Structure and Function of Novel Cellulosic, Hemicellulosic and Pectic Glycoside Hydrolases

Alan Cartmell

A Thesis Submitted for the Degree of Doctor of Philosophy

2006 – 2010

Institute for Cell and Molecular Biosciences

Newcastle University

Abstract

Cellulose is a major component of the plant cell wall and is the most abundant organic molecule in the biosphere. Efficient degradation of this polysaccharide is required if the plant cell wall is to be used as a viable source of renewable biofuels. Bacteria express an arsenal of different cellulases that catalyse the degradation of cellulose. The reason why many different cellulases are expressed rather than one highly active cellulase is unclear, but probably lies in the structural diversity displayed by cellulose, which is much greater than its invariant chemical composition suggests. Part of this work describes a novel cellulase from the plant cell wall degrading bacterium *Clostridium thermocellum*. This cellulase, *Ct*Cel119, is the first of this class of enzyme to display the α_8 helical fold, is the founding member of a new GH family, performs catalysis through a possible "Grothuss style mechanism" and shares features typical of lytic transglycoslases. Structural data also seem to suggest that the enzyme may attack a novel structure in crystalline cellulose, which could contribute to understanding why bacteria such as *C. thermocellum* employ a diverse variety of cellulases.

A study on glycoside hydrolase family (GH) 26, which consists mainly of endo-β-1,4 mannanases, was also conducted. The work presented in this thesis focused on two novel members of the family. One component of this section provided a thorough kinetic analysis of mutants of active site residues of a GH26 endo-β-1,4-1,3-glucanase. This identified crucial interactions at the -2 subsite, which contribute to the stabilisation of the $^4H_3$ transition state. The other component of this section was the identification and characterisation of an exo-acting mannanase, *Cj*Man26C, also termed a mannobiohydrolase. *Cj*Man26C is the only mannanase characterised to date to release mannobiose. The mannobiohydrolase displayed an extremely high catalytic efficiency of 3×10^9 min^{-1} M^{-1} against mannotetraose. The crystal structure revealed a -1 sugar in a $^1S_5$ pre transition state, providing further support for a $B_{2,5}$ transition state in GH26. The exo activity was conferred by a four amino acid insertion in loop 3 at the -2 subsite. Mutation of D130, to Gly or Ala, in loop 3 was enough to partially convert the enzyme to an endo-mode of action, while removal of D130 plus two flanking resides caused a full conversion to an endo-mode of action.

The gene expansion observed in family GH43 enzymes was also investigated. Eleven genes encoding GH43 enzymes were cloned, expressed and investigated for catalytic activity. Three arabinofuranosidases were characterised, two exo α-1,5-L-arabinofuranosidases and a novel sugar beet arabinan specific α-1,2-L-arabinofuranosidase that could attack both single and double substitutions named *Cj*Abf43B. The crystal structure of *Cj*Abf43B was solved in complex with ligand. The structure revealed a curved binding cleft, around a deep active site pocket, that was specific for the curved nature of sugar beet arabinan backbone. The curved binding cleft also had a groove into which α-1,3-L-arabinofuranosides could be accommodated, explaining how the enzyme has plasticity for single and double substitutions. A β-1,4 xylosidase was also characterised, while four enzymes were identified that displayed 'trace' activity against xylans. Three of these appeared to display endo-activity, while the fourth enzyme displayed very weak arabinfuranosidase activity against xylans.

Acknowledgments

I would like to thank, first and foremost, Professor Harry Gilbert for giving me the opportunity to do this Ph.D. I could not have asked for a better Ph.D supervisor and I am grateful for your patient tutoring, diligent guidance and generosity, despite my constant questioning. I am also grateful to all members of Lab M.2035 (Newcastle University) and 2075 (CCRC), who are in no particular order: Miss Joanna Norman, Dr David Bolam, Dr Cedric Montanier, Dr Tibor Nagy, Dr Claire Dumon, Dr Mark Proctor, Dr Evangelos Topakas, Dr Louise Tailford, Dr Nicola Brown, Dr Hongjun Zheng, Miss Ziyuan Liu, Dr Edith Diaz-Mireles, Dr Artur Rogowski, Dr Wade Abbot, Miss Lauren McKee, Miss Fiona Cuskin, Mr Carl Morland, Dr James Flint, Dr Elisabeth Lowe and Mr Yanping Zhu.

I would also like to thank Professor Richard Lewis, Dr Jon Marles-Wright and Dr Susan Firbank of the structural biology laboratory at Newcastle University for their generous help and support in collaborations. I am also grateful for collaborations with the Structural Biology group of Professor Gideon Davies at York University and Dr Valerie Ducros. I would also like to thank Dr William York and Dr Maria Péna, from the CCRC, for their expertise in NMR analysis. Finally I would like to also thank Professor Carlos Fontes and Miss Joana Brás for their valuable contributions to this work.

Contents

List of Figures

List of Tables

Abbreviations

[E]	Enzyme concentration
[S]	Substrate concentration
4NP	4-nitrophenolate
A_{280}	Absorbance at a wavelength of 280 nm
A_{600}	Absorbance at a wavelength of 600 nm
Amp	Ampicillin
Ara	L-Arabinose
BLAST	Basic Local Alignment Search Tool
C	Chair (sugar ring conformation)
CAZy	Carbohydrate Active Enzyme server
CBM	Carbohydrate-Binding Module
CE	Carbohydrate Esterase
CFE	Cell Free Extract
Cj	*Cellvibrio japonicus*
*Cj*Abf51	Arabinofuranosidase 51 from *Cellvibrio japonicus*
Ct	*Clostridium thermocellum*
DNSA	3,5-Dinitrosalicylic Acid
dNTP	Deoxynucleotidetriphosphate
ESRF	European Synchrotron Radiation Facility
FPLC	Fast Protein Liquid Chromatography
Fuc	L-Fucose
G2M5	6^3,6^4-di-α-D-galactosyl mannopentaose
G	Glucose
G2	Cellobiose
G3	Cellotriose
G4	Cellotetraose
G5	Cellopentaose
G6	Cellohexaose
GH	Glycoside Hydrolase
H	Half-chair (sugar ring conformation)
His_6-tag	Polyhistidine tag
HPLC	High Performance Liquid Chromatography
IEC	Ion Exchange Chromatography
IMAC	Immobilised Metal Affinity Chromatography
IPTG	Isopropyl β-D-thiogalactopyranoside
ITC	Isothermal calorimetry
LacZ	β-Galactosidase
LB	Luria-Bertani Broth
M6	Mannohexaose
M	Mannose
M2	Mannobiose
M3	Mannotriose
M4	Mannotetraose
M5	Mannopentaose
M6	Mannohexaose
MQ H_2O	MilliQ ultra pure double distilled 18.2 MΩ water

M_r	Molecular mass
OSX	Oat spelt xylan
p	Plasmid
P	Primary cell wall
PAD	Pulsed Amperometric Detection
PCR	Polymerase Chain Reaction
PDB	Protein Data Bank
PEG	Polyethylene Glycol
RAX	Rye arabinoxylan
RG	Rhamnogalacturonan
RGI	Rhamnogalacturonan I
RGII	Rhamnogalacturonan II
RMS	Root mean square
S	Secondary cell wall
SAD	Single Anomalous Diffraction
SeMet	Selenomethionine
TEMED	N,N,N',N'-Tetramethylethylenediamine
TLC	Thin layer chromatography
WAX	Wheat arabinoxylan
YSBL	York Structural Biology Laboratory

Journal Articles

During the course of this investigation four articles were published containing work conducted by the author during this Ph.D, some of which is included in this thesis:

Analysis of the structural and functional diversity of plant cell wall specific family 6 carbohydrate binding modules.
Abbott DW, Ficko-Blean E, van Bueren AL, Rogowski A, Cartmell A, Coutinho PM, Henrissat B, Gilbert HJ, Boraston AB.
Biochemistry. 2009 Nov 3;48(43):10395-404.

The active site of a carbohydrate esterase displays divergent catalytic and noncatalytic binding functions.
Montanier C, Money VA, Pires VM, Flint JE, Pinheiro BA, Goyal A, Prates JA, Izumi A, Stålbrand H, Morland C, Cartmell A, Kolenova K, Topakas E, Dodson EJ, Bolam DN, Davies GJ, Fontes CM, Gilbert HJ.
PLoS Biol. 2009 Mar 31;7(3):e71

The Cellvibrio japonicus mannanase CjMan26C displays a unique exo-mode of action that is conferred by subtle changes to the distal region of the active site.
Cartmell A, Topakas E, Ducros VM, Suits MD, Davies GJ, Gilbert HJ.
J Biol Chem. 2008 Dec 5;283(49):34403-13. Epub 2008 Sep 17.

Probing the beta-1,3:1,4 glucanase, CtLic26A, with a thio-oligosaccharide and enzyme variants.
Money VA, Cartmell A, Guerreiro CI, Ducros VM, Fontes CM, Gilbert HJ, Davies GJ.
Org Biomol Chem. 2008 Mar 7;6(5):851-3. Epub 2008 Feb 4.

Chapter I

General Introduction

1.1 Introduction

As the world's energy demands increase and non-renewable resources, such as fossil fuels, diminish, alternative, renewable sources of energy production will need to be utilised, such as bio-ethanol. First generation bio-ethanol is currently being produced from corn starch and sugar cane. This process, however, has numerous problems such as the need for fertile land, the quantity of ethanol that can be produced does not meet the demand for liquid fuel that exists, and the industrial use of starch is in direct competition with an important human and livestock food source. A more environmentally sustaining approach to bio-ethanol production is to use the plant cell wall, often referred to a ligno-cellulosic biomass, as the starting material to produce liquid fuels. The production of second generation bio-ethanol from ligno-cellulosic biomass represents one of the major industrial challenges of the 21st century and is an area of intensive research (Somerville 2007; Himmel et al. 2009). The plant cell wall represents the biggest source of organic carbon in the biosphere. It is estimated 10^{11} tonnes of carbon is recycled annually by archaea, bacteria and fungi, this has an energy equivalent to 640 billion barrels of oil. (Coughlan 1985; Gong et al. 1999; Himmel et al. 1999). These organisms perform this extraordinary feat by expressing a battery of enzymes, primarily glycoside hydrolases (GHs), which degrade the plant cell wall. It is estimated that from one ton of ligno-cellulosic biomass ~110 gallons of ethanol could be produced, depending on the plant species, while one ton of sugar cane yields only ~19.5 gallons of ethanol (Somerville 2007). This makes bio-ethanol production from ligno-cellulosic material an attractive and potentially viable option for the production of liquid fuel.

The production of ethanol from plant cell wall material is achieved by harnessing the hydrolytic power of microbial GHs. These enzymes break down the polysaccharides, which comprise the plant cell wall, into simple monomeric sugars such as glucose. These simple sugars are then fermented, by yeast, into ethanol. The plant cell wall, however, is highly recalcitrant to enzyme degradation and this imposes

a significant cost to the production of ethanol from ligno-cellulosic material compared to the production of the alcohol from corn starch, which can be produced for ~$1 per gallon (Somerville 2007). The recalcitrance of the plant cell wall is imposed largely by cellulose, a highly crystalline polysaccharide consisting of β-1,4-D-glucose, which performs a purely structural role (Park 2010). Starch is composed of α-1,4-D-glucose making it far more accessible to enzymatic degradation (Fettke et al. 2009). Advances have been made in pre-treatments that open up the plant cell wall to enzymes (Zhang et al. 2007; Moxley et al. 2008), and the cost of the GHs used in biofuel production, mainly cellulases, have been reduced by 90 % (Somerville 2007). Furthermore, yeast strains have been developed that can utilise sugars other than glucose, such as xylose, which are derived from the degradation of hemicellulosic polysaccharides (Wenger et al. 2010; Zhang et al 2010.). Fermentation strategies are also being developed that use organisms other than yeast, an example of which is *Clostridium acetobutylicum*, capable of producing molecules with higher energy contents than ethanol, such as butanol (Li et al 2010.; Liu et al 2010.; Oshiro et al 2010.). All these advances have lowered costs dramatically but, further breakthroughs are needed in all these areas to make second generation bio-fuels a competitive alternative to liquid fuel production from fossil fuels. The focus of this thesis is the study of GHs that deconstruct the plant cell wall; enzymes that are fundamental to the production of cheap bio-ethanol production.

1.2 Plant Cell Wall

1.2.1 Structure

The plant cell wall is a complex, macromolecular, extracellular matrix that is presented on the surface of the plasma membrane. It fulfils many roles providing structural integrity, intracellular communication, microbial defence and water transport (Reiter 2002; Ryden et al. 2003). The cell wall consists of multiple layers (Figure 1.1). The first layer is the middle lamella and is deposited just after cell division, which is formed at the cell plate as the two cells divide (Beguin et al. 1994). The primary cell wall (P) is then formed, over the middle lamella, and is important during growth and development. When cells have ceased growing they may deposit a secondary cell wall (S), which is formed between the plasma membrane and the

primary cell wall and affords greater structural strength (Beguin et al. 1994). The secondary cell can be deposited in distinct layers S1, S2 and S3 (Figure 1.1).

The primary cell walls are composed of cellulose, hemicellulose and pectin. Pectin is an important feature and comprise the most complex polysaccharides known. They form hydrated gels that push cellulose microfibrils apart, easing their slippage during cell growth and also locking them in place when growth ceases (Cosgrove 2005). Secondary cell walls comprise cellulose, hemicellulose and lignin, an aromatic

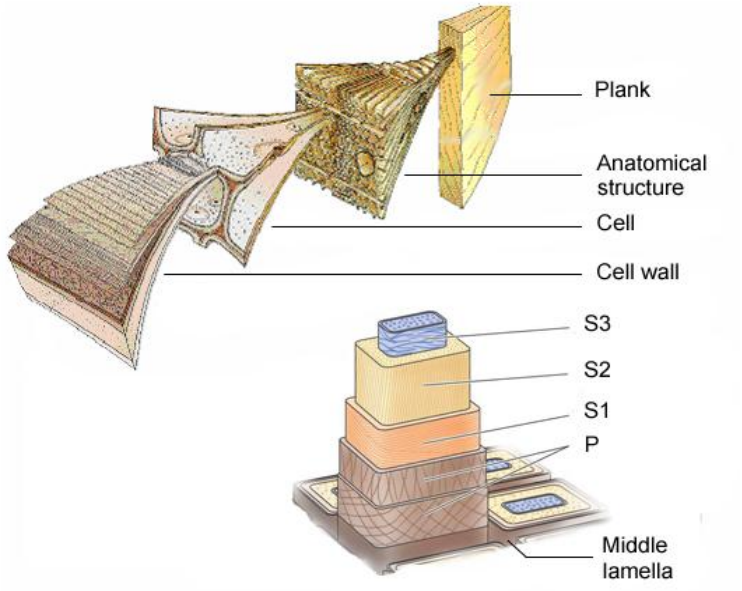


Figure 1.1 Schematic of the plant cell wall structure reproduced from Beguin and Aubert 1994. P=Primary cell wall; S1, S2 and S3=Secondary cell wall.

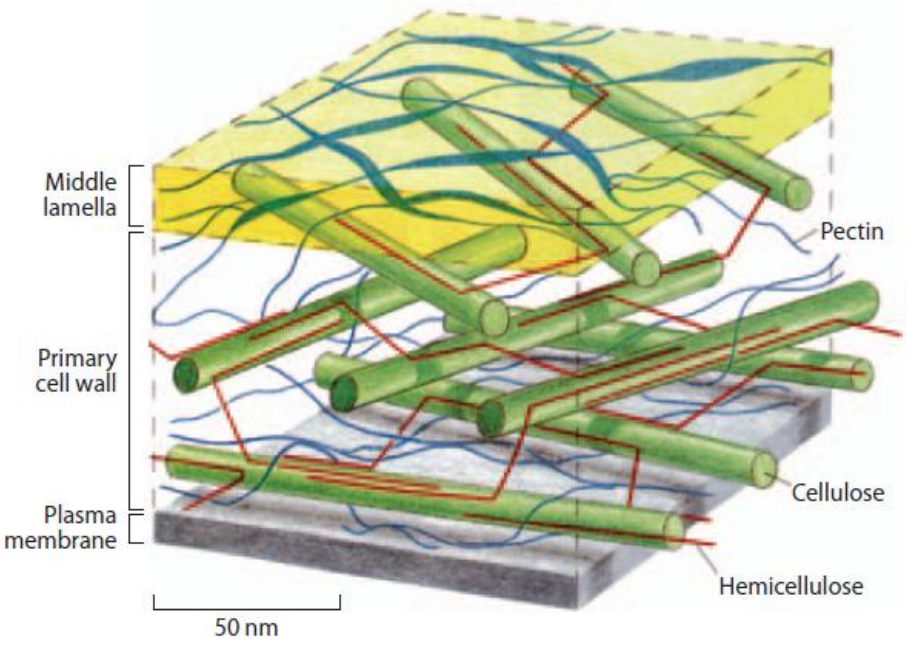


Figure 1.2 Schematic of the higher order architecture of the primary plant cell wall. Figure reproduced from (Scheller 2010).

phenolic macromolecule. Secondary cell walls contain little pectin and are laid down when growth of the cell has ceased. The presence of lignin helps strengthen and waterproof the wall, making it important for specialised cells such as xylem (Hofte 2010). The cellulose microfibrils of the secondary cell wall are arranged parallel to one another providing further structural strength (Muller et al. 2002). This is in contrast to the primary cell wall where cellulose microfibril arrangement is thought to be random (McNeil et al. 1984). The macromolecular structure of the primary plant cell wall is represented in Figure 1.2. The cellulose microfibrils form the bulk of the wall, while hemicellulose and pectin form a complex network of inter and intra molecular hydrogen bonds, coating the cellulose microfibrils, thereby reducing enzymatic accessibility to these structures (Scheller 2010).

1.2.2 Cellulose

Cellulose is not only the most abundant polysaccharide in the plant cell wall, but is the most abundant organic molecule in the biosphere. It is a homo polymer of β-D-1,4 glucose with no branches, where the repeating unit is a disaccharide displaying a two fold screw axis (Carpita 1997). Cellulose fulfils a purely structurally role in the plant cell wall. It is synthesised in long chains, which can contain as many as 15,000 glucose molecules, by cellulose synthase at the plasma membrane (Saxena et al. 2001). Individual cellulose chains, from 30-100, associate through extensive intra and inter molecular hydrogen bonds to form microfibrils 3-5 nm in diameter, providing structural strength to the cell wall (Figure 1.3) (Beguin et al. 1994). These microfibrils have a crystalline structure and six different crystal forms have been observed. Cellulose I is the dominant form and has paracrystalline/amorphous regions interspersed between highly ordered crystalline regions of the cellulose microfilbril (Teeri 1997; Park S 2010). Depending upon the plant species, however, the ratio of crystalline to paracrystalline/amorphous cellulose can vary greatly. In cotton about 70 % of the cellulose is crystalline compared to the algae *Valonia macrophysa* where the cellulose is almost 100 % crystalline (Wood 1988; Lehtio et al. 2003).

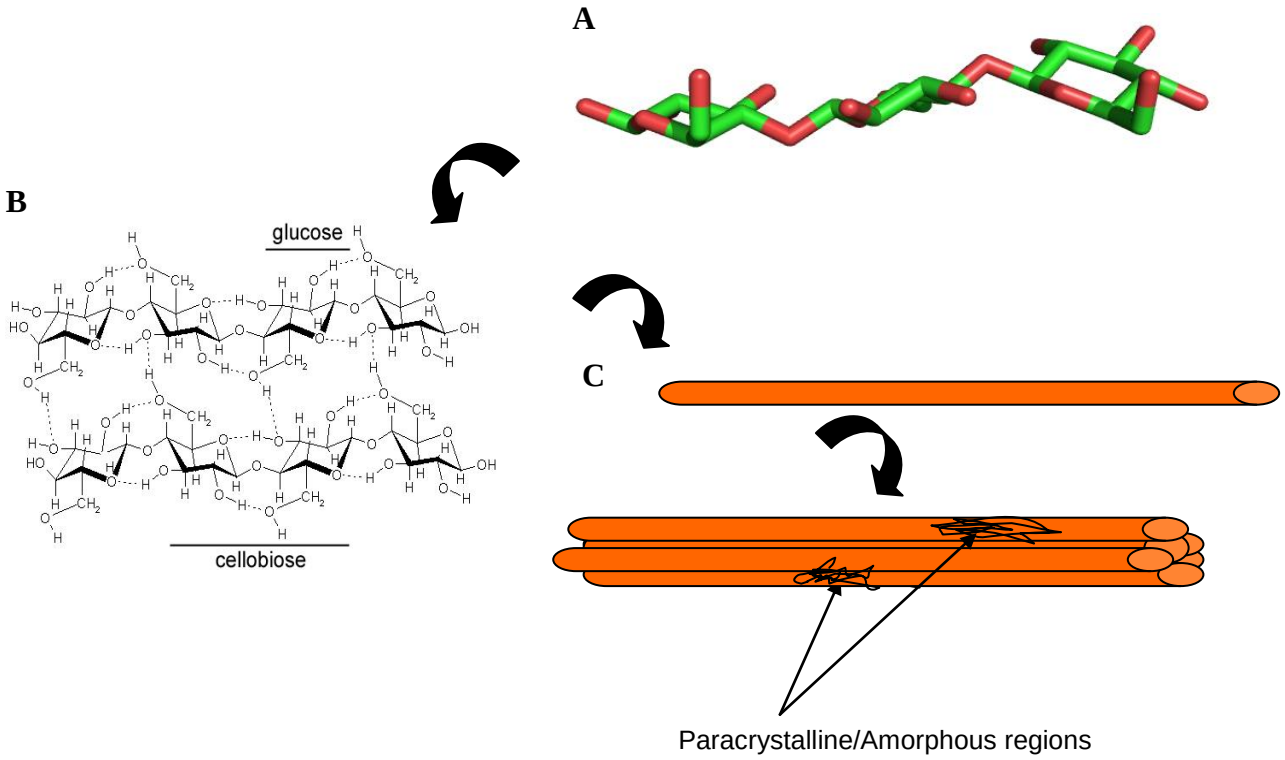


Figure 1.3 Schematic representation of the structure of cellulose microfibrils.
A=Cellotriose shows a two-fold screw axis (unpublished data); B=Diagram of extensive intra and inter molecular hydrogen bonds involved between glucan chains (Taken from Cedric YG Montaniers Thesis); C= Diagram of microfibril formation.

1.2.3 Hemicellulose

Hemicellulose is a collective term for a group of hetreogenous polysaccharides that contain a β-1,4 linked backbone. These polysaccharides include xylans, xyloglucans, mannans and β-1,3,1,4 glucans. These types of hemicellulose are present in all terrestrial plants, except β-1,3,1,4 glucans, which are almost exclusively found in Poales. The structure and types of hemicellulose present in the plant cell wall vary greatly between plant species (Scheller 2010). The role of hemicellulose is to interact with and cross link cellulose fibrils, and in some cases lignin, strenghting the cell wall. Hemicelluose is synthesised by glycosyltransferases at the Golgi membrane, packaged into vesicles and secreted at the plasma membrane. This is in contrast to cellulose which is synthesised at the plasma membrane (Lerouxel et al. 2006; Scheller 2010).

1.2.3.1 Xylan

Xylans are a diverse group of polysaccharides with the common feature of a β-1,4 linked xylose backbone that has a three fold screw axis, each xylose in the

backbone is rotated 120 ° relative to the preceding xylose (Figure 1.4A). The backbone can be substituted with α-L-arabinofuranose at O2 and/or O3 (Ebringerova 2005). Arabinose side chains can themselves be substituted with ferulic acid groups at C5. Another common modification is the presence of α-1,2-D-glucuronic acid, which can be substituted at O4 with a methyl group. Commelinid monocot xylans are mainly substituted with arabinose and are referred to as arabinoxylans, while xylan in the secondary cell walls of dicots is predominantly substituted with glucuronic acid and is referred to as glucuronoxylan. In eucalyptus α-galactose has been observed linked to the O2 position of glucuronic acid side chains (Figure 1.4B) (Shatalov et al. 1999). In dicots and conifers a conserved oligosaccharide of β-D-Xylose-(1,4)-β-D-Xylose-(1,3)-α-L-Rhamnose-(1,2)-α-D-Galactose-(1,4)-Xylose has been observed at the reducing end of xylan chains (Johansson 1977; Andersson 1983). This feature has not been observed in grasses but the genes responsible are conserved. Most xylans are acetylated to various degrees, especially at O3 (Ebringerova 2005).

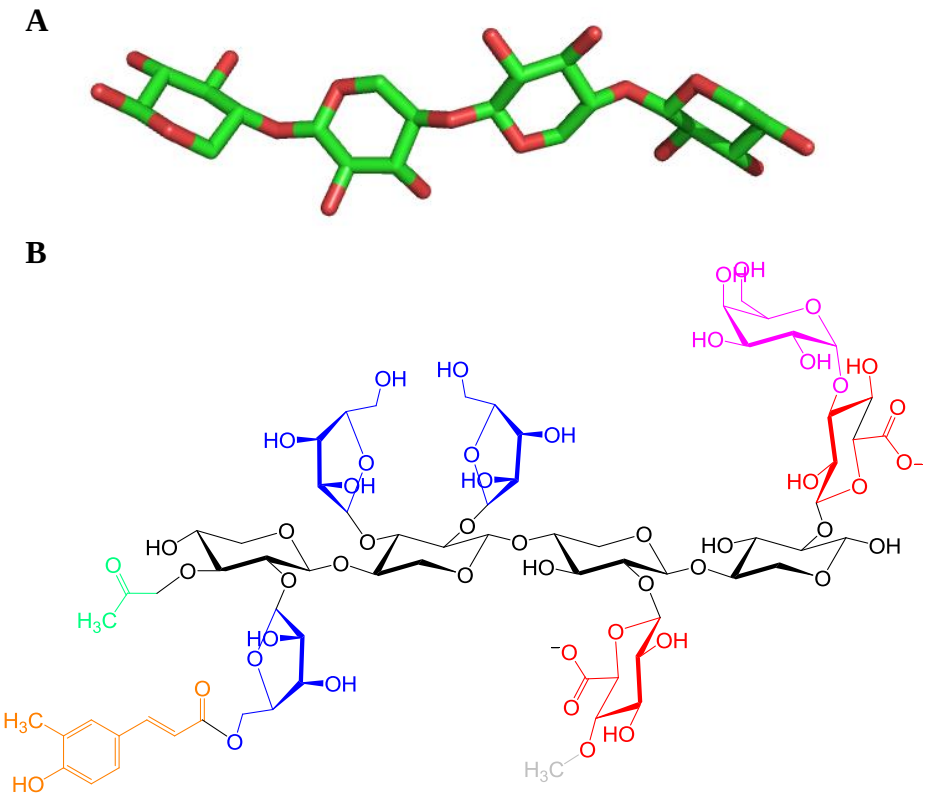


Figure 1.4 Structure and composition of arabinoxylan
A=Xylotetraose demonstrating 3 fold screw axis taken from pdb 1US2; B=Schematic of arabino/glucuronoxylan. Black=β-1,4 Xylose; Blue=α-1,2 and α-1,3 L-Arabinofuranose; Red=α-1,2 D-Glucuronic acid; Magenta=α-1,2 Galactose; Green=Acetyl group; Grey=Methyl group; Orange = Ferulic acid.

1.2.3.2 Xyloglucan

Xyloglucan is present in all terrestrial plants analysed to date but has not been found in charophytes (Moller et al. 2007; Popper 2008). It has a β-1,4 glucose backbone that is regularly substituted with α-1,6 xylose. The xylose side chains can then be substituted at O2 with β-galactose (Figure 1.5) or α-L-arabinofuranose; galactose side chains can be modified with α-1,2-L-fucose (Figure 1.6) (Scheller 2010). A one letter code has been developed for xyloglucan with G denoting unbranched glucose; X symbolises glucose substituted with xylose; L denotes a galactose appended to the xylose decoration; S shows an arabinofuranose side chain is present and F for a fucose side chain (Scheller 2010). The branching pattern of xyloglucan has taxonomic and functional significance. Lower amounts of branching correlate to a reduced solubility of the polysaccharide. Profound differences in charge have been found between xyloglucans in vascular plants and mosses, with the former having neutral side chains and the latter charged (Pena et al. 2008). Among vascular plants there are two repeating core oligosaccharides observed, these being XXGG and XXXG, with XXGG dominating in commelinoid monocots and XXXG dominating in solanaceous plants (York et al. 1996).

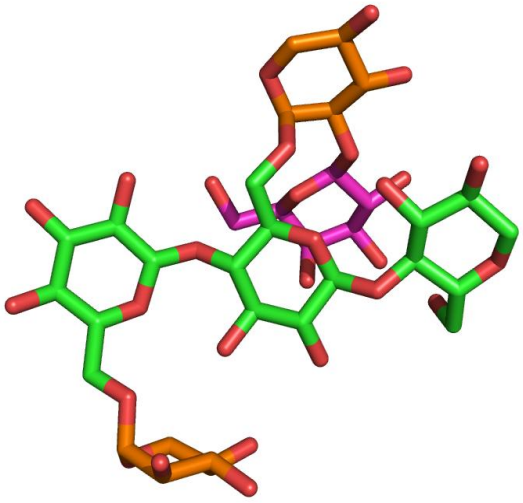


Figure 1.5 Crystal structure of a Xyloglucan hexasaccharide
Hexasaccharide of xyloglucan; β-1,4 Glucose=Green; α-1,6 Xylose=Orange; β-1,2 Galactose= Magenta. Pdb accession number 1UMZ

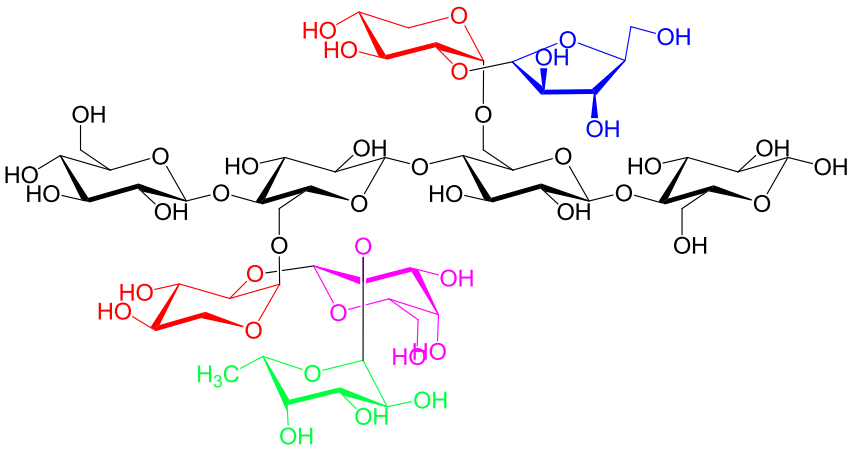


Figure 1.6 Schematic of xyloglucan structure and composition
Black = β-1,4 Glucose; Red = α-1,6 Xylose; Blue = α-1,2-L-Arabinofuranose; Magenta = β-1,2 Galactose and Green = α-1,2 L-Fucose.

1.2.3.3 Mannan

Mannan is the collective term for the polysaccharides mannan, galactomannan, glucomannan, galactoglucomannan and glucuronomannan. Mannan is an undecorated β-1,4 linked homopolymer of mannose that displays the same two fold screw axis as seen in cellulose (Figuer 1.7A) (Moreira et al. 2008). Mannan can form crystalline microfibrils much like cellulose and in some algae the mannose polymer completely substitutes for cellulose. Galactomannan comprises a β-1,4 linked backbone of mannose but contains galactose linked α-1,6. The amount of galactose substitutions varies between species but mannose:galactose ratios of 1:1 to 5:1 have been observed (Aspinall 1980; Brett 1996). This polysaccharide forms part of secondary cell wall thickening and is involved in the taking up of water in seeds (Brett 1996). Glucomannan is a heterogenous backbone of β-1,4 glucose and mannose and is a major hemicellulose found in gymnosperms (Scheller 2010). Glucomannan can also be substituted with α-1,6 galactose, on mannose residues only, and this polysaccharide is defined as galactoglucomannan (Figure 1.7B) (Moreira et al. 2008). Galactomannan, glucomannan and galactoglucomannan can be acetylated, at varying degrees, at O2 and O3 of mannose. Glucuromannan is found in small amounts in many cell walls and is composed of a α-1,4 mannose and β-1,2 glucuronic acid backbone with side chains of galactose and arabinose (Moreira et al. 2008)

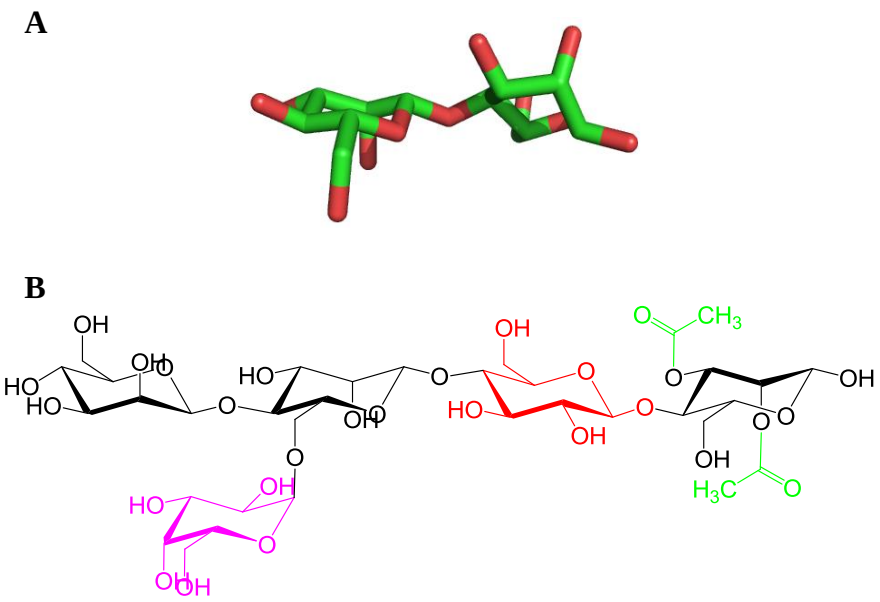
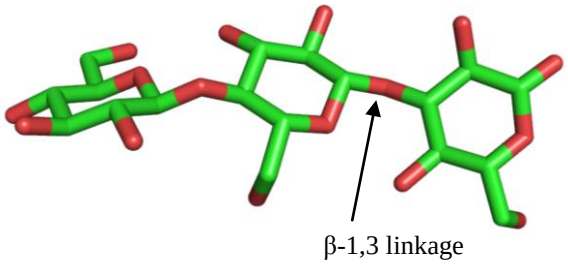


Figure 1.7 Structure of mannan polysaccharides
A=mannobiose showing two fold screw axis, pdb accession number 2VX7. B=Schematic of Galactoglucomannan Black=β-1,4 mannose; Red=β-1,4 Glucose; Magenta=α-1,6 Galactose and Green=Acetyl groups.

1.2.3.4 β-1,3, β-1,4 mixed linked glucans

Mixed linkage glucans are composed primarily of β-1,4 glucose with interspersed β-1,3 linkages (Figure 1.8A and B). These mixed linkage glucans are dominated by β-1,4 cellotriosyl and cellotetrasyl repeats linked by β-1,3 linkages, but longer stretches of β-1,4 have been observed (Stone 1992). The presence of the β-1,3 linkages prevents this homopolymer of glucose forming crystalline structures like cellulose, and at high concentrations it forms a gel. Mixed linkage glucans play a role in cell expansion in primary cell walls and there abundance is very growth stage dependent (Gibeaut et al. 2005). The presence of mixed linkage glucans is almost exclusively in monocots, and its absence in dicots, has led to the idea that mixed linkage glucans evolved independently in grasses, although recent studies suggested the presence of mixed linkage glucans in more primitive taxa. This may suggest that it is a ancient trait that dicots have lost over the course of evolution, but the former hypothesis seems more likely (Scheller 2010).

A



B

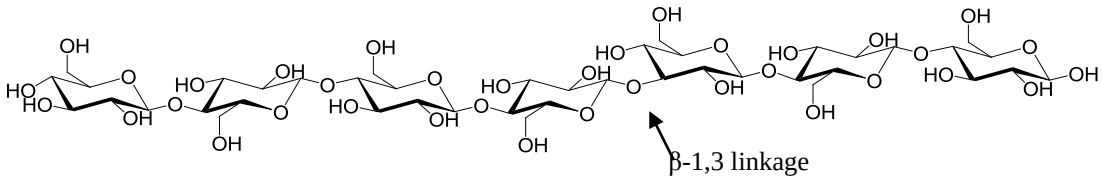


Figure 1.8 Structure of mixed linkage glucans
A=Trisaccharide of β-1,4 glucose, β-1,4 glucose, β-1,3 glucose pdb accession number 1ZM1. B=Schematic of the common structure observed in β-1,4, β-1,3 mixed linkage glucans.

1.2.4 Pectic Polysaccharides

Pectins are a family of complex polysaccharides that all contain α-1,4 galacturonic acid residues. Pectins can comprise up to 30 % of the primary cell wall in dicots and between 5 % and 10 % in monocots, such as grasses (Aspinall 1980; Brett 1996). Pectin is absent or in very low abundance in secondary cell walls, which is indicative of its important role in plant cell wall growth (Cosgrove 2005). The common pectic polysaccharides are homogalacturonan (HG), rhamnogalacturonan I (RGI) and rhamnogalacturonan II (RGII).

1.2.4.1 Homogalacturonan

Homogalacturonan (HG) is a homo polymer of α-1,4 galacturonic acid and can account for up to 60 % of the pectin found in primary cell walls. HG can be methyl esterified at the C5 carboxylate group. The amount of methyl esters can vary and HG with a high degree of methyl esterification is referred to as pectin where as HG with a low level of methyl esterification is referred to as pectic acid. The polysaccharide can also contain acetyl groups (Figure 1.9). The properties and biological role of HG is believed to be mediated through ionic interactions. High levels of methyl esterification therefore cause a reduction in negative charge, and have been associated with cell separation (Wolf et al. 2009).

1.2.4.2 Rhamnogalacturonan I

Rhamnogalacturonan I (RGI) is a pectic polysaccharide that has a backbone composed of a repeating disaccharide, consisting α-1,2 galacturonic acid and α-1,4-L-rhamnose. RGI contains polysaccharide side chains that are typically attached to rhamnose residues, however, the density of side chains is not regularly distributed leading to heavily decorated backbone regions referred to as "hairy regions" and the undecorated as "smooth regions"(Brett 1996). The side chains appended to rhamnose are linear β-1,4 galactose chains (Galactan), linear β-1,4 galactose chains substituted with β-1,6 galactose and α-1,4 arabinofuranose (arabinogalactan), linear α-1,5-L-arabinofuranose chains (Linear arabinan) and linear α-1,5-L-arabinofuranose chains substituted with monomeric α-1,2 and α-1,3-L-arabinofuranose side chains (Sugarbeet arabinan) (Figure 1.9) (Oomen et al. 2002). Little is known about the function of RGI, however, there is increasing evidence that changes in the structure of arabinan and galactan side chains are linked with different stages of cell and tissue development (Oomen et al. 2002).

1.2.4.3 Rhamnogalacturonan II

Rhamnogalacturonan II (RGII) was first described in 1978 and its structure has been shown to be conserved in monocots and dicots (Darvill et al. 1978). RGII is a low molecular weight pectic polysaccharide being 5 – 10 kDa in mass. It has a α-1,4 galacturonic acid backbone, the same as HG. There are four side chain substitutions observed in RGII two disaccharides and two complex oligosaccharides (Vidal et al. 2000; Glushka et al. 2003). RGII is the most complex polysaccharide identified to date consisting of 20 different monosaccharides, including apiose, aceric acid, 3-keto-3-deoxy-manno-octulosonic acid (Kdo) and 3-deoxy-lyxo-2-heptulosaric acid (Dha) and has 9 different linkages (Figure 1.9) (Vidal et al. 2000). Methyl groups and acetyl groups are also present in RGII further adding to its complexity. RGII exists as a dimer, being cross linked by a borate diol diester, between two apiosyl residues. The borate atom is chiral and thus two diasteroisomers can form but it is not known which isomer occurs naturally in plant cell walls (Kobayashi et al. 1996).

1.2.4.4 Xylogalacturonan

Xylogalacturonan has a α-1,4 galacturonic acid backbone that is substituted with β-1,3 xylose. The backbone can also be methylated and acetylated (Mohnen 2008). Xylogalacturonans are only present in reproductive tissues and little is known about the function of the polysaccharide.

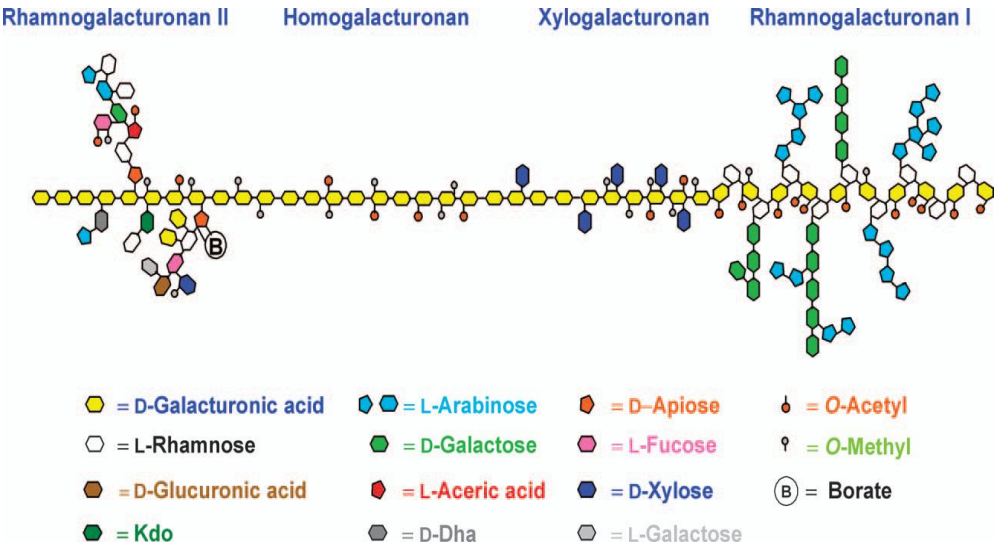


Figure 1.9 Schematic of the four common types of pectin in primary cell walls
Reproduced from Scheller 2007.

1.2.5 Lignin

Lignin is the generic term for a group of aromatic polymers resulting from the oxidative coupling of 4-hydroxyphenylpropanoids and is the second most abundant organic molecule in the biosphere after cellulose. These polymers are found in secondary cell walls imparting structural solidity and hydrophobicity. The main building blocks of lignin are coniferyl alcohol, sinapyl alcohol and relatively minor amounts of paracoumaryl alcohol, collectively known as monolignols (Figure 1.10). Lignin helps to protect cell walls against microbial degradation and adds to the recalcitrance of the wall to enzymatic breakdown (Vanholme et al 2010.).

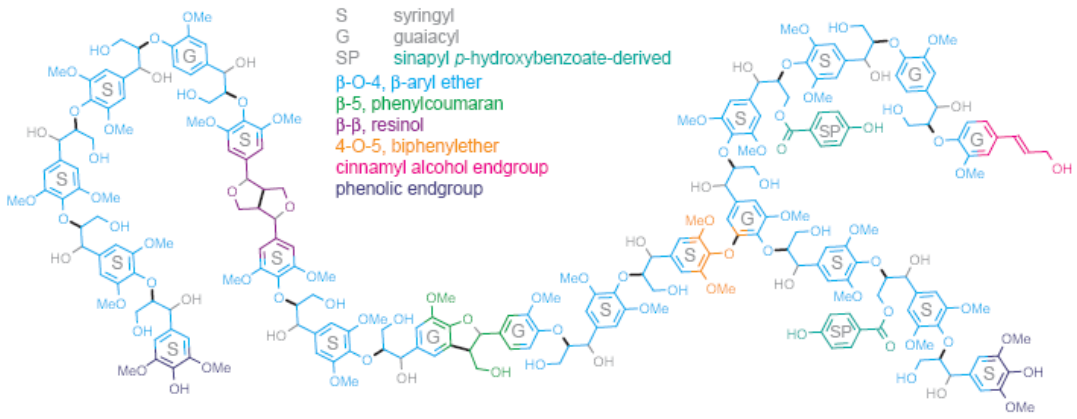


Figure 1.10 Schematic representation of lignin polymer predicted from poplar
Predicted by NMR analysis reproduced from Vanholme 2010.

1.2.6 Proteins in the Plant cell wall

Nearly 10 % of the dry weight of the plant cell wall is made up by proteins. Virtually all plant proteins are glycosylated. Plant O-glycans are usually linked to serine, theronine or hydroxyproline residues. Proteins containing glycans linked to hydroxyproline residues are termed hydroxyproline rich glycoproteins (HGRPs). A class of proteins known as expansins occur in the plant cell wall. These proteins are non-catalytic and are thought to alter plant cell wall structure by disrupting hydrogen bond interfaces between cellulose and hemicellulose. Expansins are modular proteins comprising of two domains, one that is structurally similar to Glycoside Hydrolase family 45 (GH45) and the other is related to grass pollen allergen domain two. The crystal structure of expansins reveals planar aromatic residues in both domains that are suggested to be involved in polysaccharide binding (Cosgrove 2000). Enzymes are also found in the plant cell wall and are involved in degradation and remodelling, GH16s which are xyloglucan:xyloglucosyltransferases and xyloglucanases, are elegant examples of this (Johansson et al. 2004).

1.3 Microbial Glycoside Hydrolases

Annually 10^{11} tonnes of plant structural polysaccharides are degraded by microbial glycoside hydrolases (GHs) (Somerville 2007). These GHs are expressed by saprophytic soil dwelling microbes, rumen microorganisms and bacteria residing in the gut of lower eukaryotes such as insects like termites (Zhou et al. 2007). The plant cell wall represents a major source of carbon and energy, however, the component polysaccharides are extremely stable with cellulose having an estimated half life of ~five million years (Wolfenden et al. 2008). For microbes to utilise these

polysaccharides, as an energy source, the GHs expressed by these organisms have to be highly efficient. In fact the rate enhancement displayed by cellulases (enzymatic versus non-enzymatic hydrolysis of cellulose) is believed to be greatest seen in nature (Wolfenden et al. 2001; Wolfenden et al. 2008)

1.3.1 Classification and nomenclature

Due to the importance of GHs in plant cell wall degradation, as well as other biological processes, a great number of genes encoding these enzymes have been isolated, sequenced and the encoded proteins characterized. A system to classify GH sequences was first proposed by Henrissat in 1991 (Henrissat 1991a; Henrissat 1991b). This system evolved into the CAZy database which currently contains 118 GH families based on sequence similarity and the conservation of the tertiary fold, catalytic mechanism and catalytic apparatus. Many of these families have been grouped into clans. Families within a clan display the same fold and the mechanism and catalytic residues are similarly conserved. However, there is little overall sequence similarity between different families within a specified clan (Henrissat et al. 1995; Henrissat et al. 1995; Henrissat et al. 1997; Cantarel et al. 2009). All families within a clan are believed to have evolved from a common ancestor.

There are currently seven tertiary folds displayed by GHs: (β/α_8), β-jelly roll, 6 fold β-propeller, 5 fold β-propeller, α/α_6, α+β and β-helix (Henrissat 1991a; Henrissat et al. 1995). These seven folds are organised into 14 clans GH-A to GH-N (Figure 1.11). Clan GH-A is the largest clan consisting of 18 families, this is four times the size of the next largest clan, GH-E, which contains only four families (Henrissat 1991a; Henrissat et al. 1995). Different clans may have the same fold but their divergence is based on a lack of conservation in the catalytic machinery and/or the catalytic mechanism (Henrissat et al. 1995). For example GH-A and GH-K both display the (β/α_8) fold but GH-A enzymes utilise a retaining mechanism, where the catalytic acid/base and nucleophile (see Section 1.3.2 for a description of catalytic mechanisms and catalytic residues), which are both glutamates, are found at the C-terminus of the β strands 4 and 7, respectively. Indeed, historically, clan GH-A was originally referred to as the β-4/7 superfamily in recognition of the location of the catalytic amino acids (Henrissat et al. 1996; Jenkins et al. 2001). Members of GH-K, of which there are three families, all operate via a substrate assisted mechanism where the carbonyl oxygen of the acetyl group at C2, of the substrate, acts as a catalytic

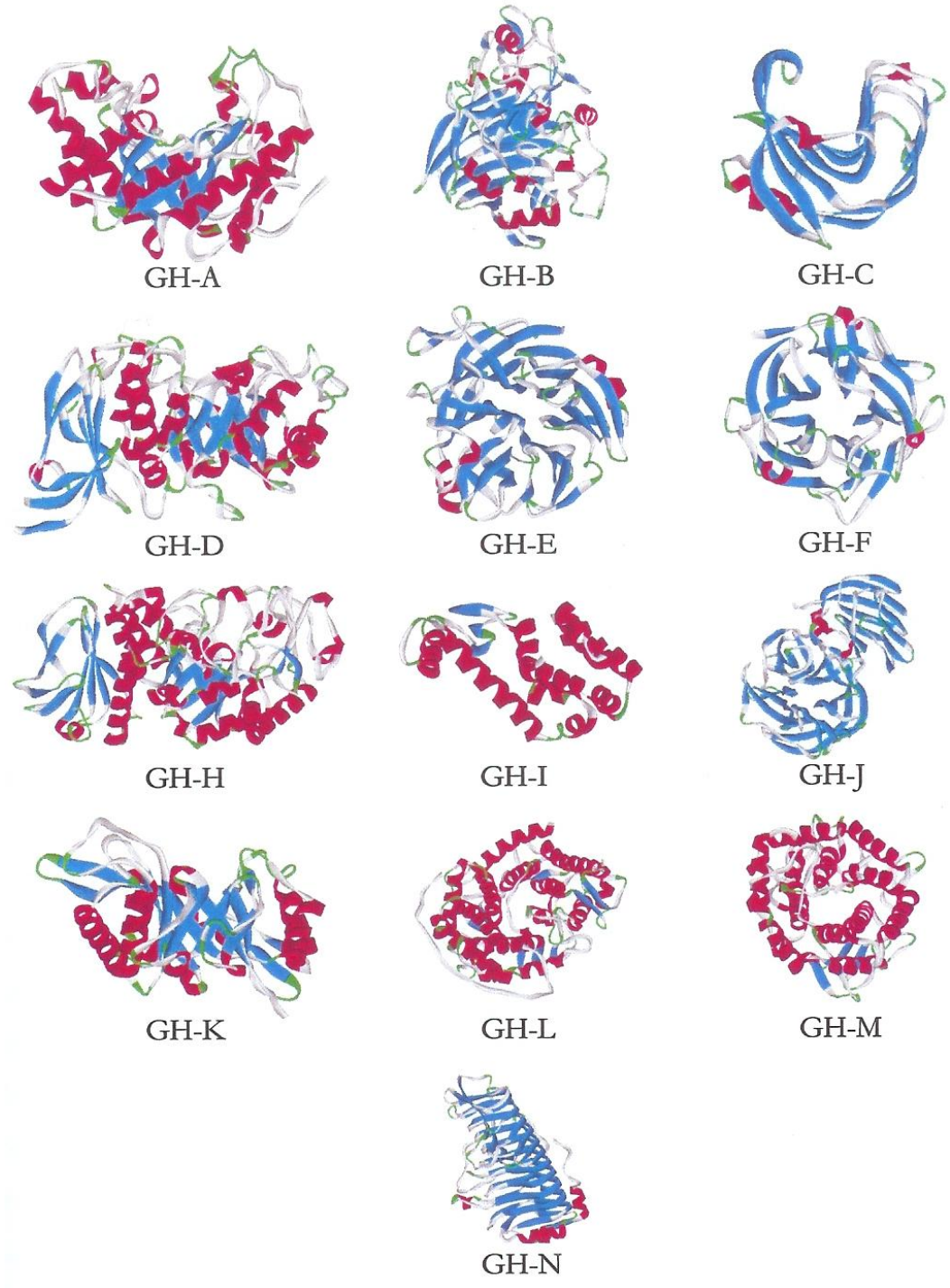


Figure 1.11 Diagram of Clan secondary struture taken from (Pell 2004).
Secondary structural elements are coloured as follows; red- α-helices, blue, β-sheet. The representative enzymes displayed are as follows for the different Clans; GH-A *Cellvibrio japonicus* xylanase Xyn10C (GH10); GH-B, *Hypocrea jecorina* cellobiohydrolase I (GH7); GH-C, *Aspergillus niger* XynA (GH11); GH-D, *H. jecorina* RuC-30 α-galactosidase 1 (GH27); GH-E, *Salmonella typhimurium* TA262 sialidase (GH33); GH-F, *C. japonicus* arabinanase Arb43A (GH43A); GH-H, *A. niger* α-amylase (GH13); GH-I, Bacteriophage T4 lysozyme (GH24); GH-J, *Thermotoga maritima* invertase (GH32); GH-K, *Flavobacterium meningospticum endo*-β-N-acetylglucosaminidase F1 (GH18); GH-L, *A. awamori* var. X-100 glucoamylase (GH15); GH-M, *Clostridium thermocellum endo*-β-1,4-glucanase; GH-N, *A. aculeatus* rhamnogalacturonase A (GH28).

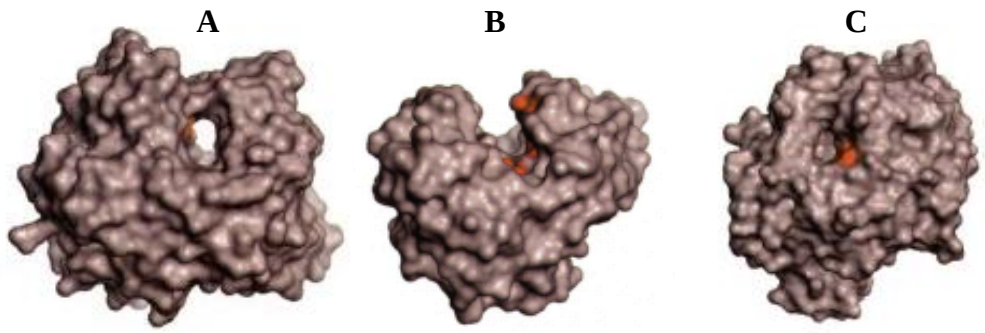


Figure 1.12 surface topography of glycoside hydrolases
Panel A 'tunnel' topography, commonly found in processive exo-acting cellobiohydrolases. Panel B 'open cleft' topography, found in endo-acting enzymes. Panel C 'pocket' topography, found in non-processive exo-acting enzymes. Adapted from Davies and Henrissat 1995.

nucleophile, and thus the enzymes only contain one catalytic residue; the acid-base. All members of GH-K act on chitin or similar substrates where C2 of the sugar is substituted with an acetyl group (Mark et al. 2001; van Aalten et al. 2001; Ling et al. 2009).

The seven different tertiary folds display three general topologies (Henrissat et al. 1997). The open cleft topology generally has a cleft which extends the length of the enzyme and can accommodate a single polysaccharide chain (Figure 1.12B). Enzymes displaying this topology are mostly endo-acting and cleave randomly along polysaccharide chains producing varying sized fragments depending on the composition of the polysaccharide and the number of subsites the enzyme utilises (Davies et al. 1995). The pocket topography is displayed by exo-acting enzymes which are non-processive (Figure 1.12C). Enzymes displaying this topology usually act on side chains of polysaccharides, providing greater access for the endo-acting enzymes, or convert oligosaccharides to their monosaccharide substituents (Davies et al. 1995). The final topology observed in GHs is the tunnel topography (Figure 1.12A). This topography has only been observed in cellobiohydrolases. It is thought to have evolved from the open cleft topology with the sides of the cleft extending to form a ceiling. The tunnel topography confers processivity to the enzymes, so cleavage of substrate can occur with product release but without release of the substrate itself. Cellobiohydrolases cleave individual glucan chains of cellulose, moving along processively releasing cellobiose, the repeating structure of cellulose chains (Davies et al. 1995).

A modern designation of GHs takes into account their activity, family, organism and isoform. For example, *Cj*Man26A is a mannanase (Man), from GH

family 26 (26), from the organism *Cellvibrio japonicus* (*Cj*) and was the first GH26 identified from this organism (A). Subsequent GH26 mannnanases are defined as *Cj*Man26B, C, D etc (Henrissat et al. 1997).

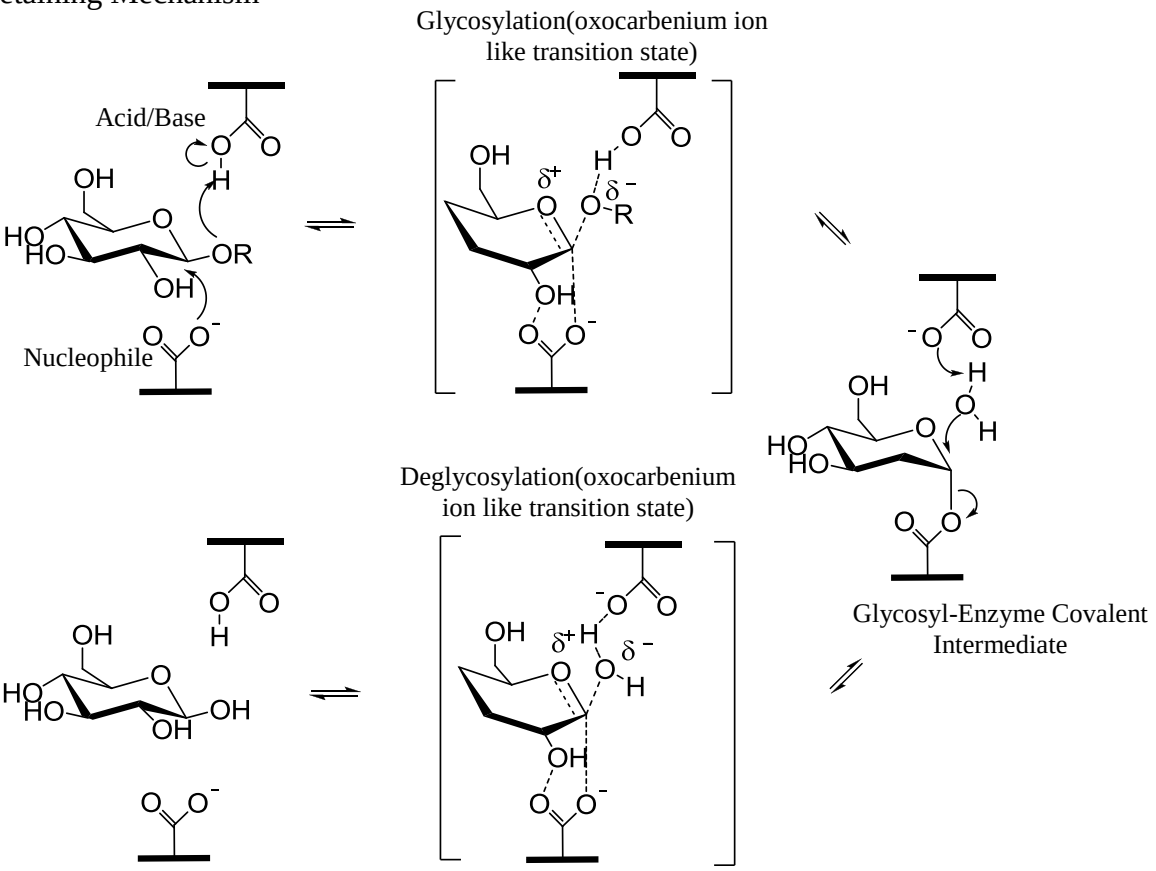
1.3.2 Main Catalytic Mechanisms

There are two main catalytic mechanisms employed by glycoside hydrolases (GHs). These are the retaining or double displacement mechanism, which proceeds through a glycosyl-enzyme covalent intermediate and an inverting mechanism or single displacement mechanism that does not form a covalent intermediate. Retaining or inverting refers to the geometry of the anomeric carbon after hydrolysis. If a β bond is cleaved and a β bond is formed the reaction has retained the configuration of the anomeric carbon, conversely if a β bond is cleaved and an α bond is formed the reaction has inverted the configuration of the anomeric carbon. α and β refers to the configuration of the anomeric carbon, C1, relative to the configuration of the chiral centre furthest away, typically C5. If the two have the same configuration then the anomeric carbon is said to be β and if they have the opposite anomeric configuration they are said to have an α configuration. Typically, but not always, α bonds are axial in orientation, while β bonds are equatorial in orientation (Koshland et al. 1954; Sinnott 1990).

1.3.2.1 β–Retaining Mechanism

Reactions catalyzed by GHs that display a retaining mechanism proceed through a glycosylation and deglycosylation step, and both components of the reaction progress through an oxocarbenium ion like transition state (Figure 1.13). The mechanism utilises a nucleophile, which carries a negative charge, and an acid/base, which needs to be protonated at the optimum pH of the enzyme, which is frequently around 7.0. The pKa of the acidic amino acids Asp and Glu is ~4.0 (MacLeod et al. 1994; Tull et al. 1994). The acid/base is thus situated in an apolar environment which alters the pKa of the acid/base so it is protonated at physiological pH. Also, the catalytic acid/base and nucleophile are ~5.5 Å apart in retaining enzymes and it is unfavourable to have two negative charges in such proximity and this ‘forces’ the acid/base, which is situated in a more apolar environment, to be protonated (Davies et al. 1995). If the nucleophile is mutated the acid/base is no longer protonated and can give rise to a class of enzyme coined ‘glycosynthases’ discussed later

β-Retaining Mechanism



α-Retaining Mechanism

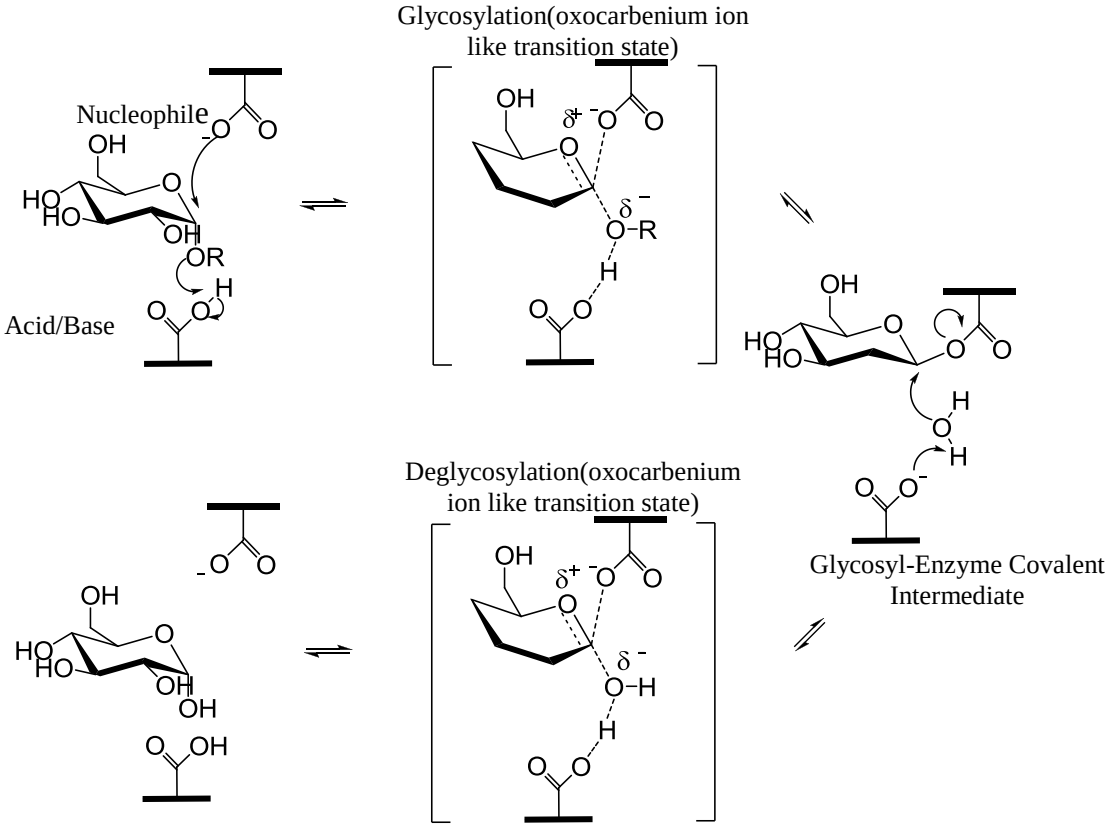


Figure 1.13 Scehmatic of the α-Retaining Mechanism and the β-Retaining Mechanism

(Ly et al. 1999; Mayer et al. 2000; Shaikh et al. 2008). In the first step of the reaction, which hydrolyses β-linkages, the catalytic nucleophile attacks the anomeric carbon, C1, from below, the α face of the sugar, at the same time the acid/base donates a proton to the leaving group, which is typically, but not always, a sugar. Catalysis proceeds through an oxocarbenium ion like transition state, which promotes the formation of a positive charge at C1. The sugar is distorted from a relaxed ground state $^{4}C_{1}$ chair to a one of four possible transition state conformations, $^{4}H_{3}$, $^{3}H_{4}$, $^{2,5}B$ or a $B_{2,5}$ ($^{4}H_{3}$ is depicted in Figure 1.13), which all place C1, C2, C5 and the endocyclic ring oxygen in the same plane (Sinnott 1990; Zechel et al. 2000). Crucially this allows the generation of the positive charge at C1 to be shared between C1 and the endocyclic ring oxygen, leading to partial sp^2 hydridisation and stabilising the transition state. The positive charge at C1 facilitates attack by the catalytic nucleophile and formation of the covalent intermediate. During catalysis the proton of O2 becomes more acidic, interacting favourably with one of the lone pair of electrons of the nucleophiles carbonyl group, helping support electrophilic migration of the sugar towards the nucleophile as the covalent intermediate is formed (MacKenzie et al. 1998). The O2 hydroxyl in β-retaining enzymes has been shown to provide 3-4 fold more energy to catalysis than any other hydroxyl (Namchuk et al. 1995). This first step, known as glycosylation, leads to an α linked glycosyl enzyme intermediate covalently linked through an ester bond.

In the second step of the reaction, known as deglycosylation, the acid/base now acts as a general base abstracting a proton from an incoming water molecule and promoting the formation of a hydroxyl ion. The hydroxyl ion then attacks C1 from above, the β face of sugar, promoting cleavage of the α linked glycosyl-enzyme intermediate and leading to the formation of a β orientated hydroxyl and the final product. This step also proceeds through an oxocarbenium ion like transition state as described above. The retaining mechanism can be thought of as two inversions leading to an overall retention. At both steps of the transition state bonds are simultaneously being made and broken making hydrolysis by GHs akin to S_N2 nucleophillic substitution. There is another possibility that can occur at the second stage which is termed transglycosylation (Sinnott 1990). Here another sugar, rather than a water molecule, enters the active site and the acid/base is able to abstract a proton from a hydroxyl group of the incoming sugar. This activates the hydroxyl group which can then attack C1 from above, the same as a water molecule described

above, leading to cleavage of the α linked glycosyl-enzyme intermediate forming a product that is β linked to another sugar. This product, however, is a substrate and is hydrolysed by the enzyme again. There are exceptions to the hydrolysis of the transglycosylation product, such as in GH16, where some members of this family favour transglycosylation over hydrolysis (Johansson et al. 2004; Baumann et al. 2007).

1.3.2.2 α-Retaining Mechanism

In the α-retaining mechanisms α linked substrates are hydrolysed and produce products with an α anomeric configuration. The reaction is very similar to that described for the β-retaining mechanism with glycosylation and deglycosylation proceeding through an oxocarbenium ion like transition state and the formation of a covalent intermediate between the two transition states (Figure 1.13). In α-retainers the covalent intermediate is β linked, and the activated water attacks the glycosyl-enzyme intermediate from the α face of the sugar (Sinnott 1990; Zechel et al. 2000). There are subtle differences at the transition state with the important interaction of O2 with the nucleophile being negated in the α-retaining mechanism, as O2 does not come into the same proximity with the nucleophile during catalysis, as occurs in the β-retaining mechanism. The endocyclic ring oxygen, however, does come into closer proximity with the lone pairs of electrons of the nucleophile carbonyl group during electrophillic migration. This causes a greater amount of the positive charge shared between C1 and the endocyclic ring oxygen to reside at the endocyclic ring oxygen, making the interaction between the two oxygen atoms more favourable (Zechel et al. 2000). A general kinetic flow applicable to initial hydrolysis for both α-retaining and β-retaining enzymes is shown in Figure 1.14.

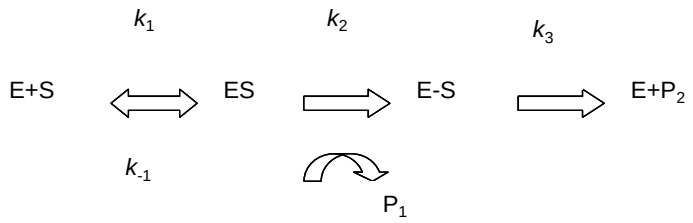
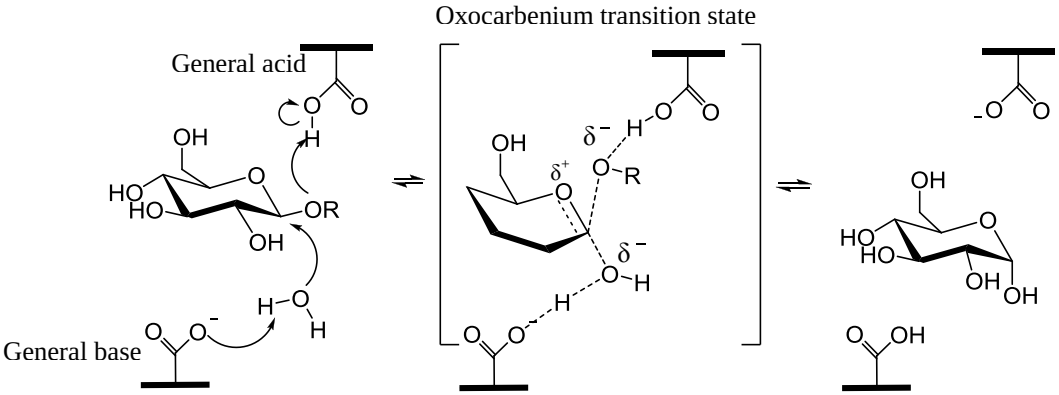


Figure 1.14 Kinetic flow for retaining enzymes
E=Enzyme; S=Substrate; ES=Michaelis complex, E-S=Covalent intermediate; P_1=Product 1; P_2=Product 2; k_1=rate constant for formation of ES complex, k_{-1}=rate constant for breakdown of ES complex, k_2= rate constant for covalent intermediate formation, k_3= Catalytic constant for hydrolysis of the covalent intermediate

1.3.3.3 Inverting Mechanism

In the β-inverting mechanism β linked substrates are hydrolysed to yield products with an α anomeric configurations. Inverting mechanisms also utilise two acidic catalytic amino acids, these being a general base and a general acid. The two catalytic residues are typically 10 Å apart (Davies et al. 1995). Again the residue acting as the general acid needs to be protonated, and this is usually achieved through the apolar environment in which the general acid is situated. In many inverting enzymes the general base is too far away to exert an influence on the general acid. Some inverting enzymes, such as those of GH43, contain a third essential acidic residue that serves to modulate the pKa of the general acid ensuring it is protonated (Nurizzo et al. 2002). The β-inverting mechanism proceeds through a single catalytic step. Here the general base abstracts a proton from a water molecule, referred to as the nucleophilic water (NW), generating a hydroxyl ion which attacks C1 from the α face of the sugar, at the same time the general acid donates a proton to the glycosidic oxygen, which assists departure of the leaving group (Figure 1.15). Again catalysis proceeds through an oxocarbenium ion like transition where the relaxed ground state $^{4}C_{1}$ chair is distorted into a $^{4}H_{3}$, $^{3}H_{4}$, $^{2,5}B$ or a $B_{2,5}$ ($^{4}H_{3}$ is depicted in Figure 1.15), which all place C1, C2, C5 and the endocyclic ring oxygen in the same plane, allowing the positive charge generated at C1, which promotes nucleophilic attack, to be shared with the endocyclic ring oxygen via partial sp^2 hybridisation (Sinnott 1990; Zechel et al. 2000). Inverting enzymes cannot perform transglycosylation, but can also be made into glycosynthases, which are discussed later (Hehre et al. 1979; Honda et al. 2006; Wada et al. 2008). In the α-inverting mechanisms α linked substrates are hydrolysed to generate reaction products that display a β anomeric configuration. Catalysis again proceeds through a single catalytic step, deploying an oxocarbenium ion transition state as described above (Figure 1.15). The general base abstracts a proton from a NW generating a hydroxyl ion, which attacks C1 from the β face of the sugar, while the general acid simultaneously protonates the leaving group sugar. Bond cleavage and formation occur simultaneously and both inverting mechanisms are akin to S_N2 nucleophilic substitution. Although inverting enzymes cannot transglycosylate they are capable of the Hehre-resynthesis-hydrolysis mechanism, which was first demonstrated with an α-amylase (Hehre et al. 1979).

β-Inverting Mechanism



α-Inverting Mechanism

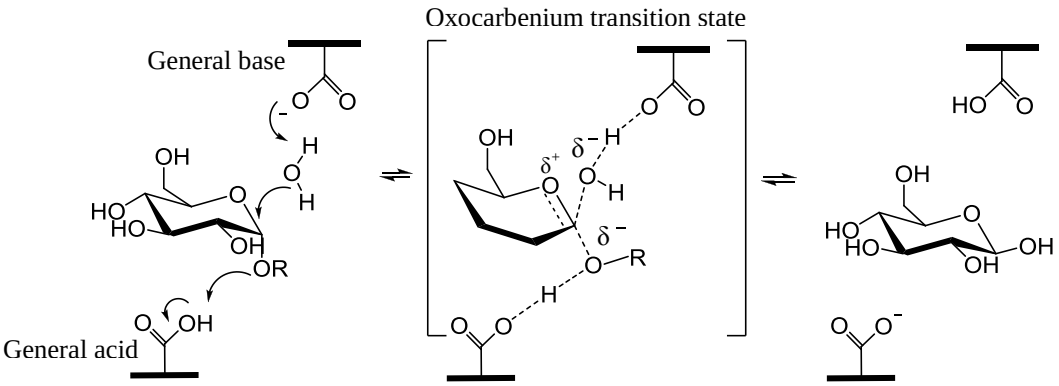


Figure 1.15 Schematic of α-Inverting Mechanism and β-Inverting Mechanism

When the enzyme was given β-maltosyl fluoride, the opposite anomeric configuration to the α glycosides normally hydrolysed, it was able to convert the substrate to maltose and fluoride ions. Later various other inverting enzymes were shown to catalyse the hydrolysis of the 'wrong' glycosyl fluoride, suggesting the reaction is not uncommon among inverting enzymes. When the glycoside fluoride, with the opposite anomeric configuration to the natural substrate, is cleaved in the presence of a sugar, resynthesis of a glycosidic bond occurs, with the anomeric configuration of the natural substrate, which is subsequently hydrolysed on enzyme (Hehre et al. 1979). Figure 1.16 shows the kinetic flow for the initial stages of both the α-inverting and β-inverting mechanisms.

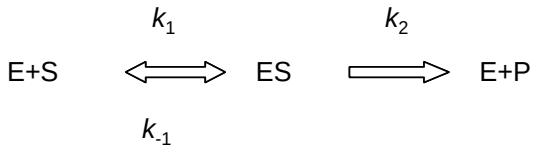


Figure 1.16 Kinetic Flow for Inverting EnzymesE=Enzyme, S=Substrate, ES=Enzyme-Subtrate complex, P=Product.k_1=rate constant for formation of ES complex, k_{-1}=rate constant for breakdown of ES complex, k_2= Catalytic Constant.

1.3.4 Glycosynthases

The term glycosynthase refers to glycoside hydrolases (GHs), which have been mutated so that they synthesise, rather than hydrolyse, glycosidic linkages. This was first demonstrated in 1998 by Mackenzie et al (MacKenzie at al. 1998) with the β-retaining enzyme *Agrobacterium* sp. β-1,4-glucosidase. Retaining enzymes, as described above, have the ability to transglycosylate and synthesise glycosidic linkages, but the product is a substrate and is therefore immediately hydrolysed (Sinnott 1990). If the rate of transglycosylation could be enhanced and the rate of hydrolysis lowered, then retaining glycosidases could be used as synthetic enzymes. Mackenzie et al (MacKenzie et al. 1998) achieved this with *Agrobacterium* sp. β-1,4-glucosidase by mutating the catalytic nucleophile to Ala, which ablates hydrolytic activity and causes the acid/base to become deprotonated. Using an α-linked fluoro glucoside acceptor and glucose donor, the mutant enzyme was now able to synthesise, but not subsequently hydrolyse, β-1,4 glycosidic bonds. The α-linked fluoro glucosides were used for two reasons; firstly, fluoride is a good leaving group and does not need protonation, which is important as the acid/base is deprotonated thus cannot act as a general acid, secondly, it mimics the reactive glycosyl-enzyme intermediate. The deprotonated acid/base is able to act as a general base abstracting a proton from O4 of glucose (donor), and O4 is then able to mount a nucleophilic attack on C1, from the β face of the sugar, cleaving the carbon-fluorine bond and generating a β-1,4 glycosidic linkage (Figure 1.17). The *Agrobacterium* glycosynthase was able to generate glucose oligosaccharides to a length where they precipitated out of solution. α-linked fluoro galactoside was also used as an acceptor by the *Agrobacterium* glycosynthase to generate lactose but further extension was not achieved as the axial O4 of galactose is not tolerated at the donor position.

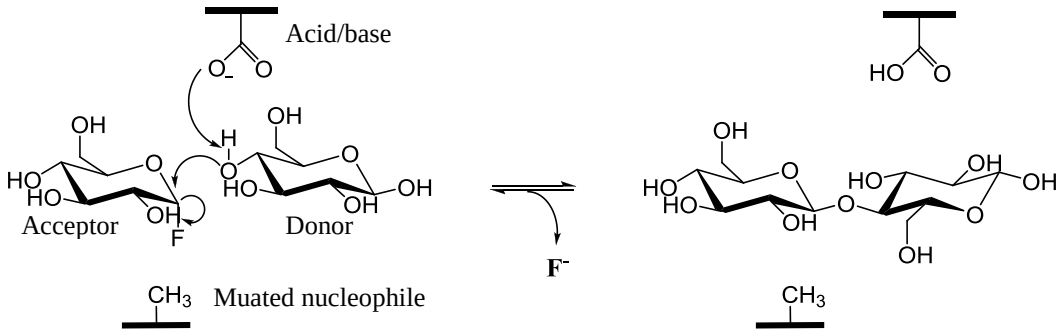


Figure 1.17 The glycosynthase mechanism

Since the first conversion of a retaining enzyme into a glycosynthase, many more retaining enzymes have been converted into synthetic enzymes, synthesising glycosidic bonds between mannosides, xylosides and galactosides (Shaikh et al. 2008). Glycosynthases have also been created that are able to transfer carbohydrates onto non-carbohydrate aglycones (Shaikh et al. 2008). Glycosynthases offer several advantages over using glycosyltransferases, nature's natural synthetic enzymes. They are easier to express and obtain by common molecular biology methods, while glycosyltransferases use expensive activated nucleotide phospho sugars donors. (Shaikh et al. 2008). Virtually all glycosynthases have been created from retaining glycoside glycoside hydrolases (MacKenzie et al. 1998; Shaikh et al. 2008) but there is one example of an inverting glycoside hydrolase being converted into a glycosynthase (Honda et al. 2006; Wada et al. 2008).

1.3.5 Alternative Catalytic Mechanisms

There are several alternative mechanisms employed by a minority of GHs to catalyse glycosidic bond cleavage, these are summarized as follows:

1.3.5.1 Neighbouring Group Participation

GHs in families 18, 20, 25, 56, 84 and 85 hydrolyse substrates containing an N-acetyl group at C2. Theses enzymes utilise a catalytic acid/base but have no catalytic nucleophile. Instead the carbonyl of the N-acetyl group at C2 acts as an intra molecular nucleophile and leads to the formation of an oxoazolinium ion intermediate (He et al.). Typically a second acidic amino acid is involved but acts as a stabilising residue for charge development during the transition state (van Aalten et al. 2001). Not all enzymes that are active on substrates substituted at C2 with N-acteyl operate via this mechanism, GH families 3 and 22 utilise a retaining mechanism (Rypniewski et al. 1993), while hexosaminidases of GH19 and GH23 operate through an inverting mechanism (Ubhayasekera et al. 2007).

1.3.5.2 Alternative Nucleophiles

GHs termed sialidases and trans-sialidases, located in GH33 and GH34, catalyse the hydrolysis and transglycosylation, respectively, of sialic acids. These enzymes have an acidic residue acting as an acid/base but utilise a Tyr as their catalytic nucleophile, which is activated by an adjacent acidic residue (Smith et al.

2002; Watts et al. 2003). A possible reason why a Tyr has evolved as the catalytic nucleophile is that sialic acids are negatively charged and the use of a negatively charged nucleophile would cause charge repulsion. Tyr is a neutral nucleophile but needs to be activated by a general base to increase its nucleophilicity (Smith et al. 2002; Watts et al. 2003).

*1.3.5.3 Exogenous Base*s

GHs found in GH1, termed myrosinases, catalyse the hydrolysis of anionic thioglycosides found in plants (Burmeister et al. 2000). GH1 enzymes have two Glu acting as the catalytic acid/base and catalytic nucleophile, but in myrosinases a Gln has replaced the position of the acid/base. This is likely due to a need to reduce electrocstatic repulsion with the anionic aglycone. The thiosulphate aglycone is a sufficiently good leaving group not to require protonation and thus the glycosylation step is unaffected. The deglycosylation step requires a base to activate a water molecule, to generate a hydroxyl ion to attack C1. This is performed by the co-enzyme L-ascorbate, which acts as an exogenous base and generates the required hydroxyl ion from water to complete the deglycosylation step (Burmeister et al. 2000).

1.3.5.4 NAD-dependent Hydrolysis

The GHs of families 4 and 109 operate via an unusual mechanism that requires an NAD cofactor (Varrot et al. 2005; Liu et al. 2007). The mechanism proceeds through an anionic transition state with elimination and redox reactions rather than an oxocarbenium ion like transition state (Varrot et al. 2005). For the substrate 6-phospho-β-glucose an oxidation of the O3 hydroxyl, by the enzyme bound NAD cofactor, increases the acidity of the O2 hydroxyl and, with the help of an enzymatic base, an E1cB elimination occurs. The anomeric centre is then attacked by a hydroxyl ion, generated from a water molecule, and the ketone group at C3 is reduced to yield the final sugar product. Although glycosidic bond cleavage occurs by elimination, water is consumed and the overall reaction is hydrolysis (Varrot et al. 2005).

1.3.5.5 Lytic Transglycosylation

This mechanism occurs in GH23 with enzymes that are active on substrates substituted at C2 with an N-acetyl group, and requires only a catalytic acid/base. In the first step of the reaction the catalytic acid/base donates a proton to the leaving group and the carbonyl of the acetyl group acts as an intra molecular nucleophile acttacking C1, forming an oxoazolinium intermediate. The acid/base then abstracts a proton from O6, which then attacks C1 relieving the oxoazolinium intermediate and generating a 1,6 anhydro product that is inverted to a $^{1}C_{4}$ chair. A second acidic residue acts to stabilise charge generation at the transition state (Thunnissen et al. 1995).

1.3.6 Subsite Nomenclature

A subsite of a GH constitutes all the amino acids which interact with a single sugar molecule. The subsites either side of the scissile glycosidic oxygen are numbered +1 toward the reducing end and -1 towards the non-reducing end of oligosaccharide. As the enzyme extends towards the reducing end of the substrate the subsites increase in positive value ie. +1, +2, +3, and in negative value, ie. -1, -2, -3, when the enzyme extends towards the non-reducing end of the substrate (Figure 1.18) (Davies et al. 1995). The -1 subsite houses the catalytic machinery, and it is the sugar molecule at this subsite that is distorted to the oxocarbenium ion like transition state.

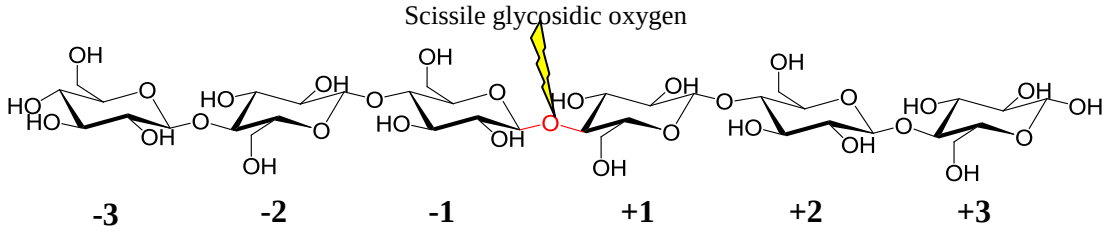


Figure 1.18 Schematic showing susbsite nomenclature of glycoside hydrolases
Scissile glycosidic oxygen is shown in red

1.3.7 Main Plant Cell Wall Glycoside Hydrolases

1.3.7.1 Cellulases

Cellulases work synergistically to degrade cellulose and its degradation products. These enzymes are currently found in GH families GH5, 6, 7, 8, 9, 12, 44, 45, 48 and 74 and utilise both inverting and retaining mechanisms (Henrissat 1991; Henrissat et al. 1997). Endo-acting cellulases cleave randomly along the cellulose

backbone producing oligosaccharides of varied sizes, while endo-processive cellulases make an initial random cleavage, then move processively along the glucan chain from the initial point of cleavage (Zverlov et al. 2005; Kitago et al. 2007). This class of cellulase displays an open cleft topology into which single glucan chains are accommodated. Endo-acting enzymes generally attack paracrystalline or amorphous regions of the cellulose microfibril, as the highly crystalline regions are difficult for the enzymes to access. Cellobiohydrolases are exo-acting, processive, enzymes and display the tunnel topography. They work from either the reducing or non-reducing end of individual cellulose chains releasing cellobiose as they progress (Guimaraes et al. 2002). The tunnel topology of cellobiohydrolases allows cellobiose to diffuse out of one end of the active site, while another cellobiose unit of the glucan chain is 'fed' in from the other end (Guimaraes et al. 2002). These enzymes are critical in cellulose degradation as their processive nature allows the enzymes to move into the highly crystalline regions of the cellulose microfibril (Birsan et al. 1998). The final class of cellulolytic enzymes are β-glucosidases, which are exo-acting enzymes that degrade small cellooligosaccharides to glucose (Sanz-Aparicio et al. 1998).

1.3.7.2 Xylanases

Xylanases are enzymes that cleave the β-1,4 xylose backbone of xylan, producing a range of xylooligosaccharides of varying size. These enzymes are mainly found in GH10 and GH11, although an example is also found in GH5. All xylanases classified to date operate via the retaining mechanism (Henrissat et al. 1997). GH5 and GH10 xylanases display an $(\alpha/\beta)_8$ fold typical of GH-A, while GH11 displays a β-jelly roll fold as expected of GH-C enzymes. GH10 and GH11 display an open cleft topology into which a single xylan chain sits (Harris et al. 1994; Vandermarliere et al. 2008). GH10 and GH11 xylanases act in synergy with α-L-arabinofuranosidases, 4-O-methyl glucuronidases, esterases and ferulic acid esterases to complete the degradation of xylan. There are some reports of GH10 and GH11 enzymes accommodating side chains of decorated xylan chains which, in the case of GH10 xylanases, enhance catalytic activity (Pell et al. 2004). The xylanase in GH5 has the unique feature of only hydrolysing the β-1,4 xylose backbone of xylan if it is substituted with α-1,2 glucuronic acid (Larson et al. 2003). GH8 also contains an example of a xylanase that is specially adapted to low temperatures (Van Petegem et al. 2003).

1.3.7.3 Mannanases

Mannanases catalyse the degradation of mannan, galactomannan, glucomannan and galactoglucomannan by hydrolysing β-1,4 mannose glycosidic linkages. Mannanases are currently found in GH families 5, 26 and 113 exhibit a retaining mechanism and display the $(\alpha/\beta)_8$ fold typical of GH-A (Henrissat et al. 1996). Endo-acting enzymes display an open cleft topology and cleave randomly along the backbone of mannan polysaccharides producing a range of mannose oligosaccharides of varying length (Hogg et al. 2001; Hogg et al. 2003). Exo-acting β-mannosidases display a pocket topography and hydrolyse mannooligosaccharides to monomeric mannose (Tailford et al. 2007). The hydrolysis and synthesis of mannosidic linkages is extremely difficult due to stereoelectronic clashes caused by the axial O2 of mannose, making mannanases a very impressive class of enzyme that is generating great interest. Several mannanases have been implicated in using the less common $B_{2,5}$ boat transition state, over the $^4H_3$ half chair, as unfavourable interactions with the axial O2 can be alleviated (Ducros et al. 2002; Tailford et al. 2008). Due to the heterogeneity of mannan polysaccharides, mannanases act in synergy with β-1,4-glucosidases, α-1,6-galactosidases and esterases to complete the breakdown of the polysaccharides.

1.3.8 Additional Plant Cell Wall Degrading Enzymes

GHs are the main, but not the only degradative enzymes that contribute to plant cell wall degradation. There are two other main classes of enzyme that contribute to this process; pectate lyases (PL) and carbohydrate esterases (CE).

1.3.8.1 Pectate Lyases

PLs are currently grouped into 21 different families based on amino acid sequence, displaying 9 different tertiary folds: Parallel β-helix, $(\alpha/\alpha)_7$ barrel, $(\alpha/\alpha)_6$ barrel, β-sandwich+β-helix, β-jelly roll, $(\alpha/\alpha)_6$ barrel+antiparallel β-sheet, $(\alpha/\alpha)_3$ barrel, β-propellor and triple strand β-helix (Coutinho 1999). The enzymes cleave pectic polysaccharides via a metal dependent, usually Ca^{2+}, β-elimination mechanism resulting in the formation of a double bond between C4 and C5 at the newly created non-reducing end (Charnock et al. 2002). The mechanism is dependent on the ability of the enzymes to abstract a proton from the C5 of galacturonic acid. The C5 proton is particularly acidic, compared to the other protons of the sugar ring, due to the

presence of the carboxylate group at C5, whose negative charge is stabilized through polar interactions with calcium (Charnock et al. 2002). The enzymatic base used to abstract the C5 proton is commonly an Arg although Lys is proposed for PL9 and some PL1 members (Coutinho 1999). Arg has a pKa of ~12.5, and Lys ~10.5, and would thus be protonated at physiological pH, this is clearly not the case in PLs due to pKa modulation of the base by its protein microenvironment, but it is curious why an acidic amino acid, which have pKas ~4.0, are not utilised instead.

1.3.8.2 Carbohydrate Esterases

CEs catalyse the deacetylation of O and N linked acetyl groups from polysaccharides found in the plant cell walls. They are currently organised into 16 families based on primary sequence identity. The 16 families display only two tertiary folds the β/α/β "serine protease" fold and the α/β-hydrolase fold (Coutinho 1999). Most families utilise an Asp-His-Ser catalytic triad and operate through a mechanism analogous to serine proteases (Stryer 2001). The Ser activated by His, which has its pKa modulated by Asp, performs a nucleophillic attack on the acetyl group, while the His donates a proton to the sugar leaving group; the reaction proceeds through a tetrahedral intermediate, which is stabilized by an oxyanion hole. This leads to the formation of an acyl-enzyme covalent intermediate. The His then deprotonates an incoming water molecule, which performs a nucleophillic attack on the acyl-enzyme covalent intermediate and completes the catalytic cycle. Hydrolysis of ester bonds by CE4 enzymes, however, involves a divalent metal ion and Asp, which act together to activate a water molecule that attacks the acetyl group (Montanier et al. 2009). Leaving group departure is facilitated by protonation by a His. The reaction transition state also proceeds through a tetrahedral intermediate generating an oxyanion.

1.4 Carbohydrate Binding Modules

GHs, and other plant cell wall degradative enzymes, are often appended to non-catalytic modules via a flexible linker. The non-catalytic modules that bind to polysaccharides are called carbohydrate binding modules (CBM). CBMs were first identified in 1986 when van Tilbeurgh showed that discrete, stable, modules, which displayed cellulose binding activity, could be proteolytically cleaved from cellulases. CBMs have since been shown to bind a wide variety of polysaccharides including

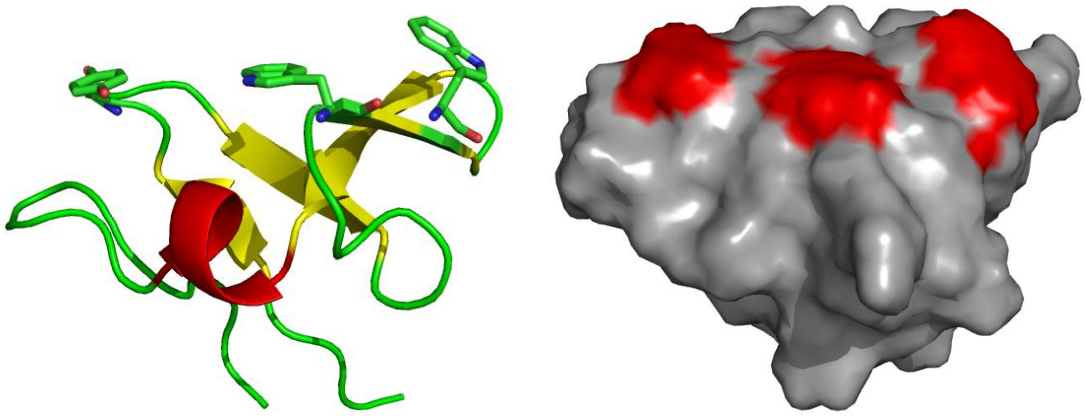
both crystalline and non-crystalline cellulose, agarose, xylan, β-1,3-β-1,4 mixed linkage glucans, mannan, laminarin and starch (Boraston et al. 2004).

1.4.1 Classification and nomenclature

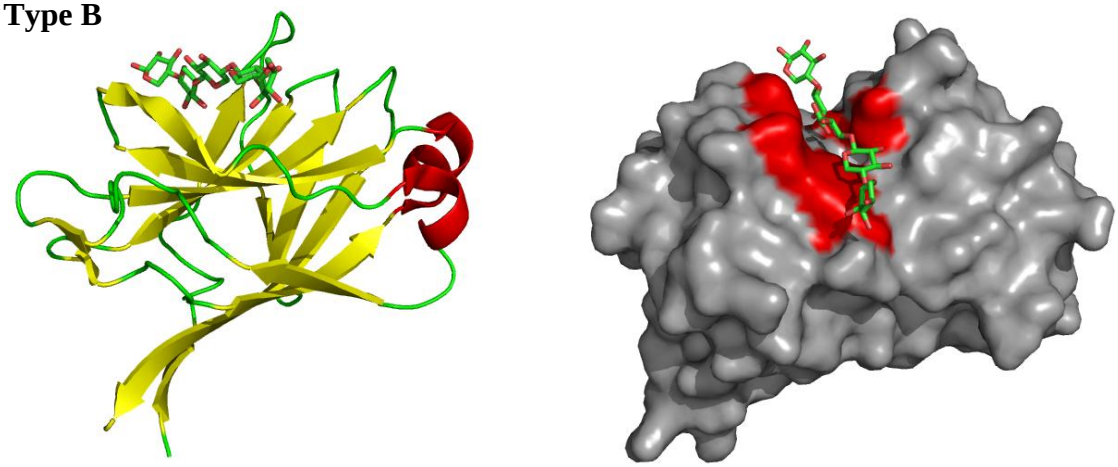
CBMs are currently classified in a similar manner to GHs in the CAZy database. There are currently 59 CBM families in the CAZy database, with 3D structures available for representatives of 45 of these families. There are seven general fold families displayed by CBMs of which the β-sandwich, consisting of two β-sheets that adopt a jelly-roll involving at least one structural divalent metal ion (fold family 1), is by far the most common

The seven fold families display three general topologies termed type A, type B and type C. Type A CBMs bind to insoluble substrates, mainly crystalline cellulose and chitin, and show little or no affinity for soluble oligosaccharides or polysaccharide chains (Tormo et al. 1996; Boraston et al. 2004). They present a flat planar face, which is usually comprised of three aromatic residues (Figure 1.19 Type A). This flat, planar, binding surface is complementary for the 110 hydrophobic face of crystalline cellulose (Boraston et al. 2004). Type B CBMs bind individual glycan chains in a single substrate binding cleft which spans the proteins surface (Figure 1.19 Type B). Type C CBMs are known as 'small sugar binders' and optimally bind mono, di and trisaccharides. They present a pocket into which the target ligand is accommodated (Figure 1.19 Type C) (Boraston et al. 2004).

Type A



Type B



Type C

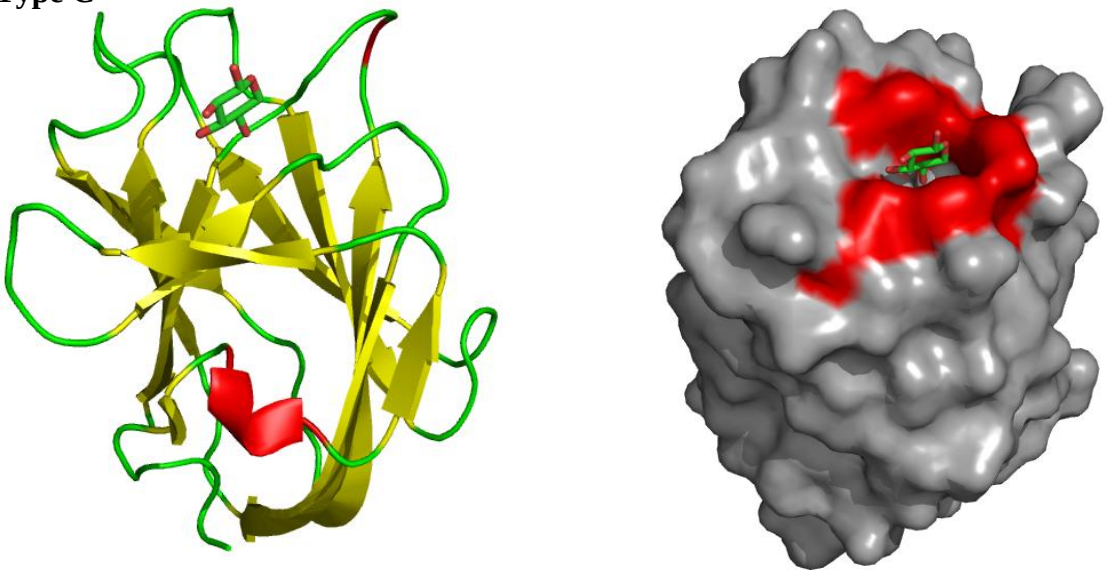


Figure 1.19 Topography of CBMs
Type A depicts *Cj*CBM10 as both a cartoon and surface representation (pdb 1E8R); Type B depicts *Cj*CBM15 in complex with xylotetraose as both a cartoon and surface representation (pdb 1GNY); Type C depicts *Cc*CBM6 in complex with xylose as both a cartoon and surface representation (pdb 2V4V). In cartoon representations α-helix = red, β-strand = yellow and loops = green. In surface representations residues that form the binding surface/cleft/pocket are shown in red. All ligands are coloured in with carbon green and oxygen red.

1.4.2 Mechanism of CBM action

CBMs increase the activity of GHs in three ways; a proximity effect, a targeting effect and by disrupting the structure of the substrate (Boraston et al. 2004). Through their ability to bind to sugars in the plant cell wall CBMs concentrate the appended enzymes on the surface of the substrate. The increase in the effective concentration of the enzyme in the vicinity of the substrate leads to an enahancement in the hydrolytic rate. It has been shown that removal of CBMs from GHs decreases the activity of the catalytic modules against insoluble but not soluble substrates (Hall et al. 1995).

Type B CBMs often mirror the activity of the catalytic module to which they are linked ie. cellulases, xylanases and mannanases are attached to CBMs which target cellulose, xylan and mannan, respectively. There is also evidence for the proposal that some CBMs target enzymes to areas susceptible to or already undergoing degradation. The family 9 CBM from xylanase 10A of *Thermotoga maritima* targets the reducing ends of polysaccharides (Notenboom et al. 2001), while some CBM35s bind to Δ4,5-anhydrogalaturonic acid but are appended to enzymes that deconstruct xylan (Montanier et al. 2009). Both ligands are products of plant cell wall degradation, especially Δ4,5-anhydrogalaturonic acid, which is only found when pectic polysaccharides are degraded by pectate lyases (Seyedarabi et al.). It should be noted that type A CBMs, which bind crystalline cellulose, are also appended to catalytic modules that do not degrade cellulose (Simpson et al. 2000). The rationale is that cellulose is the most abundant polysaccharide present in the plant cell wall, which is intimately associated with hemicellulose and pectin, so by staying in proximity to cellulose, the enzyme is likely to be presented at the surface of its target substrate.

Some CBMs seem to have the ability to disrupt polysaccharide structure. This was first demonstrated for CBM2a of Cel6A from *Cellulomonas fimi*. The CBM appeared to disrupt the structure of crystalline cellulose, which caused a very moderate increase in activity of the discrete cellulase (Gilkes et al. 1989). Stronger evidence comes from CBM33, which exist as discrete modules. These CBMs have been shown to disrupt the structure of chitin and significantly potentiate the activity of chitinases (which are not appended to the CBM33) against highly crystalline forms of the substrate (Vaaje-Kolstad et al. 2005).

1.5 Systems Hydrolysing Plant Cell Walls

The complexity displayed by the plant cell wall is mirrored in the systems employed by microbes to degrade it. A consortium of several types of cellulases, hemicellulases, pectate lyases and esterases are utilised. These enzymes can be thought of as a system and there are two general ways these systems can be organised. They can be organised into non-associating enzyme systems, usually observed in aerobic bacteria, or an associating enzyme system, which occurs in anaerobic bacteria and fungi.

1.5.1 Non-Associating Enzyme System

One of the best studied non-associating enzyme systems belongs to the bacterium *C. japonicus*, a gram negative, aerobic, saprophytic, soil dwelling organism with a pH optimum of 7-9 and a temperature optimum of 37 °C (Humphry et al. 2003). The organism was shown to be able to utilise crystalline cellulose, xylan, mannan, arabinan, galactan and pectin as its sole carbon and energy source (Hazlewood et al. 1998). In response to the complex carbohydrates in the plant cell wall the bacterium secretes extracellular GHs, which contain CBMs that target cellulose or other components of the plant cell wall. These modular GHs catalyse the initial degradation of the insoluble plant cell wall polysaccharides to shorter soluble oligosacchrides/polysaccharides. These soluble polymers are then hydrolysed by enzymes, appended on the outer membrane of the bacterium, to short chain oligosaccharides that are transported into the periplasim where they are further metabolised to mono-, di- or trisaccharides. These sugars are then transported across the inner membrane into the cytoplasm where they are utilized as an energy and carbon source. Figure 1.20 details an example of the *C. japonicus* xylan degrading system. It has been demonstrated that the bacterium expresses five xylanases in response to xylan-containing polysaccharides, as well as accessory enzymes to remove side chains (Beylot et al. 2001; Emami et al. 2002; Emami et al. 2009). The xylanases *Cj*Xyn10A, *Cj*Xyn11A and *Cj*Xyn11B all contain cellulose binding CBMs and are secreted to attack and solubilise the insoluble xylan associated with cellulose microfibrils. There is temporal regulation with *Cj*Xyn11A expressed prior to *Cj*Xyn11B. *Cj*Xyn11A also contains an esterase enabling the enzyme to deconstruct acetylated xylan. *Cj*Xyn11B is thus likely expressed later on in the degradative

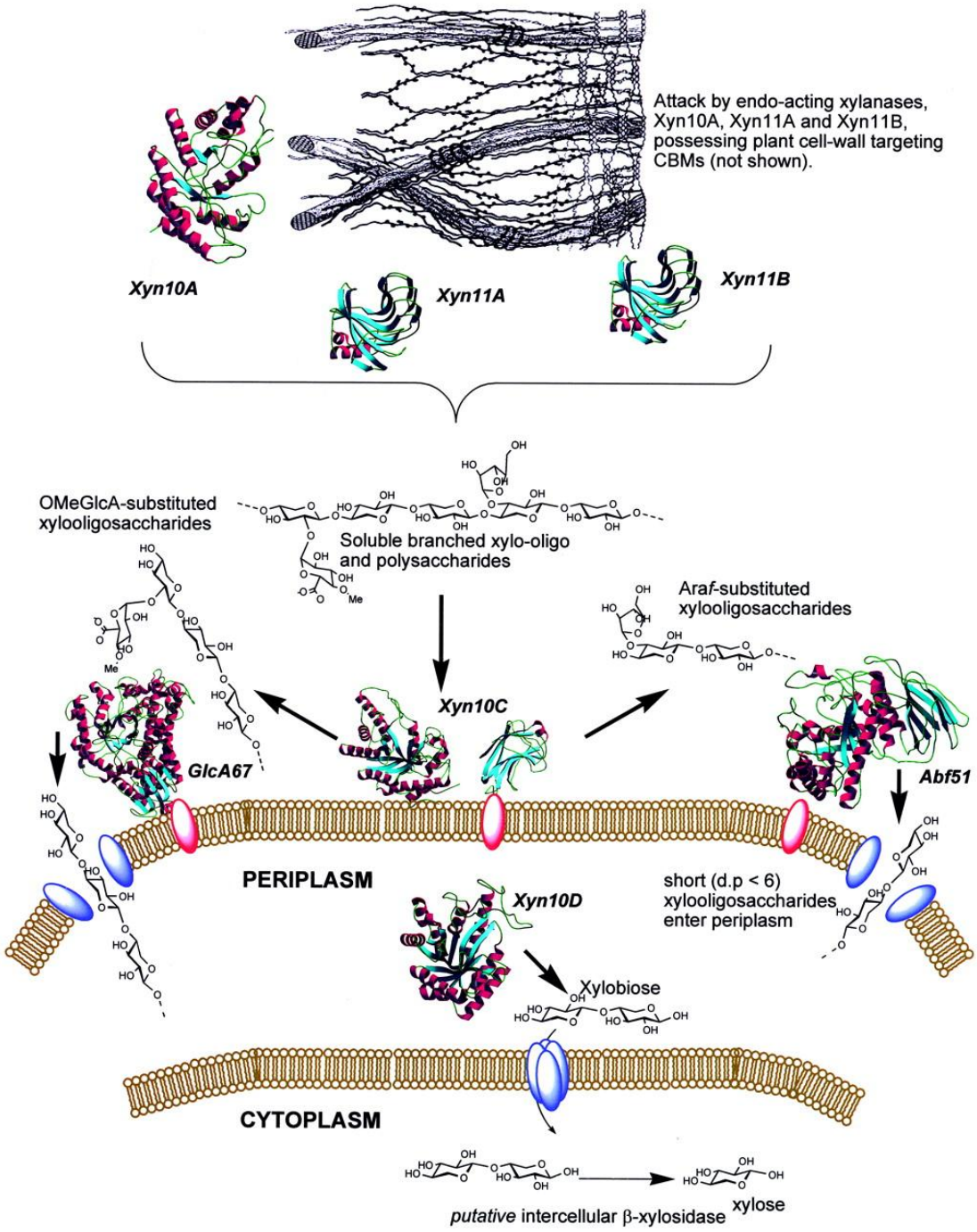


Figure 1.20 Schematic of the Non-associating enzymes system in *C. japonicus*
The structures shown for *Cj*Xyn11A and *Cj*Xyn11B are representative bacterial GH11 enzymes from *Bacillus agaradhaerens* and *Nonomuraea flexuosa*, respectively, the structure shown for *Cj*Abf51A is from *Geobacillus stearothermophilus*, and the structure shown for *Cj*Xyn10D is of the homologous (93 % similar) enzyme *Cm*Xyn10B from *C. mixtus*. The enzymes *Cj*Abf51A, *Cj*GlcA67A, and *Cj*Xyn10C are all attached to the outer membrane of the bacterium, whereas *Cj*Xyn10A, *Cj*Xyn11A, and *Cj*Xyn11B are secreted into the culture medium. The sugar side-chains released by the α-glucuronidase and arabinofuranosidase, respectively, are likely to be transported into the cytoplasm of *C. japonicus* (Pell et al. 2004).

process when greater areas of undecorated xylan become available (Emami et al. 2002). *Cj*Xyn10C contains a xylan binding CBM and is appended to the outer membrane of the bacterium. This enzyme likely hydrolyses the large soluble products from the initial xylan degradation. The membrane bound *Cj*Abf51 and *Cj*GlcA67 and CE2s remove arabinofuranose, 4-O-methyl glucuronic acid and acetyl side chains, respectively, generating small undecorated xylose oligosaccharides, which are taken up into the periplasm (Beylot et al. 2001). *Cj*GlcA67 only removes 4-O-methyl glucuronic acid from the non-reducing end of xylooligosaccharides and shows no activity on polymeric xylan (Nagy et al. 2003). *Cj*Xyn10D, which is thought to reside in the periplasm, hydrolyses small xylooligosaccharides to xylobiose, which is taken up into the cell and likely broken down to xylose by a putative intracellular xylosidase. This is an elegant paradigm for non-associating enzyme systems. Indeed the hierarchical cellular location of hydrolytic enzymes is a general feature of hemicellulosic degradation by *C. japonicus*, which is particularly well illustrated by the mannan degrading apparatus, discussed in detail in Chapter 3 of this thesis.

1.5.2 Associating Enzyme Systems

The best studied, and first discovered, associating enzyme system belongs to the bacterium *Clostridium thermocellum*. *C. thermocellum* is a gram positive, thermophilic anaerobe and is one of the most efficient cellulose degrading organisms identified to date (Fontes et al.). The organism assembles its plant cell wall catalytic machinery into a megadalton complex, termed the cellulosome, which is attached to the surface of the bacterium (Figure 1.21). The scaffold of the cellulosome, called CipA, is comprised of 140 residue tandemly arranged modules called type I cohesins. The catalytic machinery comprising GHs, PLs, esterases and CBMs are recruited onto CipA via type I dockerin domains, typically at their C-terminus, which bind tightly to the CipA type I cohesins. The cellulosome is bound to the surface of the bacterium through a membrane protein that contains a type II cohesin domain that binds to the type II dockerin present at the C-terminus of CipA (Fontes et al.).

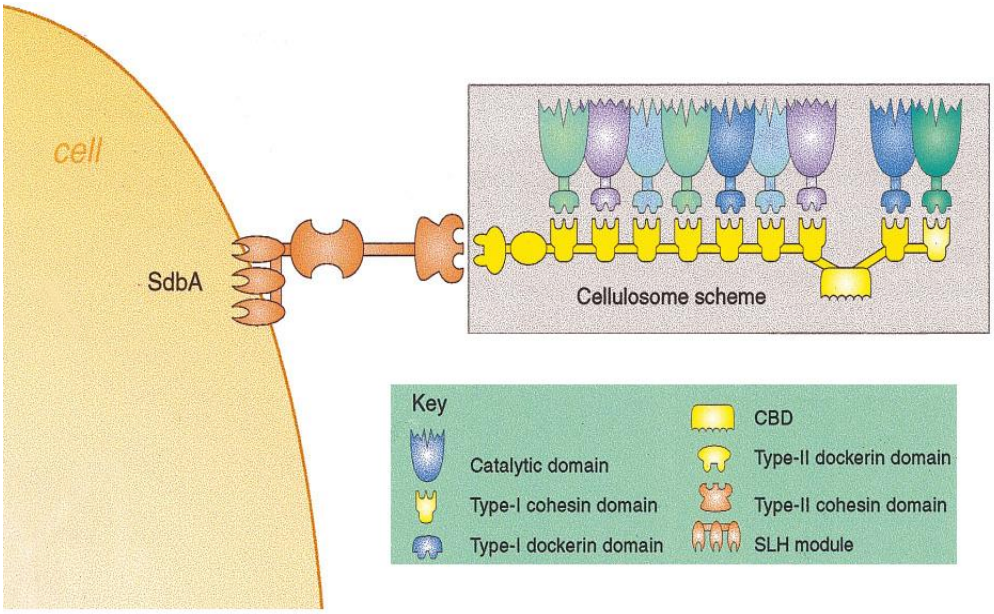


Figure 1.21 Schematic representation of cellulosome organization and attachment to the *C. thermocellum* cell surface.
The cellulosome and its associated anchoring proteins all comprise modular components. The scaffolding protein of *C. thermocellum*, shown in yellow, is composed primarily of nine copies of cohesin module, a Family-IIIa CBD and a type-II dockerin domain (Bayer et al. 1998).

Objectives

The focus of the research in this thesis was identifying and characterising novel glycoside hydrolases. Although there is a great breadth of knowledge that has been unearthed in the past decades, there is still much to be discovered. Some GH families have very narrow specificities with all members active on the same substrate. GH26 is one of these families but rare deviations from the expected activity allow a further insight into mechanism and mode of action of the family and glycoside hydrolases in general (Chapter 3).

There are currently a 118 families catalogued on the CAZy database and this is continually expanding. Work presented in thesis will contribute to this, allowing the formation of GH119. Discovery of new families is important in uncovering all the tools utilised by nature to deconstruct the plant cell wall (Chapter 4).

In some organisms there are expansions of specific GH families. This is logical when the family has diverse specificities and functions. When there is an expansion in a GH family that is not well characterised this provides an opportunity to explore activities new to the family and possibly new to glycoside hydrolases in general (Chapter 5). By gaining knowledge into why a particular GH family has been expanded, to meets the needs of the organism, a valuable insight can be gained into

plant cell wall degradation in nature. The specific objectives of this study are listed below:

Structural and functional characterisation of novel GH26 members.

Structural and functional characterisation of a novel cellulase that founds a new GH family.

Utilise structural and functional studies to elucidate the gene expansion of GH43s in *C. japonicus*.

Chapter 2

Materials and Methods

2.1 Molecular Biology

All experiments and solutions were made using high quality MQ H_2O (18.2Ω) produced by a Millipore Milli-RO 10 Plus Water Purification System.

2.1.1 Bacterial Strains and Plasmids

The *Escherichia coli* strains used in this study are listed in Table 2.1.

E .coli strain	**Description**	**Use**	**Reference**
BL21(DE3)	F$^-$ *dcm ompT hsdS*(r_B- m_B-) gal l(DE3)	Protein Expression	Studier *et al*., 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)	Seleno-methionie Protein Expression	Novagen
XL1-blue	*recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F´*proAB lacI^{q}Z* Δ M15 Tn10(Tet^r)]	DNA Replication	Bullock *et al*., 1987
One Shot™TOP10	F´*mcrA* Δ*(mrrCB-hsdRMS-mrr)* Δ 80*lacZ* Δ M15 Δ *lacX74 deoR recA1 araD139* Δ (*ara-eu*)7697 *galU galK rspL endA1 nupG*	DNA ligation	Invitrogen

Table 2.1 Bacterial strains used in this study

2.1.2 Media and growth conditions for propagation of bacteria

E. coli strains (unless stated) were grown at 37 °C in Luria-Bertani broth (LB) medium (1 % (w/v) Bacto©Tryptone, 1 % (w/v) NaCl and 0.5 % (w/v) yeast extract, pH 7.4). Conical baffle flasks two times the volume of the media were used to provide sufficient aeration by rotary shaking (180 rpm). Agar plates were produced by the addition of 2 % (w/v) Bacteriological agar N^{o}1 (Oxoid) to the appropriate volume of LB medium. This suspension was then autoclaved, allowed to cool and poured into

Petri dishes. If required an appropriate antibiotic was added prior to pouring. These agar plates could then be stored at 4°C for a maximum of 4 weeks.

2.1.3 Plasmids used for protein expression

Plasmid	Size (bp)	Phenotype/Gennotype	Reference
pET20b	3716	Amp^r, T7, pelB	Novagen
pET22b	5482	Amp^r, T7, *lac*, *lac*I^q	Novagen
pET28b	5369	Kan^r, T7, *lac*, *lac*I^q	Novagen
pET32b	5900	Amp^r, T7, *lac*, *lac*I^q	Novagen

Table 2.2 Vector plasmids used in this study

The plasmids multiple cloning sites (MCS) are shown in Appendix A.

2.1.4 Selective Media

For antibiotic selections appropriate dilutions of a stock solution (Table 2.3) were added to the media. Isopropylthio-β-D-galactoside (IPTG) was added to media to induce recombinant protein expression in strains carrying *lacI*^{q} on the plasmid and/or in the genome for induction of recombinant genes controlled by *lacO*. IPTG was made at a stock concentration of 1 M; a final concentration of 1 mM was added to BL21 and B834 strains whereas 0.1 mM was used for Tuner strains.

Antibiotic	Stock concentration	Working concentration	Storage
Ampicillin	50 mg/ml	50 µg/ml	4 °C for less than 4 days
Kanamycin	50 mg/ml	50 µg/ml	4 °C for less than 4 days

Table 2.3 Antibiotics used in this study

2.1.5 Sterilisation

All solutions and media were sterilised by autoclaving using either an Astell Hearson 2000 Series Autoclave or a Prestige© Medical Series 2100 Clinical Autoclave at 121 ^oC, 32 lb / $inch^{-2}$ for 20 min. Where necessary, solutions were sterilized by filter sterilization through a 0.22 µm sterile filter disc (Stupor© Acrodisc© 3.2 Gelman Sciences).

2.1.6 Chemicals, enzymes and media

The chemicals, enzymes and media used in this study, which are not specifically listed in the text, were either from Sigma Chemical Company, Megazyme or BDH Laboratories Ltd.

2.1.7 Centrifugation

Bacterial cultures ranging from 100-1000 ml were harvested by centrifugation at 5000 x g, in 500 ml Nalgene bottles, in a Beckman J2-21 centrifuge with the use of a JA-10 rotor. Bacterial cultures of 5 ml were harvested by centrifugation at 5000 x g, in 25 ml sterilins, in a MSE Mistral 3000i bench top centrifuge with a fixed angle rotor. Eppendorf tubes were centrifuged up to 13000 x g using a Heraeus Instruments Biofuge *pico* benchtop centrifuge.

2.1.8 Plating bacteria

Bacterial cell suspensions of 100-200 µl were dispensed onto an appropriate agar plate, in close proximity to a bunsen burner flame. A glass spreader immersed in 100 % ethanol was then passed through the bunsen burner flame and the ethanol allowed to burn off. The glass spreader was then pressed against the agar plate to ensure it was cool and then used to spread the bacterial cells uniformly across the plate.

2.1.9 Chemically competent *E. coli*

E. coli strains were made chemically competent for the uptake of plasmid DNA following the adapted procedure outlined by Cohen *et al* (Cohen et al 1972). A one ml aliquot of an overnight LB *E. coli* culture was inoculated into a one litre, unbaffled flask, containing 100 ml of LB. The 100 ml *E. coli* culture was then grown at 37 °C, 180 rpm, until it had reached an optical density of 0.4 (A_{600}=0.4). The *E. coli* culture was then placed on ice for ten min after which it was centrifuged at 5000 x g. The cell pellet was then gently re-suspended in 8 ml of ice cold $MgCl_2$. The cells were then, spun down again, as mentioned previously, and gently re-suspended in 4 ml of ice cold $CaCl_2$. The re-suspended cells were then left on ice for two hours, after which the *E. coli* had become competent for the uptake of plasmid DNA. The competent *E. coli* cells were then stored as 100 µl aliquots, in Eppendorf tubes, with 25 % (v/v) glycerol at -80 ºC

2.1.10 Transformation of chemically competent *E.coli*

A 100 µl aliquot of chemically competent *E. coli* cells was typically mixed with 1 µl (10-1000 ng) of plasmid DNA. The cells were then left on ice for 30 min, after which they were heat shocked at 42 °C for two min and then returned to ice for a further two min. The transformed cells were then plated on agar plates, containing an appropriate antibiotic, inverted and placed in a 37 °C oven for 16 h.

For the transformation of QuickChange™ products or after ligation cells were mixed with 500 µl of LB after the final two min on ice and incubated at 37 °C, 180 rpm, for one h. The cells were then centrifuged at 5000 x g, the supernatant was removed and the cells were re-suspended in 100 µl of LB and plated onto agar plates as described above.

2.1.11 Rapid small scale (±20µg) preparation of plasmid DNA

For the propagation of recombinant plasmid DNA 5 ml cultures of XL1-blue cells containing the appropriate plasmid were grown overnight from single colonies. The cultures were then centrifuged at 5000 x g and the supernatant removed. Plasmid purification was carried using a QIAGEN® QIAspin Prep kit as per the manufacturer's instructions.

2.1.12 Midi scale (±500µg) preparation of plasmid DNA

For the propagation of vector plasmid DNA 100 ml cultures of XL1-Blue cells containing the appropriate plasmid (normally a stock vector) were grown to an optical density of 0.6 (A_{600}=0.6). The cultures were then centrifuged at 5000 x g and the supernatant removed. Plasmid purification was then performed using a QIAGEN ® Plasmid Plus Midi Kit as per the manufacturer's instructions.

2.1.13 Restriction digest of DNA

The required amount of DNA, dissolved in EB buffer (10 mM Tris-HCl pH 8.5), was mixed with 10 x concentrated reaction buffer appropriate for the endonucleases used. Between one and two units (one unit is the amount of enzyme required to cleave 1 µg of DNA during 1 h at 37 °C) of endonuclease was added, and the reaction incubated at 37 °C for one h.

2.1.14 Measuring DNA concentration

DNA concentration was either estimated by visualisation on an agarose gel or determined by spectroscopy. For visualisation of DNA concentration a small amount was electrophoresed on an agarose gel and compared in intensity against a ladder of known standards. For spectroscopy DNA was appropriately diluted, placed in quartz cuvettes and an absorption reading at A_{260} was taken using a Pharmacia Ultrospec 4000 spectrophotometer. At A_{260} 50 µg/ml of double stranded DNA gives an absorption of 1.0; single stranded DNA and RNA gives an absorption of 1.0 at 33 µg/ml.

2.1.15 Agarose gel electrophoresis

The separation and determination of sizes of linear DNA molecules was carried out by electrophoresis through submerged horizontal gels. Only mini gels for rapid analysis were used in this work and proved to be sufficient. Gels were prepared by dissolving 400 mg of agarose in 50 ml (w/v) of 1 x TBE buffer (89 mM Tris Base, 89 mM Boric acid and 2 mM EDTA). The suspension was then boiled until the agarose had dissolved. The gel was then allowed to cool and 0.5 µg/ml of ethidium bromide was added to allow visualisation of electrophoresed DNA. The gel was then poured into the mini gel systems (Applied Biosystems). Once the gel had set it was covered in 50 ml of 1 x TBE and ran at 120 volts for one h (LKB Bromma 2197 Power Supply). Samples were prepared by mixing 5 µl of sample with 1 µl of 6 x loading dye. Hyperladder one from Bioline was rang alongside the samples as a size and quantity standard.

2.1.16 Visualisation of DNA and photography of agarose gels

After electrophoresis DNA was visualised using a documentation system (Bio-Rad Gel Doc 1000 using Molecular Analyst™/PC Windows Software). Photographs were produced using a Mitsubishi Video Copy Processor (Model P68B attached to the gel documentation system) with Mitsubishi thermal paper.

2.1.17 Determination of DNA fragment size by agarose gel electrophoresis

The migration of linear double stranded DNA within an agarose gel is inversely proportional to the $\log_{10}$ of its size. Therefore, by running a standard of

known sizes alongside the electrophoresed samples one can estimate the size of the sample.

2.1.18 Purification of DNA samples

2.1.18.1 Purification of vector DNA

Vector DNA that had been digested with endonucleases was purified by agarose gel electrophoresis using high purity Seachem Gold™ Agarose dissolved in 1 x TBE buffer. The vector DNA was excised from the gel using a scalpel blade and purification from the gel carried out using a QIAquick Gel extraction kit (Qiagen) as per the manufacturer's instructions.

2.1.18.2 Purification of inserts and PCR products

Insert DNA digested with endonucleases and PCR products were purified using QIAquick PCR Purification Kit (Qiagen) as per manufacturer's instructions.

2.1.19 Ligation of insert and vector DNA

Ligations were carried out by mixing digested insert and vector DNA in a molar ratio of approximately 3:1 in ligation reactions. Where problems were encountered this ratio was increased 2-3 fold. Ligation reactions were made up in 1.5 ml Eppendorf tubes and are listed in Table 2.4.

Amounts	Components
100 ng	Vector DNA
300 ng	Insert DNA
4 µl	5 × Ligation Buffer (250 mM Tris/HCl (pH 7.6), 50 mM $MgCl_2$, 5 mM ATP, 5 mM DTT, 25 % (w/v) polyethylene glycol-8000)
Up to13 µl	Sterile water
1 µl	Invitrogen®T4 DNA Ligase (4 U/µl)
20 µl	Total volume

Table 2.4 Components of ligation reactions

The ligation reaction was then left for one h at room temperature before being used to transform chemically competent One Shot™TOP10 cells.

2.1.20 Phosphorylation of DNA

Phosphorylation, where needed, was carried out on oligonulceotide primers using T4 polynucleotide kinase (Fermentas). Chemically synthesised oligonucletide primers lack a 5' phosphate and therefore need to be phosphorlylated when the PCR products are to be ligated. This does not apply to primers which are digested by endonucleases as digestion releases 5' phosphate groups. A typical phosphoryltion reaction is listed in Table 2.5.

Amount	Component
2 µl	10 X Buffer (700 mM Tris-HCl pH 7.6, 10 mM $MgCl_2$, 5 mM DTT)
1 µl	100 µM Oligonucleotide primer
2 µl	10mM ATP
1 µl	T4 polynucleotide kinase (2 units/ µl)
Up to 14 µl	Sterile water
20 µl	Total volume

Table 2.5 Components of a phosphorylation reaction

2.1.21 Polymerase chain reaction

The polymerase chain reaction (PCR) developed by Mullis and Faloona (Mullis et al 1987) was used throughout this study to amplify genes from genomic DNA and to mutate and alter recombinant genes already cloned into vector plasmids. All primers used in this study were produced using automated MWG Oligo-2000 synthesizer technology and purified by HPSF® (MWG-Biotech AG, Germany). All primers were analysed by MALDI-TOF prior to despatch from the company. All PCR reactions were made up in a 0.25 ml thin walled Eppendorf tube.

2.1.21.1 Standard PCR

This type of PCR was used to amplify genes from genomic DNA. Two primers, one complimentary to the 5' end of the forward strand and the other complimentary to the 5' end of the reverse strand, which flanked either side of the gene, were used. A thermostable DNA polymerase, in the presence of dNTPs, is able to extend the primers using the complementary DNA strand as a template. Primers were approximately 20-25 bases in length with a ~50 % GC content, sequence permitting, and where possible both started and ended with G's and C's. Primers were

designed so that they had a melting temperatures (Tm) of >50°C and within 2 °C of each other within the primer pair. The following formula was used to calculate primer Tm.

$$Tm = 81.5 + 0.41(\%GC) - 675/N$$

N is the primer length in bases and % GC is a whole number. When required restriction sites were added to the 5' end of primers. When doing so an additional six nucleotides were also added upstream of the restriction site to allow efficient cutting of the PCR product by endonucleases. A typical PCR reaction is shown in Table 2.6.

Amounts	Components
5 µl	10 × Buffer for KOD DNA Polymerase (10 × = 1.2 M Tris-HCl, 100 mM KCl, 60 mM $(NH_4)_2SO_4$, 1 % Triton X-100, 0.01% BSA, pH 8.0)
2 µl	25 mM $MgCl_2$
5 µl	dNTPs (2 mM each)
5 µl	3 µM oligonucleotide primer #1
5 µl	3 µM oligonucleotide primer #2
≈ 100 ng	Genomic DNA
up to 28 µl	PCR grade water
1 µl	Novagen®KOD DNA Polymerase (2.5 U/µl)
50 µl	Total volume

Table 2.6 Showing composition of a standard PCR reaction

The PCR machine used during this work was PX2 (Hybaid). The standard PCR cycle used is shown in Table 2.7.

No of cycles	Step temperature and time
1 cycle	95 °C for 1 min
35 cycles	95 °C for 1 min Annealing temperature for 1 min 72 °C for 1 min/kb fragment size
1 cycle	72 °C for 5 min
1 cycle	Held at 4°C

Table 2.7 Showing cycles parameters for a standard PCR

2.1.21.2 Quikchange™ site-directed mutagenesis

Mutagenesis of amino acids was carried out using QuikChange™ Site-Directed Mutagenesis Kit from Stratagene, following the manufacturer's instructions. The primers for QuikChange™ contained a codon in which two nucleotides differed from the codon of the amino acid to be mutated; this was flanked by 15 nucleotides on either side that where complementary to the DNA sequence. The Tm for primers designed for QuikChange™ was calculated using the formula below:

$$Tm = 81.5 + 0.41(\%\ GC) - 675/N - \%\ \text{mismatch}$$

A typical QuickChange™ reaction is listed in Table 2.8.

Amount	Component
10 ng	DNA template
5 µl	10 × Reaction Buffer (100 mM KCl, 100 mM$(NH_4)_2SO_4$, 200 mM Tris/HCl (pH 8.8), 20 mM $MgSO_4$ 1 % Triton® X-100, 1 mg/ml nuclease-free BSA)
5 µl	dNTPs mix 2mM of each
5 µl	3µM oligonucleotide primer #1
5 µl	3µM oligonucleotide primer #2
Up to 28 µl	PCR grade water
1 µl	*PfuTurbo*® DNA polymerase (2.5 U/ µl)
50 µl	Total Volume

Table 2.8 showing the composition of a Quikchange reaction

The PCR cycle used for QuickChange™ mutagenesis is listed in Table 2.9

No of cycles	Step temperature and time
1 cycle	95 °C for 30 s
16 cycles	95 °C for 30 s 55 °C for 1 min 68 °C for 1 min/ kb of plasmid length
1 cycle	68 °C for 5 min
1 cycle	Held at 4 °C

Table 2.9 showing cycle settings for Quikchange

After the thermocycling procedure had finished 1 µl (10 units) of *Dpn*I was added to the reaction, to digest any methylated template DNA but not the unmethylated amplified DNA, and incubated at 37 °C for one h. The DNA was then transformed into One Shot™ TOP10 competent cells.

2.2.21.3 Inverse PCR (IPCR)

Inverse PCR (IPCR) was used to delete amino acids. The method uses two primers, designed so that the 5' ends are back to back, and are separated by the region of DNA coding for the amino acids to be removed (Figure 2.1). The PCR reaction and cycle parameters used for IPCR are the same as those found in Tables 2.6 and 2.7 under standard PCR. After the thermocycling procedure had finished 1 µl (10 units) of *Dpn*I was added to the reaction to remove parental DNA as detailed above. Primers were phosphorylated (as described Section 2.1.20) prior to performing IPCR as the product is blunt ended and needs to be ligated (as described Section 2.1.19). The DNA was then transformed into One Shot™ TOP10 competent cells.

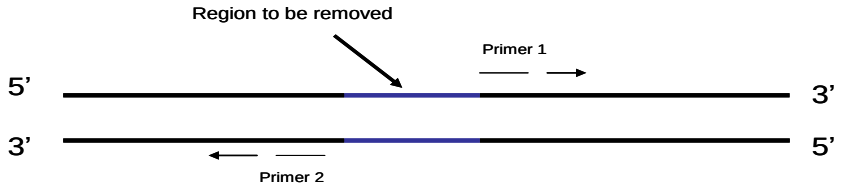


Figure 2.1 showing the principal of IPCR

2.2.22 Automated DNA sequencing

DNA sequencing employed the Value Read service from MWG Biotech AG, Ebersberg, Munich, Germany using ABI 3700 sequencers and Big dye technology (Applied Biosystems). Plasmid DNA was dried by vacuum lyophilisation at room temperature in 1.5 ml tubes. Plasmids were then sequenced with appropriate primers (Table 2.10)

Plasmid type	Forward primer	Reverse primer
pET plasmids	T7 TAATACGACTCACTATAGGG	T7 term CTAGTTATTGCTCAGCGGT

Table 2.10 Primers used in sequencing reactions

An example of the results obtained from a typical DNA sequencing reaction, using the T7 forward primer, are shown in Figure 2.2.

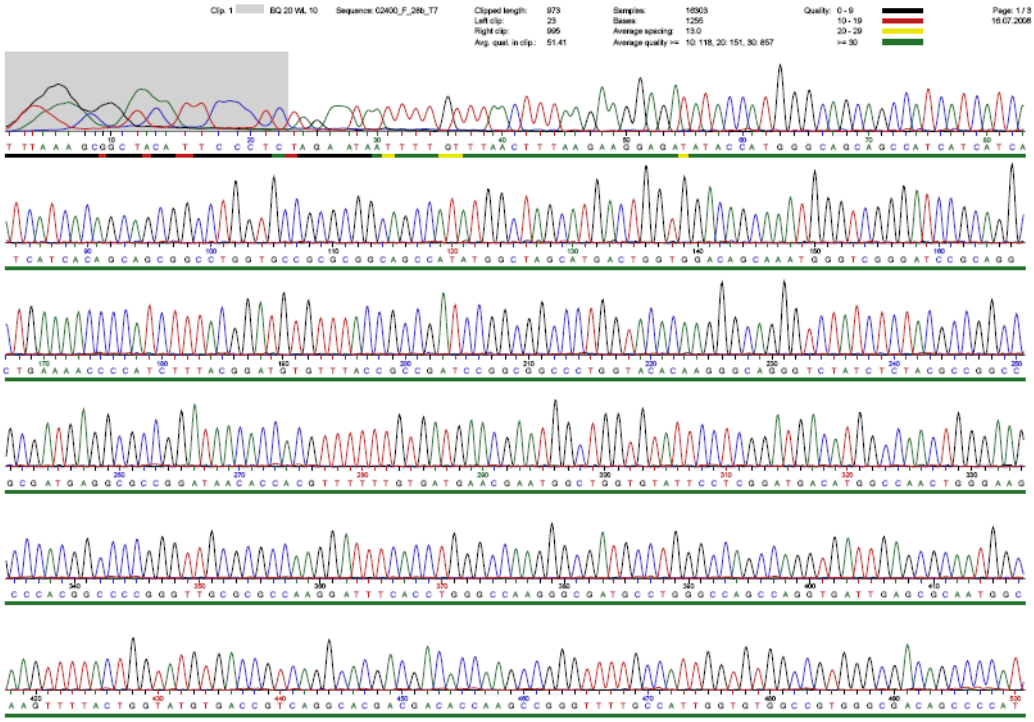


Figure 2.2 chromatogram of the raw data produced from a sequencing reaction using T7 forward primer

2.2.23 Computerised sequence alignments and analysis

DNA and protein sequence alignments, translations, reverse complementation and BLAST searches were carried out using the online tools of the ExPASY proteomics server (http://expasy.org/).

2.2.24 Over Expression of recombinant proteins in *E.coli*

Unless specified over expression of recombinant protein was as follows. A streak of *E. coli* colonies, harbouring an appropriate plasmid, were inoculated into 5 ml of LB containing the appropriate antibiotic. This 5 ml culture was grown over night for 16 h at 37 ºC and 180 rpm. The following morning the 5 ml culture was used to inoculate a 2 litre baffled flask, containing 1 litre of LB with an appropriate antibiotic. This culture was grown at 37 ºC and 180 rpm and allowed to reach an optical density of 0.6 (A_{600nm} = 0.6). For induction of over expressed protein at 37 °C IPTG was then added to a final concentration of 1 or 0.1 mM depending on the strain used. Cultures were then left 4-6 h before the cells were harvested by centrifugation

(Section 2.1.7). For protein to be induced at 16 °C, once they had reached an optical density of 0.6 (A_{600nm} = 0.6), they were cooled to 16 °C after which IPTG was added to a final concentration of 1 or 0.1 mM depending on the strain used. The cultures were then incubated at 16 °C and 180 rpm for 16 h. The following morning the cells were harvested by centrifugation (Section 2.1.7). The cell pellets were then re-suspended in Talon buffer (50mM Tris-HCl, 300mM NaCl, pH 8.0) using 10 ml per litre of original culture. The cell suspension was either stored at -20°C or cell free extract was prepared for protein purification (Section 2.2.25).

2.2.25 Preparation of cell free extracts and re-suspension of insoluble pellets from *E. coli*

Cell suspensions, as prepared in Section 2.2.24, were sonicated for two min on ice, using a B. Braun Labsonic U set at low intensity ≈45 watts and 0.5 s cycling. The lysed cell suspension was then transferred to a 50 ml centrifuge tube (Nalgene) and centrifuged at 30000 x g for 30 min at 4 °C. The supernatant comprising the cell free extract was removed and the cell pellet was re-suspended in 10 ml of Talon buffer containing 8 M urea.

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

Analysis of protein in terms of size, purity and quantity was carried out using the method first described by Lamelli (Lamelli 1972). Gels can contain different amounts of acrylamide, typically 10-15 %, depending on the size of the protein to be analysed. In this body of work 12.5 % gels were used throughout. SDS-PAGE was carried out using the AE-6450 apparatus supplied by ATTO Corporation (Genetics Research Instruments). A pair of glass plates (12 cm × 10 cm and 12 cm × 9 cm) were washed with detergent, rinsed with distilled water and cleaned with ethanol. The plates were then clipped together with a rubber gasket in between. The resolving gel (Table 2.11) was poured into the plates, covered with water, and allowed to polymerise.

Volume	Component
9.4 ml	0.75 M Tris/HCl buffer, pH 8.8 with 0.2 % SDS
5.8 ml	40 % (w/v) Acrylogel, 3 % (w/v) acrylamide, bisacrylamide
3.5 ml	Distilled water
90 µl	10 % (w/v) Ammonium Persulphate
30 µl	TEMED

Table 2.11 Composition of SDS-PAGE resolving gel

The water was then removed. A stacking gel was prepared (Table 2.12) and poured on top of the resolving gel. A comb was then inserted in to the stacking gel and the gel left to polymerise.

Volume	Component
3.75 ml	0.25 M Tris/HCl buffer, pH 6.8 with 0.2 % SDS
0.75 ml	40 % (w/v) Acrylogel, 3 % (w/v) acrylamide, bisacrylamide
3.0 ml	Distilled water
60 µl	10 % (w/v) Ammonium Persulphate
20 µl	TEMED

Table 2.12 Composition of SDS-PAGE stacking gel

Once the gel had polymerised the rubber gasket and comb were removed. Protein samples were prepared by mixing 15 µl of sample with 5 µl of loading dye (Table 2.13). The mixture was then boiled for 2 min and 10 µl loaded into the wells.

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

Table 2.13 Loading dye compostion

The gels were then placed in the tank, which was filled with running buffer (Table 2.14), and 10 µl of sample loaded into the wells. A standard, with proteins of known M_r, (Appendix D) was also loaded to correlate with proteins in the sample.

Amount	Component
32 mM	250 mM Tris/190mM Glycine, pH 8.3
0.1%	SDS

Table 2.14 Showing running buffer composition

After electrophoresis protein in the gels was visualised by staining with Coomassie blue stain (20% (v/v), methanol, 10% (v/v), glacial acetic acid and 0.04% Coomassie Brilliant Blue R) for twenty min with gentle shaking. The gel was then soaked in destain (20 % (v/v) methanol, 10 % (v/v) glacial acetic acid) until the protein bands were visual by eye. Gels were documented by photographing with a Canon PowerShoot A75 digital camera.

2.2.27 Protein Purification

2.2.27.1 Immobilised metal affinity chromatography (IMAC)

IMAC was used to purify recombinant proteins that contained a stretch of six histidines (His_6) at their N or C terminus. Protein purification by this method was carried out in Talon buffer (50mM Tris/HCl, 300mM NaCl, pH 8.0). At this pH histidine, which has a pKa of 6.0, is deprotonated and thus the nitrogens in the imidazole ring have lone pairs of electrons. It is this feature that allows the His_6 tag to bind to immobilised divalent metal ions. The interaction can be dissociated by competing off the tag with high concentrations of imidazole. The metal affinity matrix used in the work was ™ Fast Flow (Clontech Laboratories Inc) in which the His_6 tags interact with immobilised Cobalt ions.

2.2.27.2 IMAC using an isocratic flow by gravity

A 1-5 ml bed volume, depending on the quantity of protein to be purified as assessed by SDS-PAGE, was made in a Fast flow column (Clontech Laboratories Inc). The column was equilibrated with 2 times the bed volume of Talon buffer. CFE, prepared as in Section 2.2.25 and filtered using a 1.2 micron Acrodisc Gelman filter, was applied to the column and allowed to flow by gravity. The column was then washed sequentially with 1 x the bed volume of Talon buffer, 1 x bed volume of Talon containing 5 mM imidazole and finally with 5 1 x bed volumes of Talon buffer containing 100 mM imidazole. Each fraction was analysed by SDS-PAGE (Section 2.2.26). Fractions containing recombinant protein of the required purity were kept and

either subjected to further purification or dialysed extensively against an appropriate buffer.

2.2.27.3 Ion exchange chromatography

A Bio-Rad BioLogic DuoFlow™ System with a flow rate of 2 ml/min connected to a UNO™ Q12 anion exchange column (Bio-Rad), using a Bio-Rad BioFrac™ fraction collector, was used to carry out a secondary purification step if IMAC had not generated protein of the desired purity. Protein fractions were dialysed into 10 mM Tris/HCl pH 8.0 and loaded onto the UNO™ Q12 column, which had been equilibrated with 100 ml of 10 mM Tris/HCl pH 8.0. The column was washed with 60 ml of the equilibrating buffer to remove unbound protein, before elution with a 120 ml linear 0–500 mM NaCl gradient in 10 mM Tris/HCl pH 8.0. Protein elution was detected by UV-absorbance and fractions of 1.0 ml were collected above a threshold of 0.05 AU at 280nm using a Bio-Rad BioFrac™ fraction collector. Each fraction was collected and analysed by SDS-PAGE (Section 2.2.26).

2.2.27.4 Gel filtration

The same FPLC system as described above, with a flow rate of 1 ml/min, was connected to a Pharmacia HiLoad™ 16/60 Superdex™ 75 prep grade gel filtration column using a Bio-Rad BioFrac™ fraction collector. The protein samples to be loaded were concentrated to 2 ml and loaded using a 2 ml static loop. Elution from the column was done using 10 mM Tris/HCl pH 8.0 containing 150 mM NaCl. Fractions of 1.0 ml were collected and analysed by SDS-PAGE (Section 2.2.26). Figure 2.3 shows a calibration curve for a Pharmicia Hi Load^{TM} 16/60 superdex^{TM} 75 prep grade gel filtration column.

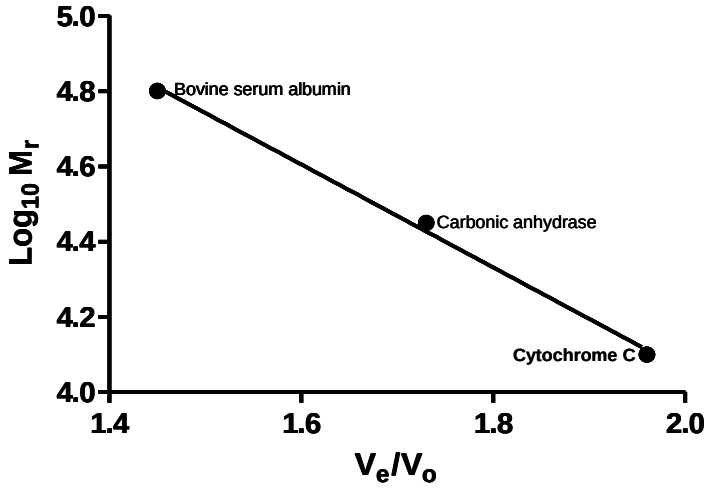


Figure 2.3 Calibration curve for Pharmacia HiLoad™ 16/60 Superdex™ 75 prep grade gel filtration column

2.2.28 Quantification of pure protein solutions

The concentration of pure protein solutions was performed using the methods described by Gill and von Hippel (Gill et al 1989) and Pace et al (Pace et al 1995). Protein solutions were diluted, if necessary, into an appropriate buffer and a scan performed from 200 nm to 400 nm using a Pharmacia Biotech Ultrospec 4000, UV/vis spectrophotometer. Readings at 280 and 320 nm were taken. The 320 nm reading was subtracted from the 280 nm reading, to account for background, and the subsequent value used to determine protein concentration through the use of the beer-lambert law shown below:

$$A = \varepsilon Cl$$

A = Absorbance reading

ε = molar extinction coefficient of sample

C = molar concentration of sample

l = path length

The molar extinction coefficient was calculated using ProtParam at the online ExPASY server (http://expasy.org).

2.2.29 Concentrating protein

Protein solutions were concentrated after filtering (0.22 μm) using 20 ml and 2 ml Vivaspin™ centrifugal concentrators (VivaScience). A 5, 10 or 30 kDa molecular weight cut offs were used as appropriate for the protein size. Centrifugation was performed at 3500 × g using a MSE Mistral 3000i bench centrifuge with a swing out rotor.

2.3 Biochemistry

2.3.1 Enzyme assays

All assays were performed at 37 °C, pH 7.0 and contained 1mg/ml bovine serum albumin unless specifically stated. All reactions were pre-warmed to 37 °C and initiated by the addition of appropriately diluted enzyme, also pre-warmed to 37 °C. The same set of volumetric pipettes was used throughout and all assays carried out in triplicate. The values of K_M and k_{cat} were determined, where possible, by using seven substrate concentrations that straddled the K_M using the formula:

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

V_0 = initial velocity, V_{max} = number of substrate molecules hydrolysed at infinite substrate concentration, K_M = concentration at which the reaction rate is half $V_{max,}$ [S] = substrate concentration

Where independent K_M and k_{cat} values could not be determined k_{cat}/K_M was determined at substrate concentrations considerably below the K_M where:

$$V_0 = (k_{cat}/K_M)[S]$$

k_{cat} = number of substrate molecules turned over per molecule of enzyme

All graphs were plotted in GraphPad Prism 4.02 and the kinetic constants were determined by non-linear regression of the plots of V_0 against [S].

2.3.1.1 4-nitophenyl α-L-arabinofuranoside and 4-nitrophenyl β-D-xylopyranoside assays

The release of 4-nitrophenolate from 4-nitrophenyl-α-L-arabinofuranoside and 4-nitrophenyl β-D-xylopyranoside was carried out at 25 ºC in 50 mM sodium

phosphate buffer, pH 7.0, using a Pharmacia Biotech Ultrospec 4000, UV/vis spectrophotometer at a wavelength of 400 nm. All reactions were performed in glass cuvettes. At pH 7.0 4-nitrophenolate has an extinction coefficient of 10500 M^{-1} cm^{-1} (Figure 2.4).

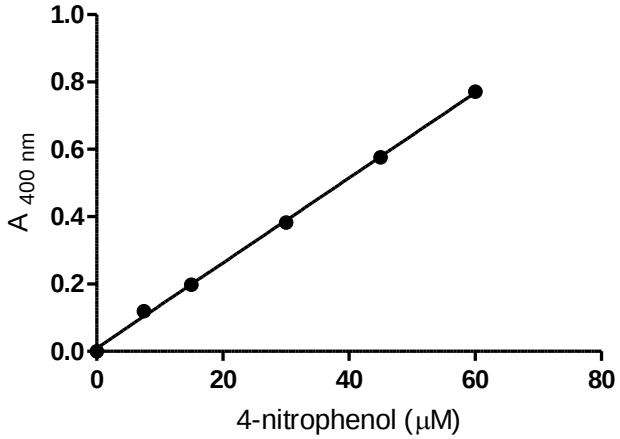


Figure 2.4 Standard curve of 4-nitrophenol at pH 7.0

2.3.1.2 pH optimum assays using 4-nitrophenyl α-L-arabinofuranoside and 4-nitrophenyl β-D-xylopyranoside

Assays were carried out as described above with the exception that a range of different buffers were used, corresponding to the pH to be tested (Table 2.15). Assays were monitored discontinuously and time points were terminated by the addition of an equal volume of 500 mM Na_2CO_3. The addition of Na_2CO_3 raises the pH to 11. At this pH 4-nitrophenolate has an extinction coefficient of 24150 (Figure 2.5). Some pH values could not be overlapped due to inhibition of the enzyme by the buffer.

Buffer	Buffer range	pH used
50mM Sodium acetate	3.7-5.6	4.5,5.0,5.5
50mM MES	5.5-6.7	5.5,6.0
50mM Sodium phosphate	6.0-8.0	6.0,6.5,7.0,7.5,8,0
50mM CAPSO	8.9-10.3	9.0,9.5,10.0

Table 2.15 pH of buffers for pH optimum determination using 4-nitrophenyl α-L-arabinofuranoside as substrate

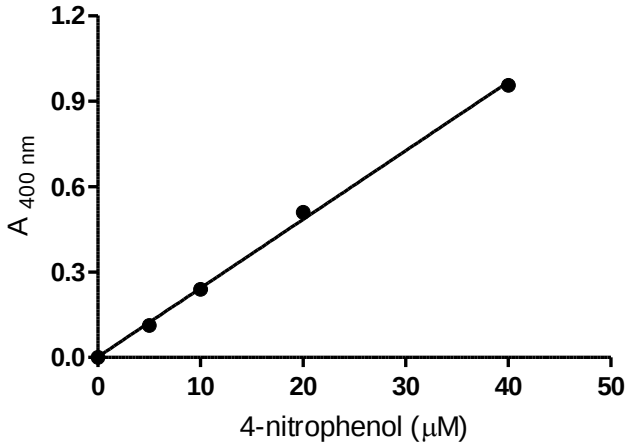


Figure 2.5 Standard curve for 4-nitrophenol at pH 11.0

2.3.1.3 2,4-Dinitrophenyl β-D-1,4 cellotriose assays

The release of 2,4-dinitrophenolate from 2,4-dinitrophenyl β-D-1,4 cellotriose was monitored at 37 ºC in 50mM MES, 20mM NaCl pH 5.5, using a Pharmacia Biotech Ultrospec 4000, UV/vis spectrophotometer at a wavelength of 400 nm. All reactions were performed in glass cuvettes. At pH 5.5 2,4-nitrophenol has an extinction coefficient of 9890 M^{-1} cm^{-1} (Figure 2.6).

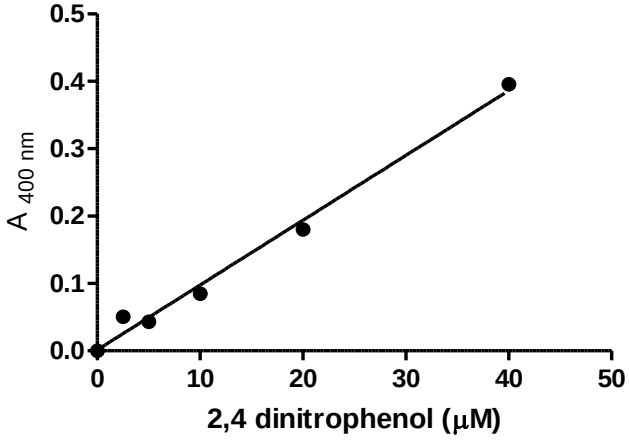


Figure 2.6 standard curve for 2,4 dinitrophenol at pH 5.5

2.3.1.4 4-methylumbelliferyl-Glc-β-D-1,4-Glc-β-D-1,3-Glc assays

The release of 4-methylumbelliferone from 4-methylumbelliferyl-Glc-β-D-1,4-Glc-β-D-1,3-Glc was monitored fluorescently using an excitation and emission wavelength of 365 nm and 440 nm respectively. The assays were carried out in 50 mM sodium phosphate buffer pH 7.0 in 5 mm × 5 mm quartz cuvettes using a Cary Eclipse Varian spectrofluorimeter. The extinction coefficient of 4-methylumbelliferone, as measured by emission at 440 nm, was 273993 M^{-1} cm^{-1} (Figure 2.7).

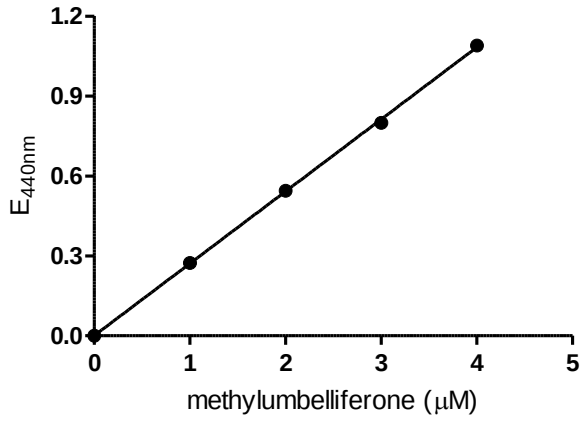


Figure 2.7 Standard curve of methylumbelliferone emission E_{440nm} at pH 7.0

2.3.1.5 Inhibition assay using 4-methylumbelliferyl-Glc-β-D-1,4-Glc-β-D-1,3-Glc

Inhibition assays were carried out as described above in Section 2.3.1.4 with 20 μM substrate in the presence of several inhibitor concentrations that straddled the K_I. The following formula was used to calculate the K_I:

$$1/ K_I = ((V_0/V_i)-1) / [I]$$

$1/ K_I$ = reciprocal of K_I, V_0 = uninhibited rate, V_i = inhibited rate, [I] = inhibitor concentration

2.3.1.6 Galactose/Arabinose and xylose dehydrogenase assay

The release of arabinose and xylose was monitored through the use of a linked assay utilising galactose or xylose dehydrogenase, respectively (galactose dehydrogenase is also able to oxidise arabinose). Galactose and xylose dehydrogenase catalyse the oxidation of galactose/arabinose and xylose, respectively, and reduces NAD^+ to generate NADH in a 1:1 molar ratio. NADH absorbs at A_{340nm} and thus its synthesis can be directly monitored. An extinction coefficient of 6230 M^{-1} cm^{-1} was

used to calculate the NADH concentration. All reactions were carried out at 25 °C in 50 mM sodium phosphate buffer pH 7.0. Glass cuvettes were used and NADH release monitored using a Biotech Ultrospec 4000, UV/vis spectrophotometer. Table 2.16 lists the reaction composition for the linked assay.

Volume	Component
50 µl	Appropriately diluted enzyme
50 µl	0.5 M sodium phosphate pH 7.0
50 µl	10 mM NAD^+
50-345 µl	Substrate
5 µl	Galactose dehydrogenase or xylose dehydrogenase 500 U/ml (1 U = oxidation of 1 µmol galactose min^{-1})
50-345 µl	Sterile water
500 µl	Total volume

Table 2.16 Showing composition of a typical galactose dehydrogenase linked assay

2.3.1.7 Substrate depletion method

This method was used to calculate k_{cat}/K_M for cellooligosaccharide and mannooligosaccharide substrates where hydrolysis could only be monitored by high pressure liquid chromatography (HPLC) detailed in Section 2.3.2. Assays using cellooligosaccharides were performed at 60 °C, when xylooligosaccharides and mannooligosaccharrides were the substrates assays were performed at 37 °C. The substrate concentration was considerably below the K_M and the following formula was used:

$$k \times t = \ln (S_0/S_t)$$

$k = k_{cat}/K_M$[enzyme], t = time, ln = natural log, S_0 = substrate concentration at time 0, S_t = substrate concentration at time t.

2.3.1.8 3, 5-Dinitrosalicyclic acid assay (DNSA)

The rate of hydrolysis of polysaccharides was monitored by DNSA using the method described by Miller (Miller 1959). Sugars with a free anomeric carbon can exist as a linear molecule with an aldehyde group that is capable of acting as a reducing agent. Upon reduction DNSA undergoes a colour change which can be

detected at 575 nm. Reactions were monitored discontinuously and time points terminated by the addition of 500 µl of DNSA reagent (1 % (w/v) DNSA, 0.2 % (v/v) phenol, 1 % (w/v) NaOH, 0.008 % glucose, 0.05 % (w/v) $NaSO_3$) to 500 µl reactions. The mixtures where then boiled for 20 min, left on ice for 2 min and then allowed to equilibrate at room temperature before absorbance readings were taken at 575 nm using a Pharmacia Biotech Ultrospec 4000, UV/vis spectrophotometer. A standard curve of 0-1000 µg/ml monosaccharide (Figure 2.8) was used to quantify the released reducing sugar. DNSA assays performed with glucose and mannose polysaccharides were carried out at 60 and 37 °C respectively.

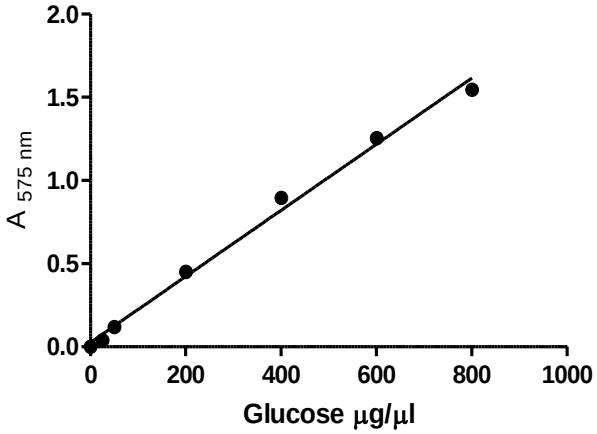


Figure 2.8 DNSA standard curve using glucose concentrations 0-800 µg/µl

2.3.2 High performance liquid chromatography (HPLC)

2.3.2.1 Oligosaccharide analysis

Oligosaccharides from polysaccharide hydrolysis, generated both by enzymatic and by acid treatment, were analysed using a CARBOPAC™ PA-100 anion exchange column (Dionex) equipped with a CARBOPAC™ PA-100 guard column. The fully automated system (LC25 Chromatography Oven, GP40 Gradient Pump, ED40 Electrochemical Detector and AS40 Autosampler) had a loop size of 200 µl, flow rate of 1.0 ml/min, operated at room temperature, a pressure of ≈ 2400 psi and sugars were detected by pulsed amperometric detection (PAD). The PAD settings were E_1 = +0.05, E_2 = +0.6 and E_3 = -0.6. Typical elution conditions used were 0-5 min 66 mM NaOH, 5-30 min 66 mM NaOH with a 0-75 mM sodium acetate linear gradient, 30-40 min 500 mM sodium acetate in 66 mM NaOH, 40-50 min 500

mM sodium hydroxide and then from 50-60 min 66 mM NaOH. For cellooligosaccharides the linear gradient was extended to 5-35 min 66 mM NaOH with 0-95 mM sodium acetate, all other steps were kept the same. Appropriate oligosaccharides were used as standards at a concentration of 50 μM. Samples to be analysed were centrifuged at 13000 x g for 10 min and if necessary diluted. Data were collected and manipulated using Chromeleon™ Chromatography Management System V.6.8 (Dionex) *via* a Chromeleon™ Server (Dionex). Final graphs were drawn with Prism 4.02 (GraphPad).

2.3.2.2 Analysis of the anomeric configuration of oligosaccharides

The anomeric configuration of newly created reducing ends from enzymatic hydrolysis was analysed using a spherisorb hydrophobicity columnm S5 ODS2; 25 cm x 4.6 mm (Waters corporation); using an adapted method first described by Braun et al (Braun et al. 1993). Two columns were run in series using water as the eluent. The same system was used as described above in Section 2.3.2.2 as well as flow rate and loop size. Oligosaccharides were detected as stated above in Section 2.3.2.2; this was made possible by the post column addition of 1 M NaOH at a flow rate of 0.5 ml/min. Appropriate oligosaccharides were used as standards at a concentration of 50 μM. Samples were not centrifuged or diluted and were loaded directly after 10 min incubation with enzyme. This was to ensure no significant mutarotaion of the newly created reducing ends had occurred.

2.3.3 Isothermal titration calorimetry (ITC)

ITC can be used to measure biomolecular interactions. It can simultaneously determine all the thermodynamic parameters for an interaction in a single experiment. It works by measuring the heat released or absorbed during biomolecular binding events, with respect to a reference cell. As heat is released or absorbed during the bimolecular interaction the ITC machine will adjust the amount of power supplied to the sample cell to keep its temperature the same as the reference cell. It is this differential power that is measured and translated into changes in heat.

During this study a Microcal ITC_{200} calorimeter was used throughout. ITC measurements were made at temperatures that ranged from 10 °C to 60 ^{o}C. An inactive catalytic acid mutant of *Ct*Cel119 (E96A) was dialysed against 50 mM MES buffer, 20 mM NaCl pH 5.5. Cellohexaose was dissolved in the dialysate and

regenerated cellulose (RC) was also resuspended in the same buffer to minimize heats of dilution. During a titration experiment, E96A 88 μM, was stirred at 500 rpm in a 200 μ1 reaction cell, which was injected with 20 successive 2 μl aliquots of 2 mM cellohexaose. For experiments with RC, prepared as described in Section 2.3.6.5, the ligand was retained in the cell at 29.4 mg/ml and the protein (812 μM) was the titrant, which was injected with 40 successive injections of 1 μl.

Integrated heat effects, after correction for heats of dilution, were analyzed by non-linear regression using a single site-binding model (Microcal Origin, version 7.0). The fitted data yield the association constant (K_A) and the enthalpy of binding (ΔH). Other thermodynamic parameters were calculated using the standard thermodynamic equation:

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

R = gas constant, ln = natural log, ΔG = change in Gibbs free energy, ΔH = change in enthalpy, T = absolute temperature (Kelvin), ΔS = change in entropy

To measure the change in heat capacity (ΔC_p), the following equation was used:

$$\Delta C_p = \Delta H_{T1\text{-}T2}/T1\text{-}T2$$

ΔC_p = change in heat capacity, ΔH = change in enthalpy, T = absolute temperature (Kelvin)

2.3.4 Depletion binding isotherms

Depletion binding isotherms were performed against regenerated cellulose (RC) in 50 mM MES, 20 mM NaCl pH 5.5. Protein concentrations between 1 and 30 μM were added to RC at 1mg/ml (final volume 0.8 ml). Samples were incubated at 40 °C for 1 h with gentle mixing and then centrifuged at 13000 x g. The supernatant was then removed and its absorbance at 280 nm was measured to quantify the amount of protein present, which was then subtracted from total protein to give the amount of protein bound to the RC.

2.3.5 Thin layer chromatography (TLC)

TLC was used as an initial screening technique for enzyme activity and, for the detection of sugars samples purified from P2 columns. A TLC plate (Silicagel 60,

20 × 20, Merck) was cut to the desired size (the minimum height should be 10 cm). Samples (2 x 1 µl drops) were spotted on plates, allowing the first spot to dry before applying the next. TLC plates were then placed in a tank containing 5-10 ml of TLC running buffer (1-butanol/acetic acid/water 2:1:1 (v/v)). The tank was allowed to equilibrate for up to 2 h before use. The TLC plate was then left to allow the samples to migrate up the plate, via capillary action, until the solvent front was 1 cm from the top of the plate. Plates were then removed and dried. Visualisation of the migrated sugars was achieved by immersing the TLC plate in developer (sulphuric acid/ethanol/water 3:70:20 v/v, orcinol 1 %) for 1-2 seconds. Plates were again dried and developed by heating between 80 and 100 °C.

2.3.6 Preparation of oligosaccharide and polysaccharide substrates

2.3.6.1 Insoluble Mannan

Insoluble mannan was prepared from carob galactomannan. Carob galactomannan (6 g) was dissolved in 400 ml of 50 mM sodium phosphate buffer, pH 7.0, to a final concentration of 1.5 % (w/v). Complete dissolution was achieved by a combination of vigorous mixing and heating to 50 °C; the mixture was also autoclaved to make sure all carob galactomannan had solubilised and was sterile. To the 1.5 % carob galactomannan 1000 units (1 unit = amount of enzyme required to hydrolyse 1µmole/min of substrate) of the *Cellvibrio japonicus* GH27 α-galactosidase (*Cj*Aga27 (Halstead et al. 2000)) was added. The reaction was then left 1-2 days until complete. A zero time point with no enzyme was taken immediately before the addition of *Cj*Aga27. The reaction was monitored and judged to be complete when the release of galactose had plateaued, measured using the galactose dehydrogenase assay (Section 2.3.1.6). When complete the insoluble mannan is visible as a white suspension. The suspension was then centrifuged, in 500 ml Nalgene bottles, at 5000 x g. The supernatant was decanted and the insoluble mannan pellet resuspended in water. This process was repeated another 4-5 times to ensure no soluble oligosaccharide or polysaccharides were present. The insoluble mannan was then freeze dried (Section 2.3.6.7).

2.3.6.2 Washing of carob galactomannan

High molecular weight oligosaccharides (7-10 sugar residues) were removed by precipitating carob galactomannan with 70 % (v/v) ethanol; high molecular weight oligosaccharides remain dissolved in this solvent. The precipitated carob galactomannan was then spun down at 5000 x g and the supernatant removed. This step was repeated two more times. The final step involved washing with 100% ethanol, then centrifuging at 5000 x g, decanting of the supernatant and allowing the pellet to air dry.

2.3.6.3 Preparation of arabinans for NMR analysis

Sugarbeet arabinan (purchased from Megazyme International) was prepared for NMR analysis using both a mild acid hydrolysis and ethanol precipitation. This was done to reduce noise and the overlapping of multiple signals. Sugarbeet arabinan at a concentration of 30 mg/ml, in a 20 ml volume, was boiled in the presence of 50 mM HCl for 30 min. The sample was then placed on ice for 2 min after which 1 ml of 1 M NaOH was added to neutralise the acid. This sample was then freeze dried (Section 2.3.6.7). The freeze dried material was resuspended in 2 ml of water, to which 8 ml of ethanol was added to a final concentration of 80 % (v/v). The material was centrifuged at 5000 x g the supernatant removed and then the pellet resuspended in 100% ethanol; centrifuged at 5000 x g again, the supernatant removed and the pellet allowed to air dry. This material was then suitable for analysis by NMR.

2.3.6.4 Preparation of Phosphoric Acid Swollen Cellulose (PASC)

PASC was prepared using the adapted method of Wood (Wood 1988). Avicel (10 g) was mixed with a small amount of ethanol to from a paste. This was then added to 200 ml of concentrated phosphoric acid (85% w/v) and left for one hour at 4 °C with occasional stirring. This mixture was then poured into 4 litres of ice cold water and left at 4 °C for 30 min. The mixture was then washed with 10 % (w/v) $NaHCO_3$ before being dialysed against water at 4 °C. The PASC was then placed in a waring blender for 60 s and 5 mM sodium azide was added as a preservative.

2.3.6.5 Preparation of Regenerated Cellulose (RC)

RC was prepared as described by Boraston *et al* (Boraston 2005). Avicel (10 g) was mixed with a small amount of ethanol to from a paste. This was then added to

200 ml of concentrated phosphoric acid (85 % w/v) and left at room temperature until dissolved (4–6 h). The RC was then precipitated by the addition of two volumes of cold acetone and then dialysed extensively against water. Sodium azide (5 mM) was added as a preservative.

2.3.6.6 Synthesis of 1,6 anhydro cellotriose

1,6 anhydro cellotriose was synthesised as described by Tanaka *et al* (Tanaka 2009). Briefly, a mixture of 50 mM cellotriose, 300 mM 2-chloro-1,3-dimethyl imidazolinium chloride (DMC) and 900 mM ethylamine (Et_3N) was mixed in 0.5 ml of water at 0 °C for 15 min. The product was then purified down a P2 bio gel resin (Bio rad) and the product confirmed by 2-D NMR.

2.3.6.7 Freeze drying of purified sugars

Purified oligosaccharides and polysaccharides were frozen to -80 °C and then lyophilised in a Christ Alpha 1-2 Freeze Drier at -80 °C.

2.4 Crystallography

In order to solve the three dimensional structures of the proteins in this study both molecular replacement and single anomalous dispersion (SAD) techniques were employed. The latter technique required modified proteins that contained heavy atoms, which cause anomalous scattering.

2.4.1 Protein sample preparation

2.4.1.1 Native Protein samples

Proteins were purified to ≥95% as assessed by SDS-PAGE. Purification was done through a combination of IMAC, anion exchange and finally gel filtration chromatography (Section 2.2.27). After purification proteins were dialysed extensively, where possible, against water. *Cj*Man26C was dialysed against 5mM Hepes pH 7.0 and *Cj*Abf43B D14A against 50 mM NaCl.

2.4.1.2 Selenomethionine derivatives

This heavy atom derivative was used to perform SAD experiments. In order to express Seleno-methionine derived forms of protein *E. coli* strain B834, a methionine auxotroph, was used. A specific medium was prepared, using a kit from Athena Enzyme Systems™ comprising a SelenoMet Medium Base™, a SelenoMet™ Nutrient Mix and a SelenoMethionine Solution, following the manufacturer's instructions. *E. coli* strain B834, transformed with the appropriate plasmid, was inoculated into 5 ml of LB and grown at 37 °C, 180 rpm for 16 h in the presence of the appropriate antibiotic. This culture was then inoculated into a 500 ml unbaffled flask containing 100 ml of LB supplemented with the appropriate antibiotic. The culture was grown to an optical density of A_{600nm} = 0.2. Cells were then harvested by centrifugation at 5000 x g for 5 min, the supernatant decanted and the cell pellet re-suspended in 5 ml of SelenoMet Medium Base™, this step was repeated 3 times to remove any trace of LB media. The final suspension of cells was used to inoculate 1 litre of SelenoMet Medium Base™ in a 2 litre baffled flask. Then 50 ml of SelenoMet™ Nutrient Mix was added followed by 4 ml of a Selenomethionine Solution (10 mg/ml). Prior to the addition of the Selenomethionine Solution 10 ml of culture was removed to act as a control; the culture should not grow without the addition of Selenomethionine Solution. Overexpression of the selenomethionine protein was then carried out as described in Section 2.2.24. Cells were harvested by centrifugation at 5000 x g and resuspended in 5 ml of Talon buffer (50 mM Tris-HCl pH 8.0, 300 mM NaCl) supplemented with 5 mM β-mercaptoethanol. CFE was prepared as described in Section 2.2.25 and selenomethionine protein purified using the same protocols as native protein. All IMAC buffers were supplemented with 5 mM β-mercaptoethanol; anion exchange and gel filtration buffers were supplemented with 10 mM DL-Dithiothreitol (DTT, Sigma). The purified selenomethionine proteins were extensively dialysed against 10 mM DTT.

2.4.2 Protein crystallisation screen

Commercially available crystallisation screens were used to find initial conditions for protein crystallisation. Crystal screen 1, 2 and PEG/Ion™ 1 were purchased from Hampton research. The PACT and JCSG+ suites were purchased from Qiagen. Initial crystal screens were typically carried out using a protein concentration of 20 mg/ml. For co-crystallisations a ligand concentration of between

10 and 70 mM was used. Optimal crystals were obtained through varying the precipitant and salt concentrations, as well as the pH of the original condition found in the initial screen. Crystallography was conducted in collaboration with Professor Gideon Davies, YSBL, University of York and Professor Richard Lewis, ICaMB, Newcastle University

2.4.3 Growth of crystals

Initial crystal screens employed sitting drop vapour diffusion and were setup using a Mosquito™ (TTP Labtech) nanolitre pipetting robot with 96-well crystallization plates (Greiner). Mother liquor, 100 µl, was pipetted into the wells, 4 µl of protein sample was then pipetted into the 4 sample reservoirs. A 100 nl protein sample was then aliquoted by the robot, to which the robot then added 100 nl of a crystallization condition. The plates were then sealed using easyxtal sealing films (Qiagen). Optimisations were done by hand and employed the hanging drop vapour diffusion method. Protein and mother liquour were mixed together, typically 1 µl of mother liquor was added to 1 µl of protein sample, on a silliconised glass cover slip 18 mm x 18 mm. Cover slips were pre-treated with Aqua Sil™ (Hampton research) and polished with a silk scarf prior to use. Cellstar 24 well tissue culture trays were used and prepared by dispensing 0.5 ml of mother liquor per well, then applying Vacuum Grease (Dow Corning) to the rim of each well. Cover slips, prepared as described above, were immediately inverted and sealed by pressing down onto the vacuum grease, coating the rim of the wells. All trays were then incubated at 20 °C and were checked daily for crystal growth for the first two weeks; and weekly there after. Drops were viewed using a using a Leica MZ-6 crystallisation microscope.

2.4.4 Crystal screening

Crystals thought to be of good quality, based on morphology and size, were fished from drops using rayon fibre loops (Hampton research) and flash frozen in liquid nitrogen. The true quality of the crystals was assessed by their diffraction pattern. Two images were taken of crystals, 0 ° and 90 ° with a 0.5 ° oscillation, using a Rigaku rotating RU-200 X-ray generator, with a Cu 1.5418 Å target operating at 50 kV and 100 mA with focusing X-ray optics from Osmics. The diffraction pattern produced was used to judge whether the crystal was suitable to be sent to a synchrotron and indexed using crystal clear to determine the space group.

2.4.5 Structure solution

Data collected was indexed using iMosflm (Leslie 1992) and subsequently scaled using Scala (Evans 2006); both programs are part of the CCP4 suite of programmes (1994).

2.4.5.1 Molecular replacement

Molecular replacement is a method that allows the determination of a new structure through the use of an existing model which shares a high degree of similarity. The two proteins should be ~40 % identical at the primary sequence level. Molecular replacement takes the existing structure, for which phase information is known, and uses its Patterson map to perform a series of complex translational, rotational and rigid body displacements onto the Patterson map of the unknown structure. Once a suitable match has been found the phases of the unknown 'new structure' can be determined. Phaser (McCoy et al. 2007) was used to perform all molecular replacements during this study and is part of the CCP4 programme suite (1994).

2.4.5.2 Single wavelength anomalous dispersion (SAD)

For structure solution where there were no suitable models to perform molecular replacement SAD was used, utilising selenomethionine protein (Section 2.4.1.2). At a wavelength of 0.98Å selenium atoms cause the anomalous scattering of X-rays. This anomalous scattering allows the construction of a Patterson map for the heavy atoms only (selenium in this case), from which the phase information can be calculated and used as a reference to calculate the phase angles of all other reflections. ShelXC/D/E (Sheldrick 2008) was used to calculate the Patterson and phase angles of the heavy atoms and subsequently all other phase information. Profess (1994) was used to work out any non-crystallographic symmetry. This information was then used for density modification using DM (1994) utilising a pdb containing the heavy atoms only. Arp/Warp (Langer et al. 2008) was used to perform automatic model building into the calculated electron density.

Refinement of all structure solutions was performed using Refmec5 (Murshudov et al. 1997) and manual correction was performed using COOT (Emsley et al. 2004) . Prior to refinement Uniqueify (1994) was used to generate a free R value

(5 % of the observations set aside for cross-validation analysis) and used to monitor the accuracy of the model as refinement proceeded.

2.4.6 Visualisation of structures

Crystal structures were viewed using PyMOL (Delano Scientific), which was also used to produce figures.

Chapter 3

Structural and functional characterisation of novel glycoside hydrolase family 26 members

3.1 Introduction

Mannan is a hemicellulose consisting of a β-1,4 backbone of mannose (in glucomannans the backbone is glucose and mannsoe) which, can be decorated with galactose at O6 and acetate at O2 and/or O3 depending on its origin. It is a major matrix polysaccharide found in the cell walls of angiosperms and soft woods (Moreira et al. 2008).

Mannose is an epimer of glucose differing solely at the C2 position where its hydroxyl group adopts an axial instead of an equatorial configuration (Figure 3.1)

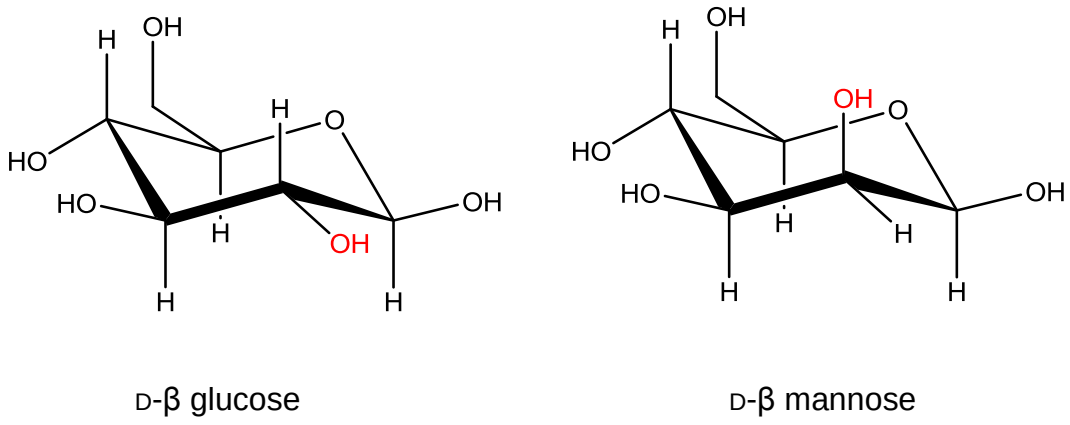


Figure 3.1 The epimers of D-β glucose and D-β mannose.
The C2 hydroxyl is highlighted in red

The configuration of the hydroxyl at C2 in mannose has two consequences; it is a specificity determinant which can be selected by proteins and it causes unfavourable stereoelectronic clashes during nucleophilic substitution at the anomeric carbon. It is for this second reason that enzymes, which hydrolyse mannan polymers, are so remarkable. They catalyse, often rapidly, what is regarded as an extremely difficult chemical substitution due to the 1,2 diaxial interactions that need to be alleviated during catalysis (Ducros et al. 2002; Tailford et al. 2008).

Glycoside hydrolase subsites are numbered around the site of bond cleavage using the nomenclature of Davies *et al* (Davies et al. 1995). The subsites directly

either side of bond cleavage are numbered +1 and -1. Subsites downstream of +1 increase in number being +2, +3 etc toward the reducing end of the substrate. Subsites upstream of -1 decrease in number being -2, -3 etc toward the non–reducing end of the substrate.

Glycoside hydrolase family 26 (GH26) consists almost exclusively of endo-β-1,4 mannanases (Henrissat et al. 1995; Henrissat et al. 1997). The family is in clan GH-A, displays a $(\alpha/\beta)_8$ fold (TIM- barrel), and performs catalysis through a double displacement mechanism leading to retention of the anomeric configuration of the substrate. In GH-A the catalytic acid/base and nucleophile (both glutamates) are located on the C-terminus of β strands 4 and 7 respectively (Cantarel et al. 2009). All members of this superfamily, except for GH26, have the NEP motif, in which an aspargine and proline flank the catalytic acid/base. In GH26 the asparagine is replaced with a histidine to give the HEP motif (Henrissat 1991). An aspargine to histidine substitution is a conserved change as both residues can act as hydrogen bond donors and acceptors.

As alluded to earlier GH26 contains mainly endo-β 1,4 mannanases, the best characterised of which is the *Cellvibrio japonicus* mannanase *Cj*Man26A. There are, however, two notable exceptions, the Alcaligenes sp. strain XY-234 endo-β-1,3-xylanase 3XynAlc (Sakaguchi 2010) and the *Clostridium thermocellum* endo-β-1,3,1,4-lichenase *Ct*Lic26A (Taylor et al. 2005).

*Ct*Lic26A was previously identified as a mixed linkage glucanase. It is part of a modular protein comprising of a GH26 (*Ct*Lic26A) and GH5 (*Ct*Cel5E) catalytic modules, a family 11 non-catalytic carbohydrate binding module (CBM11) and a type 1 dockerin module (Figure 3.2) (Taylor et al. 2005). *Ct*Cel5E was shown to be a cellulase active on crystalline polymers of β-1,4 glucose, such as Avicel, and its activity was potentiated against insoluble substrates by CBM11 (Taylor et al. 2005). The presence of a dockerin indicates that the modular enzyme is a component of the *C. thermocellum* cellulosome, which is a multi enzyme complex appended to the surface of the bacterium that catalyses plant cell wall degradation (Fontes et al 2010.).

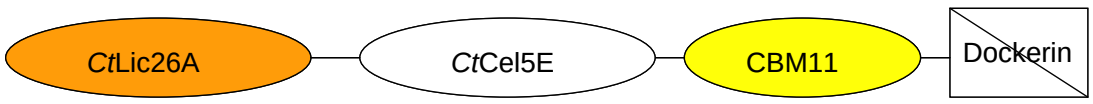


Figure 3.2 Modular architecture of mature *Ct*Lic26A -Cel5E

In a previous study the activity of *Ct*Lic26A was detailed and the crystal structure solved in complex with an inhibitor revealing the critical -1, and -2, subsite interactions (Taylor et al. 2005). Further crystallographic studies shed light on the likely transition state by trapping snapshots along the reaction pathway of methylumbelliferyl β-laminarinbioside hydrolysis. The observed conformation adopted in the Michealis complex by the substrate, which was between a $^{1}S_{3}$ skew boat and a $^{1,4}B$, boat strongly suggests a $^{4}H_{3}$ half chair as the likely transition state (Money et al. 2006). This is in contrast to the proposed $B_{2,5}$ transition state adopted by mannan substrates of GH26 enzymes and some other Clan GH-A mannanases (Ducros et al. 2002; Tailford et al. 2008). The structural characterisation of *Ct*Lic26A has provided an opportunity to explore the role of key active site residues in catalysis and substrate binding, which are explained in this chapter.

The other enzyme studied in this chapter is *Cj*Man26C. The enzyme was identified, by bioinformatic analysis of the genome of *C. japonicus*, as a β-mannanase (DeBoy et al. 2008). *Cj*Man26C is one of three GH26 mannanases appended to the outer membrane through a type 2 signal peptide (Tailford 2007). The β-mannanases of *C. japonicus* have been extensively explored and the analysis of *Cj*Man26C completes the most comprehensive mannan-degrading apparatus reported to date.

3.2 Objectives

1. To provide an analysis of the substrate binding cleft of *Ct*Lic26A, using a strategy informed by crystallographic data.

2. Explore the substrate binding features of *Cj*Man26C, through structure/function studies. Together these two goals will provide a functional context to GH26 enzymes in *C. japonicus*.

3.3 Results

3.4 *Ct*Lic26A

3.4.1 Expression and Purification

The pET28a based recombinant plasmid pCF1 **(provided by Professor Carlos Fontes)** was transformed into *Escherichia coli* strain BL21 (DE3) (section 2.1.10) and expression of the lichenase was induced by the addition of 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) when the A_{600nm} = 0.6. Induction of recombinant protein expression was performed at 37 °C for 4 hours (section 2.2.24). Cells were then centrifuged, cell free extracts prepared (section 2.2.25), and immobilised metal affinity chromatography (IMAC) was then used to purify *Ct*Lic26A (29 kDa) to electrophoretic homogeneity. An SDS-PAGE gel of the purification process is shown in Figure 3.3. The concentration of *Ct*Lic26A was calculated using the molar extinction coefficient of 109320 at A_{280nm}, and its predicted size of 29 kDa is similar to its calculated molecular weight based on the primary sequence of the mature protein.

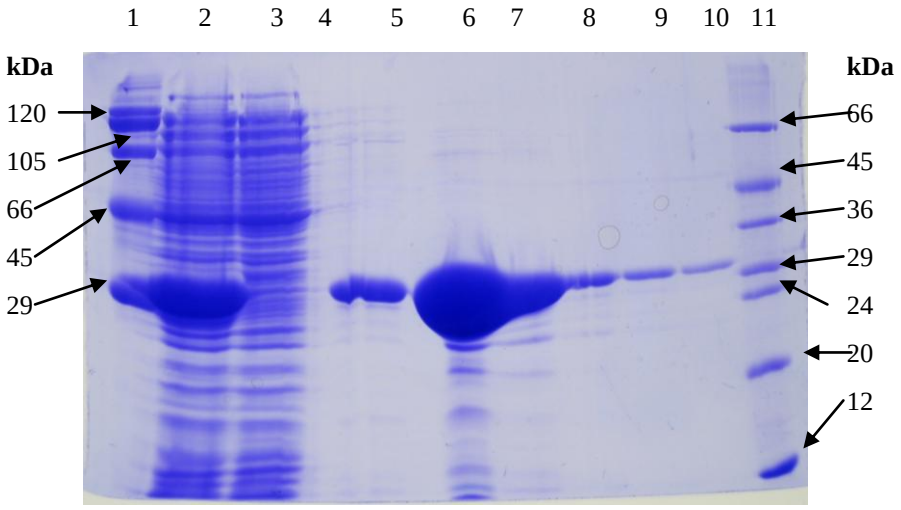


Figure 3.22 SDS-PAGE of IMAC purification of *Ct*Lic26A.
1 high molecular weight marker; 2, cell free extract; 3, Flow through; 4, Talon buffer (See section 2.2.27 for composition) wash; 5, Talon buffer supplemented with 10 mM imidazole; 6-10, Talon buffer supplemented with 100 mM imidazole; 11, Low molecular weight marker

3.4.2 K_i of the thiopentasaccharide

An inhibitor was used to obtain a crystal structure of *Ct*Lic26A with a substrate analog. The K_i of the thiopentasaccharide, determined as described in section 2.3.1.5, was calculated to be ~70 µM from the data presented in Figure 3.4. The inhibitor is a of mixed β-1,3,1,4 glucan where the scissile glycosidic oxygen is replaced with a sulphur (Figure 3.5). This prevents hydrolysis of the compound by the enzyme, *Ct*Lic26A cleaves the β-1,4 glycosidic bond between the +1 and -1 subsites but requires a β-1,3 linkage between the -1 and -2 glucose.

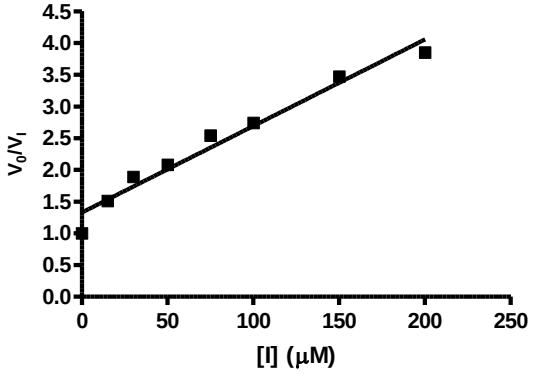


Figure 3.4 Graph showing the linear regression to calculate K_i
V_0/V_i (uninhibited/inhibited rate) was plotted against the inhibitor concentration and the reciprocal of the slope used to determine the K_i. Rate measurements were done using a substrate concentration of 20 µM.

3.4.3 Crystal structure with a thiopentasaccharride

The crystallisation and structure determination of *Ct*Lic26A were carried out by **Dr Victoria A Money from the laboratory of Professor Gideon J Davies**. Crystals of *Ct*Lic26A were grown by hanging drop vapour diffusion in 30% polyethylene glycol (PEG) 5000 MME, 0.15 ammonium sulphate and 0.1 M MES pH 6.5 with 10 mM ligand present. The crystal structure of the enzyme-inhibitor complex revealed two molecules of the thiopentasaccharide, with the reducing end of one molecule occupying the -2 and -3 susbsites, and the non-reducing end of another molecule occupying +1 and +2 (Figure 3.5 and 3.6).

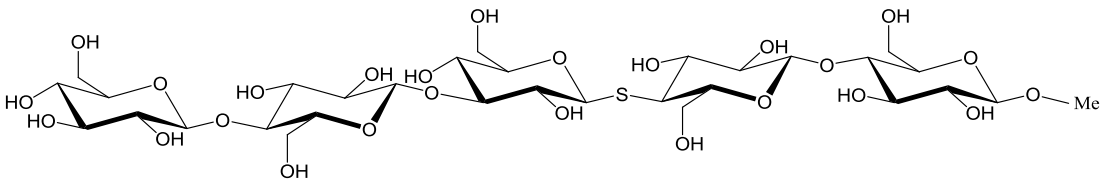


Figure 3.5 The thiopentasaccharide used in the this study.
The glycosidic oxygen between the +1 and -1 sugars has been replaced with a sulphur. The bond between the glucose residue at -1 and -2 is a β-1,3 linkage. All other bonds are β-1,4.

3.4.4 Subsite interactions

3.4.4.1 Positive subsites

N157 hydrogen bonds via its Oδ1 to O3 of the glucose at +2, and to the O6 of the glucose at +1 via Nδ2. This encompasses all the polar interactions made in the positive subsites. W114 stacks against the sugar at the +1 subsite providing a hydrophobic platform (Figures 3.7 and 3.8).

3.4.4.2 Negative subsites

The -1 subsite houses the catalytic apparatus, the acid/base, E109, and the nucleophile, E222. The other residue analysed at the -1 subsite is Y115 which makes interactions with the glycosidic oxygen between -1 and -2 in a native polymer. The hydroxyl group of Y115 interacts with O2 of the glucose at the -2 subsite. The other polar interactions at -2 come from E70 which hydrogen bonds to O2 and O3 through Oε2 and Oε1, respectively, E258 which hydrogen bonds to O6 via Oε1, and K260 hydrogen bonds to the endocyclic ring oxygen. F41 and W72 provide hydrophobic stacking interactions and complete the -2 subsite (Figures 3.7 and 3.8).

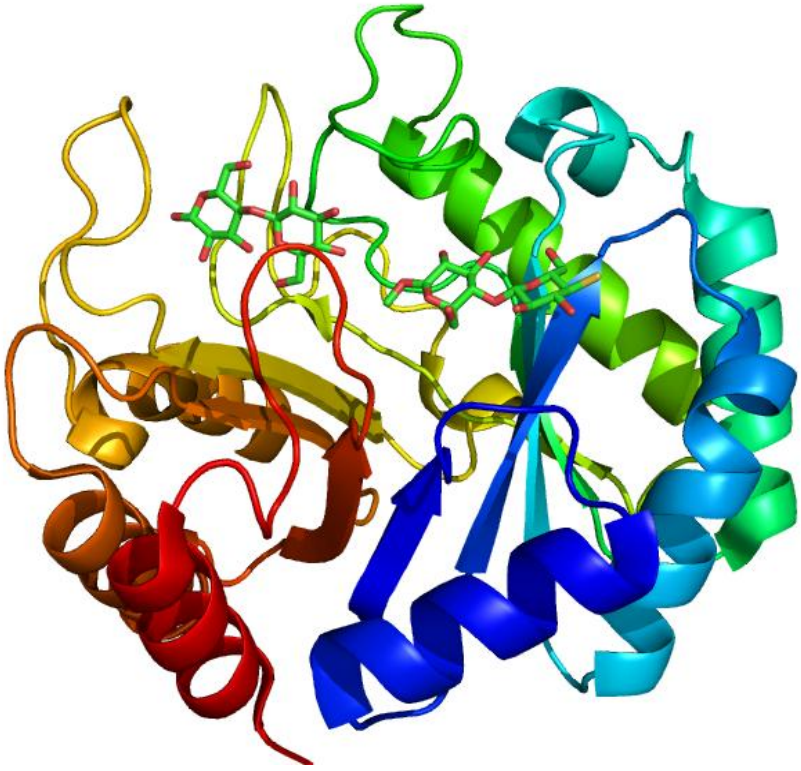


Figure 3.6 Cartoon showing the $(\alpha/\beta)_8$ fold of *Ct*Lic26A
Two molecules of the thiopentasaccharide are bound. The secondary elements are ramped from blue (N-terminus) to red (C-terminus).

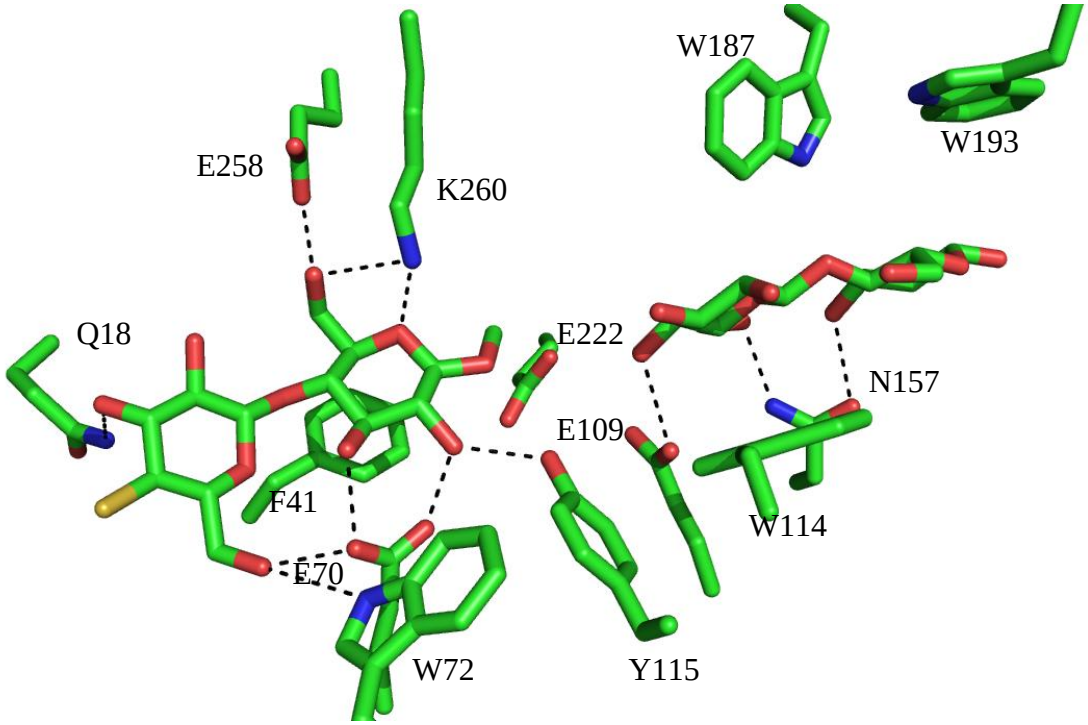


Figure 3.7 Representation of the active site interactions made by the two molecules of the thiopentasacccharride.
The +1 and +2 subsites are occupied by the non-reducing end of one molecule while -2 and -1 are bound to the reducing end of another molecule.

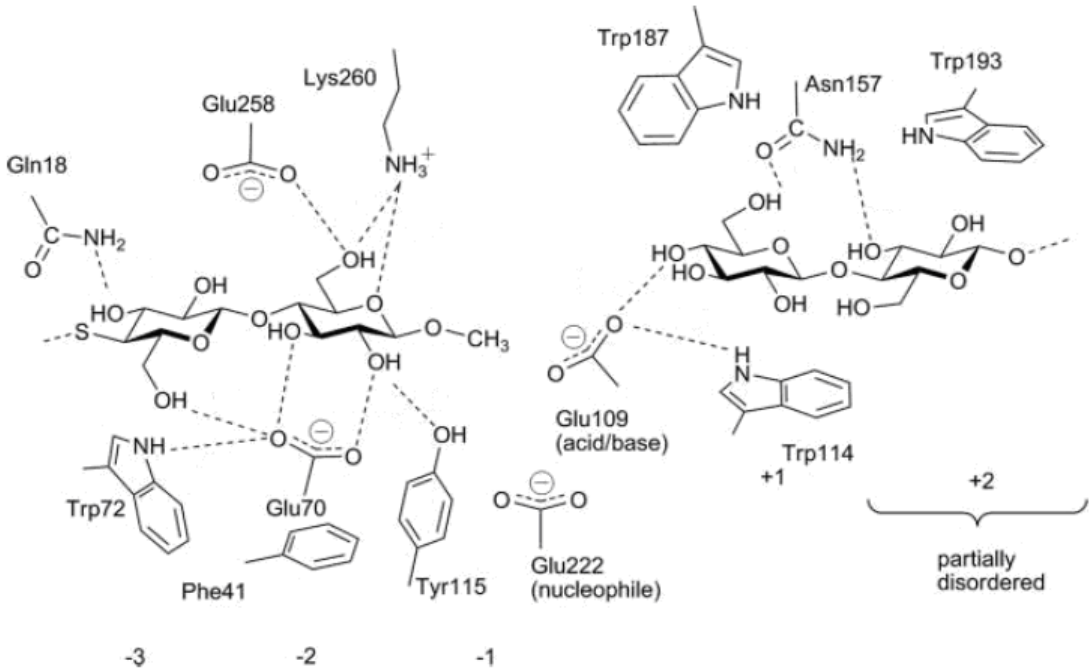


Figure 3.8 A schematic diagram detailing interactions *Ct*Lic26A makes with two molecules of the thiopentasacchride.
Reproduced from Money et al (Money et al. 2008)

This makes -2 the most extensive subsite (Figure 3.7 and 3.8). At -3 Q18 via Nε2 and E70 through Oε1 make polar contacts with O3 and O6 of the glucose, respectively. (Figures 3.7 and 3.8)

Based on the structural data provided of *Ct*Lic26A bound to the thiopentasacharride the above residues were mutated to alanine (**mutagenesis was performed by Catarina I.P.D. Guerreiro from the lab of Professor Carlos M.G.A Fontes**) and the mutant enzymes were subjected to kinetic analysis.

3.4.5 Kinetic analysis of Mutants

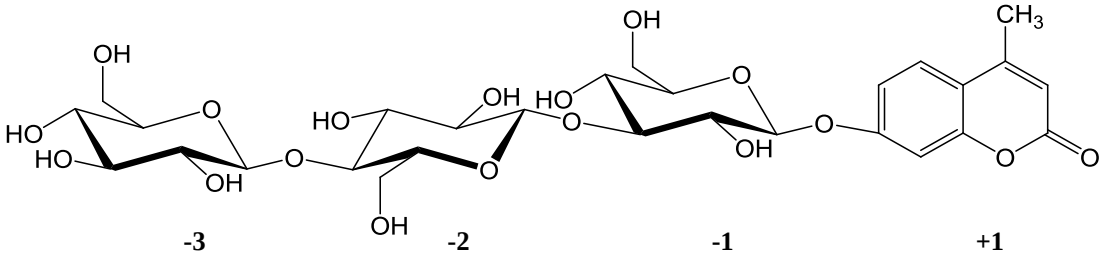


Figure 3.9 Schematic of the Aryl-glycoside substrate (β1,4-Glc- β1,3-Glc-MU) used in the kinetic analysis of *Ct*Lic26A mutants

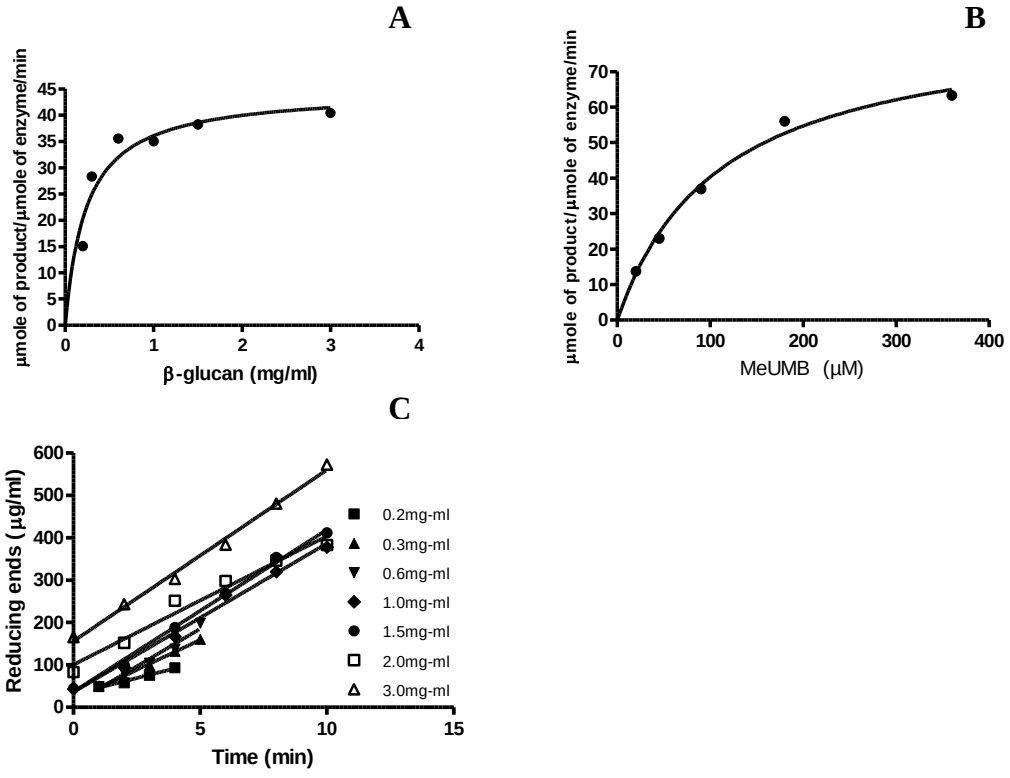


Figure 3.230 Graphs showing non-linear and linear regression.
A-an example of *Ct*Lic26A acting on β-glucan were reaction products were quantified by the DNSA method (Miller 1959). B-an example of *Ct*Lic26A hydrolysing Aryl-glycoside (β1,4-Glc- β1,3-Glc-MU), performed using fluorescence spectroscopy. C-the linear rates at each concentration of substrate used to plot the data displayed in A. Graphs were plotted in GraphPad Prism 4.02.

For aryl-glycoside analysis a trisaccharide linked to methylumbelliferryl (MeUMB) that spanned the active site subsites +1 to -3, with methylumbelliferryl lying at +1, was used (Figure 3.9). Cleavage of the methylumbelliferryl was monitored fluorescently with an excitation and emission wavelength of 365 nm and 440 nm, respectively. Polysaccharide analysis was done using barley β-glucan (high viscosity). Cleavage was monitored by the reducing sugar assay of Miller (Miller 1959), which utilises the ability of the aldehyde/ketone group of sugars to reduce 3,5-dintrosalcyclic acid (DNSA) to 3-amino-5-nitrosylicylate. This causes a colour change that can be monitored at a wavelength of 575 nm. All kinetic data described is presented in Table 3.1. Examples of the data obtained are shown in Figure 3.10 for the wild type enzyme, and all kinetic parameters were determined by non-linear regression plotting rate against substrate concentration.

3.4.5.1 +2 subsite

The N157A mutation caused similar reductions in K_M and k_{cat}, (~10 fold), against the oligosaccharide resulting in an almost wild type k_{cat}/K_M, suggesting a possible switch in the rate limiting step. This effect was not seen against the polysaccharide substrate where both K_M and k_{cat} were near wild type values.

3.4.5.2 +1 Subsite

The W114A mutation disrupted the hydrophobic platform at the +1 subsite and greatly reduced the catalytic efficiency of the enzyme against the aryl-glycoside substrate. Individual parameters of k_{cat} and K_M could not be obtained due to a linear increase in rate observed up to a substrate concentration of 2 mM, which was the maximum deployed. This suggests that a large increase in K_M (>20 fold) contributes to the substantial decrease in k_{cat}/K_M. Against the polysaccharide substrate no significant decrease in k_{cat} and K_M was observed, and there was only a 2 fold reduction in k_{cat}/K_M.

3.4.5.3 -2 Subsite

The mutation of F41 at the -2 subsite had little effect on K_M but caused a large decrease in k_{cat} against both substrates used. Against the aryl-glycoside and the polysaccharide the k_{cat} was reduced ~23000 fold and ~40 fold respectively. This

phenotype was also observed for the mutations E70A, Y115A and E258A but with smaller effects on k_{cat}.

The W72A mutant showed a significant reduction in k_{cat}/K_M against the aryl-glycoside. Once again individual values could not be obtained as the reaction rate was linear up to 2 mM. This would suggest that the elevation of K_M plays a significant role in the reduction of k_{cat}/K_M. Against the polysaccharide this mutation caused a modest increase in the catalytic efficiency due to a slight increase in k_{cat}.

The K260A mutant showed a significant elevation in K_M and decrease in k_{cat}. The effect on the polysaccharide was a smaller increase in K_M but a slightly greater decrease in k_{cat}. This caused significant reductions in the k_{cat}/K_M for both substrates.

3.4.5.4 -3 Subsite

The mutation of Q18, which makes interactions solely with the glucose residue at -3, showed an increase in K_M and a decrease in k_{cat} for the aryl-glycoside. This led to a 170 fold reduction in k_{cat}/K_M compared to wild type. Against the polysaccharide the effects were once again more modest with a ~2 fold increase in K_M and a ~3 fold decrease in k_{cat}. The modest effect on the k_{cat}/K_M for the polysaccharide was a ~6 fold reduction.

The results suggest a key role for the residues housed at the -2 subsite with the majority of the most detrimental mutations residing here. Mutagenesis in the positive subsites shows little effect against polysaccharides. W72 and W114 are important for aryl-glycoside substrates, while N157 may have an important role in deglycosylation. For all mutations the effects were much more pronounced against the aryl-glycoside substrate than the polysaccharide and, in one case, W72A, a slight increase in k_{cat}/K_M was observed.

	Aryl-Glycoside (β1,4-Glc- β1,3-Glc-MU)			β-Glucan		
Enzyme	**K_M (μM)**	**k_{cat} (min^{-1})**	**k_{cat}/K_M (min^{-1} M^{-1})**	**K_M (mg ml^{-1})**	**k_{cat} (min^{-1})**	**k_{cat}/K_M (min^{-1} ml mg^{-1})**
Wt	110 ± 18	23 ± 1.5	$2.1 \times 10^5 \pm 4.6 \times 10^4$	0.34 ± 0.06	$7.4 \times 10^4 \pm 6 \times 10^3$	$2.19 \times 10^5 \pm 1.8 \times 10^4$
Q18A	1763 ± 890	2.1 ± 0.44	$1.2 \times 10^3 \pm 8.4 \times 10^2$	0.67 ± 0.01	$2.3 \times 10^4 \pm 2 \times 10^3$	$3.5 \times 10^4 \pm 3.9 \times 10^3$
F41A	139 ± 23	0.001 ± 0.0001	$7.2 \times 10^0 \pm 1.9 \times 10^0$	1.3 ± 0.36	$1.6 \times 10^3 \pm 1 \times 10^2$	$1 \times 10^3 \pm 3.4 \times 10^2$
E70A	155 ± 18	0.005 ± 0.0003	$3.2 \times 10^1 \pm 5.8 \times 10^0$	1.3 ± 0.13	$8.1 \times 10^3 \pm 2 \times 10^3$	$6 \times 10^3 \pm 2.1 \times 10^3$
W72A	-	-	$6.6 \times 10^2 \pm 0.08$[a]	0.46 ± 0.08	$1.4 \times 10^5 \pm 5 \times 10^3$	$2.94 \times 10^5 \pm 5 \times 10^4$
W114A	-	-	$1.6 \times 10^2 \pm 0.002$[a]	1.4 ± 0.14	$8.5 \times 10^4 \pm 4 \times 10^3$	$6.2 \times 10^4 \pm 9.3 \times 10^3$
Y115A	260 ± 56	0.05 ± 0.004	$1.7 \times 10^2 \pm 5.1 \times 10^1$	1.0 ± 0.18	$1.5 \times 10^4 \pm 4 \times 10^2$	$1.5 \times 10^3 \pm 3.2 \times 10^2$
N157A	11 ± 2	0.73 ± 0.04	$6.4 \times 10^4 \pm 1.5 \times 10^4$	0.49 ± 0.09	$7.8 \times 10^4 \pm 9 \times 10^3$	$1.58 \times 10^5 \pm 4.9 \times 10^5$
E258A	188 ± 24	1.6 ± 0.06	$8.3 \times 10^3 \pm 1.4 \times 10^3$	0.55 ± 0.07	$2.6 \times 10^4 \pm 1 \times 10^3$	$4.6 \times 10^4 \pm 7.4 \times 10^3$
K260A	1162 ± 232	3.2 ± 0.25	$2.8 \times 10^3 \pm 7.8 \times 10^2$	0.65 ± 0.12	$5 \times 10^3 \pm 1 \times 10^2$	$7 \times 10^3 \pm 1.4 \times 10^3$

Table 3.1 Enzyme Kinetics of mutant and wild type *Ct*Lic26A against aryl-glycoside and polysaccharide substrates.
[a]indicates that the k_{cat}/K_M were estimated from linear regression using a substrate concentration << K_M. Blue symbolises mutations at the +2/+1 subsite boundary, lilac +1, red -2 and orange -3. Errors reported for aryl-glycoside are R^2 values. Errors reported for β-Glucan are standard errors generated from triplicate results.Unless stated data were generated by non-linear regression using the Michaelis-Menten equation in GraphPad Prism 4.02..

3.5 *Cj*Man26C

3.5.1 Expression and Purification

The plasmid pVT1 which, is derived from pET20b and encodes the mannanase *Cj*Man26C (**provided by Dr Evangelos Topokas**), was transformed into *E. coli* strain BL21 (section 2.1.10) and expression of the recombinant protein induced by the addition of 1 mM IPTG when the A_{600nm}= 0.6. Induction of recombinant mannanase synthesis was performed at 16°C (section 2.2.24). IMAC was used to purify *Cj*Man26C (45 kDa) to electrophoretic homogeneity for subsequent kinetic analysis. Where IMAC purification was not sufficient to generate the required level of purity, gel filtration was also performed (Section 2.2.27). An SDS-PAGE gel of both purification steps is shown in Figure 3.11. The concentration of *Cj*Man26C was calculated using the molar extinction coefficient of 98290 at A_{280nm}.

3.5.2 Biochemical characterisation

Recombinant *Cj*Man26C displayed activity against linear mannan, glucomannan and carob galactomannan, in which the mannan backbone is decorated with α-1,6 galactose (Table 3.2 and Figure 3.12). It did not show activity against other plant cell wall polysaccharides. The pH and temperature optima of *Cj*Man26C, using galactomannan as the substrate, were pH 8.0 and 50 °C, respectively. This data was generated by **Dr Evangelos Topokas, working in the laboratory of Professor Harry Gilbert,** whilst performing preliminary experiments on *Cj*Man26C. This study showed *Cj*Man26C hydrolysed the mannose containing polysaccharides galactomannan and glucomannan at a reduced rate compared to the endo-acting β-mannanase *Cj*Man26A (Table 3.2). The k_{cat}/K_M of the two enzymes against mannotriose was similar, while *Cj*Man26C hydrolysed mannotetraose ~210 fold faster than *Cj*Man26A. The k_{cat}/K_M of *Cj*Man26C against mannotetraose was 3.0 x 10^9 min^{-1} M^{-1} which is approaching catalytic perfection, 10^9 s^{-1} M^{-1}, where catalysis is only limited by the rate at which substrate diffuses into the active site in an aqueous environment. Both enzymes demonstrated reduced catalytic rates against insoluble mannan, reflecting the recalcitrance of insoluble polysaccharides to biological attack (Hall et al. 1995); neither enzyme hydrolysed mannobiose. No significant increase in k_{cat}/K_M was seen for *Cj*Man26C against mannopentaose and mannohexaose (Table 3.2), suggesting that *Cj*Man26C has 4 subsites.

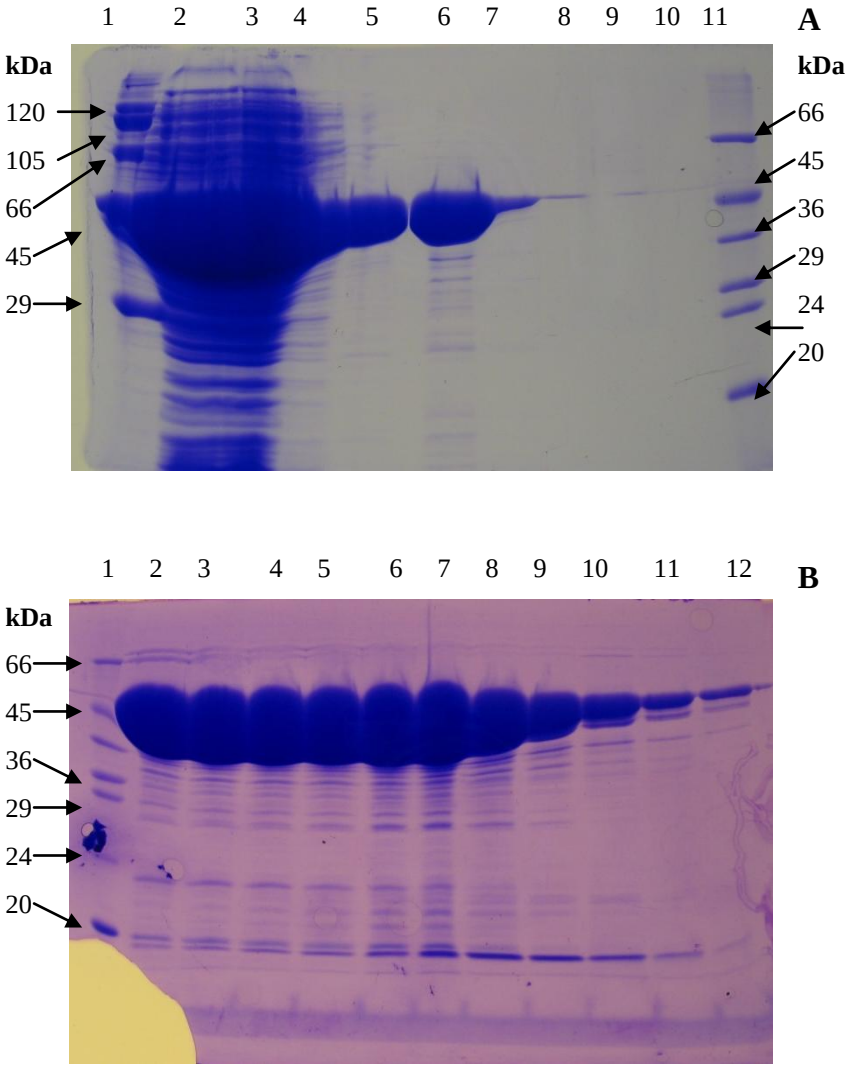


Figure 3.11 Gel A- SDS-PAGE of IMAC purification of *Cj*Man26C.
1, High molecular weight marker; 2, Cell free extract; 3, Flow through; 4, Talon buffer wash; 5, Talon buffer supplemented with 10 mM imidazole; 6-10, Talon buffer supplemented with 100 mM imidazole; 11, Low molecular weight marker. Gel B- SDS-PAGE of a gel filtration purification of *Cj*Man26C. 1, Low molecular weight marker, fractions 2-12 represents 1ml fractions from gel purification.

Substrate [a]	*Cj*Man26C k_{cat}/K_M (min^{-1} M^{-1})	*Cj*Man26A k_{cat}/K_M (min^{-1} M^{-1})
Galactomannan	$1.4 \times 10^4 \pm 0.9$ [b]	$1.3 \times 10^5 \pm 0.8$ [b]
Glucomannan	$1.1 \times 10^4 \pm 0.8$ [b]	$1.6 \times 10^5 \pm 0.6$ [b]
Insoluble Mannan	$2.2 \times 10^4 \pm 0.5$ [b]	$3.5 \times 10^4 \pm 1.0$ [b]
Mannotriose	$2.9 \times 10^6 \pm 2.2 \times 10^5$	$3.65 \times 10^6 \pm 1.4 \times 10^{-3}$
Mannotetraose	$3.0 \times 10^9 \pm 3.3 \times 10^8$	$1.8 \times 10^7 \pm 1 \times 10^{-2}$ [c]
Mannopentaose	$3.5 \times 10^9 \pm 1.2 \times 10^8$	ND
Mannohexaose	$2.7 \times 10^9 \pm 2.2 \times 10^8$	ND

Table 3.2 Catalytic activity of *Cj*Man26C and *Cj*Man26A against mannose polymers. [a]Galactomannan and glucomannan were from carob and konjac seeds respectively. Mannan was made from galactomannan. [b]Activity is specific activity (µmole of product/µmole of enzyme/min). ND denotes activity not determined. [c]Data obtained from Hogg et al (Hogg et al 2001). Substrate and enzyme concentrations are reported in the legend of Figure 3.12. All data was calculated using linear regression analysis.

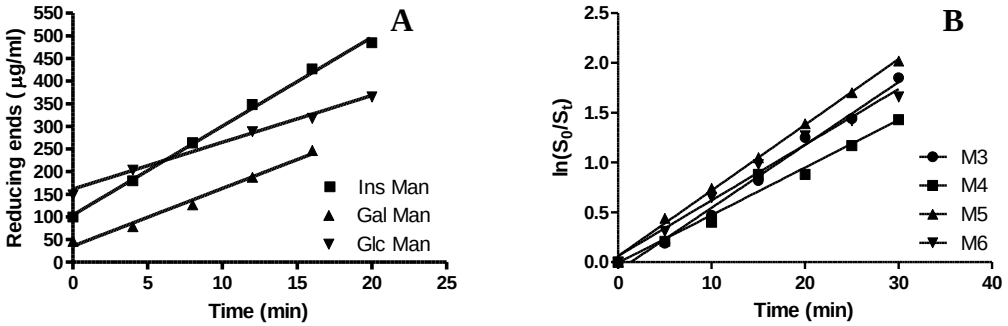


Figure 3.12 Graphs showing linear regression used to calculate values in Table 3.
A-shows 10nM *Cj*Man26C vs 5mg/ml polysaccharide substrates Ins man = Insoluble mannan, Gal Man=Galactomannan, Glc Man=Glucomannan. B-shows the rates calculated for oligosaccharide substrates using the substrate depletion method. M3 –30µM mannotrisoe vs 20nM *Cj*Man26C. M4 30µM mannotetraose vs 20pM *Cj*Man26C. M5- 30µM mannopentaose vs 20pM *Cj*Man26C. M6- M5-30µM mannopentaose vs 20pM *Cj*Man26C. Graphs were plotted in GraphPad Prism 4.02 using linear regression.

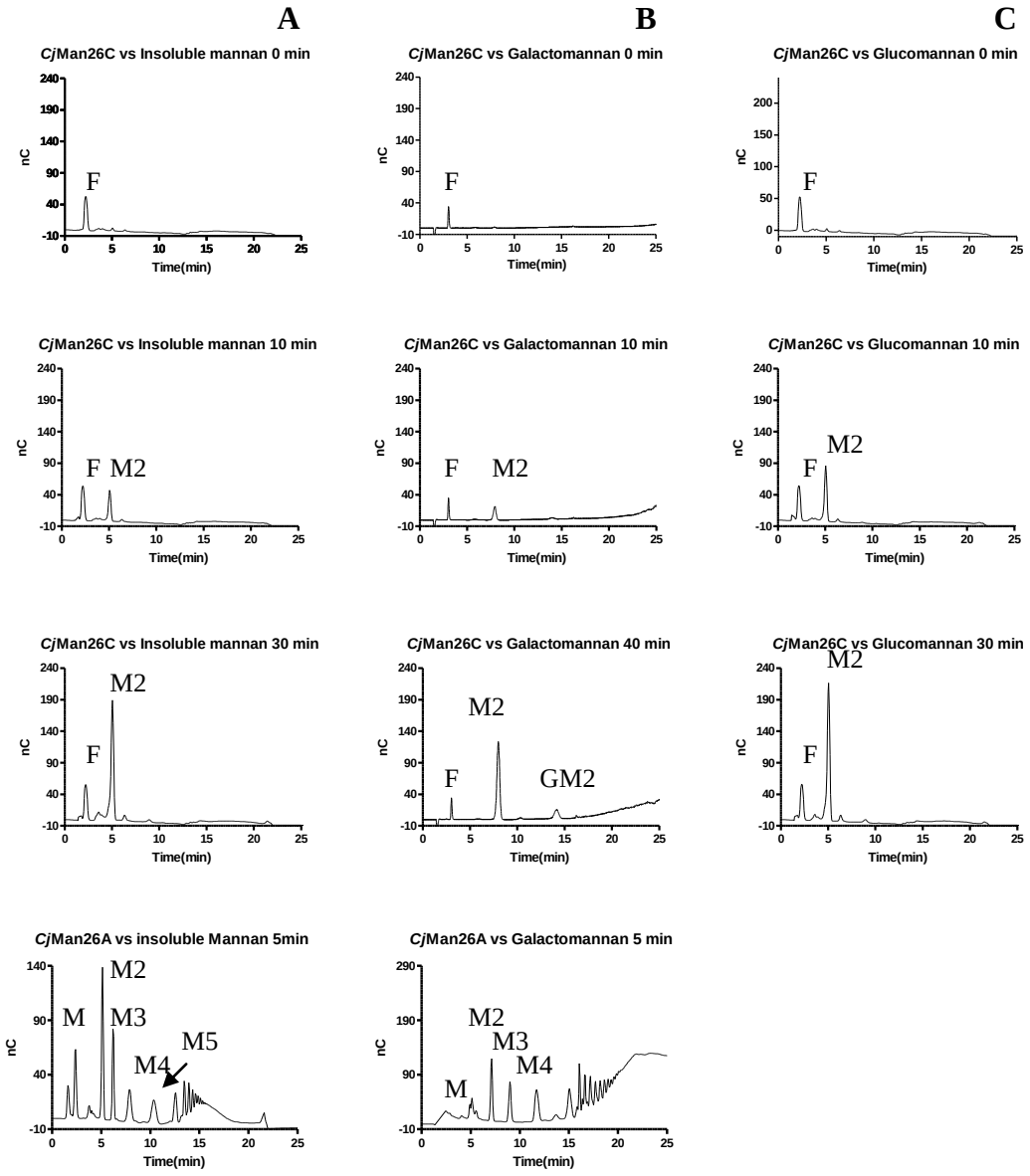


Figure 3.13 Digestion profiles of *Cj*Man26C and *Cj*Man26A versus mannose polysaccharides
Reaction were performed at 37 °C in 50 mM sodium phosphate buffer pH 7.0. Panel A shows the digestion of Insoluble mannan (2 mg/ml) by *Cj*Man26C (5 nM) with the bottom chromatogram showing the digestion of insoluble mannan (5 mg/ml) by *Cj*Man26A (10 nM). Panel B shows the digestion of galactomannan (2 mg/ml) by *Cj*Man26C (10 nM) with the bottom chromatogram showing the digestion of galactomannan (2 mg/ml) by *Cj*Man26A (5 nM). Panel C shows the digestion of glucomannan (2 mg/ml) by *Cj*Man26C (10 nM).M = Mannose, M2 = Mannobiose, M3 = Mannotrisoe, M4 = Mannotetraose, M5 = Mannopentaose GM2 = glactomannobiose F = Fucose (internal standard). Reaction products were identified by comparing their elution times with mannooligosaccharides standards. On the y axis nC denotes nanocoloumbs.

These data show *Cj*Man26C is highly active against oligosaccharide substrates, but less active against polysaccharide substrates when compared to the endo-acting β-mannanase *Cj*Man26A from the same family. The product profile generated by *Cj*Man26C against the polysaccharides insoluble mannan, galactomannan and glucomannan and against the oligosaccharides mannotriose, mannoteatraose, mannopentaose and mannohexaose were analysed by HPLC (Figures

3.13 and Figure 3.14). Only mannobiose was liberated from the polysaccharides mannan and glucomannan, while from galactomannan mannobiose and galactomannobiose (galactose α-1,6 linked to the reducing end of mannobiose) were released (Figure 3.13). It is evident that *Cj*Man26C is an exo-acting mannanase releasing mannobiose or galactomannobiose, which from crystal structure data presented later, was shown to be derived from the non-reducing end. Comparison of the product profiles of *Cj*Man26C with those of *Cj*Man26A (the two chromatograms at the bottom of Panels A and B in Figure 3.13) highlight the exo mode of action by the enzyme. *Cj*Man26A, as shown previously (Hogg et al. 2001) operates via a classic endo-mode of action, generating an array of products. This is in stark contrast to the single product released by *Cj*Man26C from insoluble mannan and glucomannan and the two products released from galactomannan due to the decorated nature of the polysaccharide. Mannotriose is hydrolysed to mannobiose and mannose, mannotetraose is converted exclusively to mannobiose, mannotriose and mannobiose are generated from mannopentaose and finally mannohexaose (which is not 100 % pure) is cleaved predominantly to mannotetraose and mannobiose (Figure 3.14). These data are entirely consistent with the exo-mode of action of *Cj*Man26C in which a disaccharide is released. The results also show that the +2 subsite from *Cj*Man26C and *Cj*Man26A contributes -4.3 kcal mol^{-1} and -1.0 kcal mol^{-1} to catalysis, respectively, accounting for the enhanced rate of the exo-acting mannanase against mannotetraose (Table 3.2). The observation that *Cj*Man26C liberates only mannobiose from glucomannan shows that only mannose can be accommodated at the -1 and -2 subsites.

The accommodation of galactose in the active site was investigated further using the substrate 6^3,6^4-Gal_2Man_5. The substrate was hydrolysed to 50% using a GH27 α-galactosidase. Assuming galactose cleavage is random a mixture that contained ~50% 6^3,6^4-Gal_2Man_5 ~25% Mannopentaose, ~12.5% 6^3-Gal-Man_5 and ~12.5% 6^4-Gal-Man_5 would be generated. Treatment of this mixture shows the *Cj*Man26C can hydrolyse 6^4-Gal-Man_5 to 6^2-Gal-Man_2 and mannotriose. No 6^3-Gal-Man_3 is produced from 6^3-Gal-Man_5 (Figure 3.15). This shows that galactose can be accommodated at the -1 but not the +1 subsite, if galactose could be accommodated at the +1 subsite 6^3-Gal-Man_3 would be generated.

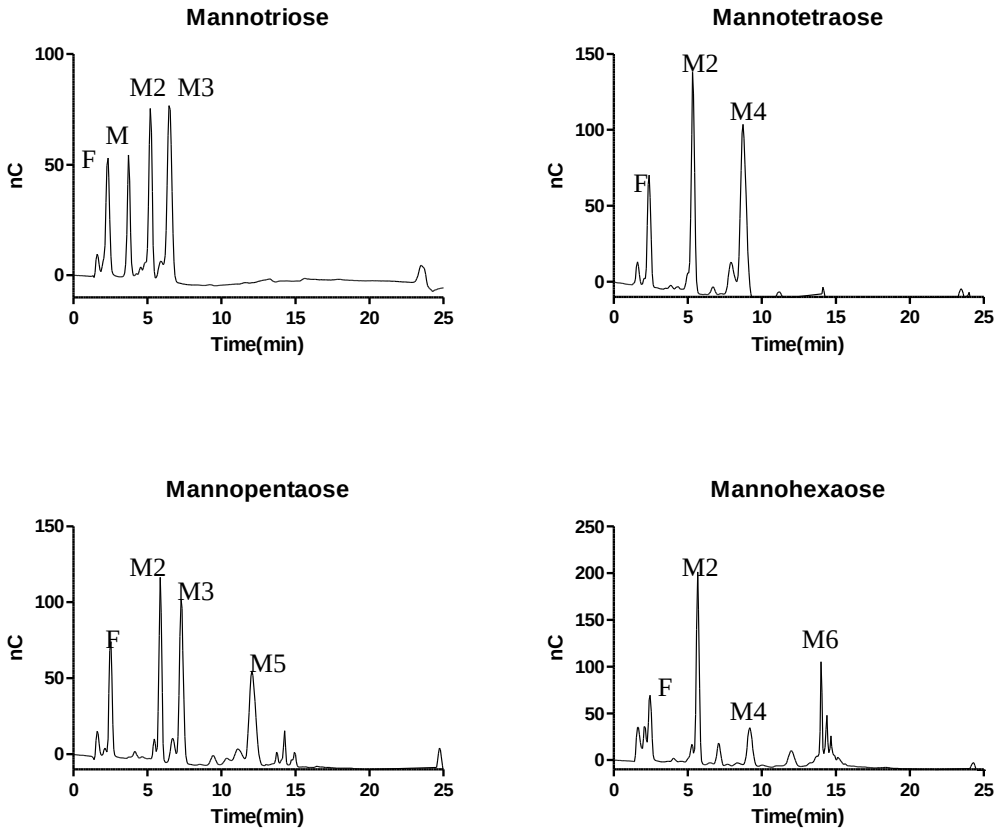


Figure 3.14 Digestion profiles of *Cj*Man26C versus mannose oligosaccharides.
M = Mannose, M2 = Mannobiose, M3 = Mannotriose, M4 = Mannotetraose, M5 = Mannopentaose, M6 = Mannohexaose and F = Fucose (Internal standard).

Digestion profiles where $6^3$$6^4$-$Gal_2Man_5$ was hydrolysed to ~50% with a GH27 then treated with *Cj*Man26C

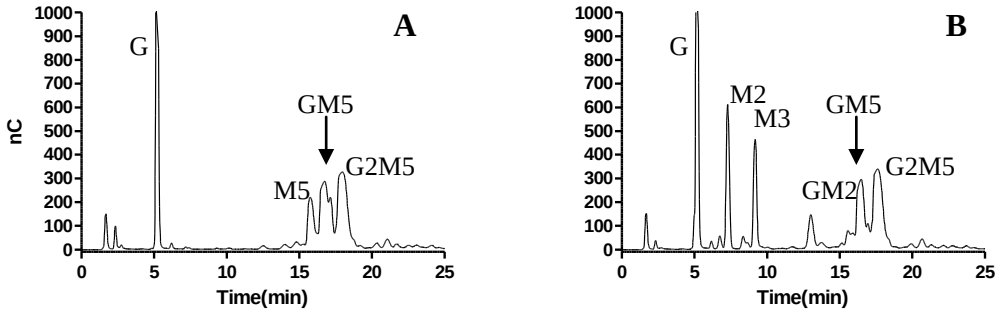


Figure 3.15 Digestion profiles detailing the accommodation of galactose by *Cj*Man26C.
A-shows the 50% digestion mixture of $6^3$$6^4$-$Gal_2Man_5$. B-shows the mixture after incubation with *Cj*Man26C. G2M5 = $6^3$$6^4$-$Gla_2Man_5$, GM5 = 6^3Gal-Man_5/6^2Gal-Man_5, M5 = Mannopentaose, GM2 = 6^2-Gal-Man_2, M2 = Mannobiose, G = Galactose.

3.5.3 Crystallisation and structure solution

The structure of *Cj*Man26C was solved in four forms. A native structure with glycerol bound at -2, a mannose complex with mannose at -2, a E338A nucleophile mutant in complex with mannobiose occupying the -1 and -2 subsites and finally a complex of the wild type enzyme with 6^3-Gal-Man_4, which spans +2 to -2.

All crystals of enzyme in complex with ligand grew in the same condition; 15% PEG 3000, 100 mM sodium citrate pH 5.6 supplemented with 10 mM ligand. The 6^3-Gal-Man_4 *Cj*Man26C complex was crystallised in the presence of 6^1-Gal-Man_2 (purchased from Megazyme International) and either represents a trace contaminant; a hexasaccharide contaminant was observed by mass spectrometry (**personal communication with Dr Tibor Nagy**), or a novel transglycosylation product. The native crystals of apo *Cj*Man26C were grown in 20 % PEG 3350 and 200 mM NaCl. Glycerol (20 % v/v) was used as a cryoprotectant and all cryoprotectants contained 10 mM of the appropriate ligand where necessary.

Data collection, structure solution and refinement were performed by **Dr Valerie M-A Ducros and Dr Micheal Suits from the laboratory of Professor Gideon J Davies**. Data were collected at the European synchrotron radiation facility using beamlines ID14-I and ID23-I. The four *Cj*Man26C structures were solved by molecular replacement using *Cj*Man26A as the search model. The data collection and refinement statistics are shown in Table 3.3.

	Apo-*Cj*Man26C	***Cj*Man26C-mannose**	***Cj*Man26C-mannobiose**	***Cj*Man26C-6^3-Gal-Man$_4$**
Data collection				
Space group	$P2_12_12_1$	$P6_122$	$P6_122$	$P6_122$
Cell dimensions (Å) (°)	a = 52.38, b = 83.6, c = 84.3 α = β = 90° γ = 120°	a = 84.7, b = 84.7, c = 244.7 α = β = 90° γ = 120°	A = 84.3, b = 84.3, c = 244.3 α = β = 90° γ = 120°	a = 84.5, b = 84.5, c = 244.3 α = β = 90° γ = 120°
Resolution (Å)	20-1.7 (1.76-1.7)	50-1.47 (1.52-1.47)	20-1.80 (1.86-1.80)	20-1.57 (1.63-1.57)
$I/\sigma I$	19.4 (3.7)	47.4 (5.3)	16.2 (2.9)	38.4 (10.4)
R_{merge}(%)	8.3 (48.6)	5.8 (28.7)	8.7 (41.1)	6.2 (24.6)
Completeness (%)	100 (100)	99.1 (99.7)	99.9 (100)	99.4 (99.3)
Redundancy	7.0 (7.0)	9.1 (7.0)	9.5 (8.8)	11.3 (11.0)
Refinement				
R_{work}/R_{free}	16/19	13/16	16/18	12/15
RMSD bond lengths (Å)	0.004	0.012	0.015	0.009
RMSD bond angles (°)	0.6	0.9	1.3	0.9

Table 3.3 Data collection and refinement statistics for *Cj*Man26C.
Data collected by Dr VA Ducros and Dr M Suits.

3.5.4 Three Dimensional Structure

The 3-D structure of *Cj*Man26C displays the canonical $(\beta/\alpha)_8$ barrel topology as expected of a GH26 and Clan GH-A member (Figure 3.16). The structure of *Cj*Man26C shows a high degree of structural similarity, as may be expected, to its homolog *Cj*Man26A. The two structures overlay with a root mean square deviation of 1.46 Å for 311 equivalent C_α atoms using SSM (Krissinel et al. 2004), and the two enzymes display 33% sequence identity. In Figure 3.16A and B the overall fold of *Cj*Man26C is shown as cartoon a representation, of the native enzyme with glycerol and mannose bound at the -2 subsite, respectively. Both molecules make multiple polar interactions with D130, which plays a key role in the mode of action of the enzyme.

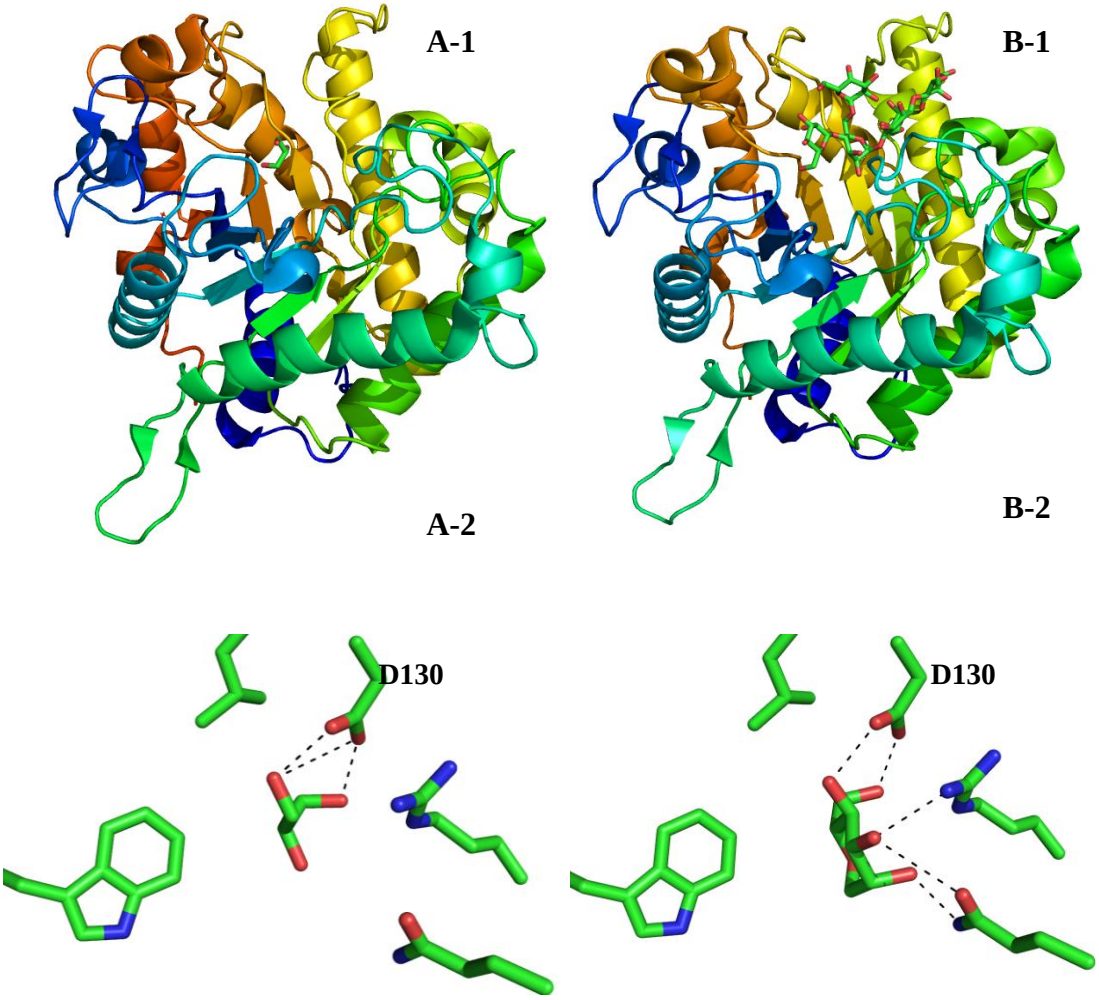


Figure 3.16 Cartoon representations of *Cj*Man26C and the -2 subsite.
A-1,B-1- Structure fold of *Cj*Man26C; A-2 glycerol at the -2 subsite B2- mannose at the -2 subsite. Colour ramped N-terminus blue to C-terminus red.

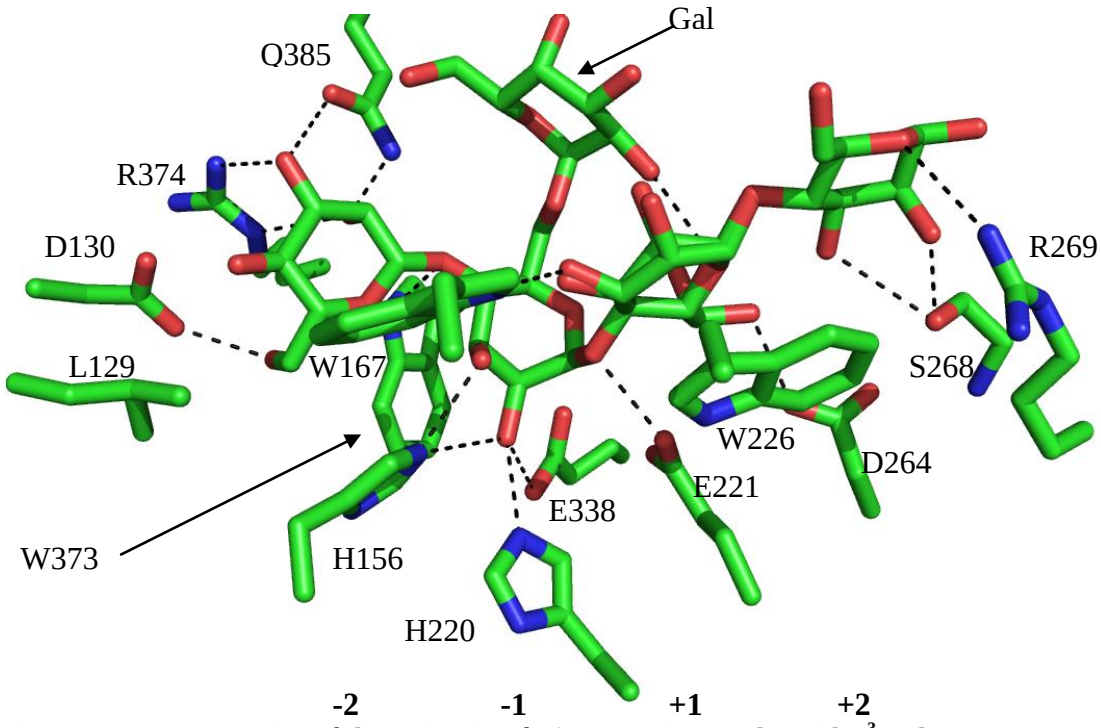


Figure 3.17 Representation of the active site of *Cj*Man26C in complex with 6^3-Gal-Man_4.
Gal signifies the α linked galactose residue.

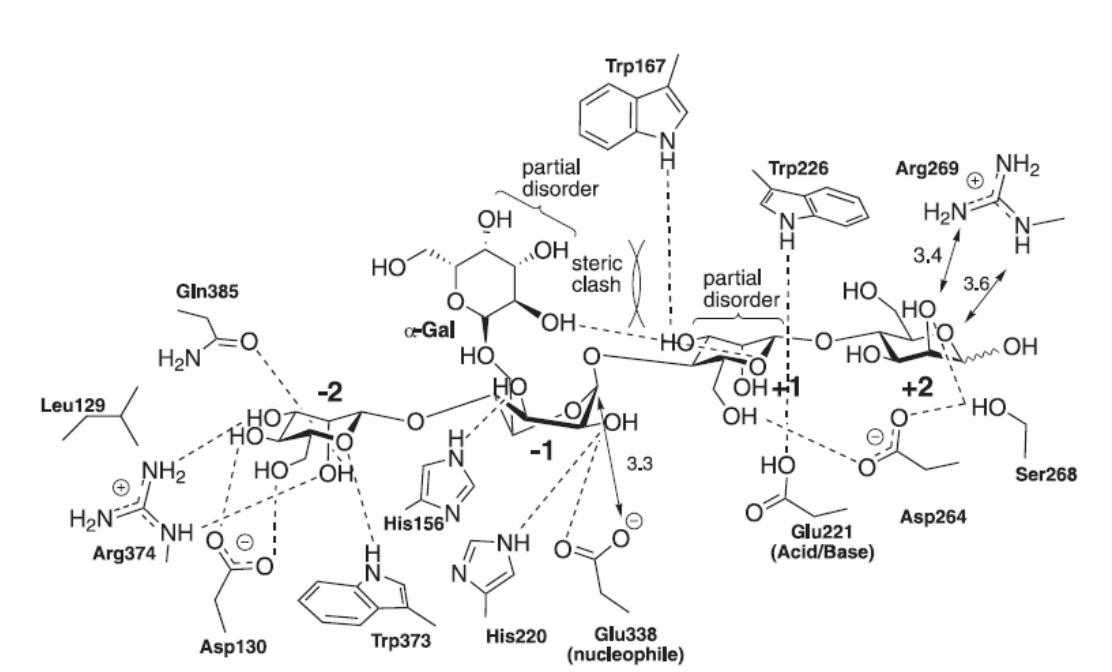


Figure 3.18 schematic representation of the active site of *Cj*Man26C in complex with 6^3-Gal-Man_4,.

Structural and sequence alignments of *Cj*Man26C with other members of GH26, and more widely within Clan GH-A allows the unambiguous assignment of the catalytic residues. *Cj*Man26C is predicted to operate via a double displacement mechanism (All Clan GH-A enzymes display this mechanism) requiring a

nucleophile, E338, and an acid/base, E221. Mutation of these residues caused a >5 x 10^6 fold decrease in catalytic activity (Table 3.4). A description of the interactions between the protein and ligand is as follows:

3.5.4.1 +2/+1 Subsites

At the +2 subsite R269 makes long range polar contacts with the endocyclic ring oxygen and O2 via Nη1 and Nε respectively. S268 forms polar contacts with O2 and O3. At the +1 subsite W226 provides an aromatic platform against which the mannose residue is able to stack. D264 makes a hydrogen bond via Oδ2 with the hydroxymethyl group of C6, while the Nε1 of W167 forms a hydrogen bond with O3. There is partial disorder at the +1 subsite due to steric clashes between O2 and the C2-C3 edge of the galactose residue linked to the -1 mannose (Figures 3.17 and 3.18).

3.5.4.2 -1 subsite

The -1 subsite is of particular interest as this is the site of catalysis, and the interactions can be described for both the pre-transition state $^1S_5$ skew boat conformation, displayed by the 6^1-Gal-Man$_4$, and the ground state $^4C_1$ chair conformation, displayed by mannobiose (Figure 3.19). In the relaxed $^4C_1$ chair Y297 interacts with the β O1 of the mannose residue at -1. E221, the catalytic acid/base, forms a hydrogen bond via Oε2 with O1 and O2 as one may expect. O1 would be involved in the glycosidic linkage and thus requires protonation during catalysis. H156 forms a hydrogen bond via Nε2 with O3 and lastly, W373 provides a hydrophobic platform and also forms a hydrogen bond through Nε1 with the glycosidic oxygen between the mannose residues at the -1 and -2 subsites. In the $^1S_5$ skew boat conformation O1 adopts a pseudo-axial conformation and the interaction with Y297 is lost, but the interaction, unsurprisingly, with the acid/base E221, is retained. O2 adopts a pseudo-equatorial conformation and its interaction with the acid/base is lost, however it now forms interactions with the nucleophile E338, via Oε2, and Nε2 of H220 O3 becomes pseudo-axial but still within hydrogen bond distance of H156. Significantly, H220 is conserved in all GH26 family members. No other differences are observed between the two ring conformations for the -1 mannose residue.

3.5.4.3 *-2 Subsite*

The most extensive interactions are made at the -2 subsite. D130 hydrogen bonds to O4 and O6 via Oδ1 and Oδ2. W373 may form a hydrogen bond with O2. R374 hydrogen bonds to O2 and O3 via Nη2 and Nε respectively. The interactions at the -2 subsite are completed by Q385, which also hydrogen bonds to O2 and O3 via Nε1 and Oε1, respectively.

Selection for the axial orientation of mannose at O2, this being the unique feature of mannose compared to other sugars in the plant cell wall, is made at +2 by R269 and at -2 by R374 and Q385. No selection for the axial orientation of O2 is made at the -1 and +1 subsites, raising issues regarding the mechanistic basis for the mannanase activity displayed by the enzyme, which is addressed in the discussion section (Figures 3.17 and 3.18).

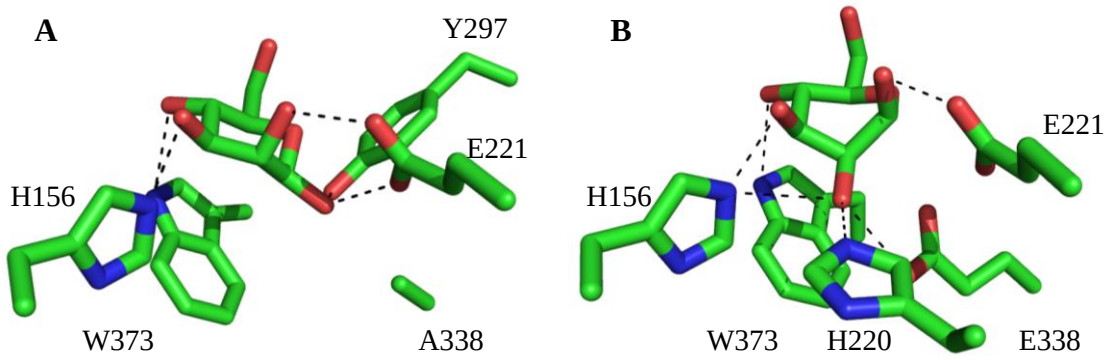


Figure 3.19 Sticks representation of the -1 subsite interactions.
A- Interactions of the -1 mannose residue in the relaxed $^{4}C_{1}$ conformation B- Interactions of the -1 mannose residue in the pre transition state $^{1}S_{5}$ skew boat conformation. In Panel A the E338A of *Cj*Man26C is shown.

3.5.4.4 The -2 subsite and its role in exo specificity

The interactions at the -2 subsite are conserved between *Cj*Man26C and *Cj*Man26A. W373, R374 and Q385 in *Cj*Man26C are equivalent to W332, R333 and H349 in *Cj*Man26A. The difference between the two enzymes is D130 in *Cj*Man26C, which interacts with O4 and O6 of the mannose, and is found in loop 3 that connects β-strand 2 to α-helix 3. The equivalent residue in *Cj*Man26A is E121, which interacts only with O6 of the mannose at the -2 subsite. The reason for this difference is a four amino acid insertion in loop 3 of *Cj*Man26C, compared to *Cj*Man26A. This results in a 2 Å shift in the position of D130 in *Cj*Man26C, relative to E121 in *Cj*Man26A, enabling the aspartate to interact with both O4 and O6 of the mannose at this distal

negative subsite. Another consequence of the four residue insertion is the presence of the backbone carbonyl of L129 in *Cj*Man26C which, in conjunction with the interactions made by D130, prevent extension of substrate beyond the -2 subsite. This is shown in Figure 3.20A and B, which display surface representations of the two mannanases with their respective ligands bound. The residues D130 and L129 of *Cj*Man26C are highlighted in blue and yellow. E121 in *Cj*Man26A is also highlighted in blue. The -2 subsite is shown and extension of the substrate beyond -2 is prevented by the occlusion of the cleft. In *Cj*Man26A the ligand has extended beyond the -2 subsite demonstrating the cleft is open. Panel C and D show overlays of both *Cj*Man26A and *Cj*Man26C. In Panel C the ligand 6^3-Gal-Man_4 (complex *Cj*Man26C 6^3-Gal-Man_4) is shown. In Panel D the ligand 2,4 dinitrophenyl 2-deoxy-2-fluoro-β-mannotriose (produced from the pdb 1gvy) is shown. Figure 3.20 Panel E and Panel F show the -2 subsite in further detail L129 and D130 are shown in blue and yellow. It is evident from Figure 3.20 Panel F the steric clash L129 and D130 cause, if a sugar is extended beyond -2 when compared to Figure 3.20 Panel E.

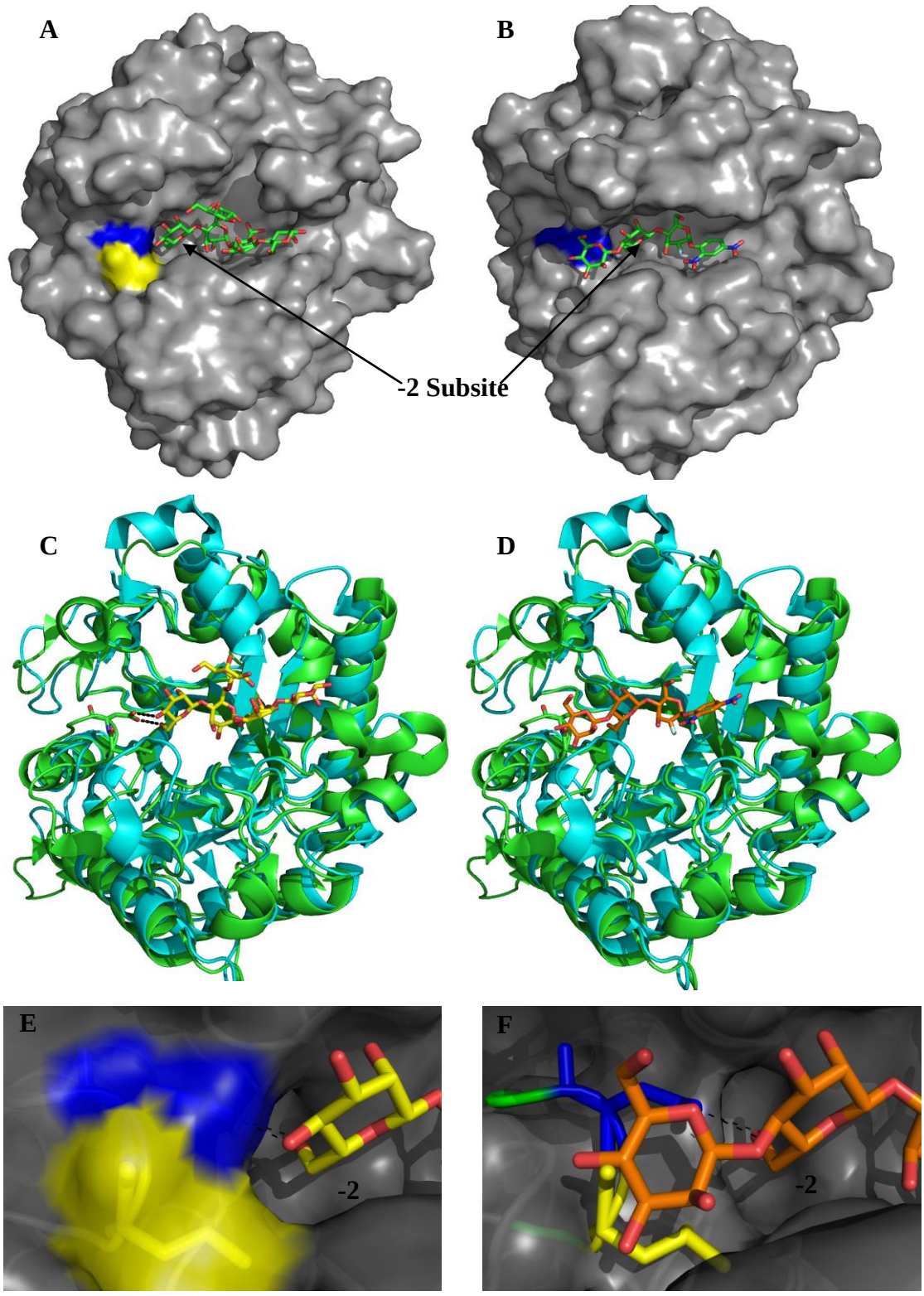


Figure 3.20 Steric hindrance provided by the -2 subsite.
A- Surface representation of *Cj*Man26C in complex with 6^3-Gal-Man_4. L130 is highlighted in yellow and D130 in blue. B- Surface representation of *Cj*Man26A in complex with 2,4 dinitrophenyl 2-deoxy-2-fluoro-β-mannotriose (pdb accession number 1gvy). E121 is highlighted in blue. C- Cartoon overlay of *Cj*Man26C (green) and *Cj*Man26A (pale blue) with 6^3-Gal-Man_4 (Yellow and red). D- Cartoon overlay of *Cj*Man26C (green) and *Cj*Man26A (pale blue) with 2,4 dinitrophenyl 2-deoxy-2-fluoro-β-mannotriose (orange and red). E-Surface of *Cj*Man26C in complex with 6^3-Gal-Man_4 (yellow and red) L129 (yellow), D130 (blue). F- Surface of *Cj*Man26A in complex with 2,4 dinitrophenyl 2-deoxy-2-fluoro-β-mannotriose (orange and red) and loop 3 of *Cj*Man26C insterted as a cartoon with L129 (yellow) D130 (blue) shown as sticks.

The insight provided from the structure of *Cj*Man26C suggested that the four residue extension of loop 3 causes the exo-mode of action. A series of mutations based around this four residue stretch were then introduced into *Cj*Man26C. L129 and D130 were mutated to both glycine and alanine, and the double mutants L129G/D130G and L129A/D130A were also created using the quikchange mutagenesis method (Section 2.1.21.2) (Figure 3.21A). The following loop deletions were also generated, using the IPCR method detailed in section 2.1.21.3: Δ(L129/D130), Δ(N128/L129/D130) and Δ(N128/L129/D130/A131) (Figure 3.21B and C). All primers used are listed in Appendix B Table AB1. All the mutants expressed in soluble form in *E. coli* and were purified to electrophoretic homogeneity as judged by SDS-PAGE. The k_{cat}/K_M was determined for these mutants against 30 µM mannotetraose, a concentration $>K_{M,}$ using the substrate depletion method (Table 3.4).

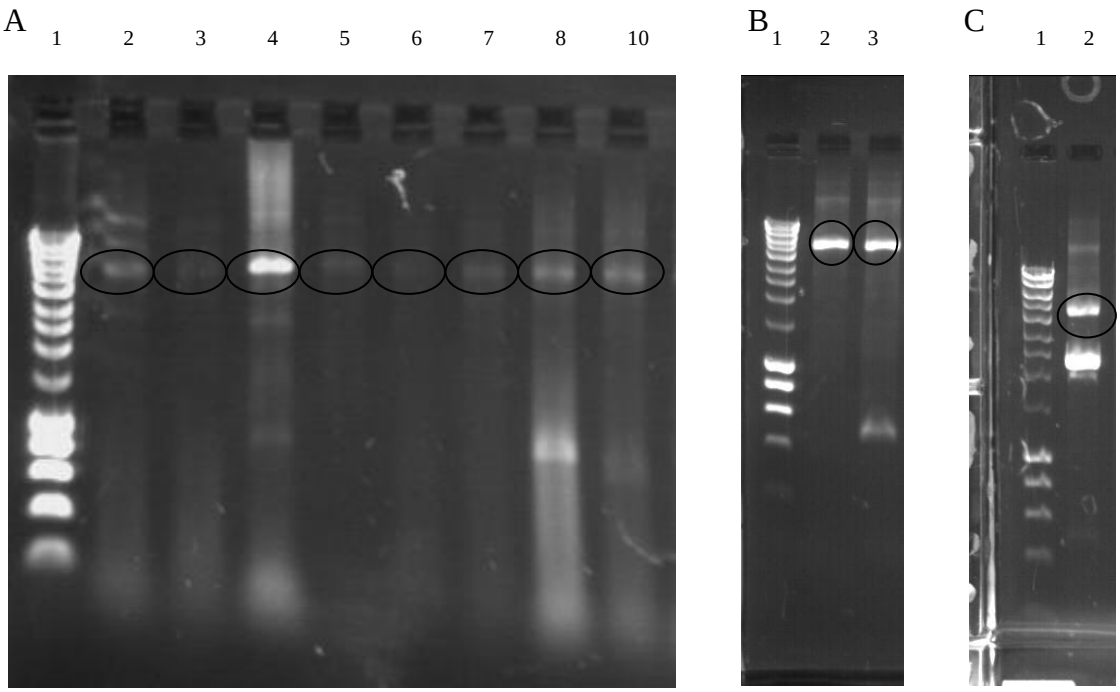


Figure 3.21 Agarose gels of mutant plasmids.
Circles indicate presence of ~6kb PCR product (vector + gene) 5 µl sample + 2 µl loading dye; All samples are after Dpn1 treatment; Gel A shows the agarose gel of the quikchange mutagenesis; 1 = Hyperladder 1 (Invitrogen) 2 = E221A; 3 = E338A; 4 = L106G; 5 = L106A; 6 = D107G; 7 = D107A; 8 = L106G/D107G; 8 = L106A/D107A. Gel B shows a agarose gel of the inverse PCR for loop deletion mutants; 1 = Hyperladder 1 (Invitrogen); 2 = ΔLD; 3 = ΔNLD. Gel C shows the agarose gel of the final loop deletion mutant; 1 = Hyperladder 1 (Invitrogen); 2 = ΔNLDA. Reactions in Gels A and B were produced using the following protocol: 1 min 95°C (1 cycle), 1 min 95°C – 1min 55°C – 6 min 68°C (25 cycles), 6 min 72°C (1 cycle), Held 4°C; reactions were performed using Kod polymerase in 1 x Kod buffer supplemented with 1mM $MgCl_2$. Reaction in Gel C were produced using the following protocol: 1 min 95°C (1 cycle), 1 min 95°C – 30 sec 65°C – 6 min 72°C (25 cycles), 6 min 68°C (1 cycle), Held 4°C; reactions were performed using Kod polymerase in 1 x Kod buffer supplemented with 0.5mM $MgCl_2$

Enzyme	k_{cat}/K_M (min^{-1} M^{-1})
*Cj*Man26C (wild type)	$3.0 \times 10^9 \pm 3.3 \times 10^8$
*Cj*Man26A (wild type)	$1.4 \times 10^7 \pm 2.8 \times 10^5$
E221A[a]	$7.5 \times 10^2 \pm 2.8 \times 10^1$
E338A[a]	$7.7 \times 10^2 \pm 2.6 \times 10^1$
L129G[a]	$4.0 \times 10^6 \pm 3.8 \times 10^5$
L129A[a]	$7.4 \times 10^6 \pm 1.5 \times 10^5$
D130G[a]	$2.7 \times 10^5 \pm 7.4 \times 10^4$
D130A[a]	$2.9 \times 10^4 \pm 3.1 \times 10^3$
L129G/D130G[a]	$8.2 \times 10^7 \pm 9.1 \times 10^6$
L129A/D130A[a]	$4.5 \times 10^4 \pm 4.5 \times 10^3$
ΔL129/D130[a]	$2.7 \times 10^5 \pm 4.5 \times 10^4$
ΔN128/L129/D130[a]	$2.7 \times 10^4 \pm 1.7 \times 10^3$
ΔN128/L129/D130/A131[a]	$3.4 \times 10^4 \pm 6.6 \times 10^2$

Table 3.4 Catalytic activity of *Cj*Man26A, *Cj*Man26C and *Cj*Man26C variants.
Mannotetraose (30 µM) was used as the substrate as this explores all subsites and enzyme concentrations ranging from 20 pM to 10 µM were used. k_{cat}/K_M was calculated using the substrate depletion method (see section 2.3.1.7). [a]indicates mutants are of *Cj*Man26C. All data were generated by non-linear regression.

E221A and E338A display extremely low activity consistent with the key role E221 and E338 play in catalysis, described above. Mutation of L129 to glycine and alanine caused a ~500 and ~1000 fold reduction in k_{cat}/K_M. The same mutations of D130 resulted in an even greater reduction in catalytic efficiency ranging from ~1 x 10^4 to 1 x 10^5 fold. The double mutant L129A/D130A showed a ~5 x 10^4 fold reduction in k_{cat}/K_M compared to wild type *Cj*Man26C, similar to that of the D130A mutation alone. Interestingly the double mutant L129G/D130G retained a significant level of catalytic efficiency, being only ~37 fold below that of the wild type mannanase. All the loop deletion mutations incurred a significant catalytic cost.

The loop mutants were assayed against galactomannan and the product profiles assessed (Figure 3.22), wild type *Cj*Man26C and *Cj*Man26A are shown as reference profiles. While mutation of L129 had no effect on the mode of action of *Cj*Man26C, substitution of D130 had a significant influence on the product profile. D130G, D130A and L129A/D130A displayed an endo-mode of action, producing a wide range of products from galactomannan, comparable to that produced by *Cj*Man26A. Shortening of loop 3 through the creation of the loop deletions ΔL129/D130, ΔN128/L129/D130 and ΔN128/L129/D130/A131 also switched the

mode of action of *Cj*Man26C from exo to endo. The only mutation of D130 that did not appear to alter the mode of action of *Cj*Man26C was L129G/D130G; this mutant retained an exo mode of action (Figure 3.22).

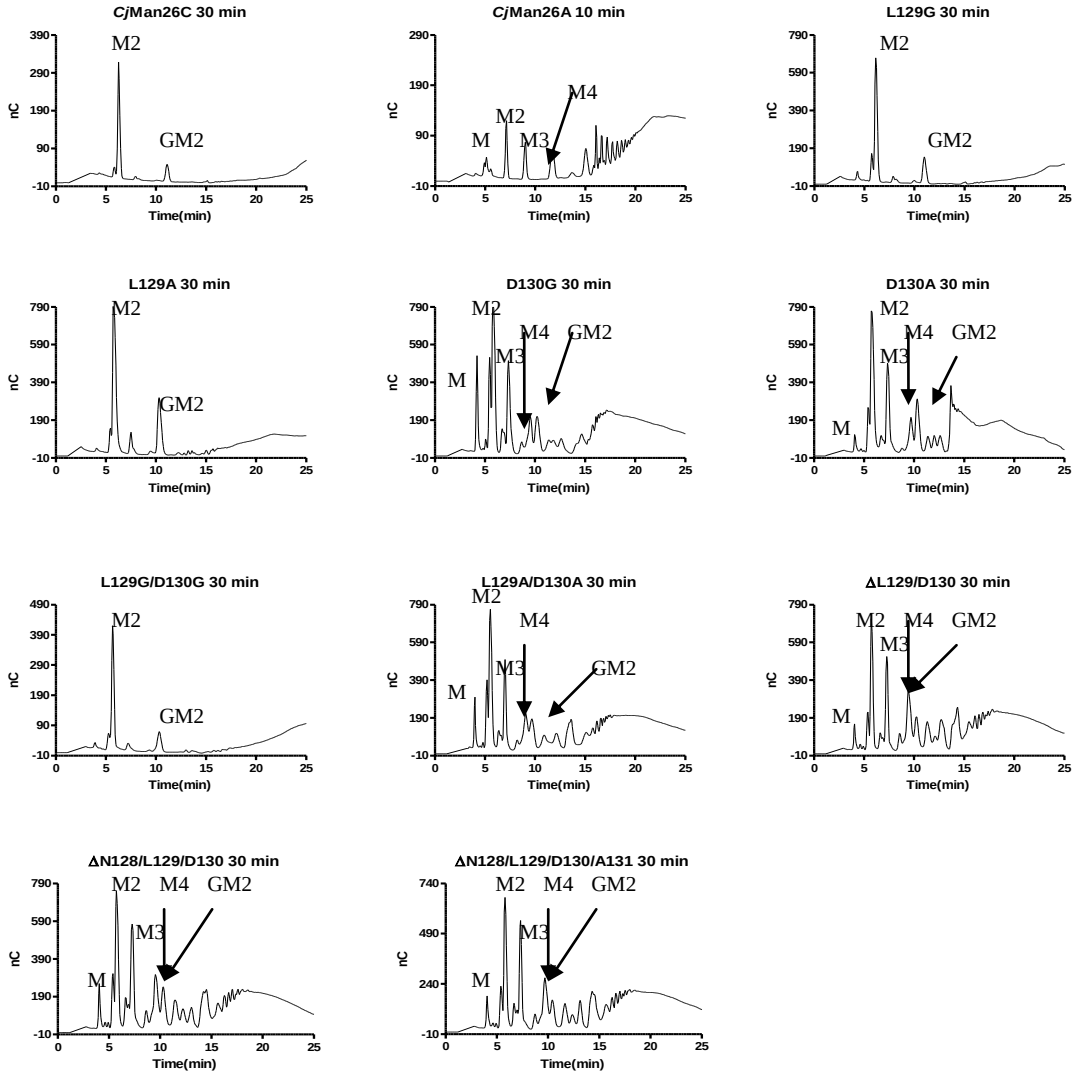


Figure 3.22 Chromatograms showing the reactions product profiles of *Cj*Man26C, *Cj*Man26A and the *Cj*Man26C variants listed in Table 4.
All enzymes (2-200 nM) were incubated with 2 mg/ml galactomannan for 30 minutes in 50 mM sodium phosphate buffer pH 7.0. M = mannose, M2 = mannobiose, M3 = mannotrisoe, M4 = mannotetraose and GM2 = galactomannobiose are indicated.

The extent of the endo character of the mutants can be assessed quantitatively by looking at the initial products of mannohexaose digestion. The endo mannanase *Cj*Man26A has 4 subsites. In the initial stages of mannohexaose hydrolysis the ratio of mannobiose and mannotetraose compared to mannotriose will be 2:1. This is because mannohexaose can occupy 4 subsites in 3 ways, (1) +2 to -2 with a single mannose residue extending beyond both +2 and -2 producing two molecules of mannotriose, (2) +2 to -2 with two mannose residues extending beyond -2 releasing mannobiose and mannotetraose, (3) +2 to –2 with two molecules extending beyond the +2 subsite producing mannobiose and mannotetraose. In contrast *Cj*Man26C, initially, produces only mannobiose and mannotetraose from mannohexaose, meaning mannohexaose must bind +2 to -2 with two mannose residues extending beyond +2 (Figure 3.23).

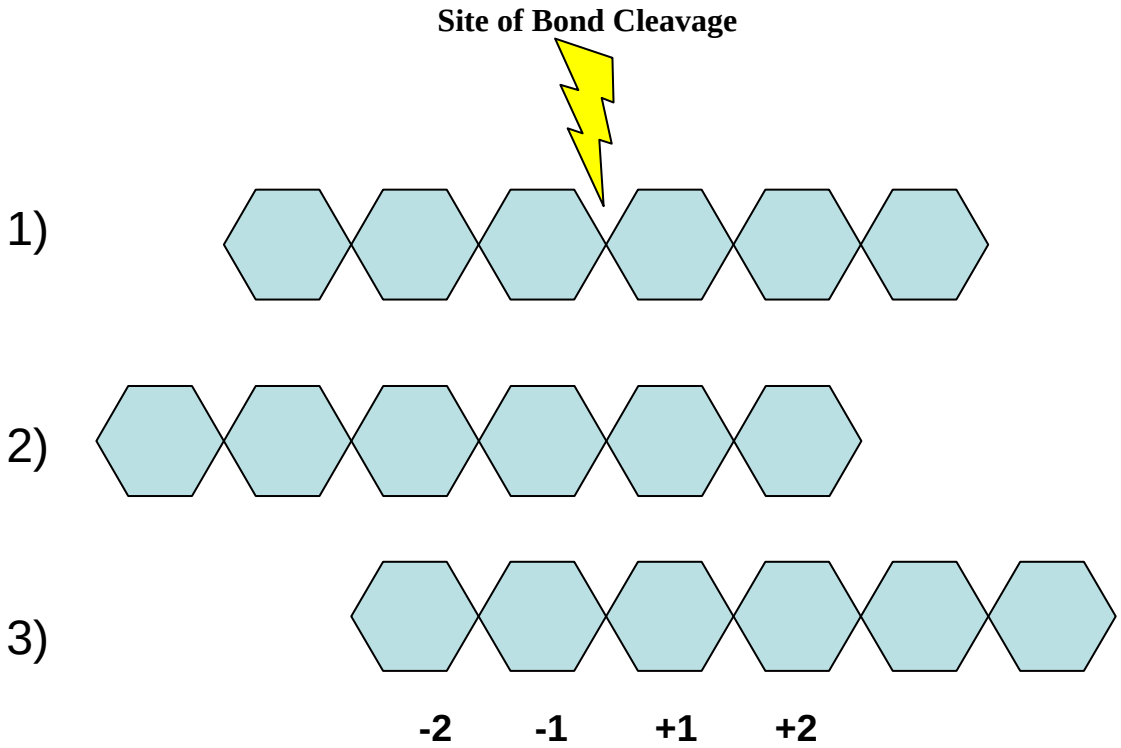


Figure 3.23 Possible binding modes of mannohexaose to an endo mannanase with 4 subsites.

Consistent with the exo activity displayed by L129G, L129A and L129G/D130G against galactomannan, all these *Cj*Man26C variants displayed the exo product ratio of 1:0 M2+M4/M3 when mannohexaose was the substrate (Figure 3.24 and Table 3.5). In contrast ΔN128/L129/D130 and ΔN128/L129/D130/A131 displayed a product ratio against mannohexaose that suggests a complete conversion to an endo mode of action. The other mutants displayed M2+M4/M3 ratios that varied from 3.0:1 to 5.6:1 suggesting that, while these enzymes can exhibit an endo mode of action, a partial steric block distal of the -2 subsite likely remains (Figure 3.24 and Table 3.5).

Enzyme	**Ratio (M2+M4/M3)[b]**
*Cj*Man26C wild type	1:0
*Cj*Man26A	2.2:1
L129G[a]	1:0
L129A[a]	1:0
D130G[a]	3.8:1
D130A[a]	3.0:1
L129G/D130G[a]	1:0
L129A/D130A[a]	5.6:1
ΔL129/D130[a]	4.2:1
ΔN128/L129/D130[a]	2.1:1
ΔN128/L129/D130/A131[a]	2.2:1

Table 3.5 Product ratios of *Cj*Man26A, *Cj*Man26C and *Cj*Man26C variants.
[a]All mutants are of *Cj*Man26C. [b]The ratio of products was measured after 10% hydrolysis using 20 pM - 2μM enzyme vs. 30 μM Mannohexaose. All reactions were performed in 50 mM sodium phosphate buffer pH 7.0. A ratio of 1:0 signifies an exo mode of action, while a ratio of 2:1 or higher signifies an endo mode of action.

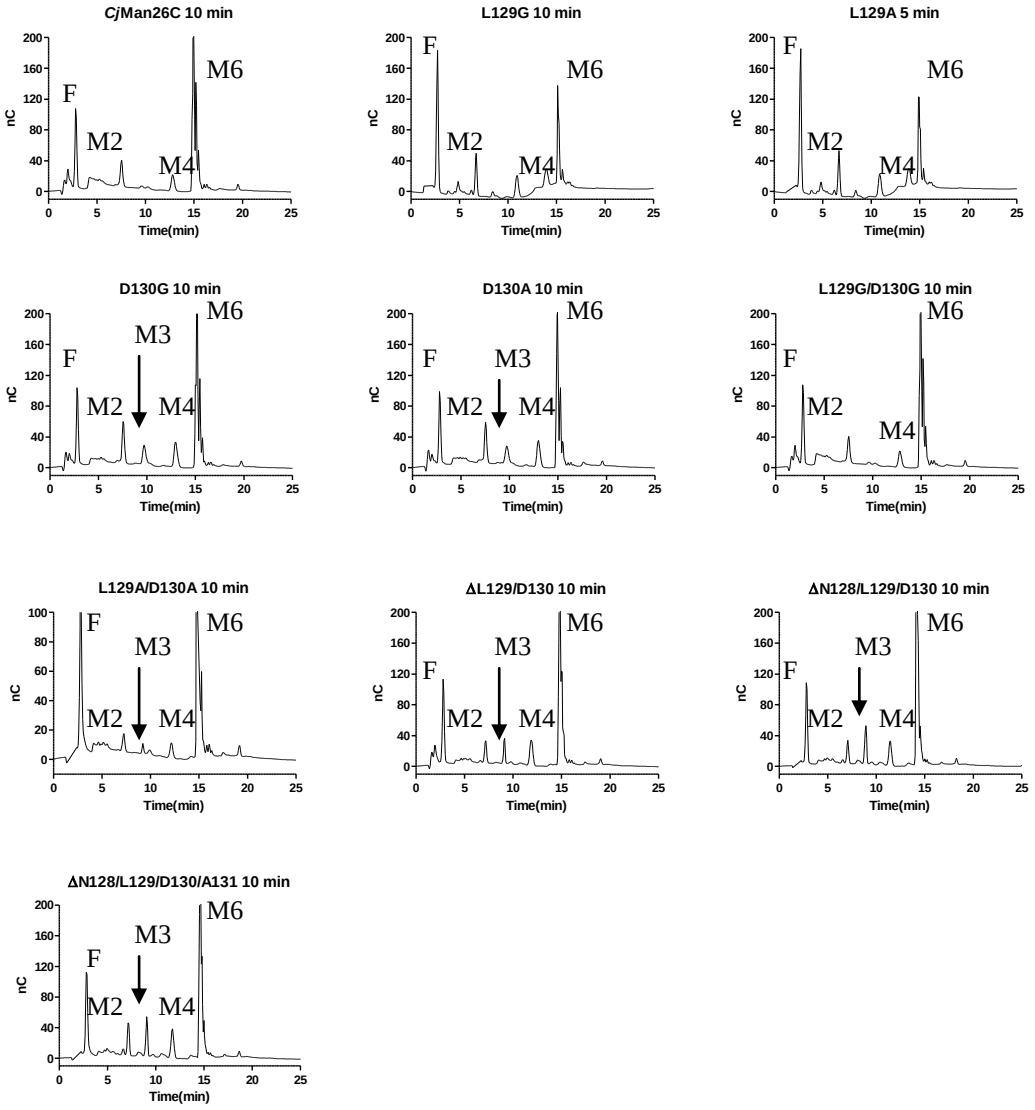


Figure 3.24 Chromatograms of *Cj*Man26C and mutants showing the products generated in the early stages of mannohexaose digestion.
Enzyme concentration rangeing from 20 pM–200 nM and 30 µM mannohexaose were used. F = Fucose (internal standard); M2 = Mannobiose; M3 = Mannotriose; M4 = Mannotetraose; M6 = Mannohexaose are shown. Reactions were carried out at 37°C in 50 mM sodium phosphate buffer pH 7.0.

3.6 Discussion

3.7 *Ct*Lic26A

3.7.1 Inhibitor binding

Unlike the majority of thio based glucan inhibitors, the thiopentasaccharide used in this study displayed a K_i that was similar to the K_M of the aryl-glycoside substrate used to perform the kinetic analysis (Money et al. 2006). It is intriguing that in the crystal structure, of *Ct*Lic26A, two molecules of the thiopentasaccharide bound to each enzyme. This could be an artefact of crystallisation or represent a true mode of inhibition. The similarity of the K_i to the K_M of the aryl-glycoside may suggest the thiopentasacchride does indeed span the active site in solution. Isothermal titration calorimetry (ITC) could be used to determine the stoichiometry of binding and elucidate whether the observed ratio of 2:1 is a crystallographic artefact.

3.7.1.1 +2/+1 subsites

The N157A mutant displays a substantial and similar reduction in both K_M and k_{cat} against the aryl-glycoside substrate compared to the wild type enzyme. This is diagnostic for a decrease in k_3, deglycosylation, relative to k_2, glycosylation, which may introduce a switch in the rate limiting step, as observed with other glycoside hydrolases (MacLeod et al. 1994). By decreasing k_3 relative to k_2, the covalent glycosyl-enzyme intermediate accumulates leading to a significant reduction in the observed K_M as the enzyme becomes saturated at a low substrate concentration. The aryl-glycoside substrate contains methyumbelliferone at the reducing end and it is this that occupies the +1 subsite. Methyumbelliferone has a pKa of 7.8, at 25°C (Graber et al. 1986), compared to a glucose molecule which has a pKa >12.0 (Tull et al. 1994). This makes methyumbelliferone a far superior leaving group and enhances the rate of glycosylation due to the reduced need for protonic assistance from E221 the acid/base during glycosylation. By contrast, irrespective of the substrate hydrolysed, the acid/base is required to activate a water molecule for the deglycosylation step. Kinetics of N157A indicates that the mutation has had a larger effect on deglycosylation than glycosylation, implying the N157 influences the function of the catalytic acid/base, particularly in the deglycosylation step.

Support for the role of N157, exclusively, in the deglycosylation step, is the observation that the mutation has little impact on the kinetic parameters of the enzyme against the polysaccharide substrate. The rate limiting step of endoglycanases is invariably glycosylation, as increased occupation of the positive subsites leads to increased catalytic efficiency (Hogg et al. 2001; Pell et al. 2004). Thus, although the N157A mutation reduces k_3, the rate limiting step, for polysaccharides, remains k_2. Therefore the reduction in the rate of deglycosylation has no impact on the activity of substrates, such as polysaccharides, which have poor leaving groups.

N157A lays 6.7 Å from the catalytic acid/base, E221. This distance is too great for N157 to be acting as a pKa modulator of the glutamate. The likely role of N157 is in helping to position a water molecule, from which a hydroxyl ion will be generated by the catalytic acid/base, to promote deglycosylation.

The mutation of W114, at the +1 subsite, caused a 4 fold elevation in K_M and had no effect on k_{cat} against the polysaccharide. This suggests the tryptophan provides no energy to transition state stabilisation but may play a minor role in forming the Michealis complex. Against the aryl-glycoside W114A had a major effect and, indeed, individual kinetic parameters could not be obtained, pointing to a K_M >>1 mM. The role of W114 therefore appears to be in the formation of the Michealis complex with the aryl-glycoside substrate, but is not important in polysaccharide binding. W114 may interact favourably with the aromatic ring of the methylumbelliferryl group through overlapping π orbitals. This interaction cannot be made by glucose as it does not contain π orbitals and thus, no effect of the mutation is observed against β-glucan. W114 may play a key role in aryl-glycoside binding by promoting the formation of productive complexes. Previous studies on *Cj*Man26A showed mutation of an aromatic residue at +1 switched binding from exclusively +2 to -2 to also allow +1 to -2 which caused a 10 fold reduction in catalytic efficiency, indicating the formation of less productive complexes (Hogg et al. 2001). This is unlikely to be the effect of W114A, as the β-1,3 linkage of the aryl-glycoside has to lie between the -1 and -2 subsites, as well as the demand that the methylumbelliferryl occupies +1 for cleavage of the substrate to occur, making unproductive complexes an unlikely explanation for the reduced activity of W114A against the aryl-glycoside when compared to the polysaccharide substrate. The size of β-glucan chains means there are many productive modes of binding and the chance of forming unproductive complexes would be low.

3.7.1.2 -2 subsite

The -2 subsite houses the bulk of the mutations analysed, and is linked to the -1 glucose molecule through a β-1,3 linkage. The F41A mutant caused a substantial reduction in k_{cat} against both the polysaccharide and the aryl-glycoside, although the effect was much greater for the latter. Little change in K_M was observed. This implies a general role for F41 in providing binding energy for transition state stabilisation. Catalysis proceeds through a $^4H_3$ half chair in *Ct*Lic26A (Money et al. 2006) as opposed to a $B_{2,5}$ boat, observed in GH26 mannanases (Ducros et al. 2002; Tailford et al. 2008). The crucial difference between these two transitions states is at O3, which lies equatorial in a $^4H_3$, but pseudo axial and up in $B_{2,5}$. F41 probably helps to promote a $^4H_3$ transition state locking the glucose at the -2 subsite into a conformation that favours the equatorial geometry of the linked O3 of the -1 subsite glucose.

The E70A, Y115A and E258A mutations displayed a similar phenotype to that of F41A. E70A makes hydrogen bonds to O2 and O3 of the -2 glucose molecule, Y115 makes a polar contact with O2 and E258A with O6. These three residues provide energy for stabilisation of the transition state, rather than binding of the substrate in the ground state. Once again implicating the importance of maximising the interactions made with the glucose molecule at -2 during the formation of the transition state.

These mutants at the -2 subsite display the same phenotype; a significant effect on k_{cat} but not on K_M. This suggests that the four amino acids mutated contribute to transition state stabilisation with the relative contributions being F41>E70A>Y115>E258. The interactions made by these residues may be compromised if a transition state other than $^4H_3$ was utilised ie. $^3H_4$ or $^{2,5}B$. For all four residues the effect was greater against the aryl-glycoside substrate than the polysaccharide, a generic issue that is addressed below.

K260 also resides at the -2 subsite and makes polar contacts with O6 and the endocyclic ring oxygen. Mutation of this residue causes a significant increase in K_M and decrease in k_{cat} for the aryl-glycoside substrate. Thus K260 appears to play a role both in the formation of the Michealis complex and provides some energy to transition state stabilisation. Against the polysaccharide substrate the K260A mutation had little effect on K_M with a reduction in k_{cat} being its main effect. This could suggest K260 is not needed to form the Michealis complex but contributes energy to the formation of the transition state.

The final residue at the -2 subsite is W72. Mutation of this amino acid has a large effect on k_{cat}/K_M against the aryl-glycoside substrate which, at least in part, reflects a substantial increase in K_M and implicates an important role in formation of the Michealis complex against the aryl glycoside. Against the polysaccharide the W72A mutation had very little effect on K_M but caused a 2 fold increase in k_{cat}. The slight increase may be due to accelerated departure of product at the end of the catalytic cycle.

3.7.1.3 -3 subsite

The final mutation is Q18A at the -3 subsite. The glutamine forms a hydrogen bond with O3 of the glucose molecule at -3. The mutation causes a ~10 fold reduction in k_{cat} and a ~10 fold increase in K_M against the aryl-glycoside substrate. This suggests that the distal subsite contributes to the formation of the Michealis complex and to transition state stabilisation. The mutation had little effect on the activity against the polysaccharide.

3.7.2 Conclusions

Mutations made in the negative subsites of retaining glycosides hydrolases are generally more detrimental than those introduced into the positive subsites (Armand et al. 2001; Hogg et al. 2001). This holds true for *Ct*Lic26A. The positive subsites are needed for glycosylation only, whereas the residues residing in the negative subsites are required for both the glycosylation and deglycosylation steps. This likely reflects the need to aid transition state distortion in the deglycosylation reaction through a reduced number of subsites. The positive subsites also have to allow departure of the leaving group after glycosylation and as such must not interact too strongly with the product. This selects for weaker glycone subsites.

In general the mutations had a far greater effect on aryl-glycoside hydrolysis compared to the polysaccharide cleavage. This trend has been observed previously for retaining glycoside hydrolases (Armand et al. 2001; Hogg et al. 2001). There are three possible reasons for this trend 1) The aryl-glycoside does not utilise all potential subsites; 2) The aryl-glycoside forms unproductive complexes; 3) The thermodynamics of binding are less favourable for aryl-glycoside versus polysaccharide substrates.

*Ct*Lic26A has 5 major subsites extending from +2 to -3. The aryl-glycoside used in this study for kinetic analysis was an aryl-trisaccharide which occupies the +1 to -3 subsites, with methylumbelliferryl occupying the +1 subsite and will not make the same interactions as glucose at this subsite. The polysaccharide will place glucose at all the subsites and therefore make more extensive interactions with the protein. These additional interactions made by the polysaccharide may 'dilute' the effects, of individual mutations, observed against the aryl-glycoside substrate.

The second possibility is unlikely and was referred to in the discussion of the W114A mutation. The methylumbelliferryl needs to be at +1 and the β-1,3 linkage can only be accommodated between -1 and -2 subsites, making the formation of unproductive complexes unlikely.

Another possible explanation for the different effect the mutations had on the two substrates is the thermodynamic penalties incurred by the two molecules. The aryl-glycoside is flexible and has substantial conformational flexibility. By contrast, the polysaccharide will have reduced conformational flexibility owing to its size and extensive intra-molecular hydrogen bonding. The more rigid conformation of the polysaccharide may be more complementary to the substrate binding cleft and thus maximise molecular interactions with the result that the effects of single mutations are dampered. The aryl glycoside is flexible and needs to be restrained into the enzymes active site and thus every interaction is crucial. Loss of a single hydrogen bond effects the ability of the molecule to be accommodated in the active site. The suggestion here is that a polysaccharide substrate maximises enthalpic interactions through the formation of optimal hydrogen bonds whereas the aryl-glycoside cannot. The most likely and easily understood explanation of those discussed is that the polysaccharide utilises the additional subsites available and thus the effects of mutagenesis are countered by more extensive interactions between the protein and the polysaccharide.

3.8 *Cj*Man26C

3.8.1 Activity of *Cj*Man26C

The lower activity of *Cj*Man26C, compared to *Cj*Man26A, against the polysaccharides galactomannan and glucomannan reflects the exo-activity of the enzyme. *Cj*Man26C can only hydrolyse galactomannan and glucomannan from the non-reducing end, whereas *Cj*Man26A is able to explore the whole polymer chain making the substrate more accessible. Furthermore, *Cj*Man26C is only able to accommodate a galactose residue at the -1 or +2 subsites but not the +1 or -2 subsites. This is due to O6 of the -1 mannose, which is linked α-1,6 to galactose, pointing into solution and there appears to be a 'void' into which the galactose residue can be tolerated (Figure 3.25). O6 of the mannose residues at +1 and -2, however, points into the core of the protein and thus a galactose decoration cannot be accommodated at these subsites due to steric hindrance. The O6 of the mannose residue at +2 points into solution and it is likely that a galactose residue will be tolerated. This is consistent with the production of mannobiose and galactomannobiose from galactomannan. The data indicate that all galactomannobiose produced will have galactose appended at the reducing end of the disaccharide.

The only product generated from glucomannan is mannobiose. This indicates that only mannose is tolerated in the -1 and -2 subsites. Selection at -1 is for an equatorial O2 and axial O3 of the $B_{2,5}$ transition state, and only a mannose residue fills this requirement, as discussed later. At the -2 and +2 subsites there is selection for an axial O2. In *Cj*Man26A R269, analogous to R374 in *Cj*Man26C, attributes the tight binding of mannose at the -2 subsite (Hogg et al. 2001; Tailford et al. 2009) and this is likely to be the role of R374 in *Cj*Man26C. There is no selection at +1 for the axial O2 of mannose and thus a glucose molecule may be accommodated at this subsite.

The above restrictions imposed on *Cj*Man26C are compounded by its exo mode of action from the non-reducing end. Only the non-reducing ends of galactomannan chains, which do not place galactose at +1 or -2, and glucomannan molecules that place mannose at the -1 and -2 susbsites, are substrates for *Cj*Man26C. The exo-mannanase will trim these chains back until it creates non-reducing ends that the enzyme cannot hydrolyse. *Cj*Man26A, which displays an endo mode of action, is able to 'scan' along galactomannan and glucomannan chains to find areas which are either lightly decorated or exclusively mannose. This greatly increases the effective

substrate concentration available to *Cj*Man26A, and readily explains its enhanced activity over *Cj*Man26C. It does not appear, however, that *Cj*Man26A can accommodate galactose at its -1 subsite, as no galactomannobiose is observed when galactomannan is treated with *Cj*Man26A. *Cj*Man26A has a small α–helix comprising residues 325 to 331 at the -1 subsite that would appear to clash with a galactose linked α-1,6 to the -1 mannose (Figure 3.25B). This helix is absent in *Cj*Man26C and instead creates a void, as discussed previously, which can be seen in the surface representation in Figure 3.25A, into which galactose can be accommodated.

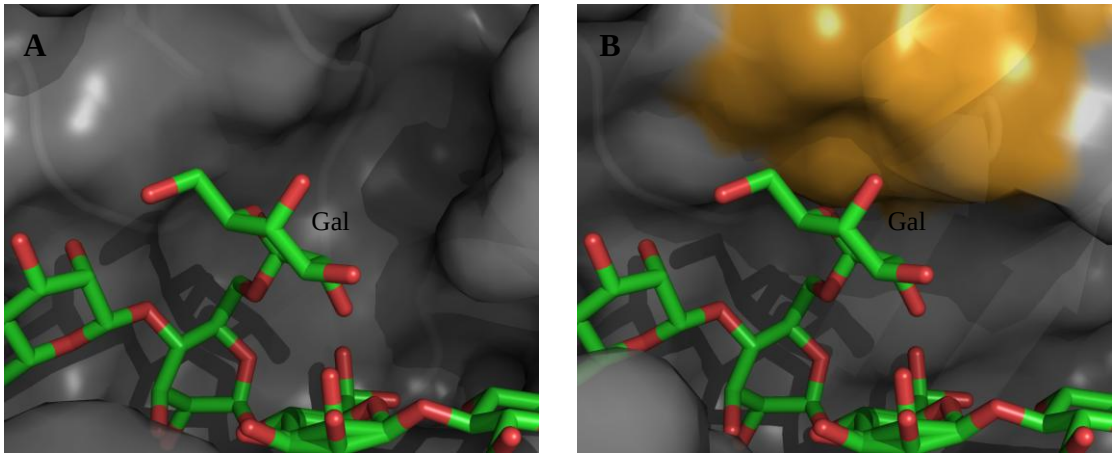


Figure 3.25 Accommodation of galactose at the -1 subsite.
A-shows *Cj*Man26C complexed with 6^3-Gal-Man_4. B-shows *Cj*Man26A overlaid with 6^3-Gal-Man_4, the additional helix is highlighted in orange.

Against mannotriose *Cj*Man26C and *Cj*Man26A display similar rates, while the exo-mannanase is ~200 fold more active against the tetrasaccharide compared to the endo enzyme. This increase in rate, which equates to an additional 2.7 kcal/mol, can be attributed to the additional interactions made by S268 and R269 at the +2 subsite of *Cj*Man26C. There are no obvious interactions made at the +2 subsite of *Cj*Man26A and it is not readily evident why this subsite confers a 10 fold increase in k_{cat}/K_M, although it is possible the increase in catalytic efficiency may reflect a preference of the +1 subsite for an internal β-D-mannose (Hogg et al. 2001).

3.8.2 Crystal structutre

The crystals of the wild type and the nucleophile mutant of *Cj*Man26C were grown in the presence of galactomannobiose and mannotetraose respectively. The resultant crystal structures did not contain the target ligand. In the nucleophile mutant,

E338A, mannotetraose had been hydrolysed to mannobiose, which occupied the -1 and -2 subsites in a relaxed $^4C_1$ chair conformation. The hydrolysis of mannotetraose may not be surprising as the E338A mutant displayed some activity despite being 5 x 10^6 fold less than that of the wild type. This residual activity may be due to trace amounts of wild type enzyme created by a translation error during expression (Shallom et al. 2002); this is unlikely as a two nucleotide change of GAA to GCC was made. *Cj*Man26C is such an efficient catalyst that a double catalytic mutant may need to be made to completely ablate activity. At pH 7.0 there are 10^{-7} molar of hydroxyl ions, these rare hydroxyl ions may be able to fill the void beneath the -1 mannose sugar left by mutation of glutamate to alanine and thus allow catalysis to occur, albeit at a much reduced rate. The structure of wild type *Cj*Man26C in complex with "galactomannobiose", obtained from Megazyme, provided a serendipitous insight. An oligosaccharide of 4 mannose residues, bound +2 to -2, decorated at -1 with a α-1,6 linked galactose was observed. There was also some evidence for a galactose appended to the +2 mannose residue but it was too disordered to be modelled correctly. This may represent a trace contaminant of the 6^1-6^3-Gal_2-Man_4 (a hexasaccharide was observed by mass spectrometry), or a novel transglycosyltaion product as has been observed with some amylases (Dauter et al. 1999; Brzozowski et al. 2000). In any event it is extremely unusual to observe substrate bound in a productive mode spanning the +1/-1 subsites to a catalytically competent glycoside hydrolase. Possible explanations for this observation are as follows: The galactose appended to the -1 mannose makes unfavourable steric clashes with the C2-C3 edge of the mannose at +1 which may hinder transition state stabilisation. Also the galactose forms a hydrogen bond with the endocyclic ring oxygen of the +1 mannose via O2, and this may slow product departure, promoting condensation of the hydrolysed products.

3.8.2.1 The -1 subsite

The -1 mannose is distorted into a $^1S_5$ skew boat conformation with the nucleophile, E338, poised for in-line attack at C1 with an angle of ~154°. This has been observed previously for GH26, as well as GH2, β-mannanases and β-mannosidases (Ducros et al. 2002; Tailford et al. 2008). This time the distortion is with a wild type enzyme bound to a natural substrate with a sugar as the leaving group. The previous work involved a 2,4 dinitrophenyl leaving group and a 2 fluoro

substituted sugar (Ducros et al. 2002). This allows the -1 subsite interactions to be observed for both the $^4C_1$ chair conformation and the $^1S_5$ skew boat conformation. The distortion into a $^1S_5$ conformation suggests catalysis proceeds $^1S_5 > B_{2,5} > {}^0S_2$ for glycosylation and $^0S_2 > B_{2,5} > {}^1S_5$ for deglycosylation, with the transition state being $B_{2,5}$. The advantage of a $B_{2,5}$ transition state for β-D-mannose is that O2 is placed pseudo equatorial, as seen in the $^1S_5$ conformation where, it hydrogen bonds to H220 and E338, and thus 1,2 diaxial interactions (caused by an axial O2) at the anomeric centre are avoided during catalysis. The equatorial orientation of O2 would not occur with a $^4H_3$ half chair, $^3H_4$ half chair or $^{2,5}B$ boat transition state and in the latter two O4 would be pseudo axial (and down), and thus the -2 mannose would make steric clashes with the protein surface. It has been shown previously that an equatorial O2 hydroxyl at the transition state is an important feature of β retaining hydrolases (Namchuk et al. 1995). During catalysis the pKa of the O2 hydroxyl is altered and its proton becomes more acidic, enabling it to associate more favourably with the lone pairs of electrons of the nucleophile carbonyl group during catalysis. Indeed work on an Agrobacterium β-glucosidase, using substrates where the various hydroxyls had been replaced by hydrogen or fluorine, demonstrated that the O2 hydroxyl contributes 3-4 fold more energy (kcal/mole) at the transition state, in glycosylation and deglycosylation, than any other hydroxyl in the pyranose ring (Namchuk et al. 1995). This interaction does not occur in α retainers or inverting enzymes. Despite the importance of O2 at the transition state and the fact it is the unique feature of mannose in the ground state, it does not drive the specificity for mannose at the -1 subsite as it makes no interactions with the protein in the $^4C_1$ ground state (Ducros et al. 2002). *Ct*Lic26A is a GH26 enzyme active on the glucose polymer β-glucan and performs catalysis using a $^4H_3$ half chair transition state which also places O2 in an equatorial position (Money et al. 2006). So if a glucose molecule occupied the -1 subsite of *Cj*Man26C it could make the same interactions with H220 and E338 through a $^4H_3$ conformation as mannose does in a $B_{2,5}$ geometry. The selection for mannose at -1 is derived from the conformation adopted by O3 at the transition state. As mannose is distorted to the $B_{2,5}$ conformation O3, which is equatorial in the $^4C_1$ ground state, becomes pseudo axial and 'up'. In contrast in a $^4H_3$ conformation, adopted by glucose, O3 would remain equatorial making this the unique feature that confers specificity for mannose at the -1 subsite (Taylor et al. 2005; Money et al. 2006; Tailford 2007; Tailford et al. 2008). Overlays of *Cj*Man26C, *Cj*Man26A and *Ct*Lic26A show

considerable conservation at the -1 subsite with the noticeable exception that no protein side chain is available to interact with O3 in *Ct*Lic26A. In families GH1 and GH5, comparison of closely related members that act on glucose and mannose configured substrates suggest that altered loop conformations in the -1 subsite resulted in a differing for selection of O3 at the transition state (Taylor et al. 2005). This is not needed in GH26 as all members that do not act on mannose polymers act on substrates where O3 of the -1 sugar forms a glycosidic linkage with the residue at -2 and is therefore tethered (Taylor et al. 2005). Recently a structural genomics program has solved the structure of a GH26 enzyme predicted to be a putative β-1,4 endo-glucanase (Structural Genomics). It would be interesting to know, if this GH26 enzyme is indeed an endo-glucanase, whether catalysis will proceed through a $^4H_3$ transition state without the tethering of O3 of the -1 subsite sugar.

3.8.2.2 Exo/Endo conversion

The most unique feature of *Cj*Man26C is its exo-mode of action. This can be attributed to a four residue insertion of the amino acids asparagine, leucine, aspartate and alanine in loop 3 that connects β-strand 2 and α-helix 3. This raises D130 in loop 3 2 Å above the position of the equivalent residue, E121, in *Cj*Man26A. This enables D130 to interact with both O4 and O6 of the -2 mannose residue. These polar interactions by D130 fix the loop in place when the substrate binds and, together with L129; present a steric block to extension beyond the -2 subsite. The action of L129 is through its backbone carbonyl and thus it is not surprising that mutation of this residue had no effect on the exo-mode of action; the backbone carbonyl will be invariant amongst all the amino acids. All mutants of D130, except the double mutant L129G/D130G, caused loss of the exo mode of action. Mutation or deletion of D130 abolishes the ability of loop to lock in place upon substrate binding, promoting an endo-mode of action (Figure 3.24). Against galactomannan the endo mutants of *Cj*Man26C appear to be comparable to the mode of action of *Cj*Man26A. Indeed, the *Cj*Man26C endo mutants even appear to produce a greater number of oligosaccharides than *Cj*Man26A. As alluded to earlier, there is no evidence that *Cj*Man26A can accommodate galactose into its binding cleft, while *Cj*Man26C can. This would allow *Cj*Man26C loop mutants to perform more extensive hydrolysis of galactomannan in the absence of α-galactosidases.

The question was then asked just how endo were the endo-acting mutants of *Cj*Man26C? This was assessed by analysing the initial products of mannohexaose hydrolysis, comparing the ratio of mannobiose to mannotetraose plus mannotriose. A true endo enzyme should generate a ratio of 2:1. This was indeed the case with *Cj*Man26A displaying a ratio of 2.2:1. A true exo, releasing mannobiose, will generate a ratio of 1:0. The two *Cj*Man26C mutants ΔNLD and ΔNLDA displayed ratios of 2.1:1 and 2.3:1 respectively. This shows that these mutants have a complete endo mode of action with binding of mannohexaose being random. The other mutants that displayed an endo-mode of action had ratios ranging from 3.0:1 to 5.6:1. This signifies that, although they have an endo-mode action, binding is not random. For a 100 % conversion to an endo acting enzyme, D130 has to be removed along with at least two flanking residues. Although D130 is needed to lock the loop in place to confer an exo-mode of action, the loop in which only D130 has been mutated, may still be able to partially block extension beyond -2.

The only mutation of D130 that did not introduce endo activity was the double mutant L129G/D130G. This mutant also retained significant catalytic activity. The introduction of consecutive glycine residues may have introduced significant conformational flexibility into the loop. This may allow another residue to substitute for the loss of D130, or possibly the backbone carbonyls of the glycine residues themselves. Either of these possibilities may help explain the surprising characteristics of the L129G/D130G mutant.

All mutations of either L129 or D130 cause a significant decrease in catalytic efficiency. This may not be surprising as both residues are in the glycone region of the active site. For retaining glycoside hydrolases mutations in this region are highly detrimental as they provide energy for both glycosylation and deglycosylation steps. Mutation or removal of D130 removes two hydrogen bond interactions between substrate and the enzyme at the -2 subsite and this readily explains the dramatic loss in activity. The mutation or removal of L129 had a significant effect of catalytic efficiency also. This could be due to a change in the conformation adopted by D130. Mutation of L129 to glycine or alanine may cause D130 to adopt a geometry where it does not form optimal hydrogen bonds with the substrate. The effect could also be due to the loss of a hydrophobic interaction between L129 and the C6 of the -2 mannose residue.

There are few examples of successful exo to endo conversions in the literature. Experiments on cellobiohydrolases, such as *Hi*Cel6A and *Tr*Cel7A, have engineered these exo acting enzymes to a more endo character. This was done by mutating residues in loops comprising the tunnel ceiling and a decrease in viscosity of hydrolysed substrate was observed, indicating a more endo character, although a significant catalytic cost was incurred (Kleywegt et al. 1997; Varrot et al. 1999). The modification of a key binding residue in the -5 subsite of a polygalacturonase changed its mode of action from endo processive to endo random (Abbott et al. 2007). More successful exo to endo conversions were that of *Cj*Arb43A and YesX;BSU07060. *Cj*Arb43A releases only arabinotriose from linear arabinan. By mutating the residues D32 and Q35 in the -3 subsite the enzyme was switched to an endo mode of action with only a 25 % loss in catalytic efficiency (Proctor et al. 2005). YesX;BSU07060 is a pectate lyase from family 11 and releases a unsaturated disaccharide from rhamnogalacturan. Mutation of a loop insertion comprising residues P439 to Y447 converted the enzyme to an endo mode of action with a ~3 fold increase in k_{cat}/K_M (Ochiai et al. 2009).

3.8.3 Conclusion

In light of the genome sequence of *C. japonicus* it is believed *Cj*Man26C represents the final mannan degrading tool in the bacterium's arsenal of mannan degrading enzymes. The bacterium expresses three GH26 β-mannanases, *Cj*Man26A, *Cj*Man26B and *Cj*Man26C, and four GH5 enzymes, *Cj*Man5A, *Cj*Man5B, *Cj*Man5C and *Cj*Man5D. *Cj*Man5A, *Cj*Man5B and *Cj*Man5C are extracellular mannanases appended to carbohydrate binding modules that target crystalline cellulose (CBMs). The likely role of these enzymes is to attack the insoluble plant cell wall and initiate the hydrolysis of mannans and glucomannans. These solubilised mannans will then be hydrolysed by *Cj*Man26A, *Cj*Man26B, *Cj*Man26C and *Cj*Man5D which are appended to the surface of *C. japonicus* by virtue of type II signal peptides (these append mature enzymes to the surface of the cell). These surface bound enzymes lack CBMs, providing support for their role in the hydrolysis of soluble mannans. *Cj*Man26B is attached to the outer membrane by a long 70 amino acid linker. This long linker will allow *Cj*Man26B to explore a greater sequence space than the other outer membrane bound mannanases. *Cj*Man26B may act as an intermediate between the extracellular GH5 enzymes and the outer membrane bound mannanases. It may break long soluble

mannans, produced from hydrolysis of the plant cell wall, into smaller oligosaccharides that are more easily dealt with by the other mannnanases. *Cj*Man26B, potentially, has 6 subsites whereas *Cj*Man26A and *Cj*Man26C possess 4. The activity of *Cj*Man26B against mannotetraose is 3-5 orders of magnitude below that of *Cj*Man26A and *Cj*Man26C adding some support to this proposal. It is not clear why *C. japonicus* expresses both *Cj*Man26A and *Cj*Man26C. There may be exo/endo synergy displayed by the two enzymes, similar to that observed in the human intestinal tract for peptide and polysaccharide degradation (Vrieze et al 2010.).

3.9 Overall Conclusion

GH26 is an exquisite example of how closely related enzymes harness different conformational itineraries to perform hydrolysis. *Cj*Man26C adds to the growing body of support for the utilisation of a $B_{2,5}$ transition state by glycoside hydrolase families acting on β-mannosides. The -2 subsite of *Ct*Lic26A is essential for efficient catalysis. This may not be surprising as *Ct*Lic26A harnesses a $^{4}H_{3}$ transition state with O3 tethered to the -2 sugar. The difference between glucose and mannose is an axial O2 in the ground state with O3 being equatorial for both. At the transition state, however, O2 occupies the same position in glucose and mannose and it is O3 that is the major specificity determinant at the -1 subsite.

*Cj*Man26C adds to the number of successful exo to endo conversions, and provides an elegant example of how subtle changes, the insertion of four amino acids, can significantly alter function. Knowledge of these subtle changes can aid in the identification and creation of novel, industrially relevant, enzymes.

Chapter 4

Characterisation of a Novel Cellulase that founds a New Glycoside Hydrolase Family

4.1 Introduction

Cellulose, a β-1,4 polymer of glucose, is the major polysaccharide found in both primary and secondary cell walls. It is the most abundant molecule in the biosphere and a major contributor to the recalcitrance of the plant cell wall to biological attack. Archea, bacteria, fungi and lower members of the animal kingdom, such as termites, express cellulases that degrade cellulose into its constituent monomer glucose. The sugar can then be utilised as an energy source. It is through harnessing of these cellulases that cellulose degradation can be coupled to the production of ethanol by yeast, which can then be used as a liquid fuel (Somerville 2007; Himmel et al. 2009). This is of great importance as it allows the plant cell wall to be used as a renewable, carbon dioxide neutral, energy source. When oil inevitably runs out it is these sources of energy that may contribute to the sustainable production of liquid of fuel.

Cellulases, which act synergistically to break down cellulose, display three modes of action. Endo acting cellulases cleave randomly along the cellulose chain to release cellooligosaccharides (Birsan et al. 1998). The activity of these enzymes tends to be low as the crystalline nature of cellulose restricts access. Endo-processive enzymes cleave randomly then work processively from the point of cleavage (Zverlov et al. 1999; Zverlov et al. 2003). Exo acting enzymes primarily comprise cellobiohydrolases and β-glucosidases (Sanz-Aparicio et al. 1998; Guimaraes et al. 2002). Cellobiohydrolases release cellobiose from either the non-reducing or the reducing end of cellulose chains; they act in a processive manner and are thus able to degrade crystalline forms of the substrate (Guimaraes et al. 2002). β-glucosidases release glucose from the non-reducing end of cellooligosaccharides (Sanz-Aparicio et al. 1998). Cellulases can be found in the following glycoside hydrolase (GH) families in the CAZy database: GH5, 6, 7, 8, 9, 12, 44, 45, 48 and 74 (Henrissat 1991; Henrissat et al. 1997).

Glycoside hydrolases bind to multiple sugar residues in polysaccharide chains, and each region of these enzymes that binds a monomer is known as a subsite. The subsites are given positive or negative numbers with bond cleavage occurring between +1/-1. As the subsites move toward the reducing end of the substrate they increase in positive value ie +2, +3, while as they move out toward the non-reducing end they increase in negative value ie. -2, -3 (Davies et al. 1995).

Clostridium thermocellum is one of the most efficient cellulose degrading organisms known (Fontes et al 2010.). The bacterium organises its plant cell wall degrading machinery into a membrane bound multi enzyme complex termed the cellulosome (Fontes et al 2010.). The non-catalytic scaffoldin of the cellulosome, called CipA, is appended to the membrane through a type II dockerin. CipA contains tandemly arranged modules called type I cohesins (Fontes et al 2010.). Enzymes are recruited to the CipA scaffoldin through the interaction of type I dockerin modules, present at the C terminus of the catalytic subunits, to the CipA type I cohesins. The affinity between dockerins and cohesins is extremely tight with K_A values in excess of $1 \times 10^9\ M^{-1}$ (Carvalho et al. 2007).

Dockerins are easily identified in the genome of *C. thermocellum* as they have a high degree of conservation. Proteins appended to dockerins are targeted to the cellulosome and are thus likely to contribute to plant cell wall degradation. This gives a method of scanning the *C. thermocellum* genome for genes encoding dockerins that share little or no sequence similarity with characterised carbohydrate active proteins. This provides a way to search for novel components of the cellulosome.

The protein, initially named Cthe0435, and subsequently referred to here as *Ct*Cel119, was identified in this way. The ~300 residue dockerin-containing protein contains a ~250 amino acid module that shows no similarity to any characterised protein. A BLAST search against the Uniprot database renders only 12 results. The top result is a protein of 606 amino acids with a score of 90, an identity of 26% over ~100 amino acids and a ε-value of 0.3, which represents very limited sequence similarity (data not shown).

*Ct*Cel119 was shown to be significantly up regulated when *C. thermocellum* was grown on crystalline cellulose, suggesting *Ct*Cel119 plays a role in the hydrolysis of this β-1,4 glucose polymer (Zverlov et al. 2005; Gold et al. 2007). Preliminary biochemical analysis of this novel protein showed that it had catalytic activity against β-1,4 glucans and β-1,4,1,3 mixed linkage glucans, while its crystal structure revealed

a novel cellulase fold. These experiments were carried out by **Miss Joana Brás from the laboratory of Professor Carlos Fontes**. In this Chapter the biochemical properties of the enzyme were explored in detail.

4.2 Objectives

To characterise *Ct*Cel119, both biochemically and structurally, providing the basis for the founding of a new glycoside hydrolase family.

4.3 Results

4.3.1 Expression and purification

The plasmid pCel119, contained nucleotides 450 – 1125 of Cthe0435 inserted into the *Nhe*I and *Xho*I sites of pET28a and the encoded protein, which comprises the ~250 amino acid module of unknown function, contains an N–terminal His tag (**provided by Professor Carlos Fontes**). pCel119 was transformed into *E. coli* strain BL21 (Section 2.1.10) and expression of *Ct*Cel119 was induced by the addition of 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) when the A_{600nm} of the culture reached 0.6. Induction of recombinant protein expression was performed at 37 °C with incubation for 4-6 h. (Section 2.2.24). Cells were then centrifuged, cell free extracts were prepared (Section 2.2.25), immobilised metal affinity chromatography (IMAC) was used to purify *Ct*Cel119 (27 kDa) to electrophoretic homogeneity for subsequent biochemical analysis. SDS-PAGE of proteins at each stage of the purification is shown in Figure 4.1. The concentration of *Ct*Cel119 was calculated using the molar extinction coefficient of 52370 M^{-1} cm^{-1} at A_{280nm}.

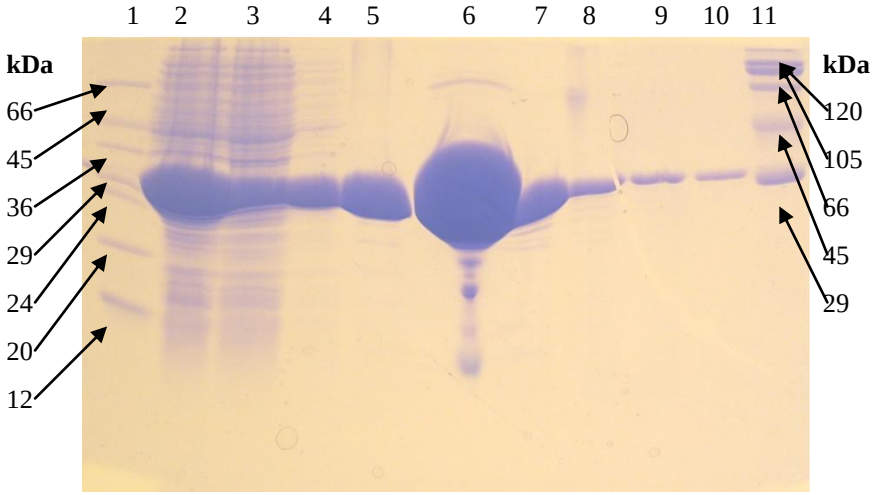


Figure 4.1 SDS-PAGE of IMAC purification of *Ct*Cel119.
1, High molecular weight marker; 2, Cell free extract; 3, Flow through; 4, Talon buffer wash; 5, Talon buffer supplemented with 10 mM imidazole; 6-10, Talon buffer supplemented with 100 mM imidazole.

4.3.2 Biochemical characterisation

Recombinant *Ct*Cel119 expressed in *E. coli* showed activity against linear β-1,4 and mixed linkage β-1,4,1,3 glucans. It did not show activity against other plant cell wall polysaccharides including β-1,3 glucans such as laminarin, xylans, mannans or the polysaccharides chitin and chitosan (data not shown). The pH and temperature optima of *Ct*Cel119, using β-glucan as the substrate, are pH 5.5 and 60 °C respectively; data were generated by **Miss Joana Brás from the laboratory of Professor Carlos Fortes**. This chapter continued the work of Miss Brás by providing a more quantitative, and in depth, analysis of the enzyme. *Ct*Cel119 displayed activity against the glucose polymers β-glucan, carboxymethyl cellulose (CMC), phosphoric acid swollen cellulose (PASC) and trace activity against Avicel (Table 4.1).

Enzyme	Substrate	Specific activity[a] (min^{-1})	k_{cat}/K_M (min^{-1} M^{-1})
*Ct*Cel119	β-glucan	$6.4 \times 10^3 \pm 2.0 \times 10^2$	-
*Ct*Cel119	CMC	$4.8 \times 10^2 \pm 6.9 \times 10^1$	
*Ct*Cel119	PASC	$1.6 \times 10^3 \pm 2.6 \times 10^2$	-
*Ct*Cel119	Cellohexaose	-	$1.01 \times 10^4 \pm 2.0 \times 10^2$
*Ct*Cel119	Cellopentaose	-	$4.50 \times 10^2 \pm 8.5 \times 10^1$
*Ct*Cel119-CBM3a	β-glucan	$2.7 \times 10^3 \pm 4.0 \times 10^2$	-
*Ct*Cel119-CBM3a	PASC	$1.7 \times 10^3 \pm 2.4 \times 10^1$	-
*Ct*Cel119-CBM3a	CMC	$3.4 \times 10^2 \pm 8.1 \times 10^1$	-

Table 4.1 Kinetic data of *Ct*Cel119 and the *Ct*Cel119-CBM3a fusion
[a] specific activity was calculated at a substrate concentration of 5mg/ml. All data were generated by linear regression; errors are standard errors generated from triplicate data.

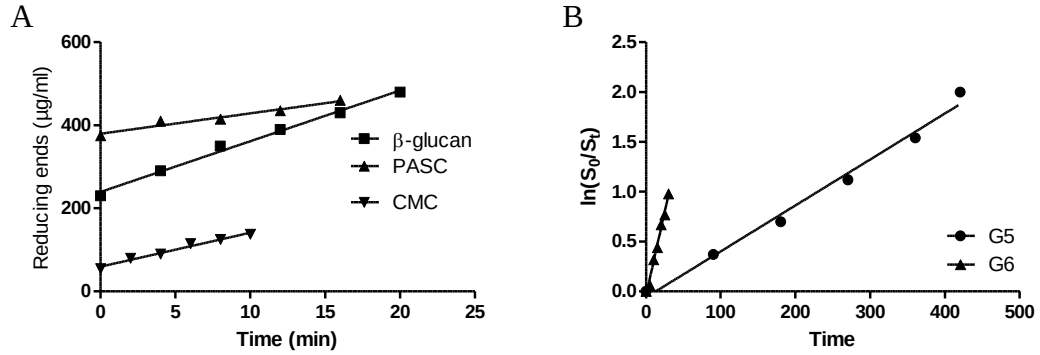


Figure 4.2 Graphs of Linear regression used to calculate values in Table 4.1
A-Rates of *Ct*Cel119 versus β-glucan, PASC and CMC (5 mg/ml) measured by reducing sugar assay using DNSA reagent and the method of Miller (1959). Concentration of *Ct*Cel119 used was 15 nM except for CMC where 80 nM was deployed. B-The substrate depletion method was used to calculate k_{cat}/K_M. Cellohexaose (G6) and cellopentaose (G5) concentrations were 15 µM vs 10 µM and 3 µM *Ct*Cel119 respectively. Graphs were plotted in GraphPad Prism 4.02 using linear regression.

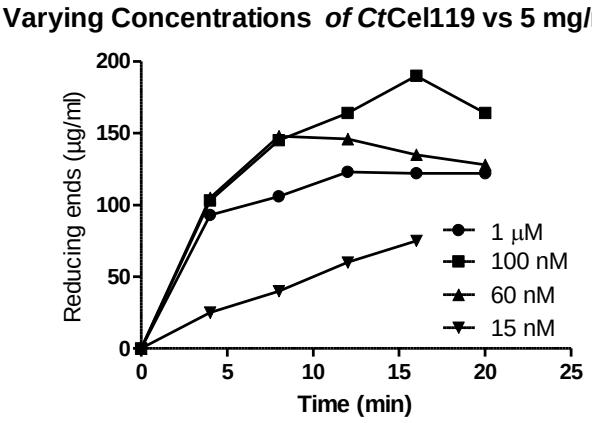


Figure 4.3 Varying concentrations of *Ct*Cel119 versus 5 mg/ml PASC
Burst kinetics of *Ct*Cel119 demonstrating initial hydrolysis after which no observable hydrolysis occurs. Graph was plotted in GraphPad Prism 4.02.

The highest activity of *Ct*Cel119 was observed against the mixed linked β-glucan. This activity was only four fold higher than the activity observed against PASC. The difference is likely a reflection of the soluble and insoluble nature of the respective substrates. The kinetics of PASC displayed an 'initial burst' after which further hydrolysis could not be reliably measured, shown in Figure 4.2A and Figure 4.3. Trace activity was observed against Avicel, which is a highly crystalline preparation of cellulose, but a specific activity could not be measured. CMC is a preparation of cellulose where either/and O2, O3 and O6 can be substituted with carboxymethyl groups at each glucose molecule. This has the effect of solubilising the cellulose chains. *Ct*Cel119 displayed low activity against CMC.

*Ct*Cel119 was fused to *Ct*CBM3a (a cellulose specific carbohydrate binding module (CBM) that is a component of CipA) to create *Ct*Cel119-CBM3a; this was performed by **Dr Artur Rogowski from the laboratory of Professor Harry Gilbert**. This was done to investigate whether the activity of *Ct*Cel119 could be potentiated by the CBM against insoluble substrates. CBMs have been shown to enhance the activity of their appended enzymes against insoluble substrates by 2-10 fold (Hall et al. 1995). *Ct*Cel119-CBM3a showed no elevated activity, compared to *Ct*Cel119, against PASC or the soluble substrates. Figure 4.4 shows the profile of the initial products released by *Ct*Cel119 against PASC. The oligosaccharides produced ranged from cellotriose to cellohexaose, cellotetraose being the most abundant product.

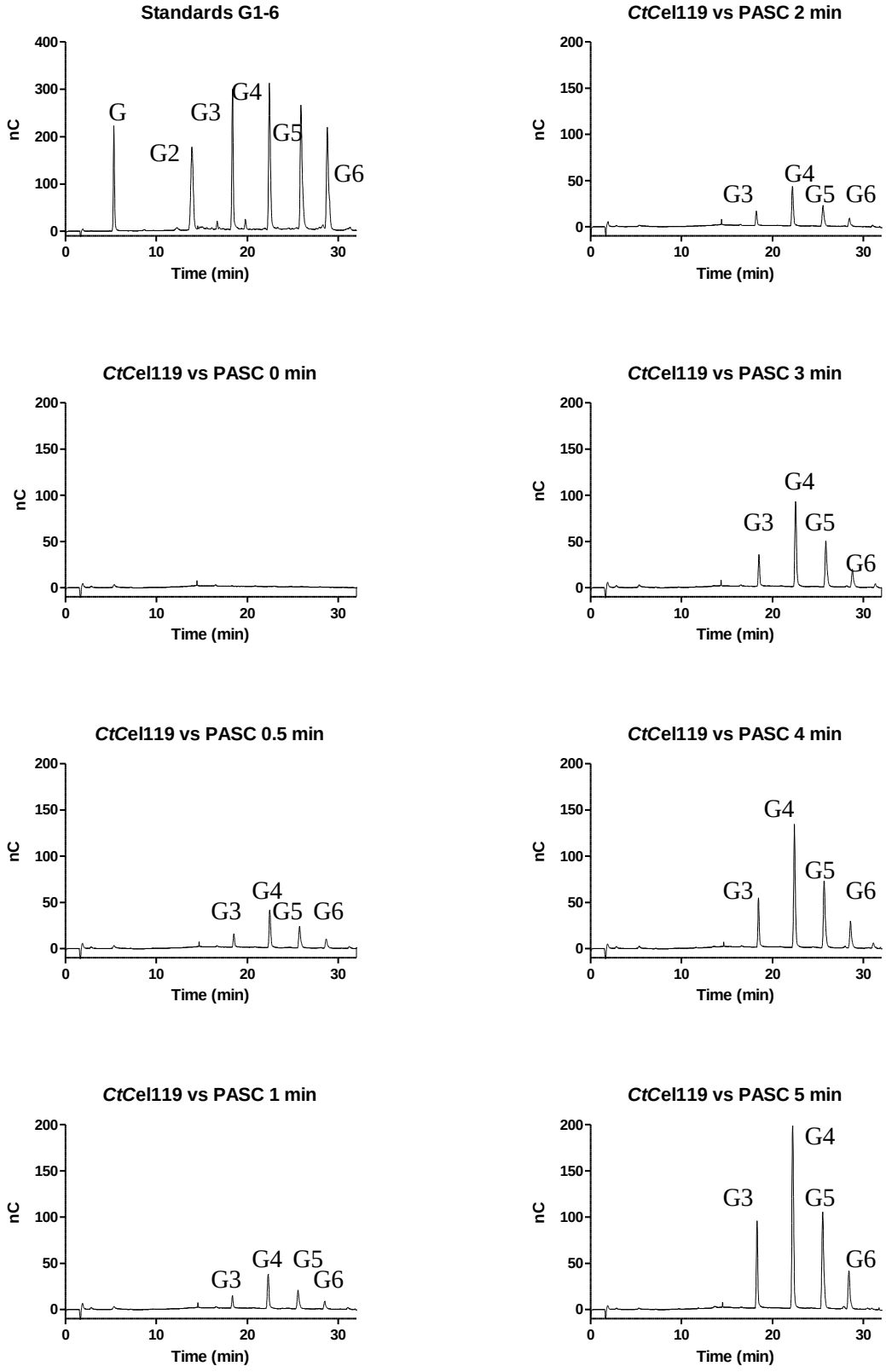


Figure 4.4 Initial oligosaccharides released by *Ct*Cel119 versus PASC.
G1 = Glucose, G2 = Cellobiose, G3 = Cellotriose, G4 = Cellotetraose, G5 = Cellopentaose and G6 = Cellohexaose. *Ct*Cel119 (5 µM) vs. PASC (2 mg/ml) was used. Reactions were carried out in 50 mM MES 20 mM NaCl pH 5.5 at 60 °C.

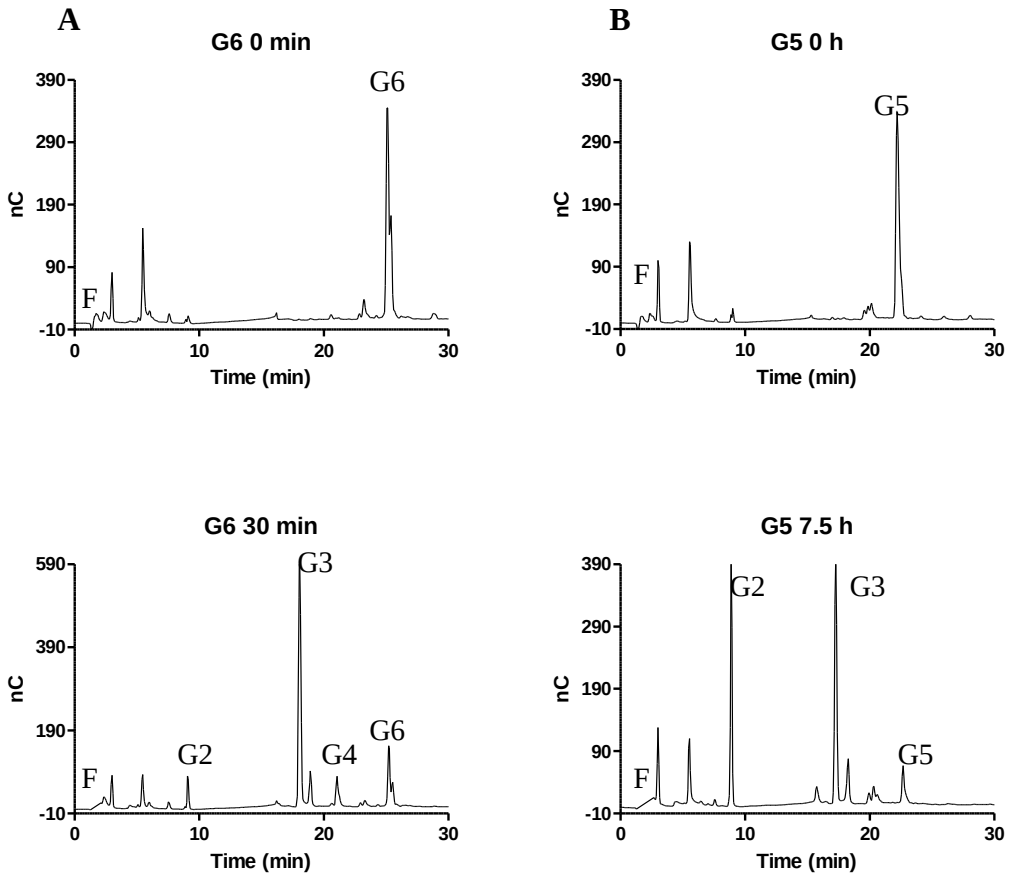


Figure 4.5 Digestion profiles of *Ct*Cel119 versus cellohexaose and cellopentaose
A-Cellohexaose at 0 min and 30 min. B-Cellopentaose 0 h and 7.5 h. F = Fucose (internal standard);G2 = Cellobiose;G3 = Cellotriose;G4 = Cellotetraose;G5 = Cellopentaose and G6 = Cellohexaose. Oligosaccharides were at 15 µM; 10 µM and 3 µM *Ct*Cel119 was used against cellopentaose and cellohexaose, respectively. Reactions were carried out in 50 mM MES, 20 mM NaCl pH 5.5 at 60 °C.

The activity of *Ct*Cel119 was assayed against cellooligosaccharides. *Ct*Cel119 had a k_{cat}/K_M value of 1.01 x 10^4 min^{-1} M^{-1} against cellohexaose, which is ~20 fold higher than against cellopentaose (Table 4.1). This indicates *Ct*Cel119 has at least six subsites. No activity was detected against cellotetraose or smaller cellooligosaccharides. The product profile of cellohexaose hydrolysis by *Ct*Cel119 (Figure 4.5) shows cellotriose is the major product. This is consistent with *Ct*Cel119 containing six subsites in which the cellohexaose extends from -3 to +3. A very small amount of cellobiose and cellotetratose was also produced. This may indicate the presence of a seventh subsite which is rarely occupied by cellohexaose. The digestion profile of *Ct*Cel119 against cellopentaose shows only cellobiose and cellotriose were generated, indicating the pentasaccharide can bind to subsites extending -3 to +2 or -2 to +3. The observation that the cellotriose produced is predominantly a single anomer (Figure 4.6) indicates that the pentasaccharide primarily occupies subsites -3 to +2.

The ~20 fold difference in k_{cat}/K_M when cellopentaose and cellohexaose are used as substrates can be attributed to the +3 subsite contributing ~2.0 kcal mol^{-1} to catalysis.

*Ct*Cel119 shows no similarity to any known glycoside hydrolase families and therefore its mechanism of action needed to be established. This was done by HPLC using the method first described by Braun et al (Section 2.3.2.2). At equilibrium glucose exists as 61:39 β anomer:α anomer (Stoddart 1971), irrespective of whether the sugar is a monosaccharide or the reducing end of a polymer, evident in Figure 4.6A, where cellotriose displays the expected anomeric ratio. Cellopentaose was treated with *Ct*Cel119 as the reaction generates the products cellobiose and cellotriose, one of which would have the anomeric configuration generated solely from hydrolysis by *Ct*Cel119. Cellohexaose was not used as two molecules of cellotriose would be produced, one with a new reducing end, the product of hydrolysis, and the other from the original cellohexaose molecule which would already be in equilibrium between the α and β anomer, making the stereochemical outcome of the reaction harder to assign. Figure 4.6B shows that the initial product produced after 10 min is the α anomer of cellotriose. This suggests *Ct*Cel119 performs catalysis via an inverting mechanism utilising a general acid, to provide protonation of the leaving group, and a general base, which activates a water molecule that attacks the anomeric carbon from below the plane of the pyranose ring. Figure 4.6C shows the reaction boiled immediately after 10 min incubation with *Ct*Cel119. It can be seen the cellotriose has undergone anomeric equilibration.

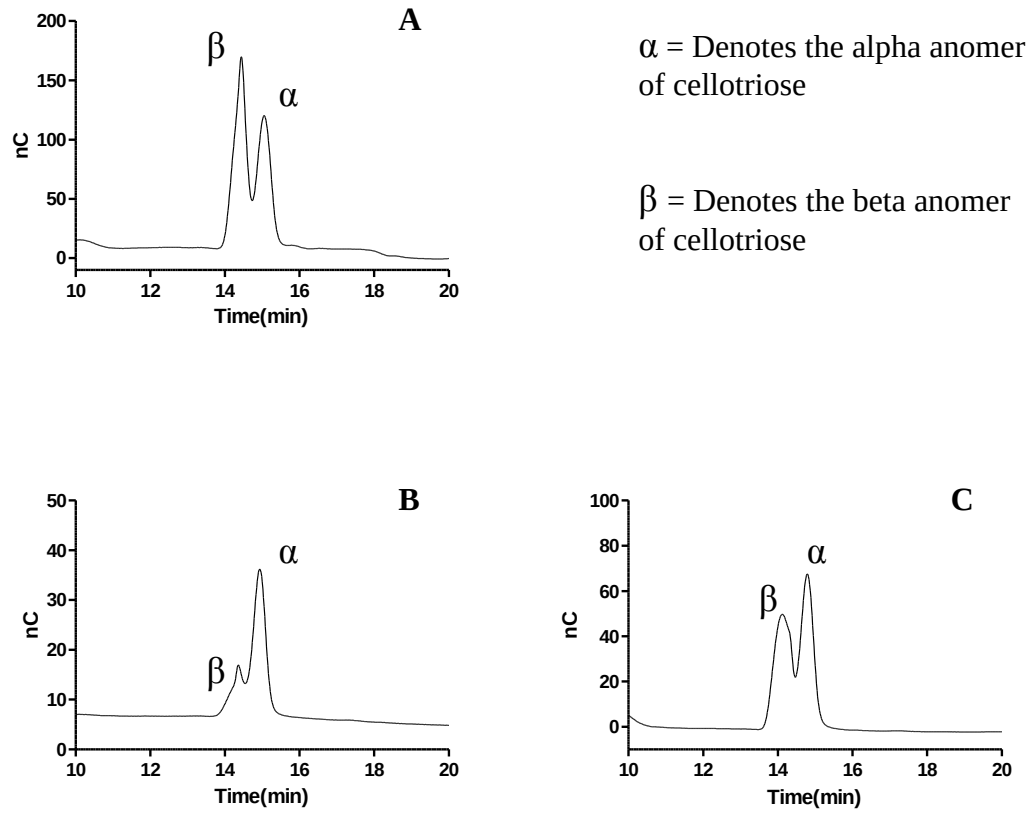


Figure 4.6 HPLC chromatograms showing the anomeric configuration of products generated by *Ct*Cel119 hydrolysis of cellopentaose.
A-cellotriose standard 100 µM. B-cellopentaose (1 mM) incubated with *Ct*Cel119 (3 µM) after 10 min. C-The product from B was boiled immediately after the 10 min enzyme substrate incubation.

4.3.3 Affinity of *Ct*Cel119 for cellulose and cellooligosaccharides

The affinity of the inactive catalytic acid mutant of *Ct*Cel119, E96A, (determined through crystallographic and biochemical analysis which is discussed later) was assessed by isothermal titration calorimetry (ITC) and depletion binding isotherms. Examples of the ITC traces of E96A against cellohexaose and regenerated cellulose (RC) are shown in Figure 4.7A and B. Figure 4.7C shows an ITC trace of *Ct*CBM3a, which was performed to compare the binding of E96A to a protein known to bind crystalline cellulose. Figure 4.7D shows the depletion isotherm of E96A against RC. All data from the binding experiments are shown in Table 4.2. E96A binds to cellohexaose and RC at 40 °C with K_A values of 1.5 x 10^4 M^{-1} and 1.39 x 10^5 M^{-1}, respectively (Table 4.2). This demonstrates E96A binds to RC with a ~10 fold higher affinity than to cellohexaose. This is confirmed by depletion

	Isothermal titration calorimetry						Depletion binding isotherms	
Ligand	**Temperature (°K)**	**K_A (M^{-1})**	**ΔG (kcal mol^{-1})**	**ΔH (kcal mol^{-1})**	**TΔS (kcal mol^{-1})**	**N**	**K_A (M^{-1})**	**N_0 (μmol g^{-1})**
RC	303	2.45 (±0.3) x 10^5	-7.5	-20.0 (±3.30)	-12.5 (±3.3)	1.04 ± 0.04	-	-
RC	313	1.39 (±0.4) x 10^5	-6.9	-40.2 (±6.8)	-33.3 (±7.5)	1.02 ± 0.05	3.9 (±0.5) x 10^5	17.4 (±1.0)
RC	323	1.02 (±0.4) x 10^5	-7.5	-83.1(±5.5)	-75.6 (±31)	0.96 ± 0.02	-	-
RC	333	1.03 (±0.5) x 10^5	-7.6	-79.5 (±5.4)	-71.9 (±9.1)	0.99 ± 0.01	-	-
Cellohexaose	283	5.5 (±0.4) x 10^4	-6.1	-7.4 (±0.3)	-1.3 (±0.3)	0.93 ± 0.03	-	-
Cellohexaose	293	4.2 (±0.1) x 10^4	-6.2	-8.2 (±0.1)	-2.0 (±0.4)	0.91 ± 0.02	-	-
Cellohexaose	303	3.1 (±0.1) x 10^4	-6.2	-8.1 (±0.2)	-1.9 (±0.2)	1.01 ± 0.04	-	-
Cellohexaose	313	1.5 (±0.1) x 10^4	-6.0	-8.9 (±0.7)	-2.9 (±0.7)	1.07 ± 0.11	-	-

Table 4.2 Binding data of E96A to RC and cellohexaose
Binding of E96A (88.2 μM) to cellohexaose (2 mM); Cellohexaose was titrated into protein. Binding of E96A (835 μM) to RC (29.4 mg/ml); E96A was titrated into RC. Depletion isotherms were performed by using a concentration of 1 mg/ml RC and 1–30 μM E96A. All reactions were carried out in 50 mM MES pH 5.5, containing 20 mM NaCl.

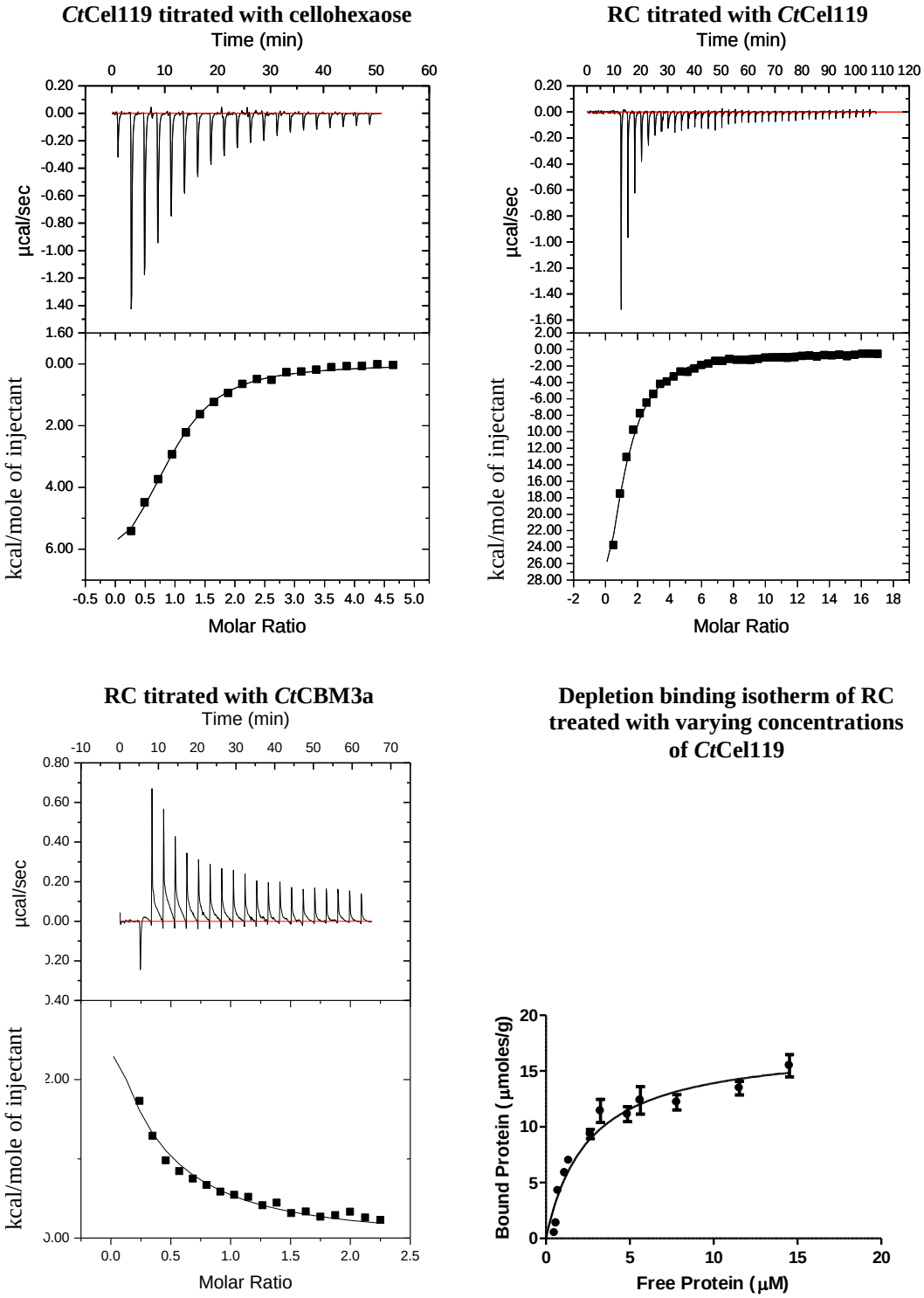


Figure 4.7 ITC and depletion binding isotherms if *Ct*Cel119 and *Ct*CBM3a binding to RC and cellohexaose.
For ITC using RC (29.4 mg/ml) protein (835 µM E96A and 200 µM *Ct*CBM3a) was titrated into the cellulose. For ITC using cellohexaose (2 mM), the oligosaccharide was titrated into E96A (88.2 µM). For binding isotherms 1 mg/ml of RC was used with concentrations of E96A ranging from 1–30 µM. All experiments using E96A were performed in 50 mM MES pH 5.5, 20 mM NaCl. Experiments with *Ct*CBM3a were performed in 10 mM Tris-HCl pH 8.0. All data shown was produced at 40 °C, except *Ct*CBM3a which was performed at 60 °C.

isotherms which were performed at 40°C and gave a K_A value of 3.9 x 10^5 M^{-1}; the three fold difference in determined affinities is small for two very different methods. The data from the ITC show that the binding of E96A to both cellohexaose and RC is enthalpathically driven and incurs an entropic cost. This cost is greater for the binding to RC than cellohexaose. This is in stark contrast to the data observed for *Ct*CBM3a, which shows that the interaction of the protein with RC is endothermic, shown by the positive heat inflections in Figure 4.7C. The data were not of an appropriate quality to fit to a binding curve. The positive heat inflections suggest that binding is entropically driven with an enthalpic cost.

The main reason the depletion binding isotherms were conducted was to obtain an accurate number of available binding sites per gram of cellulose. This was because the ITC experiment, in which RC was titrated with E96A, was not sufficiently sigmoidal to generate a reliable stoichiometry. The value obtained from the depletion isotherms of 17 µmoles g^{-1} was used to determine the concentration of E96A binding sites on the RC. The concentration of RC used in the ITC experiments was 29.4 mg/ml which, using the stoichometry obtained by depletion isotherms (17 µmoles g^{-1}), yields a concentration of 512 µM available binding sites. This value yielded a stoichiometry of ~0.03. Thus, it would appear that ~40 fold more E96A binds to RC in the depletion binding isotherm, compared to the ITC experiment. This is in contrast to the cellohexaose ITC experiments (Figure 4.7A), which did yield a stoichiometry of 1. Thus, although there is a discrepancy between the stoichiometry values, there is a good correlation of the K_A values.

Against RC the ΔH halved as the temperature was increased 10 °K, except at 333 K where the ΔH remained similar to the value obtained at 323 °K. This gave Δ*C*p of -2.02 kcal mol^{-1} for 303 °K–313 °K, -4.29 kcal mol^{-1} for 313 °K–323 °K and 0.36 kcal mol^{-1} for 323 °K–333 °K. Against cellohexaose the changes in ΔH with change in temperature were small. The Δ*C*p for cellohexaose were as follows, 283 °K–293 °K -0.08 kcal mol^{-1}, 293 °K – 303 °K 0.01 kcal mol^{-1}, 303 °K–313 °K -0.08 kcal mol^{-1}.

4.3.4 Synergy with the major exo cellulase *Ct*Cel48S

*Ct*Cel48S is the major exo acting cellobiohydrolase in the cellulosome of *C. thermocellum* (Guimaraes et al. 2002). It is well established that endo-acting enzymes can greatly potentiate the action of cellobiohydrolases (Gal et al. 1997; Parsiegla et al. 2008). In general it is thought endo-acting enzymes attack amorphous regions of the

cellulose and create new reducing and non-reducing ends, which cellobiohydrolases can utilise (Gal et al. 1997; Parsiegla et al. 2008). This can be thought of as increasing the effective substrate concentration available to cellobiohydrolases.

Encouragingly, *Ct*Cel119 shows synergy when used in combination with *Ct*Cel48S against Avicel. *Ct*Cel48S contains a dockerin which enables the enzyme to interact with CBM3a that is appended to a cohesin (Carvalho et al. 2007). The synergy experiments were performed by **Mrs Yael Vazana from the laboratory of Professor Ed Bayer**. *Ct*Cel119 showed little or no activity against Avicel when used as a discrete enzyme. Fusion to the crystalline cellulose binding CBM3a failed to potentiate the activity of *Ct*Cel119. In contrast *Ct*Cel48S displayed significant activity against Avicel, which was not significantly potentiated when fused to CBM3a. When *Ct*Cel48S was used in conjunction with *Ct*Cel119 a modest potentiation of *Ct*Cel48S, by 1.3 fold was observed compared to *Ct*Cel48S alone. Fusing CBM3a to the two enzymes resulted in a 2 fold increase in sugar production when the cellulases were combined in the same reaction (Figure 4.8).

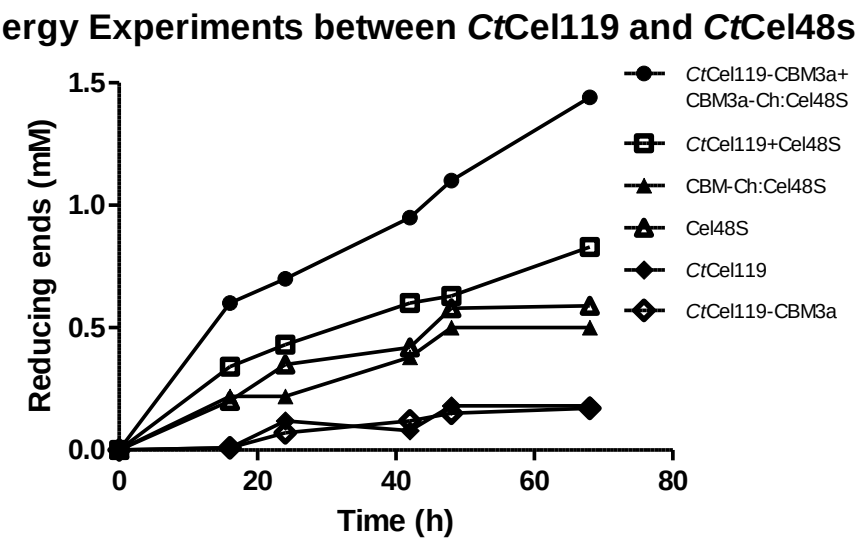


Figure 4.8 Kinetics of Avicel hydrolysis by *Ct*Cel119 and *Ct*Cel48S
CBM-Ch = *Ct*CBM3a; Ch = type 1 cohesin. *Ct*Cel48s contains a dockerin and *Ct*CBM3a contains a cohesion. When the two are mixed together they interact non-covalently through a cohesion dockerin interaction. Graphs was plotted in GraphPad Prism 4.02.

4.3.5 Crystal Structure

The structure of *Ct*Cel119 bound to two molecules of cellotriose was solved using single wavelength anomalous dispersion. Crystals, obtained by **Joana Brás from the laboratory of Professor Carlos Fontes and the structure was solved by Dr Ana Luisa Carvalho**, were grown using hanging drop diffusion in the condition 20 % PEG 3350, 8 % tascimate pH 5.0. The hanging drops contained 10 mM cellohexaose (subsequently hydrolysed to two molecules of cellotriose).

*Ct*Cel119 is an α-helical protein containing 8 α-helices comprised of the following residues: α-helix 1 K43–A51, α-helix 2 V55–D77, α-helix 3 W70–C77, α-helix 4 K82–S97, α-helix 5 E145–Y150, α-helix 6 E153–K173, α-helix 7 R177–T189 and α-helix 8 Q197–N217 (Figure 4.9). α-helix 4 contains mainly hydrophobic residues and comprises the core of the protein. The other seven helices then spiral around this hydrophobic core helix and thus collectively display a helical topology. The fold displayed by *Ct*Cel119 could be described as a α_8 superhelical fold (van Asselt et al. 1999).

The surface representation displayed in Figure 4.10 reveals two different topologies for the positive and negative subsites. The +1 to +3 subsites present a shallow cleft that interacts with the cellotriose. The cellotriose bound at these subsites displays a linear conformation with a two-fold screw axis typical of crystalline cellulose. In the negative subsites the cellotriose adopts a twisted conformation, consistent with the deep narrow cleft displayed by the negative subsites.

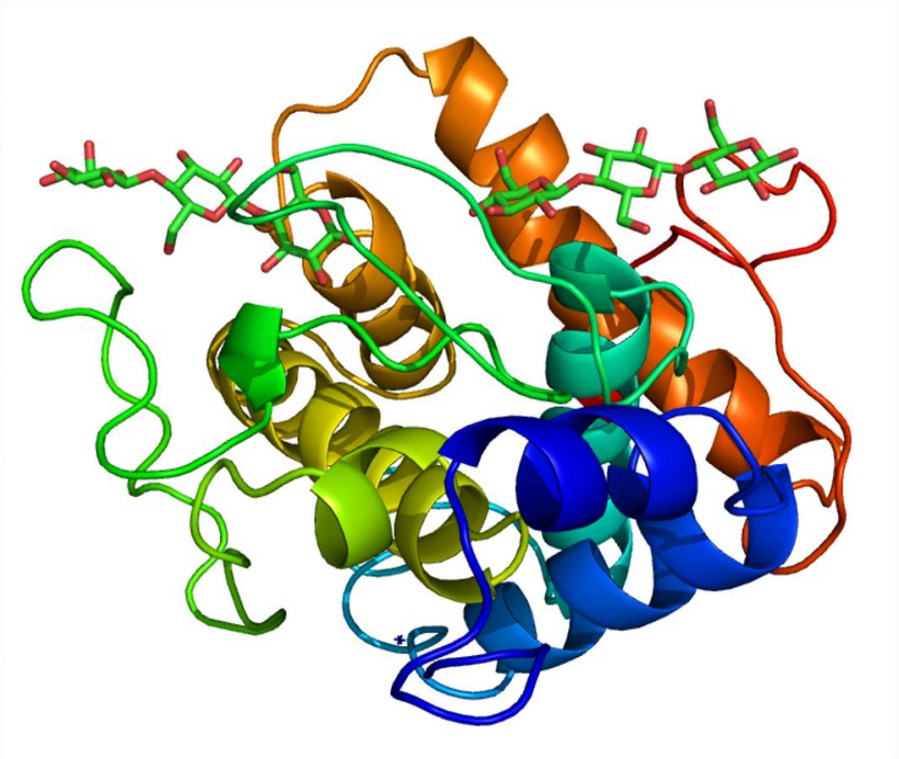


Figure 4.9 Cartoon of *Ct*Cel119
Two molecules of cellotriose bound +3 to +1 and -2 to -4. Colour is ramped blue N-terminal to red C-terminal.

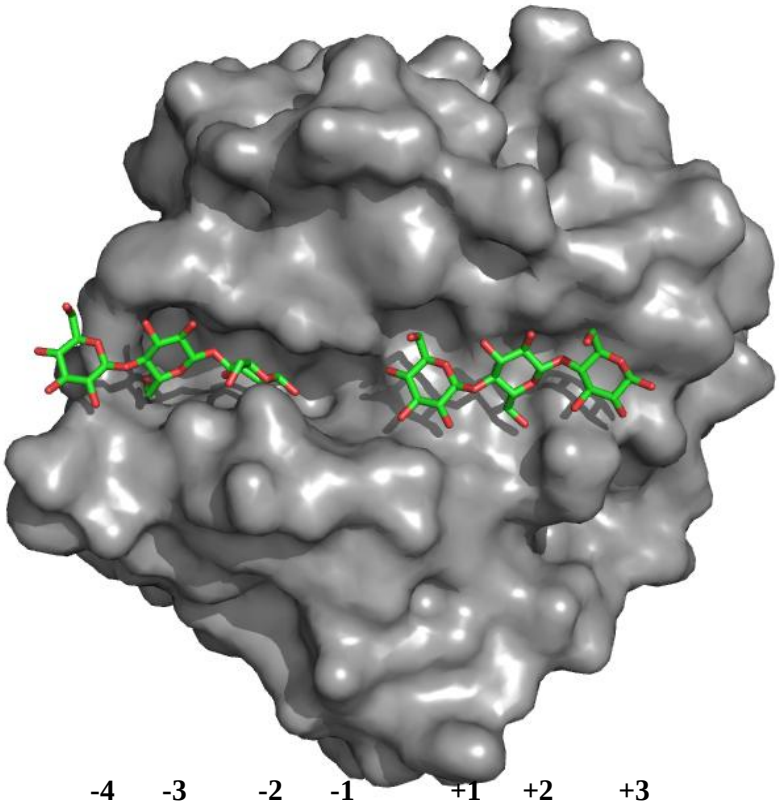


Figure 4.10 Surface representation of *Ct*Cel119
The cellotriose bound +1 to +3 displays a flat linear conformation. The cellotriose bound -2 to -4 adopts a twisted conformation.

4.3.6 Subsite interactions

4.3.6.1 Positive Subsites

At the +3 subsite R100 forms hydrogen bonds with both O2 and O3 via Nη1 and Nη2 respectively. T98 hydrogen bonds with O3 or O6 of the +2 glucose through Oγ. At the +2 subsite R95 hydrogen bonds to O3 via Nη2 and D228 hydrogen bonds to O2 via Oδ2. At the +1 subsite E96, the catalytic acid (see below), hydrogen bonds to O4 via Oε1 and the backbone carbonyl of E96 hydrogen bonds to O2. T105 interacts with O3 through its Oγ group. Q117 hydrogen bonds to O3 via Nε1 and finally N188 hydrogen bonds with O4 via Nδ2. (Figures 4.11 and 4.12)

4.3.6.2 Negative subsites

At the -2 subsite S110 hydrogen bonds through Oγ with either O6 or the endocyclic ring oxygen. This hydrogen bond is weak, indicated by the alternate conformation of the side chain seen in the crystal structure, and through mutagenesis, where mutants of the amino acid had little effect on catalytic activity (Table 4.3). T129 interacts with O3 via its backbone carbonyl. Y187 completes the interactions at the -2 subsite and hydrogen bonds to O2 via its backbone carbonyl. At the -3 subsite M131 hydrogen bonds to O6 via its backbone amine group, T189 hydrogen bonds to O2 via its backbone carbonyl and W191 makes polar contacts with O2 or O3 via its indole nitrogen. At the putative -4 subsite the only interaction is between the backbone carbonyl of N129 and O2 (Figures 4.11 and 4.12).

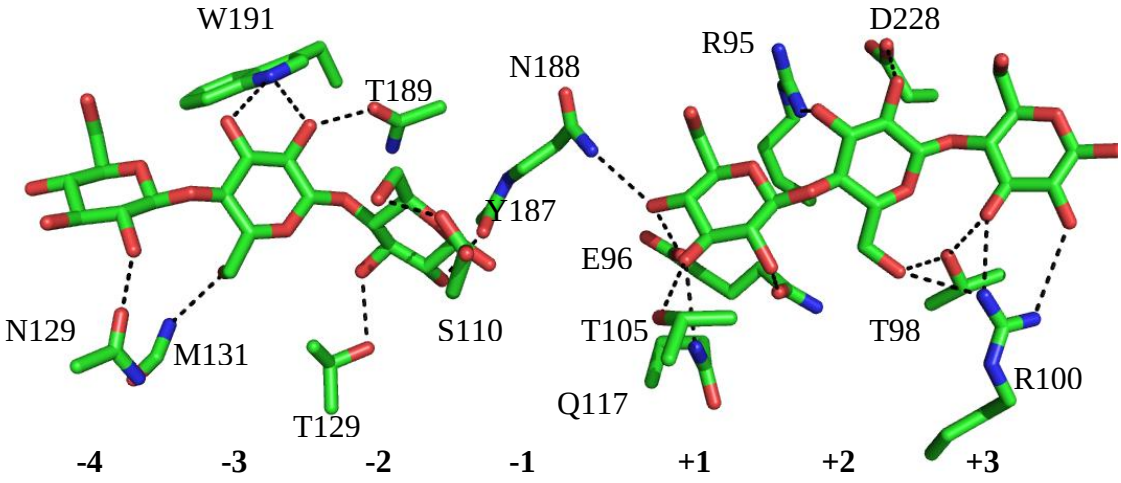


Figure 4.11 *Ct*Cel119 bound to two molecules of cellotriose with the protein-ligand interactions detailed.

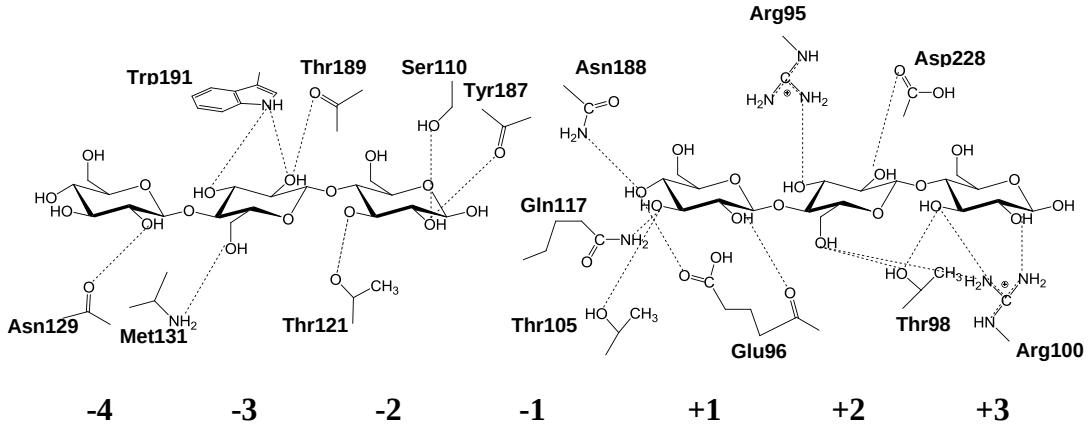


Figure 4.12 Schematic representation of the protein-ligand interactions of *Ct*Cel119 with two molecules of cellotriose

4.3.6.3 Active site centre of CtCel119 -1/+1

The -1/+1 subsite houses the catalytic machinery of glycoside hydrolases (Davies et al. 1995; Henrissat et al. 1995). Despite the absence of a glucose molecule at the -1 subsite, the identification of the catalytic acid, E96, was clear. In the crystal structure E96 lies beneath the α face of the +1 sugar and hydrogen bonds to O4. This is the perfect position to donate a proton to O4 of the +1 sugar, which would be required for departure of the cleaved sugar involved in the scissile glycosidic bond. Mutation of E96 inactivated *Ct*Cel119 and the E96A mutant displayed near wild type catalytic efficiency against the substrate 2,4 dinitrophenyl-cellotriose (Table 4.3), a

substrate that does not require protonic assistance at pH 5.5 as the leaving group 2,4 dinitrophenolate, has a pKa ~3.5 (MacLeod et al. 1994; Tull et al. 1994).

Detailed examination of the crystal structure did not show an appropriately positioned residue that could act as the catalytic base. The pH profile of *Ct*Cel119 (Figure 4.13), generated by **Miss Joana Brás from the laboratory of Professor Carlos Fortes,** had only a descending limb confirming the presence of an ionisable group with a pKa of 6.8, which is likely to be the pKa of the catalytic acid. The ascending limb was unable to be determined due to complete loss of activity below pH 4.5, which reflects precipitation (unfolding) of the protein at low pH. N188 and S110 were explored as possible candidate catalytic bases. Mutation of N188 to Ala significantly lowered the catalytic efficiency of the enzyme but it retained activity, and thus the Asn is unlikely to be the catalytic base (Table 4.3). S110 was mutated to Ala, Asp and Glu. Selection of this residue as a candidate base was based on structural similarity of *Ct*Cel119 to GH23 lysozymes (discussed below) where acidic residues equivalent, to S110, were believed to act as the catalytic base (Karlsen et al. 1996). The mutants S110A, S110D and S110E displayed catalytic efficiencies ~2-3 fold below that of wild type. This difference is very small and precludes S110 playing any significant role in catalysis (Table 4.3).

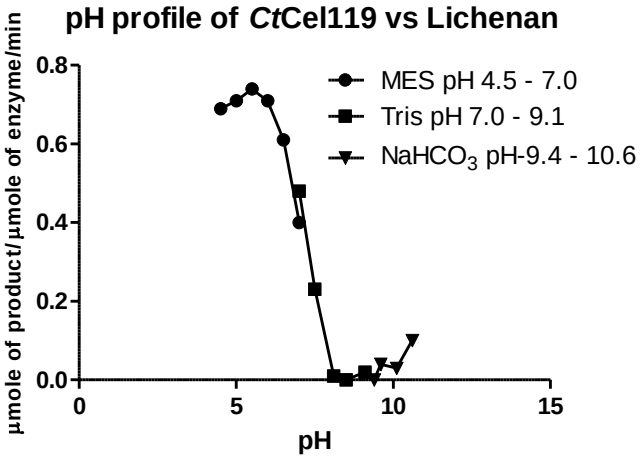


Figure 4.13 pH profile of *Ct*Cel119 versus lichenan.
All reactions were performed at 60 °C over a period of 1 hour using 1.7 nM *Ct*Cel119 and 2.5 mg/ml lichenan. All buffer concentrations were at 50 mM. Generated by Miss Joana Brás from the laboratory of Professor Carlos Fontes.

*Ct*Cel119	Substrate	Specific activity (min^{-1})	k_{cat}/K_M (min^{-1} M^{-1})
Wild type	Cellohexaose	-	$1.01 \times 10^4 \pm 2.0 \times 10^2$
Wild type	β-glucan	$6.4 \times 10^3 \pm 2.0 \times 10^2$	-
Wild type	2,4 DNP cellotriose	-	$1.11 \times 10^2 \pm 3.6 \times 10^0$
E96A	2,4 DNP cellotriose	-	$7.87 \times 10^1 \pm 1.2 \times 10^1$
E96A	β-glucan	NA	-
E96A	Cellohexaose	-	NA
N188A	Cellohexaose	-	$1.61 \times 10^1 \pm 3.9 \times 10^0$
S110A	Cellohexaose	-	$4.1 \times 10^3 \pm 1.00 \times 10^2$
S110E	Cellohexaose	-	$5.5 \times 10^3 \pm 6.61 \times 10^2$
S110D	Cellohexaose	-	$3.2 \times 10^3 \pm 3.7 \times 10^2$

Table 4.3 Catalytic activity of *Ct*Cel119 and variants.
NA = No activity detected; 2,4 DNP cellotriose = 2,4 dinitrophenyl cellotriose. All data was generated using non-linear regression analysis and errors are standard errors generated from triplicate results.

4.3.7 Structural similarity of *Ct*Cel119 to GH23

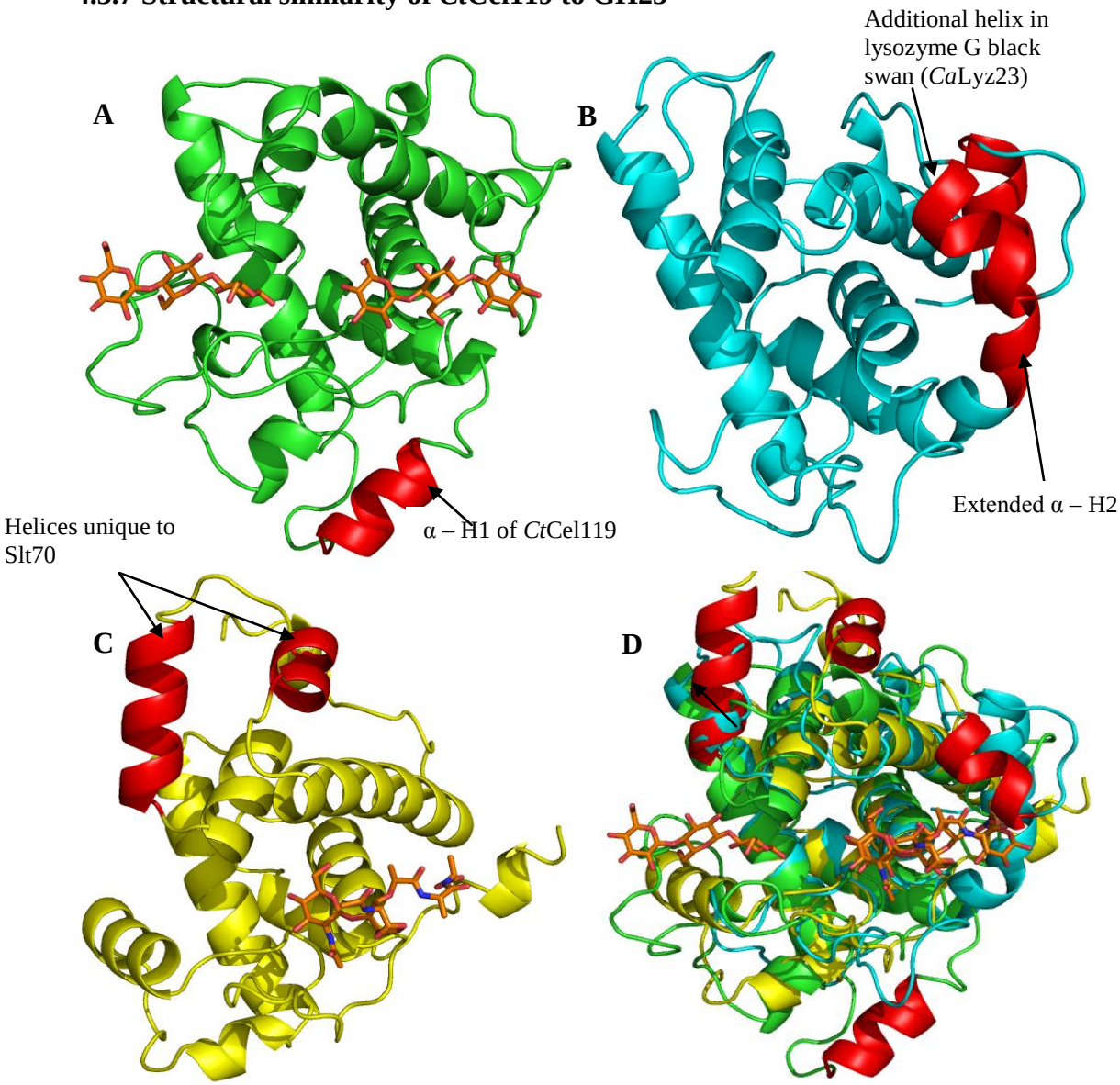


Figure 4.14 Cartoon representations of *Ct*Cel119 and two members of GH23.
A-*Ct*Cel119 (green) with ligand (ornage); B-lysozyme G black swan (CaLyz23) (cyan) (pdb 1LSP); C-Slt70 lytic transglycosylase (yellow) with ligand (orange) (1QSA) and D-All three cartoon overlaid. Helices unique to each protein are highlighted in red.

An SSM search (http://www.ebi.ac.uk/msd-srv/ssm) was performed to see if any structural homologs of *Ct*Cel119 were in the pdb database. It was hoped that any structurally related proteins may be able to shed light on a candidate for the catalytic base. The closest structural homologs were found in GH23, in particular lysozyme G black swan, which only had a sequence identity of 22 % but displayed an r.m.s.d of 1.89 Å over 110 residues. Figure 4.14 shows two members of GH23, the black swan lysozyme G (*Ca*Lyz23) and Slt70 lytic transglycoslase from *E .coli*. GH23 displays the same α-helical fold as *Ct*Cel119 and many of the helices are invariant between the

two GH23 structures and *Ct*Cel119. α–H2 to α–H6 are completely conserved between the three proteins and comprise the core structure. α–H1 is unique to *Ct*Cel119, while *Ca*GH23 contains an additional helix comprising C18–E24. This helix would clash with glucose molecules at the positive subsites of *Ct*Cel119 and provides some insight into the substrate discrimination of the two enzymes. The protein also has an extended α–H2. Slt70 has two unique helices (Figure 4.14).

At the critical +1/-1 subsites the similarity between the three enzymes is almost completely conserved (Figure 4.15). The catalytic acid E96 in *Ct*Cel119 is equivalent to E73 in *Ca*Lyz23 and E478 in Slt70; Q117 can be superimposed to Q95 in *Ca*Lyz23 and Q496 in Slt70; Y187 is analogous to Y147 in *Ca*Lyz23 and Y552 in Slt70; N188 is equivalent to N148 in *Ca*Lyz23 and N553 in Slt70; Y203 corresponds to Y169 in *Ca*GH23 and Y587 in Slt70. The only residue that shows variability in the three structures is S110 in *Ct*Cel119, in *Ca*Lyz23 this residue appears to be analogous to D97 while Slt70 presents M498 at this position.

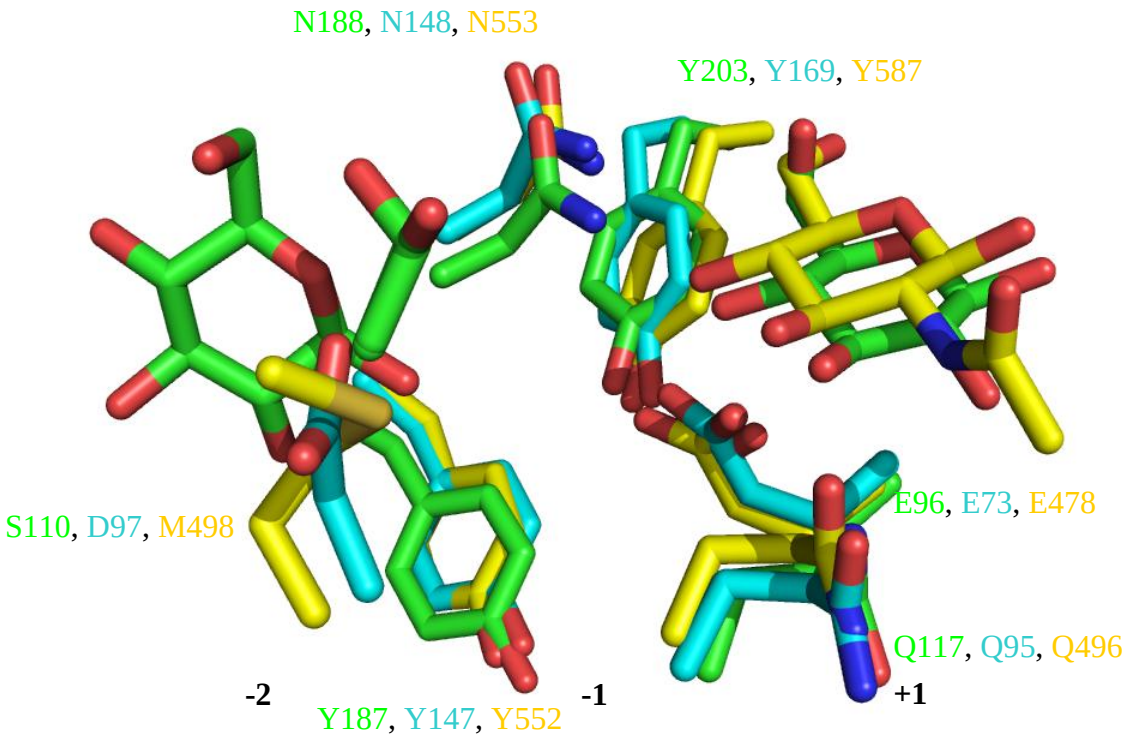


Figure 4.15 Overlay of the +1/-1 subsites of *Ct*Cel119, *Ca*Lyz23 and Slt70.
*Ct*Cel119 (green), CaLyz23 (cyan) and Slt70 (yellow). There is almost complete conservation at the +1/-1 subsites except at position S110 with the serine adopting alternate conformations.

The most interesting point of note is that despite the near absolute conservation of *Ca*Lyz23 and Slt70 at the +1/-1 subsites, they perform glycosidic bond cleavage by different mechanisms. *Ca*Lyz23 operates via a single displacement mechanism (Karlsen et al. 1996) while Slt70 is a lytic transglycoslase (van Asselt et al. 1999). The lytic transglycosylation reaction does not consume water and substrate-assisted nucleophillic attack contributes to catalysis. Here the catalytic residue E478, equivalent to E96 in *Ct*Cel119 (Figure 4.15), first acts as a general acid donating its proton to the glycosidic oxygen. As bond cleavage occurs, at the oxocarbenium ion like transition state, the acetyl group attached at C2 of the N-Acetylmuramic acid (MurNAc) attacks C1 to form an oxazolinium intermediate. The deprotanated E478 then acts as a general base abstracting a proton from the C6 hydroxymethyl group, allowing O6 to attack C1 from above generating a 1,6 anhydro product, with concomitant cleavage of the oxazolinium ring. No water is consumed in this reaction and the sugar is inverted to a $^{1}C_{4}$ chair, a general lytic transglycosylation mechanism is shown in Figure 4.16 (van Asselt et al. 1999).

By contrast lysozymes of GH23 do not form a 1,6 anhydro product but operate via an inverting mechanism, with D87 positioned to act as a catalytic base (Figure 4.17) (Karlsen et al. 1996). Thus, it is possible that GH23 members, which have an acidic residue equivalent to D87, operate via an inverting mechanism while those that lack a candidate catalytic base function as lytic transglycosylases, such as Slt70. The active site of *Ct*Cel119 resembles the lytic transglycoslase with S110 being situated where an acidic residue should be. The fact that α-cellotriose is the major product produced from hydrolyisis of cellopentaose is inconsistent with this mechanism. Also the lytic transglycoslation mechanism requires the formation of an oxazolinium intermediate, which in turn requires acetylation at C2. Glucose is not acetylated and *Ct*Cel119 is not active on the substrates chitin, chitosan or chitohexaose (data not shown), which are comprised of N-acetylglucosamine and contain differing degrees of acetylation at C2. For lytic transglycoslation to occur *Ct*Cel119 would need to facilitate the ‘same side’ attack of O6 at C1, which is highly unfavourable, and then hydrolyse the 1,6 anhydro product on enzyme to produce α-cellotriose. To explore the unlikely possibility that *Ct*Cel119 is indeed a lytic transglycoslase the molecule 1,6 anhydro cellotriose was synthesised following the method of Tanaka *et al* (Tanaka 2009) (Section 2.3.6.6). The production of 1,6 anhydro cellotriose was confirmed by

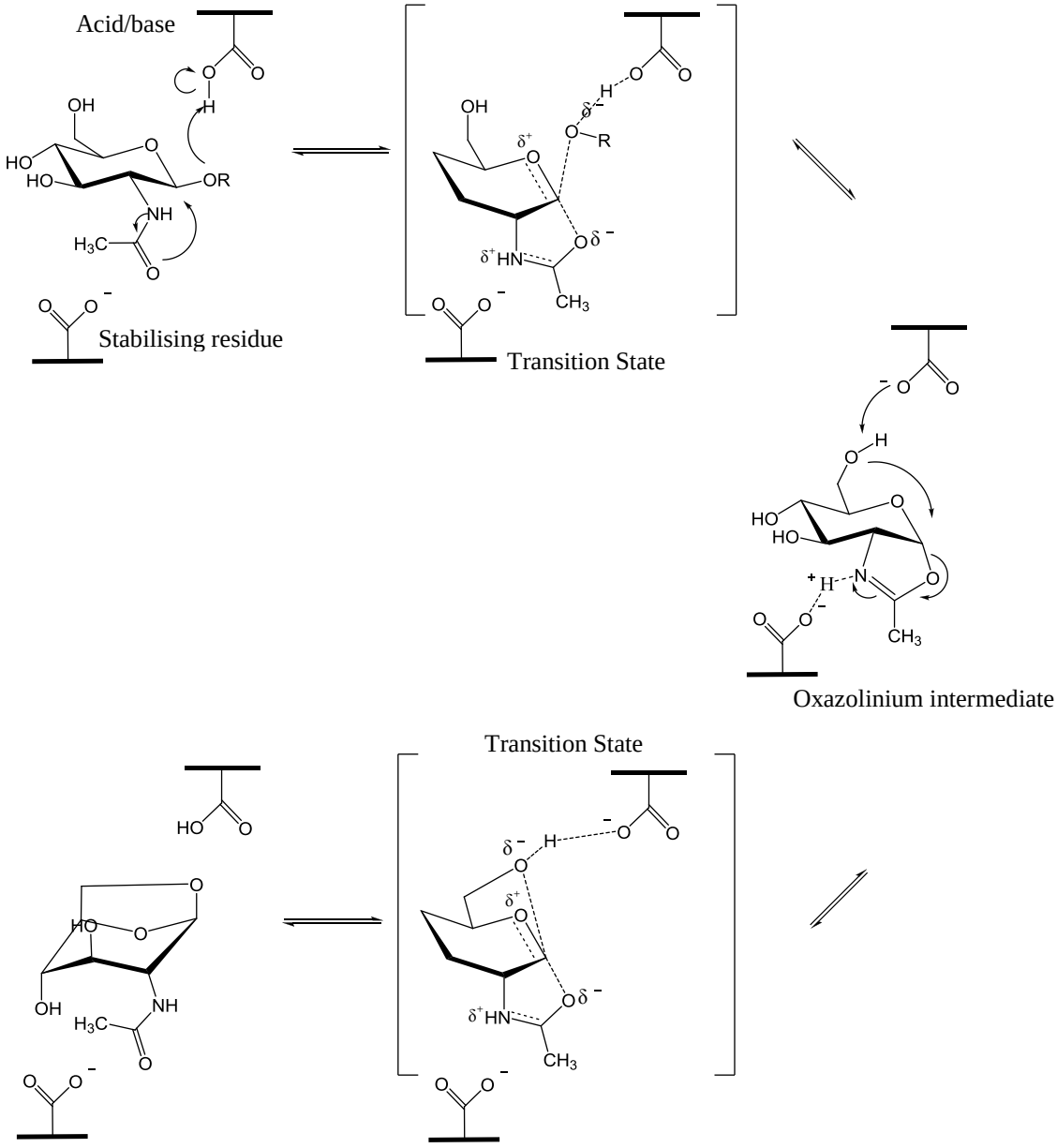


Figure 4.16 Schematic of the lytic transglycosylation mechanism.

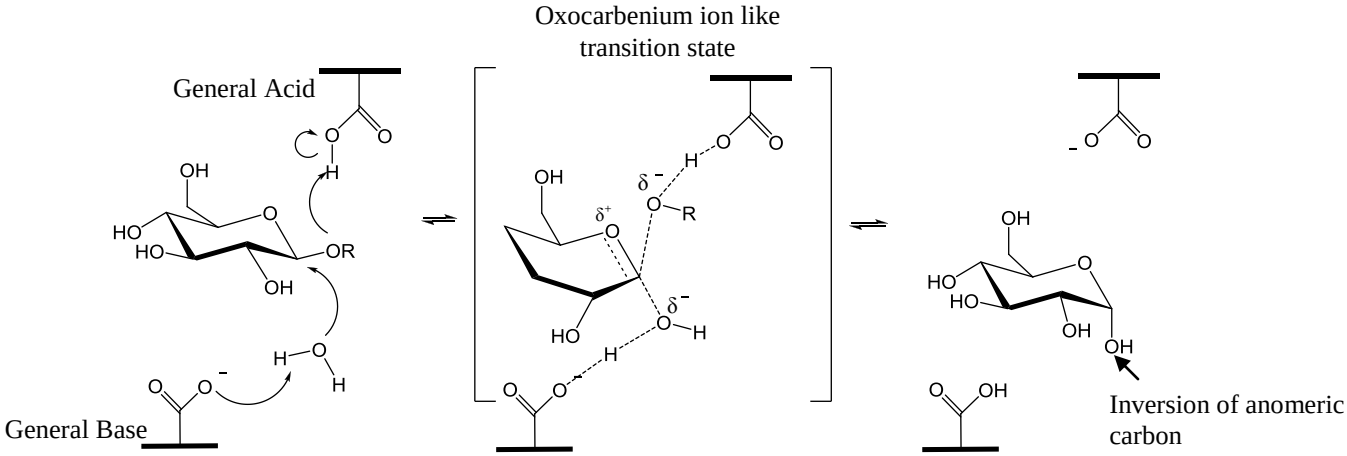


Figure 4.17 General inverting mechanism for β–glycosidases.

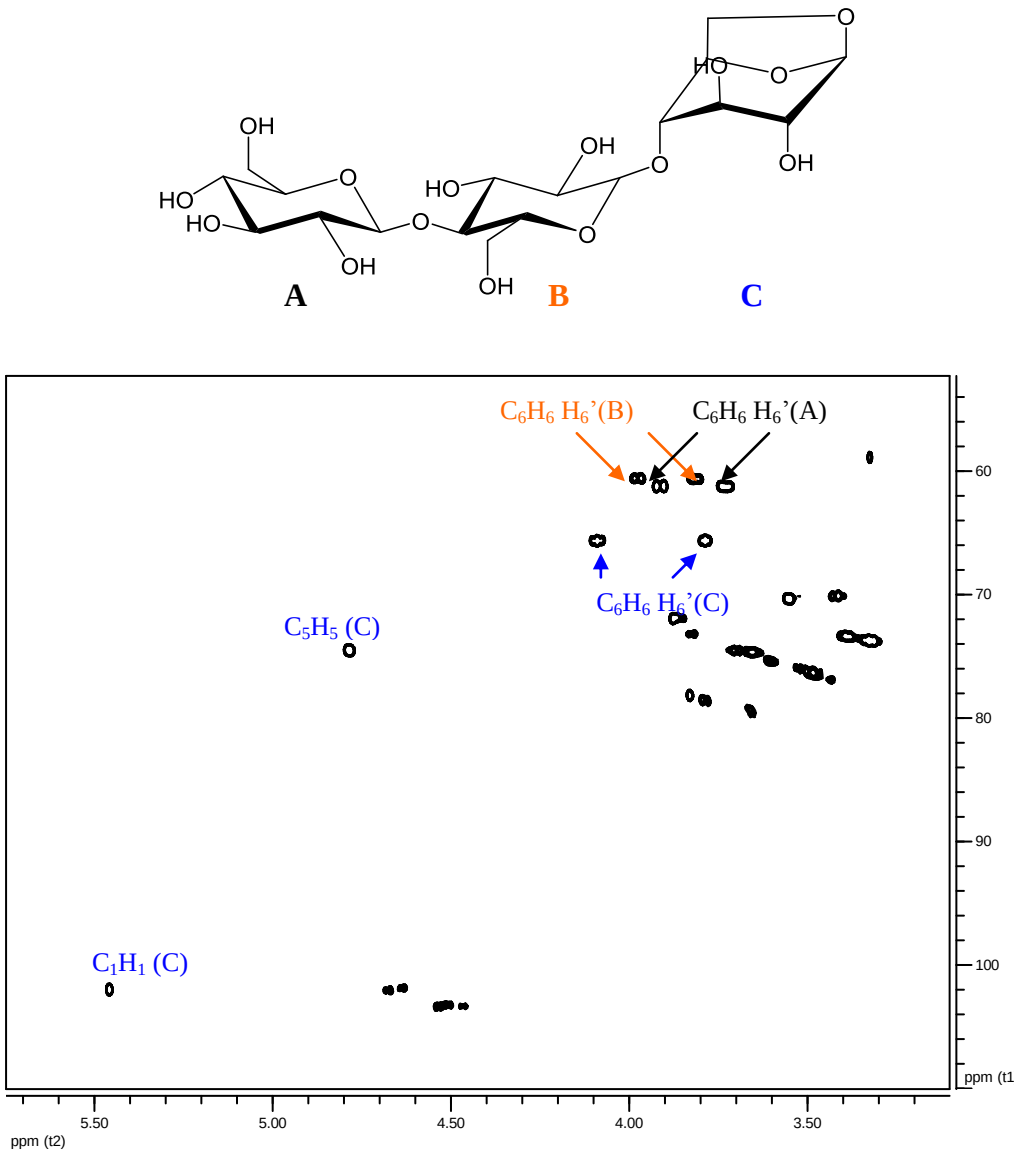


Figure 4.18 gHSQC spectrum of 1,6 - anhydro cellotriose
Black and orange arrows identify signals from β-1,4 glucose and blue identifies the presence of 1,6 anhydro glucose

gHSQC 2-D NMR, which was performed by **Dr Maria Pena from the laboratory of Dr Will York** (Figure 4.18). If *Ct*Cel119 operates as a lytic transglycosylase then the enzyme should be able to hydrolyse 1,6 anhydro cellotriose to cellotriose. *Ct*Cel119 was incubated with 1,6 anhydro cellotriose for 4 h and the reactions analysed by TLC. No cellotriose was observed confirming the enzyme cannot hydrolyse 1,6 anhydro cellotriose and it is not a lytic transglycosylase (Figure 4.19). As controls *Ct*Cel119 was simultaneously incubated with cellohexaose and cellohexaose with 1,6 anhydro cellotriose present. Both reactions produce cellotriose when analysed by TLC (Figure 4.19). These data support the view that *Ct*Cel119 is not a lytic transglycosylase. This led to the hypothesis that replacing S110 with an acidic residue, as in *Ca*Lyz23, might

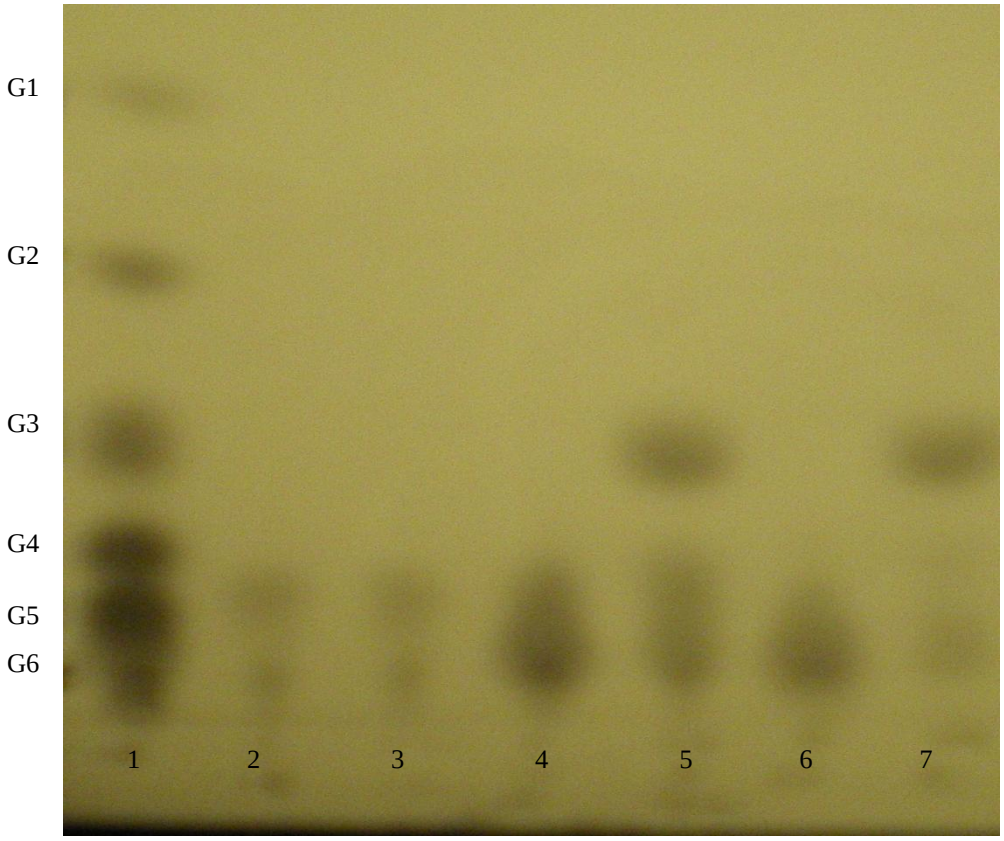


Figure 4.19 TLC of *Ct*Cel119 and various substrates.
1 = Standards; 2 = 1,6 anhydrocellotriose control; 3 = 1,6 anhydrocellotriose incubated with *Ct*Cel119 for 4 hours; 4 = Cellohexaose control; 5 = Cellohexaose incubated with *Ct*Cel119 for 4 hours; 6 = Cellohexaose and 1,6 anhydrocelltriose control; 7 = Cellohexaose and 1,6 anhydrocellotriose incubated with *Ct*Cel119 for 4 hours. All sugars were at 1 mM and 7 µM *Ct*Cel119 was used. Reactions were done in 50 mM MES pH 5.5 with 20 mM NaCl at 60 °C. G1 = Glucose; G2 = Cellobiose; G3 = Cellotriose; G4 = Cellotetraose; G5 = Cellopentaose and G6 = Cellohexaose

enhance catalysis. Mutation of S110 to Ala had little effect on catalysis and mutation of the Ser to Asp and Glu resulted in very small decreases in catalytic efficiency (Table 4.3). This precludes a significant role for S110 in catalysis

4.3.8 Crystallisation of E96A with Cellohexaose

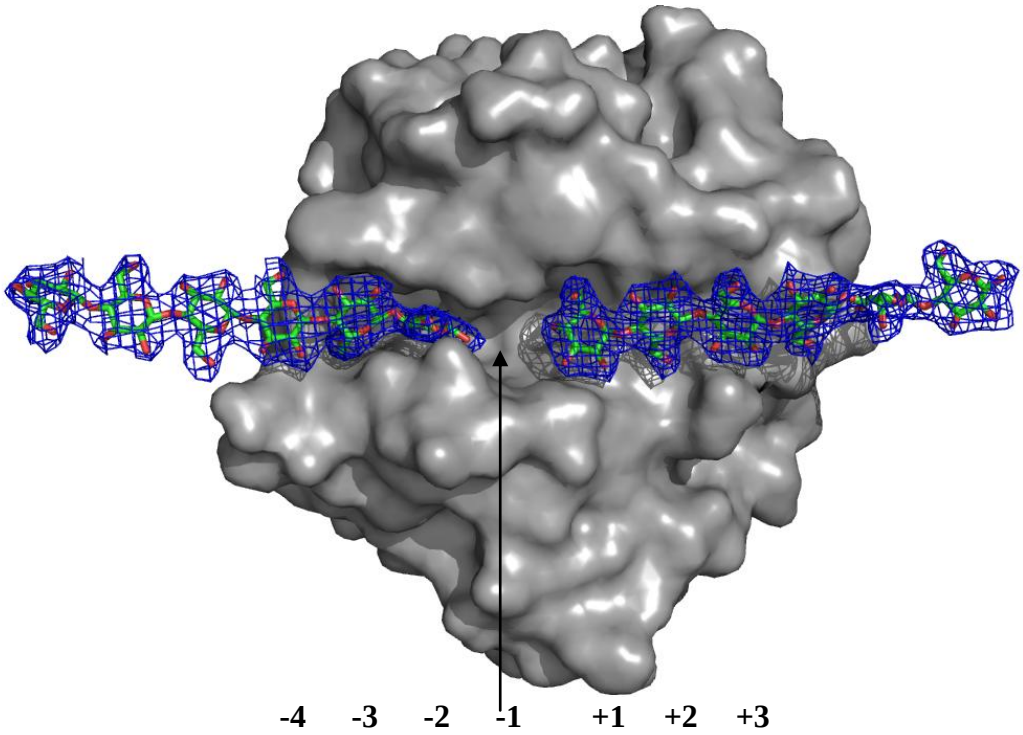


Figure 4.20 Surface representation of E96A in complex with two molecules of cellohexaose.
The $2F^o$-F^c electron density map is contoured at 1 σ around the cellohexaose molecule and clearly shows two molecules of cellohexaose either side of the -1 subsite.

	E96A-Cellohexaose
Data collection	
Space group	$P3_22_1$
Cell dimensions (Å)	a = b = 70.22, c = 105.51
(°)	α = β = 90° γ = 120°
Resolution (Å)	39.85-2.8 (2.95-2.65)
Mean $I/\sigma I$	5.7 (3.8)
R_{merge}(%)	36(50)
Completeness (%)	100 (100)
Redundancy	12.1 (12.4)
Refinement	
R_{work}/R_{free}	24/30
RMSD bond lengths (Å)	0.017
RMSD bond angles (°)	2.20

Table 4.4 Data collection and refinement statistics for E96A-Cellohexaose.

The catalytic acid mutant, at 120 mg/ml, was crystallised in 20 % PEG 3350 and 8 % tascimate pH 6.0 in the presence of 5 mM cellohexaose in the hope of mapping the conformation of the substrate bound at the critical -1 subsite. The crystals were extremely small and not of the same quality as the wild type selenomethionine derivative crystals grown previously. Data was indexed and the structure of the E96A mutant solved by molecular replacement. The structure was partially refined and data collection and refinement statistics are given in Table 4.4. The structure revealed that two molecules of cellohexaose spanned the active site cleft but did not bind to the critical -1 subsite. Each molecule of cellohexaose actually spanned two molecules of *Ct*Cel119 in the crystal lattice running -2 to -4 in one molecule and +1 to +3 in the next (Figure 4.20). Only one molecule of *Ct*Cel119 is shown for clarity.

4.4 Discussion

*Ct*Cel119 is an endo-acting cellulase that constitutes a new glycoside hydrolase family and, although it shares structural conservation with GH23, there is little sequence conservation with this family. *Ct*Cel119 is the only cellulase known to display an α_8 superhelical fold (Henrissat et al. 1996; Henrissat et al. 1997). The enzyme utilises six significant subsites to perform catalysis with cellohexaose predominantly binding to subsites +3 to -3, yielding two molecules of cellotriose. The small amount of cellotetraose and cellobiose seen (Figure 4.5) together with the observation that one of the cellotriose molecules binds from -2 to -4, may indicate the presence of a weak -4 subsite. This putative subsite makes one interaction with a glucose molecule, however the very low amount of cellotetratose and cellobiose produced from cellohexaose suggests that this interaction is weak. No cellobiose is produced during the initial stages of the reaction when *Ct*Cel119 hydrolyses PASC, only oligosaccharides with a degree of polymerisation of 3 to 6 were observed. This product pattern not only shows that *Ct*Cel119 is endo-acting, but also demonstrates that substrate is required to occupy subsites -3 to +3 for hydrolysis to occcur. Indeed, cellopentaose, which occupies the subsites +2 to -3 (Figure 4.5B and 4.6B), is hydrolysed 20 times less efficiently than cellohexaose. Cellotetratose is not hydrolysed by *Ct*Cel119 and this indicates the requirement for occupation of +2 to -3 for catalysis. It cannot be elucidated whether it is the loss of binding at the +2 or -3

subsite that results in the inability of *Ct*Cel119 to hydrolyse cellotetraose. The activity of *Ct*Cel119 was highest against β–glucan, although it should be noted that the initial rate of PASC hydrolysis by *Ct*Cel119 was only four fold lower than that of β–glucan (Table 4.1). Generally the relative activity of endo-acting cellulases against these two glucans is much larger than four fold (Hall et al. 1995), suggesting *Ct*Cel119 can access insoluble forms of the cellulose. However, only a small proportion of PASC was hydrolysed during the initial stages, ~3 %, indicating that the bulk of the insoluble polysaccharide was inaccessible to the cellulase. The enhanced rate against β–glucan likely reflects the accessibility offered by soluble substrates. The weak activity against CMC can be attributed to the low tolerance of methyl substitutions by *Ct*Cel119, observed previously for other cellulases.

It was shown that the inactive catalytic acid mutant, E96A, bound to RC with a 10 fold higher affinity than to cellohexaose (Table 4.2), which spans most of the active site. This may suggest that a putative -4 subsite accounts for this difference in affinity. The apparent lack of a significant contribution of the -4 subsite to catalysis, however, would not explain the ten fold increase in K_A. An alternate hypothesis is that *Ct*Cel119 binds to particular regions of RC tighter than it does to cellohexaose. It was observed that the active site cleft of *Ct*Cel119 appears to accommodate a linear conformation at the positive subsites, similar to that found in crystalline regions of cellulose, while at the negative subsites, a twisted conformation of the substrate is observed. In addition, the -1 subsite is also expected to display a twisted conformation. These negative subsites may therefore bind to more disordered regions of the cellulose (Figure 4.10). This led to the hypothesis that *Ct*Cel119 may target the boundaries between crystalline and paracrystalline/amorphous cellulose. Cellohexaose would have to be linear to bind to the positive subsites of *Ct*Cel119. Cellooligosaccharides, however, adopt a twisted or helical conformation in solution (Sergeyev et al. 2005), and thus a significant energetic penalty would be incurred when cellohexaose binds to the enzyme. The targeting of 'pre formed' structures in RC, which are complementary to the topology of the substrate binding cleft, may account for the ten fold increase in K_A.

The binding to both cellohexaose and RC is enthalpathically driven and incurs an entropic cost as seen previously for many protein carbohydrate interactions (Flint et al. 2005; van Bueren et al. 2005; Abbott et al. 2009). The binding of proteins to insoluble substrates has only been studied by ITC twice previously (Creagh et al.

1996; Boraston 2005). Creagh *et al* (Creagh et al. 1996) showed that binding of *Cf*CBM2a to bacterial microcrystalline cellulose (BMCC) was entropically driven, while Boraston (Boraston 2005) revealed binding of *Cc*CBM17 and *Cc*CBM28 to RC was enthalpically driven. The present study explores the binding of a cellulase to cellohexaose and cellulose. Binding to RC had a greater entropic cost than binding to cellohexaose, which was unexpected. RC is conformationally restricted, while cellohexaose is not. It would be predicted that the loss of conformational freedom of the cellohexaose, upon protein binding, would incur a greater entropic penalty and thus the result is hard to explain. This has resonance with the work by Boraston (Boraston 2005) where *Cc*CBM17 bound to the high affinity sites in RC with a greater enthalpy and entropic cost compared to cellohexaose (Boraston 2005). In contrast Creagh *et al* showed that the binding of *Cf*CBM2a to bacterial microcrystalline cellulose (BMCC) was entropically driven with very small negative heat inflections. Work presented here with *Ct*CBM3a, which could not be fitted to a binding model, showed a trace with positive heat inflections indicating that entropy would have to be positive and thus comprise the driving force for binding.

The binding of E96A to RC gave a negative heat capacity, while cellohexaose recognition resulted in a very small heat capacity compared to RC. A negative heat capacity was also observed by Boraston and Creagh *et al* (Creagh et al. 1996; Boraston 2005). A negative heat capacity is associated with dehydration effects of the protein and ligand, and hence a positive entropic contribution to binding. In the work by Creagh *et al* (Creagh et al. 1996) the negative heat capacity is consistent with the entropic driving force. The negative heat capacity observed with E96A is more difficult to understand. The increased entropic cost of binding RC compared with cellohexaose is not consistent with a lower conformational entropic cost, indicated by the respective heat capacities. The active site cleft of *Ct*Cell119 presents no obvious aromatic/hydrophobic interactions, and thus polar interactions dominate ligand recognition, again making a negative heat capacity hard to explain. Although dehydration effects are mainly associated with hydrophobic interactions, polar residues will also interact with water when ligand is absent. When ligand binds, the polar residues will no longer interact with water and it will be free to join bulk solvent causing a negative heat capacity to be observed. This would then be a favourable contribution to entropy but this is not observed in the ITC experiments. Although the K_A appears to be correct (ITC and binding isotherms agree) the exact magnitude of

ΔH and ΔS is only known if the stoichiometry is correct, and thus the thermodynamic basis for binding is uncertain.

ITC using cellohexaose as the titrate gave a stoichiometry of ~1 as expected from the crystal structure of the enzyme, which reveals a single binding site on the proteins surface. ITC experiments in which protein was titrated into RC gave a stoichiometry value much lower than observed by the binding isotherms. One explanation is that non-specific binding occurred during the binding isotherms, which is not observed by ITC. However, it seems unlikely that this "non-specific" binding, occurs, particularly as the affinity of these non-specific events would be similar to specific protein recognition revealed by ITC. Another possible explanation is that there are two classes of binding sites, one of which is heat neutral and therefore not observable by ITC except in a deviation from a stoichiometry of 1. Indeed some studies on protein carbohydrate recognition have been shown to be heat neutral by ITC, but tight binding was revealed by other methodologies (Gloster et al. 2007).

The idea that *Ct*Cel119 targets a pre-formed structure in cellulose, in particular the boundary between crystalline and paracrytsalline/amorphous regions is a tenuous but attractive hypothesis. Support for this view is provided by the observation that two very different topologies are observed in the enzymes substrate binding cleft. ITC data and PASC kinetics show a low concentration of available binding sites on the surface of the RC. These data support the targeting of rare and specific features in the cellulose, such as the boundary regions between crystalline and paracrystalline/amorphous cellulose. It should be noted, however, that the ITC data should be viewed with some caution as direct stoichiometry values could not be determined.

*Ct*Cel119 shows synergy with the major *C. thermocellum* cellobiohydrolase *Ct*Cel48S. This is similar to results described previously for the *C. cellulolyticum* system using *Cc*Cel48F and *Cc*Cel9G, where endo and exo synergy was observed (Fendri et al. 2009). The synergy was greatest when both enzymes where appended to *Ct*CBM3a, whereas fusion of the CBM to the enzymes acting individually had little effect. *Ct*CBM3a may help to target *Ct*Cel48S and *Ct*Cel119 to the same area on the cellulose chain and thus maximise synergy between the two enzymes.

The mechanism of *Ct*Cel119, at first glance, would be that of a simple inverting mechanism involving both a general acid and a general base. The position, mutation and kinetic properties of E96 lead to its unambiguous assignment as the

catalytic acid. In particular the acid mutant displays similar activity to the wild type enzyme against 2,4 dinitrophenyl-cellotriose. Seminal work by Withers (MacLeod et al. 1994; Tull et al. 1994) showed that for substrates with excellent leaving groups protonation by the catalytic acid is not needed, and this residue can be mutated with little or no effect on the overall $k_{cat}/K_{M.}$ This is indeed the observation seen here for E96A. No activity was observed against cellohexaose or β-glucan, confirming the importance of E96 in acting as the general acid against substrates with poor leaving groups. Its position below the α face of the +1 glucose, and its hydrogen bonding to O4, which would be the scissile glycosidic oxygen, also support the role of E96 as the catalytic acid.

The assignment of the catalytic base was not possible. No acidic residue lies in the vicinity of the -1 subsite that could act as the catalytic base. The only initial candidate was N188. Work with GH95, which operates via an inverting mechanism, showed that a catalytic Asn, which was activated by an Asp, may act as the general base (Nagae et al. 2007). N188 does not hydrogen bond to any acidic residue and the N188A mutant retained catalytic activity, inconsistent with a role as the general base. The lack of a candidate catalytic base has been described previously for inverting enzymes (Alzari et al. 1996; Davies 1997; Parsiegla et al. 1998; Varrot et al. 1999). In GH8 and GH48, which are both members of clan GH-M, a Tyr is the only residue that hydrogen bonds to the catalytic water and is likely to play a role in generating a hydroxyl ion from water (Alzari et al. 1996; Parsiegla et al. 1998). There is no obvious equivalent residue in *Ct*Cel119. In GH6 the identity of the catalytic base has also remained obscure and shares resonance with the work presented here (Davies 1997; Varrot et al. 1999). Several candidate residues have been suggested for some GH6 members, while a "Grotthus mechanism" is supported for other members. This mechanism allows the base to be at a distal location and generate a hydroxyl ion locally which is then transmitted via a solvent chain to the site of catalysis. *Ct*Cel119 may operate by such a mechanism but a candidate for the 'remote base' is hard to identify.

The catalytic acid mutant when crystallised with cellohexaose did not yield a structure with glucose at the -1 subsite. This was unfortunate but is not uncommon and there are many structures which do not allow the -1 subsite to be mapped. An explanation may lie in the fact that the -1 subsite of glycoside hydrolases has evolved to bind to the transition state. At the transition state bonds are simultaneously being

made and broken. This has the result of ‘stretching’ the bonds and the -1 subsite thus binds to an elongated substrate. The distance of O1 to O4 of glucose molecules in the distal subsites of *Ct*Cel119 is ~5.5 Å, while the distance between O1 of the -2 glucose and O4 of the +1 glucose is 6.05. Thus, binding of glucose at -1 in the ground state is not as favourable as at the other subsites, which have evolved to bind glucose in the ground state.

4.5 Conclusions

*Ct*Cel119 is an endo-acting cellulase that utilises six major subsites and can hydrolyse polymers of five residues or greater. It shares no sequence homology with any family in the CAZy database and as such constitutes a new family, GH119. *Ct*Cel119 shares structural homology with GH23 and as such GH119 and GH23 are probably members of the same clan. GH23 is currently not classified into a GH clan. The identity of the catalytic base remains unanswered and currently a ‘Grotthus mechanism’ seems the most plausible conclusion from the data. More investigation into the transition state is needed and the mechanism by which a hydroxyl ion is generated to attack C1.

Chapter 5

Analysis of Family 43 Glycoside Hydrolases from *Cellvibrio japonicus*

5.1 Introduction

Cellvibrio japonicus is a saprophytic, soil dwelling, gram negative bacterium. *C. japonicus* is an efficient plant cell wall degrading organism that expresses a battery of glycoside hydrolases, some of which have been extensively studied (DeBoy et al. 2008). The bacterium's genome encodes a predicted 124 glycoside hydrolases, based on sequence similarity to members of the glycoside hydrolase families (GH) within the CAZy database. The *C. japonicus* genome contains a particularly large number of genes encoding GH5, GH13 and GH43 enzymes (Figure 5.1). GH5 has an array of enzyme activities directed at β-glucans, β-mannans, β-galactans and β-xylans, providing a biological rationale for why this family has been expanded in *C. japonicus* (Henrissat 1991; Henrissat et al. 1997). To date 345 members of this family have been characterised and the crystal structure of 29 of these enzymes have been

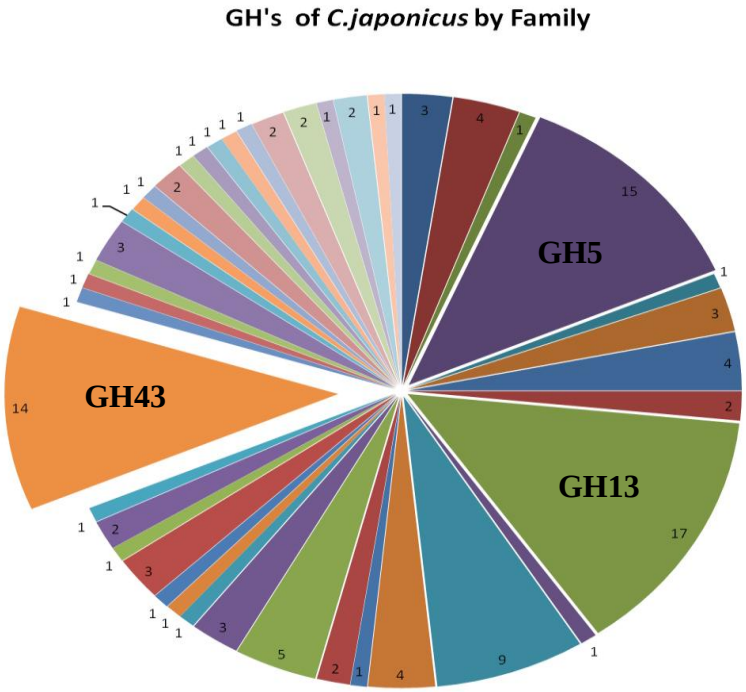


Figure 5.1 *C. japonicus* GHs grouped into families
The number of enzymes in each family are indicated. GH5, GH13 and GH43 are highlighted.

solved. The enzymes in GH13 mediate hydrolysis, phosphorolysis and remodelling (transglycosylation) of α-glucans, especially amylose a major component of starch. GH13 also hydrolyses a variety of linkages including α-1,4, α-1,6 and even α-1,1, found in trehalose. Starch is a polysaccharide that is readily utilised by bacteria as a primary carbon source explaining why there has been a large expansion of GH13 in *C. japonicus* and many other organisms. There are currently 668 characterised members of GH13 and the crystal structures of 72 enzymes have been reported (Henrissat et al. 1995).

GH43 is dominated by β-xylosidases, α-L-arabinofuranosidases and α-L-arabinanases, although the family also contains two β-1,3 galactanases. This represents a limited number of activities, and only 54 members, out of ~1200 sequences, have been characterised, while the crystal structure of 12 GH43 enzymes have been solved. Currently only one of the 14 *C. japonicus* GH43s has been characterised, *Cj*Arb43A, which was shown to be an endo-processive arabinanase (Nurizzo et al. 2002; Proctor et al. 2005). Within the remaining 13 *C. japonicus* GH43s there may be extensive functional redundancy, or the enzymes could display novel activities. GH43 is an inverting family of enzymes that utilise a catalytic triad of three acidic residues to perform catalysis. This triad comprises an Asp, acting as a general base, a Glu, acting as a general acid and a second Asp that modulates the pKa of the general acid (Nurizzo et al. 2002). The overall tertiary fold of GH43 is a five bladed β–propeller, the first enzymes shown to display this fold (Nurizzo et al. 2002). In work presented in this Chapter, 11 of the 13 *C. japonicus* GH43s were cloned, expressed in *E. coli* and their catalytic activities explored. The crystal structure of one of these enzymes was determined in complex with ligand.

5.2 Objectives

To investigate the biological rational of the expansion of the GH43 family in *C. japonicus*, utilising structure function studies to explore the enzymatic properties of the enzymes.

5.3 Results

5.3.1 Cloning

In total eleven genes encoding GH43 enzymes were amplified by PCR from *C. japonicus* genomic DNA (Figure 5.2A). The region amplified encodes the respective mature enzyme lacking its signal peptide. PCR amplified products (insert) and vector DNA were then cut with the same restriction enzymes, ligated and transformed into *E. coli* yielding ~100-500 colonies (Sections 2.1.10 and 2.1.19). Plasmid DNA was recovered from six colonies, from each ligation, and checked for the presence of insert by digesting the isolated DNA with appropriate restriction enzymes. Plasmids containing inserts (Figure 5.2B) were then sequenced. Typically ~70 % of plasmids sequenced contained the desired gene. All the plasmid constructs generated are detailed in Table 5.1.

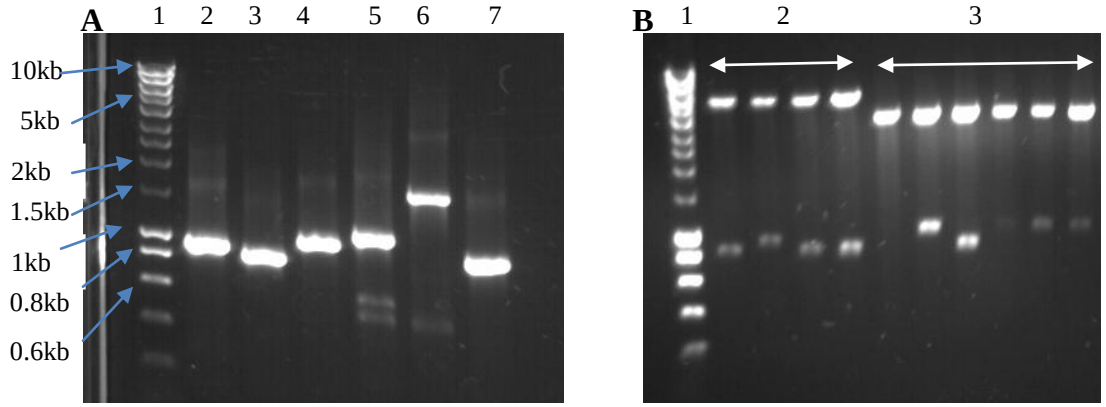


Figure 5.2 Agarsose gels of PCR cloned genes and digestion of plasmid DNA
A-PCR amplified products 1=Hyperladder 1 (Invitrogen); 2=CJA_0819; 3 =CJA_0816; 4=CJA_0799; 5=CJA_3012; 6=CJA_3594; 7=CJA_3018. B- Digest of cloned genes. 1=Hyperladder 1 (Invitrogen); 2 =CJA_3018 ligated with pET32b; 3=CJA_3018 ligated with pET28b.

Vector	Restriction sites	Plasmid	*C. japonicus* gene	Enzyme Name
pET32b	BamHI; XhoI	pAC1	CJA_0799	*Cj*Gly43E
pRSET	BamHI; HindIII	pAC2	CJA_0806	*Cj*Abf43L
pET32b	BamHI; XhoI	pAC3	CJA_0816	*Cj*Gly43C
pRSET	BamHI; HindIII	pAC4	CJA_0818	*Cj*Gly43D
pET32b	BamHI; XhoI	pAC5	CJA_0819	*Cj*Abf43M
pREST	BamHI; HindIII	pAC6	CJA_3007	*Cj*Gly43H
pET22b	NdeI; XhoI	pAC7	CJA_3012	*Cj*Abf43K
pET28b, pET32b	BamHI; XhoI	pAC8, pAC12	CJA_3018	*Cj*Abf43B
pRSET	BamHI; HindIII	pAC9	CJA_3070	*Cj*Xyl43A
pET28b	BamHI; XhoI	pAC10	CJA_3594	*Cj*Gly43B
pRSET	BamHI; HindIII	pAC11	CJA_3601	*Cj*Gly43I

Table 5.1 Gene numbers, vector, restrictions site, plasmid names, and soluble enzyme expression conditions.
The genes listed can be found in the CAZy database under GH43(http://www.cazy.org/GH43.html). The restriction sites listed were used to clone the GH43genes into the expression vectors. All recombinant proteins expressed in soluble form when being induced with 1 mM (BL21 DE3) or 100 µM isopropyl β-D-1-thiogalactopyranoside (Tuner) at 16 °C, at an A_{600nm}=0.6, and left for 16 hours. Recombinant protein expressed equally well in both BL21 (DE3) and Tuner expression strains.

5.3.2 Small scale expression tests

Plasmids with the correct insert sequence were tested for their ability to direct the synthesis of soluble recombinant protein. Small scale, 10 ml *E. coli* cultures were grown to mid-exponential phase and the recombinant gene expression was induced, through the addition of isopropyl β-D-1-thiogalactopyranoside (IPTG), at 37 °C and 16 °C as described in Materials and Methods Section 2.2.24. Cell free extracts were generated and analysed by SDS-PAGE (Figure 5.3). Conditions for generating the eleven GH43s in soluble form are listed in the legend of Table 5.1.

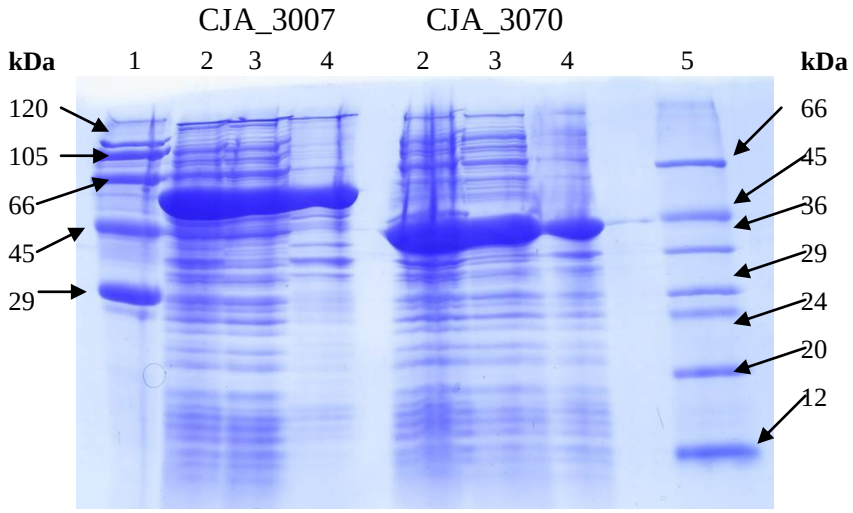


Figure 5.3 SDS-PAGE of small scale expression trials
Recombinant protein expression was induced at 16 °C in appropriate *E. coli* strain. 1=High molecular weight marker; 2=Whole cell; 3=Cell free extract; 4=Insoluble pellet; 5=Low molecular weight marker
.

5.3.3 Expression and purification of GH43 Enzymes

The 11 GH43 *C. japonicus* enzymes were expressed in soluble form in *E. coli* strain BL21 (DE3) containing the appropriate plasmid. The *E. coli* strains were cultured to mid-exponential phase (A_{600nm}=0.6) at 37 °C after which recombinant protein expression was induced by the addition of 1 mM IPTG and the cultures were incubated at 16 °C for 16 h (Section 2.2.24). Cells were harvested by centrifugation, cell free extracts prepared (Section 2.2.25), and immobilised metal affinity chromatography (IMAC) was used to purify the proteins to electrophoretic homogeneity. Examples of SDS-PAGE analysis of the IMAC purification process is shown in Figure 5.4 for two GH43 enzymes. SDS-PAGE analysis of the remaining nine GH43s is displayed in the Appendix C. The concentration of the proteins was calculated using the molar extinction coefficients of 84370 M^{-1} cm^{-1} for *Cj*Abf43B (Figure 5.4A) and 68760 M^{-1} cm^{-1} for *Cj*Xyl43A (Figuer 5.4B) at A_{280nm}, and their sizes of 35 kDa and 40 kDa, respectively, determined by SDS-PAGE, are similar to their calculated molecular weights based on the primary sequence of the mature proteins. The molecular masses and molar extinction coefficients of the remaining nine enzymes are listed in the Appendix C.

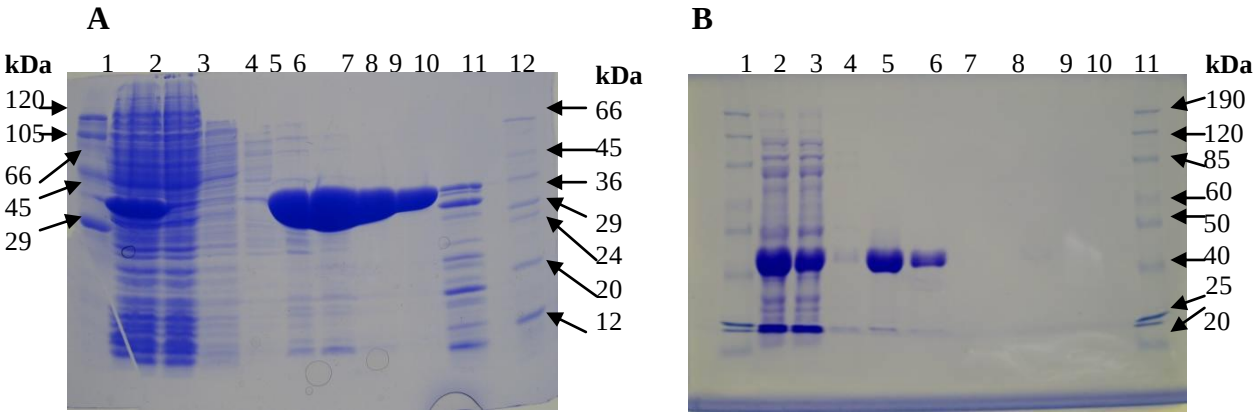


Figure 5.4 SDS-PAGE analysis of GH43 purification
A=*Cj*Abf43B in pET28b; 1 high molecular weight marker; 2 cell free extract;3 Flow through; 4 Talon buffer (see section 2.2.27 for composition) wash; 5 Talon buffer supplemented with 10 mM imidazole; 6-10 Talon buffer supplemented with 100 mM imidazole; 11 Insoluble pellet; 12 Low molecular weight marker. B=*Cj*Xyl43A Benchmark prestained protein ladder (Invitrogen) 1,11 ; 2 cell free extract;3 Flow through; 4 Talon buffer (see MM for composition) wash; 5 Talon buffer supplemented with 10 mM imidazole; 6-9 Talon buffer supplemented with 500 mM imidazole.

5.3.4 Biochemical characterisation *Cj*Abf43K and *Cj*Abf43L

Recombinant *Cj*Abf43K and *Cj*Abf43L showed activity against α-1,5-L-arabinofuranose oligosaccharides, linear arabinan and sugarbeet arabinan (Figure 5.5

and Table 5.2). Linear arabinan is a polysaccharide consisting of α-1,5-L-arabinofuranosides. Sugarbeet arabinan consists of a α-1,5-L-arabinofuranose backbone substituted with α-1,3 and α-1,2-L-arabinofuranose side chains. *Cj*Abf43K and *Cj*Abf43L showed no activity against the other plant cell wall polysaccharides tested including β-arabinoxylans, β-galactans, β-mannans and β-glucans.

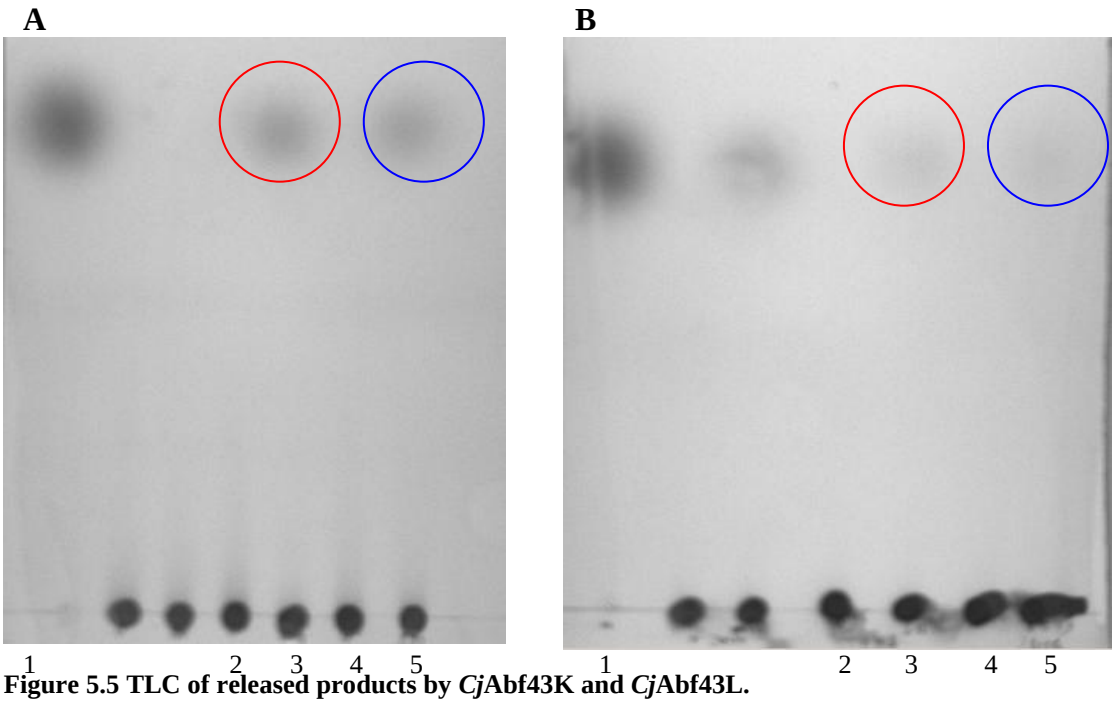


Figure 5.5 TLC of released products by *Cj*Abf43K and *Cj*Abf43L.
A – Activity against linear arabinan. 1=Arabinose standard; 2=5 mg/ml Linear arabinan control; 3=5 mg/ml Linear arabinan incubated overnight with *Cj*Abf43K; 3=5 mg/ml Linear arabinan control; 4=5 mg/ml Linear arabinan incubated overnight with *Cj*Abf43L. B – Activity against sugar beet arabinan. 1 = Arabinose standard; 2=5 mg/ml sugar beet arabinan control; 3=5 mg/ml Linear arabinan incubated overnight with *Cj*Abf43K; 4=5 mg/ml sugar beet arabinan control; 5=5 mg/ml sugar beet arabinan control incubated overnight with *Cj*Abf43L. All reactions were carried out at 37 °C in 50 mM sodium phosphate buffer pH 7.0. Red and blue circles highlight released arabinose by *Cj*Abf43K and *Cj*Abf43L from linear arabinan and its absence when sugar beet arabinan is used as the substrate.

*Cj*Abf43K has a pH and temperature optima of 6.6 and 39 ºC respectively (Figure 5.6C and D). The enzyme showed low activity against 4-nitrophenyl-α-L-arabinofuranoside, which may be indicative of important interactions with arabinose at the +1 subsite. Highest activity was observed against arabinobiose, no increase in catalytic efficiency was observed against arabinotetraose or arabinohexaose, indicating the enzyme has only two major subsites (Figure 5.6A and B and Table 5.2). *Cj*Abf43L has pH and temperature optima of 6.8 and 33 °C respectively (Figure 5.6G and H). The catalytic efficiency of the enzyme against 4-nitrophenyl-α-L-arabinofuranoside and arabinooligosacharides was similar, although individual kinetic

constants could not be calculated for the aryl glycoside. There was no significant increase in the catalytic efficiency of the enzyme against arabinobiose to arabinohexaose indicating the enzyme has only two major subsites (Figure 5.6E and F and Table 5.2). Both enzymes released arabinose from linear arabinan. Although no arabinose from sugar beet arabinan could be detected by TLC (Figure 5.5A and B), the generation of arabinose was observed spectroscopically and by HPLC (data not shown). The enzymes display a k_{cat}/K_M against sugarbeet arabinan, which is around ~30-60 fold below that for the arabinooligosaccharides and the total amount of arabinose released by *Cj*Abf43K and *Cj*Abf43L from sugarbeet arabinan was only 1.5 % and 1.9 % of the polysaccharide, respectively, which is very low (Table 5.2). These data are consistent with the view that the two enzymes target the α-1,5 linked backbone. In sugar beet arabinan the backbone is largely inaccessible to *Cj*Abf43K and *Cj*Abf43L, as the side chains create steric clashes with the two enzymes.

Enzyme	Substrate	K_M (mM)	k_{cat} (min^{-1})	k_{cat}/K_M (min^{-1} M^{-1})	Total arabinose release per mg (mg)
*Cj*Abf43K	4-nitrophenyl-α-L-arabinofuranoside	0.09 ± 0.02	0.63 ± 0.09	$7 \times 10^3 \pm 2.5 \times 10^3$	-
	Arabinobiose	0.34 ± 0.03	755 ± 100	$2.2 \times 10^6 \pm 4.8 \times 10^5$	-
	Arabinotetraose	1.08 ± 0.09	708 ± 19	$6.56 \times 10^5 \pm 6.9 \times 10^4$	-
	Arabinohexaose	1.10 ± 0.08	693 ± 11	$6.3 \times 10^5 \pm 5.7 \times 10^4$	-
	Sugar beet arabinan	-	-	$2.02 \times 10^4 \pm 1.11 \times 10^3$	0.015
*Cj*Abf43L	4-nitrophenyl Arabinofuranoside	-	-	$2.63 \times 10^4 \pm 1.96 \times 10^3$	-
	Arabinobiose	7.64 ± 1.27	220 ± 22.06	$2.88 \times 10^4 \pm 7.8 \times 10^3$	-
	Arabinotriose	3.39 ± 0.80	357 ± 18.8	$1.05 \times 10^5 \pm 3.0 \times 10^4$	-
	Arabinotetraose	3.97 ± 0.24	268 ± 11.01	$6.75 \times 10^4 \pm 6.75 \times 10^3$	-
	Arabinohexaose	4.33 ± 1.31	321 ± 41.69	$7.41 \times 10^4 \pm 3.2 \times 10^4$	-
	Sugar beet arabinan	-	-	$1.03 \times 10^3 \pm 2 \times 10^1$	0.019

Table 5.2 Kinetic values of *Cj*Abf43K and *Cj*Abf43L.
*Cj*Abf43K (2 µM) was used against 4-nitrophenyl α-L-arabinofuranoside (4-NPA) and 50 nM against arabinoligosaccharides. *Cj*Abf43L (100 nM) was used against 4-nitrophenyl α-L-arabinofuranoside and 100 nM against arabinoligosaccharides All reactions were performed at 25 ºC in 50 mM sodium phosphate buffer pH 7.0. Errors reported are standard errors from triplicate results.

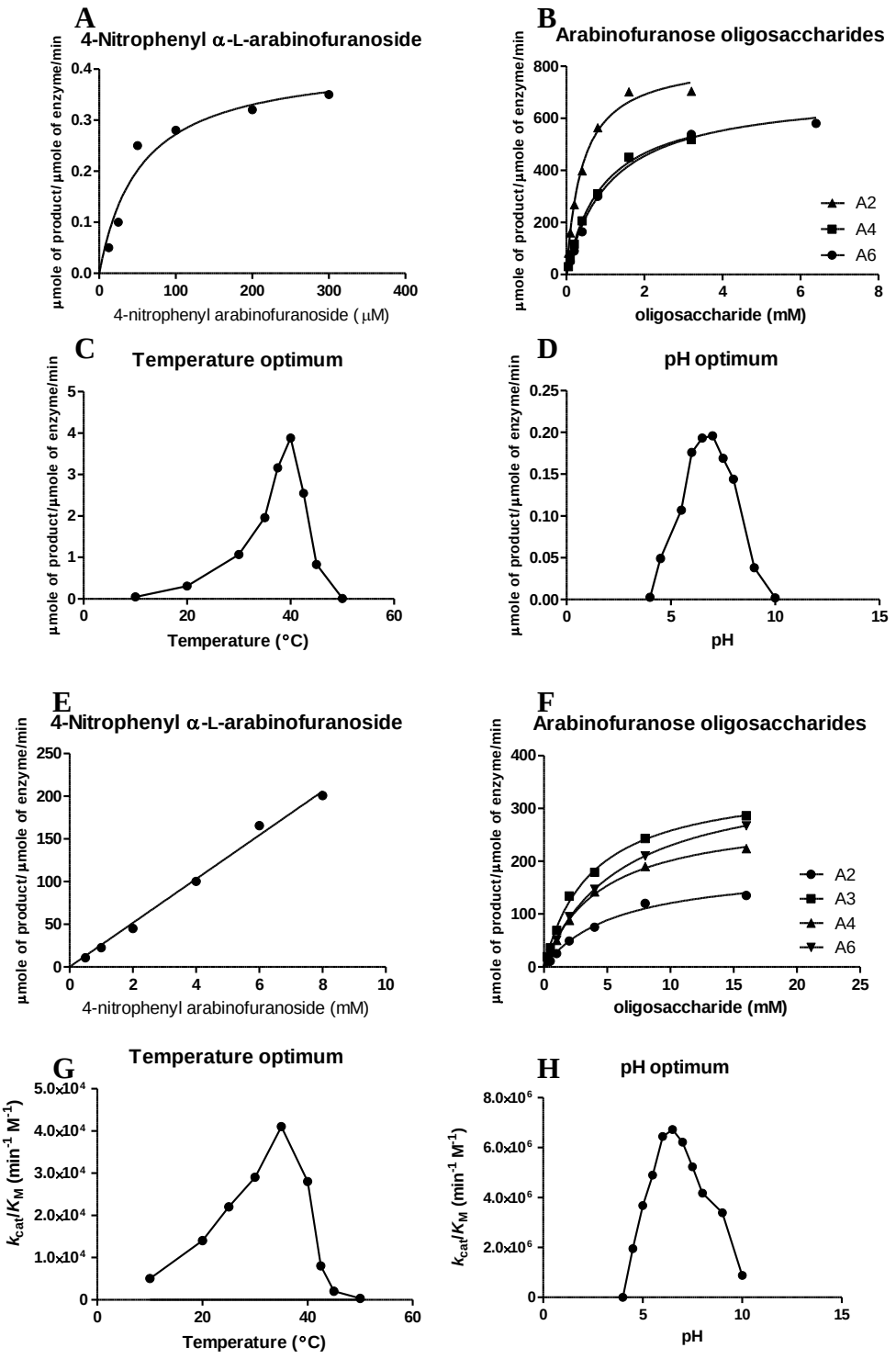


Figure 5.6 Graphs of kinetic data of *Cj*Abf43K and *Cj*Abf43L

A-*Cj*Abf43K (2 µM) vs 4-nitrophenyl-α-L-arabinofuranoside (4NP-AF). B-50 nM *Cj*Abf43K vs A2 (arabinobiose), A4 (arabinotetraose) and A6 (arabinohexaose). C and D-Temperature and pH optima 2 µM *Cj*Abf43K vs 1 mM 4-nitrophenyl Arabinofuranoside; E-100 nM *Cj*Abf43L versus 4-nitrophenyl arabinofuranoside; F-100 nM *Cj*Abf43L vesus A2, A3 (arabinotriose), A4 and A6; G and H-Temperature and pH optima were determined with 100 nM *Cj*Abf43L vs 1 mM (4NP-AF) . A, B, E and F were performed at 25 °C in 50 mM sodium phosphate buffer pH 7.0; C and G were performed at temperatures ranging from 10-50 °C in 50 mM sodium phosphate buffer pH 7.0; D and H were performed at 25 °C at pH ranging from 4.0-10.0. For pH 4.0-5.5 50 mM NaAcetate; pH 5.5-6.0 50 mM MES; pH 6.0-8-0 50 mM sodium phosphate and pH 9.0-10 50 mM CAPSO . Graphs were plotted in GraphPad Prism 4.02 using non-linear regression. Points were joined in graphs C,D,G and H.

5.3.5 Biochemical characterisation of *Cj*Xyl43A

Recombinant *Cj*Xyl43A (CJA_3070) showed activity against β-1,4 linked xylose oligosaccharides, wheat arabinoxylan, rye arabinoxylan and oat spelt xylan (Figure 5.7 and Table 5.3). *Cj*Xyl43A showed no activity against the other plant cell wall polysaccharides tested; α-L-arabinans, β-galactans, β-mannans and β-glucans.

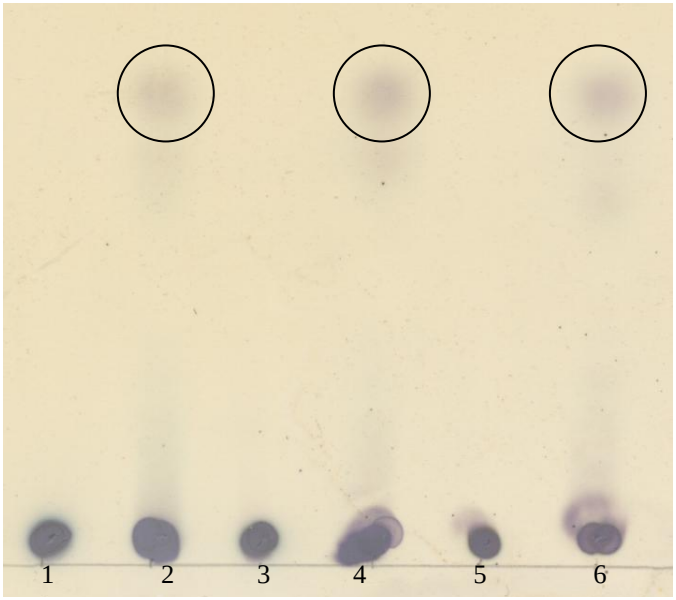


Figure 5.7 Thin layer chromatography of *Cj*Xyl43A versus xylans.
1 = 5 mg/ml wheat arabinoxylan control; 2 = 5 mg/ml wheat arabinoxylan incubated overnight with *Cj*Xyl43A; 3 = 5 mg/ml rye arabinoxylan control; 4 = 5 mg/ml rye arabinoxylans incubated overnight with *Cj*Xyl43A; 5 = 5 mg/ml oat spelt xylan control; 6 = 5 mg/ml oat spelt xylan incubated overnight with *Cj*Xyl43A. Xylose is circled.

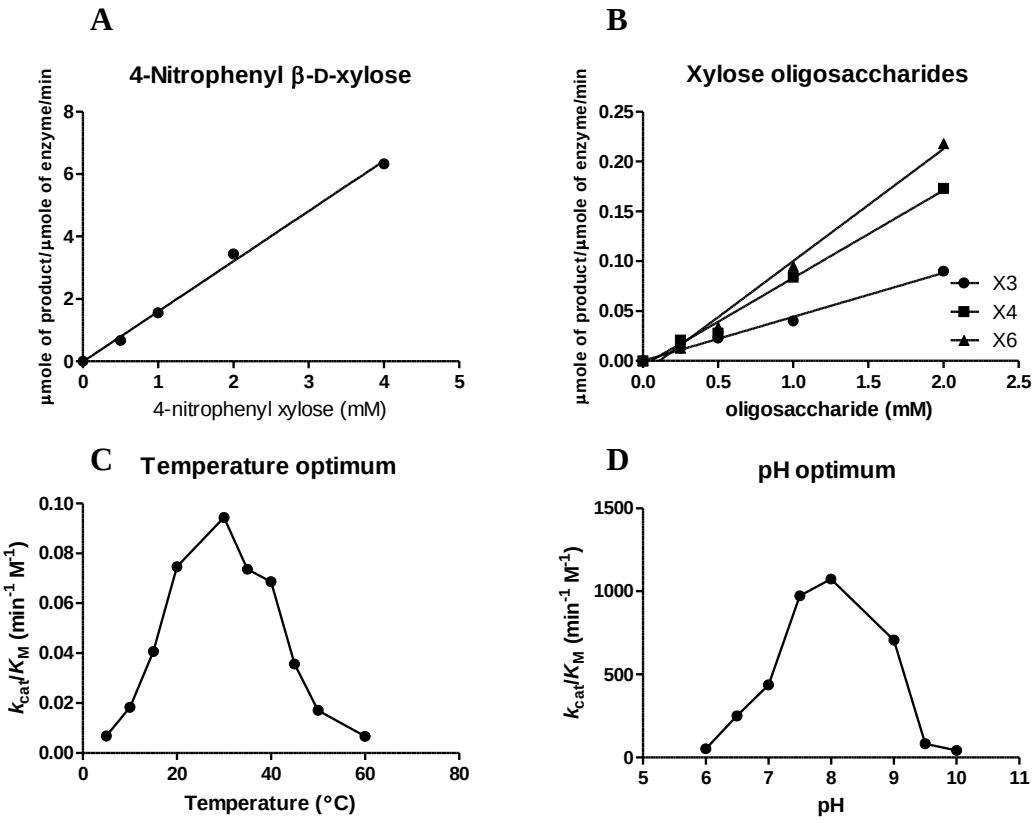


Figure 5.8 Graphs of kinetic data of *Cj*Xyl43A
A-1.8 µM *Cj*Xyl43A vs 4-nitrophenyl-β-D-xylose (4NP-X). B-1.8 µM *Cj*Xyl43A vs X2 (xylobiose), X3 (xylotriose), X6 (xylohexaose). C and D-Temperature and pH optima 1.8 µM *Cj*Xyl43A vs 1 mM (4NF-X). A and B were carried out at 25 °C in 50 mM sodium phosphate buffer pH 7.0. C was performed at a temperature range of 10-60 °C in 50 mM sodium phosphate buffer pH 7.0. D was performed at 25 °C with a pH range of 4-10, pH 4.0-5.5 50 mM Na Acetate, pH 5.5-6.0 50 mM MES, pH 6.0-8.0 50 mM sodium phosphate buffer and pH 9.0-10.0 50 mM CAPSO. Graphs were plotted in GraphPad Prism 4.02 using non-linear regression for A and B; Points were joined in graphs C and D.

Substrate	**k_{cat}/K_M (min^{-1} M^{-1})**
4-Nitrophenyl-β-D-xylose	$2.03 \times 10^3 \pm 2.78 \times 10^2$
Xylotriose	$4.09 \times 10^3 \pm 2.54 \times 10^2$
Xylotetraose	$6.93 \times 10^3 \pm 7.46 \times 10^2$
Xylohexaose	$7.15 \times 10^3 \pm 1.37 \times 10^3$

Table 5.3 Kinetic constants of *Cj*Xyl43A
*Cj*Xyl43A (1.8 µM) was used against all substrates. All reactions were done at 25 °C in 50 mM sodium phosphate buffer pH 7.0.Data was generated by linear regression using GraphPad Prism 4.02 and errors are standard errors generated from triplicate results.

The pH and temperature optima of *Cj*Xyl43A are 8.1 and 29 °C respectively (Figure 5.8C and D). The catalytic efficiency of *Cj*Xyl43A against 4-nitrophenyl-β-D-xylose was similar to that observed against xylotriose, xylotetraose and xylohexaose (Table 5.3). No hydrolysis of xylobiose was detected. These data suggest that the

enzyme has three major subsites. The enzyme released xylose from wheat arabinoxylan, rye arabinoxylan and oat spelt xylan (Figure 5.7), consistent with the generation of xylose from xylotriose, xylotetraose and xylohexaose.

5.3.6 *Cj*Gly43C, *Cj*Gly43D, *Cj*Gly43I and *Cj*Gly43H

Collectively *Cj*Gly43C (CJA_0816), *Cj*Gly43D (CJA_0818), *Cj*Gly43I (CJA_3007) and *Cj*Gly43H (CJA_0361) all displayed trace activity against wheat arabinoxylans (WAX) (Figure 5.9). *Cj*Gly43I appeared to remove only arabinose from WAX but at a slow rate (Figure 5.9). The enzyme displayed a k_{cat}/K_M against 4-nitrophenyl-α-L-arabinofuranoside of $3.73 \times 10^2 \pm 22$ min^{-1} M^{-1}. *Cj*Gly43C, *Cj*Gly43D and *Cj*Gly43H display what appeared to be endo-xylanase activity. This activity is very low when compared to the GH10 xylanase *Cj*Xyn10B. The HPLC traces shown in Figure 5.9 were derived from 24 h reactions, whereas the chromatogram of the *Cj*Xyn10B-xylan reaction was generated after 1 h. These data demonstrate that the three GH43 enzymes are very slow acting xylanases.

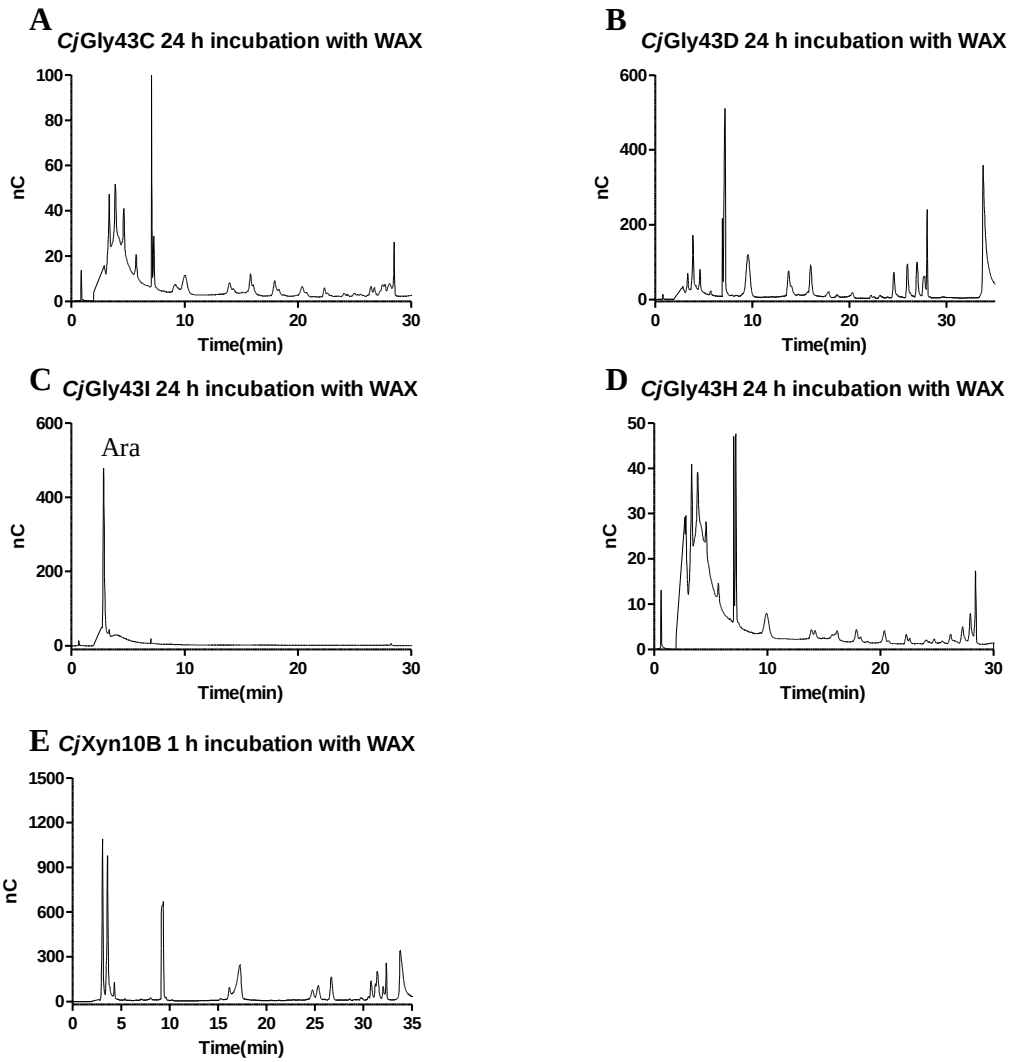


Figure 5.9 HPLC chromatograms of *Cj*Gly43C, *Cj*Gly43D, *Cj*Gly43I and *Cj*Gly43H against WAX
A–5µM *Cj*Gly43C vs 2 mg/ml wheat arabinoxylan (WAX). B-5µM *Cj*Gly43D vs 2 mg/ml wheat arabinoxylans. C-5µM *Cj*Gly43C vs 2 mg/ml wheat arabinoxylan. D-5µM *Cj*Gly43I vs 2 mg/ml wheat arabinoxylan. D-5µM *Cj*Gly43H vs 2 mg/ml wheat arabinoxylan. E-5µM *Cj*Xyn10B vs 2 mg/ml wheat arabinoxylan. Ara = Arabinose. All reactions were performed at 25 °C in 50 mM sodium phosphate buffer pH 7.0. A, B, C and D are undiluted reactions while E was a 1:100 dilution of the enzyme xylan reaction. Graphs were plotted in GraphPad Prism 4.02.

5.3.7 *Cj*Gly43B, *Cj*Gly43E and *Cj*Abf43M

The proteins *Cj*Gly43B (CJA_3594) *Cj*Gly43E (CJA_0799) and *Cj*Abf43M (CJA_0819) were expressed in soluble form but no activity could be detected against a range of plant cell wall polysaccharides including α-arabinans, β-arabinoxylans, β-mannans, β-glucans, and pectins. These proteins will not be discussed in any further detail.

5.3.8 Biochemical characterisation *Cj*Abf43B

Recombinant *Cj*Abf43B (CJA_3018) showed activity against 4-nitrophenyl-α-L-arabinofuranoside and released arabinose from sugarbeet arabinan but was not active on linear arabinan (Figure 5.10A and B). The enzyme showed no activity against any other plant cell wall polysaccharides including β-arabinoxylans, β-glucans, β-mannans or β-galactans.

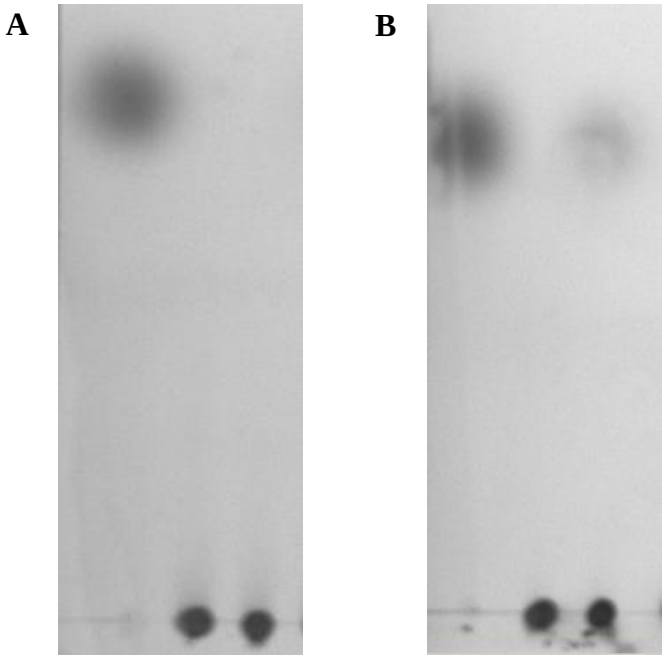


Figure 5.10 TLC of released products by *Cj*Abf43B
A–depicts activity against linear arabinan. 1=Arabinose standard; 2=5 mg/ml Linear arabinan control; 3 =5 mg/ml Linear arabinan incubated overnight with *Cj*Abf43B. B–depicts activity against sugar beet arabinan. 1=Arabinose standard; 2=5 mg/ml sugar beet arabinan control; 3=5 mg/ml sugar beet arabinan incubated overnight with *Cj*Abf43B. All reactions were carried out at 37 °C in 50 mM sodium phosphate pH 7.0.

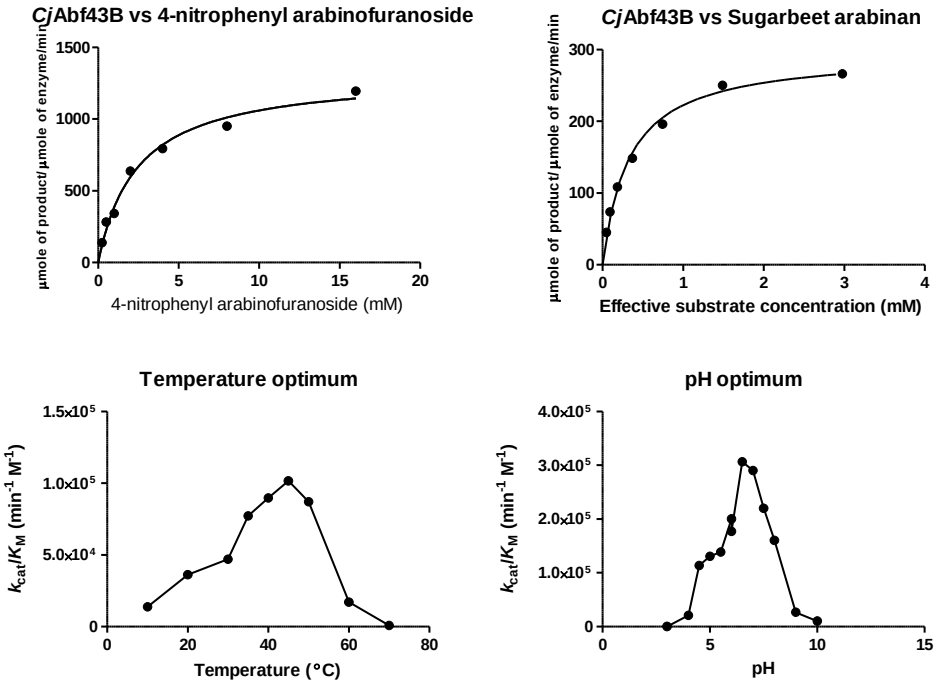


Figure 5.11 Graphs of kinetic data for *Cj*Abf43B
A-5 nM *Cj*Abf43B vs 4-nitropheny-α-L-arabinofuranoside (4NP-AF). B-40 nM *Cj*Abf43B vs sugarbeet arabinan. C and D-Temperature and pH optima 5 nM *Cj*Abf43B vs 1 mM (4NP-AF). A and B were carried out at 25 °C in 50 mM sodium phosphate buffer pH 7.0. C was carried out at a temperature range of 10-70 °C in 50 mM sodium phosphate buffer pH 7.0. D was carried out at 25 °C with a pH range of 4-10, pH 4.0-5.5 50 mM Na Acetate, pH 5.5-6.0 50 mM MES, pH 6.0-8.0 50 mM sodium phosphate buffer and pH 9.0-10.0 50 mM CAPSO. Graphs were plotted in GraphPad Prism 4.02 using non-linear regression for A and B; Points were joined in graphs C and D.

Substrate	**K_M (mM)**	**k_{cat} (min^{-1})**	**k_{cat}/K_M (min^{-1} M^{-1})**	**Total arabinose release from 1 mg (mg)**
4-nitrophenyl α-L-arabinofuranoside	3.08 ± 0.51	1896 ± 319	$6.16 \times 10^5 \pm 2 \times 10^5$	-
Sugar beet arabinan	0.28 ± 0.02*	308 ± 8	$1.09 \times 10^6 \pm 1.09 \times 10^5$	0.045
Arabinohexaose	-	-	$1.1 \times 10^2 \pm 7 \times 10^{-3}$	-

Table 5.4 Kinetic constants of *Cj*Abf43B.
*Cj*Abf43B (5 nM) vs 4-nitrophenyl α-L-arabinofuranoside; 40 nM *Cj*Abf43B vs sugarbeet arabinan; 18 µM *Cj*Abf43B vs arabinohexaose. * K_M calculated by hydrolysing 1mg/ml of sugarbeet arabinan to completion to calculate effective substrate concentration. Data was calculated by non-linear regression was using the Michealis-Menten equation; Standard errors reported are generated from triplicate results.

*Cj*Abf43B has temperature and pH optima of 42 °C and 6.7 respectively. The enzyme displayed a similar catalytic efficiency against 4-nitrophenyl-α-L-arabinofuranoside and sugar beet arabinan. The K_M and k_{cat}, however, were lower

against sugar beet arabinan than 4-nitrophenyl-α-L-arabinofuranoside (Table 5.4). The enzyme displayed a k_{cat}/K_M of 1.1 x 10^2 against arabinohexaose, an α-1,5 linked oligosaccharide, which is 10^4 fold lower than that observed against sugar beet arabinan (Table 5.4). Figures 5.10A and B show there is no detectable release of arabinose from linear arabinan but the monosaccharide is generated from sugar beet arabinan. Linear arabinan consists of a α-1,5-L-arabinofuranose backbone, while sugar beet arabinan comprises an α-1,5-L-arabinfuranose backbone that is decorated with both α-1,3-L-arabinofuranose and α-1,2-L-arabinofuranose side chains. Approximately 60 % of the backbone sugars are decorated with α-1,3-L-arabinofuranose side chains, while only a small amount of α-1,2-L-arabinofuranose side chains are present, although some backbone residues are decorated with both α-1,2 and α-1,3-L-arabinofuranose side chains (Megazyme 2010). These data suggest that *Cj*Abf43B primarily hydrolyses α-1,3 and/or α-1,2 linkages present in sugar beet arabinan. The small amount of total arabinose released from the polysaccharide is inconsistent with the hydrolysis of the α-1,3 linkages. If the enzyme removed the α-1,3 linked arabinofuranose side chains, which decorate ~60 % of the backbone, then the total amount of arabinose released from 1 mg of sugarbeet arabinan should be in the region of ~0.4 mg; only 0.045 mg of sugar was released. This suggests that the enzyme preferentially hydrolyses the α-1,2 linked arabinofuranose side chains present in sugar beet arabinan. To confirm that *Cj*Abf43B hydrolysed α-1,2 linkages in sugar beet arabinan Heteronuclear Single Quantum Coherence (HSQC) NMR was carried out.

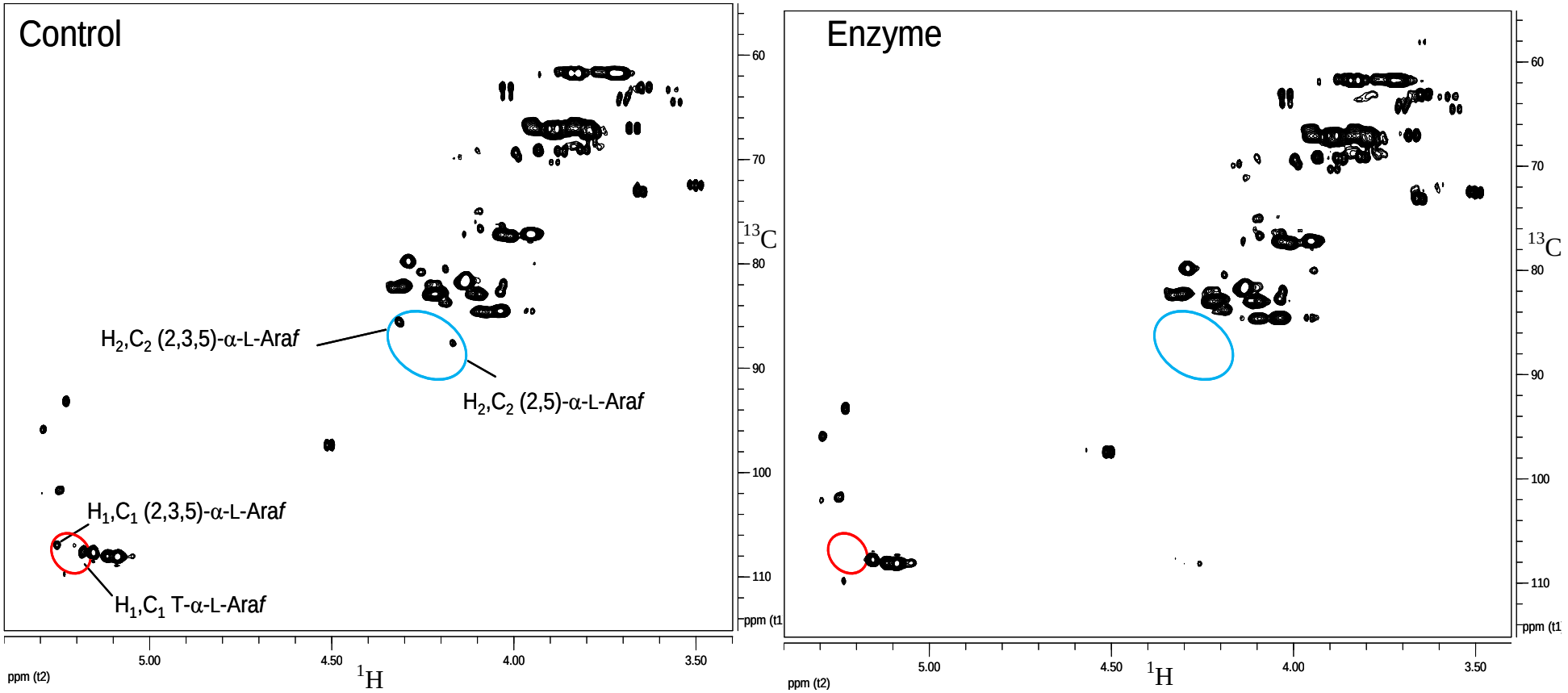


Figure 5.12 Heteronuclear Single Quantum Coherence (HSQC) 1-bond 13C–1H correlation spectroscopy.
Reactions were performed in 5 mM sodium phosphate buffer pH 7.0. *Cj*Abf43B (10 µM) was incubated with acid hydrolysed sugar beet arabinan for 1 hour. Chemical shifts are shown in Table 7. Blue circle indicates the interaction of H2C2 of α-1,2 linked arabinose in single and double substitutions. Red circle indicates H1C1 interaction of α-1,2 linked arabinose in single and double substitutions.

	H1	H2	H3	H4	H5	C1	C2	C3	C4	C5
T-α-L-1,3 arabinofuranose	5.157	4.137	3.959	4.034	3.835, 3.721	107.68	81.62	77.14	84.6	61.7
T-α-L-1,2 arabinofuranose	5.186	4.130	3.966	4.065	3.828 3.719	107.68	-	-	-	-
T-α-L-1,2 arabinofuranose	5.180	4.132	3.963	4.080	3.833 3.718	107.51	81.62	77.13	84.6	61.70
5-α-L-1,5 arabinofuranose	5.085	4.129	4.030	4.217	3.892 3.798	108.77	81.6	77.10	82.99	67.22
5-α-L-1,5 arabinofuranose	5.095	4.131	4.010	4.260	-	108.77	81.6	77.3	-	-
3,5-α-L-1,5 arabinofuranose	5.118	4.292	4.094	4.307	3.947 3.840	108.1	79.79	82.98	82.32	67.05
3,5-α-L-1,5 arabinofuranose	5.112	4.291	4.037	4.218	3.891 3.806	108.00	79.70	82.66	83.32	67.00
2,5-α-L-1,5 arabinofuranose	5.209	4.170	4.160	4.097	3.955 3.885	107.02	87.60	75.86	-	67.00
2,3,5-α-L-1,5 arabinofuranose	5.254	4.314	4.255	4.300	3.900 3.800	107.20	85.60	80.79	81.66	67.00
Free Arabinose										
α-L-arabinopyranoside	4.509	3.501	3.650	3.934	3.890 3.671	97.39	72.49	73.16	69.17	67.05
β-L-arabinopyranoside	5.229	3.808	3.872	3.999	4.018 3.641	93.07	69.05	69.21	69.38	63.10
α-L-arabinofuranose	5.247	4.031	3.980	4.110	3.790 3.687	101.74	-	-	-	-
β-L-arabinofuranose	5.293	4.085	4.040	3.834	3.800 3.659	95.93	-	-	-	-

Table 5.5 1H and 13C chemical shift (ppm) of the arabinan oligosaccharides corresponding to Figure 5.12.

The NMR analysis was performed by Dr Maria Pena from the laboratory of Dr William York at the CCRC, USA. Sugar beet arabinan was hydrolysed by controlled acid hydrolysis and the products were precipitated in 80 % (v/v) ethanol to remove large oligosaccharides and residual polysaccharides. HSQC was performed on the remaining material before and after treatment with *Cj*Abf43B. In Figure 5.12 the blue circle corresponds to the signal of C2-H2 of the backbone arabinose that interacts with α-1,2-L-arabinofuranose. The red circle corresponds to the signal of C1-H1 of the α-1,2-L-arabinofuranose (chemical shifts given in Table 5.5). Two signals are generated because the α-1,2 linkages are found in the context of single and double substitutions. These signals disappear after 1 h incubation with *Cj*Abf43B, demonstrating the enzyme hydrolyses α-1,2-L-arabinofuranose side chains from single and double substitutions.

The controlled acid hydrolysis of sugar beet arabinan produced a disaccharide fragment that was purified by size exclusion chromatography. 1D H-NMR analysis showed the disaccharide to be an arabinofuranose α-1,3 linked to arabinopyranose; *Cj*Abf43B displayed very low activity against the disaccharide taking 12 h to hydrolyse the substrate (Figure 5.13).

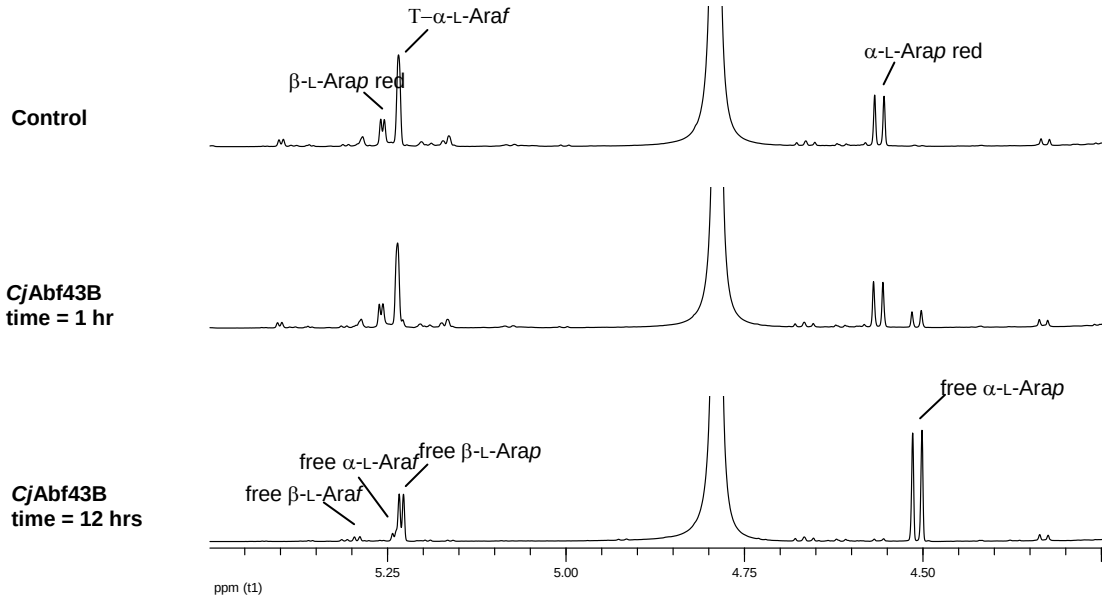


Figure 5.13 1D 1H-NMR of arabinofuranose α-1,3 linked to arabinopyranose
The disaccharide was treated with 10 µM *Cj*Abf43B in 5 mM sodium phosphate buffer pH 7.0 at 25 °C. Ara*f* = arabinofuranose Ara*p* = arabinopyranose. Red = reducing end.

Synergy experiments with the *Humicola insolens* arabinofuranosidase AbfII, also show *Cj*Abf43B acts on α-1,2-L-arabinofuranose side chains in the context of single and double substitutions. AbfII is specific for α-1,3-L-arabinofuranose side chains in double substitutions of arabinoxylan and sugar beet arabinan (Sorensen et al. 2006); it will not act on α-1,3-L-arabinofuranose side chains that are single substitutions. Treatment of 1 mg sugar beet arabinan with AbfII releases 0.033 mg of arabinose. Subsequent addition of *Cj*Abf43B released another 0.043 mg, giving a total arabinose release of 0.076 mg. Treatment of sugar beet arabinan with *Cj*Abf43B first generates 0.045 mg of arabinose and subsequent addition of AbfII failed to release any further arabinose (Table 5.6).

Total arabinose release from 1 mg (mg)		**Total arabinose release from 1 mg (mg)**	
***AbfII**	**+*Cj*Abf43B**	***Cj*Abf43B**	**+AbfII**
0.033	0.076	0.045	-

Table 5.6 Synergy between AbfII and *Cj*Abf43B.
*indicates enzyme was added first; + indicates enzyme was added second. – indicates no further release of arabinose was observed. All reaction were performed in 50 mM sodium phosphate buffer pH 7.0 at 25 °C with 5 µM of each enzyme.

5.3.9 Crystal Structure

The crystal structure of *Cj*Abf43B was solved as a selenomethine derivative, a native form and a base mutant in complex with ligand. It was necessary to grow selenomethionine crystals of *Cj*Abf43B so the single wavelength anomalous dispersion method could be employed to calculate experimental phases. No appropriate model was available for a successful molecular replacement. The selenomethionine derivative of *Cj*Abf43B was dialysed against 10 mM DL-Dithiothreitol (DTT) and was crystallised at 30 mg/ml in 1.5 M ammonium sulphate and 100 mM Tris-HCl pH 7.5. Crystals grew over night at 20 °C. To collect X-ray diffraction data glycerol was used as a cryoprotectant at 25 % (v/v). The selenomethionine crystals were in the tetragonal space group $P4_12_12$, containing four molecules in the asymmetric unit, and diffracted to a maximum resolution of 3.0 Å. The data was processed and scaled using Imosflm (Leslie 1992) and Scala (Evans 2006). The scaled data was then input in to the programs SHELX C/D/E (Sheldrick 2008) and the atomic positions of 30 selenium atoms were found and an initial electron density map generated. The initial map was improved through the use of

solvent flattening and 2 fold non-crystallographic symmetry operators using the program DM (1994), this improved the overall figure of merit (FOM) from 0.40 to 0.73. An initial model was built with ARP/wARP (Langer et al. 2008); which was improved by using Buccaneer (Cowtan 1998). Automatic refinement was performed with Refmac5 (Murshudov et al. 1997) interspersed with manual correction in Coot 0.5 (Emsley et al. 2004). Native protein was dialysed against water and crystallised at 10 mg/ml in 13 % PEG 3350 and 150 mM sodium acetate at 20 °C, with crystals growing over a four week period. The crystals were in the monoclinic space group $P2_1$ with two molecules in the asymmetric unit. The base mutant was dialysed against 50 mM sodium chloride and crystallised at 20 mg/ml in the presence of 20 mM arabinotetraose in 20 % PEG 3350 and 200 mM ammonium tartrate at 20 °C, with crystals growing over a two month period. When collecting X-ray data on these crystals ligand was added to the cryprotectant (25 % ethylene glycol v/v). Crystals were in the tetragonal $P4_12_12$ space group containing two molecules in the asymmetric unit. Native and base mutant crystals in complex with arabinotetratose diffracted to a maximum resolution of 1.64 Å and 1.80 Å, respectively. Data was processed and scaled using Imosflm (Leslie 1992) and Scala (Evans 2006). The scaled data was input into Phaser (McCoy et al. 2007) and the structures solved by molecular replacement using the selemethionine structure as the search model. Refinement was performed using Refmac5 (Murshudov et al. 1997) interspersed with manual correction in Coot 0.6 (Emsley et al. 2004). Waters were added using ARP/water (Langer et al. 2008). Data collection and refinement statistics are given in Table 5.7.

	Selenomethionine	Native	D14A-Ligand
Date collection			
Space group	$P4_12_12$	$P2_1$	$P4_12_12$
Cell dimensions			
(Å)	a = b = 192.65, c = 132.5	a = 47.29, b = 139.27, c = 50.46	a = b = 85.64, c = 195.65
(°)	α = β = γ = 90°	α = γ = 90°, β = 116°	α = β = γ = 90°
Resolution (Å)	72.17 – 2.99 (3.15 – 2.99)	36.29 – 1.64 (1.73 – 1.64)	44.73 – 1.79
No. of observations	728565	260148	1008728
No of unique reflections	50829	71488	69576
Mean $I/\sigma I$	14.9 (5.2)	11.5 (4.1)	16 (5.4)
R_{merge}(%)	0.22 (0.57)	0.10 (0.40)	0.11 (0.50)
Completeness (%)	100 (100)	99.8 (100)	100 (100)
Anomalous completeness (%)	100 (100)	-	-
Multiplicity	14.3 (14.7)	3.6 (3.7)	14.5 (14.5)
Refinement			
R_{work}/R_{free}	0.20/0.24	0.15/0.18	0.17/0.20
RMSD bond lengths (Å)	0.011	0.010	0.010
RMSD bond angles (°)	1.35	1.226	1.242
Ramachandran plot*			
Favoured (%)	94.7	96.5	96.1
Allowed (%)	99.9	99.7	100
Mean B-factor			
Wilson B (A^3)	51.5	11.20	26.50
Main chain (A^3)	21.03	9.60	25.50
Side chain (A^3)	21.82	10.18	27.03
Ligand/water (A^3)	-/-	-/22.00	32.54/37.91

Table 5.7 Data collection and refinement statistics.
* Ramachandran statistics generated using molprobity.

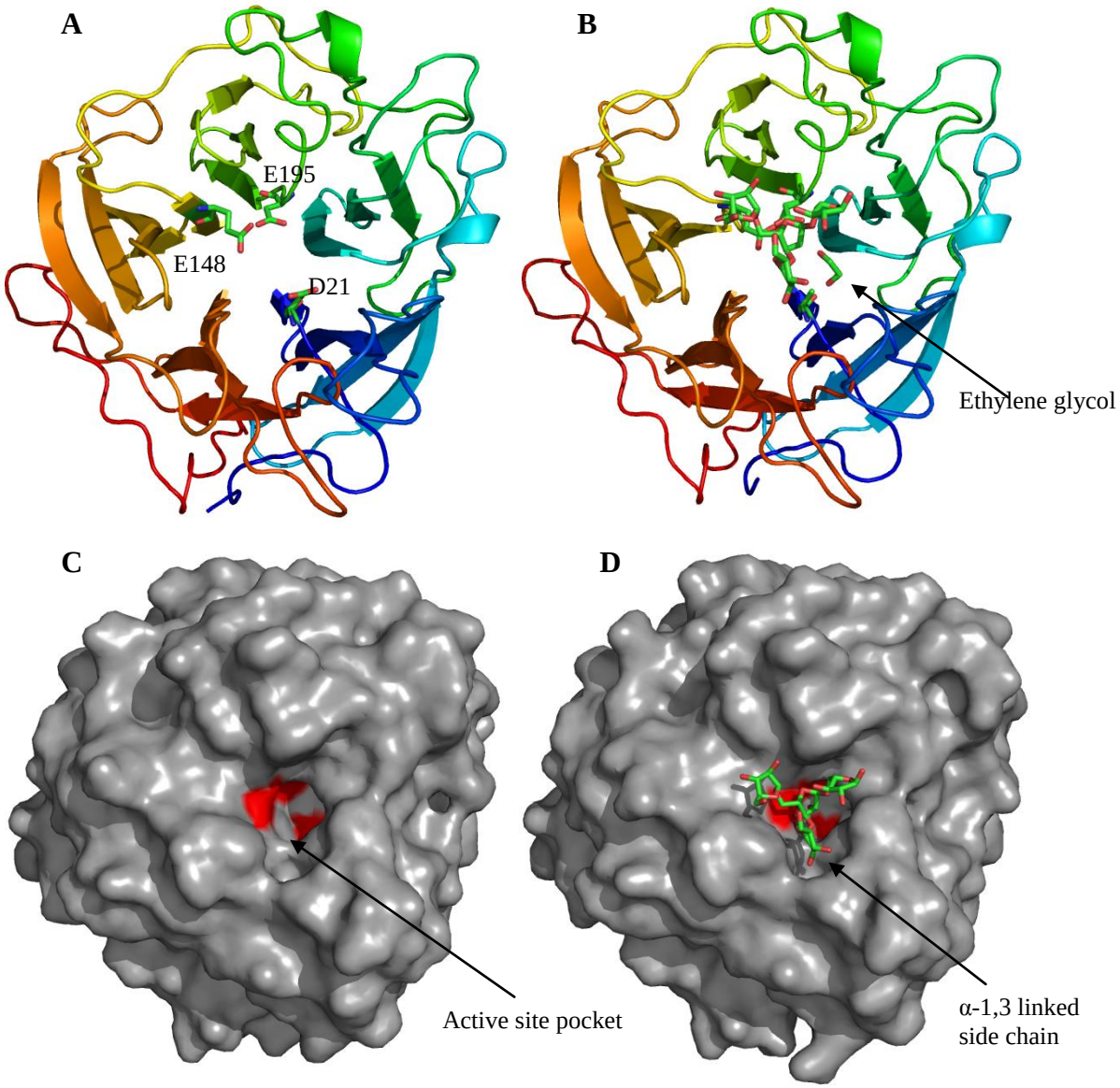


Figure 5.14 Cartoon and surface representation of *Cj*Abf43B
A–Cartoon representation of the native crystal structure of *Cj*Abf43B; B–Cartoon representation of ligand bound crystal structure of *Cj*Abf43B; colours are ramped N–terminus blue through to C–terminus red, C–Surface representation of the native crystal structure of *Cj*Abf43B; D-Surface representation of the ligand bound crystal structure of *Cj*Abf43B. The catalytic residues D21, E148 and E195 are highlighted in red. Figures drawn using pymol (http://www.pymol.org).

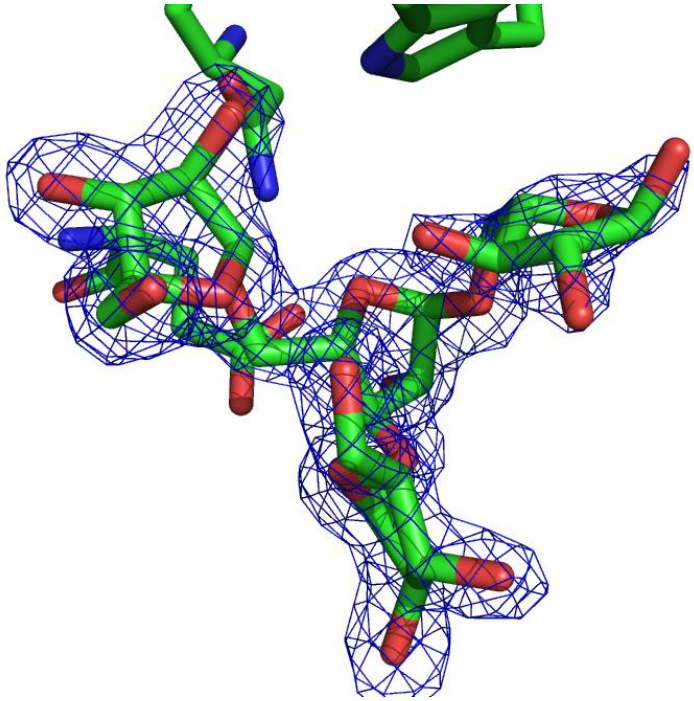


Figure 5.15 Tetrasaccharide contoured with the $2F^{o} – F^{c}$ electron density map
The ligand has an occupancy of 70 % and the $2F^{o} – F^{c}$ map is contoured at 1 σ

The crystal structures reveal *Cj*Abf43B has a five bladed β–propeller fold, typical of a GH43 enzyme (Figure 5.14A and B). The structure also reveals the catalytic residues reside in a deep pocket, which the α-1,2 linked side chain will occupy, and shows how an α-1,3 linked side chain could, potentially be accommodated (Figure 14C and D). The catalytic residues were identified through alignment against other GH43 enzymes where the catalytic residues were known (data not shown); the equivalent residues in *Cj*Abf43B are D21, E148 and E195. Replacing each of these residues with alanine inactivated the enzyme. The base mutant was crystallised with arabinotetraose, the crystal structure, however, revealed arabinotriose in both molecules in the asymmetric unit at an occupancy of 70 %. One of the arabinotriose molecules revealed density for an α-1,3 linked side chain, partially disordered (Figure 5.15), which likely represents a trace contaminant as linear arabinan oligosaccharides are manufactured by removing the side chains from sugar beet arabinan (Megazyme 2010).

5.3.10 Subsite interactions

5.3.10.1 Subsite nomenclature

The nomenclature for glycoside hydrolase subsites, described previously, is not completely adequate for exo-acting enzymes acting on the side chains of

oligosaccharides containing multiple substitutions, and there is no well defined alternative in the literature. An adaptation of the current nomenclature will be used here. Bond cleavage still occurs between the -1/+1 subsites and as the substrate extends towards the reducing end, the subsites increase in positive value ie. +2R, +3R; the R indicates that the arabinose is at the reducing end of the substrate. The -1 subsite, housing the α-1,2-L-arabinofuranose residue, is in a deep pocket and extension beyond this subsite is not possible. The +1 arabinose, however, has another two arabinofuranose residues attached, an α-1,3 linked side chain and an α-1,5 linked backbone residue. The α-1,5 linked arabinofuranose will be numbered +2NR with further backbone residues being numbered +3NR, +4NR etc. The α-1,3 linked residue will be numbered +2NR*. The addition of NR signifies the arabinose terminates with a non-reducing end and +2 symbolises both residues are attached to the +1 arabinofuranose. A * is added to differentiate between the two +2NR subsites (Figure 5.16).

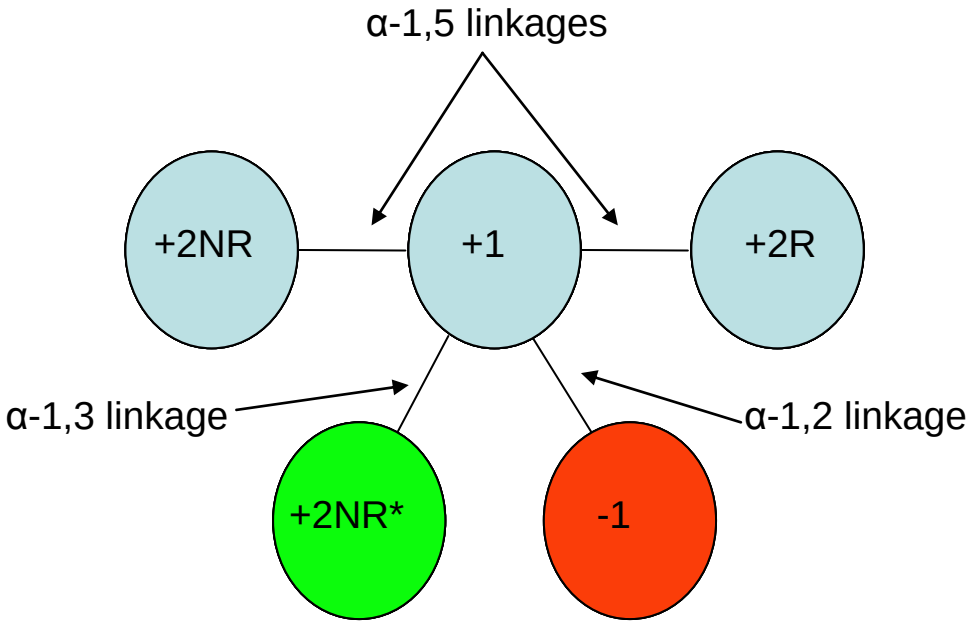


Figure 5.16 Schematic of subsites nomenclature for *Cj*Abf43B subsites

5.3.10.2 Positive subsites

The +2R arabinose is sandwiched between the aromatic side chains of F46 and W144, which are 4.3 Å and 3.7 Å respectively, from the sugar. The partial disorder of the electron density of the arabinose suggests these interactions are weak. W144 also forms a hydrogen bond with the O2 of the +2NR arabinose via its indole nitrogen. At the +1 subsite the Nδ2 of N165 interacts with the endocylic ring oxygen of the arabinose and the glycosidic oxygen of the glycosidic bond linking the arabinose at the +1 and +2NR subsites. N165 also hydrogen bonds to O2 of the +2NR arabinose via Oδ1. Oε2 of E195, the catalytic acid, interacts with O2 of the +1 arabinose, which would be involved in the scissile glycosidic bond (Figures 5.17 and 5.18). F214 may also provide a hydrophobic interaction at both the +2NR and the +2NR* subsites, which houses the α-1,3 linked arabinofuranose side chain in double substitutions. Other interactions at the +2NR* subsite are made by the hydroxyl group of Y273, which interacts with O3, and Q272 that makes a polar contact with O2 via Oε1. These interactions are likely to be weak as there is partial disorder in the electron density map of the +2NR* arabinofuranose (Figures 5.15, 5.17 and 5.18).

5.3.10.3 Negative subsite

The pocket housing the -1 subsite has no arabinose present but a molecule of ethylene glycol was observed. The ethylene glycol 'stacks' against W83 and binds to R275 via either Nη1 or Nη2. These interactions likely mimic those that would be made by arabinose (Figures 5.17 and 5.18). Based on structural information a range of mutants were created and assayed for activity against 4-nitrophenyl-α-L-arabinofuranoside and sugar beet arabinan.

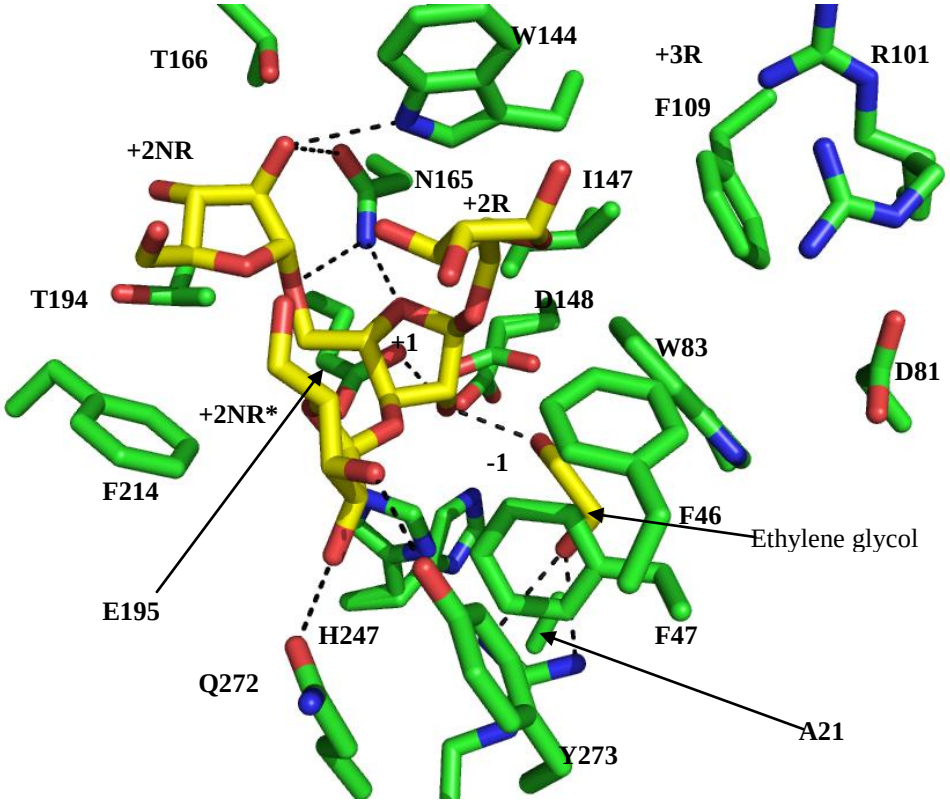


Figure 5.17 Representation of the active site of *Cj*Abf43B in complex with ligand.
Protein carbon and oxygens are shown in green and red with ligand shown in yellow and red

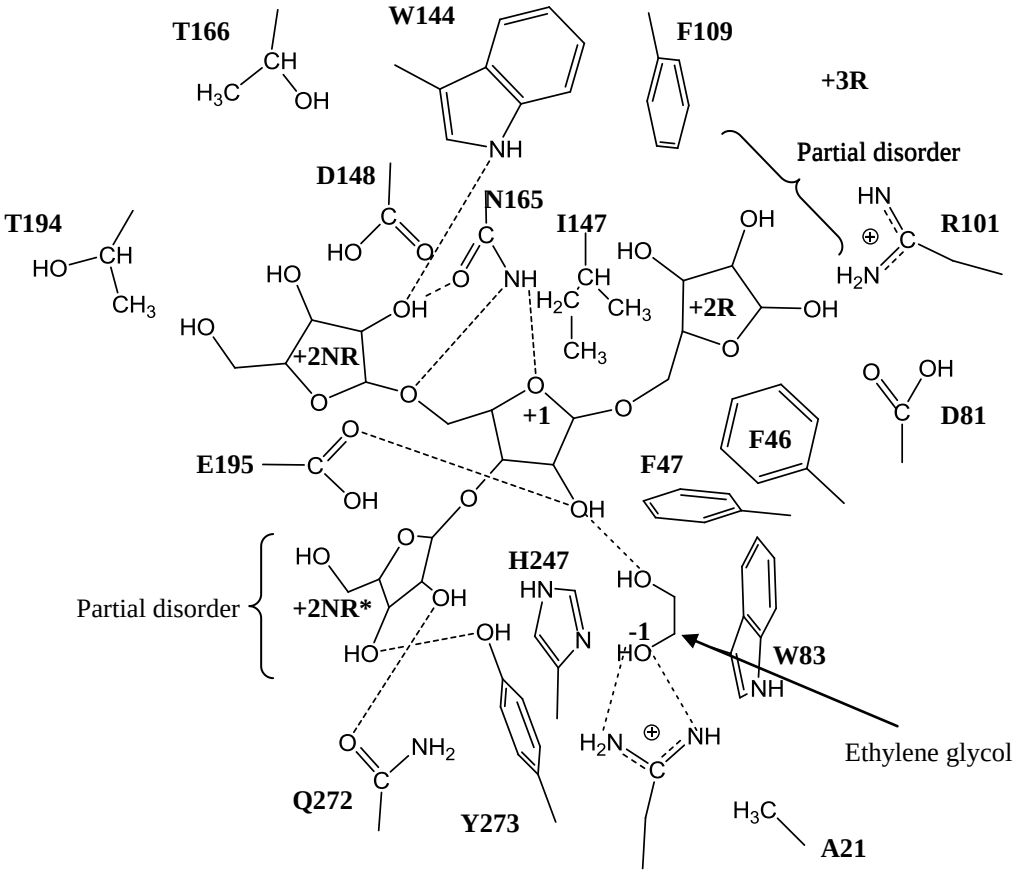


Figure 5.18 Schematic representation of the active site of *Cj*Abf43B in complex with ligand.
Alternate conformations of R101 and H247 have been left out for clarity.

5.3.11 Kinetic analysis of mutants

All kinetic data are displayed in Table 5.8. Kinetic parameters were determined by non-linear regression analysis, examples of which are displayed in Figure 5.11.

5.3.11.1 +3R subsite

The putative +3R subsite mutants were based on the possible interactions that would be made by extension of the arabinan backbone from the +2R subsite. The mutations D81A and F109A had no significant effect on the activity of the enzyme against the aryl glycoside or sugar beet arabinan substrates. The R100A mutant had little effect on K_M against either substrate but reductions in k_{cat} were ~10 fold for both substrates, this is likely due to small perturbations in tertiary structure effecting the +1/-1 subsites as the residue does not interact with the aryl glycoside.

5.3.11.1 +2R subsite

The F46A mutant had little effect on the catalytic efficiency of both substrates and the individual kinetic parameters when arabinan was hydrolysed by *Cj*Abf43B. The individual kinetic constants could not be attained for the aryl glycoside as the K_M was above the maximum substrate concentration deployed. For the W144A mutant individual kinetic constants could also not be obtained for either substrate as the K_M could not be estimated. A reduction in catalytic efficiency of ~10 and ~30 fold was observed for the aryl glycoside and polysaccharide substrates, respectively.

5.3.11.2 +1/-1 subsites

The mutation of D21, D148 and E195, which comprise the catalytic triad of Clan F enzymes completely inactivate *Cj*Abf43B against both substrates tested (Nurizzo et al. 2002). D21 is the catalytic base, E195 the catalytic acid and D148 is thought to be an acidic residue that modulates the pKa of the catalytic acid ensuring it is protonated (Nurizzo et al. 2002). F47, W83 and I147 all contribute to the -1 pocket topology, providing a microenvironment for catalysis to occur; mutation of these three hydrophobic residues cause significant reductions in k_{cat}/K_M against both substrates. The F47A mutation slightly raised the K_M but had a significant effect on k_{cat} against the polysaccharide, suggesting it mainly contributes energy to transition state distortion. A similar loss in catalytic efficiency was observed against the aryl

glycoside but individual kinetic parameters could not be obtained. The W83A mutation caused the greatest loss in activity of any single mutation (that was not a catalytic residue) and was completely inactive against the polysaccharide. This may not be surprising as aromatic residues beneath the -1 sugar have often been shown to be essential to catalysis in many glycosidases, and usually aid the formation of the transition state (Hogg et al. 2001; Money et al. 2008). I147A causes similar rate reductions in k_{cat}/K_M against both substrates and the K_M was elevated above the maximum substrate concentration used for both substrates. I147 sits at the +1/-1 boundary and helps form the top of the active site pocket. The residue lays 3.6 Å from the catalytic acid, E195, and may help to create an apolar environment, which would contribute to the pKa of the catalytic acid. H247 resides in the -1 subsite, lying between D21 and E195; mutation of H247 had a minor effect on K_M against the polysaccharide but had a larger effect on k_{cat} suggesting it mainly provides energy for transition state distortion against the polysaccharide. No effect on the catalytic efficiency was observed for the mutation against the aryl glycoside. A significant increase in K_M was observed and the value could not be estimated. This suggests that although affinity for the aryl glycoside has decreased, k_{cat} has increased. This may be due to the weaker affinity for the -1 arabinose increasing departure of the product and thus enhancing k_2. Against the polysaccharide the N165A mutation raised the K_M and the parameter could not be estimated. The effect on k_{cat}/K_M was the largest of any non-catalytic residue outside the -1 subsite, and may not be surprising as the residue forms three hydrogen bonds with the +1 endocylic ring oxygen, glycosidic oxygen between +1 and +2NR, and the O2 of the +2NR arabinose. These interactions likely contribute significantly to formation of the Michealis complex and also provide energy for transition state stabilisation. It is likely, however, that the hydrogen bond to the glycosidic oxygen is the most significant interaction between the Asn and the substrate. It is the only interaction, which is unique when the O2 of the +1 arabinose is orientated into the -1 subsite, thus is likely to be involved in driving specificity for α-1,2-L-arabinofuranose linkages. Against the aryl glycoside a slight lowering of K_M was observed and there was a marginal increase in k_{cat}/K_M. The aryl glycoside places an arabinose at -1 and a phenyl group at +1 and thus N165 is not able to interact with the substrate as it does not offer any hydrogen bond donors or acceptors. Indeed replacing the polar side chain with alanine makes the +1 subsite more hydrophobic and more able to accommodate the phenyl ring of the aryl glycoside. This increases

the affinity without compromising k_{cat}. Mutation of R275 prevents the expression of *Cj*Abf43B in *E. coli*. The residue forms a salt bridge between loops at the base of the 1st and 5th β blades of the 5 bladed β–propeller (Figure 5.19). The mutation probably effects proper folding and the protein will be degraded by the expression host, *E. coli*.

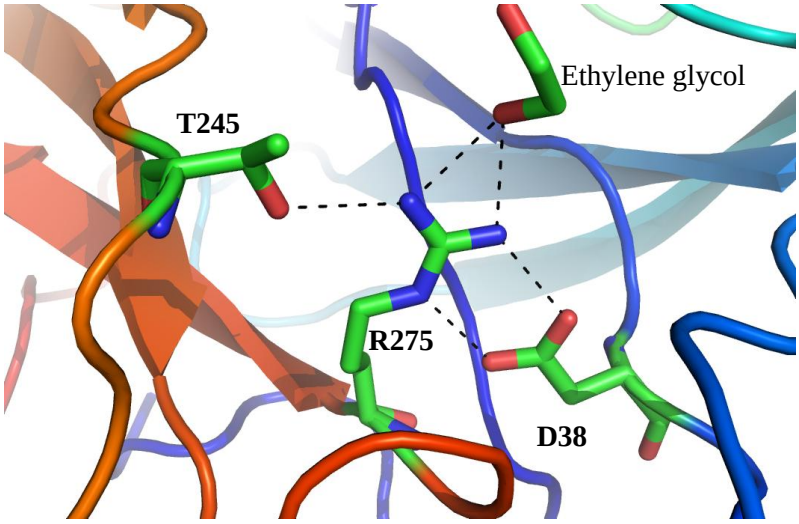


Figure 5.19 Showing the interactions of R275 with the loops at the base of β blades 1 (blue) and 5 (ornage).

5.3.11.3 +2NR subsite

The effects of mutating W144 were discussed under the +2R sub section. The main effect of the mutation may be the loss of the hydrogen bond to the O2 of the +2NR arabinose. The density for the +2NR sugar is well ordered although interactions with F214 may also contribute to this. The mutant F214A caused similar reductions in catalytic efficiency against both substrates. F214A forms part of the cleft for the backbone to sit in and makes a hydrophobic contacts with the +2NR arabinose and is ~5 Å from the +1 subsite. Mutation may cause disruption of the +1 subsite accounting for the similar reduction in catalytic efficiency against both substrates. The T166A and T194A mutants had very minor effects on catalytic efficiency against either substrate and play no significant role in enzyme function. Similarly, T194W and T166W/T194W mutants had a minor effect on catalytic efficiency against the aryl glycoside. The effect of T166W was more significant than T194W and may reflect the proximity of the residue to W144. The introduction of another Trp close to W144 may alter its geometry and thus effect interactions at the +1 subsite. Against the polysaccharide the mutations were far more severe, with the effect on catalysis following the order of T166W>T166W/T194W>T194W. The reason for this effect is

that the introduction of Trp mutations at the +2NR subsite mimics the steric hindrance that precludes the activity of the xylan specific arabinofuranosidase *Bs*AXH-m2,3 against sugar beet arabinan (Figure 5.20B). The surface cleft of *Bs*AXH-m2,3 has evolved to accommodate a linear β-1,4-D-xylose polymers and is unable to accommodate the curved conformation of α-1,5-L-arabinofuranose polymers. The opposite is true for *Cj*Abf43B. The crucial difference is the location of the +2NR subsite of *Cj*Abf43B which accommodates the non-reducing end of the arabinotriose in the crystal structure (Figure 5.20A). By mutating T166 to Trp, unfavourable steric interactions will occur with the arabinose at the +2NR subsite. Against 4-nitrophenyl-α-L-arabinfuranose, which does not occupy the +2NR there is little or no effect on catalytic efficiency. Against the polymeric substrate sugar beet arabinan, which will occupy, and extend beyond, the +2NR subsite the reduction in catalytic efficiency is substantial. If the +2NR subsite cannot be occupied then for hydrolysis to occur the non-reducing end of the polysaccharide must occupy the +1 subsite and only α-1,2-L-arabinofuranose side chains appended to the non-reducing arabinose will be a substrate, this also means the crucial interaction N165 makes with the glycosidic oxygen between +1 and +2NR cannot occur. This dramatically reduces the effective substrate concentration available to the enzyme and thus greatly reduces the k_{cat}/K_M.

5.3.11.4 +2NR subsite*

The mutation Q272A has no effect on the function of the enzyme against either substrate and the hydrogen bond to O2 is thus not an important interaction. Y273A had very little effect on the activity of the enzyme against the polysaccharide, although the mutation had a greater effect on k_{cat}/K_M when the aryl glycoside was the substrate. Y273 stacks against F47 and mutation to alanine may disrupt the geometry of F47 and thus the -1 subsite. These effects may be alleviated by the polysaccharide, as placing an arabinose at the +2NR* subsite could stabilise the geometry of F47 in the -1 subsite. These data show the interactions at the +2NR* subsite are not crucial to catalytic activity. This subsite merely accommodates the α-1,3-L-arabinofuranose side chain without making interactions that contribute to enzyme function.

Enzyme	4-nitrophenyl-α-L-arabinofranoside			Sugar beet arabinan		
	K_M (mM)	k_{cat} (min^{-1})	k_{cat}/K_M (min^{-1} M^{-1})	K_M (mM)	k_{cat} (min^{-1})	k_{cat}/K_M (min^{-1} M^{-1})
Wild type	3.08 ± 0.51	1896 ± 319	$6.16 \times 10^5 \pm 2 \times 10^5$	0.28 ± 0.02	308 ± 8	$1.09 \times 10^6 \pm 1.09 \times 10^5$
D21A	-	-	-	-	-	-
D148A	-	-	-	-	-	-
E195A	-	-	-	-	-	-
F46A	-	-	$2.15 \times 10^5 \pm 2.99 \times 10^4$	1.35 ± 0.11	427 ± 11	$3.15 \times 10^5 \pm 4.1 \times 10^4$
F47A	-	-	$1.57 \times 10^3 \pm 2.58 \times 10^2$	1.60 ± 0.50	14 ± 2	$8.95 \times 10^3 \pm 3.85 \times 10^3$
D81A	2.52 ± 0.15	524 ± 43	$2.08 \times 10^5 \pm 2.9 \times 10^4$	0.19 ± 0.02	266 ± 27	$1.46 \times 10^6 \pm 2.9 \times 10^5$
W83A	-	-	$2.45 \times 10^2 \pm 5.3 \times 10^1$	-	-	-
R100A	4.67 ± 1.56	134 ± 46	$2.87 \times 10^4 \pm 2.1 \times 10^4$	0.70 ± 0.08	47 ± 2	$6.74 \times 10^4 \pm 1.0 \times 10^4$
F109A	3.44 ± 1.18	906 ± 245	$2.63 \pm 10^5 \pm 1.6 \times 10^5$	0.51 ± 0.09	293 ± 18	$5.77 \times 10^5 \pm 1.33 \times 10^5$
W144A	-	-	$6.44 \times 10^4 \pm 1.13 \times 10^4$	-	-	$3.58 \times 10^4 \pm 3.95 \times 10^3$
I147A	-	-	$1.94 \times 10^3 \pm 2 \times 10^2$	-	-	$2.45 \times 10^3 \pm 3.92 \times 10^2$
N165A	0.92 ± 0.05	1676 ± 147	$1.82 \times 10^6 \pm 2.5 \times 10^5$	-	-	$2.98 \times 10^2 \pm 3.30 \times 10^1$
T166A	4.86 ± 0.23	727 ± 98	$1.5 \times 10^5 \pm 2.9 \times 10^4$	0.38 ± 0.03	260 ± 7	$3.36 \times 10^5 \pm 3.7 \times 10^4$
T166W	11.16 ± 0.4	586 ± 1	$5.25 \times 10^4 \pm 2.2 \times 10^3$	-	-	33 ± 1.45
T194A	3.45 ± 0.13	717 ± 57	$2.08 \times 10^5 \pm 1.7 \times 10^4$	0.77 ± 0.08	154 ± 8	$1.99 \times 10^5 \pm 2.99 \times 10^4$
T194W	4.33 ± 0.47	2315 ± 403	$5.35 \times 10^5 \pm 1.5 \times 10^5$	4.06 ± 0.43	27 ± 2	$6.61 \times 10^3 \pm 1.19 \times 10^3$
F214A	9.44 ± 0.75	354 ± 17	$3.75 \times 10^4 \pm 4.9 \times 10^3$	-	-	$2.45 \times 10^3 \pm 3.92 \times 10^3$
H247A	-	-	$8.81 \times 10^5 \pm 1.9 \times 10^4$	0.80 ± 0.15	30 ± 0.25	$3.73 \times 10^4 \pm 7.46 \times 10^3$
Q272A	2.90 ± 0.49	797 ± 210	$2.75 \times 10^5 \pm 1.2 \times 10^5$	0.51 ± 0.04	329 ± 10	$6.45 \times 10^5 \pm 7.1 \times 10^4$
Y273A	-	-	$1.66 \times 10^4 \pm 2.20 \times 10^2$	1.40 ± 0.22	415 ± 69	$3.07 \times 10^5 \pm 1.01 \times 10^5$
T166W/T194W	3.86 ± 1.01	606 ± 252	$1.57 \times 10^5 \pm 1.1 \times 10^5$	-	-	$8.78 \times 10^2 \pm 1.75 \times 10^2$

Table 5.8 Kinetic constants of *Cj*Abf43B and mutant variants.

Red indicates -1 subsite mutants; Orange indicates +1 subsite muatnts; Blue indicates +2R subsite mutants; Green indicates putative +3 subsite mutants; Plum indicates +2NR subsite mutants and Brown indicates +2NR* subsite mutants. – indicates parameter unattainable; 5 nM–10 µM of enzyme was used and all reactions were carried out at 25 °C in 50 mM sodium phosphate buffer pH 7.0. Kinetic parameters were determined by non-linear regression analysis using the Michaelis-Menten equation; All assays were performed in triplicate and standard errors report the deviation between those triplicate reulsts..

5.3.12 Sugar beet arabinan specificity

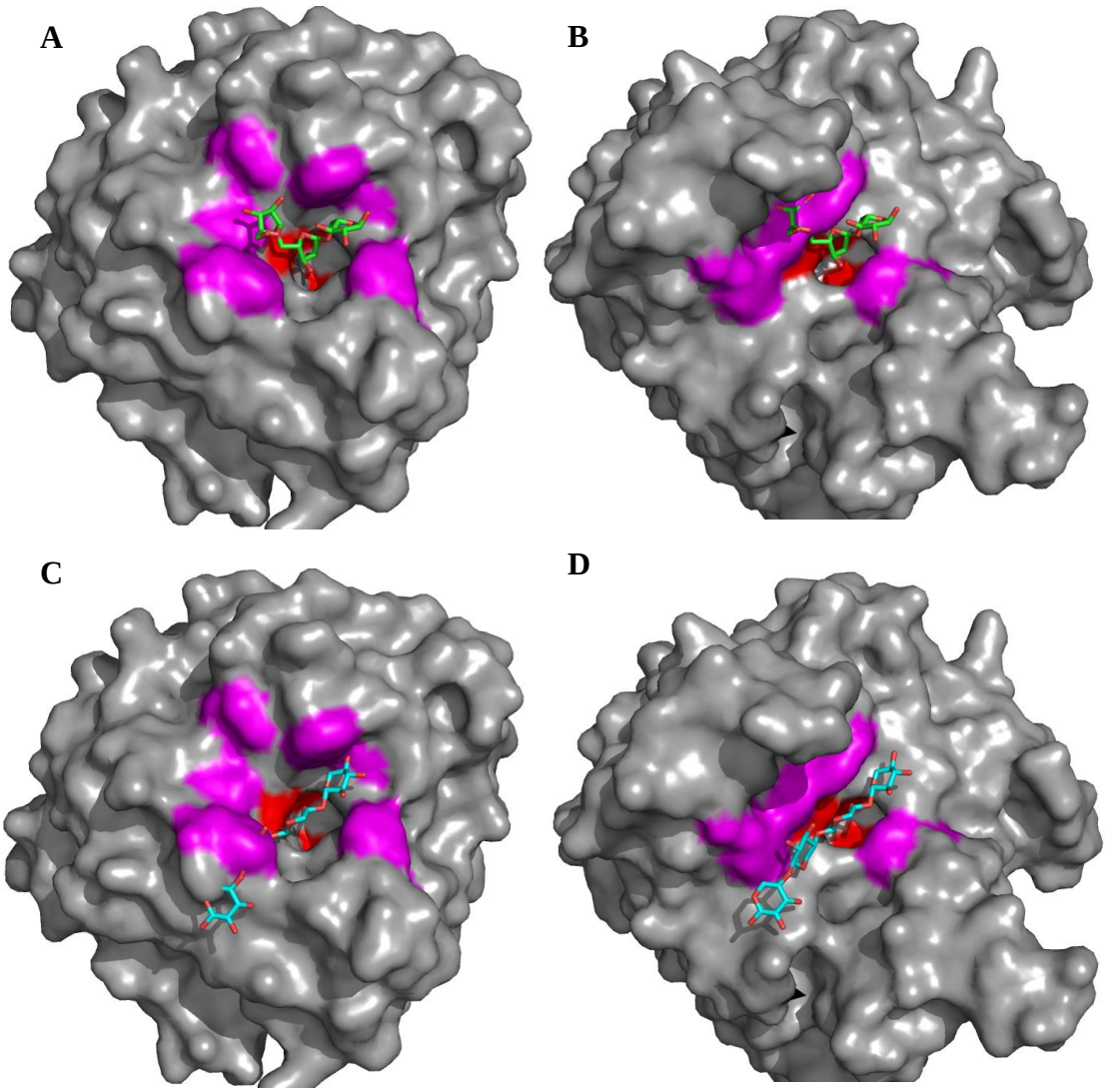


Figure 5.20 Surface representations of *Cj*Abf43B and *Bs*AXH-m2,3 complexed with arabinotriose and xylotetraose, respectively
A–Surface representation of *Cj*Abf43B in complex with arabinotriose (α-1,3 linked side chain has been left out for clarity). B–Surface representation of *Bs*AXH-m2,3 (pdb code 3C7G) overlaid with arabinotriose. C-Surface representation of *Cj*Abf43B overlaid with xylotetraose (from pdb 3C7G). D-Surface representation of *Bs*AXH-m2,3 in complex with xylotetraose (pdb code 3C7G). Red indicates catalytic residues. Magenta highlights the polysaccharide backbone binding cleft.

Figure 5.20 illustrates the specificity of *Cj*Abf43B for sugar beet arabinan through its capacity to accommodate the backbone of the polysaccharide. *Bs*AXH-m2,3 is a GH43 arabinoxylan specific arabinofuranosidase, which cleaves both α-1,2 and α-1,3-L-arabinofuranose side chains from singly substituted xylose residues (Vandermarliere et al. 2009). *Cj*Abf43B is a sugarbeet arabinan specific arabinofuranosidase that cleaves α-1,2-L-arabinofuranose side chains from singly and doubly substituted backbone residues. The active site pocket and the catalytic residues

it encompasses are almost completely conserved across the GH43 landscape (Figure 5.20 and 5.21). The surface topography of the two enzymes is, however, very different. *Cj*Abf43B presents a curved surface topography, which is very well suited to accommodate the curved nature of the α-1,5-L-arabinofuranose backbone of sugar beet arabinan (Figure 5.20A). Superimposition of xylotetraose shows that the linear nature of the backbone would cause steric clashes with the protein (Figure 20C). The opposite is true of *Bs*AXH-m2,3 which presents a linear binding cleft, capable of accommodating the linear β-1,4 linked xylose backbone of arabinoxylan that adopts a three-fold screw axis. Superimposition of arabinotriose shows the curved nature of the α-1,5-L-arabinofuranose backbone of sugar beet arabinan would cause steric clashes with the proteins surface (Figure 5.20D).

5.3.13 Major Specificity determinant for α-1,2 linked side chains

Arabinofuranose oligosaccharides are almost completely symmetrical molecules, the only unique features are the presence of the hydroxymethyl group at the non-reducing end, its absence at the reducing end and the position of the glycosidic oxygen. In the crystal structure solved there was unambiguous density for the hydroxymethyl group and the ligand was modelled with the O2 hydroxyl of the +1 arabinose orientated into the active site pocket, in agreement with NMR and kinetic data. The ligand can also be modelled with O3 of the +1 sugar orientated toward the active site pocket with almost complete conservation of protein-ligand interactions; this requires an 180^{o} reorientation of the backbone (Figure 5.22). Figure 5.22 illustrates the symmetry displayed by the ligand in terms of redundancy of the interactions it can make. The +2NR* subsite makes the exact same interactions with an α-1,2 or an α-1,3 linked arabinfuranose, and so must the -1 subsite. An overlay of the -1 subsite of *Bs*AXH-m2,3, which can hydrolyse α-1,2-L-arabinofuranose and α-1,3-L-arabinofuranose side chains from xylan with the -1 subsite of *Cj*Abf43B, shows an almost complete conservation (Figure 5.21), except I147 in *Cj*Abf43B is replaced by F162 in *Bs*AXH-m2,3. At the +2NR subsite O2 is replaced by O3 and the same interactions are made by the two hydroxyl groups. The +2R subsite makes interactions with two aromatics and once again these are preserved when the substrate is inverted. There is no positive discrimination for the hydroxymethyl group of the backbone. As attention is turned toward the +1 subsite, the interactions with E195, the

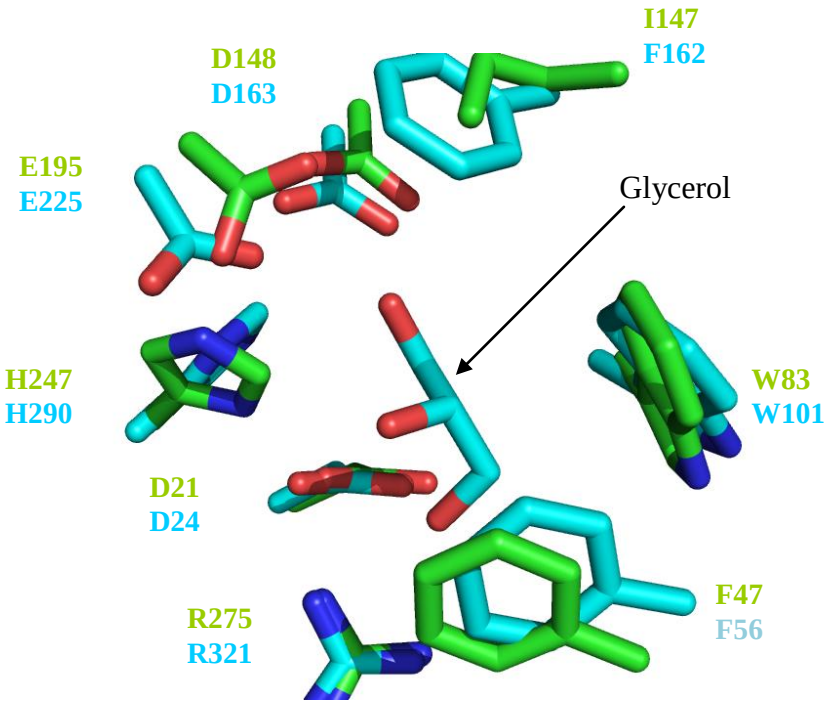


Figure 5.21 Overlay of the -1 subsite of *Cj*Abf43B (Green) and *Bs*AXH-m2,3 (Cyan pdb code 3C7G)

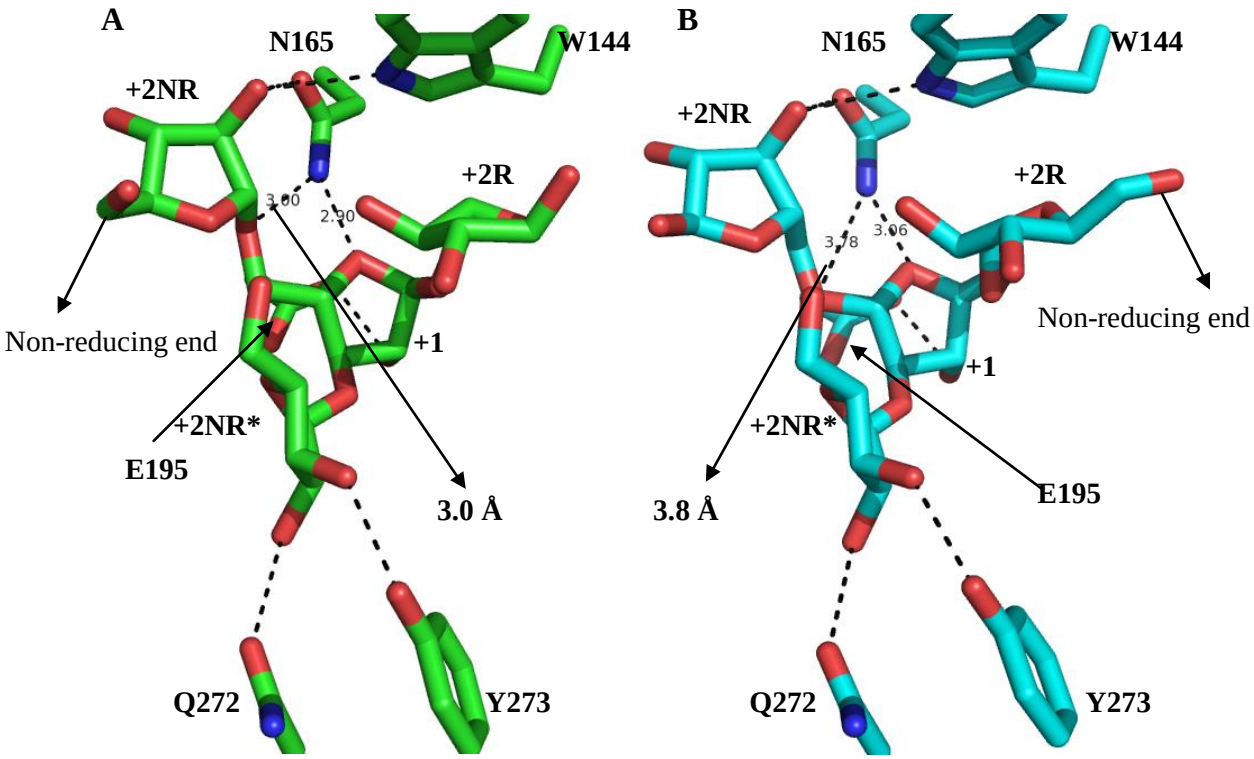


Figure 5.22 Two potential orientations of substrate recognition bound to *Cj*Abf43B
A–Interactions made when O2 of the +1 sugar is orientated toward the active site pocket and the non-reducing end is at +2NR. B-Interactions made when O3 of the +1 sugar is orientated toward the active site pocket and the non-reducing end is at +2R

catalytic acid, are preserved in both orientations, and N165 is in a position to interact with the endocyclic ring oxygen, the O2 of the +2NR arabinose and the glycosidic oxygen between the +1/+2NR subsites. In both orientations the interaction with the endocyclic ring oxygen and the +2NR arabinose are conserved with hydrogen bonding distance to the endocylclic ring oxygen being 3.0 Å with an angle of 145 ° in Figure 5.22A and 3.1 Å with an angle of 143 ° in Figure 5.22B. The interaction with the glycosidic oxygen however, is only made when the non-reducing end of the backbone resides at +2NR, orientating the O2 of the +1 arabinose into the active site pocket (Figure 5.22A). When O3 is orientated into the active site pocket, and the non-reducing end is at +2R (Figure 5.22B), the glycosidic oxygen is 3.8 Å from N165 with an angle of 146 ° and thus beyond hydrogen bonding distance, as opposed to 3.0 Å with an angle of 127 ° in Figure 5.22A.

A BLASTP search of *Cj*Abf43B against the Uniprot database returned 49 sequences with identities ranging from 82-43 %. All but five sequences contained an Asn, preceded by a Trp and a Gly, equivalent to N165 in *Cj*Abf43B, demonstrating a high level of conservation (Figure 5.23). Result 13 of the BLASTP search is the *Bacteroides thetaiotaomicron* protein *Bt*3069, which shares 66 % identity with *Cj*Abf43B. *Bt3069* contains an Asn equivalent to N165 and displays the same specificity as *Cj*Abf43B (**Personal communication with Miss LS McKee**). Figure 23 displays results 13 and 25-50 from the BLASTP search. Results 25-50 share identities ranging from 43-47 % with *Cj*Abf43B and all but three contain an Asn, preceded by a Trp and a Gly, suggesting that these protein sequences may also be sugar beet arabinan specific α-1,2-L-arabinofuranosidases.

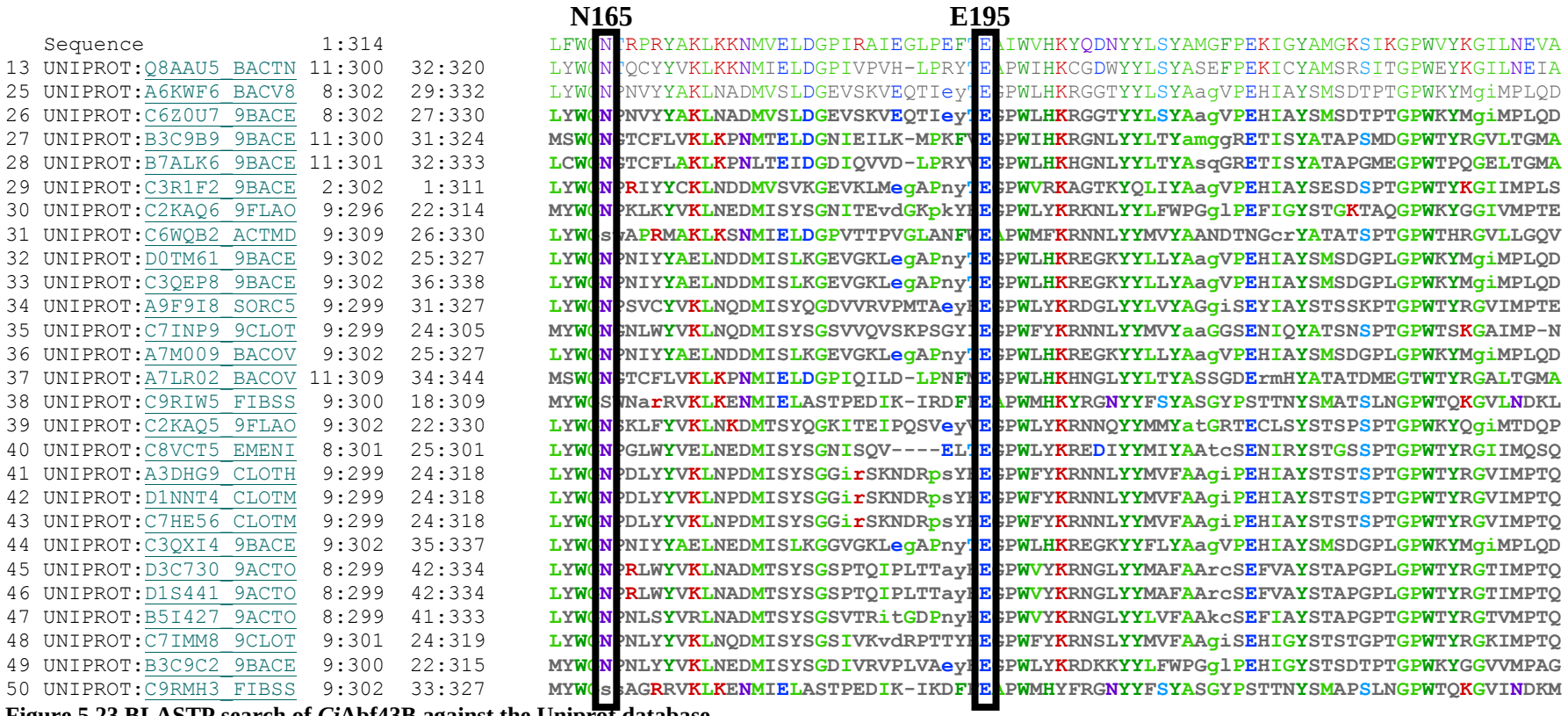


Figure 5.23 BLASTP search of *Cj*Abf43B against the Uniprot database

Results 13 (Identity 66 %), and 25–50 (Identities 43–47 %) are shown. Conservation of N165 is highlighted as is E195 the catalytic acid for comparison. All sequences except 31, 38 and 50 contain an Asn preceded by a Trp and a Gly. This WGN sequence may be a key indicator of sugarbeet arabinan specific α-1,2-L-arabinofuranosidases within GH43.

The activity of the N165A mutant against the mixture of arabinooligosaccharides, generated by controlled acid hydrolysis of sugar beet arabinan, was investigated using 1D H1-NMR (Figure 5.24). It was hoped that the 1D H1-NMR would conclusively prove that the mutation of N165 would abolish specificity for α-1,2-L-arabinofuranose side chains and the enzyme would be much more promiscuous. The wild type enzyme (40 μM) completely removed all α-1,2-L-arabinofuranose side chains within 10 min and reduced the size of the α-1,3-L-arabinofuranose peak indicating partial hydrolysis of α-1,3 linkages had also occurred. The N165A mutant also reduced the size of the α-1,3-L-arabinofuranose peak but the α-1,2-L-arabinofuranose side chains were not fully removed until the 240 min time point. As the two reactions progressed beyond 240 min there were little observable differences. After a week incubated at 4 °C, so that the reaction did not go to completion, the main difference was that the N165A mutant had hydrolysed a much greater proportion of the α-1,5-L-arabinofuranose linkages present in the mixture.

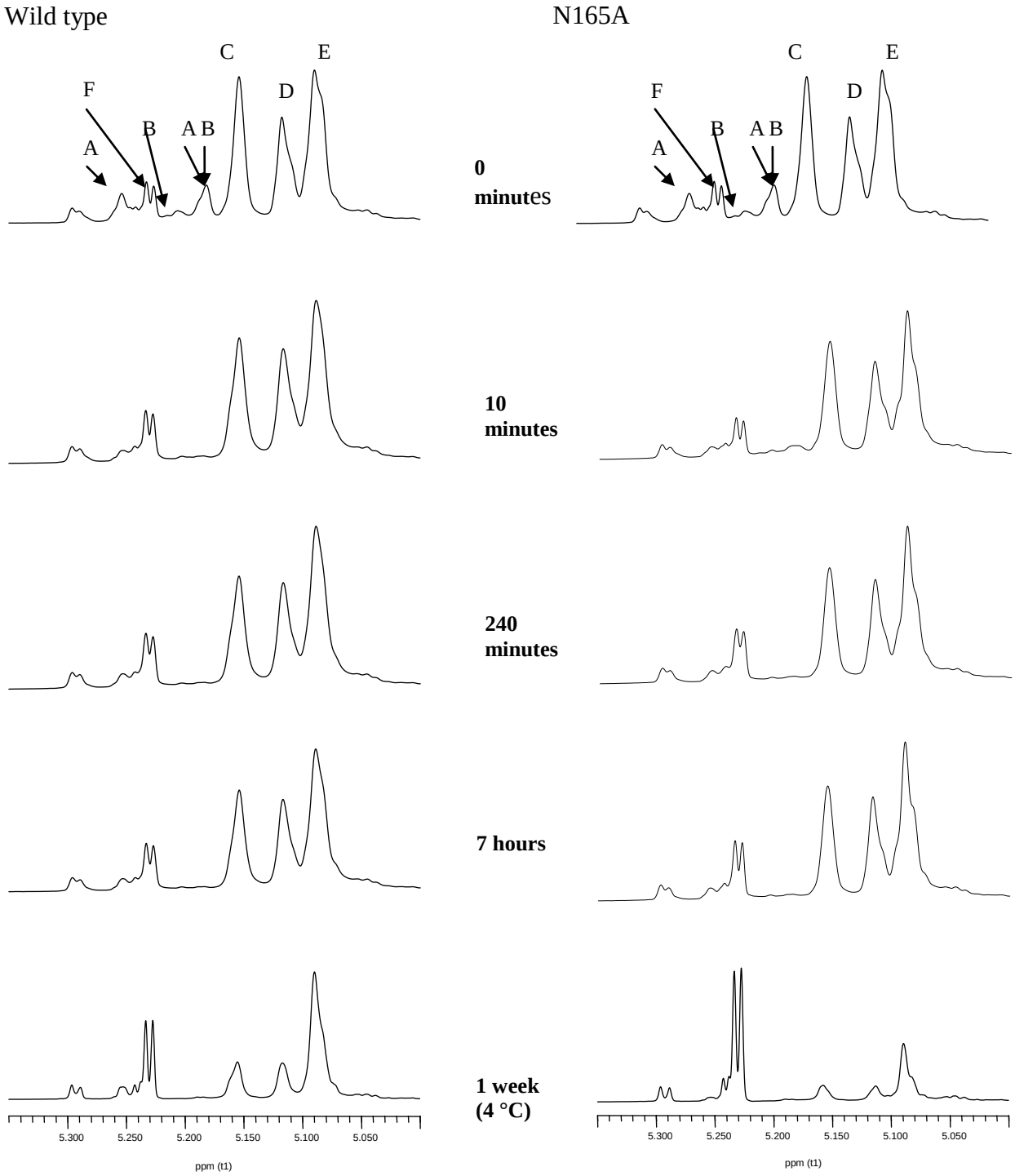


Figure 5.24 1D H1-NMR of 40 µM *Cj*Abf43B and 80 µM N165A against 7 mg of sugarbeet arabinan acid hydrolysis mixture.
A–Signal generated by α-1-2 and 2-3-5 linkages in double substitutions. B-Signal generated by α-1-2 and 2-5 linkages in single substitutions. C–Signal generated by α-1-3 linkage. D–Signal generated by -3-5 α-linkages interaction. E-Signal generated by -5 linkage. F–Free arabinopyranose. Reactions were left at 25 °C for 7 hours then 1 week at 4 °C and were carried out in 5 mM sodium phosphate buffer pH 7.0.

5.4 Discussion

5.4.1 *Cj*Abf43K and *Cj*Abf43L

The k_{cat}/K_M values of the enzymes against sugar beet arabinan were only ~30-60 fold below that of their activity against α-1,5-L-arabinofuranose oligosaccharides, yet both enzymes released very small amounts of arabinose from the polysaccharide, much less than expected if the enzymes were targeting α-1,2 or α-1,3-L-arabinofuranose side chains. This suggests that both enzymes are targeting the α-1,5-L-arabinofuranose backbone of the polysaccharide, which is heavily decorated and thus largely inaccessible, explaining the low levels of arabinose release. The high activity against α-1,5-L-arabinofuranose oligosaccharides confirms the specificity of the enzymes towards α-1,5-L-arabinofuranose linkages. Both enzymes only release arabinose from linear arabinan, sugar beet arabinan and arabinooligosaccharides indicating an exo-mode of action, likely from the non-reducing end, as has been described previously for other GH43s (Proctor et al. 2005; Brux et al. 2006; Brunzelle et al. 2008). Both enzymes utilise two major subsites to perform catalysis demonstrated by no significant difference in k_{cat}/K_M when oligosaccharides with a degree of polymerisation (d.p.) ranging from 2-6 were used as substrates.

*Cj*Abf43L displayed a catalytic efficiency against 4-nitrophenyl-α-L-arabinofuranose similar to that of the enzyme against arabinooligosaccharides, but individual kinetic parameters could not be calculated due to a K_M above the maximum substrate concentration deployed. This may suggest that the +1 subsite does not contribute significant energy to the transition state but is important in the formation of the Michealis complex. It should be noted that the pKa of the 4-nitrophenolate leaving group (~7.5) is much lower than that of arabinose, pKa >12.0, making the substrate more labile and easier to hydrolyse, as the need for protonic assistance is reduced. The +1 subsite may therefore provide energy for transition state distortion but this additional energy is not needed against more labile substrates.

*Cj*Abf43K displayed a catalytic efficiency for 4-nitrophenyl-α-L-arabinofuranose ~100 fold lower than against the arabinooligosaccharides, with a ~10 fold reduction in K_M and a ~1000 fold decrease in k_{cat}. This suggests the interactions with the +1 arabinose provide significant energy to transition state stabilisation but have less impact on formation of the Michealis complex. Substrate binding may be driven largely by the -1 subsite and the rate k_1 may be similar for both substrates; the

decrease in K_M against 4-nitrophenyl-α-L-arabinofuranose could reflect the reduction in k_2. The impaired ability to distort the sugar to the transition state, due to the loss of polar interactions at the +1 subsite, reduces k_{-1} and k_2, which causes the substrate to spend longer 'on enzyme', thus lowering both the K_M and k_{cat}.

Both enzymes are exo-acting, displaying their highest activity against short undecorated α-1,5-L-arabinofuranose oligosaccharides. This suggests their role is in the final steps of arabinan degradation. *C. japonicus* possesses a GH51, which can hydrolyse α-1,2, α-1,3 and, weakly, α-1,5-L-arabinofuranose linkages (Beylot et al. 2001), while another GH43 displays endo-processive α-1,5-L-arabinase activity releasing arabinotriose (Proctor et al. 2005). It is likely *Cj*Abf43K and *Cj*Abf43L hydrolyse the degradation products of these enzymes but in different contexts as *Cj*Abf43K is intracellular and *Cj*Abf43L is extracellular, containing a type II signal peptide (Tailford 2007). It may be plausible to suggest that *Cj*Abf43L might be responsible for largely generating arabinose and arabinobiose, which may preferentially be taken up by *C. japonicus*. The arabinobiose taken up could then be hydrolysed to arabinose by *Cj*Abf43K in the cytoplasm.

5.4.2 *Cj*Xyl43A

*Cj*Xyl43A hydrolyses 4-nitrophenyl-β-D-xylose with a similar catalytic efficiency to xylotriose, xylotetraose and xylohexaose. Individual kinetic parameters could not be obtained due to a K_M above the maximum substrate concentration used. The enzyme did not hydrolyse xylobiose suggesting the β-xylosidase has 3 major subsites. The high K_M displayed by *Cj*Xyl43A may indicate that xylotriose is not the natural substrate for the enzyme. The k_{cat}/K_M of the enzyme against xylooligosaccharides is 10-100 fold lower than other GH43 β-1,4-D-xylosidases (Brux et al. 2006; Brunzelle et al. 2008). It is possible *Cj*Xyl43A may hydrolyse a novel degradation product. The gene encoding *Cj*Xyl43A, *gly43G*, is downstream of *rspA* a starvation sensing protein, a hypothetical protein, *gly43J*, a gene encoding a potential GH43 that was not cloned and *xyn10C*, which encodes the GH10 xylanase, *Cj*Xyn10C (Figure 5.25) (DeBoy et al. 2008). *Cj*Xyn10C has been well characterised and is less efficient compared to other *C. japonicus* GH10 xylanases at degrading small oligosaccharides. It is possible that *Cj*Xyl43A has evolved to act in synergy with *Cj*Xyn10C and specifically targets the oligosaccharide degradation products of the xylanase. There is also the possibility that *Cj*Xyl43A operates from the reducing end

of the substrate and may be specific for the α or β orientation of the anomeric carbon, which would at least double the K_M, with the β:α being ~2:1 (Stoddart 1971). There is an example in GH8 of a reducing end xylosidase that has three significant subsites and releases xylose (Fushinobu et al. 2005).

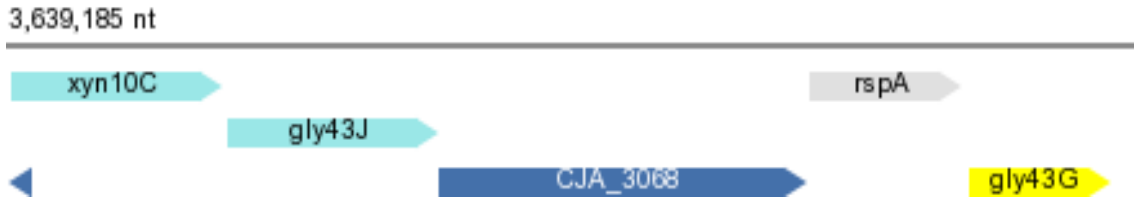


Figure 5.25 Potential operon containing the gene gly43G encoding *Cj*Xyl43A

5.4.3 *Cj*Gly43C, *Cj*Gly43D, *Cj*Gly43I and *Cj*Gly43H

Collectively these enzymes all display trace endo-xylanase activity against wheat, rye and oat spelt xylan. This activity is not sufficient to describe these enzymes as true endo-xylanases when their activity is compared to a typical endo-xylanase such as *Cj*Xyn10B. The hydrolysis of 2 mg/ml of arabinoxylan by *Cj*Xyn10B was complete in well under an hour. Using equivalent molar concentrations of *Cj*Gly43C, *Cj*Gly43D and *Cj*Gly43H against 2 mg/ml arabinoxylan it took 24 h to produce evidence of even very modest hydrolysis of the polysaccharide. *Cj*Gly43C, *Cj*Gly43D reside in the same operon as *Cj*Abf43M (DeBoy et al. 2008), a protein for which no activity has been found, and a putative GH97, Agd97A, a family that contains α-glucosidases and α-galactosidases (Figure 5.26) (Henrissat 1991). Unfortunately this information does not assist in developing a hypothesis for the potential role of the GH43 enzyme. Inspection of the genome of *C. japonicus* reveals that *gly43H* resides downstream of *abf43K*, and genes encoding a putative GH5 cellulase and a putative GH39 xylosidase (Figure 5.27). Again this information sheds very little light on the potential activity of *Cj*Gly43H.

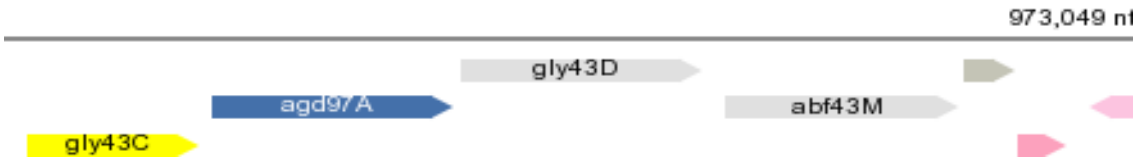


Figure 5.26 Potenatial operon containing *Cj*Gly43C, *Cj*Gly43D and *Cj*Abf43M.

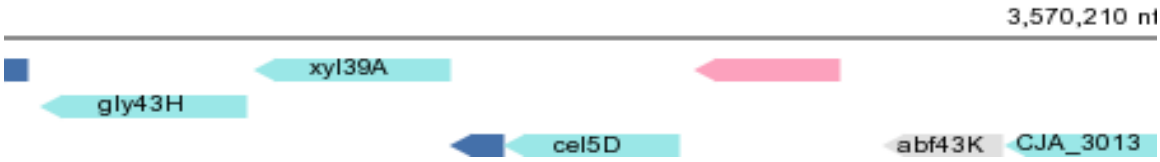


Figure 5.27 operon containing *Cj*Gly43H.

*Cj*Gly43I appears to be a xylan-specific arabinofuranosidase. Once again the activity of the enzyme is remarkably low when compared to the GH43 arabinofuranosidases in this work and in the wider literature (Beylot et al. 2001; Taylor et al. 2006). The catalytic efficiency of the enzyme against 4-nitrophenyl-α-L-arabinofuranose was also extremely low and *gly43I* is sandwiched between genes that encode hypothetical proteins, giving no indication of a possible activity (DeBoy et al. 2008).

5.4.4 *Cj*Abf43B

5.4.4.1 Activity of CjAbf43B

*Cj*Abf43B displays specificity for sugar beet arabinan exhibiting a strong prefence for α-1,2-L-arabinofuranose side chains from single and double substitutions. The fact that arabinopyranose can be accommodated at the +1 subsite raises the question of whether *Cj*Abf43B could tolerate xylose at +1. This is likely to be true but of little biological relevance as the enzyme selects against polymeric xylan as a substrate, discussed below. The ability of *Cj*Abf43B to attack single and double substations allows it to show synergy with *H. insolens* AbfII, which will only hydrolyse α-1,3-L-arabinofuranose side chains in double substitutions (Sorensen et al. 2006). This synergy only occurs if AbfII is used before *Cj*Abf43B, otherwise *Cj*Abf43B removes all the α-1,2-L-arabinofuranose side chains and thus no double substitutions remain. These synergy experiments show double substitutions comprise 3.3 % of the polysaccharide and single α-1,2-L-arabinose side chains 1.2 %.

5.4.4.2 Sugarbeet arabinan specificity

The specificity of *Cj*Abf43B for sugar beet arabinan is due to the surface topography of the enzyme. F46, W144 and F214 help form a curved 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. The enzyme cannot tolerate

xylan as the linear nature of the polysaccharide, which has a 3 fold screw axis, would clash with part of the curved cleft of the enzyme. The opposite is true for *Bs*AXH-m2,3 where sugar beet arabinan would clash with the linear nature of the enzymes substrate binding surface (Vandermarliere et al. 2009). Mutation of F46, W144 and F214 have far less significant effects on the enzymes function than mutations at the -1/+1 subsites. This suggests their role is to form a curved cleft for the backbone of the polysaccharide, without forming strong interactions. This is consistent with the view that the polysaccharide backbone must dissociate before the arabinose in the -1 pocket can then be released, which must occur before another round of catalysis can be performed. If residues forming the surface cleft interacted strongly with the backbone of sugar beet arabinan then this may slow product departure and reduce the catalytic efficiency of the enzyme.

5.4.4.3 Specificity for α-1,2 linkages

The -1 subsite of *Cj*Abf43B is almost completely conserved with *Bs*AXH-m2,3, which hydrolyses α-1,2 and α-1,3-L-arabinofuranose linkages with equal efficiency, and suggests the specificity of *Cj*Abf43B toward α-1,2 linkages is not housed at the -1 subsite (Figure 5.21) (Vandermarliere et al. 2009). Only one interaction between the ligand and *Cj*Abf43B is unique when O2 is orientated into the -1 pocket; the hydrogen bond between N165 and the glycosidic oxygen linking the sugars at the +1 and +2NR subsites (Figure 5.22). Mutation of N165 to Ala causes the catalytic efficiency of *Cj*Abf43B against α-1,2-L-arabinofuranose side chains to decrease (~3000 fold below wild type level) to a level similar to that of the enzyme against α-1,5-L arabinofuranose linkages (~10000 fold below wild type level).

The NMR data of the N165A mutant and wild type enzyme was difficult to interpret. The data showed that both enzymes had some activity against α-1,3 linkages, but the N165A still hydrolysed α-1,2 linkages preferentially despite being the least abundant linkage. Interestingly the N165A mutant was better able to hydrolyse α-1,5-L-arabinofuranose linkages. A possible explanation is that there are other subtle reasons for α-1,2 specificity, although the crystallographic and kinetic data suggests this is unlikely. There is the possibility that the N165A mutant contained trace amounts of wild type enzyme from contamination or a translation error (Shallom et al. 2002). The concentration of N165A used was 80 μM and if 0.1 % was wild type this would equate to 80 nM, enough enzyme to have a significant

effect on the experimental outcome. Another potential problem is that the substrate used was a mixture of different, alcohol soluble, oligosaccharides derived from the controlled acid hydrolysis of sugar beet arabinan. This gives rise to a highly heterogeneous and undefined substrate. This could have the potential consequence that the α-1,3 linkages are less accessible than the α-1,2 linkages and thus hydrolysed at a reduced rate by the N165A mutant relative to the α-1,2-L-arabinofuranose side chains. Collectively, the data point to N165 being the specificity determinant for the hydrolysis of α-1,2-L-arabinofuranose side chains. The observation that N165 is conserved in a cohort of GH43 enzymes, it forms a unique interaction when O2 is orientated into the -1 subsite and the reduction in catalytic efficiency of the N165A mutant, down to a level similar to non-specific α-1,5-L-arabinofuranose hydrolysis, all support the conclusion that the Asn is a key specificity determinant for α-1,2-L-arabinofuranose specificity. It achieves this through a specific hydrogen bond to the glycosidic oxygen between the +2NR and +1 subsites, which only exists when α-1,2 side chains occupy the -1 subsite.

5.5 Conclusion

Although four activities, one of them novel, were identified, three enzymes displayed no activity and four displayed trace activities which were hard to interpret. This once again clearly shows the complexity of plant cell wall degradation. Analysis of individual enzymes alone may not be sufficient to elucidate their function. Just as the plant cell wall is a composite structure, whole operons may need to be cloned, expressed and characterised for each of the individual activities of the enzymes in these loci to be uncovered.

Two exo-α-1,5-L-arabinofuranosidases were discovered, *Cj*Abf43K, which is intracellular, and *Cj*Abf43L, which is extracellular. *Cj*Abf43K helps to break down small oligosaccharides to monomeric arabinose within the cell, while *Cj*Abf43L may act in synergy with *Cj*Arb43A, an endo processive arabinanase producing arabinotriose, and *Cj*Abf51A, which hydrolyses primarily α-1,2 and α-1,3-L-arabinofuranose linkages but can also hydrolyse α-1,5-L-arabinofuranose at a reduced rate (Beylot et al. 2001).

*Cj*Abf43B displays a unique specificity for the α-1,2-L-arabinofuranose side chains of sugar beet arabinan. It also appears to display novel, flexible,

arabinofuranosidase activity as it can attack single and double substitutions with equal efficiency.

*Cj*Xyl43A appears to be a xylosidase but the catalytic efficiency of the enzyme is about ~100 fold lower than other GH43 xylosidases. This and the very high K_M against xylooligosaccharides suggest the true activity may be more complex. This is further hinted at as *xyl43a* is downstream of the *Cj*Xyn10C gene, which was shown previously to be a xylanase with a reduced ability to hydrolyse smaller oligosaccharides, compared to other GH10 enzymes. *Cj*Xyl43A may work in synergy with *Cj*Xyn10C and hydrolyse unique products generated by the GH10 xylanase.

Chapter 6

Final Discussion

The plant cell wall has the potential to be a significant industrial resource for future generations, providing a sustainable, carbon neutral, liquid fuel in the form of bio-ethanol. It is clear, however, that hurdles must be overcome and these ultimately relate to making ligno-cellulosic degradation economically viable. Microbes in nature perform this task very well. Saprophytic organisms, such as *Cellvibrio japonicus*, utilise the plant cell wall as their primary energy source. They produce a consortium of glycoside hydrolases, which enable these microorganisms, to utilise the plant cell wall as a carbon and energy source. Through the discovery and analysis of these glycoside hydrolases efficient systems for degrading the plant cell wall can be developed which, ultimately, will lead to the production of bio-ethanol as an economically viable process.

Chapter 3 presented the detailed characterisation of novel GH26 members, *Ct*Lic26A and *Cj*Man26C. *Ct*Lic26A had been shown previously to be a endo-β-1,4,1,3 glucanase (Taylor et al. 2005) a deviation from the typical endo-β-1,4 mannanase activity expected for GH26. Further studies implicated a $^{4}H_{3}$ half chair as the likely transition state (Money et al. 2006), which is different to β-1,4 mannanases where the transition state is thought to adopt a $B_{2,5}$ boat conformation (Ducros et al. 2002). A detailed kinetic analysis of the substrate binding cleft was lacking and this was provided through alanine scanning of appropriate residues. This identified amino acids, other than the catalytic residues, whose interactions with the substrate were important in catalysis. Aromatic stacking beneath the -2 sugar had a profound effect on k_{cat} and N157 appeared to play a role in helping to position a water molecule during deglycosylation. The data showed mutations had a greater effect on the aryl-glycoside than the polysaccharide, and that mutations in the negative subsites were more severe than those in the positive subsites; this has been observed previously for other retaining glycosidases (Ly et al. 1999; Hogg et al. 2001). The second GH26 enzyme investigated, *Cj*Man26C, was shown to be a mannanase that displayed an atypical exo-mode of action. The crystal structure of the enzyme revealed a four amino acid extension in loop 3 that conferred the exo-mode onto what was,

essentially, an endo-acting scaffold. Mutation of a single acidic residue, D130, which is located in loop 3 and hydrogen bonds to the -2 mannose, conferred an endo-mode of action to the enzyme. The crystal structure also revealed a $^{1}S_{5}$ skew boat conformation for the -1 mannose, supporting previous data for a $B_{2,5}$ transition state for GH26 mannanases (Ducros et al. 2002). Uniquely, the substrate distortion was observed with a wild type enzyme complexed with a native substrate. The substrate also contained a galactose decoration appended to the -1 mannose, which gave insights into how this side chain, present in mannan polymers, inhibit the catalytic performance of mannanases. The two likely mechanisms are: promoting condensation of the hydrolysed glycosidic linkage or through unfavourable steric clashes between the galactose and the +1 mannose. These unique members of GH26 allowed invaluable insights into how glycoside hydrolase mechanisms are a subtle interplay between the active site architecture and that of the substrate. The elucidation of the exo-mode of action displayed by *Cj*Man26C revealed how subtle modifications can lead to dramatic changes in enzyme function. It also suggests a possible role in exo/endo synergy on the surface of the bacterium to help ensure efficient mannan hydrolysis.

Chapter 4 presented the discovery, crystallisation and detailed characterisation of a novel cellulase, *Ct*Cel119, that constitutes a new glycoside hydrolase family, GH119. Cellulose is the major polysaccharide in the plant cell wall, and is actually the most abundant organic molecule in the biosphere. Cellulolytic organisms, such as *Clostridium thermocellum*, express large numbers of cellulases to completely degrade cellulose (Henrissat 1991; Henrissat et al. 1995). Although chemically invariant, and unbranched, cellulose microfibrils can have crystalline, paracrystalline and amorphous regions (Park 2010). Exo/endo synergy between cellobiohydrolases and endo-β-1,4 glucanases is crucial for the efficient degradation of cellulose microfibrils. *Ct*Cel119 is an endo-β-1,4 glucanases, and such enzymes are important as they attack the amorphous/paracrystalline regions creating new non-reducing and reducing ends from which cellobiohydrolases can act. *C. thermocellum* possesses a predicted 30 endo-β-1,4 glucanases from families 5, 8, 9, 44, 48 and 74. The rationale of having so many different endo-β-1,4 glucanases is not immediately evident. It is even less clear why *C. thermocellum* would also express *Ct*Cel119 when its catalytic efficiency is up to 10^{4} fold below that of other β-1,4 glycanases (Harjunpaa et al. 1996; Schulein 1997). The clue could be in the architecture of the of the *Ct*Cel119 substrate binding

cleft. The positive subsites of *Ct*Cel119 accommodate the linear two fold screw axis of crystalline cellulose, while the negative subsites binds to cellooligosaccharides that display a twisted conformation that deviates from that expected of crystalline cellulose. This led to the hypothesis that *Ct*Cel119 may target regions at the boundary between amorphous and crystalline regions of cellulose. It would be tempting to suggest that the many endo-β-1,4 glucanases expressed by *C. thermocellum* may similarly target structurally distinct regions of cellulose microfibrils. This is a difficult hypothesis to prove but if true cellulase mixtures for ligno-cellulose degradation would require far greater optimisation. Discovery of new families such as GH119 allow questions to be raised about why an organism possesses such numbers of particular enzymes, and demonstrates that we still do not fully understand the microbial degradation of cellulose.

Chapter 5 attempted to explore and explain the expansion of genes encoding GH43 enzymes in *C. japonicus,* and resonates with the idea proposed in Chapter 4 that all enzymes in an organism perform a particular role. There are 14 putative GH43 sequences in the *C. japonicus* genome (DeBoy et al. 2008), and only one had been previously characterised (Proctor et al. 2005). There is low functional diversity displayed by the characterised GH43s in the CAZy database and only 55 of ~1200 sequences have been characterised. This led to the hypothesis that the GH43 genes in *C. japonicus* may encode a variety of enzymes with novel activities. This thesis describes the cloning and attempted characterization of eleven of the remaining thirteen uncharacterised GH43 enzymes. The data revealed a novel activity; a sugar beet arabinan specific α-1,2-L-arabinofuranosidase was identified that showed plasticity for single and double substitutions. The structure of the enzyme revealed a curved binding cleft specifically evolved to bind the curved nature of sugar beet arabinan, and a groove in this binding cleft to accommodate α-1,3-L-arabinofuranosides in double substitutions. A single hydrogen bond to a glycosidic oxygen accounted for the enzymes preference for α-1,2-L-arabinofuranosides. This novel activity can be attributed to several amino acids at key positions and yet again, nature is producing dramatic changes in enzyme function through relatively minor changes to the same scaffolds. In addition, two exo-α-1,5-L-arabinofuranosidases were also discovered, one localised to the cytoplasm and the other secreted into the periplasm. This activity is not novel within GH43 but has not been described previously for any other *C. japonicus* enzyme characterised to date. Although the

enzymes display the same activity they perform their roles in different contexts. A potential β-1,4 xylosidase was also identified but the K_M of the enzyme against xylooligosaccharides suggest that its true substrate may be more complex than linear xylooligosaccharides. Another four enzymes displayed trace activity against arabinoxylan polysaccharides, indicating that their natural substrates are more complex than the polysaccharides assessed in this work. For the last three GH43 proteins no activity was detected against any substrate tested. So although these efforts identified a novel glycoside hydrolase and added new GH43 activities to the *C. japonicus* arsenal of GHs, it was not a complete success. There were seven enzymes for which no activity could be ascribed and the key to why there has been an expansion of GH43 in the genome of *C. japonicus* remains elusive. The elucidation of the GH43 conundrum may require a systems biology approach, rather than the characterisation of each individual protein. There may be a subtle interplay where GH43 activities are potentiated by other glycoside hydrolases encoded in the *C. japonicus* genome, such as that suggested for *Ct*Cel119 in Chapter 4. It could also be that within the context of the plant cell wall there are substrates that are not yet available commercially. The likelihood is that there is a reason why evolution has selected expansion of GH43 in several plant cell wall degrading microorganisms. In any event, it is probable that the arsenal of GH43 enzymes provides *C. japonicus* with a competitive edge within the context of plant cell wall degradation. It is similarly unclear how the large number of cellulases expressed by organisms such as *C. thermocellum* act synergistically to degrade cellulose? Why have some particular glycoside hydrolase families been expanded? Some of the answers to these questions can be obtained through work such as that presented in Chapter 3, where it is deviation from the expected enzyme specificities that allow us to explore and answer the real intricate details of glycoside hydrolase function. Not until we fully understand how glycoside hydrolases function, and how their activities contribute to the survival of the endogenous host, will we really find the key to using these enzymes in processes that efficiently degrade the plant cell wall.

Future Work

The endo-acting mutants of *Cj*Man26C appear to produce a greater array of oligosaccharide products, when compared to *Cj*Man26A, from the polysaccharide

galactomannan. This is likely due to the ability of the enzyme and its variants to accommodate galactose appended to the -1 mannose. This should give the endo-acting loop deletion mutants of *Cj*Man26C, ΔN128/L129/D130 and ΔN128/L129/D130/A131, a greater ability to breakdown galactomannan. The conversion of *Cj*Man26C to an endo-mode of action came at a significant catalytic cost. This catalytic cost could be significantly reduced, however, if the interaction of E121 with O6 of the -2 mannose in *Cj*Man26A, could be mimicked in the *Cj*Man26C mutants ΔN128/L129/D130 and ΔN128/L129/D130/A131. *Cj*Man26A is currenly one of the most active mannanases characterised to date, if *Cj*Man26C mutants, ΔN128/L129/D130 and ΔN128/L129/D130/A131, could be engineered to have the same, if not better, catalytic efficiency as *Cj*Man26A it could be a valuable tool in the degradation ligno-cellulosic material. If successful this would be real rational design leading to the creation GHs that are more catalytic efficient at degrading plant cell wall material. Recently in GH26 the structure of a putative β-1,4-endoglucanase, *Pb*CMcase (REF), was solved. It would be interesting to confirm the activity of this enzyme. If it is indeed a β-1,4-endoglucanase detailed characterisation may reveal further insights into GH26. An interesting question would be what transition state the enzyme would utilise, as in β-1,4 glucose polymers the O3 of the -1 sugar is not tethered to the -2 sugar. All other enzymes in GH26 which do not act on β-1,4-mannans, act on β-1,3-linked polymers.

Chapter 4 described the detailed characterisation of *Ct*Cel119, a cellulase that founds a new GH family. Although a great deal was learned about the cellulase no candidate for the catalytic base could be identified and its exact role in cellulose degradation is not completely understood. A "Grotthus mechanism" was proposed where a hydroxyl ion is generated by a remote base and transmitted by a solvent chain to the site of catalysis. If a "Grotthus mechanism" is indeed utilised it should be possible to find the remote base. The remote base is not readily identifiable upon inspection of the crystal structure but alanine scanning of the surface acidic amino acids could reveal its location. Further synergy experiments with other *C. thermocellum* cellulases also need to be conducted to attempt to elucidate the role *Ct*Cel119 plays in cellulose degradation.

Attempts to elucidate the gene expansion of GH43 in *C. japonicus* did not reveal the insights that were hoped. Three enzymes were characterised comprehensively and for eight no definitive activity could be described. Five of these

eight including *Cj*Xyl43A, a suspected xylosidase, showed trace activity against arabinoxylans but not enough for this to be their true substrate. The work highlighted that to elucidate all GH43 activities it may not be enough to clone and express the enzymes as individual entities. The GH43s studied in this work reside in a number of operons encoding a variety of putative carbohydrate active proteins. It may be necessary to clone whole operons and examine not only the individual activities but how they may work in synergy and one activity complement or reveal another.

References

(1994). "The CCP4 suite: programs for protein crystallography." Acta Crystallogr D Biol Crystallogr **50**(Pt 5): 760-3.

Abbott, D. W. and A. B. Boraston (2007). "The structural basis for exopolygalacturonase activity in a family 28 glycoside hydrolase." J Mol Biol **368**(5): 1215-22.

Abbott, D. W., E. Ficko-Blean, A. L. van Bueren, A. Rogowski, A. Cartmell, P. M. Coutinho, B. Henrissat, H. J. Gilbert and A. B. Boraston (2009). "Analysis of the structural and functional diversity of plant cell wall specific family 6 carbohydrate binding modules." Biochemistry **48**(43): 10395-404.

Alzari, P. M., H. Souchon and R. Dominguez (1996). "The crystal structure of endoglucanase CelA, a family 8 glycosyl hydrolase from Clostridium thermocellum." Structure **4**(3): 265-75.

Andersson, S., Samuelson, O, Ishihara, M, Shizium, K (1983). "Struture of the reducing end-groups in spruce xylan." Carbohydrate Research **111**: 283-288.

Armand, S., S. R. Andrews, S. J. Charnock and H. J. Gilbert (2001). "Influence of the aglycone region of the substrate binding cleft of Pseudomonas xylanase 10A on catalysis." Biochemistry **40**(25): 7404-9.

Aspinall, G. (1980). Chemistry of cell wall polysacharides. In The Biochemistry of Plants (A Comprehensive Treaty), New York: Academic Press.

Baumann, M. J., J. M. Eklof, G. Michel, A. M. Kallas, T. T. Teeri, M. Czjzek and H. Brumer, 3rd (2007). "Structural evidence for the evolution of xyloglucanase activity from xyloglucan endo-transglycosylases: biological implications for cell wall metabolism." Plant Cell **19**(6): 1947-63.

Bayer, E. A., L. J. Shimon, Y. Shoham and R. Lamed (1998). "Cellulosomes-structure and ultrastructure." J Struct Biol **124**(2-3): 221-34.

Beguin, P. and J. P. Aubert (1994). "The biological degradation of cellulose." FEMS Microbiol Rev **13**(1): 25-58.

Beylot, M. H., K. Emami, V. A. McKie, H. J. Gilbert and G. Pell (2001). "Pseudomonas cellulosa expresses a single membrane-bound glycoside hydrolase family 51 arabinofuranosidase." Biochem J **358**(Pt 3): 599-605.

Birsan, C., P. Johnson, M. Joshi, A. MacLeod, L. McIntosh, V. Monem, M. Nitz, D. R. Rose, D. Tull, W. W. Wakarchuck, Q. Wang, R. A. Warren, A. White and S. G. Withers (1998). "Mechanisms of cellulases and xylanases." Biochem Soc Trans **26**(2): 156-60.

Boraston, A. B. (2005). "The interaction of carbohydrate-binding modules with insoluble non-crystalline cellulose is enthalpically driven." Biochem J **385**(Pt 2): 479-84.

Boraston, A. B., D. N. Bolam, H. J. Gilbert and G. J. Davies (2004). "Carbohydrate-binding modules: fine-tuning polysaccharide recognition." Biochem J **382**(Pt 3): 769-81.

Braun, C., A. Meinke, L. Ziser and S. G. Withers (1993). "Simultaneous high-performance liquid chromatographic determination of both the cleavage pattern and the stereochemical outcome of the hydrolysis reactions catalyzed by various glycosidases." Anal Biochem **212**(1): 259-62.

Brett, C., Waldren K (1996). Physiology and Biochemistry of Plant Cell Walls, Chapman and Hall edn.

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

Brunzelle, J. S., D. B. Jordan, D. R. McCaslin, A. Olczak and Z. Wawrzak (2008). "Structure of the two-subsite beta-d-xylosidase from Selenomonas ruminantium in complex with 1,3-bis[tris(hydroxymethyl)methylamino]propane." Arch Biochem Biophys **474**(1): 157-66.

Brux, C., A. Ben-David, D. Shallom-Shezifi, M. Leon, K. Niefind, G. Shoham, Y. Shoham and D. Schomburg (2006). "The structure of an inverting GH43 beta-xylosidase from Geobacillus stearothermophilus with its substrate reveals the role of the three catalytic residues." J Mol Biol **359**(1): 97-109.

Brzozowski, A. M., D. M. Lawson, J. P. Turkenburg, H. Bisgaard-Frantzen, A. Svendsen, T. V. Borchert, Z. Dauter, K. S. Wilson and G. J. Davies (2000). "Structural analysis of a chimeric bacterial alpha-amylase. High-resolution analysis of native and ligand complexes." Biochemistry **39**(31): 9099-107.

Burmeister, W. P., S. Cottaz, P. Rollin, A. Vasella and B. Henrissat (2000). "High resolution X-ray crystallography shows that ascorbate is a cofactor for myrosinase and substitutes for the function of the catalytic base." J Biol Chem **275**(50): 39385-93.

Cantarel, B. L., P. M. Coutinho, C. Rancurel, T. Bernard, V. Lombard and B. Henrissat (2009). "The Carbohydrate-Active EnZymes database (CAZy): an expert resource for Glycogenomics." Nucleic Acids Res **37**(Database issue): D233-8.

Carpita (1997). Structure and biosynthesis of plant cell walls. Plant Metabolism. T. D. Dennis DT, Lefebure DD, Layzell DB. Harlow**:** 125-147.

Carvalho, A. L., F. M. Dias, T. Nagy, J. A. Prates, M. R. Proctor, N. Smith, E. A. Bayer, G. J. Davies, L. M. Ferreira, M. J. Romao, C. M. Fontes and H. J. Gilbert (2007). "Evidence for a dual binding mode of dockerin modules to cohesins." Proc Natl Acad Sci U S A **104**(9): 3089-94.

Charnock, S. J., I. E. Brown, J. P. Turkenburg, G. W. Black and G. J. Davies (2002). "Convergent evolution sheds light on the anti-beta -elimination mechanism common to family 1 and 10 polysaccharide lyases." Proc Natl Acad Sci U S A **99**(19): 12067-72.

Cohen SN, Chang ACY, Hsu L (1972) Nonchromosomal antibiotic resistance in bacteria: Genetic transformation of *Escherichia coli* by R-factor DNA. *Proceedings of the National Academy of Sciences of the United States of America* **69:** 2110-2114

Cosgrove, D. J. (2000). "Expansive growth of plant cell walls." Plant Physiol Biochem **38**(1-2): 109-24.

Cosgrove, D. J. (2000). "Loosening of plant cell walls by expansins." Nature **407**(6802): 321-6.

Cosgrove, D. J. (2005). "Growth of the plant cell wall." Nat Rev Mol Cell Biol **6**(11): 850-61.

Coughlan, M. P. (1985). "Cellulose hydrolysis: the potential, the problems and relevant research at Galway." Biochem Soc Trans **13**(2): 405-6.

Coutinho, P. M., Henrissat, B (1999). Carbohydrate-active enzymes: an integrated database approach. In "Recent Advances in Carbohydrate Bioengineering", The Royal Society of Chemistry, Cambridge.

Cowtan, K. (1998). "Modified phased translation functions and their application to molecular-fragment location." Acta Crystallogr D Biol Crystallogr **54**(Pt 5): 750-6.

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

Darvill, A. G., M. McNeil and P. Albersheim (1978). "Structure of Plant Cell Walls: VIII. A New Pectic Polysaccharide." Plant Physiol **62**(3): 418-22.

Dauter, Z., M. Dauter, A. M. Brzozowski, S. Christensen, T. V. Borchert, L. Beier, K. S. Wilson and G. J. Davies (1999). "X-ray structure of Novamyl, the five-domain "maltogenic" alpha-amylase from Bacillus stearothermophilus: maltose and acarbose complexes at 1.7A resolution." Biochemistry **38**(26): 8385-92.

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

Davies, G., Sinnott, M. L. and Withers, S. G (1997). in Comprehensive Biological Catalysis, Academic Press, London.

DeBoy, R. T., E. F. Mongodin, D. E. Fouts, L. E. Tailford, H. Khouri, J. B. Emerson, Y. Mohamoud, K. Watkins, B. Henrissat, H. J. Gilbert and K. E. Nelson (2008). "Insights into plant cell wall degradation from the genome sequence of the soil bacterium Cellvibrio japonicus." J Bacteriol **190**(15): 5455-63.

Ducros, V. M., D. L. Zechel, G. N. Murshudov, H. J. Gilbert, L. Szabo, D. Stoll, S. G. Withers and G. J. Davies (2002). "Substrate distortion by a beta-mannanase: snapshots of the Michaelis and covalent-intermediate complexes suggest a B(2,5) conformation for the transition state." Angew Chem Int Ed Engl **41**(15): 2824-7.

Ebringerova, A., Hromadkova, Z, Heinze, T (2005). "Hemicellulose." Advanced polymer science **186**: 1-67.

Emami, K., T. Nagy, C. M. Fontes, L. M. Ferreira and H. J. Gilbert (2002). "Evidence for temporal regulation of the two Pseudomonas cellulosa xylanases belonging to glycoside hydrolase family 11." J Bacteriol **184**(15): 4124-33.

Emami, K., E. Topakas, T. Nagy, J. Henshaw, K. A. Jackson, K. E. Nelson, E. F. Mongodin, J. W. Murray, R. J. Lewis and H. J. Gilbert (2009). "Regulation of the xylan-degrading apparatus of Cellvibrio japonicus by a novel two-component system." J Biol Chem **284**(2): 1086-96.

Emsley, P. and K. Cowtan (2004). "Coot: model-building tools for molecular graphics." Acta Crystallogr D Biol Crystallogr **60**(Pt 12 Pt 1): 2126-32.

Evans, P. (2006). "Scaling and assessment of data quality." Acta Crystallogr D Biol Crystallogr **62**(Pt 1): 72-82.

Fendri, I., C. Tardif, H. P. Fierobe, S. Lignon, O. Valette, S. Pages and S. Perret (2009). "The cellulosomes from Clostridium cellulolyticum: identification of new components and synergies between complexes." Febs J **276**(11): 3076-86.

Fettke, J., M. Hejazi, J. Smirnova, E. Hochel, M. Stage and M. Steup (2009). "Eukaryotic starch degradation: integration of plastidial and cytosolic pathways." J Exp Bot **60**(10): 2907-22.

Flint, J., D. N. Bolam, D. Nurizzo, E. J. Taylor, M. P. Williamson, C. Walters, G. J. Davies and H. J. Gilbert (2005). "Probing the mechanism of ligand recognition in family 29 carbohydrate-binding modules." J Biol Chem **280**(25): 23718-26.

Fontes, C. M. and H. J. Gilbert "Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates." Annu Rev Biochem **79**: 655-81.

Fushinobu, S., M. Hidaka, Y. Honda, T. Wakagi, H. Shoun and M. Kitaoka (2005). "Structural basis for the specificity of the reducing end xylose-releasing exo-oligoxylanase from Bacillus halodurans C-125." J Biol Chem **280**(17): 17180-6.

Gal, L., C. Gaudin, A. Belaich, S. Pages, C. Tardif and J. P. Belaich (1997). "CelG from Clostridium cellulolyticum: a multidomain endoglucanase acting efficiently on crystalline cellulose." J Bacteriol **179**(21): 6595-601.

Gibeaut, D. M., M. Pauly, A. Bacic and G. B. Fincher (2005). "Changes in cell wall polysaccharides in developing barley (Hordeum vulgare) coleoptiles." Planta **221**(5): 729-38.

Gill SC, Von Hippel PH (1989) Calculation of protein extinction coefficients from amino acid sequence data. Anal Biochem **182**: 319-326

Gilkes, N. R., D. G. Kilburn, R. C. Miller, Jr. and R. A. Warren (1989). "Structural and functional analysis of a bacterial cellulase by proteolysis." J Biol Chem **264**(30): 17802-8.

Gloster, T. M., P. Meloncelli, R. V. Stick, D. Zechel, A. Vasella and G. J. Davies (2007). "Glycosidase inhibition: an assessment of the binding of 18 putative transition-state mimics." J Am Chem Soc **129**(8): 2345-54.

Glushka, J. N., M. Terrell, W. S. York, M. A. O'Neill, A. Gucwa, A. G. Darvill, P. Albersheim and J. H. Prestegard (2003). "Primary structure of the 2-O-methyl-alpha-L-fucose-containing side chain of the pectic polysaccharide, rhamnogalacturonan II." Carbohydr Res **338**(4): 341-52.

Gold, N. D. and V. J. Martin (2007). "Global view of the Clostridium thermocellum cellulosome revealed by quantitative proteomic analysis." J Bacteriol **189**(19): 6787-95.

Gong, C. S., N. J. Cao, J. Du and G. T. Tsao (1999). "Ethanol production from renewable resources." Adv Biochem Eng Biotechnol **65**: 207-41.

Graber, M. L., D. C. DiLillo, B. L. Friedman and E. Pastoriza-Munoz (1986). "Characteristics of fluoroprobes for measuring intracellular pH." Anal Biochem **156**(1): 202-12.

Guimaraes, B. G., H. Souchon, B. L. Lytle, J. H. David Wu and P. M. Alzari (2002). "The crystal structure and catalytic mechanism of cellobiohydrolase CelS, the major enzymatic component of the Clostridium thermocellum Cellulosome." J Mol Biol **320**(3): 587-96.

Hall, J., G. W. Black, L. M. Ferreira, S. J. Millward-Sadler, B. R. Ali, G. P. Hazlewood and H. J. Gilbert (1995). "The non-catalytic cellulose-binding domain of a novel cellulase from Pseudomonas fluorescens subsp. cellulosa is important for the efficient hydrolysis of Avicel." Biochem J **309 (Pt 3)**: 749-56.

Halstead, J. R., M. P. Fransen, R. Y. Eberhart, A. J. Park, H. J. Gilbert and G. P. Hazlewood (2000). "alpha-Galactosidase A from Pseudomonas fluorescens subsp. cellulosa: cloning, high level expression and its role in galactomannan hydrolysis." FEMS Microbiol Lett **192**(2): 197-203.

Harjunpaa, V., A. Teleman, A. Koivula, L. Ruohonen, T. T. Teeri, O. Teleman and T. Drakenberg (1996). "Cello-oligosaccharide hydrolysis by cellobiohydrolase II from Trichoderma reesei. Association and rate constants derived from an analysis of progress curves." Eur J Biochem **240**(3): 584-91.

Harris, G. W., J. A. Jenkins, I. Connerton, N. Cummings, L. Lo Leggio, M. Scott, G. P. Hazlewood, J. I. Laurie, H. J. Gilbert and R. W. Pickersgill (1994). "Structure of the catalytic core of the family F xylanase from Pseudomonas

fluorescens and identification of the xylopentaose-binding sites." Structure **2**(11): 1107-16.
Hazlewood, G. P. and H. J. Gilbert (1998). "Structure and function analysis of Pseudomonas plant cell wall hydrolases." Prog Nucleic Acid Res Mol Biol **61**: 211-41.
He, Y., M. S. Macauley, K. A. Stubbs, D. J. Vocadlo and G. J. Davies "Visualizing the reaction coordinate of an O-GlcNAc hydrolase." J Am Chem Soc **132**(6): 1807-9.
Hehre, E. J., C. F. Brewer and D. S. Genghof (1979). "Scope and mechanism of carbohydrase action. Hydrolytic and nonhydrolytic actions of beta-amylase on alpha- and beta-maltosyl fluoride." J Biol Chem **254**(13): 5942-50.
Henrissat, B. (1991a). "A classification of glycosyl hydrolases based on amino acid sequence similarities." Biochem J **280 (Pt 2)**: 309-16.
Henrissat, B. (1991b). "Sequence homology between a beta-galactosidase and some beta-glucosidases." Protein Seq Data Anal **4**(1): 61-2.
Henrissat, B., I. Callebaut, S. Fabrega, P. Lehn, J. P. Mornon and G. Davies (1995). "Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases." Proc Natl Acad Sci U S A **92**(15): 7090-4.
Henrissat, B., I. Callebaut, S. Fabrega, P. Lehn, J. P. Mornon and G. Davies (1996). "Conserved catalytic machinery and the prediction of a common fold for several families of glycosyl hydrolases." Proc Natl Acad Sci U S A **93**(11): 5674.
Henrissat, B. and G. Davies (1997). "Structural and sequence-based classification of glycoside hydrolases." Curr Opin Struct Biol **7**(5): 637-44.
Henrissat, B. and A. Romeu (1995). "Families, superfamilies and subfamilies of glycosyl hydrolases." Biochem J **311 (Pt 1)**: 350-1.
Himmel, M. E. and E. A. Bayer (2009). "Lignocellulose conversion to biofuels: current challenges, global perspectives." Curr Opin Biotechnol **20**(3): 316-7.
Himmel, M. E., M. F. Ruth and C. E. Wyman (1999). "Cellulase for commodity products from cellulosic biomass." Curr Opin Biotechnol **10**(4): 358-64.
Hofte, H. (2010). "Plant cell biology: how to pattern a wall." Curr Biol **20**(10): R450-2.
Hogg, D., G. Pell, P. Dupree, F. Goubet, S. M. Martin-Orue, S. Armand and H. J. Gilbert (2003). "The modular architecture of Cellvibrio japonicus mannanases in glycoside hydrolase families 5 and 26 points to differences in their role in mannan degradation." Biochem J **371**(Pt 3): 1027-43.
Hogg, D., E. J. Woo, D. N. Bolam, V. A. McKie, H. J. Gilbert and R. W. Pickersgill (2001). "Crystal structure of mannanase 26A from Pseudomonas cellulosa and analysis of residues involved in substrate binding." J Biol Chem **276**(33): 31186-92.
Honda, Y. and M. Kitaoka (2006). "The first glycosynthase derived from an inverting glycoside hydrolase." J Biol Chem **281**(3): 1426-31.
Humphry, D. R., G. W. Black and S. P. Cummings (2003). "Reclassification of 'Pseudomonas fluorescens subsp. cellulosa' NCIMB 10462 (Ueda et al. 1952) as Cellvibrio japonicus sp. nov. and revival of Cellvibrio vulgaris sp. nov., nom. rev. and Cellvibrio fulvus sp. nov., nom. rev." Int J Syst Evol Microbiol **53**(Pt 2): 393-400.
Jenkins, J. and R. Pickersgill (2001). "The architecture of parallel beta-helices and related folds." Prog Biophys Mol Biol **77**(2): 111-75.

Johansson, M., Samuelson, O (1977). "Reducing end groups in birch xylan and their alkaline degradation." Wood science technology **11**: 251-263.

Johansson, P., H. Brumer, 3rd, M. J. Baumann, A. M. Kallas, H. Henriksson, S. E. Denman, T. T. Teeri and T. A. Jones (2004). "Crystal structures of a poplar xyloglucan endotransglycosylase reveal details of transglycosylation acceptor binding." Plant Cell **16**(4): 874-86.

Karlsen, S., E. Hough, Z. H. Rao and N. W. Isaacs (1996). "Structure of a bulgecin-inhibited g-type lysozyme from the egg white of the Australian black swan. A comparison of the binding of bulgecin to three muramidases." Acta Crystallogr D Biol Crystallogr **52**(Pt 1): 105-14.

Kitago, Y., S. Karita, N. Watanabe, M. Kamiya, T. Aizawa, K. Sakka and I. Tanaka (2007). "Crystal structure of Cel44A, a glycoside hydrolase family 44 endoglucanase from Clostridium thermocellum." J Biol Chem **282**(49): 35703-11.

Kleywegt, G. J., J. Y. Zou, C. Divne, G. J. Davies, I. Sinning, J. Stahlberg, T. Reinikainen, M. Srisodsuk, T. T. Teeri and T. A. Jones (1997). "The crystal structure of the catalytic core domain of endoglucanase I from Trichoderma reesei at 3.6 A resolution, and a comparison with related enzymes." J Mol Biol **272**(3): 383-97.

Kobayashi, M., T. Matoh and J. Azuma (1996). "Two Chains of Rhamnogalacturonan II Are Cross-Linked by Borate-Diol Ester Bonds in Higher Plant Cell Walls." Plant Physiol **110**(3): 1017-1020.

Koshland, D. E., Jr., Z. Budenstein and A. Kowalsky (1954). "Mechanism of hydrolysis of adenosinetriphosphate catalyzed by purified muscle proteins." J Biol Chem **211**(1): 279-87.

Krissinel, E. and K. Henrick (2004). "Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions." Acta Crystallogr D Biol Crystallogr **60**(Pt 12 Pt 1): 2256-68.

Lamelli UK (1972) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature **227**: 680-685

Langer, G., S. X. Cohen, V. S. Lamzin and A. Perrakis (2008). "Automated macromolecular model building for X-ray crystallography using ARP/wARP version 7." Nat Protoc **3**(7): 1171-9.

Larson, S. B., J. Day, A. P. Barba de la Rosa, N. T. Keen and A. McPherson (2003). "First crystallographic structure of a xylanase from glycoside hydrolase family 5: implications for catalysis." Biochemistry **42**(28): 8411-22.

Lehtio, J., J. Sugiyama, M. Gustavsson, L. Fransson, M. Linder and T. T. Teeri (2003). "The binding specificity and affinity determinants of family 1 and family 3 cellulose binding modules." Proc Natl Acad Sci U S A **100**(2): 484-9.

Lerouxel, O., D. M. Cavalier, A. H. Liepman and K. Keegstra (2006). "Biosynthesis of plant cell wall polysaccharides - a complex process." Curr Opin Plant Biol **9**(6): 621-30.

Leslie, A. G. W. (1992). "Recent changes to the MOSFLM package for processing film and image plate data " Acta Crystallogr D Biol Crystallogr **No. 26**.

Li, J., J. B. Zhao, M. Zhao, Y. L. Yang, W. H. Jiang and S. Yang "Screening and characterization of butanol-tolerant micro-organisms." Lett Appl Microbiol **50**(4): 373-9.

Ling, Z., M. D. Suits, R. J. Bingham, N. C. Bruce, G. J. Davies, A. J. Fairbanks, J. W. Moir and E. J. Taylor (2009). "The X-ray crystal structure of an Arthrobacter protophormiae endo-beta-N-acetylglucosaminidase reveals a (beta/alpha)(8)

catalytic domain, two ancillary domains and active site residues key for transglycosylation activity." J Mol Biol **389**(1): 1-9.

Liu, Q. P., G. Sulzenbacher, H. Yuan, E. P. Bennett, G. Pietz, K. Saunders, J. Spence, E. Nudelman, S. B. Levery, T. White, J. M. Neveu, W. S. Lane, Y. Bourne, M. L. Olsson, B. Henrissat and H. Clausen (2007). "Bacterial glycosidases for the production of universal red blood cells." Nat Biotechnol **25**(4): 454-64.

Liu, Z., Y. Ying, F. Li, C. Ma and P. Xu "Butanol production by Clostridium beijerinckii ATCC 55025 from wheat bran." J Ind Microbiol Biotechnol **37**(5): 495-501.

Ly, H. D. and S. G. Withers (1999). "Mutagenesis of glycosidases." Annu Rev Biochem **68**: 487-522.

MacKenzie, L. F., G. Sulzenbacher, C. Divne, T. A. Jones, H. F. Woldike, M. Schulein, S. G. Withers and G. J. Davies (1998). "Crystal structure of the family 7 endoglucanase I (Cel7B) from Humicola insolens at 2.2 A resolution and identification of the catalytic nucleophile by trapping of the covalent glycosyl-enzyme intermediate." Biochem J **335 (Pt 2)**: 409-16.

MacKenzie, L. F., Wang, Q.P., Warren, and Withers, S.G. (1998). "Glycosynthases: mutant glycosidases for oligosaccharide synthesis." J. Am. Chem. Soc **120**: 5583-5584.

MacLeod, A. M., T. Lindhorst, S. G. Withers and R. A. Warren (1994). "The acid/base catalyst in the exoglucanase/xylanase from Cellulomonas fimi is glutamic acid 127: evidence from detailed kinetic studies of mutants." Biochemistry **33**(20): 6371-6.

Mark, B. L., D. J. Vocadlo, D. Zhao, S. Knapp, S. G. Withers and M. N. James (2001). "Biochemical and structural assessment of the 1-N-azasugar GalNAc-isofagomine as a potent family 20 beta-N-acetylhexosaminidase inhibitor." J Biol Chem **276**(45): 42131-7.

Mayer, C., D. L. Zechel, S. P. Reid, R. A. Warren and S. G. Withers (2000). "The E358S mutant of Agrobacterium sp. beta-glucosidase is a greatly improved glycosynthase." FEBS Lett **466**(1): 40-4.

McCoy, A. J., R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni and R. J. Read (2007). "Phaser crystallographic software." J Appl Crystallogr **40**(Pt 4): 658-674.

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

Megazyme. (2010). "Arabinan (sugarbeet)." from http://www.megazyme.com/Dynamic.aspx?control=CSViewProduct&categoryName=Polysaccharides&productId=P-ARAB.

Miller, L. G. (1959). " Use of dinitrosalicylic acid reagent for determination of reducing sugar " Anal Chem **31**(3): 426-428.

Mohnen, D. (2008). "Pectin structure and biosynthesis." Curr Opin Plant Biol **11**(3): 266-77.

Moller, I., I. Sorensen, A. J. Bernal, C. Blaukopf, K. Lee, J. Obro, F. Pettolino, A. Roberts, J. D. Mikkelsen, J. P. Knox, A. Bacic and W. G. Willats (2007). "High-throughput mapping of cell-wall polymers within and between plants using novel microarrays." Plant J **50**(6): 1118-28.

Money, V. A., A. Cartmell, C. I. Guerreiro, V. M. Ducros, C. M. Fontes, H. J. Gilbert and G. J. Davies (2008). "Probing the beta-1,3:1,4 glucanase, CtLic26A, with a thio-oligosaccharide and enzyme variants." Org Biomol Chem **6**(5): 851-3.

Money, V. A., N. L. Smith, A. Scaffidi, R. V. Stick, H. J. Gilbert and G. J. Davies (2006). "Substrate distortion by a lichenase highlights the different conformational itineraries harnessed by related glycoside hydrolases." Angew Chem Int Ed Engl **45**(31): 5136-40.

Montanier, C., V. A. Money, V. M. Pires, J. E. Flint, B. A. Pinheiro, A. Goyal, J. A. Prates, A. Izumi, H. Stalbrand, C. Morland, A. Cartmell, K. Kolenova, E. Topakas, E. J. Dodson, D. N. Bolam, G. J. Davies, C. M. Fontes and H. J. Gilbert (2009). "The active site of a carbohydrate esterase displays divergent catalytic and noncatalytic binding functions." PLoS Biol **7**(3): e71.

Montanier, C., A. L. van Bueren, C. Dumon, J. E. Flint, M. A. Correia, J. A. Prates, S. J. Firbank, R. J. Lewis, G. G. Grondin, M. G. Ghinet, T. M. Gloster, C. Herve, J. P. Knox, B. G. Talbot, J. P. Turkenburg, J. Kerovuo, R. Brzezinski, C. M. Fontes, G. J. Davies, A. B. Boraston and H. J. Gilbert (2009). "Evidence that family 35 carbohydrate binding modules display conserved specificity but divergent function." Proc Natl Acad Sci U S A **106**(9): 3065-70.

Moreira, L. R. and E. X. Filho (2008). "An overview of mannan structure and mannan-degrading enzyme systems." Appl Microbiol Biotechnol **79**(2): 165-78.

Moxley, G., Z. Zhu and Y. H. Zhang (2008). "Efficient sugar release by the cellulose solvent-based lignocellulose fractionation technology and enzymatic cellulose hydrolysis." J Agric Food Chem **56**(17): 7885-90.

Mullis KB, Faloona FA (1987) specific sysnthesis of DNA in vitro via polymerase-catalysed chain reaction. *Methods Enzymol* **155**: 335-350

Muller, M., R. Hori, T. Itoh and J. Sugiyama (2002). "X-ray microbeam and electron diffraction experiments on developing xylem cell walls." Biomacromolecules **3**(1): 182-6.

Murshudov, G. N., A. A. Vagin and E. J. Dodson (1997). "Refinement of macromolecular structures by the maximum-likelihood method." Acta Crystallogr D Biol Crystallogr **53**(Pt 3): 240-55.

Pace CN, Vajdos F, Fee L, Grimsley G and Gray T (1995) How to measure and predict molar absorption coefficient of a protein. Protein Sci 4: 2411-2423

Nagae, M., A. Tsuchiya, T. Katayama, K. Yamamoto, S. Wakatsuki and R. Kato (2007). "Structural basis of the catalytic reaction mechanism of novel 1,2-alpha-L-fucosidase from Bifidobacterium bifidum." J Biol Chem **282**(25): 18497-509.

Nagy, T., D. Nurizzo, G. J. Davies, P. Biely, J. H. Lakey, D. N. Bolam and H. J. Gilbert (2003). "The alpha-glucuronidase, GlcA67A, of Cellvibrio japonicus utilizes the carboxylate and methyl groups of aldobiouronic acid as important substrate recognition determinants." J Biol Chem **278**(22): 20286-92.

Namchuk, M. N. and S. G. Withers (1995). "Mechanism of Agrobacterium beta-glucosidase: kinetic analysis of the role of noncovalent enzyme/substrate interactions." Biochemistry **34**(49): 16194-202.

Notenboom, V., A. B. Boraston, D. G. Kilburn and D. R. Rose (2001). "Crystal structures of the family 9 carbohydrate-binding module from Thermotoga maritima xylanase 10A in native and ligand-bound forms." Biochemistry **40**(21): 6248-56.

Nurizzo, D., J. P. Turkenburg, S. J. Charnock, S. M. Roberts, E. J. Dodson, V. A. McKie, E. J. Taylor, H. J. Gilbert and G. J. Davies (2002). "Cellvibrio japonicus alpha-L-arabinanase 43A has a novel five-blade beta-propeller fold." Nat Struct Biol **9**(9): 665-8.

Ochiai, A., T. Itoh, B. Mikami, W. Hashimoto and K. Murata (2009). "Structural determinants responsible for substrate recognition and mode of action in family 11 polysaccharide lyases." J Biol Chem **284**(15): 10181-9.

Oomen, R. J., C. H. Doeswijk-Voragen, M. S. Bush, J. P. Vincken, B. Borkhardt, L. A. van den Broek, J. Corsar, P. Ulvskov, A. G. Voragen, M. C. McCann and R. G. Visser (2002). "In muro fragmentation of the rhamnogalacturonan I backbone in potato (Solanum tuberosum L.) results in a reduction and altered location of the galactan and arabinan side-chains and abnormal periderm development." Plant J **30**(4): 403-13.

Oshiro, M., K. Hanada, Y. Tashiro and K. Sonomoto "Efficient conversion of lactic acid to butanol with pH-stat continuous lactic acid and glucose feeding method by Clostridium saccharoperbutylacetonicum." Appl Microbiol Biotechnol.

Park S, B. J., Himmel ME, Parilla PA, Johnson DK. (2010). "Cellulose crystallinity index: measurement techniques and their impact on interpreting cellulase performance." Biotechnology for biofuels **24**(3): 3-10.

Parsiegla, G., M. Juy, C. Reverbel-Leroy, C. Tardif, J. P. Belaich, H. Driguez and R. Haser (1998). "The crystal structure of the processive endocellulase CelF of Clostridium cellulolyticum in complex with a thiooligosaccharide inhibitor at 2.0 A resolution." Embo J **17**(19): 5551-62.

Parsiegla, G., C. Reverbel, C. Tardif, H. Driguez and R. Haser (2008). "Structures of mutants of cellulase Cel48F of Clostridium cellulolyticum in complex with long hemithiocellooligosaccharides give rise to a new view of the substrate pathway during processive action." J Mol Biol **375**(2): 499-510.

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

Pell, G., L. Szabo, S. J. Charnock, H. Xie, T. M. Gloster, G. J. Davies and H. J. Gilbert (2004). "Structural and biochemical analysis of Cellvibrio japonicus xylanase 10C: how variation in substrate-binding cleft influences the catalytic profile of family GH-10 xylanases." J Biol Chem **279**(12): 11777-88.

Pena, M. J., A. G. Darvill, S. Eberhard, W. S. York and M. A. O'Neill (2008). "Moss and liverwort xyloglucans contain galacturonic acid and are structurally distinct from the xyloglucans synthesized by hornworts and vascular plants." Glycobiology **18**(11): 891-904.

Popper, Z. A. (2008). "Evolution and diversity of green plant cell walls." Curr Opin Plant Biol **11**(3): 286-92.

Proctor, M. R., E. J. Taylor, D. Nurizzo, J. P. Turkenburg, R. M. Lloyd, M. Vardakou, G. J. Davies and H. J. Gilbert (2005). "Tailored catalysts for plant cell-wall degradation: redesigning the exo/endo preference of Cellvibrio japonicus arabinanase 43A." Proc Natl Acad Sci U S A **102**(8): 2697-702.

Reiter, W. D. (2002). "Biosynthesis and properties of the plant cell wall." Curr Opin Plant Biol **5**(6): 536-42.

Ryden, P., K. Sugimoto-Shirasu, A. C. Smith, K. Findlay, W. D. Reiter and M. C. McCann (2003). "Tensile properties of Arabidopsis cell walls depend on both a xyloglucan cross-linked microfibrillar network and rhamnogalacturonan II-borate complexes." Plant Physiol **132**(2): 1033-40.

Rypniewski, W. R., H. M. Holden and I. Rayment (1993). "Structural consequences of reductive methylation of lysine residues in hen egg white lysozyme: an X-ray analysis at 1.8-A resolution." Biochemistry **32**(37): 9851-8.

Sakaguchi, K., Kawamura, T., Watanabe, N., Kiyohara, M., Yamaguchi, K., Ito, M., Tanaka, I. (2010). "Atomic resolution analysis of beta-1,3-xylanase catalytic module from Vibrio sp. AX-4
".
Sanz-Aparicio, J., J. A. Hermoso, M. Martinez-Ripoll, J. L. Lequerica and J. Polaina (1998). "Crystal structure of beta-glucosidase A from Bacillus polymyxa: insights into the catalytic activity in family 1 glycosyl hydrolases." J Mol Biol **275**(3): 491-502.
Saxena, I. M., R. M. Brown, Jr. and T. Dandekar (2001). "Structure--function characterization of cellulose synthase: relationship to other glycosyltransferases." Phytochemistry **57**(7): 1135-48.
Scheller, H. V. (2010). "Hemicelluloses." Annu Rev Plant Biol. **2**(61): 263-289.
Schulein, M. (1997). "Enzymatic properties of cellulases from Humicola insolens." J Biotechnol **57**(1-3): 71-81.
Sergeyev, I. and G. Moyna (2005). "Determination of the three-dimensional structure of oligosaccharides in the solid state from experimental 13C NMR data and ab initio chemical shift surfaces." Carbohydr Res **340**(6): 1165-74.
Seyedarabi, A., T. T. To, S. Ali, S. Hussain, M. Fries, R. Madsen, M. H. Clausen, S. Teixteira, K. Brocklehurst and R. W. Pickersgill "Structural insights into substrate specificity and the anti beta-elimination mechanism of pectate lyase." Biochemistry **49**(3): 539-46.
Shaikh, F. A. and S. G. Withers (2008). "Teaching old enzymes new tricks: engineering and evolution of glycosidases and glycosyl transferases for improved glycoside synthesis." Biochem Cell Biol **86**(2): 169-77.
Shallom, D., V. Belakhov, D. Solomon, G. Shoham, T. Baasov and Y. Shoham (2002). "Detailed kinetic analysis and identification of the nucleophile in alpha-L-arabinofuranosidase from Geobacillus stearothermophilus T-6, a family 51 glycoside hydrolase." J Biol Chem **277**(46): 43667-73.
Shatalov, A. A., D. V. Evtuguin and C. Pascoal Neto (1999). "(2-O-alpha-D-galactopyranosyl-4-O-methyl-alpha-D-glucurono)-D-xylan from Eucalyptus globulus Labill." Carbohydr Res **320**(1-2): 93-9.
Sheldrick, G. M. (2008). "A short history of SHELX." Acta Crystallogr A **64**(Pt 1): 112-22.
Simpson, P. J., H. Xie, D. N. Bolam, H. J. Gilbert and M. P. Williamson (2000). "The structural basis for the ligand specificity of family 2 carbohydrate-binding modules." J Biol Chem **275**(52): 41137-42.
Sinnott, M. L. (1990). "Catalytic mechanism of glycosyl transfer." Chem Rev **90**: 1171-1190.
Smith, B. J., J. L. McKimm-Breshkin, M. McDonald, R. T. Fernley, J. N. Varghese and P. M. Colman (2002). "Structural studies of the resistance of influenza virus neuramindase to inhibitors." J Med Chem **45**(11): 2207-12.
Somerville, C. (2007). "Biofuels." Curr Biol **17**(4): R115-9.
Sorensen, H. R., C. T. Jorgensen, C. H. Hansen, C. I. Jorgensen, S. Pedersen and A. S. Meyer (2006). "A novel GH43 alpha-L-arabinofuranosidase from Humicola insolens: mode of action and synergy with GH51 alpha-L-arabinofuranosidases on wheat arabinoxylan." Appl Microbiol Biotechnol **73**(4): 850-61.
Stoddart, J. (1971). Stereochemistry of Carbohydrates
Stone, B., Clarke, AE (1992). Chemistry and Biology of (1→3)-β-Glucans, Bundoora: La Trobe Univ. Press.

Stryer, L. B., J and Tymoczko, J (2001). Biochemistry, W. H. Freeman and Company.
Tailford, L. E. (2007). The catalytic recognition of Mannose, Newcastle University.
Tailford, L. E., V. M. Ducros, J. E. Flint, S. M. Roberts, C. Morland, D. L. Zechel, N. Smith, M. E. Bjornvad, T. V. Borchert, K. S. Wilson, G. J. Davies and H. J. Gilbert (2009). "Understanding how diverse beta-mannanases recognize heterogeneous substrates." Biochemistry **48**(29): 7009-18.
Tailford, L. E., V. A. Money, N. L. Smith, C. Dumon, G. J. Davies and H. J. Gilbert (2007). "Mannose foraging by Bacteroides thetaiotaomicron: structure and specificity of the beta-mannosidase, BtMan2A." J Biol Chem **282**(15): 11291-9.
Tailford, L. E., W. A. Offen, N. L. Smith, C. Dumon, C. Morland, J. Gratien, M. P. Heck, R. V. Stick, Y. Bleriot, A. Vasella, H. J. Gilbert and G. J. Davies (2008). "Structural and biochemical evidence for a boat-like transition state in beta-mannosidases." Nat Chem Biol **4**(5): 306-12.
Tanaka, T. H., WC, Noguchi, M Kobayashi, A and Shoda SI (2009). "Direct synthesis of 1,6-anhydro sugars from unprotected glycopyranoses
by using 2-chloro-1,3-dimethylimidazolinium chloride." Tetrahedron Letters.
Taylor, E. J., A. Goyal, C. I. Guerreiro, J. A. Prates, V. A. Money, N. Ferry, C. Morland, A. Planas, J. A. Macdonald, R. V. Stick, H. J. Gilbert, C. M. Fontes and G. J. Davies (2005). "How family 26 glycoside hydrolases orchestrate catalysis on different polysaccharides: structure and activity of a Clostridium thermocellum lichenase, CtLic26A." J Biol Chem **280**(38): 32761-7.
Taylor, E. J., N. L. Smith, J. P. Turkenburg, S. D'Souza, H. J. Gilbert and G. J. Davies (2006). "Structural insight into the ligand specificity of a thermostable family 51 arabinofuranosidase, Araf51, from Clostridium thermocellum." Biochem J **395**(1): 31-7.
Teeri (1997). "Crystalline cellulose degradation: new insight into the function of cellobiohydrolases." Trends in Biotechnology **27**(Pt1): 19-24.
Thunnissen, A. M., H. J. Rozeboom, K. H. Kalk and B. W. Dijkstra (1995). "Structure of the 70-kDa soluble lytic transglycosylase complexed with bulgecin A. Implications for the enzymatic mechanism." Biochemistry **34**(39): 12729-37.
Tormo, J., R. Lamed, A. J. Chirino, E. Morag, E. A. Bayer, Y. Shoham and T. A. Steitz (1996). "Crystal structure of a bacterial family-III cellulose-binding domain: a general mechanism for attachment to cellulose." Embo J **15**(21): 5739-51.
Tull, D. and S. G. Withers (1994). "Mechanisms of cellulases and xylanases: a detailed kinetic study of the exo-beta-1,4-glycanase from Cellulomonas fimi." Biochemistry **33**(20): 6363-70.
Ubhayasekera, W., C. M. Tang, S. W. Ho, G. Berglund, T. Bergfors, M. L. Chye and S. L. Mowbray (2007). "Crystal structures of a family 19 chitinase from Brassica juncea show flexibility of binding cleft loops." FEBS J **274**(14): 3695-703.
Vaaje-Kolstad, G., D. R. Houston, A. H. Riemen, V. G. Eijsink and D. M. van Aalten (2005). "Crystal structure and binding properties of the Serratia marcescens chitin-binding protein CBP21." J Biol Chem **280**(12): 11313-9.
van Aalten, D. M., D. Komander, B. Synstad, S. Gaseidnes, M. G. Peter and V. G. Eijsink (2001). "Structural insights into the catalytic mechanism of a family 18 exo-chitinase." Proc Natl Acad Sci U S A **98**(16): 8979-84.

van Asselt, E. J., A. M. Thunnissen and B. W. Dijkstra (1999). "High resolution crystal structures of the Escherichia coli lytic transglycosylase Slt70 and its complex with a peptidoglycan fragment." J Mol Biol **291**(4): 877-98.

van Bueren, A. L., C. Morland, H. J. Gilbert and A. B. Boraston (2005). "Family 6 carbohydrate binding modules recognize the non-reducing end of beta-1,3-linked glucans by presenting a unique ligand binding surface." J Biol Chem **280**(1): 530-7.

Van Petegem, F., T. Collins, M. A. Meuwis, C. Gerday, G. Feller and J. Van Beeumen (2003). "The structure of a cold-adapted family 8 xylanase at 1.3 A resolution. Structural adaptations to cold and investgation of the active site." J Biol Chem **278**(9): 7531-9.

Vandermarliere, E., T. M. Bourgois, S. Rombouts, S. Van Campenhout, G. Volckaert, S. V. Strelkov, J. A. Delcour, A. Rabijns and C. M. Courtin (2008). "Crystallographic analysis shows substrate binding at the -3 to +1 active-site subsites and at the surface of glycoside hydrolase family 11 endo-1,4-beta-xylanases." Biochem J **410**(1): 71-9.

Vandermarliere, E., T. M. Bourgois, M. D. Winn, S. van Campenhout, G. Volckaert, J. A. Delcour, S. V. Strelkov, A. Rabijns and C. M. Courtin (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." Biochem J **418**(1): 39-47.

Vanholme, R., B. Demedts, K. Morreel, J. Ralph and W. Boerjan "Lignin biosynthesis and structure." Plant Physiol **153**(3): 895-905.

Varrot, A., S. Hastrup, M. Schulein and G. J. Davies (1999). "Crystal structure of the catalytic core domain of the family 6 cellobiohydrolase II, Cel6A, from Humicola insolens, at 1.92 A resolution." Biochem J **337 (Pt 2)**: 297-304.

Varrot, A., M. Schulein and G. J. Davies (1999). "Structural changes of the active site tunnel of Humicola insolens cellobiohydrolase, Cel6A, upon oligosaccharide binding." Biochemistry **38**(28): 8884-91.

Varrot, A., V. L. Yip, Y. Li, S. S. Rajan, X. Yang, W. F. Anderson, J. Thompson, S. G. Withers and G. J. Davies (2005). "NAD+ and metal-ion dependent hydrolysis by family 4 glycosidases: structural insight into specificity for phospho-beta-D-glucosides." J Mol Biol **346**(2): 423-35.

Vidal, S., T. Doco, P. Williams, P. Pellerin, W. S. York, M. A. O'Neill, J. Glushka, A. G. Darvill and P. Albersheim (2000). "Structural characterization of the pectic polysaccharide rhamnogalacturonan II: evidence for the backbone location of the aceric acid-containing oligoglycosyl side chain." Carbohydr Res **326**(4): 277-94.

Vrieze, A., F. Holleman, E. G. Zoetendal, W. M. de Vos, J. B. Hoekstra and M. Nieuwdorp "The environment within: how gut microbiota may influence metabolism and body composition." Diabetologia **53**(4): 606-13.

Wada, J., Y. Honda, M. Nagae, R. Kato, S. Wakatsuki, T. Katayama, H. Taniguchi, H. Kumagai, M. Kitaoka and K. Yamamoto (2008). "1,2-alpha-l-Fucosynthase: a glycosynthase derived from an inverting alpha-glycosidase with an unusual reaction mechanism." FEBS Lett **582**(27): 3739-43.

Watts, A. G., I. Damager, M. L. Amaya, A. Buschiazzo, P. Alzari, A. C. Frasch and S. G. Withers (2003). "Trypanosoma cruzi trans-sialidase operates through a covalent sialyl-enzyme intermediate: tyrosine is the catalytic nucleophile." J Am Chem Soc **125**(25): 7532-3.

Wenger, J. W., K. Schwartz and G. Sherlock "Bulk segregant analysis by high-throughput sequencing reveals a novel xylose utilization gene from Saccharomyces cerevisiae." PLoS Genet **6**(5): e1000942.

Wolf, S., G. Mouille and J. Pelloux (2009). "Homogalacturonan methyl-esterification and plant development." Mol Plant **2**(5): 851-60.

Wolfenden, R. and M. J. Snider (2001). "The depth of chemical time and the power of enzymes as catalysts." Acc Chem Res **34**(12): 938-45.

Wolfenden, R. and Y. Yuan (2008). "Rates of spontaneous cleavage of glucose, fructose, sucrose, and trehalose in water, and the catalytic proficiencies of invertase and trehalas." J Am Chem Soc **130**(24): 7548-9.

Wood, T. (1988). "Preparation of crystalline, amorphous, and dyed cellulose substrates." Methods in Enzymology **160**: 19-25.

York, W. S., V. S. Kumar Kolli, R. Orlando, P. Albersheim and A. G. Darvill (1996). "The structures of arabinoxyloglucans produced by solanaceous plants." Carbohydr Res **285**: 99-128.

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

Zhang, X., Y. Shen, W. Shi and X. Bao "Ethanolic cofermentation with glucose and xylose by the recombinant industrial strain Saccharomyces cerevisiae NAN-127 and the effect of furfural on xylitol production." Bioresour Technol **101**(18): 7104-7110.

Zhang, Y. H., S. Y. Ding, J. R. Mielenz, J. B. Cui, R. T. Elander, M. Laser, M. E. Himmel, J. R. McMillan and L. R. Lynd (2007). "Fractionating recalcitrant lignocellulose at modest reaction conditions." Biotechnol Bioeng **97**(2): 214-23.

Zhou, X., J. A. Smith, F. M. Oi, P. G. Koehler, G. W. Bennett and M. E. Scharf (2007). "Correlation of cellulase gene expression and cellulolytic activity throughout the gut of the termite Reticulitermes flavipes." Gene **395**(1-2): 29-39.

Zverlov, V. V., J. Kellermann and W. H. Schwarz (2005). "Functional subgenomics of Clostridium thermocellum cellulosomal genes: identification of the major catalytic components in the extracellular complex and detection of three new enzymes." Proteomics **5**(14): 3646-53.

Zverlov, V. V., N. Schantz and W. H. Schwarz (2005). "A major new component in the cellulosome of Clostridium thermocellum is a processive endo-beta-1,4-glucanase producing cellotetraose." FEMS Microbiol Lett **249**(2): 353-8.

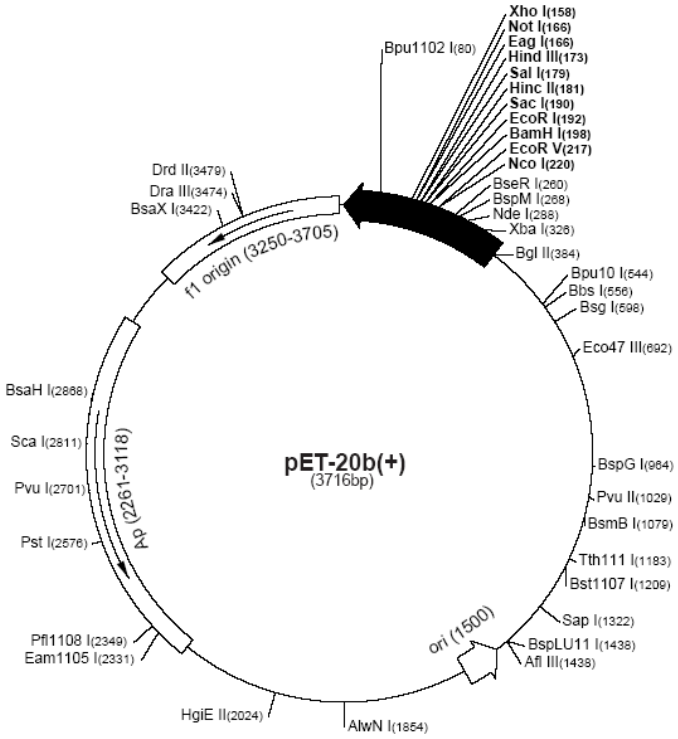
Zverlov, V. V., G. A. Velikodvorskaya and W. H. Schwarz (2003). "Two new cellulosome components encoded downstream of celI in the genome of Clostridium thermocellum: the non-processive endoglucanase CelN and the possibly structural protein CseP." Microbiology **149**(Pt 2): 515-24.

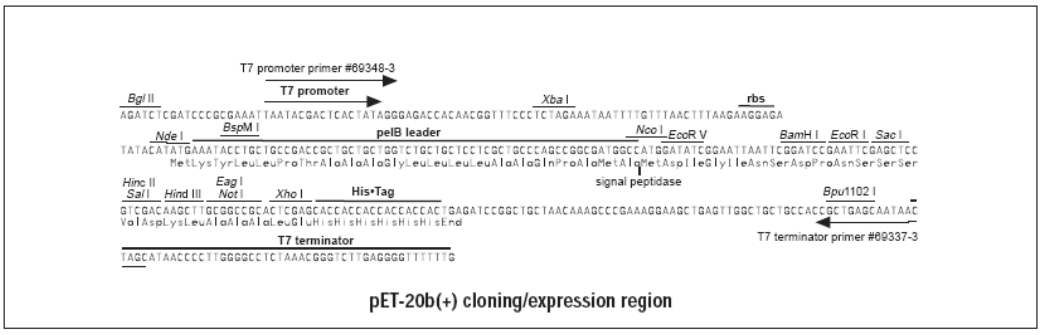
Zverlov, V. V., G. A. Velikodvorskaya, W. H. Schwarz, J. Kellermann and W. L. Staudenbauer (1999). "Duplicated Clostridium thermocellum cellobiohydrolase gene encoding cellulosomal subunits S3 and S5." Appl Microbiol Biotechnol **51**(6): 852-9.

Appendix A

pET20b

pET-20b(+) sequence landmarks	
T7 promoter	353-369
T7 transcription start	352
pelB coding sequence	224-289
Multiple cloning sites	
(*Nco* I - *Xho* I)	158-225
His•Tag coding sequence	140-157
T7 terminator	26-72
pBR322 origin	1500
bla coding sequence	2261-3118
f1 origin	3250-3705

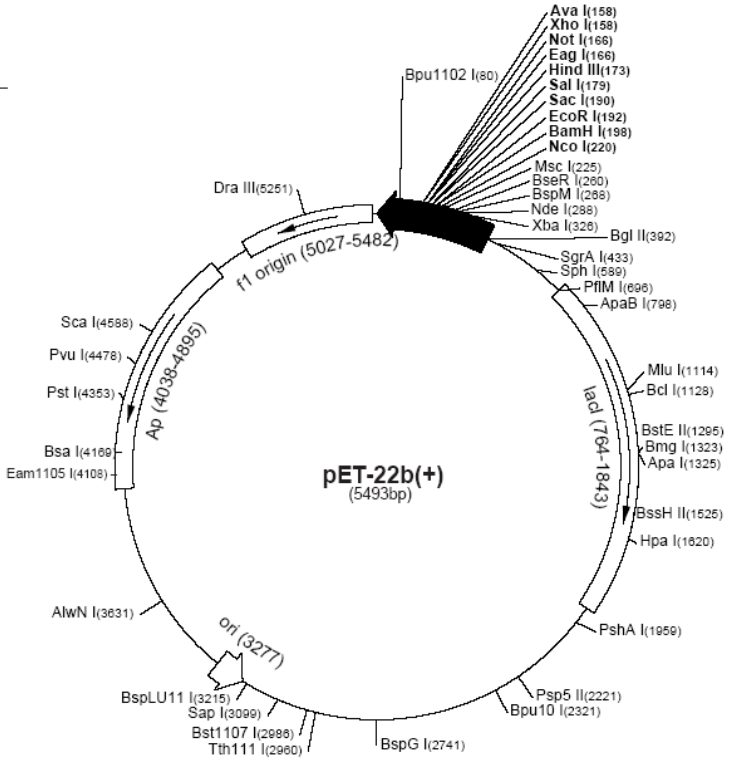




pET-20b(+) cloning/expression region

pET22b

pET-22b(+) sequence landmarks	
T7 promoter	361-377
T7 transcription start	360
pelB coding sequence	224-289
Multiple cloning sites	
(*Nco* I - *Xho* I)	158-225
His•Tag coding sequence	140-157
T7 terminator	26-72
lacI coding sequence	764-1843
pBR322 origin	3277
bla coding sequence	4038-4895
f1 origin	5027-5482

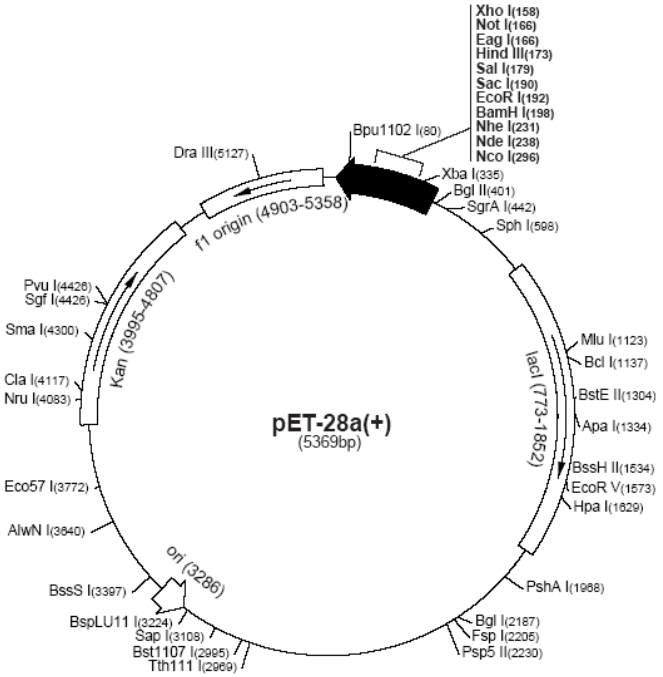


pET-22b(+) cloning/expression region

pET28b

pET-28a(+) sequence landmarks	
T7 promoter	370-386
T7 transcription start	369
His•Tag coding sequence	270-287
T7•Tag coding sequence	207-239
Multiple cloning sites	
(*Bam*H I - *Xho* I)	158-203
His•Tag coding sequence	140-157
T7 terminator	26-72
lacI coding sequence	773-1852
pBR322 origin	3286
Kan coding sequence	3995-4807
f1 origin	4903-5358

The maps for pET-28b(+) and pET-28c(+) are the same as pET-28a(+) (shown) with the following exceptions: pET-28b(+) is a 5368bp plasmid; subtract 1bp from each site beyond *Bam*H I at 198. pET-28c(+) is a 5367bp plasmid; subtract 2bp from each site beyond *Bam*H I at 198.

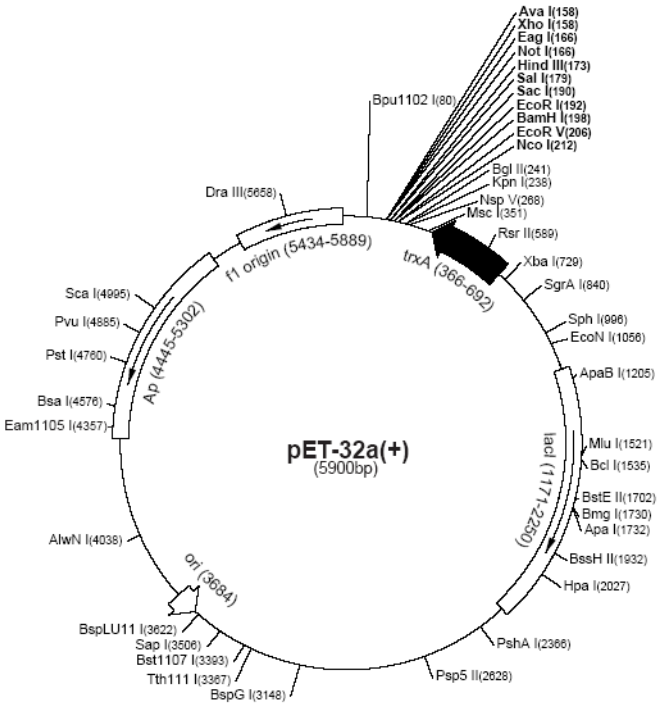


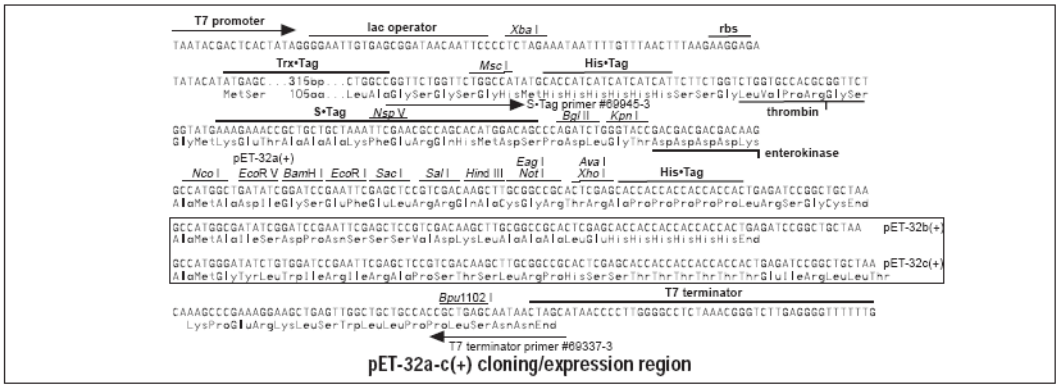
pET-28a-c(+) cloning/expression region

pET32b

pET-32a(+) sequence landmarks	
T7 promoter	764-780
T7 transcription start	763
Trx•Tag coding sequence	366-692
His•Tag coding sequence	327-344
S•Tag coding sequence	249-293
Multiple cloning sites	
(*Nco* I - *Xho* I)	158-217
His•Tag coding sequence	140-157
T7 terminator	26-72
lacI coding sequence	1171-2250
pBR322 origin	3684
bla coding sequence	4445-5302
f1 origin	5434-5889

The maps for pET-32b(+) and pET-32c(+) are the same as pET-32a(+) (shown) with the following exceptions: pET-32b(+) is a 5899bp plasmid; subtract 1bp from each site beyond *Bam*H I at 198. pET-32c(+) is a 5901bp plasmid; add 1bp to each site beyond *Bam*H I at 198 except for *Eco*R V, which cuts at 209.





pET-32a-c(+) cloning/expression region

Appendix B

Mutation	Primer
E221A	F-ggc ccc tgg cac gcc cat aac ggc gac
	R-gtc gcc gtt atg ggc gtg cca ggg gcg
E338A	F-gcc gcc ctg acc gcc acg ggc aat aat
	R-att att gcc ctg ggc ggt gac ggc ggc
L129G	F-g gga cac acc gcc aat gga gat gcg gtt aac ttt g
	R-c aaa gtt aac cgc atc tcc att ggc ggt gtg tcc c
L129A	F-g gga cac acc gcc aat gct gat gcg gtt aac ttt g
	R-c aaa gtt aac cgc atc agc att ggc ggt gtg tcc c
D130G	F-g gga cac acc gcc aat ctg gga gcg gtt aac ttt gaa aag
	R-ctt ttc aaa gtt aac cgc tcc cag att ggc ggt gtg tcc c
D130A	F-g gga cac acc gcc aat ctg gct gcg gtt aac ttt gaa aag
	R-ctt ttc aaa gtt aac cgc agc cag att ggc ggt gtg tcc c
L129G/D130G	F-gaa ttg gga cac acc gcc aat gga gga gcg gtt aac ttt gaa aag
	R-ctt ttc aaa gtt aac cgc tcc tcc att ggc ggt gtg cct aac ttc
L129A/D130A	F-gaa ttg gga cac acc gcc aat gca gca gcg gtt aac ttt gaa aag
	R-ctt ttc aaa gtt aac cgc tgc tgc att ggc ggt gtg cct aac ttc
ΔLD	F-gcg gtt aac ttt gaa aag atg cag ca
	R-aat ggc ggt gtg tcc caa ttc cag g
ΔNLD	F-gcg gtt aac ttt gaa aag atg cag ca
	R- ggc ggt gtg tcc caa ttc cag gcc
ΔNLDA	F- gtt aac ttt gaa aag atg cag cac tg
	R- ggc ggt gtg tcc caa ttc cag gcc

Table AB1 Primers used to generate *Cj*Man26C mutants

Mutation	Primer
E96A	F- gga atg gct acc aga gcc tcc act ttt aga gct
	R- agc tct aaa agt gga ggc tct ggt agc cat tcc
N188A	F- ggc ttg atg gga tac gcg aca ggt tgg att gac ggt gcg g
	R- c cgc acc gtc aat cca acc tgt cgc gta tcc cat caa gcc
N188D	F-ggc ttg atg gga tac gac aca ggt tgg att gac ggt gcg g
	R-c cgc acc gtc aat cca acc tgt gtc gta tcc cat caa gcc
N188E	F-ggc ttg atg gga tac gag aca ggt tgg att
	R- aat cca acc tgt ctc gta tcc cat caa gcc
S110A	F- acc gga agc ggg gct gcc cac gct ttt ggc cct
	R- agg gcc aaa agc gtg ggc agc ccc gct tcc ggt
S110D	F- acc gga agc ggg gct gat cac gct ttt ggc cct
	R- agg gcc aaa agc gtg atc agc ccc gct tcc ggt
S110E	F- acc gga agc ggg gct gag cac gct ttt ggc cct
	R- agg gcc aaa agc gtg ctc agc ccc gct tcc ggt

Table AB2 Primers used to generate *Ct*Cel119 mutants

Mutation	Primer
D21	F- gat gtg ttt acc gcc gcg ccg gcg gcc ctg gta cac aag ggc
	R- gcc ctt gtg tac cag ggc cgc cgg cgc ggc ggt aaa cac atc
E148	F- atc gat tgg gac gat ata gcc cct tcg gta ttt att gat gac
	R- gtc atc aat aaa tac cga agg ggc tat atc gtc cca atc gat
E195	F- ctg ccc gaa ttt acc gcc gcg atc tgg gta cat aaa tac cag g
	R- c ctg gta ttt atg tac cca gat cgc ggc ggt aaa ttc ggg cag
F46A	F- ccg gat aac acc acg gcc ttt gtg atg aac gaa
	R- ttc gtt cat cac aaa ggc cgt ggt gtt atc cgg
F47A	F- gat aac acc acg ttt gcc gtg atg aac gaa tgg
	R- cca ttc gtt cat cac ggc aaa cgt ggt gtt atc
D81A	F- acc tgg gcc aag ggc gcc gcc tgg gcc agc cag
	R- ctg gct ggc cca ggc ggc gcc ctt ggc cca ggt
W83A	F- gcc aag ggc gat gcc gcc gcc agc cag gtg att
	R- aat cac ctg gct ggc ggc ggc atc gcc ctt ggc
R100A	F- tac tgg tat gtg acc gtc gcc cac gac gac acc aag ccg ggt
	R- acc cgg ctt ggt gtc gtc gtg ggc gac ggt cac ata cca gta
F109A	F- gac acc aag ccg ggt gcc gcc att ggt gtg gcc
	R- ggc cac acc aat ggc ggc acc cgg ctt ggt gtc
W144A	F- gat acc cct atc gat gcc gac gat ata gat cct
	R- agg atc tat atc gtc ggc atc gat agg ggt atc
I147A	F- atc gat tgg gac gat gcc gat cct tcg gta ttt att
	R- aat aaa tac cga agg atc ggc atc gtc cca atc gat
N165A	F- tac ctg ttc tgg gga gcc acc agg cca cgc tac
	R- gta gcg tgg cct ggt ggc tcc cca gaa cag gta
T166A	F- tac ctg ttc tgg gga aat gca agg cca cgc tac gcc aag ctg
	R- cag ctt ggc gta gcg tgg cct tgc att tcc cca gaa cag gta
T166W	F- ctg ttc tgg gga aat tgg agg cca cgc tac gcc aag
	R- ctt ggc gta gcg tgg cct cca att tcc cca gaa cag
T194A	F- gaa ggg ctg ccc gaa ttt gca gag gcg atc tgg gta cat aaa
	R- ttt atg tac cca gat cgc ctc tgc aaa ttc ggg cag ccc ttc
T194W	F- ggg ctg ccc gaa ttt tgg gag gcg atc tgg gta cat
	R- atg tac cca gat cgc ctc cca aaa ttc ggg cag ccc
F214A	F- tcc tat gcc atg ggc gcc ccc gag aag atc ggc
	R- gcc gat ctt ctc ggg ggc gcc cat ggc ata gga
H247A	F- ggc aat acg ccc act aac gcc cag gcc att atc gaa ttc aat
	R- att gaa ttc gat aat ggc ctg ggc gtt agt ggg cgt att gcc
Y273A	F- ccc gat ggc ggc caa gcc cgt cgt tcg gtc agt
	R- act gac cga acg acg ggc ttg gcc gcc atc ggg
Q272A	F- cgc ccc gat ggc ggc gcc tac cgt cgt tcg gtc agt
	R- act gac cga acg acg gta ggc gcc gcc atc ggg gcg
R275A	F- gat ggc ggc caa tac cgt gcc tcg gtc agt atc gat gag ctg
	R- cag ctc atc gat act gac cga ggc acg gta ttg gcc gcc atc

Table AB3 Primers used to generate *Cj*Abf43B mutants

Appendix C

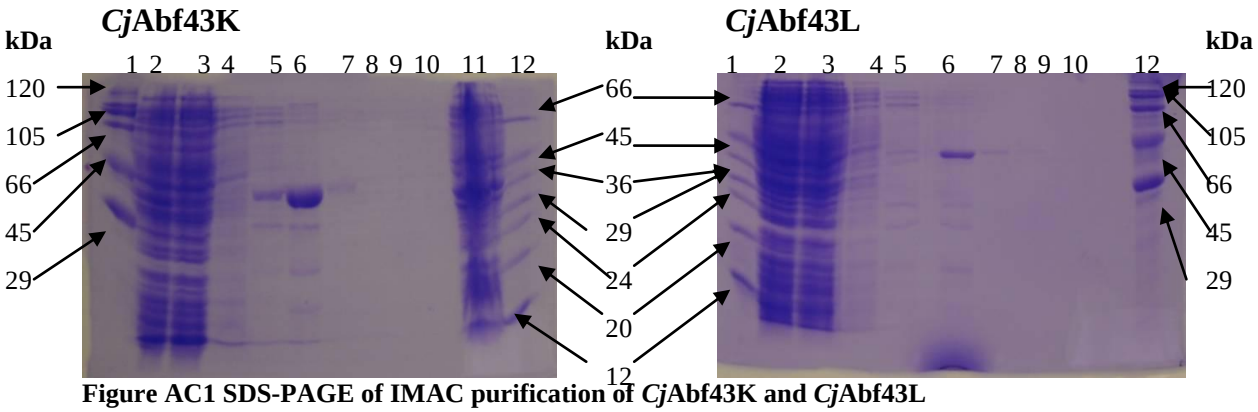
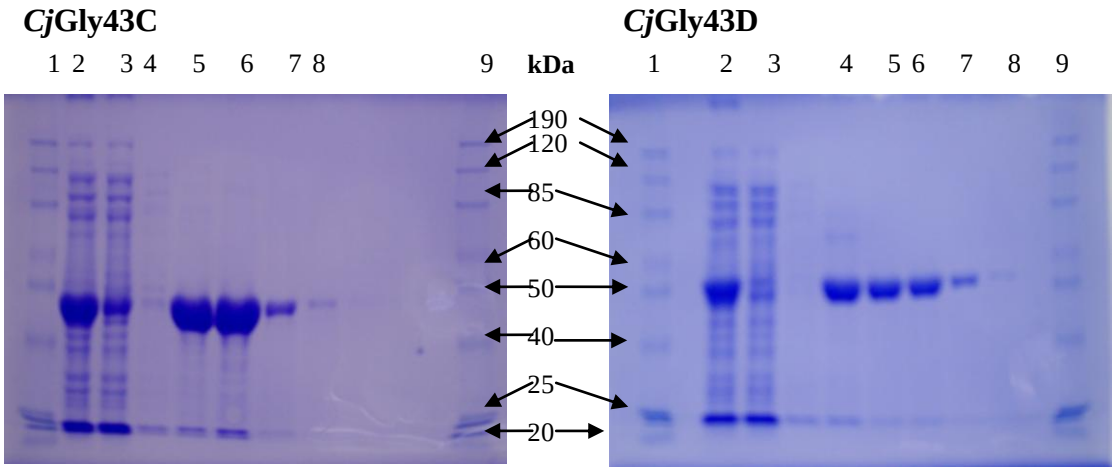


Figure AC1 SDS-PAGE of IMAC purification of *Cj*Abf43K and *Cj*Abf43L
1 high molecular weight marker; 2 cell free extract;3 Flow through; 4 Talon buffer (see MM for composition) wash; 5 Talon buffer supplemented with 10 mM imidazole; 6-10 Talon buffer supplemented with 100 mM imidazole; 11 Insoluble pellet; 12 Low molecular weight marker



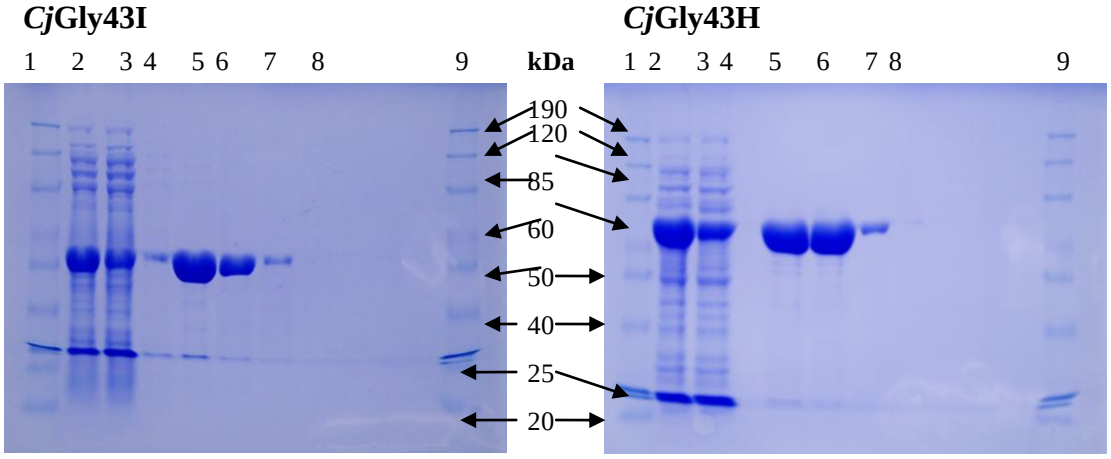


Figure AC2 SDS-PAGE of IMAC purification of *Cj*Gly43C, *Cj*Gly43D, *Cj*Gly43I and *Cj*Gly43H
1 and 9 Benchmark prestained protein ladder (invitrogen) ; 2 cell free extract;3 Flow through; 4 Talon buffer (see MM for composition) wash; 5 Talon buffer supplemented with 10 mM imidazole; 6-8 Talon buffer supplemented with 500 mM imidazole

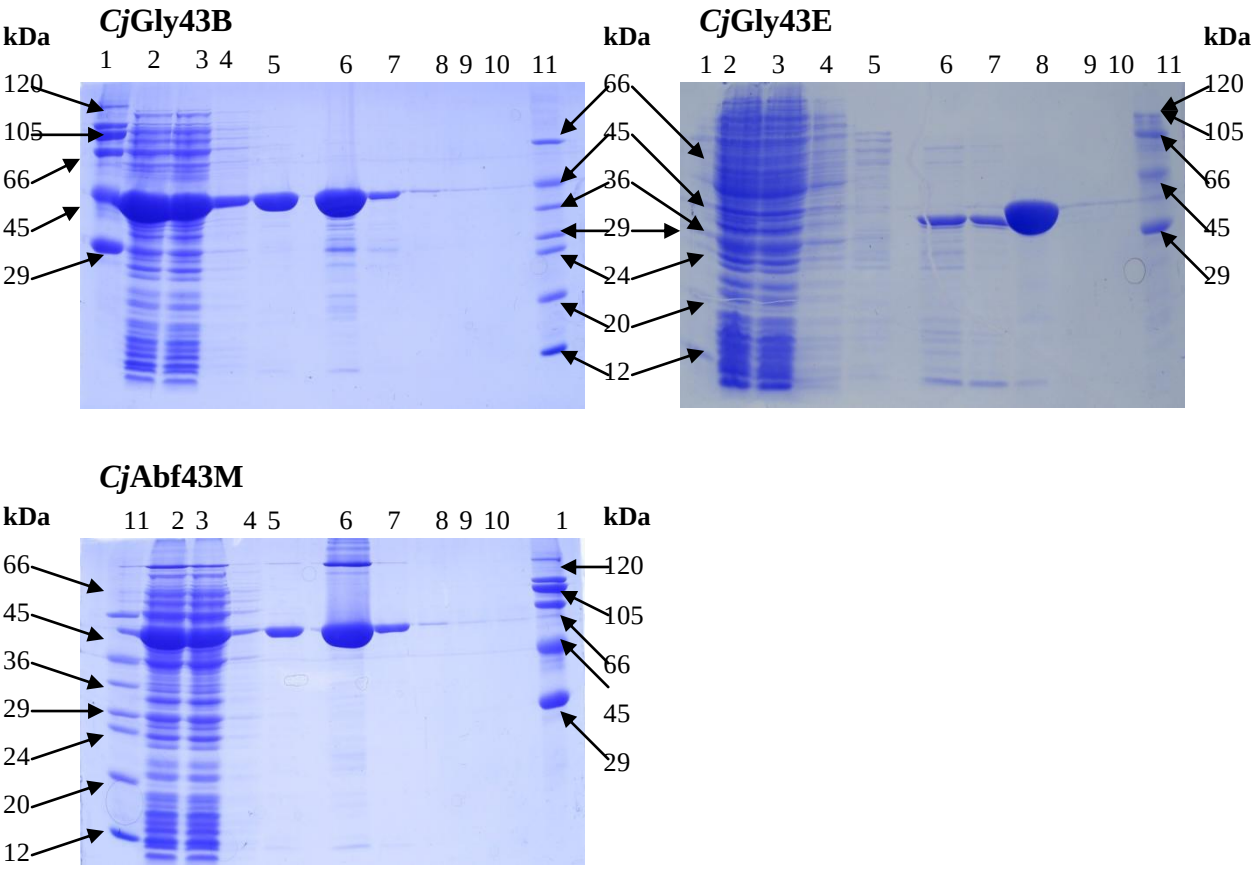


Figure AC3 SDS-PAGE of IMAC purification of *Cj*Gly43B, *Cj*Gly43E and *Cj*Abf43M
1 high molecular weight marker; 2 cell free extract;3 Flow through; 4 Talon buffer wash; 5 Talon buffer supplemented with 10 mM imidazole; 6-10 Talon buffer supplemented with 100 mM imidazole; 11 Insoluble pellet; 12 Low molecular weight marker

Protein	Molecular mass (kDa)	Extinction Coefficient (M^{-1} cm^{-1})
***Cj*Abf43K**	36	93070
***Cj*Abf43L**	38	77350
***Cj*Abf43M**	37	86875
***Cj*Gly43B**	62	141765
***Cj*Gly43C**	31	70250
***Cj*Gly43D**	36	96380
***Cj*Gly43E**	35	101090
***Cj*Gly43I**	60	149200
***Cj*Gly43H**	58	148280

Table AC1 Molecular masses and extinction coefficient of GH43 proteins

Appendix D

Low molecular weight protein standards (Sigma)	**M_r (kDa)**
Albumin, Bovine	66
Albumin, Egg	45
Glyceraldehyde-3-P Dehydrogenase	36
Carbonic Anhydrase, Bovine	29
Trypsinogen, Bovine	24
pancreasTrypsin inhibitor, Soybean	20
α-Lactalbumin, Bovine milk	14.2

High molecular weight protein standards (Sigma)	**M_r (kDa)**
Myosin, Rabbit Muscle	205
β-Galactosidase, *E coli*	116
Phosphorylase B, Rabbit Muscle	97.4
Albumin, Bovine	66
Albumin, Egg	45
Carbonic Anhydrase, Bovine Erythrocytes	29