Mechanisms by which Glycoside Hydrolases Recognize Plant, Bacterial and Yeast Polysaccharides

A thesis submitted for the degree of Doctor of Philosophy

Newcastle University

2008-2012

Fiona Cuskin

Institute for Cell and Molecular Biosciences

Acknowledgements

I would like to thank my supervisor Harry Gilbert, for the opportunity to study a PhD within his lab. I would also like to thank Harry, for his invaluable guidance, support and patience throughout my studies.

The many people past and present I have worked with in Harry's lab have provided invaluable advice and friendship and would like to thank Carl Morland, Dave Bolam, Elisabeth Lowe, Arthur Rogowski, Adam Jackson, Alan Cartmell, Lauren Mckee, Joanna Norman, Yanping Zhu and Wade Abbott to name a few. This work would not have been possible if it was not for the collaboration and expertise of Dr Arnaud Basle for the X-ray crystallography and Dr Alexandra Solovyova for the Analytical Ultracentrifugation.

Finally, I would like to thank my family and friends for the love and support over the last 4 years.

Abstract

The deconstruction of complex carbohydrates by glycoside hydrolases requires extensive enzyme consortia in which specificity is often conferred by accessory modules and domains that are distinct from the active site. The diverse mechanisms of substrate recognition were explored in this thesis using selected yeast, bacterial and plant polysaccharides as example substrates.

Carbohydrate binding modules (CBM) are non-catalytic modules that enhance the catalytic activity of their glycoside hydrolase counterparts through binding to polysaccharide. Normally CBMs are found attached to glycoside hydrolases that target insoluble recalcitrant substrates resulting in a moderate, 2-5 fold, potentiation in enzyme activity. A CBM, defined herein as CBMX40, is found at the C-terminal of a glycoside hydrolase family (GH) 32 enzyme, SacC, which displays exo-levanase activity. CBMX40 binds the non-reducing end of the levan chain targeting the disaccharide fructose-2,6-fructose unit. Removal of CBMX40 results in a >100-fold decrease in catalytic activity against levan, compared to the full length native enzyme. The truncated SacC catalytic domain acts as a non-specific exo-β-fructosidase displaying similar activity on β2,1- (inulin) and β2,6-linked fructose polymers, both polysaccharides and oligosaccharides. When CBMX40 was fused to a non-related exo-β-fructosidase, BT 3082, it conferred exo-levanase specificity on the enzyme. Thus CBMX40 is not only able to enhance catalytic activity but is also able to confer catalytic specificity. This led to the hypothesis that the CBM and the active site of the enzyme bind to different terminal residues of branched fructans such as levan. This results in enhanced affinity through avidity effects leading to the potentiation of catalytic activity.

The gut bacterium *Bacteroides thetaiotaomicron* contributes to the maintenance of a healthy human gut. *B. thetaiotaomicron* is able to acquire and utilise complex carbohydrates that are not attacked by the intestinal enzymes of the host. *B. thetaiotaomicron* dedicates a large proportion of its genome to glycan degradation with a large expansion of α-mannan degrading enzymes. The *B. thetaiotaomicron* genome encodes 23 GH92 α-mannanosidases and 10 GH76 α-mannanases. While GH92 has recently been characterised the activities displayed by GH76 relies on the characterization of a single enzyme in this family. *B. thetaiotaomicron* organises the genes required to sense, degrade, transport and utilise specific complex glycans into genetic clusters defined as Polysaccharide Utilisation Loci (PULs). Transcriptomics revealed that two PULs are up regulated in response to yeast mannan, PUL 36 and PUL 68. These PULs contain both GH76 enzymes along with GH92 enzymes and other CAZy annotated enzymes. Biochemical analysis of the GH76 enzymes found in the two PULs show they are α1, 6 mannanases capable of hydrolysing the α1, 6 mannan backbone of yeast mannan, with the putative periplasmic enzymes generating small oligosaccharides, while the surface mannanases releasing larger products. The three GH92 enzymes encoded by the two PULs have been shown to remove α1, 2 and α1, 3 linked mannose branches from yeast mannan polysaccharide. In addition PUL 68 also encodes a phosphatase that removes the phosphate from mannose-6-phosphate and glucose-6-phosphate but not from intact mannan. Therefore, this study describes the ability of *B. thetaiotaomicron* to target and degrade yeast α-mannans.

The GH5 enzyme *Ct*Xyl5A from *Clostridium thermocellum* is an arabinoxylan specific xylanase that contains a GH5 catalytic module appended to several CBMs. The apo structure of the GH5 catalytic module appended to a family 6 CBM reveals a large pocket abutted to the -1 subsite of the active site. This pocket was thought to bind the arabinose decoration appended to the O3 of the xylan backbone. Here mutational and structural studies showed that the fulfilment of arabinose is this pocket is the key specificity determinant for the novel arabinoxylanase activity. Significantly the bound arabinose displayed a pyranose conformation, rather than a furanose structure which is the typical conformation adopted by arabinose side chains in arabinoxylans. This structural information suggests that *Ct*Xyl5A may be able to exploit side chains other than arabinofuranose residues as substrate specificity determinants.

Contents

List of Figures

List of Tables

Chapter 1: Introduction

Carbohydrates are biologically important highly diverse molecules. As the term carbohydrates suggest the three defining elements of these molecules are carbon, oxygen and hydrogen. Other defining parameters of carbohydrates are the presence of a hydroxyl group and either a aldehyde or ketone group. Monosaccharides are the smallest and simplest sugars and provide the building blocks for more complicated oligosaccharides and polysaccharides. Monosaccharides with a carbon chain of 5 or more atoms adopt a cyclic form rather than an open chain conformation. Cyclic sugars can either adopt a pyranose conformation, a six membered ring, or a furanose conformation, a five membered ring. Monosaccharides are linked through a glycosidic bond where a sugar is bound through the anomeric carbon to the hydroxyl of another sugar. The complexity and diversity of polysaccharides is achieved through the range of diverse monosaccharides and the stereochemistry and location of the glycosidic linkages (Stryer , 2002).

Carbohydrates have critical roles in many biological processes including central metabolism, cell signalling, protein folding and targeting, while also fulfilling a structural and storage function. Complex carbohydrates are the primary substrates for biofuel synthesis, and are important components of the pharmaceutical and food industries. A thorough knowledge of how polysaccharides are deconstructed in nature is vital to understanding how they can be exploited within an industrial context. This review of carbohydrate structure will focus on the polysaccharides that are targeted by the proteins analysed in this thesis. In addition the proteins and bacterial systems discussed in this introduction are relevant to the material analysed in this thesis.

1.1 Complex polysaccharides structures and functions

1.1.1 Fructan polysaccharides

Fructans, polysaccharides comprising predominantly of fructose units, are found in both microbes and plants. They are predominately used as storage polymers; after starch they are the most abundant non structural polysaccharides (Hendry, 1993).

The β 2,1 fructose polymer, inulin, is found mainly in plants where is fulfils a storage role. The inulin chains can be between 20-100 fructose (Fru) units and usually have a single glucose (Glc) linked through an α2,1-glycosidic bond, capping the reducing end of the fructan chain (Roberfroid, 2005). Sucrose is used as the starting block for the synthesis of the inulin chain, which accounts for the terminal glucose. The disaccharide is cleaved by a fructosyltransferase and the released fructose is linked β2,1-to an acceptor Fru, which, in the first biosynthetic step, is a component of a sucrose (Suc) molecule to generate Fru-β2,1-Suc (1-kestose). Each subsequent round of catalysis adds a further Fru to generate, for example Fru-β2,1-Fru-β2,1-Suc in the second round of catalysis leading to (Fru-β2,1-Fru)$_n$-α2,1-Suc after *n* rounds of catalysis (French, 1989). Thus, the reducing end of the fructan chain is capped with a Glc residue. In plants fructose is generally more efficiently transferred to the acceptor 1-kestose oligosaccharide from another donor 1-kestose rather than a sucrose molecule. Figure 1.1 shows the structure of inulin. Therefore, as the chain elongates during the synthesis of inulin the efficiency to accept fructose increases, which results in the production of longer chain inulins rather than many small oligosaccharides (Franck and De Leenheer, 2005). The terminal Glc serves to prevent the inulin chains from spontaneous degradation. These two characteristics are important as inulin is exploited as an energy storage polysaccharide in plants.

Inulin is a linear polysaccharide and has the potential for branches at the C6 of the fructose units; this however is unusual in plants, but does occur in microbes, albeit at low levels. The small inulin oligosaccharides do not adopt any particular molecular conformation; however inulin chains of degree of polymerization (DP) 9 or above form a six fold helical structure that can pack into various isoforms. The molecular conformation of inulin is thought to be fairly variable and while the right handed helix is favoured it can also form left handed helices (French, 1988).

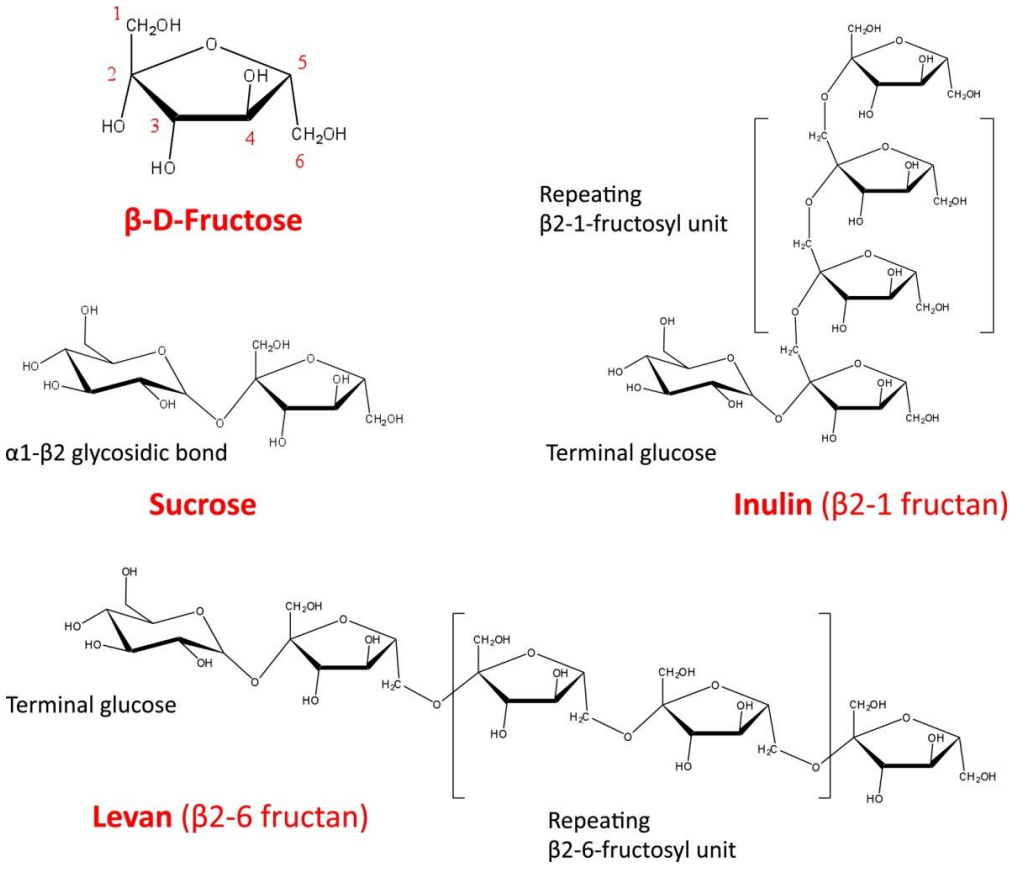


Figure 1.1 Inulin and levan structure

Figure taken from Sonnenburg et al., 2010

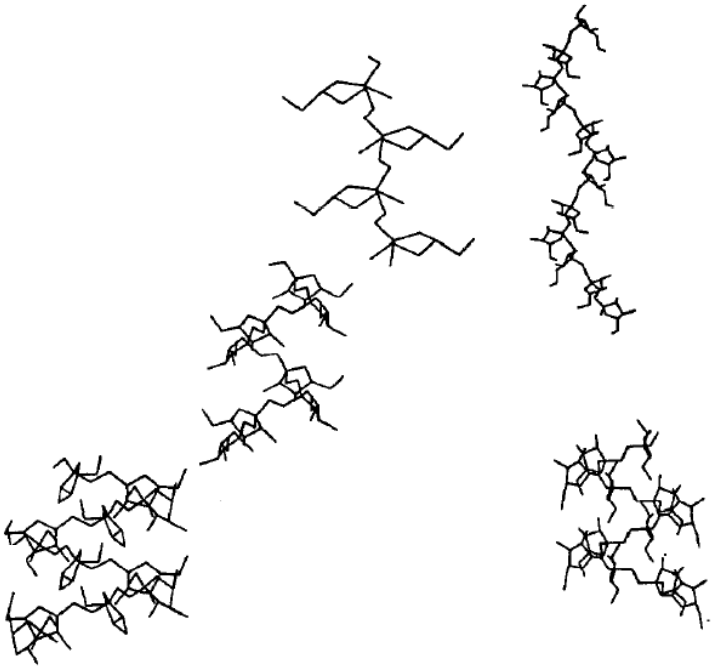


Figure 1.2 Predicted molecular conformations adopted by inulin.

Helical conformations formed by inulin in solution, demonstrates the flexibility of inulin in solution. Figure taken from (French, 1988)

Inulin, which is produced mainly in dicotyledons plants such as Jerusalem artichoke and chicory roots, has numerous applications both biologically and industrially (Hendry, 1993). Inulin traverses the human digestive system and reaches the large bowel and thus provides an energy source for the flora of the gut. A balanced gut flora is very important in human health and has been linked both with obesity and diabetes. The food industry therefore markets inulin as a prebiotic as it promotes growth of “friendly” bacteria (Lopez-Molina et al., 2005). The food industry also uses inulin as a natural sweetener in low fat foods due to its sweet taste and indigestible nature (Stevens et al., 2001). Medically inulin can be exploited in kidney test functions as, when administered intravenously, it is rapidly excreted but is not metabolised or reabsorbed, providing an effective way to measure glomerular filtration. Inulin is able to form a gel in aqueous solution, a characteristic that is

utilised by pharmaceutical companies to deliver drugs to the colon (Vervoort et al., 1997). The protective layer of inulin gel is only destroyed once it reaches the bacteria in the large intestine and so releasing the desired drugs intact to the target location. Chemical applications of inulin include its microbial fermentation to ethanol and, more frequently, as a precursor to glycerol production (Fuchs, 1987).

The β-2,6 fructose polymer levan, shown in Figure 1.1, is predominately found in microorganisms, but low molecular weight forms of the fructan can be synthesised by some monocotyledons plants (Han, 1990). Synthesis of levan occurs through the addition of fructose via a β-2,6 linkage to a sucrose acceptor molecule by levansucrases (Deonder, 1966, Fuchs, 1956). Each round of catalysis grows the fructan chain appended to sucrose as described for inulin. The smallest levan oligosaccharide is termed 6-kestose and is comprised of a sucrose molecule with a fructose attached through a β2,6 glycosidic bond. As plant levans are low molecular weight they are predominately linear chains with very little branching. However, the high molecular weight of microbial levans are highly branched through the attachment of β-2,6 fructose chains, through a β 2,1-linkage, to the polysaccharide backbone. The degree of branching and molecular weight of levans vary between microorganisms, although generally bacterial levans have a high molecular weight and are highly branched. The conformation adopted by levans is thought to be more restricted than inulin, favouring a left hand helix.

Plant levan like inulin is used as a storage polysaccharide and is also found mainly within roots (Pontis and Del Campillo, 1985). Microbial levan is secreted by bacteria to form part of the capsule and often a biofilm layer. Bacteria producing levan include soil *Bacillus sp* and oral prokaryotes such as *Streptococcus salivarius,* which produce levan in dental plaque (Higuchi et al., 1970). Biologically levan exhibits

similar properties to inulin in that it is indigestible to human gut enzymes and so can be exploited in a similar way as prebiotics and low calorie food. Levan also has industrial applications, such as in the emulsification and stabilization of foods (Han, 1990). However, in plants levan is not overly abundant and so this source of the polysaccharide is not used in biotechnological applications. Therefore levan microbes are more widely used, not just because it is more abundant but also because high and low molecular weight forms of the fructan can be produced. However, while microbial levan is favoured, this source of the polysaccharide is not entirely without difficulty. Firstly, the choice of microbe used to produce levan must take into account the complexity and molecular weight required as well as selecting an organism that produces a high yield of the fructan. The microbe must be grown in suitable sucrose medium at optimal growth conditions. The levan is extracted fairly easily by alcohol precipitation (Tanaka et al., 1980, Han and Clarke, 1990). At a small scale laboratory level levan can be produced and isolated with ease. To produce quantities of levan needed at an industrial scale the process becomes logistically complex, financially and energy costly. For these reasons levan is not utilised to its full potential and inulin is often preferred due to its availability.

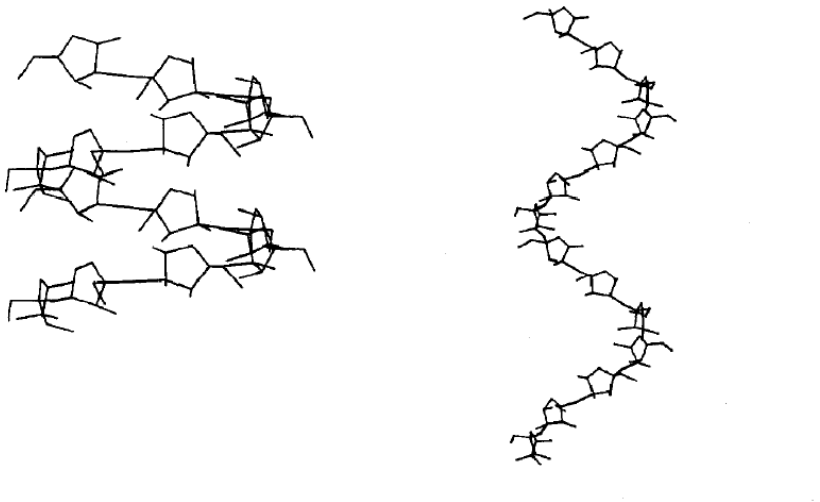


Figure 1.3 Predicted molecular conformations adopted by levan

The two allowed molecular conformations of levan in solution exhibits the restriction of levan structures in solution. Figure taken from (French, 1988)

1.1.2 Mannan from *Saccharomyces cerevisiae*

Saccharomyces cerevisiae and other yeasts produce a protective capsule termed the cell envelope. The cell envelope consists of three components: the cell wall, the periplasmic space and the plasma membrane. The cell wall is the outermost part of the cell envelope and one of the major constituents of the cell wall is polysaccharides. Glucans and α-mannans (defined henceforth as mannans) are the major polysaccharides found in yeast walls, although there is a small amount of chitin.

The mannan is attached to proteins within the cell wall mainly through an asparagine. The mannan polysaccharide comprises of two structures the core oligosaccharide and the outer chain oligosaccharide. Two N-acetylglucosamines link the polysaccharide to the protein. The first mannose of the core oligosaccharide is bound to the non-reducing N-acetylglucosamine through a β1,4 glycosidic bond. To the trisaccharide eight mannose units are attached though various α-1,6 ,α-1,3 and

α-1,2 linkages (Nakajima and Ballou, 1974). The core oligosaccharide is identical to the high mannose N-glycans found in higher eukaryotes, apart from an additional α-1,3 linked mannose unit. From the core oligosaccharide the outer chain is linked through an α-1,6 mannose backbone with α-1,2 side chains that are generally capped with an α-1,3 linked mannose. Figure 1.4 depicts the yeast mannan structure (Hernandez et al., 1989a) . The side chains of yeast mannans can also be phosphorylated. The regulation and synthesis of mannan is controlled by various glycosytransferases. The transferases localized at the endoplasmic reticulum synthesise the core oligosaccharide, while Golgi-localized transferases synthesise the outer chain (Kukuruzinska et al., 1987). The mannosyltransferases responsible for outer chain synthesis have been extensively characterised. Studies of the structures of mannan produced by mutant strains of *S. cerevisiae, in* which mannosyltransferases have been knocked out, have defined the role of each enzyme in mannan biosynthesis. The mnn1 mutation results in a mannan polysaccharide that lacks the α-1,3 linked mannose from both the core and outer chain. In the mnn2 and mnn5 mutants the transferases that synthesise, respectively, the first and second rows of the α-1,2 linked mannose branches of the outer chain are inactive. In the mnn9 mutant the transferase that elongates the outer chain α 1,6 backbone is non-functional. Phosphorylation of the mannan polysaccharide is controlled by enzymes lacking in the mnn4 and mnn6 mutants (Hernandez et al., 1989a, Hernandez et al., 1989b, Lewis and Ballou, 1991, Nakajima and Ballou, 1975, Romero and Herscovics, 1989).

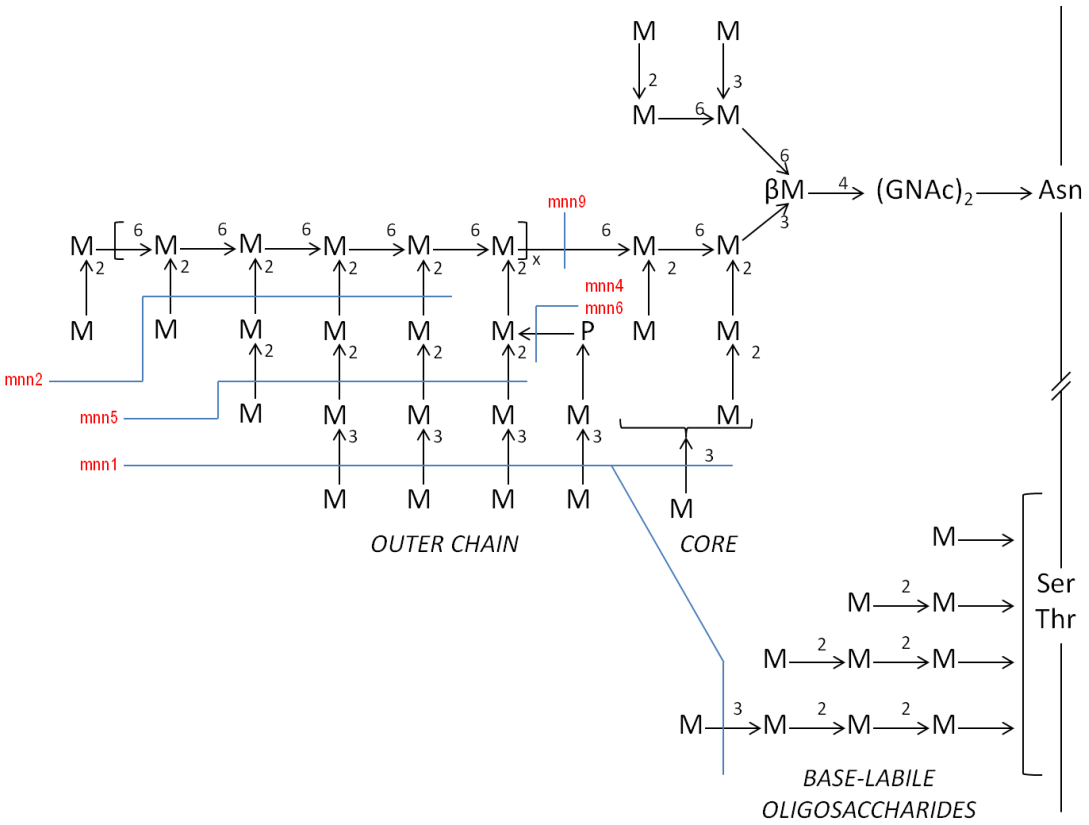


Figure 1.4 Structure of Yeast Mannan

M= mannose residues and numbers represent the linkage. Blue lines are an indication of the changes by mnn mutations which are labelled in red.

People affected by Chron's disease often have systemic antibodies against *S. cerevisiae.* These antibodies recognise the N-linked mannan found in the yeast cell wall and consequently yeast mannan is used as an antigen in a diagnostic ELISA to test for the presence of this antibody in patients (Sendid et al., 1998). Yeast derived mannan oligosaccharides are also widely used as a supplement to animal feeds (White et al., 2002). It is used both in farm animals such as pigs and chicken as well as domestic pets such as dogs. Animal feeds supplemented with mannan oligosaccharides have been shown, in a number of studies, to improve gut health and to prevent colonisation of the intestines by pathogenic bacteria. In fact mannan oligosaccharides are so effective they are now used as an alternative to the prophylactic use of antibiotics in animal feed (Baurhoo et al., 2007, Patterson and

Burkholder, 2003, Zhao et al., 2012). Mannan oligosaccharides are thought to prevent infection in the gut as they mimic the cell surface carbohydrates that pathogenic bacteria use as receptors prior to cell invasion. Thus, instead of pathogenic bacteria binding to cell surface carbohydrates and infecting the intestinal epithelium, they bind to the free mannan oligosaccharides in the gut and thus are no longer invasive.

1.1.3 Major plant structural polysaccharides

Arabinoxylan is a hemicellulosic polysaccharide found within the matrix of the plant cell wall. The plant cell wall comprises an array of polysaccharides. The major polysaccharide is cellulose, a β-1,4 linked glucose polymer that is the most abundant organic molecule on the planet. The linked glucose moieties are arranged in linear chains and are typically thousands of glucose units long (Carpita, 1997). They contain ordered inter- and intra-chain hydrogen bonds that result in the formation of microfibrils, which group together to form a highly crystalline structure. Cellulose is a structural component of the cell wall and the β linkage and crystalline nature make it highly inaccessible to enzymatic degradation (Beguin and Aubert, 1994). The cellulose fibrils are embedded in a matrix of hemicelluloses, which are shorter branched polysaccharides that are derived from a number of monosaccharides, primarily mannose, arabinose, xylose, glucose and galactose.

Xylans, a major hemicellulose in the plant cell wall, consists of a backbone of β-1,4 linked D-xylose residues that can contain a range of decorations . Arabinoxylan consists of the β-1,4 linked xylose backbone with L-arabinose side chains linked α-1,2 and/or α-1,3 to the xylose residues. In addition, the xylan backbone can also be decorated with a α-1,2-linked 4-O-methyl-D- glucuronic acid. The xylose

residues of the xylan backbone are in the pyranose configuration whereas the arabinose decorations are furanose sugars. In addition to the sugar decorations of the backbone, the polysaccharide can also be extensively acetylated and the arabinose decorations may be esterified to ferulic acid (Ebringerova et al., 2005) . Figure 1.5 shows the structure and substitutions of xylan.

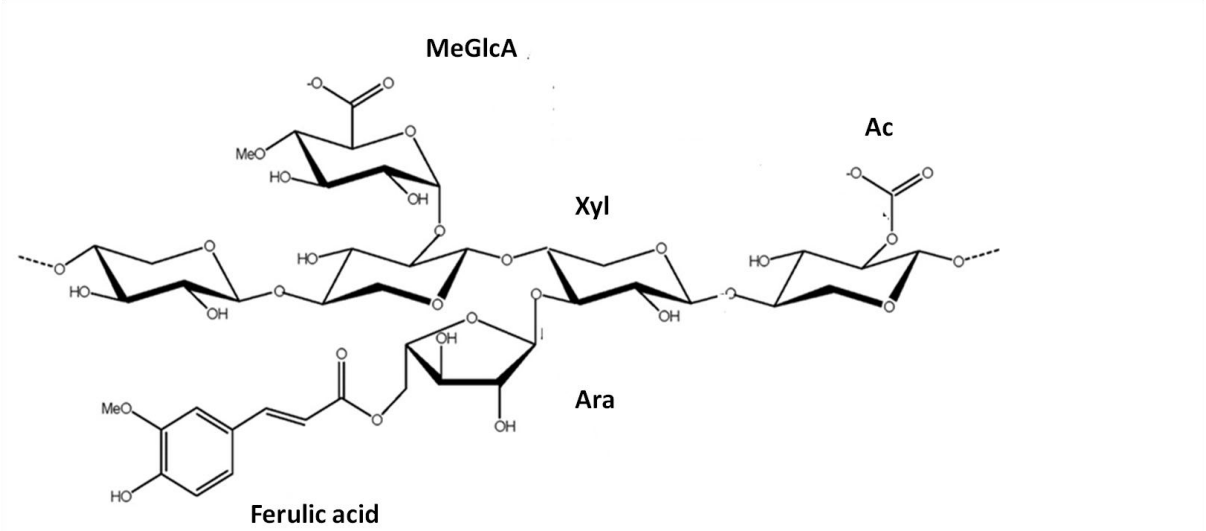


Figure 1.5 Xylan structure

Structure of xylan showing possible substitutions. MeGlcA= 4-O-Methylglucuronic acid Xyl= Xylopyranose Ara= Arabinofuranose Ac= acetyl acid. Figure adapted from (Emami et al., 2009)

Arabinoxylan is particularly prevalent in cereal crops. Wheat and rye arabinoxylan are important components in baking. The xylan backbone in wheat arabinoxylan is heavily substituted with single arabinose residues linked O3 and di-substituents at O2 and O3; a mono arabinofuranose substitution at O2 is very rare (Hoffman et al., 1992). The degree of substitution affects the water solubility and water holding capacity of the polysaccharide which is important in baking. Wheat and rye

arabinoxylan are also a major component of human dietary fibre. Arabinoxylans have been shown to be of particular interest in human health. The addition of soluble arabinoxylan to the diets of patients with diabetes has been shown to decrease the

blood glucose levels and insulin levels; however, the mechanism by which this is achieved is unknown (Lu et al., 2004). One common hypothesis is that the soluble, viscous arabinoxylan decreases the rate of gastric emptying and so delays the absorption of glucose and thus the level of glucose in the blood.

1.2 Glycoside hydrolase

Glycoside hydrolases are a prevalent group of enzymes that hydrolyse glycosidic bonds that link the sugars in complex carbohydrates, or the linkage in glyco-conjugates. Hydrolysis of the glycosidic bond proceeds through one of two acid-base assisted catalytic mechanisms (Sinnott, 1990). Bond cleavage can proceed through either a single displacement or a double displacement reaction. The single and double displacement mechanisms result in the inversion or retention, respectively, of the anomeric configuration at C1 at the site of bond cleavage. Both mechanisms require two carboxylate amino acids for catalysis, an acid/acid-base and a base/nucleophile.

The single displacement or inverting mechanism proceeds through a one step reaction. The charged base extracts a proton from water, the newly formed hydroxyl group acts as a nucleophile and attacks the anomeric carbon of the glycone sugar. The uncharged catalytic acid donates a proton to the glycosidic oxygen encouraging leaving group departure leading to hydrolysis of the glycosidic bond. The hydroxyl must attack in the opposite orientation to the glycosidic bond leading to inversion of anomeric configuration.

The double displacement mechanism is a two-step process that involves the formation of a covalently linked glycosyl-enzyme intermediate. The first reaction, glycosylation, results in the formation of the glycosyl-enzyme intermediate and release of the aglycone sugar from the active site. As with the single displacement mechanism the catalytic acid/base, which functions as the catalytic acid and is uncharged, donates a proton to the glycosidic oxygen promoting leaving group departure. The anomeric carbon is attacked by the charged amino acid nucleophile,

which is thus covalently attached to the glycone sugar through an ester bond in the opposite stereochemistry to the original glycosidic bond. Deglycosylation, the second reaction of the mechanism, results in the hydrolysis of the ester bond between the nucleophile amino acid and the glycone sugar. Deglycosylation results in the release of the second product of the reaction from the active site. Hydrolysis is achieved by the activation of a water molecule by the removal of a proton by the charged catalytic acid/base, which is now acting as the catalytic base. The newly formed hydroxyl ion can now attack the ester bond and release the glycone sugar, resulting in the regeneration of the catalytic nucleophile and replenishing the protonated catalytic acid-base.

Both mechanisms proceed through a positive oxocarbeniuim transition state as leaving group departure and anomeric attack are asynchronous. This transition state is a very high energy molecule and the geometry of the sugar is distorted from the normal chair to a half chair or boat configuration. This results in C1 and the ring oxygen being in the same plane and are thus able to share a positive charge through a partial double bond. The catalytic residues in both of the mechanisms are usually carboxylates but there is variation in the distance between these amino acids. In the single displacement mechanism the residues are less constrained and can be up to 10 Å apart, whereas in the double displacement mechanism the catalytic residues are much closer together, at about 5.5 Å, and there is a little variation in the distance. This is due to the difference in the placement of the base or nucleophile in the active site rather than the catalytic acid. In the single displacement mechanism the distance between the catalytic base and sugar must be sufficient to allow the presence of a water molecule; in contrast the nucleophile in the double displacement mechanism

needs to be much closer as it mounts a direct attack on the anomeric carbon (Rye and Withers, 2000, Zechel and Withers, 2000).

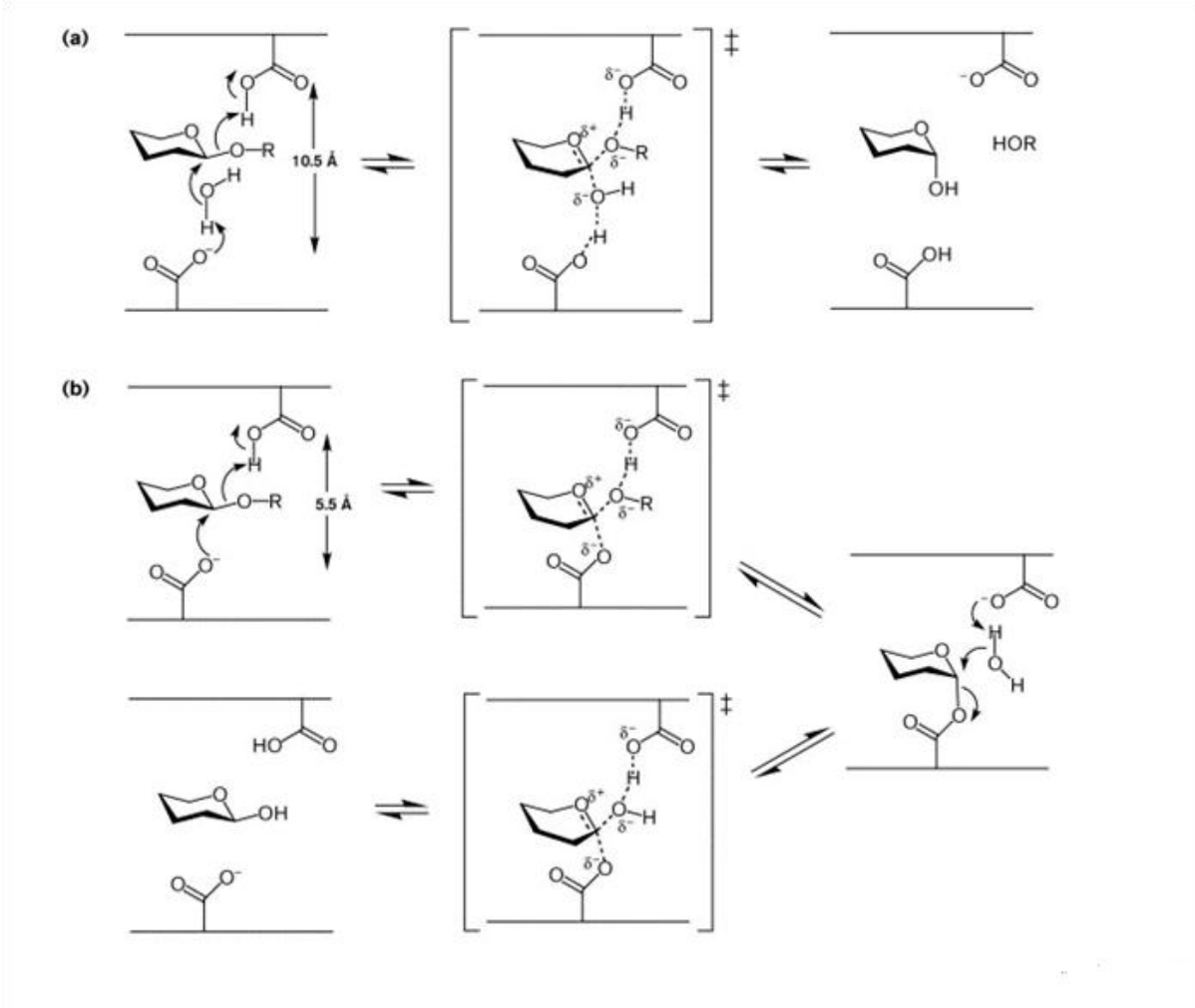


Figure 1.6 Catalytic mechanism for a) single displacement reaction, b) double displacement reaction

The transition state is shown in brackets. Figure taken from (Rye and Withers, 2000)

Glycoside hydrolase enzymes have been grouped into sequenced based families in the CAZy database (www.cazy.org)(Cantarel et al., 2009). Presently there are 130 glycoside hydrolase families (GHs). Each GH has a conserved fold, catalytic machinery and mechanism (Henrissat and Davies, 1997). The substrate specificity within some GHs, such as GH2, can be highly variable. However the stereochemistry and the regioselectivity of the target glycosidic bond are conserved.

In other GHs such as GH10 and GH11 substrate specificity is invariant. The mode of action of glycoside hydrolases, and thus to some extent their specificity is dictated by the topology of the active site. There are three main active site topologies displayed by glycoside hydrolases; (i) the open cleft, which accommodates internal regions of polysaccharides and thus these enzymes have an endo-mode; (ii) pocket topology, which accommodates typically a monosaccharide; (iii) tunnel topology, which is a feature of exo-acting processive enzymes where the polysaccharide chain is moved through the substrate binding tunnel, Figure 1.7. The endo-mode of action generates a range of oligosaccharides, exo-acting enzymes generally release monosaccharides, while exo-acting processive enzymes often produce disaccharides (Davies and Henrissat, 1995).

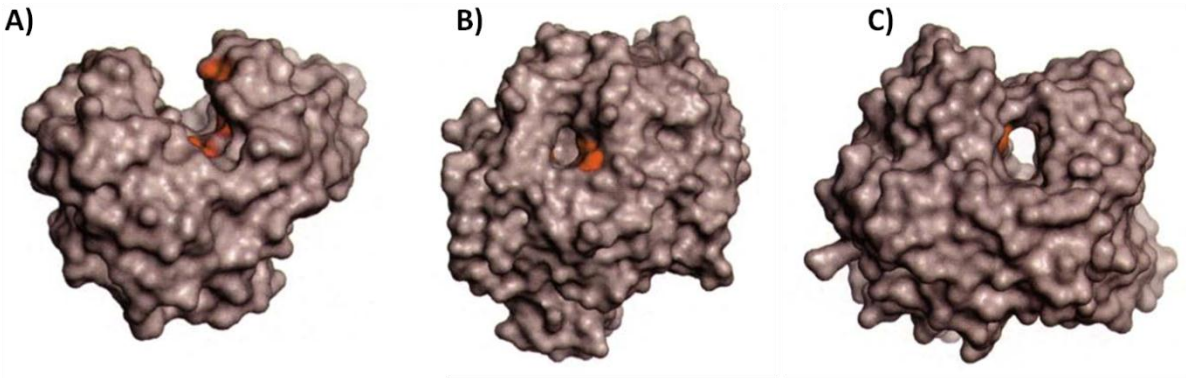


Figure 1.7 Glycoside hydrolase active site topology

A) Open cleft; B) Pocket; C) Tunnel. Active sites are shown in red. Figure adapted from (Davies and Henrissat, 1995)

The CAZy database has also grouped the GHs into clans in which the structural fold, catalytic apparatus and mechanism are conserved, but members of the different families display very low sequence similarity (Henrissat and Davies, 1997). There are 14 GH clans on the CAZy website from GH-A to GH-N. The biggest clan is GH-A which contains 19 GHs all of which display the $(\beta/\alpha)_8$ barrel fold, and catalysis is

through a retaining mechanism. It is thought that the families within each clan have a common evolutionary ancestry (Cantarel et al., 2009).

1.3 Carbohydrate Binding Modules

Carbohydrate binding modules (CBM) are non-catalytic discreet modules usually attached to glycoside hydrolases. CBMs are differentiated from other sugar binding modules, such as lectins and sugar transport proteins, as they are components of enzymes. CBMs for the most part are components of enzymes that attack insoluble recalcitrant polysaccharides (www.cazy.org). The first discovered CBMs were termed cellulose binding domains as they bound to crystalline cellulose (Tomme et al., 1988). Like glycoside hydrolases the CAZy database has grouped CBMs into, currently, 64 sequence-based families. Like GHs some CBM families display conserved ligand specificity, whereas in other families glycan recognition can be highly variable (Cantarel et al., 2009). CBMs have also been characterised into functional types. Type As are surface binding CBMs that usually target crystalline cellulose and chitin, Figure 1.8a. They have a flat hydrophobic binding surface that is thought to complement the planar surface of crystalline polysaccharides. Type A CBMs have little or no affinity for soluble ligands. CBMs grouped in the Type B category bind polysaccharide chains. The binding sites have a groove or cleft architecture capable of accommodating a single glycan chain. Biochemical analysis of Type B CBMs shows that they prefer longer chains such as hexasaccharides as opposed to small oligosaccharides, Figure 1.8b. Type C CBMs bind small sugars such as mono-, di-, and tri-saccharides. Type C CBMs have a very similar architecture to Type B but the binding clefts of Type C proteins are shorter and have more of a pocket topology, Figure 1.8c. Type C CBMs appear to have more of a lectin-like binding property than Type A and Type B modules (Boraston et al., 2004).

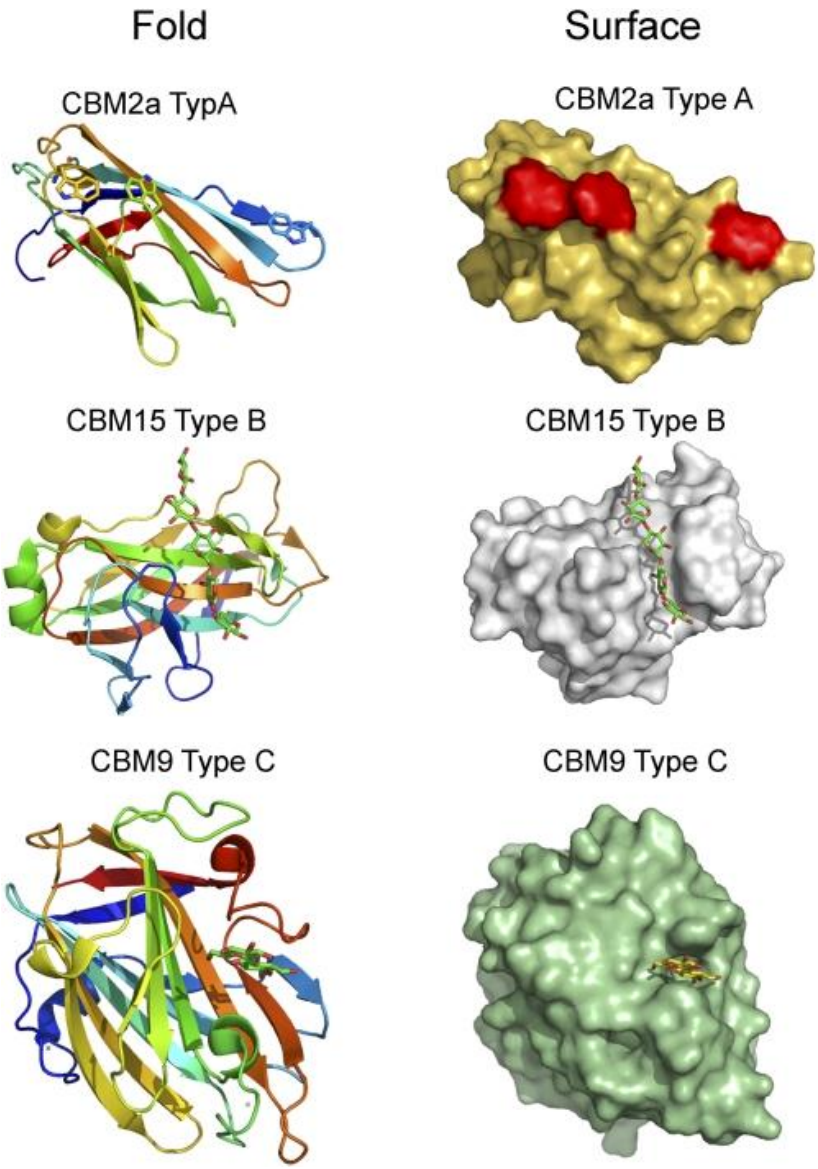


Figure 1.8 Carbohydrate binding module structures.

Type A, CBM2a from *Cellulomonas fimi* xylanase10A, the red areas show surface ligand binding sites. (PDB code 1XG). Type B, CBM15 from *Cellvibrio japonicas* xylanase10C, in complex with oligosaccharide ligand, showing the tunnel topology. (PDB code 1GNY). Type C, CBM9 from *Thermotoga maritima* GH10 xylanase, in complex with monosaccharide, showing the pocket topology. (PDB code 1I82). Figure taken from (Gilbert, 2010)

The primary role of CBMs is to enhance the activity of the appended enzyme against recalcitrant substrates. There are three routes proposed by which CBMs achieve this. The first and by far the most reported is the proximity effect. The binding of CBMs to their target ligands brings the enzyme into close and prolonged proximity to

the polysaccharide substrate. This effectively leads to an increase in enzyme concentration on the polysaccharide and thus more rapid degradation of the substrate. Removal of CBM from glycoside hydrolases results in a decrease in enzymatic activity against insoluble but not soluble polysaccharides (Bolam et al., 1998).

The second route by which CBMs can potentiate the activity of the attached enzyme is by a targeting effect. Instead of merely concentrating the enzyme onto the surface of the substrate, the CBM targets the enzyme to specific regions of the target polysaccharide that are particularly accessible to enzymatic attack. An eloquent example of a targeting CBM is the CBM9 that is a component of xylanase10A found in *Thermotoga maritima.* The CBM9 only binds the reducing ends of polysaccharides, suggesting that the CBM targets the enzyme to areas of the plant cell wall that are actively being degraded (Winterhalter et al., 1995).

The final route by which catalytic enhancement can be achieved is through a disruptive effect. There is very little data reported on the disruptive effect of CBMs on polysaccharides. When the first cellulose binding domain (CBD) was isolated it was thought that it potentiated catalytic activity by disrupting the recalcitrant structure of cellulose making it more accessible to enzyme attack. The CBM2a from *Cellulomonas fimi* demonstrated a modest capacity to disrupt the structure of crystalline cellulose (Din et al., 1994). It is thought that it may disrupt the hydrogen bonds holding the cellulose microfibers together and so make cellulose chains accessible for degradation by the attached enzyme. The disruptive effect is viewed with some caution as it is extremely difficult to observe the disrupted structure and this mode of action has only been reported for a single cellulose binding domain and

a starch binding module, however, in both instances, the enhancement in enzyme activity was only 50% (Giardina et al., 2001).

1.4 Glycoside Hydrolases acting on Fructan Polysaccharides

Glycoside hydrolases that target fructan polysaccharides are grouped into two families GH32 and GH68. Both families are the only members of the GH-J clan and so have a 5-fold β propeller structure. Clan GH-J enzymes display a retaining mechanism, and the conserved catalytic machinery comprises a catalytic acid/base aspartate residue and a catalytic nucleophile glutamate. There is also another conserved aspartate which is thought to play in a role in transition state stabilisation (www.cazy.org).

1.4.1 Family GH68 enzymes

Activities within GH68 include levansucrase, β-fructfuranosidase and inulosucrase. Levansucrases and inulosucrases act as fructosyltransferases to synthesise fructan polymers rather than mediate hydrolytic reactions. Levansucrases, at high concentrations of sucrose, catalyse transglycosylation reactions in which fructose residues are transferred to the sucrose acceptor resulting in the synthesis of levan oligosaccharides and the release of glucose (Ozimek et al., 2006). Transglycosylation occurs through the two step retaining mechanism. After the glycosylation reaction, in the deglycosylation step, instead of water acting as the acceptor, a second sucrose molecule (or fructan in subsequent cycles of catalysis) acts as the acceptor (Edelman, 2006). The first GH68 structure solved (Protein Data Bank) was a levansucrase from *Bacillus subtillis,* shown in Figure 1.9.. The structure revealed a five-bladed β- propeller fold with the active site situated in the middle of the propeller (Meng and Futterer, 2003). Structural studies on bacterial sucrases

identified, as well as the conserved catalytic residues, a number of non-catalytic conserved amino acids. In the enzymes from gram-positive bacteria there is a conserved arginine, in gram-negative bacteria enzymes a histidine is conserved in the equivalent position. These conserved residues are crucial for transfructosylation and are involved in the specificity and efficiency of the reaction. Bacterial inulosucrases are mainly found in lactic acid producing bacteria while levansucrases are found in various different bacteria. Bacterial levansucrases produce levans that are soluble and have a high DP (Lammens et al., 2009).

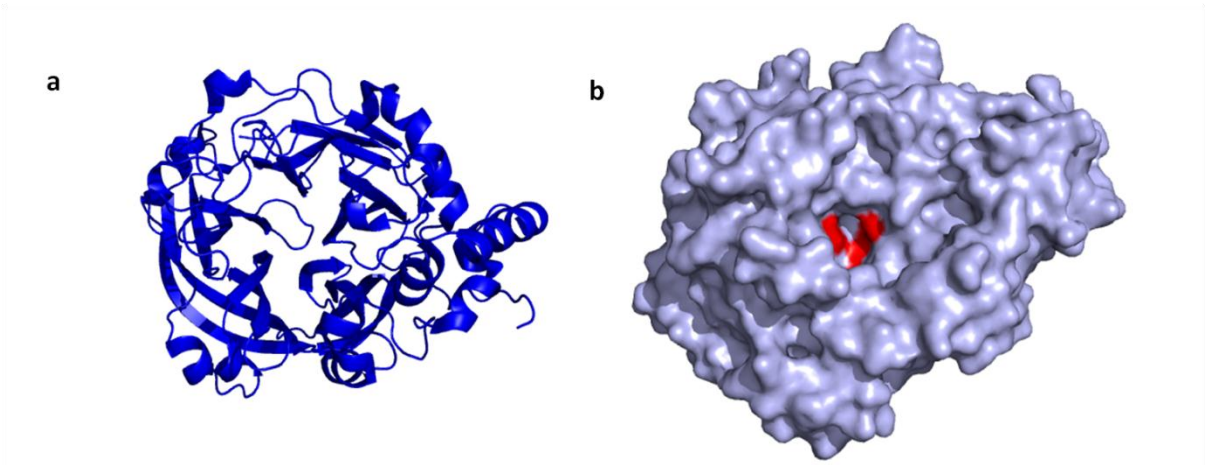


Figure 1.9 GH68 Levansucrase from *B. subtilis*

a, shows the cartoon structure of the β-propeller fold with the active site found in the middle. b, shows the surface structure with the residues involved in catalysis coloured red. PDB code 1OYG, created using PyMol.

1.4.2 GH32 enzymes

GH32 has a broader range of characterised activities compared to GH68. The main characterised activities are fructan hydrolysing, for example exo- and endo- inulases, levanases and invertases. Inulases cleave β-2,1 linkages while levanases hydrolyse the β-2,6 linkage, while invertases cleave sucrose to glucose and fructose. There are some enzymes with transglycosylating activites such as levan fructosyltransferase. The first solved structure of a GH32 family was an extracellular β-fructosidase from *Thermotoga maritime,* Figure 1.10 (Alberto et al., 2004).

Interestingly, GH32 enzymes have an extra C-terminal domain of unknown function. The C-terminal domain has a sandwich-like fold which consist of two six stranded anti-parallel β-sheets. The C-terminal domain has been shown to be essential for protein stability, however it is not known if this is the main purpose of the domain (Altenbach et al., 2004). As with the GH68 enzymes, the active site of the GH32 enzymes is in the middle of the β-propeller. Along with the three conserved residues in the active site there are various residues that play an important role in catalysis. Examples of this are the conserved tryptophan and phenylalanine residues at the rim of the active site. The hydrophobic zone created by these residues is thought to be important in substrate binding and stability. Site directed mutagenesis studies have shown a marked increase in K_m for sucrose when these aromatic residues were mutated (Le Roy et al., 2007).

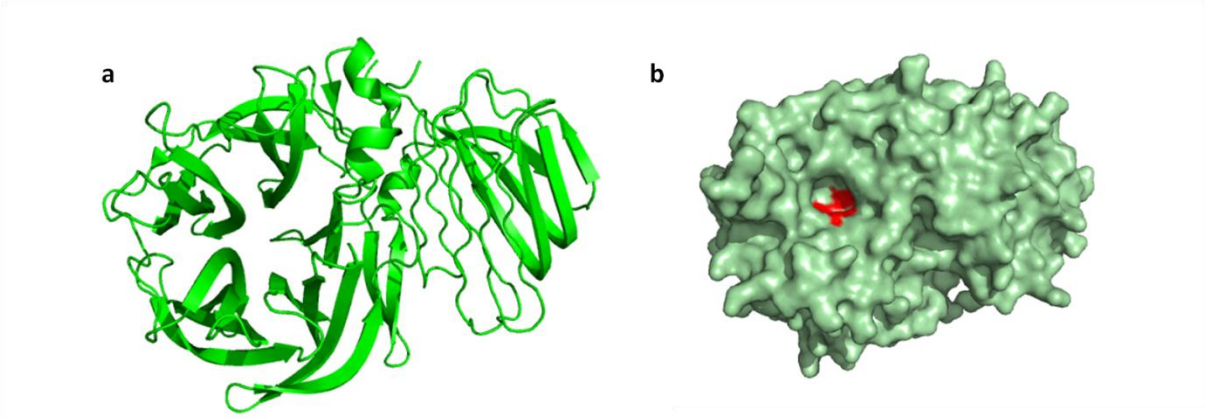


Figure 1.10 β-fructosidase from *Thermotoga martima*

a, shows cartoon structure of the β-propeller and C-terminal β-sandwich domain. b, shows surface representation with residues involved in catalysis coloured red. PDB code 1UYP, created using PyMol.

While the majority of GH32 enzymes are bacterial, there is still a large eukaryotic cohort of GH32 members, along with a small proportion found in archea. In both plants and bacteria GH32 enzymes most probably play a role in processing fructans that fulfil a storage or biofilm function.

1.5 Glycoside Hydrolase enzymes acting on Mannan polysaccharides

In plant cell walls the mannan backbone is linked through β1,4-Man bonds, while in yeast mannans the backbone contains α1,6-Man linkages. As this study is concerned with mannan derived from *S. cerevisiae*, only GHs in which the enzymes act on α-linked mannose resides are discussed here.

1.5.1 α-Mannosidases

α-Mannosidases act in an exo fashion removing single mannose residues from the non-reducing end of a α-mannan or oligosaccharides derived from the polysaccharide. Currently there are five GHs that contain characterised α-mannosidases, GH38, GH47, GH92, GH99 and GH125. Of these five families the most extensively characterised are GH38, GH47 and GH92. By contrast only a single enzyme has been extensively characterised in GH99 and GH125. (www.Cazy.org)

1.5.1.1 Glycoside hydrolase family 38

The GH38 family contains zinc dependent α-mannosidases which act through the acid-base double displacement mechanism, leading to the retention of the conformation of the anomeric carbon (Numao et al., 2003). GH38 members display a $(\alpha/\beta)_7$ fold but are not currently assigned to a glycoside hydrolase clan, Figure 1.11. The catalytic residues of these enzymes are both aspartates. The eukaryotic GH38s are defined as class II α-mannosidases, which are either golgi α-mannosidase II or lysosomal α mannosidase (Liao et al., 1996, Moremen et al., 1991) .

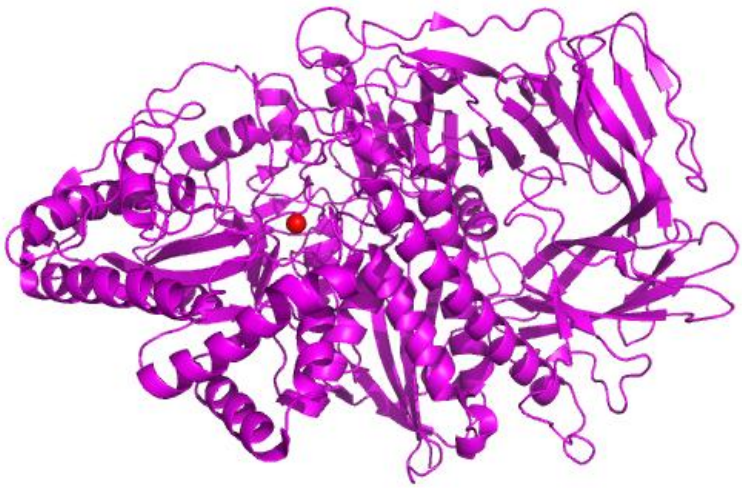


Figure 1.11 Structure of a GH38 family enzyme

Golgi α-mannosidase II from *Drosophila melanogaster,* zinc atom shown in red. PDB code 1HTY, created using PyMol.

Golgi α-mannosdiase II are capable of hydrolysing both α-1,3 and α-1,6 linked mannose residues (Bonay and Hughes, 1991). The active site of Golgi α-mannosidase II displays a requirement for a GlcNAc residue five sugar downstream from the cleavage site, for hydrolysis to occur. The Golgi α-mannosidase II enzymes are essential for N-glycan maturation in the Golgi apparatus, removing the α 1,3 and α 1,6 linked mannose from $Man_5GlcNAc_2$ leaving the $Man_3GlcNAc_2$ structure (van den Elsen et al., 2001). Studies of mutant mice deficient in Golgi α-mannosidase II display a phenotype similar to human congenital dyserythropoietic anemia type II, suggesting that the enzyme is responsible for the loss of N-glycan synthesis in erythrocytes (Chui et al., 1997).

Lysosomal α-mannosidases are either non-specific, capable of hydrolysing α-1,2, α-1,3 and α-1,6 glycosidic bonds, or are highly specific for the α-1,6 linkage (Park et al., 2005). Like the Golgi α-mannosidase II, the active site is a large open cleft, however the enzyme does not display tight specificity for structures downstream of the side of cleavage. Figure 1.12 shows an example of the GH38 active site.

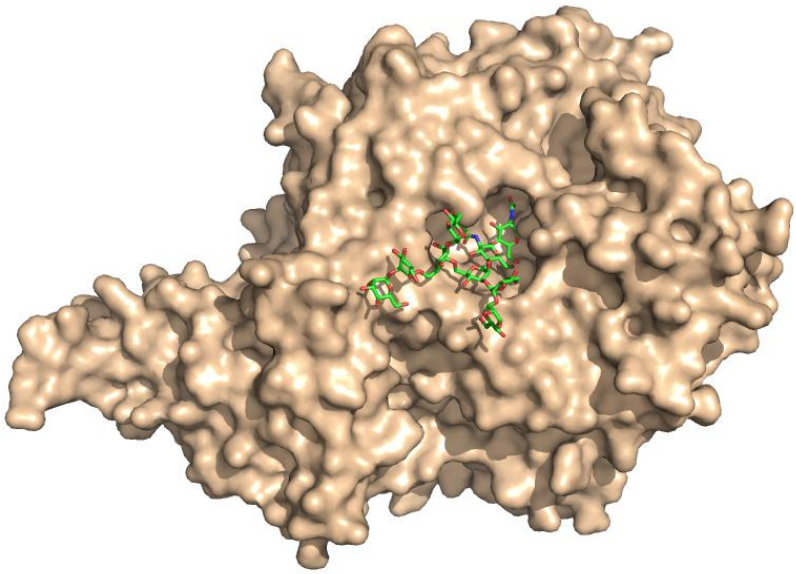


Figure 1.12 Surface representation of a lysosomal α-mannosidase

From *Bos taurus* in complex with ligand. Shows large open cleft of active site. PDB code 1O7D, created using PyMol.

In eukaryotes lysosomal α-mannosidases are important in catabolism of N-glycans during glycoprotein turnover (Cenci di Bello et al., 1983). In bacteria they have a physiological role in glycan foraging and utilization. Deficiencies in humans and livestock lysosomal mannosidases give rise to a condition called α-mannosidosis. In humans it is caused by an autosomal recessive mutation, where as in livestock it is caused by chronic poisoning by swainsonine. The complete loss or reduction in functional α-mannosidase results in the build up of sugars and glycans in the

lysosome, which impairs cell function. At its most severe form, the disease leads to death in infancy due to degeneration of the central nervous system. Milder forms of the disease results in neurological degeneration along with a myriad of other symptoms (Gotoda et al., 1998).

1.5.1.2 Glycoside Hydrolase family 47

GH47 contains a group of class I α-mannosidases, which are calcium dependant inverting enzymes that display an $(\alpha/\alpha)_7$ barrel fold, Figure 1.13. GH47 eukaryotic enzymes can be further divided into three sub-groups depending on their activity against $Man_9GlcNAc_2$ (Daniel et al., 1994, Herscovics, 2001).

Figure 1.13 Structure of a GH47 family enzyme

From *Homo sapiens* in complex with calcium (blue sphere) PDB code 1FMI, created using PyMol.

The first sub-group include human and yeast endoplasmic reticulum (ER) α-1,2 mannosidases that hydrolyse the $Man_9GlcNAc_2$ structure to the Man_8B isomer. This

is an important step in N-glycan maturation (Camirand et al., 1991, Gonzalez et al., 1999). This sub-group is also involved in protein quality control in the ER where misfolded proteins are translocated to the cytosol for degradation (Herscovics, 2001). The role of ER associated α-mannosidases has been extensively characterised in both *S. cerevisiae* and mammalian cells. Studies showed that this group of enzymes have specificity towards the terminal mannose of the middle chain of the high mannose N-glycan on misfolded proteins, Figure 1.14. This results in the transportation of the proteins to the cytosol for degradation (Jakob et al., 1998). It was found in *S. cerevisiae* that 1-deoxymannojirimycin inhibited the activity of the class I α-mannosidase resulting in the inhibition of glycoprotein degradation. This suggests that the role of the enzyme in N-glycan trimming is crucial for glycoprotein quality control and subsequent degradation (Liu et al., 1999).

The second sub-group consists of mammalian Golgi α-mannosidases IA,IB and IC along with enzymes from insect cells and filamentous fungi (Lal et al., 1994, Tremblay et al., 1998, Tremblay and Herscovics, 2000, Igdoura et al., 1999). The Golgi α-mannosidase I enzymes trim the α 1,2 linked mannose residues from $Man_{8-9}GlcNAc_2$ to $Man_5GlcNAc_2$, Figure 1.14. The difference in the three types of mammalian Golgi α-mannosidases is mainly their tissue and cell-type specific expression (Igdoura et al., 1999). Golgi-α mannosidases I preferentially remove mannose residues from the N-glycan but only display a weak affinity for the mannose residue cleaved by the ER associated enzymes. Therefore this sub-group of enzymes display a complementary specificity to the ER associated α-mannosidases.

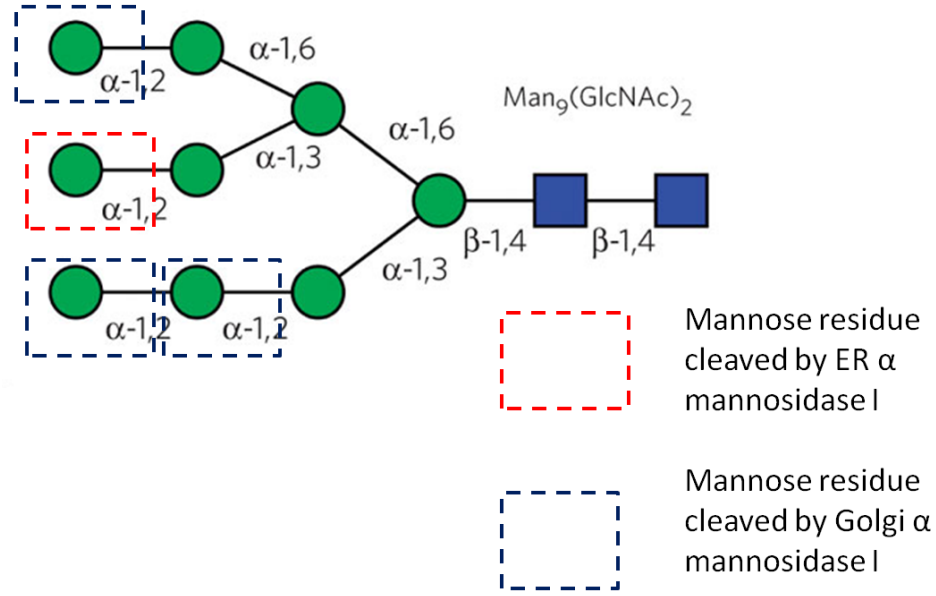


Figure 1.14 Man9 glycan trimming by α-mannosidase

Mannose residues are shown as green circles and GlcNAc residues are shown as blue squares. Mannose residues cleaved by α-mannosidases are outlined by a dashed box. Figure adapted from Zhu et al. (2010)

The third group appear to have no mannosdase activity against $Man_9GlcNAc_2$ glycan but have a role in glycoprotein degradation. These proteins have been termed ER Degradation Enhancing α-Mannosidase like proteins (EDEM) (Hosokawa et al., 2006). It has been reported that ER stress induces the overexpression of EDEM which speeds up the degradation of specific proteins (Hosokawa et al., 2001). The role and mechanism of EDEM is still unclear, the proteins while appearing to be inactive *in vitro* still posses conserved catalytic residues. Hypotheses of EDEM function include them acting as chaperones, preventing the formation of disulfide-bonded dimmers (Hosokawa et al., 2006) or aggregation of misfolded glycoproteins (Olivari et al., 2006). This is significant as inhibition of aggregation is implicated in the degradation of misfolded glycoproteins.

1.5.1.3 Glycoside Hydrolase 92

Specificity within GH92 is variable with α-1,2, α-1,3, α-1,4 and α-1,6 mannosidase activities found among the 23 *Bacteroides thetaiotaomicron* GH92 members. Like GH38 and GH47, GH92 enzymes have a requirement for a metal ion, in this case calcium, for catalysis to occur. The GH92 enzymes work through the single displacement reaction and so invert the anomeric conformation at C1. GH92 enzymes display a two domain structure; the N-terminal domain is a beta sandwich structure while the C-terminal domain is an $(\alpha/\alpha)_6$ barrel fold, shown in Figure 1.15. The catalytic acid is predicted to be a glutamate while the catalytic base is harder to recognise but is thought to be an aspartate. Mutational studies have shown that mutants of the predicted catalytic acid and catalytic base were inactive, confirming their roles in enzyme action (Zhu et al., 2010) .

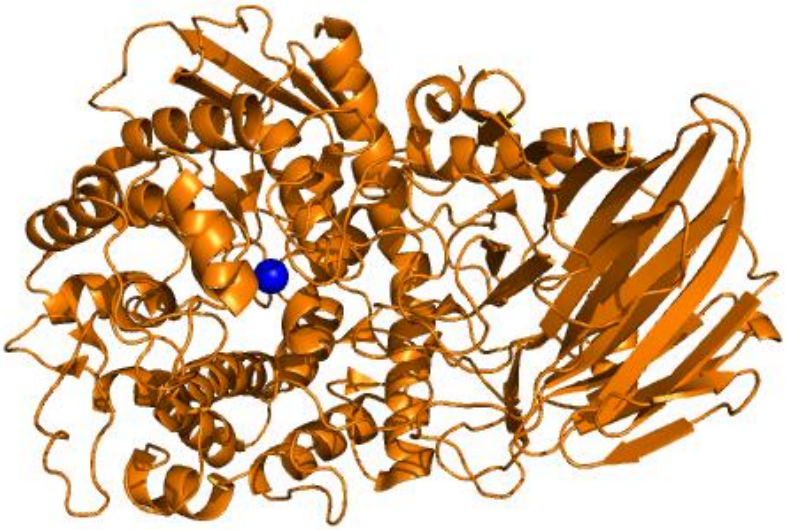


Figure 1.15 Structure of GH92 family member

From *B.thetaiotaomicron* in complex with calcium (shown in blue) PDB code 2WVX, created using PyMol.

The study of GH92 enzymes from *B. thetaiotaomicron* revealed activity against 4-nitrophenyl-α-D-mannopyranoside and α-linked mannose disaccharides. In addition, a number of the enzymes were capable of releasing mannose from $Man_9GlcNAc_2$. In the human large bowel, the environment *B.thetaiotaomicron* inhabits, the likely substrates for the 23 GH92 enzymes are the N-glycans found on host and dietary glycoproteins (Zhu et al., 2010). Furthermore, transcriptome data for *B. thetaiotaomicron* showed that when the bacterium was cultured in the presence of the N-glycan mimic yeast mannan, 50% of the GH92 enzymes were upregulated (Bjursell et al., 2006, Martens et al., 2008). Indeed the upregulated GH92 are often co-transcribed with other enzymes necessary for full degradation of N-glycans. It should be emphasised, however, that it is also possible that a proportion of the GH92 enzymes may contribute to the degradation of extended N-glycan structures presented on the surface of yeast mannan, particularly those that are co-regulated with GH76 enzymes described below.

1.5.2 α-Mannanases (Glycoside hydrolase family 76)

GH76 is currently the only family in the CAZy database annotated as having an α-mannanase activity, a *Bacillus circulans* enzyme characterised as an α1,6 mannanase. The *B. circulans* enzyme hydrolyses the α1,6 linked backbone of yeast mannan in an endo-mode of action, releasing α1,6 mannosyl oligosaccharides (Nakajima et al., 1976). The other members of the family have been annotated as α 1,6 mannanases, but have not been characterised. There is no reported information on the catalytic mechanism or the amino acids that contribute to catalysis or

substrate specificity. One structure of a GH76 enzyme from *Listeria innocua* has been solved and is shown in Figure 1.16.

The genome of *B. thetaiotaomicron* encodes several GH76 enzymes, some of which are co-regulated with GH92 enzymes and other proteins that are predicted to contribute to α-mannan degradation. This suggests that the GH76 enzymes, and the other co-regulated enzymes, may have a role in degrading yeast α-mannan (Martens et al., 2008, Bjursell et al., 2006).

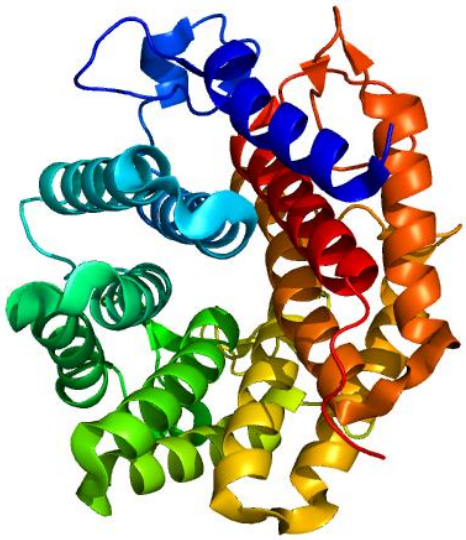


Figure 1.16 Structure of a GH76 enzyme from *Listeria innocua.*

PDB code 3K7X. Solved and deposited by Northeast Structural Genomics Consortium, created using PyMol.

1.6 Xylanases

Xylanases hydrolyse the β 1,4 xylan backbone in an endo fashion to produce a range of oligosaccharides (Harris et al., 1994). All GH10 and GH11 enzymes characterized to date are endo-acting xylanases. All but two of the characterized 235 GH10 members have β1,4 xylanase activity, the other two hydrolyse β1,3 linked xylan. GH10 contains retaining enzymes that are members of clan GH-A, displaying a typical $(\beta/\alpha)_8$ barrel fold, Figure 1.17a (Harris et al., 1994). Experimental evidence has identified both of the catalytic residues as glutamates (Tull et al., 1991, MacLeod et al., 1994). GH11 enzymes also act through a retaining mechanism; they display a β-jelly roll fold and are components of clan GH-C, Figure 1.18. As with GH10 both the catalytic residues have been reported as glutamates (Miao et al., 1994, Lawson et al., 1996). All 237 characterized GH11 enzymes were shown to be endo-β1,4-xylanases. As both the GH10 and GH11 enzymes are endo-acting xylanases and proceed through similar mechanisms, not surprisingly, their respective active sites show a similar architecture. The open cleft nature of the active site allows for multiple sugar (xylose) binding subsites, which is important for substrate recognition and catalysis (Charnock et al., 1998). While most GH10 and GH11 xylanases hydrolyse un-substituted xylan chains, there have been reports of enzymes that can tolerate, or display absolute specificity for, decorated xylan side chains.

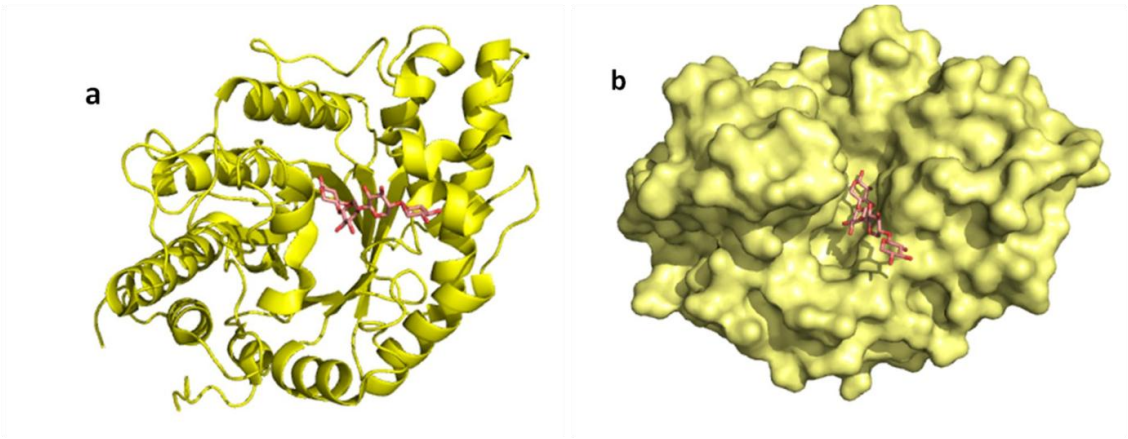


Figure 1.17 Structure of Xylanase10C from *Cellvibrio japonicas*

a, cartoon structure in complex with xylo-oligosaccharide b, surface representation shows open cleft of active site. PDB code 1US, created using PyMol.

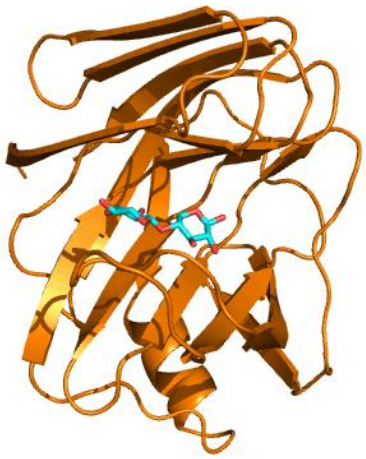


Figure 1.18 Cartoon structure of GH 11 xylanase from *Bacillus cirulans*

Substrate show in blue. PDB code 1BCX, created using PyMol.

Other GHs that include xylanase activities are GH5, GH8 and GH30. Within GH30 there are examples of highly specific xylanase enzymes. One of these enzymes is the glucuronoxylan specific xylanase found in the bacterium *Erwinia chrysanthemi*. This enzyme specifically hydrolyses xylan decorated with a 4-O-methyl-α-D-glucuronic acid (MeGlcA) at the O2 position. Structural studies showed that MeGlcA decoration of the xylose positioned in the -2 subsite (glycosidic bond cleavage occurs between the xylose at the -1 and +1 subsites) (Davies et al., 1997) is vital for

substrate distortion and thus hydrolysis of the xylan backbone (Urbanikova et al., 2011). Recently a GH5 enzyme from *Clostridium thermocellum* was reported to be arabinoxylan specific, this activity is described in more detail in Chapter 6 (Correia et al., 2011).

1.7 *Bacteroides thetaiotaomicron*

B.thetaiotaomicron is a Gram-negative bacterium and a prominent resident of the microflora of the human and murine gastrointestinal (GI) tract (Moore and Holdeman, 1974). Like almost all of the other large bowel bacteria *B. thetaiotaomicron* is an obligate anaerobe. *B. thetaiotaomicron* develops prominence in the gut during postnatal development during weaning from mother's milk to a diet rich in polysaccharides (Mackie et al., 1999). It is pivotal in the regulation of α-fucosylation of the distal intestine which directs microbial colonization of specific niches within the gut (Bry et al., 1996). This gut bacterium also affects the submucosal network of interconnected capillaries. Experiments in adult germ free mice showed that the colonization of *B.thetaiotaomicron* resulted in the full development of the capillary network within 10 days (Stappenbeck et al., 2002). *B.thetaiotaomicron* has also been shown to induce a Paneth cell protein, Ang4, that is secreted into the gut lumen and is bactericidal to Gram-positive bacteria. Ang4 expression protects the host from pathogens, while also benefiting *B.thetaiotaomicron* through its capacity to exert a degree of control of on its microbial environment (Hooper et al., 2003). *B. thetaiotaomicron* also has the ability to degrade and utilize nutrients, typically complex polysaccharides, that are not attacked by human gut enzymes. Degradation of indigestible plant polysaccharides by the human microbiome and their metabolism to short chain fatty acids such as propionate, butyrate and acetate, can account for up to 10% of human daily calorific intake (Xu and Gordon, 2003).

1.7.1 The glycophilic nature of *B. thetaiotaomicron*

B.thetaiotaomicron is a glycophile that dedicates a large proportion of its genome to polysaccharide utilisation. Currently, the genome of *B. thetaiotaomicron* is predicted

to encode 257 glycoside hydrolases, 16 polysaccharide lyases and 20 carbohydrate esterases (www.cazy.org). This impressive repertoire of enzymes suggests *B.thetaiotaomicron* is capable of hydrolysing a diverse range of glycosidic linkages. Over half of these enzymes are predicted to be extracellular or periplasmic. Whole genome transcription profiling of cecal bacteria from germ free mice inoculated with *B. thetaiotaomicron* fed on a simple sugar diet, or a rich-polysaccharide diet, detailed *B. thetaiotaomicron's* capacity to forage for glycans. Those mice fed on a simple sugar diet showed that *B.thetaiotaomicron* could utilize host glycans as an energy source. Consistent with its repertoire of glycoside hydrolases, culturing strategies showed that *B.thetaiotaomicron* is able to degrade and grow on an extensive range of plant- and mammalian-derived polysaccharides (Sonnenburg et al., 2010).

1.7.2 Polysaccharide utilisation by *B. thetaiotaomicron*

Analysis of the *B. thetaiotaomicrons* genome revealed multiple gene clusters that encode all of the necessary machinery to sense and utilise host and dietary glycans. These genetic regions are termed Polysaccharide Utilisation Loci (PULs) (Bjursell et al., 2006). A typical feature of PULs is the presence of genes encoding a SusC/SusD (Starch Utilisation System) pair which, are co-localised with the genes encoding glycoside hydrolases. Extracytoplasmic function (ECF) σ factors or hybrid two component systems (HTCS) are also encoded within the PULs and are involved in signal sensing by *B.thetaiotaomicron.*

The starch utilisation system (Sus) was the first PUL studied and characterised in *B.thetaiotaomicron,* a summary of this archetypal locus is shown in Figure 1.19. The Sus locus comprises seven genes SusABCDEFG and a transcription factor, SusR, which activates transcription of the seven Sus genes through a sensor/regulator role

(Hooper et al., 1999, Reeves et al., 1996, Reeves et al., 1997, D'Elia and Salyers, 1996b). SusR spans the cytoplasmic membrane with the N-terminal domain extended into the periplasm and the C-terminal domain into the cytoplasm. The C-terminal cytoplasmic domain contains a helix-turn-helix DNA binding motif. SusR only induces expression of the remaining Sus genes in response to the presence of maltose (D'Elia and Salyers, 1996b). SusABG are hydrolytic enzymes consisting of a neopullalanse, α-glucosidase and α-amylase, respectively. While SusG is located at the outer membrane and exposed to the external environment, SusAB are in the periplasm. SusDEFG are all lipoproteins localised to the outer membrane facing the external environment. SusD binds the incoming starch while SusG, an α-amylase, cleaves the long α-glucan chains which are subsequently transported through the membrane by the SusC protein (D'Elia and Salyers, 1996a, Shipman et al., 1999, Reeves et al., 1997, Shipman et al., 2000) .

SusC is a TonB dependent β barrel transport protein that spans the outer membrane. TonB proteins are a family of porins found in Gram-negative bacteria that span the outer membrane. Small molecules can diffuse through the protein while larger molecules are transported via active transport, the energy for which is driven by a protonmotive force through the TonB-ExbBD complex. SusC transports oligosaccharides into the periplasm for further degradation (Koropatkin et al., 2008, Reeves et al., 1996).

SusD in the starch utilisation system is responsible for binding starch at the cell surface, which is then hydrolysed by the α-amylase, SusG. SusD is a outermembrane lipoprotein tethered to the membrane by a N-terminal cysteine (Martens et al., 2008). A study of the structure of SusD reported that the protein is

optimized to bind to curved oligosaccharides. This curvature resembled the cyclic nature of starch. In addition the same study showed SusD bound cyclic ligands better than linear maltooligosaccharides. From these data it is proposed that SusD recognises the helical nature of starch rather than the individual sugars or glycosidic linkages. SusD is an α-helical protein, unlike the typical β barrel structure of CBMs, Figure 1.20 (Koropatkin et al., 2008). SusD posses an arrangement of tetratricopeptide repeat units (TPR), which have been implicated in protein-protein interactions, suggesting that the SusD TPR units could form the basis for the SusC/SusD complex. The exact role of SusD is unclear but it is thought it may hold the maltooligosaccharides for threading through the SusC protein. Genetic knockouts of SusD revealed that the protein is required for the growth of *B. thetaiotaomicron* on maltooligosaccharides of 6 glucose units or more (Sonnenburg et al., 2006).

Both SusE and SusF are thought to also be involved in polysaccharide binding but there is still uncertainty surrounding the roles of these proteins. The PULs that orchestrate the degradation of other glycans direct the synthesis of a suit of enzymes, binding proteins, transporters and sensor that are specific for the target polysaccharide (Martens et al., 2011).

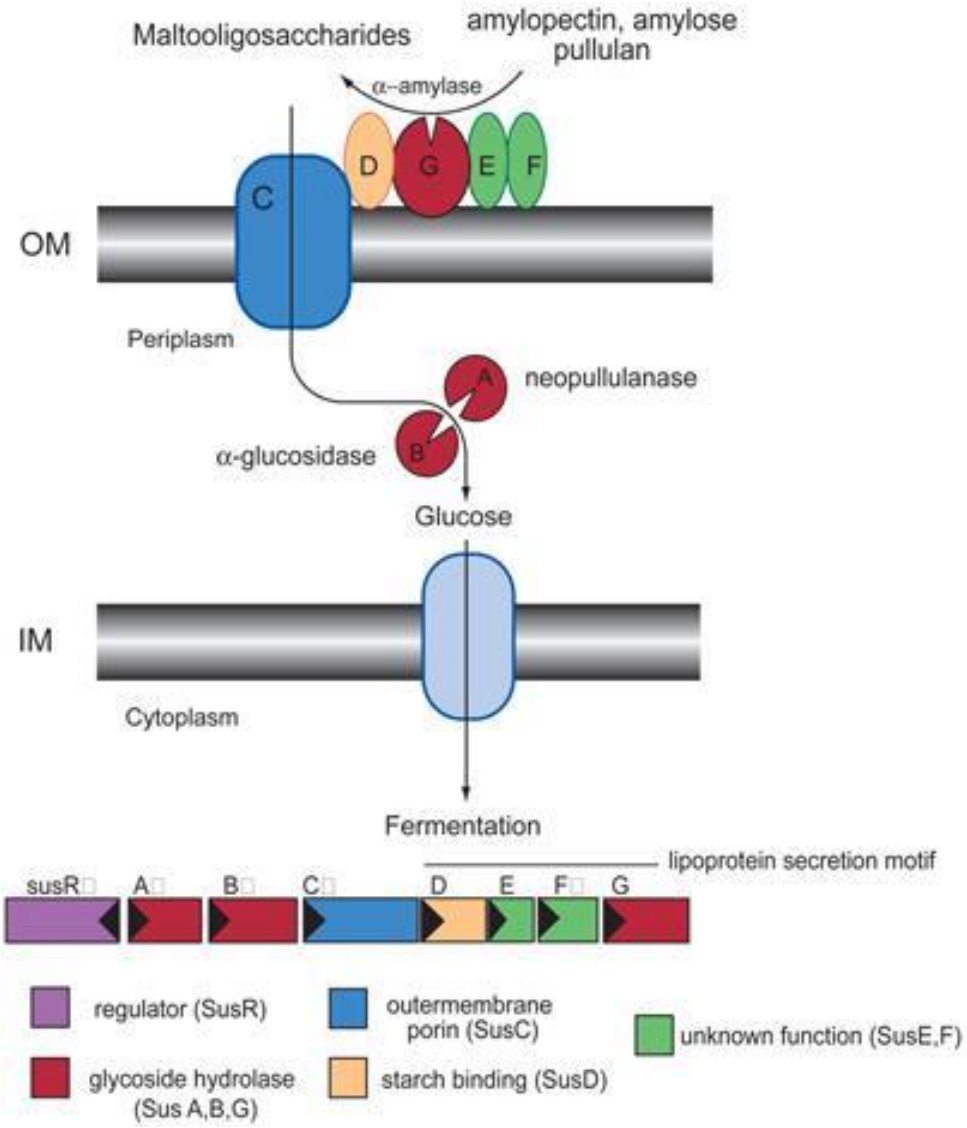
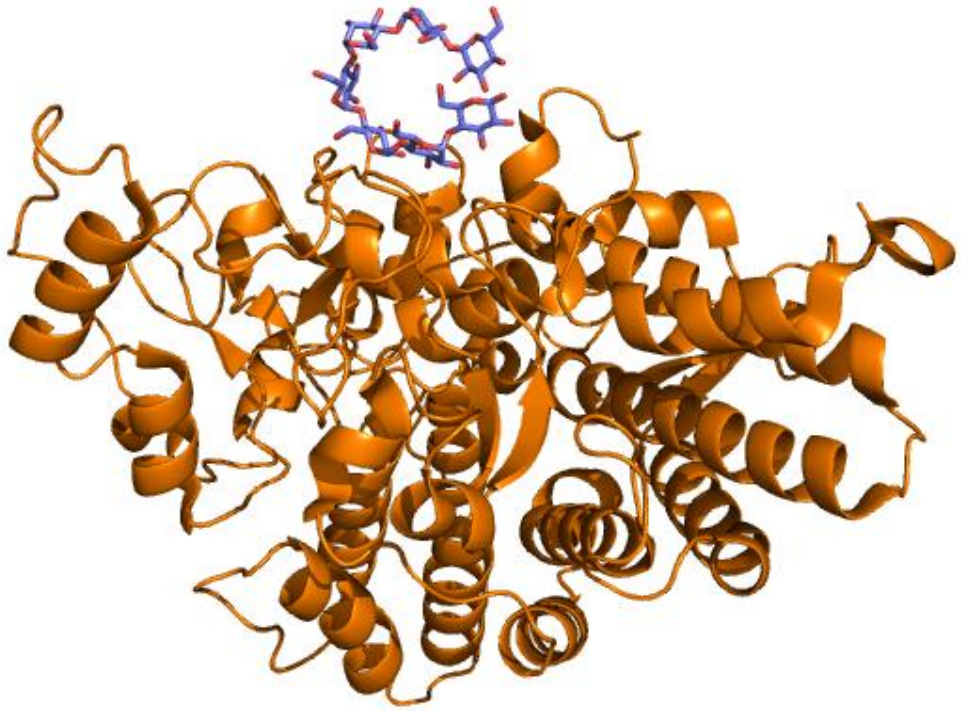


Figure 1.19 Cartoon representation of Starch Utilisation System (SUS) of *B. thetaiotaomicron*

Starch is bound at the surface by SusD and processed by SusG. The oligosaccharides are transported into the periplasm through the SusC. Further processing takes place in the periplasm by glycoside hydrolase enzymes. Finally glucose is transported through the inner membrane into the cytoplasm for utilisation. Figure taken from(Koropatkin et al., 2008)



.

Figure 1.20 SusD in complex with maltoheptose.

SusD from *B.thetaiotaomicron.* Maltoheptose is shown in blue, bound to the SusD in a helical nature. PDB code 3CK9.

1.7.2.1 Carbohydrate sensing and gene regulation

The PULs of *B. thetaiotaomicron* are tightly regulated and while there is a basal level of expression of each PUL, upregulation only occurs in response to particular environmental cues. For this to occur each PUL posses a regulation system, which is able to sense particular carbohydrates and activate gene expression in response to these ligands.

A type of regulator of PUL transcription are the Extracytoplasmic function (ECF) type σ factors that control gene expression by mediating DNA:RNA complexes. Upon detection of the activating ligand ECF type σ factor is released from its anti-σ

factor enabling the protein to interact with RNA polymerase to regulate expression of target genes (Helmann, 2002). ECF σ factors are generally grouped into two families, σ-70 and σ-54 (Wosten, 1998). The *B. thetaiotaomicron* genome encodes 54 σ factors in total; two are σ-70 proteins and two belong to the σ-54 family. The other 50 σ factors are ECF type σ factors which belong to the alternative sigma factor group of the σ-70 family. In total 12 of these ECF type σ factors are located upstream of genes encoding SusC and Sus D homologs and enzymes that catalyze complex glycan degradation. None of these 12 loci are linked to any other transcription factors (Xu and Gordon, 2003).

Hybrid two component systems (HTCS) are a molecular system for organisms to sense and respond to stimulus in the environment. HTCS incorporate all of the necessary proteins found in a classical two component system, but within a single polypeptide. This consists of the sensor kinase, histidine kinase, a phosphoacceptor domain and a DNA binding domain (Stock et al., 2000). The sensor kinase is a periplasmic domain situated at the N-terminus. It is interrupted by five predicted transmembrane domains with an additional three cytoplasmic segments, the histidine kinase, phosphoacceptor domain and the putative DNA binding domain. The putative DNA binding domain is of the AraC helix-turn-helix type (Lowe et al., 2012). It has been proposed the DNA binding domain may be released from the protein upon activation, allowing the protein domain to bind DNA and regulate expression (Miyazaki et al., 2003). A mechanism for termination of regulation has currently not been proposed. In *B. thetaiotaomicron* 32 HTCS have been identified, 23 of which are immediately upstream of polysaccharide utilisation genes. The sensor domains of HTCS are diverse binding different molecular cues that regulate gene expression, but their mechanism of signal transduction is conserved (Xu and

Gordon, 2003, Sonnenburg et al., 2006, Martens et al., 2011). A recent structural study of the periplasmic domain of BT4663 from a *B.thetaiotaomicron* HTCS reported a conformational change in structure upon ligand binding. The structure of the periplasmic domain of BT4663 displayed a β-propeller fold followed by a domain termed Y_Y_Y of unknown function. Upon binding of ligand the structure revealed a scissor blade like closing mechanism of the C-terminal Y_Y_Y domain, enabling the periplasmic domains of the protomers to mediate reciprocal phosphorylation. The study suggests a mechanism for signal transduction through the membrane that is different to the previously described piston and rotation models (Lowe et al., 2012).

1.8 Objectives

The three main aims of this study are:

1. To further elucidate the mechanisms by which CBMs contribute to glycoside hydrolase activity. By studying a novel levan binding CBM and its native GH32 enzyme, this component of the project will explore how CBMs can contribute to the enzymatic degradation of soluble polysaccharides.

2. To understand how *B. thetaiotaomicron* exploits co-regulated PULs to deconstruct α-mannans derived from *S. cerevisiae.* Characterising the GH76 and GH92 enzymes localised in α-mannan regulated PULs is designed to dissect the mechanism by which *B. thetaiotaomicron* hydrolyse and utilize α-mannans. In addition the study of hypothetical proteins encoded within the PULs is designed to reveal additional enzymes involved in α-mannan degradation.

3. To understand the active site specificity determinants that confers the specificity displayed by an arabinoxylan specific xylanase from *Clostridium thermocellum.*

Chapter 2: Materials and Methods

2.1 Molecular Biology

2.1.1 Bacterial Strains and Plasmids

All *Escherichia coli* strains and plasmids used in this study are listed in Table 2.1 and Table 2.2 respectively.

2.1.2 Media and Growth Conditions

E. *coli* were cultured at 37ºC (unless stated otherwise) with aeration in a rotary shaker (180 RPM), in sterile Luria-Bertani Broth (LB) (1 % (w/v) Bacto©Tryptone, 1 % (w/v) NaCl and 0.5 % (w/v) yeast extract, pH 7.4) (Sambrook et al., 1989) with appropriate antibiotic. Agar plates were made by adding 2 % (w/v) agar (Bacteriological) to LB, autoclaving and then allowing it to cool prior to the addition of selective antibiotic. *E*. *coli* colonies were grown by spreading bacterial cultures across the surface of the plate with a glass spreader that had been sterilized by immersion in 100 % (v/v) ethanol and passing through the hottest part (blue) of a Bunsen flame. Bacterial plates were incubated invertedly at 37°C for about 16 h in a stationary incubator (Laboratory Thermal Equipment Ltd). Longer incubation periods were avoided to prevent the formation of satellite colonies.

2.1.3 Selective Media

For antibiotic selection, antibiotics were used at 1000 fold dilution of stock solutions as shown in Table 2.3, added to the sterilizd growth media once it had cooled to less than 55°C. Isopropylthio-β-D-galactoside (IPTG) was added to strains containing *lacI*^{q} either on plasmids or in the genome for induction of transcription of recombinant

genes controlled by *lacO*. IPTG stock concentration was made at 1 M in MQ H_2O and, for most strains, added to liquid media to a final concentration of 1 mM. For Tuner *E. coli* strains IPTG was added to the media to a final concentration of 0.2 mM.

2.1.4 Storage of DNA and Bacteria

Bacterial colonies on agar plates were stored at 4 °C for a maximum of four weeks. DNA was stored at -20 °C in Elution Buffer (EB, 10 mM Tris/HCl buffer, pH 8.5).

E. coli Strain	*Description*	*Use*	*Reference*
BL21 (DE3)	F⁻ *dcm ompT hsdS*(r_B- m_B-) gal l(DE3)	Protein Expression	(Studier and Moffatt, 1986)
Tuner (DE3)	F⁻ ompT hsdSB (r_B^- m_B^-) gal dcm lacY1 (DE3)	Protein Expression	Novagen
One Shot™ TOP10	F´mcrA (mrrCB-hsdRMS-mrr) 80lacZ M15 lacX74 deoR recA1 araD139 (ara-eu)7697 galU galK rspL endA1 nupG	DNA Ligation and Replication	Invitrogen

Table 2.1 *E.coli* Strains

Plasmid	Size (kb)	Genotype	Use in this study	Reference
pET21a	5.44	amp^r, T7 lac $lacI^q$	Expression vector	Studier, 1990
pET22b	5.49	amp^r, T7 lac $lacI^q$	Expression vector	Novagen
pET28a	5.37	kan^r, T7 lac $lacI^q$	Expression vector	Novagen

Table 2.2 Bacterial Plasmids

Antibiotic	*Stock concentration*	*Working concentration*	*Storage*
Ampicillin	50 mg/ml in water	50 µg/ml	4 °C for less than 4 days
Kanamycin	50 mg/ml in water	50 µg/ml	-20 °C

***Table 2.3* Antibiotic concentrations, stock and final concentrations used in growth media**

2.1.5 Sterilisation

Solutions, media and glassware were sterilised by autoclaving in either an Astell Hearson 2000 Series Autoclave or a Prestige® Medical Series 2100 Clinical Autoclave at 121 ºC, 15 lb in^{-2} (psi) for 20 min. When necessary, solutions were filter sterilised using a 0.2 µm pore Millipore filter disc (Supor® Acrodisc® 3.2, Gelman Sciences) and a suitable sterile syringe (Plastipak®, Becton Dickinson).

2.1.6 Chemicals, enzymes and Media

Unless otherwise stated in the text all chemicals, enzymes and media are from Sigma Chemical Company. 18.2 MΩ, water produced from a Millipore Milli-RO 10 plus water purification system, was used for all experiments. Water was used as the solvent for all solutions unless otherwise stated.

2.1.7 Centrifugation

Bacterial cells from cultures of 100 –1000 ml was harvested by centrifugation at 5000 revolution per minute (RPM) for 10 min in 500 ml centrifuge tubes (Nalgene) using a Beckman J2-21 centrifuge with a JA-10 rotor. Bacterial cells from culture volumes 5-10 ml were centrifuged in 30 ml Sterilin universal tubes at 5000 g in a MSE Mistral 3000i bench centrifuge with a swing out rotor for 5-10min. Centrifugation of small volumes (<1.5ml) was carried out in 1.5 ml Eppendorf tubes at 13000 RPM in a Heraeus Instruments Biofuge pico.

2.1.8 Chemically competent *E. Coli*

E. coli strains were made competent for uptake of plasmid DNA using calcium chloride in a variation of the procedure described by Cohen *et al.* (Cohen et al., 1972). A 1 ml aliquot of a 10 ml overnight LB culture of *E. coli* was used to inoculate

1 litre non-baffle flask containing 100 ml of LB. The culture was incubated (without antibiotic) at 37 °C (180 rpm) until log phase was reached (A_{600} = 0.4). After 10 min on ice, the cells were harvested by centrifugation at 5000 × g at 4 °C for 5 min. The medium was decanted and the cells gently re-suspended in 8 ml of ice cold 100 mM $MgCl_2$. The previous step was repeated and the cell pellet re-suspended in 4 ml of ice cold 100 mM $CaCl_2$. After 2 h on ice, the cells were competent for transformation with plasmid DNA. Aliquots of 100 µl were stored at -80°C with 25 % (v/v) glycerol in 1.5 ml Eppendorf tubes.

2.1.9 Transformation of *E. coli*

A 50-100 µl aliquot of competent cell was gently thawed on ice for 5 min, and then Plasmid DNA (2 µl containing 0.1-1.0 µg DNA) was added in. The DNA and cells were gently mixed with a pipette tip and incubated on ice for 30 min. The cells were then heat-shocked at 42°C for 1.5 min and immediately placed on ice for a further 2 min. Transformations with the plasmid encoding antibiotic resistance were plated directly onto LB agar containing the appropriate antibiotic before being inverted and incubated at 37°C for 16 h. For transformation with plasmid DNA after a ligation, 250 µl of 37°C LB medium were added and the cells were recovered for 1 h (37°C, 180 rpm in a rotary shaker). Then the complete culture was plated onto appropriate selective media and incubated as before.

2.1.10 Replication of DNA and rapid small scale plasmid preparation

More plasmid DNA was produced by transformation of the DNA into Top10 cells and growth of colonies on selective LB agar plates. Single colonies were picked and

used to inoculate 5 ml of LB with appropriate antibiotic. The culture was incubated at 37 ºC with aeration overnight, as described in Section 2.1.2. These cells were then harvested by centrifugation for 5 min and the supernatant was carefully removed. Plasmid DNA was extracted using a Qiagen QIAspin Miniprep kit as described in the manufacturer's instructions.

2.1.11 Restriction Digest of DNA

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

2.1.12 Determination of DNA concentration

DNA concentration was determined either by estimation from comparison of band size/intensity to quantified DNA ladders when electrophoresed on agarose gels or by scanning the absorbance of the appropriately diluted DNA (in MQ H_2O) between 230 nm and 340 nm using a spectrophotometer (Pharmacia Biotech Ultrospec 4000, UV/vis) Double-stranded DNA at 50 µg/ml gives an A_{260} reading of 1.0, as does single-stranded DNA or oligonucleotide at 33 µg/ml.

2.1.13 Agarose gel electrophoresis of DNA

Separation and determination of the sizes of linear DNA fragments were carried out by electrophoresis through submerged horizontal gels (Myers, 1975).

0.8% (w/v) agarose gels were made by adding 0.4 g of agarose in 50 ml of 1x Tris/Borate/EDTA TBE) buffer, boiling the solution til the agarose dissolved. Once the solution had cooled to about 50°C, ethidium bromide (1 µg ml^{-1} final) was adding,

to visualise the DNA. The gels were cast in mini-gel trays (Applied Biosystems) and once set the gel was covered in 50 ml of 1x TBE Buffer. DNA samples were prepared by the addition of loading buffer and pipette into the wells of the gel. DNA standard ladders (HyperLadder I or IV from Bioline) were also loaded. The DNA was electrophoresed at 60 mA for ~1 h.

All buffer and reagent components are shown below in Table 2.4

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

Table 2.4 Buffers for agarose gel electrophoresis (10 x stocks)

2.1.14 Visualisation of DNA

An image of the gel was obtained using a gel documentation system (Bio-Rad Gel Doc 1000, Molecular Analyst™/PC Windows Software, and photographs were produced by Mitsubishi Video Copy Processor (Model P68B) with Mitsubishi thermal paper.

2.1.15 Determination of DNA fragment size

The size of linear double stranded DNA molecules can be determined because the migration rate through gel matrices is in effect inversely proportional to the $\log_{10}$ of the size of the nucleic acid. Therefore, the sizes of DNA fragments were determined by comparing their electrophoretic mobility with that of the DNA standards of known sizes.

2.1.16 Purification of DNA fragments

2.1.16.1 PCR Purification

PCR products and restriction digested inserts were purified using QIAquick PCR Purification Kit (Qiagen) as described in manufacturer's instructions.

2.1.16.2 Gel extraction

Restriction digested vectors were separated by agarose gel electrophoresis using high purity Seachem Gold™ agarose and TBE buffer. DNA was isolated from agarose gels by excising the required band from the gel with a clean scalpel blade. The DNA was then purified using a QIAquick Gel Extraction Kit (Qiagen) following the manufacturers' protocol.

2.1.17 Ligation of Insert DNA and Vector DNA

Appropriately digested DNA extracted from an agarose gel were used for ligations. DNA molecules with cohesive ends were ligated with a final molar ratio of insert: vector of 3:1. The reaction mixture included 4 µl 5x ligation buffer (250 mM Tris/HCl, pH 7.6, 50 mM $MgCl_2$, 25 µM ATP, 25 mM dithiothreitol, 25 % (w/v) polyethylene glycol 8000) and 1 µl (5U) T4 DNA ligase (Invitrogen) made up to a final reaction volume of 20 µl with MQ H_2O. Ligations were performed at room temperature for 2 h using T4 DNA ligase (Invitrogen) as per the manufacturer's instructions, then 2 µl of

ligation product was transformed into Top 10 (Invitrogen) ***E. coli*** cells as Section 2.1.9 describes.

2.1.18 Polymerase Chain Reaction

The polymerase chain reaction (PCR) developed by Mullis and Faloona (Mullis and Faloona, 1987) was used to amplify target genes throughout this study. Two oligonucleotide primers are required for PCR, one complementary to each strand of the DNA molecule, at sites that flank the region of DNA to be amplified. A thermostable DNA polymerase catalyses the synthesis of the complementary DNA strand in the presence of dNTPs. Oligonucleotide primers were designed such that complementary sequences were normally 20 bases in length with a G/C content of approximately 40 % and a melting temperature (T_m) of >45 ^oC and within 5 ^oC difference between the two primer. T_ms were calculated using the online tool "Oligonucleotide Properties Calculator" at http://www.basic.nwu.edu/biotools/oligocalc.html which uses the following calculation formular;

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

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

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

MWG-Biotech and supplied lyophilised. Primers were dissolved in sterile double distilled water to 100 pmol/µl. Control reactions lacking template DNA were always carried out. PCRs were made up in sterile 0.2 ml or 0.5 ml Eppendorf tubes as follows:

for KOD DNA Polymerase (10 × = 1.2 M Tris-HCl, 100 mM KCl, 60 mM $(NH_4)_2SO_4$, 1 % Triton X-100, 0.01% BSA, pH 8.0)
25 mM $MgCl_2$
dNTPs (2 mM each)
50 µM oligonucleotide primer #1
50 µM oligonucleotide primer #2
Template DNA
PCR Grade Water
Novagen® KOD DNA Polymerase (2.5 U/µl)
Total Volume

The standard thermocycle was as follows:

95 °C for 1 min
95 °C for 1 min
(Lower T_m of primer pair – 5 °C) for 1 min
72 °C for 1 min/kb fragment size
72 °C for 10 min and 4 °C for 24 h

After thermocycling was complete, a 5 µl aliquot of each PCR was analysed by electrophoresis on a mini-gel (Section 2.1.13).

2.1.19 Site Directed Mutagenesis

Mutagenesis of single amino acids was carried using the site directed mutagenesis method. The site-directed mutagenesis method utilizes an appropriate double-stranded recombinant plasmid DNA and two synthetic oligonucleotide primers (MWG-Biotech AG, Germany) containing the desired mutation flanked by 10-15 nucleotides that fully complemented the DNA template. The oligonucleotide primers are extended during temperature cycling by using a thermostable KOD polymerase.

1 × Buffer for KOD DNA Polymerase (10 × = 1.2 M Tris-HCl, 100 mM KCl, 60 mM $(NH_4)_2SO_4$, 1 % Triton X-100, 0.01% BSA, pH 8.0)
25 mM $MgCl_2$
dNTPs (2 mM each)
50 µM oligonucleotide primer #1
50 µM oligonucleotide primer #2
Template DNA
PCR Grade Water
Novagen® KOD DNA Polymerase (2.5 U/µl)
Total Volume 50 µl

Reactions were made up in sterile 0.5 ml thin walled Eppendorf tubes as follows

Each reaction was overlaid with 30 µl mineral oil and transferred to the Thermal cycler (Techne PHC-3) and cycling reactions started. The cycling parameters were as follows:

95 °C for 1 min
95 °C for 1 min
55 °C for 1 min
68 °C for 1 min/ kb of plasmid length

After thermocycling the reactions were placed on ice for 2 min, 1 µl of *Dpn*I (Fermentas 10 U) was pipetted directly to each amplification reaction below the mineral oil and mixed gently by pippetting the solution up and down several times. The reaction mixtures were spun down in a microcentrifuge for 1 min and incubated at 37 ºC for 1 h in order to digest parental methylated DNA but not the unmethylated amplified DNA. The DNA was then transformed into *E. Coli* TOP10 supercompetent cells (Section II.1.10).

2.1.20 Automated DNA Sequencing

DNA sequencing was conducted using the Value Read service from MWG Biotech AG, Ebersberg, Munich, Germany using ABI 3700 sequencers and Big dye technology (Applied Biosystems). Each clone was sequenced in both the forward and reverse direction. Plasmid DNA (1.5 µg) was dried by vacuum lyophilization at room temperature in 1.5 ml tubes. Plasmids were usually sequenced with T7 promoter and T7 terminator. Genes encoded in pGEX vector were sequenced with pGex forward primer and pGex reverse primer, all primer sequences are shown in Table 2.5 shown below.

Plasmid type	***Forward primer***	***Reverse primer***
pET plasmids	T7 promoter	T7 terminator
	TAATACGACTCACTATAGGG	**CTAGTTATTGCTCAGCGGT**

Table 2.5 Plasmid sequencing primer

2.1.21 Expression of Proteins

E. *coli* strain BL21 or Tuner was transformed with recombinant plasmids conferring corresponding antibiotics resistance. Cells were plated onto LB-antibiotics agar plates and incubated overnight at 37 ºC. Colonies inoculated into 5 ml of LB-antibiotics which was incubated overnight at 37 ºC in a shaking incubator. The 5 ml cultures were inoculated into 1 L LB containing appropriate antibiotics (in a 2 L baffled flask), which was grown at 37 ºC in a shaking incubator until an OD_{600} of 0.4-0.6 was reached. The flasks were then transferred into a shaking incubator at 16 ºC and allowed to equilibrate for 1 h before induction with 0.2 mM IPTG for 16 h. Cells were harvested by centrifugation at 5000 rpm and resuspended in TALON buffer (20mM Tris , 300mM NaCl pH 8.0). The cell suspension was sonicated for 2 min using a B. Braun Labsonic U sonicator set at low intensity ~42 watts and 0.5 second cycling, then transferred to a 50 ml centrifuge tube (Nalgene) and cell debris pelleted at 15000 rpm for 30 min at 4 ^{o}C, and the supernatant comprising the cell-free extract (CFE), was retained for further use. The cell debris-containing pellet (insoluble fraction) was resuspended in 10 ml of TALON Buffer. For biochemical assays target proteins was purified by Immobilised Metal Affinity Chromatography (IMAC) with TALON Resins (Clontech).

2.1.22 Purification of Proteins

2.1.22.1 Immobilised Metal Affinity Chromatography (IMAC)

His-tagged proteins can be purified by (IMAC) as the Histidine side-chain interacts with electropositive transition metal immobilised in a column, this interaction can be disrupted by imadizol. The His-tag protein can be eluted from the column matrix using an imadizole gradient.

TALON™ (Clontech Laboratories Inc.) columns with 2 ml TALON™ resin containing cobalt ions were prepared by washing in 10 volumes water and then TALON^{TM} buffer (20 mM Tris/HCl buffer, pH 8.0, containing 300 mM NaCl). Cell free extracts (CFEs) were filtered (0.45µm) and poured through the column. The column was washed with 4 x 5 ml of TALON^{TM} buffer. The protein was eluted with 2 x 5 ml of TALON™ buffer containing 10 mM imidizole followed by 2 x 5 ml of TALON^{TM} buffer containing 100 mM imidizole. Analysis by SDS-PAGE of each stage of the purification indicated which fractions contained the purified protein.

2.1.22.2 Gel Filitration Using FPLC

The column (HiLoad™ 16/60 Superdex™ 200 Prep grade, GE Healthcare) was washed with 150 ml of Buffer A containing 150 mM NaCl at 1 ml min-1. The protein was concentrated to a volume of 5.5 ml and filtered (0.45 µM) before being loaded onto the column. The automatic system (Bio-Rad) eluted the proteins in 165 ml of Buffer A containing 150 mM NaCl at 1 ml min-1. The fractions above a threshold value of 0.02 Absorbance Units at a wavelength of 280 nm were collected and analysed by SDS-PAGE.

2.1.23 SDS PAGE

SDS-PAGE method described by Laemmli (Laemmli, 1970) was used to analyse the size, relative purity and relative quantity of protein. Gels consisting of 10 – 15 % polyacrylamide (Acrylogel 3; BDH Electran®) were typically used dependent on size of protein of interest. The gels are composed of two parts, namely the resolving gel and the stacking gel (Table 2.6). The system used was the AE-6450 apparatus from ATTO Corporation (Genetic Research Instruments) which employs 12 cm x 10 cm glass plates. The resolving gel was poured into the plates covered with water and allowed to polymerize. The water is then removed and the stacking gel poured on top of the resolving gel. This is allowed to polymerize with a comb in place. The comb and the rubber seal around the glass plates were removed when the gel was set. Protein samples were prepared by mixing 15 µl of appropriately diluted protein solution, 7 µl of loading buffer (Table 2.6) and boiling for three min. The plates were placed in the gel tank that was filled with running buffer (Table 2.6). The samples were loaded into the well of the gel, which was then electrophoresed at a current of 30 A per gel.

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

2.1.23 Determination of Protein Concentration

Pure protein concentration was determined using published methods (Gill and von Hippel, 1989, Pace et al., 1995). The absorbance between 280nm and 320nm of the appropriately diluted pure proteins was measured with a spectrophotometer and the data were used to calculate protein concentrations, using the formula:

$$A = \varepsilon C l D$$

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

The ProtParam program (www.expasy.com) was used to calculate the molar extinction coefficient of protein. The length of the light path was always 1 cm.

2.1.24 Concentration Protein

Protein solutions were concentrated using 20 ml or 2 ml Vivaspin™ centrifugal concentrators (VivaScience) with 10 or 30 kDa molecular weight cut off filters as appropriate after filtering (0.22 µm). Centrifugation was performed at 3500g using a MSE Mistral 3000i bench centrifuge with a swing out rotor at 10 ºC.

Resolving gel	
0.75 M Tris/HCl buffer, pH 8.8 with 0.2 % SDS	9.4 ml
40 % (w/v) Acrylogel (BDH Electran; acrylamide, 3 % (w/v) bisacrylamide)	5.8 ml
d.d. H_2O	3.5 ml
10 % (w/v) Ammonium persulphate	90 µl
TEMED	30 µl
Stacking gel	
0.25 M Tris/HCl buffer, pH 8.8 with 0.2 % SDS	3.75 ml
40 % (w/v) Acrylogel, 3 % (w/v) acrylamide, bisacrylamide	0.75 ml
d.d. H_2O	3.0 ml
10 % (w/v) Ammonium persulphate	60 µl
TEMED	20 µl
Loading buffer	
SDS	10 % (w/v)
0.25 M Tris/HCl buffer, pH 6.8) with 0.2 % SDS	5 ml
Glycerol	25 % (w/v)
β-mercaptoethanol	2.5 ml
Bromophenol blue dye	0.1 %
Running buffer	
Tris/190 mM glycine, pH 8.3	32 mM
SDS	0.1 %

Table 2.6 Composition of resolving gel, stacking gel and loading buffer for SDS-PAGE

2.2 Bioinformatics

2.2.1 Alignments

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

2.2.2 Prediction of prokaryotic signal peptides

The presence or absence of signal peptides, and the signal peptide cleavage position, was predicted using the SignalP 3.0 software hosted at www.cbs.dtu.dk/services/SignalP/(Bendtsen et al., 2004).

2.3 Biochemistry

Unless otherwise stated all enzyme assays were carried out at 37°C. All assays were repeated at least twice.

2.3.1 Enzyme assays

Unless otherwise stated, all enzyme assays were performed at, and all reagents and cuvettes pre-warmed to, 37 °C. All assays were repeated at least three times. Graphs were plotted in GraphPad Prism 5.0 which was also used to calculate slopes, gradients and standard errors. The non-linear “one-site” binding model was used to fit kinetic data to estimate K_m and k_{cat}.

2.3.1.1 Fructose detection using Fructose assay kit (megazyme)

Unless otherwise stated substrates used in fructose detection assays were levan and inulin. The standard reaction was carried out in 500 µl at 37°C in 50 mM Sodium phosphate, 50 mM sodium acetate pH 5.5. Various substrate concentrations were used in order to calculate the kinetic parameters of the enzymes. Fructose release was measured over 20 mins using a fructose assay kit from megazyme. The kit utilises three linker enzymes to measure the fructose liberated from polysaccharide. The liberated fructose is phosphorylated to fructose-6-phosphate (F-6-P) by Hexokinase (HK). Fructose-6-phosphate is converted to glucose-6-phosphate (G-6-P) by Phosphoglucose isomerase (PGI). In the presence of glucose-6-phosphate dehydrogenase (G6P-DH) glucose-6-phosphate is oxidised by nicotinamide-adenine dinucleotide phosphate (NADP+) to gluconate-6-phosphate. This leads to the formation of reduced nicotinamide-adinine dinuclecotide phosphate (NADPH). The amount of NADPH formed is stoichometric with the amount of fructose released this is measured by the increase in absorbance at 340 nm using the extinction coefficient of 6300 M^{-1} $cm^{-1.}$

2.3.1.2 Fructose detection using HPLC analysis

Standard reactions were carried out in a total volume of 1500 µl at 37 °C in 50 mM sodium phosphate, 50 mM sodium acetate pH 5.5. Various substrate concentrations were used in order to calculate the kinetic parameters of the enzymes. Assays were incubated for 30 min and four 200 ul time points were taken, each aliquot was heat inactivated by boiling for 10 min. Following inactivation the samples were centrifuged at 13000 RPM for 10mins to pellet any undigested material. The samples were then diluted 5-fold and subjected to HPLC analysis . Each reaction contained a fucose

internal standard and all data obtained was normalised to this standard and converted to a fructose concentration using a fructose standard curve.

2.3.1.3 3, 5-Dinitrosalicyclic acid assay (DNSA)

The rate of hydrolysis was monitored by the increase in reducing sugar formed over time. The free anomeric carbon at the end of a polysaccharide can open from its more common cyclic conformation and act as a weak reducing agent. Each time a glycosidic bond is hydrolysed a new reducing end is formed, the concentration of which can be determined with the DNSA reagent using the Miller method (Miller, 1959) A 500 µl aliquot of an enzyme reaction was added to 500 µl DNSA reagent (1 % (w/v) DNSA, 0.2 % (v/v) phenol, 1 % (w/v) NaOH, 0.002 % glucose, 0.05 % (w/v) $NaSO_3$) to terminate the reaction. The tube was then boiled 20 min, placed on ice for 10 min, equilibrated to room temperature and the absorbance read at 575 nm. A standard curve of 0-1000 µg/ml monosaccharide (plus polysaccharide substrate) was used to quantify the released reducing sugar.

2.3.2 Thin Layer Chromatography (TLC)

A TLC plate (Silicagel 60, 20 × 20, Merck) was cut to the desired size (the minimum height should be 10 cm). Between 5 and 10 µl of samples were spotted on the plate, separated by 10 mm. Solvent (50 ml) comprising freshly made 1-butanol/acetic acid/water (2:1:1, v/v) was poured into a glass chromatography tank (23 × 23 × 7.5) and covered tightly. Vapours were allowed to equilibrate for at least 2 h before use. The TLC plate was placed into the tank and samples allowed to migrate until the running buffer reached ≈ 1 cm to the top of the plate. The plate was dried carefully using a hairdryer and put back in the tank for another run. The plate was dried again and immersed for a few seconds in an orcinol sulphuric acid reagent (sulphuric

acid/ethanol/water 3:70:20 v/v, orcinol 1 ‰), dried carefully and heated until sugars were developed, at 120 °C (5 -10 min). Plates were recorded using a digital camera (Canon PowerShoot A75). Standards consisting of known monosaccharides and oligosaccharides were spotted on the TLC plate . In order to detect the presence of sugar in a sample quickly, 5 µl were spotted on a section of TLC plate and immersed for a few second in orcinol sulphuric acid reagent, dried carefully and heated until sugars were revealed, as previously described.

2.3.3 High Pressure Liquid Chromatography (HPLC)

Enzyme catalysed hydrolysis reactions of polysaccharides were analysed using an analytical CARBOPAC™ PA-100 anion exchange column (Dionex) equipped with a CARBOPAC™ PA-100 guard column. The fully automated system (ICS-3000 gradient pump, detector compartment, electrochemical detector, auto sampler) had a loop size of 100 µl, flow rate of 1.0 ml/min, room temperature, pressure of ≈ 2000 psi and sugars were detected by pulsed amperometric detection (PAD). The PAD settings were E_1 = +0.05, E_2 = +0.6 and E_3 = -0.6. Typical elution conditions used were 0-10 minutes 66 mM NaOH, 10-25 min 66 mM NaOH with 0-75 mM sodium acetate linear gradient. After each run the column was washed with 500 mM sodium acetate for 10 min and then 500 mM sodium hydroxide for 10 min and then equilibrated with 66 mM sodium hydroxide for 10 min. Appropriate monosaccharides (200 µM) were used as standards and enzymatic reaction was diluted 5 fold before being loaded on to the column. Data were collected and manipulated using Chromeleon™ Chromatography Management System V.6.8 (Dionex) *via* a Chromeleon™ Server (Dionex). Final graphs were drawn with Prism 5.0 (GraphPad).

2.3.4 Isothermal Titration Calirometry (ITC)

The thermodynamics and binding affinities of carbohydrate protein interaction were investigated by ITC using a MicroCal™ VP-Isothermal Titration Calorimeter. ITC measurements were made at 25 °C following standard procedures. The proteins were dialyzed extensively against buffer and the ligands were dissolved in the dialysis buffer to minimize heats of dilution. Degassed protein solution (1.4331 ml) at high concentration (80 - 145 µM) was equilibrated in a reaction cell maintained at 25 °C. To this solution, 28 aliquots (10 µl) of ligands at an appropriate concentration were automatically injected, with rapid stirring (307 rpm), at 200 s intervals. Typically, oligosaccharides ligands were titrated at ≈ 10 mM and soluble polysaccharide ligands at 10 mg/ml. Binding was monitored by measuring heat released (exothermic binding) or absorbed (endothermic binding) for each aliquot of ligand. During the titration, the difference in electrical power required to maintain the temperature of the reaction cell versus the temperature of the reference cell was recorded, and from these differences the heat change on binding calculated.

The molar concentration of binding sites present in polysaccharide ligands was determined following the method of Szabo *et al.* (Szabo et al., 2001). After determining that the test protein contained a single binding site, by titrations with oligosaccharides of known molarity, the molar concentration of polysaccharide ligands was fitted in an iterative fashion until the n-value was as close as possible to one. Integrated heat effects, after correction for heats of dilution, were analyzed by nonlinear regression using a single site-binding model (Microcal Origin, version 7.0). The fitted data yield the association constant (K_a) and the enthalpy of binding (ΔH). Other thermodynamic parameters were calculated using the standard thermodynamic equation:

$$-RT\ln K_a = \Delta G = \Delta H - T\Delta S$$

where R is the gas constant (1.99 $cal.K^{-1}.mol^{-1}$), T is the temperature in degree absolute (298.15 K), Δ*G* is the change in free enthalpy and ΔS is the entropy of binding.

2.3.5 Affinity Gel electrophoresis

An affinity electrophoresis method using native polyacrylamide gels containing soluble ligand sugars was developed based on the method of Tomme et al (1996). A continuous gel system was used with the same gel apparatus (ATTO Corporation) as used for SDS-PAGE (Section II.1.25). Gels contained 7.5 % (w/v) acrylamide (Acrylogel 3; BDH Electran™) in 25 mM Tris, 250 mM glycine buffer, pH 8.3, which comprised the Running and Sample buffers. For ligand-containing gels, appropriate polysaccharides were added at 0.001 - 0.1 % (w/v) final concentration prior to polymerisation. Pure proteins (6 µg) in 7 µl of loading buffer, containing 5 % (v/v) glycerol and 0.0025 % Bromophenol, Blue final concentrations, were electrophoresed at 10 mA/gel at room temperature for 2.5 h. Proteins were then stained in 0.4 % (w/v) Coomassie Blue, 40 % (v/v) methanol, 10 % (v/v) glacial acetic acid for ≈ 1 h and then destained in 40 % (v/v) methanol, 10 % glacial acetic acid. Gels with and without ligands were run in the same gel box with identical samples loaded on each. BSA (15 µg) was used as a negative, non-interacting control.

2.3.6 Analytical Ultracentrifugation (AUC)

All analytical Ultracentrifugation experiments were carried out by Dr Alexandra Solovyova in the pinnale laboratory in the Institute of Cell and Molecular Biosciences. Below is a brief description of the AUC method.

Sedimentation velocity (SV) experiments were carried out in a Beckman Coulter (Palo Alto, CA, USA) ProteomeLab XL-I analytical ultracentrifuge using interference optics. All AUC runs were carried out at the rotation speed of 48,000 rpm and experimental temperature 20°C; the velocity scans were taken 1 second apart, 600 scans in total. The sample volume was 400 μl. The concentrations of BT3082 were used as follows: 200μM, 150μM, 100μM and 50μM and combined with X40 added in 1000 times in excess. The partial specific volumes ($\bar{\nu}$) for the proteins were calculated from the protein amino acid sequence, using the program SEDNTERP (Laue *et al.,* 1992) and extrapolated to the experimental temperature following the method as described in (Durchschlag 1986). The density and viscosity of the buffer (50 mM phosphate pH7.0, 200 mM NaCl) at the experimental temperature was also calculated using SEDNTERP.

The distributions of sedimenting material were modelled as a distribution of Lamm equation solutions (Schuck, 1998) where the measured boundary a(r,t) was modelled as an integral over differential concentration distributions c(s):

$$a(r,t) = \int c(s)\chi(s,D,r,t)ds + \varphi \qquad (1)$$

where φ is the noise component, r is the distance from the centre of rotation and t is time. The expression χ(s, D, r, t) denotes the solution of the Lamm equation (Lamm,

1929) for a single species by finite element methods (Demeler and Saber, 1998). Eq. 1 is solved numerically by discretisation into a grid of 200 sedimentation coefficients for both interference data and the best-fit concentrations for each plausible species are calculated according to a linear least squares fit (as implemented in SEDFIT (http://www.analyticalultracentrifugation.com)). The sedimentation velocity profiles were fitted using a maximum entropy regularisation parameter of p = 0.95. The data were fitted for the position of the meniscus, value of the friction coefficient (f/f_0) and both the time- and radial-independent noise. The weight average sedimentation coefficient was calculated by integrating the differential sedimentation coefficient distribution (Schuck, 2003). The absolute values of sedimentation coefficients and molecular masses were estimated by integration of the peaks on c(s) and c(M) distributions respectively (Dam and Schuck 2004).

The program HYDROPRO (Garcia de la Torre *et al.,* 2000) was used to calculate the sedimentation coefficient values for BT3082 and X40 the basis of their atomic coordinates. Crystallographic symmetrical dimer of BT3082 was built using the program PyMol (DeLano Scientific) and its sedimentation coefficient was calculated using HYDROPRO.

2.3.7 Purification of oligosaccharides by size exclusion chromatography

1 g of Levan was resuspended in 10 ml of H20, and incubated in 50 mM HEPES pH 7.5 with an endo β2,6 fructanase (BT1760) at 37 °C to partially digest the polysaccharide. The reaction was heat inactivated by boiling for 10 min. Aliquots were taken at appropriate time points and analysed by TLC to follow the reaction. The sample was freeze dried and resuspended in 3 ml of H_2O for size exclusion purification. The partially digested levan products were were separated by size

exclusion chromatography using P2 Bio-gel (Bio-Rad) matrix packed in 2 Glass Econo-Column™ (2.5 cm × 80.0 cm) with a flow adaptor (Bio-Rad) at 0.2 ml/min in degassed MQ H2O.The partially digested levan products was loaded directly onto the columns, which was then run with MQ H_2O as mobile phase at 0.2 ml/min using a peristaltic pump (LKB Bromma 2132 microperpex™). The 2 ml fractions were collected continuously 16 h after loading for 39 h using a Bio-Rad model 2110 fraction collector. A 2 µl aliquot of every three fraction was analysed by TLC identifying fractions of interest to be pooled and freeze dried for further analysis.

Chapter 3: The Effect of CBMX40 on the Catalytic Activity of its Native Enzyme SacC

3.1 Introduction

Many glycoside hydrolases, particularly those that target insoluble substrates, contain non-catalytic carbohydrate binding modules (CBMs) (Boraston et al., 2004). As with the glycoside hydrolase families, CBMs have also been grouped into sequence-based families in the CAZY database (www.cazy.org). Prior to the start of this project a group of sequence-related X modules (where the X signifies a module of unknown function) were identified in a subset of glycoside hydrolases by Bernard Henrissat. This family of X modules was termed X40, and Figure 3.1 illustrates the new family as a phylogenetic tree. A notable feature of the phlyogenetic tree is that many of the X40 sequences are found in glycoside hydrolase family 32 (GH32) enzymes. Numerous enzymes in GH32 have been characterised and shown to hydrolyse fructan polysaccharides, such as inulin and levan, through an exo- or endo-mode of action (Lammens et al., 2009). Although a CBM that binds specifically to the β-2,1-linked fructose polymer, inulin has been identified (Lee et al., 2004), there have been no reports of CBMs that bind levan, a predominantly β-2,6-linked fructose-containing polysaccharide. It is potentially significant, therefore, that an X40 module is appended to a *Bacillus subtilis* GH32 enzyme, SacC, which displays exo-levenase activity (Wanker et al., 1995). It is possible that the X40 module in SacC (defined henceforth as CBMX40) may comprise the first example of a levan-specific CBM. To test this hypothesis recombinant CBMX40 was expressed in *Escherichia coli* and its capacity to bind to complex carbohydrates was assessed (experiments

carried out by Dr James Flint). Isothermal titration calirometry (ITC) showed that CBMX40 bound to levan but not to inulin or fructose, Figure 3.2. The data also showed that the protein displayed higher affinity for levanbiose (K_A 16.9 X 10^3 M^{-1}), the disaccharide derived from levan, than the parent polysaccharide (K_A 3.3 X 10^3 M^{-1}), Table 3.1. The crystal structure of CBMX40 (determined by Dr James Flint), Figure 3.2, revealed a canonical jelly roll fold, while mutagenesis studies showed that the ligand binding site is between the loops connecting the β sheets, Table 3.1 and Figure 3.2. The structure of the protein in complex with ligand was not obtained.

3.2 Objectives

The objective of this chapter is to explore the ligand specificity of CBMX40 in greater detail and to characterise its role in the SacC enzyme.

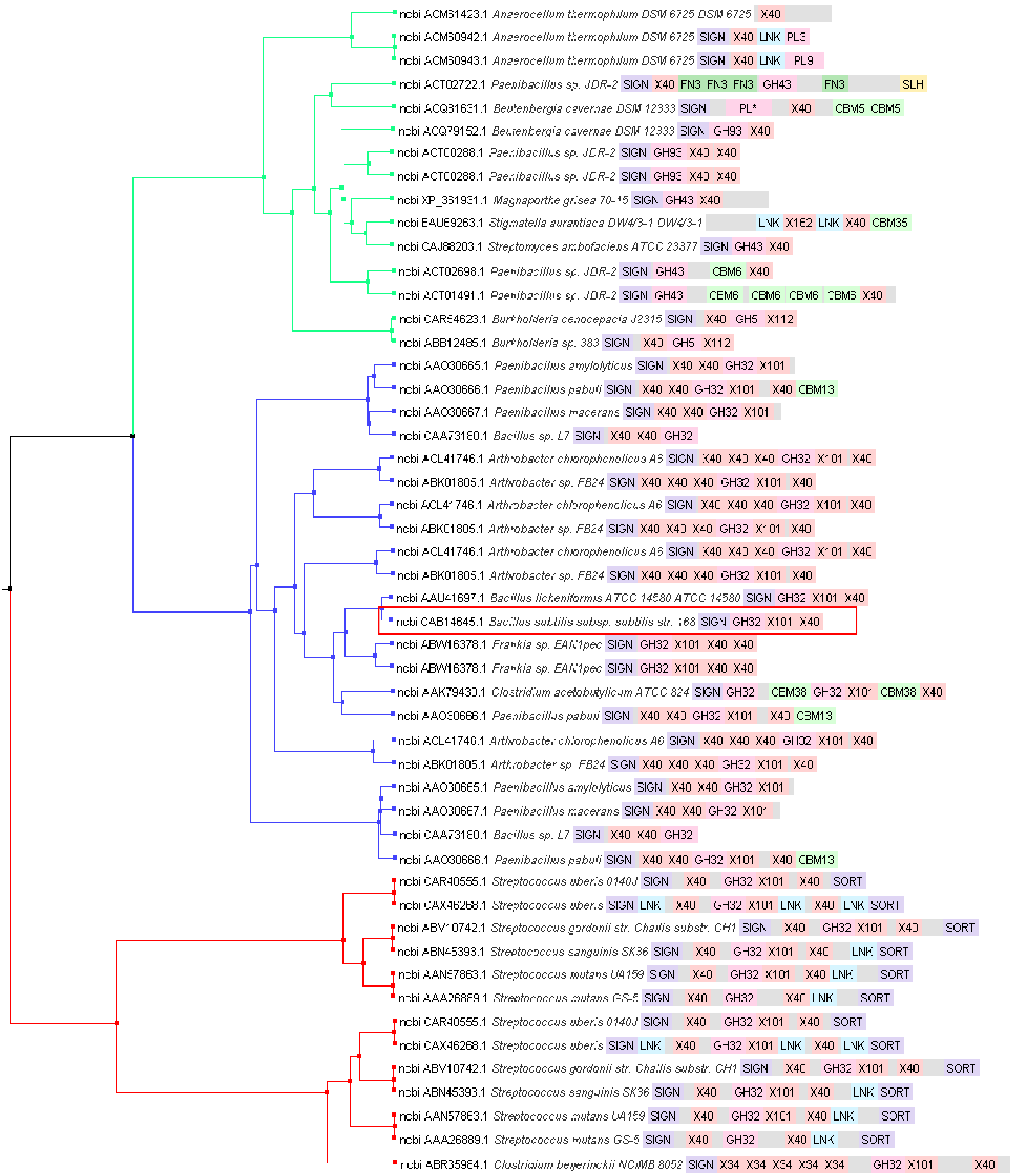


Figure 3.1 Phylogenetic tree of putative X40 containing protein sequences

SacC is highlighted in red box. Tree made with the CAZyBox (in house program interfaced with the CAZy database); sequences were aligned with MUSCLE and a distance matrix was calculated using the BLOSSUM62 substitution parameters. The matrix was then used as input for WMV-SECATOR for making clusters (subfamilies). The leaves of the tree are labelled with the accession numbers, the species, and the modular structure of the proteins. If a protein has more than one X40, it appears more than once on the tree

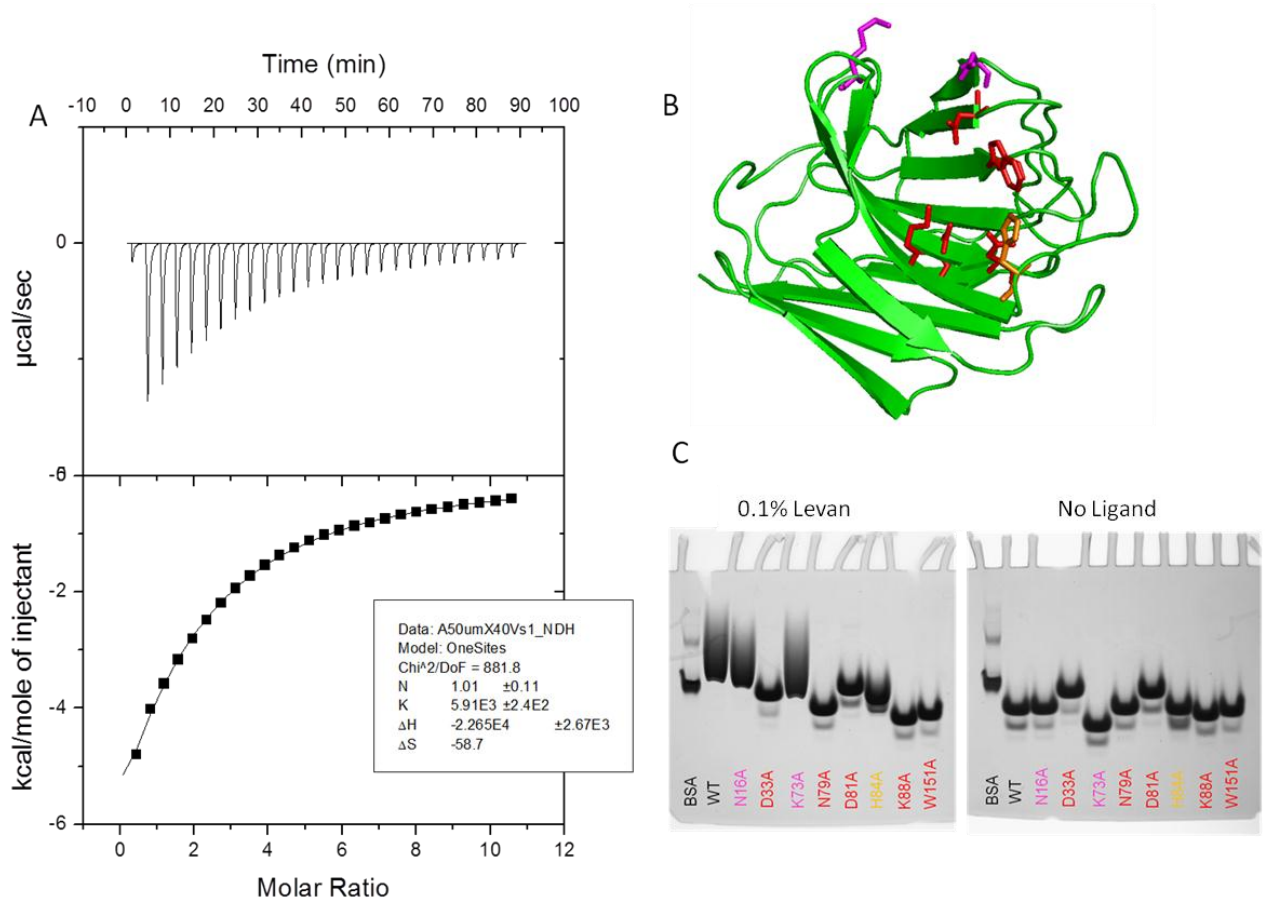


Figure 3.2 CBMX40 binding to Levan

Panel A- Example ITC of CBMX40 binding to levan using 10mg/ml Levan in the syringe and 50 µM CBMX40 in the cell. The top frame shows the raw heats and the bottom shows the integrated peak areas fitted to a one site model using Microcal origin software. Panel B- Cartoon representation of CBMX40 molecule. The protein displays a β canonical jelly roll fold, the amino acids potentially involved in binding are shown in stick formation and colour coded to correspond with the affinity gel in Panel C. Panel C- Affinity gel showing the binding of CBMX40 mutants to levan. Each amino acid is illustrated in the structure in Panel B. Amino acids coloured red are essential for binding; orange, while not essential, still effect binding; those coloured pink do not affect levan binding. BSA is shown as a non interacting control. Data in figure from a personal communication from Dr James Flint.

CBMX40	Ligand	K_a ($x10^3$ M^{-1})	ΔG ($x10^3$ kcal/mole)	ΔH ($x10^3$ kcal/mole)	TΔS (kcal/mole)	N
Wild type	Levanbiose	16.9	-5.8	-12.1	-6.3	1.1
Wild type	Levantriose	10.6	-5.5	-11.2	-5.7	1.1
Wild type	Levantetraose	8.2	-5.3	-12.3	-6.9	1.0
Wild type	Levanpentaose	8.8	-5.4	-13.4	-8.0	1.0
Wild type	Levan	3.3	-4.8	-18.4	-13.6	1.0
N16A	Levan	1.4	-4.3	-12.5	-8.2	1.0
D33A	Levan	NDB	-	-	-	-
K73A	Levan	2.9	-4.7	-13.1	-8.4	1.1
N79A	Levan	NDB	-	-	-	-
D81A	Levan	<1.0	nd	nd	nd	1.0
H84A	Levan	3.2	-4.8	-13.3	-8.5	1.0
K88A	Levan	NDB	-	-	-	-
W151A	Levan	NDB	-	-	-	-

Table 3.1 Affinity and thermodynamics of the binding of wild type and mutants of CBMX40 to levan and levan oligosaccharides.

NDB- No detectable binding. nd- not determined. ITC traces fitted to a single site model. For polysaccharides when molar concentrations were unknown the traces were fitted to an N value as close to one as possible by varying the molar concentration of the polysaccharide. All data was from a personal communication from Dr James Flint.

3.3 Results

3.3.1 Gene Cloning, Protein expression and purification.

To characterise the effect of CBMX40 on the activity of SacC two gene constructs were designed encoding mature SacC ($SacC_M$; residues Asp 23 to Ser677 of full length), lacking its signal peptide, and the GH32 catalytic domain ($SacC_{CD}$; residues Asp23 to Gly521 of full length SacC), respectively, Figure 3.3. The genes encoding full length SacC and $SacC_{CD}$ were cloned into the *E. coli* expression vector pET28a, CBMX40 (residues Thr522 to Ser677 of full length SacC) was previously cloned into the *E. coli* expression vector pET22a by Dr James Flint. The appropriate DNA fragments were amplified by the polymerase chain reaction (PCR) from genomic *B.subtllis DNA* using primers designed to bind to the 3' and 5' regions of the SacC gene, shown in Table 3.2. The PCR fragments were digested with BamHI and XhoI and ligated into the appropriate vectors as described in Section 2.1.17. Transformants containing plasmid DNA were tested for the presence of insert by digestion with BamHI and XhoI and visualised by agarose gel electrophoresis, described in Section 2.1.13. An example gel showing positive transformants containing the desired insert is shown in Figure 3.4. The clones were sequenced (by MWG; data not shown) to confirm that no mutations were introduced during PCR amplification. The proteins encoded by these plasmids contain a C-terminal His_6-tag. All the recombinant proteins were expressed in BL21(DE3) *E. coli* cells; cultures were grown at 37 °C to mid-exponential phase (A_{600nm} 0.6) and recombinant protein production was induced with 1 mM IPTG at 16 °C and incubation for 16 h. The proteins were purified by IMAC to apparent electrophoretic homogeneity. Figure 3.5 shows an SDS-PAGE of the purification of $SacC_M$ and $SacC_{CD}$ and CBMX40.A mutation (W152A), which inactivates CBMX40, was introduced into $SacC_M$. The

primers that contained the mutation (listed in Table 3.2), were used in the site-directed mutagenesis PCR (Section 2.1.19) in which $SacC_M$-pET28a (contains *$SacC_M$*) was the template.

Primer name	Sequence	Restriction site
$SacC_M$ FWD	CCG CTA CAC **GGA TCC** gat gca gcc gat tca agc tac	BamHI
$SacC_M$ REV	CTC CAG **CTC GAG** AGA CTC CTT CGT TAC ATT CTG AAA GAC	XhoI
$SacC_{CD}$ FWD	CCG CTA CAC **GGA TCC** gat gca gcc gat tca agc tac	BamHI
$SacC_{CD}$ REV	CTC CAG CTC GAG CGT TCC CCA TAC CTT TTT TAA AGG GTG	XhoI
$SacC_M$ W152A FWD	TTT GGC TTG AAT GTG GCG GAC GCG ACT GCT GTC	N/A
$SacC_M$ W152A REV	GAC AGC AGT CGC GTC CGC CAC ATT CAA GCC AAA	N/A

Table 3.2 Primers utilised to clone genes encoding $SacC_M$, $SacC_{CD}$ and $SacC_M$:W152A.
Restriction sites are shown in bold type

Figure 3.3 Cartoon representation of SacC constructs

A- $SacC_M$: Mature protein (residues 23-677) comprising both the GH32 catalytic domain and CBMX40, but lacking the signal peptide. B- $SacC_{CD}$ Truncated SacC comprising the GH32 catalytic domain only (residues 23-521). C- CBMX40 (residues 522-677) . Images made using ProSITE my domains image creator.

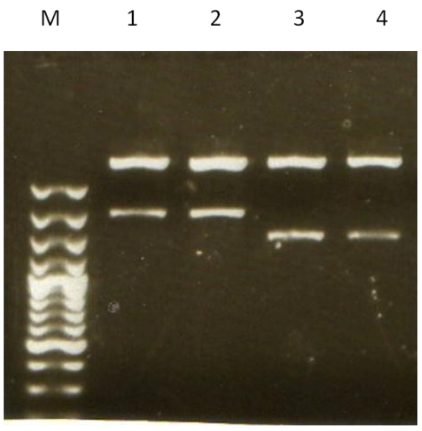


Figure 3.4 Agarose gel showing restriction digest of SacC constructs

M- DNA marker. Lanes 1 and 2 contain positive *sacC*$_M$ clones showing the insert at about 1800bp. Lanes 3 and 4 contain a truncated derivative of *sacC*$_M$ encoding $SacC_{CD}$ showing the insert at about 1500bp.

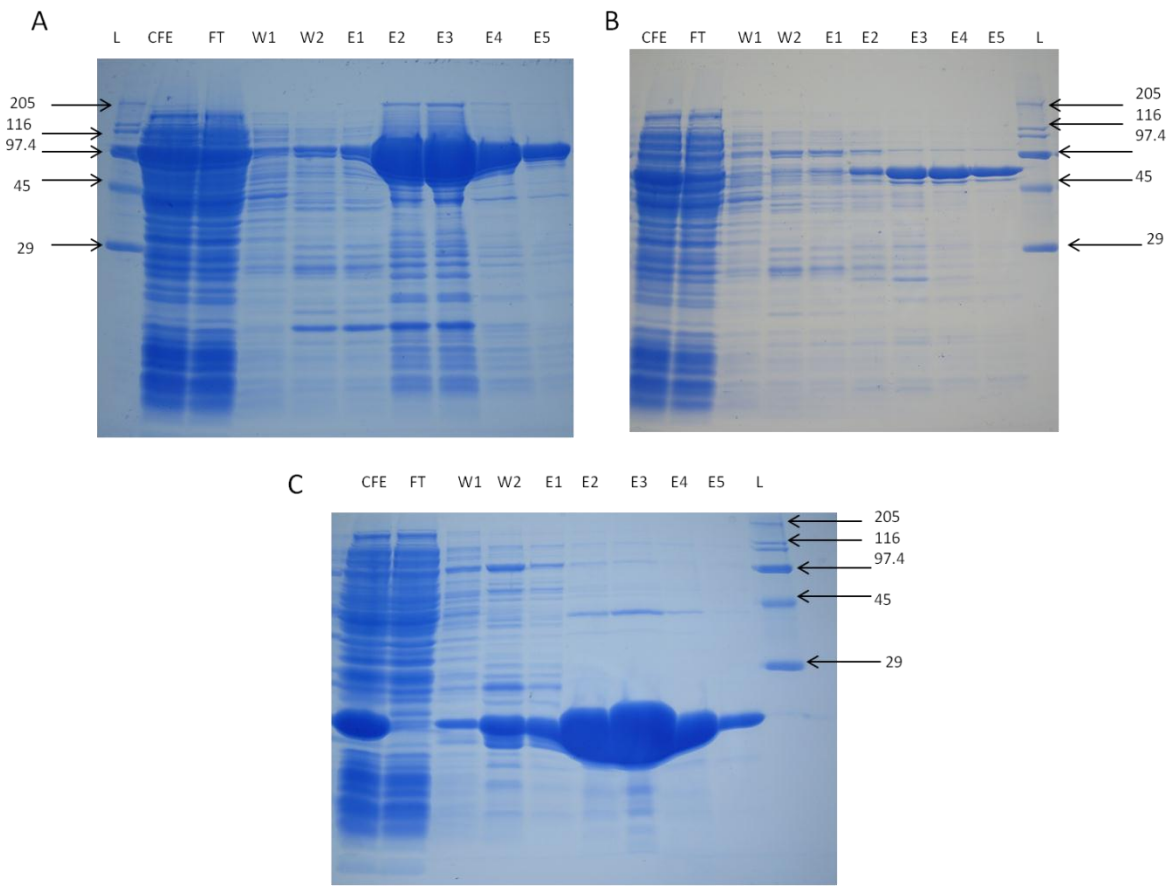


Figure 3.5 SDS PAGE gel showing IMAC purification of SacC constructs

Panel A shows IMAC purification of SacC mature protein . Panel B shows IMAC purification of SacC catalytic domain. Panel C shows IMAC purification of CBMX40. CFE= cell free extract, FT= Flow through from column, W= buffer wash, E1,E2 and E3= 10mM imadizole elution, E4 and E5= 100mM imadizole elution.

3.3.2 Further Biochemical analysis of LevX40

Prior to this project CBMX40 was characterised as a levan binding CBM that preferred the disaccharide unit of the levan polysaccharide. However, the mode of binding and polysaccharide coverage of CBMX40 was unknown. For example, it is possible, given that the appended enzyme is a exo-acting fructosidase, that the CBM targets the non-reducing end of levan, rather than binding to internal regions. To test this hypothesis CBMX40, and an inactive mutant (D147A) of the exo-acting *Bacillus* levanase BT 3082 (Sonnenburg et al., 2010), which thus binds exclusively to the non-reducing end of polysaccharide, were titrated against levan from *Serratia levanicum* using ITC, Figure 3.6. From the resultant data, it was calculated that both the BT 3082 D147A mutant and CBMX40 bound once every 24 fructose residues of the polysaccharide. To calculate the stoichiometry the concentration of ligand was changed until an N value of 1 was achieved as there is only 1 binding site for each protein. The amount of ligand used in the ITC experiment in grams per litre was divided by the molecular weight of fructose thus giving a concentration of fructose this was then divided by the concentration of ligand which gave an N-value of 1, resulting in the number of sugars. An example is shown for the BT 3082 D147A mutant below.

0.25% Levan used as ligand in ITC = 2.5 g/L

Concentration of levan to fit ITC to N-value of 1= 0.063 mM

Fructose molecular weight – molecular weight of water= 162

2.5/162= 0.0154 = 15.4 mM

15.4/0.63= 24 fructose residues

so 1 BT 3082 D147A binds per 24 fructose residues

The similar stoichiometry values for the two proteins suggest that, similar to BT 3082, CBMX40 also binds to the non-reducing end of levan chains. This view was confirmed by the resolution of the crystal structure of CBMX40 in complex with ligands (see Discussion).

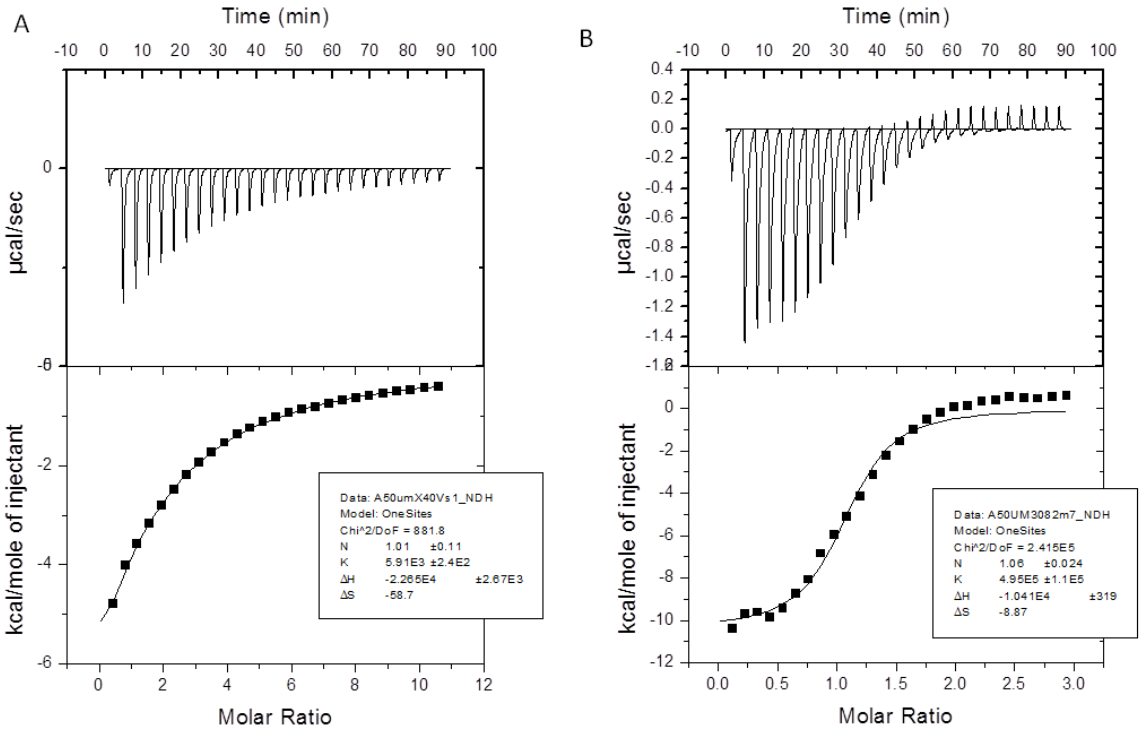


Figure 3.6 Example ITC traces

Panel A- shows 50 µM X40 in the cell and 1% Levan in the syringe. Panel B- shows 50 µM 3082 D147A in the cell and 0.25% Levan. As the concentration of levan was unknown the molarity of the polysaccharide was iteratively fitted to obtain an N value of 1.

3.3.3 Effect of CBMX40 on SacC catalytic activity against levan

3.3.3.1 SacC activity against fructan polysaccharides

In order to dissect the role of CBMX40 on the catalytic activity of the SacC enzyme, $SacC_M$,CBMX40 and $SacC_{CD}$ were including in enzyme assays in which fructose release was either measured using a fructose assay kit or by HPLC analysis as outlined in Section 2.3.1.1 and 2.3.1.2, respectively. Experiments were carried out at 37 °C in 50 mM sodium phosphate-acetate buffer pH 5.5. This buffer was used as previous studies on SacC showed that the pH optimum of the enzyme was between 5.5 and 6.

To test whether $SacC_M$ and $SacC_{CD}$ were active, the capacity of the enzymes to hydrolyse sucrose was evaluated using the glucose release assay. The data, presented in Table 3.3, showed that there was no significant difference in the activity of the two enzymes. This was expected as CBMX40 does not bind to sucrose and thus is unlikely to influence activity against the disaccharide. The data, however, confirms that the truncated catalytic module has folded correctly and that the integrity of the active site has not been compromised by the removal of the CBM.

To assess the activity of the two enzymes against levan from *Erwina herbicola*, $SacC_M$ and $SacC_{CD}$ were incubated with the polysaccharide and the fructose released was analysed by HPLC over time. The initial rates were plotted in GraphPad prism and analysed by non-linear regression to calculate the kinetic parameters of the two enzymes. Examples of the chromatography, initial rate plots and non-linear regression curves are shown in Figure 3.7 and Figure 3.8 respectively, while the kinetic data are reported in Table 3.3.

Although both $SacC_M$ and $SacC_{CD}$ were active on levan, they displayed very different kinetic parameters against the polymeric substrate. Thus, the K_m and k_{cat} of $SacC_{CD}$ are, respectively higher and lower than the corresponding kinetic parameters of $SacC_{M.}$ The catalytic efficiency k_{cat}/K_m of the truncated enzyme against *E. herbicola* levan is 120-fold lower than the full length fructosidase. Interestingly, $SacC_M$ appears to display substrate inhibition kinetics whereas $SacC_{CD}$ displays typical Mechalias-Menten kinetics. In contrast to the difference in activity against *E. herbicola* levan, the catalytic efficiency of $SacC_{CD}$ was only 20-fold lower than $SacC_M$ when the levan from *S. levanicum* was deployed as the substrate, with both enzymes displaying typical Mechalias-Menten kinetics. These data suggest that CBMX40 influences the activity of SacC against levan, although the effect of the levan binding protein modules varies depending on the source of the polysaccharide. This implies that there are significant differences in the structure of the two sources of levan. When inulin was utilised as the substrate, there was no significant difference in catalytic activity between $SacC_M$ and $SacC_{CD,}$ consistent with the observation that CBMX40 displays no affinity for the β-2,1-linked fructan . A third derivative of SacC, $SacC_M$ W152A, in which CBMX40 is inactive, displayed activity against Levan that was similar to $SacC_{C,}$ Table 3.3. This supports the view that the low activity of Sac_{CD} against levan reflects the loss of the CBM, and not through any structural changes to the catalytic module.

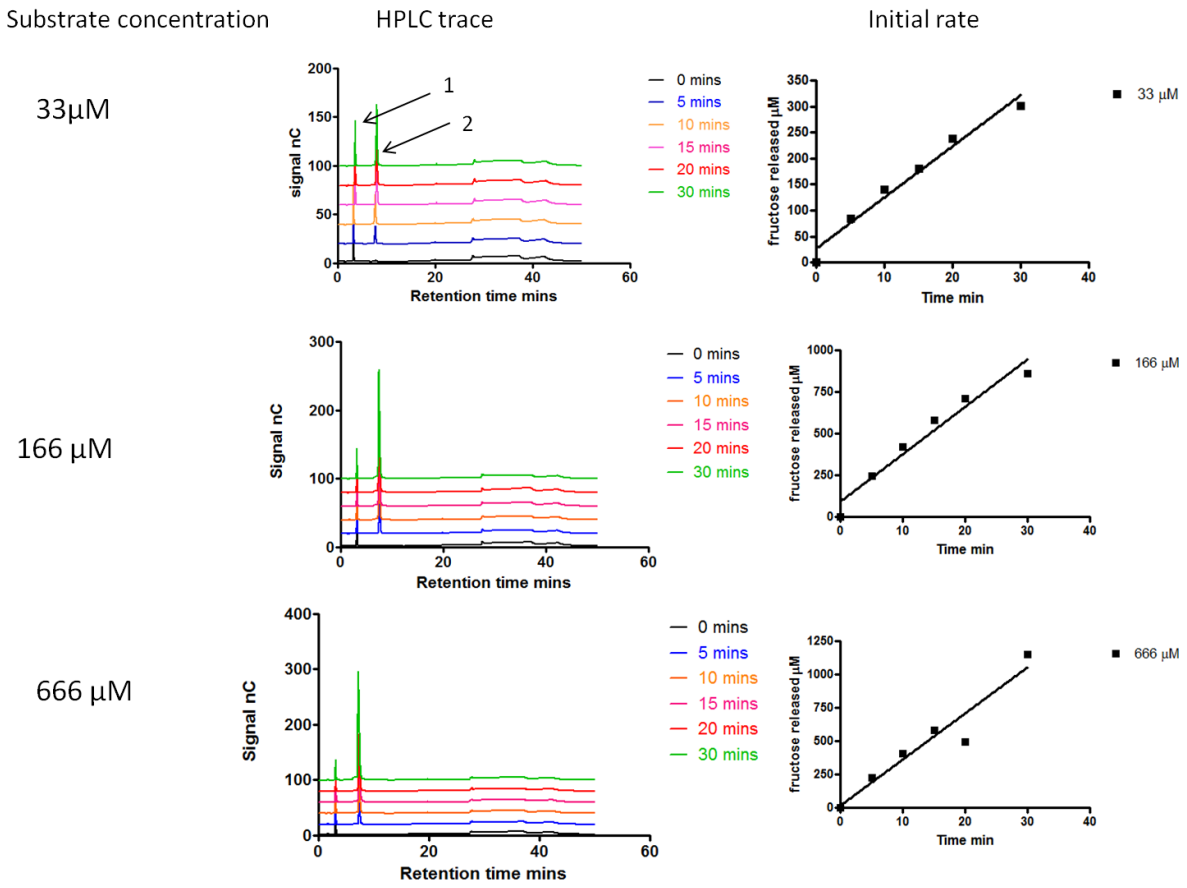


Figure 3.7 Example HPLC traces and initial rate graphs for $SacC_M$ vs Levan from *E.herbicola*

HPLC traces plotted in Graphpad prism. Each graph shows the HPLC trace for each time point for the relevant substrate concentration. Peak 1 shows the fucose internal standard, Peak 2 shows the fructose released. Initial rate graphs show the fructose released in µM as a function of time.

	$SacC_M$				$SacC_{CD}$			$SacC_M$ W151A		
	K_m µM	K_{cat} s^{-1}	K_i	K_{cat}/K_m M^{-1} s^{-1}	K_m µM	K_{cat} S^{-1}	K_{cat}/K_m M^{-1} s^{-1}	K_m µM	K_{cat} s^{-1}	K_{cat}/K_m M^{-1} s^{-1}
Substrate										
Sucrose	34400 ± 5793	152.3 ± 7.96		443	42300 ±9255	135.4 ±9.83	319			
Levan from Erwinia herbicola	272 ±154	135 ± 48.2	915 ± 595	4.9×10^5	1120 ± 472	4.6 ± 1.03	4.1×10^3	1079 ± 472	12.1 ± 1.03	1.10×10^4
Levan from Serratia levanicum	57.23 ±8.4	27± 0.43		4.7×10^5	221.7 ±124.8	4.99 ± 1.09	2.2×10^4			
Inulin	621 ±188	12.5 ±2.18		1.9×10^4	701 ± 104	7.5 ± 0.69	1.0×10^4			
Levanbiose	3400 ± 4879	149 ± 173		4.3×10^4	1900 ± 991	54 ± 19.9	2.8×10^4			
Levantriose	1537 ± 264	16.6 ± 1.3		1.0×10^4	2770 ± 1065	13.1 ± 2,3	4.7×10^3			

Table 3.3 Kinetic values for wild type and derivatives of SacC

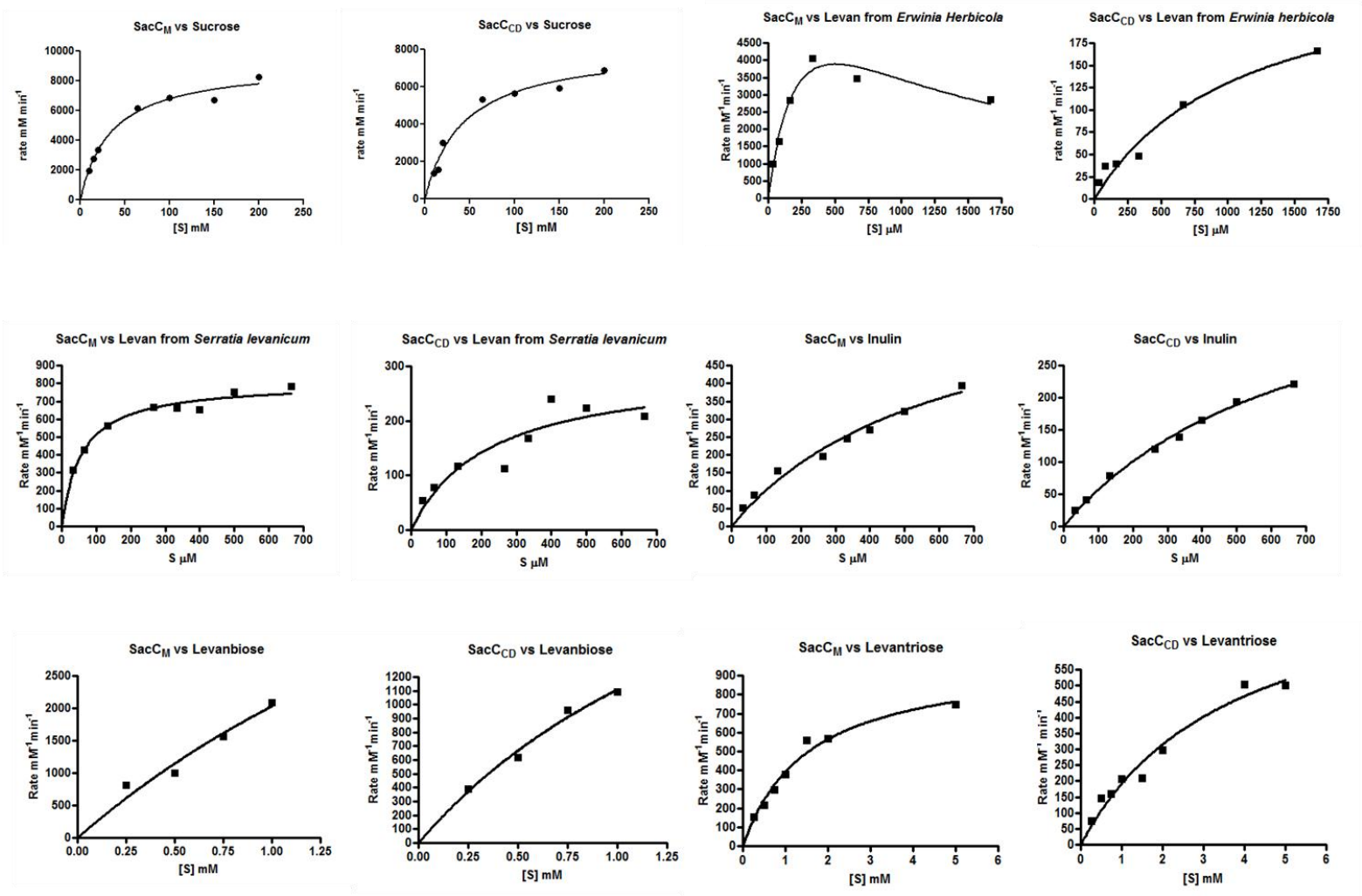


Figure 3.8 Example of kinetic analysis for SacC$_M$ and SacC$_{CD}$ against various substrates.

Initial rates derived from either glucose or fructose release using the Megazyme assay kit or by HPLC analysis, were plotted in Graphpad prism using either the Mechalias-menten equation or substrate inhibiton equation. All assays were carried out in 50mM sodium phosphate-actetate buffer pH 5.5 at 37°C

3.3.3.2 SacC activity against levan derived oligosaccharides

In order to attain levan derived oligosaccharides 1g of levan from *E. herbicola* was partially digested by 1µM Bt 1760, an endo β2,6 fructanase (Sonnenburg et al., 2010), giving a range of oligosaccharide products. This crude oligosaccharide mixture was purified by size exclusion chromatography as described in Section 2.3.7. The fractions from the purification are shown in Figure 3.9.

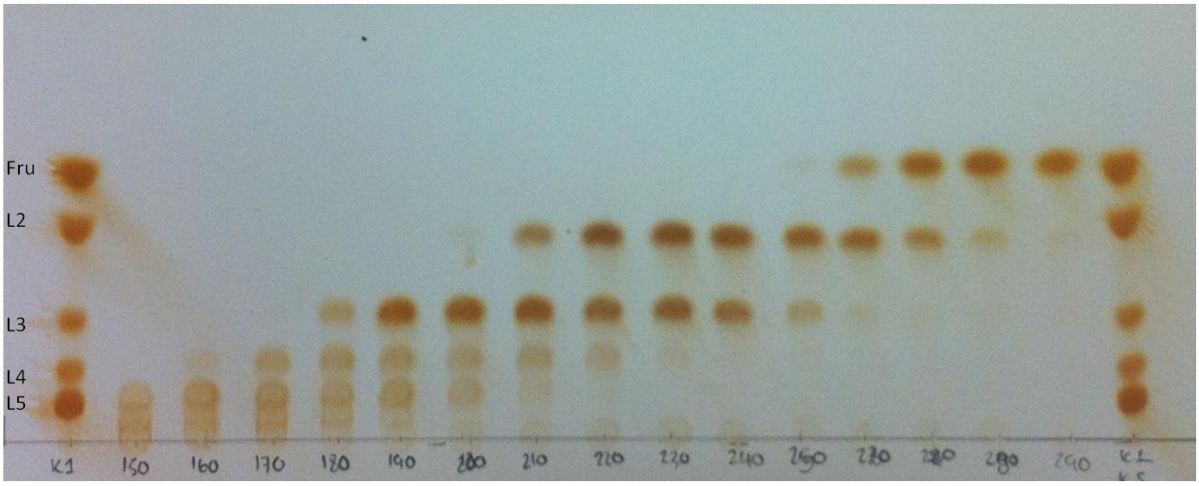


Figure 3.9 TLC analysis of P2 purified levan oligosaccharide fractions.

6 µl of every tenth (5ml) fraction from the P2 colums subjected to TLC analysis. Every second fraction was also analysed by TLC for a more accurate analysis and to decide which fractions to pool together.

The activity of $SacC_M$ against levan-derived oligosaccharides (Table 3.3) showed that the enzyme displayed similar activity against levanbiose and levantriose, indicating that the enzyme contains only two substrate binding sites, -1 and +1. The activity of Sac_{CD} against the two levan-oligosaccharides was similar to $SacC_M$, providing further support for the view that CBMX40 does not directly contribute to substrate binding in the active site (-1 and +1 subsites) of the enzyme. It should be noted, however, that the kinetic parameters for $SacC_M$ against levanbiose and levantriose are similar to the corresponding kinetic parameters of $SacC_{CD}$ against

levan from *E. herbicola*. It appears, therefore, that CBMX40, through its capacity to bind levan, potentiates the activity of the enzyme against the polysaccharide, but not against oligosaccharides.

3.3.4 Catalytic activity of $SacC_{CD}$ in the presence of CBMX40

The previous section detailed the ability of CBMX40 to potentiate the activity of the catalytic domain of SacC when in a single polypeptide. Given the potentiation of the catalytic domain by CBMX40 it is interesting to explore whether CBMX40 can enhance the activity of $SacC_{CD}$ when the two protein modules are not covalently linked. To do this $SacC_{CD}$ was incubated with levan from both *E. herbicola* and *S. levanicum* in the presence of CBMX40, and the intial rate of fructose release was measured by HPLC. A control experiment was carried out in tandem in which $SacC_{CD}$ was incubated with the two levans in the absence of CBMX40.

When Sac_{CD} was incubated with levan from *E. herbicola* in the presence of CBMX40 the rate of substrate hydrolysis increased. At the highest concentration of CBMX40 to $SacC_{CD}$, 20 µM to 10 nM, respectively, there was a 25-fold enhancement in the rate of hydrolysis, Figure 3.10. However, even at this extremely high ratio of CBMX40: $SacC_{CD}$ the rate did not reach the level of Sac_M activity against levan. Interestingly, when this experiment was repeated with levan from *S. levanicum,* the elevation in activity was only 1.5-fold, suggesting that there is a difference in the structure of the two types of levan. These data indicate that CBMX40 acts directly on levan to aid enzymatic hydrolysis, or that CBMX40 and $SacC_{CD}$ interact to form a complex, even when not physically attached.

3.3.5 $SacC_{CD}$ interacts with CBMX40 when not physically attached

Analytical ultracentrifugation (AUC) was utilised to investigate if there is an interaction between $SacC_{CD}$ and CBMX40 when the two protein modules are not covalently attached. All AUC experiments were carried out and analysed by Dr Alexandra Solovyova as described in Section 2.3.6.

The two proteins combined in a 1:1 ratio were subjected to sedimentation velocity AUC. Sedimentation velocity experiments are run at high rotor speeds and measure how proteins sediment in response to centrifugal force, which gives information on molecular weight and shape of the proteins. Calculations from the expected molecular weight and atomic coordinates of the proteins give S-values that correspond to the $SacC_{CD}$ and CBMX40 and a $SacC_{CD}$/CBMX40 complex. The data are displayed in Figure 3.11; three peaks were identified in the experiment. The first peak, at an S-value of 2.08 and a molecular weight of 19.4 kDa, corresponds to the predicted molecular mass of CBMX40. The second peak at an S-value of 4.25 and a molecular weight of 56.8 kDa corresponds to the predicted size of $SacC_{CD}$. The third peak has an experimental S-value of 4.99 and a molecular weight 72.5 KDa. Given that the calculated molecular weight of a $SacC_{CD}$/CBMX40 complex is 75.9 KDa it can be assumed this peak comprises a heterodimer of the two proteins.

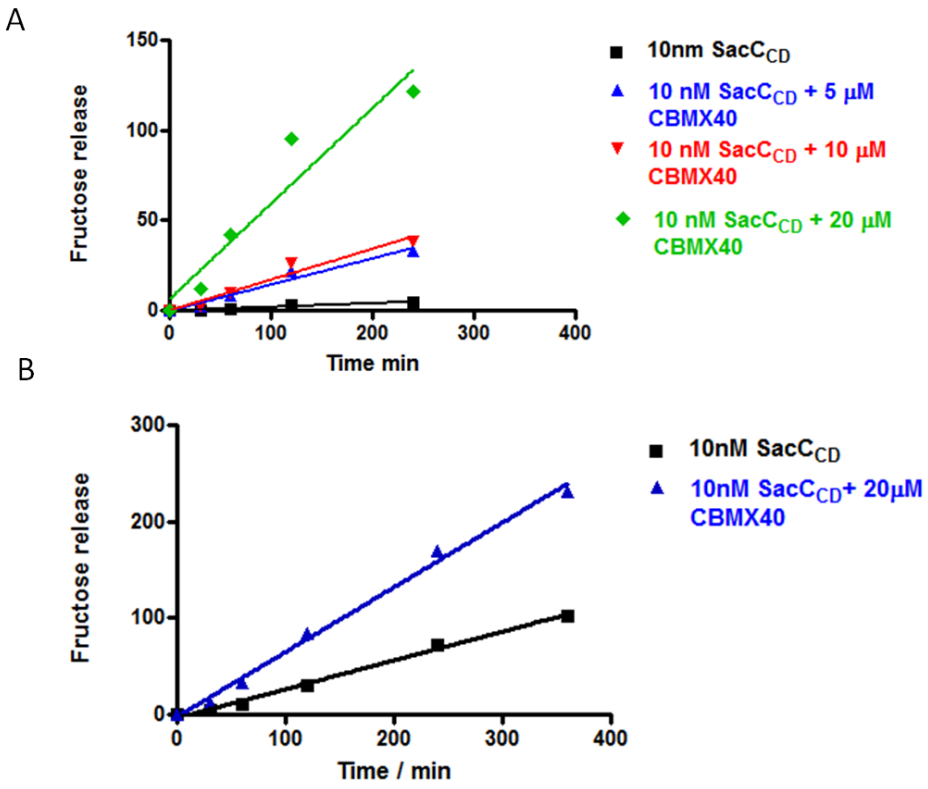


Figure 3.10 Initial rates for SacC$_{CD}$ in trans with CBMX40

Panel A shows the activity of SacC$_{CD}$ with various CBMX40 concentrations against 715 µM Levan from *E. herbicola.* Panel B shows the activity of SacC$_{CD}$ with CBMX40 against 715 µM Levan from *S. levanicum.* Both experiments were measured by HPLC analysis of fructose release.

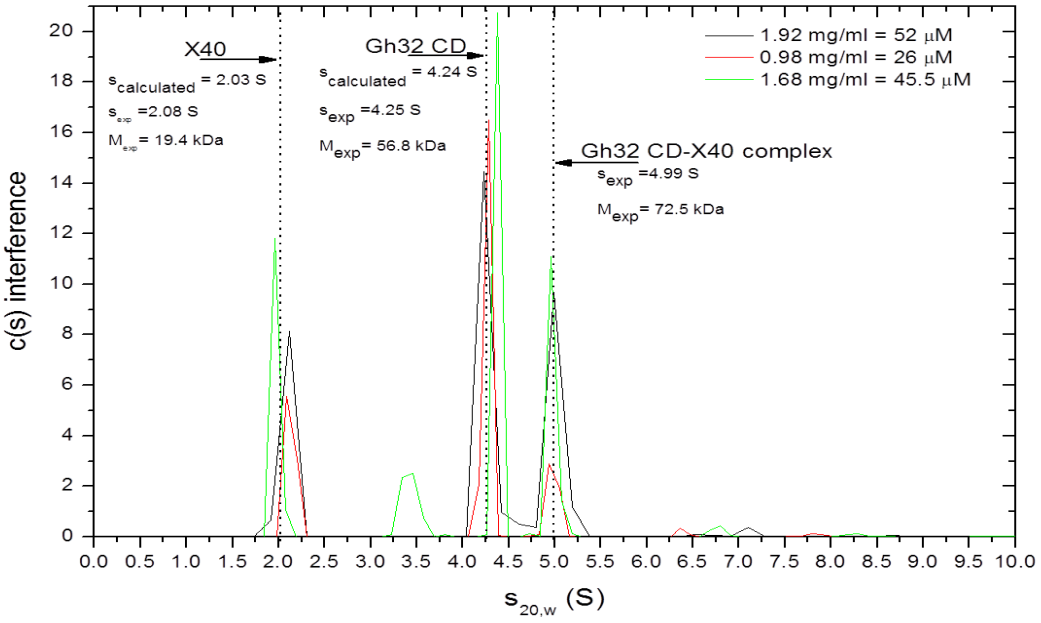


Figure 3.11 AUC showing $SacC_{CD}$/CBMX40 complex

All experiments were centrifuged run at 20 °C in 50 mM Na-HEPES, pH 7.5, containing 150 mM NaCl

3.4 Discussion

The data presented in this chapter illustrates the characterisation of CBMX40 as a novel levan binding domain. CBMX40 binds the disaccharide unit of the levan polysaccharide chain. The ITC results, described in Section 3.3.2 are indicative of CBMX40 binding to the non-reducing chain end. CBMX40 displays the same binding pattern as an inactive exo-acting levanase, BT 3082 D147A. BT 3082 D147A must bind the non reducing end of the levan chain for catalysis, therefore, CBMX40 must also bind the non reducing ends of the levan chain. A structure in complex with ligand would confirm the mode of binding. Through the duration of this project attempts were made to attain this structure, however none of these attempts were successful. To this aim a new construct was designed swapping the current C-terminal His-tag for an N-terminal clevable His-tag, primers shown in Table 3.4. The new construct was cloned and the protein purified and, after removal of the His-tag, crystallised by Mr Carl Moreland. The structure of the tagless CBMX40 in complex with fructose and levanbiose has recently been solved by Dr Arnaud Basle.

Primer name	Sequence	Restriction site
CBMX40 FWD	CCG CTA CAC CAT ATG Acg aca cct ttt atg tcc	NdeI
CBMX40 REV	CTC CGA CAT CTC GAG tta aga ctc ctt cgt tac	XhoI

Table 3.4 CBMX40 primers for pET 16b construct

The structures of CBMX40 in complex with both fructose and levanbiose are shown in Figure 3.12. The structure of CBMX40 in complex with levanbiose shows that Asp33, Asn79, Asp81 and Lys88, shown previously to be essential for binding through mutagenesis studies, Table 3.1, make direct polar contacts with the bound

fructose. Phe 91 and Trp 151 sandwich the fructose and thus make hydrophobic contacts with the two faces of the furanose ring, shown in Figure 3.13. The surface representation of CBMX40 in Figure 3.12 shows that the O6 of the fructose is buried deep into the pocket of CBMX40 and therefore CBMX40 is only capable of binding to the non-reducing end of the levan chain, consistent with the ITC-determined stoichiometry of binding.

It seems curious that while, in the crystal structure, CBMX40 binds fructose, ITC experiments revealed no evidence that the protein binds the monosaccharide. In solution fructose exists in equilibrium between the pyranose and furanose form of the sugar. This equilibrium favours the pyranose form but, in the crystal structure, fructose binds in the furanose form. To bind the furanose form in solution would be energetically unfavourable as there would be a loss of entropy. At the non-reducing end of levan polymers fructose is held in the furanose conformation, preventing a loss of entropy upon binding CBMX40 resulting in an elevation in affinity, compared to the monosaccharide.

CBMX40 makes only one polar interaction with the second fructose of levanbiose through His-84, although hydrophobic interactions between Trp-151 and the second sugar are evident. Specificity for levan rather than inulin must be conferred by the second fructose. The mutagenesis studies show that when His-84 is mutated to alanine it has very little effect on binding. If this interaction is not specific for levan than it seems puzzling that CBMX40 does not bind inulin. The specificity for levan, rather than inulin, may be conferred through hydrophobic interactions between Trp-

151 and the second fructose, although it is possible that the 2,1-linked furanose sugar makes steric clashes with the protein.

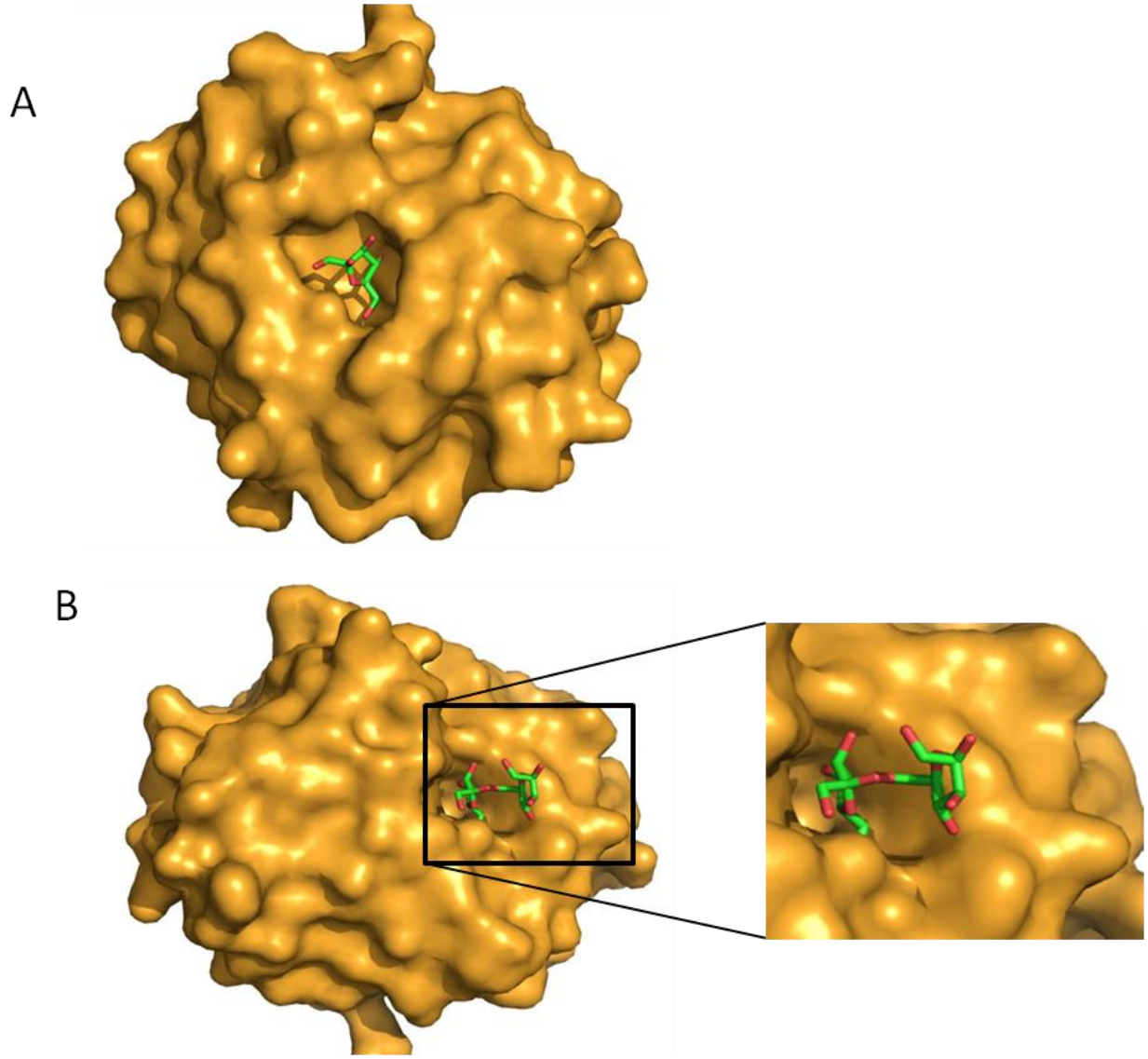
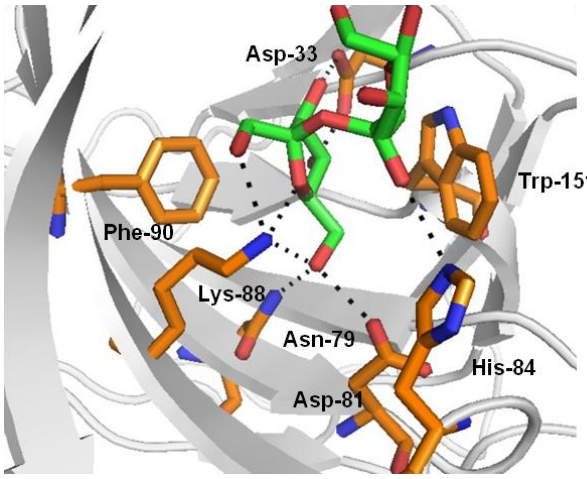


Figure 3.12 Structure of CBMX40 in complex with fructose and levanbiose

Panel A shows the surface representation of CBMX40 in complex with fructose. Panel B shows the surface representation of CBMX40 in complex with levanbiose. Created in Pymol.



3.13 Interactions to levanbiose in the binding site of CBMX40

Binding site of CBMX40 showing polar interactions with levanbiose. Asp-33, Asn-79, Asp81 and Lys-88 make polar contacts with the fructose in the binding pocket and Phe-90 makes hydrophopic contacts. His-84 makes polar contacts with the second fructose protruding out of the binding site. Trp-151 makes hydrophobic contacts with the terminal fructose and the adjacent sugar in the oligosaccharide. Created in PyMol.

The kinetic data presented in this chapter suggests that CBMX40 plays a pivotal role in the ability of the catalytic domain to hydrolyse levan, in the context of full length SacC. This potentiation is more marked with levan derived from *E. herbicola*, compared to the polysaccharide produced by *S. levanicum.* The potentiation in activity by CBMX40 reported here is considerably larger than other studies, which have studied the influence of CBMs on the activity of enzymes acting on insoluble substrates (Bolam et al., 1998). It is particularly surprising; therefore, that CBMX40 has such a dramatic influence of the hydrolysis of a soluble polysaccharide. The reduced potentiation of activity against levan from *S. levanicum* is also interesting. The actual activity of $SacC_M$ against levan from *S. levanicum* and *E. herbicola* are similar. $SacC_{CD}$, however, is higher against levan from *S. levanicum* than the corresponding *E. herbicola* polysaccharide. This result suggests that there is a difference in the structure of levan from the two microorganisms, with the *S.*

levanicum levan more accessible to enzymatic attack than the *E. herbicola* polymer. In fact the literature suggests that microbial levan is more highly branched than plant levan. Indeed, it has been reported that levan from *S. levanicum* is less branched and has a shorter chain length than typical microbial levans (Kojima et al., 1993). In contrast, levan from *E. herbicola* is characterised as having longer chains and around 1 in 30 fructose residues is branched through a β-2,1-linkage (Blake et al., 1982).

Kinetic studies using levan derived oligosaccharides as substrates show that $SacC_M$ is more active against the polysaccharide than the oligosaccharides, whereas the activity of $SacC_{CD}$ is similar against the two types of substrate. This may indicate that CBMX40 enhances affinity of the enzyme for the polysaccharide by binding, in harness with the active site of the fructosidase, to the branched regions of the levan. In essence the CBM is functioning as a distal sugar subsite, similar to the multiple subsites present in the substrate binding cleft of endo-acting glycanases, which greatly enhance affinity for the target polysaccharide (Charnock et al., 1998). The mature SacC protein displays substrate inhibition kinetics against levan from *E.herbicola*, compared to the typical Mechalias-Menten kinetics displayed with the other substrate tested. As levan from *E. herbicola* is highly branched as the concentration increases the aggregation also increases and thus making the CBMX40 less effective and the catalytic activity decreases. The *in trans* experiments showing the rescue of $SacC_{CD}$ activity by CBMX40 indicate the CBMX40 may disrupt the structure of levan increasing the access of the polysaccharide to enzyme attack, rather than just deliver the catalytic domain to the substrate. However, the AUC experiments show that, when not covalently linked, $SacC_{CD}$ and CBMX40 are

capable of forming a complex, which would support a delivery mechanism rather than CBMX40 acting directly on the levan.

From the data presented in this chapter the following hypothesis has been developed. The literature reports that levan is made up of relatively long chains with various branch points, which aggregate together to form a spherical bundle (Ingelman and Siegbahn, 1944). This aggregation could reduce the access of levan to enzyme attack. CBMX40 binds levan at the non-reducing end, bringing the catalytic domain into proximity with other chain ends, allowing it to attack the polysaccharide. The complex formed between $SacC_{CD}$/CBMX40, revealed by AUC, suggests that there are possible avidity effects driving the observed potentiation of enzyme activity. As the chain length of the levan and number of branch points increases the polysaccharide aggregates, reducing access to $SacC_{CD}$ and thus the activity of the enzyme.

3.5 Future work

Bacillus subtilis produces its own levan as part of its bacterial capsule. This levan has been characterised and found to be more highly branched and complex than the levan from *E. herbicola.* Utilising levan from *B.subtillis* as a substrate for $SacC_M$ and $SacC_{CD}$ may reveal a bigger difference in activity due to the increased complexity of the polysaccharide. Also, as SacC is thought to be a capsule processing enzyme, such experiments are more likely to reveal the biological role of CBMX40

Chapter 4: How Does CBMX40 Effect the Catalytic Activity of a Unrelated Exo- Fructosidase

4.1 Introduction

CBMX40 is a newly characterised non-catalytic carbohydrate binding domain (CBM) that binds the β-2,6 linked fructose polymer, levan. In the glycoside hydrolyse family (GH) 32 enzyme, SacC, CBMX40 potentiates the activity of the catalytic domain 120-fold. SacC is found in the soil bacterium *Bacillus subtillis*, where it is thought to be involved in capsule processing. Although there are a number of GH32 enzymes, the vast majority do not appear to have a CBMX40 module, or any other known CBM. BT 3082 is a GH32 exo-acting fructosidase, capable of hydrolysing both levan and inulin, expressed by the human gut bacterium *Bacteroides thetaiotaomicron.* BT 3082 is co-regulated with a polysaccharide utilization locus (PUL_{fru}) that targets dietary fructose-based polysaccharides. PUL_{fru} is controlled by a hybrid two component system in which fructose is the only activating ligand (Sonnenburg et al., 2010). However, the outer membrane endo-acting GH32 enzyme, BT 1760, is only active against the β2,6 linked levan with no detectable activity against β 2,1 linked inulin. This explains why *B. thetaiotaomicron* is capable of growing on levan and not inulin and, as BT 3082 is localized to the periplasm, the fructosidase will encounter levan oligosaccharides, but not reaction products of inulin degradation.

BT 3082 and SacC are expressed by very different bacteria and encounter fructan polysaccharides and/or oligosaccharides in differing biological contexts. While both enzymes are capable of hydrolysing levan and inulin polysaccharides, BT 3082 may only encounter levan oligosaccharides, and probably not long chain levans. SacC as

a capsule processing enzyme would, encounter long chain levan, but not inulin, which is generally thought to be a plant storage polymer. CBMX40 has a considerable effect on SacC activity; however, it is unclear what effect, if any, the CBM would have on BT 3082 an exo-acting fructosidase from a non related bacterium, which functions within a different biological context.

4.2 Objectives

The objective of this chapter is to explore whether CBMX40 has the capacity to potentiate the activity unrelated GH32 enzymes.

4.3 Results

4.3.1 Gene cloning, protein expression and purification

To explore the ability of CBMX40 to potentiate activity of a non-related GH32 enzyme, a number of constructs were designed and utilised. Wild type BT 3082 was cloned into pET22a by Dr Zheng (Sonnenburg et al., 2010). A fusion protein of CBMX40- BT 3082 was also designed. The primers are shown in Table 4.1. DNA encoding BT 3082 and CBMX40 were amplified by Polymerase chain reaction (PCR) and ligated into a pET22b variant, pET22jf, that contains a modified multi cloning site to give two separate multi cloning sites divided by a sequence encoding a 15 residue proline/theronine linker sequence; the map of pET22jf is shown in Figure 4.1. The modified vector was made by Dr James Flint. CBMX40 was cloned into the first multi cloning site using the KpnI and HindIII restriction sites. BT 3082 was then cloned in the second multi cloning site using the EcoRI and XhoI restriction sites. Cartoon representations are shown in Figure 4.2. All enzyme constructs were expressed in BL21(DE3) *Escherichia coli cells* by induction of mid-exponential cultures with 1mM IPTG and incubation at 16 °C overnight. The enzyme constructs contained a C-terminal His_6 Tag, encoded by the vector (thus the stop codon was omitted from the reverse primer), and thus Immobilised Metal Affinity Chromatography (IMAC) was used to purify the enzymes.

Primer Name	Sequence	Restiction site
Bt3082 Fwd	CAA CAC GAA TTC GGA GAA GTA TCT TTT AAA ATA ACC	EcoRI
Bt 3082 Rev	CAA CAC CTC GAG CCA AAT GGA TTC TAC GGA AAA GAC	Xho I
LevX40 Fwd	CAA CAC GGT ACC ATG ACG ACA CCT TTT ATG TCC AAT ATG	kpn I
LevX40 Rev	CAA CAC AAG CTT AGA CTC CTT CGT TAC ATT CTG AAA	Hind III

Table 4.1 Primers used in the cloning of CBMX40-BT 3082 fusion protein.

.

Figure 4.1 Modified pET22jf vector map

Protein fusion vector containing two multi-cloning sites separated by a PT-linker and a C-terminal His tag.

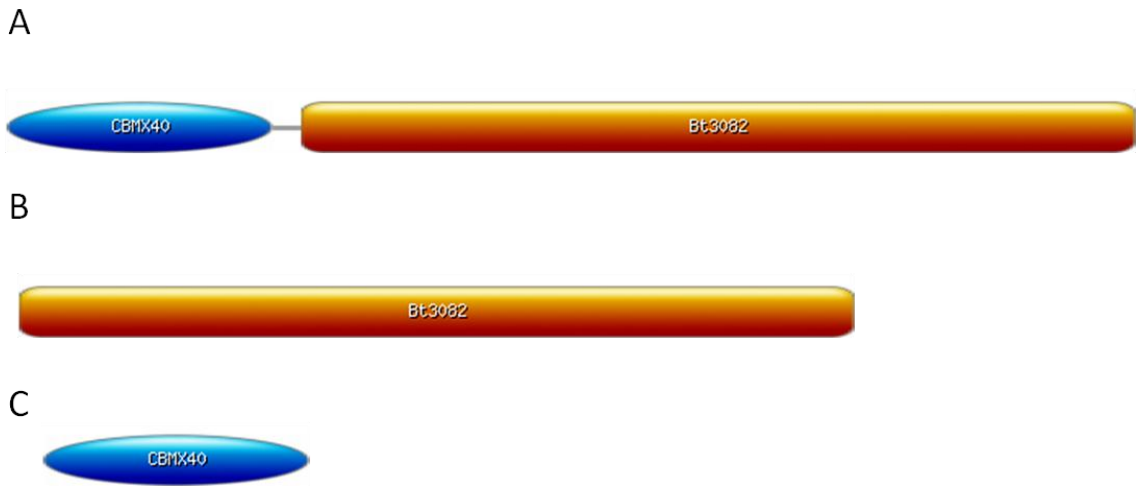


Figure 4.2 Cartoon representation of BT 3082 constructs

Panel A shows CBMX40-BT 3082 fusion protein. Panel B shows wild type BT 3082. Panel C shows CBMX40 image created with Prosite MyDomain.

4.3.2 Effect of CBMX40 on activity of BT 3082

To investigate if CBMX40 is capable of potentiating the activity of BT 3082 against levan, both the wild type BT 3082 and the CBMX40-Bt3082 fusion protein were expressed and purified. Their activity against levan from *Erwinia herbicola*, and inulin from chicory, was measured by monitoring fructose release using a fructose assay kit from Megazyme International. All experiments were carried out at 37 °C in sodium phosphate buffer pH 7.5. Non-linear regression plots are shown in Figure 4.3 and kinetic data are reported in Table 4.2.

The data showed that CBMX40-BT 3082 was considerably more active against *E. herbicola* levan than wild type BT 3082. Kinetic parameters K_m and K_{cat} were lower and higher respectively for CBMX40-BT 3082, compared to wild type BT 3082. The Kcat/*Km*, or catalytic efficiency, of CBMX40-BT 3082 was 52-fold higher than that of BT 3082. This potentiation in catalytic efficiency is less than that seen by CBMX40 in wild type SacC, but is still significant, showing that CBMX40 is able to potentiate the activity of a non-related exo-acting frucotsidase. Once again, as with $SacC_{M,}$ CBMX40-BT 3082 displayed substrate inhibition kinetics against *E. heribicola* levan, whereas BT 3082 displayed typical Michaelis-Menten kinetics, suggesting this inhibition is due to CBMX40 binding to levan. The activity of BT 3082 and CBMX40-BT 3082 against inulin showed no significant difference in activity, around 30-fold lower than the catalytic efficiency of CBMX40-BT 3082 against levan, consistent with the data from the previous chapter.

	CBMX40-BT 3082				BT 3082		
	K_m µM	K_{cat} s^{-1}	K_i	K_{cat}/K_m M^{-1} s^{-1}	K_m µM	K_{cat} s^{-1}	K_{cat}/K_m M^{-1} s^{-1}
Substrate							
Levan from *Erwinia herbicola*	26.3 ±11.4	49 ± 2.9	490 ± 266	1.8×10^6	205 ± 80.9	7 ± 1.1	3.4×10^4
Inulin	360 ±108	22 ±3.4		6.2×10^4	481 ± 129	26 ± 4	5.4×10^4

Table 4.2 Kinetic values for CBMX40-BT 3082 and BT 3082

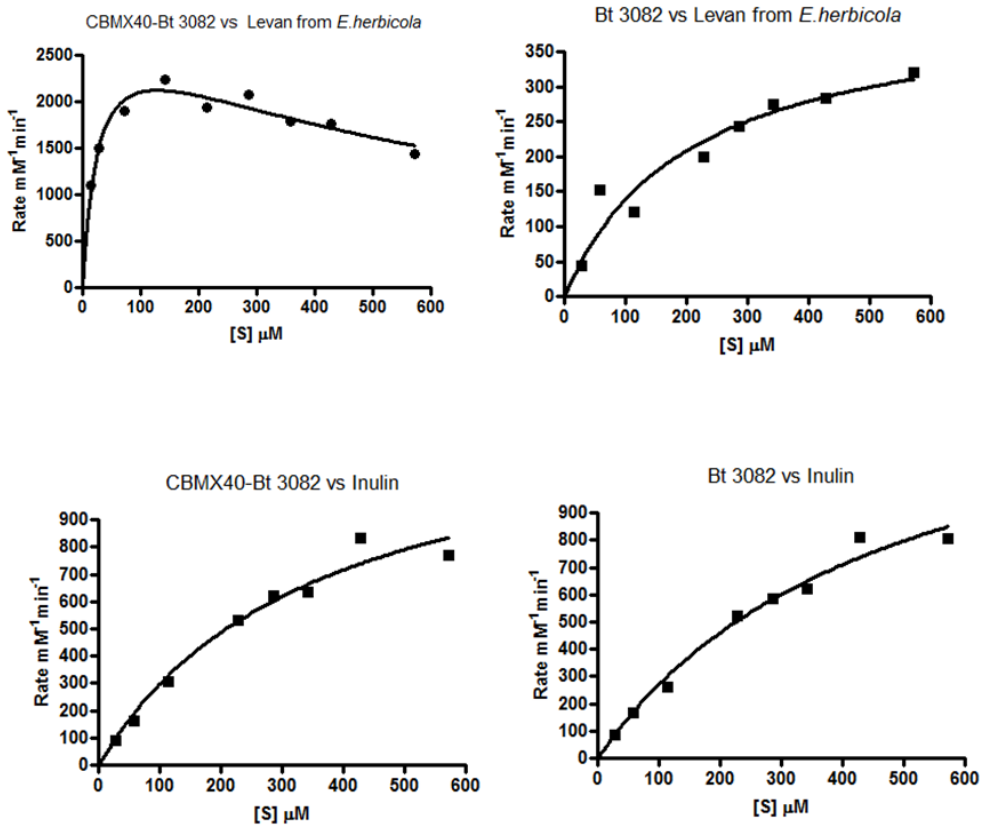


Figure 4.3 Example of kinetic analysis of CBMX40-BT 3082 and BT 3082 against levan and inulin

Initial rates derived from fructose assay kit from Megazyme International were plotted in GraphPad Prism using Michaelis-Menten or substrate inhibition equation.

4.3.3 Catalytic activity of BT 3082 in the presence of CBMX40

CBMX40 has the ability to enhance the activity of an unrelated exo-fructosidase BT 3082 when physically attached. The previous chapter showed CBMX40 could rescue the activity of $SacC_{CD}$ when not covalently attached to the enzyme. To investigate if CBMX40 is able to rescue the activity of BT 3082, both proteins were incubated with levan from *E. herbicola* and *Serratia levanicum*. The release of fructose was measured by HPLC analysis. Control experiments with BT 3082 in the absence of CBMX40 was also measured concurrently. The experiments were analysed by HPLC as the addition of high concentrations of CBMX40 resulted in a interference when measuring absorption at 340 nM using the megazyme fructose assay kit. Plotted initial rates are shown in Figure 4.4.

When BT 3082 was incubated with levan from *E.herbicola* in the presence of CBMX40, at a ratio of 200 nM to 100 µM, respectively, the rate of substrate hydrolysis increased compared to BT 3082 in the absence of CBMX40. The enhancement was estimated to be 145-fold, although this value must be viewed with some caution as the activity of BT 3082 against *E.herbicola* levan was very low and thus difficult to measure accurately. The rate of fructose release by BT 3082 in the absence of CBMX40 against levan from *S.levanicum* was much higher in comparison to the rate of release from the levan from *E.herbicola.* The enhancement in rate against levan from *S.levanicum* in the presence of CBMX40 was much smaller than the enhancement against *E.herbicola.* The actual enhancement in rate measured for BT 3082 in the presence of CBMX40 against levan from *S.levanicum* was only 1.2 fold higher compared with the rate in the absence of CBMX40. CBMX40 had little influence in potentiating the activity of BT 3082 against levan from *S.levanicum* as BT 3082 is able to effectively degrade the levan in the absence of

CBMX40. The difference in activity of BT 3082 against the two levans is due to the differing structures from *S.levanicum* and *E.herbicola.* The levan produced by *S.levanicum* has found to be short chained with very few branch points, making it highly accessible to the enzyme attack (Kojima et al., 1993). Whereas, the levan produced by *E.herbicola* is made up of longer chains that have branch points occurring at about 1 in 30 fructose residues (Blake et al., 1982). These results suggest as with the results in Chapter 3 that CBMX40 can potentiate the activity of enzyme against complex levan but not as effectively against less complex levans.

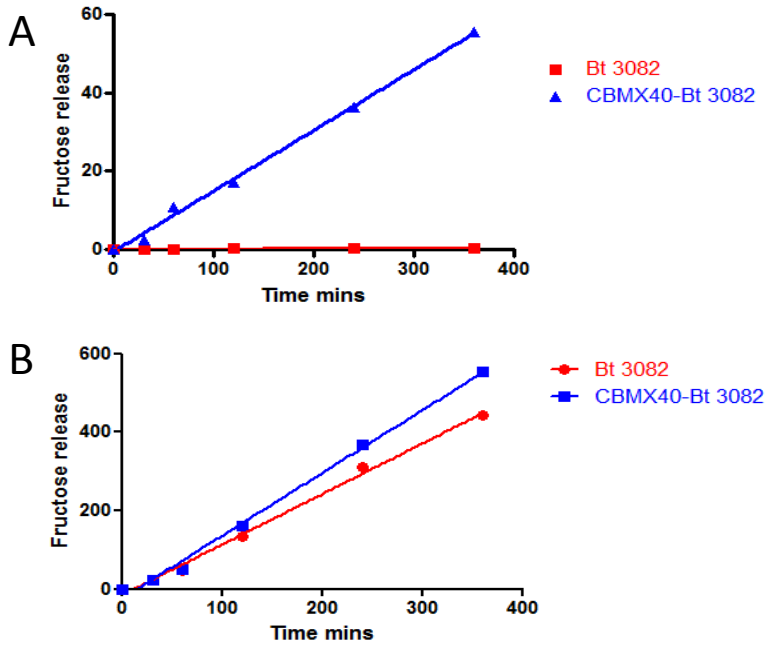


Figure 4.4 Initial rates for BT 3082 in trans with CBMX40

Panel A shows initial rates for 500 nM BT 3082 and 200 µM CBMX40 against 715 µM levan from *E.herbicola*

Panel B shows initial rates for 500 nM BT 3082 and 200 µM CBMX40 against 715 µM levan from *S.levanicum*

4.3.4 Can BT 3082 form a complex with CBMX40

CBMX40 potentiates the activity of BT 3082, both when covalently attached, and when the proteins are added to the same reaction. To determine if this potentiation is a result of BT 3082 and CBMX40 associating into a complex, analytical ultracentrifugation was utilised. All AUC experiments were carried out and analysed by Dr Alexandra Solovyova as described in Section 2.3.6. Figure 4.5 and Figure 4.6 shows the results from the AUC experiments.

BT 3082 and CBMX40 were mixed in a 1:1 ratio and subjected to sedimentation velocity AUC. Sedimentation velocity reactions give information on size and shape of the proteins. As with AUC experiments in the previous chapter S-values were calculated from atomic coordinates and molecular weight of the proteins. In this experiment there were only two peaks present, the first at an S-value of 2.009 which corresponds to CBMX40. The second peak was found at an experimental S-value of 6.4 and corresponds to a BT 3082 dimer which has a calculated S-value of 6.68. This would mean that there is no complex formed between BT 3082 and CBMX40, suggesting that the potentiation seen when CBMX40 and BT 3082 are not attached is not by CBMX40 acting on BT 3082 but by acting on the substrate, levan.

However, when BT 3082 and CBMX40 are mixed in a 1:1000 ratio and subjected to sedimentation velocity AUC three peaks are attained. The first again at an S-value of around 2 is identified as CBMX40. The second peak is found at an S-value of 4.25 and is a monomer (1:1) of BT 3082:CBMX40 in complex. The third peak at an S-value 6.5 relates to the dimer configuration (2:2) of BT 3082:CBMX40 complex. At a high ratio of CBMX40 to BT 3082 a complex is formed between the two proteins. Interestingly, when CBMX40 is at a large excess and at lower concentrations of BT

3082 not only is a dimer complex formed but a monomer complex also. When BT 3082 is subjected to AUC in the absence of CBMX40 only one peak is revealed at a S-value of 6.65 indicative of a BT 3082 dimer, displayed in Figure 4.6. The concentration of BT 3082 subjected to AUC were comparative to those used when in large CBMX40 excess. Therefore, the presence of BT 3082:CBMX40 monomer complex is due to CBMX40 being in large excess as there is no monomer BT 3082 peak displayed in the absence of CBMX40.

In addition to the AUC experiments, isothermal titration calirometry was also utilised to determine if there is a measurable interation between BT 3082 and CBMX40. CBMX40 was injected into BT 3082 which was present in the cell. Figure 4.7 shows the resultant ITC trace from the titration. ITC measures the thermodymanics of binding interactions by measuring the increase or decrease in tempreature against a reference cell. The absence of any uniform peaks show that there was no measurable association between CBMX40 and BT 3082.

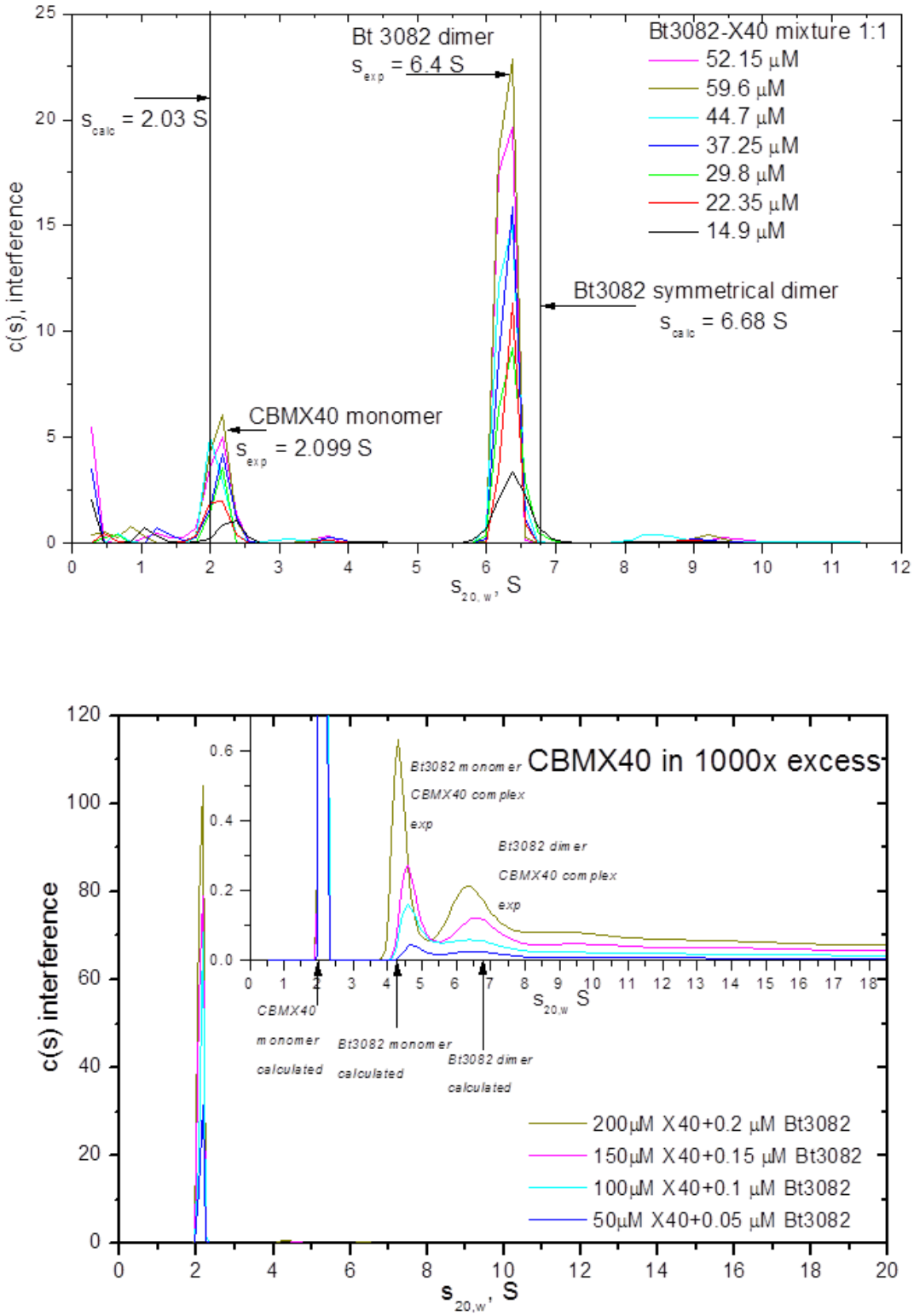


Figure 4.5 AUC trace for CBMX40 and BT 3082

All experiments were centrifuged at 20 °C in 50 mM Na-HEPES, pH 7.5, containing 150 mM NaCl.

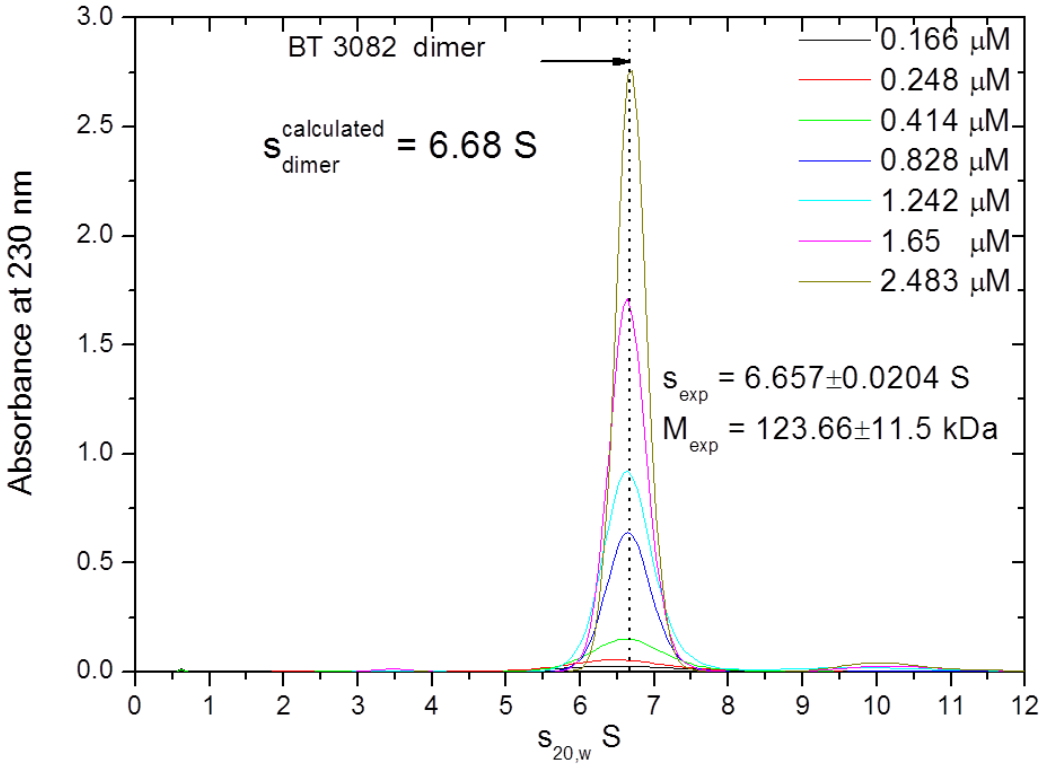


Figure 4.6 AUC of BT 3082 trace in absence of CBMX40

All experiments were centrifuged at 20 °C in 50 mM Na-HEPES, pH 7.5, containing 150 mM NaCl.

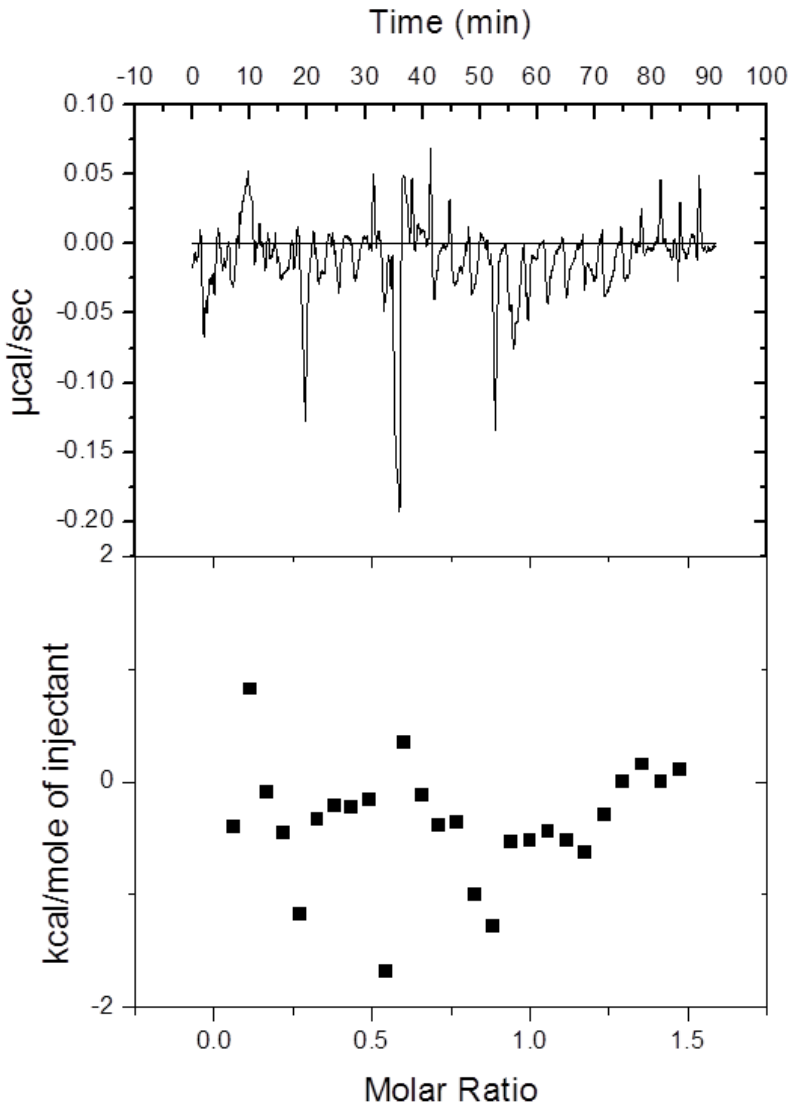


Figure 4.7 ITC trace showing CBMX40 titrated against BT 3082

77 µM Bt 3082 in cell and 580 µM CBMX40 present in syringe carried out at 25 °C

4.4 Discussion

The previous chapter detailed the binding of levan to CBMX40, and the positive effect the protein module had on the catalytic domain of its native enzyme, SacC. The data in this chapter show that when BT 3082 and CBMX40 are fused together, the CBM once again potentiates the exo-fructosidase activity of the enzyme against levan from *E.herbicola*. In its native protein, CBMX40 potentiates the activity of the catalytic domain against levan 120-fold, while, when fused to BT 3082, it potentiates the activity only 52-fold. This likely reflects the fact that BT 3082 is not the parent enzyme of CBMX40, and thus the position of the two protein modules is not optimized for achieving maximum synergy. Indeed in the artificial constructs CBMX40 and Bt3082 are separated by a flexible 15 residue linker, while in SacC there is no obvious linker suggesting that the 3D structure of the enzyme ensures that the position of the enzyme and CBM are optimized to achieve maximum synergy. Not only does CBMX40 enhance the activity of BT 3082 against levan but, by doing so, BT 3082 now has a preference for the β 2,6 linked polysaccharide, levan, rather than the inulin, a β 2,1 linked fructan. This conclusion is drawn from the kinetic data, which reveals very little difference in the kinetic parameters of wild type BT 3082 against levan and inulin. Whereas, the kinetic values for CBMX40-BT 3082 differ greatly between levan and inulin the Kcat/Km or catalytic efficiency of CBMX40-Bt 3082 for levan is 29 times that of CBMX40-BT 3082 against inulin.

Once again the In trans experiments show that CBMX40 can rescue the activity of BT 3082 against levan from *E. herbicola* when the two proteins are not physically attached. The enhancement in rate is 145 fold compared to BT 3082 in the absence of CBMX40. It should be noted that BT 3082 it less active against complex long chain, branched levans such as *E. herbicola*, compared to less highly branched

forms of the polysaccharide produced by *S. levanicum*. This is illustrated by the increased activity against the shorter chained, less branched levan from *S.levanicum.* Therefore the enhancement in rate is less notable when BT 3082 is incubated with this levan in the presence of CBMX40.

The previous chapter showed that the rate enhancement of $SacC_{CD}$ in the presence of CBMX40 was as a result of a complex formed as detailed by analytical ultracentrifugation, results are summarised in Table 4.3. When CBMX40 and BT 3082 were subjected to analytical ultracentrifugation at a 1:1 ratio, no complex was exhibited between the two proteins. However at a 1:1000 ratio of BT 3082 to CBMX40 both a monomer and a dimer complex are exhibited. In the absence of CBMX40, BT 3082 is present only as a dimer. This would suggest that the BT 3082 monomer is observed due to the interaction of CBMX40 with BT 3082. The $SacC_{CD}$:CBMX40 complex reported in the previous chapter was also a monomer complex and $SacC_{CD}$ appeared to only be present in solution as a monomer and not a dimer. This is significant as $SacC_{CD}$ displays a higher enzyme activity than BT 3082 against levan from *E.herbicola* this could be attributed to BT 3082 being a dimer in solution while $SacC_{CD}$ is a monomer. ITC experiments could not find any measurable interaction between BT 3082 and CBMX40 this suggests that any interaction formed is very weak as it must be out of the limits of the ITC experiment.

The data presented in this chapter provides information on possible mechanisms for potentiation of catalytic activity by CBMX40. Levan is a complex polysaccharide which forms long chains with various branch points. These chains aggregate to form spherical bundles that make them inaccessible to enzyme attack (Ingelman and Siegbahn, 1944). However, it is possible that, within this structure an equilibrium between a closed state, where there are very few chain ends available for attack,

and the polysaccharide chains are tightly bound, and a slightly more open structure where the chain ends are more accessible to enzyme attack. CBMX40 can bind the non reducing chain ends of the levan chain when in the open state, forcing the equilibrium in favour of the open state, allowing BT 3082 to access the chain ends for productive hydrolysis. In the absence of CBMX40, levan is less accessible to BT 3082 attack and therefore the rate of hydrolysis is slower. However as CBMX40 is capable of interacting with BT 3082 to form a complex it seems unlikely that CBMX40 would act directly on the levan polysaccharide. It seems more feasible that the complex formed delivers BT 3082 to the chain ends of the levan polysaccharide through CBMX40 binding to the non-reducing end of another levan chain. In the highly branched levan from *E. herbicola* the ability of CBMX40 to increase BT 3082 activity is increased due to the effects of the chain ends being in close proximity due to the amount of branches and chain ends. It still may be possible that while CBMX40 does not act directly on the levan polysaccharide it may increase BT 3082's access to the polysaccharide through decreased aggregation of the polysaccharide chains.

As with full length SacC, CBMX40-BT 3082 also displays substrate inhibition kinetics when incubated with levan from *E. herbicola.* This is because the levan is highly branched and so is highly aggregated. As the concentration of substrate increases the aggregation also increases and so is less accessible to attack. Despite the aggregation of the substrate there are still sufficient available ends for BT 3082 to hydrolyse as the catalytic activity of BT 3082 is so low. Therefore, BT 3082 does not display such inhibition.

	Enhancement in rate against *E. herbicola* levan	Ratio of enzyme:CBMX40 intrans experiment	Ratio of enzyme:CBMX40 in AUC experiment
$SacC_{CD}$	0		
$SacC_{M}$	120 fold*		
Intrans $SacC_{CD}$:CBMX40	25 fold*	1:2000	1:1
BT 3082	0		
CBMX40-BT 3082	52 fold**		
Intrans BT 3082 :CBMX40	145 fold** ≠	1:500	1:1000

Table 4.3 Summary of the enhancement of rate by CBMX40 on $SacC_{CD}$ and BT 3082 and the ratios of enzyme to CBMX40 in experiments
Shows the enhancement in rate for each construct against *E. herbicola* levan. *compared to rate of $SacC_{CD}$. **compared to rate of BT 3082. ≠ rate is exgerated as BT 3082 activity is very low. The ratio of enzyme:CBMX40 is shown for both the intrans assays and AUC experiments

4.5 Future Work

This chapter detailed the effect of CBMX40 on a non related GH32 exo-acting fructosidase. However, what effect would CBMX40 have on a GH32 endo acting levanase? By fusing the CBMX40 to, for example, BT 1760, the outer membrane endolevanase encoded by PUL_{fru}, would it have a positive or negative effect on the activity of the enzyme. Also, testing BT 1760 *in trans* in the presence of CBMX40 would provide insight into weather the proposed substrate disruptive action of CBMX40 can affect the endo-activity of BT 1760.Therefore exploring further the mechanism of action of CBMX40.

Chapter 5: Yeast Mannan Utilisation by *Bacteroides thetaiotaomicron*

5.1 Introduction

Bacteroides thetaiotaomicron is a beneficial bacterium found within the microbiota of the human gut. The bacterium forages on dietary and host polysaccharides found within the environment of the gut. *B. thetaiotaomicron* produces numerous glycanases that deconstruct polysaccharides, which are not attacked by enzymes produced by the human host. The genes encoding these *B. thetaiotaomicron* enzymes are organised into clusters termed Polysaccharide Utilization Loci (PULs). PULs contain all the machinery required for sensing, transport and deconstruction of specific polysaccharides encountered within the environment of the human gut (Bjursell et al., 2006). The genome of *B. thetaiotaomicron* genome encodes 10 glycoside hydrolase family (GH) 76 enzymes and 24 GH92 enzymes, annotated on the CAZy database, described in Section 1.2. The *B.thetaiotaomicron* GH92 enzymes have recently been extensively characterised as α-mannosidases able to hydrolyse a number of α-mannosidic linkages present in various glycans (Zhu et al., 2010) . GH76 enzymes have not been extensively characterised; only a single GH76 member, from *Bacillus circulans,* has been investigated, which was shown to by an endo-α1,6-mannosidase (Nakajima et al., 1976). The genes encoding these enzymes in *B. thetaiotaomicron* are located in PULs, and transcriptomic data suggest that some of these PULs are upregulated by α-mannans and host glycans (Martens et al., 2008, Bjursell et al., 2006). In this study there are two GH76-containing PULs of particular interest PUL 36 (BT 2615-BT 2623) and PUL 68 (BT 3773-BT 3792) Figure5.2. Both of these PULs are regulated by a hybrid two

component system and both contain SusC and SusD homologs. SusC and SusD were first characterised in the Starch utilisation locus. SusD is annotated as a outer membrane starch binding protein and SusC a TonB-dependent transporter which transports starch oligosaccharides into the periplasm (Koropatkin et al., 2008, Reeves et al., 1996, Reeves et al., 1997, Shipman et al., 2000). The HTCS is located across the inner membrane between the periplasm and cytoplasm and acts as a sensor regulator. In essence both PULs contain all the required elements to sense, acquire and utilize yeast mannan. Both of these PULs are largely un-characterised; although the GH92 enzymes have been investigated (Zhu et al., 2010), their activities against yeast mannan degradation have not been extensively analysed. The GH76 enzymes are yet to be characterised but could provide useful insight into yeast metabolism by *B.thetaiotaomicron.*

Yeast mannan is a complex polysaccharide produced by *Saccharomyces cerevisiae.* It is an α mannan that comprises an α1,6 mannose linked backbone with α1,2 mannose side chains that are capped with an α1,3 linked mannose residues. It can also have a varying degree of phosphorylation. *S. cerevisiae* employs a myriad of enzymes to construct the mannan polysaccharide. For this reason mutant *S. cerevisiae* strains lacking vital enzymes that catalyse the synthesis of the polysaccharide produce various different mannan structures. The wild type and the mutant mannan structures are shown in Figure 5.1

5.2 Objectives

The objective of this chapter is to investigate the enzyme activities against yeast mannan found within the two α-mannan activated *B. thetaiotaomicron* loci, PUL 36 and 68.

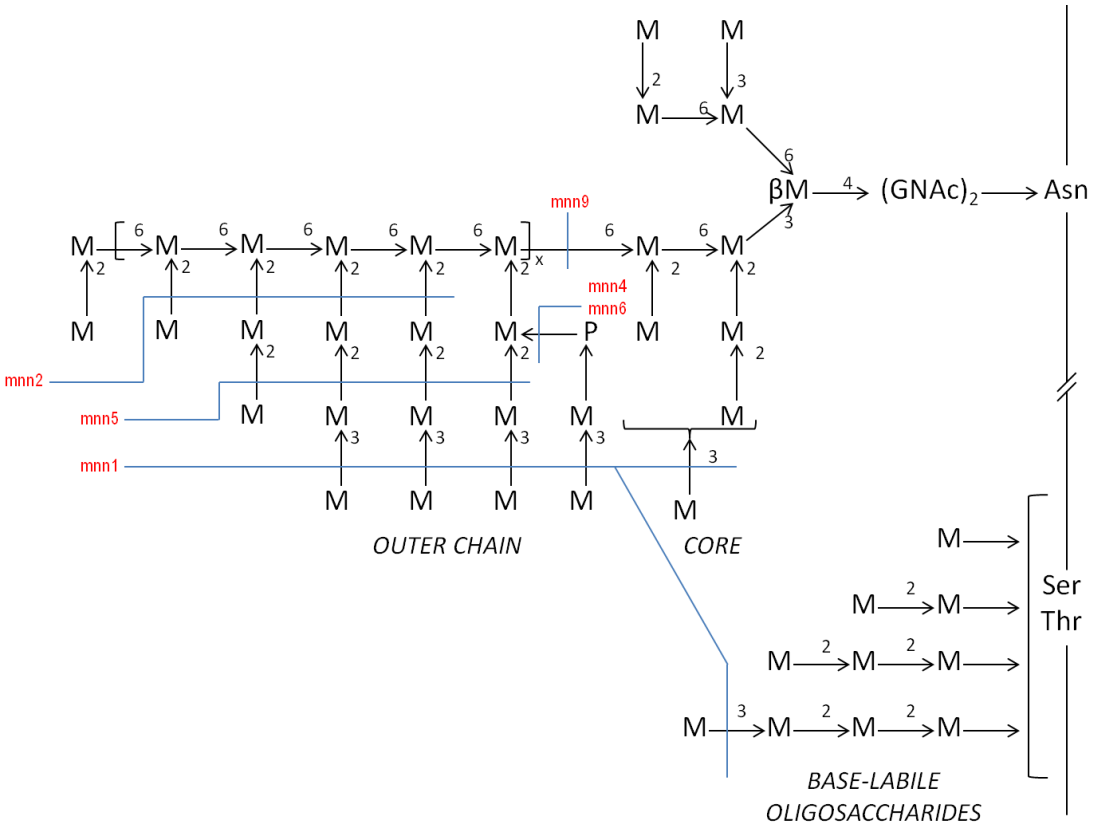
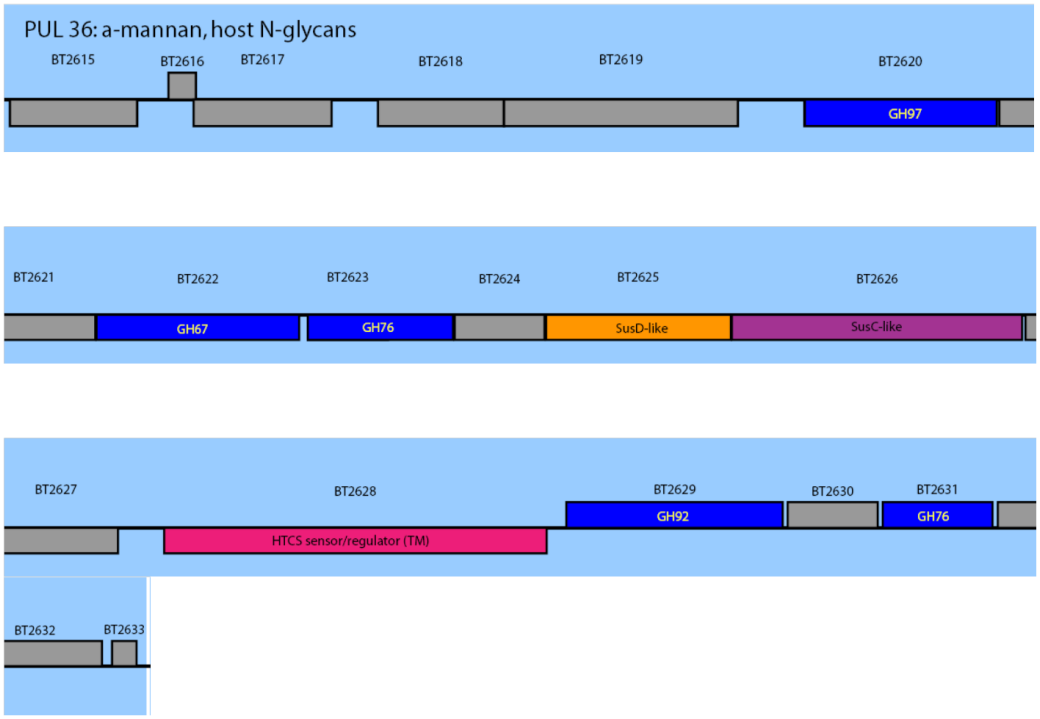


Figure 5.1 Yeast Mannan structure

Blue lines are an indication of the changes by mnn mutations which are labelled in red.

A



B

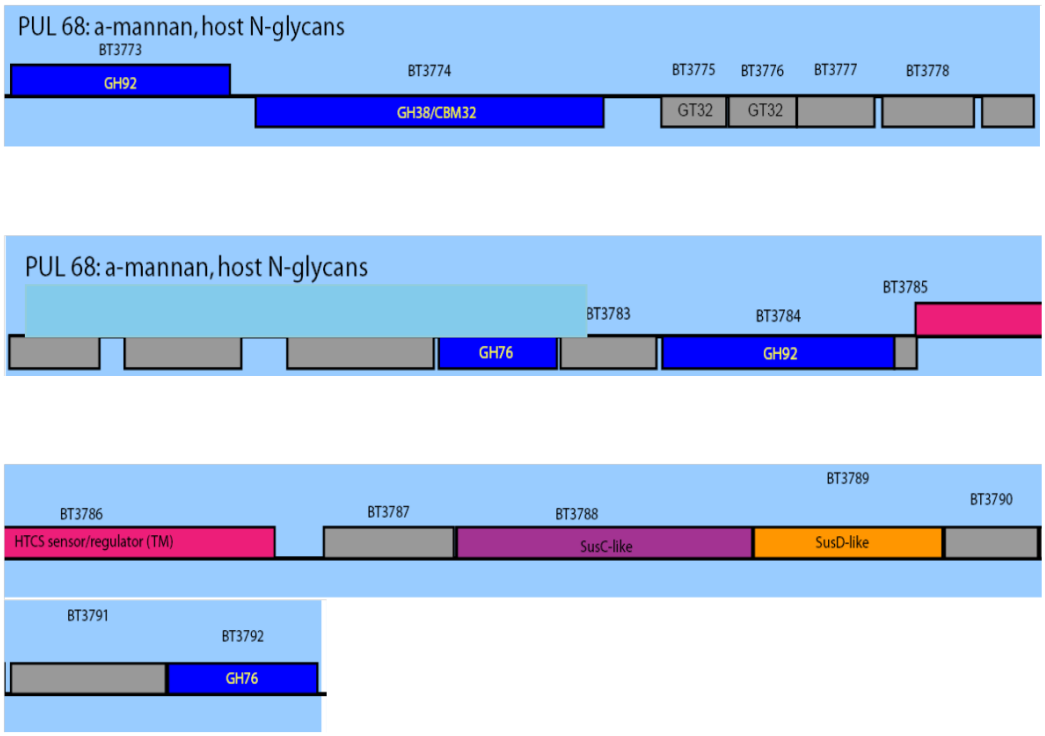


Figure 5.2 Yeast mannan degrading Polysaccharide Utilisation Loci (PUL)

A- PUL 36 map B- PUL 68 map. Grey boxes are opening reading frames. Blue boxes are CAZy enzymes. Pink boxes are HTCS. Purple boxes are SusC like proteins and orange boxes are SusD like proteins.

5.3 Results

5.3.1 Gene Cloning, Protein Expression and Purification

Of the 10 genes encoding the *B.thetaiotaomicron* GH76 enzymes 8 were amplified by PCR using the bacterium's genomic DNA as template, and cloned successfully into either of the *E. coli* expression vectors pET28a (N-terminal His_6 tag) or pET32b (N-terminal Trx Tag and His_6 Tag). An example of an agarose gel of PCR fragments is shown in Figure 5.2. Primers, which were designed to the 3` and 5` regions of the gene, are shown in Table 5.1. PCR fragments were digested with the appropriate restriction enzymes and ligated into the chosen vector. Plasmids from transformants were tested for the presence of *B.thetaiotaomicron* inserts by digestion with appropriate restriction enzymes and visualised by agarose gel electrophoreses. The expression of the proteins encoded by the recombinant plasmids in *Escherichia coli* BL21 cells were induced by the addition 1 mM IPTG at 16 °C to mid-exponential phase cultures, and incubation overnight with aeration. Proteins were purified from cells by immobilised metal ion affinity chromatography (IMAC) and purification was visualised by SDS PAGE gels. An example of an SDS PAGE showing expression and purification is shown in Figure 5.3

Other proteins utilised in this chapter were cloned into appropriate *E. coli* expression vectors by various different people. BT 2631 was cloned by Dr Joanna Norman. Dr Yanping Zhu cloned all of the GH92 enzymes used. BT 3783 was cloned by Carl Moreland. All of the above proteins were expressed and purified as described above for the GH76 enzymes.

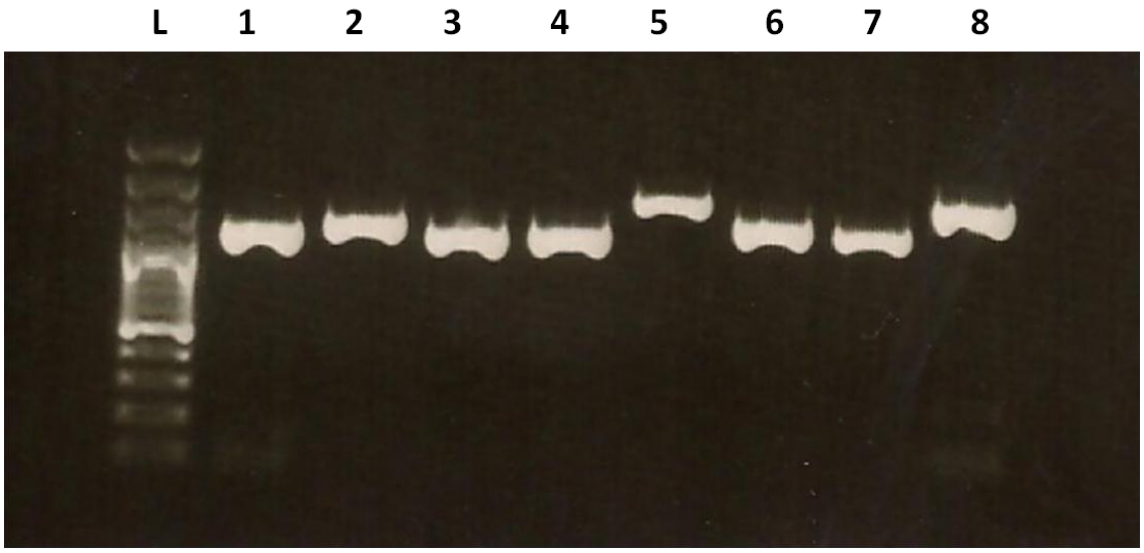


Figure 5.3 Agarose gel of GH76 PCR fragments

L- ladder, 1- BT 1883. 2-BT 3521, 3- BT 3782, 4- BT 3301, 5- BT 2623, 6- BT 3524, 7- BT 2949, 8- BT 3792

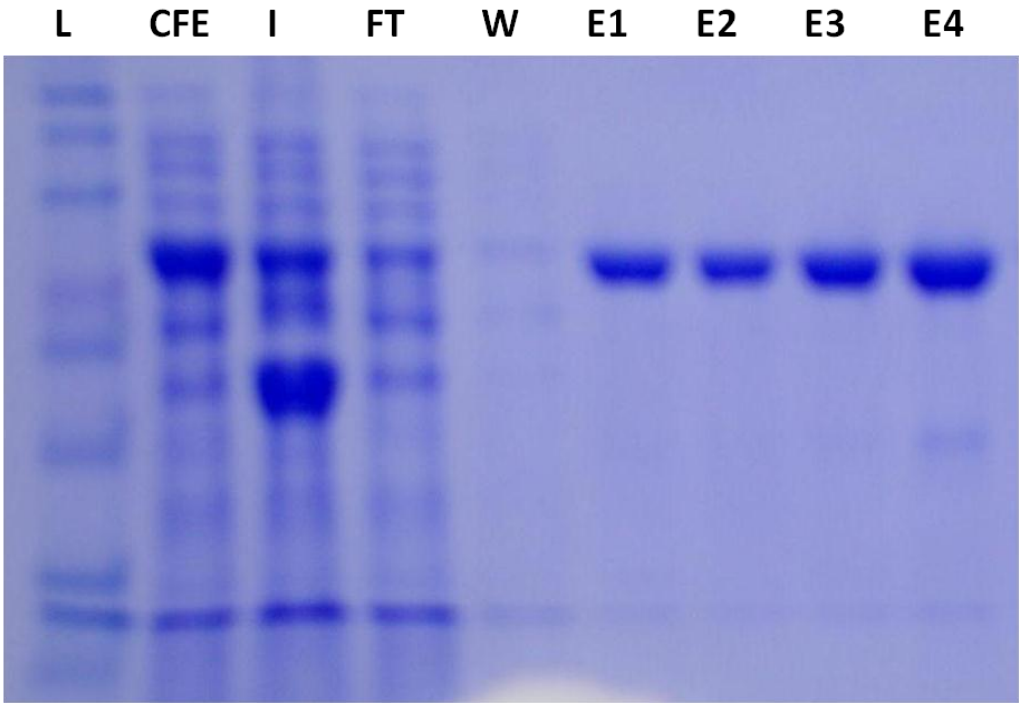


Figure 5.4 Example SDS PAGE of BT 3792 expression and purification

IMAC purification of BT 3792, purified in E1 ,E2, E3 and E4 fractions. L= Ladder CFE= cell free extract, I= Insoluble fraction, FT= Flow through from column, W= buffer wash, E1, E2 = 10mM imadizole elution, E3 and E4= 100mM imadizole elution.

Primer Name	Sequence	Restriction site	Vector
BT2949FWD	CAA CAC **GCT AGC** GGC TGT GAT GCC ACT GTA CAG GAT ATC	NheI	pET 28a
BT2949REV	CAA CAC **CTC GAG** TTA ATC GTT GAG TAT TCT TTT GTA ATT	XhoI	pET 28a
BT2623FWD	CAA CGC **GCT AGC** GAT GAA AAG TAT GAT ATA TTA GAA AGA TAC	BamHI	pEt 32b
BT2623REV	CAA CAC **CTC GAG** CTA CTC CTG AAA ACA TCC AAG TTG TTC CAA	XhoI	pET 32b
BT3301FWD	CAA CAC **GCT AGC** TCA ACA AAG CAG GCA GAC TCA GAT TCA	BamHI	pET 32b
BT3301REV	CAA CAC **CTC GAG** TTA TTG TAT AGT TGC TAG TCG GCT ATA	XhoI	pET 32b
BT3501 FWD	CAA CAC **GCT AGC** TGC GGA GAA GGA GGA AAC GAT AAT TAC	NheI	pET 28a
BT 3501 REV	CAA CAC **AAG CTT** TTA TTC TAA CCT AGC TAT CCC CTC CAT	Hind III	pET 28a
BT3521FWD	CAA CAC **GCT AGC** GAT GAT AAA GAA GCT AGA CCT TAT	NheI	pET 28a
BT3521REV	CAA CAC **CTC GAG** TTA TTT ATC GGT TGG AAA CAA TAA	XhoI	pET 28a
BT3524FWD	CAA CAC **GCT AGC** GCT GAT AGC TCA GAA AAA ACG AAC	BamHI	pET 32b
BT3524REV	CAA CAC **CTC GAG** TTA TTT TTC TGC CGG AAA CAG AAG	XhoI	pEt 32b
BT3782 FWD	CAA CAC **GCT AGC** AAA ATC ACT TAT CTG GCA GGC GGA	Nhe I	pET 28a
BT3782REV	CAA CAC **CTC GAG** TTA TTT CTC GAT TTC AGC GAA CAG	XhoI	pET 28a
BT3792 FWD	CAA CAC **GCT AGC** GCT ATC AGC AAA GCC GGC AAG	NheI	pET 28a
BT3792REV	CAA CAC **CTC GAG** CTA TCG GTT TTC ACG TTC CAA TAC	XhoI	pET 28a

Table 5.1 Primers for GH76 cloning
Restriction sites are shown in bold

5.3.2 Characterisation of Glycoside hyrdrolase 76 enzymes

Of the eight soluble GH76 enzymes produced, only four were shown to be catalytically active. While the enzymes were not active on mannan produced from wild type *S. cerevisiae* four were active on mannan produced from the mnn2 *S.cerevisiae* mutant. This mutant lacks the α 1,3 and α 1,2 branches in the outer chain of the polysaccharide leaving only the α1,6 linked mannose backbone, as shown in Figure 5.1. Thin layer Chromatography (TLC) shows that BT 2623, BT 2631, BT 3782 and BT 3792 produce a range of oligosaccharide products when incubated with the mnn2 mutant mannan, Figure 5.4. This is indicative of an endo acting mode of action, in which the enzyme cleaves the polysaccharide within the chain rather from the end of the chain, thus producing a number of different sized products. Inspection of the reaction by TLC over a 24 h reaction shows that while all of the enzymes are active on the mnn2 mutant mannan, BT 2631 and BT 3782 have a similar mode of action in which they produce smaller oligosaccharides. BT 2631 produces mainly di- and tri-saccharide sugars with a small amount of mannose and larger oligosaccharides. The reaction appears to be over after 60 min. While BT 3782 also generates mainly small oligosaccharides from the mnn2 mutant yeast mannan, the reaction product profile shows subtle differences to BT 2631. BT 3782 generates more of the monosaccharide unit along with the di- and tri-saccharide units, and the reaction is over after 30 min. However, BT 2623 and BT 3792 are also similar producing larger oligosaccharides. BT 2623 generates mostly tri-sacharides and bigger oligosaccharides which are not very well defined. The reaction seems to be over after about 2 h, suggesting the enzyme is slow acting. BT 3792 generates more defined oligosaccharides with a small quantity of di-sacchrairde product but the tri-, tetra- and pentose-oligosaccharide are the major products generated. The

reaction is finished after 1 h, at which point well defined oligosaccharide products were evident.

Kinetic data for each enzyme were determined by incubating the enzymes with the Mnn2 mutant mannan at 37 °C in 50 mM MOPS buffer pH 7.5. The rate of reducing sugar released was determined using the 3, 5-dinitrosalicyclic acid assay. In the assay the number of free reducing ends is proportional to absorbance at 575 nm due to the reduction of 3,5- dinitrosalicyclic acid to 3-amino,5-nitrosalicylic acid by the free reducing ends of the oligosaccharides created by hydrolysis of the polysaccharide. The assay is explained in more detail in Section 2.3.1.3. Initial rates were plotted in GraphPad Prism 5.0 and kinetic parameters calculated by non-linear regression. Data are shown in Figure 5.5 and Table 5.2.

The K_m and k_{cat} for BT 2631 against the mnn2 mannan could not be calculated as the K_m was beyond the limits of this assay. For this reason the K_{cat}/K_m of BT 2631 against the Mnn2 mannan was calculated. This was achieved by measuring the initial rate of the reaction against various substrate concentrations, the initial rates were plotted and k_{cat}/K_m calculated by linear regression. The K_{cat}/k_m for BT 2631 is relatively large against Mnn2 mannan compared to the other GH76 enzymes. This indicates that BT 2631 is efficient at degrading Mnn2 mannan. BT 3782 has a similar mode of action to BT 2631, however the K_m and k_{cat} were able to be calculated . BT 3782 displays a high affinity for the Mnn2 mannan substrate as indicated by the low K_m measured. The high k_{cat} indicates a good turnover of substrate by BT 3782. Thus, the ability of BT 3782 to degrade Mnn2 mannan is reflected in the high k_{cat}/K_m . The k_{cat}/K_m for these enzymes indicates that BT 3782 is more effective at deconstructing Mnn2 mutant yeast mannan than BT 2631. BT 2631 has an immeasurably high K_m's this may be attributed to the aggregation of

mnn2 mutant yeast mannan in solution causing accessibility problems for the enzyme. In addition BT 2631 and BT 3782 both generate small reaction products from the polysaccharide, and are predicted to be periplasmic, Mnn2 mutant yeast mannan may not be the preferred substrate for the two enzymes.

BT 2623 and BT 3792 hydrolyse the Mnn2 mutant yeast mannan generating long chain oligosaccharides. The data for BT 2623 show that the enzyme is fairly good at hydrolysing Mnn2 mutant yeast mannan. The low K_m of the enzyme is indicative of a high affinity for the substrate. The turnover number of the enzyme, however, is quite low, which has a negative impact on the catalytic efficiency of the enzyme. BT 3792 also displays a low K_m suggesting a high affinity for the Mnn2 mutant yeast mannan. However, the K_{cat} for BT 3792 is relatively high signifying a faster rate of hydrolysis. Thus, the catalytic efficiency of BT 3792 against Mnn2 mutant yeast mannan is far superior to that of BT 2623.

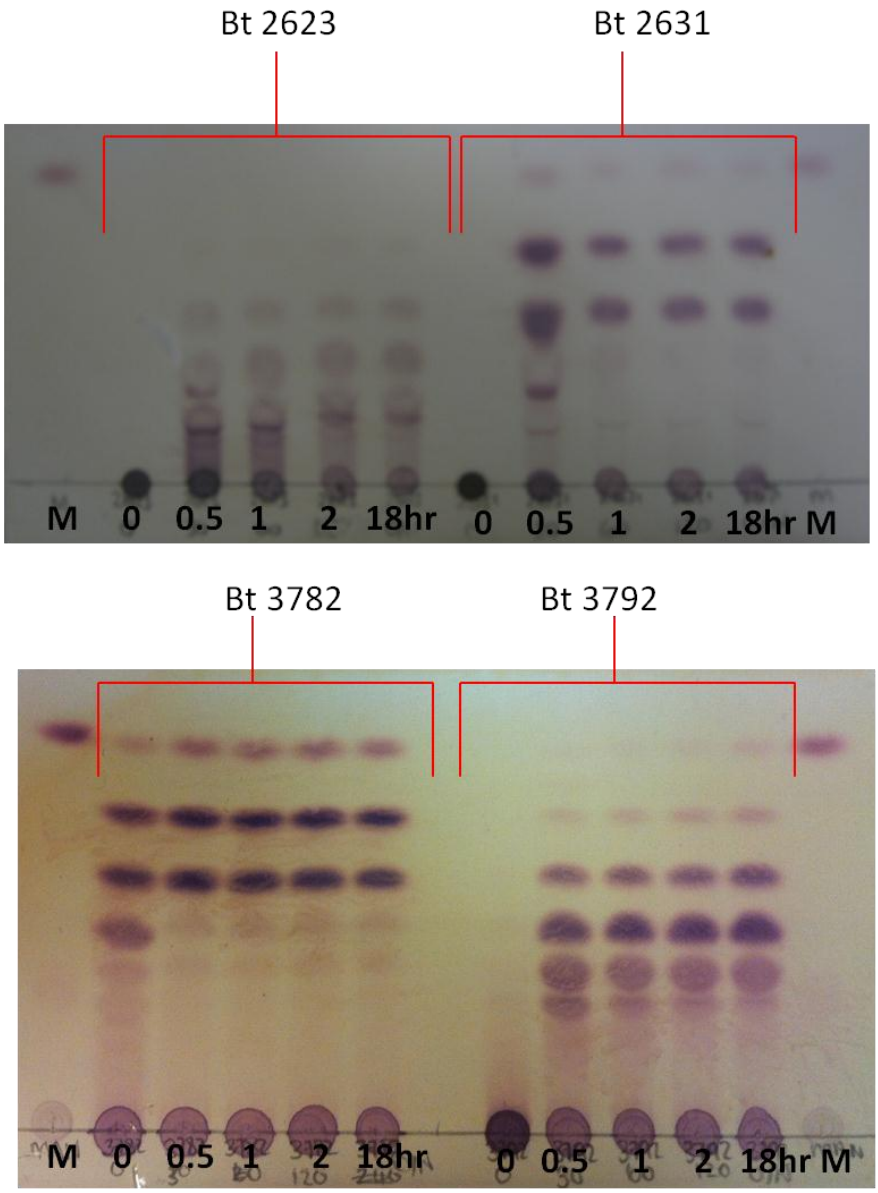


Figure 5.5 TLC of GH76 enzymes against Mnn2 18hr timeline

Each enzyme (1 µM) was incubated with Mnn2 mannan (1 mM) at 37 °C in 50 mM MOPS pH 7.5. Aliquots were taken at timed intervals and the reaction stopped by heat inactivation and subjected to TLC analysis.

Note that the activity seen in the 0 min time point for BT 3782 is due to an experimental error and the enzyme not being fully inactivated immediately. M= Mannose

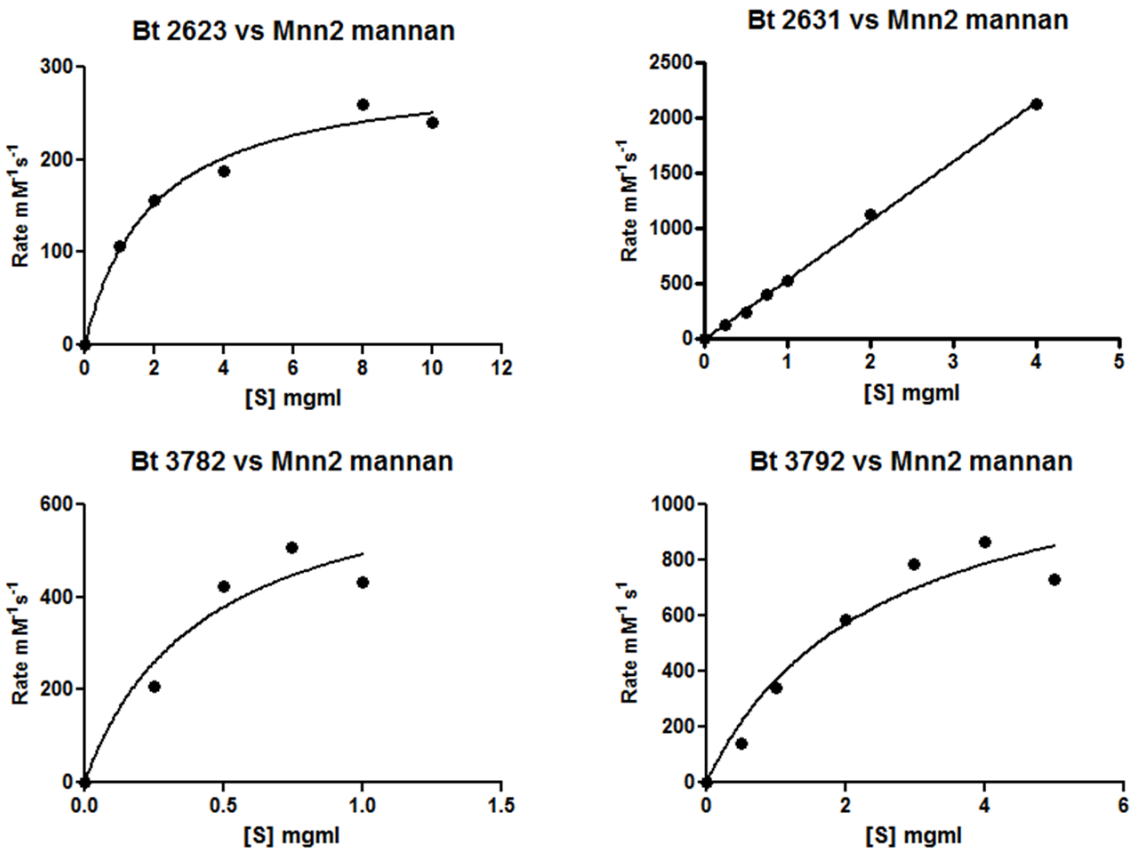


Figure 5.6 Example of kinetic analysis of GH76 enzymes against Mnn2 mutant yeast mannan

Initial rates derived from DNSA were plotted in GraphPad Prism 5.0 using Michaelis Menten equation or linear regression.

Protein Name	K_m mg ml^{-1}	k_{cat} s^{-1}	k_{cat}/K_m s^{-1} mg^{-1} ml
BT 2623	1.892 ± 0.399	297 ± 19.07	157
BT 2631	nd	nd	743 ± 8.048
BT 3782	0.4315 ± 0.333	705± 219	1633
BT 3792	2.425 ± 1.136	1263 ± 262	521

Table 5.2 Kinetic values for GH76 enzymes against Mnn2 mutant yeast mannan
nd= Not determined.

5.3.3 Characterisation of GH76 containing PULs

The characterised GH76 enzymes BT 2623, BT 2631 and BT 3782, BT 3792 are expressed as components of PULs regulated by α-mannan. This section describes the characterisation of other proteins found within PUL 36 (BT 2615-BT 2633) and PUL 68 (BT 3773-BT 3792).

5.3.3.1 PUL 36 (BT 2615 - BT 2633)

PUL 36 gene expression is regulated by a HTCS in response to α-mannan. Table 5.3 shows the proteins encoded by PUL36; it details their predicted and characterised functions (if characterised) and their predicted cellular localizations. PUL 36 contains five enzymes that are annotated as glycoside hydrolases, with two of them in GH76 BT 2623 and BT 2631. BT 2620 is annotated as a GH97 member, a family that contains α-glucosidases and α-galactosidases (www.cazy.org), BT 2622 is located in GH67, a family that contains α-glucoronidases and xylan α1,2-glucoronidases (www.cazy.org). Neither of these predicted enzymes have not currently been characterised as either part of this study or any earlier study.

BT 2629 is annotated as a GH92 enzyme, it was characterised previously as part of a wider study on the extensive *B. thetaiotaomicron* family of GH92 proteins, however a specific activity for BT 2629 could not be identified. The enzyme appeared not to be active against any disaccharide substrates or N-glycans, although it was apparently active against wild type yeast mannan (Zhu et al., 2010). To further investigate the role of BT 2629, the enzyme was incubated at 37 °C with all the mutant mannans available for 16 h and the reaction visualised by TLC. Figure 5.6 shows the results of the TLC analysis. BT 2629 was capable of releasing mannose from wild type yeast mannan, Mnn1, Mnn4, Mnn5 and Mnn6 mutant yeast mannans.

The Mnn1 mutation results in a yeast mannan produced that lacks the terminal α1,3 linked mannose, whereas the Mnn5 mutant produces yeast mannan that lacks the second row of the α1,2 linked mannose branches of the outer chain (Lewis and Ballou, 1991). Mnn4 and Mnn6 regulate the expression of the phosphorylated mannotriose of the outer chain and so this decoration is absent in both of these mutant yeast mannans (Hernandez et al., 1989a). These results suggest that BT2629 is possibly a non specific α1,2 and α1,3 mannosidase that removes the branches of the outer chain of yeast mannan. However it could also be possible that BT2629 hydrolyses a specific α1,2 mannose-α1,6 mannose structure. The enzyme is predicated to be found within the periplasm of the cell suggesting it has a processing role in the deconstruction of yeast mannan polysaccharide. As it is found in the periplasm of the cell, BT2629 is most likely to encounter mannan oligosaccharides making specificity for a α1,2 man-α1,6man oligosaccharide plausible.

Protein name	Characterised	Predicted function	Characterised function	Predicited cellular localization
BT 2618	NO	two component system response regulator		Cytoplasm
BT 2619	NO	two component system sensor histidine Kinase		Trans membrane
BT 2620	NO	GH97-α glucosidase and α galctosidase		Periplasm
BT 2621	NO	Hypothetical protein		Cytoplasm
BT 2622	NO	GH67- α glucoronidase and xylan α1,2 glucoronidase		Cytoplasm
BT 2623	YES	GH76- α 1, 6 mannanase	α 1, 6 mannanase	Outer membrane
BT 2624	NO	Hypothetical protein		Outer membrane
BT 2625	NO	SusD-like		Outer membrane
BT 2626	NO	SusC-like		Periplasm
BT 2627	NO	Cell surface protein		Outer membrane
BT 2628	NO	HTCS sensor/regulator (TM)		Trans membrane
BT 2629	YES	GH92 - α mannosidase	α 1,2 and α1,3 mannosidase	Periplasm
BT 2630	NO	Hypothetical protein		Outer membrane
BT 2631	YES	GH76- α 1, 6 mannanase	α 1, 6 mannanase	Periplasm
BT 2632	NO	GH125- exo α 1,6 mannosidase		Periplasm
BT 2633	NO	Hypothetical protein		Cytoplasm

Table 5.3 Properties of proteins found in PUL 36

Cellular locations predicted using SignalP and LipoP based on amino acid analysis

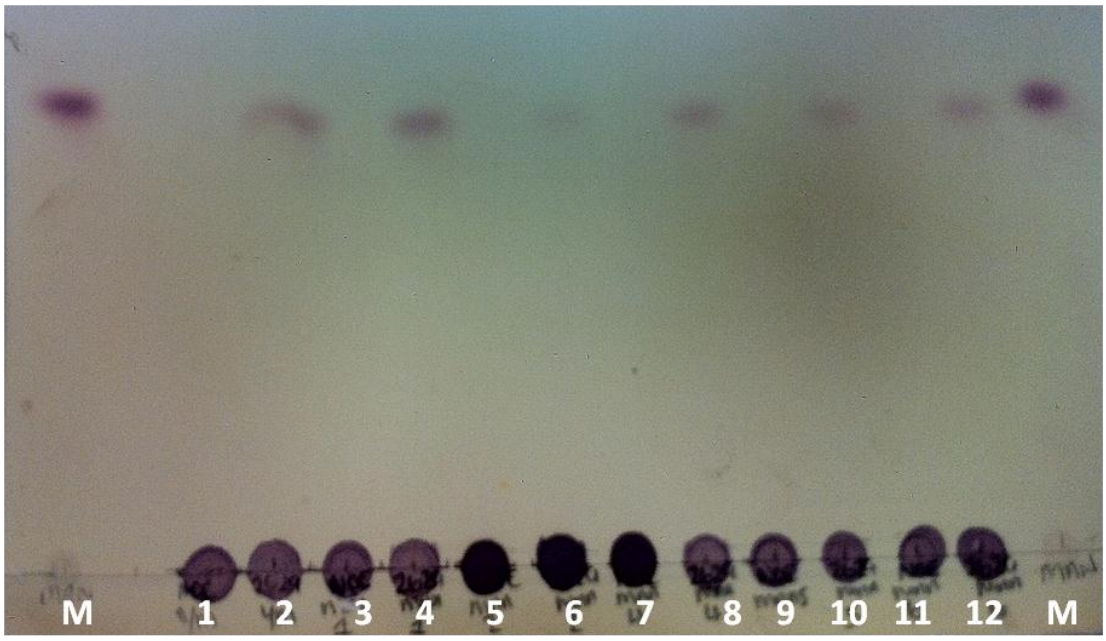


Figure 5.7 TLC analysis of BT 2629 (1 µM) vs mutant mannans (1 mM)

M = Mannose , 1= No Enzyme vs Wt mannan ,2= BT 2629 vs Wt mannan, 3= No Enzyme vs Mnn1 mannan, 4= BT 2629 vs Mnn1 mannan , 5= No Enzyme vs Mnn2 mannan , 6= BT 2629 vs Mnn2 mannan , 7= No Enzyme vs Mnn4 mannan, 8= BT 2629 vs Mnn4 mannan, 9= No Enzyme vs Mnn5 mannan, 10= BT 2629 vs Mnn5 mannan, 11= No Enzyme vs Mnn6 mannan ,12= BT 2629 vs Mnn6 mannan. All experiments were incubated at 37 °C for 16 h.

5.3.3.2 PUL 68 (BT 3773- BT 3792)

PUL 68 is the second of the GH76 containing PULs that is regulated by a HTCS that is activated by α-mannan. Table 5.4 shows the proteins encoded by PUL 68 and their predicted function and cellular localization.

PUL 68 encodes five enzymes, which are annotated as glycoside hydrolases. There are two GH76 enzymes (BT 3782 and BT 3792), two GH92 (BT 3773 and BT 3784) and one GH38/CBM32 (BT 3774).

BT 3774 was not characterised as part of this study but has been annotated as a GH38 enzyme containing a CBM32. Known activities for GH38 enzymes include mannosyl-oligosaccharide α1,3 α1,6 mannosidases and mannosyl- oligosaccharides α1,3 mannosidases (www.cazy.org). BT 3774 is predicted to be present within the periplasm of the cell, indicating that it probably plays a role in deconstruction of oligosaccharides ready for import into the cytoplasm.

PUL 68, unlike PUL 36, encodes two GH92 enzymes, which were both characterised in a previous study by Yanping Zhu (Zhu et al., 2010). The first GH92 BT 3773 was characterised as an α1,3 mannosidase active both on the disaccharide and the polysaccharide yeast mannan. However, when tested against mannan produced from *S. cerevisiae* there appeared to be no mannose released from the wild type or mutant forms of the polysaccharide, Figure 5.7. The second GH92 BT 3784 was previously characterised as an α1,2 mannosidase which is active on the disaccharide and the mannan polysaccharide. TLC analyses of reactions against the mutant mannans reveal it is capable of releasing mannose from all of the mutant mannans, Figure 5.8. This may seem a little strange as it has previously been characterised as an α1,2-mannosidase, however BT 3784 does retain some activity

against other linkages. The observation that BT 3784 appears to be highly active on both wild type yeast mannan and the mnn1 mutant mannan suggests that the enzyme's α-1,3 mannosidase activity strips the mannose side chain exposing the α1,2 linked mannose units, which BT 3784 can remove more readily. The Mnn1 mutant mannan lacks the terminal α1, 3 linked mannose therefore BT 3784 can readily degrade the remaining α1,2-linked mannose side chains. Both GH92 enzymes are found within the periplasm and so have a processing role removing side chains prior to transport into the cytoplasm.

BT 3783 is annotated as a hypothetical protein, when the protein sequence is entered in BLAST search it shows a high homology to phosphatase enzymes. To test if BT 3783 displayed phosphatase activity it was incubated with mannose-6-phosphate, glucose-6-phosphate, fructose-6-phosphate and mannose-1-phosphate. Released monosaccharide was displayed by TLC analysis as the monosaccharide migrates further up the silica plate than the phosphorylated monosaccharide. Figure 5.9 shows a time course of BT 3783 against the three sugar phosphates. As expected BT 3783 appears to be most active on mannose-6-phosphate, but is also highly active against glucose-6-phosphate. BT 3783 does not show significant activity against fructose-6-phosphate or mannose-1-phosphate. These data show that the enzyme is specific for pyranose sugars phosphorylated at O6.

To determine the kinetic parameters of BT 3783 against both mannose-6-phosphate and glucose-6-phosphate, BT 3783 was incubated with various concentrations of each substrate and aliquots taken over a fixed time period. The aliquots were treated with dowex resin to remove any remaining substrate and then analysed by high pressure anion exchange chromatography (HPAEC). Experiments were carried out in 50 mM HEPES at pH 7.0 at 37 °C. Monosaccharide release was plotted in

Graphpad prism and initial rates were analysed by non-linear regression to calculate kinetic parameters. Example HPAEC traces and initial rates are shown in Figure 5.10. Figure 5.11 shows non-linear regression plots and Table 5.5 reports the kinetic values for BT 3783 against mannose-6-phosphate and glucose -6-phosphate.

The K_m of BT 3783 against mannose-6-phosphate is higher than against glucose-6-phosphate suggesting tighter binding to the glucose configured substrate. In contrast the turnover rate (k_{cat}) was considerable higher for phosphorylated mannose compared to glucose-6-phosphate. The k_{cat}/K_m shows that the catalytic efficiency of the enzyme for mannose-6-phosphate is considerably higher than against glucose-6-phosphate, although the optimum substrate may be a phosphorylated mannooligosaccharide. The higher activity against phosphorylated mannose was also apparent in the TLC time course as the disappearance of glucose-6-phosphate is slower than mannose-6-phosphate. Mannose-6-phosphate was fully hydrolysed after 10 min whereas small amounts of glucose-6-phosphate were evident after 60 min.

BT 3786 was predicted to be the HCTS sensor/ regulator of PUL 68, the periplasmic domain was cloned, expressed and purified by Carl Morland and binding studies carried out. Isothermal titration calriometry (ITC) was utilised to assess the ability of BT 3786 sensor domain to bind sugars. BT 3786 was titrated with both Mnn2 mannan polysaccharide and a oligosaccharide assortment derived from Mnn2 mannan. Mnn2 oligosaccharides were generated by incubation of Mnn2 mannan polysaccharide with the GH76 enzyme BT 3792 at 37 °C for 1hr. Preliminary ITC revealed that BT3786 was not capable of binding mnn2 mannan but was capable of binding oligosaccharides derived from mnn2 mannan. BT 3786 displayed binding to a crude mixture of mnn2 oligosaccharides but it is not possible to calculate the

thermodynamics of binding as the exact concentration of sugars and the precise oligosaccharide BT 3786 binds to are unknown. The ITC trace in Figure 5.12 shows that the titration does not go to saturation and as the heat peaks, in the upper panel, are very small the binding appears to be fairly weak. This may be due to a low concentration of ligand present within the oligosaccharide mixture.

Protein name	Characterised	Predicted function	Characterised function	Predicted cellular localization
BT 3773	YES	GH92 - α mannosidase	α1,3 mannosidase	Periplasm
BT 3774	NO	GH38 - mannosyl-oligosaccharide α-1,3-1,6-mannosidase		Periplasm
BT 3775	NO	GT32- mannosyltranferase		Cytoplasm
BT 3776	NO	GT32- mannosyltranferase		Cytoplasm
BT 3777	NO	Hypothetical protein		Periplasm
BT 3778	NO	Hypothetical protein		Cytoplasm
BT 3779	NO	Hypothetical protein		Cytoplasm
BT 3780	NO	Hypothetical protein		Outer membrane
BT 3781	NO	GH125 – exo α1,6 mannosidase		Periplasm
BT 3782	YES	GH76 - endo-α 1,6-mannanase	Endo-α1,6-mannanase	Periplasm
BT 3783	YES	Phosphatase	mannosyl-, glucosyl phosphatase	Outer membrane
BT 3784	YES	GH92 - α mannosidase	α1,2 mannosidase	Periplasm
BT 3785	NO	Hypothetical protein		Cytoplasm
BT 3786	NO	HTCS sensor/regulator (TM)		Periplasm
BT 3787	NO	Hypothetical protein		Outer membrane
BT 3788	NO	SusC-like		Cytoplasm
BT 3789	NO	SusD-like		Outer membrane
BT 3790	NO	Hypothetical protein		Outer membrane
BT 3791	NO	Hypothetical protein		Outer membrane
BT 3792	YES	GH76- α 1,6 mannanase	Endo-α1,6-mannanase	Outer membrane

Table 5.4 Properties of Proteins found in PUL 68

Cellular locations predicted using SignalP and LipoP based on amino acid analysis.

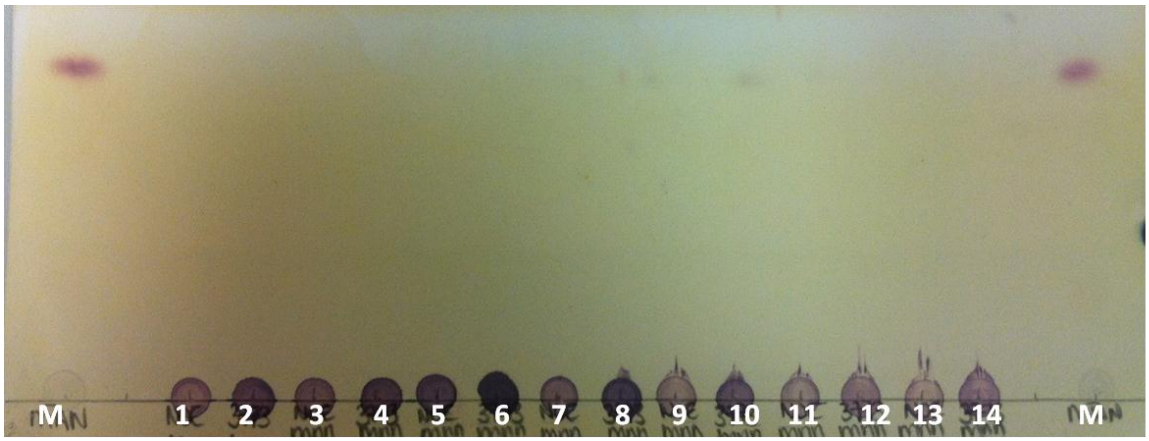


Figure 5.8 TLC analysis of BT 3773 vs mutant mannans

M = Mannose , 1= No Enzyme vs Wt mannan ,2= BT 3773 vs Wt mannan, 3= No Enzyme vs Mnn1 mannan, 4= BT 3773 vs Mnn1 mannan , 5= No Enzyme vs Mnn2 mannan , 6= BT 3773 vs Mnn2 mannan , 7= No Enzyme vs Mnn4 mannan, 8= BT 3773 vs Mnn4 mannan, 9= No Enzyme vs Mnn5 mannan, 10= BT 3773 vs Mnn5 mannan, 11= No Enzyme vs Mnn6 mannan ,12= BT 3773 vs Mnn6 mannan, 13= No Enzyme vs Mnn9 mannan, 14= BT 3773 vs Mnn9 mannan. Enzyme concentration 1 µM and substrate concentration 1mM incubated for 16 hrs at 37 °C

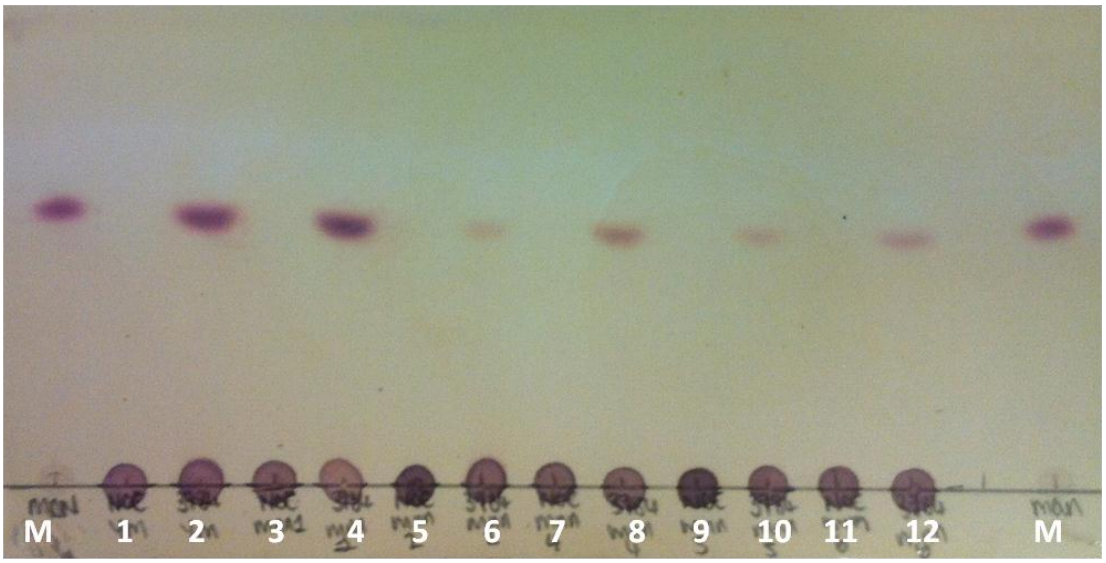


Figure 5.9 TLC analysis of BT 3784 vs mutant mannans

M = Mannose , 1= No Enzyme vs Wt mannan ,2= BT 3784 vs Wt mannan, 3= No Enzyme vs Mnn1 mannan, 4= BT 3784 vs Mnn1 mannan , 5= No Enzyme vs Mnn2 mannan , 6= BT 3784 vs Mnn2 mannan , 7= No Enzyme vs Mnn4 mannan, 8= BT 3784 vs Mnn4 mannan, 9= No Enzyme vs Mnn5 mannan, 10= BT 3784 vs Mnn5 mannan, 11= No Enzyme vs Mnn6 mannan ,12= BT 3784 vs Mnn6 mannan Enzyme concentration 1 µM and substrarte concentration 1mM incubated for 16 hrs at 37 °C

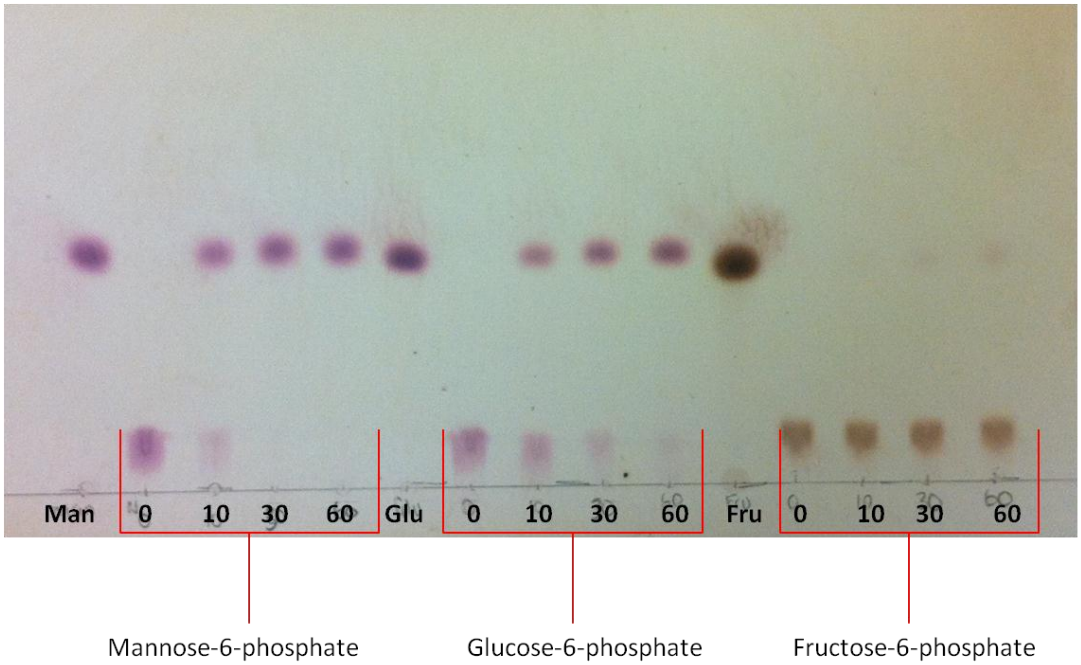


Figure 5.10 TLC analysis of BT 3783 activity over 60mins

BT 3783 (1 µM) incubated with substrates (1 mM) as labelled in 50 mM HEPES pH 7.5 at 37 °C Aliquots taken at 0, 10, 30, 60 min. The 0 min time point shows the starting monosaccharide-6-phosphate.

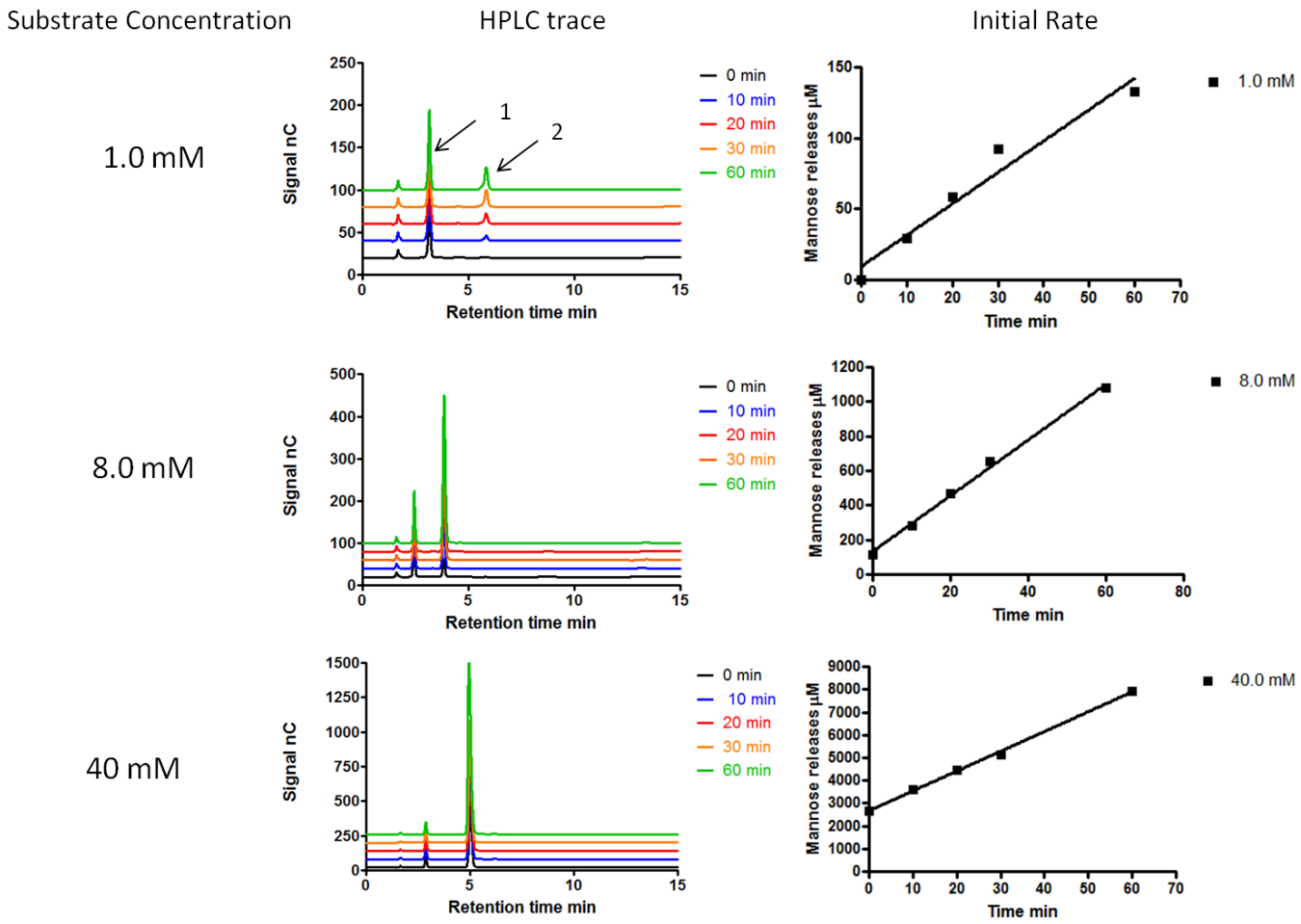


Figure 5.11 Example HPLC traces and initial rate graphs for 1 µM BT 3783 vs mannose-6-phosphate

HPLC traces plotted in Graphpad prism. Each graph shows the HPLC trace for each time point for the relevant substrate concentration. Peak 1 shows the fucose internal standard, Peak 2 shows the mannose released. Initial rate graphs show the mannose released in µM as a function of time.

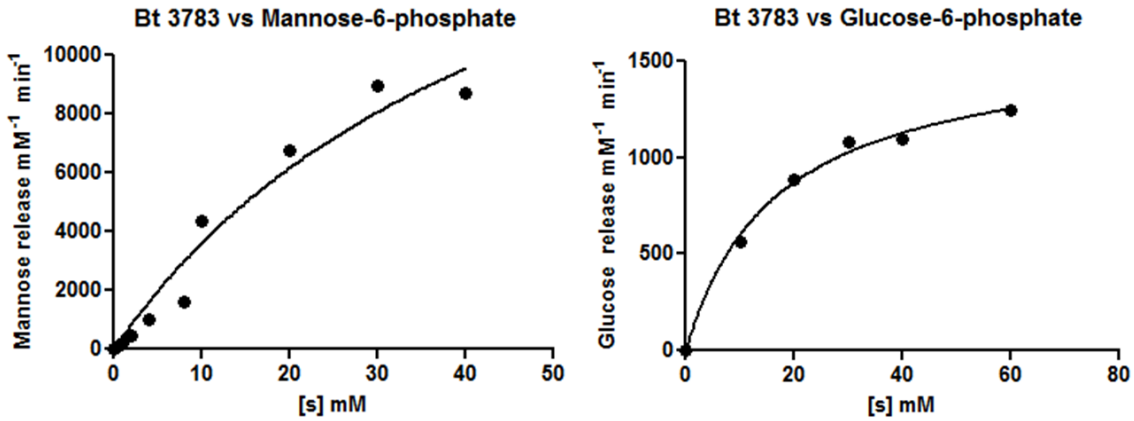


Figure 5.12 Example Non-linear regression curves for BT 3783

Initial rates were plotted in GraphPad prism using the Mechaelis-Menten equation.

Substrate	K_m mM	K_{cat} s^{-1}	K_{cat}/K_m M^{-1} s^{-1}
Man-6-P	48.4 ± 25.1	351 ± 119	7252
Glu-6-P	16.68 ± 2.412	26.68 ± 1.3	1575

Table 5.5 Kinetic values for BT 3783 against mannose-6-phosphate (Man6-P) and glucose-6-phosphate

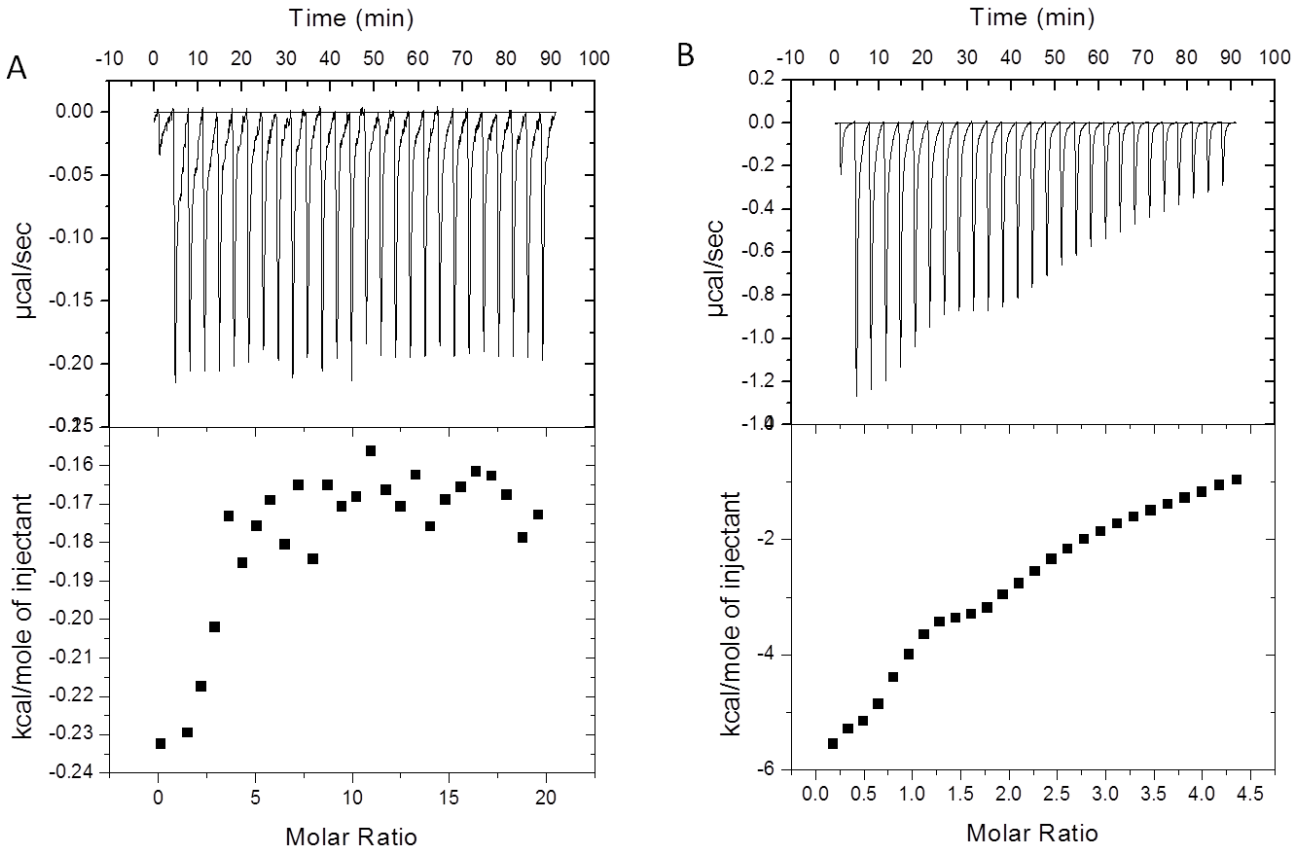


Figure 5.13 ITC trace of BT 3786

Panel A shows 50 µM BT 3786 vs 1 mg/ml Mnn2 mannan Panel B shows 50 µM BT 3786 vs 1 mg/ml Mnn2 oligosaccharides . All ITC experiments were carried out in 50mM HEPES ph 7.5 at 20 °C

5.4 Discussion

This chapter attempts to dissect how *B. thetaiotaomicron* deconstructs the mannan polysaccharide produced from the yeast *S.cerevisiae.* As yeast mannan is a complex polysaccharide with various different glycosidic linkages, understandably, an array of enzymes is required to deconstruct the polysaccharide into useable oligosaccharides. PUL 36 and 68 from *B. theataiomicron* are upregulated in the presence of yeast mannan; both PULs are controlled by a hybrid two component system (HTCS). Each PUL encodes a pair of GH76 enzymes located on the outer membrane and in the periplasm, respectively. The outer membrane localized GH76 enzymes (BT 2623 BT 3792) both release large oligosacchairdes from the mnn2 mannan, while the periplasmic GH76 enzymes (BT 2631 and BT 3782) generate small oligosaccharides from the mnn2 mannan. This difference is significant as studies on other PULs from *B. thetaiotaomicron* reveal that, along with the SusC/D pair localised at the outer membrane, a number of enzymes usually process the polysaccharide before the oligosaccharides are transported across the membrane into the periplasm (Bjursell et al., 2006, Martens et al., 2008). One of these enzymes is often an endo-acting enzyme that produces the oligosaccharides for transportation. BT 2623 and BT 3792 appear to be the outer membrane endo-acting enzymes, while BT 2631 and BT 3782 are periplasmic and thus further process the oligosaccharides which are ultimately converted to monosaccharides and transported into the cytoplasm. This type of enzyme organisation is typical of PUL-encoded degradative systems that degrade complex polysaccharides (Sonnenburg et al., 2010).

PUL 36 encodes one GH92 α- mannosidase, BT 2629, whereas PUL 68 encodes two GH92 enzymes, BT 3773 and BT 3784. BT 2629 appears to be a non-specific α

mannosidase capable of removing both the α1,2 and α1,3 mannose branches from the backbone of the main outer chain of wild type yeast mannan. The pair of GH92 enzymes encoded in PUL 68 have distinct activities. BT 3773 is an α1, 3-mannosidase, however, when BT 3773 was tested against wild type and mutant mannans there was no mannose released when visualised by TLC. The in activity observed for BT 3773 may be due to a preparation of in active protein or possibly that the enzyme prefers an oligosaccharide substrate rather than the polysaccharide substrates tested. BT 3784 shows activity against the wild type and all of the mutant mannans. This is indicative of BT 3784 having mainly an α1,2 mannosidase activity although the enzyme can hydrolyse α1,3 mannose linkages, as reported by Yanping Zhu (Zhu et al., 2010). Mannose released from the mutant mannans by BT 3784 could also be attributed to the hydrolysis of α1,2 glycosidic bonds of the core of the mannan polysaccharide as the side chains are not always capped with the α1,3 linked mannose and only the mnn1 mutant effects the structure of the core polysaccharide. The three GH92 enzymes are all predicted to be found within the periplasm of the *B.thetaiotaomicron* cell. The presence of exo acting glycoside hydrolases in the periplasm is fairly typical of Sus like PULs, whereby the GH92, in this case, can strip the mannose from the side chains. The mannose can then be transported into the cytoplasm for utilization. The lack of specificity of the periplasmic exo acting enzymes does not seem to be unique to the yeast mannan PULs as the fructan utilization locus appears to display the same characteristic (Sonnenburg et al., 2010).

The GH76 and GH92 enzymes encoded in PUL 36 and PUL 68 have the capability to hydrolyse all the major linkages found within yeast mannan; however there appears to be a flaw in the model. The two outer membrane endo acting GH76

enzymes will only produce oligosaccharides from the mannan produced by the mnn2 mutant yeast. This mutant mannan consists of the α1,6 linked mannose backbone, devoid of any branches. In the human gut the bacteria would encounter wild type yeast mannan that retains its α1,2 branches capped with an α1,3 linked mannose. The GH92 enzymes are capable of processing the side chains to reveal the α1,6 mannose linked backbone. However all three of the GH92 enzymes present within the two PULs are predicted to be localised in the periplasm. The enzymes in the periplasm will only encounter the oligosaccharides transported through the outer membrane by the SusC like protein which have been generated by the endo acting enzymes on the surface. Therefore there must be additional enzymes located at the cell surface capable of removing the side chains of yeast mannan in either an endo or exo manner.

Previously studied sensor domains of HTCS involved in polysaccharide utilisation revealed that regulation of the PUL was driven by direct binding of an oligosaccharide to the periplasmic sensor domain. The sensor domains studied bound only linear oligosaccharides, suggesting that branches were not required for recognition of polysaccharides. BT 3786 is annotated as the HTCS sensor/regulator of PUL 68 and ITC showed that the periplasmic sensor domain bound oligosaccharides derived from the Mnn2. The oligosaccharides bound by BT 3786 are linear α1,6 linked mannan oligosaccharides which, lack the α1,2 and α1,3 branches found in wild type yeast mannan. Therefore it appears BT 3786 exhibits the characteristics previously displayed by other HTCS sensor domains in binding linear not branched oligosaccharides.

Both PUL 36 and 68 encode for a number of hypothetical proteins that are yet to be characterised and could provide the crucial function of removing the side chains and

allowing the outer membrane GH76 enzymes to deconstruct the backbone. In addition PUL 68 encodes BT 3774 which is annotated as a GH38 enzyme. Unfortunately this enzyme could not be cloned and expressed in solution, so activity assays were not carried out. However, other enzymes in the GH38 family have been characterised as α mannosidases active on mannosyl-oligosaccharides. It is possible that the GH38 could remove the side chains in an endo manner producing α1,2- α1,3 oligosaccharides and revealing the backbone of the yeast mannan. As BT 3774 is predicted to reside at the periplasm of the bacterial cell it would not be able to remove the side chains to aid the outer membrane GH76 enzymes. It does however suggest that the hypothetical proteins encoded in the PUL must provide the de-branching role.

In this chapter one of the hypothetical proteins from PUL 68 has been characterised. Bioinformatic data suggested that BT 3783 may have a phosphatase activity. To test this BT 3783 was cloned and expressed and tested for phosphatase activity against a number of substrates. As explained in the results BT 3783 is indeed capable of releasing a phosphate molecule from both the mannose-6-phosphate and the glucose-6-phosphate. Kinetic data suggest that while BT 3783 is active on both substrates it prefers mannose-6-phosphate. Yeast mannan is a complicated polysaccharide that has varying degrees of phosphorylation ranging from 1 phosphate for every 19 mannose residues to 1 phosphate per 13 mannose residues. The phosphate group can be bound through the C6 or C1 of the mannose residue. BT 3783 is capable of liberating mannose from mannose-6-phosphate, but a secondary enzyme is required to hydrolyse mannose-1-phosphate. BT 3783 is predicted to reside at the outer membrane and so will most probably encounter mannose-6-phosphate as part of more complex oligosaccharides.

This chapter demonstrates how *B. thetaiotaomicron* possesses within PUL 36 and PUL 68 the enzymes needed to successfully hydrolyse all the major bonds within yeast mannan. Although not all of the annotated enzymes have been characterised as part of this study, those that have contribute to the degradation of the complex polysaccharide. Both PULs are upregulated in response to yeast mannan and most probably work simultaneously to sense, transport and utilise this polysaccharide.

5.5 Future Work

Future work crucial to understanding how *B. thetaiotaomicron* utilises yeast mannan would be to characterise the other annotated enzymes and also to screen the hypothetical proteins for potential activity against yeast mannan, Figure 5.13. Further binding studies with both the SusD like protein and the sensor domain from the HTCS of the two PULs would provide some insight into the regulation and the specificity of the PULs. Growth studies of *B. thetaiotaomicron* on simple sugars and yeast mannan with and without the SusC like protein knocked out would provide some insight into the ability of *B. thetaiotaomicron* to utilise yeast mannan and the importance of the SusC like protein. All of the above studies would provide a more complete knowledge of how *B. thetaiotaomicron* senses, transports and utilises complex carbohydrates found in the human gut.

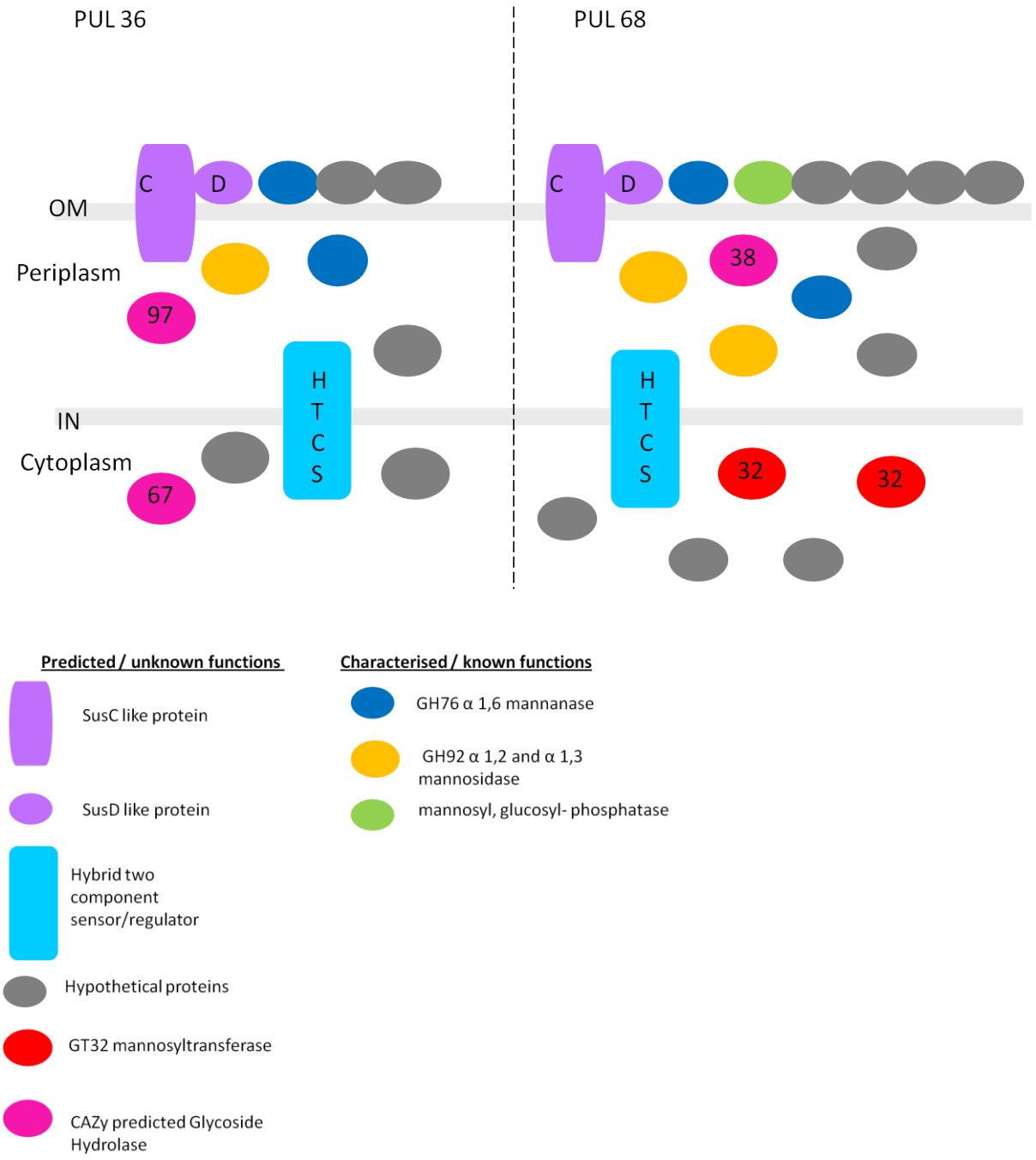


Figure 5.14 Diagram of PUL 36 and PUL 68 showing known and unknown activities

Chapter 6: Probing the Active Site of a Arabinoxylan Specific GH5 Xylanase

6.1 Introduction

The plant cell wall consists of various interlinked polysaccharides whose structures vary with respect to their sugar components and linkages. The major component of the plant cell wall is the homo-polysaccharide cellulose, which is largely crystalline and thus resistant to enzymatic hydrolysis. Hemicelluloses are a group of matrix polysaccharides that are chemically more complex than cellulose, but are not crystalline. The major hemicellulose present in the secondary cell wall of dicots is xylan, a β1,4 xylose polymer that is decorated with various sugars, primarily glucuorinc acid and arabinose units, and can also be acetylated. While the major xylanases found in nature, located in GH10 and GH11, efficiently degrade unsubstituted xylans, they are unable to deconstruct highly decorated forms of the polysaccharide such as arabinoxylan and glucuronoarabinoxylan. In arabinoxylan the xylan backbone is highly decorated with α-arabinofuranose side chains that are linked O2 and/or O3 to the xylose residues.

Previously Dr Yanping Zhu characterised a GH5 family arabinoxylan-specific xylanase enzyme, *Ct*Xyl5A, from *Clostridium thermocellum* (Correia et al., 2011). *Ct*Xyl5A has a complex molecular architecture comprising an N-terminal GH5 catalytic module abutted onto a CBM6, while the enzyme also contains a CBM13, fibronectin module, a CBM62 and a C-terminal type I dockerin that facilitates the recruitment of the protein into the *C. thermocellum* cellulosome, displayed in Figure 6.1 . The study showed that the CBM6 binds the termni of *xylo*- and *gluco*-configured

polymers suggesting that its role is to target the enzyme to areas of the plant cell wall that are being actively degraded. The CBM62 binds to terminal galactose decorations of polysaccharides, which may indicate that the enzyme might preferentially act on unusual xylans that contain galactose side chains. Structural and biochemical studies of the GH5 component of *Ct*Xyl5A revealed the mechanism by which the enzyme displayed its arabinoxylanase activity. The GH5 module accommodates arabinofuranose units appended to the O3 of the xylopyranose bound at the active site (-1 subsite), and utilizes this as a specific determinant for hydrolysis. The distal subsites of the enzyme are also able to accommodate arabinose side chains allowing for efficient hydrolysis of highly decorated arabinoxylans. An apo structure of GH5-CBM6 was solved, Figure 6.2. The data revealed a large pocket (defined as the -2NR* subsite) that abuts onto the -1 subsite, which could, potentially, accommodate a sugar decoration appended O3 to the xylose positioned in the active site. The structure also showed that GH5-CBM6 lacks the amino acid residues that interact with the O3 of the active site sugar. In addition, four amino acids in the -2NR* subsite were identified that were candidates to interact with the arabinose attached to the xylose housed in the active site. A structure of GH5-CBM6 in complex with either substrate or reaction products was not solved and consequently the mechanism by which arabinose binds to the -2NR* subsite pocket could only be modelled.

Figure 6.1 Molecular architecture of *Ct*Xyl5A protein

Created in MyDomains

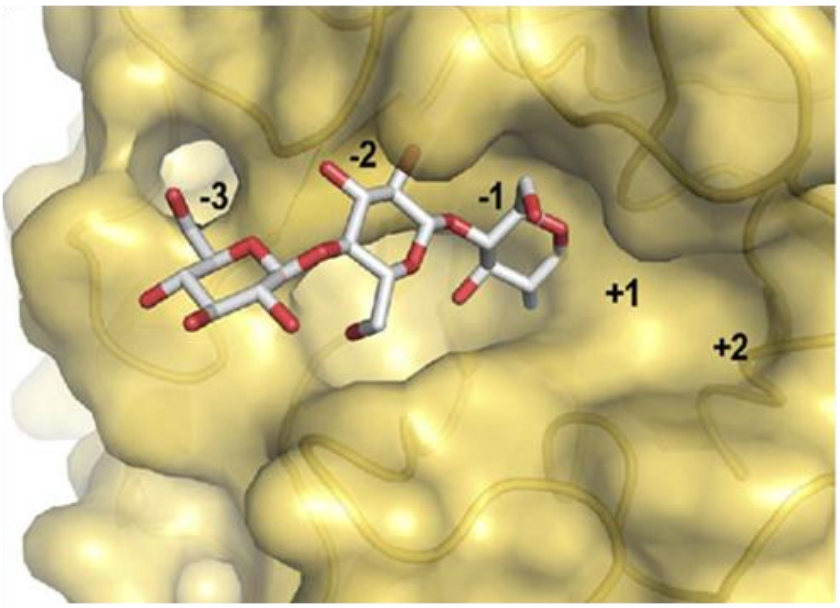


Figure 6.2 GH5 active site

GH5 active site showing subsites, superimposed with 2-deoxy-2-fluorocellotriose. Adapted from (M A Correia, 2011)

6.2 Objectives

The main aim of this chapter is to understand the mechanism by which arabinose binds to the -2NR* subsite of *Ct*Xyl5A and how these residues contributes to the arabinoxylanase activity displayed by the enzyme. This was achieved by site-directed mutagenesis of the amino acids identified as candidates for interaction with the -2NR* arabinose and activity assays of the resultant mutants to identify their importance for hydrolysis. In addition the crystallisation of GH5-CBM6 in complex with arabinose or reaction products of the enzyme was carried out to provide insight into the mechanism of substrate binding at this critical subsite.

6.3 Results

6.3.1 Site directed mutagenesis and protein expression and purification

The GH5-CBM6 wild type gene was previously cloned into the pET21a vector supplying a C-terminal His_6-tag to the recombinant protein. This construct was used as the template DNA for the site directed mutagenesis using the primers shown in Table 6.1. Site directed mutagenesis was carried out as described in Section 2.1.19. To generate the treble mutant, GH5-CBM6 E68A was used as template DNA to generate the double mutant GH5-CBM6 E68A/Y92A. This double mutant was then used as template DNA to generate the GH5-CBM6 E68A/Y92A/N135A mutant which was used as template to generate the final quadruple mutant GH5-CBM6 E68A/Y92A/N135A/N139A. Colonies obtained were sequenced to confirm that only the appropriate mutations had been incorporated into the *Ct*Xyl5A gene.

All of the recombinant proteins were expressed in *E. coli* BL21 (DE3) cells, which were cultured at 37 °C to mid-exponential phase. Recombinant protein expression was induced by the addition of 1 mM IPTG and incubation for a further 5 h at 37 °C before being harvested and purified by IMAC. For crystallography the wild type protein was further purified by size exclusion chromatography

Primer name	Sequence
Gh5 E68A FWD	CCA TAC ACT TCC ACG **GCC** TGG ACG GCG GCA GCT
Gh5 E68A REV	AGC TGC CGC CGT CCA **GGC** CGT GGA AGT GTA TGG
Gh5 Y92A FWD	GCA GTA CAC CTC **GCC** GCA GAA TGC
Gh5 Y92A REV	GCA TTC TGC **GGC** GAG GTG TAC TGC
Gh5 N135A FWD	GTA ATA ACC ATA GGC **GCC** GGT GCC AAT AAC
Gh5 N135A REV	GTT ATT GGC ACC **GGC** GCC TAT GGT TAT TAC
Gh5 N139A FWD	GGC AAC GGT GCC AAT **GCC** GGA AAT CAT AAC GCG
Gh5 N139A REV	CGC GTT ATG ATT TCC **GGC** ATT GGC ACC GTT GCC

Table 6.1 Primers used for site directed mutagenesis
Mutations are shown in bold.

6.3.2 Ability of GH5-CBM6 mutants to hydrolyse arabinoxyalan

This section describes the effects of mutations made to the -2NR* subsite on the ability of GH5-CBM6 to hydrolyse arabinoxylan. To do this the enzyme activity of each mutant was compared to that of the wild type GH5-CBM6. Enzyme activity was measured using the 3, 5-dinitrosalicyclic acid assay which measures the rate of reducing sugar released (described in more detail in Section 2.3.1.3). The K_m of wild type GH5-CBM6 enzyme appears to be very high and the catalytic rate was directly proportional to a substrate concentration up to 2 mg/ml. For this reason the K_m could not be determined and so k_{cat}/ K_m was determined using linear regression. Wild type GH5-CBM6 and the mutants were incubated with 1mg/ml arabinoxylan from wheat in 50 mM sodium phosphate pH 7.5 at 50 °C. Aliquots were taken at different times and reducing sugar released measured. All data were analysed using GraphPad Prism 5.0. The activities of the mutants were reported as a percentage of GH5-CBM6 wild type activity in Table 6.2 and the linear regression graphs are shown in Figure 6.3.

As the binding of arabinose in the -2NR* subsite is thought to be the specificity determinant for hydrolysis of arabinoxylan, alanine substitution of amino acids involved in arabinose recognition should render the enzyme inactive. However, the kinetic analysis of GH5-CBM6 single mutants, containing mutations within the arabinose binding pocket, showed a decrease in, but not a total loss, of activity. In fact only one mutation, E68A, resulted in a substantial decrease in activity; the mutant retained 23% of wild type activity. The other three amino acids Trp92, Asn135 and Asn139, when mutated to alanine, did not appear to have any significant effect on the activity of the enzyme; the three mutants retained >80% of wild type arabinoxylanase activity. This could partly be due to the hydrophobic nature of the -2NR* subsite which would provide a number of amino acids that could interact with

the bound arabinose. In an effort to counter act this possible effect, a treble mutant of E68A/Y92A/N135A and a quadruple mutant, E68A/Y92A/N135A/N139A, were generated. Both of these mutants were tested for activity against wheat arabinoxylan. The treble mutant does show a further decrease in activity compared to the respective three single mutant proteins, however E68A/Y92A/N135A still retained 50% of wild type activity. The quadruple mutant showed a further decrease in activity close to that of the GH5-CBM6 E68A mutant. This is potentially surprising as it would be expected that the quadruple mutant would be inactive due to its inability to interact with the arabinose decoration of the xylose positioned in the -1 subsite (active site).

To assess whether GH5-CBM6 possesses any xylanase activity, the wild type, treble mutant and quadruple mutant were incubated with birchwood xylan and the activity measured. Birchwood xylan is undecorated and thus testing for activity against this source of xylan would reveal if there is any background xylanase activity. The results show that indeed the wild type GH5-CBM6 did posses a small amount of xylanase activity, as did the treble and quadruple mutants. The GH5-CBM6 wild type xylanase activity compared to the arabinoxylanase activity is very low, the quadruple mutant showed a higher xylanase activity compared to its arabinoxylanase activity. Also worth noting is that the respective xylanase and arabinoxylanase activities displayed by the treble and quadruple mutants were similar. This observation would strongly suggest that the activity of the two mutants against arabinoxylan from wheat is background xylanase activity. So, in essence, the treble and quadruple mutants of GH5-CBM6 display little arabinoxylanase-specific activity. These data suggest that the proposed arabinose binding pocket, which abuts onto the -1 subsite (active site)

is the specificity determinant for the arabinoxylanase activity displayed by GH5-CBM6.

Protein Name	k_{cat}/K_m against Wheat arabinoxylan min^{-1} mg ml^{-1}	% of wild type activity against wheat arabinoxylan	K_{cat}/K_m against Birchwood xylan min^{-1} mg ml^{-1}	Birchwood/wheat arabinoxylan activity
GH5-CBM6 WT	5361	100	906	0.17
GH5-CBM6 E68A	1255	23	Nd[a]	Nd
GH5-CBM6 Y92A	4697	87	Nd	Nd
GH5-CBM6 N135A	5207	97	Nd	Nd
GH5- CBM6 N139A	4744	88	Nd	Nd
GH5-CBM6 TM	2769	51	1335	0.48
GH5-CBM6 QM	1135	21	776	0.68

Table 6.2 GH5-CBM6 and mutant activity against wheat arabinoxylan and Birchwood xylan

*Nd; not determined. GH5-CBM6 TM = GH5-CBM6 E68A/Y92A/N135A. GH5-CBM6 QM= GH5-CBM6 E68A/Y92A/N135A/N139A

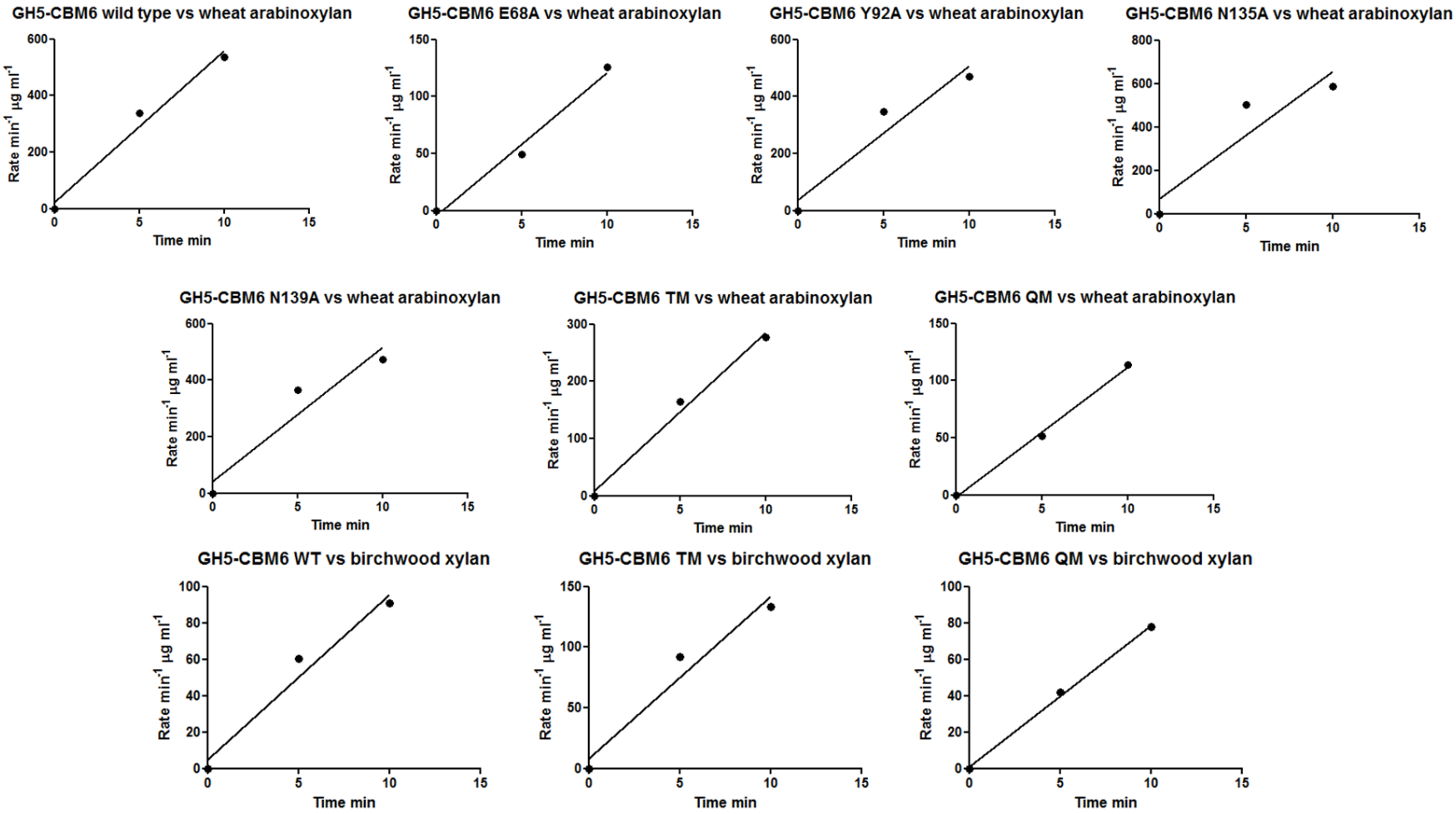


Figure 6.3 Example linear regression graphs for GH5-CBM6 wild type and mutant constructs against wheat arabinoxylan and birchwood xylan

All graphs were plotted in GraphPad Prism. All experiments were carried out with 1µM enzyme and 1 mg ml substrate in 50mM sodium phosphate pH 7.5 at 50 °C. GH5-CBM6 TM = GH5-CBM6 E68A/Y92A/N135A. GH5-CBM6 QM= GH5-CBM6 E68A/Y92A/N135A/N139A

6.3.3 Structure of GH5-CBM6 in complex with arabinose

To understand how GH5-CBM6 interacts with arabinose in the proposed -2NR* binding pocket, GH5-CBM6 (10 mgml) was co-crystallised with arabinose (300 mM) in 1M ammonium sulphate 0.1M bis tris 1% PEG 3350 pH 5.5. Crystals were grown by hanging drop vapour diffusion method at 20 °C for 14 days. The crystals were cryo-protected using the reservoir solution with the addition of 15% PEG 400. The diffraction data was collected at the diamond light source on beam line IO4-1. Diffraction images were integrated using XDS and results scaled using SCALA. The phase problem was solved by molecular replacement using the coordinates from the apo wild type structure deposited in the Protein data bank (PDB) under 2Y8K. Molecular replacement was carried out using molrep software. The final model was built and improved using COOT and refinement was carried out using REFMAC. The fishing of crystals, data collection and solving of the GH5-CBM6 structure with arabinose was carried out by Dr Anaurd Basel in the structural biology lab at Newcastle University.

The crystals grown were a plate like shape, shown in Figure 6.4, and diffracted to 1.75 Å. Crystallographic statistics are shown in Table 6.3. The structure revealed the expected $(\beta/\alpha)_8$ barrel of the GH5 domain linked to the CBM6 domain that displays a β sandwich fold architecture. Unexpectedly the arabinose was bound both in the GH5 domain and in the CBM6 domain. The cartoon structure of GH5-CBM6 in complex with arabinose is shown in Figure 6.5.

Figure 6.4 Crystals of GH5-CBM6 co-crystallised with arabinose

Crystals grown in 1M ammonium sulphate 0.1M bis tris 1% PEG 3350 pH 5.5 at 20°C for 14 days.

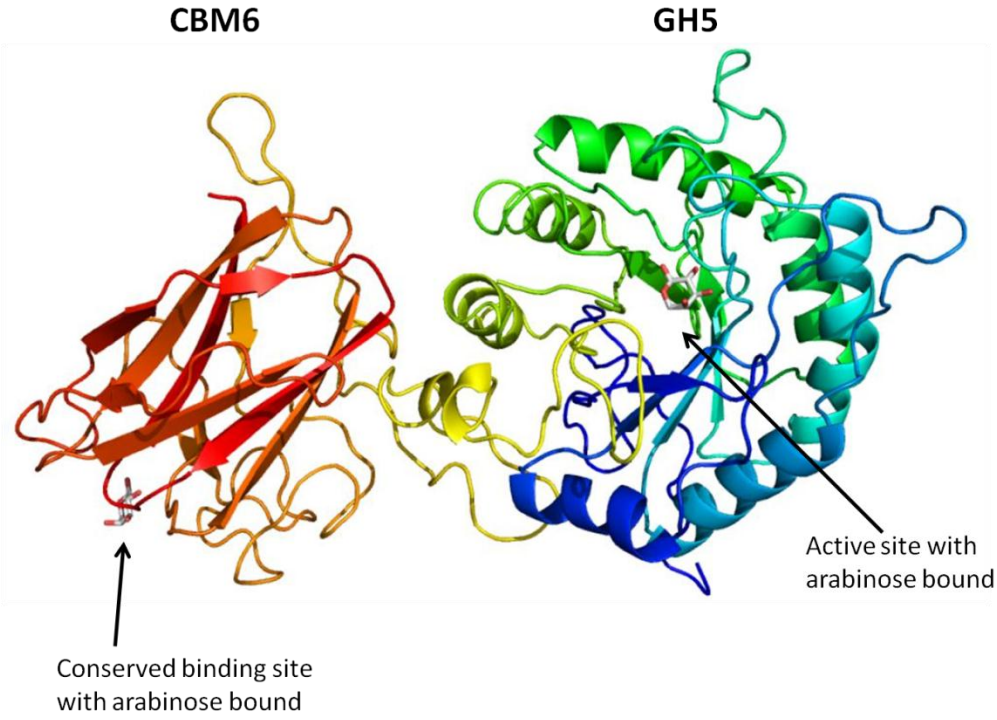


Figure 6.5 Cartoon structure of GH5-CBM6 in complex with arabinose

The protein structure shows arabinose bound within the active site of the GH5 domain and the conserved binding site of the CBM6 domain. Created using PyMol.

	GH5-CBM6 arabinose complex
Data collection	
Space group	P $2_1 2_1 2_1$
Unit cell (Å)	67.0920 72.3660 109.1130
Wavelenth (Å)	0.9173
Resolution (Å)	44.85-1.84 (1.75-1.75)
Number of observations	188803 (25099)
Total number unique	53296 (7590)
Multiplicity	3.5 (3.3)
Completeness	98.3 (97.2)
Average I/sigma I	15.3 (2.5)
Rmerge	0.066 (0.466)
Refinement	
R_{work} (%)	17.74
R_{free} (%)	21
Number of protein atoms	3766
Number of water atoms	248
Number of ligand atoms	10
Number of solvent atoms	8
Rmsd	
Bond lengths (Å)	0.124
Bond angle (°)	1.45
Ramachandran	
Allowed + favoured (%)	100
B-factors	
Protein ($Å^2$)	17
Water ($Å^2$)	24
ligand ($Å^2$)	19
other solvent ($Å^2$)	37.2

Table 6.3 Crystallographic data and refinement statistics

Binding of the arabinose to the CBM6 domain was evident in the conserved sugar binding region. The sugar was sandwiched between the two conserved residues found within the CBM6 family, Trp424 and Phe478. Both conserved residues make hydrophobic interactions with arabinose and, in addition, Trp 424 makes polar contacts with the O2 of the arabinose. The O2 of the arabinose also makes polar contacts with Asp507, while O1 interacts with Glu411. Figure 6.6 shows binding of arabinose, in its pyranose configuration, to CBM6.

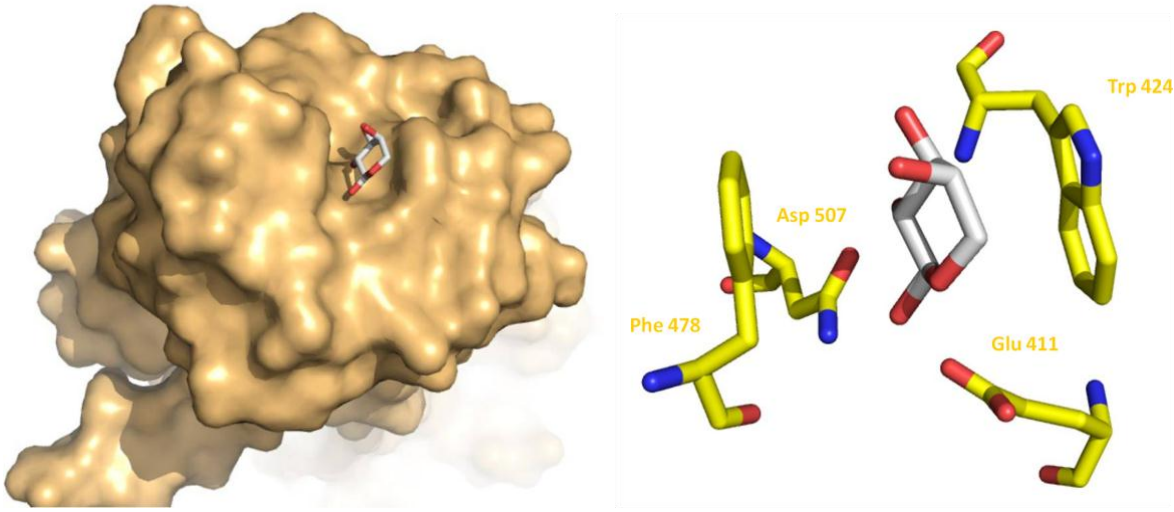


Figure 6.6 CBM6 in complex with arabinose

A shows surface representation of CBM6 in complex with arabinose. B shows amino acids that interact with arabinose. Created using PyMol.

The GH5 domain has an arabinose bound in the -2NR*subsite. Figure 6.7 shows binding of arabinose to the GH5 domain. The structure shows that the arabinose interacts with various amino acids found within the pocket; Glu68 and Gly136 make polar interaction with O3 of the sugar and Asn135 makes hydrogen bonds with the O3 and O2 of the arabinose. Asn139 and Asn170 also make polar interactions and Tyr92 makes hydrophobic contacts with the arabinose. The C1 of the arabinose ring points out towards the catalytic apparatus, Glu171 and Glu279 of the active site

where in fact it would be linked to the O3 of the xylan backbone. Unexpectedly the arabinose bound in the structure adopts a pyranose conformation, as opposed to the furanose form of the sugar found in the arabinoxylans described todate. In solution, arabinose exists in equilibrium between the pyranose and furanose conformations, which highly favours the pyranose conformation. It appears that the -2NR* subsite is not specific for either the pyranose or the furanose form of arabinose and so has bound the dominant arabinopyranose conformation.

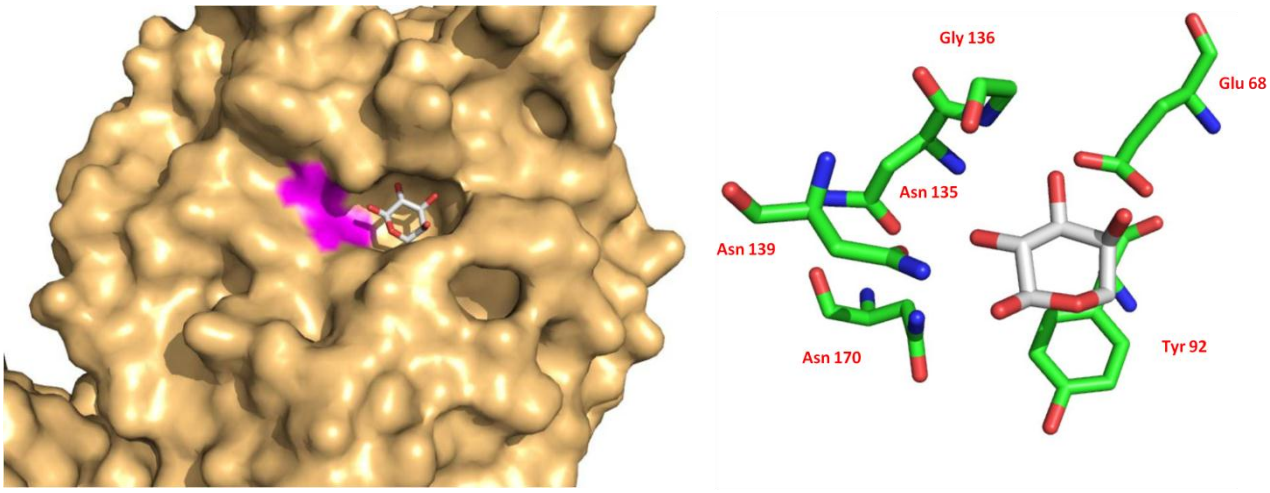


Figure 6.7 GH5 in complex with arabinose

A shows surface representation of the GH5 domain with arabinose bound in the +2NR* subsite. Catalytic residues are coloured in magenta. . B shows the amino acids interact with arabinose. Created using PyMol.

6.3 Discussion

The objective of this chapter was to assess if the arabinoxylanse activity of GH5-CBM6 was determined by the binding of arabinose to the proposed -2NR* subsite. Two approaches were deployed; site directed mutation of the arabinose binding amino acids in the -2NR* subsite and subsequent assays of the resultant mutants. In the second approach the crystal structure of GH5-CBM6 in complex with arabinose was determined.

The activity assays showed that single mutations were not capable of fully inactivating the enzyme. In fact, with the exception of E68A, the single mutations had little if any effect on catalytic activity. The GH5-CBM6 in complex with arabinose reveals why the single mutants have such little effect. The arabinopyranose bound in the -2NR* subsite makes extensive interactions with the amino acids found within the pocket. With such a wealth of interactions mutation of just one amino acid makes very little difference to the ability of arabinose to bind in the pocket. The treble and quadruple GH5-CBM6 mutants showed a considerable decrease in activity against wheat arabinoxylan compared to the wild type enzyme. While the preferred substrate for GH5-CBM6 is arabinoxylan, the enzyme still retains a background xylanase against birchwood xylan. The treble and quadruple GH5-CBM6 mutants also display xylanase activity against birchwood xylan. When the xylanase activity of the wild type enzyme is compared to the treble and quadruple mutants, it becomes evident that a significant amount of the activity displayed by the GH5-CBM6 variants is actually background xylanase activity. The treble and quadruple mutants thus display, effectively, no significant arabinoxylanase-specific activity. From the structural data it is not surprising that the treble and quadruple mutant are almost inactive as virtually all of the interactions made by arabinose to the protein are eliminated. Without the

occupation of arabinose in the -2NR* pocket the arabinoxylan specific activity is lost leaving only residual xylanase activity.

In the GH5-CBM6 complex arabinose is bound in the pyranose conformation. The pyranose conformation is the favoured form of arabinose in solution and so has bound in the pocket as there appears to be no specificity for the furanose conformation that would be found in the arabinoxylan decoration. As the protein does not confer any specificity for the conformation of arabinose it would be less energetically favourable for it to bind in the furanose form as a free monosaccharide. However in the arabinoxylan polysaccharide the arabinose decoration is held in the furanose conformation eradicating the energetic issue. Therefore the arabinose binding pocket appears to be non-specific for either the furanose or pyranose form of the monosaccharide. The active site of a GH5 cellulase from *Bacillus agaradharens* (*Ba*Cel5A) interacts with the O3 of -1Glc through a histidine and tyrosine residue. In the active site of GH5-CBM6 there is a distinct lack of residues that interact with the O3 of the active site sugar. Therefore it is necessary for the -2NR* subsite to make additional productive binding interactions with the arabinose in order to stabilise the complex.

An additional arabinose is bound in the conserved sugar binding site of the CBM6 domain of the GH5-CBM6 structure. The CBM6 domain has been characterised previously and was found to have affinity for various oligosaccharides including the reaction products of the GH5 domain, recognising the terminal region of the oligosaccharides. In essence the CBM6 domain is fairly non-specific in the oligosaccharides it binds and can accommodate a number of different sugars including arabinose and xylose.

The GH5 family of enzymes includes a range of differing activities such as cellulases, mannosidases and xylanases. However, GH5-CBM6 from *C. thermocellum* is the first arabinoxylan specific xylanase reported. The previous study on GH5-CBM6 described the arabinoxylan specific activity, whereas this chapter illustrates how the specificity of the enzyme is driven by the binding of arabinose in the -2NR* subsite. Within the GH5 family there is a highly specific enzyme from *Paenibacillus pabuli*, named XG5, which is a xyloglucan-specific endoglucanse. XG5 also confers specificity for defined xyloglucan oligosaccharide by interactions made with side chain decorations of the glucan with the negative subsites of the substrate binding cleft of the enzyme (Gloster et al., 2007). In addition to the highly specific GH5 enzymes, xylanases specific for glucuronoxylan have been reported in the GH30 family. Xylanase A from *Erwinia chrysanthemi* is a glucrononxylan specific xylanase from GH30, which confers specificity through interactions made with the methyl glucuronic acid substitution of the xylose found at the -2 sub site (Urbanikova et al., 2011, Vrsanska et al., 2007). The GH5 and GH30 both belong to clan GH-A as they share similar structural folds and closely related in sequence and function.

Within a larger scope GH5-CBM6 is part of a large modular protein termed CtXyl5A which is recruited to the *C. thermocellum* cellulosome. The cellulosome is a large multi protein complex which targets the plant cell wall. Proteins are recruited to the large CipA scafoldin protein which is tethered to the bacterial cell wall. Within this complex various proteins are employed to degrade the plant cell wall. Therefore the unique arabinoxylan specific xylanase activity of GH5-CBM6 would complement the GH10 and GH11 xylanases and cellulases, which are also part of the cellulosome. This combination offers maximum degradation and utilization of the glycans available in the environment of the bacterium.

Chapter 7: Final Discussion

Chapters 3 and 4 describe the characterisation of a novel non-catalytic carbohydrate binding module (CBMX40) that binds levan and potentiates the activity of the catalytic domain of the endogenous exo-acting enzyme (SacC) against the β2,6-fructan levan, but not against fructooligosaccharides or inulin. The binding of CBMX40 to levan results in a marked increase in the activity of SacC against levan; this was observed in the full length enzyme and also *in trans* experiments where the catalytic domain and CBMX40 are not covalently attached. A fusion protein consisting of CBMX40 appended to a β–fructosidase from *Bacteroides thetaiotaomicron,* BT3082, showed a similar potentiation in activity. Indeed in *in trans* experiments BT3082 displayed an increase in levanase activity in the presence of CBMX40. BT 3082 is a non-specific exo-fructosidase capable of hydrolysing the β 2,1 linkage of inulin and the β 2,6 linkage of levan. The fusion of CBMX40 to BT3082 appears to change the non-specific enzyme into a glycoside hydrolase specific for levan. These data show that CBMX40 is capable of having a generic effect on non-specific β–fructosidases, converting these enzymes into specific exo-levanases.

Since CBMs were first discovered in the mid 1980s, their ability to disrupt substrate to aid hydrolysis by its catalytic counterpart has been proposed. The discovery of cellulose binding domains was heralded as the “missing” C_1 factor within cellulase systems. The C_1 factor was thought to change the complex nature of cellulose making it more accessible to the catalytic domain (Reese et al., 1950, Din et al., 1994). The ability of CBMs to affect the rate of hydrolysis *in trans* has previously been viewed as evidence for CBMs potentiating enzyme function through a substrate disruptive mechanism. Past studies on CBMs have shown ~0.5 fold potentiation of

catalytic activity when studied *in trans.* A study of a starch binding domain (SBD) described a 50% increase in activity when the catalytic domain of the native glucoamylase enzyme was incubated with insoluble starch in the presence of the SBD. The increase in hydrolytic rate was attributed to the SBD disrupting the structure of starch making it more accessible to the enzyme (Giardina et al., 2001). The *in trans* experiments in Chapter 3 and 4 show an enhancement in rate of hydrolysis upwards of 25-fold .Therefore the ability of CBMX40 to enhance the rate by such levels *in trans* seems to be indicative of a disruption effect on the levan substrate. To fully evaluate the possibility that CBMX40 works through such a mechanism, potential non-covalent interactions between CBMX40 and the catalytic domains were investigated. Analytical ultracentrifugation (AUC) revealed CBMX40 was able to form complexes with both the catalytic domain of SacC and BT3082. In contrast the starch binding domain reported no measurable interaction between the two proteins when measured by thermodynamics, while Dinn et al. (1994) did not consider possible non-covalent interactions between the cellulose binding domain and the cellulase catalytic domain to explain their *in trans* data. These studies, therefore, have not rigorously evaluated possible non-covalent interactions that may explain the observed, very modest, *in trans* potentiation of enzyme activity. Furthermore, both the α–amylase and cellulases are ~1000-fold more active against soluble substrates compared to crystalline polysaccharides. Thus, the 0.5-fold in increase in activity in the *in trans* experiments indicate that the CBMs are causing an extremely modest disruption of the crystalline substrate (Bolam et al., 1998, Din et al., 1994, Giardina et al., 2001) . To conclude, the disruptive effect of the starch and cellulose binding modules should be viewed with caution, and the ability of CBMX40 to disrupt the levan structure is unlikely.

The CBMX40 interacts with levan through the non-reducing end of the levan chain. The structure of CBMX40 in complex with levanbiose revealed that the protein module bound the non-reducing terminal fructose. Examples of CBMs that target glycans at the non-reducing termini include members of CBM32 that interact with terminal Gal and GalNAc residues. A GH84 exo-GlcNAcase (GH84C) contains a CBM32 that targets the terminal Gal-GlcNAc structures presented on the surface of red blood cells. A model of the GH84C from *Clostridium perfringes,* based on crystal structures and solution conformations of the protein, revealed that the active site of the catalytic module and the binding site of the CBM32 are arranged on the same side of the protein, facing in the same direction (Ficko-Blean and Boraston, 2006). This led to a model for CBM enhancement where there is simultaneous binding and hydrolysis of the terminal glycans presented on the cell surface by the CBM32 and the catalytic module, respectively. The potentiation by CBMX40 of the GH32 catalytic domain is highest against complex branched levans. This enhancement in enzyme activity is likely achieved through an “avidity” mechanism, in which the CBMX40 binds the terminal end of one branch, while the SacC catalytic domain hydrolyses the terminal of another branch of the same polysaccharide molecule. The combined binding of the CBMX40 and the SacC catalytic domain, compared to just the SacC catalytic domain, is much tighter, hence the “avidity” effect. These data may provide a generic model for the role of CBMs that bind the termini of glycan chains in potentiating the activity of exo-acting glycoside hydrolase enzymes against branched substrates.

Chapter 5 focuses on the characterisation of the GH76 enzymes encoded by the *B. thetaiotaomicron* genome, and α-mannan utilisation by two polysaccharide utilisation loci (PULs) that contain GH76 members. The majority of α-mannan encountered

within in the human gut is as part of N-glycans found on host and dietary glycoproteins. Studies of α-mannan degrading glycoside hydrolases including enzymes from GH38, GH47, GH92 and GH125 have mainly been studied in the context of N-glycan degradation (Hosokawa et al., 2001, Igdoura et al., 1999). A recent study of the 23 GH92 enzymes encoded by the *B. thetaiotaomicron* genome described the ability of a number of the GH92 members to hydrolyse high mannose N-glycans. In total 20 of the characterised GH92 enzymes are reported as active on yeast mannan but the natural substrate in the human gut environment is proposed to be N-glycans found on host and dietary glycoproteins (Zhu et al., 2010). However it would appear that the GH76 encoding PULs target yeast mannan rather than N-glycans as outlined below.

In vivo studies of *B. thetaiotaomicron* introduced into the gut of germ free mice reported induction of expression of PUL 36 and PUL 68, when the mice were fed a diet lacking exogenous α-mannan. This induction however was modest, suggesting that these PULs were not involved in N-glycan foraging in the gut. In vitro studies used yeast mannan as a mimic for N-glycans, when *B.thetaiotaomicron* was cultured on yeast mannan again expression of PUL 36 and PUL 68 was induced to a much higher degree (Bjursell et al., 2006). Yeast mannan appears to be a good mimic for mammalian N-glycan (mN-glycan) as it contains all the linkages (α1,6, α1,3 and α1,2) found in mN-glycans. The linkages however, are found in a different context in yeast mannan to mN-glycans. The α1,6 backbone present in both glycans is much more extended in yeast mannan where the α1,2 linked mannose is capped with a α1,3 linked mannose. These branches are reversed in mN-glycans with the α1,3 linked mannose being capped with α1,2 linked mannose. Recent *in vitro* studies of *B. thetaiotaomicron* cultured in the presence of both high mannose mN-glycan and

yeast mannan, separately, revealed PUL 36 and PUL 68 were only up regulated in the presence of yeast mannan and not high mannose mN-glycans, indicating the PULs were dedicated to yeast mannan acquisition and not mN-glycan utilization.

The activity assays in Chapter 5 revealed 7 of the 13 CAZy annotated glycoside hydrolases were active against yeast mannan. The four GH76 enzymes encoded in these PULs are endo α1,6 mannanases, which hydrolyse the yeast mannan backbone into various sized oligosaccharides. In each PUL a GH76 is located at the outer membrane and the second GH76 is located in the periplasmic space. The presence of an endo acting enzyme at the cell surface is typical of Sus like PULs. The three GH92 enzymes encoded within the PULs are all located in the periplasm and are capable of removing the α1,2 and α 1,3 linked mannose branches from the backbone. Both PULs also encode a predicted GH125 enzyme; while not currently characterised, they most probably hydryolyse α 1,6-manooligosaccharides released from the mannan backbone. PUL 68 also encodes a phosphatase that most probably removes phosphates from mannan oligosaccharides at the outer membrane. Along with the two GH125 enzymes there are also three other CAZy annotated enzymes (GH97, GH67 and GH38) that have not currently been characterised within the PULs. These enzymes are most probably involved in further processing of the yeast mannan polysaccharide. Although Chapter 5 describes a number of enzymes within the two PULs that contribute to yeast mannan hydrolysis, there is still a cohort of hypothetical proteins encoded in the PUL that may contribute to the degradation of the polysaccharide. For example, the GH76 enzymes only hydrolyse the yeast mannan backbone and not mannan with α1,2 and α1,3 linked mannose branches. Therefore there needs to be an enzyme at the cell surface capable of removing the branches to expose the α1,6 backbone for the GH76 enzyme. Analysis of the amino

acid sequence of these hypothetical proteins predicts that at least five of these are present at the cell surface and so are potential candidates for enzymes that remove mannan branches. Both PUL36 and PUL68 are regulated by a hybrid two component system (HTCS) and encode a SusC/SusD pair. Preliminary results of the sensor domain of the HTCS in PUL68 show that it is specific for yeast mannan as it binds oligosaccharides generated by digestion of mannan generated by the yeast mutant mnn2 by a GH76 enzyme. The results presented in Chapter 5 provide evidence that *B. thetaiotaomicron* dedicates at least two PULs to yeast mannan acquisition and utilisation.

Chapter 6 probes the mechanism by which GH5-CBM6 from *Clostridium thermocellum* confers specificity for arabinoxylan. Structural and mutational studies revealed that the binding of arabinose to the side pocket of the active site is required for hydrolysis of the xylan backbone. GH5-CBM6 is part of a large multi modular protein, CtXyl5A, which is recruited to the cellulosome of *C. thermocellum* (Correia et al., 2011*).* The cellulosome is a large multi protein complex that targets the plant cell wall. Proteins are recruited to the large CipA scafoldin protein which is tethered to the bacterial cell wall (Gilbert, 2007). CtXyl5A consists of an N-terminal GH5 arabinoxylan specific xylanase, CBM6, CBM13, CBM62 and a type I dockerin at the C-terminal (Bras et al., 2011). The type I dockerin interacts with the CipA protein through a cohesin module to tether CtXyl5A to the cellulosome complex. CBM6 displays affinity for cellohexasoe and cellobiose along with GH5 reaction products and xylooligosaccharides, binding to the termini of these oligosaccharides. CBM6 most probably targets the enzyme to areas of the plant cell wall that are undergoing active degradation, similar to the predicted role of a CBM9 in a GH10 xylanase. CBM62 binds galactose and arabinopyranose through interactions at the O4,

distinguishing these sugars from other netural monosaccharides. The binding of CBM62 to galactose containing polysaccharides suggests that it may target the enzyme to arabinoxylans with galactose decorations such as corn stem arabinoxylan, which contains a terminal galactose. Alternatively, the CBM62 may bind to polysaccharides found in close proximity to regions of arabinoxylan in the plant cell wall. The cellulosome recruits a number of plant cell wall degrading enzymes including cellulases, GH10 and GH 11 xylanases. The unique arabinoxylan specific xylanase offers a complementary activity to other enzymes recruited to the cellulosome.

The complex nature of polysaccharides results in various forms of specificity at different levels of carbohydrate utilisation. Bacteria like *B.thetaiotaomicron* sense and utilise complex polysaccharides through genes encoded by PULs. Modular proteins can utilise each domain to drive specificity for a polysaccharide through the use of non-catalytic binding modules. Finally, glycoside hydrolases evolve subtle changes in the architecture of the active site and amino acid interactions with individual sugars of polysaccharides to confer specificity.

7.1 Future work

CBMX40 represents a new CAZy CBM family and as such future work to study other members of the family would determine if the mechanism and mode of binding is generic to the family or unique to CBMX40. The study of other CBMX40 containing GH32 family enzymes could reveal a subgroup of enzymes where catalytic activity is significantly potentiated by the presence of a CBMX40. It may also reveal a subset of levan binding CBMX40s.

To further dissect the ability of *B. thetaiotaomicron* PULs to deconstruct yeast mannan a more complete description of the enzyme activities found within the regulated PULs is necessary. This could be achieved by characterisation of all CAZy annotated encoded proteins and also by the study of hypothetical proteins which may have the potential for novel catalytic activities. To understand the importance of yeast mannan utilisation to *B.thetaiotaomicron's* survival, knockouts of the mannan PULs could be used to assess the bacterium's ability to grow on complex mannans.

To further investigate the specificity of CtXyl5A kinetic studies against a range of arabinoxylans from different sources would reveal the specificity of the enzyme and the substrate targeted.

References

ALBERTO, F., BIGNON, C., SULZENBACHER, G., HENRISSAT, B. & CZJZEK, M. 2004. The three-dimensional structure of invertase (beta-fructosidase) from Thermotoga maritima reveals a bimodular arrangement and an evolutionary relationship between retaining and inverting glycosidases. *J Biol Chem,* 279**,** 18903-10.

ALTENBACH, D., NUESCH, E., MEYER, A. D., BOLLER, T. & WIEMKEN, A. 2004. The large subunit determines catalytic specificity of barley sucrose:fructan 6-fructosyltransferase and fescue sucrose:sucrose 1-fructosyltransferase. *FEBS Lett,* 567**,** 214-8.

ALTSCHUL, S., MADDEN, T., SCHAFFER, A., ZHANG, J., ZHANG, Z., MILLER, W. & LIPMAN, D. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res.,* 25**,** 3389-3402.

BAURHOO, B., PHILLIP, L. & RUIZ-FERIA, C. A. 2007. Effects of purified lignin and mannan oligosaccharides on intestinal integrity and microbial populations in the ceca and litter of broiler chickens. *Poult Sci,* 86**,** 1070-8.

BEGUIN, P. & AUBERT, J. P. 1994. The biological degradation of cellulose. *FEMS Microbiol Rev,* 13**,** 25-58.

BENDTSEN, J. D., NIELSEN, H., VON HEIJNE, G. & BRUNAK, S. 2004. Improved prediction of signal peptides: SignalP 3.0. *J Mol Biol,* 340**,** 783-95.

BJURSELL, M. K., MARTENS, E. C. & GORDON, J. I. 2006. Functional genomic and metabolic studies of the adaptations of a prominent adult human gut symbiont, Bacteroides thetaiotaomicron, to the suckling period. *J Biol Chem,* 281**,** 36269-79.

BLAKE, J. D., CLARKE, M. L., JANSSON, P. E. & MCNEIL, K. E. 1982. Fructan from Erwinia herbicola. *J Bacteriol,* 151**,** 1595-7.

BOLAM, D. N., CIRUELA, A., MCQUEEN-MASON, S., SIMPSON, P., WILLIAMSON, M. P., RIXON, J. E., BORASTON, A., HAZLEWOOD, G. P. & GILBERT, H. J. 1998. Pseudomonas cellulose-binding domains mediate their effects by increasing enzyme substrate proximity. *Biochem J,* 331 (Pt 3)**,** 775-81.

BONAY, P. & HUGHES, R. C. 1991. Purification and characterization of a novel broad-specificity (alpha 1----2, alpha 1----3 and alpha 1----6) mannosidase from rat liver. *Eur J Biochem,* 197**,** 229-38.

BORASTON, A. B., BOLAM, D. N., GILBERT, H. J. & DAVIES, G. J. 2004. Carbohydrate-binding modules: fine-tuning polysaccharide recognition. *Biochem J,* 382**,** 769-81.

BRAS, J. L., CORREIA, M. A., ROMAO, M. J., PRATES, J. A., FONTES, C. M. & NAJMUDIN, S. 2011. Purification, crystallization and preliminary X-ray characterization of the pentamodular arabinoxylanase CtXyl5A from Clostridium thermocellum. *Acta Crystallogr Sect F Struct Biol Cryst Commun,* 67**,** 833-6.

BRY, L., FALK, P. G., MIDTVEDT, T. & GORDON, J. I. 1996. A model of host-microbial interactions in an open mammalian ecosystem. *Science,* 273**,** 1380-3.

CAMIRAND, A., HEYSEN, A., GRONDIN, B. & HERSCOVICS, A. 1991. Glycoprotein biosynthesis in Saccharomyces cerevisiae. Isolation and

characterization of the gene encoding a specific processing alpha-mannosidase. *J Biol Chem,* 266**,** 15120-7.
CANTAREL, B. L., COUTINHO, P. M., RANCUREL, C., BERNARD, T., LOMBARD, V. & HENRISSAT, B. 2009. The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. *Nucleic Acids Res,* 37**,** D233-8.
CARPITA, N. C. 1997. Structure and Biosynthesis of Plant Cell walls *In:* DENNIS, D. T., LEFEBURE, D. D. & D.B., L. (eds.) *Plant metabolism* Harlow: Wesley Longman
CENCI DI BELLO, I., DORLING, P. & WINCHESTER, B. 1983. The storage products in genetic and swainsonine-induced human mannosidosis. *Biochem J,* 215**,** 693-6.
CHARNOCK, S. J., SPURWAY, T. D., XIE, H., BEYLOT, M. H., VIRDEN, R., WARREN, R. A., HAZLEWOOD, G. P. & GILBERT, H. J. 1998. The topology of the substrate binding clefts of glycosyl hydrolase family 10 xylanases are not conserved. *J Biol Chem,* 273**,** 32187-99.
CHUI, D., OH-EDA, M., LIAO, Y. F., PANNEERSELVAM, K., LAL, A., MAREK, K. W., FREEZE, H. H., MOREMEN, K. W., FUKUDA, M. N. & MARTH, J. D. 1997. Alpha-mannosidase-II deficiency results in dyserythropoiesis and unveils an alternate pathway in oligosaccharide biosynthesis. *Cell,* 90**,** 157-67.
COHEN, S. N., CHANG, A. C. Y. & HSU, L. 1972. Nonchromosomal antibiotic resistance in bacteria: Genetic transformation of *Escherichia coli* by R-factor DNA. *Proc Natl Acad Sci U S A,* 69**,** 2110-2114.
CORREIA, M. A., MAZUMDER, K., BRAS, J. L., FIRBANK, S. J., ZHU, Y., LEWIS, R. J., YORK, W. S., FONTES, C. M. & GILBERT, H. J. 2011. Structure and function of an arabinoxylan-specific xylanse. *J Biol Chem,* 286**,** 22510-20.
D'ELIA, J. N. & SALYERS, A. A. 1996a. Contribution of a neopullulanase, a pullulanase, and an alpha-glucosidase to growth of Bacteroides thetaiotaomicron on starch. *J Bacteriol,* 178**,** 7173-9.
D'ELIA, J. N. & SALYERS, A. A. 1996b. Effect of regulatory protein levels on utilization of starch by Bacteroides thetaiotaomicron. *J Bacteriol,* 178**,** 7180-6.
DANIEL, P. F., WINCHESTER, B. & WARREN, C. D. 1994. Mammalian alpha-mannosidases--multiple forms but a common purpose? *Glycobiology,* 4**,** 551-66.
DAVIES, G. & HENRISSAT, B. 1995. Structures and mechanisms of glycosyl hydrolases. *Structure,* 3**,** 853-9.
DAVIES, G. J., WILSON, K. S. & HENRISSAT, B. 1997. Nomenclature for sugar-binding subsites in glycosyl hydrolases. *Biochem J,* 321 (Pt 2)**,** 557-9.
DEONDER, R. 1966. Levansucrase from *Bacillus subtilis*. *Methods Enzymology* 8**,** 500-505.
DIN, N., DAMUDE, H. G., GILKES, N. R., MILLER, R. C., JR., WARREN, R. A. & KILBURN, D. G. 1994. C1-Cx revisited: intramolecular synergism in a cellulase. *Proc Natl Acad Sci U S A,* 91**,** 11383-7.
EBRINGEROVA, A., HROMADKOVA, Z. & HEINZE, T. 2005. Hemicellulose. *Advanced Polymer Science* 186**,** 1-67.
EDELMAN, J. 2006. The Formation of Oligosaccharides by Enzymic Transglycosylation *In:* NORD, F. F. (ed.) *Advances in Enzymology and Related areas of Molecular Biology* Hoboken , NJ, USA: John Wiley and sons Inc.

EMAMI, K., TOPAKAS, E., NAGY, T., HENSHAW, J., JACKSON, K. A., NELSON, K. E., MONGODIN, E. F., MURRAY, J. W., LEWIS, R. J. & GILBERT, H. J. 2009. Regulation of the xylan-degrading apparatus of Cellvibrio japonicus by a novel two-component system. *J Biol Chem,* 284**,** 1086-96.

FICKO-BLEAN, E. & BORASTON, A. B. 2006. The interaction of a carbohydrate-binding module from a Clostridium perfringens N-acetyl-beta-hexosaminidase with its carbohydrate receptor. *J Biol Chem,* 281**,** 37748-57.

FRANCK, A. & DE LEENHEER, L. 2005. Inulin. *Biopolymers online*

FRENCH, A. D. 1988. Accessible Conformation of the Beta-D- (2-1)- and (2-6) Linked d-Fructans Inulin and Levan *Carbohydrate Research,* 176**,** 17-29.

FRENCH, A. D. 1989. Chemical and Physical-Properties of Fructans. *Journal of Plant Physiology* 134**,** 125-136.

FUCHS, A. 1956. Synthesis of levan by pseudomonads. *Nature,* 178**,** 921.

FUCHS, A. 1987. Potentials for Non-Food Utilization of Fructose and Inulin. *Starch-Strake,* 39**,** 335-343.

GIARDINA, T., GUNNING, A. P., JUGE, N., FAULDS, C. B., FURNISS, C. S., SVENSSON, B., MORRIS, V. J. & WILLIAMSON, G. 2001. Both binding sites of the starch-binding domain of Aspergillus niger glucoamylase are essential for inducing a conformational change in amylose. *J Mol Biol,* 313**,** 1149-59.

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

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

GILL, S. C. & VON HIPPEL, P. H. 1989. Calculation of protein extinction coefficients from amino acid sequence data. *Anal Biochem,* 182**,** 319-26.

GLOSTER, T. M., IBATULLIN, F. M., MACAULEY, K., EKLOF, J. M., ROBERTS, S., TURKENBURG, J. P., BJORNVAD, M. E., JORGENSEN, P. L., DANIELSEN, S., JOHANSEN, K. S., BORCHERT, T. V., WILSON, K. S., BRUMER, H. & DAVIES, G. J. 2007. Characterization and three-dimensional structures of two distinct bacterial xyloglucanases from families GH5 and GH12. *J Biol Chem,* 282**,** 19177-89.

GONZALEZ, D. S., KARAVEG, K., VANDERSALL-NAIRN, A. S., LAL, A. & MOREMEN, K. W. 1999. Identification, expression, and characterization of a cDNA encoding human endoplasmic reticulum mannosidase I, the enzyme that catalyzes the first mannose trimming step in mammalian Asn-linked oligosaccharide biosynthesis. *J Biol Chem,* 274**,** 21375-86.

GOTODA, Y., WAKAMATSU, N., KAWAI, H., NISHIDA, Y. & MATSUMOTO, T. 1998. Missense and nonsense mutations in the lysosomal alpha-mannosidase gene (MANB) in severe and mild forms of alpha-mannosidosis. *Am J Hum Genet,* 63**,** 1015-24.

HAN, Y. W. 1990. Microbial levan. *Adv Appl Microbiol,* 35**,** 171-94.

HAN, Y. W. & CLARKE, M. A. 1990. Production and Characterisation of Microbial Levan *Journal of Agriculture and Food Chemistry* 38**,** 393-396.

HARRIS, G. W., JENKINS, J. A., CONNERTON, I., CUMMINGS, N., LO LEGGIO, L., SCOTT, M., HAZLEWOOD, G. P., LAURIE, J. I., GILBERT, H. J. & PICKERSGILL, R. W. 1994. Structure of the catalytic core of the family F xylanase from Pseudomonas fluorescens and identification of the xylopentaose-binding sites. *Structure,* 2**,** 1107-16.

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

HENDRY, G. F. 1993. Evolutionary origins and natural functions of fructans-a climatological, biogeographic and mechanistic appraisal *New Phytologist,* 123**,** 3-14.
HENRISSAT, B. & DAVIES, G. J. 1997. Structural and sequence based classification of glycoside hydrolase *Current opinion Structural Biology* 7**,** 637-644.
HERNANDEZ, L. M., BALLOU, L., ALVARADO, E., GILLECE-CASTRO, B. L., BURLINGAME, A. L. & BALLOU, C. E. 1989a. A new Saccharomyces cerevisiae mnn mutant N-linked oligosaccharide structure. *J Biol Chem,* 264**,** 11849-56.
HERNANDEZ, L. M., BALLOU, L., ALVARADO, E., TSAI, P. K. & BALLOU, C. E. 1989b. Structure of the phosphorylated N-linked oligosaccharides from the mnn9 and mnn10 mutants of Saccharomyces cerevisiae. *J Biol Chem,* 264**,** 13648-59.
HERSCOVICS, A. 2001. Structure and function of Class I alpha 1,2-mannosidases involved in glycoprotein synthesis and endoplasmic reticulum quality control. *Biochimie,* 83**,** 757-62.
HIGUCHI, M., IWAMI, Y., YAMADA, T. & ARAYA, S. 1970. Levan synthesis and accumulation by human dental plaque. *Arch Oral Biol,* 15**,** 563-7.
HOFFMAN, R. A., KAMERLING, J. P. & VLIEGENTHART, J. F. 1992. Structural features of a water-soluble arabinoxylan from the endosperm of wheat. *Carbohydr Res,* 226**,** 303-11.
HOOPER, L. V., STAPPENBECK, T. S., HONG, C. V. & GORDON, J. I. 2003. Angiogenins: a new class of microbicidal proteins involved in innate immunity. *Nat Immunol,* 4**,** 269-73.
HOOPER, L. V., XU, J., FALK, P. G., MIDTVEDT, T. & GORDON, J. I. 1999. A molecular sensor that allows a gut commensal to control its nutrient foundation in a competitive ecosystem. *Proc Natl Acad Sci U S A,* 96**,** 9833-8.
HOSOKAWA, N., WADA, I., HASEGAWA, K., YORIHUZI, T., TREMBLAY, L. O., HERSCOVICS, A. & NAGATA, K. 2001. A novel ER alpha-mannosidase-like protein accelerates ER-associated degradation. *EMBO Rep,* 2**,** 415-22.
HOSOKAWA, N., WADA, I., NATSUKA, Y. & NAGATA, K. 2006. EDEM accelerates ERAD by preventing aberrant dimer formation of misfolded alpha1-antitrypsin. *Genes Cells,* 11**,** 465-76.
IGDOURA, S. A., HERSCOVICS, A., LAL, A., MOREMEN, K. W., MORALES, C. R. & HERMO, L. 1999. Alpha-mannosidases involved in N-glycan processing show cell specificity and distinct subcompartmentalization within the Golgi apparatus of cells in the testis and epididymis. *Eur J Cell Biol,* 78**,** 441-52.
INGELMAN, B. & SIEGBAHN, K. 1944. Dextran and levan molecules studied with the electron microscope. *Nature* 3903**,** 237-238.
JAKOB, C. A., BURDA, P., ROTH, J. & AEBI, M. 1998. Degradation of misfolded endoplasmic reticulum glycoproteins in Saccharomyces cerevisiae is determined by a specific oligosaccharide structure. *J Cell Biol,* 142**,** 1223-33.
KOJIMA, I., SAITO, T., IIZUKA, M., MINAMIURA, N. & ONO, S. 1993. Characterisation of Levan produced by *Serratia sp. Journal of Fermentation and Bioengineering,* 75**,** 9-12.
KOROPATKIN, N. M., MARTENS, E. C., GORDON, J. I. & SMITH, T. J. 2008. Starch catabolism by a prominent human gut symbiont is directed by the recognition of amylose helices. *Structure,* 16**,** 1105-15.

KUKURUZINSKA, M. A., BERGH, M. L. & JACKSON, B. J. 1987. Protein glycosylation in yeast. *Annu Rev Biochem,* 56**,** 915-44.
LAEMMLI, U. K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature,* 227**,** 680-685.
LAL, A., SCHUTZBACH, J. S., FORSEE, W. T., NEAME, P. J. & MOREMEN, K. W. 1994. Isolation and expression of murine and rabbit cDNAs encoding an alpha 1,2-mannosidase involved in the processing of asparagine-linked oligosaccharides. *J Biol Chem,* 269**,** 9872-81.
LAMMENS, W., LE ROY, K., SCHROEVEN, L., VAN LAERE, A., RABIJNS, A. & VAN DEN ENDE, W. 2009. Structural insights into glycoside hydrolase family 32 and 68 enzymes: functional implications. *J Exp Bot,* 60**,** 727-40.
LAWSON, S. L., WAKARCHUK, W. W. & WITHERS, S. G. 1996. Effects of both shortening and lengthening the active site nucleophile of Bacillus circulans xylanase on catalytic activity. *Biochemistry,* 35**,** 10110-8.
LE ROY, K., LAMMENS, W., VERHAEST, M., DE CONINCK, B., RABIJNS, A., VAN LAERE, A. & VAN DEN ENDE, W. 2007. Unraveling the difference between invertases and fructan exohydrolases: a single amino acid (Asp-239) substitution transforms Arabidopsis cell wall invertase1 into a fructan 1-exohydrolase. *Plant Physiol,* 145**,** 616-25.
LEE, J. H., KIM, K. N. & CHOI, Y. J. 2004. Identification and characterization of a novel inulin binding module (IBM) from the CFTase of Bacillus macerans CFC1. *FEMS Microbiol Lett,* 234**,** 105-10.
LEWIS, M. S. & BALLOU, C. E. 1991. Separation and characterization of two alpha 1,2-mannosyltransferase activities from Saccharomyces cerevisiae. *J Biol Chem,* 266**,** 8255-61.
LIAO, Y. F., LAL, A. & MOREMEN, K. W. 1996. Cloning, expression, purification, and characterization of the human broad specificity lysosomal acid alpha-mannosidase. *J Biol Chem,* 271**,** 28348-58.
LIU, Y., CHOUDHURY, P., CABRAL, C. M. & SIFERS, R. N. 1999. Oligosaccharide modification in the early secretory pathway directs the selection of a misfolded glycoprotein for degradation by the proteasome. *J Biol Chem,* 274**,** 5861-7.
LOPEZ-MOLINA, D., NAVARRO-MARTINEZ, M. D., ROJAS MELGAREJO, F., HINER, A. N., CHAZARRA, S. & RODRIGUEZ-LOPEZ, J. N. 2005. Molecular properties and prebiotic effect of inulin obtained from artichoke (Cynara scolymus L.). *Phytochemistry,* 66**,** 1476-84.
LOWE, E. C., BASLE, A., CZJZEK, M., FIRBANK, S. J. & BOLAM, D. N. 2012. A scissor blade-like closing mechanism implicated in transmembrane signaling in a Bacteroides hybrid two-component system. *Proc Natl Acad Sci U S A,* 109**,** 7298-303.
LU, Z. X., WALKER, K. Z., MUIR, J. G. & O'DEA, K. 2004. Arabinoxylan fibre improves metabolic control in people with Type II diabetes. *Eur J Clin Nutr,* 58**,** 621-8.
MACKIE, R. I., SGHIR, A. & GASKINS, H. R. 1999. Developmental microbial ecology of the neonatal gastrointestinal tract. *Am J Clin Nutr,* 69**,** 1035S-1045S.
MACLEOD, A. M., LINDHORST, T., WITHERS, S. G. & WARREN, R. A. 1994. The acid/base catalyst in the exoglucanase/xylanase from Cellulomonas fimi is glutamic acid 127: evidence from detailed kinetic studies of mutants. *Biochemistry,* 33**,** 6371-6.

MARTENS, E. C., CHIANG, H. C. & GORDON, J. I. 2008. Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. *Cell Host Microbe,* 4**,** 447-57.
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. & GORDON, J. I. 2011. Recognition and degradation of plant cell wall polysaccharides by two human gut symbionts. *PLoS Biol,* 9**,** e1001221.
MENG, G. Y. & FUTTERER, K. 2003. Structural Framework of Fructosyl Transfer in *Bacillus subtillis* Levansucrase. *Nature Structural Biology* 10**,** 935-941.
MIAO, S., ZISER, L., AEBERSOLD, R. & WITHERS, S. G. 1994. Identification of glutamic acid 78 as the active site nucleophile in Bacillus subtilis xylanase using electrospray tandem mass spectrometry. *Biochemistry,* 33**,** 7027-32.
MIYAZAKI, K., MIYAMOTO, H., MERCER, D. K., HIRASE, T., MARTIN, J. C., KOJIMA, Y. & FLINT, H. J. 2003. Involvement of the multidomain regulatory protein XynR in positive control of xylanase gene expression in the ruminal anaerobe Prevotella bryantii B(1)4. *J Bacteriol,* 185**,** 2219-26.
MOORE, W. E. & HOLDEMAN, L. V. 1974. Human fecal flora: the normal flora of 20 Japanese-Hawaiians. *Appl Microbiol,* 27**,** 961-79.
MOREMEN, K. W., TOUSTER, O. & ROBBINS, P. W. 1991. Novel purification of the catalytic domain of Golgi alpha-mannosidase II. Characterization and comparison with the intact enzyme. *J Biol Chem,* 266**,** 16876-85.
MULLIS, K. B. & FALOONA, F. A. 1987. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol,* 155**,** 335-50.
NAKAJIMA, T. & BALLOU, C. E. 1974. Structure of the linkage region between the polysaccharide and protein parts of Saccharomyces cerevisiae mannan. *J Biol Chem,* 249**,** 7685-94.
NAKAJIMA, T. & BALLOU, C. E. 1975. Yeast manno-protein biosynthesis: solubilization and selective assay of four mannosyltransferases. *Proc Natl Acad Sci U S A,* 72**,** 3912-6.
NAKAJIMA, T., MAITRA, S. K. & BALLOU, C. E. 1976. An endo-alpha1 leads to 6-D-mannanase from a soil bacterium. Purification, properties, and mode of action. *J Biol Chem,* 251**,** 174-81.
NUMAO, S., KUNTZ, D. A., WITHERS, S. G. & ROSE, D. R. 2003. Insights into the mechanism of Drosophila melanogaster Golgi alpha-mannosidase II through the structural analysis of covalent reaction intermediates. *J Biol Chem,* 278**,** 48074-83.
OLIVARI, S., CALI, T., SALO, K. E., PAGANETTI, P., RUDDOCK, L. W. & MOLINARI, M. 2006. EDEM1 regulates ER-associated degradation by accelerating de-mannosylation of folding-defective polypeptides and by inhibiting their covalent aggregation. *Biochem Biophys Res Commun,* 349**,** 1278-84.
OZIMEK, L. K., KRALJ, S., VAN DER MAAREL, M. J. & DIJKHUIZEN, L. 2006. The levansucrase and inulosucrase enzymes of Lactobacillus reuteri 121 catalyse processive and non-processive transglycosylation reactions. *Microbiology,* 152**,** 1187-96.
PACE, C. N., VAJDOS, F., FEE, L., GRIMSLEY, G. & GRAY, T. 1995. How to measure and predict the molar absorption coefficient of a protein. *Protein Sci,* 4**,** 2411-23.
PARK, C., MENG, L., STANTON, L. H., COLLINS, R. E., MAST, S. W., YI, X., STRACHAN, H. & MOREMEN, K. W. 2005. Characterization of a human

core-specific lysosomal {alpha}1,6-mannosidase involved in N-glycan catabolism. *J Biol Chem,* 280**,** 37204-16.
PATTERSON, J. A. & BURKHOLDER, K. M. 2003. Application of prebiotics and probiotics in poultry production. *Poult Sci,* 82**,** 627-31.
PONTIS, H. G. & DEL CAMPILLO, E. 1985. Biochemistry of Storage Carbohydrates in Green Plants. *In:* DEY, P. M. & DIXO, A. R. (eds.). Academic Press, New York
REESE, E. T., SIU, R. G. & LEVINSON, H. S. 1950. The biological degradation of soluble cellulose derivatives and its relationship to the mechanism of cellulose hydrolysis. *J Bacteriol,* 59**,** 485-97.
REEVES, A. R., D'ELIA, J. N., FRIAS, J. & SALYERS, A. A. 1996. A Bacteroides thetaiotaomicron outer membrane protein that is essential for utilization of maltooligosaccharides and starch. *J Bacteriol,* 178**,** 823-30.
REEVES, A. R., WANG, G. R. & SALYERS, A. A. 1997. Characterization of four outer membrane proteins that play a role in utilization of starch by Bacteroides thetaiotaomicron. *J Bacteriol,* 179**,** 643-9.
ROBERFROID, M. B. 2005. Introducing inulin-type fructans. *Br J Nutr,* 93 Suppl 1**,** S13-25.
ROMERO, P. A. & HERSCOVICS, A. 1989. Glycoprotein biosynthesis in Saccharomyces cerevisiae. Characterization of alpha-1,6-mannosyltransferase which initiates outer chain formation. *J Biol Chem,* 264**,** 1946-50.
RYE, C. S. & WITHERS, S. G. 2000. Glycosidase Mechanisms. *Current Opinion in Chemical Biology,* 4**,** 573-580.
SAMBROOK, J., FRITSCH, E. F. & MANIATIS, T. 1989. *Molecular cloning; a laboratory manaual* New York, Cold Springs Harbour laboratory press.
SENDID, B., QUINTON, J. F., CHARRIER, G., GOULET, O., CORTOT, A., GRANDBASTIEN, B., POULAIN, D. & COLOMBEL, J. F. 1998. Anti-Saccharomyces cerevisiae mannan antibodies in familial Crohn's disease. *Am J Gastroenterol,* 93**,** 1306-10.
SHIPMAN, J. A., BERLEMAN, J. E. & SALYERS, A. A. 2000. Characterization of four outer membrane proteins involved in binding starch to the cell surface of Bacteroides thetaiotaomicron. *J Bacteriol,* 182**,** 5365-72.
SHIPMAN, J. A., CHO, K. H., SIEGEL, H. A. & SALYERS, A. A. 1999. Physiological characterization of SusG, an outer membrane protein essential for starch utilization by Bacteroides thetaiotaomicron. *J Bacteriol,* 181**,** 7206-11.
SINNOTT, M. L. 1990. Catalytic mehcanism of enzyme glycosyl transfer *Chem Rev* 1171-1202.
SONNENBURG, E. D., SONNENBURG, J. L., MANCHESTER, J. K., HANSEN, E. E., CHIANG, H. C. & GORDON, J. I. 2006. A hybrid two-component system protein of a prominent human gut symbiont couples glycan sensing in vivo to carbohydrate metabolism. *Proc Natl Acad Sci U S A,* 103**,** 8834-9.
SONNENBURG, E. D., ZHENG, H., JOGLEKAR, P., HIGGINBOTTOM, S. K., FIRBANK, S. J., BOLAM, D. N. & SONNENBURG, J. L. 2010. Specificity of polysaccharide use in intestinal bacteroides species determines diet-induced microbiota alterations. *Cell,* 141**,** 1241-52.
STAPPENBECK, T. S., HOOPER, L. V. & GORDON, J. I. 2002. Developmental regulation of intestinal angiogenesis by indigenous microbes via Paneth cells. *Proc Natl Acad Sci U S A,* 99**,** 15451-5.

STEVENS, C. V., MERIGGI, A. & BOOTEN, K. 2001. Chemical modification of inulin, a valuable renewable resource, and its industrial applications. *Biomacromolecules,* 2**,** 1-16.
STOCK, A. M., ROBINSON, V. L. & GOUDREAU, P. N. 2000. Two-component signal transduction. *Annu Rev Biochem,* 69**,** 183-215.
STUDIER, F. W. & MOFFATT, B. A. 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J Mol Biol,* 189**,** 113-30.
Berg, J.M.T., John L.; and Stryer, Lubert., *Biochemistry* Fifth ed. 2002: W.H. Freeman and company
SZABO, L., JAMAL, S., XIE, H., CHARNOCK, S. J., BOLAM, D. N., GILBERT, H. J. & DAVIES, G. J. 2001. Structure of a family 15 carbohydrate-binding module in complex with xylopentaose. Evidence that xylan binds in an approximate 3-fold helical conformation. *J Biol Chem,* 276**,** 49061-5.
TANAKA, T., OI, S. & YAMAMOTO, T. 1980. The molecular structure of low and high molecular weight levans synthesized by levansucrase. *J Biochem,* 87**,** 297-303.
TOMME, P., VAN TILBEURGH, H., PETTERSSON, G., VAN DAMME, J., VANDEKERCKHOVE, J., KNOWLES, J., TEERI, T. & CLAEYSSENS, M. 1988. Studies of the cellulolytic system of Trichoderma reesei QM 9414. Analysis of domain function in two cellobiohydrolases by limited proteolysis. *Eur J Biochem,* 170**,** 575-81.
TREMBLAY, L. O., CAMPBELL DYKE, N. & HERSCOVICS, A. 1998. Molecular cloning, chromosomal mapping and tissue-specific expression of a novel human alpha1,2-mannosidase gene involved in N-glycan maturation. *Glycobiology,* 8**,** 585-95.
TREMBLAY, L. O. & HERSCOVICS, A. 2000. Characterization of a cDNA encoding a novel human Golgi alpha 1, 2-mannosidase (IC) involved in N-glycan biosynthesis. *J Biol Chem,* 275**,** 31655-60.
TULL, D., WITHERS, S. G., GILKES, N. R., KILBURN, D. G., WARREN, R. A. & AEBERSOLD, R. 1991. Glutamic acid 274 is the nucleophile in the active site of a "retaining" exoglucanase from Cellulomonas fimi. *J Biol Chem,* 266**,** 15621-5.
URBANIKOVA, L., VRSANSKA, M., MORKEBERG KROGH, K. B., HOFF, T. & BIELY, P. 2011. Structural basis for substrate recognition by Erwinia chrysanthemi GH30 glucuronoxylanase. *FEBS J,* 278**,** 2105-16.
VAN DEN ELSEN, J. M., KUNTZ, D. A. & ROSE, D. R. 2001. Structure of Golgi alpha-mannosidase II: a target for inhibition of growth and metastasis of cancer cells. *EMBO J,* 20**,** 3008-17.
VERVOORT, L., VAN DEN MOOTER, G., AUGUSTIJNS, P., BUSSON, R., TOPPET, S. & KINGET, R. 1997. Inulin hydrogels as carriers for colonic drug targeting: I. Synthesis and characterization of methacrylated inulin and hydrogel formation. *Pharm Res,* 14**,** 1730-7.
VRSANSKA, M., KOLENOVA, K., PUCHART, V. & BIELY, P. 2007. Mode of action of glycoside hydrolase family 5 glucuronoxylan xylanohydrolase from Erwinia chrysanthemi. *FEBS J,* 274**,** 1666-77.
WANKER, E., HUBER, A. & SCHWAB, H. 1995. Purification and characterization of the Bacillus subtilis levanase produced in Escherichia coli. *Appl Environ Microbiol,* 61**,** 1953-8.

WHITE, L. A., NEWMAN, M. C., CROMWELL, G. L. & LINDEMANN, M. D. 2002. Brewers dried yeast as a source of mannan oligosaccharides for weanling pigs. *Journal of Animal Science* 80**,** 2619-2628.

WINTERHALTER, C., HEINRICH, P., CANDUSSIO, A., WICH, G. & LIEBL, W. 1995. Identification of a novel cellulose-binding domain within the multidomain 120 kDa xylanase XynA of the hyperthermophilic bacterium Thermotoga maritima. *Mol Microbiol,* 15**,** 431-44.

WOSTEN, M. M. 1998. Eubacterial sigma-factors. *FEMS Microbiol Rev,* 22**,** 127-50.

XU, J. & GORDON, J. I. 2003. Honor thy symbionts. *Proc Natl Acad Sci U S A,* 100**,** 10452-9.

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

ZHAO, P. Y., JUNG, J. H. & KIM, I. H. 2012. Effect of mannan oligosaccharides and fructan on growth performance, nutrient digestibility, blood profile, and diarrhea score in weanling pigs. *J Anim Sci,* 90**,** 833-9.

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. & DAVIES, G. J. 2010. Mechanistic insights into a Ca2+-dependent family of alpha-mannosidases in a human gut symbiont. *Nat Chem Biol,* 6**,** 125-32.

Appendix A

A1 Chemicals

Amersham-Boehringer Mannheim

2’-Deoxyadenosine 5’-triphosphate (dATP)

2’-Deoxycytidine 5’-triphosphate (dCTP)

2’-Deoxyguanosine 5’-triphosphate (dGTP)

2’-Deoxythymidine 5’-triphosphate (dTTP)

BioGene

Electrophoresis grade Agarose

British Drug Houses (BDH)

Acetic acid (Glacial)

Acrylamide solution (40% (w/v); Electran)

Boric acid

Bromophenol Blue

Citric acid

Calcium chloride

Chloroform

Dimethylformamide

Ethanol (industrial grade)

Hydrochloric acid

Isopropanol

Magnesium chloride

Magnesium sulphate

Mineral oil

Methanol

Polyethelene glycol MW 8000 (PEG-8000)

Polyethelene glycol MW 550 monomethyl ether (PEG 550 mme)

Potassium dihydrogen orthophosphate

Sodium acetate

Sodium chloride

Sulphuric acid

Triton X-100

Fisons

46/48% (w/v) NaOH

Sodium acetate trihydrate

James Burrough (F.A.D.) Ltd

Ethanol

Megazyme

Rye Arabinoxylan (low viscosity)

Wheat Arabionoxylan (medium viscosity)

Inulin From Chicory

Melford Laboratories

Isopropyl-β-D-thiogalactoside (IPTG)

CAPS

HEPES

GE Health Care

Adenosine 5’-triphosphate (ATP)

Agarose (ultrapure)

Sigma

2,4-Dinitrophenol

3,5-Dinitrosalasylic acid (DNSA)

Ammonium persulphate

Ampicillin

Beechwood xylan

Birchwood xyaln

Bis tris propane

Bovine serum albumin, fraction V (BSA)

Chloramphenicol

Coomassie Brilliant Blue G

D-Glucose

Dimethylsulphoxide

di-Sodium hydrogen phosphate

Dithiothrietol (DTT)

D-Mannose

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

Ethidium bromide

Ethylene glycol

Glucose-6-Phosphate

Glycerol

Imidazole

Kanamycin

L-Fucose

Levan from *E.herbicola*

Mannose-6-phosphate

Mineral oil (light)

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

Nicotinamide adenine dinucleotide-reduced

Phenol

Polyethylene glycol MW 3350

Sodium bicarbonate

Sodium carbonate

Sodium dihydrogen orthophosphate

Sodium dodecyl sulphate (SDS)

Sucrose (nuclease free)

Trizma base (Tris)

Yeast mannan

β-Mercaptoethanol

A2 Media

Difco

Bacto®tryptone

Bacto®yeast extract

Oxoid

Bacteriological Agar No.1

A3 Enzymes

MBI Fermentas

DNA restriction endonucleases

Invitrogen

Bacteriophage T4 DNA ligase

Novagen

KOD HotStart DNA polymerase

Sigma

Glucose-6-phosphate dehydrogenase

Hexokinase

Phosphoglucose isomerase

Phosphomannose isomerase

Stratagene

*Pfu*Turbo DNA polymerase

*Dpn*I restriction endonuclease

A4 DNA

MBI Fermaentas

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

MWG Biotech

All primers

A5 Kits

Qiagen

Plasmid Mini kit

Plasmid Midi kit

Qiaquik Gel Extraction kit

Qiaquik PCR Purification kit

Stratagene

QuikChange™ Site-Directed Mutagenesis kit

Megazyme

D-Mannose/D-Fructose/D-Glucose assay kit