., Stalbrand,H. and Berrin, J.G. 2013. Structural and biochemical analyses of glycoside hydrolase families 5 and 26 beta-(1,4)-mannanases from *Podospora anserina* reveal differences upon manno-oligosaccharide catalysis. *J Biol Chem* 288 (20):14624-14635.

Creze, C., Castang, S., Derivery, E., Haser, R., Hugouvieux-Cotte-Pattat, N., Shevchik, V.E. and Gouet, P. 2008. The crystal structure of pectate lyase peli from soft rot pathogen *Erwinia chrysanthemi* in complex with its substrate. *J Biol Chem* 283 (26):18260-18268.

Cui, Z., Maruyama, Y., Mikami, B., Hashimoto, W. and Murata, K. 2007. Crystal structure of glycoside hydrolase family 78 alpha-L-rhamnosidase from *Bacillus* sp. GL1. *J Mol Biol* 374 (2):384-398.

Cuskin, F., Flint, J.E., Gloster, T.M., Morland, C., Basle, A., Henrissat, B., Coutinho, P.M., Strazzulli, A., Solovyova, A.S., Davies, G.J. and Gilbert, H.J. 2012. How nature can exploit nonspecific catalytic and carbohydrate binding modules to create enzymatic specificity. *Proc Natl Acad Sci U S A* 109 (51):20889-20894.

Cuskin, F., Lowe, E.C., Temple, M.J., Zhu, Y., Cameron, E.A., Pudlo, N.A., Porter, N.T., Urs, K., Thompson, A.J., Cartmell, A., Rogowski, A., Hamilton, B.S., Chen, R., Tolbert, T.J., Piens, K., Bracke, D., Vervecken, W., Hakki, Z., Speciale, G., Munōz-Munōz, J.L., Day, A., Peña, M.J., McLean, R., Suits, M.D., Boraston, A.B., Atherly, T., Ziemer, C.J., Williams, S.J., Davies, G.J., Abbott, D.W., Martens, E.C. and Gilbert, H.J. 2015. Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism. *Nature* 517 (7533):165-169.

Dassa, B., Borovok, I., Lamed, R., Henrissat, B., Coutinho, P., Hemme, C.L., Huang, Y., Zhou, J. and Bayer, E.A. 2012. Genome-wide analysis of *Acetivibrio cellulolyticus* provides a blueprint of an elaborate cellulosome system. *BMC Genomics* 13:210.

Davies, G. and Henrissat, B.. 1995. Structures and mechanisms of glycosyl hydrolases. *Structure* 3 (9):853-859.

Davies, G.,Wilson, K.S. and Henrissat, B. 1997. Nomenclature for sugar-binding subsites in glycosyl hydrolases. *Biochem J* 321 (Pt 2):557-559.

Davis, B.G., Brandstetter, T.W., Hackett, L., Winchester, B.G., Nash, R.J., Watson, A.A., and Griffiths, R.C. 1999. Tetrazoles of manno- and rhamno-pyranoses: Contrasting inhibition of mannosidases by [4.3.0] but of rhamnosidase by [3.3.0] bicyclic tetrazoles. *Tetrahedron* 55 (14):4489-4500.

de Beer, T.A., Berka, K., Thornton, J.M and Laskowski, R.A. 2014. PDBsum additions. *Nucleic Acids Res* 42 (Database issue):D292-296.

D'Elia, J.N. and Salyers, A.A. 1996a. Contribution of a neopullulanase, a pullulanase, and an alpha-glucosidase to growth of *Bacteroides thetaiotaomicron* on starch. *J Bacteriol* 178 (24):7173-7179.

D'Elia, J.N. and Salyers, A.A. 1996b. Effect of regulatory protein levels on utilization of starch by *Bacteroides thetaiotaomicron. J Bacteriol* 178 (24):7180-7186.

Delphi, L. and Sepehri, H. 2016. Apple pectin: A natural source for cancer suppression in 4T1 breast cancer cells in vitro and express p53 in mouse bearing 4T1 cancer tumors, in vivo. *Biomed Pharmacother* 84:637-644.

Dereeper, A., Audic, S., Claverie, J.M. and Blanc, G. 2010. BLAST-EXPLORER helps you building datasets for phylogenetic analysis. *BMC Evol Biol* 10:8.

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.F., Guindon, S., Lefort, V., Lescot, M., Claverie, J.M. and Gascuel O. 2008. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. *Nucleic Acids Res* 36 (Web Server issue):W465-469.

Despres, J., Forano, E., Lepercq, P., Comtet-Marre, S., Jubelin, G., Yeoman, C.J., Miller, M.E., Fields, C.J., Terrapon, N., Le Bourvellec, C., Renard, C.M., Henrissat, B., White, B.A. and Mosoni, P. 2016. Unraveling the pectinolytic function of *Bacteroides xylanisolvens* using a RNA-seq approach and mutagenesis. *BMC Genomics* 17:147.

Din, N., Damude, H.G., Gilkes, N.R., Miller, R.C.J., Warren, R.A. and Kilburn, D.G. 1994. C1-Cx revisited: intramolecular synergism in a cellulase. *Proc Natl Acad Sci U S A* 91 (24):11383-11387.

Ding, S.Y., Bayer, E.A., Steiner, D., Shoham, Y. and Lamed, R. 2000. A scaffoldin of the *Bacteroides cellulosolvens* cellulosome that contains 11 type II cohesins. *J Bacteriol* 182 (17):4915-4925.

Ding, S.Y., Rincon, M.T., Lamed, R., Martin, J.C., McCrae, S.I., Aurilia, V., Shoham, Y., Bayer, E.A. and Flint, H.J. 2001. Cellulosomal scaffoldin-like proteins from *Ruminococcus flavefaciens. J Bacteriol* 183 (6):1945-1953.

Ducros, V.M., Zechel, D.L., Murshudov, G.N., Gilbert, H.J., Szabó, L., Stoll, D., Withers, S.G. and Davies, G.J. 2002. Substrate distortion by a beta-mannanase: snapshots of the Michaelis and covalent-intermediate complexes suggest a B(2,5) conformation for the transition state. *Angew Chem Int Ed Engl* 41 (15):2824-2827.

Duncan, S.H., Scott, K.P., Ramsay, A.G., Harmsen, H.J.M., Welling, G.W., Stewart, C.S. and Flint, H.J. 2003. Effects of alternative dietary substrates on competition between human colonic bacteria in an anaerobic fermentor system. *Appl Environ Microbiol* 69 (2):1136-1142.

Eckburg, P.B., Bik, E.M., Bernstein, C.N., Purdom, E., Dethlefsen, L., Sargent, M., Gill, S.R., Nelson, K.E. and Relman, D.A. 2005. Diversity of the human intestinal microbial flora. *Science* 308 (5728):1635-1638.

Edgar, R.C. 2004a. MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinformatics* 5:113.

Edgar, R.C. 2004b. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32 (5):1792-1797

Elhenawy, W., Debelyy, M.O. and Feldman, M.F. 2014. Preferential packing of acidic glycosidases and proteases into *Bacteroides* outer membrane vesicles. *MBio* 5 (2):e00909-e00914.

Fanutti, C., Ponyi, T., Black, G.W., Hazlewood, G.P. and Gilbert, H.J. 1995. The conserved noncatalytic 40-residue sequence in cellulases and hemicellulases from anaerobic fungi functions as a protein docking domain. *J Biol Chem* 270 (49):29314-29322.

Ficko-Blean, E., Stuart, C.P., Suits, M.D., Cid, M., Tessier, M., Woods, R.J. and Boraston, A.B. 2012. Carbohydrate recognition by an architecturally complex α-N-acetylglucosaminidase from *Clostridium perfringens*. *PLoS One* 7 (3):e33524.

Finn, R.D., Coggill, P., Eberhardt, R.Y., Eddy, S.R., Mistry, J., Mitchell, A.L., Potter, S.C., Punta. M., Qureshi. M, Sangrador-Vegas, A., Salazar, G.A., Tate. J. and Bateman, A.2016. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res* 44 (D1):D279-285.

Fleischer, A., O'Neill, M.A. and Ehwald, R. 1999. The pore size of non-graminaceous plant cell walls is rapidly decreased by borate ester cross-linking of the pectic polysaccharide rhamnogalacturonan II. *Plant Physiol* 121 (3):829-838.

Flint, H.J. 2012. The impact of nutrition on the human microbiome. *Nutr Rev* 70 Suppl 1:S10-13.

Flint, H.J. and Bayer, E.A. 2008. Plant cell wall breakdown by anaerobic microorganisms from the Mammalian digestive tract. *Ann N Y Acad Sci* 1125:280-288.

Flint, J., Bolam, D.N., Nurizzo, D., Taylor, E.J., Williamson, M.P., Walters, C., Davies, G.J., Gilbert, H.J. 2005. Probing the mechanism of ligand recognition in family 29 carbohydrate-binding modules. *J Biol Chem* 280 (25):23718-23726.

Fontes, C.M. and Gilbert, H.J. 2010. Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates. *Annu Rev Biochem* 79:655-681.

Freelove, A.C., Bolam, D.N, White, P., Hazlewood, G.P. and Gilbert, H.J. 2001. A novel carbohydrate-binding protein is a component of the plant cell wall-degrading complex of *Piromyces equi*. *J Biol Chem* 276 (46):43010-43017.

Fujimoto, Z. 2013. Structure and function of carbohydrate-binding module families 13 and 42 of glycoside hydrolases, comprising a beta-trefoil fold. *Biosci Biotechnol Biochem* 77 (7):1363-1371.

Fujimoto, Z., Jackson, A., Michikawa, M., Maehara, T., Momma, M., Henrissat, B., Gilbert, H.J. and Kaneko, S. 2013. The structure of a *Streptomyces avermitilis* alpha-L-rhamnosidase reveals a novel carbohydrate-binding module CBM67 within the six-domain arrangement. *J Biol Chem* 288 (17):12376-12385.

García-Hernández, E. and Hernández-Arana, A.. 1999. Structural bases of lectin-carbohydrate affinities: comparison with protein-folding energetics. *Protein Sci* 8 (5):1075-1086.

Garron, M.L, and Cygler, M. 2010. Structural and mechanistic classification of uronic acid-containing polysaccharide lyases. *Glycobiology* 20 (12):1547-1573.

Garron, M.L, and Cygler, M. 2014. Uronic polysaccharide degrading enzymes. Curr Opin Struct Biol 28:87-95.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D. and Bairoch, A. 2005. Protein identification and analysis tools on the ExPASy server;. In The Pr*oteomics Protocols Handbook, edit*ed by T. P. P. Handbo*ok: Humana Press, 571*-607.

Georgelis, N., Yennawar, N.H. and Cosgrove, D.J. 2012. Structural basis for entropy-driven cellulose binding by a type-A cellulose-binding module (CBM) and bacterial expansin. *Proc Natl Acad Sci U S A* 109 (37):14830-14835.

Gilbert, H.J. 2010. The biochemistry and structural biology of plant cell wall deconstruction. *Plant Physiol* 153 (2):444-455.

Gilbert, H.J., Knox, J.P. and Boraston, A.B. 2013. Advances in understanding the molecular basis of plant cell wall polysaccharide recognition by carbohydrate-binding modules. *Curr Opin Struct Biol* 23 (5):669-677.

Gilkes, N.R., Warren, R.A., Miller, R.C.J. and Kilburn, D.G. 1988. Precise excision of the cellulose binding domains from two *Cellulomonas fimi* cellulases by a homologous protease and the effect on catalysis*. J Biol Chem* 263 (21):10401-10407.

Glenwright, A.J., Pothula, K.R., Bhamidimarri, S.P., Chorev, D.S., Basle, A., Zheng, H., Robinson, C.V., Winterhalter, M., Kleinekathöfer, U., Bolam, D.N. and van den Berg, B. 2017. Structural basis for nutrient acquisition by dominant members of the human gut microbiota. *Nature*.

Glushka, J.N., Terrell, M., York, W.S., O'Neill, M.A., Gucwa, A., Darvill, A.G., Albersheim, P. and Prestegard, J.H. 2003. Primary structure of the 2-*O*-methyl-alpha-L-fucose-containing side chain of the pectic polysaccharide, rhamnogalacturonan II. *Carbohydr Res* 338 (4):341-352.

Guinane, C.M. and Cotter, P.D. 2013. Role of the gut microbiota in health and chronic gastrointestinal disease: understanding a hidden metabolic organ. *Therap Adv Gastroenterol* 6 (4):295-308.

Hakulinen, N., Tenkanen, M. and Rouvinen, J. 2000. Three-dimensional structure of the catalytic core of acetylxylan esterase from *Trichoderma reesei*: insights into the deacetylation mechanism. *J Struct Biol* 132 (3):180-190.

Henrissat, B. 1991. A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J* 280 (Pt 2):309-316.

Henrissat, B. and Bairoch, A. 1993. New families in the classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J* 293 (Pt 3):781-788.

Henrissat, B. and Bairoch, A. 1996. Updating the sequence-based classification of glycosyl hydrolases. *Biochem J* 316 (Pt 2):695-696.

Henrissat, B. and Davies, G. 1997. Structural and sequence-based classification of glycoside hydrolases. *Curr Opin Struct Biol* 7 (5):637-644.

Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J.P., Davies, G. 1995. Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. *Proc Natl Acad Sci U S A* 92 (15):7090-7094.

Herrero, M., de Lorenzo, V. and Timmis, K. N. 1990. Transposon vectors containing non-antibiotic resistance selection markers for cloning and stable chromosomal insertion of foreign genes in gram-negative bacteria. *J Bacteriol* 172 (11):6557–6567.

Herron, S.R., Scavetta, R.D., Garrett, M., Legner, M. and Jurnak, F. 2003. Characterization and implications of Ca2+ binding to pectate lyase C. *J Biol Chem* 278 (14):12271-12277.

Herve, C., Rogowski, A., Blake, A.W., Marcus, S.E., Gilbert, H.J. and Knox, J.P. 2010. Carbohydrate-binding modules promote the enzymatic deconstruction of intact plant cell walls by targeting and proximity effects. *Proc Natl Acad Sci U S A* 107 (34):15293-15298.

Hidaka, M., Nishimoto, M., Kitaoka, M., Wakagi, T., Shoun, H. and Fushinobu, S. 2009. The crystal structure of galacto-N-biose/lacto-N-biose I phosphorylase: a large deformation of a TIM barrel scaffold. *J Biol Chem* 284 (11):7273-7283.

Hileman, R.E., Fromm, J.R., Weiler, J.M. and Linhardt, R.J. 1998. Glycosaminoglycan-protein interactions: definition of consensus sites in glycosaminoglycan binding proteins. *Bioessays* 20 (2):156-167.

Hold, G.L. 2014. Western lifestyle: a 'master' manipulator of the intestinal microbiota? *Gut* 63 (1):5-6.

Holm, L. and Rosenstrom, P. 2010. Dali server: conservation mapping in 3D. *Nucleic Acids Res* 38 (Web Server issue):W545-549.

Hooper, L.V., Midtvedt, T. and Gordon, J.I. 2002. How host-microbial interactions shape the nutrient environment of the mammalian intestine. *Annu Rev Nutr* 22:283-307.

Hu, S., Dong, T.S., Dalal, S.R., Wu, F., Bissonnette, M., Kwon, J.H. and Chang, E.B. 2011. The microbe-derived short chain fatty acid butyrate targets miRNA-dependent p21 gene expression in human colon cancer. *PLoS One* 6 (1):e16221.

Huang, C.H., Zhu, Z., Chen, H.C., Ko, T.P., Chen, C.C., Wang, I., Ho, M.R., Zeng, S.D, Huang, Y.F., Liu, J.R. and Guo, R.T. 2014. Structure and catalytic mechanism of a glycoside hydrolase family 127 β-L-arabinofuranosidase (HyBA 1). J Bioproces *Biotech 4 (6).*

Inngjerdingen, K.T., Patel, T.R., Chen, X., Kenne, L., Allen, S., Morris, G.A., Harding, S.E., Matsumoto, T., Diallo, D., Yamada, H., Michaelsen, T.E., Inngjerdingen, M. and Paulsen, B.S. 2007. Immunological and structural properties of a pectic polymer from *Glinus oppositifolius*. *Glycobiology* 17 (12):1299-1310.

Ishii, T. and Ono, H. 1999. NMR spectroscopic analysis of the borate diol esters of methyl apiofuranosides. *Carbohydr Res* 321 (3):257-260.

Ishii, T., Matsunaga, T., Pellerin, P., O'Neill, M.A., Darvill, A. and Albersheim, P. 1999. The plant cell wall polysaccharide rhamnogalacturonan II self-assembles into a covalently cross-linked dimer. *J Biol Chem* 274 (19):13098-13104.

Ito, T., Saikawa, K., Kim, S., Fujita, K., Ishiwata, A., Kaeothip, S., Arakawa, T., Wakagi, T., Beckham, G.T., Ito, Y. and Fushinobu S. 2014. Crystal structure of glycoside hydrolase family 127 beta-l-arabinofuranosidase from *Bifidobacterium longum*. *Biochem Biophys Res Commun* 447 (1):32-37.

Itoh, T., Ochiai, A., Mikami, B., Hashimoto, W. and Murata, K. 2006b. Structure of unsaturated rhamnogalacturonyl hydrolase complexed with substrate. *Biochem Biophys Res Commun* 347 (4):1021-1029.

Itoh, T., Ochiai, A., Mikami, B., Hashimoto, W. and Murata, K. 2006a. A novel glycoside hydrolase family 105: the structure of family 105 unsaturated rhamnogalacturonyl hydrolase complexed with a disaccharide in comparison with family 88 enzyme complexed with the disaccharide. *J Mol Biol* 360 (3):573-585.

Iwai, H., Masaoka, N., Ishii, T. and Satoh, S. 2002. A pectin glucuronyltransferase gene is essential for intercellular attachment in the plant meristem. Pro*c Natl Acad Sci U S A 99* (25):16319-16324.

Jackson, C.L., Dreaden, T.M., Theobald, L.K., Tran, N.M., Beal, T.L., Eid, M., Gao, M.Y., Shirley, R.B., Stoffel, M.T., Kumar, M.V. and Mohnen, D. 2007. Pectin induces apoptosis in human prostate cancer cells: correlation of apoptotic function with pectin structure. *Glycobiology* 17 (8):805-819.

Jamal-Talabani, S., Boraston, A.B., Turkenburg, J.P., Tarbouriech, N., Ducros, V.M. and Davies, G.J. 2004. Ab initio structure determination and functional characterization of CBM36; a new family of calcium-dependent carbohydrate binding modules. *Structure* 12 (7):1177-1187.

Jayani, R.S., Saxena, S. and Gupta, R. 2005. Microbial pectinolytic enzymes: A review. *Process Biochem* 40 (9):2931-2944.

Jenkins, J. and Cook, M. 2004. Mosquito®: an accurate nanoliter dispensing technology. *JALA* 9 (4):257-261.

Jenkins, J., Shevchik, V.E., Hugouvieux-Cotte-Pattat, N. and Pickersgill, R.W. 2004. The crystal structure of pectate lyase Pel9A from *Erwinia chrysanthemi*. *J Biol Chem* 279 (10):9139-9145.

Jensen, M.H., Otten, H., Christensen, U., Borchert, T.V., Christensen, L.L., Larsen, S., Leggio and L.L. 2010. Structural and biochemical studies elucidate the mechanism of rhamnogalacturonan lyase from *Aspergillus aculeatus*. *J Mol Biol* 404 (1):100-111.

Jindou, S., Borovok, I., Rincon, M.T., Flint, H.J., Antonopoulos, D.A., Berg, M.E., White, B.A., Bayer, E.A. and Lamed, R. 2006. Conservation and divergence in cellulosome architecture between two strains of *Ruminococcus flavefaciens. J Bacteriol* 188 (22):7971-7976.

Johansen, K.S. 2016. Lytic polysaccharide monooxygenases: the microbial power tool for lignocellulose degradation. *Trends Plant Sci* 21 (11):926-936.

Johnson, M., Zaretskaya, I., Raytselis, Y., Merezhuk, Y., McGinnis, S. and Madden, T.L. 2008. NCBI BLAST: a better web interface. *Nucleic Acids Res* 36 (Web Server issue):W5-9.

Jongkees, S.A. and Withers, S.G. 2011. Glycoside cleavage by a new mechanism in unsaturated glucuronyl hydrolases. *J Am Chem Soc* 133 (48):19334-19337.

Juncker, A.S., Willenbrock, H., Von Heijne, G., Brunak, S., Nielsen, H. and Krogh, A. 2003. Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci* 12 (8):1652-1662.

Kakiuchi, M., Isui, A., Suzuki, K., Fujino, T., Fujino, E., Kimura, T., Karita, S., Sakka, K. and Ohmiya, K. 1998. Cloning and DNA sequencing of the genes encoding *Clostridium josui* scaffolding protein CipA and cellulase CelD and identification of their gene products as major components of the cellulosome. *J Bacteriol* 180 (16):4303-4308.

Kanehisa, M. and Goto, S. 2000. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28 (1):27-30.

Kibbe, W.A. 2007. OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Res* 35 (Web Server issue):W43-46.

Koropatkin, N.M. and Smith, T.J. 2010. SusG: a unique cell-membrane-associated alpha-amylase from a prominent human gut symbiont targets complex starch molecules. *Structure* 18 (2):200-215.

Koropatkin, N.M., Cameron, E.A. and Martens, E.C. 2012. How glycan metabolism shapes the human gut microbiota. *Nat Rev Microbiol* 10 (5):323-335.

Koropatkin, N.M., Martens, E.C., Gordon, J.I. and Smith, T.J. 2008. Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices. *Structure* 16 (7):1105-1115.

Krause, D.O., Bunch, R.J., Smith, W.J.M. and McSweeney, C.S. 1999. Diversity of Ruminococcus strains: a survey of genetic polymorphisms and plant digestibility. *J Appl Microbiol* 86 (3):487-495.

Laemmli, U.K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* 227:680-685

Lamed, R., Naimark, J., Morgenstern, E. and Bayer, E.A. 1987. Specialized cell surface structures in cellulolytic bacteria. *J Bacteriol* 169 (8):3792-3800.

Larsbrink, J., Rogers, T.E., Hemsworth, G.R., McKee, L.S., Tauzin, A.S., Spadiut, O., Klinter, S., Pudlo, N.A., Urs. K., Koropatkin, N.M., Creagh, A.L., Haynes, C.A., Kelly, A.G., Cederholm, S.N., Davies, G.J., Martens, E.C. and Brumer, H. 2014. A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes. *Nature* 506 (7489):498-502.

Lattimer, J.M. and Haub, M.D. 2010. Effects of dietary fiber and its components on metabolic health. *Nutrients* 2 (12):1266-1289.

Ley, R.E., Backhed, F., Turnbaugh, P., Lozupone, C.A., Knight, R.D. and Gordon, J.I. 2005. Obesity alters gut microbial ecology. *Proc Natl Acad Sci U S A* 102 (31):11070-11075.

Ley, R.E., Hamady, M., Lozupone, C., Turnbaugh, P.J., Ramey, R.R., Bircher, J.S., Schlegel, M.L., Tucker, T.A., Schrenzel, M.D., Knight, R. and Gordon, J.I. 2008. Evolution of mammals and their gut microbes. *Science* 320 (5883):1647-1651.

Ley, R.E., Peterson, D.A. and Gordon, J.I. 2006. Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell* 124 (4):837-848.

Li, X.L., Chen, H. and Ljungdahl, L.G. 1997. Two cellulases, CelA and CelC, from the polycentric anaerobic fungus *Orpinomyces* strain PC-2 contain N-terminal docking domains for a cellulase-hemicellulase complex. *Appl Environ Microbiol* 63 (12):4721-4728.

Lin, L., Xu, W., Liang, H., He, L., Liu, S., Li, Y., Li, B. and Chen, Y. 2015. Construction of pH-sensitive lysozyme/pectin nanogel for tumor methotrexate delivery. Colloids Surf B Biointerfaces 126:459-466.

Lodish, H., Berk, A., Zipursky, S.L., Matsudaira, P., Baltimore, D. and Darnell, J. 2000. The Dynamic Plant Cell Wall, in Molecular Cell Biology. 4th edn. New York: W. H. Freeman.

Lombard, V., Bernard, T., Rancurel, C., Brumer, H., Coutinho, P.M. and Henrissat, B. 2010. A hierarchical classification of polysaccharide lyases for glycogenomics. *Biochem J* 432 (3):437-444.

Lombard, V., Golaconda Ramulu, H., Drula, E., Coutinho, P.M. and Henrissat, B. 2014. The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Res* 42 (Database issue):D490-495.

Luis, A.S., Venditto, I., Temple, M.J., Rogowski, A., Basle, A., Xue, J., Knox, J.P., Prates, J.A., Ferreira, L.M., Fontes, C.M., Najmudin, S. and Gilbert, H.J. 2013. Understanding how noncatalytic carbohydrate binding modules can display specificity for xyloglucan. *J Biol Chem* 288 (7):4799-4809.

Maksimainen, M., Paavilainen, S., Hakulinen, N. and Rouvinen, J. 2012. Structural analysis, enzymatic characterization, and catalytic mechanisms of beta-galactosidase from *Bacillus circulans sp. alkalophilus*. *FEBS J* 279 (10):1788-1798.

Marietta, E.V., Gomez, A.M., Yeoman, C., Tilahun, A.Y., Clark, C.R., Luckey, D.H., Murray, J.A., White, B.A., Kudva, Y.C. and Rajagopalan, G. 2013. Low incidence of spontaneous type 1 diabetes in non-obese diabetic mice raised on gluten-free diets is associated with changes in the intestinal microbiome. *PLoS One* 8 (11):e78687.

Markovic, O. and Janecek, S. 2001. Pectin degrading glycoside hydrolases of family 28: sequence-structural features, specificities and evolution. *Protein Eng* 14 (9):615-631.

Martens, E.C., Chiang, H.C. and Gordon, J.I. 2008. Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. *Cell Host Microbe* 4 (5):447-457.

Martens, E.C., Koropatkin, N.M., Smith, T.J. and Gordon, J.I. 2009. Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm. *J Biol Chem* 284 (37):24673-24677.

Martens, E.C., Lowe, E.C., Chiang, H., Pudlo, N.A., Wu, M., McNulty, N.P., Abbott, D.W., Henrissat, B., Gilbert, H.J., Bolam, D.N. and Gordon, J.I. 2011. Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol* 9 (12):e1001221.

Martens, E.C., Roth, R., Heuser, J.E. and Gordon, J.I. 2009b. Coordinate regulation of glycan degradation and polysaccharide capsule biosynthesis by a prominent human gut symbiont. *J Biol Chem* 284 (27):18445-18457.

Matsui, I., Ishikawa, K., Matsui, E., Miyairi, S., Fukui, S. and Honda, K. 1991. Subsite structure of Saccharomycopsis alpha-amylase secreted from *Saccharomyces cerevisiae*. *J Biochem* 109 (4):566-569.

Matsunaga, T., Ishii, T., Matsumoto, S., Higuchi, M., Darvill, A., Albersheim, P. and O'Neill, M.A. 2004. Occurrence of the primary cell wall polysaccharide rhamnogalacturonan II in pteridophytes, lycophytes, and bryophytes. Implications for the evolution of vascular plants. *Plant Physiol* 134 (1):339-351.

Maxwell, E.G., Colquhoun, I.J., Chau, H.K., Hotchkiss, A.T., Waldron, K.W., Morris, V.J. and Belshaw, N. 2016. Modified sugar beet pectin induces apoptosis of colon cancer cells via an interaction with the neutral sugar side-chains. *Carbohydr Polym* 136:923-929.

Mayans, O., Scott, M., Connerton, I., Gravesen, T., Benen, J., Visser, J., Pickersgill, R. and Jenkins, J. 1997. Two crystal structures of pectin lyase A from *Aspergillus* reveal a pH driven conformational change and striking divergence in the substrate-binding clefts of pectin and pectate lyase. *Structure* 5 (5):677-689.

McBride, M.J., Xie, G., Martens, E.C., Lapidus, A., Henrissat, B., Rhodes, R.G., Goltsman, E., Wang, W., Xu, J., Hunnicutt, D.W., Staroscik, A.M., Hoover, T.R., Cheng, Y.Q. and Stein, J.L. 2009. Novel features of the polysaccharide-digesting gliding bacterium *Flavobacterium johnsoniae* as revealed by genome sequence analysis. *Appl Environ Microbiol* 75 (21):6864-6875.

McCarter, J.D. and Withers, G.S. 1994. Mechanisms of enzymatic glycoside hydrolysis. *Curr Opin Struct Biol* 4:885-892.

McCartney, L., Blake, A.W., Flint, J., Bolam, D.N., Boraston, A.B., Gilbert, H.J. and Knox, J.P. 2006. Differential recognition of plant cell walls by microbial xylan-specific carbohydrate-binding modules. *Proc Natl Acad Sci U S A* 103 (12):4765-4770.

McKee, L.S., Peña, M.J., Rogowski, A., Jackson, A., Lewis, R.J., York, W.S., Krogh, K.B., Viksø-Nielsen, A., Skjøt, M., Gilbert, H.J. and Marles-Wright, J. 2012. Introducing endo-xylanase activity into an exo-acting arabinofuranosidase that targets side chains. *Proc Natl Acad Sci U S A* 109 (17):6537-6542.

McKie, V.A., Vincken, J.P., Voragen, A.G., van den Broek, L.A., Stimson, E. and Gilbert, H.J. 2001. A new family of rhamnogalacturonan lyases contains an enzyme that binds to cellulose. *Biochem J* 355 (Pt 1):167-177.

McNeil, N.I. 1984. The contribution of large intestine to energy supplies in man. *Am J Clin Nutr* 39 (2):338-342.

Midtvedt, T., Zabarovsky, E., Norin, E., Bark, J., Gizatullin, R., Kashuba, V., Ljungqvist, O., Zabarovska, V., Möllby, R. and Ernberg, I. 2013. Increase of faecal tryptic activity relates to changes in the intestinal microbiome: analysis of Crohn's disease with a multidisciplinary platform. *PLoS One* 8 (6):e66074.

Miyata, T., Kashige, N., Satho, T., Yamaguchi, T., Aso, Y. and Miake, F. 2005. Cloning, sequence analysis, and expression of the gene encoding *Sphingomonas paucimobilis* FP2001 alpha-L-rhamnosidase. *Curr Microbiol* 51 (2):105-109.

Mohnen, D. 2008. Pectin structure and biosynthesis. *Curr Opin Plant Biol* 11 (3):266-277.

Montanier, C., van Bueren, A.L., Dumon, C., Flint, J.E., Correia, M.A., Prates, J.A., Firbank, S.J., Lewis, R.J., Grondin, G.G., Ghinet, M.G., Gloster, T.M., Herve, C., Knox, J.P., Talbot, B.G., Turkenburg, J.P., Kerovuo, J., Brzezinski, R., Fontes, C.M., Davies, G.J., Boraston, A.B. and Gilbert, H.J. 2009. Evidence that family 35 carbohydrate binding modules display conserved specificity but divergent function. *Proc Natl Acad Sci U S A* 106 (9):3065-3070.

Montanier, C.Y., Correia, M.A., Flint, J.E., Zhu, Y., Basle, A., McKee, L.S., Prates, J.A., Polizzi, S.J., Coutinho, P.M., Lewis, R.J., Henrissat, B., Fontes, C.M. and Gilbert, H.J. 2011. A novel, noncatalytic carbohydrate-binding module displays specificity for galactose-containing polysaccharides through calcium-mediated oligomerization. *J Biol Chem* 286 (25):22499-22509.

Mullis, K.B. and Faloona, F.A. 1987. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol* 155:335-350.

Mutter, M., Beldman, G., Pitson, S.M., Schols, H.A. and Voragen, A.G. 1998. Rhamnogalacturonan alpha-d-galactopyranosyluronohydrolase. An enzyme that specifically removes the terminal nonreducing galacturonosyl residue in rhamnogalacturonan regions of pectin. *Plant Physiol.* 1998 May;117(1):153-63

Mutter, M., Beldman, G., Schols, H.A. and Voragen, A.G. 1994 Rhamnogalacturonan alpha-L-rhamnopyranohydrolase. A novel enzyme specific for the terminal nonreducing rhamnosyl unit in rhamnogalacturonan regions of pectin. *Plant Physiol* 106 (1):241-250.

Najmudin, S., Guerreiro, C.I., Carvalho, A.L., Prates, J.A., Correia, M.A., Alves, V.D., Ferreira, L.M., Romao, M.J., Gilbert, H.J., Bolam, D.N. and Fontes, C.M. 2006. Xyloglucan is recognized by carbohydrate-binding modules that interact with beta-glucan chains. *J Biol Chem* 281 (13):8815-8828.

Najmudin, S., Pinheiro, B.A., Prates, J.A., Gilbert, H.J., Romao, M.J. and Fontes, C.M. 2010. Putting an N-terminal end to the *Clostridium thermocellum* xylanase Xyn10B story: crystal structure of the CBM22-1-GH10 modules complexed with xylohexaose. *J Struct Biol* 172 (3):353-362.

Naran, R., Chen, G. and Carpita, N.C. 2008. Novel rhamnogalacturonan I and arabinoxylan polysaccharides of flax seed mucilage. *Plant Physiol* 148 (1):132-141.

Ochiai, A., Itoh, T., Maruyama, Y., Kawamata, A., Mikami, B., Hashimoto, W. and Murata, K. 2007. A novel structural fold in polysaccharide lyases: *Bacillus subtilis* family 11 rhamnogalacturonan lyase YesW with an eight-bladed beta-propeller. *J Biol Chem* 282 (51):37134-37145.

O'Neill, E.C., Stevenson, C.E., Paterson, M.J., Rejzek, M., Chauvin, A.L., Lawson, D.M. and Field, R.A. 2015. Crystal structure of a novel two domain GH78 family alpha-rhamnosidase from *Klebsiella oxytoca* with rhamnose bound. *Proteins* 83 (9):1742-1749.

O'Neill, M.A., Eberhard, S., Albersheim, P. and Darvill, A.G. 2001. Requirement of borate cross-linking of cell wall rhamnogalacturonan II for *Arabidopsis* growth. *Science* 294 (5543):846-849.

O'Neill, M.A., Ishii, T., Albersheim, P. and Darvill, A.G. 2004. Rhamnogalacturonan II: structure and function of a borate cross-linked cell wall pectic polysaccharide. *Annu Rev Plant Biol* 55:109-139.

Oomen, R.J., Doeswijk-Voragen, C.H., Bush, M.S., Vincken, J.P., Borkhardt, B., van den Broek, L.A., Corsar, J., Ulvskov, P., Voragen, A.G., McCann, M.C. and Visser, R.G. 2002. In muro fragmentation of the rhamnogalacturonan I backbone in potato (*Solanum tuberosum* L.) results in a reduction and altered location of the galactan and arabinan side-chains and abnormal periderm development. *Plant J* 30 (4):403-413.

Ottman, N., Smidt, H., de Vos, W.M. and Belzer, C. 2012. The function of our microbiota: who is out there and what do they do? *Front Cell Infect Microbiol* 2:104.

Paal, K., Ito, M. and Withers, S.G. 2004. *Paenibacillus* sp. TS12 glucosylceramidase: kinetic studies of a novel sub-family of family 3 glycosidases and identification of the catalytic residues. *Biochem J* 378 (Pt 1):141-149.

Pabst, M., Fischl, R.M., Brecker, L., Morelle, W., Fauland, A., Köfeler, H., Altmann, F. and Léonard, R. 2013. Rhamnogalacturonan II structure shows variation in the side chains monosaccharide composition and methylation status within and across different plant species. *Plant J* 76 (1):61-72.

Pagès, S., Valette, O., Abdou, L., Bélaïch, A. and Bélaïch, J.P. 2003. A rhamnogalacturonan lyase in the *Clostridium cellulolyticum* cellulosome. *J Bacteriol* 185 (16):4727-4733.

Pell, G. 2004. Structural an functional analysis of Cellvibrio hemicellulases, Institute for Cell and Molecular Biosciences, Newcastle University.

Pellerin, P., Doco, T., Vidal, S., Williams, P. and Brillouet, J.M. 1996. The pectic polysaccharide rhamnogalacturonan II is a major component of the polysaccharides present in fruit-derived products. *Progress Biotechnol* 14:67-78.

Petersen, T.N., Brunak, S., von Heijne, G. and Nielsen, H. 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods* 8 (10):785-786.

Pettipher, G.L. and Latham, M.J. 1979. Characteristics of enzymes produced by *Ruminococcus flavefaciens* which degrade plant cell walls. *J Gen Microbiol* 110:21-27.

Pickersgill, R., Smith, D., Worboys, K. and Jenkins, J. 1998. Crystal structure of polygalacturonase from *Erwinia carotovora* ssp. *carotovora. J Biol Chem* 273 (38):24660-24664.

Pires, V.M., Henshaw, J.L., Prates, J.A., Bolam, D.N., Ferreira, L.M., Fontes, C.M., Henrissat, B., Planas, A., Gilbert, H.J. and Czjzek, M. 2004. The crystal structure of the family 6 carbohydrate binding module from *Cellvibrio mixtus* endoglucanase 5a in complex with oligosaccharides reveals two distinct binding sites with different ligand specificities. *J Biol Chem* 279 (20):21560-21568.

Pitson, S.M., Mutter, M., van den Broek, L.A., Voragen, A.G. and Beldman, G. 1998. Stereochemical course of hydrolysis catalysed by alpha-L-rhamnosyl and alpha-D-galacturonosyl hydrolases from *Aspergillus aculeatus*. *Biochem Biophys Res Commun* 242 (3):552-559.

Pohlschröder, M., Leschine, S.B. and Canale-Parola, E. 1994. Multicomplex cellulase-xylanase system of *Clostridium papyrosolven*s C7. *J Bacteriol* 176 (1):70-76.

Ragauskas, A.J., Williams, C.K., Davison, B.H., Britovsek, G., Cairney, J., Eckert, C.A., Frederick, W.J.Jr., Hallett, J.P., Leak, D.J., Liotta, C.L., Mielenz, J.R., Murphy, R., Templer, R. and Tschaplinski, T. 2006. The path forward for biofuels and biomaterials. *Science* 311 (5760):484-489.

Raghothama, S., Simpson, P.J., Szabó, L., Nagy, T., Gilbert, H.J. and Williamson, M.P. 2000. Solution structure of the CBM10 cellulose binding module from *Pseudomonas xylanase A. Biochemistry* 39 (5):978-984

Rakoff-Nahoum, S., Coyne, M.J. and Comstock, L.E. 2014. An ecological network of polysaccharide utilization among human intestinal symbionts. Curr B*iol 24 (1)*:40-49.

Reeves, A.R., D'Elia, J.N., Frias, J. and Salyers, A.A. 1996. A *Bacteroides thetaiotaomicron* outer membrane protein that is essential for utilization of maltooligosaccharides and starch. *J Bacteriol* 178 (3):823-830.

Reeves, A.R., Wang, G.R. and Salyers, A.A. 1997. Characterization of four outer membrane proteins that play a role in utilization of starch by *Bacteroides thetaiotaomicron. J Bacteriol* 179 (3):643-649.

Rincon, M.T., Cepeljnik, T., Martin, J.C., Barak, Y., Lamed, R., Bayer, E.A. and Flint, H.J. 2007. A novel cell surface-anchored cellulose-binding protein encoded by the sca gene cluster of *Ruminococcus flavefaciens. J Bacteriol* 189 (13):4774-4783.

Rincon, M.T., Cepeljnik, T., Martin, J.C., Lamed, R., Barak, Y., Bayer, E.A. and Flint, H.J. 2005. Unconventional mode of attachment of the *Ruminococcus flavefaciens* cellulosome to the cell surface. *J Bacteriol* 187 (22):7569-7578.

Rincon, M.T., Dassa, B., Flint, H.J., Travis, A.J., Jindou, S., Borovok, I., Lamed, R., Bayer, E.A., Henrissat, B., Coutinho, P.M., Antonopoulos, D.A., Berg Miller, M.E. and White, B.A. 2010. Abundance and diversity of dockerin-containing proteins in the fiber-degrading rumen bacterium, *Ruminococcus flavefaciens* FD-1. *PLoS One* 5 (8):e12476.

Rincon, M.T., McCrae, S.I., Kirby, J., Scott, K.P. and Flint, H.J. 2001. EndB, a multidomain family 44 cellulase from *Ruminococcus flavefaciens* 17, binds to cellulose via a novel cellulose-binding module and to another *R. flavefaciens* protein via a dockerin domain. *Appl Environ Microbiol* 67 (10):4426-4431.

Robert, X. and Gouet, P. 2014. Deciphering key features in protein structures with the new ENDscript server. *Nucleic Acids Res* 42 (Web Server issue):W320-324.

Rogers, T.E., Pudlo, N.A., Koropatkin, N.M., Bell, J.S., Moya Balasch, M., Jasker, K. and Martens, E.C. 2013. Dynamic responses of *Bacteroides thetaiotaomicron* during growth on glycan mixtures. *Mol Microbiol* 88 (5):876-890.

Rogowski, A., Briggs, J.A., Mortimer, J.C., Tryfona, T., Terrapon, N., Lowe, E.C., Basle, A., Morland, C., Day, A.M., Zheng, H., Rogers, T.E., Thompson, P., Hawkins, A.R., Yadav, M.P., Henrissat, B., Martens, E.C., Dupree, P., Gilbert, H.J. and Bolam, D.N. 2015. Glycan complexity dictates microbial resource allocation in the large intestine. *Nat Commun* 6:7481.

Round, J. L. and Mazmanian, S.K. 2009. The gut microbiota shapes intestinal immune responses during health and disease. *Nat Rev Immunol* 9 (5):313-323.

Ryden, P., Sugimoto-Shirasu, K., Smith, A.C., Findlay, K., Reiter, W.D. and McCann, M.C. 2003. Tensile properties of *Arabidopsis* cell walls depend on both a xyloglucan cross-linked microfibrillar network and rhamnogalacturonan II-borate complexes. *Plant Physiol* 132 (2):1033-1040.

Rye, C.S. and Withers, S.G. 2000. Glycosidase mechanisms. *Curr Opin Chem Biol* 4 (5):573-580.

Sabathé, F., Bélaïch, A. and Soucaille, P. 2002. Characterization of the cellulolytic complex (cellulosome) of *Clostridium acetobutylicum*. *FEMS Microbiol Lett* 217 (1):15-22.

Salama-Alber, O., Jobby, M.K., Chitayat, S., Smith, S.P., White, B.A., Shimon, L.J., Lamed, R., Frolow, F. and Bayer, E.A. 2013. Atypical cohesin-dockerin complex responsible for cell surface attachment of cellulosomal components: binding fidelity, promiscuity, and structural buttresses. *J Biol Chem* 288 (23):16827-16838.

Salyers, A.A., Vercellotti, J.R., West, S.E. and Wilkins, T.D. 1977. Fermentation of mucin and plant polysaccharides by strains of *Bacteroides* from the human colon. *Appl Environ Microbiol* 33 (2):319-322.

Sarma, G.N. and Karplus, P.A. 2006. In-house sulfur SAD phasing: a case study of the effects of data quality and resolution cutoffs. *Acta Crystallogr D Biol Crystallogr* 62 (Pt 7):707-716.

Schachtschneider, K.M., Yeoman, C.J., Isaacson, R.E., White, B.A., Schook, L.B. and Pieters, M. 2013. Modulation of systemic immune responses through commensal gastrointestinal microbiota. *PLoS One* 8 (1):e53969.

Sénéchal, F., Wattier, C., Rustérucci, C. and Pelloux, J. 2014. Homogalacturonan-modifying enzymes: structure, expression, and roles in plants. *J Exp Bot* 65 (18):5125-5160.

Seyedarabi, A., To, T.T., Ali, S., Hussain, S., Fries, M., Madsen, R., Clausen, M.H., Teixteira, S., Brocklehurst, K. and Pickersgill, R.W. 2010. Structural insights into substrate specificity and the anti beta-elimination mechanism of pectate lyase. *Biochemistry* 49 (3):539-546.

Shipman, J.A., Berleman, J.E. and Salyers, A.A. Characterization of four outer membrane proteins involved in binding starch to the cell surface of *Bacteroides thetaiotaomicron. J Bacteriol* 182 (19):5365-5372.

Shipman, J.A., Cho, K.H., Siegel, H.A. and Salyers, A.A. 1999. Physiological characterization of SusG, an outer membrane protein essential for starch utilization by *Bacteroides thetaiotaomicron. J Bacteriol* 181 (23):7206-7211.

Silva, I.R., Jers, C., Meyer, A.S. and Mikkelsen, J.D. 2016. Rhamnogalacturonan I modifying enzymes: an update. *N Biotechnol* 33 (1):41-54.

Skorupski, K. and Taylor, R.K. 1996. Positive selection vectors for allelic exchange. *Gene* 169 (1):47–52.

Sonnenburg, E.D., Sonnenburg, J.L., Manchester, J.K., Hansen, E.E., Chiang, H.C. and Gordon, J.I. 2006. A hybrid two-component system protein of a prominent human gut symbiont couples glycan sensing in vivo to carbohydrate metabolism. *Proc Natl Acad Sci U S A* 103 (23):8834-8839.

Sonnenburg, J.L., Xu, J., Leip, D.D., Chen, C.H., Westover, B.P., Weatherford, J., Buhler, J.D. and Gordon, J.I. 2005. Glycan foraging in vivo by an intestine-adapted bacterial symbiont. *Science* 307 (5717):1955-1959.

Soriano, M., Diaz, P. and Pastor, F.I. Pectate lyase C from *Bacillus subtilis*: a novel endo-cleaving enzyme with activity on highly methylated pectin. *Microbiology* 152 (Pt 3):617-625.

Speciale, G., Thompson, A.J., Davies, G.J. and Williams, S.J. 2014. Dissecting conformational contributions to glycosidase catalysis and inhibition. *Curr Opin Struct Biol* 28:1-13.

Spellman, M.W., McNeil, M., Darvill, A.G. and Albersheim, P. 1983. Characterization of a structurally complex heptasaccharide isolated from the pectic polysaccharide rhamnogalacturonan II *Carbohydr Res* 122 (1):131-153.

Stam, M. R., Danchin, E.G., Rancurel, C., Coutinho, P.M. and Henrissat, B. 2006. Dividing the large glycoside hydrolase family 13 into subfamilies: towards improved functional annotations of alpha-amylase-related proteins. *Protein Eng Des Sel* 19 (12):555-562.

Stevenson, T.T., Darvill, A. and Albersheim,P. 1988. 3-deoxy-d-lyxo-2-heptulosaric acid, a component of the plant cell-wall polysaccharide rhamnogalacturonan-II. *Carbohydr Res* 179:268-288.

Studier, F.W. and Moffatt, B.A. 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol* 189 (1):113-130.

Tailford, L.E., Offen, W.A., Smith, N.L., Dumon, C., Morland, C., Gratien, J., Heck, M.P., Stick, R.V., Blériot, Y., Vasella, A., Gilbert, H.J. and Davies, G.J. 2008. Structural and biochemical evidence for a boat-like transition state in beta-mannosidases. *Nat Chem Biol* 4 (5):306-312.

Taiz, L. and Zeiger, E. 2010. Cell walls: structure, biogenesis, and expansion. *In Plant physiology*: Sinauer Associates, Inc, 313-338.

Tamaru, Y., Miyake, H., Kuroda, K., Nakanishi, A., Kawade, Y., Yamamoto, K., Uemura, M., Fujita, Y., Doi, R.H. and Ueda, M. 2010. Genome sequence of the cellulosome-producing mesophilic organism *Clostridium cellulovorans* 743B. *J Bacteriol* 192 (3):901-902.

Tan, L., Eberhard, S., Pattathil, S., Warder, C., Glushka, J., Yuan, C., Hao, Z., Zhu, X., Avci, U., Miller, J.S., Baldwin, D., Pham, C., Orlando, R., Darvill, A., Hahn, M.G., Kieliszewski, M.J. and Mohnen, D. 2013. An *Arabidopsis* cell wall proteoglycan consists of pectin and arabinoxylan covalently linked to an arabinogalactan protein. *Plant Cell* 25 (1):270-287.

Tancula, E., Feldhaus, M.J., Bedzyk, L.A. and Salyers, A.A. 1992. Location and characterization of genes involved in binding of starch to the surface of *Bacteroides thetaiotaomicron. J Bacteriol* 174 (17):5609-5616.

Terrapon, N., Lombard, V., Gilbert, H.J. and Henrissat, B. 2015. Automatic prediction of polysaccharide utilization loci in Bacteroidetes species. *Bioinformatics* 31 (5):647-655.

Thompson, A.J., Cuskin, F., Spears, R.J., Dabin, J., Turkenburg, J.P., Gilbert, H.J. and Davies, G.J. 2015. Structure of the GH76 alpha-mannanase homolog, BT2949, from the gut symbiont *Bacteroides thetaiotaomicron*. *Acta Crystallogr D Biol Crystallogr* 71 (Pt 2):408-415.

Tremaroli, V. and Backhed, F. 2012. Functional interactions between the gut microbiota and host metabolism. *Nature* 489 (7415):242-249.

Turnbaugh, P.J., Ley, R.E., Hamady, M., Fraser-Liggett, C.M., Knight, R. and Gordon, J.I. 2007. The human microbiome project. *Nature* 449 (7164):804-810.

van Santen, Y., Benen, J.A., Schröter, K.H., Kalk, K.H., Armand, S., Visser, J., Dijkstra, B.W. 1999. 1.68-Å crystal structure of endopolygalacturonase II from *Aspergillus niger* and identification of active site residues by site-directed mutagenesis. *J Biol Chem* 274 (43):30474-30480.

Venditto, I., Luis, A.S., Rydahl, M., Schückel, J., Fernandes, V.O., Vidal-Melgosa, S., Bule, P., Goyal, A., Pires, V.M., Dourado, C.G., Ferreira, L.M., Coutinho, P.M., Henrissat, B., Knox, J.P., Basle, A., Najmudin, S., Gilbert, H.J., Willats, W.G. and Fontes, C.M. 2016. Complexity of the *Ruminococcus flavefaciens* cellulosome reflects an expansion in glycan recognition. *Proc Natl Acad Sci U S A* 113 (26):7136-7141.

Venditto, I., Najmudin, S., Luís, A.S., Ferreira, L.M., Sakka, K., Knox, J.P., Gilbert, H.J. and Fontes, C.M. 2015. Family 46 carbohydrate-binding modules contribute to the enzymatic hydrolysis of xyloglucan and beta-1,3-1,4-glucans through distinct mechanisms. *J Biol Chem* 290 (17):10572-10586.

Verhertbruggen, Y., Marcus, S.E., Haeger, A., Ordaz-Ortiz, J.J. and Knox, J.P. 2009. An extended set of monoclonal antibodies to pectic homogalacturonan. *Carbohydr Res* 344 (14):1858-1862.

Vidal, S., Doco, T., Williams, P., Pellerin, P., York, W.S., O'Neill, M.A., Glushka, J., Darvill, A.G. and Albersheim, P. 2000. Structural characterization of the pectic polysaccharide rhamnogalacturonan II: evidence for the backbone location of the aceric acid-containing oligoglycosyl side chain. *Carbohydr Res* 326 (4):277-294.

Vincze, T., Posfai, J. and Roberts, R.J. 2003. NEBcutter: a program to cleave DNA with restriction enzymes. *Nucleic Acids Res* 31 (13):3688-3691.

White, A. and Rose, D.R. 1997. Mechanism of catalysis by retaining β-glycosyl hydrolases. *Curr Opin Struct Biol* 7:645-651.

Whited, A.M. and Johs, A. 2015. The interactions of peripheral membrane proteins with biological membranes. Che*m Phys Lipids 19*2:51-59.

Willats, W.G., McCartney, L., Mackie, W. and Knox, J.P. 2001. Pectin: cell biology and prospects for functional analysis. *Plant Mol Biol* 47 (1-2):9-27.

Withers, S.G. 2001. Mechanisms of glycosyl transferases and hydrolases. *Carbohydr Polym* 44:325-337.

Wu, G.D., Chen, J., Hoffmann, C., Bittinger, K., Chen, Y.Y., Keilbaugh, S.A., Bewtra, M., Knights, D., Walters, W.A., Knight, R., Sinha, R., Gilroy, E., Gupta, K., Baldassano, R., Nessel, L., Li, H., Bushman, F.D. Lewis, J. 2011. Linking long-term dietary patterns with gut microbial enterotypes. *Science* 334 (6052):105-108.

Xu, J., Bjursell, M.K., Himrod, J., Deng, S., Carmichael, L.K., Chiang, H.C., Hooper, L.V. and Gordon, J.I. 2003. A genomic view of the human-*Bacteroides thetaiotaomicron* symbiosis. *Science* 299 (5615):2074-2076.

Xu, J., Chiang, H.C., Bjursell, M.K. and Gordon, J.I. 2004. Message from a human gut symbiont: sensitivity is a prerequisite for sharing. *Trends Microbiol* 12 (1):21-28.

Xu, J., Mahowald, M.A., Ley, R.E., Lozupone, C.A., Hamady, M., Martens, E.C., Henrissat, B., Coutinho, P.M., Minx, P., Latreille, P., Cordum, H., Van Brunt, A., Kim, K., Fulton, R.S., Fulton, L.A., Clifton, S.W., Wilson, R.K., Knight, R.D. and Gordon, J.I. 2007. Evolution of symbiotic bacteria in the distal human intestine. *PLoS Biol* 5 (7):e156.

York, W.S., Darvill, A., McNeil, M. and Albersheim, P. 1985. 3-deoxy-D-manno-2-octulosonic acid (KDO) is a component of rhamnogalacturonan II, a pectic polysaccharide in the primary cell walls of plants. *Carbohydr Res* 138 (1):109-126.

Ze, X., Le Mougen, F., Duncan, S.H., Louis, P. and Flint, H.J. 2013. Some are more equal than others: the role of "keystone" species in the degradation of recalcitrant. *Gut Microbes* 4 (3):236-240.

Zhang, X. 2015. Insights into the enzymatic degradation of mannan (*in situ*) and pectin utilization by *Bacteroides thetaiotaomicron*, Institute for Cell and Molecular Biosciences, Newcastle University.

Zhu, Y., Suits, M.D., Thompson, A.J., Chavan, S., Dinev, Z., Dumon, C., Smith, N., Moremen, K.W., Xiang, Y., Siriwardena, A., Williams, S.J., Gilbert, H.J. and Davies, G.J. 2010. Mechanistic insights into a Ca2+-dependent family of alpha-mannosidases in a human gut symbiont. *Nat Chem Biol* 6 (2):125-132.

Zuckert, W.R. 2014. Secretion of bacterial lipoproteins: through the cytoplasmic membrane, the periplasm and beyond. *Biochim Biophys Acta* 1843 (8):1509-1516.

Appendix A

Vectors used in this study showing the different multiple cloning sites.

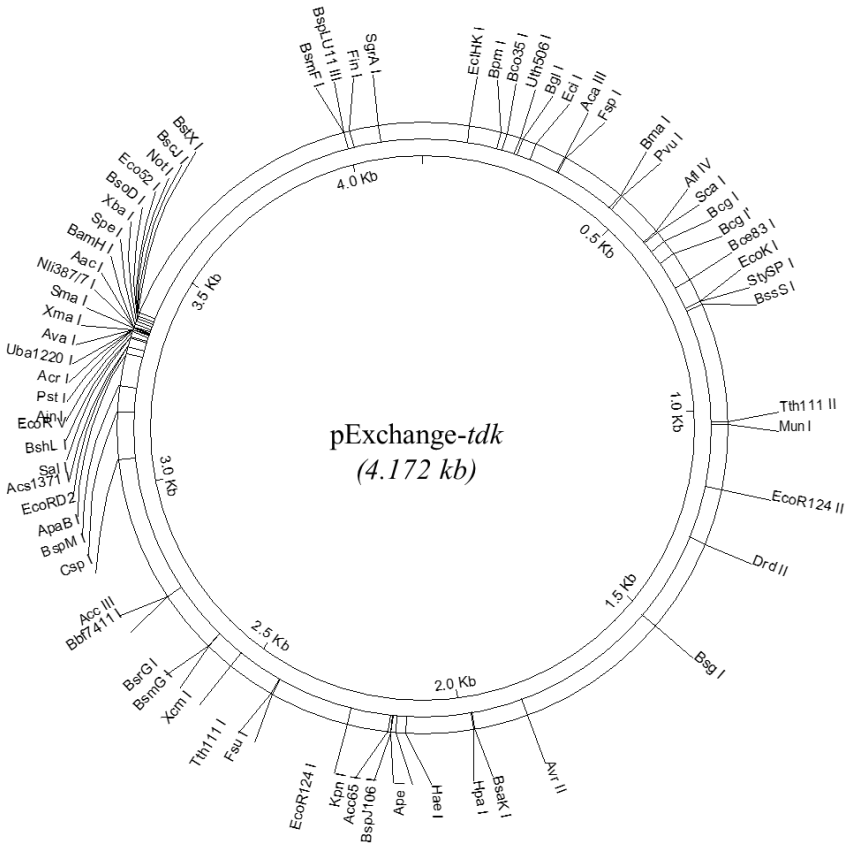
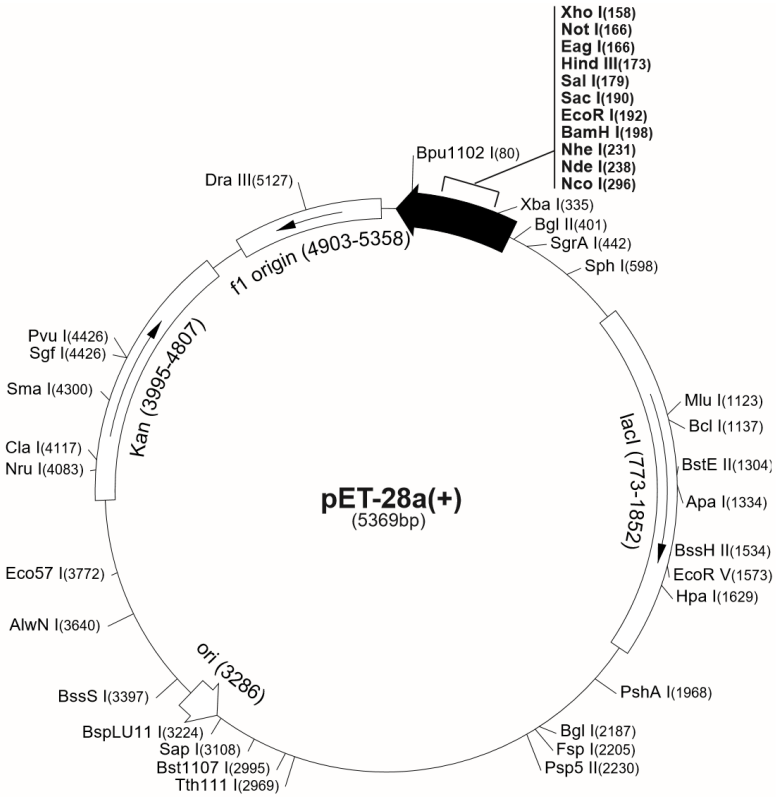


Figure A.1 Vector pExchange-tdk



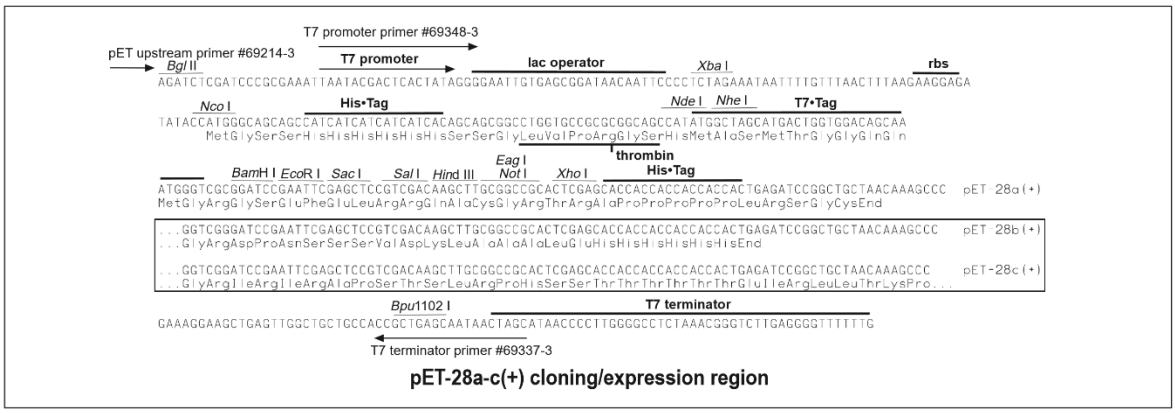


Figure A.2 pET28a

Figure adapted from Novagen.

Appendix B

Chemically synthesised protein sequences mentioned in this thesis.

$CBM77_{PL1/9}$_$CBM77_{PL1/9}$

MGSSHHHHHHSSGLVPRGSHMASPVAGNYVHDFTANGTSSSFYTIAGNLSTSKGTATYNGKTLTQCLKMETATSIS
FTAPSAGKLTLVFAEAAATAKVDGNKVTASNGIITVDLAQGAHTITKADACNLFYMEYAAPSTPVTTQQGQPVVTT
ASQPTQTTSSQQPAQTTSSSQQQTPVAGNYVHDFTANGTSSSFYTIAGNLSTSKGTATYNGKTLTQCLKMETATSI
SFTAPSAGKLTLVFAEAAATAKVDGNKVTASNGIITVDLAQGAHTITKADACNLFYMEYAA

Figure B.1 $CBM77_{PL1/9}$_$CBMPL_{1/9}$ schematic representation and protein sequence

A schematic representation of the dimer protein is presented at the top and below the respective protein sequence is showed. The CBM77 modules are represented as a green box, the yellow lines indicate the linker sequences and the N-terminal His_6-Tag is shown in purple. In black are represented the initial amino acid sequence encoded when $CBM77_{PL1/9}$_$CBMPL_{1/9}$ was cloned into pET28a utilizing the restriction sites *NheI*/*XhoI*.

GNSFEWGQVPQQPDLSWADSVGSRQLPGDKMIISANSFGAVADSTVLSTEAIQKAIDSCAVS
GGGTVILQPGYYQTGALFVKSGVNLQIGKGVTLLASPDIHHYPEFRSRIAGIEMTWPAAVIN
IVDEKNAAVSGEGTLDCRGKVFWDKYWEMRKEYVAKGLRWIVDYDCKRVRGILVERSSDITL
SGFTLMRTGFWGCQILYSDYCTIDGLTINNNIGGHGPSTDGIDIDSSCNILIENCDVDCNDD
NICIKSGRDADGLRVNRPTENVVVRNCTARKGAGLITCGSETSGSIRNILGYDLKAVGTSTV
LRLKSAMNRGGTIENIYMTRVSAENISQVLAADLNWNPSYSYSTLPKEYEGKEIPEHWKVML
TPVEPPEKGYPRFRNVYLSQVKAENVDEFISASGWNDTLRLENFYLYGIEAQTNKPGKICYT
RNFNLSGITLDVKDRDAIELKENEQSNIHFDYVKTTTDRRTAGNLPH

Figure B.2 Sequence of BT4149 homologue

Bf4149 protein sequence from *Bacteroides faecis* CAG:32 (accession code WP_010536731.)

Appendix C

Table C.1 Data collection and refinement statistics for $CBM77_{PL1/9}$

Data collection	**$CBM77_{PL1/9}$**	**$CBM77_{PL1/9}$**
Ligand	apo	apo
Source	Micromax 007, Rigaku	Diamond I04-1
Space group	$P6_122$	$P6_122$
Cell dimensions		
a, *b*, *c* (Å)	69.8, 69.8, 122.7	69.9, 69.9, 123.1
α, β, γ (°)	90, 90, 120	90, 90, 120
No. of measured reflections	1619299 (104615)	308837 (15262)
No. of independent reflections	13949 (907)	29090 (1400)
Resolution (Å)	43.05-1.93 (1.98-1.93)	43.16-1.5 (1.53-1.50)
R_{merge} (%)	6.2 (25.1)	6.1(119.6)
$CC_{1/2}$	1.00 (1.00)	1.00 (8.29)
$I/\sigma I$	84.4 (27.5)	18.3 (2.3)
Completeness (%)	100.0 (100.0)	99.8 (99.2)
Redundancy	116.1 (115.3)	10.6 (10.9)
Anomalous completeness	100.0 (100.0)	
Anomalous redundancy	64.3 (61.9)	
Refinement		
R_{work}/ R_{free}		18.62 / 22.12
No. atoms		
Protein		887
Ligand		
Water		116
B-factors		
Protein		22
Ligand		
Water		35
R.m.s deviations		
Bond lengths (Å)		0.014
Bond angles (º)		1.6
PDB code		5FU5

Highest resolution shell is shown in parenthesis.

Table C.2 Data collection and refinement statistics for BT0986

Data collection	BT0986 SeMet	BT0986	BT0986 Ligand
Date	25/07/15	24/05/15	06/03/16
Source	I04	I02	I02
Wavelength (Å)	0.979	0.913	0.979
Space group	$P2_1$	$P2_1$	$P2_1$
Cell dimensions			
a, *b*, *c* (Å)	64.4, 86.3, 121.0	64.3, 85.8, 120.9	64.0, 84.8, 120.6
α, β, γ (°)	90.0, 99.3, 90.0	90.0, 98.8, 90.0	90.0, 99.2, 90.0
No. of measured reflections	441797 (11837)	380421 (16017)	262988 (20600)
No. of independent reflections	37160 (3723)	101961 (4955)	75602 (5395)
Resolution (Å)	49.43-2.65 (2.77-2.65)	49.17-1.90 (1.93-1.90)	50.69-2.11 (2.16-2.11)
R_{merge} (%)	8.6 (24.6)	7.6 (65.3)	8.8 (81.8)
$CC_{1/2}$	0.997 (0.897)	0.996 (0.671)	0.995 (0.548)
$I/\sigma I$	22.7(4.1)	9.7 (1.7)	9.9 (1.5)
Completeness (%)		99.7 (99.1)	98.9 (99.8)
Redundancy		3.7 (3.2)	3.6 (3.8)
Anomalous Completeness (%)	85.2 (6.7)		
Anomalous Redundancy	5.7 (2.5)		
Refinement			
R_{work}/ R_{free}		17.39/20.51	19.86/24.03
No. atoms			
Protein		9278	8280
Ligand/Ions		3	14
Water		968	315
B-factors			
Protein		30.1	41.4
Ligand/Ions		33.7	49.2
Water		35.4	40.5
R.m.s deviations			
Bond lengths (Å)		0.012	0.011
Bond angles (º)		1.52	1.50
PDB code		5MQN	5MQM

Highest resolution shell is shown in parenthesis.

Table C.3 Data collection and refinement statistics for BT1003

Data collection	BT1003
Date	05/07/15
Source	I03
Wavelength (Å)	0.775
Space group	H32
Cell dimensions	
a, b, c (Å)	219.4, 219.4, 91.8
α, β, γ (°)	90.0, 90.0, 120.0
No. of measured reflections	143341 (16048)
No. of independent reflections	29265 (3300)
Resolution (Å)	109.72-2.50 (2.60-2.50)
R_{merge} (%)	13.3 (95.4)
$CC_{1/2}$	0.992 (0.339)
$I/\sigma I$	5.2 (1.5)
Completeness (%)	100.0 (100.0)
Redundancy	4.9 (4.9)
Anomalous Completeness (%)	
Anomalous Redundancy	
Refinement	
R_{work}/ R_{free}	19.41/23.90
No. atoms	
Protein	4792
Ligand/Ions	
Water	
B-factors	
Protein	67.9
Ligand/Ions	
Water	
R.m.s deviations	
Bond lengths (Å)	0.012
Bond angles (º)	1.59
PDB code	5MQO

Highest resolution shell is shown in parenthesis.

Table C.4 Data collection and refinement statistics for BT4170

Data collection	BT4170*	BT4170 Ligand 1	BT4170 Ligand 2
Date	02/07/14	12/12/14	16/10/15
Source	I04-1	I04	I04
Wavelength (Å)	0.9200	0.9794	0.9791
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Cell dimensions			
a, *b*, *c* (Å)	77.5, 123.8, 138.4	77.5, 123.9, 137.6	45.4, 76.6, 122.8
α, β, γ (°)	90	90	90
No. of measured reflections	1547996 (98909)	1712463 (15540)	155355 (11830)
No. of independent reflections	22195 (16250)	536956 (9262)	22486 (1650)
Resolution (Å)	39.65-1.48 (1.52 – 1.48)	48.4-1.07 (1.09 - 1.07)	61.7-2.20 (2.26 - 2.20)
R_{merge} (%)	7.8 (61.5)	4.9 (73.5)	9.5 (120)
$CC_{1/2}$	0.999 (0.781)	0.998 (0.378)	0.998 (0.553)
$I/\sigma I$	13.4 (2.8)	11.3 (0.9)	10.7 (1.6)
Completeness (%)	99.9 (99.9)	94.3 (33.1)	99.8 (99.5)
Redundancy	7.0 (6.1)	3.2 (1.7)	6.9 (7.2)
Anomalous Completeness (%)			
Anomalous Redundancy			
Refinement			
R_{work}/ R_{free}	15.84/18.57	14.46/16.62	24.11 / 30.77
No. atoms			
Protein	9717	10035	3257
Ligand/Ions	18	126	69
Water	1526	1741	3
B-factors			
Protein	15.0	10.7	54.2
Ligand/Ions	20.8	12.1	49.3
Water	26.0	24.8	33.2
R.m.s deviations			
Bond lengths (Å)	0.015	0.011	0.013
Bond angles (º)	1.68	1.52	1.67
PDB code	X*	X*	X*

Highest resolution shell is shown in parenthesis. (1) L-Rha-α1,4-D-GalA-α1,2-L-Rha (2) L-Rha-α1,4-D-GalA-α1,2-L-Rha and Δ4,5-GalA-α1,2-L-Rha-α1,4-D-GalA. * Not deposited in PBD database.

Appendix D

Primers used to generate mutants by site direct mutagenesis

Table D.1 Primers designed to generate the $CBM77_{PL1/9}$ mutants

Mutant	Primer Name	Primer sequence
K1092A	K1092A_Fw	ggc aat ctt tca act tca gcg gga aca gct aca tac aac
	K1092A_Rv	gtt gta tgt agc tgt tcc cgc tga agt tga aag att gcc
Y1097A	Y1097A_Fw	caa agg gaa cag cta cag cca acg gca aga ctc tta c
	Y1097A_Rv	gta aga gtc ttg ccg ttg gct gta gct gtt ccc ttt g
K1100A	K1100A_Fw	gct aca tac aac ggc gcg act ctt aca cag tgc c
	K1100A_Rv	ggc act gtg taa gag tcg cgc cgt tgt atg tag c
K1107A	K1107A_Fw	ctc tta cac agt gcc tgg cga tgg aga cag cta cat c
	K1107A_Rv	gat gta gct gtc tcc atc gcc agg cac tgt gta aga g
K1123A	K1123A_Fw	gca cct tca gca ggc gcg ctt act ctg gta ttc
	K1123A_Rv	gaa tac cag agt aag cgc gcc tgc tga agg tgc
K1136A	K1136A_Fw	gca gca gct aca gct gcg gtt gac ggc aac aag
	K1136A_Rv	ctt gtt gcc gtc aac cgc agc tgt agc tgc tgc
K1141A	K1141A_Fw	cta agg ttg acg gca acg cgg taa cag ctt caa acg g
	K1141A_Rv	ccg ttt gaa gct gtt acc gcg ttg ccg tca acc tta g
K1162A	K1162A_Fw	gca cac act atc acc gcg gca gat gcc tgc aac
	K1162A_Rv	gtt gca ggc atc tgc cgc ggt gat agt gtg tgc
Y1170A	Y1170A_Fw	cct gtt cta cat gga agc tgc agc aca cca cc
	Y1170A_Rv	ggt gtg ctg cag ctt cca tgt aga aca gg

Fw (forward primer); Rv (reverse primer).

Table D.2 Primers designed to generate the BT1003 mutants

Mutant	Primer Name	Primer sequence
H296A	H296AFw	caa tat cgt gca atg gga gct gcc gta cgt gcc aat tac
	H296ARv	gta att ggc acg tac ggc agc tcc cat tgc acg ata ttg
Q365A	Q365AFw	agc att cag aaa gtg cat gcg agc tat ggt cgt ccg tat
	Q365ARv	ata cgg acg acc ata gct cgc atg cac ttt ctg aat gct
Q365E	Q365EFw	cat tca gaa agt gca tga gag cta tgg tcg tcc gta t
	Q365ERv	ata cgg acg acc ata gct ctc atg cac ttt ctg aat g
E381A	E381AFw	cag tac ggc aca taa tgc aac ctg tgc caa cat ag
	E381ARv	cta tgt tgg cac agg ttg cat tat gtg ccg tac tg
E381Q	E381QFw	cag tac ggc aca taa tca aac ctg tgc caa cat ag
	E381QRv	cta tgt tgg cac agg ttt gat tat gtg ccg tac tg
C383A	C383AFw	tac ggc aca taa tga aac cgc tgc caa cat agg taa tat gc
	C383ARv	gca tat tac cta tgt tgg cag cgg ttt cat tat gtg ccg ta
C455A	C455AFw	gga cgg aat ata tca gtg ctt tct gtt gtc cgc cca ata cat t
	C455ARv	aat gta ttg ggc gga caa cag aaa gca ctg ata tat tcc gtc c
C457A	C457AFw	cgg aat ata tca gtt gtt tcg ctt gtc cgc cca ata cat tgc g
	C457ARv	cgc aat gta ttg ggc gga caa gcg aaa caa ctg ata tat tcc g
C457S	C457SFw	cgg aat ata tca gtt gtt tct ctt gtc cgc cca ata cat tgc g
	C457S Rv	cgc aat gta ttg ggc gga caa gag aaa caa ctg ata tat tcc g
C458A	C458AFw	aat ata tca gtt gtt tct gtg ctc cgc cca ata cat tgc gta c
	C458ARv	gta cgc aat gta ttg ggc gga gca cag aaa caa ctg ata tat t

Fw (forward primer); Rv (reverse primer)

Table D.3 Primers designed to generate the BT4145 mutants

Mutant	Primer Name	Primer sequence
D250A	D250A_Fw	ccg cat aca ttc ttc aac gcc tct tac gaa gtc tat cag
	D250A_Rv	ctg ata gac ttc gta aga ggc gtt gaa gaa tgt atg cgg
E253A	E253A_Fw	ctt caa cga ctc tta cgc agt cta tca ggc aga ct
	E253A_Rv	agt ctg cct gat aga ctg cgt aag agt cgt tga ag
E349A	E349A_Fw	gac atc ccc gaa tgt gca gga ttc ggt ttg tcc
	E349A_Rv	gga caa acc gaa tcc tgc aca ttc ggg gat gtc
E395A	E395A_Fw	aaa cct tat act tct acc gca aca ttt acc tgg ctg acg
	E395A_Rv	cgt cag cca ggt aaa tgt tgc ggt aga agt ata agg ttt
Q328A	Q328A_Fw	agc atc acc cgt aat gcg gca cac gga tcg cct
	Q328A_Rv	agg cga tcc gtg tgc cgc att acg ggt gat gct

Fw (forward primer); Rv (reverse primer).

Table D.4 Primers designed to generate the BT4170 mutants

Mutant	Primer Name	Primer sequence
R125A	R125A_Fw	gca ggt aaa acc gaa aga tca ggc tat tac ggt gtt tta tgt tac ag
	R125A_Rv	ctg taa cat aaa aca ccg taa tag cct gat ctt tcg gtt tta cct gc
H153A	H153A_Fw	gca agt gac tat tac cgg tgc tac gca atc aga gtg ttt cc
	H153A_Rv	gga aac act ctg att gcg tag cac cgg taa tag tca ctt gc
Q155A	Q155A_Fw	gac tat tac cgg tca tac ggc atc aga gtg ttt ccg tat tg
	Q155A_Rv	caa tac gga aac act ctg atg ccg tat gac cgg taa tag tc
Y184A	Y184A_Fw	gat ggt atg gca att ggt ttc gct ttg ctt gga gga agc aat aat c
	Y184A_Rv	gat tat tgc ttc ctc caa gca aag cga aac caa ttg cca tac cat c
Y184F	Y184F_Fw	gat ggt atg gca att ggt ttc ttt ttg ctt gga gga agc aat aat c
	Y184F_Rv	gat tat tgc ttc ctc caa gca aaa aga aac caa ttg cca tac cat c
G212A	G212A_Fw	cga agg tgg taa ggg ggc caa tgt tga tgg ttt tg
	G212A_Rv	caa aac cat caa cat tgg ccc cct tac cac ctt cg
N213A	N213A_Fw	aag gtg gta agg ggg gcg ctg ttg atg gtt ttg gcg
	N213A_Rv	cgc caa aac cat caa cag cgc ccc cct tac cac ctt
D215A	D215A_Fw	aag ggg ggc aat gtt gct ggt ttt ggc ggt cac
	D215A_Rv	gtg acc gcc aaa acc agc aac att gcc ccc ctt
H220A	H220A_Fw	gtt gat ggt ttt ggc ggt gcc att aac tca tct tca gtt gg
	H220A_Rv	cca act gaa gat gag tta atg gca ccg cca aaa cca tca ac
D246A	D246A_Fw	cgc gca tgg tac aat agt gct gat ggt ttc gat ttg atc
	D246A_Rv	gat caa atc gaa acc atc agc act att gta cca tgc gcg
D247A	D247A_Fw	gcg cat ggt aca ata gtg atg ctg gtt tcg att tga tca att g
	D247A_Rv	caa ttg atc aaa tcg aaa cca gca tca cta ttg tac cat gcg c
D250A	D250A_Fw	caa tag tga tga tgg ttt cgc ttt gat caa ttg ctt cga gg
	D250A_Rv	cct cga agc aat tga tca aag cga aac cat cat cac tat tg
I252A	I252A_Fw	gat gat ggt ttc gat ttg gcc aat tgc ttc gag gct gtg
	I252A_Rv	cac agc ctc gaa gca att ggc caa atc gaa acc atc atc
N253A	N253A_Fw	gat ggt ttc gat ttg atc gct tgc ttc gag gct gtg aag
	N253A_Rv	ctt cac agc ctc gaa gca agc gat caa atc gaa acc atc
D280A	D280A_Fw	caa aga agt tgc tgg tgc cgg tac cgg ttt taa agc
	D280A_Rv	gct tta aaa ccg gta ccg gca cca gca act tct ttg
K285A	K285A_Fw	ggt gac ggt acc ggt ttt gca gcc gga ggc tat gga at
	K285A_Rv	att cca tag cct ccg gct gca aaa ccg gta ccg tca cc
L318A	L318A_Fw	gtt tgg ctt att aca ata gag cga gag gat ttt atg cca atc atc
	L318A_Rv	gat gat tgg cat aaa atc ctc tcg ctc tat tgt aat aag cca aac
R319A	R319A_Fw	ggc tta tta caa tag att ggc agg att tta tgc caa tca tc
	R319A_Rv	gat gat tgg cat aaa atc ctg cca atc tat tgt aat aag cc
Y322A	Y322A_Fw	caa tag att gag agg att tgc tgc caa tca tca ttt agg tgg
	Y322A_Rv	cca cct aaa tga tga ttg gca gca aat cct ctc aat cta ttg
Y322F	Y322F_Fw	caa tag att gag agg att ttt tgc caa tca tca ttt agg tgg
	Y322F_Rv	cca cct aaa tga tga ttg gca aaa aat cct ctc aat cta ttg
N324A	N324A_Fw	gat tga gag gat ttt atg ccg ctc atc att tag gtg gaa tta
	N324A_Rv	taa ttc cac cta aat gat gag cgg cat aaa atc ctc tca atc
H325A	H325A_Fw	gag agg att tta tgc caa tgc tca ttt agg tgg aat tat att tg
	H325A_Rv	caa ata taa ttc cac cta aat gag cat tgg cat aaa atc ctc tc

Fw (forward primer); Rv (reverse primer).

Appendix E

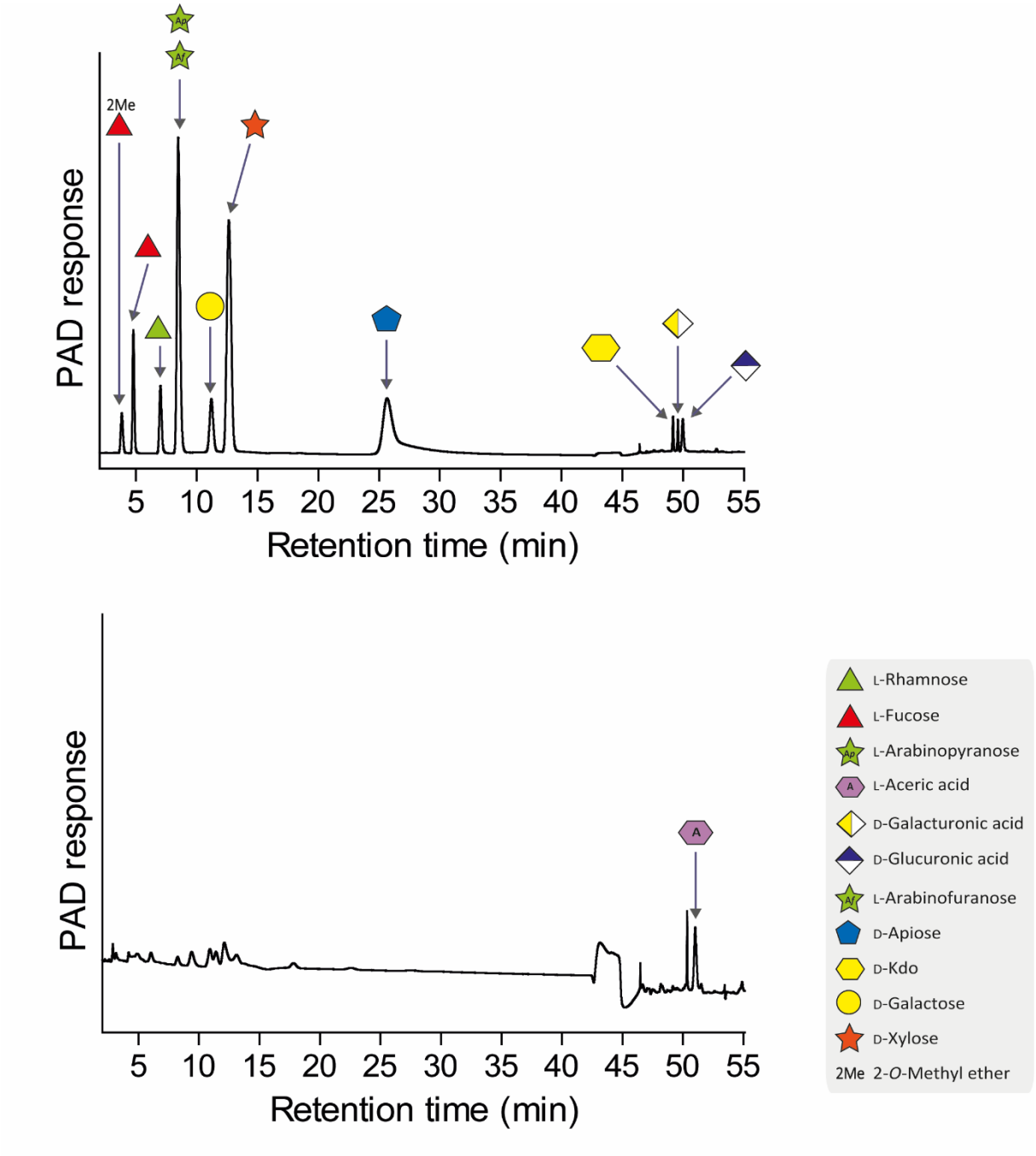


Figure E.1 HPAEC chromatogram for standards

Chromatogram showing the peaks of the different standards utilized to identify the reaction products in Chapter 4.

Appendix F

Schematic representation of BT1003 and BT0986 structures topology determined using the online software PDBsum (de Beer *et al*. 2014).

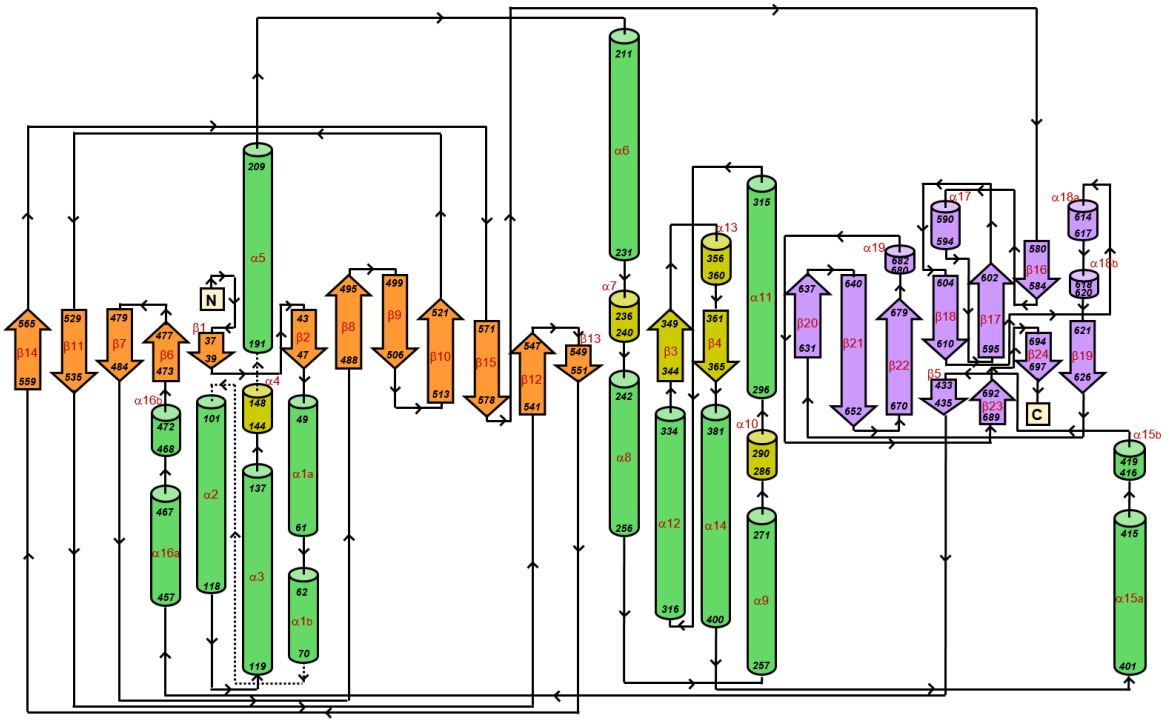


Figure F.1 BT1003 topology

The interrupted line represents the missing structural data. A domain is represented in green, B domain is shown in orange and C-domain is coloured in purple. The α-helixes (cylinders) and β-strands (arrows) present connecting the $(\alpha/\alpha)_6$ barrel elements are coloured in lemon. The β-strands and α-helixes are numbered conforming to the BT1003 sequence from the N-terminal (β1 or α1) to the C-terminal (β24 or α19).

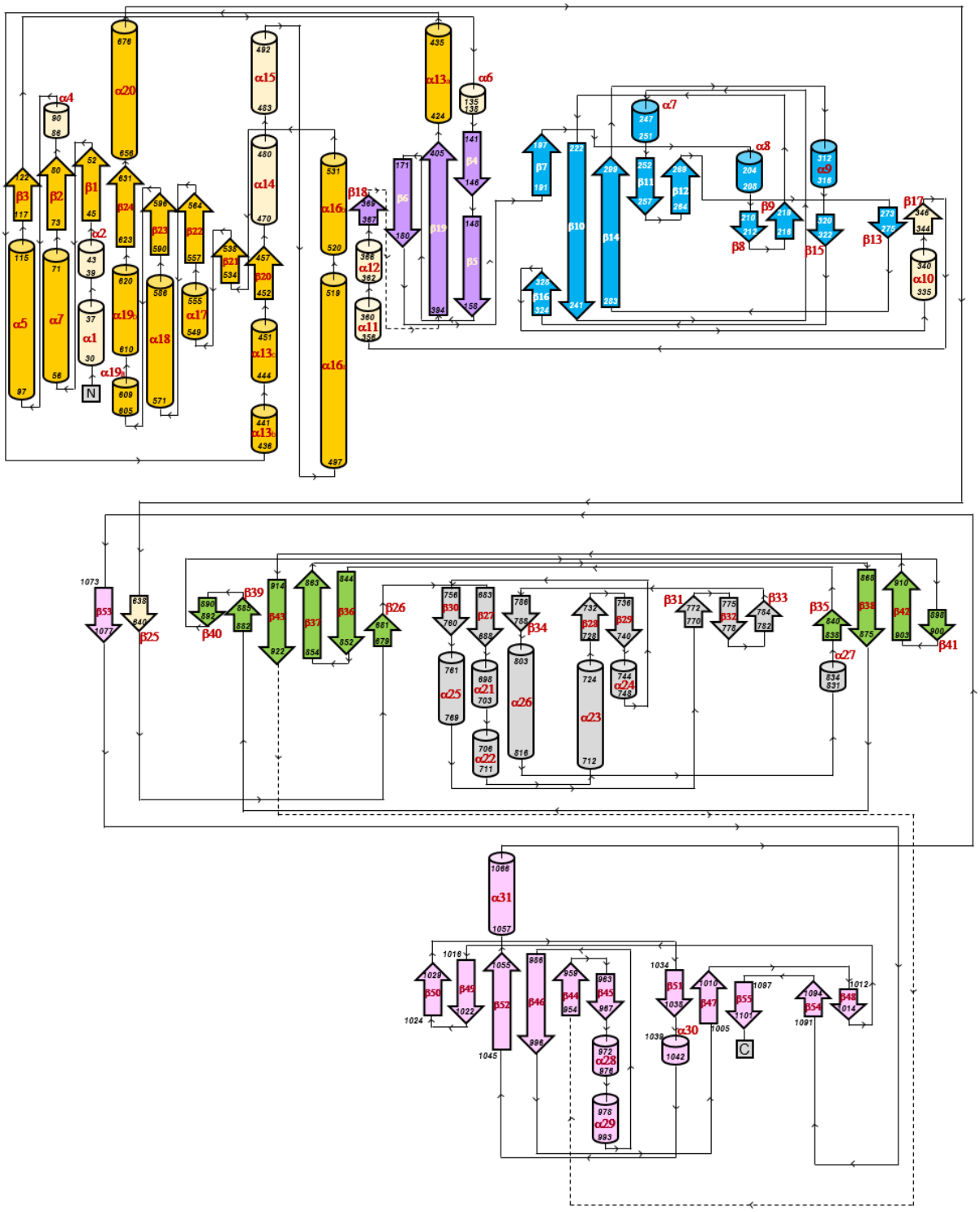


Figure F.2 BT0986 topology

The different domains are represented in different colours: A (yellow), B (purple), C (blue), D (grey), E (green) and F (pink). The β-strands (arrows) and α-helixes (cylinders) are numbered conforming to the BT0986 sequence from the N-terminal (β1 or α1) to the C-terminal (β55 or α31). The α-helixes and β-strands that form the $(\beta/\alpha)_8$ -barrel are coloured in dark yellow. The interrupted lines represent the missing structural data.

Appendix G

List of interactions protein-ligand or protein-ion described in this thesis. The list of interactions was generated using the online tool PDBsum (de Beer *et al*. 2014).

List of protein-ligand interactions

Protein BT0986 Ligand Tet

Hydrogen bonds

	<----- A T O M 1 ----->					<----- A T O M 2 ----->					
	Atom no.	Atom name	Res name	Res no.	Chain		Atom no.	Atom name	Res name	Res no.	Chain	Distance
1.	178	OH	TYR	49	A	-->	8284	O4	Tet	1	C	2.72
2.	3340	OG	SER	459	A	-->	8286	O2	Tet	1	C	2.81
3.	3363	OE2	GLU	461	A	<--	8288	N4	Tet	1	C	3.09
4.	4685	NE2	HIS	627	A	-->	8284	O4	Tet	1	C	2.71

Non-bonded contacts

	<----- A T O M 1 ----->					<----- A T O M 2 ----->					
	Atom no.	Atom name	Res name	Res no.	Chain		Atom no.	Atom name	Res name	Res no.	Chain	Distance
1.	179	CE2	TYR	49	A	---	8284	O4	Tet	1	C	3.38
2.	177	CZ	TYR	49	A	---	8284	O4	Tet	1	C	3.46
3.	178	OH	TYR	49	A	---	8284	O4	Tet	1	C	2.72
4.	777	CG	PHE	126	A	---	8282	C3	Tet	1	C	3.89
5.	778	CD1	PHE	126	A	---	8285	C2	Tet	1	C	3.66
6.	782	CD2	PHE	126	A	---	8284	O4	Tet	1	C	3.82
7.	779	CE1	PHE	126	A	---	8285	C2	Tet	1	C	3.86
8.	781	CE2	PHE	126	A	---	8292	C5	Tet	1	C	3.81
9.	780	CZ	PHE	126	A	---	8292	C5	Tet	1	C	3.82
10.	3332	CG	ASP	458	A	---	8281	O3	Tet	1	C	3.74
11.	3333	OD1	ASP	458	A	---	8281	O3	Tet	1	C	3.90
12.	3334	OD2	ASP	458	A	---	8281	O3	Tet	1	C	2.94
13.	3334	OD2	ASP	458	A	---	8282	C3	Tet	1	C	3.75
14.	3334	OD2	ASP	458	A	---	8285	C2	Tet	1	C	3.84
15.	3334	OD2	ASP	458	A	---	8286	O2	Tet	1	C	3.03
16.	3340	OG	SER	459	A	---	8286	O2	Tet	1	C	2.81
17.	3363	OE2	GLU	461	A	---	8285	C2	Tet	1	C	3.44
18.	3363	OE2	GLU	461	A	---	8286	O2	Tet	1	C	3.39
19.	3363	OE2	GLU	461	A	---	8287	C1	Tet	1	C	3.42
20.	3363	OE2	GLU	461	A	---	8288	N4	Tet	1	C	3.09
21.	4177	CZ2	TRP	563	A	---	8289	N3	Tet	1	C	3.54
22.	4177	CZ2	TRP	563	A	---	8290	N2	Tet	1	C	3.55
23.	4176	CH2	TRP	563	A	---	8289	N3	Tet	1	C	3.80
24.	4176	CH2	TRP	563	A	---	8290	N2	Tet	1	C	3.46
25.	4240	NZ	LYS	571	A	---	8288	N4	Tet	1	C	3.66
26.	4403	CD	GLU	593	A	---	8281	O3	Tet	1	C	3.45
27.	4404	OE1	GLU	593	A	---	8281	O3	Tet	1	C	3.02
28.	4405	OE2	GLU	593	A	---	8281	O3	Tet	1	C	3.38
29.	4686	CD2	HIS	627	A	---	8284	O4	Tet	1	C	3.43
30.	4684	CE1	HIS	627	A	---	8284	O4	Tet	1	C	3.80
31.	4685	NE2	HIS	627	A	---	8283	C4	Tet	1	C	3.48
32.	4685	NE2	HIS	627	A	---	8284	O4	Tet	1	C	2.71
33.	4685	NE2	HIS	627	A	---	8293	C6	Tet	1	C	3.87

Number of hydrogen bonds: 4

Number of non-bonded contacts: 33

Figure G.1 List of interaction between BT0986 and the ligand D-rhamnopyranose tetrazole

Amino acids numbered conforming to the BT0986 sequence. (Tet) D-rhamnopyranose tetrazole

List of protein-metal interactions

Protein BT0986 Metal Calcium

Hydrogen bonds

	<----- A T O M 1 ----->					<----- A T O M 2 ----->					
	Atom no.	Atom name	Res name	Res no.	Chain		Atom no.	Atom name	Res name	Res no.	Chain	Distance
1.	3334	OD2	ASP	458	A	<->	8278	CA	CA	1	B	2.47
2.	3340	OG	SER	459	A	<->	8278	CA	CA	1	B	2.62
3.	3979	OE1	GLU	538	A	<->	8278	CA	CA	1	B	2.42
4.	4151	OE1	GLU	561	A	<->	8278	CA	CA	1	B	2.79
5.	4404	OE1	GLU	593	A	<->	8278	CA	CA	1	B	2.50

Non-bonded contacts

	<----- A T O M 1 ----->					<----- A T O M 2 ----->					
	Atom no.	Atom name	Res name	Res no.	Chain		Atom no.	Atom name	Res name	Res no.	Chain	Distance
1.	3332	CG	ASP	458	A	:-:	8278	CA	CA	1	B	3.56
2.	3334	OD2	ASP	458	A	:-:	8278	CA	CA	1	B	2.47
3.	3339	CB	SER	459	A	:-:	8278	CA	CA	1	B	3.84
4.	3340	OG	SER	459	A	:-:	8278	CA	CA	1	B	2.62
5.	3978	CD	GLU	538	A	:-:	8278	CA	CA	1	B	3.42
6.	3979	OE1	GLU	538	A	:-:	8278	CA	CA	1	B	2.42
7.	3980	OE2	GLU	538	A	:-:	8278	CA	CA	1	B	3.68
8.	4150	CD	GLU	561	A	:-:	8278	CA	CA	1	B	3.59
9.	4151	OE1	GLU	561	A	:-:	8278	CA	CA	1	B	2.79
10.	4403	CD	GLU	593	A	:-:	8278	CA	CA	1	B	3.40
11.	4404	OE1	GLU	593	A	:-:	8278	CA	CA	1	B	2.50
12.	4405	OE2	GLU	593	A	:-:	8278	CA	CA	1	B	3.59

Number of hydrogen bonds: 5

Number of non-bonded contacts: 12

Figure G.2 List of interaction between BT0986 and the calcium ion

Amino acids numbered conforming to the BT0986 sequence. (Ca) calcium ion.

List of protein-ligand interactions

Protein:BT4170 Ligand Rha-GalA-Rha (negative subsites)

Hydrogen bonds

	<----- A T O M 1 ----->					<----- A T O M 2 ----->					
	Atom no.	Atom name	Res name	Res no.	Chain		Atom no.	Atom name	Res name	Res no.	Chain	Distance
1.	780	NH1	ARG	125	A	-->	3274	O62	GalA	2	C	2.91
2.	781	NH2	ARG	125	A	-->	3272	O61	GalA	2	C	3.21
3.	995	NE2	HIS	153	A	-->	3272	O61	GalA	2	C	2.58
4.	1011	OE1	GLN	155	A	<--	3268	O3	Rha	3	C	2.68
5.	1011	OE1	GLN	155	A	<--	3271	O4	Rha	3	C	2.70

Non-bonded contacts

	<----- A T O M 1 ----->					<----- A T O M 2 ----->					
	Atom no.	Atom name	Res name	Res no.	Chain		Atom no.	Atom name	Res name	Res no.	Chain	Distance
1.	779	CZ	ARG	125	A	---	3272	O61	GalA	2	C	3.77
2.	780	NH1	ARG	125	A	---	3272	O61	GalA	2	C	3.42
3.	780	NH1	ARG	125	A	---	3273	C6	GalA	2	C	3.54
4.	780	NH1	ARG	125	A	---	3274	O62	GalA	2	C	2.91
5.	781	NH2	ARG	125	A	---	3272	O61	GalA	2	C	3.21
6.	996	CD2	HIS	153	A	---	3272	O61	GalA	2	C	3.52
7.	994	CE1	HIS	153	A	---	3268	O3	Rha	3	C	3.26
8.	994	CE1	HIS	153	A	---	3272	O61	GalA	2	C	3.54
9.	994	CE1	HIS	153	A	---	3276	O5	GalA	2	C	3.88
10.	995	NE2	HIS	153	A	---	3268	O3	Rha	3	C	3.72
11.	995	NE2	HIS	153	A	---	3272	O61	GalA	2	C	2.58
12.	995	NE2	HIS	153	A	---	3273	C6	GalA	2	C	3.73
13.	995	NE2	HIS	153	A	---	3276	O5	GalA	2	C	3.53
14.	1010	CD	GLN	155	A	---	3268	O3	Rha	3	C	3.81
15.	1010	CD	GLN	155	A	---	3271	O4	Rha	3	C	3.57
16.	1011	OE1	GLN	155	A	---	3266	C4	Rha	3	C	3.48
17.	1011	OE1	GLN	155	A	---	3267	C3	Rha	3	C	3.47
18.	1011	OE1	GLN	155	A	---	3268	O3	Rha	3	C	2.68
19.	1011	OE1	GLN	155	A	---	3271	O4	Rha	3	C	2.70
20.	1012	NE2	GLN	155	A	---	3271	O4	Rha	3	C	3.88
21.	1241	CE1	TYR	184	A	---	3271	O4	Rha	3	C	3.51
22.	1242	CZ	TYR	184	A	---	3271	O4	Rha	3	C	3.88
23.	1243	OH	TYR	184	A	---	3266	C4	Rha	3	C	3.53
24.	1243	OH	TYR	184	A	---	3271	O4	Rha	3	C	3.50
25.	1452	OD1	ASN	213	A	---	3267	C3	Rha	3	C	3.85
26.	1501	CD2	HIS	220	A	---	3265	C6	Rha	3	C	3.65
27.	1499	CE1	HIS	220	A	---	3265	C6	Rha	3	C	3.46
28.	1500	NE2	HIS	220	A	---	3265	C6	Rha	3	C	3.10
29.	1747	ND2	ASN	253	A	---	3283	O2	GalA	2	C	3.87

Number of hydrogen bonds: 5

Number of non-bonded contacts: 29

Figure G.3 List of interaction between BT4170 and the α-L-Rha-α1,4-D-GalA-α1,2-L-Rha

Amino acids numbered conforming to the BT4170 sequence. Sugars were numbered from the non-reducing end to the reducing end. (Rha) L-rhmanose; (GalA) D-galacturonic acid.

Protein:BT4170 Ligand UGalA-Rha-GalA (positive subsites)

Hydrogen bonds

	<----- A T O M 1 ----->					<----- A T O M 2 ----->				
	Atom no.	Atom name	Res name	Res no.	Chain		Atom no.	Atom name	Res name	Res no.	Chain	Distance
1.	2249	NH2	ARG	319	A	-->	3319	O5	GalA	3	D	3.35
2.	2274	OH	TYR	322	A	-->	3317	O62	GalA	3	D	3.06
3.	2289	ND2	ASN	325	A	-->	3314	O4	Rha	2	D	2.74
4.	2298	NE2	HIS	325	A	-->	3315	O61	GalA	3	D	3.21

Non-bonded contacts

	<----- A T O M 1 ----->					<----- A T O M 2 ----->				
	Atom no.	Atom name	Res name	Res no.	Chain		Atom no.	Atom name	Res name	Res no.	Chain	Distance
1.	1692	OD2	ASP	246	A	---	3296	O62	UGalA	1	D	3.00
2.	1700	OD2	ASP	247	A	---	3296	O62	UGalA	1	D	3.55
3.	1721	CG	ASP	250	A	---	3294	O61	UGalA	1	D	3.39
4.	1722	OD1	ASP	250	A	---	3294	O61	UGalA	1	D	3.49
5.	1723	OD2	ASP	250	A	---	3294	O61	UGalA	1	D	2.58
6.	1723	OD2	ASP	250	A	---	3295	C6	UGalA	1	D	3.13
7.	1723	OD2	ASP	250	A	---	3296	O62	UGalA	1	D	2.81
8.	1736	CB	ILE	252	A	---	3294	O61	UGalA	1	D	3.61
9.	1739	CG2	ILE	252	A	---	3294	O61	UGalA	1	D	3.13
10.	1738	CD1	ILE	252	A	---	3294	O61	UGalA	1	D	3.81
11.	1738	CD1	ILE	252	A	---	3314	O4	Rha	2	D	3.73
12.	1955	OD1	ASP	280	A	---	3326	O2	GalA	3	D	3.53
13.	1956	OD2	ASP	280	A	---	3296	O62	UGalA	1	D	3.81
14.	1956	OD2	ASP	280	A	---	3298	O5	UGalA	1	D	2.93
15.	1956	OD2	ASP	280	A	---	3304	C1	UGalA	1	D	3.45
16.	1956	OD2	ASP	280	A	---	3312	C2	Rha	2	D	3.45
17.	1956	OD2	ASP	280	A	---	3313	O2	Rha	2	D	3.45
18.	2019	N	MET	291	A	---	3308	C6	Rha	2	D	3.52
19.	2021	CB	MET	291	A	---	3308	C6	Rha	2	D	3.62
20.	2022	CG	MET	291	A	---	3308	C6	Rha	2	D	3.64
21.	2237	CD1	LEU	318	A	---	3327	O1	GalA	3	D	3.11
22.	2247	CZ	ARG	319	A	---	3317	O62	GalA	3	D	3.57
23.	2248	NH1	ARG	319	A	---	3317	O62	GalA	3	D	3.39
24.	2248	NH1	ARG	319	A	---	3319	O5	GalA	3	D	3.55
25.	2249	NH2	ARG	319	A	---	3316	C6	GalA	3	D	3.60
26.	2249	NH2	ARG	319	A	---	3317	O62	GalA	3	D	3.47
27.	2249	NH2	ARG	319	A	---	3319	O5	GalA	3	D	3.35
28.	2275	CE2	TYR	322	A	---	3317	O62	GalA	3	D	3.58
29.	2273	CZ	TYR	322	A	---	3317	O62	GalA	3	D	3.77
30.	2274	OH	TYR	322	A	---	3315	O61	GalA	3	D	3.30
31.	2274	OH	TYR	322	A	---	3316	C6	GalA	3	D	3.66
32.	2274	OH	TYR	322	A	---	3317	O62	GalA	3	D	3.06
33.	2287	CG	ASN	324	A	---	3314	O4	Rha	2	D	3.35
34.	2288	OD1	ASN	324	A	---	3311	O3	Rha	2	D	3.78
35.	2288	OD1	ASN	324	A	---	3314	O4	Rha	2	D	3.41
36.	2289	ND2	ASN	324	A	---	3309	C4	Rha	2	D	3.53
37.	2289	ND2	ASN	324	A	---	3310	C3	Rha	2	D	3.25
38.	2289	ND2	ASN	324	A	---	3311	O3	Rha	2	D	3.28
39.	2289	ND2	ASN	324	A	---	3314	O4	Rha	2	D	2.74
40.	2299	CD2	HIS	325	A	---	3307	C5	Rha	2	D	3.72
41.	2299	CD2	HIS	325	A	---	3308	C6	Rha	2	D	3.73
42.	2299	CD2	HIS	325	A	---	3314	O4	Rha	2	D	3.49
43.	2297	CE1	HIS	325	A	---	3315	O61	GalA	3	D	3.77
44.	2298	NE2	HIS	325	A	---	3307	C5	Rha	2	D	3.87
45.	2298	NE2	HIS	325	A	---	3315	O61	GalA	3	D	3.21

Number of hydrogen bonds: 4

Number of non-bonded contacts: 45

Figure G.4 List of interaction between BT4170 and the ligand Δ4,5-GalA-α1,2-L-Rha-α1,4-D-GalA

Amino acids numbered conforming to the BT4170 sequence. Sugars were numbered from the non-reducing end to the reducing end. (Rha) L-rhmanose; (GalA) D-galacturonic acid; (UGalA) Δ4,5-GalA or unsaturated GalA.

List of protein-metal interactions

```
-----------------------------------
                  Protein:BT4170   Metal  CA (1)
                ---------------------------

Hydrogen bonds
--------------
       <----- A T O M   1 ----->      <----- A T O M   2 ----->

       Atom Atom Res  Res             Atom Atom Res  Res
        no. name name no.  Chain       no. name name no.  Chain  Distance
  1.   1468  OD2 ASP  215    A   <->  3259 CA   CA     1    B      2.35
  2.   1692  OD2 ASP  246    A   <->  3259 CA   CA     1    B      2.17
  3.   1700  OD2 ASP  247    A   <->  3259 CA   CA     1    B      2.53
  4.   1723  OD2 ASP  250    A   <->  3259 CA   CA     1    B      2.47

Non-bonded contacts
-------------------
       <----- A T O M   1 ----->      <----- A T O M   2 ----->

       Atom Atom Res  Res             Atom Atom Res  Res
        no. name name no.  Chain       no. name name no.  Chain  Distance
  1.   1466  CG  ASP  215    A   :-:  3259 CA   CA     1    B      2.80
  2.   1467  OD1 ASP  215    A   :-:  3259 CA   CA     1    B      2.53
  3.   1468  OD2 ASP  215    A   :-:  3259 CA   CA     1    B      2.35
  4.   1690  CG  ASP  246    A   :-:  3259 CA   CA     1    B      3.39
  5.   1692  OD2 ASP  246    A   :-:  3259 CA   CA     1    B      2.17
  6.   1698  CG  ASP  247    A   :-:  3259 CA   CA     1    B      3.37
  7.   1700  OD2 ASP  247    A   :-:  3259 CA   CA     1    B      2.53
  8.   1720  CB  ASP  250    A   :-:  3259 CA   CA     1    B      3.82
  9.   1721  CG  ASP  250    A   :-:  3259 CA   CA     1    B      3.53
 10.   1723  OD2 ASP  250    A   :-:  3259 CA   CA     1    B      2.47

Number of hydrogen bonds:          4

Number of non-bonded contacts:    10



                   PDB code: q987   Metal  CA (2)
                ---------------------------
Hydrogen bonds
--------------
       <----- A T O M   1 ----->      <----- A T O M   2 ----->

       Atom Atom Res  Res             Atom Atom Res  Res
        no. name name no.  Chain       no. name name no.  Chain  Distance
  1.   1447  O   GLY  191    A   <->  3260 CA   CA     2    B      2.48
  2.   1692  OD2 ASP  225    A   <->  3260 CA   CA     2    B      2.60
  3.   1956  OD2 ASP  259    A   <->  3260 CA   CA     2    B      2.72

Non-bonded contacts
-------------------
       <----- A T O M   1 ----->      <----- A T O M   2 ----->

       Atom Atom Res  Res             Atom Atom Res  Res
        no. name name no.  Chain       no. name name no.  Chain  Distance
  1.   1446  C   GLY  191    A   :-:  3260 CA   CA     2    B      3.64
  2.   1447  O   GLY  191    A   :-:  3260 CA   CA     2    B      2.48
  3.   1690  CG  ASP  225    A   :-:  3260 CA   CA     2    B      2.91
  4.   1691  OD1 ASP  225    A   :-:  3260 CA   CA     2    B      2.71
  5.   1692  OD2 ASP  225    A   :-:  3260 CA   CA     2    B      2.60
  6.   1954  CG  ASP  259    A   :-:  3260 CA   CA     2    B      3.57
  7.   1956  OD2 ASP  259    A   :-:  3260 CA   CA     2    B      2.72

Number of hydrogen bonds:          3

Number of non-bonded contacts:     7
```

Figure G.5 List of interaction between BT4170 and the calcium

Amino acids numbered conforming to the BT4170 sequence. (Ca) calcium ion.

Appendix H

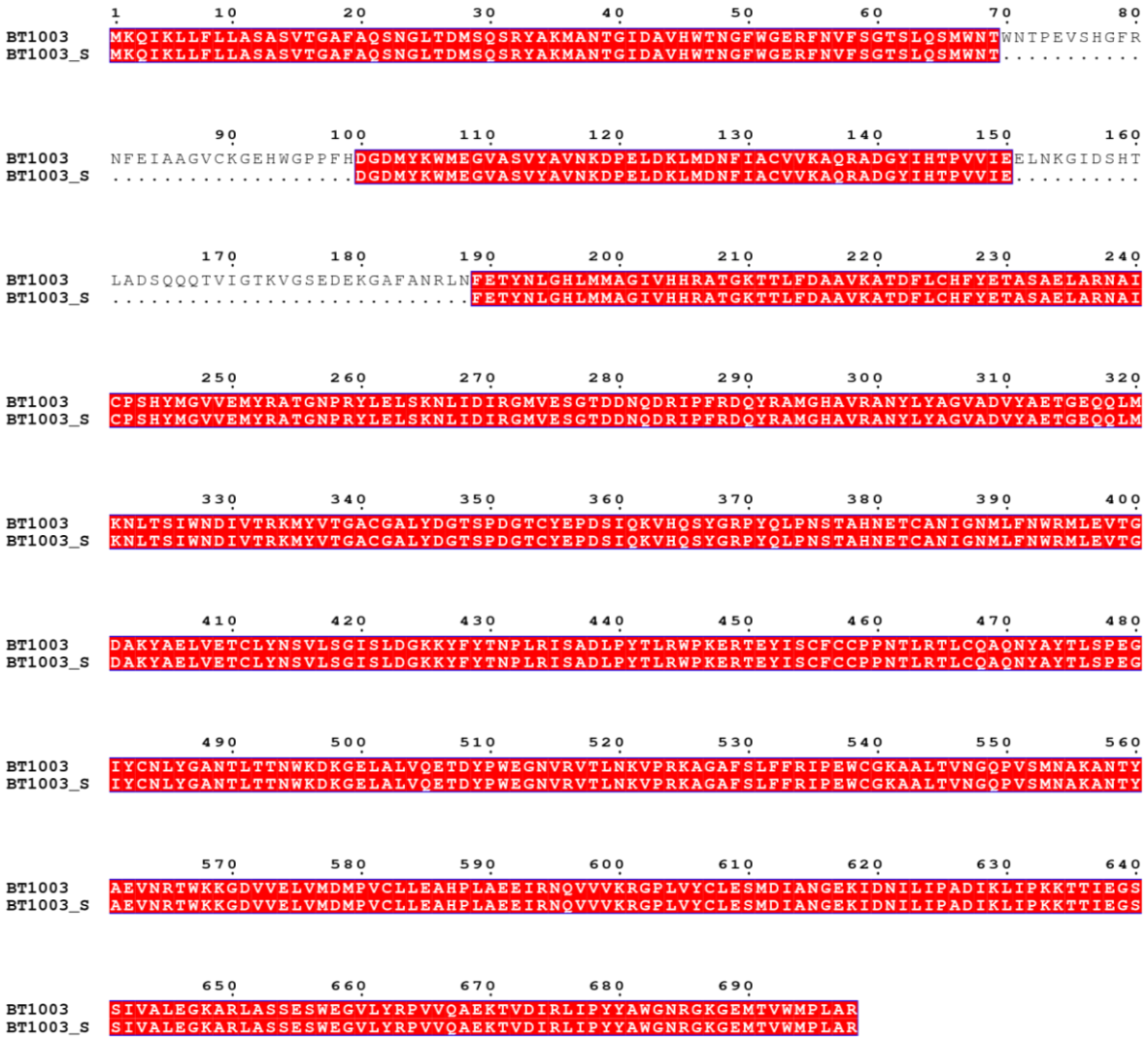


Figure H.1 Alignment of BT1003 protein sequence with the sequence in the crystal structure (BT1003_s)

The non-conserved amino acids (W70 to H99 and E151 to N188) correspond to the missing sequence at BT1003 structure. Sequences were aligned in Clustal MUSCLE (Edgar 2004a; Edgar 2004b) and visualized using ESPrit 3.0 (Robert and Gouet 2014). Amino acids with 100% of conservation are highlighted in a red background.

Appendix I

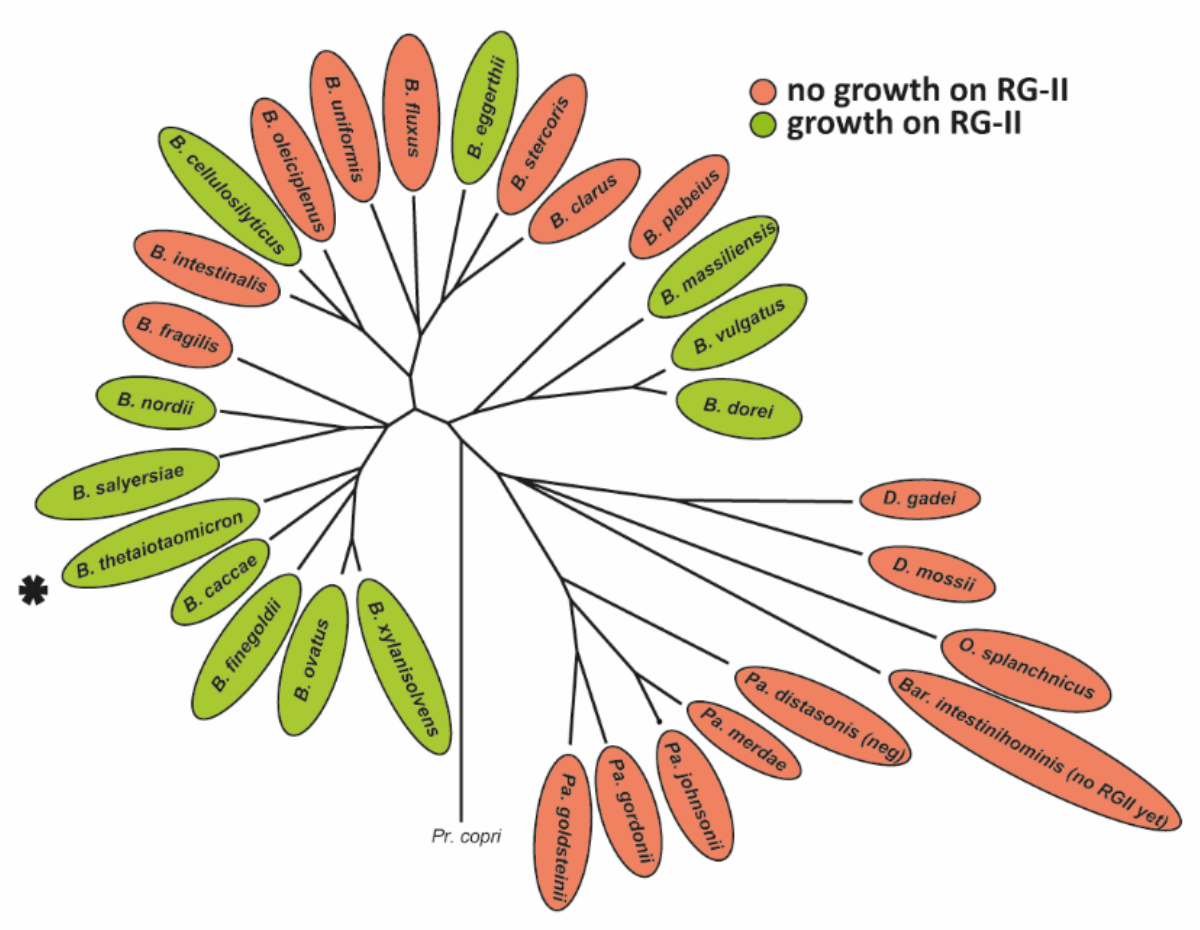


Figure I.1 Growth of selected *Bacteroidetes* species on RG-II

The 29 type strains of *Bacteroidetes* isolated from the human gut microbiota were inoculated into minimal media containing RG-II as the carbon source. The growth of the cultures was recorded at 600 nm. In Chapter 5.3.3.3 the aligned GH106 proteins are encoded by *B. thetaiotaomicron, B. xylanisolvens, B. ovatus, B. eggerthii, B. vulgatus and B. dorei.* Figure modified with the courtesy of Dr Artur Rogowski (ICaMB, Newcastle University).

Appendix J

List of primers used to generate different gene deletions in *Bacteroides thetaiotaomicron* and example of an agarose gel with the PCR products.

Table J.1 Primers designed to create different gene deletions

Gene deletion	Primer name	Primer sequence (5' to 3')	Restriction site
Δbt4164/4165	64/65_F1_Sal_Fw	cgc ggt cga cat ggc tgc gta cac tac cac	*SalI*
	64/65_F1_Rv	ctt aca ata ttc ttc ata gta tat aga tta att aga tat tca gtt a	-
	64/65_F2_Fw	taa ctg aat atc taa tta atc tat ata cta tga aga ata ttg taa g	-
	64/65_F2_Xba_Rv	ctc tct aga ccg tca att ggt tgg ata tat cc	*XbaI*
Δbt4168/4169	68/69_F1_Sal_Fw	cgc ggt cga cca gag aac att acc gct aca ttg	*SalI*
	68/69_F1_Rv	ttt gat tga aaa aca tgc tat atc ctg ttt atc aag gtt aat ac	-
	68/69_F2_Fw	gta tta acc ttg ata aac agg ata tag cat gtt ttt caa tca aa	-
	68/69_F2_Xba_RV	ctc tct aga gag agc aat acg gct gtc aat agc	*XbaI*

Restriction sites are shown underlined; F1 (fragment 1); F2 (fragment 2); Fw (forward primer); Rv (reverse primer)

Table J.2 Primers designed to sequencing the different gene deletions in *B. thetaiotaomicron*

Gene deletion	Primer name	Primer sequence (5' to 3')
Δbt4164/4165	64_65KO_Btg_Fw	gat aag agt atc tgt gtg aac ggt ag
	64_65KO_Btg_Rv	gca gtg tac gta tca tca tcg tca c
Δbt4168/4169	68_69KO_Btg_Fw	gag ccg tac aca aag agt att tg
	68_69KO_Btg_Rv	cac cat tac aat atg ctt act ac

BTg (*B. thetaiotaomicron* genomic DNA); Fw (forward primer); Rv (reverse primer)

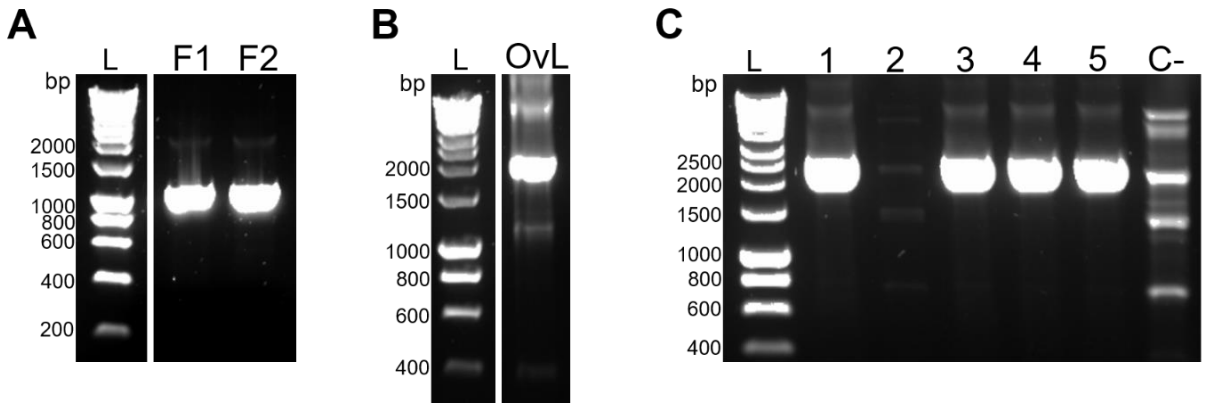


Figure J.1 Example of PCR-derived fragments of different gene deletion constructs

The figure presents the different PCRs utilized to construct the mutant *⊿bt4164/4165*. After the PCR (Chapter 2.1.11), 5 µl of products with 2 µl of loading dye were electrophoresed in a 1% (w/v) agarose gel (Chapter 2.1.12). **A.** shows the PCRs of the flanks fragment (F1 for upstream and F2 for downstream). The 1000 bp fragments were generated utilizing the primers 64/65_F1_Sal_Fw + 64/65_F1_Rv (F1) and 64/65_F2_Fw + 68/69_F2_Xba_RV (F2). **B.** presents the PCR overlap (OvL, 2000 bp) of F1 and F2 utilizing the primers 64/65_F1_Sal_Fw + 68/69_F2_Xba_RV. **C.** presents the screen by PCR to identify *B. thetaiotaomicron* mutants containing the gene deletion in the genome. The screen was performed with the primers 64_65KO_Btg_Fw + 64_65KO_Btg_Rv. The clones 1, 3, 4 and 5 represent putative positive clones (band with ~ 2200 bp). C- represents the PCR performed with genomic DNA from *B. thetaiotaomicron* wild-type. The respective DNAs were labelled and L represents the size marker in base pairs (bp).

Appendix K

List of primers used to FLAG-tag genes in *Bacteroides thetaiotaomicron* and example of an agarose gel with the PCR products.

Table K.1 Primers designed to introduce a C-terminal FLAG-Tag

Gene Flagged	Primer name	Primer sequence (5' to 3')	Restriction site
bt4170	4170F_F1_Bam_Fw	cgc ggg atc cgg tca cat taa ctc atc ttc ag	*BamHI*
	4170F_F1_Rv	tta **ttt gtc gtc atc gtc ttt gta gtc** tct ttt aaa ttt aag cgc aat ag	-
	4170F_F2_Fw	**gac tac aaa gac gat gac gac aaa** taa aac gga taa act tag aat aat	-
	4170F_F2_Xba_Rv	cgc gtc tag agc gtt gag tcc gga ggt t	*XbaI*
bt4175	4175F_F1_Bam_Fw	cgc ggg atc cca atc gta gcg acc gtt atc tg	*BamHI*
	4175F_F1_Rv	tta **ttt gtc gtc atc gtc ttt gta gtc** ctt ttt agg tat ttt aca tcc tc	-
	4175F_F2_Fw	**gac tac aaa gac gat gac gac aaa** taa att atg aat tat cgt atg atg	-
	4175F_F2_Xba_Rv	cgc gtc tag aaa tcc ttc tcc act ttg caa c	*XbaI*
bt4176	4176F_F1_Sal_Fw	cgc ggg atc cgg aag tcg ctc cgc att tc	*SalI*
	4176F_F1_Rv	tta **ttt gtc gtc atc gtc ttt gta gtc** ttc act atc tac cgc aaa g	-
	4176F_F2_Fw	**gac tac aaa gac gat gac gac aaa** taa ata aaa gat taa ctt taa aat ata gg	-
	4176F_F2_Xba_RV	cgc gtc tag acg ctc cgt ctt aaa ttc cg	*XbaI*

The FLAG-TAG sequences are highlighted in bold and the restriction sites are shown underlined; F1 (fragment 1); F2 (fragment 2); Fw (forward primer); Rv (reverse primer).

Table K.2 Primers utilized to sequencing the different gene deletions in *B. thetaiotaomicron*

Gene Flagged	Primer name	Primer sequence (5' to 3')
bt4170	4170F_Btg_Fw	ggg att tga tgt cat cgg tac gc
	4170F_ Btg_Rv	cgc cat ttc tca cgt ttc cag
bt4175	4175F_ Btg_Fw	atg gtc ttc gat ctg gac ggt g
	4175F_ Btg_Rv	ccg cat ata cat tga ccg gac gat
bt4176	4176F_ Btg_Fw	gcg tat ata gat aag gag aca cc
	4176F_ Btg_RV	ccc gta ggt gac gtc tta ccc gca

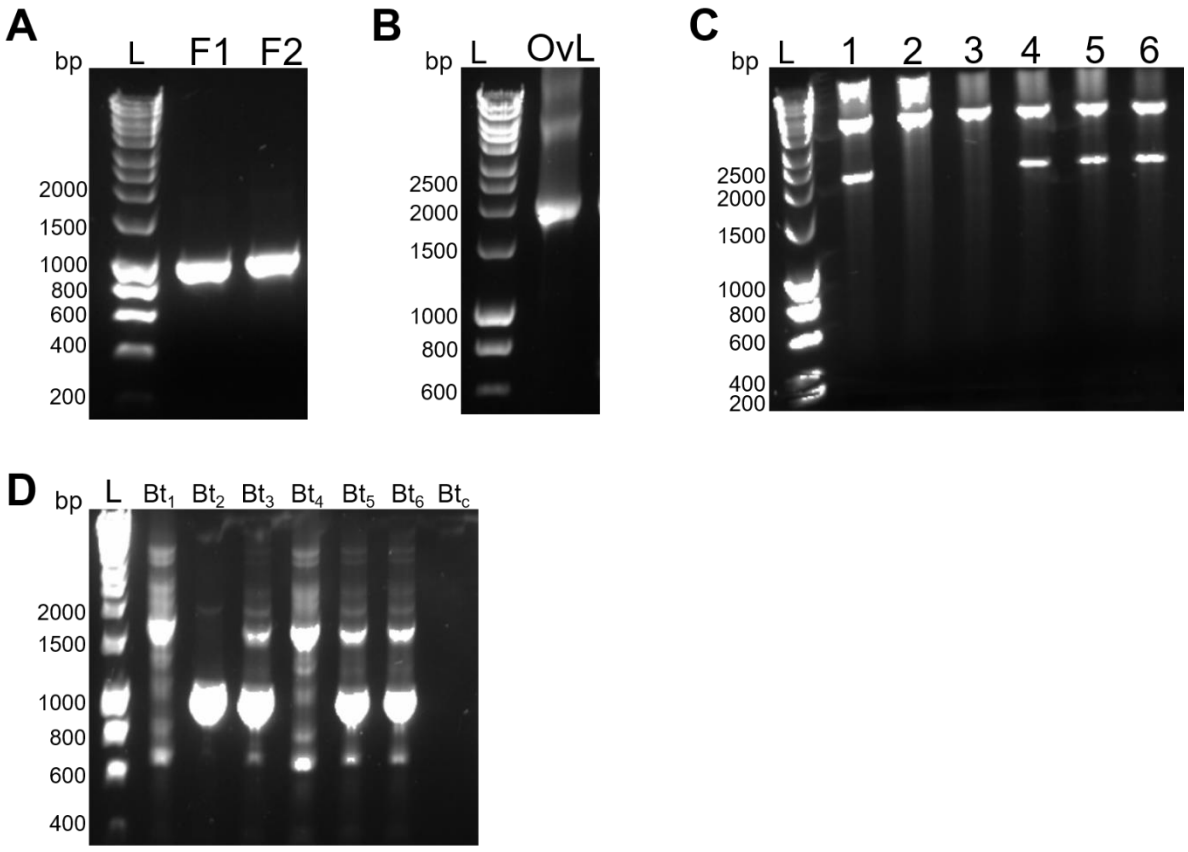


Figure K.1 Example of PCR-derived fragments of different FLAG-tag constructs analysed by agarose gel electrophoresis

The figure shows the different PCRs utilized to construct the mutant the *bt4170* FLAG-tag. After the PCR (Chapter 2.1.11), 5 µl of products with 2 µl of loading dye were electrophoresed in a 1% (w/v) agarose gel (Chapter 2.1.12). **A.** PCRs of the flanks fragment (F1 for upstream and F2 for downstream). The 1000 bp fragments were generated utilizing the primers 4170F_F1_Bam_Fw + 4170F_F1_Rv (F1) and 4170F_F2_Fw + 4170F_F2_Xba_Rv (F2). **B.** PCR overlap (OvL, 2000 bp) of F1 and F2 utilizing the primers 4170F_F1_Bam_Fw + 4170F_F2_Xba_Rv. **C.** Colony PCR screen after transformation of ligation pExchange-*tdk*-PCR. Screen was performed with the primers utilized in B. Clones 1, 4, 5 and 6 represent putative positive clones (band with ~ 2200 bp). Clones 1, 3, 4 and 5 represent putative positive clones (band with ~ 2200 bp). **D.** PCR screen to identify *B. thetaiotaomicron* mutants containing the gene FLAG-tagged in the genome. Screen was performed with the primers FLAG_Fw (gac tac aaa gac gat gac gac aaa) + 4170F_ Btg_Rv. The clones Bt_2, Bt_3, Bt_5 and Bt_6 represent putative positive clones (band with ~ 1000 bp). Bt_c represents the negative control PCR done with genomic DNA from *B. thetaiotaomicron* wild-type. The respective DNAs were labelled and L represents the size marker in base pairs (bp).