Diversity in structure and substrate specificity of family 43 glycoside hydrolase enzymes

Lauren Sara McKee

A Thesis Submitted for the Degree of Doctor of Philosophy

2008-2011

Institute for Cell and Molecular Biosciences

Newcastle University

Declaration

The work presented in this thesis is solely the work of the student Lauren S McKee, except where stated.

NMR analysis was performed by Dr. Maria Peña, Assistant Research Scientist in Prof. William York's laboratory at the Complex Carbohydrate Research Centre at the University of Georgia, Athens. Samples were prepared and lyophilised by the student, but all NMR analysis was carried out by M. Peña. Identification of peaks and interpretation of data were also performed by M. Peña.

Solution of the selenomethionine structure presented in Chapter Four by the single wavelength anomalous dispersion method was first performed by Dr. Jon Marles-Wright, University Research Fellow in the Structural Biology Laboratory at Newcastle University. Subsequent to instruction by Dr. Marles-Wright, the student repeated the structure solution – crystallographic statistics and structure figures shown in this thesis are from the structure solved by the student.

Solution of the ligand-bound structures presented in Chapter Four were solved by the student using techniques of molecular replacement, following instruction by Dr. D Wade Abbott, a Post-Doctoral Research Associate in Prof. Harry Gilbert's laboratory at the CCRC in Athens.

Acknowledgements

At the completion of my PhD thesis I must extend heartfelt thanks to my supervisor Harry Gilbert, for allowing me to spend three wonderful years working in his group. Thanks to his teaching and encouragement, I have gained confidence in my skills as a scientist and I feel I have finally found my voice.

Thanks to Harry's temporary relocation to the USA, I have been lucky enough to work with two separate but equally special groups of people. From the team at Newcastle University, I'd like to express my gratitude to Joanna Norman, Elisabeth Lowe, Carl Morland, Fiona Cuskin, David Bolam, Ziyuan Liu and Artur Rogowski. Of the group we left behind at the Complex Carbohydrate Research Centre, I am grateful for the friendship and advice of Lei Zhao, Yanping Zhu, Supriya Ratnaparkhe, Annabel Hao, Wade Abbott and Alan Cartmell.

Finally, much of the work presented in this thesis would not have been possible without collaboration with individuals more experienced and more capable than I. In particular, I'd like to thank Jon Marles-Wright and Maria Peña for their respective contributions to the x-ray crystallography and NMR experiments which were invaluable to this study.

Abstract

There are strong drives to find a viable alternative to the use of petroleum as a transport fuel. Bioethanol presents an attractive option, but the long-term costs of producing this fuel from corn starch are now apparent. Second generation biofuels, derived from cellulosic plant cell walls, are a more acceptable alternative. A limiting factor in the economical utilisation of plant biomass is efficient saccharification of carbohydrates. The process is slowed by the chemical complexity of the substrate, notably the recalcitrant double substitution structures in arabinan and arabinoxylan.

Reflecting this complex chemistry, microorganisms that degrade the wall synthesise an array of glycoside hydrolases. Several such organisms contain a large number of genes encoding family 43 (GH43) glycoside hydrolases. To better understand the biological rationale behind the expansion in this family, the biochemical properties of the GH43 enzymes of a human gut symbiont, *Bacteroides thetaiotaomicron*, were investigated. Through cloning experiments, soluble protein was obtained for 25 enzymes. Activity screens uncovered several enzymes with a weak xylanase activity, three arabinoxylan-specific arabinofuranosidases, two *endo*-arabinanases and a novel arabinofuranosidase with specificity for α-1,2 side chains of singly and doubly substituted backbone residues. The crystal structure of a close homologue of the novel arabinofuranosidase is reported here. These data show how *B. thetaiotaomicron* deploys a combination of *endo*-acting and side chain-cleaving hydrolases to metabolise arabinan polysaccharides.

Two GH43 enzymes (designated AXHd3s) have been found to target the double substitution structure in arabinoxylan. The crystal structure of the *Humicola insolens* AXHd3 was sought to understand this specificity, and is presented in complex with reaction products. Structural and mutagenic data were used to identify the mechanism by which the enzyme houses the O3-linked arabinofuranose in the active site, while exploiting the O2 appended arabinofuranose and asymmetrical xylan backbone as specificity determinants. Analysis of these data showed that orientation of the backbone, mediated by interactions with a conserved Tryptophan, positions the O3 arabinose into the active site. Modification of the rim of the active site pocket generated an AXHd3 variant that displayed both *endo*-xylanase and AXHd3 arabinofuranosidase activities. The introduction of additional catalytic functions into a biotechnologically relevant glycoside hydrolase provides a platform for evolving further, industrially significant, activities into the AXHd3 scaffold.

Table of Contents

Table of Figures

Table of Tables

List of Commonly Used Abbreviations

AG – arabinogalactan

AGP – arabinogalactan-protein

AX – arabinoxylan

CAZy – Carbohydrate Active enZymes database

CBM – carbohydrate binding module

CCRC – Complex Carbohydrate Research Centre, University of Georgia

d.p. – degree of polymerisation

ECFσ – extracytoplasmic function sigma factor

GAX – glucuronoarabinoxylan

GH – glycoside hydrolase

GX – glucuronoxylan

HCA – hydrophobic cluster analysis

HG – homogalacturonan

HPLC – high performance liquid chromatography

HTCS – hybrid two component system

linA – linear arabinan

PUL – polysaccharide utilisation locus/loci

RG – rhamnogalacturonan

sbA – sugar beet arabinan

SelMet - selenomethionine

Sus – starch utilisation system

TLC – thin layer chromatography

wAX – wheat arabinoxylan

XG - xyloglucan

Journal Articles

Publications containing work conducted by the author during the course of this investigation are under preparation or in press, at the time of writing:

L S McKee*, A Cartmell*, M J Peña, J Larsbrink, H Brumer, R J Lewis, A Vikso-Nelson, H Ichinose, S Kaneko, H J Gilbert and J Marles-Wright. (2011) The structure and function of an arabinan-specific α-1,2-arabinofuranosidase identified from screening the activities of bacterial glycoside hydrolases. Journal of Biological Chemistry *286*, 15483-15495. *contributed equally.

L S McKee, J Marles-Wright, M J Peña, W S York, K B R M Krogh, A Viksø-Nelson, M Skjøt and H J Gilbert. (2011) The crystal structure of an arabinofuranosidase that targets complex xylan decorations provides a platform for evolving additional catalytic functions into the β-propeller fold. Manuscript in preparation.

C Y Montanier, M A Correia, J E Flint, Y Zhu, L S McKee, J A Prates, S J Polizzi, P M Coutinho, R J Lewis, B Henrissat, C M Fontes and H J Gilbert. (2011) A novel, non-catalytic carbohydrate-binding module displays specificity for galactose-containing polysaccharides through calcium-mediated oligomerization. Journal of Biological Chemistry. Paper in press

Other articles containing work conducted by the author which is not presented in this thesis:

Y Verhertbruggen, S E Marcus, A Haeger, R Verhoef, H A Schols, B V McCleary, L McKee, H J Gilbert and J P Knox. (2009) Developmental complexity of arabinan polysaccharides and their processing in plant cell walls. Plant Journal *59*, 413-425.

CHAPTER ONE

General Introduction

The plant cell wall is the most significant source of renewable carbon in the biosphere (Coughlan, 1985), with annual synthesis in the order of 10^{11} tonnes, representing an energy equivalent of 640 billion tonnes of oil (Gong et al., 1999). As petrol prices are once again making headline news in the UK and the price of a barrel of oil is returning to $100 per barrel, there is a financial drive to find alternatives to this costly fuel, the value of which has a wide ranging impact upon individuals and upon our whole economy. Additionally, the environmental costs of intensive fossil fuel burning are at the centre of heated debates over global warming and the causes of the freak weather patterns being observed worldwide, with floods, droughts and harsh winters affecting communities and economies.

While the use of petroleum products as transport fuel is not the only area of heavy fossil fuel use, it may represent the simplest conversion to more sustainable fuels. The investment required to convert a nation's power supply to renewable resources is huge, but a typical car engine can be converted to run on bioethanol for ~ $600 (around £370, $1 buys £0.62 February 2011) and rising petrol prices may convince customers that this is a worthwhile one-time expenditure. Efforts, particularly in the USA, have so far focussed on developing bioethanol derived from corn starch as a viable alternative to petrol. In 2007, corn fields expanded by 15 % to meet the demands of the bioethanol industry (Martinelli and Filoso, 2008). However, in recent years the full impact of corn ethanol has become apparent. Taking corn for fuel deprives the food and feed markets and we have seen prices rise steeply as a result. Furthermore, it is now clear that corn ethanol is not as environmentally advantageous as it seemed (Hill et al., 2006). Although the fuel itself does burn more cleanly than petrol, the production of corn ethanol has many adverse impacts upon the environment, including the release of carbon from converted agricultural land, and the large volumes of water required to produce the fuel (Gerbens-Leenes et al., 2009; Martinelli and Filoso, 2008). Overall, using bioethanol derived from corn offers

only a very modest saving in greenhouse gas emissions compared to the use of petroleum fuel (Sheehan, 2009).

So-called second generation biofuels, derived from cellulosic plant biomass represent a more environmentally acceptable alternative (Sheehan, 2009). This biomass can be derived from bioenergy-specific crops grown on waste land, agricultural by-products and even household waste, which largely consists of plant biomass, as the paper and cardboard portion of municipal solid waste (MSW) can be 30-60 %. Land conversion is therefore not required to obtain the raw material (Lal, 2008). Cellulose, and other polysaccharides of the plant cell wall, can be fermented to ethanol using processes and infrastructure very similar to those currently used for corn. However, due to the complexity and recalcitrance of its structure, the plant cell wall is very resistant to chemical and biological hydrolysis (Himmel et al., 1999). The major limiting factor in the economical utilisation of the plant cell wall for fuel is efficient saccharification of the carbohydrate components of the wall. It is therefore essential that we possess a detailed knowledge of the structure of the plant cell wall and of the enzymes which hydrolyse it, in order to inform future efforts to utilise lignocellulosic plant biomass. This chapter aims to review current knowledge of the structure and components of the plant cell wall, as well as the enzymes which hydrolyse the wall. A discussion will also be made of the relevance of plant polysaccharides to human health and industry.

1.1 The plant cell wall

1.1.1 Structure of the plant cell wall

Many diverse methods are now employed to probe the structure of the plant cell wall. Atomic force microscopy (AFM) can provide a close look at the wall; this is often supplemented by probing the cell wall with antibodies, carbohydrate binding modules (CBMs) and inactivated enzymes (Adams et al., 2004; Knox, 2008; Verhertbruggen et al., 2009).

The structures of the polysaccharides found within the plant cell wall will be discussed below (Section 1.1.2). A model structure for the plant cell wall is shown in Figure 1.1. The plant cell wall provides the mechanical properties that enable the

plant to deal with tensile and compressive forces which result from upright growth on land. The cell wall also regulates cell expansion and adhesion (Knox, 2008). The importance of the cell wall to healthy plant growth and development is underlined by the fact that plants dedicate around 10 % of their genomes to proteins involved in construction and rearrangement of the wall during growth. For instance, the poplar genome encodes around 1600 carbohydrate active enzymes (cazymes) (McCann and Carpita, 2008). The plant cell wall, represented as a static network of polymers in early models, is in fact a highly dynamic organelle. Plant cells continue to divide throughout the whole lifetime of a plant, while retaining many characteristics of a young cell. The thickness of the wall (0.1 – 1 μm in cross-section) must be maintained even as modifications to the wall allow the cell to elongate or divide, as a turgor pressure of hundreds of mega Pascals is exerted (Perez et al., 2003). This flexibility is achieved by reorganisation of the molecular components of the wall via rearrangement, cleavage and cross-linking of polysaccharide structures.

After a cell divides, a layer called the middle lamella is deposited. This layer is comprised almost wholly of pectin. The early cell wall has surface markers which indicate the future development of the cell and signal the location of the cell within the plant (Carpita and Gibeaut, 1993). The middle lamella is then supplanted by the primary cell wall which is pectocellulosic and contains some proteins and hemicelluloses (Perez et al., 2003). Eventually, if a cell differentiates and requires additional structural support, a secondary cell wall will be laid down. The primary wall, which surrounds dividing cells, is chemically and structurally very distinct from the secondary wall, which provides a cell with its differentiated, functional form (Carpita and Gibeaut, 1993). Thus, physical and chemical differences between the layers of the cell wall reflect different functions within the plant tissue.

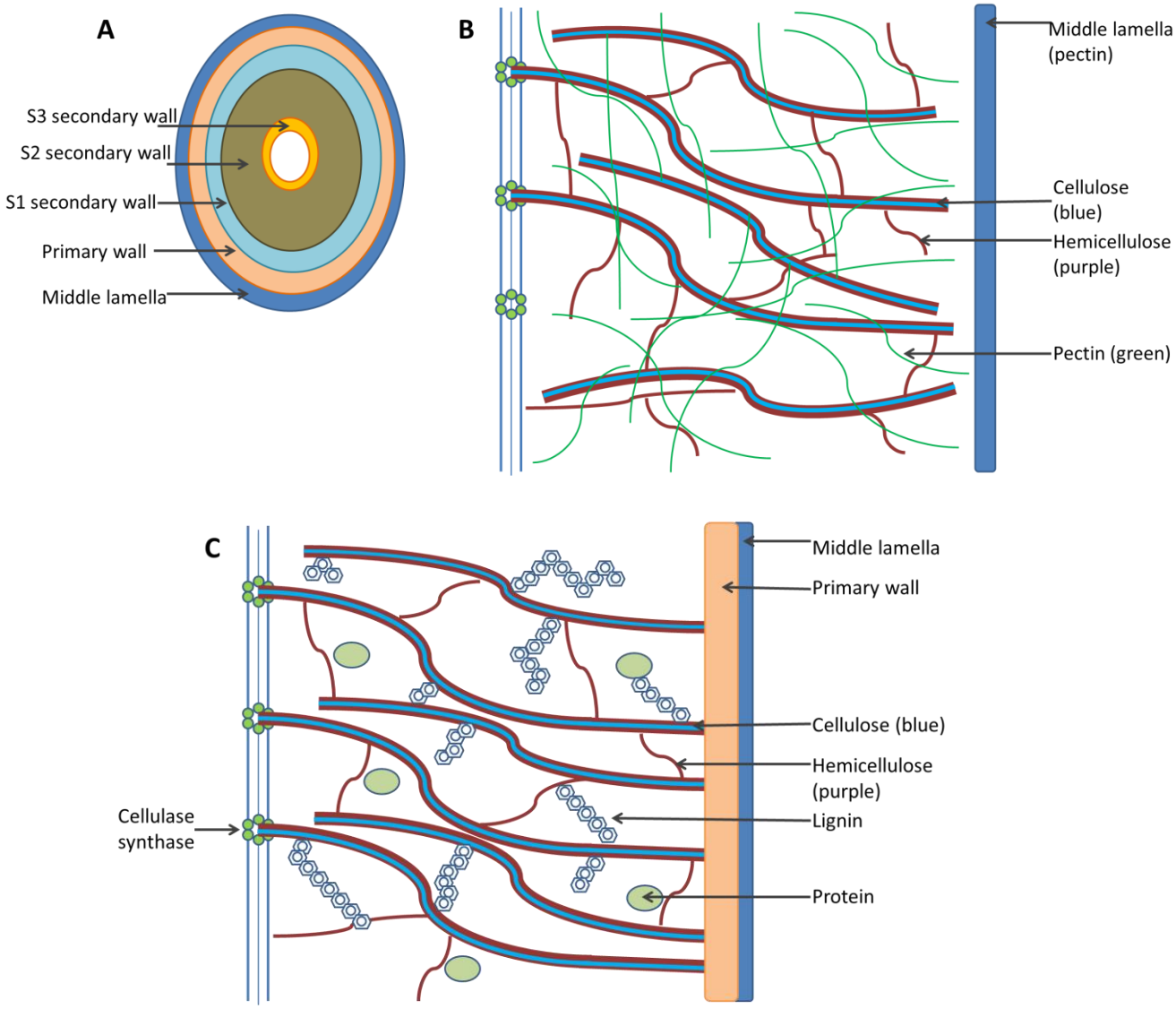


Figure 1.1 Model structure of the cell wall of a differentiated plant cell.

A: the schematic shows how the layers of the cell wall build as a cell develops and differentiates. The cell is the white space at the centre of the schematic. The first layer deposited is the lamella (blue). Subsequent layers are deposited sequentially within the previous layer. Based on a figure at www.ccrc.uga.edu.

B: the primary cell wall comprises cellulose microfibrils (light blue) coated with hemicellulose (purple). Hemicellulose and pectin (green) cross-link these fibrils.

C: the secondary cell wall is made of cellulose microfibrils (light blue) coated in hemicellulose (purple). Hemicelluloses also cross-link the microfibrils. The network is embedded with lignin. Some proteins are also present.

[4]

1.1.1.i The primary cell wall

The primary cell wall is found in cells of all higher plants (Mohnen, 2008a), and comprises up to 30 % cellulose. The role of the primary wall, which surrounds growing cells (Figure 1.1), is to generate and constrain turgor pressure from within the cell, providing structural support to the plant (Knox, 2008). In Type I primary walls, cellulose cross-links with xyloglucan (XG) and this matrix interacts closely with pectins and some proteins. Cellulose microfibrils are 5 – 15 nm wide and spaced 20 – 40 nm apart (Carpita and Gibeaut, 1993).

In Type II primary walls, which are found in grass species, xyloglucan is replaced with glucuronoarabinoxylans (GAX) which cross-link cellulose in a network of hydrogen-bonds (McCann and Carpita, 2008). As these grass cells mature, phenylpropanoids are substituted into the primary cell wall, via ether and ester linkages initiating from arabinose substitutions in GAX (McCann and Carpita, 2008).

Pectic polymers, the most structurally complex plant cell wall polysaccharides, are abundant in all primary cell walls (Knox, 2008). Areas which are especially rich in pectin include growing or dividing cells, soft plant tissues, cell corners and junctions between cells which possess secondary walls (Mohnen, 2008a). The diverse structural features of the pectins correlate with stages of cell development, as processes such as cell expansion require significant remodelling of the wall. The primary walls of dicots, grasses and woods comprise approximately 35 %, 10 % and 5 % pectin, respectively. Pectin is thought to have evolved as plants adapted to upright growth on land, when the forces of gravity became more significant and lignified secondary walls became common, as the pectin ‘glue’ provides the cell wall with much strength (Matsunaga et al., 2004).

1.1.1.ii The secondary cell wall

As an organ develops, species-specific patterns of cell types differentiate and some will develop secondary cell walls (Figure 1.1). The secondary wall is deposited upon maturity and differentiation of cells such as those of the xylem, phloem and transfer cells, which are richly abundant in forest crops. The secondary wall typically comprises 60-98 % cellulose, along with large amounts of xylan, some lignin and

sometimes a minor amount of pectin (McCann and Carpita, 2008). The secondary wall is very strong and has a set of hemicelluloses distinct from those in the primary wall. It may comprise up to 30 % lignin by mass in some species (Scheller and Ulvskov, 2010). This allows the wall to resist compressive stress. Secondary walls are structurally and functionally diverse at a molecular level, likely reflecting the mechanisms of development of different cell types (Knox, 2008).

1.1.2 Plant cell wall polysaccharides

Polysaccharides, as the major components of all plant cell walls, serve several important functions. They limit porosity, transduce environmental signals into cells and cement cells together at adhesion points and the middle lamella (Perez et al., 2003). The major classes of plant cell wall polysaccharide are cellulose, hemicellulose and pectin; these are discussed in turn below.

1.1.2.i Cellulose

Cellulose is a major component of the plant cell wall, and represents the most abundant biosynthesised polymer on Earth. The primary structure of cellulose is a β-1,4 linked glucan chain, where successive residues are rotated 180° to form a flat ribbon with cellobiose as the repeating unit (Mohnen et al., 2008b, Figure 1.2). Glucan chains of between 500 and 14,000 glucose residues hydrogen-bond to form microfibrils, insoluble structures comprising approximately thirty-six parallel chains (Mohnen et al., 2008b). This network of inter- and intra-molecular hydrogen-bonds between hydroxyl groups leads to crystallisation as microfibrils form even larger fibrils (Park et al., 2010). The crystalline microfibril is very large compared to the individual glycan chains from which it is built. A degree of polymerisation (d.p.) in the glucan chain of 2000 corresponds to around 1 μm of length. As primary wall cellulose fibrils are frequently observed to be much longer than this, it is likely that glucan chains begin and end at different points in the microfibril (Carpita and Gibeaut, 1993; Mohnen et al., 2008b).

Cellulose microfibrils provide rigidity and resistance to tensile and compressive forces. They are tethered together in the primary cell wall by extensive interactions with hemicelluloses, primarily xyloglucan (XG) but also lower amounts of mannans.

The cellulose-XG framework accounts for around 50 % of the mass of the primary cell wall (Carpita and Gibeaut, 1993). This interconnected network is in turn embedded within a matrix of pectic polymers. The importance of cellulose as a central scaffold of the plant cell wall is underpinned by the observation that mutants deficient in secondary wall cellulose show very irregular deposition of other wall components (Turner and Somerville, 1997).

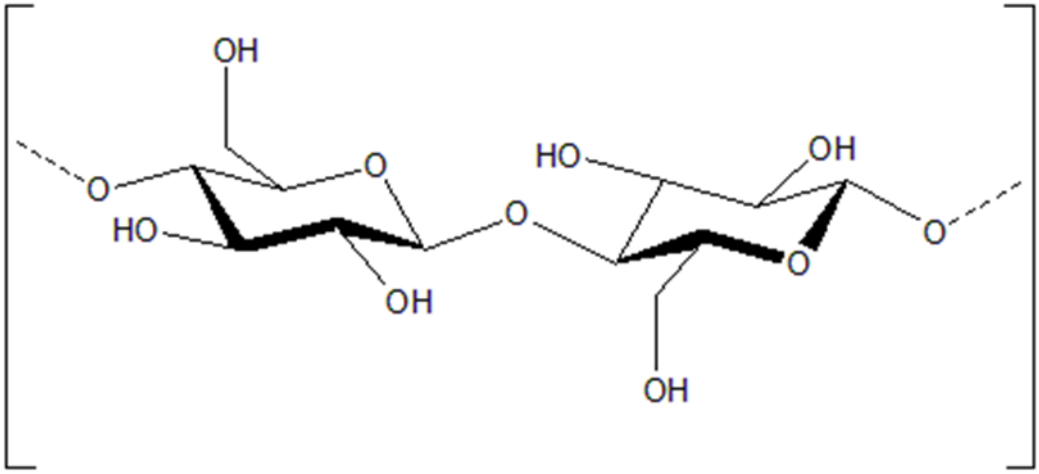


Figure 1.2 Cellobiose is the foundational unit of cellulose microfibrils.

The repeating unit of the cellulose chain is β-1,4-linked cellobiose. Intramolecular hydrogen bonds form between these cellobiose residues and hold the crystalline form of cellulose together. Glucan chains hydrogen-bond to each other and form microfibrils, which crystallise in the wall.

1.1.2.ii Hemicelluloses

A more appropriate name for this class of plant cell wall polysaccharide is ‘cross-linking glycans’. Nonetheless, the term hemicellulose, referring to polysaccharides of the plant cell wall which are neither cellulose nor pectin, persists in the literature and so is utilised in this report. A broad definition of hemicelluloses is polysaccharides with a β-1,4 linked backbone in the equatorial configuration (Scheller and Ulvskov, 2010, Figure 1.3). Hemicelluloses can be extracted from the wall by alkali treatment and prominent examples include xylans and xyloglucan, which are discussed below, as well as mannans, glucomannans and some mixed linkage glycans. Hemicelluloses are found in all terrestrial plants but the structure and abundance of the polymers within the plant cell wall varies by cell type and plant species. These

polysaccharides, particularly xyloglucan, interact with cellulose microfibrils to give strength to the wall.

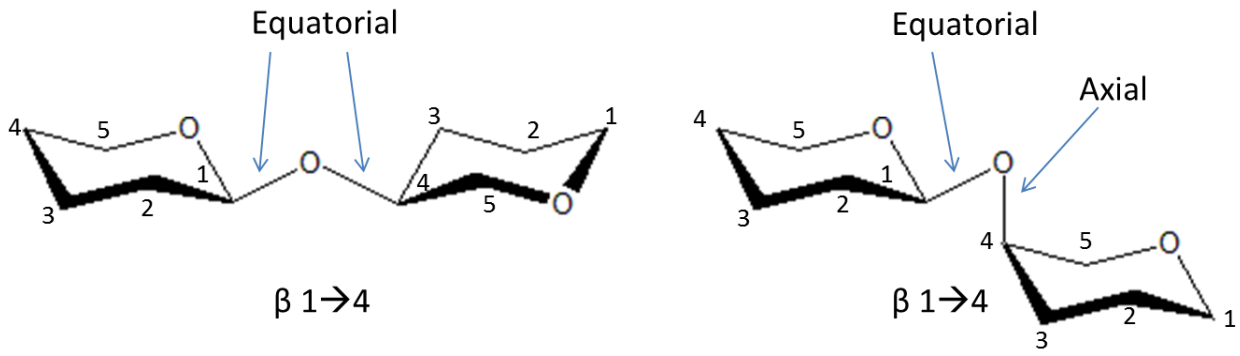


Figure 1.3 Hemicelluloses have a β-1,4-linked equatorial configuration.

The schematic (A) shows pyranose sugars in the $^4C_1$ conformation bonded in the equatorial configuration followed by hemicellulosic polysaccharides. Polymers with the alternative axial configuration at C4 (B) include galactan, a pectic polysaccharide.

Xyloglucan (XG)

Xyloglucan is found in all land plants and is the most abundant hemicellulose of the primary wall of non-grass species. It localises at points of cell adhesion after cytokinesis. The backbone structure of β-1,4 linked glucopyranosyl residues is heavily and specifically substituted (Figure 1.4). A code is used to indicate substitution of backbone glucose units: X residues are substituted with xylose, F residues are substituted with xylose, galactose and fucose, and L residues are substituted with xylose and an additional galactose, which is replaced by arabinose in S residues (Scheller and Ulvskov, 2010). These branches occur in specific patterns. For instance, XXFG is an important signalling molecule which counteracts cell expansion induced by auxin (Fry et al., 1993). Analysis of the sequences of these substitutions shows that a major divide among vascular plants is whether the predominant core repeat of XG is XXGG or XXXG (Scheller and Ulvskov, 2010). These side chains, together with the helical nature of the XG backbone, prevent self-aggregation. Broadly speaking, less extensively branched XG is less soluble and is found in expanding cells. Xyloglucan functions as a ‘tethering glycan’ in the primary cell wall by close association with cellulose microfibrils. Possible mechanisms for this

association are an extensive network of hydrogen bonds or by the XG being trapped by the microfibrils during crystallisation, due to close proximity in the cell wall (Scheller and Ulvskov, 2010). The tightly inter-connected cellulose-XG network is required to loosen to allow the plant cell to grow; this is achieved by enzymatic modification of XG in combination with internal osmotic pressure, which pushes microfibrils apart.

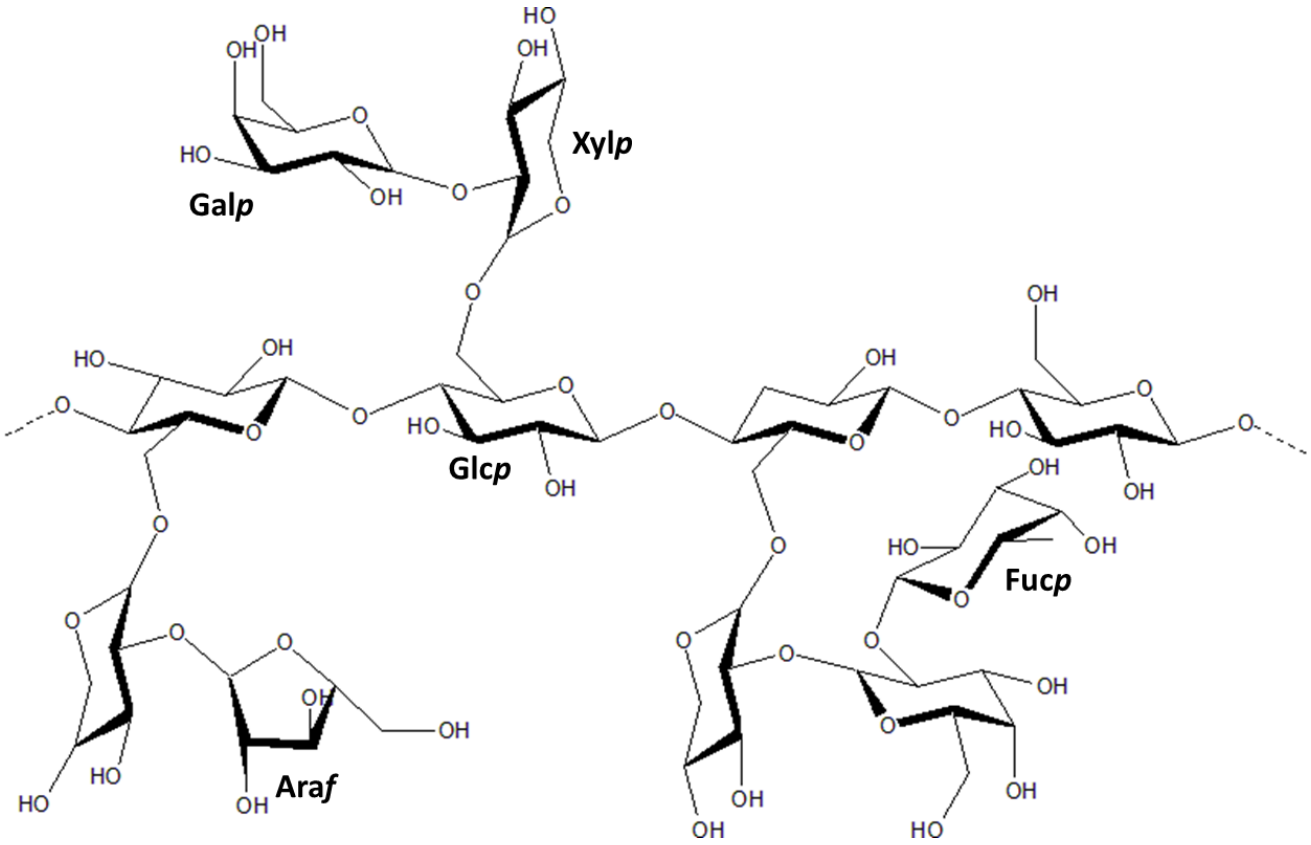


Figure 1.4 The structure of xyloglucan.

The hemicellulose xyloglucan has a backbone of β-1,4 linked glucopyranose residues. 75 % of these are substituted at the O6 position with a xylose residue, which may be additionally substituted with galactose or other sugars.

Glc*p* = glucopyranose. Fuc*p* = fucopyranose. Xyl*p* = xylopyranose. Gal*p* = galactopyranose. Ara*f* = arabinofuranose.

Xylans

Xylans are built on a backbone of β-1,4 linked xylopyranosyl residues (Figure 1.5). Glucuronoxylans (GX), which also include glucuronic acid and 4-*O*-methyl-

glucuronic acid, are the dominant non-cellulose polysaccharide in many dicot secondary cell walls (Scheller and Ulvskov, 2010). Xylans are also the major non-cellulosic polysaccharide in the primary wall of grasses, where backbone xylose residues are heavily substituted with arabinose (arabinoxylan (AX) or glucuronoarabinoxylan (GAX)) at O2, O3 or both. The pattern of arabinose substitution, and most common linkage type, varies by species (Scheller and Ulvskov, 2010) and is believed to vary even within a single species (Adams et al., 2004; Izydorczyk and Biliaderis, 1995). Substitution patterns have been studied by AFM in conjunction with probing by inactivated enzymes; the binding of a catalytically inactive GH11 xylanase, which requires three unsubstituted xylose residues to bind, changes significantly after arabinoxylan treatment with arabinofuranosidases (Adams et al., 2004).

Xylans do not possess repeated structures of side chain decoration, as seen in XG. Other than the single residue substitutions described above, the polysaccharide is generally considered to be linear, with varying degrees of acetylation, usually at O3 of xylose residues (Scheller and Ulvskov, 2010). It is likely that ‘smooth’ regions of two to five xylose residues occur between substituted stretches (Chanliaud et al., 1995; Dervilly et al., 2000). Strengthening the inter-connected nature of the wall, arabinoxylans can be cross-linked via 5-5’-diferulic acid bridges connecting arabinose side chains (Lempereur et al., 1997). Ferulic esters can be oxidatively cross-linked between GAX molecules and also with lignin (discussed below), providing a covalent association between lignin and the hemicellulose domain, increasing the complexity and strength of the plant cell wall to increase cellular defence against attack by pathogens or herbivores, and also increasing resistance to saccharification by industrial enzymes (Scheller and Ulvskov, 2010). Xylan is a major load-bearing structure of the secondary cell wall, as demonstrated by deficient mutants, which have collapsed xylem vessels and associated defects in growth and fertility (Scheller and Ulvskov, 2010).

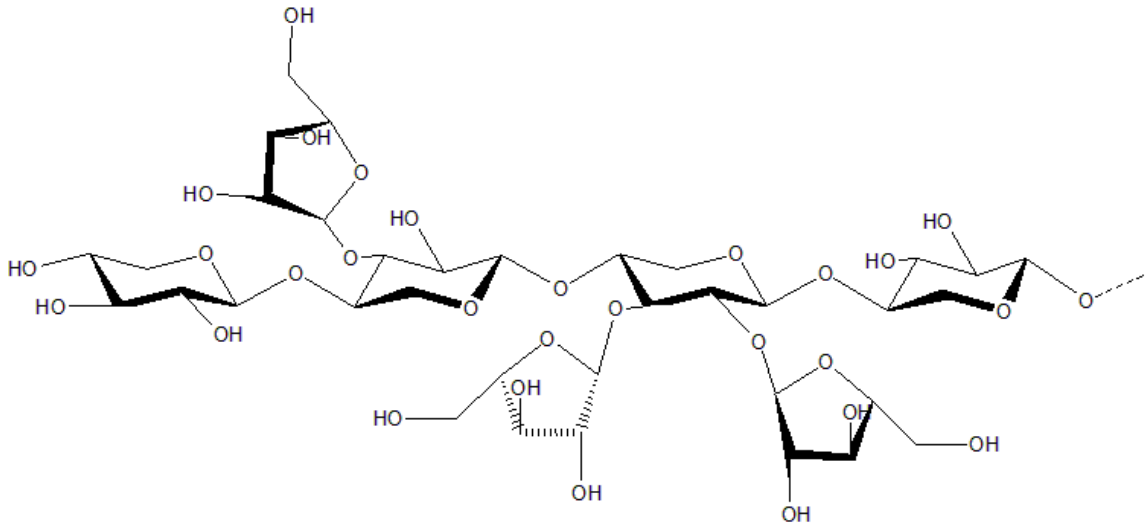


Figure 1.5 The structure of arabinoxylan.

The hemicellulose xylan is built on a backbone of β-1,4 linked xylopyranose residues. A proportion of these residues are decorated at O2, O3 or both with α-L-arabinofuranose residues. The arabinose linkages are highly flexible. There may also be substitution with glucuronic acid, and some degree of acetylation.

1.1.2.iii Pectins

Pectins are complex, acidic heteropolysaccharides found in the primary cell walls of growing plants with roles in growth, morphology, development and defence (Mohnen, 2008a). Galacturonic acid contributes around 70 % of all pectin, by mass. Pectin comprises three domains (Figure 1.6): homogalacturonan (HG), rhamnogalacturonan I (RG I) and rhamnogalacturonan II (RG II). The pectic matrix provides an environment for deposition and extension of the cellulose-glycan network (Perez et al., 2003, Figure 1.6). The lack of a pectic network in the secondary cell wall is a significant feature, as it causes this layer to be more rigid and much less capable of extension and modification than the primary wall. The distribution of the three domains within pectin is not known. Around fifty transferase enzymes are required to build the carbohydrate portions of pectin (Perez et al., 2003) if the 'one linkage – one enzyme' theory holds true (Mohnen et al., 2008b). Additional enzymes are required for modifications such as acetylation and methyl esterification, as well as those which transport nucleotide sugars to the site of pectin biosynthesis, the Golgi lumen (Mohnen, 2008a). The different pectic polysaccharides are covalently linked via their backbones (Figure 1.6) and may also be covalently

linked to xyloglucan and hemicelluloses, principally xylan. Xylan chains with a degree of polymerisation (d.p.) of up to seven may be connected at O3 to HG, and xylobiose may be located on RG I (Mohnen, 2008a). These interactions between layers of the cell wall further enhance the strength of the whole structure.

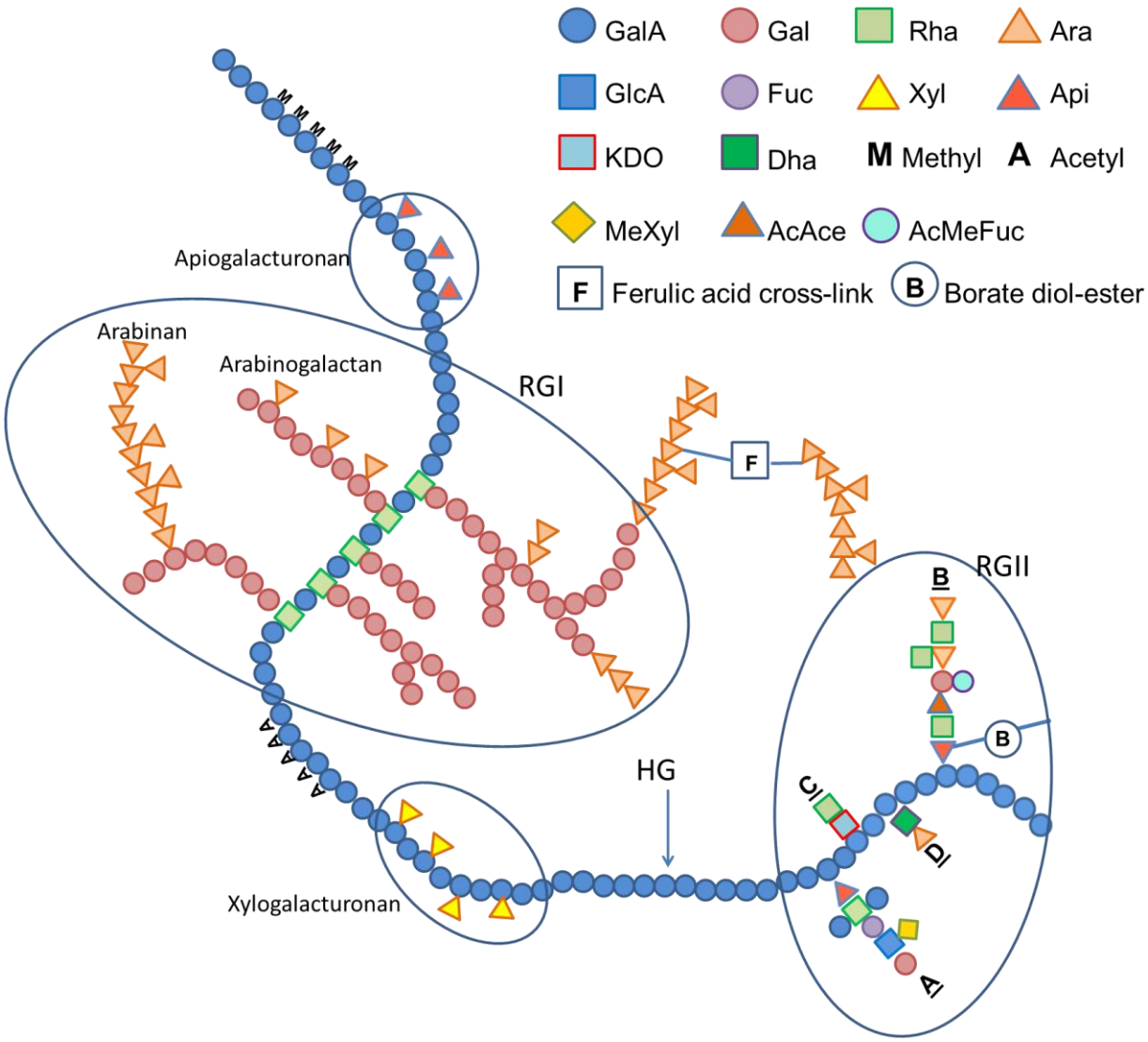


Figure 1.6 The primary structure of pectin.

The schematic represents the canonical primary structure of pectin, with the homogalacturonan (HG) backbone interspersed with decorated regions of rhamnogalacturonan I (RG I) and occasionally substituted with four conserved side chains to make rhamnogalacturonan II (RG II). This model assumes that the three domains are covalently linked. Also shown are the locations of ferulic acid cross-links between arabinan chains (which can also link to hemicelluloses) and the site of RG II dimerisation by borate diol-ester formation at an apiose residue in side chain B. Figure adapted from (Perez et al., 2003). Me = methylated. Ac = acetylated. GalA = galacturonic acid. Gal = galactose.

Rha = rhamnose. Ara = arabinose. GlcA = glucuronic acid. Fuc = fucose. Xyl = xylose. Api = apiose. Dha, KDO = rare sugars.

Homogalacturonan (HG)

Around 65 % of pectin is in the form of homogalacturonan (Mohnen, 2008a). HG comprises a linear backbone of α-1,4 linked galacturonic acid residues. Some carboxyl groups are acetylated or methyl-esterified at C6 and the degree of these substitutions is important in pectin matrix interactions (Perez et al., 2003). The HG chain is covalently attached to RG I (Figure 1.6), which is built on a backbone of a repeating disaccharide: α-L-rhamnosyl residues are 1,2 linked to α-D-galacturonosyl residues. The HG backbone is helical in structure and these helices can associate via interactions with calcium ions (Perez et al., 2003). Overall, regions of HG of up to 100 residues are interspersed with shorter RG I regions. While an HG chain proceeds in a single direction, the presence of alternating rhamnose residues in RG I slightly alters the direction of the pectin chain (Perez et al., 2003, Figure 1.6).

Rhamnogalacturonan I (RG I)

RG I makes up 20 – 35 % of pectin in the primary cell wall and is branched in a manner which is dependent upon cell type and developmental stage (Mohnen, 2008a). 20 – 80 % of backbone rhamnosyl residues have linear or branched side chains; the reason for this variation in degree of substitution is not well understood (Mohnen, 2008a). Side chains of RG I include arabinan, galactan and arabinogalactan (Perez et al., 2003) and may have a d.p. of up to forty-seven (Mohnen, 2008a). These oligomeric side chains are flexible and mobile. It is possible that they maintain pores in the cell wall via interactions with water molecules.

There are multiple defined forms of arabinogalactan, with subtle structural differences. Type I arabinogalactan (AG) comprises a β-1,4 linked backbone of galactopyranose residues, with arabinofuranosyl substitutions on certain backbone residues. There is species-dependent variation in this structure. Recent examination of a Type I AG from potato has shown that 0.6 % of galactopyranose residues in the polysaccharide backbone are β-1,3 linked, and a Type I AG isolated from soy has some α1,5 linked arabinofuranose residues in the backbone (Hinz et al., 2005). Type

I AG is linked to RG I regions of pectin, as described above, and there may be covalent associations with some other polysaccharides (Hinz et al., 2005).

Type II AGs are more widespread in plants. In addition to the RG I side chains, these structures are found in arabinogalactan proteins (AGPs) and in some excreted gums (Hinz et al., 2005). They are built on a backbone of β-1,3 linked galactopyranose residues and can be substituted with galactose, arabinose, rhamnose and others.

Other RG I branches are α-1,5 arabinan chains, which can also exist as free polymers in the primary wall, not connected to any pectic domain. The backbone structure of arabinan, which is helical, is decorated with additional arabinose residues at O2, O3 or both positions of backbone residues (Verhertbruggen et al., 2009). Arabinan contributes to important wall properties such as porosity and flexibility, particularly the flexibility of guard cells which control gas exchange and water loss by maintaining pores known as stomata (Verhertbruggen et al., 2009). Branched arabinan chains can cross-link through ferulic acid linkages attached at O5 of arabinofuranosyl residues (Levigne et al., 2004). These ferulic esters can form dimers to cross-link polysaccharides and in some cell walls arabinan is in close association with cellulose, further strengthening the connections between the layers of the cell wall (Vignon et al., 2004; Zykwinska et al., 2008; Zykwinska et al., 2005).

Rhamnogalacturonan II (RG II)

The third domain of pectin is RG II, a region of HG with four specific and highly conserved side-chains (Figure 1.6). While HG and RG I are relatively heterogeneous, RG II is notable for an apparently absolute conservation of structure. Undoubtedly the most complex plant polysaccharide, the structure of RG II is identical in all land plants analysed to date (Perez et al., 2003). This high conservation indicates a major role in higher plants. A single stretch of RG II, which constitutes approximately 10 % of pectin in the primary cell wall, typically has a molecular weight of 5 – 10 kDa (Perez et al., 2003). RG II can be released from primary cell walls by treatment with an *endo*-α-1,4-polygalacturonanase. This polymer is found widely in primary cell walls but is thought to be largely absent from the pectin-rich middle lamella. Although it accounts for only around 5 % by mass of the primary wall, RG II is a source of frustration to saccharification processes, as it

resists degradation by all known pectinolytic enzymes (Perez et al., 2003). By providing such strong resistance to enzymatic attack, RG II represents a cell wall structure that will be stable for the lifetime of the plant. This contrasts with the time-dependent chain pairings of HG chains which stabilise pectin as a gel when the cell wall is required to expand (Perez et al., 2003).

As Figure 1.6 shows, the RG II monomer is very well characterised. It comprises an HG backbone of seven to nine galacturonosyl residues and four oligomeric side chains. These side chains comprise twelve sugars connected by over twenty linkages, and are acetylated and methyl esterified at certain points. The general three-dimensional shape of the monomer (mRG II) is a flat disc with a diameter of approximately 37 Å and a thickness of approximately 17 Å (Pellerin et al., 1996; Perez et al., 2003). The B side chain (Figure 1.6) in particular is highly flexible. In the plant cell wall, two mRG II monomers will dimerise via a 1:2 borate:diol ester cross-linkage (Ishii et al., 1999; Kobayashi et al., 1996). Dimerisation covalently connects two mRG II monomers and their associated HG chains (Ishii et al., 2001). The dimerising cross-link connects the apiose sugar residues in side chain B (Figure 1.6). Mutation of this residue via genetic manipulation of the available complement of glycosyltransferases leads to reduced dimer formation and a phenotype of dwarfism (Mohnen, 2008a).

As most of the boron within the plant cell is not present in a bioavailable form, due to the formation of the RG II-boron dimer (dRG IIB), it is thought that a function of RG II may be to control the supply of boron to the rest of the plant. Boron deficiency leads to disorganised cell expansion and abnormal cell walls, presumably due to inefficient dimerisation of RG II (O'Neill et al., 1996). It has been demonstrated that dimer formation determines cell wall thickness and that absence of the dimer causes growth defects (Perez et al., 2003). Experiments have shown that dimer formation is pH-dependent *in vitro*; the dimer forms between pH 2.2 and 4.5 but is stable above pH 4. A specific putative role for the borate:diol ester cross-linkage is as a load-bearing structure which gives rigidity to the wall. During auxin-induced cellular expansion, this linkage would be hydrolysed by a decrease in wall pH, leading to a breakage of the dRG IIB dimer, and relaxation of the wall (O'Neill et al., 1996). This fits with the relatively low abundance of RG II in the cell wall of dicots and monocots,

which suggests that a purely structural role is unlikely to be the primary function of the polymer. The acid-labile nature of the dimerising linkage makes the dRG IIB structure dynamic. In addition, as a component of the pectic matrix, RG II helps to regulate the rate of growth and the passage of enzymes into the cell (O'Neill et al., 1996; Thompson, 2005).

1.1.3 Non-carbohydrate components of the plant cell wall

Lignin is an extremely heterogeneous hydrophobic aromatic polymer found in the secondary plant cell wall, where the impenetrable nature of this molecule confers properties of rigidity, waterproofing and defence against attack. Lignin provides structural support to allow upright growth and vertical transport of water and nutrients in the xylem and phloem vessels (Davin and Lewis, 2005). The compound is particularly important in water transport, as its hydrophobicity contrasts with the largely hydrophilic nature of the polysaccharide components of the plant cell wall. Ferulic acid cross-links are thought to covalently connect lignin to hemicelluloses, further strengthening the matrix and preventing water absorption (Davin and Lewis, 2005). Lignin is also deposited at wound sites, presumably to resist cellular invasion by pathogens. Lignin is synthesised on the cell wall as phenylpropanoid monomers randomly polymerise after export from the cytoplasm. These monomeric units connect via ether linkages, which are very difficult to break. The monomers are radicals and as such are toxic to the cell (Boerjan et al., 2003; Liu et al., 2011).

Lignin impacts different uses of plant material in different ways. Highly lignified wood is very durable and long-lasting, so makes an ideal building material. Conversely, this same durability and recalcitrance hinder paper production and biomass saccharification. Lignin is a major component of insoluble dietary fibre, which is metabolically inert but which is an essential part of the human diet as it absorbs water throughout the digestive system, a process which eases waste elimination (Asp, 1987).

In addition to lignin and the carbohydrate structures described above, a major component of the plant cell wall, which is overlooked in some models of the wall, is protein. It is estimated that there may be up to several hundred proteins in the wall (Showalter, 2001a). Proteins tend to be specific to a certain tissue type, such as the

reproductive gametes (Wu et al., 2001) and while most have a largely structural function, it is becoming clear that many may have active roles in processes such as cell elongation and extension. Most are rich in glycine, proline or hydroxyproline (Ringli et al., 2001; Wu et al., 2001). The largest class of plant cell wall proteins are the hydroxyproline-rich glycoproteins, including extensins and arabinogalactan-proteins (AGPs).

Extensins, as the name suggests, have long been thought to have a role in extension and elongation of the cell wall. The extensins are abundant and dynamic proteins which self-assemble into a cross-linked network in the immature cell wall due to the amphipathic nature of their structure (Lamport, 1965). Recent evidence suggests a role for extensins in loosening of the cell wall during cell division or elongation; this view is supported by the observation that some extensins may have protease activity (Showalter, 2001a).

AGPs are also highly abundant in the plant cell wall and are believed to participate in plant growth and development (Showalter, 2001b). AGPs are heavily glycosylated at specific sequences; the protein component represents only 2 – 10 % of total mass (McNeil et al., 1984). The carbohydrate portion of AGPs is largely comprised of galactose, arabinose, rhamnose, mannose and uronic acids, with an overall structure very similar to Type II arabinogalactans (Section 1.1.2.iii). Some are attached to the plasma membrane, while others are released to the cell wall, from which they may be secreted to the extracellular environment, indicating a possible role in cell signalling events. A putative structural role for AGPs as a ‘cushion’ between the plasma membrane and the plant cell wall has been proposed (Showalter, 2001b).

1.2 Glycoside hydrolases

The half-life for spontaneous hydrolysis of the glucose-glucose bond in cellulose and starch is in the region of five million years. Glycoside hydrolases are enzymes which catalyse the hydrolysis of the glycosidic bond by general acid-base assisted catalysis which requires a proton donor and a nucleophile or base (Davies and Henrissat, 1995). The enzyme-catalysed reaction proceeds via one of two possible mechanisms, which are discussed below. These enzymes accelerate hydrolysis up to 10^{17} fold faster than the spontaneous reaction (Zechel and Withers, 2000).

1.2.1 Family classification and enzyme structure

Selective hydrolysis of glycosidic bonds by living organisms is a vital step in the processes of energy uptake, cell wall expansion and production of signalling molecules (Davies and Henrissat, 1995). For this reason, glycoside hydrolases have evolved which are generally highly specific for sugars, linkages and polysaccharides, although all of these enzymes catalyse the same reaction: hydrolysis of a glycosidic bond. Specificity is driven by tertiary structure. This observation led to an attempt to improve enzyme classification, organising glycoside hydrolases by structure and hence, to an extent, by substrate specificity. A direct relationship was discovered between primary sequence and protein fold, revealed by hydrophobic cluster analysis (HCA) of enzyme sequences. This technique has shown that a structural fold can be displayed by very divergent primary sequences (Callebaut et al., 1997). HCA is a powerful method of predicting protein structure from the primary amino acid sequence, throughout which hydrophobic residues are distributed in a specific manner (Callebaut et al., 1997). Structure prediction by HCA is achieved through the production of a ‘two-dimensional’ representation of the primary sequence, which is displayed as an extended α-helix. The HCA method has developed from the observation that hydrophobic residues cluster when the sequence is represented in this way and that these hydrophobic clusters correspond significantly to elements of secondary structure, particularly α-helices and β-strands (Henrissat et al., 1995).

Hydrophobic clusters are identified by the ‘hydrophobic alphabet’ VILFMYW, as these residues are typically buried within a protein. A minimum ‘connectivity distance’ is also utilised in HCA and is defined as the minimum number of hydrophilic residues which should be present between clusters (Eudes et al., 2007). These clusters are important to protein structure as folding is driven by the requirement to bury hydrophobic regions in globular proteins, and create a hydrophilic protein surface. Thus, the protein folding process begins with hydrophobic clusters (Woodcock et al., 1992). Drawing on this relationship between sequence and structure, HCA allows reliable predictions of tertiary structure, and also permits a comparison of the size, shape and orientation of hydrophobic clusters within divergent protein sequences (Callebaut et al., 1997; Woodcock et al., 1992).

A project using the HCA method was undertaken to classify glycoside hydrolases by sequence similarity and structure (Henrissat, 1991). Structural regions surrounding the catalytic residues can be used to define conserved motifs, and these can be used to predict the location and role of the catalytic residues in structural homologues identified by HCA (Henrissat et al., 1995). Enzyme sequences which align over an entire domain are classified into a single GH family (Henrissat et al., 1995). As the method is sufficiently powerful to reveal structural similarities even when primary sequence identity is very low, predictions of structure and, in some instances, function can be made even when sequence similarity is barely detectable by conventional methods of sequence analysis (Callebaut et al., 1997). An early success of the project was the classification of 21 cellulase genes into six sub-families with shared structural fold, as identified by HCA of primary sequence (Henrissat et al., 1989). Within a family, glycoside hydrolases share a common structural fold, catalytic residues and catalytic mechanism. These classifications are shown in the publicly accessible Carbohydrate Active Enzymes (CAZy) database (http://www.cazy.org/, (Cantarel et al., 2009)).

At the time of writing, there are 124 families of glycoside hydrolases, encompassing more than one thousand enzyme sequences. Many families appear to exhibit one or a very few activities, exemplified by families GH10 and GH11, where only xylanases have been reported. The high degree of structural similarity within families predicts a limited range of specificities, but polyspecific families do exist, such as family 43. Within these families, the acquisition of new substrate specificities is a common evolutionary event (Henrissat, 1991). Families with low sequence similarity which share a common fold, catalytic apparatus and catalytic mechanism are grouped into clans. There are currently fourteen clans (A-N). The families within a clan, similar to members within a family, likely have common ancestry, leading to significant similarities in structure, including conservation of catalytic residues and mechanism (Henrissat and Bairoch, 1996).

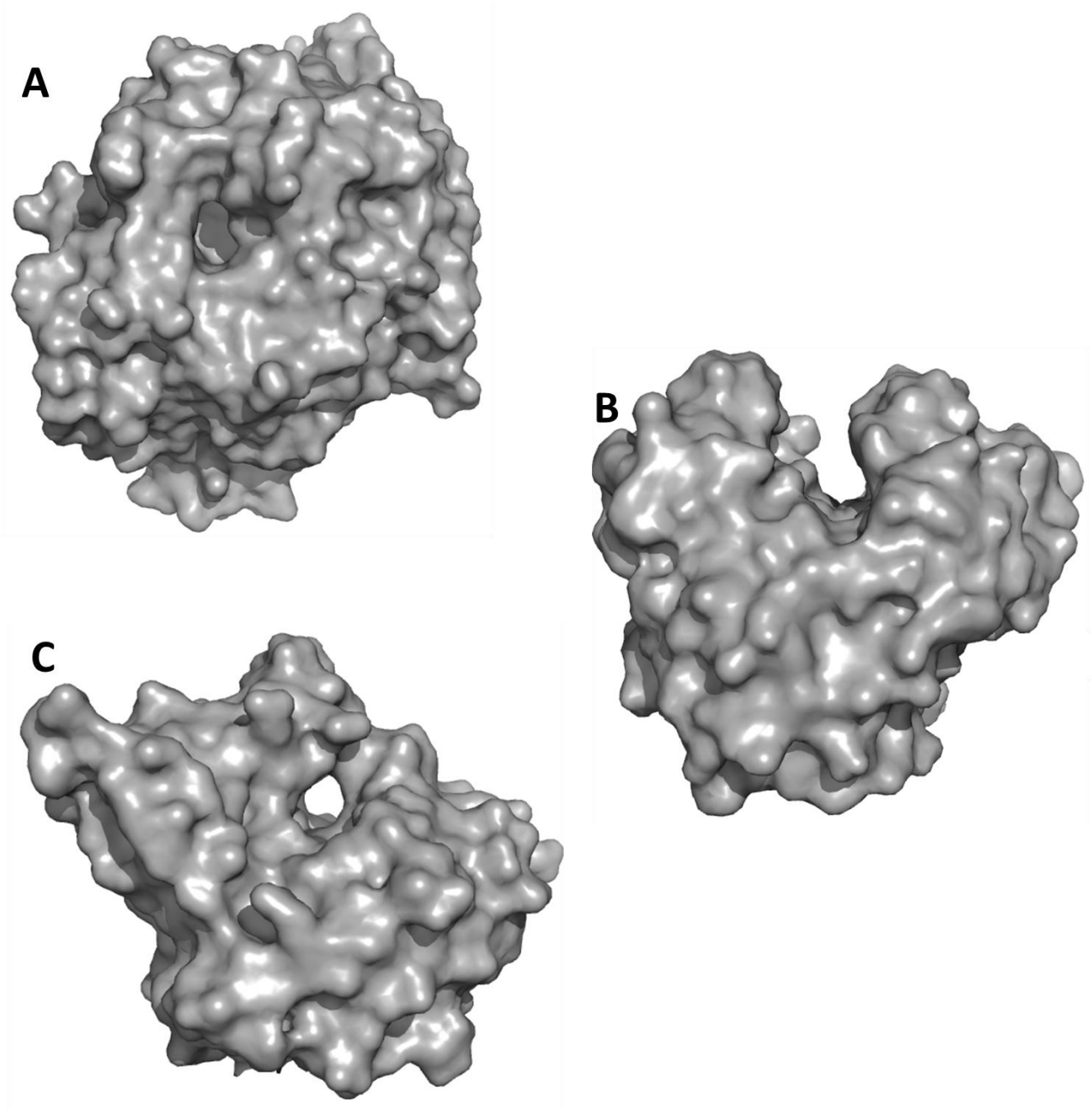


Figure 1.7 Three structural types of active site found in glycoside hydrolases.

A: the pocket topology binds a single sugar residue, as in *exo*-acting enzymes and side-chain cleaving hydrolases. Taken from an *Aspergillus awamori* GH5 glucoamylase (PDB code 1AGM) (Aleshin et al., 1994).

B: *endo*-acting enzymes may have a cleft topology. The polysaccharide will lie relatively flat within this cleft. Taken from the *Thermobifida fusca* GH6 endoglucanase Cel6A (PDB code 1TML) (Spezio et al., 1993).

C: certain processive cellulases possess a tunnel topology, allowing the substrate to be cleaved and product to be released without the need to release the substrate. Taken from a *Mycobacterium tuberculosis* GH6 endoglucanase (PDB code 1UOZ) (Varrot et al., 2005).

Figure 1.7 shows the three general topologies of the catalytic domains of glycoside hydrolases. The pocket topology (Figure 1.7) is found in enzymes targeting monosaccharides, such as β-galactosidases, and *exo* acting enzymes which target polysaccharides. The shallow active site depression accommodates a single sugar residue. The cleft topology (Figure 1.7) is a more open structure which can bind several sugar residues. Enzymes with this structure are usually *endo* acting polysaccharidases, such as xylanases. The tunnel topography has developed from the cleft, part of which is covered by a long loop (Figure 1.7). The substrate polymer is threaded through the tunnel. This allows the enzyme to release the product of hydrolysis while maintaining tight binding to the substrate, resulting in a processive mode of action (Davies and Henrissat, 1995). Most enzymes with the tunnel topography are cellulases, perhaps reflecting the need to prevent the substrate glucan chain from re-annealing to the cellulose microfibril. This processive type of activity may require an accompanying *endo* activity, from the same enzyme or a second, cooperative enzyme, to cleave the cellulose fibril, creating a chain end for the tunnel to bind (Davies and Henrissat, 1995).

1.2.2 Catalytic mechanisms

Glycoside hydrolases function via one of two possible catalytic mechanisms. The retaining mechanism results in a net retention of stereochemistry at the anomeric carbon of the glycon sugar, while the inverting mechanism leads to a net inversion of anomeric configuration. Both mechanisms require, generally, two carboxylic acids in the active site and proceed via a transition state with significant oxocarbenium character (Kelly et al., 1987, McCarter and Withers, 1994). Both mechanisms also require the formation of a transition state with significant positive charge, which is delocalised between the anomeric carbon and the endocyclic oxygen as the C1-O5 bond acquires partial double bond character (Karaveg et al., 2005). This requires that the pyranose sugar ring is distorted from the low free-energy $^{4}C_{1}$ chair conformation (Figure 1.8) to bring C5, O5, C1 and C2 into a shared plane. Four possible conformations satisfy this requirement for co-planarity: the $^{2,5}B$ and $B_{2,5}$ boat conformations and the $^{4}H_{3}$ and $^{3}H_{4}$ conformations (Figure 1.8). A large body of evidence, taken from crystal structures of enzymes with trapped intermediates shows that these conformations are employed by GH families (Barnett et al., 2010;

Ducros et al., 2002; Karaveg et al., 2005; Vocadlo and Davies, 2008; Zhu et al., 2010b). It is believed that substrate distortion in this manner helps to lower the energy barrier of the hydrolysis reaction by moving the substrate closer to the structure of the transition state and drawing the glycosidic oxygen into a more favourable position for protonation (Davies and Henrissat, 1995; Zechel and Withers, 2000). Substrate distortion is generally driven by enzyme-substrate interactions at the +1 subsite.

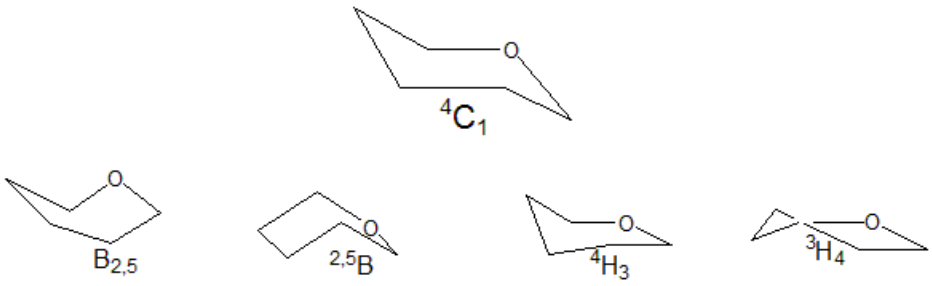
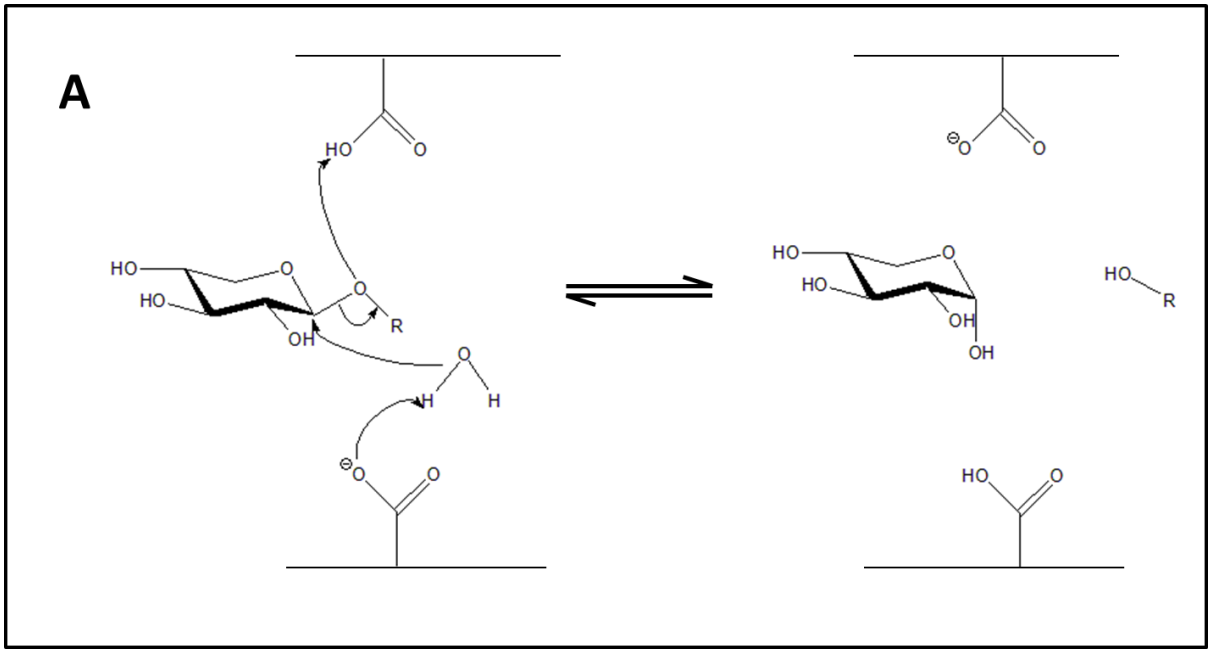
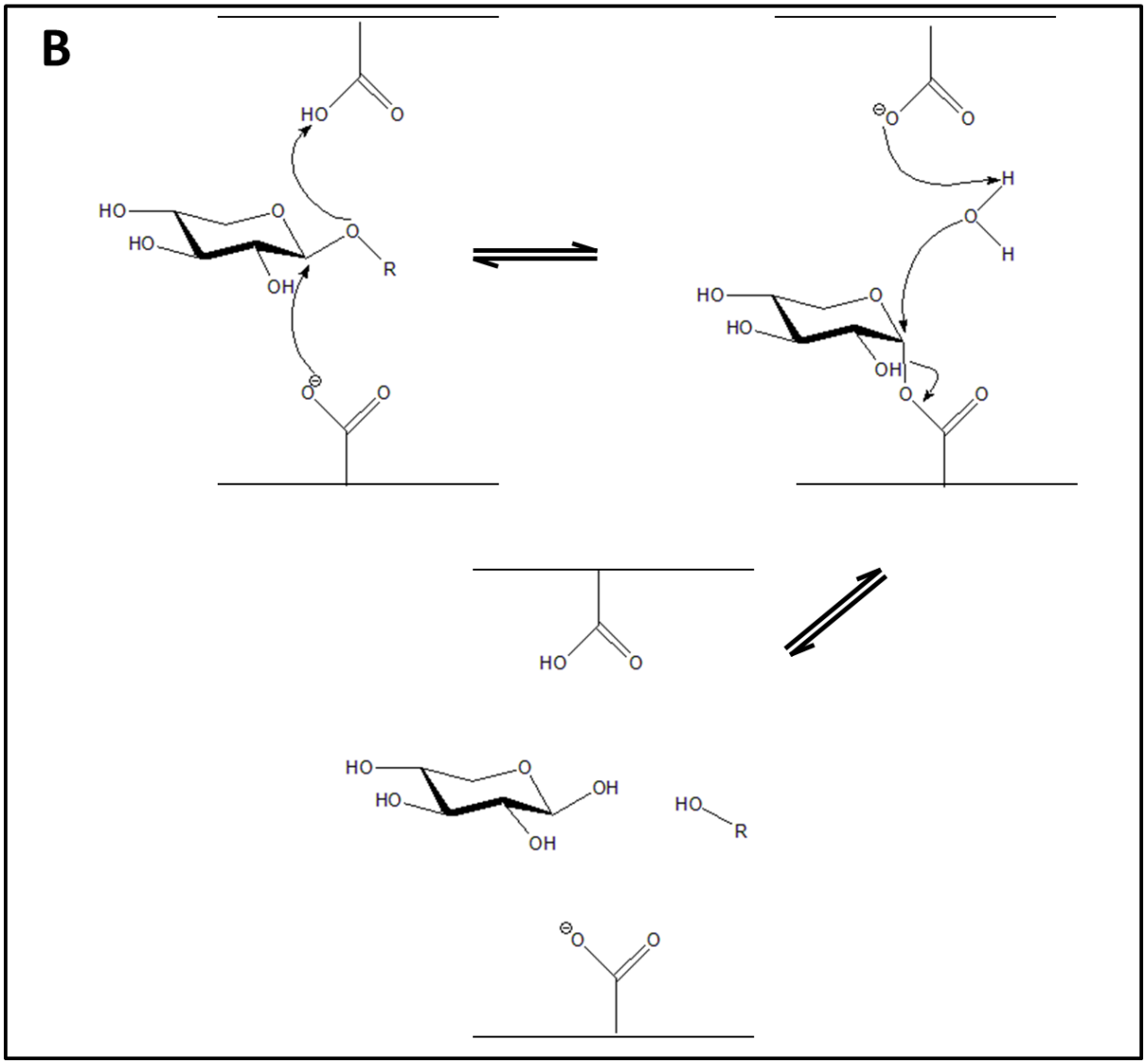


Figure 1.8 Interconversions between pyranose sugar ring conformations.

The partial map of pyranose sugar ring conformations shows the low free energy $^4C_1$ chair and various boat (B) and half-chair (H) conformations which have been shown by x-ray crystallography to be employed by glycoside hydrolases in catalysis. The endocyclic ring oxygen is shown by the O in each structure. The boat and chair conformers shown have a structure in which C-5, O-5, C-1, and C-2 are co-planar. Reproduced from (Karaveg et al., 2005).

The inverting mechanism of glycoside hydrolases (Figure 1.9) involves direct displacement of the leaving group by a water molecule. A general base activates a water molecule and subsequently a general acid protonates the leaving group (aglycon). As shown in Figure 1.9, the charged general base activates a water molecule, which launches a nucleophilic attack on the anomeric carbon. The glycosidic oxygen is then protonated by the general acid, and the aglycon departs (McCarter and Withers, 1994). The acid and base are spaced ~ 10 Å apart to allow space for a water molecule. This single nucleophilic substitution yields a product with opposite stereochemistry to the substrate.


A
HO
O
HO
HO
O
R
OH
O
H
H
O
O
HO
O
HO
HO
OH
OH
HO
R
HO
O


B
HO
O
HO
HO
O
R
OH
O
O
H
O
H
HO
O
HO
OH
O
O
HO
O
HO
HO
OH
OH
HO
R
O
O

[23]

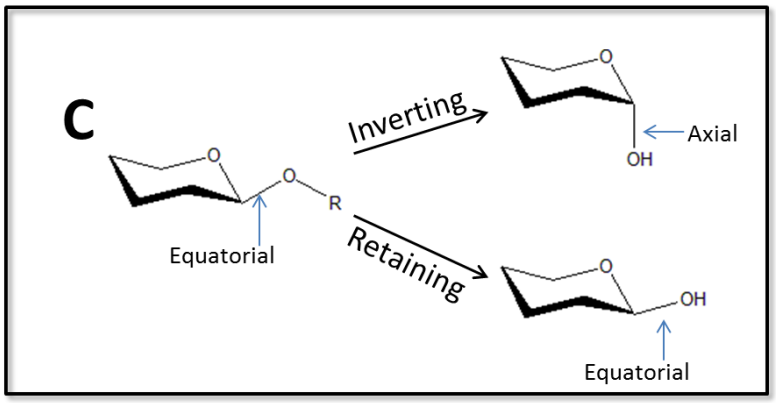


Figure 1.9 The inverting and retaining mechanisms of glycoside hydrolases.

The generalised sugars shown are in the $^4C_1$ conformation.

A: the inverting mechanism of glycoside hydrolases proceeds via a single displacement step and results in a product with the opposite anomeric configuration from the substrate.

B: the retaining mechanism generates a product with the same configuration as the substrate and proceeds via a double displacement mechanism.

C: the two mechanisms of glycoside hydrolases result in either an overall inversion or overall retention of anomeric configuration.

The retaining mechanism differs in that it proceeds via a double displacement mechanism which includes a covalent glycosyl-enzyme intermediate (Figure 1.9). The two catalytic amino acids function as a general acid/base and a nucleophile. The nucleophile directly attacks the anomeric carbon, leading to the formation of a covalent intermediate while the acid/base protonates the glycosidic oxygen and thus assists leaving group departure (Figure 1.9). The glycosyl-enzyme is hydrolysed by water activated through proton extraction by the catalytic acid/base, leading to a second nucleophilic substitution at the anomeric carbon. The glycosidic oxygen is protonated and the aglycon departs. The two carboxylate residues are ~ 5.5 Å apart. This double displacement mechanism gives rise to a product with the same stereochemistry as the substrate (McCarter and Withers, 1994), Figure 1.9).

1.2.3 Carbohydrate binding modules

The inaccessibility of insoluble polysaccharides within the complex interlocking matrices of the plant cell wall reduces the efficiency of many glycoside hydrolases.

For this reason, many of these enzymes are appended to one or more carbohydrate binding modules (CBMs) in order to promote enzyme-substrate association.

CBMs are classified in the CAZy database in a similar manner to that used for enzymes. CBMs have been described which bind a wide variety of ligands and perform several possible functions by which they facilitate enzyme catalysed hydrolysis of the glycosidic bond (Boraston et al., 2004). Firstly, the sugar binding activity of CBMs assists enzyme catalysis by concentrating the enzyme on the surface of the polysaccharide substrate, leading to more rapid degradation. These CBMs do not always bind to the target substrate of their associated enzyme; for instance, there are CBMs which bind crystalline polysaccharides, principally cellulose, which can be appended to enzymes with a variety of specificities (Boraston et al., 2004). In some cases, it is thought that a CBM may directly disrupt the structure of a polysaccharide, thereby facilitating enzyme access by increasing the availability of the substrate. However, the extent to which this phenomenon can be generalised is unclear (Boraston et al., 2004).

1.3 Glycoside hydrolase family 43

1.3.1 Structural features and mechanism

The first structure of a GH43 enzyme was the *Cellvibrio japonicus* α-L-arabinanase *Cj*Arb43A (Nurizzo et al., 2002). The enzyme hydrolyses linear arabinan into arabinotriose in a chain-end *exo* fashion, but also has some weak *endo* activity, suggesting a processive mode of action following initial cleavage within the chain. The crystal structure of the enzyme was solved to 1.9 Å and revealed a five-bladed β-propeller structure which had not previously been seen in an enzyme. The propeller is built from five β-sheets, which are highly twisted and arranged radially around the central axis (Nurizzo et al., 2002). The propeller is strengthened by hydrogen bonds between the first and fifth blades. The axial cavity forms a pocket which is filled with water and is situated at the centre of a long V-shaped cleft which extends over the surface of the enzyme (Nurizzo et al., 2002) Figure 1.10). This long depression, which has a right-angled cleft to accommodate the twisted arabinan chain, provides six sugar binding sites.

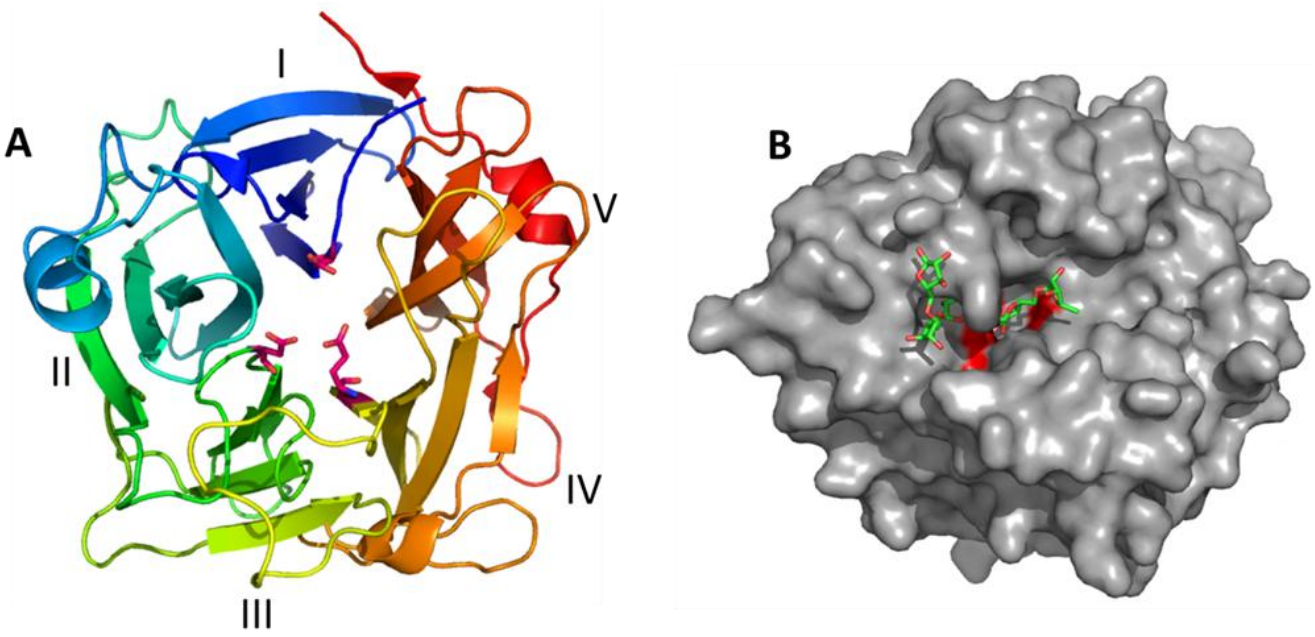


Figure 1.10 The canonical GH43 structure of *Cj*Arb43A.

A: the crystal structure of *Cj*Arb43A reveals a five-bladed β-propeller with a central cavity housing three catalytic residues, shown in stick form and coloured red.

B: the active site, and catalytic residues, of *Cj*Arb43A lie within a curved cleft over the surface of the protein, shown here in complex with arabinohexaose.

The glycosidic bond is cleaved between sugars at subsites -1 and +1, according to the naming system of Davies and colleagues (Davies et al., 1997), and the *Cj*Arb43A subsites extend from -3 to +3. The active site (-1 subsite) houses a constellation of acidic amino acids (Asp38, Asp158 and Glu221). The Glu221 is adjacent to the glycosidic bond between arabinofuranosyl residues at -1 and +1, ideally positioned to act as the catalytic acid. Asp38 is 6 Å from the anomeric carbon of the sugar residue at -1, where it functions as the catalytic base by activating a water molecule. The distance between the catalytic acid and base is ~ 7.2 Å, less than previously described for inverting enzymes. At the time of publication, the function of Asp158 was uncertain (Nurizzo et al., 2002), but the residue was thought to be involved in pKa modulation of the catalytic acid.

Subsequent crystal structures and alignments have shown that all GH43s possess this triad of carboxylate residues in the active site (Pons et al., 2004). The structure of XynB3, a β-xylosidase from *Geobacillus stearothermophilus* revealed the pKa modulating role of the additional Aspartate (Brux et al., 2006). The enzyme

comprises the canonical GH43 β-propeller and an additional C-terminal β-sandwich domain. The active site of the enzyme has a pocket topology, within which are found a triad of catalytic residues: Asp15, Asp128 and Glu187. These residues point to the centre of the axial cavity, as described for the *C. japonicus* arabinanase. The structure of XynB3 was solved in complex with xylobiose, which shed light on the role of each of these amino acids.

Asp15 is perfectly situated to act as catalytic base and activate a water molecule in the inverting single displacement mechanism. The residue is located 5.2 Å from the anomeric carbon of the xylose at the -1 subsite (glycon sugar, Figure 1.11). Glu187, the catalytic acid, is positioned ~ 2.6 Å from the glycosidic oxygen, suitable for protonation of the departing aglycon sugar (+1 subsite). Asp15 and Glu187 are nearly perpendicular, forming a 100 ° angle with the anomeric carbon of the glycon sugar (Brux et al., 2006). The distance between Asp15 and Glu187 is ~ 7.5 Å.

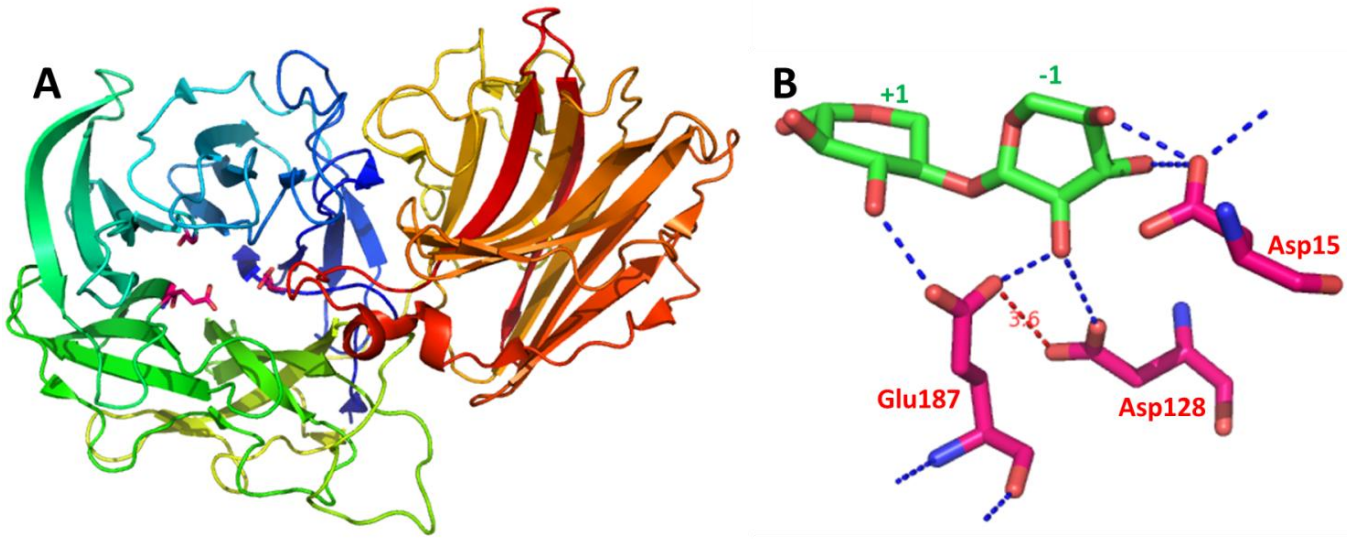


Figure 1.11 The crystal structure of XynB3 in complex with xylobiose reveals the role of the second Aspartate in five-bladed β-propeller glycoside hydrolases.

A: cartoon representation shows that the *Geobacillus stearothermophilus* comprises the β-propeller module and an additional β-sandwich module. The catalytic residues are coloured red and shown in stick form.

B: the catalytic triad is shown in stick form with xylobiose. All three amino acids make direct contacts with the xylose at -1, including Asp128; these contacts are indicated by dashed blue lines. Asp128 is also within range (3.6 Å – shown by the dashed red line) of the catalytic acid (Glu187) to modulate the pKa of that residue and orient it correctly with respect to the substrate.

The third catalytic residue, Asp128, is situated ~ 4 Å from Glu187. The pKa of Glutamate is generally 4.07, but needs to be much higher to function as the catalytic acid; the proximity of Asp128 raises this pKa to 7.1 and also orients the acid correctly relative to the substrate (Brux et al., 2006). Structures of the enzyme in complex with xylobiose showed that Asp128 also has a role in substrate binding and transition state stabilisation, making multiple hydrogen bonds with the -1 xylose. This explains the invariance of the second active site Aspartate in five-bladed β-propeller glycoside hydrolases (Families 43 and 62 in Clan F and Families 32 and 68 in Clan J). Mutation of this residue has shown that it is essential for activity in many GH43 enzymes (Alhasid et al., 2009; McKee et al., 2011; Shallom et al., 2005). Crystal structures of mutant forms of XynB3 show that in the absence of Asp128 the general acid rotates to an unfavourable position; the general base is apparently unaffected (Brux et al., 2006).

1.3.2 Major activities

Glycoside hydrolase family 43 (GH43) is polyspecific. Activities identified to date include α-L-arabinofuranosidase, β-D-xylosidase, β-D-galactosidase, arabinanase and xylanase. As of early 2011, twenty crystal structures have been deposited for this family. All share the same β-propeller structure for the catalytic module, and the triad of catalytic residues is conserved in all active enzymes. Such a wide variety of activities within a single family of enzymes may not be common, but nor is it surprising; the stereochemical similarities between substrates means that often only a small structural change can lead to the acquisition of a new specificity (Henrissat, 1991).

1.3.3 Familial expansion

A notable feature of family GH43 is the extent to which it is expanded over other families in very many microbial species, derived from a wide variety of habitats. Figure 1.12 shows the family distribution of the complement of glycoside hydrolases encoded by the genomes of six bacterial species, and gives brief notes on each species. It is clear that the expansion of the family is not limited to one similar group of species, or to one natural environment. A hypothesis is that this expansion in a family of enzymes with multiple specificities improves the totality of degradation of

carbohydrate material, as a range of enzymes with highly varied hydrolytic activities seems useful in matching the wide range of linkages within the plant cell wall.

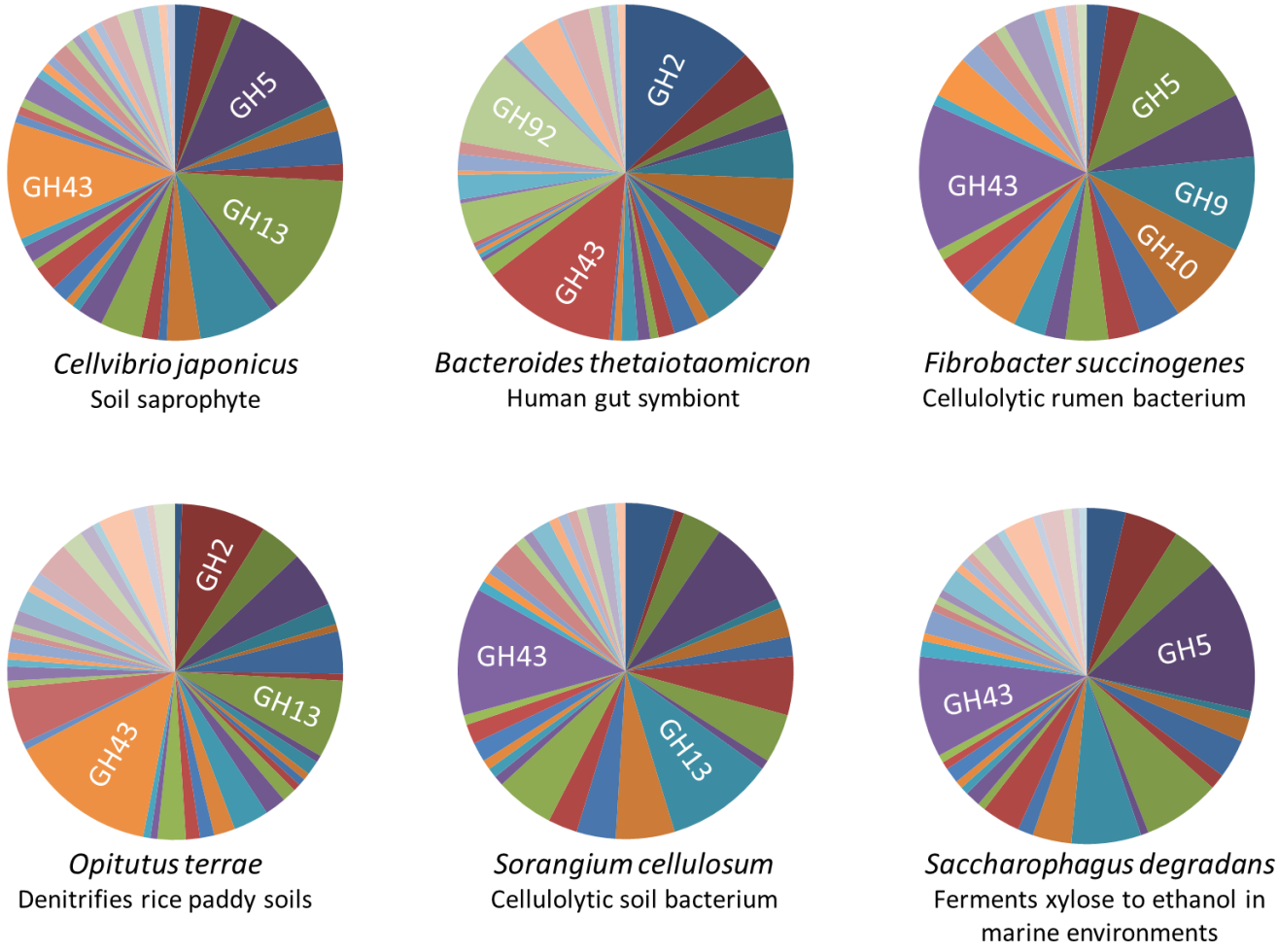


Figure 1.12 Glycoside hydrolase family distribution in six bacterial species.

The pie charts show that family GH43 is expanded in several different species from many different habitats. The size of the segments reflects the number of glycoside hydrolases encoded in each family, according to the CAZy database. Families with zero members are not shown.

1.4 Symbiotic microbiota of the human gut

The average adult human body is built of approximately 10 billion (10^{10}) cells. The gut microbiota of an individual can represent ~ 1.5 kg of aggregate biomass (Xu and Gordon, 2003b), reaching cell densities of 10^{11} - 10^{12} in the gut, the highest recorded for any ecosystem (Ley et al., 2006). This means that there can be as many as ten times more microbial cells residing within the gut of a healthy individual than all of

the somatic and germline cells which comprise the body of the individual. The 'metagenome' of this microbial community includes approximately one hundred times more genes than the human genome. These staggering numbers contribute to an emerging picture of the extent to which our lives are interconnected with an invisible ecosystem within us.

Underlining the importance of the gut microbiota to human health and development is the intensity of research currently focussing on this area, occurring in academic, clinical and industrial institutions. A recent National Institute of Health (NIH) initiative is the Human Microbiome Project (http://commonfund.nih.gov/hmp/, Fujimara et al., 2010). This project is studying the extent to which changes to the microbiota are associated with health and disease, and are making developments in many areas, including metagenomics (Lazarevic et al., 2009). These provide a broad genetic perspective of a microbial community using genetic material extracted from an environmental sample (the metagenome). These and other efforts are beginning to provide a picture of the symbiotic relationship between humans and the intestinal bacterial ecosystem, and the benefits it brings to both partners.

1.4.1 Colonisation of the gut

Colonisation of the sterile post-natal gastro-intestinal (GI) tract begins very soon after birth. Bacteria are acquired from the mother and then from the surrounding environment. The colony stabilises after around one month, or six months if birth was achieved by Caesarean section, perhaps hinting at one mode of mother to baby transmission (Ley et al., 2006). Initially, *Escherichia coli* and *Streptococcus* species dominate; these create a reducing environment for species of *Bacteroides*, *Bifidobacterium*, *Clostridium* and *Ruminococcus*, obligate anaerobes which dominate the community for most of our lives. In general, both cell density and biodiversity increase as the digestive system proceeds from the small intestine (~ 10^3 organisms per millilitre of lumen contents) to the colon (~ 10^{11} organisms per gram of contents).

This host-bacterial symbiosis is well established and has been observed in animal species. In fact, it is hypothesised that bacteria may have been helping to shape eukaryotic evolution in this way for around 10^9 years (Xu and Gordon, 2003b). The symbiotic relationship between gut microbiota and their host is most often studied

using gnotobiotic germ-free mice colonised with a specific bacterial species during or after post-natal gut development and fed from birth on a strictly controlled diet (Xu and Gordon, 2003b).

1.4.2 Species distribution of gut microflora

The adult human GI tract is populated by bacteria, archaea and some eukaryotes. The bacterial portion of the community is particularly abundant, representing a dense and diverse ecosystem (Backhed et al., 2005). There are a set of requirements which must be met to qualify for residency in the human intestine. To be a member of the gut microbiota, a species must possess a suite of glycoside hydrolases targeted to the available carbohydrates and be able to attach to surfaces where appropriate nutrient sources are abundant. The genetic ability to mutate quickly and adaptively is important, and rapid growth is vital to avoid the loss of a whole cell lineage by washout (Ley et al., 2006).

While there are generally in the region of five hundred bacterial species within the gut, 99 % of all cells are from just 30 – 40 anaerobic species. There are representatives of only eight divisions of bacteria, compared with over twenty found in soil. The population is more diverse at the level of species (Backhed et al., 2005). Variation between individuals in the precise make-up of the gut microflora is high, but one person's intestinal community remains remarkably constant over their lifetime. This suggests that there are mechanisms to suppress blooms of sub-populations, or conditions which favour cooperativity (Backhed et al., 2005). The stability of the gut ecosystem depends upon the ability of its members to diversify their responses to a fluctuating environment (Chang et al., 2004). The prominent sources of nutrition in the gut are undigested plant polysaccharides and glycoproteins, and host glycans associated with the epithelium (Xu et al., 2007). The food web in the gut must remain stable to ensure the longevity of the community, so species are often able to adapt to changes in the range of carbohydrates which are available in the gut. This adaptability is often ensured by functional redundancy encoded in the genomes of species (Xu et al., 2007). A good example is *Bacteroides thetaiotaomicron*, whose genome codes for 981 proteins involved in polysaccharide metabolism (Xu et al., 2007).

1.4.2.i Microbiota of the infant gut

As discussed earlier, the human gut is initially colonised by species such as *E. coli*, which create a reducing environment which is later dominated by anaerobic bacterial species. In infants, the dominant species are *Bifidobacteria.* These are gram positive staining anaerobic microorganisms which can represent up to 90 % dry weight of faecal matter from breast-fed babies but make up only 3 – 6 % of adult faeces. This is likely because some species, particularly *Bifidobacterium infantis*, can ferment oligosaccharides found in human milk. In young babies, these bacteria are important in regulating microbial homeostasis in the budding gut community, and modulate the response of the immature immune system to non-pathogenic species of bacteria (Gronlund et al., 2007). *B. infantis* is the dominant species in the GI tracts of new-born babies and is also found abundantly in yoghurt and other fermented dairy products, reflecting the primary metabolic targets of the species. This species, as well as *Bifidobacterium adolescentis*, the gut population of which is stable from birth until late adulthood, have been added as probiotics to dairy products, and have been shown to be capable of inducing dramatic reductions in the symptoms of irritable bowel syndrome (IBS). These reductions may be mediated via the anti-inflammatory properties of these bacteria (Brenner and Chey, 2009; Whorwell, 2009). The use of probiotics in diet modification is discussed in greater detail below.

1.4.2.ii Microbiota of the adult gut

As we age, and our diet changes, the primary source of carbohydrate available to the gut microbiota shifts from human milk oligosaccharides to plant polysaccharides, as we begin to consume fruits, vegetables and so on. This leads to a shift in the gut population away from species of *Bifidobacteria*, as new metabolic niches develop due to the new variety of carbohydrate nutrition available. The species which take over are more adapted to a wide variety of polysaccharides, and are dominated by *Bacteroides* species of bacteria, which comprise around 30 % of all gut bacteria and have been found at densities of 10^{10} – 10^{11} cells per gram of human faecal matter. *Bacteroides* is a genus of gram-negative, rod-shaped obligate anaerobes. These glycophiles can utilise simple sugars when available. However, most simple sugars do not occur at significant levels in the large intestine as the host absorbs all but ~ 2

% of these in the small intestine (Sonnenburg et al., 2010). The main energy source for *Bacteroides* in the gut is therefore plant polysaccharides from the diet for which humans do not possess appropriate degradative enzymes. The bacteria can ferment carbohydrates and transform bile acids and other steroids.

The dominant *Bacteroides* species in the gut is *Bacteroides vulgatus* which mainly utilises starch. A more versatile user of polysaccharides is *B. thetaiotaomicron*, whose proteome provides the structural means to specifically endocytose and hydrolyse polysaccharides indigestible by the host, systems which work in concert with an extensive environmental sensing network of outer membrane proteins (Reeves et al., 1997). These carbohydrate binding and hydrolysing enzymes allow *B. thetaiotaomicron* to adaptively forage for glycans in the gut, which stabilises the intestinal food web and prolongs the community as a whole (Sonnenburg et al., 2005). The mechanisms by which *B. thetaiotaomicron* performs these vital tasks are discussed in greater detail below. Briefly, a series of cell-associated enzymes hydrolyse polysaccharides to monosaccharides and oligosaccharides, which are translocated into the periplasm, where they are subjected to further hydrolysis, with the terminal products transported into the cytoplasm and further metabolised.

1.4.3 Human – microbe symbiosis in the gut

The gut microbiota can be considered a single discrete entity, a multi-functional organ with differentiated cell lineages working cooperatively to provide metabolic capabilities not provided by the human genome, including digestion of plant polysaccharides, transformation of bile acids and synthesis of some vitamins, such as Vitamin K (Xu and Gordon, 2003b). These cell lineages can be considered cooperative, as competition is minimised by the development of metabolic niches within the intestine.

The primary function served by the gut microbiota is to ferment otherwise indigestible energy substrates from our diet. The 6.3 Mb genome of *B. thetaiotaomicron* codes for 256 glycoside hydrolases and 16 pectate lyases. In contrast, the 2.85 Gb genome of *Homo sapiens* codes for just 95 glycoside hydrolases; we are particularly deficient in enzymes which can degrade xylan and arabinose-containing pectic polysaccharides, major components of soluble fibre (Backhed et al., 2005).

Bacterial degradation of recalcitrant plant polysaccharides provides the human host with carbon and energy, while the bacteria are provided with a rich glycan diet and a protected anoxic environment (Backhed et al., 2005). It is now becoming clear that the symbiosis brings us many other health-promoting benefits. For instance, these cells train our immune system to respond only to pathogenic species, a tolerance which develops in infancy. The bacteria also have a role in regulation of gut development.

1.4.4 Effects on human health and fitness

Many of the benefits to humans of the gut microbiota are related to utilisation of dietary material. Germ-free rodents, whose GI tracts are sterile, are required to consume approximately 30 % more calories to maintain the same weight as properly colonised mice (Sears, 2005). Importantly, our gut bacteria can ferment short chain fatty acids (SCFAs) including acetic acid, propionic acid and butyric acid (Guarner and Malagelada, 2003). This allows these materials to be utilised as an energy source for humans, providing as much as 10 % of our daily caloric intake, and aids absorption of cations such as Ca^{2+}, Mg^{2+} and Fe^{2+}. Furthermore, fermentation of SCFAs lowers gut pH, which broadly favours non-pathogenic species of bacteria (Beaugerie and Petit, 2004).

Our symbiotic partners in the gut also aid digestion by directly influencing the functionality of the intestine. There is evidence that certain species have a role in turnover and differentiation of epithelial cells, possibly mediated by making changes to glycan cell surface markers (O'Hara and Shanahan, 2006).

The gut microbiota can protect us from disease by out-competing pathogenic species for nutrition and attachment sites. This is known as a barrier effect (Guarner and Malagelada, 2003; Ohland and Macnaughton, 2010). Competition prevents colonisation by yeast and bacterial species such as *Clostridium difficile,* which can cause intestinal disease and diarrhoea, and which often strikes patients who are compromised when ‘friendly’ gut bacteria are knocked down by broad-spectrum antibiotics (Beaugerie and Petit, 2004; O'Keefe, 2010). Working in concert with this barrier function, the gut bacteria are thought to stimulate lymphoid tissue associated with the mucosal epithelium of the intestine to produce antibodies.

Many of these commensal species of bacteria can themselves be pathogenic in certain circumstances. For example, perforation of the gut by injury or surgery can allow members of the microbiota to invade the rest of the body, which can be potentially fatal (Goldstein et al., 2009; Simmon et al., 2008).

1.5 Pectin utilisation by gut microorganisms

Pectin (see Section 1.1.2.iii, Figure 1.6) is a major component of the plant cell wall and of soluble dietary fibre, a viscous composite material which also comprises arabinoxylans and cellulose. Pectin is not digestible by humans, as we lack the glycoside hydrolases and lyases required to degrade these polysaccharides (www.cazy.org). However, members of the gut microbiota can degrade pectin, fermenting the sugars released into short chain fatty acids (SCFAs) (Nofrarías et al., 2007).

1.5.1 *Bacteroides thetaiotaomicron*

Bacteroides thetaiotaomicron is a major component of the adult intestinal microbiota and is studied as a model gut symbiont. It degrades plant polysaccharides which are indigestible by the host and is implicated in processes such as stimulation of gut angiogenesis. *B. thetaiotaomicron* becomes dominant in the gut after a dietary switch is made from mother's milk to a diet rich in plant material, which is abundant with indigestible polysaccharides (Xu and Gordon, 2003b).

Unlike other species, *B. thetaiotaomicron* does not possess adhesive organelles (Figure 1.13); this lack could be detrimental in a fluctuating environment such as the gut where localisation close to a nutrient source is vital for success. Food acquisition is instead achieved by binding of glycan-specific outer membrane proteins to polysaccharides. Attachment in this way can prevent wash out, and increases the efficiency of glycan harvesting.

Figure 1.13 Electron micrograph of *Bacteroides thetaiotaomicron*

A highly magnified view of a *Bacteroides thetaiotaomicron* cell, showing the tightly woven polysaccharide capsule. The composition of this capsule is coordinated with the glycans available as carbon source. Reproduced from (Martens et al., 2009c).

Another way in which *B. thetaiotaomicron* maintains dominance in the gut ecosystem is by adapting its response to changes in the dietary intake of the host. As mentioned above, the ability to adapt to these changes is encoded in bacterial genomes which are often functionally redundant. The variety of glycoside hydrolases encoded by *B. thetaiotaomicron* is high even compared to other *Bacteroides* species, making it an important generalist (Xu et al., 2007). Analysis of the CAZy database reveals that the *B. thetaiotaomicron* genome codes for 256 glycoside hydrolases, 16 pectate lyases and 20 carbohydrate esterases (www.cazy.org). This totals 292 enzymes which can degrade carbohydrates, 64 of which are annotated as targeting xylan and pectin, for which there are no suitable human enzymes (Sonnenburg et al., 2005), although the fact that *B. thetaiotaomicron* is unable to grow on xylan points to a problem with this form of gene annotation. More than half of the cazymes provided by this bacterium are predicted to be extracellular or periplasmic, ensuring access to glycan substrates (Sonnenburg et al., 2005). Such a large number of enzymes would appear to suggest some redundancy in the *B. thetaiotaomicron* genome, but subtle differences

in substrate specificity and cellular location provide a complex suite of enzymes which enable degradation of the highly complex plant cell wall (Xu and Gordon, 2003b). In addition, *B. thetaiotaomicron* possesses 1035 proteins involved in membrane transport and 1813 which constitute carbohydrate-specific environmental sensing complexes, which are discussed below. These allow the bacterium to display adaptive foraging, switching its primary source of nutrition from plant polysaccharides to host glycans in times of scarcity (Sonnenburg et al., 2005).

1.5.2 Polysaccharide utilisation loci

When *B. thetaiotaomicron* is grown on starch, a suite of outer membrane proteins is expressed which are essential for utilisation of the polysaccharide and mid-sized oligosaccharides derived from the polymer. These proteins are encoded by a single operon which has been termed the Starch Utilisation System (Sus) (Figure 1.14). The Sus proteins are glycoside hydrolases, transcriptional regulators and proteins which bind and translocate carbohydrates (Reeves et al., 1997). Transcription of the genes in the Sus operon is induced by growth on starch or maltose.

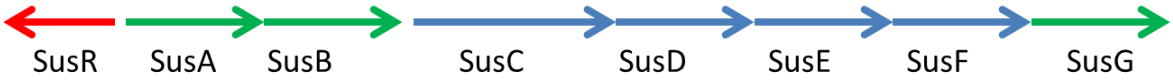


Figure 1.14 The Starch Utilisation System.

The elements of the Sus are encoded by a single operon in the *B. thetaiotaomicron* genome, shown here as a schematic. Glycoside hydrolases are shown in green; SusA and SusG are neopullulanases and SusB is an α-glucosidase. Members of the substrate binding complex (SucC-F) are shown in blue. SusE and SusF have been shown not to be essential members of this complex (Reeves et al., 1997). The regulatory element SusR is constitutively expressed and is shown in red in the schematic.

The SusC/D complex of the Sus operon binds starch and maltooligosaccharides (Xu and Gordon, 2003b). SusC and its homologues are similar to TonB-dependent β-barrel outer-membrane receptor proteins which sense and transmit environmental signals by high-affinity binding and energy-dependent uptake of substrates (Kadner et al., 2003). SusC homologues bind specific elements of carbohydrate substrates and are involved in transport of these glycans into the periplasmic space (Xu et al., 2007). SusD homologues are outer membrane proteins that are attached to a co-

expressed SusC partner (Xu et al., 2007). The *B. thetaiotaomicron* genome reflects the species' capacity to target a wide range of substrates, coding for 163 paralogues of SusC and SusD (Sonnenburg et al., 2005), which bind a variety of carbohydrate targets, increasing efficiency of glycan harvesting via nutrient attachment.

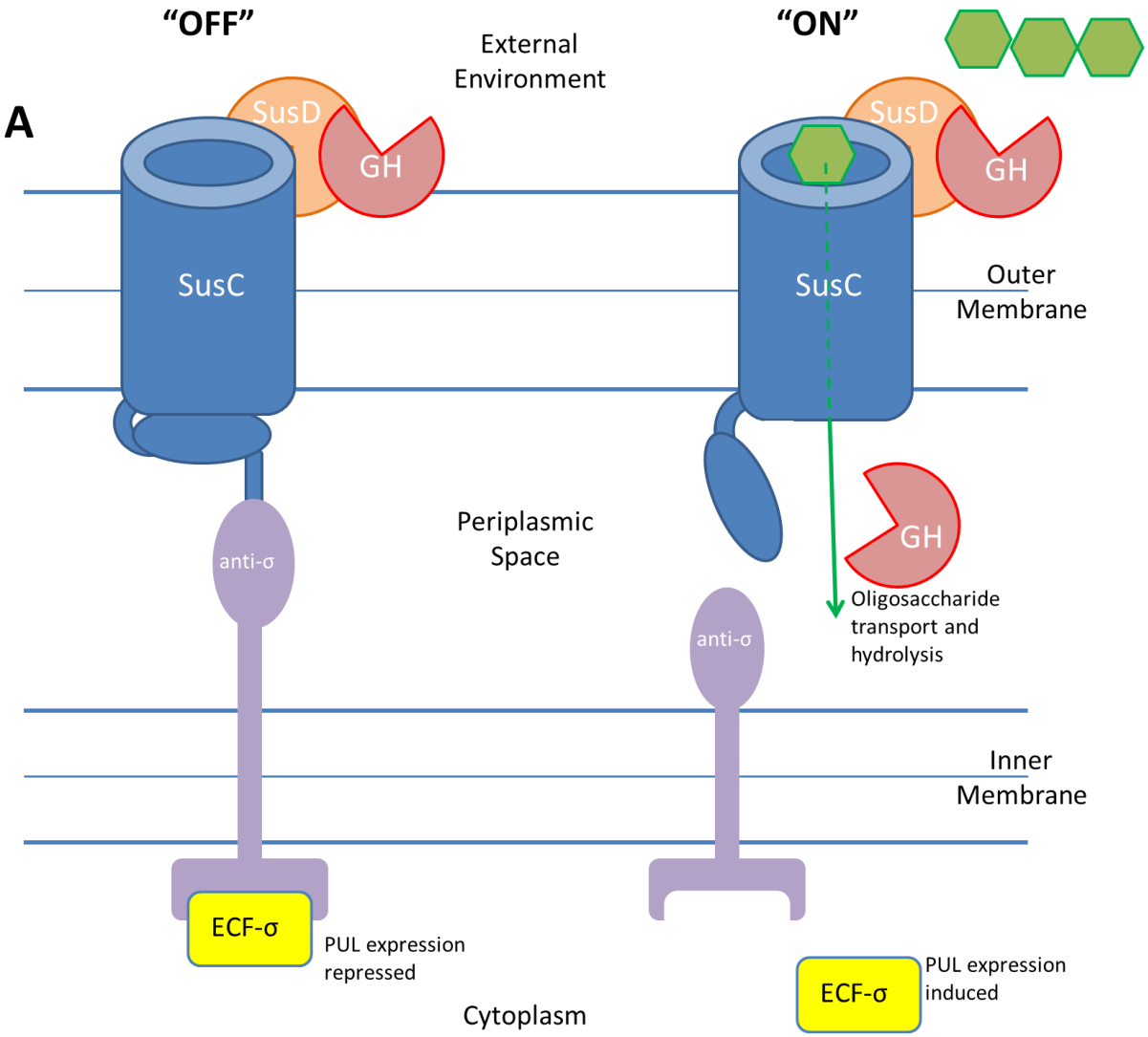
The *B. thetaiotaomicron* genome codes for at least eighty-eight Sus-like systems which target specific carbohydrates. These systems include transcriptional regulatory elements, carbohydrate binding and translocating outer-membrane proteins and glycoside hydrolases, all targeted to the same glycan. These operons, known as Polysaccharide Utilisation Loci (PUL), represent around 18 % of the *B. thetaiotaomicron* genome.

These PUL operons encode the elements required to regulate transcription of the carbohydrate binding and hydrolysis proteins, and are all broadly similar to the Sus system illustrated in Figure 1.14, with the appropriate complements of degradative enzymes and carbohydrate binding complexes. The PUL system is tightly regulated to match the nutrient glycans available in the gut. The *B. thetaiotaomicron* genome includes an unprecedented expansion of two transcriptional regulatory elements, which are often included within a PUL operon, physically linking the modulators of expression with the carbohydrate metabolising proteins (Xu and Gordon, 2003b). First of these elements are hybrid two component systems (HTCS). A single HTCS protein performs the functions of both a sensor kinase and a response regulator, and most have a DNA binding domain. HTCS proteins regulate transcription by binding specific glycans in the environment and transducing the signal to regulate gene expression (Sonnenburg et al., 2006). The second regulatory element common in *B. thetaiotaomicron* is ECF σ and anti-σ factors, which are part of the RNA polymerase complex. The receipt of an appropriate carbohydrate signal causes the ECF σ to dissociate from its respective membrane-bound anti-σ (Brooks and Buchanan, 2008). This initiates expression of binding and hydrolysing proteins specific to the carbohydrate which induced expression (Xu and Gordon, 2003b). For instance, the regulatory element (SusR) in the original Sus system binds maltose and larger oligosaccharides to activate expression of enzymes which degrade starch. Figure 1.15 shows a schematic of this positive feedback model of PUL regulation (Martens et al., 2009c).

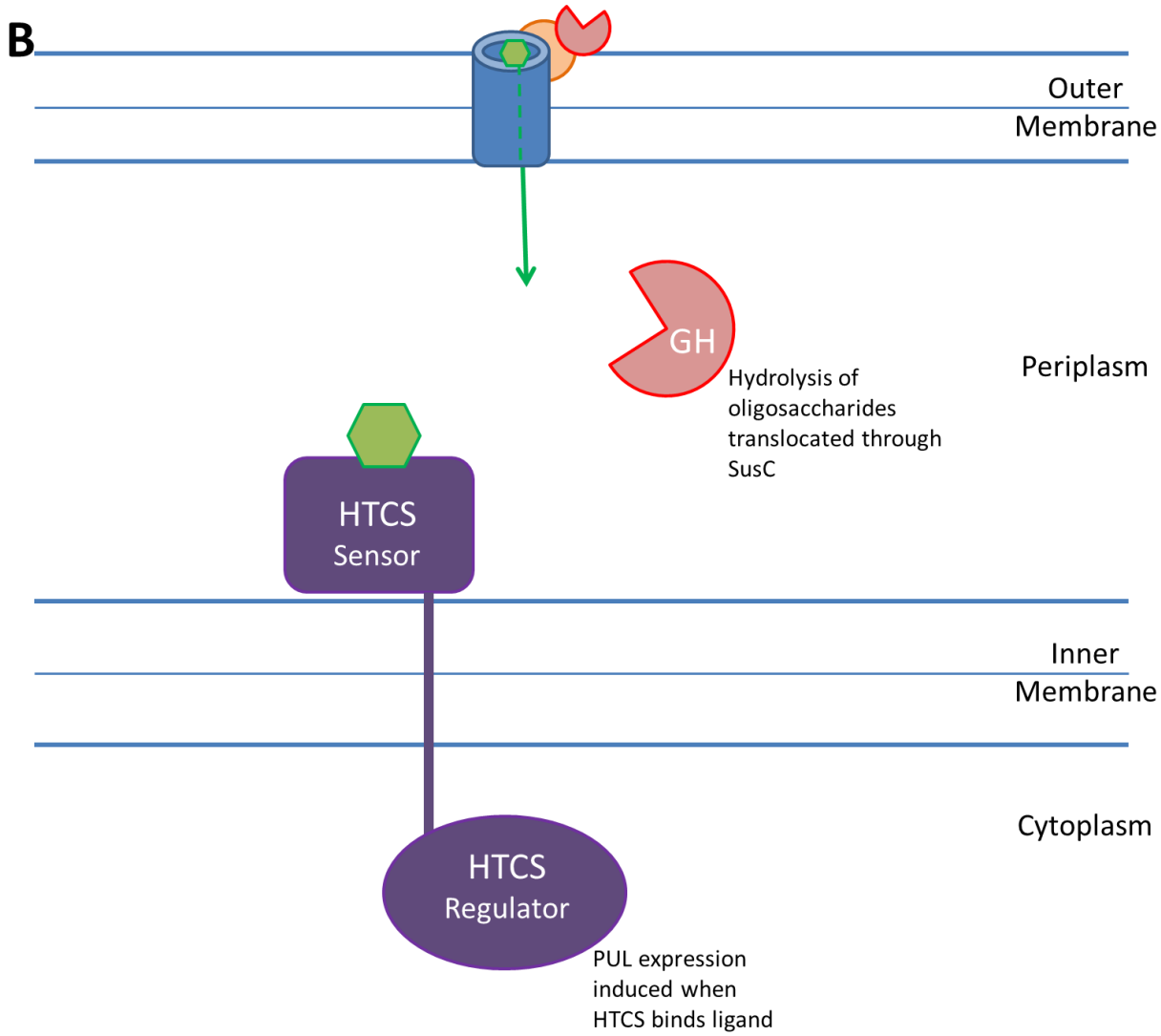
The PULs are important because the glycan binding targets of the HTCS proteins and associated SusC/D complexes determine the substrates which *B. thetaiotaomicron* can use for growth. The bacteria can grow on a variety of pectic polysaccharides as sole carbon source, including homogalacturonan, galactan and arabinan. The Sus proteins often bind elements of these polymers. This system of matching the expression of substrate-specific glycoside hydrolases and SusC/D translocation complexes with the available range of carbohydrates allows *B. thetaiotaomicron* to graze efficiently on nutrients, and to adapt flexibly to changes in the environment (Xu and Gordon, 2003b).

1.5.3 Adaptive glycan foraging

As mentioned above, *B. thetaiotaomicron* is able to forage for glycans in an adaptive manner. When plant polysaccharides are scarce, this species can switch to host glycans as a primary nutrient source. This adaptability has been demonstrated in gnotobiotic mice. Germ-free mice were fed either a polysaccharide-rich diet or a diet of simple sugars, and were colonised with *B. thetaiotaomicron* at seven weeks old (Sonnenburg et al., 2005). Electron microscopy showed that *B. thetaiotaomicron* cells assemble on food particles and are found embedded in the mucosal lining of the intestine, as attachment is required to remain close to an adequate source of nutrition. As expected, genes coding for outer-membrane carbohydrate binding proteins (SusC and SusD) and associated glycoside hydrolases were upregulated in all mice, indicating high expression of elements required for carbohydrate transport and metabolism (Sonnenburg et al., 2005).



A: in the "off" position, constitutive expression of an ECF-σ factor results in a low level of PUL expression which is repressed by the anti-σ factor via physical interaction with ECF-σ and SusC. A complex carbohydrate (polysaccharide) is hydrolysed to a simpler oligosaccharide, which binds SusC. Binding of a specific glycan switches the system to the "on" position. The glycan is transduced by the SusC/D complex. A signal is transmitted which causes the anti-σ factor to release repression of expression of the remaining PUL components. Propagation of the signal results in ECF-σ activation and increased expression of the operon, including glycoside hydrolases.



B: the HTCS is attached to the inner membrane, with domains facing both the periplasm and the cytoplasm. The periplasmic sensor domain recognises a specific glycan, and this signal is transmitted to regulatory elements including a DNA binding module. Expression of the PUL is upregulated.

Figure 1.15 Substrate-dependent regulation of the expression of PUL components.

These schematic diagrams of a hypothetical PUL model how the regulatory element is activated by the binding of a carbohydrate signal from the environment to induce expression of other PUL components. Figure is adapted from (Martens et al., 2009c) and (Sonnenburg et al., 2006). Panel A shows the two states observed for a fictional PUL under the control of an ECF-σ factor. Panel B illustrates the mode of action of an HTCS. Carbohydrates are indicated by green hexagons.

Analysis of caecal contents before and after colonisation with *B. thetaiotaomicron* revealed a marked reduction in the abundance of hexose sugars, as these were selectively metabolised by the bacteria. In addition, host expression of a major

glucose uptake protein increased upon colonisation, as more monosaccharide was liberated from the polysaccharide-rich diet and was utilised by the host (Sonnenburg et al., 2005). GeneChip analysis of the caecal contents of mice which were fed simple sugars prior to colonisation showed that there was no significant expression of glycoside hydrolases which target pectin-like polysaccharides, such as arabinanases, arabinofuranosidases and pectate lyases. There was little difference in cell density compared with those mice fed on a polysaccharide-rich diet, indicating that these bacteria had adapted to the dietary change. Instead, there was upregulation of hexosaminidases, fucosidases and sialidases, which can target N-linked glycans on the surface of host epithelial cells, indicating that the bacteria in these mice were primarily harvesting host glycans (Sonnenburg et al., 2005). The epithelial mucus provides a constant endogenous source of glycans, upon which *B. thetaiotaomicron* can rely in times of scarcity of plant polysaccharides. Species which show this kind of adaptive foraging allow the gut population of an individual to remain constant even when there are significant dietary fluctuations, by allowing the metabolic activity of the community to change (Sonnenburg et al., 2005).

1.6 Applications of plant cell wall components and degradative enzymes

1.6.1 Probiotic diet

Gut bacteria are a focus of research into health management via changes to diet in two main areas: prebiotics and probiotics. Prebiotics are small oligosaccharides which generally serve to promote the growth of specific desirable species. Probiotics are live microorganisms found indigenously in the human gut which are introduced into foods to attempt to manipulate the composition and therefore the functionality of the bacterial gut community (Vanderhoof, 2008). In theory, providing the gut community with a greater capacity to degrade dietary carbohydrates can in turn improve the digestive health of humans. Much recent research has focussed on bacterial members of the gut community and their potential for use as probiotic dietary supplements. Dairy products including bacterial probiotics, most commonly *Bifidobacteria*, are now commonplace in supermarkets, are heavily promoted in the media and are readily available in stores without a prescription. Despite this, a lot of the research supporting specific claims for these products is based on trials on very

few individuals, lacks properly randomised control experiments and is often performed, or funded, by companies with a vested interest in a specific outcome (Brenner et al., 2009). Systematic reviews are available which have assessed a series of publications using tools such as meta-analyses to give a more reliable view on the efficacy of using gut bacterial species as probiotics. *B. infantis* has been shown by such reviews to be genuinely effective in reducing the symptoms of IBS, which is often triggered by dairy products (Brenner et al., 2009). Another very common probiotic is *Bifidobacterium bifidum*, which comprises only 2 % of the gut *Bifidobacteria* community where it occupies the niche of fermenting galactooligosaccharides, the indigestible carbohydrate portion of milk products. Some strains of this species have been shown to reduce the damaging effects of Coeliac disease, possibly via direct interaction with the immune system (Palma et al., 2010).

1.6.2 Biofuels

As discussed at the beginning of this chapter, efforts are underway to commercialise bioethanol transport fuel produced from lignocellulosic biomass, as this is more sustainable and environmentally benign than ethanol derived from corn. Typical 'energy crops' which will be used to generate second generation biofuels are renewable, sustainable feedstocks, the use of which will not impact the provision of food to humans or of feed for livestock. Examples are grasses such as sorghum and switchgrass, and fast-growing trees such as poplar and willow (Lal, 2008; McCann and Carpita, 2008).

Currently, raw biomass from energy crops is pre-treated with extreme conditions of acidity, alkalinity, temperature and/or pressure (Kim et al., 2009; Yang and Wyman, 2009; Zhu et al., 2010a). It is hoped that these chemical and physical methods of pre-treatment will be supplemented and eventually supplanted by enzymatic saccharification. Glycoside hydrolases exist which break down specific elements of the plant cell wall, expressed by microorganisms which take nutrition from the wall. A diverse range of yeast, bacteria and fungi possess the enzymes required to break down the highly complex plant cell wall and ferment the resulting monosaccharides into ethanol. Harnessing these enzymes for industrial processes is a focus of much

research, as the high cost and inefficiency of biomass saccharification is currently a major bottle-neck preventing the large-scale commercialisation of lignocellulosic bioethanol fuel (Hill et al., 2006). One cost-effective option being sought is to consolidate the saccharification and fermentation of biomass to ethanol in a 'one-pot' conversion strategy (Elkins et al., 2010). This would be a highly cost-effective and efficient process, but no single organism has so far been discovered with a complete repertoire of necessary enzymes. An artificial ecosystem, resembling that found in the human intestine, may be capable of very extensive degradation of plant biomass, if supplemented with fermentative species. Alternatively, microorganisms could be re-engineered with the capacity for general hydrolysis of biomass with simultaneous production of combustible fuel (Elkins et al., 2010).

1.7 Objectives of this study

1.7.1 Understanding the expansion of glycoside hydrolase family 43

As discussed in this chapter, in the human gut symbiont *B. thetaiotaomicron*, versatility of polysaccharide use is ensured by the apparent redundancy of a suite of 256 glycoside hydrolases. Some of these enzymes are unique to this species, while certain classes and families of enzymes are expanded more significantly than can be adequately explained by the suite of glycans available in the gut. For example, the *B. thetaiotaomicron* genome encodes thirty-two GH2 enzymes, which are often β-galactosidases, compared to just one or two expressed by strains of *Bifidobacterium longum*, which metabolises lactose (Xu and Gordon, 2003b). Similarly, *B. thetaiotaomicron* shows expansion in glycoside hydrolase family 43. This family is also expanded in many other bacterial species, including the soil saprophyte *Cellvibrio japonicus*. As family GH43 is polyspecific, it is hypothesised that the expansion in this family in *B. thetaiotaomicron* is matched by a variety of substrate specificities, allowing the bacterium to achieve more complete degradation of the plant cell wall. Many of these enzymes are annotated with activities in the database, and include putative arabinanases, arabinofuranosidases and xylosidases. Proper biochemical characterisation of the enzymes is required to gain a more accurate picture of the metabolic capability of the species. To this end, a project was undertaken to characterise all GH43 enzymes encoded by the genome of *B.*

thetaiotaomicron. The primary aim of this project was to characterise the biochemical activity of this repertoire of enzymes to improve our understanding of expansion in family GH43. Specifically, cloning experiments were performed with the aim of obtaining soluble protein for the enzymes, which were subjected to activity assays to determine the substrate for each enzyme. Sequences of the enzymes were also submitted to phylogenetic analysis. These experiments are discussed in Chapter Three of this thesis and in the light of these new data, a potential evolutionary history for family GH43 is considered.

1.7.2 Structure – function analysis of an industrially important GH43

Humicola insolens is a fungal organism of the *Ascomycetes* phylum. It is a source of industrially significant cellulases and possesses a single identified GH43 enzyme, which has been described by scientists at Novozymes in Denmark. The enzyme has a highly specific activity on arabinoxylan: it selectively cleaves the α1,3 linked arabinofuranose from backbone xylopyranose residues which are doubly substituted at O2 and O3 (Sorensen et al., 2006). As presented in this thesis, a project was undertaken which aimed to understand the nature of the tight specificity of this unusual enzyme, and the molecular basis for specificity against a pseudosymmetrical substrate. Chapter Four presents crystal structures of this enzyme in complex with substrates which illustrate the mode of specificity. Kinetic analysis and mutagenesis studies have identified those amino acids which are important in coordinating this specific activity against arabinoxylan and arabinan. In addition, it is demonstrated that a minor structural modification of the active site of the enzyme introduces the capacity for degradation of the xylan backbone while retaining wildtype arabinofuranosidase activity.

CHAPTER TWO

Materials and methods

2.1 Bioinformatics and Sequence Analysis

2.1.1 Computation of protein properties

Glycoside hydrolases are classified into families on the CAZy database (http://www.cazy.org/; Cantarel et al., 2009; Henrissat, 1991) and GenBank accession numbers from the CAZy website were used to obtain protein and DNA sequences from NCBI (http://www.ncbi.nlm.nih.gov/pubmed/). Molar extinction coefficients and precise molecular weights for proteins were calculated from the primary structure using the ProtParam calculator on the Expasy molecular biology server (http://www.expasy.ch/tools/protparam.html; Gasteiger et al., 2005).

Protein sequences were input into the servers at SignalP and LipoP at http://www.cbs.dtu.dk/services/SignalP/ and http://www.cbs.dtu.dk/services/LipoP/ (Emanuelsson et al., 2007). SignalP predicts the presence and location of signal peptide cleavage sites in amino acid sequences, and LipoP produces predictions for lipoproteins. Cleavage sites given by these online algorithms were used to determine start sites for cloning recombinant proteins geared for soluble expression in the cytoplasm of *Escherichia coli* cells. PSortB was also utilised (http://www.psort.org/psortb/; Zhao et al., 2010). PSortB is a subcellular localisation prediction tool which analyses protein motifs and signal peptide sequences to give probability scores for localisation to five cellular compartments (cytoplasmic, cytoplasmic membrane, periplasm, outer membrane and extracellular).

2.1.2 Alignment and phylogenetic analysis of protein sequences

Homologues of protein sequences were identified by alignment against the NCBI protein database using the blastp server (http://www.ncbi.nlm.nih.gov/blast/Blast.cgi?PAGE=Proteins&PROGRAM=blastp).
Gene sequences were analysed by alignment against the NCBI genome database at http://www.ncbi.nlm.nih.gov/genomes/MICROBES/microbial_taxtree.html.

Alignments of protein sequences was undertaken using ClustalW (http://www.ebi.ac.uk/Tools/clustalw2/index.html; Thompson et al., 1994). Alignment outputs from ClustalW were used with the PhyML software (http://atgc.lirmm.fr/phyml/; Guindon and Gascuel, 2003) to generate phylogenetic trees which were viewed and edited using Mega4 (http://www.megasoftware.net/; Kumar et al., 2008; Tamura et al., 2007).

2.2 Molecular Biology

All solutions were prepared using water as the solvent unless otherwise specified. High quality H_2O was produced by a Millipore Milli-RO 10 Plus Water Purification System.

2.2.1 Bacterial strains and plasmids

All *Escherichia coli* strains utilised in this study are described in Table 2.1. Plasmids used in cloning experiments are outlined in Table 2.2.

Cell strain	Description	Purpose	Reference
One Shot™ TOP10	F´*mcrA* Δ*(mrrCB-hsdRMS-mrr) Φ*80*lacZ*Δ M15 Δ*lacX74 deoR recA1 araD139* Δ(*ara-eu*)7697 *galU galK rspL endA1 nupG*	Plasmid DNA ligation	Invitrogen
XL1-blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F´*proAB lacI*^{q}*Z*ΔM15 Tn10(Tet^{r})]	Plasmid DNA replication	(Bullock et al., 1987)
BL21 (DE3)	F^{-} *dcm ompT hsdS*(r_B- m_B-) gal l(DE3)	Protein expression	(Studier and Moffatt, 1986)
Tuner™ (DE3)	F^{-} *ompT hsdS*B (r_B^{-} m_B^{-}) *gal dcm lacY1* (DE3)	Protein expression	Novagen
B834	F^{-} *ompT hsdS$_B$* (r_B^{-} m_B^{-}) *gal dcm met* (DE3)	Selenomethionine protein expression	Novagen

Table 2.1 Cell strains utilised in this study.

Plasmid	Size (bp)	Antibiotic resistance phenotype	Genotype	Source
mini-pRSETA	2895	Amp^R	T7, *lac*, *lacI*^{q}	Modified version of pRSETA (Invitrogen)
mini-pRGST	3522	Amp^R	T7, *lac*, *lacI*^{q}	mini-pRSETA modified to include GST from pGEX (GE Healthcare)
pET21(a)	5443	Amp^R	T7, *lac*, *lacI*^{q}	Novagen
pGEX	4968	Amp^R	T7, *lac*, *lacI*^{q}	GE Healthcare

Table 2.2 Predominant plasmid vectors utilised in this study.

GST = glutathione-S-transferase. Multiple cloning regions of plasmids utilised in this study are given in Appendix A.

2.2.2 Media and growth conditions for bacteria

Unless otherwise stated in protein expression protocols, *Escherichia coli* strains were cultured at 37 °C in liquid Luria-Bertani (LB) broth medium (1 % (w/v) Bacto© Tryptone, 1 % (w/v) NaCl and 0.5 % (w/v) yeast extract, pH 7.2). Conical flasks with a volume twice that of the culture were used to provide sufficient aeration during rotary shaking at 180 rpm. Media were autoclaved at 120 °C for 30 minutes prior to addition of appropriate antibiotic and subsequent inoculation with bacteria.

A media kit from Molecular Dimensions™ was used (SelenoMethionine Medium Complete, comprising base medium, nutrient mix and selenomethionine solution) to express selenomethionine derivatives of proteins for crystallographic experiments.

For the preparation of solid media plates, bacteriological agar was added to media at 2 % (w/v) prior to autoclaving. Appropriate antibiotics were added before pouring plates. The plates were dried at room temperature for approximately 30 minutes before being stored at 4 °C for a maximum of four weeks.

2.2.3 Selective media

Plasmids used in this study coded for resistance to the antibiotic ampicillin (Table 2.2). A stock solution of ampicillin was prepared at 50 mg ml^{-1} in water and appropriate volumes were added to autoclaved media that had been cooled to below 55 °C to give a final working concentration of 50 µg ml^{-1}.

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*. A stock solution of 1 M IPTG was prepared in water and added to liquid media to a final concentration of 0.2 mM, unless otherwise stated.

2.2.4 Storage of DNA and bacteria

Glycerol stocks of each *E. coli* strain, containing 25 % glycerol (v/v), were stored at -80 °C and re-streaked approximately every thirty days onto solid agar media containing appropriate antibiotics where required. Bacterial colonies on agar plates were stored at 4 °C and discarded after no more than two weeks. Plasmids were stored at -20 °C in EB Buffer (10 mM Tris/HCl Buffer, pH 8.5).

2.2.5 Sterilisation

All media, buffers, and certain polysaccharide substrates were sterilised by autoclaving using an Amsco Renaissance Remanufactured 3021 Gravity Steam Steriliser (Steris).

2.2.6 Centrifugation

Bacterial cells from 5-15 ml cultures were harvested at 4 °C in 50 ml universal containers by centrifugation at 4400-5000 rpm for 6-10 minutes in an Eppendorf 5702 bench centrifuge with swing-out rotor. Cells from larger cultures, of 100-1000 ml, were harvested at 4 °C by centrifugation at 5000 rpm for 10 minutes in 500 ml centrifuge pots (Nalgene) using a Beckman Avanti centrifuge with a JA-10 rotor. Small samples were harvested in 1.5 ml Eppendorf tubes by centrifugation at up to 15,000 rpm using an Eppendorf 5424 bench top centrifuge.

2.2.7 Plating bacteria

A glass spreader was sterilised by immersion in 100 % (v/v) ethanol followed by passage through a Bunsen burner flame to burn off the ethanol. The sterile spreader was used to spread bacterial suspensions of 100 µl onto the surface of agar plates, which were incubated inverted overnight (16 hours) at 37 °C in a stationary incubator.

2.2.8 Production of chemically competent *Escherichia coli*

In a variation of the procedure described by Cohen et al., 1972, *E. coli* cells were made competent for the uptake of plasmid DNA using calcium and magnesium chlorides. A 1 ml aliquot of a 10 ml overnight culture of cells was inoculated into 100 ml of sterile non-selective LB media in a 1 litre conical flask. The culture was incubated with shaking (180 rpm) at 37 °C until log phase growth was reached, as evidenced by an absorbance at 600 nm of 0.4. A shaking incubator (Innova® Shaker Series I26, New Brunswick Scientific) was utilised for growing cultures. The cells were rested on ice for 10 minutes then harvested at 4 °C by centrifugation at 5000 rpm for 5 minutes. The medium was decanted and cells were gently resuspended by inversion in 8 ml of ice cold 100 mM $MgCl_2$. After harvesting again in the same way the cells were similarly resuspended in 4 ml of ice cold 100 mM $CaCl_2$. After two hours on ice the cells were competent for transformation with plasmid DNA. Sterile glycerol was added to 25 % (v/v) and cells were stored in 100 µl aliquots at -80 °C. Each fresh batch of competent cells was immediately tested for competency by transformation with plasmid DNA.

2.2.9 Transformation of competent *E. coli*

A 100 µl aliquot of the desired strain of competent cell was gently thawed on ice after which 2-5 µl of plasmid DNA were added. In cloning experiments, where cells were transformed with newly ligated DNA, cells were then incubated on ice for a further 30 minutes. The suspension was then heat shocked by incubation in a water bath at 42 °C for two minutes and immediately returned to ice for a further five minutes. For cloning experiments, 500 µl of non-selective LB were added and the cells were incubated with shaking at 37 °C for one hour. Transformations were then plated

directly onto LB agar containing 50 µg ml^{-1} ampicillin before inversion and incubation at 37 °C for 16 hours.

2.2.10 Small scale plasmid preparation

For plasmid preparations, DNA was transformed into *E. coli* XL1-blue or TOP10 cells (Table 2.1). A streak of colonies from the plate was inoculated into 5 ml selective LB and incubated with shaking at 37 °C overnight. Overnight 5 ml cultures were harvested by centrifugation at 5000 rpm for 8 minutes, and the supernatant decanted. Plasmid extraction and purification was performed using a Qiagen® QIAspin Prep kit as described in the manufacturer's instructions.

2.2.11 Restriction digestion of DNA

Restriction enzymes used in this study were purchased from Fermentas and New England BioLabs. Digestion of DNA by these enzymes was performed in appropriate reaction buffers following the manufacturer's instructions. The required amount of DNA (plasmid or PCR product) in EB buffer was mixed with the appropriate volume of reaction buffer. Endonuclease enzyme was added at 5-10 units per µg DNA and the digest incubated in a 37 °C water bath for 90-120 minutes, or just 20 minutes for rapid digest enzymes. One unit of enzyme is defined as the amount of enzyme required to cleave 1 µg of DNA in one hour at 37 °C.

2.2.12 Agarose gel electrophoresis of DNA

The electrophoresis of DNA molecules was carried out according to the method of Meyers et al., 1975 using submerged horizontal gels, and was used to determine the size (base pairs) of linear DNA molecules. Linear double stranded DNA molecules migrate through gel matrices at a rate that is inversely proportional to the Log_{10} of the size of the nucleic acid. Therefore, the sizes of DNA fragments were determined by comparing electrophoretic mobility with that of standards of known sizes.

Gels were prepared with 0.8 % agarose in 1 X TBE buffer (10.8 g L^{-1} Tris base, 5.5 g L^{-1} Boric acid, 40 ml L^{-1} 0.5 M EDTA pH 8.0) and boiled using a microwave oven until fully dissolved. After cooling to approximately 50 °C, 0.5 µg ml^{-1} ethidium bromide was added to allow visualisation of DNA and the gel was cast by pouring into a

sealed gel tray (BioRad). Once set, the gel was overlaid with 50 ml 1 X TBE buffer. Samples were prepared for electrophoresis by addition of loading buffer (0.25 % w/v Xylene cyanol FF, 50 % v/v glycerol in 10 X TBE) and transfer into the wells of the gel. DNA was electrophoresed at 140 V, ~60 mA for approximately 45 minutes. Standard DNA samples (GeneRuler™ 1 kb ladder, Fermentas) totalling 4 µl were also loaded and electrophoresed to determine the size of DNA fragments in the samples. In addition, band size and intensity of the standard samples was used to estimate concentration of DNA samples.

DNA concentration was also determined by spectrophotometry. The absorption of appropriately diluted DNA, between 200 nM and 400 nM, was measured in quartz cuvettes using a Varian CARY 100 Bio spectrophotometer (Agilent Technologies, USA). Double-stranded DNA at 50 µg ml^{-1} gives an A_{260} reading of 1.0. For single-stranded DNA and oligonucleotides, this absorbance reading indicates a concentration of 33 µg ml^{-1}.

2.2.13 Visualisation of agarose gels

Following electrophoresis, the agarose gel was removed from the apparatus and visualised using an AlphaImager system (Cell Biosciences). Photographs were printed using thermal paper.

2.2.14 Purification of DNA fragments

PCR products were purified using QIAquick PCR Purification Kit (Qiagen) according to the manufacturer's instructions. Restriction digested vector and insert DNA fragments were purified in the same way or alternatively were isolated from agarose gels by excision of the required band from the gel on a UV illuminating light box with a clean scalpel blade. Purification of DNA from agarose was performed using the Qiagen Gel Extraction Kit, a modification of the previously described PCR purification protocol.

2.2.15 Ligation of insert and vector DNA

Following restriction digestion with appropriate endonucleases, insert and expression vector DNA molecules with compatible cohesive ('sticky') ends were mixed at an approximate molar ratio of 3:1. A typical ligation was performed in a 200 µl Eppendorf tube as outlined in Table 2.3. Each reaction was mixed gently and incubated at room temperature for 30 minutes.

100 ng (1 µl)	Vector DNA
300 ng (2-3 µl)	Insert DNA
4 µl	5 X Ligation buffer (400 mM Tris-HCl, 100 mM $MgCl_2$, 100 mM DTT, 5 mM ATP)
1 µl	T4 DNA ligase (Fermentas) (5 U μl^{-1})
Up to 11 µl	Sterile water
20 µl	Total volume

Table 2.3 Reaction mix for insert-vector ligation.

2.2.16 Polymerase Chain Reaction

The polymerase chain reaction (PCR) (Mullis and Faloona, 1987) was used to amplify specific DNA fragments from a bacterial genome. Primers were produced using automated MWG Oligo-2000 synthesiser technology and purified by HPSF® (MWG-Biotech AG, Germany) or by the in-house DNA synthesis machines at Integrated DNA Technologies (IDT,USA) and purified by Rapid HPLC™, followed by desalting to remove small molecule impurities. Primers were analysed by MALDI-TOF by the respective manufacturer prior to dispatch.

The standard PCR protocol requires two oligonucleotide primers, one complementary to each strand of the DNA molecule at sites that flank the region to be amplified. A thermostable DNA polymerase catalyses the synthesis of the complementary DNA strand using dNTPs provided in the reaction mix. Primers were designed such that complementary sequences were 15-18 base pairs in length and had a G/C content of approximately 40 %. Ideally primers ended (3') in a relatively

G/C rich-region to improve annealing. Restriction sites were added to the 5'-ends of primers, as well as any additional bases required to keep the complementary sequence in frame. In addition, the non-palindromic sequence GTCGCC was included at the extreme 5'-end to ensure restriction enzymes cleaved their target sequences in the PCR products. Annealing temperatures (i.e. melting points, T_M) were calculated using the OligoCalc server using Equation 2.1 (http://www.basic.northwestern.edu/biotools/oligocalc.html). PCR reactions were prepared in 200 µl Eppendorf tubes as described in Tables 2.4 and 2.5.

Equation 2.1 $T_M = 64.9 + 41*(yG+zC-16.4)/(wA+xT+yG+zC)$

where w, x, y and z are the numbers of the bases A, T, G and C respectively.

2.5 µl	KOD buffer (10 X)
2.5 µl	dNTP mix (2.5 mM each)
1 µl	Template genomic DNA (50-100 ng μl^{-1})
0.5 µl	KOD DNA polymerase (10 U μl^{-1})
2.5 µl	Forward primer (~100 pmol μl^{-1})
2.5 µl	Reverse primer (~100 pmol μl^{-1})
13.5 µl	PCR-grade H_2O

Table 2.4 Reaction mix for standard PCR reactions.

A master mix (typically 20 X or 10 X) of components was generally prepared and dispensed between aliquots. KOD reaction buffer (10 X) comprises 1.2 M Tris-HCl, 15 mM $MgSO_4$, 100 mM KCl, 60 mM $(NH_4)_2SO_4$, 1 % Triton-100 and 1 mg ml^{-1} BSA.

5 µl	iProof HF buffer (5 X)
0.5 µl	dNTP mix (2.5 mM each)
0.5 µl	Template genomic DNA (~250 ng)
0.2 µl	iProof DNA polymerase (0.2 U μl^{-1})
4 µl	Forward primer (2.5 µM)
4 µl	Reverse primer (2.5 µM)
10.8 µl	PCR-grade H_2O

Table 2.5 Reaction mix for PCR reactions utilising the iProof system (BioRad).

The iProof system requires significantly shorter reaction times in the thermo cycle. iProof high fidelity (HF) buffer includes 7.5 mM $MgCl_2$ in the 5 X stock.

Two PCR machines were used in this study. Broadly speaking, for cloning experiments a PX2 (Hybaid) was used and for mutagenesis (Section 2.2.17, below) an Eppendorf Mastercycler was utilised. The standard thermo cycle for amplification of a DNA fragment is given in Tables 2.6 and 2.7. The precise duration and temperature of specific steps in the thermo cycle were varied to improve yield of PCR product. After thermo cycling was complete a 4 µl aliquot of each reaction was analysed by electrophoresis (Section 2.2.12).

1 cycle	95 °C 2 minutes
40 cycles	94 °C 1 minute Lowest T_M of primer pair – 5° 1 minute 68 °C 2 minutes
1 cycle	68 °C 5 minutes
Hold	4 °C

Table 2.6 Thermo cycling reaction protocol for a typical PCR for cloning.

1 cycle	98 °C 30 seconds
40 cycles	98 °C 30 seconds Lowest T_M of primer pair + 3° 15 seconds 72 °C 30 seconds
1 cycle	72° 1 minute
Hold	4 °C

Table 2.7 Thermo cycling protocol for an iProof reaction, with considerably shorter reaction times.

2.2.17 QuikChange site directed mutagenesis

Mutagenesis of single amino acids was carried out using a QuickChange™ Site-Directed Mutagenesis kit from Stratagene, following the manufacturer's instructions. The method uses double-stranded recombinant plasmid DNA as template and two oligonucleotide primers (Integrated DNA Technologies, USA) containing the desired codon mutation flanked on either side by 15-18 nucleotides. The primers are fully complementary to each other. The polymerase enzyme used in site-directed mutagenesis is a high fidelity archaeal enzyme called *Pfu*Pol. Reactions were prepared in 200 µl Eppendorf tubes as outlined in Table 2.8. Table 2.9 gives the mutagenic PCR thermo cycle.

5 µl	Reaction buffer (10 X)
5 µl	Plasmid DNA template (50 – 500 ng)
5 µl	Forward primer (~1.2 µg)
5 µl	Reverse primer (~1.2 µg)
1 µl	dNTP mix (2.5 mM each)
1 µl	Pfu Ultra HF DNA polymerase (2.5 U μl^{-1})
28 µl	PCR-grade H_2O

Table 2.8 Reaction mix for QuikChange mutagenesis PCR.

Reaction buffer comprises 200 mM Tris-HCl, 100 mM KCl, 100 mM $(NH_4)_2SO_4$, 20 mM $MgSO_4$, 1 % Triton-100 and 1 mg m l^{-1} BSA.

1cycle	95 °C 30 seconds
18 cycles	95 °C 30 seconds 55 °C 1 minute 68 °C 6 minutes
1 cycle	68 °C 10 minutes
HOLD	4 °C

Table 2.9 Thermo cycle protocol for QuikChange mutagenesis PCR

After thermo cycling, 1 µl of the endonuclease Dpn1 (New England BioLabs) was added to each reaction. After mixing by vortex, the reactions were incubated at 37 °C in a water bath for one hour to digest parental (non-mutated) DNA which is methylated. Dpn1 treatment leaves amplified DNA (carrying the mutation) undigested as it is not methylated. The DNA was then transformed into competent TOP10 cells using the protocol for freshly ligated plasmids outlined in Section 2.2.9.

2.2.18 Automated DNA sequencing

DNA sequencing employed the Value Read service from MWG Biotech (Germany) or the Custom Sequencing Service from Macrogen (USA). Approximately 10 µg of plasmid DNA in a 1.5 ml Eppendorf tube were shipped directly to the respective company for sequencing with primers listed in Table 2.10.

Plasmid	Forward primer	Reverse primer
pET plasmids	T7	T7 term
	TAATACGACTCACTATAGGG	CTAGTTATTGCTCAGCGGT
pGEX	pGEX 5'	pGEX 3'
	GGGCTGGCAAGCCACGTTTGGTG	CCGGGAGCTGCATGTGTCAGAGG

Table 2.10 Primers used in automated sequencing reactions.

2.2.19 Gene expression of recombinant protein in *E. coli*

The following protocol was used for over-expression of all proteins unless otherwise stated. A streak of *E. coli* colonies transformed with the appropriate plasmid was inoculated into 10-15 ml selective LB media and grown with rotary shaking (180 rpm) at 37 °C for 16 hours. This culture was inoculated directly into one litre of selective LB in a two litre conical flask which was incubated with aeration (180 rpm) at 37 °C until absorbance at 600 nm reached 0.6 (A_{600} = 0.6), indicating log phase growth. The incubator was then cooled to 16 °C and after one hour IPTG was added under sterile conditions to a final concentration of 0.2 mM; cultures were maintained at this temperature for a further 16 hours. Cells were harvested by centrifugation at 5000 rpm for 10 minutes at 4 °C. After the supernatant was decanted the cell pellet was re-suspended in 30 ml of Talon buffer (20 mM Tris/HCl buffer pH 8.0 containing 300 mM NaCl) per litre of original culture.

The methionine auxotrophic strain of *E. coli* B834 (Table 2.1) was used to express a selenomethionine-containing form of proteins intended for crystallographic analysis. Plasmid DNA encoding the protein was transformed into B834 cells (section 2.2.9); a streak of these colonies was inoculated into 10 ml selective LB and grown with

aeration (180 rpm) at 37 °C for 16 hours. The culture was then inoculated into 100 ml selective LB in a 500 ml conical flask and incubated with aeration (180 rpm) at 37 °C until A_{600nm} = 0.2. Cells were harvested by centrifugation at 4400 rpm for five minutes and re-suspended in 5 ml of SelenoMet Medium Base™ (Molecular Dimensions). This step was repeated three times to ensure the removal of all LB media. The re-suspended pellet was inoculated into one litre of SelenoMet Medium Base™ in a two litre conical flask and 50 ml of filtered (0.4 µm) SelenoMet™ Nutrient Mix was added to the culture along with 4 ml of a SelenoMethionine Solution (10 mg ml^{-1}).

2.2.20 Fractionation of *E. coli* cells

Cell suspensions (Section 2.2.19) were lysed by sonication for 2 minutes on ice using a sonic dismembrator set at low intensity (~ 40 watts) and 0.5 second cycling. Lysed cell suspensions were transferred to a 50 ml centrifuge tube (Nalgene) and cell debris pelleted by centrifugation at 17,000 rpm for 30 minutes at 4 °C. The supernatant (cell-free extract, CFE) was retained for subsequent purification (Section 2.2.21). In expression trials the pellet containing cell debris was re-suspended in 10 ml Talon buffer by pipetting.

2.2.21 Purification of recombinant proteins

2.2.21.i Immobilised metal ion affinity chromatography

Proteins derived from recombinant forms of the vectors pRSETA and pET21(a) contain a stretch of six Histidine residues at the *N*- or *C*-terminus (His_6), respectively. Immobilised metal ion affinity chromatography (IMAC) was used to purify recombinant proteins by selecting for the electron-rich residues in the His_6 tag. These residues interact with the electropositive transition metals cobalt and nickel, which were bound to a solid chromatographic matrix. This interaction can be dissociated with imidazole, which mimics and out-competes the Histidine side chain in the His_6 tag for binding to the column. The metal affinity matrix used was Talon™ Fast Flow (Clontech Laboratories Inc.) in which the His_6 tags interact with cobalt, or a His select HF Nickel Affinity gel matrix (Sigma Aldrich, USA) in which the His_6 tags interact with nickel. Subsequent to every use, the nickel resin was regenerated by successive washing with ~ 3 column volumes of 6 M guanidine hydrochloride, water,

0.5 M EDTA, water, 10 mg ml-1 NiSO4 and water, followed by storage in 20 % ethanol.

2.2.21.ii Ion exchange chromatography (IEC)

Proteins without a His_6 tag, or with an inaccessible tag, were purified from cell free extract by ion exchange chromatography (IEC) using an ÄKTA Purifier Core System with a flow of 1 ml min^{-1}, connected to an AEC HiPrep™ IEX FF anion exchange column (GE Healthcare). A Frac-920 fraction collector (GE Healthcare) was used to collect 1 ml fractions during the run.

As required, an ÄKTA Purifier system with a flow of 3 ml min^{-1} connected to a Q12 anion exchange column (BioRad) was used to carry out a secondary purification step after IMAC purification or IEC from cell free extract.

Prior to loading onto the column, protein fractions were dialysed into 10 mM Tris/HCl pH 8.0; protein was then loaded onto the equilibrated column which was washed with 4 column volumes of 10 mM Tris/HCl buffer pH 8.0 (Buffer A) to remove unbound protein, before elution with a linear 0-150 mM NaCl gradient in Buffer A. Protein elution was detected by UV absorbance and 1 ml fractions collected above a threshold of 50 mAU at 280 nm using a Biofrac fraction collector (BioRad). Each fraction was analysed by SDS-PAGE (Section 2.2.22) and those containing significant pure protein were dialysed against four litres of an appropriate buffer.

2.2.21.iii Gel filtration fast performance liquid chromatography (FPLC)

Using the ÄKTA Purifier system described above, connected to a Superdex 200 gel filtration column, a final round of purification was carried out to improve homogeneity of protein samples when required. The protein was dialysed into Buffer A and concentrated to <3 ml, then loaded onto the column using a 2 ml static loop. The protein was eluted from the column using Buffer A containing 200 mM NaCl (Buffer B). Fractions (1 ml) were collected and analysed by SDS-PAGE (Section 2.2.22); subsequently pure fractions were pooled and concentrated for further analysis.

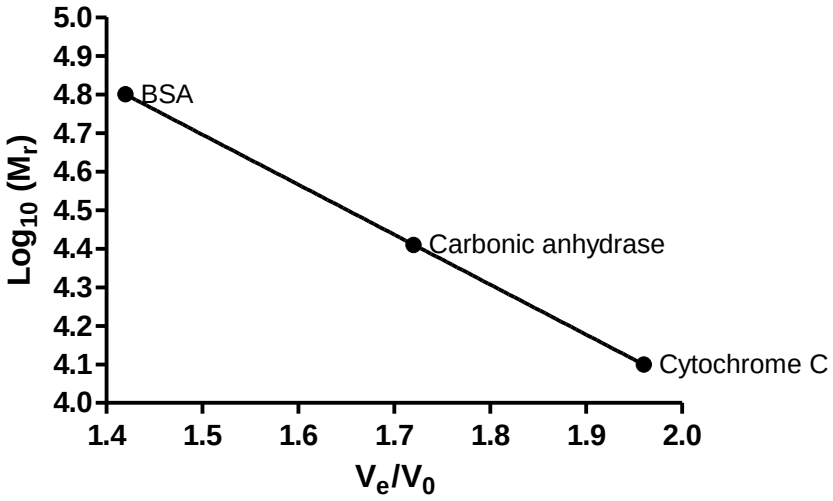


Figure 2.1 Calibration of the gel filtration column.

Calibration was performed using three proteins of known molecular weight (M_r): Bovine serum albumin (BSA, 66 kDa), Carbonic anhydrase (29 kDa) and Cytochrome C (12.4 kDa). V_e = elution volume of the protein. V_0 = volume of the column.

2.2.22 Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE)

Analysis of over-expressed proteins was carried out using SDS-PAGE according to the method of Laemmli, 1970. Gels were used to analyse the size, relative purity and quantity of protein. Gels of 12.5 % (v/v) polyacrylamide were typically used in conjunction with the XCell SureLock® Mini-Cell CE apparatus produced by Invitrogen, USA. Plastic disposable gel cassettes, 1.0 mm thickness (Invitrogen, USA) were used in producing the gels according to the reaction mixtures given in Tables 2.11 and 2.12. The resolving gel was poured into the plates then covered with water to maintain a level surface and allowed to polymerise at room temperature. After the water was removed the stacking gel was poured on top of the resolving gel. A 12-well comb was embedded into the stacking layer which was then allowed to polymerise at room temperature.

1.75 ml	dH_2O
4.7 ml	0.75 M Tris/HCl pH 8.8 with 0.2 % SDS
2.9 ml	40 % polyacrylamide
45 µl	10 % APS
15 µl	TEMED

Table 2.11 Reaction mix for resolving layer of SDS-PAGE gels.

Typically, gels were made in bulk and this reaction mix was expanded to make 9 or 18 gels together. APS = ammonium persulphate. TEMED = Tetramethylethylenediamine.

1.5 ml	dH_2O
1.9 ml	0.25 M Tris/HCl pH 6.8 with 0.2 % SDS
0.38 ml	40 % polyacrylamide
30 µl	10 % APS
10 µl	TEMED

Table 2.12 Reaction mix for stacking layer of SDS-PAGE gels.

Typically, gels were made in bulk and this reaction mix was expanded to make 9 or 18 gels together.

After they were set, gels were kept moist at 4 °C for long term storage. In preparation for use, the combs were removed and the gels placed in the electrophoresis tank and covered with running buffer (Table 2.13). Protein samples were prepared for loading by mixing 40 µl of sample and 20 µl of loading buffer (Table 2.14) and boiling for three minutes.

30.3 g	Tris base
144 g	Glycine
10 g	SDS
to 1 litre	H_2O

Table 2.13 Reaction mix for 10 X stock of SDS running buffer (Laemmli buffer).

10 % (w/v)	SDS
5 ml	0.25 M Tris/HCl pH 8.8 with 0.2 % SDS
25 % (w/v)	Glycerol
2.5 ml	β-mercaptoethanol
0.1 %	Bromophenol blue dye

Table 2.14 Reaction mix for running buffer used with samples for SDS-PAGE.

After electrophoresis, the gel was stained by soaking in Coomassie Blue Stain (0.4 % (w/v) Coomassie Brilliant Blue, 10 % (v/v) glacial acetic acid, 40 % (v/v) methanol) for 1-2 hours at room temperature with gentle shaking. The gel was destained by soaking in a solution comprising 40 % (v/v) methanol and 10 % (v/v) glacial acetic acid until the blue background faded sufficiently for all protein bands to be visible. The gel was recorded using a digital camera or using the transilluminating box of an AlphaImager system. The molecular weights (M_r) of proteins separated by SDS-PAGE were estimated by comparing their electrophoretic mobility with protein standards of known M_r (Appendix A).

2.2.23 Quantification of purified protein

Concentration of pure protein was determined according to the methods of Gill and Hippel, 1989 and Pace et al., 1995. Protein was diluted in water and A_{280} and A_{320} determined by scanning from 200 to 400 nm using a Varian CARY 100 Bio spectrophotometer (Agilent Technologies). Final absorbance was calculated by subtracting A_{320} from A_{280} and protein concentration determined according to

Equation 2.2. Molar extinction coefficients were calculated using the ProtParam calculator on the Expasy molecular biology server (Section 2.1.1).

Equation 2.2 $A = \varepsilon C l D$

where $A = A_{280} - A_{320}$

ε = molar extinction coefficient of protein at 280 nm

C = molar concentration of sample

l = length of light path in spectrophotometer

D = dilution factor

2.2.24 Concentrating protein

Protein solutions were concentrated after filtration through a 0.2 µm membrane using 20 ml or 2 ml Vivaspin™ centrifugal concentrators (VivaScience) with cut-off filters of 5, 10 or 30 kDa molecular weight, as appropriate to the protein in question. Centrifugation was performed at 4400 rpm using an Eppendorf 5702 bench centrifuge with swing-out rotor.

2.3 Biochemistry

2.3.1 Preparation of polysaccharide substrates and ligands

2.3.1.i Water-soluble polysaccharides

Unless otherwise stated, all of the polysaccharides used in this study were dissolved at appropriate concentrations in water, as were monosaccharides and oligosaccharides. Polysaccharides were generally dissolved slowly with continuous mixing by magnetic stir bar and moderate heating. Highly concentrated solutions, in the region of 40 mg ml^{-1} for wheat arabinoxylan, and solutions which would require long term storage or were known to be unstable, such as sugar beet arabinan, were autoclaved (Section 2.2.5).

2.3.1.ii Enzymatic digestion of polysaccharides

For use in applications pertaining to NMR and crystallography, enzymatic digestion of wheat arabinoxylan was performed to produce a mixture of arabinoxylooligosaccharides (AXOS). 1 g of Arabinoxylan (Megazyme) was dissolved in 40 ml of 50 mM sodium phosphate buffer pH 7.0 to which was added 2 ml BSA (1mg ml^{-1} final) and 1 ml GH10 xylanase (50 nM final). The reaction was incubated for 1 hour and mixed by frequent inversion. Aliquots of 250 µl were taken every five minutes, boiled for 10 minutes and analysed by TLC and HPLC (diluted 1 in 100) to monitor the course of the reaction and to ascertain when digestion was complete.

2.3.1.iii Partial acid hydrolysis of polysaccharides

To generate a mixture of oligosaccharides from water-soluble polysaccharide for use in enzyme assays and NMR experiments, sugars were subjected to partial acid hydrolysis. To a starting solution at 20 mg ml^{-1} was added hydrochloric acid (HCl) to a final concentration of 50 mM, 0.5 M or 1 M. The mixture was heated to 100 °C and maintained at this temperature for 15 minutes, followed by cooling to room temperature. The mixture was held at room temperature for a further 20 minutes. To neutralise the solution, sodium hydroxide (NaOH) was added to a final molarity matching that of the acid. Samples were then analysed by HPLC (Section 2.3.3).

2.3.1.iv Ethanol precipitation

When short oligosaccharides were generated from polysaccharides by acid- or enzyme-catalysed hydrolysis, remaining large polymers were removed by ethanol precipitation. Ethanol was added to solutions to a final concentration of 65-80 %. The solution was mixed carefully and incubated at 4 °C for 2-16 hours. Centrifugation at 10,000 rpm removed precipitated material. Following vacuum drying to remove ethanol from the soluble fraction, the pellet was dried and concentrated by lyophilisation.

2.3.1.v Concentrating purified sugars by lyophilisation

Oligosaccharide and polysaccharide sugars were frozen to -80 °C and lyophilised in a freeze drier at -60 °C. Freeze dried fractions were resuspended in a small volume of water or, for use in NMR, in deuterium oxide (D_2O).

2.3.1.vi Oligosaccharide size exclusion chromatography

Digested products of wheat arabinoxylan (Section 2.3.1.ii) were separated by size exclusion chromatography using P2 Bio-gel (Bio-Rad) matrix packed in a Glass Econo-Column (2.5 cm X 80 cm, total volume 500 ml, BioRad) by gravity flow at 0.2 ml min^{-1} in degassed water. The column was equilibrated with three column volumes (CVs) of water. The mix of AXOS was concentrated by lyophilisation (Section 2.3.1.v) and resuspended in 2 ml of water, and this was loaded directly onto the column bed. The column was run by gravity flow for a further 2 column volumes, with water as the eluting buffer. The standard dead volume of a P2 column is ~40 % (400 ml), after which 5 ml fractions were collected continuously using a Frac-920 fraction collector (GE Healthcare). A 10 µl aliquot of each fraction was tested for sugar content by TLC (Section 2.3.2) and subsequent rounds of HPLC (Section 2.3.3) identified fractions of interest, which were pooled and concentrated for further analysis.

2.3.2 Thin layer chromatography (TLC)

An aluminium-backed silica TLC plate (Silicagel 20 X 20, Merck VWR) was cut to the desired size and a line drawn 1 cm from the bottom of the plate. Samples were spotted in 2 µl volumes separated by a minimum of 8 mm to prevent cross-contamination. To identify products of poly- and oligosaccharide degradation, size marker standards were run in parallel on TLC plates. Monosaccharide standards were spotted at 20 mM, and polysaccharide standards typically at 2 mg ml^{-1}. For assays of xylanase activity, a ‘ladder’ of xylooligosaccharides was used as a standard, comprising xylose, xylobiose, xylotriose, xylotetraose, xylopentaose and xylohexaose (Sigma) at 20 mM, 10 mM, 7 mM, 4 mM, 3 mM and 2 mM, respectively. Enzyme reactions were spotted multiple times, with drying in between, to give an approximate final concentration of products matching that of the standards. Running

buffer (50 % (v/v) n-butanol, 25 % (v/v) acetic acid) was poured into a glass chromatography tank to a height of less than 1 cm. The tank was covered tightly to allow vapours to equilibrate, which required around two hours incubation. The TLC plate was placed into the tank and samples allowed to migrate until the buffer reached approximately 1 cm from the top of the plate. The plate was removed and dried thoroughly using a hairdryer under a fume hood. After drying, an optional second run allowing the buffer to reach the very top of the plate gave better resolution for some complex mixtures of oligosaccharides. When the plate was completely dry, an orcinol sulphuric acid reagent (3 % (v/v) sulphuric acid, 75 % (v/v) ethanol, 0.1 % (w/v) orcinol monohydrate) was used to stain sugars. The plate was immersed in the developing solution for a few seconds, dried carefully and heated to at least 70 °C for 5-10 minutes, or until sugar spots were revealed.

TLC plates were also used to quickly detect the presence of sugars in a sample, when identification was not important. In this case, 10 µl volumes of sample were spotted onto a plate which was immediately developed as described above.

2.3.3 High pressure liquid chromatography (HPLC)

Products of hydrolysis reactions of polysaccharides and oligosaccharides were analysed using analytical CARBOPAC™ PA-100 or PA-200 anion exchange columns (Dionex) equipped with a CARBOPAC™ PA-10 guard column (Dionex). The fully automated system (LC25 chromatography oven, GP40 gradient pump, ED40 electrochemical detector and AS40 autosampler) had a loop size of 10 µl, used a flow rate of 1.0 ml min^{-1} at room temperature with a pressure ~ 2300 psi. 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.

Elution conditions were 0-15 minutes 18 mM NaOH, 15-65 minutes 66 mM NaOH with 0-250 mM sodium acetate linear gradient. After each run the column was washed with 300 mM sodium hydroxide for ten minutes then equilibrated with 18 mM sodium hydroxide for ten minutes.

Appropriate monosaccharides were included as standards in separate HPLC runs at a concentration of 10-100 µM. Reaction products were identified by co-migration with

these standards. Before loading onto the column, polysaccharide hydrolysis reactions were centrifuged to remove precipitated material and diluted as required to bring sugar concentrations to within the range quoted for the standards. Data were collected and manipulated using Chromeleon™ Chromatography Management System (Dionex) via a Chromeleon™ server (Dionex). Final graphs as presented in this thesis were prepared using Prism v5 (GraphPad).

2.3.4 Enzyme assays

Unless otherwise stated, all enzyme assays were performed at 37 °C in buffer at pH 7.0. Assays for kinetic analysis were always performed in triplicate. One set of Eppendorf pipettes was used throughout each assay to minimise error. Graphs were plotted in GraphPad Prism v5, which was used to calculate slopes and error values. The non-linear regression function in Prism was used to determine parameters (K_M and k_{cat}) of Michaelis-Menten kinetics.

2.3.4.i Aryl glycoside hydrolysis assays

The 4-nitrophenyl (4NP) sugar substrates used in initial enzyme assays were as follows: 4NP β-D-galactopyranoside, 4NP-β-D-glucopyranoside, 4NP-β-D-xylopyranoside, 4NP-α-L-arabinofuranoside and 4NP-α-L-arabinopyranoside (Sigma). The appearance of a yellow colour indicates enzyme catalysed hydrolysis of the substrate. A 20 mM stock was prepared in water and stored at -20 °C until use. In a 1.5 ml Eppendorf tube reactions were performed in 50 mM sodium phosphate buffer, pH 7.0, in the presence of bovine serum albumin (BSA) at 1 mg ml^{-1}. Enzyme concentration for these assays was typically 200 nM. In negative control reactions the volume of enzyme was replaced with additional buffer. Reactions were incubated in a 37 °C water bath for 30 minutes, or several hours if no colour change was apparent.

4NP-α-L-arabinofuranoside was used in kinetic analysis of enzymes concentrations ranging from 0.1 mM – 8 mM. Reactions were prepared in glass cuvettes, pre-warmed to 37 °C in a Varian CARY 100 Bio UV/vis spectrophotometer. A fresh stock of the substrate was prepared at 10-20 mM and stored on ice until use. BSA and buffer stock concentrations were 10 mg ml^{-1} and 500 mM, respectively. The final

volume of reactions was 500 µl. Table 2.15 details reaction components. 4NP-α-L-arabinofuranoside undergoes spontaneous cleavage at 37 °C, so reactions were monitored for 5 minutes before the addition of enzyme at appropriate concentration. The release of 4-nitrophenolate (4NP) from the substrate was measured at an absorbance of 400 nm (Figure 2.2). Under the conditions outlined here, the extinction coefficient (ε) of 4NP at pH 7.0 was 2000. Kinetic parameters K_m and k_{cat} were then determined.

pNP volume	X	10	25	50	100	125	150	200	250
Buffer	50	50	50	50	50	50	50	50	50
BSA	50	50	50	50	50	50	50	50	50
Water	350- χ	340	325	300	250	225	200	150	100
Enzyme	50	50	50	50	50	50	50	50	50
pNP concentration (mM)	-	0.2	0.5	1	2	2.5	3	4	5

Table 2.15 Reaction components for enzyme assay of pNP activity, using a substrate concentration of 10 mM.

Typical enzyme concentrations in 4NP kinetic assays were 50 nM – 50 µM. All quoted volumes are units of µl.

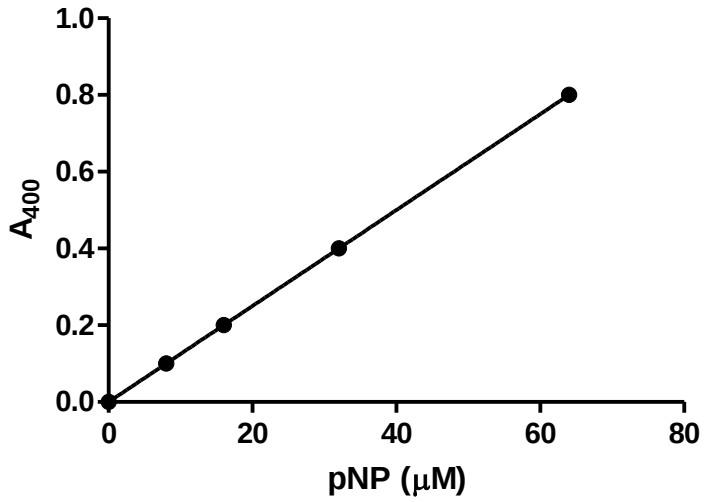


Figure 2.2 4NP standard curve at 37 °C.

2.3.4.ii 3,5-Dinitrosalicylic acid reducing sugar assay (DNSA)

Rates of hydrolysis of polysaccharide substrates were measured by the increase in reducing sugars over time. A reducing sugar in solution has an open chain with an aldehyde or ketone group, allowing the sugar to act as a reducing agent. Monomeric aldose sugars such as glucose can be reduced directly, while ketose sugars such as fructose can act as reducing agents if the functional group is converted to an aldehyde via tautomeric shift. In addition, the free anomeric carbon at the end of a polysaccharide chain can act as a weak reducing agent as it exists in equilibrium between cyclic and open chain forms. When an enzyme cleaves a glycosidic bond a new reducing end is formed. The concentration of reducing ends can be determined using the DNSA assay, described by Miller, 1959. 3,5-dinitrosalicylic acid (2-hydroxy-3,5-dinitrobenzoic acid) reacts with reducing sugars to form 3-amino-5-nitrosalicylic acid.

Time points (500 µl) of an enzyme reaction were taken and added to an equal volume of DNSA reagent (1 % (w/v) DNSA, 0.2 % (v/v) phenol, 1 % (w/v) NaOH, 0.002 % (w/v) glucose and 0.05 % (w/v) $NaSO_3$) to terminate the reaction. Glucose is included in the mixture to raise the overall reducing sugar concentration to within a range detectable by DNSA assay. The tube was then boiled for 20 minutes to initiate the colour change indicative of the presence of reducing sugars. Tubes were then placed on ice for 10 minutes, equilibrated to room temperature and centrifuged briefly to remove precipitated material. Absorbance was read at 575 nm. A standard curve of 0-1000 µg ml^{-1} monosaccharide in the presence of 10 mg ml^{-1} polysaccharide was used to quantify the reducing sugar released during the reaction (Figure 2.3).

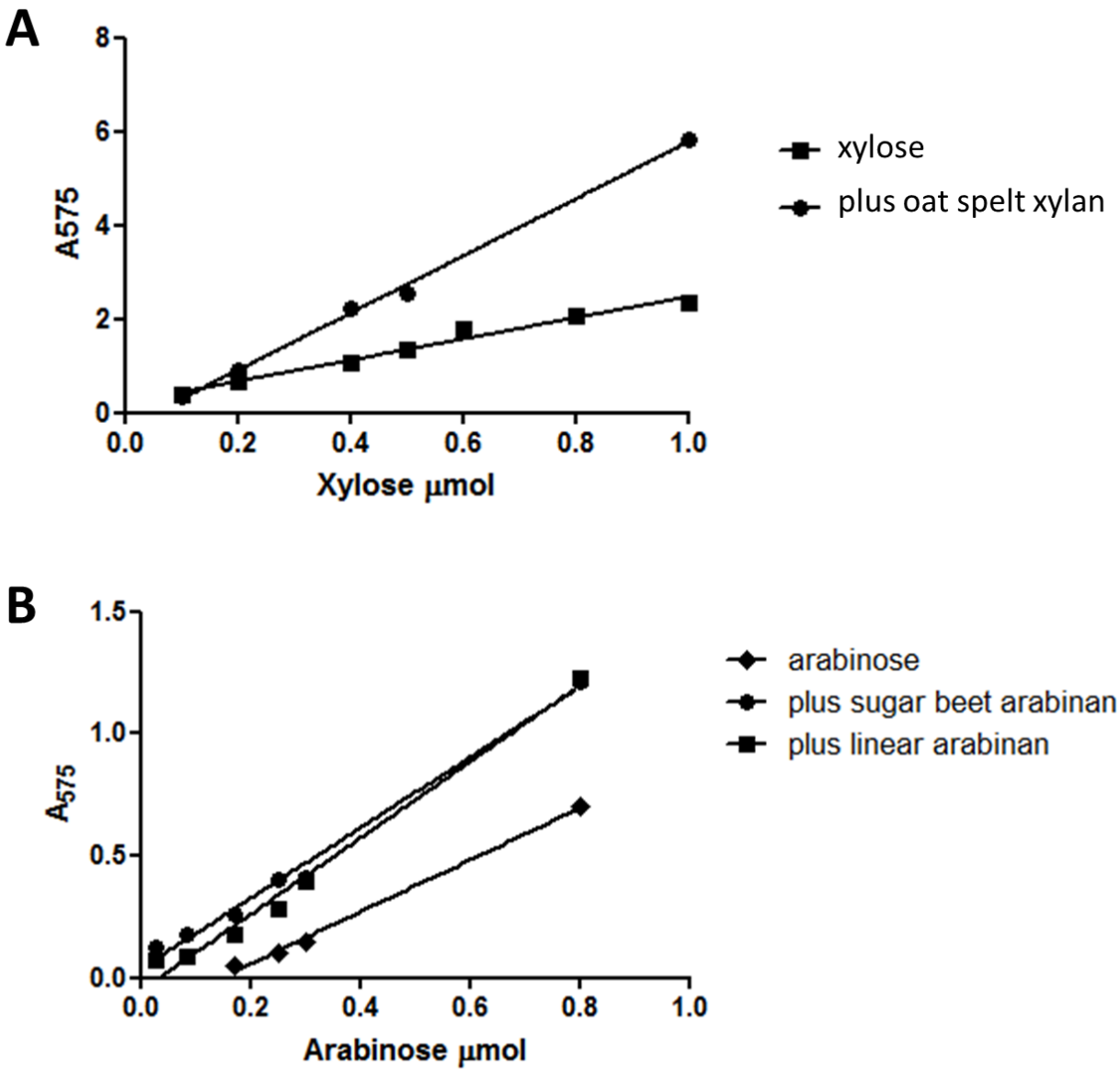


Figure 2.3 Reducing sugar standard curves for arabinose and xylose.

A: standard curve for xylose in the presence and absence of 10 mg ml^{-1} oat spelt xylan.

B: standard curve for arabinose in the presence and absence of 10 mg ml^{-1} sugar beet arabinan and linear arabinan.

2.3.4.iii Galactose/arabinose dehydrogenase assay

A linked assay was used to quantify release of galactose or arabinose from polysaccharide. The enzyme galactose dehydrogenase catalyses the reaction shown

in Equation 2.3, but has relaxed specificity at C5 and thus can accommodate arabinose.

Equation 2.3 D-galactose + NAD^+ ⇌ D-galactono-1,4-lactone + NADH + H^+

The formation of NADH in the presence of NAD by galactose dehydrogenase was measured by absorbance at A_{340}. The molar extinction coefficient (ε) used to determine NADH concentration was 6223. Reactions were prepared in quartz cuvettes pre-warmed to 37 °C as shown in Table 2.16. The reaction was monitored for five minutes prior to the addition of enzyme to ensure a stable baseline and to account for any free arabinose or galactose in the starting solution.

Substrate volume	χ	10	25	50	100	250
Galactose dehydrogenase (500 U ml^{-1})	5	5	5	5	5	5
NAD^+ (1 mM)	50	50	50	50	50	50
BSA (10 mg ml^{-1})	50	50	50	50	50	50
Buffer (500 mM)	50	50	50	50	50	50
Water	295 – χ	285	270	245	195	145
Substrate concentration (mg ml^{-1})	-	0.2	0.5	1	2	5

Table 2.16 Components of the galactose dehydrogenase linked enzyme assay.

Enzyme concentration in galactose dehydrogenase kinetic assays was typically 50 nM – 50 µM. Stock substrate concentration was usually 10 mg ml^{-1}. All quoted volumes are in units of µl.

2.3.5 Nuclear Magnetic Resonance (NMR) Spectroscopy

Prior to analysis by NMR samples were lyophilised and resuspended in 100-300 µl D_2O to minimise the water peak in the spectra. ^{1}H-NMR spectra were recorded with a Varian Inova NMR spectrometer operating at 600 MHz and with a sample temperature of 298 K. Two-dimensional spectra were recorded using standard Varian pulse programs. Chemical shifts were measured relative to internal acetone (□H 2.225, □C 30.89). Data were processed using MestRe-C software (Universidad de Santiago de Compostela, Spain). NMR was performed and interpreted by Maria

Peña of the York lab at the Complex Carbohydrate Research Centre in Athens, Georgia, USA.

2.4 Crystallography

In order to solve the crystal structures presented in this study, a single wavelength anomalous dispersion method at a wavelength optimised for the anomalous f'' signal of selenium was used where molecular replacement was not possible. A derivative form of protein containing selenomethionine was purified (Section 2.2.2, 2.2.19) for the heavy metal experiment.

Prior to refinement, 5 % of observations were set aside for cross-validation analysis and were used to monitor refinement strategies. The programmes iMosflm, Scala and Refmac5 (Murshudov et al., 1997) were obtained as part of the CCP4 software package (Project, 1994). The ShelxC/D/E software was used to locate heavy atoms and calculate phases from these sites (Sheldrick, 2008). Models were completed with manual correction in Coot (Emsley et al., 2010) in conjunction with refinement in Refmac5. Crystal structures were studied using PyMol (Delano Scientific), which was also used to generate figures for this thesis.

2.4.1 Preparation of protein samples

2.4.1.i Native form

Through purification by IMAC, IEC and gel filtration (Section 2.2.21), proteins destined for crystallisation were generated in a highly pure and homogeneous form, as assessed by SDS-PAGE. Proteins were concentrated to at least 25 mg ml^{-1} for initial screening of crystallographic conditions and buffer exchanged into molecular biology grade water (Sigma). Where appropriate, oligosaccharide ligand was added to the required concentration for screening of co-crystallisation conditions. Immediately prior to the establishment of crystallisation screens (Section 2.4.2) proteins and ligands were centrifuged at 13,000 rpm for five minutes to remove precipitated material.

2.4.1.ii Selenomethionine derivative

The protein was over-expressed and the cell pellet re-suspended in 15 ml of Talon buffer containing 2 mM β-mercaptoethanol to maintain the selenium metal in a reduced oxidation state. Cell free extract was prepared (Section 2.2.20) and selenomethionine derived protein purified as for native protein except that all buffers contained 10 mM β-mercaptoethanol as a reducing agent. Protein was concentrated to at least 15 mg ml^{-1} for initial screening and extensively buffer exchanged into molecular biology grade water containing 10 mM DL-Dithiothreitol (DTT, Sigma) as the reducing agent.

2.4.2 Sparse matrix screen of crystallisation conditions

Preliminary crystallisation conditions were determined using commercial screens of 48 or 96 conditions each. The screens PEG/Ion™, Crystal Screen™ 1 and Crystal Screen™ 2 were obtained from Hampton Research. The PACT, Classics and JCSG+ screens were obtained from Qiagen. Conditions for these screens are available online (www.qiagen.com). Initial screens were performed at the highest possible concentration, aiming for precipitation in approximately 50 % of conditions. Hanging drops (Section 2.4.3.ii) were established with a varying ratio (2:1 or 1:1) of protein to crystallography reagent (mother liquor). Where appropriate, ligand was included in drops at 10 mM for co-crystallisation trials. Conditions producing the best crystals in terms of size and morphology were optimised in subsequent screens by manipulating the original commercial conditions. Factors which were optimised include protein concentration, salt, precipitant and buffer concentration, buffer pH and incubation temperature.

2.4.3 Growth of crystals

Initial crystal screens of 96 conditions were established using the sitting drop vapour diffusion method and subsequent optimisation screens were established using hanging drop vapour diffusion.

2.4.3.i Sitting drop vapour diffusion

Sitting drop sparse matrix screens of 96 commercial conditions were established using a Mosquito™ (TTP Labtech) nanolitre pipetting robot, or by hand using automated pipettes. 96 well crystallisation plates were obtained from Greiner and sealed with sealing film with contact adhesive (EasyXtal Sealing Tapes, Qiagen). A 100 µl volume of crystallisation mother liquor was dispensed into each well and 100 nl of protein was dispensed onto each crystallisation shelf (Figure 2.4) from the robot's four sample reservoirs. 100 nl of each mother liquor was then added to each crystallisation shelf. Plates were incubated at 20 °C in a temperature-controlled cabinet. Trays were monitored daily for the appearance of crystals using a Leica MZ-6 crystallisation microscope with colour change filter lens.

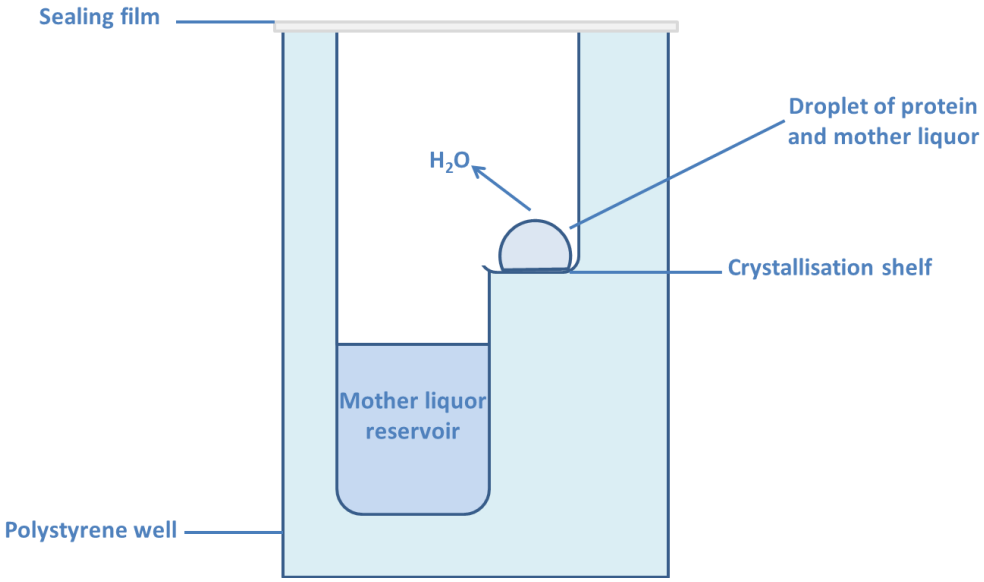


Figure 2.4 Sitting drop vapour diffusion method of crystallisation.

Protein and mother liquor are dispensed onto a crystallisation shelf which sits above the well of mother liquor. The well is sealed and diffusion between the drop and reservoir occurs.

2.4.3.ii Hanging drop vapour diffusion

For the hanging drop vapour diffusion method, 1-2 µl of protein (or protein/ligand mix) were mixed with 1-2 µl of crystallising mother liquor on a siliconised cover slip (18 X 18 mm No.2, Scientific laboratory supplies). 24-well plates (Qiagen) were

prepared by applying High Vacuum Grease (Dow Corning) to the rim of each well using a 1 ml Gilson pipette tip. Then, 500 µl of mother liquor, mixed to homogeneity, was dispensed into each well. Cover slips were set up with drops of protein and mother liquor, then immediately inverted and sealed above the well containing the corresponding mother liquor by firmly pressing the cover slip against the vacuum grease (Figure 2.5). Plates were incubated at 20 °C or 4 °C in the incubator described above. Drops were viewed using the same microscope, monitored daily for crystal growth.

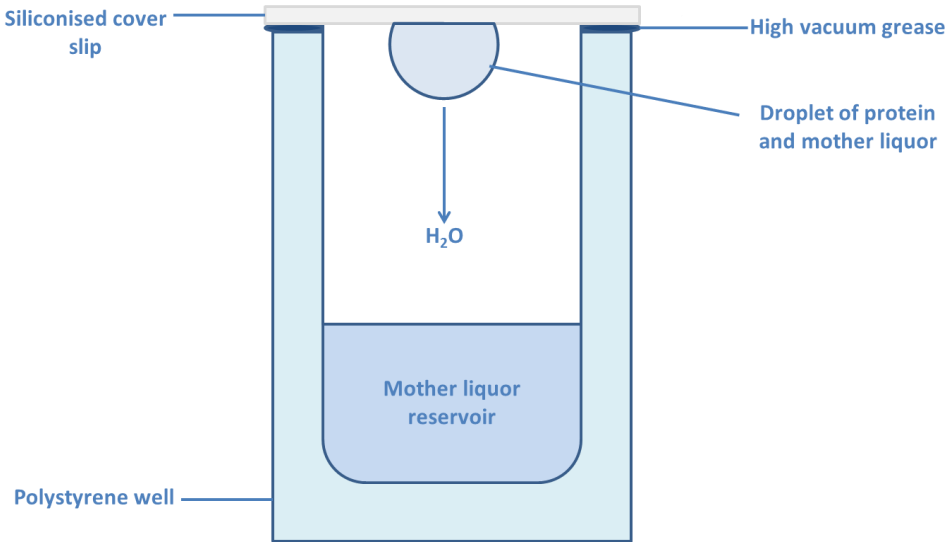


Figure 2.5 Hanging drop vapour diffusion method of crystallisation.

The drop of protein and mother liquor is suspended above the mother liquor reservoir, hanging from the sealing cover slip, attached by surface tension. The use of vacuum grease seals the well so that diffusion can take place.

2.4.4 Crystal soaks

To obtain a ligand-bound protein complex in crystal form, crystals were fished using a magnetic wand and loop (Hampton) and first soaked into fresh mother liquor to check crystal stability. If the crystals were stable, a small amount of pre-prepared ligand (oligosaccharide) was dissolved into mother liquor at near-saturation. A drop (5-10 µl) of this ligand solution was dispensed onto a cover slip. A fished crystal was transferred into this drop, which was immediately inverted and placed over an empty

well in a glass tray with shallow wells. Crystals were soaked at room temperature for between one minute and two hours before being snap frozen in liquid nitrogen.

2.4.5 Collection of X-ray diffraction data

Single three-dimensional crystals were harvested on rayon fibre loops and flash frozen in liquid nitrogen with cryoprotectant comprising mother liquor and 15-25 % glycerol (v/v) to prevent the formation of water ice crystals. Potentially useful crystals were screened on an in-house source (Rigaku rotating RU-200 X-ray generator with a Cu 1.5418 Å target operating at 50 kV and 100 mA with focussing X-ray optics from Osmics), using a Rigaku RAXIS-IV image plate detector. Typically two images per crystal at 0 ° and 90 ° with an oscillation range of 0.5 ° per image to a maximum resolution of 2.2 Å were collected and indexed using DENZO (Otwinowski and Minor, 1997) to assess the diffraction potential of the crystal and to determine space group.

2.4.6 Structure solution

2.4.6.i Single-wavelength anomalous dispersion

Full datasets were collected on a native and a selenomethionine form of the protein. Diffraction images were indexed using iMosflm, a programme which determines crystal orientation and cell parameters, and suggests a likely space group by finding spots in X-ray diffraction images (Leslie 1999, Battye et al., 2011). iMosflm generates reflection lists and integrates images. The output .mtz file, containing integrated intensities and the refined crystal orientation, was input into Scala, which scales and merges intensities (Evans 2005). The logfile generated by the Scala job contains data reduction statistic required for structure deposition, including the redundancy, systematic absences, I/sigI and R_{symm}. Scala was used to output both a scaled and an unmerged .mtz file.

For the heavy atom dataset, the unmerged intensities, converted to scalepack format, were input directly into the ShelxC/D/E programme (Sheldrick 2008). ShelxC checks and scales data and checks for the anomalous signal. This programme also determines the appropriate resolution cut-off for useful anomalous signal. ShelxD then performs a Patterson search for heavy atoms, looking for Patterson correlations

between F^+ and F^- paired reflections with a high anomalous difference. Finally, ShelxE phases the heavy atom sites and uses density modification to determine the remaining phases. Two maps are generates, one for each enantiomorph of the structure; visual inspection of the connectivity of the maps was used to confirm the correct handedness of the solution. Phases were output in .phs format, and were converted to an .mtz file using the phs2mtz script. The programme CAD was used to add the phases output by Shelx to the scaled native dataset, bringing the solution up to the full resolution of the native dataset.

The CCP4 programme Buccaneer (Cowtan, 2006) was used to build the protein model from experimental phases (CAD .mtz file) and a user-provided sequence file in fasta format. Buccaneer performs statistical chain tracing by identifying the positions of Cα atoms in the density. Five internal cycles of model building and refinement were performed. Table 2.17 describes the ten-stage model building process utilised by Buccaneer. Finally the model was refined using Refmac5 (Section 2.4.7).

1	Find Cα by searching electron density
2	Grow chain fragments by adding to identified Cα atoms
3	Join fragments
4	Link N and C termini of chain fragments by adding loops
5	Assign sequence (side chains) to Cα backbone
6	Correcting sequence
7	Remove residues with poor density
8	Build non-crystallographic symmetry (NCS) elements
9	Prune clashes
10	Rebuild side-chain atoms and carbonyl oxygen atoms

Table 2.17 Model-building process of the Buccaneer programme.

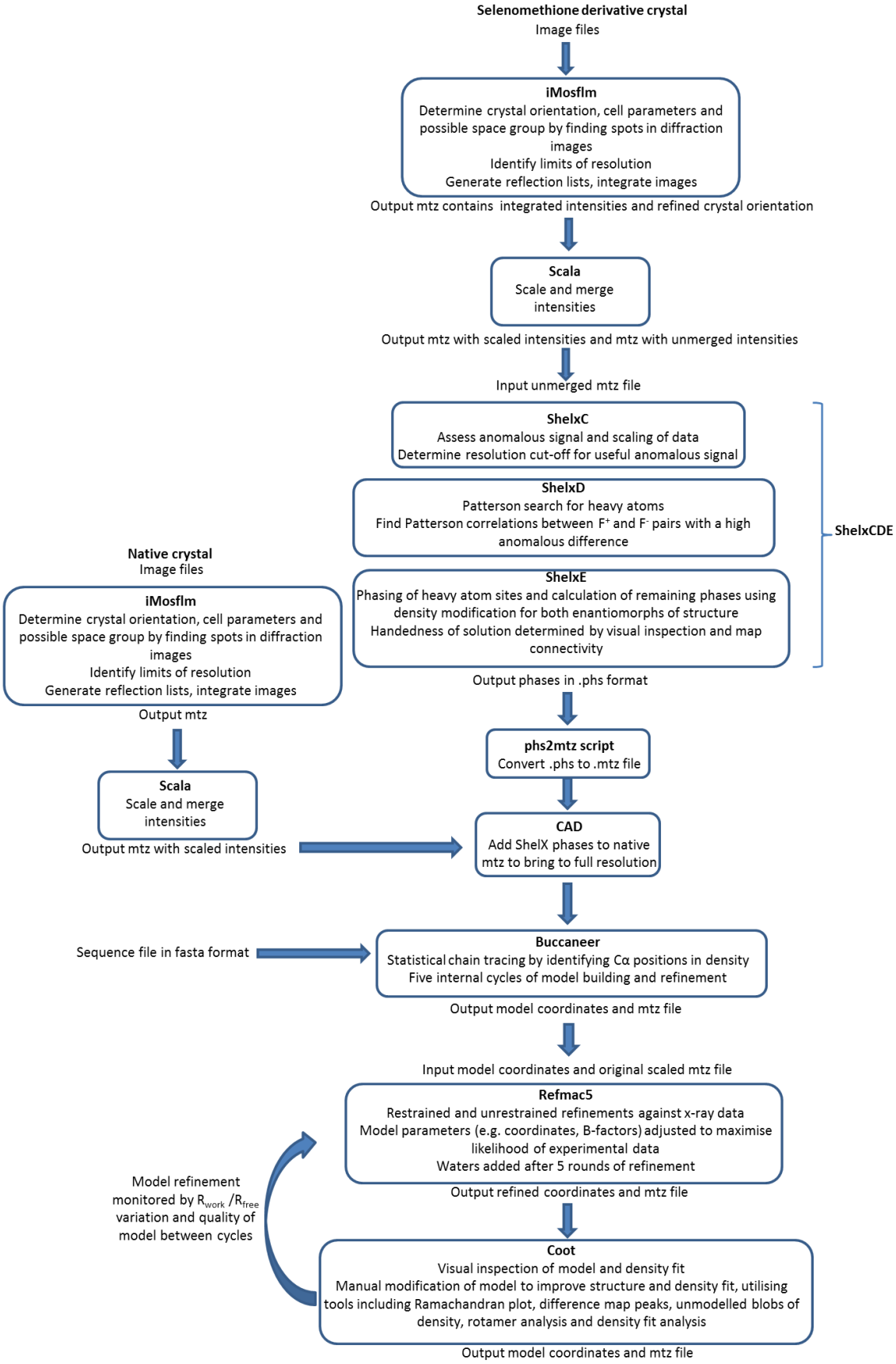


Figure 2.6 Flow chart for structure solution of selenomethionine-derivative protein.

The flow chart describes the process for solution of the structure of a selenomethionine-derived protein using the heavy selenium atoms in single wavelength anomalous dispersion.

2.4.6.ii Molecular replacement

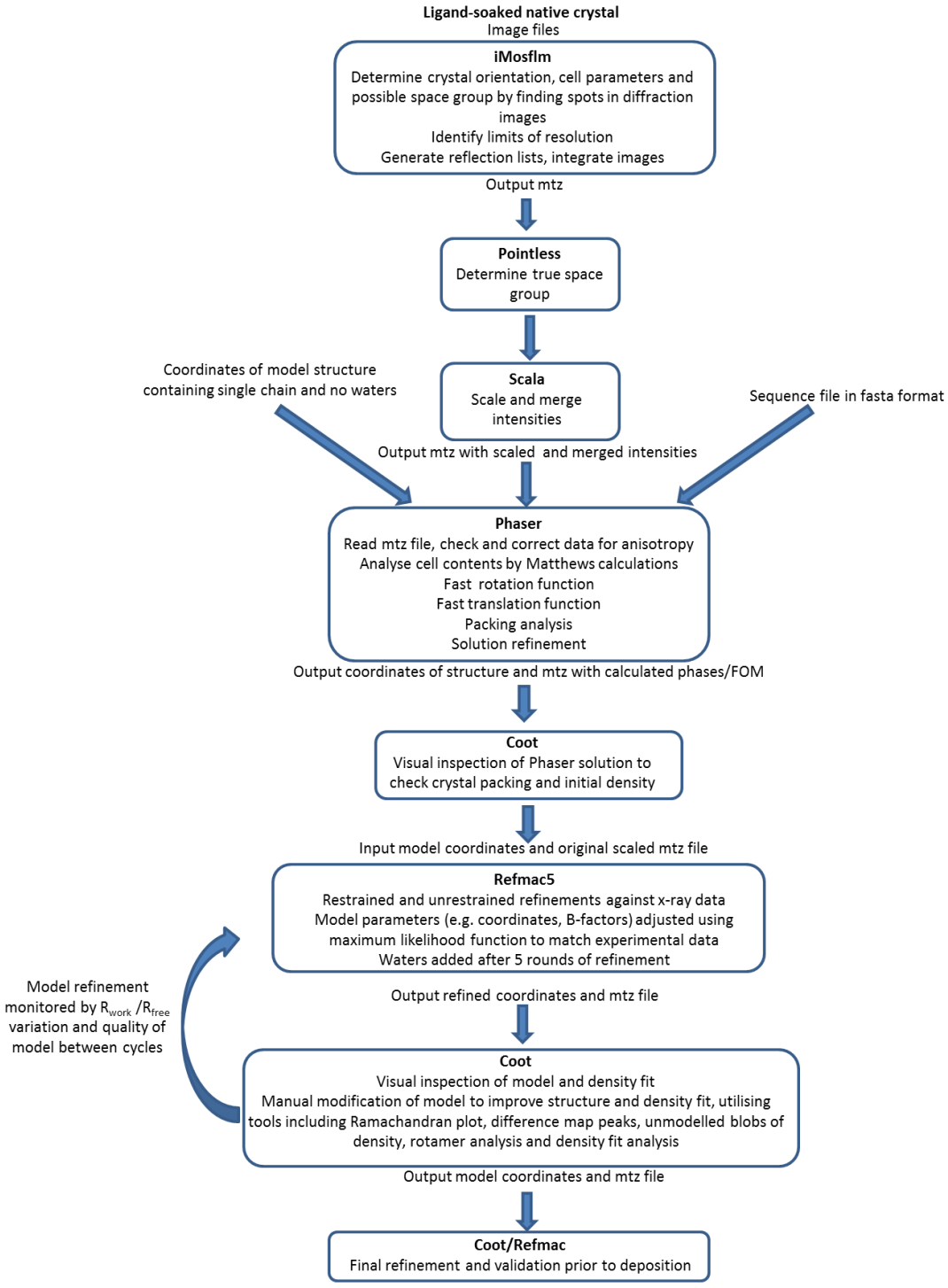


Figure 2.7 Flow chart for structure solution of ligand-bound proteins by molecular replacement.

The flow chart describes the process for solution of the structure of a native protein using molecular replacement.

After indexing in iMosflm and scaling in Scala, as described above (Section 2.4.6.i), scaled and merged intensities (.mtz file) were submitted to the CCP4 Phaser programme (McCoy et al., 2007) along with a sequence file in fasta format and the coordinates of a model structure for molecular replacement. The model coordinates were prepared to contain a single protein chain and no waters or ligands. Phaser reads the .mtz file to check and correct the data for anisotropy. Cell contents are analysed by Matthews calculations. The experimental data is then compared with the model structure, which is moved to align with the experimental cell contents. First, Phaser performs a fast rotation function, which generates multiple solutions, each given a Z score. The top solutions are then submitted to a fast translation function. After analysis of crystal packing and refinement of the solution, experimental phases are output in an .mtz file, along with coordinates of the solution. The structure and initial density were inspected visually in Coot, prior to refinement in Refmac5 (Section 2.4.7).

2.4.7 Structure refinement

A structure obtained by molecular replacement was typically submitted to five initial rounds of refinement in Refmac5 prior to close visual inspection and manual modification of the coordinates (Murshudov et al., 1997). Refmac5 performs rigid body, restrained and unrestrained refinement of a structure against X-ray data. Model parameters are adjusted using a maximum likelihood function to match experimental data. The programme outputs a refined .mtz electron density map and refined structural coordinates (.pdb file).

The starting mtz file was always used in refinement to avoid model bias. Any necessary amino acid changes (for example, replacing Selenomethionine with Methionine) were performed in Coot (Emsley et al., 2010), and this was followed by one or two rounds of restrained refinement. After this, the structure was hydrated using Arp/wArp's Solvate function (Langer et al., 2008). Manual refinement and building of missing amino acids was undertaken in Coot, followed by more rounds of restrained refinement in Refmac5. Any ligands were introduced at this point. Sugar

structures were imported from the library in Coot or downloaded from the Glycam Biomolecule Builder and Oligosaccharide Library (http://glycam.ccrc.uga.edu/CCRC/biombuilder/biomb_index.jsp). Final statistics to assess the quality of structures were obtained by submitting pdb files to the Molprobity server (http://molprobity.biochem.duke.edu/, Vincent et al., 2010).

2.4.8 Visualisation of structures

The software programmes Coot (Emsley et al., 2010) and PyMol (Delano Scientific) were used to view and analyse crystal structures. PyMol was also used to prepare figures presented in this thesis.

CHAPTER THREE

Family 43 glycoside hydrolases of *Bacteroides thetaiotaomicron*

The plant cell wall is a chemically complex composite structure that is an important biological and industrial substrate. Reflecting this diverse chemistry, microorganisms that degrade the plant cell wall synthesise an elaborate array of degradative enzymes, primarily glycoside hydrolases, which display complementary activities. Recent genomic data has shown that several organisms that utilise the plant cell wall as an important nutrient source have the genetic potential to synthesise a large number of glycoside hydrolase family 43 (GH43) enzymes. These organisms hail from many different phyla, show distinct metabolic capacities and are found in very varied habitats. Diverse examples include *Sorangium cellulosum*, a gram-negative cellulose-degrading proteobacterium found in soils all over the world (Lampky, 1971)**,** *Saccharophagus degradans*, another gram-negative proteobacterium which is able to degrade a wide range of complex polysaccharides (Ekborg et al., 2005), and many bacterial symbionts of the human digestive tract. While some GH families have quite limited specificities (for instance, GH92 enzymes are α-mannosidases (Zhu et al., 2010b)), GH43 enzymes have demonstrated activities against a broad range of substrates, with characterised enzymes to date comprising β-xylosidases, α-L-arabinofuranosidases, arabinanases, β-galactosidases and, recently, a xylanase (Zhao et al., 2010). This broad range of activities is hypothesised to be one factor that has led to the great expansion of GH43 in many microbial species.

One particular microorganism showing expansion in GH43 is *Bacteroides thetaiotaomicron*, a gram-negative bacterium which resides in the distal intestine of humans, where it is a dominant microbial species (Xu and Gordon, 2003b). *B. thetaiotaomicron* specialises in the hydrolysis of pectic and host polysaccharides, but is unable to utilise the major plant cell wall polysaccharides cellulose and xylan. According to the CAZy database, the 6.2 Mb *B. thetaiotaomicron* genome codes for enzymes from 42 GH families. As Figure 3.1 shows, some of these families contain a

large number of members, exemplified by GH2 (32 members), GH43 (31 members) and GH92 (23 members).

Within the *B. thetaiotaomicron* genome there is a recurring organisational motif, where multiple genes encoding glycoside hydrolases, carbohydrate binding proteins and other proteins are physically linked and whose transcription is induced by a specific carbohydrate structure (Reeves et al., 1997). These regions are termed Polysaccharide Utilisation Loci (PULs) and are unique to *Bacteroides* species.

In order to test the prediction that the expansion of GH43 in *B. thetaiotaomicron* reflects a diverse range of plant cell wall targets for these enzymes, experiments were undertaken to characterise this family of enzymes, with the expectation of discovering novel specificities, to better understand the evolutionary drivers that have led to the GH43 expansion. To this end, a sister project was also undertaken to characterise the GH43s from *Cellvibrio japonicus*, a saprophytic soil-dwelling organism capable of hydrolysing all major plant cell wall polysaccharides that shows a similar pattern of expansion of GH families (Figure 3.1, (Cartmell, 2010)).

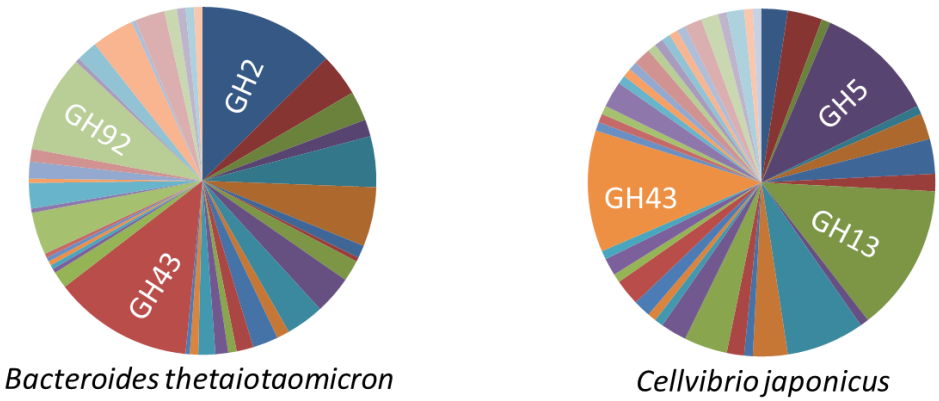


Figure 3.1 Expansion of GH family 43 in two bacterial species.

Pie charts show the distribution of glycoside hydrolases between families in *Bacteroides thetaiotaomicron* and *Cellvibrio japonicus*. Larger segments represent families with more members. Families with zero members are not shown. Data taken from www.cazy.org. In both species, GH43 is one of the most significantly expanded families.

The plant cell wall represents the largest source of recyclable carbon in the biosphere, a significant potential source of raw materials for the production of novel

synthetic chemicals and, importantly, the transport fuel bioethanol (Pauly and Keegstra, 2010). An important bottleneck in the economic utilisation of the plant cell wall is the saccharification of wall components, currently undertaken by expensive or environmentally hazardous pre-treatment processes. Enzymatic degradation of the wall is therefore an important research target, and the discovery of novel glycoside hydrolases is a significant first step towards making petroleum alternatives economically viable.

Enzyme screening projects can often uncover novel, useful enzyme activities which were not readily apparent from sequence analysis and alignment. The screening project described in this chapter aimed to clone, express and biochemically characterise the GH43 enzymes encoded by the genome of *B. thetaiotaomicron.* The findings presented in this chapter showcase an attempt to understand an apparent redundancy in the 'cazome' of *B. thetaiotaomicron* by fully characterising the bacterium's consortium of GH43 enzymes, building upon clues afforded by genomic analysis and induction studies of the PULs. Through analysis of the genome and gene sequences in conjunction with newly acquired knowledge of enzyme activities, certain degradative pathways have become apparent, and a novel specificity has been found, which is presented with the crystal structure of a close homologue from *C. japonicus*. Phylogenetic analyses of protein sequences are presented and discussed in the light of activities described here, and a possible evolutionary history of family GH43 is also discussed.

3. RESULTS

3.1 Bioinformatics

3.1.1 Genomic organisation

3.1.1.i Polysaccharide Utilisation Loci

Polysaccharide Utilisation Loci (PULs) are found throughout the *B. thetaiotaomicron* genome and all have the same general composition. A regulatory protein, which is often a hybrid two component system (HTCS) polypeptide, senses a specific carbohydrate structure and initiates transcription of the PUL. HTCS proteins

incorporate all elements found in classical two-component environment sensing signal transducing systems in one single polypeptide (Sonnenburg et al., 2006). PULs also contain two binding proteins, SusC and SusD, which bind carbohydrates targeted for translocation through the membrane (Reeves et al., 1997; Shipman et al., 2000). SusC is an integral membrane protein, while SusD is thought to be anchored to the membrane through interactions with SusC. These components can be identified by sequence analysis but a third carbohydrate binding partner, SusE, shows no homology to any known proteins and can only be identified, when present at all, by its location immediately downstream of SusD. SusE is thought to be part of the carbohydrate binding complex, increasing the affinity of the protein assembly for the appropriate ligand. Figure 3.2 outlines the typical organisation of the components of a PUL.

In addition to the binding proteins, one or more glycoside hydrolases are also present. Table 3.1 details those PULs which include GH43s, the carbohydrates which induce expression of these PULs, other GHs present and the number of SusC/D carbohydrate binding complexes present.

3.1.1.ii Other GH43 sequences

For the 27 GH43s which reside in PULs, the inducing carbohydrate served as a starting point in activity assays for the enzymes, but four GH43s are not located in PUL operons. These are Bt2959, Bt3655, Bt3656 and Bt4185.

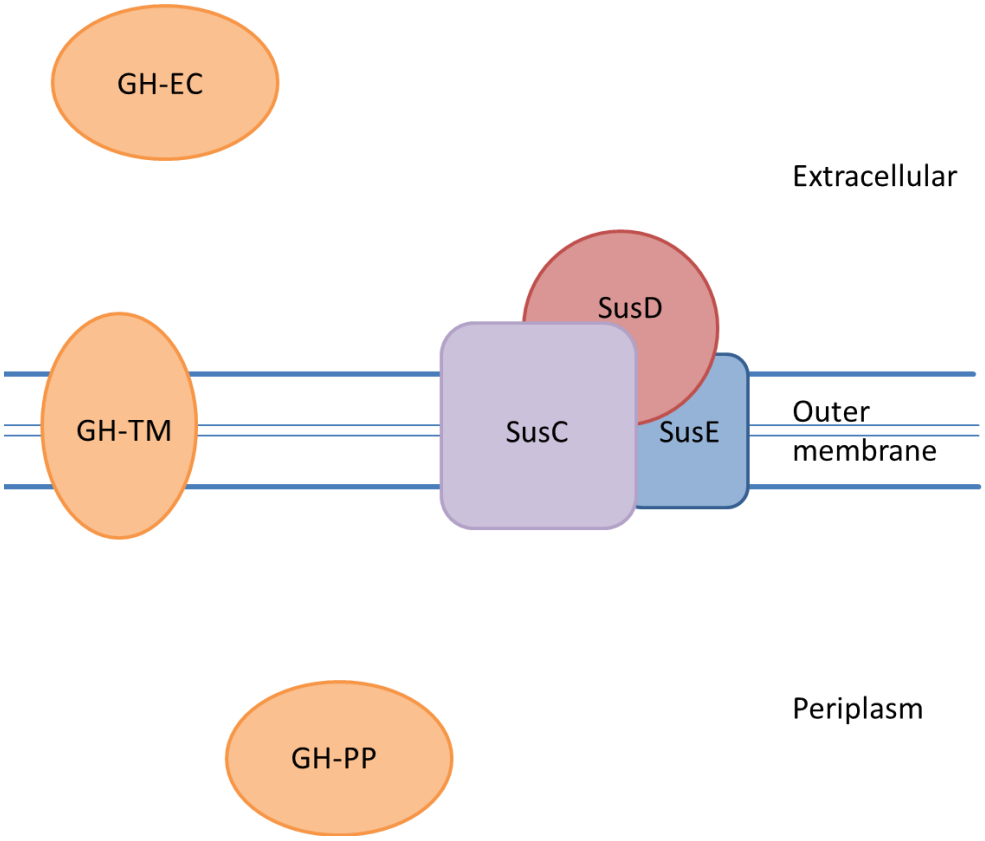


Figure 3.2 Typical cellular organisation of a Polysaccharide Utilisation Locus.

The schematic outlines the typical cellular organisation of the components of a Polysaccharide Utilisation Locus. A complex of Sus binding proteins (SusCDE) are bound to the outer membrane of the cell, with carbohydrate binding surfaces pointing into the extracellular space. Glycoside hydrolase enzymes (GH) encoded by the PUL operon may be extracellular (EC), trans-membrane (TM) or periplasmic (PP). These generate shorter glycans from the inducing polysaccharide, which are translocated into the periplasm, where binding causes the HTCS to induce high-level expression of the PUL.

PUL	ORF Limits	Inducer	GH43 ORFs	Other GHs	Binding proteins	Other proteins
2	0139-0146	n.d.	0145	GH88	SusCD	
5	0262-0290	AG	0264 0265	GH35/CBM32	HTCS SusCD SusE	
7	0348-0369	sbA pGal	0360 0367 0369	2 x GH51	HTCS 2 x SusCD SusE	Arabinose isomerase Aldose epimerase Xylulose kinase
13	0977-1030	RG II	1021	PL1 GH28 GH79	2 x SusCD	
23	1768-1781	n.d.	1781	2 x GH2 GH95 2 x GH3 GH92	SusCD SusE	
24	1871-1878	n.d.	1873	GH3 GH92 GH97	-	
26	2103-2113	n.d.	2112	2 x GH92	SusCD SusE	
39	2851-2923	n.d.	2852	GH2 GH3 GH36	HTCS SusCD SusE	
40	2892-2897	n.d.	2895	GH93	HTCS SusCD SusE	
41	2898-2910	n.d.	2898 2900	-	SusCD SusE	
42	2912-2923	n.d.	2912	-	SusCD SusE	
49	3092-3109	n.d.	3094 3108	GH2 GH51	HTCS SusCD SusE	Sulphatases

PUL	ORF Limits	Inducer	GH43 ORFs	Other GHs	Binding proteins	Other proteins
59	3461-3507	n.d.	3467	GH76	HTCS 4 x SusCD SusE	Sulphatases
60	3515-3532	n.d.	3515 3516	2 x GH76 2 x GH92	SusCD	
64	3662-3672	n.d.	3662 3663	GH29 GH97	SusCD	
65	3674-3687	AG	3675 3683 3685	-	HTCS SusCD SusE	
74	4069-4096	n.d.	4095	2 x GH92	2 x SusCD SusE	
77	4145-4183	n.d.	4152	2 x PL9 PL11 3 x GH2 GH27 4 x GH28 GH35	2 x HTCS 2 x SusCD	

Table 3.1 Components of those PULs which include GH43s.

The *B. thetaiotaomicron* genome codes for over 100 PULs, of which 18 contain GH43 enzymes. The metabolic components of these are summarised in the table. The limits of the PULs refer to gene locus numbers in the format Btxxxx. The predicted substrate for several of the PULs is host and/or residual dietary glycans. The inducing polysaccharide ('Inducer') for certain of the PULs is known. These are coded in the table as follows; AG: arabinogalactan, sbA: sugar beet arabinan, pGal: pectic galactan, RGII: rhamnogalacturonan II. n.d. = not determined.

3.2 Cloning and Expression

GH43 genes were amplified using the polymerase chain reaction (PCR) and the resultant DNA cloned in to appropriate expression vectors by ligation of complementary ‘sticky ends’, generated by restriction endonuclease enzymes.

3.2.1 Preliminary predictions of characteristics

Prior to cloning, nucleotide and amino acid sequences of the *B. thetaiotaomicron* GH43 enzymes were submitted to several forms of analysis, using the tools outlined in Chapter Two, Section 2.1. Table 3.2 gives the size (base pairs) of each gene, any internal restriction endonuclease recognition sites, predicted protein size, the nature of any signal peptides, and cleavage points for the generation of the mature protein.

3.2.2 Cloning and expression of GH43s

Primers were designed to clone the GH43 genes into the vectors mini-pRSETA and mini-pRGST (Appendix A). All primers (forward and reverse) began with the non-palindromic GC-rich sequence GTCGCC at the 5’ end to promote annealing. This was followed by the appropriate six base restriction site, with additional bases to either side of this sequence to ensure the following sequence was in the correct frame for cloning. Following the restriction site, typically 18-21 bases of the gene sequence were included, ideally culminating in a GC-rich area (‘GC clamp’) which helps promote correct binding at the 3’ end of the sequence. mini-pRSETA encodes an N-terminal His_6 tag while mini-pRGST encodes an N-terminal glutathione-S-transferase fusion protein, which can improve expression and protein solubility. These plasmids were chosen as they contain the same multiple cloning sites (MCS) so that the same pair of PCR primers could be used to clone a gene into both vectors. For some genes, internal restriction sites meant that these plasmids were not appropriate; these genes were therefore cloned into pET21(a) which carries a C-terminal His_6 tag, or pGEX which provides an N-terminal GST tag (Appendix A).

Figure 3.3 shows an example of an agarose gel analysis of several PCR reactions. Following amplification, PCR products were purified and submitted to restriction digestion with appropriate endonuclease enzymes. The plasmid vector was digested

concurrently, and both were purified and assessed for purity and concentration by agarose gel electrophoresis. Vector and insert were ligated and transformed into Top10 cells. If colonies were obtained, then 3 – 5 were inoculated separately into 5 ml of LB containing the appropriate antibiotic for overnight growth at 37 °C. Plasmid DNA was extracted from the resultant cultures by miniprep and a small portion (5 µl) submitted to restriction digestion by the endonuclease enzymes used in cloning. These reactions were subjected to agarose gel electrophoresis to check for the presence of DNA inserts (Figure 3.4). Plasmids showing the presence of insert were sent for automated sequencing by MWG or Macrogen. If no insert was found for a particular gene, additional colonies were screened before the cloning experiment was repeated.

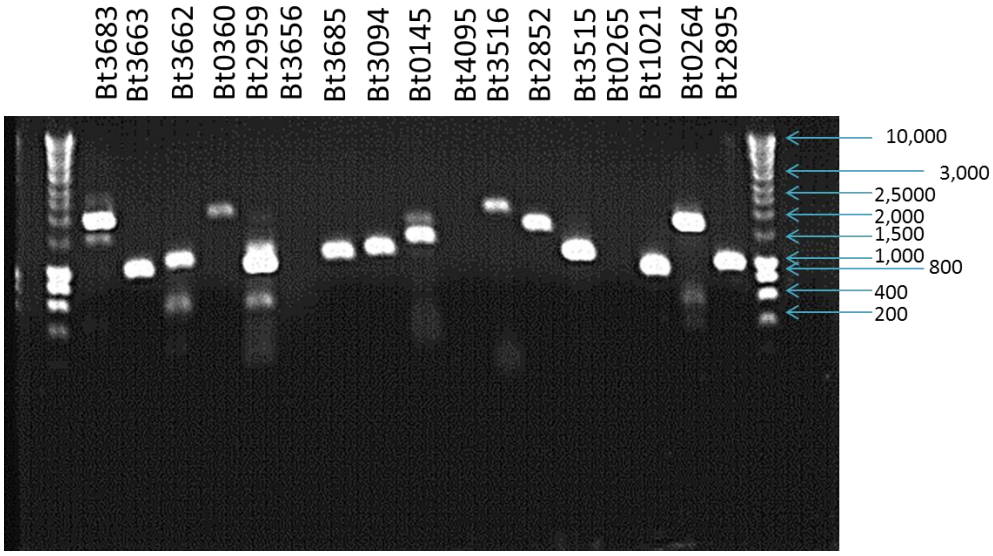


Figure 3.3 PCR amplification of seventeen genes.

Example of a typical agarose gel electrophoretic run to analyse a PCR experiment aiming to amplify seventeen GH43 genes from the *B. thetaiotaomicron* genome. Amplification in this case was successful for all but three genes (Bt3656, Bt4095 and Bt0265), which were obtained at a later date in optimised PCR experiments. Size markers shown are derived from Hyperladder I (BioLine) and indicate the size in base pairs (b.p.) of DNA fragments.

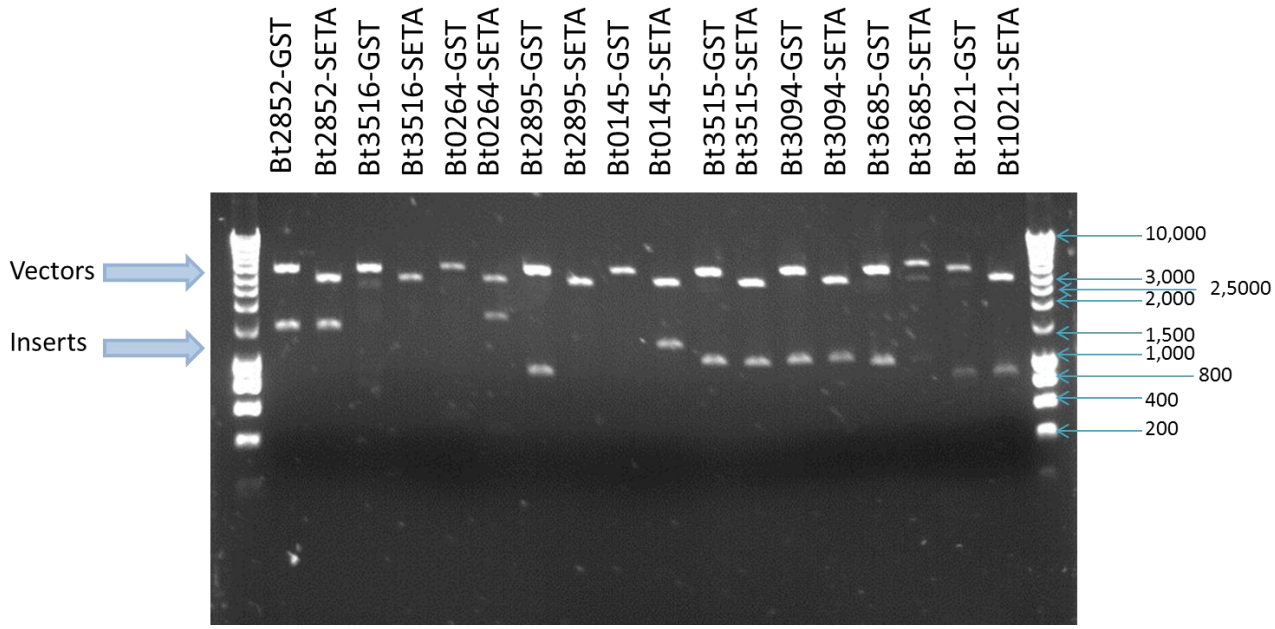


Figure 3.4 Screening plasmid DNA for inserts.

Restriction digestion of plasmid DNA extracted from Top10 cells was used to identify successful cloning experiments by the presence of an insert of the correct size. In each lane, the larger band corresponds to the vector and the smaller band (where present) corresponds to the cloned gene as an insert in the vector. The two vectors shown for each gene (mini-pRGST and mini-pRSETA) are slightly different sizes. Those plasmids showing inserts of the correct size were submitted for automated sequencing. Additional colonies were screened for genes where the initial plasmids extracted contained no inserts.

GH43	Gene size (bp)	Relevant internal restriction sites	Signal peptide	Cleavage site	Protein size (kDa)
0145	1038	None	None	-	46.9
0264	1749	HindIII	SpI	30-31 FAT-QSK	62.0
0265	1482	HindIII	SpI	18-19 IFS-QNT	54.6
0360	1926	None	SpII	19-20 LAA-CSD	71.6
0367	1542	NcoI	SpII	25-26 FVS-CTS	58.7
0369	966	NdeI	SpI	21-22 LSA-QND	36.4
1021	852	NcoI	None	-	32.6
1781	2040	BamHI EcoRI NdeI	None	-	77.6
1873	984	EcoRI	SpII	21-22 IVS-CSN	34.1
2112	1038	EcoRI HindIII NcoI	SpII	17-18 LSC-QSS	39.0
2852	1545	HindIII	SpI	20-21 AGA-QTK	56.1
2895	993	None	SpII	20-21 FTC-CGE	34.1
2898	984	NcoI NdeI	SpI	22-23 VNS-STV	37.5
2900	978	EcoRI HindIII	SpI	19-20 LMA-EDP	36.5
2912	4635	EcoRI NdeI	SpI	20-21 VYG-QEH	172.8
2959	1383	EcoRI HindIII	SpI	22-23 ARK-TEK	52.0
3094	1131	NcoI NdeI	SpI	23-24 ASA-QAY	40.4
3108	990	None	SpI	23-24 KTA-KEE	38.0
3467	1098	None	SpII	17-18 NIS-CNS	41.7
3515	1170	NcoI	SpII	36-37 GVS-CQH	40.6

GH43	Gene size (bp)	Relevant internal restriction sites	Signal peptide	Cleavage site	Protein size (kDa)
3516	1914	HindIII NdeI	SpI	20-21 ISA-QNK	69.4
3655	972	NcoI	SpI	22-23 GYS-QQS	36.9
3656	2436	None	SpI	19-20 AYS-QEY	92.7
3662	1386	EcoRI NcoI NdeI	SpI	21-22 VKA-QHN	52.5
3663	1362	HindIII NcoI	None	-	52.3
3675	1002	EcoRI	SpI	23-24 VTA-QNK	34.9
3683	1896	BamHI	SpI	21-22 IFA-QDD	72.4
3685	1068	NcoI	SpI	21-22 MRA-QKN	38.2
4095	1539	HindIII NcoI	SpI	19-20 MVA-QEQ	56.1
4152	2961	EcoRI NdeI	None	-	113.3
4185	1677	EcoRI	SpI	20-21 AVA-QKN	60.4

Table 3.2 Sequence analysis of *B. thetaiotaomicron* GH43s undertaken prior to cloning experiments.

Characteristics of the GH43 genes and proteins were established prior to cloning, and are summarised in the table. Gene sequences were analysed for internal recognition sites for the most commonly used restriction enzymes: BamHI, EcoRI, HindIII, NcoI, NdeI and XhoI. Protein sizes were predicted using the ProtParam programme at the ExPasy server.

Signal peptides were detected by the LipoP prediction tool. SpI = signal peptide. SpII = lipoprotein signal peptide.

Those plasmids which contained the gene of interest in the correct reading frame and with no mutations were transformed into Tuner and BL21 cell strains for small-scale expression trials.

A streak of colonies was inoculated into 10 ml of selective LB and grown at 37 °C until an absorbance at 600 nm of 0.6. At this point, protein expression was induced by addition of IPTG to 0.2 mM; cultures were subsequently incubated at 37 °C for 4 hours or at 16 °C overnight, to assess the effects of temperature upon protein expression. Cells were harvested by centrifugation and resuspended in 1 ml of Talon buffer (20 mM Tris/HCl buffer pH 8.0 with 300 mM NaCl). A 40 µl sample of this whole cell fraction was retained for analysis by SDS-PAGE. Following sonication, the suspension was again centrifuged to separate soluble and insoluble fractions; samples (40 µl) of both of these were retained for later analysis. A scaled-down version of the protocol for protein purification by immobilised metal affinity chromatography (IMAC) described in Chapter Two, Section 2.2.21 was used to purify these small samples. Again 40 µl of the fraction eluted with imidazole was retained for analysis by SDS-PAGE.

Figure 3.5 shows a typical SDS-PAGE of a small-scale expression trial. Figure 3.6 shows the same protein scaled up to large-scale expression in a 1 litre culture. Table 3.3 summarises successful cloning experiments, including primers and plasmids used, and grades expression of each enzyme. While correctly sequenced clones were obtained for all 31 genes, only 27 expressed protein in *E. coli*. Of these, 25 gave soluble protein, while Bt2898 and Bt4152 could not be solubilised and thus were not pursued. Three proteins (Bt3675, Bt1873 and Bt2912) when expressed with a His_6 tag could not be purified by IMAC due to processing or burial of the tag, and so were purified by ion exchange chromatography. Certain genes were cloned into multiple plasmids in an attempt to improve purification. As an example of the multiple cloning experiments which were performed for some genes, Figure 3.7 charts the cloning, expression and purification profiles of Bt3675.

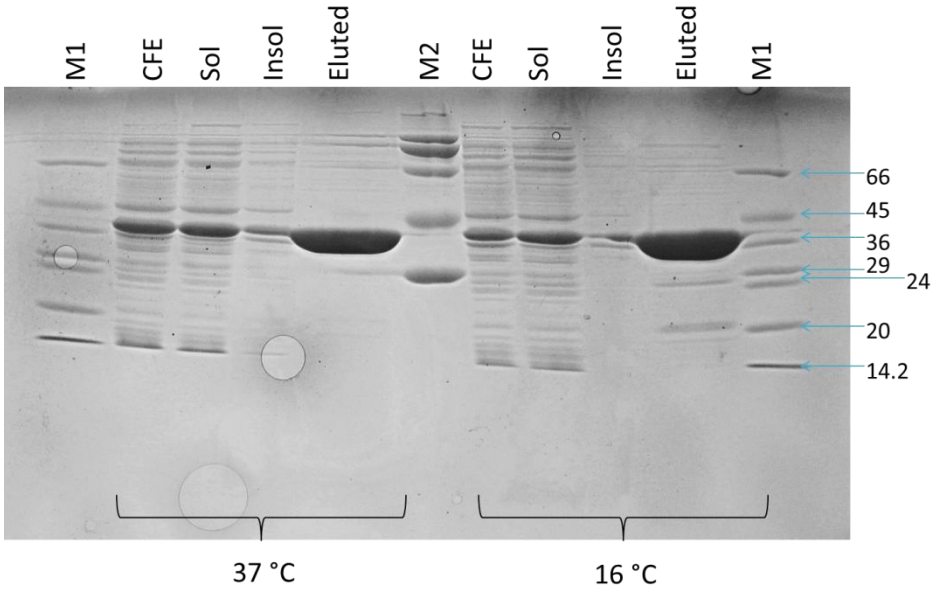


Figure 3.5 Small scale expression trial of Bt3094-pRSETA.

Small-scale expression trial of Bt3094 in Tuner cells with post-induction growth at 37 °C and 16 °C. A band corresponding to the over-expressed protein (~41 kDa) is visible in cell-free extract (CFE), soluble fraction (Sol) and the eluted fraction. The protein is not significantly present in the insoluble fraction (Insol). M1 and M2 are molecular weight ladders. Sizes of standards in M1 are shown in the gel. Those in M2 are of sizes 24, 45, 66, 97.4, 116 and 205 kDa, beginning at the bottom of the gel.

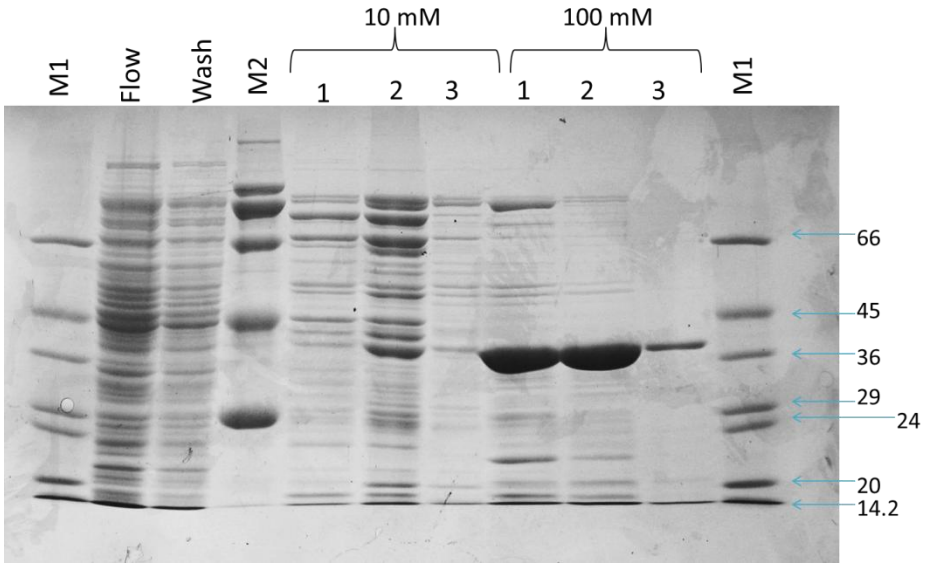


Figure 3.6 Large scale expression of Bt3094-pRSETA.

Bt3094 expression scaled up to a 1 litre culture of Tuner cells. M1 and M2 are protein standards as described for Figure 3.5. Over-expressed protein primarily eluted in 100 mM fractions of imidazole. Flow: flow-through from IMAC column. Wash: wash-through from column.

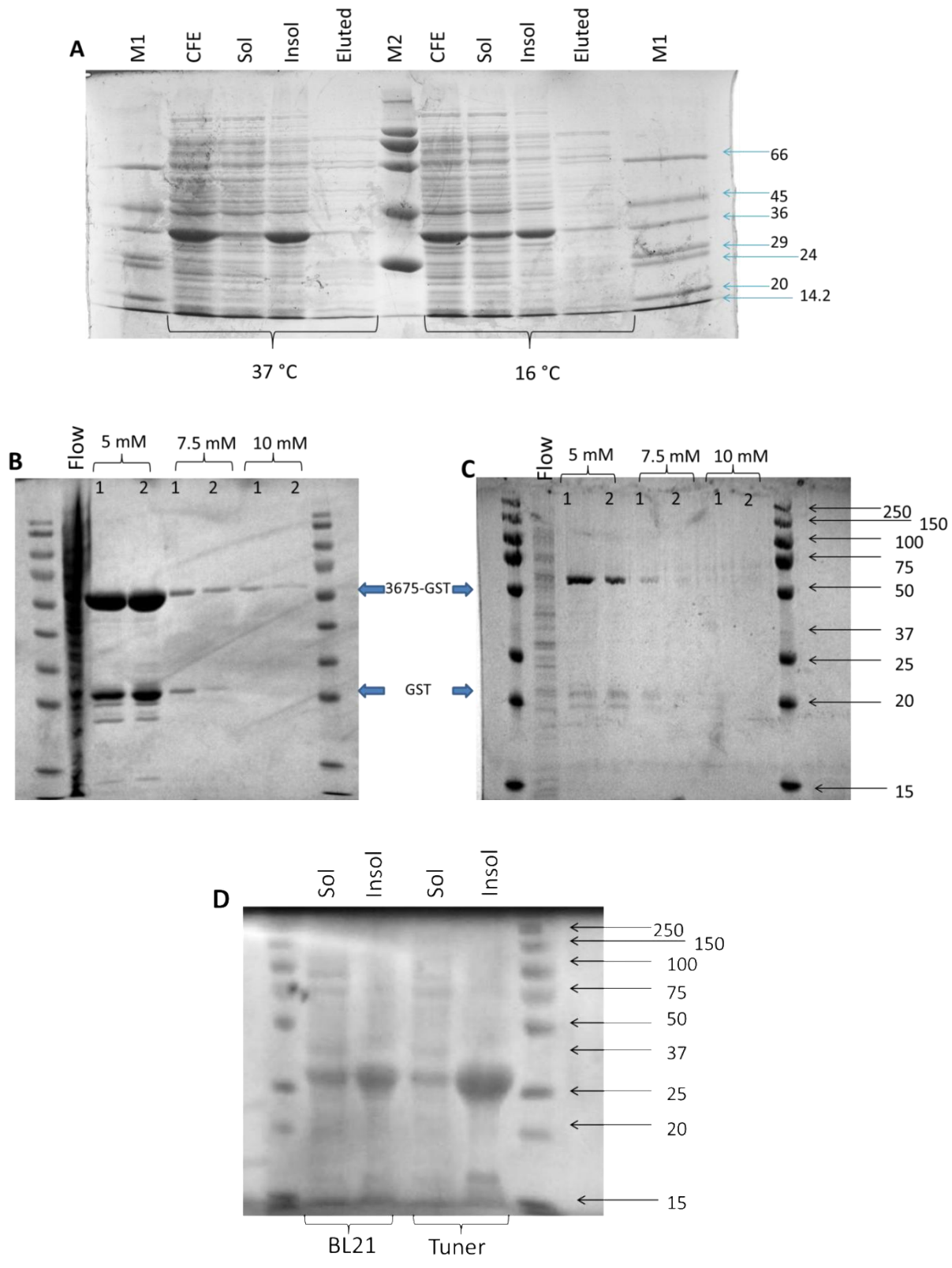


Figure 3.7 Expression profiles of Bt3675.

Bt3675 was cloned into several plasmids to attempt to obtain pure, soluble protein.

A: plasmid pRSETA gives mainly insoluble protein. M1 and M2 size markers are as described above. Attempts to purify the soluble obtained with growth at 16 °C by IMAC were not successful, as the protein failed to adhere to the column (not shown).

B: plasmid pRGST gives a low yield of soluble Bt3675 fused to GST. This purification gel shows that there is extensive degradation and loss of the fusion tag. Protein which has lost its tag becomes insoluble (not shown). The majority of protein (3675-GST and GST alone) elutes in 5 mM glutathione.

C: reduced growth time of Tuner cells transformed with the pGEX construct leads to reduced expression but also allows reduces processing of the fusion tag. A small amount of soluble protein can be purified by affinity chromatography.

D: plasmid pET21(a) gives a reasonable yield of poorly soluble protein which appears to be roughly 27 kDa, smaller than expected for Bt3675, possibly caused by processing due to the presence of a C-terminal His_6 tag. The small amount of soluble protein obtained from the pET21(a) construct again could not be purified by IMAC. The protein was purified by ion exchange chromatography and the pure protein quickly became insoluble.

Gene	Vector	Forward primer	Reverse primer	Restriction sites	Expression profile	Affinity purification
BT0145	pRSETA	ctccag ggatcc atgataaagaccagagatccg	ctccag gaattc tcatttagttacccggaaccagtc	BamH1, EcoR1	n.e.	n.a.
BT0145	pRGST	ctccag ggatcc atgataaagaccagagatccg	ctccag gaattc tcatttagttacccggaaccagtc	BamH1, EcoR1	+	+
BT0264	pRSETA	ctccag ggatcc caaccagctttcgcaactcaatcg	ctccag gaattc ttattgtggttttccgtaagcagcaaattc	BamH1, EcoR1	++	++
BT0265	pRSETA	ctccag ggatcc caaaacacccagatcagtccg	ctccag gaattc ttattgggctttcatctgttttagattattagc	BamH1, EcoR1	++++	++++
BT0360	pRSETA	ctccag ggatcc gcgtgc agcgacgacgatgaaaactcc	ctccag gaattc ttggagtttcttcccccaatatgtc	BamH1, EcoR1	+	+
BT0367	pET21a	ctccag ggatcc aattcttgg gatgataattatttatcag	ctccag gaattc tttgattctttttccccagac	BamH1, EcoR1	+++	+++
BT0369	pRSETA	ctccag ggatcc cagaatgacactacttttgtagc	ctccag gaattc ttgtacgccttccgttgtcatctgaatgcg	BamH1, EcoR1	++++	++++
BT1021	pRSETA	ctccag ggatcc atgttcacttcatttcatgagccg	ctccag gaattc ttatttcttcttgctttcttccag c	BamH1, EcoR1	Insol	n.a.

Gene	Vector	Forward primer	Reverse primer	Restriction sites	Expression profile	Affinity purification
BT1021	pRGST	ctccag ggatcc atgttcacttcatttcatgagccg	ctccag gaattc ttatttcttcttgctttcttccag c	BamH1, EcoR1	++	++
BT1781	pET21a	ctccag cc aagctt g gacggctcgggccgtggttggc	ctccag c ctcgag gg tta tttagtttgcttcaccatttttatagtgc	HindIII, Xho1	n.e.	n.a.
BT1873	pET21a	ctccag gccatgg cc aatgatgatggacagacttcc	ctccag gaagctt cc attcagagctacaaattttccatc	Nco1, HindIII	++	n.p.
BT2112	pRSETA	ctccag ggatcc gacacatcacacaagtatacagc	ctccag c ctcgag gg gaattcgtatgtatcaggtgaccagc	BamHI, XhoI	++	++
BT2852	pRSETA	ctccag ggatcc caaacaaagaatgtgacatgg	ctccag gaattc tcagttaaagtcgtaagtgaaccag	BamH1, EcoR1	++	++
BT2895	pRSETA	ctccag ggatcc gagagtgcaacagatgacgag	ctccag gaattc ttacttcagaactctcggcttaattacag	BamH1, EcoR1	+++	+++
BT2895	pRGST	ctccag ggatcc gagagtgcaacagatgacgag	ctccag gaattc ttacttcagaactctcggcttaattacag	BamH1, EcoR1	+	+
BT2898	pRSETA	ctccag ggatcc agtacagtggataaaggtgataacaatgc	ctccag gaattc tta ctttataactttcggtttaataacc	BamH1, EcoR1	n.e.	n.a.

Gene	Vector	Forward primer	Reverse primer	Restriction sites	Expression profile	Affinity purification
BT2898	pRGST	ctccag ggatcc agtacagtggataaaggtgataacaatgc	ctccag gaattc tta ctttataactttcggtttaataacc	BamH1, EcoR1	Insol	n.a.
BT2900	pRSETA	ctccag ggatcc gaa gatcccgctaagacttataagaacc	ctccag c ctcgag gg aaaatgaggtttggggg	BamHI, XhoI	++	++
BT2912	pET21a	ctccag gccatgg cc caagaacatgatggtgattatac	ctccag gaagctt cc ttttaattctatcttctttgcaaaaac	Nco1, HindIII	+	n.p.
BT2959	pRSETA	ctccag ggatcc cggaaaacggaaaaagtagtaaataacg	ctccag c ctcgag gg cagttttacaatatcagtaatgg	BamHI, XhoI	+++	+++
BT3094	pRSETA	ctccag ggatcc caagcatatggaactgctgatac	ctccag gaattc tcattttgcctgataattgtgattagg	BamH1, EcoR1	++++	++++
BT3094	pRGST	ctccag ggatcc caagcatatggaactgctgatac	ctccag gaattc tcattttgcctgataattgtgattagg	BamH1, EcoR1	+++	+++
BT3108	pRSETA	ctccag ggatcc atgaaaacactcatccaatttttattggcg	ctccag gaattc gaaaatgttatcaagccaatgtggaaa	BamH1, EcoR1	+++	+++
BT3467	pRSETA	ctccag ggatcc ggaaaaaaagaagatgttgaaaaag	ctccag gaattc ttgatttggaactgtgccagatgg	BamH1, EcoR1	+	+

Gene	Vector	Forward primer	Reverse primer	Restriction sites	Expression profile	Affinity purification
BT3467	pRGST	ctccag ggatcc ggaaaaaaagaagatgttgaaaaag	ctccag gaattc ttgatttggaactgtgccagatgg	BamH1, EcoR1	+	+
BT3515	pRSETA	ctccag ggatcc cagcataagaaggctacaacagag	ctccag gaattc ttagttctgttttttatgaataggggaacc	BamH1, EcoR1	++	++
BT3515	pRGST	ctccag ggatcc cagcataagaaggctacaacagag	ctccag gaattc ttagttctgttttttatgaataggggaacc	BamH1, EcoR1	+	+
BT3516	pRSETA	ctccag ggatcc caaaacaaaagtaagcttccg	ctccat gaattc ttatctgaagtccggagcagc	BamH1, EcoR1	+	+
BT3516	pRGST	ctccag ggatcc caaaacaaaagtaagcttccg	ctccat gaattc ttatctgaagtccggagcagc	BamH1, EcoR1	+	+
BT3655	pET21a	ctccag ggatcc tacctgtttgtctatttcaccgg	ctccat gaattc catcttgtcgggctttccgtaggcc	BamH1, EcoR1	++	++
BT3656	pET21a	ctccag ggatcc attcgcaagaatacccgaaagtc	ctccat gaattc tttcctgttttcccaatgacttctcag	BamH1, EcoR1	++	++
BT3662	pRSETA	ctccag ggatcc cacaataatccttttggcaatgc	ctccag cc aagctt g tta ataaatattccattcccagataccg	BamH1, HindIII	n.e.	n.a.

Gene	Vector	Forward primer	Reverse primer	Restriction sites	Expression profile	Affinity purification
BT3662	pET21a	ctccag gagctc cacaataatccttttggcaatgc	ctccag ac ctcgag a ttaataaatattccattcccagatac	SacI, XhoI	n.e.	n.a.
BT3662	pRGST	ctccag ggatcc cacaataatccttttggcaatgc	ctccag cc aagctt g tta ataaatattccattcccagataccg	BamH1, HindIII	n.e.	n.a.
BT3663	pRSETA	ctccag ggatcc agacagatcagtgaagaaatgaactgg	ctccag gaattc tta atataaataaataaactcatcc	BamH1, EcoR1	n.e.	n.a.
BT3663	pET21a	ctccag gagctc atggtgatca tacattccaa gg	ctccag ac ctcgag a ttaatataaataaataaactcatcc	SacI, XhoI	n.e.	n.a.
BT3663	pRGST	ctccag ggatcc agacagatcagtgaagaaatgaactgg	ctccag gaattc tta atataaataaataaactcatcc	BamH1, EcoR1	n.e.	n.a.
BT3675	pET21a	ctccag gccatgg cc cagaataagaaatccggcaatcc	ctccag gaagctt cc tggtgtagggattacctgtttg	Nco1, HindIII	Insol	n.a.
BT3675	pRSETA	ctcacg gaattc cagaataagaaatccggcaatcc	ctcacg gaattc tggtgtagggattacctgtttg		Largely insol	n.p.
BT3675	pRGST	ctcacg gaattc cagaataagaaatccggcaatcc	ctcacg gaattc tggtgtagggattacctgtttg		++	++

Gene	Vector	Forward primer	Reverse primer	Restriction sites	Expression profile	Affinity purification
BT3683	pET21a	ctcacg gaattc caggacgactggcaacttgtttgg	ctcacg actcgagca atcaaaattccacctatccatcc	EcoRI, XhoI	n.e.	n.a.
BT3683	pRSETA	ctccag ggatcc caggacgactggcaacttgtttgg	ctccag ggatcc atcaaaattccacctatccatcc	BamH1, EcoR1	n.e.	n.a.
BT3683	pRGST	ctccag ggatcc caggacgactggcaacttgtttgg	ctccag ggatcc atcaaaattccacctatccatcc	BamH1, EcoR1	n.e.	n.a.
BT3685	pRSETA	ctccag ggatcc cagaagaatagttatattattcccgggaag	ctccag gaattc ctatttttctgtttgtcagatttcttgtatttgtc	BamH1, EcoR1	+	+
BT3685	pRGST	ctccag ggatcc cagaagaatagttatattattcccgggaag	ctccag gaattc ctatttttctgtttgtcagatttcttgtatttgtc	BamH1, EcoR1	n.e.	n.a.
BT4095	pRSETA	ctccag ggatcc caagaacaaaactacttcacaaatcc	ctccag gaattc ttattttaaattcttcattttaaaataagaaaactc	BamH1, EcoR1	++	++
BT4152	pRSETA	ctccag ggatcc ctgcatcttgcctacagttatgatggg	ctccag cc aagctt g tta cattgaaaacaccagtactcctgcgg	BamH1, HindIII	n.e.	n.a.
BT4152	pRGST	ctccag ggatcc ctgcatcttgcctacagttatgatggg	ctccag cc aagctt g tta cattgaaaacaccagtactcctgcgg	BamH1, HindIII	Insol	n.a.

Gene	Vector	Forward primer	Reverse primer	Restriction sites	Expression profile	Affinity purification
BT4185	pET21a	ctccag gccatgg cc cagaaaaactatgtatccgaagtatgg	ctccag gaagctt cc tttccttgtaattcggaaccagtc	Nco1, HindIII	+	+

Table 3.3 Primers utilised in cloning experiments and expression success of each cloned construct.

The table summarises cloning experiments undertaken for the *B. thetaiotaomicron* GH43s. No soluble protein was obtained for Bt1781, Bt2898, Bt3662, Bt3663, Bt3683 or Bt4152. n.e. = not expressed. n.p. = could not be purified. Insol = no soluble protein obtained. n.a. = not applicable.

3.3 Activity screen

Pure soluble protein was submitted to a hierarchy of activity screens, the results of which will be discussed in turn in subsequent sections of this chapter. Activity screens began by incubating a high concentration of protein with a range of isolated plant polysaccharides overnight at 37°C and with a range of artificial coloured substrates for 1-2 hours. Table 3.4 summarises results of this initial screen by grading activity of each enzyme against each substrate. Activities identified in this initial screen were investigated more closely in subsequent rounds of experimentation, which are discussed below.

First, regardless of the availability of PUL induction data, all enzymes were incubated overnight at ~100 nM with a range of polysaccharides (sugar beet arabinan, linear arabinan, arabinogalactan, wheat arabinoxylan, rye arabinoxylan, oat spelt xylan and rhamnogalacturonan (Megazyme and Sigma)) at 2 mg ml^{-1}. These reactions were analysed by TLC; Figure 3.8 shows results of this screen for Bt2852. For those enzymes where carbohydrate induction data was available (see Section 3.6 Discussion, below), this served as a focal point for activity screens. For instance, Bt0360, Bt0367 and Bt0369 are part of a PUL which is induced by sugar beet arabinan and pectic galactan so they were incubated overnight with these substrates at 37 °C. Figure 3.9 shows TLC analysis of products of these reactions. This TLC screen showed that three enzymes (Bt0360, Bt0367 and Bt0369) have activity against sugar beet arabinan, two of which (Bt0360 and Bt0367) are also active against linear arabinan. Bt3675 expressed with a His_6 tag released two products from arabinogalactan. In addition, 11 enzymes released a range of products from all xylan substrates.

The next aspect of the initial screen was incubation with five aryl glycoside substrates: 4NP-β-D-galactopyranoside, 4NP-β-D-glucopyranoside, 4NP-β-D-xylopyranoside, 4NP-α-L-arabinofuranoside and 4NP-α-L-arabinopyranoside (Sigma). Pure enzyme (~ 100 nM) was incubated with the substrates (1 mM) for 2 hours. A control reaction was performed concurrently, where the substrate was incubated with buffer and no enzyme. A positive result was indicated by the

appearance of a yellow colour in the enzyme reaction. Positive results against 4NP-α-L-arabinofuranoside were obtained for Bt0369, Bt2852, Bt3094 and Bt3655.

Following this initial screen, activities identified by TLC were further explored by HPLC analysis to identify reaction products. Typically, substrate at 2 mg ml^{-1} was incubated with pure enzyme at ~ 50 nM in 50 mM sodium phosphate buffer, pH 7.0. Samples were taken at frequent time points and were analysed by HPLC to track the progress of the reaction.

Finally, where appropriate, polysaccharide and aryl glycoside assays were analysed kinetically. Assays against aryl glycoside substrates could be monitored directly at A_{400nm}. For enzymes which released arabinose or galactose from a polysaccharide, the galactose dehydrogenase linked assay was utilised (Chapter Two, Section 2.3.4.iv). A reducing sugar assay with DNSA reagent was used to assay the reactions of enzymes which cleave internal linkages in polysaccharide backbones.

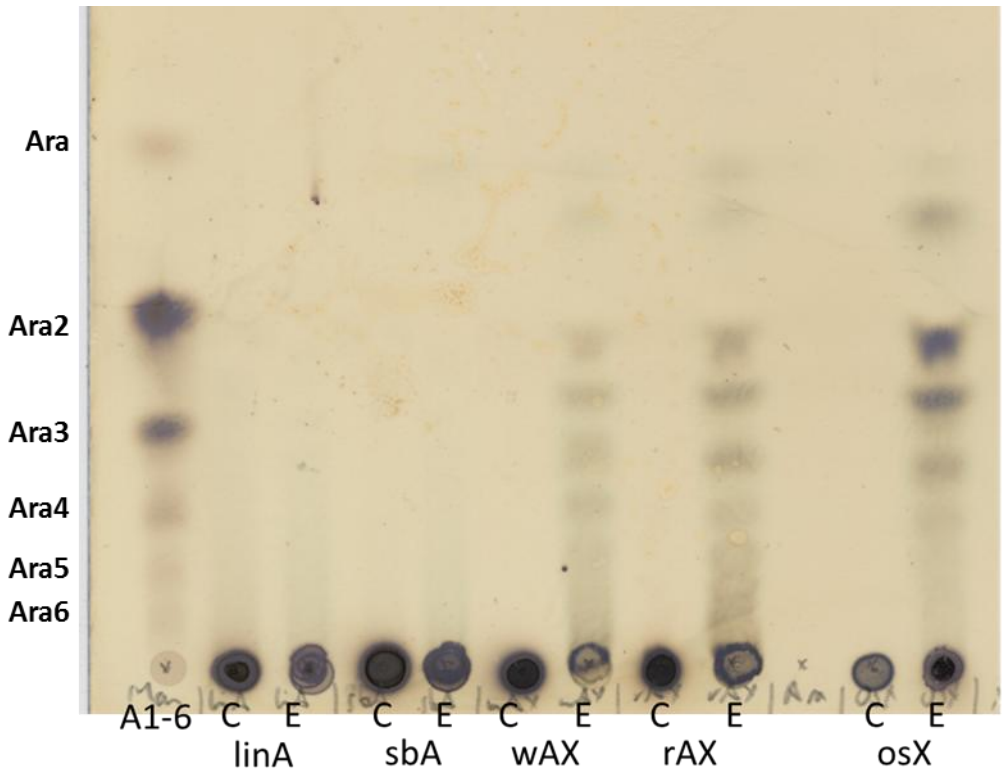


Figure 3.8 Polysaccharide screen of Bt2852 activity, analysed by TLC.

TLC plate shows results of initial activity screen for Bt2852. The enzyme was incubated overnight with a range of substrates. A range of products were generated from each xylan substrate: wheat

arabinoxylan (wAX), rye arabinoxylan (rAX) and oat spelt xylan (osX). No products were generated from Bt2852 incubation with linear arabinan (linA) and sugar beet arabinan (sbA). A size marker ladder of arabinooligosaccharides (A1-6) was included on the TLC plate. The ladder comprised oligomers of d.p. 1-6. Alongside enzyme reactions (E) were spotted control samples (C), where the polysaccharide was incubated at 37 °C in the same buffer but without enzyme.

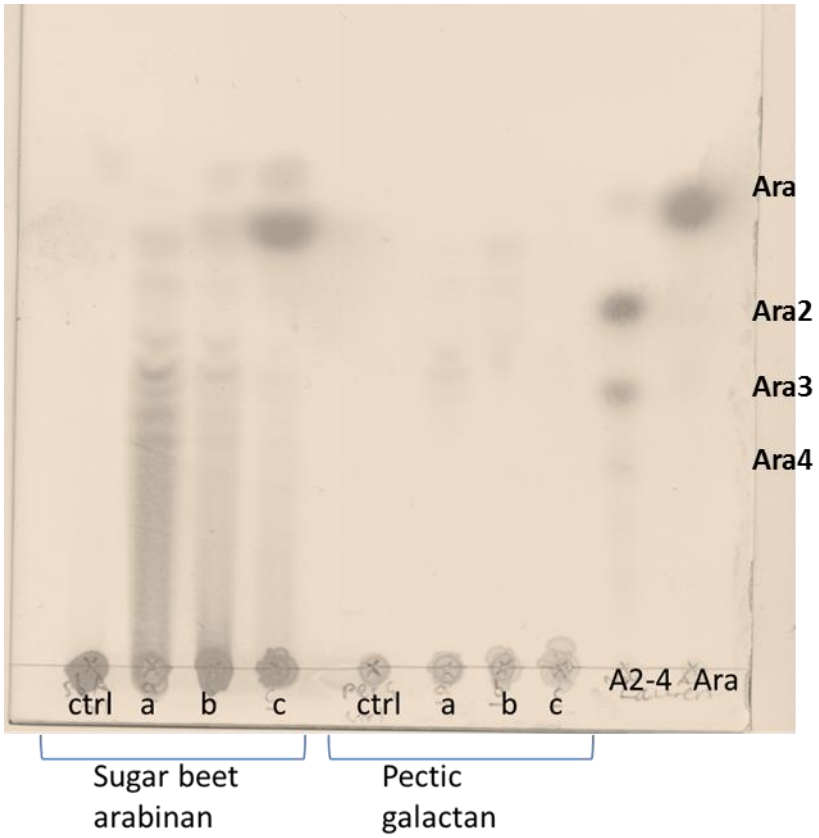


Figure 3.9 TLC analysis of three enzymes upregulated by sugar beet arabinan and pectic galactan.

Bt0360 (a), Bt0367 (b) and Bt0369 (c) were each incubated overnight with sugar beet arabinan and pectic galactan. Control reactions (ctrl) of each reaction were performed, where the polysaccharide was incubated in buffer but without enzyme. Standards were also run on the TLC plate. These were arabinose (Ara) and a ladder of arabinooligosaccharides: arabinobiose, arabinotriose and arabinotetraose (A2-4). From this initial screen, Bt0360 and Bt0367 appear to release a range of products from sugar beet arabinan, while Bt0369 releases arabinose from sugar beet arabinan. These results were later confirmed by HPLC (discussed below). None of the enzymes are active on pectic galactan.

Enzyme	Wheat arabinoxylan*	Galactan	RG I	Sugar beet arabinan	Linear arabinan	Arabinogalactan	4NPA	4NPX	4NPGal
BT0145	-	-	-	-	-	-	-	-	-
BT0264	-	-	-	-	-	-	-	-	-
BT0265	-	-	-	-	-	-	-	-	-
BT0360	-	-	-	++	+	-	-	-	-
BT0367	-	-	-	+	++	-	-	-	-
BT0369	-	-	-	+++	-	-	+++	-	-
BT1021	+/-	-	-	-	-	-	-	-	-
BT1873	-	-	-	-	-	-	-	-	-
BT2112	+/-	-	-	-	-	-	-	-	-
BT2852	++/-	-	-	-	-	-	++	-	-
BT2895	+/-	-	-	-	-	-	-	-	-
BT2900	+/-	-	-	-	-	-	-	-	-

Enzyme	Wheat arabinoxylan*	Galactan	RG I	Sugar beet arabinan	Linear arabinan	Arabinogalactan	4NPA	4NPX	4NPGal
BT2912	-	-	-	-	-	-	-	-	-
BT2959	-	-	-	-	-	-	-	-	-
BT3094	**+/-**	-	-	-	-	-	**+**	-	-
BT3108	**+/-**	-	-	-	-	-	-	-	-
BT3467	-	-	-	-	-	-	-	-	-
BT3515	-	-	-	-	-	-	-	-	-
BT3516	-	-	-	-	-	-	-	-	-
BT3655	**+/-**	-	-	-	-	-	**++**	-	-
BT3656	**+/-**	-	-	-	-	-	-	-	-
BT3675**	-	-	-	-	-	**+**	-	-	-
BT3685	-	-	-	-	-	-	-	-	-
BT4095	**+/-**	-	-	-	-	-	-	-	-
BT4185	**+/-**	-	-	-	-	-	-	-	-

Table 3.4 Summary of results of activity screen on all soluble GH43s.

The table summarises results of the initial activity screen for all GH43s obtained in a pure, soluble form. 4NPA = 4NP-α-L-arabinofuranoside. 4NPX = 4NP-β-D-xylopyranoside. 4NP-β-D-galactopyranoside. RG I = rhamnogalacturonan I.

*Enzymes were assayed against three xylan substrates: wheat arabinoxylan, rye arabinoxylan and oat spelt xylan. These have decreasing levels of arabinofuranose decoration. GH43s showing weak xylanase activity (+/-) were active on all three of these substrates. The notation ++/- for Bt2852 indicates that, while still displaying only weak xylanase activity, this enzyme was more active on the xylan substrates than others showing the same activity.

**Bt3675 polysaccharide assays were performed using cell-free extract of the protein with a His_6 tag, as soluble protein could not be adequately purified (Figure 3.7). All assays were performed alongside cell-free extract of cells transformed with pET21(a) plasmid to check for background *E. coli* activity. Bt3675-His_6 cell free extract released two products from arabinogalactan; this was not observed for the control reaction.

3.4 Activities identified

3.4.1 Weak xylanases

As demonstrated by TLC and HPLC, 11 of the enzymes showed very similar profiles of degradation of multiple xylan substrates (wheat arabinoxylan, rye arabinoxylan, oat spelt xylan and birchwood xylan) which vary in extent of arabinofuranosyl substitution. Figures 3.10 and 3.11 show HPLC analysis of the degradation of wheat arabinoxylan and the less extensively arabinofuranosylated birchwood xylan by Bt2852. Other enzymes displaying the same xylanase activity were much less active than Bt2852, making it difficult to identify any products (examples are shown in Appendix B). Arabinose is a significant early product in the Bt2852 reaction, followed by oligosaccharides of various sizes. Two other enzymes (Bt3094 and Bt3655) also show arabinose release, but it is harder to identify individual peaks in these HPLC traces than for Bt2852. This activity is very low when compared with a classical GH10 xylanase (Figure 3.10); the rate of catalysis is significantly slower and final degradation products display a greater size range which could indicate that specific, rare bonds are cleaved, or that the xylanase action is a side activity. If a rare bond were cleaved, then a product profile established early in the reaction would not change as time progresses. Hydrolysis by the enzymes described here is likely to simply be very slow hydrolysis as products continue to be generated throughout incubation. As Figures 3.10 and 3.11 show, the profile of products generated by Bt2852 stabilises at ~ 24 hours for wheat arabinoxylan and ~ 16 hours for birchwood xylan.

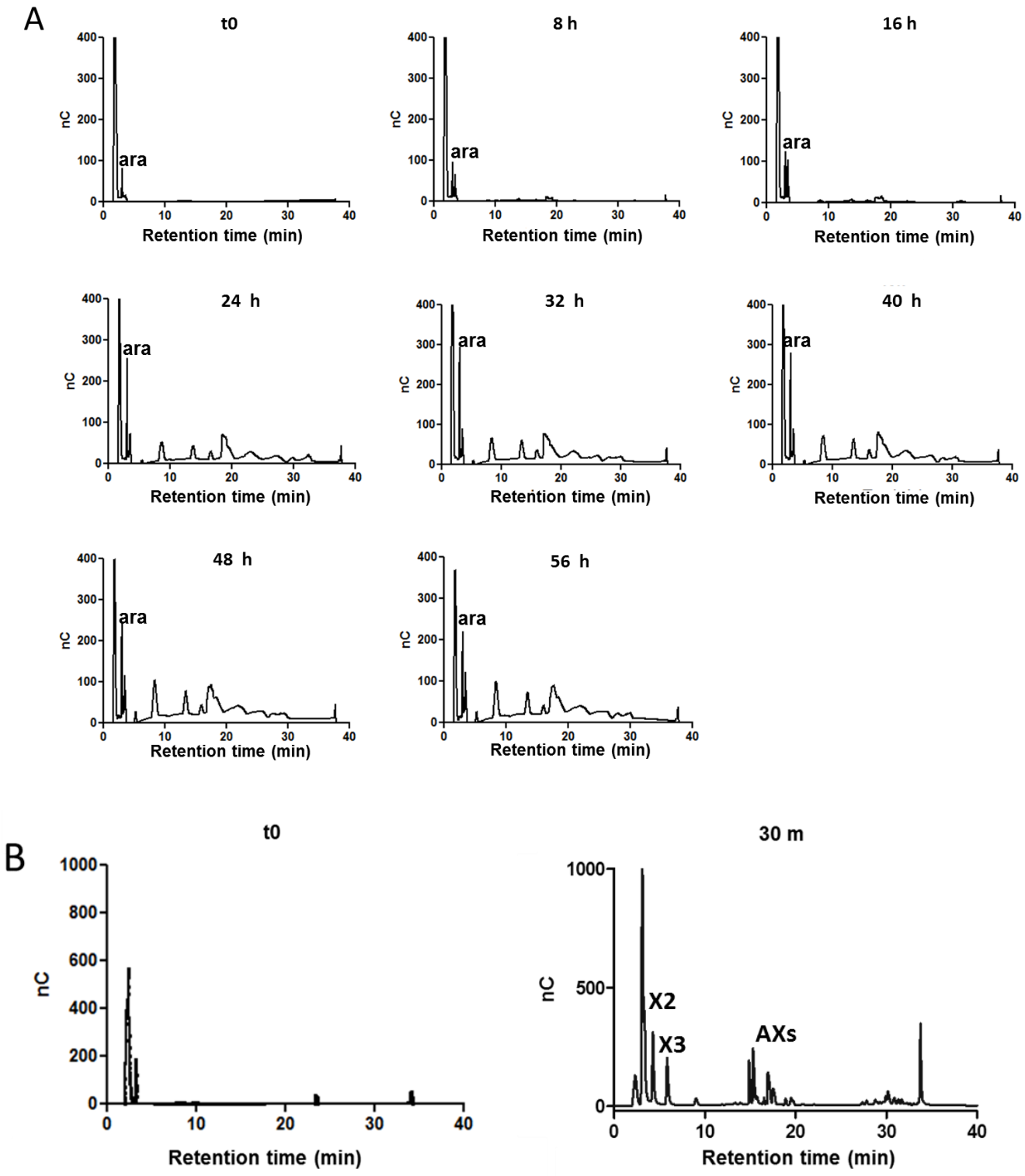


Figure 3.10 Bt2852 degradation of wheat arabinoxylan by Bt2852 and a GH10.

A: HPLC analysis of degradation of wheat arabinoxylan by Bt2852. Enzyme at 2 μM was incubated with substrate at 2 mg ml^{-1}. Arabinose (ara) is a significant product. Oligosaccharides begin to appear

after several hours of incubation. The decorated nature of the polysaccharide meant that these products could not be accurately identified.

B: HPLC analysis of degradation of wheat arabinoxylan by the xylanase *Cj*Xyn10C (Pell et al., 2004b). Enzyme was incubated at 50 nM with 2 mg ml^{-1} of substrate. The product profile is much better defined than for Bt2852, and the reaction proceeded much more quickly. X2 = xylobiose. X3 = xylotriose. AXs = arabinoxylooligosaccharides.

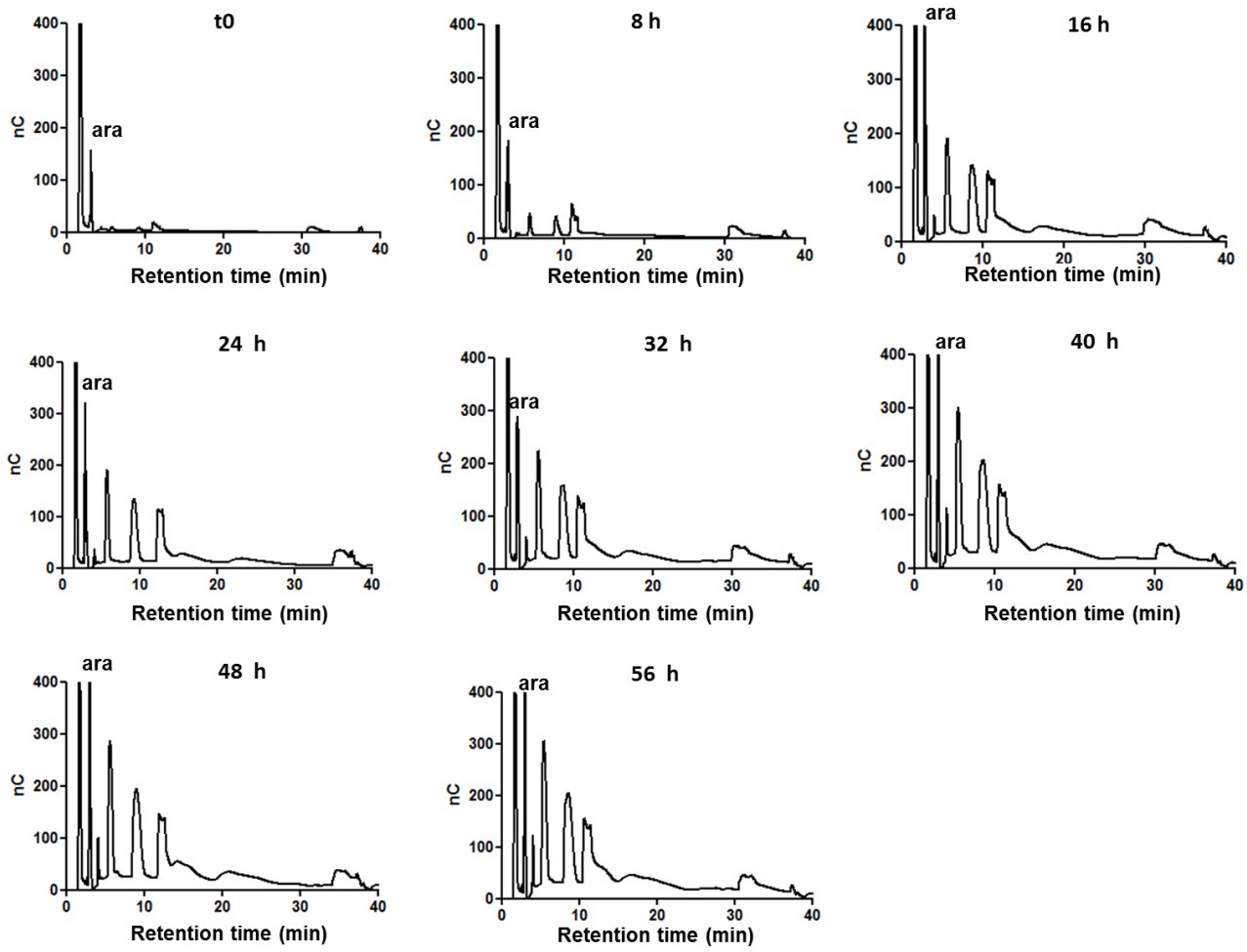


Figure 3.11 Bt2852 degradation of birchwood xylan.

HPLC analysis of degradation of birchwood xylan by Bt2852. Enzyme at 2 μM was incubated with substrate at 2 mg ml^{-1}. Again, arabinose (ara) is a significant product. Oligosaccharides begin to appear after several hours of incubation. The major products do appear to be smaller than those generated from wheat arabinoxylan, suggesting that they are undecorated xylooligosaccharides. This may be reflective of the increased accessibility of the less decorated xylan backbone.

3.4.2 Arabinofuranosidases

Three of the weak xylanases from *B. thetaiotaomicron* also display activity against 4NP-α-L-arabinofuranoside. These are Bt2852, Bt3094 and Bt3655. Table 3.5 gives kinetic parameters for these activities. As Figures 3.10 and 3.11 show, Bt2852 releases arabinose from arabinoxylans with varying degrees of substitution and the initial activity screen (Figure 3.8) showed that this enzyme is not active on arabinan substrates. Thus, Bt2852 can be defined as an arabinoxylan specific arabinofuranosidase (AXH) which also has the capacity to hydrolyse the xylan backbone. HPLC of arabinoxylan degradation showed that Bt3655 and Bt3094 also release arabinose from this substrate, although all products are produced to a lesser degree than in reactions containing Bt2852. It is likely that these enzymes can also be classified as AXHs, although they are less efficient enzymes than Bt2852. This AXH activity could not be explored kinetically for any of the three enzymes as the rate of arabinose release was too slow. As Bt2852 seems to release an approximately equivalent amount of arabinose from both wheat arabinoxylan and birchwood xylan, it can be inferred that the enzyme does not cleave arabinose residues from doubly substituted backbone xylose; data presented in Chapter Four shows that the double substitution is essentially absent from birchwood xylan.

Enzyme	K_M (mM)	k_{cat} (min^{-1})	k_{cat}/K_M (min^{-1} M^{-1})
Bt2852	0.36 ± 0.02	155.5 ± 19.2	4.32×10^5
Bt3094	2.32 ± 0.24	14.56 ± 6.14	6.28×10^3
Bt3655	0.97 ± 0.11	101.3 ± 3.4	1.04×10^5

Table 3.5 Kinetic analysis of three α-L-arabinofuranosidases.

Kinetic parameters for three GH43 arabinofuranosidases as determined by continuous assay performed in triplicate. All of these enzymes also showed xylanase activity against three substrates in TLC and subsequence HPLC analyses. Errors shown are standard errors of the mean (SEM).

3.5 Arabinan utilisation by *Bacteroides thetaiotaomicron*: PUL 7

In this study, PUL 7 has been studied especially closely, as it contains three GH43 enzymes. Figure 3.12 shows a schematic of the operon. As well as the three GH43s (Bt0360, Bt0367 and Bt0369) the locus codes for two GH51 enzymes, one hybrid two component system (HTCS) and two SusC-SusD pairs. Other metabolic enzymes are also encoded, which are not shown in the figure. GeneChip experiments have shown that PUL 7 is upregulated by sugar beet arabinan and pectic galactan. In these experiments, carried out by Prof. Eric C Martens at the University of Michigan, USA, *B. thetaiotaomicron* is grown on a specific carbohydrate and the level of expression of genes is compared with that during growth on glucose. The GeneChip method can only quantify up to a 50-fold increase, so anything above this value gives the same maximum signal. However, qPCR experiments performed in Newcastle by Dr Elisabeth C Lowe show that PUL 7 is activated to a much greater extent by arabinan and arabinooligosaccharides than by galactan. Figure 3.9 shows initial screening results of the PUL 7 GH43s; the screen revealed two arabinanases and one arabinofuranosidase, with no enzyme showing activity against pectic galactan. These data, along with the qPCR results, show that PUL 7 is an arabinan degradative locus.

Figure 3.12 Schematic of the PUL 7 operon.

Genetic elements of the PUL 7 operon are represented by arrows. Only those genes encoding elements of the carbohydrate binding complex (blue), regulators of PUL expression (red) and enzymes involved in carbohydrate metabolism (green) are shown. Not shown are an arabinose isomerase, an aldose epimerase and a xylulose kinase.

3.5.1 Endo-acting arabinanases

As was indicated by the initial TLC screen (Figure 3.9), both Bt0360 and Bt0367 show *endo*-arabinanase activity against linear and sugar beet arabinan. Consistent with other arabinanases, neither showed any activity on 4NP-α-L-arabinofuranoside. Further analysis by HPLC confirmed that both enzymes generate a range of

products from both substrates (Figures 3.13 and 3.14). The decorated nature of sugar beet arabinan prevented accurate identification of reaction products by co-migration with oligosaccharide standards. The major products for Bt0360 are observed in the HPLC trace at approximate retention times of 35, 18, 9 and 5 minutes. For Bt0367, the major products elute at approximately the same times, although peak size is considerably smaller. Mass spectrometric analysis of sugar beet arabinan products generated by Bt0360 (Figure 3.15) shows that products range in size from 2 – 11 pentose residues. These sizes will likely include both backbone and side chain residues.

Kinetic analysis of reducing sugar release by these enzymes was undertaken by the DNSA assay, which was hampered in some cases by inhibition above ~ 8 mg ml^{-1}, possibly due to substrate aggregation at high concentrations. For this reason, only the linear portion of the Michaelis-Menten curve could be accurately measured. Values were therefore obtained for the k_{cat}/K_M ratio, but not the individual values of k_{cat} and K_M. Standard curves of arabinose in the presence of appropriate polysaccharide (10 mg ml^{-1}) were used to determine the concentration of reducing sugars in these assays. Table 3.6 gives k_{cat}/K_M (min^{-1} mg^{-1} ml) for the reactions, which show that the enzymes differ in specificity; Bt0360 shows preference for branched arabinan while Bt0367 displays higher activity against linear arabinan.

Enzyme	Sugar beet arabinan	Linear arabinan
Bt0360	508.50 ± 56.88	57.82 ± 10.22
Bt0367	52.39 ± 3.24	323.91 ± 21.79

Table 3.6 Kinetic analysis of the *endo*-arabinanases of PUL 7.

Values of k_{cat}/K_M for Bt0360 and Bt0367 against sugar beet arabinan and linear arabinan. Units are min^{-1} mg^{-1} ml. The reactions were performed in triplicate. Errors shown are standard errors of the mean (SEM). Individual values for k_{cat} and K_M are not available due to substrate inhibition or a K_M in excess of the limit of practical solubility of the substrate.

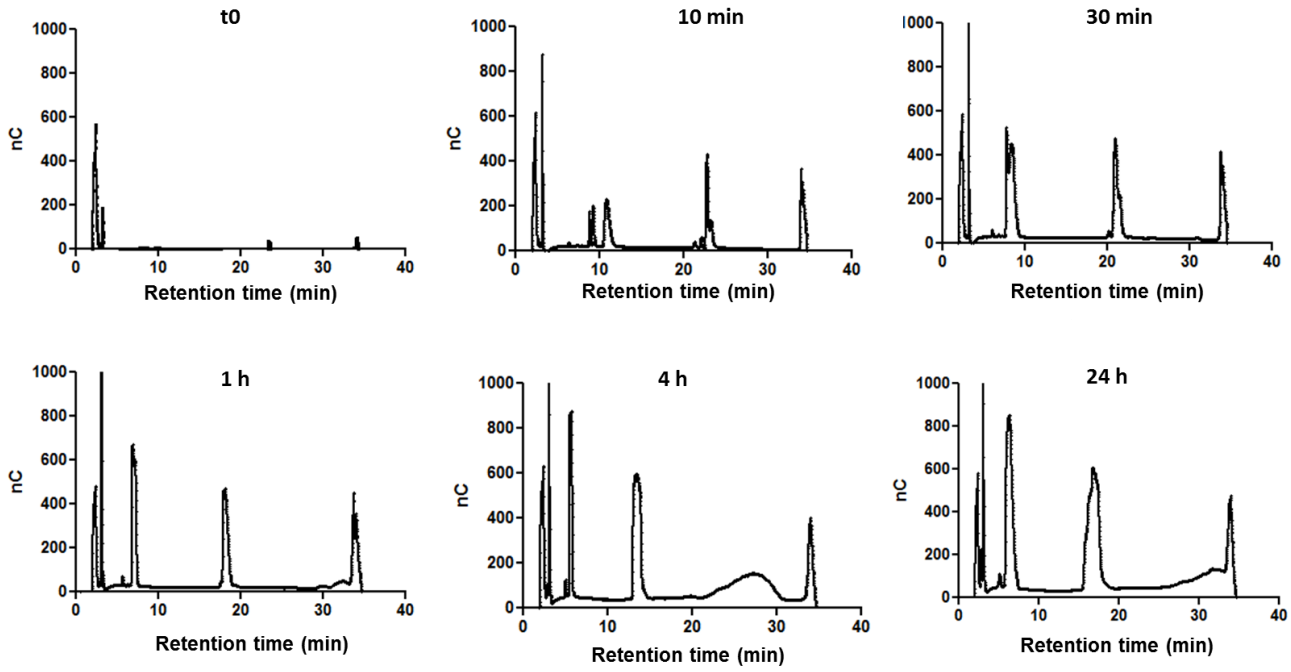


Figure 3.13 Degradation of sugar beet arabinan by *endo*-arabinanase Bt0360.

HPLC analysis of degradation of sugar beet arabinan by Bt0360, an *endo*-arabinanase encoded by PUL 7. Enzyme was incubated at a concentration of 5 μM with substrate at 10 mg ml^{-1}. A range of products is generated over 24 hours.

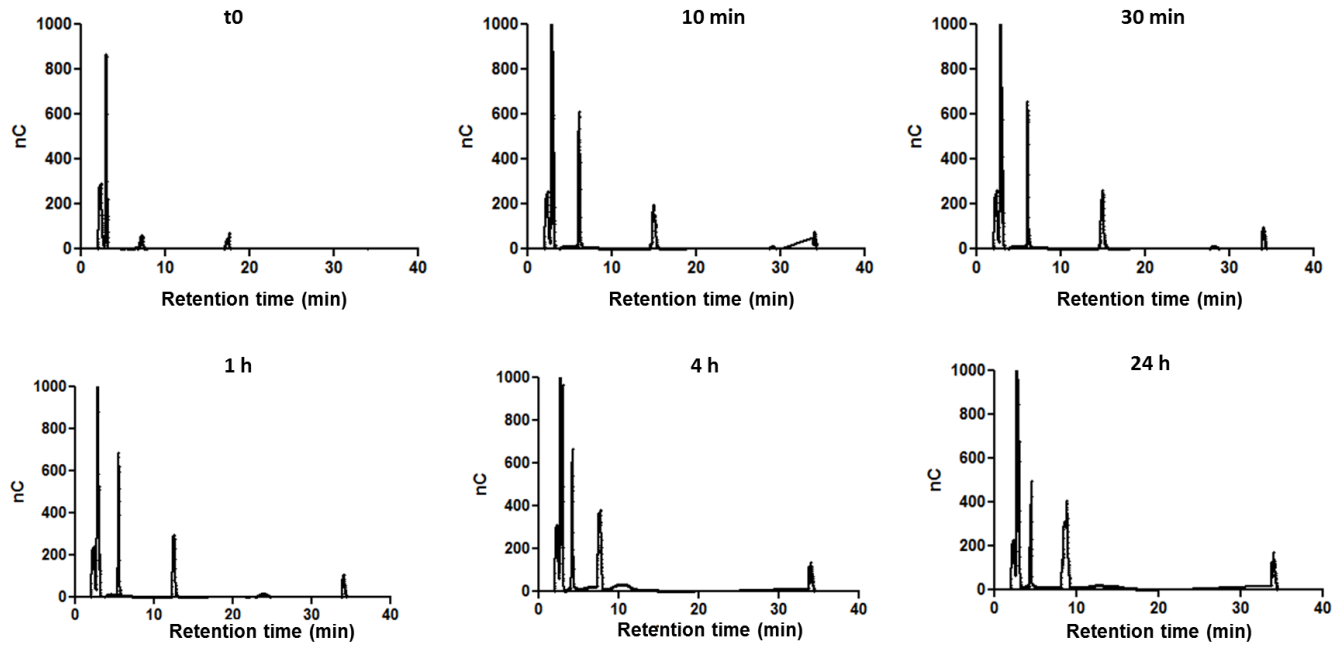


Figure 3.14 Degradation of sugar beet arabinan by *endo*-arabinanase Bt0367.

HPLC analysis of degradation of sugar beet arabinan by Bt0367, an *endo*-arabinanase encoded by PUL 7. The enzyme was again incubated at a concentration of 5 μM with substrate at 10 mg ml^{-1}. A range of products is generated over 24 hours, although degradation is less extensive than with Bt0360.

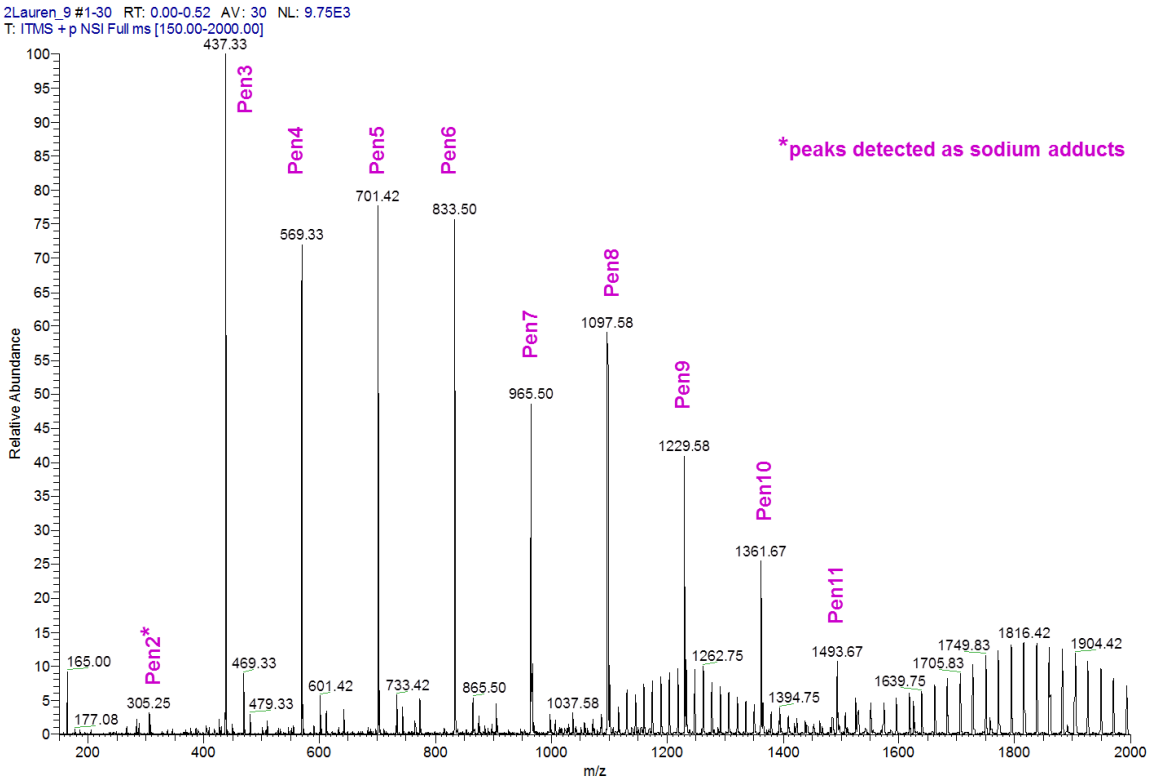


Figure 3.15 Mass spectrometry analysis of degradation of sugar beet arabinan by *endo*-arabinanase Bt0360.

Analysis of degradation of sugar beet arabinan by Bt0360 was undertaken by mass spectrometry (MS), performed by Dr R Naran at the Complex Carbohydrate Research Centre in Athens, Georgia. The substitution patterns of these oligosaccharides are not clear from these data.

3.5.2 Arabinofuranosidase de-branching enzyme

Bt0369, which is found in PUL 7 alongside the *endo*-arabinanases described above, displayed high activity against 4NP-α-L-arabinofuranoside (4NP-A*f*) and sugar beet arabinan, releasing monomeric arabinose as the sole product from the polysaccharide (Figure 3.16), while showing no activity against any other plant cell wall polysaccharides such as linear arabinan, xylans or pectins. Kinetic analysis against sugar beet arabinan, linear α-L-1,5-arabinooligosaccharides and 4NP-A*f* (Chapter Two, Section 2.3.4) showed that the enzyme was ~ 10^5 -fold more active against branched arabinan than against linear oligosaccharides. Although Bt0369 displayed similar activity for arabinan and 4NP-A*f*, the enzyme exhibited greater affinity for the polysaccharide. Table 3.7 summarises the kinetic parameters for Bt0369.

Substrate	K_M (mM)	k_{cat} (min^{-1})	k_{cat}/K_M (min^{-1} M^{-1})
Sugar beet arabinan	$2.1 \times 10^{-2} \pm 1.36 \times 10^{-3}$	2681 ± 246	1.28×10^{7}
4NP-α-L-arabinofuranoside	8.03 ± 2.35	40150 ± 7272.5	5.00×10^{6}
Arabinotriose	17.78 ± 5.79	16.03 ± 2.39	9.02×10^{2}
Arabinoheptaose	n.d.	n.d.	$3.61 \times 10^{3} \pm 3.2 \times 10^{2}$

Table 3.7 Kinetic parameters for Bt0369 hydrolysis of four substrates.

Assays were performed in triplicate. Errors shown are standard errors of the mean. Effective substrate concentration in units of M was obtained from concentration of polysaccharide (mg ml^{-1}) by submitting the polysaccharides to complete degradation by Bt0369. The final absorbance was used to calculate the concentration of target arabinose residues in 1 mg of polysaccharide.

n.d.= not determined: individual values of k_{cat} and K_M were not determined when K_M was too high to accurately measure rate beyond the linear portion of the Michaelis-Menten curve.

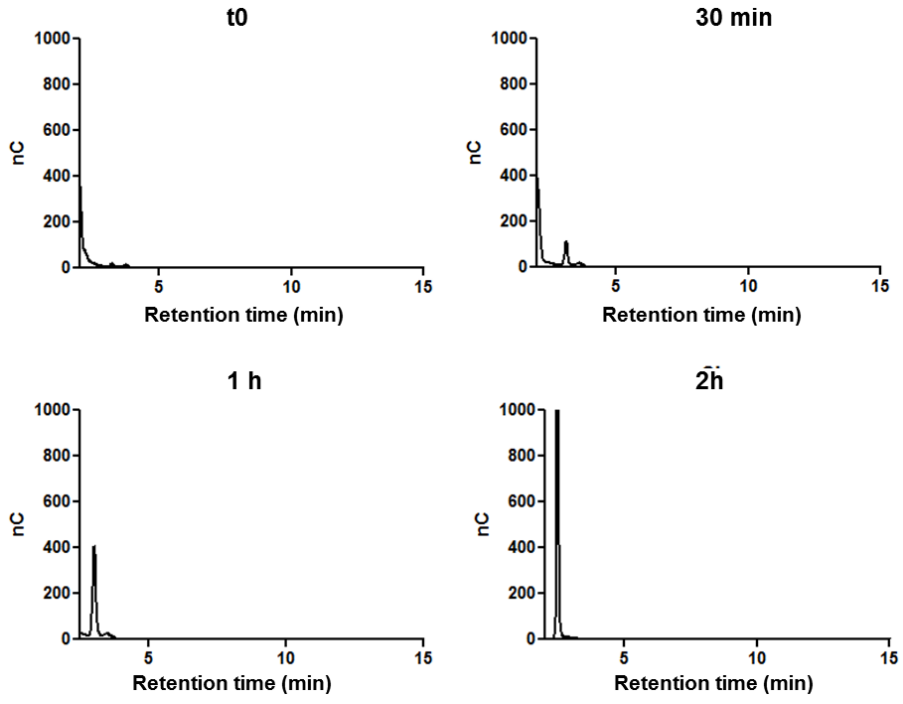


Figure 3.16 Reaction products of Bt0369 incubated with sugar beet arabinan.

HPLC analysis of sugar beet arabinan incubated with Bt0369 for 2 hours shows that arabinose is released as the only product, which was identified by co-migration with monomeric arabinose obtained from Sigma-Aldrich.

Sugar beet arabinan has an α-1,5 linked L-arabinofuranose backbone. Approximately 60 % of backbone residues are substituted with α-1,3-L-arabinofuranose side chains. Some backbone residues are doubly substituted with both α-1,2 and α-1,3 side chains, while a small proportion of single α-1,2-L-arabinofuranose side chains are also present (Swamy and Salimath, 1991). The high activity of Bt0369 against sugar beet arabinan and comparatively very low activity against α-1,5 arabinooligosaccharides suggested that the enzyme hydrolyses the polysaccharide side chains. As ~ 60 % of backbone residues in arabinan are believed to be substituted at O3, the α-1,3 side chains predicted to be present in 1 mg of sugar beet arabinan represent ~ 0.4 mg of arabinose (i.e. 40 % of total mass of polysaccharide), while the amount of arabinose released by Bt0369 from 1 mg of sugar beet arabinan is only 0.045 mg (i.e. 4.5 % of total mass). This was determined by measuring total arabinose released by the enzyme during a one hour incubation; the reaction was monitored to accurately identify the endpoint. This discrepancy suggests that Bt0369 is hydrolysing the rare α-1,2-L-arabinofuranose side chains, or the α-1,3 side chains within a specific structural context.

To test this hypothesis, the specificity of Bt0369 for sugar beet arabinan was probed by Nuclear Magnetic Resonance spectroscopy, performed and interpreted by Maria Peña at the Complex Carbohydrate Research Centre in Athens, Georgia (USA). To simplify the spectra, arabinooligosaccharides were generated by partial acid hydrolysis of sugar beet arabinan, followed by size exclusion chromatography to remove polysaccharide and large oligosaccharides and concentration by lyophilisation. Bt0369 was incubated with arabinooligosaccharides at 2 mg ml^{-1} overnight at 37 °C. Water was removed from the samples by lyophilisation. Immediately prior to NMR analysis, samples were resuspended by M. Peña in a small volume of D_2O.

Chemical shifts for the spectra are given in Appendix E, where the NMR spectra are also shown. The identities of these peaks were identified by M. Peña during this project. Interpretation of the NMR data revealed that the signals for the O2-linked arabinose side chain, in both the single and double substitution, are lost or reduced after enzyme treatment (Appendix E). Signals corresponding to α-1,3 and α-1,5 linked arabinose residues did not disappear. These data demonstrate that the

enzyme targets the O2-linked arabinose side chains of arabinan in the context of single and double substitutions, a plasticity not previously observed for other arabinofuranosidases.

The specificity of the enzyme for the O2 linkage to backbone arabinose residues that carry two substitutions (at O2 and O3) was further examined through synergy experiments with *Hi*AXHd3. This is an arabinofuranosidase that exclusively targets O3 linkages in double substitutions of arabinan and arabinoxylan (Chapter Four; Sorensen, Jorgenen et al. 2006). Arabinan which had been pre-treated with Bt0369 was not hydrolysed by *Hi*AXHd3, as no double substitutions were present. Bt0369, however, was able to release arabinose residues from arabinan pre-treated with *Hi*AXHd3. These results are consistent with the NMR data in showing that Bt0369 targets O2 linkages in backbone arabinose residues that are singly and doubly substituted.

3.5.3 Other PUL 7 components

In addition to the GH43s described above, PUL 7 encodes two GH51 enzymes, Bt0348 and Bt0368 (Figure 3.12). GH51s are invariably non-specific α-L-arabinofuranosidases (Beylot et al., 2001b; Hovel et al., 2003; Taylor et al., 2006). Attempts were made to clone these enzymes to provide a more complete picture of the degradative action of the PUL; Table B.1 in Appendix B details the primers and plasmids used in cloning experiments, as well as the results of expression trials. Soluble protein was not obtained for either enzyme. One clone (Bt0348-pET21) was lethal to *E. coli* (Figure 3.17).

Dr E C Lowe of the Bolam laboratory at Newcastle University is currently engaged in an on-going project to characterise the binding proteins of PUL 7. Data gathered so far show that the HTCS binds arabinoheptaose as its smallest unit. No binding information is so far available for the Sus proteins.

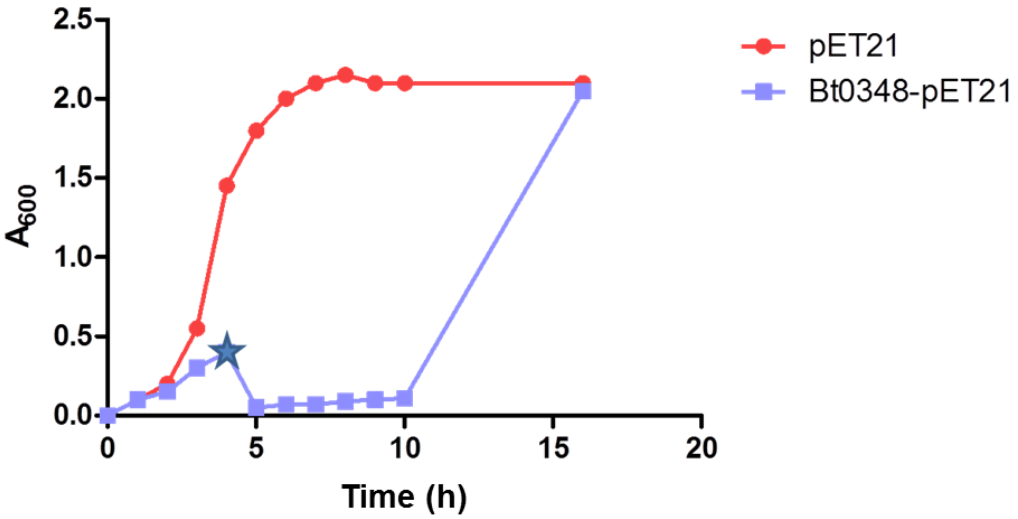


Figure 3.17 Lethality of Bt0348-pET21 to XL1-blue strain *E. coli*.

Presence of the insert slows initial growth until the point marked by the star (4 hours), when A_{600} drops sharply. Subsequent growth is extremely slow, but final absorbance after overnight growth is equivalent to the control culture. Minipreps were performed on samples of the Bt0348-pET21 culture every hour; the plasmid is not found in cells sampled after 4 hours, indicating that the cells have rejected the plasmid.

3.6 Discussion

3.6.1 Metabolism of arabinose-containing polysaccharides

In *B. thetaiotaomicron*, genes encoding enzymes of related function are clustered into polysaccharide utilisation loci (PULs) (Bjursell et al., 2006). PUL 7, which is upregulated by sugar beet arabinan and orchestrates the metabolism of this polysaccharide, contains three GH43 and two GH51 enzymes (E C Martens personal communication). As part of this project, the GH43s from this PUL were cloned, and soluble protein was obtained for all of the enzymes. Activity screens revealed that Bt0360 and Bt0367 are *endo*-arabinanases, while Bt0369 is a novel α-1,2-L-arabinofuranosidase de-branching enzyme. Soluble protein could not be obtained for the GH51s.

3.6.1.i Two *endo*-arabinanases target different regions of arabinan

The GH43 endo-arabinanases Bt0360 and Bt0367 display a preference for decorated and linear arabinan, respectively. These enzymes share 28 % sequence

identity. The catalytic efficiencies of both enzymes are low compared to other arabinanases. k_{cat}/K_M (min^{-1} mg^{-1} ml) for preferred substrate is 509 for Bt0360 and 324 for Bt0367. Classical arabinanases, such as *Bs*Arb43A and *Cj*Arb43A, hydrolyse linear arabinan with k_{cat}/K_M values of 6.0 x 10^6 min^{-1} mM^{-1} and 8.3 x 10^4 min^{-1} mg^{-1} ml, respectively (Proctor et al., 2005). The evolutionary rationale behind these low activities is consistent with data from Dr E C Lowe at Newcastle University which shows that the HTCS of PUL 7 binds arabinooligosaccharides with a minimum d.p. of seven. Thus, restricting the degradation of arabinan before it reaches the regulatory element ensures that expression of the PUL is switched on in the presence of the appropriate carbohydrate. More efficient arabinanases would produce smaller oligosaccharides, and thus the PUL would not be activated. This is an example of the finely tuned nature of the carbohydrate utilisation systems possessed by *B. thetaiotaomicron* which have allowed the bacteria to thrive in the human intestinal community. The cooperative action of the arabinanase enzymes and the carbohydrate binding proteins to bring large oligosaccharides into the periplasm ensures that most hydrolysis and metabolism occurs within the cell, preventing utilisation of the substrate by other bacteria.

Of these arabinanases, Bt0360 in particular is interesting as it shows a clear preference for branched arabinan, while all other arabinanases characterised to date are hindered by the presence of side chains, as occurs in Bt0367. The differing specificity of the enzymes is intriguing, and may be caused by different topologies around the active site. Attempts to crystallise the proteins were not successful so structures are not available, but both were submitted to the Swiss-Model workspace, a powerful automated comparative protein modelling server at ExPasy (Arnold et al., 2006; Schwede et al., 2003). No sufficiently homologous structures are available in the protein database for either *B. thetaiotaomicron* GH43, so reliable structure models were not obtained.

As no structural models could be produced, the protein sequences of the *endo*-arabinanases were analysed for clues as to the nature of the differences in specificity. As shown in Figure 3.18, in the mature protein (lacking signal peptide) both enzymes feature stretches of amino acids which do not correlate with the GH43 domain. It is possible that these regions of sequence produce structural features

which contribute to the differences in specificity between Bt0360 and Bt0367. Figure 3.18 shows schematics of the modular organisation of the *endo*-arabinanases and of the canonical GH43 *Cj*Arb43A arabinanase as a control, and shows the unidentified regions of sequence annotated as 0360-NTD, 0360-CTD and 0367-CTD. These protein sequences were analysed for conserved protein domains by a BLAST search.

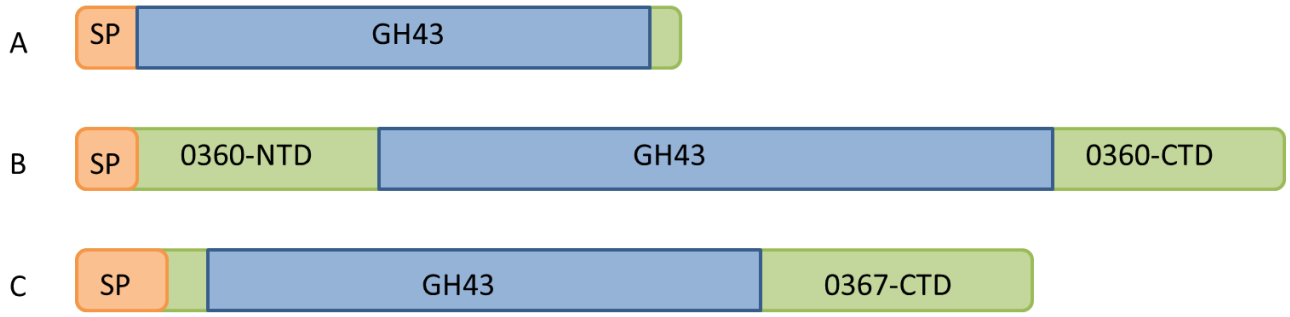


Figure 3.18 Schematic of the modular sequences of three GH43 enzymes.

A: CjArb43A. This classical GH43 arabinanase comprises solely the β-propeller domain found in all members of the family. Residues 1-32 form the signal peptide, which is absent in the mature protein.

B: Bt0360. Residues 1-28 comprise the signal peptide, which is cleaved to produce the mature lipoprotein. The GH43 catalytic domain comprises residues 163-529. Two other significant regions of sequence are also present: 0360-NTD comprises residues 28-163, while 0360-CTD covers residues 529-641.

C: Bt0367. The signal peptide here is composed of residues 1-37. The β-propeller includes residues 62-372, leaving a significant stretch (0367-CTD) at the end of the sequence, comprising residues 373-513.

SP: signal peptide. GH43: canonical GH43 β-propeller fold. Areas shaded green are unidentified modules.

The Blast search revealed that the C-terminal insertions of both enzymes (0360-CT and 0367-CT) align with various other GH43s, including arabinofuranosidases, xylosidases and arabinanases. These regions of sequence are therefore not likely to contribute to the differences in specificity between these *endo*-arabinanases. Sequence 0360-NT, the N-terminal insertion of Bt0360, showed similarity with several GH43s including three putative *endo*-arabinanases (from *Prevotella copri, Gramella forsetii* and *Prevotella oris*) but the alignment also highlighted several

Fibronectin Type III (Fn3) domains. The Fn3 domain is typically around 100 amino acids long and forms a 7-bladed β-sandwich structure. The Fn3 domain is very common in animal proteins but is also found in bacteria, where it is exclusively located in extracellular or membrane-bound glycoside hydrolases (Little et al., 1994). The distribution of Fn3s in bacterial species is very broad and is taken as an example of inter-kingdom transfer of genetic material. Rather than hailing from a common ancestor it is thought that bacterial species acquire the Fn3 gene from an animal source, likely by transformation. In the human gut, bacteria such as *B. thetaiotaomicron* are likely to encounter large amounts of eukaryotic DNA, from host cells and dietary material such as meat. The presence of Fn3 exclusively in glycoside hydrolases is likely due to selective retention of the module in these sequences when Fn3 is scattered through the genome. This implies that there is a selective advantage conferred by Fn3. Putative roles for Fn3 in carbohydrate active enzymes include direct binding and modification of polysaccharides; evidence of this has been taken from the observation that one or more of the fibronectin units in Fn3 can bind to heparin and that an Fn3-like repeat in the *Clostridium thermocellum* cellobiohydrolase CbhA directly modifies the surface of cellulose, facilitating hydrolysis (Kataeva et al., 2002). Furthermore, an Fn3 module was recently shown to be critical for binding to chitin (Martin-Garcia et al., 2010).

If Bt0360 does contain a β-sandwich structure or Fn3-type module which is absent in Bt0367, this may contribute topologically to the active site or polysaccharide binding cleft; other arabinanases comprise solely the GH43 β-propeller domain and are inhibited by arabinose side chains in arabinan, as is Bt0367 (Proctor et al., 2005). The Fn3 module may directly bind to arabinan and facilitate hydrolysis of the backbone by improving enzyme access to substrate for Bt0360. In this manner, the differences in specificity shown by Bt0360 and Bt0367 may not be conferred by modifications to the GH43 catalytic domain tailoring the enzymes to more and less decorated regions of the arabinan backbone, respectively, but by a separate module assisting catalysis by Bt0360, possibly by binding more tightly to certain regions of the polysaccharide, or associated cell wall components. Alternatively, the presence of an additional non-catalytic module may have directly altered the structure of the

active site of Bt0360, creating a pocket adjacent to the active site which can accommodate arabinose side chains on the polysaccharide.

3.6.1.ii A novel α-1,2-L-arabinofuranosidase

As part of PUL 7, these *endo*-arabinanases act in consort with the arabinan-specific arabinofuranosidase Bt0369. Bt0369 has been shown by NMR (Figures 3.17 and 3.18) to selectively cleave O2-linked arabinose side chains. Kinetic analysis shows the enzyme also has reduced activity against linear α-1,5 linked arabinooligosaccharides, so there may be some cleavage of terminal backbone residues following removal of these side chains. NMR and synergy experiments with *Hi*AXHd3, which is specific for doubly substituted arabinose side chains in arabinoxylan and arabinan, showed that Bt0369 can cleave the O2 side chain in both single and double substitutions.

A sister project undertaken by PhD student Alan Cartmell had the similar aim of characterising all GH43s expressed by the soil saprophyte *Cellvibrio japonicus*. One enzyme described by this research, *Cj*Abf43A, shares 65 % sequence identity with Bt0369 and was shown by 2D NMR, again performed by Maria Peña, to exhibit the same specificity as Bt0369 for the O2-linked arabinose side chain in sugar beet arabinan, with tolerance for this linkage in both single and double substitutions. The crystal structure of a catalytic mutant of *Cj*Abf43A was solved to 1.8 Å. The structure of *Cj*Abf43A reveals a five bladed β-propeller fold, Figure 3.19, typical of GH43 enzymes (Nurizzo et al., 2002). Similar to many arabinan-specific enzymes (Alhasid et al., 2009; Nurizzo et al., 2002; Proctor et al., 2005), the enzyme comprises only this catalytic domain, and does not possess the C-terminal β-sandwich domain observed in xylan-specific arabinofuranosidases (Vandermarliere et al., 2009).

The propeller is built by a 5-fold repeat of ‘blades’: four antiparallel β-strands comprising a β-sheet. The blades are radially arranged from the centre of the propeller. For the ligand structure, *Cj*Abf43A was crystallised with arabinotetraose (Megazyme) but the structure revealed good density for an α-1,5 arabinotriose molecule with an α-1,3 linked arabinofuranose side chain on the middle arabinose (3^2-Ara-Ara_3).

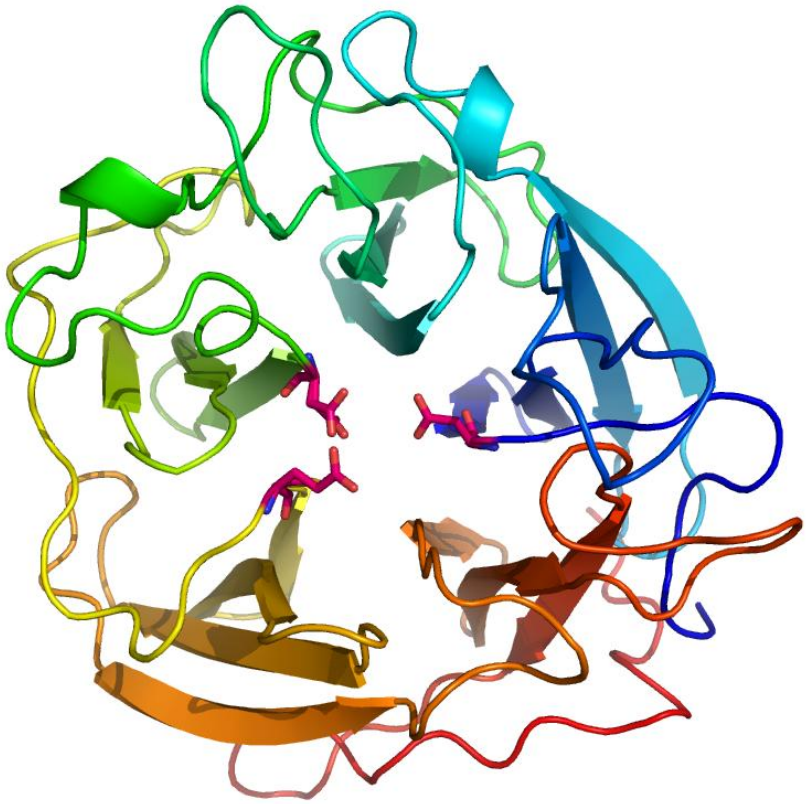


Figure 3.19 Cartoon representation of *Cj*Abf43A.

The structure of *Cj*Abf43A is coloured to show the N-terminus (blue) and C-terminus (red). A single domain, the β-propeller, is visible. The triad of catalytic residues (Asp21, Asp148 and Glu195) is shown in stick form, coloured red; these are located at the heart of the domain. Figure modelled on Nurizzo et al., 2002.

The pseudosymmetry of arabinofuranosides makes it difficult to confidently orient the ligand as oligosaccharide chains built in either direction fit equally well into density. The sole criterion used to delineate orientation in this case is the requirement for the O2-linked side chain to sit in the active site. The surface representation of *Cj*Abf43A, Figure 3.20, reveals a deep active site pocket in the centre of a highly curved cleft. The rim of the pocket abuts onto a shelf-like structure that accommodates the O3-linked arabinose side chain, demonstrating the structural basis for the enzyme's tolerance for α-1,2-L-arabinofuranose side chains in both single and double substitutions. The α-1,5-linked trisaccharide and, by inference, the arabinan backbone, lies in the central region of the curved surface which is thereby identified as the polysaccharide backbone binding cleft. This pocket contains three acidic

residues, Asp21, Asp148 and Glu195, which were identified by sequence alignments and structural analysis as the triad of catalytic residues found in GH43 enzymes.

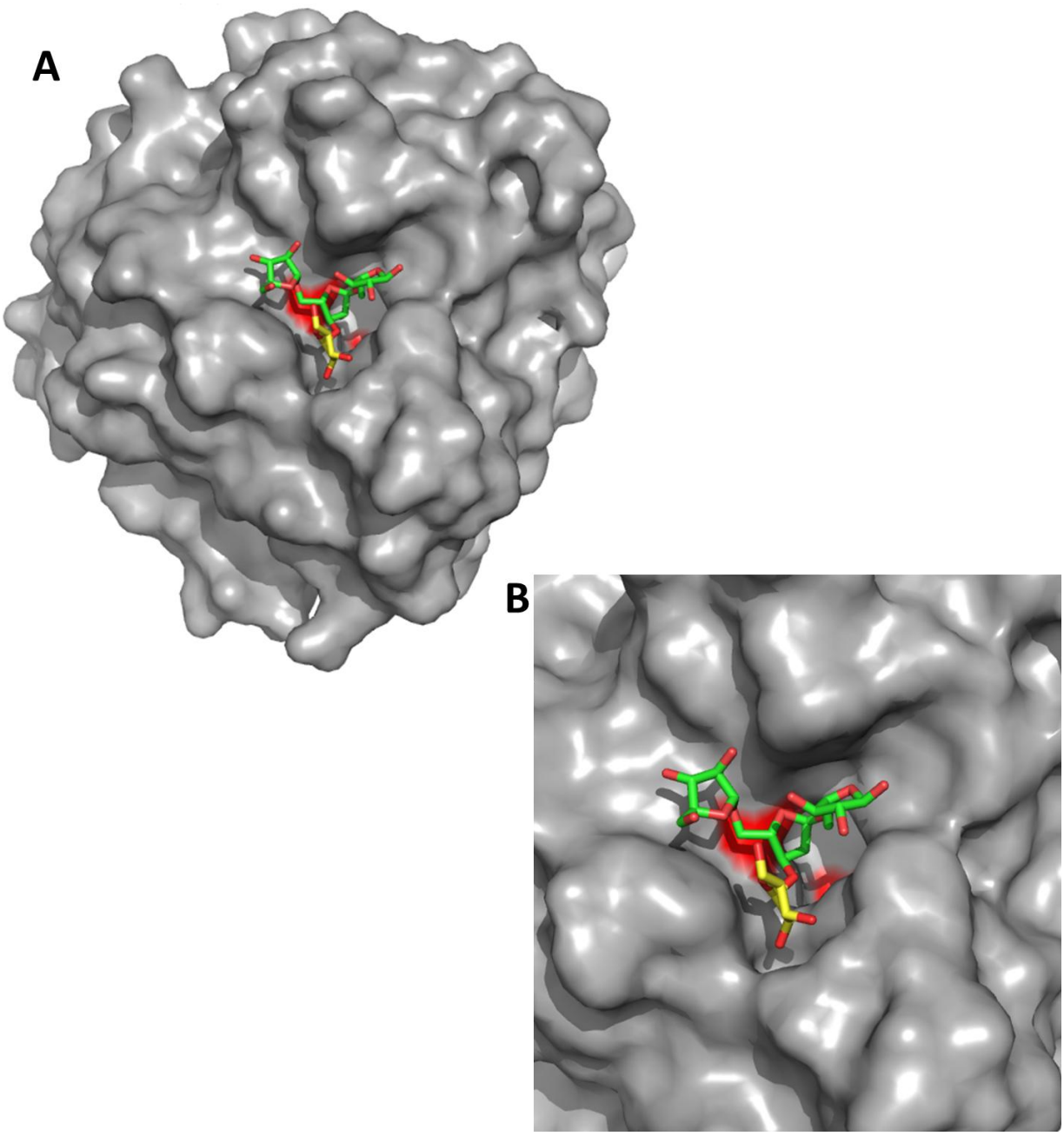


Figure 3.20 Surface representations of *Cj*Abf43A in complex with ligand.

A: surface representation of the enzyme in complex with ligand shows that the substrate is bound in the active site pocket, which houses the catalytic residues (shown in red). Also visible is the curved arabinan binding cleft, which forms a V-shape over the surface of the protein to accommodate the twisted conformation of the arabinan backbone (green). The O3-linked arabinose side chain which lies in the pocket adjacent to the active site is shown in yellow. Oxygen atoms within the carbohydrate are yellow.

B: a closer look at the active site of the ligand-bound complex shows that the α-1,5 backbone (green arabinotriose in this ligand) lies within the cleft while an O3-linked arabinose side chain (yellow) points into the depression adjacent to the active site pocket. The orientation of the central backbone sugar shows how an O2-linked side chain will point directly into the active site pocket, marked out by the red coloured catalytic residues. The ligand is coloured as for panel A.

The subsite topology of *Cj*Abf43A is defined as follows: the scissile bond is between the O2-linked arabinose decoration at -1 and the backbone arabinose at +1, using the nomenclature system developed by Davies and colleagues (Davies et al., 1997). Subsites extending towards the reducing end of the arabinan backbone (from the +1 subsite) are defined as +2R, +3R etc., while subsites extending to the non-reducing end of the polymer are designated +2NR, +3NR and so forth. The subsite accommodating the arabinose linked O3 to the +1 subsite is defined as the +2NR* subsite. Thus, the bound ligand occupies subsites +2NR, +2NR*, +1 and +2R (Figure 3.20). Based on the topology of the enzyme, the substrate binding cleft is unlikely to extend distally to the +2R subsite, although the enzyme is likely to contain at least one additional non-reducing subsite (+3NR).

An analysis of the ligand bound structure shows that the enzyme makes several contacts with the substrate. The +2R arabinose is positioned between Phe46 and Trp144. The indole nitrogen of this Tryptophan residue forms a hydrogen bond with the +2R arabinose, and mutagenesis studies have shown that this contact is important for catalysis. The position of Trp144 in the polysaccharide binding cleft suggests that it is important for substrate backbone binding. Phe214 makes hydrophobic contacts with C5 of the sugars at +2NR and +2NR*. In addition, Asn165 makes several important contacts with the substrate. Oδ1 of Asn165 hydrogen bonds with +2NR sugar, while Nδ2 makes polar contacts with the endocyclic ring oxygen of the +1 arabinose and with the glycosidic oxygen linking the +1 and +2NR sugars (Figure 3.21). If the ligand is flipped so that the O3 linked side chain points into the active site, this residue is unable to make a hydrogen bond with the glycosidic oxygen between the sugars at +2NR and +1. This indicates that Asn165 is particularly important for orienting the substrate to bring the O2-linked arabinose into the active site, consistent with data from mutagenesis studies performed by Dr A Cartmell which showed that the mutant N165A has very low activity against sugar beet arabinan but retains near wildtype activity against 4NP-α-L-arabinofuranoside.

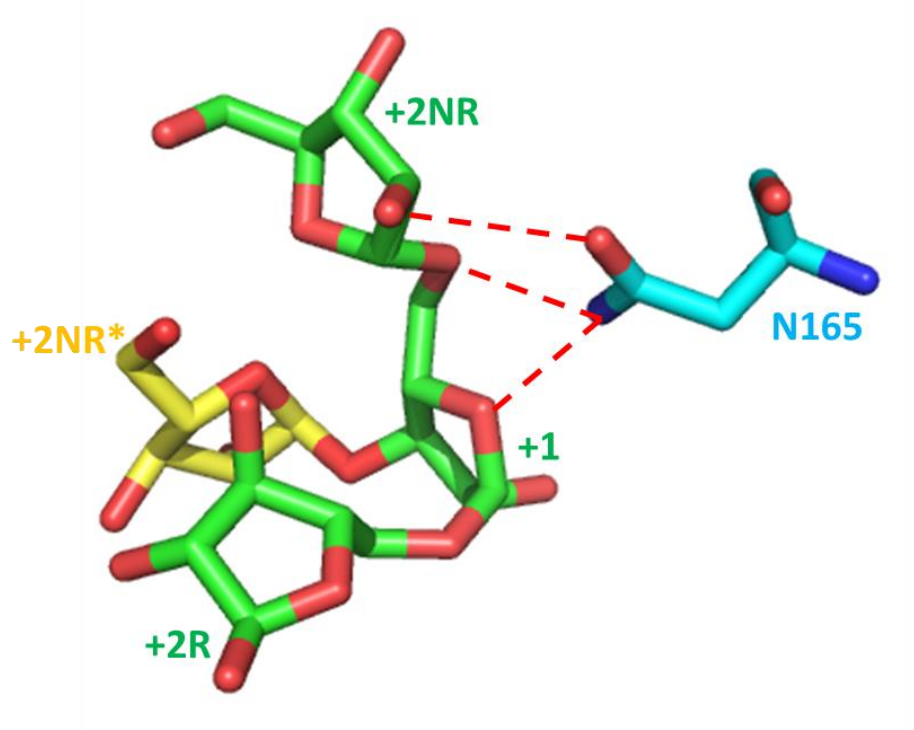


Figure 3.21 Asn165 makes critical substrate contacts for specificity.

In *Cj*Abf43A specificity is determined by ligand contacts made by Asn165 with the endocyclic ring oxygen of the +1 arabinose, and with the glycosidic oxygen connecting arabinose residues +1 and +2NR.

Due to their high sequence identity and shared specificity for α-1,2 linked arabinose side chains in sugar beet arabinan, it was expected that Bt0369 and *Cj*Abf43A would also show great structural similarity. Attempts to crystallise Bt0369 were not successful, but the protein sequence was submitted to the Swiss-Model workspace at ExPasy (Arnold et al., 2006; Schwede et al., 2003) and a model structure of Bt0369 was obtained by threading onto the crystal structure of *Cj*Abf43A. Automated validation of the homology model was undertaken by ANOLEA (Atomic Non-Local Environment Assessment), which is a tool used to perform energy calculations on protein chains and assess structure models (Melo and Feytmans, 1998). As Figure 3.22 shows the ANOLEA of the Bt0369 model was very favourable, due to the high similarity between the target and homology model (66 % identity). Figure 3.23 shows that the model comprises a single domain that displays the five-bladed β-propeller common to all GH43 enzymes.

Validation of the model was also undertaken by analysis of the structure in PyMol and comparison with the *Cj*Abf43A crystal structure. The residues Asp42, Asp169 and Glu215 were identified as the catalytic apparatus in Bt0369 by sequence alignment, and the homology model structure places these residues within a pocket topology which lies within the polysaccharide binding cleft (Figure 3.24). This is the active site of the enzyme, and is adjacent to a second depression, which accommodates the second arabinose (O3-linked) of the double substitution. These structural features correlate very well with those observed in the crystal structure of *Cj*Abf43A, indicating that the two enzymes employ highly similar strategies for substrate binding and catalysis.

A comparison of the crystal structure of *Cj*Abf43A with the homology model of Bt0369 reveals that the nature of specificity in both enzymes is determined in much the same way (Figure 3.24). The critical Asn165 in the *C. japonicus* enzyme overlays exactly with Asn186 in the *B. thetaiotaomicron* model. Similarly, Phe214 and Trp144 in *Cj*Abf43A have counterparts in Bt0369 (Phe234 and Trp165, respectively).

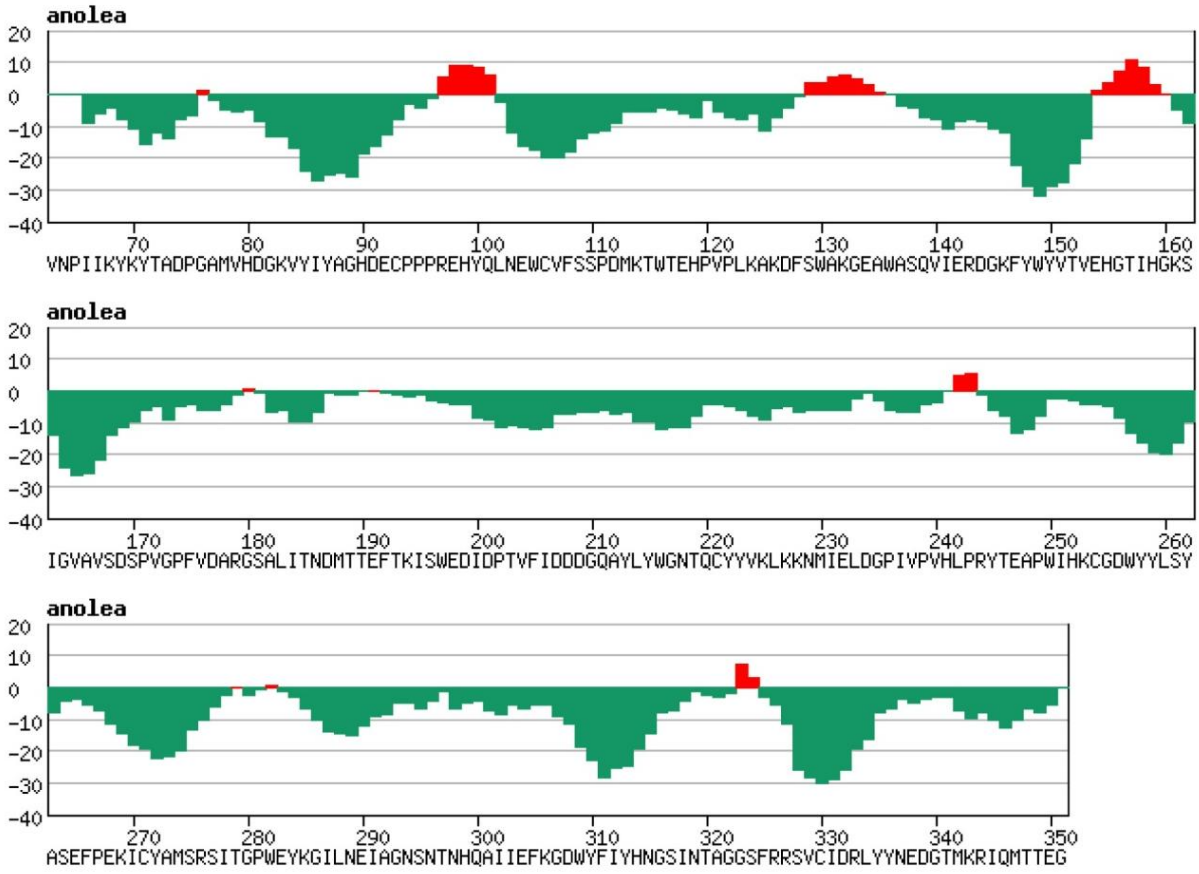


Figure 3.22 ANOLEA energy assessment of structural model of Bt0369.

Automated assessment of the homology model structure of Bt0369, generated by threading the Bt0369 sequence onto the crystal structure of *Cj*Abf43A is very favourable. Green areas indicate areas of sequence where the predicted structure has favourable calculated free energy, while red areas have unfavourable energy calculations.

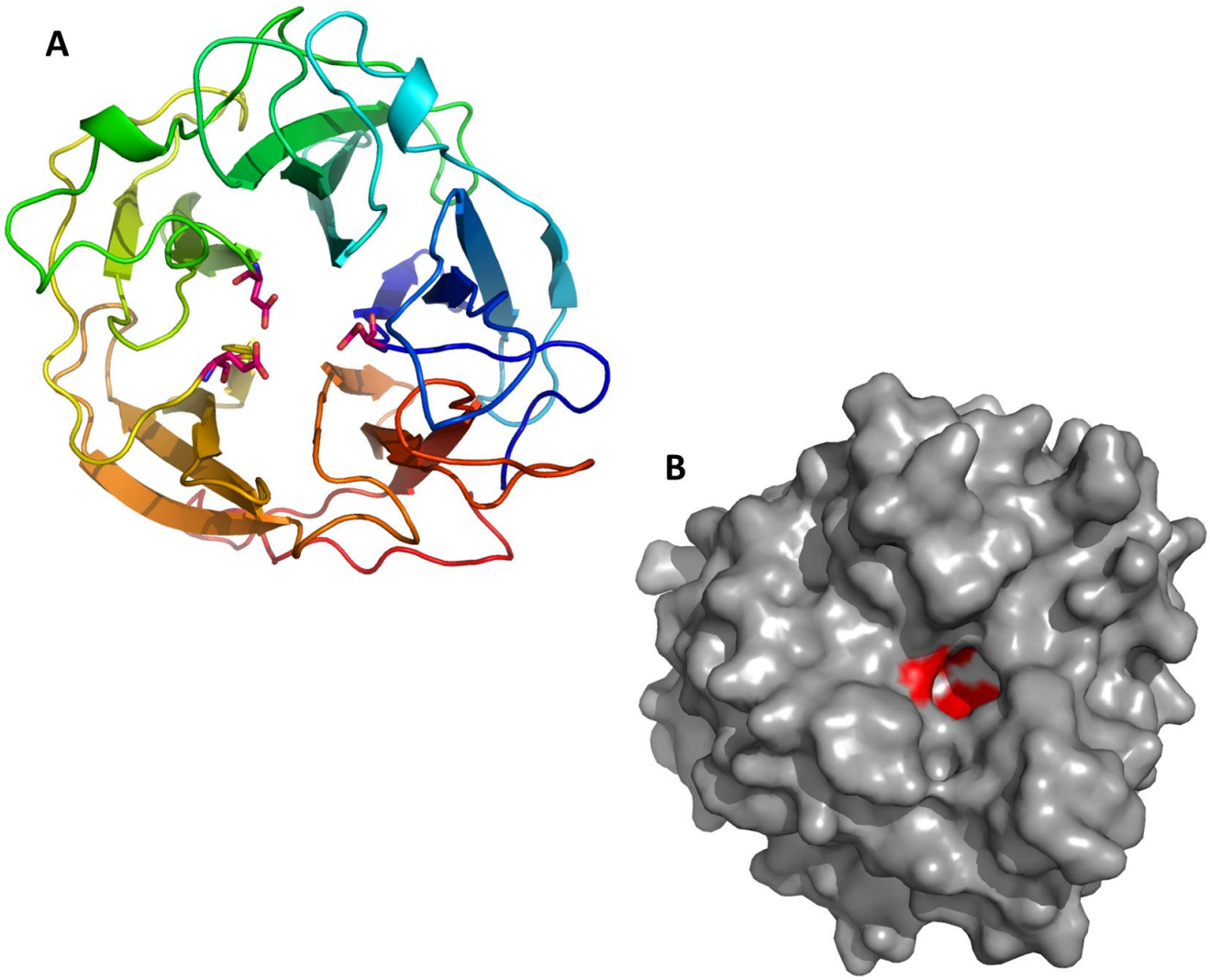


Figure 3.23 Homology model structure of Bt0369, generated by threading onto the crystal structure of *Cj*Abf43A.

A: cartoon representation of the Bt0369 model shows a single domain, the canonical GH43 β-propeller. The three catalytic residues are shown in stick form and coloured red at the heart of the propeller.

B: surface representation of the model of the enzyme shows a V-shaped cleft over the surface of the protein, centring on a pocket structure which contains the catalytic residues (red). This pocket is adjacent to a second pocket structure, as is seen in the crystal structure of *Cj*Abf43A.

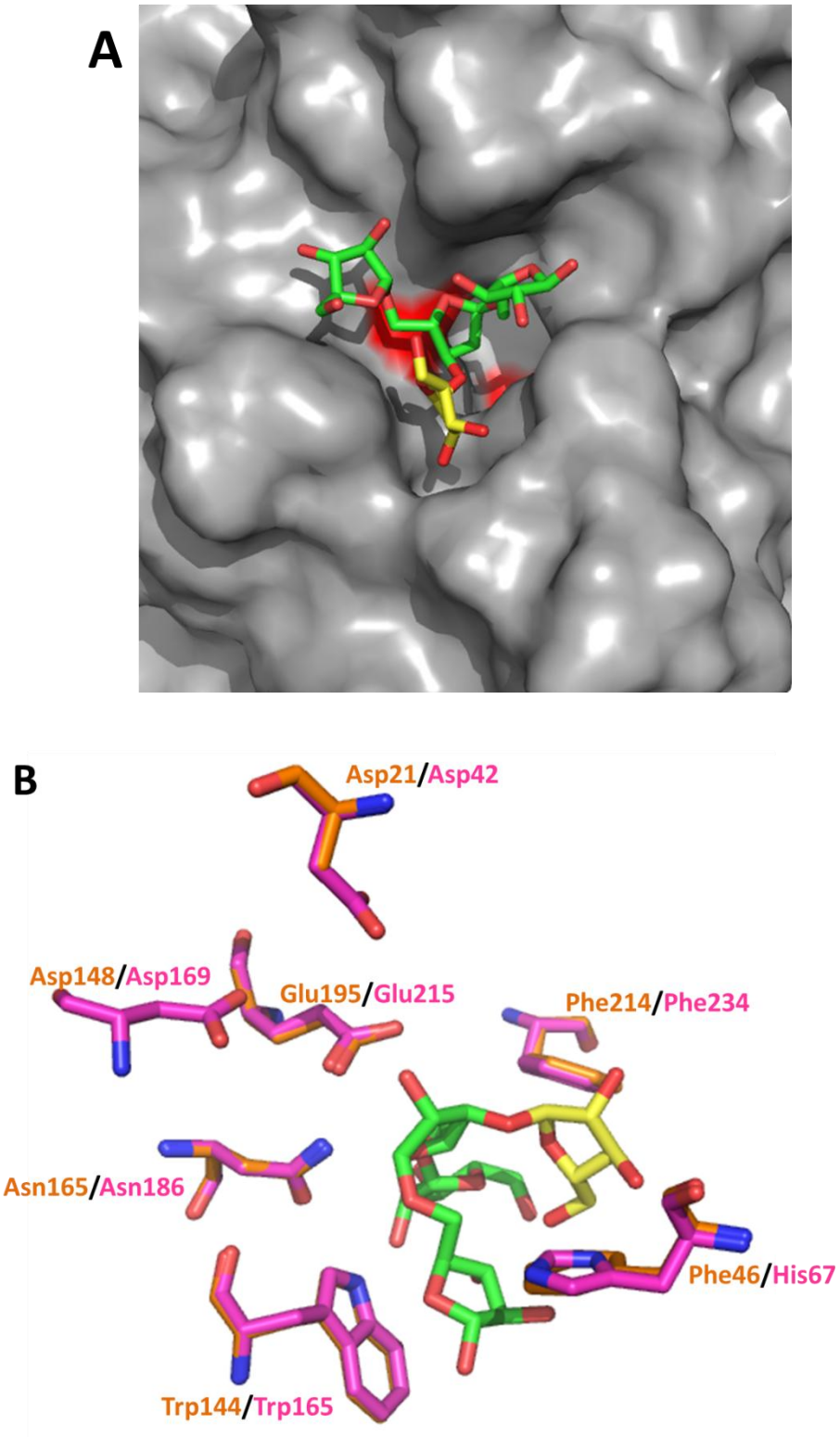


Figure 3.24 Comparing the crystal structure of *Cj*Abf43A with the homology model of Bt0369.

A: surface representation of the active site of the Bt0369 homology model overlaid with the branched ligand from the *Cj*Abf43A crystal structure. As seen for the *C. japonicus* enzyme (Figure 3.20), the

backbone of the ligand lies in the V-shaped cleft while the O3-linked side chain (yellow) points into the pocket adjacent to the active site, which is marked out by the red catalytic residues.

B: an overlay of the key residues coordinating substrate binding and specificity shows good alignment between the crystal structure of *Cj*Abf43A and the homology model of Bt0369. Only the non-critical Phe46 is not exactly conserved.

Generally, there is a paucity of information on enzymes that hydrolyse the double substitution in arabinan. The double substitution is inhibitory to most α-L-arabinofuranosidases, which are specific for singly substituted arabinose decorations (Beylot et al., 2001b). Thus, the hydrolysis of one of the linkages in these double substitutions must be a key early feature of arabinan metabolism, in order to create the single substitution side-chains which are accessible to other *exo*-acting enzymes. Removal of these side-chains may increase substrate access for *endo*-acting enzymes. Indeed, the only other published enzyme known to hydrolyse such structures is the GH43 arabinofuranosidase *Ba*AXHd3 from the colonic bacterium *Bifidobacterium adolescentis* (Lambertus et al., 2005); a highly similar enzyme is explored in greater detail in Chapter Four of this thesis. AXHd3, however, displays broader specificity than Bt3069 or *Cj*Abf43A as it hydrolyses double substitutions in both arabinoxylan and arabinan, although it will not hydrolyse singly substituted arabinose. Such an activity is not required by *B. thetaiotaomicron* as the bacterium does not metabolise xylan. By contrast *C. japonicus* has an extensive xylan degrading apparatus (Pell et al., 2004b), and thus likely contains an additional arabinofuranosidase that targets xylose residues decorated at O2 and O3.

While the utility of the specificity of Bt0369 and *Cj*Abf43A in removing the hindrance of the double substitution is clear, it is intriguing that the enzyme discriminates between the two linkages in the structure, and specifically that it removes the O2-linked component, rather than the O3. Single α-1,3 arabinose side chains are common in arabinan, while single α-1,2 side chains are very rare (Swamy and Salimath, 1991). An enzyme which could remove O3-linked arabinose side chains from both single and double substitutions would release much more arabinose from the polysaccharide than the enzyme presented here. The lack of functionally significant contacts with the +2NR* arabinose means that the double substitution is not a requirement for hydrolysis, and an enzyme which targeted O3 side chains may

be diverted from removal of the double substitutions by an abundance of the single linkages. The evolutionary rationale behind the selection of the O2 may be that it limits cleavage of single substitutions, maintaining a catalytic focus on the double substitution. It should be emphasised, however, that until the specificity of the two PUL 7 GH51 enzymes is established, the biological significance of the activity displayed by Bt0369 remains uncertain.

The crystal structure of *Cj*Abf43A, the *C. japonicus* homologue of Bt0369, reveals the nature of specificity for O2-linked arabinose side chains and also gives insight into why the enzyme is tailored exclusively for de-branching activity on arabinan and not arabinoxylan, unlike AXHd3 enzymes. As shown in Figure 3.25, the specificity of *Cj*Abf43A for sugar beet arabinan is due to the surface topography of the enzyme. Phe46, Trp144 and Phe214 contribute to the formation of a curved surface cleft around the -1 pocket. This topology is complementary to the extended helical structure of the α-1,5-L-arabinofuranose backbone of sugar beet arabinan explaining why the enzyme targets this polysaccharide. The enzyme is unable to hydrolyse arabinoxylan, which also contains α-1,2-L-arabinofuranose side chains, as the xylose backbone, which has a 3 fold screw axis, would make steric clashes with the curved surface of the substrate binding cleft. Conversely, the GH43 arabinoxylan-specific arabinofuranosidase, *Bs*AXHm2,3 (Vandermarliere et al., 2009), which cleaves only singly substituted side chains, is unable to hydrolyse arabinan, despite structural conservation with *Cj*Abf43A at the -1 subsite, as the arabinan backbone cannot be accommodated by the linear substrate binding cleft that houses the xylan backbone, Figure 3.25. Thus, it is the architecture of the surface substrate binding cleft, curved in *Cj*Abf43A and linear in *Bs*AXHm2,3, that dictates the specificity of these enzymes. Aside from Phe214, residues that line the substrate binding cleft of *Cj*Abf43A distal to the +1 subsite (which plays a key role in providing the binding energy required for substrate distortion in the active site) bind weakly to the arabinan backbone. This is consistent with the requirement for the polysaccharide backbone to dissociate prior to the release of arabinose from the active site pocket.

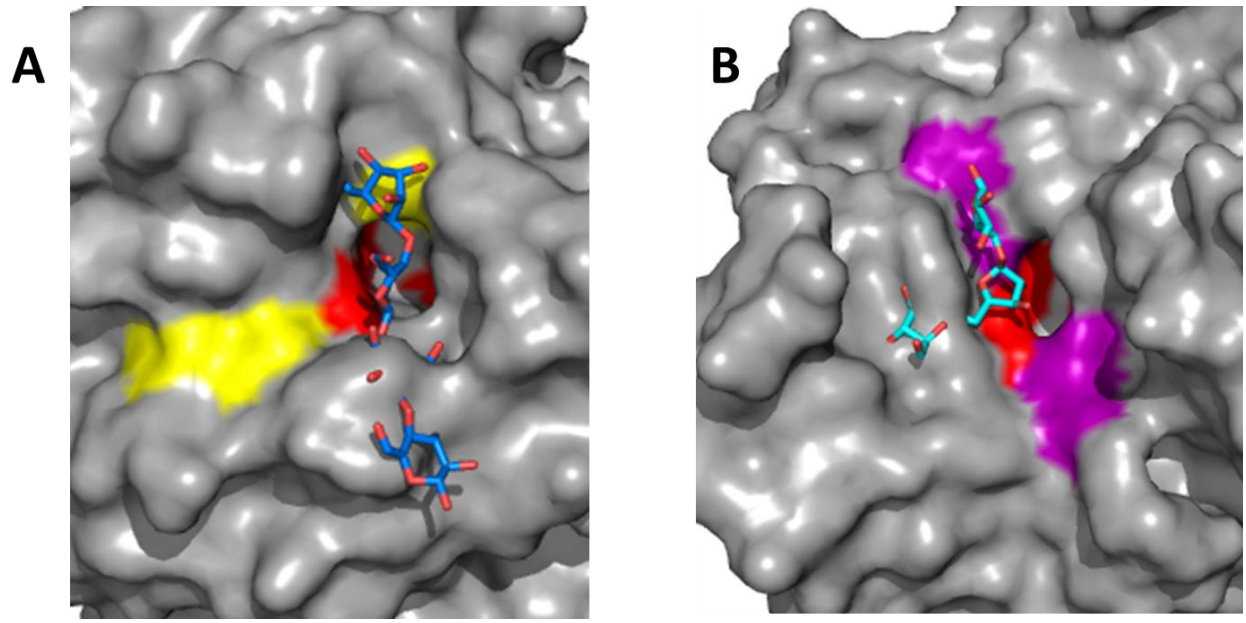


Figure 3.25 Comparison of crystal structures of *Cj*Abf43A and *Bs*AXH.

A: an overlay of *Cj*Abf43A with xylotetraose from the *Bs*AXH structure shows xylan is occluded from the active site by a major clash with the surface of the enzyme. The ligand is shown in blue while the catalytic residues of *C. japonicus* enzyme are highlighted in red. The arabinan-binding cleft proceeds in the direction indicated by the yellow shading.

B: likewise, an overlay of *Bs*AXH with arabinotriose from a *Cj*Abf43A structure shows how arabinan is excluded from the linear substrate binding cleft of this enzyme. The ligand is shown in cyan while the catalytic residues are highlighted red. The xylan-binding cleft of the *B. subtilis* enzyme is coloured purple.

3.6.1.iii Cooperativity of GH43s in arabinan metabolism

Based on an analysis of signal peptides, the *endo*-arabinanases of PUL 7 are predicted to be outer membrane-bound proteins; this is consistent with their role in the initial hydrolysis of polysaccharide arabinan. Prediction tools are unable to give a confident assessment of the cellular location of the arabinofuranosidase Bt0369, but an analysis of the enzyme activities described for the enzymes of PUL 7 indicates that it may function in the periplasm. A putative metabolic pathway for PUL 7 is shown in Figure 3.26 and proceeds as follows.

The *endo*-arabinanases Bt0360 and Bt0367 degrade arabinan to arabinooligosaccharides that are either linear or contain arabinose substitutions. The resulting oligosaccharides, with presumably varying branching patterns, are then

taken up into the cell by two membrane-bound protein complexes comprising homologues of SusC and SusD. It is quite unusual for a single PUL to include two SusC/D pairs. This may reflect a need to accommodate a diverse range of products generated by the outer membrane arabinanases.

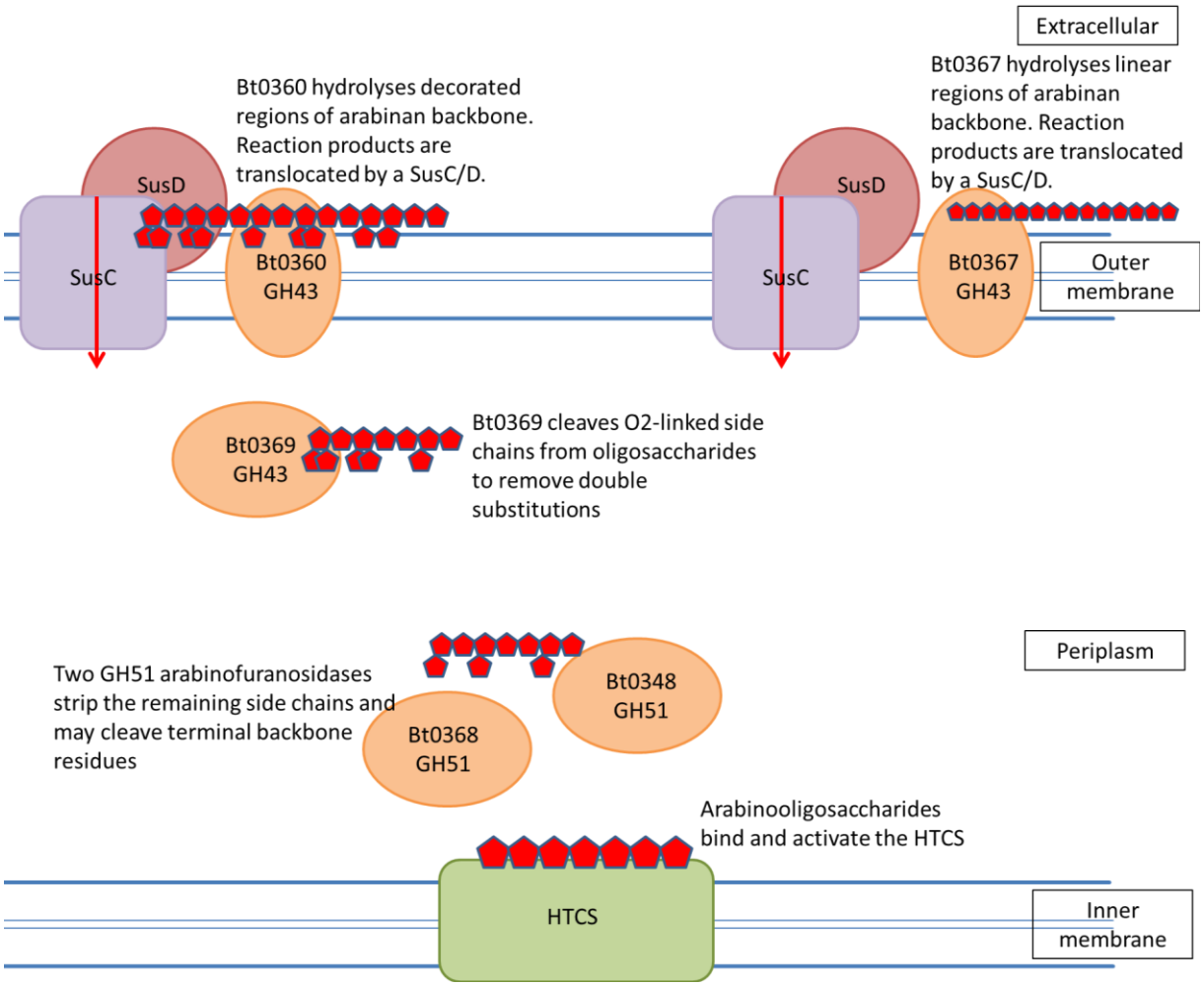


Figure 3.26 Putative cellular pathway for the components of PUL 7.

Schematic shows the putative cellular organisation of the components of the arabinan-metabolising PUL 7. The GH43s Bt0360 and Bt0367 are membrane-bound *endo*-arabinanases. GH43 Bt0369 is an α-1,2-L-arabinofuranosidase and is thought to be periplasmic. Bt0348 and Bt0368 are GH51 α-L-arabinofuranosidases of unknown specificity; they are located in the periplasm. One hybrid two component system (HTCS) resides within the inner membrane. Two SusC-D complexes are present in the outer membrane.

The α-1,2-L-arabinofuranosidase Bt0369 represents a ‘pre-treatment’ stage of arabinan processing, converting all the doubly substituted arabinose into single O3 decorations, as well as removing all single O2 arabinose side chains, so that branched arabinooligosaccharides can be stripped of side chains prior to binding to the activating HTCS. In the periplasm are two GH51 enzymes encoded by PUL 7, which are predicted to be α-L-arabinofuranosidases. At least one of the two periplasmic GH51 arabinofuranosidases is likely to remove the single O3-arabinose decorations to generate linear oligomers, a feature typical of enzymes in this family (Beylot et al., 2001a). It is possible that the other GH51 enzyme converts the linear oligosaccharides into arabinose, although no GH51 enzyme has been reported to display α-1,5-*exo*-arabinanase activity. Together, the three arabinofuranosidases generate linear arabinooligosaccharides which bind and activate the HTCS, leading to high-level expression of PUL 7 components. It is likely the outer membrane arabinanases and Sus binding proteins are expressed at a basal level, to generate the substrate required to switch on the HTCS, which upregulates expression of all PUL components in a positive feedback network designed to optimise arabinan utilisation.

This chapter reports on the capacity of the large number of GH43 enzymes produced by the colonic human symbiont *B. thetaiotaomicron* to hydrolyse arabinose-containing polysaccharides. A sister project with similar aims led to very similar findings, showing that the soil saprophyte *C. japonicus* has many GH43 enzymes with a weak xylanase activity and some with well-defined activities against arabinan (McKee et al 2011). While *C. japonicus* does not possess the organisational PUL system, genomic locations of many of the GH43 genes, in concert with newly described activity data have provided insight into carbohydrate-based metabolic pathways, most notably the arabinan degradative pathway.

In *C. japonicus* the α-1,2-L-arabinofuranosidase *Cj*Abf43A, a homologue of Bt0369, converts the double substitutions into single O3 linked decorations, which are then removed by an extracellular, membrane associated GH51 arabinofuranosidase (Beylot et al., 2001a). The linear arabinan generated is then hydrolysed exclusively to arabinotriose by an *endo*-processive arabinanase (McKie et al., 1997). The trisaccharide is metabolised, likely in the periplasm, by a GH43 α-1,5-*exo*-

arabinanase that displays a moderate preference for arabinotriose. The arabinobiose generated is transported into the cytoplasm, where it is hydrolysed by a second GH43 α-1,5-*exo*-arabinanase that appears to display a preference for the disaccharide.

As described above and illustrated in Figure 3.26, a different pathway is followed by the proteins of PUL 7 in *B. thetaiotaomicron*. The membrane-bound *endo*-arabinanases Bt0360 and Bt0367 respectively cleave those regions of the polysaccharide backbone which are relatively more and less decorated. The α-1,2-L-arabinofuranosidase Bt0369, which is predicted to be a periplasmic enzyme, then works in concert with two GH51 enzymes to remove side-chains and generate linear arabinooligosaccharides.

As discussed previously, the double substitution found in arabinan is inhibitory to most α-L-arabinofuranosidases which are specific for singly substituted arabinose decorations. Thus, the hydrolysis of one of the linkages in these double substitutions is an early feature of arabinan metabolism which increases substrate access for downstream enzymes. Indeed, it is interesting that while *C. japonicus* and *B. thetaiotaomicron* have adopted different strategies to hydrolyse the arabinan backbone, the two bacteria utilise the same debranching activity to cleave double substitutions, a critical early stage of the degradative hierarchy.

3.6.2 Expansion of glycoside hydrolase family 43

In total, soluble protein was obtained for twenty-five of the thirty-one GH43 enzymes encoded by the *B. thetaiotaomicron* genome. Each was subjected to activity screens, the results of which are summarised in Table 3.4. A sister project obtained soluble protein for eleven of the thirteen GH43s expressed by *Cellvibrio japonicus*.

In summary, ten of the *B. thetaiotaomicron* GH43s expressed in a soluble form in *E. coli* showed no measurable activity against any of the polysaccharide or aryl glycosides tested. Eleven showed a weak *endo*-xylanase activity against xylan polysaccharides with a variety of arabinofuranosyl branching patterns (wheat arabinoxylan, rye arabinoxylan, oat spelt xylan and birchwood xylan). Three of these enzymes (Bt2852, Bt3094 and Bt3655) also showed activity against 4NP-α-L-

arabinofuranoside and arabinose appears to be a significant product in arabinoxylan degradation by these GH43s (Figures 3.10 and 3.11). Two of the GH43s (Bt0360 and Bt0367) were shown to be *endo*-arabinanases while Bt0369, and a homologue from *C. japonicus*, display a novel α-1,2-L-arabinofuranosidase activity.

3.6.2.i Phylogenetic analysis of GH43s

The *B. thetaiotaomicron* and *C. japonicus* GH43 proteins were subjected to phylogenetic analysis. An alignment of all of the sequences was performed using ClustalW. The analysis incorporated selected enzymes from other organisms with known activities. These were:

- *Hi*AXHd3 and *Ba*AXHd3, which cleave α1,3 linked arabinose residues from double substitutions in arabinan and arabinoxylan (Lambertus et al., 2005)
- *Gs*XynB3, a β-xylosidase (Brux et al., 2006)
- *Bs*AXHm2,3, an arabinoxylan-specific arabinofuranosidase which cleaves the single substitution (Vandermarliere et al., 2009)
- SAVAraF43A, an *exo* α-1,5-L-arabinofuranosidase (Fujimoto et al., 2010)
- *Bs*Arb43A (Proctor et al., 2005) and *Gs*AbnB (Alhasid et al., 2009), *endo*-acting arabinanases
- *Cj*Arb43A, an *endo* processive-arabinanase (Nurizzo et al., 2002)
- *Pp*Xyn43A (Gosalbes et al., 1991) and UXOrf66 (Zhao et al., 2010), which display xylanase activity.

Most of these published enzymes are very well characterised, many with crystal structures in complex with one or more ligands. An exception is *Pp*Xyn43A, until recently the only published GH43 xylanase (Gosalbes et al., 1991). This activity was demonstrated by zymography, a highly sensitive but non-quantitative electrophoretic technique for detecting enzyme activity which utilises a modified version of the protocol for SDS-PAGE, incorporating substrate in the polyacrylamide gel. The zymogram showed a very faint halo, indicating weak xylanase activity. The enzyme was also shown to have activity against 4NP-α-L-arabinofuranoside. The xylanase UXOrf66, from an uncultured bacterium, was identified by metagenome screening (Zhao et al., 2010). Evidence for xylanase activity in this case is more compelling, as

it is derived from comprehensive analysis by reducing sugar (DNSA) assay. The enzyme also shows lesser activity on 4NP-α-L-arabinofuranoside and carboxymethyl-cellulose.

The phylogeny, presented in Figure 3.27, shows that newly described GH43s displaying *endo*-arabinanase and α-1,5-*exo*-arabinanase activity are in clades with previously described enzymes that exhibit similar activities. By contrast, and consistent with their novel activity, the well characterised α-L-1,2-arabinofuranosidases *Cj*Abf43A and Bt0369 formed a clade that contains no other enzymes with known catalytic properties. As one particular amino acid (Asn165 in *Cj*Abf43A and Asn186 in Bt0369) has proven to be a critical determinant of this specificity, as discussed above, it should be possible to identify other enzymes with this specificity by the presence of this residue. Indeed, a Clustal alignment of the top 30 homologues of Bt0369, as identified by Blast search, reveals that all possess this Asparagine residue. It is likely that at least some of these enzymes share the same specificity for O2-linked arabinose side chains in arabinan.

As *B. thetaiotaomicron* is unable to grow on xylan (Cooper et al., 1985; Tannock, 1977), and the genome of the bacterium does not encode GH10 or GH11 xylanases, it is interesting that so many GH43 enzymes with apparent xylanase action have been retained. It may be that for at least some of these enzymes, particularly where the xylan degradation is the only activity identified, that the correct substrate has simply not been identified in this study. In this case, the unidentified activity is presumed to be the primary function of the enzyme, with the xylanase action a weak side activity.

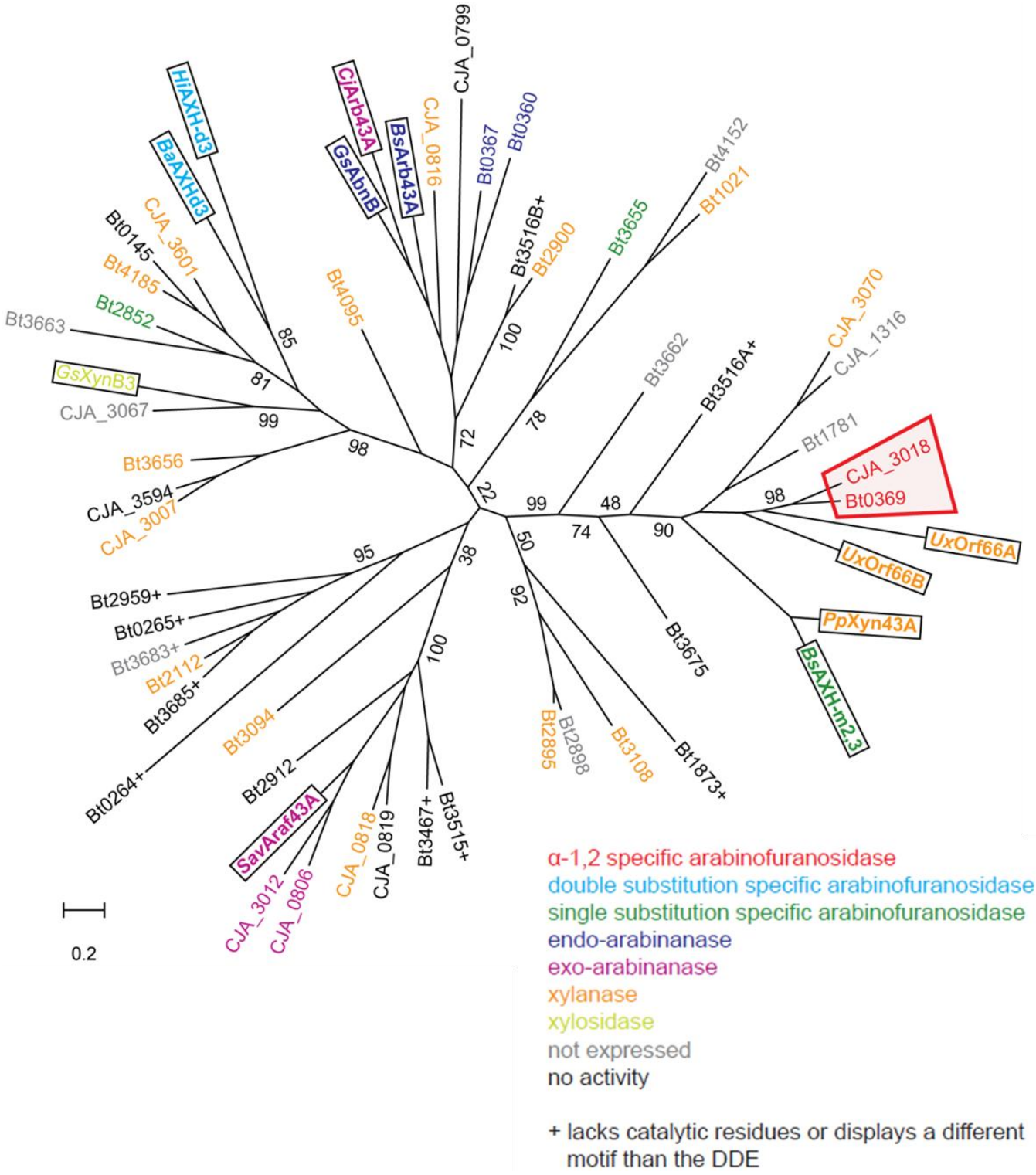


Figure 3.27 Phylogenetic tree of GH43 proteins.

Those enzymes highlighted with boxes, for example *Cj*Arb43A, are previously published examples of activities. Enzymes are colour coded according to activity, as shown in the figure. Enzymes described in this project largely cluster with previously characterised examples of the same activity. The novel α-1,2-L-arabinofuranosidases are highlighted with a red box; they are in a clade with no other enzymes.

Interestingly, the many enzymes displaying “trace” xylanase activity are not clustered into a specific region of the phylogenetic tree, suggesting that this minor activity may be a generic feature of GH43. Until recently, a GH43 from *Paenibacillus polymyxa* was the only published xylanase from the family. The evidence for this activity is rather weak, based on a faint band on zymogram, and may reflect the trace xylanase activity commonly seen in GH43, as a second look reveals that this study shows accord with the results presented in this chapter. The publication identifies *Pp*Xyn43A as a novel type of xylanase with some arabinofuranosidase activity. Despite the limitations of the methodological approach to this work, it does have resonance with dual activity GH43s identified in the work presented in this chapter, particularly the enzyme Bt2852 (Figures 3.10 and 3.11), which is capable of degradation of the xylan backbone and removal of arabinose side chains. It seems likely that *Pp*Xyn43A should be classified similarly to Bt2852, as an arabinofuranosidase with weak xylanase activity.

The apparently wide-spread weak xylanase activity in the family is interesting. One might speculate that an ancestral GH43 enzyme displayed significant xylanase activity, the capacity for which is retained in the family-wide β-propeller fold. Over time, cell wall degrading organisms acquired GH10 and/or GH11 xylanase enzymes (Gilbert, 2010; Gilbert et al., 2008; Henrissat et al., 1995; Tull and Withers, 1994; Wicki et al., 2007), which are much more efficient and operate via different substrate binding modes and catalytic mechanisms to the GH43s. Thus, as new specificities were introduced into the GH43 fold by the development of pocket and cleft topologies, the endogenous capacity to hydrolyse xylan in an efficient manner was lost. Bt2852 in particular provides an interesting example of this, as an enzyme which weakly hydrolyses the xylan backbone while displaying moderate arabinofuranosidase activity against artificial substrates (4NP-α-L-arabinofuranoside) and against xylan polysaccharides (Figures 3.10 and 3.11). Bt2852 is defined as an arabinoxylan-specific arabinofuranosidase (AXH) of unknown specificity for arabinose substitution with vestigial xylanase activity. As discussed earlier, the GH43 *Pp*Xyn43A is likely highly similar in action to Bt2852, although the nature of the arabinofuranosidase activity is unclear as it was only demonstrated on 4NP-α-L-arabinofuranoside, and not on any arabinose-containing polysaccharides. This

theory is explored again in later chapters, as a single mutation to the active site pocket of an arabinoxylan-specific arabinofuranosidase is shown to introduce *endo*-xylanase activity while retaining wild-type specificity.

A crystal structure of a weak xylanase (with no apparent additional activities), Bt2895 (PDB 3KST, unpublished), was deposited in November 2010. As Figure 3.28 shows, the protein comprises solely the canonical GH43 β-propeller domain, at the heart of which is a shallow cleft that houses an active site pocket containing the three catalytic residues. The shallow nature of this cleft, assumed to be the site of polysaccharide binding, may explain the weak activity of the enzyme. However, this apparent cleft does not overlay with xylotetraose taken from the crystal structure of *Bs*AXHm2,3; the ligand in fact lies perpendicular to the cleft. The unexpected orientation of the substrate binding cleft in Bt2895 may explain the very weak nature of the activity of this enzyme.

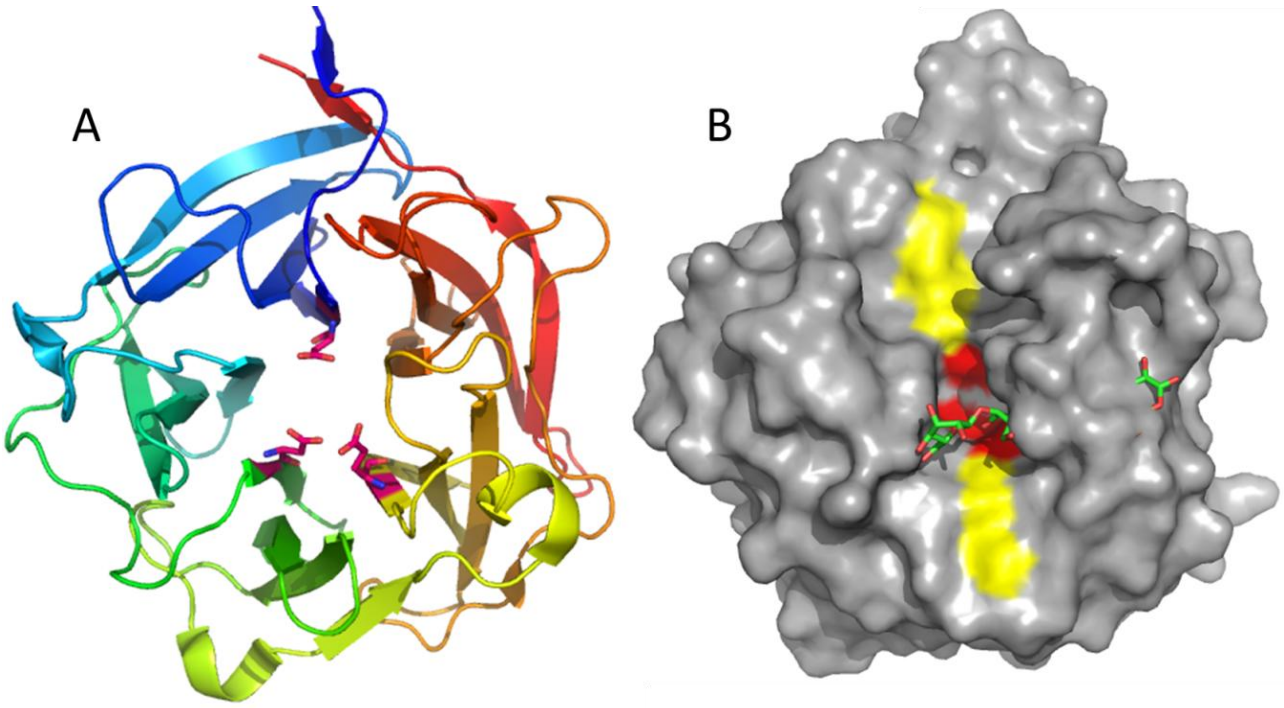


Figure 3.28 Crystal structure of Bt2895.

A: cartoon representation of Bt2895 ramped from blue (N-terminus) to red (C-terminus) with the catalytic residues shown in stick form and coloured red. The molecule does appear to be stabilised by 'molecular velcro', as blades from the first and fifth blade of the β-propeller interact closely.

B: surface representation of the protein with catalytic residues highlighted in red shows a shallow active site pocket. A surface cleft, highlighted in yellow, is also apparent. This would appear to be a polysaccharide binding cleft. However, when xylotetraose taken from the crystal structure of *Bs*AXHm2,3 is overlaid with Bt2895, a significant clash is observed.

3.6.2.ii Inactive enzymes

It should be noted that ten of the twenty-five recombinant proteins from *B. thetaiotaomicron* obtained in soluble form displayed no biologically significant enzyme activity in the screens described in this chapter. It is possible that, through redundancy, there was no requirement for the bacterium to retain functional forms of these enzymes. Alignment of these enzymes showed that six lack one or more of the catalytic residues (Bt3467, Bt3515, Bt1873, Bt0264, Bt2959 and Bt3685); this provides support for the view that they are not catalytically active. This view is further illustrated by the structure of Bt2959 (PDB 3NQH, unpublished, deposited June 2010), which shows that the protein lacks the canonical catalytic base. In addition, a long loop from the non-catalytic module in the β-propeller completely blocks the putative active site of Bt2959 (Figure 3.29). Analysis of this structure seems to confirm biochemical data to suggest that this is not an active enzyme. As Bt2959 is not found in a PUL in the *B. thetaiotaomicron* genome, it is unlikely that the protein has a biological role as a glycoside hydrolase, or that it is even expressed by *B. thetaiotaomicron* in nature. Why the gene persists in the genome is unclear. It seems likely that other GH43s lacking activity and catalytic residues may also not be biologically active. However, it is also possible that for those enzymes which do possess the expected catalytic residues, which display no apparent activity (Bt0145, Bt0265, Bt2912 and Bt3516), the appropriate substrate was simply not evaluated in this study.

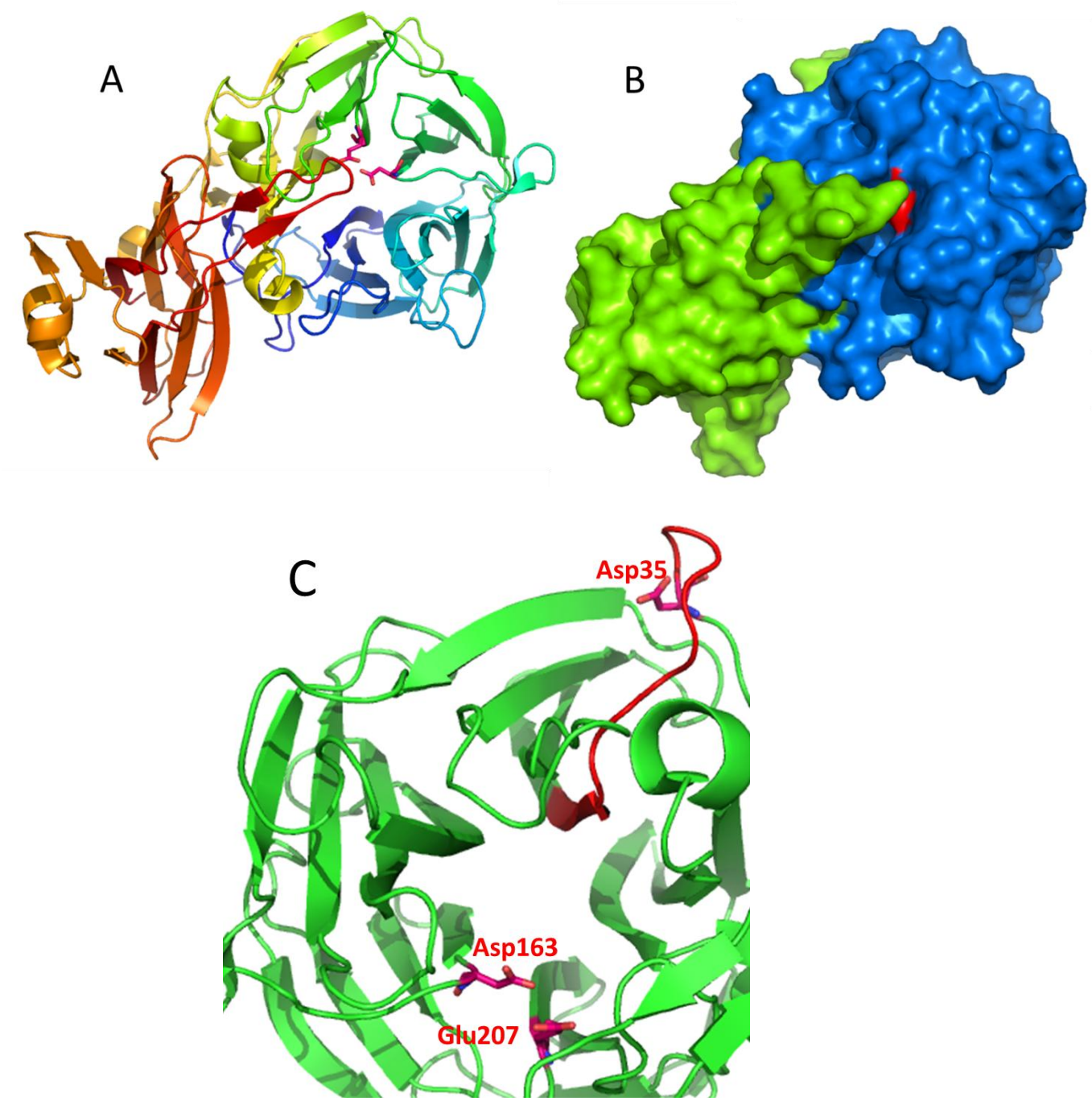


Figure 3.29 Crystal structure of Bt2959.

A: the cartoon representation shows that Bt2959 comprises the β-propeller module and a secondary β-sandwich type module. Amino acids shown in stick form (coloured red) are candidates for the general acid and pKa modulator; no general base is apparent in the active site.

B: a surface representation shows that the active site of the β-propeller domain (blue) is blocked by a loop projecting from the β-sandwich domain (green).

C: a cartoon representation of the β-propeller domain of Bt2959 shows how the catalytic base is pushed away from the active site by a loop, shown in red.

A closer look at the sequences of those GH43s which apparently lack the catalytic base by sequence alignment suggests that for some, the relevant Aspartate may be present but is in a sub-optimal location for catalysis due to the insertion of a ten amino acid loop (Figure 3.29). This loop is seen in five GH43s: Bt2959, Bt0265, Bt3683, Bt2112 and Bt3685, and is quite highly conserved, as all contain the amino acid sequence INAHG. All but Bt3683 were expressed in soluble form and assayed against polysaccharides. Only Bt2112 showed any activity, and this enzyme was one of the weakest xylanases. This would suggest that this loop disrupts the critical catalytic base, forcing it into a very unfavourable conformation, which has the effect of inactivating the enzyme.

The expansion of GH43 enzymes in microorganisms from very varied habitats points to a complex array of specificities within this family. Through the analysis of GH43s from *B. thetaiotaomicron* and *C. japonicus* some definable activities have been described, but a function could not be assigned for many of the proteins. This may partly reflect the loss in activity in these proteins due to functional redundancy. It is also possible, however, that the function of active enzymes could not be assigned as the substrates for these biocatalysts were not used, or the activities of these GH43s are only apparent when they are acting in synergy with other degradative enzymes. Despite the problems in assigning functions to all the GH43 proteins, the data presented here indicate that the main chain of arabinan, a target substrate for numerous GH43 enzymes, is degraded by an *endo* and *exo* mechanism in the gut symbiont and soil saprophyte, respectively. Although distinct, the mechanisms of arabinan degradation in the two bacteria display an element of convergence; both *B. thetaiotaomicron* and *C. japonicus* remove O2-linked arabinose decorations, in the context of single or double substitutions, through the action of an arabinan-specific □-1,2-arabinofuranosidase, an activity that has not previously been reported. The biological rationale for such an activity likely reflects the capacity of the enzyme to convert double substitutions into single decorations, which will then be accessible to the GH51 arabinofuranosidases, which are expressed by these bacteria. The crystal structure of *Cj*Abf43A reveals several topological features, such as a curved

substrate binding cleft, a shelf-like structure adjacent to the active site, and the targeting of the only asymmetric oxygen in arabinan, which confer the specificity displayed by the enzyme. The arabinan-specific □-1,2-arabinofuranosidase activity identified here will add to the toolbox of biocatalysts and probes required to deconstruct and understand the molecular architecture of plant cell walls.

CHAPTER FOUR

Structure and specificity of the *Humicola insolens* GH43 α-L-arabinofuranosidase, *Hi*AXHd3

The complexity and recalcitrance of the plant cell wall is demonstrated by the structure of xylan polysaccharides (Mohnen, 2008a), which are components of the hemicellulosic portion of the wall. Xylan consists of a β-1,4-xylopyranose backbone that is heavily decorated. L-arabinofuranose residues can decorate backbone xylose residues at one or both of the available hydroxyls (Viëtor et al., 1994). Such a complex substrate requires a consortium of enzymes for complete degradation, beginning with de-branching enzymes, particularly arabinofuranosidases, and culminating in cleavage of the backbone by *endo*-acting xylanases. A similar situation applies in the case of the pectic polysaccharide arabinan where, again, removal of the α-1,2- and α-1,3-linked arabinose side chains by arabinofuranosidases is required before *endo*-acting arabinanases can efficiently hydrolyse the arabinan backbone, which comprises α-1,5-L-arabinofuranose units (Gilbert, 2010).

The majority of arabinofuranosidases are found in glycoside hydrolase (GH) families GH43, GH51, GH54 and GH62 (Henrissat and Bairoch, 1996). The GH51 and GH54 enzymes display specificity for single arabinose decorations (i.e. O2 or O3 linked) in both arabinan and xylan (Beylot et al., 2001b). GH62 arabinofuranosidases, which are specific for arabinoxylans, also hydrolyse single O2 or O3 linked arabinose residues (Beylot et al., 2001b) (Kellett et al., 1990).

The GH43 arabinofuranosidases which have been evaluated against polysaccharide substrates (see Chapter Three) target single arabinose units (linked O2 or O3) in arabinoxylans (designated AXHm2,3 (Vandermarliere et al., 2009)), or, more rarely, release O3 linked arabinosyl residues that are components of double substitutions (xylose residues decorated at both O2 and O3 with arabinose*)*; these enzymes are defined as AXHd3 (Sorensen et al., 2006) .

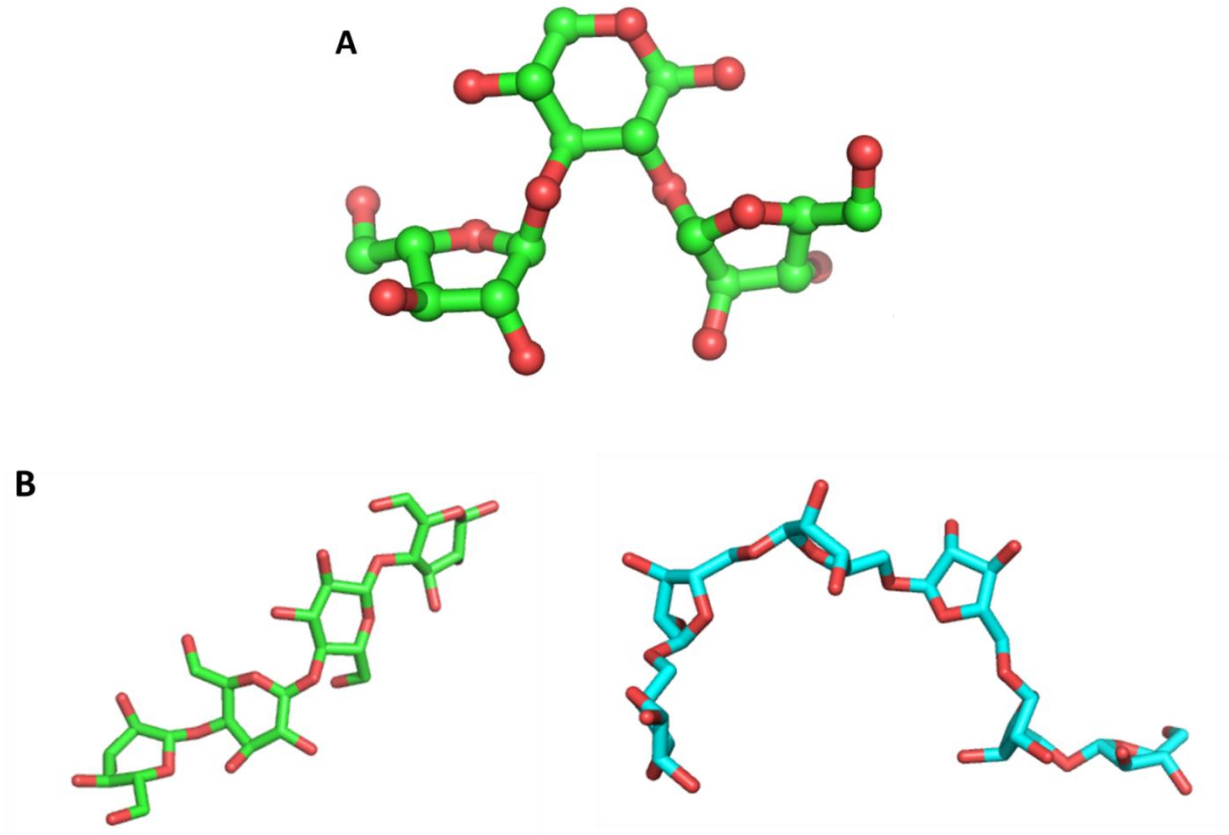


Figure 4.1 The pseudosymmetrical substrates of AXHd3 enzymes.

A: xylopyranosyl backbone residue of arabinoxylan, doubly substituted with two arabinofuranosyl residues at positions O2 and O3. This represents the 'core' substrate of AXHd3 enzymes. The α-1,2 and α-1,3 linkages are highly flexible, so the orientation of the arabinose residues is not fixed. Figure generated using the Glycam Biomolecule Builder (www.glycam.org; Kirschner et al., 2008).

B: the backbones of xylan (green) and arabinan (cyan) polysaccharides adopt very different structural conformations. Xylotetraose is taken from the crystal structure of *Bs*AXHm2,3 (PDB code 3C7E, (Vandermarliere et al., 2009)). Arabinohexaose is taken from the crystal structure of *Cj*Abf43A (PDB code 1GYD (Nurizzo et al., 2002)).

The double substitution remains a significant enzymatic challenge to efficient saccharification of plant cell wall polymers. Arabinofuranosidases able to remove single arabinose substitutions from xylan are numerous, while only two enzymes, both from GH43, are known to display AXHd3 activity. Both are highly specific and remove only the O3 linked arabinose from xylan. These enzymes are found in the gastrointestinal bacterium *Bifidobacterium adolescentis* (*Ba*AXHd3 (Lambertus et al., 2005, Laere et al., 1997)) and the thermophilic fungus *Humicola insolens* (*Hi*AXHd3

(Sorensen et al., 2006)). *Ba*AXHd3 has also been shown to release arabinose from arabinan (Lambertus et al., 2005), although it is unclear whether it retains specificity for double substitutions against this substrate. Sugar beet arabinan consists of an α-1,5 linked L-arabinofuranose backbone that is ~60 % monosubstituted with α-1,3-L-arabinofuranose side chains and, less frequently, with single α-1,2-L-arabinofuranose (Ragauskas et al., 2006). Some backbone residues are doubly substituted with both α-1,2 and α-1,3 side chains (Caffall and Mohnen, 2009). The core substrate for AXHd3 enzymes is essentially symmetrical (Figure 4.1), particularly so when considering the flexibility of the arabinose moieties, so the nature of the specificity of these enzymes is intriguing.

AXHd3 arabinofuranosidases are potentially of biotechnological significance as the hydrolysis of doubly substituted xylose residues is a critical first step in the deconstruction of arabinoxylans. Removal of the double substitution potentiates the action of other arabinofuranosidases, and complete de-branching of the polysaccharide allows for efficient backbone degradation by *endo*-acting enzymes. The clear utility of this unusual enzyme prompted a closer examination of the nature of specificity of *Hi*AXHd3 through structural analysis, site-directed mutagenesis and biochemical study. The primary aim of the project described in this chapter was to understand the molecular basis for selection for the O3-linked arabinose in the pseudosymmetrical double-substitution structure.

This chapter will present evidence for the specificity of *Hi*AXHd3 against arabinoxylan and arabinan, with kinetic analysis of both activities. Biochemical analysis proceeded via a similar to that described for the GH43 enzymes of *B. thetaiotaomicron* in the previous chapter, beginning with overnight incubation with polysaccharide analysed by TLC and HPLC. Kinetic analysis was then undertaken using the galactose dehydrogenase-linked assay, to measure release of arabinose. To investigate the structural basis for the unusual specificity displayed by AXHd3 enzymes, the crystal structure of *Hi*AXHd3 was determined with and without appropriate ligands. The nature of specificity was examined through mutagenesis studies, which also led to an exploration of possible mechanisms by which novel substrate specificities can be introduced into *Hi*AXHd3.

4. RESULTS

4.1 Wild type activity

The GH43 arabinofuranosidase *Hi*AXHd3 was previously shown to hydrolyse α-L-1,3-arabinofuranose residues from the backbone xylose units of arabinoxylan, which were substituted at both O2 and O3 (Sorensen et al., 2006). To explore the biochemical properties of the enzyme in more detail the gene encoding *Hi*AXHd3 was synthesised by Blue Heron Biotechnology Inc. (USA) and inserted into the *Escherichia coli* expression vector pET21(a) at the NcoI and XhoI restriction sites. In the original research on *Hi*AXHd3 (Sorensen et al., 2006) the protein was expressed without tags in *Aspergillus oryzae (Kauppinen et al., 1995)*. The *E. coli* derived *Hi*AXHd3 construct utilised here carried a C-terminal His_6-tag. The protein was purified to electrophoretic homogeneity by immobilised metal ion affinity chromatography (IMAC) (see Chapter Two, Section 2.2.21). Figure 4.2 shows a typical SDS-PAGE gel of the purification of the enzyme. The protein size was ~ 59 kDa, consistent with the predicted molecular weight, and was the only major peptide following purification.

An initial activity screen (Figure 4.3) indicated that *Hi*AXHd3 released arabinose as the sole product from both wheat arabinoxylan and sugar beet arabinan, as had been shown for the corresponding *Bifidobacter* AXHd3 (Lambertus et al., 2005). To explore the specificity of this activity, samples were prepared for 1D and 2D NMR. NMR spectra of polysaccharides are very complex due to the high number of constituent monosaccharides and linkage types. The spectrum for arabinoxylan, which consists of two different sugar components (arabinose and xylose) was complicated by the presence of reducing end xylose residues from contaminating oligosaccharides of various lengths. Peaks for these reducing ends masked other peaks corresponding to arabinose side chains. These oligosaccharides were therefore removed by ethanol precipitation prior to analysis. Following removal of the reducing end peaks, the spectrum for arabinoxylan was clear and simple to interpret. Conversely, arabinan generates spectra which are difficult to interpret, as many different peaks overlap. Thus, it was not possible to determine the bond(s) cleaved by *Hi*AXHd3 in arabinan as the NMR spectra of the polysaccharide were too

complex. This polymer was therefore subjected to partial acid hydrolysis (approximately 25 % of bonds were cleaved), and the oligosaccharides generated were enriched for substrates of *Hi*AXHd3 by ethanol precipitation and subsequent concentration by lyophilisation.

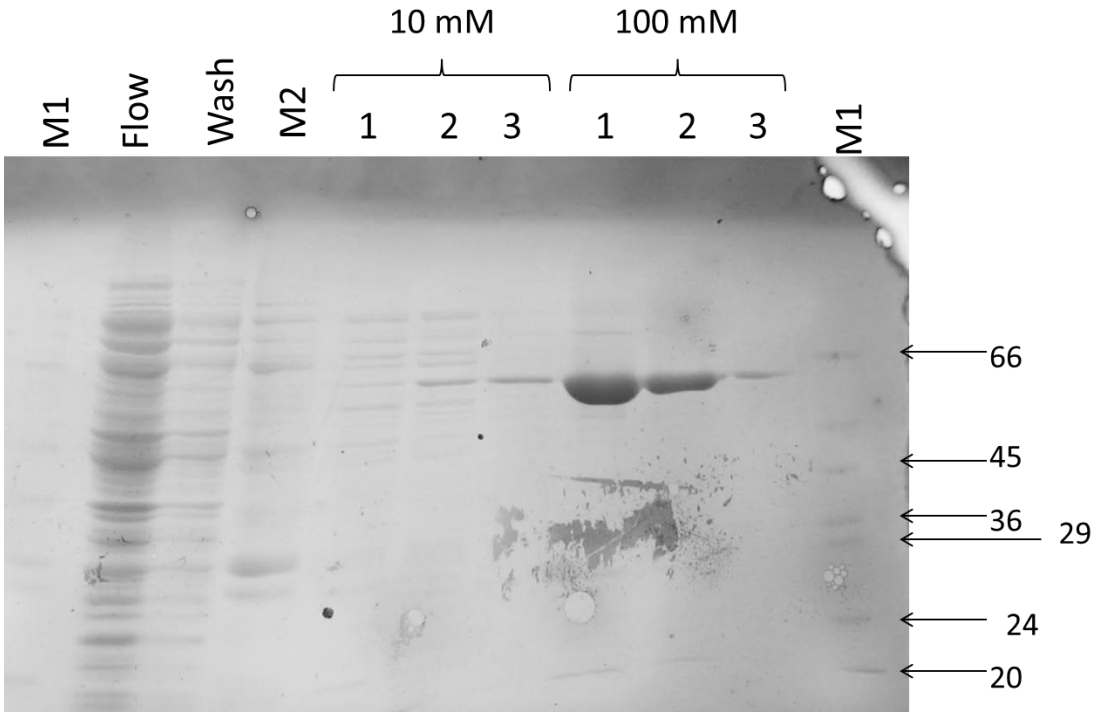


Figure 4.2 SDS-PAGE analysis of *Hi*AXHd3 expression and purification.

SDS-PAGE gel showing typical level of expression and elution profile of *Hi*AXHd3 expressed in *E. coli*. Protein was expressed in Tuner cells with initial growth at 37 °C followed by induction and overnight growth at 16 °C. M1 and M2 are size markers of peptides of known sizes (kDa). Sizes of standards in M1 are shown on the gel. Those in M2 are of sizes 24, 45, 66, 97.4, 116 and 205 kDa, beginning at the bottom of the gel. The bulk of the protein eluted in 100 mM imidazole fractions.

As for those experiments described in Chapter Three, NMR analysis of *Hi*AXHd3 was performed and interpreted by Maria Peña at the CCRC. For analysis by NMR, *Hi*AXHd3 was incubated with wheat arabinoxylan or arabinooligosaccharides generated from arabinan. Substrate at 2 mg ml^{-1} was incubated with enzyme at 100 nM at 37 °C for 16 hours; control reactions without enzyme were also prepared. Water was removed from the samples, after incubation, by lyophilisation. Immediately prior to NMR analysis, samples were resuspended in a small volume of D_2O by M. Peña.

Consistent with *Hi*AxHd3 expressed from *Aspergillus,* NMR analysis showed that the *E. coli* form of the enzyme released arabinose from doubly substituted xylose residues in wheat arabinoxylan, Appendix E. The signals corresponding to the anomeric protons of α-L-arabinose residues linked at the O2 and O3 position to doubly substituted xylose residues are present in the spectrum of arabinoxylan prior to enzyme treatment, but were absent after the polysaccharide was incubated with *Hi*AXHd3. A signal diagnostic for an α-L-arabinose linked at the O2 position to singly substituted xylose residues, was only present in the spectrum of arabinoxylan after treatment with *Hi*AXHd3. These results indicate that xylosyl residues decorated at O2 with arabinose were not present in untreated arabinoxylan but were generated by enzymatic hydrolysis of the α-L-arabinose residue linked at O3 to double substituted xylose units.

1D-NMR analysis of the arabinooligosaccharide samples was difficult to interpret due to the complexity of the spectra, so 2D-NMR was performed to aid with assignation of peak identities. Analysis of the spectrum of the arabinan oligosaccharides after incubation with *Hi*AXHd3 (Appendix E) showed that the signal diagnostic of the doubly substituted residues disappeared. However, signals corresponding to the α-1,5 linked L-arabinofuranose residues substituted at O2 were more intense after treatment with the *Hi*AXHd3 enzyme. Signals for the α-1,5 linked L-arabinofuranose residues substituted at O3 remained unchanged after enzyme treatment. These data showed the signals corresponding to the anomeric protons of arabinose units that were substituted at both O2 and O3, and the respective□□-L-arabinose residues attached to O2 of 2,3,5 linked arabinose (backbone arabinose units that are doubly substituted), were lost after enzyme treatment. There was increased intensity of the signals corresponding to the anomeric protons of □-L-arabinose residues that have a single O2 substitution, and arabinose units linked O2 to the arabinan backbone. These data show that *Hi*AXHd3 displays specificity for the □-1,3-arabinose linkages in arabinose moieties that are substituted at both O2 and O3. It is likely, therefore, that the *Bifidobacterium* AXHd3 will also target the 1,3 linkage in backbone arabinose units that are doubly substituted.

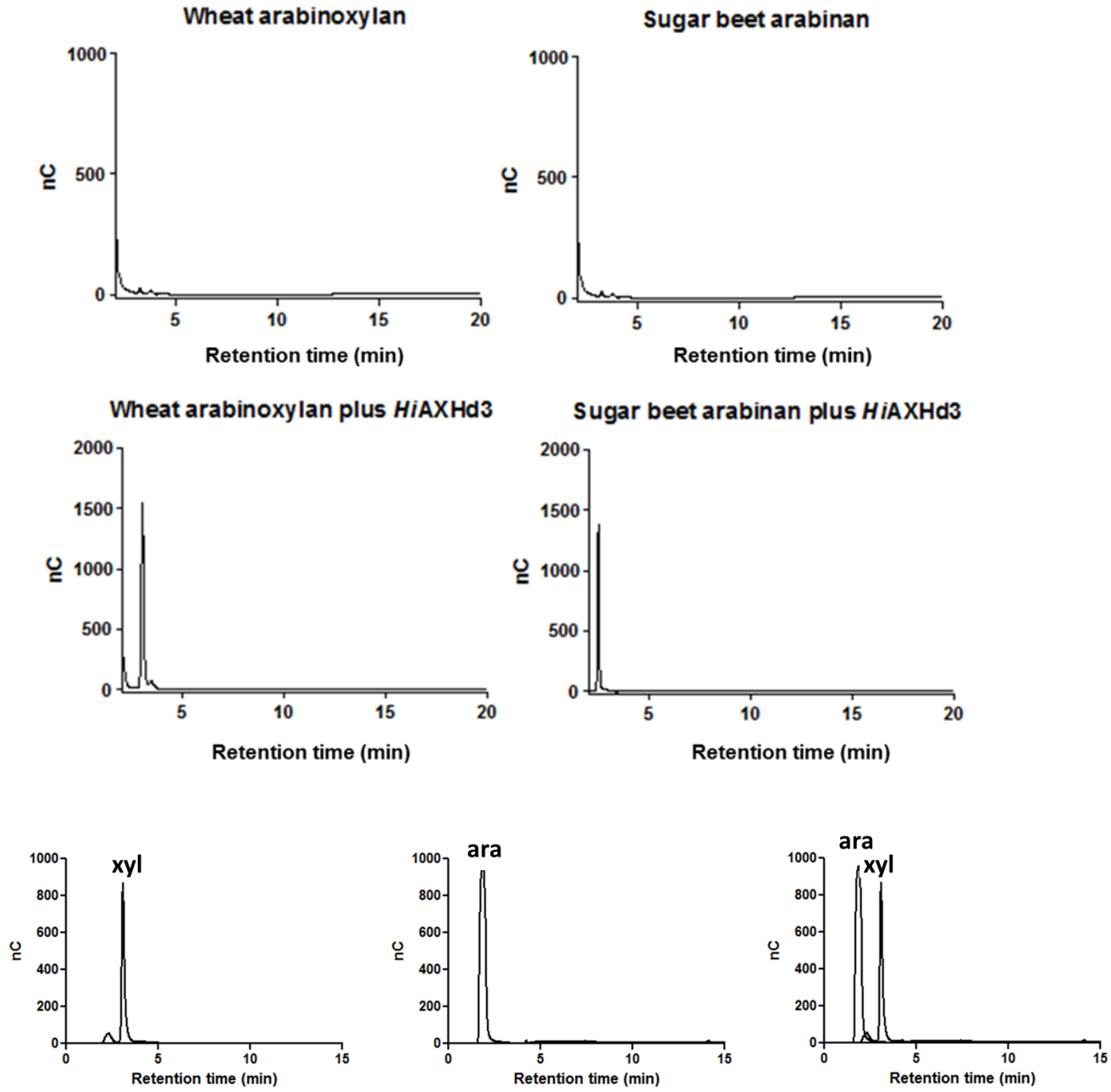


Figure 4.3 HPLC showing products generated by *Hi*AXHd3 incubation with wheat arabinoxylan and sugar beet arabinan.

HPLC chromatograms confirm activity of *Hi*AXHd3 against two substrates. These traces show that one product, arabinose, is released from wheat arabinoxylan and sugar beet arabinan upon incubation with *Hi*AXHd3. The reaction product was identified by co-migration with arabinose (Megazyme).

A continuous arabinose detection assay (galactose dehydrogenase assay, see Chapter Two, Section 2.3.4) was used to determine the kinetics of *Hi*AXHd3 action against a range of polysaccharide substrates. Consistent with the NMR data, the enzyme showed no activity against linear arabinan, in which the □-L-1,5-arabinose backbone is not decorated. The data, Table 4.2, showed that the enzyme was highly active against wheat arabinoxylan and indeed its K_M is around 10-fold lower than other enzymes that deconstruct arabinoxylans (xylanases and arabinofuranosidases active against monosubstituted arabinose residues (Beylot et al., 2001a)). Kinetic analysis of *Hi*AXHd3 activity against sugar beet arabinan is also presented. The catalytic efficiency (k_{cat}/K_M) for arabinoxylan is ~3.7 times that for arabinan.

Substrate	k_{cat} **(min^{-1})**	K_M **(M)**	k_{cat}/K_M **(min^{-1} M^{-1})**
Wheat arabinoxylan	$1.0 \times 10^4 \pm 1.3 \times 10^3$	3.5×10^{-4} 1.4×10^{-4}	2.9×10^7
Sugar beet arabinan	$1.7 \times 10^3 \pm 1.1 \times 10^2$	$1.3 \times 10^{-4} \pm 2.5 \times 10^{-5}$	1.4×10^7

Table 4.2 Kinetic analysis of wildtype *Hi*AXHd3 against wheat arabinoxylan and sugar beet arabinan.

Reactions were performed in triplicate to obtain Standard Error of the Mean (SEM). Effective substrate concentration in units of M was obtained from concentration of polysaccharide (mg ml^{-1}) by submitting the polysaccharides to complete degradation by *Hi*AXHd3. The final absorbance was used to calculate the concentration of target arabinose residues (O3-linked arabinose in the double substitution) in 1 mg of polysaccharide. Wheat arabinoxylan contains 4.61×10^{-4} moles of target arabinose per milligram. Sugar beet arabinan contains 1.84×10^{-4} moles of target arabinose per milligram.

4.2 Wildtype structure

Non-tagged *Hi*AXHd3 was expressed in *Aspergillus* and deglycosylated by Novozymes. The protein was exchanged into water and concentrated to 15 mg ml^{-1}. Four preliminary crystal screens were set up and incubated at 20 °C. Crystals grew in several conditions within one week in the Qiagen PACT screen and were optimised by manipulation of mother liquor components. All crystals were of a long three-dimensional needle shape (Figure 4.4). Optimum growth was found in 22 % PEG-1500 and 1 M MMT buffer (pH 7.0). A full dataset was collected on an in-house

generator at Newcastle University to a resolution of 1.6 Å. The crystal indexed to space group P1 21 1.

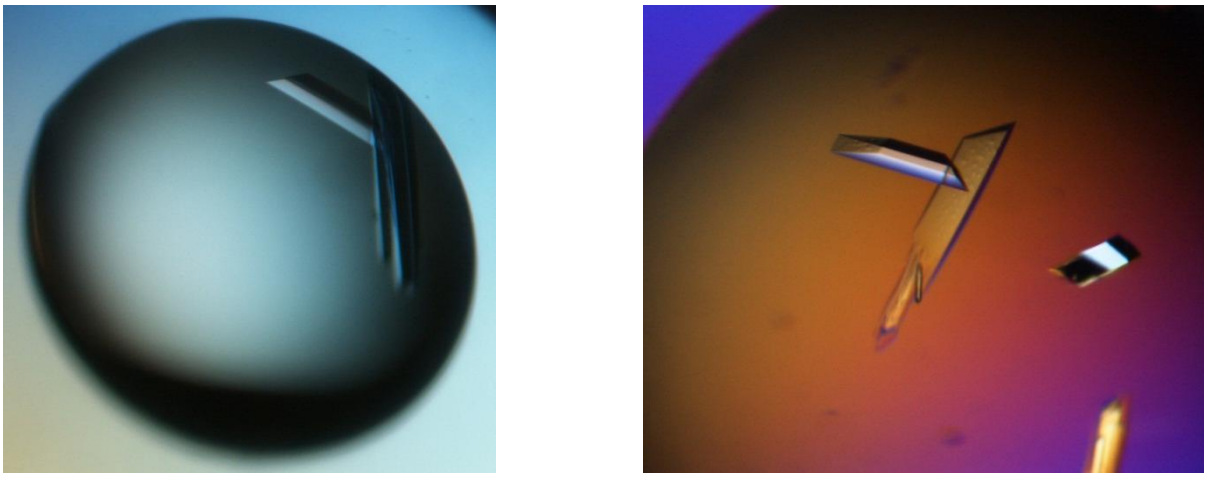


Figure 4.4 Crystals of deglycosylated untagged wildtype *Hi*AXHd3 expressed in an *Aspergillus* system.

Crystals shown were grown in 22 % PEG-1500 and 1 M MMT buffer (pH 7.0). Maximum growth was obtained within one week of incubation at 20 °C.

As no suitable homologous search model was available for molecular replacement, a heavy atom technique was employed to solve the structure of *Hi*AXHd3. A selenomethionine derivative protein was expressed in *E. coli* strain B834, a methionine auxotroph. Cells were incubated at 37 °C prior to induction and overnight growth at 16 °C. Expression of the selenomethionine derivative was roughly comparable to wildtype expression in Tuner cells and Figure 4.5 shows that IMAC purification was successful. Purity and homogeneity were further improved by subsequent anion exchange chromatography and gel filtration (Figure 4.5) using an FPLC system (see Chapter Two, Section 2.2.21). After pooling all pure fractions, the protein was exchanged into molecular biology grade water (Sigma) and concentrated to 20 mg ml^{-1}. The protein was placed in three 96-condition crystallisation screens (Qiagen) and incubated at 20 °C. Precipitation could be observed in some conditions after three-four days, and crystals began to appear within a week in several conditions. Optimal growth was obtained in 20-22 % PEG-3350, 0.1 M Bis-Tris (pH 6.5) and 2 M ammonium sulphate.

The crystal structure of *Hi*AXHd3 was solved using selenomethionine single wavelength anomalous dispersion data obtained at the Diamond Light Source synchrotron to a resolution of 1.85 Å, in harness with native data to 1.6 Å (see experimental statistics in Table 4.3). The selenomethionine structure was solved according to the method described in Chapter Two, Section 2.4.6.i. The crystals were in the P1 21 1 space group and contained two molecules in the asymmetric unit. The ShelxCDE suite of programmes (Sheldrick, 2008) was used to detect the anomalous signals of the selenium-containing residues, and these signals were used to calculate structure phases. After phase extension and improvement, the protein structure was built using the Buccaneer programme (Cowtan, 2006). The structure had final R_{work} and R_{free} values of 16.47 % and 20.95 %, respectively. The structure consisted of 525 amino acids and 1064 water molecules. The N-terminal residue Gln27 was disordered, displaying no electron density and thus was not included in the final model. *Hi*AXHd3 consists of two distinct domains that are linked by a flexible linker sequence, Val329 to Gly338. The linker was disordered in crystal, yielding weak density for the side chains, leading to somewhat higher B values for these residues. The two domains in *Hi*AXHd3 make extensive interactions.

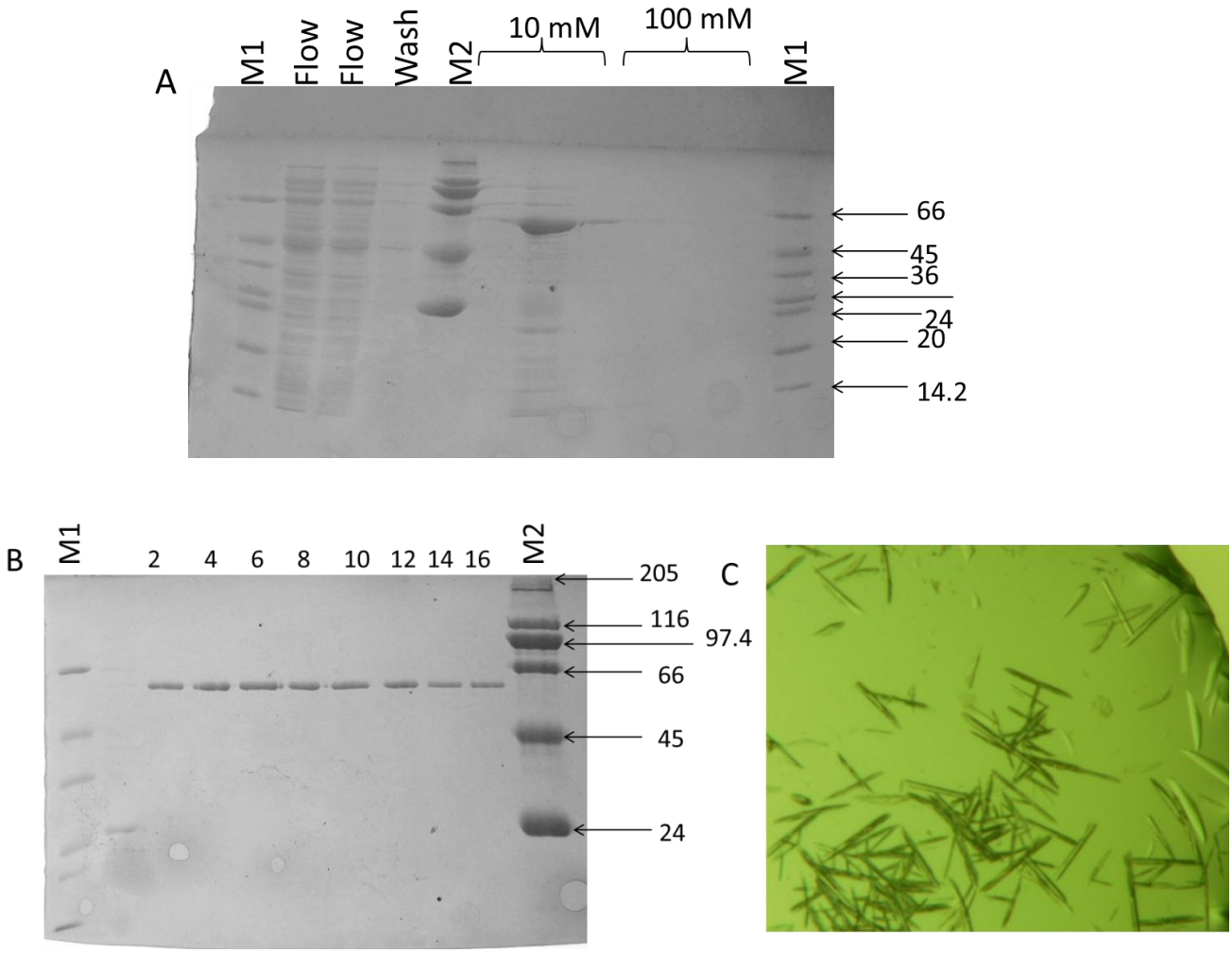


Figure 4.5 SDS-PAGE analysis of expression and purification of selenomethionine *Hi*AXHd3 and subsequent crystallisation of the protein.

Gels showing typical expression/purification profile (A) and subsequent gel filtration (B) of the selenomethionine derivative of *Hi*AXHd3, expressed in *E. coli* cells, strain B834.

C: crystals of the His-tagged selenomethionine protein were typically thin needles, smaller than seen for the non-tagged native protein.

The N-terminal domain, consisting of residues 27-328 in the full length protein, displays a five-bladed β-propeller fold, typical of GH43 enzymes. It has a cylindrical shape with a diameter and height of ~45 Å. The *Hi*AXHd3 propeller is based upon a 5-fold repeat of blades that adopt the classical “W” topology of four anti-parallel β-strands that comprise a β-sheet. The blades are radially arranged from the centre of the propeller. Within each blade, particularly in the 5th blade, the β-strands are highly

twisted with respect to their neighbours, resulting in an angle of approximately 90 ^{o} between the first and last β-strand of each blade. The N-terminal domain contains a single disulphide bond between Cys163, in the loop connecting blades 2 and 3 and Cys125 in blade 2. In some proteins that display a five-bladed β-propeller fold, the fifth blade consists of strands from the N and C terminus, closing the β-propeller in what is termed 'molecular velcro', which appears to provide considerable stabilisation to the fold (Neer and Smith, 1996). This is not present in *Hi*AXHd3, where the fifth strand comprises only β-strands from the C-terminal sequence of the domain, so that the propeller fold is not 'closed' by the fifth blade. The N- and C-terminal blades of the enzyme are, however, in close association and stability is provided by strong hydrogen bonds between Arg47 in blade 1 and Asp280 in blade 5. In addition, a small anti-parallel β-sheet consisting of the N- and C-termini, which is distinct from the β-propeller structure, is likely to also stabilise the overall fold of the domain. For this small β-sheet, the N-terminal β-propeller domain donates residues 30-33 to form one strand, while the second strand comprises residues 324-326, which lie immediately upstream of the inter-domain loop (Figure 4.6).

The C-terminal domain, extending from residues 339-558 in the full length enzyme, displays a canonical β-sandwich fold. The convex β-sheet consists of β-strands in the order β-1, β-2, β-12, β-4, β-9 and β-10, while the order of the β-strands in the concave β-sheet is β-3, β-11, β-5, β-6, β-7 and β-8. A solvent-exposed shallow cleft is presented at the interface between the two domains, Figure 4.7. The N-terminal β-propeller domain provides the bulk of the cleft, primarily blades 4 and 5 and their associated loops. One of the faces of the cleft, however, is presented by components of the C-terminal domain; the loop connecting β-strands 9 and 10, and the central region of the extended loop that joins β-strands 10 and 11 (Figure 4.7).

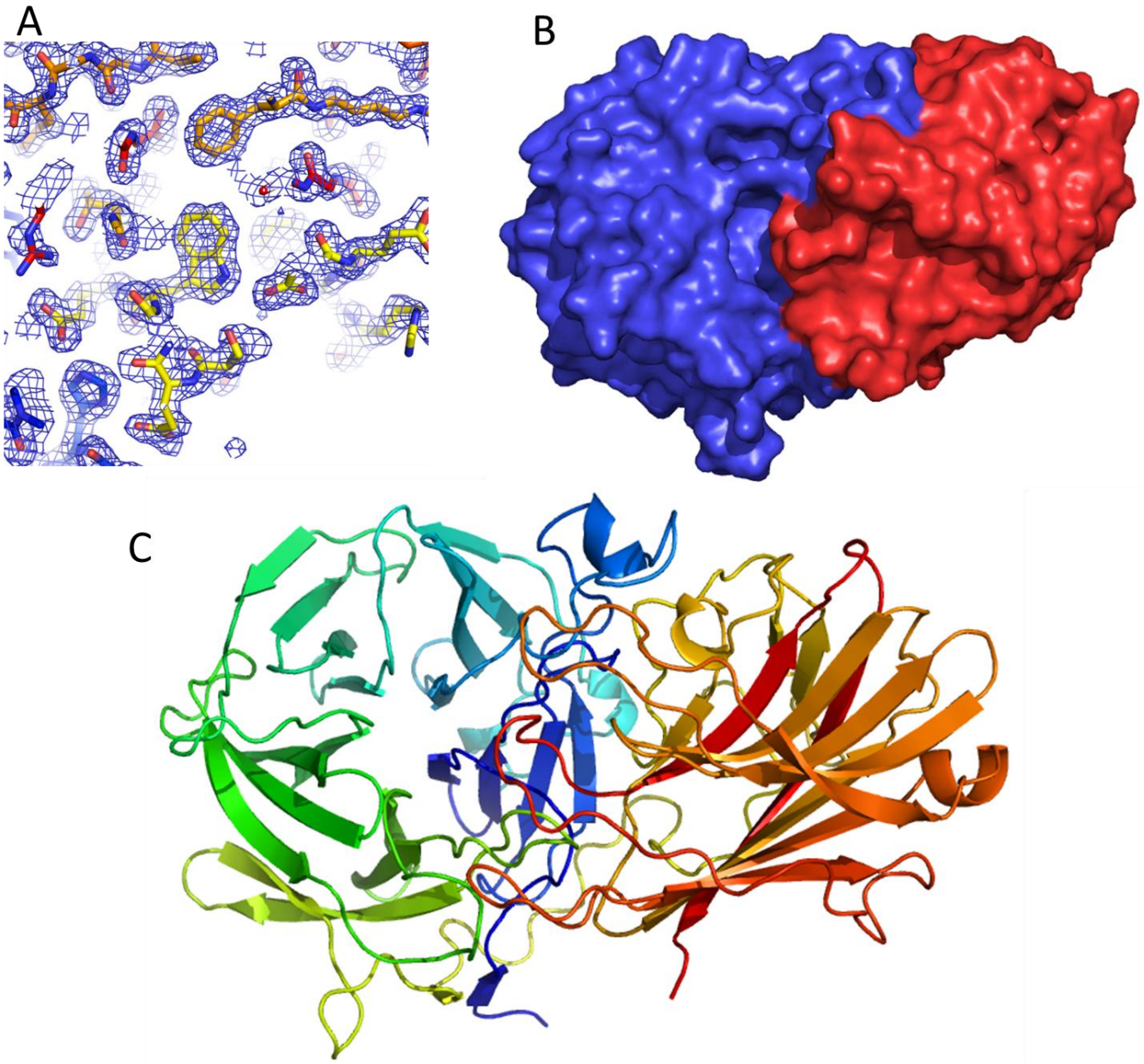


Figure 4.6 Representations of the structure of selenomethionine *Hi*AXHd3.

A: electron density fit in a representative region of the structure. The figure displays the weighted maximum likelihood Fobs – Fcalc map in blue mesh. The map is contoured to 1.0 σ.

B: surface representation of the enzyme coloured to indicate the separation between the catalytic domain (blue) and the C-terminal domain of unknown function (red).

C: cartoon representation of the secondary structure of *Hi*AXHd3. The β-propeller fold of the catalytic domain (N-terminal, coloured blue to green) and the β-sandwich fold (C-terminal, coloured yellow to red) are visible.

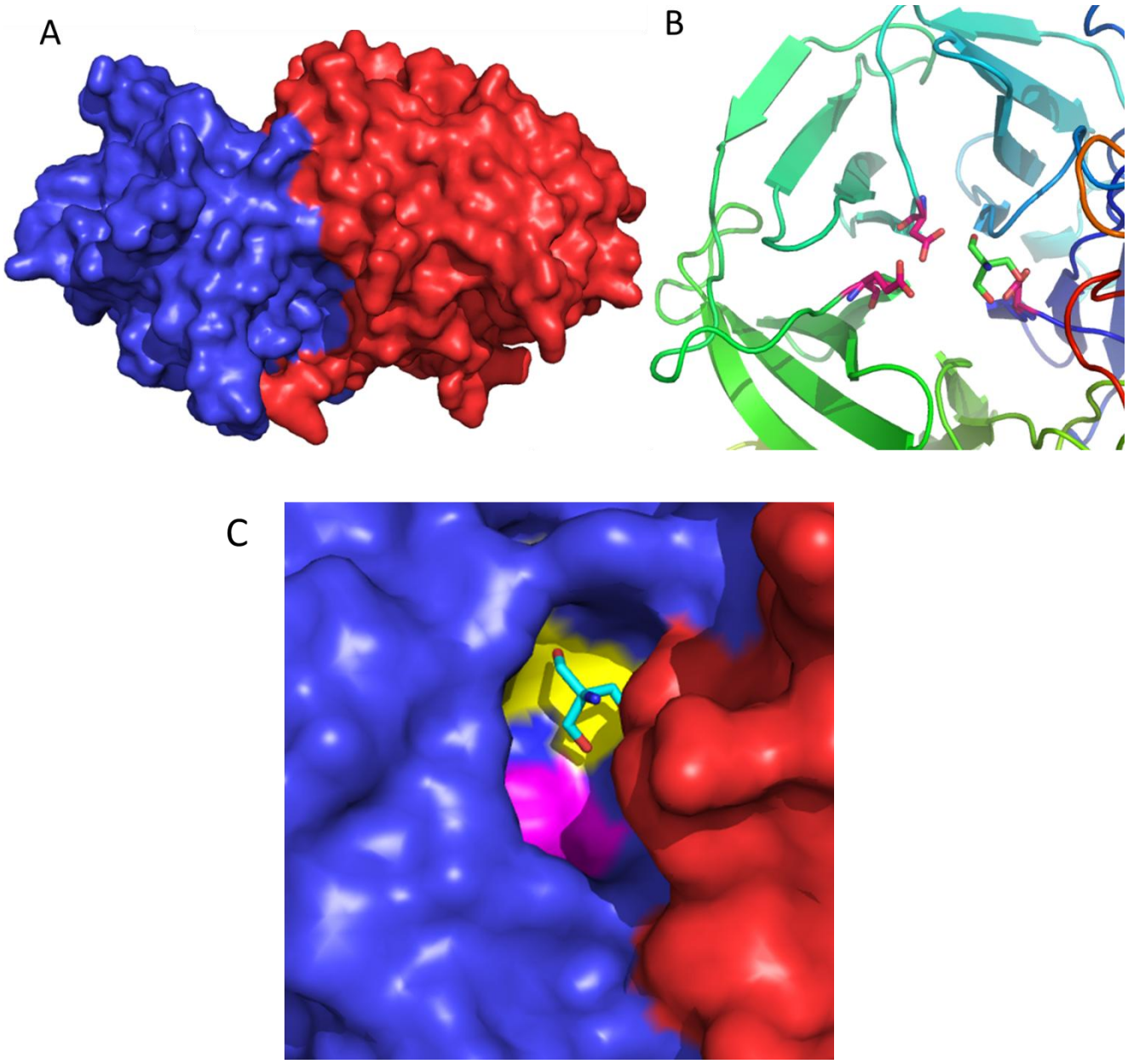


Figure 4.7 Structural representations of *Hi*AXHd3, focussing on the region of the active site.

A: surface representation of the enzyme. The polysaccharide binding cleft (viewed side-on) is built from the N-terminal (blue) and C-terminal (red) modules of the protein.

B: cartoon representation of the enzyme with Tris in the active site highlights the locations of the catalytic residues within the β-propeller. The catalytic residues are shown in stick form and coloured red.

C: surface representation of the active site of *Hi*AXHd3 shows the two layered pocket structure of the active site. The active site of the pocket, which houses a Tris molecule, is indicated by the yellow residues, while the secondary pocket is shown by the magenta residues.

In the central region of the cleft is a heart shaped depression that consists of a shelf-like structure abutted to a relatively shallow pocket, Figure 4.7. Strong density for a single molecule of Tris, the buffer used in crystallisation, is present within this pocket in both molecules in the asymmetric unit (Figure 4.7). The pocket houses a constellation of carboxylate residues (Asp42, Asp166 and Glu215) which are invariant within GH43, and have been shown to comprise the catalytic amino acids in several members of the family (Brux et al., 2006; Nurizzo et al., 2002; Proctor et al., 2005). Asp42 is the general base and Glu215 is the general acid. Asp166 is believed to play a role in modulating the pKa of the general acid and orienting the catalytic residue and substrate to facilitate catalysis. To confirm the identity of the catalytic residues, which were also pinpointed by sequence alignments with previously characterised enzymes, all three amino acids were mutated to Alanine by site-directed mutagenesis. All mutants were inactive against wheat arabinoxylan and sugar beet arabinan, supporting the view that these three acidic residues comprise the catalytic apparatus of *Hi*AXHd3.

4.3 Ligand crystallography

To further explore the mechanism of substrate recognition in *Hi*AXHd3, the crystal structure of the enzyme in complex with an appropriate substrate was sought. An oligosaccharide ligand was generated by partial digestion of wheat arabinoxylan with the GH10 xylanase *Cj*Xyn10A (Harris et al., 1996). Figure 4.8 comprises HPLC chromatograms showing the product profile of this reaction. Arabinoxylooligosaccharide (AXOS) substrates for *Hi*AXHd3 were identified by adding the enzyme to a small amount of the saccharide mixture and observing shifts in the HPLC peaks (Figure 4.8). The figure shows that GH10 degradation of the polysaccharide yields a range of peaks, including one at ~22 minutes retention time. Upon subsequent addition of *Hi*AXHd3, this peak shifts to ~15 minutes, and this is accompanied by the appearance of a peak at ~3 minutes, identified as arabinose by co-migration with a standard. Thus, it would seem that the product which elutes at 22 minutes is a target for hydrolysis by *Hi*AXHd3. After long-chain polysaccharide was removed by ethanol precipitation (65 %), the desired product was purified by size exclusion chromatography on a P2 biogel column. Figures C.1 and C.2 in Appendix C show the results of this purification as analysed by TLC. Fractions were tested for

*Hi*AXHd3 substrate by incubation with the GH43 enzyme followed by an arabinose detection assay (arabinose/galactose dehydrogenase). Fractions containing the appropriate AXOS substrate were those where *Hi*AXHd3 released arabinose; these were pooled and concentrated by lyophilisation.

To obtain a substrate-bound form of the enzyme in crystal, inactive forms of the protein were generated by site-directed mutagenesis. The mutants D42A and E215C were provided by Novozymes, again expressed in a non-tagged form in the *Aspergillus* system. Both mutants were exchanged into water and concentrated to 10 mg ml^{-1}. A crystal screen was set up for both proteins focussing on twelve conditions from the PACT screen which gave good results for the non-tagged unliganded wildtype protein (PACT 17, 29, 30, 31, 39, 40, 52, 53, 54, 63, 64, 65). The screens were set up in duplicate, with protein in the presence and absence of oligosaccharide ligand, respectively. No crystals grew for either mutant in the presence of ligand, suggesting that the oligosaccharides disrupted the formation of crystal contacts or altered protein solubility in the mother liquor. Observation of the wildtype structure indicated that solvent channels within the crystal lattice would permit entry of ligand. Therefore, as good quality crystals of small size were observed in all conditions without ligand, the conditions that gave the best crystals were optimised for use in ligand soaks. Optimum growth for D42A was found in 22.5 % PEG-4000, 0.1 M HEPES buffer (pH 7.5) and 0.1 M sodium chloride. Optimum growth for E215C was found in 20 % PEG-4000, 0.1 M HEPES buffer (pH 7.5) and 0.2 M sodium chloride.

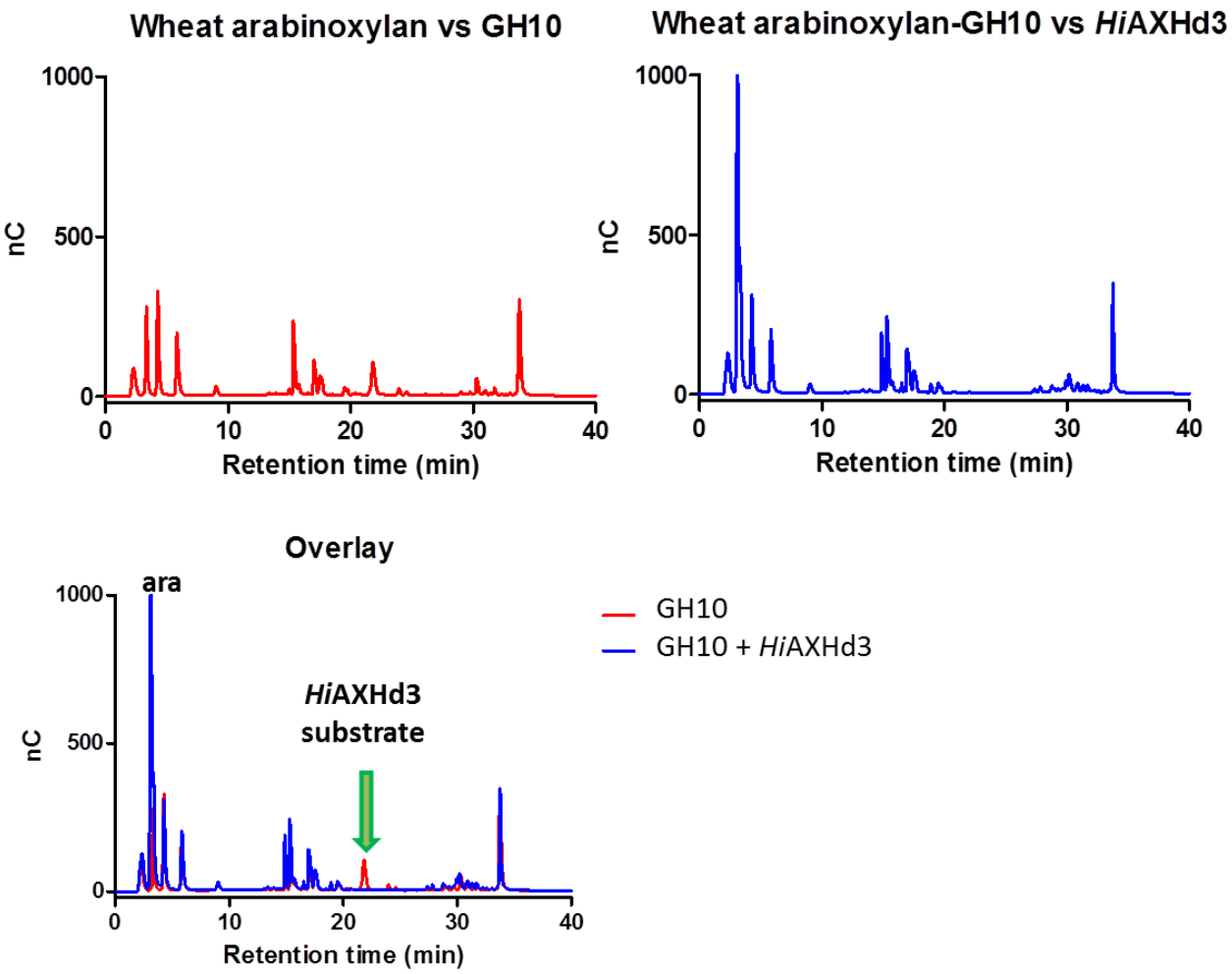


Figure 4.8 HPLC chromatograms detailing the production and identification of a suitable AXOS ligand for *Hi*AXHd3.

Wheat arabinoxylan was degraded by GH10, generating a range of products of various sizes (red). Subsequent treatment with *Hi*AXHd3 (blue) generated a large amount of arabinose (~3 minutes). An overlay of these profiles reveals that one red peak (at ~22 minutes) is not present in the blue profile, indicating that this is an AXOS which contains the double substitution targeted by *Hi*AXHd3. This peak is indicated in the figure by a green arrow. Subsequent purification techniques were employed to isolate this product.

In preparation for ligand soaking, the stability of crystals was tested by fishing and transferring to an 'alternative mother liquor' of 27 % PEG-3350 and 0.1 M HEPES buffer (pH 7.5). Crystals tested in this way were very stable, remaining unchanged for the duration of the incubation period (fifteen minutes to two hours). For ligand soaks, a few milligrams of the purified AXOS were resuspended in as little mother liquor as possible. A 4 µl aliquot of this mixture was dispensed into a soaking tray

well, into which a fished crystal was placed. Crystals were monitored for stability under a microscope and were soaked for between fifteen minutes and two hours. Finally, crystals were fished with cryoprotectant containing ligand and snap frozen in liquid nitrogen. Datasets for soaked crystals were collected at the SER-CAT synchrotron; all were of resolution between 1.6 and 1.9 Å.

In total, four datasets were collected on crystals soaked for 15 minutes, 30 minutes, one hour and two hours. All ligand-soak datasets indexed to space group P1 21 1, as did the wildtype crystals. Using the selenomethionine structure as the search model, these structures were solved by molecular replacement using the Phaser programme in CCP4. After extension of the electron density by several rounds of refinement in Refmac, density for sugar was found in the active site of only one crystal (see experimental statistics in Table 4.3). In D42A crystals that had been soaked with the ligand for 2 hours, clear electron density corresponding to xylotriose with a single arabinose side chain (Figure 4.9), linked O2 to the xylose at Xyl-2 (Xyl-1 is the reducing end xylose, Xyl-3 is the non-reducing end xylose), was observed in the active site of both molecules in the asymmetric unit. Extremely weak electron density extending beyond the reducing and non-reducing end of the xylotriose molecule was observed, indicating that the ligand consists of a decorated xylopentaose unit, although the terminal residues could not be appropriately refined as xylopyranosides. The disordered structure of the terminal xylose molecules suggests that interactions of these sugar units with the enzyme are very weak. The backbone structure of the ligand is therefore defined as xylotriose.

		SelMet *Hi*AXHd3	*Hi*AXHd3/Araf-Xyl_3	*Hi*AXHd3/$Araf_2$-Xyl
	Space group	$P2_1$	$P2_1$	$P2_1$
	Wavelength	0.98	1.00	1.54
	Unit cell	a = 65.3	a = 65.4	a = 64.9
		b = 78.3	b = 52.9	b = 53.1
		c = 95.8	c = 146.9	c = 148.1
		α = 90.0	α = 90.0	α = 90.0
		β = 103.1	β = 101.7	β = 101.5
		γ = 90.0	γ = 90.0	γ = 90.0
	Resolution	47.7-1.9	143.8-1.4	144.1-2.6
	No. observations	139366 (20370)	426983 (61153)	82432 (12635)
	No. unique observations	19083 (2731)	177611 (25617)	30444 (4418)
	Multiplicity	7.3 (7.5)	2.4 (2.4)	2.7 (2.9)
	Anomalous multiplicity	3.7 (3.8)		
	Completeness	100.0 (100.0)	99.2 (98.3)	98.5 (99.0)
	Anomalous completeness	99.6 (99.6)		
	Average I/sigma I	23.5 (11.4)	16.2 (9.8)	11.5 (7.3)
	R_{sym}	0.072 (0.158)	0.084 (0.172)	0.074 (0.137)
Refinement				
	No. reflections	80208	167521	28911
	R_{cryst}	16.47 %	18.41 %	18.34 %
	R_{free}	20.95 %	21.83 %	27.83 %
	No. atoms	9098	9306	8914
Rmsd				
	Bond lengths	0.011	0.028	0.012
	Bond angles	1.332	2.318	1.481
Ramachandran				
	Favoured	95.46 %	95.73 %	94.27 %
	Allowed	100.00 %	99.90 %	99.40 %
B-factors				
	Wilson B	18.053	11.231	6.786
	Main chain	16.321	8.356	6.207
	Side chains and waters	19.576	13.615	7.306

Table 4.3 Experimental statistics for three structure solutions of *Hi*AXHd3.

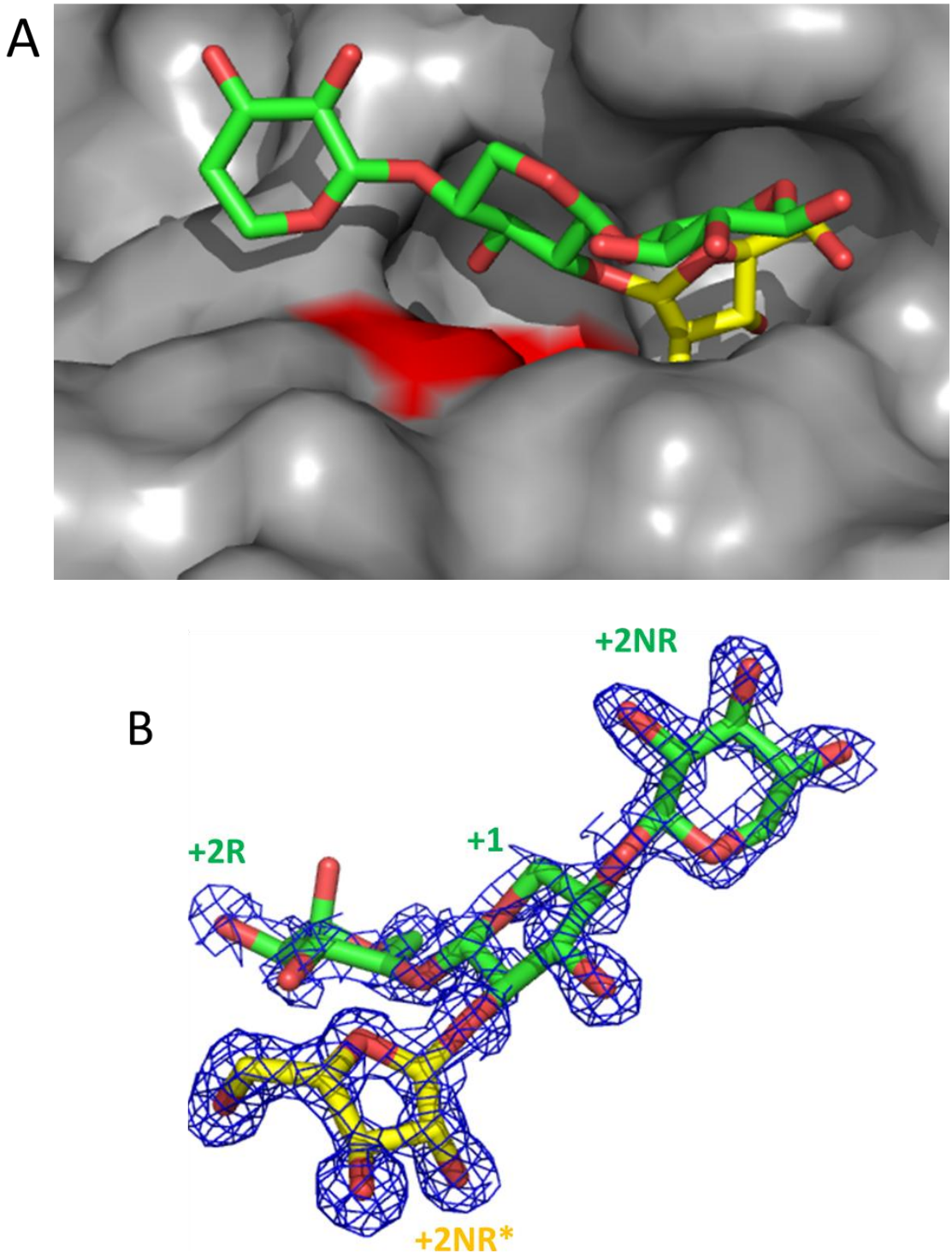


Figure 4.9 Structure representations of the high resolution *Hi*AXHd3/Ara*f*-Xyl$_3$ structure.

A: surface representation of the enzyme-substrate complex show the O2-linked arabinose in the +2NR* pocket adjacent to the active site. The arabinose at +2NR* is coloured yellow to distinguish it from the green xylose residues of the ligand. Oxygen atoms in the ligand are coloured red. The catalytic residues of the enzyme are shown in red.

B: view of the electron density surrounding the oligosaccharide ligand. The figure shows the weighted $2F_{obs} - 2F_{calc}$ map (blue mesh) generated by refinements with the oligosaccharide included. The map is contoured to 1 σ. Ligand subsites are identified on the figure. Oxygen atoms in the ligand are coloured red.

A nomenclature system for identifying ligand subsites in glycoside hydrolases is widely used and is adapted from the system of Davies and colleagues (Davies et al., 1997). According to this nomenclature, subsites are labelled –n to +n, where –n is at the non-reducing end of the substrate and +n is at the reducing end. Cleavage occurs between subsites -1 and +1. Using this nomenclature to identify the subsites in *Hi*AXHd3 that bind to substrate, xylotriose occupies subsites +2NR to +2R with the arabinose positioned at subsite +2NR*, Figure 4.9. No arabinose moiety was observed in the active site (-1 subsite). The xylan chain is pseudo-symmetrical and fits equally well into electron density in the two opposite orientations. The orientation assigned was based on three criteria: the observed hydrogen bonds to solvent and protein from endocyclic O5 atoms (whose equivalent would be C5 in the reverse orientation), the behaviour of the crystallographic temperature factors after refinement, which showed discrepancies at C5 and O5 in the reverse orientation, and the requirement for O3 to be pointing into the active site (-1 subsite). The observed ligand represents one of the reaction products of *Hi*AXHd3 action on wheat arabinoxylan, the other being free arabinose. The trapping of the reaction products in crystal, rather than substrate, likely reflects trace activity displayed by the mutant which was not observed in the spectroscopic assay over thirty minutes, but which was sufficient for hydrolysis to occur during a two hour incubation with highly concentrated substrate (ligand). However, as the ligand was not formally characterised prior to crystal soaking, by methods such as NMR or mass spectrometry, it may be that the oligosaccharide did comprise only the O2-linked arabinose, and that binding to this sugar residue was sufficient to allow the singly substituted substrate to bind the active site.

More information about the specificity of *Hi*AXHd3 would be obtained from a structure featuring an arabinose residue in the -1 subsite, so a second ligand structure was sought. The soaking method described above was employed with crystals of E215C but significantly shorter soak times were used, to minimise the

chance of hydrolysis. Although several crystals were soaked, again only one showed sugar in the active site. This structure was solved by molecular replacement using a dataset collected on an in-house generator at the University of Georgia to a resolution of 2.8 Å (Figure 4.10; see experimental statistics in Table 4.3). In crystals of E215C that had been soaked for 5 minutes, a xylose was located at the +1 subsite, which was decorated with an α-L-arabinose at O2 that again was situated at the +2NR* subsite. Weak density was observed for an additional xylose residue at each end of the ligand, but this could not be properly refined. The full ligand is likely xylotriose with the double arabinose substitution, but weak interactions between the enzyme and the xylan backbone mean that the terminal xylose residues are disordered in the crystal structure. The ligand density was visible for only one of the two molecules in the asymmetric unit. An arabinose was also observed in the -1 subsite of the E215C mutant, which is orientated to make an α-1,3-linkage with the xylose at +1, however, the distance between the two sugars is too long to represent a covalent bond, Figure 4.10 The saccharide observed in this structure represents the two reaction products generated by *Hi*AXHd3. Together, the structures of the two enzyme variants in complex with the reaction products provide insight into how *Hi*AXHd3 recognises its substrate.

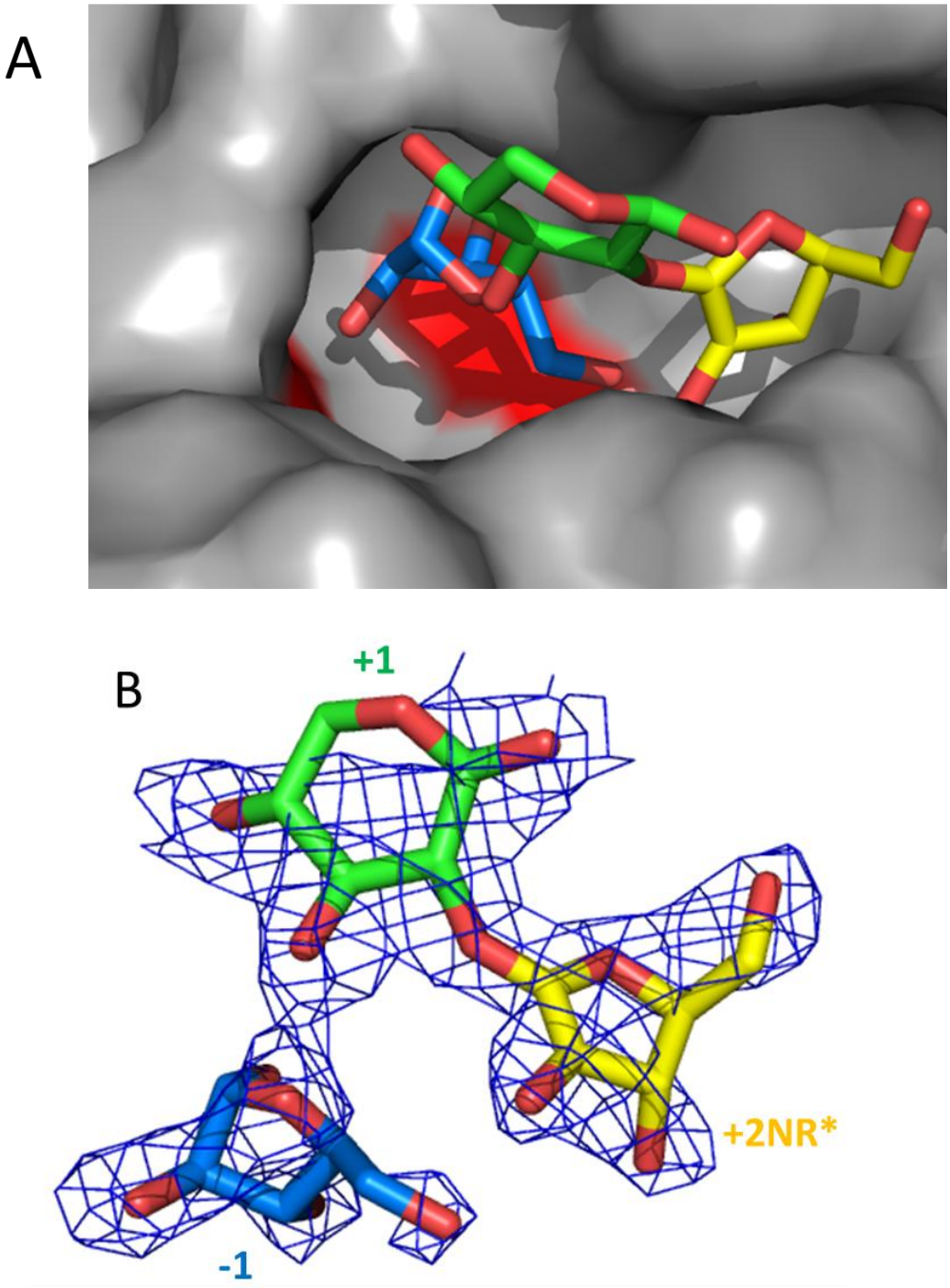


Figure 4.10 Structure representations of the low resolution *Hi*AXHd3/Araf_2-Xyl structure.

A: surface representations of the complex showing the arabinose residues buried in the pocket while the xylose (+1 subsite) stands proud of the protein surface. Catalytic residues are coloured red. The O2-linked arabinose is coloured yellow while the O3-linked arabinose (-1) in the active site is coloured blue. Oxygen atoms in the ligand are coloured red.

B: electron density surrounding the ligand. The figure shows the weighted $2F_{obs} - F_{calc}$ map (blue mesh) generated by refinements with the oligosaccharide included. The map is contoured to 1 σ. Ligand subsites are identified in the figure. Oxygen atoms in the ligand are coloured red.

4.4 Mutagenesis studies

Based on analysis of structures of *Hi*AXHd3 with bound reaction products, several amino acids were identified as being potentially important in the specificity of the enzyme for arabinose bound at O3 in the context of a double substitution. The sugar at the -1 subsite (O3-linked arabinose) interacts with the catalytic residues Asp42, Asp166 and Glu215, as well as Arg296 and Gln272. The O2 arabinose at the +2NR* subsite makes polar contacts with the residues Arg296, His271 and Asp290 while Phe288 lies on the floor of the O2 binding pocket. In addition, Trp525 and Tyr165 interact with the xylan backbone. These amino acids were each mutated to Alanine to assess the effects of the mutations on catalysis and specificity. As the hydroxyl group of Tyr165 makes a polar contact with the endocyclic oxygen of the xylose at the +2NR subsite, while the bulk of the phenolic side chain forms a wall of the active site pocket, this residue was also mutated to Phenylalanine (Y165F) and Tryptophan (Y165W) in order to disrupt substrate binding but maintain the structural integrity of the pocket. Figure 4.11 shows the locations of these residues.

Each mutant was assessed by continuous arabinose release assay for activity against wheat arabinoxylan and sugar beet arabinan. Table 4.4 summarises the effects of each mutation on enzyme activity. Mutation of any residue of the catalytic triad completely inactivated the enzyme, as did mutation of the active site residues Arg296, Gln272 and Asp290. Mutation of the residues His271 and Phe288, located at the +2NR* subsite, led to a reduction in k_{cat} with little or no change in K_M. Mutation of the residues Tyr165 or Trp525, which contact the xylan backbone, again led to a reduced k_{cat} with little or no change in K_M.

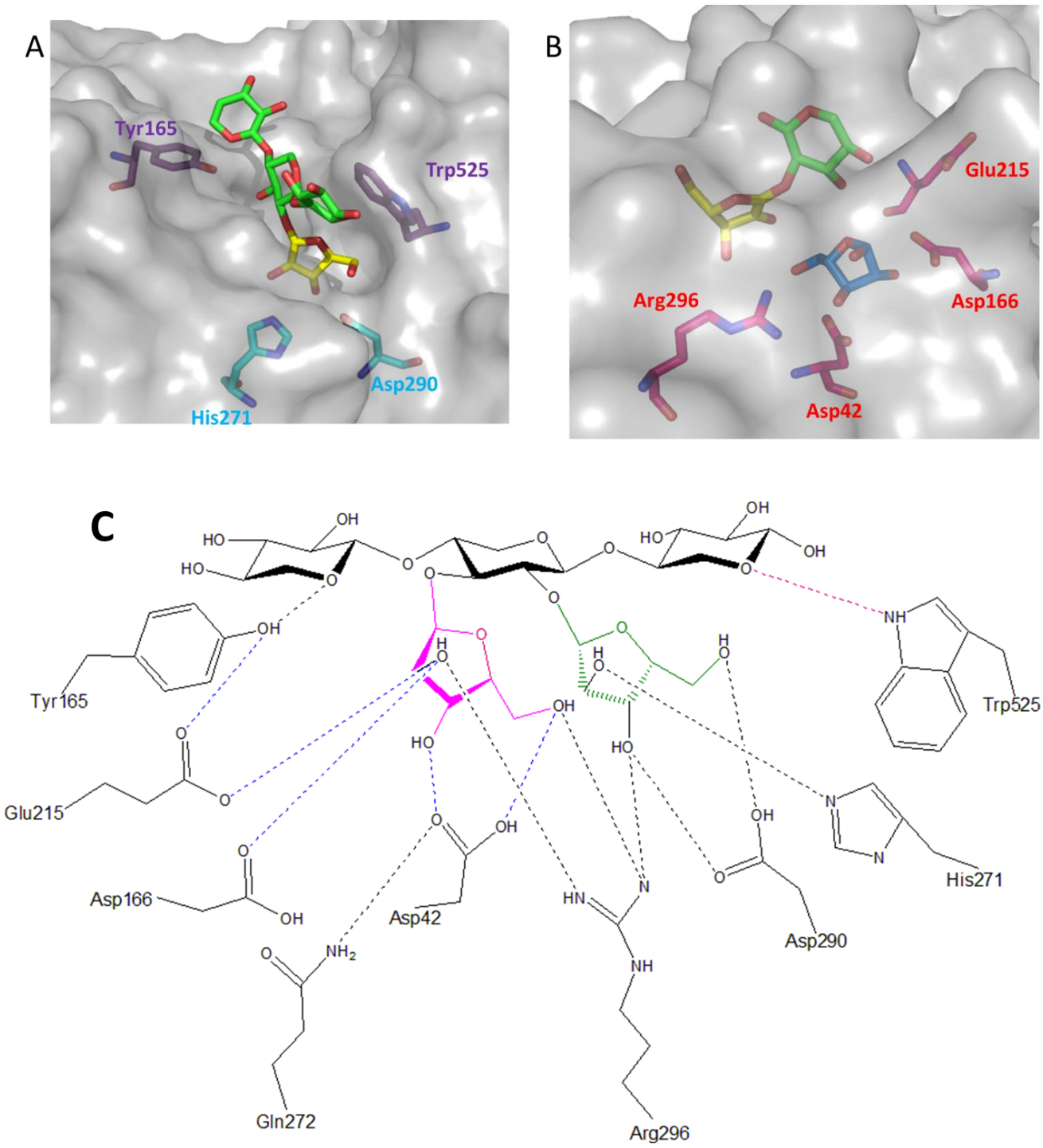


Figure 4.11 Enzyme-substrate contacts in *Hi*AXHd3.

A: the cyan residues (Asp290 and His271) makes contact with the O2-linked arabinose (+2NR* subsite). The purple resides (Trp525 and Tyr165) contact the xylan backbone of the substrate. The O2 arabinose is shown in yellow while the xylotriose backbone is in green.

B: four residues make polar contacts with the O3-linked arabinose (-1 subsite). This arabinose is shown in blue while the amino acids (catalytic residues Asp42, Asp166 and Glu215, and Arg296) are shown in red.

C: the schematic shows all contacts between the enzyme and substrate. The O3-linked arabinose (-1 subsite is shown in magenta while the O2-linked arabinose (+2NR* subsite) is green. Hydrogen bonds are not shown to scale.

It is possible that mutation of those residues with key roles in substrate specificity could lead to an altered specificity, or increased catalytic flexibility, perhaps allowing the enzyme to remove either O2 or O3 linked arabinose within the double substitution. Thus, the specificity of the mutants against wheat arabinoxylan was analysed by NMR. While it is likely that Arg296 and Asp290, as they bind to the O2-linked arabinose, are important in specificity determination, the effects of their mutation could not be assessed, as neither mutant (R296A and D290A) was active. All mutants making contact with the arabinose residues displayed wildtype specificity and only W525A showed any change in specificity (Figure 4.12). Trp525 appears to be important in coordination of the xylan backbone via a stacking interaction and hydrogen bond with the endocyclic oxygen of the xylose at +2R (see Figure 4.9 for ligand subsite nomenclature). This mutant was shown by 1D NMR to cleave the O2 linked arabinose from the double substitution (Figure 4.12). In the spectra for arabinoxylan incubated with wildtype enzyme, peaks corresponding to the single O2 and O3-linked arabinose side chains are roughly equal in size (ratio 1:0.84), as cleavage of the double substitution has generated single O2 substitutions which are approximately as numerous as the pre-existing single O3 substitution. In the spectra for arabinoxylan incubated with the W525A mutant, the same profile is observed but the ratio of peak size for the single substitutions is around 1:0.47. This indicates that random cleavage of either arabinose residues in the double substitution by the mutant has produced both single O3 and O2 substitutions. Due to the large size of the peak for O3 linked arabinose in single substitutions, it could not be ruled out that W525A can also cleave the O3 from the double substitution. Thus, it appears that loss of backbone coordination by Trp525 allows the substrate (double substitution) to bind in either orientation, pointing either the O2 or the O3 linked arabinose into the active site (-1 subsite).

In short, the double substitution substrate of HiAXHd3 is oriented so that the O3-linked arabinose points into the active site of the enzyme by interactions with the xylan backbone. In addition to Trp525, the residue Tyr165 also hydrogen bonds to the backbone. It is intriguing therefore that this Tyrosine appears not to be important to specificity. To better understand why Trp525 is important to specificity while Tyr165 is not, the structure of ligand-bound HiAXHd3 was analysed when the orientation of the ligand is flipped 180 °, so that the O2-linked arabinose points into the active site (Figure 4.13). Tyr165 is within range to form a hydrogen bond with the endocyclic oxygen of a xylose residue in both orientations of the ligand, indicating that this residue does not assist the enzyme in discriminating between the arabinose residues in the double substitution. Conversely, Trp525 makes a weak hydrogen bond to a xylose residue when the O3-linked arabinose is in the active site. In the other orientation Trp525 is too distant to the xylose to make any polar contact. The requirement for Trp525 to form a hydrogen bond is therefore the likely mechanism by which wildtype HiAXHd3 maintains specificity for the O3-linked arabinose of the double substitution. Loss of this hydrogen bond in the W525A mutant permits the substrate to enter the active site in either orientation.

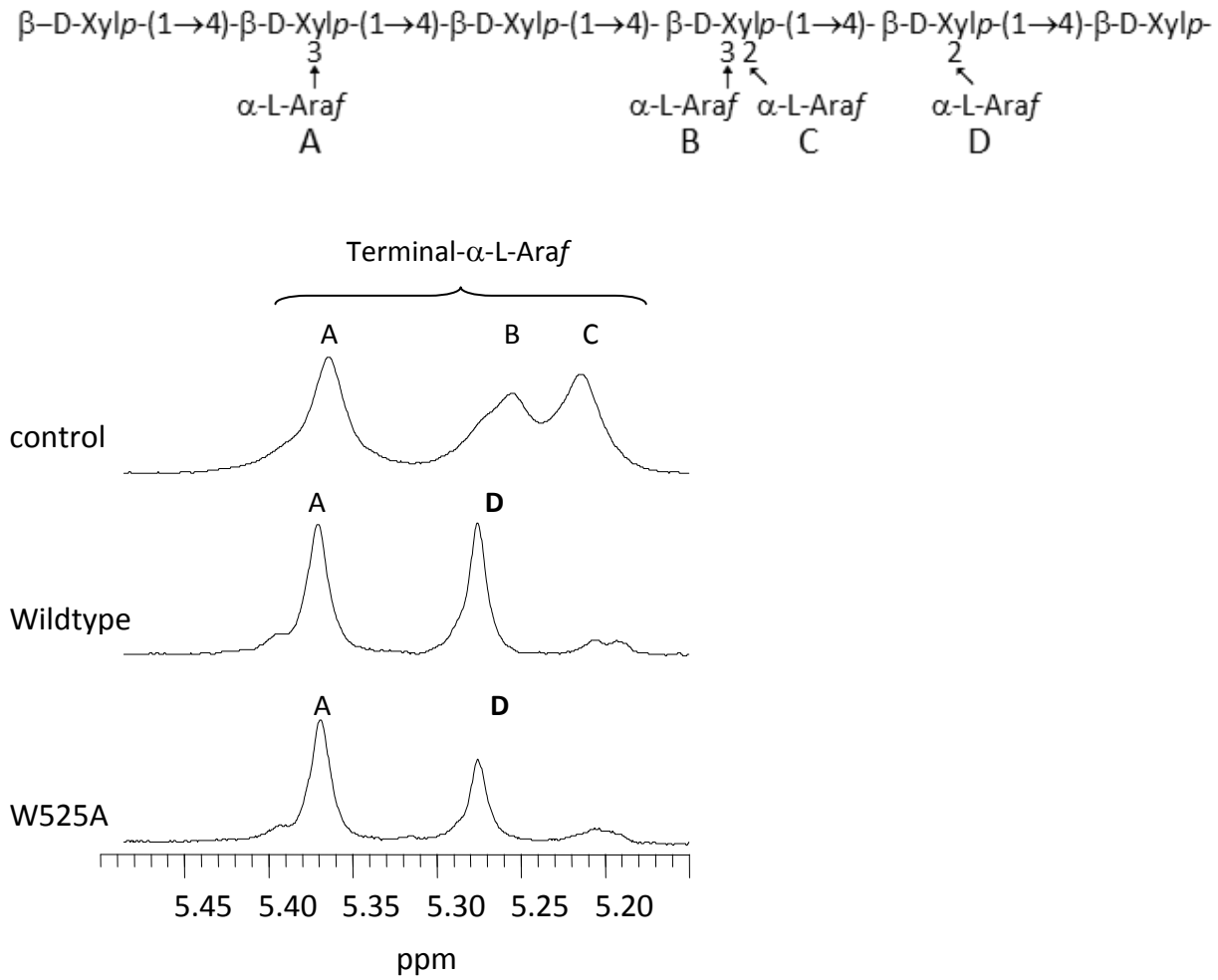


Figure 4.12 1D NMR of *Hi*AXHd3 wildtype and W525A variant wheat arabinoxylan reaction products.

In the schematic, linkages within the polysaccharide (A-D) are labelled to correspond with peaks in the NMR spectra. The spectra represent assays performed using ethanol precipitated polysaccharide (Figure 4.4). Peaks corresponding to the anomeric protons of α-L-Ara*f* residues C and B are present in the spectrum of arabinoxylan prior to enzyme treatment, but are absent after incubation. A peak corresponding to residue D is present only after treatment with *Hi*AXHd3.

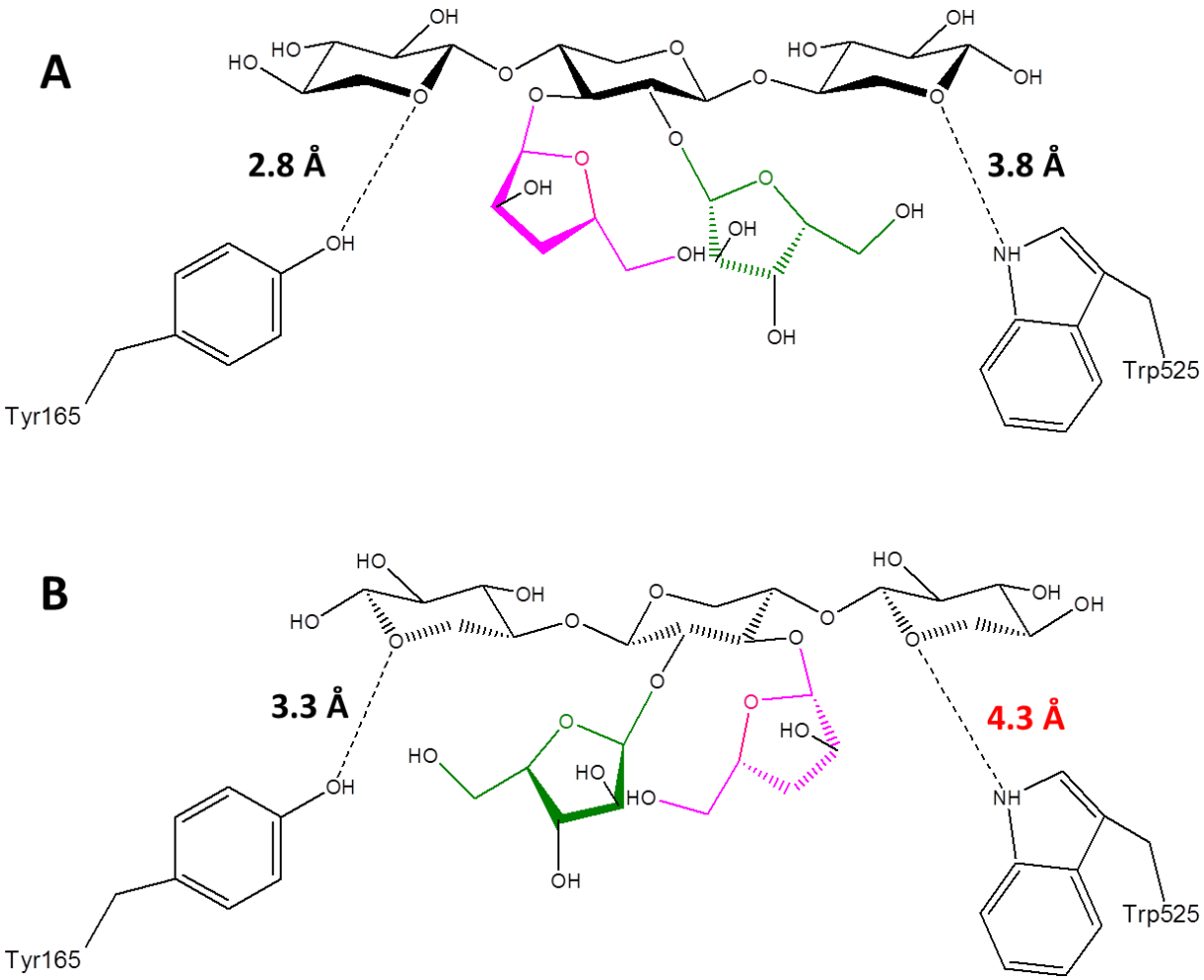


Figure 4.13 Trp525 and Tyr165 hydrogen bond to the xylan backbone.

A: when the substrate is in the correct orientation to bring the O3-linked arabinose of the double substitution into the active site, both Trp525 and Tyr165 are within range to make a hydrogen bond with the endocyclic oxygen of a xylose residue.

B: when the ligand is flipped horizontally so that the O2-linked arabinose points into the active site, Trp525 is no longer sufficiently close to hydrogen bond to the substrate.

	Subsite	Wheat arabinoxylan			Sugar beet arabinan		
		k_{cat} (min^{-1})	K_M (M)	k_{cat}/K_M (min^{-1} M^{-1})	k_{cat} (min^{-1})	K_M (M)	k_{cat}/K_M (min^{-1} M^{-1})
Wildtype		$1.0 \times 10^{4} \pm 1.3 \times 10^{3}$	3.5×10^{-4} 1.4×10^{-4}	2.9×10^{7}	$1.7 \times 10^{3} \pm 1.1 \times 10^{2}$	$1.3 \times 10^{-4} \pm 2.5 \times 10^{-5}$	1.4×10^{7}
D42A	-1	Inactive			Inactive		
D166A		Inactive			Inactive		
E215A		Inactive			Inactive		
Q272A		Inactive			Inactive		
R296A		Inactive			Inactive		
H271A	+2NR*	11.0 ± 2.4	$3.9 \times 10^{-4} \pm 2.7 \times 10^{-4}$	2.9×10^{4}	3.2 ± 1.0	$2.8 \times 10^{-4} \pm 3.2 \times 10^{-5}$	1.2×10^{4}
F288A		25.7 ± 2.7	$1.9 \times 10^{-4} \pm 6.7 \times 10^{-5}$	1.3×10^{5}	15.5 ± 2.4	$1.1 \times 10^{-4} \pm 7.7 \times 10^{-5}$	1.4×10^{5}
T231A		17.7 ± 2.8	$8.1 \times 10^{-5} \pm 5.3 \times 10^{-5}$	2.2×10^{5}	4.7 ± 1.1	$2.3 \times 10^{-5} \pm 3.7 \times 10^{-6}$	2.0×10^{5}
D290A		Inactive			Inactive		
W525A	b.b.	25.4 ± 6.6	$1.9 \times 10^{-4} \pm 1.7 \times 10^{-4}$	1.4×10^{5}	40.4 ± 4.8	$1.2 \times 10^{-3} \pm 3.6 \times 10^{-4}$	3.4×10^{4}
Y165A		38.5 ± 8.0	$2.1 \times 10^{-4} \pm 2.0 \times 10^{-4}$	1.8×10^{5}	3.0 ± 0.8	$6.2 \times 10^{-5} \pm 5.7 \times 10^{-6}$	4.8×10^{4}
Y165F		n.d.	n.d.	$2.87 \times 10^{3} \pm 9.16 \times 10^{2}$	Inactive		
Y165W		16.6 ± 2.5	$2.6 \times 10^{-5} \pm 4.1 \times 10^{-6}$	1.1×10^{6}	n.d.	n.d.	$2.07 \times 10^{3} \pm 3.4 \times 10^{2}$

Table 4.4 Alanine mutants of *Hi*AXHd3 were tested by continuous arabinose release assay.

n.d. = not determined due to high K_M. b.b. = backbone.

4.5 Engineering novel specificity

NMR analysis of the *Hi*AXHd3 mutant Y165A against wheat arabinoxylan confirmed retention of wildtype arabinofuranosidase specificity, albeit at reduced catalytic efficiency, but also revealed an abundance of peaks corresponding to reducing end xylose, which had not been observed with the wildtype enzyme. This observation prompted further analysis of the mutant. Initial TLC showed that the mutant released arabinose from arabinoxylan, as expected, but also generated a range of oligosaccharides from arabinoxylan (Appendix D). Incubation with sugar beet arabinan produced arabinose as the only product, and incubation with linear arabinan yielded no products, indicating that the Y165A mutation introduces the capacity for *endo*-type xylanase activity, but not arabinanase activity (Appendix D).

To explore in greater detail how this single mutation has influenced the activity of the enzyme, the reaction products generated from wheat arabinoxylan were analysed by HPLC (Figure 4.14). Following the release of arabinose from the double substitution, several oligosaccharide products are produced during incubation. To explore whether the newly introduced xylanase activity is dependent upon the arabinose substitutions, the capacity of Y165A to hydrolyse birchwood xylan was assessed. This polysaccharide contains very limited arabinose substitution, and the lack of activity of wildtype *Hi*AXHd3 against this substrate, as shown by spectroscopic assay, shows that the double substitution is essentially absent. The data, Figure 4.18, revealed a large number of reaction products that were identified as xylooligosaccharides based on their co-migration with appropriate standards. A small amount of arabinose was observed by HPLC for the mutant and wildtype (not shown) enzymes, but this was significantly reduced compared to arabinose release from arabinoxylan (Figures 4.17 and 4.18). The xylooligosaccharides had a degree of polymerisation (d.p.) of three or greater, although some xylobiose was observed as the reaction neared completion. The capacity of Y165A to hydrolyse xylooligosaccharides was investigated by HPLC. TLC experiments (Appendix D) and subsequent HPLC showed that while the enzyme variant displayed no activity against oligosaccharides with a d.p. of 2-4 (xylobiose, xylotriose and xylotetraose), xylopentaose was converted to xylotriose and xylobiose (Figure 4.15), while xylohexaose was cleaved exclusively into xylotriose (Figure 4.16). This indicates that

the enzyme has six subsites for xylan, of which five are critical. These HPLC data were used to calculate the catalytic efficiency of Y165A using substrate depletion (Table 4.5). For determination of depletion of substrate, enzyme was incubated with substrate (xylopentaose or xylohexaose) and samples were taken throughout the duration of the reaction. Each sample was analysed by HPLC and peak area of the substrate was recorded. A progress curve was generated for each reaction by plotting the amount of product remaining (peak area) over the time measured. A linear rate was obtained. This method assumes that the substrate concentration is below the K_M of the enzyme; this was proven by performing the experiment with each substrate at three concentrations (0 μM, 30 μM and 60 μM) to ensure a true linear relationship between rate of hydrolysis and substrate concentration.

Degradation of birchwood xylan by Y165A was assessed kinetically by reducing sugar assays (DNSA, see Chapter Two Section 2.3.4) using a standard curve of xylose. Table 4.5 summarises kinetic parameters for all assays undertaken for the Y165A variant. These data demonstrate that the Y165A mutation confers xylanase activity on *Hi*AXHd3, while the enzyme retains its arabinofuranosidase function, albeit at a significantly reduced level.

	k_{cat}	K_M	k_{cat}/K_M
Wheat arabinoxylan arabinose release	38.5 ± 3.1 min^{-1}	2.1 x 10^{-4} M ± 7.5 x 10^{-5} M	1.8 x 10^{5} min^{-1} M^{-1}
Sugar beet arabinan arabinose release	2.9 ± 0.4 min^{-1}	6.1 x 10^{-5} ± 2.5 x 10^{-5} M	4.8 x 10^{4} min^{-1} M^{-1}
Birchwood xylan reducing sugar	15.8 ± 3.7 min^{-1}	3.5 ± 2.7 mg ml^{-1}	4.5 min^{-1} mg^{-1} ml
Xylohexaose substrate depletion	n.d.	n.d.	3.5 x 10^{4} min^{-1} M^{-1}
Xylopentaose substrate depletion	n.d.	n.d.	1.3 x 10^{3} min^{-1} M^{-1}

Table 4.5 Kinetic analysis of Y165A.

*Hi*AXHd3 variant Y165A was analysed by assays for continuous arabinose release, reducing sugar release and depletion of oligosaccharide substrate. n.d. = not determined

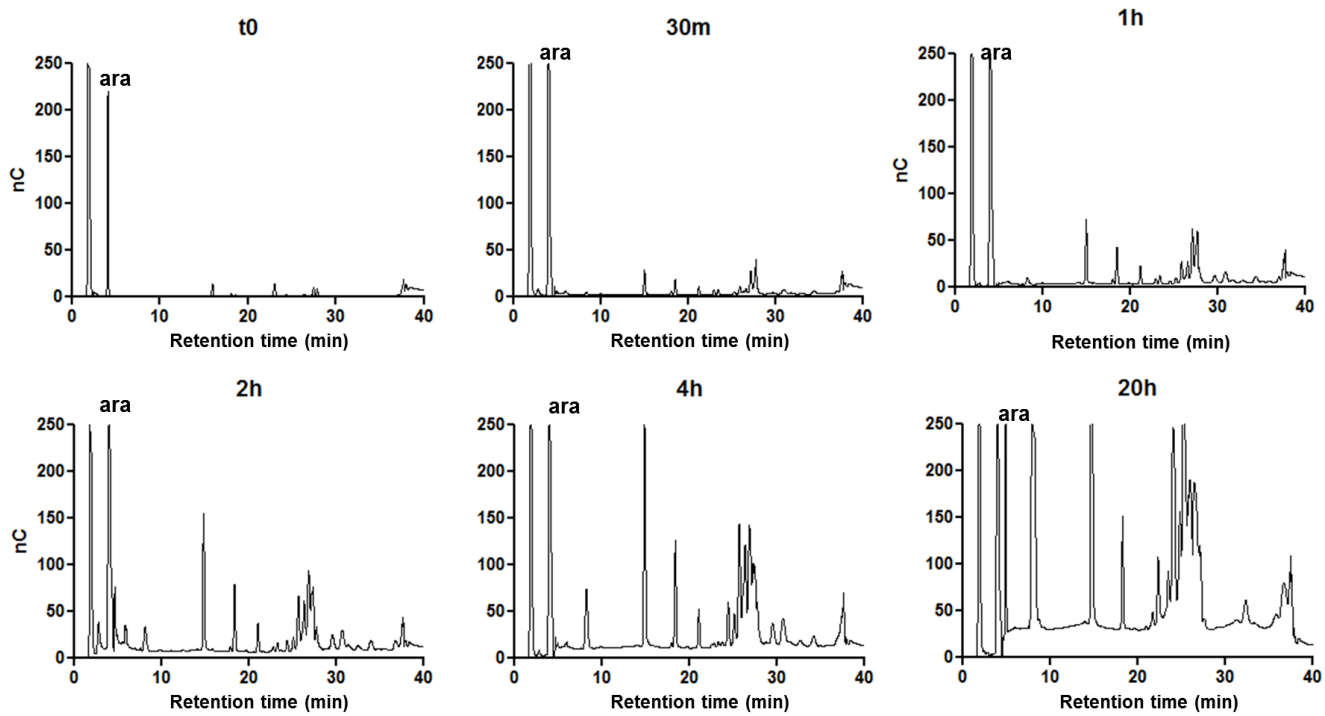


Figure 4.14 Reaction products generated by the *Hi*AXHd3 Y165A variant from wheat arabinoxylan were analysed by HPLC.

Product profile demonstrated in time points taken over 20 hours (overnight incubation). The primary product visible in early time points is arabinose (ara). Other products increase during incubation. Due to the decorated nature of the polysaccharide substrate, appropriate standards were not available for this experiment, so peaks could not be accurately identified. The arabinose peak is much larger than indicated in this figure (~ 1000 nC), as the spectra were cut off at 250 nC to improve resolution of other peaks.

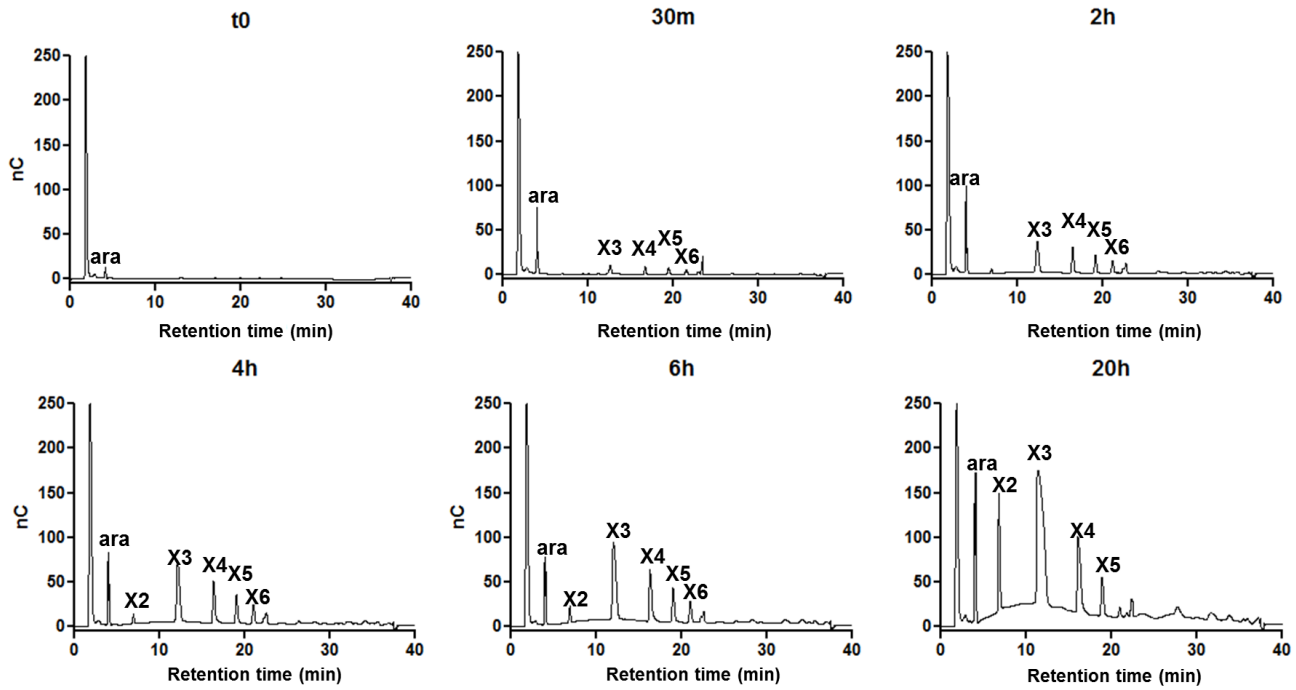


Figure 4.15 Reaction products generated by the *Hi*AXHd3 Y165A variant from birchwood xylan were analysed by HPLC.

Product profile is demonstrated by time points taken over 20 hours (overnight incubation). Very little arabinose is produced during incubation due to the rarity of the double substitution in this polysaccharide. Reaction products were identified by co-migration with xylooligosaccharide standards (Megazyme). Peak labels X2-X6 correspond to xylobiose, xylotriose, xylotetraose, xylopentaose and xylohexaose, respectively. Oligosaccharides of degree of polymerisation (d.p.) 3-6 increase over time. After four hours, xylotriose (X3) is the dominant product and a small peak corresponding to xylobiose (X2) is observed. After overnight (o/n) incubation, X2 and X3 are the dominant products. A peak for arabinose is present, but is significantly smaller than observed for wheat arabinoxylan (Figure 4.14).

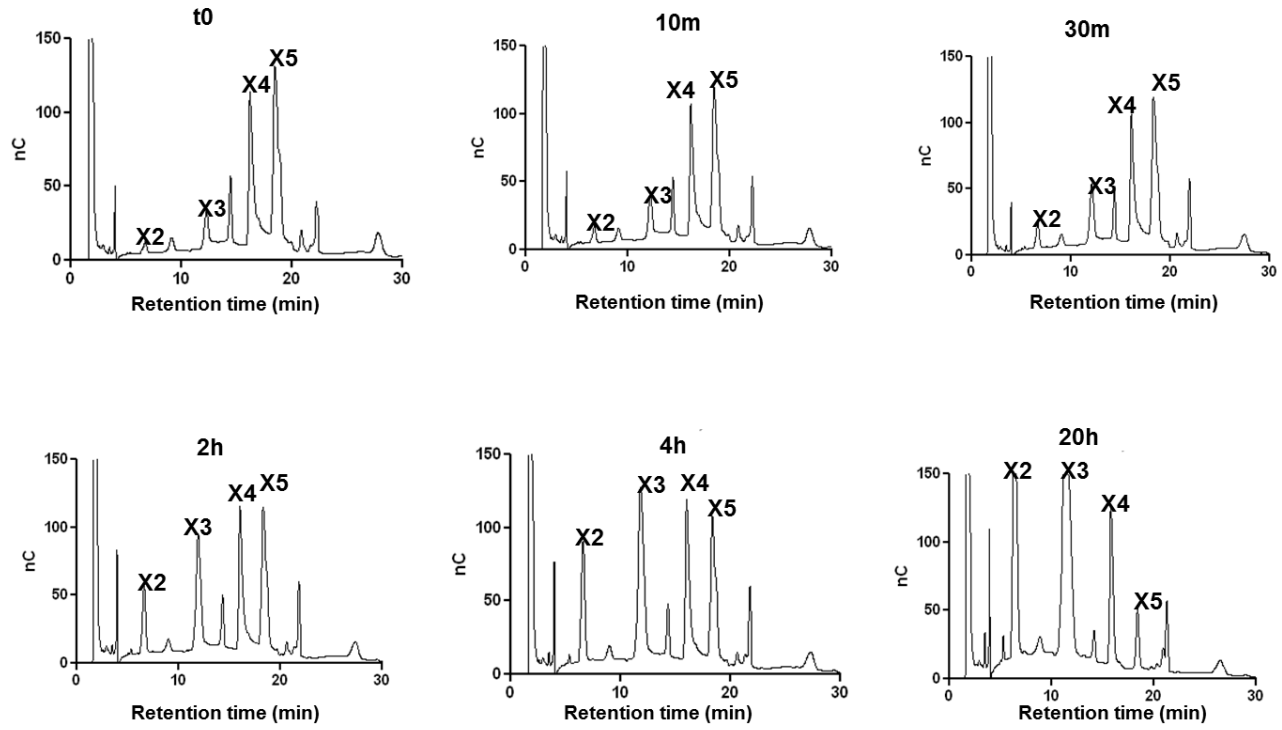


Figure 4.16 Reaction products generated from xylopentaose by Y165A were analysed by HPLC.

Xylopentaose (X5) (Megazyme) was contaminated with smaller xylooligosaccharides. X5 was incubated with Y165A for up to 20 hours with time points taken throughout. There is no observed change to the xylotetraose (X4) peak, a decrease in the X5 peak and increases to the xylotriose (X3) and xylobiose (X2) peaks, indicating that Y165A cleaves X5 into X3 and X2. Substrate at 30 μM was incubated with enzyme at 10 μM. This experiment was also performed with substrate at 60 μM (not shown).

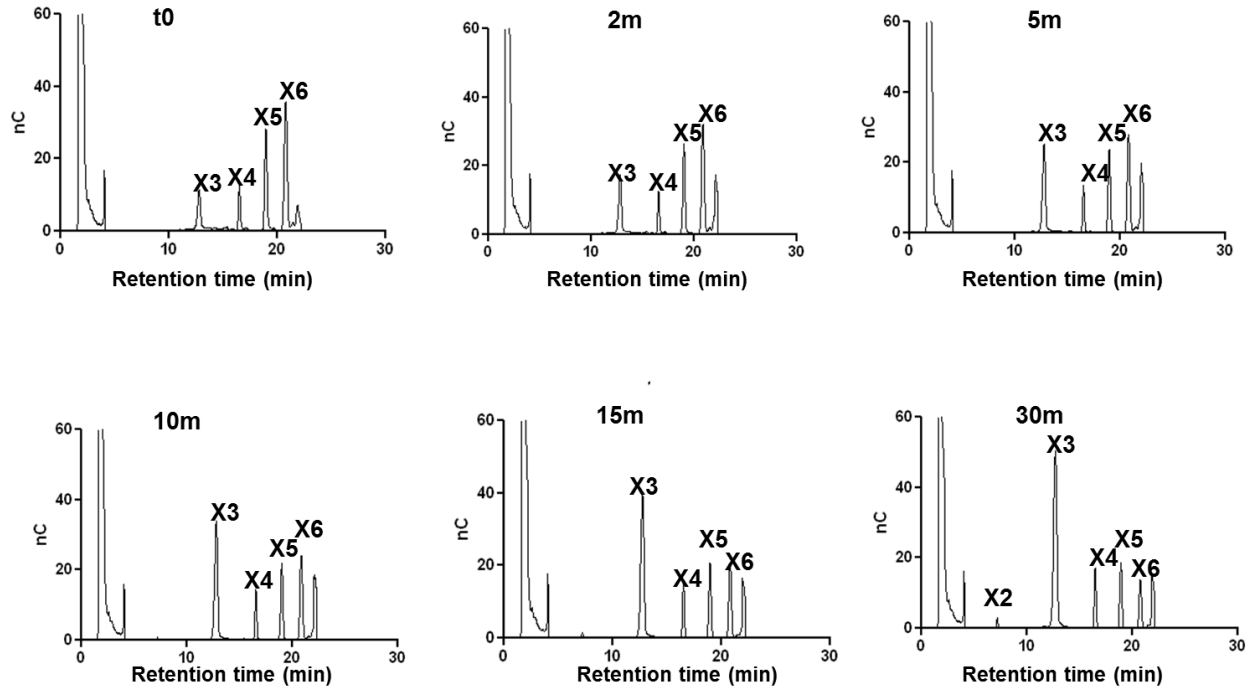


Figure 4.17 Reaction products generated from xylohexaose by Y165A were analysed by HPLC.

Xylohexaose (X6) (Megazyme) was consistently found to be contaminated with smaller xylooligosaccharides. Y165A was incubated with X6 for 30 minutes, with time points taken throughout. Analysis of the peak areas shows no change to the xylotetraose (X4) peak and, towards the end of the reaction, a very slight decrease in the xylopentaose (X5) peak. The xylohexaose (X6) peak decreases while the xylotriose (X3) peak increases. This indicates the Y165A cleaves X6 into two molecules of X3. The slight decrease in the X5 peak is matched by the appearance of a very tiny peak in a region corresponding to xylobiose. This indicates the contaminating X5 is again being cleaved into X3 and X2. Substrate at 30 μM was incubated with enzyme at 10 μM. This experiment was also performed with substrate at 60 μM (not shown).

4.6 Structure of the Y165A mutant

To fully understand the nature of the newly engineered xylanase activity of the Y165A variant of *Hi*AXHd3, a structure was sought in complex with an appropriate ligand. Inactive forms of the enzyme were generated by individually mutating Asp42, Asp166 and Glu215 to Alanine. These mutants were shown to be inactive by continuous arabinose release assay with wheat arabinoxylan and by TLC analysis of reaction with birchwood xylan. These data also confirm that the original active site of

wildtype *Hi*AXHd3 catalyses the hydrolysis of both the xylan backbone and arabinose side chains in Y165A.

Each mutant was expressed in *E. coli* Tuner cells and purified first by IMAC and then by gel filtration to increase homogeneity. After buffer exchange into water, the mutants Y165A/D42A, Y165A/D166A and Y165A/E215A were concentrated to 25, 15 and 20 mg ml^{-1}, respectively. Each mutant was placed in the full 96-condition PACT screen and in addition was set up in larger volume in six conditions which gave good crystals with the His-tagged selenomethionine wildtype protein (PACT 39, 40, 53, 29, 30 and 24). This second screen was set up in duplicate, in the absence and presence of 10 mM xylotetraose. As with the wildtype, the presence of ligand in crystallising conditions prevented crystal growth, but Y165A/D42A showed good crystal growth in many PACT conditions, including several which showed a cubic morphology not observed for the wildtype protein. Optimum crystal growth was observed in 20 % PEG-6000, 0.1 M MES buffer (pH 6.0) with 0.01M zinc chloride and 20 % PEG-3350, 0.1 M Bis-Tris propane (pH 8.5) with 0.2 M sodium fluoride (Figure 4.18). After two further weeks of incubation, Y165A/E215A showed good growth of crystals in both the cubic form and the original needle form of the wildtype protein in this second condition, but these crystals did not diffract.

A cubic crystal of Y165A/D42A from the PEG-3350 condition was taken to the Ser-CAT synchrotron (South-East Regional Collaborative Access Team, operated by the University of Georgia) and data was collected at 1.8 Å. The data indexed to $P2_1$, as did the wildtype crystals. Using the selenomethionine wildtype structure as the search model, the structure was solved (Figure 4.19, see experimental statistics in Table 4.7). Strong density for a single molecule of the buffer Bis-Tris propane was observed in the active site cleft. As shown in Figure 4.19, the loss of the side chain of Tyr165 has transformed the tight ‘heart-shaped’ pocket of the wildtype protein into an open cleft, which can accommodate a xylan chain. The Bis-Tris propane molecule occupies the active site and points along the cleft, nestled within the active site architecture.

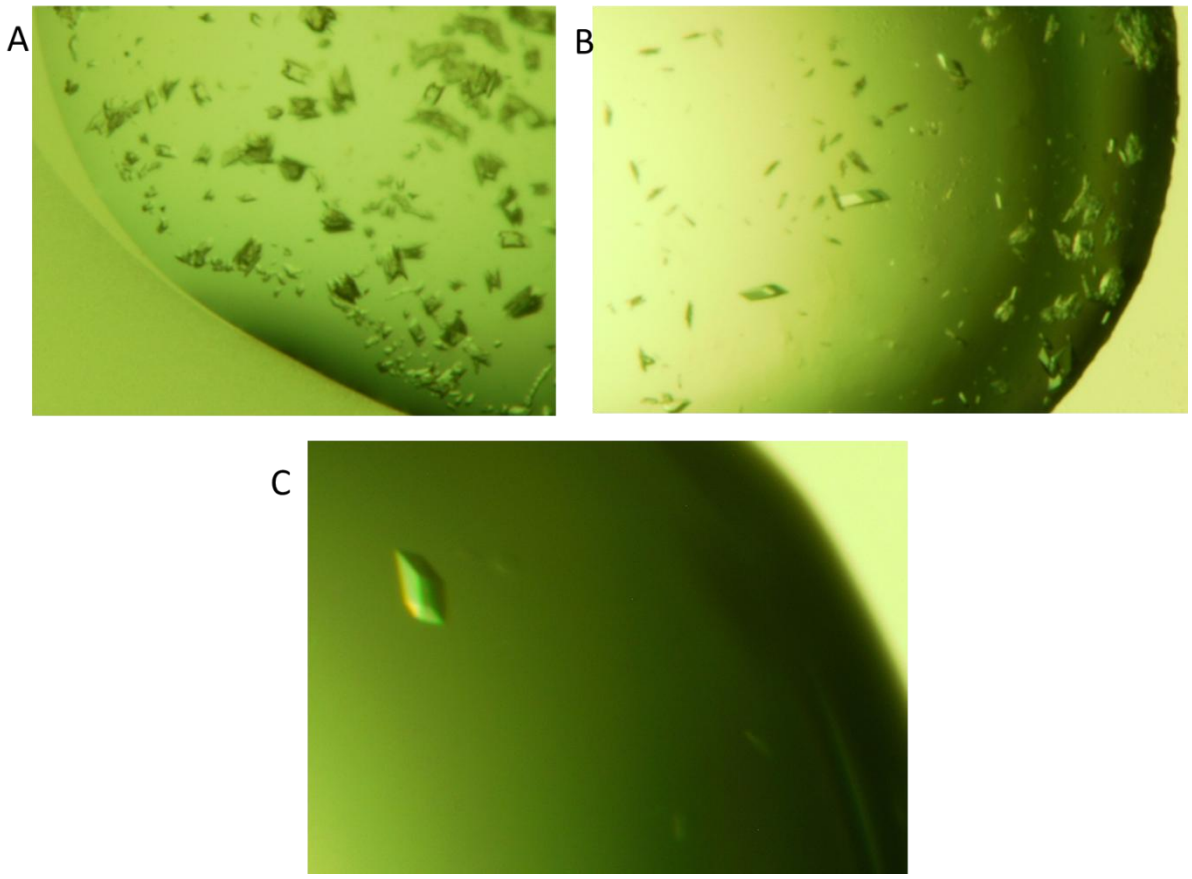


Figure 4.18 Cubic crystals of the *Hi*AXHd3 Y165A variant during progressive stages of optimisation.

Growth conditions for the crystals shown were as follows.

A: 0.01 M zinc chloride, 0.1 M MES pH 6.0, 20 % PEG-6000.

B: 0.2 M sodium fluoride, 0.1 M Bis-Tris propane pH 8.5, 20 % PEG-3350.

C: 0.3 M sodium fluoride, 0.1 M Bis-Tris propane pH 6.0, 27 % PEG-3350.

To improve our understanding of the novel activity of Y165A, structures were sought with substrate bound to the enzyme variant. Crystals of Y165A/D42A and Y165A/E215A were soaked in a variety of ligands. A few milligrams of ligand were dissolved in 10 µl of appropriate mother liquor, and this was further diluted 1 in 10 in additional mother liquor. Crystals were fished into a drop of this final solution and incubated for around ten hours at 20 °C. After soaking, crystals were transferred to a cryoprotectant mother liquor containing appropriate ligand and frozen. In total, two crystals were obtained soaked with xylohexaose, four with xylotetraose, three with

AXOS (generated in Section 4.2) and three with arabinohexaose. All were taken to the synchrotron. Both xylohexaose crystals were damaged or were split crystals, but full datasets were collected for at least one crystal for the other ligands. Crystals of the Y165A/E215A mutant did not diffract. Structures were solved by molecular replacement, using the earlier structure of Y165A as the search model, but none contained ligand. All crystals grown in the presence of Bis-Tris propane had strong density for the buffer in their active site.

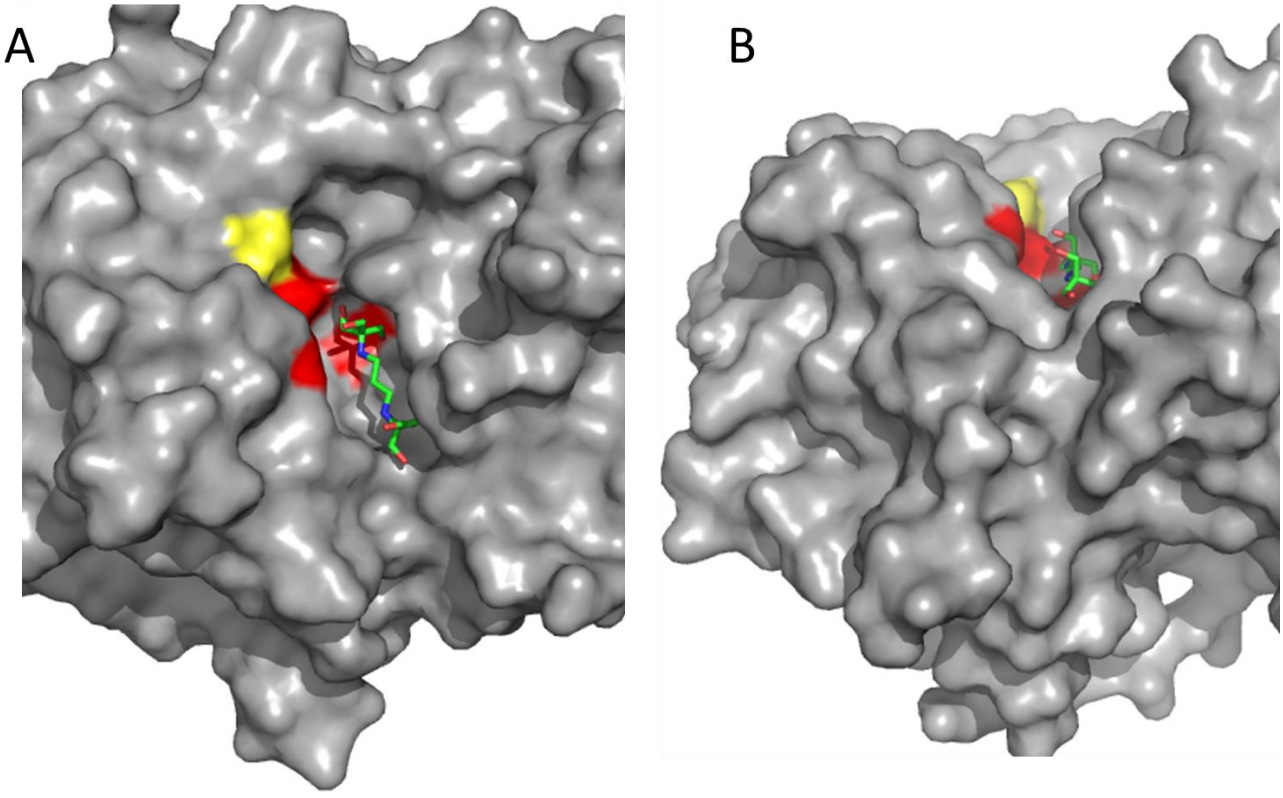


Figure 4.19 Surface representations of the structure of *Hi*AXHd3 variant Y165A in complex with Bis-Tris propane (BTP).

A: view from directly above the cleft, showing the active site and newly opened cleft above it. Catalytic residues are shown in red while the mutated residue Ala165 is in yellow. A molecule of Tris is shown in green in the active site.

B: view from the side of the protein, looking into the cleft towards the active site.

4.7 Further mutagenesis of the Y165A mutant

Based on analysis of the structure of Y165A, several residues were identified as being potentially significant to the newly identified xylanase activity. The residues His271 and Phe288 were previously shown to be important in wildtype activity so were again mutated to Alanine in the Y165A variant to assess their significance for the newly acquired xylanase activity. His271 makes polar contact with the O2 arabinose at the +2NR* subsite, while Phe288 lines the O2 binding pocket. In addition, four residues (Asn183, Pro233, Phe269 and Phe482) were selected for mutation as they appear to form part of the wall of the cleft. Figure 4.20 illustrates the locations of these residues. Each was mutated in turn to Alanine, and the effects of these mutations was assessed by continuous arabinose release assay, HPLC analysis against wheat arabinoxylan and birchwood xylan, and reducing sugar release from birchwood xylan. Table 4.6 shows kinetic parameters for these double mutants. Comparing the double mutants with the Y165A variant, Y165A/F269A had the least significant impact on any activity. Y165A/F288A had a moderate impact on all activities, as did Y165A/P233A. The double mutant Y165A/H271A showed no measurable arabinofuranosidase activity against arabinan or arabinoxylan, but retained xylanase levels similar to the Y165A variant. The mutants Y165A/N183A and Y165A/F492A both led to significant increases in k_{cat} for the xylanase activity, with Y165A/F492A also improving arabinofuranosidase activity on both arabinoxylan and arabinan.

It should be noted that when studying Y165A, and other mutants containing this mutation, xylanase activity was only observed when these proteins were purified using Nickel columns that had gone through the guanidine EDTA regeneration process.

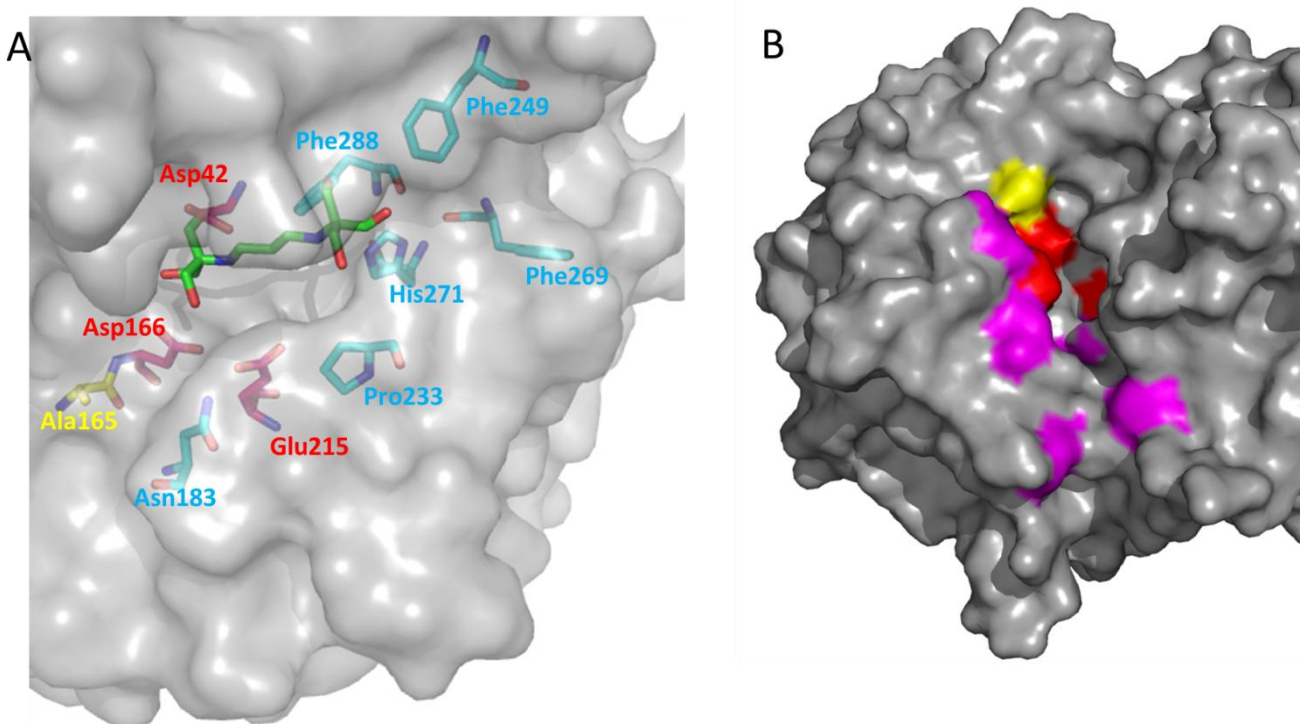


Figure 4.20 Structure representation of *Hi*AXHd3 variant Y165A indicating those residues selected for mutation.

A: a transparent rendering of the active site of Y165A shows those residues close to the ligand binding site which were selected for mutagenesis. Bis-Tris propane is shown in green. The catalytic residues are shown in red, Ala165 is yellow, and the residues which were mutated are shown in cyan.

B: surface representation showing the xylan binding cleft of Y165A in the context of the whole enzyme. The walls of the cleft are marked out by the residues shown in purple. Ala165 is shown in yellow and the catalytic residues are shown in red.

	Wheat arabinoxylan arabinose release			Sugar beet arabinan arabinose release			Birchwood xylan reducing sugar release		
	k_{cat} (min^{-1})	K_M (M)	k_{cat}/K_M (min^{-1} M^{-1})	k_{cat} (min^{-1})	K_M (M)	k_{cat}/K_M (min^{-1} M^{-1})	k_{cat} (min^{-1})	K_M (mg ml^{-1})	k_{cat}/K_M (min^{-1} mg^{-1} ml)
Wildtype	$1.0 \times 10^{4} \pm 1.3 \times 10^{3}$	$3.5 \times 10^{-4} \pm 1.4 \times 10^{-4}$	2.9×10^{7}	$1.7 \times 10^{3} \pm 1.1 \times 10^{2}$	$1.3 \times 10^{-4} \pm 2.6 \times 10^{-5}$	1.4×10^{7}	-	-	-
Y165A	38.4 ± 3.1	$2.1 \times 10^{-4} \pm 1.9 \times 10^{-4}$	1.8×10^{5}	2.9 ± 0.8	$6.2 \times 10^{-5} \pm 5.7 \times 10^{-6}$	4.8×10^{4}	15.8 ± 3.7	3.5 ± 2.7	4.5
Y165A /N183A	26.3 ± 1.5	$9.9 \times 10^{-5} \pm 2.9 \times 10^{-5}$	2.6×10^{5}	12.5 ± 2.0	$7.1 \times 10^{-4} \pm 3.2 \times 10^{-4}$	1.8×10^{4}	78.4 ± 15.8	2.2 ± 1.7	35.5
Y165A /H271A	-	-	-	-	-	-	11.6 ± 2.6	0.9 ± 0.7	13.5
Y165A /P233A	5.0 ± 0.8	$3.0 \times 10^{-4} \pm 1.7 \times 10^{-4}$	1.6×10^{4}	1.7 ± 0.3	$8.3 \times 10^{-4} \pm 4.0 \times 10^{-4}$	2.0×10^{3}	30.2 ± 8.2	6.8 ± 4.7	4.4
Y165A /F492A	121.5 ± 40.3	$4.9 \times 10^{-4} \pm 2.4 \times 10^{-4}$	2.5×10^{5}	50.4 ± 10.6	$9.4 \times 10^{-4} \pm 4.5 \times 10^{-4}$	5.4×10^{4}	98.0 ± 20.8	5.4 ± 3.3	18.1
Y165A /F288A	11.8 ± 1.8	$1.9 \times 10^{-3} \pm 7.4 \times 10^{-4}$	6.3×10^{3}	2.2 ±0.7	$7.2 \times 10^{-4} \pm 5.8 \times 10^{-4}$	3.1×10^{3}	26.8 ± 4.5	4.0 ± 2.1	6.7
Y165A /F269A	16.6 ± 1.7	$2.7 \times 10^{-4} \pm 9.3 \times 10^{-5}$	6.1×10^{4}	2.2 ± 0.2	$2.3 \times 10^{-4} \pm 6.2 \times 10^{-5}$	9.4×10^{3}	16.8 ± 4.0	4.2 ± 1.8	4.0

Table 4.6 Kinetic analysis of double mutants of *Hi*AXHd3.

Mutants of *Hi*AXHd3 variant Y165A were analysed by continuous arabinose release assay and reducing sugar assay. Kinetic parameters for wildtype and Y165A are included for comparison. – indicates no measurable activity.

		Y165A Bis-Tris propane
	Space group	$P2_1$
	Wavelength	1.00
	Unit cell	a = 65.3
		b = 77.9
		c = 95.2
		α = 90.0
		β = 102.7
		γ = 90.0
	Resolution (Å)	92.8-1.8
	No. observations	195734 (35142)
	No. unique observations	36699 (5137)
	Multiplicity	6.4 (6.6)
	Completeness	99.6 (99.4)
	Average I/sigma I	21.4 (9.8)
	R_{sym}	0.092 (0.174)
Refinement		
	No. reflections	84542
	R_{cryst}	15.74 %
	R_{free}	20.84 %
	No. atoms	9022
Rmsd		
	Bond lengths	0.024
	Bond angles	2.027
Ramachandran		
	Favoured	94.80 %
	Allowed	99.8 %
B-factors		
	Wilson B	15.24
	Main chain	11.07
	Side chains and waters	31.79

Table 4.7 Experimental statistics for the solution and refinement of the crystal structure of the *Hi*AXHd3 Y165A variant in complex with Bis-Tris propane.

4.8 Discussion

This chapter describes the crystal structure and function of an unusual arabinofuranosidase that displays absolute specificity for the O3 linked arabinose side chain of doubly substituted (O2 and O3) backbone xylopyranose (in xylan) or arabinofuranose (in arabinan) residues. Table 4.2 gives kinetic parameters for the wildtype activity of the enzyme against wheat arabinoxylan and sugar beet arabinan. The catalytic efficiency (k_{cat}/K_M) for arabinoxylan is ~ 3.7 times that for arabinan. Thus, the enzyme would seem to be tailored for arabinoxylan. While the 'core substrate' of the double substitution is highly similar in both substrates, the backbones are very different (Figure 4.1). Arabinoxylan adopts a largely linear conformation while arabinan displays a helical structure (Janaswamy and Chandrasekaran, 2005; Yui et al., 1995). These differences may be significant to the disparity in rates against these two substrates due to a relative lack of contacts between the arabinan backbone and the enzyme.

4.8.1 Specificity of wildtype *Hi*AXHd3

Analysis of reaction products by NMR showed that the wildtype enzyme shows the same specificity for the O3 arabinose attached to doubly substituted backbone residues in both arabinoxylan and arabinan. As arabinofuranose side chains can rotate around the glycosidic linkage, and the exocyclic C5 is also free to rotate, it is difficult to understand how the enzyme is able to select a specific linkage within the double substitution, a pseudosymmetrical target (Figure 4.1). It is possible that interactions with the backbone polysaccharide orient the O3 linked arabinofuranose towards the -1 subsite. Structures were sought of the enzyme in complex with substrate to shed light on this problem.

4.8.2 Structural analysis and mutagenesis of *Hi*AXHd3

The structures presented in this chapter show that, in contrast to the narrow active site pockets of *exo*-acting GH43 enzymes that target single substitutions, *Hi*AXHd3 features a much wider active site pocket containing a shelf like structure. Despite containing many of the residues that are conserved in the active site of GH43 arabinofuranosidases acting on monosubstitutions, *Hi*AXHd3 is highly specific for the

double substitution. It is possible that the spatially restricted active site of mono-acting GH43 arabinofuranosidases creates a micro environment that is optimised for binding and hydrolysis of arabinose side chains. In *Hi*AXHd3, however, the wider active site allows more rapid solvent exchange and limits the formation of such micro environments, causing the enzyme to rely on the additional binding energy provided by the O2 side chain.

The structures of the wildtype protein in complex with ligand revealed the nature of the tight specificity displayed by *Hi*AXHd3. These structures identified Arg296, His271 and Asp290 as the residues that interact with the O2 arabinose, while Trp525 and Tyr165 interact with the xylan backbone (Figure 4.21). As shown in the lower resolution ligand structure with arabinose at the -1 subsite, Arg296 also makes a polar contact with the O3 arabinose, which is appropriately positioned to make contacts with the catalytic apparatus of the enzyme.

In the E215C-ligand complex (*Hi*AXHd3/Ara*f*$_2$-Xyl) an arabinofuranose residue is located at the -1 subsite, in addition to a xylose at +1, which is linked through O2 to an arabinose positioned at the +2-NR* subsite. The structure, however, is of low resolution (2.7 Å) and the electron density for the -1 arabinose is weak (Figure 4.10). The arabinose is clearly in a furanose conformation and the exocyclic C5 is evident. As arabinose equilibrates towards the pyranose conformation in solution, it would appear that the -1 sugar is held in its furanose structure through interactions with the enzyme. This structure shows that the -1 arabinose makes several interactions with the enzyme: O2 makes hydrogen bonds with the backbone nitrogen of Ala108 and O□1 of Asp188, O3 forms polar contacts with O□2 of Asp64 and both N□1 and N□2 of Arg296, while the extensive interactions made by O5 with the protein include O□1 of Asp64, N□1 of Arg296 and both O□1 and N□2 of Gln272. The polar contacts are augmented with hydrophobic interactions, notably a highly conserved Tryptophan (Trp107 in *Hi*AXHd3), which provides a hydrophobic platform at the -1 subsite. The importance of the conserved active site residues in substrate binding is confirmed by mutagenesis studies. The mutants D290A, Q272A and R296A, in addition to mutants of the three catalytic amino acids, display no measurable activity, Table 4.3.

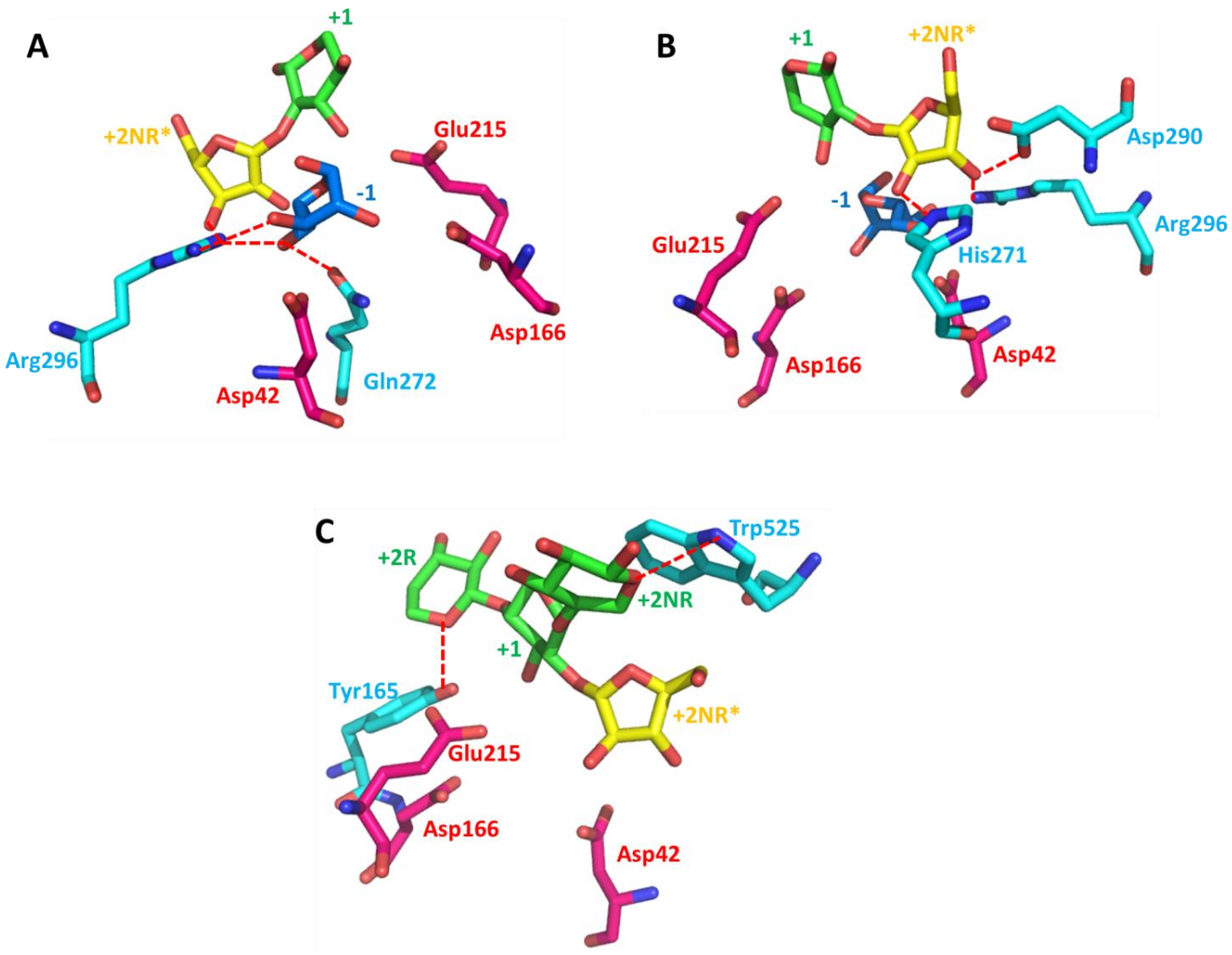


Figure 4.21 Polar contacts observed between *Hi*AXHd3 and its substrate.

Catalytic residues are shown in red in each figure, while other residues of interest are coloured cyan. Substrate contacts with the catalytic residues are not shown.

A: -1 subsite. Arg296 makes contacts with both the -1 and +2NR* arabinose residues, while Gln272 contacts the active site (-1) arabinose.

B: +2NR* subsite. His271, Arg296 and Asp290 make polar contacts with the O2 linked arabinose.

C: Tyr165 and Trp525 coordinate the substrate via interactions with the backbone. Tyr165 makes a direct polar contact, while the orientation of Trp525 indicates a stacking interaction with the +2NR xylose and its distance from the substrate suggests a hydrogen bond with the endocyclic oxygen.

The sugar at the -1 site in this structure does not appear to display the envelope (*E)* conformation typical of arabinofuranose residues and is too distant from the +1 xylose to make a covalent bond, suggesting that the glycosidic bond was cleaved during crystal soaking. The sugar appears to be perpendicular to the arabinose and xylose residues seen in the -1 subsite of other GH43 enzymes (Brux et al., 2006; Fujimoto et al., 2010; Nurizzo et al., 2002; Vandermarliere et al., 2009). Nonetheless the -1 arabinose does make polar interactions with residues that are conserved across the GH43 landscape, and have been shown to play an important role in substrate recognition in two arabinanases, an arabinofuranosidase and a xylosidase (Brux et al., 2006; Fujimoto et al., 2010; Nurizzo et al., 2002; Proctor et al., 2005). Thus, the arabinose in the active site of *Hi*AXHd3 does appear to be in the correct conformation in this structure. The unexpected orientation of the sugar is likely due to the lack of a covalent bond with the -1 arabinose and the +1 xylose.

In the D42A mutant structure (*Hi*AXHd3/Ara*f*-Xyl_3) the xylotriose ligand makes relatively few interactions with the enzyme, which likely aids product departure, as the cleaved arabinose cannot be released from the active site until the xylose polymer dissociates from the enzyme. The absence of the active site arabinose in the crystal structure may reflect more on the nature of the ligand than on the activity of the enzyme. It was initially hypothesised that the arabinose is absent due to cleavage of the ligand during the prolonged duration of the crystal soak, but it may be that the ligand initially comprised only the monosubstitution, as the oligosaccharide was not characterised by methods such as mass spectrometry prior to use in crystallography. Multiple interactions with the O2 arabinose (discussed below) could be sufficient to bind such a monosubstituted ligand.

The structures show that Glu215 is in an ideal position to function as the catalytic acid, donating a proton to the scissile glycosidic oxygen, as the O3 of the +1 xylose makes a hydrogen bond with O□1 of the Glutamate. The elevated pKa required of a catalytic acid is afforded by a hydrogen bond with Asp166, and the Glutamate side chain also makes polar contacts with Tyr165 and Asn183, which may assist in orienting the residue. The hydrophobic environment of Glu215, provided by the methyl side chains of Thr113 and Thr231, and the aliphatic ring of Pro233, may also

encourage the Glutamate to retain its hydrogen. Indeed, mutation of Thr231 to Alanine leads to a reduction in k_{cat}/K_M of around 230.

The structures and mutagenesis studies presented in this chapter confirm that the O2-linked arabinofuranose is a critical specificity determinant. The structure of both ligand complexes reveals an arabinose nestled onto the shelf-like structure extending from the active site pocket. This sugar makes several polar contacts with the enzyme. Asp290 makes bidentate hydrogen bonds with O5 and O3, His271 makes a polar contact with O2, while the N□1 of Arg296 also makes a hydrogen bond with O3 (Figure 4.21). The extensive interactions between the O2 linked arabinose and *Hi*AXHd3 explain why the furanose sugar is a critical specificity determinant. The importance of these polar interactions is underlined by the observation that the mutants R296A and D290A are completely inactive, while the H271A mutation causes a 940-fold decrease in k_{cat}, although it had no obvious effect on K_M, Table 4.4.

In addition to contacts with the arabinose residues, the enzyme makes some direct contacts with the xylan backbone of the substrate. The xylose at the +2NR subsite makes a single interaction with the enzyme, a hydrogen bond between the endocyclic oxygen and the -OH of Tyr165. The xylose at the +2R subsite makes a hydrophobic contact with Trp525, which is a component of the C-terminal β-sandwich domain; a polar contact is formed between Nδ1 of the indole ring and the endocyclic oxygen of the +2R xylose.

The importance of Tyr165 and Trp525 in substrate recognition is underlined by the ~ 200-fold decrease in catalytic efficiency when the W525A or Y165A mutations are introduced into *Hi*AXHd3, Table 4.4. It should be noted however, that the two mutations effect k_{cat} but do not significantly alter K_M. This would suggest that Trp525 and Tyr165 play a more important role in binding the transition state than the substrate in its ground state conformation. The Y165A mutation, however, may also have a significant effect on the function of the catalytic acid, due to a loss of the interactions with Glu215 described above and this could contribute to the observed decrease in enzyme activity displayed by the *Hi*AXHd3 variant.

Mutation of Tyr165 to residues other than Alanine had unexpected effects on activity. The mutant Y165F was intended to eradicate contacts between the –OH group and the xylan backbone while preserving the pocket architecture; Y165W was also expected to preserve overall structure while disrupting specific contacts. Both mutations were therefore predicted to be less deleterious to activity than Y165A, which removes direct substrate contacts and disrupts the pocket architecture. However, k_{cat}/K_M for Y165W is around six times greater than for Y165A against wheat arabinoxylan, but is significantly lower than for the Alanine substitution against sugar beet arabinan. The large increase in K_M against arabinan (too high to determine) contributes to the decrease in activity of the Y165W mutant. It is likely that the introduced Tryptophan clashes with the arabinan backbone.

Conversely, Y165F displays a very high K_M for arabinoxylan, with a catalytic efficiency ~ 64-fold less than for Y165A, and displayed no detectable activity against sugar beet arabinan (Table 4.4). This would suggest that the ability of the Y165F mutant to interact with the polysaccharide backbone is impaired for the linear xylan and completely lost for the twisted arabinan chain. It is therefore likely that the OH of Tyr165 plays an important role in binding both arabinoxylan and arabinan.

4.8.3 Comparison of *Hi*AXHd3 structure with similar enzymes

The crystal structures of *Hi*AXHd3 presented in this chapter were compared with other GH43 structures. An overlay with the arabinanases *Cj*Arb43A (Nurizzo et al., 2002) and *Bs*Arb43A, and the arabinoxylan-specific arabinofuranosidase *Bs*AXHm2,3 (Vandermarliere et al., 2009) shows structural conservation of the catalytic residues (Figure 4.22). Structural comparison of *Hi*AXHd3 with the PDB database, using DaliLite v.3 (http://ekhidna.biocenter.helsinki.fi/dali_server/start), revealed that the closest structural homologue to the arabinofuranosidase was the GH43 □-xylosidase XynB3 from *Geobacillus stearothermophilus* ((Brux et al., 2006) PDB code 2exk) with a Z-score of 35.5, rmsd of 2.6 Å over 470 aligned Cα atoms and a sequence identity of 26 %. When using just the catalytic domain or the C-terminal β-sandwich domain as the search structure, the closest homolog was again the *Geobacillus* β-xylosidase. The closest structural homolog to *Hi*AXHd3 that displayed exclusively arabinofuranosidase activity was with the *Bacillus subtilis*

GH43 enzyme *Bs*AXHm2,3, an AXH enzyme that removes arabinose side chains exclusively from xylan ((Vandermarliere et al., 2009) PDB code 3c7e). The structural comparison of the two catalytic domains had a Z-score of 25.4, rmsd of 2.7 Å over 268 aligned Cα atoms and a sequence identity of 20 %.

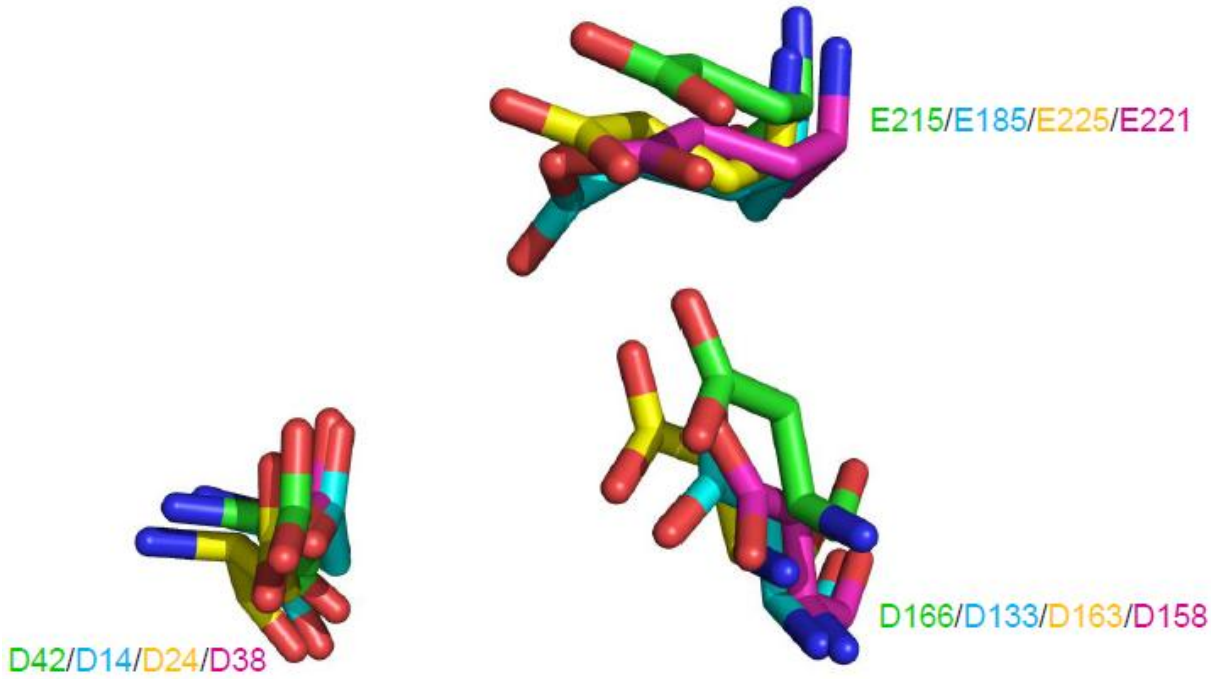


Figure 4.22 Overlay of catalytic residues in multiple crystal structures.

Structural overlay shows conservation of catalytic residues between *Hi*AXHd3 (green), *Bs*Arb43A (cyan), *Bs*AXHm2,3 (yellow) and *Cj*Arb43A (magenta).

4.8.3.i Structural conservation at the active site in arabinanases, xylosidases and arabinofuranosidase of family GH43

In addition to arabinofuranosidases GH43 also contains xylosidases, several of which have been shown to display limited arabinofuranosidase activity. By comparing the active site (-1 subsite) of the *Geobacillus* GH43 xylosidase, XynB3, with *endo* and *endo*-processive arabinanases, it has been suggested that four residues (His249, Arg288, Phe32 and Leu265) which are invariant in characterised GH43 xylosidases but are not conserved in the arabinanases, confer specificity for xylopyranose in preference to arabinofuranose at the -1 subsite. Recent structures including that of *Hi*AXHd3, however, show that all of the active site residues that interact with substrate in its ground state conformation are conserved in both

xylosidases and arabinofuranosidases, Figure 4.23. It is possible therefore, that specificity within the active site of these *exo*-acting enzymes is mediated through different interactions with the oxocarbenium transition state. In a GH51 arabinofuranosidase, the arabinose in the active site was shown to be distorted from its E_3 ground state conformation into a ^4E conformation at the transition state. In xylanases the active site xylopyranose residue is distorted from their relaxed $^4C_1$ conformation into either a $^4H_3$ or $^{2,5}B$ structure at the transition state, in GH10 and GH11 xylanases, respectively (Notenboom et al., 1998; Pell et al., 2004a; Sidhu et al., 1999). It is possible, therefore, that the active site of GH43 xylosidases and arabinofuranosidases may, initially, bind the ground state sugar through the same interactions, but then utilise additional interactions to distort the arabinofuranose and xylopyranose residues into their respective transition states.

Comparing *Hi*AXHd3 with the GenBank database showed that in the 50 proteins that display the closest sequence identity (e values of e^{-181} to e^{-60} with sequence identities varying from 60 to 31 %), the five key specificity determinants (Tyr165, His271, Asp290, Arg296 and Trp525) are highly conserved; Arg296 is invariant while Tyr165, His271, Asp290 and Trp525 are conserved in 94 %, 84 %, 92 % and 90 % of the proteins, respectively. It is likely that the majority, if not all, of these enzymes are AXHd3s, particularly when one considers that there is a lack of conservation of His271 and Trp525 in the only other confirmed AXHd3 (Lambertus et al., 2005). The recent report of the structure of an *exo*-α1,5-arabinofuranosidase also reveals a constricted deep active site pocket that could be described as a funnel (Fujimoto et al., 2010). The -1 subsite displays substantial conservation in the GH43 arabinofuranosidases, although in *Hi*AXHd3 a conserved Histidine is replaced with the critical Gln272. In fact, all the conserved residues interact with the arabinofuranose in the active site of *Hi*AXHd3, as described above.

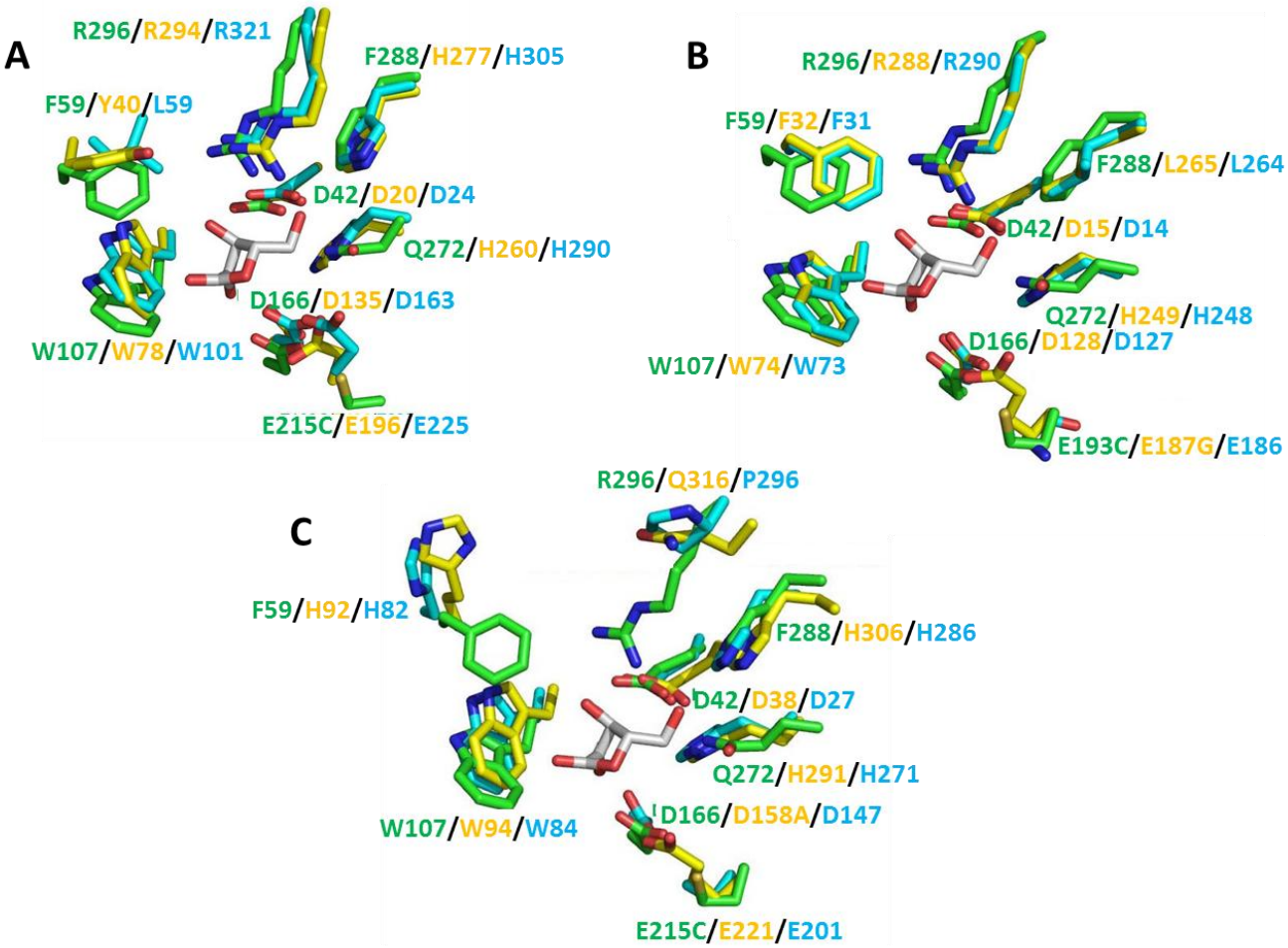


Figure 4.23 Overlays of active site residues of *Hi*AXHd3 and other enzymes show a high level of conservation.

A: an overlay of the active site of *Hi*AXHd3 (green) and other arabinofuranosidases reveals conservation of the active site residues. The arabinoxylan-specific arabinofuranosidase *Bs*AXHm2,3, which targets single substitutions, is shown in cyan (Vandermarliere et al., 2009). Yellow residues are derived from an *exo*-1,5-α-L-arabinofuranosidase from *Streptomyces avermitilis* (Fujimoto et al., 2010).

B: an overlay of *Hi*AXHd3 (green) with those of two xylosidase enzymes again reveals structural similarity at the active site. Residues in cyan are derived from a *Selenomonas ruminantium* β-xylosidase (Brunzelle et al., 2008). Yellow residues belong to a β-xylosidase from *Geobacillus stearothermophilus* (Brux et al., 2006).

C: an overlay of *Hi*AXHd3 (green) with the active sites of two arabinanase enzymes reveals some key structural differences. Yellow residues are derived from *Cj*Arb43A (Proctor et al., 2005), while residues in cyan are taken from the α-1,5-L-arabinanase *Gs*AbnB (Alhasid et al., 2009). The major

differences can be observed in the lack of alignment with *Hi*AXHd3 residues Phe37, Arg296 and Phe288.

Figure 4.23C highlights those residues which differentiate arabinofuranosidases and xylosidases from arabinanases. Arg296 has been shown to be critical for activity in *Hi*AXHd3, but it has no equivalent in the arabinanase enzymes shown in the figure. Likewise, the *Humicola* residues Phe288 and Phe59 do not align with the arabinanase active site. It seems likely that loss of these residues, particularly the binding energy conferred by Arg296, causes a different set of substrate interactions to be required in arabinanase enzymes to compensate for the loss of this energy. Thus, in arabinanase enzymes, the loss of substrate contacts in the active site has led to the development of a set of interactions with a more extensive region of the polysaccharide backbone, thereby introducing different specificity.

Given the degree of conservation in the -1 subsite of GH43 arabinofuranosidases, it is perhaps surprising that *Hi*AXHd3 displays no activity against single O3 linked arabinofuranose decorations. It is evident that the additional binding energy provided by the interaction of the O2 linked arabinose with several amino acids is critical for catalysis to occur. In the GH43 arabinofuranosidases that attack monosubstitutions, the binding pocket is much tighter than in *Hi*AXHd3. For example, in *Bs*AXHm2,3 the +2NR* pocket is occluded by Asn288, while in the arabinanase *Cj*Arb43A (Nurizzo et al., 2002) the extended loop that connects β-strands 2 and 3 in blade 4, which is much shorter in *Hi*AXHd3, would make steric clashes with the O2 linked arabinose. In the *Streptomyces avermitilis* arabinofuranosidase *Sa*Araf43A (Fujimoto et al, 2010; PDB code 3AKF) the side chain and backbone of Thr216 occlude the O2-linked arabinose binding site. It is possible that the tighter pocket observed in *Bs*AXHm2,3 and *Sa*Araf43A creates a microenvironment that is more favourable for substrate binding than in *Hi*AXHd3. It is also likely that there is less solvent exchange and thus the release of water molecules upon substrate binding may increase the affinity for arabinofuranose through enhanced entropic effects.

4.8.3.ii Structural features around the active site of GH43 enzymes confer preference for polysaccharide backbone structures

Validation of the structure of *Hi*AXHd3 presented here is provided by overlays with other GH43 enzymes. The *Bacillus subtilis* GH43 enzyme *Bs*AXHm2,3 that removes arabinose side chains exclusively from xylan (Vandermarliere et al., 2009, PDB code 3c7e) was particularly useful in this regard. Comparisons with this and other structures reveal that while there is high conservation in the active site, it is the structural elements surrounding the catalytic centre that confer the differences in substrate specificity displayed by GH43 glycoside hydrolases. In *Hi*AXHd3 the xylotriose backbone, and by inference the xylan backbone, extends over β-blade 4 and the loop connecting β-blades 2 and 3. The orientation of xylotetraose in the arabinoxylan-specific arabinofuranosidase *Bs*AXHm2,3, and xylotriose in *Hi*AXHd3, is very similar. By contrast, in the arabinan-specific enzymes *Cj*Arb43A, AbnB and *Sa*Araf43A, the backbone chain is in a perpendicular orientation with respect to the xylooligosaccharide molecules in *Bs*AXHm2,3 and *Hi*AXHd3, Figure 4.24. This radical difference in the location of these polymers is due to the extension of β-strand 1 in blade 3 and its associated loops in the three arabinanases, which would make steric clashes with the xylan chain. The surface topologies of these enzymes are tailored towards the backbone conformations of their target substrates (Figure 4.24).

The NMR data on W525A-mediated hydrolysis of arabinoxylan suggests that coordination of the polysaccharide backbone is important in orienting arabinose side chains into the active site of *Hi*AXHd3. Xylan is a highly symmetrical molecule; only the endocyclic oxygen provides any asymmetry. In the arabinan backbone, arabinofuranose rings are also symmetrical, apart from the angle of the endocyclic oxygen, but the glycosidic oxygen and C5 will adopt different positions in the two orientations of the backbone. As described above, *Hi*AXHd3 makes two polar contacts with the xylan pyranose ring; Tyr165 and Trp525 form hydrogen bonds with the endocyclic oxygen of the xylose at +2NR and +2R, respectively. By contrast, *Bs*AXHm2,3 makes no polar contacts with the endocyclic oxygen of bound xylotetraose, and thus the tetrasaccharide can bind to the enzyme in either orientation (Vandermarliere et al., 2009). This explains why the *Bacillus* arabinofuranosidase is able to hydrolyse either O2 or O3 linked single arabinose

decorations, while *Hi*AXHd3 is specific for only the O3 in the double substitution. It would, however, seem unlikely that polar residues that target the endocyclic oxygen of xylose residues in xylan so precisely would also make hydrogen bonds with the glycosidic oxygen of arabinan. Further investigation by crystallography is required to ascertain the nature of specificity in arabinan by probing for contacts with the backbone of this polymer; such experiments are included under recommendations for future work, discussed in Chapter Five of this thesis.

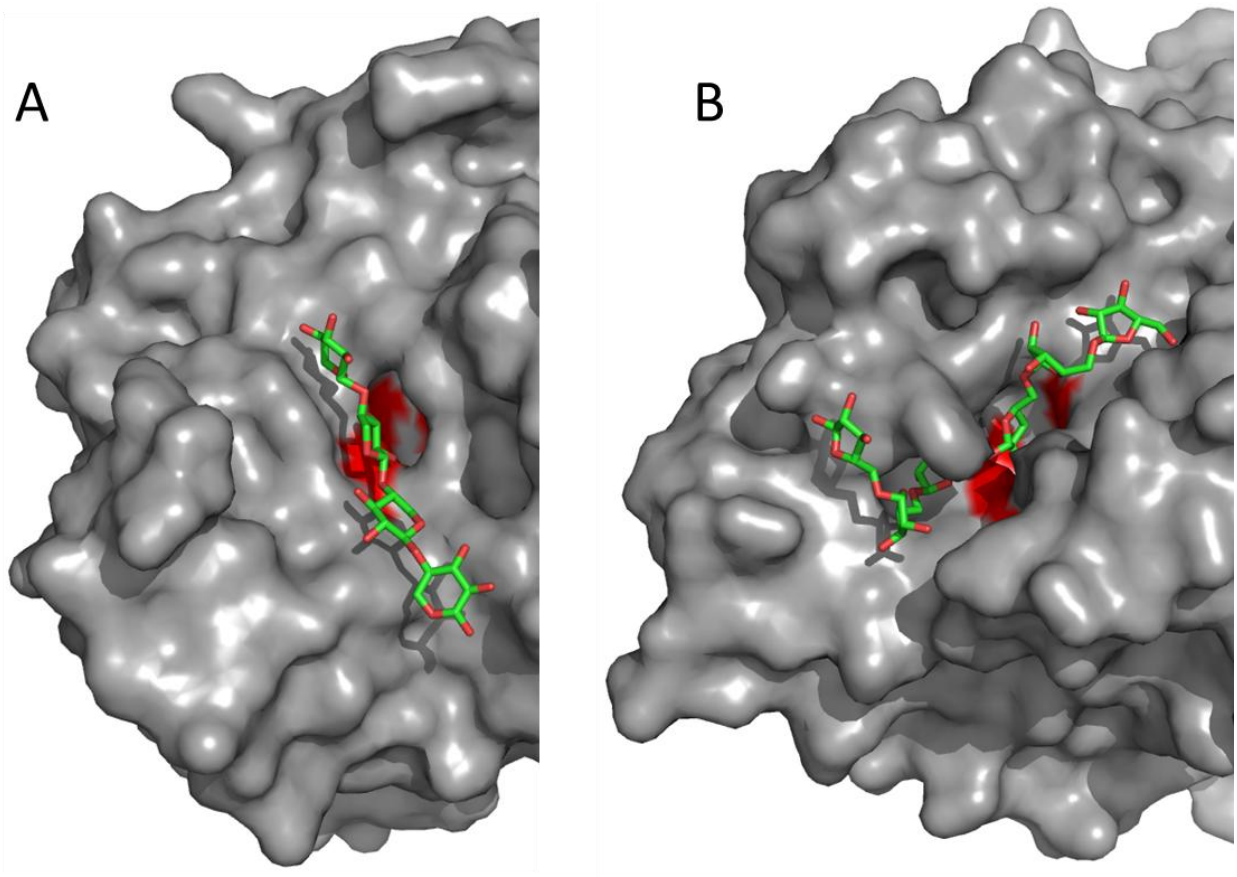


Figure 4.24 Surface representations of two GH43s in complex with substrate illustrate the differences in binding to xylan and arabinan.

A: *Bs*AXHm2,3 in complex with xylotetraose, which adopts a linear conformation and binds along a straight cleft. This linear cleft would occlude a helical arabinan backbone.

B: *Cj*Arb43A in complex with arabinohexaose, which twists around the active site in a cleft which turns over the surface of the protein. This curved cleft occludes the linear xylan backbone.

4.8.4 A conserved aromatic residue is critical to specificity of *Hi*AXHd3

Considering all of the evidence presented here, the specificity displayed by *Hi*AXHd3 appears to be conferred by orienting the polysaccharide backbone, which is mediated by Trp525, located on a loop in the C-terminal β-sandwich domain.

Trp525 is thus key to the specificity of *Hi*AXHd3; the residue is predicted to make hydrophobic contacts with the +2R xylose and the arabinose appended at O2 of the +1 xylose, as well as a polar contact between Nδ1 of the indole ring and the endocyclic oxygen of the +2R xylose. NMR experiments have demonstrated that mutation of Trp525 to Alanine causes the enzyme to become much more flexible, able to cleave either the O2 or O3 linked arabinose in the double substitution. Interestingly, this residue is contributed not by the catalytic β-propeller domain but by a loop within the C-terminal β-sandwich domain (Figure 4.25). A recent study (Yoshida et al., 2010) which looked at GH43 sequences that include this β-sandwich module (termed the XX module) examined the role of the module in substrate binding and recognition in an AXH enzyme expressed by *Fibrobacter succinogenes*, which targets monosubstituted arabinofuranose side chains in arabinoxylan. Several conserved aromatic residues within the β-sandwich domain, which were thought to be critical for carbohydrate binding, were mutated to Alanine. Only one such mutant lost binding activity, indicating that a Tyrosine residue makes a significant contribution to ligand recognition. This Tyrosine (Tyr484) is contributed by a loop in the β-sandwich module equivalent to the loop containing Trp525 in *Hi*AXHd3 (Figure 4.25). Thus, this loop in the XX module of *Hi*AXHd3 contributes directly to specificity by supplying a key residue that interacts with the substrate.

4.8.5 A conserved loop contributes topologically to specificity

As described above, a key specificity determinant for *Hi*AXHd3 is contributed by a loop in the β-sandwich module of the enzyme. The topology of this loop, which is conserved, also appears to influence specificity. The *Geobacillus stearothermophilus* β-xylosidase XynB3 (Brux et al., 2006) contains a loop in the XX domain which intrudes into the β-propeller in the vicinity of the active site pocket (Figure 4.25). This enzyme demonstrates an *exo*-type activity on xylooligosaccharide substrates and is also active on *p*NP-β-D-xylopyranoside. Enzyme activity is impaired against xylan

and longer oligosaccharides, and the authors suggest that the loop may restrict access of longer substrates to the active site. Indeed, analysis of the structure overlaid with xylotetraose shows that while the loop does not directly clash with the xylan backbone, it does contribute surface architecture to a tight pocket which excludes *endo*-type attack on xylan (Figure 4.25). A similar long loop is found in an α-L-arabinofuranosidase from *Selenomonas ruminantium*, SXA (Brunzelle et al., 2008) which displays activity primarily on 4NP-β-D-xylopyranoside, although it is also active on 4NP-α-L-arabinofuranoside. As the equivalent loops in FSUAXH1 and *Hi*AXHd3 are much shorter, it seems likely that the length of this loop is key to determining preference for substrate length; the short loops in these AXH enzymes correspond to activity on polysaccharide while steric clashes with the longer loops in XynB3 and SXA cause the enzymes to act on shorter, simpler substrates by inhibiting *endo*-type hydrolysis of polysaccharides. *Hi*AXHd3 and FSUAXH1 cleave doubly and singly substituted arabinofuranose residues from arabinoxylan, respectively. As shown in Figure 4.25, an overlay of the FSUAXH1 homology model with the *Hi*AXHd3 ligand shows how the conserved loop from the β-sandwich module excludes the double substitution; the O3-linked arabinose is positioned in the active site of FSUAXH1 while the O2-linked arabinose clashes with the protein surface. As the loop is much shorter in *Hi*AXHd3, this clash does not occur. Thus, this loop structure contributes topologically to specificity in GH43 enzymes by influencing the length and decorated nature of the substrate.

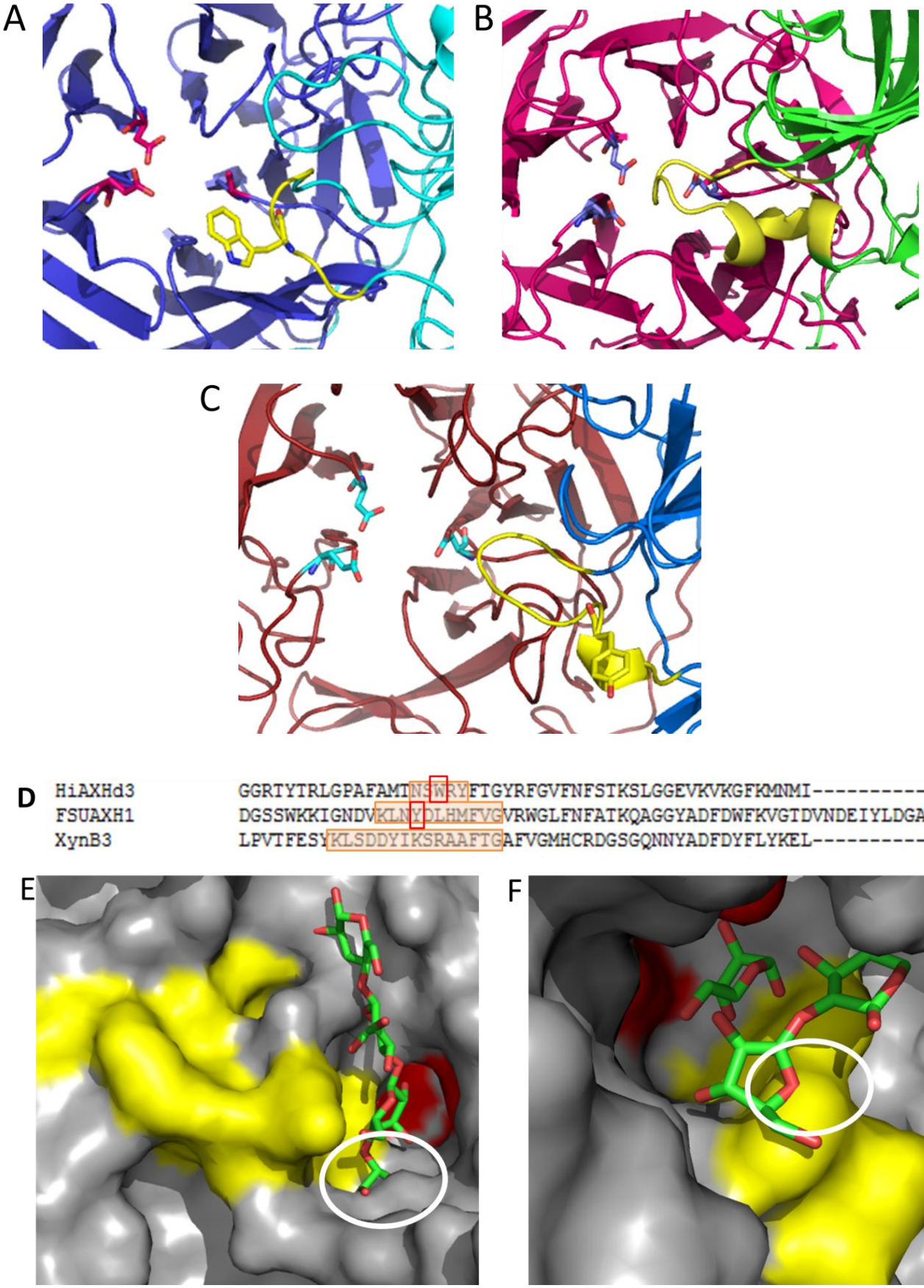


Figure 4.25 A loop in the β-sandwich module contributes a key aromatic residue.
[210]

Analysis of several GH43 structures which include a conserved loop in the β-sandwich domain gives insight into the role of the loop in determining specificity.

A: in *Hi*AXHd3 the critical specificity determinant Trp525 is contributed by a short loop in the β-sandwich. The β-propeller is blue while the β-sandwich is cyan. Trp525 and the associated loop are shown in yellow, while the catalytic residues are shown in red.

B: in XynB3 shows that the equivalent loop (yellow) is much longer and extends further into the active site. The β-propeller is magenta while the β-sandwich is green. The catalytic residues are in blue.

C: a homology model of FSUAXH1 shows that the loop (yellow) again extends towards the active site and contributes a critical Tyrosine. The β-propeller is shown in dark pink while the β-sandwich is blue. The catalytic residues are shown in cyan.

D: a sequence alignment between *Hi*AXHd3, FSUAXH1 and XynB3 show that the loops containing Trp525 and Tyr484 (marked with red boxes) align in the sequence but are different in length and do not share sequence similarity.

E: XynB3 overlaid with xylotetraose shows how the loop contributes topologically to a tight pocket which excludes polysaccharides. The loop is shown in yellow and catalytic residues are red. The clash with the xylan backbone is circled.

F: the FSUAXH1 homology model overlaid with the *Hi*AXHd3 ligand shows how this loop excludes the double substitution; the O3-linked arabinose is in the active site while the O2-linked arabinose clashes with the protein surface. As the loop is much shorter in *Hi*AXHd3, this clash does not occur. The loop is shown in yellow, and the catalytic residues are highlighted in red. The clash with the O2-linked arabinose is circled.

4.8.6 The C-terminal β-sandwich of AXH enzymes influences specificity

A phylogenetic analysis of GH43s by Yoshida, Hespen et al. identified a cohort of enzymes from the family which all display β-xylosidase and /or arabinofuranosidase activity and contain the C-terminal β-sandwich module defined as XX (Yoshida et al., 2010). Truncation studies of a xylan-specific arabinofuranosidase from *Fibrobacter succinogenes*, FSUAXH1, confirmed that the XX module is important for activity, as removal of the module significantly impaired the activity of the enzyme. The XX module, although having no direct part in catalysis, was shown to be important in binding to arabinoxylan. Indeed, the work demonstrated the interdependence between the two modules in substrate binding, indicating that the interface between

the GH43 module and the XX module represents a new form of carbohydrate binding module.

Structural analysis reveals that the β-sandwich domain can influence specificity via its orientation with respect to the catalytic GH43 module. In *Bs*AXHm2,3, which cleaves monosubstituted arabinofuranose side chains from arabinoxylan, the C-terminal domain is displaced relative to the *Hi*AXHd3 domain (Figure 4.26). Conversely, in the *Selenomonas* β-xylosidase /arabinofuranosidase *Sr*Xyn43A (PDB code 3c2u) and the *Geobacillus* β-xylosidase, XynB3, the β-sandwich domain overlaps with the equivalent region of the *H. insolens* arabinofuranosidase. It is only in *Hi*AXHd3 that a clear cleft, which accommodates a xylooligosaccharide, is evident between the two domains; Figure 4.26 shows how this cleft lies at the interface between the β-propeller and the β-sandwich. In *Bs*AXHm2,3, this cleft is contributed solely by the β-propeller domain (Figure 4.26). In XynB3 and *Sr*Xyn43A the distal region of the extended active site pocket is, in part, contributed by the loop connecting β-strands 10 and 11 in the β-sandwich domain, again suggesting that this domain plays a role in substrate binding.

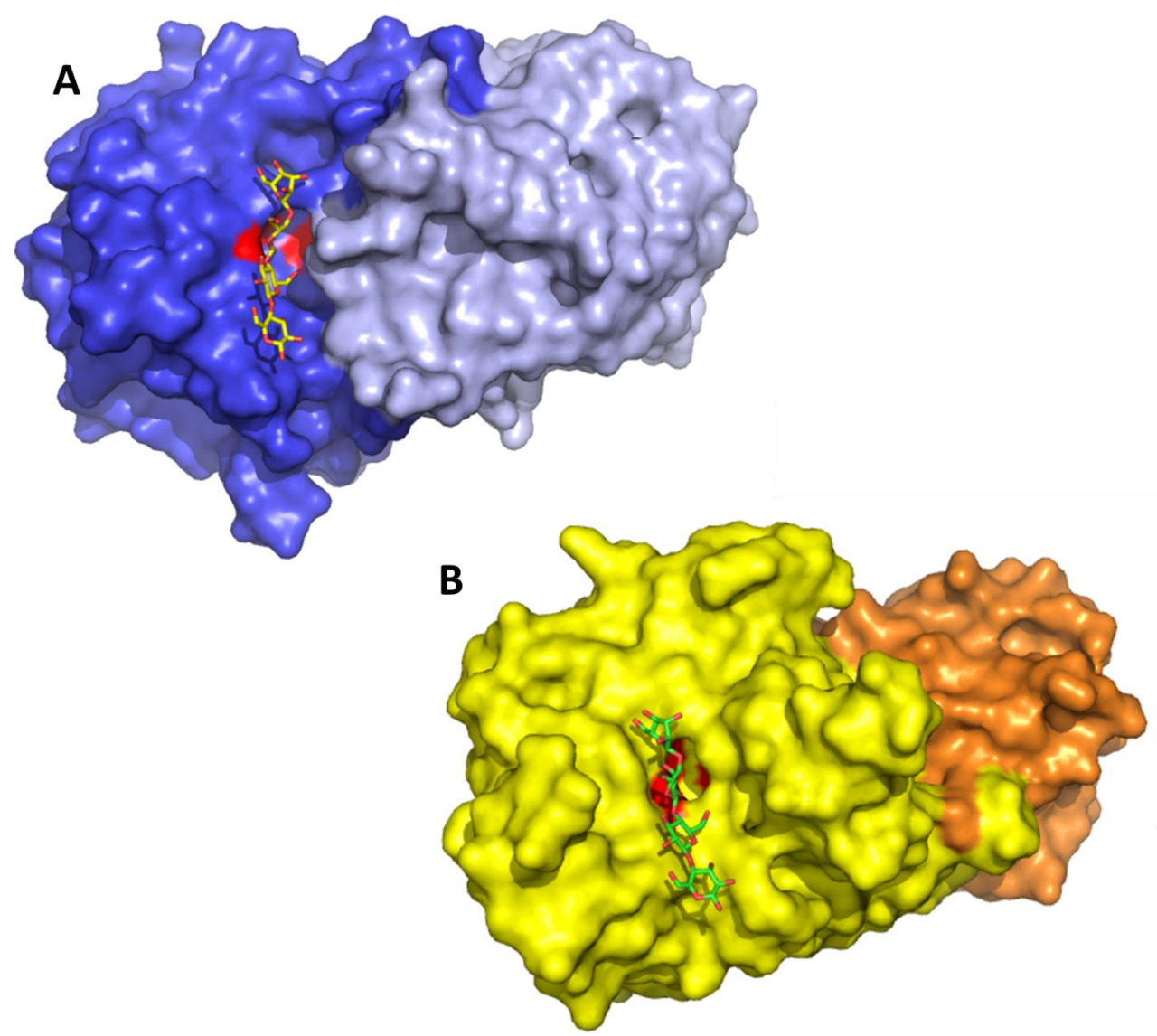


Figure 4.26 Orientation of the C-terminal β-sandwich domain leads to topological differences around the active site.

Surface representations of *Hi*AXHd3 (A) and *Bs*AXHm2,3 (B) show how the change in orientation of the β-sandwich module leads to differences in active site architecture. The substrate binding cleft of *Hi*AXHd3 (A) is located in the interface between the two domains, whereas the equivalent cleft in *Bs*AXHm2,3 (B) is formed solely by the β-propeller domain. Catalytic residues are shown in red in both. The xylan binding clefts are indicated by xylotetraose, taken from the *Bs*AXHm2,3 crystal structure. The ligand is shown in yellow in panel A and in green in panel B.

4.8.7 A single mutation, Y165A, introduces a novel activity to *Hi*AXHd3

Extensive mutagenesis of *Hi*AXHd3 has been undertaken, as described above. Of particular note is the mutation Y165A, which not only reduced the

arabinofuranosidase activity of the enzyme but also introduced the capacity to hydrolyse internal linkages in the xylan backbone. Inspection of the Y165A mutant reveals that the removal of the phenolic ring disrupts the walls of the active site pocket. Figure 4.27 shows the change to active site architecture in the Y165A variant. The expansion in the active site pocket allows entry of a xylan backbone chain into the active site. The capacity to hydrolyse the glycosidic linkages in the backbone of a pyranose configured polysaccharide, and to remove furanose side chains from the same polysaccharide, has not previously been observed in nature, nor generated through rational design or forced protein evolution. The mechanism by which an enzyme can be converted into a glycoside hydrolase that displays both *exo*-acting arabinofuranosidase and *endo*-xylanase activities is intriguing, particularly as the same active site catalyses the two reactions. In *Hi*AXHd3, the pocket is wider than other *exo*-acting GH43 enzymes and thus may be more amenable to modification such that it can accommodate the backbone sugar of a polysaccharide.

Kinetic analysis of the Y165A mutant (Table 4.5) was performed against a variety of substrates. Substrate depletion assays showed that the catalytic efficiency of the xylanase activity of the enzyme was 28-fold greater against xylohexaose compared to xylopentaose, Table 4.5, with no measurable activity on shorter oligosaccharides, indicating that either the -3 or +3 subsite confers 1.96 kcal/mole of binding energy, calculated according to Equation 4.1. This correlates with reaction products generated by Y165A from the relatively undecorated birchwood xylan (Figure 4.15), which showed that the initial products had d.p. values greater than three, which were subsequently degraded to oligosaccharides as small as xylobiose.

Equation 4.1 $(R\ T\ \ln k_{cat}/K_M)_{xylohexaose} - (R\ T\ \ln k_{cat}/K_M)_{xylopentaose}$

where R = ideal gas constant (8.314 J K^{-1} mol^{-1})

T = temperature (298 K)

4.8.8 Structural analysis and mutagenesis of the Y165A mutant

The structure of the Y165A variant of *Hi*AXHd3 shows that the removal of the phenolic side chain disrupts the face of the pocket directly below the glycosidic oxygen between the +1 and +2R xylose residues, Figure 4.27. This will enable the

xylan backbone to adopt a deeper position in the enzyme so that it can enter the active site and occupy the -1 subsite. As Y165A displays xylanase activity, and catalysis requires the original catalytic apparatus of the enzyme, it is likely that the xylose at the -1 subsite can make appropriate interactions with *Hi*AXHd3 so that the sugar can distort into an oxocarbenium transition state conformation, enabling catalysis to proceed, albeit inefficiently.

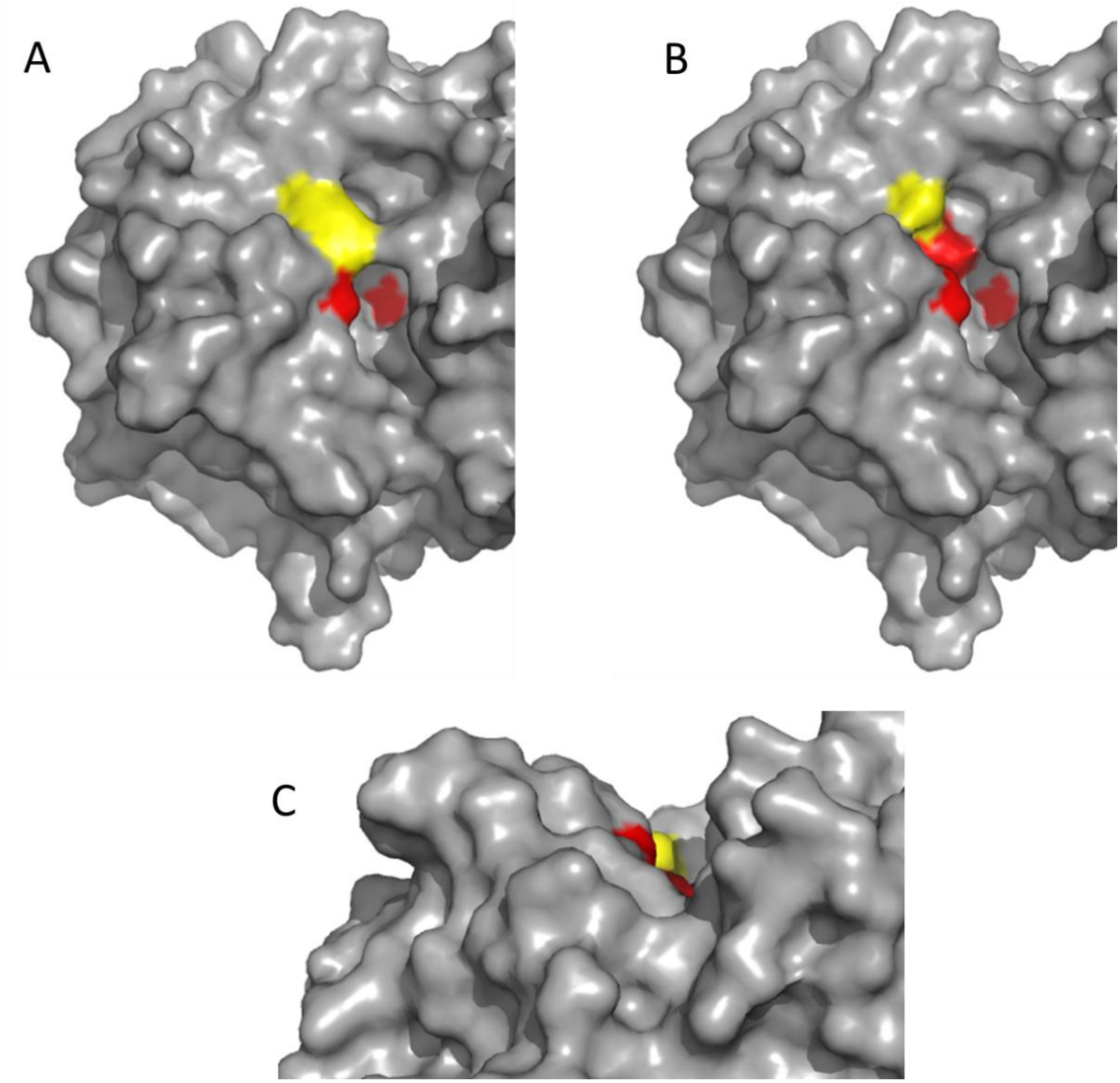


Figure 4.27 Comparison of wildtype and Y165A variant *Hi*AXHd3.

Surface representations of the active sites of wildtype (A) and Y165A variant (B) *Hi*AXHd3 demonstrate the expansion in the pocket topology caused by the loss of a single phenolic side chain, creating a much more open left in the mutant (B). Residue 165 is picked out in yellow in both

structures (Tyr165 in the wildtype, Ala165 in the mutant). The three catalytic residues are shown in red in both structures. Panel C shows a side-on view of the Y165A mutant cleft.

Figure 4.28 shows a classical GH10 xylanase in complex with xylotetraose. This enzyme is a much more efficient xylanase than Y165A. Values of k_{cat}/K_M for a *Cellvibrio japonicus* GH10 xylanase against xylohexaose and xylopentaose are 1.0 x 10^7 and 2.9 x 10^6 min^{-1} M^{-1}, respectively. For *Hi*AXHd3-Y165A, these values are 3.5 x 10^4 and 1.3 x 10^3 min^{-1} M^{-1}, respectively. These differences may be due to the strategies of substrate binding employed by the two enzymes. Where xylan is predicted to lie relatively flat in the open cleft of Y165A, the polysaccharide bends to fit into the active site of the GH10 (Charnock et al., 1998; Pell et al., 2004b). In GH10 enzymes, the -2 and +2 subsites in particular make many interactions which are absent in Y165A, which therefore needs to bind more extensively to the rest of the xylan chain. Kinetic analysis of the Y165A mutant has shown that binding to five subsites is required for activity (Table 4.5). The importance of the +2 subsite in the GH10 is underlined by the difference in k_{cat}/K_M of this enzyme against xylotetraose and xylotriose: hydrolysis of X4 is around 100 times faster than hydrolysis of X3 (8.5 x 10^4 versus 7.3 x 10^2 k_{cat}/K_M).

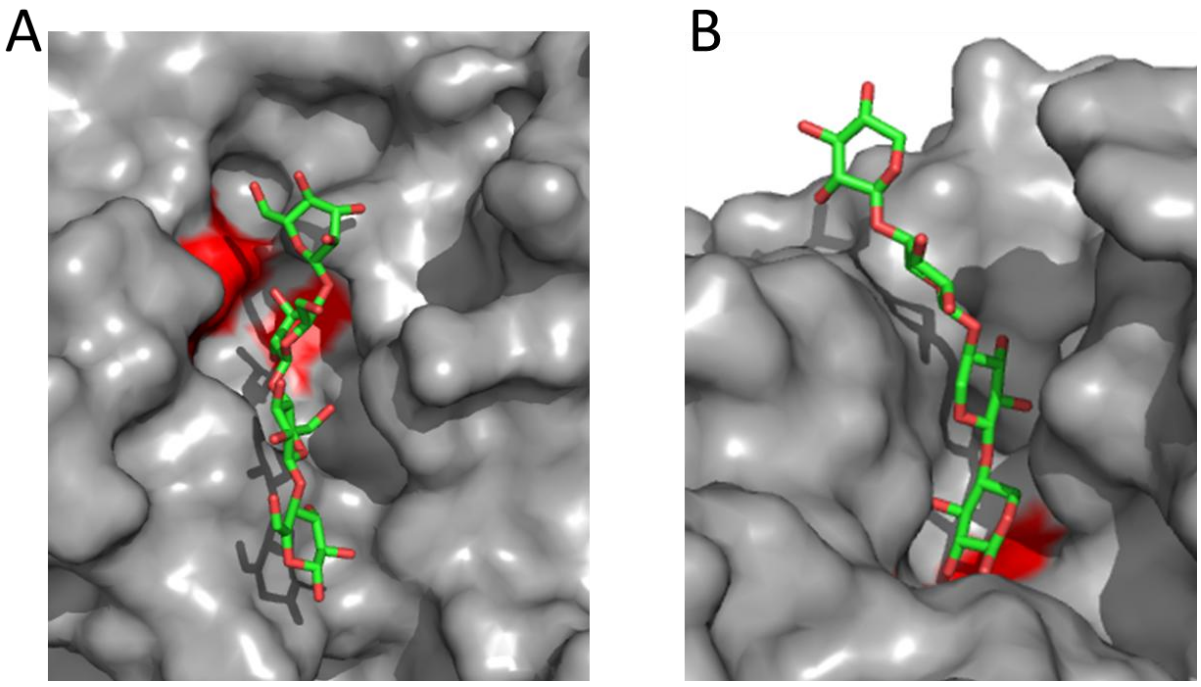


Figure 4.28 Comparison of the substrate binding clefts of Y165A and a GH10 xylanase.

Surface representations of the *Hi*AXHd3 variant Y165A and a GH10 xylanase (Harris et al., 1996; Pell et al., 2004b) illustrate the differences in substrate binding. Catalytic residues are shown in red in both panels.

A: crystal structure of *Hi*AXHd3 mutant Y165A overlaid with xylotetraose from the crystal structure of *Bs*AXHm2,3 shows how the polysaccharide lies along the shallow substrate binding cleft. It is inferred that for the xylanase activity of Y165A, the xylan backbone lies along this cleft.

B: the GH10 in complex with xylotetraose shows that here, the polysaccharide points into the active site and is twisted into the cleft.

To further explore catalysis of the xylanase and arabinofuranosidase reactions by the *Hi*AXHd3 variant Y165A, six additional Alanine mutations were introduced into the Y165A mutant. Table 4.6 summarises the kinetic parameters obtained for each mutant. His271 was previously shown to be important for recognition of the O2 linked arabinose substitution, as mutation of this residue impaired catalytic efficiency by a reduction in k_{cat}. The data show that the double mutant Y165A/H271A has lost its arabinofuranosidase activity but displays xylanase activity comparable to Y165A, although a greatly reduced K_M suggests tighter binding to the xylan backbone. This demonstrates that while the same active site and catalytic machinery are responsible for each reaction, different residues are important in coordinating the two substrates.

Other residues (Asn183, Pro233, Phe288 and Phe492) were proposed to form part of the wall of the newly exposed substrate cleft (Figure 4.29) and so were also mutated individually to Alanine to explore their role in this novel activity. Compared to the Y165A variant, arabinofuranosidase activity of the Y165A/F288A double mutant was relatively unchanged, while k_{cat}/K_M for the xylanase activity against birchwood xylan increased slightly. Pro233 was identified in wildtype structures as providing hydrophobic support to Glu215, aiding catalysis. The double mutant Y165A/P233A showed xylanase efficiency comparable to that of the Y165A variant, but arabinofuranosidase activity was reduced by ~10-fold. The different orientations of the target glycosidic bond in the substrates for these two activities may place different demands on the microenvironment of the O2 arabinose pocket, so that loss of Pro233 as an accessory residue impacts the two activities differently.

Of particular note are the double mutants Y165A/N183A and Y165A/F492A, both of which show a significant increase in xylanase activity compared to the original Y165A variant. While K_M for all three enzymes is comparable, increases in k_{cat} (5-fold for Y165A/N183A and 6-fold for Y165A/F492A) led to improved catalytic efficiency. Asn183 is located towards the reducing end of the substrate, around 4-5 Å from the non-reducing xylose in the *Hi*AXHd3/Araf-Xyl_3 structure. Phe492 is located at the floor of the cleft towards the non-reducing end of the ligand, approximately 6 Å from its nearest xylose in the xylotriose structure, but predicted to be much closer to a xylose in the full-length polymer. Structure models of these mutants (Figure 4.29) indicate that loss of these side chains simply opens the cleft even further, presumably allowing the xylooligosaccharide to nestle more deeply into the active site, improving the positioning of the glycosidic bond cleaved and facilitating generation of the transition state.

It is interesting to note that the Y165A/F492A mutant increases arabinofuranosidase activity compared to the Y165A variant by 50 %. This is intriguing as Phe492 makes no contacts with either arabinose residue; the amino acid is located some 4 Å from the O2 arabinose and possibly contributes a small portion of the hydrophobic surface of the O2 binding pocket. Loss of the phenolic side chain of this Phenylalanine is thought to improve xylanase activity by providing greater space and thereby facilitating substrate binding, but this does not seem to explain the increased catalytic efficiency for the arabinofuranosidase activity. Compared to the single mutant Y165A, the double mutant Y165A/F492A appears to increase k_{cat} for both arabinofuranosidase and xylanase activities, possibly as product departure is easier and faster when the side chain of Phe492 is absent.

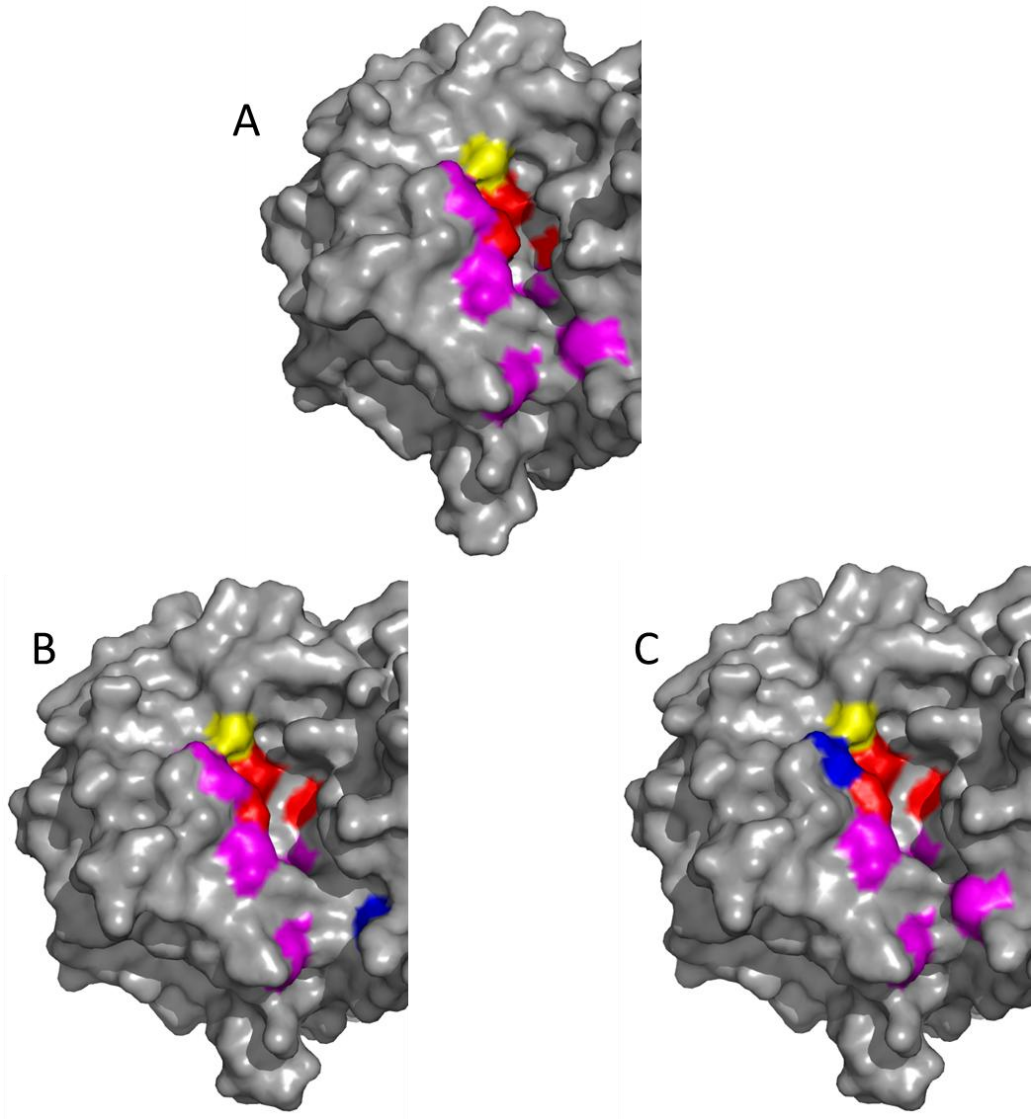


Figure 4.29 Structural representation of those residues of the substrate binding cleft of Y165A selected for further mutagenesis.

A: Surface representation of *Hi*AXHd3 variant Y165A showing those residues selected for further study by mutagenesis. Catalytic residues are shown in red, while Ala165 is picked out in yellow. Residues highlighted magenta were mutated as they form part of the walls of the xylan-binding cleft.

B: model of the Y165A-F492A double mutant. Phe492 is picked out in blue and lies at the base of the xylan cleft.

C: model of the Y165A-N183A double mutant. Asn183 is picked out in blue at the top of the cleft.

Introduction of the F492A mutation into the Y165A variant of *Hi*AXHd3 leads to improvement in both wildtype and xylanase activities. Conversely, introducing the N183A mutation (to generate Y165A/N183A) leads to little change in

arabinofuranosidase activity but a marked increase in k_{cat} for xylanase activity. Observation of the structure of *Hi*AXHd3 and the Y165A mutant shows that Asn183 hydrogen bonds to the catalytic acid Glu215 (Figure 4.30). It is predicted that this hydrogen bond helps to orient the acid correctly for wildtype arabinofuranosidase activity. The scissile bond for xylanase activity is likely to be in a different position than for arabinofuranosidase activity, so that Glu215 is not optimally placed for cleavage of the backbone. When this hydrogen bond is lost by mutating Asn183 to Alanine the acid shifts to a more favourable position for hydrolysis of the backbone. This has resonance with the view that loss of Pro233, which provides hydrophobic support to Glu215, reduces arabinofuranosidase activity of the Y165A mutant while xylanase activity is relatively unchanged, suggesting that interaction with Pro233 is more important for the catalytic acid in cleavage of the arabinose side chain than for hydrolysis of the backbone.

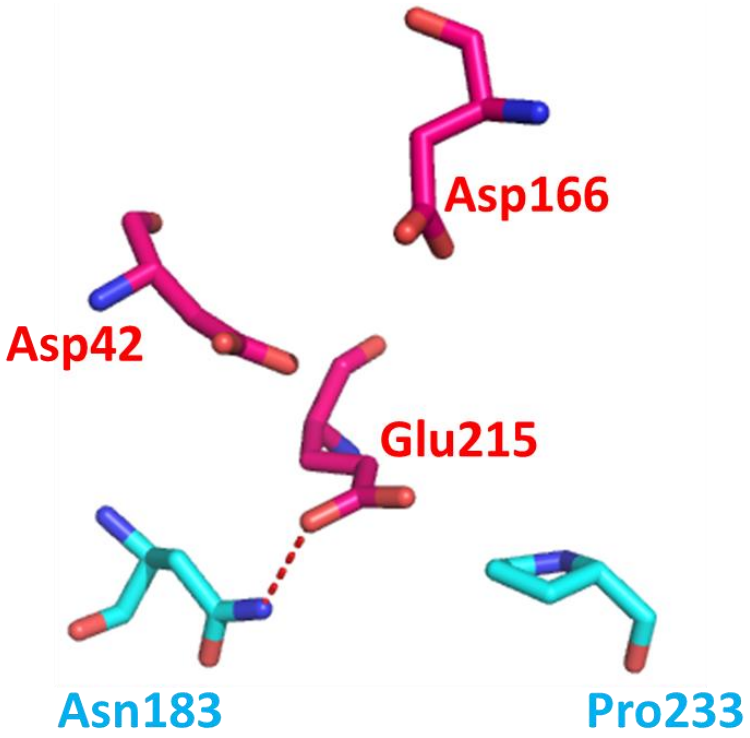


Figure 4.30 Asn183 and Pro233 support Glu215, the catalytic acid.

Catalytic residues are shown in red while Asn183 and Pro233 are in cyan. The hydrogen bond between Asn183 and Glu215 (red dashed line) is predicted to assist in orienting the catalytic acid for arabinofuranosidase activity. Loss of this hydrogen bond in the N183/Y165A double mutant allows the acidic residue to move to a position more favourable for xylanase activity. Pro233 is in position to

provide hydrophobic support to Glu215. Loss of this side chain impairs arabinofuranosidase activity but has little effect on xylanase activity.

The Y165A mutation of *Hi*AXHd3 significantly alters the activity of a highly specific enzyme. The observation that an additional catalytic activity can be introduced into a glycoside hydrolase gives insight into the catalytic activity displayed by GH43 family enzymes. In a way, the Y165A mutation does not in fact introduce a novel activity but rather reveals a pre-existing activity, which may be considered a 'default' for the GH43 family. The open cleft of the Y165A variant is present in the wildtype enzyme, but its floor contains a heart-shaped pocket which can be considered a secondary layer of architecture which is built on top of the typical GH43 fold that confers a low-level *endo*-xylanase activity. This view has resonance with findings discussed in Chapter Three of this thesis, where many GH43 enzymes from a human gut symbiont and, in a sister project, from a soil saprophyte displayed similar weak *endo*-xylanase activity.

The GH43 family is expanded in many microorganisms derived from varied habitats, and the family includes examples of many activities. Each new structure contributes to our understanding of the family, and the structure of *Hi*AXHd3 and its variant provide significant insight. The flexibility of Y165A in substrate recognition is consistent with the range of activities displayed by GH43 enzymes that include *exo*, *endo* and *endo*-processive modes of action against pyranose and furanose substrates where the stereochemistry at C4 varies. The capacity to introduce a novel specificity into *Hi*AXHd3 through a single amino acid substitution indicates that the 5-bladed □-propeller fold, displayed by GH43 enzymes, provides a structural scaffold that can be harnessed to bind a range of different sugars and catalyse the hydrolysis of glycosidic bonds through distinct modes of action.

As discussed above it is possible that the active site of GH43 xylosidases and arabinofuranosidases may, initially, bind the ground state sugar through the same interactions, but then utilise additional interactions to distort the arabinofuranose and xylopyranose residues into their respective transition states. Thus, while the active site of *Hi*AXHd3 is likely capable of binding either arabinofuranose or xylopyranose residues in their ground state conformations, the limited xylanase activity displayed

by the Y165A mutant may reflect its limited capacity to distort the sugar into its preferred transition state conformer.

4.8.9 Conclusion

The structure of this biotechnologically relevant arabinofuranosidase, in harness with the observation that additional catalytic functions can be introduced into this enzyme, provides a platform for evolving further, industrially significant, activities into *Hi*AXHd3 and other members of GH43. Additional discussion of the potential technological and economic implications of this finding, as well as recommendations for future work, is presented in Chapter Five of this thesis.

CHAPTER FIVE

General Discussion

This Chapter presents a summary of the work presented in this thesis regarding two major projects undertaken throughout the course of this study. An assessment of the family 43 glycoside hydrolases of the human gut symbiont *Bacteroides thetaiotaomicron* sheds some light on the biological rationale behind the expansion often seen in this family, and gives insight into a possible evolutionary history of the family. Structure-function analysis of a highly specific GH43 enzyme from the fungus *Humicola insolens* shows how tight specificity for a polysaccharide substrate is driven by multiple structural features and that a novel specificity can be engineered into an enzyme of this family.

5.1 GH43 enzymes of *Bacteroides thetaiotaomicron*

5.1.1 Conclusions

Chapter Three of this thesis details efforts to characterise all glycoside hydrolase family 43 (GH43) enzymes expressed by the human gut symbiont *Bacteroides thetaiotaomicron*, in an attempt to understand the biological rationale behind the common expansion in this family. Soluble protein was obtained for 25 of the 31 GH43 enzymes and each was subjected to an activity screen, testing for hydrolysis of polysaccharides and artificial aryl glycosides (Chapter Three, Section 3.3).

The activity screens of these enzymes were informed, where possible, by data on the upregulation of operons encoding GH43s when *B. thetaiotaomicron* is grown in the presence of a specific carbohydrate. These operons, known as Polysaccharide Utilisation Loci (PULs), allow a consortium of cooperative metabolic enzymes to be co-expressed in the presence of an appropriate substrate.

5.1.1.i Arabinan metabolism by GH43 enzymes

The most thoroughly described PUL is induced by sugar beet arabinan and codes for three GH43s (Section 3.5). Two of these GH43s (Bt0360 and Bt0367) are *endo*-

arabinanases (Section 3.5.1) with differing specificities for branched and linear regions of arabinan. Bt0367 is similar to most published arabinanases in that it is inhibited by arabinose side chains in the polymer, but Bt0360 is unusual as it is more active on branched polysaccharide (Table 3.6).

These arabinanases work in conjunction with Bt0369, an α-L-arabinofuranosidase with novel specificity for O2-linked side chains, which can be tolerated in the context of both single and double substitutions of the backbone. Cleavage of the double substitution removes a structure which is inhibitory to most arabinofuranosidases, allowing a more complete degradation of arabinan, but it is perhaps surprising that the O2-linked arabinose is removed. The amount of arabinose liberated from arabinan by Bt0369 is low, as this side chain is very rare outside of the double substitution. An enzyme which cleaved O3-linked side chains would release more arabinose, but may also be diverted from removal of the double substitutions by an abundance of the single substitutions. Thus, the evolutionary rationale behind the selection of the α-1,2-linked side chain may be that it limits cleavage of single substitutions, allowing the enzyme to maintain a catalytic focus on the double substitution. Downstream GH51 arabinofuranosidases are likely to remove the α-1,3-linked side chains which remain.

The crystal structure of a close homologue of Bt0369, which displays the same specificity, was solved in complex with ligands by Dr A Cartmell of the Gilbert laboratory (CCRC and Newcastle University). The structure of this enzyme, *Cj*Abf43A, reveals a five bladed β-propeller fold, Figure 3.19, as the sole domain of the enzyme. The structure of the ligand complex (Figure 3.20) reveals the nature of the specificity of the enzyme for the O2-linked side chain and the tolerance shown by the enzyme for the double substitution. The surface representation of *Cj*Abf43A, Figure 3.20, reveals a deep pocket in the centre of a highly curved cleft. The rim of the pocket abuts onto a shelf-like structure that accommodates the O3-linked arabinose side chain of the double substitution. The O2 of the central arabinose which participates in the target glycosidic bond points directly into the active site pocket. The orientation of this cleft is adapted to the twisted structure of the arabinan chain and excludes the linear xylan backbone, as demonstrated in overlays with an arabinoxylan-specific arabinofuranosidase.

These arabinan-active GH43s of *B. thetaiotaomicron* are hypothesised to function cooperatively in a degradative pathway which generates oligosaccharides from sugar beet arabinan (Section 3.6.1.iii, Figure 3.26). The arabinanases Bt0360 and Bt0367 cleave the arabinan backbone, generating oligosaccharides which are linear or decorated with arabinofuranose side chains. These short oligomers are taken into the cell by two SusC/D carbohydrate binding complexes. Subsequently Bt0369, the α-1,2-L-arabinofuranosidase, functions as a de-branching enzyme and strips O2-linked side chains from arabinan, thereby removing all instances of the double substitution structure which is inhibitory to the vast majority of arabinofuranosidases. Finally, in the periplasm these oligosaccharides are subjected to final metabolism by as yet uncharacterised enzymes.

Arabinan degradation in *C. japonicus* proceeds via a different pathway which nonetheless shows convergence with the *B. thetaiotaomicron* system in the use of a highly specific α-1,2-L-arabinofuranosidase in degradation. Following removal of double substitutions by *Cj*Abf43A, an extracellular GH51 arabinofuranosidase removes the remaining single O3-linked side chains. The resulting linear backbone is then cleaved exclusively to arabinotriose by an *endo*-processive arabinanase (McKie et al., 1997). The trisaccharide is metabolised by two GH43 α-1,5-*exo*-arabinanases.

5.1.1.ii A weak xylanase activity is displayed by many GH43 enzymes

The activity screen of soluble *B. thetaiotaomicron* GH43s revealed that eleven show weak *endo*-xylanase activity against xylans with a variety of arabinofuranosyl branching patterns (wheat arabinoxylan, rye arabinoxylan, oat spelt xylan and birchwood xylan). Three of these enzymes (Bt2852, Bt3094 and Bt3655) also showed activity against 4NP-α-L-arabinofuranoside and arabinose appears to be a significant product in arabinoxylan degradation by these xylanases (Figures 3.10 and 3.11). It is likely that these enzymes, particularly the most active, Bt2852, are arabinoxylan-specific arabinofuranosidases (AXH) of unknown specificity. The amount of arabinose released from birchwood xylan is comparable to that released from wheat arabinoxylan; this is in stark contrast with the double-substitution specific *Hi*AXHd3 (Chapter Four), which liberates an almost negligible amount of arabinose from birchwood xylan, suggesting that the double substitution is extremely rare in

birchwood xylan. This indicates that Bt2852 is cleaving singly substituted arabinose side chains.

The arabinofuranosidase activity displayed by Bt2852 appears to be more efficient than the xylanase action of the enzyme, which is likely a side activity. Other GH43s displaying xylanase action without complementary arabinofuranosidase activity may appear to be weakly acting as they hydrolyse a rare bond within the polysaccharide backbone. However, the enzymes do not appear to discriminate between the xylan substrates tested, which differ greatly in degrees of decoration. Therefore it seems likely that the xylanase function is a side activity. These enzymes may not be biologically active, or it may simply be that the correct substrate was not evaluated in this study. If so, it is difficult to envision what these correct substrates may be.

5.1.1.iii Phylogenetic analysis of family GH43

The GH43s of *B. thetaiotaomicron* and *C. japonicus* were submitted to phylogenetic analysis together with previously characterised examples from the family (Section 3.6.2.i, Figure 3.27). Newly described GH43s displaying *endo*-arabinanase and α-1,5-*exo*-arabinanase activity are in clades with previously described enzymes that exhibit similar activities. By contrast, and consistent with their novel activity, the well characterised α-L-1,2-arabinofuranosidases *Cj*Abf43A and Bt0369 formed a clade that contains no other enzymes with known catalytic properties. The many enzymes displaying “trace” xylanase activity are not clustered into a specific region of the family tree, suggesting that this minor activity may be a generic feature of GH43.

5.1.2 Future work

While many GH43 enzymes from *B. thetaiotaomicron* have been characterised in this study, relatively few have demonstrated novel activities. One particularly interesting enzyme characterised here is Bt0360, an *endo*-arabinanase which shows preference for branched arabinan structures. This sets Bt0360 apart from other characterised arabinanases, which are inhibited by side chains, as is Bt0367. The nature of the different specificity of Bt0360 is intriguing. Attempts to obtain crystal structures and homology models were unsuccessful, but sequence analysis indicates that this enzyme may possess an additional N-terminal module not seen in

other arabinanase enzymes (Figure 3.18). If so, this structural feature may contribute to the unusual action of the enzyme. More intensive crystallisation trials may be fruitful and yield an informative crystal structure. Another useful experiment would be the characterisation of the reaction products generated from sugar beet arabinan by Bt0360. Mass spectrometry on permethylated samples of reaction products would provide details on the decorated structures of the arabinooligosaccharides. In this way, we could learn whether the enzyme cleaves the arabinan backbone in a specific structural context, or if it is merely sufficiently tolerant of side chains that a randomised mixture of variously decorated products is generated. Of relevance to this work is an on-going project in the Bolam laboratory at Newcastle University aiming to characterise the binding proteins (SusCD complexes) of the arabinan PUL.

Another interesting GH43 identified in this study which may warrant closer examination is Bt2852, which has been defined as an arabinoxylan-specific arabinofuranosidase with background xylanase activity, as the enzyme has activity on 4NP-α-L-arabinofuranoside and releases arabinose from arabinoxylan but not arabinan, as well as a weak *endo*-xylanase activity. NMR experiments similar to those described for *Hi*AXHd3 in Chapter Four could elucidate the specificity of the arabinofuranosidase activity of the enzyme. In addition, a crystal structure of the enzyme in complex with ligand (a xylooligosaccharide or a branched arabinoxylooligosaccharide) could be very informative to our hypothesis that the GH43 fold carries the capacity for xylanase action. However, Bt2852 is only a weakly active enzyme, and many more efficient arabinofuranosidases and xylanases have been described, so the value of the labour required for these efforts may be questionable.

Similarly, the enzyme Bt3675 could be said to warrant further examination. This enzyme proved highly problematic in terms of expression and purification of soluble protein (Figure 3.7) but an activity screen using impure cell-free extract of over-expressed Bt3675 indicated that the enzyme was active on 4NP-β-D-galactopyranoside and released two products from arabinogalactan, a monosaccharide and a second, larger product. This could be an interesting activity but was not pursued due to the difficulty of working with the enzyme. According to the CAZy database (Cantarel et al., 2009), family GH43 contains *exo*-β-1,3-

galactanases (Ichinose et al., 2005; Kotake et al., 2009) and at least one β-galactosidase (Beloqui et al., 2010); Bt3675 is likely to belong to one of these classifications. Indeed, the gene for Bt3675 is found on the *B. thetaiotaomicron* genome in a PUL which is upregulated in the presence of arabinogalactan. The only other enzymes present in this PUL are the GH43s Bt3683 and the non-functional Bt3685, which lacks catalytic residues (Section 3.6.2.ii). Bt3683 is a bimodular GH43 linked to a GH16 module, a family that includes β-galactosidases (Tempel et al., 2005). In this study, soluble protein was not obtained for Bt3683, but one can imagine cooperativity between this enzyme and Bt3675 in the hydrolysis of arabinogalactan.

Those enzymes whose substrate has not been identified in this study remain a puzzle. The search for substrates is complicated when one considers that the activity of enzymes which work synergistically in biology, as is quite likely for the components of the PULs of *B. thetaiotaomicron*, may only be apparent when assayed in biologically realistic conditions, such as in the presence of other appropriate enzymes. One possible avenue of exploration is the hydrolysis of complex glycans. GeneChip data indicate that four of the PULs containing GH43s (Table 3.1) are upregulated when *B. thetaiotaomicron* is grown in the presence of N-linked glycans. In the gut, the bacteria would certainly be exposed to such structures, from dietary intake of eukaryotic tissue (meat and plant material) and attached to host cells of the intestine. Bacterial gut symbionts are hypothesised to play a role in the rapid turnover of epithelial cells in the digestive tract (Martens et al., 2008; Sonnenburg et al., 2005) and this may be mediated by hydrolysis of surface glycans. *B. thetaiotaomicron* has been shown to be capable of turning to host glycans in times of scarcity of other nutrient sources (Sonnenburg et al., 2005). However, the structure and composition of human glycans do not represent a substrate for GH43. Complex N-linked plant glycans contain a β-linked xylopyranose residue and as such may be hydrolysed by members of family GH43, as there are characterised β-xylosidases within this family.

5.1.3 Implications

The major finding of the work presented in Chapter Three is an understanding of the metabolism of arabinan by *B. thetaiotaomicron* by description of the enzymes of PUL 7, which function as a cooperative system. GeneChip and carbohydrate binding data show that the hybrid two component system that regulates transcription of the PUL binds arabinan, thus the enzyme activities described in this study are likely to be biologically significant and give insight into the functioning of the metabolic PUL system as a whole.

Additionally, the work presented here gives insight into the GH43 family. By demonstrating that a weak xylanase activity is found in distantly related sequences from varied bacterial species, credence is given to an evolutionary theory of family GH43. The data suggest that an early member of the family displayed a biologically significant xylanase activity. It is hypothesised that this was rendered redundant by the acquisition of more efficient xylanases (GH10 and GH11) which resulted in GH43 enzymes developing additional structural features to alter the function of these proteins or, in some instances, biological relevance was lost. The presence in *B. thetaiotaomicron* of multiple inactive enzymes which appear not to possess catalytic machinery supports this. The phylogenetic analysis shown in Chapter Three (Figure 3.27) also suggests that it is likely that weak xylanases will be found in other bacterial species, including *Cellvibrio japonicus*.

The weak xylanase Bt2852 has been designated as an arabinoxylan-specific arabinofuranosidase; the potential implications of this dual function are interesting. Complete degradation of arabinoxylan is a convoluted process, traditionally requiring a consortium of glycoside hydrolases from multiple families to remove all side chains and efficiently hydrolyse the backbone. Arabinoxylan is a highly abundant polymer in the plant cell wall, which is increasingly viewed as an important raw material for the production of chemicals and transport fuel. However, the polysaccharide is complex, and degradation requires enzymes including, but not limited to, arabinofuranosidases and xylanases. The complexity of the process to saccharify arabinoxylan, particularly the requirement for multiple enzymes to be produced, is one factor which hampers the economic utilisation of the plant cell wall. The

economic viability of the process would be increased if fewer enzymes could be used. Thus, a glycoside hydrolase such as Bt2852 with demonstrable dual degradative functions against arabinoxylan is an attractive proposition. However, the enzyme is very slow, which clearly reduces its utility. Nonetheless, knowledge of this enzyme may inform future experiments in protein engineering which aim to produce similar dual function enzymes for uses in the biofuel and bioprocessing sectors.

5.2 *Hi*AXHd3, a fungal arabinofuranosidase

5.2.1 Conclusions

5.2.1.i Specificity of the wildtype enzyme

In Chapter Four of this thesis, work is presented which describes the structure and specificity of *Hi*AXHd3, a fungal arabinofuranosidase of industrial significance. The enzyme had previously been shown to be specific for α1,3 linked arabinose side chains found in the double substitution of arabinoxylan (Sorensen et al., 2006). The protein has now been shown to possess the same specificity for arabinan; kinetic analysis has shown that rates are roughly comparable against both substrates. NMR experiments demonstrated specificity for the O3-linked arabinose in the double substitution in arabinan and arabinoxylan.

Crystal structures were sought to better understand the specificity displayed by *Hi*AXHd3. The structure of the enzyme shows that the protein comprises two modules linked by a short flexible loop (Chapter Four, Figure 4.9). The N-terminal module is the canonical β-propeller common to all GH43s, while at the C-terminal is a β-sandwich module not seen in all GH43s but which is common for arabinoxylan-specific arabinofuranosidases (Vandermarliere et al., 2009; Yoshida et al., 2010). Analysis of the surface structure of the enzyme shows that the active site (containing the three catalytic residues Asp42, Asp166 and Glu215) is a pocket in the surface, adjacent to which is a second depression. Together, these features make a heart-shaped pocket, which accommodates the double substitution of the substrate. Another notable feature is a substrate binding cleft that accommodates xylan and likely arabinan, which extends over the protein and is formed at the interface between the β-propeller and β-sandwich modules.

The structure of *Hi*AXHd3 in complex with xylotriose, in which the central xylose has an O2-linked arabinose, showed that the side chain points into the pocket structure adjacent to the active site (defined as the +2NR* subsite). This structure provides information on enzyme contacts with the O2 arabinose and on the orientation and interactions of the xylan backbone, which are inferred from the positioning of the xylotriose ligand. The structure of a second protein-ligand complex has density for a single xylose residue (+1 subsite) linked α-1,2 to an arabinose in the pocket adjacent to the active site, and an arabinofuranose is also present in the active site (-1 subsite) that is not covalently linked to the +1 xylose.

These structures give insight into the nature of the specificity of *Hi*AXHd3. Arg296 makes direct polar contacts with both arabinose residues, while Gln272 contacts the O3-linked arabinose at the -1 subsite. His271 makes a polar contact with the O2-linked arabinose, which is also coordinated by hydrophobic contacts with Phe288 and Thr231, residues which line the O2 binding pocket (+2NR*). The xylan backbone is positioned in the substrate binding cleft by polar contact between the endocyclic oxygen of the xylose at subsite +2NR and Tyr165. In addition, Trp525, which belongs to the β-sandwich module, makes stacking interactions and a weak hydrogen bond with the xylose at +2R (Figure 4.13).

Each of these residues has been mutated to Alanine. NMR analysis of the activity of these mutants has shown that only the W525A mutation leads to a change in specificity. This mutant is able to cleave either the O2 or O3 linked arabinose from the double substitution, indicating that coordination of the polysaccharide by Trp525 is critical for orientation of the backbone such that the α-1,3 linked arabinose sits in the active site. Specificity is likely aided by interactions with the O3-linked arabinose; mutation of Arg296 or Gln272, which contact the -1 arabinose directly, is lethal to the enzyme.

Removal of the double substitution by *Hi*AXHd3 is likely an early step in degradation of arabinoxylan by *H. insolens*. Removal of the α-1,3-linked component of the double substitution leaves a roughly equal mix of singly substituted O2 and O3-linked arabinose side chains. The organism possesses a GH51 α-L-arabinofuranosidase which shows synergy with *Hi*AXHd3 and has the capacity to catalyse removal of α-

1,2 and α-1,3-linked singly substituted side chains (Sorensen et al., 2006). This flexible, synergistic GH51 is also seen in *Bifidobacterium adolescentis*, which expresses the only other characterised AXHd3 enzyme. The stages involved in the subsequent metabolism of de-branched arabinoxylan are currently unclear, as the *H. insolens* genome has yet to be fully sequenced. According to the CAZy database (www.cazy.org), *H. insolens* possesses xylanases from families GH10 and GH11 (Dalboge and Heldt-Hansen, 1994). These enzymes will degrade xylan into shorter oligosaccharides, which may be metabolised by other downstream enzymes. More information is available on the metabolic capacity of *B. adolescentis*, due to a fully sequenced genome and several characterised enzymes. *B. adolescentis* is able to grow on arabinoxylooligosaccharides as the sole nutrient source due to the combined degradative action of *Ba*AXHd3, the synergistic GH51, and a GH8 reducing end xylose-releasing *exo*-oligoxylanase (Lambertus et al., 2005). Thus, *B. adolescentis* can utilise oligosaccharides generated from arabinoxylan but does not appear to have the capacity to degrade the polysaccharide backbone. Other bacteria found in the gut, such as *Bacteroides ovatus*, do have this capacity (Cooper et al., 1985; Weaver et al., 1992), suggesting that *B. adolescentis* utilises oligosaccharides released by other commensal species. The xylanase enzymes so far identified in *H. insolens* indicate that the fungus does not display this dependency upon another species. Nonetheless, *H. insolens* and *B. adolescentis* do display convergence as they remove arabinose decorations from xylooligosaccharides via the same mechanism.

5.2.1.ii A single mutation introduces dual functionality to *Hi*AXHd3

In mutagenesis studies of *Hi*AXHd3, one mutation in particular gave very interesting results. The mutant Y165A retained specificity for the O3-linked arabinose of the double substitution, albeit at reduced catalytic efficiency, but also gained xylanase activity. Kinetic analysis of the mutant (Section 4.5) showed that it possesses six subsites for xylanase activity, of which five are critical. Further mutagenesis studies confirmed that the same catalytic apparatus (Asp42, Asp166 and Glu215) is responsible for both the arabinofuranosidase and xylanase activities of the Y165A variant of *Hi*AXHd3.

Structural analysis of Y165A shows that the heart-shaped pocket of the wildtype enzyme has been altered to reveal an open cleft topology (Figure 4.27). Removal of the phenolic side chain of Tyr165 disrupts the face of the pocket directly below the glycosidic oxygen between the +1 and +2NR xylose residues. Subsequent mutagenesis studies revealed that introducing the mutations N183A and F492A into the Y165A variant significantly increase xylanase activity (Table 4.6). The Y165A/F492A mutant in particular is of interest, as it displays greater xylanase and arabinofuranosidase activity than the single Y165A mutant. Furthermore, it is possible to convert the enzyme to only display xylanase activity: the Y165A/H271A mutant released no arabinose from arabinoxylan but showed xylanase activity roughly comparable to the Y165A single mutant.

5.2.2 Future work

The crystal structures, kinetic analysis, mutagenesis and NMR experiments detailed in Chapter Four give insight into the mechanisms underlying the specificity displayed by *Hi*AXHd3. Nonetheless, gaps in our understanding do persist, which warrant further examination.

Mutagenesis studies, coupled with NMR experiments and structural data, have demonstrated that Trp525 is critical to specificity in arabinoxylan. The same information is not available for *Hi*AXHd3 activity on arabinan; contacts with this polysaccharide are likely to be different than for xylan, as the backbone adopts a very different conformation (Figure 4.1). A crystal structure is required to determine whether Trp525 makes the same vital interactions with this polysaccharide. For this reason, a structure of *Hi*AXHd3 in complex with a decorated arabinooligosaccharide is currently being sought by members of the Gilbert laboratory at Newcastle University. This structure should also shed light on why mutations of Tyr165 (Y165F and Y165W) seem to affect arabinofuranosidase activity differently on arabinoxylan and arabinan (Table 4.4).

A crystal structure is also currently being pursued of *Hi*AXHd3 variant Y165A in complex with a xylooligosaccharide. Attempts to date have been hampered by the presence of Bis-Tris propane in the active site of the protein. With crystals of the wildtype enzyme, a molecule of Tris was present in the active site, but this was

displaced in some crystals by ligand during soaking experiments. However, the Bis-Tris propane found in the Y165A active site was not displaced by ligand in soaking experiments. Attempts will therefore be made to crystallise Y165A in the absence of any Tris buffers.

Mutagenesis of the Y165A variant has revealed two additional mutations (N183A and F492A) which have increased xylanase activity while retaining arabinofuranosidase activity. Despite this, while the mutant demonstrates an exciting dual function, it remains only a weakly active enzyme. Further efforts to improve activity could be undertaken by error-prone PCR to introduce random changes to the cleft structure of the enzyme. In addition, a triple mutant of Y165A/N183A/F492A should show the improvements displayed by both double mutants (Y165A/N183A and Y165A/F492A).

5.2.3 Implications

Comparing *Hi*AXHd3 with the GenBank database showed that in the 50 proteins that display the closest sequence identity, the five key specificity determinants (Tyr165, His271, Asp290, Arg296 and Trp525) are highly conserved; Arg296 is invariant while Tyr165, His271, Asp290 and Trp525 are conserved in 94 %, 84 %, 92 % and 90 % of the proteins, respectively. It is likely that the majority, if not all, of these enzymes are AXHd3s, particularly when one considers that there is a lack of conservation of His271 and Trp525 in the only other confirmed AXHd3 (Lambertus et al., 2005).

The Y165A mutation of *Hi*AXHd3 shows that a glycoside hydrolase from family GH43 with very tight specificity for an arabinose residue in an unusual structural context (the double substitution) can be altered to display xylanase activity while retaining wildtype specificity. As was discussed above, arabinoxylan is potentially a very important industrial raw material, but economic utilisation of this polysaccharide is hindered by the production costs involved in producing the consortium of enzymes required for complete saccharification. The double substitution in particular has long been problematic, as most arabinofuranosidase enzymes are inhibited by this structure. For this reason, wildtype *Hi*AXHd3 is of industrial significance as the first stage in arabinoxylan degradation, as removal of the double substitution permits other arabinofuranosidases to hydrolyse the remaining single substitutions, after

which xylanases can hydrolyse the backbone. The use of a dual function enzyme such as Y165A reduces costs involved in enzyme production and so can facilitate the economic utilisation of the plant cell wall. The enzyme is weakly acting, but activity could potentially be improved (Section 5.2.2). In addition, the Y165A mutant serves as proof of concept that a GH43 can be engineered quite easily to perform a dual function.

BIBLIOGRAPHY

Adams, E.L., Kroon, P.A., Williamson, G., Gilbert, H.J., and Morris, V.J. (2004). Inactivated enzymes as probes of the structure of arabinoxylans as observed by atomic force microscopy. Carbohydrate Research *339*, 579-590.

Aleshin, A.E., Firsov, L.M., and Honzatko, R.B. (1994). Refined structure for the complex of acarbose with glucoamylase from *Aspergillus awamori* var. X100 to 2.4-A resolution. Journal of Biological Chemistry *269*, 15631-15639.

Alhasid, A., Ben-David, A., Tabachnikov, O., Wolf, D., Naveh, E., Zolotnitsky, G., Shoham, Y., and Shoham, G. (2009). Crystal structure of an inverting GH43 1,5-α-L-arabinanase from *Geobacillus stearothermophilus* complexed with its substrate. Biochemical Journal *422*, 73-82.

Arnold, K., Bordoli, L., Kopp, J., and Schwede, T. (2006). The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. Bioinformatics *22*, 195-201.

Asp, N.G. (1987). Dietary fibre--definition, chemistry and analytical determination. Molecular Aspects of Medicine *9*, 17-29.

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

Barnett, C.B., Wilkinson, K.A., and Naidoo, K.J. (2010). Pyranose ring transition state is derived from cellobiohydrolase I induced conformational stability and glycosidic bond polarization. Journal of the American Chemical Society *132*, 12800-12803.

Battye, T.G., Kontogiannis, L., Johnson, O., Powell, H.R. and Leslie, A.G. (2011) iMosflm: a new graphical interface for diffraction-image processing with Mosflm. Acta Crystallographica D *67*, 271-281.

Beaugerie, L., and Petit, J.-C. (2004). Antibiotic-associated diarrhoea. Clinical Gastroenterology *18*, 337-352.

Beloqui, A., Nechitaylo, T.Y., López-Cortés, N., Ghazi, A., Guazzaroni, M.E., Polaina, J., Strittmatter, A.W., Reva, O., Waliczek, A., Yakimov, M.M.,Golyshina, O.V., Ferrer, M. and Golyshin, P.N. (2010). Diversity of glycosyl hydrolases from cellulose-depleting communities enriched from casts of two earthworm species. Applied and Environmental Microbiology *76*, 5934-5946.

Beylot, M.-H., Emami, K., McKie, V.A., Gilbert, H.J., and Pell, G. (2001a). *Pseudomonas cellulosa* expresses a single membrane-bound glycoside hydrolase family 51 arabinofuranosidase. Biochemical Journal *358*, 599-605.

Beylot, M.-H., McKie, V.A., Voragen, A.G.J., Doeswick-Voragen, C.H.L., and Gilbert, H.J. (2001b). The *Pseudomonas cellulosa* glycoside hydrolase family 51 arabinofuranosidase exhibits wide substrate specificity. Biochemical Journal *358*, 607-614.

Bjursell, M.K., Martens, E.C., and Gordon, J.I. (2006). Functional genomic and metabolic studies of the adaptations of a prominent adult human gut symbiont, *Bacteroides thetaiotaomicron*, to the suckling period. Journal of Biological Chemistry *281*, 36369-36279.

Boerjan, W., Ralph, J., and Baucher, M. (2003). Lignin biosynthesis. Annual Review of Plant Biology *54*, 519-546.

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

Brenner, D.M., and Chey, W.D. (2009). *Bifidobacterium infantis* 35624: a novel probiotic for the treatment of irritable bowel syndrome. Reviews in Gastroenterological Disorders *9*, 7-15.

Brenner, D.M., Moeller, M.J., Chey, W.D., and Schoenfeld, P.S. (2009). The utility of probiotics in the treatment of irritable bowel syndrome: a systematic review. American Journal of Gastroenterology *104*, 1033-1049.

Brooks, B.E., and Buchanan, S.K. (2008). Signaling mechanisms for activation of extracytoplasmic function (ECF) sigma factors. Biochimica and Biophysica Acta *1778*, 1930-1945.

Brunzelle, J.S., Jordan, D.B., McCaslin, D.R., Olczak, A., and Wawrzak, Z. (2008). Structure of the two subsite β-D-xylosidase from *Selenomonas ruminantium* in complex with 1,3-bis[tris(hydroxymethyl)methylamino]propane. Archives of Biochemistry and Biophysics *474*, 157-166.

Brux, C., Ben-David, A., Shallom-Shezif, D., Leon, M., Niefind, K., Shoham, G., Shoham, Y., and Schomburg, D. (2006). The structure of an inverting GH43 β-xylosidase from *Geobacillus stearothermophilis* with its substrate reveals the role of the three catalytic residues. Journal of Molecular Biology *359*, 97-109.

Bullock, W.O., Fernandez, J.M., and Short, J.M. (1987). XL1-Blue: A high efficiency plasmid transforming recA *Escherichia coli* strain with β-galactosidase selection. Biotechniques *5*, 376-379.

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

Callebaut, I., Labesse, G., Durand, P., Poupon, A., Canard, L., Chomilier, J., Henrissat, B., and Mornon, J.P. (1997). Deciphering protein sequence information through hydrophobic cluster analysis (HCA): current status and perspectives. Cellular and Molecular Life Sciences *53*, 621-645.

Cantarel, B.L., Coutinho, P.M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009). The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics. Nucleic Acids Research *37*, 233-238.

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

Cartmell, A. (2010). Structure and function of novel cellulosic, hemicellulosic and pectic glycoside hydrolases. PhD thesis. Institute of Cell and Molecular Biosciences (Newcastle upon Tyne, Newcastle University).

Chang, D.E., Smalley, D.J., Tucker, D.L., Leatham, M.P., Norris, W.E., Stevenson, S.J., Anderson, A.B., Grissom, J.E., Laux, D.C., Cohen, P.S. and Conway, T. (2004). Carbon nutrition of *Escherichia coli* in the mouse intestine. Proceedings of the National Academy of Science USA *101*, 7427-7432.

Chanliaud, E., Saulnier, L., and Thibault, J.-F. (1995). Alkaline extraction and characterisation of heteroxylans from maize bran. Journal of Cereal Science *21*, 195-203.

Charnock, S.J., Spurway, T.D., Xie, H., Beylot, M.-H., Virden, R., Warren, R.A., Hazlewood, G.P., and Gilbert, H.J. (1998). The topology of the substrate binding clefts of glycosyl hydrolase family 10 xylanases are not conserved. Journal of Biological Chemistry *273*, 32187-32199.

Cohen, S.N., Chang, A.C., and Hsu, L. (1972). Nonchromosomal antibiotic resistance in bacteria: genetic transformation of *Escherichia coli* by R-factor DNA. Proceedings of the National Academy of Science USA *69*, 2110-2114.

Cooper, S.W., Pfeiffer, D.G., and Tally, F.P. (1985). Evaluation of xylan fermentation for the identification of *Bacteroides ovatus* and *Bacteroides thetaiotaomicron*. Journal of Clinical Microbiology *22*, 125-126.

Coughlan, M.P. (1985). In Biotechnology and Genetic Engineering Reviews, G.E. Russell, ed. (Paris, Intercept), pp. 39-109.

Cowtan, K. (2006). The Buccaneer software for automated model building. Acta Crystallographica D *62*, 1002-1011.

Dalboge, H., and Heldt-Hansen, H.P. (1994). A novel method for efficient expression cloning of fungal enzyme genes. Molecular and General Genetics *243*, 253-260.

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

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

Davin, L.B., and Lewis, N.G. (2005). Lignin primary structures and dirigent sites. Current Opinion in Biotechnology *16*, 407-415.

Dervilly, G., Saulnier, L., Roger, P., and Thibault, J. (2000). Isolation of homogeneous fractions from wheat water-soluble arabinoxylans. Influence of the structure on their macromolecular characteristics. Journal of Agricultural and Food Chemistry *48*, 270-278.

Ducros, V.M., Zechel, D.L., Murshudov, G.N., Gilbert, H.J., Szabo, L., Stoll, D., Withers, S.G., and Davies, G.J. (2002). Substrate distortion by a β-mannanase: snapshots of the Michaelis and covalent-intermediate complexes suggest a $B_{2,5}$ conformation for the transition state. Angewendte Chemie Internationl Edition, English *41*, 2824-2827.

Ekborg, N.A., Gonzalez, J.M., Howard, M.B., Taylor, L.E., Hutcheson, S.W., and Weiner, R.M. (2005). *Saccharophagus degradans* gen. nov., sp. nov., a versatile marine degrader of complex polysaccharides. International Journal of Systematic and Evolutionary Microbiology *55*, 1545-1549.

Elkins, J.G., Raman, B., and Keller, M. (2010). Engineered microbial systems for enhanced conversion of lignocellulosic biomass. Current Opinion in Biotechnology *21*, 657-662.

Emanuelsson, O., Brunak, S., Heijne, G.v., and Nielsen, H. (2007). Locating proteins in the cell using TargetP, SignalP, and related tools. Nature Protocols *2*, 953-971.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallographica D *66*, 486-501.

Eudes, R., Tuan, K.l., Delettre, J., Mornon, J.P., and Callebaut, I. (2007). A generalized analysis of hydrophobic and loop clusters within globular protein sequences. BMC Structural Biology *7*.

Evans, P.R. (2005) Scaling and assessment of data quality. Acta Crystallographica D *62*, 72-82.

Fry, S.C., Aldington, S., Hetherington, P.R., and Aitken, J. (1993). Oligosaccharides as signals and substrates in the plant cell wall. Plant Physiology *103*, 1-5.

Fujimara, K.E., Slusher, N.A., and Cabana, M.D. (2010). Role of the gut microbiota in defining human health. Expert Review of Anti-Infective Therapy *8*, 435-454.

Fujimoto, Z., Ichinose, H., Maehara, T., Honda, M., Kitaoka, M., and Kaneko, S. (2010). Crystal structure of an *exo*-1,5-α-L-arabinofuranosidase from *Streptomyces avermitilis* provides insights into the mechanism of substrate discrimination between *exo*- and *endo*-type enzymes in glycoside hydrolase family 43. Journal of Biological Chemistry *285*, 34134-34143.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., and Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server;. In The Proteomics Protocols Handbook, J.M. Walker, ed. (Humana Press), pp. 571-607.

Gerbens-Leenes, W., Hoekstra, A.Y., and Meer, T.H.v.d. (2009). The water footprint of bioenergy. Proceedings of the National Academy of Science USA *106*, 10219-10223.

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

Gilbert, H.J., Stalbrand, H., and Brumer, H. (2008). How the walls come crumbling down: recent structural biochemistry of plant polysaccharide degradation. Current Opinion in Plant Biology *11*, 338-348.

Gill, S.C., and Hippel, P.H.v. (1989). Calculation of protein extinction coefficients from amino acid sequence data. Analytical Biochemistry *182*, 319-326.

Goldstein, E.J., Citron, D.M., Merriam, C.V., and Abramson, M.A. (2009). Infection after elective colorectal surgery: bacteriological analysis of failures in a randomized trial of cefotetan vs. ertapenem prophylaxis. Surgical Infections *10*, 111-118.

Gong, C.S., Cao, N.J., Du, J., and Tsao, G.T. (1999). Ethanol production from renewable resources. Advances in Biochemical Engineering and Biotechnology *65*, 207-241.

Gosalbes, M.J., Perez-Gonzalez, J.A., Gonzalez, R., and Navarro, A. (1991). Two β-glycanase genes are clustered in *Bacillus polymyxa*: molecular cloning, expression and sequence analysis of genes encoding a xylanase and an *endo*-β-1,3-1,4-glucanase. Journal of Bacteriology *173*, 7705-7710.

Gronlund, M.M., Gueimonde, M., Laitinen, K., Kociubinski, G., Gronroos, T., Salminen, S., and Isolauri, E. (2007). Maternal breast-milk and intestinal bifidobacteria guide the

compositional development of the bifidobacterium microbiota in infants at risk of allergic disease. Clinical and Experimental Allergy *37*, 1764-1772.

Guarner, F., and Malagelada, J.-R. (2003). Gut flora in health and disease. The Lancet *361*, 512-519.

Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Systematic Biology *52*, 696-704.

Harris, G.W., Jenkins, J.A., Connerton, I., and Pickersgill, R.W. (1996). Refined crystal structure of the catalytic domain of xylanase A from *Pseudomonas fluorescens* at 1.8A resolution. Acta Crystallographica D *52*, 393-401.

Henrissat, B. (1991). A classification of glycosyl hydrolases based on amino acid sequence similarities. Biochemical Journal *280*, 309-316.

Henrissat, B., and Bairoch, A. (1996). Updating the sequence-based classification of glycosyl hydrolases. Biochemical Journal *316*, 695-696.

Henrissat, B., Callebaut, I., Fabrega, S., Lehn, P., Mornon, J.P., and Davies, G.J. (1995). Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases. Proceedings of the National Academy of Science USA *92*, 7090-7094.

Henrissat, B., Claeyssens, M., Tomme, P., Lemesle, L., and Mornon, J.P. (1989). Cellulase families revealed by hydrophobic cluster analysis. Gene *81*, 83-95.

Hill, J., Nelson, E., Tilman, D., Polasky, S., and Tiffany, D. (2006). Environmental, economic and energetic costs and benefits of biodiesel and ethanol biofuels. Proceedings of the National Academy of Science USA *25*, 11206-11210.

Himmel, M.E., Ruth, M.F., and Wyman, C.E. (1999). Cellulase for commodity products from cellulosic biomass. Current Opinion in Biotechnology *10*, 358-364.

Hinz, S.W.A., Verhoef, R., Schols, H.A., Vincken, J.-P., and Voragen, A.G.J. (2005). Type I arabinogalactan contains β-D-Gal*p*-(1-->3)-β-D-Gal*p* structural elements. Carbohydrate Research *340*, 2135-2143.

Hovel, K., Shallom, D., Niefind, K., Belakhov, V., Shoham, G., Baasov, T., Shoham, Y., and Schomburg, D. (2003). Crystal structure and snapshots along the reaction pathway of a family 51 α-L-arabinofuranosidase. EMBO Journal *22*, 4822-4932.

Ichinose, H., Yoshida, M., Kotake, T., Kuno, A., Igarashi, K., Tsumuraya, Y., Samejima, M., Hirabayashi, J., Kobayashi, H., and Kaneko, S. (2005). An *exo* β-1,3-galactanase having a

novel β-1,3-galactan binding module from *Phanerochaete chrysosporium*. Journal of Biological Chemistry *280*, 25820-25829.

Ishii, T., Matsunaga, T., and Hayashi, N. (2001). Formation of Rhamnogalacturonan II-Borate Dimer in Pectin Determines Cell Wall Thickness of Pumpkin Tissue. Plant Physiology *126*, 1698-1705.

Ishii, T., Matsunaga, T., Pellerin, P., O'Neill, M.A., Darvill, A., and Albersheim, P. (1999). The Plant Cell Wall Polysaccharide Rhamnogalacturonan II Self-assembles into a Covalently Cross-linked Dimer. Journal of Biological Chemistry *274*, 13098-13104.

Izydorczyk, M.S., and Biliaderis, C.G. (1995). Cereal arabinoxylans: advances in structure and physicochemical properties. Carbohydrate Polymers *28*, 33-48.

Janaswamy, S., and Chandrasekaran, R. (2005). Polysaccharide structures from powder diffraction data: molecular models of arabinan. Carbohydrate Research *340*, 835-839.

Kadner, R.J., Chimento, D.P., and Wiener, M.C. (2003). The *Escherichia coli* outer membrane cobalamin transporter BtuB: structural analysis of calcium and substrate binding, and identification of orthologous transporters by sequence/structure conservation. Journal of Molecular Biology *332*, 999-1014.

Karaveg, K., Siriwardena, A., Tempel, W., Liu, Z.-J., Glushka, J., Wang, B.-C., and Moremen, K.W. (2005). Mechanism of Class 1 (Glycosylhydrolase Family 47) α-Mannosidases Involved in N-Glycan Processing and Endoplasmic Reticulum Quality Control. Journal of Biological Chemistry *280*, 16197-16207.

Kataeva, I.A., III, R.D.S., Shah, A., West, L.T., Li, X.L., and Ljungdahl, L.G. (2002). The fibronectin type 3-like repeat from the *Clostridium thermocellum* cellobiohydrolase CbhA promotes hydrolysis of cellulose by modifying its surface. Applied and Environmental Microbiology *68*, 4292-4300.

Kauppinen, S., Christgau, S., Kofod, L.V., Halkier, T., Dörreich, K., and Dalbøge, H. (1995). Molecular cloning and characterization of a rhamnogalacturonan acetylesterase from *Aspergillus aculeatus*. Synergism between rhamnogalacturonan degrading enzymes. Journal of Biological Chemistry *270*, 27172-27178.

Kellett, L.E., Poole, D.M., Ferreira, L.M., Durrant, A.J., Hazlewood, G.P., and Gilbert, H.J. (1990). Xylanase B and an arabinofuranosidase from *Pseudomonas fluorescens* subsp. *cellulosa* contain identical cellulose-binding domains and are encoded by adjacent genes. Biochemical Journal *272*, 369-376.

Kelly, M.A., Sinnott, M.L. and Herchen, M. 1987 Purification and mechanistic properties of an extracellular α-L-arabinofuranosidase from *Monilinia fructigena*. Biochemical Journal *245*, 843-849.

Kim, Y., Hendrickson, R., Mosier, N.S., and Ladisch, M.R. (2009). Liquid hot water pretreatment of cellulosic biomass. Methods in Molecular Biology *581*, 93-102.

Kirschner, K.N., Yongye, A.B., Tschampel, S.M., González-Outeiriño, J., Daniels, C.R., Foley, B.L., and Woods, R.J. (2008). GLYCAM06: a generalizable biomolecular force field. Carbohydrates. Journal of Computational Chemistry *29*, 622-655.

Knox, J.P. (2008). Revealing the structural and functional diversity of plant cell walls. Current Opinion in Plant Biology *11*, 1-6.

Kobayashi, M., Matou, T., and Azuma, J.-I. (1996). Two Chains of Rhamnogalacturonan II Are Cross-Linked by Borate-Diol Ester Bonds in Higher Plant Cell Walls. Plant Physiology *110*, 1017-1020.

Kotake, T., Kitazawa, K., Takata, R., Okabe, K., Ichinose, H., Kaneko, S., and Tsumuraya, Y. (2009). Molecular cloning and expression in *Pichia pastoris* of a *Irpex lacteus exo*-β-(1-->3)-galactanase gene. Bioscience, Biotechnology and Biochemistry *73*, 2303-2309.

Kumar, R., Singh, S., and Singh, O.V. (2008). Bioconversion of lignocellulosic biomass: biochemical and molecular perspectives. Journal of Industrial Microbiology and Biotechnology *35*, 377-391.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature *227*, 680-685.

Laere, K.M.v., Beldman, G., and Voragen, A.G. (1997). A new arabinofuranohydrolase from *Bifidobacterium adolescentis* able to remove arabinosyl residues from double-substituted xylose unites in arabinoxylan. Applied Microbiology and Biotechnology *47*, 231-235.

Lal, R. (2008). Crop residues as soil amendments and feedstock for bioethanol production. Waste Management *28*, 747-758.

Lambertus, A., Broek, M.v.d., Lloyd, R.M., Beldman, G., Verdoes, J.C., McCleary, B.V., and Voragen, A.G.J. (2005). Cloning and characterisation of arabinoxylan arabinofuranohydrolase-D3 (AXHd3) from *Bifidobacterium adolescentis* DSM20083. Applied Microbiology and Biotechnology *67*, 641-647.

Lampky, J.R. (1971). Distribution of *Sorangium cellulosum*. Applied Microbiology *22*, 937-938.

Lamport, D.T.A. (1965). The protein component of primary cell walls. Advances in Botanical Research *2*, 151-218.

Langer, G., Cohen, S.X., Lamzin, V.S., and Perrakis, A. (2008). Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7. Nature Protocols *3*, 1171-1179.

Lazarevic, V., Whiteson, K., Huse, S., Hernandez, D., Farinelli, L., Osterås, M., Schrenzel, J., and François, P. (2009). Metagenomic study of the oral microbiota by Illumina high-throughput sequencing. Journal of Microbiological Methods *79*, 266-271.

Lempereur, I., Rouau, X., and Abecassis, J. (1997). Genetic and Agronomic Variation in Arabinoxylan and Ferulic Acid Contents of Durum Wheat (*Triticum durum*.) Grain and Its Milling Fractions. Journal of Cereal Science *25*, 103-110.

Leslie, A.G.W. (1999) Integration of macromolecular diffraction data. Acta Crystallographica D *55*, 1696-1702.

Levigne, S.V., Ralet, M.-C.J., Quemener, B.C., Pollet, B.N.-L., Lapierre, C., and Thibault, J.-F.J. (2004). Isolation from Sugar Beet Cell Walls of Arabinan Oligosaccharides Esterified by Two Ferulic Acid Monomers. Plant Physiology *134*, 1173-1180.

Ley, R.E., Peterson, D.P., and Gordon, J.I. (2006). Ecological and evolutionary forces shaping microbial diversity in the human intestine. Cell *124*, 837-848.

Little, E., Bork, P., and Doolittle, R.F. (1994). Tracing the spread of fibronectin type III domains in bacterial glycohydrolases. Journal of Molecular Evolution *39*, 631-643.

Liu, C.J., Miao, Y.C., and Zhang, K.W. (2011). Sequestration and transport of lignin monomeric precursors. Molecules *18*, 710-727.

Martens, E.C., Chiang, H.C., and Gordon, J.I. (2008). Mucosal glycan foraging enhances fitness and transmission of a saccharolytic human gut bacterial symbiont. Cell Host and Microbe *4*, 447-457.

Martens, E.C., Roth, R., Heuser, J.E., and Gordon, J.I. (2009c). Coordinate Regulation of Glycan Degradation and Polysaccharide Capsule Biosynthesis by a Prominent Human Gut Symbiont. Journal of Biological Chemistry *284*, 18445-18457.

Martin-Garcia, R., Leon, N.d., Sharifmoghadam, M.R., Curto, M.A., Hoya, M., Bustos-Sanmamed, P., and Valdivieso, M.H. (2010). The FN3 and BRCT motifs in the exomer component Chs5p define a conserved module that is necessary and sufficient for its function. Cellular and Molecular Life Sciences. *In press*.

Martinelli, L.A., and Filoso, S. (2008). Expansion of sugarcane ethanol production in Brazil: environmental and social challenges. Ecological Applications *18*, 885-898.

Matsunaga, T., Ishii, T., Matsumoto, S., Higuchi, M., Darvill, A., Albersheim, P., and O'Neill, M.A. (2004). Occurrence of the Primary Cell Wall Polysaccharide Rhamnogalacturonan II in Pteridophytes, Lycophytes, and Bryophytes. Implications for the Evolution of Vascular Plants. Plant Physiology *134*, 339-351.

McCann, M.C., and Carpita, N.C. (2008). Designing the deconstruction of plant cell walls. Current Opinion in Plant Biology *11*, 314-320.

McCarter, J.D., and Withers, S.G. (1994). Mechanisms of enzymatic glycoside hydrolysis. Current Opinion in Structural Biology *4*, 885-892.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Staroni, L.C. and Read R.J. (2007) *Phaser* crystallographic software. Journal of Applied Crystallography *40*, 658-674.

McKee, L.S.*, Cartmell, A*, Peña, M.J., Larsbrink, J, Brumer, H, Lewis, R.J., Vikso-Nelson, A, Ichinose, H, Kaneko, S, Gilbert, H.J. and Marles-Wright, J. (2011) The structure and function of an arabinan-specific α-1,2-arabinofuranosidase identified from screening the activities of bacterial glycoside hydrolases. Journal of Biological Chemistry *286*, 15483-15495. *contributed equally.

McKie, V.A., Black, G.W., Millward-Sadler, S.J., Hazlewood, G.P., Laurie, J.I., and Gilbert, H.J. (1997). Arabinanase A from *Pseudomonas fluorescens* subsp. *cellulosa* exhibits both an *endo*- and and *exo*- mode of action. Biochemical Journal *323*, 547-555.

McNeil, M., Darvill, A.G., Fry, S.C., and Albersheim, P. (1984). Structure and function of primary cell walls of plants. Annual Review of Biochemistry *53*, 625-663.

Melo F and Feytmans E. (1998). Assessing protein structures with a non-local atomic interaction energy. Journal of Molecular Biology *17*, 1142-1152.

Meyers, J.A., Sanchez, D., Elwell, L.P., and Falkow, S. (1975). Simple agarose gel electrophoresis method for the identification and characterisation of plasmid deoxyribonucleic acid. Biochemical Journal *127*, 1529-1538.

Miller, G.L. (1959). The use of dinitrosalicylic acid for the determination of reducing sugar. Analytical Chemistry *31*, 426-428.

Mohnen, D. (2008a). Pectin structure and biosynthesis. Current Opinion in Plant Biology *11*, 266-277.

Mohnen, D., Bar-Peled, M., and Somerville, C. (2008b). Biosynthesis of plant cell walls. In Biomass Recalcitrance, M. Himmel, ed. (Oxford, Blackwell Publishing), pp. 94-187.

Mullis, K.B., and Faloona, F.A. (1987). Specific synthesis of DNA *in vitro* via a polymerase-catalysed chain reaction. Methods in Enzymology *155*, 335-350.

Murshudov, G.N., Vagin, A.A. and Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallographica D *53*, 540-255.

Neer, E.J., and Smith, T.F. (1996). G protein heterodimers: new structures propel new questions. Cell *84*, 175-178.

Nofrarías, M., Martínez-Puig, D., Pujols, J., Majó, N., and Pérez, J.F. (2007). Long-term intake of resistant starch improves colonic mucosal integrity and reduces gut apoptosis and blood immune cells. Nutrition *23*, 861-870.

Notenboom, V., Birsan, C., Nitz, M., Rose, D.R., Warren, R.A., and Withers, S.G. (1998). Insights into transition state stabilization of the β-1,4-glycosidase Cex by covalent intermediate accumulation in active site mutants. Nature Structural Biology *5*, 812-818.

Nurizzo, D., Turkenburg, J.P., Charnock, S.J., Roberts, S.M., Dodson, E.J., McKie, V.A., Taylor, E.J., Gilbert, H.J., and Davies, G.J. (2002). *Cellvibrio japonicus* α-L-arabinanase 43A has a novel five-blade β-propeller fold. Nature Structural Biology *9*, 665-668.

O'Hara, A.M., and Shanahan, F. (2006). The gut flora as a forgotten organ. EMBO Reports *7*, 688-693.

O'Keefe, S.J. (2010). Tube feeding, the microbiota, and Clostridium difficile infection. World Journal of Gastoenterology *14*, 139-142.

O'Neill, M.A., Warrenfeltz, D., Kates, K., Pellerin, P., Doco, T., Darvill, A.G., and Albersheim, P. (1996). Rhamnogalacturonan-II, a Pectic Polysaccharide in the Walls of Growing Plant Cell, Forms a Dimer That Is Covalently Cross-linked by a Borate Ester. *In Vitro* Conditions For The Formation And Hydrolysis Of The Dimer. Journal of Biological Chemistry *271*, 22923-22930.

Ohland, C.L., and Macnaughton, W.K. (2010). Probiotic bacteria and intestinal epithelial barrier function. American Journal of Physiology G *298*, 807-819.

Otwinowski, Z., and Minor, W. (1997). Processing of x-ray diffraction data collected in oscillation mode: Macromolecular crystallography Part A. Methods in Enzymology *276*, 307-326.

Pace, C.N., Vajdos, F., Fee, L., Grimsley, G., and Gray, T. (1995). How to measure and predict the molar adsorption coefficient of a protein. Protein Science *4*, 2411-2423.

Palma, G.d., Cinova, J., Stepankova, R., Tuckova, L., and Sanz, Y. (2010). Pivotal Advance: *Bifidobacteria* and Gram-negative bacteria differentially influence immune responses in the proinflammatory milieu of celiac disease. Journal of Leukocyte Biology *87*, 765-778.

Park, S., Baker, J.O., Himmel, M.E., Parilla, P.A., and Johnson, D.K. (2010). Cellulose crystallinity index: measurement techniques and their impact on interpreting cellulase performance. Biotechnology for Biofuels *3*, 10.

Pauly, M., and Keegstra, K. (2010). Plant cell wall polymers as precursors for biofuels. Current Opinion in Plant Biology *13*, 1-8.

Pell, G., Szabo, L.S., Charnock, S.J., Xie, H., Gloster, T.M., Davies, G.J., and Gilbert, H.J. (2004b). Structural and biochemical analysis of *Cellvibrio japonicus* xylanase 10C: how variation in substrate-binding cleft influences the catalytic profile of family GH10 xylanases. Journal of Biological Chemistry *279*, 11777-11788.

Pell, G., Taylor, E.J., Gloster, T.M., Turkenburg, J.P., Fontes, C.M.G.A., Ferreira, L.M.A., Nagy, T., Clark, S.J., Davies, G.J., and Gilbert, H.J. (2004a). The mechanisms by which family 10 glycoside hydrolases bind decorated substrates. Journal of Biological Chemistry *279*, 9597-9605.

Pellerin, P., Doco, T., Vidal, S., Williams, P., Brillouet, J.M., and O'Neill, M.A. (1996). Structural characterization of red wine rhamnogalacturonan II. Carbohydrate Research *290*, 183-197.

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

Pons, T., Naumoff, D.G., Martinez-Fleites, C., and Hernandez, L. (2004). Three acidic residues are at the active site of a β-propeller architecture in glycoside hydrolase families 31, 43, 62 and 68. Proteins *54*, 424-432.

Proctor, M.R., Taylor, E.J., Nurizzo, D., Turkenburg, J.P., Lloyd, R.M., Vardakou, M., Davies, G.J., and Gilbert, H.J. (2005). Tailored catalysts for plant cell wall degradation: redesigning the *exo/endo* preference of *Cellvibrio japonicus* arabinanase 43A. Proceedings of the National Academy of Science USA *102*, 2697-2702.

Project, C.C. (1994). The CCP4 Suite: Programs for Protein Crystallography. Acta Crystallographica D *50*, 760-763.

Ragauskas, A., Williams, C., Davidson, B., Britovsek, G., Cairney, J., Eckert, C., Jr, W.F., Hallet, J., Leak, D., Liotta, C., Mielenz, J.R., Murphy, R., Templer, R. and Tschaplinski, T. (2006). The path forward for biofuels and biomaterials. Science *311*, 484-489.

Reeves, A.R., Wang, G.-R., and Salyers, A.A. (1997). Characterisation of four outer membrane proteins that play a role in utilisation of starch by *Bacteroides thetaiotaomicron*. Journal of Bacteriology *179*, 643-649.

Ringli, C., Keller, B., and Ryser, U. (2001). Glycine-rich proteins as structural components of plant cell walls. Cellular and Molecular Life Sciences *58*, 1430-1441.

Scheller, H.V., and Ulvskov, P. (2010). Hemicelluloses. Annual Review of Plant Biology *61*, 263-289.

Schwede, T., Kopp, J., Guex, N., and Peitsch, M.C. (2003). SWISS-MODEL: an automated protein homology-modeling server. Nucleic Acids Research *31*, 3381-3385.

Sears, C.L. (2005). A dynamic partnership: Celebrating our gut flora. Anaerobe *11*, 247-251.

Shallom, D., Leon, M., Ben-David, A., Zaide, G., Belakhov, V., Shoham, G., Schomburg, D., Baasov, T., and Shoham, Y. (2005). Biochemical characterisation and identification of the catalytic residues of a family 43 β-D-xylosidase from G*eobacillus stearothermophilus* T-6. Biochemistry *44*, 387-397.

Sheehan, J.J. (2009). Biofuels and the conundrum of sustainability. Current Opinion in Biotechnology *20*, 318-324.

Sheldrick, G.M. (2008). A short history of SHELX. Acta Crystallographica A *64*, 112-122.

Shipman, J.A., Berleman, J.E., and Salyers, A.A. (2000). Characterisation of four outer membrane proteins involved in binding starch to the cell surface of *Bacteroides thetaiotaomicron*. Journal of Bacteriology *182*, 5365-5372.

Showalter, A.M. (2001a). Introduction: Plant cell wall proteins. Cellular and Molecular Life Sciences *58*, 1361-1362.

Showalter, A.M. (2001b). Arabinogalactan-proteins: structure, expression and function. Cellular and Molecular Life Sciences *58*, 1399-1417.

Sidhu, G., Withers, S.G., Nguyen, N.T., McIntosh, L.P., Ziser, L., and Brayer, G.D. (1999). Sugar ring distortion in the glycosyl-enzyme intermediate of a family GH11 xylanase. Biochemistry *38*, 5346-5354.

Simmon, K.E., Mirrett, M., Reller, L.B., and Petti, C.A. (2008). Genotypic diversity of anaerobic isolates from bloodstream infections. Journal of Clinical Microbiology *46*, 1596-1601.

Sonnenburg, E.D., Sonnenburg, J.L., Manchester, J.K., Hansen, E.E., Chiang, H.C., and Gordon, J.I. (2006). A hybrid two-component system protein of a prominent human gut symbiont couples glycan sensing *in vivo* to carbohydrate metabolism. Proceedings of the National Academy of Science USA *103*, 8834-8839.

Sonnenburg, E.D., Zheng, H., Joglekar, P., Higginbottom, S.K., Firbank, S.J., and Sonnenburg, J.L. (2010). Specificity of polysaccharide use in intestinal *Bacteroides* species determines diet-induced microbiota alterations. Cell *141*, 1241-1252.

Sonnenburg, J.L., Xu, J., Leip, D.D., Chen, C.-H., Westover, B.P., Weatherford, J., Buhler, J.D., and Gordon, J.I. (2005). Glycan foraging *in vivo* by an intestine-adapted bacterial symbiont. Science *307*, 1955-1959.

Sorensen, H.R., Jorgensen, C.T., Hansen, C.H., Jorgensen, C.I., Pedersen, S., and Meyer, A.S. (2006). A novel GH43 α-L-arabinofuranosidase from *Humicola insolens*: mode of action and synergy with GH51 α-L-arabinofuranosidase on wheat arabinoxylan. Applied Microbiology and Biotechnology *73*, 850-861.

Spezio, M., Wilson, D.B., and Karplus, P.A. (1993). Crystal structure of the catalytic domain of a thermophilic endocellulase. Biochemistry *32*, 9906-9916.

Studier, F.W., and Moffatt, B.A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. Journal of Molecular Biology *189*, 113-130.

Swamy, N.R., and Salimath, P.V. (1991). Arabinans from *Cajanus cajan* cotyledon. Phytochemistry *30*, 263-265.

Tamura, K., Dudley, J., Nei, M., and Kumar, S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution *24*, 1596-1599.

Tannock, G.W. (1977). Characteristics of *Bacteroides* isolates from the cecum of conventional mice. Applied and Environmental Microbiology *33*, 745-750.

Taylor, E.J., Smith, N.L., Turkenburg, J.P., d'Souza, S., Gilbert, H.J., and Davies, G.J. (2006). Structural insight into the ligand specificity of a thermostable family 51 arabinofuranosidase, Ara*f*51, from C*lostridium thermocellum*. Biochemical Journal *395*, 31-37.

Tempel, W., Liu, Z.-J., Horanyi, P.S., Deng, L., Lee, D., Newton, M.G., Rose, J.P., Ashida, H., Li, S.-C., Li, Y.-T.*, et al.* (2005). Three-dimensional structure of GlcNAc-α1-4Gal releasing *endo*-beta-galactosidase from *Clostridium perfringens*. Proteins: Structure, Function and Genetics *59*, 141-144.

Thompson, D.S. (2005). How do cell walls regulate plant growth? Journal of Experimental Botany *56*, 2275-2285.

Thompson, J.D., Higgins, D.G., and Gibson, T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. . Nucleic Acids Research *22*, 4673-4680.

Tull, and Withers, S.G. (1994). Mechanisms of cellulases and xylanases: A detailed kinetic study of the *exo*-β-1,4-glycanase from *Cellulosa fimi*. Biochemistry *3*, 6363-6370.

Turner, S.R., and Somerville, C.R. (1997). Collapsed xylem phenotype of *Arabidopsis* identifies mutants deficient in cellulose deposition in the secondary cell wall. Plant Cell *9*, 689-701.

Vanderhoof, J.A. (2008). Probiotics in allergy management. Journal of Pediatric Gastoenterology and Nutrition *47*, 38-40.

Vandermarliere, E., Bourgois, T.M., Winn, M.D., Campenhout, S.v., Volckaert, G., Delcour, J.A., Strelkov, S.V., Rabijns, A., and Courtin, C.M. (2009). Structural analysis of a glycoside hydrolase family 43 arabinoxylan arabinofuranohydrolase in complex with xylotetraose reveals a different binding mechanism compared with other members of the same family. Biochemical Journal *418*, 39-47.

Varrot, A., Leydier, S., Pell, G., Macdonald, J.M., Stick, R.V., Henrissat, B., Gilbert, H.J., and Davies, G.J. (2005). *Mycobacterium tuberculosis* strains possess functional cellulases. Journal of Biological Chemistry *280*, 20181-20184.

Verhertbruggen, Y., Marcus, S.E., Haeger, A., Verhoef, R., Schols, H.A., McCleary, B.V., McKee, L., Gilbert, H.J., and Knox, J.P. (2009). Developmental complexity of arabinan polysaccharides and their processing in plant cell walls. Plant Journal *59*, 413-425.

Viëtor, R.J., Kormelink, F.J.M., Angelino, S.A.G.F., and Voragen, A.G.J. (1994). Substitution patterns of water-unextractable arabinoxylans from barley and malt. Carbohydrate Polymers *24*, 113-118.

Vignon, M.R., Heux, L., Malainine, M.E., and Mahrouz, M. (2004). Arabinan-cellulose composite in *Opuntia ficus-indica* prickly pear spines. Carbohydrate Research *339*, 123-131.

Vincent, B., Chen, W., III, B.A., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2010). MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallographica D *66*, 12-21.

Vocadlo, D.J., and Davies, G.J. (2008). Mechanistic insights into glycosidase chemistry. Current Opinion in Chemical Biology *12*, 539-555.

Weaver, J., Whitehead, T.R., Cotta, M.A., Valentine, P.C., and Salyers, A.A. (1992). Genetic Analysis of a Locus on the *Bacteroides ovatus* Chromosome Which Contains Xylan Utilization Genes. Applied and Environmental Microbiology *58*, 2764-2770.

Whorwell, P.J. (2009). Do probiotics improve symptoms in patients with irritable bowel syndrome? Therapeutic Advances in Gastroenterology *2*, 37-44.

Wicki, J., Schloegl, J., Tarling, C.A., and Withers, S.G. (2007). Recruitment of both uniform and differential binding energy in enzymatic catalysis: xylanases from families 10 and 11. Biochemistry *46*, 6996-7005.

Woodcock, S., Mornon, J.P., and Henrissat, B. (1992). Detection of secondary structure elements in proteins by hydrophobic cluster analysis. Protein Engineering *5*, 629-635.

Wu, H., Graaf, B.d., Mariani, C., and Cheung, A.Y. (2001). Hydroxyproline-rich glycoproteins in plant reproductive tissues: structure, functions and regulation. Cellular and Molecular Life Sciences *58*, 1418-1429.

Xu, J., and Gordon, J.I. (2003b). Honor thy symbionts. Proceedings of the National Academy of Science USA *100*, 10452-10459.

Xu, J., Mahowoald, M.A., Ley, R.E., Lozupone, C.A., Hamady, M., Martens, E.C., Henrissat, B., Coutinho, P.M., Minx, P., Latreille, P., Cordum, H., van Brunt, A. Kim, K., Fulton, R.S., Fulton, L.A., Clifton, S.W., Wilson, R.K., Knight, R.D. and Gordon, J.I. (2007). Evolution of symbiotic bacteria in the distal human intestine. PLoS Biology *5*, 1574-1586.

Yang, B., and Wyman, C.E. (2009). Dilute acid and autohydrolysis pretreatment. Methods in Molecular Biology *581*, 103-114.

Yoshida, S., Hespen, C.W., Beverly, R.L., MckieI, R.I., and Cann, K.O. (2010). Domain analysis of a modular α-L-arabinofuranosidase with a unique carbohydrate binding strategy from the fibre degrading bacterium *Fibrobacter succinogenese* S85. Journal of Bacteriology *192*, 5424-5236.

Yui, T., Imada, K., Shibuya, N., and Ogawa, K. (1995). Conformation of an arabinoxylan isolated from the rice endosperm cell wall by X-ray diffraction and a conformational analysis. Bioscience, Biotechnology and Biochemistry *59*, 965-968.

Zechel, D.L., and Withers, S.G. (2000). Glycosidase mechanisms: anatomy of a finely tuned catalyst. Accounts of Chemical Research *33*, 11-18.

Zhao, S., Wang, J., Bu, D., Liu, K., Zhu, Y., Dong, Z., and Yu, Z. (2010). Novel glycoside hydrolases identified by screening a Chinese Holstein dairy cow rumen-derived metagenome library. Applied and Environmental Microbiology *76*, 6701-6705.

Zhu, J.Y., Pan, X., and Jr, R.S.Z. (2010a). Pretreatment of woody biomass for biofuel production: energy efficiency, technologies, and recalcitrance. Applied Microbiology and Biotechnology *87*, 847-857.

Zhu, Y., Suits, M.D.L., Thompson, A.J., Chavan, S., Dinev, Z., Dumon, C., Smith, N., Moremen, K.W., Xiang, Y., Siriwardena, A., Williams, S.J, Gilbert, H.J and Davies, G.J. (2010b). Mechanistic insights into a Ca^{2+}-dependent family of α-mannosidases in a human gut symbiont. Nature Chemical Biology *6*, 125-132.

Zykwinska, A., Thibaulta, J.-F., and Ralet, M.-C. (2008). Competitive binding of pectin and xyloglucan with primary cell wall cellulose. Carbohydrate Polymers *74*, 957-961.

Zykwinska, A.W., Ralet, M.-C.J., Garnier, C.D., and Thibault, J.-F.J. (2005). Evidence for *In Vitro* Binding of Pectin Side Chains to Cellulose. Plant Physiology *139*, 397-407.

APPENDIX A

Materials and suppliers

A.1 Protein standards

Sigma low and high molecular weight standards:

Low molecular weight protein standards	**Molecular weight (kDa)**
Albumin, bovine	66
Albumin, egg	45
Glyceraldehyde-3-P dehydrogenase	36
Carbonic anhydrase, bovine	29
Trypsinogen, bovine	24
Pancreas trypsin inhibitor, soybean	20
α-lactalbumin, bovine milk	14.2

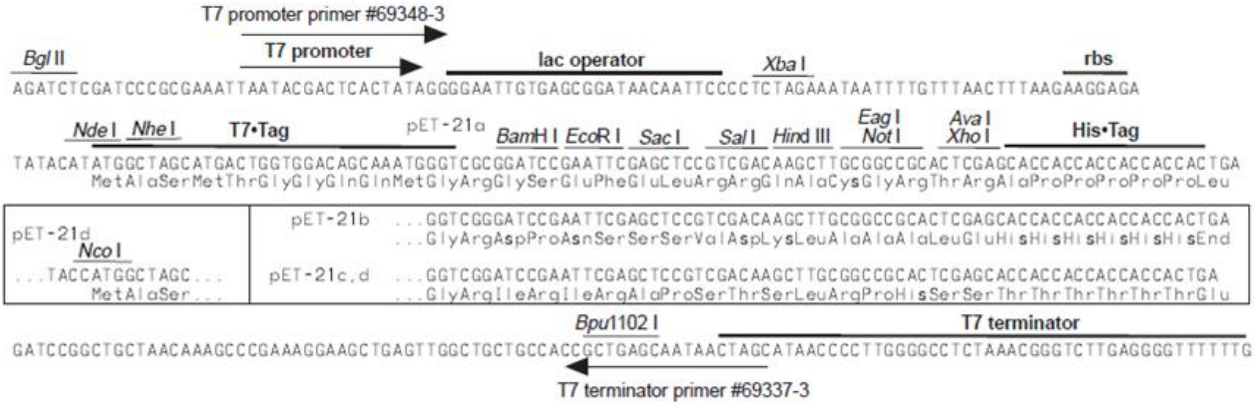
Table A.1 Sigma low molecular weight standards

High molecular weight protein standards	**Molecular weight (kDa)**
Myosin, rabbit muscle	205
B-galactosidase, *E. coli*	116
Glyceraldehyde-3-P dehydrogenase	97.4
Albumin, bovine	66
Albumin, egg	45
Carbonic anhydrase, bovine	29

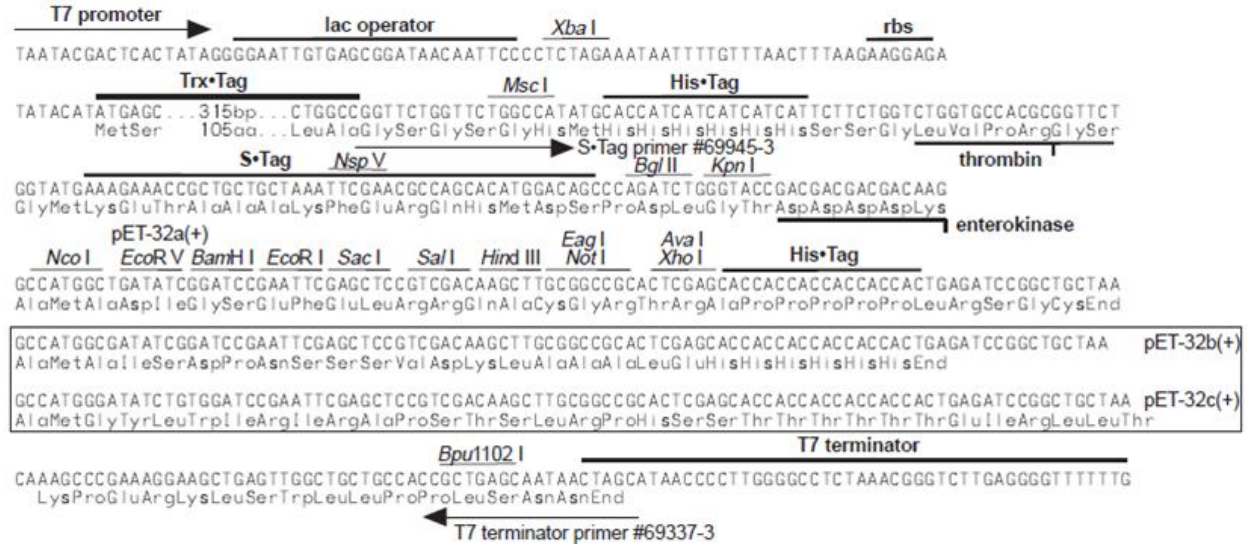
Table A.2 Sigma high molecular weight standards

A.2 Vector maps

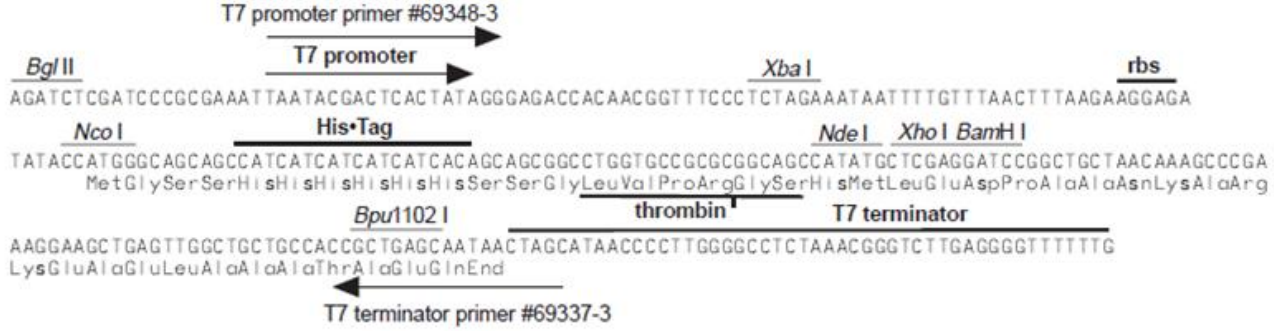
A.2.1 pET21a-d cloning/expression region



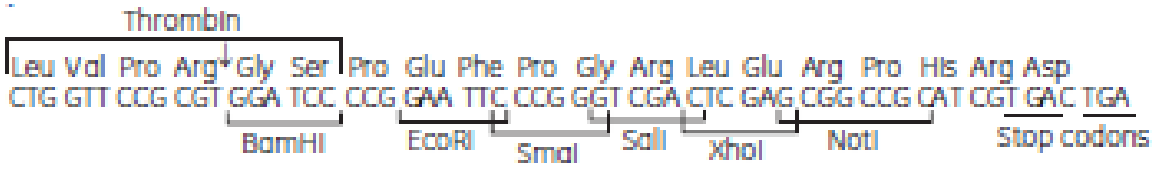
A.2.2 pET32a-c cloning/expression region



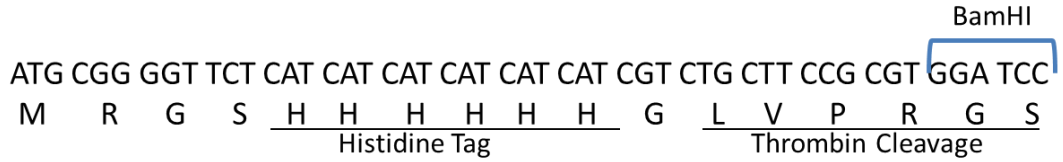
A.2.3 pET14b cloning/expression region

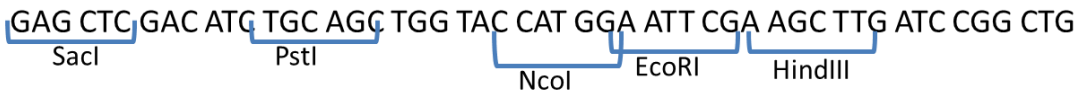


A.2.4 pGEX cloning/expression region



A.2.5 mini- pRSETA cloning/expression region

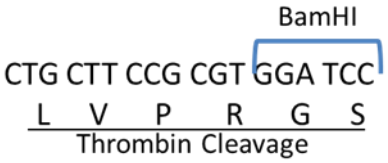


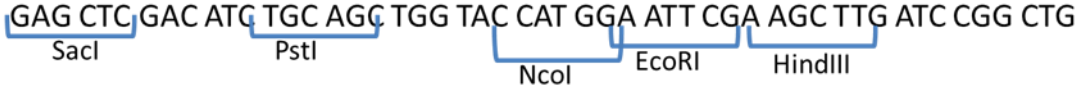


A.2.6 mini-pRGST cloning/expression region

Amino acid sequence of the glutathione-S-transferase fusion protein:

SPILGYWKIKGLVQPTRLLLEYLEEKYEEHLYERDEGDKWRNKKFELGLEFPNLPYY
IDGDVKLTQSMAIIRYIADKHNMLGGCPKERAEISMLEGAVLDIRYGVSRIAYSKDFE
TLKVDFLSKLPEMLKMFEDRLCHKTYLNGDHVTHPDFMLYDALDVVLYMDPMCLD
AFPKLVCFKKRIEAIPQIDKYLKSSKYIAWPLQGWQATFGGGDHPPKSDLVPR





APPENDIX B

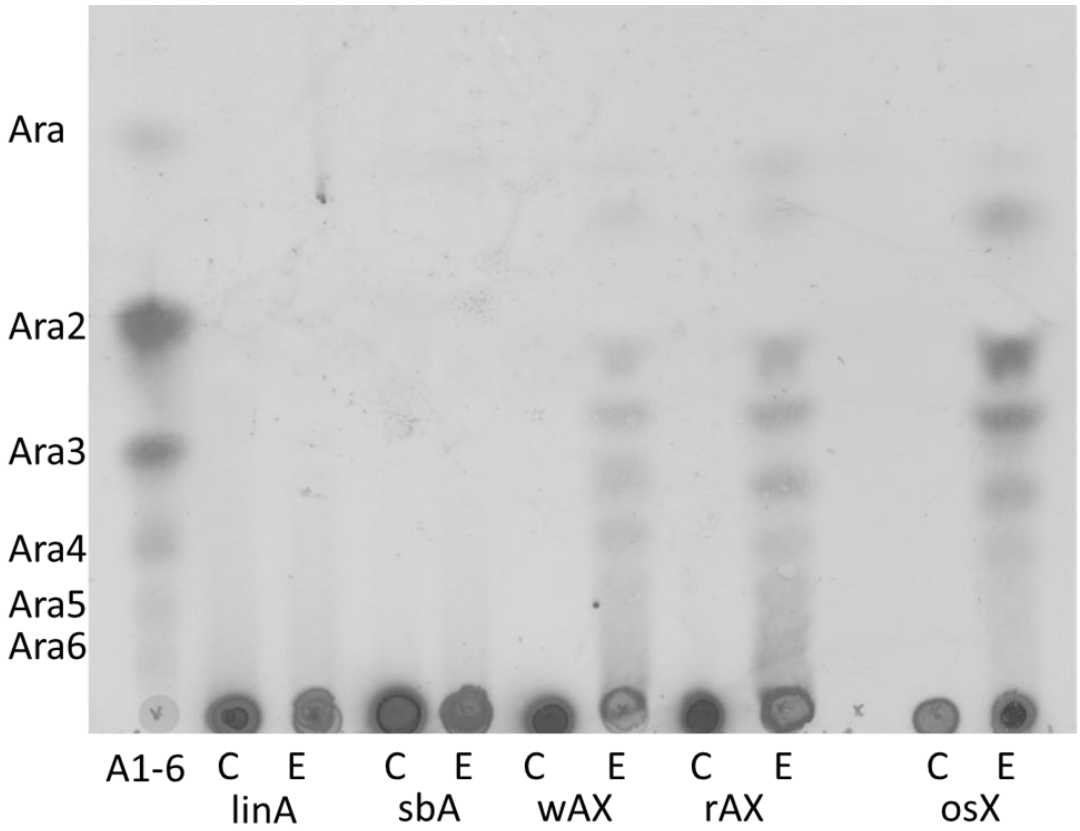
Supplemental data – Chapter Three

B.1 Polysaccharide activity screens

All GH43s obtained in a soluble form were submitted to a screen for polysaccharide activity, analysed by TLC. Several of these enzymes displayed a weak *endo*-xylanase activity.

B.1.1 Weakly acting xylanases – TLC

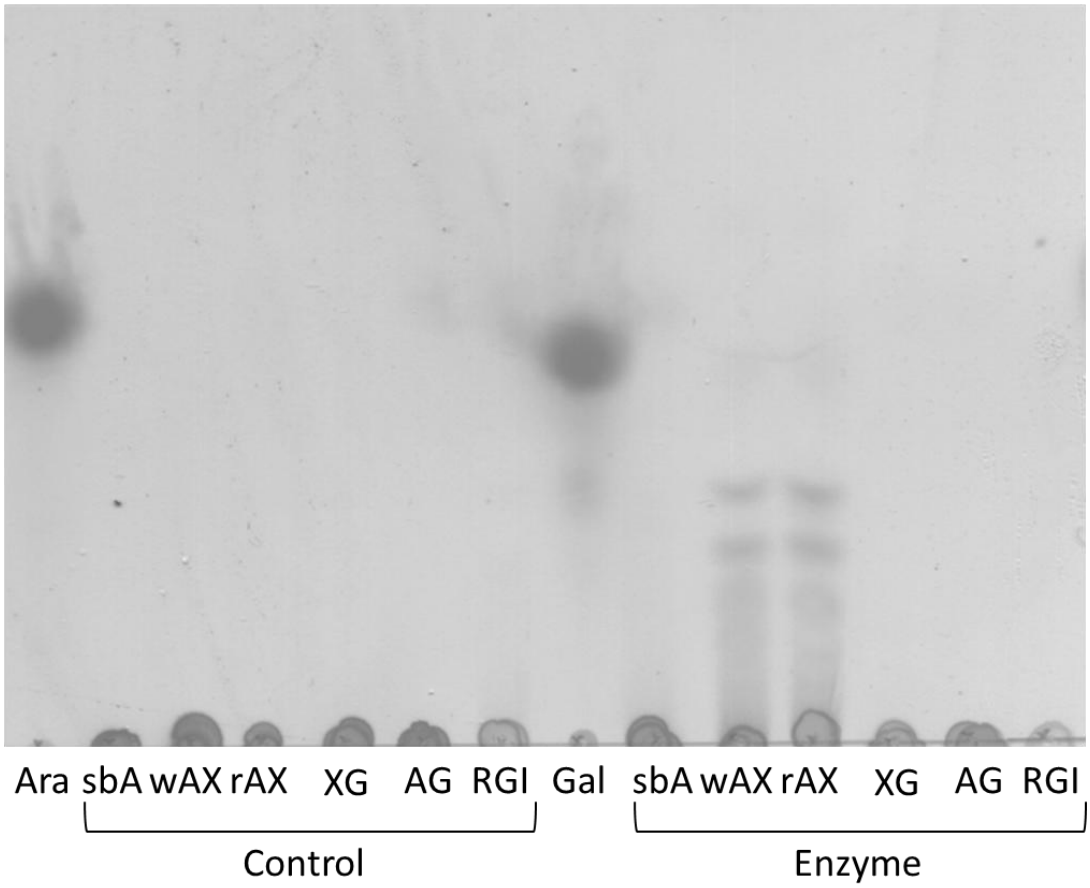
B.1.1.i Bt0265 polysaccharide screen



C = control reaction. E = enzyme reaction. A1-6 = arabinooligosaccharide standards.

linA = linear arabinan. sbA = sugar beet arabinan. wAX = wheat arabinoxylan. rAX = rye arabinoxylan. osX = oat spelt xylan.

B.1.1.ii Bt4095 polysaccharide screen

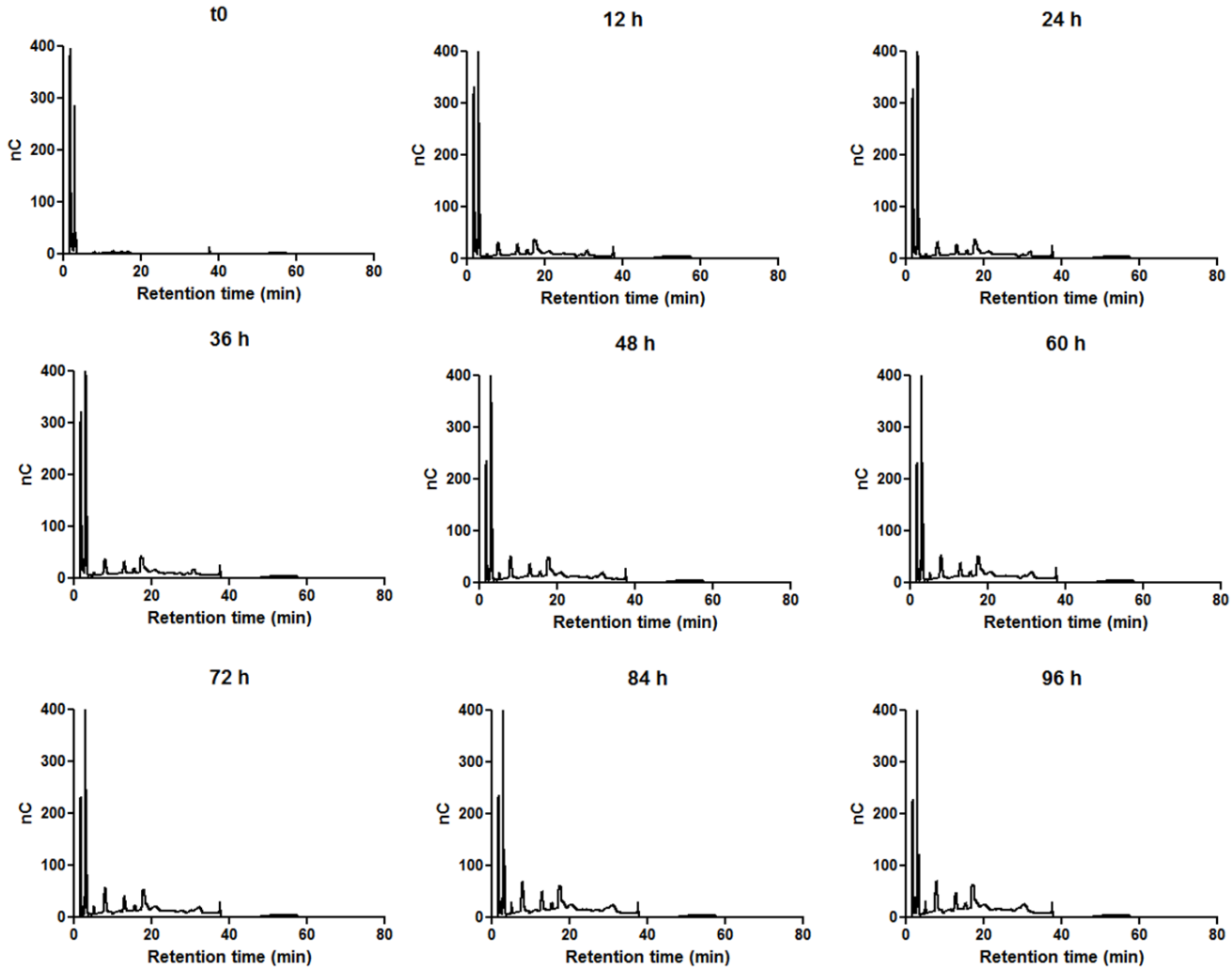


sbA = sugar beet arabinan. wAX = wheat arabinoxylan. rAX = rye arabinoxylan. XG = xyloglucan. AG = arabinogalactan. RGI = rhamnogalacturonan I.

Ara = arabinose standard. Gal = galactose standard.

B.1.2 Weakly acting xylanases – HPLC

B.1.2.i Bt3467 degradation of birchwood xylan



B.2 PUL 7 cloning experiments

B.2.1 Bt0348 and Bt0368 are GH51 enzymes encoded by the PUL 7 operon

Gene	Plasmid	Forward primer	Reverse primer	Restriction enzymes	Outcome
0348	pET21(a)	ctcgtc ggatcc caaaagagtgctaccatcactgtac	ctcgtc caagcttgc ctgtaactctaaagtaacaattgacttagccgg	BamHI HindIII	Lethal to *E. coli* (Figure 3.17)
0348	pGEX	ctcgtc gtggatccc caaaagagtgctaccatcactgtacatgc	ctcgtc actcgagcg ttactgtaactctaaagtaacaattgac	BamHI XhoI	No expression
0348	pRGST	ctcgtc ggatcc caaaagagtgctaccatcactgtacatgc	ctcgtc cgaagcttg ttactgtaactctaaagtaacaattgac	BamHI HindIII	No expression
0348	pET14(b)	ctccag catatg caaaagagtgctaccatcac	ctccag ctcgag ttactgtaactctaaagtaacaattgac	Nde1 Xho1	Some insoluble protein
0348	pET32(a)	ctccag ggatcc caaaagagtgctaccatcac	ctccag actcgagca ttactgtaactctaaagtaacaattgac	BamHI XhoI	No expression
0368	pET21(a)	ctcg ggatcc caaaccaatgaaatggtgatccagac	ctcg caagcttgc ttttttcgtgaatttataaactgcaaacgtgttcgg	BamHI HindIII	No expression
0368	pGEX	ctcgtc gtggatccc caaaccaatgaaatggtgatccagactaagaag	ctcgtc actcgagcg ttatttttttcgtgaatttataaactg	BamHI XhoI	Only GST expressed
0368	pRGST	ctcgtc ggatcc caaaccaatgaaatggtgatccagactaagaag	ctcgtc cgaagcttg ttatttttttcgtgaatttataaactg	BamHI HindIII	No expression

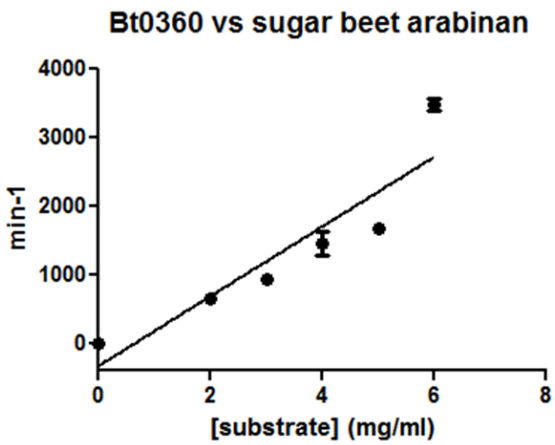
Gene	Plasmid	Forward primer	Reverse primer	Restriction enzymes	Outcome
0368	pET14(b)	ctccag ctcgag caaaccaatgaaatggtgatcc	ctccag gctgagcaa ttatttttcgtgaatttataaactgc	XhoI Bpu11021	No expression
0368	pET32(a)	ctccag ggatcc caaaccaatgaaatggtgatcc	ctccag actcgagca ttatttttcgtgaatttataaactgc	BamHI XhoI	No expression

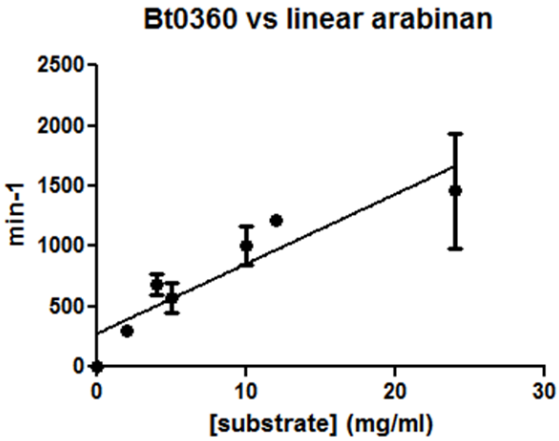
Table B.1 Plasmids, primers and restriction enzymes used in GH51 cloning experiments.

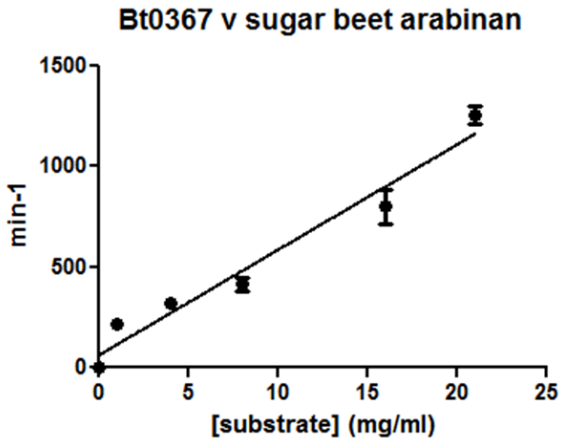
Neither Bt0348 nor B0368 was obtained in a soluble form.

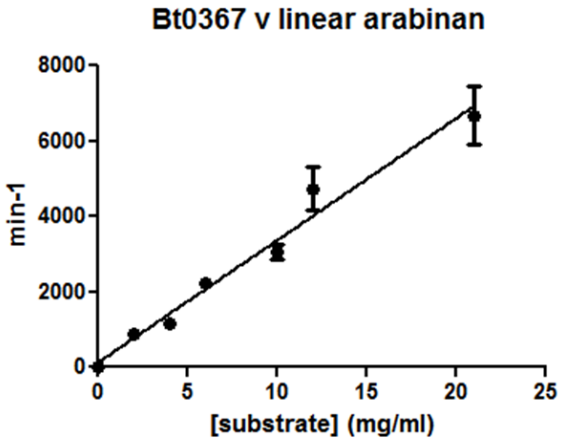
B.3 Example kinetics

B.3.1 *Endo*-arabinanases









APPENDIX C

Purification of an arabinoxylooligosaccharide

C.1 Purification of a ligand for crystallisation with HiAXHd3

An oligosaccharide ligand was generated for crystallisation with *Hi*AXHd3 by partial digestion of wheat arabinoxylan with a GH10 xylanase (Chapter Four, Section 4.3). Reaction products were purified by size exclusion chromatography. 5 ml fractions were collected and assayed for *Hi*AXHd3 activity to determine the location of a suitable substrate for the enzyme.

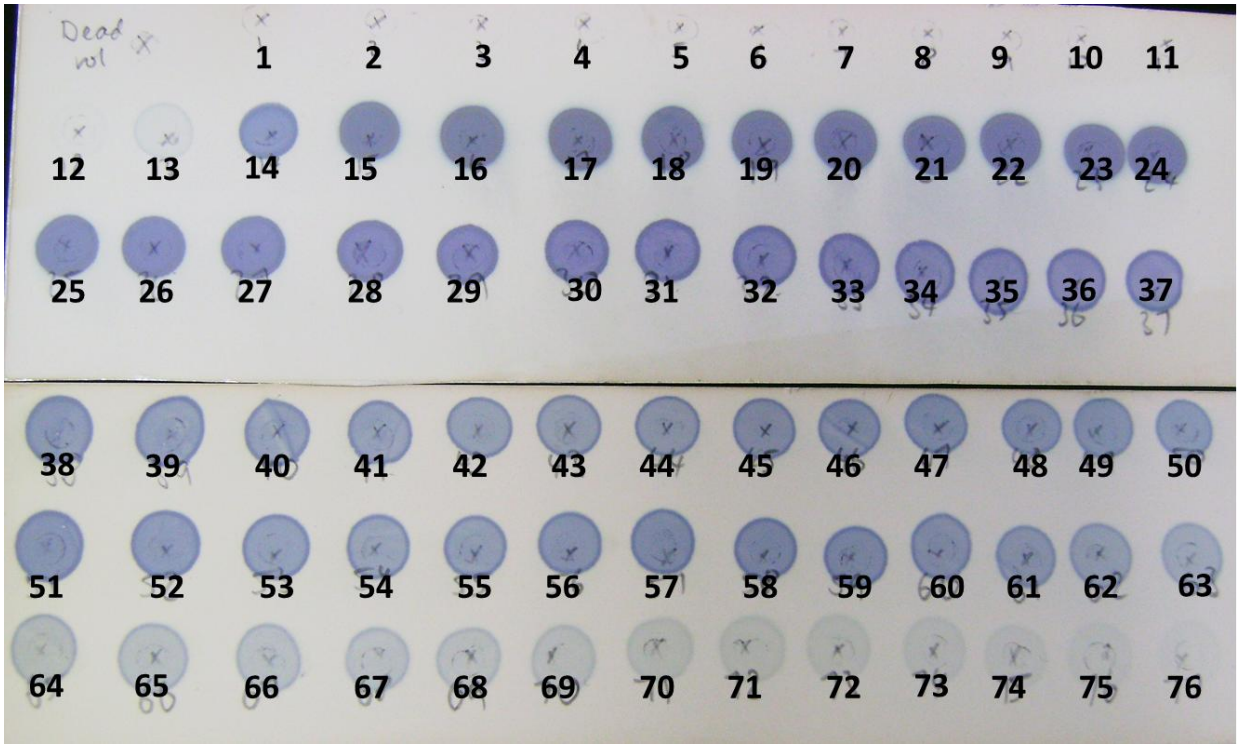
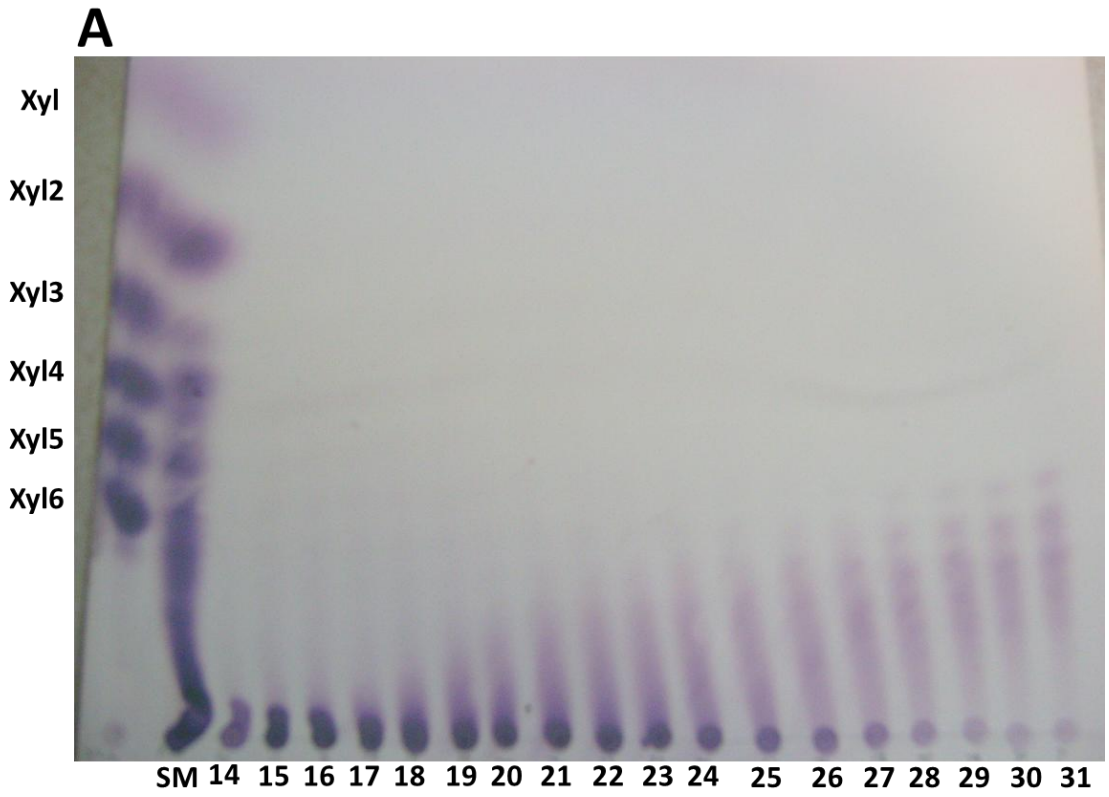
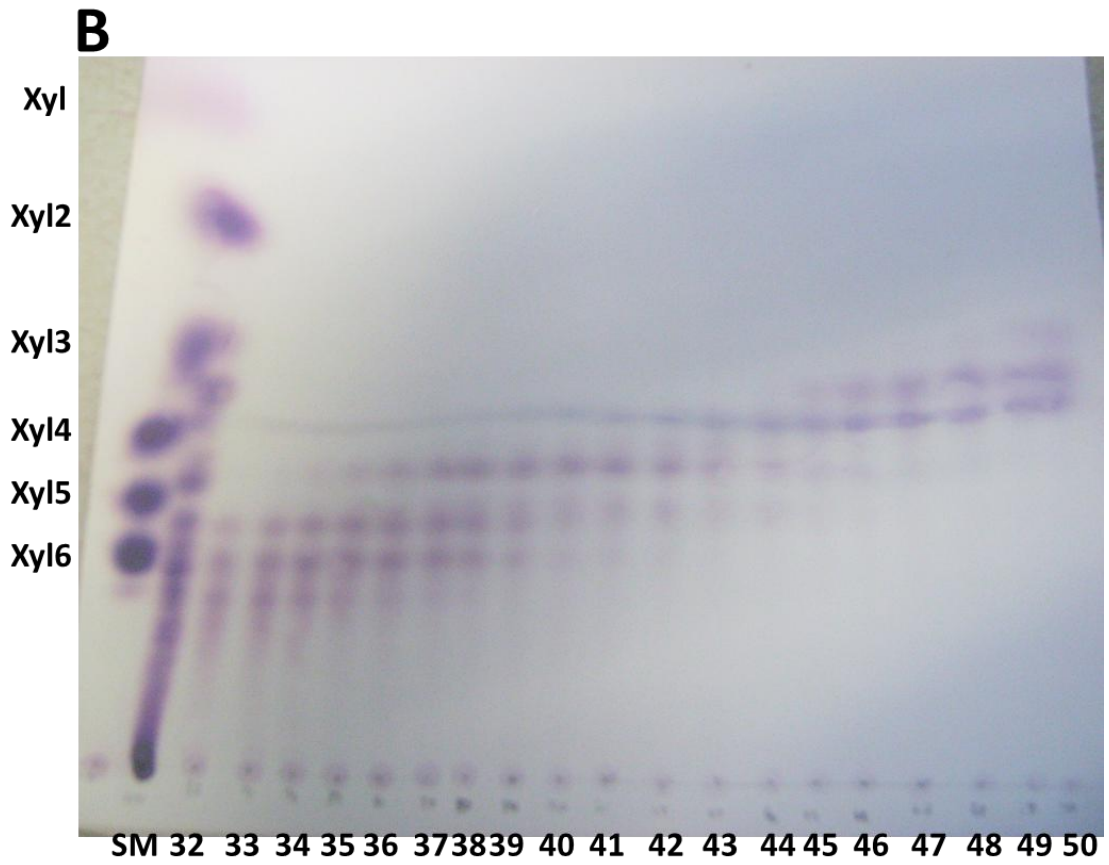


Figure C.1 TLC analysis of fractions collected from a P2 biogel column

Fractions were collected from the column after 150 ml of water had flowed through. 10 µl of each fraction was spotted onto a TLC plate, which was dried and stained to check for the presence of sugar. Significant sugar is present in fractions 14 – 63. Some sugar is still visible up to fraction 74.


A
Xyl
Xyl2
Xyl3
Xyl4
Xyl5
Xyl6
SM 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31


B
Xyl
Xyl2
Xyl3
Xyl4
Xyl5
Xyl6
SM 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50

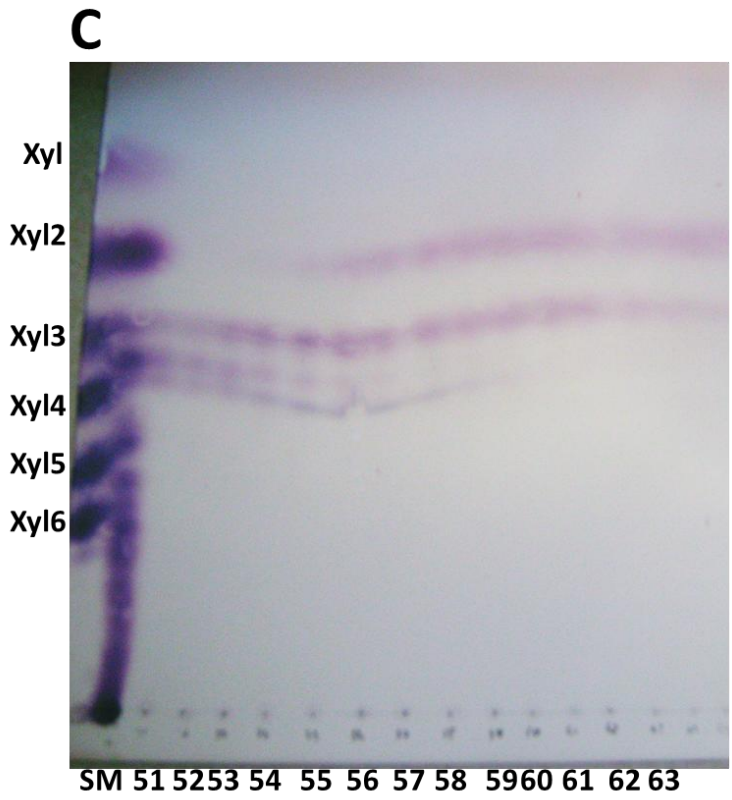


Figure C.2 TLC analysis of purified arabinoxylooligosaccharides

Fractions which had been identified as containing significant amounts of sugar were analysed by TLC as a quick assessment of the success of purification. There is a clear pattern through the fractions, with larger oligosaccharides eluting in earlier fractions, indicating that products have been successfully separated by size.

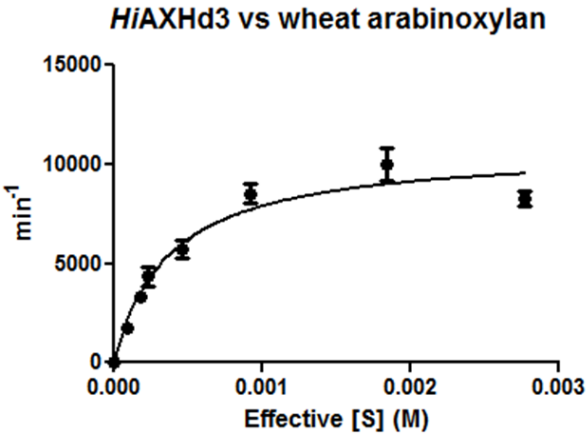
SM = starting material. This is the original mixture of arabinoxylooligosaccharides which was loaded onto the column.

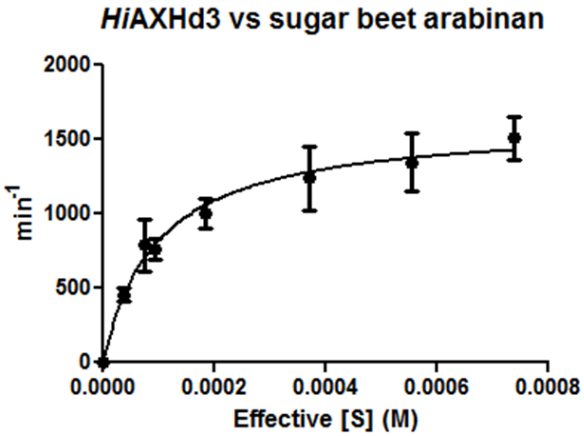
APPENDIX D

Supplemental data – Chapter Four

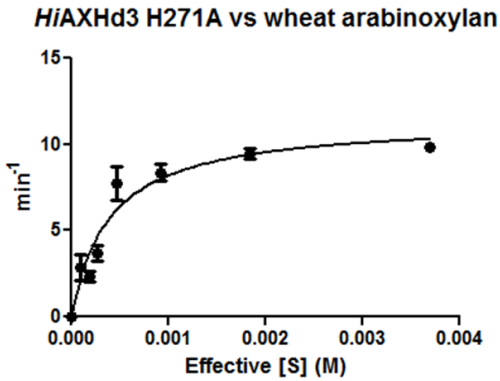
D.1 Example kinetics

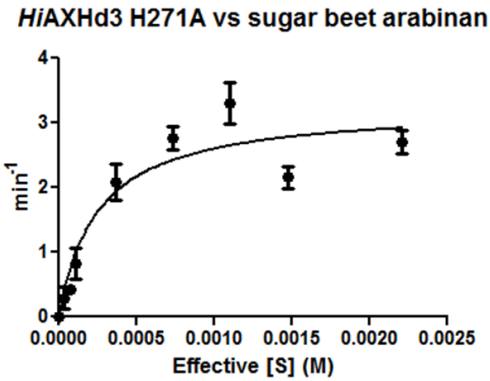
D.1.1 Wildtype *Hi*AXHd3

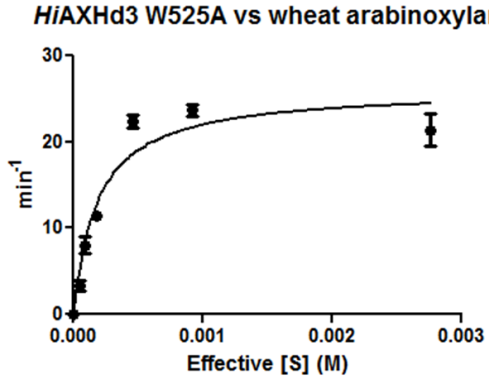


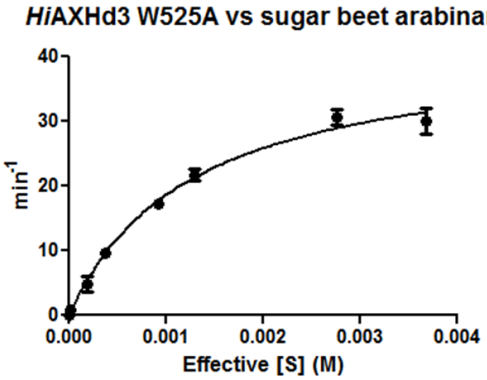


D.1.2 Mutant *Hi*AXHd3

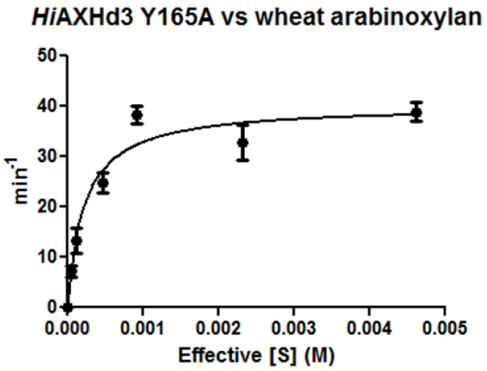


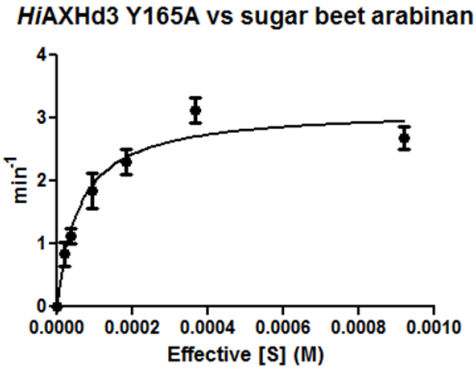


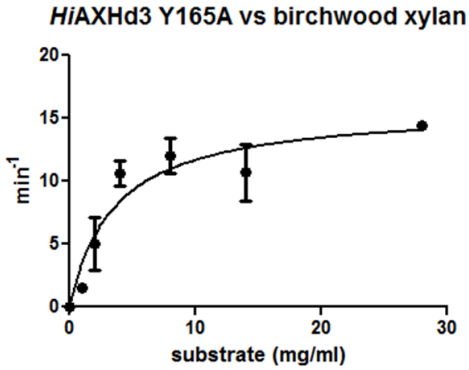




D.1.3 *Hi*AXHd3 Y165A variant

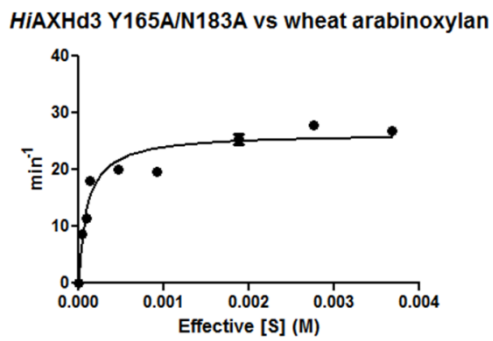


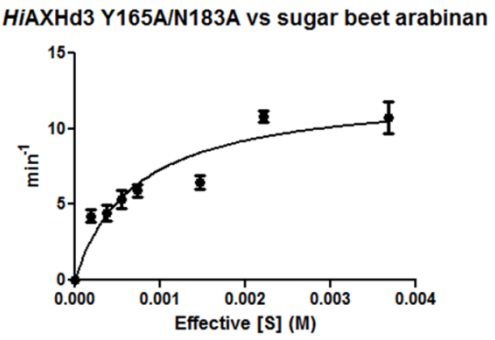


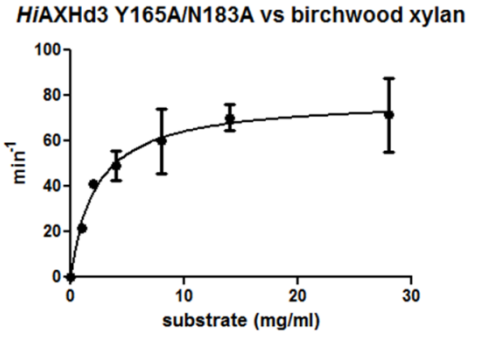


D.1.4 Double mutants of *Hi*AXHd3

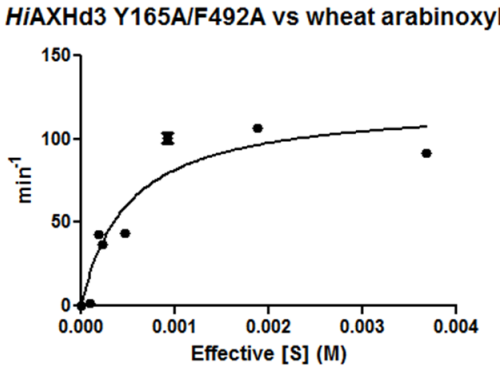
D.1.4.i Y165A/N183A

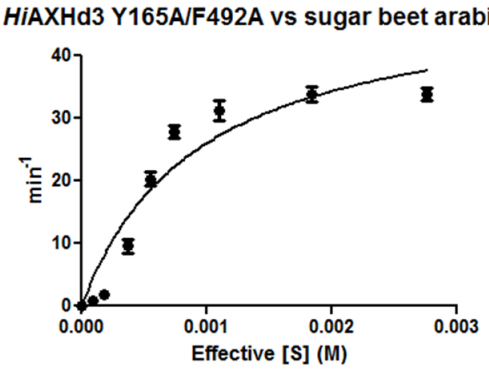


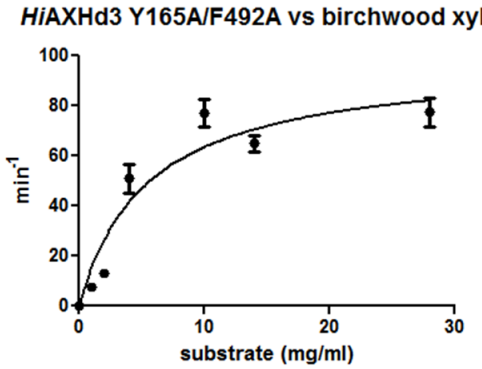




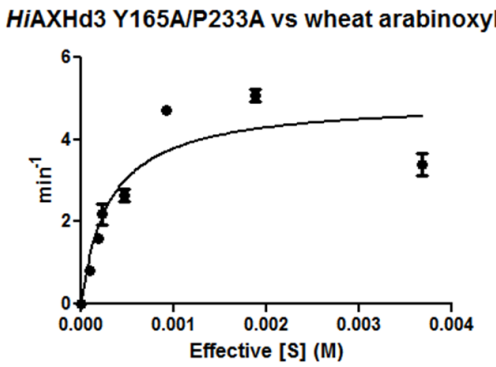
D.1.4.ii Y165A/F492A

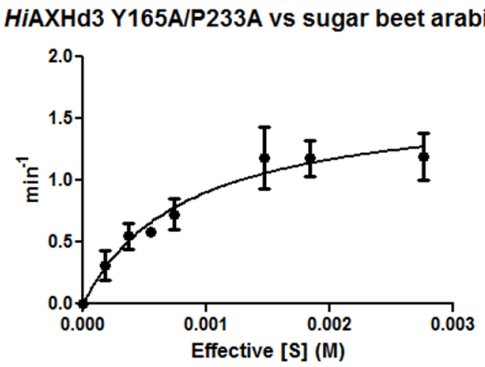


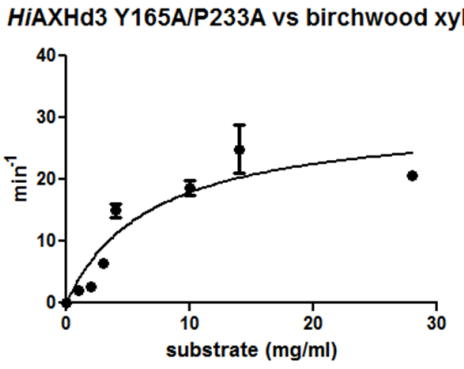




D.1.4.iii Y165A/P233A







D.2 Mutant Y165A activity screen

The Y165A mutant of *Hi*AXHd3 (Chapter Four, Section 4.5) was screened for activity on several substrates by TLC. The positive results presented here were confirmed by subsequent HPLC. These data show that the Y165A mutant releases arabinose from wheat arabinoxylan and sugar beet arabinan, as does the wildtype enzyme. In addition, the mutant releases a range of products from arabinoxylan and xylooligosaccharides with a degree of polymerisation greater than four.

D.2.1 Polysaccharide screening

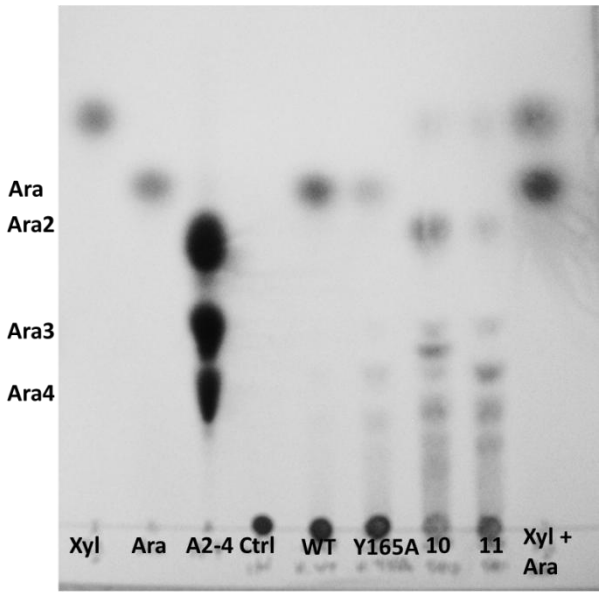


Figure D.1 TLC screen of activity against wheat arabinoxylan.

Wheat arabinoxylan (2 mg ml^{-1}) was incubated for four hours with wildtype and mutant *Hi*AXHd3, and classical xylanases from families GH10 and GH11 (*Hi*AXHd3 wildtype (WT), GH10 and GH11 at 50 nM, *Hi*AXHd3 mutant Y165A at 10 μM). A control reaction (Ctrl) was also performed, where the substrate was incubated in buffer with no enzyme. No product is released during the control reaction. Both wildtype and mutant *Hi*AXHd3 release arabinose from the substrate, although the mutant releases less of the sugar. The mutant also releases small amounts of other products, indicating some degree of xylanase activity. Degradation by the GH10 and GH11 xylanases is more significant.

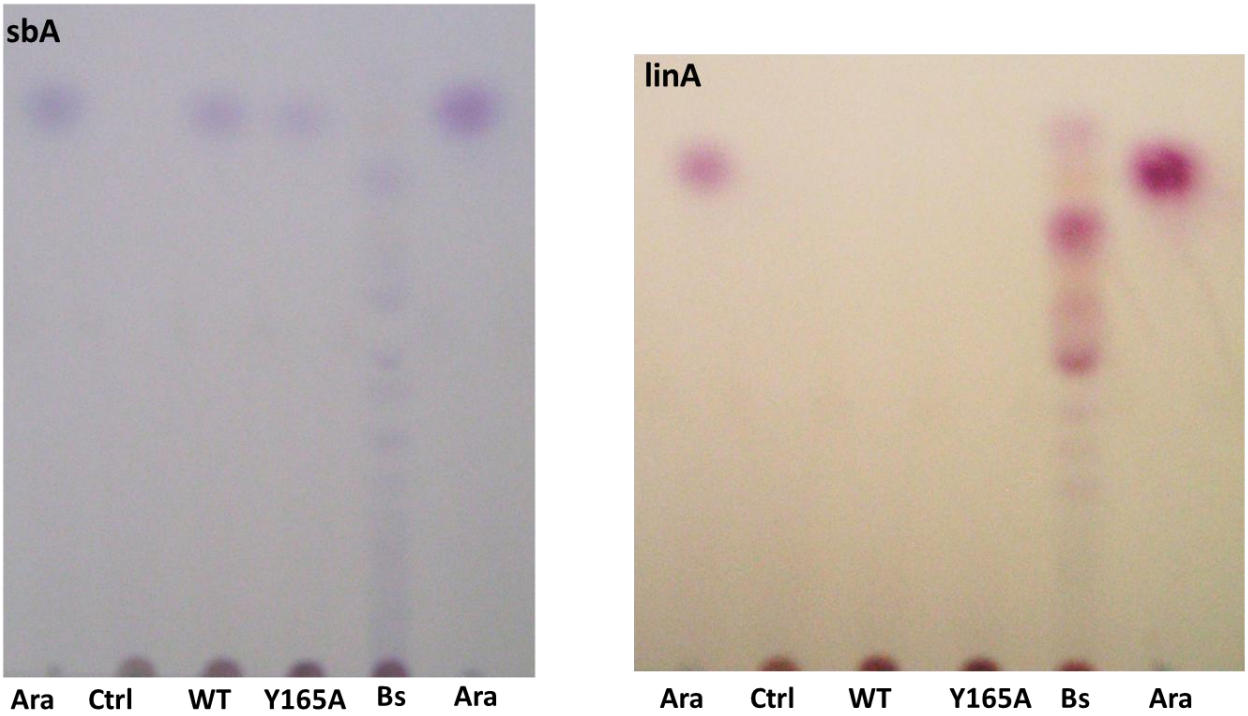


Figure D.2 TLC screen of activity against sugar beet and linear arabinan.

Substrate (2 mg ml^{-1}) was incubated for four hours with wildtype and mutant *Hi*AXHd3, and the arabinanase *Bs*Arb43A (*Hi*AXHd3 wildtype (WT) and *Bs*Arb43A at 50 nM, *Hi*AXHd3 mutant Y165A at 10 μM). A control reaction (Ctrl) was also performed, where the substrate was incubated in buffer with no enzyme. No product is released during the control reactions. Both wildtype and mutant *Hi*AXHd3 release arabinose from the sugar beet arabinan (sbA), although the mutant releases less of the sugar. Neither enzyme hydrolyses linear arabinan (linA). The previously characterised arabinanase releases oligosaccharides from both substrates. This shows that the Y165A mutant does not possess arabinanase activity.

D.2.2 Oligosaccharide screening

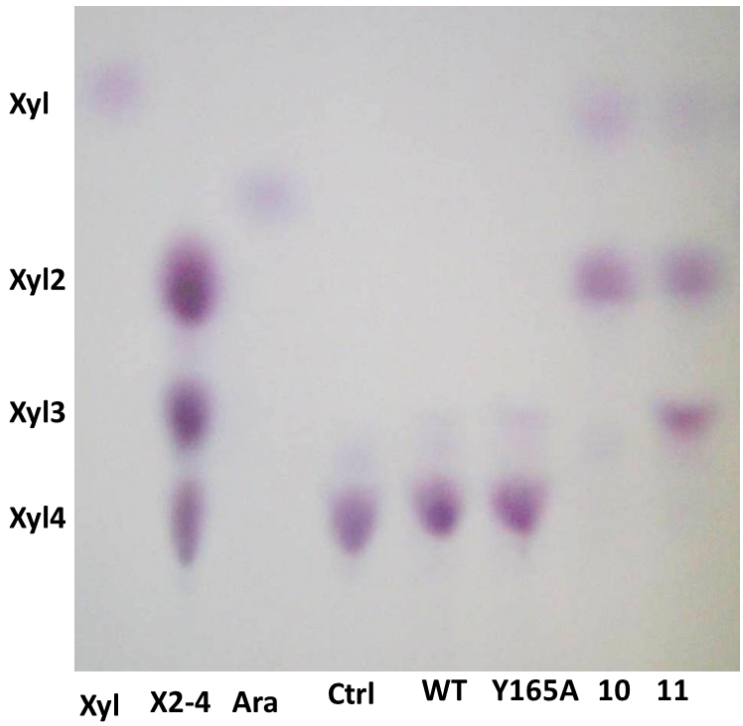


Figure D.3 TLC screen of activity against xylotetraose.

Xylotetraose (10 mM) was incubated for four hours with wildtype and mutant *Hi*AXHd3, and classical xylanases from families GH10 and GH11 (*Hi*AXHd3 wildtype (WT), GH10 and GH11 at 50 nM, *Hi*AXHd3 mutant Y165A at 10 μM). A control reaction (Ctrl) was also performed, where the substrate was incubated in buffer with no enzyme. No product is released during the control reaction. Neither wildtype nor mutant *Hi*AXHd3 hydrolyses xylotetraose, indicating that Y165A is active on longer substrates. The GH10 and GH11 xylanases release a range of products.

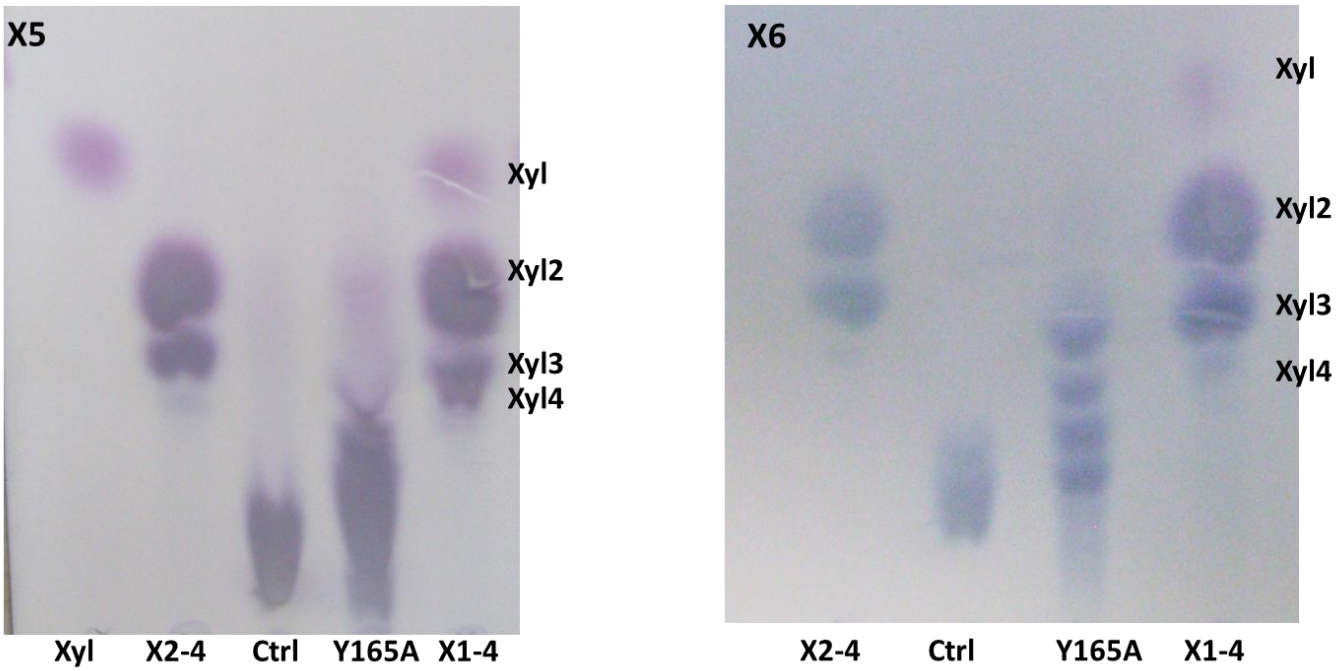


Figure D.4 TLC screen of activity against xylotetraose and xylohexaose.

Xylopentaose (X5) and xylohexaose (X6) (10 mM) were incubated for four hours with Y165A mutant *Hi*AXHd3(10 µM). A control reaction (Ctrl) was also performed, where the substrate was incubated in buffer with no enzyme. The substrates are impure, but there is a clear difference in both assays between the control samples and the enzyme reactions. This indicates that Y165A has xylanase activity on xylooligosaccharides with a degree of polymerisation of five or greater.

APPENDIX E

NMR Spectra

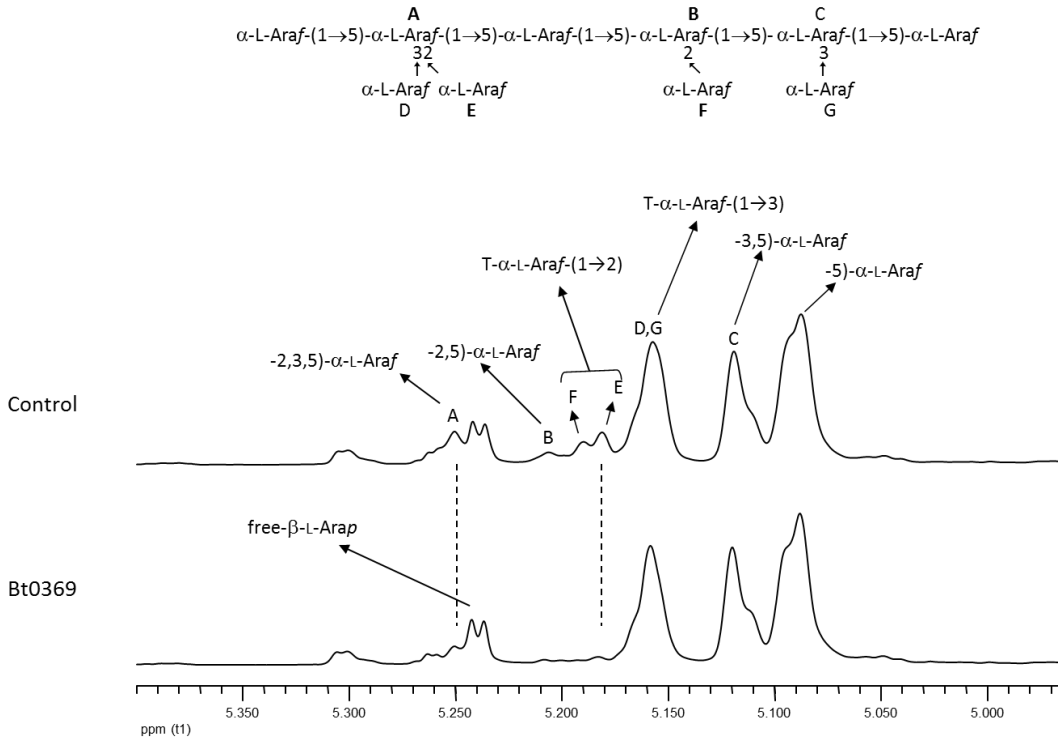


Figure E.1 1D-NMR analysis of arabinooligosaccharides incubated with Bt0369.

Linkages are identified in the schematic of arabinan by single letter codes; these correspond to the appropriately labelled peak in the spectrum. The spectrum is highly complex. Peaks F and E correspond to O2-linked arabinose in single and double substitutions, respectively. Peaks B and A correspond to backbone residues which are decorated with an O2-linked arabinose in a single and double substitution, respectively. Peaks F, E, B and A are present in the control reaction but are lost during incubation with the enzyme (peak A is somewhat obscured by overlap with the free arabinose peak produced during incubation). It is logical that loss of the O2-linked arabinose in the double substitution would lead to an increase in singly substituted O3-linked arabinose (peak G and corresponding backbone peak C) and loss of the signal for doubly substituted O3 arabinose (peak D). However, peaks D and G overlap and the high abundance of the single O3 substitution in the starting mixture makes this difficult to observe in a 1D spectrum.

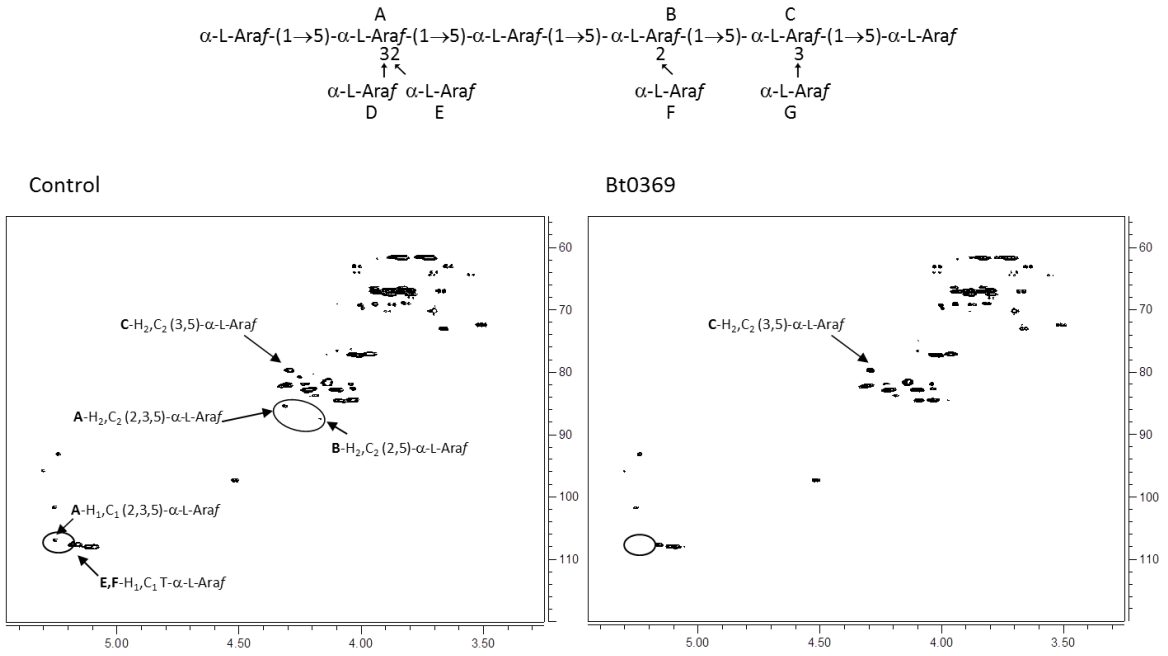


Figure E.2 2D gHSQc spectrum showing arabinooligosaccharides incubated with Bt0369.

Cross-peaks corresponding to O2-linked arabinose in single and double substitutions are circled in the Control spectrum. These are lost during incubation with the enzyme. Peaks corresponding to O3-linked arabinose remain after enzyme treatment.

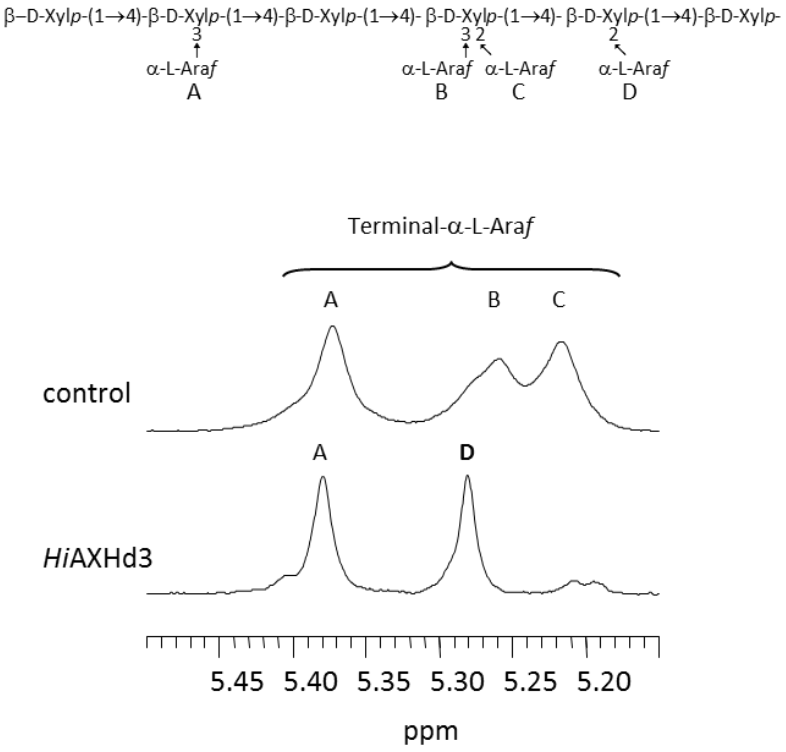


Figure E.3 1D NMR spectra of wheat arabinoxylan before and after incubation with *Hi*AXHd3.

1D NMR of *Hi*AXHd3 wheat arabinoxylan reaction products. In the schematic, linkages within the polysaccharide (A-D) are labelled to correspond with peaks in the NMR spectra. Peaks corresponding to the anomeric protons of α-L-arabinofuranose residues linked at the O2 (C) and O3 (B) position to doubly substituted xylose residues are present in the spectrum of arabinoxylan prior to enzyme treatment, but are absent after incubation. A peak corresponding to the anomeric proton of α-L-arabinofuranose linked at the O2 (D) position to single substituted xylose residues, is visible only in the spectrum of arabinoxylan after treatment with *Hi*AXHd3, having been generated by enzymatic cleavage of arabinose residues at position B. A small peak for single O2 (D) substitutions is likely present in the control spectrum but is obscured by other peaks.

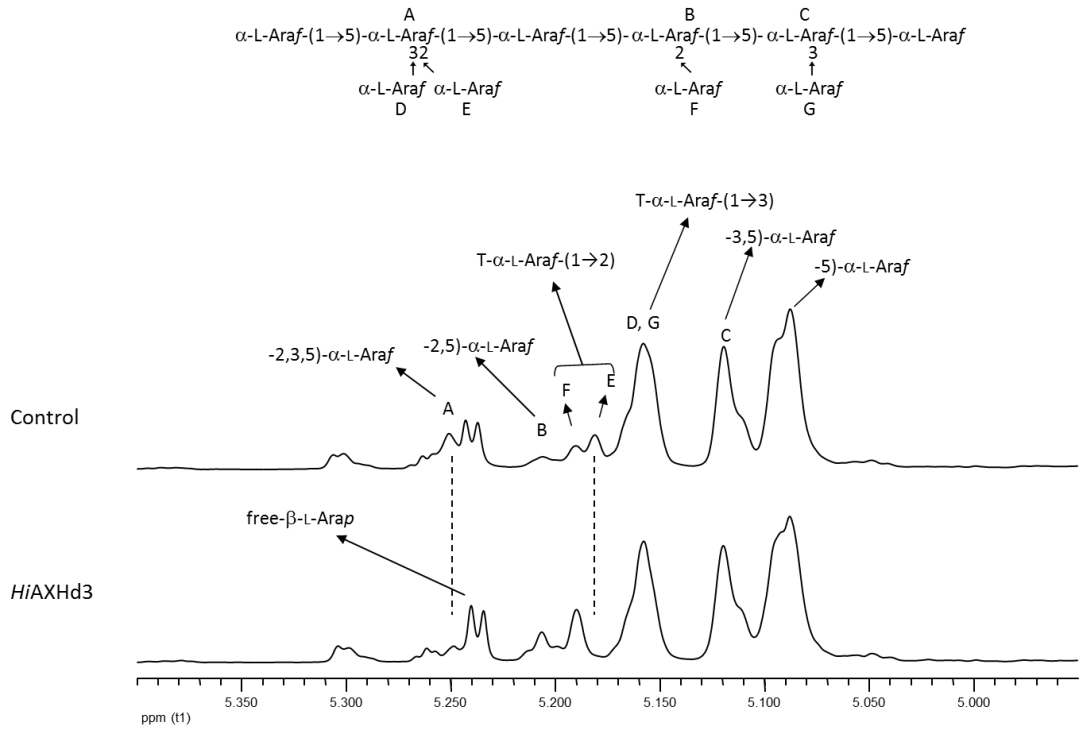


Figure E.4 1D NMR spectra of sugar beet arabinan before and after incubation with *Hi*AXHd3.

1D NMR of *Hi*AXHd3 sugar beet arabinan reaction products. In the schematic, linkages within the polysaccharide (A-G) are labelled to correspond with peaks in the NMR spectra. As described in the text, arabinooligosaccharides generated by partial acid hydrolysis of arabinan were used as substrate in this experiment. Peaks corresponding to the anomeric protons of arabinose units that were substituted at both O2 and O3, and the respective α-L- arabinofuranose residues attached to O2 of 2,3,5 linked arabinose (backbone arabinose units that are doubly substituted), were lost after enzyme treatment. Intensity of the signals corresponding to the anomeric protons of α-L- arabinofuranose residues that contain a single O2 substitution and arabinose units linked O2 to the arabinan backbone increased.

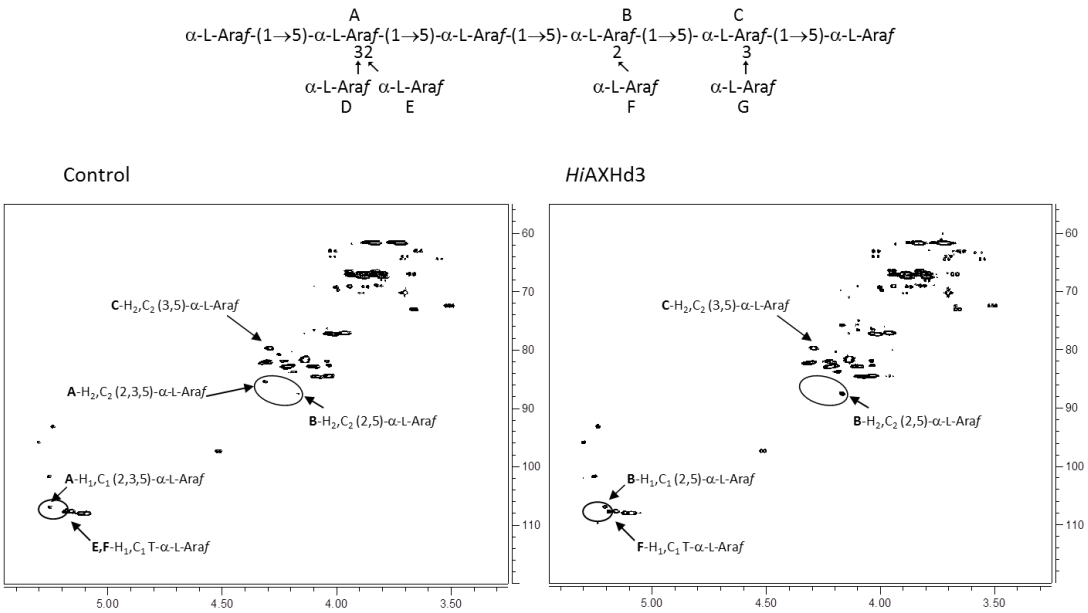


Figure E.5 gHSQC spectra of arabinooligosaccharides before and after incubation with *Hi*AXHd3.

2D NMR (gHSQC) of *Hi*AXHd3 arabinooligosaccharide reaction products. Peaks corresponding to the anomeric protons of arabinose units that were substituted at both O2 and O3, and the respective α-L-arabinofuranose residues attached to O2 of 2,3,5 linked arabinose (backbone arabinose units that are double substituted), were lost after enzyme treatment. Intensity of the signals corresponding to the anomeric protons of α-L- arabinofuranose residues that contain a single O2 substitution and arabinose units linked O2 to the arabinan backbone increased, as these were generated by cleavage of the O3-linked arabinose in the double substitution.

		H_1	H_2	H_3	H_4	H_5		C_1	C_2	C_3	C_4	C_5
T–α-L-Ara*f* -(1→3)		5.157	4.137	3.959	4.034	3.835, 3.721		107.7	81.6	77.1	84.6	61.7
T–α-L-Ara*f* -(1→2)	F	5.186	4.130	3.966	4.065	3.828,3.719		107.7	n.a.	n.a.	n.a.	n.a.
T–α-L-Ara*f* -(1→2)	E	5.180	4.132	3.963	4.08	3.833, 3.718		107.5	81.6	77.1	84.6	61.7
-5)-α-L-Ara*f* -(1→5)		5.085	4.129	4.03	4.217	3.892, 3.798		108.8	81.6	77.1	83.0	67.2
-5)-α-L-Ara*f* -(1→5)		5.095	4.131	4.01	4.26	n.a.		108.8	81.6	77.3	n.d.	n.d.
-3,5)-α-L-Ara*f* -(1→5)	C	5.118	4.292	4.094	4.307	3.947, 3.840		108.0	79.8	83.0	82.3	67.1
-3,5)-α-L-Ara*f* -(1→5)	C	5.112	4.291	4.037	4.218	3.891, 3.806		108.0	79.7	82.7	83.8	67.0
-2,5)-α-L-Ara*f* -(1→5)	B	5.209	4.170	4.16	4.097	3.955, 3.885		107.0	87.6	75.9	n.a.	67
-2,3,5)-α-L-Ara*f* -(1→5)	A	5.254	4.314	4.255	4.3	3.9, 3.8		107.0	85.6	80.8	81.66	67
free -L-Ara												
α-L-Ara*p*		4.507	3.504	3.650	3.934	3.890, 3.671		97.4	72.5	73.2	69.2	67.1
β-L-Ara*p*		5.230	3.810	3.872	3.999	4.018, 3.641		93.1	69.0	69.2	69.4	63.1
α-L-Ara*f*		5.241	4.031	3.98	4.11	3.790, 3.687		101.7	82.0	76.1	83.6	61.7
β-L-Ara*f*		5.293	4.085	4.04	3.834	3.8, 3.659		95.9	76.9	74.8	81.9	61.8

Table E.1 Chemical shifts of sugar beet arabinan components.