The Role of Enzymes and Binding Modules in the Degradation of Eukaryotic, Microbial and Plant cell walls

A thesis submitted for the degree of Doctor of Philosophy

Newcastle University

2011-2015

Max Temple

Institute for Cell and Molecular Biosciences

Acknowledgments

First and foremost, I am extremely grateful to Prof. Harry Gilbert for his guidance (and patience) throughout my PhD. I am very appreciative for the opportunity Harry provided me, and for what I have been able to achieve under his tutelage.

I would also like to thank Dr. David Bolam for his numerous contributions and helpful advisement during my PhD.

Mr. Carl Morland has been of tremendous technical support during my PhD. As well as cloning many of the genes used in this study, Carl is a seemingly infinite source of technical wisdom. His experience and advice has been instrumental during this PhD.

I would also like to acknowledge and thank Dr. Arnaud Baslé for his structural biology expertise during my PhD, performing data collection and solving/refining the crystal structures presented in this thesis.

I would like to thank everyone in lab M2035 who I have had the pleasure of knowing, both past and present (and far too numerous to mention!). I am grateful not only for their scientific support during this PhD, but also for their friendship which has been a constant source of support during the sometimes testing moments that science can provide.

Finally, thank you to Emmie and my family for all the love and support you have given me throughout the last four years.

Abstract

The microbial enzymes that depolymerize complex carbohydrates are of industrial significance particularly in the biofuels and biorefinery sectors. In the human large bowel glycan utilization plays a critical role in defining the composition of the human gut microbial community (microbiota) which, in turn, has a significant impact on health. A central feature of these processes is the specificity of the enzymes and the non-catalytic carbohydrate binding modules (CBMs) that contribute to glycan degradation. This thesis describes research designed to understand the mechanisms by which CBMs and glycoside hydrolases contribute to glycan degradation and how this impacts on the structure of the microbiota.

The first results chapter describes the biochemical properties and structural basis for the specificity displayed by two CBMs appended to the glucanase of a rumen bacterium. The sequence of the two CBMs are >75% identical and display essentially identical ligand specificities. Isothermal titration calorimetry revealed that the two proteins bound to a range of β1,4-glucans (cellulose) and, β1,3-β1,4-mixed linked glucans, displaying highest affinity for xyloglucan, a β1,4-glucan decorated with α1,6-xylose residues. The structures of the two CBMs reveal a β-sandwich fold. The ligand binding site comprises the β-sheet that forms the concave surface of the proteins. Binding to the backbone chains of β-glucans is mediated primarily by five aromatic residues that also make hydrophobic interactions with the xylose side chains of xyloglucan, conferring the distinctive specificity of the CBMs for the decorated polysaccharide. Significantly, and in contrast to other CBMs that recognize β-glucans, CBM65A utilizes different polar residues to bind cellulose and mixed linked glucans. Thus, Gln106 is central to cellulose recognition, but is not required for binding to mixed linked glucans. This chapter reveals the mechanism by which β-glucan-specific CBMs can distinguish between linear and mixed linked glucans, and show how these CBMs can exploit an extensive hydrophobic platform to target the side chains of decorated β-glucans.

In the second chapter the enzymes that contribute to the degradation of α-mannan, a prominent component of yeast cell walls, were studied. These enzymes were derived from *Bacteroides thetaiotaomicron*, a member of the microbiota. The data showed that the GH76 endo-α1,6-mannanases presented on the bacterial surface displayed significantly less activity against small mannooligosaccharides compared to the equivalent periplasmic enzymes. All the endo-α1,6-mannanases were only active on the linear backbone of a-mannan, the enzymes were unable to accommodate any side chains. These decorations were partially removed by a poorly expressed and slow acting surface GH92 α-mannosidase. In contrast, in the periplasm a highly active GH38 α-mannosidase rapidly debranched the imported yeast mannan oligosaccharides. The manooligosaccharides generated by the GH76 enzymes were then depolymerized into mannose by a pair of periplasmic exo-acting α1,6-mannosidases that contained only two substrate binding subsites. The biochemical characterization of these enzymes led to the selfish hypothesis in which *B. thetaiotaomicron* maximises deconstruction of yeast mannan in the periplasm, ensuring that the mannose generated will not be available to other organisms in the microbiota. This hypothesis were verified by showing that *B. thetaiotaomicron* was unable to support the growth of other *Bacteroides* sp. (that were able to grow on mannose and, in the case of *B. xylanisolvens*, also on debranched α-mannan) on yeast α-mannan.

In the final results chapter the mechanism by which *B. thetaiotaomicron* utilized β1,6-glucan, a component of the yeast wall, was analysed. Transcriptomic analysis identified a Polysaccharide Utilization Locus (PUL) that was transcribed in response to β1,6-glucan. The PUL encoded two enzymes and two surface glycan binding proteins (SGBPs), one of which was a SusD homologue. The two SGBPs displayed tight specificity for β1,6-glucan over other β-glucans, displaying a preference for ligands that contained >3 glucose units. The surface GH30 enzyme, BT3312, was shown to be an endo-β1,6-glucanase, while the periplasmic GH3 exo-acting β-glucosidase displayed a preference for β1,6-linkages. *B. thetaiotaomicron* accumulated β1,6-glucobiose, which was due to the low activity of the GH3 enzyme against the disaccharide and poor expression of the β-glucosidase. The crystal structure of BT3312 revealed a deep pocket that mirrored the U-shaped typology of β1,6-glucan, revealing the mechanism of substrate specificity. Finally the catalytic amino acids of both the GH30 and GH76 enzymes were identified by site-directed mutagenesis.

Contents

List of Figures

List of Tables

Chapter 1: Introduction

Carbohydrates, consisting of carbon, oxygen and hydrogen, are extremely diverse molecules, the existence of which is absolutely vital to life and a number of biological processes. Carbohydrates possess hydroxyl groups and an aldehyde or ketone group. Individual monosaccharides (single sugars), can adopt a ring conformation, which usually consists of a carbon chain of five or above atoms, existing as a six membered or five membered ring. These cyclic conformations are known as pyranose and furanose respectively. Individual monosaccharides can be connected by glycosidic linkages via the anomeric carbon to the hydroxyl group of another monosaccharide, forming oligosaccharides with varying degrees of polymerisation. Extensive polymerisation of monosaccharides results in polysaccharides - large molecules, the composition, structure and stereochemistry of which can be highly complex.

The eluded to biological roles that carbohydrates facilitate include cell metabolism and cell signalling, as well as acting as potential energy stores. Indeed, the innate biophysical properties of many polysaccharides endow the polysaccharide with functional characteristics, and so polysaccharides are often observed fulfilling a structural role in nature, such as in the cell walls of plants, bacterial cells and eukaryotes.

In addition to biological roles, carbohydrates also have extensive industrial applications and are exploited in a number of key processes and industries (pharmaceutical and food). Indeed, cellulose, the major constituent of the plant cell wall, is the most abundant source of hydrocarbons on the planet. The saccharification of cellulose is therefore of extreme relevance to the biofuel sector. Despite the recalcitrance of cellulose to enzymatic degradation, a number of bacteria exploit

cellulose as an energy source and encode for enzymes capable of deconstructing the polysaccharide.

The aforementioned example emphasises that the acquiring of knowledge of microbial exploitation of carbohydrates, mediated by the enzymes they encode, can be of extreme industrial and health (see Section 1.4.4) importance.

1.1 Glycoside Hydrolases

Glycoside hydrolases catalyse the hydrolysis of the glycosidic bonds linking saccharide residues or glyco-conjugates. The hydrolysis of glycosides results in the generation of a saccharide hemiacetal or hemiketal as well as the appropriate aglycon. Glycoside hydrolases possess the ability to catalyse glycosidic bonds where the linking atom is generally oxygen.

Glycoside hydrolases are often assigned a prefix of exo or endo, reflecting their ability to cleave terminal (most often at the non-reducing end) or internal glycosidic linkages, respectively. Substrate binding within the glycosidase active site is described using sub-sites, with the site of enzymatic cleavage as a reference point. Catalytic sub-sites accommodating sugar moieties in the direction of the non-reducing end (from the point of enzymatic cleavage) are denoted by sequential negative numbers (-1, -2, -3, etc.), with the active site located at -1 and the scissle bond between sugars located at -1 and +1 (Davies and Henrissat, 1995). Sub-sites of the enzyme interacting with glycan residues towards the reducing end are indicated with a positive number (+1, +2, +3, etc.).

1.1.1 Catalytic mechanisms

The hydrolysis of glycosidic linkages, catalysed by glycoside hydrolases, most commonly proceeds via a single or double displacement acid-base assisted

mechanism. Reactions which proceed with a single displacement mechanism results in a net inversion of anomeric configuration at C1 post-catalysis, at the site of bond cleavage. Conversely, bond cleavage which operates via a double displacement mode of catalysis results in an overall net retention of anomeric configuration. Glycoside hydrolases which operate with either a single or double displacement mechanism are therefore deemed ‘inverting’ or ‘retaining’ enzymes respectively. Irrespective of the mechanism, both modes of catalysis require two carboxylate amino acids for hydrolysis, an acid/acid-base and a base/nucleophile respectively. In inverting glycoside hydrolases the two catalytic residues, the general acid and general base, are generally aspartate and/or glutamate and are situated 6 – 11 Å apart (McCarter and Withers, 1994). The reaction proceeds via an oxocarbenium ion-like transition state. Hydrolysis by retaining glycoside hydrolases, occurring via a two-step, double displacement mechanism, involves a covalent glycosyl-enzyme intermediate, proceeding via oxocarbenium ion-like transition states. Acid/base and nucleophile assistance is given by two amino acid side chains (glutamate or aspartate) which are typically situated 5.5 Å apart. Phase one of the retaining process of catalysis is referred to as the glycosylation step. The catalytic nucleophile displaces the aglycon by attacking the anomeric centre. A glycosyl enzyme intermediate is thereby formed. The acid/base residue acts as an acid and provides protonic assistance by protonating the glycosidic oxygen encouraging leaving group departure. Phase two entails a deglycosylation step, whereby the glycosyl-enzyme is hydrolysed by a water molecule which is deprotonated by the acid/base amino acid now functioning as a general base (Rye and Withers, 2000; Zechel and Withers, 2000). A schematic of both the inverting and retaining mechanisms of hydrolysis is shown in Figure 1.1.

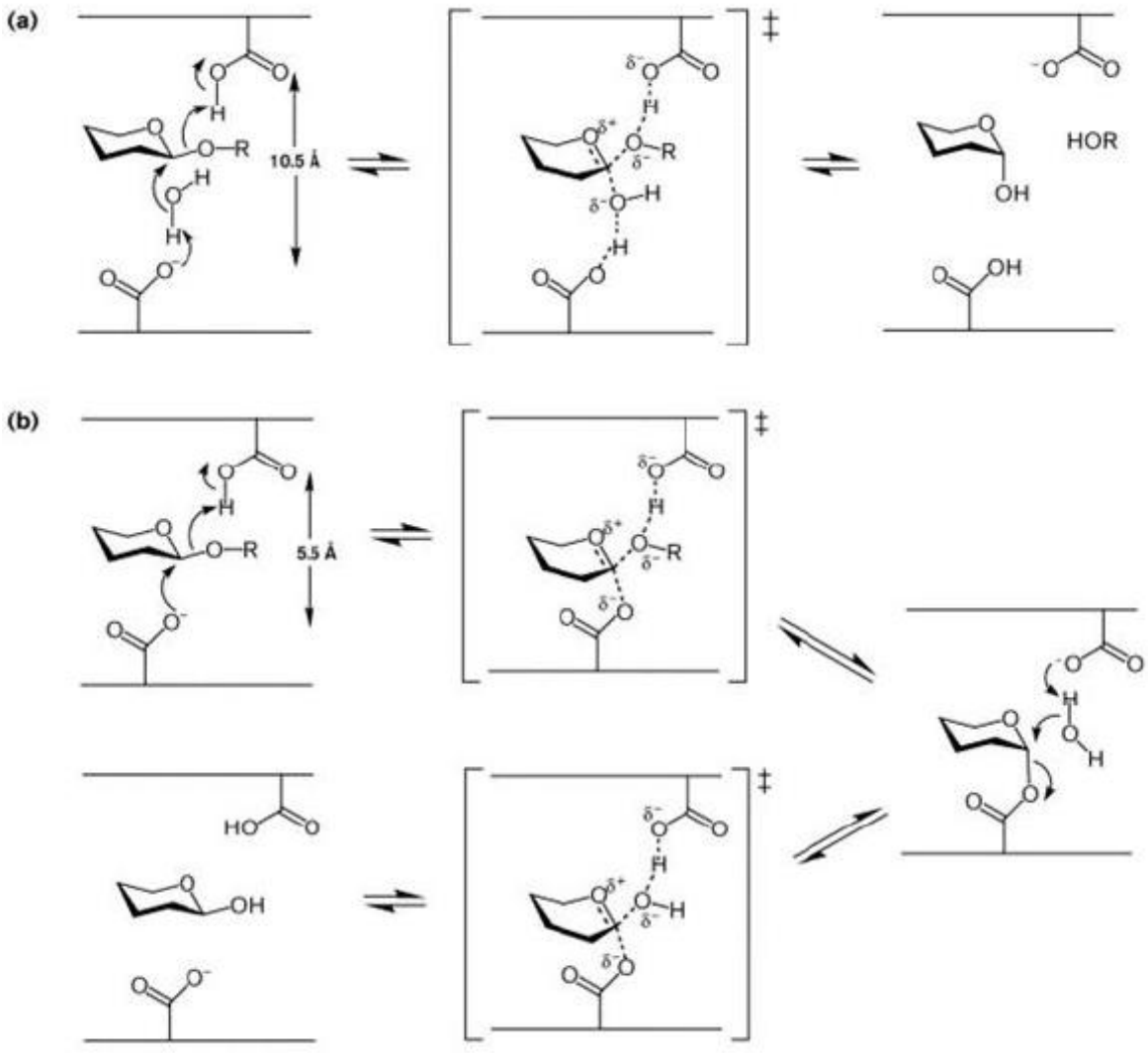


Figure 1.1. Catalytic mechanisms of (A) single displacement, inverting mechanism and (B) double displacement, retaining mechanism. The transition state of the reactions is highlighted in brackets. Taken from Rye and Withers, 2010.

1.1.2 Glycoside Hydrolase sequence based families

Glycoside hydrolases (GHs) are classified into sequence based families (defined as GHXX where XX is a number) in the CAZy database (www.cazy.org) (Cantarel *et al.*, 2009). Family members display the same fold, catalytic mechanism and catalytic apparatus (an exception to this general rule occurs in GH97). Such classification provides predicted activities of respective hypothetical glycoside hydrolases. There are currently 135 families of glycoside hydrolases on the CAZy database (as of January 2016). Members of the same glycoside hydrolase family can display the same

specificity, however, this is not always the case although the stereochemistry of the target glycosidic bond is conserved (Henrissat and Davies, 1997). GH families can be grouped into larger groups, known as clans. Clans possess extensive similarities in protein structure, catalytic residues and mechanism and help to provide structural and functional assignments to GHs.

The exo/endo mode of action of glycoside hydrolases is complementary to the topology of the active site of the enzyme. Three topologies are consistently observed in glycoside hydrolases. The open cleft topology accommodates the internal regions of the substrate and is indicative of an endo-mode of hydrolysis, cleaving sugar polymers internally. The pocket topology usually accommodates single saccharide residues (usually at the non-reducing end) and liberates monosaccharides from oligosaccharides/polysaccharides. Tunnel topology is most prevalent in exo-processive enzymes, where the polysaccharide/oligosaccharide is transported through the 'tunnel' of the enzyme, and subsequently cleaved in an exo-acting manner, liberating disaccharides (Davies and Henrissat, 1995)

1.1.3 GH families which target α-mannan

1.1.3.1 α-mannosidases

α-mannosidases are glycoside hydrolases which liberate mannose from polysaccharides/oligosaccharides containing accessible α-mannosidic linkages in an exo-mode of action. As this thesis focusses on α-mannosidases encoded by the α-mannan and high mannose N-glycan degrading systems of *Bacteroides thetaiotaomicron (B. theta)*, the α-mannosidases encoded by the genome of *B. theta* are described here.

1.1.3.1.1 GH92

GH92 α-mannosidases display variation in α-mannosidic link preference, with activity observed against α1,2, α1,3, α1,4 and α1,6 linkages (Zhu *et al.*, 2010). NMR

experiments revealed that GH92s which hydrolyse either an α1,2, α1,3 or α1,4 mannosidic linkage, generated β-mannose, indicating that an inverting mechanism of hydrolyse is utilized by members of GH92. The GH92 α-mannosidases are Ca^{2+} dependent, being one of three glycoside hydrolase families requiring a metal ion for catalytic activity. Interestingly all three families (GH38, 47 and 92) contain only α-mannosidases. Metal ion interacts with O2 and O3 of the mannose bound in the -1 active site, possibly providing the conformational flexibility required for the enzyme to recognise the ground state and transition state of the substrate, most likely due to the lack of distorting binding energy provided by the -2 sub-site as these enzymes are exo-acting (Zhu *et al.*, 2010)

The three-dimensional structure of the GH92 family of α-mannosidases reveals a two domain structure. The small N-terminal domain of the enzyme is comprised of a β-sandwich, whilst the larger C-terminal domain consists of an α/α_6 barrel fold. The active site of the GH92 family adopts a pocket topology which is formed by both the N- and C-terminal domains.

A number of GH92 α-mannosidases are encoded by the genome of *B. theta* (23), with 23 encoded by the α-mannan targeting apparatus of the bacterium (Cuskin *et al.*, 2015b). The GH92 enzymes encoded by *B. theta* have been previously characterised by Zhu et al. (2010), demonstrating activity against a variety of α-mannosidase linkages present in yeast mannan and mammalian high mannose N-glycans. Around 50% of the GH92 α-mannosidases encoded by *B. theta* were up-regulated when the bacterium was cultured on yeast α-mannan or mammalian glycans, indicating the likely importance of GH92 α-mannosidase activity in the degradation of these glycans

1.1.3.1.2 GH38

GH38 α-mannosidases utilize a retaining, double displacement mechanism of hydrolysis and require a Zn^{2+} ion for catalysis. The crystal structure of a GH38 derived from *Streptococcus pyogenes* (SpGH38) in complex with a transition state mimic shows that the zinc ion is coordinated at the active site of the enzyme in an overall T_6 octrahedral coordination configuration and interacts with the mannose residue bound in the active site, most likely enabling the distortion of the substrate towards the transition state ($^1S_5$ 'skew-boat'), reminiscent of the role of Ca^{2+} in GH92 α-mannosidases. GH38 enzymes display variation in specificity with respect to the α-mannosidic linkage and the substrate bound at the positive subsites (Suits *et al.*, 2010)

1.1.3.1.3 GH125

GH125 enzymes derived from *Clostridium perfringens* (CpGH125) and *Streptococcus pneumoniae* (SpGH125) hydrolyse α1,6-mannobiose with a net inversion of stereochemistry, indicating a single displacement mechanism of hydrolysis. Catalysis mediated by GH125 enzymes is metal-independent. This demonstrates that there is no *a prior* reason why the enzyme mediated exo cleavage of α-mannosidic linkages requires divalent metal ions. The three dimensional structure of SpGH125 and Cp125 demonstrates the family consists of an $(\alpha/\alpha)_6$ fold. Ligand recognition in the -1/+1 catalytic sub-sites of SpGH125 has been explored using thio-mannobiose, a non-cleavable substrate, revealing that aspartate 218 and glutamate 391 are the catalytic general acid and base of the enzyme respectively. The catalytic residues are structurally conserved in CpGH125 as well as the other clan members, GH15 and GH65. The resolved crystal structure of SpGH125 shows that O1 of the mannose bound at +1 is solvent exposed indicating that the enzyme can accommodate extended α1,6-mannooligosaccharides (Gregg *et al.*, 2011). A crystal structure of a

GH125 *B. theta* enzyme has been solved by a structural genomics consortium and its comparison to SpGH125 is described in Chapter 4, Discussion.

1.1.3.2 α-mannanases

α-mannases are endo-acting enzymes cleave internal α-mannosidic linkages. There are currently two families of endo-α-mannanases featured on the CAZy database, GH76 and GH99.

1.1.3.2.1 GH76

GH76 enzymes have been shown to be endo-α1,6-mannanases. It has recently been elucidated that the GH76 family act with an overall retention of stereochemistry, utilizing a double displacement method of hydrolysis. The GH76 enzymes are metal independent. The resolved crystal structure of a GH76 endo-α1,6-mannanase derived from *Bacillus circulans* (BcGH76) reveals an $(\alpha/\alpha)_6$ helical barrel fold. 3D structures of enzyme-substrate complexes show a solvent accessible, long, curved cleft like structure constituting the active site of the *Bacillus circulans* enzyme to which α1,6-mannopentaose is bound in the -4 to +1 sub-sites of the enzyme. The mannose residue bound to the -1 sub-site of the enzyme adopts a $^{O}S_5$ conformation. Interestingly, an enzyme-inhibitor complex revealed that the azasugar of α-mannosyl-1,6-isofagamine was distorted to a $B_{2,5}$ conformation in the same sub-site, therefore suggesting catalysis proceeds through a boat-conformation in the GH76 family. BcGH76 displays endo-mannanase activity against α1,6-mannooligosaccharides with a degree of polymerisation (DP) ≥3 (Thompson *et al.*, 2015).

1.1.3.2.2 GH99

GH99 mammalian and bacterial enzymes display endo-α1,2-mannanase activity (bacterial). The mammalian Golgi enzymes target glucose substituted mannosyl moieties found in the structure of immature mammalian high-mannose N-glycans, cleaving the α1,2-mannoside linkage internally and accommodating up to three

glucose substitutions proximal to the site of enzymatic cleavage. The bacterial GH99 enzymes release α1,3-mannobiose from the chain termini of *Saccharomyces cerevisiae* yeast α-mannan through the cleavage of an α1,2-linkage. Both bacterial enzymes show a strong preference for D-mannose at the -2 subsite reflecting steric clashes with an equatorial O2 (O2 is axial in mannose). Interestingly, eukaryote derived GH99s which display a preference for glucose in the -2 subsite possess a tyrosine residue capable of interacting with the equatorial O2 of glucose, conveying specificity that is complementary to the biological activity of the eukaryotic enzymes (Hakki *et al.*, 2015)

The GH99 family of endo-α-mannosidases act with a net retention of anomeric configuration and utilize a double, displacement mechanism of hydrolysis. Interestingly, the catalytic nucleophile of the family is yet to be elucidated. Instead, a catalytic process whereby catalysis is assisted by substrate has been proposed, involving the 2OH residue which is deprotonated by the general acid/base of the enzyme and proceeds via a 1,2 anhydrosugar intermediate with neighbouring group participation (Hakki *et al.*, 2015)

1.1.4 GH families which target β1,6-glucans

Reflecting the presence of β1,6-glucans in the yeast cell wall, glycoside hydrolase families which may potentially target β1,6-glycosidic linkages, are also encoded by the genome of *B. theta* and are described here.

1.1.4.1 β-glucosidases

β-glucosidases cleave β-glucosidic linkages in an exo-acting fashion.

1.1.4.1.1 GH3

GH3 contains enzymes with an plethora of functions that include β-glucosidases, β1,4-xylosidases. Members of this family often have an array of specificities. GH3 enzymes cleave β-glycosidic linkages through a retaining, double-displacement mechanism of catalysis (Paal *et al.*, 2004). Characterised GH3 enzymes that target β-glucans are exo-hydrolases, which act at the non-reducing termini of β-D-glucan polymers and glucooligosaccharides. Characterised GH3 β-glucosidases target β1,3, β1,4 and β1,6 linkages, with some enzymes capable of cleaving both β1,3 and β1,4 glucosidic bonds (Hrmova *et al.*, 1998).

The three-dimensional structure of GH3 enzymes reveals significant diversity in domain number/arrangement, as well as seemingly low sequence identity amongst some family members. However, structural overlays of crystallised GH3 β-glucosidases has shown that the architecture and amino acids in the active site are conserved, despite low sequence identity (Bohlin *et al.*, 2013)

1.1.4.2 β-glucanases

β-glucanases are endo-acting glycoside hydrolases that cleave β-linked glucosidic linkages. Described in this thesis is a representative of GH30 subfamily 3. Previously characterised members of this subfamily display β1,6-glucanase activity.

1.1.4.2.1 GH30

The GH30 family utilizes a retaining mechanism of hydrolysis. Crystal structures of GH30 members revealed a TIM barrel $(\alpha/\beta)_8$ fold. Currently, no 3D structure of a GH30 subfamily 3 β1,6-glucanase has been reported.

GH30 β1,6-glucanase of eukaryotic origin have been investigated, but not extensively characterised. The GH30 subfamily 3 fungal enzymes display endo-β1,6-glucanase activity. These enzymes target fungal cell wall β-D-glucans liberating small amounts of glucooligosaccharides (Oyama *et al.*, 2002)

1.2 Carbohydrate Binding Modules

Glycoside hydrolases that target complex recalcitrant substrates such as the plant cell wall often contain non-catalytic carbohydrate binding modules (CBMs). CBMs potentiate the catalytic activity of their cognate catalytic domains through targeting or proximity effects (Bolam *et al.*, 1998). CBMs that bind to single glycan chains often share the same specificity as their cognate enzyme, and direct the biocatalyst to its substrate. In heterogeneous complex structures, such as the plant cell wall, targeting enables efficient protein-substrate interactions. CBMs that bind insoluble carbohydrates increase the concentration of enzymes in the vicinity of these complex glycan superstructures leading to more efficient degradation. A third proposed mechanism by which CBMs potentiate the degradative capacity of their cognate enzyme is through non-catalytic disruption of the target substrate. This mechanism, however, has recently been questioned. This mechanism was founded primarily by the observation that CBP21, a member of CBM33, although not tethered to enzymes was able to substantially increase the efficiency of chitin degradation by endo-acting chitinases. CBP21 was later shown not to be a CBM but was an enzyme (lytic polysaccharide monooxygenase) that attacked chitin. Thus the poster boy of the CBM disruption school of thought has been discredited (Vaaje-Kolstad *et al.*, 2010) .

CBMs, Like glycoside hydrolases, are grouped in 71 sequence-based families on the CAZy database. The binding specificities of CBMs have been elucidated via biochemical and structural studies, with CBMs being observed to bind to a number of major plant cell wall polymers (cellulose, xylans, β-mannans, pectins) (Zhang *et al.*, 2014), storage polysaccharides (glycogen, starch and fructans), chitin and mammalian glycans.

CBM ligand binding site topography is generally complementary to the conformation adopted by the target glycan and can often indicate the preference of the CBM for crystalline or soluble carbohydrates. As such, CBMs are grouped into three types based on binding site topology, A, B or C respectively, in attempts to provide functional classification. Type A CBMs bind to crystalline substrates. The ligand binding site of type A is planar and houses three aromatic amino acid residues. The hydrophobic surface presents a platform-like architecture, complementary to the crystalline surface presented by insoluble cellulose and chitin (Kraulis *et al.*, 1989). Ligand binding in type A CBMs is entropy driven, believed to result from the release of trapped water molecules from the hydrophobic faces of crystalline ligands and the CBM itself (Creagh *et al.*, 1996). Example CBM families which feature type A CBMs include 1,2, 3, 5 and 10.

The binding site of type B CBMs presents a cleft like topology that targets the internal regions of individual glycan chains (>15 Å in length) (Charnock *et al.*, 2002). Type B CBM-ligand interactions are predominantly achieved through extensive polar contacts formed between ligand and polar amino acid residues housed within the binding cleft of the CBM. Additional specificity is conferred by aromatic amino acids, which most frequently form parallel hydrophobic stacking interactions with the sugar rings of the ligand, but can also adopt a sandwich conformation (Boraston *et al.*, 2002). Such hydrophobic residues are often orientated to accommodate the target ligand. For example, CBMs that bind to xylan (which adopts a 3-fold screw axis conformation), often have two tryptophan residues located in the ligand binding cleft that are orientated 120^{o} degrees to each other, thereby conferring xylan specificity (Simpson *et al.*, 2000). Examples of CBM families which included type B CBMs are 4, 15, 16, 20, 28 and 36.

The ligand binding site of exo-acting type C CBMs display a pocket topology and target the termini (primarily the non-reducing terminus) of glycan chains. The binding sites of type C CBMs accommodate 1 to 3 sugars. Interestingly, appending a type C CBM (CBM66) which binds to the non-reducing termini of fructans, significantly enhanced the activity of its parent, non-specific β-fructosidase enzyme (GH32) against the highly branched fructan, levan, but not inulin, a linear fructose polymer. This example demonstrates how CBMs can confer enzymatic substrate specificity (Fujimoto *et al.*, 2013). Examples of the three functional topologies demonstrated by CBMs are shown in Figure 1.2.

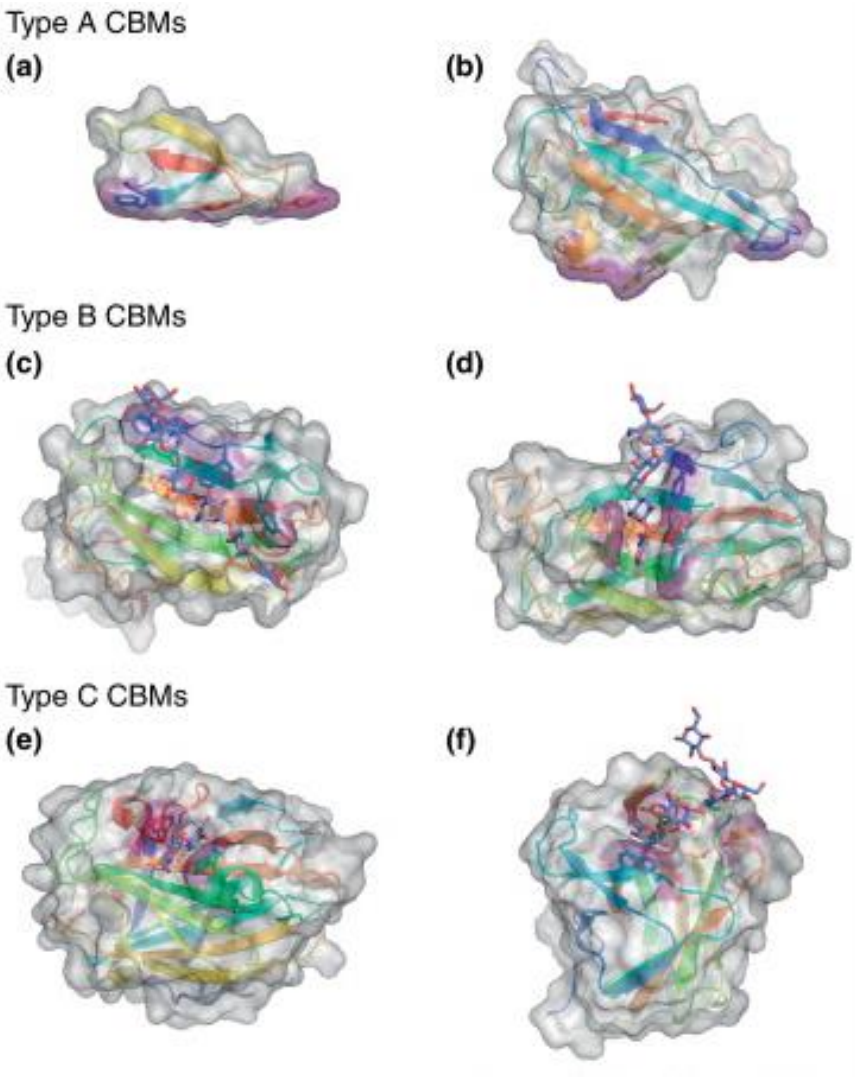


Figure 1.2. Examples of type A, ,B and C CBMs. **(A)** Structure of CBM1 from *Hypocrea jecorina* **(B)** Structure of CBM2 from *Cellulomonas fimi*. **(C)** Structure of CBM29 from *Piromyces equii* in complex with mannohexaose **(D)** Structure of CBM4 from *Cellulomonas fimi* in complex with cellopentaose **(E)** Structure of CBM9 from *Themotoga maritima* in complex with cellobiose, recognising the non-reducing terminus of the sugar molecule. **(F)** Structure of CBM6 from *Bacillus halodurans* in complex with laminarihexaose, recognising the non-reducing termini of the sugar molecule. Structures are continuously colour ramped with aromatic amino acid residues contributing to binding shown as sticks. Solvent accessible surface is coloured gray, whilst surfaces contributed to by aromatic amino acid residues is coloured purple. Co-complexed ligands are shown as blue sticks. Figure taken from Gilbert et al., 2013.

Whilst usually a more prevalent feature of lectins, binding proteins which are not too dissimilar to CBMs with regards to ligand binding mechanisms, avidity effects have been observed in enzymes that possess more than one CBM. The CBMs of these

enzymes can demonstrate different binding specificities respectively, and can result in a greater degradative capacity of the parent enzyme (Cuskin *et al.*, 2012)

Calcium has been increasingly shown to play significant roles in CBM-ligand binding, mediating direct ligand recognition. The divalent metal ion can stabilize the conformation of residues that play a key role in ligand recognition, it can play a direct role in glycan binding or it can cause increased affinity through avidity effects by mediating CBM dimerization (Montanier *et al.*, 2011).

1.3 Complex polysaccharide – structure and function

1.3.1 Major plant cell wall polysaccharides

The plant cell wall is rich in polysaccharides and presents a complex and heterogeneous structure which is extremely recalcitrant to enzymatic degradation. The major components of plant cell wall are cellulose, hemicellulose, pectin and lignin, which form a network of polysaccharides, providing mechanical support to the plant cell.

Cellulose consists of linear, undecorated β1,4-glucan chains, comprised of repeating cellobiose units (each sequential glucose monomer is orientated 180^{o} to one another). Such glycan chains are arranged in parallel and from contacts via hydrogen bonding and van der Waals forces, resulting in crystalline structures known as microfibrils, comprising 36 or 24 cellulose molecules (Fernandes *et al.*, 2011).

Hemicellulose polysaccharides most commonly consist of β1,4-linked backbones which are extensively decorated, thereby preventing the formation of crystalline structures. Such glycan chains interact with cellulose microfibrils via hydrogen bonding and function to provide strength to the plant cell wall. The major polysaccharide components of hemicellulose are xyloglucan, xylan, mannan and mixed-linked

glucans, with the branching and abundance of each polysaccharide varying dramatically between differing plant species and cell types. The polysaccharide xyloglucan, consisting of a β1,4-linked glucan backbone substituted with α1,6-xylose residues, constitutes the most abundant glycan in dicot primary cell walls (20-25 % of cell wall) (Scheller and Ulvskov, 2010). Up to 70 % of the xyloglucan backbone can be substituted. Xyloglucans of the XXXG classification have three decorated glucose residues, with a fourth unbranched backbone residue. The XXXG motif repeats regularly in the polysaccharide. The structure of xyloglucan is not ubiquitous amongst species however, with permutations of the XXXG type of xyloglucan being observed which harbour additional mono or disaccharide residues capping the canonical xylose side chains such as galactose, arabinose and fucose (Larsbrink *et al.*, 2014) . Within the plant cell wall, xyloglucan forms non-covalent interactions with cellulose, 'coating' the microfibrils, and is believed to act as a cross-linker, providing structural support during cell expansion. Another possible role xyloglucan provides is the prevention of cellulose microfibril aggregation, with xyloglucan acting as a spacer (Whitney *et al.*, 2000).

Xylans are a major constituent of the hemicellulose of grass cell walls and are extensively substituted. Up to 90 % of the xylose backbone of glucuronarabinoxylan, can be decorated with an array of saccharide residues, predominantly L-arabinose and D-(methyl)glucuronic acid but also D/L galactose D-xylose, acetate and ferulate groups, presenting a highly complex substrate (Albersheim *et al.*, 1994) . Methyl-glucuronic acid substitutions are more prevalent in dicots (10 %) than grasses (Pauly *et al.*, 2013).

Mixed linked glucans are prevalent in grass cell walls, consisting of a linear β—linked glucan. The proportion of β1,4 to β1,3 linkages is 70:30 respectively.

Heteromannans, such as β-mannans and glucomannans, represent a lower proportion of hemicellulose and act as a potential energy source in angiosperms. β-mannans are comprised of β1,4-linked mannosidic residues, whilst glucomannan consists of random sequences of glucose and mannose residues β1,4-linked in a linear fashion. Galactoglucomannan and galactomannan are polysaccharides in which galactose side chains are appended (α1,6) to the respective β1,4-linked backbones (Vogel, 2008).

Three major pectins are present in the plant cell wall; homoglacturonan (HG), rhamnogalactturonan I (RG-I) and rhamnogalactturonan II (RG-II), the proportion of which is ~60%, ~30% and 10% respectively. RG-II, a sub-domain of the pectic fraction, represents a structurally complex carbohydrate and consists of an α1,4-linked homogalacturonan backbone decorated with two structurally distinct disaccharides and two distinct oligosaccharides (Buffetto *et al.*, 2014). RGII exists as a dimer within the plant cell wall, and is cross linked via a borate molecule, forming a 1,2 borate-diol diester with the apiosyl residue containing side chain of RG-II (Kobayashi *et al.*, 1996). Pectins are proposed to act as spacers within the plant cell wall, preventing the aggregation of cellulose microfibrils (Chanliaud and Gidley, 1999)

Lignin, a component of secondary cell walls, provides strength to the plant cell wall. Lignin is highly hydrophobic and is comprised of hundreds of aromatic monomers, providing an extremely chemically complex structure. Lignin molecules are composed mainly of three monolignols; *p*-coumaryl, coniferyl and sinapyl alcohols which are polymerised via an oxidative process via free radicals (Davin and Lewis, 2000). As

such, the variety posed by lignin both chemically and sterochemically, renders the polymer extremely recalcitrant to enzymatic targeting.

1.3.2 The yeast cell wall

Yeast cells, typified by *Saccharomyces cerevisae,* are protected from environmental stresses via a cell wall. The wall consists of a variety of glucan and chitin polysaccharides cross linked covalently with proteins to form a matrix. Chitin contributes only 1% by dry weight of the total polysaccharide content of the wall of *Saccharomyces* species (Aguilar-Uscanga and Francois, 2003). Described below are the more prominent polysaccharides, typical of the yeast cell wall, Figure 1.3. Significantly, the polysaccharide and mannoproteins components of the yeast cell wall have been found to display an array of bioactive properties in both humans and animals, predominantly through interaction with mammalian pattern recognition receptors. Pattern recognition receptors, such as the Toll-like and mannose receptors located on the cell surface of immune cells bind to both β-glucans and mannoproteins respectively, recognising these molecules as pathogen associated molecular patterns (Latge, 2010). Indeed, the polysaccharides of the yeast cell wall have been attributed to the stimulation of immunoregulatory cells (Chen and Seviour, 2007) and anti-bacterial effects (Ganner *et al.*, 2010).

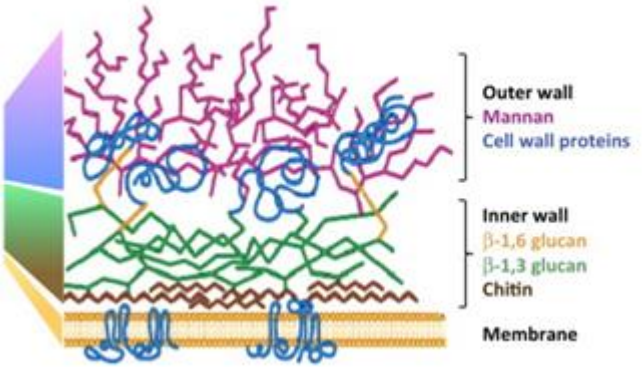


Figure 1.3. A cartoon representation of the polysaccharide network constituting the yeast cell wall. Figure adapted from Brown et al., 2014.

1.3.2.1 Yeast mannan

1.3.2.1.1 *Saccharomyces cerevisiae* yeast mannan and mutant mannans

Attached to the yeast cell wall proteins are α-mannans, bound via asparagine residues, which protrude extracellularly from the yeast cell. Manno-proteins constitute up to 50% of the yeast wall (Klis *et al.*, 1998). The structure and saccharide constituents of yeast mannan can vary amongst species.

Yeast α-mannans (yeast mannan) consist of an inner core region and an outer chain, Figure 1.4. The structure of the inner core region of yeast mannan is reminiscent of mammalian high mannose N-glycans. The inner core is composed of two β1,4 linked N-acetylglucosamine (GlcNAc) residues linked to a mannose residue via a β-1,4 bond. Appended to this trisaccharide structure are two branches, the first of which consists of two α1,6 linked mannose residues that are capped with α1,3 and α1,2 mannosyl residues, respectively. The other branch contains an α1,3-linked mannose. (Nakajima and Ballou, 1974). The outer chain of yeast mannan from *S. cerevisae*, consists of an α1,6 linked mannose backbone, decorated with α1,2 linked mannose side chains capped with α1,3 linked mannose residues. Phosphate bridges also exist within the polysaccharide and add additional mannose side chains to the

polysaccharide. The synthesis of yeast mannan is mediated by glycosyltransferases. The core region of yeast mannan is synthesised in the endoplasmic reticulum and the outer chain in the Golgi apparatus. The glycosyltransferases responsible for the biosynthesis of the outer chain in *S. cerevisae* has been characterized through deletion studies in yeast cells. Mutants missing various outer chain glycosyltransferases produce truncated mannan structures that are shown in Figure 1.4.

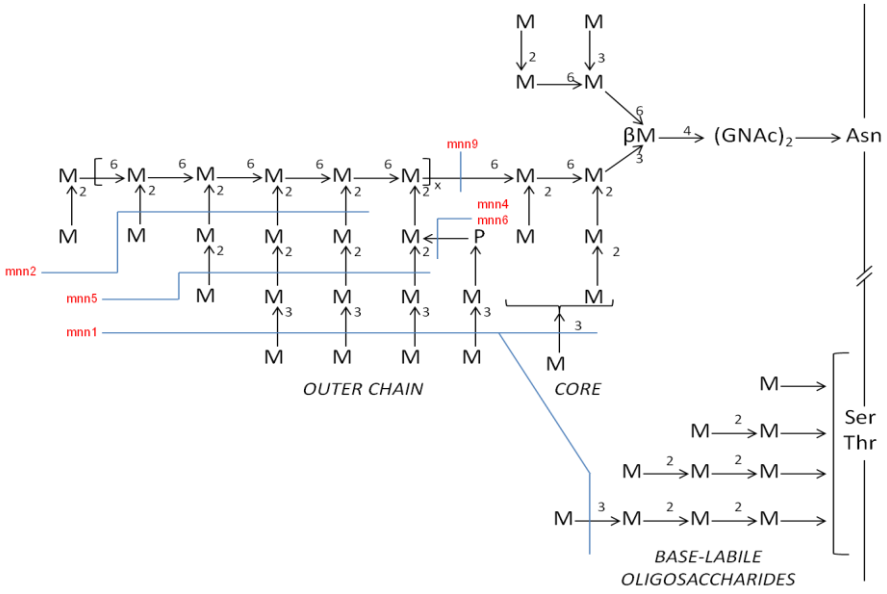


Figure 1.4. The structure of *Saccharomyces cerevisiae* Yeast mannan. M represents mannose residues and P represents phosphate. All arrows unless stated otherwise, denote an α-bond. The number above the arrow shows linkage. Mutations of glycosyltransferases are coloured red. All residues which follow/are below the blue lines are missing in that respective mutation. Figure adapted from Ballou (1990). The *mnn9* mutation completely abrogates synthesis of the yeast mannan outer chains

1.3.2.1.2 *Schizosaccharomyces pombe* yeast mannan

Schizosaccharomyces pombe possesses a morphologically unique Golgi apparatus, which differs from other eukaryotes (Ziegler *et al.*, 1994; Gremmill and Trimble, 1998). Two galactosyl transferases localized to the Golgi append galactose residues to mannans which are elongated from the core mannan structure (Ziegler *et al.*, 1994). As such, the outer chain of *S. pombe* consists of an α-1,6 linked mannose backbone, with 98% of the backbone 2-O-substituted with α-1,2 linked, pyranose galactose residues. Around 10% of these galactose side-chains are capped with an additional β-linked galactose residue. Terminal β-linked galactose residues possess a pyruvate residue linked in a 4,6-acetal-(ketal-) manner.

1.3.2.2 β-glucan fractions of the yeast cell wall

1.3.2.2.1 β1,3-glucans

Two fractions of β-glucan exist within the yeast cell wall. β1,3-glucans account for over half of the cell wall polysaccharides in *S. cerevisiae*. The β1,3-glucans of the yeast cell wall contain moderate β1,6-linked glucosdic decorations, and can extend up to ~1,500 residues in length. The polysaccharide adopts a single or triple helical confirmation, stabilised by inter-chain hydrogen bonding, and present 'spring' like structures, believed to endow the yeast cell wall with a degree of tensile strength (Kopecka, 2013).

1.3.2.2.2 β1,6-glucans

β-1,6-glucans have an important role in cell wall structure by cross-linking other constituents of the cell wall, including manno-proteins and chitin (Kapteyn *et al.*, 1999). Indeed, mutations resulting in disruptions to β1,6-glucan synthesis have lethal effects on yeast cells, owing the β1,6-glucan fraction acting to 'hold' the cell wall polymers intact (Kopecka, 2013). The structure of the mature β-1,6-glucan polysaccharide in *S. cerevisiae* has only recently been fully elucidated by Aimanianda *et al.* (2009). The β1,6-glucan polysaccharide consists of a glucose backbone linked via β-1,6 bonds.

The backbone is substituted with β1,3 linked glucose units with sporadic capping of an additional β1,6 linked glucose appended to the glucose side chains (Figure 1.5). The polymer can extend up to 200 glucose residues in length (Aimanianda et al., 2009), with the degree of branching of the polysaccharide varying between yeast species. In *S. cerevisae* β-1,6 glucans up to 7% of the polymer is branched, while 75% branching occurs in *S. pombe* (Aimanianda et al., 2009).

β1,6-glucans purified from *S. cerevisae* have previously been shown to possess anti-carcinogenic properties inhibiting metastasis in a number of cancers, including colon carcinoma cells. This is believed to occur through the modulation of macrophages and natural killer cells (Yoon *et al.*, 2008). β1,6-glucans interact with C11b/CD18 and Dectin-1 receptors, presented by neutrophils and macrophages, respectively. β1,6-glucans, therefore, activate cytokine production and adaptive immunity, which, potentially may attenuate the impact of colitis (Jawhara *et al.*, 2012). Indeed, Jawhara et al. (2012), demonstrated that yeast β-glucans demonstrated a beneficial effect on inflammation and the colonization of *C. albicans* (a potentially pathogenic yeast), in mouse models.

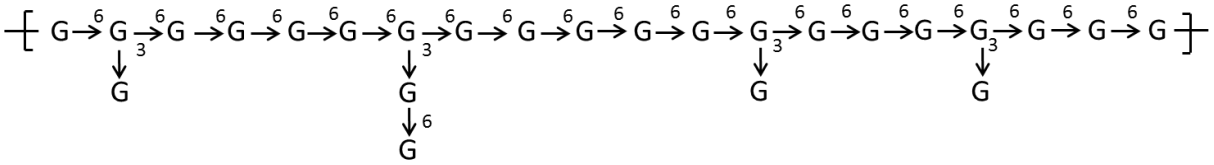


Figure 1.5. The structure of *Saccharomyces cerevisiae* β-1,6-Glucan. G indicates a glucose residue. All Arrows denote β-bonds and the numbers above the arrows represent linkage. Figure adapted from Aimanianda et al. (2009)

1.3.2.3 High mannose N-glycans

High mannose mammalian N-glycans (HMNGs) are similar in composition to the core structure of yeast mannan. The trisaccharide motif of GlcNAc-β1,4-GlcNAc-β1,4-Mannose is preserved. Appended to this trisaccharide are eight mannose residues, linked via an assortment of alpha- linkages found in yeast mannan. HMNGs, however, are distinct from the yeast mannan core structure due to additional α1,2 linked mannose residues capping the termini of the mannose branches and their lack of an outer chain.

HMNGs decorate the surface of intestinal epithelial glycoproteins and are present in secreted mucus. *Bacteroides* species, such as *B. theta* can scavenge these glycans (Martens *et al.*, 2011)

1.4 The human gut microbiome

1.4.1 The composition of human gut microbiota and environment

The human large bowel is inhabited by a vast array of micro-organisms which directly influence host health and nutrition. This community of microorganisms is referred to as the microbiota and encompasses commensal, symbiotic and pathogenic bacteria present within the gut (for review see (Bakhtiar *et al.*, 2013). The microbiota contains ~10^{14} bacteria grouped into around 1000 different species. The most prominent members of the distal gut belong to the phyla Firmicutes and Bacteroidetes, with 90% of the microbiota belonging to these two genera. The composition of the microbiota can fluctuate dramatically between individuals; however, a core ensemble of gut microbes is possessed by the majority of people (Qin *et al.*, 2010)

The composition of the gut microbiota can fluctuate dramatically due to diet and most astonishingly, the distribution of bacteria can alter in a time-scale of hours as a result

of selective feeding (Turnbaugh *et al.*, 2009). A diet high in fibre and containing abundant plant polysaccharides results in the proliferation of Bacteroidetes species and a decrease in the levels of Firmicutes (De Filippo *et al.*, 2010). Calorie restricted mice display increased levels of Bacteroidetes than those whose calorie intake is low (Ley *et al.*, 2005). Likewise, obese *ob/ob* mice, which are genetically predisposed for over-eating, have a much lower abundance of Bacteroidetes species in their microbiota and an increased level of Firmicutes species than their wild-type counterparts.

The relationship between the human host and the microbiota is symbiotic in healthy individuals (Hwang *et al.*, 2012). Commensal bacteria inhabiting the gut prevent the colonisation of potentially pathogenic bacteria on the intestinal epithelium and so constitutes a barrier to infection (Turner, 2009). Dendritic cells, generated in the gut-associated lymphoid tissue, sample commensal bacteria via pattern recognition receptors such as Toll-like receptors. This enables the innate and adaptive immune system of the gut to distinguish between commensal and pathogenic bacteria, maintaining intestinal homeostasis (Quigley, 2010). A change in the composition of the microbiota results in dysbiosis and potentially disease. Inflammatory Bowel Disease (IBD) is the chronic inflammation of the gastro-intestinal tract and has been associated with dysbiosis. Crohn's disease, a prevalent IBD, has been linked to a decrease in the number of *Firmicutes* species within the microbiota (Sokol *et al.*, 2008). The microbiota is crucial for gut homeostasis, the imbalance of which has the potential to induce disease states.

1.4.1.1 Bacteroides

The *Bacteroides* genus of the Bacteroidetes phyla are glycan-degrading generalists. *Bacteroides thetaiotaomicron* (*B. theta*), is a Gram-negative bacterium belonging to the phylum Bacteroidetes (Rogowski *et al.*, 2015). *B. theta* is a prominent member of

the human colonic microbiota and is one of the primary bacteria involved in glycan degradation in the gut (Sonnenburg *et al.*, 2010). The microbiota presents a highly competitive environment. The success and fitness of the Bacteroides can be attributed to the capacity of this phylum to degrade the major classes of plant polysaccharides and host derived glycans. Transcription and growth studies (Martens *et al.*, 2011) demonstrate that *B. theta*, in addition to metabolising major plant polysaccharides such as starch, fructans and the pectins, is also capable of utilizing host O- and N-linked mammalian glycans and α-mannans derived from yeast and other microbial eukaryotes (Cuskin *et al.*, 2015b). The closely related species, *Bacteroides ovatus* (*B.ovatus*), in addition to degrading pectins and fructans, also utilizes the major plant hemicelluloses such as xylans, β-mannans and xyloglucans (Larsbrink *et al.*, 2014; Rogowski *et al.*, 2015). The ability of the *Bacteroides* genera to degrade such glycans is attributed to the many glycanases encoded by the organism's genome. The CAZy database predicts that the *B. theta* genome encodes 261 glycoside hydrolases. Indeed, the strategies by which *B. theta* degrades glycans, beyond the proto-typical starch utilization system, have recently been further elucidated (See 1.6 Characterised PULs of *Bacteroides*).

1.4.2 Complex glycan utilization and nutrient sharing

The catabolism of complex dietary carbohydrates in the human alimentary underpins the structure of the human gut microbiota. The metagenome of the gut microbiota encodes for ~100-fold more genes than that of the human genome, many of which are targeted to polysaccharide degradation. As much as 18% of the *B. theta* genome is dedicated to glycan degradation and utilization (Rogers and Bruce, 2012). The human gut microbiota is exposed to a plethora of complex glycans which pass through the gastrointestinal tract largely undigested. Furthermore, the colonic microbiota is

exposed to host derived glycans in the form of secreted mucins and surface glycans of epithelial cells that are shed (Rogers and Bruce, 2012). These dietary and host glycans represent a major potential nutrient source for gut bacteria. Indeed, prebiotic strategies (See Section 1.4.4), used to manipulate the composition of the human gut microbiota via the promotion of prebiotic bacteria, rely upon the use of complex carbohydrates only accessible to the targeted bacterial subset. Such a strategy is heavily reliant on the glycan food web of the human gut microbiota (Rogowski *et al.*, 2015). For example, studies have shown that organisms, such as *Ruminococcus gnavus* modify host sialic acids such that the product is used exclusively by the *Ruminococcus* (Crost *et al.*, 2013). Apart from these few reports, how complex glycan degradation and utilization is orchestrated between the numerous members of the human gut microbiota is little understood, as is the potential for nutrient sharing in such a complex, highly competitive environment.

1.4.3 Short chain fatty acids and their influence on host health and nutrition

Short chain fatty acids (SCFA), the end products of bacterial fermentation of dietary glycans, are absorbed through the intestinal gut epithelium and can contribute up to 10% of our total calorific intake (Ley *et al.*, 2005). The role of the gut microbiota in energy harvesting is emphasized by the observation that germ-free rats display a lower abundance of SCFAs in the gut and excrete two-fold the amount of calories than colonized rats fed on identical, polysaccharide rich diets (Hoverstad and Midtvedt, 1986). Germ-free Rats must compensate for this lack of energy absorption via increased feeding (Tremaroli and Backhed, 2012). This evidence suggests that the gut microbiota can be described as a functional metabolic organ and so underpins the mutualistic dependence which exists between the human host and gut microbiota.

The three main SCFA products of bacterial fermentation are butyrate, acetate and propionate and have repeatedly been shown to have a positive impact on gut health, some of these roles include acting as epithelial energy sources, vasodilators and modulators of the inflammatory response (For review see Tremaroli and Backhed (2012). The ratio and abundance of SCFA produced within the gut is dependent on the level of carbohydrates consumed in the diet and the composition of the microbes which colonize the gut. For example, the co-colonisation of *B. theta* and *Methanobrevibacter smithii* is optimal for the fermentation and utilization of dietary fructans due to *M. smithii* using the products of *B. theta* fructan degradation and fermentation. Mice which are co-colonised with *M. smithii* and *B. theta* displayed a dramatic increase is adiposity, demonstrating an increased efficiency of fermentation and SCFA absorption compared to animals colonised only with *B. theta* (Samuel and Gordon, 2006). Butyrate is a major energy source for colonocytes. In germ-free mice colonocytes are starved of energy (Donohoe *et al.*, 2011). Butyrate is an inhibitor of histone deacetylase, and so plays a significant role in cellular proliferation, differentiation and modulation of gene expression within the host colonic epithelium. As such, butyrate is believed to regulate up to 2% of transcription within mammalian colonocytes (Davie, 2003). Acetate and propionate are utilized in the liver and are important substrates in lipogenesis and gluconeogenesis. This is consistent with the increased levels of stored triglycerides in the livers of mice which harbour a gut microbiota..

1.4.4 Probiotic and prebiotic strategies

Dysbiosis of the gut-microbiota results in the induction of disease states, including obesity and Inflammatory Bowel Diseases (IBDs) (Quigley, 2010). In addition, due to the numerous health benefits conferred by SCFAs, dietary strategies designed to

target subsets of bacteria are being devised in the hope of engineering the microbiota to maintain human health. This is achieved by introducing or inducing sub-populations of bacteria to proliferate within the gut using probiotics or prebiotics.

Probiotics are orally administered live strains of microorganisms, intended to colonise the GI tract to induce health benefits, including aiding in digestion. *Clostridium butyricum,* a butyrate synthesising bacterium present within the intestines of healthy adults, is utilized as a probiotic in the treatment of non-antimicrobial induced diarrhoea and constipation in humans (Okamoto *et al.*, 2000). One recent study administered *Clostridium butyricum* to rats where it was observed that fatty liver disease progression was severely reduced as a result of the bacterium indirectly modulating lipogenesis leading to a reduction in triglyceride content and insulin resistance (Endo *et al.*, 2013).

Prebiotic strategies entail the inclusion of complex glycans resistant to host degradation within the diet, which are utilized by a specific subset of bacteria in the microbiota. Such strategies are cheaper than probiotics and are utilized readily in the animal rearing industries (Goffin *et al.*, 2011). Evidence suggests that the composition of the gut microbiota can shift in response to short and long term dietary intervention (Turnbaugh *et al.*, 2009). Reduced inflammation was observed in genetically altered mice, predisposed for obesity when administered with prebiotics (Delzenne and Cani, 2011).

The findings of Rogowski et al. (2015) described in Section 1.4.2, illustrates the importance of understanding the mechanisms by which complex glycans are degraded by different members of the microbiota. For instance, the complex glucuronoarabinoxylan, utilized ‘selfishly’ by *B. ovatus*, would result in the proliferation of *B. ovatus* exclusively, a documented producer of propionate. However, if the diet

was supplemented with the less complex wheat arabinoxylans or glucuronoxylans, both *B. ovatus* and *Bifidobacteria*, which uses the oligosaccharides generated by the *Bacteroides*, could potentially thrive. *Bifidobacteria* mediate the production of the beneficial SCFA butyrate in the microbiota (Wachtershauser and Stein, 2000). Understanding these interactions with respect to glycan metabolism is critical when designing bespoke prebiotics.

Yeast is a key component of the brewing and baking industries. The inclusion of mannan oligosaccharides and yeast cultures in diet has been observed to have a positive impact on intestinal health in a number of animal studies. Weaning pigs whose diet was supplemented with yeast cultures demonstrated increased growth, improved intestinal health and a better immune function. These improvements were comparable to pigs fed antibiotic growth promoters and so has the potential to replace antibiotics in animal feed (Shen *et al.*, 2009). Both yeast culture and manno-oilgosaccharide supplementation in diet have been observed to improve gut morphology. Increased jejunal villus height and an improved villus height to crypt depth ratio has been observed. This process is thought to be modulated by intestinal microbe colonisation, explaining increased growth rates through better absorption of nutrients in the gut (Dimitroglou *et al.*, 2009).

1.5 Polysaccharide utilization loci

When *B. theta* was cultured with polysaccharides, it was observed that transcription of specific genetic loci were up-regulated. These gene clusters, defined as Polysaccharide Utilization Loci (PULs), encode an array of proteins tailored for the sensing, degradation and utilization of the polysaccharide that activates the locus. The defining aspect of a PUL is the presence of an extracellularly located homologs of the

SusD and SusC proteins (defined henceforth as $SusD_h$ and$SusC_h$, respectively) present in the starch utilization system. $SusD_h$ binding proteins bind environmental glycans and facilitate the translocation of captured glycans into the periplasmic space through delivery to the energy dependant $SusC_h$ transporter (Sonnenburg et al., 2010; Martens et al., 2011). There are 208 homologs of SusC and SusD pairs encoded by *B. theta*, indicating that the genome encodes an extensive array of glycan utilizing systems (Xu *et al.*, 2007). Encoded by most PULs are other extracellular binding proteins, such as surface glycan binding proteins (SGBPs), which sequester glycans from the environment as a prelude to enzymatic digestion and translocation to $SusD_h$ proteins prior to import through $SusC_h$ pores. A myriad of extracellular and periplasmic glycanases, encoded by the PUL degrade captured glycans in a sequential manner, rendering short chain oligosaccharides and monosaccharides, which can be metabolized by the bacterium. Polysaccharide sensing and signal transduction is achieved via sensors encoded by the PUL. Hybrid two component systems (HTCS) and extra cytoplasmic function (ECF) σ factors which span the internal cytosolic membrane are utilized in the up-regulation of the PUL and are activated by specific glycans (Miyazaki *et al.*, 2005).

The cellular localization of cell envelope associated proteins is inferred by the presence of a bacterial signal peptide. Proteins harbouring a type I signal peptide recognition motif are secreted to the periplasmic compartment of the cell. Proteins which encode a type II signal peptide motif are predicted lipoproteins and are localised to the extracellular membrane. Lipo-proteins harbour a canonical N-terminal cysteine residue to which an N-acyl-S-diaglyceryl molecule is bound following cleavage of the signal peptide, therefore anchoring the protein to the lipid membrane (Bos *et al.*, 2007).

As stated above the PUL paradigm was built upon the starch utilization system (sus) first characterised in *B. theta*. The PUL encoding the sus system is composed of eight genes *susRABCDEFG*. SusE and SusF bind and sequester starch from the environment, enabling the extracellular α-amylase, SusG, to cleave the captured starch, hydrolysing the polysaccharide into long chain oligosaccharides. SusC and SusD work in tandem to import the products of SusG degradation into the periplasmic space where enzymes, encoded by *susA* and *susB,* act upon imported oligosaccharides, degrading them into small oligosaccharides and the monosaccharide glucose through neopullulanase and α-glucanase activities. SusR is a sensor/regulator element which spans the cytosolic membrane and is responsible for the further up-regulation of the sus system. A model of starch degradation by the sus system is shown in Figure 1.6.

1.5.1 Extracellular binding proteins

1.5.1.1 SusD-like

The extracellularly located SusD, characterized in the starch utilization system of *B. theta*, has previously been shown to bind starch (Shipman *et al.*, 2000). The 3D structure of SusD indicates that the 551 amino acid residue protein is composed of 22 α-helicies, eight of which constitute four highly conserved tetra-trico peptide repeats (TPRs)(Koropatkin *et al.*, 2009). These TPR motifs are implicated in the formation of protein-protein interactions, and are thought to mediate the SusC/SusD complex formed at the extracellular membrane (Cho and Salyers, 2001).

The 3D structure of the SusD reveals a shallow-surface binding site which can interact with three individual glucose moieties of the starch molecule. Binding studies indicate that the curved starch analogue β-cyclodextrin binds to SusD with a ~10 fold higher affinity than the linear starch derivative, maltoheptaose (Koropatkin et al., 2008). Starch forms a cyclic complex in solvent, owing to the α1,4-glycosidic bonds which link

the glucose monomers of the polysaccharide, indicating that SusD displays preference for starch surrogates which possess a helical conformation. This suggests that starch recognition by SusD is structure dependant and is not conferred by the stereochemistry of the monomeric constituents of the polysaccharide. Conversely, the $SusD_h$ protein, BT1043, encoded by the O-glycan PUL of *B. theta*, binds to the reducing end of Di-β-N-acetyl glucosamine through hydrogen bonding. This is mediated via polar amino acid residues located in the ligand binding site (Koropatkin *et al.*, 2009) and so ligand recognition is conferred by interactions with individual β-N-acetyl glucosamine units. This indicates there is diversity in ligand recognition displayed by $SusD_h$ proteins.

$SusD_h$ proteins have been shown to play a substantial role in polysaccharide utilization in *B. theta* (Koropatkin et al., 2008). The deletion of SusD in the sus system abolishes the ability of *B. theta* to utilize starch and its derivatives. Surprisingly, it has been found that SusD is vital for the utilization of linear malto-oligosaccharides, even with malto-oligosaccharides with a D.P <6, for which the protein has undetectable affinity for. This suggests that SusD is not only required for starch binding and sequestering, but serves other functions vital to polysaccharide utilization. Martens et al. (2009) speculate this could be due to SusD being required to stabilise the SusC/SusD complex located to the outer-membrane in addition to delivering malto-oligosaccharides to the outer membrane SusC transporter. Indeed, SusD-like proteins have been observed that seemingly possess no binding function, suggesting that a proportion of $SusD_h$ proteins may act solely to stabilise the outer-membrane complex (Cuskin *et al.*, 2015b). These hypotheses are consistent with the observation that other outer membrane proteins, such as SGBPs, are encoded by PULs and are more likely candidates for glycan

sequestering and initial binding of environment polysaccharides, reflecting their high affinity for these ligands (Bolam and Koropatkin, 2012).

1.5.1.2 Surface glycan binding proteins

Unlike $SusD_h$ proteins, SGBPs are not vital to polysaccharide utilization (Cho and Salyers, 2001). *In vitro* studies have shown that the deletion of SGBP genes does not impact on the capacity of *B. theta* to grow on the respective polysaccharide (Shipman *et al.*, 2000). Whilst the position of the SGBP ORF is generally conserved within the PUL, there is little sequential/structural homology shared between these binding proteins.

1.5.2 SusC-like

The $SusC_h$ transporter belongs to a family of TonB-dependant receptors found in gram-negative bacteria. The protein consists of seven domains, forming a porin at the extracellular membrane through which molecules can be transported into the cell in an energy dependant manner. Energy for transportation is generated through the formation of the TonB-ExbBD complex and proton-motif forces. Whilst energy is required for the translocation of macro-molecules, smaller molecules are able to diffuse freely through the porin and into the periplasmic space (Ferguson and Deisenhofer, 2002)

1.5.3 Carbohydrate sensing and PUL up-regulation

Carbohydrate sensing and PUL up-regulation is mediated primarily through hybrid two-component systems (HTCS) and extracytoplasmic function (ECF) sigma/anti-sigma factors which span the cytosolic membrane of the bacterial cell. PULs are perpetually expressed at low levels. In response to the appropriate molecular cue, such as ligand binding to the sensor/regulator protein, specific PULs are up-regulated (Sonnenburg *et al.*, 2010)

HTCS are composed of the domains found in the typical bacterial two-component system, and include, within a single polypetide chain, an N-terminal extracellular sensor, cytoplasmic histidine kinase, response regulator, phosphoacceptor domain and DNA binding domain (Sonnenburg *et al*., 2010; Bolam and Koropatkin, 2012). The degradation products generated by extracellular and periplasmic enzymes encoded by the PUL bind to the sensor domains of the HTCS, which protrude into the periplasmic space. Once activated, it has been proposed that the cytoplasmic DNA-binding domain may be released, binding DNA and therefore activating transcription (Miyazaki *et al.*, 2005). Two classes of HTCS exist, the Reg_prop class and Periplasmic Binding Protein (PBP) class, grouped according to the structure of their respective periplasmic sensor domains. The Reg_prop class of HTCS bind oligosaccharides with a D.P ≥ 2 (Martens et al., 2011). Conversely, the only characterized PBP HTCS binds monomeric fructose (Sonnenburg et al., 2010). Nevertheless, the mechanism of signal transduction appears to be conserved (Sonnenburg et al., 2006; Martens et al., 2011).

Recently, the structure of the periplasmic domain of a HTCS (BT4663), encoded by the heparin PUL of *B. theta*, reveals that ligand binding induces a conformational change in protein structure, which differs from the previously observed piston/rotation mechanisms of the sensor histidine kinases of two component systems. Upon ligand binding, the C-terminal domain of the protein appears to adopt a scissor blade-like closing mechanism. This enables the cytoplasmic histidine kinase domain to catalyse reciprocal phosphorylation for signal transduction (Lowe *et al.*, 2012)

Extracytoplasmic function (ECF) sigma/anti-sigma factors are regulatory elements which are encoded by some PULs. When the appropriate signal is received, the ECF sigma/anti-sigma release their sigma factor which binds to RNA polymerase and

stimulate transcription (Helmann, 2002). These regulators are mainly distributed in PULs dedicated to the degradation of O-linked mucins (Bolam and Koropatkin, 2012).

1.5.4 Enzymes

PULs encode an array of glycanases, which are tailored for the deconstruction of the polysaccharide. A myriad of enzyme are required to hydrolyse complex polysaccharides, reflecting the different linkages and sugar monomers that comprise such complex glycans. This is exemplified by the complex polysaccharide rhamnogalacturonan II (RGII). RGII is composed of 20 different linkages. To fully degrade RGII, no less than 20 glycanases would be required. *B. theta* has the capacity to degrade RGII. When *B. theta* was cultured on RGII, 30 ORFs were up-regulated, constituting the RGII PUL of *B. theta* (Martens *et al.*, 2011). This evidence suggests the number of glycanases encoded by PULs correlates to the complexity of the targeted glycan.

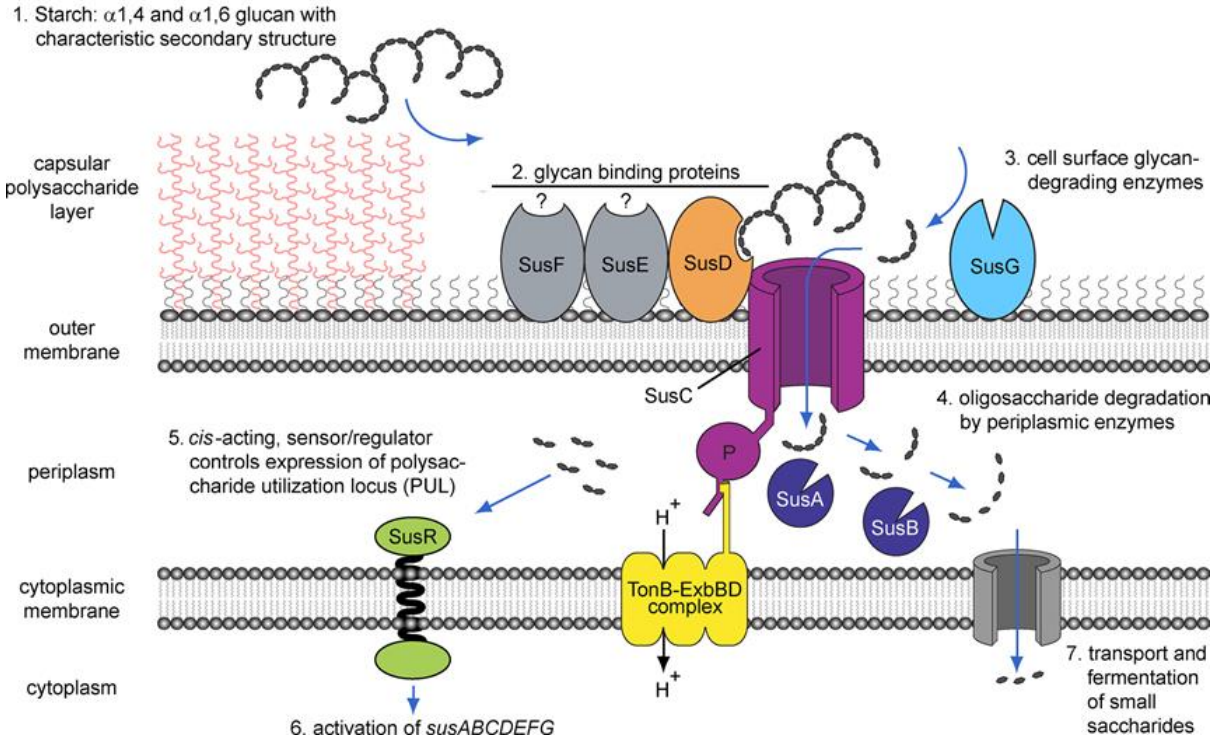


Figure 1.6. The starch utilization system of *B. theta.* A diagram of the starch utilization system of B. theta, the archetypal polysaccharide utilization loci. Environmental starch is sequestered via the extracellular binding proteins; SusFED, whereby it is hydrolysed via SusG, an extracellular α-amylase. Products of SusG degradation are transported into the periplasmic space through the extracellular membrane spanning porin, SusC. Long chain oligosaccharides are further degraded by the periplasmically located enzymes, SusAB, via neopullulanase and α-glucanase activity respectively and are transported into the cell through an inner membrane sugar permease whereby they are catabolised by the cell. The products of periplasmic degradation also result in the further up-regulation of the PUL, mediated via the senor/regulatory element, SusR. Figure adapted from Martens et al. (2009).

1.6 Characterised PULs of *Bacteroides*

Recent studies of the polysaccharide utilization systems of *Bacteroides*, highlights the complexity of the enzymes that mediate glycan degradation and are described below.

1.6.1 The xyloglucan PUL of *Bacteroides ovatus*

A *B. ovatus* PUL was up-regulated when the bacterium was cultured on xyloglucan, a relatively abundant dietary polysaccharide consisting of a β1,4-glucan backbone decorated with α1,6-xylose side chains, in a repeating XXXG or XXGG motif depending on the polysaccharide source. Side-chains can be capped with monosaccharides of galactose (β1,2-linked) or a β-D-1,2-galactose-L-α1,2-fucose disaccharide respectively. The xyloglucan PUL, referred to as XyGUL, encoded for a SusD-like protein, a SusC-like protein, a HTCS as well as eight predicted glycoside hydrolases based on their location in CAZy sequence based families. A schematic of XyGUL and a model of xyloglucan utilization by *B. ovatus* is displayed in Figure 1.7 and Figure 1.8 respectively. The extracellular degradation of xyloglucan is dominated by surface endo-xyloglucanase activity, mediated by two enzymes BoGH5A and BoGH9A respectively, generating short oligosaccharides from the polysaccharide. Genetic knock-outs of the BoGH5A ORF abrogated *B. ovatus* xyloglucan utilization, highlighting the importance of the enzyme in xyloglucan utilization. BoGH9A displayed low activity and deleting the gene encoding the endoglucanase did not affect *B. ovatus'* ability to grow on xyloglucan. Interestingly, no homologs of BoGH9A were encoded by orthologous xyloglucan PULs, perhaps providing an example of functional redundancy in a PUL system.

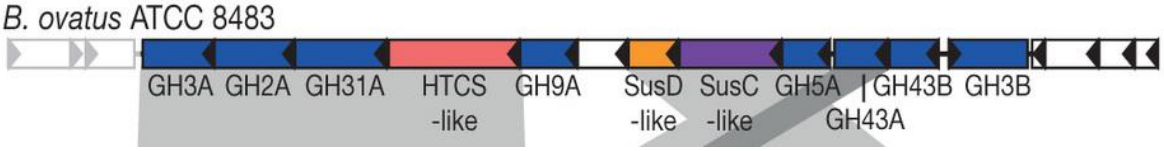


Figure 1.7. The xyloglucan PUL (XyGUL) of *B. ovatus*. ORFs encoding enzymes are coloured blue. The function of the protein encoding ORFs are labelled accordingly according to function. Figure adapted from Larsbrink et al., 2014.

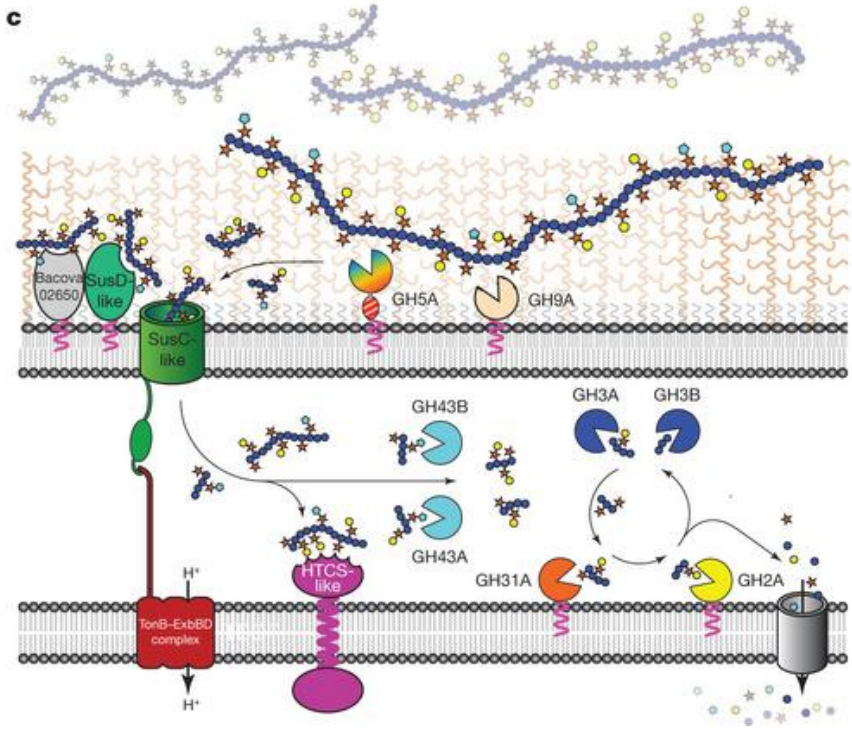


Figure 1.8. Proposed model of xyloglucan utilization by *B. ovatus*. Proteins are labelled according to function/GH family. Saccharide residues are represented by shapes and colours. Blue circles represent glucose, orange stars represent xylose, turquoise pentagons represent arabinofuranose and yellow circles represent galactose residues. The enzymes which remove individual monomers are coloured according to the saccharide residue they target. Figure adapted from La.rsbrink et al., 2014.

The remaining enzymes, predicted to be localised to the perisplasmic space, demonstrated the following exo-acting activities; GH2 β-galactosidase; GH3(A/B) β-glucosidase; GH31A α-xylosidase and GH43 (A/B) α-L-arabinofuranosidase,

respectively, and act sequentially to debranch and depolymerise imported oligosaccharides into monomeric saccharides. Interestingly, the two GH3 β-glucosidases encoded by XyGUL share specificities and act upon β-glucosidic linkages in identical contexts. The requirement for two seemingly identical activities in the PUL system is not understood, but again, suggests a degree of functional redundancy (Larsbrink *et al.*, 2014)

Interestingly, evidence suggests the XyGUL systems encoded by different *Bacteroides* species, harbour ORFs, which enable the use of xyloglucans derived from other plant sources. For example, orthologous XyGULs encode GH29 and GH95 α-fucosidases, enabling the bacteria to efficiently utilize the xyloglucan of dicots. The *B. ovatus* XyGUL does not, however, encode α-fucosidases.

The PUL described above provides an example whereby the depolymerisation of the targeted glycan, specifically a relatively abundant glycan, is mediated primarily by surface, endo-acting enzymes which show tolerance for decorations and release a plethora of short oligosaccharides into the extracellular space for import by the bacterial cell. Decoration removal is exclusively mediated in the periplasmic space.

1.6.2 The xylan PULs of *Bacteroides ovatus*

A study by Rogowski et al. (2015) demonstrates *B. ovatus* utilization of complex xylans and how different forms of xylan are degraded through the recognition of unique specificity determinants by the encoded protein ensemble. The genome of *B. ovatus* contains two PULs, implemented in the deconstruction of complex xylans. A schematic of the xylan PULs of *B. ovatus* is shown in Figure 1.9. Interestingly, the two loci appear to be regulated by different signalling molecules. The xylan PULs, designated PUL-XylS and PUL-XylL, were up-regulated by undecorated xylans (birchwood glucuronoxylan (BGX)) or by more complex GAXs and arabinoxylans. Interestingly,

PUL-XylS encodes for fewer enzymes, reflecting the lower complexity of the glycans targeted by the locus. Both PULs encode a plethora of glycoside hydrolases belonging to CAZy families with known activities against xylans. However, PUL-XylL encodes glycoside hydrolases belonging to families which were not implicated in xylan deconstruction (GH31, GH95, GH97 and GH98), and highlights the potential difficulties faced when predicting enzyme specificities based on CAZy family location.

Three surface enzymes, located in GH10 (BACOVA_04390), GH30 (BACOVA_03432) and GH98 (BACOVA_03433), were implicated in the cleavage of the xylan backbone at the cell surface. Interestingly, the activity of the surface GH10 enzyme appeared to be tailored to generate xylooligosaccharides with a DP >5. Previously characterised GH10 family members appear better suited to targeting smaller xylooligosaccharides. Indeed, the typical GH10 $(\beta/\alpha)_8$-fold barrel of the enzyme appeared to be interrupted by two xylan binding CBMs, and it is speculated to form the extended substrate binding cleft of the enzyme (Rogowski *et al.*, 2015). Such evidence suggests BACOVA_04390 had evolved to accommodate longer oligosaccharides at the cell surface, reducing release of short chain oligosaccharides which could potentially be lost to the extracellular space. The surface GH30 enzyme, BACOVA_03432 targets glucuronoxylan, conferred by the infrequent (Me)GlcA side chains positioned in the -2 subsite of the enzyme, which thus generates long oligosaccharides with a DP ~10. Similarly, the surface GH98, BACOVA_03433 generates oligosaccharides specifically from CX; the first GH98 family member to display endo-xylansae activity. The GH98 enzyme required unique specificity determinants for activity, requiring xylopyranose and arabinofuranose decorations linked O3 and O2 to the xylan backbone. The three surface xylanases of the xylan PULs therefore require infrequent specificity determinants for activity, recognising the

differing decorations of a multitude of xylans and are seemingly adapted to target different substrates.

The binding specificity of the SusD-like surface binding proteins of PUL-XylS and PUL-XylL reflect the enzyme and signalling specificities of their respective PULs. BACOVA_04390 of PUL-XylS bound to xylooligosaccharides with a a DP >6, consistent with the xylooligosaccharide products liberated by GH10 endo-xylanase activity. BACOVA_03427, encoded by PUL-XylL, displayed affinity for BGX and CX, but not WAX or undecorated xylooligosaccharides.

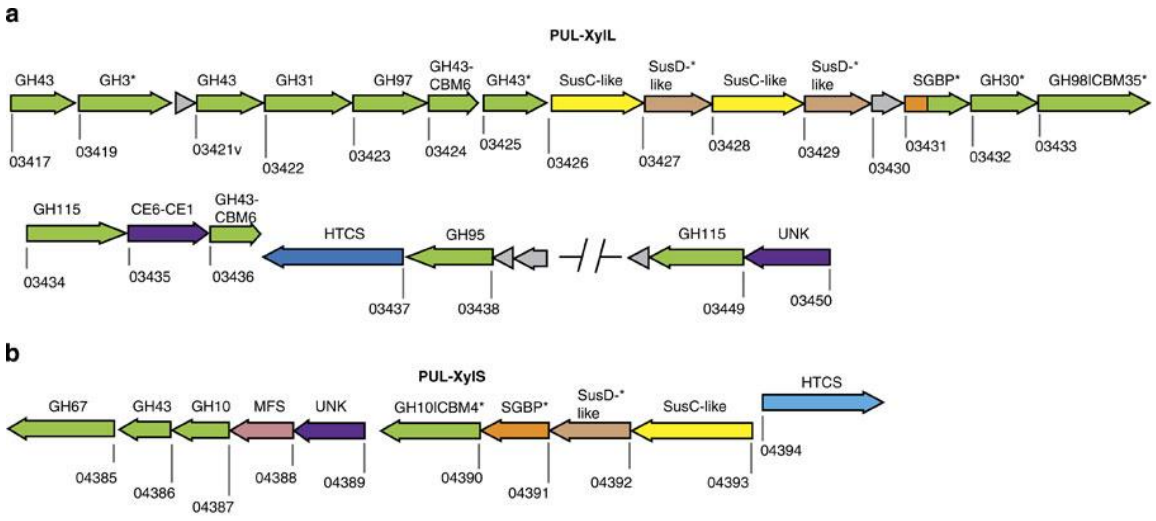


Figure 1.9. The xylan PULs encoded by *B. ovatus.* A schematic of the two xylan PULs encoded by B. ovatus, PUL-XylL and PUL-XylS respectively. Protein encoding ORFs are labelled according to protein function or CAZy sequence based family (if applicable). Figure adapted from Rogowski et al., 2015.

Three SGPB are also encoded by the xylan PULs of *B. ovatus*. The SGPB encoded by PUL-XylS displayed identical specificity as the SusD-like protein encoded by the same PUL, binding to undecorated xylooligosaccharides with a degree of DP >6 with affinities higher than the $SusD_h$ protein.

Only one SGBP encoded by PUL-XylL demonstrated binding activity. BACOVA_03431, which bound to both BGX and CX. Interestingly, BACOVA_03431 appears to be related to the GH10 family of xylanases but lacks the conserved catalytic nucleophile of the family. Similarly, the -2 subsite appeared to be abrogated. The SGBP, BACOVA_03431, appears to be an 'inactive' GH10 enzyme and is the first example of a catalytically inactive enzyme functioning as an SGBP.

Similar to the xyloglucan utilization model (Larsbrink et al., 2014), the bulk of xylan decoration removal occurs in the periplasm, and is performed by an array of exo-acting enzymes. Such enzymes display activities against xylan decorations in different contexts. For example, BACOVA_03422, the first GH31 family member to display α-xylosidase activity, liberates α-D-xylose units from the side chains of corn xylan but displays no activity against xyloglucan.

Deletion studies performed by Rogowski et al. (2015) again reinforce the differing roles of PUL-XylS and PUL-XylL in degrading different xylans. Deletion of PUL-XylS abrogated B. ovatus growth on simple xylans, but not complex xylans. Conversely, deletion of the larger PUL, PUL-XylL, prevented B. ovatus growth on complex xylans but not on linear xylans.

1.7 Objectives

The aims of this study are as follows;

1. To further elucidate the specificity the CBM65 family demonstrates for xyloglucan over linear glucans through the biochemical and structural characterization of the second CBM65 module (CBM6BB), encoded by the endoglucanase EcCel5A.
2. To explore the hierarchal degradation and mechanisms of yeast α-mannan utilization by *B. theta* through the biochemical characterisation of glycoside hydrolases which target α-mannosidic linkages. Elucidate the biochemical mechanisms by which *B. theta* targets the α-mannan of different yeast species, and to understand how the enzymatic ensemble encoded by the yeast α-mannan PULs and high mannose N-glycan PUL of *B. theta* displays specificity for similar, but structurally distinct glycans.
3. Explore β1,6-glucan utilization and degradation by *Bacteroides* through the biochemical and structural characterisation of the enzymes/proteins encoded by the β1,6-glucan degrading apparatus.

Chapter 2: Materials and Methods

2.1 Molecular biology

2.1.1 Bacterial strains and plasmids

The strains of *Escherichia coli* and plasmids utilized in this study are listed in the Tables 2.1 and 2.2.

Strain	Description	Usage	Reference
BL21 (DE3)	F'ompT,hsdS_B (r_B-m_B), gal.dcm. (DE3)	Protein Expression	Studier and Moffat (1986)
TUNER (DE3)	F'ompT,hsdS_B (r_B-m_B), gal.dcm. lacY1, (DE3)	Protein Expression	Novagen
TOP 10	F-*mcr*A Δ(*mrr-hsd*RMS-*mcr*BC)±80*lacZ*ΔM15 Δ*lac*X74 *rec*A1 *end* A1 *ara*D139 Δ(*ara*, *leu*)7697 *gal*U *gal*K Δλ- *rps*L *nup*G *ton*A *hsd*R	DNA manipulation	Invitrogen

Table 2.1. Bacterial Strains used in this study

Plasmid	Size (kb)	Genotype	Supplier
pET21a	5.4	Amp^{r*}, T7 lac laclq	Novagen
pET28a	5.4	Kan^{r**}, T7 lac laclq	Novagen
pET32b	5.9	Amp^{r*}, T7 lac laclq	Novagen
PET43b	5.9	Kan^{r**}, T7 lac laclq	Novagen

*Ampicillin resistant. **Kanamycin resistant.

Table 2.2. List of plasmids used in this study

2.1.2 Growth media

Growth media used in this study is presented in Table 2.3. Once fully dissolved, all growth media utilized were subjected to autoclaving, as detailed in section XX. Unless stated otherwise, growth media, like all chemicals and reagents in this study, were resuspended in 18.2 MΩ H_2O produced by a Millipore Milli-RO 10 purification system.

Medium	Reagents	Amount per litre	Notes
Lucia-Bertani (LB)	Bacto®tryptone	10 g	Final pH adjusted to 7.4 using NaOH.
	Bacto®yeast extract	5 g	
	NaCl	10 g	
LB-agar	Agar	2 g	Agar added to pre-prepared 100 mL LB media (2%), sterlilised and cooled before antibiotic addition
TYG media	Tryptone Peptone	10 g	Appropriate volumes of media were aliquoted into individual vessels prior to autoclaving. Media was left to cool before His-heam solution addition (1 µM per 1 mL)
	Bacto Yeast extract	5 g	
	Glucose	2 g	
	Cysteine (free base)	0.5 g	
	1 M KPO_4 pH 7.2	100 mL	
	Vitamin K solution, 1 mg/mL	1 mL	
	TYG salts	40 mL	
	0.8% $CaCl_2$	1 mL	
	$FeSO_4$, 0.4 mg/mL	1 mL	
	Rasazurin, 0.25 mg/mL	4 mL	
	His-Heam solution (0.2 M Histidine, pH 8.0	See notes	
Minimal Media	NH_4SO_4	1 g	For growth experiments, excluding the appropriate controls, concentrated monosaccharide/polysaccharide solutions were added for a final concentration of 5 mg/mL following sterilization and cooling. His-heam solution was added like TYG media.
	Na_2CO_3	1 g	
	Cysteine (free base)	0.5 g	
	1 M KPO4 pH 7.2	100 mL	
	Vitamin K solution, 1 mg/mL	1 mL	
	FeSO4 , 0.4 mg/mL	10 mL	
	Rasazurin, 0.25 mg/mL	4 mL	
	Vitamin B_{12} 0.01 mg/mL	0.5 mL	
	Mineral salts for defined medium	50 mL	
	His-Heam solution (0.2 M Histidine, pH 8.0)	See notes	

Table 2.3. Growth media

2.1.3 Selective media

Selection of bacterial cells harbouring the plasmid of interest was performed by the addition of the appropriate antibiotic stock solution (1000-fold dilution) to growth media < 55 ^{o}C. Antibiotic stocks used are displayed in Table 2.4.

Antibiotic	Stock Concentration	Final antibiotic concentration	Storage
Ampicillin	100 mg/mL in H_20	100 µg/mL	4 ^{o}C < 5 days
Kanamycin	10 mg/mL in H_2-0	10 µg/mL	4 ^{o}C < 5 days

Table 2.4. Antibiotic stocks

2.1.4 Sterilisation

Growth media, apparatus and solutions utilized in the growth of bacterial cells were sterilized via autoclaving at 121 ^{o}C, 32 lb/ inch^{-1} for 20 min. Heat sensitive reagents were sterilised by filtering through 0.22 – 1 µM pore, filter discs (Supor® Acrodiscs (Millipore)) using sterile syringes.

2.1.5 Storage of DNA and bacterial cells

Plasmid DNA was stored at -20 ^{o}C in 10 mM Tris/HCl buffer, pH 8.5 (Elution buffer, (Qiagen)). Bacterial colonies grown on agar plates were stored at 4 ^{o}C for a maximum of one week. Bacterial cells subject to long term storage were kept at -80 ^{o}C in 25% glycerol (v/v).

2.1.6 Plating of bacterial cells

Prior to spreading, a metal utensil was sterilised via submersion in 100% ethanol and passing through a flame and subsequently allowed to cool (~30 sec). Bacterial cells, suspended in media (100 µL), were pipetted onto the surface of the media-agar plate and spread evenly over the plate surface. Plated cells were incubated for 16 h at 37^{o}C

2.1.7 Anaerobic growth of *B. thetaiotaomicron* and *B. ovatus*

Bacteroides species were cultured in tryptone, yeast extract, glucose (TYG) media when optimal growth was required and to generate a 'viable' cell population following cryo-storage. Typically, a glass test tube containing 5 mL of TYG media (prepared as described in Section 2.1.3) was inoculated with 50 - 100 µL of *Bacteroides* cells from a glycerol stock (25% (v/v)) or directly from a previous culture. The tubes were subsequently incubated in an anaerobic cabinet at 37^{o}C until stationary phase was reached (2.0 OD_{600nm}).

For experiments assessing the ability of *Bacteroides* species to utilize differing carbon sources, minimal media was utilized (prepared as described in Section 2.13). Prior to experiments, cells were first washed via repeated centrifugation (5,000 x g, 10 min) and resuspension steps. Cells were pelleted, supernatant discarded and cells resuspended in minimal media (5 mL) at least three times. Sterile tubes containing 5 mL minimal media were inoculated with 200 µL of washed cells. The appropriate carbon source was added for a final concentration of 5 mg/ml via the addition of a concentrated sugar stock, sealed with cotton wool and incubated in an anaerobic cabinet (37^{o}C). Cell growth was monitored by measuring the optical density of cells within the anaerobic cabinet via the use of a spectrophotometer.

2.1.8 Centrifugation

Cultured bacterial cells (100 – 1000 mL) were harvested via low speed centrifugation (5,000 x g for 10 min) in 500 mL vessels at 4^{o}C using a JA-10 rotor and a Beckman

Coulter, J2-21 centrifuge. Bacterial cultures which were < 10 mL were centrifuged in 25 mL sterilin tubes at 5,000 x g for 10 min using a Hettich Zentrifugen bench-top centrifuge. For Cultures < 1 mL, Eppendorf tubes were used as vessels and pelleted using a HEREAEUS PICO17 (Thermo Scientific) bench-top centrifuge.

2.1.9 Chemically competent *E. coli*

The chemically competent E. coli cells utilized in this study were prepared utilizing a protocol derived from Cohen et al. (1972) (Cohen *et al.*, 1972). A single E. coli colony was used to inoculate 5 mL of LB media and incubated at 37^{o}C for 16 h with 180 RPM shaking. The overnight culture (1 mL) was subsequently used to inoculate 100 mL of LB media and incubated under the previously described conditions until an OD_{600nm} of 4.0. The cells were put on ice for 20 min, and pelleted via gentle centrifugation (280 x g for 5 min). The resultant supernatant was discarded and pelleted cells were resuspended in 3 mL of ice cold 0.1 M $CaCl_2$. The aforementioned process was repeated but with cell resuspension in 1 mL of ice cold 0.1 M $MgCl_2$. Cells were subsequently kept on ice for 2 h, completing the process. Competent cells were stored as 100 µL aliquots with 25 % (v/v) glycerol at -80^{o}C for long term storage.

2.1.10 Transformation of *E. coli*

Plasmid DNA (4 ng) was pipetted directly into competent *E. coli* cells (Section 2.1.9), mixed via gentle tapping and put on ice for 30 min under flame sterile conditions. Cells were subsequently heat shocked at 42^{o}C for 1.5 min, before being replaced back into ice for a further 2 min. LB media (500 µL) was added to cells before incubation at 37^{o}C/180 RPM shaking for 1 h. Following recovery, bacterial cells were subjected to gentle centrifugation (1.5 k RPM for 2 min). Upon centrifugation, 500 µL of supernatant was discarded, and transformed cells were gently resuspended in the remaining media supernatant by pipetting. Resuspended cells were then plated onto the appropriate antibiotic containing LB-agar plate as detailed in Section 2.1.3.

2.1.11 Replication of DNA

The *E. coli* TOP 10 cell strain was utilized to generate plasmid DNA. TOP 10 cells were transformed, as described in Section 2.1.10 and plated on selective LB plates as described previously. Following incubation and TOP 10 colony formation, a single colony was used to inoculate 5 mL of LB-media containing the appropriate antibiotic. Cells were allowed to incubate at 37^{o}C for 16 h with 180 RPM shaking. The cultured cells were subsequently harvested via centrifugation at 5,000 x g for 10 min. Resultant supernatant was gently removed and discarded before plasmid DNA was extracted from cells using a Qiagen QIAspin miniprep kit according to manufacturer's instructions.

2.1.12 Determination of DNA and protein concentration

The estimated concentration of purified DNA and protein was assessed by absorbance at 260nm and 280nm respectively using a spectrophotometer (NanoDrop 2000 UV-Vis spectrophotometer (Thermo Fisher Scientific Inc, USA) and applying the Beer-Lambert formula;

$$A = \varepsilon CI$$

Where by A = absorbance 280nm or 260nm; ε = molar extinction coefficient; I = length of light path (cm) and C = molar concentration of sample. The extinction coefficient of recombinant proteins was predicted using the following online tool; http://web.expasy.org/protparam/.

2.1.13 Extraction of genomic DNA from *Bacteroides cells*

Extraction of genomic DNA from *Bacteroides* cells was performed using the Sigma GenElute™ Genomic DNA kit, according to manufacturer's guidelines.

2.1.14 mRNA extraction and protection from *Bacteroides* cells

For experiments where-by the up-regulation of open reading frames in response to molecular cues needed to be ascertained in *Bacteroides*, *Bacteroides* cells were cultured anaerobically in minimal media and the appropriate carbon source as described in Section 2.1.7. Once mid-exponential phase (0.6 – 0.8 OD_{600nm}) was reached, the RNA of the bacterial cells was protected and extracted using the RNAprotect (Qiagen) reagent and the RNeasy Plus minikit (Qiagen), respectively, in accordance with manufacturer's guidelines.

2.1.15 Reverse transcription of mRNA to cDNA

The reverse transcription of mRNA to cDNA (extracted according to Section 2.1.14), was performed using the QIAGEN QuantiTect Reverse Transcription Kit, in accordance with manufacturer's guidelines.

2.1.16 Polymerase Chain Reaction

DNA amplification was performed by PCR, using the Novagen Hot start PCR kit (Novagen). A typical PCR reaction is shown in Table 2.5. In this study, standard PCR was used primarily to amplify the DNA of protein encoding open reading frames (ORFs) via specific primers, complementary to each strand of the DNA molecule and designed to the flanks of the DNA region to be amplified (Mullis and Faloona, 1987). Primers were designed such that ~20 bases were complementary to the target DNA sequence. Where possible, target sequences were extended/reduced to permit for optimal G/C content; ensuring a melting temperature (T_m) of ≥ 50^oC and a difference in T_m between the two primers not exceeding 5 ^oC. For cloning experiments, endonuclease restriction sites were engineered onto the 5' end of the primers. An additional six bases were added onto the primer, preceding the restriction site, to facilitate restriction enzyme cleavage and subsequent ligations in similarly cut vectors. The online service, Webcutter 2.0 (http://rna.lundberg.gu.se/cutter2/), was used to

ensure desired restriction sites were not encoded internally within the DNA sequence to be amplified. OligoCalc (http://www.basic.northwestern.edu/biotools/oligocalc.html) was used to calculate primer parameters using the following formula;

$$T_m = 64.9 + 41* (yG + zC -16.4)/ wA +xT +yG +zC)$$

Where w, x, y, z are the number of bases A, T, G, C in the sequence, respectively

Primers were synthesised and provided in lyophilized form by Eurofins MWG (Germany), and subsequently resuspended in highly distilled H_20 (Sigma-Aldrich, UK) with a volume appropriate to provide the desired concentration. Amplifications were performed using a PHC-3 thermocycler (Biorad). A typical thermo-cyclic reaction is shown in Table 2.6. The success of the amplification and integrity of the amplified products was analysed using agarose gel electrophoresis (Section 2.1.18)

Reagents	Volume
Highly pure H_2O	24 µL
10 x KOD buffer (10 X 1.2 M Tris-HCL, 100 mM KCl, 50 mM $(NH_4)_2SO_4$, 1% Triton X-100, 0.01% BSA, pH 8.0)	5 µL
dNTPs (2 mM)	5 µL
$MgSO_4$ (25 Mm)	4 µL
Template DNA (~70 ng/µL)	1 µL
Novagen KOD DNA polymerase (2.5 U/µL)	1 µL
Forward oligonucleotide primer (5 µM)	5 µL
Reverse oligonucleotide primer (5 µM)	5 µL
Total Volume	50 µL

Table 2.5. Typical PCR reaction

Program name	Event	Temperature	Duration	Number of cycles
Program 1	Denaturation	95°C	1 min	1
Program 2	Denaturation	95 °C	1 min	30
	Annealing	50 °C	1 min	
	Extension	68 °C	1 min/ 1 kbp	
Program 3	Final extension	68 °C	10 min	1
Program 4	Storage	10 °C	16 h	1

Table 2.6. Typical PCR thermocycle

2.1.17 Site-directed mutagenesis

Site directed mutagenesis (SDM) is a technique utilizing PCR to introduce nucleotide changes that encode a single, specific, amino acid change in the target protein. Forward and reverse oligonucleotide primers harbouring the mutation (in the centre of the primer) were designed to be complementary to the plasmid DNA template and codon optimised for *E. coli*. Primer parameters were calculated as described in Section 2.1.16. For SDM applications, primers were typically designed to ~30 bp and to

possess melting temperatures of ≥ 50 ^{o}C. The standard SDM-PCR reaction, utilizing a Novagen hot-start KOD DNA polymerase, is displayed in Table 2.7. The thermo-cyclic program is shown in Table 2.8. The success of the amplification and integrity of the amplified products was probed as described in Section 2.1.12

5 µL	10 x Reaction buffer (20 mM Tris-HCL, pH 7.5)
1 µL	dNTP mix (2 mM each)
1 µL	KOD DNA polymerase (2.5 U/µL)
20 ng	Plasmid dsDNA template
125 ng	Forward oligonucleotide primer
125 ng	Reverse oligonucleotide primer
	Ultra-pure sigma H_2O to a final volume of 50 µL

Table 2.7. A typical SDM-PCR reaction mixture

1 cycle at	95^{0}C for 2 min
18 cycles at	95^{0}C for 1 min
	55^{0}C for 1 min
	68^{0}C for 2 min/kb of plasmid length (vector + insert)

Table 2.8. Thermocycler program for SDM-PCR reaction

After successful PCR amplification, 1 µL (10 units) of Dpn1 restriction enzyme was added to the reactions and incubated for 1 h at 37 ^{o}C. The enzyme digested parental methylated DNA but not the newly amplified, unmethylated DNA (harbouring the mutation). The plasmid DNA containing the mutation was then transformed into *Escherichia coli* TOP10 competent cells. SDM-plasmid DNA was subjected to sequencing analyses by Eurofins MWG, and read in the forward and reverse direction using primers designed to the T7 promoter and T7T terminator regions contained in the pET vectors (Novagen).

2.1.18 Analysis of PCR result by agarose gel electrophoresis of DNA

Agarose gel electrophoresis was used to probe the success of PCR, allowing for the detection and separation of amplified DNA fragments by size. Agarose gels (1 %) were created by the addition of 0.5 agarose (low grade (Melford Ltd) or SeaKem® Gold Agarose (Lonza) into 50 mL of 1 x 0.9 mM Tris base/ 0.9 mM boric acid/ 0.2 mM EDTA, pH 8.0 (TBE) buffer and subjected to heating via microwave. Once sufficiently cooled (< 50^{o}C), ethidium bromide (1 µg/mL final concentration) was added, allowing visualisation of DNA. Gels were cast in mini-gel trays (Biorad) with an appropriate comb to provide wells. Once set, the gel-containing tank was filled with ~50 mL 1 x TBE buffer. Samples to be ran were prepared via the addition of 5 µL loading buffer (0.25% (w/v) bromophenol blue, 50% (v/v) glycerol, 10 x TBE buffer) to 5 µL of PCR reaction/DNA and loaded into wells. Hyperladder™ I markers were ran as DNA fragment size standards. Electrophoresis was ran at a constant voltage of 70 V for ~45 min. Gels were visualised using Bio-Rad Gel Doc 1000 system (Bio-rad)

2.1.19 Purification of PCR products

Products of PCR amplification were purified using the Qiagen QIAquick PCR purification Kit (Qiagen), according to manufacturer's instructions. To separate and purify amplicons of interest, DNA was extracted from the agarose gel (Seakem® Gold

Agarose) and purified using the Qiagen QIAquick Gel Extraction Kit (Qiagen), in accordance with manufactures' guidelines

2.1.20 Digestion of DNA using restriction enzymes

Restriction digest of double stranded DNA harbouring endonuclease restriction sites was performed according to manufacturer's guidelines (MBI Fermentas), providing optimal conditions for endonuclease activity. DNA subjected to endonuclease activity was purified as detailed in Section 2.1.19, preceding subsequent ligation steps.

2.1.21 Ligation of digested insert and vector DNA

Before being subjected to ligation, digested insert DNA and vector DNA were first purified as described in Section 2.1.19. DNA concentrations were first ascertained as detailed in Section 2.1.12, in order to provide a molar ratio of insert to vector of 3:1 in ligation reactions. Ligations were performed using the Novagen rapid ligation kit (Novagen) in accordance with manufacturer's instructions. A typical ligation reaction is shown in Table 2.8. Reactions were left at room temperature for 30 min and subsequently used (10 µL) to transform *E. coli* TOP 10 cells as described in Section 2.1.10.

Reagents	Volume
Vector (X ng/µL)	Concentration dependant
Insert DNA (Y ng/µL)	Concentration dependant
5 x Ligase buffer (250 mM Tris/HCL, pH 7.6, 50 mM $MgCl_2$, 25 µM ATP, 25 mM dithiothretol, 25% (w/v) polyethylene glycol 8000)	4 µl
T_4 DNA ligase (4 U/µL)	1 µl
Ultra-pure H_2O	Up to total volume
Total volume	20 µL

Table 2.9. Typical Ligation reaction

2.1.22 Automated DNA sequencing

DNA sequencing was performed using the MWG Biotech (Germany) Value Read service, to ensure the successful introduction of the desired mutation into the protein encoding DNA sequence (SDM) and/or to assess plasmid DNA integrity. Plasmid DNA samples (1.5 µg) were lyophilized using a vacuum pump. Typically, plasmids were sequenced from the forward and reverse direction using T7 promoter (TAATACGACTCACTATAGGG) and T7 terminator (CTAGTTATTGCTCAGCGGT) primers respectfully. If Internal regions of inserted DNA sequences were beyond flanking sequencing range, complementary internal primers were designed to probe sequence integrity. Bespoke primers for sequencing were also synthesised and utilized by MWG Biotech (Germany). Generated sequencing data was analysed using a multiple sequence alignment tool such as ClustalW (http://www.ebi.ac.uk/Tools/msa/clustalw2/).

2.1.23 Quantitative PCR

Quantitative PCR was performed using a Roche Light Cycler 480 real-time PCR system. For each sample of DNA to be quantified, 10 ng of cDNA (Section 2.1.15) was used and ran in duplicate. A typical qPCR reaction consisted of 5 µL SYBER green I master mix (Roche), 1 µL of forward and reverse primer (5 µM), 2 µL of cDNA template and 1 µL ultra-pure H_20. A typical qPCR reaction consisted of three cycles. 1) denaturing – 1 cycle, at 95^oC for 5 min.; 2) cycling – 45 cycles at 95^oC (5 sec), 50 ^oC (5 sec) and 72 ^oC (10 sec), respectively and 3) melt curve – 1 cycle at 95 ^oC (5 sec) and 65 ^oC for 1 min. The resultant data was analysed using the LightCycler® 480 version 1.5.0.39 software (Roche)

2.1.24 Over-expression recombinant proteins in *E. coli*

E.coli BL21 or TUNER cells were transformed as described in Section 2.1.10 with the plasmid encoding antibiotic resistance and gene of interest, and plated onto Luria-Bertani (LB) media-agar with the appropriate antibiotic (10 µg/mL (kanamycin) or 100 µg/mL (ampicillin)). The transformed cells were grown for 16 h at 37 ^oC, and were subsequently used to inoculate 1 L baffled flasks of LB broth media containing the selective antibiotic (10 -100 µg/mL final concentration). Bacterial cells were cultured at 37^oC with 180 RPM shaking to mid-exponential phase (A_{600} = 0.6). After cooling of bacterial cultures to room temperature, recombinant protein expression was induced via the addition of 1 mM (*E. coli* BL21) or 0.2 mM (*E. coli* TUNER) isopropyl β-D-1-thiogalactopyranoside (IPTG). Induced cells were incubated overnight at 16 ^oC with 150 RPM shaking. Cells were harvested via centrifugation at 5000 x g for 10 min. The bacterial pellet was subsequently resuspended in 10 mL Talon buffer (20 mM Tris/HCl, 300 mM NaCl, pH 8.0) and sonicated, lysing the cells. The sonicated pellet was centrifuged at 15, 000 x g for 30 min, leaving a supernatant containing the soluble fraction (cell free extract, (CFE)) and an insoluble fraction (the

pellet). The CFE containing the recombinantly expressed protein was then subjected to Immobilized Metal Affinity Chromatography (IMAC).

2.1.25 Purification of protein

2.1.25.1 Immobilised metal affinity chromatography (IMAC)

Immobilized metal affinity chromatography was utilized to purify recombinant protein. The column, containing 2 mL bed volume of Cobalt Talon Resin (ClonTech), was first equilibrated by passing 20 mL Talon buffer through the column. The cell free extract (CFE), containing the His-tagged recombinant protein, was loaded onto the column through a 1.2 AcroDisc syringe filter (Pall Corporation), binding the tagged protein to the resin. The column was washed using 20 mL Talon buffer, before the protein was eluted in a series of 10 mL elution steps using increasing concentrations of Talon buffer containing imidazole (5 mM and 100 mM respectively). Each fraction was subsequently analysed using SDS-PAGE for quality control. Elution fractions of sufficient purity were dialysed overnight in 4 L of buffer at an appropriate pH (selected to reduce protein precipitation), removing imidazole from the purified fraction. Protein concentration was analysed post-dialysis, as described in Section 2.1.12.

2.1.25.2 Ion-exchange and gel filtration chromatography

Recombinant proteins of insufficient purity were subjected to further purification via ion-exchange chromatography and/or gel filtration, as were proteins entering crystal trails. An Akta pure system (GE healthcare) was utilized for both ion-exchange and gel filtration chromatography. Anion exchange chromatography was performed using a CaptoQ (GE healthcare) column. Proteins subjected to anion-exchange were first dialysed overnight into 20 mM TRIS/HCL pH 8.0 (Buffer A), concentrated if required (See Section 2.1.27) and subsequently loaded onto the pre-equilibrated column using a 1 mL loop (maximum of 10 mg total protein). A reservoir of 20 mM Tris/HCl,1 M

NaCl, pH 8.0 solution (Buffer B) was utilized to provide an increasing salt gradient over time. Fractions (1.5 mL) above an absorbance threshold of 0.02 absorbance units (280nm) were collected and subsequently probed by SDS-PAGE analysis.

Gel filtration was performed using a HiLoad™ 16/60 Superdex™ 200 Prep grade column (GE Healthcare). Preceding the injection of protein, the column was first equilibrated by passing two column volumes of buffer A (20 mM Tris, 150 mM NaCl, pH 8.0) through the column. The protein sample (concentrated to 1 mL), was subsequently loaded into a 1 mL loop and injected onto the column. Buffer A was passed through the column at a flow rate of 1 mL min^{-1}. Protein fractions (1.5 mL) exceeding the set absorbance threshold were analysed in the same manner as described for anion-exchange.

Following SDS-PAGE analysis, samples of sufficient purity were pooled and buffer exchanged into a buffer of choice through repeated concentration/dilution steps (See Section 2.1.27).

2.1.26 SDS-PAGE analysis

The relative purity and size of recombinantly expressed proteins was analysed via SDS-PAGE. Polyacrylamide gels of 12.5 % (v/v) were utilised in this study. The polyacrylamide gels consist of two layers, a stacking layer and a larger resolving later respectively. The composition of each respective layer is displayed in Table 2.9. Gels were cast in 12 cm x 10 cm glass plates (ATTO Corporation (Genetic Research Instruments) and sealed via rubber bands. The resolving layer was first poured into the glass cassette, filling two 2/3s of the cassette, and topped off with water. Once the resolving layer had polymerised, additional water was discarded before the final stacking layer was poured. A comb (providing wells) was placed into the stacking layer, and allowed to set. Before gel use, the rubber seal and comb were removed.

Protein samples were prepared by the addition of 7 µL SDS loading buffer (Table 2.9) to 10 µL of purified protein fractions. Quality control samples generated during a typical IMAC purification, such as the insoluble fraction, cell free extract, un-bound flow through and the column wash, were prepared by adding 7 µL of SDS loading buffer to 1 µL of the respective fraction. Samples were boiled in a heat block for 3 min and centrifuged at 13,000 RPM for 5 min before being loaded onto the gel. Protein standards (Sigma), providing M_r markers, were ran with protein samples to provide an estimation of protein size.

SDS-PAGE was performed by placing the gel containing plates within the electrophoresis tank, which is filled with running buffer (Table 2.9). Samples were loaded onto the gel via pipetting and subjected to electrophoresis at a current of 35 mA per gel. Protein bands were stained upon the completion of electrophoresis via submerging the acrylamide gel in Coomassie Blue stain (0.4% Coomassie Brilliant Blue R, 10% (v/v) glacial acetic acid) for ≥ 30 min and visualised using a Bio-Rad Gel Doc 1000 system (Bio-rad).

Reagent	***Volume/Amount***
Resolving Gel (12.5%)	For 4 gels
0.75 M Tris/HCl buffer, pH 8.8, 0.2 % SDS	9.4 mL
40% Acrylamide	5.8 mL
H_20	3.5 mL
10% (w/v) Ammonium persulphate	90 µL
TEMED	30 µL
Stacking gel	
0.25 M Tris/HCl buffer, pH 8.8, 0.2 % SDS	3.75 mL
40% Acrylamide	0.75 mL
H_20	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, 0.2% SDS	5 mL
Glycerol	25% (w/v)
β-mercaptoethanol	2.5 mL (w/v)
Bromophenol blue dye	0.1%
Running Buffer	
32 mM Tris/ 190 mM glycine, pH 8.3	350 mL
SDS	0.1%

Table 2.9 SDS-PAGE buffers

2.1.27 Concentrating protein and buffer exchange

Purified protein fractions were concentrated using 20 mL Vivaspin™ centrifugal concentrators (Vivasceince) with appropriate M_r cut offs, dependant on protein size (30 or 10 kDa respectively). Centrifugation (5,000 x g) was performed using a swing out – MSE Mistral 3000i desktop centrifuge (MSE, UK). For crystal screens where buffer exchanging was a necessity, the process was repeated via extensive dilution/concentration cycles, where by additional buffer/H_2O was added to the protein-containing compartment and centrifuged until a low volume (< 1 mL). The buffer exchange process was repeated at least four times.

2.1.28 Protein crystallization screens

The sitting vapour diffusion method was utilized for all crystal screens performed in this study. Typically, three commercial screens were utilized unless stated otherwise (JCSG+, PACT and STRUCTURE (Qiagen)). Screens were pipetted into the 96- well plates by hand. In general, 80 µL of screen solution was added to the respective main-well. The protein of interest (mg/mL) was pipetted using a mosquite™ robot (TTP Labetch), into the two sub-wells per condition with a 1+1 and 2+1 program, providing 1 µL of protein plus 1 µL of screen solution and 2 µL of protein samples plus 1 µL of screen solution, respectively. Attempts at the co-crystallisation of protein and ligand were performed via the addition of the respective ligand (concentration variable) to the protein of interest, before robot distribution. Plates were visualised using a Leica MZ-6 microscope (Leica MICROSYSTEMS)

2.2 Biochemistry

2.2.1 Enzyme assays

Unless explicitly stated, all enzyme assays in this study were performed at 37^{o}C and performed in triplicate in the presence of bovine serum albumin (0.1 mg/mL (w/v)). Kinetic graphs were plotted using GraphPad Prism 5.0 Likewise, the kinetic parameters generated were analysed and calculated using the aforementioned software using the appropriate form of regression

2.2.1.1 Glucose and Mannose detection using mannose assay kit (Megazyme)

A megazyme detection kit (D-Mannose, D-Fructose and D-glucose (K-MANGL)) was used for assays whereby the release of glucose or mannose by glucosidases or mannosidases was quantified, respectively. For the detection of glucose, glucose released by the enzyme of interest is phosphorylated by hexokinase and adenosine-5'-triphosphate (ATP) to glucose-6-phosphate (G-6-P). G-6-P is subsequently oxidised by $NADP^{+}$, in the presence of G-6-P dehydrogenase, to gluconate-6-phosphate thereby simultaneously reducing NADP+ to NADPH. The formation of NADPH can be quantified at A_{340nm} . The measurement of mannose by the kit relies on two additional enzymes to those described above, in addition to the hexokinase and G-6-P hydrogenase. Liberated mannose is converted to mannose-6-phosphate by a hexokinase, before being converted to fructose-6-phosphate (F-6-P) by phosphomannose isomerase (PMI). The generated F-6-P, is then converted to G-6-P by phosphoglucose isomerase. The subsequent steps, leading to the reduction of $NADP^{+}$, are the same as described for glucose detection.

The reactions were performed with the appropriate buffer for the enzyme of interest. The reaction mixture, minus enzyme, was aliquoted into glass cuvettes (500 µL) and allowed to equilibrate at 37^{o}C, as was enzyme before being added. Absorbance was measured overtime at the aforementioned OD using a Pharmacia Ultrospec 400

spectrophotometer. The rate of monosaccharide (glucose/mannose) release is stoichiometric to $NADP^+$ reduction and can be quantified using the rate of NADP+ reduction over time and the extinction coefficient of NADPH (6300 M^{-1} cm^{-1}) respectively.

2.2.1.2 Galactose detection using galactose assay kit (Megazyme)

A linked enzyme assay kit (Megazyme, (K-ARGA)) was used to quantify galactose released by galactosidases, and therefore deduce the kinetic parameters of the enzyme of interest. The methodology when performing these assays was as described in Section 2.2.1.1.

The enzymes utilized in this assay result in the reduction of NAD^+ to NADH, the absorbance of which can be measured at A_{340nm}. Galactose, liberated by the enzyme of interest, is converted to β-D-galactose by galactose mutarotase. β-D-galactose is subsequently oxidised by NAD^+ in the presence of β-galactose dehydrogenase . NAD^+ is thereby reduced to NADH which is stoichiometric to galactose release. The rate of galactose release can therefore be quantified, as detailed in Section 2.2.1.1, using the extinction coefficient (6200 M^{-1} cm^{-1})

2.2.1.3 Determining enzyme kinetic parameters using pNP-substrates

The kinetic paramters of exo-acting enzymes against pNP-substrates were deduced spectrophotometrically, with the same methodology to Section 2.2.1.1. Enzymatic cleavage of the pNP group results in a stoichiometric increase in A_{400nm} when the glyosidic linkage is cleaved. The resultant 4-nitrophenolate group has a molar extinction coefficient of 10,000 at A_{400nm} at pH 7.0.

2.2.1.4 5-Dinitrosalicyclic acid assay (DNSA)

Reducing sugar assays measure the reducing ends of sugars. As a polysaccharide is cleaved, a new reducing end is exposed. This can be quantified using DNSA reagent. In this study, a reducing sugar assay was used to assay the activity of endo-acting enzymes against polysaccharides. Reactions included 1 mg/mL BSA. At the appropriate time point, 500 µL of DNSA reagent (1% w/v) DNSA, 0.2% (v/v) phenol, 1% (w/v) NaOH, 0.002% glucose, 0.05% (w/v) $NaS0_3$) was added to a 500 µL aliquot of the reaction, terminating enzymatic activity. The mixture was then boiled for 20 min, placed on ice for 10 min and left to equilibrate at room temperature. The absorbance of the reactions was measured at 575 nm. A standard curve containing 0, 50, 100, 150, 200, 250, 300, and 400 µg of the appropriate monosaccharide was utilized to quantify the formation of reducing ends. Excluding the addition of monosaccharide, standard curve samples were identical to reaction samples, including polysaccharide to control for background reducing ends.

2.2.2 Fluorescent labelling of substrates

Substrate labelling with 2-aminobenzamide, enabling detection via HPAEC with a fluorescent detection module (Section 2.2.4), was performed using a Sigma GlycoProfile 2-AB labelling kit as per manufacturer's instructions.

2.2.3 Thin layer chromatography (TLC)

In this study thin layer chromatography was used to analyse the digestion products of an enzymatic degradation of polysaccharides and oligosaccharides and to analyse purified oligosaccharides. Briefly, 3 µL of sample and the appropriate standards (1 mM) is spotted onto a 10 cm tall TLC plate (Silicagel 60, 20 x 20, Merk). The width of the plate was sample dependant. Samples were spotted 1 cm from the bottom of the plate with 1 cm spacing between each sample and dried. The plate was placed into a 1 cm deep pool of running buffer (butan-1-ol/acetic acid/water (2:1:1 (v/v))) inside a

covered, glass tank. Running buffer was left to migrate within 1 cm of the top of the plate, before being dried and replaced into the tank to run again as done previously. The plate was subsequently dried thoroughly and submerged in developer (sulphuric acid/ethanol/water (3:70:20 (v/v)), 1 % (w/v) orcinol) for 30 s before again being dried thoroughly. The plate was developed through heating at ~80 ^{o}C for 5 min. Enzyme digestions to be analysed via Thin Layer Chromatography (TLC) typically included BSA (1 mg/mL) and were heat deactivated by boiling for 10 min. Prior to analysis by TLC, assays would be centrifuged at 13,000 x g for 10 min.

2.2.4 High Pressure Liquid Chromatography

High pressure liquid chromatography (HPLC) was performed in this study for both qualitative and quantitative purposes. HPLC was used to elucidate the products of the enzymatic degradation of polysaccharides/oligosaccharides and to verify the homogeneity of oligosaccharides generated via size exclusion chromatography purification (Section 2.1.25.2 Ion-exchange and gel filtration chromatography). Substrate depletion assays were analysed using HPLC to ascertain the catalytic efficiency of an endo-acting glycoside hydrolases against oligosaccharides. Samples were analysed using a pre-equilibrated, analytical CARBOPAC^{TM} PA-200 anion exchange column and BorateTrap (Dionex) connected to an automated Dionex DX500 and ICS3000 system (Dionex), which included a gradient pump, detector compartment, electrochemical detector and auto-sampler. Mono/oligosaccharides were detected using pulsed amperometric detection (PAD) using the following settings; E1 = +0.05, E2= +0.6, E3= -0.6. Typically, samples were eluted with 0-5 min 100 mM NaOH, 5-25 min 100 mM NaOH and a 0-75 mM NaAC linear gradient, with a flow rate of 0.5 mL/min and a maximum pressure of 2000 psi. The appropriate monosaccharide and oligosaccharides were used as standards at a concentration of

0.1 mM unless stated otherwise. For quantitative assays, an internal standard of fucose (0.03 mM) was included in reactions. Fluorescent detection of 2-aminobenzamide labelled substrates was achieved using a fluorometric detector module. Data was collected and manipulated using Chromeleon™ Chromatography Management System V.6.8 (Dionex) via a Chromeleon™ server (Dionex) and GraphPad Prism 5.0.

2.2.5 Binding Assays

2.2.5.1 Isothermal titration calorimetry (ITC)

The thermodynamic parameters of protein-ligand interactions were ascertained via isothermal titration calorimetry (ITC), performed using a MicroCal™ VP-isothermal titration calorimeter. Heats of dilution were minimised during ITC experiments by dialysing the protein of interest into buffer. Said buffer was subsequently used as a solvent for ligand. Protein and ligand samples were first centrifuged (13,000 RPM for 10 min) to pellet any impurities. Protein (typically 50 µM – 100 µM) was equilibrated in the cell of the machine (1.8 mL) at a constant temperature (25^{o}C). Ligand was injected into the protein containing cell in 28 titrations of 10 µL with 200 s spacing between each injection, with stirring. Temperature fluctuations are measured during ligand injection. An increase (exothermic) or decrease (endothermic) in temperature per injection over a titration, is indicative of a protein-ligand interactions.

The thermodynamic parameters of ligand binding generated via ITC, including the association constant K_a, binding enthalpy ΔH and the stoichiometry of binding (n) were interpreted using MicroCal Origin 7.0 software. The change in Gibbs free energy ΔG^0 and change in entropy ΔS^0 were derived using the following equation;

$$-RTIn K_a = \Delta G = \Delta H - T\Delta S$$

R is the gas constant (1.99 $cal.K^{-1}$ $.mol^{-1}$), T is the temperature in degree absolute (298.15k), ΔG is free energy and ΔS is the entropy of binding.

2.2.5.2 Affinity gel electrophoresis

Affinity gel electrophoresis was used as a qualitative screen of binding to assess protein-polysaccharide interactions. Single layer, non-denaturing, polyacrylamide gels were produced and ran utilizing the same apparatus and equipment as described in Section 2.1.26. A typical non-denaturing, native gel consisted of 7.5 % (w/v) acrylamide and native buffer (25 mM Tris and 250 mM glycine buffer, pH 8.3). Ligand binding was assessed via electrophoresing the protein/s of interest through a non-ligand containing and a ligand containing gel. Polysaccharides were incorporated into gels prior to polymerisation, typically at a concentration of 0.1 % (w/v) unless stated otherwise.

Protein samples were prepared via the addition of protein (5 µg total) to loading buffer (7 µL, 5 % (v/v) glycerol and 0.0025 % bromophenol blue)) and gently mixed. Both gels (ligand and non-ligand) were ran concurrently in the same tank, filled with native buffer and electrophoresed with a current of 10 mA per gel for ~ 2 h. BSA (5 µg total) was ran parallel to protein samples as a negative binding control. The resultant gels were stained and visualised following the same protocol as Section 2.1.26

2.2.6 Generation and verification of substrates

2.2.6.1 Purification of yeast α-mannan and mutant derivatives

Overnight cultures of yeast (*S. cerevisiae* or *S. pombe*) were used to inoculate 1 L non-baffled flasks of Yeast extract/peptone/dextrose (YPD) media. Yeast cells were cultured at 30 ^{o}C with 150 RPM shaking until fully grown. Cells were harvested via centrifugation at 6000 RPM for 10 min. The resultant yeast pellet was resuspended in

10 mL 20 mM citrate buffer, pH 7.0 per litre of cell culture. The resuspended yeast pellets were subsequently pooled and autoclaved at 126 ^{o}C for 90 min. Autoclaved cells were subjected to centrifugation at 10,000 RPM for 10 min, giving a supernatant containing extracted polysaccharides and a pellet. The supernatant (supernatant 1) was collected. The remaining pellets were again resuspended in 20 mM citrate buffer, pH 7.0 (1.5 x original resuspension volume) and autoclaved again (125^{o}C, 90 min). Following being autoclaved, the above centrifugation step was repeated and the resulting supernatant (supernatant 2) collected. Supernatants 1 and 2 were subsequently pooled and an equal volume of Fehling's reagent was added to the pooled supernatants. The mixture was incubated at 40 ^{o}C for 2 h with stirring. Following incubating, the solution was divided into sterilins and the resulting precipitate harvested by centrifugation (5,000 x g for 10 min), rendering a pellet and supernatant. The supernatant was discarded and the white pellets resuspended with 3 M Hydrochloric acid, giving a light green solution (~2 mL of 3 M HCl per pellet). HCl (1 mL at a time) was added to pellets if not fully dissolved. The resultant green resuspension was pooled together and slowly added to 100 mL of methanol/acetic acid (8:1) with gentle stirring and incubated for 15 min, giving a thick, white precipitate. The precipitate was subsequently pelleted via centrifugation (5000 RPM, 10 min) and the supernatant discarded. The pellets were further washed (x3) with 100 mL methanol/acetic acid (8:1) by evenly distributing methanol/acetic acid mixture amongst pellets, vortexing briefly and centrifuging for 1500 RPM for 5 min. A final methanol wash step (100% methanol) was performed before pellets were left to dry overnight . Once dry, each pellet was dissolved in ultra-pure H_20 and dialysed into H_20. When dialysis was complete, the solution was freeze dried, rendering purified α-mannan.

2.2.6.2 Purification of oligosaccharides by size exclusion chromatography

The respective oligosaccharide mixtures was freeze dried and re-suspended in a minimal volume of 50 mM acetic acid (<5 mL). Size exclusion chromatography was utilized to separate the mixture according to molecular mass. Shorter oligosaccharides (D.P <6) were separated using P2 Bio-Gel (Bio-Rad), whilst longer oligosaccharides (D.P >5) were separated using P4 Bio-Gel. The respective Bio-Gels were packed into two Glass Econo-columns™ (2.5 cm x 80.0 cm), and connected in series with flow-adapters (Bio-Rad) before being equilibrated with degassed 50 mM acetic acid, pumped through the columns at 0.2 mL min^{-1} using a peristaltic pump (LKB Bromma 2132 microperpex™). The digestion mixture was loaded directly onto the column bed and allowed to settle before the acetic acid reservoir was pumped through the columns at 0.2 mL min^{-1}. After the column dead volume (200 mL) had run through both columns, 2 mL fractions were collected with 3 µL of every 10th fraction (or more frequently for regions of interest) via TLC, and the D.P. of the oligosaccharides were estimated by comparing their migration with appropriate oligosaccharide standards. Selected pooled fractions were then freeze dried and analysed further through MALDI mass spectrometry (Section 2.2.6.3) and High Pressure ion exchange Chromatography (HPAEC) (Section 2.2.4).

2.2.6.3 Mass spectroscopy

Purified manno-oligosaccharides were analysed via MALDI-TOF MS (Matrix-Assisted Laser Desorption Ionisation – Time Of Flight Mass Spectrometry) to measure the molecular weight of the purified oligosaccharide. The purified oligosaccharide (100 µM) was dissolved in water and diluted 10 fold with matrix (2,5-dihydroxybenzoic acid (DHB) 10 mg/mL (w/v) . The mixture (2 µL) was then spotted onto a MALDI plate and dried to co-crystallize the analyte and matrix before being analysed with the help of

Dr. Joe Grey and Dr. Adam Jackson. MALDI-TOF mass spectrometry analyses were performed with an AB4700 MS instrument (Applied Biosystems) in positive ion, reflector mode with 100 shots taken per spectrum.

2.3 Bioinformatics

2.3.1 Alignments

The sequences of protein encoding ORFs analysed in this study were acquired via the integrated microbial genomes (IMG) gene search tool (https://img.jgi.doe.gov/cgi-bin/mer/main.cgi?section=FindGenes&page=geneSearch). Alignments of DNA and protein sequences in this study were performed using ClustlW2 (http://www.ebi.ac.uk/Tools/msa/clustalw2/)

2.3.2 Prediction of prokaryotic signal peptides

The presence/type of signal peptide encoded by gram negative bacterial proteins explored in this study was predicted utilizing LipoP 1.0 software, hosted at http://www.cbs.dtu.dk/services/LipoP/

Chapter 3: Understanding How Non-catalytic Carbohydrate Binding Modules Can Display Specificity for Xyloglucan

3.1. Introduction

Plant biomass represents a major nutrient source for the bacterial ecosystems of the rumen and large bowl of mammals (Mackie and White, 1990). This abundant carbon source is exploited through an array of microbial plant cell wall degrading glycanases encoded by such bacterium. The plant cell wall however, is highly recalcitrant to enzymatic degradation owing to its heterogeneous and complex structure (See Introduction Section 1.3.1. Because of such complexity, many of the plant cell wall degrading glycoside hydrolases encoded by ruminant bacteria often encode for non-catalytic carbohydrate binding modules (CBMs). Primarily, CBMs potentiate catalytic activity through targeting or proximity effects (Bolam *et al.*, 1998) .

CBMs are grouped into amino acid sequence based families in the CAZY database (www.cazy.org) (See Introduction, Section 1.2). Recognition of target ligands by CBMs can be highly specific or promiscuous (Montanier *et al.*, 2011). Usually, in CBMs which display a degree of plasticity, specificity is conferred through the recognition of a structure common to the target ligands. For instance, all CBMs which bind β-glucans (β1,4-glucans) also bind xyloglucan (β1,4-glucose backbone, α1,6-xylose single substitutions). However, as of yet, no evidence exists to suggest that such CBMs target the O6 α-xylose decorations of the polysaccharide, but merely accommodate them (Najmudin *et al.*, 2006). Ligand recognition in CBMs is therefore diverse and often compliments the activity/biological context of its cognate catalytic domain.

A previous report identified an endoglucanase (*Ec*Cel5A) encoded by *Eubacterium cellulosolvens.* The enzyme two CBM domains that could not be designated into a

family based on sequence based similarities (molecular architecture shown in Figure 3.1). Reported target substrates of the *Ec*Cel5A enzyme include lichenin, a β1,3-β1,4 mixed linked glucan, carboxymethylcellulose and to a lesser extent, oat spelled xylan (poly-β1,4-xylose) (Yoda *et al.*, 2005). The two CBMs, CBM65A and CBM65B share 73 % sequence identity. It was previously reported that the two CBM65 modules of EcCel5a displayed binding to acid swollen cellulose and lichenin (β1,3/β1,4 glucan), barley β-glucan (β-1,3/β-1,4 glucan) and xyloglucan, indicating an unquantified preference for xyloglucan.

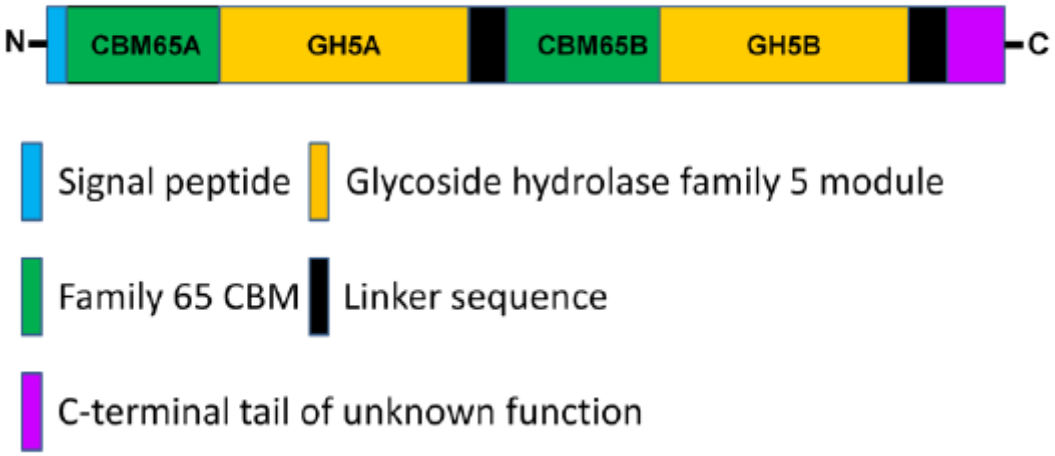


Figure 3.1. The molecular architecture of *Ec*Cel5a. CBM65A comprises residues 27 – 120 of *Ec*Cel5a. CBM65B comprises residues 581 – 713 of *Ec*Cel5a.

Qualitative binding experiments suggest that the CBM65s of *Ec*Cel5a display a preference for xyloglucan. This was explored during my Masters of Research degree. CBM65A and CBM65B were expressed recombinantly and the specificity and mechanism of polysaccharide recognition was investigated through quantitative binding assessments and a structure informed strategy to probe the ligand binding site of CBM65A. No protein in complex with ligand was obtained during the Masters project

and so the exact mechanism of ligand recognition by the CBM65s was unclear. The data generated throughout the Masters project is incorporated into the following results chapter to provide a clear narrative.

3.2 Objectives

The objectives of this chapter are to explore the binding specificity and ligand recognition of the CBM65 family, informed by pre-existing biochemical data generated during the Masters of Research degree, and to provide further insight into how differential ligand recognition of β-glucans is achieved in this newly characterised CBM family.

3.3 Results

3.3.1 Biochemical characterisation of CBM65 specificity

3. 3.1.1 Protein expression and purification

In order to investigate the binding specificities of the two CBM65s of *Ec*Cel5a, DNA encoding CBM65A and CBM65B, residues 27 – 170 and residues 581 – 713 of ful length EcCel5A respectively, were synthesised with codon usage optimised for expression in *Escherichia coli* and sub-cloned into a pET28a derivative by NZYTech Ltd, Portugal. The CBM encoding plasmids were used to transform *E. coli* TUNER competent cells. Cells were cultured and recombinant protein expression was induced according to standard protocol as outlined in Materials and Methods, Section 2.1.24. The recombinant protein contained an N-terminal His_6-tag and were thus purified by immobilised metal affinity chromatography. Purity was assessed by SDS-PAGE. Both CBM65A and CBM65B ran with an apparent M_r of 17 kDa and 18 kDa respectively, consistent with their predicted size. For crystallography, CBM65A and CBM65B were purified further using size exclusion chromatography using a HiLoad 16/60 Super-dex 75 colomn (GE healthcare). Chromatography was carrired out in 20 mM Tris-HCl

buffer, pH 7.5 containing 150 mM NaCl. Following purification, confirmed by SDS-PAGE, recombinant protein was buffer exchanged into Sigma-ultra pure H_2O using a 10 kDa molecular mass centrifugal concentrator (Amicon).

3.3.2. Specificity of the CBM65s

3.3.2.1 Qualitative Evaluation of CBM65 ligand binding.

The binding specificities of CBM65A and CBM65B were explored during the Master of Research degree using affinity gel electrophoresis (AGE) to provide a qualitative assessment of ligand binding. The results of AGE are displayed in Figure 3.2. Recombinant protein was electrophoresed through a non-denaturing, polyacrylamide gel in the presence and absence of ligand (varying concentrations (% w/v)). Protein-ligand interactions retard recombinant protein migration and are indicative of binding when compared to the migration of a non-binding standard of bovine serum albumin (BSA). Recombinant CBM65A and CBM65B bound to a several soluble β-linked glucans. Binding was observed to barley β-glucan, a β1,3/β1,4 mixed linked glucan and to decorated β1,4-linked glucans, such as xyloglucan (α1,6-xylose sidechain substitutions) and hydroxyethylcellulose (HEC). Neither CBM65A nor CBM65B bound to other β-linked polysaccharides, including xylans, glucomannans or galactans. No binding was observed to α-linked glucans, indicating that the two CBMs preferentially target β1,4 – linked glucans. The inability of the CBMs to bind xylan suggests that the O6 moieties of the glucose residues which comprise the backbone of the β1,4 – linked glucans are specificity determinants, whilst the ability of the two CBM65s to bind to the decorated xyloglucan suggest a degree of side-chain tolerance or indeed specificity in the binding site of the two proteins.

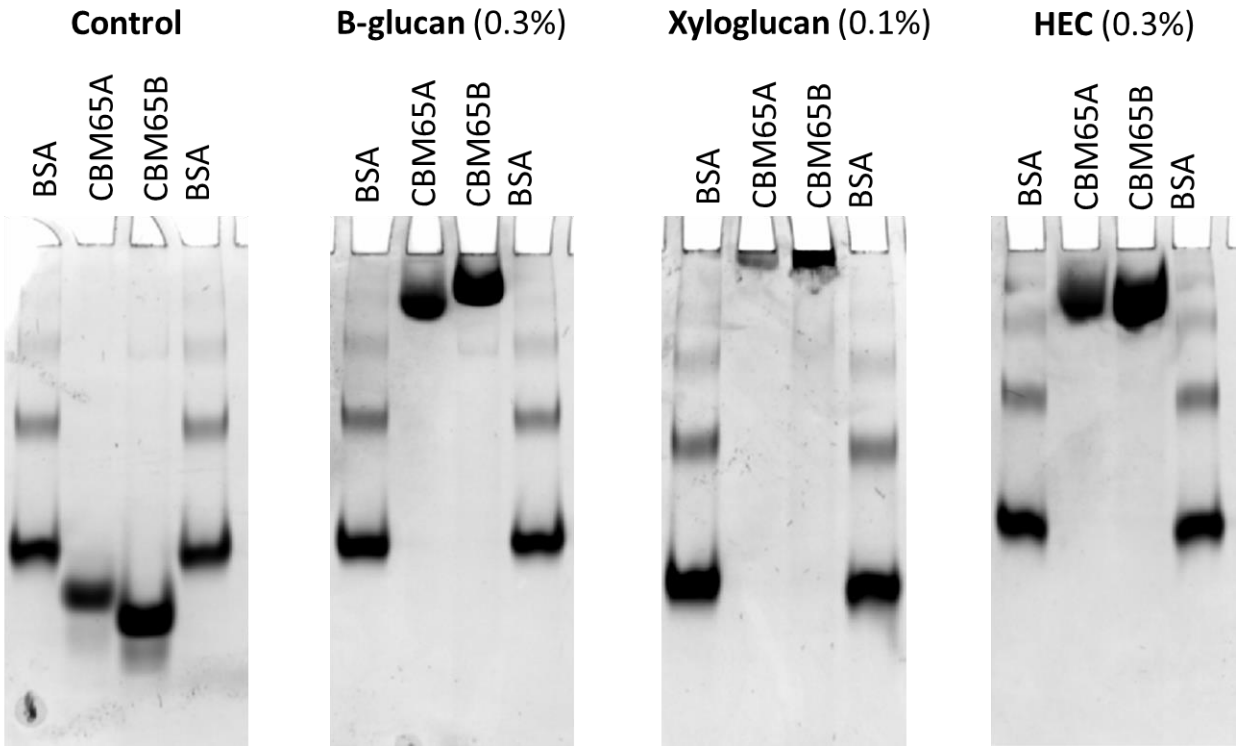


Figure 3.2. Affinity gel electrophoresis of CBM65A and CBM65B against soluble polysaccharides. CBM65A and CBM65B were electrophoresed through a non-denaturing, polyacrylamide gel in the presence and absence (Control) of ligand. Bovine serum albumin (BSA) acts as the non-binding control. Figure adapted from Luis et al., 2012.

3.3.2.2. Quantitative Evaluation of CBM65 ligand binding

To further probe the binding specificities of the two CBM65 modules and to quantitatively assess the thermodynamic parameters of binding, isothermal titration calorimetry (ITC) was performed as described in Materials and Methods, Section 2.2.5.1. The thermodynamic parameters of ligand binding generated via ITC, including the association constant K_a, binding enthalpy ΔH and the stoichiometry of binding (n) were interpreted using MicroCal Origin 7.0 software. The change in Gibbs free energy ΔG^0 and change in entropy ΔS^0 were derived using the equations, also described in Materials and Methods, Section 2.2.5.1.

Protein	Ligand	K_a (M^{-1})	ΔG (kcal $mole^{-1}$)	ΔH (kcal $mole^{-1}$)	$T\Delta S$ (kcal $mole^{-1}$)	n
CBM65A	β-Glucan	1.5 (±0.0) x 10^4	-5.7	-11.8 ± 0.1	-6.1	1.01 ± 0.0
CBM65A	Xyloglucan	1.7 (±1.4) x 10^5	-7	-18.0 ± 0.3	-11.0	1.03 ± 0.0
CBM65A	Hydroxyethylcellulose	1.2 (±0.0) x 10^4	-5.5	-11.1 ± 0.2	-5.6	1.02 ± 0.0
CBM65A	Cellohexaose	2.1 (±0.3) x 10^3	-4.3	-11.3 ± 5.5	-6.9	1.0 ± 0.1
CBM65A	Xyloglucan Heptasaccharide (XXXG)	5.62 (±0.1) x 10^3	-5.1	- 9.6 ± 0.4	-4.5	1.03 ± 0.0
CBM65B	β-Glucan	8.2 (±0.2) X 10^3	-5.3	-12.1 ± 0.1	-6.8	1.0 ± 0.0
CBM65B	Xyloglucan	1.01 (±0.1) X 10^5	-6.8	-4.2 ± 0.1	-7.7	1.0 ± 0.0
CBM65B	Hydroxyethylcellulose	1.42 (±0.4) X 10^4	-5.6	-7.0 ± 0.1	-1.4	1.0 ± 0.0
CBM65B	Cellohexaose	2.31 (±0.2) X 10^3	-4.6	-2.8 ± 1.0	-8.2	1.0 ± 0.3
CBM65B	Xyloglucan Heptasaccharide (XXXG)	1.7 (±0.0) x 10^3	-4.4	-10.2 ± 0.0	-5.8	1.0 ± 0.0

Table 3.1. Affinity and thermodynamics of CBM65A and CBM65B binding to soluble polysaccharides and oligosaccharides

The results of ITC of CBM65A and CBM65B, generated during the Masters of Research project are shown in Table 3.1. Both CBM65A and CBM65B display identical binding specificities, consistent with the AGE data. The CBM65s demonstrated affinity for the polysaccharide xyloglucan with a (K_a ~10^5 M^{-1}), whilst showing a ~10-fold reduction in affinity to the undecorated, mixed linked barley β-glucan and to hydroxyethylcellulose.

Binding of the two CBM65s was quantifiably assessed against oligosaccharides. Moderate binding affinities were observed against cellohexaose, with K_a values of ~10^3 M^{-1} for both CBMs. Unquantifiable binding to the oligosaccharides cellopentaose and cellotetraose was observed for the CBM65s with an estimated K_a value of << ~10^3 M^{-1}. No binding to cello-oligosaccharides with a degree of polymerisation (D.P.) less than four was observed. The CBMs displayed preferential binding affinities (Ka of ~10^3 M^{-1}) for xyloglucan heptasaccharide (Glc_4Xyl_3 or XXXG), an oligosaccharide consisting of four β1,4-linked glucose residues with single xylose decorations at the O6 position of three glucose residues; the reducing end glucose is unsubstituted. The binding affinities and specificities of the two CBM65s, preferentially targeting β1,4-linked and mixed β1,3-β1,4-linked ligands, is consistent with the documented activities of the CBM65s cognate enzyme, EcCel5a.

Typical of CBM binding to soluble ligands, the thermodynamic parameters of CBM-ligand binding suggests that binding was primarily enthalpy driven with a canonical entropic penalty on binding affinity, indicated by a decrease in entropy (Boraston *et al.*, 2004). The stoichiometry of CBM binding to soluble polysaccharides was calculated by altering the concentration of ligand until an *n* value of 1 was achieved (*n*

= number of binding sites), assuming a single site binding model. To determine the number of sugar residues occupied per CBM protomer, the following calculation was performed;

$$\text{Occupancy (No. of sugar residues per protein molecule)} = \frac{(\text{Concentration of polysaccharide used in ITC titration (g/L)} \div \text{Mr of sugar monomer (Da)}) \times 1000\ (\text{mM})}{\text{Concentration of polysaccharide for n=1 (mM)}}$$

CBM65A and CBM65B on average bound once per ~11 sugar residues in tandem, thereby indicating that the CBM65s bind to target ligands internally.

3.3.3 Crystal Structure of the CBM65s

To investigate ligand recognition and specificity determinants in the binding site of the CBM65s, attempts to co-crystallise the CBM65s with cellohexaose or XXXG were performed. Co-crystallisation screens of CBM65A (80 mg/mL) using the hanging drop vapour diffusion method (described in xxxxx) with cellohexaose (10 mM) and XXXG (10 mM) produced crystals in the conditions 0.2 M ammonium sulphate, 0.1 M sodium acetate, pH 4.6, 30 % v/v PEG 2000 MME and 0.2 M lithium sulfate, 0.1 M Tris, pH 8.5, 25 % w/v PEG 1500 respectively. The crystals of CBM65A in the presence of cellohexaose diffracted with a resolution of 1.6 Å and the data set was solved by Dr. Arnuad Baslé using molecular replacement of a selenomethionine-SAD derived CBM65A structure previously generated Dr. Shabir Najmudin (Lisbon, Portugal). No ligand was bound to the protein *in crystal.*

CBM65B (80 mg/mL) was co-crystallised with XXXG (100 mM) in 200 mM ammonium acetate, 100 mM tri-sodium citrate, pH 5.6, 22 – 30 % (w/v) PEG 4000. The CBM65B-XXXG crystals are shown in Figure 3.3. Upon harvesting, the crystals were cryo-protected in reservoir solution and 20 % PEG 400 containing XXXG (100 mM). The crystals were subjected to X- ray diffraction at the DIAMOND light source (Harwell, UK), beam line IO2 and diffracted at a resolution of 2.4 Å, generating a data set which was subsequently processed using the CCP4 suite. The CBM65B-XXXG protein-ligand complex had a $P4_32_12$ space group, belonging to the hexagonal system. The phase problem was solved utilizing the apo-crystal structure of CBM65B (previously solved by Dr. Shabir Najmudin (Lisbon, Portugal), in conjunction with MOLREP. Crystal fishing, data collection, structure building and refinement were all performed by Dr. Arnaud Baslé, structural biology lab, Newcastle University. Data collection and refinement statistics are presented in Appendix A.

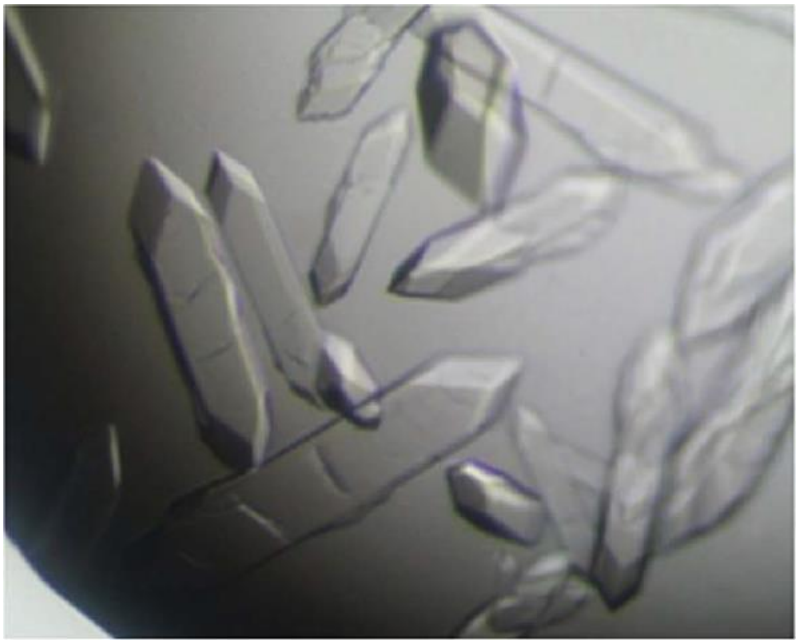


Figure 3.3. Crystals of CBM65B and XXXG obtained via hanging drop vapour diffusion method. CBM65B (80 mg/mL)/XXXG (100 mM) crystals were obtained in the following conditions; 200 mM ammonium acetate, 100 mM tri-sodium citrate, pH 5.6, 22 – 30 % (w/v) PEG 4000 . Figure taken from Venditto et al.(Venditto *et al.*, 2013).

3.3.3.1. The apo-crystal structure of CBM65

The solved structure of apo-CBM65A reveals that CBM65A protein adopts a β-sandwich fold. Interestingly, unlike most CBMs which assume a β-sandwich fold, the CBM65s seemingly lack a stabilising calcium ion coordinated by loops connecting the β-strands (Simpson et al., 2002). The β-sandwich is comprised of two β-sheets connected by loops. β-sheet 1 is formed by five anti-parallel β-strands (termed β1, β9, β3, β7, β6) and is convex in shape. The concave β-sheet 2, which comprises the solvent accessible surface of the protein, consists of four anti-parallel β-strands (β2, β8, β4, β5 respectively). The loop connecting β4 and β5 in CBM65A is longer than the structurally equivalent loop in CBM65B and is seemingly of functional significance (See 3.3.4.2 Understanding ligand recognition in the CBM65s). An α-helix (Lys^{125} to Tyr^{132}/ Lys^{668} to Tyr^{675}) completes the structures of CBM65A and CBM65B respectively. A protein schematic of the apo-structures of CBM65A and CBM65B is shown in Figure 3.4A

3.3.3.2 The ligand binding site of CBM65

The crystal structure of CBM65B in complex with XXXG shows a distinct cleft like structure, 20 – 25 Å in length presented by a β-sheet, typical of type-B CBMs which bind internally to target ligands (Figure 3.4B). The co-crystallised structure reveals the binding site harbours a significant number of tryptophan residues (Figure 3.4C). An overlay of the apo-CBM65A structure indicates these residues are strictly conserved (Figure 3.4D). Indeed, site directed mutagenesis of these residues to alanine diminished ligand recognition (see 3.3.4. Site Directed mutagenesis of CBM65A and CBM65B), thereby undoubtedly confirming that location of the ligand binding site within this cleft like structure.

3.3.3.3 β1,4-backbone recognition of xyloglucan heptasaccharide by CBM65B

Analysis of the CBM65B-XXXG co-crystallised structure revealed that four tryptophan residues (Trp^{602}, Trp^{607}, Trp^{646}, and Trp^{651}) located inside the ligand binding cleft, coordinate recognition of the β1,4-linked glucose backbone of XXXG. Recognition of the backbone is mediated primarily through apolar contacts. Trp^{646} makes a parallel apolar interaction with the undecorated, reducing end glucose henceforth termed Glc-1 (residues downstream from this residue are termed Glc-2,3 and 4 respectively, with Glc-4 being the non-reducing end) . Trp^{651} coordinates both Glc-2 and Glc-3 whilst Trp^{602} interacts with Glc-4 through the formation of parallel hydrophobic contacts. The β1,4-backbone of XXXG is further coordinated via perpendicular apolar interactions between Glc-3 and Trp^{607}. The orientation of the tryptophan residues within the cleft induce a twisting conformation between Glc-2 and Glc-4 of the tetrasaccharide backbone. The pyranose ring of Glc-1 is coordinated at an 180º angle relative to glucose-2. The only polar contact between CBM65B and the glucose backbone of XXXG is mediated via Gln^{653}, which forms a hydrogen bond with the O2 and O3 moieties of the non-reducing glucose residue (Glc-4).

3.3.3.4 CBM65B recognition of the α1,6-xylose substitution of xyloglucan heptasaccharide.

Xyl-2 (corresponding to the O6 linked xylose residue to Glc-2, Xyl-3 is O6 linked to glucose-3 etc.) is coordinated by Trp^{607}, Trp^{646}, Trp^{651}, and Tyr^{685}. Xyl-3 forms hydrophobic interactions with Trp^{651}. Polar interactions are mediated between O2 and the endocyclic oxygen of the Xyl-2 residue, coordinated by the nitrogen moiety of Trp^{607} and the $N\zeta$ group of the Lys^{689} residue. (Figure 3.4) The observed ligand recognition between CBM65B and XXXG and can be extended to CBM65A with a high degree of certainty, albeit with observed differences in β1,4 backbone recognition (see Section 3.3.4.2).

3.3.4. Site Directed mutagenesis of CBM65A and CBM65B

3.3.4.1 CBM65 Mutant generation, expression and purification

Mutant variants of CBM65A (Master of Research degree) and CBM65B (D649A), were generated via single site directed mutagenesis as described in Materials and Methods, Section 2.1.17. The primers used to generate CBM65B variant, D649A are shown in Table 3.2. PCR amplifications, harbouring the single amino acid substitution were treated with Dpn1 in order to digest methylated-DNA, leaving the cDNA intact. Successfully amplified products were sent for sequencing (MWG, Germany) to check the integrity of the cDNA sequence, confirming the single base substitution to alanine was successful. CBM65 mutant variants were expressed and purified with identical methodology to that of the wild type proteins.

CBM65B variant	Sequence (5´→3´)	Direction
D649A	GCACGTTGGGATAAAGCCATTTGGGCGCAG	Forward
	CTGCGCCCAAATGGCTTTATCCCAACGTGC	Reverse

Table 3.2. Oligonucleotide primer sequences (5' -> 3') used to generate a single site substitution of Asp^{649} to alanine.

3.3.4.2. Understanding Ligand Recognition in the CBM65s

The apo-CBM65A crystal structure and the co-crystallised CBM65B-XXXG structure were used to inform which amino acid residues located in the ligand binding cleft of

the CBM65s may mediate ligand recognition and specificity. The predicated aromatic ligand binding residues were conserved between CBM65A and CBM65B. Thus, mutations which were predominantly performed on CBM65A during the Masters of Research project to explore ligand recognition can also be applied to CBM65B. A number of single site substitutions of amino acid residues (Trp^{55}. Trp^{60}, Trp^{99} and Trp^{108}) were previously generated by Ms S. Ana Luis (Lisbon, Portugal) in the ligand binding cleft of CBM65A

AGE (polysaccharides) and ITC (polysaccharides and oligosaccharides) were performed to assess the binding capacity of CBM65A and CBM65B mutant variants to an array of ligands (Table 3.3). Substituting Trp^{55}. Trp^{60}, Trp^{99} to alanine in CBM65A, equivalent to Trp^{602}, Trp^{607} and Trp^{646} in CBM65B, disrupted binding to barley β-glucan and cellohexaose but not to xyloglucan. Only mutating Trp^{108} to alanine (Trp^{651} in CBM65B) abrogated all ligand recognition. These data confirm the importance of these four aromatic amino acid residues to the recognition of the β1,3/β1,4 backbone of β-glucan and cellohexaose and by extension, cellulose. The observation that xyloglucan recognition is maintained in all mutants except CBM65A W108A suggests that xyloglucan recognition is primarily preserved through other specificity determinants unique to xyloglucan, such as interactions with the α-xylose substitutions of the polysaccharide.

CBM65A variant	Ligand	K_a (M^{-1})	ΔG (kcal $mole^{-1}$)	ΔH (kcal $mole^{-1}$)	TΔS (kcal $mole^{-1}$)	n
T58A	β-Glucan	8.3 (±1.0) x 10^3	-5.3	-12.9 ± 0.2	-7.6	1.0 ± 0.1
Q106A	β-Glucan	1.1 (±0.03) x 10^4	-5.5	-8.0 ± 0.2	-2.5	1.0 ± 0.0
Q110A	β-Glucan	3.9 (±0.4) x 10^3	-4.9	-10.0 ± 0.2	-5.1	1.0 ± 0.1
K146A	β-Glucan	5.3 (±0.2) x 10^3	-5.1	-9.9 ± 0.5	-4.8	1.0 ± 0.0
Y70A	β-Glucan	1.1 (±0.08) x 10^4	-5.3	-11.7 ± 0.8	-6.4	1.0 ± 0.0
Y114A	β-Glucan	1.1 (±0.02) x 10^4	-5.5	-12.8 ± 0.3	-7.3	1.0 ± 0.0
Y142A	β-Glucan	8.9 (±0.04) x 10^3	-5.4	-13.0 ± 0.6	-7.6	1.0 ± 0.0
W55A W60A W99A W108A	β-Glucan and cellohexaose	No binding				
T58A	Cellohexaose	2.5 (±0.4) x 10^3	-4.6	-11.4 ± 5.5	-6.8	1.01 ± 0.4
Q106A	Cellohexaose	No binding				
Q110A	Cellohexaose	Weak binding (unquantifiable)				
K146A	Cellohexaose	~9.0(±1.2) x 10^2	Binding too weak to quantify			
Y70A	Cellohexaose	2.9 (±0.3) x 10^3	-4.7	-11.8 ± 3.8	-7.1	1.0 ± 0.2
Y142A	Cellohexaose	1.6 (±0.007) x 10^3	-4.3	-14.7 ± 3.0	-10.4	1.01 ± 0.1
Q106A	Xyloglucan	3.5 (±0.5) x 10^4	-7.5	-26.7 ± 0.7	-19.2	1.0 ± 0.0
Q110A	Xyloglucan	1.3 (±0.2) x 10^5	-6.9	-16.6 ± 0.3	-9.7	1.0 ± 0.0
W55A	Xyloglucan	6.6 (±0.3) x 10^4	-6.5	-22.4 ± 0.2	-15.9	1.0 ± 0.0
W60A	Xyloglucan	3.1 (±0.06) x 10^3	-4.7	-22.8 ±0.3	-18.1	1.0 ± 0.0
W99A	Xyloglucan	5.0 (±0.9) x 10^3	-5.0	-22.7 ± 0.3	-17.7	1.0 ± 0.0
W108A	Xyloglucan	No binding				
CBM65B variant	**Ligand**	**K_a (M^{-1})**	**ΔG (kcal $mole^{-1}$)**	**ΔH (kcal $mole^{-1}$)**	**TΔS (kcal $mole^{-1}$)**	**n**
D649A	Cellohexaose	1.5 (±0.05) x 10^3	-4.3	-8.1 ± 0.2	-3.8	1.0 ± 0.2

Table 3.3. Affinity and thermodynamics of CBM65A and CBM65B variant binding to soluble polysaccharides and oligosaccharides

The recognition of XXXG by the two CBM65s is generally conserved, as described in Sections 3.3.3.3 and 3.3.3.4 respectively. The connecting loop between β4 and β5 of CBM65A (harbouring residues Trp^{99} to Gln^{106}), is longer than the equivalent region in

CBM65B. Inspection of an overlay of the two CBM65 protein structures indicates that Gln^{106} makes polar interactions with the O2 and O3 moieties of Glc-2 of the tetrasaccharide β1,4 backbone of XXXG. However, the equivalent residue Asp^{649}, of CBM65B, is beyond hydrogen bond forming range and is unlikely to make polar contact with ligand, suggesting differential modes of cellulose recognition between the highly conserved CBM65s. Mutants, substituting Gln^{106} (CBM65A) and Asp^{649} (CBM65B) for alanine displayed identical binding affinities to WT proteins for β-glucan and xyloglucan. The CBM65B mutant, D649A, did not abrogate cellohexaose recognition. However, the CBM65A mutant, Q106A, displayed no affinity for cellohexaose as confirmed by ITC. Gln^{106} is crucial for cellulose recognition in CBM65A.

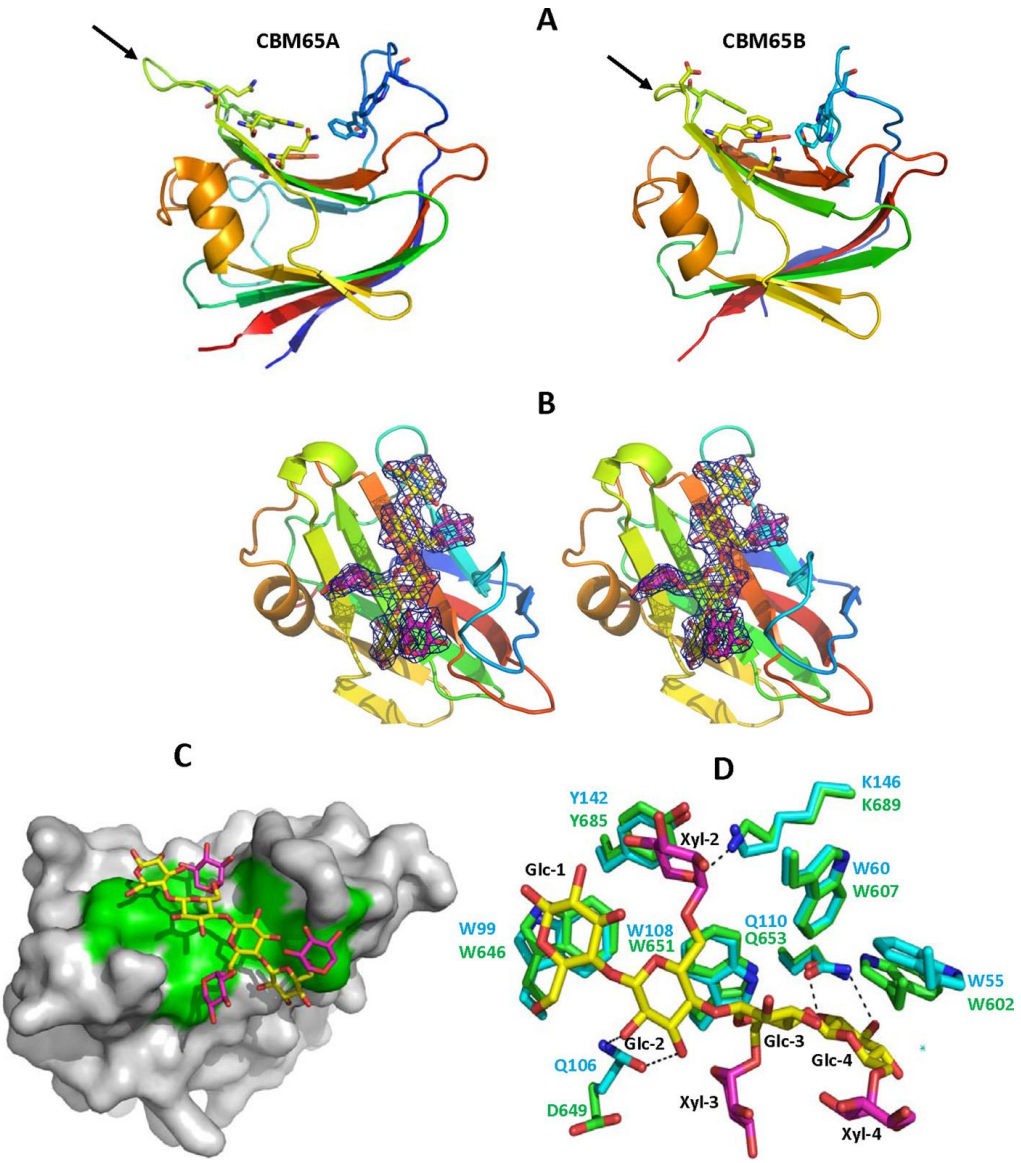


Figure 3.4. Structure of the CBM65s. A. CBM65A and CBM65B, continuously colour ramped, are shown in a schematic form. The loop containing residues Gln^{106} and Asp^{649} of CBM65A and CBM65B respectively, are highlighted by an arrow. Gln^{106} is seemingly crucial for cellohexaose recognition in CBM65A; Asp^{649} (CBM65B) is not key in cellulose recognition. **B.** CBM65B is depicted as a schematic (colour ramped as A), with ligand electron density ($2F_0 - F_c$) at 1.5 σ shown. XXXG is represented in stick format. Sugar residues are coloured according to element. Carbon moieties of glucose residues are coloured yellow, whilst the carbon moieties of the xylose decorations are yellow. **C.** Surface representation of the solvent accessible surface of CBM65B with XXXG in complex (coloured as in B). The aromatic residues key to ligand recognition are highlighted green. **D.** Overlay of the ligand binding sites of CBM65A and CBM65B showing conserved residues involved in ligand binding. The carbons of CBM65A amino acid residues are highlighted and labelled in cyan. The carbons of CBM65B amino acid residues are labelled and highlighted in green. Hydrogen bonds are represented by dashed lines. The diagrams were drawn in PyMOL. Figure taken from Luis et al., 2012.

3.4 Discussion

This chapter reports the apo structure of CBM65A and CBM65B in complex with XXXG. The biochemical and structural data described demonstrates how the differential recognition of structures within the same ligand is mediated in the CBM65 family. Recognition of β-glucans, as shown by the CBM65B-XXXG complex, is primarily mediated through hydrophobic interactions between the tryptophan residues housed in the ligand binding cleft and the glucopyranose rings of the ligand backbone. These extensive hydrophobic interactions contribute significantly to binding affinity.

The CBM65s displayed a significant preference for polysaccharides (xyloglucan and barley β-glucan) and decorated oligosaccharides (XXXG). The association constants generated by ITC indicate that both the CBM65s bind to cellotetraose with a lower affinity than cellohexaose The β1,4-backbone of cellotetraose however, fully occupies the ligand binding site. It is plausible that the increased interactions between the ligand binding site of the CBMs and the two additional glucose moieties of cellohexaose are responsible for the increase in observed affinity. However, the amino acids which mediate these extra ligand contacts are unknown. Potential candidates include Tyr^{70} and Tyr^{617} of CBM65A and CBM65B respectively, located at the entrance of the ligand binding cleft, and a glutamine (Gln^{67} in CBM65A and Gln^{659} in CBM65B). However, mutant variants of CBM65A where the aforementioned residues were substituted for alanine demonstrate similar affinities for the hexasaccharide to that of the WT protein, excluding their possible involvement in cellulose recognition. An alternative explanation for the seemingly low affinity displayed for cellotetraose could be due to mutarotation of the reducing end glucose, which would adopt multiple conformations and result in a loss of entropy. Ligands

with a D.P > 4 would instead be in a fixed conformation and thereby restrict mutarotation, reducing the associated entropic penalty. However, the fact that mixed linked glucans or β1,4-gluco polysaccharides bind with higher affinity than cellohexaose suggests that the conformational fixing of the reducing end glucose is not the sole reason for the observed decrease in affinity for cellotetraose. It has been observed that ligands which extend out of the ligand binding cleft of the CBM can bind with greater affinity than a ligand which occupies the binding site fully (Boraston et al, 2011) due to intra-chain hydrogen bonding formed between the glycan chains and may explain the displayed increased affinity the CBM65s share for longer, linear β-glucans. Thermodynamic parameters generated in this study must be viewed with a degree of caution. The *c* – value, product of protein concentration and the binding constant (K_a), provides an indication of reliability of the thermodynamic data generated by ITC. A *c* - value of < 1 is potentially indicative of unreliable data (Turnbull and Daranas, 2003). Many of the protein-ligand interactions investigated in this study have *c* - values < 1. However, for low affinity systems, such as protein-carbohydrate interactions it is not possible to achieve *c* – values in optimal ranges due to limited solubility of ligands. Nevertheless, reliable K_a and ΔH values can be generated in low-affinity systems through measures deployed in this study as described by Turnbull and Daranas (Turnbull and Daranas, 2003).

Mutating Gln^{106} in CBM65A completely abrogated cellohexaose binding. However the affinity of the mutant CBM Q106A for xyloglucan and the mixed linked, β-glucan was retained. This observation implies that CBM65A displays plasticity in ligand recognition, recognizing both linear β1,4 glucans and β1,3-1,4 mixed linked glucans independently and that Gln^{106} is only necessary for cellulose recognition. Interestingly, CBM65B lacks a functional equivalent to Gln^{106}. Mutating the perceived

functional equivalent in CBM65B, Asp^{649}, the only residue observed to form polar contacts with the glucose backbone of XXXG, did not impact upon cellohexaose recognition. Additionally, if the mechanism of CBM65B recognition of linear β1,4 glucans can be disrupted independently of β1,3-1,4 mixed linked glucan recognition is not known.

The data presented in the chapter shows how the CBM65s preferentially target xyloglucan. The increased affinity demonstrated by the CBM65s for decorated ligands can be attributed to both the apolar and polar interactions between the ligand binding site residues of CBM65B and the O6-linked xylose residues of XXXG (and by inference CBM65A). These interactions significantly increasing the affinity of the protein modules for the decorated ligands. Ligand binding is primarily achieved by hydrophobic contacts of the sugar rings with aromatic amino acids (McCartney *et al.*, 2006). Indeed, hydrophobic stacking of aromatic residues against the pyranose/furanose rings of sugar residues are common features of β-glucan binding CBMs, which is usually coordinated by three aromatic residues housed within the ligand binding cleft (Simpson *et al.*, 2000). However, five aromatic residues are located within the binding site of the CBM65s which make apolar interactions with the glucose backbone or xylose decorations of XXXG.

The importance of the aromatic residues of the ligand binding cleft to ligand binding is emphasised by the observation that mutating Trp^{108} in CBM65A, equivalent to Trp^{651} in CBM65B destroys ligand recognition, including the recognition of xyloglucan. This is consistent with the observation that the CBM65B-XXXG complex reveals that significant coordination of the bound ligand is mediated by Trp^{651}. The aromatic residues interacts with the Glc-2 and Glc-3 residues and the Xyl-2 and Xyl-3 decorations of XXXG. Trp^{60} and Trp^{99} alanine substitutions in CBM65A (Trp^{607} and

Trp^{646} in CBM65B) also result in reductions in binding affinity, highlighting the importance of these of aromatic residues in recognising the glucan backbone and xylose decorations of XXXG. Conversely, mutating Trp^{55} (Trp^{602} in CBM65B) to alanine did not diminish CBM65A affinity for xyloglucan, most likely due to the lack of interactions with xylose-4 of XXXG. These data help to provide a model for the recognition of decorated polysaccharides.

3.5 Future Work

CBM65A and CBM65B, new members of the CBM family 65 share limited sequence similarities to two endoglucanases encoded by *Cellulosilyticum ruminicola* and *Clostridium lentocellum*. The endoglucanase of *C. ruminicola*, displays two tandem repeat sequences with homology to *Ec*Cel5a. Significantly, three of the four tryptophan residues which are critical for β-glucan recognition in CBM65A and CBM65B are conserved in the related protein modules the *C. ruminicola* endoglucanase (Figure 3.5). One can hypothesise that these conserved residues likely play a similar role and are key in mediating β-glucan recognition. A better understanding of how this new family of CBMs recognise the decorations of substituted polysaccharides in lieu of conserved structural elements may enable the engineering of CBMs with multiple specificities in tandem with higher affinities which may potentiate the activity of enzymes of industrial significance.

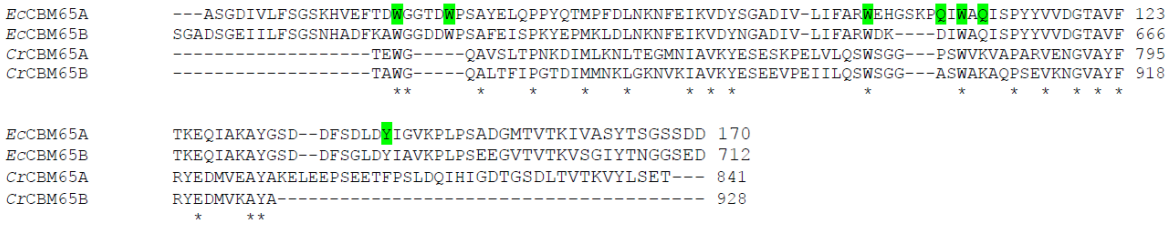


Figure 3.5. Amino acid sequence alignments of CBM family 65 proteins. The amino acid sequences labelled *Ec*CBM65A(/B) are from the *Eubacterium cellulovorans* (*Ec*) GH5 endoglucanase. The protein sequences labelled *Cr*CBM65A(/B) are from the *Cellulosilyticum ruminicola* (*Cr)* GH5 endoglucanase. Key residues necessary for ligand recognition are coloured green. Fully conserved residues are denoted by an *. Figure adapted from Luis et la., 2012.

Chapter 4: α-mannan degradation and utilization by *Bacteroides thetaiotaomicron.*

4.1. Introduction

The human gut microbial community, termed the microbiota, plays a pivotal role in human health and nutrition (Arpaia *et al.*, 2013). Key glycan degraders such as *Bacteroides thetaiomicron* (*B. theta*), are prevalent in the microbiota of 'healthy' humans. Complex glycan utilization is achieved through a protein/enzyme ensemble up-regulated to orchestrate the deconstruction of a specific glycan. These glycan degrading systems are encoded by polysaccharide utilization loci (PULs). Each PUL encodes a system that degrades a specific glycan (Bolam and Koropatkin, 2012)

Yeast α-mannan is an important component of the cell wall of yeasts such as *Saccharomyces cerevisiae* (*S. cerevisiae*) A ubiquitous feature of the glycan is an α1,6-mannose backbone that comprises the outer chain of the polysaccharide. The mannose containing backbone is decorated with a range of side chains that can vary significantly amongst yeast species. The α1,6-mannan backbone of *S. cerevisiae* is decorated with α1,2-linked mannose side chains and is capped with α1,3-linked mannose residues. Phosphate bridges provide additional complexity to this substrate, linking an α1,3-mannobiose moiety to mannose residues appended directly to the α1,6-mannan backbone of the glycan. The side chains decorating the α-mannan backbone of *Schizosaccharomyces pombe* (*S. pombe*) are less extensive, and consists of single α1,2-linked galactose residues and single α1,2-linked mannose residues appended directly to the backbone (Ziegler *et al.*, 1994).

The core region of *S. cerevisiae* α-mannan, from which the outer chain extends, is nearly identical in composition and structure to that of mammalian high mannose N-

glycans (HMNG), which decorate mammalian cells. HMNG, consist of two β1,4-linked N-acetyl glucosamine residues, to which a β1,4-linked mannose residue is appended. Extending from the central mannose residue are two branches which consists of mannose residues linked by various α-linkages. The structures of yeast α-mannans and HMNG are shown in Figure 4.1.

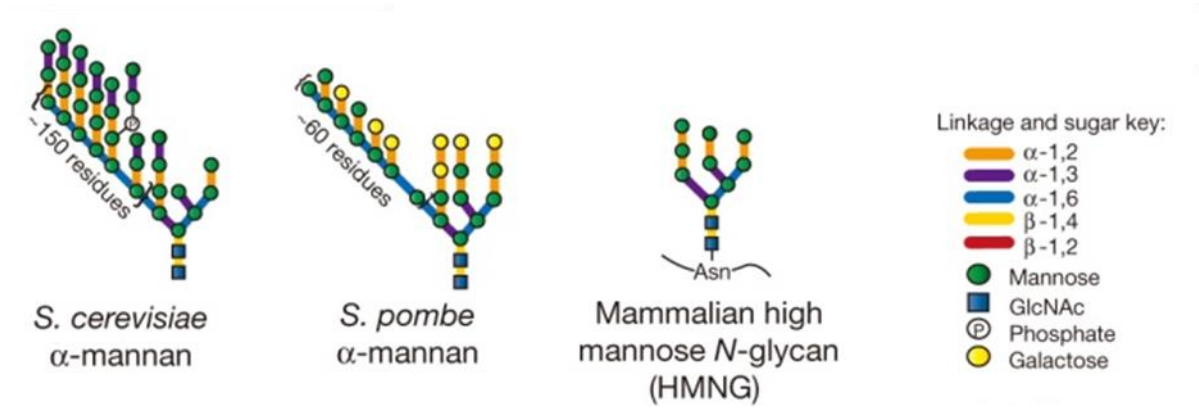


Figure 4.1. Structure of yeast α-mannans and mammalian high mannose N-glycans. A cartoon representation of the structure of the yeast α-mannan of *Saccharomyces cerevisiae* (*S. cerevisiae*) and Schizosaccharomyces pombe *(S. pombe)* α-mannan and mammalian high mannose N-glycans (HMNG). The saccharide and linkage composition of each polysaccharide is denoted by a key. Asn, to which HMNGs are covalently bound, represents asparagine. (Figure adapted from Cuskin et al., 2015).

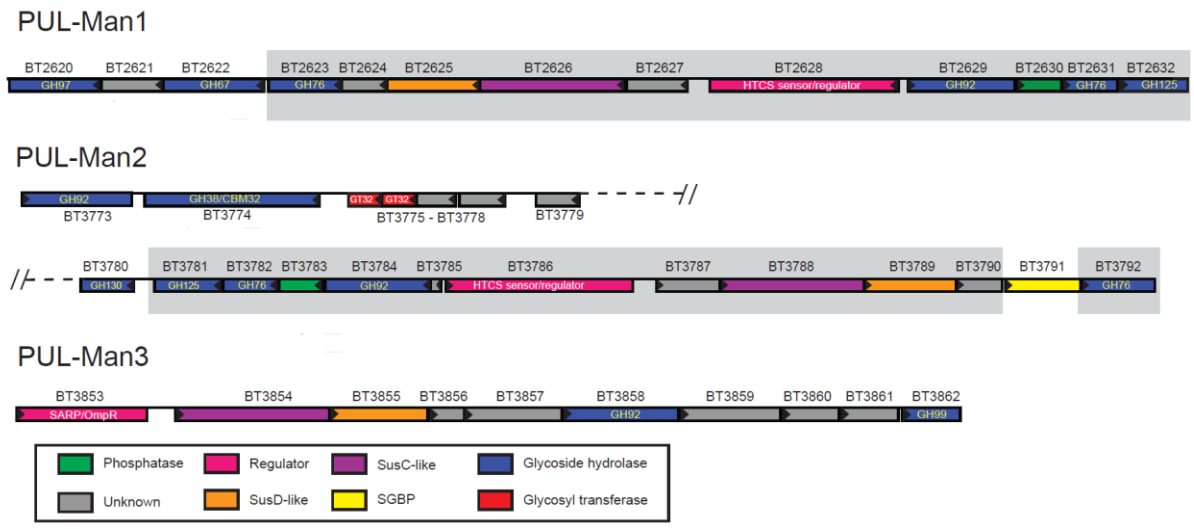
The genome of *B. theta* encodes 36 glycoside hydrolases predicted to target α-mannosidic linkages due to sequence similarities to previously characterised glycoside hydrolase families (GHs) known to possess α-mannosidase (GH38, GH47, GH92 and GH125) or α-mannanase (GH76) activities (Xu *et al.*, 2007). Three PULs, designated PUL-Man1, PUL-Man2 and PUL-Man3 (Figure 4.2) (Cuskin *et al.*, 2015b) were shown to be upregulated in *B. theta* when the bacterium was cultured with *S. cerevisiae* α-mannan as a sole carbon source. Intriguingly, PUL-Man1 and Pul-Man2 display a significant degree of synteny in the 'core' region of these loci PUL, encoding

proteins with seemingly identical predicted activities. Thus, both PUL-Man1 and PUL-Man2 encode two predicted GH76 endo-α1,6-mannanases, a GH125 α1,6-mannosidase and GH92 α-mannosidase, a phosphatase, as well as the importation machinery of the PUL (SusD-like and SusC-like). The syntenic region of the two PULs also contain proteins of unknown function that flank the SusC/D pairs. Interestingly, the open reading frames (ORFs) encoded by the peripheries of PUL-Man1 and PUL-Man2 have seemingly divergent activities. Man-PUL1 encodes for a predicted GH97 α-galactosidase and so plays a potential role in *S. pombe* α-mannan degradation. Likewise, PUL-Man2 encodes for a GH38 α-mannosidase, BT3774, an enzyme which was shown by Dr Cuskin to display broad specificity. Indeed Cuskin showed that BT3774 plays a pivotal role in the periplasmic side chain removal of S. cerevisiae yeast α-mannan; it is the only enzyme capable of efficiently removing the sterically restricted α1,2-linked mannose residues appended directly to the backbone of yeast α-mannan (Cuskin *et al.*, 2015b). PUL-Man3 shares no discernible similarities in PUL organisation to PUL-Man1 and Pul-Man2, but encodes for a GH99 endo-α1,2-mannosidase, which has known involvement in S. cerevisiae yeast α-mannan side-chain removal (Hakki *et al.*, 2015)

Surprisingly, despite the significant similarities in the structure and composition shared between the core region of yeast α-mannan and HMNGs, HMNGs up-regulate a single PUL, (PUL-HMNG), which is distinct from the mannan loci. PUL-HMNG mediates HMNG deconstruction (Cuskin *et al.*, 2015b) and encodes four enzymes, three GH92 α-mannosidase and a GH18 endo-β-N-acetylglucosaminidase, and two surface glycan binding proteins that target the glycan (Figure 4.2)

Transcriptomic data showed that yeast α-mannans are not surrogates for HMNG (Cuskin *et al.*, 2015b), but are nutrient sources for *B. theta*. Given the limited number of different α-mannosyl linkages in α-mannans and HMNGs, the biological rationale for such an extensive array of *B. theta* enzymes that target these bonds is unclear. It is possible that the branched structure of glycans containing these α-mannosyl linkages predicates a complex hierarchical degradative process, in which the various α-mannosidases display specificity for structures downstream of the target bond. This may explain the requirement for such a large number of enzymes and PULs which target similar but distinct substrates. It is likely that the HMNG-specific PUL encodes for an enzyme profile tailored to HMNG degradation, but not yeast α-mannan, deconstruction.

A



B

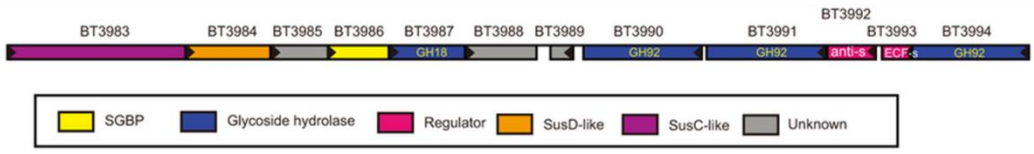


Figure 4.2. The polysaccharide utilization loci encoded by *Bacteroides thetaiotaomicron* which are up-regulated by, and target, α-mannan. A schematic representation of the polysaccharide utilization loci encoded by the genome of *B. theta* which target (**A**) yeast α-mannans (PUL-Man1, PUL-Man2 and PUL-Man3) and (**B**) high mannose N-glycans (PUL-HMNG). The genes encoding known or predicted function are colour coded and are annotated according to CAZy family where applicable. Areas of PUL-Man1 and PUL-Man2 which display a high degree of synteny are highlighted with a grey background. SGPB represents surface glycan binding protein. (Figure adapted from Cuskin et al., 2015).

4.2 Objectives

In this chapter *B. theta* is used as a model for dissecting the mechanism by which a key member of the microbiota metabolizes complex carbohydrates containing α-mannosyl linkages, achieved primarily through the characterization of enzymes encoded by PUL-Man1/2/3 and PUL-HMNG, providing a biochemical rationale for the distinct yeast α-mannan and HMNG utilization apparatus. The biochemical data generated in this chapter, in conjunction with growth studies and pre-exisitng genetic/*in vivo* data provides a model for the utilization of glycans that contain α-mannosidic linkages.

4.3 Results

4.3.1. Elucidating the α-mannosidases encoded by *Bacteroides thetaiotaomicron* capable of de-branching yeast α-mannan

The enzymatic degradation of α-mannan is greatly restricted through steric constraints imposed through mannose side-chains appended to the α1,6-mannan backbone. Previous work performed by Dr. Fiona Cuskin found that the four GH76 endo-α1,6-mannanases encoded by PUL-Man1 and PUL-Man2 were unable to cleave the mannan backbone until undecorated regions are exposed. As stated above Dr Cuskin showed that the GH38 α-mannosidase BT3774, encoded by PUL-Man2, removes the most challenging, sterically restricted, mannosidic bond in which the mannose at the base of the mannan side chains is linked α1,2 to the mannan backbone. However, the three mannan PULs do not encode a candidate surface α-mannosidase. Such an enzyme must exist otherwise the surface endo-α1,6-mannanases would not be able to cleave the mannan backbone. The surface deconstruction of yeast α-mannan is therefore poorly understood. The capacity of 11 *B. theta* α-mannosidases (Zhu *et al.*, 2010), predicted to be localised to the cell surface and belonging to GHs (ten GH92s

members and one GH38 member) with known α-mannosidase activity, to remove mannose side-chains was assessed. The localisation of the α-mannosidases explored in this section were predicted due to the presence of a type II single peptide, assessed via LipoP (http://www.cbs.dtu.dk/services/LipoP/). However, some α-mannosidases were assayed, despite being predicted to harbour a type I signal peptide, owing to the presence of a canonical cysteine residue encoded early in the N-terminal protein sequence (< 20 residues from start codon); a potential indicator of a signal II peptide. An example signal peptide analysis is shown in Figure 4.3. The remaining analysis are displayed in Appendix A.

MVTRLFFLSICFPFLLTSCQQSKKTVEFVDYVNPLMGTESTFAFSHGNTY
PAVAVPWGMNFWSPQTGENGSGWMYTYTDSLMRGFRQTHQPSPWINDYGT
FSIMPLAGELKMSHKERLVPFSHQQEKATPYNYSVTFNNGLQTSLSATSR
GAVFEVSFPEKEDQYVVVDAYNGGSSITIEPEKRLVKGATRYNNGGVPDN
FANYFMMEFSHPVIEYGTYNGDTLLHHQTDVAADYTCAYLKFDVPAGEKL
TIRTASSFISPEQAAINFNREVADADVQLISGKAREQWNNYLGRVEAEGG
TDEQLRTFYSCLYRTLLFPREFYEFDSQGNPVYYSPYDGNVHDGYMYTDN
GFWDTFRAVHPLFTLLYPEVSERVTQSIINAYNESGFMPEWASPGHRGCM
IGNNSVSLLVDAWMKGIQTVDAEKALEAMIHQTQARHAEIASVGRDGFEY
YDKLGYVPYPEVPEATAKTLEYAYADWCIARFAESLGKQDIADQYYQKAP
NYRNLYYPEHGFMWTKDAKGNWRDRFDATEWGGPFTEGSSWHWTWSVFHD
PEGLSELMGGHEPMIARLDSMFVAPNTYNYGTYGFVIHEIAEMVALNMGQ
YAHGNQPVQHAIYLYDYIGQPWKTQYHLRNVMDKLYNSGSKGYCGDEDNG
QTSAWYVFSAMGFYPVCPGMPEYAIGSPLFKKVTLHLPEGKNFVVSAADN
AADRPYIRKALLNGQEFTRNYLTHDELKQGGELNLSMDSVPNQQRGTQPA
DFPYSYSK

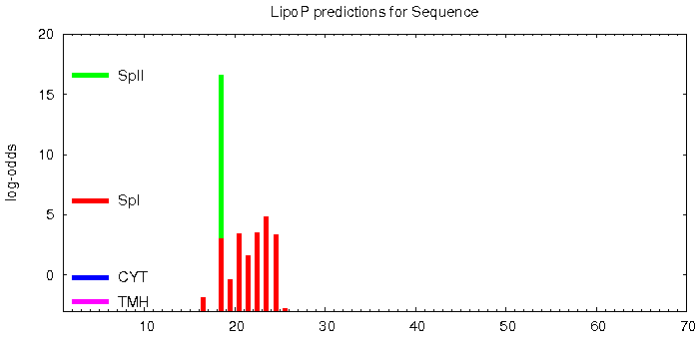


Figure 4.3. LipoP signal peptide prediction of BT2199. The left hand panel shows the full, wild-type amino acid sequence of BT2199, a GH92 α-mannosidase. The right hand panel shows a LipoP signal peptide analysis of the BT2199 protein sequence. BT2199 is predicted to have a signal II (SpII) peptide, indicating the protein is localised to the cell surface. The point of signal peptide cleavage is predicted to be between residues 18 and 19, the position of which is denoted by red and green colouring of the amino acid residues at the site of cleavage, S18 and C19, respectively.

4.3.1.1 Expression and Purification

The constructs (generated previously by Dr Zhu;(Zhu *et al.*, 2010)) encoding the enzymes described above ere expressed *in E. coli* TUNER cells and purified by IMAC as described in 2.1.25.1 , generating pure recombinant protein, confirmed via SDS-PAGE. An example of an SDS-PAGE analysis of a typical IMAC purification is shown in Figure 4.4.

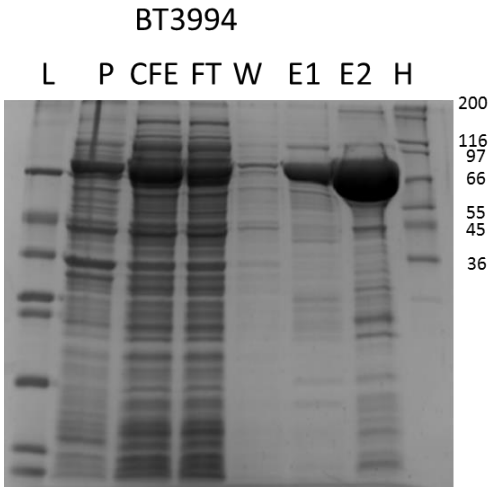


Figure 4.4. Example SDS-PAGE analysis of an IMAC purification. All SDS-PAGE gels were 12.5 % (w/v). L = low size marker H= high size marker (kDa); P = insoluble fraction; CHE = cell free extract; FT = flow through; W = wash (1x TALON buffer); E1 = elution 1 (5 mM imidazole,1 x TALON buffer); E2 = elution 2 (100 mM imidazole, 1 x TALON buffer). A. BT3994 ran with an apparent M_r of ~80 kDa, consistent with the predicted sizes of the recombinant proteins

4.3.1.2 The α-mannosidases which may potentiate side-chain removal.

The capacity of each α-mannosidase to de-branch yeast α-mannan was performed in the presence of characterised enzymes believed to facilitate side chain removal and included the GH99 endo-α1,2-mannosidase, BT3862 (Hakki *et al.*, 2015) and the sugar-6-monophosphatase, BT2630. Also incorporated into assays was the GH76 endo-α1,6-mannanase, BT2631 (see below) which can act upon the α1,6-mannan backbone exposed following side-chain removal, generating α1,6-linked

mannooligosaccharides visualised by TLC analysis. Seven out of the ten α-mannosidases were active against intact yeast α-mannan, liberating mannose (Figure 4.5). Of the exo-acting α-mannosidases tested only BT2199 and BT3990, not encoded by PUL-Man1-3 of *B. theta*, were capable of removing the side chains of yeast mannan, enabling the GH76 endo-mannanase to attack the α1,6-mannose backbone and evidenced by the production of oligosaccharides (Figures 4.6). BT3990 is encoded by HMNG-PUL whilst BT2199 appears to be encoded by a PUL seemingly distinct from the yeast α-mannan degrading apparatus of *B. theta*. Subsequent cell localization studies by Lowe and Cuskin showed that BT3990 was located in the periplasm. Thus, BT2199 is the candidate surface enzyme which generates limited undecorated regions of the yeast mannan backbone, which could thus be cleaved by the outer membrane GH76 endo-mannanases.

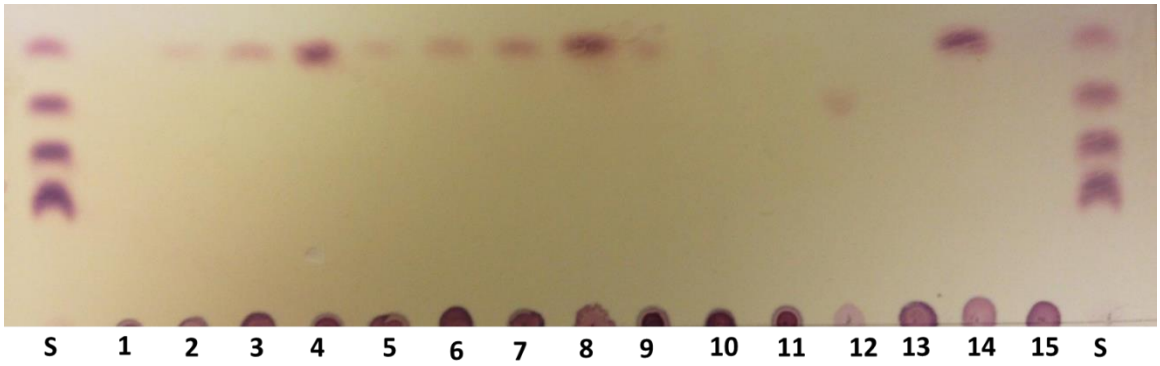


Figure 4.5. Thin Layer Chromatography (TLC) of predicted extracellular α-mannosidases activity against intact S. cerevisiae α-mannan. TLC analysis of *S. cerevisiae* yeast α-mannan (1 mg/ml (w/v)) incubated individually with 50 µM of each respective GH92 α-mannosidase, and the GH38 α-mannosidase, BT4072. Assays were performed in 20 mM Tris/HCl, 300 mM NaCl, pH 8.0 at 37 ^{o}C for 1 h. Digests included 1 µM GH99 endo-α1,2-mannanase, 1 µM sugar-6-monophosphatase (BT2630) to determine the activity of the GH92/GH38 against yeast α-mannan sans GH76 endo-α1,6-mannanase activity. Samples were run against α-mannooligosaccharide standards, D.P 1 to 4, denoted by S. Lane 1 - No enzyme, yeast α-mannan; Lane 2 - BT1769 (GH92);Lane 3 – BT1878 (GH92); Lane 4 – BT2199 (GH92); Lane 5 - BT3130 (GH92); Lane 6 - BT3773 (GH92); Lane 7 - BT3858 (GH92); Lane 8 - BT3990 (GH92); Lane 9 - BT3991 (GH92); Lane 10 - BT3994 (GH92); Lane 11 -BT4073 (GH92); Lane 12 - BT3862 (GH99); Lane 13 -BT2630 (phosphatase); Lane 14 – BT3774 (GH38); Lane 15 – BT4072 (GH38).

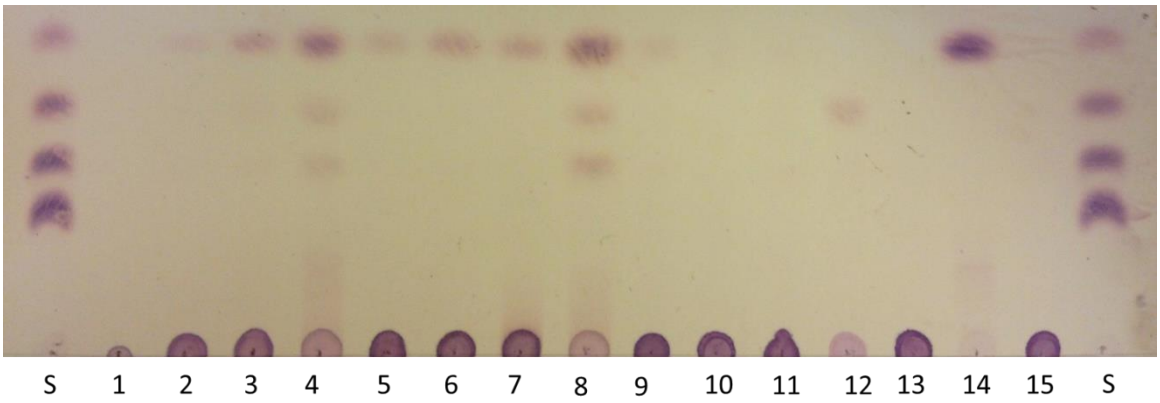


Figure 4.6. The capacity of endo-mannanases to act on mannan treated with α-mannosidases analysed by TLC. The experiment was carried out as described in Figure 4.3 except that the reactions were supplemented with 1 µM GH99 endo-α1,2-mannanase, 1 µM sugar-6-monophosphatase (BT2630) and 1 µM GH76 endo-α1,6-mannanase(BT2631) to determine de-branching activity of the GH92/GH38. The lanes are labelled as in Figure 4.3.

4.3.1.3 Comparing extracellular side-chain removal with periplasmic side-chain removal of yeast α-mannan.

The GH92 α-mannosidase, BT2199, was deemed to be the most likely putative candidate enzyme to facilitate the extracellular de-branching of yeast α-mannan, owing to the enzyme's predicted localisation to the cell surface. Due to the periplasmically located GH38 α-mannosidase BT3774's key role in side-chain removal and deconstruction of yeast α-mannan, assays were performed to determine the relative differences in the rate of surface and periplasmic side-chain depolymerisation, providing potential biological context.

The ability of the α-mannosidases to de-branch yeast α-mannan at various concentrations was assessed (Figure 4.7). The GH76 endo-α1,6-mannanase, BT3792 was present in the assays to determine debranching activity of the α-mannosidases. The relative rate of debranching was deduced via TLC analysis. Both enzymes were able to facilitate GH76 endo-mannanase attack, observed by the presence of α1,6-

manno-oligosaccharides and confirming the ability of these enzymes to de-branch yeast α-mannan. The GH38, BT3374, was significantly more active, enabling backbone access at concentrations as low as 0.2 µM. BT2199 was only able to adequately facilitate endo-mannanase activity at concentrations of 5 µM, a ~25-fold lower activity than BT3774 in its ability to expose the backbone of yeast mannan (Figure 4.8). The GH99 endo-α1,2-mannosidase (BT3862), and the sugar-6-phosphtase (BT2630), were not incorporated in these assays, demonstrating that the respective activity of these enzymes is not required for side chain removal. Surface side chain removal appears to be significantly limited relative to periplasmic depolymerisation and is consistent with the low quantities of mannose released into the culture supernatant when *B. theta* is cultured with yeast mannan as a sole carbon source (Cuskin *et al.*, 2015b).

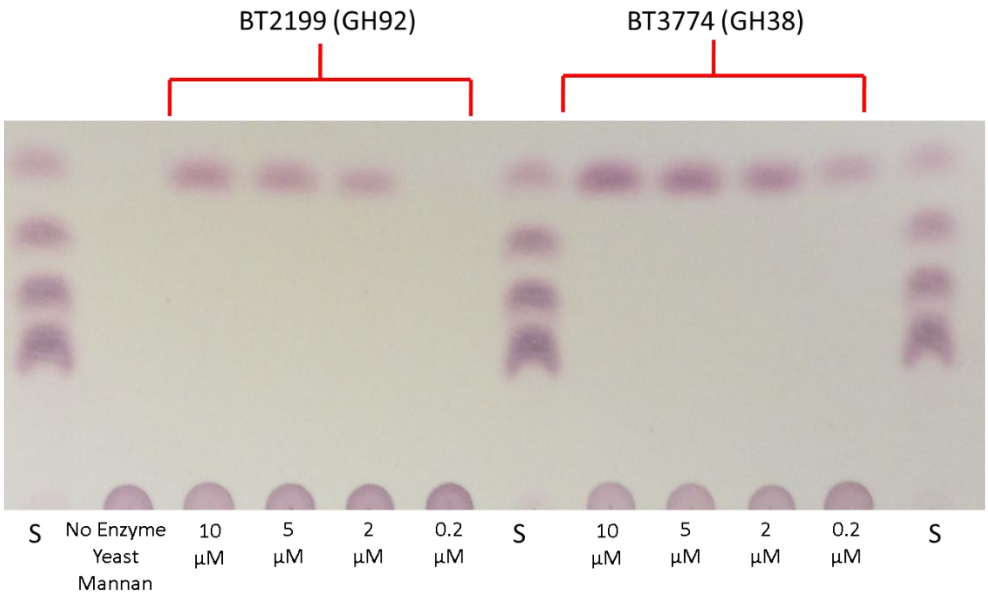


Figure 4.7. Thin Layer Chromatography comparing BT2199, GH92 α-mannosidase activity with the dominant periplasmic debranching enzyme, BT3774 GH38 α-mannosidase, against yeast α-mannan. TLC analysis comparing the putative extracellularly located GH92 (BT2199) with the periplasmically located GH38 (BT3774). The relative activity of the enzymes against *S. cerevisiae* yeast α-mannan at various concentrations was assayed by incubating enzyme with the polysaccharide (1 mg/ml (w/v)). The concentration of the enzyme being assayed is displayed below the respective lane. Assays were performed in 20 mM Tris/HCl, 300 mM NaCl, pH 8.0 at 37 ^{o}C for 1 h. Samples were run against α-mannooligosaccharide standards, D.P 1 to 4, denoted by S.

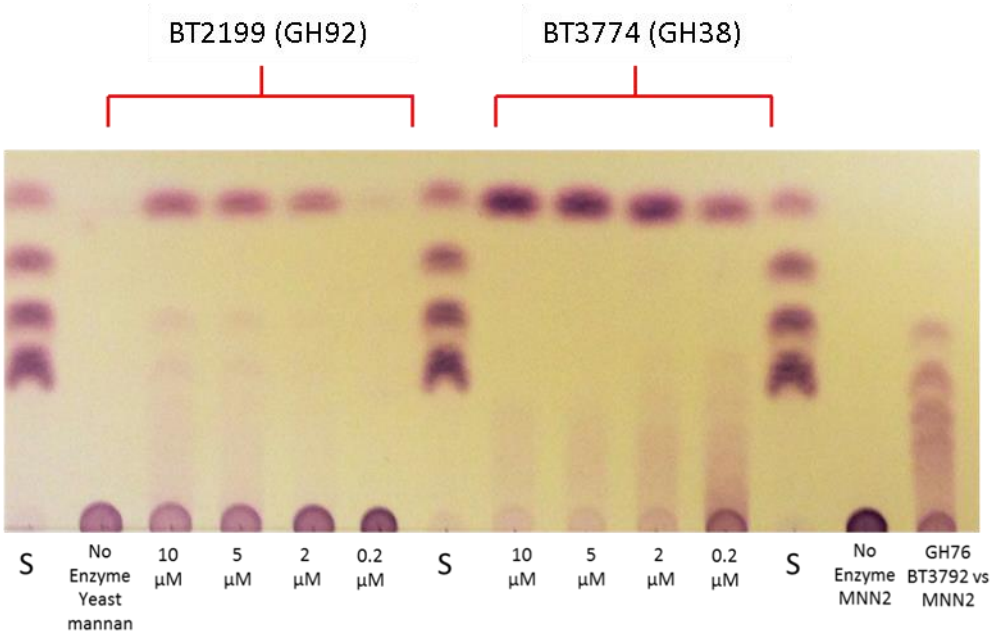


Figure 4.8. Thin Layer Chromatography comparing BT2199, GH92 α-mannosidase yeast α-mannan debranching activity with the dominant periplasmic debranching enzyme, BT3774 GH38 α-mannosidase. A. TLC analysis comparing the putative extracellularly located GH92 (BT2199) with the periplasmically located GH38 (BT3774). The ability of the enzymes to debranch *S. cerevisiae* yeast mannan at various concentrations when incubated with the polysaccharide (1 mg/ml (w/v)) was assayed in the presence of the extracellular GH76 endo-α1,6-mannanase, BT3792 (1 µM). The concentration of the enzyme being assayed is displayed below the respective lane. Assays were performed in 20 mM Tris/HCl, 300 mM NaCl, pH 8.0 at 37 ^{o}C for 1 h. Samples were run against α-mannooligosaccharide standards, D.P 1 to 4, denoted by S. The GH76 endo-α1,6-mannanase, BT3792 (1 µM) was incubated with the mutant mannan, mnn2 (1 mg/mL (w/v)) lacks mannose side-chains) as a positive control of GH76 activity.

4.3.2. Characterisation of the enzymes which degrade the α1,6-mannan backbone of yeast α-mannan

4.3.2.1.1The cellular location of the four GH76 endo-α1,6-mannanases (mannanases) encoded by the yeast mannan PULs of *B. theta*

The mannan PULs (PUL-Man) of *B. theta* encode four GH76 enzymes; two derived from both PUL-Man1 (BT2623 and BT2631) and PUL-Man2 (BT3782 and BT3792. Sequence analysis of the four enzymes by LipoP predicts that BT2623 and BT3782 have a type I signal peptide and so are located to the periplasmic space, with BT2623 and BT2631 predicted to contain a type II signal peptide, suggesting that these membrane-bound enzymes are on the surface of the bacterium. Localisation experiments utilizing fluorescently labelled antibodies performed by Drs Cuskin and Lowe, confirmed that BT3792 was indeed present on the cell surface and therefore corroborates these predictions (Cuskin et al.2015).

4.3.2.1.2. Expression and purification

The four GH76 enzymes were previously cloned into pET-based expression vectors, which supplied a His_6 tag, by Dr Cuskin. The four proteins were purified here using IMAC and their purity analysed by SDS-PAGE, Figure 4.9. BT2623, BT3782 and BT3792 were cloned by Dr. Fiona Cuskin and BT2631 was cloned by Dr. Joanna Norman. The four GH76 mannanases, like all *B. theta* proteins investigated in this study, were cloned without signal peptides. The genes encoding BT3782 and BT3791 were cloned into pET28a with N-terminal truncations of 154 and 74 amino acid residues, respectively, and harboured an N-terminal His_6-tag. To aid in solubility, DNA encoding mature BT2623 and BT2631 were cloned into pET32b and pET43b, which supplies a thioredoxin and NUS tag, respectively, and a C-terminal His_6-tag to the recombinant proteins. BT3792, BT2623 and BT2631 were expressed in *E. coli* BL21

cells and were produced is soluble form, as seen upon SDS-PAGE analysis. BT3782 was expressed in *E. coli* TUNER cells to maximise soluble protein production.

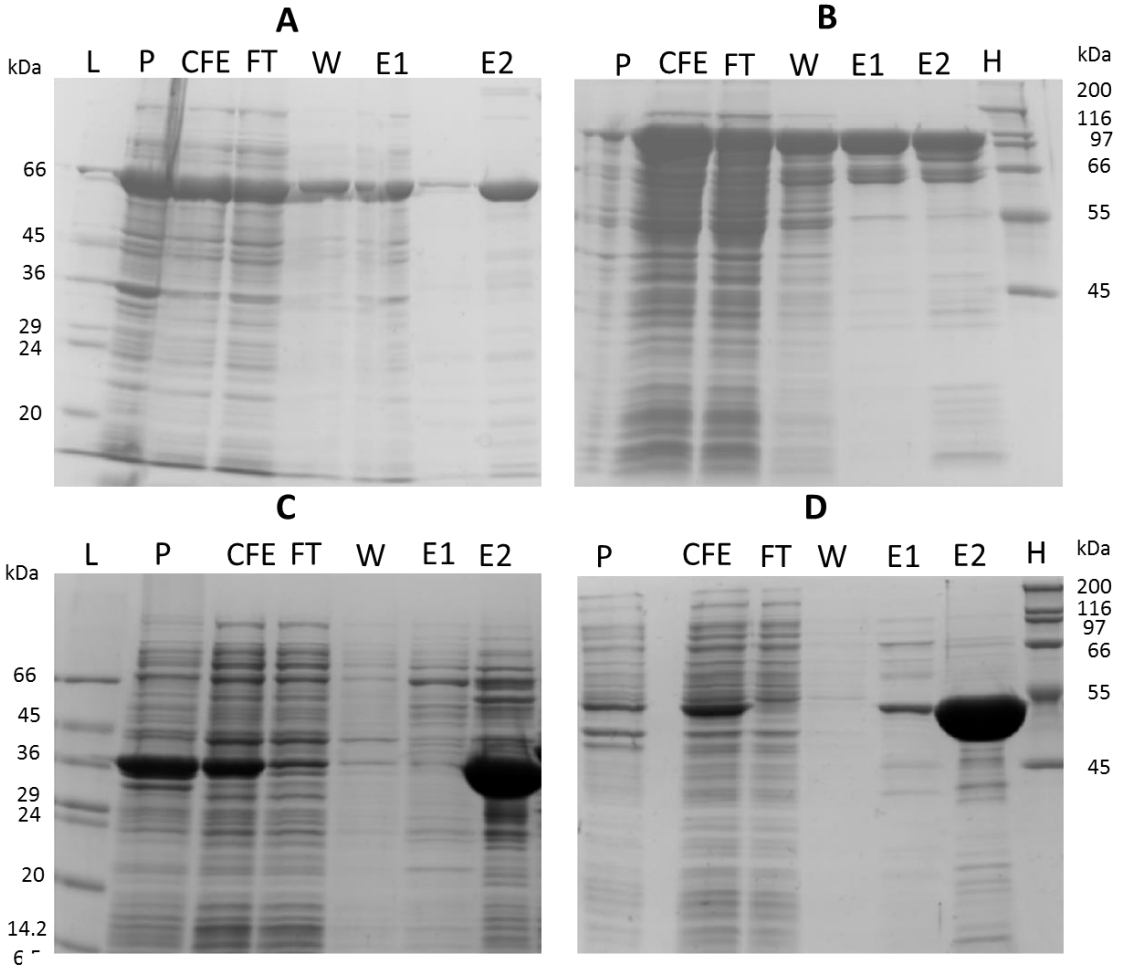


Figure 4.9. SDS- PAGE analysis of the protein purification of the four GH76s encoded by yeast mannan PULs one and two. SDS-PAGE data of BT2623 (A), BT2631 (B), BT3782 (C) and BT3792 (D) following IMAC. All SDS-PAGE gels were 12 % (w/v). L= low molecular weight marker; P = insoluble fraction; CFE = cell free extract; FT= flow through; W = wash; E1 = elution 1, 5 mM imidazole; E2 = elution 2, 100 mM imidazole. **A**. BT2623 ran with an apparent M_r of 66 kDa. **B.** BT2631 ran with an apparent M_r of 116 kDa. **C.** BT3782 ran with an apparent M_r of 40 kDa. **D**. BT3792 ran with an apparent M_r of 48 kDa. These values are consistent with the predicted size of the proteins.

4.3.2.1.3. The four GH76 are endo-α1,6-mannanases with activity against mnn2 mutant yeast mannan

As demonstrated previously by Dr Cuskin, the four recombinant GH76 endo-mannanases encoded by PUL-Man1 and PUL-Man2 were active against the undecorated α1,6-mannan backbone derived from the mnn2 mutant of *S. cerevisiae*. In this PhD project TLC analysis revealed that the surface located mannanases, BT2623 and BT3792, produced larger oligosaccharide products (D.P ≥ 4) than their periplasmic counterparts, BT2631 and BT3782 (D.P ≤3) Kinetic data showed that the periplasmic GH76 enzymes were ~3 to 10 times more active than their surface counterparts (Table 4.1)

	K_M (mg ml^{-1})	k_{cat} (s^{-1})	k_{cat}/K_m (s^{-1} mg^{-1} ml)
BT2623	1.892 ± 0.399	297 ± 19.07	157
BT2631	nd	nd	743 ± 8.048
BT3782	0.4315 ± 0.333	705 ± 219	1633
BT3792	2.425 ± 1.136	1263 ± 262	521

Table 4.1. Kinetic parameters of GH76 endo-α1,6-mannanase activity against the mnn2. Generated by Dr. Cuskin.

4.3.2.1.4. Activity of the GH76 mannanases versus α1,6-mannooligosaccharides

In order to elucidate the activity of the GH76 enzymes against oligosaccharides, high D.P (>4) and low D.P (<4) α1,6-mannooligosaccharides were generated by treatment of mnn2 yeast mannan with BT3792 and BT2631, respectively. Enzyme (10 uM) was incubated with the polysaccharide (10 mg/mL (w/v) for 2 h at 37 ^{o}C , in the presence of 20 mM Na-HEPES, pH 7.5. Enzymatic digests were subsequently pooled, lyophilized and subjected to size exclusion chromatography using P2-gel columns

(BioRad) as described in Materials and Methods, Section 2.2.6.2, rendering pure oligosaccharides of single species. The integrity of purified oligosaccharides was confirmed by HPAEC-PAD and MALDI-TOF mass spectrometry (Figure 4.10). Full methodology of the preparation of the oligosaccharides is detailed in Materials and Methods, Section 2.2.6.3.

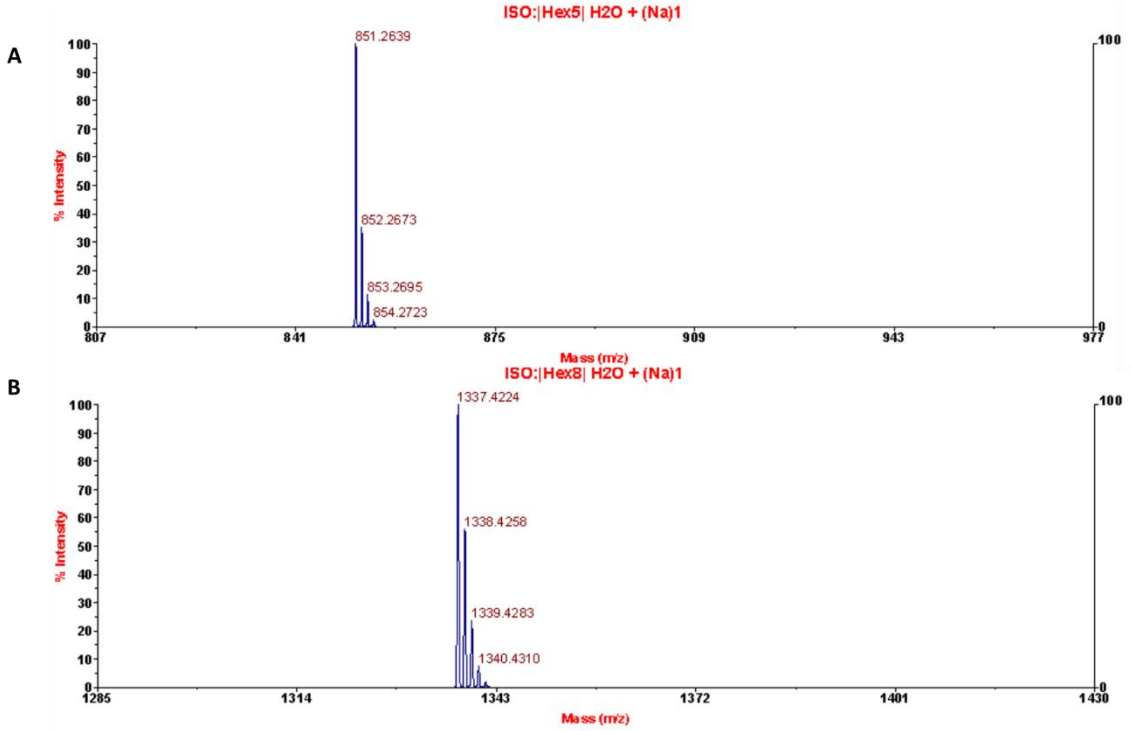


Figure 4.10. Mass Spectrometry result of purified α1,6-mannooligosaccharides. Details of MALDI-TOF spectrometry (mass spec) used to verify purified α1,6-mannooligosaccharides are described in Materials and Methods, Section 2.2.6.3. (A) mass spec result of α1,6-mannopentose (M5) and (B) α1,6-mannooctose (M8). A mass/charge ratio (m/z) of 851 was given for M5, corresponding to the mass of M5 (828) plus a sodium adduct (23). The m/z of M8 was observed at 1337, corresponding to the mass of M8 (1314) plus Sodium (23).

The catalytic activity of the four GH76 enzymes against α1,6-mannooligosaccharides was assessed by incubating each enzyme with an α-1,6-mannooligosaccharide with a D.P. of 2-8 (defined as M2-M8, respectively). Enzyme concentration was adjusted to provide ≥ 80 % degradation within the maximum time point and aliquots of the reaction were taken and the enzyme inactivated by boiling. Substrate depletion was observed by HPAEC-PAD (Figure 4.11) and plotted using GraphPad Prism 5.0. The gradient (velocity (min)) was ascertained via linear regression analysis. The equation utilized is shown below;

Velocity (min) = $In[S_{t0}] / In[S_{tx}]$*

K_{cat}/K_m (M^{-1} min^{-1}) = Velocity (min) / Enzyme concentration (M)

*In represents natural log. peak area of oligosaccharide when time = 0 min. tx is peak area of oligosaccharide at time point

The data are displayed in Table 4.2 The four GH76 enzymes displayed comparable rates against longer oligosaccharides (D.P 6-8), degrading α-1,6-mannooctose (D.P 8), with an efficiency of ~1 x 10^6 M^{-1} min^{-1}. The activities of the surface endo-mannanases, BT2623 and BT3792, demonstrated a significant decline in activity against oligosaccharides with a D.P < 6, with a k_{cat}/K_m ~1 x 10^3 M^{-1} min^{-1} against mannotetraose (D.P 4), and no activity observed against mannotriose over 24 h. Conversely, the periplasmic GH76 enzymes, BT2623 and BT3782, maintained catalytic efficiency against manno-oligosaccharides with a D.P as low as 4, with only a 2-fold decrease in activity observed between mannotetraose and mannotriose. These data demonstrate that the substrate binding cleft of the periplasmic enzymes are optimised to bind small oligosaccharides compared to the surface endo-mannanases.

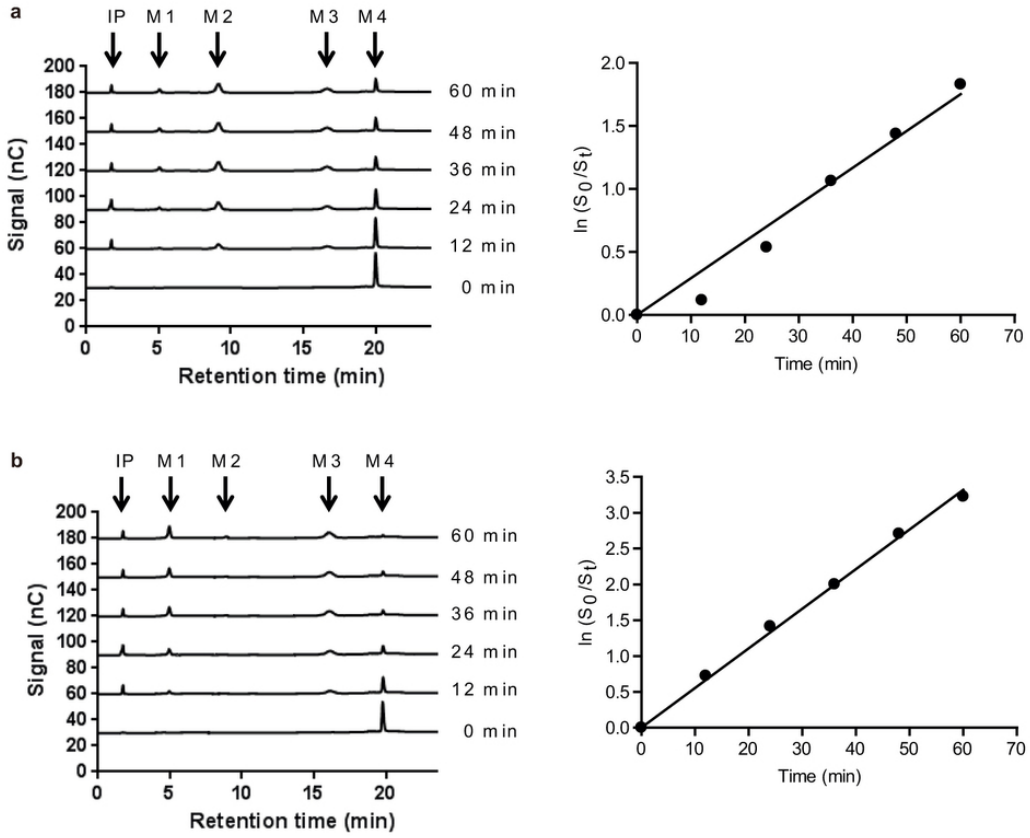


Figure 4.11. GH76 α-mannanases substrate depletion data generation. In Panel a and b, BT3782 (0.05 µM) and BT3792 (30 µm) respectively, were incubated with α1,6-mannotetraose at a concentration <<K_m.. Assays were performed in 50 mM Na-phosphate buffer, pH 7.5 in the presence of 0.01 mg/mL (w/v) BSA and incubated at 37°C for 1h. Reaction aliquots were taken at the respective time-point and heat deactivated. Substrate depletion was measured using HPAEC and the rate (b) was plotted using GraphPad Prism 5.0, enabling k_{cat}/K_M to be determined. α1,6-mannooligosaccharides were identified by their degree of polymerization (M1, mannose; M2, mannobiose; M3, mannotriose; M4, mannotetraose) and IP is injection peak.

Oligosaccharide d.p.	k_{cat}/K_m (M^{-1} min^{-1})			
	BT2623(s)	BT2631(p)	BT3792(s)	BT3782(p)
3	No Activity	$1.0 \times 10^3 \pm 1.0\ 10^2$	No Activity	$5.6 \times 10^3 \pm 2.0 \times 10^2$
4	$2.2 \times 10^2 \pm 0.1 \times 10^2$	$1.3 \times 10^5 \pm 2.0 \times 10^2$	$1.8 \times 10^3 \pm 0.3 \times 10^2$	$5.8 \times 10^5 \pm 3.0 \times 10^4$
5	$7.2 \times 10^2 \pm 0.1 \times 10^2$	$1.3 \times 10^6 \pm 6.6 \times 10^4$	$2.4 \times 10^3 \pm 0.5 \times 10^2$	$3.9 \times 10^6 \pm 1.8 \times 10^5$
6	$1.8 \times 10^4 \pm 1.3 \times 10^3$	$1.2 \times 10^6 \pm 5.2 \times 10^4$	$1.1 \times 10^5 \pm 5.9 \times 10^3$	$6.1 \times 10^6 \pm 1.6 \times 10^5$
7	$5.0 \times 10^4 \pm 3.7 \times 10^3$	$1.7 \times 10^6 \pm 1.8 \times 10^5$	$1.3 \times 10^6 \pm 9.7 \times 10^4$	$4.5 \times 10^6 \pm 2.3 \times 10^5$
8	$1.5 \times 10^5 \pm 5.6 \times 10^3$	$1.9 \times 10^6 \pm 1.3 \times 10^5$	$1.5 \times 10^6 \pm 1.3 \times 10^5$	$1.0 \times 10^7 \pm 5.6 \times 10^5$

Table 4.2. Kinetic data of the four GH76 α-1,6-mannanases VS α-1,6-manno-oligosaccharides (d.P 3-8).

4.3.2.1.5. Sequence and structural analysis of the GH76s encoded by *Bacteroides thetaiotaomicron*

Alignment of the amino acid sequences of the four GH76s revealed a conserved motif (YDD) (Figure 4.12). Comparison of the 3D crystal structure of a *Listeria innocua* GH76, Lin0763, with that of the GH76 enzyme, BT3792 (crystal structure solved by Dr. Mike D. Suits, Victoria University, Canada), revealed that the two aspartate (Asp) residues, Asp^{258} and Asp^{259} of the YDD motif are also structurally conserved (Figure 4.13). These conserved Asp residues are housed within a curved, cleft-like structure, possibly indicating the substrate binding cleft of the enzyme (See Discussion section). These data suggest that Asp^{258} and Asp^{259} are candidate catalytic residues of the GH76 family of glycoside hydrolases.

Figure 4.12. Sequence alignment of the four GH76 α1,6-mannanases encoded PUL-Man1 and PUL-Man2 of *B. theta*. Alignment of the amino acid sequences of the four GH76 enzymes; BT2623, BT2631, BT3782 and BT3792 respectively. The highly conserved YDD motif, comprising of the tentative catalytic residues of the enzymes, is highlighted in green. * represent conserved residues, : indicates conservation between groups of strongly similar properties and . indicates conservation between groups of weakly similar properties.

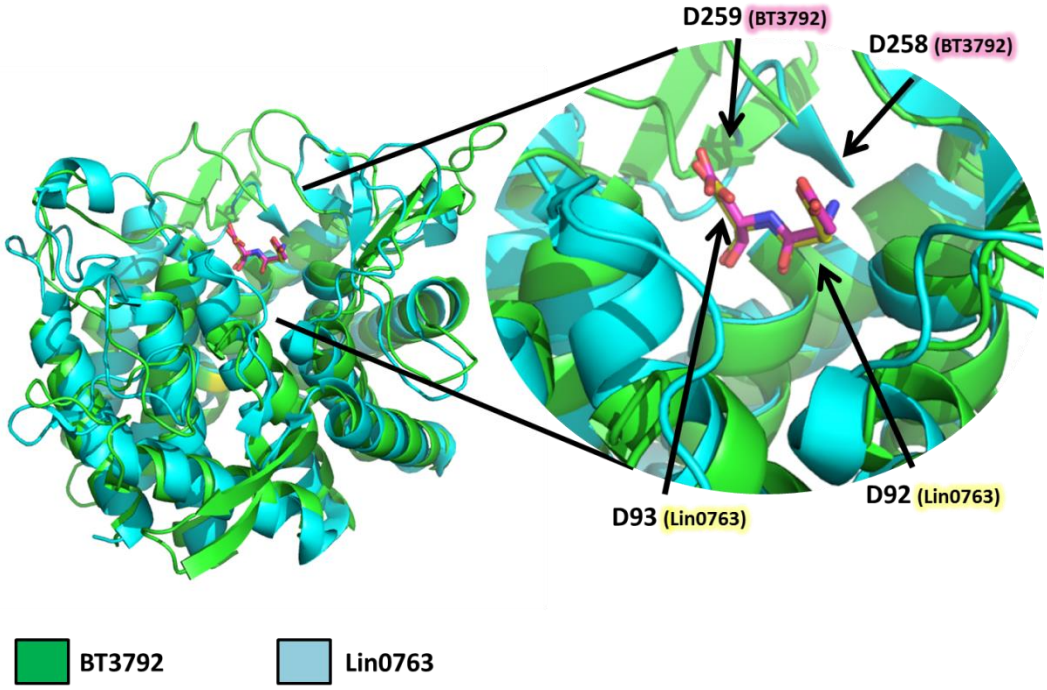


Figure 4.13. **Structural alignment of the 3D structures of BT3792 and Lin0763.** A structural alignment of the Bacteroides thetaiomicron GH76 (BT3792) with the *Listeria innocua* GH76 (Lin 0763), showing conserved aspartate residues. The structure of BT3792 is coloured green and the structure of Lin0763 is coloured blue. BT3792 aspartate residues are highlighted magenta whilst the Lin0763 aspartate residues are highlighted yellow.

4.3.2.1.6. Site directed mutagenesis of the catalytic residues of the GH76, BT3792.

To determine if the conserved aspartate residues of the GH76 enzyme BT3792 were the catalytic residues of the GH76 family of glycoside hydrolases, site directed mutagenesis was utilized. DNA encoding the single amino acid substitution was generated via PCR amplification using the primers displayed in Table 4.3 and analysed by electrophoresis, ensuring amplification of the DNA was successful (Figure 4.14). Parental DNA was degraded with Dpn1 a restriction enzyme that attacks only methylated DNA, and the introduction of the mutation into the cloned amplified DNA was confirmed by sequencing (MWG Eurofins), confirming that the candidate catalytic residues, Asp^{258} and Asp^{259}, had been individually substituted with alanine in the encoded proteins. Mutant derivatives of BT3792 were defined D258A and D259A respectively.

3792 variant	Sequence (5' -> 3')	Direction
D258A	GTTTTTACGCCGATAACGTATGGATCGGG	Forward
	CCCGATCCATACGTTATCGGCGTAAAAAC	Reverse
D259A	GTTTTTACGATGCCAACGTATGGATCGGG	Forward
	CCCGATCCATACGTTGGCATCGTAAAA	Reverse

Table 4.3. Primers used to generate BT3792 mutants.

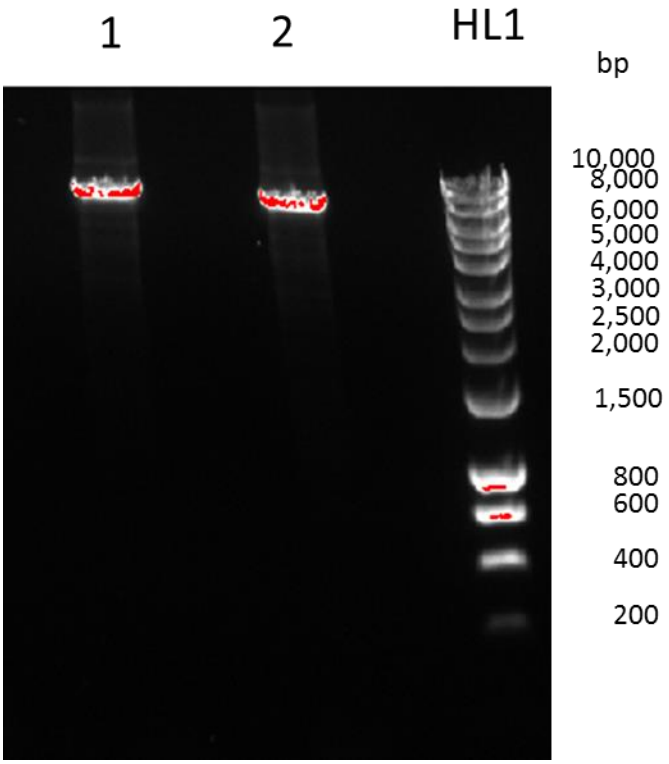


Figure 4.14. Quality control of the Site-Directed Mutagenesis of BT3792. An agarose gel of the site-directed mutagenesis PCR amplification of BT3792. Lane 1 = PCR product containing D258A mutation; Lane 2 = PCR product of D259A mutation. Invitrogen Hyper Ladder 1 (HL1) was run as a molecular weight marker. Agarose gels were 1% and electrophoresed for 1 h. Molecular weight markers indicate an amplication product of ~8000 bp, consistent with the size of vector plus the BT3792 insert.

4.3.2.1.6.1. Mutant expression and purification

D258A and D259A, were expressed as described in Section 3.1.2 with near-identical yields as the wild-type enzyme. An example SDS-PAGE of the purification of D258A and D259A is shown in Figure 4.15.

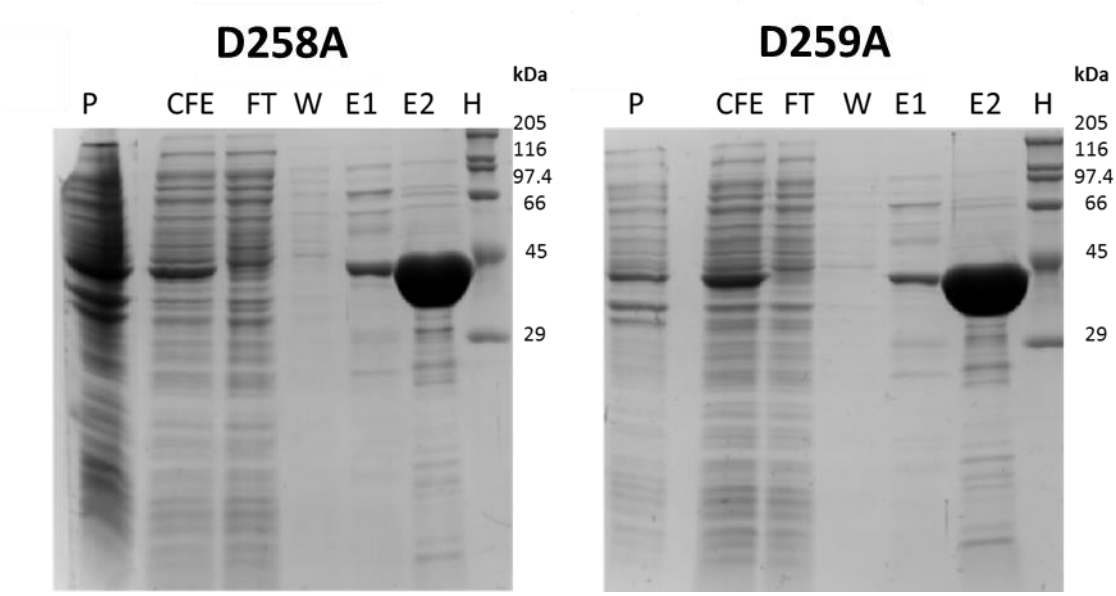


Figure 4.15. SDS- PAGE analysis of the protein purification of mutant variants of BT3792. SDS-PAGE data of BT3792 mutant variants D258A and D259A following IMAC. All SDS-PAGE gels were 12 % (w/v). L= low molecular weight marker; P = insoluble fraction; CFE = cell free extract; FT= flow through; W = wash; E1 = elution 1, 5 mM imidazole; E2 = elution 2, 100 mM imidazole. Both proteins ran with an apparent M_r of 45 kDa, consistent with their expected size.

4.3.2.1.6.2. Assaying the ability of BT3792 derivatives to degrade mnn2 mutant yeast mannan.

The catalytic activity of the BT3792 mutants, D258A and D259A, was assessed using a reducing sugar assay against the mutant yeast mannan, mnn2. The reducing sugar assay assesses the ability of the enzyme to hydrolyse glycosidic bonds, and ergo, reveal new reducing ends, which can be quantified using a standard curve of mannose. Mutant activity was assayed concurrently with wild-type BT3792. The assay was conducted in 50 mM Na-HEPES buffer, pH 7.5. Reactions were incubated at 37^0C for 30 min at a final substrate concentration of 5 mg/mL (w/v). Both of the mutant enzymes displayed no measurable activity against mnn2 yeast mannan at an enzyme concentration of 25 µM, whilst wild-type enzyme displayed activity at concentrations of 0.1 µM (a 250 fold difference). These findings suggest that Asp^{258} and Asp^{259} are the catalytic residues of the enzyme.

4.3.2.2. Investigating the activity of the GH125 enzymes, BT2632 and BT3781, encoded by the yeast α-mannan PULs of Bacteroides thetaiotaomicron.

4.3.2.2.1. Expression and purification

Both members of GH125 encoded by PUL-Man1 and PUL-Man2 were cloned by Mr. Carl Morland into pET28a. Recombinant forms of mature BT2632 and BT3781, consisting of amino acid residues 22 – 460 and 24 – 460 respectively, were expressed in *E. coli* BL21 cells, purified by IMAC and analysed via SDS-PAGE. Both GH125 enzymes were expressed in soluble form (Figure 4.16).

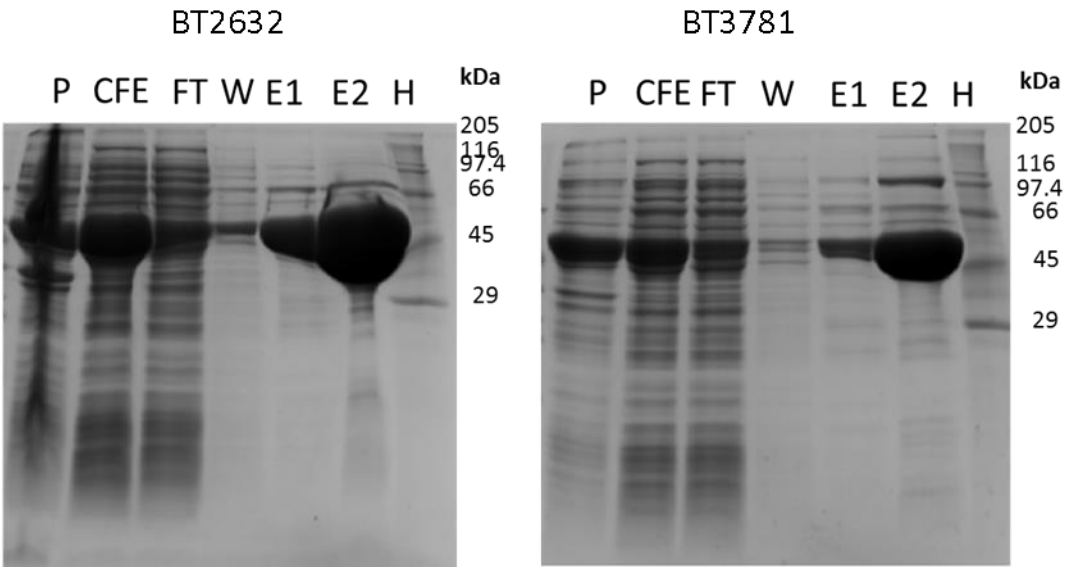


Figure 4.16. SDS-PAGE analysis of the protein purification of the GH125s, BT2632 and BT3781. The two proteins were purified by IMAC. All SDS-PAGE gels were 12 % (w/v). L= low molecular weight marker; P = insoluble fraction; CFE = cell free extract; FT= flow through; W = wash; E1 = elution 1, 5 mM imidazole; E2 = elution 2, 100 mM imidazole. Both proteins migrated according to their predicted size. BT2632 and BT3781 migrated with an apparent M_r of ~55 kDa.

4.3.2.2.2. The GH125s are α1,6-mannosidases

The GH125 enzymes encoded by PUL-Man1 and PUL-Man2 (BT2632 and BT3781 respectively), share 92% sequence homology and are predicted α1,6-mannosidases. Sequence analysis suggests both of the GH125 harbour a type I signal peptide, and so are predicted to be located to the periplasmic space. As such, the two enzymes were assayed against wild type and mnn2 mutant *S. cerevisiae* mannan. TLC and HPAEC-PAD analysis revealed that both enzymes were unable to degrade wild type yeast mannan (data not shown) but released mannose from the undecorated α1,6 backbone of mnn2 mannan, confirming that the two GH125 enzymes are exo-α1,6-mannosidases (Figure 4.17).

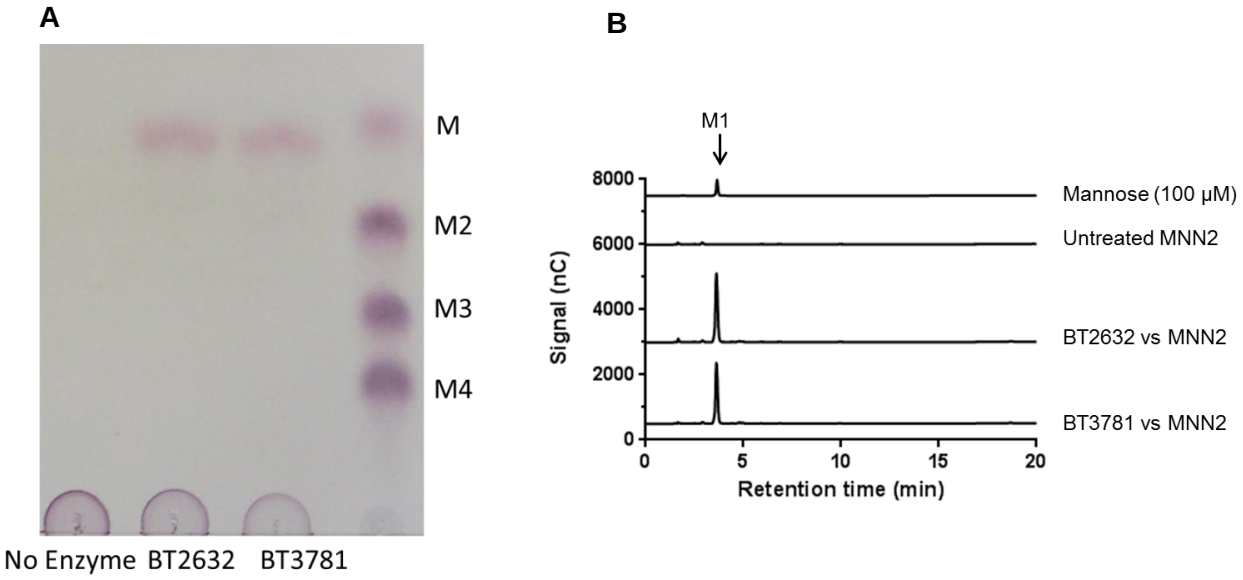


Figure 4.17. The activity of the GH125 α-mannosidases, encoded by PUL-Man1 and PUL-Man2 of *Bacteroides thetaiotaomicron*. BT2632 and BT3781 (0.1 µM) were incubated with 1 mg/ml (w/v) debranched mannan (mnn2) for 1 h in 50 mM Na-HEPES buffer, pH 7.5 at 37 °C. Panel A displays TLC analysis of the reactions, while the corresponding samples were analysed by HPAEC in Panel B. α1,6-mannooligosaccharides were identified by their degree of polymerization (M1, mannose; M2, mannobiose; M3, mannotriose; M4, mannotetraose) and IP is injection peak.

4.3.2.2.3. GH125 α1,6-mannosidase activity against α1,6-mannooligosaccharides.

Kinetic data of the two GH125 enzymes were generated versus α-1,6 linked mannobiose (D.P 2), mannotriose (D.P 3) and mannotetraose (D.P 4), using a mannose detection assay kit (Megazyme) as described in Materials and Methods, Section 2.2.1.1. Reactions were carried out in 50 mM Na-HEPES buffer, pH 7.5 at 37°C.

The kinetic data generated is displayed in Table 4.4 and Figure 4.18. Both of the GH125 enzymes (BT2623 and BT3781) displayed similar affinities (K_M) for each substrate, demonstrating a small preference for mannobiose over mannotriose and mannotetraose, indicating that both enzymes contain two subsites, -1 and +1.

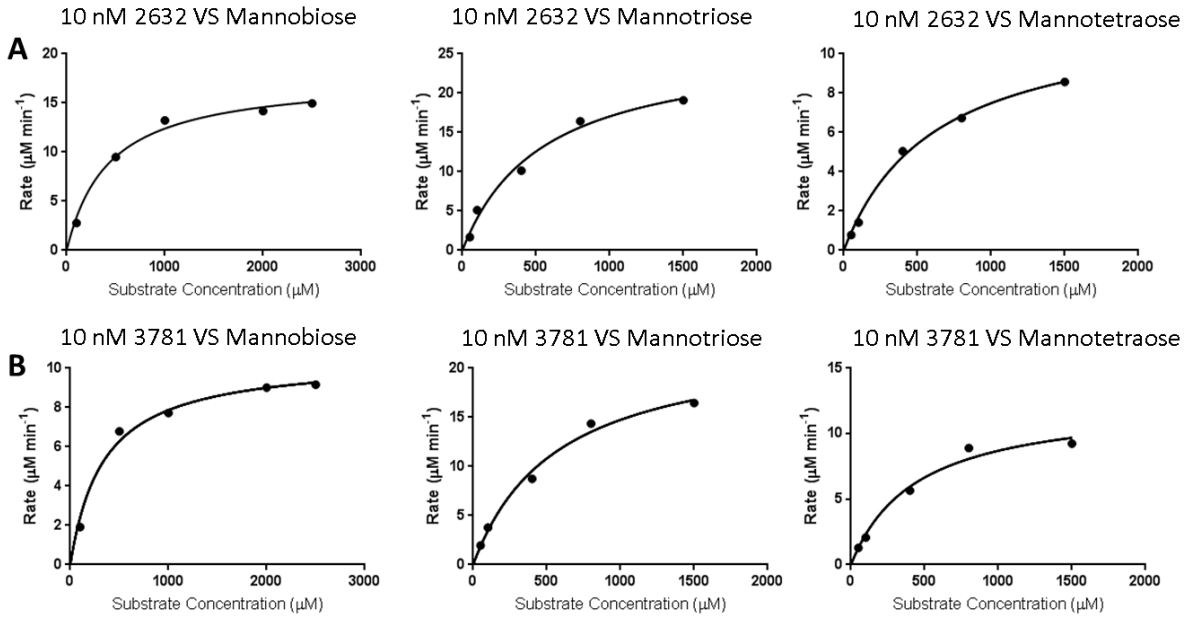


Figure 4.18. Graphs of the kinetic data generated for the two GH125s VS α1,6-manno-oligosaccharides (D.P 2-4). Kinetic data of BT2632 (**A**) and BT3781 (**B**) versus mannooligosaccharides. Initial rates were derived using the mannose-detection kit (Megazyme). Rate of absorption (A min^{-1}) was converted to rate of mannose released (μM min^{-1}) using the extinction co-efficient of NADPH and plotted in GraphPad Prism 5.0 using Michaelis-Menten kinetics.

GH125	Substrate	K_M (mM)	k_{cat} (min^{-1})	K_{cat}/K_M (mM^{-1} min^{-1})
BT2632	α1,6-mannobiose	0.43 ± 0.04	1777 ± 44	4133
	α1,6-mannotriose	0.53 ± 0.07	2613 ± 132	4930
	α1,6-mannotetraose	0.67 ± 0.07	1260 ± 60	1881
BT3781	α1,6-mannobiose	0.34 ± 0.03	1051 ± 22	3091
	α1,6-mannotriose	0.59 ± 0.05	2313 ± 90	3920
	α1,6-mannotetraose	0.42 ± 0.06	1239 ± 59	2960

Table 4.4. Kinetic data of the two GH125 α-1,6-mannasidases VS α-1,6-manno-oligosaccharides (D.P 2-3).

4.3.2.3. Assaying the activity of the GH97 BT2620, an α-galactosidase encoded by PUL-MAN1 of B. theta

BT2620 is a member of GH97 a family that contains α-galactosidases and α-glucosidases (Cantarel *et al.*, 2009). Mannan from the yeast *S. pombe* contains α-galactose side chains side chains and thus the polysaccharide is potential substrate for BT2620. To test this hypothesis the ORF encoding mature BT2620 was cloned into the *E. coli* expression vector pET28a. The protein was expressed in *E. coli* and subsequently purified by IMAC as described for the other *B. theta* enzymes described in this chapter, the methodology of which is detailed in Materials and Methods, Section 2.1.25.

4.3.2.3.2 Characterization of GH97 BT2620, an α-galactosidase

The data displayed in Figure 4.19 showed that when incubation of BT2620 with *S. pombe* α-mannan liberated galactose, demonstrating that the enzyme cleaved the α1,2-linked galactosyl units appended to the α1,6-mannan backbone. Thus, BT2620 is an α-galactosidase. No galactose was liberated from *S. cerevisiae* α-mannan by BT2620. The specificity of BT2620 was explored against several substrates containing α-linked galactose moieties using TLC. BT2620 displayed extremely low catalytic activity against raffinose (O-α-D-galactopyranosyl-(1,6)-α-D-glucopyranosyl-β-D-fructofuranoside) and melibiose (6-O-α-D-galactopyranosyl-D-glucose). No activity was observed against stachyose (α-D-galactopyranosyl-(1,6)-α-D-galactopyranosyl-(1,6)-α-D-glucopyranosyl-β-D-Fructofuranoside) or carob galactomannan, a polysaccharide consisting of a β1,4-linked mannose backbone with α1,6-galactose side-chains. Kinetic parameters were unattainable due to the extremely low catalytic activity of BT2620 against these substrates and *S. pombe* α-mannan

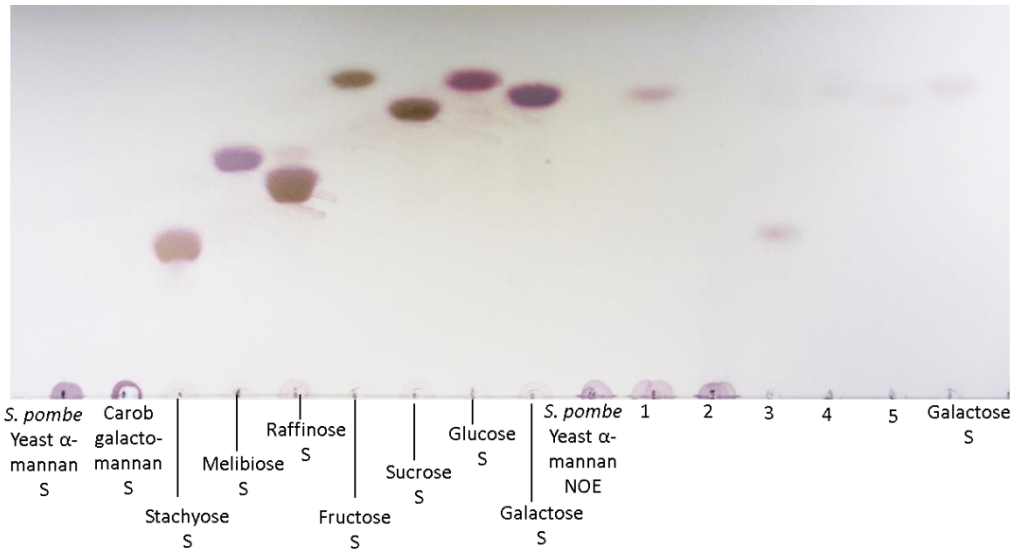


Figure 4.19. Thin Layer Chromatography assessing GH97 α-galactosidase activity. TLC analysis of GH97 BT2620 (1 µM), encoded by PUL-Man1, incubated with various polysaccharides/oligosaccharides containing α-linked galactose moieties. Assays were performed in 50 mM Na-HEPES, pH 7.0 at 37 ^{o}C for 1 h. Polysaccharide concentrations were 1 mg/mL (w/v) and oligosaccharide concentrations were 1 mM respectively. Samples were run against multiple standards, denoted by S. NOE denotes no enzyme control. Numbered lanes show GH97 BT2620 incubations with; 1 – *S. pombe* yeast α-mannan; 2 – Carob galactomannan; 3 – stachyose; 4 –melibiose and 5 – raffinose.

BT2620, displayed activity against pNP-α-D-galactose but not pNP-β-D-galactose or pNP- α-D-glucose. The enzyme was extremely slow with a catalytic efficiency (k_{cat}/K_M) of 56 min^{-1} M^{-1} (Figure 4.20). Individual kinetic parameters could not be determined as the K_M was >>3 mM, after which substrate solubility was a problem.

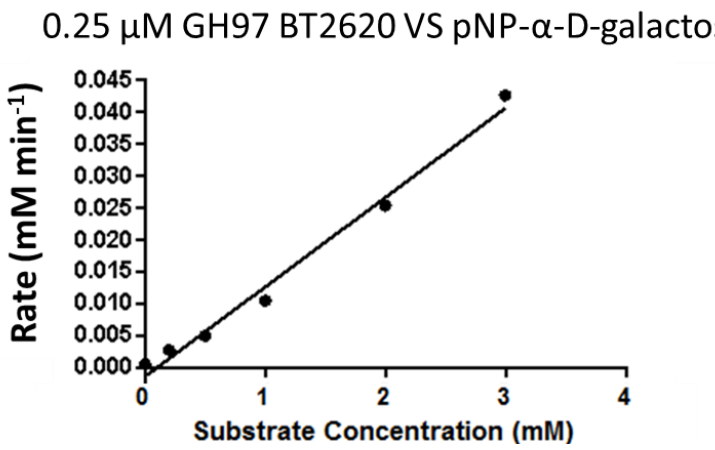


Figure 4.20. Graph of GH97 α-galactosidase BT2620 vs pNP-α-D-galactose to generate catalytic efficiency constant. Rates were generated spectophotometrically (A min^{-1}) at OD_{400nm}. The rate of pNP group cleavage and release, stoichiometric to glucose release (mM min^{-1}), was calculated using the extinction coefficient of pNP at pH 7.0. Activity against pNP-α-D-galactose was generated by incubating enzyme (0.25 µM) at 37 ^{o}C in the presence of Na- HEPES, pH 7.0. Rates were analysed using linear regression in GraphPad Prism 5.0.

4.3.2.3.3 De-branching activity of BT2620 against *S. pombe* α-mannan

The ability of BT2620, to de-branch *S. pombe* α-mannan and enable GH76 endo α-mannanase accesses to the α1,6-mannan backbone was assessed by incubating BT2620 (1 µM) in the presence of the periplasmic (BT3782) or surface (BT3792) GH76 endo-mannanases. TLC analysis of the reactions showed that in in the absence of α-galactosidase activity, the GH76 endo-mannanases were not able to attack the decorated α1,6-mannan backbone, as no mannooligosaccharides were generated. Conversely, in assays incorporating the α-galactosidase BT2620, the GH76 mannanases were able to attack the mannan backbone releasing mannooligosaccharides (Figure 4.21).

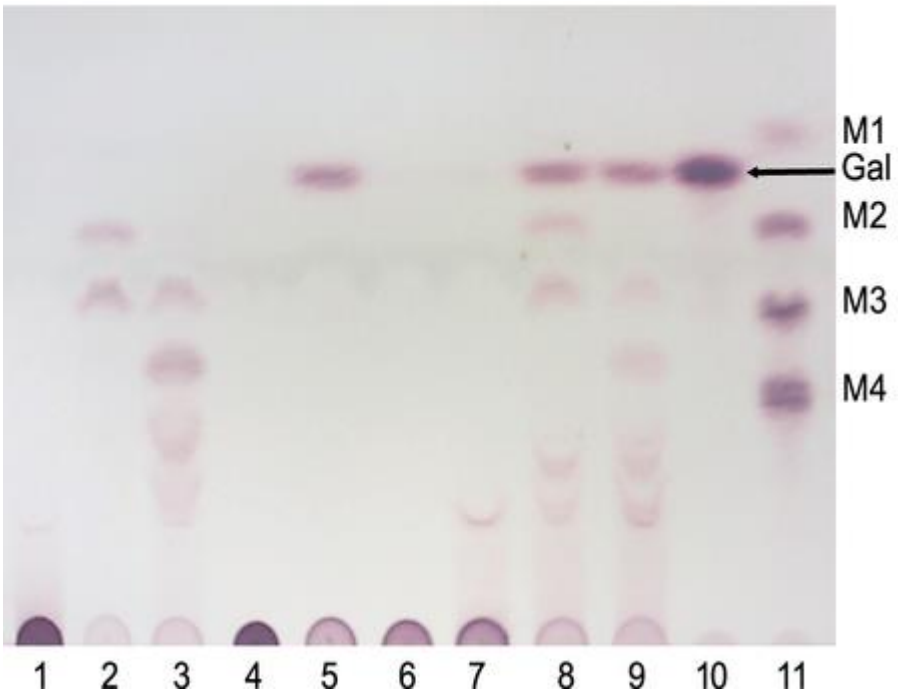


Figure 4.21. TLC of *S. pombe* mannan treated with BT2620 and other mannan degrading enzymes. TLC demonstrating that BT2620, a GH97 α-galactosidase, is capable of facilitating GH76 endo-α1,6-mannanase attack on the *S. pombe* α1,6-mannan backbone. Lanes 1, 2 and 3 show debranched mannan, mnn2 (1 mg/mL (w/v)) incubated without enzyme, with the periplasmic (BT3782) and extracellular (BT3792) GH76 endo-mannanase, respectively. Lanes 4 to 7 show individual incubations of *S. pombe* α-mannan (1 mg/mL (w/v)) with 4 – no enzyme; 5 – 1µM BT2620; 6 – 1µM BT3782; 7 – 1µM BT3792). Lanes 8 and 9 show incubations of *S. pombe* α-mannan (1 mg/mL (w/v)) with the BT2620 (1 µM) in tandem with 8 – BT3782 (1 µM) and 9 – BT3792 (1 µm). Lane 10 - galactose standard (Gal); Lane 11 – α1,6-mannooligosaccharides standards, D.P 1 – 4 (M1, M2, M3 and M4). Reactions were performed in 50 mM Na-HEPES, pH 7.0 and proceeded for 1 h at 37°C

4.3.3. Can the degradation products liberated by *B. theta* cultured on α-mannan support the growth of other non-mannan utilising *Bacteroides*?

To assess if the surface break down products of yeast α-mannan degradation generated by *B. theta* can sustain the growth of other non-mannan utilizing *Bacteroides* species, co-culturing experiments (in collaboration with Dr. Fiona Cuskin and Dr. Elisabeth Lowe) were performed with two additional *Bacterodies* species; *Bacteroides cellulosilyticus* WH2 (*B. cellulo*) and *Bacteroides xylanisolvens NLAE-zI-*

p352 (B. xylani). *B. cellulo* is unable to degrade and utilize α-mannan, but is able to grow on mannose. *B. xylani* contains a PUL syntenic with *B. theta* PUL-Man1. *B. xylani* exhibits very limited growth on *S. cerevisiae* mannan (Figure 4.22) but, in addition to growing on mannose, can depolymerise and utilize unbranched α-mannan, consisting solely of the α1,6-mannan backbone (Cuskin *et al.*, 2015b). *B. theta*, prepared as described in Materials and Methods 2.1.7, was co-cultured with either *B. xylani* or *B. cellulo* independently, in the presence of either mannose (5 mg/mL (w/v)) or yeast α-mannan (5 mg/mL (w/v)) as the sole carbon source. All experiments were performed in triplicate. Samples of bacterial growth (1 mL) were taken at the point of inoculation, early exponential phase, late exponential phase and stationary phase (Assessed by OD at 600nm). Cultures were plated onto rich media to determine the number of colonies at each phase of growth. The proportions of the Bacteroides species relative to each other were determined using quantitative PCR from genomic DNA, using unique markers for each strain (performed by Dr. Elisabeth Lowe). Both *B. cellulo* and *B. xylani* were also grown individually on *S. cerevisae* α-mannan. The results of the co-culturing experiment are depicted in Figure 4.22.

When *B. theta* was co-cultured with either *B. cellulo* or *B. xylani* in the presence of mannose, the relative proportion of the two co-cultured *Bacteroides* species remained constant from the point of inoculation to stationary phase, suggesting that the three *Bacteroides* species are capable of importing and catabolising mannose equally. When co-cultured on *S. cerevisiae* α-mannan, however, *B. theta* out-competed both *B. cellulo* and *B. xylani*; the relative proportion of the respective co-cultured organisms changed from ~1:1 at inoculation to almost 100% *B. theta* at stationary phase. These data suggest that the limited mannose released by *B. theta* during extracellular mannan degradation cannot support the growth of *Bacteroides* species.

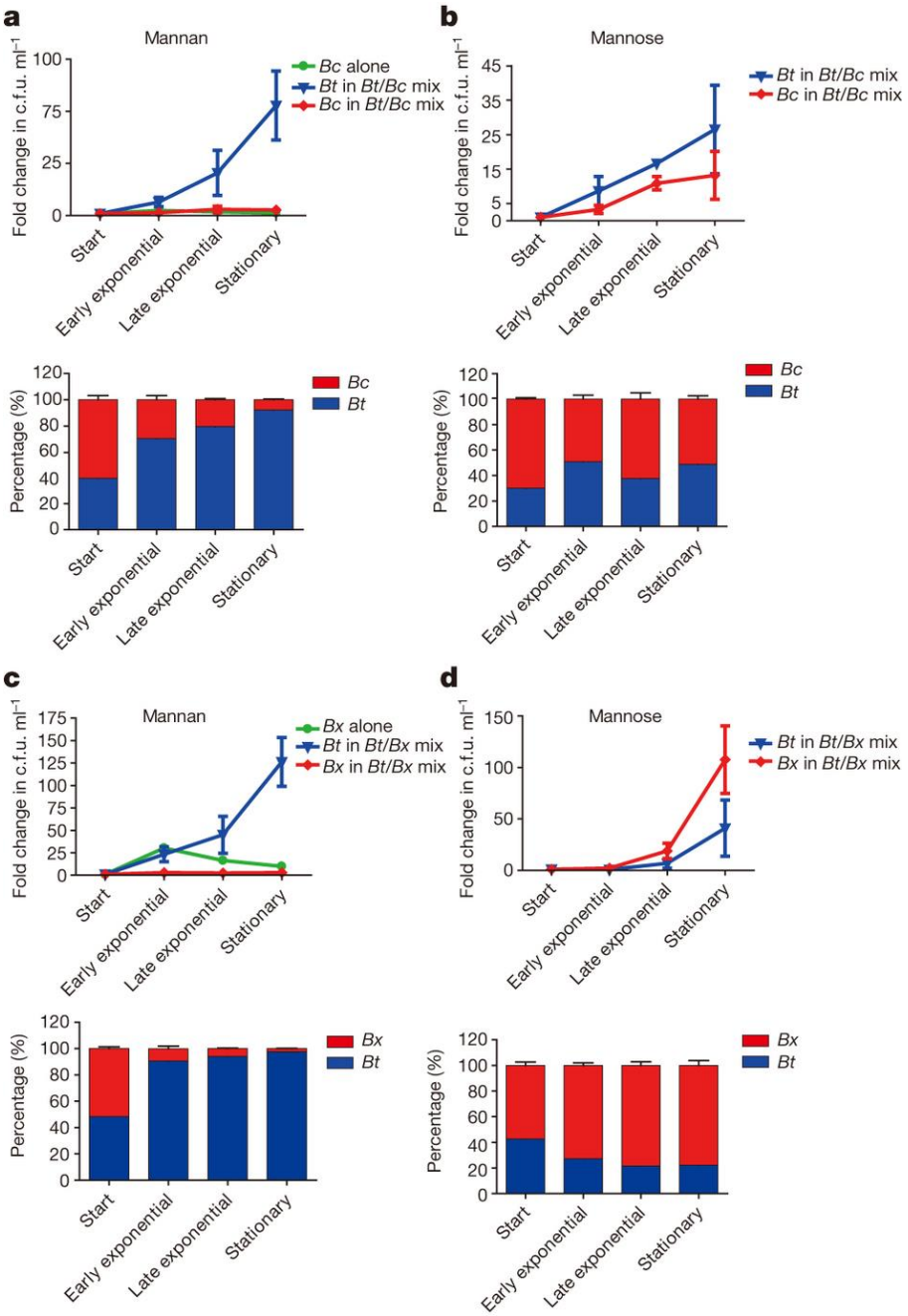


Figure 4.22. Assessing if yeast α-mannan degradation products, generated by *B. theta* can support the growth of other *Bacteroides* species. *B. theta* (*Bt*) was co-cultured anaerobically with *B. cellulo WHZ* (**a,b**) or *B. xylani* (*Bx*) *NLAE-zl-p352* (**c,d**), in the presence of mannose (**b,d**) or *S. cerevisiae* α-mannan (**a,c**) as the sole carbon source (5 mg/mL (w/v)). The upper graph of each panel depict the colony forming units mL^{-1} (c.f.u. mL^{-1}) of each strain, relative to the c.f.u mL^{-1} at inoculation. Total c.f.u. mL^{-1} was determined by colony counts. The relative proportion of each *Bacterodies* species was determined by qPCR (Dr. Elisabeth Lowe) of marker genes from genomic DNA (lower graph of each panel). Error bars represents the standard deviation of three biological replicates (Cuskin et al., 2015).

4.3.4. Further elucidating the enzymatic deconstruction of high mannose N-glycan by α-mannosidases encoded by the high mannose N-glycan polysaccharide utilization locus of *Bacteroides thetaiotaomicron.*

PUL-HMNG, encoded by *B. theta,* is activated solely by HMNG and not yeast α-mannan, despite the similarities in core structure and composition of the respective glycans (Cuskin *et al.*, 2015b). PUL-HMNG encodes three GH92 α-mannosidases that hydrolyse α1,2- (BT3990) α1,3- (BT3991) and α1,6- (BT3994) mannosidic linkages (Zhu et al., 2010). As reported by Zhu et al. (2010) using qualitative assays, BT3994 is able to hydrolyse the terminal α1,6-mannosidic linkage in HMNG treated with BT3990, which generates $Man_5GlcNAc_2$, or both BT3990 and BT3991 that produces $Man_3GlcNAc_2$ (Figure 4.23). BT3994, however, was shown not to be active against an α1,6-linked disaccharide of mannose. This observation suggests a specificity determinant is required for BT3994 activity, which is present in both $Man_5GlcNAc_2$ and $Man_3GlcNAc_2$, but absent in α1,6-mannobiose. It can be hypothesized that either BT3994 requires an essential mannose residue at the +2 subsite or that BT3994 displays specificity for a GlcNAc moeity at the +3 subsite of the enzyme for catalysis to occur. This hypothesis was probed in the following experiments.

4.3.4.1 A non-reducing end N-acetyl glucosamine is a specificity determinant, essential for GH92 BT3994 α-mannosidase activity.

To provide quantitative detailed information on the specificity of BT3994, a GH92 α1,6-mannosidase (Zhu et al., 2010), the activity of the enzyme was determined using fluorescent labelled substrate. A $Man_9GlcNAc_2$ oligosaccharide was labelled with 2-aminobenzamide (AB) as described in Chapter 2 Section 2.2.2. Such labelling compromises the integrity of the modified reducing end GlcNAc. $Man_5GlcNAc_2$-AB and $Man_3GlcNAc_2$-AB oligosaccharides were generated from $Man_9GlcNAc_2$-AB (Structures shown in Figure 4.23) using BT3990 and both BT3991 and BT3990,

respectively. This was verified by HPAEC-PAD using a fluorescence detection system. Likewise, α1,6-mannotetraose and α1,6-mannotriose were 2-AB labelled (Figure 4.24) to explore BT3994 specificity for mannooligosaccharides.

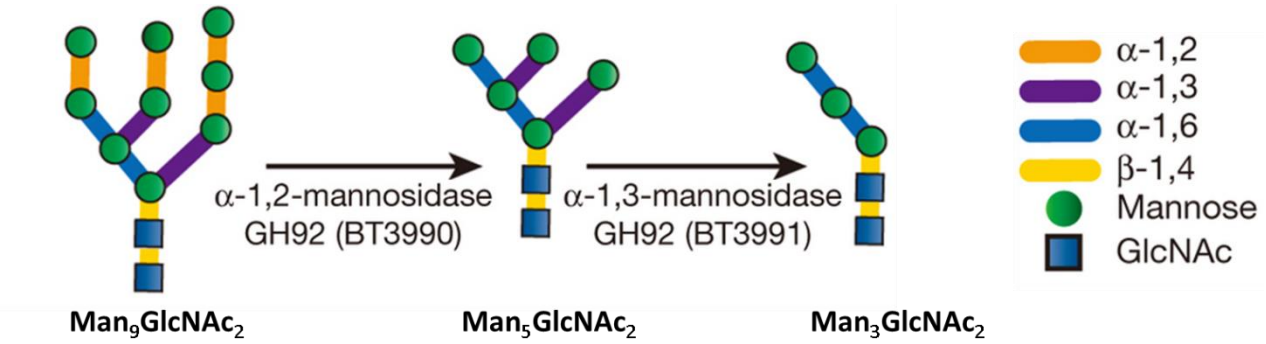


Figure 4.23. Structure of $Man_9GlcNAc_2$ and oligosaccharides generated to probe GH92 BT3994 α-mannosidase specificity. The structure of $Man_9GlcNAc_2$ is shown. $Man_5GlcNAc_2$ is generated by incubation of $Man_9GlcNAc_2$ with the GH92 α1,2-mannosidase, BT3990. $Man_3GlcNAc_2$ can be subsequently generated via treatment with the GH92 α1,3-mannosidase , BT3991. The specificity of the α1,6-mannosidase, BT3994 was tested against 2-Aminobenzine labelled $Man_5GlcNAc_2$ and $Man_3GlcNAc_2$ and analysed by a fluorescence detection system using HPAEC-PAD. The sugar/linkage composition of the oligosaccharides is denoted by a key.

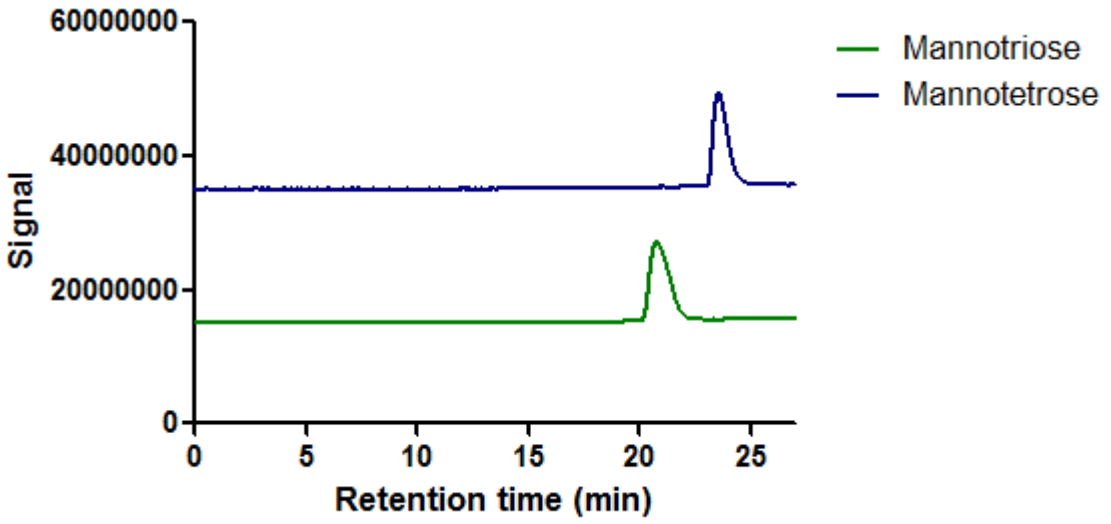


Figure 4.24. HPAEC-PAD flurescence detection of 2-aminobenzne labelled α1,6-mannotetraose and α1,6-mannotriose. An example of the detection of 2-aminobenzene (AB) labelled α1,6-mannotetraose (Man_4-AB) and α1,6-mannotriaose (Man_3-AB) using a florescence detection system (HPAEC-PAD). The respective labelled α1,6-oligosaccharide is distinguished by colour and denoted by a key.

The GH92 α1,6-mannosidase, BT3994, was active against both 2-aminobenzamide labelled $Man_5GlcNAc_2$-AB and $Man_3GlcNAc_2$-AB. Efficiency constants (k_{cat}/K_M) of BT3994 against these substrates were generated by substrate depletion using the same methodology as described in Section 4.3.2.1.4.and analysed by HPAEC-PAD using fluorescence detection. The data are reported in Table 4.5 and example analyses are shown in Figure 4.25. Interestingly, the GH92 α1,6-mannosidase was not active against α1,6-mannotetraose-AB (effectively mannotriose) at enzyme concentrations as high as 10 µM, strongly suggesting that BT3994 requires GlcNAc at the reducing end +3 subsite for catalytic activity. Likewise, the lack of activity BT3994 displays against α1,6-mannotriaose-AB (effectively mannobiose), confirms the observations of Zhu et al., (2010)

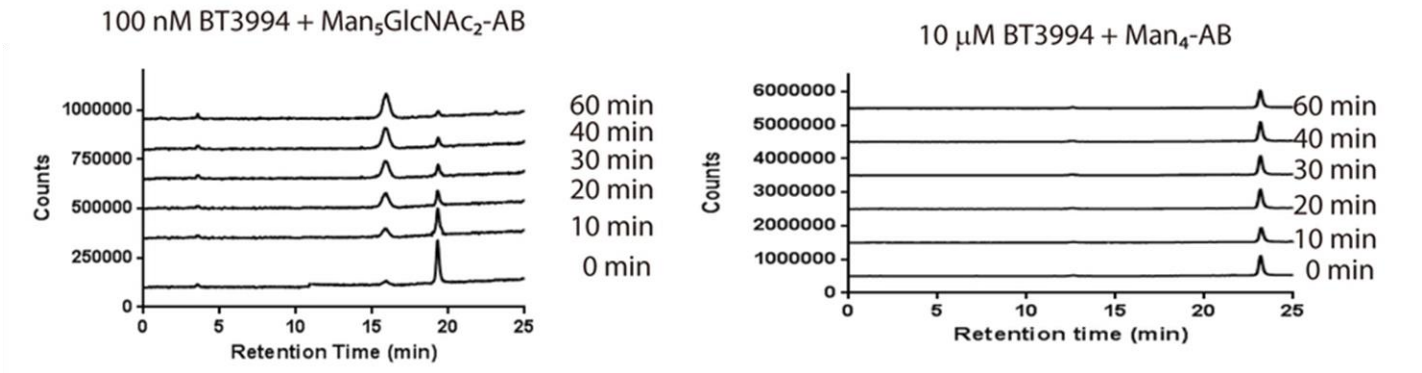


Figure 4.25. HPAEC-PAD fluorescent detection to monitor substrate depletion and activity of the GH92 α1,6-mannosidase, BT3994. Representative example of the substrate depletion data provided by GH92 α1,6-mannosidase BT3994 incubation with $Man_5GlcNac_2$-AB and α1,6-mannotetraose-AB (Man_4-AB), analysed by fluorescent detection (HPAEC-PAD). k_{cat}/K_M values were generated by determining the rate of substrate depletion using GraphPad Prism 5.0, if applicable. Assays were performed in 20 mM Na-citrate buffer, pH 7.5.

Enzyme	Substrate	K_{cat}/K_M (min^{-1} mM^{-1})
BT3994	α1,6-Mannobiose*	Nd
	α1,6-Mannotriaose**	Nd
	$Man_5GlcNAc_2$-AB	165
	$Man_3GlcNAc_2$-AB	391

Table 4.5. Catalytic efficiency of BT3994 against α1,6-mannooligosaccharides and exposed α1,6-linked mannose residues of HMNG. *Substrate is α1,6-Mannotriaose-AB and represents α1,6-Mannobiose. ** Substrate is α1,6-Mannotetraose-AB and represents α1,6-Mannotriaose.

4.4. Discussion

The biochemical data generated in this chapter attempts to elucidate the hierarchal degradation of yeast α-mannan orchestrated by PUL-Man1-3 of *B. theta*. As such a model of yeast α-mannan degradation by *B. theta* is displayed in Figure 4.29.

GH76 endo-α1,6-mannanase attack on the α1,6-mannan backbone is restricted due to steric constraints imposed by the mannan side chains of yeast α-mannan. Thus, limited side-chain removal must occur extracellularly to facilitate surface GH76 α-mannanases cleaving mannosidic linkages in the backbone. Attempts to elucidate the α-mannosidases localised to the cell surface which could fulfil this function revealed two such candidates; BT3990 and BT2199, which were able to de-branch yeast α-mannan sufficiently to enable the GH76 α-mannanases to attack the backbone. Subsequently, localisation experiments performed by Dr Fiona Cuskin and Dr Elisabeth Lowe showed that BT3990 is a periplasmic enzyme. Excluding the potential role of the enzyme in extracellular side-chain deconstruction (Cuskin *et al.*, 2015b). Instead, the α-mannosidase BT2199 is the most likely candidate enzyme. The BT2199 ORF appears to be encoded by a PUL, owing to the presence of a SusC/D-pair. This PUL and *bt2199* specifically were not upregulated by yeast α-mannan. However, only the fold up-regulation of *bt2199* in response to α-mannan was explored, and so the basal expression levels of BT2199 are unknown. Conversely, the key periplasmic debranching enzyme, BT3774, is up-regulated 50~fold compared to glucose grown cells, when *B. theta* cells were confronted with α-mannan (Cuskin *et al.*, 2015b). Uniquely, *B. theta* has seemingly evolved a strategy to utilize yeast α-mannan through deploying surface exo-acting enzymes which are not encoded, or up-regulated by, the cognate PUL or polysaccharide degradation products respectively (Cuskin *et al.*, 2015b). Instead, side-chain removal is mediated by a constitutively

expressed α-mannosidase, which will thus infrequently expose the mannan backbone. Indeed the limited removal of the mannan side chains at the surface of *B. theta* is also reflected by the very low activity of BT2199 compared to the functionally equivalent GH38 α-mannosidase BT3774 (Cuskin *et al.*, 2015b). *B. theta* has seemingly evolved a strategy which minimises nutrient loss at the cell surface, illustrated by the lack of oligosaccharides observed when *B. theta* is cultured with yeast α-mannan as a sole carbon source (Cuskin *et al.*, 2015b). The inability of *B. theta* support the growth of *B. xylani* or *B. cellulo* when co-cultured on yeast mannan supports the view that the breakdown of the glycan at the surface of *B. theta* is minimal. Whilst BT2199 is the candidate enzyme which mediates extracellular side chain removal, additional experiments must be performed to validate this claim. This hypothesis could be verified by knocking out *bt2199* and exploring whether the mutant is able to grow on yeast mannan.

The GH76 glycoside hydrolases encoded by PUL-Man1 and PUL-Man2 are endo-α1,6-mannanases. Site-directed mutagenesis of the two highly conserved aspartate residues, housed in a curved, cleft-like structure of the enzyme, abolished catalytic activity. Asp258 and Asp259 are therefore the candidate catalytic residues of the enzyme. Indeed, the topography of the active site is complementary to the conformation adopted by α-mannans as illustrated by Thompson et al. (2015). Indeed an overlay of the structure of BT3792 with a close GH76 homolog (*Bacillus ciculans (Bc)* endo-α1,6-mannanase; PDB code 5AGD) in complex with mannopentaose revealed that the ligand is located in a curved cleft containing the two catalytic residues, confirming the location of the substrate binding region and the active site Figure 4.26. The distance separating the two catalytic aspartate residues of the

enzyme of ~5 Á is typical of glycoside hydrolases that utilize a retaining, double displacement mechanism of catalysis (Rye and Withers, 2000), suggesting that the GH76 enzyme BT3792 and, by inference, all members of the family, are retaining enzymes. This hypothesis was confirmed by NMR performed by Dr. Alan Cartmell, revealing that the α-mannooligosaccharide products generated by the GH76 enzyme BT3792 demonstrated an overall net retention in anomeric configuration (Cuskin *et al.*, 2015b). In the structural overlay described above, Figure 4.27 Asp258 is below the plane of the mannose bound in the active site (-1 subsite) and Oδ1 of the carboxylate is 2.6 Å from the anomeric carbon demonstrating its role as the catalytic nucleophile. Oδ1 of Asp259 is within hydrogen bonding distance (2.9 Å) indicating that this carboxylate residue functions as the catalytic acid/base.

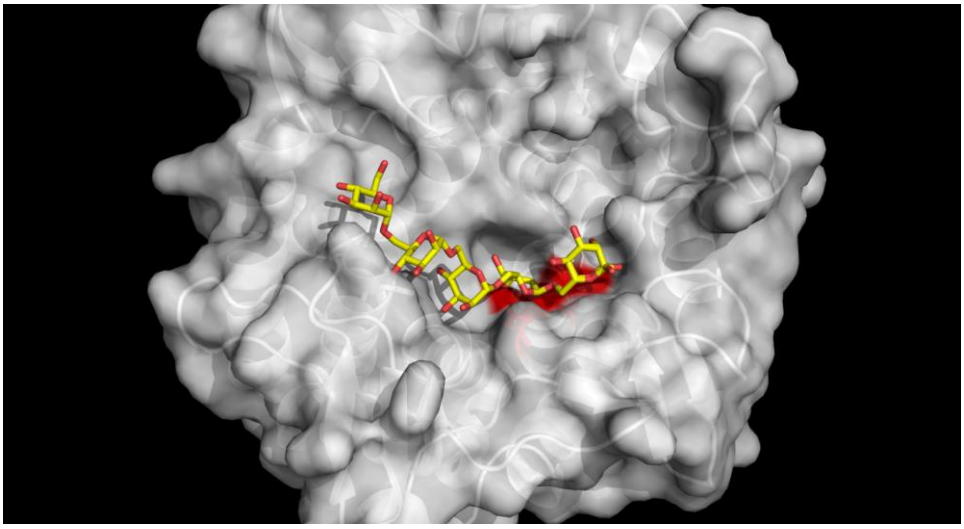


Figure 4.26. Crystal structure of the GH76 endo-α,16-mannanase, BT3792, in complex with α1,6-mannopentose. α1,6-mannopentaose is displayed in the ligand binding cleft of BT3792, and is coloured yellow. Oxygen moieties are coloured pink respectively. The surface of the enzyme is displayed and coloured grey. The conserved catalytic residues of the enzyme, Asp^{258} and Asp^{259} are coloured red. PDB code: 5AGD.

Endo-α1,6-mannanase activity is key to the sequential degradation of yeast α-mannan, illustrated by the observed differences in activity displayed between the surface and periplasmic GH76 enzymes. The two extracellular GH76 endo-α1,6-mannanases (BT2623 and BT3792) generate longer oligosaccharides from the mannan backbone than the corresponding periplasmic enzymes (BT2631 and BT3782), and are catalytically slower against the polysaccharide. Kinetic data of the GH76 enzymes versus α1,6-mannooligosaccharides suggest that the periplasmic endo-mannanases are more adapted for the degradation of small chain mannooligosaccharides than the corresponding extracellular enzymes. These findings offer a potential model by which the backbone of yeast mannan is processed by *B. theta*. The surface GH76 enzymes are catalytically slower and make infrequent cuts along the exposed backbone of yeast mannan, generating large oligosaccharides which can be imported into the periplasm of the cell. Once within the cell, periplasm the GH76 enzymes rapidly degrade the imported mannooligosaccharides. The longer oligosaccharides generated by the surface mannanases ensures minimal extracellular metabolism and prevents loss of substrate to the human gut microbiota. Indeed, limiting the extracellular cleavage of long chain oligosaccharides would help synchronise the rate of import with cleavage; maximising substrate acquisition through reduced extracellular metabolism

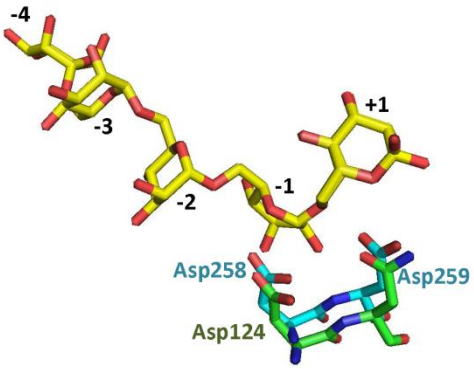


Figure 4.27. Overlay of the catalytic residues of the GH76 endo-α16-mannnases, BT3792 and BcGH76, in complex with mannopentaose. The catalytic residues of BT3792 are depicted in blue. The homologous residues, located within the BcGH76 (PDB code: 5AGD) active site are coloured green respectively. The bound ligand, α1,6-mannopentaose is coloured yellow, with pink oxygen moieties. The sub-sites of the enzyme in which each respective pyranose ring of mannopentoase interacts is labelled accordingly.

The observed lower rates of surface GH76 endo-mannanase activity is consistent with limited biochemical data of surface *Bacteroides* enzymes (arabinanases and xylanases), which also display low catalytic activity (Cartmell *et al.*, 2011). The periplasmic GH76 endo-mannanases, by generating numerous short chain α1,6-oligosaccharides, maximise the substrate available to the periplasmic located GH125 α1,6-mannosidases.

The two GH125 α-mannosidases, BT2632 and BT3781, encoded by PUL-Man1 and PUL-Man2 respectively, are exo-α1,6-mannosidases which are capable of degrading α1,6-mannooligosaccharides and demonstrate the highest affinity for mannobiose. The preference for the disaccharide suggests that two sugar binding sub-sites are utilized by the enzymes, with Man-α1,6-Man motif occupying the -1 and +1 subsites. The crystal structure of the apo form of BT3781 and the *Clostridium perfringens* GH125 exo-α1,6-mannosidase, CpGH125, were previously determined, the latter in complex with α1,6-mannobiose (Gregg *et al.* (2011). In the crystal-structure of CpGH125 the non-reducing end of mannobiose occupied the -1 subsite and the

reducing end mannose was bound at the +1 subsite. The CpGH125 catalytic site has the potential to accommodate longer oligosaccharides, with the reducing end sugar protruding into solvent. The structure of the catalytic site of CpGH125 indicates an inability for the enzyme to accommodate α1,2-linked side chains as O2 of the mannose at the +1 and/or -1 subsites are pointing at the enzyme surface. Indeed an overlay of the apo crystal structure of BT3781 and CpGH125 show that the active site (-1 subsite) and the +1 subsite are completely conserved in the two enzymes side chains and is consistent with the inability of the *B. theta* enzyme to act on α1,6-mannan backbones that contain side chains appended to O2, Figure 4.28. The structural conservation between the two enzymes also indicates that BT3781 only contains two subsites explaining why the enzyme displays maximum activity against mannobiose. The observation that the GH125 *B. theta* enzyme could not act upon the highly decorated yeast mannan, indicate that side chain removal of oligosaccharides imported into the periplasm occurs prior to backbone depolymerisation by the GH76 and GH125 enzymes.

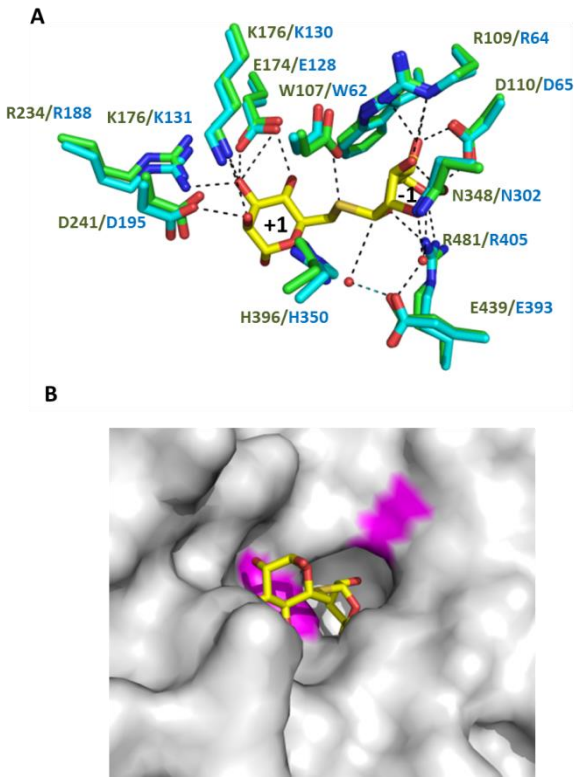


Figure 4.28. The structures of GH125 a1,6-mannosidases that play a key role in yeast mannan degradation. Panel A shows an overlay of BT3781 (green; PDB code 2P0V) with the substrate and catalytic residues of the *Clostridium perfringens* GH125 a-mannosidase CpGH125 (cyan; PDB code 3QT9), in which the ligand 6-S-alpha-D-mannopyranosyl-6-thio-a-D-mannopyranose (Man-S-Man) is shown in yellow. Panel B shows the solvent exposed surface of BT3781 in the vicinity of the active site in which the catalytic residues (Glu174 and Glu439) are depicted in magenta. The position of Man-S-Man is based on the overlay shown in Panel A. The figure was prepared using PyMOL.

B. theta has seemingly adapted a 'selfish' mechanism of yeast α-mannan utilization, supported by biochemical characterisation of the enzymes encoded by PUL-Man1/2/3 and the inability of *B. theta* mannan degradation products to sustain the growth of other *Bacteroides* species. These data are entirely consistent with the observation that wild type *B. theta* is out-competed by a *B. theta* PUL-Man1/2/3 deletion strain in animals fed a diet not containing α-mannan, highlighting the significant energy cost/investment targeting yeast α-mannan entails, most likely owing the significant number of enzymes, transporters and binding proteins that must be expressed to fully degrade the polysaccharide (Cuskin *et al.*, 2015b). These results reinforce the hypothesis that yeast α-mannan is a consistent nutrient source for the human gut microbiota. Indeed the selfish utilization strategy of high complexity/low abundance glycans extends to other PUL systems, such as the *Bacteroides ovatus* utilization of complex xylans (Rogowski et al., 2015).

PUL-Man1 encodes a GH97 α-galactosidase, BT2620, which removes the α-galactose decorations appended to the backbone of *S. pombe* α-mannan to facilitate endo-mannanase attack. This indicates that *B. theta* has evolved the ability to target the α-mannan of different yeast species. The importance of BT2620 in *B. theta S. pombe* α-mannan utilization is emphasized by observed growth defect of the Δ*PUL-Man1 B. theta* mutant cultured on *S. pombe* mannan (Cuskin *et al.*, 2015b). Interestingly, PUL-Man2 encodes a GH130 glycoside hydrolases capable of targeting the β1,2-mannosidic linkages which cap the mannose side-chains of *Candida albicans* α-mannan, a pathogenic yeast and commensal member of the gut microbiota (Cuskin *et al.*, 2015a). This reinforces the variety of fungal mannans targeted by *B. theta.*

The GH92 α1,6-mannosidase, BT3994, encoded by PUL-HMNG, targets the α1,6-mannosidic linkages of $Man_5GlcNAc_2$-AB and $Man_3GlcNAc_2$-AB. The AB labelled

substrates utilized to assay BT3994 specificity in this chapter destroyed the terminal reducing end GlcNAc and thus the substrate targeted by these enzymes were $Man_5GlcNAc$ and $Man_3GlcNAc$. Given that the enzyme did not attack □1,6-mannotriose GlcNAc positioned in the +3 subsites is an absolute specificity determinant. Although BT3994 can hydrolyse substrates extended past the +3 subsite it is likely that $Man_5GlcNAc$ and $Man_3GlcNAc$ are the biologically substrates, as the GH18 endo-β-N-acetylglucosaminidase encoded by PUL-HMNG is localised to the cell surface. This enzymes releases HMNG by hydrolysis of the β1,4-linkage of the terminal GlcNAc-GlcNAc moiety of $Man_9GlcNAc_2$, before importation into the cell. The periplasmic localised enzymes of the HMNG-PUL would therefore only be presented with $Man_9GlcNAc$. These data show that BT3994, is specific for HMNG, and supports the evidence that the HMNG-PUL apparatus is distinct from the α-mannan degradation apparatus, encoding an enzyme profile tailored to HMNG degradation, but not yeast α-mannan, deconstruction. Indeed, the inability of a mutant strain of *B. theta* lacking bt3993 (extracellular sigma-factor regulator of PUL-HMNG) to grow on $Man_9GlcNac_2$ demonstrates that the HMNG-PUL targets HMNG exclusively (Cuskin *et al.*, 2015b).

While α-mannan represents a nutrient to the human microbiota, the polysaccharide may also have clinical implications. Patients with Crohn's disease often have circulating anti-*S. cerevisiae* antibodies (ASCA). The epitope for this antibody is a mannan with α terminal Man-α1,3-Man disaccharide (Peeters *et al.*, 2001);(Quinton *et al.*, 1998), suggesting a relationship between the polysaccharide and Crohn's disease. This view is supported by the observation that the ASCA epitope, expressed by a number of microorganisms inhibits the killing of *Escherichia coli* by macrophages and monocytes, suggesting that the suppression of mucosal phagocyte function by

microbial α-mannan may contribute to Crohn's disease pathogenesis (Mpofu *et al.*, 2007). Intriguingly mannan induces the presentation of the endo-α1,2-mannosidase BT3862 and the α1,3-mannosidase on the surface of *B. thetaiotaomicron* that, collectively, would degrade the ASCA epitope and thus prevent the inhibition of mucosal phagocytes by α-mannans. Thus, in addition to utilizing yeast mannan as a nutrient, the capacity of *B. thetaiotaomicron* to express a repertoire of α-mannosidases that target α-mannans may also protect the large bowel from Crohn's disease and inflammatory bowel disease generally.

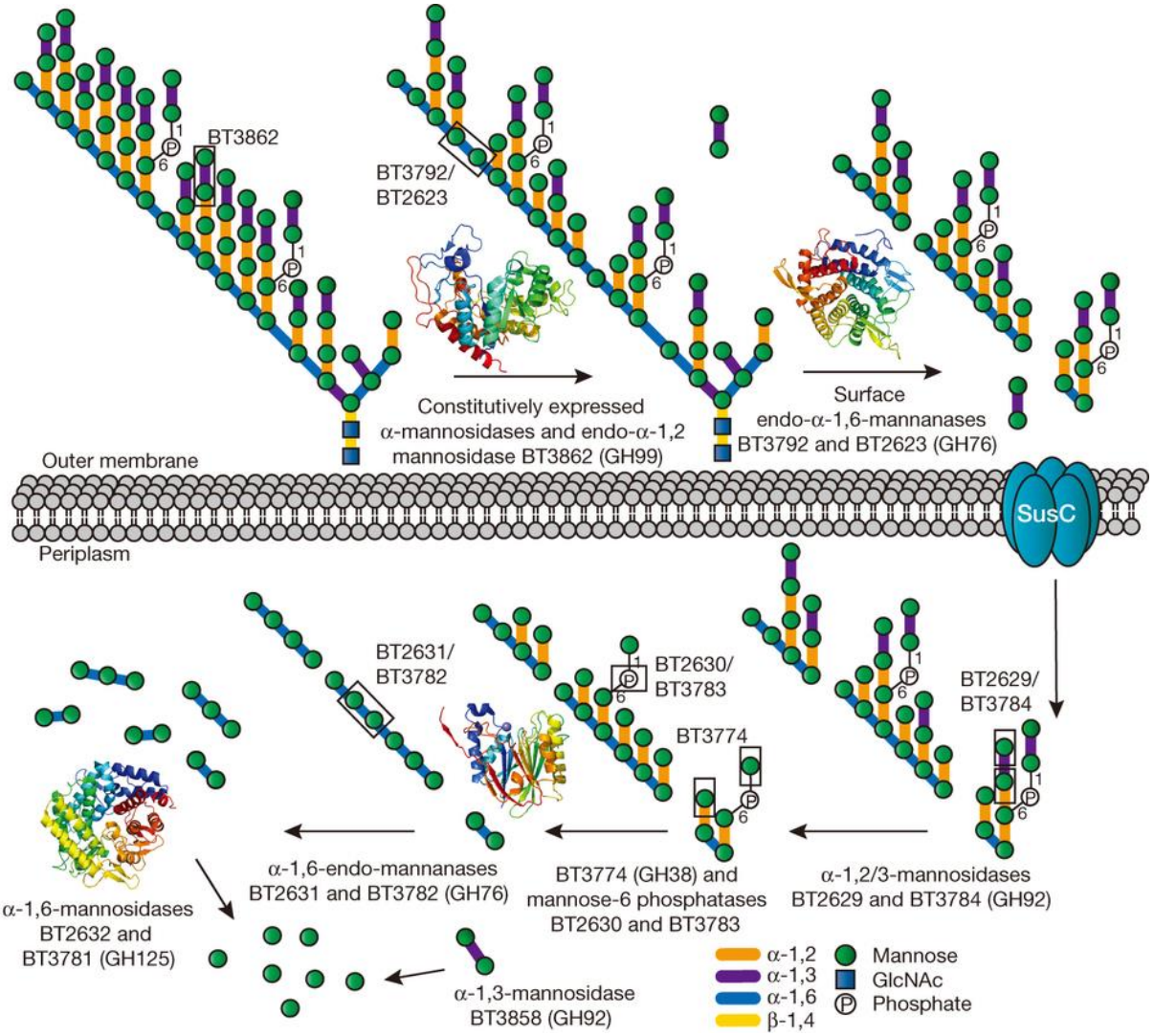


Figure 4.29. Model of Yeast α-mannan degradation by *Bacteroides thetaiotaomicron*. In this model limited degradation of the mannan occurs at the surface and the bulk of glycan degradation occurs in the periplasm. Yeast α-mannan is partially de-branched at the cell surface by constitutively expressed α-mannosidases (most likely the GH92 α-mannosidase, BT2199 and potentially with GH99 endo-α1,2-mannosidase assistance), providing access for the extracellular GH76s endo-α1,6-mannanases (BT2623 and BT3792) to the α1,6-mannan backbone. High molecular weight oligosaccharide products are transported through the SusC homolog into the periplasm. Following concerted side-chain removal via GH38 α-mannosidase (BT3774), GH92 α-mannosidase (BT2629/BT3784) and phosphatase (BT2630 and BT3783) activity, periplasmic GH76 enzymes (BT2632 and BT3782), hydrolyse the linear α-1,6-mannooligosaccharides into small chain products (D.P ≤3). The α1,6-mannooligosaccharides upregulated the mannan PULs prior to their hydrolysis into mannose by the GH125 α1,6-mannosidases BT2632 and BT3781, which can then be imported and fermented by the bacterium.

4.5. Further work

Further work must be performed in order to elucidate if BT2199 is the only extracellular candidate α-mannosidase capable of debranching yeast α-mannan. This could be investigated via knocking out the ORF encoding BT2199 and assessing the ability of the mutant to grown on yeast α-mannan and the linear mannan backbone.

Furthermore, to further assess if the break down products of *B. theta* growth with yeast α-mannan can support the growth of other *Bacteroides*, a more compelling experiment would be to co-culture wild type *B. theta* with a Pul-Man1/2/3 deletion strain of *B. theta*, reducing any possible inter-species differences in mannan catabolism. Similarly, a deletion strain of the surface GH76 endo-mannanases encoded by PUL-Man1 and PUL-Man2 could be co-cultured with wild type *B. theta* to assess if the release of high molecular weight oligosaccharides during α-mannan depolymerisation are able to support growth of other potential mannan utilizers.

Chapter 5: *Bacteroides thetaiotaomicron* and *Bacteroides ovatus* can utilize the β1,6-glucan fraction of the yeast cell wall as an energy source

5.1 Introduction

The Bacteroides genus, exemplified by the gram negative *Bacteroides thetaiotaomicron* (*B. theta*), represents a significant proportion of the human gut microbiota. Around 40% of the *B. theta* genome encodes carbohydrate active enzymes (CAZymes) enabling this bacterium to utilize a myriad of complex dietary and host-derived glycans (Rogowski *et al.*, 2015). Such CAZymes, many of which belong to sequence based glycoside hydrolase families (GHs), are often encoded in distinct genetic loci that target and are up-regulated by specific glycans. These polysaccharide utilization loci (PULs) encode periplasmic and cell envelope associated proteins that mediate the acquisition, degradation, and utilization of specific polysaccharides. All PULs encode homologues to the SusC/SusD pair present in the starch utilization locus. These proteins, which are localised to the extracellular membrane, mediate glycan import into the periplasm. Hypothetical PULs are identified by the presence of homologues of SusC and SusD (defined as $SusC_h$ and $SusD_h$, respectively), which are upregulated by the depolymerised products of their cognate polysaccharide. The portfolio of PULs is not conserved amongst the Bacteroidetes genera, and thus the glycan degrading capacity varies in the organisms that belong to this phylum (Larsbrink, 2014).

It was previously reported that when *B. theta* was cultured with yeast extract, peptone, dextran (YPD) media, the expression of a PUL (defined henceforth as PUL_{btPUS}) containing *bt3310* and *bt3311* that encode a $SusC_h$ and $SusD_h$, respectively, were upregulated (Sonnenburg *et al.*, 2010). Bioinformatic analysis of PUL_{btPUS} identified a number of potential protein encoding genes including a homologue to the regulator of

the starch PUL, defined here as $SusR_h$ (BT3309), a GH30 glycoside hydrolase (BT3312), a surface glycan binding protein or SGBP (BT3313) and a GH3 glycoside hydrolase (BT3314). It is possible that the proteins encoded by PUL_{btPUS} orchestrate the deconstruction of a polysaccharide present in the YPD medium. A schematic of this predicted PUL is shown in Figure 5.1.

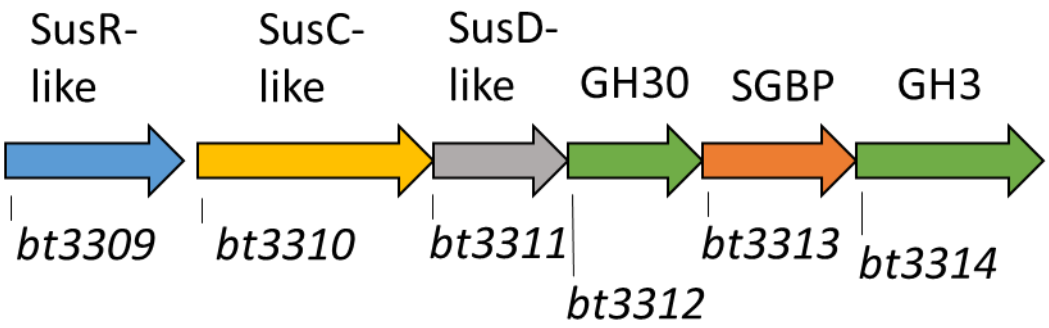


Figure 5.1. Schematic of PUL_{btPUS}. Hypothetical protein encoding open reading frames (ORFs) are shown as arrows (to scale), the direction of which indicates the orientation of the ORF. Predicted activity is annotated above the ORF, whilst the locus tag of the ORF is annotated below. Sus-like indicates the predicted role of the encoded protein based on sequence similarities with the homologous protein in the prototypic starch utilization system (Sus). GH (glycoside hydrolase) followed by a number indicates the amino-acid sequence based family the protein is categorised in the CAZy database. SGBP represents surface glycan binding protein.

Yeast cell wall polysaccharides present in the yeast extract of YPD media are potential target substrates for PUL_{btPUS}. The yeast cell wall, consisting of a number of complex polysaccharides, includes chitins, α-mannans and β-glucans that differ in linkage (See Chapter 1, Section 1.3.2). *B. theta* can utilize yeast α-mannan as an energy source (Cuskin *et al.*, 2015b), suggesting yeast cell wall polysaccharides are viable targets for the human gut microbiota. Indeed, the acquisition and evolution of the PULs which target yeast mannan coincides with the human domestication of yeast (Cuskin *et al.*, 2015b). The predominant glycan in the yeast wall, comprising ~80%, is β1,3-glucan decorated with β1,6-linked glucose side chains. These β-glucans comprise up to 1500 glucose residues and plays a structural role in the yeast cell wall. The helical conformation adopted by the β1,3-glucans provides tensile strength to the cell wall, protecting the yeast cell against turgor induced by osmatic stress. The minor β1,6-glucans, constituting 10 – 15% of the *Saccharomyces cerevisiae* (*S. cerevisiae*) cell wall, consist of a β1,6-glucan backbone substituted with β1,3-glucose decorations, and are estimated to extend to ~200 glucose residues in length. The proportion of β1,3-glucose side chains is species specific, being 7 % in *S. cerevisiae* but can be as much as up to 75 % in *Schizosaccharomyces pombe* (Aimanianda *et al.*, 2009). The β1,6-glucan fraction of the cell wall is covalently linked to the β1,3-glucan fraction and cross-links the yeast cell wall (Aimanianda *et al.*, 2009). The importance of the β1,6-glucan fraction to cell wall structure and yeast survival, is demonstrated by the lack of viability shown by yeast mutants which lack the transferases responsible for the assembly of the β1,6-glucan fraction *in vivo* (Kurita et al., 2011*)*. A representation of the structure of the β1,6-glucan fraction of the yeast cell wall is displayed in Figure 5.2

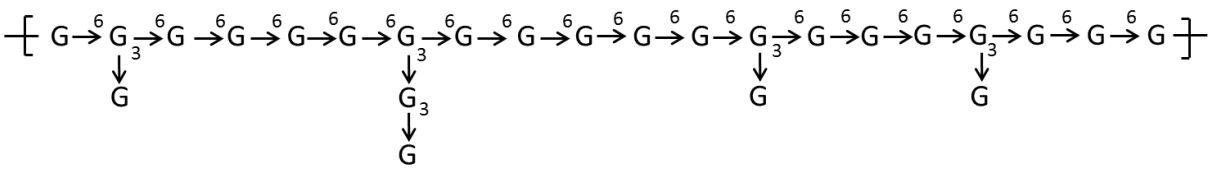


Figure 5.2. Structure of the β1,6-glucan fraction of the *Saccharomyces cerevisiae* cell wall. G indicates glucose residues. All arrows represent β-bonds, with numbers besides the arrows denoting linkage. Figure adapted from Aimanianda et al. (2009)

Previously characterised GH30 glycoside hydrolases which belong to the sub-family 3 are at present exclusively endo-β1,6-glucanases. Reported activity of GH3 glycoside hydrolases includes β-glucosidase activity, suggesting that these enzymes target β-linked glucans and that PUL_{btPUS} may orchestrate β-glucan deconstruction, specifically β1,6-glucans.

5.2 Objectives

The objectives of this chapter is to elucidate the degradation of β1,6-glucans by *Bacteroides* through the biochemical characterisation of the proteins encoded by the PULs believed to target this substrate.

5.3 *Bacteroides thetaiotaomicron* and *Bacteroides ovatus* are capable of utilizing β1,6-glucans as a carbon source.

5.3.1. Culturing *B. theta and B. ovatus* with β-glucans.

To establish the yeast β-glucan(s) utilized by *B. theta* and *B. ovatus,* the two Bacteroides were cultured anaerobically in the presence of the linear β1,3-linked glucan, laminarin and pustulan, a linear β1,6-glucan polysaccharide. Growth curves are displayed in Figure 5.3. Both *Bacteroides* species were able to utilize pustulan as an energy source, growing to an optical density of 2.0 (OD_{600nm}). No growth was observed when *B. theta* and *B. ovatus* were incubated in minimal media containing

laminarin for 24 h period indicating that the two organisms were unable to target β1,3-glucans.

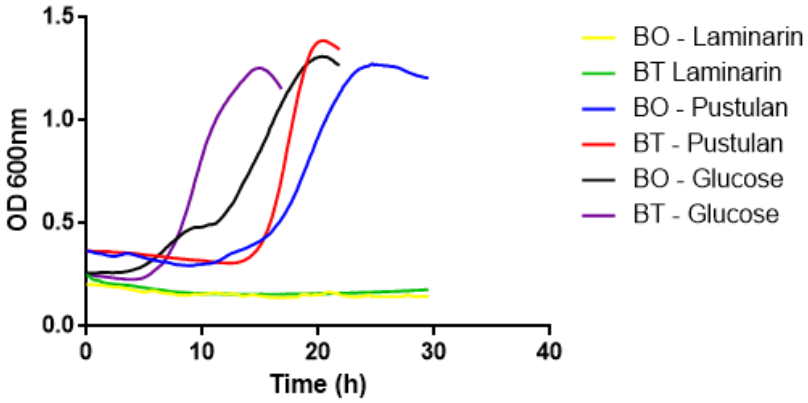


Figure 5.3. Growth curves of anaerobic *B. theta* and *B. ovatus* cultured on glucose and β-glucans. Minimal media (2 mL) containing glucose (5 mg/mL), the linear β1,3-glucan laminarin (5 mg/mL) or the linear β1,6-glucan, pustulan (5 mg/mL), was inoculated with *B. theta (*BT) or *B. ovatus* (BO)). Inoculated bacteria were incubated anaerobically in a 24-well plate. Bacterial growth was continuously monitored (represented by OD_{600nm}), until stationary phase was reached. Bacterial growths were performed in triplicate, the average optical density is plotted on GraphPad Prism 5.0.

5.3.2. Transcription of PUL_{btPUS} and PUL_{boPUS} is upregulated by β1,6-glucans.

Given that *B. theta* can grow on yeast β1,6-glucans and the genes encoded by PUL_{btPUS} are predicted to encode PUL_{PUS}, it is likely that transcription of the locus is activated by pustulan. As described below, *B. ovatus contains a PUL (*PUL_{boPUS}) that is synthenic with PUL_{boPUS} and thus is predicted to also orchestrates the degradation of pustulan. To test the hypothesis that PUL_{btPUS} and PUL_{boPUS} were upregulated by pustulan, confirming their role in the degradation of the β1,6-glucan, quantitative polymerase chain reactions (qPCRs) were performed on the transcripts derived from the genes encoded by the two loci. To ensure that no observed activation of the PULs

was due to unspecified components in TYG medium, *Bacteroides* cells (stored in glycerol (25 % v/v)/ TYG) were sub cultured twice over night into minimal media containing glucose (5 mg/mL), before being washed and inoculated into minimal media (3 mL) containing glucose (5 mg/mL) and pustulan (5 mg/mL), respectively. Cells were grown anaerobically until early-exponential phase (0.6 OD_{600nm}) and RNA was extracted and reverse transcribed to cDNA as described in Materials and Methods, Sections 2.1.14 and 2.1.15, respectively. The primers, complementary to regions of the ORFs and designed to produce amplicons between 100 – 150 base pairs are displayed in Table 5.1. qPCR revealed that, compared to glucose grown cells, pustulan mediated a ~100-fold upregulation of *bt3310 – bt3313* in *B. theta* and ~200-fold up-regulation of *bacova_00942 – 00945* in *B.ovatus*, Figure 5.4. No upregulation of *bt3309* and *bacova_00941*, the $SusR_h$ regulators, was observed as seen in previous gene upregulation studies of PULs (Sonnenburg et al., 2010). Interestingly, the ORFs encoding for BT3314 and BACOVA_00946, the hypothetical GH3 β-glucosidases, were only upregulated ~10- and ~25-fold in *B. theta* and *B. ovatus* respectively (relative to glucose grown cells) (Figure 5.4). The exact signalling molecule that induce the up-regulation of the β-glucan PUL in both species in not known.

***Bacteroides* species**	**Open Reading Frame**	**Nucleotide sequence (5' -> 3')**	**Direction**
Bacteroides thetaiotaomicron	*bt3309*	TGCCTCTTCGGCTTCTATTT	Forward
		ATGGAATTAGGAGCCACGTC	Reverse
	bt3310	ACGTGTTGCTTACCTTGCTG	Forward
		GTCGGAGGCTTCACCATACT	Reverse
	bt3311	TCCGTAGCCCGTACTCTTCT	Forward
		AAACCATCAGCCTTCACCTC	Reverse
	bt3312	ATGGAAGAGGTAGCATTGGG	Forward
		TGACAACCACCTTCACGATT	Reverse
	bt3313	GGCTGGGAAGACAATACGAT	Forward
		CGGACTCGCTTCATAGTTGA	Reverse
	bt3314	GAGTTCAACCGTCACACCAC	Forward
		ACCGGAGTCGGTATCTTCAC	Reverse
Bacteroides ovatus	*bacova_00941*	TATCTGCGACTCTGCAATCC	Forward
		CGTGTACATTCCCGTTGAAG	Reverse
	bacova_00942	GGTAACCAGTCGGGTATTGG	Forward
		AGCATTTGCGTAATCAGTCG	Reverse
	bacova_00943	GGACTGGAACCGGTATATGG	Forward
		CATGGCACGGAATATCTTTG	Reverse
	bacova_00944	GGCGCCTACTTCCATTACAT	Forward
		TCAGAGAACGTTTCCGTCAG	Reverse
	bacova_00945	ATCTGGGCTATTGGAACAGG	Forward
		CCGGCTATGAATGTCAGTTG	Reverse
	bacova_00946	AGTGGCTAAAGATGCCGACT	Forward
		TGCCAGTTCGCTTATCACTC	Reverse

Table 5.1. Nucleotide sequences of primer utilized in quantitative PCR experiment to assess gene expression levels of the genes in PUL_{btPUS} and PUL_{boPUS}

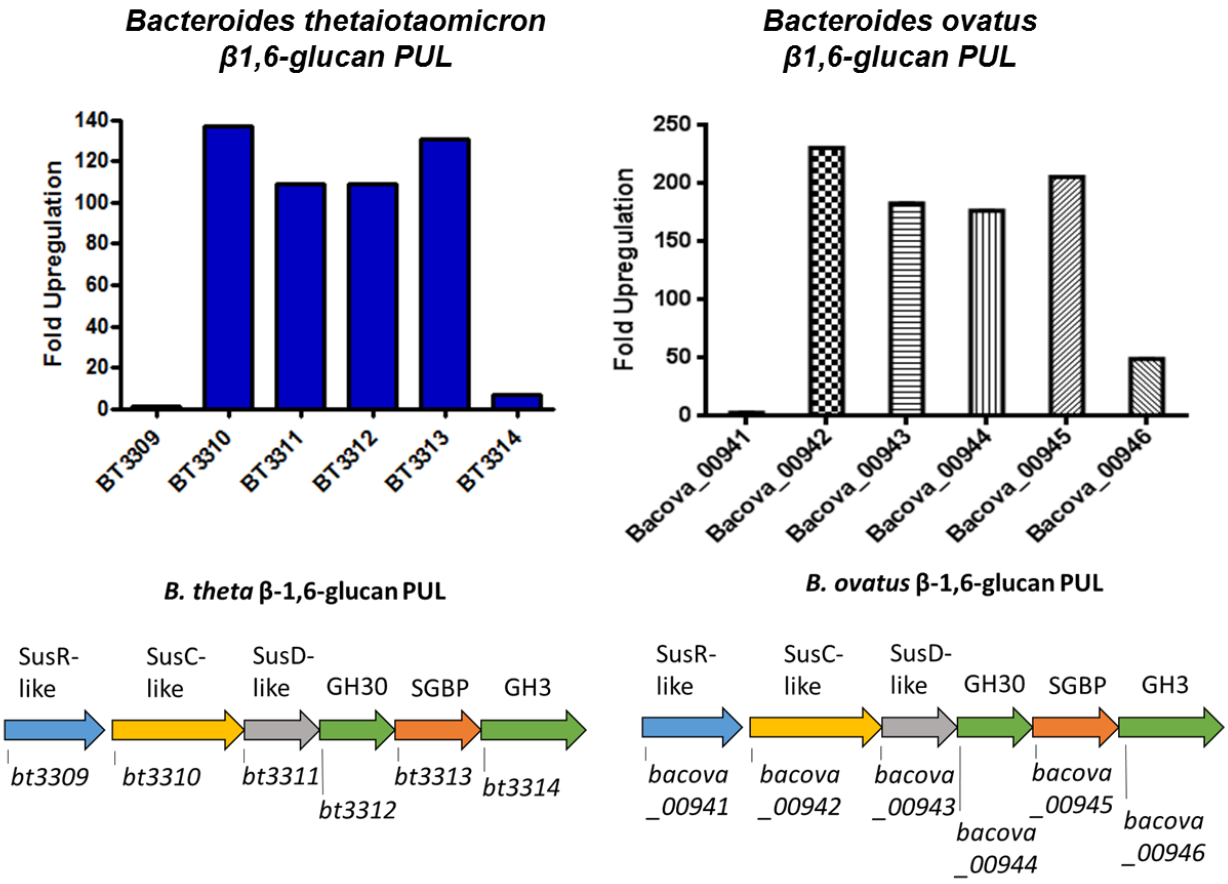


Figure 5.4. Quantitative PCR of the transcripts derived from PUL_{btPUS} and PUL_{boPUS} when *B. theta* and *B. ovatus* were cultured on pustulan and glucose. Anaerobic cultures in minimal media (3 mL) containing the respective carbon source (5 mg/mL) were grown until 0.6 OD_{600nm} before mRNA was harvested. Growths were performed in triplicate. The fold up-regulation compared to glucose grown cells of each gene is labelled with the respective encoded labelled according to predicted function.

5.3.3. PUL_{btPUS} is wide-spread amongst the Bacteroides genus.

Analysis of the genomes of Bacteroides revealed that PUL_{btPUS} is conserved amongst all other *B. theta* strains currently sequenced from clinical settings, and numerous other Bacteroides species including *B. ovatus* strain ATCC-8483, where the locus is defined as PUL_{boPUS}. Thus, PUL_{btPUS} and PUL_{boPUS} are syntenic, while the corresponding encoded proteins display a high degree of sequence identity and are thus highly likely to display the same function (Figure 5.5).

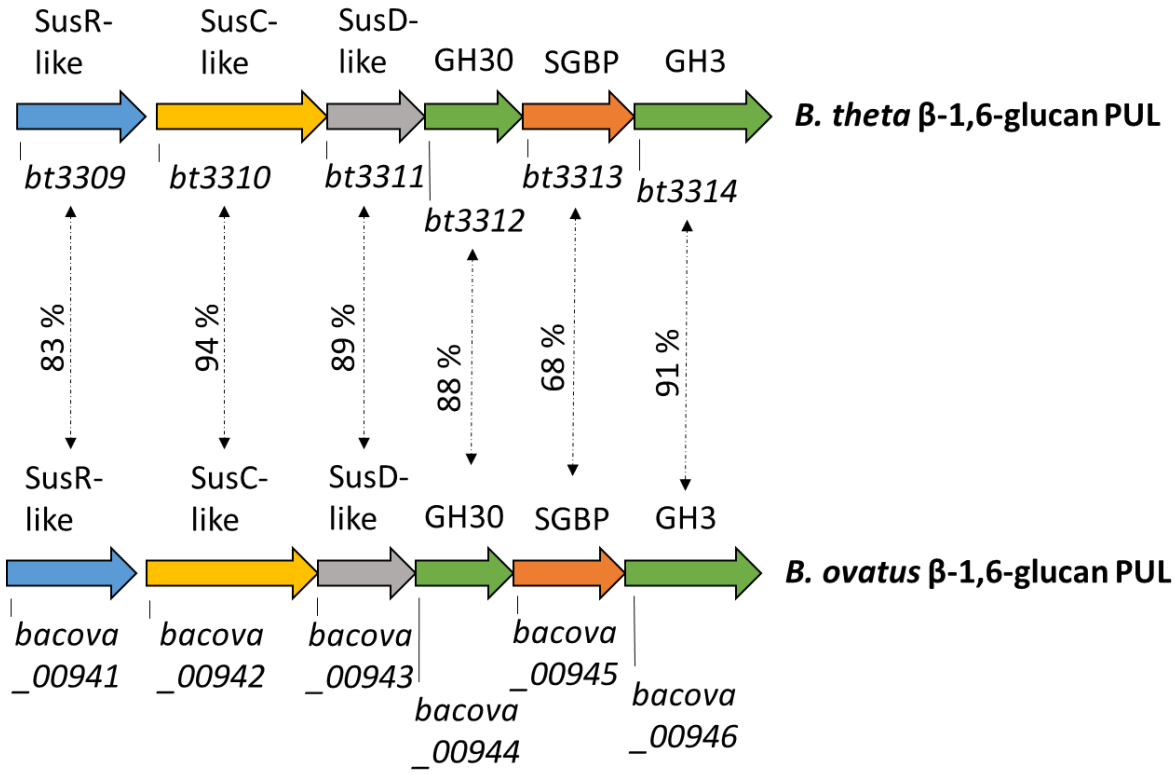


Figure 5.5. Schematic representation of the synteny and homology shared between PUL_{btPUS} and PUL_{boPUS}. The two PULs are labelled by *Bacteroide*s species. The locus tag of the predicted protein encoding ORFs are displayed beneath their respective representation (*bt* for *B. theta* and *bacova* for *B. ovatus*). The amino-acid sequence identity shared between the encoded hypothetical proteins with the same predicted function (denoted by a dashed line) is displayed as a percentage.

5.4. Characterisation of the proteins encoded by PUL_{btPUS}

The mechanism of β1,6-glucan utilization by *Bacteroides* was initially explored in *B. theta* and accomplished through characterising the proteins encoded by PUL_{btPUS}.

5.4.1. Characterisation of the extracellular binding proteins, $SusD_h$ and SGBP encoded by PUL_{btPUS}

5.4.1.1. Expression and purification

DNA encoding $SusD_h$ and extracellular SGBP, BT3311 and BT3313 respectively, were cloned via a high-throughput cloning service (NZYtech, Portugal). The proteins (like all *B. theta* proteins investigated in this study), were cloned without signal peptides. The genes were cloned into a pET28a derivative and supplied an N-terminal His_6-tag to BT3311 and BT3313. The two proteins, which were expressed in *E. coli* TUNER cells were produced in soluble form and purified by IMAC, Figure 5.6. The observed size of the purified proteins were consistent with their predicted size.

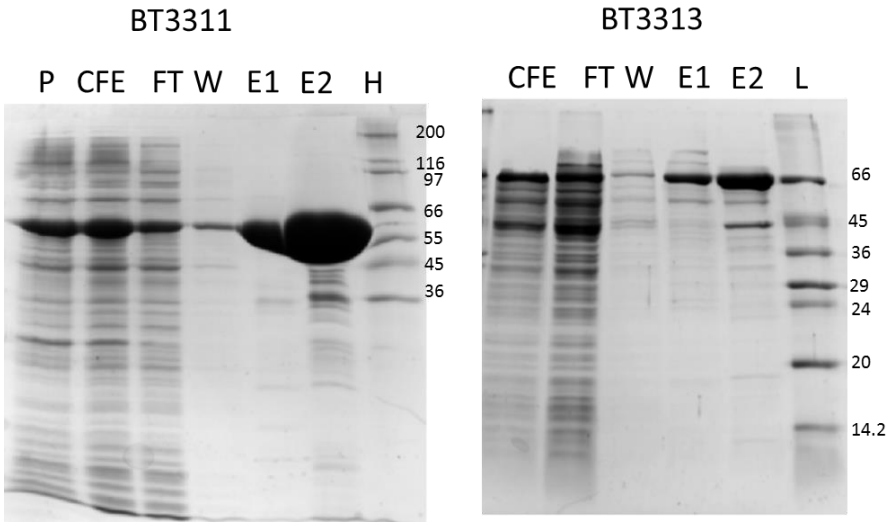


Figure 5.6. SDS-PAGE of recombinant expression and purification of BT3311 and BT3313. SDS-PAGE of typical BT3311 and BT3313 purification following IMAC. All SDS-PAGE gels were 12 % (w/v). Marker M_r is displayed in kDa. L = low size marker; P = insoluble fraction; CHE = cell free extract; FT = flow through; W = wash (1x TALON buffer); E1 = elution 1 (5 mM imidazole,1 x TALON buffer); E2 = elution 2 (100 mM imidazole, 1 x TALON buffer). BT3311 ran with an apparent M_r of 58 kDa. BT3313 ran with an apparent M_r of 70 kDa. Both values are consistent with the predicted sizes of the recombinant proteins, at 55.5 kDa and 72 kDa respectively.

5.4.1.2 The Extracellular binding proteins demonstrate preferential binding to Pustulan, a linear β1,6-glucan polysaccharide.

The activity of the potential carbohydrate surface binding proteins to soluble polysaccharides was first assessed via affinity gel electrophoresis as a qualitative screening method, which is described in Materials and Methods, Section 2.2.5.2. The non-binding negative control protein was bovine serum albumin (BSA). Both BT3311 and BT3313 displayed binding to pustulan (Figure 5.7). No binding to the linear β1,3 linked glucan laminarin was observed.

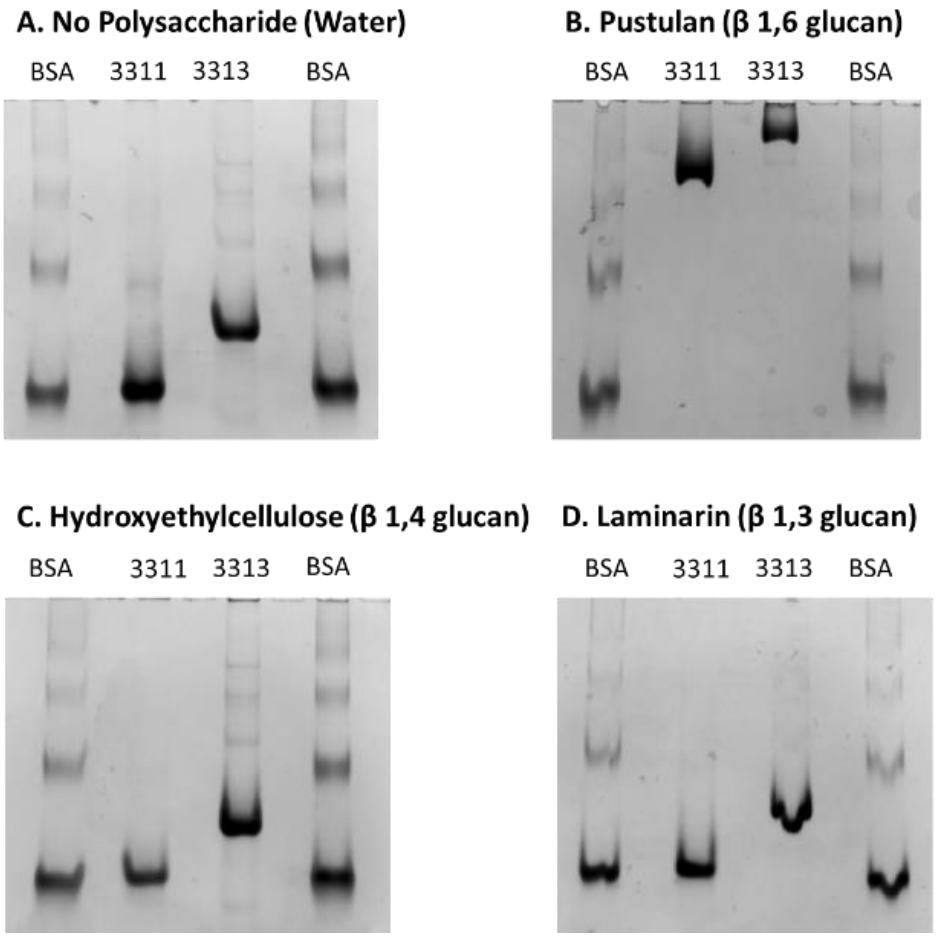


Figure 5.7. Affinity gel electrophoresis of BT3311 and BT3313 against soluble β-glucans. Recombinant BT3311 ($SusD_h$) and BT3313 (SGBP) were electrophoresed through a non-denaturing, polyacrylamide gel in the presence and absence of ligand (0.1 % (w/v). Bovine serum albumin (BSA) was used as a non-binding control.

To provide a quantitative assessment of glucan recognition, isothermal titration calorimetry (ITC) was utilized to determine the thermodynamic parameters of ligand binding. Using ITC the association constant (K_a) change in enthalpy and stoichiometry of binding (n) were determined using MicroCal Origin 7.0 software, as described in Materials and Methods, Section 2.2.5.1 The data-set is reported in Table 5.2 and example titrations are shown in Figure 5.8. The $SusD_h$ (BT3311) and SGBP (BT3313) displayed K_a values of 4.0 x 10^4 M^{-1} and 5.1 x 10^5 M^{-1}, respectively, for pustulan.

Protein	**Ligand**	**K_a (M^{-1})**	**n**
BT3311	Pustulan	4.0 (±0.25) x 10^4	1.1 ± 0.0
	β1,6-glucoheptaose	1.8 (±0.03) x 10^4	1.0 ± 0.0
	β1,6-glucohexaose	1.6 (±0.02) x 10^4	1.0 ± 0.0
	β1,6-glucopentaose	Unquantifiable Binding	
	β1,6-glucotetraose		
	β1,6-glucobiose	No Binding	
BT3313	Pustulan	5.1 (±0.4) x 10^5	1.0 ± 0.0
	β1,6-glucoheptaose	6.0 (±0.04) x 10^4	1.0 ± 0.0
	β1,6-glucohexaose	5.8 (±0.06) x 10^4	1.0 ± 0.0
	β1,6-glucopentaose	Unquantifiable Binding	
	β1,6-glucotetraose		
	β1,6-glucobiose	No Binding	

Table 5.2. Association constant values (K_a) generated by isothermal titration calorimetry for BT3311 and BT3313 against pustulan and β1,6-oligosaccharides

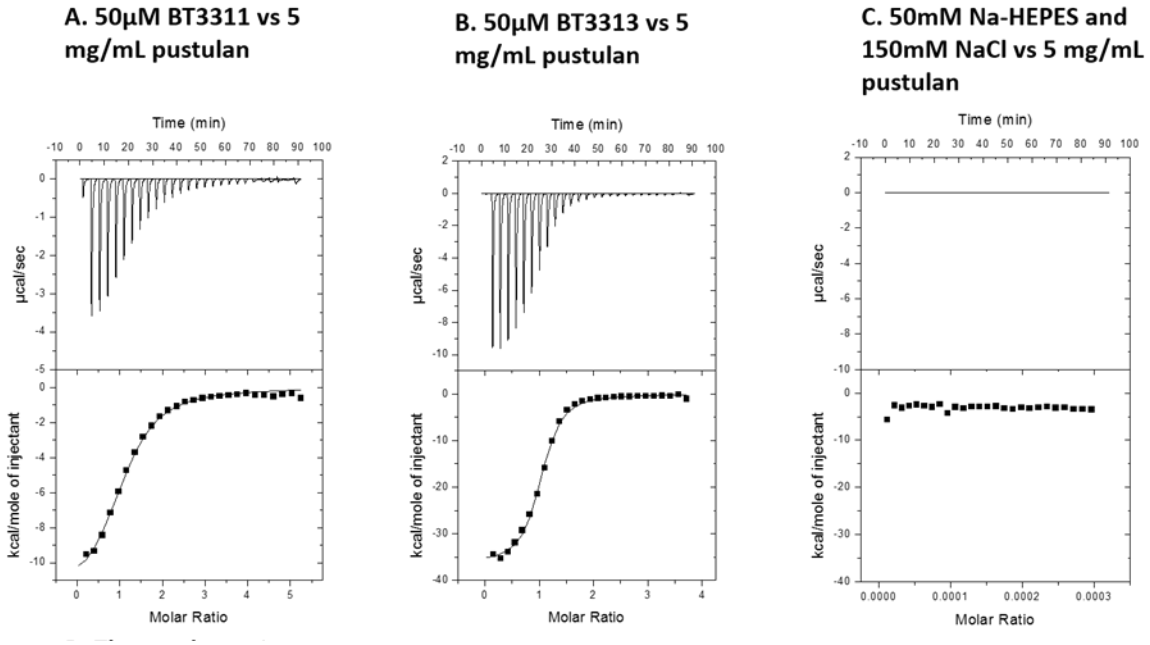


Figure 5.8. Example ITC data of BT3311 and BT3313 binding to pustulan. Panel A and B show an ITC trace of BT3311 and BT3313, respectively, binding to pustulan. Panel C demonstrates a control titration of ligand into buffer. Ligand (5 mg/mL) in the syringe was titrated into the protein containing cell (50 µM) or buffer for control titration. The top half of each ITC trace shows the raw heats generated per titration; the bottom half shows the integrated peak areas, fitted using a single-site binding model by MicroCal Origion software. All ITC was performed in 50 mM Na-HEPES buffer, 150 mM NaCl, pH 7.5

To deduce the ability of the surface binding proteins to bind to β1,6-gluco-oligosaccharides with differing degrees of polymerisation (D.Ps), the pustulan polysaccharide (1 g) was subjected to acid hydrolysis (20 mM HCl for 2 h) giving β1,6-oligosaccharides of varying D.Ps. Following neutralisation, the oligosaccharide mixture was subjected to size exclusion chromatography (see Section 2.2.6.2. for methodology), rendering purified oligosaccharides of single species ranging from glucobiose to glucooctaose (Shown in Figure 5.9). The size of the oligosaccharides was based on the assumption that successive eluted glycans differed by a D.P. of 1.

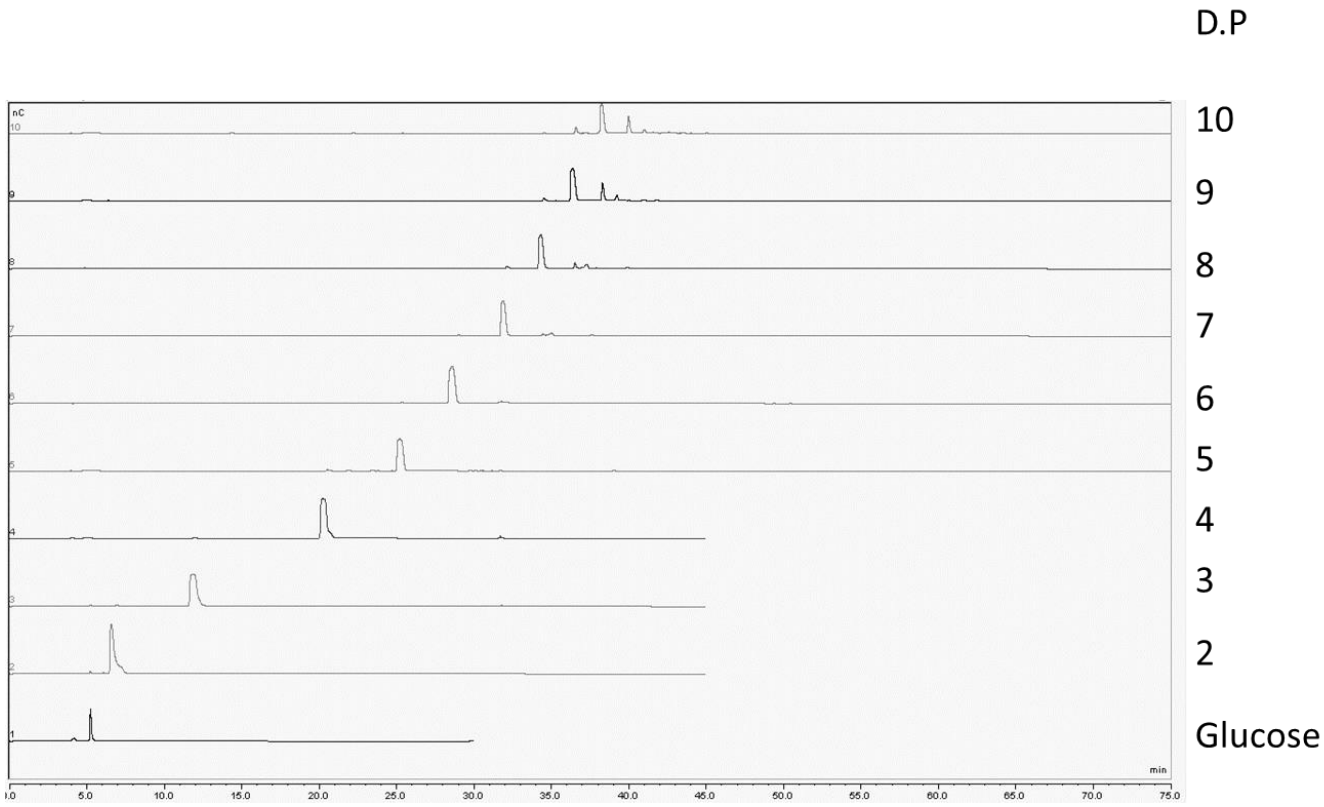


Figure 5.9. HPAEC-PAD analysis of purified β1,6-oligosaccharides. Oligosaccharide fractions rendered by P2-column size exclusion chromatography were first analysed for purity and degree of polymerisation (D.P) by HPAEC-PAD. Samples were diluted in water (100-fold dilution) for analysis. A glucose (0.05 mM) standard was ran concurrently to provide information on D.P of purified oligosaccharides relative to glucose. Fractions are labelled according to D.P (2-10). Once confirmed to be homogenous in composition, purified oligosaccharide fractions were pooled and subjected to lyophilisation.

Both proteins displayed maximum binding to oligosaccharides with a D.P. of 6 or 7. No measurable binding was detected for ligands with a D.P. ≤5. BT3313 displayed an affinity for glucohexaose and glucoheptaose around 3.5-fold greater than BT3311, Table 5.2.

5.4.2. Characterisation of the enzymes encoded by the β1,6-glucan PUL of *Bacteroides thetaiotaomicron*

5.4.2.1. Expression and Purification

The ORFs encoding mature BT3312 and BT3314 (lacking signal peptides) were cloned into a pET28a expression vector derivative as described in Materials and Methods, Section 2.1.21. The His-tagged proteins were purified by IMAC to electrophoretic homogeneity. The size of the purified proteins, as judged by SDS-PAGE, Figure 5.10 were consistent with their predicted molecular weight based on primary sequence. BT3312, a candidate for crystal trails, was subjected to further purification by size-exclusion chromatography

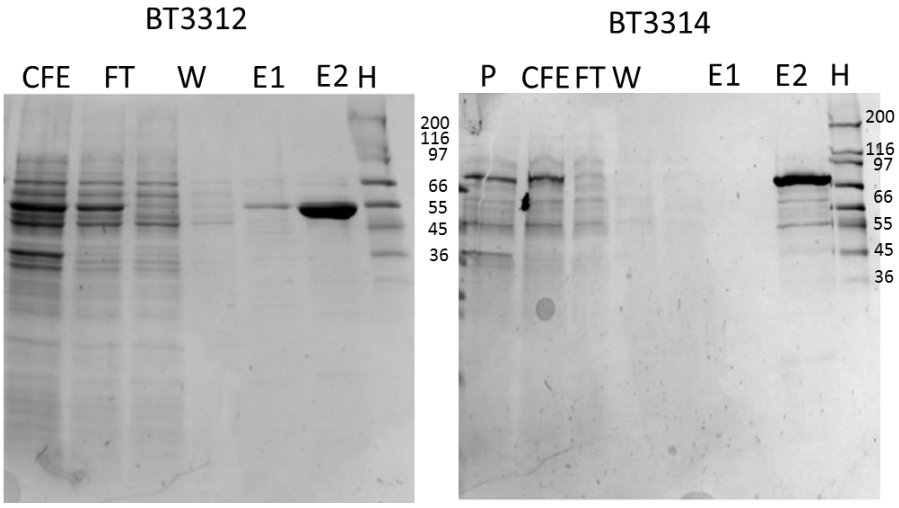


Figure 5.10. SDS-PAGE of recombinant expression and purification of BT3312 and BT3314. SDS-PAGE of typical BT3312 and BT3314 purification following IMAC. All SDS-PAGE gels were 12 % (w/v). Marker M_r is displayed in kDa. L = low size marker; P = insoluble fraction; CHE = cell free extract; FT = flow through; W = wash (1x TALON buffer); E1 = elution 1 (5 mM imidazole,1 x TALON buffer); E2 = elution 2 (100 mM imidazole, 1 x TALON buffer). BT3312 ran with an apparent M_r of 55 kDa. BT3314 ran with an apparent M_r of 80 kDa. Both values are consistent with the predicted sizes of the recombinant proteins, at 53 kDa and 82 kDa respectively.

5.4.2.2. Sequence analysis and prediction of the cellular location of, BT3312, a hypothetical GH30 endo-β1,6-glucanase.

BT3312 is a member of sub-family 3 of GH30. All previously characterized members of this sub-family have been shown to display β1,6-glucanase activity. LipoP (http://www.cbs.dtu.dk/services/LipoP/) analysis of BT3312 predicts that the protein is localised to the outer membrane of the cell due to the presence of a type II signal peptide (Figure 5.11).

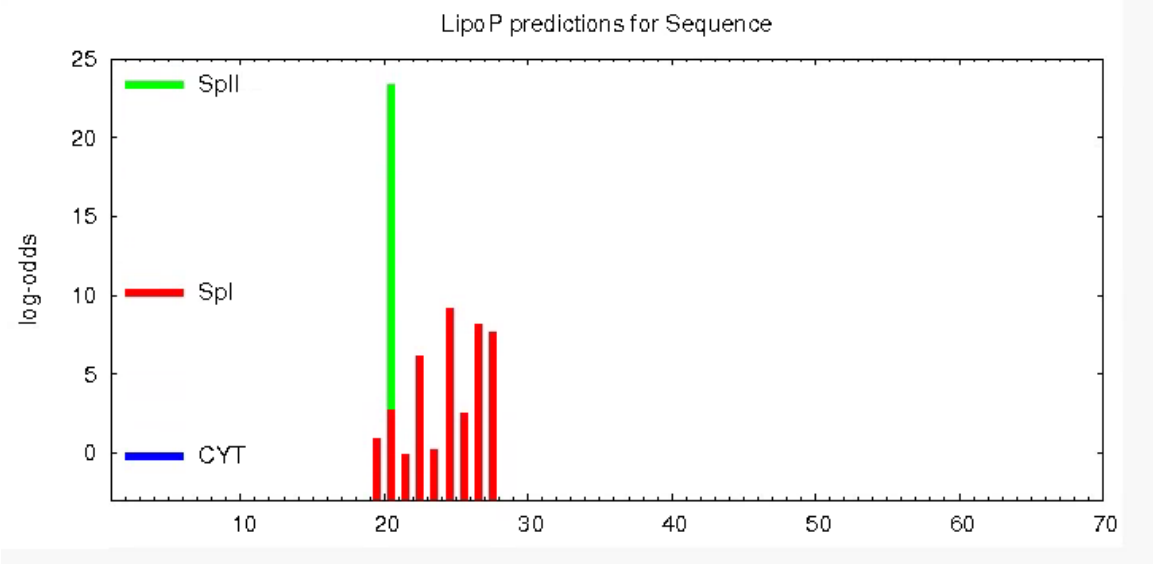


Figure 5.11. LipoP analysis of the predicted GH30 β1,6-glucananse. LipoP analysis (http://www.cbs.dtu.dk/services/LipoP/) indicates the likelihood of a SpII signal peptide (lipoprotein signal peptide) and a predicted cleavage site at amino acid residue 20 of the protein sequence. These data indicate BT3312 is localised to the outer membrane of the cell.

5.4.2.3. The GH30, BT3312, is an endo-β1,6-glucanase with activity against pustulan, a linear β1,6 glucan polysaccharide.

In order to investigate the specificity and activity of BT3312, the capacity of the enzyme to hydrolyse various glyosidic linkages was assessed. Qualitative assessments of activity were analysed by TLC. Michaelis-Menten kinetics was generated via reducing sugar assays (polysaccharide substrates) or by substrate depletion using HPLC (oligosaccharide substrates) as described in 2.2.1.4.1 and 2.2.4, respectively. Assays incorporating BT3312 were conducted at 37 ^{o}C in 20 mM sodium-phosphate buffer, pH 7.5. Rates were plotted using GraphPad Prism 5.0 and analysed via non-linear or linear regression where appropriate.

BT3312 was found to be catalytically active against the linear β1,6-linked glucan pustulan, rendering glucose and glucobiose as limit products of digestion (Figure 5.12), thereby confirming that BT3312 is a β1,6-glucanase. During the initial stages of the reaction a series of oligosaccharides were generated demonstrating that the

enzyme has an endo mode of action (Figure 5.13). No activity against the other polysaccharides assessed including β1,3-glucan, β1,4-glucan and β1,6-galactan was observed.

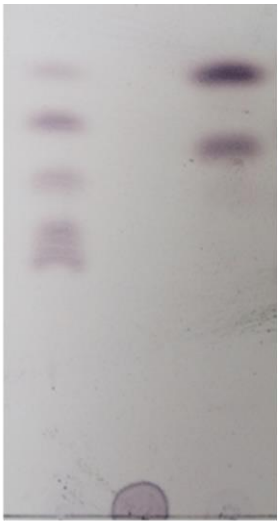


Figure 3.12. Thin layer chromatography (TLC) of limit products generated by BT3312 against pustulan. TLC showing a 16 h digest of BT3312 (1 µM) with pustulan (1 mg/mL), ran against a glucose and β1,4-glucooligosaccharide standard (G -> G4). NOE = no enzyme control.

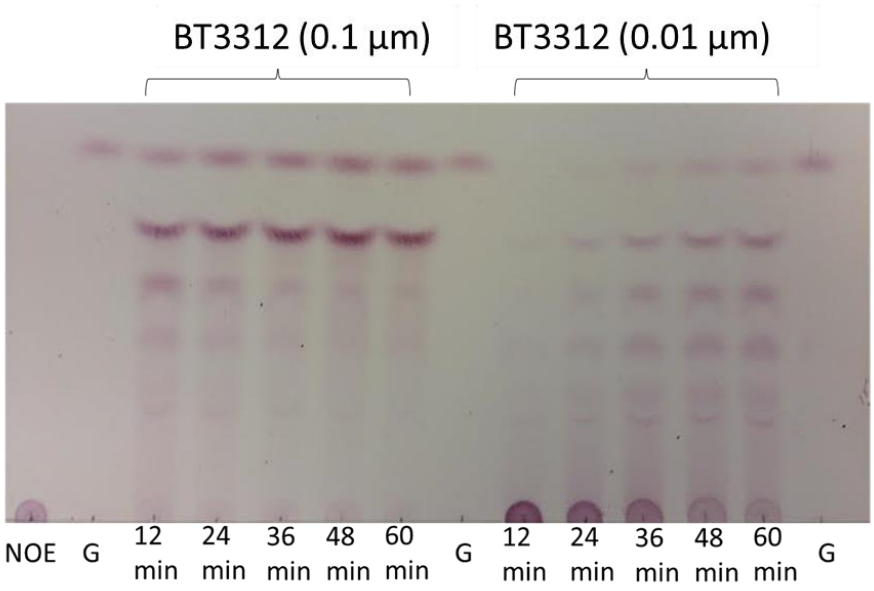


Figure 5.13. TLC analysis of GH30 endo-β1,6-glucananse ,BT3312, activity against pustulan. TLC showing BT3312 (0.1 µM or 0.01 µM) activity against pustulan (1 mg/mL) over a time course totalling 1 h. Samples of the reactions were taken at the respective time points and deactivated by boiling for 10 min. NOE = No enzyme control. G = glucose (1 mM) which was ran as a standard.

The full Michaelis-Menten kinetic parameters of BT3312 against pustulan are reported in Table 5.3 and Figure 5.14 and were determined via a reducing sugar assay as described in Materials and Methods, Section 2.2.1.4.1. Briefly, as the glycoside hydrolase cleaves sugar chains internally, new reducing ends are formed, inducing a colour change in the assay reagents which can be measured spectrophotometrically at $OD_{540nm.}$ A standard curve of glucose was utilized to quantify the formation of new reducing ends. Enzyme (50 nM) was incubated with a range of pustulan concentrations (0.5 – 8 mg/mL) with aliquots taken of the reaction over a series of time points. Reactions were deactivated by the addition DNSA reagent. The initial linear rate was plotted against substrate concentration. BT3312 had a k_{cat} of 13824 min^{-1} against the polymeric substrate and a K_m of 5.6 mg/mL for the polysaccharide.

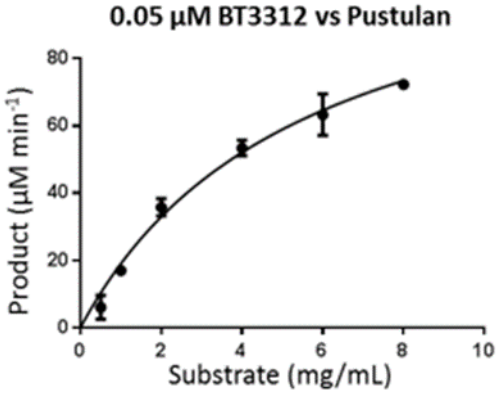


Figure 5.14. Activity of BT3312 against pustulan. Rates were generated via a DNSA reducing sugar assay using a fixed enzyme concentration (0.05 µM) against a range of substrate concentrations Rates were plotted using GraphPad Prism 5.0 and analysed using Michaelis-Menten kinetics.

Enzyme	Substrate	K_m (mg/mL)	V_{max} (µM min^{-1})	K_{cat} (min^{-1})	K_{cat}/K_m (mg^{-1} mL min^{-1})
BT3312	Pustulan	5.6	69.1	1382	247

Table 5.3. Kinetic parameters generated from BT3312 VS Pustulan

5.4.2.4. BT3312 activity vs pustulan-derived oligosaccharides.

To ascertain the activity of BT3312 against the oligosaccharide β1,6-glucooctose, BT3312 (0.01 µM) was incubated with β1,6-glucooctaose (0.03 mM) to ≤ 30 % of the starting quantity of substrate (represented by peak area) at T_0 (0 min) after a 25 min time course. Aliquots of the reactions were taken at 5 min time points (T_{1-6}) and analysed using HPAEC-PAD to deduce peak area. BSA (0.1 mg/mL) and an internal standard of fucose (0.01 mM) were incorporated into all assays. Substrate depletion was plotted using Graphpad Prism 5.0 and the gradient (velocity (min)) was ascertained via linear regression analysis. The K_{cat}/K_m (M^{-1} min^{-1}) of β1,6-glucooctose hydrolysis by BT3312 was subsequently deduced as performed in Chapter 4, Section 4.3.2.1.4.. BT3312 demonstrated a k_{cat}/K_m value of 6.4 x 10^6 M^{-1} min^{-1} against β1,6-glucooctaose. The graph used to generate the catalytic efficiency of the enzyme is depicted in Figure 5.15.

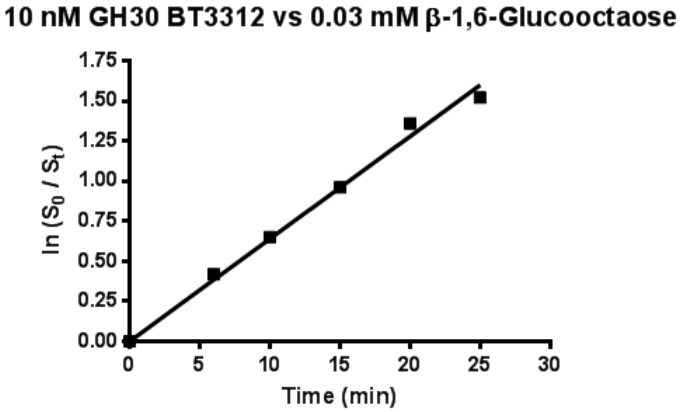


Figure 5.15. Graph showing substrate depletion of β1,6-glucooctaose by BT3312. BT3312 (10 nM) was used to treat β1,6-glucooctaose (0.03 mM). Aliquots were taken at 5 min intervals over a 25 min time-course. Reactions were deactivated by boiling for 10 min. The area of the glucooctasoe peak was determined at the respective time point and standardised to an internal standard of fucose. $\ln(S_0/S_t)$ values were plotted using GraphPad Prism 5.0, where S_0 = peak area of glucooctaose at T_0 (0 min) and S_t =peak area of glucooctaose at time point (X min). Linear regression was used to ascertain the gradient (rate) which was subsequently divided by enzyme concentration to derive the K_{cat}/K_m efficiency constant.

Enzyme	Oligosaccharide	K_{cat}/K_m (M^{-1} min^{-1})
BT3312	β1,6-glucooctaose	6.39 (± 0.2) x 10^6

Table 5.4.Substrate depletion data of BT3312 vs β1,6-glucooctaose

5.4.2.5. Crystallisation of BT3312.

Purified protein (20 mg/mL) was distributed into 96 well plates by a Mosquito microliter pipetting robot (TTP labtech) with a total of two drops per well; the ratio of protein mixture to reservoir, being 1:1 and 2:1 respectively. Eight commercially available 96-well screens were used. Crystals of BT3312 were obtained in the commercial 96-well screen PACT, the conditions being 0.2 M Sodium Bromide, 0.1 Bis Tris Propane, pH 6.5, 20 % w/v PEG 3350. The structure of the GH30 enzyme was solved by molecular replacement by Arnaud Baslé using the GH30 human glucosylceramidase (displays

26% identity over 86% of sequence with BT3312) as the search model. Subsequent to the PhD project crystals (grown by Dr Elisabeth Lowe) were transferred to a 10 µl drop consisting of 7 µl reservoir solution, 2 µl 20% PEG 400 and 1 µl glucose-β1,6-deoxynijiromycin (kindly provided by Dr Spencer Williams, University of Melbourne) and allowed to soak for 5 min prior to freezing. Data was collected at Diamond Light Source, on beamline IO3 and the inhibitor-bound structure was solved by Dr Arnaud Baslé, as described before. Data collection and refinement statistics are displayed in Appendix A.

5.4.2.6. Analysis of the apo-structure of BT3312, a GH30 encoded by B. theta

BT3312 comprises two domains. The catalytic domain displays a $(\beta/\alpha)_8$ barrel fold or TIM barrel, extending from residues Asp82 to Lys427, and a β-sandwich domain comprising sequences from both the N- and C-termini. The TIM barrel contains a central eight stranded β-barrel (strands 1 and 5 were not represented in the β-strand schematic using the pymol programme). Extending from each β-strand is an α-helix. In general extended loops link the α-helices with the β-barrel, although a small β-sheet comprising three anti-parallel strands is in the loop connecting β-strand-8 and α-helix-8. The β-sandwich domain contains seven antiparallel β-strands in one sheet, two contributed by the N-terminal region of the enzyme. The other β-sheet comprises three antiparallel strands with one derived from the N-terminus. As BT3312 is located in GH30, the enzyme is a member of clan GH-A in which the fold and catalytic apparatus and mechanism are conserved. According to these criteria BT3312 is predicted to display a double displacement acid-base assisted mechanism in which anomeric configuration is retained after bond cleavage. Similarly the catalytic acid/base and catalytic nucleophile are predicted to be in β-strands 4 and 7 of the barrel, respectively, and highly conserved within the clan. Thus, the candidate catalytic residues are

Glu236 (acid-base) and Glu337 (nucleophile). The catalytic roles of Glu236 confirmed by the observation that the mutant E236A rendered the enzyme completely inactive.

Consistent with other clan GH-A glycoside hydrolases the substrate binding is predicted to be positioned on top of the β-barrel. Indeed, inspection of this region of the enzyme reveals a deep U-shaped pocket that houses the catalytic residues, Figure 5.16**.** The active site pocket of BT3312 is very different to the topology of the substrate binding regions of other GH-A endo-acting enzymes, which display a linear or curved cleft that is open at both ends. The architecture of the substrate binding region of BT3312 is adapted to the structure of β1,6 glucan. NMR analysis of this short glucan revealed a hooked, u-shaped configuration (Lowman et al., 2011) that is entirely complementary to the typology of the substrate binding region of BT3312 (see below).

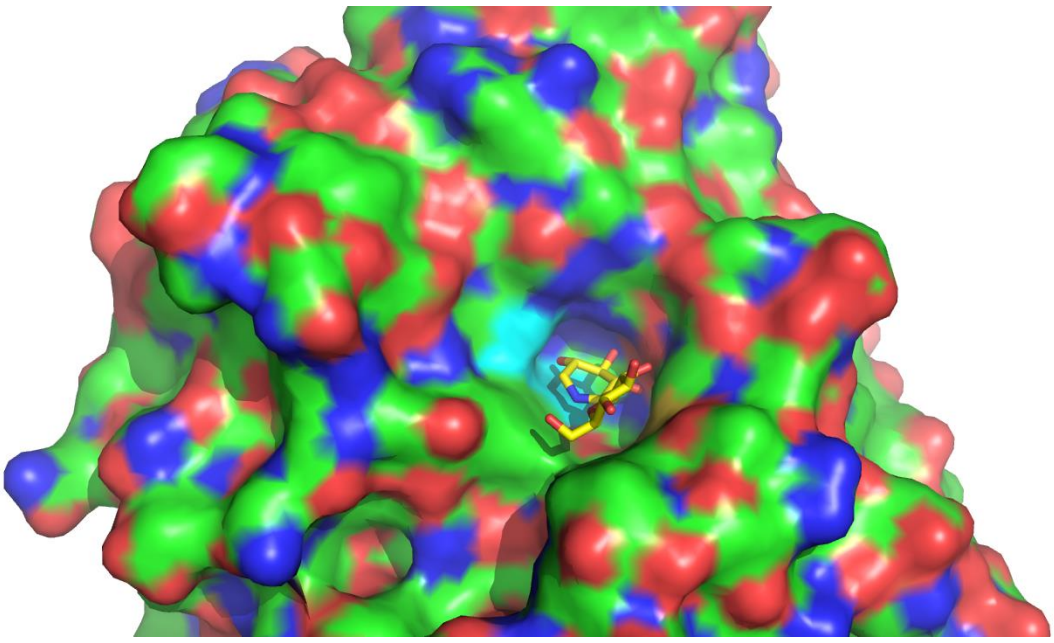


Figure 5.16. The pocket topology of the GH30, BT3312 active site. The surface of the enzyme is shown. The catalytic residues of the enzyme, residing in the pocket-like active site of the enzyme are coloured light blue, The bound inhibitor, glucose-β1,6-deoxynijiromycin, is coloured yellow.

5.4.2.7. Analysis of the BT3312-inhibitor complex.

The apo structure of BT3312 bound to the transition state mimic glucose-β1,6-deoxynijiromycin revealed details of the active site and the distal substrate binding subsites. The deoxynijiromycin bound in the active site (-1 subsite) adopts an ^{4}E (*E* refers to envelope) conformation, Figure 5.17. While this may comprise the transition state geometry, most *β-gluco*-configured substrates adopt a $^{3}H_{4}$ conformation at the transition state. Given that ^{4}E and $^{3}H_{4}$ (*H* refers to a half-chair geometry) are adjacent to each other in the reaction coordinate pathways, and nitrogen in the deoxynijiromycin ring may skew the conformation adopted by the molecule in active site, it is likely that $^{3}H_{4}$ is the transition state conformation during β1,6-glucan hydrolysis by BT3312.

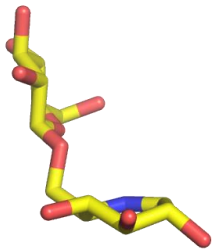


Figure 5.17. The 'envelope' conformation adopted by the transition state mimic, glucose-β1,6-deoxynijiromycni, when in complex with BT3312. The deoxynijiromycin bound in the active site (-1 subsite) adopts an ^{4}E (*E* refers to envelope) conformation. The endocyclic nitrogen of the molecule is coloured blue, whilst oxygen atoms are coloured red. Carbons are coloured yellow

The polar interactions between the active site of BT3312 and deoxynijiromycin are displayed in Figure 5.18. The carboxylate of Glu337 and Nδ2 of Asn235 made polar contacts with O2. O3 interacted with Nε1 of Trp177, Nδ2 of Asn242 and the carboxylate of Asp129. The Oδ2 of Asp129 and Nε1 of Trp375 made hydrogen bonds with O4. The polar interactions were completed by the endocyclic nitrogen donating a hydrogen bond to Oδ2 of Glu337. Binding of deoxynijiromycin in the active site was

also mediated through apolar contacts between the sugar ring and Trp375, which provides the hydrophobic platform in the -1 subsite. The catalytic nucleophile is ~3 Å from the anomeric carbon of deoxynijiromycin and is thus in an optimum position to mount a nucleophilic attack on the chiral centre. Glu236 is in an optimal position to donate a proton to the glycosidic oxygen appended to C1, and thus is able to function as the acid-base catalyst.

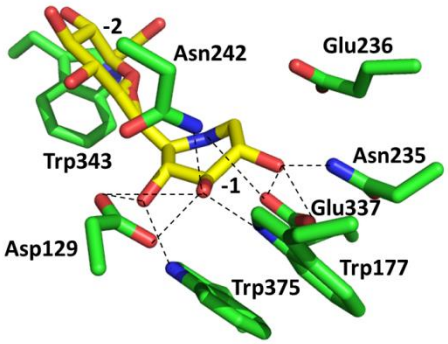


Figure 5.18. The polar interactions which coordinate the inhibitor complex in the BT3312 active site. Amino acids which interact with the deoxynijiromycin inhibitor are coloured green and labelled accordingly. The inhibitor is labelled as in Figure 5.20. The catalytic subsite in which the respective cyclic ring of the inhibitor binds is also labelled accordingly. Hydrogen bonds are denoted by dotted lines.

The -2 subsite makes only apolar interactions with the glucose at the non-reducing end of glucose-β1,6-deoxynijiromycin. Trp343 stacks against the pyranose rings while the disulphide formed by Cys391 and Cys394 also makes apolar contacts with the -2 glucose. Given that O6 of the non-reducing glucose is pointing into solvent it is unlikely that there are additional negative subsites in BT3312. It is difficult to establish the number of positive subsites without bound ligand in this region of the enzyme. Based

on the height of the wall of the substrate binding cleft, however, there are likely to be ~2 positive sites in which Glu236, His279, Asn280 and Asp284 contribute to substrate binding.

BT3312 is the first structure of GH30 subfamily 3. Comparison of the structure of BT3312 with GH30 enzymes with different specificities sheds light on the structural basis for substrate specificity, which is dominated by the topology of the substrate binding regions of the respective proteins. Thus, when BT3312 is compared with the glucuronoxylanase from *Dickeya chrysanthemi* D1 (PDB code 2y24), The *B. theta* enzyme contains a large loop extending from residues 379-396 that sterically occludes the -2 and -3 subsites in the glucuronoxylanase, Figure 5.19A,B. This loop is stabilised by a disulfide bond between Cys391 and Cys394, and three short beta strands behind the loop (coloured orange in Figure 5.19). BT3312 contains a second extended loop comprising Gln234 to Trp250 that prevents access to the positive subsites in the glucuronoxylanase, Figure 5.20A,B. Indeed the lack of residues that target O6 of substrate bound at -1 or -2 reinforces the view that it is the topology of the proximal substrate region of BT3312 that confers specificity for β1,6-glucan over β1,4-xylan. The xylose polymer, which lacks O6 moieties, adopts a linear three-fold screw axis conformation that is distinct from the highly curved, U-shaped structure exhibited by the β1,6-glucan. It should be emphasised that β1,3-glucans are highly curved and adopt a conformation that is similar to β1,6-glucans ref. The active site of BT3312, however, is not able to accommodate a β1,3-linked glucose as O6 would clash with the protein surface.

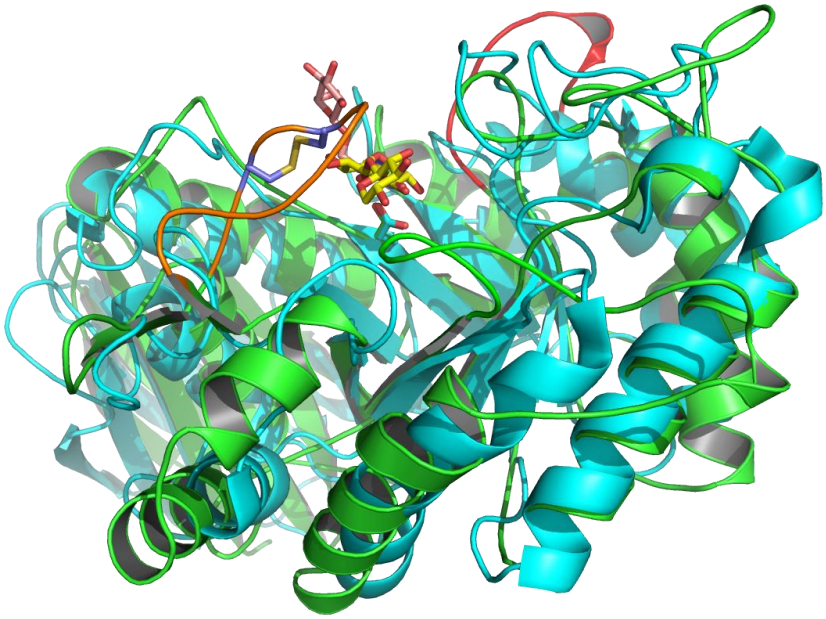


Figure 5.19A. Overlay of BT3312 with a glucuronoxylanase from *Dickeya chrysanthemi.* Both enzymes are depicted in schematic form. BT3312, the GH30 encoded by *B. theta* is shown in green. The Dickeya chrysanthemi glucuronoxylanase (PDP code: 2y24) is coloured cyan. Co-complexed inhibitor is shown in yellow. The extended loop of BT3312 which sterically occludes the -2 and -3 subsites in the glucuronoxylanase, is shown in orange

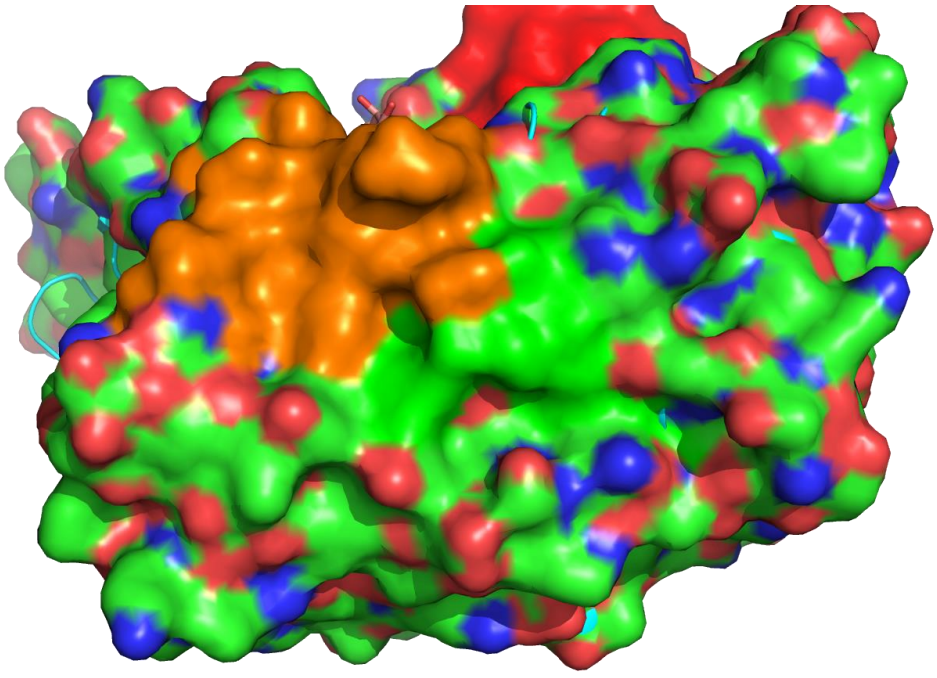


Figure 5.19B. Overlay of BT3312 with a glucuronoxylanase from *Dickeya chrysanthemi* - Surface. Depicted as described for Figure 5.19A, with the surface of the enzymes displayed.

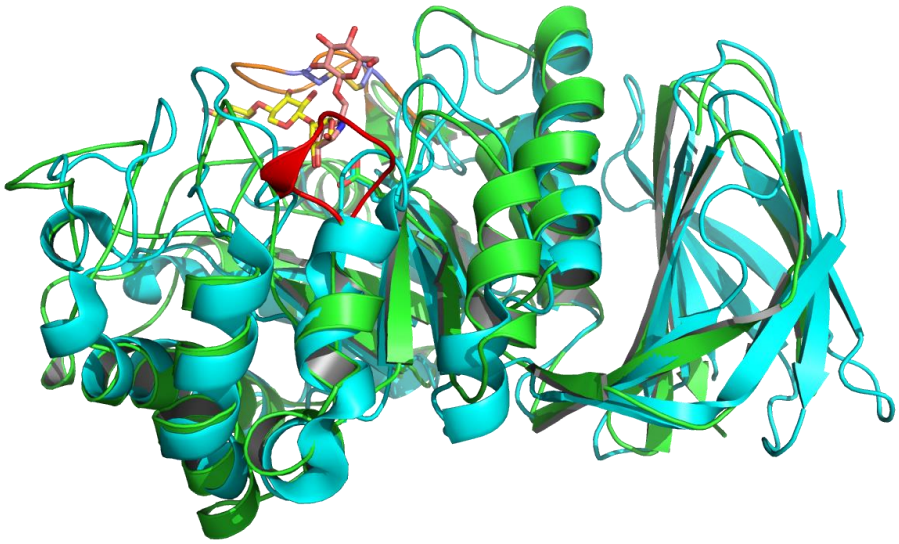


Figure 5.20A. Overlay of BT3312 with a glucuronoxylanase from *Dickeya chrysanthemi* showing the extended loop of BT3312. Proteins and bound inhibitor are represented as described in Figures 5.22. A second extended loop of BT3312 (shown in red), is depicted and is predicted to prevent access to the positive subsites in the glucuronoxylanase

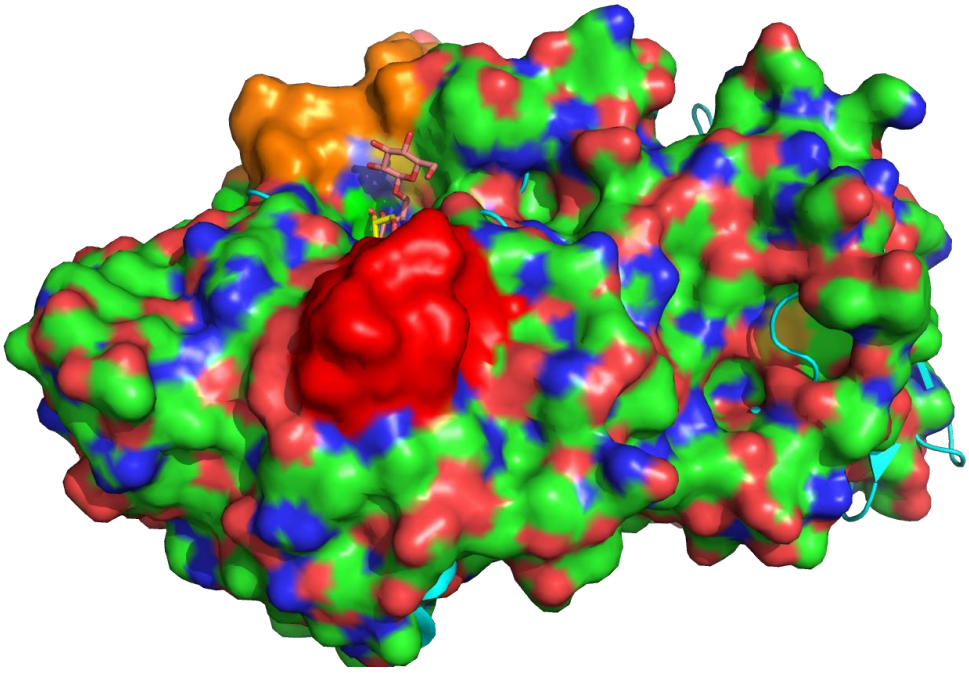


Figure 5.20B. Overlay of BT3312 with a glucuronoxylanase from *Dickeya chrysanthemi* showing the extended loop of BT3312 - Surface. Proteins and bound inhibitor are as represented as in Figures 5.23A with a surface representation. The second extended loop of BT3312 (shown in red), is predicted to prevent access to the positive subsites in the glucuronoxylanase

The active site (-1) subsite of the GH30 subfamily 3 enzyme BT3312 was compared with the homo-sapiens β-glucosidase (PDB code 2V3D) and the *D. chrysanthemi* glucuronoxylanase (PDB code 2Y24), representatives of subfamilies 1 and 8, respectively, Figure 5.21. In addition to the catalytic nucleophile and acid-base residues many of the other substrate binding residues were conserved in the three enzymes. Indeed the only two residues were not conserved. Asn142 in BT3312 was replaced in the other two enzymes by an aromatic residue that makes apolar contacts with the substrate. The loop containing Asp129 adopts a different position in the glucuronoxylanase (to accommodate the linear extended substrate), and thus there is no residue equivalent to the aspartate in this enzyme. The additional distal subsites in the glucuronoxylanase likely compensate for this loss of binding energy in the active site.

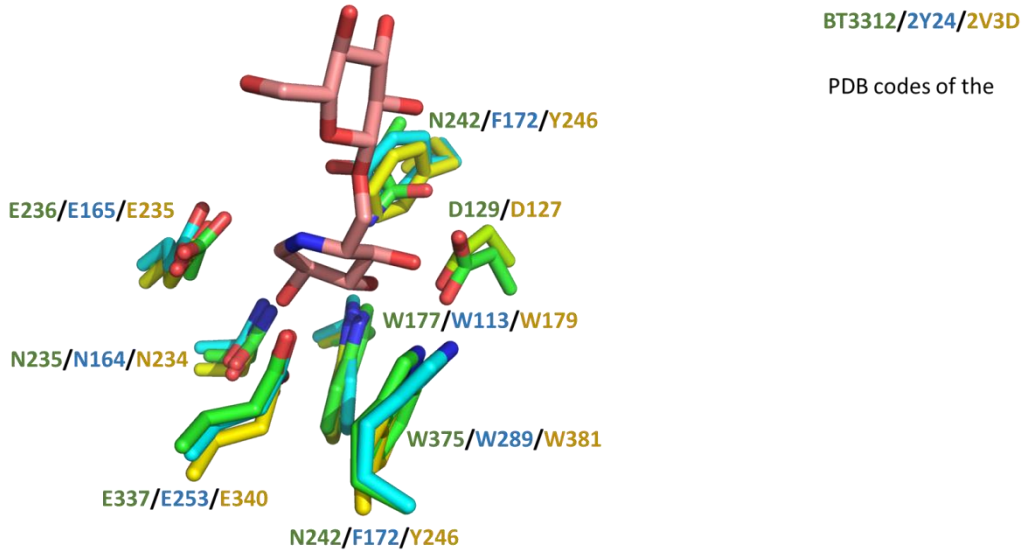


Figure 5.21. An overlay of the active site residues of GH30 BT3312, a *homo-sapiens* β-glucosidase and the *D. chrysanthemi* glucuronoxylanase. BT3312 amino acid residues are shown in green, the *D. chrysanthemi* glucuronoxylase (PDB code: 2y24) are depicted in blue and the amino acid residues of the homo-sapiens β-glucosidase are shown in yellow, respectively. Amino acid residues are labelled according and are believed to form interactons with the bound glucose-β1,6-deoxynijiromycin inhibitor (coloured pink).

5.4.2.8. Cellular location of, BT3314, a predicted GH3 exo-glucanase.

LipoP analysis suggests that the GH3 BT3314 is predicted to have a signal I peptide, indicating that the protein is most likely localised to the periplasmic space.

5.4.2.9. Determining the activity BT3314, a GH3 enzyme encoded by PUL$_{btPUS}$

The ability of BT3314 to hydrolyse various glyosidic linkages was assessed by incubating recombinant enzyme (10 µM and 0.01 µM) at 37 ^{o}C for 1 h against β1,4 – linked galactobiose, mannobiose and xylobiose as well as β-linked gluco-dissacharides (β1,4/ β1,3/ β1,6). TLC analysis revealed that BT3314 was only able to hydrolyse β-D-gluco-disaccharides, demonstrating a preference for the β1,6-linkage. BT3314 was also active against p-nitrophenyl-β-D-glucopyranoside (PNP-β-D-glucose), confirming that BT3314 was a β-glucosidase as predicted, and hydrolysing gluco-configured sugars from the non-reducing end with an exo-mode of action.

The observed preference for β-linkage was further explored quantitatively using a linked-assay kit that detected the production of glucose. Briefly, released glucose is phosphorylated and is oxidised by glucose-6-phosphate dehydrogenase, which converts NADP+ to NADPH, which can be detected at A_{340nm}. The assay is fully described in Methods and Methodology, Section 2.2.1.1. The activity of BT3314 against pNP-β-D-glucose was similarly assessed spectrophotometrically owing to the stoichiometric increase in A_{400nm} when the glyosidic linkage is cleaved. The assay is described in Methods and Methodology, Section 2.2.1.3. Kinetic parameters of BT3314 vs substrates were conducted in 50 mM Na-HEPES buffer, pH 7.0. Rates were plotted using GraphPad Prism 5.0 and analysed via linear regression or non-linear regression (against pNP-β-D-glucose). The full-data set is shown in Table 5.5 and example graphs displayed in Figure 5.22.

The β-glucosidase BT3314 showed a distinct preference for the β1,6-linked disaccharide, demonstrating a k_{cat}/K_m value of 5.6 μM^{-1} min^{-1} compared to 0.18 μM^{-1} min^{-1} for laminaribiose (β1,3) and 0.05 μM^{-1} min^{-1} for cellobiose (β1,4). The K_m values for these substrates were too high to determine. Full Michaelis-Menten kinetics were obtained for BT3314 against pNP-β-D-glucose. BT3314 had an extremely low catalytic efficiency against this substrate, with a k_{cat}/K_m of 0.02 μM^{-1} min^{-1}, highlighting the potential significance of the +1 sub-site of the enzyme in substrate binding. The GH3 enzyme displayed a K_m value of 2.4 mM and a k_{cat} (number of substrate molecules, hydrolysed per active site, per unit of time) of 40.8 $min^{-`1}$.

Enzyme	Substrate	K_m (μM)	K_{cat} (min^{-1})	K_{cat}/K_m ($\mu M^{-1}min^{-1}$)
BT3314	pNP-β-D-Glu	2403	40.82	0.02
	β-D-Glc-[1->3]-D-Glc	ND	ND	0.18
	β-D-Glc-[1->4]-D-Glc	ND	ND	0.05
	β-D-Glc-[1->6]-D-Glc	ND	ND	5.6

Table 5.5. Kinetic parameters of BT3314, against β-linked gluco-disaccharides. ND = No data.

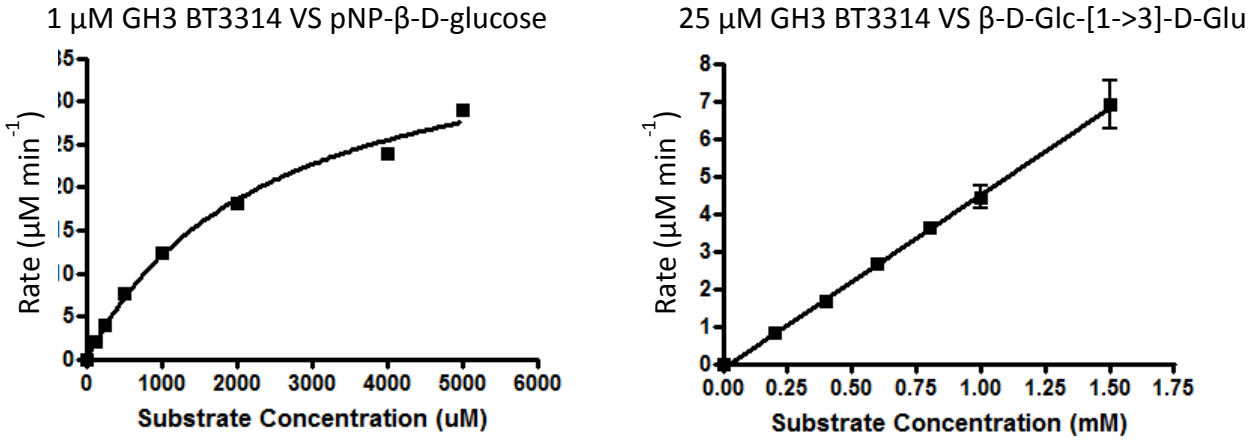


Figure 5.22. Example graphs of kinetic data generated for BT3314, VS pNP-β-D-glucose and β1,3-laminabiose. Kinetic data of BT3314 versus pNP-β-D-glucose (A) and β1,3-laminabiose (B). Against pNP-β-D-glucose, initial rates were derived spectrophotometrically, where-by rate of absorption (A min^{-1}) at OD_{400nm} was converted to rate of glucose released, inferred through pNP group cleavage and release (min^{-1}) using the extinction coefficient of pNP at pH 7.0. Rates were analysed in GraphPad Prism 5.0 using Michaelis-Menten kinetics. Similarly, against β-linked disaccharides, rates of absorption at 340 nm were ascertained and converted to rate of glucose release (min^{-1}) using the extinction coefficient of NADPH and plotted in GraphPad Prism 5.0 before being analysed using linear regression.

5.5. The *Bacteroides ovatus* GH3 β-glucosidase, BACOVA_00946, displays significantly greater activity than the homologous Bacteroides thetaiotaomicron enzyme.

5.5.1. Cloning and Expressions

DNA encoding for the GH3 β-glucosidase, BACOVA_00946, minus signal peptide, was generated by PCR using primers which encoded for Nhe1 and Xho1 restriction sites at the 5' and 3' ends respectively. Digested DNA was subsequently ligated into a pET28a expression vector in a manner which ensured the inclusion of an N-terminal poly-His_6 tag was appended to the GH3 β-glucosidase. The validity of the cloned DNA sequence was confirmed by DNA sequencing (MWG, Germany) using primers probes

commentary to the T7 promoter and T7 terminator encoded by the expression plasmid. BACOVA_00946 was expressed in *E. coli* TUNER cells according to standard protocol and purified by IMAC. SDS-PAGE analysis showed that the enzyme was had been highly purified, Figure 5.23.

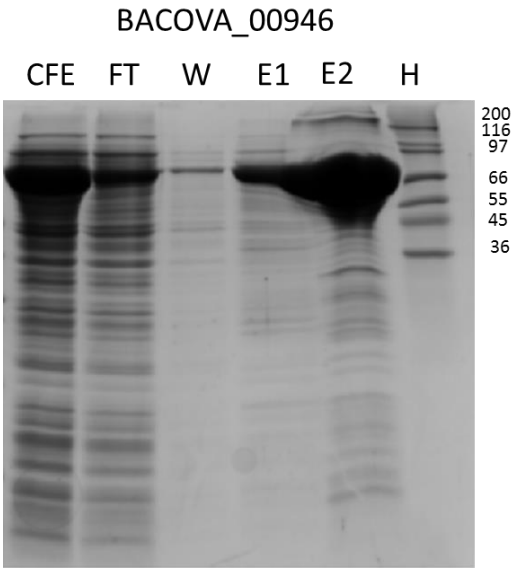


Figure 5.23. SDS-PAGE of recombinant expression and purification of BACOVA_00946. SDS-PAGE of a typical BACOVA_00946 purification following IMAC. All SDS-PAGE gels were 12 % (w/v). Marker M_r is displayed in kDa. L = low size marker; P = insoluble fraction; CHE = cell free extract; FT = flow through; W = wash (1x TALON buffer); E1 = elution 1 (5 mM imidazole,1 x TALON buffer); E2 = elution 2 (100 mM imidazole, 1 x TALON buffer). Bacova_00946 ran with an apparent M_r of 80 kDa, consistent with the recombinant protein's predicted sizes of 53 kDa

5.5.2.2 Evaluation of BACOVA_00946, a GH3 β-glucosidase vs β-linked gluco-disaccharides

The GH3 encoded by *B. ovatus*, BACOVA_00946, displayed identical specificities as the characterised GH3 of B. theta, BT3314, confirming that BACOVA_00946 is a β-glucosidase. The kinetic parameters of BACOVA_00946 against β-linked gluco-disaccharides and pNP-β-D-glucose was ascertained using the same methodology as the *B. theta* enzyme and are reported in Table 5.6 respectively. Example graphs are displayed in Figure 5.24.

BACOVA_00964 displayed the same trend in linkage preference as the *B. theta* enzyme, demonstrating the highest catalytic efficiency for the β1,6-linkage with a k_{cat}/K_m value of 38.8 μM^{-1} min^{-1}. The enzyme had a ~10-fold lower catalytic efficiency against the β1,3-linked disaccharide (3.7 min^{-1} μM^{-1}) and displayed a K_{cat}/K_m of 0.65 μM^{-1} min^{-1} for the β1,4-linkage, a ~60-fold reduction in activity compared to the enzyme's efficiency when cleaving the β1,6-linkage. The *B. ovatus* GH3 is significantly more active against all tested linkages than the equivalent enzyme in *B. theta*. BACOVA_00964 was ~7-fold more efficient at cleaving the β1,6-linked disaccharide, ~20-fold more efficient against the β1,3-linkage and ~10 fold more efficient when hydrolysing the β1,4-linkage. BACOCA_00964 demonstrated extremely low catalytic activity against pNP-β-D-glucose, although the enzyme was 10-fold more efficient that the corresponding *B. theta* β-glucosidase.

Enzyme	Substrate	K_m (µM)	K_{cat} (min^{-1})	K_{cat}/K_m ($µM^{-1}min^{-1}$)
Bacova_00946	pNP-β-D-Glu	1052	195.8	0.19
	β-D-Glc-[1->3]-D-Glc	ND	ND	3.73
	β-D-Glc-[1->4]-D-Glc	ND	ND	0.65
	β-D-Glc-[1->6]-D-Glc	ND	ND	38.8

Table 5.6. Kinetic parameters of BACOVA_00964, against β-linked gluco-disaccharides. ND = No data.

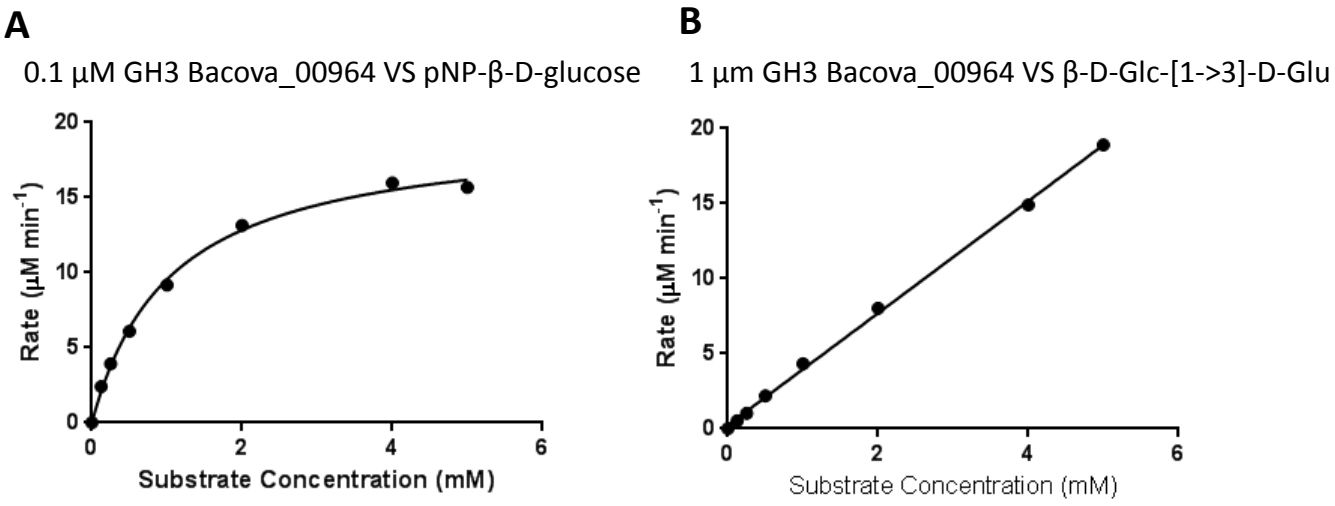


Figure 5.24. Example graphs of kinetic data generated for BACOVA_00964, VS pNP-β-D-glucose and β1,3-laminabiose. Kinetic data of BACOVA00964 versus pNP-β-D-glucose (A) and β1,3-laminabiose (B).

5.5.2.3. Amino Acid sequence comparison between the GH3 β-glucosidases, BT3314 and BACOVA_00946

To explore the possible biochemical rationale between the observed differences in catalytic activity demonstrated between the GH3 β-glucosidases of *B. theta* (BT3314) and *B. ovatus* (BACOVA_00946), an alignment of the amino acid sequences of the two proteins was performed, and is displayed in Figure 5.25. The two GH3 β-

glucosidases share a 91 % sequence similarity. Interestingly, the only major differences in amino acid residues is a 12 amino acid insertion in the *B. theta* protein, encompassing residues Ser^{239} – Tyr^{250}. Sequence alignments of previously characterised GH3 family β-glucosidases show that this insertion is, at least in sequence, in close proximity to the catalytic nucleophile of the enzyme, Asp^{265} and Asp^{279} in the *B. theta* and *B. ovatus* enzymes, respectively. Whilst highly speculative, it is possible that this amino acid insertion near the catalytic nucleophile of the enzyme is responsible for the observed difference in catalytic activity. Interestingly, alignments of the GH3 amino acid sequences encoded by the β1,6-glucan locus of all sequenced *B. theta* clinical isolates reveals that the insertion of 12 amino acids is 100 % conserved, whilst significantly, all alignments of the GH3 from the β-1,6-glucan PUL of other sequenced *Bacteroides* species are more similar to the *B. ovatus* GH3 enzyme, lacking this insertion.

Alignment – BT3314 and BACOVA_00946) – homology 91 %

Figure 5.25. Full amino acid sequence alignment of the GH3 β-glucosidases BT3314 and BACOVA_00946. Catalytic residues of the enzyme, based on homology, are highlighted in green. N= catalytic nucleophile. The 12 amino acid insertion (residues Ser^{239} – Tyr^{250}) in the *Bacteroides theta* enzyme, hypothesized to be responsible for retardation in catalytic activity, is highlighted by a blue box.

5.6. Culture supernatant analysis.

Thin layer chromatography (TLC) of early stationary phase (2.0 OD_{600nm}) culture supernatants of *B. theta* and *B. ovatus* growth with pustulan revealed the accumulation of a glycan the growth media (Figure 5.26), confirmed to be the disaccharide β1,6-glucobiose by HPAEC-PAD analysis. The potential significance for the accumulation of the disaccharide is explored in the Discussion.

Analysis of culture supernatants of *B. theta* growth with pustulan revealed a large number of β1,6-oligosaccharides during the exponential phase (Figure 5.27). This is consistent with the endo-mode of action of the GH30 enzyme BT3312

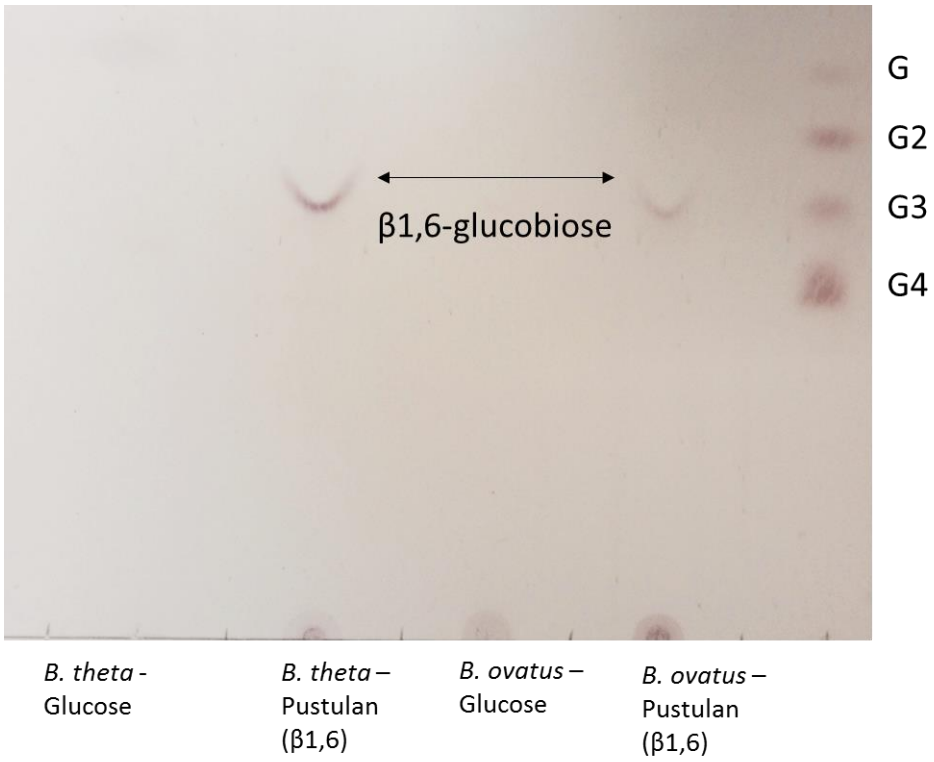
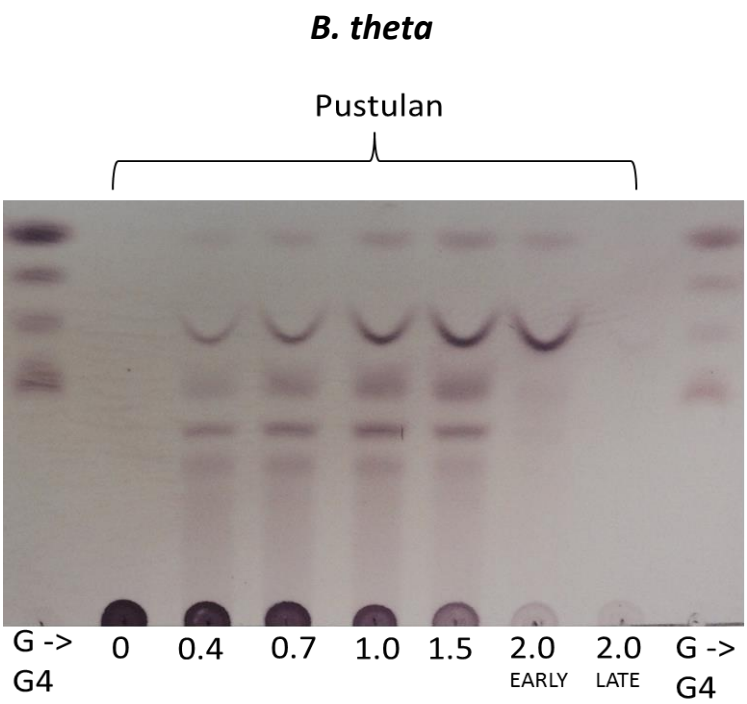


Figure 5.26. Thin layer chromatography (TLC) of culture supernatants of *B. theta* and *B. ovatus* grown on pustulan and glucose to stationary phase. Culture supernatants from stationary phase cultures were subjected to TLC using the following markers: glucose (G) and available β1,4-linked glucooligosaccharides (cellobiose (G2) – cellotetraose (G4).

A



B

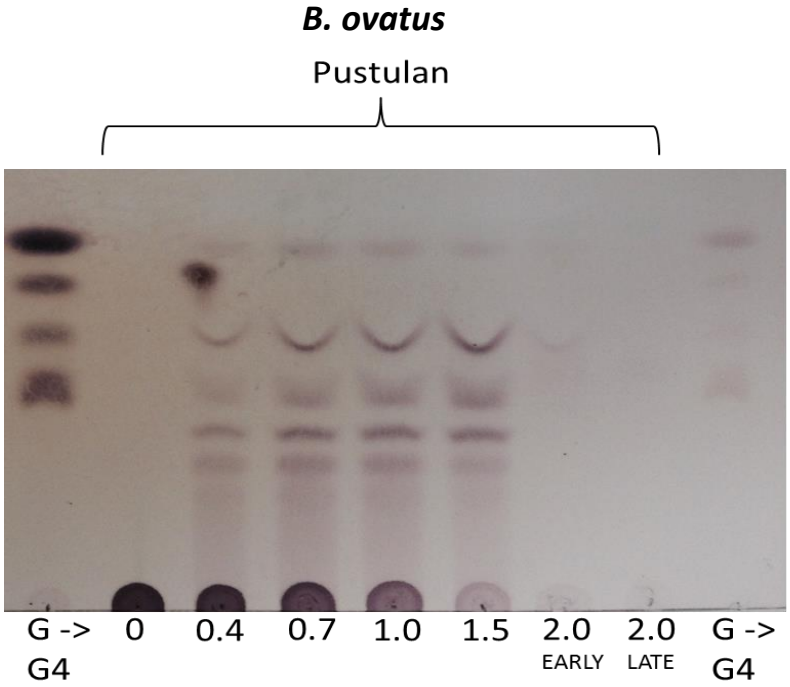


Figure 5.27. TLC analysis of culture supernatants of *B. theta* and *B. ovatus* grown on the β1,6-polysaccharide, pustulan. Overnight grown minimal media with glucose (5 mg/ mL)) were used to inoculate minimal media (5 mL) containing pustulan (5 mg/mL). Anaerobic growth was measured spectrophotometrically (OD_{600nm}), which is representative of different stages of bacterial growth. Samples of the culture supernatant were taken at various optical densities and analysed by TLC. Samples are labelled according to the optical density (OD_{600nm}) at which they were taken. EARLY/LATE represent stationary phase samples taken immediately when stationary phase was reached or ~6 h into stationary phase respectively. Glucose and β1,4-glucooligosaccharides (d.p. 2 to 4) were used as standards (denoted by G -> G4). Panel A and B displays growth of *B. theta* and *B. ovatus,* respectively,with pustulan.

5.7. Discussion

This chapter provides evidence that the PUL_{btPUS} targets the utilization of β1,6-glucans. This PUL appears to be widespread amongst a number of *Bacteroides* species, suggesting that β1-6-glucans are a viable energy source. Whilst β1,6-glucans upregulate PUL_{btPUS} and PUL_{boPUS} the exact signalling molecule which activates these loci has not been elucidated. Owing to the observation that *B. theta* and *B. ovatus* cells cultured with glucose as a sole carbon source display no increase in gene expression, it can be concluded with a high degree of certainty that monomeric glucose does not activate the β1,6-glucan PUL. Instead, the activating molecule is most likely a β1,6-oligosaccharide, the D.P of which was not explored.

The data presented in this chapter attempts to construct a model of β1,6-glucan utilization in Bacteroides through the characterisation of the proteins encoded by PUL_{btPUS}. The surface binding proteins encoded by PUL_{btPUS}, and by inference the equivalent proteins derived from PUL_{boPUS}, target the β1,6-glucan pustulan preferentially, and display similar affinities for β1,6-glucooligosaccharides with a D.P > 7 as shown by ITC. β1,6-glucans have been shown to adopt a hook-like conformation in solution, with four β-D-1,6-glucopyranose residues forming one hook (Lowman *et al.*, 2011). It is possible that the two surface binding proteins not only recognise and interact with the individual glucopyranose residues of the β1,6-glucan, but also the secondary structure formed by the target glycan. Previously characterised $SusD_h$ proteins have been observed to recognise glycan structures (Koropatkin *et al.*, 2009);(Koropatkin *et al.*, 2008). The binding specificities of both $SusD_h$, BT3311, and the SGBP, BT3313, are consistent with the activity of the surface GH30 endo-β1,6-glucanase. Complementary to the biological function of SGBPs, sequestering target glycans to the cell surface, a number of SGBPs have been observed to bind to target

polysaccharides and their constituent oligosaccharides following surface enzyme degradation (Cuskin *et al.*, 2015b). Indeed, this is consistent with the longer oligosaccharide products generated by the surface GH30 endo-β1,6-glucanase generated during the initial stages of degradation as observed *in vivo*. $SusD_h$ proteins bind oligosaccharides for transportation through their cognate outer membrane $SusC_h$ into the periplasmic space where further enzymatic deconstruction occurs, reiterating the hypothesis that both extracellular binding proteins target long chain oligosaccharides and is complemented by studies of other PUL systems (Rogowski et al., 2015).

Genomic data indicate that the core enzymatic machinery necessary to fully degrade β1,6-glucans is seemingly conserved amongst *Bacteroides* species, and consists solely of a GH30 endo-β1,6-glucanase and a GH3 β-glucosidase. As such, PUL_{btPUS} presents a relatively simple model of polysaccharide utilization by a PUL system. Indeed, the complexity of a PUL (complexity meaning the number of protein encoding ORFs required to fully degrade and utilize the target substrate) correlates to the complexity and accessibility of the target substrate (Rogowski et al., 2015). The GH30 endo-β1,6-glucanase, predicted to be localised to the cell surface, targets extracellular polysaccharides initially, cleaving the sugar polymer into β1,6-olgiosaccharides. The presence of an endo-acting enzyme on the cell surface, generating oligosaccharides which can be further processed (usually following transportation into the periplasm), is typical of other PULs which target 'less-complex' substrates (galactan. Jonathon Briggs personal communication).

The capability of the surface GH30 endo-β1,6-glucanase to hydrolyse short chain oligosaccharides to a D.P of three (not quantified) is interesting. Previously characterised surface endo-acting enzymes preferentially target longer

oligosaccharides and display little/no catalytic activity against oligosaccharides with a d.p < 5, owing to weak binding at sub-sites proximal to the active site of the enzyme (Cuskin et al., 2014, Rogowski et al., 2015), and perhaps suggests a significant proportion of oligosaccharides are degraded at the cell surface. The observation that culture supernatants of *B. theta* and *B. ovatus* grown in the presence of pustulan results in the accumulation of β1,6-glucobiose in the growth media, suggests that the bulk of degradation does indeed occur extracellularly and is concurrent with the activity of the recombinant GH30 enzyme. It has recently been proposed that *Bacteroides* utilizes a 'selfish' mechanism of polysaccharide utilization when presented with a complex glycan, which would require a significant energy investment by the bacterium to degrade. The bulk of depolymerisation occurs within the bacterial cell and therefore reduces the quantity of depolymerised products released extracellularly for potential scavenging by competitors (Cuskin et al., 2014, Rogowski et al., 2015). Conversely, Bacteroides has been shown to employ strategies which liberate oligosaccharides extracellularly during degradation, to the extent of being able to sustain the growth of co-cultured bacterial species *in vivo* (Rogowski et al., 2015). The abundance of extracellular β1,6-oligosaccharides released by the surface GH30 during growth suggest that the bacterium has not evolved to utilize β1,6-glucans in an overtly selfish manner. The significance of this is not clear, although a 'self-less' model of polysaccharide utilization in *B. theta* is usually utilized for high-abundance/low-complexity substrates (unpublished data). Interestingly, β1,6-glucans would not be abundant, although the complexity of this glucan fraction can vary drastically.

The exo-acting GH3 β-glucosidase, predicted to reside within the perisplasm, most likely further depolymerises β1,6-linked oligosaccharides by liberating glucose from imported GH30 endo-glucanase products, which can be transported into the

cytoplasm and catabolized by the bacterium. However, the necessity for this enzyme can be questioned owing to the ability of the surface GH30 glucanase to generate glucose from the polysaccharide. The *B. theta* GH3, BT3314, has been shown to target a number of β-linked disaccharides, and displayed the greatest preference for the β1,6-linkage, albeit with a severely reduced efficiency when compared to the *B. ovatus* GH3 β-glucosidase. The difference in catalytic activity observed between the recombinant GH3 enzymes of *B. theta* and *B. ovatus* is surprising owing to the high degree of amino acid sequence homology shared between the two enzymes (91 %). It should be noted, however, that due to availability of substrate, the activity of the GH3 enzymes could not be tested against β-oligosaccharides with a D.P > 2. It is plausible that the *B. theta* enzyme, or indeed both enzymes, may be catalytically more efficient against the tri or tetrasaccharide, accommodating additional β-D-glucose moieties in potential +2/+3 catalytic sub-sites, lowering the affinity constant (K_m). Regardless, the need for the GH3 enzyme in the model of β1,6-glucan degradation can be questioned; the low catalytic activity of the *B. theta* GH3 enzyme in conjunction with the ability of the surface GH30 to produce glucose suggests a level of redundancy in the *B. theta* β1,6-glucan PUL system. To reiterate, gene up-regulation data suggest that the activating molecule of the β1,6-glucan PUL is a β1,6-linked gluco-oligosaccharide. Catalytically low levels of β-glucosidase activity would therefore be necessary to degrade the signalling molecule, preventing continuous up-regulation of the PUL, providing one potential reason for the lack of total-redundancy of the GH3 enzyme in *B. theta*.

The degree of β1,3-linked glucose side-chains of the β1,6-glucan fraction of the yeast cell wall can vary significantly amongst yeast species. Studies report branching of up to 75 % in *Schizosaccharomyces pombe*, a yeast utilized in the brewing industry.

BT3314, the endo-β1,6-glucanase can tolerate a degree of decoration of the β1,6-backbone and can hydrolyse purified yeast β1,6-glucans from *Saccharomyces cerevisiae*, which on average has β1,3-linked glucose decorations every 19 glucose residues (Aimanianda *et al.*, 2009). The degree of tolerance the GH30 endo-β1,6-glucanase possesses for highly decorated β1,6-glucans is not known. It can be speculated that a highly decorated β1,6-glucan, such as the β-glucans possessed by *S. pombe* (*B. theta* targets the α-mannan of *S. pombe (Cuskin et al., 2015b)*), would result in a lower frequency of cuts making it a less-accessible backbone, potentially altering the proposed model of β1,6-glucan degradation dramatically. Whilst highly speculative, it can be hypothesised that when *Bacteroides* is challenged with highly decorated β 1,6-glucans, the GH3 β-glucosidase would play a much more pivotal role and may even provide a competitive advantage to *B. ovatus* which encodes for a significantly more active GH3 β-glucosidase owing to the enzyme's ability to remove the β1,3-linked glucose decorations of imported oligosaccharides more efficiently. Indeed, *B. theta* is capable of degrading and utilizing yeast α-mannan, whilst *B. ovatus* cannot, lacking the PULs necessary to target the highly complex glycan (Cuskin *et al.*, 2015b). *B. theta* requires a significant energy investment to target the highly complex α-mannans of the yeast cell wall*. In vivo* studies have shown that wild-type *B. theta* is outcompeted by a mutant lacking the PULs needed to target α-mannan when fed a non-mannan diet, thereby suggesting α-mannan utilization is only 'cost-effective' when the target substrate is available. The evolutionary persistence of the α-mannan PULs in *B. theta* therefore suggests that α-mannans are valuable target substrates for the gut bacterium, despite the 'risk'. Due to the role of β1,6-glucans in the structural integrity of the yeast cell wall, it can be speculated that enzymatic disruption of this glucan may potentially liberate other structural polysaccharides. Tentatively, *B. theta*

may have evolved a strategy to disrupt the yeast cell wall, gaining access to a preferential, more abundant, substrates as a nutrient sources. This view is supported by the proposed redundancy of the *B. theta* GH3 glucosidase and its similarity with GH3 enzymes from other *B. theta* strains. *B. ovatus*, which is unable to utilize yeast α-mannans and encodes for a significantly more active GH3 β-glucosidase may be a more efficient β1,6-glucan utilizer as a result, due to the organism's requirement to access β1,6-glucans solely as an energy source.

5.8 Further Work.

Future work is necessary to further explore the observed differences in GH3 glucosidase activity between the *B. theta* and *B. ovatus* enzyme, in order to fully ascertain if such discrepancies in activity extend to *in vivo*, and to fully elucidate the biological rationale for such differences in activity. To achieve this goal, competition experiments must be performed to see if the more active GH3 of *B. ovatus* provides a competitive advantage over *B. theta* when cultured with pustulan and yeast β1,6-glucans of various species to investigate if the degree of side-chains affects β1,6-glucan utilization. Assessing the ability of Bacteroides species to degrade the β1,6-glucans of other yeast species would help to inform our knowledge of complex glycan utilization and sharing with the gut.

Yeasts, such as C*andida albicans* are commensal members of the gut, and can be pathogenic in immunocompromised patients. The ability of *B. theta* to degrade yeast cell wall glycans, and the potential of the organism to disrupt yeast cell walls may help to modulate yeast levels within the gut, preventing opportunistic yeast infections and pathogenesis. The ability of the bacterium to access and degrade intact yeast cell walls must be assessed.

Further health implications may arise from the release of yeast cell wall molecules into the gut lumen, induced by *B. theta* hydrolysis. Both β1,3 and β1,6 glucans have been observed to be immunostimulatory and may interact with the immune system of the host. Indeed, β1,6 glucans are potent cytokine inducers in human blood (Noss et al., 2013). As such, the degradation of yeast cell wall polysaccharides by Bacteroides may result in inflammation of the gut which would need to be explored.

Chapter 6.0: Final Discussion

The data presented in this study provides example mechanisms by which organisms that inhabit the highly competitive environments of both the rumen and the human distal gut have evolved unique strategies to facilitate the degradation and utilization of complex carbohydrates.

Bacteria which inhabit the rumen have evolved to target plant biomass primarily through deploying glycoside hydrolases which often encode for non-catalytic CBMs. CBMs enhance the activity of their parent catalytic domain (As described in the Introduction, Section 1.2). Understanding the molecular basis of CBM-ligand interactions informs enzyme based saccharification strategies to disrupt and degrade the highly complex and heterogeneous plant cell wall. As such, engineering the specificity of enzymes and CBMs is of industrial relevance. Chapter 3 describes the continued structural and biochemical characterisation of a non-catalytic binding module of the CBM65 family, derived from an endoglucanase encoded by *Eubacterium cellulosolvens*, an organism which is consistently dominant in the rumen (Prins *et al.*, 1972). The CBM65 family uniquely displays specificity for the side-chains of the decorated β1,4-glucan xyloglucan, over linear β1,4-glucans. The crystal structure of CBM65B in complex with ligand provides insight into side-chain specificity, revealing that specificity for the decorated xyloglucan is conferred through an extended hydrophobic platform, consisting of five aromatic residues, forming interactions between both the glucan backbone and xylose decorations of the polysaccharide. ITC revealed that the CBM65s displayed a higher affinity for the decorated XXXG oligosaccharide than the undecorated equivalent oligosaccharide, cellotetraose. Substituting the back-bone specific residue in CBM65A abrogated

cellulose recognition but did not impact upon xyloglucan affinity, therefore emphasising the importance of side-chain recognition, which is in contrast to other previously characterised CBMs (Najmudin *et al.*, 2006). Surprisingly, mutating the structurally equivalent residue in CBM65B did not impact upon cellulose recognition, despite no other obvious candidate residues being apparent in the ligand binding cleft. The mechanism by which CBM65B recognises linear β1,4-glucans is some-what unclear. A co-complexed structure of CBM65B with cellooligosaccharides is necessary to elucidate whether CBM65A and 65B recognise cellulose via differening mechanisms or to identify a functionally equivalent residue in the CBM65B which provides a polar contact with the glucose backbone of linear β1,4-glucans.

X-2 L110F, a CBM variant engineered from a xylan specific *Rhodothermus marinus* CBM, has recently been shown to bind xyloglucan via interactions with both the backbone and decorations of xyloglucan (von Schantz *et al.*, 2014). X-2 L110F displays similarities in xyloglucan recognition to CBM65B, both displaying a preference for the branched polysaccharide over linear β1,4-glucans. Likewise, the aromatic amino acid residues housed within the ligand binding cleft have a crucial impact on ligand binding in both CBM65 and X-2 L110F. The binding interface of X-2 L110F is rich in aromatic side chains, forming hydrophobic interactions with XXXG. However, whilst CBM65 binding of XXXG is primarily mediated by apolar and hydrophobic contacts, X-2 L110F binding of XXXG requires substantial binding site plasticity in order to achieve polar contacts between the xylose decorations of XXXG, with several amino acids undergoing conformational changes in the ligand binding cleft. In contrast, the ligand binding site of CBM65 is pre-arranged and undergoes little rearrangement of amino acid chains upon ligand binding. CBM65B-XXXG complex reveals only one polar contact between the protein and xylose decorations of XXXG, whilst up to nine

polar contacts are formed between X-2 L110F and the xylose sidechains, Interestingly, X02 L110F and CBM65s recognise differing faces of the XXXG oligosaccharide. Whilst the CBM65 display no binding to xylans, X-2 L110F displays higher affinity for oat-spelt xylan over tamarind seed xyloglucan, again suggesting differing mechanism of ligand recognition. Such differences observed in xyloglucan recognition between the CBM65s and X-2 L110F, as well as differences in ligand binding cleft architecture, may enable the targeting of xyloglucan in differing plant cell wall contexts and be of industrial benefit. Indeed, CBMs can recognise identical ligands in solution, but demonstrate differing specificities for cell walls (McCartney *et al.*, 2006). Likewise, whilst the primary role of CBMs is to increase enzyme substrate proximity, some CBMs may enable targeting to accessible regions of the plant cell wall (Montanier *et al.*, 2009). The multiple binding specificities demonstrated by the CBM65s may enable more efficient targeting in heterogeneous plant cell walls. The structure of xyloglucan can vary between cell type and plant species (Fry *et al.*, 1993). The mechanism of ligand recognition displayed by CBM65 may allow for subtle variations in ligand structure and may accommodate the xyloglucans of taxonomically diverse cell walls. These data presented in Chapter 3 provides a model by which decorated polysaccharides are recognised by CBMs, and how a CBM can demonstrate a distinct preference for xyloglucan over β-glucans (Luis *et al.*, 2013).

Chapters 4 and 5 provides insight into yeast cell wall polysaccharide utilization by *B, theta* through the biochemical characterisation of the molecular machinery encoded by the respective PUL systems. The ability of *Bacteroides* to utilize yeast cell polysaccharides is entirely consistent with the organisms perceived role as a ‘glycan generalist’ (Martens *et al.*, 2009), enabling *Bacteroides* to flourish in a highly-competitive eco-system. The strategies utilized by *B. theta* to target different fractions

of the yeast cell wall, α-mannan and β1,6-glucans respectively, are in stark contrast. Indeed, α-mannan utilization appears to be restricted to limited *Bacteroides* species, (Cuskin *et al.*, 2015b), whilst PUL_{PUS} is seemingly wide-spread. β1,6 glucan utilization represents a seemingly simple model of polysaccharide utilization. Extracellular deconstruction is dominated by an endo-acting β1,6-glucanase, with a high efficiency against the polysaccharide and high d.p β1,6-oligosaccharides. This is reminiscent of xyloglucan utilization by *B. ovatus* (Larsbrink, 2014), whereby extracellular metabolism of the polysaccharide is mediated by endo-acting enzymes. Likewise, the $SusD_h$ and SGPB encoded by PUL_{PUST} bind polysaccharide and β1,6-oligosaccharides with a d.p >6 preferentially. Further parallels can be drawn between the xyloglucan system and pustulan utilization by *B. theta*, due to the speculated redundancy of exo-acting enzymes encoded by the respective PULs. The periplasmic GH3 β-glucosidase of PUL_{btPUS} is catalytically feeble. The surface endo-glucanase can generate glucose from the polysaccharide and so the necessity of this enzyme for β1,6-glucan utilization is unknown. It can be speculated that this enzyme may serve to hydrolyse the molecular cues which up-regulate the PUL. The significance of the GH3 needs to be elucidated by assessing the ability of GH3 deficient *B. theta* strains to utilize β1,6 glucans. Such observations are further emphasised by the fact that the homologous GH3 β-glucosidase encoded by PUL_{boPUS} displays significantly greater activity against tested disaccharides.

Evidence provided in Chapter 4 shows that B. theta as evolved to utilize the α-mannan of different yeast species. An α-galactosidase encoded by PUL-Man1 was able to sufficiently remove the α-galactosyl residues appended to the backbone of S. pombe α-mannan to enable GH76 endo-mannanase attack of the backbone.

Numerous attempts to manipulate the composition of the human gut microbiota have been attempted. However, the lack of knowledge of complex glycan utilization and sharing within the gut environment by key glycan degraders has slowed such efforts (Hooper *et al.*, 2002). Evidence provided in Chapter 4 demonstrates how the α-mannan apparatus of *B. theta* is tailored to minimise substrate loss to the environment, and by extension, the gut lumen. Conversely, oligosaccharides are freely liberated during β1,6-glucan utilization by *B. theta*. Extracellular side chain removal of α-mannan is mediated by a constitutively expressed α-mannosidases, not encoded by PUL-Man1-3. Extracellular mannose-side chain removal is severely limited, but sufficient enough to enable GH76 endo-α1,6-mannanase access to the sterically restricted α1,6-mannanase back-bone. Mannose liberated during *B. theta* α-mannan deconstruction could not support the growth of other *Bacteroides* species. Indeed, biochemical characterisation of the surface GH76 endo-α1,6-mannanases reveals that they are adapted to hydrolyse longer α1,6-oligosaccharides, and are slower than their periplasmically located counterparts. This strategy would reduce loss of substrate to the gut lumen, synergising substrate cleavage with importation. The periplasmic α-mannosidases were ~20-50-fold more efficient at mannose side-chain deconstruction than the enzyme believed to facilitate extracellular debranching. These data suggest *B. theta* has adopted a selfish mechanism of α-mannan degradation, owing the significant energy investment targeting α-mannan entails (Cuskin *et al.*, 2015b). Such observations resonate with the findings of (Rogowski *et al.*, 2015), where-by the by products of *B. ovatus* complex xylan utilization were unable to support the growth of co-cultured xylooligosaccharide users. Conversely, the products liberated during *B. ovatus* degradation of simpler xylans were able to the support growth of co-cultured species. These findings suggest that glycan complexity dictates the strategy of

utilization and contributes to our understanding of bacterial interactions within the gut microbiota, potentially informing prebiotic strategies.

The data presented in Chapter 4 also shows how *B. theta* has evolved to target the α-mannan of differing yeast species. The α-galactosidase encoded by PUL-Man1 was able to sufficiently remove the α-galactosyl residues appended to the backbone of S. pombe yeast mannan to enable GH76 endo-mannanase attack. PUL-Man2 encodes a GH130 glycoside hydrolase which targets the β1,2-mannosidic linkages that cap the mannose side-chains of *Candida albicans* α-mannan, a pathogenic yeast and commensal member of the gut microbiota (Cuskin *et al.*, 2015b). The ability of *B. theta* to target alternate fungal mannans most likely arose due to exposure to dietary mannans or commensal yeasts in the gut.

To conclude, evidence presented in this thesis demonstrates that *Bacteroides* has evolved the ability the target yeast cell wall polysaccharides, and exploits α-mannans and β1,6-glucans through differing strategies. *B. theta* encodes for PUL-Man1-3 and $PUL_{PUS.}$ It can be speculated that *B. theta* may disrupt yeast cell walls through β1,6-glucan hydrolysis, thereby providing access to a niche and preferential substrate (α-mannan). Conversely, *B. ovatus* does not encode for PUL-Man, but encodes $PUL_{PUS.}$ *B. ovatus* may potentially solely target β1,6-glucans as an energy source, a hypothesis which may be supported by the greater activity displayed by the GH3 encoded by PUL_{boPUS} than the *B. theta* homologue. This needs to be assessed by competition experiments. Nevertheless, the impact of yeast modulation within the gut (especially of pathogenic yeasts) may be of importance to human health.

References

Aguilar-Uscanga, B. and Francois, J.M. (2003) 'A study of the yeast cell wall composition and structure in response to growth conditions and mode of cultivation', *Lett Appl Microbiol*, 37(3), pp. 268-74.

Aimanianda, V., Clavaud, C., Simenel, C., Fontaine, T., Delepierre, M. and Latge, J.P. (2009) 'Cell wall beta-(1,6)-glucan of *Saccharomyces cerevisiae:* structural characterization and in situ synthesis', *J Biol Chem*, 284(20), pp. 13401-12.

Albersheim, P., An, J., Freshour, G., Fuller, M.S., Guillen, R., Ham, K.S., Hahn, M.G., Huang, J., O'Neill, M., Whitcombe, A. and et al. (1994) 'Structure and function studies of plant cell wall polysaccharides', *Biochem Soc Trans*, 22(2), pp. 374-8.

Arpaia, N., Campbell, C., Fan, X.Y., Dikiy, S., van der Veeken, J., deRoos, P., Liu, H., Cross, J.R., Pfeffer, K., Coffer, P.J. and Rudensky, A.Y. (2013) 'Metabolites produced by commensal bacteria promote peripheral regulatory T-cell generation', *Nature*, 504(7480), pp. 451-+.

Bakhtiar, S.M., LeBlanc, J.G., Salvucci, E., Ali, A., Martin, R., Langella, P., Chatel, J.M., Miyoshi, A., Bermudez-Humaran, L.G. and Azevedo, V. (2013) 'Implications of the human microbiome in inflammatory bowel diseases', *FEMS Microbiol Lett*, 342(1), pp. 10-7.

Ballou, C.E. (1990) 'Isolation, characterization, and properties of *Saccharomyces cerevisiae* mnn mutants with nonconditional protein glycosylation defects', *Methods in Enzymology*, 185, pp. 440-470.

Bohlin, C., Praestgaard, E., Baumann, M.J., Borch, K., Praestgaard, J., Monrad, R.N. and Westh, P. (2013) 'A comparative study of hydrolysis and transglycosylation activities of fungal beta-glucosidases', *Appl Microbiol Biotechnol*, 97(1), pp. 159-69.

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

Bolam, D.N. and Koropatkin, N.M. (2012) 'Glycan recognition by the Bacteroidetes Sus-like systems', *Curr Opin Struct Biol*, 22(5), pp. 563-9.

Boraston, A.B., Bolam, D.N., Gilbert, H.J. and Davies, G.J. (2004) 'Carbohydrate-binding modules: fine-tuning polysaccharide recognition', *Biochemical Journal*, 382, pp. 769-781.

Boraston, A.B., Nurizzo, D., Notenboom, V., Ducros, V., Rose, D.R., Kilburn, D.G. and Davies, G.J. (2002) 'Differential oligosaccharide recognition by evolutionarily-related beta-1,4 and beta-1,3 glucan-binding modules', *J Mol Biol*, 319(5), pp. 1143-56.

Bos, M.P., Robert, V. and Tommassen, J. (2007) 'Biogenesis of the gram-negative bacterial outer membrane', *Annu Rev Microbiol*, 61, pp. 191-214.

Buffetto, F., Ropartz, D., Zhang, X.J., Gilbert, H.J., Guillon, F. and Ralet, M.C. (2014) 'Recovery and fine structure variability of RGII sub-domains in wine (Vitis vinifera Merlot)', *Ann Bot*, 114(6), pp. 1327-37.

Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V. and Henrissat, B. (2009) 'The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics', *Nucleic Acids Res*, 37(Database issue), pp. D233-8.

Cartmell, A., McKee, L.S., Pena, M.J., Larsbrink, J., Brumer, H., Kaneko, S., Ichinose, H., Lewis, R.J., Vikso-Nielsen, A., Gilbert, H.J. and Marles-Wright, J. (2011) 'The structure and function of an arabinan-specific alpha-1,2-arabinofuranosidase identified from screening the activities of bacterial GH43 glycoside hydrolases', *J Biol Chem*, 286(17), pp. 15483-95.

Chanliaud, E. and Gidley, M.J. (1999) 'In vitro synthesis and properties of pectin/*Acetobacter xylinus* cellulose composites', *Plant J*, 20(1), pp. 25-35.

Charnock, S.J., Bolam, D.N., Nurizzo, D., Szabo, L., McKie, V.A., Gilbert, H.J. and Davies, G.J. (2002) 'Promiscuity in ligand-binding: The three-dimensional structure of a Piromyces carbohydrate-binding module, CBM29-2, in complex with cello- and mannohexaose', *Proc Natl Acad Sci U S A*, 99(22), pp. 14077-82.

Chen, J. and Seviour, R. (2007) 'Medicinal importance of fungal beta-(1-->3), (1-->6)-glucans', *Mycol Res*, 111(Pt 6), pp. 635-52.

Cho, K.H. and Salyers, A.A. (2001) 'Biochemical analysis of interactions between outer membrane proteins that contribute to starch utilization by *Bacteroides thetaiotaomicron*', *J Bacteriol*, 183(24), pp. 7224-30.

Cohen, S.N., Chang, A.C.Y. and Hsu, L. (1972) 'Nonchromosomal Antibiotic Resistance in Bacteria - Genetic Transformation of *Escherichia-Coli* by R-Factor DNA', *Proceedings of the National Academy of Sciences of the United States of America*, 69(8), pp. 2110-&.

Creagh, A.L., Ong, E., Jervis, E., Kilburn, D.G. and Haynes, C.A. (1996) 'Binding of the cellulose-binding domain of exoglucanase Cex from *Cellulomonas fimi* to insoluble microcrystalline cellulose is entropically driven', *Proc Natl Acad Sci U S A*, 93(22), pp. 12229-34.

Crost, E.H., Tailford, L.E., Le Gall, G., Fons, M., Henrissat, B. and Juge, N. (2013) 'Utilisation of Mucin Glycans by the Human Gut Symbiont *Ruminococcus gnavus* Is Strain-Dependent', *Plos One*, 8(10).

Cuskin, F., Basle, A., Ladeveze, S., Day, A.M., Gilbert, H.J., Davies, G.J., Potocki-Veronese, G. and Lowe, E.C. (2015a) 'The GH130 Family of Mannoside Phosphorylases Contains Glycoside Hydrolases That Target beta-1,2-Mannosidic Linkages in *Candida Mannan*', *Journal of Biological Chemistry*, 290(41), pp. 25023-25033.

Cuskin, F., Flint, J.E., Gloster, T.M., Morland, C., Basle, A., Henrissat, B., Coutinho, P.M., Strazzulli, A., Solovyova, A.S., Davies, G.J. and Gilbert, H.J. (2012) 'How nature can exploit nonspecific catalytic and carbohydrate binding modules to create enzymatic specificity', *Proc Natl Acad Sci U S A*, 109(51), pp. 20889-94.

Cuskin, F., Lowe, E.C., Temple, M.J., Zhu, Y., Cameron, E.A., Pudlo, N.A., Porter, N.T., Urs, K., Thompson, A.J., Cartmell, A., Rogowski, A., Hamilton, B.S., Chen, R., Tolbert, T.J., Piens, K., Bracke, D., Vervecken, W., Hakki, Z., Speciale, G., Munoz-Munoz, J.L., Day, A., Pena, M.J., McLean, R., Suits, M.D., Boraston, A.B., Atherly, T., Ziemer, C.J., Williams, S.J., Davies, G.J., Abbott, D.W., Martens, E.C. and Gilbert, H.J. (2015b) 'Human gut Bacteroidetes can utilize yeast mannan through a selfish mechanism', *Nature*, 517(7533), pp. 165-9.

Davie, J.R. (2003) 'Inhibition of Histone Deacetylase Activity by Butyrate', *J. Nutr.*, 133, pp. 2485S-2493S.

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

Davin, L.B. and Lewis, N.G. (2000) 'Dirigent proteins and dirigent sites explain the mystery of specificity of radical precursor coupling in lignan and lignin biosynthesis', *Plant Physiol*, 123(2), pp. 453-62.

De Filippo, C., Cavalieri, D., Di Paola, M., Ramazzotti, M., Poullet, J.B., Massart, S., Collini, S., Pieraccini, G. and Lionetti, P. (2010) 'Impact of diet in shaping gut microbiota revealed by a comparative study in children from Europe and rural Africa', *Proc Natl Acad Sci U S A*, 107(33), pp. 14691-6.

Delzenne, N.M. and Cani, P.D. (2011) 'Interaction between obesity and the gut microbiota: relevance in nutrition', *Annu Rev Nutr*, 31, pp. 15-31.

Dimitroglou, A., Merrifield, D.L., Moate, R., Davies, S.J., Spring, P., Sweetman, J. and Bradley, G. (2009) 'Dietary mannan oligosaccharide supplementation modulates intestinal microbial ecology and improves gut morphology of rainbow trout, Oncorhynchus mykiss (Walbaum)', *J Anim Sci*, 87(10), pp. 3226-34.

Donohoe, D.R., Garge, N., Zhang, X., Sun, W., O'Connell, T.M., Bunger, M.K. and Bultman, S.J. (2011) 'The microbiome and butyrate regulate energy metabolism and autophagy in the mammalian colon', *Cell Metab*, 13(5), pp. 517-26.

Endo, H., Niioka, M., Kobayashi, N., Tanaka, M. and Watanabe, T. (2013) 'Butyrate-producing probiotics reduce nonalcoholic Fatty liver disease progression in rats: new insight into the probiotics for the gut-liver axis', *PLoS One*, 8(5), p. e63388.

Ferguson, A.D. and Deisenhofer, J. (2002) 'TonB-dependent receptors—structural perspectives', *Biochimica et Biophysica Acta*, 1565, pp. 318-332.

Fernandes, A.N., Thomas, L.H., Altaner, C.M., Callow, P., Forsyth, V.T., Apperley, D.C., Kennedy, C.J. and Jarvis, M.C. (2011) 'Nanostructure of cellulose microfibrils in spruce wood', *Proc Natl Acad Sci U S A*, 108(47), pp. E1195-203.

Fry, S.C., York, W.S., Albersheim, P., Darvill, A., Hayashi, T., Joseleau, J.P., Kato, Y., Lorences, E.P., Maclachlan, G.A., Mcneil, M., Mort, A.J., Reid, J.S.G., Seitz, H.U., Selvendran, R.R., Voragen, A.G.J. and White, A.R. (1993) 'An Unambiguous Nomenclature for Xyloglucan-Derived Oligosaccharides', *Physiologia Plantarum*, 89(1), pp. 1-3.

Fujimoto, Z., Jackson, A., Michikawa, M., Maehara, T., Momma, M., Henrissat, B., Gilbert, H.J. and Kaneko, S. (2013) 'The structure of a *Streptomyces avermitilis* alpha-L-rhamnosidase reveals a novel carbohydrate-binding module CBM67 within the six-domain arrangement', *J Biol Chem*, 288(17), pp. 12376-85.

Ganner, A., Stoiber, C., Wieder, D. and Schatzmayr, G. (2010) 'Quantitative in vitro assay to evaluate the capability of yeast cell wall fractions from *Trichosporon mycotoxinivoran*s to selectively bind gram negative pathogens', *J Microbiol Methods*, 83(2), pp. 168-74.

Goffin, D., Delzenne, N., Blecker, C., Hanon, E., Deroanne, C. and Paquot, M. (2011) 'Will isomalto-oligosaccharides, a well-established functional food in Asia, break through the European and American market? The status of knowledge on these prebiotics', *Crit Rev Food Sci Nutr*, 51(5), pp. 394-409.

Gregg, K.J., Zandberg, W.F., Hehemann, J.H., Whitworth, G.E., Deng, L., Vocadlo, D.J. and Boraston, A.B. (2011) 'Analysis of a new family of widely distributed metal-independent alpha-mannosidases provides unique insight into the processing of N-linked glycans', *J Biol Chem*, 286(17), pp. 15586-96.

Gremmill, T.R. and Trimble, B.R. (1998) 'All pyruvylated galactose in *Schizosaccharomyces pombe* N-glycans is present in theterminal disaccharide, 4,6-O-[(R)-(1-carboxyethylidine)]-Galb1,3Gala1-', *Glycobiology*, 8(11), pp. 1087-1095.

Hakki, Z., Thompson, A.J., Bellmaine, S., Speciale, G., Davies, G.J. and Williams, S.J. (2015) 'Structural and kinetic dissection of the endo-alpha-1,2-mannanase activity of bacterial GH99 glycoside hydrolases from Bacteroides spp', *Chemistry*, 21(5), pp. 1966-77.

Helmann, J.D. (2002) 'The extracytoplasmic function (ECF) sigma factors', *Adv Microb Physiol*, 46, pp. 47-110.

Henrissat, B. and Davies, G. (1997) 'Structural and sequence-based classification of glycosidehydrolases', *Curr Opin Struct Biol*, 7, pp. 637-644.

Hooper, L.V., Midtvedt, T. and Gordon, J.I. (2002) 'How host-microbial interactions shape the nutrient environment of the mammalian intestine', *Annual Review of Nutrition*, 22, pp. 283-307.

Hoverstad, T. and Midtvedt, T. (1986) 'Short-Chain Fatty Acids in GermfreeMice and Rats', *J. Nutr.*, 116, pp. 1772-1776.

Hrmova, M., MacGregor, E.A., Biely, P., Stewart, R.J. and Fincher, G.B. (1998) 'Substrate binding and catalytic mechanism of a barley beta-D-Glucosidase/(1,4)-beta-D-glucan exohydrolase', *J Biol Chem*, 273(18), pp. 11134-43.

Hwang, J.S., Im, C.R. and Im, S.H. (2012) 'Immune disorders and its correlation with gut microbiome', *Immune Netw*, 12(4), pp. 129-38.

Jawhara, S., Habib, K., Maggiotto, F., Pignede, G., Vandekerckove, P., Maes, E., Dubuquoy, L., Fontaine, T., Guerardel, Y. and Poulain, D. (2012) 'Modulation of intestinal inflammation by yeasts and cell wall extracts: strain dependence and unexpected anti-inflammatory role of glucan fractions', *PLoS One*, 7(7), p. e40648.

Kapteyn, J.C., Van Den Ende, H. and Klis, F.M. (1999) 'The contribution of cell wall proteins to the organization ofthe yeast cell wall', *Biochimica et Biophysica Acta*, 1426, pp. 373-383.

Klis, F.M., Ram, A.F. and Montijn, R.C. (1998) 'Posttranslational modifications of secretory proteins ', *Methods in Microbiology*, 26, pp. 223-238.

Kobayashi, M., Matoh, T. and Azuma, J. (1996) 'Two Chains of Rhamnogalacturonan II Are Cross-Linked by Borate-Diol Ester Bonds in Higher Plant Cell Walls', *Plant Physiol*, 110(3), pp. 1017-1020.

Kopecka, M. (2013) 'Yeast and fungal cell-wall polysaccharides can self-assemble in vitro into an ultrastructure resembling in vivo yeast cell walls', *Microscopy*, 62(3), pp. 327-339.

Koropatkin, N., Martens, E.C., Gordon, J.I. and Smith, T.J. (2009) 'Structure of a SusD homologue, BT1043, involved in mucin O-glycan utilization in a prominent human gut symbiont', *Biochemistry*, 48(7), pp. 1532-42.

Koropatkin, N.M., Martens, E.C., Gordon, J.I. and Smith, T.J. (2008) 'Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices', *Structure*, 16(7), pp. 1105-15.

Kraulis, J., Clore, G.M., Nilges, M., Jones, T.A., Pettersson, G., Knowles, J. and Gronenborn, A.M. (1989) 'Determination of the three-dimensional solution structure of the C-terminal domain of cellobiohydrolase I from *Trichoderma reesei.* A study using nuclear magnetic resonance and hybrid distance geometry-dynamical simulated annealing', *Biochemistry*, 28(18), pp. 7241-57.

Kurita, T., Noda, Y., Takagi, T., Osumi, M. and Yoda, K. (2011) 'Kre6 protein essential for yeast cell wall beta-1,6-glucan synthesis accumulates at sites of polarized growth', *J Biol Chem*, 286(9), pp. 7429-38.

Larsbrink, J. (2014) 'A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes', *Nature*, 506, pp. 498-502.

Larsbrink, J., Rogers, T.E., Hemsworth, G.R., McKee, L.S., Tauzin, A.S., Spadiut, O., Klinter, S., Pudlo, N.A., Urs, K., Koropatkin, N.M., Creagh, A.L., Haynes, C.A., Kelly, A.G., Cederholm, S.N., Davies, G.J., Martens, E.C. and Brumer, H. (2014) 'A discrete genetic locus confers xyloglucan metabolism in select human gut Bacteroidetes', *Nature*, 506(7489), pp. 498-502.

Latge, J.P. (2010) 'Tasting the fungal cell wall', *Cell Microbiol*, 12(7), pp. 863-72.

Ley, R.E., Backhed, F., Turnbaugh, P., Lozupone, C.A., Knight, R.D. and Gordon, J.I. (2005) 'Obesity alters gut microbial ecology', *Proc Natl Acad Sci U S A*, 102(31), pp. 11070-5.

Lowe, E.C., Basle, A., Czjzek, M., Firbank, S.J. and Bolam, D.N. (2012) 'A scissor blade-like closing mechanism implicated in transmembrane signaling in a Bacteroides hybrid two-component system', *Proceedings of the National Academy of Sciences of the United States of America*, 109(19), pp. 7298-7303.

Lowman, D.W., West, L.J., Bearden, D.W., Wempe, M.F., Power, T.D., Ensley, H.E., Haynes, K., Williams, D.L. and Kruppa, M.D. (2011) 'New Insights into the Structure of (1 -> 3,1 -> 6)-beta-D-Glucan Side Chains in the *Candida glabrata* Cell Wall', *Plos One*, 6(11).

Luis, A.S., Venditto, I., Temple, M.J., Rogowski, A., Basle, A., Xue, J., Knox, J.P., Prates, J.A.M., Ferreira, L.M.A., Fontes, C.M.G.A., Najmudin, S. and Gilbert, H.J. (2013) 'Understanding How Noncatalytic Carbohydrate Binding Modules Can Display Specificity for Xyloglucan', *Journal of Biological Chemistry*, 288(7), pp. 4799-4809.

Mackie, R.I. and White, B.A. (1990) 'Recent Advances in Rumen Microbial Ecology and Metabolism - Potential Impact on Nutrient Output', *Journal of Dairy Science*, 73(10), pp. 2971-2995.

Martens, E.C., Koropatkin, N.M., Smith, T.J. and Gordon, J.I. (2009) 'Complex glycan catabolism by the human gut microbiota: the Bacteroidetes Sus-like paradigm', *J. Biol. Chem.*, 284, pp. 24673-24677.

Martens, E.C., Lowe, E.C., Chiang, H., Pudlo, N.A., Wu, M., McNulty, N.P., Abbott, D.W., Henrissat, B., Gilbert, H.J., Bolam, D.N. and Gordon, J.I. (2011) 'Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts', *PLoS Biol*, 9(12), p. e1001221.

McCarter, J.D. and Withers, S.G. (1994) 'Mechanisms of enzymatic glycoside hydrolysis', *Curr Opin Struct Biol*, 4(6), pp. 885-92.

McCartney, L., Blake, A.W., Flint, J., Bolam, D.N., Boraston, A.B., Gilbert, H.J. and Knox, J.P. (2006) 'Differential recognition of plant cell walls by microbial xylan-specific carbohydrate-binding modules', *Proceedings of the National Academy of Sciences of the United States of America*, 103(12), pp. 4765-4770.

Miyazaki, K., Hirase, T., Kojima, Y. and Flint, H.J. (2005) 'Medium- to large-sized xylo-oligosaccharides are responsible for xylanase induction in Prevotella bryantii B14', *Microbiology*, 151(Pt 12), pp. 4121-5.

Montanier, C., van Bueren, A.L., Dumon, C., Flint, J.E., Correia, M.A., Prates, J.A., Firbank, S.J., Lewis, R.J., Grondin, G.G., Ghinet, M.G., Gloster, T.M., Herve, C., Knox, J.P., Talbot, B.G., Turkenburg, J.P., Kerovuo, J., Brzezinski, R., Fontes, C.M.G.A., Davies, G.J., Boraston, A.B. and Gilbert, H.J. (2009) 'Evidence that family 35 carbohydrate binding modules display conserved specificity but divergent function', *Proceedings of the National Academy of Sciences of the United States of America*, 106(9), pp. 3065-3070.

Montanier, C.Y., Correia, M.A., Flint, J.E., Zhu, Y., Basle, A., McKee, L.S., Prates, J.A., Polizzi, S.J., Coutinho, P.M., Lewis, R.J., Henrissat, B., Fontes, C.M. and Gilbert, H.J. (2011) 'A novel, noncatalytic carbohydrate-binding module displays specificity for galactose-containing polysaccharides through calcium-mediated oligomerization', *J Biol Chem*, 286(25), pp. 22499-509.

Mpofu, C.M., Campbell, B.J., Subramanian, S., Marshall-Clarke, S., Hart, C.A., Cross, A., Roberts, C.L., Mcgoldrick, A., Edwards, S.W. and Rhodes, J.M. (2007) 'Microbial mannan inhibits bacterial killing by macrophages: A possible pathogenic mechanism for Crohn's disease', *Gastroenterology*, 133(5), pp. 1487-1498.

Mullis, K.B. and Faloona, F.A. (1987) 'Specific Synthesis of DNA Invitro Via a Polymerase-Catalyzed Chain-Reaction', *Methods in Enzymology*, 155, pp. 335-350.

Najmudin, S., Guerreiro, C.I.P.D., Carvalho, A.L., Prates, J.A.M., Correia, M.A.S., Alves, V.D., Ferreira, L.M.A., Romao, M.J., Gilbert, H.J., Bolam, D.N. and Fontes, C.M.G.A. (2006) 'Xyloglucan is recognized by carbohydrate-binding modules that interact with beta-glucan chains', *Journal of Biological Chemistry*, 281(13), pp. 8815-8828.

Nakajima, T. and Ballou, C.E. (1974) 'Degradation cerevisiae Mannan by Alkaline Fragments Obtained from *Saccharomyces*', *J. Biol. Chem.*, 249, pp. 7679-7684.

Okamoto, T., Sasaki, M., Tsujikawa, T., Fujiyama, Y., Bamba, T. and Kusunoki, M. (2000) 'Preventive efficacy of butyrate enemas and oral administrationof *Clostridium butyricum* M588 in dextran sodium sulfate-induced colitis in rats', *J Gastroenterol*, 35, pp. 341-346.

Oyama, S., Yamagata, Y., Abe, K. and Nakajima, T. (2002) 'Cloning and expression of an endo-1,6-beta-D-glucanase gene (neg1) from *Neurospora crassa*', *Biosci Biotechnol Biochem*, 66(6), pp. 1378-81.

Paal, K., Ito, M. and Withers, S.G. (2004) 'Paenibacillus sp. TS12 glucosylceramidase: kinetic studies of a novel sub-family of family 3 glycosidases and identification of the catalytic residues', *Biochem J*, 378(Pt 1), pp. 141-9.

Pauly, M., Gille, S., Liu, L., Mansoori, N., de Souza, A., Schultink, A. and Xiong, G. (2013) 'Hemicellulose biosynthesis', *Planta*, 238(4), pp. 627-42.

Peeters, M., Joossens, S., Vermeire, S., Vlietinck, R., Bossuyt, X. and Rutgeerts, P. (2001) 'Diagnostic value of anti-*Saccharomyces cerevisiae* and antineutrophil cytoplasmic autoantibodies in inflammatory bowel disease', *American Journal of Gastroenterology*, 96(3), pp. 730-734.

Prins, R.A., Hungate, R.E., Vanvugt, F. and Vanvorst.Cj (1972) 'Comparison of Strains of Eubacterium-Cellulosolvens from Rumen', *Antonie Van Leeuwenhoek Journal of Microbiology and Serology*, 38(2), pp. 153-&.

Qin, J.J., Li, R.Q., Raes, J., Arumugam, M., Burgdorf, K.S., Manichanh, C., Nielsen, T., Pons, N., Levenez, F., Yamada, T., Mende, D.R., Li, J.H., Xu, J.M., Li, S.C., Li, D.F., Cao, J.J., Wang, B., Liang, H.Q., Zheng, H.S., Xie, Y.L., Tap, J., Lepage, P., Bertalan, M., Batto, J.M., Hansen, T., Le Paslier, D., Linneberg, A., Nielsen, H.B., Pelletier, E., Renault, P., Sicheritz-Ponten, T., Turner, K., Zhu, H.M., Yu, C., Li, S.T., Jian, M., Zhou, Y., Li, Y.R., Zhang, X.Q., Li, S.G., Qin, N., Yang, H.M., Wang, J., Brunak, S., Dore, J., Guarner, F., Kristiansen, K., Pedersen, O., Parkhill, J., Weissenbach, J., Bork, P., Ehrlich, S.D., Wang, J. and Consortium, M. (2010) 'A human gut microbial gene catalogue established by metagenomic sequencing', *Nature*, 464(7285), pp. 59-U70.

Quigley, E.M.M. (2010) 'Probiotics in gastrointestinal disorders', *Hosp. Pract.*, 38, pp. 122-129.

Quinton, J.F., Sendid, B., Reumaux, D., Duthilleul, P., Cortot, A., Grandbastien, B., Charrier, G., Targan, S.R., Colombel, J.F. and Poulain, D. (1998) 'Anti-*Saccharomyces cerevisiae* mannan antibodies

combined with antineutrophil cytoplasmic autoantibodies in inflammatory bowel disease: prevalence and diagnostic role', *Gut*, 42(6), pp. 788-791.

Rogers, G.B. and Bruce, K.D. (2012) 'Exploring the parallel development of microbial systems in neonates with cystic fibrosis', *MBio*, 3(6), pp. e00408-12.

Rogowski, A., Briggs, J.A., Mortimer, J.C., Tryfona, T., Terrapon, N., Lowe, E.C., Basle, A., Morland, C., Day, A.M., Zheng, H.J., Rogers, T.E., Thompson, P., Hawkins, A.R., Yadav, M.P., Henrissat, B., Martens, E.C., Dupree, P., Gilbert, H.J. and Bolam, D.N. (2015) 'Glycan complexity dictates microbial resource allocation in the large intestine', *Nature Communications*, 6.

Rye, C.S. and Withers, S.G. (2000) 'Glycosidase mechanisms', *Curr Opin Chem Biol*, 4(5), pp. 573-80.

Samuel, B.S. and Gordon, J.I. (2006) 'A humanized gnotobiotic mouse model of host-archaeal-bacterial mutualism', *Proc Natl Acad Sci U S A*, 103(26), pp. 10011-6.

Scheller, H.V. and Ulvskov, P. (2010) 'Hemicelluloses', *Annu Rev Plant Biol*, 61, pp. 263-89.

Shen, Y.B., Piao, X.S., Kim, S.W., Wang, L., Liu, P., Yoon, I. and Zhen, Y.G. (2009) 'Effects of yeast culture supplementation on growth performance, intestinal health, and immune response of nursery pigs', *J Anim Sci*, 87(8), pp. 2614-24.

Shipman, J.A., Berleman, J.E. and Salyers, A.A. (2000) 'Characterization of Four Outer Membrane Proteins Involved in Binding Starch to the Cell Surface of *Bacteroides thetaiotaomicron*', *J. Bacteriol.*, 182(19), pp. 5365-5372.

Simpson, P.J., Xie, H., Bolam, D.N., Gilbert, H.J. and Williamson, M.P. (2000) 'The structural basis for the ligand specificity of family 2 carbohydrate-binding modules', *J Biol Chem*, 275(52), pp. 41137-42.

Sokol, H., Lay, C., Seksik, P. and Tannock, G.W. (2008) 'Analysis of bacterial bowel communities of IBD patients: what has it revealed?', *Inflamm Bowel Dis*, 14(6), pp. 858-67.

Sonnenburg, E.D., Zheng, H., Joglekar, P., Higginbottom, S.K., Firbank, S.J., Bolam, D.N. and Sonnenburg, J.L. (2010) 'Specificity of polysaccharide use in intestinal bacteroides species determines diet-induced microbiota alterations', *Cell*, 141(7), pp. 1241-52.

Suits, M.D., Zhu, Y., Taylor, E.J., Walton, J., Zechel, D.L., Gilbert, H.J. and Davies, G.J. (2010) 'Structure and kinetic investigation of *Streptococcus pyogene*s family GH38 alpha-mannosidase', *PLoS One*, 5(2), p. e9006.

Thompson, A.J., Speciale, G., Iglesias-Fernandez, J., Hakki, Z., Belz, T., Cartmell, A., Spears, R.J., Chandler, E., Temple, M.J., Stepper, J., Gilbert, H.J., Rovira, C., Williams, S.J. and Davies, G.J. (2015) 'Evidence for a boat conformation at the transition state of GH76 alpha-1,6-mannanases--key enzymes in bacterial and fungal mannoprotein metabolism', *Angew Chem Int Ed Engl*, 54(18), pp. 5378-82.

Tremaroli, V. and Backhed, F. (2012) 'Functional interactions between the gut microbiota and host metabolism', *Nature*, 489(7415), pp. 242-9.

Turnbaugh, P.J., Ridaura, V.K., Faith, J.J., Rey, F.E., Knight, R. and Gordon, J.I. (2009) 'The effect of diet on the human gut microbiome: a metagenomic analysis in humanized gnotobiotic mice', *Sci Transl Med*, 1(6), p. 6ra14.

Turnbull, W.B. and Daranas, A.H. (2003) 'On the value of c: Can low affinity systems be studied by isothermal titration calorimetry?', *Journal of the American Chemical Society*, 125(48), pp. 14859-14866.

Turner, J.R. (2009) 'Intestinal mucosal barrier function in health and disease', *Nat Rev Immunol*, 9(11), pp. 799-809.

Vaaje-Kolstad, G., Westereng, B., Horn, S.J., Liu, Z., Zhai, H., Sorlie, M. and Eijsink, V.G. (2010) 'An oxidative enzyme boosting the enzymatic conversion of recalcitrant polysaccharides', *Science*, 330(6001), pp. 219-22.

Venditto, I., Basle, A., Luis, A.S., Temple, M.J., Ferreira, L.M., Fontes, C.M., Gilbert, H.J. and Najmudin, S. (2013) 'Overproduction, purification, crystallization and preliminary X-ray characterization of the C-terminal family 65 carbohydrate-binding module (CBM65B) of endoglucanase Cel5A from Eubacterium cellulosolvens', *Acta Crystallogr Sect F Struct Biol Cryst Commun*, 69(Pt 2), pp. 191-4.

Vogel, J. (2008) 'Unique aspects of the grass cell wall', *Curr Opin Plant Biol*, 11(3), pp. 301-7.

von Schantz, L., Hakansson, M., Logan, D.T., Nordberg-Karlsson, E. and Ohlin, M. (2014) 'Carbohydrate binding module recognition of xyloglucan defined by polar contacts with branching xyloses and CH-Pi interactions', *Proteins-Structure Function and Bioinformatics*, 82(12), pp. 3466-3475.

Wachtershauser, A. and Stein, J. (2000) 'Rationale for the luminal provision of butyrate in intestinal diseases', *European Journal of Nutrition*, 39(4), pp. 164-171.

Whitney, S.E., Gidley, M.J. and McQueen-Mason, S.J. (2000) 'Probing expansin action using cellulose/hemicellulose composites', *Plant J*, 22(4), pp. 327-34.

Xu, J., Mahowald, M.A., Ley, R.E., Lozupone, C., Hamady, M., Martens, E.C., Henrissat, B., Coutinho, P.M., Minx, P., Latreille, P., Cordum, H., Van Brunt, A., Kim, K., Fulton, L.A., Clifton, W.S., Wilson, R.K., Knight, R. and Gordon, J.I. (2007) 'Evolution of Symbiotic Bacteriain the Distal Human Intestine', *PLoS Biol*, 5(7).

Yoda, K., Toyoda, A., Mukoyama, Y., Nakamura, Y. and Minato, H. (2005) 'Cloning, sequencing, and expression of a *Eubacterium cellulosolvens* 5 gene encoding an endoglucanase (Cel5A) with novel carbohydrate-binding modules, and properties of Cel5A', *Applied and Environmental Microbiology*, 71(10), pp. 5787-5793.

Yoon, T.J., Kim, T.J., Lee, H., Shin, K.S., Yun, Y.P., Moon, W.K., Kim, D.W. and Lee, K.H. (2008) 'Anti-tumor metastatic activity of beta-glucan purified from mutated *Saccharomyces cerevisiae*', *Int Immunopharmacol*, 8(1), pp. 36-42.

Zechel, D.L. and Withers, S.G. (2000) 'Glycosidase mechanisms: anatomy of a finely tuned catalyst', *Acc Chem Res*, 33(1), pp. 11-8.

Zhang, X., Rogowski, A., Zhao, L., Hahn, M.G., Avci, U., Knox, J.P. and Gilbert, H.J. (2014) 'Understanding how the complex molecular architecture of mannan-degrading hydrolases contributes to plant cell wall degradation', *J Biol Chem*, 289(4), pp. 2002-12.

Zhu, Y., Suits, M.D., Thompson, A.J., Chavan, S., Dinev, Z., Dumon, C., Smith, N., Moremen, K.W., Xiang, Y., Siriwardena, A., Williams, S.J., Gilbert, H.J. and Davies, G.J. (2010) 'Mechanistic insights into a Ca2+-dependent family of alpha-mannosidases in a human gut symbiont', *Nat Chem Biol*, 6(2), pp. 125-32.

Ziegler, F.D., Gemmill, T.R. and Trimble, B.R. (1994) 'Glycoprotein Synthesis in Yeast', *JBC*, 269, pp. 12527-12535.

Appendix A

Protein	GH Family	Signal peptide (SpI or SpII)	Predicted cleavage posistion in protein sequence
BT1769	GH92	SpII	18/19
BT1878		SpII	23/24
BT2199		SpII	18/19
BT3130		SpI	20/21
BT3773		SpI	22/23
BT3858		SpII	23/24
BT3990		SpI	19/20
BT3991		SpII	19/20
BT3994		SpII	16/17
BT4072	GH38	SpI	20/21
BT4073	GH92	SpII	19/20

lipoP analysis of *B. theta* encoded, putative extracellular α-mannosidases used to determine debranching activity on yeast α-mannan

Data collection	CBM65B-XXXG
Source	I04
Wavelength (Å)	0.9795
Space group	$P4_3\,2_1\,2$
Cell dimensions	
a, *b*, *c* (Å)	57.9, 57.9, 116.7
α, β, γ (°)	90.0, 90, 90.0
No. of unique observations	8,856
Multiplicity	6.2 (4.7)
Resolution (Å)	58.37-2.35 (2.48-2.35)
R_{merge} (%)	7.8 (66.1)
$I/\sigma I$	13.4 (2.0)
Completeness (%)	100.0 (100.0)
Refinement	
R_{work}/ R_{free}	22.44 (28.03)
No. atoms	
Protein	955
Ligand	72.
Water	1
B-factors	
Protein	50.4
Ligand	47.6
Water	39.5
R.m.s deviations	
Bond lengths (Å)	0.012
Bond angles (º)	1.56

Data collection and refinement statistics for CBM65B

Data collection	BT3312	BT3312-Inhibitor
Source	I02	I03
Wavelength (Å)	0.9794	0.7749
Space group	$P2_1$	$P2_1$
Cell dimensions		
a, *b*, *c* (Å)	62.4, 156, 78.0	62.4, 78.8, 145.5
α, β, γ (°)	90.0, 100.4, 90.0	90.0, 94.9, 90.0
No. of measured reflections	429705 (10374)	445926 (22895)
No. of independent reflections	114209 (4871)	116665 (5762)
Resolution (Å)	48.24-1.90 (1.93-1.90)	48.79-1.85 (1.88-1.85)
R_{merge} (%)	6.7 (57.8)	9.9 (78.4)
$CC_{1/2}$	1.00 (0.47)	1.00 (0.64)
$I/\sigma I$	9.0 (1.8)	7.8 (1.6)
Completeness (%)	99.2 (86.2)	97.4 (97.6)
Redundancy	3.8 (2.1)	3.8 (4.0)
Refinement		
R_{work}/ R_{free}	19.8 (24.8)	18.7 (23.2)
No. atoms		
Protein	10497	10675
Ligand	N.A.	66
Water	705	616
B-factors		
Protein	34.3	30.6
Ligand	N.A.	29.15
Water	39.6	39.9
R.m.s deviations		
Bond lengths (Å)	0.017	0.018
Bond angles (º)	1.82	1.79

Data collection and refinement statistics for BT3312 (GH30).

Appendix B – Chemicals, media and enzymes used in this study

B1 Chemicals

Biogene

- Electrophoresis grade Agarose

British Drug Houses (BDH)

- Acetic acid (glacial)
- Acrylamide solution (40 % (w/v); Electran)
- Bromophenol blue
- Citric acid
- Calcium chloride
- Ethanol
- Hydrochloric acid
- Isopropanol
- Magnesium chloride
- Magnesium sulphate
- Methanol
- Potassium dihydrogen orthophosphate
- Sodium acetate
- Sodium chloride
- Sulphuric acid
- Triton X-100

Dextra

- $Man_9(GlcNAc)_2$

Fisons

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

Megazyme

- Barley β-glucan
- Cellobiose
- Cellotriose
- Cellotetraose

Cellopentaose

Cellohexaose

Hydroxyethylcellulose (HEC)

Laminarbiose

Melibiose

Pustulan

Raffinose

Stachyose

Xyloglucan heptasaccharide (XXXG)

Xyloglucan

Melford Laboratories

Isopropyl-β-D-thiogalactoside (IPTG)

HEPES

GE Health Care

Agarose (ultrapure)

Sigma

3,5-Dinitrosalasylic acid (DNSA)

4-nitophenyl substrates (pNP)

Ammonium persulphate

Ampicillin

Bovine serum albumin (BSA)

Coomassie Brilliant Blue G

Copper sulphate

D-Galactose

D-Glucose

D-Mannose

di-Sodium hydrogen phosphate

Ethelene diamine tetra-acetid acid, disodium salt (EDTA)

Ethidium bromide

Glycerol

Imidazole

Kanamycin

L-Fucose

N,N,N',N,- Tetramethylethylene diamine (TEMED)

Potassium Chloride

Potassium thiocyanate

Rochelle salts

Sodium bicarbonate

Sodium carbonate

Sodium dodecyl sulphate (SDS)

Yeast mannan

Zinc Sulphate

β-Mercaptoethanol

B2 Media

Difco

Bacto®tryptone

Bacto®yeast extract

Oxioid

Bacteriological Agar No. 1

B3 Enzymes

MBI Fermentas

DNA restriction endonucleases

Invitrogen

Bacteriophage T4 DNA ligase

Megazyme

Glucose-6-phosphate dehydrogenase

Hexokinase

Phosphoglucose isomerase

Phosphomannose isomerase

Novagen

KOD HotStart DNA polymerase

Stratagene

Dpn1 restriction endonuclease

B4 DNA

MWG Biotech

All primers

B5 Kits

Qiagen

Plasmid mini kit

Plasmid midi kit

Qiaquick Gel extraction kit

Qiaquick PCR purification kit