Insights into the enzymatic degradation of mannan (*in situ*) and pectin utilization by *Bacteroides thetaiotaomicron*

A thesis submitted for the degree of Doctor of Philosophy
Newcastle University
2010-2014

Xiaoyang (Jeff) Zhang

Institute for Cell and Molecular Biosciences
Newcastle University

Acknowledgements

First of all, I would like to thank Harry for his supervision, both in Complex Carbohydrate Research Center, USA and Newcastle, without which I would have taken many "scientific detour". I am also grateful for being a part of the Marie-curie family and had such a great opportunities working closely with people in similar fields. It's been such a good time working with the following members of Harry's laboratory, both past and present in no particular order, for their company and advice: Yanping Zhu, Carl Morland, Dave Bolam, Artur Rogowski, Wade Abbott, Adam Jackson, Elisabeth Lowe, Alan Cartmell, Lauren Mckee, Sarah Shapiro, Didier Ndeh, Jonathon Briggs, Max Temple, Ana Luis, Lucy Crouch, Fiona Cuskin, and Jose Munoz.

As collaborative the nature of PhD and particularly the Marie-curie are, a huge amount of kind and generous support has been given by Michael Hahn (CCRC, USA) and Paul Knox (Leeds, UK) for the mannan project, and Marie-Christine Ralet (INRA, France) for the pectin project. My big and sincere Thank You goes unreservedly to all these people. I also acknowledge the support by the European Union Seventh Framework Programme (FP7 2007–2013) under Grant 263916 (WallTraC, Marie Curie Initial Training Network).

On a personal level I am grateful for having a great supportive family. Thank you, my dear parents, grandparents and all my friends.

(老爸老妈，我要毕业啦。哈。在此用中文写上一段，以作纪念。祝你们和姥姥姥爷身体健康， 万事如意。)

Abstract

The cell walls of plants are extremely complex and dynamic, and serve as the most abundant organic carbon source on the planet. These cell walls represent a potential alternative to fossil substrate for several industries such as the biofuel and chemical sectors. Plant cell walls also represent a significant nutrient for the microorganisms that inhabit the human distal gut. Understanding how the plant cell wall is degraded by microbial enzymes is critical to the industrial use of these composite structures, but also in the design of human promoting novel pre- and probiotics. In this thesis work is described that has analysed the importance of non-catalytic carbohydrate binding modules (CBMs) in the enzymatic degradation of β-mannanas in plant cell walls, and the mechanism by which complex pectins are utilized by gut bacteria.

Plant cell wall degrading enzymes often contain one or more CBMs. Work presented in this thesis evaluated the contribution of the specificities displayed by the catalytic modules and CBMs of endo-β1,4-mannanases (mannanases) and esterases, to the enzymatic degradation of intact plant cell walls of both *Nicotiana tabacum* (tobacco) and *Physcomitrella patens (moss)*. The data showed that CBMs greatly potentiated the activity of mannanases and esterases against intact plant cell walls. The cellulose and mannan binding CBMs have the greatest impact on the removal of mannan from tobacco and *Physcomitrella* cell walls, respectively. Finally, rather than improving the efficiency of mannan degradation by promoting enzyme substrate proximity, CBM32, which binds to the non-reducing end of mannan chains, enhanced the hydrolytic activity of a mannanase by preventing transglycosylation reactions. This work provides insights into the biological significance *in vivo* for the complex array of hydrolytic enzymes expressed by plant cell wall degrading organisms.

Chapter 4 and 5 focused on the mechanism by which a prominent human gut bacterium, *Bacteroides thetaiotaomicron*, degraded the complex pectins rhamnogalacturonan I (RGI) and rhamnogalacturonan II (RGII). RGII is particularly complex consisting of 13 different monosaccharides and over 20 linkages and is believed to be highly resistant to microbial digestion. In this study proteins encoded by the RGI and RGII utilization loci of *B. thetaiotaomicron* were expressed in *Escherichia coli*, and their activity against RGI and RGII were assessed. With respect to RGII BT1010 was shown to be an α-L-galactosidase that removes the terminal sugar on chain A. BT3662is an α-L-arabinofuranosidase that hydrolysed the α-1,3-linkage between arabinofuranosyl units and the polygalacturonic acid backbone. Both enzymes displayed novel activities for their respective CAZy families, GH95 (BT1010) and GH43 (BT3662). With respect to RGI a model for the complete depolymerization of the polysaccharide was generated. Briefly, the enzyme consortium that degraded the GalA-Rha backbone included three rhamnogalacturonan lyases and four glycoside hydrolases; two GH105 4,5Δunsaturated rhamnogalacturonase, a rhamnosidase and an RGI-specific galacturonosidase. The kinetic parameters of the lyases and GH105 enzymes revealed distinct but complementary specificities. The short galactan side chains remaining after endo-galactanase attack were removed by three β-galactosidases. The model is completed by the identification of an esterase that was shown to play an important role in the capacity of the lyases and glycoside hydrolases to access the RGI backbone. Whole cell assays showed that significant RGI degradation occurred on the bacterial surface and unsaturated tetrasaccharides induced the expression of the RGI degrading enzymes.

The data presented in this thesis underpins the use of CBMs in the industrial utilization of integral cell wall polysaccharides, while the data presented for pectin degradation by a prominent gut bacterium identifies opportunities for developing novel functional foods.

Publications resulted from the work in this thesis:

Zhang, X., Rogowski, A., Zhao, L., Hahn, M., Avci,U, Knox, P., Gilbert, H. (2013). Understanding how the complex molecular architecture of mannan degrading hydrolases contributes to plant cell wall degradation. The Journal of Biological Chemistry, 2002-12.

Buffetto, F., Ropartz, D., **Zhang, X.**, Gilbert, HJ., Guillon, F., Ralet, MC.(2014) Recovery and fine structure variability of RGII sub-domains in wine (Vitis vinifera Merlot). Annals of Botany, 1327-37.

Mizutani, K., Fernandes, V. O., Karita, S., Luis, AS., Sakka, M., Kimura, T., Jackson, A., **Zhang, X.**, Fontes, CM., Gilbert, HJ., Sakka, K. (2012). Influence of a mannan binding family 32 carbohydrate binding module on the activity of the appended mannanase. Applied and Environmental Microbiology, 4781-87.

Abbreviations

2A, 6M	Acetyl or Methyl and their substitution position on glycans
AA	Amino Acid
ABC	ATP-binding cassette
ABP	arabinose-binding protein
AG	arabinogalactan
AGE	Affinity Gel Electrophoresis
Amp	Ampicillin
Api	Apiose
Ara	Arabinose
BLAST	Basic Local Alignment Search Tool
BSA	Bovine serum albumin, fraction V
BT	*Bacteroides thetaiotaomicron*
CAZymes	Carbohydrate-Active enZymes
Capso	3-(Cyclohexylamino)-2-hydroxy-1-propanesulfonic acid
DHA	3-deoxy-D-lyxo-2-heptulosaric acid
DNA	Deoxyribonucleic acid
DNSA	3,5-Dinitrophenol
DP	Degree of polymerisation
DTT	Dithiothrietol
ECF	Extracytoplasmic function
EDTA	Ethelene diamine tetra-aceticacid, disodium salt
FEH	Fructan exohydrolase
FPLC	Fast Protein Liquid Chromatography
Frc	Fructose
Fuc	Fucose
Gal	Galactose
GalA	Galacturonic acid
GBP	Glucose binding protein
GH	glycoside hydrolase (family)
GI	gastrointestinal
Glc	D-glucose
GlcA	Glucuronic acid
HBP	histidine-binding protein
HEPES	N-(2-Hydroxyethyl) piperazine-N′-(2-ethanesulfonic acid) sodium salt
HGA	Homogalacturonan
HK	Histidine kinase
HPLC	High Performance Liquid Chromatography
HTCS	Hybrid two component system
IM	Inner Membrane
IMAC	Immobilised Metal Affinity Chromatography
IPTG	Isopropyl-β-D-thiogalactopyranoside
ITC	Isothermal Calorimetry

IUBMB	International Union of Biochemistry and Molecular Biology
Kan	Kanamycin
KDO	keto sugar 3-Deoxy-D-manno-oct-2-ulosonic acid
LB	Luria-Bertani Broth
MBP	Maltose binding protein
MOPS	3-(*N*-Morpholino) propane sulfonic acid
MW	Molecular Weight
ND	Not determined
OM	Outer membrane
PAD	pulsed amperometric detection
PAGE	Polyacrylamide Gel Electrophoresis
PBP	Periplasmic binding protein
PCR	Polymerase Chain Reaction
PDB	Protein data bank
PEG	Polyethylene Glycol
PEG550mme	Polyethylene Glycol 550 monomethyl ether
PG	Pectic Galactan
PINNACLE	Proteome investigations at Newcastle
PL	Polysaccharide Lyases
PULs	Polysaccharide utilisation loci
RBP	Ribose binding protein
Rha	Rhamnose
RGI	Rhamnogalacturonan I
RGII	Rhamnogalacturonan II
RNAP	RNA polymerase
RR	Response regulator
SDS	Sodium Dodecyl Sulphate
Se-Met	Seleno methionine
Sus	Starch utilisation system
TCEP	Tris (2-carboxyethyl) phosphine hydrochloride
TCS	Two component system
TEMED	N,N,N',N'-Tetramethylethylenediamine
TLC	Thin layer chromatography
TM	Transmembrane
WT	Wild type
XG	Xyloglucan
Xyl	Xylose

Chapter 1 Introduction

1.1 Plant cell wall

The cell walls of plants, composed predominantly of the polysaccharides cellulose, hemicellulose, and pectin, are extremely complex and dynamic, and serve as the most abundant organic carbon source on the planet. Each year approximately 10^{11} tonnes of plant biomass is produced by photosynthesis with an estimated total energy content equivalent to 640 billion tons of oil (Coughlan 1985). Therefore, the plant cell wall represents a significant substrate for several biotechnological processes such as the use of pectins and probiotics in the food sector, the generation of biofuels and in supplying feeder molecules for the chemical industry. However, due to its recalcitrant nature, knowledge leading to complete conversion of biomass to industrial products is very limited. Thus, there is an urgent need to identify enzyme consortia that efficiently hydrolyse the plant cell wall, and the production of these biocatalysts at economically viable levels.

The plant cell wall consists of distinct layers (**Figure 1.1**): the middle lamella, the primary cell wall and the secondary cell wall (Beguin and Aubert 1994). It provides the plant with excellent structural support, conferred by a complex framework of cellulose microfibrils interconnected by a hemicellulose matrix of polymers that include xylan, xyloglucan, mannan, arabinoxylan and glycoproteins (Carpita and Gibeaut 1993).

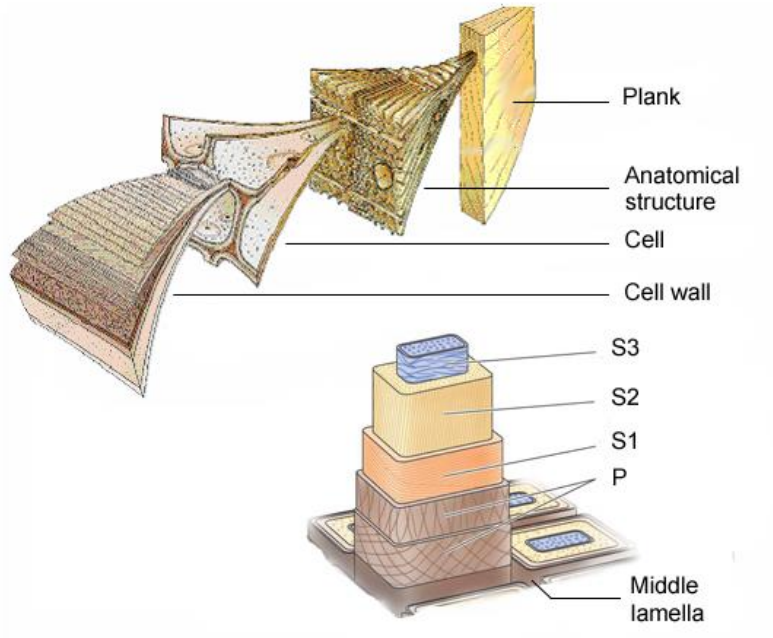


Figure 1.1 Structure of the plant cell wall
Schematic of the plant cell taken from Béguin and Aubert (1994). S3, S2, S1 and P represent for different layers of the secondary cell wall (S) and the primary cell wall (P), respectively.

Compared with the primary cell wall composed of cross-linked cellulose, hemicellulose and pectin (Cosgrove 2005, Brett CT 1996), the secondary cell walls contain a higher concentration of hemicellulose and a reduced amount of pectin. Phenolic compounds in the secondary walls, such as lignin, also provide additional rigidity and strength, as well as hydrophobic properties (Whetten and Sederoff 1995).

1.1.1 Cellulose

As the most abundant polysaccharide on Earth, cellulose comprises at least 15000 glucose linked by β-1,4-glycosidic bonds (Carpita and Gibeaut 1993) . The repeating unit is the disaccharide cellobiose with the second residue rotated 180º compared to the first. About 30-100 glucan chains align in through an ordered hydrogen bonding network to form microfibrils 3-5 nm diameter (Beguin and Aubert 1994, Lehtio et al. 2003a). The β1-4 linkage gives cellulose a unique linear structure resulting in a crystalline property, which confers high tensile strength and chemical stability. In contrast, the α-linkages of starch form

an accessible helix that is much more susceptible to enzymatic digestion and is thus ideal for carbon storage (Beguin and Aubert 1994).

1.1.2 Hemicellulose

Hemicellulosic polysaccharides all contain a β1,4-linked sugar backbone: Xylans, mannans, and xyloglucans contain β-1,4-D-Xyl, β-1,4-D-Man, and β-1,4-D-Glc backbone sugars, respectively, while in glucomannan, the backbone is randomly dispersed with β-1,4-Glc and β-1,4-Man sugars (**Figure 1. 2**). Xylan is the major hemicellulose in many land plants (Gilbert and Hazlewood 1991), while mannan is the primary hemicellulose in gymnosperms (Puls and Schuseil 1993b).

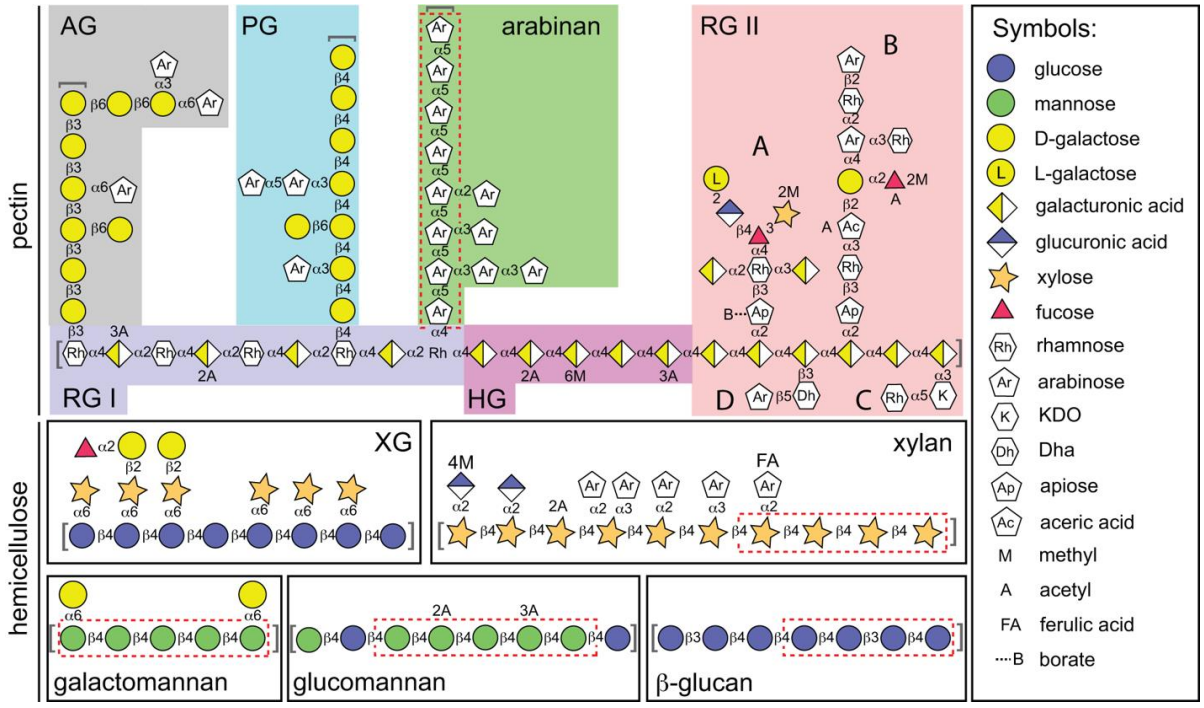


Figure 1. 2 The structure of hemicellulose and pectin (Martens et al. 2011). AG, arabinogalactan; PG, pectic Galactan; RGI, rhamnogalacturonan I; RGII, rhamnogalacturonan II; XG, xyloglucan; 2A or 6M, acetyl or methyl and their substitution position.

1.1.2.1 Xylan

Xylans is composed of β1, 4 linked xylose (Xyl) backbone and species-dependent side chains, and can account for 30% of the plant dry weight (Joseleau, Comtat and Ruel 1992). Arabinoxylans, for example, are xylans from monocot primary cell walls decorated with Ara units. The Ara units in arabinoxylans can also be substituted with ferulic acid. (Brett and Waldron 1990). In dicotyledonous plants (dicots), glucuronoxylan is the dominant type of xylan (Brett and Waldron 1990).

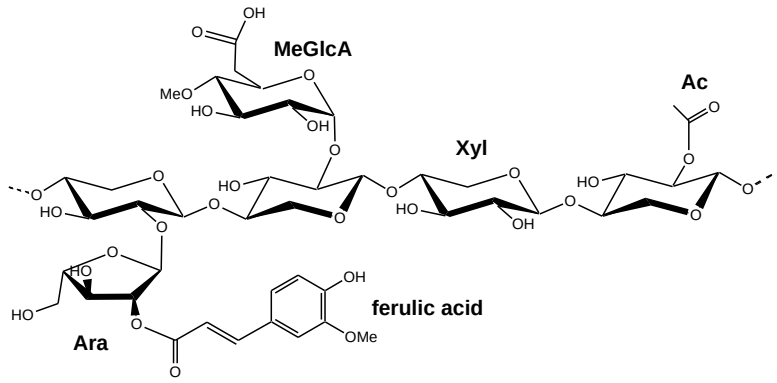


Figure 1.3 Structure of xylan
Structure of a hypothetical xylan chain showing the major types of possible substituents. Ac: acetyl group; Ara: arabinofuranose; MeGlcA: 4-O-methylglucuronic acid; Xyl: xylopyranose.

1.1.2.2 Xyloglucan

Being the main hemicellulose in the primary cell wall of dicotyledonous plants, xyloglucan consists of a β-1,4-Glc linked backbone decorated at regular intervals with α1,6-Xyl units that can be further decorated as follows: β1,2-D-Gal-α1,6-D-Xyl, α1,2-L-Fuc-β1,2-D-Gal-α1,6-D-Xyl and α1,2-L-Ara*f*-α1,6-D-Xyl. Xyloglucan is cross-linked with cellulose microfibrils and form hydrogen bonds with amorphous cellulose chains, or at the edges of cellulose microfibrils (Hayashi, Ogawa and Mitsuishi 1994a, Ogawa, Hayashi and Okamura 1990, Hayashi, Ogawa and Mitsuishi 1994b), and is thus thought to contribute to the flexibility of the cell wall (Pauly et al. 1999).

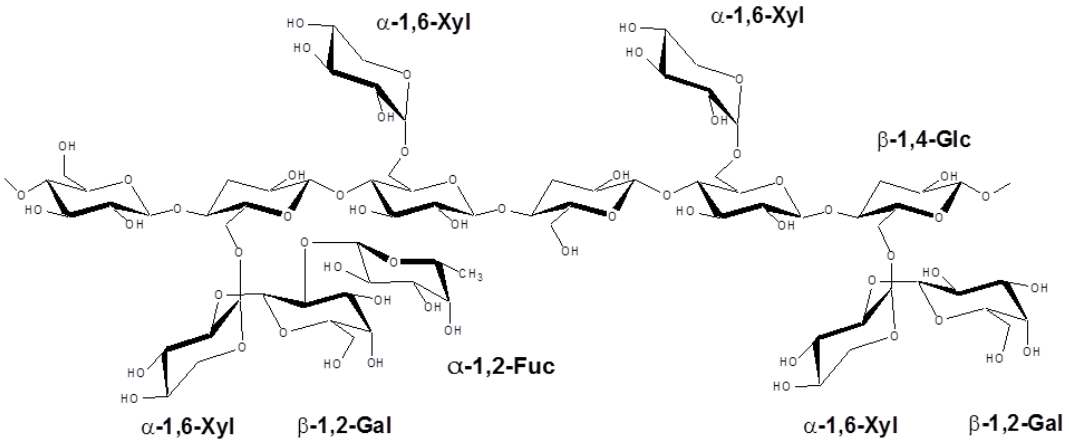


Figure 1. 4 Structure of xyloglucan (Dr. Cedric Montanier personal communication).

Structure of a hypothetical xyloglucan chain showing the major types of possible substituents. Abbreviations are: D-Xyl, xylopyranose; D-Glc, glucopyranose; D-Gal, galactopyranose; L-Fuc, fucopyranose.

1.1.2.3 Mannan

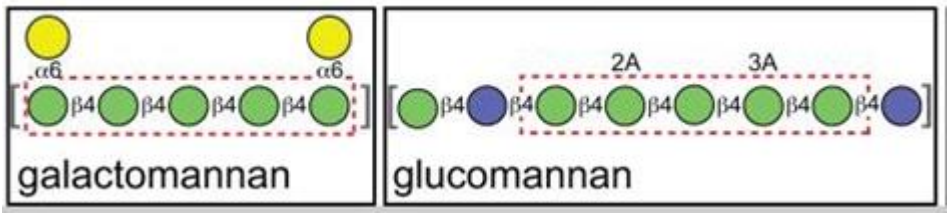
Mannan is a complex set of hemicellulosic heteroglycans containing β-1,4-mannosyl residues and are abundant in the lignified secondary cell walls of gymnosperms, as well as endosperm walls such as carob, ivory nuts and coffee beans (Puls and Schuseil 1993a). It consists of undecorated mannan, galactomannan, glucomannan, galactoglucomannan and glucuronomannan (Brett and Waldron 1990, Brett CT 1996). Similar to cellulose, undecorated mannan can form crystalline microfibrils. Usually mannan is considered to have both storage (especially in the seeds) and structural functions in cell walls. (Handford et al. 2003, Marcus et al. 2010). Distribution of mannans within *Arabidopsis thaliana* includes the thickened secondary cell walls of xylem structures, parenchyma, the interfascicular (between the vascular bundles of the stem) fibres in inflorescence stems, and in lower concentrations in thickened cell walls of leaves and stems (Handford et al. 2003).

Glucomannan contains a random order of β-1,4-linked Man and Glc residues in proportions that vary between plant species. It is the main hemicellulose component of the secondary cell wall of the gymnosperms, accounting for 25% of the dry weight of wood (Puls and Schuseil

1993a). Glucomannans can be acetylated to confer solubility at O2 and/or O3, similar to galactomannan (Capek et al. 2000).

In galactomannan and galactoglucomanan (usually found in small amount in many cell walls) a proportion of the Man residues, which varies between plant species, contain a single α-1,6-galactose side chain. The soluble galactomannan is a part of the secondary cell wall thickening and is thought to be involved in the imbibing and taking up of water by seeds (Brett CT 1996).

A



B

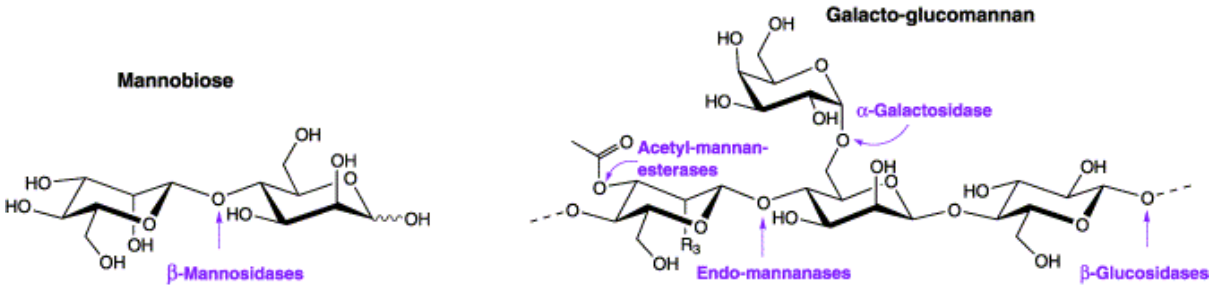


Figure 1. 5 Schematic structure of galactomannan and glucomannans (A) and enzymes that could degrade mannan related polysaccharides (B) (Shallom and Shoham 2003). (Panel A: Green circle, mannose; yellow circle, D-galactose; blue circle, glucose; A, acetyl group; 2A or 3A, acetyl group and their substitution position. Panel B: enzymes that attack different mannans)

1.1.3 Pectin

Pectic polysaccharides are important determinants of wall porosity and wall thickness while they also play structural roles to "glue" cells together in the adhesive middle lamella layer (Cosgrove 2005, Brett CT 1996). Previous research also pointed out the potential function of pectins in the modulation of cell wall pH and ion balance by providing charged surfaces (Mukhiddinov et al. 2000).

There are three major forms of pectin: homogalacturonan (HGA), rhamnogalacturonan I (RGI), and rhamnogalacturonan II (RGII). Homogalacturonan consists of a polygalacturonic acid backbone while RGI displays a backbone composed of an alternating disaccharide, [(α-1,4)-D-GalA/(α-1,2)-L-Rha]n. The Rha residues contain extensive side chain decorations. RGII is the most structurally complex of the three pectic polysaccharides, and its detailed structure is discussed later.

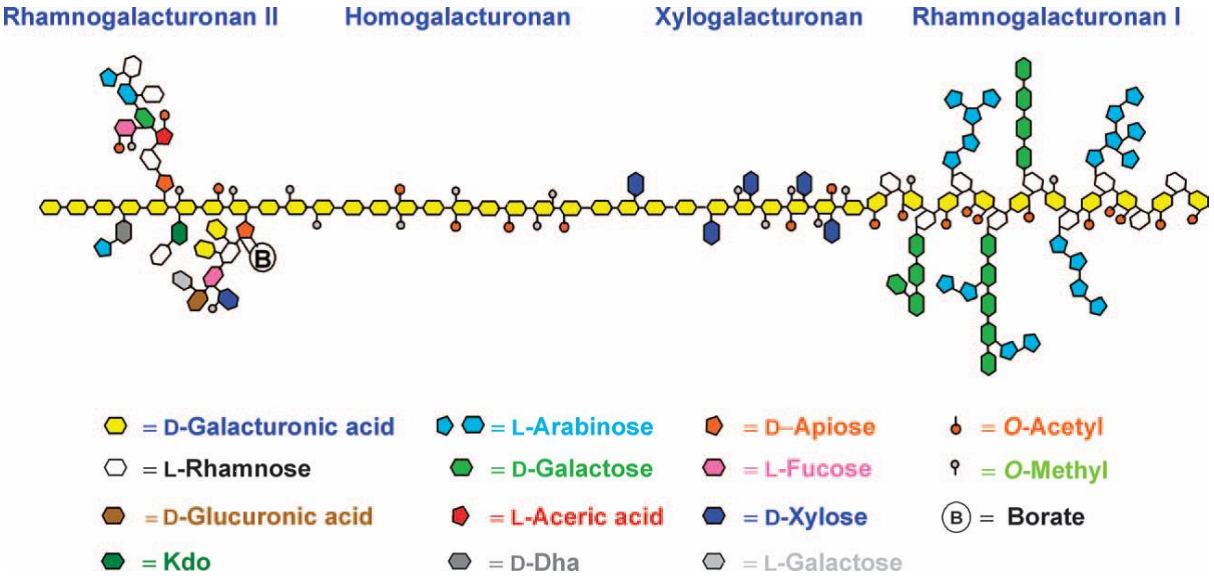


Figure 1. 6 Schematic structure of pectin

(Scheller et al. 2007).

1.1.3.1 Homogalacturonan (HG)

HGA is a homopolymer of 100-200 residues composed of α-(1,4)-linked galacturonic acid (GalA) units. The carboxylate of the GalA units is often methyl-esterified and sometimes acetyl esterification occurs at O2 or O3 (Ralet et al. 2001). The degree of methyl and acetyl esterification is variable depending on the plant species and growth phase, and this affects the physicochemical properties of the pectin *in vivo* (especially the formation of calcium-mediated interactions between HG chains) and *in vitro* (forming gels in food industry) (Harholt, Suttangkakul and Scheller 2010).

1.1.3.2 Rhamnogalacturonan I (RGI)

As the only type of pectin not built on a pure galacturonan backbone, RGI consists of an alternating GalA (α-1,4-linked) and Rha (α-1,2-linked) backbone, that is decorated at O4 of the Rha units with several different side chains. Around 20-80% of the Rha residues contain long side chains that vary depending on the plant species, although they primarily comprise arabinans and/or galactans (Willats et al. 2001). The term "hairy regions" was used (Brett CT 1996) to describe the RGI side chains, which occurred in blocks alternating with smooth side-chain free regions.

The arabinan side chains, which are water soluble, consists of a backbone of α-L-1,5-Ara*f* units that are in turn decorated primarily at O3 with α-L-Ara*f* residues. Arabinan is found mainly in apple juice and sugar beet (Karimi and Ward 1989, Tanaka, Abe and Uchida 1981). Galactans are undecorated polymers of β1,4-linked D-Gal residues. Galactans are particularly abundant in RGI derived from potato and lupins (Ridley, O'Neill and Mohnen 2001, Mohnen 2008, McKie et al. 2001, Oomen et al. 2002, Ji et al. 2004). The backbone of both arabinans and galactan containing 100-200 sugars (Oomen et al. 2002, Ji et al. 2004, Mohnen 2008).

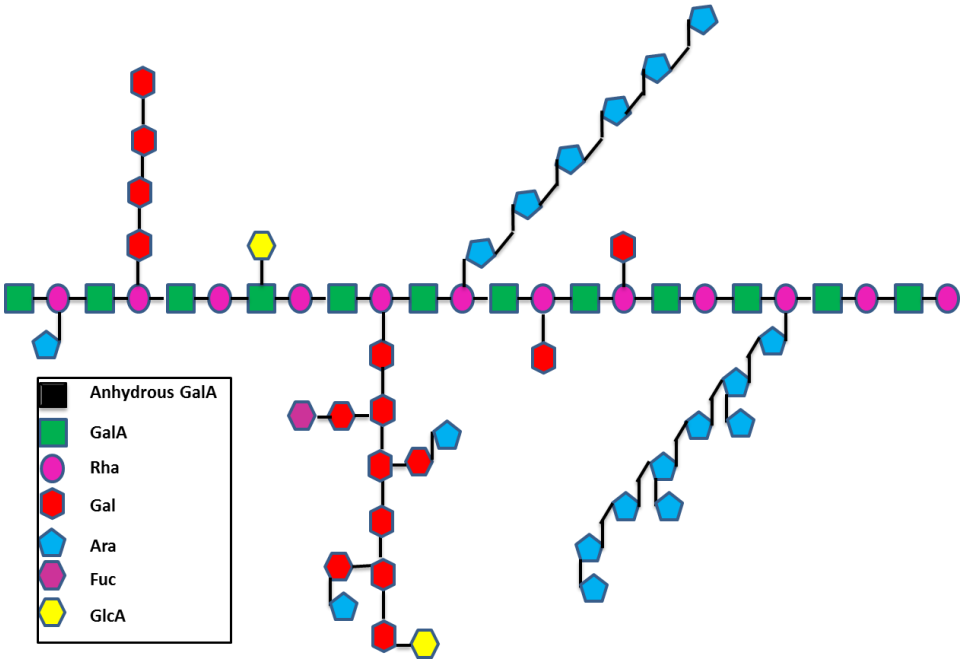


Figure 1.7 Schematic structure of RGI. (GalA, galacturonic acid; Rha, Rhamnose; Gal, Galactose; Ara, Arabinose; Fuc, Fucose; GlcA, Glucuronic acid.)

In the leaf and root of *Trifolium pretense L.*, RGI is located in the middle lamella, 80%-90% of which is associated with the expanded region of the middle lamella in the corner junctions between cells (Moore and Staehelin 1988). The Golgi apparatus is the main site of synthesis of the non-cellulosic cell-wall polysaccharides.

Recently, increasing evidence has indicated that changes in the structures of the arabinan and galactan side chains are associated with different stages of development (McCartney et al. 2000, Oomen et al. 2002, Gomez et al. 2009). Transgenic studies involving the *in vivo* expression of endoglycanases showed that cleaving the RGI backbone and/or side chains can have significant effect on plant growth and development (Oomen et al. 2002).

1.1.3.3 Rhamnogalacturonan II (RGII).

1.1.3.3.1 RGII is the most complex polysaccharide known

The complexity of RGII is unmatched by any other plant polysaccharides so far described. RGII is highly branched pectin (Figure 1.8). It consists of a backbone of □-1,4 linked GalA residues with four discrete oligosaccharide side chains. It contains 11 different glycosyl residues and 20 different glycosyl linkages (**Figure 1.9**). Rhamnose (Rha) is the most common sugar found in RGII side chains (14 mol %) (Harholt et al. 2010, ONeill et al. 1996).

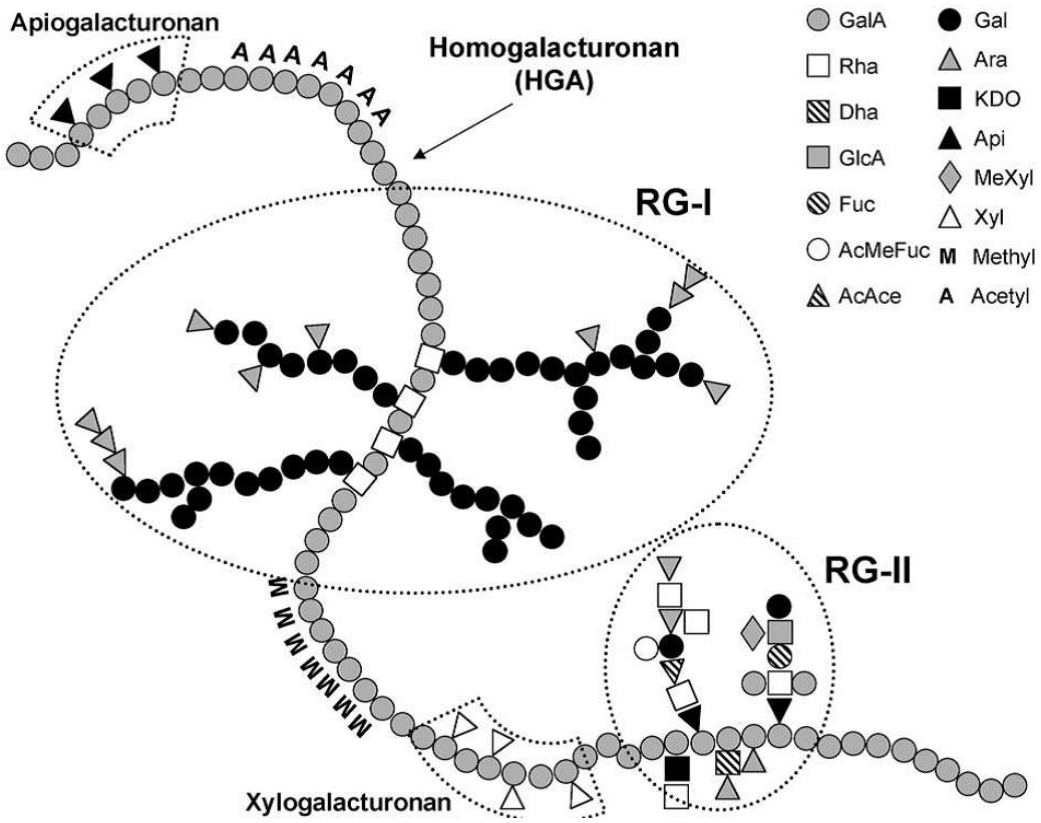


Figure 1.8 Schematic representation of the structure of pectin. HGA, homogalacturonan. **(Perez, Rodriguez-Carvajal and Doco 2003) (GalA, galacturonic acid; Rha, Rhamnose; Gal, Galactose; Ara, Arabinose; Fuc, Fucose; Xyl, xylose; Api, Apiose; GlcA, Glucuronic acid; KDO, keto sugar; DHA, 3-deoxy-D-lyxo-2-heptulosaric acid.)**

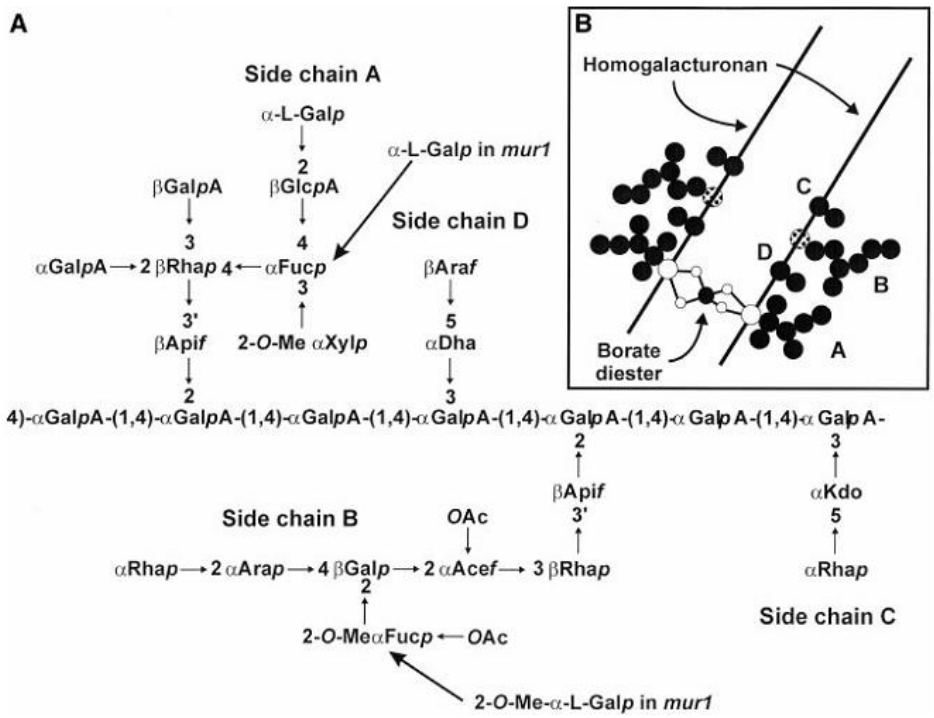


Figure 1.9 The structure of RGII . (A) RGII structure. (B) A borate ester covalently links two RGII molecules between the apiosyl residues.(O'Neill et al. 2001)

RGII contains several unusual sugar components. Aceric acid is the only branched, acidic deoxy monosaccharide in nature. The KDO (keto sugar), methyl-Xyl, methyl-Fuc and DHA (3-deoxy-D-lyxo-2-heptulosaric acid), in plants, are only found in RGII (Harholt et al. 2010). The complex and unique features of RGII (5-10 kDa) are highly conserved among dicotyledons, monocotyledons, gymnosperms, pteridophytes, and lycophytes, which indicates that the polysaccharide may play a very important role in plant growth. In plant cell walls RGII exists either in a monomer or as a dimer in which two molecules of the polysaccharide are held together by a borate cross link (**Figure 1.9**) (Harholt et al. 2010, Albersheim 1989, Melton et al. 1986, Mohnen 2008). Dimerized RGII also contains heavy metals such Pb^{2+}, Ba^{2+}, and St^{2+}. In red wine, RGII exists as a dimer at pH 3.5-3.8. (Perez et al. 2003).

1.1.3.3.2 RGII is essential for plant growth

Studies have shown that the lack of RGII dimer causes severe growth defects. The *Arabidopsis* pleiotropic mutant (mur1) induces dwarfism and fragile cell walls (**Figure 1. 10**) (O'Neill et al. 2001). In mur1 the gene responsible for GDP-L-fucose synthesis in RGII is mutated. Watering the mutant with fucose or boron restored normal growth, which suggests that lack of fucose in the RGII structure affects dimer formation and boron can restore oligomerization (O'Neill et al. 2001).

A recent study showed that the formation of RGII dimer with boron is directly related to the thickness of the plant cell wall (Ishii, Matsunaga and Hayashi 2001). Growing pumpkin in the absence of boron leads to significant swelling of the middle lamellae of cell walls and dissociates the dimeric from of RGII into monomers (**Figure 1. 11**). Glycosyl residue and glycosyl linkage analysis showed that there is no significant difference in the primary structures of RGII dimers and monomers. Adding boric acid to the boron-deficient pumpkins rescues the abnormal plants and helps the conversion of RGII monomer to dimer within five hours (Ishii et al. 2001).

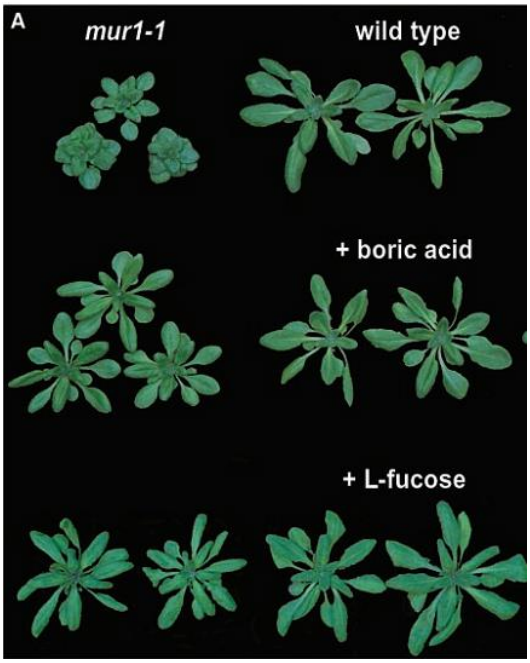


Figure 1. 10 Knockout mutant showed that RGII is essential for plant healthy growth. (The gene for GDP-L-fucose synthesis is mutated in mur1-1) **(O'Neill et al. 2001)**

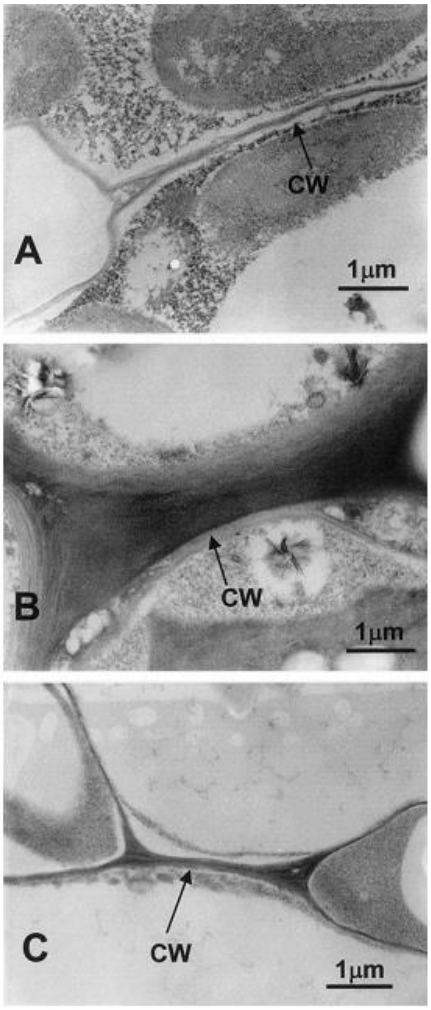


Figure 1. 11 Borate treatment to the pumpkin middle lamellae area. A, B-normal; B, boron-deficient (thickened cell wall in dark shadow is visible, as pointed by the arrow labeled as CW); C, boron-deficient treated with boron for 5 h. CW, cell wall. **(Ishii et al. 2001).**

1.1.3.3.3 RGII is highly recalcitrant to degradation

Due to the complex structure, RGII is very difficult for microbes to degrade. An important example of this is the red wine produced from prolonged fermentation, during which almost all the polysaccharides in the grape cell wall are degraded except for RGII; this is why red wine is a very good source from which to extract large quantities of the branched pectin (Pellerin et al. 1996). White wine, however, generally comes from the fermentation of grape juice without any skins and thus does not contain much RGII.

Recent studies, however, showed that the human gut bacterium, *Bacteroides thetaiotaomicron*, can grow in media where RGII is the only carbon source, and transcriptomics identified the bacterial genes upregulated by the polysaccharide (Martens et al. 2011). However, the mechanism by which *B. thetaiotaomicron* mediates RGII degradation is unknown. Such information is not only important in understanding how the human large bowel microbes can utilize dietary glycans as growth substrates, but may provide a platform for developing novel prebiotics.

1.2 Carbohydrate-Active enZYmes and its Database CAZY

The online database CAZy (http://www.cazy.org/) groups carbohydrate related enzymes and non-catalytic carbohydrate binding modules (CBMs) into sequence-based families. All members of a given family display the same fold, while catalytic apparatus and mechanism are similarly conserved within a family (Henrissat and Davies 1997, Coutinho et al. 2003, Cantarel et al. 2009b). Substrate specificity, however, can be highly variable, exemplified by glycoside hydrolase family (GH) 2 and 43, although the stereochemistry of the target linkages is conserved within a GH. Four types of polysaccharide modifying enzymes are categorized: glycoside hydrolases, which hydrolyse glycosidic bonds; glycosyl transferases (GTs), which catalyze the synthesis of glycosidic bonds; polysaccharide lyases (PLs), which mediate non-hydrolytic cleavage of glycosidic bonds and carbohydrate esterases (CEs) that hydrolysis carbohydrate esters. The CAZy database also groups CBMs into sequence-based families (Boraston et al. 2004). With respect to CBMs the fold and, generally, the site of ligand recognition is conserved. In some CBM families ligand specificity is invariant, while in other families, exemplified by CBM6, the target glycans can vary between different members (Montanier et al. 2010, Correia et al. 2010, Cantarel et al. 2009b).

Currently, 44 of the 133 GHs contain enzymes that deconstruct plant cell walls. Crystal structures of relevant enzymes in 41 of these 44 GHs have been reported. Structural information is also available for 54 out of the 71 CBM families.(Herve et al. 2010, Gilbert, Knox and Boraston 2013). Indeed, structural data are available for the five families; 1, 2a, 3, 5, 10 and 63 that mediate binding to crystalline cellulose (Boraston et al. 2004, Gilbert et al. 2013).

1.2.1 Glycoside hydrolases

Glycoside hydrolases are enzymes that catalyze the hydrolysis of glycosidic bonds. Those that cleave substrate at termini or within the middle of a polysaccharide chain are called exo- and endo- enzymes, respectively. Glycoside hydrolases cleave glycosidic bonds principally via two different mechanisms leading to inversion (α to β or vice versa) or retention (α to α or β to β) of configuration at the anomeric carbon. The retaining mechanism operates via a two-step, double-displacement mechanism. In the first step a carboxylate functions as the nucleophile and attacks the anomeric carbon of the glycone sugar, while leaving group departure is promoted by the glycosidic oxygen accepting a proton from the second catalytic carboxylate of the enzyme that serves as a general acid in this instance. As a result a covalent glycosyl-enzyme intermediate is formed. In the second step the proton donating carboxylate now acts as a general base extracting a proton from the incoming water molecule, which then acts as a nucleophile and hydrolyzes the glycosyl-enzyme intermediate (Zechel and Withers 2000). (**Figure 1. 12**). Inverting hydrolases cleave glycosidic bonds with net inversion of anomeric configuration, and catalysis is mediated through a one-step, single-displacement mechanism. In both mechanisms the transition state is believed to be an oxocarbenium ion (Zechel and Withers 2000). The distance between the two catalytic carboxylates can be used to suggest the mechanism of the reaction: 5.5 Å and ~10 Å in retaining and inverting enzymes, respectively.

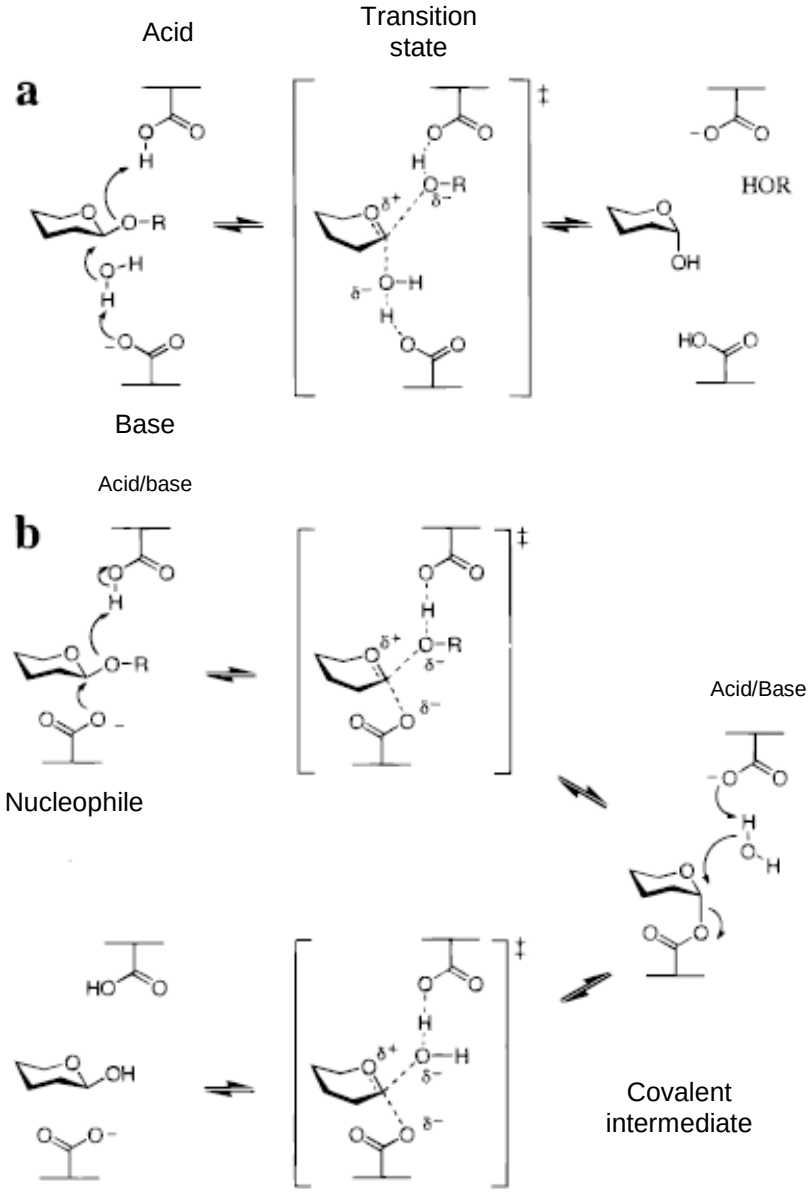


Figure 1. 12 General mechanisms for inverting (a) and retaining (b) glycosidases. Adapted from (Zechel and Withers 2000).

Currently there are 133 GHs, many of which have been grouped into 14 clans (**Figure 1. 13**). Each clan contains a number of GHs that are believed to have shared a common ancestral sequence. As the largest clan, GH-A contains enzymes from GH1, 2, 5, 10, 17, 26, 30, 35, 39, 42, 50, 51, 53, 59, 72, 79, 86 and 113. The enzymes in GH-A display a $(\beta/\alpha)_8$-barrel fold, target β-linkages via a double displacement mechanism (Jenkins et al. 1995, Henrissat et al. 1995).

As shown in **Figure 1. 14**, the active site of glycoside hydrolases usually adopt three general topologies (Henrissat and Davies 1997): exo-acting enzymes display a pocket topology, endo-acting enzymes contain an open active site cleft which can accommodate a single polysaccharide chain, and cellobiohydrolases exhibit a tunnel topologies that are similar to the open cleft topology of endo-acting enzymes, except that extended loops from the two sides of the cleft interact to form the ceiling of the tunnel (Henrissat and Davies 1997). Exo-acting enzymes usually remove terminal decorations from the side chains of polysaccharides providing better access for the endo-acting glycoside hydrolases, which cut randomly along polysaccharide chains generating oligosaccharides with different sizes, as best illustrated on a TLC plate as a "product smear"(Henrissat et al. 1995). Cellobiohydrolases move along cellulose chains, displaying a processive mode of action releasing cellobiose after each catalytic event (Rouvinen et al. 1990). However, the loops forming the tunnel in these enzymes can open, and thus cellobiohydrolases mediate infrequent cleavage of internal glycosidic linkages (Armand et al. 1997).

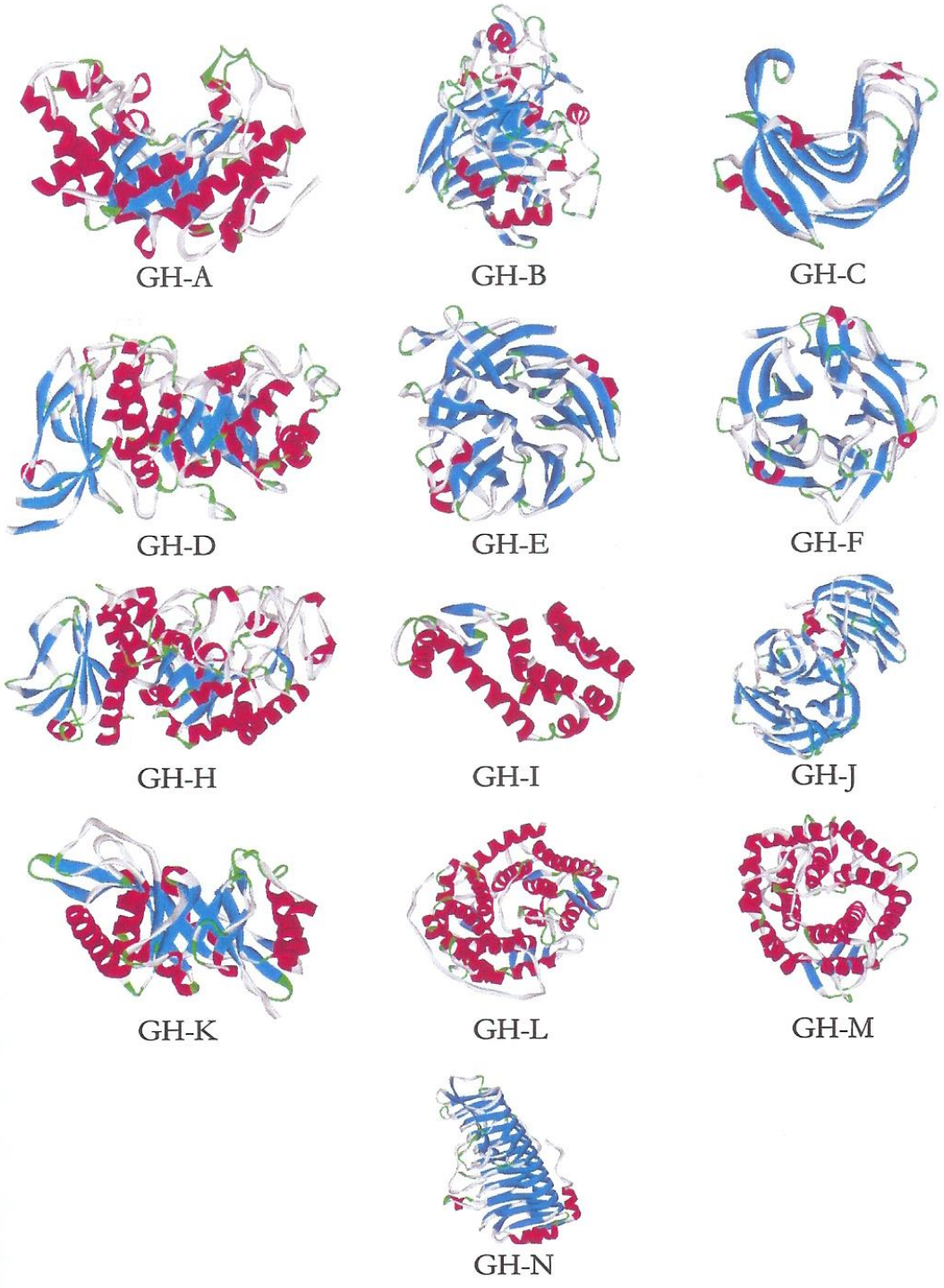


Figure 1. 13 Cartoon representation shows example structures of GH clans A-N.
β strands are shown in blue and α helices in red. The rest and the loop are labeled in green and grey (Pell 2004).

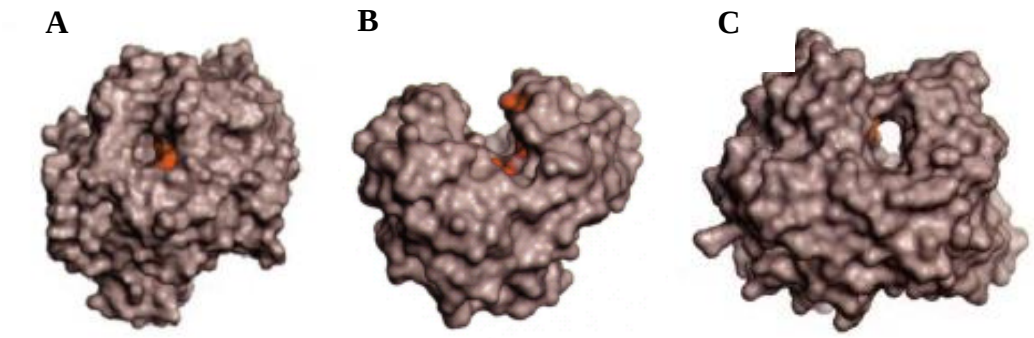


Figure 1. 14 The three general active site topologies adopted by glycoside hydrolases. (A) 'pocket' topography displayed by non-processive exo-acting enzymes (B) ' open cleft' topography, found in endo-acting enzymes. (C) 'tunnel' topography, presented by processive exo-acting cellobiohydrolases (Henrissat and Davies 1997, Davies and Henrissat 1995). Catalytic residues are shaded in red.

Subsites of the substrate binding clefts (which houses the active site and binds sugars distal to the catalytic centre) of ***endo***-acting glycoside hydrolases can accommodate more than one sugar residue in an oligo- or polysaccharide. A nomenclature for consistently describing these subsites, which vary in number and topography from enzyme to enzyme, has been developed (Davies and Henrissat 1995). Subsites are numbered from –n to +n with –n binding the non-reducing end of the sugar polymer, and +n the reducing end. Cleavage occurs between the -1 (active site) and +1 subsites.

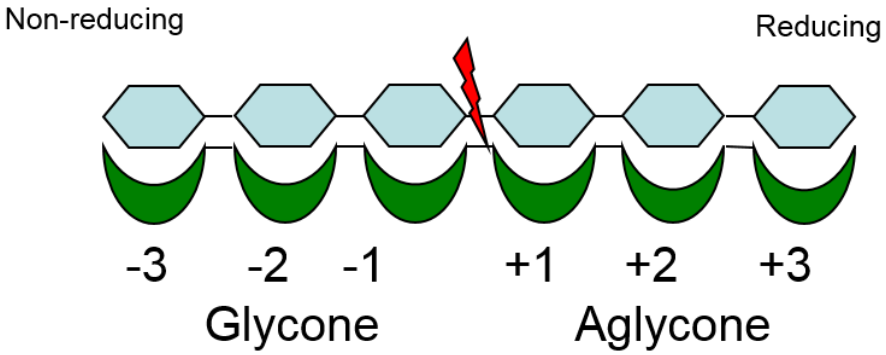


Figure 1. 15 Sugar residues bind to subsites within the active site of glycoside hydrolases. The polarity of the sugars (blue) and the subsites (green) is indicated. The glycosidic bond, which is hydrolysed, is indicated in red.

1.2.2 Polysaccharide lyases and Δ4,5-GalA specific glycoside hydrolases

Polysaccharide lyases are enzymes that cleave uronic acid-containing polysaccharide chains via a β -elimination mechanism to generate an unsaturated uronic acid and a new reducing end (**Figure 1.16**). These enzymes are produced by many bacteria and some phytopathogenic fungi. Endo-acting pectate lyases are more abundant than exo-pectate lyases. So far 22 polysaccharide lyase families (PLs) are available on the CAZy website (http://www.cazy.org/Polysaccharide-Lyases.html), and pectate lyases are located in PLs 1, 2, 3, 9 and 10 (Lombard et al. 2010b, Garron and Cygler 2010). A parallel β-helical topology is common to pectate lyases (Garron and Cygler 2010), although several novel structures have also been reported including a family 10 pectate lyase displaying a (α/α)6 toroid fold (Charnock et al. 2002, Novoa De Armas et al. 2004) and a family 2 lyase from *Yersinia enterocolitica* displays a rare (α/α)7 barrel fold (Abbott and Boraston 2007). Significant structural similarities have been observed when superimposing the catalytic centre of structurally unrelated pectate lyases, highlighting the conserved β-elimination mechanism (Charnock et al. 2002). Ca^{2+} ions are also required to bind directly to the enzymes and link the substrate to the proteins (Herron et al. 2003), but a recent study also suggested the trend for utilizing Mn^{2+} or Ni^{2+} instead of Ca^{2+} in the general transition metal-assisted β-elimination(Abbott and Boraston 2007). The products generated by pectate lyases contain a Δ4,5-GalA at the non-reducing end, and thus unsaturated uronic acids are removed by specific glycoside hydrolases belonging to GH105, which display a unique mechanism of action (Jongkees and Withers 2014).

.

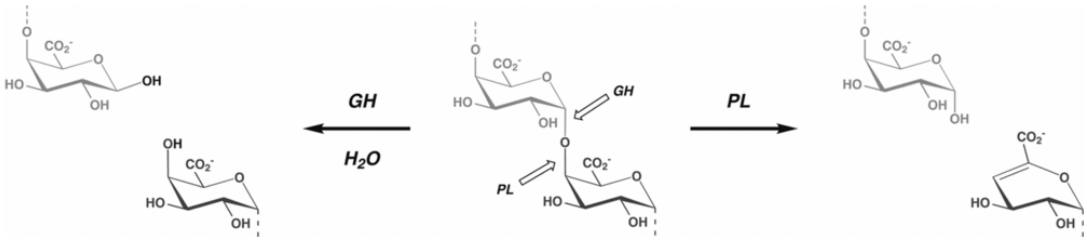


Figure 1.16 Comparison of the products of PL and GH, exemplified by polygalacturonate (pectate) cleavage (Lombard et al. 2010b).

1.2.3 Carbohydrate Esterases

Carbohydrate esterase is a group of enzymes that catalyse the hydrolysis de-O or de-N-acylation of substituted saccharides. Currently 16 families have been categorized in the CAZy database (http://www.cazy.org/Carbohydrate-Esterases.html). The crystal structures of carbohydrate esterases reveal mainly a classical β/α/β "serine protease" or α/ β-hydrolase fold (Davies, Gloster and Henrissat 2005). The most common catalytic is a Ser-His-Asp catalytic triad that catalyzes deacetylation analogous to the action of classical lipase and serine proteases, but other mechanisms such as a Zn^{2+} catalyzed deacetylation prevail in some families (Hakulinen, Tenkanen and Rouvinen 2000).

1.2.4. Carbohydrate binding modules (CBMs)

CBMs are often appended to carbohydrate active enzymes that attack recalcitrant polysaccharides such as those found in the plant cell wall. To date the CAZy database have categorized 67 different CBM families (http://www.cazy.org/Carbohydrate-Binding-Modules.html). CBMs are functionally classified into three categories (types A, B, and C) based on the topology of their ligand-binding sites and their mode of ligand recognition (Boraston et al. 2004).

Type A CBMs bind to insoluble, highly crystalline cellulose and/or chitin, via a "planar platform" binding site that is composed of aromatic amino acids. The planar architecture is

believed to be complementary to the flat surfaces of its ligand (McLean et al. 2000, Tormo et al. 1996). Although type A CBMs are present in a range of enzymes (cellulases, mannanases, xylanases, pectinases and esterases) and CBM families (family 1, 2a, 3, 5 and 10), currently the specificity of type A modules is believed to be conserved (Boraston et al. 2004). Type B CBMs bind to single glycan chains. Their binding sites can interact with up to six sugars but display negligible binding to oligosaccharides with a degree of polymerization (DP) of three or less. The binding sites are normally grooves or clefts, located on the concave surface (see below) of the proteins with a topology complementary to the conformation of the ligand. The depth of the groove varies from very shallow in *Cc*CBM17, to sufficiently deep to accommodate the entire width of a pyranose ring (Boraston et al. 2002, Notenboom et al. 2001). Unlike type A modules, the specificity of type B CBMs is highly variable, targeting amorphous cellulose, xylan, β-1,3-glucans, β-1,3-1,4- mixed linkage glucans, galactomannan, glucomannan and starch (Boraston et al. 2004). Type C CBMs bind to mono-, di- or tri-saccharides, and a review in 2013 reclassified type C CBMs as the group that binds to the termini of glycans (exo-type), in contrast with Type B CBMs which bind internally on glycan chains (endo-type) (Gilbert et al. 2013). Type C CBMs have been reported in families 6, 9, 13, 14, 18, 32, 35, 62(Boraston et al. 2004, Gilbert et al. 2013, Montanier et al. 2011).

The crystal or NMR structures of CBMs to date have revealed seven folds (Figure 1.17) (Boraston et al. 2004). The most common is a β-sandwich structure, composed of two β-sheets of three to six antiparallel β-strands, which is stabilized by a structural calcium ion. These β-sandwich structures are found in CBM families 2, 3, 4, 6, 9, 11, 15, 16, 17, 20, 21, 22, 25, 26, 27, 28, 29, 32, 34, 35, 36, 40, 41, 42, 44, 47, 48, 58, 60, 61 and 68 (Cantarel et al. 2009a, Lombard et al. 2014). The second most frequent fold is the β-trefoil fold (Murzin, Lesk and Chothia 1992), exemplified by family 13 CBMs, and has 12 stranded β-sheet that form six

hairpin turns. These proteins contain three distinct sugar binding sites and can display a type B or type C binding mode (Cantarel et al. 2009a, Lombard et al. 2014, Gilbert 2003, Gilbert et al. 2013, Boraston et al. 2004).

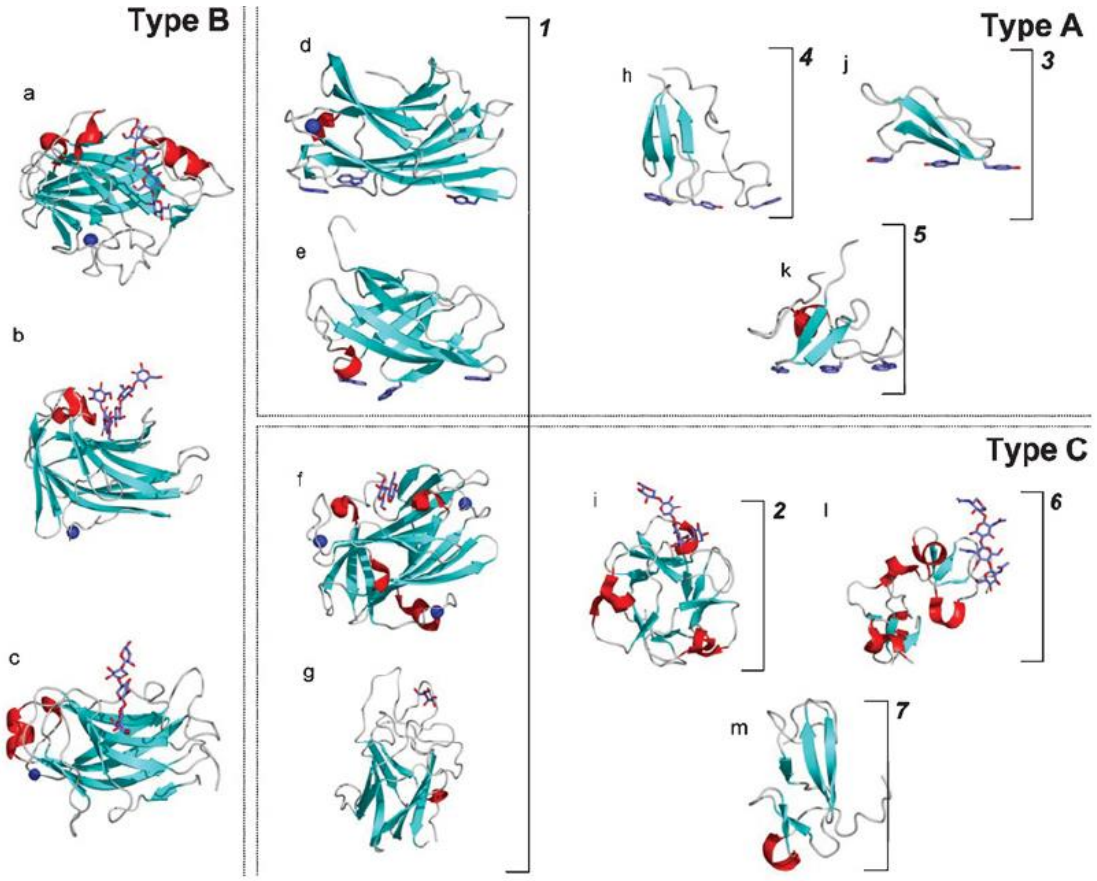


Figure 1.17 Fold relationships among CBMs.

Dotted boxes surround examples of CBMs belonging to the functional types A, B, and C. Brackets with numbers indicate examples of CBMs belonging to fold families 1–7. CBMs shown are as follows: (a) family 17 CBM, CcCBM17, from *Clostridium cellulovorans* in complex with cellotetraose ; (b) family 4 CBM, TmCBM4-2, from *Thermotoga maritima* in complex with laminariohexaose; (c) family 15 CBM, *Cj*CBM15, from *Cellvibrio japonicus* in complex with xylopentaose; (d) family 3 CBM, CtCBM3, from *Clostridium thermocellum*; (e) family 2 CBM, CfCBM2, from *Cellulomonas fimi*; (f) family 9 CBM, TmCBM9-2, from *Thermotoga maritima* in complex with cellobiose; (g) family 32 CBM, MvCBM32, from *Micromonospora viridifaciens* in complex with galactose; (h) family 5 CBM, EcCBM5, from *Erwinia chrysanthemi*; (i) family 13 CBM, SlCBM13, from *S. lividans* in complex with xylopentaose; (j) family 1 CBM, TrCBM1, from Trichoderma reesi; (k) family 10 CBM, *Cj*CBM10, from *Cellvibrio japonicus*; (l) family 18 CBM from *Urtica dioca* in complex with chitotriose; (m) family 14 CBM, tachychitin, from *Tachypleus tridentatus*. Bound ligands are shown as 'liquorice' representations, while bound metal ions are shown as a blue spheres (reproduced from Boraston et al, 2004)(Boraston et al. 2004)

The mechanism by which CBMs potentiate enzyme catalysis is not completely understood. It has been suggested that type A modules slide across the surface of cellulose (Jervis, Haynes and Kilburn 1997), which facilitates substrate access. Type B and type C CBMs lock onto specific ligands and thus direct the enzyme to its target substrate (Kellett et al. 1990). A recent study also confirmed that CBMs boosted enzyme activity (pectinases and xylanases*) in vivo* by targeting and proximity effects (Herve et al. 2010), as the action of the cognate catalytic module toward polysaccharides, in intact cell walls, could be potentiated through CBM recognition of both substrate and non-substrate polysaccharides. This is consistent with the numerous observations that removal of the CBM reduces the activity of glycoside hydrolases and polysaccharide lyases against insoluble substrates, but not against soluble substrates (Boraston et al. 2004, Henshaw et al. 2004). A CBM35 was reported to bind to Δ4,5–GalA, a signature product of pectin degradation, and it was suggested that this enabled the cognate catalytic module to be targeted to regions of the plant cell wall that is being actively digested (Gilbert et al. 2013, Montanier et al. 2009b). In 2012, a type C CBM from *Bacillus subtilis* was also reported to confer enzymatic substrate specificity to a non-specific GH32 fructosidase (Cuskin et al. 2012). The model proposed that CBM and the catalytic module bind in concert to the multiple termini on branched polysaccharides and thus enhance affinity, specificity and therefore catalytic activity (Cuskin et al. 2012, Gilbert et al. 2013). Besides, a "disruption" mechanism has also been proposed for some CBMs that increases enzyme access to highly crystalline polysaccharides such as cellulose and crystalline form of starch (amylose) (Din et al. 1994, Southall et al. 1999). It should be noted, however, that the increase in activity mediated by these discrete CBMs was very modest (less than 30%). In 2005, a CBM33 was reported to disrupt the surface of chitin and potentiate the activity of chitinases (Vaaje-Kolstad et al. 2005), but this family of CBM33 was later shown to be a copper-dependent lytic

polysaccharide monooxygenase and thus reclassified (Vaaje-Kolstad et al. 2010). Therefore, on balance it seems that, in general, CBMs potentiate enzyme activity by bringing the biocatalysts into close proximity with the target substrate.

Mannan specific CBMs are generally type B modules that recognize single glycan chains CBMs (Boraston et al. 2004), consistent with nature of their target ligands. CBM35, for example, is *from Cellvibrio japonicus* Man5C and binds to galactomannan, glucomannan, manno-oligosaccharides and amorphous insoluble mannan, but not crystalline mannan. Other mannan specific CBMs exemplified by CBM27 from ***Thermotoga maritima*** binds to mannan with affinity 100 times higher than CBM35 and was used in the following mannan project.

1.3 Mannanase

1.3.1 *Cj*Man5A and *Cj*Man26A mannanases play different roles

β-1,4-Mannanases hydrolyse mannans and glucomannans, and are generally located in GH5 and GH26. The molecular architecture and biochemical properties of the *Cellvibrio. japonicus* mannanases have been extensively studied (Hogg et al. 2003). The genome of *Cellvibrio. japonicus* encodes six mannanases, three from both GH5 (Man5A, Man5B and Man5C) and GH26 (Man26A, Man26B and Man26C), which all contain signal peptides and are thus secreted out of the cytoplasm (Hogg et al. 2003, Braithwaite et al. 1995, Bolam et al. 1996, Cartmell et al. 2008). Man26A, Man26B and Man26C consist of a single catalytic module and contain type II signal peptides suggesting that these enzymes are bound to the lipid membrane and are likely presented on the surface of *Cellvibrio. japonicus*. Man5A, Man5B and Man5C contain type A CBMs belonging to families 2a (Man5B), 5 (Man5C) and 10 (all three GH5 mannanases), while Man5C also contains a type B CBM35 that binds mannan chains (Bolam et al. 2004). The family 10 and 2a CBMs bind to crystalline cellulose and ivory nut crystalline

mannan. The catalytic domains of Man5A, Man5B and Man26B degrade galactomannan and glucomannan, but not crystalline mannan or cellulosic substrates. Man5C can degrade crystalline ivory nut mannan but its activity against glucomannan and galactomannan was lower (Hogg et al. 2003). Man5A, Man5B, Man5C, Man26A and Man26B exhibit classical endo-activity. By contrast Man26C is an exo-acting mannanase that releases mannobiose from the non-reducing end of mannans (Cartmell et al. 2008). It is proposed that since the GH5 enzymes contain crystalline polysaccharide binding CBMs, they are likely to target mannans that are integral to the plant cell wall, while the GH26 enzymes, which lack CBMs, target the storage form of the polysaccharide, galactomannan, and manno-oligosaccharides (Hogg et al. 2003).

1.3.2 Recognition of Heterogeneous Substrates by Diverse β-Mannanases

It is not clear how the mannanases manifest specificity toward heterogeneous substrates, exemplified by the plant structural polysaccharide glucomannan, which comprises a backbone of a random sequence of β-1,4-linked glucose and mannose units. β-Mannanases, from both GH5 and GH26 hydrolyze glucomannan. However, the mechanism by which these enzymes select for glucose or mannose at distal subsites is currently unclear. (Tailford et al. 2009)reported the biochemical properties and crystal structures of both a GH5 & GH26 mannanase, and described the contributions of distal subsites to substrate specificity in these enzymes. Two types of enzyme were studied: BaMan5A, derived from *Bacillus agaradhaerens*, and Man26B from *Cellvibrio. japonicus*, can accommodate glucose or mannose at both their -2 and +1 subsites. In contrast, Man26A and Man26C from *Cellvibrio. japonicus* display tight specificity for mannose at their two negative subsites (Tailford et al. 2009).

The crystal structure of BaMan5A reveals that a polar residue at the -2 subsite makes productive contact with O2 of the bound sugar in either its axial (as in mannose) or its

equatorial (as in glucose) configuration, while other distal subsites do not exploit O2 as a specificity determinant. Thus, BaMan5A is able to hydrolyze glucomannan with a variable sequence of glucose and mannose (Tailford et al. 2009). The crystal structure of Man26A and Man26C reveal that tight specificity for mannose at the -2 subsite is mediated by polar interactions, dominated by a conserved arginine with the axial O2 of the sugar (Tailford et al. 2009, Cartmell et al. 2008, Hogg et al. 2001). Variants of Man26A in which the residues in the -2 subsite that interact with O2 of bound Man was unable to distinguish between Man and Glc at this distal subsite, but these mutations resulted in a 1000-fold reduction in activity (Tailford et al. 2009).

1.4 Pectin degrading enzymes

The study of pectin degradation started several decades ago focusing on microbial plant parasites and the infestation that occurs at the plant surface (Carpita and Gibeaut 1993, Alghisi and Favaron 1995, Ried and Collmer 1986). During colonisation, breakdown of the plant cell wall is essential not only for parasite invasion of tissues, but also for release of usable carbohydrates, or accessibility to protoplast and cell killing (Alghisi and Favaron 1995). A number of endo-and exo-acting polygalacturonases and pectate lyases were studied and characterized (Alghisi and Favaron 1995, Dow et al. 1989). In 2010, Professor Bernard Henrissat reviewed all the pectic lyases categorized within the online database CAZy (Lombard et al. 2010a), **Table 1.1**.

CAZy family	Subfamily	Known activities	Characterized enzymes
PL1			
	1	Endo-polygalacturonate (pectate) lyase (EC 4.2.2.2)	4/173
	2	Endo-polygalacturonate (pectate) lyase (EC 4.2.2.2)	1/53
	3	Endo-polygalacturonate (pectate) lyase (EC 4.2.2.2)	10/38

	4	Pectin lyase (EC 4.2.2.10)	18/40
	5	Endo-polygalacturonate (pectate) lyase (EC 4.2.2.2)	16/38
		Exo-polygalacturonate lyase (EC 4.2.2.9)	4/38
	6	Endo-polygalacturonate (pectate) lyase (EC 4.2.2.2)	19/66
	7	Endo-polygalacturonate (pectate) lyase (EC 4.2.2.2)	7/22
	8	Pectin lyase (EC 4.2.2.10)	4/22
PL2			
	1	Endo-polygalacturonate (pectate) lyase (EC 4.2.2.2)	4/21
	2	Exo-polygalacturonate lyase (EC 4.2.2.9)	2/23
PL3			
	1	Endo-polygalacturonate (pectate) lyase (EC 4.2.2.2)	3/137
	2	Endo-polygalacturonate (pectate) lyase (EC 4.2.2.2)	16/115
	3	Endo-polygalacturonate (pectate) lyase (EC 4.2.2.2)	1/8
PL4			
	1	Rhamnogalacturonan lyase (EC 4.2.2.-)	2/21
	2		0/15
	3	Rhamnogalacturonan lyase (EC 4.2.2.-)	1/7
	4	Rhamnogalacturonan lyase (EC 4.2.2.-)	1/6
PL5			
	1	Poly(β-D-mannuronate) lyase (alginate lyase) (EC 4.2.2.3)	9/27
PL6		Poly(β-D-mannuronate) lyase (alginate lyase) (EC 4.2.2.3)	1/21
		Chondroitin-sulfate-ABC endolyase (EC 4.2.2.4)	1/21
PL7			
	1	Poly(β-D-mannuronate) lyase (alginate lyase) (EC 4.2.2.3)	1/20
	2		0/7
	3	Poly(α-L-guluronate) lyase (EC 4.2.2.11)	2/8
	4		0/5
	5	Poly(α-L-guluronate) lyase (EC 4.2.2.11)	4/17
PL8			
	1	Hyaluronate lyase (EC 4.2.2.1)	11/85
	2	Chondroitin-sulfate-ABC endolyase (EC 4.2.2.20)	3/30
	3	Chondroitin AC lyase (EC 4.2.2.5)	3/7
PL9			
	1	Endo-polygalacturonate (pectate) lyase (EC 4.2.2.2)	6/66
		Exo-polygalacturonate lyase (EC 4.2.2.9)	2/66
	2		0/18
PL10			

	1	Endo-polygalacturonate (pectate) lyase (EC 4.2.2.2)	5/47
PL11			
	1	Rhamnogalacturonan lyase (EC 4.2.2.-)	4/64
PL12			
	1		0/35
	2	Heparin-sulfate lyase (EC 4.2.2.8)	2/10
PL13		Heparin lyase (EC 4.2.2.7)	2/7
PL14			
	1	Glucuronate lyase (EC 4.2.2.-)	1/8
	2		0/6
	3	Poly(β-D-mannuronate) lyase (alginate lyase) (EC 4.2.2.3)	1/13
		Exo-oligo-alginate lyase (EC 4.2.2.-)	1/13
PL15			
	1	Oligo-alginate lyase (EC 4.2.2.-)	3/5
PL16			
	1		0/14
	2	Hyaluronate lyase (EC 4.2.2.1)	2/8
	3	Hyaluronate lyase (EC 4.2.2.1)	1/10
PL17			
	1		0/14
	2	Poly(β-D-mannuronate) lyase (alginate lyase) (EC 4.2.2.3)	1/15
PL18		Poly(β-D-mannuronate) lyase (alginate lyase) (EC 4.2.2.3)	3/6
		Poly(α-L-guluronate) lyase (EC 4.2.2.11)	3/6
PL20		Endo-β-1,4-glucuronan lyase (EC 4.2.2.14)	1/11
PL21		Heparin lyase (EC 4.2.2.7)	1/9
		Heparin-sulfate lyase (EC 4.2.2.8)	1/9
PL22		Oligogalacturonate lyase (EC 4.2.2.6)	1/47

Table 1.1 Activities in PL families and subfamilies. Number of characterized enzymes in each subfamily is listed on the right column.

The backbone of the pectin can be digested by glycoside hydrolases (especially GH28) and polysaccharide lyases. Endo- and exo-polygalacturonase (GH28) cleave the backbone of the smooth regions (homogalacturonan), while the more intricate backbone of the hairy regions (RGI) are attacked by endo- and exo-rhamnogalacturon hydrolases (GH28) and other related enzymes (Martens-Uzunova and Schaap 2009). Enzymes in fungi have been studied intensively. *A. niger*, for example, encodes seven methylation sensitive pectate lyases that differ in their mode of action (Martens-Uzunova and Schaap 2009, Benen et al. 2000, van den

Brink and de Vries 2011, Bussink et al. 1991, Parenicova et al. 2000). In another organism , *Rhizopus. oryzae*, 15 of the 18 putative GH28 polygalcturonases showed different substrate specificities for polygalacturonan, trigalacturonan, and digalacturonan, and their activity against polygalacturonan ranged from less than 1 to 200 μmol/min/mg protein (Mertens and Bowman 2011). Two rhamnogalacturonan hydrolases have also been reported in *A. niger*, and the different digestion pattern suggested that each enzyme acted on a structurally distinct regions of the substrate (Suykerbuyk et al. 1997).

Six pectin lyases and only one pectate lyase have been identified and partially characterized in *A. niger* (Benen et al. 2000, Harmsen, Kusters-van Someren and Visser 1990). Pectin lyases have preference for substrates with a high degree of methylesterification, whereas pectate lyases prefer substrates with a low degree of esterification and need Ca^{2+} for catalysis. Rhamnogalacturonan lyases differ substantially in their structure from pectin and pectate lyases, and cleave within the hairy regions of pectin. This group of lyases usually belongs to two families, PL4 and PL11, where PL4 lyases have a much lower optimum pH than PL11 lyases (Jensen et al. 2010). One PL4 rhamnogalacturonan lyase from *Aspergillus aculeatus* has been characterized, and the cleavage preferably occurs four residues from L-rhamnose at the reducing end. The activity of the enzyme is severely affected by the presence of acetyl groups in the backbone of Rhamnogalacturonan (Mutter et al. 1998, de Vries et al. 2000). Therefore it is speculated that esterases are needed to deconstruct the backbone of RGI. In concert with the action of the polysaccharide lyases, as described above, GH105 glycoside hydrolases are required to remove the Δ4,5-GalA at the non-reducing end. Finally, with respect to RGI rhamnosidases are required to remove terminal L-Rha units generated by the endo-acting enzymes. Although GH78 and GH106 contain α-L-rhamnosidases, the capacity of enzymes from these families to attack RGI is unknown.

RGI depolymerization also require accessory enzymes (for example, GH43 endo-α-1,5-arabinanases, exo-arabinanases, and GH53 endo-β-1,4-galactanases) to remove the side chains and provide access for the main-chain hydrolyzing pectinolytic enzymes. To date, a model for the complete digestion of RGI, has not been reported, which is a major focus of this thesis.

1.5. Key bacteria used in the project to degrade polysaccharide.

A significant number of bacteria contain polysaccharide-degrading genes and can thus utilize a variety of complex carbohydrates from the environment. This PhD programme will focus on two of these organisms, *Cellvibrio. japonicus* and *Bacteroides thetaiotaomicron*.

1.5.1 *Cellvibrio japonicus*

C. japonicus (NCIMB 10462), an aerobic Gram negative bacterium first isolated from Japanese soil in the 1950s, is one of the best studied systems of plant cell wall hydrolysis (Gilbert et al. 1987, Gilbert et al. 1988, Hall et al. 1988).

Genomic sequence analysis showed that *C. japonicus* contains the complete repertoire of enzymes required to degrade plant cell wall and storage polysaccharides: approximately 130 predicted glycoside hydrolases, one-third of which were predicted to contain CBMs derived from 13 of the 71 known families (DeBoy et al. 2008). *C. japonicus* has also been shown, experimentally, to degrade most plant cell wall polysaccharides, including crystalline cellulose (DeBoy et al. 2008, Gardner et al. 2014), mannan (Hogg et al. 2003), and xylan (Andrews et al. 2004, Pell et al. 2004, Larsbrink et al. 2011) (reviewed by Gilbert, 2010). Unlike *Clostridium thermocellum*, an anaerobic plant cell wall degrading organism, *C. japonicus* enzymes that target polysaccharides integral to the plant cell wall, are secreted into the extracellular milieu, and do not assemble into large cellulosome-like complexes on the bacterial

surface (Millward-Sadler et al. 1995, Gilbert et al. 1987, Hall et al. 1988, Hall et al. 1989). It is also noteworthy that *C. japonicus* is predicted to encode a large number of GH43 enzymes, a family that targets arabinose decorations appended to "hairy" pectins and xylans, and thus these side chains may represent a major nutrient source (DeBoy et al. 2008).

1.5.2 *Bacteroides thetaiotaomicron*

B. thetaiotaomicron, a Gram-negative obligate anaerobe, is a prominent member of the normal human (and murine) gastrointestinal tract microbiota. The genome of *B. thetaiotaomicron* is predicted to encode 256 glycoside hydrolases, which are distributed into 43 families in the CAZy database, 16 polysaccharide lyases and 20 carbohydrate esterases. *B. thetaiotaomicron* is capable of hydrolyzing most glycosidic linkages identified in nature (including RGII). Comparing to the other prominent member of the gut bacterial community, *Firmicutes*, Bacteroides contain more than four times carbohydrate degradation genes per genome, (El Kaoutari et al. 2013). In contrast, there are only 98 known or putative glycoside hydrolases encoded by the human genome (Martens, Chiang and Gordon 2008, Koropatkin et al. 2008, Martens et al. 2011).

Analysis of the *B. thetaiotaomicron* genome showed that genes encoding polysaccharide degrading enzymes, and related binding proteins and outer membrane porins (SusC- and SusD-like), are usually clustered and co-regulated. (Martens et al. 2008, Martens et al. 2009b, Martens et al. 2011). Thus, these functionally related genes comprise "Polysaccharide Utilization Loci" (PULs) (**Figure 1.18**) that encode all of the requisite functions required degrade and sense specific glycans (Bjursell, Martens and Gordon 2006, Martens et al. 2008, Martens et al. 2009a, Martens et al. 2009b). The first PUL studied was the starch utilization system (Sus) of *B. thetaiotaomicron* as shown in **Figure 1.18**. A typical degradation pathway

is as follows: The complex glycan from the "environment" is bound by a SusD-like protein (and other binding proteins) attached to the outer membrane; glycoside hydrolases encoded by the PUL, located on the outer membrane, attack the attached glycan, hydrolyzing the polysaccharide into oligosaccharides which are then transported by the SusC-like protein into the periplasm. The imported oligosaccharides are then further degraded by periplasmic enzymes and the resultant monosaccharides are transported into the cytoplasm and metabolized. PULs also encode the regulatory proteins such as extra-cytoplasmic functioning (ECF) σ factors or hybrid two-component systems (HTCS), which activate expression of the genes that comprise the cognate locus. HTCS comprise single polypeptides that contain all of the domains of a classical two component system phosphorelay (Sonnenburg et al. 2006). These sensor/regulatory proteins (in the starch PUL the SusR regulator is an ECF-σ), straddle the inner member and are activated by oligosaccharides generated in the periplasm (**Figure 1.18**) (Martens et al. 2008, Martens et al. 2009a, Martens et al. 2009b).

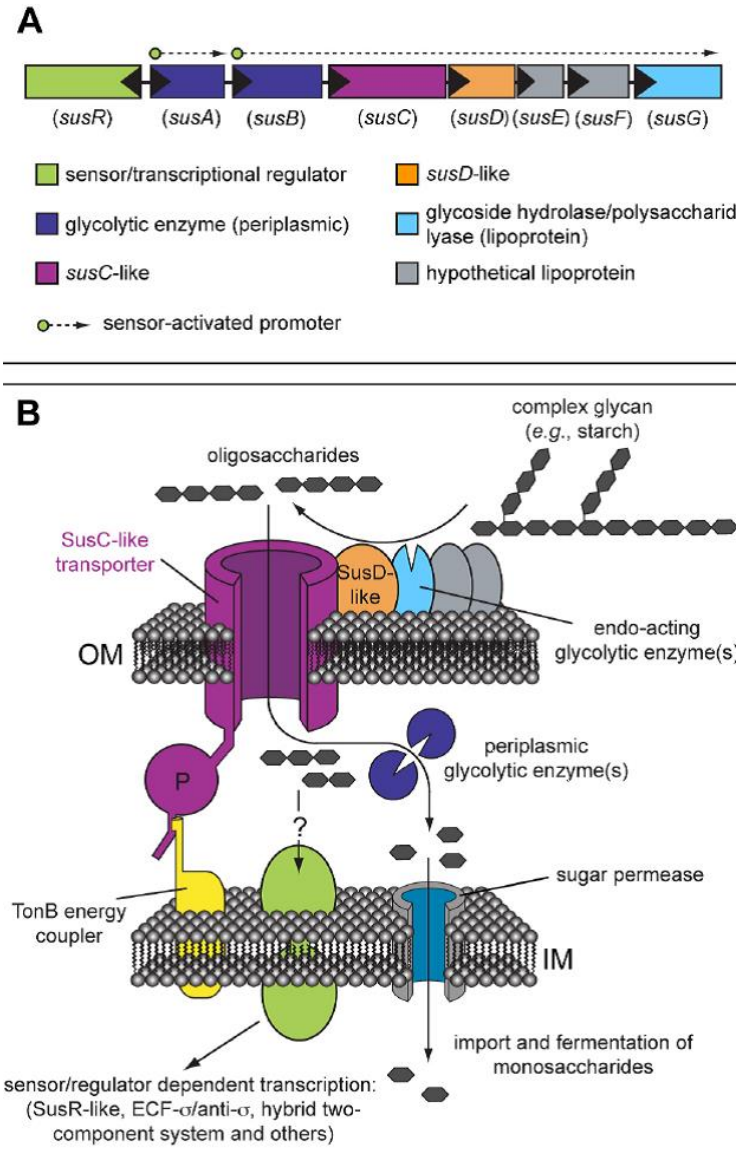


Figure 1.18 Examples of polysaccharide utilization locus (PUL) from *B. thetaiotaomicron*. This model is based on the prototypic starch utilization system (Sus). (A) The genomic architecture of the Sus system. (B) A functional model of how PUL gene products participate in glycan catabolism. Figure cited from **(Martens et al. 2008).**

While polysaccharides are known to activate PUL-associated HTCS, there is a paucity of information about the actual molecular cues recognized, and the mechanisms by which these inducers are protected by the degradative system to mediate their effect (Martens et al. 2011). In this study, we have not only characterized many enzymes related to RGI utilization, but also qualitatively identified the signaling molecule that binds to the sensor and up-regulates the RGI

PUL. A mechanism is also proposed by which the integrity of the inducing molecule is protected from being digested.

1.6. Objectives of the thesis.

Generally, enzyme activity is determined through *in vitro* assays using purified polysaccharides. This approach, however, does not consider the context in which these enzymes function *in vivo*, where the complex structure of the plant cell wall is likely to influence the function of these biocatalysts. In Chapter 3 of this thesis, the activities of GH5 and GH26 mannanases and family 2 carbohydrate esterases (CE2s) in vivo (against intact plant cell walls) were compared, and the capacity of CBMs, which display different specificities, to enhance the activity of these enzymes against intact plant cell walls was investigated. The data provide insight into whether the specificity of the CBMs or the topology of the catalytic modules are the primary drivers for the specificity of these enzymes *in vivo*.

Since the important role played by the community of the gut bacteria especially *B. thetaiotaomicron* (El Kaoutari et al. 2013) has been widely recognized, Chapter 4 and 5 studied the mechanism by which the pectin (especially RGI and RGII) is depolymerized and utilized by the predominant gut bacterium *B. thetaiotaomicron*. A number of pectic specific enzymes from this organism have been studied and characterized, which provide a first insight into how this predominant gut bacterium helps human to digest pectin.

Chapter 2. Materials and methods

2.1 Molecular Biology

2.1.1 Bacterial strains and plasmids

The *Escherichia coli* strains and plasmids used in this study are described in Table 2 - 1 and Table 2 - 2.

2.1.2 Media and growth conditions

E. coli cultures were grown at 37 ºC (unless indicated otherwise) with aeration in a rotary shaker (200 rpm), in sterile Luria-Bertani Broth (LB) (Sambrook, Fritsch and Maniatis 1989) or Selenomethionine (SeMet) medium (supplied by Molecular Dimensions Limited) with appropriate antibiotic, as described in Table 2 - 3 and Table 2 - 4.

To grow *E. coli* colonies, agar plates were made by adding 2 % (w/v) agar (Bacteriological) to LB, autoclaving and then allowing it to cool prior to the addition of selective antibiotic. The LB-agar was then poured into Petri dishes to solidify. Once solid the surface of the agar plate was dried by leaving the plates inverted at room temperature for ~24 h. The *E. coli* were spread across the surface of the plate with a glass spreader that had been sterilized by immersion in 100 % (v/v) ethanol, the excess of which was removed by passing through the hottest part (blue) of a Bunsen flame. Once the culture had dried the lid of the plate was replaced and the plate incubated, inverted, at 37 °C for 16 h in a stationary incubator (Laboratory Thermal Equipment Ltd). Longer incubation periods were avoided to prevent the formation of satellite colonies.

***E. coli* Strain**	**Description**	**Use example**	**Reference**
BL21 (DE3)	F' ompT, $hsdS_B(r_B\text{-}m_B)$, gal, dcm, (DE3)	Native protein expression (*Cj*Man5A and *Cj*Man26A)	(Studier and Moffatt 1986)
XL1-Blue	recA1, endA1, gyrA96, thi1, hsdR17, supE44, relA1, lac, [F'proAB$lacI^qZ\Delta$15, Tn10,(tet^r)]	DNA manipulation	Stratagene
Tuner (DE3)	F' ompT, $hsdS_B(r_B\text{-}m_B)$, gal, dcm, lacY1(DE3)	Native protein expression (*Cj*Man5A)	Novagen
B834	F^- ompT $hsdS_B$ (r_B^- m_B^-) gal dcm met (DE3)	Expression trial	Novagen
C41	Protease deficient strain of BL21(DE3) lysogen with a tighter control on protein expression than BL21 DE3; permits the expression of more toxic proteins under control of a T7 promoter with IPTG.	Protein trial expression and DNA manipulation	(Miroux and Walker 1996)
JM83 (DE3)	recA-, (DE3) ara, Δ(lac-proAB), rspL, ϕ80, lacZΔM15	DNA manipulation	(Norrander, Kempe and Messing 1983)
Origami™ pLysS	ara-leu7697, lacX74, phoAPvuII, phoR, araD139, ahpC, gal,E galK, rpsL, F´[lac+(lacIq)pro] gor522 ::Tn10 (Tc^r) trxB::kan (DE3) pLysS (Cm^r)	Expression trial	Novagen

Table 2 - 1 *E.coli* strains

*Cj*Man26A- *Cellvibrio japonicus endo* β-1,4-mannanase Man26A, *Cj*Man5A- *Cellvibrio japonicus endo*-β-1,4-mannanase Man5A. IPTG-Isopropyl-*β*-D-thiogalactopyranoside.

Plasmid	Size (kb)	Genotype	Use in this study	Reference
pET21a	5.44	amp^r, T7 lac $lacI^q$	*Cj*Man26A,	Novagen
pET22b	5.49	amp^r, T7 lac $lacI^q$	MGS,	Novagen
pET28a	5.37	kan^r, T7 lac $lacI^q$	RGI PUL enzyme,	Novagen

Table 2 - 2 Description of plasmids

Abbreviations; amp^r- Ampicillin resistance gene, kan^r-Kanamycin resistance gene, $lacI^q$- β-galactosidase repressor gene, Zn^r-Zeocin™ resistance, *ccd*B-lethal gene disrupted by an insert. For protein abbreviations see
Table 2 - 1 legend.

Medium	Composition for 1 litre	
Luria-Bertani Medium	10 g	Bacto®tryptone
	5 g	Bacto®yeast extract
	10 g	NaCl
	The pH was adjusted to 7.4 with NaOH and the final volume was made up to 1 litre before autoclaving	
M9 minimal Medium (supplied by Molecular Dimensions limited)	100 ml	10 x M9 salts (see below)
	20 ml	20 % (w/v) glucose
	2 ml	1 M $MgSO_4$
	1 ml	100 mM $CaCl_2$
10 x M9 Salts	128 g	$Na_2HPO_4.7H_20$
	30 g	KH_2PO_4
	5 g	NaCl
	10 g	NH_4Cl
	The pH was adjusted to 7.4 with NaOH and the final volume was made up to 1 litre before autoclaving	

Table 2 - 3 Growth media

2.1.3 Selective media

Antibiotics were used at 1000 fold dilution of stock solutions as given in Table 2 - 4; they were added to the sterilized growth media once it had cooled to ~50 ºC.

Antibiotic	Stock concentration	Storage details
Ampicillin (Amp)	50 mg ml^{-1} in water	+4 ºC for < 2 days
Kanamycin (Kan)	50 mg ml^{-1} in water	-20 ºC
Chloramphenicol (Cm)	34 mg ml^{-1} in 100 % (v/v) ethanol	-20 ºC

Table 2 - 4 Antibiotic concentrations, stock and final concentrations used in growth media

2.1.4 Storage of DNA and bacteria

Bacterial cultures (e.g. *Bacteroides thetaiotaomicron*) were stored at -80 ºC in 25 % (v/v) glycerol in cryo-vials. Bacterial colonies on agar plates were stored at 4 ºC for a maximum of four weeks. DNA was stored frozen at -20 ºC in Elution Buffer (EB, 10 mM Tris/HCl buffer, pH 8.5).

2.1.5 Sterilisation

Solutions, media and glassware were usually sterilised by autoclaving in either an Astell Hearson 2000 Series Autoclave or a Prestige® Medical Series 2100 Clinical Autoclave at 121 ºC, 15 lb in-2 (psi) for 20 min. When necessary, solutions were filter sterilised using a 0.45 µm pore Millipore filter disc (Supor® Acrodisc® 3.2, Gelman Sciences) and a suitable sterile syringe (Plastipak®, Becton Dickinson).

2.1.6 Chemicals, enzymes and media

Chemicals, enzymes and media that are not detailed in the text are listed in Appendix A. The water used to make all media and buffers was double distilled water purified to 18.2 Ω with a Millipore Milli-RO 10 Plus Water Purification System.

2.1.7 Centrifugation

Harvesting of bacterial cells from cultures of 50 –1000 ml was carried out by centrifugation at 5000 g for 10 min in 500 ml centrifuge tubes (Nalgene) using a Beckman J2-21 centrifuge with

a JA-10 rotor. Bacterial cells from culture volumes 5-10 ml were harvested in 30 ml Sterilin universal containers at 3000 g in a MSE Mistral 3000i bench centrifuge with a swing out rotor. Centrifugation of small volumes was carried out in 1.5 ml tubes (Treff) at 13000 g in a Haraeus Instruments Biofuge pico.

2.1.8 Chemically competent *E. coli*

E. coli strains were made competent for uptake of plasmid DNA using calcium chloride in a variation of the method described by Cohen (Cohen, Chang and Hsu 1972). An aliquot (1 ml) of a 10 ml culture of *E. coli* grown overnight in LB with appropriate antibiotic was used to inoculate 1000 ml of LB (with no antibiotic) in a 2 litre non-baffled flask. The culture was incubated at 37 ºC with aeration in a shaking incubator rotating at 180 rpm until log phase was reached (A_{600}=0.4). After 10 min on ice the cells were harvested by centrifugation at 5000 g at 4 ºC for 5 min. The supernatant was removed and the cells gently resuspended with a pipette in 8 ml of ice cold 100 mM $MgCl_2$. The cells were harvested again by centrifugation, the supernatant removed and the cell pellet resuspended in 4 ml of ice cold 100 mM $CaCl_2$. After incubation for 2 h on ice the cells were either used immediately or stored at -80 ºC in 200 µl aliquots with 15 % (v/v) glycerol in 1.5 ml cryovials (Eppendorf).

2.1.9 Transformation of competent *E. coli*

A 50-100 µl aliquot of competent *Escherichia coil* cells was gently thawed on ice for 5 min, and then plasmid DNA (1 µl containing 0.1-1.0 µg DNA) was added. The DNA and cells were gently mixed with a pipette tip and incubated on ice for 30 min. The cells were then heat-shocked at 42 °C for 1.5 min and immediately kept on ice for a further 2 min. Next, 200 µl of Luria broth (LB) was added, and then the Eppendorf tubes were secured to the shaker platform of a 37 °C orbital shaker and incubated for 1 h. Finally, 200 µl transformed cells were plated

onto selective LB agar media and incubated at 37 °C overnight. Controls were carried out by transforming the *E. coli* with standard plasmids or with buffer containing no plasmid.

2.1.10 Replication of DNA

When necessary, more DNA was produced by transformation of the DNA into Top10 chemically competent cells and growth of colonies on selective LB agar plates. The DNA was prepared on a small or medium scale using Qiagen® kits essentially according to the manufacturer's instructions. The buffer compositions, when publicly available, are shown in Appendix A.

2.1.11 Small scale (20 ng) preparation of plasmid DNA from *E. coli* (Qiagen miniprep)

Single colonies were used to inoculate a universal containing 5 ml of LB with appropriate antibiotic. The culture was incubated at 37 ºC with aeration overnight, as described before. These cells were then harvested by centrifugation for 10 min and the supernatant carefully removed. The cell pellet was re-suspended in 250 µl Buffer P1 by pipetting and transferred to a 1.5 ml tube. To this suspension 250 µl of Buffer P2 was added and mixed by inverting 5 times. This was incubated at room temperature for 3-5 min to allow cell lysis to proceed. The cell lysis reaction is halted by the addition of 350 µl of Buffer N3 and the suspension mixed immediately by gentle inversion. The cloudy white precipitate which forms was removed by centrifugation for 10 min. The supernatant was applied to a QIAspinTM prep column. This was centrifuged for 1 min and the eluant discarded. The column was washed with 500 µl Buffer PB, followed by two washes with 750 µl Buffer PE with 1 min centrifugations between each step (discarding the eluant) and after the final wash to remove residual Buffer PE. DNA was eluted into a fresh 1.5 ml tube by the addition of 50 µl EB to the column bed, incubated at room temperature for 1 min followed by a 1 min centrifugation.

2.1.12 Restriction digests of DNA

Digestion of double stranded DNA with restriction endonucleases was carried out as directed by the manufacturers (MBI Fermentas). Endonuclease restriction sites were identified using the tool on-line at http://rna.lundberg.gu.se/cgi-bin/cutter2/cutter.

The required amount of DNA dissolved in EB was mixed with the appropriate volume of 10 times concentrated reaction buffer. Endonuclease was then added at 5-10 units μg^{-1} DNA and the digest incubated at 37 ºC for 2 h (one unit is the amount of enzyme required to digest 1 µg of DNA in 1 h). For double digests where both enzymes retained activity and specificity in the same reaction buffer the reactions were carried out simultaneously. When the reaction conditions for two restriction enzymes were incompatible the DNA was digested with one endonuclease, the buffer removed with the Qiagen PCR-clean-up kit then digested with the other endonuclease. Controls of undigested and single endonuclease-digested DNA were also carried out.

2.1.13 Ligation of insert DNA and vector DNA

Appropriately digested DNA extracted from an agarose gel (using standard QIAquick Gel Extraction Kit) or cleaned using the Qiaquick PCR purification kit was used for ligations. DNA molecules with cohesive ends were ligated with a final molar ratio of insert:vector of 3:1 whereas for DNA molecules with blunt-ended termini the ratio was at least 10:1. The reaction mixture included 4 µl 5x ligation buffer (250 mM Tris/HCl, pH 7.6, 50 mM $MgCl_2$, 25 µM ATP, 25 mM dithiothreitol, 25 % (w/v) polyethylene glycol 8000) and 1 µl (5U) T4 DNA ligase (Invitrogen) made up to a final reaction volume of 20 µl with H2O. Ligations were performed at room temperature for 2 h using T4 DNA ligase (Invitrogen) as per the manufacturer's instructions, then 2 µl of ligation product was transformed into TOP10 chemically competent cells (Invitrogen).

2.1.14 Determination of DNA concentration

DNA concentration was determined either by estimation from comparison of band size/intensity compared to quantified DNA ladders when electrophoresed on agarose gels or by scanning the absorbance of the appropriately diluted DNA (in MQ H_2O) between 230 nm to 340 nm using a NanoDrop™.

2.1.15 Agarose gel electrophoresis of DNA

The separation and determination of the sizes of linear DNA molecules were carried out by electrophoresis through submerged horizontal gels. The compositions of all buffers and reagents are given in Table 2 - 5.

The gels consist of agarose (0.4 g) in 1 x Tris/Borate/EDTA (TBE) buffer (50 ml, see below) and were cast in mini-gel trays (Applied Biosystems). The gels were submerged in 1 x TBE buffer containing ethidium bromide (0.5 μg ml^{-1}) so that the DNA could be visualised. The samples were prepared by the addition of loading buffer and pipetted into the wells of the gel. A lane of markers, such as HyperLadder III (Bioline) was also loaded. The DNA was electrophoresed at 60 mA for ~1 h.

Buffer	**Ingredient**	**Amount**
Electrophoresis Buffer-TBE	Tris base	108 g l^{-1}
(pH 8.3 checked)	Boric acid	55 g l^{-1}
	0.5 M EDTA pH 8.0	40 ml l^{-1}
DNA Loading Buffers		
DNA Sample Buffer	Bromophenol Blue	0.25 % (w/v)
	Glycerol	50 % (v/v)
	TBE	10 x
PCR DNA Sample Buffer	Xylene Cyanol FF	0.25 % (w/v)
	Glycerol	50 % (v/v)
	TBE	10 x

Table 2 - 5 Buffers for agarose gel electrophoresis (10 x stocks)

2.1.16 Visualisation of DNA and photography of agarose gels

An image of the gel was obtained using a gel documentation system (Bio-Rad Gel Doc EZ, and photographs were either produced by Mitsubishi Video Copy Processor (Model P68B) with Mitsubishi thermal paper, or electronically saved.

2.1.17 Recovery of DNA from agarose gels

Agarose gels for gel extraction were made using high purity Seachem Gold™ agarose and TBE buffer. DNA was isolated from agarose gels by excision of the required band from the gel with a clean scalpel blade. The DNA was purified using a QIAquick Gel Extraction Kit (Qiagen) essentially following the manufacturers' protocol.

The gel slice was weighed and added to 3 volumes (100 mg gel ~100 μl) of Buffer QG (ingredients not stated by manufacturer) before incubating at 50 ^{o}C until the gel had completely dissolved (~10 min). The melted agarose solution was then applied to a QIAquick column and centrifuged at 13000 *g* for 1 min. The flow-through was discarded, and this centrifugation step repeated with 0.5 ml Buffer QG followed by 0.75 ml Buffer PE (ingredients not stated by manufacturer) before spinning again to remove any residual Buffer PE. The DNA was then eluted from the column with 30-50 μl of water or Elution buffer.

2.1.18 Automated DNA sequencing

DNA sequencing was conducted using the Value Read service from MWG Biotech AG, Ebersberg, Munich, Germany using ABI 3700 sequencers and Big dye technology (Applied Biosystems), each clone was sequenced in both the forward and reverse direction. Plasmid DNA (1.5 μg) was dried by vacuum lyophilization at room temperature in 1.5 ml tubes. Plasmids were then sequenced with appropriate primers listed in Table 2 - 6.

Plasmid type	Forward primer	Reverse primer
pET plasmids	T7 TAATACGACTCACTATAGGG	T7 term CTAGTTATTGCTCAGCGGT
pCR® Blunt	M13 uni (-21) TGTAAAACGACGGCCAGT	M13 reverse (-29) CAGGAAACAGCTATGACC

Table 2 - 6 Plasmid sequencing primers

2.1.19 Polymerase Chain Reaction (PCR)

The polymerase chain reaction (PCR) developed by Mullis and was used in this study to amplify DNA fragments and, in a modified form, to introduce amino-acid mutations .

Two oligonucleotide primers are required for PCR, one complementary to each strand of the DNA molecule, at sites that flank the region of DNA to be amplified. A thermostable DNA polymerase catalyses the synthesis of the complementary DNA strand in the presence of dNTPs. Oligonucleotide primers were designed such that complementary sequences were normally 20 bases in length with a G/C content of approximately 40 % and a melting temperature (Tm) >45 °C and within 5 °C difference between the two primer. Tms were calculated using the online tool "Oligonucleotide Properties Calculator" at http://www.basic.nwu.edu/biotools/oligocalc.html

If possible, primers were also designed such that they contain two or three G or C bases at both ends of the primer, which would make the primer termini anneal well to the template strand and increase the amplification efficiency. When required, restriction site sequences were added to 5´-ends of primers with the addition of the sequence CTCCAG at the extreme 5´-end to facilitate the restriction cutting of the PCR product for ligation into appropriate vectors.

PCRs were made up in sterile 0.2 ml or 0.5 ml Eppendorf tubes as follows:

1 µl	dNTP mix (100 mM)
1 µl	50 µM Forward oligonucleotide primer
1 µl	50 µM Reverse oligonucleotide primer
~ 60 ng	DNA template
1 µl	DNA polymerase (2.5U μl^{-1})
5 µl	10 x reaction Buffer (200 mM Tris/HCl, pH 8.8, 20 mM MgSO4, 100 mM KCl, 100 mM $(NH_4)2SO_4$, 1 % Triton® X-100, 1 mg ml^{-1} BSA)
	sterilised ddH_2O to 50µl

Control reactions lacking target DNA were always carried out.

The standard thermocycle was as follows:

1 cycle	95°C for 3 min
25 cycles	95°C for 30 sec, (Lower T_m of primer pair – 5) °C for 1 min 72°C for 2 min kb^{-1} fragment size
1 cycle	72°C for 10 min and 5°C for 24 h

The period of the extension reaction at 72 °C was generally 2 min for every kb of DNA to be amplified for the *Pfu*Turbo™ and KOD™ polymerases. After thermocycling was complete, a 10 µl aliquot of each PCR was analysed by electrophoresis on a mini-gel.

2.1.20 Purification of PCR products

DNA was purified using QIAquick PCR Purification Kits (Qiagen) following the manufacturer's protocol.

To 1 volume (30 µl) of the PCR reaction, 5 volumes (150 µl) of Buffer PB (ingredients not stated by manufacturer) were added. The mix was applied to a QIAquick column and centrifuged at 13000 g for 1 min. After discarding the flow through, 0.75 ml of Buffer PE (ingredients not stated by manufacturer) was added and centrifugation repeated in order to wash the column before spinning again to remove any residual Buffer PE. The DNA was then eluted from the column with 30-50 µl of EB.

2.2 Protein purification

2.2.1 Over expression of protein in *E. coli*

Colonies from plates containing freshly transformed (maximum 2 days) cells were used to inoculate LB (10 ml) containing the appropriate antibiotic. These were incubated at 37 °C with aeration (orbital incubator shaking at ~150 rpm) overnight, and the resultant cultures were used to inoculate 1 litre of LB supplemented with the appropriate antibiotic. After incubation at 37 °C with aeration to mid-growth phase (absorption observed at 600 nm, A_{600}=0.6), the cultures were transferred to an incubator at the appropriate temperature (with aeration) and recombinant protein expression induced with the required amount of the lactose analogue IPTG. The flasks were incubated at this temperature for the necessary length of time and then the cells harvested by centrifugation.

The protein was extracted from the cell pellet and purified either by making periplasms or cell free extracts depending on the location of the protein. The mannanase from *Cellvibrio japonicus*, *Cj*Man26A, was the only protein found in the periplam while the other proteins used in this study were cytoplasmic.

2.2.2 Preparation of cell-free extracts (CFEs)

Cell free extracts (CFEs) were usually prepared from 1-4 litres of *E. coli* that were harvested by centrifugation at 6000 *g* for 10 min. The cell pellet was resuspended in 10 ml litre^{-1} of 10 mM Tris/HCl pH 8.0 containing 300 mM NaCl (TALON^{TM} Buffer) and stored in 10 ml aliquots in Sterilin tubes for 16 h at -20 ^{o}C. After thawing, the cell suspension was sonicated for 2 min using a B. Braun Labsonic U sonicator set at low intensity ~42 watts and 0.5 second cycling, then transferred to a 50 ml centrifuge tube (Nalgene) and cell debris pelleted at 30000 *g* for 30 min at 4 ^{o}C, and the supernatant comprising the CFE, was retained for further use. The cell

debris-containing pellet (insoluble fraction) was resuspended in 10ml litre^{-1} ml of TALON Buffer.

2.2.3 Immobilised Metal Affinity Chromatography (IMAC)

His-tagged proteins can be purified by IMAC because the imidazole side-chain interacts with electropositive transition metal immobilised in a column. TALON™ (Clontech Laboratories Inc.) columns with 2 ml TALON™ resin containing cobalt ions were prepared by washing in 10 volumes water and then TALON buffer. Cell free extracts (CFEs) were filtered (0.45µm) and poured through the column. The column was washed with 4 x 5 ml of TALON buffer followed by 2 x 5 ml of TALON buffer containing 5 mM imidizole. The protein was eluted with 4 x 5 ml of TALON buffer containing 100 mM imidazole. Analysis by SDS-PAGE of each stage of the purification indicated which fractions contained the required purified protein.

2.2.4 Determination of protein concentration

Pure protein concentration was determined using published methods--- spectrophotometric assay (Gill and von Hippel 1989, Pace et al. 1995). The absorbance between 280nm and 340nm of the appropriately diluted pure proteins was measured with a spectrophotometer and the data were used to calculate protein concentrations, using the formula:

$$A = \varepsilon ClD$$

Where A = absorbance 280 nm-absorbance 320 nm, ε = molar extinction coefficient, l = length of light path (cm), D = dilution factor and C = molar concentration of sample.

Calculation of the molar extinction coefficient was done by use of the ProtParam program hosted at www.expasy.com. The length of the light path was always 1 cm.

2.2.5 SDS-PAGE

SDS-PAGE was used to analyse the size, relative purity and quantity of protein. Gels consisting of 10 – 15 % polyacrylamide were typically used dependent on size of protein of interest. The gels are composed of two parts, the resolving gel and the stacking gel (Table 2 - 7). The system used was the AE-6450 apparatus from ATTO Corporation (Genetic Research Instruments) which employs 12 cm x 10 cm glass plates. The resolving gel was poured into the plates layered with water and allowed to polymerize. The water is then removed and the stacking gel poured on top of the resolving gel. This is allowed to polymerize with a comb in place. The comb and the rubber seal around the glass plates were removed when the gel was set. Protein samples were prepared by mixing 15 µl of appropriately diluted protein solution, 7 µl of loading buffer (Table 2 - 7) and boiling for three min. The plates were placed in the gel tank that was filled with running buffer. The samples were loaded into the well of the gel, which was then electrophoresed at a current of 30 A per gel.

After electrophoresis, the gel was soaked in Coomassie Blue stain (0.4 % Coomassie Brilliant Blue R, 10 % (v/v) glacial acetic acid, 40 % (v/v) methanol) for 60 min with gentle shaking. The gel was then destained in a solution consisting of 40 % (v/v) methanol and 10 % (v/v) glacial acetic acid until all of the protein bands were visible and recorded using a digital camera (Canon PowerShoot A75). The Mr of proteins separated by SDS-PAGE was estimated by comparing their electrophoretic mobility with protein standards of known Mr.

Component	Volume/Amount
Resolving gel	
0.75 M Tris/HCl buffer, pH 8.8 with 0.2 % SDS	9.4 ml
40 % (w/v) Acrylogel (BDH Electran; acrylamide, 3 % (w/v) bisacrylamide)	5.8 ml
d.d. H_2O	3.5 ml
10 % (w/v) Ammonium persulphate	90 μl
TEMED	30 μl
Stacking gel	
0.25 M Tris/HCl buffer, pH 8.8 with 0.2 % SDS	3.75 ml
40 % (w/v) Acrylogel, 3 % (w/v) acrylamide, bisacrylamide	0.75 ml
d.d. H_2O	3.0 ml
10 % (w/v) Ammonium persulphate	60 μl
TEMED	20 μl
Loading buffer	
SDS	10 % (w/v)
0.25 M Tris/HCl buffer, pH 8.8) with 0.2 % SDS	5 ml
Glycerol	25 % (w/v)
β-mercaptoethanol	2.5 ml
Bromophenol blue dye	0.1 %
Running buffer	
Tris/190 mM glycine, pH 8.3	32 mM
SDS	0.1 %

Table 2 - 7 Composition of resolving gel, stacking gel and loading buffer for SDS-PAGE

2.2.6 Ion exchange chromatography using a Fast Performance Liquid Chromatography (FPLC) system

The pooled fractions of protein purified by IMAC were dialysed against 2 x 4 litres of 10 mM Tris/HCl buffer, pH 8.0 (Buffer A) at 4 ºC in dialysis visking tube with a 10 KDa cut off to remove the salt. The column (Biorad UNO^{TM} Q12) was washed with 6 M guanidium/HCl then equilibrated with 400 ml of Buffer A. The buffers used for this column were filtered (0.2 μm) and degassed. The dialysed protein was filtered (0.2 μm), loaded onto the column and eluted in an automatic programme (Biorad) in which 20 ml of Buffer A is passed through the column, followed by 60 ml of a linear gradient from 0 to 250 mM NaCl, in Buffer A. This is achieved by increasing the percentage of Buffer B, which is Buffer A containing 250 mM NaCl. The pump runs at 1 ml min^{-1}. The fractions were only collected when the protein concentration was above a threshold value of 0.05 Absorbance Units at a wavelength of 280 nm. These

fractions were analysed by SDS-PAGE and those containing the purest protein were pooled and subjected to the final stage of purification.

2.2.7 Gel filtration

The column (HiLoa^{TM} 26/60 Superde^{TM} 75 Prep grade, Amersham Pharmacia Biotech) was washed with 400 ml of Buffer A containing 150 mM NaCl at 2 ml min^{-1}. The protein was concentrated to a volume of 5.5 ml and filtered (0.2 μM) to load onto the column. The automatic system (Akta *prime*, Amersham Pharmacia Biotech) eluted the proteins in 400 ml of Buffer A containing 150 mM NaCl at 2 ml min^{-1}. The fractions were collected and analysed by SDS-PAGE.

2.2.8 Concentrating protein

Protein solutions were concentrated when necessary during the purification process, and prior to use for enzyme assays or crystallization. Filtered (0.2 μm) protein solutions were concentrated in centrifugal concentrators with a suitable size (usually 10 kDa) molecular weight cut-off (Filtron MacroSep^{TM}) in a MSE Mistral 3000i bench centrifuge with a swing out rotor at a force of 3000 *g*, this concentrates approximately 15 ml to 1.5 ml in 2 h.

2.3 Bioinformatics

2.3.1 Alignments

Amino acid sequence searches were carried out using the Basic Local Alignment Search Tool (BLAST (Altschul et al. 1997), using the NCBI (National Center for Biotechnology Information) version hosted at the European Bioinformatics Institute (EBI) website (http://www.ebi.ac.uk). Amino acid sequences were aligned using either ClustalW or T-Coffee, at the EBI website.

2.3.2 Prediction of prokaryotic signal peptides

Signal peptides target proteins out of the cytoplasm to the cell membrane or for secretion. The presence or absence of signal peptides, and the site of signal peptide cleavage, was assessed using the SignalP software hosted at www.expasy.com (Nielsen and Krogh 1998, Nielsen et al. 1997). The sequence of the 50 N-terminal amino acids of the enzymes under investigation were submitted to the web-server for analysis using neural networks and Hidden Markov Models (HMM) trained on Gram-negative or Gram-positive bacteria as appropriate. The results display the probability (between 0 and +1) that the enzyme contains a signal peptide. The results obtained are explained as follows (summarised from www. expasy.com).

C-score (raw cleavage site score). This output score from neural networks is trained to recognise cleavage sites versus other sequence positions, which is trained to be high at the residue immediately after the cleavage site, low at all other positions.

S-score (signal peptide score). This output score from networks is trained to recognise signal peptide versus non-signal-peptide positions, which is trained to be high at all residues before the cleavage site and low for 30 positions after the cleavage site, and in the N-termini of proteins without a signal peptide.

Y-score (combined cleavage site score). The prediction of cleavage site location is optimised by observing where the C-score is high and the S-score changes from a high to a low value. The Y-score formalises this by combining the height of the C-score with the slope of the S-score.

The results of the HMM provide a probability of the presence of a signal peptide and a prediction of the three regions of the signal peptide; the basic N-terminus (n-), the α-helix (h-) and the polar cleavage site (c-).

Cleavage sites resulting in a cysteine residue at the N-terminus of the mature enzyme are Type-II signal peptidase sites (Inouye, Duffaud and Inouye 1986), and most likely result in lipid-linked, and thus membrane-associated, protein.

2.4 Biochemistry

Unless otherwise stated, all enzyme assays were performed at, and all reagents and cuvettes pre-warmed to, 37 ºC. All assays were repeated at least twice. The same set of Gilson pipettes was used throughout each assay, and these were calibrated periodically.

2.4.1 3, 5-Dinitrosalicyclic acid assay (DNSA)

The rate of hydrolysis of polysaccharides was monitored by the increase in reducing sugar over time. The free anomeric carbon at the end of polysaccharide chains can open from its more common cyclic sugar conformation and act as a weak reducing agent. When a glycosidic bond is cleaved by enzymatic hydrolysis, a new reducing end is formed, the concentration of which can be determined by the DNSA assay, using the method of Miller (Miller 1959). A 500 µl aliquot of an enzyme reaction was added to 500 µl DNSA reagent (1 % (w/v) DNSA, 0.2 % (v/v) phenol, 1 % (w/v) NaOH, 0.002 % glucose, 0.05 % (w/v) NaSO3) to terminate the reaction. The tube was then boiled for 20 min, placed on ice for 10 min, equilibrated to room temperature and the absorbance read at 575 nm. A standard curve of 0-1000 µg/ml monosaccharide (plus polysaccharide substrate) was used to quantify the released reducing sugar.

2.4.2 Thin layer chromatography (TLC)

A TLC plate (Silicagel 60, 20 × 20, Merck) was cut to the desired size (the minimum height should be 10 cm). Between 5 and 10 µl of samples were spotted on the plate, separated by 7 mm. Solvent (50 ml) comprising freshly made 1-butanol/acetic acid/water (2:1:1, v/v) was

poured into a glass chromatography tank (23 × 23 × 7.5) and covered tightly. Vapours were allowed to equilibrate for at least 2 h before use. The TLC plate was placed into the tank and samples allowed to migrate until the running buffer reached ≈ 1 cm to the top of the plate. The plate was dried carefully using a hairdryer and put back in the tank for another run. The plate was dried again and immersed for a few seconds in an orcinol sulphuric acid reagent (sulphuric acid/ethanol/water 3:70:20 v/v, orcinol 1 ‰), dried carefully and heated until sugars were revealed, at 120 °C (5 -10 min). Plates were recorded using a digital camera (Canon PowerShoot A75). To estimate the size of the sugars, different size marker standards were run in parallel, depending on the nature of the samples. Standards consisting of known monosaccharides and oligosaccharides were spotted on the TLC plate. In order to detect the presence of sugar in a sample quickly, 5 µl were spotted on a section of TLC plate and immersed for a few second in orcinol sulphuric acid reagent, dried carefully and heated until sugars were revealed, as previously described.

2.4.3 High performance liquid chromatography (HPLC)

The HPLC machine (Dionex ISC3000) is fully automated and has a loop size of 200 µl, and runs at a flow rate of 1 (or 0.5) ml min^{-1} and a pressure of ~2300 (or ~2800) psi (lb in2). The sugars were detected by pulsed amperometric detection (PAD) with settings E1= +0.05, E2= +0.6, E3= -0.6. The column is a CARBOPAC™ PA-100 (or PA-200 and PA-1) anion exchange column equipped with a CARBOPAC™ PA 100 (or PA-200) guard column (Dionex). At least two separate standards (e.g. D-mannose and mannobiose, 10 µM each) were analysed in each run, one near the beginning and the other at the end to control for the variation in the elution times of the same sugar over the run. In a standard run the column was equilibrated for 5 min with 66 mM sodium hydroxide, and then the sugars were eluted in a linear gradient from 0-15 % of 500 mM sodium acetate, in 66 mM sodium hydroxide, for 30 min. The column was

then washed with 500 mM sodium acetate in 66 mM sodium hydroxide for 5 min, and with sodium hydroxide 500 mM for 10 min then re-equilibrated with 66 mM sodium hydroxide for 15 min. The total run lasts 65 min for each sample.

2.4.4 Isothermal titration calorimetry (ITC)

The thermodynamics and binding affinities of carbohydrate protein interaction were investigated by ITC using an iTC200 ITC machine (MicroCal). ITC measurements were made at 25 °C following standard procedures. The proteins were dialyzed extensively against buffer and the ligands were dissolved in the dialysis buffer to minimize heats of dilution. Degassed protein solution (250 µl) at high concentration (80 - 145 µM) was equilibrated in a reaction cell maintained at 25 °C. To this solution, 20 aliquots (2 µl) of ligands at an appropriate concentration were automatically injected, with rapid stirring (1000 rpm), at 300 s intervals. Typically, oligosaccharides ligands were titrated at about 10 mM and soluble polysaccharide ligands at 10 mg/ml. Binding was monitored by measuring heat released (exothermic binding) or absorbed (endothermic binding) for each aliquot of ligand. During the titration, the difference in electrical power required to maintain the temperature of the reaction cell versus the temperature of the reference cell was recorded, and from these differences the heat change on binding calculated.

The molar concentration of binding sites present in polysaccharide ligands was determined following the method. After determining that the test protein contained a single binding site, by titrations with oligosaccharides of known molarity or through structural biology, the molar concentration of polysaccharide ligands was fitted in an iterative fashion until the n-value was as close as possible to one. Integrated heat effects, after correction for heats of dilution, were analyzed by nonlinear regression using a single site-binding model (Microcal Origin, version

7.0). The fitted data yield the association constant (KA) and the enthalpy of binding (ΔH). Other thermodynamic parameters were calculated using the standard thermodynamic equation: -RTlnKA = ΔG = ΔH-TΔS where R is the gas constant (1.99 cal.K-1.mol-1), T is the temperature in degree absolute (298.15 K), ΔG is the change in free energy and ΔS is the change in entropy upon binding. When required, to remove any trace of calcium, purified protein at a concentration of 300 μM was passed through a 20 × 1.5 cm column containing a 10 ml bed volume of Chelex-100 (Sigma) under the flow of gravity. The buffer used for ITC was treated in the same way to remove any trace of calcium.

2.4.5 Affinity gel electrophoresis (AGE)

An affinity electrophoresis method uses native polyacrylamide gels containing soluble ligand sugars. A continuous gel system was applied with the same gel apparatus (ATTO Corporation) as used for SDS-PAGE. Gels contained 7.5 % (w/v) acrylamide (Acrylogel 3; BDH Electran™) 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, containing 5 % (v/v) glycerol and 0.0025 % Bromophenol, Blue final concentrations, were electrophoresed at 10 mA/gel at room temperature for 2.5 h. Proteins were then stained in 0.4 % (w/v) Coomassie Blue, 40 % (v/v) methanol, 10 % (v/v) glacial acetic acid for ≈ 1 h and then destained in 40 % (v/v) methanol, 10 % glacial acetic acid. 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.

2.4.6 Production and purification of RGI backbone (mucilage)

Mucilage was extracted in this project from *Arabidopsis thaliana* seeds as a source of undecorated RGI backbone. The method used was as follows: 80 g of *A. thaliana* were

resuspended in 80 ml of distilled water and stirred at ~15 rpm overnight at 4 ^{o}C. The suspension was then centrifuged at 15000 rpm for 5 min and the supernatant was filtered through a G1 glass filter (Pore size: 15–40 micron), Whatman filter paper (Catalog. 3030917), or a BDH porosity 3 filter funnel (Pore size: 15–40 micron), and then dialyzed at least twice against 40 volumes of water before freeze drying the glycan. From 80 g of seeds the yield of RGI was typically 1 g.

The RGI generated from *A. thaliana* mucilage contained no side chains comprising only of the Rha-GalA backbone. Mucilage extracted from flax seeds contained RGI decorated with L-Gal and L-Fuc (Naran, Chen and Carpita 2008). The method was identical to that described for *A. thaliana* with a yield of 1.5g of RGI from 80 g of seeds.

2.4.7 Purification of oligosaccharides by size exclusion chromatography

RGI backbone (*A. thaliana* mucilage) was digested by either BT4170 (PL9) or a GH28 endo-galacturonase (U29SH8) from Novozyme to generate oligosaccharides comprising 4,5ΔGalA-$[Rha\text{-}GalA]_n$ or $[Rha\text{-}GalA]_n$, respectively. BT4170 was used at 600nM against substrate at 20mg/ml in 20mM Tris-HCl buffer pH8.5 with 2mM $CaCl_2$ at 37 ^{o} C. Samples were taken at various time points and analyzed to make sure the production of a wide range of oligosaccharides. U29SH8 was applied at a concentration of 0.2μM against substrate of 10mg/mL in 50mM sodium acetate buffer pH5 at 37 ^{o} C and the digestion was terminated after 30min by boiling for 5min. The digests derived from the two enzymes were freeze dried or concentrated to <5 ml dried, and then separated by size exclusion chromatography using P2 or P4 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 in degassed MQ H_2O. Fully digested and freeze dried samples was diluted and loaded directly onto the columns, which was then run with MQ H_2O 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. After analysis by TLC and HPLC, identified fractions of interest were pooled and freeze dried for further analysis. The size of the oligosaccharides was determined by MALDI-TOF Mass spectrometry (see Chapter 4.2.3).

From 0.5 g of RGI, the yield of ΔRha-GalA-Rha-GalA and ΔRha-GalA are 80mg and 20mg, respectively and the yield of Rha-GalA-Rha-GalA and Rha-GalA-Rha-GalA-Rha-GalA 64 mg and 54 mg, respectively.

2.4.8 Extraction of RGII from red wine in large scale.

Pure RGII is required as the initial substrate to analyze the enzymatic digestion of RGII. The polysaccharide was extracted from red wine (Merlot) from Nantes, France during my secondment/relocation. Wine and primary cell wall RGII have identical structures.(Doco and Brillouet 1993) RG II is present in wine as a monomer and a dimer that is cross-linked by a 1:2 borate-diol ester.(Kobayashi, Matoh and Azuma 1996) Extraction was performed during my secondment in INRA (French National Institute for Agricultural Research) following the protocol below: 21 liters of red wine were condensed into 5 litres by ultrafiltration and rotary evaporation. RGII was then precipitated by the addition of five volumes of ethanol. After dissolving the precipitate in water, three resins were used to remove pigment from the red wine: two strong cation exchangers and one anion exchanger. Both bound and unbound fractions contain RGII, which may be caused by minor structural difference of RGIIs and slight binding ability of RGII to the anion exchanger. Both fractions were subjected to DEAE anion exchange column chromatography and eluted with acetate buffer. Finally the RGII was subjected to S-400 size exclusion chromatograph to remove neutral sugar impurities. (**Figure 2.1**).

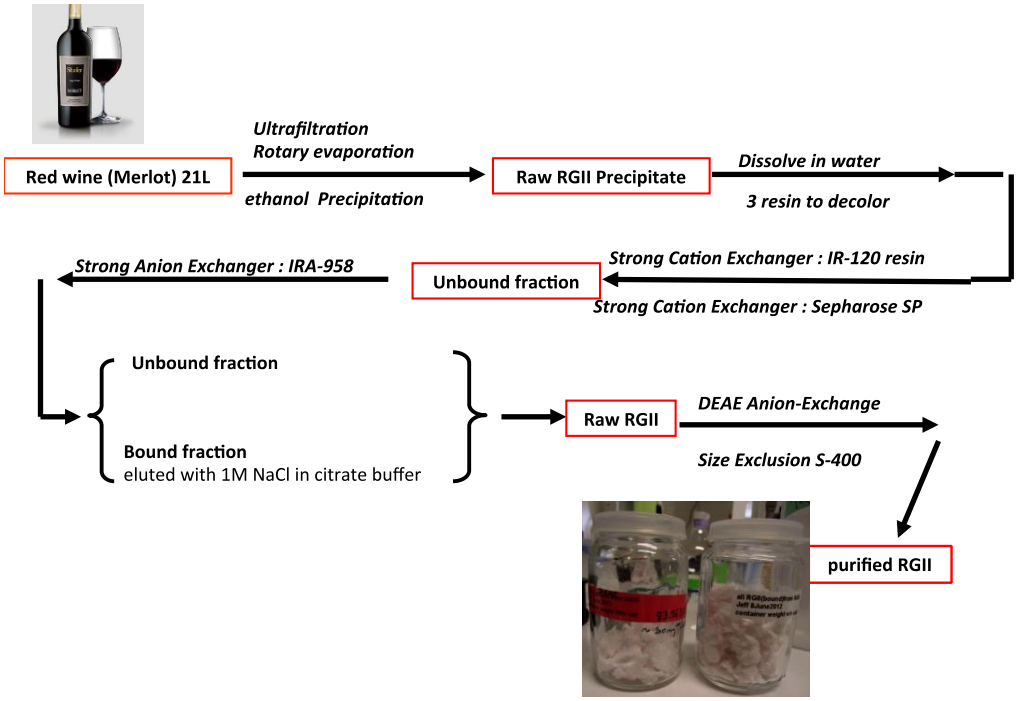


Figure 2.1 A flow chart of the extraction of 1.5 gram RGII from 21 litres of red wine.

2.5 Immunohistochemical labeling

Molecular probes such as monoclonal antibodies (MABs) and carbohydrate-binding modules (CBMs) are important tools to detect and dissect cell wall structures in plant materials and intensively used in this study.

2.5.1 Polysaccharide specific antibodies (MABs)

A large range of rodent MABs to plant cell wall polysaccharides are now available as detailed at the online sites of Biosupplies (www.biosupplies.com.au), Carbosource Services (www.ccrc.uga.edu/~carbosource/CSS_home.html) and PlantProbes (www.plantprobes.net). Biosupplies and Carbosource MABs are derived using mouse hybridoma systems and thus the probes require anti-mouse secondary reagents whereas those at PlantProbes are mostly rat and require anti-rat secondary reagents. In this study, antibodies from both CCRC (Hahn's lab) and Leeds (Knox's lab) were tested and used.

2.5.2 Preparation and fixation of plant material

4% solution of formaldehyde in PEM buffer (50 mM Pipes, 5 mM EGTA, 5 mM MgSO4; pH adjusted to 7.0 with KOH) was prepared, followed by a 12% or 16% (w/v) stock solution of paraformaldehyde in water by heating up to 70°C and adding 1 M NaOH dropwise until the cloudy solution turns clear. Ethanol (30% to 100%) was used for the dehydration procedures. Steedman's wax, an ethanol-soluble low melting point polyester wax (35-37°C), is prepared for embedding (Sigma-Aldrich, Gillingham, UK). For LR White resin embedding (hard grade, containing 0.5% of the catalyst benzoin methyl ether, Agar Scientific), disposable base moulds (15 x 15 x 5 mm, Electron Microscopy Sciences, Hatfield, USA), Gelatine capsules (Agar Scientific) and embedding cassettes (Simport, Beloeil, Canada) were used for embedding.

2.5.3 Wax-embedding protocol

The Steedman's wax was used due to its low melting point and good sectioning properties. Washing of the fixed material took place in in PEM buffer or PBS buffer twice for 10 min, followed by dehydrating by incubation in an ascending ethanol series (30, 50, 70, 90, 100%) for 30 min for each change at 4°C. Samples were then moved to 37°C, incubated in molten wax and ethanol (1:1, O/N) and then 100% wax (2 x 1 h), and placed in the mould filled with molten wax. Leave at room temperature overnight to solidify. Samples were ready to be used 12 –24 h after embedding or can be stored in a cool, dry place indefinitely.

2.5.4 Sectioning wax-embedded material

These instructions are for the use of a HM 325 rotary microtome (Microm, Bicester, UK) but they will be readily adapted to other systems. Sections are cut to a thickness of ~10 - 12 μm to produce ribbons which are transferred to paper. Sections are selected and placed on polysine slides over a drop of water to promote spreading. Slides are allowed to dry in air overnight. To de-wax and re-hydrate sections incubate slides with 100% ethanol 3 x 10 min, 90%

ethanol/water 10 min, 50% ethanol/water 10 min, water 10 min, water 90 min. Slides are then air-dried and can be stored at RT indefinitely.

2.5.5 Embedding protocol for LR White resin

Samples were washed in buffer minus fixative for 10 min for 3 times (or overnight at 4°C) and then dehydrated using an ascending ethanol series (30, 50, 70, 90, 100%) with 30 min each change. Samples were then infiltrate with resin at 4°C by increasing from 10% resin in ethanol 1 h, 20% 1 h, 30% 1 h, 50% 1 h, 70% 1 h, 90% 1 h, 100% resin overnight before being transferred to gelatine capsules. Fill the capsules with resin to the top and seal to exclude air. Allow polymerize of resin either at 37°C for 2 days or by action of UV light at 4°C.

2.5.6 Sectioning of resin-embedded materials.

These instructions relate to the use of Reichert-Jung Ultracut Ultramicrotome.

For light microscopy sections were cut to a thickness of 1 -2 μm onto water and transfered to a drop of water on Vectabond-coated slides and allow to dry in air. For electron or fluorescent microscopy samples were cut into ultrathin sections to a thickness of ~80 nm when they should be silvery gold in colour.

2.5.7 Immunolabelling of plant cell walls using monoclonal antibodies

Labeling experiments were performed both in tobacco and *Physcomitrella patens*. Tobacco slides are usually imbedded in the wax thus a dewaxing step is required. To de-wax and re-hydrate sections, slides were incubated with 100% ethanol 3 x 10 min, 90% ethanol/water 10 min, 50% ethanol/water 10 min, water 10 min, water 90 min. Slides were then air-dried and can be stored at room temperature indefinitely.

Since mannan is masked by pectin in tobacco (Marcus et al. 2010), a 2-h sodium carbonate treatment (0.1 M, pH 11.4) are performed. After washing with 2 x 10 min phosphate buffered saline (PBS), a 2 h-incubation with pectate lyase (10μg/ml) in CAPS buffer was used to remove the pectin masking layer. Sections were ready for immunolabelling after wash with 3 changes PBS.

Typical labeling experiments in this project began with a 30 min treatment by the mannanase catalytic module and its CBM fusion at the same concentration (typically 0.1μM, 20uL) at 37 °C. In the case of esterase and its CBM constructs, 0.38uM enzymes were applied for 30min digestion. After washing away the enzyme, 3% (w/v) milk protein (such as Marvel Milk) was loaded to the section to minimize non-specific binding. After 30 min, to determine the amount of mannan remaining, 10 μL CBM-GFP (10 μM, with 2 mM Ca^{2+}) or primary antibody was applied for 1h. In the case of the antibody, a secondary anti-mouse antibody conjugated with FITC would then be applied for 1 h after washing off the primary antibody. Slides were then ready to be observed under microscope after adding one drop of anti-fade reagents (Citifluor glycerol) and putting on covering glass. All samples were examined using an Olympus BX-61 microscope with epifluorescence irradiation. Images were captured using a Hamamatsu ORCA285 camera. The Velocity software were used to count the pixels in order to quantify the enzyme effect. The following parameters were set for the quantification: find objects by intensity, 50-400 ; exclude objects by size, < 5 uM^2; measure objective, by mean, min, max voxel count, volum, sum; color objects, uniform , orange.

Chapter 3. The complex molecular architecture of mannan degrading hydrolases contributes to plant cell wall degradation

3.1 Introduction

Each year photosynthesis deposits a vast amount of the solar energy (3×10^{21} J) into plant cell walls (Ragauskas et al. 2006) making it the most abundant source of organic carbon in the biosphere. It consists mainly of an elaborate array of chemically distinct interlocking polysaccharides, and the recycling of this biomass, mainly through microbial degradation, plays a vital role in the global carbon cycle (Himmel et al. 2007, Himmel and Bayer 2009). Annually about 10^{11} tonnes of plant biomass is recycled, and the sugars released by these enzymes comprise a major energy and carbon resource within the biosphere (Brett CT 1996, Hogg et al. 2003). It is also evident that the microbial enzymes that catalyze the degradative process have significant industrial applications, particularly in the biofuels section where plant cell walls presents a renewable and thus sustainable substrate..

The plant cell wall is degraded by an array of glycoside hydrolases, carbohydrate esterases, polysaccharide lyases and lytic polysaccharide monooxygenases, which have been grouped into sequence-based families on the CAZy database (Henrissat 1991, Henrissat and Davies 1997, Jongkees and Withers 2014, Gardner et al. 2014). The database can be accessed online at http://www.cazy.org. β-1,4-Mannanases (defined henceforth as mannanase) are one of the key enzyme groups that attack the plant cell wall (Mccleary 1988). Grouped into three of the 133 sequence-based glycoside hydrolase families (GHs), 5, 26 and 113, these mannanases cleave either mannans or glucomannans. The protein fold [$(\alpha/\beta)_8$-barrel], catalytic machinery and mechanism of glycosidic bond cleavage (acid-base-assisted catalysis via a double displacement mechanism) are conserved in these three enzyme families (Bolam et al. 1996,

Hogg et al. 2001, Hogg et al. 2003). The two mannanases were used in this project, *Cj*Man5A and *Cj*Man26A (Braithwaite et al. 1995), originated from *Cellvibrio japonicus* (formerly *Pseudomonas fluorescens* subsp. *cellulosa*), which degrades mannan, xylan, cellulose, arabinan, galactan and pectins (McKie et al. 2001, Hazlewood and Gilbert 1998, Emami et al. 2002). Recent studies have shown *Cj*Man5A displays a more promiscuous specificity than *Cj*Man26A. While both enzymes bind mannose (Man) in the active site only the GH5 mannanase can accommodate glucose (Glc) in the distal regions of its substrate binding cleft (Hogg et al. 2003, Cartmell et al. 2008)

Enzymes that degrade plant cell walls often have a complex molecular architecture consisting of catalytic modules and one or more non-catalytic carbohydrate-binding modules (CBMs). These CBMs have been classified into sequence-based families in the CAZy database. For example, *Cj*Man5A contains two family 10 CBMs, CBM10-1 and CBM10-2. The tandem CBMs bind to crystalline polysaccharides of mannan or cellulose while CBM10-1 alone binds weakly to these ligands. (Hogg et al. 2003) Thus *Cj*Man5A is considered to attack mannans that either form crystalline structures or are integral components of the plant cell wall. In contrast a cohort of glycoside hydrolases that attack plant cell wall polysaccharides do not contain CBMs. An example of this type of enzyme is *Cj*Man26A (Hogg et al. 2003). Recent studies have suggested CBMs enhance the efficiency of enzymes by increasing enzyme substrate proximity, mediating prolonged and intimate contact between the respective catalytic module and its target substrate. (Bolam et al. 1998, Carrard et al. 2000, Herve et al. 2010)

Comparison of enzyme activity is usually determined through *in vitro* assays using purified polysaccharides. This approach, however, does not consider the context in which these enzymes function *in vivo*, where the complex structure of the plant cell wall is likely to

influence the function of these biocatalysts. In this chapter the activities of GH5 and GH26 mannanases and family 2 carbohydrate esterases (CE2s) *in vivo* were compared, and the capacity of CBMs, which display different specificities, to enhance the activity of these enzymes against intact plant cell walls was investigated. Specifically, CBM3a and CBM27 were used in this study to explore the role of cellulose and mannan specific CBMs in potentiating enzyme action against intact plant cell walls. The data provide insight into whether the specificity of the CBMs or the topology of the catalytic modules are the primary drivers for the specificity of these enzymes *in situ* (against intact plant cell walls).

3.2 Results

3.2.1 The preliminary results: Gene cloning, protein expression, fluorescent probes characterization and comparison

Construction of enzyme fusion and GFP-CBM probes.

In total six mannanase-CBM fusions, four esterase-CBM hybrid proteins and two GFP mannan-specific CBM (CBM27 and CBM35) conjugants were used in this project, **Figure 3.1**. All the CBM fusions were constructed by myself (**Table 3.1Figure 3.2**) except for *Cj*Man5A, *Cj*Man26A and *Cj*CE2B, which were generated by Dr Lei Zhao. The clones encoding the catalytic modules of the esterases and mannanases were generated previously. (Montanier et al. 2009a, Gilbert 2010, Cartmell et al. 2008)

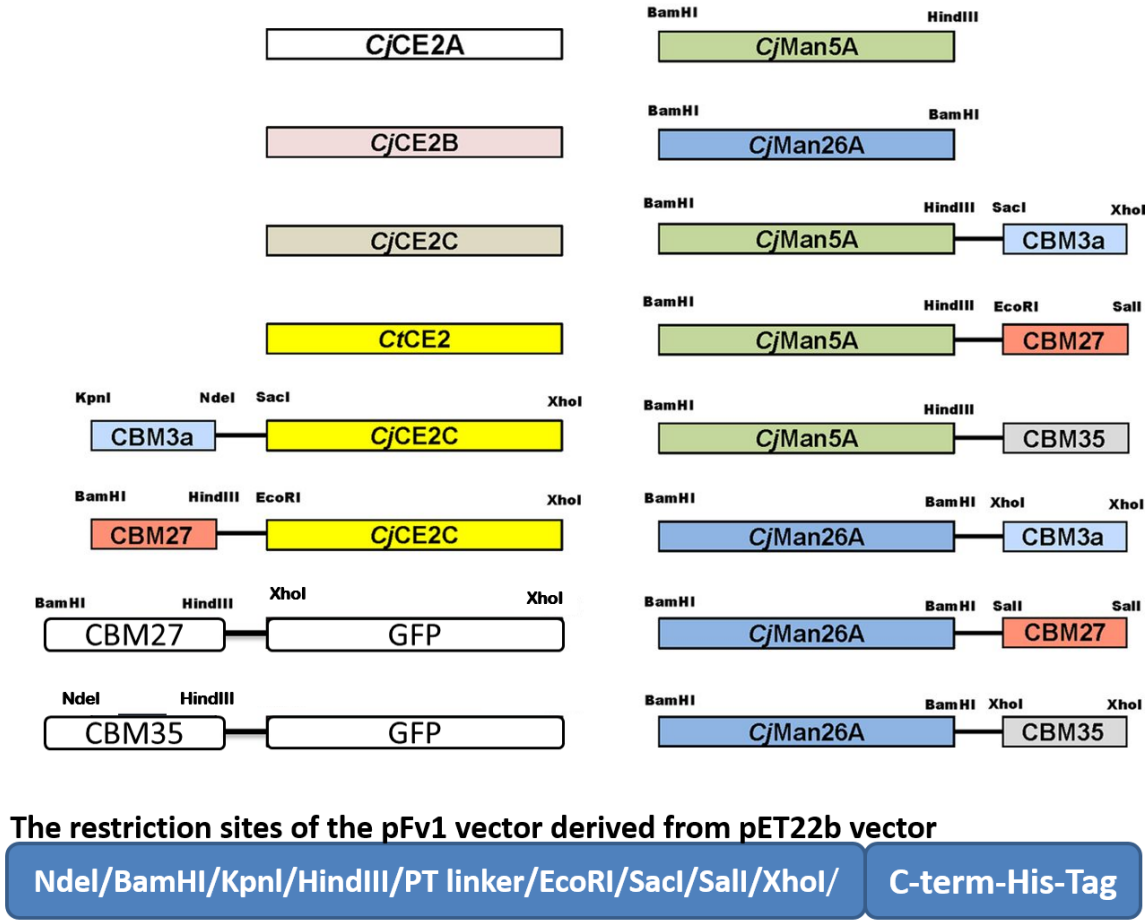


Figure 3.1 Schematic of the mannanases, esterases and GFP-CBM probes used in this study. The line joining the catalytic modules and CBMs represents a 15-residue Thr-Pro linker. The restriction sites used to clone the DNAs encoding the catalytic modules and CBMs are indicated. Different CBMs were cloned into pFV1 vectors containing *Cj*Man5A plus PT linker or *Cj*Man26A plus PT linker with a His tag in the C terminal.

Enzyme construct	Primers
CjMan5A-CBM3a	F- CCG**GAGCTC**GTATCAGGCAATTTG R- CCG**CTCGAG**ACCGGGTTCTTTACCC
CjMan5A-CBM27	F- CCG**GAATTC**AACGAAGCACGGTACG R- ACGC**GTCGAC**CATACCTCCTG
CjMan5A-CBM35	F- CGCG**GAATTC**GTACCGGAAGGCAATA R- ATAT**CTCGAG**GGTCAGGCCGCCCAG
CjMan26A-CBM3a	F- CCG**CTCGAG**GTATCAGGCAATTTG R- CCG**CTCGAG**ACCGGGTTCTTTACCC
CjMan26A-CBM27	F- TATA**GTCGAC**AACGAAGCACGGTACG R- ACGC**GTCGAC**CATACCTCCTG
CjMan26A-CBM35	F- CCG**CTCGAG**GTACCGGAAGGCAATA R- ATAT**CTCGAG**GGTCAGGCCGCCCAG
CjCE2C-CBM3a	F- ATTATA**CTCGAG**ATGGCCCAGGCGGAGCCG R- CGACGT**CTCGAG**CAACGCATTTTTTCCCGGATAAATGCC
CjCE2C-CBM27	F- CTTATA**GAATTC**ATGGCCCAGGCGGAGCCGGC R- CGACGT**CTCGAG**CAACGCATTTTTTCCCGGATAAATGCC

Table 3.1 Primers used for cloning in this study. (F, forward primer; R, the reverse one) The restriction sites have been highlighted.

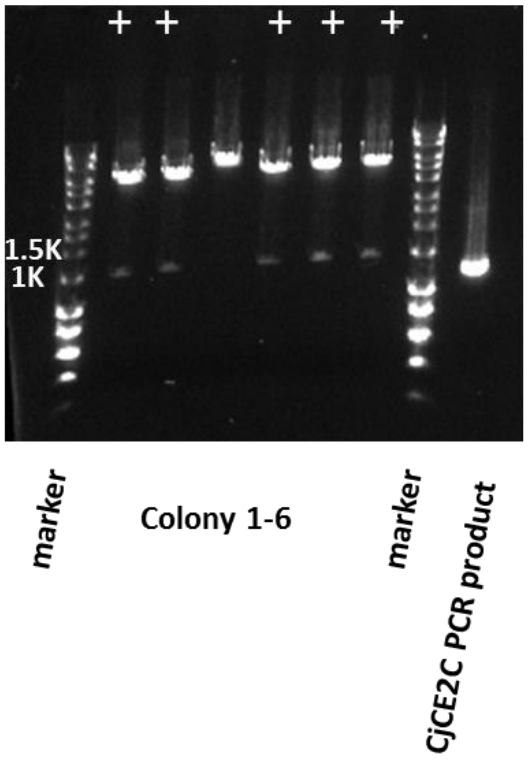


Figure 3.2 An example of cloning DNA encoding *Cj*CE2C into a vector encoding CBM3a. In this example plasmid DNA extracted from six colonies were digested with Xho I and EcoR I restriction enzymes. Colonies containing potential clone (size of 1.2 kb) were labelled with "+".

Overexpression of recombinant proteins.

All the recombinant proteins used in this project were expressed in *Escherichia coli* strain TUNER competent cell using standard Immobilized Metal Ion Affinity Chromatography (IMAC) unless otherwise specified, **Figure 3.3**. In several cases the proteins were not 100% pure, particularly the CE2 esterase conjugants, and the precise "active concentrations" of all the enzymes was required to compare the activity of these biocatalysts against plant cell walls. The esterase constructs were determined by assaying against the soluble substrate 4-nitrophenyl acetate esterase substrate, while the activity of the mannanases was evaluated using soluble carob galactomannan (low viscosity) as the substrate. This approach is predicated on the assumption that the CBMs deployed here would not affect the enzyme activity against soluble substrate (Mizutani et al. 2012, Herve et al. 2010, Gilbert 2010, Guillen, Sanchez and Rodriguez-Sanoja 2010, Gilbert, Stalbrand and Brumer 2008). The enzyme assay data are presented in Figure 3.4.

Name	Predicted molecular weight (Da)	Experimentally determined size (Da)	Mannose release from carob galactomannan (μg/ml/min/nmol)	Active concentration (μM) calibrated by Figure 3.4
CtCE2	23,456	~23,000	N/A	N/A
CjCE2A	39,778	~39,000	N/A	N/A
CjCE2B	40,063	~40,000	N/A	N/A
CjCE2C	41,557	~41,000	N/A	N/A
*Cj*Man5A	33,343	~38,000	1.81	1.00
*Cj*Man5A-$(PT)_{15}$-CBM3a	56,633	~45,000	1.32	0.72
*Cj*Man5A-$(PT)_{15}$-CBM27	57,471	~47,000-57,000	0.98	0.54
*Cj*Man5A-$(PT)_{15}$-CBM35	52,132	~46,000-53,000	1.03	0.56
*Cj*Man26A	43,573	~43,000	1.67	1.00
*Cj*Man26A-$(PT)_{15}$-CBM3a	66,932	~66,000	1.66	0.99
*Cj*Man26A-$(PT)_{15}$-CBM27	69,792	~69,000	1.52	0.91
*Cj*Man26A-$(PT)_{15}$-CBM35	62,673	~62,000	1.62	0.97

Table 3.2 Predicted and experimentally determined size of the proteins used in this study.

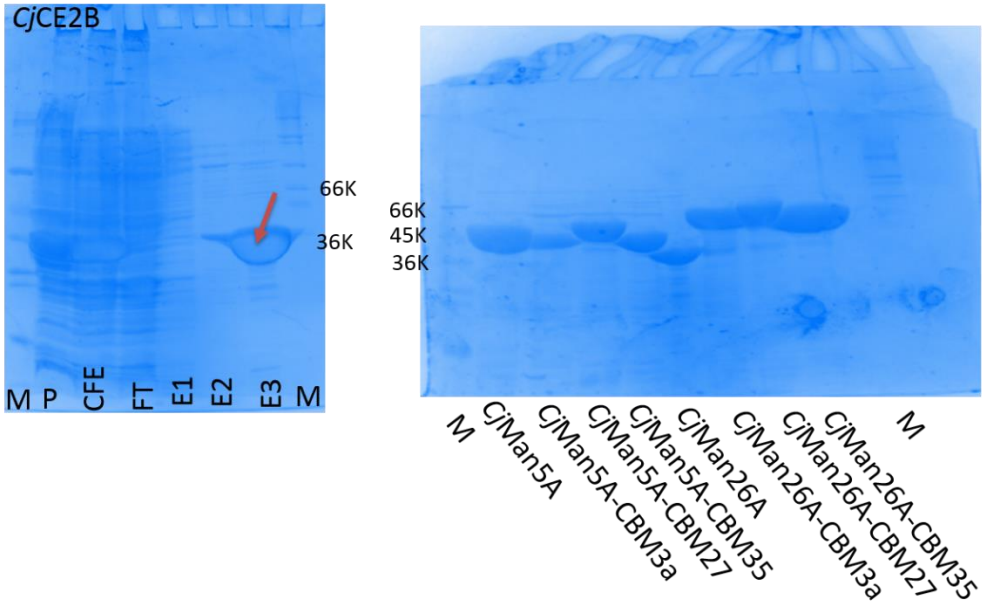


Figure 3.3 Example of purified recombinant enzymes subjected to SDS-PAGE using a 10% polyacrylamide gel. Protein expression was carried out using *E. coli* Tuner competent cells and purified by IMAC, except for *Cj*Man26A which lacked a His_6-tag, as detailed in Chapter 2 Section 2.2. The lanes were labeled as follows: M, Marker; P, insoluble pellet; CFE, Cell free extract; E1-3, Elutions by imidazole gradient of 0mM, 5mM and 150mM respectively in Talon buffer.

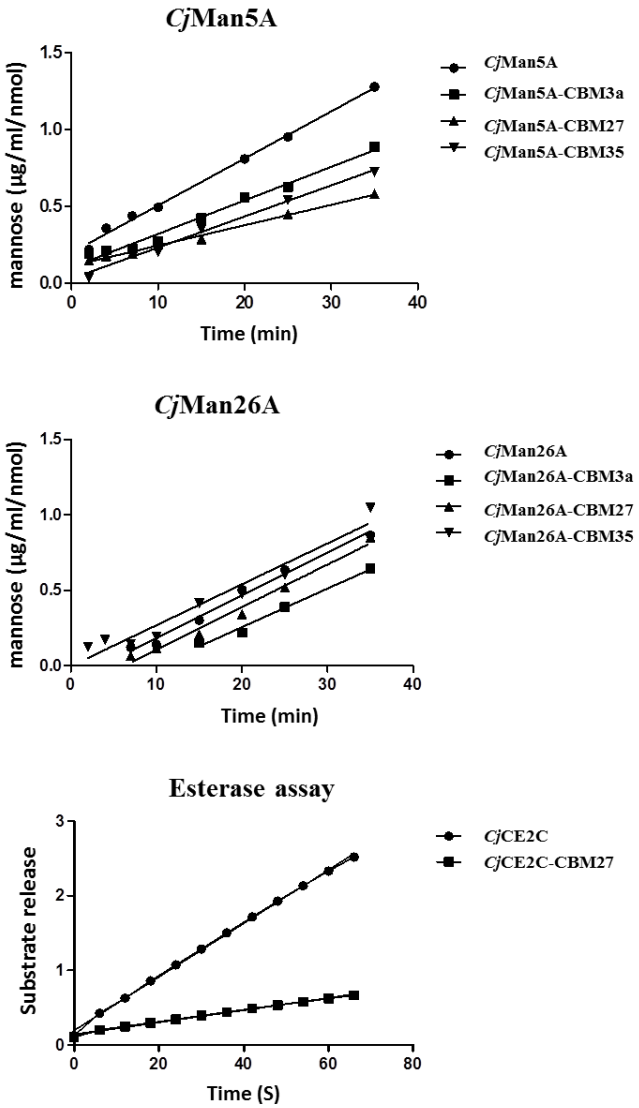


Figure 3.4 Examples of assays used to determine the "active" concentration of mannanase and esterases. The substrate used was medium viscosity carob galactomannan at 10mg/ml in 20mM Tris buffer at 37^{o}C. Cleavage was monitored by the production of reducing sugar determined using DNSA reagent and 4-Nitrophenyl acetate esterase substrate as described in Chapter 2.4.

CBM27-GFP is a much better probe than CBM35-GFP.

Two CBMs were conjugated with GFP to generate the mannan probes CBM27-GFP (tight mannan binder) and CBM35-GFP. The two probes were then used to detect mannan in the context of plant cell walls *in vivo*. As shown in **Figure 3.5**, CBM27-GFP labeled the pith, xylem and epidermis/cortical regions of tobacco cell wall sections considerably better than CBM35-GFP, probably due to the fact that CBM27 binds to mannan ten-fold more tightly than

CBM35 (Bolam et al. 2004, Marcus et al. 2010). For this reason only CBM27-GFP was used to probe mannan in cell walls during this project.

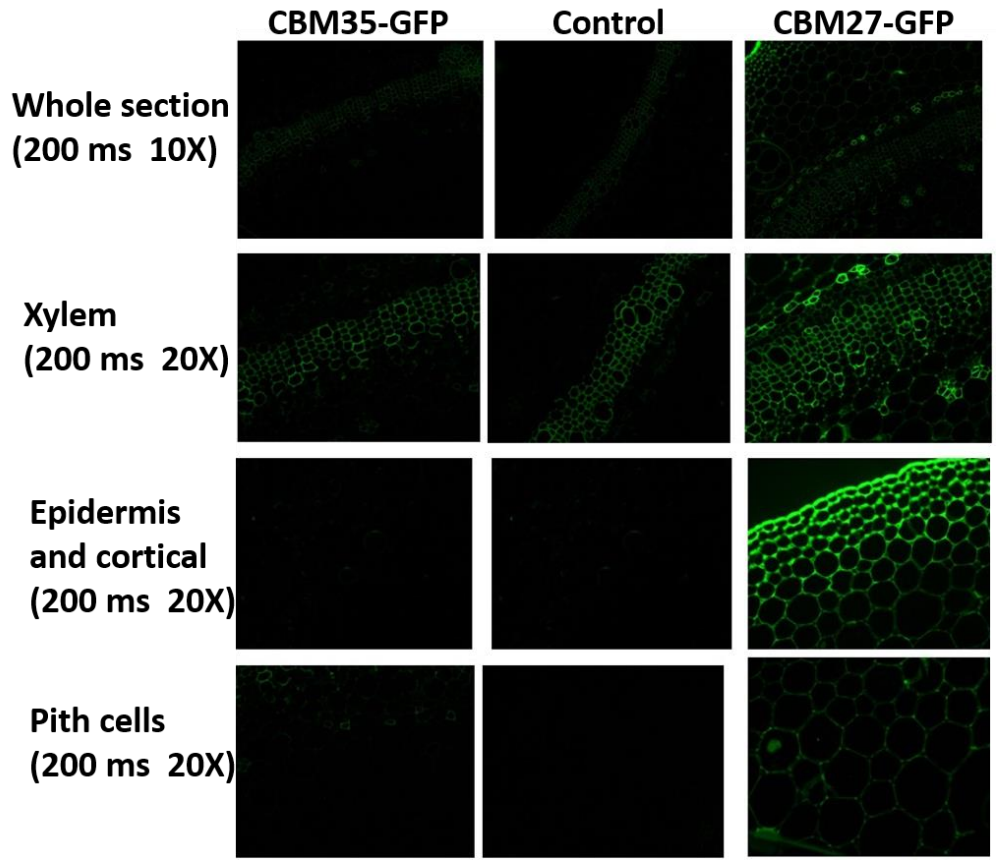


Figure 3.5 Comparison of the two mannan-specific probes labeling tobacco sections: CBM27-GFP and CBM35-GFP at 10 μM in PBS for 1.5 h. Magnification: 10X or 20X; exposure time: 200 ms.

Evidence that CCRC-M170 antibody is acetyl-mannan specific and *Physcomitrella* cell wall contains acetyl-mannan

The mouse monoclonal antibody CCRC-M170, from the Complex Carbohydrate Research Center USA, was believed to bind to mannan (M. Hahn personal communication) based on the antigen used to raise the antibody was acetylated mannan. To explore the specificity of CCRC-M170 in more detail, *Physcomitrella* cell walls, pretreated with different buffers, were probed with the antibody. The data, **Figure 3.6**, showed that CCRC-M170 bound to *Physcomitrella* pretreated with CAPSO (pH 9.0), but did not recognize cell walls pretreated with strong alkali such as NaOH or Na_2CO_3. As strong alkali removes acetyl groups from polysaccharide, these

data suggest that CCRC-M170 recognizes acetylated polysaccharides, likely mannan. Further evidence supporting the specificity of CCRC-M170 for acetylated mannan, was provided when freshly cut *Physcomitrella* sections treated with the family 2 carbohydrate esterase *Cj*CE2C, which removes acetyl groups from mannan (Zhang et al. 2014, Gilbert 2010, Montanier et al. 2009a, Correia et al. 2008), was probed with CCRC-M170. The data, presented in **Figure 3.6**, showed that *Cj*CE2C destroyed the epitope recognized by CCRC-M170, confirming that the antibody specifically targets an acetylated polysaccharide, likely mannan.

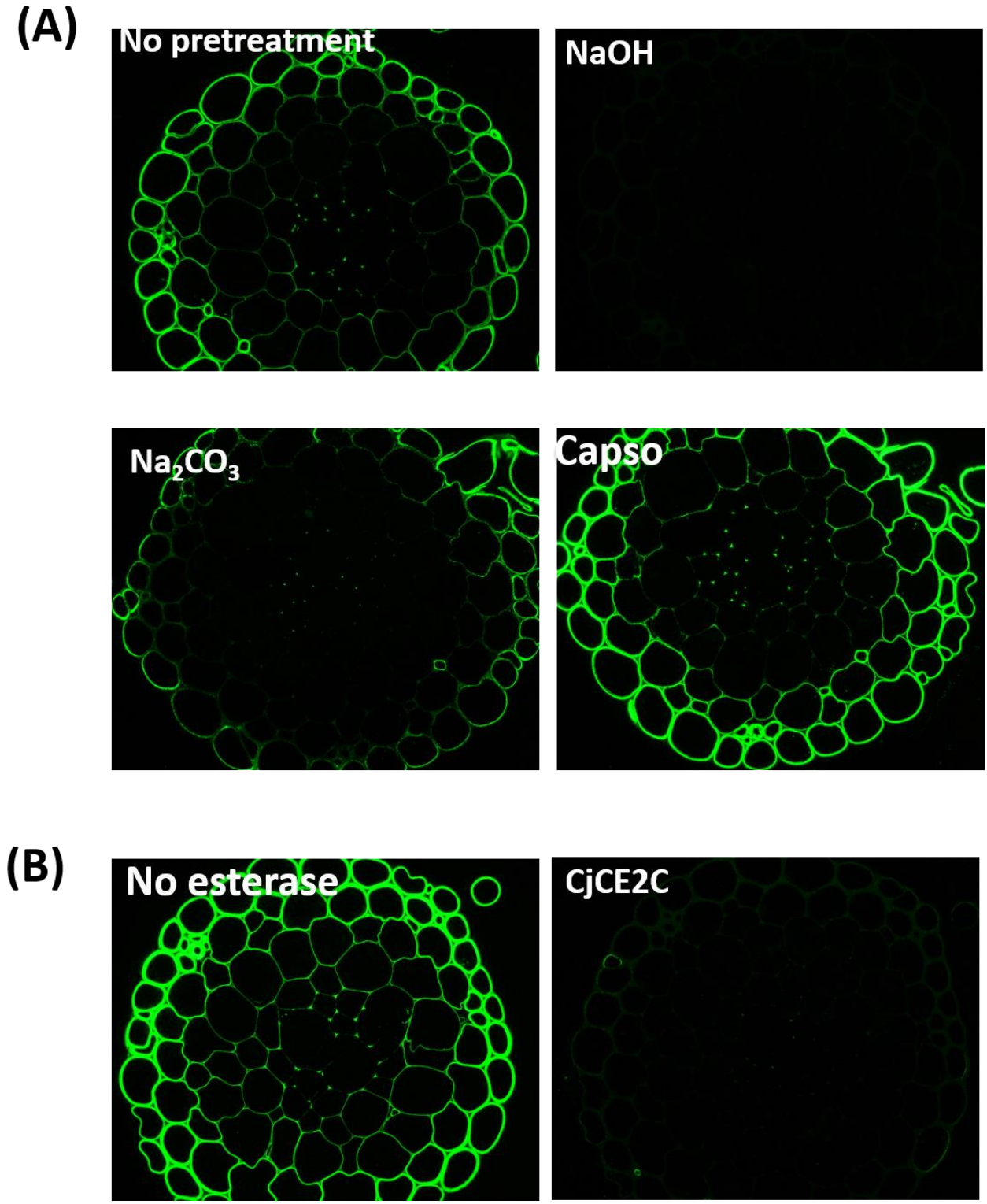


Figure 3.6 CCRC-M170 epitope probing. Freshly cut *Physcomitrella* sections were labelled with CCRC-M170 after different pre-treatment for 1 h with NaOH 500 mM, 100 mM Na_2CO_3 pH 11.0, 50 mM CAPSO buffer, pH 9.6 in (A)or 26 μM *Cj*CE2 esterase in phosphate buffered saline in (B).

Issues when using air-exposed *Physcomitrella* cell wall sections and CCRC-M170 antibody.

Before this study, it was assumed at the Complex Carbohydrate Research Center that sectioned slides can be stored at room temperature and used for a long time (months) after sectioning. However, CCRC-M170 no longer bound to *Physcomitrella* sections that had been generated three weeks prior to the experiment, **Figure 3.7**. The loss of the epitope is likely caused by the oxidation of the acetyl groups on the mannan. For this reason, all the *Physcomitrella* sections used in this study were either used immediately or stored at 4 ^{o}C for no longer than 5 days.

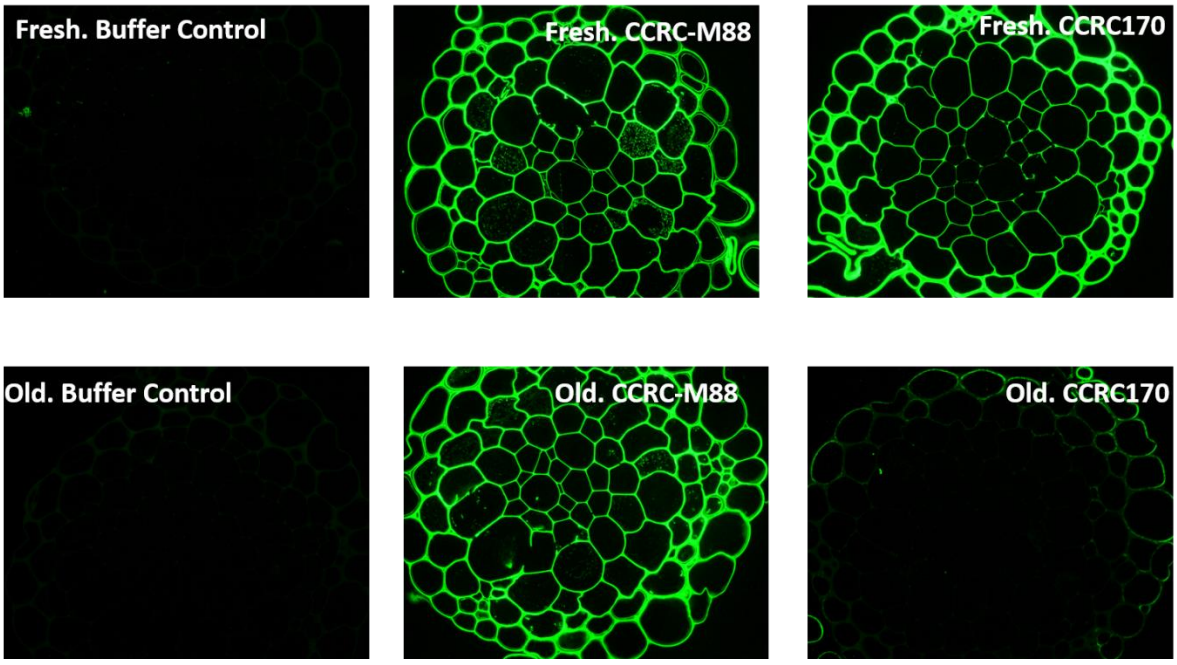


Figure 3.7 slides exposed in air for 3 weeks lost binding capacity to CCRC-M170. CCRC-M88, a xyloglucan specific antibody, was used as a positive control to label *Physcomitrella*. Old means the slides were 3 weeks old and fresh means less than 5 days. Incubation time: 1 h; buffer, PBS buffer; antibody, 1 in 5 dilution in PBS buffer.

Masking of mannan to antibodies

In tobacco cell walls, mannan is masked to the external milieu by pectin and thus enzymatic removal of the galacturonic acid-containing polysaccharide is required to label mannan with antibodies or CBMs that recognize the backbone of the mannose polymer (Zhang et al. 2014, Marcus et al. 2010, Herve et al. 2010). In contrast, the data presented in **Figure 3.8**A showed CBM27-GFP bound to *Physcomitrella* cell walls prior and subsequent to alkali and pectate lyase treatment. Thus, in *Physcomitrella* cell walls the mannan was not masked by any pectin that may be present. This suggests that in non-vascular plants such as moss the cell wall architecture may bear some significant difference with the advanced plants. It should also be noted that **Figure 3.8**B showed that the removal of acetyl groups by the esterase CjCE2C did not promote mannan recognition by the CBM27-GFP probe. This is consistent with previous observations that CBM27 binds to unacetylated mannan, highlighting the existence of a large amount of unacetylated mannan even in lower plant species like *Physcomitrella* (Correia et al. 2010, Zhang et al. 2014, Bolam et al. 2004, Hogg et al. 2003).

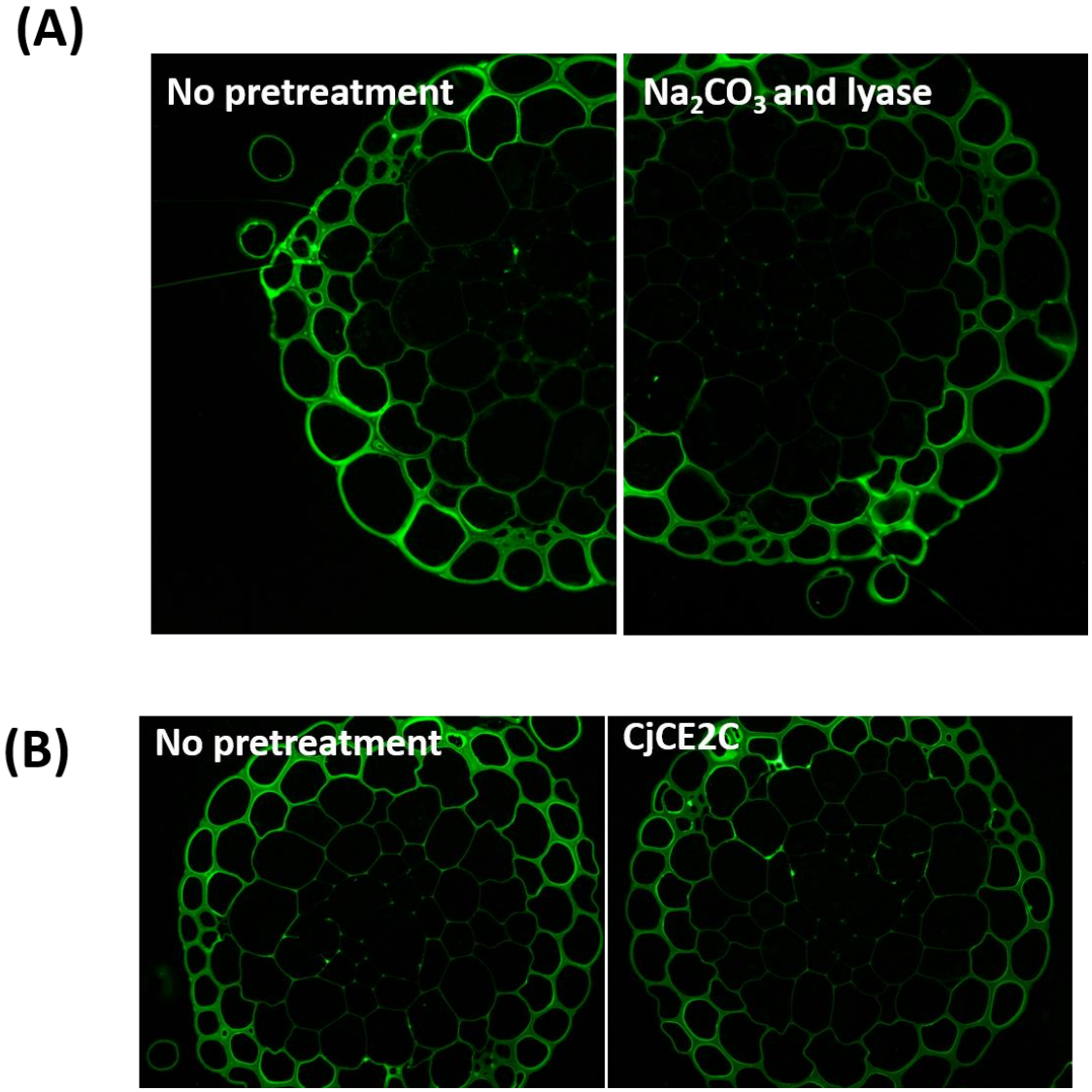


Figure 3.8 Probing *Physcomitrella* wall with CBM27-GFP. Freshly cut *Physcomitrella* sections were labelled with 10μM CBM27-GFP for 1h after pre-treating for 1 h with 100 mM Na_2CO_3 buffer, pH 11.0, containing 200 nM pectate lyase (*Cj*PL2A) in (A) or 26 μM *Cj*CE2C in (B).

Screening of other mannan-related antibodies

Besides CCRC-M170, Professor Michael Hahn's Lab in the Complex Carbohydrate Research Center has several other mannan-related antibodies (M. Hahn personal communication), which were screened for CCRC-M170-like binding profiles. CCRC-M166, 167, 168, 172-175 were generated in mice using Guar galactomannan antigen as the antigen, CCRC-M169 and CCRC-M170 were raised against acetylated mannan, and CCRC-M171 was generated using linseed

(flax seeds) mucilage as the antigen. NaOH at 0.5 M was used to remove the acetyl groups on mannan, and fresh sections were labeled before and after treatment with alkali. The results, **Figure 3.9**, showed that CCRC-M169 is also an acetyl mannan specific antibody as it does not bind to cell walls pretreated with alkali. The binding properties of the other antibodies, although not influenced by NaOH pretreatment, showed a great deal of diversity exemplified by CCRC-M173 and CCRC-M174, which only bound to outer epidermis areas of the plant sections that were not distinguished by the other probes. This suggests that the mannan in the *Physcomitrella* cell wall is not homogeneous and the antibodies target their unique and distinct epitopes within the context of the intact plant cell walls.

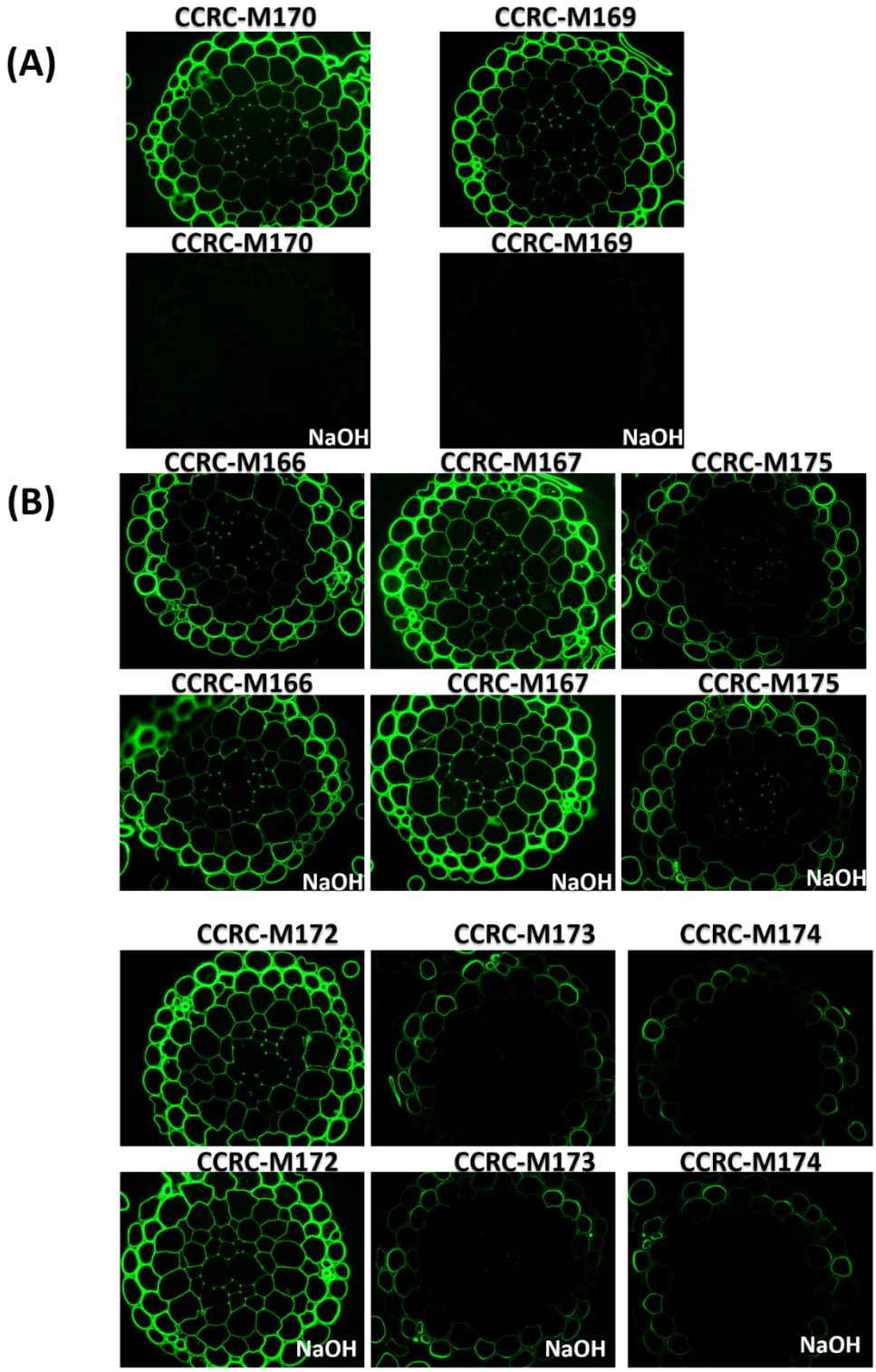


Figure 3.9 Screening of mannan-related antibodies. Freshly cut *Physcomitrella* sections were labeled with various antibodies raised against mannan before and after NaOH pre-treatment (0.5 M, 1 h). Panel A, two possible acetyl mannan-specific antibodies; Panel B, antibodies whose binding properties were not affected by the removal of acetyl groups. (Detailed methodology described in Chapter 2.5)

3.2.2 CBMs significantly potentiate mannanases and esterases activities in different plant cell walls

*Cj*Man5A and *Cj*Man26A digestion of mannan in tobacco walls.

To explore the capacity of the catalytic module of *Cj*Man5A (*Cj*Man5A$_{CM}$) to attack tobacco cell walls, different concentrations of the enzyme were incubated with sections of the plant pretreated with alkali and a pectate lyase to remove the pectins, which are known to mask mannan from proteins recognition. The activity of the mannanase was quantified by determining the amount of mannan remaining after enzyme treatment, by probing the sections with LM21, a mannan specific antibody (Marcus et al. 2010, Ordaz-Ortiz, Marcus and Knox 2009). The data presented in **Figure 3.10** revealed an exponential correlation between the amount of enzyme incubated with the plant cell wall and the extent of mannan degradation. These results suggest that the mannan in tobacco cell walls is equally accessible to the GH5 mannanase.

The catalytic module of *Cj*Man26A (*Cj*Man26A$_{CM}$) was also incubated with tobacco cell walls at different concentrations. Unlike *Cj*Man5A$_{CM}$, *Cj*Man26A$_{CM}$ did not present a direct correlation between the amount of enzyme used and the extent of mannan degradation. On the contrary, data showed that up to 60% of tobacco mannan in the wall was not accessible to *Cj*Man26A$_{CM}$, even at an enzyme concentration of 600 nM, **Figure 3.11**. One possible reason could be that the active site of *Cj*Man5A can accommodate glucose in distal subsites while *Cj*Man26A displays a strict requirement for mannose at both the negative binding subsites and thus limits the accessibility of the substrate. (Tailford et al. 2009, Hogg et al. 2003)

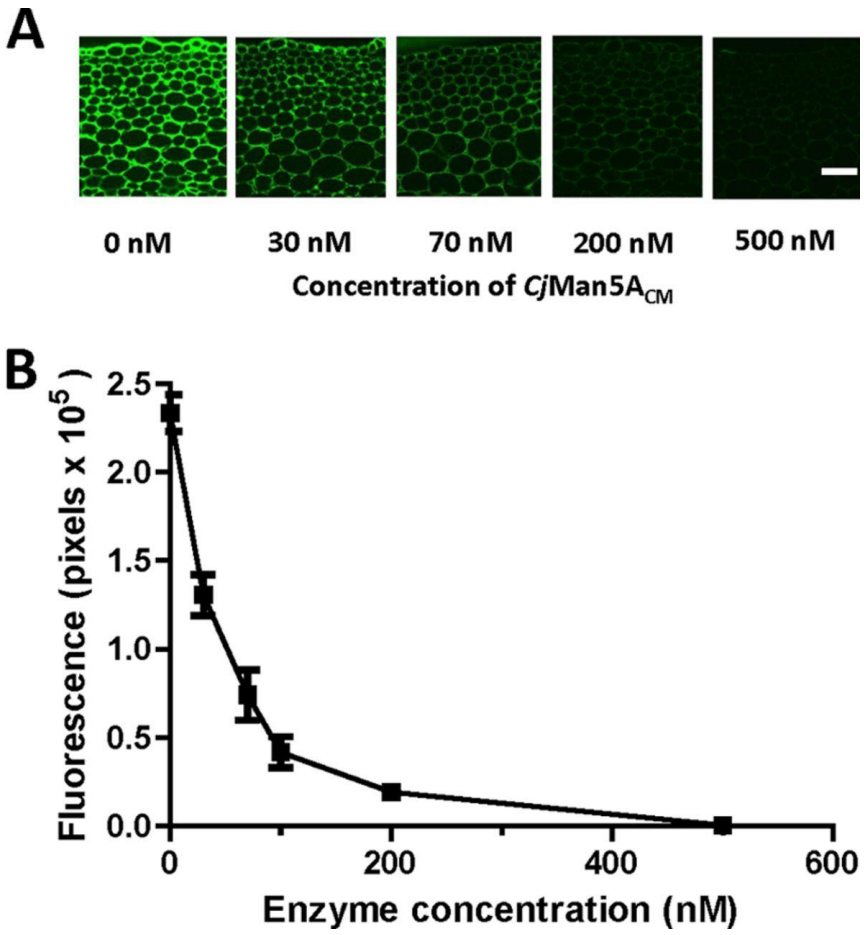


Figure 3.10 The activity of *Cj*Man5A$_{CM}$ against tobacco cell walls. Sections were incubated with *Cj*Man5A$_{CM}$ at the concentrations indicated for 20 min at 25 °C. The mannan remaining after enzymatic treatment was determined by probing the sections with the mannan-specific antibody LM21. Panel A displays representative immunofluorescence micrographs of LM21 binding to enzyme-treated sections. Panel B shows the fluorescence intensities of the enzyme treated sections. Each datum point presented in this figure is derived from three or four different sections. The scale bar is 50 μm. Quantification was done here and in future experiments by quantifying the pixels in the fluorescent image using Volocity software. (Error bars show the standard error. Detailed methodology described in Chapter 2.5.)

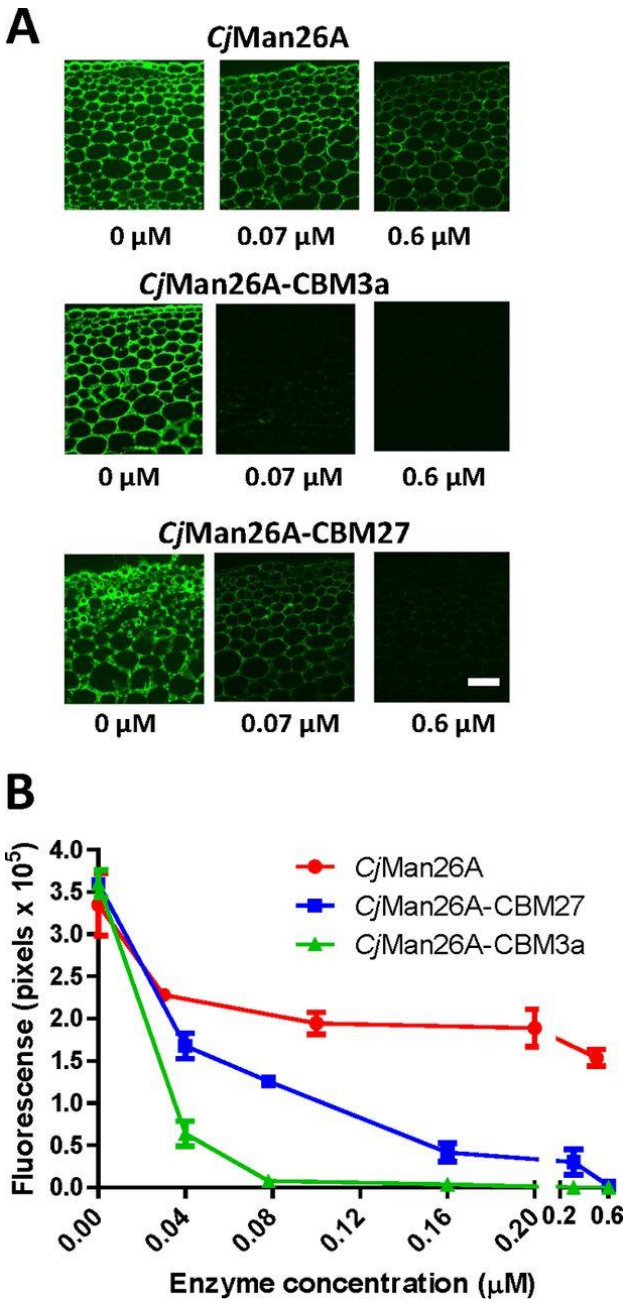


Figure 3.11 The activity of *Cj*Man26A$_{CM}$ against tobacco cell walls. Stem sections were incubated with *Cj*Man26A$_{CM}$ and the enzyme fused to cellulose (CBM3a) and mannan (CBM27) binding CBMs at the concentrations indicated for 20 min at 25 °C. The mannan remaining after enzymatic treatment was determined using the mannan-specific antibody LM21. Panels A and B display representative immunofluorescence micrographs and the quantified fluorescence intensities of all the sections, respectively. The scale bar is 50 µm. Error bars show the standard error.

CBM3a assists *Cj*Man5A and *Cj*Man26A in degrading mannan in tobacco walls.

To determine whether CBMs contributed to mannan degradation, the GH5 and GH26 mannanases used here were fused to different CBMs and the activities of the resultant enzymes were assessed. When the GH5 mannanase was fused to cellulose (CBM3a) and mannan (CBM27 and CBM35) binding CBMs, the resultant enzymes were more active against tobacco cell walls than *Cj*Man5A$_{CM}$, (Figure 3.12). At 20 nM, *Cj*Man5A$_{CM}$ removed 25% of the mannan in tobacco walls during a 5 min incubation, while *Cj*Man5A-CBM3a, *Cj*Man5A-CBM27 and *Cj*Man5A-CBM35 degraded 85%, 62% and 50% of the cell wall mannan, respectively, under the same reaction conditions used for *Cj*Man5A$_{CM}$. Thus, CBM3a caused the largest increase in enzyme activity (about five times compared to *Cj*Man5A$_{CM}$) against mannan in tobacco cell walls. The observation that CBM35 binds to mannan less tightly than CBM27, may be the reason why the family 35 module was less effective in potentiating the activity of *Cj*Man26A than CBM27.

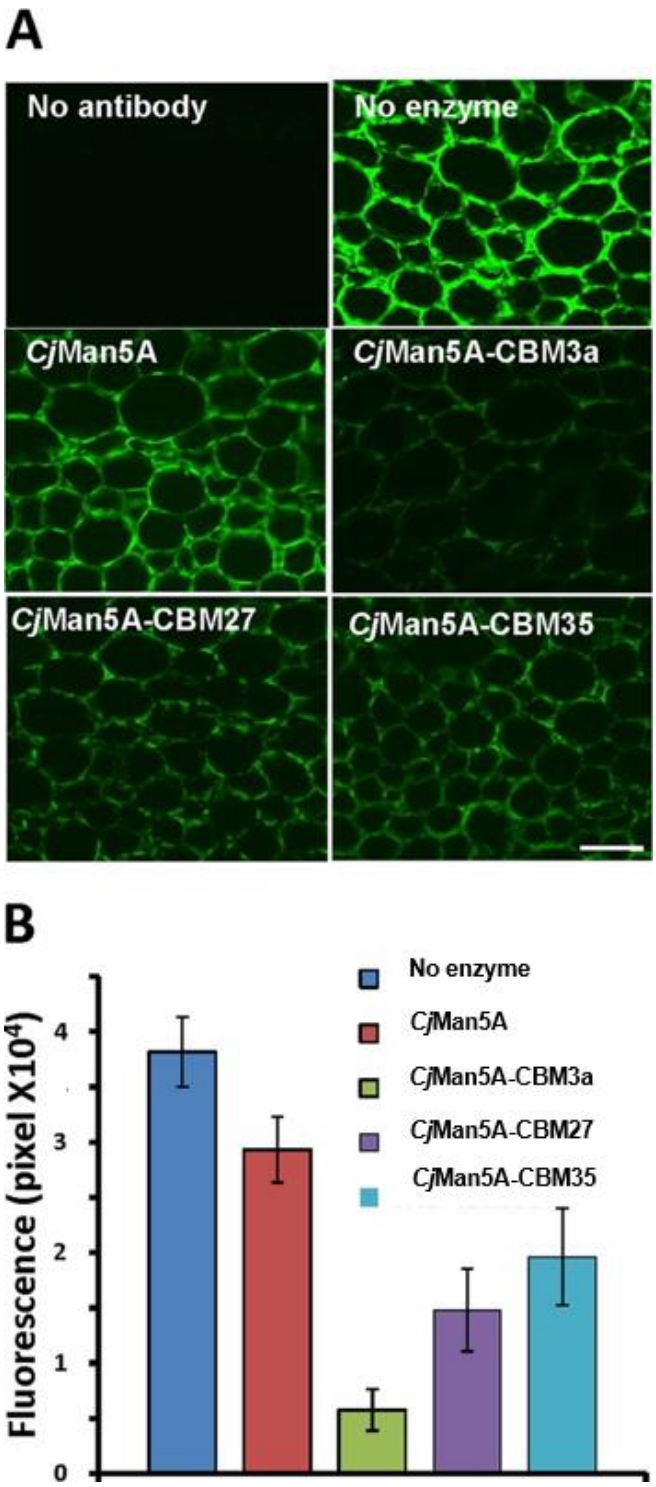


Figure 3.12 Comparison of the different roles of CBMs in potentiating *Cj*Man5A activity against tobacco cell walls. Sections of tobacco stems were incubated with 0.02 μM *Cj*Man5A and derivatives of the enzyme containing CBMs for 5 min (Chapter 2.5). Representative fluorescence images of enzyme-treated tobacco sections labelled by the mannan-specific antibody LM21 are shown in Panel A. Quantification of the fluorescence intensities of the equivalent images are shown in Panel B. The scale bar is 50 μm. Three replicates were performed and error bars showed the standard error.

In a similar series of experiments using *Cj*Man26A as the catalytic entity, both CBM3a and CBM27 greatly enhanced the activity of the enzyme, **Figure 3.11**. Similar to the *Cj*Man5A conjugates, the cellulose binding module was more effective than the mannan-specific CBM in increasing the activity of the GH26 mannanase. Significantly, *Cj*Man26A, when fused to either CBM, degraded the regions of mannan that was inaccessible to *Cj*Man26A$_{CM}$. It indicates that some CBMs can guide the enzyme to the area that is inaccessible to the catalytic module. These data, which are discussed in detail in Section 3.3 and can also be applied to the *Cj*Man5A results, may reflect the extensive cellulose network in the tobacco wall, the mobility of CBM3a on the surface of cellulose as well as the close association between mannan and cellulose.

*Cj*Man5A and *Cj*Man26A digestion of mannan in *Physcomitrella*.

In a similar series of experiments to those describe above using tobacco cell walls as substrate, the activity of *Cj*Man5A$_{CM}$ and *Cj*Man26A$_{CM}$ against *Physcomitrella* cell walls was also evaluated. The data, **Figure 3.13**, showed that up to a concentration of 300 nM *Cj*Man5A$_{CM}$ digested significantly more mannan than *Cj*Man26A$_{CM}$. It appears, therefore, that the mannan in *Physcomitrella* cell walls is more accessible to *Cj*Man5A than *Cj*Man26A. However, *Cj*Man26A$_{CM}$ plateaus at ~20% degradation unless very high levels of enzyme are added ($\geqslant$ 2.5 µM) after which degradation increases to about 60%. With *Cj*Man5A$_{CM}$ mannan degradation plateaus at 60% degradation at 100 nM and remains constant up to 10 µM.

A detailed comparison of two CBMs potentiating of the activity of *Cj*Man26A and *Cj*Man5A against *Physcomitrella* cell walls.

To evaluate the influence of CBMs on the activity of the two mannanases against *Physcomitrella* cell walls, the enzymes, fused to CBM3a or CBM27, were titrated against the plant. As shown in **Figure 3.13**, fusing CBM3a or CBM27 to *Cj*Man26A$_{CM}$ enhanced the

activity of the enzyme up to a concentration of 100 nM by as much as 6-fold. Increasing enzyme concentration up to 10 μM resulted in no further increase in the extent of mannan degradation. Although the effect of CBM27 and CBM3a were broadly similar, at very low enzyme concentration, 50 nM, the mannan binding module was slightly more effective than the cellulose binding CBM at potentiating mannanase activity. These data indicate that the CBMs target enzymes to the "degradable" regions of mannan, but do not enhance access to the areas of the polysaccharides in *Physcomitrella* cell walls that are completely recalcitrant to attack by *Cj*Man26A$_{CM}$.

In the case of *Cj*Man5A, the catalytic module alone could not digest all mannan in the cell wall, but fusion to both CBM3a and CBM27 enhanced enzyme activity *in vivo,* **Figure 3.13**. With respect to the CBM3a a similar pattern emerges to what was observed for the *Cj*Man26A-CBM constructs; the cellulose binding module increases the efficiency by which the enzyme degrades the accessible 75% of the mannan (mediated by 50 nM of *Cj*Man5A when fused to a CBM versus 600 nM for *Cj*Man5A$_{CM}$), **Figure 3.13**. CBM27, however, targeted the enzyme to both the accessible and inaccessible regions of mannan, with 10 μM of enzyme completely degrading the polysaccharide.

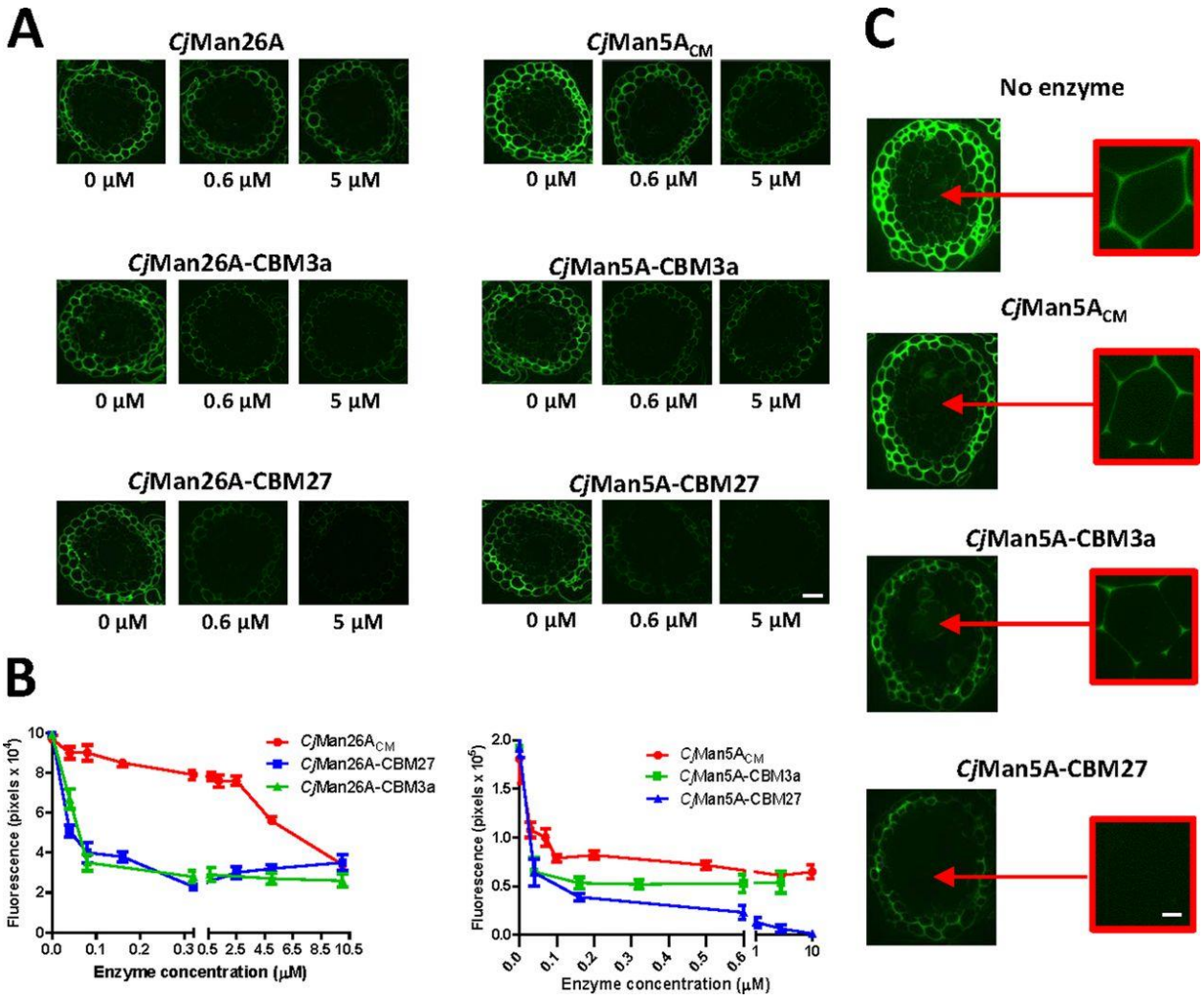


Figure 3.13 Differential roles of the CBMs in the potentiation of *Cj*Man26A and *Cj*Man5A against *Physcomitrella* cell walls. Panel A shows representative immunofluorescence micrographs and Panel B shows the quantified fluorescence intensities of the cells in the perimeter of all the sections. Panel C shows immunofluorescence micrographs in the red boxes of the representative examples of the inner cells (the pith area), magnified 5 more times, of Physcomitrella treated with *Cj*Man5A and derivatives of the enzyme containing the various CBMs. The reactions were carried out in PBS for 20 min at 37 ºC. A standard protocol for these experiments is described in Chapter 2.5. The scale bar is 25 µm in A and 5 mm in the red boxed images in C. (Zhang et al. 2014)

CBM27 also potentiates esterase action on acetylated mannan in *Physcomitrella*.

The demonstration that CCRC-M170 can be used to probe for the presence of acetylated mannan (**Figure 3.6**), facilitated the evaluation of the capacity of CBMs to influence the activity of mannan esterases *in situ*. To explore this, CBMs were fused to the family 2 carbohydrate esterase (CE2) *Cj*CE2C to generate CBM27-*Cj*CE2C and CBM3a-*Cj*CE2C.

Fresh cut moss sections were treated with the three esterase constructs and then probed with the acetylated mannan-specific antibody. These results demonstrate that CBM27 increased the capacity of the esterase to remove acetyl groups from mannan by ~four-fold (**Figure 3.14**), while CBM3a only slightly increased enzyme activity, by less than 35%. These data show that, in addition to glycoside hydrolases, CBMs can also enhance the activity of esterases, and are consistent with the observation that a cohort of carbohydrate esterases are appended to CBMs in nature (Kroon et al. 2000, Gordillo et al. 2006, Ferreira et al. 1993).

In addition to *Cj*CE2C, the activity of three other CE2 esterases against the *Physcomitrella* cell wall was assessed, Figure 3.14. Apart from *Cj*CE2C, none of the other esterases showed significant activity towards acetylated mannan in *Physcomitrella* cell walls. These enzymes have been characterized against a series of model substrates including partially acetylated gluco-, manno- and xylopyranosides, which showed that the CE2 enzymes have a strong preference for deacetylation at the 6-position in aldohexoses (Topakas et al. 2010).

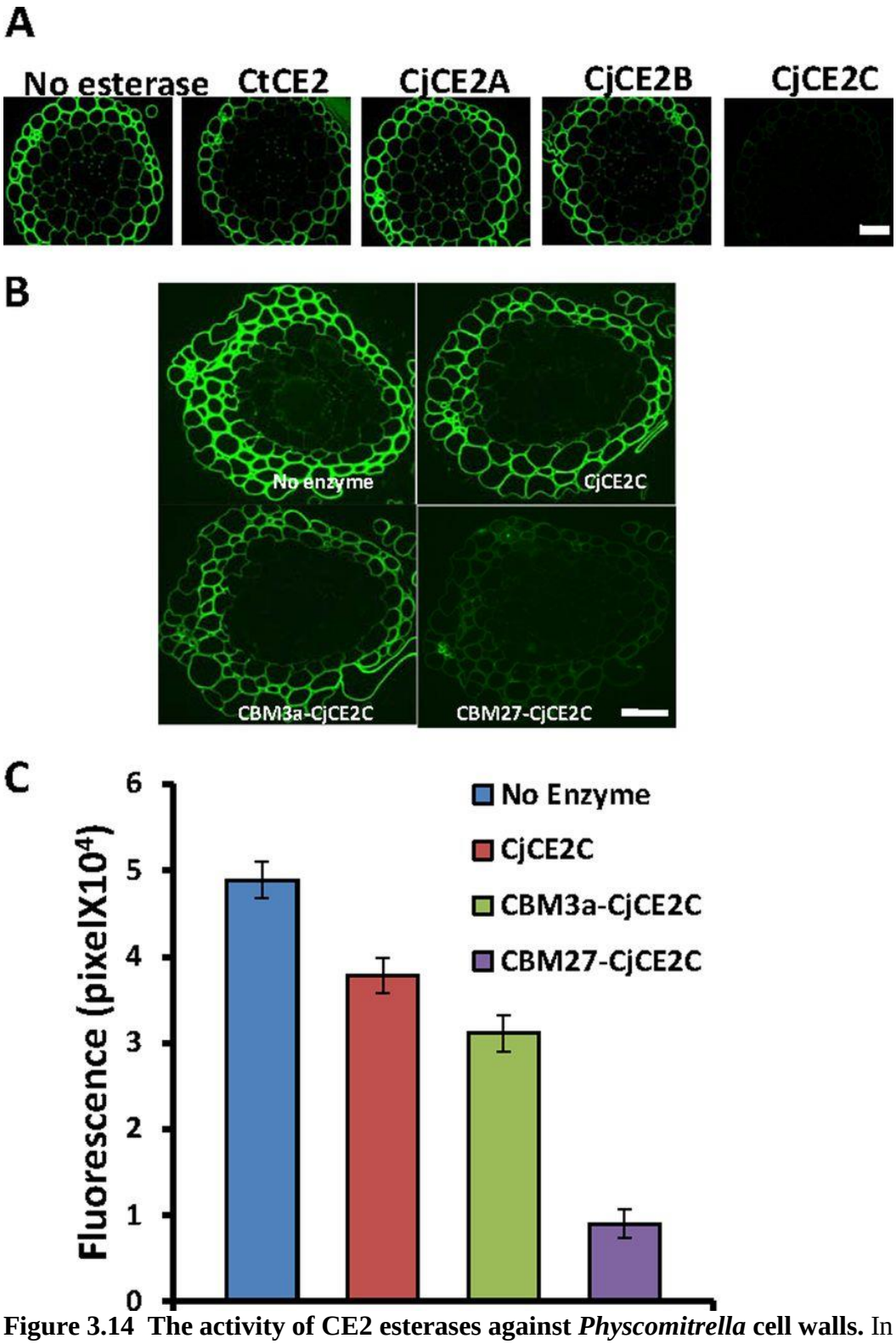


Figure 3.14 The activity of CE2 esterases against *Physcomitrella* cell walls. In Panel A the four CE2 esterases (lacking a CBM) at 6 μM were incubated with *Physcomitrella* sections for 1 h as described in Chapter 2.5. The presence of acetylated mannan was detected using the antibody CCRC-M170 and immunofluorescence microscopy. In Panels B and C *Physcomitrella* sections were treated with 0.13 μM *Cj*CE2C, *Cj*CE2C-CBM27, and *Cj*CE2C-CBM3a for 70 min and the presence of acetylated mannan was detected as described above. Panel B shows the fluorescence images of the sections, and Panel C the quantified fluorescence intensities. The scale bar is 25 μm. Error bars show the standard error.

3.2.3 Influence of a non-reducing-end mannan binding CBM on mannanase activity

In collaboration with the Sakka Group in Japan (Mizutani et al. 2012), the *Clostridium thermocellum* hypothetical protein, Cthe_0821, defined as *Ct*Man5A here, has been studied. The enzyme is a major component of the *C. thermocellum* cellulosome when the bacterium is grown on cellulose (Mizutani et al. 2012). The enzyme was shown, by the Sakka group, to be a modular endo-mannanase comprising a N-terminal signal peptide, a GH5 catalytic module, a family 32 CBM (CBM32), and a C-terminal type I dockerin module. Professor Sakka's group and Dr Adam Jackson at Newcastle showed that the CBM32 targets the non-reducing ends of mannooligosaccharides, and influenced the product profile of the enzyme against mannooligosaccharides (Mizutani et al. 2012). The data presented in (**Figure 3.15**showed that in reactions catalyzed by the catalytic module of *Ct*Man5A there was an imbalance of products exemplified by the hydrolysis of mannotetraose, where the generation of mannotriose was not associated with the production of mannose (**Figure 3.15**). Such an imbalance of products is diagnostic of transglycosylation reactions (see Charnock et al., 1997). When the catalytic module was fused to CBM32 there was no imbalance of the reaction products using mannoligosaccharides as substrates, suggesting an absence of transglycosylation. The CBM32 may bind to the non-reducing end of mannooligosaccharides in the vicinity of the +1 subsite, preventing these molecules from attacking the glycosyl-enzyme intermediate (*Ct*Man5A is a retaining glycoside hydrolase where the double displacement mechanism involves a covalent enzyme-glycosyl intermediate) (Mizutani et al. 2012, Mizutani et al. 2014, Rye and Withers 2000) and thus participating in transglycosylation reactions (Mizutani et al. 2012). It is possible, therefore, that the CBM32 may prevent futile transglycosylation reactions (where these is no net hydrolysis of the substrate) from occurring when the enzyme is attacking mannans. To test this hypothesis, in this project, recombinant forms of the catalytic module of *Ct*Man5A (*Ct*Man5A$_{CM}$) and the enzyme containing CBM32 (*Ct*Man5A$_{CM}$-CBM32) were

purified to electrophoretic homogeneity, **Figure 3.16**, and then evaluated against soluble and insoluble purified mannans, and the polysaccharide present in intact plant cell walls.

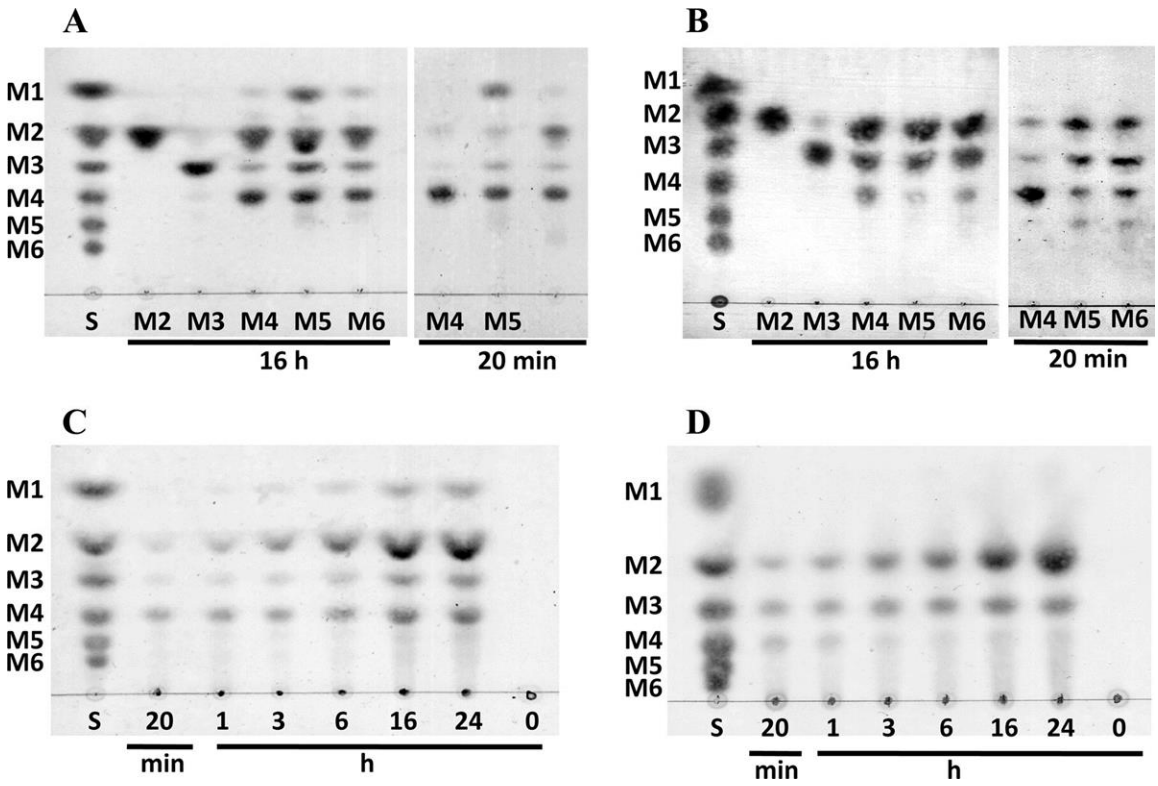


Figure 3.15 TLC analysis of hydrolysis products from mannooligosaccharides by CBM32-*Ct*Man5A$_{CM}$ (A) and *Ct*Man5A$_{CM}$ (B) and from ivory nut mannan by CBM32-*Ct*Man5A (C) and *Ct*Man5A (D). Each mannooligosaccharide in Panel A and B, identified by its degree of polymerization (100 µg, M2 to M6), was incubated with purified CBM32-*Ct*Man5A or *Ct*Man5A (0.18 unit each) for 16 h or 20 min, and the hydrolysates were analyzed by TLC. (C and D) Ivory nut mannan (50 µg each) was treated with CBM32-*Ct*Man5A or *Ct*Man5A (0.54 unit each) up to 24 h. Samples were taken at intervals, and the hydrolysates were analyzed by TLC. These data were acquired from the co-authors in (Mizutani et al. 2012).

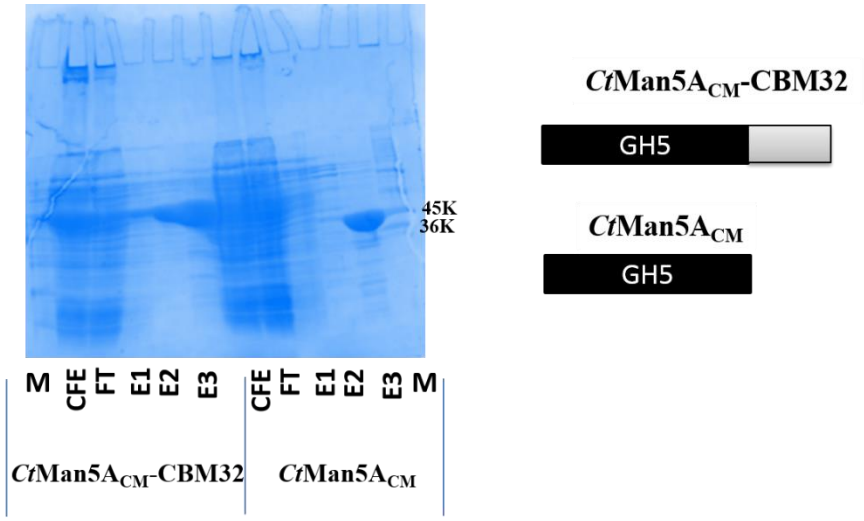


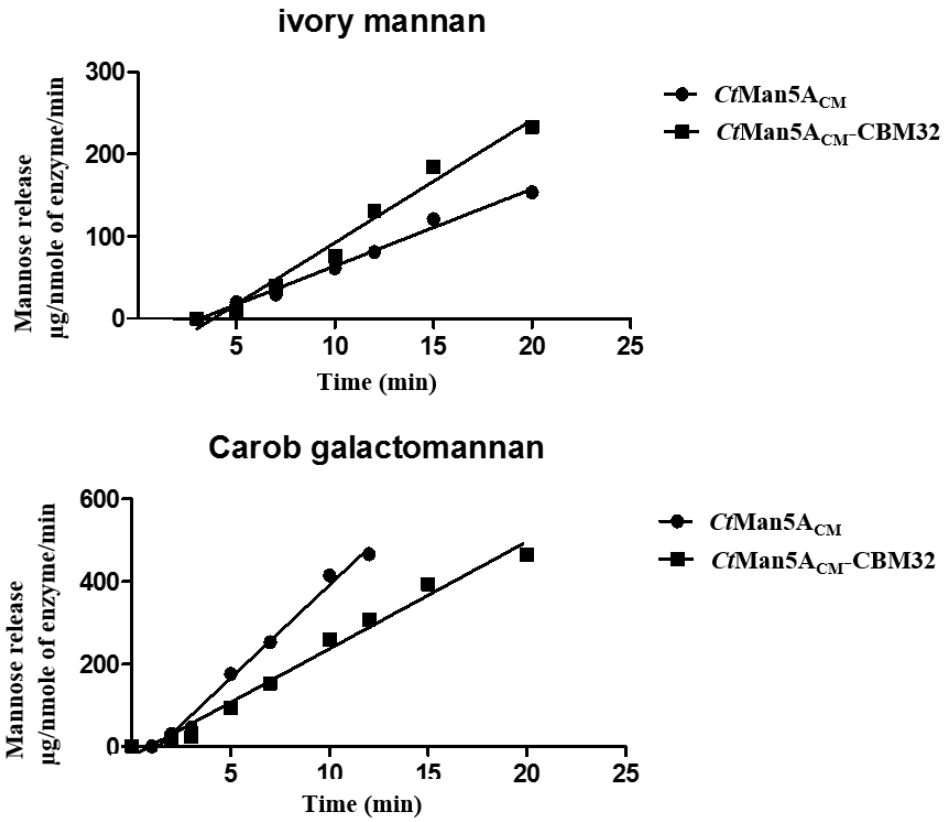
Figure 3.16 Recombinant enzymes subjected to SDS-PAGE. Schematic of the recombinant protein is on the right. The two proteins were purified by IMAC and the various fractions analyzed by SDS-PAGE using a 10% polyacrylamide gel. The lanes were labeled as follows: M, Marker with sizes in kDa; CFE, cell free extract; FT, flow through; E1-3, elutions by a stepwise imidazole gradient of 0 mM, 5 mM and 150 mM, respectively in Talon buffer. Predicted molecular weights (KDa) for *Ct*Man5A and *Ct*Man5A-CBM32 are 36K and 52K.

Figure 3.17 showed that the catalytic module of the mannanase *Ct*Man5A (designated *Ct*Man5A$_{CM}$) was more active against the soluble mannan high viscosity carob galactomannan than *Ct*Man5A$_{CM}$-CBM32. In contrast, *Ct*Man5A$_{CM}$-CBM32 was more efficient than *Ct*Man5A$_{CM}$ in digesting insoluble mannan such as ivory nut mannan or β-1,4-Mannan (**Figure 3.17** upper). It is possible that ivory nut mannan and beta-mannan contain more non-reducing ends than carob galactomannan, which are potential targets for the CBM32. Perhaps ivory nut mannan chains have a low (degree of polymerization), (Chanzy et al. 1984) while beta-mannan, generated in part by the digestion of carob galactomannan with mannanases, will have a lower d.p. than its parent polysaccharide. When *Ct*Man5A$_{CM}$-CBM32 is bound to polymers with a low degree of polymerization the enzyme can access most of the substrate as the majority of the mannosidic linkages will be available to the catalytic module when tethered to the non-reducing end through its CBM32. Thus the benefit of preventing transglycosylation is not offset by restricted enzyme access. The reduced activity of *Ct*Man5A$_{CM}$-CBM32 against soluble carob galactomannan, compared to *Ct*Man5A$_{CM}$, may reflect limited access to the complete

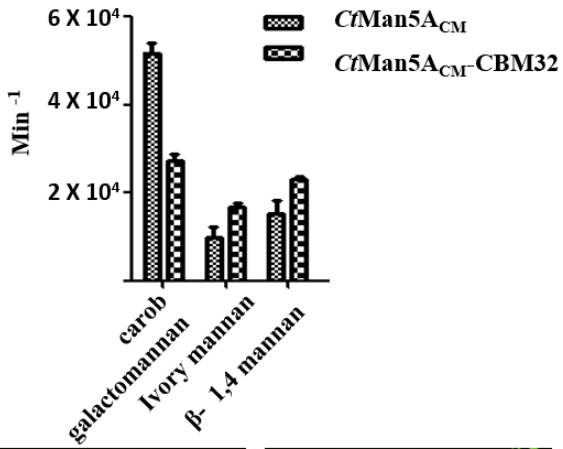
substrate by the *endo*-acting mannanase when it is tethered to the non-reducing end through its CBM32.

In situ, CBM32 contributed very little, if any, to the degradation of *Physcomitrella* mannan by *Ct*Man5A (**Figure 3.17** lower), which is probably caused by the different chemical environment between mannan in solution and within the complex architecture of plant cell walls, where the non-reducing end of the polysaccharide, targeted by the CBM32, may be in complex with other wall polymers and thus not accessible to the protein module. Another possible explanation for the observation is that the CBM would bind infrequently to cell walls where the mannan would have a very high degree of polymerization and thus the CBM32 would have little influence on the *endo* activity of the mannanase. It is possible that *Ct*Man5A is a component of the *C. thermocellum* cellulosome that hydrolyzes oligosaccharides or short polysaccharides released by other mannanases, and it is only when acting on these substrates, where transglycosylation is more likely due to the increased molar concentration of non-reducing ends, that the CBM32 contributes to enzyme function by preventing the futile reaction.

(A)


ivory mannan
Mannose release
µg/nmole of enzyme/min
300
200
100
0
5
10
15
20
25
Time (min)
CtMan5ACM
CtMan5ACM-CBM32
Carob galactomannan
Mannose release
µg/nmole of enzyme/min
600
400
200
0
5
10
15
20
25
Time (min)
CtMan5ACM
CtMan5ACM-CBM32

(B)


6 X 10⁴
4 X 10⁴
2 X 10⁴
Min⁻¹
CtMan5ACM
CtMan5ACM-CBM32
carob galactomannan
Ivory mannan
β- 1,4 mannan

(C)

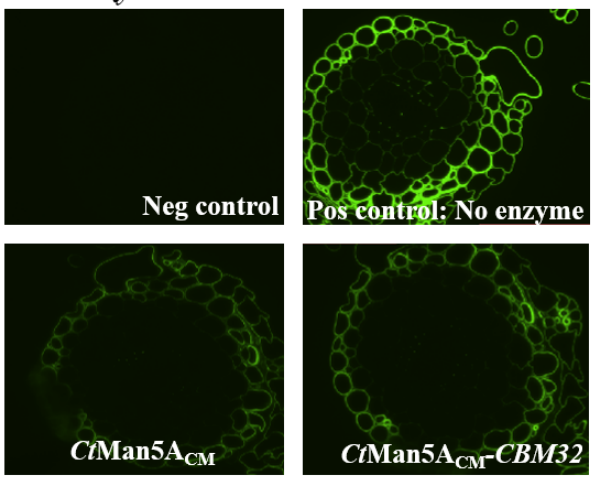

Neg control
Pos control: No enzyme
CtMan5ACM
CtMan5ACM-CBM32

Figure 3.17 Enzyme activity assay of *Ct*Man5A$_{CM}$-CBM32 and *Ct*Man5A$_{CM}$ *in vitro* by DNSA method (A and B) and *in situ* probed by CCRC-M170 in *Physcomitrella* (C). For the kinetic assay, reaction was carried out at 60 ^{o}C, in 50mM MES buffer pH 6.0; enzyme concentration 10 nM; substrate concentration 5 mg/ml. For fluorescent labelling, 0.5µM enzyme was used in a one-hour incubation.

3.3 Discussion

This work explores how different CBMs potentiate both endo-mannanases and esterases *in vivo*. In the preliminary phase, enzyme constructs containing various CBMs were generated and their activities against different mannans were evaluated. Mannanase activity *in situ* was evaluated using antibody probes to determine the amount of substrate that remained after enzyme treatment. In general, polysaccharide specific monoclonal antibodies (mAb) are excellent and powerful tools to study the structure of the plant cell wall due to their high affinity, but their exact epitopes are difficult to define compared to CBM-GFP probes, due to the lack of pure antibody in large quantities. Monoclonal antibodies typically bind to 4-6 sugars with high affinity (Pattathil et al. 2010) but the small culture volumes of the mAb producing cells prevents large scale protein purification suitable for studying detailed ligand specificity using isothermal titration calorimetry (ITC) and protein X-ray crystallography. Thus, out of the hundreds of antibodies that target plant cell walls, available in Professor Michael Hahn's lab in CCRC, few have had their epitopes extensively characterized (Pattathil et al. 2010, DeMartini et al. 2011, Steffan et al. 1995, Hahn et al. 1993). The CBM-GFP probes, however, can be overexpressed in *E. coli* making it easy to determine the precise ligands for these proteins through binding assays, typically by ITC and, ultimately, through visualization of the CBM in complex with its ligand(s) using X-ray crystallography. The CBM-based probes, however, generally display low affinity compared to monoclonal antibodies (DeMartini et al. 2011, Pattathil et al. 2010, Gilbert 2010, Gilbert 2003, Gilbert et al. 2008, Boraston et al. 2004), which can result in a weak binding pattern leading to false negatives. Thus, CBMs and monoclonal antibodies fulfill complementary roles when probing the structural architecture of

plant cell walls. For example, there are no reported CBMs that bind to the internal regions of pectins, while a repertoire of antibodies that interact with these glycans have been characterized (Pattathil et al. 2010, Blake et al. 2006, Herve et al. 2010, Knox et al. 1990, Willats et al. 2000, Jones, Seymour and Knox 1997). In contrast, while CBMs that bind specifically to crystalline cellulose have been described (CBM2a, CBM10 and CBM3a/b) (McLean et al. 2002, McCartney et al. 2004, Jing et al. 2009, Jindou et al. 2007, Blake et al. 2006), antibodies specific for this recalcitrant polysaccharide have not been generated.

The characterization of CCRC-M170 indicated that more work is needed to characterize the extensive array of plant cell wall monoclonal antibodies available at CCRC (Complex Carbohydrate Research Center, USA). The fact that CCRC-M170 binds only to acetylated mannan makes it very unique and was invaluable in characterizing the activity of mannan esterases *in situ*. Significantly, the data showed that the acetyl groups on mannan are unstable once the cell wall sections have been exposed to air for a few weeks, even after they have been pre-embedded in resin. Thus, standard protocols need to be revised when dealing with acetylated polysaccharides and researchers working with CCRC-M170 and related antibodies are encouraged to use slides as fresh as possible.

The data presented here show that CBMs potentiate mannanase activity against plant cell walls of advanced and non-vascular plants. The results are consistent with the hypothesis, proposed by Hervė et al. (2010) that CBMs promote the enzymatic deconstruction of intact plant cell walls by targeting and proximity effects. An interesting feature of the data is the differential activity of the GH5 and GH26 mannanases against plant cell walls, and how CBMs modulate the enzymatic degradation of these polymers. For example, from the data presented in **Figure 3.10** it is evident that *Cj*Man5A$_{CM}$ can access, equally, all the mannan in the tobacco cell wall,

while *Cj*Man26A$_{CM}$ can only hydrolyse ~50% of the polysaccharide in these cell walls. This may reflect differences in the specificity of the enzymes. *Cj*Man5A can accommodate glucose in subsites distal to the active site, while *Cj*Man26A displays a strict requirement for mannose at all negative subsites (Tailford et al. 2009). The more relaxed specificity of *Cj*Man5A enables the enzyme to access regions of glucomannan where there is a particularly high ratio of Glc:Man. In contrast *Cj*Man26A is unlikely to efficiently degrade these regions of glucomannan. It is evident that additional GH26 and GH5 enzymes need to be evaluated against tobacco cell walls to explore whether the different activities against this substrate, observed here for the two *Cellvibrio. japonicus* enzymes, is limited to these two mannanases, or is a general feature of glycoside hydrolases from these two families.

The observation that a cellulose binding CBM had the largest impact on the activity of *Cj*Man5A against tobacco walls could be explained by the close association of mannan with cellulose, especially in the epidermis/cortical tissues. In this plant, cellulose is more abundant than mannan (up to 45% of tobacco callus cell walls comprises cellulose) and the ratio of cellulose to mannan can reach 7 to 1. (Albersheim 1975, McNeil et al. 1984). Thus cellulose represents a more abundant receptor for CBM3a compared to the ligand targeted by the mannan binding CBMs. It is also possible that binding to a cellulose allows the CBM to slide freely along the cellulose microfibrils and thus maximize substrate access without dissociating from the cell wall (Tomme et al. 1998). This sliding motion is afforded by predominantly parallel hydrophobic contact between the CBM and the pyranose rings in cellulose (Blake et al. 2006, Lehtio et al. 2003b, Liu et al. 2011, Boraston et al. 2004). In contrast, polar contacts dominate the interaction of mannan binding CBMs and the polysaccharide and thus to access different regions of the cell wall requires dissociation and the reassociation reducing the amount of enzyme in close contact with substrate.

Fusing either CBM3a or CBM27 to *Cj*Man26A$_{CM}$ makes all the mannan in tobacco walls accessible to the enzyme (**Figure 3.11**). It is possible that in the structures targeted by the CBMs, mannan and cellulose are in intimate contact enabling both CBMs (CBM3a and CBM27 bind cellulose and mannan, respectively) to interact with these regions of the wall. I propose that the mannan in these structures contains a high concentration of glucomannan which is generally a poor substrate for *Cj*Man26A. However, when held in close proximity with glucomannan, through either the CBM3a or CBM27, the substrate binding cleft can access the glucomannan chains and mediate glycosidic bond cleavage. The rationale for why CBM3a is more effective than CBM27 in potentiating the activity of *Cj*Man26A and *Cj*Man5A may be similar, and is discussed above. However, it is also possible that CBM27 is less effective than CBM3a in accessing the recalcitrant glucomannans because it only binds to limited regions of the polysaccharide, where there are consecutive mannose residues. Maybe the CBM27 is not so effective as it binds less frequently to the wall and may not be able to efficiently target glucomannan. It would be very useful to explore this hypothesis using a glucomannan specific probe but, currently, no probe that specifically targets the polysaccharide has been described. Professor Hahn's antibody collection at CCRC may contain candidates, but further work is required.

Mannan digestion in lower non-vascular plants exemplified by *Physcomitrella,* as shown in **Figure 3.13**, indicated that neither *Cj*Man5A nor *Cj*Man26A can access all the mannan in moss, even when fused to the cellulose binding CBM, CBM3a. CBM27, however, does enable *Cj*Man5A to access all the mannan in *Physcomitrella*, but this requires a very high concentration of enzyme. It is possible that the mannan in *Physcomitrella* is heavily acetylated, and that the acetyl groups may create a steric block that restricts the ability of the two

mannanases from attacking the polysaccharide. This hypothesis could be tested by evaluating the action of the two mannanases against *Physcomitrella* walls that had been deacetylated by the esterase *Cj*CE2C (this argument does not apply to tobacco cells, which was deacetylated prior to probing mannanase action). This hypothesis is consistent with the relaxed specificity displayed by the distal subsites of *Cj*Man5A, which may enable the enzyme to access heavily acetylated mannan, providing the enzyme is held in close proximity with the polysaccharide through the CBM27. Indeed, the observation that CCRC-M170 binds to *Physcomitrella* walls (**Figure 3.6**) demonstrates the presence of acetylated mannan. It is also possible that the "inaccessible" mannan in *Physcomitrella* walls is glucomannan that has a particularly high Glc:Man ratio and is thus not accessible to *Cj*Man26A, and is a very poor substrate for *Cj*Man5A. Only when fused to CBM27 can the GH5 enzyme access this glucose-rich glucomannan. This may be explained by the abundance of mannan and the lack of cellulose in the *Physcomitrella* wall (Zhang et al. 2014, Wise, Saxena and Brown 2011, Fu, Yadav and Nothnagel 2007, Liepman et al. 2007), which enabled the CBM27 to target the enzyme to more extensive regions of the substrate.

Because in this project a number of CBMs have been fused to the same catalytic module and compared *in situ*, our research can be used to determine how specificity of a CBM influences its capacity to potentiate the activity of enzymes against cell walls. The data presented here indicate that CBMs, which bind to the most abundant polysaccharide in the wall, or the polymer most closely associated with the substrate of the enzyme, are likely to maximize enzymatic activity. Thus, in the tobacco cell wall cellulose is the most abundant polysaccharide and is in close association with mannan, explaining why cellulose binding CBM3a is more effective at enhancing mannanase activity than a mannan-specific CBM. It is apparent, however, that there is no universal formula instructing us about the best CBM or CBM combination that optimizes

enzyme activity, as CBM selection also depends on the topology and specificity of the target enzyme and the context of the substrate within the plant cell wall. Overall the data reflect the complexity of the intact cell wall and the difficulty of enzymatic digestion *in situ*. This study demonstrates that to achieve complete digestion of the mannan component of plant cell walls *in situ*, an appropriate CBM is required, linked to a highly active mannanase with a topology and substrate specificity that is optimized for the target cell wall. Indeed, as exemplified by *Cellvibrio. japonicus* (DeBoy et al. 2008), mannan degrading bacteria produce multiple mannanases, often with different combinations of CBMs, which may reflect the requirement for different enzyme combinations to achieve optimal degradation of diverse cell walls.

Although, previously, modular carbohydrate esterases with CBMs were found in nature (Kroon et al. 2000, Gordillo et al. 2006, Ferreira et al. 1993, Montanier et al. 2009b), the capacity of these CBMs to potentiate esterase activity has not been reported to our knowledge, especially *in vivo*. Data presented here show that CBMs modulate the activity of carbohydrate esterases against plant cell walls *in vivo*. In **Figure 3.14** the activity of *Cj*CE2C was potentiated by three different CBMs, with CBM27 being the most effective against *Physcomitrella*. It appears that the CBMs enhanced esterase activity through proximity effects, similar to the function of these protein modules in multi-modular glycoside hydrolases. It should be noted, however, that some CBMs, exemplified by CBM2b-1-2, which targets xylan, can have a negative impact on esterase activity (McCartney et al. 2006) (data acquired by Dr Artur Rogowski). One explanation is that the CBM binds to non-acetylated regions (Herve et al. 2010) where the substrate for the catalytic module is scarce, and therefore guides the enzyme away from its real substrate. It suggests that the overall CBM-mediated potentiation depends on the components and context of the cell wall targeted by the enzymes.

A striking feature of the esterase component of this study is the observation that while *Cj*CE2C can deacetylate *Physcomitrella* mannan *in situ*, the other three CE2 esterases characterized to date, *Ct*CE2, *Cj*CE2A and *Cj*CE2B, did not attack the intact plant cell wall. This is somewhat surprising as all four CE2 were active against acetylated xylan, both purified and within intact tobacco cell walls, and acetylated mannan. The 3D structures of *Ct*CE2, *Cj*CE2A and *Cj*CE2B have been solved (**Figure 3.18**) and made available on the protein database (http://www.rcsb.org) (Berman et al. 2000, Montanier et al. 2009a). All three 3D structures reveal a bi-domain enzyme in which an N-terminal β-sheet "jelly roll" domain (around 130 residues) is linked to a C-terminal domain of approximately 220 residues. Superimposition of *Ct*CE2 with the two *Cellvibrio* esterases reveals that both the N-terminal β-sheet domain and the C-terminal catalytic domain are structurally conserved (Montanier et al. 2009a). Due to the lack of the crystal structure of *Cj*CE2C, it is not clear why only *Cj*CE2C removes acetyl group from *Physcomitrella* mannan. Alignment (**Figure 3.19**) shows that both distal aromatics are absent in *Cj*CE2C and there is likely to be a lot of perturbation in the substrate binding cleft as there are deletions in the vicinity of where the aromatics should be, suggesting a very different substrate binding cleft to the other enzymes.

Speculation is that the lack of one distal aromatic residues indicated by the alignment of these esterase sequence (**Figure 3.19**) results in a much more open substrate binding cleft, making *Cj*CE2C accessible to acetylated mannans within the *Physcomitrella* cell walls (Zhang et al. 2014, Montanier et al. 2009a). It is speculated that this is the enzyme that attacks cell walls directly while the other *Cellvibrio. japonicus* esterases go for acetylated oligosaccharides. More studies are required on this.

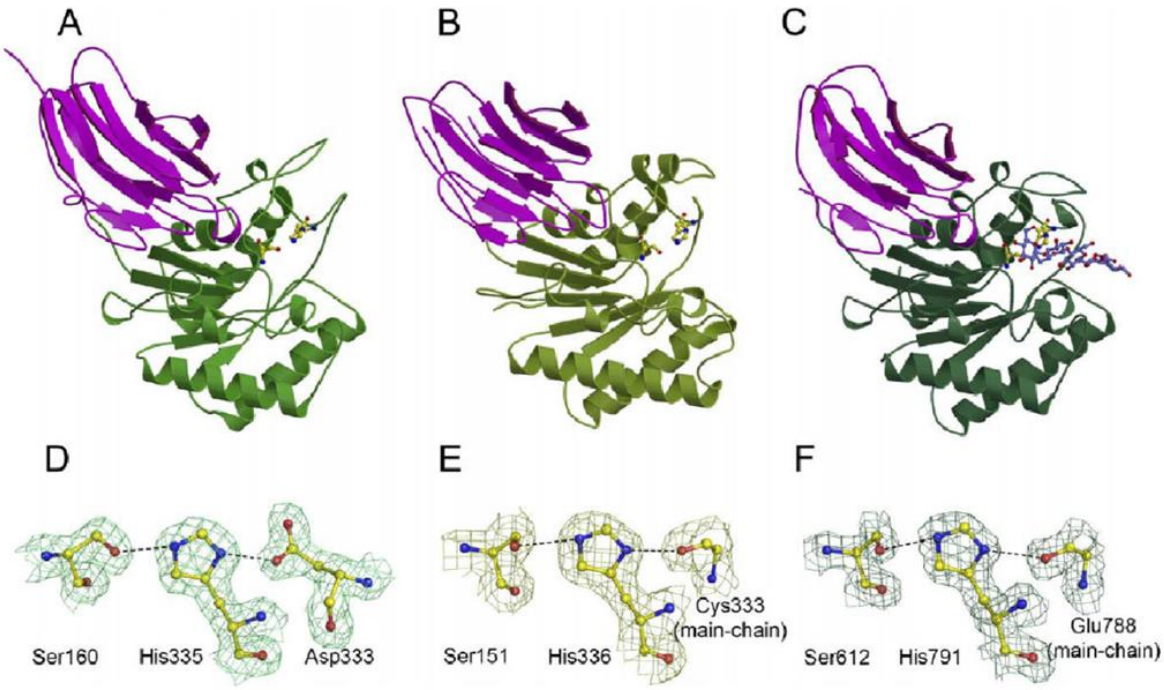


Figure 3.18 Crystal structure of the carbohydrate esterases of family 2.
(A) 3-D structure of *Cj*CE2B with the catalytic domain in green and the β-sheet domain in magenta. The catalytic Ser and His are shown in ball-and-stick representation.
(B) *Cj*CE2A, drawn as above.
(C) *Ct*CE2 as above and with cellopentaose in blue.
(D) The catalytic Ser-His-Asp triad of *Cj*CE2A.
(E) The catalytic Ser-His dyad with the main-chain carbonyl interaction from Cys333 of *Cj*CE2B.
(F) The catalytic dyad and main-chain carbonyl of *Ct*CE2.
This figure was is cited from (Montanier et al. 2009a)

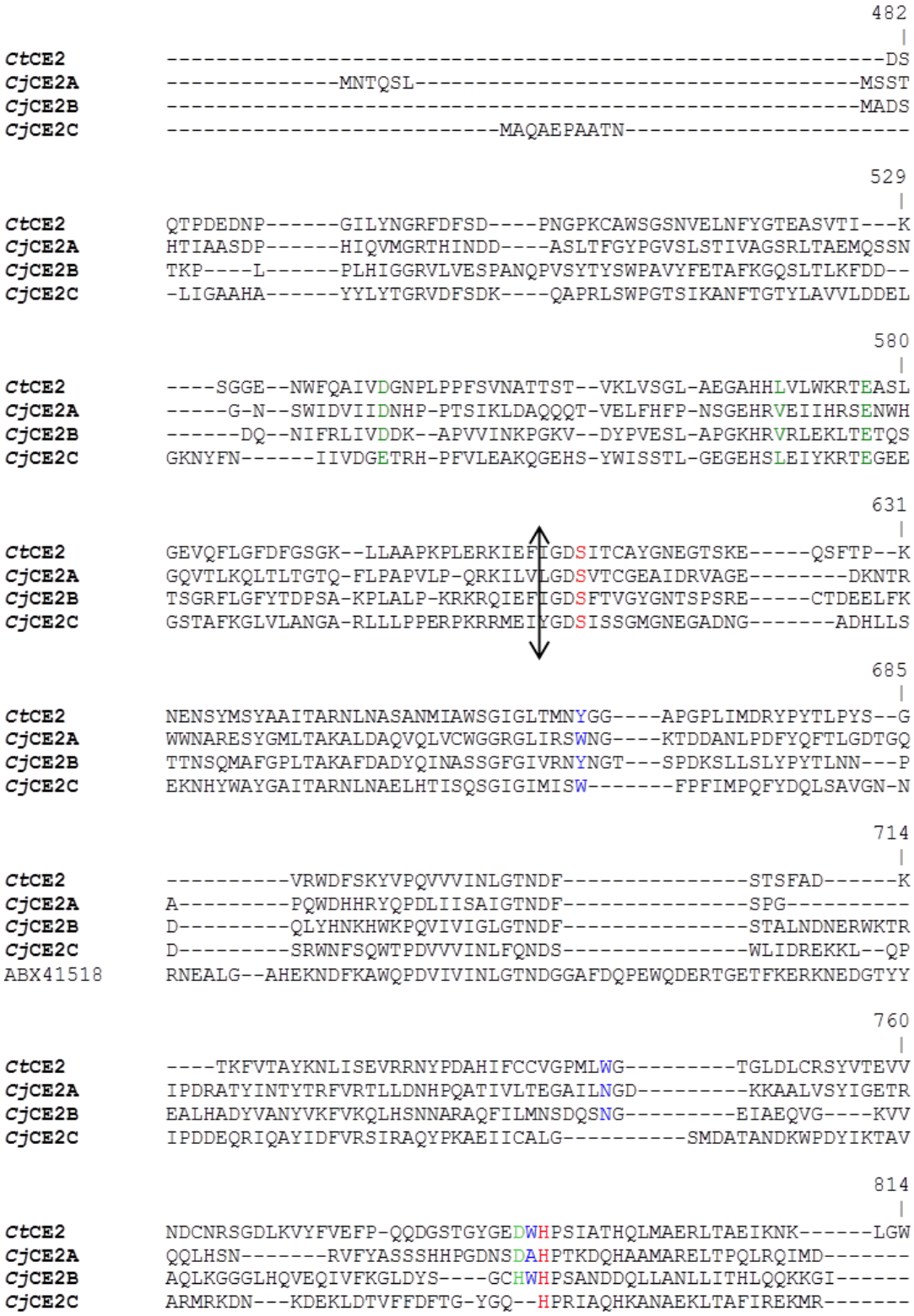


Figure 3.19 Alignment of CE2 Enzymes. The catalytic histidine and serine are in red, the residue equivalent to the aspartate that completes the catalytic triad is green, while the equivalent residues to the three aromatic amino acids in *Ct*CE2 that binds to cellulose are blue. The vertical arrow marks the division between the β-jelly roll fold and the α/β-hydrolase catalytic domain. The residues are labeled according to the *Ct*CE2 sequence. (Montanier et al. 2009a).

3.4 Future work

An interesting future experiment would be to attach multiple CBMs that display different specificities to one or multiple enzymes (including glycoside hydrolases, pectic lyases and esterases), to create an artificial cellulosomic machinery. One reason for doing that is to study how multiple CBM combinations are able to modulate the activity of enzymes that target plant cell walls. It would be particularly interesting to explore whether a novel artificial cellulosome, containing CBMs from organisms that do not synthesize these enzyme complexes, can enhance the activity of both an artificially engineered cellulosome and its natural counterpart, against a variety of recalcitrant substrates. It would be interesting to investigate how many CBMs are required minimally to fully digest specific cell walls, which may provide an evolutionary rationale for the presence of these protein modules in specific enzymes.

In addition to cellulose, pectin is also closely associated with mannan, evidenced by the requirement to remove the galacturonic acid polymer to expose mannans to CBMs or enzymatic attack. An interesting question to be asked, therefore, is whether a pectin specific CBM may be a candidate for enhancing the activity of *Cj*Man5A and *Cj*Man26A against tobacco cell walls *in situ*. Since a pectin binding CBM was recently identified by Anna Sofia Luis project (working in Professor Harry Gilbert's lab in University of Newcastle upon Tyne), this experiment is now tractable.

Chapter 4. Enzymatic degradation and utilization of Rhamnogalacturonan I (RGI) by *Bacteroides thetaiotaomicron*

4.1 Introduction

The human diet contains a high proportion of complex carbohydrates that are primarily plant storage polysaccharides, such as starch and fructans, and cell wall structural polysaccharides. Apart from accessible forms of starch, these complex polysaccharides are not depolymerized by human digestive enzymes and thus pass into the large bowel, where symbiotic bacteria degrade and utilize these glycans, generating short chain fatty acids as the major fermentation products. (Hooper, Midtvedt and Gordon 2002, Martens et al. 2011) Pectins, galacturonic acid containing plant glycans, are a major component of the human diet. These polysaccharides are present in many foods, particularly fruit, and are also widely used in the food industry as gelling agents; it is estimated that in industrialized countries 5 g of pectins are consumed per capita per day (Ridley et al. 2001, Thakur, Singh and Handa 1997). Pectins are thus an important growth substrate for the human large bowel microbial community (microbiota). Indeed, these glycans are used as prebiotics that stimulate the microbiota as well as, potentially, benefiting humans by reducing blood cholesterol levels and chelating lead and mercury (Ridley et al. 2001, Thakur et al. 1997).

Physiological and genomic data show that glycan utilization is a key evolutionary driver of the microbiota (Arumugam et al. 2011, Backhed et al. 2005). Studies of cultivated microbiota species indicate that the *Bacteroidetes*, composed largely of members of the genus *Bacteroides*, exhibit the capacity to metabolize a variety of plant and animal-derived glycans (Martens et al. 2011, Salyers et al. 1977, Martens et al. 2008, Koropatkin et al. 2008). *Bacteroides thetaiotaomicron*, a Gram-negative obligate anaerobe, is a prominent member of the normal

human (and murine) microbiota. The genome of *B. thetaiotaomicron* is predicted to encode 256 glycoside hydrolases, 16 polysaccharide lyases and 20 carbohydrate esterases distributed in which are distributed into 43, 8 and 9 families in the CAZy database. The genes encoding the enzymes, binding proteins (defined as SusD-like proteins), transporters (defined as SusC like proteins that mediate transport across the outer membrane) and sensors (generally hybrid two component systems defines as HTCS) that orchestrate the degradation of specific glycans are physically linked in the genome, and are defined as Polysaccharide Utilization Loci (PULs), **Figure 4.1**. (Martens et al. 2008, Koropatkin et al. 2008, Martens et al. 2011). The glycan(s) targeted by a PUL activates transcription of the locus by binding to the relevant HTCS. In total genome of *B. thetaiotaomicron* contains ~100 PULs.

B. thetaiotaomicron is capable of hydrolyzing a range of plant and host glycans including the highly branched pectin rhamnogalaturonan I (RGI) (Martens et al. 2011). The backbone of RGI, comprising galacturonic acid (GalA) and rhamnose (Rha), is decorated with extensive side chains referred to as arabinan and galactan, **Figure 4.2**. Arabinan consists of a backbone of α-1,5-linked L-arabinofuranose (Ara) residues found mainly in apple juice and sugar beet (Verhertbruggen et al. 2009). Galactan comprises linear chains β-1,4-linked galactopyranose (Gal) units. Current and previous work in Professor Harry Gilbert's lab in university of Newcastle upon Tyne is exploring the biochemical and genetic basis for arabinan and galactan degradation by *B. thetaiotaomicron*. The focus of this chapter is to determine mechanism by which the Rha and GalA containing backbone of RGI is depolymerized. Previous studies showed that RGI activates the transcription of a large PUL, defined as PUL-RGI extending from the genes *bt4145* to *bt4183*, **Figure 4.2**. To understand the mechanism by which this PUL orchestrates the degradation of RGI, the enzymes encoded by this locus were produced in

recombinant form and their biochemical properties characterized. The data enabled a model for RGI degradation to be generated.

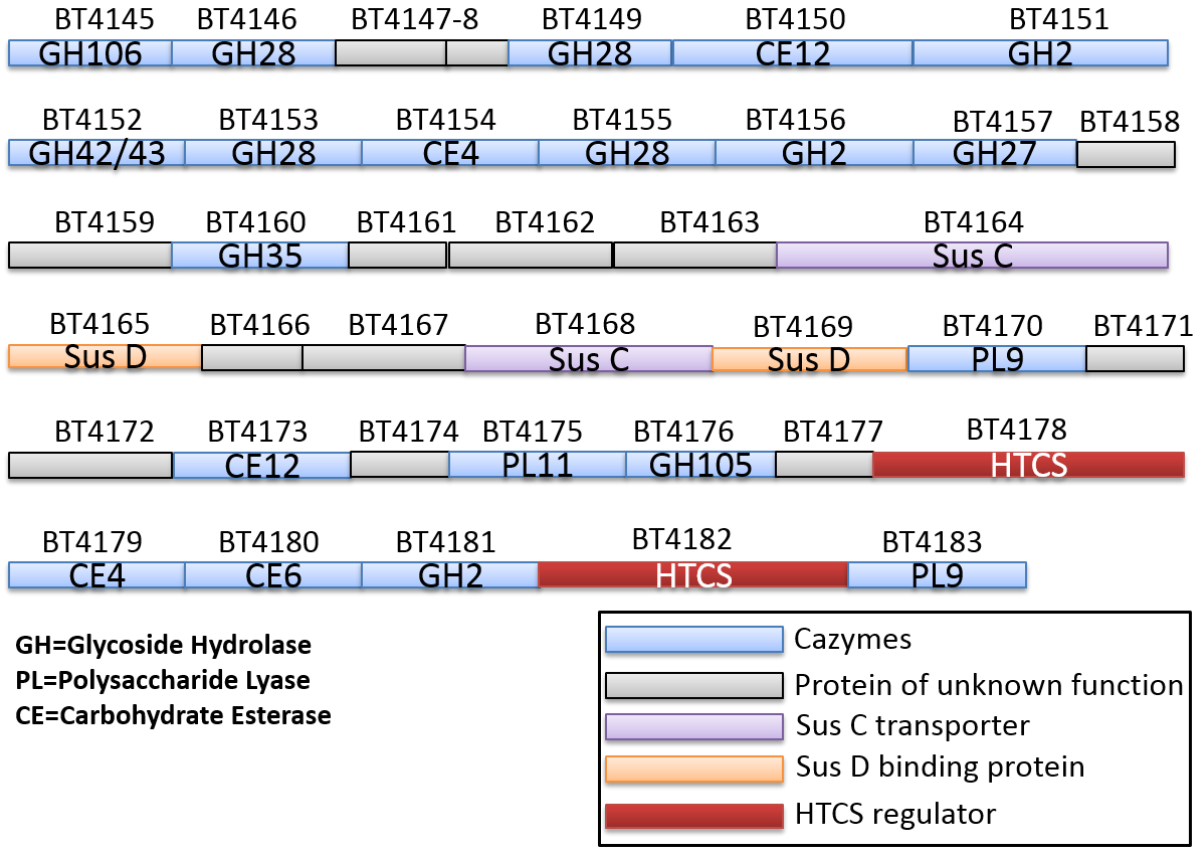


Figure 4.1 Map of PUL-RGI from *B. thetaiotamicron*. The genes encoding CAZymes are annotated by the family in which the cognate protein is located.

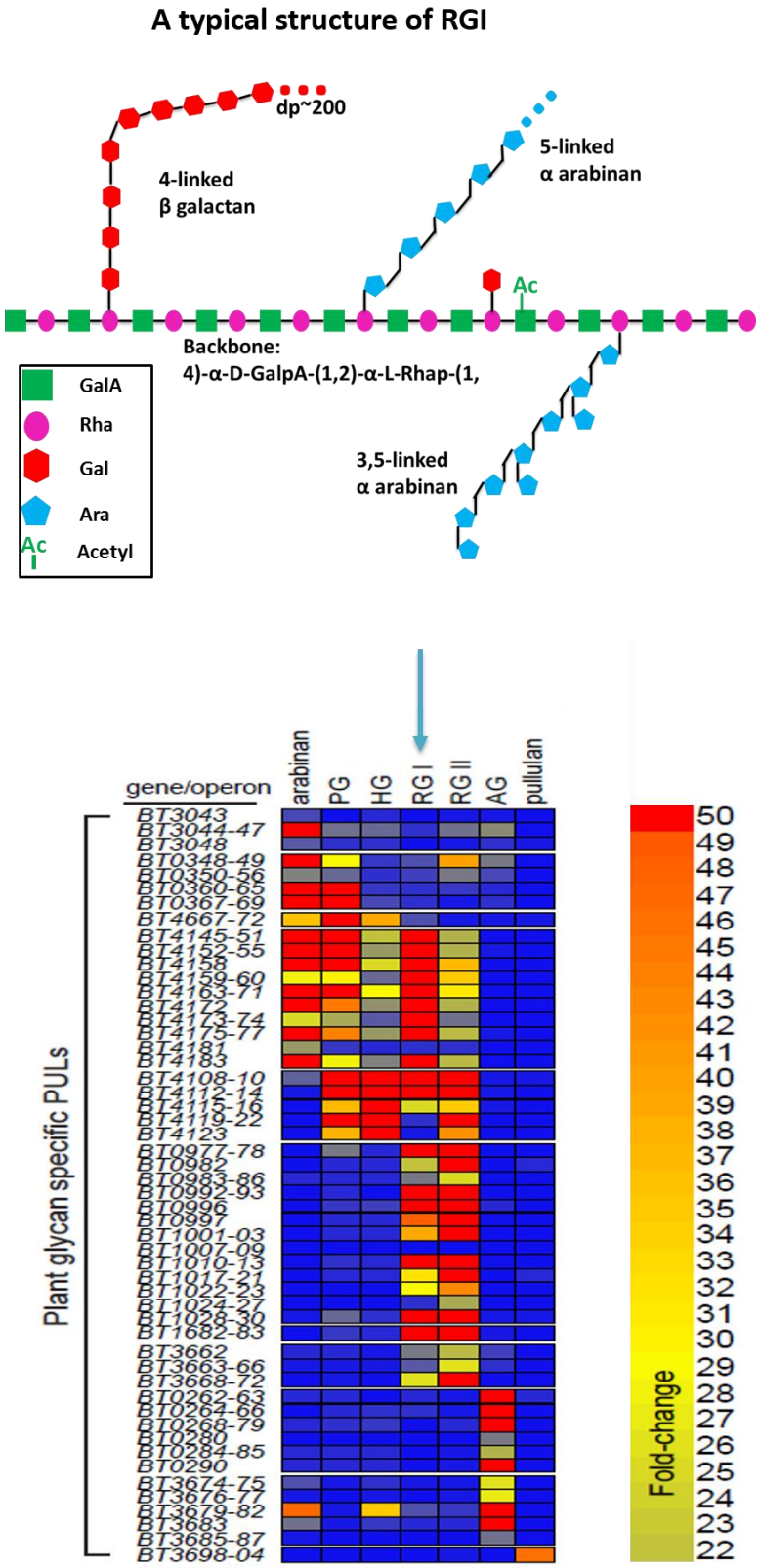


Figure 4.2 A schematic of RGI structure and the transcriptional data when *B. thetaiotaomicron* was grown on different plant glycans (Martens et al. 2011). PUL-RGI corresponds to *bt4145-bt4183.*

4.2 Results

4.2.1. Cloning of the genes in the RG I utilization locus.

In total 31 out of the 36 genes in PUL-RGI were cloned by NZYTech (1649-038 Lisboa, Portugal) into the *E. coli* expression vector pET28a from Novagen, as detailed in **Table 4.1**. In total 21 genes encode proteins in glycoside hydrolase, polysaccharide lyase or carbohydrate esterase families (GHs, PLs and CEs, respectively) annotated on the CAZy database (http://www.cazy.org/), (Henrissat 1991, Cantarel et al. 2009b), four genes encode two SusC/D pairs, two genes encode hybrid two component systems (HTCSs) and the remaining 9 genes encode secreted proteins (contain signal peptides) of unknown function.

Name	*CAZy* family	Signal peptides	Predicted molecular weight (kDa)	Experimentally determined size (kDa)
BT4145	**GH106**	SPI	102	100
BT4146	**GH28**	SPI and SPII	49	50
BT4149	**GH28**	SPI and SPII	36	Not cloned
BT4150	**CE12**	SPI	44	46
BT4151	**GH2**	SPI	106	100
BT4152	**GH42/43**	SPI	111	100
BT4153	**GH28**	SPI	47	47
BT4154	**CE4**	SPI	35	Insoluble
BT4155	**GH28**	SPI	49	49
BT4156	**GH2**	SPI	106	100
BT4157	**GH27**	SPI	69	Insoluble
BT4160	**GH35**	SPI	76	70
BT4170	**PL9**	SPI	55	55
BT4173	**CE12**	SPI and SPII	58	55
BT4174	**GH105**	SPI	42	42
BT4175	**PL11**	SPI	66	60
BT4176	**GH105**	SPI and SPII	50	50
BT4179	**CE4**	SPI	65	55
BT4180	**CE6**	SPI	27	27
BT4181	**GH2**	SPI	110	102
BT4183	**PL9**	SPI and SPII	46	45
BT4147	Uncategorized	SPI	20	20
BT4158	Uncategorized	SPI	24	24
BT4159	Uncategorized	SPI	61	Insoluble
BT4163	Uncategorized	SPI	84	66
BT4164	**SusC**	SPII	125	Not cloned
BT4165	**SusD**	SPII	58	60
BT4166	Uncategorized	SPII	62	62
BT4167	Uncategorized	SPI and SPII	64	60
BT4168	**SusC**	SPII	120	Not cloned
BT4169	**SusD**	SPII	64	62
BT4171	Uncategorized	SPII	22	22
BT4172	Uncategorized	SPI	123	Insoluble
BT4177	Uncategorized	SPI	12	Not cloned
BT4178	**HTCS**	SPI	162	70
BT4182	**HTCS**	SPI	164	Not cloned

Table 4.1 Genes in PUL-RGI.

All genes have been cloned into the vector pET28 vector unless labeled as "Not cloned". Upper part contains genes predicted to encode enzymes located in families within the CAZy database. The lower section contains genes that encode proteins, which are not components of the CAZy database. Prediction of lipoproteins and signal peptides were performed by LipoP 1.0 Server (www.cbs.dtu.dk/services). SpI: signal peptide (signal peptidase I). SpII: lipoprotein signal peptide (signal peptidase II) (Juncker et al. 2003). SusC and SusD are putatively transporters and binding proteins, respectively, while HTCSs are Hybrid Two-Component Systems that

potentially bind to the signaling molecule to up-regulate the PUL. “Uncategorized” means the encoded protein has no known function. Only the sensor part of the gene *bt4178* was cloned (protein size, 70 kDa).

4.2.2Enzyme expression of the RGI utilization Locus

To characterize PUL-RGI the encoded proteins were produced in recombinant form in *E. coli* and purified by immobilized metal ion affinity chromatography (IMAC) as they all contained a His_6-tag provided by the vector. In total 27 of the proteins were produced in soluble with the exception of BT4154 (CE4), BT4157 (GH27), BT4159 and BT4172 (uncategorized by CAZy), which either could not be produced in *E. coli* or formed inclusion bodies. Examples of protein purifications are shown in **Figure 4.3**, while the remaining protein purifications are displayed in Appendix A6. The sizes of the purified proteins, **Table 4.1**, were consistent with their predicted molecular weights based on amino acid sequence.

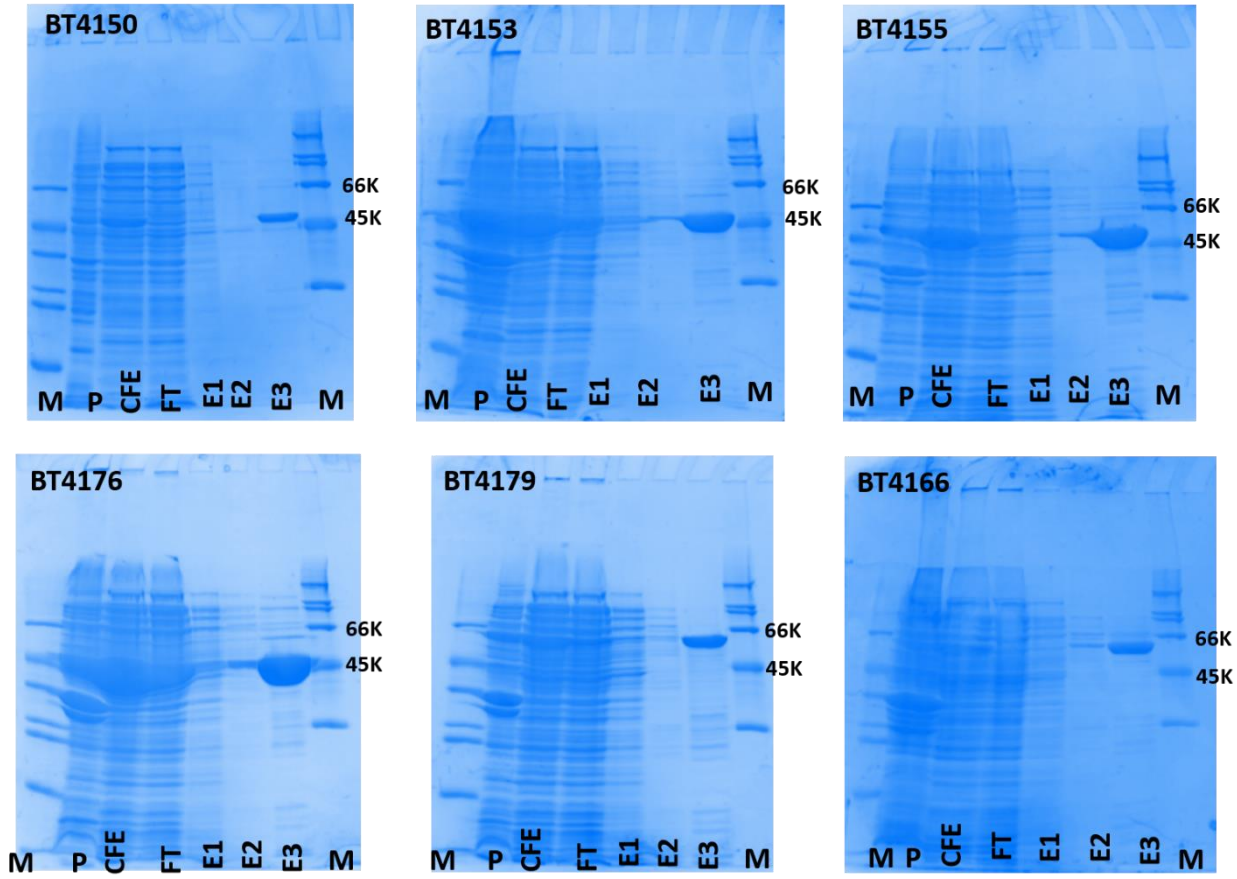


Figure 4.3 Example of IMAC purified recombinant enzymes subjected to SDS-PAGE. Recombinant expression was in *E. coli* Tuner cells under conditions detailed in Chapter 2.2.; M, Marker; P, insoluble pellet; CFE, Cell free extract; E1-3, Elutions by imidazole gradient of 0 mM, 5 mM and 150 mM, respectively, in Talon buffer (10 mM Tris-HCl pH 8.0 containing 300 mM NaCl).

4.2.3 Polysaccharide and oligosaccharides used in this component of the thesis

To explore RGI degradation the catalytic activity of the proteins (that could be generated in soluble form in *E. coli*) encoded by PUL-RGI were initially evaluated against RGI from *Arabidopsis* mucilage (RGI-AM), prepared as described in Chapter 2.4. This form of RGI was selected as it contains no side chains so the enzymatic depolymerization of the backbone is not complicated by the presence of arabinose or galactose decorations, which may sterically hinder enzyme function. Subsequently, selected enzymes were evaluated against potato RGI from Megazyme, defined henceforth as RGI-PG, which contains short galactose side chains, and

also against sugar beet arabinan in which the RGI backbone is decorated with arabinose side chains.

Oligosaccharides that contain an anhydrous GalA on the non-reducing ends, **Figure 4.4**, were produced by treating RGI-AM with the polysaccharide lyase BT4170 and the products purified by size exclusion chromatography as described in Chapter 2.4. The identity of the smallest oligosaccharide ΔGalA-Rha (defined as di henceforth) was confirmed by incubation with the 4,5Δ-galacturonase BT4176 that generated only rhamnose (Rha), **Figure 4.4**. Larger oligosaccharides were shown by mass spectrometry (**Figure 4.5**) to comprise the tetrasaccharide 4.5ΔGalA-Rha-GalA-Rha (defined as tetra) or a mixture of tetra, 4,5ΔGalA-Rha-GalA-Rha-GalA-Rha (hexa) and 4,5ΔGalA-Rha-GalA-Rha-GalA-Rha-GalA-Rha (octa).

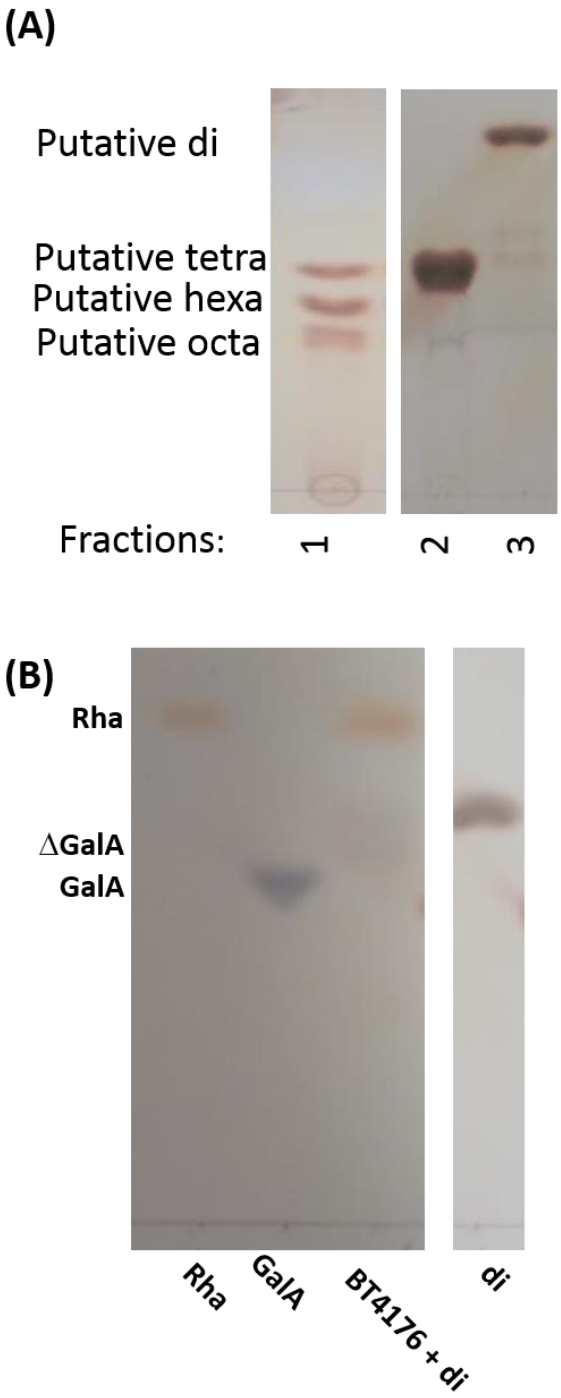


Figure 4.4 The oligosaccharides derived from lyase treated RGI backbone on TLC and evidence that the putative di is ΔGalA-Rha. (A) Putative oligosaccharides correspond to 4,5ΔGalA-Rha (di), 4,5ΔGalA-Rha-GalA-Rha (tetra), 4,5ΔGalA-Rha-GalA-Rha-GalA-Rha (hexa) and 4,5ΔGalA-Rha-GalA-Rha-GalA-Rha-GalA-Rha (octa). Detailed methodology is described in Chapter 2.4; (B) TLC of BT4176 (GH105) releasing Rha from di. Enzyme concentration was 1 μM, substrate concentration 4mg/ml and the reaction was carried out overnight in 20 mM Tris-HCl buffer, pH 7.0.

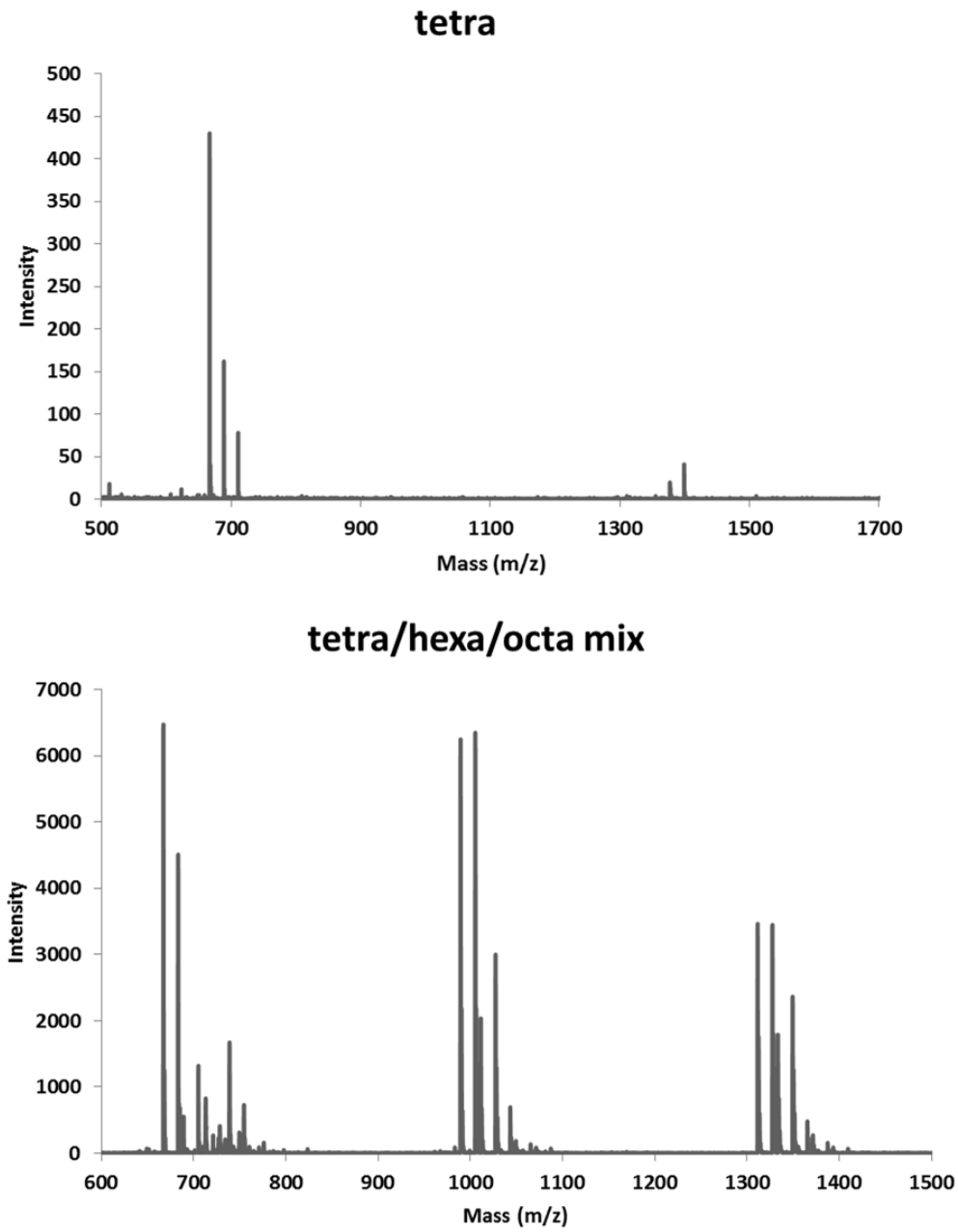
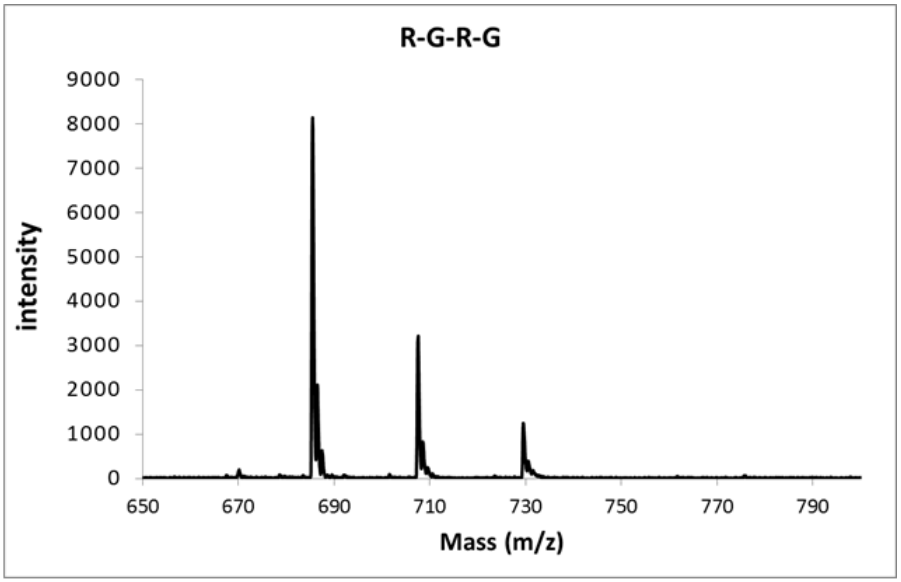


Figure 4.5 Mass spectrometric confirmation of the molecular weights of the tetra, hexa and octa oligosaccharides. 4.5ΔGalA-Rha-GalA-Rha (tetra), about 644 g/mol; 4,5ΔGalA-Rha-GalA-Rha-GalA-Rha (hexa), about 966 g/mol; 4,5ΔGalA-Rha-GalA-Rha-GalA-Rha-GalA-Rha (octa), about 1288 g/mol. The calculation is made based on the molar mass of Rha, GalA and ΔGalA of 164.16 g/mol, 194.14 g/mol and 176.14 g/mol, respectively. The molecular weight is right when considering the mass of sodium that flies with the oligos.

Oligosaccharides with a Rha at the non-reducing end was obtained by partially digesting RGI-AM with the GH28 endo-galacturonase (U29SH) from Novozymes. The products Rha-GalA-Rha-GalA-Rha-GalA, Rha-GalA-Rha-GalA and Rha-GalA were purified by P2 size exclusion chromatography and confirmed by mass spectrometry (details in Chapter 2.4), **Figure 4.6**.



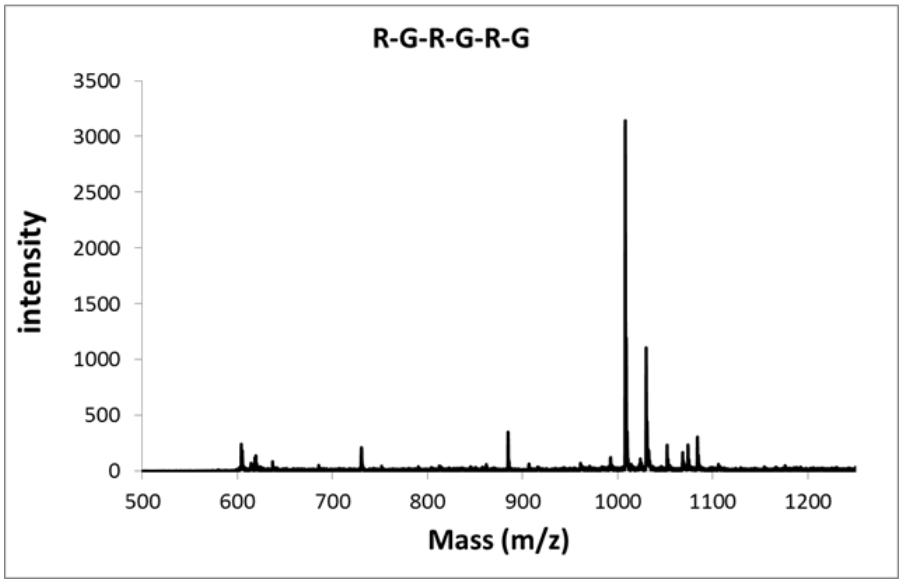


Figure 4.6 Mass spectrometric confirmation of the molecular weights of the oligosaccharides with a Rha at the non-reducing end. R, rhamnose; G, galacturonic acid; R-G-R-G (about 662 g/mol); R-G-R-G-R-G (about 984 g/mol); the molecular weight is right when considering the mass of a sodium that flies with the oligos.

4.2.4 The pectic lyases (PL9 and PL11) in the RGI PUL

Initially the three enzymes that are predicted to be polysaccharide lyases, based on their location in families PL9 and PL11, were evaluated against RGI-AM using TLC to detect reaction products. All three enzymes, BT4170 (PL9), BT4183 (PL9) and BT4175 (PL11), were active against RGI-AM displaying an endo mode of action evidenced by the release of different oligosaccharides rather than a single reaction product, (**Figure 4.7**).

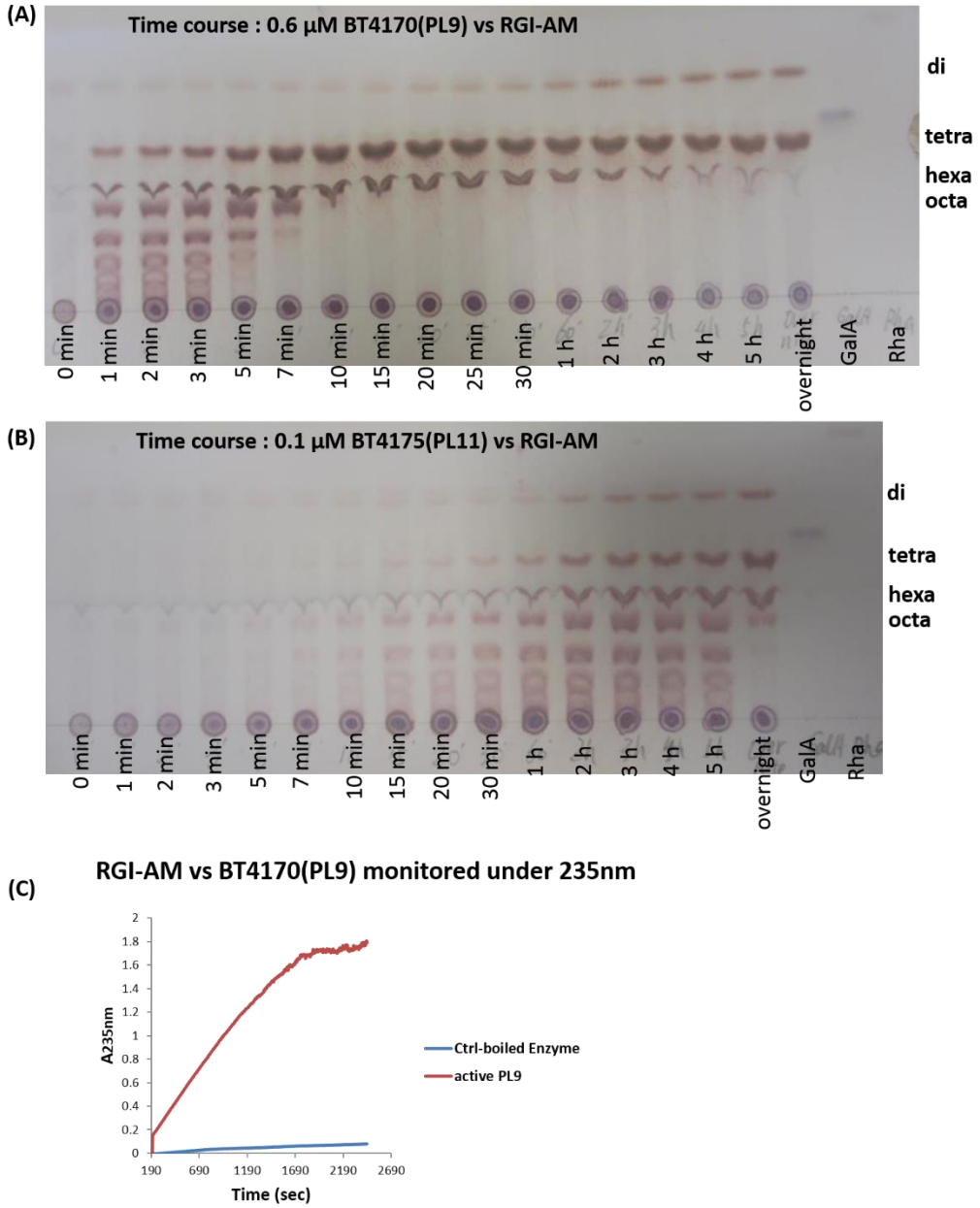


Figure 4.7 Comparison of the two lyases BT4170 and BT4175 against RGI-AM (A and B). The experiment was carried out in 50 mM CAPSO buffer pH 8.5 with 2 mM $CaCl_2$ at 37 ^oC. Oligosaccharides were expected to contain 4,5ΔGalA at the non-reducing end. A scan of the reaction under 235 nm wavelength is on the bottom (C).

Substrate cleavage was associated with an increase in A_{235nm}, consistent with bond cleavage through a lyase mechanism that generates a double bond between C4 and C5 of GalA (4,5ΔGalA) leading to an increase in ultraviolet absorbance. Interestingly BT4175 (PL11) also cleaved RGI-PG, while the two PL9 lyases (BT4170 and BT4183) did not digest this

polysaccharide (**Figure 4.8** and **Figure 4.9**), indicating that the enzymes are unable to accommodate Gal side chains. Data were confirmed by both TLC and HPLC.

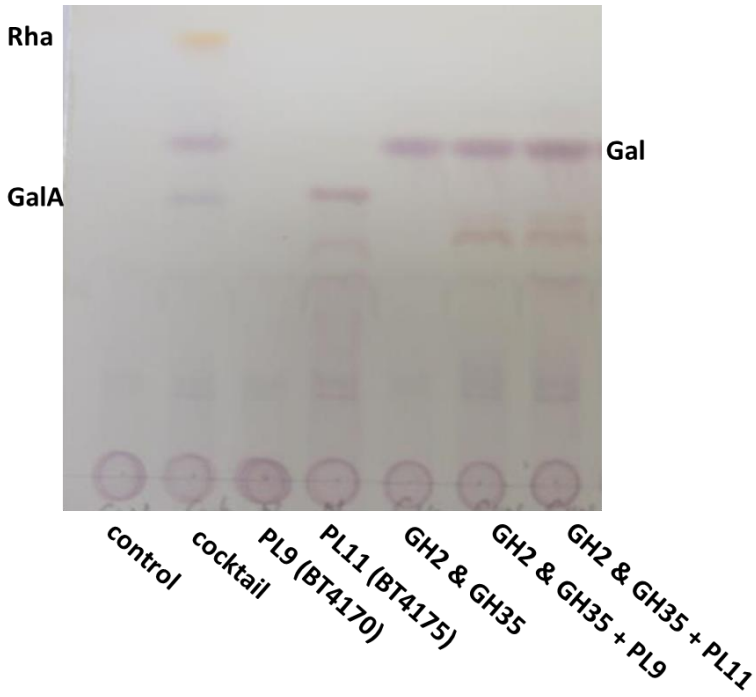


Figure 4.8 Digestion of RGI-PG by BT4170 and BT4175. Reactions were carried out in 50 mM Na-HEPES buffer pH 7.5 containing 2 mM $CaCl_2$ at 37 ^{o}C for 16 h. The substrate was at 5 mg/ml and enzyme at 1 µM. The cocktail contained all the enzymes available from RGI PUL.

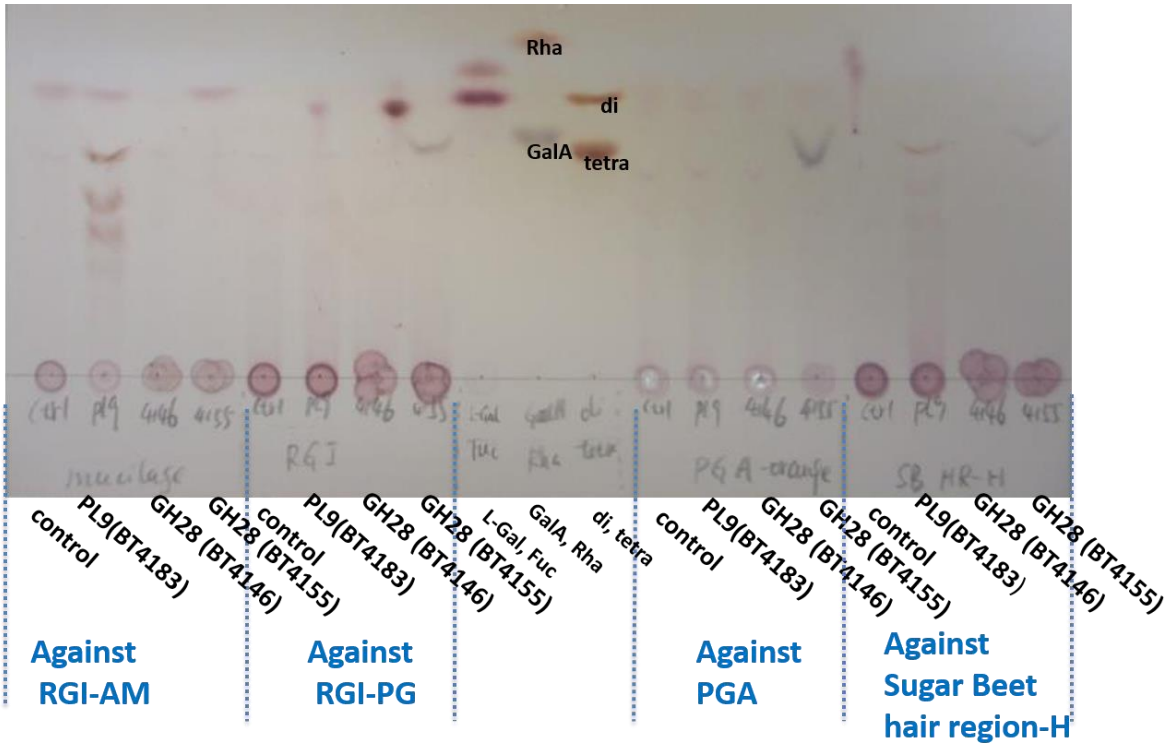


Figure 4.9 Activity of BT4183, BT4146 and BT4155 against various pectins. The enzymes at 1 μM were incubated with RGI and PGA (polygalacturonic acid) at 3 mg/ml in 50 mM sodium phosphate buffer, pH 7.0, containing 2 mM $CaCl_2$ for 16 h at 37 ^{o}C. The reactions were monitored by TLC. Sugar beet hair region-H is purified by INRA in Nantes.

The pH profile of the lyases BT4175 and BT4170 revealed a pH optimum of 8.5 for both enzymes, **Figure 4.10**, while BT4183 displays a pH optimum of 7.0, which may, in part, explain why the bacterium expresses two RGI lyases from the same family. Kinetic data, **Figure 4.11,** showed that BT4175 was significantly more active against the RGI backbone (RGI-AM) than BT4170 (PL9). BT4170, BT4183 and BT4175 displayed no activity against homogalacturonic acid, demonstrating that all three enzymes are endo-acting rhamnogalacturonan I (RGI) lyases

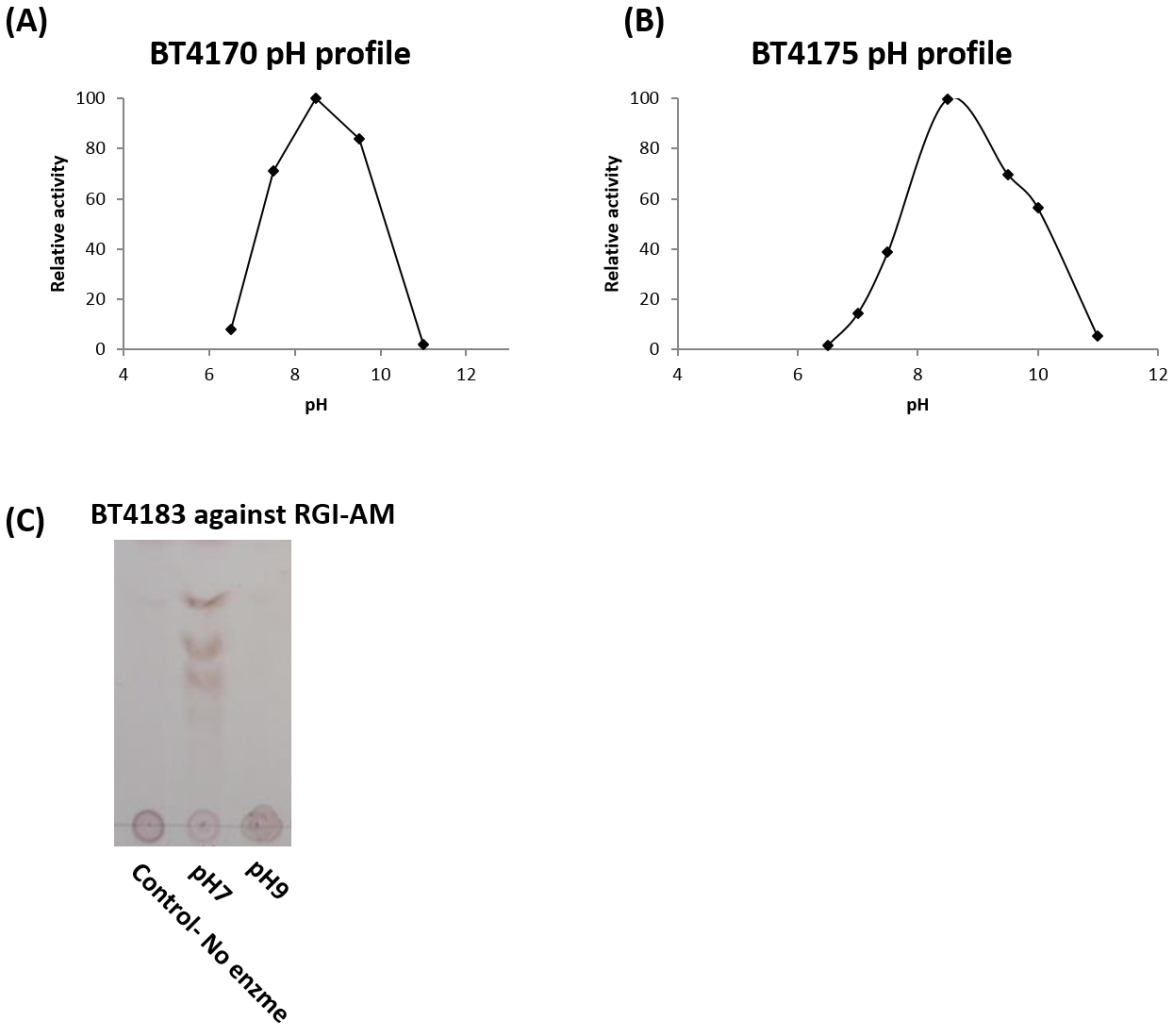


Figure 4.10 The pH profiles of the lyases BT4170 (A), 4183 (PL9) and BT4175 (PL11)(B) against RGI-AM. The buffers used were at 20 mM and consisted of Phosphate/Citrate for pH 6.5, Tris-HCl for pH 7, Na-HEPES for pH 7.5, Tris-HCl for pH 8.5, CAPSO for pH >9. Preliminary TLC data showing BT4183 (PL9) prefers a neutral pH are at the bottom (C). Activity at the optimal pH was defined as 100% in (A) and (B).

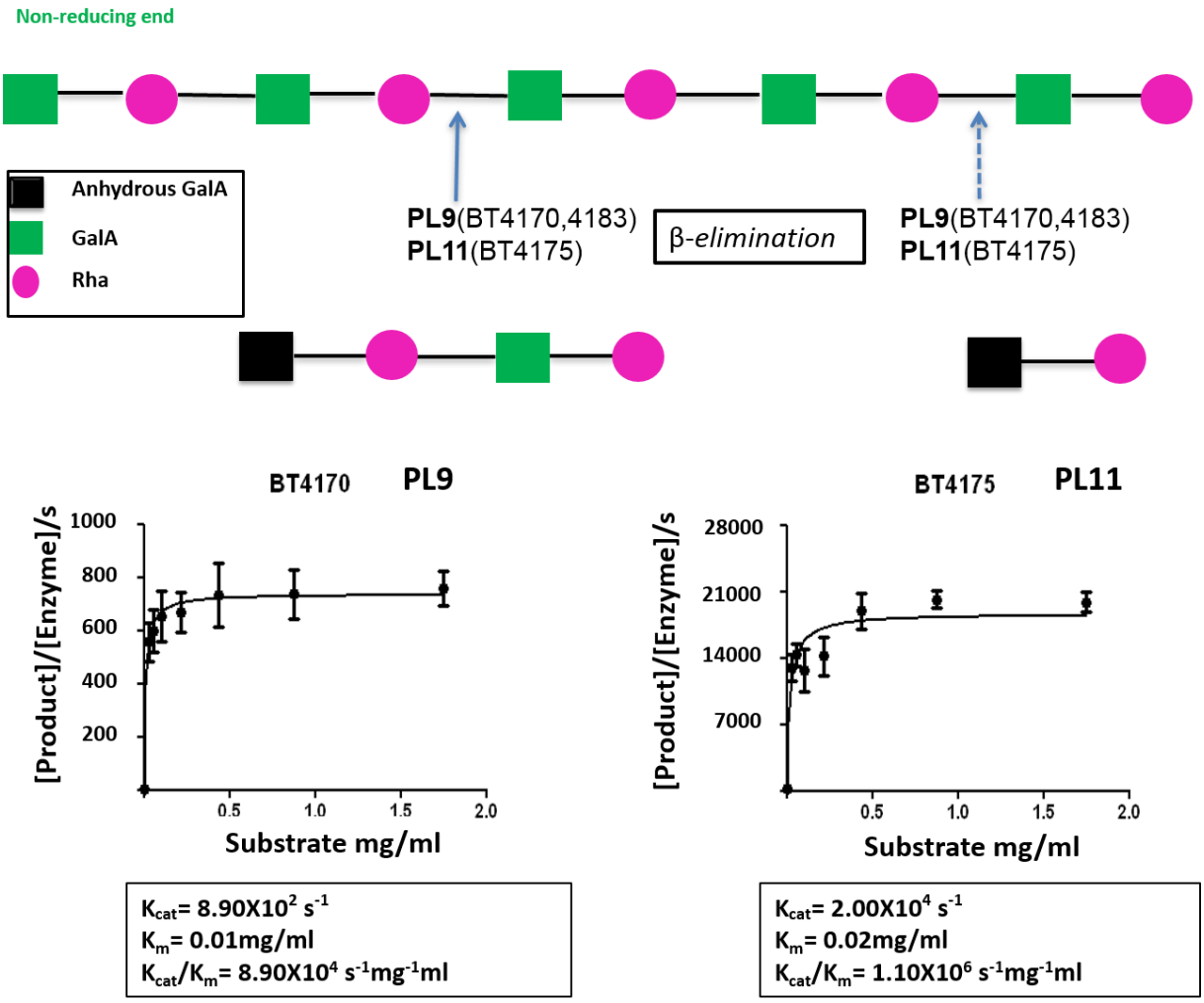


Figure 4.11 The activity of BT4170 (PL9) and BT4175 (PL11) against RGI-AM. The assays were carried out in 20 mM Tris-HCl buffer pH 8.5 at 37 °C. Note that the K_m could only be estimated as catalytic rate and could not be measured at a substrate concentration below this kinetic constant. Enzyme concentration for BT4170 (PL9) and BT4175 (PL11) is 100 nM and 3 nM, respectively. Three replicates were performed and error bars showed the standard error.

The activity of BT4170 (PL9) and BT4175 (PL11) against different oligosaccharides purified from polysaccharide lyase digested RGI-AM (the preparation of these molecules is described in Chapter 2.4) was evaluated. These oligosaccharides, which varied in their degree of polymerization (d.p.) from 2 to 8, contain an unsaturated GalA and saturated Rha at the non-reducing and reducing ends, respectively. The di-saccharide (4,5ΔGalA-Rha) was confirmed by digestion with the lyase BT4170, which released only Rha (4,5ΔGalA is not detected by either HPLC or TLC), **Figure 4.4**. The other oligosaccharides are defined as tetra-, hexa- and octa- oligosaccharides based on their sizes determined by mass spectrometry, **Figure 4.5**. The

data presented in **Figure 4.12** showed that both enzymes digested hexa- and octa-oligosaccharides but displayed no activity against the tetrasaccharide. Again BT4175 (PL11) was much more active than BT4170 against substrates with a degree of polymerizationof 6 or 8.

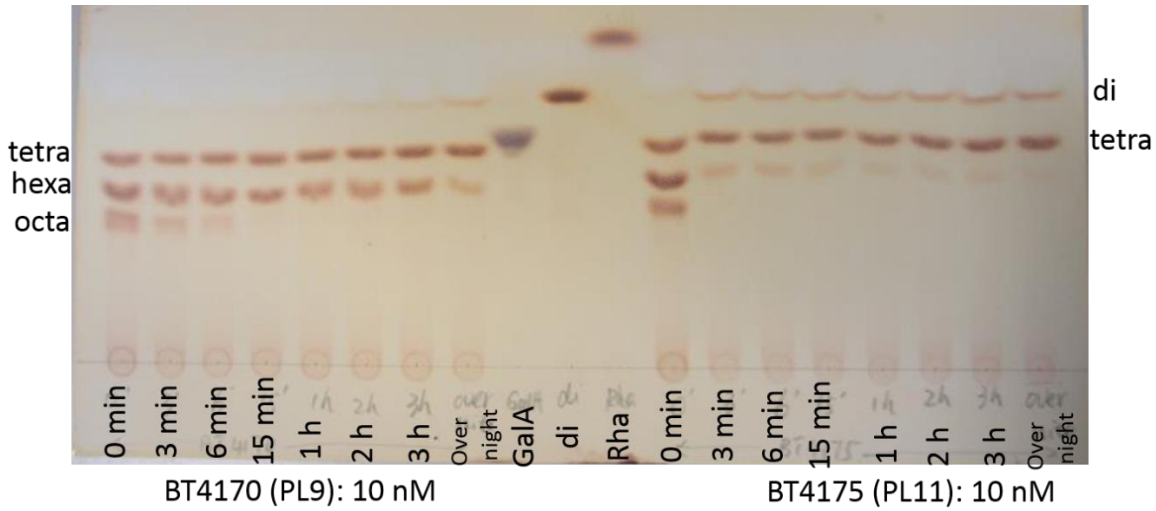


Figure 4.12 Activity of BT4170 (PL9) and BT4175 (PL11) against oligosaccharides. The oligosaccharides were derived by lyase treatment (Chapter 2.4) and thus contain an unsaturated GalA at the non-reducing end. For example, tetra means 4,5ΔGalA-Rha-GalA-Rha. The enzyme reactions were carried out in 50 mM CAPSO buffer pH 8.5 with 2 mM $CaCl_2$ at 37 ºC.

4.2.5 The two GH105 unsaturated rhamnogalacturonyl hydrolases display substrate selectivity

After characterizing the polysaccharide lyases against RGI-AM, the remaining enzymes encoded by PUL-RGI were evaluated against the polysaccharide partially depolymerized with the RGI lyase BT4170; the products generated have an elevated A_{235nm} due to the double bond between C4 and C5 of GalA (4,5ΔGalA) upon cleavage of the Rha-GalA glycosidic bonds. Treatment of lyase digested RG-AM with either of the two GH105 enzymes, BT4174 and BT4176, resulted in a significant reduction in A_{235nm}, reflecting the cleavage of the terminal 4,5ΔGalA residue, which is associated with the loss of A_{235nm} due to the collapse of the double bond, **Figure 4.13** (Deng, O'Neill and York 2006).

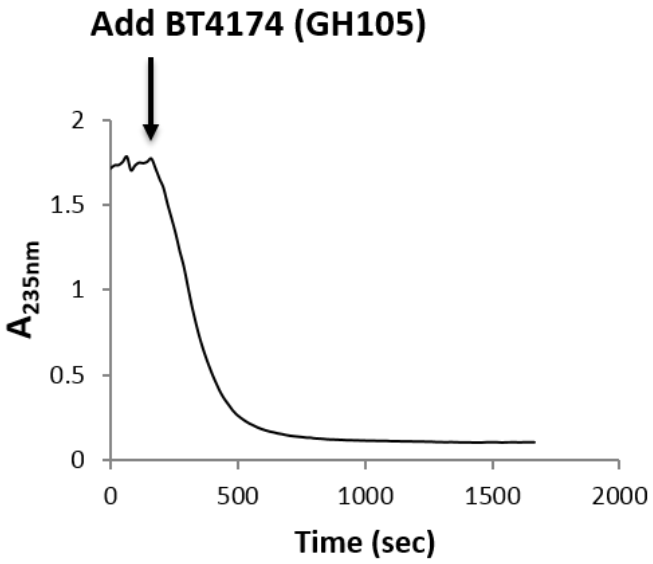


Figure 4.13 Monitoring tetra (4,5ΔGalA-Rha-GalA-Rha) vs BT4174 (GH105) under 235nm. Reaction condition, 20 mM Tris-HCl buffer pH 7.0. Enzyme concentration was 100 nM and substrate concentration was 4 mg/ml.

The two GH105 enzymes displayed no activity against homogalacturonic acid, and thus BT4174 and BT4176, reflecting their specificity for RGI are defined as Δ4,5-unsaturated rhamnogalacturonyl hydrolases (Δ4,5-RGIases). The pH optimum of both GH105 enzymes was 7.0, **Figure 4.14**, although BT4176 displayed a much broader pH activity profile than BT4174. Kinetic data, presented in **Figure 4.15**, showed that BT4176 and BT4174 selectively target 4,5ΔGalA-Rha and ΔGalA-Rha-GalA-Rha, respectively.

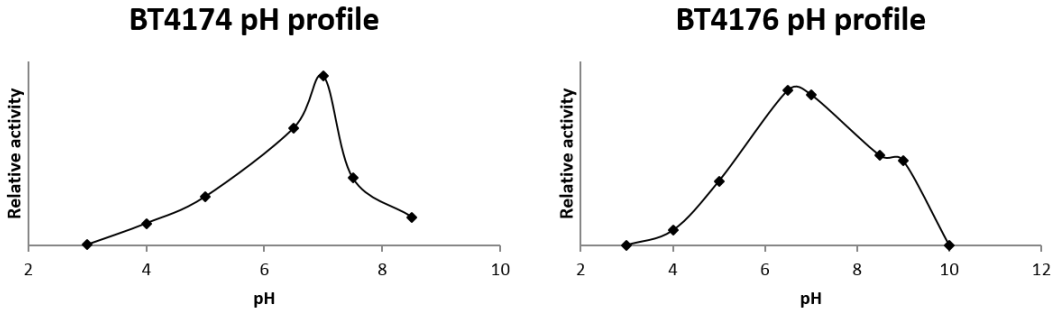
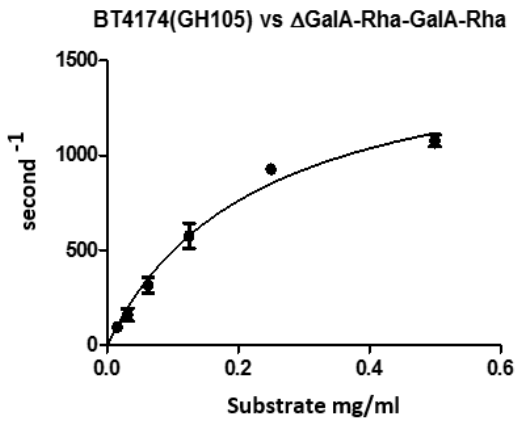
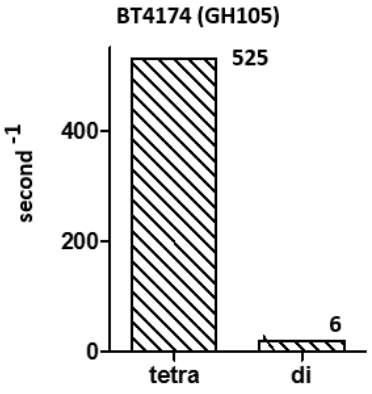
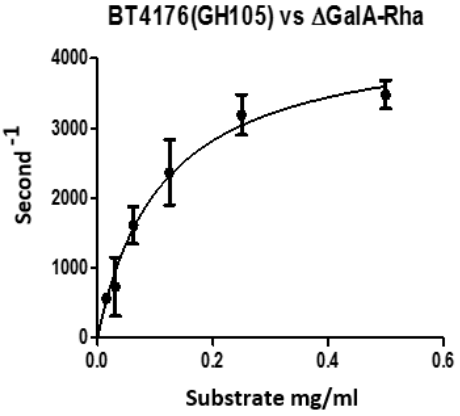


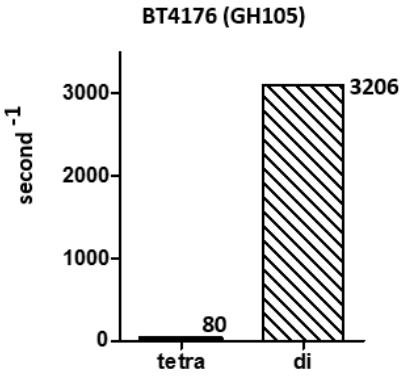
Figure 4.14 pH profiles of the two GH105 enzymes. BT4174 was assayed against 4,5ΔGalA-Rha-GalA-Rha and BT4176 against 4,5ΔGalA-Rha. The buffers (20 mM) used at the different pH values were as follows: citrate phosphate buffer for pH 3-5, PC buffer for pH 6.5, Tris-HCl for pH7, Na-HEPES for pH 7.5, Tris-HCl for pH 8.5, CAPSO for pH >9. Activity at the optimal pH was defined as relative 100%.





K_{cat}=1627 s^{-1}
K_m=0.33 mM
K_{cat}/K_m=4930 $s^{-1}mM^{-1}$





K_{cat}=4420 s^{-1}
K_m=0.35 mM
K_{cat}/K_m=12628 $s^{-1}mM^{-1}$

Figure 4.15 Kinetics of BT4174 and BT4176 against oligosaccharides. The activity of the two enzymes against 4,5ΔGalA-Rha-GalA-Rha (tetra) and 4,5ΔGalA-Rha (di) were evaluated in 20 mM Tris-HCl buffer pH 7.0. Enzyme concentration for BT4174 and BT4176 were 100 nM and 10 nM, respectively. The relative activity of the enzymes (100 nM) against the two substrates was determined at a single substrate concentration of 4 mg/ml. Three replicates were performed and error bars showed the standard error.

4.2.6 The specificity of BT4145 (GH106) and BT4153 (GH28)

Incubation of the two RGI lyases and Δ4,5-RGIases with RGI-AM generated an oligosaccharide presumed to be Rha-GalA-Rha and a trace amount of Rha, consistent with the activities of the four enzymes, **Figure 4.16**. Incubation of RGI-AM with all the soluble recombinant enzymes (defined as the "enzyme cocktail") encoded by PUL-RGI generated exclusively GalA and Rha. To identify the enzymes that converted the trisaccharide into the two sugars, an "individual enzyme knockout" experiment was performed, in which each enzyme was sequentially omitted from the "enzyme cocktail", which was then incubated with RGI-AM overnight. The data, **Figure 4.16**, showed that removal of BT4145 or BT4153 from the "enzyme cocktail" greatly reduced the amount of GalA and Rha generated from RGI-AM, suggesting that these enzymes contribute to the degradation of the Rha-GalA-Rha trisaccharide.

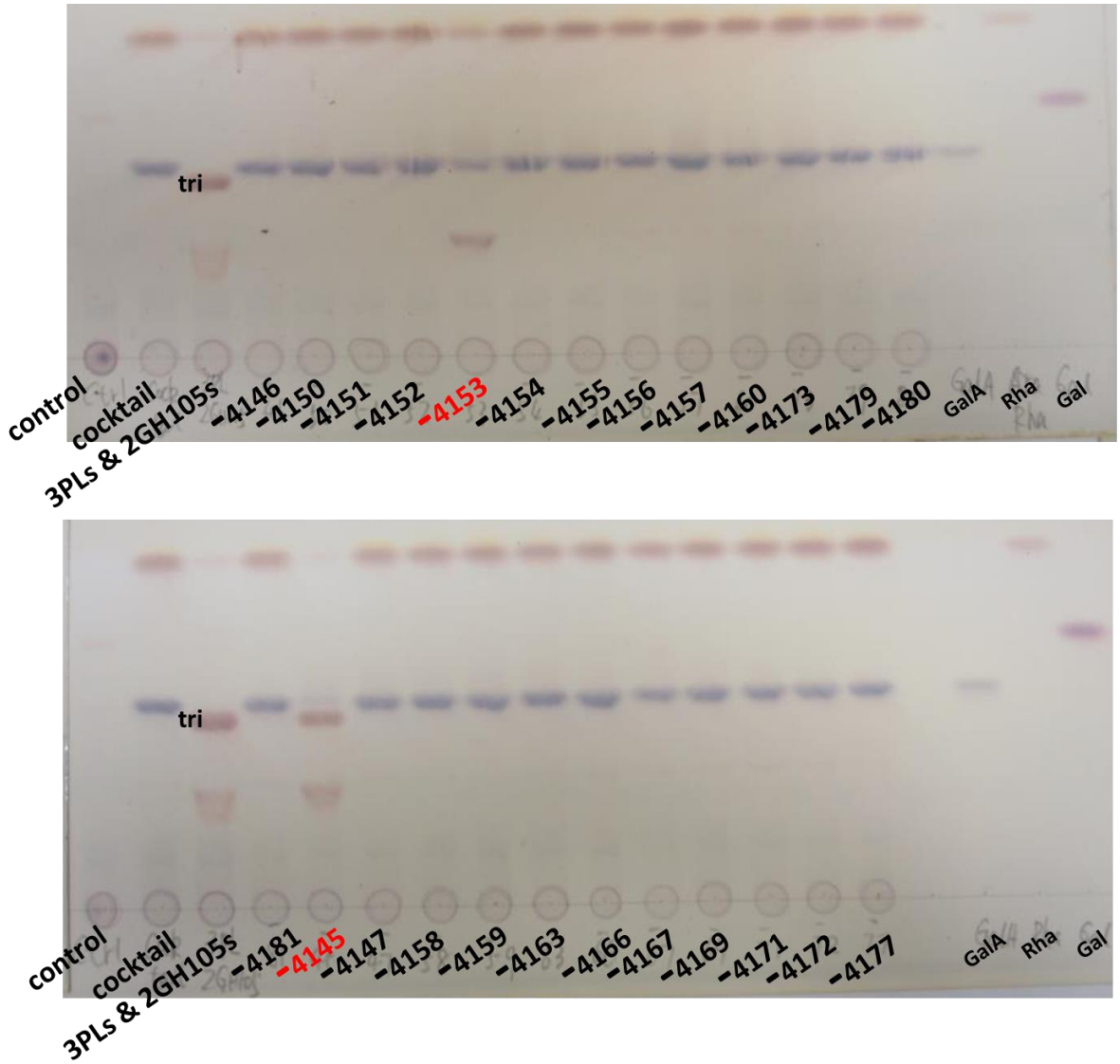


Figure 4.16 Identification of enzymes that contribute to RGI-AM depolymerization. The enzyme cocktail either intact or deficient in a single enzyme (e.g. -4158 means that BT4158 was omitted from the cocktail; each enzyme in the cocktail was at 1 μM) was incubated with RGI-AM at 5 mg/ml in 50 mM Na-HEPES buffer, pH7.5, with 2 mM of $CaCl_2$ at 37 ºC for 16 h. The enzyme mixture in the lane labeled 3PLs & 2GH105s contains the three RGI lyases (PLs) and the two Δ4,5-RGIases (GH105s). The reactions were analyzed by TLC. The putative trisaccharide was labeled as "tri".

This view is consistent with the observation that Rha and GalA and 4,5ΔGalA are the only products generated when RGI-AM was incubated with BT4145 or BT4153 in conjunction with the two RGI lyases and Δ4,5-RGIases, **Figure 4.17**. Given that BT4145 and BT4153 are in rhamnosidase (GH106)- and galacturonase (GH28)-specific families, it is likely that the two

enzymes act sequentially on Rha-GalA-Rha (generated by the lyases and Δ4,5-RGIases) to convert the trisaccharide into its constituent monosaccharides, Rha and GalA.

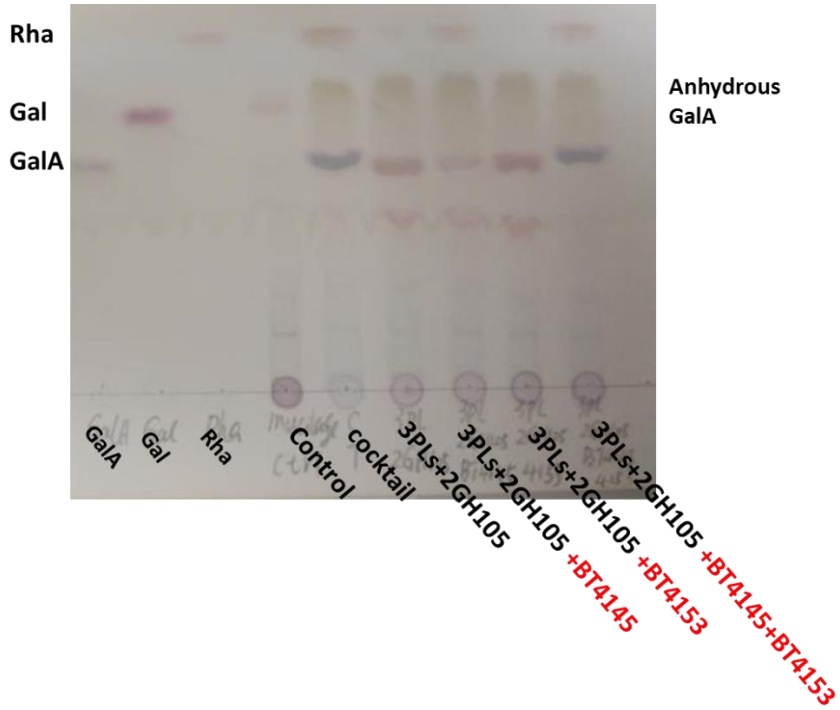


Figure 4.17 The two newly found enzymes BT4145 (GH106) and BT4153 (GH28) contribute to the complete digestion of the RGI-AM. Reaction was carried out in 50 mM Na-HEPES buffer pH 7.5 with 2 mM of $CaCl_2$ at 37 ºC; substrate concentration, 5mg/ml and enzyme concentration, 1 µM; PL, pectic lyases; the 2GH105s are BT4174 and BT4176.

4.2.7 Preliminary analysis of the activity of BT4145 against oligosaccharides

The oligosaccharides Rha-GalA-Rha-GalA-Rha-GalA, Rha-GalA-Rha-GalA and Rha-GalA, generated from RGI-AM as described in Chapter 2.4, were incubated with the putative rhamnosidase BT4145 and the products were analyzed by HPLC. The data, displayed in **Figure 4.18**, showed that all three oligosaccharides were degraded within 1 min, implying that the enzyme displayed similar activity against these substrates, although further kinetic experiments are required to support this conclusion (see Discussion Section). The observation that Rha was generated against all the oligosaccharides but not GalA supported the view that BT4145 is a rhamnosidase.

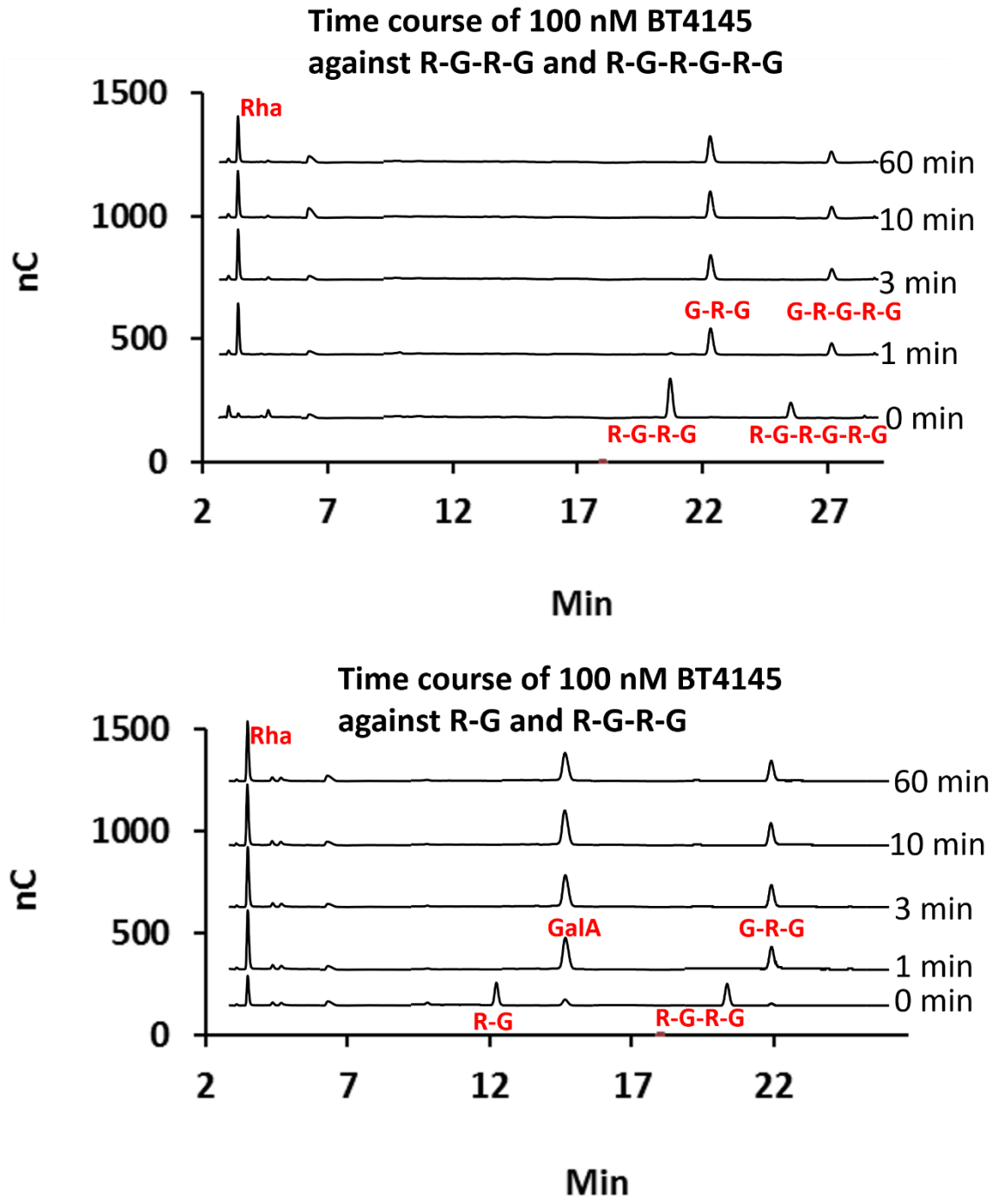


Figure 4.18 Activity of BT4145 (GH106) against oligosaccharides containing Rha at the non-reducing end. The reactions were carried out in 20 mM Na-HEPES pH 7.5 at 37 ºC and monitored by HPLC. The oligosaccharide structures are putative except for R-G-R-G and R-G-R-G-R-G which were confirmed by mass spectrometry in **Figure 4.6**.

4.2.8 Synergy between the rhamnosidase BT4145 and the galacturonidase BT4153

The saturated oligosaccharides Rha-GalA-Rha-GalA and Rha-GalA-Rha-GalA-Rha-GalA, generated as described in Section 4.2.3 were treated with the rhamnosidase BT4145 and the products analyzed by HPLC. The data, **Figure 4.19**, revealed the release of Rha and the oligosaccharide products GalA-Rha-GalA and GalA-Rha-GalA-Rha-GalA. These oligosaccharides were then hydrolyzed by the galacturonidase BT4153 to generate Rha-GalA and Rha-GalA-Rha-GalA, demonstrating that the hydrolysis of RGI requires the sequential action of the two enzymes. It should be noted that the endo-acting RGIase used to generate the original oligosaccharides generated two minor products that appeared to comprise GalA-Rha-GalA and GalA-Rha-GalA-Rha-GalA. Given that the enzyme cleaves galacturonosidic bonds (GalA is positioned upstream of the glycosidic bond), the data may indicate that RGI does not comprise entirely of repeating Rha-Gal units; the polysaccharide appears to contain successive GalA residues. The position of these GalA-GalA units in the polysaccharide is unclear. The experiment confirmed that all the enzymes used were active and the mucilage substrate can be digested step by step if controlled carefully. This makes it possible to generate bespoke oligosaccharides from RGI with Rha at the non-reducing end and GalA at the reducing end, products that can not be produced by the use of pectic lyases. A scaled-up experiment was carried out later to purify the following oligosaccharides: Rha-GalA-Rha-GalA-Rha-GalA, Rha-GalA-Rha-GalA and Rha-GalA.

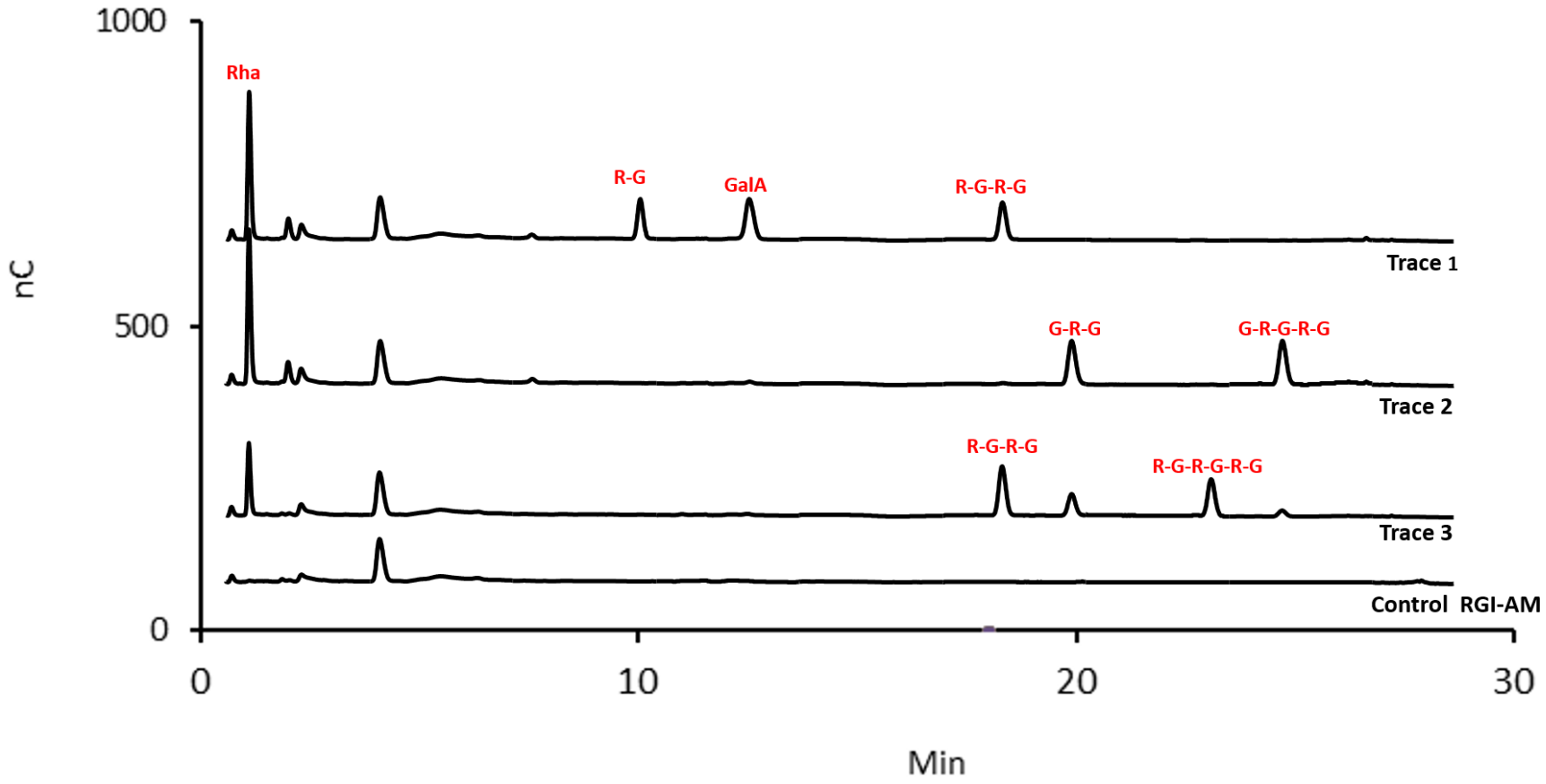


Figure 4.19 HPLC analysis of the enzymatic degradation of RGI-AM derived oligosaccharides by an endo-RGIGH106 (BT4145) rhamnosidase and GH28 (BT4153) rhamnogalacturonase. In trace 3, oligosaccharides were produced by incubating RGI-AM with U29SH for 3h. In trace 2 the oligosaccharides were treated with BT4145 and the enzyme was then inactivated by boiling. In trace 1 the products generated by BT4145 were treated with BT4153. The enzymes were at 1 μM; and the enzymatic reactions were carried out in 50 mM Tris-HCl buffer, pH 7.5., for 3 h.

4.2.9 Synergy between all the backbone cleaving enzymes.

Individually, neither the rhamnosidase BT4145 (GH106) nor the rhamnogalacturonase BT4153 (GH28) is active against the RGI backbone. But when mixed together, the two enzymes release GalA and Rha from RGI-AM (**Figure 4.20**), indicating the expected enzymatic synergy. When BT4170 (PL9) and the two GH105s (BT4174 and BT176) were included with the GH28 and GH106 enzymes, the production of Rha and GalA was increased further. Quantification by HPLC showed that 2.5 and 1.6 more Rha and GalA were generated if all the enzymes were used in the reactions, **Figure 4.20**, compared to when reactions contained just the BT4145 and BT4153. Based on the calculated ratio of GalA (anhydrous): GalA: Rha = (2.5-1.6) 0.9: 1.6: 2.5, which approximates to 1:2:3, it is deduced that for every repeating unit of hexasaccharide ($[Rha\text{-}GalA\text{-}Rha\text{-}GalA\text{-}Rha\text{-}GalA]_n$), BT4170, the PL9 RGI lyase, can only cleave once, producing one 4,5ΔGalA, while the rest of the enzyme cocktail releases the other two GalA and three Rha units.

The fact that the BT4145 and BT4153 releases 2.5 times less product than the cocktail of enzymes containing RGI lyases and Δ4,5-RGIases indicates a possible block or modification on the presumably smooth RGI backbone extracted from *Arabidopsis* seed mucilage. It would appear that non-reducing end of 40% of the chains were capped and were thus not accessible to the exo-acting enzymes.

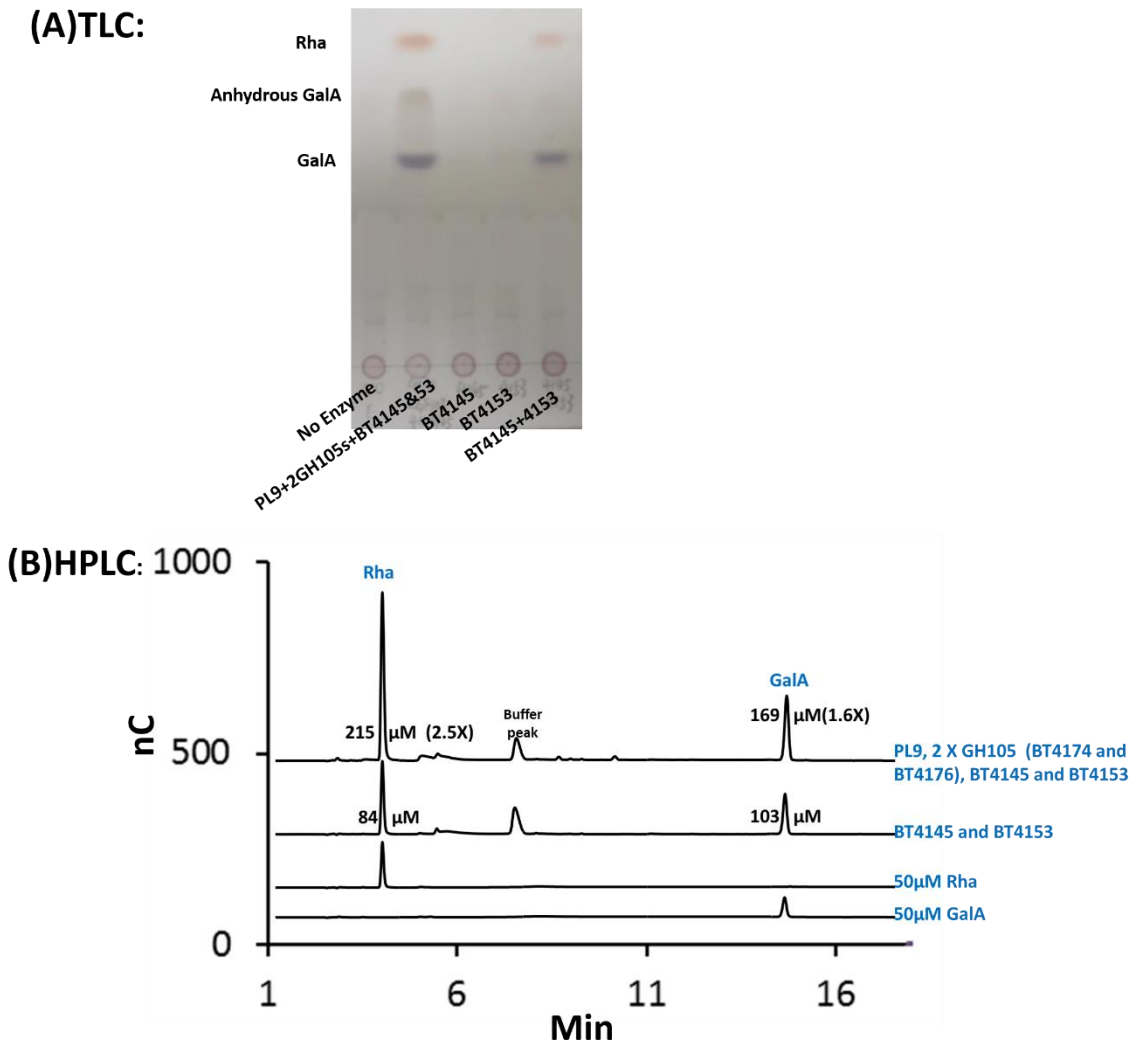


Figure 4.20 Synergy between the rhamnosidase BT4145 (GH106) and the rhamnogalacturonase BT4153 (GH28) against RGI-AM. (A), TLC of the RGI-AM digestion by different enzymes combinations; (B), HPLC quantification of the synergy between these two enzymes, compared with the complete digestion involving all enzymes in RGI-AM degradation. Digestion was carried out in 50 mM Na-HEPES buffer pH 7.5 with 2 mM $CaCl_2$; quantification of the products was calibrated against the standards.

4.2.10 Enzymatic degradation of RGI from potato (RGI-PG)

Although the bulk of the galactan and arabinan side chains are removed by *B. thetaiotamicron*, it is highly likely that small side chains of the two polysaccharides will remain appended to the RGI backbone as the endo and exo galactanases and arabinanases would be sterically restricted by the Rha-GalA chain. In turn any such side chain remnants would impede the action of the

backbone degrading enzymes. To explore this in more detail, the action of the enzymes encoded by PUL-RGI against RGI-PG was explored.

All the available soluble recombinant enzymes encoded by PUL-RGI were incubated with RGI-PG and the products were evaluated by HPLC. The data revealed the release of primarily Gal, Rha, GalA and a little Ara, **Figure 4.21**. To explore the synergistic interactions within the enzymatic cocktail, the effect of removing single enzymes, or related enzyme combinations, from the mixture were assessed. The data showed that removal of all the RGI lyases and the Δ4,5GalA-RGIases caused the largest reduction in Rha release, while the generation of Gal was also affected by these enzyme omissions, **Figure 4.21**.

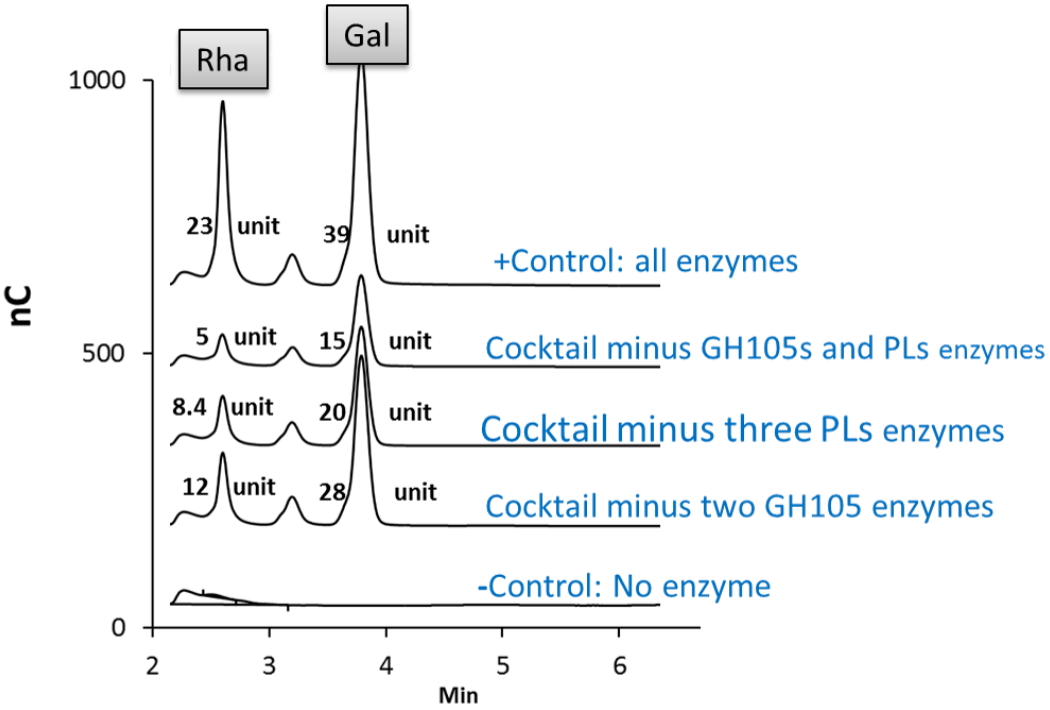


Figure 4.21 Effects of the deletion of GH105 enzymes and RGI lyases from the enzyme cocktail encoded by PUL-RGI. Peaks are integrated and labelled.

To explore how the galactan side chain is degraded, glycoside hydrolases from GH2 and GH35 (populated by β-galactosidases) were removed, together with the RGI lyases and the Δ4,5GalA-RGIases, from the enzyme cocktail, which was then used to digest RGI-PG. The data, presented in **Figure 4.22**, showed that removal of the GH2 enzymes from the cocktail

eliminated Rha, while also reducing the generation of Gal, confirming their β-galactosidase function. Removal of the GH35 enzyme BT4160 did not affect Rha production, but had a moderate influence on Gal release, suggesting that BT4160 is also a β-galactosidase. Deletion of both the GH2s and GH35 enzymes eliminated all Gal released from RGI-PG and prevented the appearance of Rha. These data suggest that the β-galactosidase function of the GH2, and possibly the GH35, enzymes influence the release of Rha from RGI-PG.

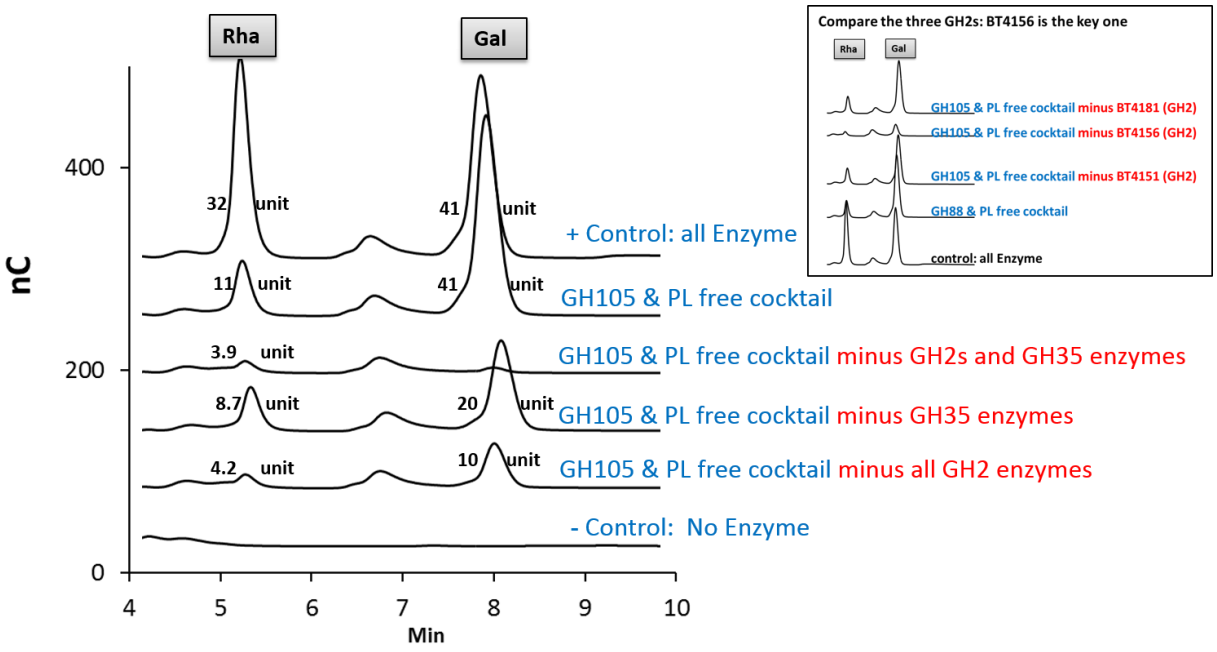
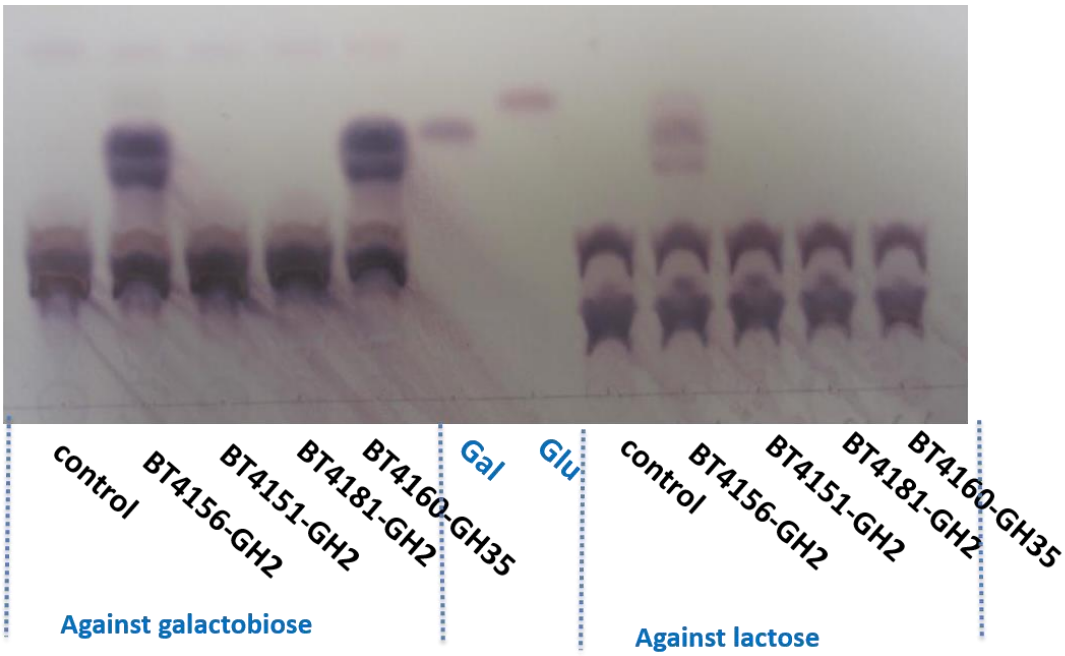


Figure 4.22 Evaluation of the effect of deletion of the GH2 and GH35 β-galactosidases from GH105 and PL-free enzyme cocktails on the enzymatic degradation of RGI-PG. Peaks are integrated and labeled. BT4151, BT4156, BT4181 are in GH2 and BT4160 is in GH35.

To explore the role of the individual GH2 enzymes on both Gal and Rha release, each enzyme was sequentially removed from the cocktail incubated with RGI-PG. The data, **Figure 4.22**, showed that only the removal of BT4156 caused a substantial reduction in Rha release, indicating that this β-galactosidase played a particularly important role in facilitating the

depolymerization of the RGI backbone. To further explore the specificity of the GH2 and GH35 enzymes, their activity against a range of other glycans was assessed. The data, **Figure 4.23**, showed that BT4156-GH2 and BT4160-GH35 were active against β-1,4-galactobiose and only BT4156-GH2 was able to cleave lactose, albeit very slowly. The GH2s BT4151 and BT4156 were very active against 4-nitrophenyl-β-D-galactopyranoside (4NP-Gal), while the GH35 enzyme, BT4160, hydrolyzed 4NP-Gal extremely slowly, with a yellow color only appearing after incubation overnight. No activity was observed for BT4181, suggesting that this enzyme is not a β-galactosidase.



	Galactobiose	Lactose	PNP-α-D-gal	PNP-β-D-gal
BT4151 (GH2)	inactive	inactive	inactive	active
BT4156 (GH2)	active	active	inactive	active
BT4181 (GH2)	inactive	inactive	inactive	inactive
BT4160 (GH35)	active	inactive	inactive	active(overnight)

Figure 4.23 Activity of GH2s and GH35 in the PUL-RGI against galactobiose, lactose and related 4NP substrates. The overnight reaction was carried out in 50 mM Na-HEPES buffer pH 7.5 with 2 mM of $CaCl_2$ at 37 ^oC; substrate concentration, 10 mM and enzyme concentration, 1 µM; Gal, galactose; Glu, glucose.

To explore the mechanism of Gal release further, the β-galactosidases were incubated with RGI-PG individually and together. The data, **Figure 4.24**, showed that the three GH2 enzymes and the GH35 β-galactosidase released a similar amount of Gal but, collectively, the enzymes released more Gal indicating possible synergy between these biocatalysts.

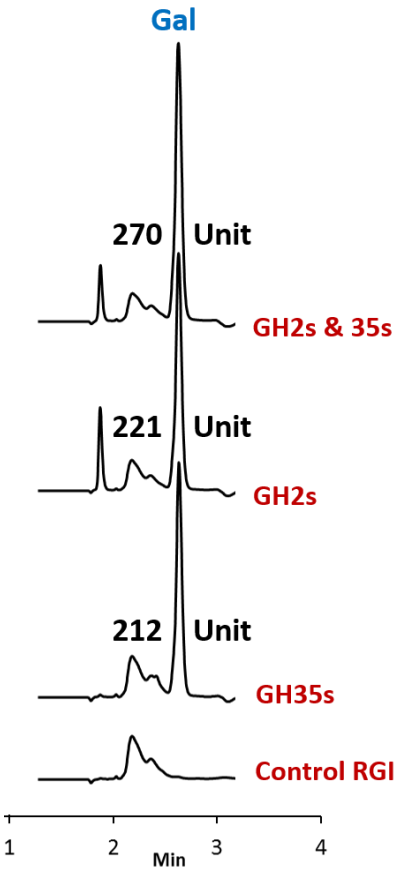


Figure 4.24 Galactose released by GH2 and GH35 β-galactosidases in the digestion of RGI from potato. HPLC peaks are integrated and labeled. BT4151, BT4156, BT4181 are in GH2 and BT4160 is in GH35.

4.2.11 Analysis of the "non-CAZy" proteins encoded by PUL-RGI

To explore the possible contribution of the "non-CAZy" proteins to RGI depolymerization, RGI-PG was incubated with an enzyme cocktail containing different combinations of the CAZy enzymes and all the non-CAZy hypothetical proteins. The data, **Figure 4.25**, showed that no Rha was released in the absence of the hypothetical proteins. Thus, one or more of the non-CAZy enzymes must act in synergy with the backbone cleaving enzymes to release Rha.

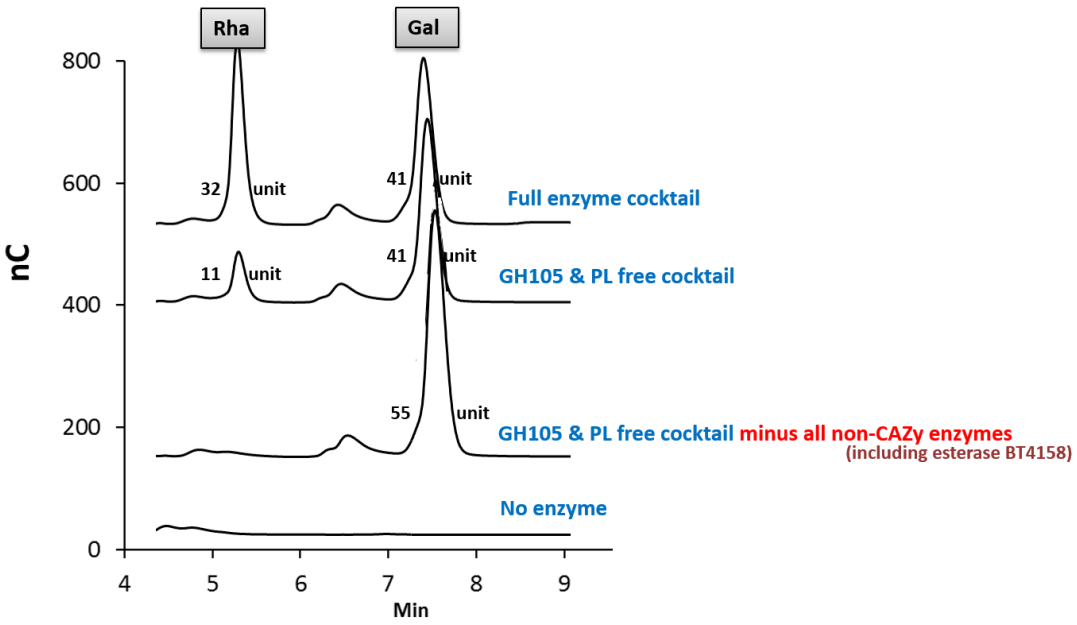


Figure 4.25 Evaluation of the deletion of non-CAZy enzymes from the GH105 and PL-free enzyme cocktail incubated with RGI-PG. The reactions were analyzed by HPLC. The integrated peaks were assigned arbitrary units and their identity was based on their co-migration with standards.

To explore the identity of the non-CAZy protein(s) that contributed to RGI degradation, these proteins, individually, were omitted from an enzyme cocktail lacking the GH105 unsaturated rhamnogalacturonyl hydrolases and the PL9 and PL11 polysaccharide lyases, and the effect on the products generated was evaluated by HPLC. The omission of BT4158 from the enzyme cocktail resulted in the generation of no detectable Rha, however, removal of the other non-CAZy proteins from the mixture of biocatalysts did not influence the product profile. To identify the activity displayed by BT4158, the enzyme was incubated with a range of polysaccharides and aryl-glycosides and 4-nitrophenyl-acetate (4NPA). BT4158 was only active against 4NPA, **Figure 4.26**, demonstrating that the enzyme is an esterase. The role of an acetyl esterase in the depolymerization of RGI is analyzed in detail in the Discussion Section.

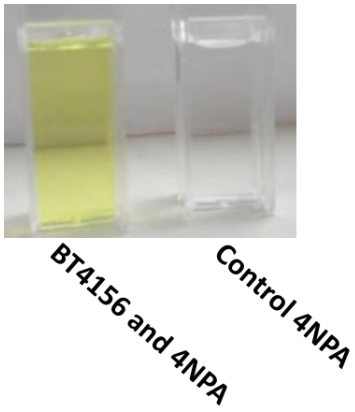


Figure 4.26 Visible colour change indicated that BT4158 was active against 4-nitrophenyl-acetate (4NPA). Enzyme concentration was 1 µM, and substrate concentration 1 mM. The reaction was carried out in 20 mM Tris-HCl buffer, pH 7.0 for 10min.

4.2.12 Removal of the galactan side chain enables the lyase BT4170 (PL9) to access the RGI-PG

Previously BT4170 (PL9) was shown to digest RGI-AM but not RGI-PG. To establish whether access to the RGI backbone was restricted by the galactose side chains, the lyase was incubated with RGI-PG together with the GH2 and GH35 β-galactosidases. The data, presented in **Figure 4.8**, showed that BT4170 displayed activity against RGI-PG in the presence of the β-galactosidases, demonstrating that the galactose side chains impeded access of the lyase to the polysaccharide backbone.

4.2.13 The recent discovery of a GH28 galacturonase (BT4155) in the PUL-RGI

A further enzyme encoded by PUL-RGI, BT4155 (GH28), released a small amount of GalA from polygalacturonic acid and RG-PG, but not from RG-AM, **Figure 4.9** The enzyme appears to be an exopolygalacturonase. The observed activity against RGI-PG likely reflects polygalacturonic acid impurities in the RGI preparation.

4.2.14 Regulation of the RGI utilization by *Bacteroides thetaiotaomicron*

The two-component regulatory system serves as a basic stimulus-response coupling mechanism to allow organisms to sense and respond to changes in many different environmental conditions.(Stock, Robinson and Goudreau 2000). A Previous report (Martens et al. 2011) has also revealed the role such a system plays in gut bacteria such as *B. thetaiotaomicron.* To activate transcription of the RGI-PUL, the sensor of the regulatory protein (hybrid two component system; HTCS) binds to a degradation product released from RGI. In previous studies it was proposed that BT4178 comprises the HTCS that activates RGI-PUL (Martens et al. 2011). To investigate the identity of the activating ligand, RGI-AM was digested by the PL9 lyase BT4170, and the capacity of the reaction products to bind to the sensor domain of BT4178 was evaluated by isothermal titration calorimetry (ITC). The data identified binding of digested RGI-AM to BT4178 indicating the activating ligand was present in this mixture, **Figure 4.27**. Individual oligosaccharides containing 4,5ΔGalA at the non-reducing end, derived from RGI-AM (see Chapter 2.4), were titrated against the sensor domain of BT4178. The data showed that only the tetrasaccharide bound to BT4178. Removal of 4.5ΔGalA from the tetrasaccharide through the action of the GH105 enzyme BT4174 abolished binding of the oligosaccharide to the sensor domain, indicating the importance of ΔGalA in protein-ligand recognition, **Figure 4.27**.

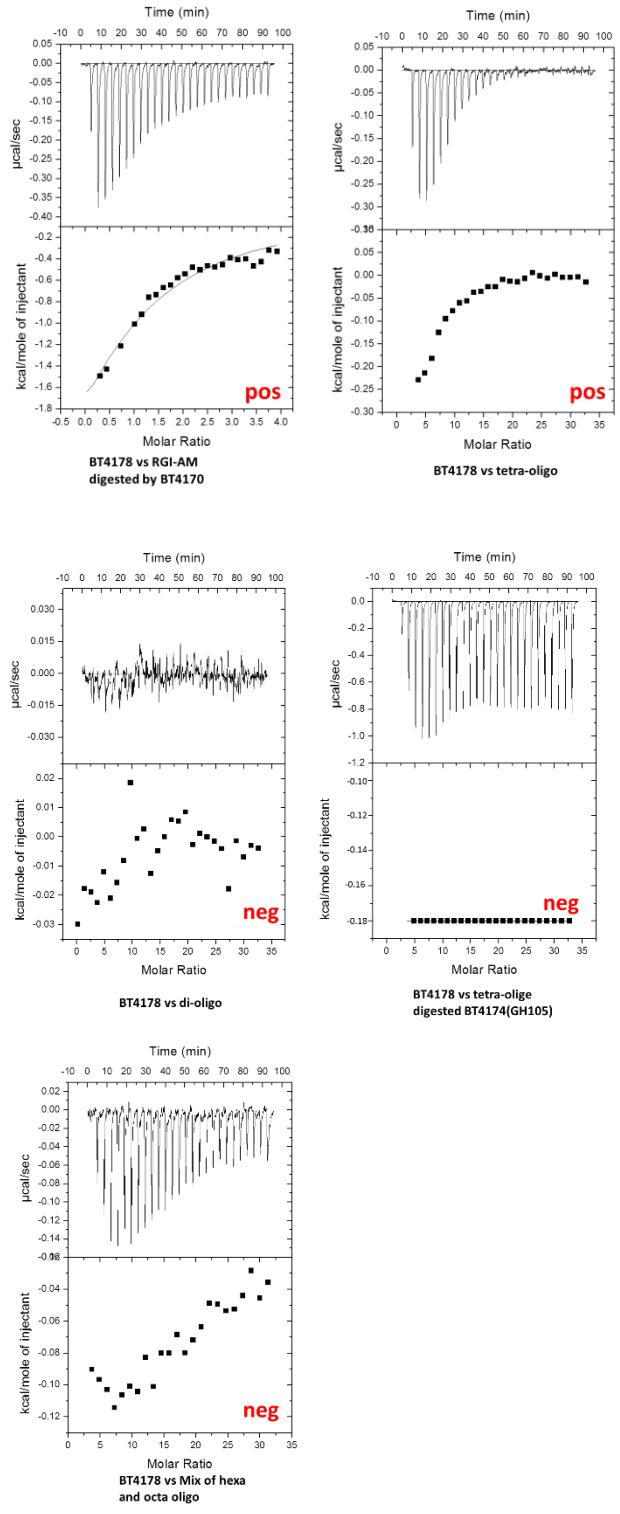


Figure 4.27 Binding of BT4178 to RGI derived oligosaccharides generated by the PL9 lyase BT4170. Titrations indicating protein-ligand binding are labelled in red "pos" and the lack of an interaction is denoted by red "neg". The ligand (5 mM) in the syringe was titrated into the sensor domain of BT4178 (50 µM) in the cell. Titration of saccharide into buffer is subtracted to remove the heats of solute dilution. Parameters for the tetra oligosaccharide binding: K_A 8.7 × 10^3 M^{-1}; ΔH -4818± 174.5 cal mol^{-1}; ΔS = -1.87; N^a= 1.00 ([a]The ITC data were fitted to a single site binding model for all ligands.)

4.2.15 The whole cell assay of the *Bacteroides thetaiotamicron* on RGI

To investigate the possible role of the enzymes on the outer membrane of *B. thetaiotamicron* in RGI degradation, whole cell assays were carried out. *B. thetaiotamicron* was grown to mid exponential phase on RGI (to induce the RGI degrading apparatus) and glucose (to act as a control) and the cells were exposed to RGI in a solution lacking a nitrogen source and thus preventing growth through the lack of energy derived from intracellular metabolism. It is assumed that no import of polysaccharides or oligosaccharides can occur so any observed glycan degradation (monitored by TLC) is mediated by surface enzymes. The data, **Figure 4.28** showed no degradation of RGI by cells cultured on glucose. However, when *B. thetaiotamicron* was grown in RGI-AM or RGI-PG, surface enzymes were capable of digesting both substrates. As shown in **Figure 4.28**, GalA was released from RGI-PG in the first 30 min, although this is possibly due to the presence of contaminating polygalacturonic acid (PGA). Interestingly, when whole cells were incubated with RGI-AM, after ~1 h, both GalA and Rha were evident, together with a dominant oligosaccharide, probably ΔGalA-Rha-Gal-Rha. It is unlikely that the degradation of RGI-AM occurred on the surface as the GH106 rhamnosidase (BT4145) and GH28 galacturonidase (BT4153) contain typical type 1 signal peptides and are thus predicted to be located in the periplasm. Thus RGI-AM may be imported into the periplasm through passive diffusion. It is possible, however, that surface enzymes, which are not encoded by PUL-RGI, mediate surface degradation of RGI-PUL, although no such candidate glycoside hydrolases or lyases were encoded by the other PULs involved in pectin degradation. In contrast, given that Gal, Rha or small oligosaccharides were not released from RGI-PG, this polysaccharide was not imported into the periplasm, suggesting active transport is an absolute requirement for the transport of this pectin molecule across the outer membrane.

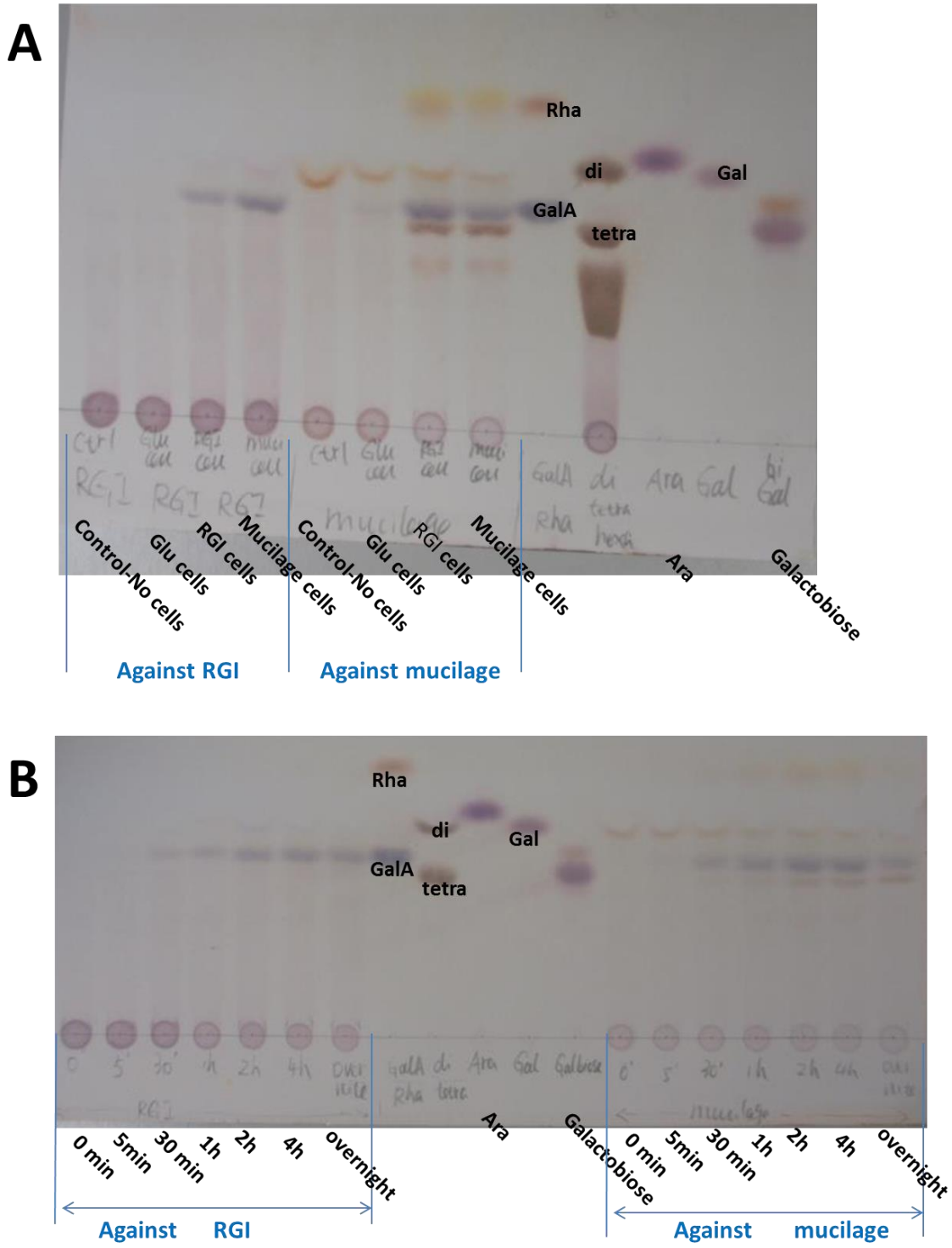


Figure 4.28 TLC analysis of RGI degradation in whole cell assays of *B. thetaiotamicron* grown on glucose, RGI-AM and RGI-PG. In Panel A, "Glu cells" refers to glucose grown cells, "tetra" and "di" oligosaccharides are standard oligosaccharides with ΔGalA on the non-reducing end. In Panel A whole cells were incubated with RGI, while Panel B shows a time course of RGI-PG grown whole cells exposed to RGI-PG or RGI-AM.

4.2.16 The growth curve of wild type *B. thetaiotamicron* and the BT4178 mutant lacking a functional RGI HTCS

To understand the ability of the *B. thetaiotamicron* to utilize pectic polysaccharides, the bacterium was grown on minimal media containing RGII and RGI from different sources. The experiment showed (**Figure 4.29**) that RGII supported the growth of *B. thetaiotamicron* the most, followed by RGI-PG, RG-AM and finally RGI from soybean.

To explore the significance of the HTCS (BT4178) that bound to ΔGalA-Rha-GalA-Rha, the growth profile of the *Δbt4178* was determined. The data, **Figure 4.29**, showed that the gene knockout did not affect growth on glucose, but significantly reduced initial growth on RGI-AM. This is consistent with the specificity of the sensor of BT4178 for oligosaccharides derived from RGI-AM, indicating that the HTCS regulates PUL-RGI.

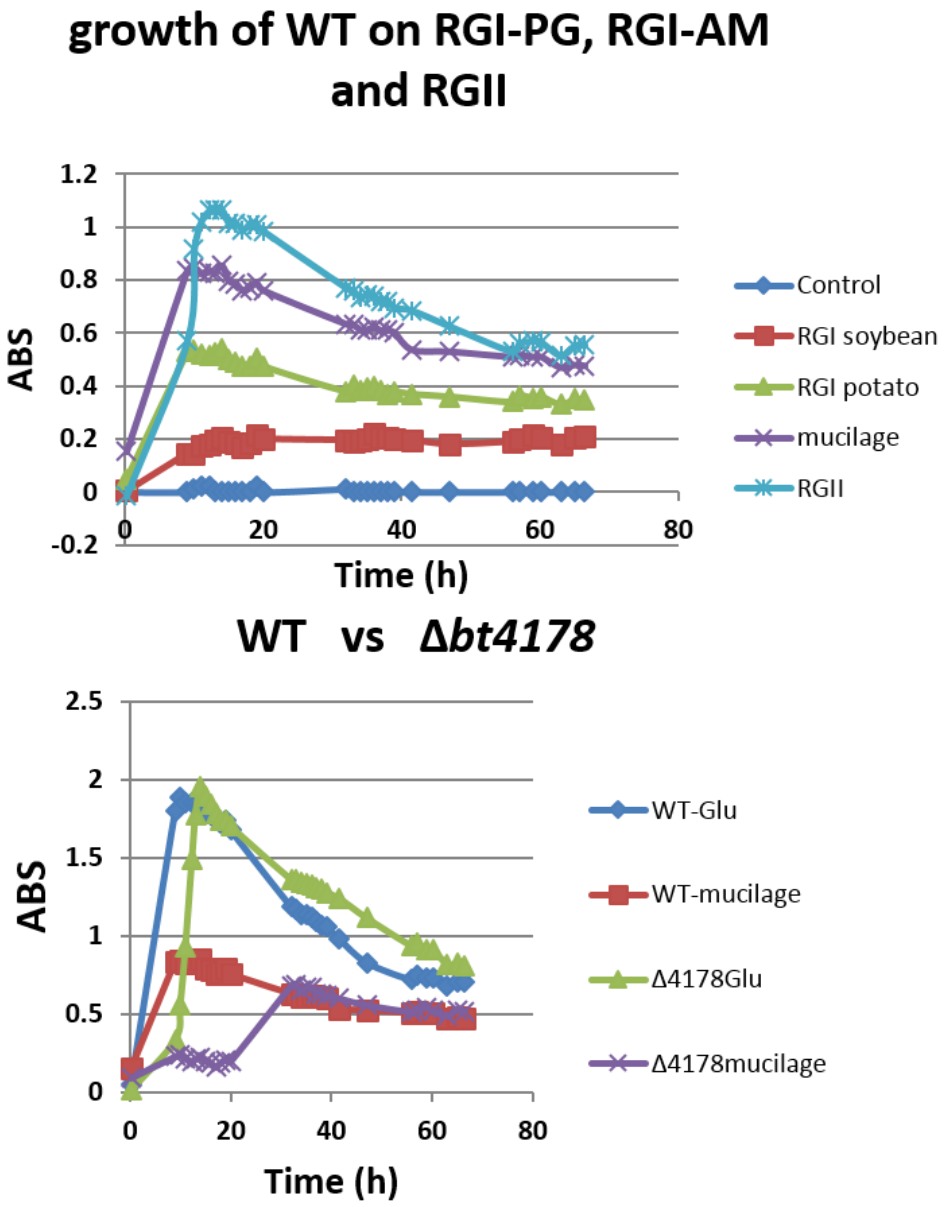


Figure 4.29 The growth profile of wild type *B. thetaiotaomicron* and the *Δbt4178*. Experiments were carried out in 5 ml test tubes and at least three replicates were performed. WT, wild type.

4.2.17 Up-regulation of the *susC* genes in the RGI PUL

SusC-like proteins are outer membrane protein porins that mediate glycan active transport into the periplasm. Genes encoding these proteins are often used as markers for the transcription of PULs as they are the most highly expressed genes in these loci. Growth of *B. thetaiotaomicron* on RGI-AM is predicted to activate transcription of the RGI PUL as a tetrasaccharide derived from RGI activates the HTCS BT4178 associated with this locus. To test this hypothesis

transcription of the two SusC-like genes located in the RGI PUL were determined by qRT-PCR when *B. thetaiotaomicron* was grown on RGI-AM. Surprisingly, while *bt4168* was upregulated 40-fold, there was no increase in the transcription of the other SusC gene, *bt4164*, **Figure 4.30**. It is possible that the different SusCs transport different forms of RGI, and are only activated in response to the precise ligand imported by the transporter.

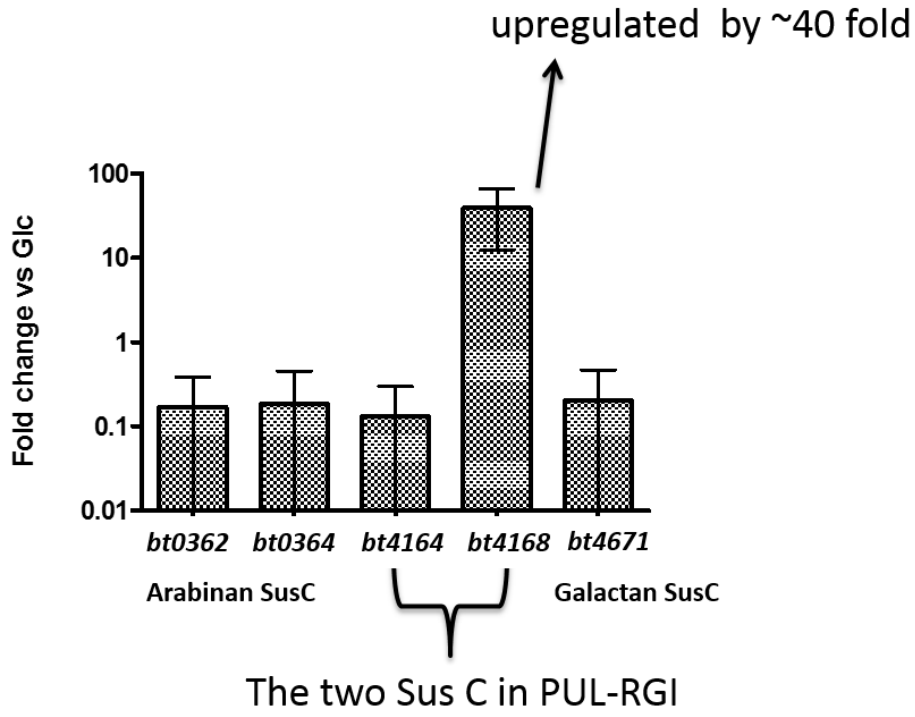


Figure 4.30 qRT-PCR of *susCs* in PUL-RGI when *B. thetaiotaomicron* was grown on RGI-AM. *bt4164* and *bt4168* are the two Sus C genes in the RGI PUL, *bt0362* and *bt0364* are the SusC genes from the arabinan PUL and BT4671 is the Sus C gene from the galactan PUL. Three replicates were performed and error bars showed the standard error.

4.2.18 Finding the binding ligands for the two SusD in the RGI PUL

BT4165 and BT4169 are the two SusD-like proteins encoded by PUL-RGI with potential roles in binding RGI and delivering these molecules to the SusC outer membrane transporters. The capacity of different RGI oligosaccharides, derived by enzymatic digestion, to bind to the two SusD proteins was evaluated by ITC (**Table 4.2**). No binding of these glycans was observed. It is possible that the SusD molecules are only able to interact with RGI molecules in complex with their SusC partners. Alternatively, given the complexity of RGI molecules from different

plant species, it is possible that the SusD proteins bind to complex forms of RGI which are distinct from RGI-PG or RGI-AM.

Ligand	BT4169(Sus D)	BT4165(Sus D)
intact mucilage	-	-
mucilage +BT4170(PL9)	-	N/A
mucilage +BT4170(PL9)+2GH105s	-	N/A
mucilage +BT4175(PL11)	-	-
mucilage +BT4175(PL11)+2GH105s	-	-
RGI potato	-	N/A
RGI potato +BT4175(PL11)	-	N/A
mucilage + U29SH (GH28)	-	-
RGI potato + U29SH (GH28)	-	-

Table 4.2 Attempt to find the right binding ligand for the two SusD (BT4165 and BT4169) in the RGI PUL: all result are negative. Left column: list of potential ligands tested by iTC. "-", No binding. N/A, not applicable.

4.3 Discussion

4.3.1 The three lyases in the same PUL

From the biochemical data presented in the results section, the most active lyase, BT4175, was shown to depolymerize both decorated and undecorated forms of RGI. In contrast, the two PL9 lyases, which are significantly less active, only target undecorated forms of the polysaccharide, **Figure 4.8** and **Figure 4.9**. The function of the two PL9 lyases is unclear. Theoretically *B. thetaiotaomicron* could deploy a single "super lyase" that displays high catalytic activity and can accommodate the myriad of side chains that decorate RGI. However, current evolutionary theory argues that the co-localization of diverse activities within an enzyme is likely to be rare because it would compromise the catalytic activity of the protein. (Montanier et al. 2009a, O'Brien and Herschlag 1999) In other words, substrate specificity contributes to enzymatic efficiency. The PL11 RGI lyase BT4175, however, is the most active lyase and yet displays more plasticity in substrate specificity than the PL9 enzymes. It should be appreciated, however, that this plasticity simply reflects an inability to recognize Rha at O4 at subsites distal to the active site. Another reason for the multiple lyases could be a consequence of the continuing

evolution of the structure of polysaccharides, and thus the gain of a new gene (in this case the lyase genes), likely through horizontal gene transfer, could confer a selection advantage (Lejohn 1971, Waltz et al. 2012, Martone et al. 2009). It is possible that the enzyme encoded by the acquired gene displays differences in specificity on substrates that were not used in this study.

It is also formally possible that the function of the different lyases reflects their cellular location. Enzymes expressed by *B. thetaiotaomicron* that contain a type I signal peptide (SP1) are believed to be located in the periplasm. Although all three RGI lyases appear to contain SP1s and are thus assumed to be located in the periplasm, the proposed cellular location of these enzymes requires confirmation by cellular localization studies. For example, if the PL11 RGI lyase, BT4175, was secreted, the enzyme would digest RGI-PG into large oligosaccharides, with cleavage restricted by both acetyl and galactose side chains. It is these oligosaccharides that are imported into the periplasm for further degradation by the PL9 RGI lyases and the repertoire of exo-acting enzymes. (Lejohn 1971, Waltz et al. 2012, Martone et al. 2009) To confirm the cellular location of the key enzymes, polyclonal antibody can be generated against them and used through methods like western blot. To summarize, the existence of three lyases with, potentially, different specificities and/or cellular locations highlights the complexity of the RGI pectin utilization system expressed by *B. thetaiotaomicron.*

4.3.2 Deploying a plethora of enzymes to digest a simple structure

The structure of RGI-AM comprises alternating units of Rha and GalA, and thus, in theory, the polysaccharide requires only two enzymes to completely break down the polysaccharide. However, evolutionary pressures imposed on *B. thetaiotaomicron* has resulted in a plethora of enzymes that contribute to the complete depolymerization of the RGI backbone, **Figure 4.31**. To explain this complexity, one reason is that by employing endo-acting lyases, the pectin

chains are cleaved into numerous oligosaccharides, providing more substrate for the exo-acting enzymes, enhancing the rate of substrate degradation. Furthermore, the RGI lyases create a unique product, 4,5ΔGalA-Rha-GalA-Rha, that induces transcription of PUL-RGI. It is the unique structure of this tetrasaccharide that confers the required specificity for the target sensor. For example, if Rha or GalA comprises the signaling molecule, such sugars can be generated from several non-RGI glycans and thus PUL-RGI could be activated in the absence of the target polysaccharide. This would be energetically unfavourable as the bacterium would synthesize a complex enzyme consortium when the appropriate substrate was not available. Since ΔGalA-Rha-GalA-Rha is a critical signaling molecule, how the tetrasaccharide is protected from the digestive machinery is an important issue. Given that the terminal 4,5ΔGalA protects the molecule from rhamnosidases and galacturonidases, slow release of this unsaturated sugar by BT4174 (GH105) would ensure that the molecule is long lived. Kinetic data in **Figure 4.15** showed that the activity of the GH105 enzyme is indeed much lower compared to its 4,5ΔGalAase counterpart, BT4176, which targets the disaccharide ΔGalA-Rha. Indeed, it would be interesting to crystalize the two GH105 enzymes to explore the functional basis for their different specificities.

Another factor that may contribute to the signal molecule protection hypothesis is the pH profile of the GH105 enzymes, **Figure 4.14**. BT4174 has a sharp pH profile while the other GH105 tolerates a wider pH range. Considering the regulatory proteins are usually located on the inner membrane with other transporters, the pumping of ions and other chemical groups may cause a pH fluctuation near the site where the signaling tetrasaccharide binds to the sensor domain of the HTCS (Sonnenburg et al. 2006, Martens et al. 2011, Chang and Stewart 1998). The narrow pH range of BT4174 may, however, restrict the degradation of the signaling

molecule to microenvironments within the periplasm, which again may confer protection from degradation.

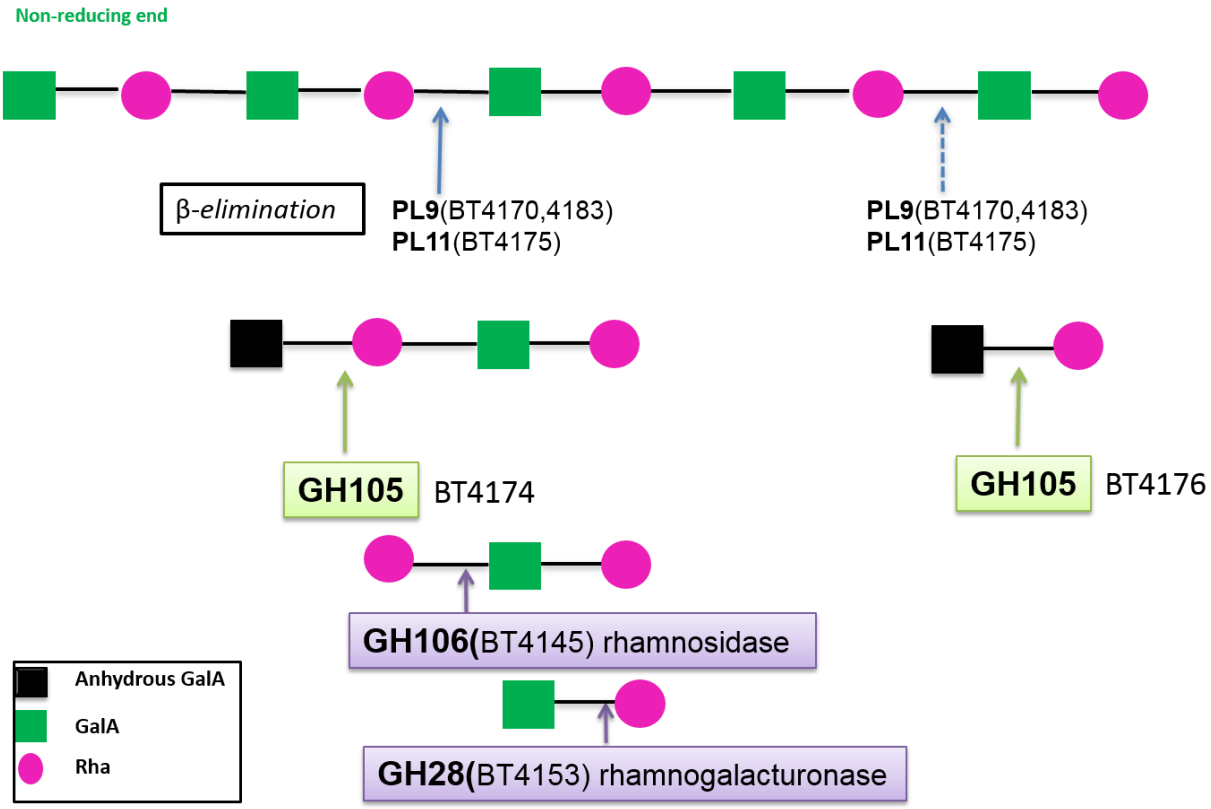


Figure 4.31 Overview of step by step degradation of the RGI backbone (RGI-AM).

4.3.3 De-branching of RGI by the Galactan & Arabinan Utilisation System (work by Briggs and McKeen in Prof Gilbert's lab in University of Newcastle)

RGI may contain a variety of side chains, principally galactan and arabinan, which can be species and cell specific (Mohnen 2008). These side chains pose additional challenges as they may prevent the backbone degrading enzymes from accessing the substrate. Work by Jonathon Briggs and Lauren McKeen have shown the mechanism by which the arabinan and galactan PULs of *B. thetaiotamicron* orchestrate the removal of these side chains from RGI.

The backbone of branched arabinan and unbranched arabinan is partially cleaved on the outer membrane by two GH43 arabinanases (BT0360 and BT0367). The products are transported into the periplasm by SusC-like porins and the side chains of the branched oligosaccharides are removed by a arabinofuranosidase (BT0369) to remove the α1,2-linked decorations and a GH51 arabinofuranosidase (BT0348) to hydrolyse the α1,3-arabinofuranose units. A periplasmic exo-acting GH51 arabinofuranosidase (BT0368) then hydrolyses the liner α1,5-linked backbone into arabinofuranose, which is then transported into the cytoplasm to be metabolized. During the digestion, a regulatory protein HTCS binds to arabino-heptasaccharide to up-regulate the Arabinan Utilisation Locus (**Figure 4.32**).

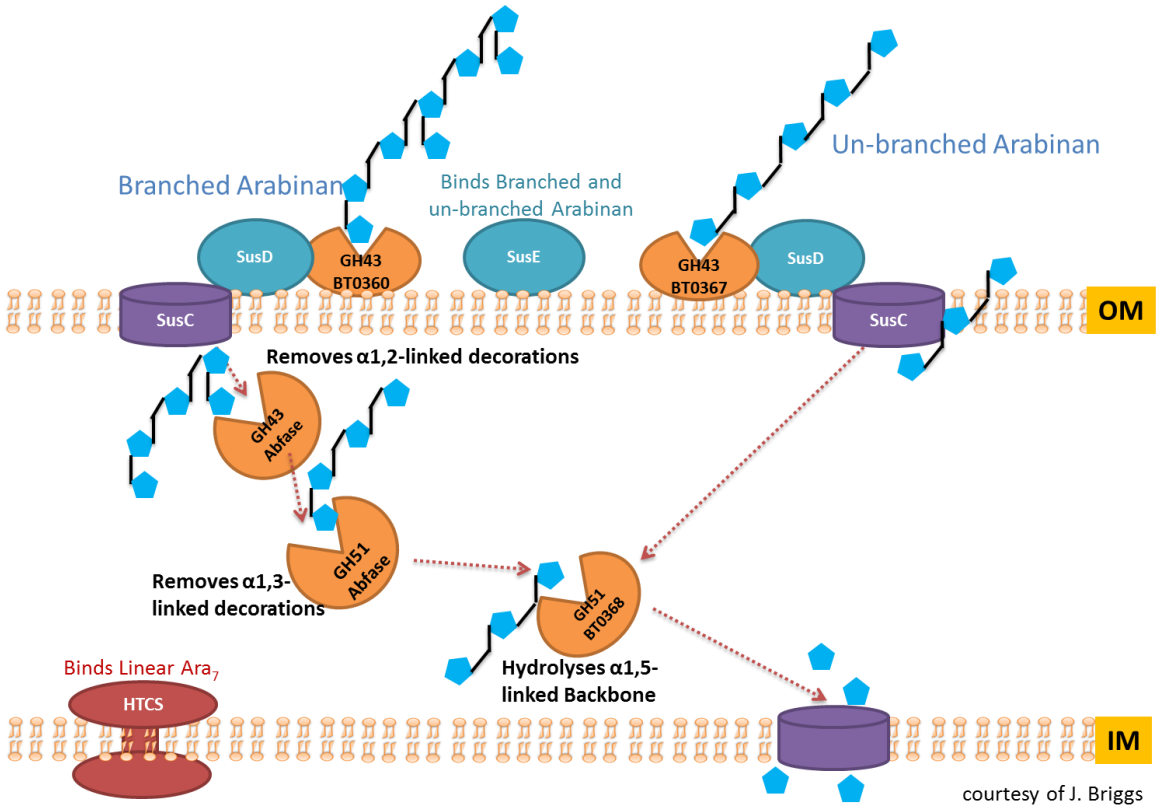


Figure 4.32 The Arabinan Utilisation System in *Bacteroides thetaiotamicron*. OM, Outer membrane; IM, inner membrane. HTCS, regulatory protein. (personal communication from J. Briggs and L. McKeen,)

Similarly, galactan is processed by the Galactan Utilisation System with the first step involving a GH53 endo-galactanase, which partially hydrolyses the polysaccharide on the outer

membrane. After being transported into periplasm, the galactooligosaccharides are hydrolyzed into galactose by an exo-acting GH2 β-galactosidase. Galactotetrasaccharide, generated during the degradation process, binds to and activates the HTCS that up-regulates the Galactan Utilisation Locus (**Figure 4.33**).

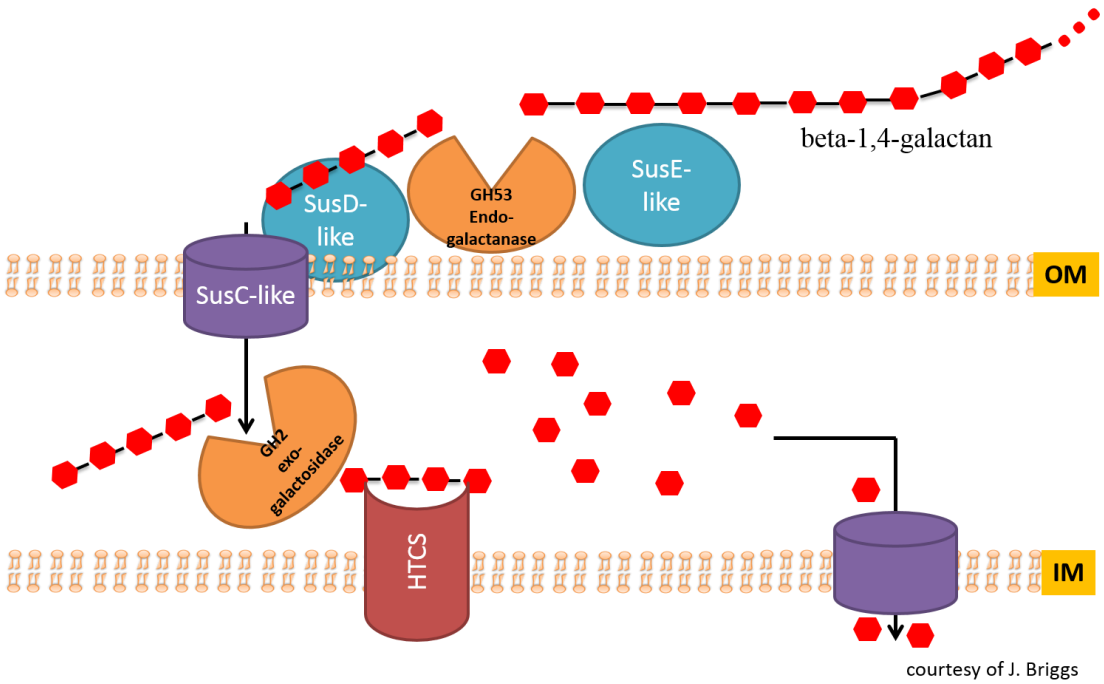


Figure 4.33 The Galactan Utilisation System in *Bacteroides thetaiotamicron*. OM, Outer membrane; IM, inner membrane. HTCS, regulatory protein. . (personal communication from J. Briggs and L. McKeen)

4.3.4 Proposed model of RGI-PG degradation.

The galactan and arabinan utilisation systems may depolymerize most of the RGI side chain, but the experiments described here on the GH2 and GH35 β-galactosidases showed that these galactan stubs remain attached to the RGI backbone, which are not removed by the galactanase or the exo-galactosidase encoded by the galactan PUL. Therefore, RGI-PG creates additional challenges by the presentation of decorations (especially the remaining galactan side chain) that prevent the backbone degrading enzymes from accessing the substrate. The enzymes required to process RGI-PG into a completely linear RGI molecule are displayed in the model,

shown in **Figure 4.34**, and is predicated on the "Enzyme cocktail knockout digestion data" described before, and the observed synergy between specific enzyme combinations.

The galactosidase GH2 and GH35 enzymes are capable of removing components of the Gal side chains, while data in **Figure 4.22** and **Figure 4.23** suggested that BT4156 (GH2) is likely to be the enzyme that removes the Gal directly linked to the backbone as it is the only galactosidase that is active against galactobiose, lactose and 4NP-β-D-Gal, revealing promiscuity in the recognition of the moiety bound at the +1 subsite. This view is supported further by the data presented in **Figure 4.22** showing that deleting this enzyme from the GH105 and PL-free enzyme cocktail had a much larger effect on the release of Gal and Rha, compared to deletion of the other galactosidases.

During the enzyme activity screening, one protein that is not categorized within CAZy, BT4158 was shown to affect the release of Rha, **Figure 4.25**. The enzyme was shown to hydrolyse 4NPA and is thus an acetyl esterase. It is therefore speculated that the backbone of the RGI is acetylated (likely at the GalA residue) as shown previously (Ridley et al. 2001, Mohnen 2008), and the removal of the acetyl groups is likely required for the enzymes, such as the rhamnosidase and RGI galacturonase to attack the backbone.

After removing the remaining galactan side chain, previous data in **Figure 4.8** revealed that the pectic lyases PL9 (BT4170) could then digest on RGI-PG. The esterase also helps to produce a smooth backbone of RGI which will be targeted by the orchestrated enzymes in **Figure 4.31**.

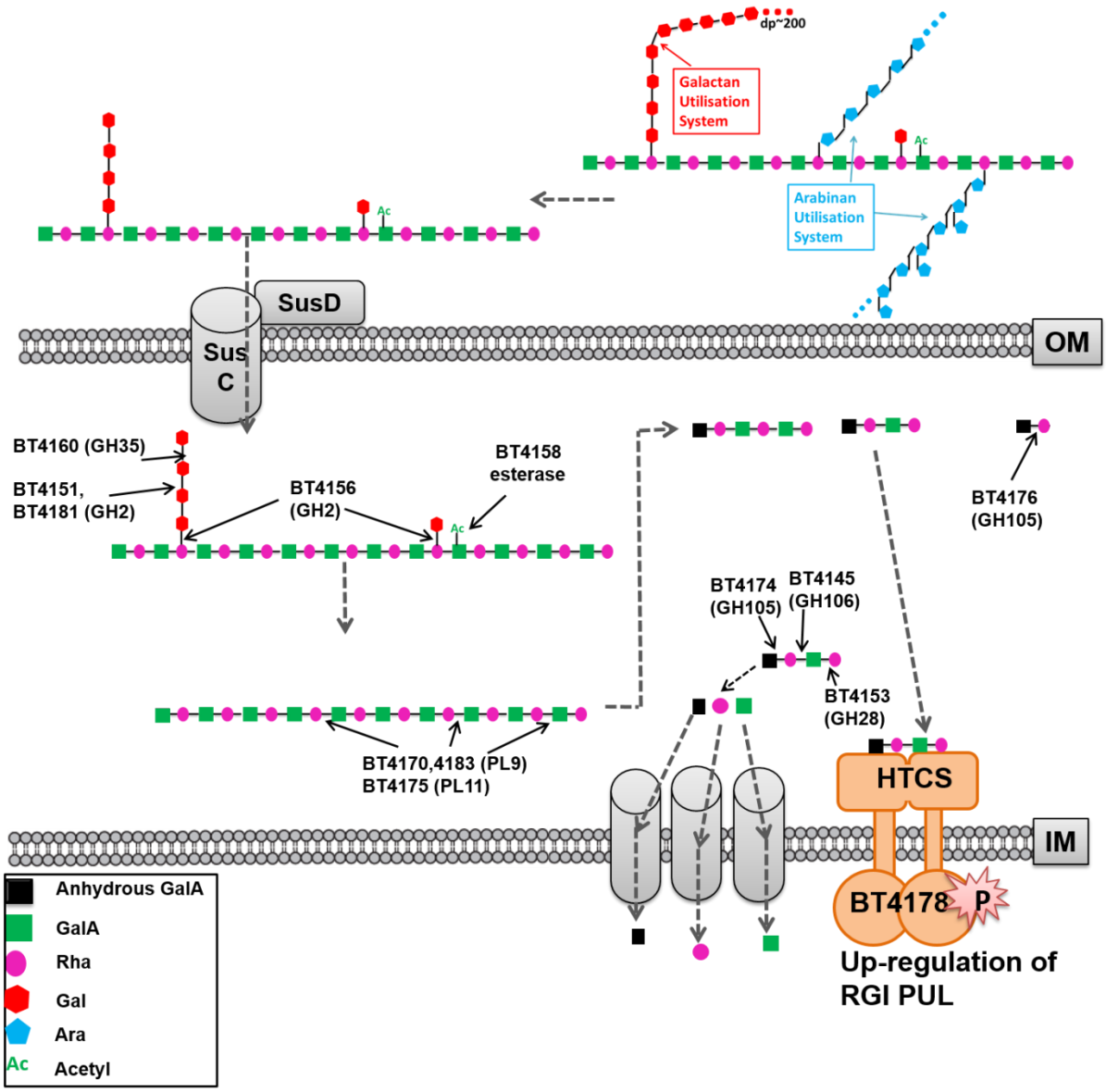


Figure 4.34 The final model of RGI utilization within the context of *Bacteroides thetaiotaomicron* membranes. HTCS, Hybrid Two Component System (sensor); OM, outer membrane; IM, inner membrane; Ac, acetyl.

4.3.5 Bespoke RGI-derived oligosaccharides

Through this project, methodology has been established and the enzymes generated, that enable RGI to be hydrolyzed delivering a Rha, GalA or ΔGalA at the non-reducing end of oligosaccharides with a defined d.p.(degree of polymerization) Thus, several RGI lyases can produce oligosaccharides with a terminal ΔGalA, while the RGIase can be used to produce

molecules with a Rha at the non-reducing end. Two other enzymes from the RGI PUL, the rhamnosidase BT4145 and the rhamnogalacturonase BT4153 can be used in a step by step digestion to further modify the oligosaccharides with a defined d.p.(degree of polymerization), which can be purified by size exclusion chromatography. This study provides the methodology and reagents that can be used to generate defined oligosaccharides whose importance in microbiota and human nutrition can be evaluated.

4.3.6 The growth profile of *Bacteroides thetaiotamicron* and up-regulation

Figure 4.29 shows that the order of branched pectin utilization by *B. thetaiotamicron* was as follows: RGII>RGI-PG>RGI-AM>>RGI from soybean. A possible reason for the preference for RGII is the great diversity of sugars in the polysaccharide making it more "nutritious". Furthermore, the RGII used in this project, which was extracted from red wine, was highly soluble compared to RGI-PG and RGI-AM. The structure of RGI soybean possibly made it difficult for the bacterium to utilize, since the Megazyme product brochure showed that its content of GalA, Rha and Gal was 51.0%, 6.3% and 12.3%, respectively. Other sugars such as fucose and xylose were also present in the polysaccharide individually at a level >12%, and it is possible that PUL-RGI does not have the enzymes required to remove the chemically heterogeneous side chains.

Figure 4.30 showed that when *B. thetaiotamicron* was growing on RGI-AM, only one of the two SusC genes was up-regulated (~40-fold). The data suggest that BT4168 (SusC) is involved in the transport of de-branched RGI, exemplified by the RGI-AM, while the other SusC, BT4164, might only transport the more branched or intact RGI structure, depending on the source or plant species in the diet. Similar experiment on RGI-PG may provide extra information on this point but the impurity of RGI-PG may be an issue. Mutant *Δbt4178* may be used to study the up-regulation of the two SusC genes but the data in **Figure 4.29** suggested

that the knockout only delayed the growth curve and the bacterium managed to compensate for the loss of the HTCS gene.

4.4 Future work

Figure 4.35 highlights the RGI degrading enzymes currently identified in this work. The enzymes with question marks remain to be characterized include BT4146 and 4149 (GH28), BT4152 and BT4157 (GH27), and some putative esterases. It is possible that the enzymes with unknown activities target unusual RGI structures, such as RGI from flax seeds or other uncommon food sources. Therefore, in the future more types of RGI should be purified and tested against the enzymes encoded by the RGI PUL. A possible function for BT4155 (GH28) that does not attack RGI directly is that they may target the junction between RGI and homogalacturonan. However, a characterized substrate containing a highly defined junction sequence, such as GalA-GalA-Rha-GalA is not currently available. Further research can be focused on this topic.

It would be very interesting to continue the structural biology of selected RGI enzymes. For example, the structure of the three lyases in the RGI PUL can be examined and compared to better understand the mechanism by which these enzymes display their different specificities. Similarly, the crystal structures of the two GH105 enzymes (BT4174 and BT4176) would enable us to understand how these enzymes have evolved to display high selectivity for di and tetrasaccharides.

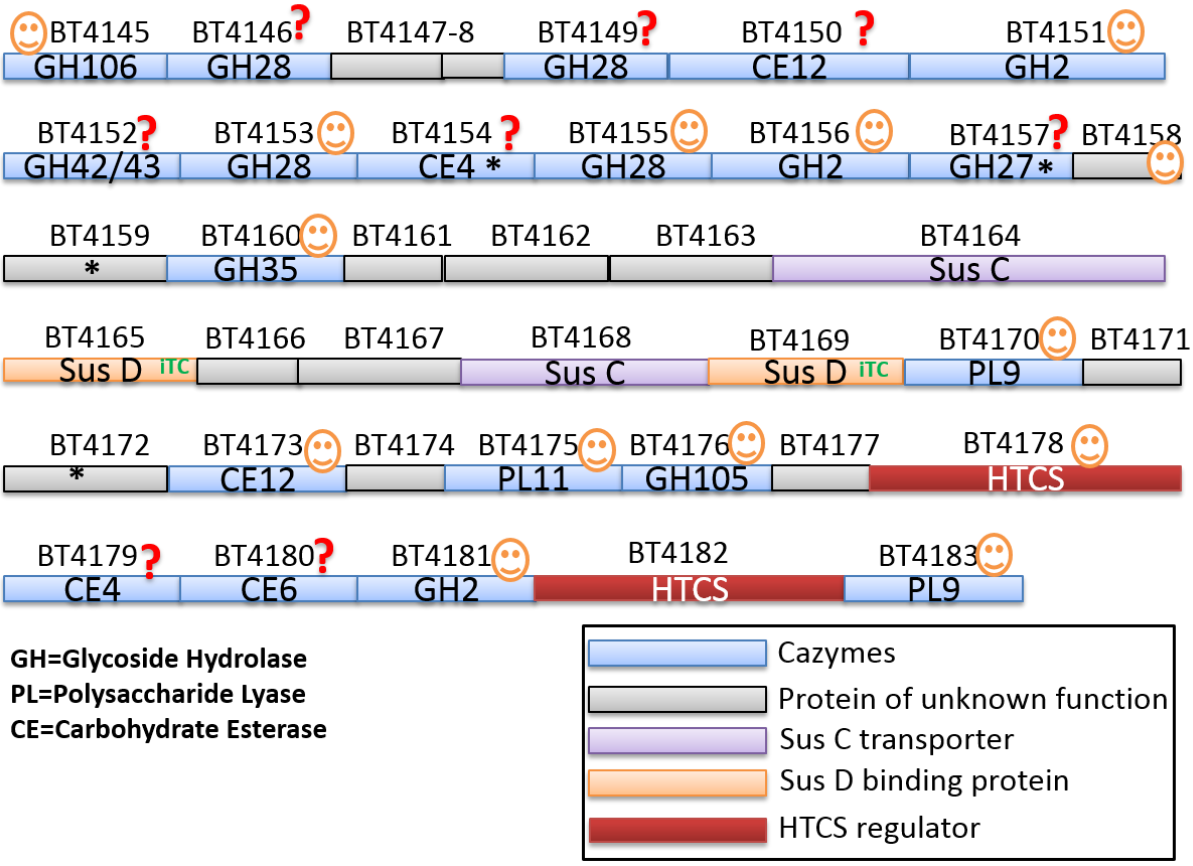


Figure 4.35 Overview of the characterized enzymes in the map of the RGI PUL (polysaccharide utilization loci) of the *Bacteroides thetaiotamicron*. Smiley faces mean the enzyme activity has been identified while question marks denote enzymes that have not been characterized. Asterisk shows insoluble protein in *E.coli*. Green "iTC" label the two SusD proteins which displayed no binding to RGI or RGI-derived oligosaccharides.

Other interesting work is to test the esterase BT4158 against a range of acetylated monosaccharides or polysaccharides to see if it is specific to RGI or a non-specific acetyl esterase. Characterization of other enzymes encoded by the RGI PUL is also very important, such as the GH2 and GH35 β-galactosidases.

It is also intriguing to evaluate if 4,5ΔGalA is crucial for the binding of the sensor BT4178 to and the signaling tetrasaccharide. Removal of the 4,5ΔGalA from the oligosaccharide by the GH105 enzyme abolished binding to the sensor domain **Figure 4.27**. However, a more convincing way to test this is to create a new version of tetra oligo, with a normal GalA at the

non-reducing end (namely GalA-Rha-GalA-Rha) and test it against the sensor protein. To achieve this, a scale-up experiment is required to produce this oligosaccharide in a carefully designed step by step digestion of large oligosaccharides produced by the polysaccharide lyases.

Chapter 5. Enzymatic degradation of rhamnogalacturonan II by *Bacteroides thetaiotaomicron*

5.1 Introduction

Rhamnogalacturonan II (RGII) consists of 13 different monosaccharides and over 20 linkages, and is regarded as the most complex carbohydrate in nature. (Mohnen 2008). This pectic polysaccharide contains an α-1,4 linked homogalacturonan backbone and five distinct side-chains, which vary in the degree of polymerization, branching, and sugar chemistry, **Figure 5.1**. RGII is very resistant to microbial digestion due to its complexity exemplified by red wine. During the fermentation nearly all polysaccharides from the grape cell wall are digested, leaving RGII, being highly recalcitrant, as the dominant glycan in the red wine. It is estimated that in red wine the RGII content is 100 - 400 mg/L. (Perez et al. 2003)

RGII is one of the most conserved polysaccharides throughout the plant kingdom, and has been shown to be critical for plant growth and development (Ahn et al. 2006, O'Neill et al. 2001, Ishii et al. 1999). Within the cell wall two apiose subunits from distinct RGII molecules are cross-linked by borate, leading to the formation of higher order pectin structures.

Enzymes that are active against RGII have not been reported. Unlike cellulose, whose recalcitrance is due to its physical properties, RGII is believed to resist digestion because of its inherent chemical complexity, the presence of rare sugars and linkages, and borate cross-linking of the glycan. Thus, deconstruction of RGII requires the assembly of an extensive array of enzyme activities.

The observation (Martens et al. 2011) that *Bacteroides thetaiotaomicron* can grow on RGII has provided the first insights into the bacterial processing of the glycan. This phenotype indicates that the *B. thetaiotaomicron* genome encodes a specialized set of enzymes dedicated to RGII degradation and utilization. Transcriptomic profiling of cultures of *B. thetaiotaomicron* grown on RGII identified a Polysaccharide Utilization Locus (PUL) that is expressed in response to the glycan (Martens et al. 2011). This RGII PUL consists of ~60 genes, encoding an estimated 32 glycan modifying enzymes, based on their sequence similarity with other proteins categorized by the CAZy database (http://www.cazy.org/), transporers and regulatory proteins, (see **Table 5.1**). The high density of, potentially, novel enzymes encoded by this PUL suggests that either a high proportion of the linkages in RGII are cleaved, or there is a high level of functional enzymatic redundancy. Furthermore, it is unclear whether the RGII is degraded by an endo, exo or endo-exo mechanism. The objective of the project is to characterize the enzymes encoded by the RGII PUL which could lead to the possible establishment of model for RGII.

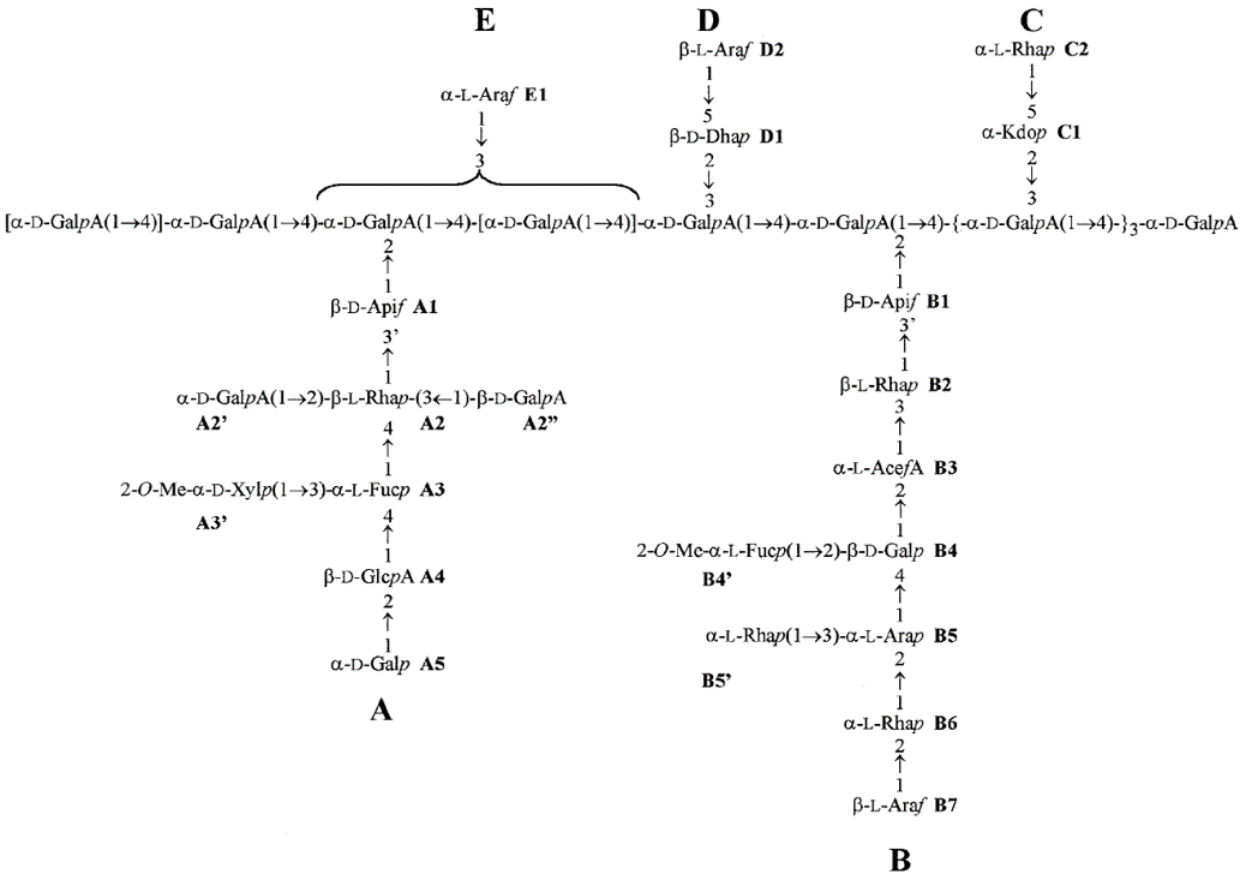


Figure 5.1 A schematic structure of rhamnogalacturonan II (RGII). Examples of abbreviations: Ara*f*, arabinose furanose; Rha*p*, rhamnose pyranose. Size chains are labeled in bold. E1, side chain E position 1.

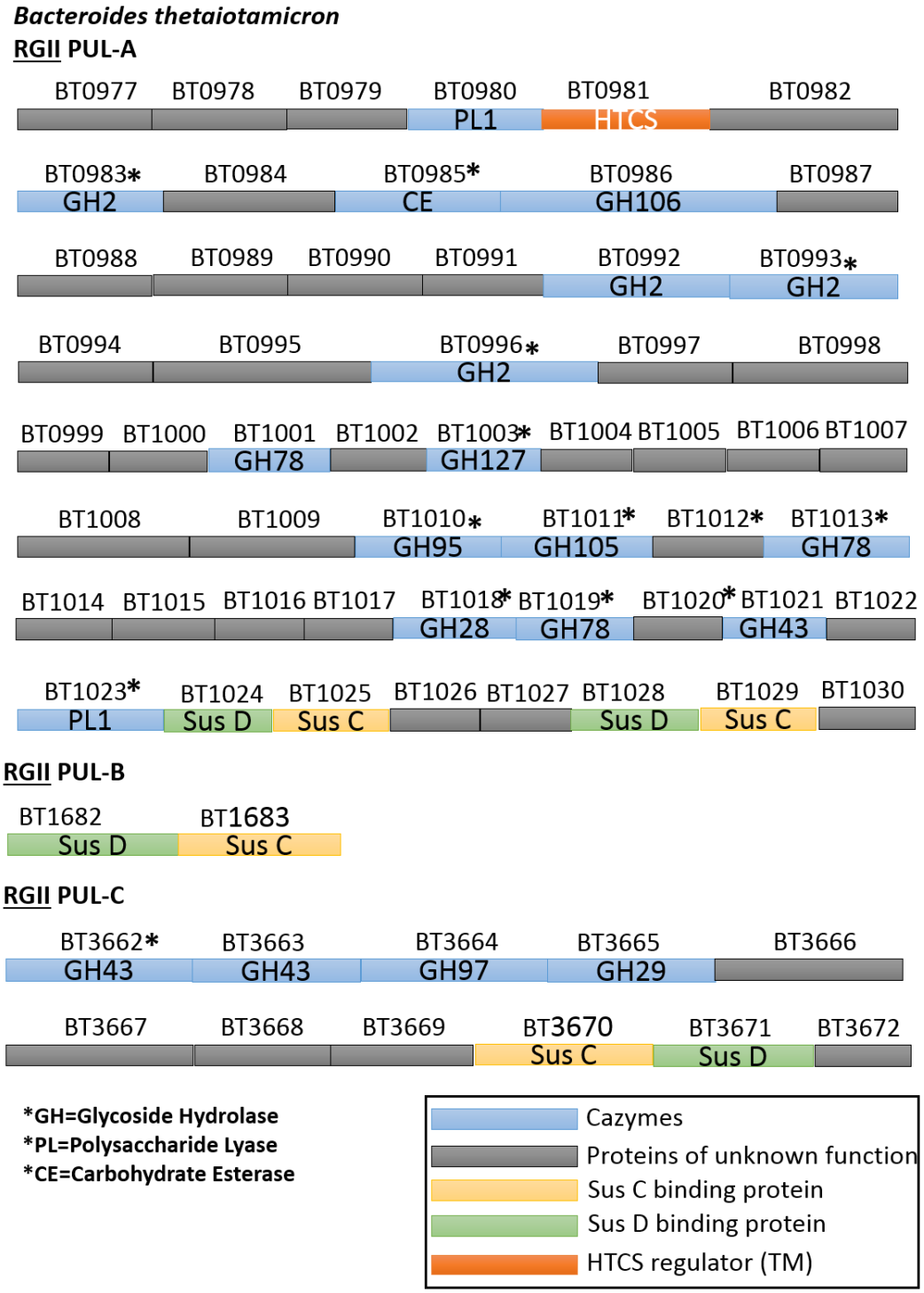


Figure 5.2 Map of the RGII PUL (polysaccharide utilization loci) of the *Bacteroides thetaiotamicron*. Asterisks label enzymes that were studied in this thesis.

5.2 Results

5.2.1 Gene cloning and protein expression

Production of enzymes encoded by RGII PULs.

RGII activated three *B. thetaiotaomicron* PULs, with RGII PUL-A being the major one, with the other two loci defined as RGII PUL-B and RGII PUL-C. So far 32 genes from RGII PUL-A and six genes from RGII PUL-C have been amplified by PCR and cloned into the *E. coli* expression vectors pET28a or pET32b by Dr Yanping Zhu and NZYTech (1649-038 Lisboa, Portugal) (**Table 5.1**). Nearly half of the proteins encoded by these genes are located in CAZy families (http://www.cazy.org/).

In this project the expression of all 32 genes from RGII PUL-A in *E. coli* was assessed, and those that did not generate detectable recombinant protein (four genes) were re-sequenced. As a result these genes were re-amplified by PCR and cloned into expression vectors. In total, protein expression experiments showed that 29 of the 32 genes generated soluble protein in *E. coli.* As these proteins all containing a His_6-tag they were purified by IMAC (see **Figure 5.3** for examples of expression and IMAC purification of selected proteins, and the final purification of all the proteins) with the three insoluble proteins being the N-terminal domain of BT1013 (GH33), the C-terminal domain of BT0979 (a potential lyase) and BT1026.

GHs	**CAZy**	**Function predicted by GenBank**	**Cloned in pET28a or pET32a**	**Soluble/insoluble expression**	**Signal peptide prediction**	**Predicted size (kDa)**	**Experimental size (kDa)**
BT0983	GH2	beta-galactosidase	pET28a	soluble	SpI	97	97
BT0986	GH106	alpha-rhamnosidase	pET28a	soluble	SpI	122	120
BT0992	GH2	beta-galactosidase	pET28a	soluble	SpI	108	100
BT0993	GH2	beta-galactosidase	pET28a	soluble	SpI	124	116

BT0996NT	GH2	beta-galactosidase	pET28a	soluble	SpI, SpII	52	52
BT0996CT			pET28a	soluble	SpI, SpII	119	119
BT1001	GH78	alpha-L-rhamnosidase	pET28a	soluble	SpI	81	81
BT1010	GH95	alpha-L-fucosidase	pET28a	soluble	SpI, SpII	93	90
BT1011	GH105	unsaturated rhamnogalacturonside hydrolase	pET28a	soluble	SpI, SpII	45	40
BT1013NT	GH78	alpha-L-rhamnosidase	pET28a	insoluble	SpI	97	insoluble
BT1013CT	GH33	sialidase	pET28a	soluble	SpI	45	45
BT1018	GH28	endo-polygalacturonase	pET28a	soluble	SpI	50	50
BT1019	GH78	alpha-L-rhamnosidase	pET28a	soluble	Neither	105	105
BT1021	GH43	arabinosidase	pET28a	soluble	SpII	33	37
BT1003	GH127	β-L-arabinofuranosidase	pET28a	soluble	SpI	78	68
PLs							
BT0980	PL 1	pectate lyase	pET28a	soluble	SpI, SpII	70	70
BT1023	PL 1	pectate lyase	pET28a	soluble	SpI	75	75
ORFs							
BT0979NT		pectin lyase fold	pET28a	soluble	SpI	57	59
BT0979CT		C-term unknown fold	pET28a	insoluble	SpI	77	insoluble
BT0984		GH95-like	pET28a	soluble	SpI	89	89
BT0985		putative sialyate-O-acetylesterase	pET28a	soluble	SpI, SpII	61	61
BT0997		hypothetical protein	pET32a	soluble	SpI, SpII	100	100
BT1002		Pectin Lyase	pET28a	soluble	SpI	80	80
BT1012		hypothetical protein	pET28a	soluble	SpI	56	56
BT1017		hypothetical protein	pET28a	soluble	SpI	74	70
BT1020NT		hypothetical protein	pET28a	soluble	SpI	70	70
BT1020CT		hypothetical protein	pET32a	soluble	SpI	54	54
BT1022		outer Membrane Protein	pET28a	soluble	SpI, SpII	25	25
BT1026		hypothetical protein	pET28a	insoluble	SpI, SpII	62	insoluble
BT1030		hypothetical protein	pET28a	soluble	SpI, SpII	54	54
BT1031		xylose isomerase-like	pET28a	soluble	SpI	34	34
BT0981(-NT)		HTCS	pET28a	soluble	SpI	101	101
BT0988		Mg^{2+} transport ATPase protein	pET28a	soluble	Neither	100	100
outside the RGII PUL A							
BT3661		GH97	pET28a	soluble	SpI	71	71
BT3662		GH43	pET28a	soluble	SpI	50	50
BT3663		GH43	pET28a	soluble	Neither	52	52
BT3665		GH29	pET28a	soluble	SpI	61	61
BT3671		SusD	pET28a	soluble	SpII	75	75
BT3672		hypothetical protein	pET28a	soluble	SpII	34	34

Table 5.1 Genes cloned in the RGII project. Predicted activities by GenBank and where appropriate the CAZy family also shown. GHs, Glycoside Hydrolases; PLs, Polysaccharide Lyases; ORFs, Open Reading Frames; CAZy, Carbohydrate-Active enZYmes database developed by the Glycogenomics group at AFMB in Marseille, France (http://www.cazy.org/); HTCS, Hybrid Two Component System; the prediction of signal peptides used the LipoP 1.0 Server; SP1, type I signal peptide; SPII, type II signal peptide (http://www.cbs.dtu.dk/services/LipoP/).

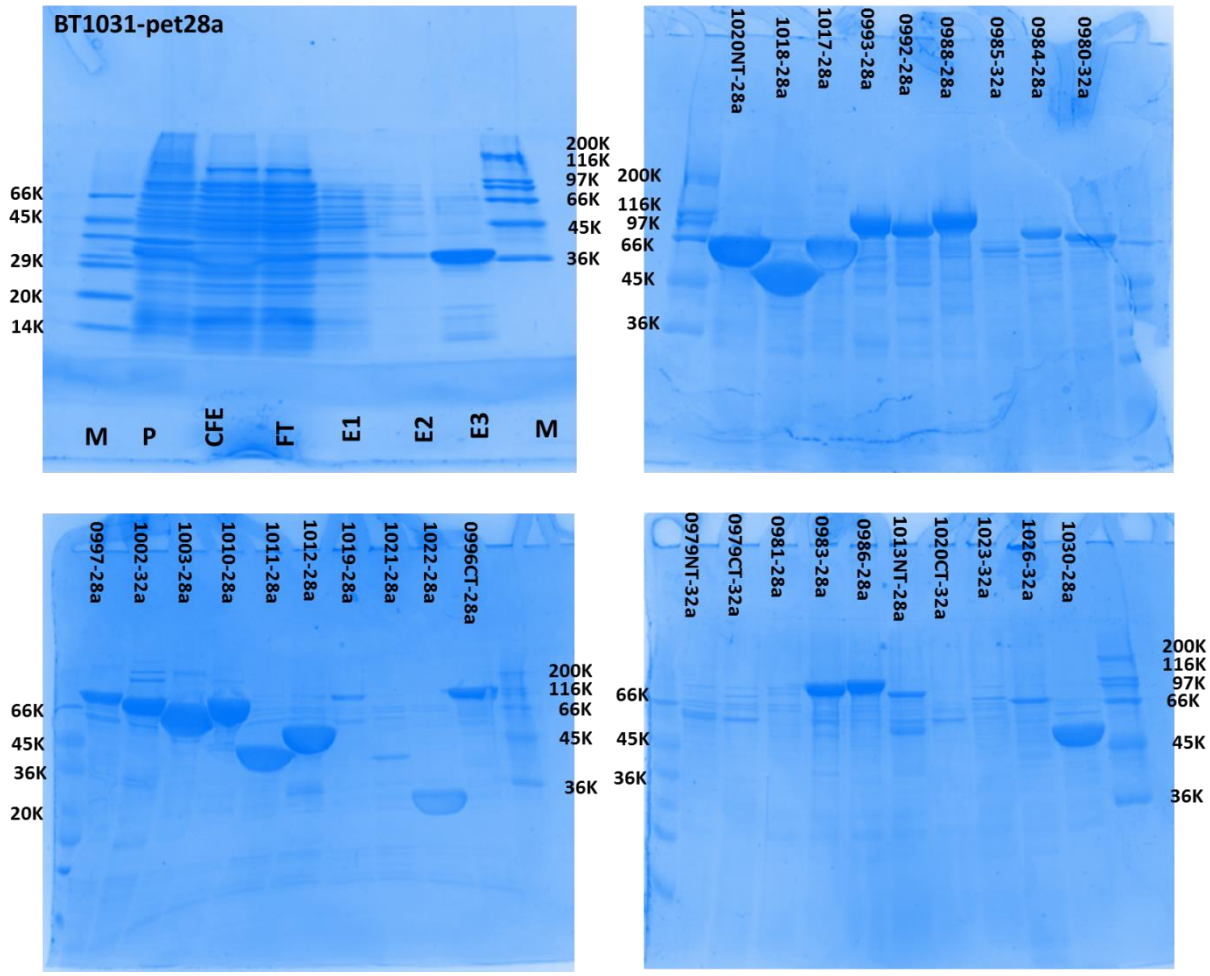


Figure 5.3 Examples of IMAC purified recombinant enzymes used in the RGII project subjected to SDS-PAGE. Protein expression and purification was as described in Chapter 2.2. The observed size of the individual purified proteins was consistent with their molecular mass predicted from amino acid sequence. The proteins were identified from their genomic locus and the vector into which the corresponding gene was cloned. For example "BT1020NT-28a" signifies the protein comprises the N-terminal domain of BT1020 and the encoding gene was inserted into pET28a. M, Marker; P, insoluble pellet; CFE, Cell free extract; E1-3, Elutions by imidazole gradient of 0 mM, 5 mM and 150 mM respectively in Talon buffer.

5.2.2 Enzymatic degradation of RGII by *Bacteroides thetaiotaomicron*

Degradation of RGII by an enzyme cocktail and individual enzymes

A mixture of all the soluble recombinant enzymes (see **Table 5.1)** was incubated with the RGII and the products released were analyzed by TLC and HPLC. The data showed that Rha, Gal, Ara and GlcA were released (**Figure 5.4**).

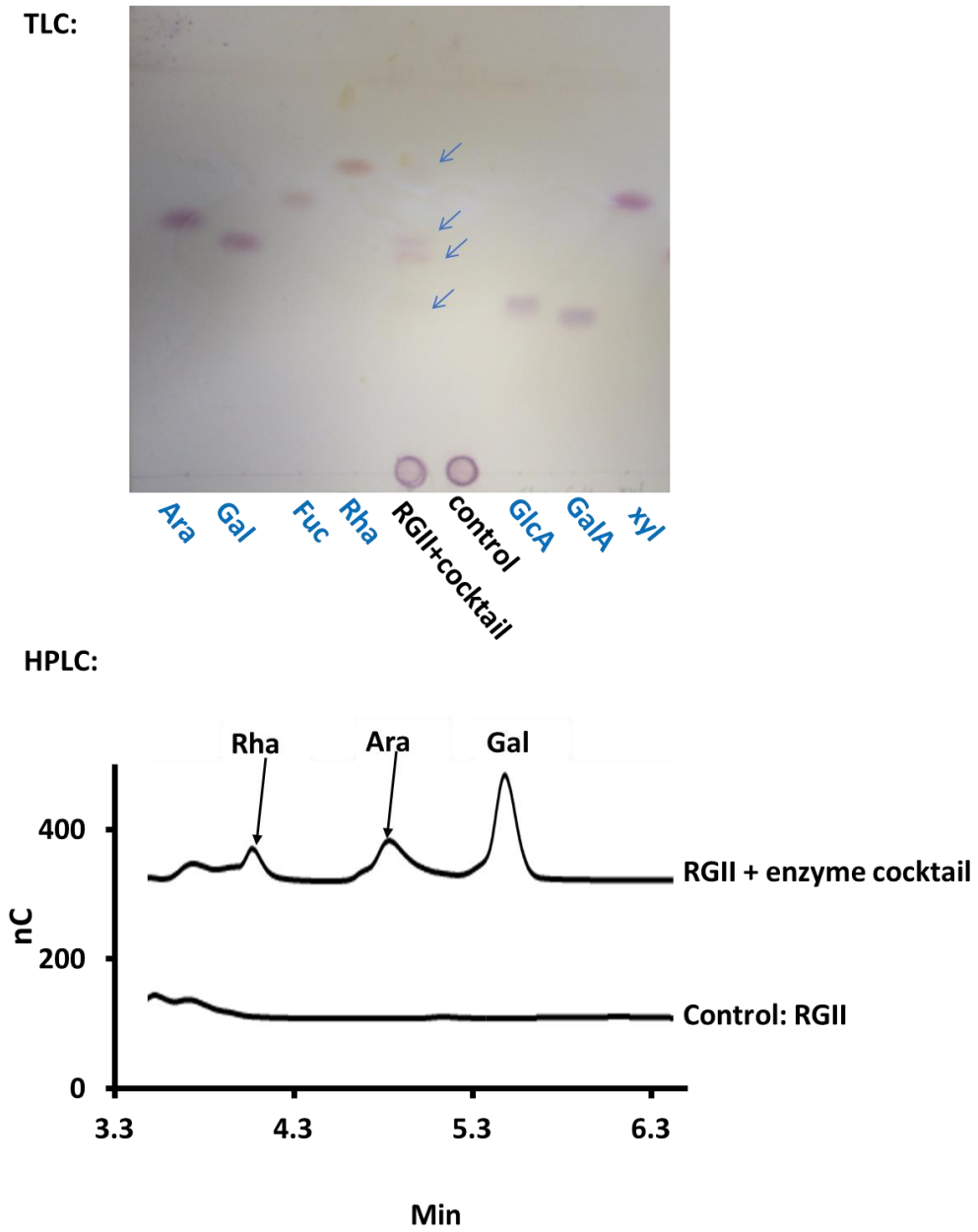


Figure 5.4 Digestion of RGII by the enzyme cocktail. The reaction was carried out at 37 ^{o}C overnight, 50 mM Na-HEPES buffer pH 7.0; RGII was 0.5 mg/ml and each enzyme was at 1 µM. Products were analyzed by TLC (products labeled by arrows) and HPLC.

Experiments of individual enzyme incubations showed that apart from two enzymes, BT3662 (GH43) and BT1010 (GH95), all other proteins displayed very limited activity against RGII. For example, HPLC data showed the release of a small quantity of galactose by BT0983 (GH2) or BT0985, and a small quantity of rhamnose by BT1019 (GH78) and BT0988, **Figure 5.5** and **Figure 5.6**. The release of galactose is likely to be from the B4 position on side chain B as Dr Marie C Ralet in INRA (with whom we collaborated to purify RGII) had unpublished data that showed that part of the side chain B may not be intact in the RGII. Therefore some galactose can be cleaved as an exposed external sugar and this could explain the low yield of the sugar. The released rhamnose can be from either side chain B or C as shown in **Figure 5.6**.

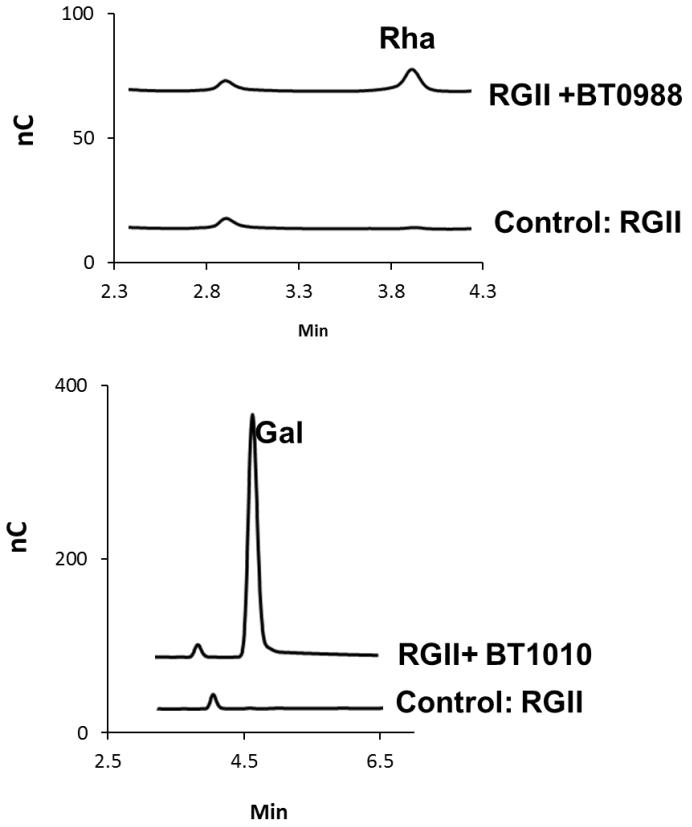


Figure 5.5 HPLC examples of an enzyme with low activity against RGII and an enzyme with high activity. The reaction was carried out at 37 ºC overnight, 50 mM HEPES buffer pH 7.0; substrate concentration, 0.5 mg/ml and enzyme concentration, 1 µM.

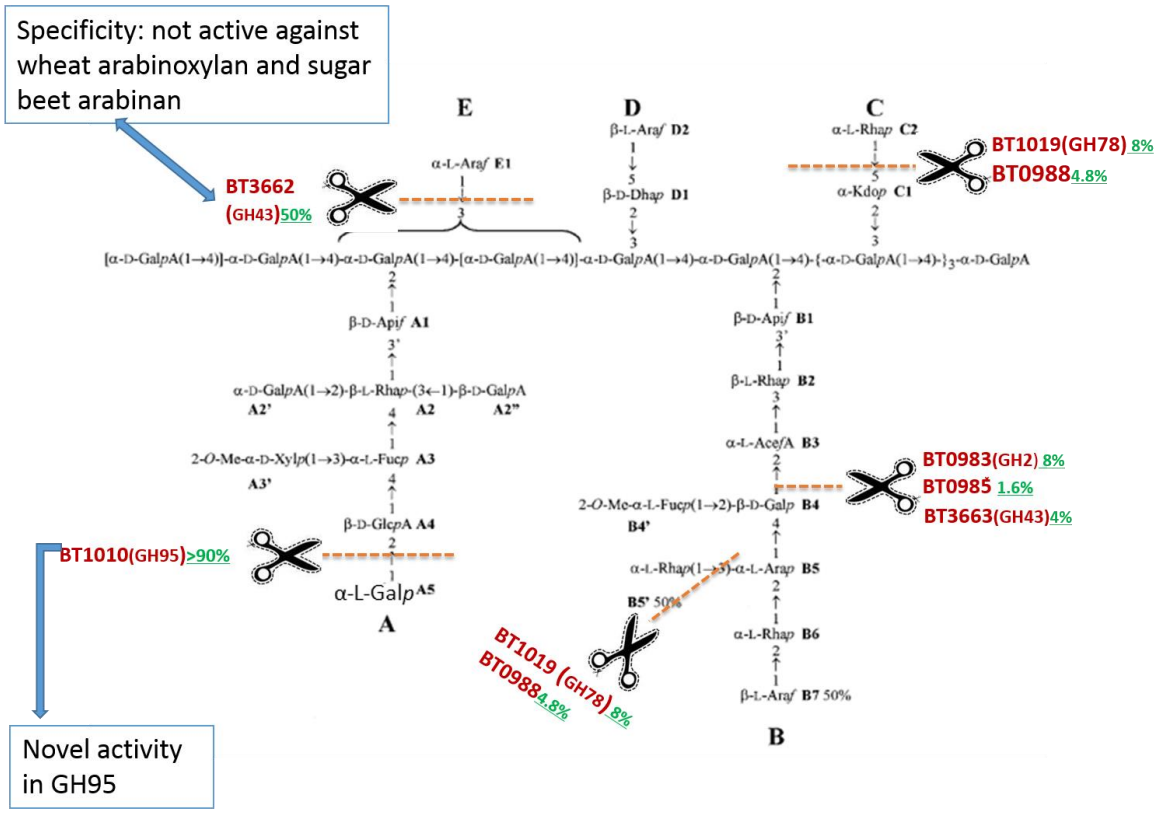


Figure 5.6 Enzymes that are active against RGII. The green percentage shows how much of the total target bond was cleaved by the individual enzyme, at 1 µM concentration after 16 h of incubation with 1.5 mg/ml RGII in 50 mM Na-HEPES buffer pH 7.0 (the calculation is based on the HPLC quantification of the released products and an estimated molecular weight of 8 KD for the substrate RGII). The products were identified by HPLC. The specificity of the enzymes that were highly active against RGII, BT3662 and BT1010, are highlighted in the blue box.

BT1010 is highly specific for RGII

Of the two enzymes that displayed the highest activities against RGII, BT1010 generated Gal from intact RGII, which suggested a novel specificity for a CAZy family as the protein belongs to GH95, a family that hitherto displayed exclusively fucosidase activity. The stereochemistry of the released Gal was confirmed as the L-form by the inability of D-galactose dehydrogenase to oxidise the sugar. The enzyme removed virtually all the α-L-Gal from the terminus of RGII Chain A, **Figure 5.6** . Recent work by Dr Rogowski (Newcastle University) has shown that a second GH95 enzyme, from *B. ovatus*, displays α-L-galactosidase activity. The *B. ovatus* enzymes was active against both RGII and corn xylan, while BT1010 was specific for the pectic

polysaccharide (**Figure 5.7**). The specificity of BT1010 for RGII is further illustrated by its lack of activity against RGI in flax mucilage (**Figure 5.8**), the only other glycan known to contain L-Gal; in the flax polysaccharide L-Gal is attached O3 to the non-reducing terminal Rha.

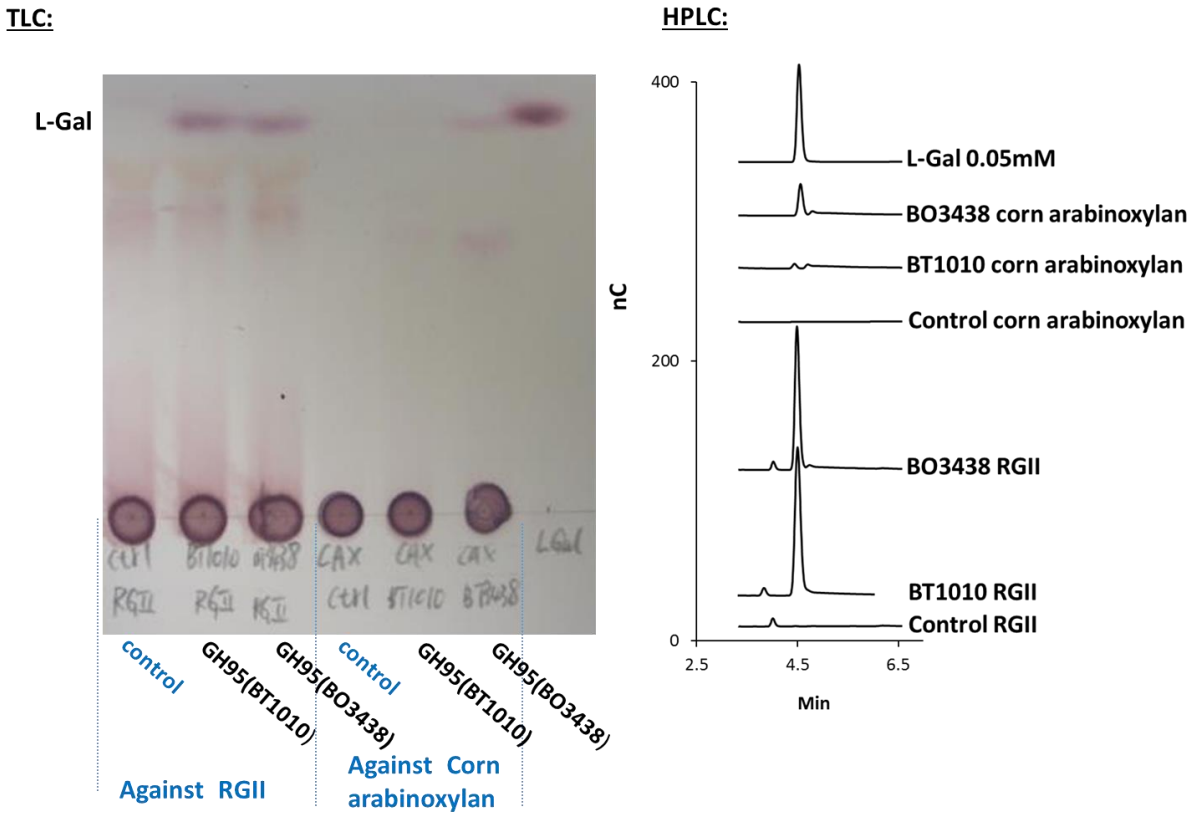


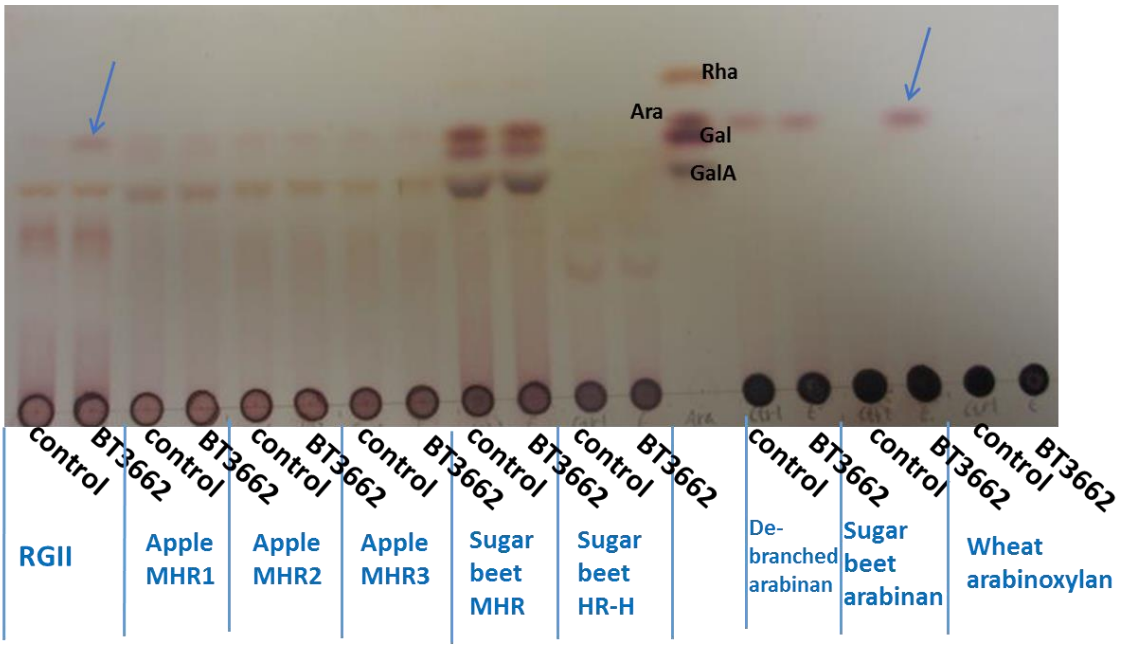
Figure 5.7 The specificity of BT1010 (GH95) and BO3438 (GH95) against RGII and corn arabinoxylan. The reaction was carried out at 37 ^{o}C in 50 mM sodium phosphate buffer pH7.0 for 16 h; substrate concentration, 3 mg/ml and enzyme concentration, 1 µM.

Figure 5.8 TLC images of the specificity of BT1010 (GH95) against RGII and flax mucilage (containing slightly branched RGI backbone with the rare sugar L-galactose attached at the O-3 position of the rhamnosyl residues) at pH 7.0. The 16 h reaction was carried out at 37 ºC, 50 mM sodium phosphate buffer pH 7; substrate concentration, 3 mg/ml and enzyme concentration, 1 µM.

BT3662 is also highly specific for RGII

BT3662 released Ara from RGII as shown in **Figure 5.9**. Because BT3662 is categorized as a GH43 α-L-arabinofuranosidase and only side chain E has a terminal α-L-Ara linkage, it is very likely that this enzyme specifically targets the side chain E of RGII. The activity of the enzyme against a number of other glycans was evaluated. The data presented in **Figure 5.9** showed that BT3662 does not cleave apple RGI, wheat arabinoxylan or debranched arabinan, however, the enzyme displayed activity against intact, and thus highly branched, sugar beet arabinan. Quantitative assays, however, showed that BT3662 was ~30-fold more active against RGII

than sugar beet arabinan, **Table 5.2**, indicating that the enzyme displays significant specificity for RGII.



Sugar analysis of the substrates from INRA, Nantes (%)

	GalA	Rha	Fuc	Ara	Xyl	Man	Gal	Glc	DM	DA
Sugar beet										
MHR	40.4	15	0	17.5	0	0	19.8	5.6	27	44
HR-H	41.3	30.8	0	1.4	0	0	25.2	1.3	0	0
Apple										
RG-I-MHR1	42.3	17.4	3.4	14.7	11.1	1.2	8.1	1.8	16	8
RG-I-MHR2	50	20.1	4.6	10.3	6	0	7.2	1.9	17	5
RG-I-MHR3	52.2	18.5	3.8	9.4	10.3	0	5.8	0	0	0

Figure 5.9 Screening the activity of BT3662 (GH43) against different glycans. The products are labelled in blue arrows; the reaction was carried out at 37 ºC, in 20 mM sodium phosphate buffer pH 7 for overnight; substrate concentration, 5 mg/ml and enzyme concentration, 1 µM; "MHR", modified hair region of RGI from INRA, Nantes.

Enzyme concentration	control	10 nM	100 nM	1 µM
RGII	0.90	3.85	31.85	100.00
Sugar beet arabinan	0.06	0.18	1.02	2.11

Table 5.2 Comparison of the relative activity of BT3662 (GH43) against RGII and sugar beet arabinan. Enzyme activity at 1µM against RGII was defined as 100%. Assays were carried out by using the Lactose-Galactose Assay Kit from Megazyme. Experiment was carried out in 50 mM HEPES buffer pH 7 with substrates concentration of 1 mg/ml.

Screening enzyme activities using aryl-glycosides

p-Nitrophenyl (PNP) conjugated glycosides are commercially available and offer a rapid method to screen for the possible activities of enzymes that contribute to RGII degradation. Substrate hydrolysis can easily be monitored by the release of 4-nitrophenolate that has an intense yellow colour, however, these aryl-glycosides provide no insight into the role of the sugar located immediately downstream of the glycosidic oxygen (i.e. the sugar bound at the +1 subsite). The following glycosides conjugated to PNP were used to screen the enzymes encoded by the RGII PULs: α-L-rhamnopyranoside, α-D-galactopyranoside, β-D-galactopyranoside, α-L-fucopyranoside, β-L-fucopyranoside, α-D-glucopyranoside, β-D-glucopyranoside, α-L-arabinofuranoside, β-L-arabinofuranoside. The results showed that BT0993 (GH2) and BT3661 (GH97) were highly active against PNP-β-D-galactopyranoside and PNP-α-D-galactopyranoside, respectively. A non-CAZy enzyme, BT0988, (predicted to be a Mg^{2+} transport ATPase protein) was moderately active against PNP-α-L-rhamnopyranoside. Weak activities were also observed for BT0986 (GH106) and BT3663 (GH43), which showed activity against PNP-α-L-rhamnopyranoside and β-D-galactopyranoside, respectively, after overnight incubation. These data provided preliminary insight into the possible linkages targeted by the potential enzymes, for example, the galactosidase could target the galactosidic linkage in RGII. The activities of these enzymes are

also influenced by the sugar at the +1 subsite which may explains the low activity of BT0988, for example.

	Active substrate	Activity
BT0993 (GH2)	PNP-β-D-galactopyranoside	+++
BT3661 (GH97)	PNP-α-D-galactopyranoside	+++
BT0986 (GH106)	PNP-α-L-rhamnopyranoside	+
BT3663 (GH43)	PNP β-D-galactopyranoside,	+
BT0988	PNP-α-L-rhamnopyranoside	++

Table 5.3 Activities against aryl-glycosides. "+++" means high/ instant activity and "+" means activity was observed after overnight incubation at 37 ^{o}C overnight, 50 mM Na-HEPES buffer pH 7.0.

Fucosidase BT3665 (GH29) is active against fucosyllactose.

The fucosidase BT3665 (GH29) was tested against fucosyllactose and flax RGI (derived from flax mucilage), which contains Fuc side chains. The results (**Figure 5.10**) showed that the enzyme was only active against fucosyllactose, the trisaccharide was hydrolyzed to lactose and Fuc. TFA (Trifluoroacetic acid) hydrolyzed RGII was also tested against this enzyme but no product was released. Inspection of the Mass Spectrometry analysis of the TFA treated polysaccharide failed to identify a peak containing a terminal Fuc unit, which is likely required for BT3665 to display activity.

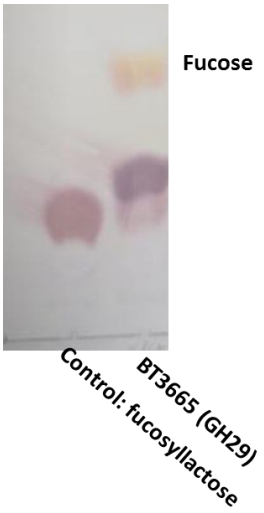


Figure 5.10 TLC images of BT3665 against fucosyllactose. The reaction was carried out at 37 ºC, 50 mM sodium phosphate buffer pH 7; Fucosyllactose concentration, 2 mM and enzyme concentration, 1 µM.

Two enzymes in the RGII PUL are active against chemically hydrolyzed RGII

The enzymes that displayed no activity against RGII was evaluated against the polysaccharide hydrolysed with 2 M TFA at 110 ºC for 1h. The data revealed that two of the proteins, BT0992 (GH2) and BT1018 (GH28), released a glycan that was not identified (**Figure 5.11**). It is speculated that the acid hydrolysis exposed glycosidic bonds that were now accessible to these enzymes. The product of BT1018 (GH28) is predicted to be GalA related, because of the activities displayed by enzymes from this CAZy family, the colour of the glycan released, its migration on TLC, and the fact that it did not elute from the HPLC column with any neutral sugars. The product released by BT0992 (GH2) was likely to be an oligosaccharide rather than a monosaccharide, based on the HPLC elution time. The fact that BT0992 (GH2) displayed no activity against PNP-galactopyranoside or PNP-arabinofuranoside suggested that the activity of this enzyme is likely to be influenced by the sugar at the +1 subsite. It is very possible that the TFA hydrolysis exposed some of the linkages buried deeply in the RGII structure, and thus facilitated the attack from certain enzymes.

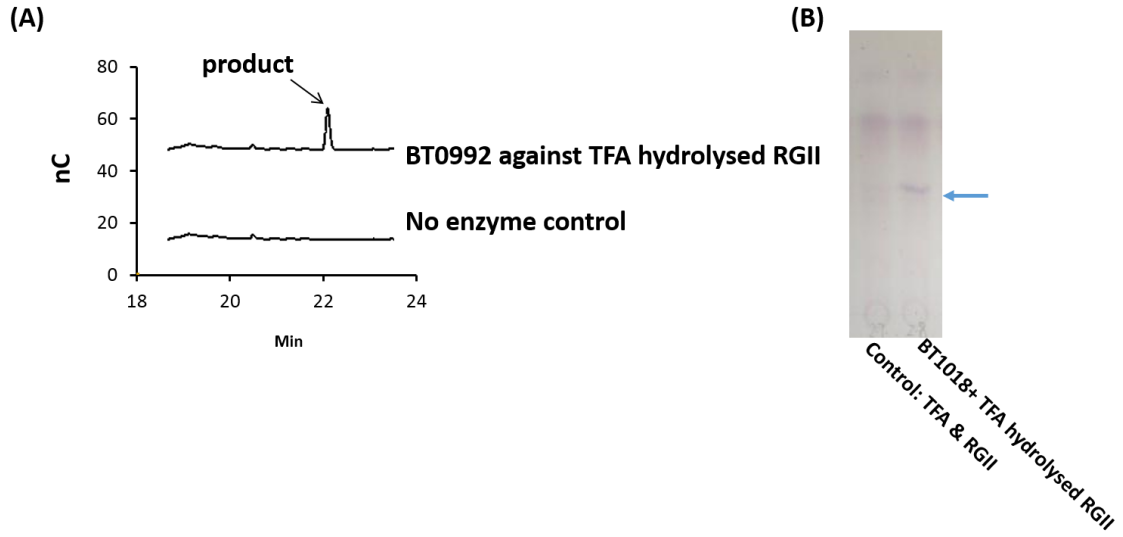


Figure 5.11 BT0992 (GH2) and BT1018 (GH28) incubated with RGII treated with TFA. (A), HPLC of the product released by BT0992; (B), TLC of the product of BT1018 labelled by a blue arrow. Enzyme reactions were carried out in 20mM Na-HEPES pH 7.5 with 1 µM enzyme at 37 ^{o}C for 16 h. Hydrolysis condition, 2 M TFA at 110 ^{o}C for 1 h.

Whole cell and sonicated assays of *Bacteroides thetaiotamicron* against RGII

The whole cell assay was used in this project to study the possible enzymes on the surface of *B. thetaiotaomicron* that contribute to RGII degradation. In this experiment *B. thetaiotaomicron* was grown on RGII to up-regulate the RGII PULs and the cells were harvested at the mid-exponential phase. The intact cells were incubated with RGII in an aerobic environment to prevent metabolism from occurring and thus active transport. Therefore, any depolymerization of RGII would have occurred at the surface of the bacterium. The degradation profile of these cells was compared to RGII degradation by sonicated cells where surface, periplasmic and even cytoplasmic enzymes contribute to the catabolic process. Finally, metabolites derived from RGII incubated with sonicated cells provide insights into which of the glycosidic linkages are not cleaved by the *B. thetaiotaomicron* enzymes.

When non-disrupted *B. thetaiotaomicron* cells were incubated with RGII, GalA was clearly generated **Figure 5.12**. These results indicate the presence of surface exo-enzymes that attack the polygalacturonic acid backbone or, possibly, remove the GalA from the Rha in Chain A, **Figure 5.1**.

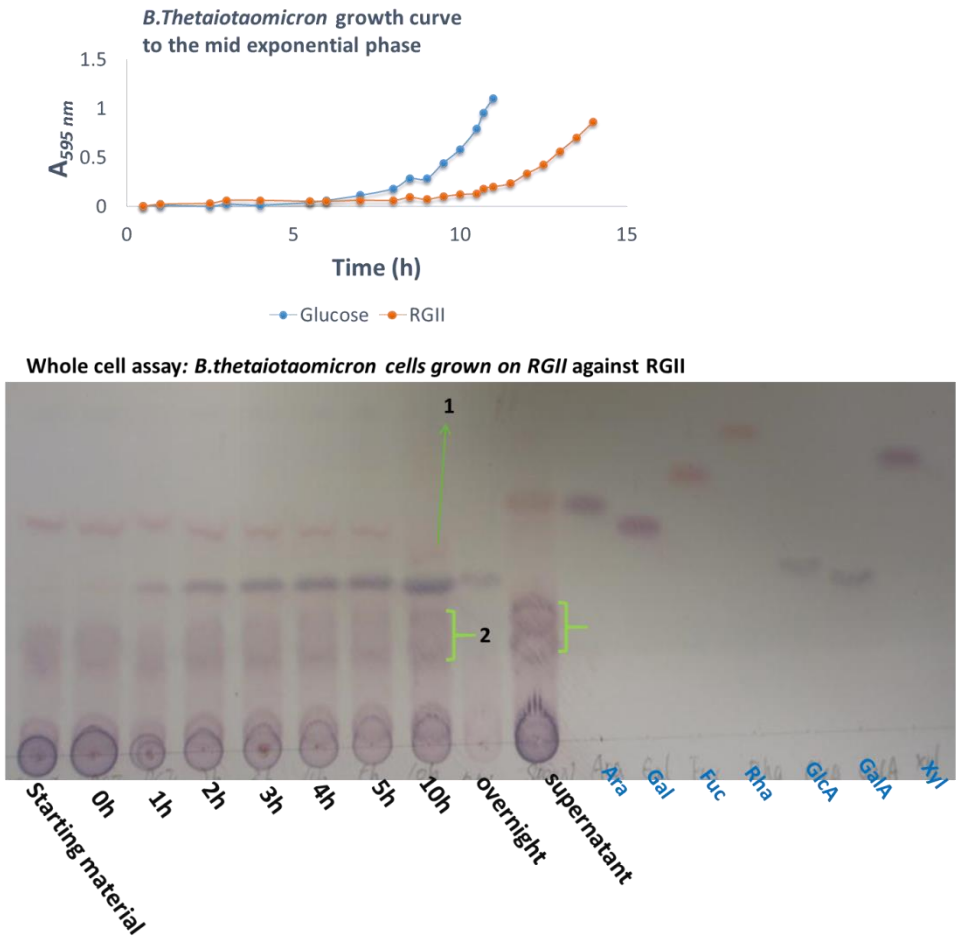


Figure 5.12 Whole cells assay of *B. thetaiotaomicron* incubated with RGII. Upper Panel displays the growth curve of *B. thetaiotaomicron* to mid exponential phase. Bottom Panel shows TLC images of whole cells incubated with RGII in Phosphate buffered saline pH 7.0 containing 2 mM $CaCl_2$ at 37 ºC. The two possible oligosaccharide products are labeled as 1 and 2. "Supernatant" means the growth media leftover after the cells were harvested.

With respect to sonicated cells of *B. thetaiotaomicron* incubated with RGII, the most clearly observed products monosaccharides released were GalA, Gal, Rha and likely Ara were released (**Figure 5.13**). These sugars are present in the various chains in RGII, and in the polygalacturonic acid backbone. Although it is difficult to identify the precise linkages cleaved

in this experiment, they are likely to be the terminal β-L-Ara*f* units that cap chain D and B, the α-L-Ara*f* units of chain E, L-Gal that is the terminal sugar in chain A and Rha, which caps chain C.

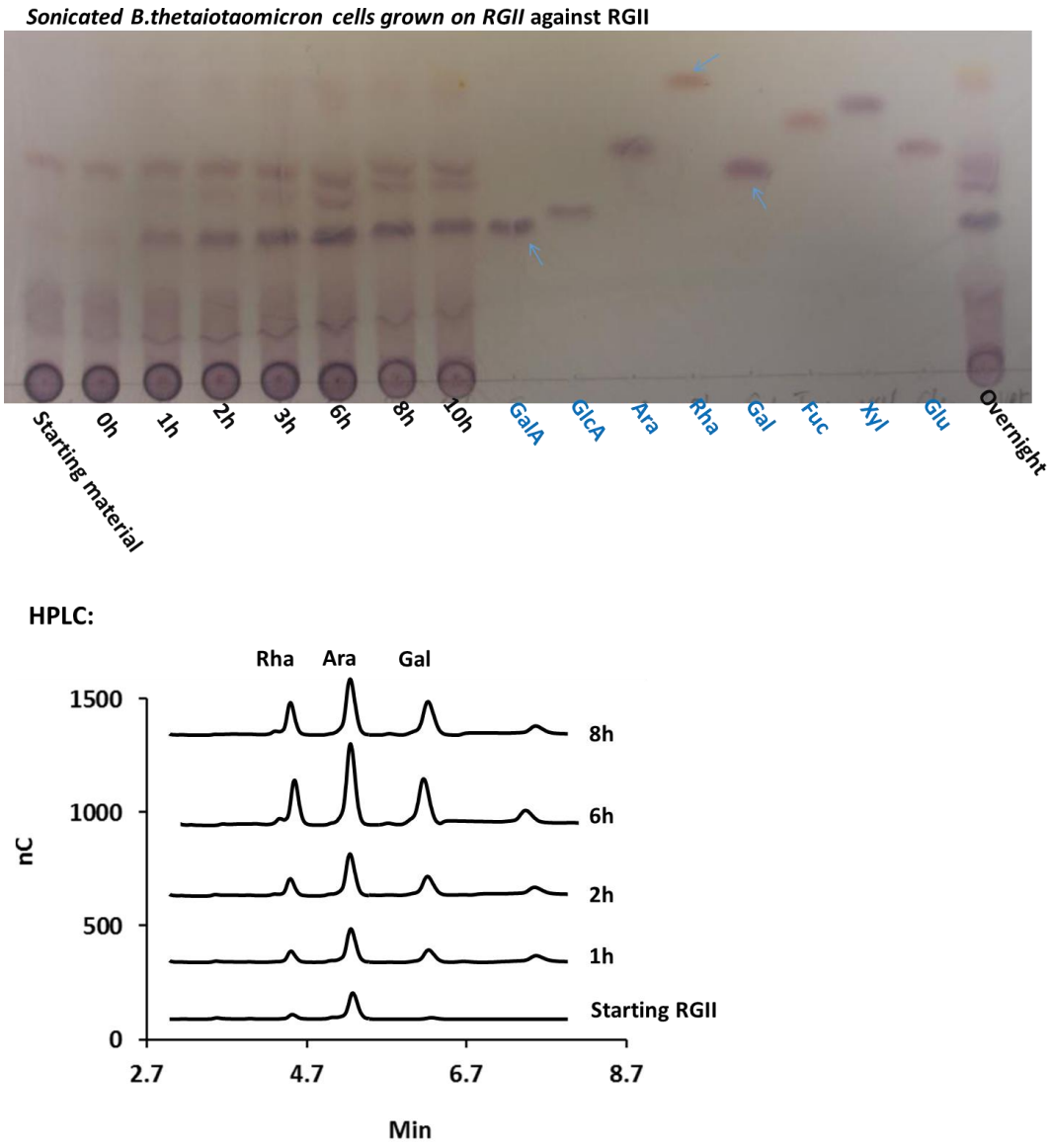


Figure 5.13 Activated of sonicated bacteria cells against RGII. HPLC confirmed the release of Gal, Rha, and possibly Ara. Digestion was in Phosphate buffered saline buffer pH 7.0 with 2 mM $CaCl_2$ at 37 ºC.

To understand the ability of *B. thetaiotamicron* to metabolize monosaccharides (especially rare sugars like Kdo) in RGII, the bacterium was grown on minimal media containing different sugars present in RGII. The experiment showed that *B. thetaiotamicron* could utilize all the tested monosaccharides except for apiose and Kdo (3-Deoxy-D-manno-oct-2-ulosonic acid),

Figure 5.14. With respect to the constitutents of RGII, the bacterium displays a preference for D-Gal and L-Ara, as these sugars support rapid growth to $A_{600\ nm}$ 2.0 The second tier of the favorable sugars contains L-Rha, D-GlcA, D-GalA and L-Gal as they support relatively slow growth of *B. thetaiotamicron* up to A_{600nm} 1.5. Monosaccharides such as Kdo and Apiose (Api) do not support the growth of *B. thetaiotamicron*. Other monosaccharides such as DHA, aceric acid, methyl-xylose and methyl-fucose were not tested due to the lack of commercial availability.

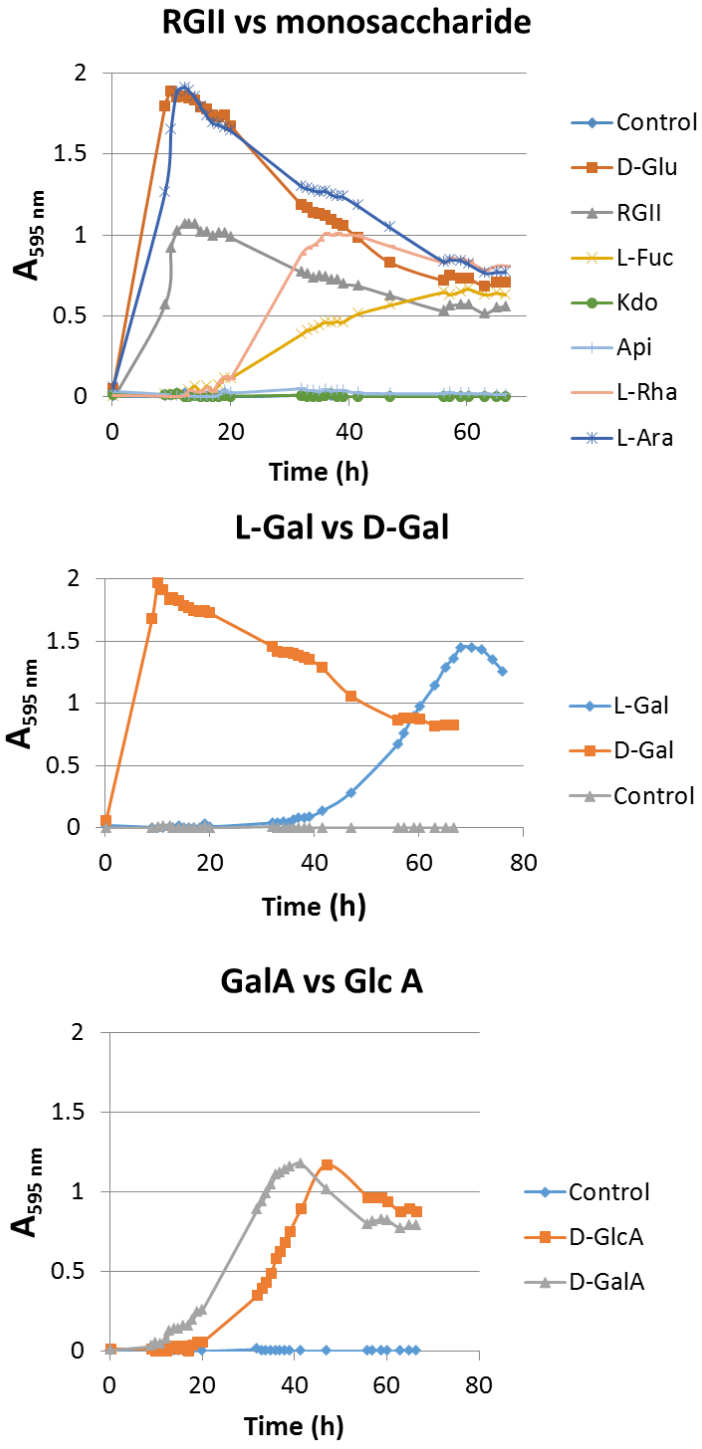


Figure 5.14 Growth of *B. thetaiotaomicron* on RGII and related monosaccharides. Experiments were carried out in 5 ml test tubes and at least two replicates were performed for each sugar. In the control, no carbon source was added.

5.3 Discussion

The ability of *B. thetaiotaomicron* to grow on RGII supplemented minimal media provided the first evidence that a single bacterium can utilize the polysaccharide. Two driving forces may explain the selective pressures that have resulted in RGII metabolism by the bacterium. Firstly, the human large bowel represents a highly competitive environment for the trillions of bacteria that occupy this microbial niche. RGII catabolism may enable a small cohort of microbial symbionts, such as *B. thetaiotaomicron*, to colonize the human gut under nutrient limiting conditions, thereby providing them with a selective advantage. Secondly, the depolymerization of RGII may releases certain molecules that may provide health benefits to the host and even other bacteria, contributing to the mutual benefits provided by the symbiotic-host relationship. For example, boron, which can be released from RGII, has been shown to be important for bone and kidney physiology but more research is required to study the transportation of Boron in Colon (Devirian and Volpe 2003, Pahl, Culver and Vaziri 2005, Hunter 2009).

The data derived from the incubation of the enzyme cocktail, encoded by RGII PULs with the polysaccharide indicated that the catabolism of the glycan may proceed through a coordinated hierarchy of sugar release by a series of exo-acting enzymes. Thus, depolymerization is initiated by the removal of the sugars at the termini of the side-chains. These terminal sugars may have evolved to "shield" the branched polysaccharide from enzymatic attack. However, the proposed hierarchical degradation of RGII complicates the characterization of individual enzymes, assuming they proceed through an exo-mechanism, as the majority of the substrates would remain inaccessible until exposed by upstream reactions. In contrast, RGI utilization by *B. thetaiotaomicron* features several endo-acting enzymes exemplified by the lyases, possibly because endo enzymes expose more reducing and non-reducing termini for the downstream exo-acting enzymes. The lack of obvious endo-acting enzymes in the RGII degrading apparatus

could reflect the highly branched structure of the polysaccharide, which would impose significant constrains on endo-acting enzymes. This could be one reason why the characterization of enzymes encoded by the RGII PULs has progressed more slowly than the RGI degrading apparatus. Another reason is that in some cases the cloned genes (encoding the RGII degrading enzymes) should have included more upstream sequence to make sure that the integrity of the protein sequence was not compromised. Correction has been made to this point by re-cloning certain genes by NZYTech (1649-038 Lisboa, Portugal). Additionally, as illustrated for the analysis of RGI degradation described in Chapter 4, it would be very helpful if certain oligosaccharides derived from RGII could be generated and used to characterize the enzymes encoded by the RGII PULs.

Although the signaling molecule that activates the RGII PUL has not, as yet, been identified, it is speculated that because of exo-mediated digestion, the signal molecule is very likely to be unique monosaccharides such as Kdo and Dha. These two sugars are not deeply hidden inside the structure yet unique enough to specifically up-regulate the RGII PULs, and not other similar PULs such as the RGI PUL. Although **Figure 5.14** shows that *B. thetaiotaomicron* does not grow on Kdo, this does not rule out the possibility of Kdo acting as a signaling sugar. A modification of the signal molecule could potentially switch off the signaling, instead of completely metabolizing the molecule. A more possible signaling candidate would be an oligosaccharide that contains Kdo or Dha. The ability of the bacterium to grow on Dha has not been tested due to lack of commercially available Dha. Another possibility for the signaling molecule is a component of Chain A or B, such as L-Gal from Chain A, or an oligosaccharide after preliminary enzymatic trimming. To explore further the nature of the signaling molecule Chain A and Chain B could be purified and incubated with *B. thetaiotaomicron* followed by transcriptomic analysis to elucidate whether Chain A, Chain B or both chains contribute to the

signaling molecule. Further experiments could focus on the sequential removal of terminal sugars and evaluation of whether the truncated chains can activate PUL-RGII. This approach would identify the component of the side chains that activate the locus.

An interesting speculation concerns BT3662, an enzyme that specifically targets Chain E of RGII, but not terminal L-Ara*f* residues in other polysaccharides, (**Figure 5.9**and **Table 5.2**). It should be emphasized that Chain E is only present in certain types of RGII, exemplified by the polysaccharide from red wine (grape) (Seveno et al. 2009), **Figure 5.15**. RGII from other plants lack Chain E. Of potential significance is the observation that the BT3662 gene is in RGII PUL-C, which is a small PUL remote from the main RGII PUL, RGII PUL-A, ranging from BT0980 to BT1030, **Figure 5.2**. It is speculated that RGII PUL-A has evolved to target primitive forms of RGII found in less advanced plant species, such as non-vascular plants, a version that lacks Chain E. However, as plants evolved, Chain E appears in advanced plants such as *Vitis vinifera Merlot* (grape), due to the potential requirement for extra strength and rigidity of the cell wall. *B. thetaiotaomicron* seems to have adapted to its competitive niche by acquiring genes through horizontal transfer that are incorporated into a PUL, distinct from its original locus, expanding the function of the encoded enzyme consortium. It is also possible that changes in diet that contain different forms of RGII is the evolutionary driver that has selected for genes encoding enzymes such as BT3662. The proposed theory, once confirmed, could provide first clues into the co-evolution of gut bacteria and polysaccharides. It also provides an explanation for the observation that several PULs can be simultaneously up-regulated by the same substrate exemplified by yeast mannan (Martens et al. 2011). Based on the premise that genes with related function are co-localized within the genome, it is possible that the two variable sugars on Chain B (shown in *parenthesis* in **Figure 5.15**) are likely targeted by enzymes encoded by RGII PUL-C as occurs for Chain E.

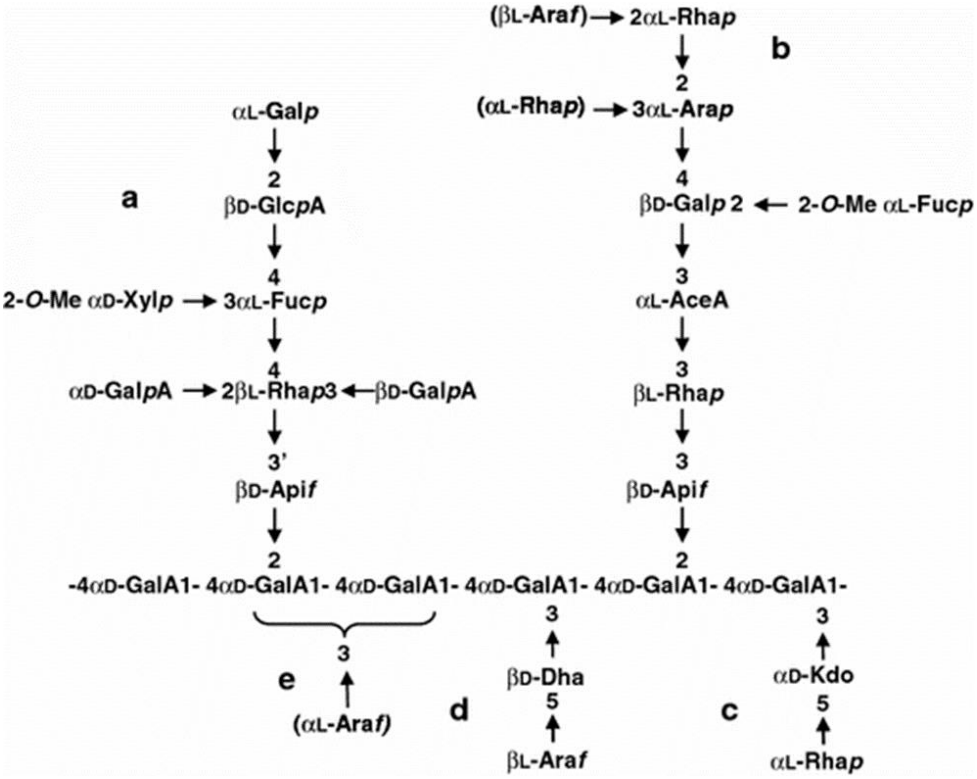


Figure 5.15 Structure of RG-II isolated from angiosperm cell walls. Residues within brackets are absent in *Arabidopsis*. **(Seveno et al. 2009)** Examples of abbreviations: Araf, arabinose furanose; Rhap, rhamnose pyranose. Size chains are labeled as a-e.

A final model of the degradation of RGII is proposed in **Figure 5.16**. Apart from the two enzymes that are highly active against RGII, BT3662 (GH43) and BT1010 (GH95), a fucosidase BT3665 (GH29) is predicted to attack side chain A as it is active against fucosyllactose. BT1019(GH78) and BT0988 released small amount of rhamnose and thus believed to attack the terminal Rha residue on side chain B or C. BT0983(GH2), BT0985 and BT3663(GH43) released trace amount of galactose and thus may attack truncated forms of side chain B in which D-Gal is the terminal sugar (the integrity of a small percentage of side chain B may have been compromised during the purification; Dr Ralet's unpublished data and personal communication from INRA, Nantes, France). BT0993 is also possible to attack the B4 position as it showed strong activity against PNP-β-D-galactopyranoside.

BT0986(GH106) displayed no activity against intact RGII but slow activity against PNP-α-L-rhamnopyranoside, suggesting its activity may be influenced by the sugar at the +1 subsite. Therefore unlike BT1019(GH78) and BT0988, BT0986(GH106) is likely to attack a new type of α-L Rha linkage, such as the one in the B6 position of the side chain B, **Figure 5.16**. The strong activity of BT3661 (GH97) against PNP-α-D-galactopyranoside is puzzling due to the lack of such linkage in the RGII structure. It is possible that because BT3661 is encoded by RGII PUL-C, this enzyme may be specialized to attack rare RGII linkage that is not present in the RGII molecule used in this study.

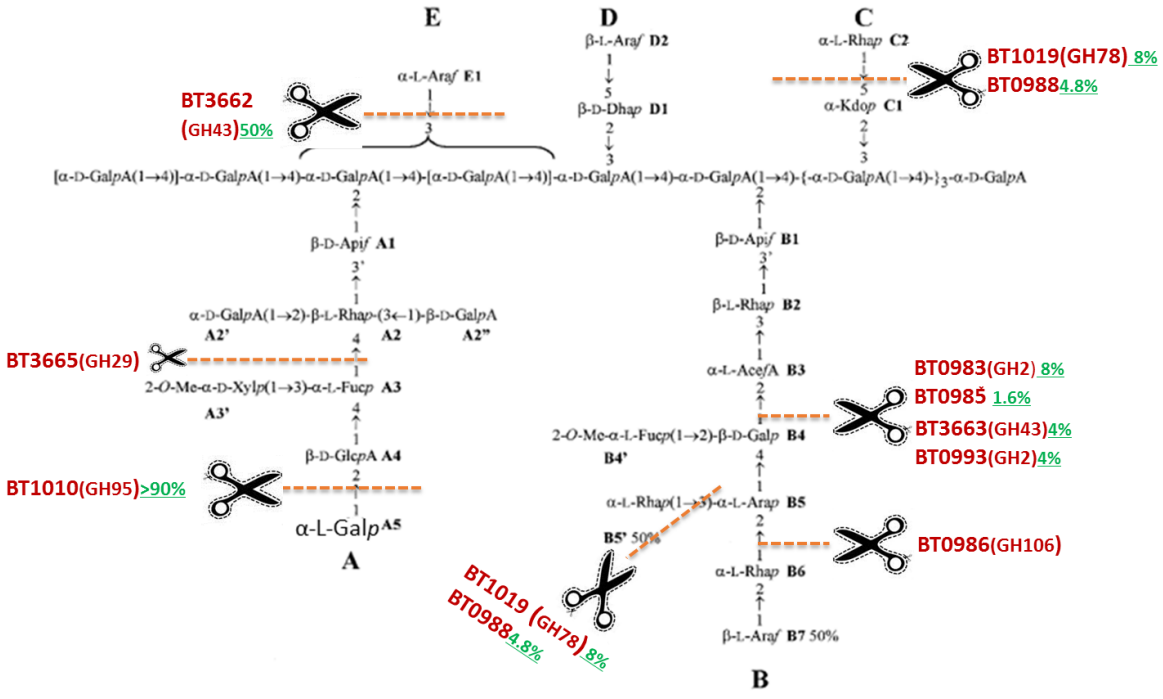


Figure 5.16 Final model of the degradation of RGII. The green percentage shows how much of the total target bond was cleaved by the individual enzyme based on the calculation described before. The products were identified by HPLC unless the green percentage was unlabeled.

5.4 Future work

1. Continue to characterize the biochemical properties of all enzymes encoded by PUL-RGIIs that orchestrate RGII degradation. It is possible that this strategy may not be possible as the hierarchical nature of the degradative process may restrict the biochemical approach. Thus, enzyme characterization should be complemented with a genetic approach to dissect the function of the enzymes encoded by the RGII PULs.

2. Crystallize interesting enzymes such as BT1010 and BT3662 and analyse the structural basis for their observed substrate specificities.

Chapter 6. Final discussion

This final discussion chapter will attempt to draw together the data presented in the thesis. This Chapter will focus on how the work presented in this thesis has contributed to understanding how carbohydrate binding modules (CBMs) influence the specificity and activity of carbohydrate active enzymes, particularly mannanases and esterases, and the mechanisms by which the dominant human gut bacterium *Bacteroides thetaiotaomicron* utilizes highly branched pectins.

Plant cells are surrounded by an extremely complex extracytoplasmic network of polysaccharides, which is mainly composed of cellulose, hemicelluloses and pectins (Albersheim 1975, McNeil et al. 1984, Brett and Waldron 1990, Gilbert 2007, Gilbert et al. 2008). Although the molecular details of the association of hemicelluloses and pectins with cellulose microfibrils is not fully understood, the interactions between cellulose chains both within microfibrils and with matrix polysaccharides restrict their accessibility to enzyme attack. In Chapter 3 the accessibility of mannans in plant cell walls to enzymatic attack was assessed using two mannanases, *Cj*Man5A and *Cj*Man26A (Braithwaite et al. 1995, Hogg et al. 2003). Both enzymes originated from the saprophytic soil bacterium *Cellvibrio japonicus* (formerly *Pseudomonas fluorescens* subsp. *cellulosa*), which degrades mannan, xylan, cellulose, arabinan, galactan and pectins (McKie et al. 2001, Hazlewood and Gilbert 1998, Emami et al. 2002, DeBoy et al. 2008, Gardner et al. 2014). Recent studies have shown that *Cj*Man5A displays more promiscuous specificity than *Cj*Man26A. While both enzymes bind mannose (Man) in the active site only the GH5 mannanase can accommodate glucose (Glc) in the distal regions of its substrate binding cleft (Hogg et al. 2003, Cartmell et al. 2008). Chapter 3 explores how different CBMs potentiate both endo-mannanases and esterases *in vivo*. The results are consistent with the hypothesis, proposed by Hervé et al. (2010) that CBMs promote the

enzymatic deconstruction of intact plant cell walls by targeting and proximity effects. The catalytic module *Cj*Man5A$_{CM}$ can access, equally, all the mannan in the tobacco cell wall, while *Cj*Man26A$_{CM}$ can only hydrolyse ~50% of the polysaccharide in these cell walls. In the cell wall of *Physcomitrella patens*, neither of the two catalytic modules can fully hydrolyse all the mannan in the cell wall. This reflects the need for CBMs and differences in the specificity of the enzymes as *Cj*Man5A can accommodate glucose in subsites distal to the active site, while *Cj*Man26A displays a strict requirement for mannose at all negative subsites (Tailford et al. 2009). The observation that a cellulose binding CBM had the largest impact on the activity of *Cj*Man5A against tobacco walls could be explained by the close association of mannan with cellulose, and the observation that cellulose, which comprises 45% of the wall of tobacco callus cells, is ~7-fold more abundant than mannan in this plant species (Albersheim 1975, McNeil et al. 1984). In *Physcomitrella*, CBM27, which binds specifically to mannan, enables *Cj*Man5A to access all the mannan. It is possible that the mannan in *Physcomitrella* is heavily acetylated, and that the acetyl groups may create a steric block that restricts the ability of the two mannanases from attacking the polysaccharide, but does not prevent CBM binding. Indeed, the observation that CCRC-M170 binds to *Physcomitrella* walls demonstrates the presence of acetylated mannan. It is also possible that the "inaccessible" mannan in *Physcomitrella* walls is glucomannan that has a particularly high Glc:Man ratio and is thus not accessible to *Cj*Man26A. Overall, the data in Chapter 3 indicate that CBMs, which bind to the most abundant polysaccharide in the wall, or the polymer most closely associated with the substrate of the enzyme, are likely to maximize enzymatic activity. This also applies to esterases; CBM27 potentiated the activity of the esterase *Cj*CE2C against *Physcomitrella* walls, while the cellulose binding CBM3a had little impact.

In Chapters 4 and 5 pectin degradation and utilization by the dominant gut bacterium *B. thetaiotaomicron* was studied. The human large bowel is colonized by a community of microbes that makes a significant contribution to the nutrition and health of its host (Ley, Peterson and Gordon 2006, Siezen and Kleerebezem 2011, Arumugam et al. 2011, Kamada et al. 2012, Koropatkin, Cameron and Martens 2012, McNulty et al. 2013). While there have been numerous attempts to manipulate the composition of this ecosystem through dietary and probiotic intervention (Wollowski, Rechkemmer and Pool-Zobel 2001, Cotillard et al. 2013), such endeavors have been hampered by an underlying lack of knowledge of the biochemistry of these microorganisms, particularly with respect to their capacity to utilize both dietary and host carbohydrate polymers as significant nutrients (Hooper et al. 2002). Interestingly the genome of *B. thetaiotaomicron* encodes a very large number of glycoside hydrolases, and thus is predicted to target host and dietary glycans. (Martens et al. 2011, Xu et al. 2003) It is estimated that in industrialized countries 5 g of pectins are consumed per capita per day and are thus represents an important growth substrate for the human large bowel microbial community (microbiota) (Ridley et al. 2001, Thakur et al. 1997). The data in Chapter 4 has helped to develop a model for RGI (especially its backbone) utilization by *B. thetaiotaomicron.* Many interesting enzymes active against RGI and RGII have been found and characterized, exemplified by the three lyase and the two GH105 enzymes that hydrolyse Δ4,5GalA from the resultant unsaturated RGI oligosaccharides. The three lyases target different RGI structures and thus are not "functionally redundant" as substrate specificity contributes to enzymatic efficiency. From an evolutionary point of view, these lyases may have exist to cater for the variety of pectins from different food sources. The data in Chapter 4 also show how the bacterium deploys a plethora of enzymes to digest simple structures exemplified by the RGI backbone containing only two kinds of linkage. It is very possible that the involvement of endo acting enzymes speeds up the digestion and the participation of lyases creates a unique product,

such as the unsaturated tetrasaccharide that specifically up-regulates the RGI PUL, instead of similar PULs that orchestrate the degradation of glycans containing Rha and GalA. Considering Chapter 3, where the enzymatic degradation of integral plant cell wall glycans is potentiated by CBMs, the plethora of enzymes required to digest the simple structures presented by RGI may be necessary to compensate for the lack of CBMs in these enzymes. It is possible that under certain evolutionary pressures saprophytic bacteria such as *Cellvibrio. japonicus* use different CBMs to ensure complete enzymatic degradation of plant cell walls, as the carbon source in the soil, which is utilized by a large number of microorganisms, is not in constant supply especially in the winter. Human gut bacteria, on the other hand, enjoy a more abundant and regular supply of carbon derived from the diet and host glycans, and, to effectively compete, need to rapidly digest and utilize the resultant monosaccharides sugars. This rapid digestion requires that the bacteria target highly accessible polysaccharides, such as pectins, rather than integral components of the plant cell wall which are degraded very slowly. This may explain the lack of CBM containing enzymes expressed by *B. thetaiotaomicron*, as the bacterium targets only accessible dietary glycans. Indeed, CBMs are lacking in the hemicellulases expressed by human gut bacteria (Martens et al. 2011), suggesting that these glycoside hydrolases are targeting accessible forms of these polysaccharides, which likely act as storage polymers, rather than the xylans and mannans that are integral components of the plant cell wall.

In Chapter 4 a mechanism named "signaling molecular protection" was proposed, as the data indicated that the activity of the GH105 enzyme BT4174, which targets the unsaturated tetrasaccharide signaling molecule, is significantly lower than the GH105 enzyme, BT4176, that hydrolyses the unsaturated disaccharide. It would appear that evolution has regulated the signaling pathway by tuning enzyme activity. More work is required to explore whether tuning

the activity of degrading enzymes is a common mechanism for preserving the signaling molecules that activate PULs.

The fact that *B. thetaiotaomicron* can grow on RGII proves that this bacterium is not just capable of rapidly turning over accessible dietary polysaccharide, but also able to digest one of the most complicated structures found in nature, even when it is not the main component of the plant cell wall. This may help to explain its dominance in the human gut. The data presented in Chapter 5 identified only two enzymes that attack RGII, which remove the terminal sugars in chain A and chain E. However, the RGII PUL encodes a very large number of putative enzymes, located in CAZy families where members are known to hydrolyse some of the linkages in RGII. Furthermore, some of these enzymes were shown to be active against aryl-glycosides. Their lack of activity against RGII suggests that the glycan is degraded exclusively by a hierarchical exo-mechanism. Thus, if a single enzyme was either not functional when produced in recombinant form from *E. coli*, or possibly not encoded by the RGII PUL explored here, then the degradative pathway would be blocked, preventing the activity of enzymes acting downstream of this "block" to be identified. It is also possible that *B. thetaiotaomicron* only "nibbles" on RGII, utilizing a few terminal sugars. This possibility seems unlikely, however, given the large number of proteins, several of which are in CAZy families, encoded by the RGII PULs. Clearly more research is needed to understand the digestion, transport and utilization of RGII by *B. thetaiotaomicron.*

Reference

Abbott, D. W. & A. B. Boraston (2007) A family 2 pectate lyase displays a rare fold and transition metal-assisted beta-elimination. *J Biol Chem,* 282**,** 35328-36.

Ahn, J. W., R. Verma, M. Kim, J. Y. Lee, Y. K. Kim, J. W. Bang, W. D. Reiter & H. S. Pai (2006) Depletion of UDP-D-apiose/UDP-D-xylose synthases results in rhamnogalacturonan-II deficiency, cell wall thickening, and cell death in higher plants. *J Biol Chem,* 281**,** 13708-16.

Albersheim, P. (1975) The walls of growing plant cells. *Sci Am,* 232**,** 80-95.

--- (1989) Cell-Walls - the Battleground of Plant-Microbe Interactions - a Citation Classic Commentary on Biochemistry of the Cell-Wall in Relation to Infective Processes by Albersheim,P., Jones,T.M., and English,P.D. *Current Contents/Agriculture Biology & Environmental Sciences***,** 14-14.

Alghisi, P. & F. Favaron (1995) Pectin-Degrading Enzymes and Plant-Parasite Interactions. *European Journal of Plant Pathology,* 101**,** 365-375.

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

Andrews, S. R., E. J. Taylor, G. Pell, F. Vincent, V. M. Ducros, G. J. Davies, J. H. Lakey & H. J. Gilbert (2004) The use of forced protein evolution to investigate and improve stability of family 10 xylanases. The production of Ca2+-independent stable xylanases. *J Biol Chem,* 279**,** 54369-79.

Armand, S., S. Drouillard, M. Schulein, B. Henrissat & H. Driguez (1997) A bifunctionalized fluorogenic tetrasaccharide as a substrate to study cellulases. *Journal of Biological Chemistry,* 272**,** 2709-2713.

Arumugam, M., J. Raes, E. Pelletier, D. Le Paslier, T. Yamada, D. R. Mende, G. R. Fernandes, J. Tap, T. Bruls, J. M. Batto, M. Bertalan, N. Borruel, F. Casellas, L. Fernandez, L. Gautier, T. Hansen, M. Hattori, T. Hayashi, M. Kleerebezem, K. Kurokawa, M. Leclerc, F. Levenez, C. Manichanh, H. B. Nielsen, T. Nielsen, N. Pons, J. Poulain, J. Qin, T. Sicheritz-Ponten, S. Tims, D. Torrents, E. Ugarte, E. G. Zoetendal, J. Wang, F. Guarner, O. Pedersen, W. M. de Vos, S. Brunak, J. Dore, H. I. T. C. Meta, M. Antolin, F. Artiguenave, H. M. Blottiere, M. Almeida, C. Brechot, C.

Cara, C. Chervaux, A. Cultrone, C. Delorme, G. Denariaz, R. Dervyn, K. U. Foerstner, C. Friss, M. van de Guchte, E. Guedon, F. Haimet, W. Huber, J. van Hylckama-Vlieg, A. Jamet, C. Juste, G. Kaci, J. Knol, O. Lakhdari, S. Layec, K. Le Roux, E. Maguin, A. Merieux, R. Melo Minardi, C. M'Rini, J. Muller, R. Oozeer, J. Parkhill, P. Renault, M. Rescigno, N. Sanchez, S. Sunagawa, A. Torrejon, K. Turner, G. Vandemeulebrouck, E. Varela, Y. Winogradsky, G. Zeller, J. Weissenbach, S. D. Ehrlich & P. Bork (2011) Enterotypes of the human gut microbiome. *Nature,* 473**,** 174-80.

Backhed, F., R. E. Ley, J. L. Sonnenburg, D. A. Peterson & J. I. Gordon (2005) Host-bacterial mutualism in the human intestine. *Science,* 307**,** 1915-20.

Beguin, P. & J. P. Aubert (1994) The biological degradation of cellulose. *FEMS Microbiol Rev,* 13**,** 25-58.

Benen, J. A., H. C. Kester, L. Parenicova & J. Visser (2000) Characterization of *Aspergillus niger* pectate lyase A. *Biochemistry,* 39**,** 15563-9.

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

Bjursell, M. K., E. C. Martens & J. I. Gordon (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**,** 36269-79.

Blake, A. W., L. McCartney, J. E. Flint, D. N. Bolam, A. B. Boraston, H. J. Gilbert & J. P. Knox (2006) Understanding the biological rationale for the diversity of cellulose-directed carbohydrate-binding modules in prokaryotic enzymes. *J Biol Chem,* 281**,** 29321-9.

Bolam, D. N., A. Ciruela, S. McQueen-Mason, P. Simpson, M. P. Williamson, J. E. Rixon, A. Boraston, G. P. Hazlewood & H. J. Gilbert (1998) *Pseudomonas* cellulose-binding domains mediate their effects by increasing enzyme substrate proximity. *Biochem J,* 331 (Pt 3)**,** 775-81.

Bolam, D. N., N. Hughes, R. Virden, J. H. Lakey, G. P. Hazlewood, B. Henrissat, K. L. Braithwaite & H. J. Gilbert (1996) Mannanase A from *Pseudomonas fluorescens* ssp. cellulosa is a retaining glycosyl hydrolase in which E212 and E320 are the putative catalytic residues. *Biochemistry,* 35**,** 16195-204.

Bolam, D. N., H. Xie, G. Pell, D. Hogg, G. Galbraith, B. Henrissat & H. J. Gilbert (2004) X4 modules represent a new family of carbohydrate-binding modules that display novel properties. *J Biol Chem,* 279**,** 22953-63.

Boraston, A. B., D. N. Bolam, H. J. Gilbert & G. J. Davies (2004) Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem J,* 382**,** 769-81.

Boraston, A. B., D. Nurizzo, V. Notenboom, V. Ducros, D. R. Rose, D. G. Kilburn & G. J. Davies (2002) Differential oligosaccharide recognition by evolutionarily-related beta-1,4 and beta-1,3 glucan-binding modules. *J Mol Biol,* 319**,** 1143-56.

Braithwaite, K. L., G. W. Black, G. P. Hazlewood, B. R. Ali & H. J. Gilbert (1995) A non-modular endo-beta-1,4-mannanase from *Pseudomonas fluorescens* subspecies cellulosa. *Biochem J,* 305 (Pt 3)**,** 1005-10.

Brett, C. T. & K. Waldron. 1990. *Physiology and biochemistry of plant cell walls*. London ; Boston: Unwin Hyman.

Brett CT, W. K. (1996) Physiology and Biochemistry of Plant Cell Walls. *Chapman and Hall edn*.

Bussink, H. J., K. B. Brouwer, L. H. de Graaff, H. C. Kester & J. Visser (1991) Identification and characterization of a second polygalacturonase gene of *Aspergillus niger*. *Curr Genet,* 20**,** 301-7.

Cantarel, B. L., P. M. Coutinho, C. Rancurel, T. Bernard, V. Lombard & B. Henrissat (2009a) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Research,* 37**,** D233-D238.

--- (2009b) The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res,* 37**,** D233-8.

Capek, P., M. Kubackova, J. Alfoldi, L. Bilisics, D. Liskova & D. Kakoniova (2000) Galactoglucomannan from the secondary cell wall of Picea abies L. Karst. *Carbohydr Res,* 329**,** 635-45.

Carpita, N. C. & D. M. Gibeaut (1993) Structural models of primary cell walls in flowering plants: consistency of molecular structure with the physical properties of the walls during growth. *Plant J,* 3**,** 1-30.

Carrard, G., A. Koivula, H. Soderlund & P. Beguin (2000) Cellulose-binding domains promote hydrolysis of different sites on crystalline cellulose. *Proc Natl Acad Sci U S A,* 97**,** 10342-7.

Cartmell, A., E. Topakas, V. M. Ducros, M. D. Suits, G. J. Davies & H. J. Gilbert (2008) The *Cellvibrio japonicus* mannanase CjMan26C displays a unique exo-mode of action that is conferred by subtle changes to the distal region of the active site. *J Biol Chem,* 283**,** 34403-13.

Chang, C. & R. C. Stewart (1998) The two-component system. Regulation of diverse signaling pathways in prokaryotes and eukaryotes. *Plant Physiol,* 117**,** 723-31.

Chanzy, H. D., A. Grosrenaud, R. Vuong & W. Mackie (1984) The crystalline polymorphism of mannan in plant cell walls and after recrystallisation. *Planta,* 161**,** 320-9.

Charnock, S. J., I. E. Brown, J. P. Turkenburg, G. W. Black & G. J. Davies (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**,** 12067-72.

Cohen, S. N., A. C. Chang & L. Hsu (1972) Nonchromosomal antibiotic resistance in bacteria: genetic transformation of *Escherichia coli* by R-factor DNA. *Proc Natl Acad Sci U S A,* 69**,** 2110-4.

Correia, M. A., D. W. Abbott, T. M. Gloster, V. O. Fernandes, J. A. Prates, C. Montanier, C. Dumon, M. P. Williamson, R. B. Tunnicliffe, Z. Liu, J. E. Flint, G. J. Davies, B. Henrissat, P. M. Coutinho, C. M. Fontes & H. J. Gilbert (2010) Signature active site architectures illuminate the molecular basis for ligand specificity in family 35 carbohydrate binding module. *Biochemistry,* 49**,** 6193-205.

Correia, M. A., J. A. Prates, J. Bras, C. M. Fontes, J. A. Newman, R. J. Lewis, H. J. Gilbert & J. E. Flint (2008) Crystal structure of a cellulosomal family 3 carbohydrate esterase from *Clostridium thermocellum* provides insights into the mechanism of substrate recognition. *J Mol Biol,* 379**,** 64-72.

Cosgrove, D. J. (2005) Growth of the plant cell wall. *Nat Rev Mol Cell Biol,* 6**,** 850-61.

Cotillard, A., S. P. Kennedy, L. C. Kong, E. Prifti, N. Pons, E. Le Chatelier, M. Almeida, B. Quinquis, F. Levenez, N. Galleron, S. Gougis, S. Rizkalla, J. M. Batto, P. Renault, J. Dore, J. D. Zucker, K. Clement, S. D. Ehrlich & A. M. Consortium (2013) Dietary intervention impact on gut microbial gene richness. *Nature,* 500**,** 585-+.

Coughlan, M. P. (1985) The Properties of Fungal and Bacterial Cellulases with Comment on Their Production and Application. *Biotechnology & Genetic Engineering Reviews,* 3**,** 39-109.

Coutinho, P. M., E. Deleury, G. J. Davies & B. Henrissat (2003) An evolving hierarchical family classification for glycosyltransferases. *J Mol Biol,* 328**,** 307-17.

Cuskin, F., J. E. Flint, T. M. Gloster, C. Morland, A. Basle, B. Henrissat, P. M. Coutinho, A. Strazzulli, A. S. Solovyova, G. J. Davies & H. J. Gilbert (2012) How nature can exploit nonspecific catalytic and carbohydrate binding modules to create enzymatic specificity. *Proceedings of the National Academy of Sciences of the United States of America,* 109**,** 20889-20894.

Davies, G. & B. Henrissat (1995) Structures and Mechanisms of Glycosyl Hydrolases. *Structure,* 3**,** 853-859.

Davies, G. J., T. M. Gloster & B. Henrissat (2005) Recent structural insights into the expanding world of carbohydrate-active enzymes. *Current Opinion in Structural Biology,* 15**,** 637-645.

de Vries, R. P., H. C. M. Kester, C. H. Poulsen, J. A. E. Benen & J. Visser (2000) Synergy between enzymes from *Aspergillus* involved in the degradation of plant cell wall polysaccharides. *Carbohydrate Research,* 327**,** 401-410.

DeBoy, R. T., E. F. Mongodin, D. E. Fouts, L. E. Tailford, H. Khouri, J. B. Emerson, Y. Mohamoud, K. Watkins, B. Henrissat, H. J. Gilbert & K. E. Nelson (2008) Insights into plant cell wall degradation from the genome sequence of the soil bacterium Cellvibrio japonicus. *J Bacteriol,* 190**,** 5455-63.

DeMartini, J. D., S. Pattathil, U. Avci, K. Szekalski, K. Mazumder, M. G. Hahn & C. E. Wyman (2011) Application of monoclonal antibodies to investigate plant cell wall deconstruction for biofuels production. *Energy & Environmental Science,* 4**,** 4332-4339.

Deng, C., M. A. O'Neill & W. S. York (2006) Selective chemical depolymerization of rhamnogalacturonans. *Carbohydr Res,* 341**,** 474-84.

Devirian, T. A. & S. L. Volpe (2003) The physiological effects of dietary boron. *Critical Reviews in Food Science and Nutrition,* 43**,** 219-231.

Din, N., H. G. Damude, N. R. Gilkes, R. C. Miller, Jr., R. A. Warren & D. G. Kilburn (1994) C1-Cx revisited: intramolecular synergism in a cellulase. *Proc Natl Acad Sci U S A,* 91**,** 11383-7.

Doco, T. & J. M. Brillouet (1993) Isolation and Characterization of a Rhamnogalacturonan-Ii from Red Wine. *Carbohydrate Research,* 243**,** 333-343.

Dow, J. M., D. E. Milligan, L. Jamieson, C. E. Barber & M. J. Daniels (1989) Molecular-Cloning of a Polygalacturonate Lyase Gene from Xanthomonas-Campestris Pv Campestris and Role of the Gene-Product in Pathogenicity. *Physiological and Molecular Plant Pathology,* 35**,** 113-120.

El Kaoutari, A., F. Armougom, J. I. Gordon, D. Raoult & B. Henrissat (2013) The abundance and variety of carbohydrate-active enzymes in the human gut microbiota. *Nat Rev Microbiol,* 11**,** 497-504.

Emami, K., T. Nagy, C. M. Fontes, L. M. Ferreira & H. J. Gilbert (2002) Evidence for temporal regulation of the two *Pseudomonas cellulosa* xylanases belonging to glycoside hydrolase family 11. *J Bacteriol,* 184**,** 4124-33.

Ferreira, L. M., T. M. Wood, G. Williamson, C. Faulds, G. P. Hazlewood, G. W. Black & H. J. Gilbert (1993) A modular esterase from *Pseudomonas fluorescens* subsp. cellulosa contains a non-catalytic cellulose-binding domain. *Biochem J,* 294 (Pt 2)**,** 349-55.

Fu, H., M. P. Yadav & E. A. Nothnagel (2007) *Physcomitrella patens* arabinogalactan proteins contain abundant terminal 3-O-methyl-L: -rhamnosyl residues not found in angiosperms. *Planta,* 226**,** 1511-24.

Gardner, J. G., L. Crouch, A. Labourel, Z. Forsberg, Y. V. Bukhman, G. Vaaje-Kolstad, H. J. Gilbert & D. H. Keating (2014) Systems biology defines the biological significance of redox-active proteins during cellulose degradation in an aerobic bacterium. *Mol Microbiol*.

Garron, M. L. & M. Cygler (2010) Structural and mechanistic classification of uronic acid-containing polysaccharide lyases. *Glycobiology,* 20**,** 1547-73.

Gilbert, H. J. (2003) How carbohydrate binding modules overcome ligand complexity. *Structure,* 11**,** 609-10.

--- (2007) Cellulosomes: microbial nanomachines that display plasticity in quaternary structure. *Mol Microbiol,* 63**,** 1568-76.

--- (2010) The biochemistry and structural biology of plant cell wall deconstruction. *Plant Physiol,* 153**,** 444-55.

Gilbert, H. J. & G. P. Hazlewood (1991) Genetic modification of fibre digestion. *Proc Nutr Soc,* 50**,** 173-86.

Gilbert, H. J., G. Jenkins, D. A. Sullivan & J. Hall (1987) Evidence for multiple carboxymethylcellulase genes in *Pseudomonas fluorescens* subsp. cellulosa. *Mol Gen Genet,* 210**,** 551-6.

Gilbert, H. J., J. P. Knox & A. B. Boraston (2013) Advances in understanding the molecular basis of plant cell wall polysaccharide recognition by carbohydrate-binding modules. *Current Opinion in Structural Biology,* 23**,** 669-677.

Gilbert, H. J., H. Stalbrand & H. Brumer (2008) How the walls come crumbling down: recent structural biochemistry of plant polysaccharide degradation. *Curr Opin Plant Biol,* 11**,** 338-48.

Gilbert, H. J., D. A. Sullivan, G. Jenkins, L. E. Kellett, N. P. Minton & J. Hall (1988) Molecular cloning of multiple xylanase genes from *Pseudomonas fluorescens* subsp. cellulosa. *J Gen Microbiol,* 134**,** 3239-47.

Gill, S. C. & P. H. von Hippel (1989) Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem,* 182**,** 319-26.

Gomez, L. D., C. G. Steele-King, L. Jones, J. M. Foster, S. Vuttipongchaikij & S. J. McQueen-Mason (2009) Arabinan metabolism during seed development and germination in *Arabidopsis*. *Mol Plant,* 2**,** 966-76.

Gordillo, F., V. Caputo, A. Peirano, R. Chavez, J. Van Beeumen, I. Vandenberghe, M. Claeyssens, P. Bull, M. C. Ravanal & J. Eyzaguirre (2006) Penicillium purpurogenum produces a family 1 acetyl xylan esterase containing a carbohydrate-binding module: characterization of the protein and its gene. *Mycological Research,* 110**,** 1129-1139.

Guillen, D., S. Sanchez & R. Rodriguez-Sanoja (2010) Carbohydrate-binding domains: multiplicity of biological roles. *Appl Microbiol Biotechnol,* 85**,** 1241-9.

Hahn, M. G., J. Puhlmann, M. J. Swain, W. Steffan, P. Albersheim & A. G. Darvill (1993) Generation and Characterization of Monoclonal-Antibodies against Plant-Cell Wall Polysaccharides. *Journal of Cellular Biochemistry***,** 14-14.

Hakulinen, N., M. Tenkanen & J. Rouvinen (2000) Three-dimensional structure of the catalytic core of acetylxylan esterase from Trichoderma reesei: Insights into the deacetylation mechanism. *Journal of Structural Biology,* 132**,** 180-190.

Hall, J., G. P. Hazlewood, P. J. Barker & H. J. Gilbert (1988) Conserved reiterated domains in *Clostridium thermocellum* endoglucanases are not essential for catalytic activity. *Gene,* 69**,** 29-38.

Hall, J., G. P. Hazlewood, N. S. Huskisson, A. J. Durrant & H. J. Gilbert (1989) Conserved serine-rich sequences in xylanase and cellulase from *Pseudomonas fluorescens* subspecies cellulosa: internal signal sequence and unusual protein processing. *Mol Microbiol,* 3**,** 1211-9.

Handford, M. G., T. C. Baldwin, F. Goubet, T. A. Prime, J. Miles, X. Yu & P. Dupree (2003) Localisation and characterisation of cell wall mannan polysaccharides in *Arabidopsis thaliana*. *Planta,* 218**,** 27-36.

Harholt, J., A. Suttangkakul & H. V. Scheller (2010) Biosynthesis of Pectin. *Plant Physiology,* 153**,** 384-395.

Harmsen, J. A., M. A. Kusters-van Someren & J. Visser (1990) Cloning and expression of a second *Aspergillus niger* pectin lyase gene (pelA): indications of a pectin lyase gene family in A. niger. *Curr Genet,* 18**,** 161-6.

Hayashi, T., K. Ogawa & Y. Mitsuishi (1994a) Characterization of the adsorption of Xyloglucan to Cellulose. *Plant Cell Physiol.,* 35**,** 1199-1205.

Hayashi, T., K. Ogawa & Y. Mitsuishi (1994b) Characterization of the Adsorption of Xyloglucan to Cellulose. *Plant and Cell Physiology,* 35**,** 1199-1205.

Hazlewood, G. P. & H. J. Gilbert (1998) Structure and function analysis of *Pseudomonas* plant cell wall hydrolases. *Prog Nucleic Acid Res Mol Biol,* 61**,** 211-41.

Henrissat, B. (1991) A classification of glycosyl hydrolases based on amino acid sequence similarities. *Biochem J,* 280 (Pt 2)**,** 309-16.

Henrissat, B., I. Callebaut, S. Fabrega, P. Lehn, J. P. Mornon & G. Davies (1995) Conserved Catalytic Machinery and the Prediction of a Common Fold for Several Families of Glycosyl Hydrolases. *Proceedings of the National Academy of Sciences of the United States of America,* 92**,** 7090-7094.

Henrissat, B. & G. Davies (1997) Structural and sequence-based classification of glycoside hydrolases. *Curr Opin Struct Biol,* 7**,** 637-44.

Henshaw, J. L., D. N. Bolam, V. M. Pires, M. Czjzek, B. Henrissat, L. M. Ferreira, C. M. Fontes & H. J. Gilbert (2004) The family 6 carbohydrate binding module CmCBM6-2 contains two ligand-binding sites with distinct specificities. *J Biol Chem,* 279**,** 21552-9.

Herron, S. R., R. D. Scavetta, M. Garrett, M. Legner & F. Jurnak (2003) Characterization and implications of Ca2+ binding to pectate lyase C. *J Biol Chem,* 278**,** 12271-7.

Herve, C., A. Rogowski, A. W. Blake, S. E. Marcus, H. J. Gilbert & J. P. Knox (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**,** 15293-8.

Himmel, M. E. & E. A. Bayer (2009) Lignocellulose conversion to biofuels: current challenges, global perspectives. *Curr Opin Biotechnol,* 20**,** 316-7.

Himmel, M. E., S. Y. Ding, D. K. Johnson, W. S. Adney, M. R. Nimlos, J. W. Brady & T. D. Foust (2007) Biomass recalcitrance: engineering plants and enzymes for biofuels production. *Science,* 315**,** 804-7.

Hogg, D., G. Pell, P. Dupree, F. Goubet, S. M. Martin-Orue, S. Armand & H. J. Gilbert (2003) The modular architecture of *Cellvibrio japonicus* mannanases in glycoside

hydrolase families 5 and 26 points to differences in their role in mannan degradation. *Biochem J,* 371**,** 1027-43.

Hogg, D., E. J. Woo, D. N. Bolam, V. A. McKie, H. J. Gilbert & R. W. Pickersgill (2001) Crystal structure of mannanase 26A from *Pseudomonas cellulosa* and analysis of residues involved in substrate binding. *J Biol Chem,* 276**,** 31186-92.

Hooper, L. V., T. Midtvedt & J. I. Gordon (2002) How host-microbial interactions shape the nutrient environment of the mammalian intestine. *Annual Review of Nutrition,* 22**,** 283-307.

Hunter, P. (2009) Not boring at all Boron is the new carbon in the quest for novel drug candidates. *Embo Reports,* 10**,** 125-U10.

Inouye, S., G. Duffaud & M. Inouye (1986) Structural requirement at the cleavage site for efficient processing of the lipoprotein secretory precursor of *Escherichia coli*. *J Biol Chem,* 261**,** 10970-5.

Ishii, T., T. Matsunaga & N. Hayashi (2001) Formation of rhamnogalacturonan II-borate dimer in pectin determines cell wall thickness of pumpkin tissue. *Plant Physiol,* 126**,** 1698-705.

Ishii, T., T. Matsunaga, P. Pellerin, M. A. O'Neill, A. Darvill & P. Albersheim (1999) The plant cell wall polysaccharide rhamnogalacturonan II self-assembles into a covalently cross-linked dimer. *Journal of Biological Chemistry,* 274**,** 13098-13104.

Jenkins, J., L. Lo Leggio, G. Harris & R. Pickersgill (1995) Beta-glucosidase, beta-galactosidase, family A cellulases, family F xylanases and two barley glycanases form a superfamily of enzymes with 8-fold beta/alpha architecture and with two conserved glutamates near the carboxy-terminal ends of beta-strands four and seven. *FEBS Lett,* 362**,** 281-5.

Jensen, M. H., H. Otten, U. Christensen, T. V. Borchert, L. L. Christensen, S. Larsen & L. L. Leggio (2010) Structural and biochemical studies elucidate the mechanism of rhamnogalacturonan lyase from *Aspergillus aculeatus*. *J Mol Biol,* 404**,** 100-11.

Jervis, E. J., C. A. Haynes & D. G. Kilburn (1997) Surface diffusion of cellulases and their isolated binding domains on cellulose. *Journal of Biological Chemistry,* 272**,** 24016-24023.

Ji, Q., R. J. Oomen, J. P. Vincken, D. N. Bolam, H. J. Gilbert, L. C. Suurs & R. G. Visser (2004) Reduction of starch granule size by expression of an engineered tandem starch-binding domain in potato plants. *Plant Biotechnol J,* 2**,** 251-60.

Jindou, S., S. Petkun, L. Shimon, E. A. Bayer, R. Lamed & F. Frolow (2007) Crystallization and preliminary diffraction studies of CBM3b of cellobiohydrolase 9A from *Clostridium thermocellum*. *Acta Crystallographica Section F-Structural Biology and Crystallization Communications,* 63**,** 1044-1047.

Jing, H. Q., D. Cockburn, Q. X. Zhang & A. J. Clarke (2009) Production and purification of the isolated family 2a carbohydrate-binding module from *Cellulomonas fimi*. *Protein Expression and Purification,* 64**,** 63-68.

Jones, L., G. B. Seymour & J. P. Knox (1997) Localization of pectic galactan in tomato cell walls using a monoclonal antibody specific to (1->4)-beta-D-galactan. *Plant Physiology,* 113**,** 1405-1412.

Jongkees, S. A. & S. G. Withers (2014) Unusual enzymatic glycoside cleavage mechanisms. *Acc Chem Res,* 47**,** 226-35.

Joseleau, J.-P., J. Comtat & K. Ruel. 1992. Chemical structures of xylans and their interactions in the plant cell walls. In *Xylans and Xylanases,* eds. J. Visser, G. Beldman, M. A. Someren & A. G. Voragen, 1-15. Amsterdam: Elsevier.

Juncker, A. S., H. Willenbrock, G. Von Heijne, S. Brunak, H. Nielsen & A. Krogh (2003) Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci,* 12**,** 1652-62.

Kamada, N., Y. G. Kim, H. P. Sham, B. A. Vallance, J. L. Puente, E. C. Martens & G. Nunez (2012) Regulated virulence controls the ability of a pathogen to compete with the gut microbiota. *Science,* 336**,** 1325-9.

Karimi, S. & O. P. Ward (1989) Comparative-Study of Some Microbial Arabinan-Degrading Enzymes. *Journal of Industrial Microbiology,* 4**,** 173-180.

Kellett, L. E., D. M. Poole, L. M. Ferreira, A. J. Durrant, G. P. Hazlewood & H. J. Gilbert (1990) Xylanase B and an arabinofuranosidase from *Pseudomonas fluorescens* subsp. cellulosa contain identical cellulose-binding domains and are encoded by adjacent genes. *Biochem J,* 272**,** 369-76.

Knox, J. P., P. J. Linstead, J. King, C. Cooper & K. Roberts (1990) Pectin Esterification Is Spatially Regulated Both within Cell-Walls and between Developing-Tissues of Root Apices. *Planta,* 181**,** 512-521.

Kobayashi, M., T. Matoh & J. Azuma (1996) Two chains of rhamnogalacturonan II are cross-linked by borate-diol ester bonds in higher plant cell walls. *Plant Physiology,* 110**,** 1017-1020.

Koropatkin, N. M., E. A. Cameron & E. C. Martens (2012) How glycan metabolism shapes the human gut microbiota. *Nat Rev Microbiol,* 10**,** 323-35.

Koropatkin, N. M., E. C. Martens, J. I. Gordon & T. J. Smith (2008) Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices. *Structure,* 16**,** 1105-15.

Kroon, P. A., G. Williamson, N. M. Fish, D. B. Archer & N. J. Belshaw (2000) A modular esterase from Penicillium funiculosum which releases ferulic acid from plant cell walls and binds crystalline cellulose contains a carbohydrate binding module. *European Journal of Biochemistry,* 267**,** 6740-6752.

Larsbrink, J., A. Izumi, F. M. Ibatullin, A. Nakhai, H. J. Gilbert, G. J. Davies & H. Brumer (2011) Structural and enzymatic characterization of a glycoside hydrolase family 31 alpha-xylosidase from *Cellvibrio japonicus* involved in xyloglucan saccharification. *Biochem J,* 436**,** 567-80.

Lehtio, J., J. Sugiyama, M. Gustavsson, L. Fransson, M. Linder & T. T. Teeri (2003a) The binding specificity and affinity determinants of family 1 and family 3 cellulose binding modules. *Proc Natl Acad Sci U S A,* 100**,** 484-9.

--- (2003b) The binding specificity and affinity determinants of family 1 and family 3 cellulose binding modules. *Proceedings of the National Academy of Sciences of the United States of America,* 100**,** 484-489.

Lejohn, H. B. (1971) Enzyme Regulation, Lysine Pathways and Cell Wall Structures as Indicators of Major Lines of Evolution in Fungi. *Nature,* 231**,** 164-&.

Ley, R. E., D. A. Peterson & J. I. Gordon (2006) Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell,* 124**,** 837-48.

Liepman, A. H., C. J. Nairn, W. G. T. Willats, I. Sorensen, A. W. Roberts & K. Keegstra (2007) Functional genomic analysis supports conservation of function among cellulose synthase-like a gene family members and suggests diverse roles of mannans in plants. *Plant Physiology,* 143**,** 1881-1893.

Liu, Y. S., J. O. Baker, Y. N. Zeng, M. E. Himmel, T. Haas & S. Y. Ding (2011) Cellobiohydrolase Hydrolyzes Crystalline Cellulose on Hydrophobic Faces. *Journal of Biological Chemistry,* 286**,** 11195-11201.

Lombard, V., T. Bernard, C. Rancurel, H. Brumer, P. M. Coutinho & B. Henrissat (2010a) A hierarchical classification of polysaccharide lyases for glycogenomics. *Biochemical Journal,* 432**,** 437-444.

--- (2010b) A hierarchical classification of polysaccharide lyases for glycogenomics. *Biochem J,* 432**,** 437-44.

Lombard, V., H. G. Ramulu, E. Drula, P. M. Coutinho & B. Henrissat (2014) The carbohydrate-active enzymes database (CAZy) in 2013. *Nucleic Acids Research,* 42**,** D490-D495.

Marcus, S. E., A. W. Blake, T. A. Benians, K. J. Lee, C. Poyser, L. Donaldson, O. Leroux, A. Rogowski, H. L. Petersen, A. Boraston, H. J. Gilbert, W. G. Willats & J. P. Knox (2010) Restricted access of proteins to mannan polysaccharides in intact plant cell walls. *Plant J,* 64**,** 191-203.

Martens-Uzunova, E. S. & P. J. Schaap (2009) Assessment of the pectin degrading enzyme network of *Aspergillus niger* by functional genomics. *Fungal Genet Biol,* 46 Suppl 1**,** S170-S179.

Martens, E. C., H. C. Chiang & J. I. Gordon (2008) Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. *Cell Host Microbe,* 4**,** 447-57.

Martens, E. C., N. M. Koropatkin, T. J. Smith & J. I. Gordon (2009a) Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm. *J Biol Chem,* 284**,** 24673-7.

Martens, E. C., E. C. Lowe, H. Chiang, N. A. Pudlo, M. Wu, N. P. McNulty, D. W. Abbott, B. Henrissat, H. J. Gilbert, D. N. Bolam & J. I. Gordon (2011) Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol,* 9**,** e1001221.

Martens, E. C., R. Roth, J. E. Heuser & J. I. Gordon (2009b) Coordinate regulation of glycan degradation and polysaccharide capsule biosynthesis by a prominent human gut symbiont. *J Biol Chem,* 284**,** 18445-57.

Martone, P. T., J. M. Estevez, F. C. Lu, K. Ruel, M. W. Denny, C. Somerville & J. Ralph (2009) Discovery of Lignin in Seaweed Reveals Convergent Evolution of Cell-Wall Architecture. *Current Biology,* 19**,** 169-175.

McCartney, L., A. W. Blake, J. Flint, D. N. Bolam, A. B. Boraston, H. J. Gilbert & J. P. Knox (2006) Differential recognition of plant cell walls by microbial xylan-specific carbohydrate-binding modules. *Proc Natl Acad Sci U S A,* 103**,** 4765-70.

McCartney, L., H. J. Gilbert, D. N. Bolam, A. B. Boraston & J. P. Knox (2004) Glycoside hydrolase carbohydrate-binding modules as molecular probes for the analysis of plant cell wall polymers. *Anal Biochem,* 326**,** 49-54.

McCartney, L., A. P. Ormerod, M. J. Gidley & J. P. Knox (2000) Temporal and spatial regulation of pectic (1-->4)-beta-D-galactan in cell walls of developing pea cotyledons: implications for mechanical properties. *Plant J,* 22**,** 105-13.

Mccleary, B. V. (1988) Carob and Guar Galactomannans. *Methods in Enzymology,* 160**,** 523-527.

McKie, V. A., J. P. Vincken, A. G. Voragen, L. A. van den Broek, E. Stimson & H. J. Gilbert (2001) A new family of rhamnogalacturonan lyases contains an enzyme that binds to cellulose. *Biochem J,* 355**,** 167-77.

McLean, B. W., A. B. Boraston, D. Brouwer, N. Sanaie, C. A. Fyfe, R. A. J. Warren, D. G. Kilburn & C. A. Haynes (2002) Carbohydrate-binding modules recognize fine substructures of cellulose. *Journal of Biological Chemistry,* 277**,** 50245-50254.

McLean, B. W., M. R. Bray, A. B. Boraston, N. R. Gilkes, C. A. Haynes & D. G. Kilburn (2000) Analysis of binding of the family 2a carbohydrate-binding module from *Cellulomonas fimi* xylanase 10A to cellulose: specificity and identification of functionally important amino acid residues. *Protein Eng,* 13**,** 801-9.

McNeil, M., A. G. Darvill, S. C. Fry & P. Albersheim (1984) Structure and function of the primary cell walls of plants. *Annu Rev Biochem,* 53**,** 625-63.

McNulty, N. P., M. Wu, A. R. Erickson, C. Pan, B. K. Erickson, E. C. Martens, N. A. Pudlo, B. D. Muegge, B. Henrissat, R. L. Hettich & J. I. Gordon (2013) Effects of diet on resource utilization by a model human gut microbiota containing Bacteroides cellulosilyticus WH2, a symbiont with an extensive glycobiome. *PLoS Biol,* 11**,** e1001637.

Melton, L. D., M. Mcneil, A. G. Darvill, P. Albersheim & A. Dell (1986) Structure of Plant-Cell Walls .17. Structural Characterization of Oligosaccharides Isolated from the Pectic Polysaccharide Rhamnogalacturonan-Ii. *Carbohydrate Research,* 146**,** 279-305.

Mertens, J. A. & M. J. Bowman (2011) Expression and characterization of fifteen *Rhizopus oryzae* 99-880 polygalacturonase enzymes in Pichia pastoris. *Curr Microbiol,* 62**,** 1173-8.

Miller, G. L. (1959) The use of dinitrosalicylic acid reagent for thr determination of reducing sugar. *Anal. Chem.,* 31**,** 426-428.

Millward-Sadler, S. J., K. Davidson, G. P. Hazlewood, G. W. Black, H. J. Gilbert & J. H. Clarke (1995) Novel cellulose-binding domains, NodB homologues and conserved

modular architecture in xylanases from the aerobic soil bacteria *Pseudomonas* fluorescens subsp. cellulosa and Cellvibrio mixtus. *Biochem J,* 312 (Pt 1)**,** 39-48.

Miroux, B. & J. E. Walker (1996) Over-production of proteins in *Escherichia coli*: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J Mol Biol,* 260**,** 289-98.

Mizutani, K., V. O. Fernandes, S. Karita, A. S. Luis, M. Sakka, T. Kimura, A. Jackson, X. Zhang, C. M. Fontes, H. J. Gilbert & K. Sakka (2012) Influence of a mannan binding family 32 carbohydrate binding module on the activity of the appended mannanase. *Appl Environ Microbiol,* 78**,** 4781-7.

Mizutani, K., M. Sakka, T. Kimura & K. Sakka (2014) Essential role of a family-32 carbohydrate-binding module in substrate recognition by *Clostridium thermocellum* mannanase CtMan5A. *FEBS Lett,* 588**,** 1726-30.

Mohnen, D. (2008) Pectin structure and biosynthesis. *Curr Opin Plant Biol,* 11**,** 266-77.

Montanier, C., J. E. Flint, D. N. Bolam, H. Xie, Z. Liu, A. Rogowski, D. P. Weiner, S. Ratnaparkhe, D. Nurizzo, S. M. Roberts, J. P. Turkenburg, G. J. Davies & H. J. Gilbert (2010) Circular permutation provides an evolutionary link between two families of calcium-dependent carbohydrate binding modules. *J Biol Chem,* 285**,** 31742-54.

Montanier, C., V. A. Money, V. M. Pires, J. E. Flint, B. A. Pinheiro, A. Goyal, J. A. Prates, A. Izumi, H. Stalbrand, C. Morland, A. Cartmell, K. Kolenova, E. Topakas, E. J. Dodson, D. N. Bolam, G. J. Davies, C. M. Fontes & H. J. Gilbert (2009a) The active site of a carbohydrate esterase displays divergent catalytic and noncatalytic binding functions. *PLoS Biol,* 7**,** e71.

Montanier, C., A. L. van Bueren, C. Dumon, J. E. Flint, M. A. Correia, J. A. Prates, S. J. Firbank, R. J. Lewis, G. G. Grondin, M. G. Ghinet, T. M. Gloster, C. Herve, J. P. Knox, B. G. Talbot, J. P. Turkenburg, J. Kerovuo, R. Brzezinski, C. M. Fontes, G. J. Davies, A. B. Boraston & H. J. Gilbert (2009b) Evidence that family 35 carbohydrate binding modules display conserved specificity but divergent function. *Proc Natl Acad Sci U S A,* 106**,** 3065-70.

Montanier, C. Y., M. A. Correia, J. E. Flint, Y. Zhu, A. Basle, L. S. McKee, J. A. Prates, S. J. Polizzi, P. M. Coutinho, R. J. Lewis, B. Henrissat, C. M. Fontes & H. J. Gilbert (2011) A novel, noncatalytic carbohydrate-binding module displays specificity for galactose-containing polysaccharides through calcium-mediated oligomerization. *J Biol Chem,* 286**,** 22499-509.

Moore, P. J. & L. A. Staehelin (1988) Immunogold Localization of the Cell-Wall-Matrix Polysaccharides Rhamnogalacturonan-I and Xyloglucan during Cell Expansion and Cytokinesis in Trifolium-Pratense L - Implication for Secretory Pathways. *Planta,* 174**,** 433-445.

Mukhiddinov, Z. K., D. K. Khalikov, F. T. Abdusamiev & C. C. Avloev (2000) Isolation and structural characterization of a pectin homo and ramnogalacturonan. *Talanta,* 53**,** 171-6.

Murzin, A. G., A. M. Lesk & C. Chothia (1992) beta-Trefoil fold. Patterns of structure and sequence in the Kunitz inhibitors interleukins-1 beta and 1 alpha and fibroblast growth factors. *J Mol Biol,* 223**,** 531-43.

Mutter, M., G. Beldman, S. M. Pitson, H. A. Schols & A. G. J. Voragen (1998) Rhamnogalacturonan alpha-D-galactopyranosyluronohydrolase - An enzyme that specifically removes the terminal nonreducing galacturonosyl residue in rhamnogalacturonan regions of pectin. *Plant Physiology,* 117**,** 153-163.

Naran, R., G. Chen & N. C. Carpita (2008) Novel rhamnogalacturonan I and arabinoxylan polysaccharides of flax seed mucilage. *Plant Physiol,* 148**,** 132-41.

Nielsen, H., J. Engelbrecht, S. Brunak & G. von Heijne (1997) A neural network method for identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Int J Neural Syst,* 8**,** 581-99.

Nielsen, H. & A. Krogh (1998) Prediction of signal peptides and signal anchors by a hidden Markov model. *Proc Int Conf Intell Syst Mol Biol,* 6**,** 122-30.

Norrander, J., T. Kempe & J. Messing (1983) Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. *Gene,* 26**,** 101-6.

Notenboom, V., A. B. Boraston, P. Chiu, A. C. Freelove, D. G. Kilburn & D. R. Rose (2001) Recognition of cello-oligosaccharides by a family 17 carbohydrate-binding module: an X-ray crystallographic, thermodynamic and mutagenic study. *J Mol Biol,* 314**,** 797-806.

Novoa De Armas, H., C. Verboven, C. De Ranter, J. Desair, A. Vande Broek, J. Vanderleyden & A. Rabijns (2004) Azospirillum irakense pectate lyase displays a toroidal fold. *Acta Crystallogr D Biol Crystallogr,* 60**,** 999-1007.

O'Brien, P. J. & D. Herschlag (1999) Catalytic promiscuity and the evolution of new enzymatic activities. *Chem Biol,* 6**,** R91-R105.

O'Neill, M. A., S. Eberhard, P. Albersheim & A. G. Darvill (2001) Requirement of borate cross-linking of cell wall rhamnogalacturonan II for *Arabidopsis* growth. *Science,* 294**,** 846-9.

Ogawa, K., T. Hayashi & K. Okamura (1990) Conformational analysis of xyloglucans. *Int J Biol Macromol,* 12**,** 218-22.

ONeill, M. A., D. Warrenfeltz, K. Kates, P. Pellerin, T. Doco, A. G. Darvill & P. Albersheim (1996) Rhamnogalacturonan-II, a pectic polysaccharide in the walls of growing plant cell, forms a dimer that is covalently cross-linked by a borate ester - *In vitro* conditions for the formation and hydrolysis of the dimer. *Journal of Biological Chemistry,* 271**,** 22923-22930.

Oomen, R. J., C. H. Doeswijk-Voragen, M. S. Bush, J. P. Vincken, B. Borkhardt, L. A. van den Broek, J. Corsar, P. Ulvskov, A. G. Voragen, M. C. McCann & R. G. Visser (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**,** 403-13.

Ordaz-Ortiz, J. J., S. E. Marcus & J. P. Knox (2009) Cell wall microstructure analysis implicates hemicellulose polysaccharides in cell adhesion in tomato fruit pericarp parenchyma. *Mol Plant,* 2**,** 910-21.

Pace, C. N., F. Vajdos, L. Fee, G. Grimsley & T. Gray (1995) How to measure and predict the molar absorption coefficient of a protein. *Protein Sci,* 4**,** 2411-23.

Pahl, M. V., B. D. Culver & N. D. Vaziri (2005) Boron and the kidney. *Journal of Renal Nutrition,* 15**,** 362-370.

Parenicova, L., J. A. Benen, H. C. Kester & J. Visser (2000) pgaA and pgaB encode two constitutively expressed endopolygalacturonases of *Aspergillus niger*. *Biochem J,* 345 Pt 3**,** 637-44.

Pattathil, S., U. Avci, D. Baldwin, A. G. Swennes, J. A. McGill, Z. Popper, T. Bootten, A. Albert, R. H. Davis, C. Chennareddy, R. Dong, B. O'Shea, R. Rossi, C. Leoff, G. Freshour, R. Narra, M. O'Neil, W. S. York & M. G. Hahn (2010) A comprehensive toolkit of plant cell wall glycan-directed monoclonal antibodies. *Plant Physiol,* 153**,** 514-25.

Pauly, M., P. Albersheim, A. Darvill & W. S. York (1999) Molecular domains of the cellulose/xyloglucan network in the cell walls of higher plants. *Plant J,* 20**,** 629-39.

Pell, G. 2004. Structural and functional analysis of *Cellvibrio* hemicellulases. In *School of Cell and Molecular Biosciences*, 285. Newcastle-upon-Tyne: Newcastly-upon-Tyne.

Pell, G., L. Szabo, S. J. Charnock, H. Xie, T. M. Gloster, G. J. Davies & H. J. Gilbert (2004) Structural and biochemical analysis of *Cellvibrio japonicus* xylanase 10C: how variation in substrate-binding cleft influences the catalytic profile of family GH-10 xylanases. *J Biol Chem,* 279**,** 11777-88.

Pellerin, P., T. Doco, S. Vidal, P. Williams, J. M. Brillouet & M. A. ONeill (1996) Structural characterization of red wine rhamnogalacturonan II. *Carbohydrate Research,* 290**,** 183-197.

Perez, S., M. A. Rodriguez-Carvajal & T. Doco (2003) A complex plant cell wall polysaccharide: rhamnogalacturonan II. A structure in quest of a function. *Biochimie,* 85**,** 109-121.

Puls, J. & J. Schuseil (1993a) Chemistry of Hemicelluloses - Relationship between Hemicellulose Structure and Enzymes Required for Hydrolysis. *Hemicellulose and Hemicellulases,* 4**,** 1-27.

---. 1993b. Chemistry of hemicelluloses: relationship between hemicellulose structure and enzymes required for hydrolysis. In *Hemicellulose and Hemicellulases,* eds. M. P. Coughlan & G. P. Hazlewood. London: Portland Press Ltd.

Ragauskas, A. J., C. K. Williams, B. H. Davison, G. Britovsek, J. Cairney, C. A. Eckert, W. J. Frederick, Jr., J. P. Hallett, D. J. Leak, C. L. Liotta, J. R. Mielenz, R. Murphy, R. Templer & T. Tschaplinski (2006) The path forward for biofuels and biomaterials. *Science,* 311**,** 484-9.

Ralet, M. C., V. Dronnet, H. C. Buchholt & J. F. Thibault (2001) Enzymatically and chemically de-esterified lime pectins: characterisation, polyelectrolyte behaviour and calcium binding properties. *Carbohydrate Research,* 336**,** 117-125.

Ridley, B. L., M. A. O'Neill & D. Mohnen (2001) Pectins: structure, biosynthesis, and oligogalacturonide-related signaling. *Phytochemistry,* 57**,** 929-67.

Ried, J. L. & A. Collmer (1986) Comparison of pectic enzymes produced by Erwinia chrysanthemi, Erwinia carotovora subsp. carotovora, and Erwinia carotovora subsp. atroseptica. *Appl Environ Microbiol,* 52**,** 305-10.

Rouvinen, J., T. Bergfors, T. Teeri, J. K. Knowles & T. A. Jones (1990) Three-dimensional structure of cellobiohydrolase II from Trichoderma reesei. *Science,* 249**,** 380-6.

Rye, C. S. & S. G. Withers (2000) Glycosidase mechanisms. *Curr Opin Chem Biol,* 4**,** 573-80.

Salyers, A. A., J. R. Vercellotti, S. E. West & T. D. Wilkins (1977) Fermentation of mucin and plant polysaccharides by strains of Bacteroides from the human colon. *Appl Environ Microbiol,* 33**,** 319-22.

Sambrook, J., E. F. Fritsch & T. Maniatis. 1989. *Molecular cloning; a laboratory manual.* New York: Cold Springs Harbour Laboratory Press.

Scheller, H. V., J. K. Jensen, S. O. Sorensen, J. Harholt & N. Geshi (2007) Biosynthesis of pectin. *Physiologia Plantarum,* 129**,** 283-295.

Seveno, M., A. Voxeur, C. Rihouey, A. M. Wu, T. Ishii, C. Chevalier, M. C. Ralet, A. Driouich, A. Marchant & P. Lerouge (2009) Structural characterisation of the pectic polysaccharide rhamnogalacturonan II using an acidic fingerprinting methodology. *Planta,* 230**,** 947-957.

Shallom, D. & Y. Shoham (2003) Microbial hemicellulases. *Curr Opin Microbiol,* 6**,** 219-28.

Siezen, R. J. & M. Kleerebezem (2011) The human gut microbiome: are we our enterotypes? *Microb Biotechnol,* 4**,** 550-3.

Sonnenburg, E. D., J. L. Sonnenburg, J. K. Manchester, E. E. Hansen, H. C. Chiang & J. I. Gordon (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**,** 8834-9.

Southall, S. M., P. J. Simpson, H. J. Gilbert, G. Williamson & M. P. Williamson (1999) The starch-binding domain from glucoamylase disrupts the structure of starch. *FEBS Lett,* 447**,** 58-60.

Steffan, W., P. Kovac, P. Albersheim, A. G. Darvill & M. G. Hahn (1995) Characterization of a Monoclonal-Antibody That Recognizes an Arabinosylated (1-]6)-Beta-D-Galactan Epitope in Plant-Complex Carbohydrates. *Carbohydrate Research,* 275**,** 295-307.

Stock, A. M., V. L. Robinson & P. N. Goudreau (2000) Two-component signal transduction. *Annu Rev Biochem,* 69**,** 183-215.

Studier, F. W. & B. A. Moffatt (1986) Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol,* 189**,** 113-30.

Suykerbuyk, M. E., H. C. Kester, P. J. Schaap, H. Stam, W. Musters & J. Visser (1997) Cloning and characterization of two rhamnogalacturonan hydrolase genes from *Aspergillus niger*. *Appl Environ Microbiol,* 63**,** 2507-15.

Tailford, L. E., V. M. Ducros, J. E. Flint, S. M. Roberts, C. Morland, D. L. Zechel, N. Smith, M. E. Bjornvad, T. V. Borchert, K. S. Wilson, G. J. Davies & H. J. Gilbert (2009)

Understanding how diverse beta-mannanases recognize heterogeneous substrates. *Biochemistry,* 48**,** 7009-18.

Tanaka, M., A. Abe & T. Uchida (1981) Substrate-Specificity of Alpha-L-Arabinofuranosidase from Plant Scopolia-Japonica Calluses and a Suggestion with Reference to the Structure of Beet Araban. *Biochimica Et Biophysica Acta,* 658**,** 377-386.

Thakur, B. R., R. K. Singh & A. K. Handa (1997) Chemistry and uses of pectin--a review. *Crit Rev Food Sci Nutr,* 37**,** 47-73.

Tomme, P., A. Boraston, B. McLean, J. Kormos, A. L. Creagh, K. Sturch, N. R. Gilkes, C. A. Haynes, R. A. Warren & D. G. Kilburn (1998) Characterization and affinity applications of cellulose-binding domains. *J Chromatogr B Biomed Sci Appl,* 715**,** 283-96.

Topakas, E., S. Kyriakopoulos, P. Biely, J. Hirsch, C. Vafiadi & P. Christakopoulos (2010) Carbohydrate esterases of family 2 are 6-O-deacetylases. *FEBS Lett,* 584**,** 543-8.

Tormo, J., R. Lamed, A. J. Chirino, E. Morag, E. A. Bayer, Y. Shoham & T. A. Steitz (1996) Crystal structure of a bacterial family-III cellulose-binding domain: a general mechanism for attachment to cellulose. *EMBO J,* 15**,** 5739-51.

Vaaje-Kolstad, G., S. J. Horn, D. M. F. van Aalten, B. Synstad & V. G. H. Eijsink (2005) The non-catalytic chitin-binding protein CBP21 from *Serratia marcescens* is essential for chitin degradation. *Journal of Biological Chemistry,* 280**,** 28492-28497.

Vaaje-Kolstad, G., B. Westereng, S. J. Horn, Z. L. Liu, H. Zhai, M. Sorlie & V. G. H. Eijsink (2010) An Oxidative Enzyme Boosting the Enzymatic Conversion of Recalcitrant Polysaccharides. *Science,* 330**,** 219-222.

van den Brink, J. & R. P. de Vries (2011) Fungal enzyme sets for plant polysaccharide degradation. *Appl Microbiol Biotechnol,* 91**,** 1477-92.

Verhertbruggen, Y., S. E. Marcus, A. Haeger, R. Verhoef, H. A. Schols, B. V. McCleary, L. McKee, H. J. Gilbert & J. P. Knox (2009) Developmental complexity of arabinan polysaccharides and their processing in plant cell walls. *Plant J,* 59**,** 413-25.

Waltz, F., G. Wissmann, J. Lippke, A. M. Schneider, H. C. Schwarz, A. Feldhoff, S. Eiden & P. Behrens (2012) Evolution of the Morphologies of Zinc Oxide Mesocrystals Under the Influence of Natural Polysaccharides. *Crystal Growth & Design,* 12**,** 3066-3075.

Whetten, R. & R. Sederoff (1995) Lignin Biosynthesis. *Plant Cell,* 7**,** 1001-1013.

Willats, W. G., L. McCartney, W. Mackie & J. P. Knox (2001) Pectin: cell biology and prospects for functional analysis. *Plant Mol Biol,* 47**,** 9-27.

Willats, W. G. T., G. Limberg, H. C. Buchholt, G. J. van Alebeek, J. Benen, T. M. I. E. Christensen, J. Visser, A. Voragen, J. D. Mikkelsen & J. P. Knox (2000) Analysis of pectin structure part 2 - Analysis of pectic epitopes recognised by hybridoma and phage display monoclonal antibodies using defined oligosaccharides, polysaccharides, and enzymatic degradation. *Carbohydrate Research,* 327**,** 309-320.

Wise, H. Z., I. M. Saxena & R. M. Brown (2011) Isolation and characterization of the cellulose synthase genes PpCesA6 and PpCesA7 in Physcomitrella patens. *Cellulose,* 18**,** 371-384.

Wollowski, I., G. Rechkemmer & B. L. Pool-Zobel (2001) Protective role of probiotics and prebiotics in colon cancer. *Am J Clin Nutr,* 73**,** 451S-455S.

Xu, J., M. K. Bjursell, J. Himrod, S. Deng, L. K. Carmichael, H. C. Chiang, L. V. Hooper & J. I. Gordon (2003) A genomic view of the human-Bacteroides thetaiotaomicron symbiosis. *Science,* 299**,** 2074-6.

Zechel, D. L. & S. G. Withers (2000) Glycosidase mechanisms: anatomy of a finely tuned catalyst. *Acc Chem Res,* 33**,** 11-8.

Zhang, X. Y., A. Rogowski, L. Zhao, M. G. Hahn, U. Avci, J. P. Knox & H. J. Gilbert (2014) Understanding How the Complex Molecular Architecture of Mannan-degrading Hydrolases Contributes to Plant Cell Wall Degradation. *Journal of Biological Chemistry,* 289**,** 2002-2012.

Appendix A. Chemicals, media and enzymes used in this study

A1 Chemicals

Amersham-Boehringer Mannheim
- 2'-Deoxyadenosine 5'-triphosphate (dATP)
- 2'-Deoxycytidine 5'-triphosphate (dCTP)
- 2'-Deoxyguanosine 5'-triphosphate (dGTP)
- 2'-Deoxythymidine 5'-triphosphate (dTTP)

BioGene
- Electrophoresis grade Agarose

British Drug Houses (BDH)
- Acetic acid (Glacial)
- Acrylamide solution (40% (w/v); Electran)
- Boric acid
- Bromophenol Blue
- Citric acid
- Calcium chloride
- Chloroform
- Dimethylformamide
- Ethanol (industrial grade)
- Hydrochloric acid
- Isopropanol
- Magnesium chloride
- Magnesium sulphate
- Mineral oil
- Methanol
- Polyethelene glycol MW 8000 (PEG-8000)
- Polyethelene glycol MW 550 monomethyl ether (PEG 550 mme)
- Potassium dihydrogen orthophosphate
- Sodium acetate
- Sodium chloride
- Sulphuric acid
- Triton X-100

Calbiochem
- Kifunensine

Dextra
- GlcNAcβ1-2Man
- Man3(a)
- Man3(b)
- $Man_9(GlcNAc)_2$
- N-acetyl lactosamine
- α-1,2-Mannobiose
- α-1,3-Mannobiose
- α-1,4-Mannobiose
- α-1,6-Mannobiose

Fisons
- 46/48% (w/v) NaOH
- Sodium acetate trihydrate

Hampton Research
- Aqua Sil

James Burrough (F.A.D.) Ltd
- Ethanol

Megazyme
- Rye Arabinoxylan (low viscosity)
- Tamarind Xyloglucan
- Wheat Arabionoxylan (medium viscosity)
- Xylobiose
- Xylohexaose
- Xylopentaose
- Xylotetraose
- Xylotriose

Melford Laboratories
- Isopropyl-β-D-thiogalactoside (IPTG)
- Urea (ultrapure)
- HEPES

GE Health Care
- Adenosine 5'-triphosphate (ATP)
- Agarose (ultrapure)

Sigma
- 1,4-Dideoxy-1,4-imino-D-Mannitol
- 1-Deoxymannojirimycin
- 2,4- methane-pentanediol (MPD)
- 2,4-Dinitrophenol
- 3,5-Dinitrosalasylic acid (DNSA)
- 4-nitrophenyl α-D-mannopyranoside (PNP-Man)
- Ammonium persulphate
- Ampicillin
- Beechwood xylan
- Birchwood xyaln
- Bis tris propane
- Bovine serum albumin, fraction V (BSA)
- Castanospermine
- Chicken ovalbumin
- Chloramphenicol
- Coomassie Brilliant Blue G
- Copper sulphate
- DEAE-Trisacryl

Deoxynojirimycin
D-Glucose
Dimethylsulphoxide
di-Sodium hydrogen phosphate
Dithiothrietol (DTT)
D-Lactate
D-Mannose
Ethelene diamine tetra-aceticacid, disodium salt (EDTA)
Ethidium bromide
Ethylene glycol
Gadolinium chloride
Glyceraldehyde-3-phosphate
Glycerol
Guanosine 5’ diphospho-D-mannose
Guanosine 5’ diphosphoglucose
Guanosine 5’ diphospho-β-L-fucose
Imidazole
Kanamycin
L-Fucose
Lithium acetate
Lithium chloride
$Man_9(GlcNAc)_2$
Manganese Chloride
Mineral oil (light)
N,N,N’,N,-Tetramethylethylene diamine (TEMED)
N-Butyl-deoxynojirimycin
Nicotinamide adenine dinucleotide-reduced
Para-nitrophenyl glucopyranoside
Para-nitrophenyl mannopyranoside
Phenol
Phosphoenol pyruvate potassium salt (PEP-K)
Polyethylene glycol MW 3350
Potasium acetate
Potassium Chloride
Potassium thiocyanate
Rochelle salts
Sodium bicarbonate
Sodium carbonate
Sodium dihydrogen orthophosphate
Sodium dodecyl sulphate (SDS)
Sodium fluoride
Sodium sulphate
Sucrose (nuclease free)
Swinsonine
Trizma base (Tris)
Uridine 5’ diphospho-D-mannose
Yeast mannan
Zinc sulphate
β-Mercaptoethanol

A2 Media

Difco
- Bacto®tryptone
- Bacto®yeast extract

Oxoid
- Bacteriological Agar No.1

A3 Enzymes

MBI Fermentas
- DNA restriction endonucleases

Invitrogen
- Bacteriophage T4 DNA ligase

Novagen
- KOD HotStart DNA polymerase

Sigma
- Glucose-6-phosphate dehydrogenase
- Hexokinase
- Phosphoglucose isomerase
- Phosphomannose isomerase

Stratagene
- *Pfu*Turbo DNA polymerase
- *Dpn*I restriction endonuclease

A4 DNA

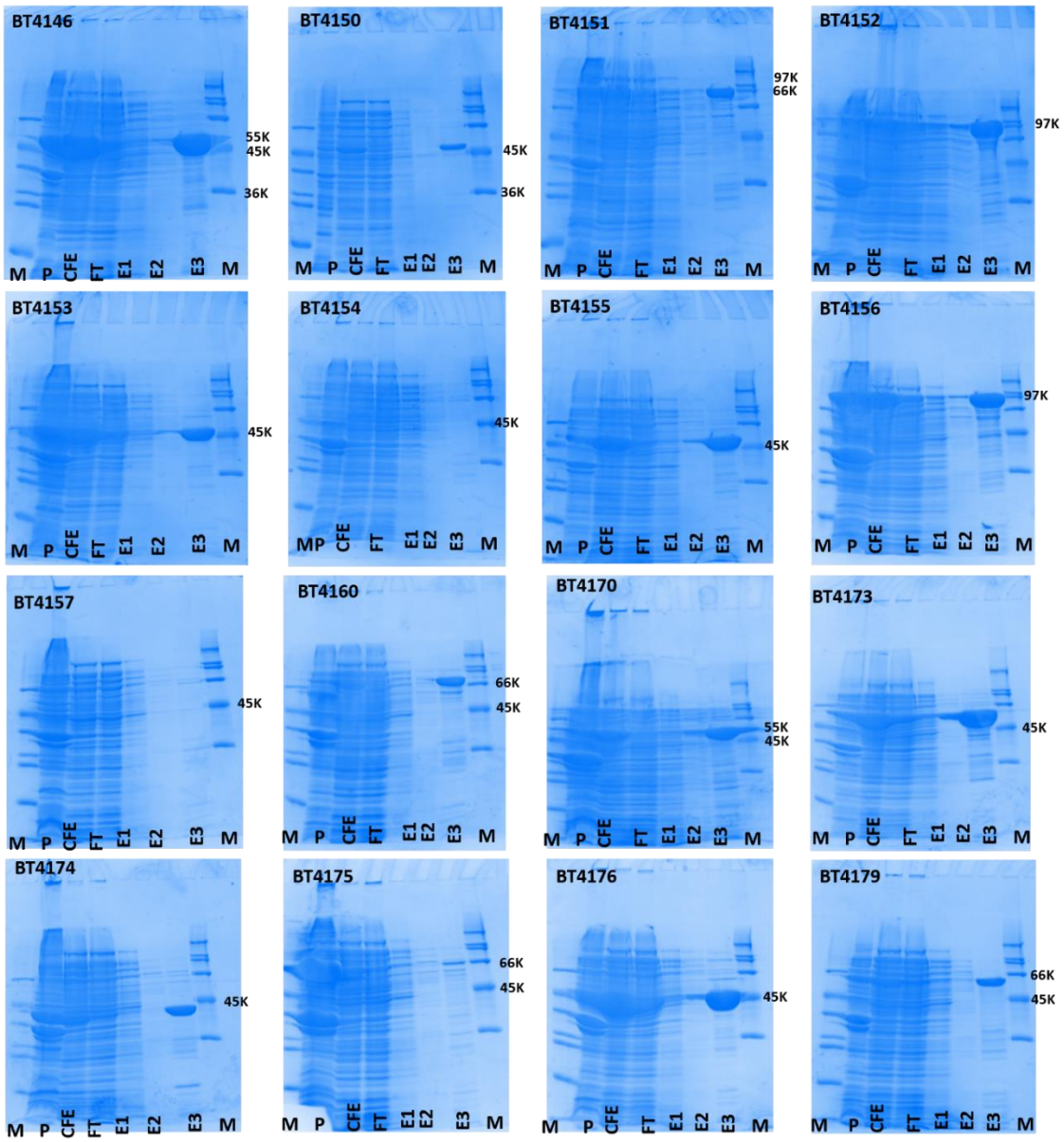
MBI Fermaentas
- *Lambda* bacteriophage DNA, digested with *Eco*RI and *Hin*dIII

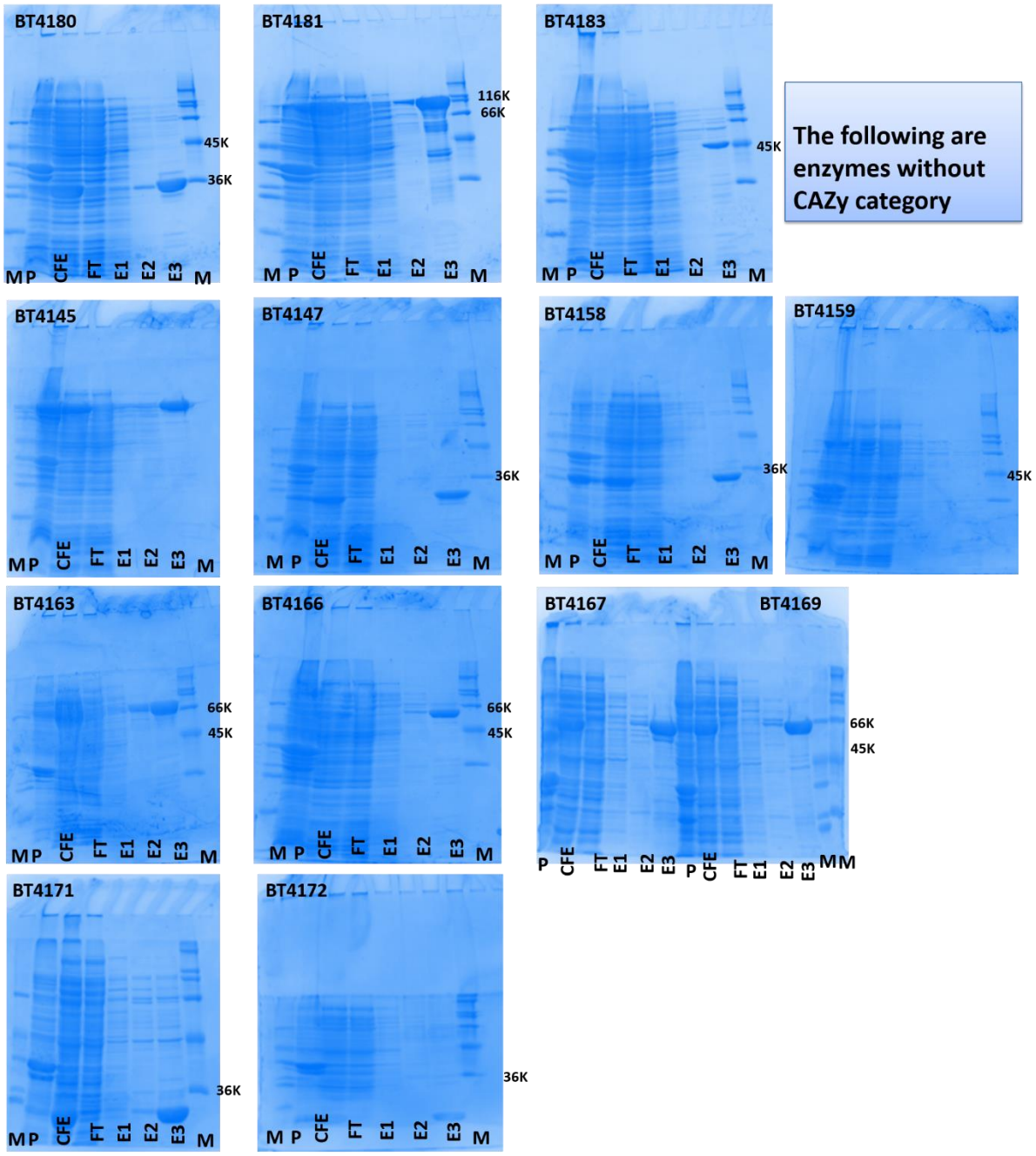
MWG Biotech
- All primers

A5 Kits

Qiagen
- Plasmid Mini kit
- Plasmid Midi kit
- Qiaquik Gel Extraction kit
- Qiaquik PCR Purification kit

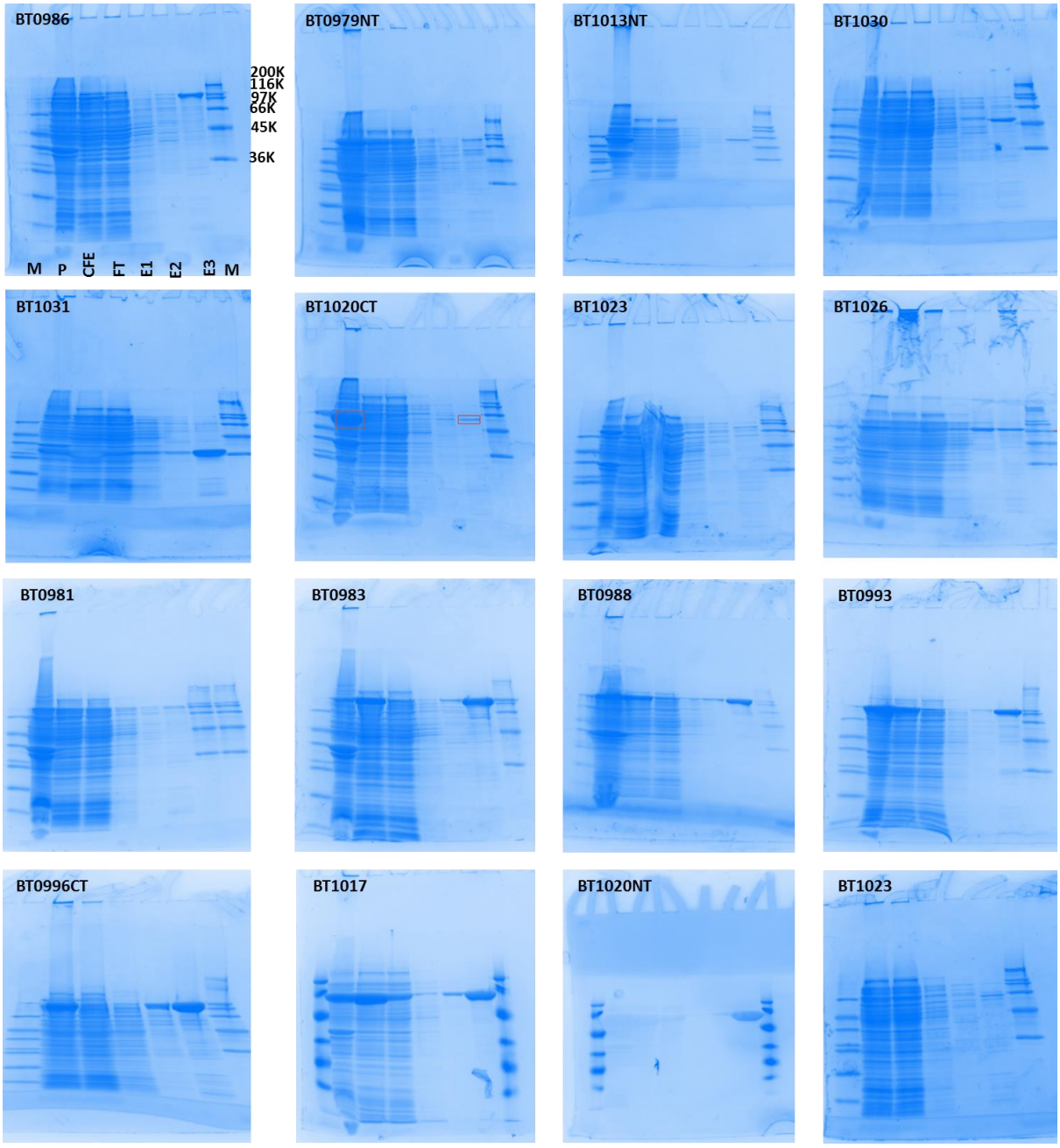
Appendix B. Example of IMAC purified recombinant enzymes used in the RGI project subjected to SDS-PAGE.

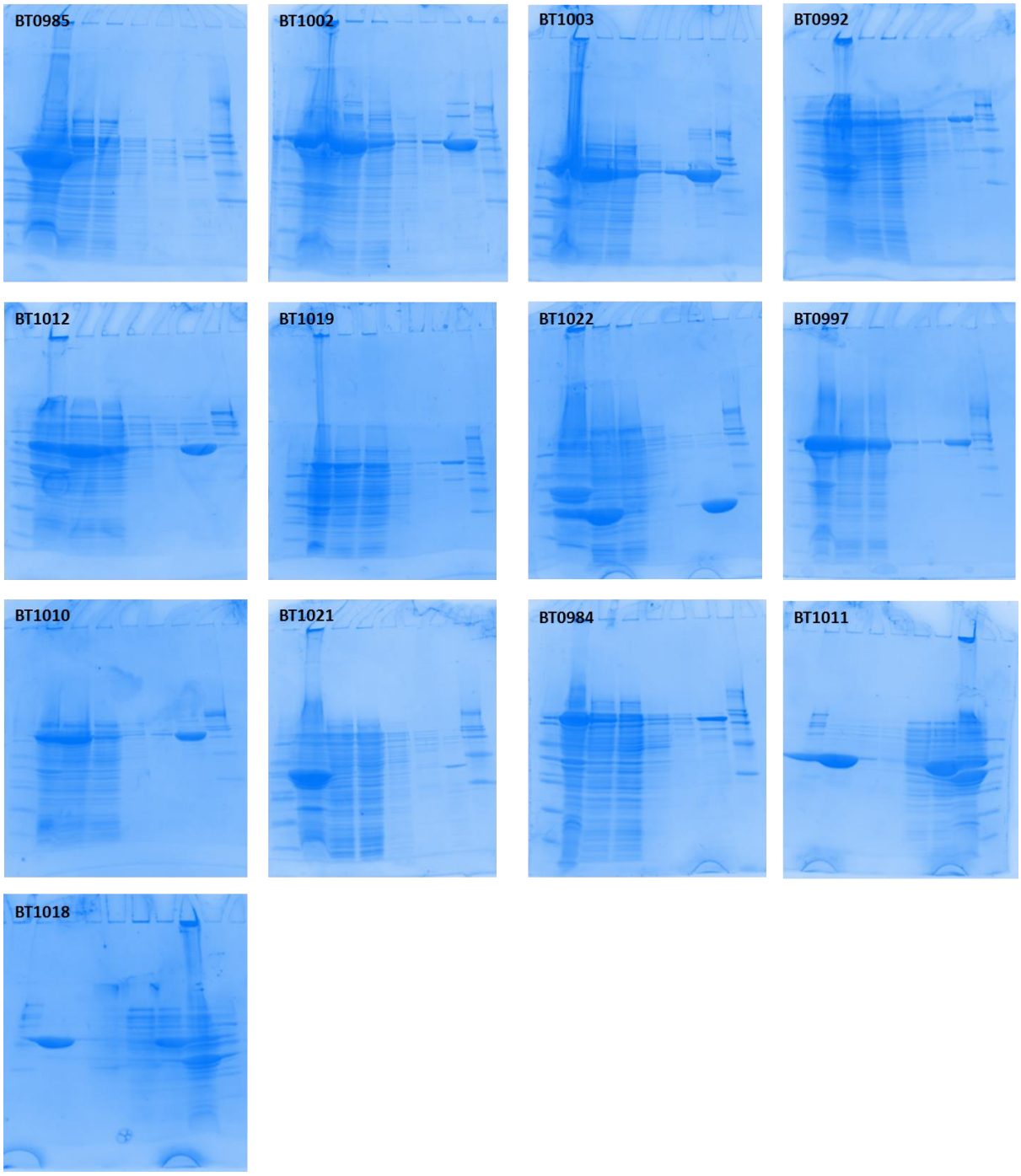




Recombinant expression was in *E. coli* Tuner cells under conditions detailed in Chapter 2.2.; M, Marker; P, insoluble pellet; CFE, Cell free extract; E1-3, Elutions by imidazole gradient of 0 mM, 5 mM and 150 mM, respectively, in Talon buffer (10 mM Tris-HCl pH 8.0 containing 300 mM NaCl).

Appendix C. Example of IMAC purified recombinant enzymes used in the RGII project subjected to SDS-PAGE.





Recombinant expression was in *E. coli* Tuner cells under conditions detailed in Chapter 2.2.; M, Marker; P, insoluble pellet; CFE, Cell free extract; E1-3, Elutions by imidazole gradient of 0 mM, 5 mM and 150 mM, respectively, in Talon buffer (10 mM Tris-HCl pH 8.0 containing 300 mM NaCl).